JMIR Human Factors

Impact Factor (2023): 2.6
Volume 13 (2026) ISSN 2292-9495 Editor in Chief: Andre Kushniruk, BA, MSc, PhD, FACMI

Contents

Original Papers

Anesthesia Providers' Perspectives on the Redesigned Philips Acoustic Alarm System: Qualitative Pre- and Postimplementation Study (e82703)
Greta Gasciauskaite, Stefanie Senn, Cynthia Hunn, Christoph Nöthiger, David Tscholl, Tadzio Roche. 23

User Perceptions of a Large Language Model–Based Chatbot for Cognitive Reappraisal of Stress: Mixed Methods Feasibility Study (e89736)
Ananya Bhattacharjee, Jina Suh, Mohit Chandra, Javier Hernandez. 37

Redesign of Bedside Supply Carts to Improve Emergency Department Workflows: Mixed Methods Participatory Design (e80861)
Kat Hefter, Sammy Dreibelbis, Theresa Haupt, Amelia McIver, Agnes Wang, Julia Dwight, Ogechi Nwodim, Neil Ray. 54

Development of the ERATbi App, a Clinical Decision Support System for Early Recovery After Traumatic Brain Injury in the ICU: Usability Study (e79981)
Hsiao-Ching Yen, I-Hui Wu, Wei-Ling Hsiao, Sheng-Ru Lai, Chen-Hao Yang, Hsien-Chi Liao, Yin-Yi Han. 67

Overview of a User-Centered, Mixed-Methods Process for Designing Interconnected and Focused Mobile Applications on Patient Care Environment (InterFACE): Augmented-Reality Decision Support System for Pediatric Resuscitation (e78144)
Frederic Ehrler, Ana Rajic, Alexandre De Masi, Sharleen Olanka, Marco Generelli, Jennifer Davidson, Yiqun Lin, Kangsoo Kim, Pierre-Louis Rebours, Marc Ibrahim, Donovan Duncan, Ryan Kang, Sergio Manzano, Adam Cheng, Johan Siebert. 77

Designing App Interfaces to Elicit Specific Emotional Responses and Improve Attention and Short-Term Memory in Patients With Insomnia Undergoing Brief Cognitive Behavioral Therapy: Within-Subject Eye-Tracking Experimental Pilot Study (e79883)
Kuan-Chu Su, Hsiao-Yean Chiu, Ko-Chiu Wu, Chia-Chi Chang. 95

Reflections on the Inclusive Co-Design Process of a Virtual Assistant for Individuals With Complex Care Needs: Mixed Methods Study (e81941)
Julia van Calis, Kirsten Bevelander, Anneke van der Cruijsen, Jenneken Naaldenberg, Geraline Leusink. 114

The Development of a Web-Based Application to Support Informal Caregivers of Individuals With Head and Neck Cancer: Human-Centered Design Approach (e81896)
Awais Ahmad, Åsa Cajander, Ulrica Langegård, Birgitta Johansson, Anna Henriksson, Mona Pettersson, Waqas Ahmad, Ylva Tiblom Ehrsson.
1 3 0

XSL•FO
RenderX

Improving Usability of the Interrupting Prolonged Sitting With Activity Virtual Teacher Training Modules: Case Study (e83427)
Rebecca Hasson, Tiffany Tam, Lucy Myers, Anna Schwartz, Alissa Li. . . . 147

Personalizing Mobile Apps for Health Behavioral Change According to Personality: Cross-Sectional Validation of a Preference Matrix (e78939)
Laetitia Gosetto, Gilles Falquet, Frederic Ehrler. . . . 162

On the Design of a Sign Language Corpus of Medical Terms for Automatic Translation Systems: Mixed Methods Approach (e72789)
Milena Marcolino, Luiza Rosa, Elidea Bernardino, Mario Campos. . . . 180

Development, Implementation, and Preliminary Results of a Novel Social Network for Rare Disease Communities: Mixed Methods Study (e80230)
Lara Fendrich, Henriette Högl, Ralf Schmidt, David Bascom, Armin Berger, Max Lübbering, Gundula Ernst, Justus Schumacher, Sabrina Hölzer, Lorenz Grigull. . . . 192

A Mobile App (Tpro) for Symptom Management in Patients With Deep Vein Thrombosis Based on Patient-Reported Outcomes: Design and Development Using an Iterative Convergent Mixed Methods Approach (e92738)
Qiaodan Lu, Yafei Liu, Manna Shao, Lei Wang. . . . 212

An Online Epilepsy Self-Management Tool for Patients With Epilepsy: A Formative Usability Study (e79540)
Katarzyna Czerniak, Ross Shegog, Refugio Sepulveda, Robert Addy, Youngran Kim, Sahiti Myneni, Alejandra Garcia-Quintana, Kimberly Martin, David Labiner. . . . 226

Designing a Trauma-Informed Safety by Design Framework for Nonconsensual Intimate Image Removal Platforms: A 2-Phase Sequential Exploratory Study (e94438)
Ji-yeon Lee, Yejun Koh. . . . 249

Optimizing Usability of Digital Health Interventions for Nondigitally Native Adults: A Framework-Based Approach to Develop Just-in-Time Adaptive Interventions (JITAIs) and Other Adaptations (e87365)
Abby Cheng, Christine Gou, Adriana Martin, Sarah Hartz, Joanna Abraham. . . . 264

Personalized Intelligent Chatbot Based on AI-Generated Content Assists Memoir Writing for Older Adults With Cognitive Impairment: Mixed Methods Study (e83428)
Yibo Meng, Yuan Que, Zhe Yan, Bingyi Liu, Zixin Wang, Mandi Yang, Huidi Lu. . . . 283

User-Centered Virtual Clinic Services for Remote, Rural, and Underserved Sub-Saharan Africa: Development and Usability Study (e84272)
Abby Blocker, Ahmed Biyabani, Joyce Mwangama, Mohammed Datay, Bessie Malila. . . . 297

Mental Health Monitoring in Esports Athletes Using an mHealth App: Requirements Analysis and Usability Study (e90140)
Leona Stolberg, Daniel Hieber, Peter Kuhn, Michael Hebel, Johannes Schobel. . . . 316

Co-Designing a Digital Platform for the Management of Modifiable Risk Factors and Comorbidities in Patients With Atrial Fibrillation: Mixed Methods Usability Study (e95810)
Sofia Klavebäck, Frieder Braunschweig, Matthias Lidin, Emma Svennberg. . . . 330

Effect of Pharmacy Student Peer Supervision on the Accuracy of Admission Medication Reconciliation: Prospective Pre-Post Observational Study (e77486)
Basma Ammor, Morgane Masse, Anne Toulemonde, Laurine Cadart, Jean-Baptiste Beuscart, Marc Lambert, Pascal Odou, Bertrand Décaudin.
5 8 6

XSL•FO
RenderX

Health Information Technology–Related Loss of Central Surveillance Data in a Heart Intensive Care Unit: Multi-Framework Case Report (e92560)
Md Jabin. . . . 595

Dynamic Balance Control and Postural Adaptation in Human-Robot Collaborative Manipulation: Within-Subject Experimental Study (e79930)
Alessia de Nobile, Daniele Bibbo, Simone Ranaldi, Maurizio Schmid, Giovanni Corvini, Silvia Conforto. . . . 614

Prevalence of Work-Related Pain or Discomfort Among Urologists in the State of Florida: Results From the Florida Urologic Society Task Force on Ergonomic Challenges Experienced by Its Members (e88848)
Justin David, Austin Hill, Jared Schommer, Eric Regele, Colleen Ball, Sanoj Punnen, Joseph Costa, Adam Ball, Jamin Brahmbhatt, David Thiel, Raymond Pak, Ram Pathak. . . . 630

C8 Health, a Platform for the Implementation of Best Practices: Survey-Based Usability Study (e83547)
Faria Nisar, Nicolas D Alessandro, Jessica Suratkal, Ido Zamberg, Samantha Pope, Ali Ali, Luis Tollinche. . . . 636

Validation of the Taiwan Chinese Version of the Assistive Technology Usability Questionnaire for People With Neurological Diseases for Wearable Robotic Exoskeletons: Usability Study (e89556)
Hander Wang, Jian-Jia Huang, Yu-Cheng Pei, Huey-Wen Liang. . . . 643

Utility of Augmented Reality Glasses With Waveguide Optics and Facial Recognition Technology for Patient Verification in a Simulated Outpatient Setting: Physician Questionnaire Survey Study (e95664)
Ryo Ogasawara, Soichiro Yoshida, Kentaro Okamoto, Eiki Nagaoka, Hirotaka Sagawa, Kazutaka Sumita, Hajime Tanaka, Satoshi Shiraga, Tomohiro Amemiya, Yasuhisa Fujii. . . . 655

Feasibility and Preferences to Adopt mHealth-Based Interventions for HIV Prevention Among High-Risk Groups: Cross-Sectional Study (e81111)
Faizur Rehman, Muhammad Shahid, Mehran Riaz, Malik Umair, Farah Azhar, Adeel Siddiqui, Ali Ahmed. . . . 665

Evaluation of a Contactless Sleep Monitoring Device for Sleep Stage Detection at Home in a Healthy Population: Prospective Study in Free-Living Conditions (e77033)
Marie-Ange Stefanos, Guillaume de Laboulaye, David Campo, Martin de Gourcuff, Pierre Escourrou, Boris Matrot, Anna Islind, Pierre Geoffroy.
6 8 3

Designing for Knowledge Generalization in Medical Device Instructions Among Health Care Professionals: Qualitative Interview and Observation Study (e82405)
Qing Xia, Jeremy Opie, Tom Brookshaw, Eloy Barbarin, Ann Blandford, Clare Selden. . . . 708

Assistive Technologies for Pulse-Based Tele-Practice in Traditional Chinese Medicine: Two-Phase, Dual-Method Feasibility and Pilot Study (e77740)
Jia Ruan, Shucheng Chen, Minru Wu, Shuhan Li, Yuan Ho, Ka Fan, Wing Chow, Leung Chiu, Wing Yeung. . . . 723

Wearable Device Photoplethysmography As a Viable Tool to Longitudinally Monitor Vasoconstriction Biomarkers for Predicting Vaso-Occlusive Crisis in Sickle Cell Disease: Feasibility and Validation Study (e75465)
Payal Shah, Sabrina Sy, Mingjing Chen, Michael Khoo, Thomas Coates, Saranya Veluswamy. . . . 737

Opportunities for Improved Device Design Based on Central Line Placement Practices: Contextual Inquiry Study (e84621)
Mary Privitera, Sameer Khan, Bilal Irfan, Shayan Ali, Cecelia Arredondo, Kyrsten Sanderson, Jordan Bonomo. . . . 750

An Early-Stage Digital Therapeutic Intervention to Enhance Affective Response During Physical Activity Among Adults With Overweight or Obesity: Benchmark-Driven Formative Testing Study (e71472)
Rachel Crosley-Lyons, Lori Hatzinger, Micaela Hewus, Wei-Lin Wang, Delfien Van Dyck, Jimi Huh, Eric Hekler, Genevieve Dunton. . . . 766

XSL•FO
RenderX

A Summative Usability Evaluation of an Infusion Pump Through Simulation-Based Testing With Nurses: Mixed Approach Study (e86443)
Yingjun Wan, Yun Zhang, Yunlong Zhao, Han Yu, Kaifeng Liu. . . . 783

A Comprehensive Self-Medication Management System to Reduce Medication Errors Among People Living With Dementia: Mixed Methods Feasibility and Usability Study (e86828)
Santosh Basapur, Charles Gellman, Jamie Plenge, Amelia Troutman, Erin Yurko, Brandon Woolsey, Justin McClendon, Lisa Marceau, Neelum Aggarwal. . . . 798

Exploring Factors Influencing Nursing Task Prioritization for Supportive Information System Design: Qualitative Study With Thematic Analysis (e89940)
Kodai Iwamoto, Mayumi Toyama, Goshiro Yamamoto, Chang Liu, Kazumasa Kishimoto, Yukiko Mori, Tomohiro Kuroda. . . . 810

Health Care Professionals' Experience in a Digitally Supported Hospital-at-Home Program: Cross-Sectional Study (e106615)
Carlos Bibiano Guillén, Eva Arroyo Urea, Olga Estremera López, Sonia López Lallave, María Aranda Higuero, Cristina Ubeda García, Silvia Sánchez Sierra, Almudena Calvo Cruz, María Mir Montero. . . . 821

Contextual Assessments for Chronic Obstructive Pulmonary Disease Transition of Care Bundle Implementation Planning for the Reduce REVISITS Study: Rapid Sequential Explanatory Mixed Methods Approach (e82078)
Mahima Akula, Kim Erwin, Leah Traeger, Hannah Pick, Fei Gao, Laura Damschroder, Valerie Press. . . . 831

Improving the Efficiency of the Intravenous Medicine Preparation Pathway With an Intravenous Workflow Software Solution in Full-Capacity Pharmacy Units at Watford General Hospital: Observational Study and Economic Analysis (e85408)
Bethany Umpleby, Nick Hex, Teresa Martins, Husanain Soori. . . . 847

Organizational Tensions in the Implementation of Modifiable Off-the-Shelf Technologies in a University Hospital: Qualitative Multimethod Study (e84841)
Alessia Nowak, Carolina Gaul, Elena Hinz, Christine Knoll, Begüm Kül, Anja Wels, Julia Berkmann, Cathrin de Pasquale, Gloria Kremser, Daniel Fürstenau, Akira-Sebastian Poncette, Felix Balzer, Lina Mosch. . . . 858

Developing mHealth IT for Older Adult Medication Safety: Remote Participatory Co-Design Using the RAPID Method (e82366)
Jordan Hill, Aaron Ganci, Noll Campbell, Andrew Pickett, Michelle Chui, Ephrem Abebe, Richard Holden. . . . 918

Surface Electromyography as a Method for Characterizing Mammogram Discomfort: Cross-Sectional Questionnaire Study of Procedural Stress (e83971)
Krystyna Gielo-Perczak, Riley McNaboe, Hugo Posada-Quintero. . . . 931

Development of a Data-Enabled Mixed Method for Designing for Patients From a Human-Centered Design Perspective: Explorative Case Study (e87677)
Yingtao Sun, Jiwon Jung. . . . 950

What Users Say About Reimbursable Digital Therapeutics in Germany: Large-Scale App Store Review Analysis Using a Large Language Model (e92415)
Benjamin Kinast, Henrik Rohde, Björn Schreiweis, Hannes Ulrich. . . . 970

Smartphone-Based Ecological Momentary Assessment to Monitor Opioid Use and Overdose Among People Who Use Opioids: Prospective Observational Feasibility Study (e95655)
Kamal Gautam, Kiran Paudel, Jeffrey Wickersham, Antoine Khati, Anushka Thapa, Sandesh Bhusal, Md Sujan, Sherry Pagoto, Toan Ha, Roman Shrestha. . . . 984

XSL•FO
RenderX

Expert Views on Criteria for Evaluation of Human Factors Methods: Qualitative Interview Study (e73324)
Selvana Awad, Rachel Begg, Thomas Loveday, Andrew Baillie, Melissa Baysari. 998

Interest in and Predictors of Engagement With a Virtual Reality Intervention Among People With Chronic Pain: Cross-Sectional Survey Study (e70598)
Genevieve Bryant, Samuel Holzman, Hector Perez. 1007

Consensus on Core Competencies of Undergraduate Life Sciences Students for Entry-Level Participation in Clinical Trials: RAND-Modified Delphi Study (e89449)
Ángela Cano, Patricia Cosano-Pérez, Juan Castón, Diana Corona-Mata, Rosa Ballesteros-Olmedo, Eva Venegas-Moreno, Jaime Perea-Cortés, Concepcion Aranda-Valera, Isabel Machuca, Julian Torre-Cisneros. 1015

Change Adaptation Resilience Evaluation Questionnaire for Industry 5.0 Among Workers in High-Tech Manufacturing Companies: Questionnaire Development and Pilot Evaluation Study (e83754)
Giulia Bassi, Angelo Valente, Serena Tassoni, Silvia Salcuni. 1025

Feasibility of the “OA Coach” Mobile App to Support Individuals with Osteoarthritis: Development and Usability Testing (e84857)
Vicky Duong, Naomi Bloul, Jocelyn Bowden, Karen Bracken, Kate Bryce, Leticia Deveza, Jillian Eyles, Abdolhay Farivar, Robin Huang, Sarah Kobayashi, Na Liu, Carin Pratt, Charlotte Strong, Venkatesha Venkatesha, Shirley Yu, David Hunter. 1050

Telemedicine Service Experience Questionnaire for Chinese Outpatients: Development and Validation Study (e60551)
Peicheng Wang, Kun Zhang, Yanhua Chen, Minan Zhao, Peiyao Li, Yi Kuang, Yanrong Du, Weiguo Zhu, Xiao Long, Leiyu Shi, Jiming Zhu. . .
1 0 6 3

Design and Performance of an Email-Based Patient Recruitment Campaign in Primary Care Research: Formative Secondary Analysis (e67088)
Vanessa Vaillancourt, Marie-Dominique Poirier, Amélie Fournier, Sabrina Wong, Marie-Eve Poitras. 1076

Usability Across 3 mHealth Problem-Solving Training Interventions for Diverse Neurodevelopmental and Neurological Populations: Multicase Usability Evaluation (e90379)
David Ogundairo, Shannon Juengst, Avani Modi, Shari Wade, Matthew Schmidt. 1086

Determinants of Adoption of a Mechanical Ventilation Dashboard in Intensive Care: Measurement Instrument for Determinants of Innovations–Based, Single-Center, Cross-Sectional Evaluation (e92313)
Sarah Mugge, Tobias van Leijsen, Merijn Reuland, Annelies Visser, Frederique Paulus, Dave Dongelmans. 1112

AI-Based Rehabilitation System for Congenital Muscular Torticollis: Formative Usability Evaluation (e101427)
Seojin Hong, Heewon Nam, Seongwoo Jeong, Hyun Choi, Hyosun Kweon. 1125

A Geospatially Customizable Culturally Tailored Just-in-Time Adaptive Intervention for Violence-Affected People Living With HIV: Mixed Methods Acceptability, Feasibility, and User Experience Study (e80705)
Simone Skeen, Stephanie Tokarz, Rayna Gasik, Ethan Smith, Katherine Theall, Gretchen Clum. 1141

Smartphone App Using Reinforcement Learning for Obesity: Single-Arm Feasibility Study (e77323)
Ken Kurisu, Yoshiharu Yamamoto, Tomohisa Aoyama, Toshimasa Yamauchi, Kazuhiro Yoshiuchi. 1161

Postural Education in School-Aged Populations: Development and Usability Evaluation of a Mobile Biofeedback App (EduBack) (e79282)
Marco García-Luna, Miguel García-Jaén, Daniel Ruiz-Fernández, Carmen Manchado, Juan Cortell-Tormo. 1172

Real-World Evaluation of the Eye+Dot Online Triage Support Tool in Community Optometry Practices: Mixed Methods Evaluation Study (e77278)
Jen Lim, Sarah Farrell, Samuel Newlands, Louise Allen. 1186

XSL•FO
RenderX

Aspects Supporting and Hindering Type 2 Diabetes Self-Management in Web-Based Educational Portals: Usability Testing Study With Updated Framework in Razavi-Khorasan, Iran (e78903)
Javad Jafari, Klas Karlgren, Hossein Karimi Moonaghi, Mohammad Mazaheri Habibi, Saeid Eslami, Italo Masiello, Stefano Bonacina. 1199

Impact of Rural Trauma Team Development Education on Prehospital Time, Referral-to-Dispatch Interval, and Neurological and Musculoskeletal Injury Outcomes: Cluster Randomized Controlled Trial (e82591)
Herman Lule, Micheal Mugerwa, Anne Abio, Benson Oguttu, Andrew Kakeeto, Fiona Walsh, Hervé Lekuya, Robinson Ssebuufu, Patrick Kyamanywa, Andreas Deckert, Till Bärnighausen, Jussi Posti, Michael Wilson. 1217

Quality of Informed Consent and Interface Usability in Primary Care e-Consultation: Cross-Sectional Study (e78483)
Caitlin Parfitt-Ford, Lisa Ballard, Adriane Chapman. 1244

Grip Strength Estimation Using Input Data From a Commodity Smartphone: Model Development and Validation Study (e90316)
Komei Tajima, Kaori Ikematsu, Toshiya Isomoto, Kunihiro Kato, Yuta Sugiura. 1258

Technology Acceptance of 2 mHealth Apps During an Observational Technology Evaluation Study: Qualitative Results of a Mixed Methods Study (e70873)
Johanna Graeber, Elke Warmerdam, Svenja Aufenberg, Christopher Bull, Kristen Davies, Juliane Döhring, Kirsten Emmert, Claire Judd, Corina Maetzler, Nikolay Manyakov, Ralf Reilmann, Wan-Fai Ng, Victoria Macrae, Walter Maetzler, Hanna Kaduszkiewicz. 1271

A Smartphone App to Improve Intuitive Eating and Diet Quality: Design and Usability Study (e81439)
Mandy Korpusik, Delaram YazdanSepas, Hawley Almstedt. 1284

Examining Telemedicine Adoption After the COVID-19 Pandemic and Evaluating the Impact of a Virtual Hospital on Clinician Engagement: Longitudinal Survey Study (e93911)
Galia Barkai, Shany Edelman-Yaron, Hadar Amir, Havi Inbar-Lankry, Rachel Sarafraz. 1297

Usability and Workflow Integration of a Machine Learning–Derived Neonatal Risk Predictor in Kenyan Neonatal Units: Multisite User-Centered Pilot Evaluation (e94828)
Ronald Nyatuka, Paul Macharia, Esther Kakhata, Faith Siva, Betsy Muriithi, Md Jabin. 1305

Patient Perspectives on a Symptom Assessment App in Emergency Care: Interview Study Among Patients in the AkuSym Trial (e101285)
Claudia Schorr, Malte Schmieding, Doreen Kuschick, Felix Balzer, Myrto Bolanaki, Christoph Heintze, Marvin Kopka, Martin Möckel, Anna Slagman, Alica Thissen, Cornelia Wäscher, Konrad Schmidt, Hendrik Napierala. 1321

Workforce Perceptions of a Conceptual Hospital-at-Home Digital Twin: Qualitative Study (e96645)
Faiza Yahya, Matthew Cooper, Lauren Lawson, Hamde Nazar. 1336

Feasibility and Acceptability of the SISKA System for Continuous Student Emotional and Mental Health Guidance: Mixed Methods Study (e87647)
Tantut Susanto, Heni Windarwati, Niken Ati, Rizkiyani Istifada, Liasari Armaijn, Andi Pranata, Hanik Nihayati, Meidiana Dwidiyanti, Thika Marliana, Ermanto Fahamsyah, Slamin, Yunita Arihandayani. 1353

Insights on Designing the MINT (Micro-Interventions for Navigating Tension) Chatbot System for Stress Management: Collaborative Autoethnography (e89876)
Marco Bolpagni, Sofia Piffer, Marco Pozza, Munesu Maminimini, Silvia Gabrielli. 1371

Patient and Designer Collaboration in Cocreating Technical Innovations in a Hospital-Based Makerspace: Qualitative Interview Study (e85926)
Jiske van der Giessen, Onno Helder, Thijs van Houwelingen. 1387

XSL•FO
RenderX

Perspectives and Needs Regarding Remote Monitoring Technologies Among South Asian Individuals Living With Long-Term Conditions in the United Kingdom: Semistructured Interview and Focus Group Study (e82333)
Syed Ali, Yumna Masood, Karen Staniland, William Dixon, Sabine van der Veer, Caroline Sanders. 1399

Digital Assistive Technology Acceptance and Use by Caregivers of Older Adults With Cognitive Impairment: Qualitative Interview Study (e80614)
Xingyue Su, Honglin Chen, Pasi Karjalainen. 1412

Exploring Patients' Experiences in a Blended Transdiagnostic Group Treatment: Qualitative Study (e86016)
Noelia Jiménez-Orenga, Alba Fadrique-Jiménez, Azucena García-Palacios, Jorge Osma, Amanda Díaz-García, Juana Bretón-López. 1425

User Perspectives on Conversational Agents for Preventive Alcohol Self-Help: Qualitative Exploratory Study (e86752)
Marissa De Vries, Melissa Begeman, Marloes Derksen, Anna Goudriaan, Matthijs Blankers. 1451

Video-Algorithmic Patient Monitoring in Mental Health Inpatient Settings: Qualitative Study of Patient or Consumer, Clinician, and Vendor Perspectives (e85518)
Piers Gooding, Hamilton Kennedy, Simon D'Alfonso, Timothy Kariotis, Catherine Daniel, Bridget Hamilton. 1461

Feasibility of a Mobile-Based Home Monitoring System for Patients With Heart Failure: Mixed Methods Pilot Study (e84309)
Finna Indriany, Kemal Siregar, Bambang Siswanto, Budhi Purwowiyoto, Rarsari Soerarso, Indrajani Sutedja, Benny Setiawan, Hendy Wijaya, Raymond Wong, Salvatore Di Somma, Peter MacDonald. 1477

Using Theory-Based Frameworks to Identify Barriers, Facilitators, Expectations, and Willingness to Pay for Online Medical Consultation Services in China: Qualitative Study (e96448)
Shujie Dong, Qiushi Cai, Jingyi Ye, Yanya Chen, Rongsheng Zhao, Wai-kit Ming. 1488

Evaluation of Wellby, a Cocreated Mobile App and Wearable to Support Stress Management and Overall Well-Being: Mixed Methods Acceptability and Usability Study (e79381)
Justin Laiti, Sina Javadpour, Jenna Mullen, Claire Conway, Elaine Byrne, Pádraic Dunne. 1504

Acuity, Diagnosis, and Community Factors in Virtual and In-Person Treatment: Qualitative Study (e82863)
Brianna Cerrito, Alexa Connors, Amanda Fialk, Jamie Xiao, Josephine Usow, Frank Buono. 1524

The Role of Explanations in AI-Generated Alerts: Qualitative Study of Clinical Views on Explainable AI in Predictive Tools (e81460)
Jessica Rahman, Alana Delaforce, DanaKai Bradford, Jane Li, Farah Magrabi, David Cook, Aida Brankovic. 1535

Challenges and Strategies of Managing Mental Health Medication in Real-World Practices: Interview Study Among University Students (e67579)
Jiachen Li, Justin Steinberg, Elizabeth Mynatt, Varun Mishra. 1552

Health Care Professionals' and Patients' Perceptions and Experiences of Who Uses Video Consultations, and Why: Qualitative Study (e68658)
Irene Muli, Åsa Cajander, Nadia Davoody, Lovisa Hagstedt, Helena Hvitfeldt, Maria Hägglund, Marina Taloyan. 1569

Enhancing Data-Driven Decision-Making in HIV Care With Viral Load and Early Infant Diagnosis Data Dashboards in Côte d'Ivoire: Qualitative Study (e76550)
Yao He, Yves-Rolland Kouabenan, Paul Assoa, Nancy Puttkammer, Stephen Gloyd, Noah Hoffman, Bradley Wagenaar, Casey Iiams-Hauser, Pascal Komena, N'zi Kamelan, Adama Pongathie, Jan Flowers, Nadine Abiola, Natacha Kohemun, Jean-Bernard Amani, Christiane Adje-Toure, Lucy Perrone. 1584

XSL•FO
RenderX

Patients' and Health Care Professionals' Experiences of a Digital Self-Management System for Asthma: Qualitative Study (e79866)
Lovisa Jäderlund Hagstedt, Helena Hvitfeldt, Sara Riggare, Göran Petersson, Nadia Davoody, Maria Hägglund. 1596

Using the Theoretical Domains Framework to Identify Barriers to and Enablers of Patient Telemedicine Services Use in China: Qualitative Study (e78457)
Ke Liu, Yuting Yang, Zixuan Song, Huixian Li, Yanli Lyu, Ke Zhang, Xinxia Wu, Zheng Hou, Yipei Wang. 1607

Hospital-at-Home for South Asian Communities in British Columbia, Canada: Qualitative Interview Study (e79675)
Emma Wong, Mahabhir Kandola, Kamal Arora, Harroop Sharda, Roman Deol, Mary Jung, Robert Paquin, Maria Montenegro, Megan MacPherson. 1621

Bridging Gaps in Women's Heart Health: User-Centered Needs Assessment Informed by Patient and Clinician Interviews (e82916)
Christine Jacob, Sangeetha-Rose Puthanveettil, Patrick Vavken, Emel Kaplan, Christine Zuern. 1635

Exploring Video Consultations Across the Public and Private Sectors in Norway: Semistructured Interview Study (e80812)
Mari Skoge, Sofie Aminoff, Henrik Ihler, Kari Kværner, Linn Støme, Kristin Romm. 1654

Perspectives of Frontline Clinicians and End-Line Users on Smartphone-Based Photography for Assessing Traumatic Dental Injuries: Focus Group Interview Study and Thematic Analysis (e82668)
Emily Schultz, Boyen Huang, Margaret Shenouda, Mohamed Estai, Sarbin Ranjitkar, Jeffrey Louie, Patimaporn Pungchanchaikul. 1670

Evaluation of the Implementation of a Mobile Health App to Support Dutch Primary Care for Diabetes: Qualitative Study (e54431)
Liselot van den Berg, Lisenka te Lindert, Jiska Aardoom, Anke Versluis, Sofie Willems, Niels Chavannes, Marise Kasteleyn. 1686

Facilitators and Barriers to Over-the-Counter Hearing Aid Use in People With Dementia: Semistructured Interview Study (e83857)
Dana Urbanski, Alexander Hungs, Peggy Nelson, Joseph Gaugler. 1701

e-Mental Health Intervention Preferences Among Informal Caregivers of Adults Living with Chronic Kidney Disease: Semistructured Interview Study (e80962)
Chelsea Coumoundouros, Abla Sami, Paul Farrand, Robbert Sanderman, Louise von Essen, Joanne Woodford. 1723

A Chronic Pain Self-Management Mobile App (Dolodoc): Cross-Sectional Acceptability Study (e77163)
Julie Guebey, Laetitia Gosetto, Benno Rehberg, Christian Lovis, Frederic Ehrler, Aude Molinard-Chenu. 1738

Applying the Information-Motivation-Behavioral Skills Model to a Virtual Reality–Assisted Cognitive Health Education Intervention to Promote Screening Uptake: Qualitative Study (e91657)
Tingting Yin, Yulei Song, Bingqing Yang, Jingjing Yan, Yan Wang, Haiyan Yin, Xueqing Zhang, Guihua Xu, Yamei Bai. 1749

Stakeholder Perspectives on Information About Digital Health Applications in German Outpatient Mental Health Care: Qualitative Interview Study (e82514)
Klemens Höfer, Pascal Raszke, Jürgen Wasem, Felix Plescher, Anna Bußmann, Sarah Schlierenkamp, Michael Minor, Luisa Friedrich, Carsten Volland, Sophia Zander, Anja Wadeck, Udo Schneider, Josepha Katzmann, Enno Maaß, Carina Abels. 1763

Reducing Information Uncertainty in Hospital Wayfinding for People With Specific Needs: Findings From a Qualitative, Multimethod Study (e106882)
Ladislava Zbiejczuk Suchá, Kate ina Jochecová, Radek Pavlí ek, Petr ervenka, Karel Sobol, Henrieta Greplová, Gabriela Godišková, Alica Kúdelová, Pavla Pospíšilová, Ond ej Kvarda, Natálie Ká ová, Michaela Holubec Birtusová. 1775

XSL•FO
RenderX

Dynamic Augmented Reality Cues for Telementoring in Laparoscopic Surgery: Usability Study (e86566)
Jhasketan Padhan, Hawa Hamza, Dehlela Shabir, Julien Abinahed, Panagiotis Tsiamyrtzis, Zhigang Deng, Nikhil Navkar. 1793

An Automated Curriculum to Support Behavioral Health Counseling Among Pediatric Residents: Usability Study (e77438)
Liam Fleck, Rachel Herbst, Andrea Meisman, Thomas Talbot, Michael Glisson, Max Remington, Micah Goldson, Francis Real. 1806

Evaluating the Viability of Virtual Reality for Children's Food Choice Research: Comparative Mixed Methods Study (e88476)
Deepti Aggarwal, Thuong Hoang, Sze-Yen Tan, Mohammadreza Mohebbi, Catherine Russell. 1814

Challenges and Benefits of Virtual Reality in Home-Based Poststroke Rehabilitation: Co-Design Qualitative Study (e78798)
Trust Saidi, Ann Hestetun-Mandrup, Nenad Pavel, Ingvild Rosseland, Kathinka Selmer-Olsen, Nora Mørk, Åse Bergheim, Minna Pikkarainen.
1 8 3 7

Multimodal Virtual Reality Assessment of Medication Effects in Attention-Deficit/Hyperactivity Disorder and Its Distinction From Depression: Cross-Sectional Study (e85351)
Laura Asché, Julian Pakos, Hannah Schrage, Johanna Schuder, Luisa Jung, Dario Sanchez, Benjamin Selaskowski, Annika Wiebe, Alexandra Philipsen, Niclas Braun. 1856

Development and Validation of the Large Language Model Privacy Concern Assessment Questionnaire: Psychometric Evaluation Study on a Brief Instrument for Assessing User Privacy Concerns With Large Language Model Applications (e102910)
Leming Zhou, Hongtao Wang, Yi Xu, I Setiawan, Haomin Hu, Andi Saptono, Yong Choi, Bambang Parmanto. 1877

Mode Effects on Self-Reported Health Across Web-Based, Interviewer-Administered, and Nurse-Administered Modes: Population-Based Study (e89896)
Yunlong Liang, Tarek Al Baghal, Jonathan Burton, Meena Kumari. 1897

Japanese Health Information Technology Usability Evaluation Scale for Sexually Transmitted Infection–Related Chatbots: Development and Psychometric Validation Study (e90483)
Tomoko Hato, Hirono Ishikawa, Kense Todo, Keisuke Harada, Atsushi Yoshikawa, Yoshiharu Fukuda, Rebecca Schnall. 1911

Monitoring Health Status: Development and Preliminary Validation of a Personal Health Index Using the International Classification of Functioning, Disability and Health (e84802)
Ilkka Rautiainen, Lauri Parviainen, Veera Jakoaho, Sami Äyrämö, Jukka-Pekka Kauppi. 1924

Development and Psychometric Validation of a Usability Instrument Based on ISO 25010 for Electronic Health Record Systems in Peruvian Health Care Settings: Methodological Study (e81377)
David Mamani-Pari, Alex Peñaloza-Hualpa, Milton Humpiri-Flores, Eder Gutierrez-Quispe, Jorge Otazu-Luque. 1946

"Helpful in Thinking Through the Question": Redesigning the Patient Health Questionnaire-8 for Depression Assessment to Enhance Patient Engagement and Understanding (e76230)
Hannah Dorris, Jenna Sheldon, Courtney Sepucha, Emmanuel Makau, Pape Traoré, Elizabeth Murnane, Nicholas Jacobson. 1960

Mode Effects Between Telephone and Web Interviews in the Post-COVID-19 Questionnaire Survey CoVerlauf: Exploratory Study (e80631)
Paula Herrera-Espejel, Hermann Pohlabeln, Lisa Kühne, Stefan Rach. 1976

Chatbot Usability Scale in Chinese Users: Cross-Cultural Adaptation and Validation Study (e84971)
Haoming Ma, Runyuan Pei, Sijia Li, Aoqi Wang, Xingyi Tang, Meihua Piao. 1994

XSL•FO
RenderX

Web-Based Application for Cognitive and Functional Assessments in Dementia Screening: Mixed Methods, User-Centered Development Approach (e85454)
Nattaporn Piyaamornpan, Suwat Srisuwannanukorn, Kosit Tangthamrongthanawat, Pornsawan Mekhasingharak, Chatchawan Rattanabannakit, Saowalak Hunnangkul, Natthamon Wongkom, Vorapun Senanarong. . . . 2009

Identifying Biomarkers Using a Portable, Home-Based Eye-Tracking System to Predict Short-Term Visual Fatigue Deterioration: Prospective Observational Feasibility Study (e84479)
Fan Song, Guangyu Li, Jian Zhang, Zhen Tian, Mingge Li, Mengying Lai, Mingguang He, Yanxian Chen. . . . 2027

Proximal and Long-Term Participants' Perspectives on Approach Bias Modification for Smoking Cessation: Qualitative Study (e93238)
Alla Machulska, Johanna Bück, Esra Sünkel, Marie Neubert, Tim Klucken. . . . 2046

Evaluating a Digital Mental Health Tool for Implementation Into New Zealand's Integrated Primary Mental Health and Addictions Service: Usability Study (e84412)
Vincent Allen, Danielle Lottridge, Karolina Stasiak. . . . 2067

Association Between Web-Based Patient Portal Messaging and Disease Activity in Inflammatory Bowel Disease: Retrospective Pilot Cohort Study (e87518)
Pavan Vemulakonda, Shriansh Singh, Liza Khutsishvili, Anushka Deogaonkar, Valerie Stark, Marie Borum. . . . 2081

Multicenter Usability Evaluation and Co-Development of a Digital Decision-Support Tool for Labor Triage: Mixed Methods Study (e87589)
Mariana Tome, Xavier Laurent, Kristiyan Georgiev, John Tolladay, Sarah Collins, Deborah Hedgecott, Lyuba Bozhilova, Jane Hirst, Lawrence Impey, Antoniya Georgieva. . . . 2086

Barriers to Population Data Use in Overdose Fatality Reviews: Observational and Interview Study of Dashboard Deployment (e98864)
Amey Salvi, Cynthia Holladay, Allyson Dir, Bradley Ray, Matthew Aalsma, Khairi Reda. . . . 2107

Sentence-Level Provenance for AI Medical Record Summarization in a Click-to-Inspect Interface: Formative Usability Evaluation (e95644)
Andrew Parambath, Giordana Pulpo, Vince Hartman. . . . 2120

Understanding Delivery and Engagement With a Digital Self-Management Intervention for Chronic Obstructive Pulmonary Disease Across Two Clinical Settings: Qualitative Study (e88108)
Martin Ruddock, Alison Blythin, Bethany Cliffe, Lucy Yardley, James Dodd, Rachel Williams, Katherine Bradbury, Tom Wilkinson, Ben Ainsworth. . . . 2129

Physicians' Complex Relationships With Health Information Technology and Burnout: Cross-Sectional Survey Study (e93266)
Mavis Jones, Timothy Jason, Dara Liu, Jeannette Comeau, Chandi Chandrasena, Noni MacDonald, Janice Graham. . . . 2143

From Prototype to Practice: Mixed Methods Study of Human-Centered Design and Usability Evaluation of Augmented Reality for Complex Oncology Surgical Workflows (e86725)
Maria Matsangidou, Mikolaj Kowal, Jakob Yianni, Maria Krini, Samir Pathak, Katie Gordon, Malcolm Luker, Deborah Stocken, David Jayne, Constantinos Pattichis. . . . 2161

Privacy Fact Sheets for Mitigating Disease-Related Privacy Concerns and Facilitating Equal Access to the Electronic Health Record: Randomized Controlled Trial (e71124)
Niklas von Kalckreuth, Markus Feufel. . . . 2177

XSL•FO
RenderX

Exploring Medical Information Needs and Accessibility in Swedish Dental Care by Analysis of Documentation Workflows and Electronic Dental Records in Dalarna: Sociotechnical Qualitative Study (e82691)
Sahid Rahim, Nadia Davoody, Stefano Bonacina. . . . 2193

Factors Influencing User Satisfaction in Accessing Health Data: Cross-Sectional Survey of United Kingdom Adults (e75935)
Maria Xenou, Omar Al-Ubaydli. . . . 2214

Real-World Performance of a New Online Eye Symptom Triage Tool (Eye+Dot) in an Emergency Eye Clinic: Mixed Methods Evaluation Study (e81343)
Louise Allen, Jen Lim, Fengyi Liu, Nimesha Alex. . . . 2223

Evaluating a Shared Decision Support Tool for Pediatric Cardiopulmonary Arrest: Mixed Methods Usability Study (e78736)
Ana Rajic, Sharleen Kayne Olanka, Marco Generelli, Jennifer Davidson, Yiqun Lin, Ryan Kang, Kangsoo Kim, Pierre-Louis Rebours, Marc Ibrahim, Donovan Duncan, Sergio Manzano, Adam Cheng, Alexandre De Masi, Johan Siebert, Frederic Ehrler. . . . 2236

Development and Usability of a Dashboard for Quality Monitoring and Resident-Centered Care in Australian Residential Long-Term Care: Mixed-Methods Study (e80478)
Ronald Dendere, Gillian Stockwell-Smith, Michelle Lang, Murray Hargrave, Sara Mayfield, Leonard Gray. . . . 2259

Emotion-Adaptive Large Language Model–Driven Clinical Decision Support: User Evaluation of the Empathic Clinical Decision Support System Framework for Trust and Explainability (e89005)
Tongze Zhang, Sang Bae, Tammy Chung, Anind Dey. . . . 2272

Constructing a Theoretical Hybrid Care Pathway From Stakeholder Perspectives: User-Centered Design Study (e93735)
Ruben Knevels, Jelle Schils, Sarah Haesevoets, Julie Vranken, Chris Burtin, David Ruttens, Femke Wouters. . . . 2293

Development and User Testing of an Online Platform to Deliver School-Based Inclusive Physical Education Solutions: Mixed Methods Study (e89476)
Sangeetha Mohanraj, Laurie Malone, Sunkanmi Arogbokun, Lauren Lieberman, Mohanraj Thirumalai. . . . 2307

Designing Coloring-Based Digital Art Therapy to Treat Alexithymia in Chinese College Students: Qualitative Study (e82128)
Shuzhan Liu, Haoyong Deng, Jing Chen, Qi Wang. . . . 2322

User Experience of a Web-Based Mobile Health App Supporting Low-Income Pregnant Individuals With Diabetes: Mixed Methods Study (e84061)
Sydney Raucher, Layna Lu, Tazim Merchant, Elizabeth Soyemi, Charlotte Niznik, Rana Saber, Chen Yeh, Lynn Yee. . . . 2337

Novel Web-Based Technology to Promote Goal-Setting in Complex Chronic Illness: Randomized Controlled Trial (e70402)
Jody Lin, Bernd Huber, Ofra Amir, Shiri Assis-Hassid, Sebastian Gehrmann, Krzysztof Gajos, Barbara Grosz, Lee Sanders. . . . 2350

Digital Health Literacy and Tool Adoption in Postoperative Care in a Safety-Net Hospital Population: Mixed Methods Study (e75496)
Christopher Awad, Alexander Martinek, Chunhao Zou, Rosalind Byrd, Hui Ooi, Kimberly Do, Henry Young, Bhavin Adhyaru, Muhammed Idris, Rosa Arriaga. . . . 2363

Evaluating User Engagement and Satisfaction With Digital Mental Health Interventions: Randomized Controlled Trial of a Text Messaging Program and e-Mental Health Resources (e76587)
Gloria Obuobi-Donkor, Reham Shalaby, Belinda Agyapong, Samuel Obeng Nkrumah, Medard Adu, Ejemai Eboreime, Lori Wozney, Vincent Agyapong. . . . 2374

XSL•FO
RenderX

Designing a Substance Misuse Data Dashboard for Overdose Fatality Review Teams: User-Centered Design Approach (e79407)
Marie Pisani, Madeline Oguss, Julia Dickson-Gomez, Constance Kostelac, Amy Parry, Starr Moss, Elizabeth Salisbury-Afshar, Brian Patterson, Michael Spigner, Megan Gussick, Alison Krautkramer, Timothy Gruenloh, Askar Safipour Afshar, Preeti Gupta, Anoop Mayampurath, Majid Afshar. . . . 2389

Navigation, Adoption, and Use of Digital Health Technologies for Irritable Bowel Syndrome Self-Management: Focus Group Study of Patient Experience and Decision-Making (e75012)
Adrijana D'Silva, Nicolle Hua, Mary Modayil, Judy Seidel, Deborah Marshall. . . . 2403

Adolescent Perceptions of an Online Safety Chatbot: Survey Study (e71498)
Meriel Charles, James Sauer, Erin Roehrer, Jeremy Prichard, Paul Watters, Joel Scanlan. . . . 2418

Simulated Workflow Feasibility Evaluation of a Web-Based Periorbital Measurement Platform: Development and Usability Study (e82859)
George Nahass, Jacob Ende, Sasha Hubschman, Benjamin Beltran, Bhavana Kolli, Caitlin Berek, James Edmonds, RV Chan, Pete Setabutr, James Larrick, Darvin Yi, Ann Tran. . . . 2435

Perceptions of Health Care Professionals on the Integration and Use of AI in Clinical Cancer Care: Interview Study (e83240)
Akhona Khumalo, Ala Alaqra. . . . 2445

Digital Health Literacy and Health Technology Readiness Among People With Epilepsy or Multiple Sclerosis: Cross-Sectional Study (e85625)
Anna Vahteristo, Virpi Jylhä, Hanna Kuusisto. . . . 2466

Digital Mental Health Promotion Services for Youth: A Qualitative Study of Help-Seeking Through Mindhelper.dk (e91017)
Amalie Kusier, Caroline Dalsgaard, Sofie Hoffmann, Lau Thygesen, Anna Folker. . . . 2478

Factors Associated With Health Perception and Psychological Well-Being Among Third-Age University Students: Cross-Sectional Study on the Role of Cyberchondria (e83371)
Semra Gündo du, Özlem Özgür, Asli Kilavuz, smail Tufan. . . . 2494

Effectiveness of a Brief Digitized Contact-Based Intervention in Improving Mental Health Stigma and Help-Seeking in Young Adults: Mixed Methods Study (e74391)
Annabel Songco, Natalie Dahora, Andrew Essen, Andrew Mackinnon, Gemma Sicouri, Elizabeth Pellicano, Jennifer Hudson. . . . 2509

Digital Episodic Future Thinking Intervention (Luminaut): Co-Design and Iterative Development Study (e74099)
Naomi Hoffmann, Caitlin Howlett, Kate Little, Ian Gwilt, Megan Rebuli, Paige Brooker, Aaron Davis. . . . 2525

Young People's Perceptions of Signposting in a Digital Mental Health Helpline: Mixed Methods Analysis of Cross-Sectional Data (e73369)
Angel Au, Anya Jan, Shyn Phua, Leslie Gutman. . . . 2544

Identifying Skill and Usability Barriers to Digital Health Tool Use Among Older Adult Patients in US Safety Net Clinics: Mixed Methods Study (e78430)
Taylor Rapson, Magaly Ramirez, Sandy He, Jeanette Wong, Hyunjin Kim, Isabel Luna, Andersen Yang, Junhong Li, Paul Fishman, James Ralston, Courtney Lyles, Elaine Khoong. . . . 2561

XSL•FO
RenderX

Cocreating a Digital Patient Preparedness Tool for International Students Accessing Primary Care in Germany: A Convergent Methods Study Within the Health CASCADE Network (e87519)
Vinayak Anand Kumar, Margrit Schreier, Maria Giné-Garriga, Maria Ortmann, Fatima-Zohra Belmokhtar, Simona Grineviciute, Tran Quan, Likhita Aluru, Sonia Lippke. . . . 2581

Association of eHealth Literacy With Cognitive Function, Frailty, and Functional Status in Hospitalized Geriatric Patients: Cross-Sectional Observational Study (e92045)
Chantal Flemm, Rainer Wirth, Sarah El Bataioui, Diana Daubert, Ida Haussleiter, Matthias Hermenau, Jan Dieris-Hirche, Judith Kuhlmann, Katharina Kunde, Ulrike Trampisch, Bianca Ueberberg, Horst Vollmar, Theresa Busse. . . . 2600

Telemedicine Adoption Among Health Care Professionals in Israeli Geriatric Medical Centers: Moderated Mediation Cross-Sectional Questionnaire Study (e95371)
Gizell Green, Alena Lochmannová, Riki Tesler, Tammy Porat-Packer. . . . 2618

Exploring Common and Novel Actualized Affordances of Fitbit: Mixed Methods Study (e85412)
Moayad Alshawmar, Bengisu Tulu, E Wilson, Adrienne Hall-Phillips, Ibrahim Aljadani, Emmanuel Agu. . . . 2633

Designing for Amplifying Voices—Supporting Bangladesh's Low-Income Ready-Made Garment Workers Through Technology Amid COVID-19 Pandemic: Interview Study Among Workers (e90675)
Rahat Rony, Nova Ahmed. . . . 2645

Evaluating Acceptance and Use of Digital Health Platforms for Consultation and Medication Delivery Among Men Who Have Sex With Men and Transgender People Living With HIV: Cross-Sectional Survey Informed by an Extended UTAUT Approach (e93129)
Emmanuel Oluokun, Festus Adedoyin, Huseyin Dogan, Nan Jiang. . . . 2659

Physician-Patient Communication and Physicians' Acceptance of a Tailored Digital Health Information Service: Quantitative Online Survey (e80620)
Paula Memenga, Elena Link. . . . 2680

Near Miss Reporting and Organizational Learning in Health Care: Conceptual Framework Development Study (e87846)
Mohammed As'ad. . . . 2691

"To Choose a Band-Aid" - Children's Preferences for Participation in Health Care Situations: Q Methodology Study (e89802)
Merja Hietanen, Maksims Kornevs, Catarina Nahlén Bose, Jayanth Raghothama, Sebastiaan Meijer. . . . 2703

Anticipated Barriers and Facilitators to the Implementation of an AI Tool Advising on the Placement of Mitigation Strategies to Improve Air Quality in the Hospital Setting: Qualitative Study (e93420)
Yolanda Barrado-Martín, Sanjoli Mathur, Zarnie Khadjesari, Sigrún Clark, Cecilia Vindrola-Padros, Laurence Lovat, Fiona Stevenson. . . . 2720

Conversational AI in Hereditary Cancer Care: Sociotechnical Study of Responsible Design Requirements (e97869)
Dominic Lammert, Jacqueline Lammert, Justin Hofenbitzer, Tristan Radtke, Yingping Sun, Martin Haimerl, Stefanie Betz, Doran Scholl, Jürgen Pfeffer. . . . 2742

Risk Visualization With Petal-X and Continuous Glucose Monitoring to Support Person-Centered Care and Lifestyle Change in Cardiovascular Prevention: Qualitative Descriptive Study (e92101)
Adrijana Svenšek, Diego Rojo, Mateja Lorber, Zalika Klemenc-Ketis, Lucija Gosak, Katrien Verbert, Gregor Štiglic. . . . 2762

Real-World Health Monitoring Using Wrist-Worn Wearables in Community-Dwelling Older Adults in China: Mixed Methods Feasibility Study (e98592)
Ying Bai, Yi Tang, Yiwei Zhao, Lixin Ma, Yanjun Ma, Zhibin Wang. . . . 2778

XSL•FO
RenderX

Body Image and Eating Disorder Education Chatbot, JEM, in Australia and Canada: First 6-Month Real-World Survey Evaluation (e90783)
Gemma Sharp, Sara Marini, Emily Tam, Hao Hu. . . . 2791

Strategic Use of Negative Emojis in Messaging-Based Interventions for Public Health Communication on Social Media: Mixed Methods Study (e78824)
Yue Luo, Shubin Yu. . . . 2810

Effectiveness of Mobile-Delivered Exercise and Yoga Programs on Depressive Symptom Reduction in Employees: Randomized Controlled Trial (e80287)
Jeong-Hyun Kim, So Yoo, Sohee Oh, Sun-Young Moon, Kyesan Lee, Ki-Hyeok Nam, Heyeon Park, Se Jung, Hyo-Joon Choi, Hyo Kim, Seok Lee, Namhee Kim, Myoungnam Song, Beomjun Min, Ji-Hye Lee, Jeong Shin. . . . 2825

Evaluating a Dual Digital Cognitive Behavioral Therapy and Health and Wellness Coaching Intervention for Anxiety and Depression: Single-Arm Pilot Study (e92448)
Alex Kirk, Christopher King, Robert Gallop, Brian Doan, Lorelei Simpson Rowe, Laura Randa. . . . 2850

Patient Experiences of Digital Technology Use in Interstitial Lung Disease (PRODIGY-ILD Study): Qualitative Study Using Reflexive Thematic Analysis (e91533)
Emer Gunne, Alessandro Franciosi, Cormac McCarthy, Michael Keane, Peter Doran, Sinead Holden. . . . 2869

Evaluation of the Feasibility and Acceptability of Perfect Fit, a Virtual Coach–Based mHealth Intervention for Smoking Cessation and Physical Activity in Adults: Mixed Methods Study (e83456)
Milon van Vliet, Eline Meijer, Nele Albers, Kristell Penfornis, Walter Baccinelli, Bouke Scheltinga, Roxy van Eersel, Niels Chavannes, Anke Versluis, The Perfect Fit Consortium. . . . 2887

Exploring Use, Usability, and Functionality in the Iterative Testing of a Co-Designed Preventative Mental Health Smartphone App for University Students: Comparison of Monash Thrive App Development Trials (e67818)
Melinda McCabe, Michelle Byrne, Judith Gullifer, Isabelle Smart, Chunlei Chang, Sanket Singh, Gagan Khara, Kim Cornish. . . . 2908

The Role of Rating Valence in AI Skin Cancer App Acceptance: Eye-Tracking and Questionnaire Study (e93489)
Inga Jagemann, Sabrina Hegner, Gerrit Hirschfeld. . . . 2918

The Quality and Characteristics of Digital Mental Health Apps: Mixed Methods Study (e67944)
Maciej Zych, Raymond Bond, Maurice Mulvenna, Jorge Martinez Carracedo, Lu Bai, Tim Andrews. . . . 2936

Participant Engagement With a Digital Behavioral Health App for Chronic Pain: Descriptive Secondary Analysis of a Feasibility Randomized Controlled Trial (e88122)
Susan Zbikowski, Jo Masterson, Yohali Burrola-Mendez, Chialing Hsu, Kris Ma, Ying Zhang, Deanna Waters, Kari Stephens. . . . 2956

User Acceptance of Smart Home Emergency Response Systems: Mixed Methods Study (e93003)
Michael Pantförder, Oliver Krüger, Annika Rietz, Colin Dzieia, Sven Meister. . . . 2971

AI-Enhanced Conversational Agents for Personalized Asthma Support in People With Asthma: Factors for Engagement, Value, and Efficacy in a Cross-Sectional Survey Study (e80979)
Laura Moradbakhti, Dorian Peters, Jennifer Quint, Björn Schuller, Darren Cook, Rafael Calvo. . . . 2990

A User-Centered Interface Design Framework for the DELONELINESS System in Older Adults: Design Indicator Development and Prioritization (e88263)
Yi Zhou, Jessica Rees, Faith Matcham, Michela Antonelli, Sebastien Ourselin, Wei Liu. . . . 3010

XSL•FO
RenderX

Usability Evaluation of a Macular Quantitative Square Grid Self-Examination Application in Patients With Macular Disease: Mixed Methods Study (e79699)
Shu Li, Enming Zhang, Jiani Pan, Yan Xu, Kairong Zheng, Xun Xu, Qiong Fang. 3027

Evaluating a Smartphone App to Monitor Blood Pressure in Normotensive Pregnancies, High-Risk Pregnancies, and Women With Preeclampsia: Prospective Longitudinal Feasibility Study (e70370)
Maria Andersson, Christine Rubertsson, Elia Psouni, Lena Erlandsson, Camilla Edvinsson, Stefan Hansson. 3038

Acceptability of a Digital Care App in Patients Undergoing Hip and Knee Arthroplasty: Prospective Cohort Study (e79682)
Yacine Louni, Matthew Laroche, Abdulrhman Alnasser, Mohammad Abuhaneya, Eric Belzile, Sandhya Baskaran, Jennifer Mutch, Anthony Albers. 3054

Providing Psychological Therapy Support and Improving Postdischarge Data Collection: Preliminary Evaluation of the “Paddle” App (e68671)
Graham Thew, Michelle Lee, Ineke Wolsey, Fiona Giles, Emma Hayden, Grace Jell, Jamie Peacock, John Pimm, Joanne Ryder, Alison Salvadori, David Clark. 3073

Deployment of an Activity Monitoring Program to Complement a Clinical Intervention for Veterans With Gulf War Illness: Qualitative Study (e82177)
Selene Mak, Pauline McManus, Steven Greer, Kari Haws, Carrie Carlson, Helena Chandler, Omowunmi Osinubi, Garrett Ash. 3085

Continuous Glucose Monitoring for Personalized Nutrition in Real-World Vively App Users: Retrospective Observational Study (e80734)
Michelle Jospe, Kelli Richardson, Susan Schembre. 3095

Support, Monitoring, and Reminder Technology for Mild Dementia (SMART4MD) for People With Mild Cognitive Impairment and Their Informal Caregivers: Cost-Effectiveness Analysis (e77808)
Zartashia Ghani, Johan Jarl, Peter Anderberg, Johan Sanmartin Berglund, Maria Aparicio, Pilar Barnestein-Fonseca, Selim Cellek, Fermín Mayoral-Cleries, Maite Garolera, Gloria Guerrero-Pertiñez, Karen Hayden, Carmel Moore, Jufen Zhang, Dominic Trepel, Sanjib Saha. 3108

Early Deployment of an Integrated Digital Platform (shamiriOS) for Scalable Youth Mental Health Service Delivery in Kenya: Development and Usability Study (e79107)
Shadrack Lilan, Tom Osborn, Wendy Mmbone, Jean Kasudi, Benny Otieno, Ichami Etyang, Jianing Tu, Edmund Korley, Christine Wasanga. . 3 1 2 5

Effect of a Personalized Mobile App on Glucose Control in Adults With Prediabetes and Type 2 Diabetes: Exploratory Pilot Randomized Controlled Trial (e87692)
Elena Lalama, Marta Csanalosi, Stefan Kabisch, Saskia Wilson-Barnes, Lazaros Gymnopoulos, Kosmas Dimitropoulos, Konstantinos Rouskas, Argiriou Anagnostis, Ioannis Oikonomidis, Leontios Hadjileontiadis, Veronique Cornelissen, Maria Hassapidou, Ioannis Pagkalos, Sofia Balula Dias, Kathryn Hart, Andreas Pfeiffer. 3148

Mobile Health for Postpartum Contraception Knowledge in India (mMitra): Realist Evaluation Using Moderation Analysis and Structural Equation Models (e97650)
Abinaya Chandrasekar, Emma Slaymaker, John Ford, Neha Madhiwalla, Emily Warren, Caroline Free, Ona McCarthy. 3167

Long-Term Effects of Multimedia Education and “Foot Alerts” Through the MyU App on Diabetic Foot Care in Individuals With Moderate-High Risk: Pilot Randomized Controlled Trial (e78261)
Ameenah Alawadhi, Kay Khan, Grace Messenger, Alice Cusworth, Mohammad Assi, Stuart Gray, Ebaa Al-Ozairi. 3183

Co-Created Digital Pretherapy Psychoeducation for Outpatients in Specialized Mental Health Care: Usability Evaluation and Patient Satisfaction Study (e80130)
Henrik Pedersen, Tatiana Skliarova, Liv Engvik, Arthur Mandahl, Carlos De las Cuevas, Audun Havnen, Mariela Lara-Cabrera. 3199

Utilization of AI Among Medical Students and Development of AI Education Platforms in Medical Institutions: Cross-Sectional Study (e81652)
Xiaokang Shi, Zewu Jiang, Li Xiong, Ka-Chun Siu, Zhen Chen. 3218

XSL•FO
RenderX

Evaluation of Combined Educational Methods on Motivational Interviewing for Final-Year Medical Students: Mixed Methods Study (e89126)
Isaraporn Thepwongsa, Radhakrishnan Muthukumar, Pat Nonjui. 3233

Understanding Barriers to Effective Injury Care by Medical Trainees and Traffic Law Enforcement First Responders in Low-Income Contexts in Uganda (Motor Registry Project Part 2): Convergent Mixed Methods Analysis (e84774)
Herman Lule, Benson Oguttu, Micheal Mugerwa, Michael Wilson, Jussi Posti. 3254

E-Learning Modules Based on Bloom Taxonomy and the Miller Pyramid for First-Year Indian Medical Students: Randomized Controlled Study in Medical Education (e84339)
Archana Prabu Kumar, Padmavathi R, Maheshkumar K, Abirami Omprakash, Prabu Chokkalingam Mani, B Sathiyasekaran, Maruthy K N. . . 3 2 7 6

Gamified Formative Assessment in a First-Year Medical Morphology Course: Qualitative Study (e90651)
Álvaro Herrera Alcaíno, Sofía Araya Figueroa, Valentina Riquelme Vilches, Eduardo Vargas Zurita, Renata Cozzi Ahumada, Camila Riquelme Bahamondes, Carlos Rosas Chuñil. 3291

Human-AI Interaction With AI-Assisted Tumor Overlays in Pediatric Whole-Body Magnetic Resonance Imaging: Exploratory Reader Study (e81066)
Abhishek Moturu, Olurotimi Komolafe, Sayali Joshi, Andrea Doria, Anna Goldenberg. 3303

Effects of AI Assistance Timing on Pharmacists' Trust in Automated Pill Recognition Technology: Within-Participants Experimental Study (e92822)
Jin Kim, Brigid Rowell, Megan Whitaker, Qiyuan Chen, Raed Al Kontar, Corey Lester, Xi Yang. 3314

Clinician Trust and Human Factors in AI-Enabled Clinical Decision Support in Acute Care: Mixed Methods Study (e95472)
Meghana Darla, Danielle Miltz, Khushboo Chandnani, Saptarshi Purkayastha, John Diehl, Sivasubramanium Bhavani. 3329

Public Perceptions and Determinants of Preference for Robotic-Assisted Surgery in the United Arab Emirates: Cross-Sectional Study (e97975)
Yasir Elhadi, Aminu Abdullahi, Abdelrahman Alblooshi, Abdulla Alhosani, Arwa Al Khalidi, Mohammed Saleem, Shamma Alrashdi, Reem Aljaberi, Azhar Rahma. 3347

Clinicians' Trust in AI-Based Clinical Recommendations Across Controlled Primary Care–Style Clinical Vignettes: Nurse-Dominant Experimental Pilot Study (e97649)
Avishek Choudhury, Yeganeh Shahsavar, Ayse Gurses. 3362

Age-Related Differences in Mental Health Status at Psychiatric Discharge Among Patients Across 10 Acute Care Facilities in Alberta, Canada: Cross-Sectional Study (e92360)
Wanying Mao, Ernest Owusu, Reham Shalaby, Hossam Elgendy, Belinda Agyapong, Ejemai Eboreime, Peter Silverstone, Pierre Chue, Wesley Vuong, Arto Ohinmaa, Andrew Greenshaw, Vincent Agyapong. 3385

Factors Influencing Patient Satisfaction Within the Multistage Online Consultation Process in Internet Hospitals in Beijing: Empirical Analysis (e79213)
Yi Wang, Yujia Han, Mengyu Duan, Yani Wang. 3400

Exploring English and Swedish General Practitioners' Behavioral Intentions to Use Telemedicine: Comparative Study (e73609)
Michael Harris, Gordon Taylor, Miriam Pikkemaat, Hans Thulesius, Veronica Milos Nymberg. 3411

XSL•FO
RenderX

Comparison of Mask-Wearing Behavior on Social Media and Its Relationship With Demographic Characteristics During the COVID-19 Pandemic: Social Media Data Analysis Between the United States and Japan (e78236)
Kiki Ferawati, Wan She, Shoko Wakamiya, Eiji Aramaki. . . . 3420

Using Digital Mental Health and Measurement-Based Care With Youth Experiencing Suicidal Thoughts and Behaviors: Qualitative Study of Clinician Barriers and Facilitators (e81505)
Lia Norman, Katherine Bright, Karina Pintson, Julia Hews-Girard, Adrijana D'Silva, Courtney Habina, Diana Pricop, Geneca Henry, Sarah Daniel, Lauren Volcko, Katelyn Greer, Jason Gondziola, Leanne Stamp, Melanie Fersovitch, Karen Moskovic, Jessica Bradley, Haley LaMonica, Ian Hickie, Frank Iorfino, David Johnson, Paul Arnold, Gina Dimitropoulos. . . . 3436

Service Users' Views on Digital Remote Monitoring for Psychosis: Survey Study (e86152)
Xiaolong Zhang, Emily Eisner, Daniela Di Basilio, Cara Richardson, Joseph Firth, Sandra Bucci. . . . 3459

The Impact of Tobacco Smoking on Treatment Response Among Patients With Psoriasis Undergoing Biologic Treatment: Prospective Observational Study (e90963)
Fanlingzi Shen, Yuning Ding, Xiuqi Zhang, Quanruo Xu, Zhen Duan, Ruiqi Cai, Rui Zhang, Xiangjin Gao, Ruiping Wang. . . . 3473

Sex Representation and User Preferences in Pain Drawing Body Charts in Back Pain Research: Multimethod Study (e76175)
Sonja Schläpfer, Casey Murphy, Yanick Lukic, Concepción Campos-Asensio, Giovanni Colacicco, Beatrice Lauber, Sophie Masneuf, Luana Nyirö, Georg Kajdi, Reto Sutter, Christoph Laux, Philipp Ackermann, Jesús López-Alcalde, Claudia Witt. . . . 3485

Parental Influence on Children's Media Use in South Korea: National Population-Based Study (e75292)
Ji Kim, Ah Yang, Hye Lee. . . . 3503

An Internet-Delivered Intervention to Reduce LGBTQ+ Prejudice Among Romanian Teachers: Randomized Controlled Trial (e63787)
Nastasia S l gean, Ioana Latu, Torill Larsen, Andreea Isb oiu, Florin Sava. . . . 3525

Understanding Psychologists' Usage, Knowledge, and Attitudes Toward Digital Mental Health Solutions for Refugees and Migrants: Exploratory Cross-Sectional Survey in Sweden (e75263)
Kristofer Vernmark, Anahita Geranmayeh, Naira Topooco, Gerhard Andersson, Shervin Shahnavaz. . . . 3537

Exploring Response Patterns to Motivational Messages Supporting Physical Activity: Latent Class Analysis (e91917)
Chihiro Moriishi, Takeyuki Oba, Keisuke Takano, Kentaro Katahira, Kenta Kimura. . . . 3552

Exploratory Analysis of the Digital Health Gender Paradox in Consumer-Facing Digital Health Technology Adoption: Cross-Sectional Survey Study (e101963)
Jessica Wallace, Mehak Batra, Tafheem Wani, Michael Liem, Kylie Ovenden, James Boyd, Therese Keane, Urooj Khan. . . . 3569

The Use of Digital Technology to Support the Emotional Well-Being of Older People Who Receive Care at Home From the Perspective of Care Recipients: Qualitative Study (e93823)
Tanya Davison, Hilary Davis, Marissa Dickins, Sonja Pedell, Tuan Nguyen, Deborah Carveth, Therese Den Dulk, Karen Smith, Julie Kelly, Sunil Bhar. . . . 3584

Attitudes Toward Digital Health Technology in Neurological Conditions: Qualitative Study of Patients, Caregivers, and Health Care Professionals (e107517)
Sarah Fearn, John Spreadbury, Christopher Kipps. . . . 3598

AI-Assisted Clinical Documentation in Routine Danish General Practice: Quantitative Pre-Post Study (e94431)
Louise Olsen, Philipp Harbig, Anna Laurberg, Jacob Laurberg, Morten Charles. . . . 3611

XSL•FO
RenderX

Large Language Model–Based Simplification of Digital Therapeutics Explanations for Insomnia and Nicotine Dependence: Two Randomized Online Experiments (e89451)
JunYoung Seo, Moses Yook, Dai Kim, JunHee Lee, Jae Yoo, GiHwan Byeon, In Choi. . . . 3620

Demystifying Quality Metrics and Unveiling the True Measure of Quality of Care in Nursing Homes: Mixed Effects Analysis (e72770)
Suhas Bharadwaj, Haneen Ali. . . . 3648

Radiologist Perceptions of an AI Tool for Intracranial Hemorrhage Detection in Teleradiology: Cross-Sectional Survey Study (e92145)
Andrew Del Gaizo, Jackson Del Gaizo, Troy Shahoumian. . . . 3669

Preliminary Support for a Technology-Enhanced Suicide Prevention Intervention for College Students: Mixed Methods Preimplementation Study (e79537)
Jocelyn Meza, Joan Asarnow, Natalia Jaramillo, Juliane Martinez, Candice Biernesser, Brandie George-Milford, David Brent. . . . 3678

Integrating Wearable Technology in Digital Therapy for People With Psychosis (SloMo): Mixed Methods Co-Design and Feasibility Study (e91188)
Kathryn Taylor, Thomas Ward, Matteo Cella, Philippa Garety, Amy Grant, Katie Philps, Anna Wojdecka, Melanie Flory, Amy Hardy. . . . 3697

Effects of a Group-Based Cognitive Behavioral Therapy for Sleep in Nursing Students: Pilot Randomized Controlled Trial (e85021)
David Pérez-Manchón, Clara Azpeleta, Beatriz Gal-Iglesias, Cayetana Ruiz Zaldibar. . . . 3713

Using Intervention Mapping to Codevelop Orchid, a Digital Tool for Reproductive Life Planning: Development and Feasibility Study (e87650)
Catherine Stewart, Helen Carr, Maitri Tursini, Alice Howe, Jennifer Hall. . . . 3730

Examining the Effect of Online Engagement on Older Adults' Subjective Memory Capability: Cross-Sectional Path Analysis (e73018)
Soohyoung Lee, Hang Liu. . . . 3746

High Screen Exposure and Its Association With Physical and Mental Well-Being Among School-Going Children and Adolescents in Bangladesh: Cross-Sectional Study (e73524)
Shahria Kakon, Tanjir Soron, Mohammad Hossain, Biplob Hossain, Fahmida Tofail, Rashidul Haque. . . . 3755

Telehealth Use Among Older Adults Receiving Home- and Community-Based Services: Cross-Sectional Analysis Using the National Core Indicators–Aging and Disabilities Survey (e77522)
Dana Urbanski, Romil Parikh, Benjamin Langworthy, Jack Wolf, Chanee Fabius, Janette Dill, Eric Jutkowitz, Tetyana Shippee. . . . 3766

Comparing Images of Depression in Mass Media and AI-Generated Pictures: Mixed Methods Study (e81230)
Nuria Saladie, Carolina Llorente, Renée Joosten, Gema Revuelta. . . . 3784

Feasibility and Acceptability of a Prevention-Focused Screener for Perinatal Depression Risk: Mixed Methods Cohort Study (e81638)
Tamar Krishnamurti, Samantha Rodriguez, Leah Cope, Lara Lemon, Priya Gopalan, Cara Nikolajski, Hyagriv Simhan, Kelly Williams. . . . 3798

Using a Digital-Based Mindfulness Curriculum to Enhance Healthy Aging Outcomes in Community-Dwelling Older Adults in Taiwan: Mixed Methods Feasibility Study (e84161)
Yu-Rung Wang, Pei-Lun Hsieh, Chia-Hsiu Chang, Chia-Chi Hsiao, Mei-Lien Hu. . . . 3809

XSL•FO
RenderX

Adaptation and Acceptability of a Low-Intensity Cognitive Behavioral Therapy App to Support Low Mood and Worry Management in Female Forces Veterans: Mixed Methods Study (e84365)
Paul Farrand, Andy Bacon, Melika Janbakhsh, Natalie Flay, Elizabeth Turnbull, Jonathan Baker. 3828

Mental Health Profiles Based on Self-Regulation and Technology Use in the Digital Era in a Spanish-Speaking Sample: Latent Profile Analysis (e77167)
Angélica Garzón Umerenkova, Marisleidy Alba Cabañas, Erika Malpica-Chavarria. 3846

Online Tiered Screening for Mental Health Problems Among Refugees in Sweden: Validation Study (e82763)
Jennifer Meurling, Elisabet Rondung, Youstina Demetry, Anahita Geranmayeh, Anna Leiler, Gerhard Andersson, Anna Bjärtå. 3861

Investigating the Impact of Lived Experience Stories on Self-Harm, Mood, and Help-Seeking Intentions: Web-Based Between-Participants Experimental Study in Individuals With Recent Self-Harm (e71280)
Jennifer Ferrar, Lizzy Winstone, Ian Penton-Voak, Lucy Biddle, Paul Moran, Lydia Grace, Becky Mars. 3877

Use of Digital Peer Support for Employee Well-Being: Retrospective Analysis Across Five Large Employers (e90431)
Farbod Sedaghati, Anya Stetsenko, Ilayda Ozsan McMillan, Harpreet Nagra, Zara Dana. 3892

Risk Factors for Noninitiation and Dropout in Blended Therapy in Inpatient Psychiatric Patients: Retrospective Cohort Study (e85816)
Nikita Gupta, Fabian Gardin, Thomas Berger, Wolfram Kawohl. 3912

Wellness4Teachers Supportive Text Messaging Program Association With Observed Improvement in Psychological Symptoms Among Teachers in Canada: Single-Arm Prospective Longitudinal and Naturalistic Controlled Study (e92142)
Belinda Agyapong, Yifeng Wei, Pamela Brett-MacLean, Vincent Agyapong. 3924

Objective Monitoring of Tablet Use–Related Optical Exposure and Its Association With Axial Length in Preschool Children: Cross-Sectional Intelligent Monitoring Study (e79266)
Yidong Zhu, Hao Chen, Senlin Lin, Hong Jiang, Mingdao Zhang, Yi Sun, Chenshu Li, Yingnan Jia. 3941

Good Clinical Practice Guidance for Line Materials, Filtration, and Light Protection in Intravenous Medication Administration: Modified Delphi Consensus Study (e88333)
Andrew Dickman, Penelope Tuffin, Rania Al-Jaber, Mona El-Harmeel, Jennifer Schneider, Irene Taladriz-Sender, Adam Sutherland, Robert Terkola, James Waterson. 3956

Development of a Blended Physical Activity Intervention for Office Employees Using Intervention Mapping: Intervention Development Study (e87328)
Yan Sun, Alison Ou, Tony Shi, Yang Gao, Martin Hagger, Zhennan Xiong, Xiang-yan Chen, Shirley Fong, Linda Zhong, Yao Xie, Julien Baker.
3 9 6 8

Investigating Context in How Clinicians Make Sense of AI: Qualitative Interview Study (e99117)
Simone Mingels, Hannah Piehl, Rianne Fijten, Ricky Janssen, Bart Penders. 3979

A Log-Level Data-Driven Precision Education Tool for Pediatrics Trainees: Human-Centered Development and Validation Study (e79952)
Alexander Fidel, Mark Mai, Naveen Muthu, Adam Dziorny. 3994

Shifting Skills in Robot-Assisted Surgery: Case Study With Interdisciplinary Perspectives for Human Factors and Career Research (e87735)
Thomas Ellwart, Julia Birke, Johanna Paul. 4010

Health Care Workers' Perspectives on Virtual Reality in Mental Health: Qualitative Interview Study (e80851)
Anna Szczegielniak, Jakub Mo aryn, Łukasz Kunert, Szymon Florek, Robert Pudlo. 4026

XSL•FO
RenderX

Interactive, Personalized Patient Decision Aid for COVID-19 Vaccination in Canada: User-Centered Design Approach (e86283)
Doriane Etienne, Patrick Archambault, Isaac Bogoch, Christine Chambers, Andrea Chittle, Juliette Demers, S Driedger, Ève Dubé, Marie-Pierre Gagnon, Teresa Gavaruzzi, Anik Giguère, Nathalie Grandvaux, Kelly Grindrod, Hina Hakim, Samira Jeimy, Jason Kindrachuk, Annie LeBlanc, Shannon MacDonald, Ruth Ndjaboue, Magniol Noubi, Rita Orji, Jean-Sébastien Paquette, Elizabeth Parent, Jean-Sébastien Renaud, Beate Sander, Monica Taljaard, Dana Tannenbaum Greenberg, Marie-Claude Tremblay, Sabina Vohra-Miller, Vivian Welch, Holly Witteman. 4038

Enhancing Understanding and Acceptance of Equipment Localization: Mixed Methods Study With Clinic Staff and Potential Patients (e79583)
Angela Fiedler, Sabine Patzl, Melissa Schütz, Astrid Schütz. 4061

Learning From the Adoption of a Readmission Clinical Decision Support Tool: Group Model Building Approach (e87522)
Nina Sperber, Sarah Haas, Jiaxin Gao, Samantha Hamelsky, Theresa Kiki-Teboum, Afraaz Malick, Rishab Pulugurta, Jacqueline Rodriguez, Hana Shafique, Eden Singh, Kriti Vasudevan, Scott Rockart, David Gallagher, Adam Johnson. 4072

Reviews

Engagement-Supportive Mechanisms in Stress-Focused Digital Mental Health Interventions With Adaptive, Conversational, and Context-Aware Systems: Narrative Review (e90855)
Munesu Maminimini, Sofia Piffer, Silvia Gabrielli. 345

Mobile Apps Supporting Medication Adherence in Substance Use Disorder Treatment: Narrative Review (e104420)
Jane Kim, Lua Boschman, Pouya Azar, Borna Noureddin. 373

Persuasive Systems Design Features of Smartphone Apps for Psychosis: Systematic Review (e81101)
Theresa Taylor, Jasmin Peat, David Kell, Shadi Daryan, Pamela Jacobsen. 386

Soma Design for Digital Mental Health and Well-Being Interventions: Scoping Review (e79400)
Sachiyo Ito-Jaeger, Aysegul Kafadar, Steve Benford, Elvira Perez Vallejos. 399

Engaging Older Adults With Neurocognitive Disorders in Digital Health Technologies: Scoping Review (e70157)
Sié Da, Maxime Sasseville, Marie-Soleil Hardy, Idrissa Beogo, Amédé Gogovor, Samira Amil, Achille Yameogo, Florian Naye, Farzaneh Yousefi, Frédéric Bergeron, Anik Giguère, Annie LeBlanc, James Plaisimond, Carole Rivard-Lacroix, Marie-Pierre Gagnon, Chloé Cachinho. 416

Physical and Psychological Effects of Smartphone App–Based Walking Interventions in Community-Dwelling Older Adults: Systematic Review and Behavior Change Technique–Informed Analysis (e78042)
Hiroki Abe, Tomohiro Isinuki, Rinako Ono, Naoki Kanaya, Masaya Tanno, Michiyo Hirano. 430

How and Why Telehealth Interventions Improve Self-Care Among Vulnerable Groups of Patients With Heart Failure: Scoping Review and Rapid Realist Synthesis (e78859)
Saleema Allana, Colleen Norris, Armish Hussain, Nada Alaidarous, Raza Haider, Alexander Clark. 448

Machine Learning for the Analysis of Healthy Lifestyle Data: Scoping Review and Guidelines (e78648)
Tony Estrella, Lluis Capdevila, Carla Alfonso, Josep-Maria Losilla. 460

Predicting Electronic Health Record Usability: Scoping Review of Adoption Models, Metrics, and Future Directions (e86076)
Dillon Chrimes, Alex Thomo, Mu-Hsing Kuo, Elizabeth Borycki, Andre Kushniruk. 496

XSL•FO
RenderX

Psychometric Properties of Digital Divide Measurement Instruments in Health Contexts: Systematic Review (e82005)
Noelia Navas-Echazarreta, Elena Chover-Sierra, César Rubio-Tarazón, Raquel Martínez-Pascual, Pablo Del Pozo-Herce, Jesús Martínez-Tofé, Raúl Juárez-Vela, Antonio Martínez-Sabater. . . . 513

User-Centered Design for Digital Patient-Navigation Tools in Oncology: Scoping Review (e87686)
Saba Kheirinejad, Brianna White, Parnian Kheirkhah Rahimabad, Janet Zink, Soheil Hashtarkhani, Fekede Kumsa, Rezaur Rashid, Lokesh Chinthala, Christopher Brett, Robert Davis, David Schwartz, Arash Shaban-Nejad. . . . 524

Process Evaluation of Digital Mental Health Interventions for Psychosis: Scoping Review With Framework Synthesis (e88032)
Chloe Hampshire, Charlotte Dack, Shadi Daryan, Carolina Fialho, Rayan Taher, Ashley-Louise Teale, Jenny Yiend, Pamela Jacobsen. . . . 538

Usability and User Experience Assessment of Health Care Conversational Agents Using Validated Subjective Instruments: Systematic Review and Comparative Analysis (e102911)
João Pavão, Rute Bastardo, Anabela Silva, Nelson Rocha. . . . 559

Associations of Health Literacy and eHealth Literacy With Patient Portal Access and Use: Systematic Review (e104485)
Philipp Haessner, Mahitha Abraham, Aleena Peter, Abigail Rosario, Samika Kanaskar, Crystal Romero, Megan Gregory. . . . 3631

Viewpoints

The Cognitive Transaction: Toward a Human Factors Research Agenda for AI in Anesthesia and Perioperative Care (e102683)
Sheena Warner, Christopher Stucky, Tamara Haegerich, Young Yauger. . . . 882

How to Evaluate the Accuracy of Symptom Checkers and Diagnostic Decision Support Systems: Symptom Checker Accuracy Reporting Framework (SCARF) (e76168)
Marvin Kopka, Markus Feufel. . . . 889

Understanding the Limits of Patient Safety Classification Systems for Health Information Technology–Related Incidents (e91783)
Md Jabin. . . . 896

Prioritizing Human Factors in the Clinical Translation of Mixed Reality Surgical Guidance (e101995)
Shayan Bahadori, Peter Buckle. . . . 909

Corrigenda and Addendas

Correction: Auxiliary Teaching and Student Evaluation Methods Based on Facial Expression Recognition in Medical Education (e89669)
Xueling Zhu, Roben Juanatas. . . . 2738

Correction: Acceptability of a Digital Care App in Patients Undergoing Hip and Knee Arthroplasty: Prospective Cohort Study (e101097)
Yacine Louni, Matthew Laroche, Abdulrhman Alnasser, Mohammad Abuhaneya, Eric Belzile, Sandhya Baskaran, Jennifer Mutch, Anthony Albers. . . . 2739

Correction: Risk Factors for Noninitiation and Dropout in Blended Therapy in Inpatient Psychiatric Patients: Retrospective Cohort Study (e111245)
Nikita Gupta, Fabian Gardin, Thomas Berger, Wolfram Kawohl. . . . 2740

XSL•FO
RenderX

Correction: Quality of Informed Consent and Interface Usability in Primary Care e-Consultation: Cross-Sectional Study (e103250)
Caitlin Parfitt-Ford, Lisa Ballard, Adriane Chapman. 3301

Letters to the Editor

Authors' Reply: Empathy Cuts Both Ways in Clinical AI (e110739)
Sang Bae, Tongze Zhang, Tammy Chung, Anind Dey. 3163

Empathy Cuts Both Ways in Clinical AI (e108778)
Vera Sorin, Eyal Klang. 3165

Research Letters

Development and Initial Usability Evaluation of SSEndoSim, a Free, Web-Based Endodontic Diagnostic Simulator: Cross-Sectional Study (e100407)
Sarang Suresh, Priya Rani, Feroze Kalhoro. 3376

A Digital App-Based Intervention (Platoniq) to Reduce Loneliness: Single-Arm Pilot Study (e99374)
Silvan Hornstein, Jasmina Schmidt, Lara Sachs, Tobias Krieger, Daniel Stachnik, Ulrike Lueken. 3381

XSL•FO
RenderX

Anesthesia Providers' Perspectives on the Redesigned Philips Acoustic Alarm System: Qualitative Pre- and Postimplementation Study

Greta Gasciauskaite[1], MD; Stefanie Senn[2], MA; Cynthia Hunn[1], MD; Christoph B Nöthiger[1], MD; David W Tscholl[1], MD; Tadzio Raoul Roche[1], MD

[1]Institute of Anaesthesiology and Perioperative Medicine, University Hospital Zurich, University of Zurich, Raemistrasse 100, Zurich, Switzerland
[2]Clinic for Anaesthesiology, Winterthur Cantonal Hospital, Winterthur, Switzerland

Corresponding Author:
Tadzio Raoul Roche, MD
Institute of Anaesthesiology and Perioperative Medicine, University Hospital Zurich, University of Zurich, Raemistrasse 100, Zurich, Switzerland

Abstract

Background: Alarm fatigue caused by frequent or false alarms poses a persistent threat to patient safety. Despite technological progress, alarm acoustics remain largely unchanged and are often perceived as disruptive. To address this, Philips redesigned its patient monitoring alarm sounds through a user-centered approach aimed at improving priority differentiation and reducing emotional strain.

Objective: This study provides insights into human-technology interaction by examining anesthesia providers' experiences with the original and updated alarms, with a focus on emotional responses, usability, and guidance for the user-centered design of future clinical alarm systems.

Methods: This single-center qualitative study involved anesthesia providers who completed an online questionnaire before and after the implementation of the updated Philips alarms. Only those who completed the pre-implementation phase participated in the postimplementation phase. The questionnaire included 4 open-ended questions addressing perceptions of the current alarm sounds, suggestions for improvement, design expectations, and attitudes toward an alarm-free operating room. Responses were analyzed using thematic analysis to identify key usability and emotional response themes.

Results: A total of 90 eligible anesthesia providers participated in the preimplementation phase, and 77 (85.6%) participated in the postimplementation phase. Positive emotional responses increased in the postimplementation phase, whereas concerns regarding alarm functionality also became more prominent. Before the introduction of the updated alarm sounds, participants predominantly called for softer sounds. Following implementation, the most frequently expressed concern shifted to the need for clearer prioritization of alarms. Across both phases, the primary expectation remained the alarms' ability to effectively capture attention. The concept of an alarm-free operating room elicited concerns about increased workload and potential risks to patient safety.

Conclusions: The redesigned alarm sounds were perceived more positively in terms of emotional acceptance; however, they did not improve the recognition of alarm priority. The modest acoustic changes did not address the broader issue of alarm overload. Suggestions such as visual-only alerts for low-priority alarms show potential but must be balanced with patient safety standards. Future alarm development should combine user feedback with expert-driven and evidence-based approaches to improve both usability and clinical effectiveness.

(JMIR Hum Factors 2026;13:e82703) doi:10.2196/82703

KEYWORDS

anesthesia; auditory perception; clinical alarms; equipment design; patient monitoring; patient safety; sound; user-centered design; human-computer interaction

Introduction

Alarm sounds are essential safety mechanisms in modern clinical environments, designed to alert caregivers to potentially critical changes in a patient's condition [1]. However, high exposure to alarms, particularly those that are clinically irrelevant or false positives, can lead to alarm desensitization—a phenomenon known as alarm fatigue [2]. This is associated with delayed or missed responses to true alarms and represents a significant threat to patient safety [3,4]. Beyond the sheer volume of alarms, the lack of clarity and contextual relevance of alarm sounds

XSL•FO
RenderX

further exacerbates the problem [5]. Poorly differentiated alarm sounds or those lacking intuitive meaning may cause confusion among clinicians and impair the timely interpretation of patient status [6-8], particularly in already crowded acoustic environments where multiple devices compete for clinicians' attention.

Despite significant advances in patient monitoring technology, the acoustic design of alarms has seen little innovation over the past decades, despite extensive research in this area [9-11]. Historically, alarm sounds were designed to be as loud and attention-grabbing as possible, often by using auditory roughness and very salient, high-pitched frequencies [12]. The goal was to ensure that clinical staff would notice the alarms immediately. However, clinical staff often perceive these sounds as disruptive and emotionally taxing, with studies reporting that they are frequently experienced as stressful, difficult to interpret, and hard to distinguish from one another [8,13-15]. However, in many cases, subtle cue tones may be more appropriate as not all alarms require immediate action by health care professionals.

Philips (Koninklijke Philips N.V.), a leading manufacturer of patient monitors, has recognized these issues. In response, the company initiated a user-centered redesign process to revise the auditory characteristics of its monitor alarms. The updated system introduced changes to the acoustic cues with the aim of improving the distinction between alarm priorities and optimizing the balance between attention-grabbing and pleasant auditory stimulation [16]. The underlying logic of when and why alarms are triggered remained unchanged.

The redesigned alarm sounds attempt to address issues of poor clarity, limited contextual relevance, and negative emotional responses. However, it remains unclear whether such user-centered design changes translate into meaningful improvements in clinical practice. While the inclusion of end users in the design process is crucial, it may also limit innovation if feedback is primarily shaped by existing habits and expectations rather than by exposure to novel alarm paradigms [17]. More progressive approaches to alarm sonification, such as auditory icons or spoken cues that are easily distinguishable and convey information about the underlying cause of the alarm in addition to attracting attention, do not appear to have been considered in this redesign [9,10,18].

This qualitative study explored anesthesia providers' perceptions of patient monitoring alarms before and after the implementation of the redesigned Philips sounds. Specifically, we investigated emotional responses, perceived usability, and participants' visions for future alarm systems while reflecting on the broader implications for alarm design, innovation strategies, and patient safety.

Methods

Ethical Considerations

The Cantonal Ethics Committee Zurich reviewed the study protocol and issued a declaration of nonjurisdiction (Req-2022-00689) [19]. Before participation, written informed consent was obtained from all participants, granting permission for the use of the collected data for research purposes. Participation in the study was voluntary and without compensation. All data were deidentified prior to analysis. The data are presented in aggregated form only, and no conclusions about individual participants can be drawn from the published results.

Updated Philips Patient Monitor Alarm Sounds: Design and Implementation

While the design process behind the updated Philips alarm sounds was not part of this study, we briefly summarize its key objectives to provide context. The redesign followed a human-centered approach, combining qualitative methods (interviews and workshops) with quantitative surveys across 2 iterative development phases. During the original design process, anesthesia professionals (including nurses, residents, and attending physicians) from 7 countries contributed their perspectives on what constitutes an improved alarm sound.

Participants consistently highlighted the need to reduce the harshness and unpleasantness of alarm cues, expressing a preference for sounds that are more pleasant, natural, and less stressful. They also emphasized the importance of clearer differentiation between alarm priority levels and better distinction from other clinical acoustic signals.

In the second iteration, participants were asked to evaluate and select from a range of newly developed alarm sound options. On the basis of this feedback, the sounds were revised as follows: low-priority alarms were softened and spaced with longer tone intervals, medium-priority alarms were given distinctive pitch and timing characteristics, and high-priority alarms were adjusted to reduce shrillness while maintaining urgency. A more detailed description of the design process can be found in the publication by Hunn et al [20]. Both the original and updated alarm sounds are available for comparison in Multimedia Appendix 1.

Study Design

This investigator-initiated, single-center qualitative descriptive study was conducted at University Hospital Zurich, Switzerland, in 2 phases: before (January 2024 to March 2024) and after (August 2024 to October 2024) the implementation of the updated Philips patient monitoring alarm sounds. The aim was to explore anesthesia providers' (trainee nurses, nurses, resident physicians, and attending physicians) perceptions of the original and updated alarm sounds following their implementation and a subsequent adaptation period of 4 months.

Before implementation, clinical staff participated in a 1-hour training session to familiarize themselves with the new alarm system.

An online questionnaire consisting of both open- and closed-ended questions was completed by participants as part of a face-to-face study procedure. In the prephase, participants were eligible if they worked regularly in the operating room and were familiar with the existing alarm sounds; no additional inclusion or exclusion criteria were applied. Only participants who had taken part in the prephase took part in the postphase.

XSL•FO
RenderX

Survey Questionnaire

The questionnaire was developed using Google Forms and administered to participants at 2 time points: once during the prephase (original Philips alarms) and again during the postphase (updated Philips alarms). It comprised 2 main sections.

The first section included 4 open-ended questions designed to elicit free-text responses regarding participants' perceptions of current patient monitoring alarm sounds, suggested improvements and their underlying reasoning, expectations for alarm cue design, and vision of an operating room without acoustic alarms.

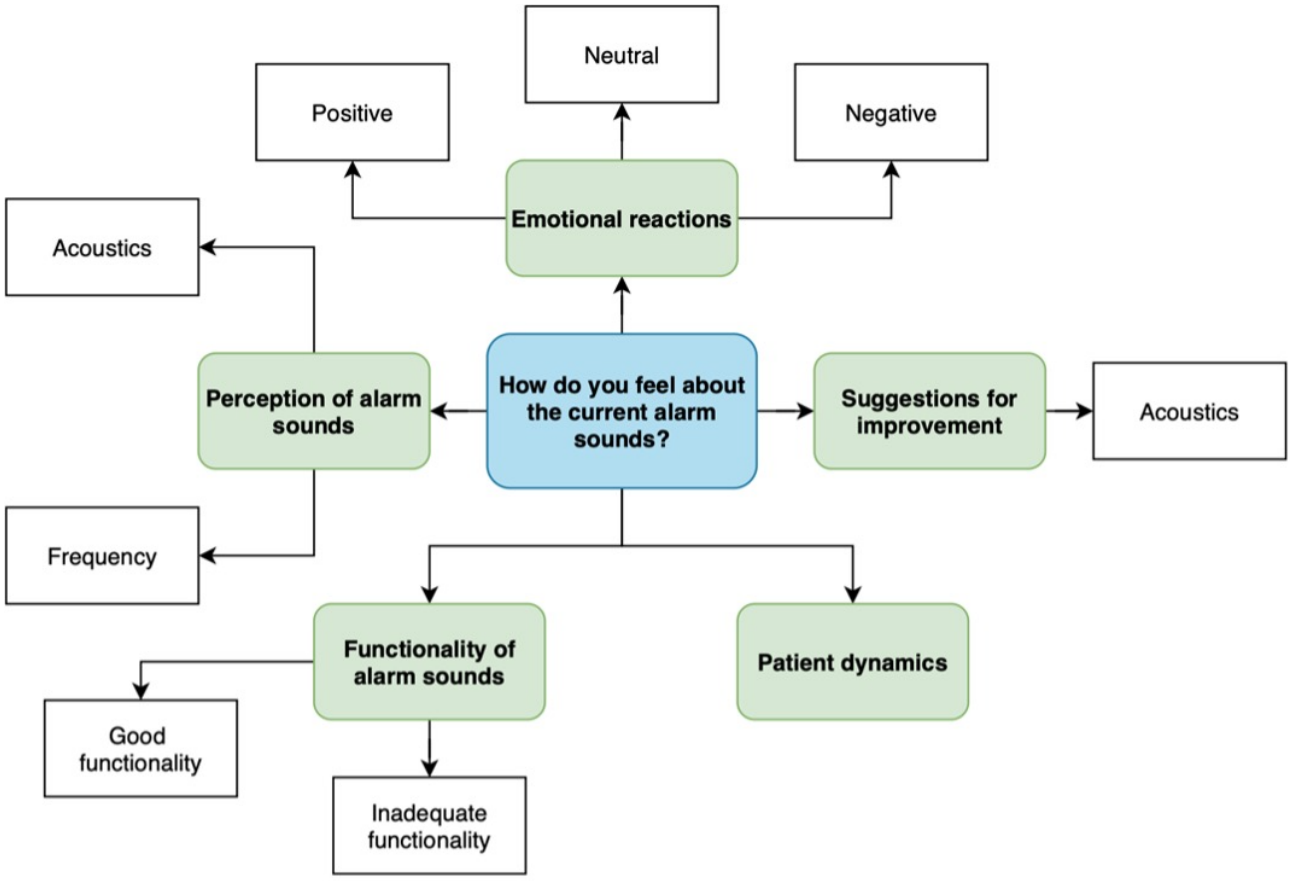
The second section contained 4 closed-ended questions to collect demographic data (age, gender, years of experience in anesthesia, and professional role) during both phases. In the prephase only, 2 additional questions were included to assess whether participants regularly played a musical instrument and whether they had attended an introductory lecture on the redesigned alarm sounds before their clinical implementation.

Data Analysis

A 6-step thematic analysis method was used to examine the open-ended responses and identify dominant themes [21]. After thoroughly reviewing the data, the research team convened to share impressions and explore potential thematic concepts. On the basis of this discussion, 4 coding templates were developed for each question to facilitate further analysis (Figures 1-4). Participants' responses were independently coded by 3 team members (GG, SS, and TRR). Any discrepancies in coding were resolved by the third coder (GG).

Closed-ended responses were analyzed using Microsoft Excel and presented as numerical values with percentage distributions or as medians with IQRs and minimum and maximum values.

Figure 1. Coding tree illustrating the main themes and subthemes identified when participants were asked how they felt about the current alarm sounds.



XSL•FO
RenderX

Figure 2. Coding tree illustrating the main themes and subthemes identified when participants were asked what they would change about the current alarm sounds and why.

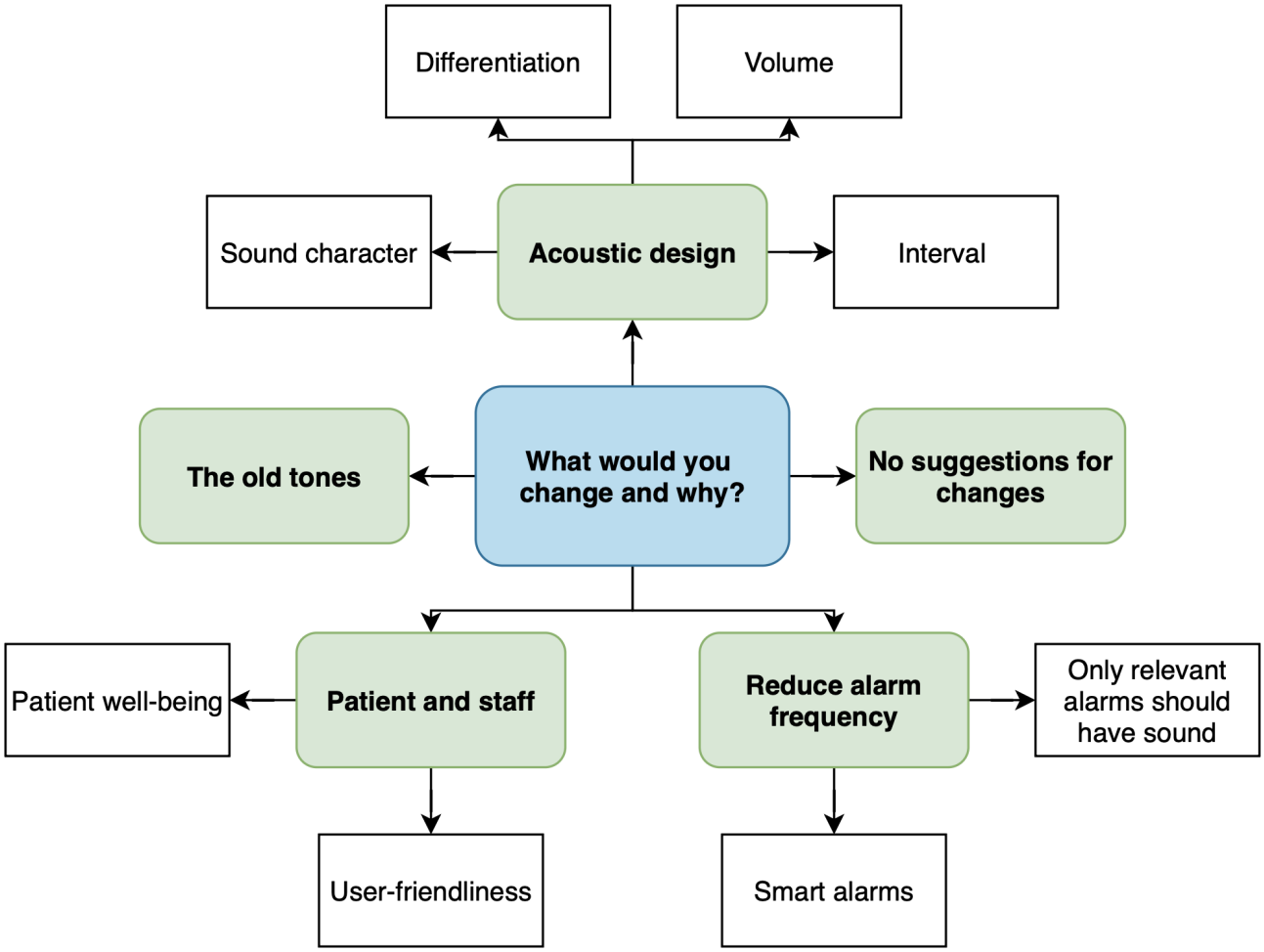
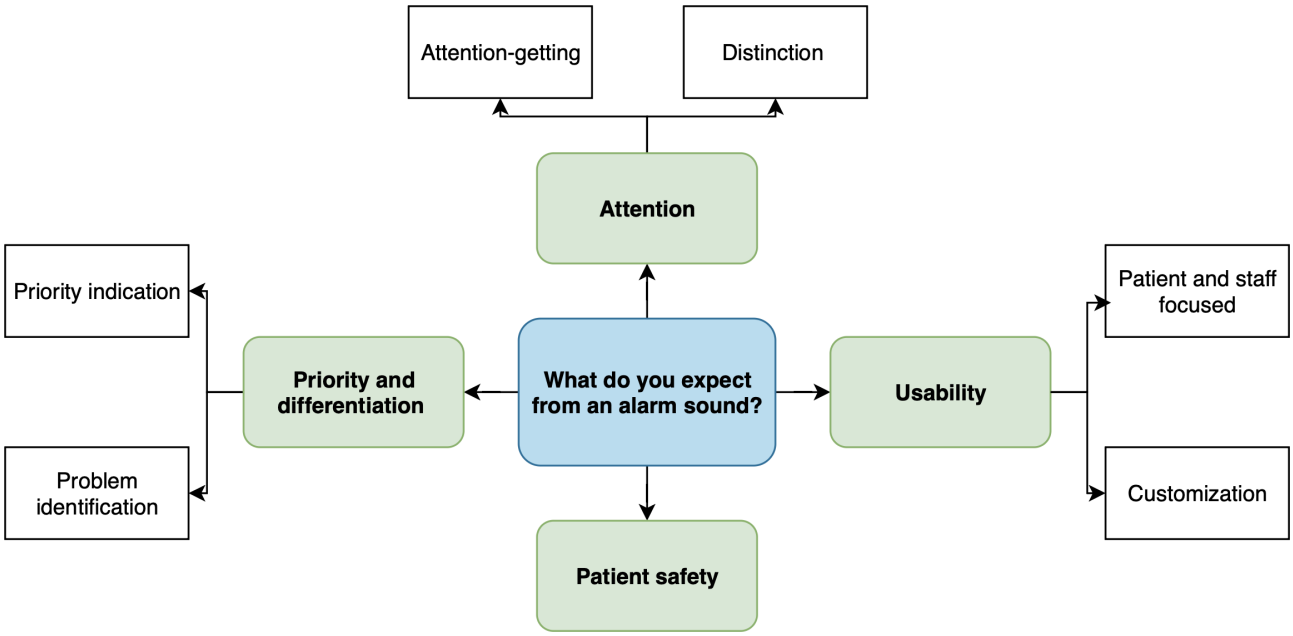
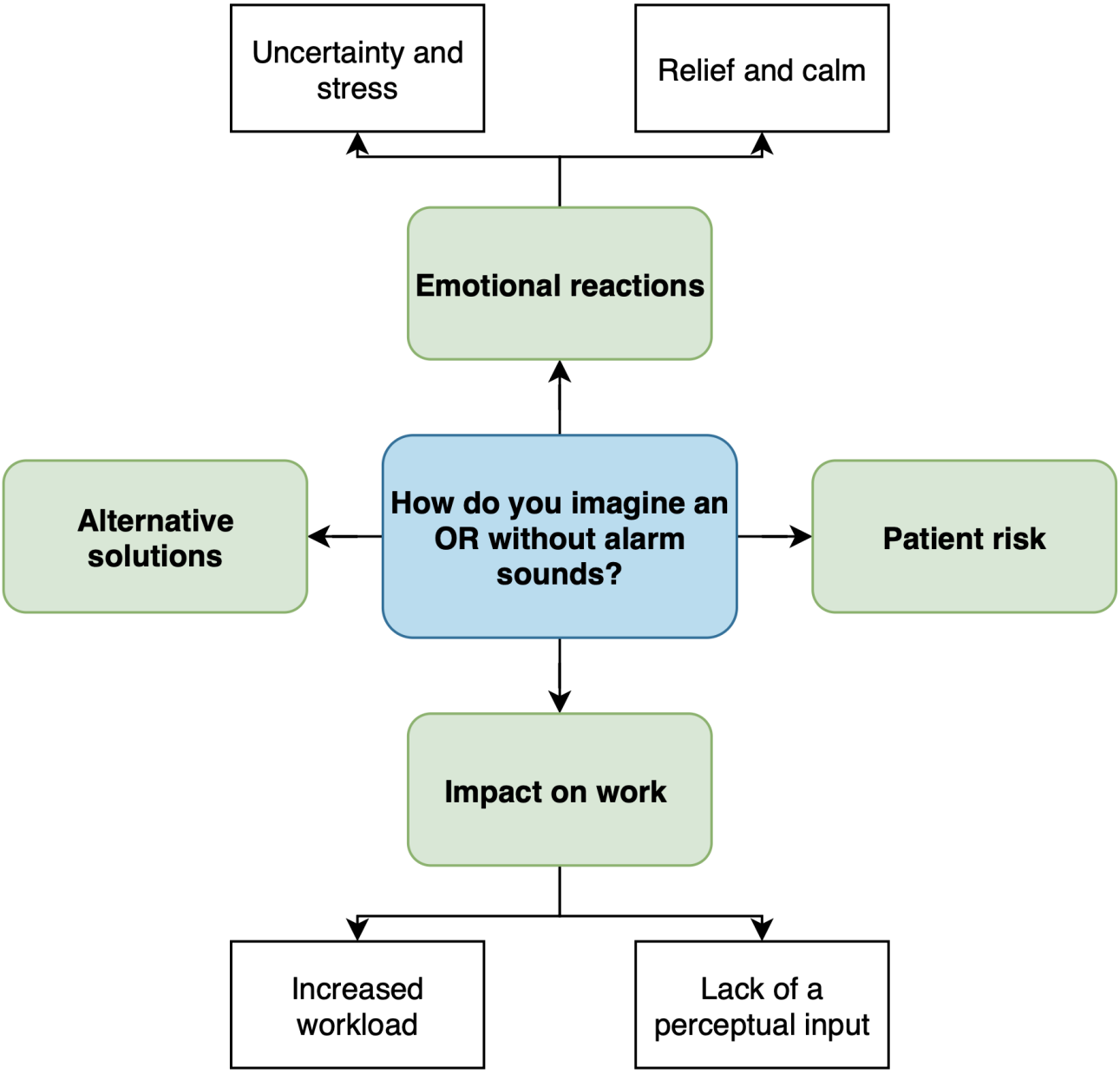


Figure 3. Coding tree illustrating the main themes and subthemes identified when participants were asked what they expected from an alarm sound.



XSL•FO
RenderX

Figure 4. Coding tree illustrating the main themes and subthemes identified when participants were asked how they imagined an operating room (OR) without alarm sounds.



Results

Participant and Study Characteristics

Of the 90 anesthesia providers initially contacted, all (n=90, 100%) completed the questionnaire during the prephase. In the postphase, 85.6% (77/90) of the potential participants submitted the questionnaire, as 14.4% (13/90) were unavailable due to a change in employment. A total of 888 comments were collected: 509 (57.3%) during the prephase and 379 (42.7%) during the postphase. A detailed study and participant characteristics are provided in Table 1.

XSL•FO
RenderX

Table . Participant and study characteristics.

	Values	
	Prephase (n=90)[a]	Postphase (n=77)
Participants who completed the questionnaire, n (%)	90 (100)	77 (100)
Collected comments, n	509	379
Age (y), median (IQR; range)	34 (30-41; 23-61)	34 (30-42; 24-62)
Sex, n (%)		
Female	45 (50)	34 (44.2)
Male	45 (50)	43 (55.8)
Experience (y), median (IQR; range)	5 (2.3-11; 0-40)	6 (2.8-12.5; 0-35)
Role, n (%)		
Nurse anesthetist in training	5 (5.6)	2 (2.6)
Certified nurse anesthetist	23 (25.6)	22 (28.6)
Resident—1-2 y of training	16 (17.8)	14 (18.2)
Resident—3-5 y of training	17 (18.9)	12 (15.6)
Resident—>5 y of experience	8 (8.9)	6 (7.8)
Staff anesthesiologist	21 (23.3)	21 (27.3)
"I play a musical instrument regularly (once or more per week)," n (%)	13 (14.4)	—[b]
"I attended the introductory lecture," n (%)	39 (43.3)	—

[a]A total of 90 participants were contacted in both phases.

[b]The question was not repeated in the postphase.

Main Themes From Responses to Open-Ended Questions: Impressions Before and After Implementing the Updated Philips Patient Monitor Alarm Sounds

Overview

Table 2 presents the themes and subthemes corresponding to the coding trees (Figures 1-4), as identified in participants' responses, along with illustrative examples and their numerical and percentage distributions.

XSL•FO
RenderX

Table . Identified themes and subthemes in participants' responses with examples, numerical values, and percentage distributions.

Theme and subthemes	Prephase comments, n/N (%)	Postphase comments, n/N (%)	Examples	
			Prephase (original alarms)	Postphase (updated alarms)
Question: "How do you feel about the current alarm sounds?"				
Emotional reactions				
Positive	37/139 (26.6)	46/89 (51.7)	"I am satisfied with the current alarm sounds as they help me alert to dangerous situations." [Participant 72]	"I feel less stressed with the new alarm sounds." [Participant 85]
Neutral	21/139 (15.1)	3/89 (3.4)	"Familiar alarm sounds that are well integrated into my daily workflow." [Participant 46]	"I have gotten used to it." [Participant 55]
Negative	30/139 (21.6)	7/89 (7.9)	"The more frequently and longer they sound, the more it increases my stress level while working." [Participant 22]	"I am familiar with the old sounds and prefer them." [Participant 73]
Perception of alarm sounds				
Acoustics	9/139 (6.5)	2/89 (2.2)	"The sounds are very high-pitched." [Participant 27]	"The new sounds are more high-pitched, otherwise similar to the previous ones." [Participant 22]
Frequency	10/139 (7.2)	0/89 (0)	"High-priority alarms sound too often." [Participant 85]	—[a]
Functionality of alarm sounds				
Good functionality	9/139 (6.5)	9/89 (10.1)	"I perceive them as a high level of security, and they allow me to focus on the surgical tasks as well." [Participant 38]	"The new alarm sounds capture appropriate attention." [Participant 75]
Inadequate functionality	16/139 (11.5)	20/89 (22.5)	"The alarm sounds with 'low priority' are often not noticed." [Participant 87]	"I find that the new alarm sounds often go unnoticed in hectic situations." [Participant 84]
Patients				
Patient dynamics	5/139 (3.6)	2/89 (2.2)	"I think that the alarm sounds are very unsettling for the patients." [Participant 66]	"For the patients, the new alarm sounds are certainly much more pleasant and less startling." [Participant 3]
Suggestions for improvement				
Acoustics	1/139 (0.7)	0/89 (0)	"The alarm sounds could be slightly lower in pitch." [Participant 24]	—
Question: "What would you change and why?"				
No suggestions for changes[b]	26/112 (23.2)	22/84 (26.2)	"I would not change anything. The alarms fit into my workflows." [Participant 75]	"I do not want to change anything." [Participant 90]
Acoustic design				
Sound character	26/112 (23.2)	7/84 (8.3)	"I would wish for softer tones. The saturation tone drops too quickly to a lower pitch. From 100% to 98%, it startles me." [Participant 33]	"The new alarm sounds are more pleasant." [Participant 82]

XSL•FO
RenderX

Theme and subthemes	Prephase comments, n/N (%)	Postphase comments, n/N (%)	Examples	
			Prephase (original alarms)	Postphase (updated alarms)
Differentiation	11/112 (9.8)	27/84 (32.1)	"Alarms that indicate high priority should stand out more from the 'standard tones.' A better differentiation for emergencies to quickly grab the attention of the team." [Participant 46]	"Low- and medium-priority alarm sounds are difficult to distinguish." [Participant 70]
Volume	5/112 (4.5)	4/84 (4.8)	"The volume should gradually increase." [Participant 29]	"Low-level alarms could be louder and more attention-grabbing." [Participant 65]
Interval	10/112 (8.9)	6/84 (7.1)	"For lower alarm priorities, the frequency should be lower." [Participant 6]	"I would like to increase the frequency of the middle tone a little." [Participant 89]
Reduction of alarm frequency				
Smart alarms	7/112 (6.3)	2/84 (2.4)	"Filter settings and auto-limits; more 'intelligence' in the alarm sounds." [Participant 20]	"Reduce the Apnoea alarm during anaesthesia induction from high alarm priority to medium priority." [Participant 29]
Only relevant alarms should have sound	17/112 (15.2)	8/84 (9.5)	"Most alarms are artifacts or clinically irrelevant." [Participant 71]	"I only need two alarm priorities; technical alarms (low priority) are unnecessary." [Participant 52]
Patient and staff				
Patient well-being	6/112 (5.4)	2/84 (2.4)	"For patients in the induction phase, the alarm sounds are very startling and sometimes frightening." [Participant 51]	"Some patients found the sound unpleasant and asked if it would be heard the entire time." [Participant 53]
User-friendliness	4/112 (3.6)	3/84 (3.6)	"Easily acknowledgeable alarms." [Participant 22]	"I would like a simpler way to silence the alarm tone." [Participant 46]
The old tones[b]	—	3/84 (3.6)	—	"The old alarm tones." [Participant 53]
Question: "What do you expect from an alarm sound?"				
Attention				
Attention grabbing	49/134 (36.6)	46/104 (44.2)	"A good alarm tone grabs attention and does not allow for ignoring." [Participant 5]	"Direct attention to the monitor and to the patient." [Participant 22]
Distinction	22/134 (16.4)	11/104 (10.6)	"A good alarm tone must be easily distinguishable from other sounds." [Participant 14]	"Important alarms should stand out. A low tone is distracting." [Participant 2]
Priority and differentiation				
Priority indication	18/134 (13.4)	15/104 (14.4)	"I expect a simple classification of the alarm's importance (i.e., distinction of priority)." [Participant 66]	"The ability to assess the alarm priority without directly looking at the monitor." [Participant 4]
Problem identification	14/134 (10.4)	12/104 (11.5)	"I expect a precise differentiation of the problem." [Participant 81]	"The corresponding meaning of the tone and its consequences are important." [Participant 5]
Usability				

XSL•FO
RenderX

Theme and subthemes	Prephase comments, n/N (%)	Postphase comments, n/N (%)	Examples	
			Prephase (original alarms)	Postphase (updated alarms)
Patient and staff focused	24/134 (17.9)	17/104 (16.3)	"It should convey the information but be pleasant." [Participant 31]	"The alarm tones should fulfil their purpose and not create additional stress." [Participant 87]
Customization	5/134 (3.7)	2/104 (1.9)	"Alarm only when there is a real clinical emergency. Do not alarm for any technical issues." [Participant 40]	"Alarm tones should be adjustable to the individual patient." [Participant 3]
Patient safety[b]	2/134 (1.5)	1/104 (1.0)	"Safety for the patient." [Participant 33]	"Patient safety." [Participant 3]
Question: "How do you imagine an operating room without alarm sounds?"				
Emotional reactions				
Uncertainty and stress	49/124 (39.5)	37/102 (36.3)	"It would be a very large burden for me." [Participant 38]	"It is not possible; it cannot be done." [Participant 26]
Relief and calm	29/124 (23.4)	19/102 (18.6)	"A more pleasant work atmosphere where one can stay focused for longer periods." [Participant 46]	"A calm, pleasant working atmosphere." [Participant 3]
Impact on work				
Increased workload	8/124 (6.5)	10/102 (9.8)	"It is exhausting for anaesthesiologists, as the monitor must be constantly checked." [Participant 66]	"It would make everyday tasks significantly more difficult, as it would require more frequent visual checks of the vital parameters, potentially preventing full concentration on other tasks." [Participant 84]
Lack of a perceptual input	16/124 (12.9)	6/102 (5.9)	"I am used to working a lot with my hearing, and it helps in an emergency to quickly assess and focus on the patient." [Participant 43]	"No focus because you would have to constantly look at the monitor." [Participant 62]
Alternative solutions[b]	4/124 (3.2)	6/102 (5.9)	"Visual alarms, colours on the monitor." [Participant 42]	"Further exploration/research into new ways of presenting alarms (e.g., tactile/visual) would definitely be advisable, to possibly achieve this in the near or somewhat distant future. It would certainly be desired." [Participant 68]
Patient risk[b]	18/124 (14.5)	24/102 (23.5)	"However, patients would likely be at risk." [Participant 11]	"I believe that would pose a risk to patient safety; I consider the alarm tones essential for staying focused, especially during long surgeries." [Participant 89]

[a]Not applicable.

[b]Distinct themes without subthemes.

How Do You Feel About the Current Alarm Sounds?

We identified 2 main themes in participants' responses: emotional reactions and alarm sounds' functionality. Before the implementation of the updated Philips patient monitoring alarm sounds, 26.6% (37/139) of the collected statements reported a positive emotional response. After the update, this proportion increased to 51.7% (46/89), indicating a notable improvement in the perceived emotional acceptance of alarm sounds. However, concerns about the functionality of alarm sounds remained prevalent across both phases, with related issues

XSL•FO
RenderX

reported in 11.5% (16/139) of preimplementation responses and 22.5% (20/89) of postimplementation responses, making it the second most frequently mentioned theme.

What Would You Change and Why?

The main themes identified in participants' responses differed slightly between the pre- and postimplementation phases. In the prephase, the most frequently mentioned themes were no suggestions for changes (26/112, 23.2%), sound character (specifically the need to reduce high-pitched tones and make sounds softer and more harmonious; 26/112, 23.2%), and a suggestion to present only high-priority alarms with sound while representing low-priority alarms visually (17/112, 15.2%).

In the postphase, the dominant theme was the difficulty in distinguishing alarm sounds based on priority (27/84, 32.1%), highlighting the need to make tones for different priority levels (low, medium, and high) more distinct. The second most frequently mentioned theme, as in the prephase, was that no changes were required (22/84, 26.2%).

What Do You Expect From an Alarm Sound?

The responses in the pre- and postimplementation phases differed slightly. In both phases, expectations focused primarily on an alarm sound's ability to capture attention (49/134, 36.6% in the prephase vs 46/104, 44.2% in the postphase), emphasizing the need for effectiveness even in noisy environments. The second most frequently mentioned theme in both phases was priority indication (18/134, 13.4% in the prephase vs 15/104, 14.4% in the postphase), meaning that an alarm sound should clearly convey the urgency of a situation.

In the prephase, another dominant theme was the distinction of alarm sounds from other sounds (22/134, 16.4%). In contrast, in the postphase, a new theme emerged: usability in terms of alarms addressing both staff and patient needs (17/104, 16.3%), meaning that they should be easily perceptible for staff without causing stress for patients.

How Do You Imagine an Operating Room Without Alarm Sounds?

The same themes dominated participants' responses to this question: uncertainty and stress due to increased effort and strain from the lack of auditory cues (49/124, 39.5% in the prephase vs 37/102, 36.3% in the postphase), relief and calm associated with reduced noise and a more pleasant work environment (29/124, 23.4% in the prephase vs 19/102, 18.6% in the postphase), and patient risk as the absence of alarm sounds may lead to overlooked critical situations or delayed reactions to changes (18/124, 14.5% in the prephase vs 24/102, 23.5% in the postphase).

Discussion

In this investigator-initiated, single-center qualitative descriptive survey study, we explored anesthesia providers' perceptions of patient monitoring alarm sounds before and after the clinical introduction of the redesigned Philips alarms.

While the updated alarm sounds were associated with a greater proportion of positive emotional responses among the study participants, the perceived issue of clearly distinguishing between alarm priorities persisted. Before the update, participants predominantly reported difficulties in differentiating between medium- and high-priority alarms. However, after the update, they indicated increased difficulty distinguishing between low- and medium-priority alarms.

Alternative strategies, such as user training for alarm threshold adjustment; reduction of overmonitoring [2,22]; and advanced technical solutions, such as smart alarms, multiparameter integration, and short-delay algorithms, have shown greater promise in reducing alarm fatigue than alarm design alone [23]. In addition, innovative acoustic approaches, including auditory icons, spearcons, and voice alerts, may further enhance alarm systems. Studies suggest that these formats improve speed and accuracy of interpretation while lowering cognitive load [9,24-26]. Importantly, the International Electrotechnical Commission (IEC) explicitly recommended the use of auditory icons in its revised global standard for medical equipment alarms (IEC 60601-1-8) published in 2020 [27].

Beyond acoustic design, the ergonomics of the patient bedside are also important. Limited visibility of alarm triggers may lead clinicians to rely on multiple auditory cues from multiple devices. Optimizing screen layout, integrating networked devices with hierarchical alarms, and using smart display strategies such as a patient avatar could reduce cognitive load and improve timely responses [28]. These ergonomic and acoustic considerations provide a useful framework for understanding how clinical staff actually perceive and interact with alarm sounds in practice.

In our study, participants primarily perceived the role of acoustic alarms as drawing attention to the patient monitor rather than conveying specific clinical information. It is noteworthy that participants did not suggest innovative ways in which acoustic alarms could communicate specific deviations in vital signs. Similarly, established alternative sonification approaches, as discussed in the literature [29] and reflected in the IEC 60601-1-8 international standard, were not referenced in the responses. This may suggest limited awareness or perceived relevance in the clinical context [27]. This observation may be explained by the innovation paradox inherent to user-centered design: users tend to articulate needs and preferences based on familiar concepts and experiences [17]. Thus, if design processes rely exclusively on user feedback, opportunities for more radical or transformative innovations may be overlooked. Therefore, to foster meaningful advancements in alarm systems, user-centered design should be complemented with expert-driven innovation strategies and evidence-based frameworks.

This study also explored the radical notion of eliminating acoustic alarms altogether. Surprisingly, only a small proportion of participants considered this to be a significant risk to patient safety, despite the fact that acoustic alarms are mandated as essential patient safety tools in various guidelines and industry standards [1,4,7]. This finding underscores the need for ongoing education about the safety functions of alarms and the complex trade-offs between alarm fatigue and patient risk.

XSL•FO
RenderX

One particularly interesting and easily implementable idea was repeatedly mentioned by participants: anesthesia providers suggested eliminating the auditory presentation of low-priority alarms, opting instead for visual-only indications. A closer examination reveals that low-priority alarms in Philips patient monitors predominantly signal technical issues unrelated to the patient's clinical status and typically do not require intervention of the caregiver. We believe that, depending on the clinical context—whether intensive care units, operating rooms, recovery rooms, or telemetry wards—the function and role of alarms vary significantly. Different rooms, different rules: what works in the operating room may not suit intensive care. User feedback is invaluable for evaluating these different use profiles. However, any changes to alarm management strategies, particularly the reduction of acoustic cues, must be carefully evaluated against existing safety guidelines and subjected to thorough risk assessments to ensure that patient safety is not compromised.

This study has several notable strengths. It was conducted in a real-world clinical environment at a large tertiary university hospital, enhancing the external validity and practical relevance of the findings [30]. The pretest-posttest design allowed for a direct comparison of perceptions before and after the implementation of the updated Philips patient monitoring alarms [31]. Furthermore, open-ended responses were systematically analyzed using a well-established 6-step thematic analysis method [21], with independent coding by multiple researchers to ensure analytical rigor and minimize bias. The provision of original study materials, including the questionnaire and multimedia supplements, further strengthens the transparency and reproducibility of the research.

Nonetheless, several limitations must be acknowledged. The single-center pretest-posttest design without a control group precludes causal inference, does not account for time-dependent effects, and limits generalizability [30]. However, the within-subject design, standardized training, and extended adaptation period were chosen to enhance internal and ecological validity for the evaluation of alarm sound perception in a real-world clinical setting. Participation was voluntary, which may have introduced selection bias, as individuals with greater interest in alarm systems or new technologies might have been more likely to participate. The questionnaire, although carefully designed, was not externally validated, potentially affecting the reliability of the findings. It should also be noted that while the patient monitor was evaluated in a clinical setting in the operating room, the study did not assess the monitor within the full operating room soundscape, which may influence real-world perception. Finally, perceptions were self-reported and, therefore, subject to biases such as social desirability or recall bias. Moreover, while improvements in emotional acceptance of alarm sounds were achieved, it is critical to emphasize that positive emotional responses alone do not guarantee enhanced patient safety. Functional effectiveness, particularly the rapid and accurate identification of alarm priorities, remains paramount.

Despite these limitations, this study provides valuable insights into the user experience with updated patient monitoring alarms and highlights important considerations for future alarm design, development, and evaluation efforts. When developing new technological solutions to reduce alarm frequency and false positives and improve the distinguishability and intuitive meaning of alarm signals, it is crucial not to rely solely on user-centered design [17]. Current evidence [9,24,25], international standards [27], and expert input must also be incorporated to achieve an optimal balance among usability, safety, and innovation. Future patient monitoring alarm systems could benefit from a multimodal approach that integrates conventional alarm sounds with auditory icons, voice alerts, and advanced visual displays to better support clinical decision-making and effectively mitigate alarm fatigue for both caregivers and patients [29].

Acknowledgments

The authors are grateful to the participants of the study for their investment and time. During the preparation of this work, the authors used GPT-4o (OpenAI) to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

Philips provided support with personnel for the activation of the updated alarm sounds. The licensing of the sounds was carried out independently by University Hospital Zurich and was not related to this study. The research that led to these results was supported by the Institute for Anesthesiology and Perioperative Medicine of University Hospital Zurich with funds from the University of Zurich.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files (Multimedia Appendix 2).

Authors' Contributions

Conceptualization: CH, CBN, DWT, TRR
Data curation: GG, SS, TRR
Formal analysis: GG, SS, TRR

XSL•FO
RenderX

Investigation: CH, CBN, TRR
Methodology: CH, CBN, DWT, TRR
Project administration: CBN, DWT, TRR
Resources: CBN, DWT, TRR
Software: CH
Supervision: CBN, DWT, TRR
Visualization: GG
Writing—original draft: GG, TRR
Writing—review and editing: GG, SS, CH, CBN, DWT, TRR

Conflicts of Interest

CBN and DWT are among the inventors of the Visual Patient Avatar and Visual Patient Predictive technologies, for which the University of Zurich and Philips hold various patents, design rights, and trademarks. CBN, DWT, and TRR are listed as inventors of Visual Patient Predictive. These technologies are being developed jointly with Philips companies in Germany, the Netherlands, and the United States under agreements that provide CBN and DWT with research funding, travel support, consulting and lecturing honoraria, and royalties. TRR receives research funding, travel support, and honoraria and may receive royalties if Visual Patient Predictive is successfully commercialized. DWT also serves on the Philips Patient Safety Advisory Board. CH, CBN, and DWT are inventors of Visual Clot, and CBN, DWT, and TRR are inventors of Visual Blood—technologies for which the University of Zurich holds related intellectual property. Royalties may be awarded to the inventors if these technologies are commercialized. In addition, DWT is the inventor of Visual Hemofilter and Visual Heart, with similar intellectual property held by the University of Zurich and potential royalties pending commercialization. DWT has also received travel support and honoraria from Instrumentation Laboratory – Werfen (United States), the Swiss Foundation for Anesthesia Research (Switzerland), and the International Symposium on Intensive Care and Emergency Medicine (Belgium). None of the authors have any financial interest in the alarm sounds evaluated in this study.

Multimedia Appendix 1
Original and updated Philips patient monitoring alarm sounds.
[PPTX File, 14209 KB - humanfactors_v13i1e82703_app1.pptx]

Multimedia Appendix 2
Data generated and analyzed during this study.
[XLSX File, 172 KB - humanfactors_v13i1e82703_app2.xlsx]

References

1. WHO Guidelines for Safe Surgery 2009: Safe Surgery Saves Lives: World Health Organization; 2009.
2. Ruskin KJ, Hueske-Kraus D. Alarm fatigue: impacts on patient safety. Curr Opin Anaesthesiol 2015 Dec;28(6):685-690. [doi: 10.1097/ACO.0000000000000260] [Medline: 26539788]
3. Sendelbach S, Funk M. Alarm fatigue: a patient safety concern. AACN Adv Crit Care 2013;24(4):378-386. [doi: 10.1097/NCI.0b013e3182a903f9] [Medline: 24153215]
4. National Performance Goals Effective January 2026 for the Critical Access Hospital Program. The Joint Commission. 2025. URL: https://digitalassets.jointcommission.org/api/public/content/0e0ed74ecb53426c93fbcf9362ab003a?v=97b90dfc [accessed 2026-02-27]
5. Mumaw RJ, Roth EM, Patterson ES. Lessons from the glass cockpit: innovation in alarm systems to support cognitive work. Biomed Instrum Technol 2021 Jan 1;55(1):29-40. [doi: 10.2345/0899-8205-55.1.29] [Medline: 33662139]
6. Atyeo J, Sanderson PM. Comparison of the identification and ease of use of two alarm sound sets by critical and acute care nurses with little or no music training: a laboratory study. Anaesthesia 2015 Jul;70(7):818-827. [doi: 10.1111/anae.13020] [Medline: 25735731]
7. Sanderson PM, Wee A, Lacherez P. Learnability and discriminability of melodic medical equipment alarms. Anaesthesia 2006 Feb;61(2):142-147. [doi: 10.1111/j.1365-2044.2005.04502.x] [Medline: 16430567]
8. McNeer RR, Bennett CL, Dudaryk R. Intraoperative noise increases perceived task load and fatigue in anesthesiology residents: a simulation-based study. Anesth Analg 2016 Feb;122(2):512-525. [doi: 10.1213/ANE.0000000000001067] [Medline: 26797555]
9. Roche TR, Braun J, Ganter MT, et al. Voice alerting as a medical alarm modality for next-generation patient monitoring: a randomised international multicentre trial. Br J Anaesth 2021 Nov;127(5):769-777. [doi: 10.1016/j.bja.2021.07.015] [Medline: 34454710]
10. McNeer RR, Horn DB, Bennett CL, Edworthy JR, Dudaryk R. Auditory icon alarms are more accurately and quickly identified than current standard melodic alarms in a simulated clinical setting. Anesthesiology 2018 Jul;129(1):58-66. [doi: 10.1097/ALN.0000000000002234] [Medline: 29698253]

XSL•FO
RenderX

11. Edworthy J. Medical audible alarms: a review. J Am Med Inform Assoc 2013 May 1;20(3):584-589. [doi: 10.1136/amiajnl-2012-001061] [Medline: 23100127]
12. Burdick KJ, Yang S, Lopez AE, Wessel C, Schutz M, Schlesinger JJ. Auditory roughness: a delicate balance. Br J Anaesth 2023 Oct;131(4):649-652. [doi: 10.1016/j.bja.2023.07.003] [Medline: 37537119]
13. Ryherd EE, Waye KP, Ljungkvist L. Characterizing noise and perceived work environment in a neurological intensive care unit. J Acoust Soc Am 2008 Feb;123(2):747-756. [doi: 10.1121/1.2822661] [Medline: 18247879]
14. Honan L, Funk M, Maynard M, Fahs D, Clark JT, David Y. Nurses' perspectives on clinical alarms. Am J Crit Care 2015 Sep;24(5):387-395. [doi: 10.4037/ajcc2015552] [Medline: 26330431]
15. Lacherez P, Seah EL, Sanderson P. Overlapping melodic alarms are almost indiscriminable. Hum Factors 2007 Aug;49(4):637-645. [doi: 10.1518/001872007X215719] [Medline: 17702215]
16. Philips sounds technology. Philips. URL: https://www.philips.co.uk/healthcare/technology/alarm-sounds [accessed 2026-02-27]
17. van Velsen L, Ludden G, Grünloh C. The limitations of user-and human-centered design in an eHealth context and how to move beyond them. J Med Internet Res 2022 Oct 5;24(10):e37341. [doi: 10.2196/37341] [Medline: 36197718]
18. Li SY, Tang TL, Hickling A, Yau S, Brecknell B, Sanderson PM. Spearcons for patient monitoring: laboratory investigation comparing earcons and spearcons. Hum Factors 2017 Aug;59(5):765-781. [doi: 10.1177/0018720817697536] [Medline: 28570832]
19. Federal Act on Research Involving Human Beings (Human Research Act, HRA) of September 30, 2011. Fedlex. URL: https://www.fedlex.admin.ch/eli/cc/2013/617/en [accessed 2026-03-13]
20. Hunn C, Nöthiger CB, Braun J, et al. User-centered redesign of monitoring alarms: a pre-post study on perception, functionality, and recognizability following real-life clinical implementation. Healthcare (Basel) 2025 Nov 24;13(23):3033. [doi: 10.3390/healthcare13233033] [Medline: 41373249]
21. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
22. Bosma S, Christopher R. Implementing a unit-based alarm management bundle for critical care nurses. Crit Care Nurse 2023 Apr 1;43(2):36-45. [doi: 10.4037/ccn2023418] [Medline: 37001877]
23. Cvach M. Monitor alarm fatigue: an integrative review. Biomed Instrum Technol 2012;46(4):268-277. [doi: 10.2345/0899-8205-46.4.268] [Medline: 22839984]
24. Edworthy JR, Talbot N, Martin N. Responding to clinical alarms in unfolding simulated clinical scenarios: auditory icons perform better than tonal alarms. Br J Anaesth 2025 Jun;134(6):1773-1778. [doi: 10.1016/j.bja.2024.12.047] [Medline: 40118671]
25. Sanderson PM, Brecknell B, Leong S, et al. Monitoring vital signs with time-compressed speech. J Exp Psychol Appl 2019 Dec;25(4):647-673. [doi: 10.1037/xap0000217] [Medline: 30883150]
26. Bennett C, Dudaryk R, Crenshaw N, Edworthy J, McNeer R. Recommendation of new medical alarms based on audibility, identifiability, and detectability in a randomized, simulation-based study. Crit Care Med 2019 Aug;47(8):1050-1057. [doi: 10.1097/CCM.0000000000003802] [Medline: 31135498]
27. IEC 60601- 1-8:2006/AMD 2: 2020: medical electrical equipment—part 1–8: general requirements for basic safety and essential performance— collateral standard: general requirements, tests and guidance for alarm systems in medical electrical equipment and medical electrical systems. International Organization for Standardization. 2020. URL: https://www.iso.org/standard/72634.html [accessed 2026-02-27]
28. Gasciauskaite G, Lunkiewicz J, Roche TR, Spahn DR, Nöthiger CB, Tscholl DW. Human-centered visualization technologies for patient monitoring are the future: a narrative review. Crit Care 2023 Jun 28;27(1):254. [doi: 10.1186/s13054-023-04544-0] [Medline: 37381008]
29. Ebensperger M, Roche TR, Rahrisch A, Webster CS, Tscholl DW. The future of sound in patient monitoring: a call to action to embrace the evidence and move forward. Br J Anaesth 2025 Jun;134(6):1616-1619. [doi: 10.1016/j.bja.2025.03.012] [Medline: 40355168]
30. Friedman LM, Furberg CD, DeMets DL, Reboussin DM, Granger CB. Fundamentals of Clinical Trials: Springer; 2015. [doi: 10.1007/978-3-319-18539-2]
31. Grimes DA, Schulz KF. Descriptive studies: what they can and cannot do. Lancet 2002 Jan 12;359(9301):145-149. [doi: 10.1016/S0140-6736(02)07373-7] [Medline: 11809274]

Abbreviations

IEC: International Electrotechnical Commission

XSL•FO
RenderX

Edited by A Kushniruk, J Sarvestan; submitted 20.Aug.2025; peer-reviewed by J Schlesinger, M Hellaby; revised version received 09.Feb.2026; accepted 11.Feb.2026; published 01.Apr.2026.

Please cite as:
Gasciauskaite G, Senn S, Hunn C, Nöthiger CB, Tscholl DW, Roche TR
Anesthesia Providers' Perspectives on the Redesigned Philips Acoustic Alarm System: Qualitative Pre- and Postimplementation Study
JMIR Hum Factors 2026;13:e82703
URL: https://humanfactors.jmir.org/2026/1/e82703
doi:10.2196/82703

© Greta Gasciauskaite, Stefanie Senn, Cynthia Hunn, Christoph B Nöthiger, David W Tscholl, Tadzio Raoul Roche. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 1.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

User Perceptions of a Large Language Model–Based Chatbot for Cognitive Reappraisal of Stress: Mixed Methods Feasibility Study

Ananya Bhattacharjee[1], PhD; Jina Suh[2], PhD; Mohit Chandra[3], MS; Javier Hernandez[2], PhD

[1]King Center on Global Development, Stanford University, 366 Galvez St, Stanford, CA, United States

[2]Microsoft Research, Microsoft (United States), Redmond, WA, United States

[3]School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, United States

Corresponding Author:
Ananya Bhattacharjee, PhD
King Center on Global Development, Stanford University, 366 Galvez St, Stanford, CA, United States

Abstract

Background: Cognitive reappraisal is a widely studied emotion regulation strategy that helps individuals reinterpret stressful situations in ways that reduce their emotional impact. Digital mental health (DMH) tools often struggle to support this process because scripted templates fail to adapt to the varied and incomplete ways users describe their stressors. Large language models (LLMs) offer opportunities to increase conversational flexibility while preserving structured intervention steps.

Objective: This study aimed to examine the feasibility of an LLM-based single-session intervention (SSI) for stress reappraisal. We assessed whether the activity was associated with short-term improvements in stress-related outcomes, and explored the design tensions that arose during user interactions.

Methods: We conducted a feasibility study with 100 employees from a large US technology company. Participants completed a structured cognitive reappraisal session delivered by a GPT-4o–based chatbot within Qualtrics. Pre-post measures included situational stress intensity, stress mindset, perceived demand, and perceived resources (all 5-point scales). Paired Wilcoxon signed-rank tests were used with Benjamini-Hochberg correction. To complement self-reports, we analyzed sentiment and stress trajectories across conversation tertiles using a Robustly Optimized BERT Pretraining Approach (RoBERTa; BERT: Bidirectional Encoder Representations from Transformers) sentiment classifier, a RoBERTa stress classifier, and an LLM-based stress rater. We also conducted a thematic analysis of participants' open-ended responses covering stress impact, perspective change, liked and disliked aspects, and improvement suggestions, alongside their conversation logs.

Results: Significant reductions were observed in situational stress intensity (mean 0.29, SD 0.83, r_{rb}=0.54; P=.002) and significant improvements in stress mindset (mean 1.70, SD 4.37, r_{rb}=0.44; P=.002). Sentiment and stress classifiers showed consistent declines in negative sentiment and stress from conversation start to end (all omnibus Friedman tests P<.001). Thematic analysis of open-ended responses surfaced several patterns in how participants engaged with the activity: they valued the structured prompts for organizing their thoughts, described shifts in how they interpreted their stressors, used the interaction to identify concrete next steps, and experienced the chatbot as a nonjudgmental space for reflection. Reported design tensions included perceived scriptedness, variable preferences for conversation length, and mixed reactions to AI-driven empathy.

Conclusions: An LLM-enhanced cognitive reappraisal activity showed promise to be delivered as a brief digital intervention and is associated with short-term improvements in situational stress and stress mindset. Participants appreciated the clarity and reflection supported by the structured sequence, while noting important design challenges in balancing structure with conversational naturalness and contextual depth. These findings highlight both the promise and the design constraints of integrating LLMs into DMH interventions.

(JMIR Hum Factors 2026;13:e89736) doi:10.2196/89736

KEYWORDS

cognitive reappraisal; reflection; LLM; generative AI; stress; large language model

Introduction

Cognitive reappraisal is an emotion regulation strategy that involves reframing the meaning of a situation in order to change its emotional impact [1-3]. In the context of stress, this entails altering one's interpretation of a stressful event to reduce its harmful emotional effects. It is considered a form of cognitive appraisal that enables individuals to reinterpret an event, often finding new meaning or viewing the outcome as more benign [1]. By changing the way one thinks about a stressor, cognitive reappraisal can shift the emotional response from anxiety or anger to a more neutral or even positive state. For example,

XSL•FO
RenderX

interpreting a professional setback as a growth opportunity rather than a failure can diminish distress and foster adaptive coping [4].

A substantial body of research demonstrates that cognitive reappraisal is a beneficial strategy for psychological and emotional well-being. Individuals who regularly use reappraisal tend to report lower levels of depression and anxiety [2,3,5]. Over time, greater reappraisal ability has also been linked to a reduced risk of developing depressive symptoms, particularly under conditions of high stress [1]. Reappraisal has been associated with more frequent positive affect and greater life satisfaction [2]. In addition, people who use reappraisal often report better interpersonal relationships [6,7]. These benefits have been observed across diverse populations and contexts, with studies showing that habitual use of reappraisal correlates with improved psychological health over time [3]. Together, these findings highlight the substantial and wide-ranging benefits of cognitive reappraisal for well-being.

However, existing strategies for cognitive reappraisal of stress, similar to many digital mental health (DMH) interventions, often struggle to provide support that feels applicable to users' situations [8]. An important challenge lies in their reliance on scripted responses and fixed templates that proceed through a predetermined sequence without adjusting to the way users naturally describe their experiences [8,9]. Individuals may give partial descriptions, mention several elements of the situation at once, or provide information that is either too broad or too narrow for the next required step in the intervention. Scripted tools [10-12] typically move forward regardless of these variations, which can create moments when the guidance feels rigid or misaligned with what the user is attempting to communicate. Prior chatbots such as Woebot [13] and Wysa [14] demonstrated that structured cognitive behavioral therapy (CBT)-style exchanges could produce measurable reductions in stress and negative affect, though their scripted templates limited contextual responsiveness. These limitations have highlighted the need for systems that maintain a consistent intervention structure while allowing conversational adjustments within each step. Approaches that combine structured reappraisal steps with conversational responsiveness may help reduce the sense of mismatch that users often report with scripted digital interventions [15,16].

Large language models (LLMs) offer an emerging opportunity to address the limitations of traditional stress reappraisal tools that rely on fixed scripts and rigid templates [17,18]. Instead of producing predetermined responses, LLMs can engage in brief conversational exchanges that acknowledge users' descriptions and request clarification when needed [19,20]. This capacity supports a smoother dialogue within a structured intervention, helping the system remain connected to what the user has already expressed. For example, if a user introduces a stressor in broad terms, an LLM-based agent can provide a simple acknowledgment and ask for a small detail required for the next step. If a user highlights a particular source of tension, the agent can recognize that element before guiding the user forward in the predefined sequence. These conversational elements can help maintain coherence between the user's wording and the transitions across steps, which may reduce the sense of rigidity that often occurs in scripted tools [17,21].

While LLMs offer exciting possibilities, they cannot simply be inserted into well-being interventions such as cognitive reappraisal without careful design. These interventions are inherently structured, and preserving that structure is essential for delivering appropriate support. Without it, there is a risk of generating content that may be unhelpful or even counterproductive. Rather than viewing LLMs as replacements for structured interventions, we posit that their promise lies in augmenting these techniques, enabling context-sensitive interaction through natural language. By acknowledging the user's situation and adapting responses accordingly, LLMs may help structured interventions feel tailored, while still upholding the core principles that guide them [17,19,22].

Natural language processing (NLP) research has used language models to support reframing of negative thoughts, including positive reframing as style transfer [23] and identification and reframing of cognitive distortions [21]. More recent work has directly examined LLM-based cognitive reappraisal: Sharma et al [22] showed that LLMs can support identification of thinking traps and provision of contextually appropriate reframes, and Zhan et al [16] showed that LLMs guided by psychologically grounded constitutions can generate targeted reappraisals rated by clinical psychologists as more aligned with established principles than human-written responses. However, these approaches have largely focused on generating individual reappraisal responses, with less attention to the multiturn conversational process through which reframing often unfolds in real-world supportive interactions.

Motivated by these opportunities and challenges, we examined the feasibility of an LLM-enhanced cognitive reappraisal intervention delivered as a single-session intervention (SSI). SSIs are brief, structured activities that can produce rapid shifts in thoughts or feelings and have shown benefits for stress, negative thinking, and related outcomes across varied populations [8,22,24,25]. They are well-suited for early-stage exploration because they allow researchers to assess whether a short, structured activity can support users as they work through a stressful scenario. In this study, we focus specifically on situational stress, the immediate psychological tension a person experiences when confronted with a specific stressor, such as a difficult interaction with a colleague, an unresolved conflict, or an overwhelming workload. This is distinct from general perceived stress, which reflects a person's cumulative sense of feeling overwhelmed or unable to cope across their daily life over a period of weeks or months [26].

Our work was guided by the following research questions (RQs):

1. RQ1: How does engaging with an LLM-based chatbot for cognitive reappraisal of a stressful scenario impact individuals' situational stress levels?
2. RQ2: What design tensions arise when individuals interact with an LLM-based chatbot to cognitively reappraise a stressful scenario?

To examine these questions, we designed an SSI that used an LLM to guide users through a structured cognitive reappraisal

XSL•FO
RenderX

activity [24,25]. The LLM was instructed to follow a predefined sequence similar to prior work [8]. Within each step, it incorporated conversational adaptivity by acknowledging user input and requesting brief clarifications when needed, while preserving the structure of the intervention. We chose a feasibility study design to examine whether this approach, in which the LLM facilitates a structured reappraisal process instead of generating reappraisals on the user's behalf, could be delivered reliably and produce short-term shifts in situational stress. Our aim was to gather early evidence of perceived benefits and preliminary impact, and to identify design tensions that emerge when structured psychological interventions are delivered through conversational interfaces. Accordingly, our focus was on understanding user experiences and implementation considerations rather than conducting a controlled evaluation. Feasibility work is important when introducing new technological mechanisms because it can reveal practical factors that influence real-world deployment and highlight design considerations [27,28].

In a feasibility study involving 100 employees from a large US technology company, it showed that participants found the activity helpful for working through stressful situations. Qualitative feedback indicated that the structured sequence supported clearer thinking by helping participants slow down, articulate their thoughts, and reinterpret their stressors in a more constructive way. Many also described feeling validated, noting that the interaction helped them organize their emotions and provided a nonjudgmental space for reflection. These qualitative insights were complemented by promising quantitative results, with significant reductions in immediate stress (r_{rb}=0.54; P=.002) and improvement in stress mindset (r_{rb}=0.44; P=.002), further corroborated by sentiment analysis of conversation data. However, the integration of LLMs also introduced design tensions: contextualizing the experience required additional user input, and following a predetermined structure could reduce the perceived conversational quality of the interaction.

Our contributions include (1) a structured LLM-enhanced cognitive reappraisal intervention for managing stress, (2) empirical insights into how the intervention impacts users' experiences of stress, and (3) design tensions to inform future LLM-based DMH interventions. Our study demonstrates the potential of LLM-enhanced reappraisal as a brief, scalable intervention, while also drawing attention to design challenges, including how to balance conversation length, contextual richness, and expression of empathy.

Methods

Design of the Intervention

Theoretical Grounding and Sequence Design

Our intervention was designed as an SSI [24,25]. SSIs are structured, stand-alone activities intended to produce immediate effects on targeted psychological outcomes. Typically lasting 10 - 20 minutes, they offer a time-efficient way to promote reflection, reappraisal, or symptom relief, making them suitable for a wide range of populations. In addition to offering rapid support, SSIs can serve as a gateway to more comprehensive mental health care by helping individuals build insight or skills that increase their willingness to seek further support [8]. Prior research has demonstrated their effectiveness in addressing negative thought patterns, reducing stress, and alleviating symptoms of anxiety and depression in both clinical and nonclinical settings [8,22,24,25].

Unlike multisession interventions, which typically unfold over multiple weeks through repeated check-ins, skill-building exercises, and progressive assignments [12,29], SSIs are completed in a single sitting without requiring prior sessions or follow-up engagement. While multisession programs rely on cumulative exposure and gradual skill development over time, SSIs are designed to deliver focused support within a bounded interaction. Prior work illustrates how SSIs can compress core therapeutic elements into a single experience. For example, Schleider and Weisz [30] tested a brief computer-based SSI in which participants engaged with psychoeducational content and a short reflective writing task within one sitting, and observed reductions in depression and anxiety symptoms at follow-up. More recently, Sharma et al [20] demonstrated that cognitive reframing can be supported through structured, single-session interactions with language models. In contrast to multisession formats that distribute these elements across time, SSIs integrate learning, reflection, and application into a single, self-contained interaction.

We drew particular inspiration from a prior SSI reported by Bhattacharjee et al [8]. The activity was designed to be easily referenced or practiced by individuals when encountering negative thoughts and emotions in their everyday lives. Their work adapted core concepts from CBT, specifically thought records and behavioral chaining [31,32], into a guided digital reflection activity. In CBT, thought records involve documenting a situation; the thoughts, emotions, and behaviors it evokes; and then generating alternative thoughts [31]—behavioral chaining focuses on tracing the sequence from trigger to reaction to identify patterns [32]. Both techniques help individuals become more aware of their internal responses and prepare for similar situations in the future by promoting insight and reflection. Given that cognitive reappraisal is not inherently context-specific, our activity was similarly designed to be broadly applicable across a range of stressful situations, with the workplace serving as the deployment context for this study.

Our activity follows a similar structure, consisting of 11 questions. The activity was developed through an iterative process involving researchers with expertise in human-computer interaction (HCI), psychology, cognitive science, generative AI, and DMH. Members of the research team had prior experience designing DMH interventions both with and without the use of AI and have published extensively in leading HCI, psychology, and AI venues. The first set of prompts (Questions 1 - 7) operationalizes the cognitive model underlying CBT thought records, in which a situation gives rise to automatic thoughts that in turn drive emotional and behavioral responses [31,32]. Moving from the situation (Q1) to its most troubling element (Q2), to automatic thoughts (Q3-Q4), to feelings (Q5), to behaviors (Q6), and finally to an integrated trigger → thought → feeling → behavior summary (Q7) mirrors the standard structure used to build awareness of the links among cognition,

XSL•FO
RenderX

affect, and action before any restructuring is attempted [15]. Questions 8 - 11 then enact cognitive reappraisal as defined in the stress-appraisal and emotion-regulation literatures [3,5]. Q8 corresponds to examining the evidence for and proportionality of an appraisal, a core component of cognitive restructuring. Q9 draws directly on the challenge-versus-threat distinction in transactional models of stress appraisal [4], inviting users to reframe a stressor as a challenge to be met rather than a threat to be avoided. Q10 reflects reappraisal in Gross's process model of emotion regulation, in which the meaning of a situation is reinterpreted to alter its emotional impact [5]. Q11 supports generalization and consolidation, encouraging users to carry the reappraisal into future situations [3]. Reappraisal comes last in the sequence by design: examining and reframing an appraisal is difficult until the cognition, emotion, and behavior tied to the trigger have been laid out, which is the work of Questions 1-Q7.

Table 1 shows the full sequence of prompts and their intended purpose. After completing all steps, the chatbot was instructed to provide a brief summary of the conversation and clearly state that the structured part has concluded.

Table . Reflection questions used in the intervention and their intended purpose.

Question number	Questions	Purpose
1	What is the situation? Feel free to explain it in as much detail as you would like.	Provides context for the activity
2	What part of the situation is the most troubling?	Sets an agenda for the rest of the activity
3	What are you thinking to yourself?	Identifies troubling thoughts
4	What thought is the most troubling?	Focuses attention on the most troubling thought
5	What do you feel when you think this?	Reinforces the core CBT[a] principle that thoughts trigger feelings
6	When you have these feelings, what actions do you take? What do you avoid?	Identifies behaviors that are caused by the cascading effect of thoughts and feelings
7	Retype the summary of the situation in the following format: trigger→thought→feeling→behavior	Synthesizes past reflection by highlighting the connection between the trigger and its manifestations
8	Do you believe that the initial trigger justifies the intensity of your thoughts and feelings?	Challenges potentially negative thought patterns
9	How can interpreting the trigger as a challenge rather than a threat alter your response?	Encourages cognitive reinterpretation of the situation
10	What new perspectives could you adopt to view these challenges as opportunities?	Promotes a reappraisal frame focused on growth
11	How might this change in perspective influence your future reactions to similar challenges?	Encourages transfer of reappraisal strategies to future scenarios

[a]CBT: cognitive behavioral therapy.

LLM-Driven Implementation

Motivated by recent advances in LLM-based interventions that support context-aware support [17,22,33], we adapted the sequence in Table 1 to allow content to align more closely with users' specific circumstances. To implement this, we used a prompt engineering approach. The prompt was developed through an iterative process by the same research team. The complete system prompt is provided in Section A in Multimedia Appendix 1. We structured the prompt using GPT-4o with the following components:

1. Structured reflection process: The prompt was designed to guide users through a step-by-step conversation aligned with Table 1.
2. Conversational interaction through natural dialogue: Instead of presenting fixed or survey-like questions, the chatbot was instructed to respond naturally and acknowledge the user's situation. If a response was vague or incomplete, it was instructed to ask clarification questions to gather more detail. The chatbot was also instructed to avoid repeating the same phrases or prompts. The chatbot was designed as a structured facilitator; it did not produce advice, suggestions, or reframings on the user's behalf. The reappraisal questions (Q8-Q11) were designed to prompt users to arrive at these insights through their own reflection, with the chatbot's role being to acknowledge, clarify, and guide rather than to supply content.
3. Reflective instructional prompt: Prior work has shown that in longer conversations, LLMs may deviate from the intended behavior, even when following the prompt, because they try to adjust to the user's specific responses, clarifications, or context [34]. In our case, this could sometimes lead the model to drift away from the specific question or theme it was meant to address. To handle this, we added an instruction for the model to first generate a response, then check if it aligns with the current question's theme, and revise it if needed [35].

The chatbot was implemented as a text-based conversational agent and was embedded directly within the Qualtrics survey platform. It did not have a named persona or profile picture.

XSL•FO
RenderX

The chatbot opened each session with a fixed introductory message welcoming the participant and describing the nature of the interaction (Section A in Multimedia Appendix 1). The language style was conversational and reflective, with a temperature setting of 0.7 to allow for naturalistic variation in responses while maintaining coherence. A screenshot of the interface is provided in Section C in Multimedia Appendix 1.

Participants

We recruited information workers from a large US-based technology company with a broad portfolio spanning software, hardware, cloud services, and cybersecurity. Employees based in the United States were invited via email to participate in a voluntary survey. We recruited from this setting because workplace stress is well-documented as particularly prevalent in the technology sector [36,37] and because our affiliation with a technology company provided internal access to a large, geographically distributed employee population.

Participants were informed that the study focused on AI conversational agents for supporting workplace stress scenarios. A total of 100 participants completed the study. They represented multiple gender identities (69 men, 27 women, and 4 undisclosed; more options were offered) and varied in age (22 participants aged 18 - 35 years, 61 aged 36 - 55 years, and 17 aged 56 years and older). Prior to engaging with the intervention, participants completed the 10-item Perceived Stress Scale (PSS) [26] to assess their stress levels over the past month. The average score was 21.26 (SD 3.67), which falls within the moderate stress range (typically defined as 14 - 26). The PSS served as a baseline measure to characterize participants' general stress levels at the time of the study; no minimum score was required for participation. Scores on the PSS range from 0 to 40, with higher scores indicating greater perceived stress (0 - 13=low, 14 - 26=moderate, and 27 - 40=high).

Participants also reported their prior experience with text-based AI conversational agents: 34% (34/100) had 6 months to 1 year of experience, 33% (33/100) had 1 to 2 years, 22% (22/100) had 2 to 5 years, 6% (6/100) had more than 5 years, and 5% (5/100) had less than 6 months, suggesting that most participants had prior familiarity with conversational agents. A detailed breakdown of participant demographics, along with distributions of PSS scores, is available in Section B in Multimedia Appendix 1.

Procedure

Ethical Considerations

This study was approved by the Microsoft Research Institutional Review Board (IRB) through the Microsoft Research Ethics Review Program under ERP ID 10893. Participants were invited to engage in an SSI in which they reflected on a workplace stress situation through a conversation with a chatbot.

Procedure Overview

The study was disseminated through a one-time email invitation, and participants could complete the activity at any time of their choosing after receiving it. The intervention was designed as a single, one-time activity to be completed within a continuous sitting; participants were not able to return to complete additional sessions. The chatbot guided them through a series of questions designed to support reflection on their experience. Participants were encouraged to respond genuinely to each prompt. At the end of the interaction, the chatbot indicated that the conversation was complete, after which participants proceeded to the next page. They were informed in advance that the conversation would likely involve around 15 conversational turns, based on the 11 core questions programmed into the chatbot, along with potential follow-up and clarification prompts. Section C in Multimedia Appendix 1 demonstrates an example conversation between the chatbot and a user. All components of the study, including the chatbot interaction, were completed within the Qualtrics platform.

To assess the impact of the SSI, participants completed a set of self-report questions both before and after the session. To evaluate short-term changes following the intervention, we selected measures that capture both immediate subjective stress responses and theoretically relevant appraisal-related constructs. Specifically, we assessed situational stress intensity, stress mindset, perceived demand, and perceived resources, as these reflect how individuals interpret and respond to stressful situations. All responses were recorded on a 5-point Likert scale, where 1 indicated low agreement or intensity (eg, "strongly disagree" and "very low") and 5 indicated high agreement or intensity (eg, "strongly agree" and "very high"). The measures included:

Situational Stress Intensity

A single question assessing participants' stress level in relation to a workplace situation [8]. We selected this brief measure to capture immediate subjective stress relevant to the focal scenario and to minimize burden within a single-session study.

Stress Mindset

Eight statements capturing beliefs about the potential enhancing or debilitating effects of stress (eg, "Experiencing this stress facilitates my learning and growth;" "The effects of this stress are negative and should be avoided") [38]. These included both positively and negatively framed statements. We included this measure because cognitive reappraisal may shift how individuals interpret the meaning and consequences of stress.

Perceived Demand

A rating of how demanding the stressor felt [36,39]. This measure was included as an exploratory outcome grounded in the job demands-resources model [39].

Perceived Resources

A rating of the participant's perceived ability to cope with the stressor [36,39]. Alongside perceived demand, this measure captured whether the intervention shifted participants' sense of their own coping capacity [39].

Situational stress intensity served as the primary outcome measure, as it most directly captures the intended effect of the reappraisal activity. The remaining measures—stress mindset, perceived demand, and perceived resources—were exploratory and included to provide a broader characterization of participants' stress-related experiences before and after the intervention.

XSL•FO
RenderX

To gather user feedback about the intervention, participants also responded to open-ended questions about how the activity affected their stress levels and their view of the situation, as well as what aspects they liked or disliked. All open-ended questions were provided immediately at the end of the session, within the same Qualtrics survey. These questions included the following:

1. How did this activity affect your stress levels? Feel free to explain in as much depth as you would like.
2. Please explain whether and how interacting with the chatbot has changed your view of the stressful situation.
3. Please comment on what aspects of the chatbot interaction you liked or found helpful. Why?
4. Please comment on what aspects of the chatbot interaction you did not like or did not find helpful. Why?
5. Please suggest how we can further improve this activity.

All open-ended questions were mandatory except for the improvement suggestion question, which was optional. All 100 participants provided responses to the 4 mandatory questions, and 66 provided responses to the optional improvement suggestion question.

Because the study used a one-time, survey-based format, participants who did not complete the study did not submit responses and therefore were not included in the final sample. We were unable to assess dropout rates among participants who may have started but not completed the study.

Data Analysis

Overview of Data Analysis

As a feasibility study, our objective was to examine whether engagement with an LLM-based chatbot for cognitive reappraisal could lead to observable improvements in stress-related outcomes, such as situational stress intensity and stress mindset. Early-stage feasibility studies serve a distinct role in intervention research [27]; they help determine whether the conceptual approach is worth pursuing, whether users engage meaningfully with the system, and whether measurable changes occur in theoretically relevant constructs. Demonstrating even modest improvements under these conditions can support the potential viability of LLM-based reappraisal as a foundation for more controlled investigations. To contextualize these quantitative findings, we complement them with qualitative analyses that explore how users experienced and interpreted the chatbot's support, providing insight into the underlying mechanisms and perceived value of the interaction.

We conducted paired Wilcoxon signed-rank tests to assess pre-post differences across the 4 measures described in the "Procedure" section. We applied the Benjamini-Hochberg (BH) procedure to account for multiple comparisons.

To complement user-reported scores, we analyzed linguistic features of user-chatbot conversations. Each conversation was divided into 3 segments (T1=beginning, T2=middle, and T3=end), and we examined how predicted stress and sentiment varied across them using 3 approaches.

Robustly Optimized BERT Pretraining Approach Sentiment Classifier

A pretrained Robustly Optimized BERT Pretraining Approach (RoBERTa)–based model [40] (BERT: Bidirectional Encoder Representations from Transformers) that outputs a probability distribution over positive and negative sentiment for each conversation segment.

RoBERTa Stress Classifier

A pretrained RoBERTa-based model [41] estimates stress probability ranging from 0 to 1 for each conversation segment.

LLM Stress Rater

A prompting-based approach that assigns a stress score to each user message using a structured prompt informed by prior work on LLMs as evaluators [42-46]. The rater outputs a score from 1 to 5, where 1=no observable stress and 5=high stress (characterized by pronounced emotional strain or difficulty coping). The rating criteria were developed through a review of prior literature and iterative design, enabling finer distinctions in stress intensity. Further details about the prompt design are provided in Section D in Multimedia Appendix 1.

While these methods provide useful approximations of users' affective and psychological states, we recognize that their outputs can be shaped by modeling assumptions and may miss contextual nuance. We therefore interpret them as supportive evidence for our quantitative analyses rather than as definitive indicators of users' internal experiences.

For the qualitative analysis, we first anonymized and cleaned participants' responses, then conducted a thematic analysis [47]. The first author independently reviewed all participant responses and developed an initial codebook using an open-coding method. This process reflected an iterative approach that integrated both deductive and inductive techniques: the analysis of RQ1 was guided by prior work on stress reappraisal interventions, while RQ2 was approached inductively, allowing themes to emerge directly from the data. The first author then used axial coding to refine categories and develop higher-order themes. Throughout the analysis, the codes and themes were regularly discussed with other members of the research team to ensure multiple perspectives were considered. The development of themes was reviewed through detailed collaborative discussions. This approach is consistent with qualitative coding practices used in prior studies [48,49].

Ethical Considerations

This research was reviewed and approved by the IRB at the technology company where the study was conducted. Given the sensitive nature of the intervention, which included reflection on negative thoughts and potentially distressing life situations, we were attentive to ethical considerations throughout the research process and implemented safeguards to support participant well-being. All participants provided informed consent before beginning the study. To protect confidentiality, they were explicitly requested not to include any personally identifiable information in their responses. Participants were also informed that they could skip any question or withdraw from the study at any time without consequence. Additionally,

XSL•FO
RenderX

participants were provided with information about internal support resources in case they experienced stress, anxiety, depression, or emotional discomfort during or after the session.

Results

Interaction Characteristics

Participants wrote an average of 12.81 (SD 1.66) messages per conversation, contributed 283.74 (SD 243.16) words, and spent 23.09 (SD 23.99) minutes engaging with the chatbot. These results suggest that participants were responsive to the prompts and invested a reasonable level of time and effort in composing substantive replies. Building on these observations, we now examine how our study addressed the RQs.

How Does Engaging With an LLM-Based Chatbot for Cognitive Reappraisal of a Stressful Scenario Impact Individuals' Situational Stress Levels?

Overview

Participants reported significant reductions in situational stress intensity following the intervention (mean 0.29, SD 0.83, r_{rb}=0.54; *P*=.002) and significant improvements in stress mindset (mean 1.70, SD 4.37, r_{rb}=0.44; *P*=.002). The latter reflects a shift toward viewing stress as more enhancing than debilitating. Perceived resources showed improvement (mean 0.17, SD 0.83, r_{rb}=0.32; *P*=.07), and perceived demand showed a reduction (mean 0.12, SD 0.83, r_{rb}=0.22; *P*=.17), although these changes were not statistically significant. Table 2 shows these results.

Table . Pre and postintervention means (SDs) for all outcome measures. The improvement column reports change in the beneficial direction, computed as Post-pre for measures where higher is better (↑) and Pre-post for measures where lower is better (↓), so that positive values always indicate improvement. Rank-biserial correlations (r_{rb}) are likewise signed so that positive values favor the intervention. *P* values are corrected using the Benjamini-Hochberg (BH) procedure (α=.05).

Measures	Pre, mean (SD)	Post, mean (SD)	Intended direction	Improvement, mean (SD)	W	*P* value	r_{rb}
Situational stress intensity	3.38 (0.95)	3.09 (0.91)	↓	0.29 (0.83)	216	.002[a]	0.54
Stress mindset	20.71 (6.64)	22.41 (7.38)	↑	1.70 (4.37)	994	.002[a]	0.44
Perceived demand	3.40 (0.91)	3.28 (0.89)	↓	0.12 (0.83)	385	.17	0.22
Perceived resources	2.97 (0.97)	3.14 (1.00)	↑	0.17 (0.83)	306.5	.07	0.32

[a]*P*<.01.

We examined the progression of average negative sentiment and stress scores across conversation tertiles (T1-T3) for all 3 classifiers, followed by nonparametric statistical analyses. Results are presented in Figure 1 and Table 3. Across all 3 classifiers, users' messages exhibited a consistent pattern: sentiment became more positive, and indicators of stress decreased from the beginning to the end of the interaction. Friedman tests revealed statistically significant differences across tertiles for each classifier. Pairwise Wilcoxon tests showed significant changes between T1 and T3, as well as between T2 and T3, in all 3 cases. These findings suggest that as users engaged with the chatbot, their expressions of negative sentiment and stress declined over time, aligning with the observed reduction in self-reported stress scores.

XSL•FO
RenderX

Figure 1. Changes in sentiment and stress scores across conversation tertiles (T1-T3). Error bars represent SDs. LLM: large language model; RoBERTa: Robustly Optimized BERT Pretraining Approach.

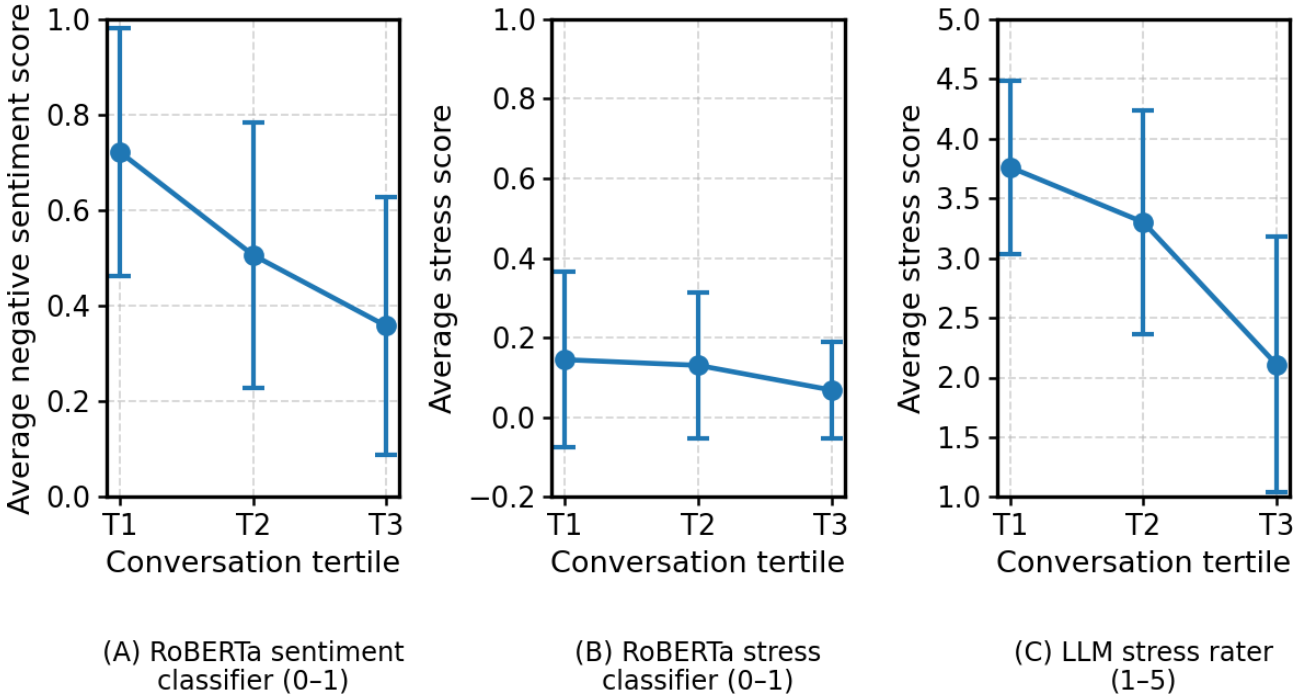


Table . Statistical tests accompanying Figure 1, assessing changes in negative sentiment and stress scores across conversation tertiles (T1-T3) for all 3 classifiers. For each classifier, we report an omnibus Friedman test across T1-T3 and 3 pairwise Wilcoxon signed-rank tests. All *P* values are corrected using the BH procedure.

Statistics	RoBERTa[a] sentiment classifier	RoBERTa stress classifier	LLM[b] stress rater
Omnibus test: Friedman across T1-T3	<.001[c]	<.001[c]	<.001[c]
Pairwise tests (Wilcoxon signed-rank)			
T1 vs T2	<.001[c]	.69	<.001[c]
T1 vs T3	<.001[c]	<.001[c]	<.001[c]
T2 vs T3	<.001[c]	<.001[c]	<.001[c]

[a]RoBERTa: Robustly Optimized BERT Pretraining Approach (BERT: Bidirectional Encoder Representations from Transformers).

[b]LLM: large language model.

[c]*P*<.001.

General Appreciation for Structured Reflection

Participants in general highlighted the value of the chatbot's step-by-step approach, noting that the ordered prompts helped them organize thoughts that might otherwise remain scattered. P55 described the session as "a good exercise in laying out the issue, my responses, and getting to why it's that way," and P42 emphasized how the sequence of questions encouraged structured analysis rather than a quick vent. Several also echoed that the structure offered a clear path through the conversation, making it easier to track what had been covered and where to focus next.

Participants often remarked that the value of the interaction lay in how it helped them arrive at their own insights through deliberate unpacking of the stressor. They acknowledged that the structured opportunity to express themselves brought a sense of relief. The act of "thinking aloud" through writing or guided dialogue was also perceived as familiar and useful. Drawing a comparison to journaling, P15 reflected:

> *This activity did decrease my stress level by allowing for a structured conversation regarding a specific stressor. It helped me reflect on my own thoughts without spiraling into anxiety. Just by explaining and going through the situation it made me feel more calm, similar to journaling but more structured. I have previously liked journal prompts in the past and this feels similar.*

Some participants also appreciated that the structured format helped shift their focus from problems to potential solutions. They noted that the prompts encouraged them to slow down and articulate their thoughts more deliberately. As P89 explained:

XSL•FO
RenderX

> *I often struggle for words, but having a chatbot to ask these questions makes me feel like I can take my time to find the right way to say what I want to say.*

These comments suggest that participants valued not just the content of the activity, but the structured way in which it guided their thinking.

Reinterpreting Stressful Situations

The shift observed in stress mindset scores, toward viewing stress as more enhancing than debilitating, was echoed in participants' qualitative comments. Several described how the conversation helped them reassess their situation from a new angle. P79 shared that it "shifted my perspective just enough to see it more as a challenge than a frustration," while P21 noted that the chatbot offered an alternative way of thinking when they felt mentally stuck, helping them move beyond unproductive thought patterns.

The conversation logs further reflect this reappraisal process. P4, who began the session feeling overwhelmed and unsupported by colleagues, wrote at Q9 that "perhaps the anger and anxiety are the result of repeated cycles and maybe I need to consider I am burnt out on the role"—a reinterpretation that emerged through their own responses to the structured questions. P65, working through anxiety about executing quickly on a difficult technical project, arrived by Q10 at a reframe of the work as "an exploration, an open mindset that allows me to explore boundaries," shifting from perceiving the situation as a threat to viewing it as an opportunity for learning.

However, not every participant experienced dramatic shifts in perspective. P30, for example, remarked that "while it did not change my viewpoint 100%, it was nice to write it out," suggesting that even partial reframing can be helpful. Others, including P15 and P48, described subtle changes in how they interpreted or emotionally responded to the stressor. These comments highlight the potential of the interaction to support gradual cognitive shifts, helping participants reinterpret their stressors and relate to stress in a more constructive way. It is worth noting that these reframings were generated by participants themselves through the structured reappraisal questions; the chatbot's role was to guide and facilitate this process rather than to supply alternative interpretations directly.

Identifying Concrete Next Steps

Beyond perspective shifts, some participants described how the dialogue helped translate reflection into concrete action. P66 appreciated "the structured approach to break down the root of my stress and ideate on solutions," while P83 noted that the activity helped them "channel stress into action." P70 shared that although their stress level did not change, the activity "helped to structure my approach to the situation," and P76 commented that the conversation "helped lead me to some useful options to help me fix my problem." These comments suggest that, for some participants, the value of the activity lay not only in processing emotions and perspectives but in arriving at a clearer sense of what to do next.

The conversation logs further illustrate how this process unfolded during the later stages of the activity. Participants who initially described interpersonal tension or coordination difficulties sometimes progressed toward specific action plans by the final reappraisal prompts. For example, P73, who began the session feeling overwhelmed by disorganized planning and strained colleague relationships, later described intentions to foster better communication and establish more regular check-ins with their team before the next planning cycle. Similarly, P61, who entered the conversation frustrated by a difficult team dynamic, moved toward articulating concrete steps such as involving a new team member in daily discussions and investing in a stronger working relationship. P59, who entered the session anxious about presenting to a large group of external customers in a nonnative language and worried about losing the audience's attention, arrived by the final prompts at a concrete plan to use auto-captioning technology in future presentations and to build in pauses for audience questions. These examples suggest that the structured questioning sequence could help some users move from diffuse stress toward more actionable coping plans.

Validation of Personal Experiences

Some participants described the chatbot as helpful in recognizing or affirming their existing efforts and emotional responses. Even when the chatbot did not introduce new ideas, it was still seen as useful for reinforcing what participants were already feeling or doing. As P16 explained:

> *Just because it didn't change my view or give me any new data to think about doesn't mean it wasn't useful… it's validation that there isn't something I'm missing, that I've explored every opportunity and tried everything I can.*

Others similarly appreciated moments where the chatbot reflected back their current state of mind. P74 noted:

> *It was great at validating my feelings and guiding me in its approach, even when I was reluctant to think about a solution.*

P82 shared that "the chatbot acknowledged my efforts as productive" in dealing with a difficult situation, and P92 mentioned it "affirmed steps planned."

For a few, this sense of reinforcement was seen as limited. As P46 commented, "Lots of validation. It might have been better if it had challenged me." Still, for many participants, the interaction was valued for offering a sense of confirmation rather than redirection.

Appreciation for Nonjudgmental Interaction

Several participants appreciated the chatbot's nonjudgmental tone, describing it as a space where they could express themselves freely. For some, this quality made the interaction feel less pressuring than conversations with people. P42 remarked that it "allowed me to vent, without fear of judgment," while P65 shared, "I liked that it was objective."

Some others described the experience as supportive precisely because it did not involve being evaluated. P66 emphasized the benefit of having "an unbiased conversation" that helped them "focus on a task at hand." This perception of the chatbot as nonjudgmental contributed to some participants' willingness to open up or think through the situation more deliberately.

XSL•FO
RenderX

Perceived Misalignment With Context or Needs

Some participants described instances when the chatbot's responses felt disconnected from the realities of their situation. In particular, a few noted that the interaction seemed to assume ideal conditions, such as collaboration or personal control, that did not reflect their experiences. As P36 explained:

> *The chatbot assumes that all the people involved in the situation are reasonable, collaborative, and want to reach a solution that's acceptable to everyone. But the nature of my situation is that the other team has already established that they are none of these things.*

P94 similarly expressed that the prompts did not adequately reflect the complexity of their stressor, stating, "I felt that I needed to double down and explain more about the stressful situation, to say it's not that easy."

Some participants also described discomfort with having to revisit a stressful situation in detail, particularly when they had already discussed it elsewhere. P59 remarked, "I already had to explain the situation to myself, then my team, and now this bot wants to know about it too, causing me to relive the situation in my mind." Others felt that, despite going through the interaction, there was no tangible outcome. As P24 noted, "I dislike the fact that it goes nowhere and there is no action to be done." These reflections suggest that for some participants, retelling the situation without any follow-up or resolution made the experience less satisfying.

What Design Tensions Arise When Individuals Interact With an LLM-Based Chatbot to Cognitively Reappraise a Stressful Scenario?

Scripted Structure Versus Conversational Naturalness

As discussed in the previous sections, participants noted that the structured nature of the chatbot's prompts helped them consider alternative perspectives and identify potential solutions. However, several participants also reflected on how this structure may have come at the expense of a more natural, conversational flow. P6 remarked that "it wanted to ask certain questions as opposed to engage in conversation," highlighting a perception that the interaction was being guided more by a fixed sequence than by mutual exchange.

Some participants further sensed that the chatbot was following a script, which contributed to a feeling of artificiality. As P75 put it, the interaction felt "not human," due to what appeared to be a scripted, predetermined flow. While certain elements of the structure, such as follow-up questions or requests for clarification, were appreciated by some, they were not universally welcomed. For example, P94 noted that these follow-ups "helped pull out some important information," but others, like P96, found them repetitive and frustrating:

> *I didn't necessarily like that it always yielded additional questions, some of which I felt I had already answered earlier in the conversation.*

Together, these responses underscore a tension between the benefits of a structured, goal-directed dialogue and the desire for a more natural and responsive conversational experience.

Balancing Contextual Understanding and Conversation Length

Some participants expressed a desire for more questions to allow them to elaborate on their situation—particularly when their stressors were complex or nuanced. P7 noted the difficulty of condensing their experience, remarking, "It's the writing it down, trying to cover the important topics with brevity. Tough nut to crack." Similarly, P57 wished for a longer "getting to know each other" period to help the chatbot better personalize its responses.

Conversely, some participants felt that even the current level of interaction was somewhat extended. P12 commented that the exchange lasted "much longer than I cared for." They expressed that the number of questions, often accompanied by clarifications and follow-ups, made the experience feel more effortful than necessary. This concern was particularly salient among participants dealing with less intense stressors, where a longer interaction felt disproportionate. As P23 explained, "My stress trigger was basic and didn't need to have that many turns." These perspectives reveal a core design tradeoff: while deeper context can enable more tailored and relevant responses, excessive questioning may lead to user fatigue or disengagement in some cases.

Perceptions Toward AI and Artificial Empathy

Although the study did not explicitly ask about participants' attitudes toward AI, some participants shared comments that reflected preexisting skepticism about the use of AI and technology for well-being support. Several participants indicated that their discomfort stemmed not from the content of the conversation, but from their underlying beliefs about what makes an interaction feel helpful. P29 shared that "the fact that it was a chatbot made it annoying," and P37 expressed discomfort with "the fact that it was not a real person." These comments suggest a reluctance to view nonhuman agents as appropriate partners for discussing personal or emotional matters. P51 further expressed mistrust toward the system, stating that they do not trust a software entity that cannot feel emotions.

These views extended to how participants interpreted the chatbot's attempts at empathy. Some reacted negatively to what they perceived as artificial emotional cues. P14 stated, "The whole mimicking empathy is annoying," and P41 described the interaction as "trying too hard, like mimicking empathy but not quite getting there." In contrast, others responded more positively to the chatbot's empathic tone. P44 noted that the message felt "kind," even though they knew it was automated. P58 similarly shared, "I could tell it was a bot, but it still felt like it was trying to be understanding." These contrasting reactions highlight how users' preconceptions about AI—and particularly their expectations around empathy from AI—impacted their perceptions of AI support.

Discussion

Overview

In this work, we present a cognitive reappraisal intervention design that leverages LLMs to provide conversational interaction. Our SSI incorporated a structured conversational

XSL•FO
RenderX

scaffold to guide the reappraisal process, enabling individuals to revisit initial interpretations, examine different aspects of their challenges, and consider alternative explanations. This combination of a structured SSI format with LLM-driven conversational flexibility addresses a well-documented limitation of scripted tools, which typically cannot adapt to the varied ways users describe their experiences. Our deployment showed that this approach can retain many benefits of past cognitive reappraisal interventions, including reducing negative emotions, increasing clarity about one's situation, and encouraging reflection on potential growth opportunities [2,8,22]. However, integrating LLMs introduced several design tensions, particularly in balancing contextualization with the number of questions asked, maintaining a structured flow while preserving conversational quality, and managing perceptions of AI tone and empathy.

In the remainder of this section, we reflect on the key findings of our study, discuss how they align with or extend existing literature, and propose relevant directions for future research. We conclude with a discussion of the limitations of the study.

Principal Results

Our findings suggest that the LLM-based cognitive reappraisal activity can retain many of the benefits associated with more traditional reappraisal techniques. Participants reported reductions in situational stress and improvements in stress mindset after engaging with the intervention. These shifts were supported by the structured nature of the activity, which guided users in breaking down the elements of their stressful experiences. This process enabled them to reassess their initial interpretations and consider alternative perspectives, an outcome consistent with prior research on the role of cognitive reappraisal in fostering more adaptive emotional responses [1,2,6,50].

In addition to promoting perspective shifts, the activity helped some participants reflect on possible actions they could take to address their situations. Several described recognizing specific next steps or reframing their experiences as opportunities for growth [2,22]. These qualitative observations were further corroborated by 3 independent classifier-based measures, which showed consistent reductions in expressed stress and negative sentiment across conversation tertiles, providing convergent evidence beyond self-report. This highlights how structured LLM-enhanced interventions can support cognitive reframing while helping users identify concrete ways to respond to their stressors.

Even when substantial cognitive shifts did not occur, participants described experiencing subtle emotional changes through their engagement with the structured, LLM-enhanced activity. In particular, the process of externalizing a stressful situation (ie, expressing internal thoughts and emotions outwardly [51]) was seen as helpful in itself. Several participants likened the experience to journaling, which can offer a sense of release and promote clarity [52-54]. These findings suggest that the value of reappraisal-based interventions may extend beyond cognitive reframing, by also offering structured opportunities for reflection and emotional processing through expressive dialogue.

At the same time, participants' responses surfaced a tension between the benefits of a structured sequence and the desire for a more naturalistic conversational flow. While the structure was widely valued for organizing reflection, some described the interaction as feeling scripted, particularly when follow-up questions appeared to repeat information already provided. Designing LLM interactions that move smoothly between structured progression and responsive acknowledgment, without drifting from the intended sequence, should be an active research area for future work.

However, we also observed that cognitive reappraisal may not be appropriate in all scenarios. Participants pointed out that, although support was directed at a single individual, the source of stress could be more complex and involve multiple people within their network. This challenge is not unique to cognitive reappraisal but reflects a broader limitation in much of the literature on health interventions, which often emphasizes individual-level support [55,56]. In response to this gap, a growing line of research is exploring group-level interventions that engage multiple individuals in jointly addressing shared or relational challenges [57-59]. Building on this momentum, future work could explore the design of technology-mediated cognitive reappraisal interventions that support group-level reflection [60]. For example, digital platforms could facilitate shared appraisal by prompting teams to surface collective stressors, reflect on interpersonal dynamics, and co-construct alternative interpretations. By embedding reappraisal within team interactions, these technologies may have the potential to foster shared understanding, strengthen social support, and address the relational nature of stress more effectively than individually oriented interventions.

Considerations for Broader Applications

Our designed activity can function as an SSI, offering a brief, structured opportunity for individuals to engage in cognitive reappraisal in response to everyday stressors. SSIs have been shown to produce meaningful psychological benefits with minimal time commitment [24,25], so they could be well-suited for integration into demanding professional environments. We emphasize that LLM-based cognitive reappraisal interventions need not follow the exact structure used in our study. Our results point to the trade-off between contextual depth and conversation length. Some participants wished for more opportunities to elaborate on complex stressors, while others found the number of prompts excessive when their stressor felt straightforward. Future LLM-based interventions could explore adaptive mechanisms that modulate the depth and length of the interaction based on the complexity of the stressor [9], while ensuring that shortening the interaction does not come at the cost of the reappraisal process itself.

We note that while this study was conducted in a workplace setting, the cognitive reappraisal sequence was designed to be applicable across a broad range of stressful situations rather than tailored exclusively to occupational stressors. The workplace context shaped our recruitment, participant population, and the types of stressors participants brought to the activity, but the generality of the intervention was intentional: cognitive reappraisal as a strategy is not inherently

XSL•FO
RenderX

context-specific, and a broadly applicable sequence may be more easily adapted across deployment settings in future work.

We also note that these interventions alone are not sufficient to address the broader challenges of well-being and may serve as modular components within a more comprehensive framework alongside other strategies [29]. These SSIs could be embedded within larger DMH programs spanning multiple weeks and complement practices such as mindfulness activities or psychoeducational content [29]. When our activity is repeated over time, future work should explore how to maintain a sense of freshness and avoid monotony by incorporating variations in structure, tone, or content of the responses. However, the use of LLM-generated content for more severe mental health conditions, including posttraumatic stress disorder or suicidal ideation, would likely require additional safeguards such as clinical review and supervision prior to delivery.

A related design concern that warrants attention as these interventions scale is the tendency of LLM-based systems toward overvalidation [61,62]. While participants in our study generally appreciated the nonjudgmental and affirming tone of the chatbot, some noted that the interaction offered little challenge or pushback. This reflects a broader pattern documented in the literature, where LLMs acting in therapeutic or supportive roles have been found to overvalidate users' beliefs without promoting the kind of cognitive flexibility that can underlie reappraisal [62]. Future design efforts should explore how to calibrate the balance between validation and constructive challenge, for instance by incorporating adaptive feedback mechanisms that respond to the depth and quality of a user's reappraisal attempts.

More broadly, as LLM-based mental health tools become more widely adopted, important questions arise about their relationship to professional care. These tools have the potential to extend support to individuals who would not otherwise access services, but if users come to rely on AI for emotional support in lieu of professional help, there is a risk of eroding motivation to seek human care precisely among those who may need it most. Future work should examine how LLM-based interventions can be designed to complement rather than substitute professional relationships, for example by incorporating referral pathways or escalation mechanisms for users who may require more intensive support [19].

Perceptions of AI Tone and Empathy

Participants shared a range of preferences and reactions regarding the tone of the chatbot and its attempts to convey empathy. Many appreciated the nonjudgmental tone, noting that it offered a space where they could articulate their thoughts without fear of being judged, similar to prior benefits of nonjudgmental care [63,64]. At the same time, the structured sequence of prompts, while helpful for reflection, sometimes made the interaction feel more like a scripted exchange than a fluid conversation. This sort of perception can reduce the potential of DMH interventions [65].

A related tension emerged around the expression of empathy. While prior work has emphasized the importance of empathetic responses in mental health interventions [33,66], some participants were skeptical of AI's capacity to provide genuine emotional support. In some cases, these preexisting beliefs contributed to how participants interpreted the chatbot's attempts at empathy, with some finding them helpful and others perceiving them as forced or artificial.

Future work could address these tensions by reducing the perceived scriptedness of the interaction and more deliberately shaping the agent's persona. A growing body of research has explored the use of human-like traits in conversational agents, and incorporating elements such as tone, conversational style, and empathy level based on models of anthropomorphism [67] or personality frameworks (eg, the big five [68]) could improve user satisfaction. These traits could be either predefined or adapted dynamically based on user preferences and expectations [69]. Designing agents that are also sensitive to how users perceive AI's role in providing support, particularly in sensitive contexts, may help ensure that the tone and level of empathy feel appropriate rather than artificial.

Limitations

Our study has several limitations. First, although pre-post measures allowed us to capture within-subject changes, such designs may be susceptible to response shift bias [70] or demand characteristics [71]. Participants' awareness of being measured before and after the intervention may have influenced how they interpreted or reported their experiences. In addition, several outcomes were assessed using brief or single-item self-report measures selected to reduce participant burden in a single-session study, which may provide less precision than longer validated instruments. Future work deploying similar interventions in a longitudinal or multisession format should prospectively track engagement and assess reasons for noncompletion. Additionally, as a feasibility study, the absence of a control condition means we cannot isolate the contribution of the structured reappraisal sequence from the general effects of expressive writing or conversational engagement.

In our study, we used the GPT-4o model, a state-of-the-art LLM at the time of the study, to generate responses for the reappraisal activity. We designed custom prompts to structure the interaction, which participants generally found appropriate and helpful. However, different models may vary in how they respond to similar prompts, and these differences could influence the perceived quality and usefulness of the interaction. Smaller or less capable models may also struggle to interpret user input or maintain coherence across multiturn conversations. As such, caution is warranted when generalizing the effectiveness of our prompt design or findings across other LLMs. Future research should examine how different models perform in similar interventions and explore alternative strategies, such as fine-tuning, to ensure consistency and relevance in generated content.

Our sample was drawn from a single large technology company, which may limit the generalizability of our findings to other populations and organizational contexts. Future work should examine whether similar results hold across different industries, demographic groups, and cultural settings, and should also prospectively collect time zone and contextual information to examine how the time and setting of engagement influence

XSL•FO
RenderX

outcomes. Additionally, future longitudinal deployments should assess reasons for noncompletion to better understand barriers to sustained use.

Finally, our study primarily examined participants' perceptions of the LLM-enhanced cognitive reappraisal activity, their reflections on the experience and its perceived usefulness. We did not systematically evaluate the quality or fidelity of the chatbot's responses. While the activity was designed to follow a structure informed by evidence-based psychological strategies, future work should assess whether the generated content aligns with those strategies and avoids content that may be inappropriate or counterproductive. Involving clinicians or domain experts in reviewing responses could help ensure that the content remains consistent with established psychology principles.

Conclusions

This study demonstrates the promise of LLMs in delivering structured, adaptive support for stress. Our single-session cognitive reappraisal intervention helped employees reinterpret stressful experiences, leading to significant reductions in immediate stress and improvements in stress mindset. Participants appreciated the guided structure, which offered clarity and validation, while also enabling personalized reflection through contextual responses by the chatbot.

At the same time, integrating LLMs surfaced important design tensions, particularly around the balance between contextualization and user effort, and between structured guidance and conversational naturalness. These insights underscore the need for careful orchestration of structure and flexibility in LLM-based mental health tools. Future work should explore how such interventions can evolve beyond single sessions, adapt dynamically to individual and organizational contexts, and extend to other domains. Our findings contribute both empirical evidence and practical guidance for designing LLM-enhanced DMH interventions.

Acknowledgments

We acknowledge Kori Inkpen, Andy Wilson, Kael Rowan, Ann Paradiso, and Jieun Kim for their help in testing early versions of the research prototype.

Generative AI Use

During manuscript preparation and revision, the authors used OpenAI ChatGPT and Anthropic Claude to assist with language refinement, copyediting, and formatting. The authors reviewed and edited all AI-assisted text and take full responsibility for the content of the manuscript.

Funding

This work was supported by Microsoft Research through institutional resources. No external grant funding was received for this work.

Authors' Contributions

Conceptualization: AB, JH
Data curation: AB
Formal analysis: AB, MC
Investigation: AB
Methodology: AB, JH
Project administration: AB, JH
Software: AB
Supervision: JS, JH
Visualization: AB
Writing – original draft: AB
Writing – review & editing: AB, MC, JS, JH

Conflicts of Interest

AB and MC were research interns at Microsoft Research at the time the study was conducted. JS was an employee of Microsoft at the time of the research, and JH is currently an employee of Microsoft. The authors declare no other competing financial or personal interests related to this work.

Multimedia Appendix 1
Chatbot implementation and study details.
[DOCX File, 3159 KB - humanfactors_v13i1e89736_app1.docx]

XSL•FO
RenderX

References

1. Ford BQ, Karnilowicz HR, Mauss IB. Understanding reappraisal as a multicomponent process: the psychological health benefits of attempting to use reappraisal depend on reappraisal success. Emotion 2017 Sep;17(6):905-911. [doi: 10.1037/emo0000310] [Medline: 28358559]
2. Kitson A, Slovak P, Antle AN. Supporting cognitive reappraisal with digital technology: a content analysis and scoping review of challenges, interventions, and future directions. Presented at: Proceedings of the CHI Conference on Human Factors in Computing Systems; May 11-16, 2024; Honolulu, HI p. 1-17. [doi: 10.1145/3613904]
3. Troy AS, Shallcross AJ, Brunner A, Friedman R, Jones MC. Cognitive reappraisal and acceptance: effects on emotion, physiology, and perceived cognitive costs. Emotion 2018 Feb;18(1):58-74. [doi: 10.1037/emo0000371] [Medline: 29154585]
4. Lazarus RS, Folkman S. Stress, Appraisal, and Coping: Springer Publishing; 1984. URL: https://books.google.com/books?hl=en&lr=&id=i-ySQQuUpr8C&oi=fnd&pg=PR5&dq=Richard+S+Lazarus+and+Susan+Folkman.+1984.+Stress,+appraisal,+and+coping.+Springer+publishing+company.&ots=DhIRhpdePi&sig=-BlhUbXxMuquj1uy77XZ41EdiXQ [accessed 2026-07-27]
5. Gross J. Emotion regulation: conceptual and empirical foundations. In: Handbook of Emotion Regulation: The Guilford Press; 2014:3-20.
6. Stover AD, Shulkin J, Lac A, Rapp T. A meta-analysis of cognitive reappraisal and personal resilience. Clin Psychol Rev 2024 Jun;110:102428. [doi: 10.1016/j.cpr.2024.102428] [Medline: 38657292]
7. Augustine AA, Hemenover SH. On the relative effectiveness of affect regulation strategies: a meta-analysis. Cogn Emot 2009 Sep;23(6):1181-1220. [doi: 10.1080/02699930802396556]
8. Bhattacharjee A, Chen P, Mandal A, et al. Exploring user perspectives on brief reflective questioning activities for stress management: mixed methods study. JMIR Form Res 2024 Feb 8;8:e47360. [doi: 10.2196/47360] [Medline: 38329800]
9. Bhattacharjee A, Williams JJ, Meyerhoff J, Kumar H, Mariakakis A, Kornfield R. Investigating the role of context in the delivery of text messages for supporting psychological wellbeing. Proc SIGCHI Conf Hum Factor Comput Syst 2023 Apr;2023:1-19. [doi: 10.1145/3544548.3580774] [Medline: 37223844]
10. Bhattacharjee A, Williams JJ, Beltzer M, et al. Investigating the role of situational disruptors in engagement with digital mental health tools. Proc ACM Hum-Comput Interact 2025 Oct 18;9(7):1-35. [doi: 10.1145/3757487]
11. Agyapong VIO, Hrabok M, Vuong W, et al. Changes in stress, anxiety, and depression levels of subscribers to a daily supportive text message program (Text4Hope) during the COVID-19 pandemic: cross-sectional survey study. JMIR Ment Health 2020 Dec 18;7(12):e22423. [doi: 10.2196/22423] [Medline: 33296330]
12. Aguilera A, Hernandez-Ramos R, Haro-Ramos AY, et al. A text messaging intervention (StayWell at Home) to counteract depression and anxiety during COVID-19 social distancing: pre-post study. JMIR Ment Health 2021 Nov 1;8(11):e25298. [doi: 10.2196/25298] [Medline: 34543230]
13. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 6;4(2):e19. [doi: 10.2196/mental.7785] [Medline: 28588005]
14. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR mHealth and uHealth. Preprint posted online on 2018. [doi: 10.2196/preprints.12106] [Medline: 30470676]
15. O'Leary K, Schueller SM, Wobbrock JO, Pratt W. "Suddenly, we got to become therapists for each other": designing peer support chats for mental health. Presented at: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems; Apr 21-26, 2018; Montreal, Canada p. 1-14. [doi: 10.1145/3173574.3173905]
16. Zhan H, Zheng A, Lee YK, Suh J, Li JJ, Ong DC. Large language models are capable of offering cognitive reappraisal, if guided. arXiv. Preprint posted online on Aug 8, 2024. [doi: 10.48550/arXiv.2404.01288]
17. Bhattacharjee A, Zeng Y, Mohr DC, et al. Perfectly to a tee: understanding user perceptions of personalized LLM-enhanced narrative interventions. DIS (Des Interact Syst Conf) 2025 Jul;2025:1387-1416. [doi: 10.1145/3715336.3735810] [Medline: 40873971]
18. Li Y, Ding X, Chen Y, Li Y, Ma N. Customizable AI for depression care: improving the user experience of large language model-driven chatbots. Presented at: Proceedings of the 2025 ACM Designing Interactive Systems Conference; Jul 5-9, 2025; Madeira, Portugal p. 1844-1866. [doi: 10.1145/3715336.3735795]
19. Bhattacharjee A, Zeng Y, Xu SY, et al. Understanding the role of large language models in personalizing and scaffolding strategies to combat academic procrastination. Proc SIGCHI Conf Hum Factor Comput Syst 2024 May;2024:1-18. [doi: 10.1145/3613904.3642081] [Medline: 38863598]
20. Sharma A, Rushton K, Lin IW, Nguyen T, Althoff T. Facilitating self-guided mental health interventions through human-language model interaction: a case study of cognitive restructuring. Presented at: Proceedings of the CHI Conference on Human Factors in Computing Systems; May 11-16, 2024; Honolulu, HI p. 1-29. [doi: 10.1145/3613904.3642761]
21. Maddela M, Ung M, Xu J, Madotto A, Foran H, Boureau YL. Training models to generate, recognize, and reframe unhelpful thoughts. Presented at: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); Jul 9-14, 2023. [doi: 10.18653/v1/2023.acl-long.763]

XSL•FO
RenderX

22. Sharma A, Rushton K, Lin I, et al. Cognitive reframing of negative thoughts through human-language model interaction. In: Rogers A, Boyd-Graber J, Okazaki N, editors. Presented at: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); Jul 9-14, 2023; Toronto, Canada p. 9977-10000. [doi: 10.18653/v1/2023.acl-long.555]
23. Ziems C, Li M, Zhang A, Yang D. Inducing positive perspectives with text reframing. Presented at: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); May 22-27, 2022. [doi: 10.18653/v1/2022.acl-long.257]
24. Schleider JL, Dobias ML, Sung JY, Mullarkey MC. Future directions in single-session youth mental health interventions. J Clin Child Adolesc Psychol 2020;49(2):264-278. [doi: 10.1080/15374416.2019.1683852] [Medline: 31799863]
25. Schleider JL, Weisz JR. Little treatments, promising effects? Meta-analysis of single-session interventions for youth psychiatric problems. J Am Acad Child Adolesc Psychiatry 2017 Feb;56(2):107-115. [doi: 10.1016/j.jaac.2016.11.007] [Medline: 28117056]
26. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [doi: 10.2307/2136404] [Medline: 6668417]
27. Aschbrenner KA, Kruse G, Gallo JJ, Plano Clark VL. Applying mixed methods to pilot feasibility studies to inform intervention trials. Pilot Feasibility Stud 2022 Sep 26;8(1):217. [doi: 10.1186/s40814-022-01178-x] [Medline: 36163045]
28. Bowen DJ, Kreuter M, Spring B, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457. [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
29. Meyerhoff J, Beltzer M, Popowski S, et al. Small steps over time: a longitudinal usability test of an automated interactive text messaging intervention to support self-management of depression and anxiety symptoms. J Affect Disord 2024 Jan 15;345:122-130. [doi: 10.1016/j.jad.2023.10.119] [Medline: 37866736]
30. Schleider J, Weisz J. A single-session growth mindset intervention for adolescent anxiety and depression: 9-month outcomes of a randomized trial. J Child Psychol Psychiatry 2018 Feb;59(2):160-170. [doi: 10.1111/jcpp.12811] [Medline: 28921523]
31. Beck JS. Cognitive Behavior Therapy, 3rd edition: Guilford Press; 2020.
32. Rizvi SL, Ritschel LA. Mastering the art of chain analysis in dialectical behavior therapy. Cogn Behav Pract 2014 Aug;21(3):335-349. [doi: 10.1016/j.cbpra.2013.09.002]
33. Sharma A, Lin IW, Miner AS, Atkins DC, Althoff T. Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. Nat Mach Intell 2023;5(1):46-57. [doi: 10.1038/s42256-022-00593-2]
34. Kim M, Kim T, Vo THA, Jung Y, Lee U. Exploring modular prompt design for emotion and mental health recognition. Presented at: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems; Apr 26 to May 1, 2025; Yokohama Japan p. 1-18. [doi: 10.1145/3706598.3713888]
35. Madaan A, Tandon N, Gupta P, et al. Self-refine: iterative refinement with self-feedback. In: Oh A, Naumann T, Globerson A, Saenko K, Hardt M, Levine S, editors. Presented at: Advances in Neural Information Processing Systems 36; Dec 10-16, 2023. [doi: 10.52202/075280-2019]
36. Morshed MB, Hernandez J, McDuff D, et al. Advancing the understanding and measurement of workplace stress in remote information workers from passive sensors and behavioral data. Presented at: 2022 10th International Conference on Affective Computing and Intelligent Interaction (ACII); Oct 18-21, 2022; Nara, Japan p. 1-8. [doi: 10.1109/ACII55700.2022.9953824]
37. Howe E, Suh J, Bin Morshed M, et al. Design of digital workplace stress-reduction intervention systems: effects of intervention type and timing. Presented at: CHI Conference on Human Factors in Computing Systems; Apr 29 to May 5, 2022. [doi: 10.1145/3491102.3502027]
38. Crum AJ, Salovey P, Achor S. Rethinking stress: the role of mindsets in determining the stress response. J Pers Soc Psychol 2013 Apr;104(4):716-733. [doi: 10.1037/a0031201] [Medline: 23437923]
39. Demerouti E, Bakker AB, Nachreiner F, Schaufeli WB. The job demands-resources model of burnout. J Appl Psychol 2001 Jun;86(3):499-512. [doi: 10.1037/0021-9010.86.3.499] [Medline: 11419809]
40. Hartmann J, Heitmann M, Siebert C, Schamp C. More than a feeling: accuracy and application of sentiment analysis. Int J Res Mark 2023 Mar;40(1):75-87. [doi: 10.1016/j.ijresmar.2022.05.005]
41. Dstefa/roberta-base_stress_classification. Hugging Face. 2025 Dec 7. URL: https://huggingface.co/dstefa/roberta-base_stress_classification [accessed 2026-10-13]
42. Baysan MS, Uysal S, İşlek İ, Çığ Karaman Ç, Güngör T. LLM-as-a-judge: automated evaluation of search query parsing using large language models. Front Big Data 2025;8:1611389. [doi: 10.3389/fdata.2025.1611389] [Medline: 40761620]
43. Chandra M, Sriraman S, Khanuja HS, Jin Y, De Choudhury M. Reasoning is not all you need: examining llms for multi-turn mental health conversations. arXiv. Preprint posted online on May 28, 2025. [doi: 10.48550/arXiv.2505.20201]
44. Chandra M, Sriraman S, Verma G, et al. Lived experience not found: llms struggle to align with experts on addressing adverse drug reactions from psychiatric medication use. In: Chiruzzo L, Ritter A, Wang L, editors. Presented at: Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics; Apr 29 to May 4, 2025; Albuquerque, NM p. 11083-11113. [doi: 10.18653/v1/2025.naacl-long.553]
45. Gao M, Ruan J, Sun R, Yin X, Yang S, Wan X. Human-like summarization evaluation with ChatGPT. arXiv. Preprint posted online on Apr 5, 2023. [doi: 10.48550/arXiv.2304.02554]

XSL•FO
RenderX

46. Han B, Yau C, Lei S, Gratch J. Knowledge-based emotion recognition using large language models. Presented at: 2024 12th International Conference on Affective Computing and Intelligent Interaction (ACII); Sep 15-18, 2024; Glasgow, Scotland p. 1-9. [doi: 10.1109/ACII63134.2024.00005]
47. Clarke V, Braun V. Thematic analysis. J Posit Psychol 2017 May 4;12(3):297-298. [doi: 10.1080/17439760.2016.1262613]
48. McDonald N, Schoenebeck S, Forte A. Reliability and inter-rater reliability in qualitative research: norms and guidelines for CSCW and HCI practice. Proc ACM Hum Comput Interact 2019;3:1-23. [doi: 10.1145/3359174]
49. Yang L, Neustaedter C. Our house: living long distance with a telepresence robot. Proc ACM Hum Comput Interact 2018;2:1-18. [doi: 10.1145/3274459]
50. Jurkiewicz O, McGarrigle CB, Oveis C. How to improve others' emotions: reappraise and be responsive. Affect Sci 2023 Jun;4(2):233-247. [doi: 10.1007/s42761-023-00183-4] [Medline: 37293680]
51. Batum P, Yağmurlu B. What counts in externalizing behaviors? The contributions of emotion and behavior regulation. Curr Psychol 2007;25:272-294. [doi: 10.1007/BF02915236]
52. Bhattacharjee A, Gong Z, Wang B, et al. "Actually I Can Count My Blessings": user-centered design of an application to promote gratitude among young adults. Proc ACM Hum-Comput Interact 2024 Nov 7;8(CSCW2):1-29. [doi: 10.1145/3686936]
53. Nepal S, Pillai A, Campbell W, et al. Contextual AI journaling: integrating LLM and time series behavioral sensing technology to promote self-reflection and well-being using the MindScape app. Ext Abstr Hum Factors Computing Syst 2024 May;2024:86. [doi: 10.1145/3613905.3650767] [Medline: 39072254]
54. Ullrich PM, Lutgendorf SK. Journaling about stressful events: effects of cognitive processing and emotional expression. Ann Behav Med 2002;24(3):244-250. [doi: 10.1207/S15324796ABM2403_10] [Medline: 12173682]
55. Bhattacharjee A, Sultana S, Amin MR, Iqbal Y, Ahmed SI. "What's the Point of Having This Conversation?": from a telephone crisis helpline in Bangladesh to the decolonization of mental health services. ACM J Comput Sustain Soc 2023 Dec 31;1(2):1-29. [doi: 10.1145/3616381]
56. Fernando S. Mental Health Worldwide: Culture, Globalization and Development: Springer Nature; 2014. URL: https://www.researchgate.net/profile/Suman-Fernando/publication/311295777_Mental_health_worldwide_Culture_globalization_and_development/links/5be88b1a4585150b2bb03b5a/Mental-health-worldwide-Culture-globalization-and-development.pdf [accessed 2026-07-27]
57. Fox KE, Johnson ST, Berkman LF, et al. Organisational- and group-level workplace interventions and their effect on multiple domains of worker well-being: a systematic review. Work Stress 2022 Jan 2;36(1):30-59. [doi: 10.1080/02678373.2021.1969476]
58. Hoddinott P, Allan K, Avenell A, Britten J. Group interventions to improve health outcomes: a framework for their design and delivery. BMC Public Health 2010 Dec 31;10:800. [doi: 10.1186/1471-2458-10-800] [Medline: 21194466]
59. Kilanowski JF. Breadth of the socio-ecological model. J Agromedicine 2017;22(4):295-297. [doi: 10.1080/1059924X.2017.1358971] [Medline: 28742433]
60. Yang T, Choi I. Reflection as a social phenomenon: a conceptual framework toward group reflection research. Education Tech Research Dev 2023 Apr;71(2):237-265. [doi: 10.1007/s11423-022-10164-2] [Medline: 35669695]
61. Iftikhar Z, Ransom S, Xiao A, Nugent N, Huang J. Therapy as an NLP task: psychologists' comparison of llms and human peers in CBT. arXiv. Preprint posted online on Jun 25, 2025. [doi: 10.48550/arXiv.2409.02244]
62. Iftikhar Z, Xiao A, Ransom S, Huang J, Suresh H. How LLM counselors violate ethical standards in mental health practice: a practitioner-informed framework. Proc AAAI/ACM Conf AI Ethics Soc 2025;8(2):1311-1323. [doi: 10.1609/aies.v8i2.36632]
63. Siette J, Cassidy M, Priebe S. Effectiveness of befriending interventions: a systematic review and meta-analysis. BMJ Open 2017 Apr 26;7(4):e014304. [doi: 10.1136/bmjopen-2016-014304] [Medline: 28446525]
64. Koh A. Non-judgemental care as a professional obligation. Nurs Stand 1999;13(37):38-41. [doi: 10.7748/ns1999.06.13.37.38.c2612] [Medline: 10455688]
65. Borghouts J, Eikey E, Mark G, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387. [doi: 10.2196/24387] [Medline: 33759801]
66. Saha T, Gakhreja V, Das AS, Chakraborty S, Saha S. Towards motivational and empathetic response generation in online mental health support. Presented at: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval; Jul 11-15, 2022. [doi: 10.1145/3477495.3531912]
67. Kuzminykh A, Sun J, Govindaraju N, Avery J, Lank E. Genie in the bottle: anthropomorphized perceptions of conversational agents. Presented at: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems; Apr 25-30, 2020. [doi: 10.1145/3313831.3376665]
68. Soldz S, Vaillant GE. The big five personality traits and the life course: a 45-year longitudinal study. J Res Pers 1999 Jun;33(2):208-232. [doi: 10.1006/jrpe.1999.2243]
69. Bhattacharjee A, Suh J, Ershadi M, Iqbal ST, Wilson AD, Hernandez J. Understanding communication preferences of information workers in engagement with text-based conversational agents. arXiv. Preprint posted online on Nov 1, 2024. [doi: 10.48550/arXiv.2410.20468]

XSL•FO
RenderX

70. Howard GS, Dailey PR. Response-shift bias: a source of contamination of self-report measures. J Appl Psychol 1979;64(2):144-150. [doi: 10.1037/0021-9010.64.2.144]
71. Nichols AL, Maner JK. The good-subject effect: investigating participant demand characteristics. J Gen Psychol 2008 Apr;135(2):151-165. [doi: 10.3200/GENP.135.2.151-166] [Medline: 18507315]

Abbreviations

BERT: Bidirectional Encoder Representations from Transformers
CBT: cognitive behavioral therapy
DMH: digital mental health
HCI: human-computer interaction
IRB: institutional review board
LLM: large language model
NLP: natural language processing
PSS: Perceived Stress Scale
RoBERTa: Robustly Optimized BERT Pretraining Approach
RQ: research question
SSI: single-session intervention

Edited by A Kushniruk; submitted 05.Jan.2026; peer-reviewed by B Suh, C Schillings, R Chaudhury; revised version received 18.Jun.2026; accepted 23.Jun.2026; published 22.Oct.2026.

Please cite as:
Bhattacharjee A, Suh J, Chandra M, Hernandez J
User Perceptions of a Large Language Model–Based Chatbot for Cognitive Reappraisal of Stress: Mixed Methods Feasibility Study
JMIR Hum Factors 2026;13:e89736
URL: https://humanfactors.jmir.org/2026/1/e89736
doi:10.2196/89736

© Ananya Bhattacharjee, Jina Suh, Mohit Chandra, Javier Hernandez. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 22.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Redesign of Bedside Supply Carts to Improve Emergency Department Workflows: Mixed Methods Participatory Design

Kat Hefter[1], MSE, JD; Sammy Dreibelbis[2], BFA; Theresa Haupt[1], MSE; Amelia McIver[1], MSE; Agnes Wang[3], MSE; Julia Dwight[4], BS, BSN; Ogechi Nwodim[5], MD; Neil Ray[5], MD

[1]Department of Bioengineering, School of Engineering and Applied Sciences, University of Pennsylvania, 210 S. 33rd St., Suite 240, Skirkanich Hall, Philadelphia, PA, United States

[2]Department of Integrated Product Design, School of Engineering and Applied Sciences, University of Pennsylvania, Philadelphia, PA, United States

[3]Department of Data Sciences, School of Engineering and Applied Sciences, University of Pennsylvania, Philadelphia, PA, United States

[4]Emergency Department, University of Pennsylvania Health System, Philadelphia, PA, United States

[5]Department of Emergency Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Kat Hefter, MSE, JD
Department of Bioengineering, School of Engineering and Applied Sciences, University of Pennsylvania, 210 S. 33rd St., Suite 240, Skirkanich Hall, Philadelphia, PA, United States

Abstract

Background: Emergency departments are often chaotic environments where delays can significantly impact patient care. Key items are stored in supply carts in or near patient rooms to promote efficiency and enable nurses to spend more time assisting patients. However, disorganization, lack of standardization, and lack of stocking can cause significant delays and negatively impact the quality of care.

Objective: This study utilized human-centered and participatory design to improve the workflow for supply acquisition in an emergency department.

Methods: Using a mixed methods, participatory design approach following the double diamond framework, the team worked with nursing staff and physicians in an urban emergency department to understand the root causes of frustrations with the current supply carts. Qualitative findings about bedside nursing workflows were integrated with quantitative observations of inventory and supply usage to drive a rapid-cycle prototyping process to optimize supply management in the bedside cart.

Results: A lack of clinical staffing exacerbates preexisting challenges with restocking the medical supplies in the bedside carts. This problem is compounded by the misallocation of supplies, with high-frequency items underrepresented and low-frequency items overrepresented in the bedside carts. This leads to wastage of the seldom-used supplies and a lack of access to the most used supplies. The reorganization of the cart through co-design with nursing staff sped up supply acquisition by approximately 20% overall, tripled the availability of the most important supplies, and reduced the need for restocking from once per shift to once per 3 shifts, thus producing tangible improvements even within institutional limitations.

Conclusions: A participatory design process, using human factors principles in tandem with extensive input from end users, enables improvements to stocking. Implications for practice include (1) lack of easy access to appropriate supplies negatively impacts patient care and contributes to nurse burnout and frustration, (2) human factors engineering can improve access to patient care supplies through redesigning the layout of hospital supply carts to better align with workflows, and (3) co-design with frequent collaboration from stakeholders and end users ensures that solutions address the issues that matter most in a sustainable way.

(JMIR Hum Factors 2026;13:e80861) doi:10.2196/80861

KEYWORDS

cluster analysis; emergency department; ergonomic workflow; human-centered design; inventory; medical supply cart; patient care

Introduction

In a busy emergency department (ED), every second counts. Medical professionals require access to key supplies and medication in an efficient and timely manner [1,2]. These key items are often stored in supply carts in or near patient rooms in order to promote efficiency [3-6]. However, disorganization and lack of standardization can delay patient care and negatively impact outcomes [1-3,7-9]. The lack of supplies is also a frequent cause of operational failures. Studies have shown that

XSL•FO
RenderX

access to supplies is a leading factor in "time wasted" in an emergency setting [3,9,10], and a lack of familiarity with a cart system has been identified as one of the most common factors adversely affecting the quality of care in a high-stakes medical setting [11].

Human factors engineering has emerged as a dominant method to address challenges in stocking [1,2,4,7,8,12-15]. For instance, one study found that reorganizing and standardizing the contents of a resuscitation cart in a children's hospital according to the 5S Lean design method reduced the time needed to acquire supplies by 46% [7]. Similarly, the reorganization of a hospital medication cart based on usability testing and simulation reduced the number of wasteful actions taken while searching for the proper supply [12].

While groups vary in the methods they used, one thing remains true in each case: the first step requires understanding who is most affected by the challenges caused by the existing system.

The burden of a disorganized supply system often falls heavily on nursing staff, who are already struggling with staffing shortages and heavy workloads [16,17]. Nurses report frequently having to leave rooms to search for supplies, increasing job frustration, delaying patient care, and reducing the time available to spend with patients [9,18]. One study found that more than one-third of nurses' time is spent looking for equipment, with an annual estimated cost of US $1 million in lost time per year [10].

Nurse feedback and input are therefore a crucial component of any redesign [14,15], both to optimize any new system based on nurses' lived experience [17] and to promote approval and uptake of a new system [1,19,20]. Nurse feedback and participation also give nurses a voice in their workplace, increasing satisfaction, improving work performance, and helping to provide an environment for safe, quality patient care [14,21].

Here, we discuss the use of a participatory human factors approach to redesign the bedside cart supply in the ED of a low-resourced urban community hospital. The aim of this mixed methods co-design intervention was to understand the challenges associated with restocking and access to patient care supplies and implement a strategy to improve reliable and efficient access to supplies for nurses in an ED. Our novel approach, which combined the use of the double diamond design framework with a participatory design, allowed us to reduce the time required to access supplies by reorganizing the carts to hold more of the most crucial supplies and minimizing wasted space.

Methods

Overview

In this paper, we report on the use of a data-driven and participatory design approach to address the supply shortage in ED bedside carts. Our approach followed the double diamond design framework [22]. The double diamond design framework is a human-centered framework that breaks a design into 2 main phases, each composed of a divergent step of discovery and ideation, followed by a convergent step distilling findings into a concise and workable next step. In addition, we relied heavily on principles of co-design [23], working closely with stakeholders at every step of the process to ensure any solutions fit within the constraints of the setting and appropriately addressed their concerns.

Following this framework, the design process was broken into 3 main phases. In the first phase of the double diamond, the team immersed itself in the clinical environment and engaged with stakeholders to gather information. Based on these observations, they then performed a root cause analysis to discover and define the true problems affecting supply accessibility. Finally, they underwent a co-design process with hospital staff to develop and deliver a new bedside cart prototype that successfully addressed the issue. The results are organized according to the SQUIRE (Standards for Quality Improvement Reporting Excellence) reporting guidelines for quality improvement in health care [24].

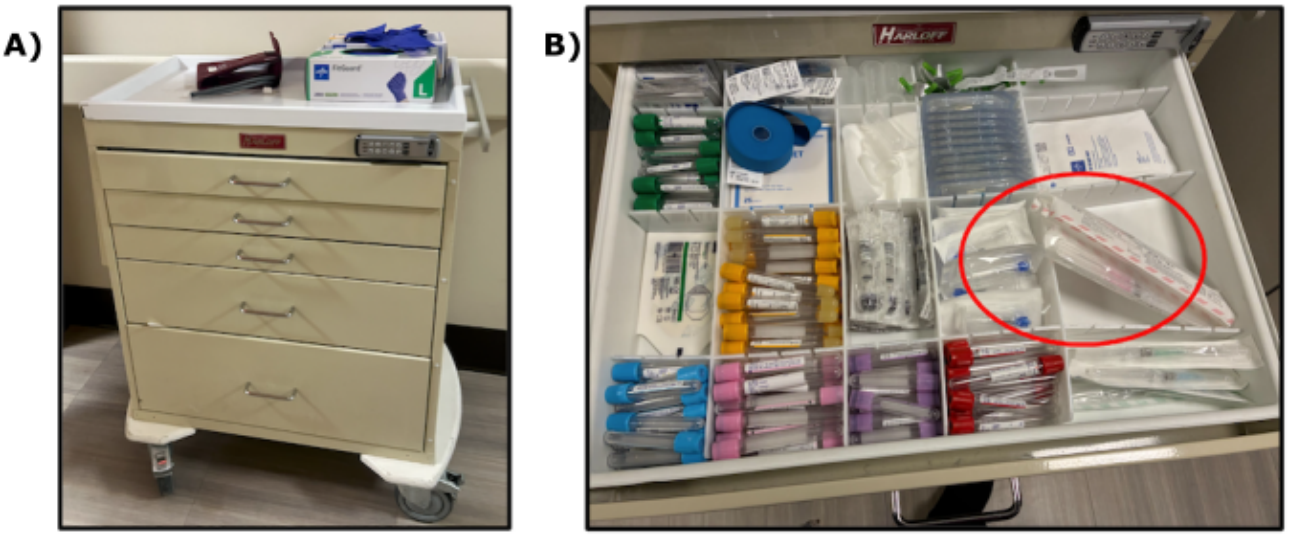
Study Background and Location

This study was conducted in the Emergency Department of the Hospital of the University of Pennsylvania's campus at Cedar Avenue (HUP-Cedar), a low-resourced urban ED in eastern Pennsylvania. The ED serves more than 48,000 patients every year and has inpatient medical and psychiatric services, with very little specialty coverage. Children, patients who are critically injured, or patients with trauma are generally sent to other hospitals in the region, and the admission rate for patients is 10% to 13%. The department is split into a 23-bed high acuity area staffed with a 1:4 nursing to patient ratio and a 10-bed low acuity area staffed with 1 to 2 nurses.

The ED utilizes a lockable 5-drawer Harloff ''bedside" cart positioned in the hallway outside each patient room. Drawers have varied heights; the top 3 are each approximately 2.5″ tall, the fourth drawer is 5.5″ tall, and the bottom drawer is 8.5″ tall. Supplies are separated by thin, $\frac{1}{8}$″ acrylic dividers in compartments that often fail to accommodate longer items such as needles and syringes (Figure 1). In addition, heavy items such as intravenous (IV) bags have deformed many of the carts over the years. Carts are often understocked given limitations in staff availability, causing nurses to search 3 or 4 carts to find necessary supplies. In addition, the underutilized supplies may linger in the cart for years well past their expiration dates. The research team was embedded within the ED from January 2025 to May 2025; the full process can be seen in Figure S1 in Multimedia Appendix 1. The team members included multiple engineers with specialties in mechanical design, product design, data science, human factors, and systems engineering and 2 emergency medicine physicians. The project also included a nurse who provided significant input and helped the embedded team interface with other members of the nursing staff.

XSL•FO
RenderX

Figure 1. Original cart system. (A) Lockable 5-drawer Harloff "bedside" cart positioned in the hallway of the emergency department. (B) Original layout and dividers in drawer 1 (topmost drawer) of the bedside carts. Note that needles (indicated by the red circle) are stored at a diagonal because the allocated space is too short.



Clinical Immersion and Key Stakeholder Interviews

The team began by observing users interacting with the bedside cart and interviewing key stakeholders (Multimedia Appendix 2) to understand the key challenges and concerns that they saw with the current carts, and to understand the supplies they needed the most. Interviews were semistructured [25] with a predetermined list of questions, but conversations varied according to participant interests. Interviews were conducted with multiple team members present; 1 person led the conversation, while others took detailed notes to ensure no information was missed. After 60 hours of observation, and after the insights from observation and interviewing had reached saturation [26,27], the team used a thematic analysis [28] to identify the key unmet needs (Table 1). These needs were subsequently validated through conversations with nursing staff.

Table . Sample thematic analysis of user statements.

Insight	Example statements
Unreliable or unavailable supply puts patient lives at risk.	• *[When I try to use the carts,] the things I need to save my patient's life aren't there.* • *I'd rather save patient lives than have a clean cart. I can clean the cart later, but I can't bring them back.* • *I can't leave a seizing patient to go look for oxygen masks because my cart didn't have any.*
Stocking is time-consuming and irritating, and there is often no one to do it.	• *The techs are pulled to do other tasks like 1-on-1, and no one has time to restock the carts.* • *If we had safer ratios, I'd be more willing to stock the carts.* • *I just want my carts stocked. The rest, I can work with.* • *Anything to make things a little easier or streamlined, especially with refilling the carts, would be huge.*
Nurses do not trust the carts to have the supplies they need and often use workarounds.	• *If cart supplies are low, I'll get a basin of the most used things and put them on a cart so I know where to find them.* • *I run from cart to cart all day, then say 'F** it, I'm just going to the stockroom'.* • *I get things from the Clean Supply rather than the carts so I don't waste my time looking.*
There is not a perfect consensus over what should be in the carts.	• *I've been going from drawer to drawer looking for socks for my patient.* • *After the emergency, that's when we can put on socks.*
Cart contents do not match what nurses actually need.	• *A lot of things in the cart are rarely used - they're a waste of space.* • *I really want a space for the ultrasound needles.* • *When there are things in the cart that aren't used, they expire and we have to throw them out.*

XSL•FO
RenderX

Co-Design With Key Stakeholders and Initial Solution Selection

Focused solutions to the supply problem were addressed in a series of co-design sprints to determine how to reform the supply distribution system. A multidisciplinary group of stakeholders (including the research team, additional physicians, and nurses) underwent a series of design thinking brainstorming exercises to generate over 40 potential solutions. From there, ideas were screened for viability (whether the solution addressed the core problem), desirability (whether the stakeholders would like and use the solution), and feasibility (whether the solution would be possible given other constraints). From the session, the team identified 6 key domains—accessibility, legality, reliability, ease of restocking, capacity, and feasibility—required for a successful solution based on the unmet needs. Subsequently, screening and scoring matrices were implemented to identify the top solution [29]. Although the original solution selected was to pull out the key items used most often by nurses and put them in their own designated carts, obtaining additional carts would have become a challenge. The team sought to determine whether a similar result could be obtained merely by reallocating the proportions of various items in the existing carts to align with usage and need.

Quantitative Analysis of Current Inventory System and Item Usage

An inventory of 6 different carts throughout the ED was obtained to understand the current supply state. Maximum capacity was calculated as the highest value from each of the carts. During the observation period, the team also collected data on the supplies used to care for 17 different patients with different chief complaints (16 h of observation). This allowed the team to see supplies being used in real time and observe nursing micromovements as well as get a sense of overall inventory throughout the department.

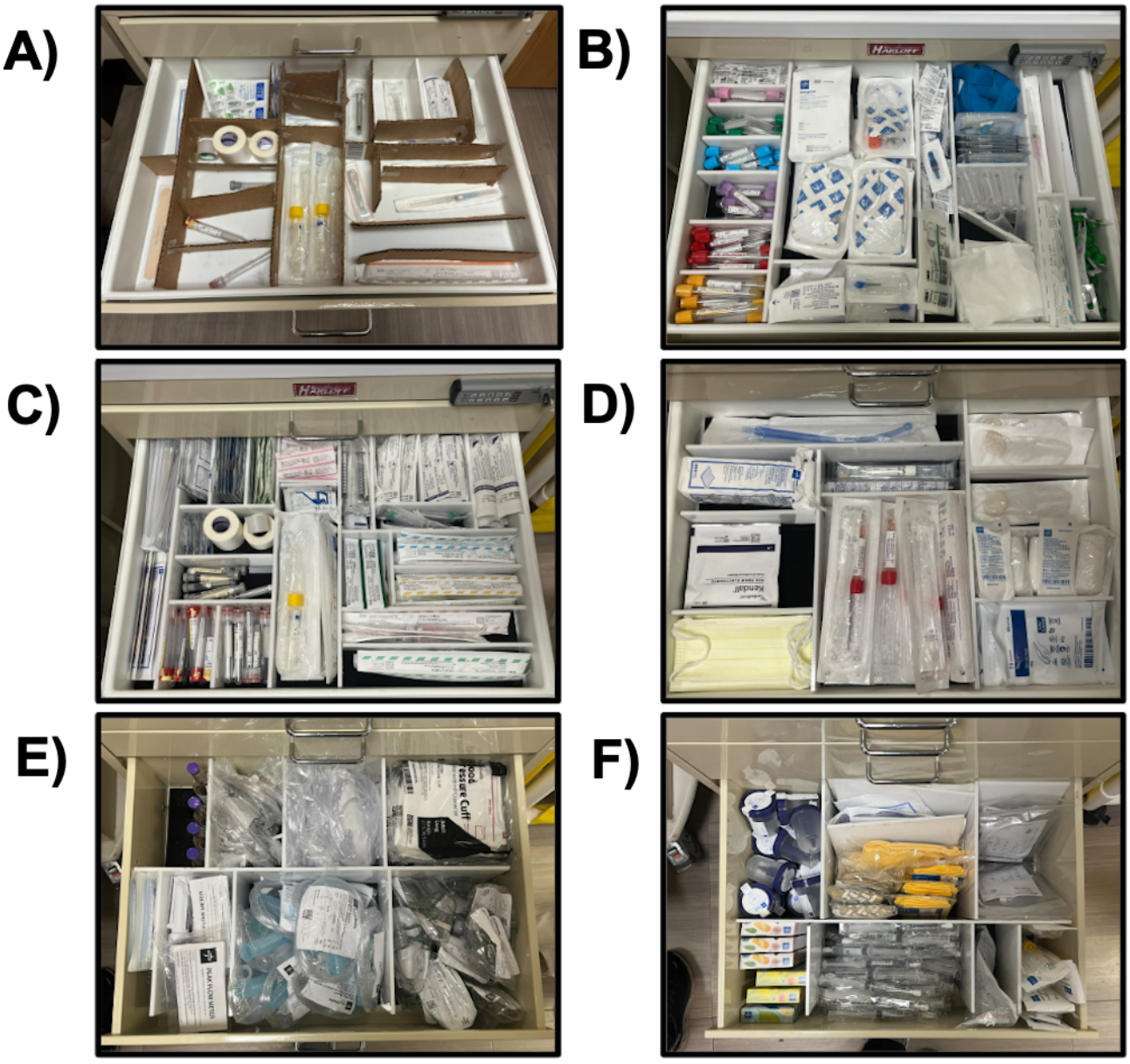
From there, the team ran a k-means clustering analysis on the supply usage data to identify if there were any clusters of supplies that were frequently used together. While preliminary data were too limited to generate many solid clusters, the initial results using k-means with 2 clusters revealed a small cluster of high-use items and a much longer list of less frequent items. These high-use items were then compared with reports given by nurses (Multimedia Appendix 3).

Rapid Prototyping of New Drawer Organization System

Through conversation with nursing staff, the team obtained (1) a list of all of the supplies nurses wanted in the supply carts, (2) suggestions for changes in quantity for items already present, and (3) requests for layout changes to improve workflow. The team compared the wish lists to the inventory lists and eliminated unnecessary items, then integrated the qualitative suggestions with their previous quantitative results to create prototypes for drawer dividers to better allocate cart space. This was done in a “jigsaw” fashion [8]: the team built early prototypes from cardboard and tape using real supplies as a guide, then asked the nurses to move around the dividers as needed (Figure 2A). The team also asked nurses to model the movements needed to acquire items so that the new compartments had an ergonomic flow.

XSL•FO
RenderX

Figure 2. Prototype to finalized product. (A) Cardboard prototype used in an interactive design process with nurses (drawer 1); (B) finalized, fully stocked prototype for drawer 1; (C) drawer 2; (D) drawer 3; (E) drawer 4; and (F) drawer 5.



Prototypes were designed according to the following principles, all based in human factors engineering: (1) items that were used together should be located in the same drawer and in close proximity, (2) higher-frequency items should be on higher shelves to optimize body posture, (3) compartments should conform to the size and shape of the item, (4) items should fit enough supply for at least 1 full shift for a nurse, and (5) designs should minimize exterior changes to the cart to minimize relearning.

Once nurses provided feedback on the cardboard prototypes, the team incorporated those changes into new prototypes (Figure 2B-E). Team members created computer-assisted design drawings using OnShape [30] for each of the drawer bases and inserts based on the final cardboard models, then laser-cut prototypes made from ¼″ acrylic using a laser cutter.

Simulation and Evaluation of Design

To ensure that the dividers were an improvement over the original models, the team ran a variety of simulations with the help of nursing staff. The team created four scenarios that would require different combinations of supplies: (1) a patient ready for discharge, (2) a patient with fever and a recurrent headache, (3) a patient with chest and abdominal pain, and (4) a patient with multiple constitutional complaints. The number of scenarios was based on nurses' estimates of how many patients they would see in each room throughout the duration of their 12-hour shift; scenarios were intentionally left vague.

The team fully stocked both an existing cart and the prototype and asked on-shift nurses to collect the necessary supplies for each of the patient scenarios. Items removed from the cart were not replaced but instead set aside; nurses were told that if an item was not present in the cart when they needed it, they should go about their normal routine, whatever that was. This often involved searching another cart, or going to the clean supply,

XSL•FO
RenderX

substituting a different item or collection of items, or foregoing that item altogether. Nurses were quasi-randomly assigned, in equal numbers, to start the simulation with either the prototype cart or the current standard to grab items from either the old or new carts first so that the time required to figure out what items to select did not bias the overall time. When all 4 scenarios were conducted with 1 cart, they would then run the same scenarios with the other cart.

Carts were not restocked between sets of simulations ("shifts"). This was to mimic real-time usage, where stockers often would not be able to resupply carts until multiple shifts had elapsed. After 3 shifts had elapsed without restocking, the carts were fully restocked, and the simulations were run again with different nurses.

Ethical Considerations

The University of Pennsylvania's Institutional Review Board (IRB) waived the need for ethics approval and participant consent for the collection, analysis, and publication of the anonymized data for this quality improvement initiative. Informed consent was obtained verbally before participation. All data have been anonymized. No identifying information about the nurses was collected, and the team was not assessing their responses for clinical accuracy. Nurses were not compensated for their participation.

Results

Clinical Immersion and Key Stakeholder Interviews

The team spoke with a wide range of stakeholders, hoping to engage with people throughout the entire supply workflow. This included nurses, doctors, and physician assistants who used the carts; technicians who stocked the carts; and even the supply manager who oversaw stocking and ensured that the department remained in compliance with hospital policy and external regulations. The full chart of interviewees can be found in Multimedia Appendix 2.

During the observation period, the team was told that the bedside carts were fully stocked by emergency technicians and were positioned outside patient rooms so that nurses could efficiently provide patient care. In practice, however, carts were constantly understocked, and nurses frequently visited multiple carts searching for supplies. As a result of this, some nurses no longer relied on the carts and instead went to the central supply closet every time they needed something or carried a personal bucket of supplies they thought they might need, contrary to hospital policy.

After roughly 60 hours of observations and thematic analysis of statements obtained during the interviews (Table 1), it became clear that the current setup did not meet the supply and demand for current nursing staff. For example, each cart could hold only 2 IV start kits, while nurses reported using at least 1 per patient. In addition, the cart could only hold up to three 1-L bags of IV fluids, when a nurse went through an average of 5 to 10 bags per shift across all of their rooms. As a result, these key supplies were constantly in short supply and frequently depleted.

Second, while technicians were responsible for restocking these carts, they lacked sufficient time to do so amid widespread staffing shortages and a host of other, often more critical responsibilities such as sitting with patients who required continuous monitoring or helping to transport patients. Nurses could assist with restocking when time permitted, but ED shifts were often busy, and this was usually impossible. Carts often remained unstocked for several shifts, often for days on end.

In addition, while there was a photographic guide on top of the cart to assist with stocking, there was a lack of consistency in the items inside the cart. Over time, nurses had begun to stock the carts with their preferred items and formats. The clean supply room was aptly labeled with items in their respective categories—IV, wound care, comfort, respiratory, fluids, and so forth—but this system did not correspond to the organization in drawers.

Finally, the observations and interviews demonstrated the importance of buy-in for any potential solution. Nurses expressed concern that any additional work from a proposed solution would fall upon their already-overloaded shoulders and stated outright that there were certain proposals that they would simply not implement.

Co-Design With Key Stakeholders and Initial Solution Selection

The interdisciplinary team held a design sprint workshop with several nurses to generate over 40 potential solutions. Solutions ranged from developing new technologies and processes to detect when a cart was empty, to hiring a new person with the sole job of restocking carts, to altering the workflows of staff members in other departments to help assist with restocking, to even assigning each nurse their own cart to be responsible for and stock as they wished. However, solutions were limited by resources and by staffing shortages and workload. The hospital was unable to allocate money to hire a new person or replace the existing carts, and the team did not identify anyone in the environment with the extra time and ability to add restocking to their workflow. In addition, policy restrictions meant that each cart needed to be standardized. Additionally, nurses and doctors rejected any solution that involved making restocking their responsibility entirely. Given these limitations, the most feasible solution involved a reorganization of the cart system, utilizing the existing space but reconfiguring it to increase capacity for high-use items.

Quantitative Analysis of Current Inventory System and Item Usage

Before making any attempts to redesign the cart capacity, the team needed to understand both the original layout and typical patterns of item usage. The team thus developed an inventory to track supply usage for nurses across a total of 16 hours to capture the relative frequency of supply usage, as well as the types of supplies often used in conjunction with each other. From this information, and through consultation with nursing staff, the team was able to isolate 3 key clusters of highly used supplies: IV, diagnostics, and respiratory supplies. The contents of each of these clusters can be found in Multimedia Appendix 3. They also obtained an approximate estimate of the amount

XSL•FO
RenderX

of each of these supplies needed per day. IV start kits were the most highly used item, as they were used with nearly every patient. Along with the kits, nurses also used flushes, heplocks, and vacutainers.

Rapid Prototyping of New Drawer Organization System

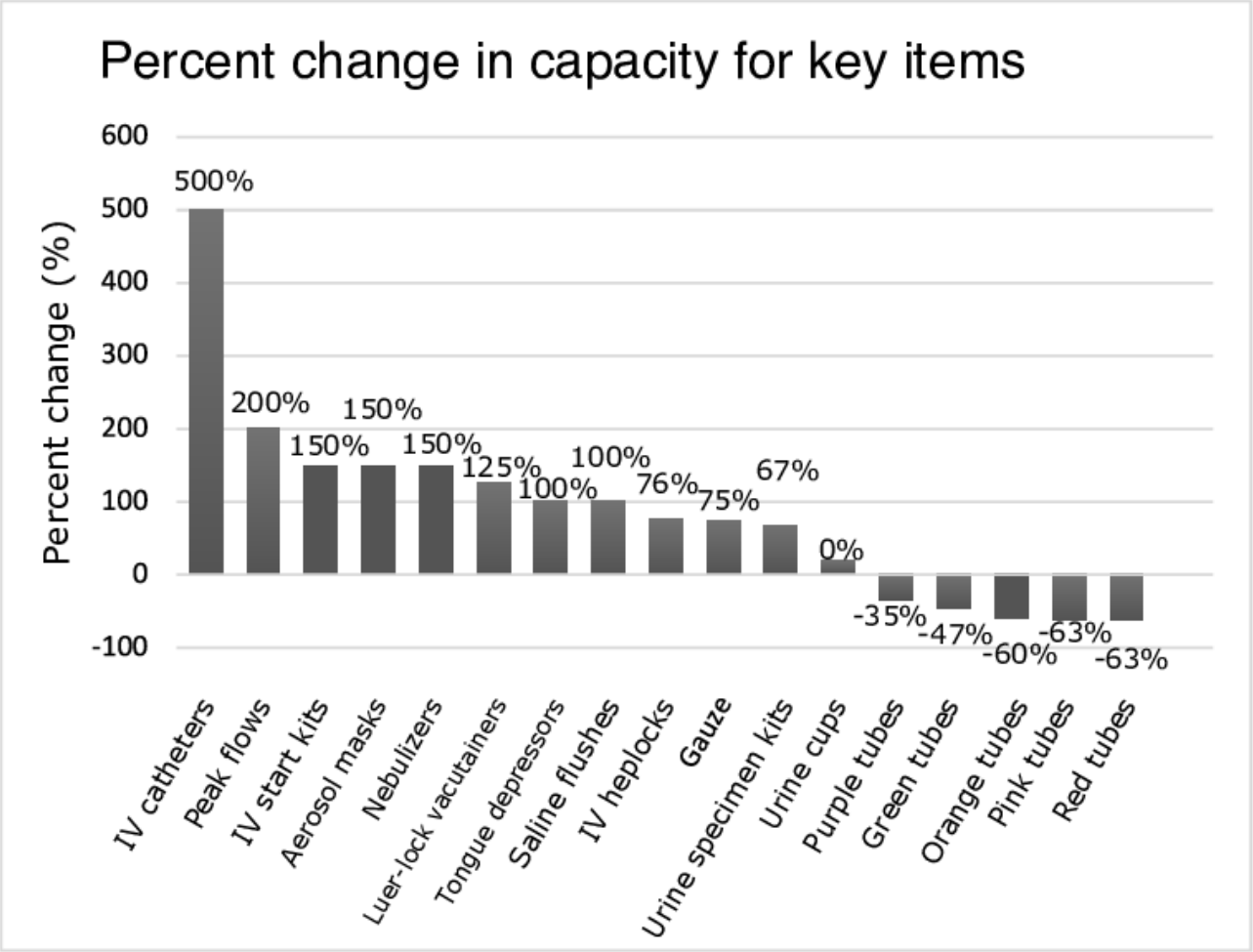
After undergoing the rapid prototyping process described above and shown in Figure 2, the new design provided a marked improvement over the old model. The layouts for the old versus original designs can be seen in Figure 3, and the difference in capacity between the 2 designs can be seen in Figure 4 (key supplies only; full list in Multimedia Appendix 4). Specifically, for the IV kits, the new carts comfortably held 9 kits, while previous models held only 2 or 3—a 200% increase for the most highly used supply. Co-occurring supplies, such as the flushes (100% increase), vacutainers (125% increase), and heplocks (75% increase), similarly increased in capacity. Given the total number of carts on the floor, this meant that the number of IV kits stored in the carts jumped from approximately 100 to approximately 300—enough to make it through multiple shifts without restocking. The new cart design similarly doubled the size of spaces allocated for key respiratory supplies such as peak flow meters (200% increase), nebulizers (150% increase), and aerosol masks (150% increase).

Figure 3. Drawer 1 original versus new layout. Colored boxes highlight key changes that were made. Blue: Needles are no longer stored at a diagonal with no decrease in capacity. Purple: Hep-lock capacity increased by 75%. Green: Intravenous start kit capacity increased by 200%. Orange: Storing all tubes in 1 column makes for a more ergonomic workflow; capacity decreased to avoid wastage.



XSL•FO
RenderX

Figure 4. Cart redesign corrects supply allocation. After redesign, critical items like those used with intravenous (IV) or for urine collection and testing increased in capacity, while overrepresented test tubes decreased in capacity.



In many cases, this was achieved by removing low-frequency items that were rarely used and could be more effectively stored in the central supply closet. In other cases, it was achieved by ensuring each compartment fit the size of the tool it was meant to contain, eliminating wasted space and reducing the chances of needles being damaged or bent by attempting to force them into the cart. The new design also allowed space for the addition of other supplies, such as a speculum, that were highly sought after by physicians and nurses and could not fit in the cart previously. The design met with approval from nursing and administrative staff at all stages of the process.

In addition, the finalized model was both sturdier and less prone to breakage than the original model. This was the result of several iterations: the first draft, which used an acrylic solvent, proved unstable and snapped upon construction, so the team reconfigured the designs to use a press-fit layout instead. With this approach, drawers were easy to assemble and sturdy enough to withstand constant use.

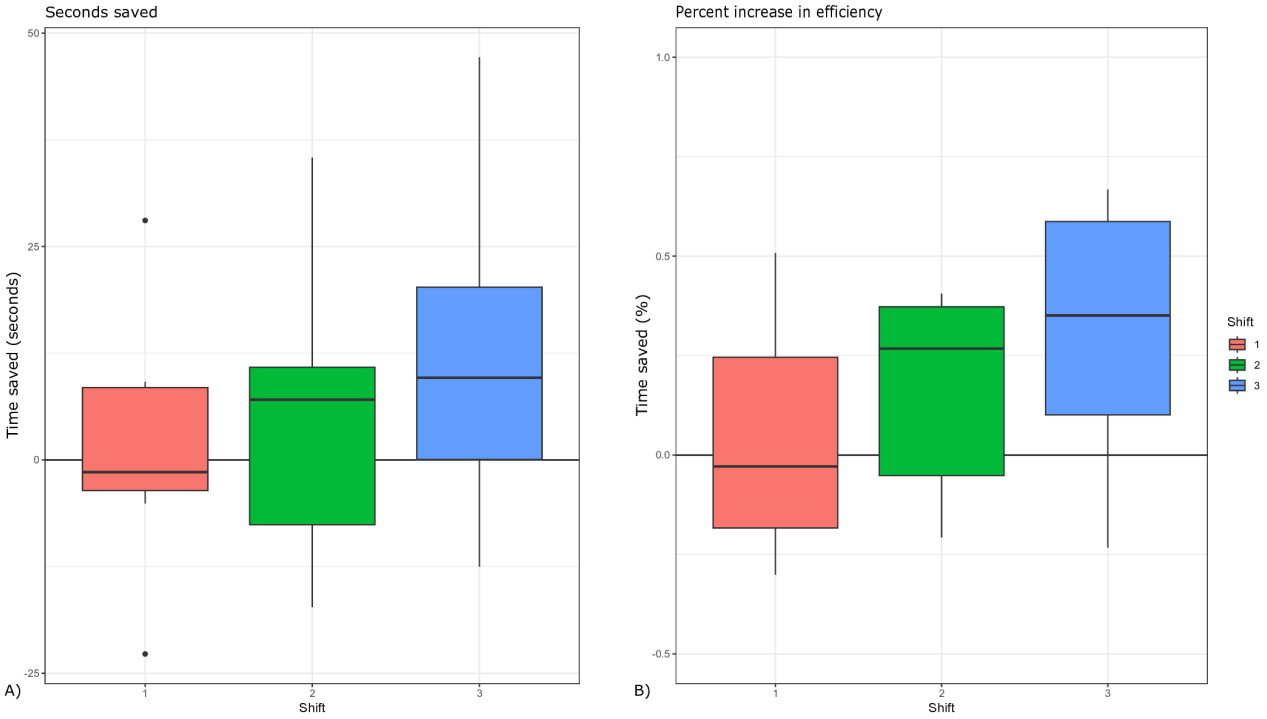
Simulation and Evaluation of Design

Once the prototypes were completed, they were tested for practical usage. After completing the simulation described above, the novel carts outperformed their older counterparts. On average, nurses were able to acquire needed supplies from the new cart nearly 6 seconds faster (21% faster) than they were able to get supplies from the old cart (2-tailed paired *t* test, time to acquire item from old cart=38.2 seconds; time to acquire item from new cart=32.4 seconds; degrees of freedom=23; P=.097). The lack of statistical significance was likely a limitation of low sample size; in addition, this simulation likely overestimated the time required to get samples from the newer cart because nurses were unfamiliar with the new configuration.

In addition, the time saved only increased as the carts went longer intervals without needing to be restocked. During the first shift, the new cart only saved about 1.5 seconds on average. During the second shift, the new cart saved an average of 4.7 seconds. On the third shift, the cart saved an average of 11.3 seconds (Figure 5). This time is a direct reflection of an improvement in capacity. While the old cart ran out of IV start kits by the third patient, forcing nurses to search for kits in nearby carts for every subsequent patient, the new cart contained all of the needed supplies for every patient. On average, a trip from a cart to the clean supply room (the preferred solution for nurses who did not trust that supply carts contain the proper equipment) took 1 minute; these trips were entirely eliminated by the use of the new prototype.

XSL•FO
RenderX

Figure 5. Time savings as a function of new cart. (A) The new cart saved an average of 5.8 seconds per simulated patient, and the time saved increased as the cart went longer without being restocked. (B) As the cart went longer without restocking, the percentage of time saved while accessing needed supplies increased, from about even when both were fully stocked, to 40% after 2 shifts. Boxplots show distribution of data (n=8 per shift). Sample size was too low to assess statistical significance.



Furthermore, the team noticed that the number of supplies taken from the new cart was much higher than those taken from the old cart. Specifically, there were several items that nurses were not willing to go to another cart for, such as speculums or certain test kits, but were happy to grab if they were present. This implied that having carts stocked with these supplies meant that nurses would actually use them, providing more timely patient care.

Plans for and Barriers to Full Implementation

Due to resource constraints, the new drawers are not fully implemented throughout the department. Full implementation would require a large up-front cost: due to the reorganization of supplies, some supplies were moved to different drawers. Because hospital policy requires that all carts contain the same organization, it is not possible to gradually phase in new carts over an extended period. In addition, because of this reorganization, it is not possible to replace the cart drawers for all carts, 1 drawer at a time, since incomplete adoption would mean that some items would no longer be included. This is certainly a limitation of the team's design process—had considerations about gradual or modular adoption been considered earlier in the process, this would not have been an issue. However, the team is working to find the lowest-cost way to implement these new designs.

They are also working to mitigate unintended consequences arising from adjustments to the new cart system. Specifically, to improve the ease of adjustment, they have created a pamphlet showing an updated schematic of the cart with pictures of the contents of each full cart. A written list of the cart's contents and approximate capacity will also be provided to keep at each nursing station for training purposes and reference. They will label carts with the contents of each drawer (ie, drawer 1 will be labeled with IV supplies; drawer 2 will be labeled with wound care, diagnostics, and extra needles; drawer 3 will be labeled with miscellaneous small supplies; drawer 4 will be labeled with respiratory supplies; and drawer 5 will be labeled with flushes, urine cups, catheters, tubing), and the bottom of each section of the drawer will have a photo of the supply that belongs inside it. These improvements will facilitate stocking, assist with the transition, and make it easier to onboard new nurses. In addition, it will decrease the likelihood of drift over time, ensuring that staff use the partitions as directed.

Discussion

Principal Findings

This participatory design process highlighted the importance of co-design for improving workflow and supply management in an ED. By observing, interviewing, and directly working with the people who use the system the most, the team was better able to understand the key challenges that influenced the effectiveness and constrained the solution space. The final prototype increased the capacity for key items by up to 200%, enabling cart capacity to last 36 hours (3 shifts) rather than requiring restocking within the first 12-hour shift. In addition, similar to other studies [15,31], the reduction in response time finding supplies did not reach statistical significance, but trended toward a statistical result. This is in spite of the fact that the new design meant participants spent more time looking for objects that had been moved, a metric that would improve as users became more familiar with the new system. Notably, the improvement here is likely due to the new design reducing the number of trips to the clean supply room or searching for

XSL•FO
RenderX

supplies from other carts, rather than from the more logical organization of the cart itself.

Overall, the findings of improved supply accessibility are in line with reports from other human factors–based redesigns of supply carts in other departments of the hospital [1,2,4,7,12-14]. While the constraints of this setting are not entirely analogous to other settings—here, the challenge was less about disorganization and more about a limited availability of stockers—they nevertheless suggest that cart organization, tailored to the needs of the staff and built with human factors principles in mind, plays a critical role in departmental workflows.

Even when the carts were fully stocked, nurses noted that the items that were present were there in the wrong quantities. Low-frequency items were present in large quantities, while there were hardly any of the most commonly used supplies. In fact, the supply manager told us that she had gone through the carts and had to toss hundreds of dollars' worth of unused supplies that had expired, an enormous waste of money and space for a department that had little to spare of either. This finding of overstocked items leading to wastage has been found in other sites as well [3,32] and suggests that periodic evaluation to ensure that carts can contain appropriate quantities of each supply is necessary. Our discovery of clusters of items frequently used together also matches the findings reported elsewhere [33].

In addition, the success of the quality improvement demonstrated the importance of participatory design, particularly the frequent consultation with stakeholders during the ideation and prototyping process, ensuring that the team's proposed solution was both effective and acceptable. Stakeholders were glad to be able to address a problem that mattered a great deal to them and helped to ensure that the new designs addressed their particular concerns about the old model. Co-design processes and design sprints are widely used in applied health research and are critical for ensuring that solutions focus on outcomes that are most important to end users [23,34,35].

Tailoring the cart's contents to the user workflow, and specifically minimizing the number of drawers that needed to be opened, improved user experiences. Although the cart's organization changed slightly, and users were less familiar with it, they still found it easier to use and preferred it over the standard cart. This is similar to prior findings that nurses preferred to use a cart that only required them to open 1 drawer per patient rather than a cart where everything was organized by tool but required multiple accesses [13].

In addition, nurses discussed how eager they were to be able to participate in the process. While this has been a persistent problem, engaging nurses in the redesign process gave them a voice, helping them feel like change was possible. This matches with previous research that found that nurse involvement in the co-design process improves nurse satisfaction [14].

The prototype created by this team is an example of how participatory design can create substantial improvement in a nursing workspace. The team was able to identify the key items nurses needed most and provide enough of those items to last multiple shifts without restocking, substantially improving workflow. The organization was designed according to nurses' priorities, focusing on ergonomics and efficiency using the same principles found in other studies [14,15]. While the solution could not solve the root causes of the issue—low funding and staffing shortages—it still made a meaningful, actionable contribution within the preexisting constraints by reducing the frequency of needed restocking. Once fully implemented, it is expected to save nurses a substantial amount of time per shift looking for items, reducing frustration and increasing the amount of time available for providing quality patient care. In addition, the durability of the acrylic, the work done to document and improve the ease of replicability of the system, and the buy-in from stakeholders all help to cement its durability and reduce the drift in restocking practices.

While the final design of the drawers in this project is customized to the specific hospital, the co-design process used demonstrates how stakeholder contribution and co-design can enable teams to make a meaningful impact in a low-resourced setting, working around structural barriers to improve nurse workflow and patient care.

Limitations

This project, while useful, was unable to address the root cause of the problem of supply shortages: the lack of a dedicated person who could simply restock all of the carts on a reliable basis. Despite this challenge, the implemented solution managed to ameliorate these effects by decreasing the time needed to restock and increasing the time the cart could go without restocking.

The project also encountered challenges regarding implementation—specifically, that because some items were moved to different drawers, and institutional policy required that all carts be organized the same way, implementation would need to occur all at once rather than being phased in gradually, requiring a larger up-front cost.

In addition, the data collection for the simulations was limited due to team availability and the availability of nursing staff to assist during their own busy shifts. The simulations also relied on convenience sampling: working with whatever staff were available at the time. Staff received the same prompts for both carts in the simulation; it is possible that thinking through the supplies beforehand made them faster at accessing the supplies the second time. We attempted to mitigate this response by randomizing which cart the team used first. Variable levels of experience with both the old and revised carts may also have led to bias in how quickly nurses found supplies in each cart [15]. Despite the lack of significance in time saved, the team was able to validate the use of the intervention through the increase in capacity for key supplies and through positive qualitative feedback from nursing staff indicating that, based on their past experiences, the cart would make a substantial improvement.

Finally, the designs for the cart interiors created during this project were tailor-made for this specific department and the cart system already in place, and would not necessarily generalize to other EDs. However, the processes used to design

XSL•FO
RenderX

and validate the solution are readily transferable, and the designs could be tweaked to fit the unique contexts of other departments.

Conclusion

The co-design process used in this study demonstrates the importance of stakeholder engagement and contribution to quality improvement and the use of the double diamond approach in a health care setting. Stakeholder input was crucial for navigating around institutional barriers and ensuring that any proposed solutions would be accepted and implemented, rather than ignored. Indeed, the buy-in generated from participatory design ensured that the changes made would be sustainable. While the team's designs were specific to the ED's context, the process and insights can be applied to any other type of workplace improvement process. Specifically, we recommend engaging with stakeholders at every step of the way, particularly during the rapid prototyping phase. However, we suggest that future teams consider the ease of implementation, especially gradual implementation for budgetary reasons, while evaluating their designs.

Acknowledgments

The content is solely the responsibility of the authors and does not necessarily represent the official views of any funding agency. This work could not have been done without the incredible staff at the emergency department, whose insight and perspective were crucial. We greatly appreciate their time and willingness to speak with us. In addition, we are grateful to Helen Zhang for her perspectives on design.

Funding

KH is supported generally by the National Institutes of Health (5T32NS091006-10). This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. No external organization had any role in the data for this study.

Data Availability

Full anonymized data may be made available upon request.

Authors' Contributions

Conceptualization: AM, AW, KH, NR, SD, TH
Data curation: AM, AW, KH, SD, TH
Investigation: AM, AW, KH, SD, TH
Methodology: AM, AW, KH, SD, TH
Project administration: JD, NR, ON
Resources: AM, SD
Review & editing: AM, AW, KH, TH
Supervision: JD, NR, ON
Writing – original draft: AM, AW, KH, TH
Writing – review & editing: JD, NR, ON, SD

Conflicts of Interest

None declared.

Multimedia Appendix 1
Timeline of steps for cart intervention project.
[PNG File, 58 KB - humanfactors_v13i1e80861_app1.png]

Multimedia Appendix 2
List of interviewees for primary stakeholder interviews.
[DOCX File, 15 KB - humanfactors_v13i1e80861_app2.docx]

Multimedia Appendix 3
High-use items identified by clustering algorithm and nurse report.
[DOCX File, 18 KB - humanfactors_v13i1e80861_app3.docx]

Multimedia Appendix 4
Comparison of the number of items in the current nursing carts to the modified amounts in the prototype (complete inventory).
[DOCX File, 26 KB - humanfactors_v13i1e80861_app4.docx]

XSL•FO
RenderX

References

1. McLaughlin RC. Hospital extra: redesigning the crash cart: usability testing improves one facility's medication drawers. Am J Nurs 2003 Apr;103(4):64A-64F. [doi: 10.1097/00000446-200304000-00037]
2. Lefebvre MS, Milloy S, Joynt C. Optimal crash cart configuration for a surgical NICU: utilizing human factors principles. Adv Neonatal Care 2021 Aug 1;21(4):289-396. [doi: 10.1097/ANC.0000000000000814] [Medline: 33278105]
3. Silva VFD, Lazzari DD, Reisdorfer N, Michaelsen SC, Kuhnen AE. Analyzing the operational conditions of crash carts in clinical and surgical hospitalization units. Rev Esc Enferm USP 2021;55:e03693. [doi: 10.1590/S1980-220X2019040003693] [Medline: 33886918]
4. Cajigas I, Silva MA, Chang H, et al. EVD supply cart decreases time to ventriculostomy placement in the emergency department. Neurosurgery 2020;67(Supplement_1). [doi: 10.1093/neuros/nyaa447_301]
5. Fahrenkopf MP, Eichhorn MG. Development of a plastic surgery supply cart: patient outcomes and quality of care. Plast Reconstr Surg Glob Open 2019 Feb;7(2):e2111. [doi: 10.1097/GOX.0000000000002111] [Medline: 30881838]
6. Johnson K, Smith C, Barker LG, et al. Implementation of a resident-designed procedure cart in a busy emergency department. Cureus 2023 Jul;15(7):e42211. [doi: 10.7759/cureus.42211] [Medline: 37602021]
7. Maul E, Latham B, Westgate PM. Saving time under pressure: effectiveness of standardizing pediatric resuscitation carts. Hosp Pediatr 2016 Feb;6(2):67-71. [doi: 10.1542/hpeds.2015-0161] [Medline: 26794403]
8. Shultz J, Davies JM, Caird J, Chisholm S, Ruggles K, Puls R. Standardizing anesthesia medication drawers using human factors and quality assurance methods. Can J Anaesth 2010 May;57(5):490-499. [doi: 10.1007/s12630-010-9274-8] [Medline: 20143279]
9. Tucker AL, Spear SJ. Operational failures and interruptions in hospital nursing. Health Serv Res 2006 Jun;41(3 Pt 1):643-662. [doi: 10.1111/j.1475-6773.2006.00502.x] [Medline: 16704505]
10. Storfjell JL, Ohlson S, Omoike O, Fitzpatrick T, Wetasin K. Non-value-added time: the million dollar nursing opportunity. J Nurs Adm 2009 Jan;39(1):38-45. [doi: 10.1097/NNA.0b013e31818e9cd4] [Medline: 19104286]
11. Citolino CMF, Santos ES, Silva RDCG, Nogueira LDS. Factors affecting the quality of cardiopulmonary resuscitation in inpatient units: perception of nurses. Rev Esc Enferm USP 2015 Dec;49(6):908-914. [doi: 10.1590/S0080-623420150000600005] [Medline: 27419673]
12. Rousek JB, Hallbeck MS. Improving medication management through the redesign of the hospital code cart medication drawer. Hum Factors 2011 Dec;53(6):626-636. [doi: 10.1177/0018720811426427] [Medline: 22235525]
13. Agarwal S, Swanson S, Murphy A, Yaeger K, Sharek P, Halamek LP. Comparing the utility of a standard pediatric resuscitation cart with a pediatric resuscitation cart based on the Broselow tape: a randomized, controlled, crossover trial involving simulated resuscitation scenarios. Pediatrics 2005 Sep;116(3):e326-e333. [doi: 10.1542/peds.2005-0320] [Medline: 16061568]
14. Crimlisk JT, Doherty MM, Fernandes E, Leblanc E, Guarino R, Costello KV. Adult code cart redesign: clinical implications. Nursing (Auckl) 2018 Jul;48(7):58-61. [doi: 10.1097/01.NURSE.0000532760.94860.2e] [Medline: 29923923]
15. Frazier M, Webster K, Dewan M, et al. Improving usability of the pediatric code cart by combining lean and human factors principles. Pediatr Qual Saf 2023;8(4):e676. [doi: 10.1097/pq9.0000000000000676] [Medline: 37551262]
16. Probst CA, Carter M, Cadigan C, et al. Utilizing a human factors nursing worksystem improvement framework to increase nurses' time at the bedside and enhance safety. J Nurs Adm 2017;47(2):94-100. [doi: 10.1097/NNA.0000000000000446] [Medline: 28067682]
17. Tucker AL, Singer SJ, Hayes JE, Falwell A. Front-line staff perspectives on opportunities for improving the safety and efficiency of hospital work systems. Health Serv Res 2008 Oct;43(5 Pt 2):1807-1829. [doi: 10.1111/j.1475-6773.2008.00868.x] [Medline: 18522667]
18. Ford S. Nurses waste "an hour a shift" finding equipment. Nurs Times 2009;105(5):1. [Medline: 19266954]
19. Rochais É, Atkinson S, Bussières JF. Nursing perception of the impact of medication carts on patient safety and ergonomics in a teaching health care center. J Pharm Pract 2013 Apr;26(2):131-137. [doi: 10.1177/0897190012451908] [Medline: 22797833]
20. Lindsay MR, Lytle K. Implementing best practices to redesign workflow and optimize nursing documentation in the electronic health record. Appl Clin Inform 2022 May;13(3):711-719. [doi: 10.1055/a-1868-6431] [Medline: 35668677]
21. Graham-Dickerson P, Houser J, Thomas E, et al. The value of staff nurse involvement in decision making. J Nurs Adm 2013 May;43(5):286-292. [doi: 10.1097/NNA.0b013e31828eec15] [Medline: 23615371]
22. The Field Guide to Human-Centered Design, 1st edition: IDEO.org; 2015. URL: https://www.ideo.com/journal/design-kit-the-human-centered-design-toolkit [accessed 2026-01-08]
23. Peters S, Guccione L, Francis J, et al. Evaluation of research co-design in health: a systematic overview of reviews and development of a framework. Implement Sci 2024 Sep 11;19(1):63. [doi: 10.1186/s13012-024-01394-4] [Medline: 39261956]
24. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0—Standards for Quality Improvement Reporting Excellence—revised publication guidelines from a detailed consensus process. J Am Coll Surg 2016 Mar;222(3):317-323. [doi: 10.1016/j.jamcollsurg.2015.07.456] [Medline: 26385723]
25. Jamshed S. Qualitative research method-interviewing and observation. J Basic Clin Pharm 2014 Sep;5(4):87-88. [doi: 10.4103/0976-0105.141942] [Medline: 25316987]

XSL•FO
RenderX

26. Grady MP. Qualitative and Action Research: A Practitioner Handbook: Phi Delta Kappa Educational Foundation; 1998.
27. Saunders B, Sim J, Kingstone T, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907. [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
28. Naeem M, Ozuem W, Howell K, Ranfagni S. A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. Int J Qual Methods 2023 Oct;22:WOS. [doi: 10.1177/16094069231205789]
29. Zenios SA, Makower J, Yock PG. Biodesign: The Process of Innovating Medical Technologies: Cambridge University Press; 2010. [doi: 10.1017/CBO9780511757853]
30. Onshape. URL: https://www.onshape.com/en/ [accessed 2026-01-17]
31. Chen DW, Chase VJ, Burkhardt ME, Agulto AZ. Using human factors design principles and industrial engineering methods to improve accuracy and speed of drug selection with medication trays. Jt Comm J Qual Patient Saf 2016;42(10):473-477. [doi: 10.1016/s1553-7250(16)42065-9] [Medline: 27712606]
32. Tsima BM, Rajeswaran L, Cox M. Assessment of cardiopulmonary resuscitation equipment in resuscitation trolleys in district hospitals in Botswana: a cross-sectional study. Afr J Prim Health Care Fam Med 2019 Oct 17;11(1):e1-e7. [doi: 10.4102/phcfm.v11i1.2029] [Medline: 31714118]
33. Hansoti B, Hahn E, Rao A, et al. Calibrating a chief complaint list for low resource settings: a methodologic case study. Int J Emerg Med 2021 May 19;14(1):32. [doi: 10.1186/s12245-021-00347-8] [Medline: 34011284]
34. Wu L, Zhu C, Seguin JP, et al. Designing culturally adapted digital mental health support tool for Chinese-speaking international students in Australia: a qualitative co-design study. JMIR Form Res 2025 Oct 21;9:e76695. [doi: 10.2196/76695] [Medline: 41186423]
35. Hopstaken JS, Koeneman M, Hooijer R, et al. How to improve pancreatic cancer network care using a human-centered design sprint. J Med Internet Res 2025 Sep 5;27:e55598. [doi: 10.2196/55598] [Medline: 40911870]

Abbreviations

ED: emergency department
IV: intravenous
SQUIRE: Standards for Quality Improvement Reporting Excellence

Edited by E Borycki; submitted 07.Aug.2025; peer-reviewed by EP Larsen, K Webster; revised version received 10.Nov.2025; accepted 13.Nov.2025; published 28.Jan.2026.

Please cite as:
Hefter K, Dreibelbis S, Haupt T, McIver A, Wang A, Dwight J, Nwodim O, Ray N
Redesign of Bedside Supply Carts to Improve Emergency Department Workflows: Mixed Methods Participatory Design
JMIR Hum Factors 2026;13:e80861
URL: https://humanfactors.jmir.org/2026/1/e80861
doi:10.2196/80861

© Kat Hefter, Sammy Dreibelbis, Theresa Haupt, Amelia McIver, Agnes Wang, Julia Dwight, Ogechi Nwodim, Neil Ray. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 28.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Development of the ERATbi App, a Clinical Decision Support System for Early Recovery After Traumatic Brain Injury in the ICU: Usability Study

Hsiao-Ching Yen[1*], PT, MS; I-Hui Wu[2,3,4*], MD, PhD; Wei-Ling Hsiao[5], RN, MS; Sheng-Ru Lai[6], MS; Chen-Hao Yang[7], MS; Hsien-Chi Liao[2], MD, MS; Yin-Yi Han[2,8], MD, PhD

[1]Division of Physical Therapy, Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, Taipei, Taiwan
[2]Department of Traumatology, National Taiwan University Hospital, No. 7, Chung Shan South Road, Taipei, Taiwan
[3]Department of Cardiovascular Surgery, National Taiwan University Hospital, Taipei, Taiwan
[4]Graduate Institute of Clinical Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan
[5]Department of Nursing, National Taiwan University Hospital, Taipei, Taiwan
[6]Department of Dietetics, National Taiwan University Hospital, Taipei, Taiwan
[7]Division of Respiratory Therapy, Department of Integrated Diagnostics & Therapeutics, National Taiwan University Hospital, Taipei, Taiwan
[8]Department of Anesthesiology, National Taiwan University Hospital, Taipei, Taiwan
[*]these authors contributed equally

Corresponding Author:
Yin-Yi Han, MD, PhD
Department of Traumatology, National Taiwan University Hospital, No. 7, Chung Shan South Road, Taipei, Taiwan

Abstract

Background: Early rehabilitation in neurocritical care is often underutilized due to fragmented workflows, interdisciplinary coordination challenges, and the absence of structured digital decision support. Traditional clinical decision support systems (CDSS) often address single domains and lack adaptability to the dynamic, multiprofessional workflows of intensive care units (ICUs).

Objective: To develop and evaluate the usability of the ERATbi App (Early Recovery After Traumatic Brain Injury App), a modular, tablet-based CDSS was designed to streamline early rehabilitation planning and strengthen interdisciplinary coordination for patients with moderate-to-severe traumatic brain injury (TBI) in intensive care settings.

Methods: The ERATbi app integrates four functional modules—delirium risk management, precision nutrition, stepwise early mobilization, and respiratory care for rib fractures—into a unified interface. A simulation-based usability study was conducted with 18 ICU clinicians. Evaluation metrics included System Usability Scale (SUS) scores, task completion rates, error rates, and task durations. Additional user feedback was collected via a 5-point Likert satisfaction survey and semi-structured qualitative interviews.

Results: The app demonstrated high usability (mean SUS score 83.6, SD 7.4), a 100% (18/18 participants) task completion rate, and a low error rate (4.2%). Average module completion time was 6.5 minutes, and user satisfaction was high (mean 4.7, SD 0.5). Users highlighted the value of the app's visual logic, real-time alerts, adaptive thresholds, and modular workflow integration for enhancing team coordination and decision consistency.

Conclusions: The ERATbi app demonstrated excellent usability, high user satisfaction, and clinical relevance in simulated ICU workflows. Its logic-driven, workflow- integrated design may support scalable, interdisciplinary implementation of early rehabilitation in neurocritical care settings.

(JMIR Hum Factors 2026;13:e79981) doi:10.2196/79981

KEYWORDS

brain injuries, traumatic; decision support systems; human engineering; intensive care units; rehabilitation, early; usability testing

XSL•FO
RenderX

Introduction

Epidemiology and Clinical Burden of Traumatic Brain Injury

Traumatic brain injury (TBI) is a major global public health concern and a leading cause of death and long-term disability worldwide [1]. In Taiwan, the annual incidence of hospitalized head injuries is approximately 126.1 per 100,000 population, equivalent to nearly 29,000 cases each year. Among these, an estimated 20.5% involve moderate-to-severe TBI, which is frequently associated with impaired consciousness, mechanical ventilation, and polytrauma [2].

Barriers to Early Rehabilitation in the Acute Phase

Although the benefits of early rehabilitation are increasingly supported by evidence [3,4], its implementation in the acute phase remains limited. Major barriers include safety concerns, fragmented workflows, and the absence of structured clinical decision-support systems [5-8]. Clinical hesitation often stems from hemodynamic instability, fluctuations in intracranial pressure (ICP), and the complexity of managing multiple invasive devices. These challenges force clinicians to balance therapeutic benefits against physiological risks, contributing to variability in decision-making and delayed initiation of rehabilitation.

Gaps in Multidisciplinary Integration and Decision Support

Early mobilization, nutritional therapy, delirium prevention, and respiratory management show positive impacts on outcomes of critically ill patients [9-15]. However, these interventions are typically delivered by separate disciplines with limited cross-team coordination, resulting in inconsistent implementation and delayed functional recovery. Most existing electronic medical record (EMR) systems do not provide process-based visualization or real-time interdisciplinary decision support. As a result, clinicians often depend on paper-based tools, fragmented documentation, and manual communication, further hindering timely and coordinated early rehabilitation.

Study Rationale and Objectives

To address these challenges, we developed the Early Recovery After Traumatic Brain Injury (ERATbi) app, a modular, web-based clinical decision support system (CDSS) designed to assist ICU clinicians in delivering structured early rehabilitation for patients with moderate-to-severe TBI. The system integrates four evidence-based modules: (1) delirium risk management, (2) precision nutrition therapy, (3) stepwise early mobilization, and (4) respiratory care for rib fractures. It incorporates standardized safety thresholds (eg, mean arterial pressure ≥65 mm Hg, ICP <20 mm Hg), automated alerts, and decision checkpoints to support safe and consistent clinical decision-making.

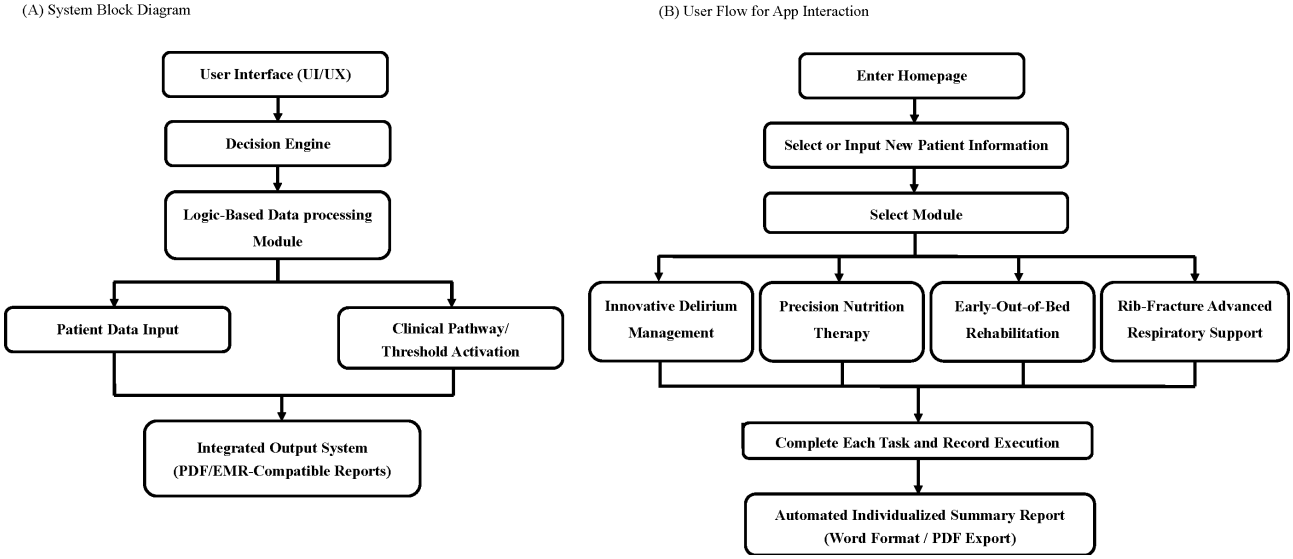
Grounded in user-centered design and interdisciplinary collaboration, this study describes the ERATbi app's theoretical framework, system architecture, and simulated clinical workflows, and evaluates its usability, clinical relevance, and potential to enhance safety, standardization, and interdisciplinary coordination in early neurocritical rehabilitation.

Methods

System Design and Development Approach

The development of the ERATbi app followed a user-centered design methodology, which is widely recommended for CDSS development in high-acuity settings due to its emphasis on workflow integration, context alignment, and stakeholder engagement [16-18] (Figure 1). A multidisciplinary co-design team—comprising intensivists, physical therapists, dietitians, critical care nurses, and software engineers—collaboratively identified workflow gaps and defined the functional requirements for a neuro-ICU–specific decision-support system. The development process consisted of three iterative stages:

Figure 1. System architecture and user workflow diagram of the ERATbi (Early Recovery After Traumatic Brain Injury) app.



XSL•FO
RenderX

Needs Assessment Phase

Direct observations and stakeholder interviews were conducted to map existing ICU rehabilitation workflows for patients with moderate-to-severe TBI and to identify barriers to timely intervention. Key gaps included inconsistent interdisciplinary coordination and the absence of real-time patient stability assessment tools.

Prototype Development Phase

Modular logic trees and interactive wireframes were constructed based on key clinical variables—such as Glasgow Coma Scale (GCS) score, time since admission, vital signs, and device constraints (eg, chest tubes, external ventricular drains). These algorithms were designed to emulate expert reasoning while supporting real-time, safety-focused decision-making.

Simulated Use Phase

Preliminary usability testing was conducted using mock clinical scenarios, enabling iterative refinement of the interface, navigation structure, and backend decision logic in response to user feedback [17].

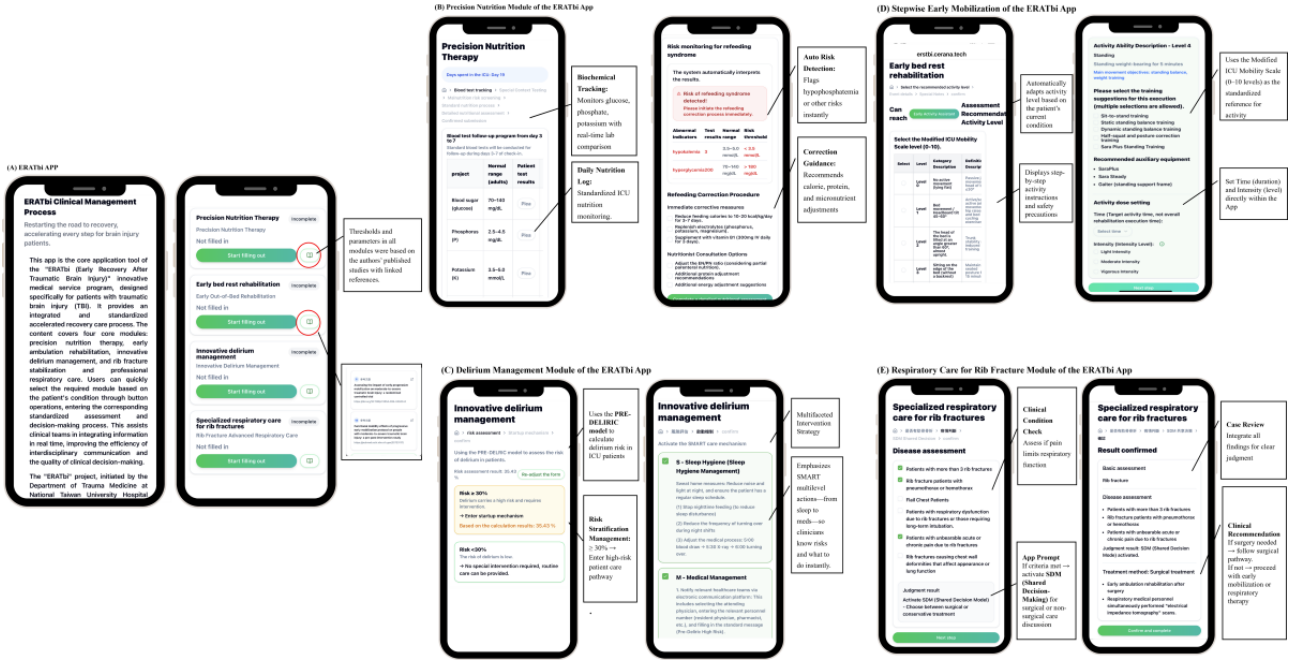
Ethical Considerations

This study did not involve human participants or the collection of personally identifiable information. Usability evaluations were conducted exclusively in a simulated environment using standardized case scenarios developed from deidentified patient data. These data were originally obtained from an institutional review board-approved study at National Taiwan University Hospital (IRB No. 202306107RIND). The original informed consent and IRB approval covered secondary analyses of the existing data without requiring additional consent. All data used in this secondary analysis were deidentified prior to analysis, and no personally identifiable information was accessible to the study team. All participants were ICU clinicians (physicians, nurses, therapists, and dietitians) who evaluated only the app's interface and workflow. No patient contact occurred, and no clinical interventions were performed. In accordance with institutional policy, the study was deemed exempt from additional IRB review, as it did not constitute human subjects research.

Modular Architecture and Embedded Clinical Logic

The ERATbi app consists of four functional modules, each corresponding to a core domain of early neurocritical rehabilitation (Figure 2). This modular structure supports process-based clinical reasoning and enhances interdisciplinary alignment.

Figure 2. Interface storyboard of the ERATbi (Early Recovery After Traumatic Brain Injury) app demonstrating the stepwise workflow across four functional modules: precision nutrition, delirium management, stepwise early mobilization, and respiratory care.



Precision Nutrition Module

This module identifies ICU patients at risk of nutritional deficiency within the first 7 days of admission using real-time GCS and BMI calculations. Consistent with ESPEN (European Society for Clinical Nutrition and Metabolism) and ASPEN (American Society for Parenteral and Enteral Nutrition) guidelines, it stratifies patients into intervention tiers and provides individualized caloric, protein, and micronutrient recommendations [19-21]. Embedded calculators further ensure accurate risk estimation and streamline nutritional decision-making [10,22,23].

Delirium Management Module

This module incorporates the "Sweet SMART Home" protocol [24] to guide delirium risk stratification based on patient age, cognitive function, and sedation level. It emphasizes nonpharmacologic strategies such as circadian regulation and reorientation, consistent with contemporary ICU guidelines for delirium prevention [11,25].

XSL•FO
RenderX

Stepwise Early Mobilization Module

Algorithmic logic trees stratify patients according to consciousness level (GCS), physiological thresholds, and device attachments. Based on these criteria, the system generates graded activity recommendations and automated alerts for contraindicated conditions. The framework draws on established ICU mobilization models emphasizing phased recovery and functional readiness [3,26,27].

Respiratory Care for Rib Fracture Module

This module addresses comorbid thoracic trauma by supporting shared decision-making for surgical fixation and integrating electrical impedance tomography data to guide respiratory management. It includes structured prompts for pulmonary hygiene, analgesia assessment, and individualized weaning plans, aligning with ICU pain and ventilation protocols [28,29].

User Interface and Decision-Support Design

The ERATbi interface uses a process-based visual workflow that enables clinicians to navigate each module in a structured, stepwise manner. The dashboard displays real-time patient status across the four domains. Core features include auto-calculated clinical thresholds and alert mechanisms triggered by physiological or procedural contraindications.

Although full EMR integration is not yet available, the system generates structured visual summaries and consolidated recommendations to support clinical review. Data-entry fields incorporate real-time validation—such as missing-value alerts and format checks—particularly within the nutrition and mobilization modules. Guided prompts and sequential navigation are designed to reduce cognitive load and enhance data completeness during clinical use.

Simulation-Based Usability Evaluation

A simulation-based usability evaluation was conducted in the trauma ICU of a tertiary medical center. Eighteen clinicians participated, including physical therapists (n=8), ICU nurses (n=5), attending physicians (n=2), respiratory therapists (n=2), and one dietitian (n=1), all with previous experience in neurocritical care and early mobilization.

Each participant completed two standardized case vignettes reconstructed from de-identified patient records, representing moderate and severe TBI. Each vignette contained five consecutive ICU days of physiological parameters (eg, GCS, MAP [mean arterial pressure], FiO_2 [fraction of inspired oxygen], and ICP), nutritional data, and mobility progression to simulate continuous monitoring. Participants initiated a "new patient entry" and sequentially updated data to reflect daily changes, mirroring real-world decision-making workflows. Although each vignette contained data for multiple ICU days, task-level analysis was based on discrete user interactions, which were counted independently of the number of simulated days.

Case vignettes were tailored to professional scope: physical therapists emphasized activity planning, nurses focused on safety and delirium monitoring, and physicians emphasized interdisciplinary coordination. Participants used hospital-issued tablets or smartphones to complete workflows for at least two modules. All interactions were recorded via screen-capture software for subsequent analysis.

Assessment Approach

Usability was assessed using a mixed-methods approach integrating quantitative and qualitative data. Quantitative measures included the System Usability Scale (SUS) [30], task completion rate, error rate (eg, incorrect pathway selection, data-entry mistakes), and average task duration. After each simulation, participants completed a customized satisfaction questionnaire with a 5-point Likert scale (Multimedia Appendix 1) and participated in a 10 - 15-minute semi-structured interview (Multimedia Appendix 2).

Interviews were conducted face-to-face by a trained observer, audio-recorded with participant consent, and designed to explore usability facilitators, barriers, and improvement opportunities. Example prompts included the following questions. "Which parts of the interface were most intuitive or confusing?" "Were there any steps that slowed your workflow?" "What features would you modify or add to enhance clinical applicability?"

Interview recordings were transcribed verbatim and analyzed thematically following Braun and Clarke's inductive approach. Two independent researchers coded themes related to facilitators, barriers, and design recommendations. Quantitative and qualitative findings were integrated to develop a comprehensive understanding of user experience and system performance. Sessions were monitored using validated ICU CDSS evaluation frameworks [16,17].

Methodological Rigor

To further enhance the rigor of this mixed-methods design, several strategies were implemented. For the quantitative component, task-completion metrics and SUS scores were cross-validated by two reviewers to ensure data accuracy. For the qualitative component, two researchers coded the open-ended responses using an inductive thematic approach. Coding discrepancies were discussed and resolved through consensus, and an audit trail was maintained to document analytic decisions.

Data triangulation was achieved by comparing usage metrics, task-performance patterns, and qualitative feedback. Integration of quantitative and qualitative findings followed a convergence model, whereby qualitative themes were compared with performance trends to enhance interpretive depth. Collectively, these approaches strengthened the credibility, dependability, and confirmability of the study.

Results

System Architecture Overview

Eighteen ICU professionals participated in the simulation-based usability evaluation using standardized case vignettes. All participants independently completed workflows in at least two ERATbi modules, with their interactions fully recorded for analysis. The ERATbi App integrates four core modules, Delirium Management, Precision Nutrition, Stepwise Early Mobilization, and Respiratory Care, into a unified, logic-driven clinical decision support system tailored for neurocritical

XSL•FO
RenderX

rehabilitation. Each module automatically adapts recommendations based on patient-specific variables such as vital signs, level of consciousness (GCS), and device conditions. Safety features include real-time alerting, visual cue signaling deviations from physiological thresholds, and auto-generated clinical summaries. Together, these mechanisms support interdisciplinary coordination and help clinicians maintain a consistent understanding of patient status and safety priorities.

Usability Metrics

The ERATbi App demonstrated strong usability across all quantitative indicators. The System Usability Scale (SUS) yielded a mean score of 83.6 (SD 7.4), corresponding to the "excellent" usability range [31]. All participants successfully completed their assigned workflows, resulting in a 100% (18/18) task completion rate. The mean time to complete a single module was 6.5 minutes (SD 1.3). The overall error rate was 4.2% (2 errors out of 54 total interactions), primarily involving minor data-entry or selection issues.

User satisfaction, measured via a 5-point Likert scale, averaged 4.7 (SD 0.5), indicating high perceived usefulness and acceptance. Participants consistently highlighted the interface's logical visual organization, which facilitated intuitive navigation and supported clinical reasoning, particularly beneficial for junior staff. Features such as real-time alerts, automated summaries, and visualized physiological thresholds were frequently cited as improving confidence and enhancing interdisciplinary discussions.

Qualitative Feedback

Open-ended feedback and observer field notes revealed three major themes: (1) Clarity and cognitive alignment: Participants reported that the stepwise logic structure closely aligned with their real-world ICU decision-making processes, reducing cognitive burden and supporting rapid clinical reasoning; (2) Enhanced interdisciplinary coordination: The integration of rehabilitation, nutrition, delirium, and respiratory considerations within a single interface promoted a shared mental model, enabling more coherent team discussions and planning. In this study, "shared mental model" refers to the alignment of understanding among ICU physicians, nurses, dietitians, and physical therapists regarding safety priorities, patient status, and rehabilitation goals, allowing teams to interpret clinical information consistently and coordinate care more efficiently; (3) Implementation potential: Users expressed confidence in the system's clinical applicability and highlighted the value of future EMR integration to support handoffs, documentation consistency, and safety review workflows.

Overall, the ERATbi App demonstrated strong feasibility, high user acceptance, and operational efficiency in simulated ICU environments, supporting its potential for broader deployment in neurocritical care workflows.

Discussion

Principal Findings

This study developed and evaluated the ERATbi App, a modular CDSS designed to support early rehabilitation planning for ICU patients with moderate-to-severe TBI. In simulation testing with 18 multidisciplinary ICU clinicians, the system demonstrated excellent usability (mean SUS 83.6, SD 7.4), complete task success (100%), a low error rate (4.2%, 2/54 tasks), and high user satisfaction (mean 4.7, SD 0.5). Participants reported that the app's structured, modular interface enhanced clarity, reduced cognitive burden, and facilitated interdisciplinary communication and real-time decision-making. Collectively, these findings indicate that ERATbi is highly feasible, operationally efficient, and well-positioned for clinical implementation in neurocritical care environments.

Overcoming the Fragmentation of Early ICU Rehabilitation Support

Prior CDSS research in ICU settings has predominantly focused on single-domain applications, including delirium screening [32,33], primary palliative care [34], rare disease diagnosis [35], fall prevention [36], or early mobility-related decision-making [37,38]. Although these systems have demonstrated benefits within their respective domains, most operate as stand-alone tools lacking interoperability, limiting their ability to support the complex, interdisciplinary workflows required in neurocritical care.

This fragmentation is compounded by EMR-integrated CDSS platforms that largely emphasize documentation or retrospective data retrieval rather than offering adaptive, real-time clinical guidance. These systems rarely adjust recommendations in response to rapidly changing physiological parameters or evolving clinical contexts, reducing their effectiveness in high-acuity decision-making.

To address these persistent gaps, the ERATbi app was conceptualized as a workflow-embedded, multi-domain CDSS specifically tailored for neurocritical care. Its design reflects the multifactorial demands of early rehabilitation in moderate-to-severe TBI—including fluctuating consciousness, invasive device management, physiological instability, and the interactions of multiple disciplines.

Evidence supporting this approach is aligned with recent studies. In 2024, Dunn et al [37] emphasized the need for standardized terminology and safety thresholds in ICU mobility tools, demonstrating high expert agreement (Content Validity Index=0.93). Similarly, in 2025, Wilson-Jene et al [38] demonstrated that algorithm-based guidance improved adherence to safety protocols and reduced mobilization delays. These studies reinforce the necessity of structured, logic-driven, and adaptive CDSS frameworks such as ERATbi.

Modular Integration and Workflow-Embedded Design in the ERATbi App

The ERATbi app was intentionally engineered as a multidomain CDSS to meet the interdisciplinary and dynamically evolving needs of neurocritical care. Unlike traditional single-domain tools, ERATbi integrates four pillars of early rehabilitation—delirium management, precision nutrition, stepwise mobilization, and respiratory care—within a unified modular interface. This architecture enables clinicians to navigate shifting consciousness levels, invasive device

XSL•FO
RenderX

constraints, physiological instability, and interdisciplinary workflows with improved clarity and consistency.

Aligned with evidence-based guidelines and developed using a user-centered design methodology, ERATbi reflects contemporary recommendations that prioritize usability, workflow integration, and team-wide accessibility [39,40]. At the system's core is a decision engine that synthesizes validated physiological thresholds and individualized risk factors to generate real-time, context-aware recommendations—marking a shift from retrospective documentation tools to proactive, adaptive CDSS design.

Visual logic pathways, real-time alerts, and dynamic thresholds were incorporated to reduce cognitive load while preserving decision accuracy. Prior studies demonstrate that such features improve protocol adherence [41], reduce ICU errors [42], and enhance decision consistency during high-acuity transitions [43]. Additionally, intuitive interfaces with clearly presented thresholds have been shown to strengthen user trust, situational awareness, and adoption across multidisciplinary teams [44-46].

By embedding these technical and human-centered principles within a modular architecture, the ERATbi App provides a scalable and interoperable platform capable of standardizing early rehabilitation workflows and improving safety and communication in complex neurocritical care settings.

Clinical Implications

This study highlights several important clinical implications. First, the structured, logic-driven design of ERATbi supports consistent clinical decision-making and reduces variability across shifts and providers. Second, its visual dashboards and auto-generated summaries enhance interdisciplinary communication by providing a shared reference for physicians, nurses, dietitians, and physical therapists. Third, the app demonstrates utility as both an educational and an implementation tool, particularly for onboarding new personnel, reinforcing protocol adherence, and supporting quality improvement initiatives. Participants also noted that the modular framework could be adapted to other neurocritical populations, including stroke or post-neurosurgical patients, suggesting the system's broader clinical scalability.

Scalability and Data Integration Potential

The ERATbi app was designed for future interoperability with the hospital's HIS through HL7-FHIR–based data exchange. Structured data entered into the app (eg, GCS, MAP, ICP, FiO_2 , mobility level, and nutritional parameters) is planned to be stored using standardized FHIR Observation and Encounter resources. Physiological and biochemical data, such as vital signs, blood gas analyses, and laboratory results, are planned for future automated retrieval from the HIS within the preceding 24 - 48 hours to reduce manual input.

Upon full integration, the system is planned to generate AI-assisted summary reports that consolidate the four rehabilitation modules into a single interface. These structured reports are intended to be documented in the EMR and made accessible during interdisciplinary rounds. This planned automated data exchange is expected to improve accuracy, reduce clinician workload, enhance scalability across ICUs, and enable predictive analytics by linking rehabilitation trajectories with clinical outcomes such as mobilization timing, extubation readiness, and ICU length of stay.

Limitations

This study has several limitations. First, the evaluation was conducted in a simulation-based environment rather than during real-time clinical deployment, which may affect generalizability. Although case scenarios reflected real moderate-to-severe TBI encounters, certain dynamic or emergent ICU conditions may not have been fully captured. Second, the ERATbi app was assessed as a stand-alone prototype without EMR integration, which may limit immediate scalability and clinical adoption.

Future Work

Beyond its clinical decision support function, the ERATbi system will be expanded into an educational and quality improvement platform. A planned B2C extension will provide simplified dashboards and interactive modules for patient- and family-centered education to support health literacy and post-ICU self-management. A real-world pilot study is planned at the neuro-ICU of National Taiwan University Hospital, followed by multicenter evaluations examining outcomes such as mobilization timing, delirium incidence, ICU length of stay, and safety events. Future iterations will incorporate a continuous, data-driven feedback mechanism linking clinician decisions with patient outcomes to refine decision thresholds. This iterative framework will enable adaptive modeling and sustained quality improvement in neurocritical rehabilitation.

Conclusion

The ERATbi app is a modular clinical decision support system designed to address the complex demands of early rehabilitation in neurocritical care. Simulation-based usability testing demonstrated high user acceptance, excellent usability, and strong potential to support workflow standardization, enhance interdisciplinary coordination, and promote patient safety. By integrating evidence-based clinical logic with user-centered design, the system helps to close critical gaps in ICU rehabilitation for patients with moderate-to-severe traumatic brain injury. With planned EMR integration and real-world implementation, the ERATbi app offers a scalable foundation for disseminating early rehabilitation strategies across broader neurocritical care populations.

Acknowledgments

The authors thank (1) the Department of Traumatology, National Taiwan University Hospital; (2) the National Taiwan University, College of Medicine; and (3) the Taiwanese Society of Phlebology for covering the article processing charge.

XSL•FO
RenderX

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request and with appropriate institutional approvals, where applicable.

Authors' Contributions

Hsiao-Ching Yen and I-Hui Wu designed the study and contributed equally to this work. Hsiao-Ching Yen led the system development, study conception, and drafted the manuscript. I-Hui Wu contributed to the development of the clinical framework and critically revised the manuscript for important intellectual content.
Wei-Ling Hsiao, Sheng-Ru Lai, and Chen-Hao Yang contributed to module design, usability evaluation, and data interpretation. Hsien-Chi Liao provided clinical oversight and interdisciplinary coordination.
Yin-Yi Han supervised the overall project and finalized the manuscript. All authors reviewed and approved the final manuscript and are accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
ERATbi (Early Recovery After Traumatic Brain Injury) app Usability Evaluation Questionnaire.
[DOCX File, 15 KB - humanfactors_v13i1e79981_app1.docx]

Multimedia Appendix 2
Qualitative feedback coding framework.
[DOCX File, 14 KB - humanfactors_v13i1e79981_app2.docx]

References

1. Masel BE, DeWitt DS. Traumatic brain injury: a disease process, not an event. J Neurotrauma 2010 Aug;27(8):1529-1540. [doi: 10.1089/neu.2010.1358] [Medline: 20504161]
2. Hsu IL, Li CY, Chu DC, Chien LC. An epidemiological analysis of head injuries in Taiwan. Int J Environ Res Public Health 2018 Nov 4;15(11):2457. [doi: 10.3390/ijerph15112457] [Medline: 30400380]
3. Yen HC, Han YY, Hsiao WL, et al. Functional mobility effects of progressive early mobilization protocol on people with moderate-to-severe traumatic brain injury: a pre-post intervention study. NeuroRehabilitation 2022;51(2):303-313. [doi: 10.3233/NRE-220023] [Medline: 35723117]
4. Kumar MA, Romero FG, Dharaneeswaran K. Early mobilization in neurocritical care patients. Curr Opin Crit Care 2020 Apr;26(2):147-154. [doi: 10.1097/MCC.0000000000000709] [Medline: 32068582]
5. Dirkes SM, Kozlowski C. Early mobility in the intensive care unit: evidence, barriers, and future directions. Crit Care Nurse 2019 Jun;39(3):33-42. [doi: 10.4037/ccn2019654] [Medline: 31154329]
6. Dafoe S, Chapman MJ, Edwards S, Stiller K. Overcoming barriers to the mobilisation of patients in an intensive care unit. Anaesth Intensive Care 2015 Nov;43(6):719-727. [doi: 10.1177/0310057X1504300609] [Medline: 26603796]
7. Yao Y, Dunn Lopez K, Bjarnadottir RI, et al. Examining care planning efficiency and clinical decision support adoption in a system tailoring to nurses' graph literacy: national, web-based randomized controlled trial. J Med Internet Res 2023 Aug 11;25:e45043. [doi: 10.2196/45043] [Medline: 37566456]
8. Krupp A, Steege L, Lee J, Lopez KD, King B. Supporting decision-making about patient mobility in the intensive care unit nurse work environment: work domain analysis. JMIR Nurs 2022 Sep 27;5(1):e41051. [doi: 10.2196/41051] [Medline: 36166282]
9. Adler J, Malone D. Early mobilization in the intensive care unit: a systematic review. Cardiopulm Phys Ther J 2012 Mar;23(1):5-13. [Medline: 22807649]
10. Singer P, Blaser AR, Berger MM, et al. ESPEN guideline on clinical nutrition in the intensive care unit. Clin Nutr 2019 Feb;38(1):48-79. [doi: 10.1016/j.clnu.2018.08.037] [Medline: 30348463]
11. Devlin JW, Skrobik Y, Gélinas C, et al. Clinical Practice Guidelines for the Prevention and Management of Pain, Agitation/Sedation, Delirium, Immobility, and Sleep Disruption in Adult Patients in the ICU. Crit Care Med 2018 Sep;46(9):e825-e873. [doi: 10.1097/CCM.0000000000003299] [Medline: 30113379]
12. Lewis K, Balas MC, Stollings JL, et al. A focused update to the Clinical Practice Guidelines for the Prevention and Management of Pain, Anxiety, Agitation/Sedation, Delirium, Immobility, and Sleep Disruption in adult patients in the ICU. Crit Care Med 2025 Mar 1;53(3):e711-e727. [doi: 10.1097/CCM.0000000000006574] [Medline: 39982143]

XSL•FO
RenderX

13. Prins JTH, Van Lieshout EMM, Ali-Osman F, et al. Surgical stabilization versus nonoperative treatment for flail and non-flail rib fracture patterns in patients with traumatic brain injury. Eur J Trauma Emerg Surg 2022 Aug;48(4):3327-3338. [doi: 10.1007/s00068-022-01906-1] [Medline: 35192003]
14. Franssen AJPM, Daemen JHT, Luyten JA, et al. Treatment of traumatic rib fractures: an overview of current evidence and future perspectives. J Thorac Dis 2024 Aug 31;16(8):5399-5408. [doi: 10.21037/jtd-23-1832] [Medline: 39268098]
15. Kao CC, Chen KC, Chiang XH, et al. Clinical outcomes of rib fracture stabilization and conservative treatment in a high-volume Asian trauma center: a propensity score-matched retrospective study. World J Emerg Surg 2025 May 19;20(1):40. [doi: 10.1186/s13017-025-00620-8] [Medline: 40390007]
16. Downie AS, Hancock M, Abdel Shaheed C, et al. An electronic clinical decision support system for the management of low back pain in community pharmacy: development and mixed methods feasibility study. JMIR Med Inform 2020 May 11;8(5):e17203. [doi: 10.2196/17203] [Medline: 32390593]
17. Ratwani RM, Savage E, Will A, et al. A usability and safety analysis of electronic health records: a multi-center study. J Am Med Inform Assoc 2018 Sep 1;25(9):1197-1201. [doi: 10.1093/jamia/ocy088]
18. Azadi A, García-Peñalvo FJ. Optimizing clinical decision support system functionality by leveraging specific human-computer interaction elements: insights from a systematic review. JMIR Hum Factors 2025 May 6;12:e69333. [doi: 10.2196/69333] [Medline: 40327851]
19. Díaz Chavarro BC, Romero-Saldaña M, Reveiz J, Molina-Recio G, et al. Impact of nutritional screening on mortality and intensive care unit length of stay. Front Nutr 2019;25(2):132-149. [doi: 10.3389/fnut.2025.1474039] [Medline: 31068745]
20. Singer P, Blaser AR, Berger MM, et al. ESPEN practical and partially revised guideline: clinical nutrition in the intensive care unit. Clin Nutr 2023 Sep;42(9):1671-1689. [doi: 10.1016/j.clnu.2023.07.011] [Medline: 37517372]
21. Compher C, Bingham AL, McCall M, et al. Guidelines for the provision of nutrition support therapy in the adult critically ill patient: The American Society for Parenteral and Enteral Nutrition. J Parenter Enteral Nutr 2022 Jan;46(1):12-41 [FREE Full text] [doi: 10.1002/jpen.2267]
22. Lee HY, Oh BM. Nutrition management in patients with traumatic brain injury: a narrative review. Brain Neurorehabil 2022 Mar;15(1):e4. [doi: 10.12786/bn.2022.15.e4] [Medline: 36743843]
23. Li Y, Liu C, Luo X, et al. Controlling nutritional status score and prognostic nutrition index predict the outcome after severe traumatic brain injury. Nutr Neurosci 2022 Apr;25(4):690-697. [doi: 10.1080/1028415X.2020.1804097] [Medline: 32778001]
24. Chou YH, Hsiao WL, Chen CJ, et al. Association between prosthesis contour and peri-implantitis in patients compliant with supportive periodontal therapy: a retrospective cohort study. J Prosthodont Res 2025 Jan 10;69(1):68-75. [doi: 10.2186/jpr.JPR_D_23_00236] [Medline: 39198202]
25. Pun BT, Balas MC, Barnes-Daly MA, et al. Caring for critically ill patients with the ABCDEF Bundle: results of the ICU Liberation Collaborative in over 15,000 adults. Crit Care Med 2019 Jan;47(1):3-14. [doi: 10.1097/CCM.0000000000003482] [Medline: 30339549]
26. Hodgson CL, Stiller K, Needham DM, et al. Expert consensus and recommendations on safety criteria for active mobilization of mechanically ventilated critically ill adults. Crit Care 2014 Dec 4;18(6):658. [doi: 10.1186/s13054-014-0658-y] [Medline: 25475522]
27. Zhang L, Hu W, Cai Z, et al. Early mobilization of critically ill patients in the intensive care unit: a systematic review and meta-analysis. PLoS ONE 2019;14(10):e0223185. [doi: 10.1371/journal.pone.0223185]
28. Lian KH, Yang CC, Hu FC, et al. Quality of life outcomes after surgical intervention in patients with multiple rib fractures: a prospective cohort study. Surgery 2023 Apr;173(4):1066-1071. [doi: 10.1016/j.surg.2022.12.006] [Medline: 36658082]
29. Liu HY, Lin TH, Chen KC, Hsiao WL, Hu RH, Liao HC. Comparison between non-surgical and surgical management of rib fractures in major trauma patients without brain injuries. Am J Surg 2023 Sep;226(3):350-355. [doi: 10.1016/j.amjsurg.2023.05.025] [Medline: 37263888]
30. Brooke J. SUS: a quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, editors. Usability Evaluation In Industry: CRC Press; 1996:189-194 URL: https://www.taylorfrancis.com/chapters/edit/10.1201/9781498710411-35/sus-quick-dirty-usability-scale-john-brooke [accessed 2026-01-28]
31. Bangor AK, Kortum P, Miller JT. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009(3):114-123 [FREE Full text]
32. Wassenaar A, Schoonhoven L, Devlin JW, et al. Delirium prediction in the intensive care unit: comparison of two delirium prediction models. Crit Care 2018 May 5;22(1):114. [doi: 10.1186/s13054-018-2037-6] [Medline: 29728150]
33. Zhang S, Ding S, Cui W, Li X, Wei J, Wu Y. Evaluating the effectiveness of a clinical decision support system (AI-Antidelirium) to improve nurses' adherence to delirium guidelines in the intensive care unit. Intensive Crit Care Nurs 2025 Apr;87:103933. [doi: 10.1016/j.iccn.2024.103933] [Medline: 39787945]
34. Tan A, Durbin M, Chung FR, et al. Design and implementation of a clinical decision support tool for primary palliative Care for Emergency Medicine (PRIM-ER). BMC Med Inform Decis Mak 2020 Jan 28;20(1):13. [doi: 10.1186/s12911-020-1021-7] [Medline: 31992301]
35. Schaaf J, Sedlmayr M, Sedlmayr B, Prokosch HU, Storf H. Evaluation of a clinical decision support system for rare diseases: a qualitative study. BMC Med Inform Decis Mak 2021 Feb 18;21(1):65. [doi: 10.1186/s12911-021-01435-8] [Medline: 33602191]

XSL•FO
RenderX

36. Jung H, Park HA, Lee HY. Impact of a decision support system on fall-prevention nursing practices. J Patient Saf 2023 Dec 1;19(8):525-531. [doi: 10.1097/PTS.0000000000001168] [Medline: 37922246]
37. Dunn H, da Costa Ferreira Oberfrank N, Krupp A. Preimplementation of critical care early mobility clinical decision support: a content validation study. Comput Inform Nurs 2024 Jun 1;42(6):463-469. [doi: 10.1097/CIN.0000000000001121] [Medline: 38512354]
38. Wilson-Jene H, Cowan RE, Roy Z, Pearlman J. Development of the RightWheel manual wheelchair wheel rolling resistance clinical decision support system through iterative clinician interviews. J Rehabil Assist Technol Eng 2025;12(20556683251336184):20556683251336184. [doi: 10.1177/20556683251336184] [Medline: 40291810]
39. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3(17):17. [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
40. Bright TJ, Wong A, Dhurjati R, et al. Effect of clinical decision-support systems: a systematic review. Ann Intern Med 2012 Jul 3;157(1):29-43. [doi: 10.7326/0003-4819-157-1-201207030-00450] [Medline: 22751758]
41. Nguyen AS, Yang S, Thielen BV, et al. Clinical decision support intervention and time to imaging in older patients with traumatic brain injury. J Am Coll Surg 2020 Sep;231(3):361-367. [doi: 10.1016/j.jamcollsurg.2020.05.023] [Medline: 32561447]
42. Tignanelli CJ, Shah S, Vock D, et al. A pragmatic, stepped-wedge, hybrid type II trial of interoperable clinical decision support to improve venous thromboembolism prophylaxis for patients with traumatic brain injury. Implement Sci 2024 Aug 5;19(1):57. [doi: 10.1186/s13012-024-01386-4] [Medline: 39103955]
43. Umer A, Mattila J, Liedes H, et al. A decision support system for diagnostics and treatment planning in traumatic brain injury. IEEE J Biomed Health Inform 2019 May;23(3):1261-1268. [doi: 10.1109/JBHI.2018.2842717] [Medline: 29993563]
44. Dayan PS, Ballard DW, Tham E, et al. Use of traumatic brain injury prediction rules with clinical decision support. Pediatrics 2017 Apr;139(4):e20162709. [doi: 10.1542/peds.2016-2709] [Medline: 28341799]
45. Greenberg JK, Otun A, Kyaw PT, et al. Usability and acceptability of clinical decision support based on the KIIDS-TBI tool for children with mild traumatic brain injuries and intracranial injuries. Appl Clin Inform 2022 Mar;13(2):456-467. [doi: 10.1055/s-0042-1745829] [Medline: 35477149]
46. Lang E, Neuschwander A, Favé G, et al. Clinical decision support for severe trauma patients: machine learning based definition of a bundle of care for hemorrhagic shock and traumatic brain injury. J Trauma Acute Care Surg 2022 Jan 1;92(1):135-143. [doi: 10.1097/TA.0000000000003401] [Medline: 34554136]

Abbreviations

CDSS: Clinical Decision Support System
EMR: electronic medical record
ERATbi: Early Recovery After Traumatic Brain Injury
GCS: Glasgow Coma Scale
ICP: intracranial pressure
ICU: intensive care unit
SUS: system usability scale
TBI: traumatic brain injury

Edited by C Jacob; submitted 02.Jul.2025; peer-reviewed by KH Lin, M Hellaby, M Gagnon-Roy; revised version received 13.Dec.2025; accepted 08.Jan.2026; published 06.Feb.2026.

Please cite as:
Yen HC, Wu IH, Hsiao WL, Lai SR, Yang CH, Liao HC, Han YY
Development of the ERATbi App, a Clinical Decision Support System for Early Recovery After Traumatic Brain Injury in the ICU: Usability Study
JMIR Hum Factors 2026;13:e79981
URL: https://humanfactors.jmir.org/2026/1/e79981
doi:10.2196/79981

© Hsiao-Ching Yen, I-Hui Wu, Wei-Ling Hsiao, Sheng-Ru Lai, Chen-Hao Yang, Hsien-Chi Liao, Yin-Yi Han. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 6.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly

XSL•FO
RenderX

cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Overview of a User-Centered, Mixed-Methods Process for Designing Interconnected and Focused Mobile Applications on Patient Care Environment (InterFACE): Augmented-Reality Decision Support System for Pediatric Resuscitation

Frederic Ehrler[1], PhD; Ana Rajic[1], MSc; Alexandre De Masi[2], Phd; Sharleen Olanka[3], MSc; Marco Generelli[3], MSc; Jennifer Davidson[4,5], BSc, RN; Yiqun Lin[4], MD, MHSc, PhD; Kangsoo Kim[6], PhD; Pierre-Louis Rebours[1], MSc; Marc Ibrahim[1], MSc; Donovan Duncan[4,5,7], MD; Ryan Kang[6], MSc; Sergio Manzano[2,8], MD; Adam Cheng[5], MD, FRCPC; Johan N Siebert[2,8], MD

[1]Direction of information science, University Hospital of Geneva, Gabrielle Perret Gentil 4, Geneva, Switzerland

[2]Faculty of Medicine, University of Geneva, Geneva, Switzerland

[3]Educational Technologies and Learning Sciences (TECFA), Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland

[4]KidSIM-ASPIRE Simulation Research Program, Alberta Children's Hospital, University of Calgary, Calgary, AB, Canada

[5]Departments of Pediatrics and Emergency Medicine, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[6]Department of Electrical and Software Engineering, Schulich School of Engineering, University of Calgary, Calgary, AB, Canada

[7]Pediatric Intensive Care Unit, Alberta Children's Hospital, University of Calgary, Calgary, AB, Canada

[8]Department of Pediatric Emergency Medicine, Geneva Children's Hospital, University Hospital of Geneva, Geneva, Switzerland

Corresponding Author:
Frederic Ehrler, PhD
Direction of information science, University Hospital of Geneva, Gabrielle Perret Gentil 4, Geneva, Switzerland

Abstract

Background: Pediatric cardiopulmonary resuscitation (CPR) is a highly complex and time-critical process that demands precise team coordination and strict adherence to pediatric advanced life support (PALS) guidelines. In real-world practice, adherence often deteriorates due to cognitive overload, fragmented communication, and disruption of information flow under stress. Although digital cognitive aids have shown potential to improve adherence, existing tools are often limited to single tasks, lack team-wide integration, or fail to adapt in real time to dynamic clinical environments.

Objective: This study aimed to design and evaluate InterFACE (Interconnected and Focused Mobile Applications on Patient Care Environment), an integrated, augmented reality (AR)–enabled digital health system developed to support real-time PALS adherence and enhance team coordination during pediatric resuscitation.

Methods: A structured, mixed methods, user-centered design process was used. Persona development and spatial analysis characterized the needs and positions of key resuscitation roles. A 3-round Delphi process with experts identified critical information elements for display. Iterative user experience (UX) prototyping was performed, followed by simulation-based evaluations of three system components: (1) TeamScreen, a wall-mounted team display providing a shared overview of the resuscitation process; (2) Guiding Pad (developed by Pierre Louis Rebours and Marc Ibrahim), a tablet-based app for documentation and algorithm navigation; and (3) AR head-mounted displays (HMDs) for team leaders and medication nurses, delivering role-specific, context-aware guidance. Usability was assessed with standardized instruments, including the System Usability Scale (SUS), Technology Acceptance Model (TAM), and User Experience Questionnaire (UEQ).

Results: The Delphi study achieved consensus on 20 core information elements, distributed across the 3 interfaces. Usability testing demonstrated high acceptance across all modalities. The Guiding Pad supported effective navigation of resuscitation algorithms with a 78%-100% task completion rate. The TeamScreen achieved an overall task success rate of 81%, improving situational awareness despite some confusion in high-density regions. AR HMDs received favorable evaluations, with SUS scores rated “Good” to “Excellent,” and UEQ ratings indicating high intuitiveness, stimulation, and attractiveness. Participants consistently described InterFACE as intuitive, useful for real-time decision-making, and supportive of team synchronization. Reported challenges included interface complexity, incomplete integration with patient monitors, and potential cognitive load from simultaneous information streams.

Conclusions: InterFACE represents a significant advancement in digital cognitive aids by combining shared displays, tablets, and AR guidance into a synchronized, role-specific ecosystem. The system shows promise in enhancing adherence to PALS,

XSL•FO
RenderX

reducing cognitive load, and improving team coordination in simulated pediatric resuscitations. While results demonstrate strong usability and acceptance, further research is needed to evaluate clinical effectiveness in real-world settings, including randomized controlled trials, integration with hospital information systems via Fast Healthcare Interoperability Resources (FHIR) standards, and potential artificial intelligence–driven decision support to optimize adaptability and long-term skill retention.

(JMIR Hum Factors 2026;13:e78144) doi:10.2196/78144

KEYWORDS

implementation; rescucitation; augmented reality; pediatric; team coordination; digital cognitive aids; situational awareness; user-centered design

Introduction

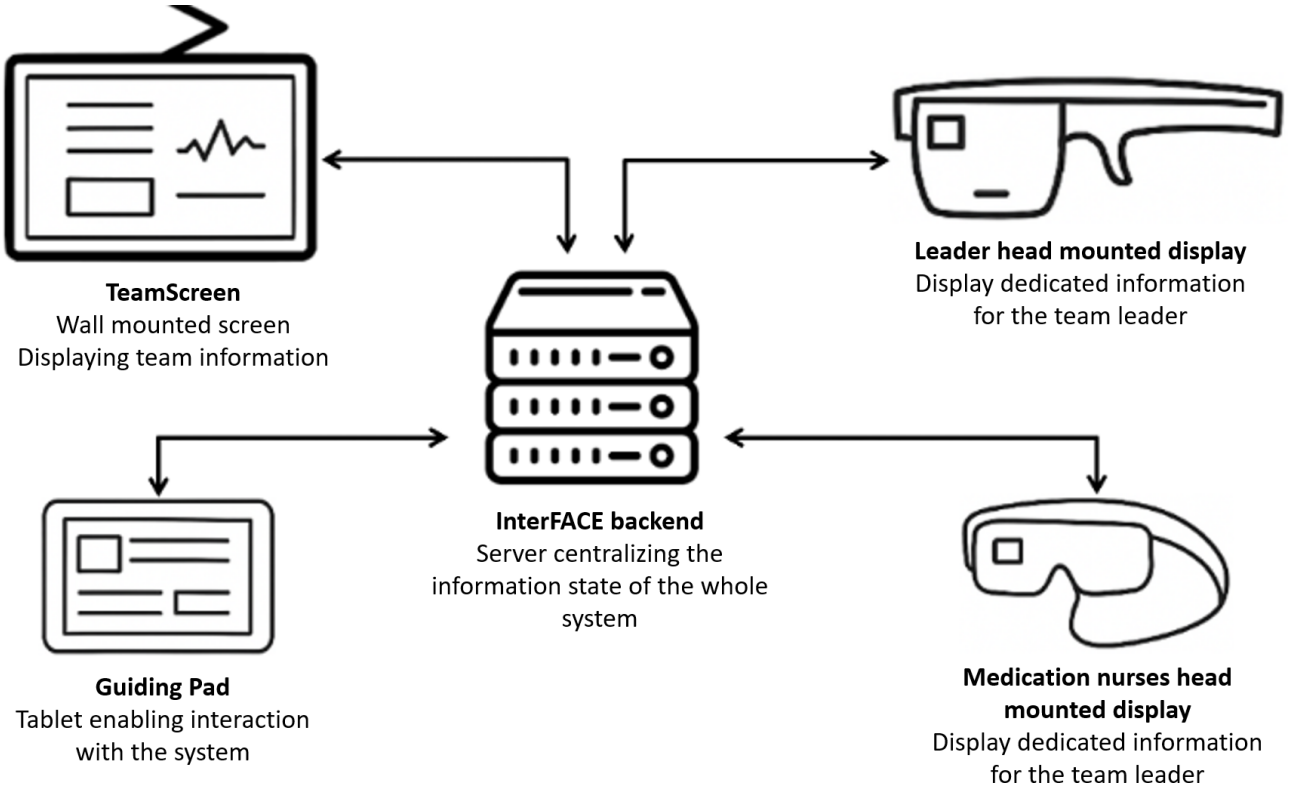
Cardiopulmonary resuscitation (CPR) in pediatric and adult emergencies is a complex, time-sensitive process requiring precise team coordination. Adherence to pediatric advanced life support (PALS) guidelines, such as those from the American Heart Association (AHA) [1] and European Resuscitation Council [2], is crucial for optimizing CPR quality and patient outcomes [3]. Yet, the high-stress environment of CPR often disrupts guideline adherence due to cognitive overload, miscommunication, and fragmented information flow, impairing decision-making and delaying critical interventions.

Digital health technologies offer promising solutions to support resuscitation teams in maintaining guideline compliance and coordination. Previous research has investigated digital resuscitation tools like head-mounted displays (HMDs) [4], tablet-based guidance systems [5,6], and mobile apps to support caregivers, enhance adherence to resuscitation guidelines, and minimize drug preparation errors [7,8]. While these technologies have demonstrated potential benefits, they also reveal limitations in usability, team-wide information sharing, and real-time adaptability to dynamic clinical scenarios.

To address these limitations, we developed InterFACE (Interconnected and Focused Mobile Applications on Patients Care Environment), an integrated digital suite designed to enhance pediatric CPR performance through real-time, role-specific guidance and improved team coordination aligned with pediatric PALS algorithms. InterFACE comprises three core components: (1) a wall-mounted display delivering a dynamic, real-time resuscitation overview for the team; (2) a tablet app, operated by the documenting (script) nurse, enabling CPR algorithm navigation and task documentation; and (3) mixed reality (augmented reality, AR) HMDs, worn by the team leader and medication nurse, providing context and role-specific, interactive guidance (Figure 1). Unlike traditional AR, which overlays static digital data, InterFACE leverages AR technology to integrate spatially anchored, interactive virtual elements into the physical environment to enhance situational awareness and enable hands-free access to critical information. InterFACE aims to deliver accurate, context-specific information to each team member in real time, enhancing CPR efficiently through improved team alignment and PALS guidelines adherence.

XSL•FO
RenderX

Figure 1. System architecture of InterFACE, illustrating the conceptual flow of information between the backend and user interfaces. InterFACE: Interconnected and Focused Mobile Applications on Patients Care Environment.



Unlike existing solutions that typically address a single aspect of the resuscitation process, such as medication preparation or algorithm visualization, InterFACE proposes a comprehensive, multirole synchronized system integrating wall-mounted displays, tablets, and head-mounted AR interfaces. This integration aims to improve shared cognition, role-specific guidance, and real-time adaptability—key elements often missing in previous digital resuscitation tools.

This paper aimed to present the user-centered design process and formative usability evaluation of InterFACE, an integrated, AR-based decision support system for pediatric resuscitation. Rather than providing exhaustive detail on each methodological step, we offer a cohesive, high-level overview of the mixed methods process, including Delphi consensus, usability testing of AR interfaces, and informatics architecture, which guided the system's development. Each component has been or will be the subject of dedicated manuscripts; here, our goal is to highlight how these complementary approaches were integrated into a coherent design framework for a complex clinical decision support system.

Methods

Overview

The InterFACE was designed and developed using a structured, multiphase, mixed methods approach, blending qualitative and quantitative methodologies to create a user-centered resuscitation decision support system aligned with clinical needs and real-world dynamics.

Persona Development and Spatial Analysis

Personas representing key pediatric resuscitation roles (eg, team leader, medication nurse, proceduralist) were developed based on direct field observations and a series of semistructured interviews with clinical experts. Each persona captured role-specific responsibilities, cognitive workload, and critical information needs during high-stress resuscitation scenarios. The purpose of these personas was not to be directly evaluated but to serve as design scaffolds that informed the overall architecture of the InterFACE system. Specifically, they were used to: define role-specific interface requirements; guide how information was distributed across the 3 system components (tablet, AR head-mounted display, and team screen); ensure that each interface aligned with the cognitive workflow and situational demands of its corresponding clinical role.

In parallel, a spatial analysis was conducted within the actual resuscitation environment. This consisted primarily of mapping the typical positions and movement patterns of team members, assigning them to defined spatial zones (eg, head of bed, medication zone, and documentation area). This approach allowed us to identify where each team member typically stands or moves, and how they interact during various phases of care. Combining persona insights with spatial analysis ensured that digital information was presented in a way that matched both the cognitive and physical workflows of pediatric resuscitation.

Expert Consensus Process: Delphi Study

A 3-round Delphi study established essential system information elements via expert consensus.

XSL•FO
RenderX

Round 1: Item Generation

Resuscitation experts from our research team reviewed 4 personas (team leader, scribe, medication nurse, and other team members) and current PALS guidelines, identifying critical real-time resuscitation data elements via open-ended responses. These data elements, such as heart rate, oxygen saturation, time of drug administration, and the next expected algorithm step, were thematically analyzed to generate an initial item list representing the minimal critical information required for decision-making and situational awareness during pediatric resuscitation.

Round 2: Prioritization

Each item was rated for importance on a pediatric 5-point Likert scale by resuscitation experts (emergency doctors and nurses) from Geneva Children's Hospital and Alberta Children's Hospital. Items with mean scores <3 were excluded; those with high variability (SD>1.5) were discussed further for agreement to decide their inclusion. Qualitative feedback refined the list.

Round 3: Final Consensus

Experts finalized the list, with ≥80% agreement defining inclusion.

Iterative User Experience (UX) Design

A human-centered process ensured usability and workflow alignment.

Static Prototyping

Initial prototypes visualized layout, role-specific interfaces, information architecture, and functionality, undergoing internal testing for readability, functionality, and clarity.

Dynamic Prototyping

Refined static prototypes added interactivity, minimizing cognitive workload and aligning with resuscitation constraints, validated through iterative testing.

Simulation-Based Evaluation

The usability and feasibility of the InterFACE suite was evaluated through simulated resuscitation scenarios designed to mimic real-world pediatric cardiopulmonary arrests. Each system component was independently tested to assess its performance, usability, and integration into clinical workflows. For the wall-mounted display, namely TeamScreen, participants completed 21 information-retrieval tasks, with performance gaged by task success rates, completion times, and score from the Post-Study System Usability Questionnaire (PSSUQ) [9]. The tablet underwent testing with 27 interaction tasks, where effectiveness was measured through task success, time-to-completion, and PSSUQ scores, ensuring the app's alignment with dynamic resuscitation demands. The AR HMDs were assessed following a simulated CPR scenario, during which the team leader and medication nurse received guidance via Microsoft HoloLens 2. Usability was evaluated using the System Usability Scale (SUS), a 10-item questionnaire measuring perceived usability (effectiveness, efficiency, satisfaction) [10]; Technology Acceptance Model (TAM), a 12-item questionnaire assessing perceived usefulness and ease of use [11]; and User Experience Questionnaire (UEQ), a 26-item questionnaire evaluating 6 user experience dimensions [12] to capture both practical and experiential feedback. This stepwise evaluation provided comprehensive insights into each component's strengths and areas for refinement, prioritizing clarity, accessibility, and real-time utility in high-pressure settings.

Informatics Development and Integration

Following the usability evaluations, informatics development synthesized insights from previous testing phases into a cohesive final system. This process emphasized user needs, cognitive ergonomics, and compatibility with clinical workflows, achieved through iterative refinement rather than a single overhaul. By addressing feedback on usability and functionality, the development phase ensured that InterFACE seamlessly integrated into resuscitation practices, balancing technical sophistication with practical application. The result was a system designed to minimize cognitive workload while maximizing support for pediatric CPR teams, ready for broader validation in future studies.

Ethical Considerations

This study did not involve patients, patient representatives, or the collection or use of identifiable personal health data. All study activities were limited to health care professionals participating in system design workshops, semi-structured interviews, and simulation-based usability evaluations of the InterFACE system. According to institutional and regional policies at the Geneva University Hospitals (HUG) and the University of Calgary, studies that exclusively involve health care professionals for the purposes of system design, usability testing, and simulation-based evaluation, without involving patients or identifiable personal data are not classified as human subject research and are therefore exempt from formal review by a Research Ethics Committee / Institutional Review Board. All participants were informed about the study objectives, procedures, and voluntary nature of participation and provided informed consent prior to participation. Data collected during usability sessions and interviews were anonymized, and no personally identifiable information was recorded or retained. No financial compensation was provided.

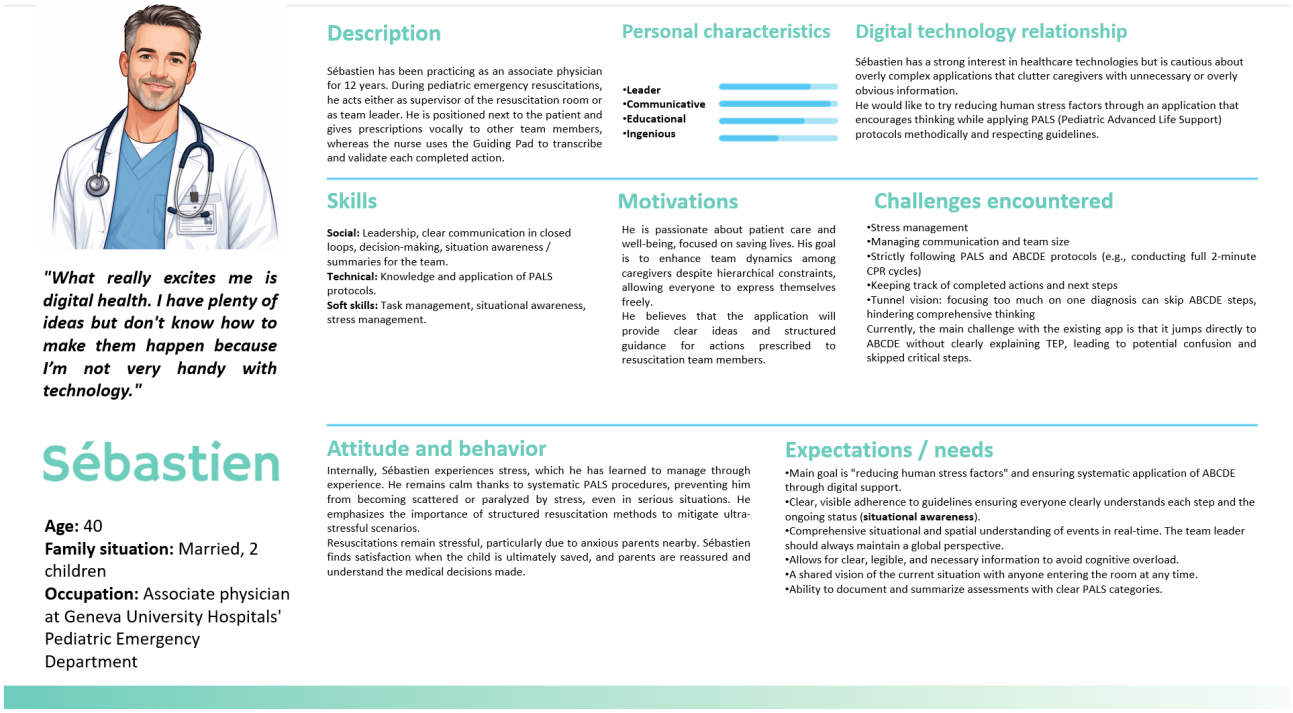
Results

Persona Development and Spatial Analysis

To explore the behaviors, information needs, communication methods, and expectations of key actors involved in pediatric resuscitation, we conducted in-depth face-to-face interviews with 4 nurses and 4 physicians. These interviews offered valuable insights into team roles and interactions during resuscitation, identifying areas where effective information flow enhances performance and where communication gaps may hinder outcomes. A total of 4 distinct personas were developed, each representing a key resuscitation role. First, the team leader, a central coordinator, focuses on decision-making, information synthesis, and communication, requiring real-time, comprehensive patient data and clear channels, as detailed in the persona profile example in Figure 2.

XSL•FO
RenderX

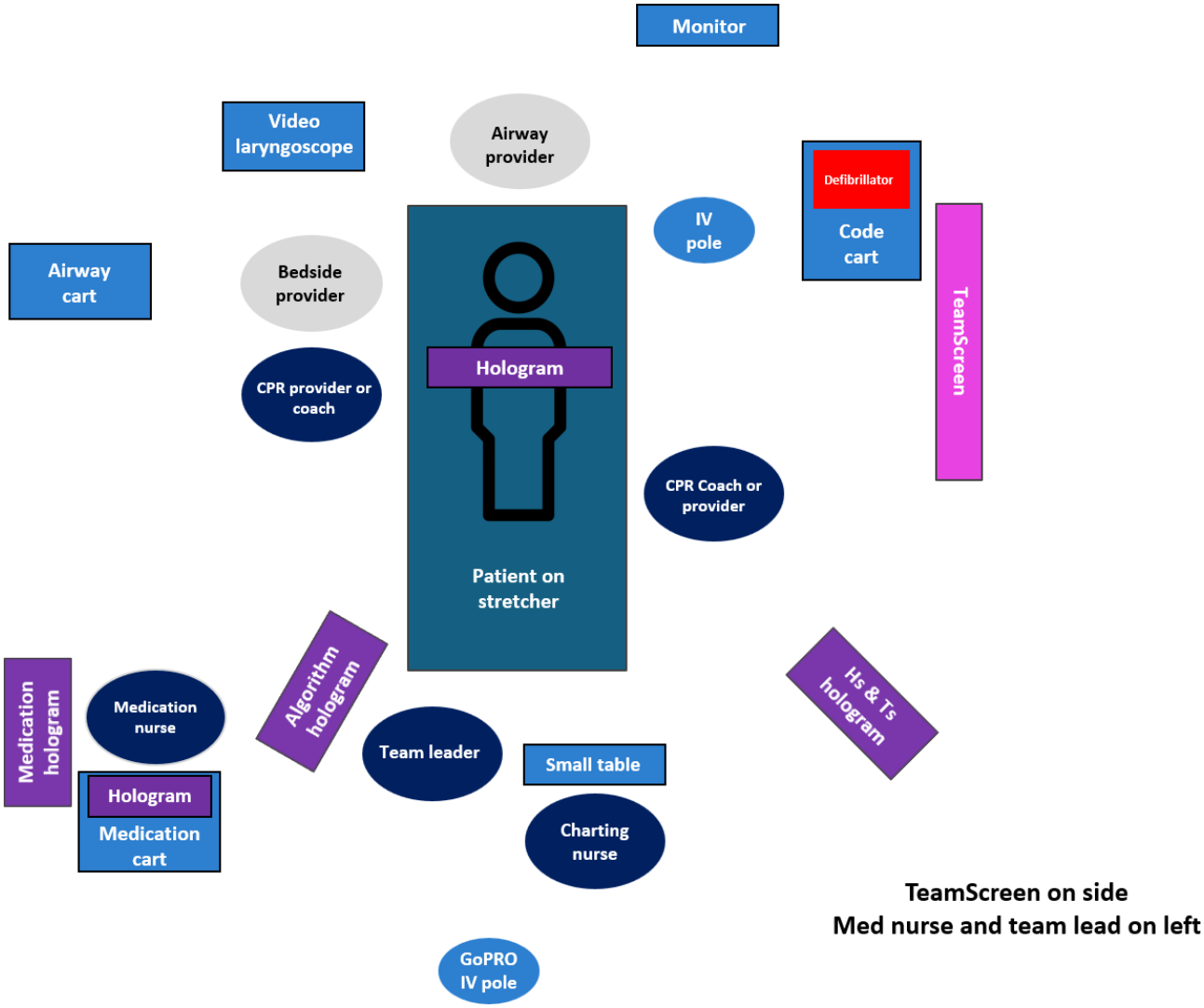
Figure 2. Persona profile of the team leader in pediatric resuscitation. This figure depicts a pediatric emergency physician with 12 years of experience as the team leader, detailing his responsibilities, competencies, motivations, challenges, and interaction with digital health technologies. The team leader plays a pivotal role in decision-making, coordination, communication, guideline adherence, and team performance under high-stress conditions.



Second, the documenting nurse (scribe), responsible for real-time documentation, emphasized the need for a streamlined interface to record events while tracking verbal directives. Third, the medication nurse, tasked with drug preparation and administration, highlighted the importance of quick access to drug information and dosage guidelines through an error-minimizing interface. Fourth, the rest of the team, encompassing supporting roles with dynamic responsibilities, stressed the need for flexible communication tools to enable real-time collaboration, task coordination, and problem-solving. A spatial analysis, conducted through mapping and observation during two pediatric cardiac arrest simulated resuscitation scenarios, revealed consistent positioning patterns (Figure 3): the team leader was centrally located at the head of the bed for optimal oversight, the scribe was nearby for rapid documentation access, the medication nurse was positioned close to medication for swift access, and the rest of the team adapted dynamically to support primary roles.

XSL•FO
RenderX

Figure 3. Cardiopulmonary resuscitation (CPR) provider placement in the study setting. Arrangement of team members and equipment around the manikin on a stretcher for effective resuscitation.



Delphi Study

The Delphi study's first round generated 40 distinct information items, grouped into 8 thematic categories: Patient status (7 items), Change of status (6 items), Presentation of clinical task (4 items), Running timer (4 items), Drugs (5 items); CPR and defibrillation (4 items); PALS (7 items); and Teamwork (3 items; Figure 3). In the second round, a panel of 15 pediatric emergency nurses and 15 physicians rated these items on a 5-point Likert scale (Table 1). Strong consensus emerged for most items, except for displaying intravenous drug doses on the AR HMD where nurses and physicians held divergent views, prompting further discussion on optimal presentation. No significant differences were observed between groups, with both preferring the team-wide screen over AR HMD for real-time information, suggesting the TeamScreen's role as the primary interface despite AR's potential.

XSL•FO
RenderX

Table . Selection of 40 information items from the first round of the Delphi study. An item was considered selected if a participant rated it above 3 on the Likert scale. The selected items are organized into 8 thematic categories related to effective cardiopulmonary resuscitation (CPR) and pediatric advanced life support (PALS) management.

Thematic categories and selected items	Glasses		TeamScreen	
	MD[a]	RN[b]	MD	RN
Patient status, n (%)				
HR[c]	5 (33.33)	8 (53.33)	11 (73.33)	14 (93.33)
RR[d]	5 (33.33)	7 (46.67)	11 (73.33)	14 (93.33)
SpO_2[e]	5 (33.33)	7 (46.67)	11 (73.33)	14 (93.33)
BP[f]	5 (33.33)	6 (40)	11 (73.33)	14 (93.33)
Temperature	3 (20)	2 (13.33)	9 (60)	9 (60)
$EtCO_2$[g]	5 (33.33)	3 (20)	10 (66.67)	13 (86.67)
Pulse	6 (40)	8 (53.33)	10 (66.67)	14 (93.33)
Change of status, n (%)				
Cardiac rhythm	9 (60)	9 (60)	12 (80)	14 (93.33)
HR<60[h]	10 (66.67)	10 (66.67)	11 (73.33)	15 (100)
RR<8[i]	10 (66.67)	11 (73.33)	10 (66.67)	15 (100)
SpO_2<80%	11 (73.33)	10 (66.67)	12 (80)	15 (100)
Loss of pulse	12 (80)	12 (80)	14 (93.33)	15 (100)
Return of pulse	9 (60)	12 (80)	11 (73.33)	15 (100)
Presentation of clinical tasks, n (%)				
Previous tasks	7 (46.67)	8 (53.33)	12 (80)	11 (73.33)
Current task	9 (60)	14 (93.33)	14 (93.33)	15 (100)
Next task	12 (80)	14 (93.33)	13 (86.67)	14 (93.33)
Running timer, n (%)				
CPR[j] timer (2-minute countdown)	12 (80)	10 (66.67)	15 (100)	15 (100)
Pause in CPR timer (hands off time)	7 (46.67)	8 (53.33)	13 (86.67)	12 (80)
Epinephrine timer (4-minute countdown)	11 (73.33)	14 (93.33)	14 (93.33)	15 (100)
Overall event duration	5 (33.33)	7 (46.67)	13 (86.67)	13 (86.67)
Medication, n (%)				
IV/IO[k] insertion location	0 (0)	0 (0)	4 (26.67)	5 (33.33)
Epinephrine dose	9 (60)	15 (100)	14 (93.33)	13 (86.67)
Amiodarone/ Lidocaine dose	7 (46.67)	13 (86.67)	12 (80)	14 (93.33)
Fluid bolus dose	4 (26.67)	12 (80)	9 (60)	13 (86.67)
Other medication dose	5 (33.33)	12 (80)	9 (60)	14 (93.33)
CPR and defibrillation, n (%)				
Initiate / resume CPR	9 (60)	8 (53.33)	15 (100)	13 (86.67)
Stop CPR / pulse check	9 (60)	7 (46.67)	15 (100)	14 (93.33)
Defibrillation timing (time to next shock)	10 (66.67)	7 (46.67)	12 (80)	14 (93.33)
Defibrillation dose	6 (40)	8 (53.33)	14 (93.33)	14 (93.33)
Advanced life support, n (%)				
Entire AHA[l] clinical algorithm (whole flow diagram)	6 (40)	5 (33.33)	11 (73.33)	13 (86.67)

XSL•FO
RenderX

Thematic categories and selected items	Glasses		TeamScreen	
	MD[a]	RN[b]	MD	RN
Advanced airway timing	4 (26.67)	7 (46.67)	7 (46.67)	9 (60)
ETT/LMA[m] size	2 (13.33)	4 (26.67)	11 (73.33)	10 (66.67)
Check tube placement	2 (13.33)	2 (13.33)	5 (33.33)	7 (46.67)
Review reversible causes	9 (60)	3 (20)	13 (86.67)	12 (80)
Bedside echo	4 (26.67)	3 (20)	4 (26.67)	7 (46.67)
Page ECMO[n] team	8 (53.33)	3 (20)	10 (66.67)	7 (46.67)
Teamwork, n (%)				
Role allocation	1 (6.67)	3 (20)	7 (46.67)	9 (60)
Case summary	2 (13.33)	6 (40)	10 (66.67)	10 (66.67)
Ask for team input	4 (26.67)	2 (13.33)	5 (33.33)	6 (40)
Mean (SD)	6.64 (2.76)	7.69 (3.19)	10.87 (2.22)	12.31 (2.36)

[a]MD: medical doctor.
[b]RN: registered nurse.
[c]HR: heart rate.
[d]RR: respiratory rate.
[e]SpO2: pulse oximetry.
[f]BP: blood pressure.
[g]EtCO2: end-tidal carbon dioxide.
[h]HR<60: heart rate less than 60.
[i]RR<8: respiratory rate less than 8.
[j]CPR: cardiopulmonary resuscitation.
[k]IV/IO: intravenous/intraosseous.
[l]AHA: American Heart Association.
[m]ETT/LMA: endotracheal tube/laryngeal mask airway.
[n]ECMO: extracorporeal membrane oxygenation.

The third round finalized a refined list, achieving over 80% consensus for inclusion. Although over two-thirds of experts strongly supported including patient vital signs, these were excluded due to technical constraints requiring monitor integration, which was beyond this version's scope. A total of 7 items, primarily medication, timer, and tasks information, were selected for the medication nurses' HMD, while the team leader's HMD included these plus CPR and advanced life support data, totaling 20 items. The TeamScreen displayed the most items, adding 3 more—previous task, overall resuscitation duration, and ETT/LMA size. While physicians and nurses showed similar preferences, nurses favored broader visibility on the TeamScreen for monitoring patient status and rhythm changes, aligning with their assessment role, whereas physicians were more accepting of HMDs for immediate alerts. Both groups prioritized shared screens for critical and retrospective data, underscoring team coordination's importance.

Static Prototyping

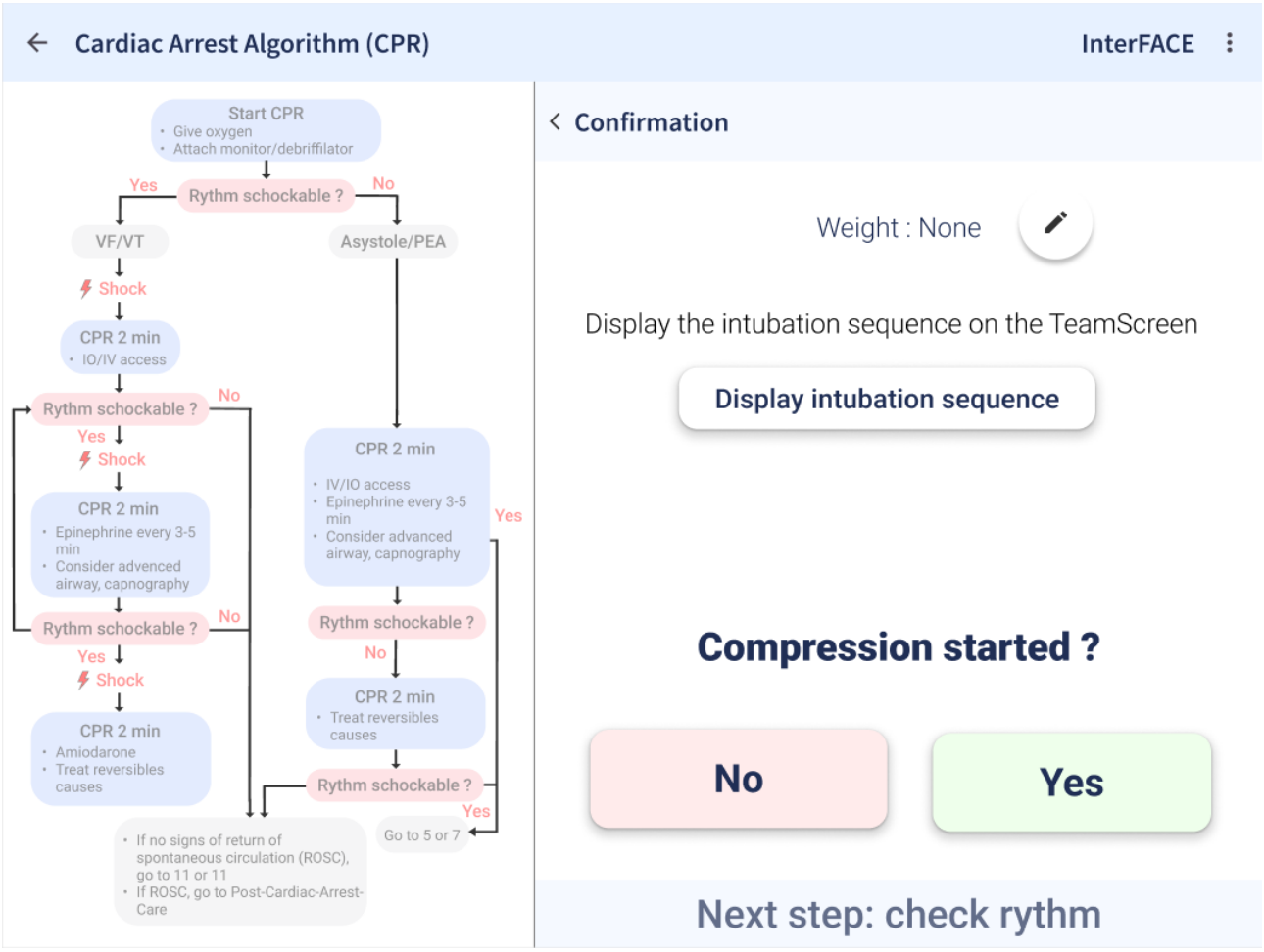
The initial system design was developed using Figma, a cloud-based collaborative graphic design and user interface prototyping software. The interface was structured around role-specific needs and PALS guidelines. Through internal testing and iterative refinement, the focus was on optimizing readability, functionality, and clarity, ensuring the prototype was ready for end user evaluation.

Mobile Tablet App Design

The Guiding Pad tablet app was designed based on user interaction and cognitive workload analysis, resulting in a 2-zone interface (Figure 4). The left zone supports advanced interactions, such as navigating PALS algorithms, accessing clinical data, and adjusting to evolving patient conditions, while the right zone focuses on primary task execution, offering real-time, context-sensitive prompts aligned with PALS algorithms.

XSL•FO
RenderX

Figure 4. Screenshot of the guiding pad tablet app. This screenshot illustrates the guiding pad interface, with the left panel displaying the cardiac arrest algorithm as a visual workflow guide and the right panel presenting an interactive confirmation prompt to ensure adherence to critical steps, such as verifying chest compressions before proceeding.

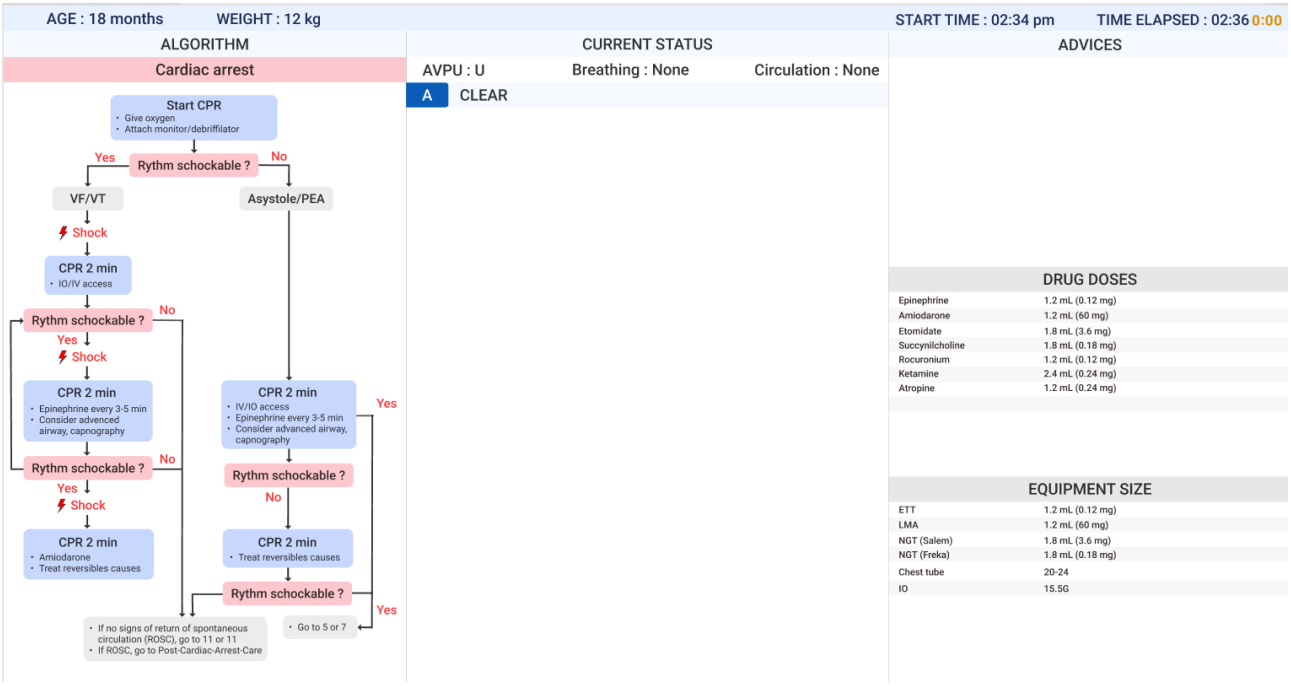


Team Screen Design

The team-wide display was organized into 3 functional sections to enhance situational awareness, prioritize actions, and streamline information retrieval (Figure 5). The left section visually represents the full PALS algorithm, enabling the team to track progress and anticipate steps. The center section highlights current required actions, ensuring team-wide awareness. The right section provides rapid access to critical drug information and ABCDE assessment results, minimizing workflow disruptions.

XSL•FO
RenderX

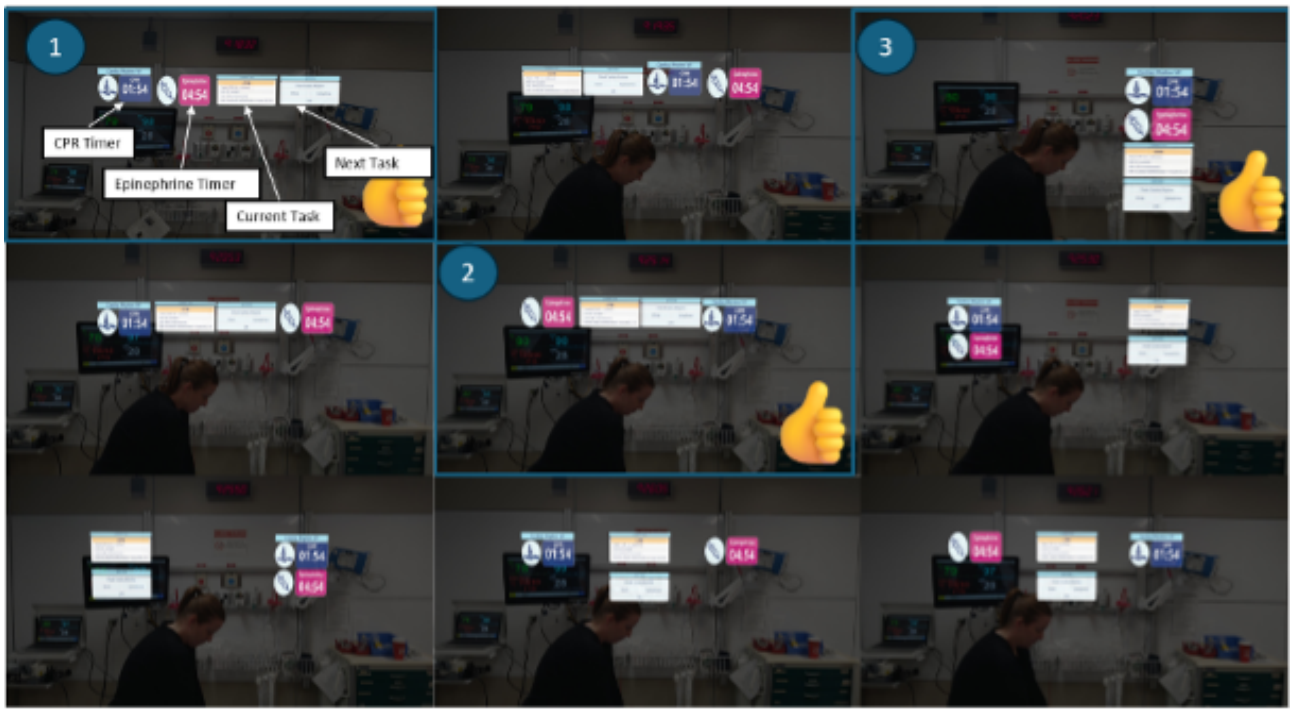
Figure 5. Early design for TeamScreen interface for PALS Cardiac arrest algorithm. Algorithm panel: cardiac arrest algorithm flowchart. Current status panel: progression in the ABCDE evaluation. Advices panel: advice for the current resuscitation step and information about drug doses and equipment size.



AR HMD Design

Based on the information prioritized during the initial design phase, several static layout prototypes were developed to explore different ways of presenting content within the AR display for the team leader (Figure 6). These mock-ups aimed to evaluate the most effective positioning of information in the field of view, with a focus on readability, prioritization of critical data, and minimizing cognitive load. Particular attention was paid to key AR design heuristics, including peripheral clarity (ensuring that essential information remains accessible even when not in central vision) and anchoring consistency (maintaining spatial predictability of interface elements).

Figure 6. Static layout propositions for the augmented reality (AR) head-mounted display (HMD) of the team leader.



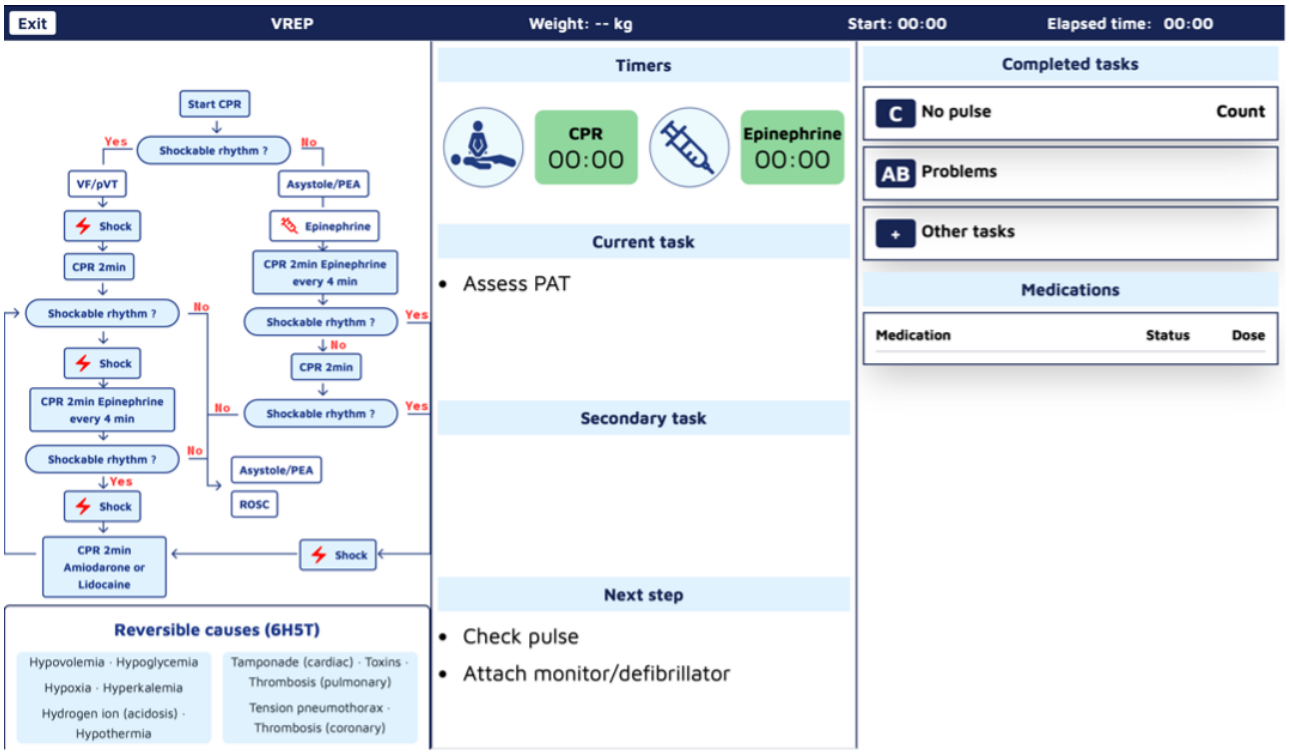
XSL•FO
RenderX

Dynamic Prototyping

Mobile Tablet App and Team Screen

Dynamic prototypes, developed in Figma, incorporated interactive elements simulating structured resuscitation workflows, task progression, and patient data visualization (Figure 7). A dozen weekly reviews with pediatric emergency and intensive care physicians, nurses, and system developers gathered qualitative feedback on information hierarchy, interaction efficiency, role-specific views, and guidance clarity. This ensured critical information was accessible, key actions required minimal steps, interfaces were tailored to roles, and system prompts supported clinical decision-making effectively.

Figure 7. Screenshot of the TeamScreen. This screenshot depicts the interface for cardiac arrest management, comprising multiple panels. Left panel: pediatric advanced life support (PALS) cardiac arrest algorithm detailing steps for VF/pVT and asystole/PEA, including emergency drugs and shock decisions. Center panel: visual representation of the patient with cardiopulmonary resuscitation (CPR) and epinephrine timers, current tasks, secondary tasks, and next steps. Right panel: completed tasks log with categories (no pulse, AB problems, other tasks) and medication dosing section. Bottom panel: reversible causes (6H5T). AB : airway/breathing; PEA: pulseless electrical activity; VF: ventricular fibrillation; pVT: pulseless ventricular tachycardia.



User feedback drove key adjustments and interaction logic. The Guiding Pad adopted a 2-zone layout with the left zone for advanced interactions and the right zone for primary task execution and information access. The TeamScreen featured algorithm visualization on the left, prioritized actions in the center, and real-time ABCDE assessment and drug data on the right. Interaction flows were streamlined for seamless transitions between algorithm steps, enhancing usability and clinical alignment.

AR HMD

In the dynamic prototyping phase, the static layouts were implemented within a functional AR environment to evaluate their behavior under real-time conditions. This phase focused on assessing how interface elements responded to user movement, task switching, and environmental complexity. A particular emphasis was placed on occlusion management, ensuring that critical real-world visual cues—such as team members, medical devices, and patient access—remained unobstructed by virtual overlays. The prototypes were tested using simulated clinical scenarios, allowing the team to refine spatial placement, interaction timing, and adaptive content behaviors based on contextual relevance and physical positioning. These evaluations informed iterative adjustments to optimize both usability and clinical safety in dynamic, high-pressure environments.

Usability Testing of Individual Components

Guiding Pad

The Guiding pad usability test showed overall effectiveness, with the 15 participating nurses demonstrating ease of use during 27 tasks. Of these, 21 tasks achieved a 100% completion rate, while the remaining six ranged from 67% to 93% completion, indicating areas for refinement. The most frequently encountered issue involved advanced interactions, such as selecting the appropriate endotracheal tube size and insertion depth. Some participants were unaware that default values had been preselected. Another commonly failed task was the manual prescription of drugs; since the system automatically prescribed certain drugs, participants were sometimes uncertain about the required actions. Tasks completion times varied significantly, ranging from 2.93 seconds to 33.53 seconds on average, reflecting differences in task complexity.

XSL•FO
RenderX

TeamScreen

The TeamScreen evaluation, conducted with 5 clinicians and 15 nurses, highlighted the clear organization and meaningful presentation of information, with an 81% overall task success rate, reflecting participants' ability to locate and interpret essential data under time pressure. Interface elements like shock logs, intubation timestamps, and reversible - causes guidance were clear and easy to retrieve, but challenges in distinguishing ongoing CPR status and impending interventions (eg, epinephrine administration) highlighted the need for more intuitive visual structuring or prompts in high-density regions, as participants reported confusion over critical updates. Tasks completion times varied significantly, ranging from 2.35 seconds to 18.45 seconds on average, reflecting differences in task complexity.

AR HMD

In the usability evaluation of the AR HMD with 5 pediatric emergency physicians and 5 participating nurses participating, the system was found to be intuitive and efficient. Both the team leaders and medication nurses reported "Good" and "Excellent" SUS scores, respectively, with medication nurses benefiting from clear, task-specific prompts. The UEQ results indicated high attractiveness and positive evaluations in pragmatic qualities (intuitiveness, efficiency, and dependability), facilitating smooth task execution, and hedonic qualities (stimulation and novelty). The TAM confirmed high acceptance, with users perceiving the system as useful and easy to use during resuscitation scenarios.

Informatics and System Architecture Development With Final Integration

The InterFACE system was designed as a real-time digital support tool for pediatric resuscitation, following a task tree model akin to a business process modeling notation (BPMN) [13], with each PALS algorithm step represented as a cursor (open, in progress, completed, or canceled) to support sequential execution and decision-driven adjustments based on clinical inputs. The system comprises a backend microservice (emergency-guideline-service) and role-tailored frontend applications. The backend, built as a Spring Boot microservice [14], handles business logic via a Representational State Transfer (REST) application programming interface (API) for data exchange and a server-sent events (SSE) API for real-time updates, following a domain-driven design approach with a root aggregate managing PALS algorithms, drugs, and clinical observations. An event-driven architecture [15] with event sourcing (ES) ensures data consistency and low latency, logging events in MongoDB (version 8.0) for real-time broadcasting via SSE, without requiring continuous polling.

Frontend modules, developed in Angular 17 JavaScript framework, include the Guiding Pad user interface (UI) for the scribe (supporting documentation, algorithm adjustments, and both sending commands and listening backend events; Figure 8), the TeamScreen UI for the team (displaying algorithm steps, prompts, and parameters in read-only mode while passively receiving real-time updates via SSE; Figure 9), and the AR HMD UI for the medication nurse and team leader (enabling hands-free data access; Figures 10 and 11). Event sourcing transmits incremental changes to frontend clients—reducing data load by avoiding the need to send the entire state—with events timestamped and logged in MongoDB for full process reconstruction and auditability. For deployment, the backend runs as a Docker container, frontends are served via an NGINX server, and a short code identifies each resuscitation process, supporting concurrent sessions within the hospital network and remote guidance, allowing experts to assist or lead resuscitations from a distance. However, for performance optimization, real-time data is stored in memory and not persisted after server shutdown, as this proof-of-concept targets simulated patients in future studies, without authentication.

XSL•FO
RenderX

Figure 8. Implemented version of the guiding pad. The left side shows initially the CPR algorithm but enables also the access to advanced interaction such as medication prescription, weight input, airways documentation, and reversible causes documentation. The right part of the screen is dedicated to documenting the main tasks of the current step of the algorithm.

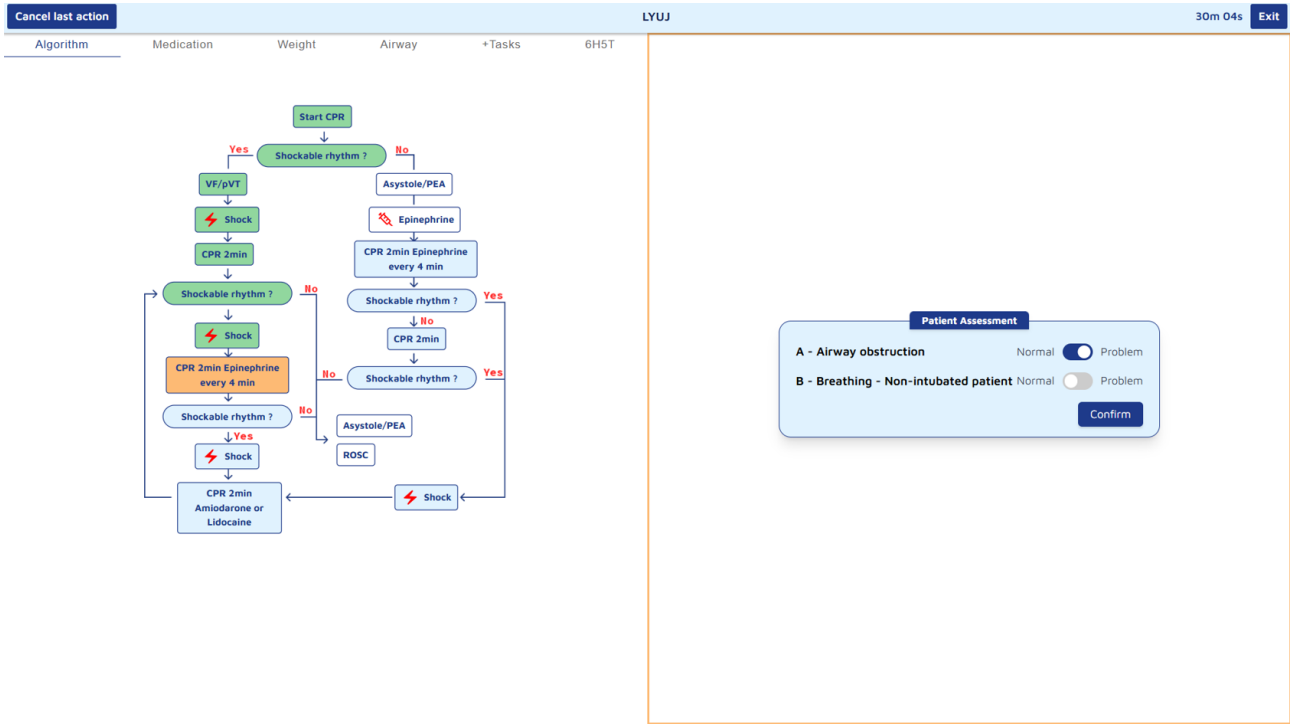
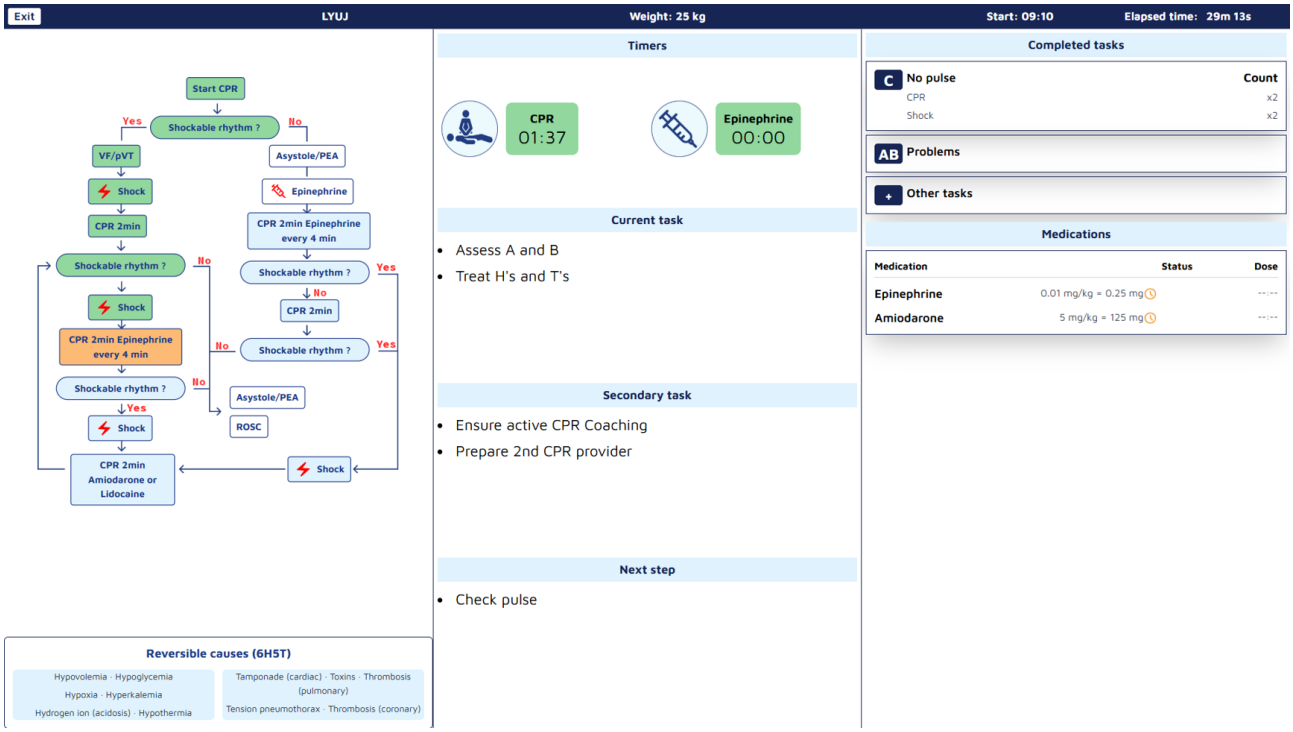


Figure 9. Implemented version of the TeamScreen. The left part contains the cardiopulmonary resuscitation algorithm showing the executed steps in green and the current one in orange. Underneath, we found the reversible causes. The middle column displays the most important information. Two counters indicate the cardiopulmonary resuscitation remaining time and the time to epinephrine administration. The middle part displays the current primary and secondary tasks, and the last part is dedicated to the next step. Finally, the right column provides information about the ABCDE assessment as well as a view on the administered medication.



XSL•FO
RenderX

Figure 10. Vision of the team leader in the augmented reality (AR) display. The same information displayed in TeamScreen, such as the algorithm, the reversible causes, the current and next tasks, as well as the counter, is available on the field of vision.



Figure 11. Final augmented reality (AR) head-mounted display (HMD) interface for the medication nurse showing role-specific information, such as drug doses, timers, and administration prompts.



Discussion

InterFACE System as a Cognitive Aid in Pediatric Resuscitation

InterFACE represents an advancement in digital cognitive aids for pediatric resuscitation, as evidenced by its design and evaluation through simulation-based approaches. The system integrates AR HMDs, tablets, and large displays within the resuscitation room, aiming to enhance shared cognition and temporal coordination among teams. This support is intended to ensure adherence to PALS guidelines, manage cognitive workload, and augment, rather than replace, clinical judgment, addressing key needs in high-stakes emergency situations. Although the personas themselves were not directly evaluated in the usability testing phase, they provided the conceptual

XSL•FO
RenderX

scaffolding for the system's architecture. By structuring the distribution of information and interaction logic around role-specific personas, the subsequent prototypes and evaluations could directly test whether these design assumptions translated into improved usability, workflow fit, and team coordination.

Comparison With Existing Digital Resuscitation Tools

Several digital tools have been developed to support pediatric resuscitation, focusing on improving guideline adherence, reducing cognitive overload, and enhancing coordination. Previous research, such as studies on HMDs [4,16], team situation displays [17,18], mobile apps for medication administration [7,8], CPR corrective feedback devices [19,20], and tablet-based resuscitation guidance tools [5,6], highlights their benefits in specific areas. However, these technologies often lack comprehensive team-wide information sharing, role-specific guidance, and real-time adaptability to evolving clinical situations. InterFACE addresses these limitations by synchronizing multiple role-specific interfaces through an event-driven architecture with event sourcing, ensuring continuous real-time, context-adaptive updates without manual intervention. This facilitates seamless communication and coordination. Compared to PediAppRREST [6] and similar tools, InterFACE stands out by offering synchronous, role-specific guidance across multiple interfaces, including immersive AR displays, and by leveraging an event-driven architecture to ensure consistent, low-latency updates across the care team. This enables shared situation awareness and real-time decision support that is both context-sensitive and visually anchored in the physical care environment.

Advancing Resuscitation Support With Mixed Reality

AR has shown promise in both resuscitation training and real-time procedural support. A recent systematic review highlighted the potential of AR in these areas ([21]). For example, an AR CPR feedback system provided real-time visual feedback on chest compression quality, significantly improving performance with compliance rates increasing from 18% - 21% to 87% - 90% ([16]). However, such devices are limited to specific tasks, like compression feedback, and do not offer comprehensive guidance for the entire resuscitation process. AR extends beyond AR by offering spatially anchored, interactive digital elements that provide hands-free, dynamic, and context-aware prompts. This technology enables a more holistic approach to resuscitation support. Via AR HMDs, InterFACE delivers real-time, role-specific decision support for critical tasks such as airway management, drug administration, and adherence to PALS algorithms. By providing a shared digital platform, InterFACE enables seamless team coordination to support the delivery of high-quality CPR.

While AR and virtual reality have been explored for resuscitation training, demonstrating potential in skill acquisition and knowledge retention, current evidence does not consistently show their superiority over traditional training methods [21]. This inconsistency limits their direct application in clinical settings. InterFACE addresses this gap by transitioning from static training models to real time, adaptive clinical support. By structuring digital support around team coordination and integrating directly into emergency workflows, InterFACE aims to optimize resuscitation performance beyond what is achievable through simulation alone, ultimately improving patient outcomes.

Technical Contributions and Innovations

InterFACE is built on a microservices architecture to structure critical care workflows effectively, with a key innovation being its event-driven model using SSE for real-time updates, minimizing latency, reducing unnecessary data polling, and enhancing scalability in time-sensitive settings. It replaces BPMN-based execution with a flexible task-tree model, acknowledging the dynamic, nonsequential nature of resuscitation workflows in high-intensity environments.

Challenges and Limitations

Several limitations must be acknowledged. First, the current evaluation focused exclusively on usability and acceptability in simulated environments and did not assess clinical outcomes in real resuscitation cases. While simulations provide a controlled setting to test usability and workflow integration, they cannot fully reproduce the stress, variability, and unpredictability of real emergencies. This may limit the generalizability of the findings. To mitigate this, scenarios were designed to be as realistic as possible, involving experienced clinicians. Future research should include randomized controlled trials in clinical settings.

Second, the complexity of pediatric resuscitation inherently increases cognitive load, particularly given the variability in patient age, size, and drug dosing. ([22]). Balancing comprehensiveness with simplicity in the interface was challenging. Iterative prototyping and multiple rounds of user feedback were used to refine the information hierarchy, but residual risks of information overload may persist. Future versions should explore adaptive content delivery and integration with patient-specific data.

Third, technical constraints limited system integration. The current prototype did not connect with live patient monitors or hospital information systems, which prevented inclusion of real-time vital signs—an item highly requested during the Delphi study. This may have reduced ecological validity. To mitigate this, placeholders and manual data inputs were provided to simulate the functionality. Future iterations should incorporate interoperability standards such as Fast Healthcare Interoperability Resources (FHIR) to enable seamless integration.

Fourth, the system was deployed in a single-instance Docker server without load balancing, restricting scalability for broader trials or multi-site use. Although this setup was sufficient for proof-of-concept testing, it may limit performance in high-demand contexts. Cloud-based deployment with redundancy and load balancing should be considered for larger clinical implementations.

In addition, the current version of InterFACE does not include core CPR quality metrics, such as compression-to-ventilation ratio, ventilation rate and volume, or end-tidal CO_2 , all of which are crucial indicators of resuscitation performance and highly relevant for team leader oversight. This omission reflects

XSL•FO
RenderX

the early-stage focus on workflow coordination and feasibility in simulation contexts, as well as technical constraints around integration with monitor data and CPR feedback devices. Future iterations of the system should prioritize the incorporation of these metrics through integration with patient monitors, compression sensors, and $ETCO_2$ waveforms, in order to provide the team leader with a more comprehensive, real-time view of resuscitation quality.

Fifth, some participants reported discomfort or distraction during prolonged use of AR HMDs. While refinements prioritized concise alerts and minimized unnecessary overlays, usability in extended real-world resuscitations may be affected. Gesture-based or multimodal interactions may improve user comfort in future versions.

Taken together, these limitations emphasize the need for further development, technical integration, and large-scale clinical validation before InterFACE can be widely adopted in routine pediatric resuscitation.

Future Directions

Future work should prioritize clinical validation through randomized controlled trials, assessing impact on guideline adherence, response times, and error reduction across different hospital settings. Artificial intelligence–powered decision support, leveraging machine learning on large-scale CPR data, could offer predictive guidance by dynamically adapting to patient-specific factors, historical resuscitation patterns, and evolving clinical conditions. These advancements may reduce clinicians' cognitive burden load while enhancing decision-making in time-sensitive situations. Scalability and multisite deployment, supported by cloud-based architectures with load balancing, are essential for broader adoption. Finally, integration with hospital information systems via FHIR standards would enable seamless data exchange with electronic health records and medical device data streams, positioning InterFACE as a transformative tool in pediatric resuscitation.

Conclusion

This study provides a high-level overview of the user-centered, mixed methods process used to design and formatively evaluate InterFACE, an integrated augmented-reality decision support system for pediatric resuscitation. Usability testing across tablets, wall-mounted displays, and AR head-mounted devices demonstrated good acceptance, feasibility, and perceived usefulness in simulated scenarios. While clinical outcomes were not assessed, the results indicate that InterFACE can enhance team coordination and adherence to pediatric resuscitation guidelines in a controlled environment. These findings support the feasibility of deploying multimodal, AR-enhanced cognitive aids and provide a blueprint for further clinical evaluation.

Acknowledgments

The authors declare that they used a generative artificial intelligence tool (ChatGPT v5.1; OpenAI) in a limited manner, exclusively for language-related tasks such as linguistic reformulation, readability improvement, and grammatical checking of certain paragraphs. All scientific content, including the ideas, methodology, results, analyses, and conclusions, was entirely written, validated, and assumed by the authors.

Funding

This study was funded by grant support provided by the Private Foundation of the Geneva University Hospitals, the Alberta Children's Hospital Research Institute, the Alberta Children's Hospital Foundation, the Department of Pediatrics at the University of Calgary, The University of Calgary Clinical Research Fund Grant, and the NSERC Alliance – Alberta Innovates Advance Program.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: FE, AR, JNS
Methodology: FE, AR, JNS, AC, SM, RK, YL, JD, DD, ADM, SO, KK
Software: MI, PLR
Investigation: FE, AR, JNS, AC, SM, RK, YL, JD, DD, ADM, SO, KK
Writing – original draft: FE, JNS
Writing – review & editing: YL, JD, DD, JNS, SM, ADM, AR, SO
Supervision: FE
Project administration: FE

Conflicts of Interest

SM received research grant support from the HUG Foundation for work related to this project. AC received research grant support from the Alberta Children's Hospital Foundation, Alberta Children's Hospital Research Institute, Department of Pediatrics at the

XSL•FO
RenderX

University of Calgary, and the HUG Foundation for work related to this project. JS received research grant support from the Geneva University Hospitals (HUG) Private Foundation for work related to this project. None declared by other authors.

References

1. Topjian AA, Raymond TT, Atkins D, et al. Part 4: Pediatric Basic and Advanced Life Support: 2020 American Heart Association Guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2020 Oct 20;142(16_suppl_2):S469-S523. [doi: 10.1161/CIR.0000000000000901] [Medline: 33081526]
2. Van de Voorde P, Turner NM, Djakow J, et al. European Resuscitation Council Guidelines 2021: Paediatric Life Support. Resuscitation 2021 Apr;161:327-387. [doi: 10.1016/j.resuscitation.2021.02.015]
3. Wolfe HA, Morgan RW, Zhang B, et al. Deviations from AHA guidelines during pediatric cardiopulmonary resuscitation are associated with decreased event survival. Resuscitation 2020 Apr;149:89-99. [doi: 10.1016/j.resuscitation.2020.01.035] [Medline: 32057946]
4. Siebert JN, Ehrler F, Gervaix A, et al. Adherence to AHA Guidelines when adapted for augmented reality glasses for assisted pediatric cardiopulmonary resuscitation: a randomized controlled trial. J Med Internet Res 2017 May 29;19(5):e183. [doi: 10.2196/jmir.7379] [Medline: 28554878]
5. Siebert JN, Lacroix L, Cantais A, Manzano S, Ehrler F. The impact of a tablet app on adherence to American Heart Association guidelines during simulated pediatric cardiopulmonary resuscitation: randomized controlled trial. J Med Internet Res 2020 May 27;22(5):e17792. [doi: 10.2196/17792] [Medline: 32292179]
6. Corazza F, Arpone M, Tardini G, et al. Effectiveness of a novel tablet application in reducing guideline deviations during pediatric cardiac arrest: a randomized clinical trial. JAMA Netw Open 2023 Aug 1;6(8):e2327272. [doi: 10.1001/jamanetworkopen.2023.27272] [Medline: 37535352]
7. Siebert JN, Bloudeau L, Combescure C, et al. Effect of a mobile app on prehospital medication errors during simulated pediatric resuscitation: a randomized clinical trial. JAMA Netw Open 2021 Aug 2;4(8):e2123007. [doi: 10.1001/jamanetworkopen.2021.23007] [Medline: 34459905]
8. Siebert JN, Ehrler F, Combescure C, et al. A mobile device application to reduce medication errors and time to drug delivery during simulated paediatric cardiopulmonary resuscitation: a multicentre, randomised, controlled, crossover trial. Lancet Child Adolesc Health 2019 May;3(5):303-311. [doi: 10.1016/S2352-4642(19)30003-3] [Medline: 30797722]
9. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Hum Comput Interact 1995 Jan;7(1):57-78. [doi: 10.1080/10447319509526110]
10. Brooke J. SUS - a quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation in Industry 1996:189-194.
11. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep 1;13(3):319-340. [doi: 10.2307/249008]
12. Schrepp M, Hinderks A, Thomaschewski J. Construction of a benchmark for the User Experience Questionnaire (UEQ). IJIMAI 2017;4(4):40. [doi: 10.9781/ijimai.2017.445]
13. Pufahl L, Zerbato F, Weber B, Weber I. BPMN in healthcare: challenges and best practices. Inf Syst 2022 Jul;107:102013. [doi: 10.1016/j.is.2022.102013]
14. Dhalla HK. A performance comparison of RESTful applications implemented in Spring Boot Java and MS.NET Core. J Phys: Conf Ser 2021 Jun 1;1933(1):012041. [doi: 10.1088/1742-6596/1933/1/012041]
15. Denecke K, Kirchner G, Dolog P, et al. Event-driven architecture for health event detection from multiple sources. Stud Health Technol Inform 2011;169:160-164. [doi: 10.3233/978-1-60750-806-9-160] [Medline: 21893734]
16. Kleinman K, Hairston T, Smith B, et al. Pediatric chest compression improvement via augmented reality cardiopulmonary resuscitation feedback in community general emergency departments: a mixed-methods simulation-based pilot study. J Emerg Med 2023 Jun;64(6):696-708. [doi: 10.1016/j.jemermed.2023.03.058] [Medline: 37438023]
17. Ehrler F, Del Zotto M, Rouyer F, Weinhold T, Lovis C, Siebert J. Design of InterFACE: a tool to improve collaborative work and decision making during rescucitation. Stud Health Technol Inform 2018;255:117-121. [doi: 10.3233/978-1-61499-921-8-117] [Medline: 30306919]
18. Parush A, Mastoras G, Bhandari A, et al. Can teamwork and situational awareness (SA) in ED resuscitations be improved with a technological cognitive aid? Design and a pilot study of a team situation display. J Biomed Inform 2017 Dec;76:154-161. [doi: 10.1016/j.jbi.2017.10.009] [Medline: 29051106]
19. Nicolau A, Jorge I, Vieira-Marques P, Sa-Couto C. Influence of training with corrective feedback devices on cardiopulmonary resuscitation skills acquisition and retention: systematic review and meta-analysis. JMIR Med Educ 2024 Dec 19;10:e59720. [doi: 10.2196/59720] [Medline: 39699935]
20. Bennett DJ, Itagaki T, Chenelle CT, Bittner EA, Kacmarek RM. Evaluation of the augmented infant resuscitator: a monitoring device for neonatal bag-valve-mask resuscitation. Anesth Analg 2018 Mar;126(3):947-955. [doi: 10.1213/ANE.0000000000002432] [Medline: 28863023]
21. Cheng A, Fijacko N, Lockey A, et al. Use of augmented and virtual reality in resuscitation training: a systematic review. Resusc Plus 2024 Jun;18:100643. [doi: 10.1016/j.resplu.2024.100643] [Medline: 38681058]

XSL•FO
RenderX

22. Luten R, Wears RL, Broselow J, Croskerry P, Joseph MM, Frush K. Managing the unique size-related issues of pediatric resuscitation: reducing cognitive load with resuscitation aids. Acad Emergency Med 2002 Aug;9(8):840-847. [doi: 10.1111/j.1553-2712.2002.tb02175.x]

Abbreviations

AHA: American Heart Association
API: application programming interface
AR: augmented reality
BPMN: business process modeling notation
CPR: cardiopulmonary resuscitation
ES: event sourcing
FHIR: Fast Healthcare Interoperability Resources
HMD: head-mounted display
InterFACE: Interconnected and Focused Mobile Applications on Patient Care Environment
PALS: pediatric advanced life support
PSSUQ: Post-Study System Usability Questionnaire
REST: Representational State Transfer
SSE: server-sent event
SUS: System Usability Scale
TAM: Technology Acceptance Model
UEQ: User Experience Questionnaire
UI: user interface
UX: user experience

Edited by A Kushniruk; submitted 27.May.2025; peer-reviewed by GJ Noordergraaf, S Brooks; revised version received 03.Oct.2025; accepted 01.Dec.2025; published 13.Feb.2026.

Please cite as:
Ehrler F, Rajic A, De Masi A, Olanka S, Generelli M, Davidson J, Lin Y, Kim K, Rebours PL, Ibrahim M, Duncan D, Kang R, Manzano S, Cheng A, Siebert JN
Overview of a User-Centered, Mixed-Methods Process for Designing Interconnected and Focused Mobile Applications on Patient Care Environment (InterFACE): Augmented-Reality Decision Support System for Pediatric Resuscitation
JMIR Hum Factors 2026;13:e78144
URL: https://humanfactors.jmir.org/2026/1/e78144
doi:10.2196/78144

© Frederic Ehrler, Ana Rajic, Alexandre De Masi, Sharleen Olanka, Marco Generelli, Jennifer Davidson, Yiqun Lin, Kangsoo Kim, Pierre-Louis Rebours, Marc Ibrahim, Donovan Duncan, Ryan Kang, Sergio Manzano, Adam Cheng, Johan N Siebert. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 13.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Designing App Interfaces to Elicit Specific Emotional Responses and Improve Attention and Short-Term Memory in Patients With Insomnia Undergoing Brief Cognitive Behavioral Therapy: Within-Subject Eye-Tracking Experimental Pilot Study

Kuan-Chu Su[1*], MA; Hsiao-Yean Chiu[2,3,4,5*], PhD; Ko-Chiu Wu[6], PhD; Chia-Chi Chang[6], MDes

[1]College of Design, National Taipei University of Technology, Taipei, Taiwan

[2]School of Nursing, College of Nursing, Taipei Medical University, Taipei, Taiwan

[3]Department of Nursing, Taipei Medical University Hospital, Taipei, Taiwan

[4]Research Center of Sleep Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

[5]Research Center of Sleep Medicine, Taipei Medical University Hospital, Taipei, Taiwan

[6]Department of Interaction Design, National Taipei University of Technology, Rm.701-4, Design Building, No.1, Sec.3, Chung-Hsiao E. Rd, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Ko-Chiu Wu, PhD
Department of Interaction Design, National Taipei University of Technology, Rm.701-4, Design Building, No.1, Sec.3, Chung-Hsiao E. Rd, Taipei, Taiwan

Abstract

Background: Patients with insomnia have difficulty in both falling asleep and maintaining sleep. Individuals with long-term sleep deprivation are prone to poor concentration and impaired memory; however, these problems can be alleviated following brief behavioral treatment for insomnia (BBT-I). This study involved the design of an app called “Sleep Well” that enables individuals with insomnia to easily record their sleep behavior. The app guides users to recall and record sleep-related information, acquire sleep hygiene knowledge, and communicate with therapists online.

Objective: This study examined how specific sleep diary interface design features in a brief cognitive behavioral therapy for insomnia (BBT-I) app influence users’ attention and short-term memory. Using a combination of objective eye-tracking measures and subjective attention assessments, the study compared 3 interface designs to determine how visual layout, input modality, and interaction style interact with insomnia symptoms to affect attentional performance, memory accuracy, and user preference.

Methods: Three sleep diary interfaces were designed, varying background mode (day vs night), color scheme (blue vs green), box shape (circular, rounded rectangular, or rectangular), and input method (slide-in, tap, or type-in). A total of 33 participants completed standardized diary-entry tasks while eye movements were recorded using an eye tracker to capture gaze trajectories and visual attention patterns during app interaction. User experience, subjective attention, and interface preferences were assessed using structured questionnaires. Data were analyzed using descriptive statistics, nonparametric tests, Pearson correlation analysis, cross-tabulation analysis, and exploratory factor analysis to examine associations among interface design, attentional performance, memory accuracy, and user characteristics.

Results: A total of 33 participants (n=13, 39.4% male and n=20, 60.6% female) aged 20 to 64 years completed this study. Based on the Insomnia Severity Index, 6 of 33 (18.2%) participants had clinical insomnia and 13 of 33 (39.4%) reported insomnia symptoms. Most participants reported staying up late (22/33, 66.7%), and more than half of participants reported drinking tea (17/33, 51.5%). Interface design significantly influenced objective attentional performance, as measured by eye-tracking indicators of task efficiency and visual allocation. Sleep quality and insomnia symptoms were consistently associated with attentional and short-term memory outcomes, with memory accuracy varying across interfaces and showing particular sensitivity to sleep maintenance difficulties. Subjective attentional control was strongly associated with both eye-tracking metrics and memory performance, and interface preferences differed by insomnia status.

Conclusions: Interface design significantly modulates attention and short-term memory performance in users with insomnia. Eye-tracking revealed that insomnia symptoms and sleep quality influence visual attention and task efficiency, whereas subjective attentional control showed stronger and more consistent associations with memory accuracy than physiological eye-movement indicators. These findings suggest that cognitive processing during sleep diary completion relies more on internal attentional

XSL•FO
RenderX

states than on observable gaze behavior. Designing low-load, attention-supportive interfaces may therefore improve usability and data accuracy in digital BBT-I interventions.

(JMIR Hum Factors 2026;13:e79883) doi:10.2196/79883

KEYWORDS

cognitive behavioral therapy; emotional design; memory; eye tracking; attention; interface design; insomnia

Introduction

Background

Good sleep quality (SlpQ) not only improves learning, memory, retention, and recall, but it also facilitates the creation and innovation of solutions from new information [1]. Sleep further consolidates memories by transferring them from short-term to long-term memory. Sleep deprivation can affect the basal state of attention when performing simple tasks [2]. In addition to affecting memory consolidation, sleep deprivation also reduces attention, alertness, concentration, and problem-solving ability, thereby affecting work and daily life [3-6].

Cognitive behavioral therapy is currently the recommended first-line treatment for insomnia [7]. However, nearly half of patients prefer to receive such treatment online, which is often more convenient and efficient [8]. Cognitive behavioral therapy for insomnia (CBT-I) has been applied to different cohorts of patients with insomnia, such as mobile phone app–based CBT-I for adolescents of different genders [9] and internet-based CBT-I for patients with insomnia who continue to experience symptoms after drug treatment [10]. As mobile technology becomes more integrated into daily life, research on the use of mobile phones to treat insomnia has increased. In addition to strengthening sleep retraining through smartphones [11], CBT-I apps and platforms have also been developed, and the usability of such systems has been thoroughly reviewed [12].

Brief cognitive behavioral therapy for insomnia (BBT-I) retains the behavioral therapy component of CBT-I but shortens the treatment duration from approximately 6 to 8 weeks to just 4 weeks. BBT-I and CBT-I share the use of 2 key behavioral principles (ie, sleep restriction and stimulus control) to enable patients to sleep in a predictable and reliable manner. Sleep restriction limits the time a person spends in bed, whereas stimulus control refers to creating a clear and positive association between bed and sleep.

Owing to the shortened treatment time, BBT-I has been widely adopted for insomnia treatment. The benefits of BBT-I are comparable to those of CBT-I [13]; thus, BBT-I has been suggested as an effective intervention for insomnia management [7]. However, accurate information recording by patients with insomnia and clinicians remains key to effective treatment.

During the COVID-19 pandemic, researchers developed a digital brief therapy for insomnia and explored its adherence and efficacy [14]. In this treatment, patients with insomnia are required to operate a mobile phone interface, recall details of their daily life and sleep from the previous day, and complete a sleep diary. Individuals with insomnia therefore prefer specialized interface designs that allow them to maintain control of their pace of life and sleep patterns [15].

Because diagnostic criteria for categorizing insomnia and cognitive measures vary widely, there is a need to standardize methods for assessing insomnia and cognitive performance in research [5]. Currently, there is limited research on the usability of BBT-I app interfaces for information input and retrieval among patients with insomnia who experience memory and attention impairments.

Interface Design and Cognitive Considerations in Insomnia

According to the theory of working memory, individuals can remember only 4 to 6 pieces of information within a 20-second timeframe. In mobile application design, the commonly used waterfall layout may lead to reading times that exceed the limitations of human memory. In previous work, we employed a grid layout to highlight important information and demonstrated significant improvements in interface content memory [16]. Moreover, symmetry, order, simplicity, and regularity of interface layout reduce the difficulty of memory formation [17]. These qualities enhance esthetics, facilitate memorization, and reduce cognitive load. When developing an interface, designing elements symmetrically and arranging them in an ordered pattern based on size are recommended. In particular, aligning the edges and starting points of elements as much as possible also improves memorization.

From a physiological perspective, reading in dark mode under low ambient lighting reduces eyestrain. Visual fatigue in nighttime environments is related to color and screen brightness, such that lower screen brightness can reduce visual fatigue. In contrast, under normal lighting conditions, there appear to be no differences between using dark or light mode [18].

The colors green, blue, and black result in the least visual fatigue under low screen brightness conditions, whereas red results in the highest visual fatigue. In a study reviewing the effects of color on visual fatigue, blue was rated as the most favorable color for overall interface design [19]. Blue conveys calmness and stability, white conveys purity and peace, and green is considered a healthy color that can reduce stress and anxiety and elicit positive emotions [20,21].

Interface shape is an important factor in the emotional responses to interface design [22]. Many studies have used corner shape (eg, circular vs angular) to reduce cognitive load in the brain and obtain a better user experience. Patients rate circular elements higher because circles convey tenderness and caring emotions, whereas squares evoke a stronger sense of power, which may create pressure and is less likely to inspire positive emotions [23]. The use of boxes can change the spatial distribution of attention, affect the apparent size of objects, change potential visual perception, and thus affect user attention,

XSL•FO
RenderX

maintain device controllability, and reduce cognitive load [24,25].

For the design of a medical app interface, a previous study recommended that the page be operated using click and slide gestures and that instant suggestions be provided during the operation. To enhance readability, clear text in a font size no smaller than 14 points should be used. To distinguish function buttons, a blue-and-white background is recommended, and placing an icon in the upper-left corner of the page can improve visibility. This ensures that there is sufficient distance between function keys for user operations [26].

Eye Tracking and Cognitive Measures in Interface Evaluation

During information retrieval, eye movements, particularly fixation rate, are predictive for recovering previously remembered details but not for newly generated details. This suggests that during memory recall, the primary function of eye movements is to reproduce already perceived information rather than to generate entirely new details [27]. Moreover, blink frequency, total number of blinks, maximum blink duration, and pupil diameter are positively correlated with mental workload [28]. Thus, eye tracking provides insight into information processing during intensive tasks as the user manipulates interface features. Functional evaluation of the interface is combined with finding the best functionality to achieve the best design [29]. Eye-tracking metrics such as first fixation time, total fixation time, and total number of fixations allow detailed measurement of the psychological characteristics of online cognitive processes [30-32].

Study Aims

The aim of this study was to investigate how sleep diary interface design in a BBT-I app influences users' attention, visual attention patterns, and short-term memory. Three sleep diary interfaces with systematically varied visual layouts, color schemes, box shapes, and input modalities were developed and evaluated. A mixed methods approach was used: objective eye-tracking measures were used to assess distribution of attention and task efficiency during interface interaction, whereas subjective attention scales, memory recall tasks, and user preference questionnaires captured cognitive performance and user experience. The study further examined how insomnia severity, SlpQ, and related symptoms moderated these outcomes. It was expected that interfaces with lower cognitive load and clearer interaction structures would support more stable attention, higher memory accuracy, and greater user satisfaction, particularly among individuals with stronger insomnia-related cognitive vulnerability.

Methods

Introduction to Sleep App Functions

The functions of our app were developed according to the requirements of BBT-I. Under the guidance of specialist clinical psychologists and physicians, patients with insomnia were asked to complete a sleep diary and the Dysfunctional Beliefs and Attitudes about Sleep: Validation of a Brief Version-16, which helped therapists to diagnose and understand the distinct causes and types of insomnia for each patient. According to the type of insomnia, a sleep-aid behavioral plan was tailored, including sleep hygiene education, stimulus control, and sleep restriction. The content of sleep hygiene education included paying attention to the sleeping environment, adjusting indoor temperature and humidity, timing of caffeine intake, ensuring sufficient physical activity during the day, and controlling the length of naps.

In this study, we developed an app front-end and back-end management system named "Sleep Well." The front-end structure for patients consisted of the following components: (1) sleep timer, (2) sleep diary, (3) BBT-I health education, and (4) system settings. The sleep timer calculated bedtime and wake-up time. If the patient was unable to fall asleep, they could choose relaxation guidance videos or sleep-inducing music for assistance. The sleep diary was used to record various sleep-stage times, daytime habits from the previous day, and self-assessments of SlpQ and mental state (MS) after waking up. BBT-I health education conveyed correct knowledge related to sleep, which consisted of sleep hygiene, sleep behavior therapy, relaxation guidance, and sleep-inducing music. System settings included basic settings and message functions. During treatment, users could consult and interact online with dedicated therapists through the messaging area, and therapists could observe patients' usage status in the background (Figure 1). The back-end structure of the app consisted of the following components: (1) data center, (2) patient management, (3) message management, and (4) system settings. The data center collected statistics on physiological characteristics, lifestyle habits, recovery time, and sleep conditions of all patients with insomnia, enabling an understanding of the treatment results for a certain period or a certain group of people. Patient management reviewed daily sleep data, diary records, and consultation results. Message management was used by patients to send messages and by therapists to respond to patient queries. System settings contained therapist account management, permissions, and notification settings.

XSL•FO
RenderX

Figure 1. Association diagram of the front-end and back-end components of "Sleep Well."



Sleep Diary: Three Interface Design Features

Sleep diaries primarily record the sleep time and daily routine of patients with insomnia, providing therapists with an assessment of the factors affecting sleep and planning sleep hours for patients.

This study extracted the core function of the app (ie, the sleep diary record page) and extended it into 3 different interface designs by varying background (day vs night), color scheme, box shape, input information, and interactive operation methods. Interface A used night mode with a blue color scheme, circular boxes, and a card-style appearance; the time input format was mainly sliding. Interface B used day mode with a blue color scheme, rounded rectangular box, a pop-up appearance, and tap-based input operation. Interface C used day mode with a green color scheme, rectangular boxes, a keyboard-based appearance, and direct manual input (tape-in) interaction (Figure 2).

The day-night background used a light background to imply daytime and a dark background to simulate nighttime. Matching icon colors and box shapes could stimulate the users' feelings toward the interface, and the input information and interactive operation methods could measure operation preferences. Furthermore, participants completed subjective questionnaires, and an eye tracker was used to measure the gaze duration, gaze position, saccade duration, number of saccades (NS), and saccade direction (SDir) to provide an integrated analysis of the factors that affected the differences in attention and memory for the 3 interfaces.

XSL•FO
RenderX

Figure 2. Visual comparison of sleep diary interfaces (from left to right: interfaces A, B, and C).



Experiment Methods

In the first phase, participants completed standardized recall tasks using all 3 sleep diary interfaces (A, B, and C). The order of interface presentation was counterbalanced across participants using a Latin square design to control for potential order and learning effects, which are common concerns in within-subject experimental designs [33,34]. Each interface was operated twice, and both task accuracy and differences in completion time between the first and second trials were recorded to assess learning-related performance changes.

Eye-tracking data were collected during a predefined task round for each participant, with the specific interface assigned according to the counterbalanced sequence. This approach ensured consistent measurement conditions while minimizing fatigue and carryover effects across interfaces, in line with established eye-tracking methodology in human–computer interaction and cognitive task research [29,30,32].

Primary and Secondary Outcomes

The primary outcome of this pilot study was attentional performance during sleep diary interface interaction. Attentional performance was assessed using both objective eye-tracking indicators, including task completion time, NS, gaze stability, and SDir, and subjective attentional control, measured by the Attention Control Scale (ATTC). These measures were selected to capture complementary aspects of visual attention allocation, cognitive load, and attentional regulation during interface use.

Secondary outcomes included short-term memory performance during diary recall tasks and interface preference and user experience. Memory outcomes were examined to explore downstream cognitive effects associated with attentional demands, whereas preference-related measures were included to inform practical interface design implications for digital BBT-I applications.

Participant Recruitment and Eligibility Criteria

The study was conducted between April 1 and May 15, 2023. A total of 33 participants aged 20 to 63 years were voluntarily recruited through Taipei Medical University. All participants provided written informed consent and completed a self-administered questionnaire that included demographic information and the Insomnia Severity Index (ISI) [35]. Clinical insomnia was defined as meeting at least one of the following criteria more than 3 times per week for a minimum duration of 3 months: (1) difficulty initiating sleep, defined as taking more than 30 minutes to fall asleep after going to bed; (2) difficulty maintaining sleep, defined as being awake for more than 30 minutes after sleep onset (SlpOn); or (3) early morning awakening, defined as waking at least 30 minutes earlier than desired. In addition, participants with an ISI total score of 15 or higher were classified as having clinical insomnia [35]. Participants without clinically significant insomnia were defined as those with ISI total score of 14 or lower and no subjective complaints of insomnia symptoms.

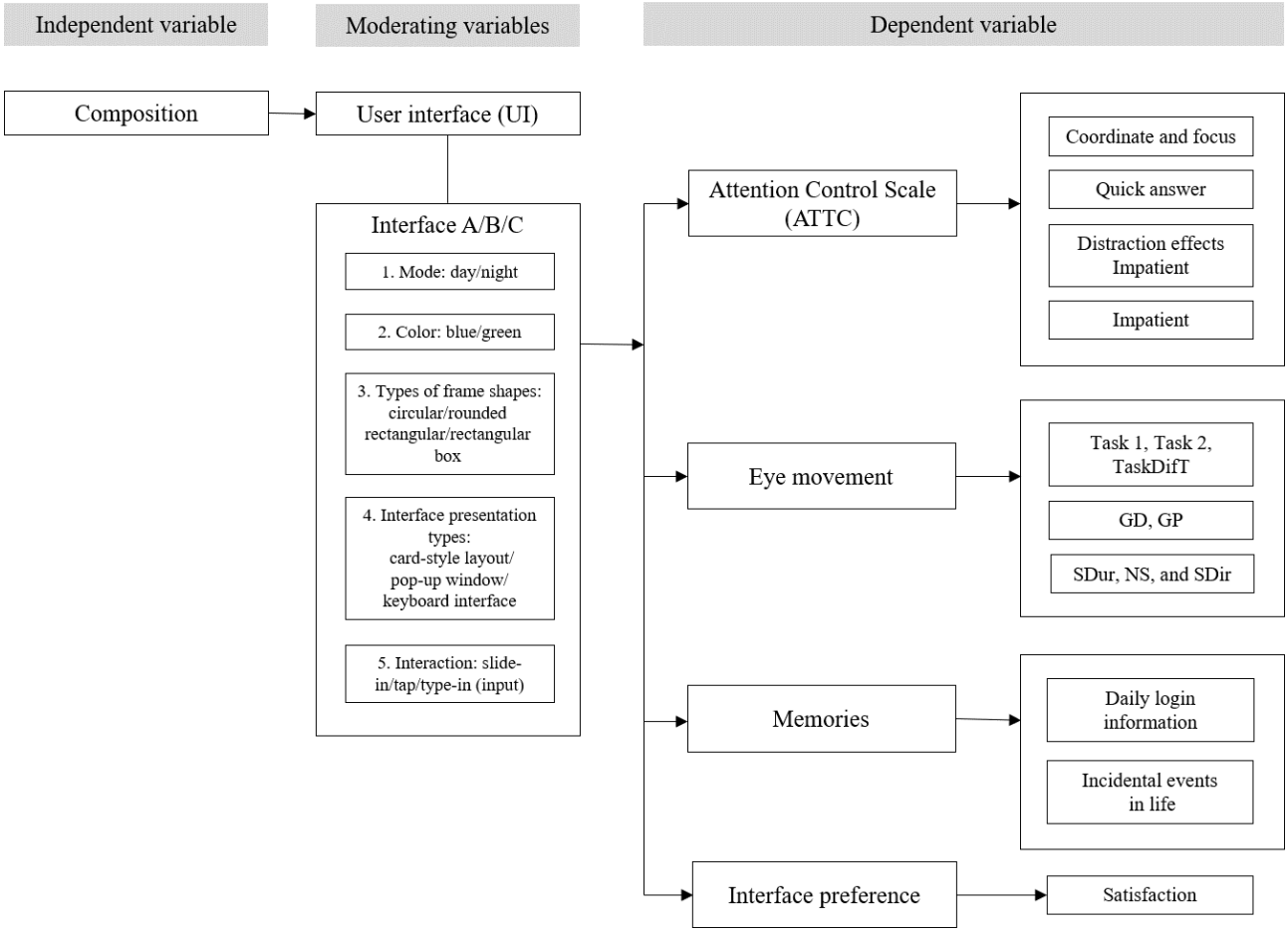
Data Collection

Independent variables included demographic characteristics (personal characteristics, lifestyle habits, sleep assessment [ISI-1a to 1c items], and insomnia [ISI-2 to 5 items]) and a MS questionnaire were analyzed through interface characteristics (screen polarity, box design, and layout, and information input method) and 4 dependent variables: (1) interface preference, (2) subjective attention (including coordinate and focus, quick answer, distraction effects, and impatient), (3) eye movement

XSL•FO
RenderX

data (operation time 1 [task 1], operation time 2 [task 2], difference in operation time [TaskDifT], gaze duration, gaze position, saccade duration, NS, and SDir), and (4) memories (daily login information and incidental events in life) (Figure 3). All study variables were categorized into conceptual domains, including population characteristics, lifestyle factors, sleep assessments, eye-tracking indicators, and memory recall measures. A complete classification of variables by name and type is provided in Multimedia Appendix 1.

Figure 3. Correlations between population characteristics (independent variables) and 4 dependent variables (ATTC, eye movements, memory recall, and interface preference). GD: gaze duration; GP: gaze position; NS: number of saccades; SDir: saccade direction; SDur: saccade duration; TaskDifT: difference in task operation time.



We examined whether the demographic characteristics in the 4 dimensions contained other factors that could influence a user's operation. Subjective attention measured the mentality of a user and their response to the operating interface. The eye tracker presented the attentiveness of a user and other physiological signals. Recall involved short-term and long-term memory as well as cognitive load.

Age, gender, lifestyle factors, sleep status (sleeping <7 h or below average sleep duration), severity of insomnia, and attention level were collected based on questionnaires and attention scales (Table 1). Eye movements were measured using an eye tracker (GP3 webcam, Gazepoint), and subjective opinions of the users for all 3 interfaces were collected through a postquestionnaire.

The recorded variables of the app operation were recall tasks that required inputting sleep information daily, including SlpOn, out-of-bed time, and wake after sleep onset (WASO; life events such as drinking (drinking items and drinking time [DrkTi]); exercise items (ExcsIm), exercise start time, and exercise end time, and nap start time and nap end time.

XSL•FO
RenderX

Table . Factor classification of the Attention Control Scale, with ATN[a] denoting item numbers, based on exploratory factor analysis.

Classification and question	Content	Detailed description
Coordinate and focus		
ATN5	Difficulty in coordinating recall	When filling out the sleep diary, I find it difficult to coordinate attention between completing optional information and recalling the task description
ATN6	Ease of responding after distraction	After being interrupted or distracted, I can easily refocus on the content I am filling out
Quick answer		
ATN2	Difficulty recalling information quickly	It is difficult to quickly recall the description of the task
ATN3	Filling in the form to generate interest	When I have a sleep task, I quickly become interested in filling out the sleep diary
ATN4	Time-consuming adaptation to operation	It takes me a while to get used to the mode of operation of filling out the sleep diary
Distraction effects		
ATN7	Unable to think simultaneously	It is impossible to think about two fields in the sleep diary at the same time
ATN8	Distraction from others	I easily become distracted when someone is in the same space as I am when filling out the sleep diary or doing a recollection
ATN9	Rapid switch mode	I can quickly switch from the previous interface operation mode to the current interface operation mode
ATN1	Inattention to surroundings	I can easily concentrate when filling out the sleep diary without being aware of what is happening around me
Impatient		
ATN10	Impatience while answering questions	I rush to answer the questions in the sleep diary before I finish reading them.

[a]ATN: Attention Control Scale item number.

Ethical Considerations

The TMU Joint Institutional Review Board from Taipei Medical University approved this study, which was planned, carried out, analyzed, and interpreted independently of any industry partners (approval number N202302007). All participants provided voluntary, informed, and written consent. The capability of the patients to participate in the study was ensured. During data analysis, the information of the participants was pseudonymized and coded using identification numbers. Only researchers assigned to the study had access to the data. No compensation was provided to participants for their participation and no images of individual participants or identifiable features are included in the paper or Multimedia Appendix 2.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics (version 26; IBM Corp). Analyses included descriptive statistics, nonparametric tests, Pearson product–moment correlation coefficient analysis, cross-tabulation analysis, and exploratory factor analysis. Each participant completed the questionnaire 3 times, corresponding to the 3 different interfaces. To control for multiple comparisons, the Bonferroni correction was applied. The significance level (α=.05) was divided by 3, resulting in an adjusted *P* value threshold of .017 for each individual test. Eye-tracking data were collected using the GP3 webcam eye tracker (Gazepoint). Recorded data were processed and analyzed using Gazepoint Analysis software, which is supported exclusively on the Windows operating system.

Results

Participant Characteristics

A total of 33 participants (n=13, 39.4% male and n=20, 60.6% female) of age 20 to 64 years completed this study. Based on the ISI, 27 out of 33 (81.8%) participants were classified as those without clinically significant insomnia, and 6 out of 33 (18.2%) were classified as having clinical insomnia. Based on subjective reports, 13 out of 33 (39.4%) participants were categorized as having self-reported insomnia (SRI), and 20 out of 33 (60.6%) were categorized as having no self-reported insomnia (NSRI). Demographic and sleep-related characteristics are summarized in Tables 2 and 3.

In total, 22 out of 33 (66.7%) participants reported staying up late, including 12 NSRI and 10 SRI individuals. Daily tea

XSL•FO
RenderX

drinking was reported by 17 out of 33 (51.5%) participants, of whom 11 were NSRI. Daily coffee drinking was reported by 15 out of 33 (45.5%) participants. In addition, 5 out of 33 (15.2%) participants were current smokers, 2 out of 33 (6.1%) reported alcohol drinking, 7 out of 33 (21.2%) reported long-term medication use, and 18 out of 33 (54.5%) reported having a regular sleep schedule.

Table . Cross-analysis of gender and age group (N=33).

Age group (y)	Male, n	Female, n	Total, n (%)
20 - 29	4	6	10 (30.3)
30 - 44	6	10	16 (48.5)
45 - 64	3	4	7 (21.2)
Total, n (%)	13 (39.4)	20 (60.6)	33 (100)

Table . ISI[a] and cross-analysis table of patients with SRI[b] and NSRI[c].

ISI score	NSRI, n (%)	SRI, n (%)	Total, n (%)
0-7 (no insomnia)	14 (70)	0 (0)	14 (42.4)
8-14 (mild insomnia)	6 (30)	7 (53.8)	13 (39.4)
15-21 (moderate insomnia)	0 (0)	5 (38.5)	5 (15.2)
22-28 (severe insomnia)	0 (0)	1 (7.7)	1 (3)
Total	20 (100)	13 (100)	33 (100)

[a]ISI: Insomnia Severity Index.

[b]SRI: self-reported insomnia.

[c]NSRI: no self-reported insomnia.

Attentional Performance (Eye-Tracking Measures)

Task efficiency differed by demographic and sleep factors. Age was associated with longer task completion times on Interface B (H=8.943, P=.01) and Interface C (task 1: H=10.075, P=.006; task 2: H=12.472, P=.002). Gender differences were observed in saccade counts on Interface C (H=7.294, P=.007).

Sleep characteristics were consistently associated with eye tracking outcomes. Poorer SlpQ was associated with a higher NS on Interface B (r=−0.448, P=.01) and Interface C (r=−0.491, P=.005). Sleep maintenance difficulty (ISI-1b) was associated with longer optimal operation time on Interface A (r=0.462, P=.007) and increased NS on Interface C (r=0.430, P=.02). Worry about sleep (ISI-5) was negatively associated with operation time 2 (task 2) on Interface C (r=−0.426, P=.01), indicating altered task dynamics.

Lifestyle factors also influenced gaze behavior. Smoking and alcohol drinking were associated with SDir variability on Interfaces A and B (χ^2=18.942 - 28.562, P≤.02; Figure 4).

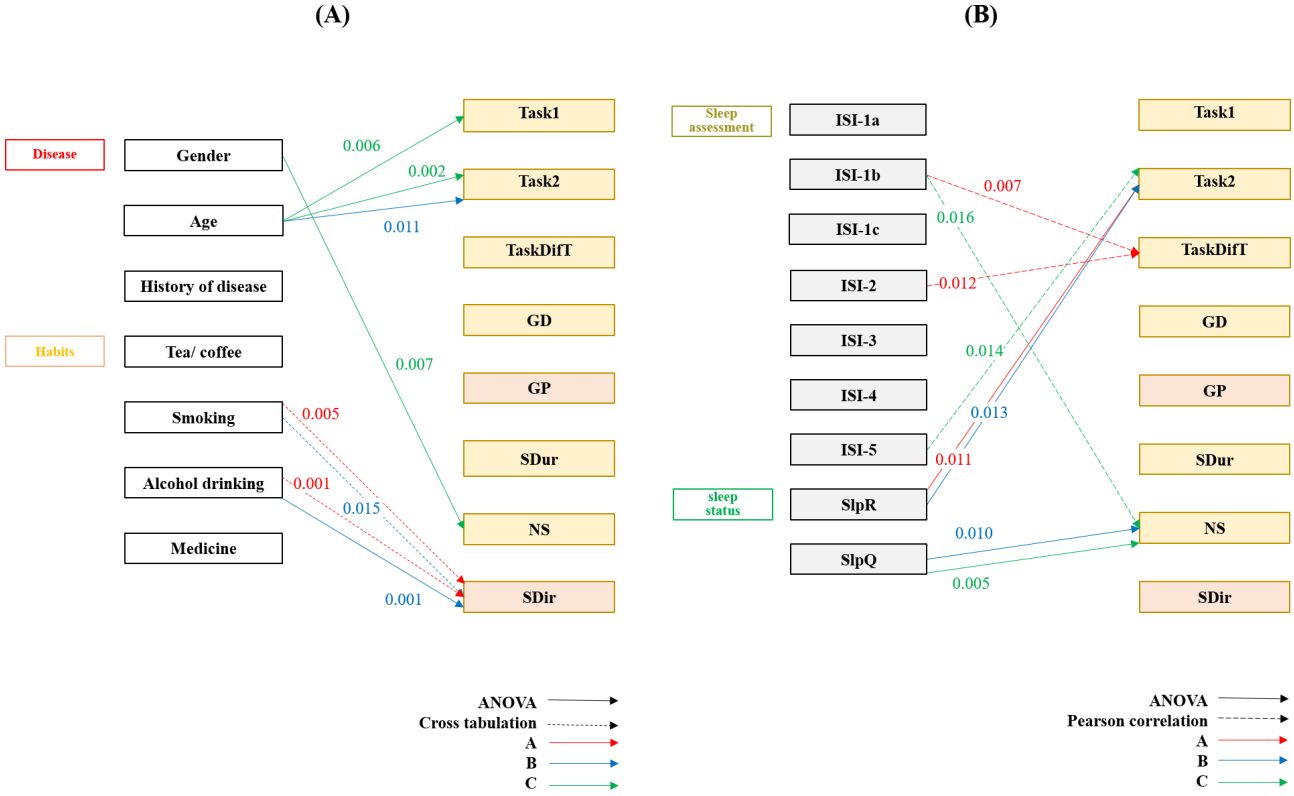
Eye-tracking results indicated that visual attention allocation and task efficiency differed across interfaces. Older age was consistently associated with longer task operation times, particularly on Interfaces B and C, indicating age-related reductions in interaction efficiency. Gender differences were observed in NS in Interface C, suggesting distinct visual scanning strategies.

Sleep-related factors played a central role in attentional performance. Poorer SlpQ and irregular sleep patterns were associated with increased NS and longer operation times, reflecting higher cognitive workload during interface interaction. Lifestyle behaviors, including smoking and alcohol drinking, were associated with altered SDir on Interfaces A and B, indicating changes in visual exploration patterns.

Among the 3 designs, Interface A (night mode, circular boxes, sliding input) was characterized by fewer saccades and more stable gaze transitions, suggesting reduced attentional dispersion and lower cognitive demand during task execution.

XSL•FO
RenderX

Figure 4. Relationships between (A) population characteristics and (B) insomnia symptoms (ISI) and eye-tracking information. GD: gaze duration; GP: gaze position; ISI: Insomnia Severity Index; NS: number of saccades; SDir: saccade direction; SDur: saccade duration; SlpQ: sleep quality; SlpR: sleep regularity; TaskDifT: difference in task operation time.



Short-Term Memory Performance

Age was significantly associated with the task of remembering ExcsIm in Interface A (H_2=11.886, P=.003).

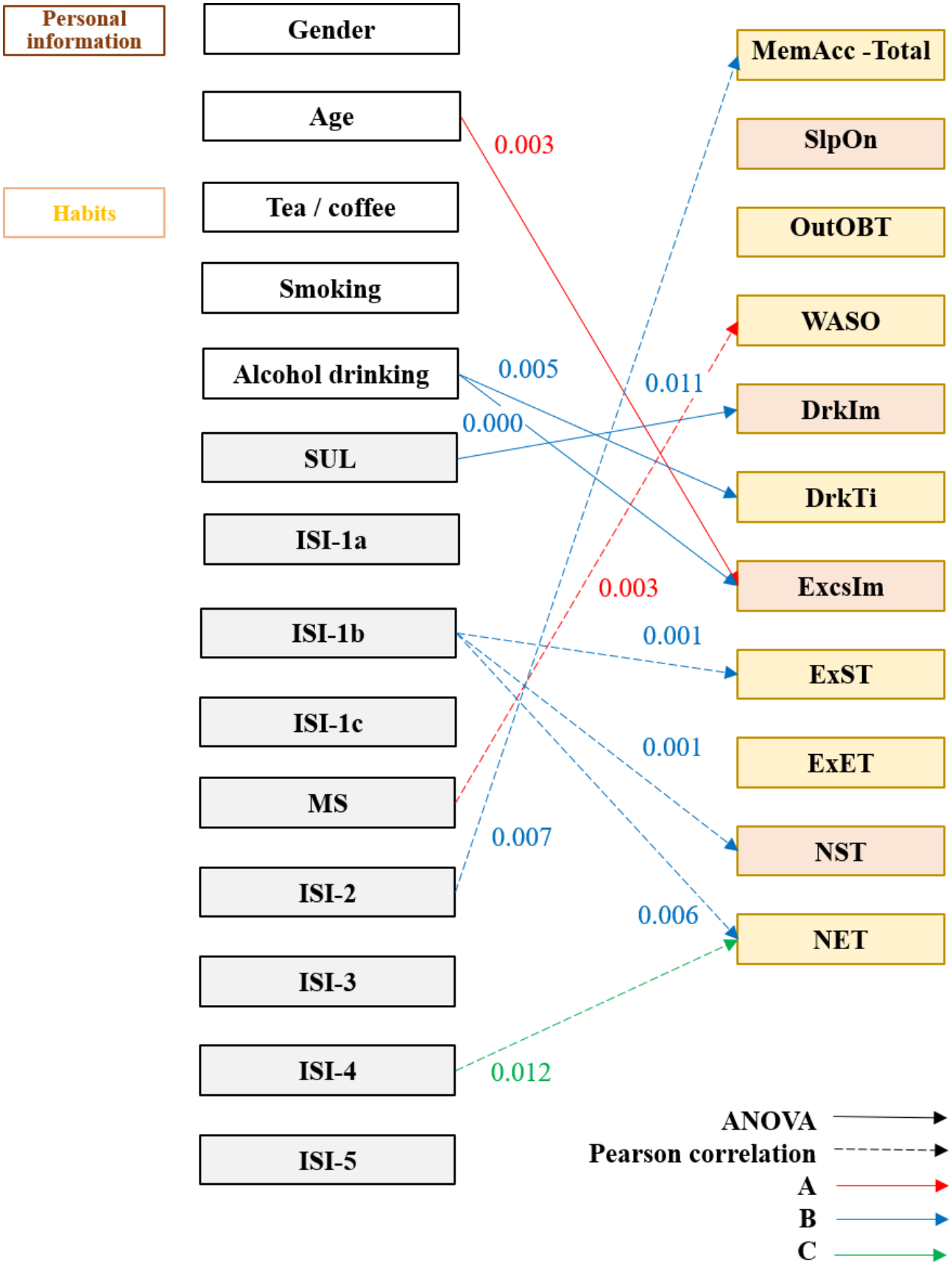
Alcohol drinking showed significant effects on the task of DrkTi in Interface B (H_2=10.551, P=.005) and the task of ExcsIm in Interface B (H_2=15.500, P<.001). Staying up late was also associated with drinking items in Interface B (H_1=6.400, P=.01).

Current sleep status was related to memory accuracy (total) on Interface A (r=–0.505, P=.003) and Interface A (r=–0.558, P<.001), as well as nap end time on Interface B (r=–0.468, P=.006).

ISI-2 was associated with memory accuracy total (r=–0.407, P=.007), and ISI-4 was associated with DrkTi in Interface C (r=–0.431, P=.01; Figure 5)

XSL•FO
RenderX

Figure 5. Relationships between population characteristics, insomnia symptoms (ISI), and memory recall scores. DrkIm: drinking items; DrkTi: drinking time; ExcsIm: exercise items; ExET: exercise end time; ExST: exercise start time; ISI: Insomnia Severity Index; MemAcc-Total: total memory accuracy; MS: mental state; NET: nap end time; NST: nap start time; OutOBT: out-of-bed time; SlpOn: sleep onset; SUL: stay up late; WASO: wakefulness after sleep onset.



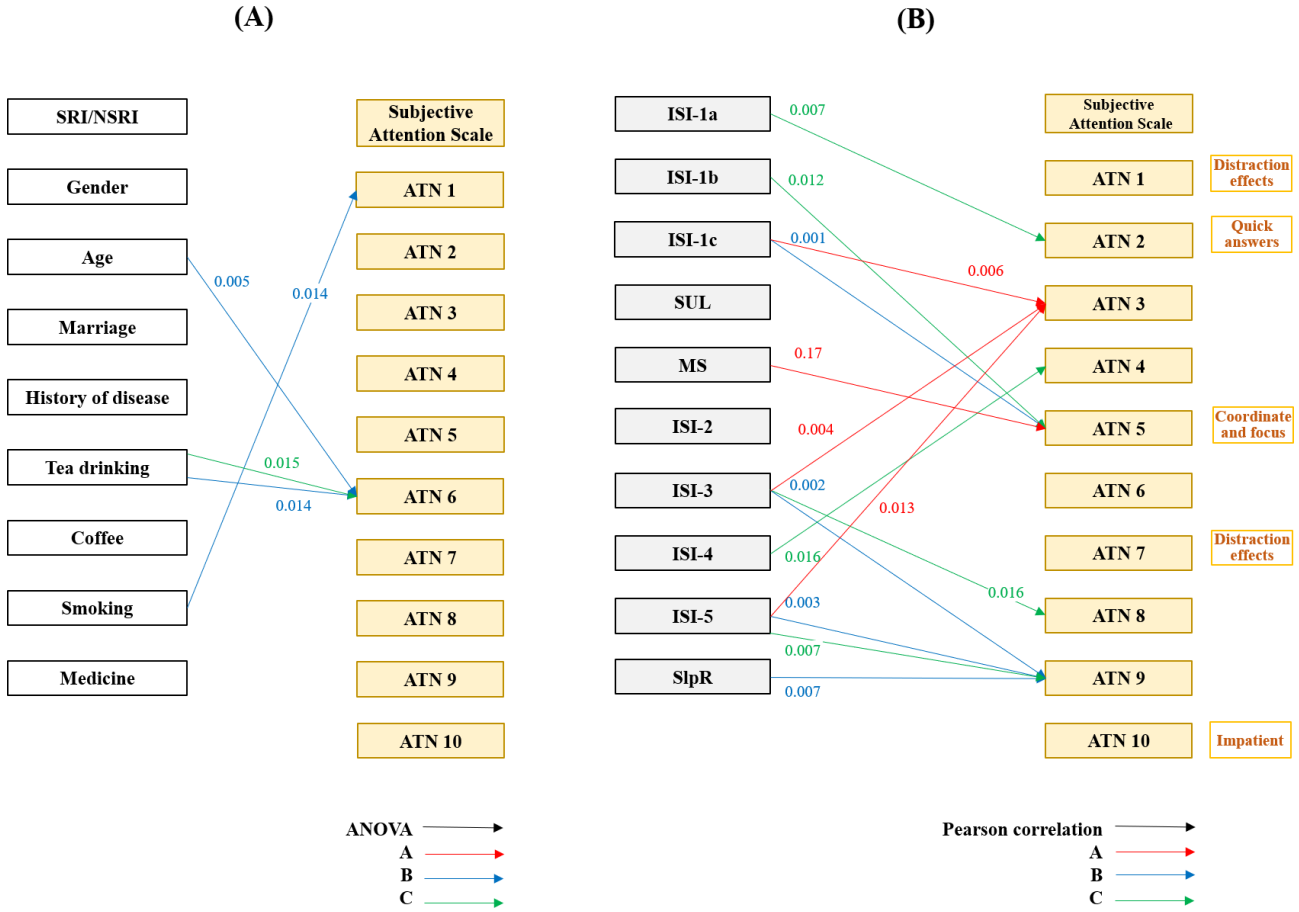
XSL•FO
RenderX

Subjective Attentional Control (ATTC)

The Kruskal-Wallis *H* test showed that age was significantly associated with ATN6 (ATTC item number 6) on Interface B (*H*=10.643, *P*=.005). Among lifestyle factors, smoking was significantly associated with ATN1 on Interface B (*H*=8.585, *P*=.01). In addition, tea drinking was significantly associated with ATN6 on both Interface B (*H*=10.641, *P*=.01) and Interface C (*H*=10.519, *P*=.02).

Associations were observed between sleep-related factors, insomnia symptoms, and specific subjective attention (ATN) domains across interfaces. Sleep regularity was positively associated with ATN9 (distraction effects) on Interface B (*r*=0.460, *P*=.007). Regarding sleep initiation difficulties, ISI-1a was positively associated with ATN2 (quick answers) on Interface C (*r*=0.461, *P*=.007), and ISI-1b was positively associated with ATN5 (coordinate and focus) on Interface C (*r*=0.434, *P*=.01). Early morning awakening (ISI-1c) was positively associated with ATN3 (quick answers) on Interface A (*r*=0.467, *P*=.006) and with ATN5 (coordinate and focus) on Interface B (*r*=0.563, *P*<.001). In contrast, sleep interference with daily functioning (ISI-3) was negatively associated with ATN3 (quick answers) on Interface A (*r*=−0.487, *P*=.004), but positively associated with ATN9 (distraction effects) on Interface B (*r*=0.528, *P*=.002) and ATN8 (distraction effects) on Interface C (*r*=0.417, *P*=.02). ISI-4 showed a positive association with ATN4 (quick answer) on Interface C (*r*=0.417, *P*=.02). Worry about sleep (ISI-5) was negatively associated with ATN3 (quick answers) on Interface A (*r*=−0.426, *P*=.01), but positively associated with ATN9 (distraction effects) on Interface B (*r*=0.505, *P*=.003) and Interface C (*r*=0.463, *P*=.007). In addition, MS was negatively associated with ATN5 (coordinate and focus) on Interface A (*r*=−0.412, *P*=.02; Figure 6).

Figure 6. Associations between (A) population characteristics and (B) insomnia symptoms (ISI) and subjective attention scale across interfaces. ATN: Attention Control Scale item number; ISI: Insomnia Severity Index; MS: mental state; NSRI: no self-reported insomnia; SlpR: sleep regularity; SRI: self-reported insomnia; SUL: stay up late.



Specific ATTC subdomains were selectively associated with demographic and lifestyle factors. Age was related to attention refocusing ability in Interface B, while smoking and tea drinking were associated with distraction-related indices.

Insomnia symptoms demonstrated robust associations with subjective attention. Greater sleep initiation and maintenance difficulties, early morning awakening, and daytime dysfunction were associated with increased distractibility and reduced attentional coordination across interfaces. Worry about sleep was consistently associated with impulsive responding and reduced sustained attention.

MS showed a negative association with attentional coordination, highlighting the interaction between emotional condition and cognitive control during interface use.

Interface Preference and User Experience

Under normal daylight conditions, no significant differences were observed between SRI and NSRI participants in preference

XSL•FO
RenderX

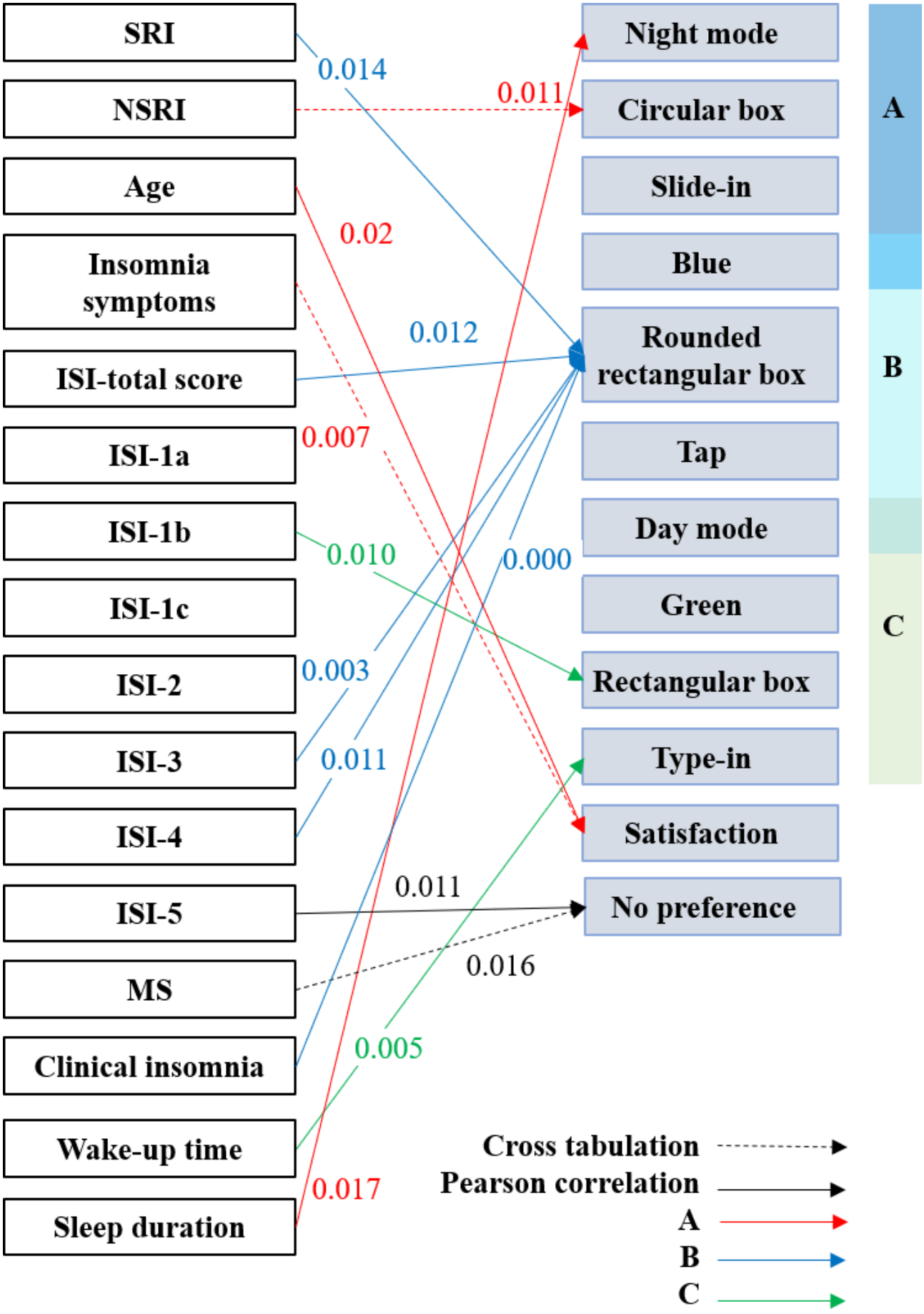
for night versus day mode. NSRI participants showed a greater preference for circular boxes (χ^2_1=6.421, *P*=.01), whereas SRI participants preferred rounded rectangular boxes (*r*=0.424, *P*=.01). Insomnia symptoms participants reported higher satisfaction with Interface A (χ^2_{12}=27.094, *P*=.007), while those with clinical insomnia showed a lower preference for rounded rectangular boxes (*r*=0.593, *P*<.001).

The ISI total score was positively correlated with preference for rounded rectangular box (*r*=0.431, *P*=.01). ISI-1b was associated with rectangular box (*r*=0.440, *P*=.01), whereas ISI-3 and 4 were associated with rounded rectangular boxes (*r*=0.498, *P*=.003; *r*=0.437, *P*=.01). ISI-5 was associated with indifference toward interface interaction design (*r*=0.439, *P*=.01). Wake-up time was negatively correlated with preference for type-in (*r*=−0.479, *P*=.005), and sleep duration was positively correlated with preference for night mode (*r*=0.414, *P*=.02).

Between MS and the interface interaction mode, there was no preference (χ^2_3=10.313, *P*=.02; Figure 7).

XSL•FO
RenderX

Figure 7. Pearson correlation and cross-tabulation analyses of insomnia measures and interface preferences. ISI: Insomnia Severity Index; MS: mental state; NSRI: no self-reported insomnia; SRI: self-reported insomnia.



XSL•FO
RenderX

Using K-means clustering, participants were first divided into 2 clusters based on those without clinically significant insomnia and those with clinical insomnia to examine differences in preferences among Interfaces A, B, and C. Significant between-cluster differences were observed in satisfaction with Interface A ($F_{1,31}$=18.164, P<.001), overall interface satisfaction ($F_{1,31}$=7.691, P=.009), interaction duration ($F_{1,31}$=38.214, P<.001), and consistency of interface interaction ($F_{1,31}$=44.574, P<.001). Day mode also differed significantly between clusters ($F_{1,31}$=10.333, P=.003).

A second K-means analysis divided participants into 4 clusters based on insomnia symptoms. Significant cluster effects were found for satisfaction with Interface A ($F_{3,29}$=4.726, P=.008), Interface B ($F_{3,29}$=11.824, P<.001), and Interface C ($F_{3,29}$=6.519, P=.002). Clusters also differed significantly in interaction duration ($F_{3,29}$=5.800, P=.003), overall interface satisfaction ($F_{3,29}$=15.501, P<.001), and consistency of interface interaction ($F_{3,29}$=36.099, P<.001). In addition, significant differences were observed in preferences for color modes (eg, color mode 1: $F_{3,29}$=3.643, P=.02; color mode 2: $F_{3,29}$=4.233, P=.01) and interaction methods, including slide-in ($F_{3,29}$=3.522, P=.03) and tap input ($F_{3,29}$=5.446, P=.004).

Participants without SRI showed a significantly greater preference for rounded rectangular visual designs and reported higher satisfaction with Interface A. In contrast, participants with clinical insomnia preferred rounded rectangular designs. Sleep characteristics further influenced preferences. Longer sleep duration was associated with preference for night-mode displays. Cluster analyses based on insomnia status and symptom severity revealed significant differences in interface satisfaction, perceived task duration, and preferred interaction methods, with Interface A most frequently rated as the most efficient overall.

Integrated Relationships Between Attention, Eye Tracking, and Memory

Across the 3 interfaces, several subjective attention (ATN) indicators showed significant negative associations with task memory performance during operational tasks. On Interface A, task memory performance was negatively correlated with ATN7 (distraction effects; r=−0.446, P=.009), ATN1 (distraction effects; r=−0.498, P=.003), and ATN10 (impatient; r=−0.456, P=.008), indicating that greater distraction and impatience were associated with poorer task recall. On Interface B, a significant negative association was observed between task memory performance and ATN5 (coordinate and focus; r=−0.410, P=.02), ATN2 (quick answers; r=−0.564, P<.001), and ATN7 (distraction effects; r=−0.477, P=.005), suggesting that reduced attentional coordination, quick answers, and distraction were related to lower recall accuracy. On Interface C, task memory performance was negatively correlated with ATN7 (distraction effects; r=−0.488, P=.004), ATN1 (distraction effects; r=−0.511, P=.002), and ATN5 (coordinate and focus; r=−0.430, P=.01; Table 4).

Table . Associations between subjective attention (ATN[a]) indicators and task memory performance across interfaces.

Interface	ATN indicator (domain)	*r*	*P* value
A	ATN7 (distraction effects)	−0.446	.009[b]
A	ATN1 (distraction effects)	−0.498	.003[b]
A	ATN10 (impatient)	−0.456	.008[b]
B	ATN5 (coordinate and focus)	−0.410	.02
B	ATN2 (quick answers)	−0.564	<.001[b]
B	ATN7 (distraction effects)	−0.477	.005[b]
C	ATN7 (distraction effects)	−0.488	.004[b]
C	ATN1 (distraction effects)	−0.511	.002[b]
C	ATN5 (coordinate and focus)	−0.430	.01

[a]ATN: Attention Control Scale item number.

[b]P<.02.

Overall, higher levels of subjective attentional difficulty (eg, distraction, impatience, reduced coordination) were consistently associated with lower task memory accuracy. Eye-tracking measures supported these findings: faster operation times and more stable saccade patterns were associated with better recall of time-based diary entries.

These convergent results indicate that interface designs that reduce attentional fragmentation also facilitate more accurate short-term memory during sleep diary completion.

Associations were observed between eye-tracking metrics, task memory performance, and subjective attention (ATN) indicators across interfaces. On Interface A, SDir was negatively correlated with SlpOn (r=−0.475, P=.006) and significantly associated with WASO (H_4=14.500, P=.006), indicating that greater directional variability in gaze was related to longer memory-encoding time and nocturnal sleep fragmentation. On Interface B, task 2 was negatively correlated with memory recall duration (r=−0.411, P=.02), suggesting improved task efficiency during secondary operations. With respect to subjective

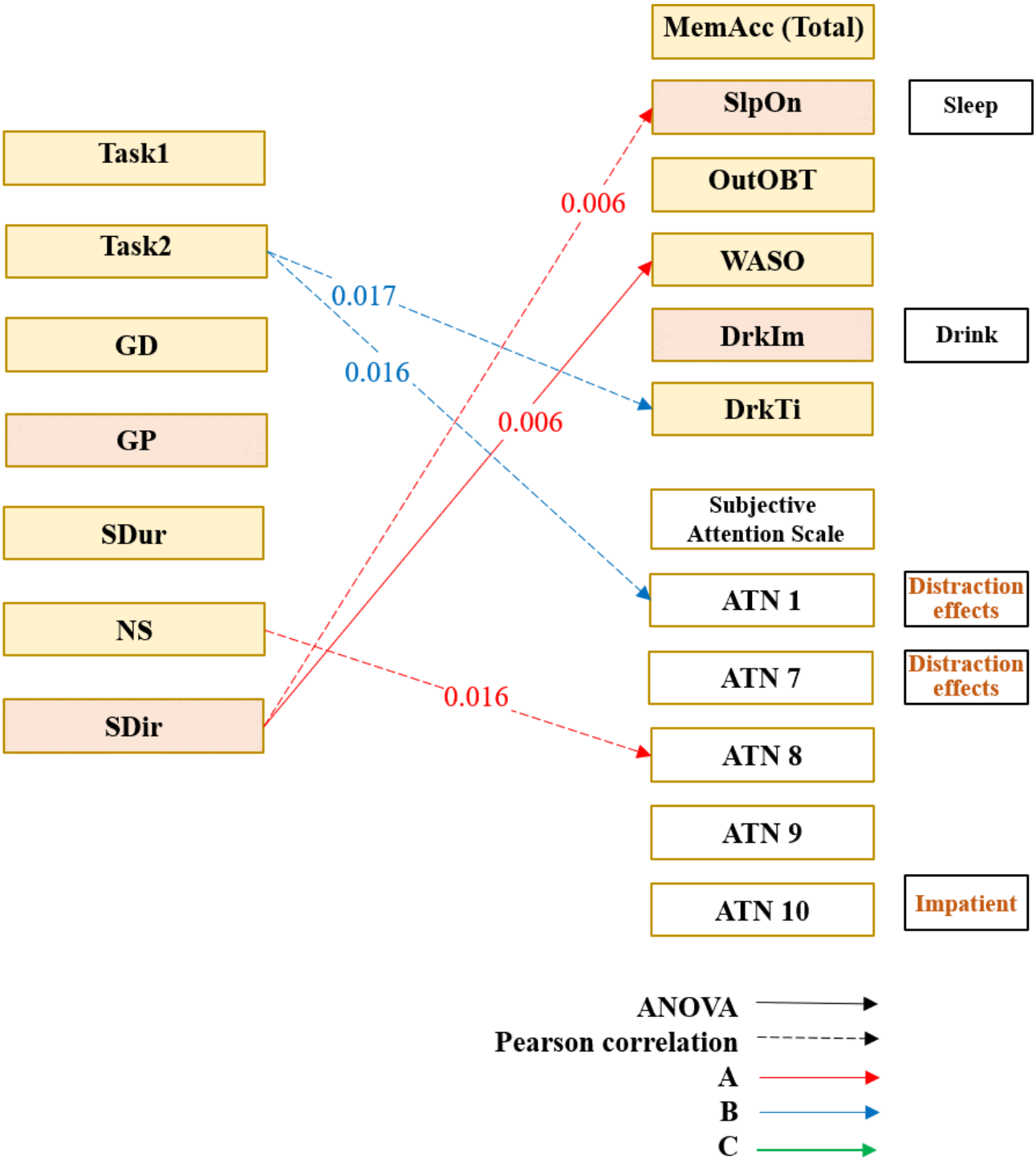
XSL•FO
RenderX

attention, task 2 on Interface B was negatively correlated with ATN1 (distraction effects; r=−0.415, P=.02), whereas on Interface A, the NS was positively correlated with ATN8 (distraction effects; r=0.424, P=.02). Collectively, these findings indicate that interface-specific eye-movement patterns are associated with both cognitive workload during task execution and subjective attentional control, with gaze stability reflecting more efficient memory processing and reduced attentional disruption.

Eye-tracking metrics showed selective associations with memory: greater SDir variability on Interface A was associated with longer sleep-onset recall (r=−0.475, P=.006) and WASO (H_4=14.500, P=.006). Faster second-trial task times on Interface B were associated with improved recall (r=−0.411, P=.02; Figure 8).

Detailed statistical results for all nonparametric and correlation analyses are provided in Multimedia Appendix 1.

Figure 8. Associations between eye-tracking metrics, task memory performance, and subjective attention. ATN: Attention Control Scale item number; DrkIm: drinking items; DrkTi: drinking time; GD: gaze duration; GP: gaze position; MemAcc-Total: total memory accuracy; NS: number of saccades; OutOBT: out-of-bed time; SDir: saccade direction; SDur: saccade duration; SlpOn: sleep onset; WASO: wakefulness after sleep onset.



XSL•FO
RenderX

Discussion

Principal Results

This study investigated how interface design features of a BBT-I sleep diary app influence attention, visual processing, and short-term memory in individuals with varying insomnia severity. By integrating eye-tracking metrics, subjective attention ratings, behavioral memory accuracy, and interface preferences, the study provides a multidimensional understanding of how insomnia-related cognitive vulnerabilities interact with digital interface design.

Interface design was associated with differences in objective attentional performance, as reflected in task completion time, saccade count, and gaze stability. Second, SlpQ and insomnia symptoms—particularly sleep maintenance difficulty and sleep-related worry—were consistently associated with impaired attention and reduced memory accuracy. Third, short-term memory performance varied across interface designs and was especially sensitive to sleep disruption during event-based recall. Fourth, subjective attentional control showed strong convergence with both eye-tracking indicators and memory outcomes. Finally, interface preferences differed by insomnia status, with participants without clinically significant insomnia favoring visually softer, lower-load designs.

Interface Design and Attentional Load

Differences in eye-movement patterns across interfaces indicate that layout structure and interaction modality directly shape attentional demands. Interfaces requiring more visual scanning or manual input elicited higher saccade activity and longer task durations, particularly among participants with poorer SlpQ. In contrast, designs characterized by simpler layouts and guided interactions were associated with more stable gaze patterns and reduced attentional dispersion. These findings support cognitive load theory, suggesting that insomnia amplifies sensitivity to interface complexity and visual fragmentation.

Insomnia Symptoms and Memory Vulnerability

Memory accuracy was particularly vulnerable to sleep maintenance difficulties and daytime dysfunction, especially for event-based diary items such as exercise and nap timing. Routine sleep information (eg, bedtime and wake time) was less affected. This selective impairment aligns with prior evidence that insomnia disproportionately affects episodic and temporally ordered recall. Interface designs that increase cognitive effort during data entry may therefore exacerbate memory errors in users with insomnia.

Subjective Attention as a Key Predictor

One of the most robust findings was that subjective attentional difficulties—such as distraction, impatience, and reduced coordination—were stronger predictors of memory accuracy than eye-tracking measures. While eye-tracking captured physiological effort and scanning behavior, subjective attention better reflected the internal cognitive state governing successful recall. This suggests that memory recall during sleep diary completion is primarily an internally driven cognitive process, rather than one dependent on overt visual behavior.

Limitations

The following limitations should be considered when interpreting the results of this study:

1. Our findings represented regional population characteristics. The participants were of Asian descent between the ages of 20 and 64. Their lifestyle habits and insomnia conditions may have been influenced by factors such as age, occupation, and region of residence.
2. One limitation of the present study was the difficulty in recruiting participants with insomnia, which led to an underrepresentation of individuals with severe insomnia. Future studies with larger samples are warranted to enhance statistical power and generalizability.
3. The sleep diary interface of the app used in our study was based on the log recording method of CBT-I, and we only investigated user experience as well as the preference, attention, and memory status of participants. The therapeutic effect of our interfaces on insomnia was not within the scope of our study.
4. The mobile device used was running iOS. Other mobile application systems would require additional designs.

Conclusions

This study demonstrates that interface design plays a critical role in shaping attention, visual processing, and short-term memory performance in users with insomnia. Insomnia symptoms—particularly sleep maintenance difficulty and sleep-related worry—were consistently associated with attentional disruption and reduced memory accuracy during sleep diary completion. Notably, subjective attentional control was a stronger and more reliable predictor of memory performance than eye-tracking measures, indicating that cognitive processing during recall is driven primarily by internal attentional states rather than observable gaze behavior. For designers of digital insomnia interventions, these findings highlight the need to reduce cognitive load through simple, structured layouts; prioritize low-effort input methods such as tapping over typing; support sequential recall with guided, step-by-step data entry; and use visually soft, predictable designs that minimize distraction. Interfaces should also be adaptable to insomnia severity, offering more guided and low-complexity interactions for users with greater sleep disruption. By aligning interface design with the cognitive characteristics of insomnia, digital BBT-I applications may improve usability, data accuracy, and user engagement.

XSL•FO
RenderX

Funding

This study was supported by the Taipei University System Academic Cooperation Project and represents a collaborative effort between the National Taipei University of Technology and Taipei Medical University, under project number USTP-NTUT-TMU-114-01.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to security purposes requested by the academic institution but are available from the corresponding author on reasonable request.

Authors' Contributions

KCS and HYC contributed equally to the study conception and design, data analysis, and manuscript drafting. KCS additionally led the coordination and integration of the manuscript. KCW provided methodological guidance and supervision throughout the study. CCC contributed to interface design development and data interpretation. All authors critically reviewed the manuscript and approved the final version for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Classification of variables by name and type.
[PDF File, 118 KB - humanfactors_v13i1e79883_app1.pdf]

Multimedia Appendix 2
All statistical results.
[PDF File, 185 KB - humanfactors_v13i1e79883_app2.pdf]

Checklist 1
CONSORT 2010 checklist.
[PDF File, 200 KB - humanfactors_v13i1e79883_app3.pdf]

References

1. The impact of sleep on learning and memory. Scribd. URL: https://www.scribd.com/document/682852013/The-Impact-of-Sleep-on-Leaming-and-Memory-Chronobiology-and-Sleep-Institute-Perelman-School-of-Medicine-at-the-University-of-Pennsylvania [accessed 2026-01-29]
2. Terán-Pérez GJ, Ruiz-Contreras AE, González-Robles RO, et al. Sleep deprivation affects working memory in low but not in high complexity for the N-Back test. Neurosci Med 2012;03(4):380-386. [doi: 10.4236/nm.2012.34047]
3. Klinzing JG, Niethard N, Born J. Mechanisms of systems memory consolidation during sleep. Nat Neurosci 2019 Oct;22(10):1598-1610. [doi: 10.1038/s41593-019-0467-3] [Medline: 31451802]
4. Brownlow JA, Miller KE, Gehrman PR. Insomnia and cognitive performance. Sleep Med Clin 2020 Mar;15(1):71-76. [doi: 10.1016/j.jsmc.2019.10.002] [Medline: 32005351]
5. Wardle-Pinkston S, Slavish DC, Taylor DJ. Insomnia and cognitive performance: a systematic review and meta-analysis. Sleep Med Rev 2019 Dec;48:101205. [doi: 10.1016/j.smrv.2019.07.008] [Medline: 31522135]
6. Acosta MT. Sleep, memory and learning [Article in Spanish]. Medicina (B Aires) 2019;79 Suppl 3:29-32. [Medline: 31603840]
7. Edinger JD, Arnedt JT, Bertisch SM, et al. Behavioral and psychological treatments for chronic insomnia disorder in adults: an American Academy of Sleep Medicine clinical practice guideline. J Clin Sleep Med 2021 Feb 1;17(2):255-262. [doi: 10.5664/jcsm.8986] [Medline: 33164742]
8. Sato D, Sutoh C, Seki Y, Nagai E, Shimizu E. Treatment preferences for internet-based cognitive behavioral therapy for insomnia in Japan: online survey. JMIR Form Res 2019 May 15;3(2):e12635. [doi: 10.2196/12635] [Medline: 31094319]
9. Li SH, Graham BM, Werner-Seidler A. Gender differences in adolescent sleep disturbance and treatment response to smartphone app–delivered cognitive behavioral therapy for insomnia: exploratory study. JMIR Form Res 2021 Mar 23;5(3):e22498. [doi: 10.2196/22498] [Medline: 33755029]
10. Sato D, Yoshinaga N, Nagai E, Nagai K, Shimizu E. Effectiveness of internet-delivered computerized cognitive behavioral therapy for patients with insomnia who remain symptomatic following pharmacotherapy: randomized controlled exploratory trial. J Med Internet Res 2019 Apr 11;21(4):e12686. [doi: 10.2196/12686] [Medline: 30973344]
11. Mair A, Scott H, Lack L. Intensive sleep retraining treatment for insomnia administered by smartphone in the home: an uncontrolled pilot study. J Clin Sleep Med 2022 Jun 1;18(6):1515-1522. [doi: 10.5664/jcsm.9892] [Medline: 35641889]

XSL•FO
RenderX

12. Pulantara IW, Parmanto B, Germain A. Development of a just-in-time adaptive mHealth intervention for insomnia: usability study. JMIR Hum Factors 2018 May 17;5(2):e21. [doi: 10.2196/humanfactors.8905] [Medline: 29773529]
13. Bramoweth AD, Lederer LG, Youk AO, Germain A, Chinman MJ. Brief behavioral treatment for insomnia vs. cognitive behavioral therapy for insomnia: results of a randomized noninferiority clinical trial among veterans. Behav Ther 2020 Jul;51(4):535-547. [doi: 10.1016/j.beth.2020.02.002] [Medline: 32586428]
14. Liu X, Li Y, Yan R, et al. The platform development, adherence and efficacy to a digital brief therapy for insomnia (dBTI) during the COVID-19 pandemic. Methods 2022 Sep;205:39-45. [doi: 10.1016/j.ymeth.2022.04.016] [Medline: 35526723]
15. Suh SA, Jeon H, Ong J. Diagnostic tools for insomnia. In: Kushida CA, editor. Encyclopedia of Sleep and Circadian Rhythms, 2nd edition: Academic Press; 2023:140-148. [doi: 10.1016/B978-0-12-822963-7.00015-3]
16. Xu J, Liu S, Fu X. Applying working memory theory to redesign a mobile application user interface: take a handicraft self-learning page as an example. Presented at: 6th International Conference, Design, User Experience, and Usability: Designing Pleasurable Experiences (DUXU 2017), held as part of HCI International 2017; Jul 9-14, 2017. [doi: 10.1007/978-3-319-58637-3_26]
17. Ma C, Wang H, Xue C. Research on the relationship between the aesthetics of interface elements layout and visual working memory. In: HCI International 2022 - Late Breaking Papers Design, User Experience and Interaction:597-607. [doi: 10.1007/978-3-031-17615-9_42]
18. Pedersen LA, Einarsson SS, Rikheim FA, Sandnes FE. User interfaces in dark mode during daytime – improved productivity or just cool-looking? 2020 Presented at: 14th International Conference on Universal Access in Human-Computer Interaction Design Approaches and Supporting Technologies; Jul 19-24, 2022. [doi: 10.1007/978-3-030-49282-3_13]
19. Tian P, Xu G, Han C, et al. Effects of paradigm color and screen brightness on visual fatigue in light environment of night based on eye tracker and EEG acquisition equipment. Sensors (Basel) 2022 May 27;22(11):4082. [doi: 10.3390/s22114082] [Medline: 35684700]
20. Clarke T, Costall A. The emotional connotations of color: a qualitative investigation. Color Res Appl 2008 Oct;33(5):406-410. [doi: 10.1002/col.20435]
21. Hussain AR. Colour psychology in art: how colour impacts mood. Art Des Rev 2021;09(4):301-308. [doi: 10.4236/adr.2021.94025]
22. Wang J, Hsu Y. The relationship of symmetry, complexity, and shape in mobile interface aesthetics, from an emotional perspective—a case study of the smartwatch. Symmetry (Basel) 2020;12(9):1403. [doi: 10.3390/sym12091403]
23. Zhou C, Wang Z, Jiang Z. Interactive interface design for telemedicine and the emotional needs of patients. Presented at: 2021 16th International Conference on Computer Science & Education (ICCSE); Aug 19-21, 2021 p. 554-559. [doi: 10.1109/ICCSE51940.2021.9569545]
24. Kirsch W, Kunde W. The size of attentional focus modulates the perception of object location. Vision Res 2021 Feb;179:1-8. [doi: 10.1016/j.visres.2020.11.004] [Medline: 33264688]
25. Zhuo S, Oh J. Rounding for warmth, angling for fluency: how shapes, cognitive load, and privacy risk influence app adoption. Int J Hum Comput Interact 2025 Jun 18;41(12):7202-7217. [doi: 10.1080/10447318.2024.2390755]
26. Zhang Q, Liu Y. Smart user experience medical app interface design based on mobile devices. Expert Syst 2022;39(5):e12808. [doi: 10.1111/exsy.12808]
27. Bhanap R, Oberauer K, Rosner A. The role of eye movements and covert shifts of attention in working memory retrieval. PsyArXiv. Preprint posted online on Sep 12, 2024. [doi: 10.31234/osf.io/5d73r]
28. Volden F, Alwis Edirisinghe V, Fostervold KI. Human gaze-parameters as an indicator of mental workload. Presented at: Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018); Aug 26-30, 2018. [doi: 10.1007/978-3-319-96059-3_23]
29. Wegner S, Lohmeyer Q, Wahlen D, Neumann S, Groebli JC, Meboldt M. Value of eye-tracking data for classification of information processing-intensive handling tasks: quasi-experimental study on cognition and user interface design. JMIR Hum Factors 2020 Jun 3;7(2):e15581. [doi: 10.2196/15581] [Medline: 32490840]
30. Park B, Flowerday T, Brünken R. Cognitive and affective effects of seductive details in multimedia learning. Comput Human Behav 2015 Mar;44:267-278. [doi: 10.1016/j.chb.2014.10.061]
31. Park S, Lee SW, Han S, Cha M. Clustering insomnia patterns by data from wearable devices: algorithm development and validation study. JMIR mHealth uHealth 2019 Dec 5;7(12):e14473. [doi: 10.2196/14473] [Medline: 31804187]
32. Stark L, Brünken R, Park B. Emotional text design in multimedia learning: a mixed-methods study using eye tracking. Comput Educ 2018 May;120:185-196. [doi: 10.1016/j.compedu.2018.02.003]
33. Bartlett FC. Remembering: A Study in Experimental and Social Psychology: Cambridge University Press; 1995. [doi: 10.1017/CBO9780511759185]
34. Lin YH, Chiang TW, Lin YL. Increased internet searches for insomnia as an indicator of global mental health during the COVID-19 pandemic: multinational longitudinal study. J Med Internet Res 2020 Sep 21;22(9):e22181. [doi: 10.2196/22181] [Medline: 32924951]
35. Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001 Jul;2(4):297-307. [doi: 10.1016/s1389-9457(00)00065-4] [Medline: 11438246]

XSL•FO
RenderX

Abbreviations

ATN: Attention Control Scale item number
ATTC: Attention Control Scale
BBT-I: brief cognitive behavioral therapy for insomnia
CBT-I: cognitive behavioral therapy for insomnia
CONSORT: Consolidated Standards of Reporting Trials
DrkTi: drinking time
ExcsIm: exercise items
ISI: Insomnia Severity Index
MS: mental state
NS: number of saccades
NSRI: no self-reported insomnia
SDir: saccade direction
SlpOn: sleep onset
SlpQ: sleep quality
SRI: self-reported insomnia
WASO: wake after sleep onset

Edited by A Kushniruk; submitted 02.Jul.2025; peer-reviewed by A Azizi, D Sato; accepted 18.Jan.2026; published 19.Feb.2026.

Please cite as:
Su KC, Chiu HY, Wu KC, Chang CC
Designing App Interfaces to Elicit Specific Emotional Responses and Improve Attention and Short-Term Memory in Patients With Insomnia Undergoing Brief Cognitive Behavioral Therapy: Within-Subject Eye-Tracking Experimental Pilot Study
JMIR Hum Factors 2026;13:e79883
URL: https://humanfactors.jmir.org/2026/1/e79883
doi:10.2196/79883

© Kuan-Chu Su, Hsiao-Yean Chiu, Ko-Chiu Wu, Chia-Chi Chang. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 19.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Reflections on the Inclusive Co-Design Process of a Virtual Assistant for Individuals With Complex Care Needs: Mixed Methods Study

Julia F E van Calis[1,2], MSc; Kirsten E Bevelander[1,2], PhD; Anneke W C van der Cruijsen[1,2]; Jenneken Naaldenberg[1,2], PhD; Geraline L Leusink[1,2], MD, MBA, PhD

[1]Department of Primary and Community Care, Radboud University Medical Center, PO Box 9101, Nijmegen, The Netherlands
[2]Academic Collaborative Intellectual Disability and Health - Sterker op Eigen Benen (SOEB), Nijmegen, The Netherlands

Corresponding Author:
Julia F E van Calis, MSc
Department of Primary and Community Care, Radboud University Medical Center, PO Box 9101, Nijmegen, The Netherlands

Abstract

Background: The digitalization of society has transformed daily life and health care, offering opportunities for accessibility and independence for individuals with complex care needs. However, users with limited digital skills still experience challenges because the technologies do not to align with their needs. Inclusive research and design approaches can improve technology by actively involving end users and stakeholders.

Objective: This study investigated the experiences of co-researchers with a mild intellectual disability or autism spectrum disorder and other key stakeholders over time regarding the inclusive design process for a digital tool for individuals with complex care needs that was developed in a transdisciplinary consortium.

Methods: The project that was examined applied an inclusive design process to develop a sensitive virtual assistant using the Vision in Product Design method and the design thinking approach. Nine consortium members, including 3 co-researchers, participated in semistructured interviews and a group discussion about the inclusive design process after each of the project's 5 work packages (WPs). This resulted in 31 interviews and 5 group discussions in total. Individual experiences were gathered during interviews, and group discussions facilitated collective reflection. During the interviews, an adapted questionnaire was used for each WP with Likert scales and open-ended questions. The data analysis was conducted using a thematic approach and descriptive statistics for the questionnaire data.

Results: Quantitative findings from questionnaires were complemented with qualitative insights from interviews and group discussions, with results presented chronologically per WP. The qualitative analysis resulted in 3 main themes: project approach, collaborative dynamics, and co-design in practice. Project approach showed how the team adapted its inclusive collaboration through expectation management, structured processes, and accessible materials. Collaborative dynamics described how communication and support evolved and how inclusive design principles were applied in practice. Co-design in practice outlined co-researcher involvement and content adaptations across the 5 WPs, highlighting how experiential knowledge directly informed design decisions. These findings show that inclusive collaboration developed over time and contributed meaningfully to both process and content.

Conclusions: This study shows that, to accommodate an inclusive research and design process, tensions between project efficiency and meaningful inclusion need to be addressed, underlining the importance of continuous coordination, collaboration, and flexibility in transdisciplinary settings. Further, applying a stepwise approach in inclusive collaborations supports coordination, continuous evaluation, and flexibility. Inclusive methods, like preparatory activities, clear role division, accessible materials, and iterative feedback, enabled active co-researcher participation. These methods contributed to a shift in ownership, allowing co-researchers to gain greater influence and co-shape both the development process and the content. The findings provide insights into how to enhance equity and relevance in inclusive technology design for individuals with complex care needs, such as individuals with a mild intellectual disability or autism spectrum disorder.

(JMIR Hum Factors 2026;13:e81941) doi:10.2196/81941

KEYWORDS

mild intellectual disabilities; autism spectrum disorder; inclusive design; citizen science; inclusive research; sensitive virtual assistant

XSL•FO
RenderX

Introduction

The rapid digitalization of society has transformed the way in which individuals interact with technology in health care and in their daily lives [1,2]. As digital innovations become more integrated within society, it is essential that they meet the needs of diverse user groups such as those requiring complex care. These individuals often rely on a wide range of health and social support services, making accessibility to digital health tools particularly relevant. However, such tools can be difficult to use and less accessible for individuals with limited digital skills. For example, individuals with a mild intellectual disability (MID) or autism spectrum disorder (ASD) often experience difficulties acquiring digital literacy skills and using digital devices or the internet [3,4] because of limited cognitive skills and challenges with processing information efficiently [5]. Both groups frequently struggle with understanding complex instructions, problem-solving, and adapting to new technologies [5,6]. This can place them in a vulnerable position, as it limits their ability to fully participate in today's complex society [7,8]. Digital tools designed to fit their living situation may reduce challenges, increase their empowerment, and contribute to more effective care, ultimately decreasing health disparities in these populations [9,10].

Although digital health tools can enhance accessibility and independence, they are often not developed to align with these users' specific needs and skills [2,11,12]. Despite the growing inclusion of these target groups in technology research and design projects [11,13,14], the participation of end users with MID and ASD remains limited [15-17]. This gap highlights the need for inclusive research and design approaches that incorporate the perspectives of these end users in the development of health technologies [18-20]. In this study, inclusive design is defined as an approach to make the design and development process accessible, equitable, and participatory for individuals with diverse abilities and experiences. Drawing from the Vision in Product Design (ViP) methodology [21] and design thinking [22], inclusive design refers to a flexible, iterative process that values multiple forms of expertise, including experiential knowledge, and adapts its methods, materials, and facilitation strategies to ensure meaningful involvement of all participants.

User-centered approaches [23] place user groups at the center of the design process by using tailored methods to address their specific needs [24,25]. These approaches follow an iterative process, applying rapid prototyping and prioritizing end user experiences by collaborating within interdisciplinary teams [26]. Inclusive, flexible, and adaptive approaches have been shown to be well-suited to involving people without prior experience in technological design [11,14]. Recent work in inclusive participatory design further demonstrates how design processes can be adapted to respect diverse sensory, cognitive, and communicative modes of engagement [27-30]. These studies provide important methodological insights into how participation can evolve into equitable collaboration, ensuring accessibility and involvement across all stages of technological development. In research, inclusive approaches such as inclusive research or citizen science actively involve end users in all phases of technological research and development processes. In these approaches, end users are experts-by-experience and co-researchers [31,32] who serve as advisors, cooperators, or nonprofessional researchers alongside those with academic or design training to add value through the sharing of their lived experiences [13,33]. User-centered and inclusive research and design approaches can be applied to develop (health) care technologies that are not only functional but also accessible and usable for all [14,34]. However, no research has been conducted into how end users with limited digital skills experience collaboration in technological research and development processes or into the adaptations that are necessary to involve them.

In addition to the participation of co-researchers and experts-by-experience, the involvement of other stakeholders in technological research and development is essential [20,26]. Examples of these stakeholders include (informal) caregivers, support workers, IT developers, or other specialists to align with the end user context and support in practice [35]. This gives technologies a greater chance of successful implementation and use [18,36,37]. However, previous studies have found that less attention is given to involving key stakeholders and end users throughout all design iterations or to the continuous monitoring of the design process during implementation [20,26]. Reflecting on the collaboration during technology development and implementation phases can provide practical insights valuable for (future) development [26]. Therefore, this study aimed to examine how inclusive design methods facilitate the development of digital health tools, focusing on the experiences of co-researchers with MID or ASD and other stakeholders over time on the inclusive design process of a digital tool for individuals with complex care needs, developed within a transdisciplinary consortium. Rather than assessing the design outcomes of the digital tool itself, the research focused on the process-level learning, that is, the iterative adaptations, facilitation strategies, and collaborative dynamics that enabled or hindered inclusive participation over time. The research question guiding the analysis was: *What process modifications and content adjustments were made over time to accommodate an inclusive design process with co-researchers with MID or ASD in the development and implementation of a digital tool for individuals with complex care needs?*

Methods

Study Setting

Within this study, the inclusive design process for a digital tool in the form of a sensitive virtual assistant (SVA) for individuals with complex care needs was investigated. A consortium of transdisciplinary partners aimed to develop an SVA that could guide individuals with complex care needs toward appropriate support for questions related to mental and physical care [10]. Artificial intelligence technology was used to enable the SVA to adapt its responses in a sensitive manner.

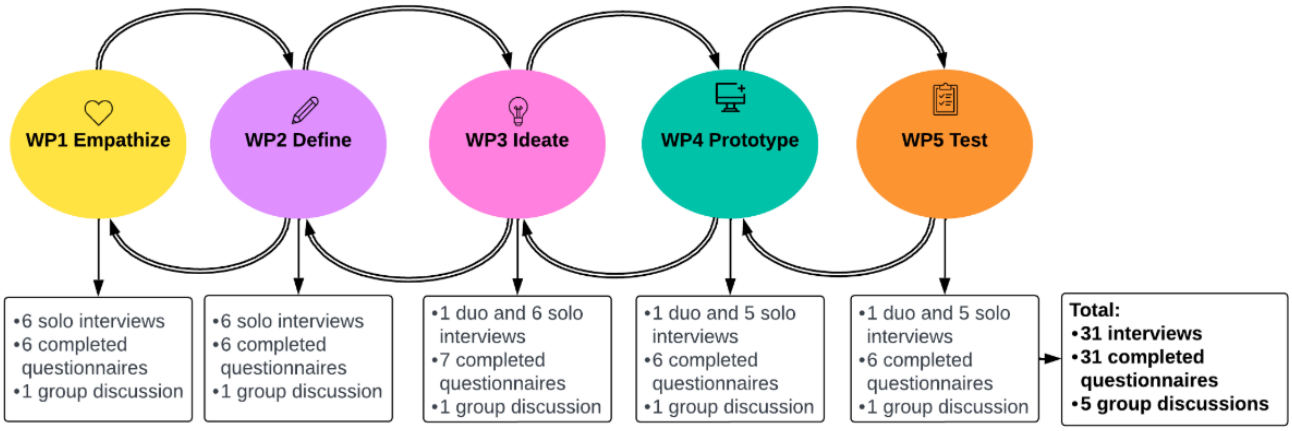
The consortium consisted of 5 partners: (1) a university department with expertise in technology and chatbots in service settings; (2) an academic collaborative conducting research with and about people with intellectual disabilities; (3) a social design

XSL•FO
RenderX

agency; (4) a foundation for social care organization informing, advising, supporting, and employing people with complex care needs such as individuals with ASD and intellectual disability; and (5) an IT company. In this project, 3 individuals with ASD or MID, representing the target group, participated as co-researchers and experts-by-experience throughout the entire research and design process. The co-researchers work for the academic collaborative and the social care organization. Their involvement enabled the consortium to consider the challenges underlying the complexity of different living and care situations. Their roles varied across project phases such as sharing lived experiences, co-designing prototypes, recruiting participants, co-moderating focus groups, and preparing data collection materials. By combining inclusive research and user-centered design, the project aimed to ensure meaningful participation.

The SVA was developed using the ViP method [21], incorporating the design thinking approach [22]. Each design thinking phase was operationalized as a separate work package (WP1-WP5), aligned with the steps of the ViP method (Figure 1) [22]. In the Empathize phase (WP1), the focus was on identifying and mapping the needs of the end users through focus groups and a literature review. The Define phase (WP2) analyzed the insights from WP1 to develop a user profile, which included user needs and desired experiences. In the Ideate phase (WP3), a wide range of creative solutions was generated. Among these activities, personas were created, representing fictional user profiles designed to represent typical users' behaviors, motivations, and needs. Additionally, the development of user scenarios and user journeys helped define functional requirements and typical user experiences. In the Prototype phase (WP4), these ideas were turned into early-stage functional designs (ie, low-fidelity prototypes) and tested for usability and performance. This phase also included making the ethical approval application for user testing. Finally, the Test phase (WP5) involved developing 3 functional versions of the SVA, each with a different interaction style. These were tested with real users in actual care settings to evaluate and refine the solution.

Figure 1. Flowchart of the sensitive virtual assistant (SVA) project and data collection for the reflection on the inclusive design approach.



Study Design

This study used a mixed methods approach to reflect iteratively with co-researchers and consortium members on the SVA's inclusive design process over time; see Figure 1. The SVA functioned as a context and illustrative case that allowed the research team to analyze inclusive processes in practice rather than to evaluate the final design outcome. Qualitative data were gathered through (1) semistructured interviews with co-researchers and key stakeholders followed by (2) group discussions with co-researchers at the end of the 5 phases or respective WPs. During the group discussions, the co-researchers reflected together on the insights from the interviews. Quantitative data were collected during the semistructured interviews by filling in a digital questionnaire adapted for each WP. The integration of qualitative and quantitative methods ensured structural reflection over time across the entire process with all participants. Further, this study was conducted in collaboration with a co-researcher with an MID, whose perspective was actively integrated into the research process.

Participants

Representatives from all 5 consortium partners participated in the study, with a total of 9 participants, including the 3 co-researchers. The other 6 participants were consortium partners involved in various roles: researchers, social designers, project leader, and technology developers. Table 1 shows the participants' distribution in the interviews and questionnaires. The first 3 group discussions took place with 3 co-researchers, and the last 2 discussions took place with 2 co-researchers.

XSL•FO
RenderX

Table . Participant distribution in the interviews.

Total participants per work package (WP)	Co-researchers, n	Researchers, n	Social designers, n	Project leader, n	Technology developers, n
WP1 (n=6)	3	1	1	1	0
WP2 (n=6)	3	1	1	1	0
WP3 (n=8)	3	1	1	1	2
WP4 (n=7)	2	1	1	1	2
WP5 (n=7)	2	1	2	1	1

Ethical Considerations

Primary data collection was conducted in compliance with the current data protection regulations of the General Data Protection Regulation and the Declaration of Helsinki. Each participant received an information letter and provided informed consent via a comprehensible consent form. Both documents were co-created with the co-researcher with MID to enhance accessibility (eg, by incorporating simple text, graphic support, large font, and increased white space). Data collection was anonymous, and no conclusions can be drawn about individuals in the presentation of results. Since the participants were part of the consortium and were compensated for their input as members, they were not compensated separately for this study. The entire study was submitted for ethical review to the Research Ethics Committee CMO Radboudumc and waived from further ethical assessment (reference number 2023 - 16400) because it does not constitute medical scientific research according to the Medical Research Involving Human Subjects Act (WMO) [38]. This waiver is consistent with national and institutional policies that specify that research falling outside the scope of the WMO does not require further review. Relevant guidance can be found in the CCMO framework for non-WMO research (CCMO, *Niet-WMO-onderzoek*)[39], as well as the CMO Radboudumc policy on non-WMO assessment[40].

Procedure and Data Collection

The aim of this study was to identify key adaptations in both the collaborative process and the content of inclusive research practices over time (May 2023–May 2024). After each WP was completed, semistructured interviews were conducted with the consortium members, including the co-researchers. These semistructured interviews focused on participants' engagement, lessons learned, and overall experiences and were conducted with 6 to 8 participants per WP. As the interviews were conducted after each phase, experiences were still recent and therefore easier to retrieve. During the interviews, quantitative data were collected using digital questionnaires adapted for each WP. The interviewer filled in the quantitative responses using *I Co-research* (*IkOnderzoekMee*; Crowdience), an inclusive digital research platform for accessible dissemination of questionnaires [14,41]. The questionnaires consisted of quantitative elements using Likert-scale items (see Table 2), incorporating simplified text, graphic support, large fonts, and increased white space to facilitate the participants [14]. Examples of how the scales were used in the questionnaires can be found in Multimedia Appendix 1. During the interviews, open-ended questions were asked regarding specific activities of the WP and the clarity of roles and tasks within these activities, experiences within the collaboration in the consortium, goals of the WPs, lessons learned, and individual reflections. The open-ended questions provided a deeper understanding of how adaptations were implemented and experienced. Multimedia Appendix 1 provides the interview guide combining the interview questions and a WP1 questionnaire. Following the interviews with questionnaires, a group discussion was held with the 3 co-researchers involved in the project. During these discussions, insights, and areas of interest from the interviews and questionnaires such as participant engagement, challenges, and overall perceptions were discussed. An independent researcher (JFEC) moderated these sessions. In line with the paradigm "nothing about us without us" of the international treaty from the United Nations Convention on the Rights of Persons with Disabilities [42], the data collection process was designed to ensure that the voices of co-researchers with MID or ASD were central. At the same time, consortium members without disabilities were included to investigate how inclusive collaboration was experienced and facilitated across roles. This comprehensive inclusion allowed us to identify how the participatory process was shaped. Importantly, although all perspectives were collected, the reflections of co-researchers in the group discussions formed the primary reference point for interpreting and discussing results in later stages of the analysis.

Table . Scales used in the questionnaire.

Type of scale	Color	Points	Distribution (lowest to highest)
Visual analog scale	Black and white	10	1=Totally not clear, 10=Very clear; 1=Totally not nice, 10=Very nice
Stars	Yellow	5	1=Very negative, 5=Very positive; 1=Little contribution, 5=Great contribution
Smileys	Green, yellow, orange, and red	4	1=Very good, 4=Bad
Thumbs	Green, yellow, and red	3	1=Yes, 2=Partially, 3=No

XSL•FO
RenderX

Data Analysis

Descriptive statistics were applied to analyze the quantitative data from the digital questionnaires. Microsoft Office Excel was used to calculate frequencies, median scores, and range distributions, allowing for a clear overview of participants' response patterns, medians, and range distributions.

All interviews and group discussions were recorded and transcribed for analysis. Table 3 presents the consecutive steps of the qualitative data analysis, which was conducted in multiple phases using a thematic approach. After each WP, the transcripts of the semistructured interviews were rapidly coded [43,44] to gain an initial impression of participant engagement, challenges, and overall perceptions. The results from this phase were included in the subsequent group discussions with the co-researchers. During these discussions, key themes based on the research questions were identified, forming an initial framework to guide the analysis of the semistructured interview data of the corresponding WP. This framework guided further rapid coding of the interview and discussion data to identify overarching themes and subthemes [43-45]. The analysis process was conducted separately for each WP, ensuring a structured and iterative evaluation of the data.

In addition to the thematic analysis of interview and group discussion data, a structured analysis of co-researcher involvement and activities per design phase was conducted (Table 3, step 8). This analysis, based on data extracted from the interviews, focus groups, and a logbook, resulted in a table summarizing the activities conducted within each WP, the content adaptations made, and the involvement of co-researchers in these activities, supported by participants' experiences.

Table . Consecutive steps, actions, and aims of the coding process.

Step	Action	Aim
1	Reading transcripts of interviews	Become familiar with data and sections of the transcripts
2	Initial rapid coding (JFEC, AWCC) of the semistructured interviews and questionnaire responses after each work package (WP)	Gain first impression of interview data and explore areas of interest such as participant engagement, challenges, and overall perceptions
3	Using first impression on which to reflect in the subsequent group discussions with the 3 involved co-researchers (JFEC, AWCC)	Identify key themes and reflections based on research question and create initial coding framework with thematic categories
4	Applying initial coding framework and using rapid coding approach to guide second analysis	Conduct bottom-up coding, identify overlap and connections between codes, and ensure collective reflection
5	Clustering of codes into broader themes relating to the research question (JFEC, AWCC, KEB)	Identify emerging and overarching themes and subthemes
6	Discussing clusters of codes using insights from group discussions (JFEC, AWCC, KEB, JN)	Refine subthemes and analyze their relationships within each main theme, creating the final coding structure
7	Repeating steps 1-6 for each WP separately	Generate an overview of insights per WP
8	Collectively analyzing WPs	Develop a comprehensive overview of the full process and conduct qualitative data analysis, which resulted in 3 main themes: project approach, collaborative dynamics, and co-design in practice

Results

Themes

The thematic analysis resulted in 3 main themes that together describe how inclusive collaboration was shaped and experienced throughout the design process. The thematic coding map of the inclusive design process is visualized in Figure 2. The first 2 themes show how the team adapted methods, facilitation, and collaboration over time. The subthemes are first represented by quantitative results from the questionnaires then further illustrated by qualitative results with quotes from the interviews and group discussions. The last theme gives an overview of the core activities, co-researcher involvement, and content adaptations to the inclusive research and design process during each WP illustrated by participants' quotes about experiences. The results are described chronologically structured by the 5 WPs. Because the SVA functioned as a use case, the emphasis here is on process adaptations rather than design evaluation of the tool itself. Where relevant, the results indicate whether adaptations were particularly beneficial for the co-researchers with MID or ASD or more generally valuable for all consortium members.

All quantitative outcomes have been integrated into a single comprehensive table. Table 4 presents these results collectively, offering a consolidated overview of the rating ranges and median scores across all WPs.

XSL•FO
RenderX

Figure 2. Thematic coding map of the inclusive design process.

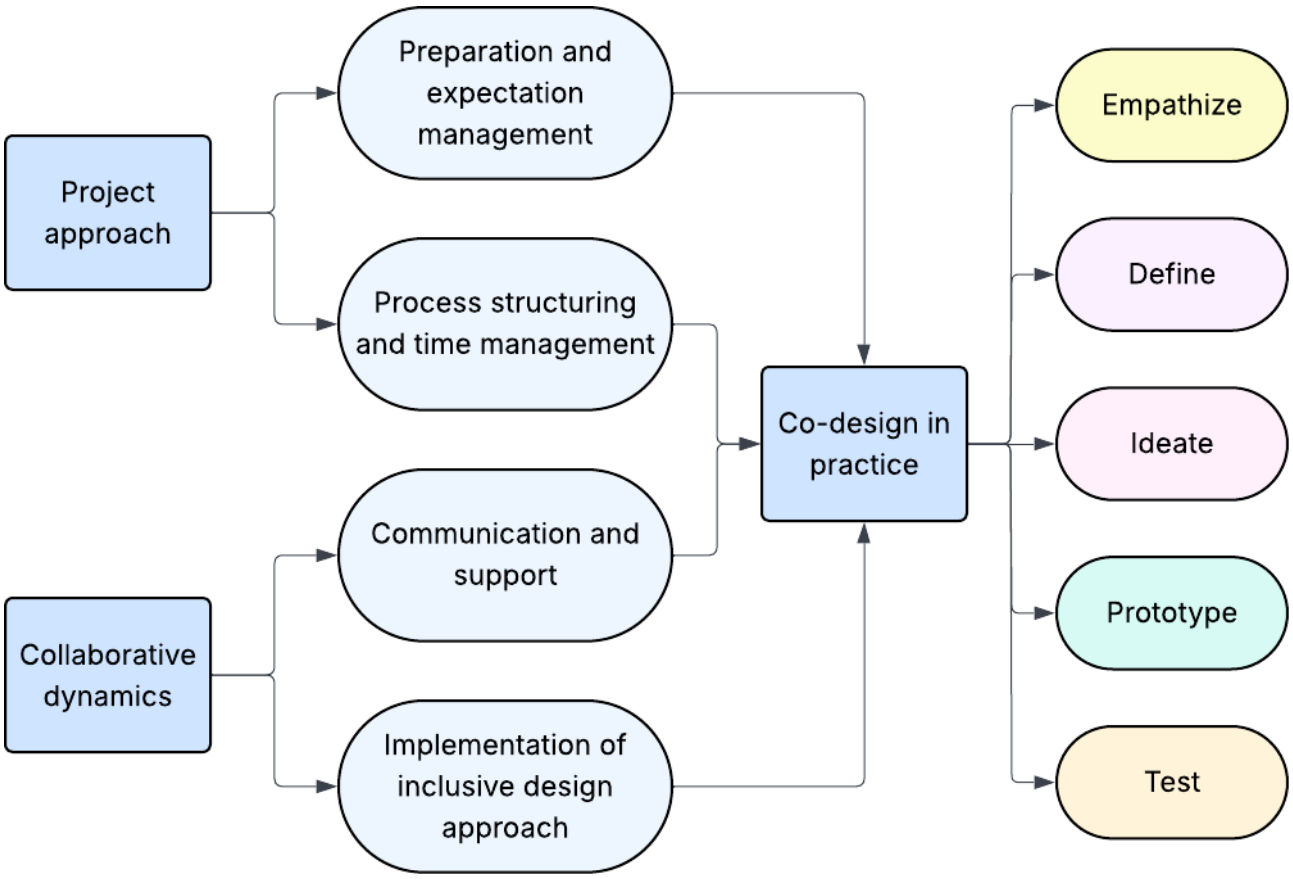


Table . Quantitative outcomes of the grading of all work packages (WPs).

Grading for	Scale	WP1 grade (n=6)		WP2 grade (n=6)		WP3 grade (n=7)		WP4 grade (n=6)		WP5 grade (n=6)	
		Range	Median	Range	Median	Range	Median	Range	Median	Range	Median
Contributions from the co-researchers to achieving the goals of the WP	1 - 5 stars (Low to High)	2 - 3	3	3 - 4	3	3 - 4	3.5	2 - 5	4	3 - 5	4.5
Clarity of tasks within each WP	Visual analog scale 1 - 10 (Low to High)	5 - 8	7.5	2 - 10	6.5	2 - 8	7	3 - 10	7	5 - 9	7.5
Overall experience of the collaboration within each WP	Visual analog scale 1 - 10 (Low to High)	3 - 5	4	3 - 5	4	3 - 5	4	3 - 5	4	3 - 5	4.5

Project Approach

Three subthemes describe how the team adapted the project approach to facilitate inclusive collaboration: preparation and expectation management, process structuring and time management, and the accessibility and clarity of materials and meetings.

Preparation and Expectation Management

Table 4 shows the grading for the co-researchers' contributions to achieving WP's goals. The median ratings went from low to

XSL•FO
RenderX

medium to high across the WPs, with a widening grade range particularly in WP4, suggesting both improvements as well as ongoing difficulties in the contributions in achieving the goals of each WP.

In WP1, considerable time was spent preparing the project and designing the research approach. This led to a strong methodological foundation by combining the ViP approach with clearly defined research questions aimed at developing the SVA in a grounded and well-substantiated way. This preparatory phase was experienced as valuable by the participants: "A lot of work before the start, especially substantiating the fundamental research. That intensive pre-application phase is paying off now" [Group discussion WP1, co-researcher 1]. Although these efforts were experienced as beneficial, participants emphasized that role clarity and mutual expectations were not sufficiently addressed in this phase. Co-researchers noted that roles and meeting purposes were not always clear, thereby hindering inclusive participation: "The division of roles and their descriptions should be clearer in advance, and the purpose behind a meeting should also be better explained" [Group discussion WP1, co-researcher 3].

In WP2, the co-researchers' role was more clearly defined, and preparatory work was focused on aligning tasks with their expertise and ensuring that they were adequately informed to contribute meaningfully to the research activities. The participants stated that, in WP3, issues identified in WP1 and WP2 were addressed, such as more structure and explicit planning, leading to process improvements. In WP4, expectations were more explicitly defined, as indicated by co-researchers. Preparations for WP5 were also experienced as well-structured. However, uncertainty about project progress remained a recurring point, especially regarding coordination with external partners. Despite the high workload and challenges, WP5 was experienced as successfully completed. Nevertheless, reflections on the full project indicated that expectation management on roles and expectations should have been addressed earlier. Clarifying roles and meeting purposes proved beneficial for all team members, while the clearer explanation of individual tasks and meeting goals was particularly important for the co-researchers to understand their expected contribution.

Process Structuring and Time Management

Process structuring and time management varied across the WPs. In WP1, a co-researcher was present at all focus groups acting as a second chairperson, contributing to continuity. In WP2, the team revisited earlier steps when clarification was needed and repeated certain elements of the methodology; this was experienced as supportive and necessary: "More attention was paid to repeating and explaining the steps, so planning those extra sessions was really helpful" [Interview WP2, consortium member 1]. These repetitions and slower pacing were adaptations that specifically supported the co-researchers by providing predictability and reinforcing understanding, while improved scheduling and process visibility benefited the entire consortium.

Co-researchers reported that certain steps in the process had been skipped or insufficiently addressed. They felt they had not been adequately informed, which led to uncertainty about their role in specific sessions. This was especially apparent in WP2, where feedback loops discussing and revisiting important aspects of the care setting were not applied consistently over time, leading to confusion, a feeling of unpreparedness, and missing structure among some co-researchers: "I didn't get enough explanation about what was expected of me, so it felt like some steps were missing and I wasn't well prepared. It felt like the structure was missing a bit" [Interview WP2, co-researcher 3]. In WP3, the design process was structured using personas and value scenarios. These tools helped visualize user needs and supported decision-making. The use of visual tools was initially introduced to make abstract information more accessible for the co-researchers but later proved equally valuable for all team members to align perspectives and trace decisions. However, not all choices made during the design process were clearly traceable to earlier input, consequently raising questions about the origin of decisions: "It's still unclear where some decisions and information come from, but I do have confidence in the process and what it will lead to" [Group discussion WP3, co-researcher 2]. In WP5, the fast pace of data collection limited opportunities for deeper engagement, particularly among co-researchers not involved in day-to-day activities. This moderate level of involvement indicates the tension between project deadlines and inclusive participation.

Accessibility of Meetings and Clarity of Materials

The low to high gradings indicated varying experiences with the clarity of the tasks within each WP (Table 4). These variations are reflected in both improvements and persistent challenges in the clarity of materials and meetings.

Throughout the project, accessibility of documents and the clarity of meeting structures were important for structuring inclusive collaboration. In WP1, the importance of aligning language and terminology across disciplines was raised. One co-researcher explained how this led to misunderstandings: "There was a moment I got stuck on the terminology, some explanations were missing, so I misinterpreted terms like 'advisory group' and focus group'" [Interview WP1, co-researcher 1].

In WP2, the team experienced difficulties with access to shared documentation because of unclear file structures and inconsistent uploading of materials. Despite this, valuable insights were still generated through collaborative analysis sessions. In addition, some participants found the abstract language used by the social designer difficult to follow, particularly during the initial explanation of the design concepts and framework. This suggests that the issue was not limited to participants with MID or ASD. Sessions about the design requirements were considered too complex or intensive, thereby negatively impacting participation. There was a clear need for better guidance and coordination: "That session was much harder than I expected. I underestimated the approach, reading all the design requirements aloud was too much for 1 co-researcher, especially with all 3 co-researchers in 1 session. Better preparation would have helped" [Interview WP2, consortium member 3]. Simplifying terminology and using concrete examples primarily supported the co-researchers, whereas improvements in file organization, consistent upload

XSL•FO
RenderX

practices, and meeting agendas were generally beneficial for all participants.

During WP3, difficulties in managing large amounts of information were expressed, particularly regarding working with personas. This emphasized the need for better preparation and briefing. Collaboration in smaller groups improved clarity, and strategic selection of who attended meetings helped structure discussions more effectively: "It was helpful that we could build on earlier experiences and realized that sometimes it's better to meet in smaller groups and have a clear agenda with a focused question for the co-researchers" [Interview WP3, consortium member 4]. This adaptation was initially designed to meet the communication preferences of the co-researchers but was later appreciated by the entire team for improving efficiency and focus. In WP4, the inconsistent involvement of co-researchers in different sessions revealed the need for a more transparent selection and planning process; this was subsequently addressed by clarifying the tasks and communicating planning and agendas more openly and in advance.

Collaborative Dynamics

This theme describes how the team experienced and shaped collaboration throughout the project. We identified 2 subthemes: (1) communication and support and (2) the implementation of the inclusive design approach.

Communication and Support

Over the course of the project, the importance of structured communication and tailored support became more evident, with improvements throughout all phases. In WP1, the involvement of a researcher with experience working with the target group contributed significantly to inclusive interaction and understanding. This early investment in a collaborative dynamic created a foundation that contributed to more effective coordination and inclusive dialogue in later phases, although some team members indicated that communication within the broader group was occasionally fragmented in WP2 and WP3. The team became more aware of the need for regular status updates and accessible coordination, particularly for co-researchers. These structured updates were introduced to help co-researchers stay connected and informed. In WP3, this learning resulted in more targeted communication and the introduction of structured feedback sessions. One co-researcher reflected: "I sometimes miss status updates, and, as a result, I feel less connected to the group and the project" [Group discussion WP3, co-researcher 3].

From WP4 onwards, lessons from earlier phases were applied. Project roles became more clearly defined, and coordination across team members and with co-researchers improved. For example, responsibilities for preparing test sessions were allocated in advance, and co-researchers were actively involved in reviewing prototypes, leading to smoother collaboration and more targeted feedback. In WP5, external factors such as delays in receiving feedback from the tests posed challenges, partly because of scheduling difficulties and the time required for internal approvals. The team members emphasized the value of having expertise with inclusive approaches embedded within the team from the start: "Looking back, I now see how important it is to have knowledge and experience with the target group. This must be secured in advance" [Interview WP5, consortium member 2]. Overall, the experiences with communication and support illustrate a learning trajectory in which the team increasingly recognized and applied the conditions required for sustainable inclusive collaboration. The growing emphasis on inclusive communication was therefore both a targeted accessibility measure for the co-researchers and a general facilitator of cohesive teamwork.

Implementation of an Inclusive Design Approach

Consistent gradings were present for the overall experience of the collaboration within each WP (Table 4), indicating that this was experienced as fairly stable.

In WP1, the team paid attention to aligning tasks with co-researchers' strengths and availability. These early decisions, such as involving co-researchers in defining project goals and choosing methods, helped create a foundation for shared ownership and equitable participation. This was reflected in how co-researchers took the initiative during planning meetings and felt empowered to provide critical feedback on design choices. In WP2, several team members and co-researchers described stronger group dynamics and increased peer support during meetings, suggesting a growing sense of cohesion. The team became more responsive to individual needs and made deliberate choices about when and how to involve co-researchers. This thoughtful approach contributed to a respectful and adaptive working environment: "There is a conscious decision whether or not to involve specific consortium members or co-researchers in certain activities. Their added value or potential overload is always considered carefully" [Group discussion WP2, co-researcher 2]. Adjusting task complexity and workload was particularly beneficial for the co-researchers, supporting autonomy and confidence in their role.

In WP3, the inclusive approach was further strengthened. Co-researchers suggested setting up a test group; this marked a shift toward more proactive engagement. Their suggestion was adopted and built upon in WP4 and WP5, where co-researchers played active roles in testing and presenting outcomes to various stakeholders, including care and support organizations, technology developers and intended target users. These contributions were experienced as motivating and valuable: "For me, it was a real highlight that I got to join and answer questions at an event. It made me feel even more part of the team" [Interview WP5, co-researcher 1]. Throughout the project, the team also gained insight into how different forms of experiential input could be used effectively. The distinction between experiential expertise (derived from personal lived experience) and experiential knowledge (the ability to reflect on and articulate that experience for others) became more visible, leading to more intentional use of individual contributions. For example, during user testing phases, some co-researchers contributed best by evaluating content based on their personal experience, whereas others translated those insights into concrete design suggestions. This initiative specifically empowered the co-researchers, reinforcing their

XSL•FO
RenderX

agency within the design process while also fostering a stronger sense of shared ownership of the team.

Co-Design in Practice

A summary of the core activities within the design phases, co-researcher involvement, and content adaptations to the inclusive design process made across the 5 different WPs is provided in Table 5. The adaptations were informed by the input from and collaboration with co-researchers. The following section elaborates on the rationale behind these adaptations resulting from this input, and reflects on participants' experiences with the content adaptations made.

XSL•FO
RenderX

Table . Summary of core activities, co-researcher involvement levels, and content adaptations.

Phase, core activities, and level of co-researcher involvement	Description of co-researcher involvement	Content adaptations to inclusive research practices
Empathize (map the end users' needs): conducted focus groups to gather insights from end users and performed a literature review to identify user challenges and context		
High	Co-researchers were actively engaged throughout all phases of the focus group (preparation, recruitment, data collection, and evaluation).	Themes for the focus groups were refined collaboratively with co-researchers to ensure inclusiveness by representing diverse care perspectives.
Limited	In the literature review, co-researchers were kept informed, contributed to discussing the research question, and reviewed summaries.	The literature review content was contextualized using insights raised by co-researchers during earlier discussions.
Define: (analyze insights from Empathize phase to create a user profile of the intended end users with the needs and desired experiences): identified and clustered contextual factors into core care intentions and developed an interest framework to capture users' underlying motivations		
Low	Co-researchers contributed to reviewing the contextual factors but were not involved in clustering them.	Terminology and categories were adjusted in response to co-researcher feedback to improve clarity and accessibility for all consortium members.
Moderate	Reasonably active contributions in developing the interest framework were supported through peer conversations and guided sessions.	Interest profiles (summaries of key user motivations) and design missions (goals guiding the design process) were co-developed with co-researchers and refined based on their feedback to ensure alignment with users' real needs.
Ideate (generate a wide range of creative solutions): created user scenarios based on complex care experiences, mapped user journeys and defined functional needs and designed personas (fictional profiles representing typical users), and formed test groups to explore concepts collaboratively		
High	The co-researchers co-created scenarios.	Scenarios were modified to reflect lived experiences and concrete examples provided by co-researchers.
Moderate	Feedback was provided by the co-researchers during sessions on the mapped journey and functional requirements.	Personas and requirements were refined based on feedback from the co-researchers to ensure alignment with real-world user experiences. Tools were used to visualize the input of the co-researchers.
High	Test groups were formed and initiated by a co-researcher.	Recruitment effort via the consortium network resulted in diverse groups, including individuals with lived experience and professionals.
Prototype (create functional designs and explore them through performance and usability tests): created and tested low-fidelity prototypes (simplified early versions) of the sensitive virtual assistant and prepared and submitted an ethical approval application for user testing		
High	The co-researchers actively interpreted and responded to user feedback.	Prototype elements, such as navigation, instructions, and tone of voice, were adapted by the co-researchers during user feedback sessions.
Moderate	A co-researcher supported the drafting of the ethical application to test the prototypes.	Ethical approval documents, such as the information and informed consent letter, were edited in consultation with a co-researcher.
Test (evaluate functional designs with real users to refine and improve the solution): developed 3 functional sensitive virtual assistant versions with different interaction styles and evaluated these versions in real care environments with end users		
Moderate	Co-researchers contributed insights during evaluation sessions, which informed improvements to the sensitive virtual assistant.	Adjustments to interaction styles (eg, tone, responsiveness) in the 3 tool versions were explicitly informed by earlier co-researcher feedback regarding communication preferences.
Moderate	Co-researchers played a key role in guiding the evaluation, interpreting user feedback, and expressing a strong sense of ownership in the process.	Adjustments were made to enhance usability, clarity of instructions, and emotional accessibility based on real-world feedback from participants and co-researchers.

In WP1, the focus group themes and literature review content were collaboratively refined with co-researchers, enabling diverse care perspectives to shape the discussions and ensure inclusive representation and contextual relevance. One consortium member reflected that they unintentionally excluded co-researchers from the later stages of the literature review, as they still saw it as a task that they had to carry out individually: "I did discuss with the co-researchers what would be a relevant

XSL•FO
RenderX

research question, but after that, I unconsciously no longer really involved them because I still had the idea that this was something I had to do myself" [Interview WP1, consortium member 1].

WP2 showed that structured feedback rounds were used to discuss and cluster contextual factors. The process and individual roles were initially unclear to some co-researchers, which led to adaptations to improve task explanation and structure for greater accessibility: "My task was not explained well. This made it unclear what my specific contribution was. We discussed this, and it was clearer afterwards" [Interview WP2, co-researcher 2]. The iterative co-development of an interest framework was supported by facilitation from the researcher; these sessions involved a process of adaptations: "It was very helpful that the postdoc supported the translation of the interest framework so that it could be properly presented, even though it took some trial and error" [Interview WP2, consortium member 3]. Terminology, categories, interest frameworks, and design missions were adapted based on co-researcher input to enhance clarity and alignment with user perspectives.

The scenarios and personas were informed by the co-researchers' concrete input; their contributions reflected real-life situations and ensured the inclusion of diverse user experiences. Tools, like an interactive online board where team members could brainstorm and collaborate, enhanced clarity and enabled their lived experiences to be visibly integrated: "There was a strong sense of contribution; you can see that on the board with all the input we have generated" [Group discussion WP3, co-researcher 3]. One co-researcher illustrated the challenges, caused by strict ethical requirements, regarding the planning of test groups: "I found setting up this plan challenging, especially because of the strict requirements of the ethics committee" [Interview WP3, co-researcher 1].

In WP4, prototype elements and ethical documents were co-developed with co-researchers to support usability, comprehensibility, and accessible communication. Co-researchers' clear understanding of their role, as exemplified during feedback sessions, appeared to support their engagement and involvement in the testing process, as suggested by their proactive participation and comments on ownership: "When testing, I knew exactly what to do, that worked very well" [Group discussion WP3, co-researcher 2]. However, ethical procedures caused frustrations among the team, with 1 co-researcher describing his frustrations: "Meeting all these requirements was time-consuming and sometimes frustrating" [Interview WP4, co-researcher 1].

In WP5, interaction styles and instructional clarity were refined based on participant and co-researcher feedback to improve responsiveness, accessibility, and emotional engagement. The initial design of the functional tool versions was shaped more by prior input than by direct co-creation in this phase, contributing to stress and frustration when delays in deployment occurred: "I really had no idea if it would work; that was stressful" [Interview WP5, co-researcher 1]. Despite this, the testing phase was positively evaluated because of adequate preparation: "My tasks during the testing were clear because the postdoc had provided good explanations in advance" [Interview WP5, co-researcher 2].

Together, the adaptations and accompanying reflections demonstrate that co-researcher involvement shaped not only the content but also the form and accessibility of the research process. They highlight how continuous feedback and flexible collaboration were essential to aligning the design with real-world needs. Overall, accessibility-focused adaptations such as simplified language, smaller group sessions, and pacing adjustments were particularly supportive of the co-researchers, whereas measures like role clarification, structured planning, and consistent documentation were more generally beneficial for all consortium members.

Discussion

General Findings

This study reflects on an inclusive and iterative design approach involving co-researchers with MID or ASD in the development of a digital health tool, examining how inclusive design methods facilitated the development of the tool and shaped collaboration within the consortium. Through adaptations, such as thorough preparation, role clarity, accessible communication, and iterative feedback, co-researchers made meaningful contributions to the inclusive research and design process. Three overarching themes, project approach, collaborative dynamics, and co-design in practice, provided insights into how inclusive collaboration was accommodated throughout different design phases. The inclusive collaboration was supported by structured engagement, cognitive accessibility strategies, and iterative methods, which collectively facilitated a greater sense of ownership and active involvement over time.

This study showed that the use of a flexible, inclusive framework deepened stakeholder engagement and enabled concrete content adaptations through its stepwise approach, in this case through design thinking [22,46]. However, design thinking is not the only framework suitable for inclusive development. Inclusive design frameworks such as design for the whole population by Clarkson et al [47], universal design by Goldsmith [48], and user-sensitive inclusive design by Newell et al [25], offer complementary approaches that emphasize accessibility, user diversity, and context-sensitive solutions. In addition, complementary frameworks such as the CeHRes (Center for eHealth Research) Roadmap [26] and the NASSS (Non-adoption, Abandonment, Scale-up, Spread, and Sustainability) framework [49] are other valuable approaches emphasizing stakeholder alignment and long-term implementation contexts, respectively. The ViP method used to develop the digital tool draws from these traditions and emphasizes collaborative approaches, shared ownership, and flexibility across project phases [21]. Involving co-researchers in the design process made it possible for them to actively influence both the collaborative workflow and the development of the tool itself. This participatory approach not only shaped the structure of the design meetings and the inclusivity of the design materials but also enhanced the value and the applicability of the outcomes [32,50]. This study contributes experiential and theoretical insights into how inclusive design

XSL•FO
RenderX

methods can be operationalized within transdisciplinary (health care) technology development.

Within the flexible and iterative approach, the tension between project efficiency and meaningful inclusion emerged as an ongoing challenge, as it required increased time, support, and adaptive capacity from the team. This tension has also been shown in previous studies, particularly in the context of complex design environments where timelines are tight and deliverables fixed [51,52]. Future research could therefore investigate how inclusive design practices affect collaborative dynamics and decision-making processes in project environments such as planning or iterative technology development. This would help clarify how to balance inclusivity with time constraints and resource limitations and offer practical strategies for managing this tension [31,51]. Understanding the value of shared expectations [53] and how teams navigate this can inform inclusive design processes that aim for meaningful inclusion throughout all phases of research and design [32].

Further, a shift toward co-researcher ownership emerged in the collaboration. In the earlier phases of the project, co-researchers were involved primarily in predefined tasks, whereas, in later phases, they initiated concrete activities such as forming user test groups. This illustrates how lived experience can transition from consulting to initiating in the design process [54]. The active involvement of co-researchers enhanced both the usability and the real-world relevance of the tool. This finding aligns with previous research suggesting that long-term, repeated engagement with end users allows them to influence the design process more deeply and ensures outcomes that better reflect their needs [50,55]. Cognitive accessibility, through visual aids, simplified materials, repetition, and small group formats, was particularly important in early phases and supported the meaningful contributions from the co-researchers [56,57]. This relates to prior literature on accessible formats [52] and dedicated support time [56] in inclusive research and technology development. Future research could further explore how to embed these inclusive practices in technology development projects with interdisciplinary teams and the involvement of co-researchers while maintaining methodological integrity and meaningful participant roles [13,31,32].

Strengths and Limitations

A key strength of this study is its design, which included structured reflection moments after each of the 5 design phases. This approach enabled the participants to share their experiences immediately after each phase while also identifying changes and adaptations across the full design trajectory. Another strength lies in the mixed methods approach, which combined qualitative interviews and group discussions with tailored quantitative questionnaires executed within an inclusive digital platform [14]. Using both experiential insights and continuous group reflections allowed the team to adapt the collaboration dynamically. As a result, the co-researchers could engage more meaningfully, which led to deeper, more context-specific findings [54]. Some limitations must be acknowledged. The study was conducted within a single consortium, working on 1 digital (health) tool in a specific context. Additionally, the relatively small sample size may limit the diversity of perspectives captured and the broader applicability of the results [58]. Therefore, future research should examine other and larger settings.

Recommendations

Drawing from the findings of this study, a set of practical recommendations for researchers and designers aiming to conduct inclusive design with individuals with complex care needs, such as individuals with MID or ASD [59,60], was formulated. These recommendations translate the experiential insights of the inclusive design process into actionable guidance for future research and practice:

- *Engage end users early and continuously*. Involve individuals with MID or ASD from the start of the project and maintain their involvement across all design phases. Early participation fosters ownership, while continuous engagement ensures that adaptations remain relevant as the project evolves.
- *Adapt sessions to fit participants' abilities and needs*. Prepare accessible materials and ensure that meetings are structured, predictable, and easy to follow. Use clear and jargon-free language, concrete examples, and visual supports. Naming, visuals, and text should be recognizable and understandable.
- *Build long-term and trusting relationships*. Sustainable collaboration requires time to establish mutual trust. Create a safe and comfortable setting where participants can meaningfully shape the design process. Especially in early stages, consider including trusted stakeholders (eg, family members or support workers) to facilitate familiarity and confidence.
- *Address power dynamics and clarify roles*. Discuss expectations, roles, and decision-making responsibilities early and explicitly. Move beyond consultation by showing participants how their input has influenced decisions. Continuous reflection and transparent communication strengthen mutual respect and shared accountability.
- *Use methods close to lived experience*. Apply participatory methods that connect to everyday contexts, such as contextual interviews, observations with discussion, cognitive walk-throughs, or think-aloud sessions. Conducting co-creation activities in familiar environments ensures that insights reflect participants' real-life experiences and capacities.
- *Monitor group dynamics*. Be attentive to how participants experience group settings. Some individuals with MID or ASD may find large groups overwhelming. Adjust group size and facilitation to participants' comfort levels and communication styles.

Conclusions

Reflection on an inclusive research and design process revealed that attention must be paid to tensions between project efficiency and meaningful inclusion, underlining the importance of ongoing coordination of the design process, collaboration, and flexibility in transdisciplinary settings. A stepwise approach, such as design thinking, proved helpful in structuring the process, continuous evaluation, and reducing these tensions. Inclusive methods, such as preparatory work, clear roles, accessible materials, and

XSL•FO
RenderX

iterative feedback, supported active engagement throughout the project. These methods led to a shift in ownership, allowing co-researchers to gain more influence and actively shape both the development process and the content of the technology. These findings extend understanding of how inclusive design methods facilitate development processes and provide insights into how to enhance equity and relevance in inclusive technology design for individuals with complex care needs, such as individuals with MID or ASD.

Acknowledgments

We thank the team for participating in the interviews with questionnaires and group discussions.

Funding

This study was funded by the Ministry of Health, Welfare, and Sport, the Netherlands (reference: 325418) and ZonMw (reference: 427003016). The funding sources had no role in the study design, data collection, data analysis, data interpretation, or writing of this report.

Data Availability

Anonymized data is available within the Radboud Data Repository upon reasonable request (eg, for research purposes on a similar topic).

Authors' Contributions

JFEC, KEB, AWCC, JN, and GLL designed the study. JFEC collected the data. JFEC and AWCC conducted the analysis. JFEC and KEB drafted the manuscript. AWCC, KEB, JN, and GLL provided critical feedback, and all authors provided critical inputs on multiple iterations. All authors have approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide and questionnaire Work Package.
[DOCX File, 201 KB - humanfactors_v13i1e81941_app1.docx]

References

1. Clifford Simplican S, Shivers C, Chen J, Leader G. “With a Touch of a Button”: staff perceptions on integrating technology in an Irish service provider for people with intellectual disabilities. J Appl Res Intellect Disabil 2018 Jan;31(1):e130-e139. [doi: 10.1111/jar.12350] [Medline: 28332313]
2. van Tuyl L, Keuper J, Standaar L, Alblas E, Keij B. Toegankelijkheid van digitale zorg: signalen uit de e-healthmonitor 2023. TSG Tijdschr Gezondheidswet 2025;103(S1):3-7. [doi: 10.1007/s12508-025-00451-7]
3. Schouten DGM, Deneka AA, Theune M, Neerincx MA, Cremers AHM. An embodied conversational agent coach to support societal participation learning by low-literate users. Univ Access Inf Soc 2023 Nov;22(4):1215-1241. [doi: 10.1007/s10209-021-00865-5]
4. Wright J. Technology in social care: a review of the UK policy landscape. sustainable care paper 2. : University of Sheffield; 2020 URL: https://drive.google.com/file/d/1xn2Gfs-cDsBXDAuIPTjiic02NfsTaUvh/view?pli=1 [accessed 2026-02-11]
5. Murphy E, Shiels O, Fiori S, et al. Bridging the digital divide for individuals with intellectual disabilities: implications for well‐being and inclusion. Brit J Learn Disabil 2025 Mar;53(1):17-31 [FREE Full text] [doi: 10.1111/bld.12613]
6. Chadwick D, Richards C, Molin M, Strnadová I. Digital inclusion and people with learning disabilities. Brit J Learn Disabil 2023 Jun;51(2):119-124 [FREE Full text] [doi: 10.1111/bld.12530]
7. Hamblin KA. Technology in care systems: displacing, reshaping, reinstating or degrading roles? New Technol Work Employ 2022 Mar;37(1):41-58. [doi: 10.1111/ntwe.12229] [Medline: 35911255]
8. Dekelver J, Daems J, Solberg S, Bosch N, Van de Perre L, De Vliegher A. Viamigo: a digital travel assistant for people with intellectual disabilities: modeling and design using contemporary intelligent technologies as a support for independent traveling of people with intellectual disabilities. : Institute of Electrical and Electronics Engineers; 2015 Presented at: 2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA); Corfu, Greece. [doi: 10.1109/IISA.2015.7388014]
9. Mistry SK, Shaw M, Raffan F, et al. Inequity in access and delivery of virtual care interventions: a scoping review. Int J Environ Res Public Health 2022 Aug 1;19(15):9411. [doi: 10.3390/ijerph19159411] [Medline: 35954768]
10. Kraaijeveld SR, van Heijster H, Bol N, Bevelander KE. The ethics of using virtual assistants to help people in vulnerable positions access care. J Med Ethics 2025 Dec 15;52(1):26-31. [doi: 10.1136/jme-2024-110464] [Medline: 39961668]

XSL•FO
RenderX

11. Lussier-Desrochers D, Normand CL, Romero-Torres A, et al. Bridging the digital divide for people with intellectual disability. CP 2017 May;11(1):Article. [doi: 10.5817/CP2017-1-1]
12. Alfredsson Ågren K, Kjellberg A, Hemmingsson H. Access to and use of the Internet among adolescents and young adults with intellectual disabilities in everyday settings. J Intellect Dev Disabil 2020 Jan 2;45(1):89-98. [doi: 10.3109/13668250.2018.1518898]
13. Vlot - van Anrooij K, Frankena TK, van der Cruijsen A, Jansen H, Naaldenberg J, Bevelander KE. Shared decision making in inclusive research: reflections from an inclusive research team. Brit J Learn Disabil 2022 Dec;50(4):554-563 [FREE Full text] [doi: 10.1111/bld.12450]
14. van Calis JFE, Naaldenberg J, van der Cruijsen AWC, Koks-Leensen MCJ, Leusink GL, Bevelander KE. Inclusive digital platforms: designing for and with users with mild intellectual disabilities or low literacy skills. Computers in Human Behavior Reports 2025 Mar;17:100617. [doi: 10.1016/j.chbr.2025.100617]
15. Følstad A, Araujo T, Law ELC, et al. Future directions for chatbot research: an interdisciplinary research agenda. Computing 2021 Dec;103(12):2915-2942. [doi: 10.1007/s00607-021-01016-7]
16. Leslie D, Mazumder A, Peppin A, Wolters MK, Hagerty A. Does “AI” stand for augmenting inequality in the era of COVID-19 healthcare? BMJ 2021 Mar 15;372:n304. [doi: 10.1136/bmj.n304] [Medline: 33722847]
17. Clemensen J, Rothmann MJ, Smith AC, Caffery LJ, Danbjorg DB. Participatory design methods in telemedicine research. J Telemed Telecare 2017 Oct;23(9):780-785. [doi: 10.1177/1357633X16686747] [Medline: 28027678]
18. Faber JS, Al-Dhahir I, Kraal JJ, et al. Guide development for eHealth interventions targeting people with a low socioeconomic position: participatory design approach. J Med Internet Res 2023 Dec 4;25:e48461. [doi: 10.2196/48461] [Medline: 38048148]
19. Guffroy M, Guerrier Y, Kolski C, Vigouroux N, Vella F, Teutsch P. Adaptation of user-centered design approaches to abilities of people with disabilities. In: Miesenberger K, Kouroupetroglou G, editors. Computers Helping People with Special Needs ICCHP 2018 Lecture Notes in Computer Science(), Vol 10896: Springer; 2018:462-465. [doi: 10.1007/978-3-319-94277-3_71]
20. van Calis JFE, Bevelander KE, van der Cruijsen AWC, Leusink GL, Naaldenberg J. Toward inclusive approaches in the design, development, and implementation of eHealth in the intellectual disability sector: scoping review. J Med Internet Res 2023 May 30;25:e45819. [doi: 10.2196/45819] [Medline: 37252756]
21. Hekkert P, van Dijk M. Vision in Design: A Guidebook for Innovators: Laurence King Publishing; 2011.
22. Ferreira FK, Song EH, Gomes H, Garcia EB, Ferreira LM. New mindset in scientific method in the health field: design thinking. Clinics (Sao Paulo) 2015 Dec;70(12):770-772. [doi: 10.6061/clinics/2015(12)01] [Medline: 26735214]
23. Nielsen J. Usability Engineering: Elsevier Science; 1993.
24. Hendriks N, Slegers K, Duysburgh P. Codesign with people living with cognitive or sensory impairments: a case for method stories and uniqueness. CoDesign 2015 Jan 2;11(1):70-82. [doi: 10.1080/15710882.2015.1020316]
25. Newell AF, Gregor P, Morgan M, Pullin G, Macaulay C. User-sensitive inclusive design. Univ Access Inf Soc 2011 Aug;10(3):235-243. [doi: 10.1007/s10209-010-0203-y]
26. van Gemert-Pijnen J, Nijland N, van Limburg M, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 5;13(4):e111. [doi: 10.2196/jmir.1672] [Medline: 22155738]
27. Frauenberger C, Makhaeva J, Spiel K. Blending methods: developing participatory design sessions for autistic children. Presented at: IDC ’17: Proceedings of the 2017 Conference on Interaction Design and Children; Jun 27-30, 2017; Stanford, CA p. 39-49. [doi: 10.1145/3078072.3079727]
28. Burtscher S, Gerling K. Neurodivergence and work in human-computer interaction: mapping the research landscape. 2024 Jun 25 Presented at: CHIWORK 2024; Jun 25-27, 2024 URL: https://dl.acm.org/doi/proceedings/10.1145/3663384 [doi: 10.1145/3663384.3663386]
29. Effert JS, Petersdorf L, Aktas S, Mörike F. Re-aligning the building blocks of usability testing: methodological insights in cognitive accessibility research for work related technologies. 2025 Aug 31 Presented at: MuC ’25; Aug 31 to Sep 3, 2025; Chemnitz Germany p. 700-704 URL: https://dl.acm.org/doi/proceedings/10.1145/3743049 [doi: 10.1145/3743049.3748573]
30. Suchanek O, Meissner JL, Angelini R, Spiel K. From participation to solidarity: a case study on access of maker spaces from deaf and hearing perspectives. 2025 Apr 26 Presented at: CHI 2025; Apr 26 to May 1, 2025 URL: https://dl.acm.org/doi/proceedings/10.1145/3706598 [doi: 10.1145/3706598.3713202]
31. Frankena TK, Naaldenberg J, Cardol M, et al. A consensus statement on how to conduct inclusive health research. J Intellect Disabil Res 2019 Jan;63(1):1-11. [doi: 10.1111/jir.12486] [Medline: 29642277]
32. Frankena TK, Naaldenberg J, Cardol M, Linehan C, van Schrojenstein Lantman-de Valk H. Active involvement of people with intellectual disabilities in health research - a structured literature review. Res Dev Disabil 2015;45-46:271-283. [doi: 10.1016/j.ridd.2015.08.004] [Medline: 26280692]
33. van Heumen L, Krueger C, Mihaila I. The development of a co-researcher training with and for people with intellectual and developmental disabilities. J Appl Res Intellect Disabil 2024 May;37(3):e13200. [doi: 10.1111/jar.13200] [Medline: 38389315]

XSL•FO
RenderX

34. van Leersum CM, Jaschinski C, Bults M, van der Zwart J. Citizen involvement in research on technological innovations for health, care or well-being: a scoping review. Health Res Policy Syst 2024 Sep 2;22(1):119. [doi: 10.1186/s12961-024-01152-4] [Medline: 39223606]
35. van Dyk L. A review of telehealth service implementation frameworks. Int J Environ Res Public Health 2014 Jan 23;11(2):1279-1298. [doi: 10.3390/ijerph110201279] [Medline: 24464237]
36. van Calis JF, van Cooten VJ, Smeets O, et al. The implementation of smart continence care for people with disabilities: a qualitative study of key stakeholders' first-hand experiences. Digit HEALTH 2024;10:20552076241290399. [doi: 10.1177/20552076241290399] [Medline: 39421307]
37. Greenhalgh T, Abimbola S. The NASSS Framework - a synthesis of multiple theories of technology implementation. Stud Health Technol Inform 2019 Jul 30;263(193-204):193-204. [doi: 10.3233/SHTI190123] [Medline: 31411163]
38. Central Committee on Research Involving Human Subjects (CCMO). URL: https://english.ccmo.nl/investigators/legal-framework-for-medical-scientific-research/laws/medical-research-involving-human-subjects-act-wmo [accessed 2026-02-16]
39. Non-WMO research. Centrale Commissie Mensgebonden Onderzoek (CCMO). URL: https://www.ccmo.nl/onderzoekers/aanvullende-informatie-over-bepaalde-soorten-onderzoek/niet-wmo-onderzoek [accessed 2026-02-16]
40. Radboudumc. Commissie Mensgebonden Onderzoek (CMO). URL: https://www.radboudumc.nl/over-het-radboudumc/kwaliteit-veiligheid/toetsen-van-medisch-wetenschappelijk-onderzoek/bij-welke-commissies-kunt-u-terecht/cmo-radboudumc [accessed 2026-02-16]
41. IkOnderzoekMee.: Crowdience URL: https://www.ikonderzoekmee.nl [accessed 2026-02-11]
42. Convention on the rights of persons with disabilities. United Nations. 2006 Dec. URL: https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-persons-disabilities [accessed 2026-02-11]
43. Vindrola-Padros C, Johnson GA. Rapid techniques in qualitative research: a critical review of the literature. Qual Health Res 2020 Aug;30(10):1596-1604. [doi: 10.1177/1049732320921835] [Medline: 32667277]
44. Tessier S. From field notes, to transcripts, to tape recordings: evolution or combination? Int J Qual Methods 2012 Sep;11(4):446-460. [doi: 10.1177/160940691201100410]
45. Watkins DC. Rapid and rigorous qualitative data analysis: the "RADaR" technique for applied research. Int J Qual Methods 2017;16(1). [doi: 10.1177/1609406917712131]
46. Culén AL, Følstad A. Innovation in HCI: what can we learn from design thinking? 2014 Presented at: NordiCHI '14: Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational; Oct 26-30, 2014; Helsinki, Finland p. 849-852. [doi: 10.1145/2639189.2654845]
47. Clarkson PJ, Coleman R, Keates S, Lebbon C. Inclusive Design: Design for the Whole Population 2013.
48. Goldsmith S. Universal Design: A Manual of Practical Guidance for Architects: Architectural Press; 2000.
49. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367. [doi: 10.2196/jmir.8775] [Medline: 29092808]
50. Bigby C, Frawley P, Ramcharan P. Conceptualizing inclusive research with people with intellectual disability. J Appl Res Intellect Disabil 2014 Jan;27(1):3-12. [doi: 10.1111/jar.12083] [Medline: 24390972]
51. Bayor AA, Brereton M, Sitbon L, et al. Toward a competency-based approach to co-designing technologies with people with intellectual disability. ACM Trans Access Comput 2021 Jun 30;14(2):1-33. [doi: 10.1145/3450355]
52. Farr M. Power dynamics and collaborative mechanisms in co-production and co-design processes. Crit Soc Policy 2018 Nov;38(4):623-644. [doi: 10.1177/0261018317747444]
53. Frauenberger C, Good J, Fitzpatrick G, Iversen OS. In pursuit of rigour and accountability in participatory design. Int J Hum Comput Stud 2015 Feb;74:93-106. [doi: 10.1016/j.ijhcs.2014.09.004] [Medline: 26109833]
54. Fletcher-Watson S, Adams J, Brook K, et al. Making the future together: shaping autism research through meaningful participation. Autism 2019 May;23(4):943-953. [doi: 10.1177/1362361318786721] [Medline: 30095277]
55. Walmsley J, Strnadová I, Johnson K. The added value of inclusive research. J Appl Res Intellect Disabil 2018 Sep;31(5):751-759. [doi: 10.1111/jar.12431] [Medline: 29231273]
56. Northway R, Howarth J, Evans L. Participatory research, people with intellectual disabilities and ethical approval: making reasonable adjustments to enable participation. J Clin Nurs 2015 Feb;24(3-4):573-581. [doi: 10.1111/jocn.12702] [Medline: 25266081]
57. Guedes LS, Johnstone J, Ellis K, Landoni M. Creative technologies in action: empowering individuals with intellectual disabilities. In: Miesenberger K, Peňáz P, Kobayashi M, editors. Computers Helping People with Special Needs ICCHP 2024 Lecture Notes in Computer Science, Vol 14751: Springer; 2024. [doi: 10.1007/978-3-031-62849-8_24]
58. Ahmed SK. How to choose a sampling technique and determine sample size for research: a simplified guide for researchers. Oral Oncology Reports 2024 Dec;12:100662. [doi: 10.1016/j.oor.2024.100662]
59. van Calis JFE, Koks-Leensen MCJ, Naaldenberg J, Bevelander KE. Inclusive digital platforms: Designing for and with users with mild intellectual disabilities or low literacy skills. Radboud University. 2025. URL: https://doi.org/10.34973/nw4q-5677 [accessed 2026-02-11]

XSL•FO
RenderX

60. Hermsen S, van Calis J, Bevelander K. Inclusive design, co-creation and evaluation of digital health for individuals with a mild intellectual disability. Presented at: Design With Or Design By, 19th EAI International Conference on Pervasive Computing Technologies for Healthcare 2025; Oct 15-17, 2025 URL: https://sander-hermsen.nl/wp-content/themes/portfolio/img/positioningpaper_InclDes_MID.pdf [accessed 2026-02-11]

Abbreviations

ASD: autism spectrum disorder
CeHRes: Center for eHealth Research
MID: mild intellectual disability
NASSS: Non-adoption, Abandonment, Scale-up, Spread, and Sustainability
SVA: sensitive virtual assistant
ViP: Vision in Product Design
WP: work package

Edited by A Bamgboje-Ayodele, A Kushniruk; submitted 06.Aug.2025; peer-reviewed by F Mörike, W Shi; accepted 21.Jan.2026; published 25.Feb.2026.

Please cite as:
van Calis JFE, Bevelander KE, van der Cruijsen AWC, Naaldenberg J, Leusink GL
Reflections on the Inclusive Co-Design Process of a Virtual Assistant for Individuals With Complex Care Needs: Mixed Methods Study
JMIR Hum Factors 2026;13:e81941
URL: https://humanfactors.jmir.org/2026/1/e81941
doi:10.2196/81941

© Julia F E van Calis, Kirsten E Bevelander, Anneke W C van der Cruijsen, Jenneken Naaldenberg, Geraline L Leusink. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 25.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Development of a Web-Based Application to Support Informal Caregivers of Individuals With Head and Neck Cancer: Human-Centered Design Approach

Awais Ahmad[1], PhD; Åsa Cajander[1], PhD; Ulrica Langegård[2,3], PhD; Birgitta Johansson[4], PhD; Anna Henriksson[5], PhD; Mona Pettersson[6], PhD; Waqas Ahmad[7], MSc; Ylva Tiblom Ehrsson[8], PhD

[1]Department of Information Technology, Uppsala University, Lägerhyddsvägen 1, Uppsala, Sweden
[2]Department of Oncology, Sahlgrenska University Hospital, Gothenburg, Sweden
[3]Institute of Health and Care Sciences, University of Gothenburg, Gothenburg, Västra Götaland, Sweden
[4]Department of Immunology, Genetics, and Pathology, Uppsala University, Uppsala, Sweden
[5]Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden
[6]Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden
[7]Department of Software Engineering, A2Z Technology, Stockholm, Sweden
[8]Department of Surgical Sciences, Uppsala University, Uppsala, Sweden

Corresponding Author:
Awais Ahmad, PhD
Department of Information Technology, Uppsala University, Lägerhyddsvägen 1, Uppsala, Sweden

Abstract

Background: Informal caregivers (ICs), often family members or close friends, provide essential support to individuals with head and neck cancer. However, they are frequently unprepared for the emotional, practical, and medical challenges involved. Web-based applications offer promising opportunities to support ICs, but their long-term adoption and acceptance remain uncertain.

Objective: This paper presents the development of Carer eSupport, a web-based application to support ICs' well-being and preparedness for caregiving. We detail the design and functionality of the Carer eSupport application and explain how it responds to both the functional and psychological needs of ICs. Additionally, we report findings from the pilot study and highlight the initial challenges ICs faced when engaging with the application, along with the strategies used to overcome them.

Methods: The study involved a multicenter research trial across ear, nose, and throat clinics and oncology and radiotherapy clinics at 4 university hospitals in Sweden. The application was developed through 3 human-centered design (HCD) iterations involving ICs, health care professionals, and researchers in human-computer interaction and cancer care.

Results: The results present an overview of the current version of Carer eSupport (developed during the third design iteration), with a focus on features that address the psychological needs of ICs, including competence, autonomy, and a sense of connection to others. The pilot study achieved a 66.7% (20/30) consent rate, a 75% (9/12) successful login rate among participants, and a 13.3% (4/30) attrition rate, meeting the established criteria. The pilot study confirmed the application's readiness for further evaluation in an ongoing randomized controlled trial. It also identified challenges, including the time constraints of ICs, login and authentication issues, limited IT infrastructure, and gaps in digital literacy.

Conclusions: Findings from the HCD process and pilot study indicate that a personalized, interactive application like Carer eSupport can provide meaningful support for ICs of individuals with head and neck cancer. The integration of HCD and health care science offers early guidance for developing digital tools that are both evidence-based and empathetic, with potential relevance beyond caregiving contexts.

Trial Registration: ClinicalTrials.gov NCT05028452, https://clinicaltrials.gov/study/NCT05028452 and NCT06307418, https://clinicaltrials.gov/study/NCT06307418

International Registered Report Identifier (IRRID): RR2-10.1186/s12885-024-12273-y

(JMIR Hum Factors 2026;13:e81896) doi:10.2196/81896

KEYWORDS

human-centered design; designing for well-being; eudaimonic motivation; head and neck cancer; informal caregivers

XSL•FO
RenderX

Introduction

Background

Health care systems worldwide face mounting pressure as they contend with limited resources and the increasing care needs of an aging population [1,2]. These strains are particularly evident in home-based care, where access to professional support is often limited. In cancer care, the shift from hospital-based treatment to outpatient and home settings has transferred much of the day-to-day responsibility to family members and close friends [3,4]. These individuals, commonly referred to as informal caregivers (ICs), provide unpaid support and are often the primary source of informal care for people with cancer [5,6]. They often struggle to navigate their roles because they lack access to the essential information, guidance, and resources needed to provide adequate care at home [7-10].

The emotional and physical toll of caregiving can be significant, affecting ICs' overall well-being. It often leads to mental health challenges such as anxiety, depression, and symptoms of post-traumatic stress [11-15]. Over time, ICs may begin to lose a sense of their own identity as their lives become increasingly centered around the illness of their patients. These difficulties are made worse when the support systems and information needed to manage care at home are limited or hard to access [16]. Therefore, ICs often express a strong need for external support and guidance to feel more confident and capable in their roles [17,18]. This is where the concept of preparedness for caregiving becomes relevant. Preparedness refers to ICs' perceived ability to provide physical and emotional support while also managing the stress of caregiving [19]. Higher levels of preparedness have been linked to lower caregiver burden and better outcomes for both ICs and patients [3]. Given these challenges, digital tools such as web-based applications offer promising opportunities to support ICs. These eHealth systems can serve as accessible platforms for delivering tailored information, emotional support, and practical guidance to those providing cancer care at home [12,16]. However, acceptance and use of eHealth systems remain uncertain, highlighting the need for further research on design strategies that ensure the meaningful involvement of relevant stakeholders throughout the development process [20-23].

Human-centered design (HCD) is a foundational approach for developing technology in the human-computer interaction (HCI) field. HCD emphasizes placing users' needs, preferences, and contextual factors at the core of the design and development process to create effective and user-empowering systems [23]. Even though HCD is widely accepted and used, it often falls short of meeting users' specific needs, making them feel disconnected and unsupported [24,25]. One major reason for this gap is how HCD is applied. When used in a limited or inconsistent way, especially without genuine and ongoing stakeholder involvement, it can become more of a label than a clear design method. Consequently, although HCD holds significant potential to enhance the usefulness and user-friendliness of technology, its impact remains limited when it is not implemented consistently and systematically in practice [25-28].

The HCD process has primarily focused on usability and functionality, but there is growing recognition that technology should also support users' well-being [29-34]. Well-being is a multifaceted concept that includes physical health, mental and emotional stability, social relationships, economic security, access to education and health care, environmental conditions, cultural aspects, personal safety, and overall life satisfaction [35]. People tend to use technologies that satisfy their functional needs and enhance their physical and psychological well-being [34]. In the field of psychology, well-being is often understood in two ways. Hedonic well-being is about feeling good, seeking pleasure, and avoiding discomfort. On the other hand, eudaimonic well-being is about personal growth, having a sense of purpose, and reaching one's full potential [36,37]. Eudaimonic motivation is a person's inner drive to grow, learn, and seek meaningful experiences in life [36]. Research on hedonic aspects is more established. For instance, the well-known extended Unified Theory of Acceptance and Use of Technology (UTAUT2) model evaluates technology acceptance by integrating hedonic motivation as a key factor in technology acceptance and use [38]. However, there is a noticeable scarcity of research that explicitly examines how eudaimonic motivation can be systematically integrated into the HCD process [39-42].

Some recent studies have sought to evaluate eudaimonic motivation quantitatively [33,42,43]. For instance, Woźniak et al [42] combined insights from psychology to develop the Eudaimonic Technology Experience Scale, which aims to quantify the meaningful interaction of users. Similarly, Kuriakose and Nagasubramaniyan [39] extended UTAUT2 to assess the acceptance of digital entertainment applications. Despite some attempts to evaluate eudaimonic motivation quantitatively, a significant gap remains in the qualitative understanding of how eudaimonic motivation influences technology acceptance and use, particularly in complex and emotionally sensitive domains such as caregiving for relatives with serious diseases. Moreover, these studies focus exclusively on the evaluation phase of the HCD process. This study integrates eudaimonic perspectives from the early stage of the HCD process, with the aim of supporting users' psychological well-being.

The Carer eSupport Project

Carer eSupport is a multidisciplinary project combining HCI, software engineering, cancer nursing, and medical and health care science expertise. It seeks to develop a web-based application that improves ICs' preparedness for caregiving and promotes their overall well-being [3,44]. By equipping ICs with targeted, accessible, and user-friendly resources, the project aims to reduce caregiver burden and positively support the overall health of individuals with head and neck cancer (HNC).

We applied the integrated HCD process with a eudaimonic perspective to guide the development and evaluation of Carer eSupport. The HCD process involved 3 iterative cycles, engaging ICs, health care professionals (HCPs), an expert group of ICs, and technology designers and developers. The collaboration of different stakeholders enabled us to continuously refine the application to understand and address

XSL•FO
RenderX

the practical and psychological needs of ICs. The project is currently conducting a randomized controlled trial (RCT) to evaluate the effects of Carer eSupport on ICs' self-reported preparedness for caregiving, caregiver burden, and well-being, compared to support as usual [44]. A pilot study was conducted at the early stage of the RCT to investigate the feasibility of the procedures and the current version of Carer eSupport. This paper presents the results of the pilot study, which determined whether it was appropriate to proceed with the full-scale RCT.

Study Aim

This paper aims to present how a web-based application can be designed to support ICs' well-being and preparedness for caregiving through the HCD process. We detail the design and functionalities of the Carer eSupport application and explain how it responds to both the functional and psychological needs of ICs. By integrating HCD with a eudaimonic perspective, this research also intends to provide insights into design approaches that prioritize user well-being in web-based application development.

Additionally, we report findings from the pilot study and highlight the initial challenges ICs faced when engaging with the application and the strategies used to overcome these issues.

Methods

Development of Carer eSupport

Human-Centered Design Process

We conducted 3 iterations of the HCD process to develop the Carer eSupport application. HCD is an iterative approach that prioritizes users and other stakeholders at the core of the design process [45]. This iterative process ensured that the application addressed the needs of ICs by incorporating their direct feedback in different design phases. As highlighted in previous research [46,47], continuous refinement based on user input enhanced usability, relevance, and support for both the practical and psychological challenges of caregiving.

The first step in the HCD process involves understanding users and their context through comprehensive research methods, including interviews, observations, surveys, and focus groups, to define user requirements. Thereafter, the design team defines the problem and establishes specific requirements based on these insights. They then produce design solutions through brainstorming and concept development, which are evaluated through user testing to assess usability, functionality, and overall satisfaction. The design is iteratively refined based on feedback until the final product effectively meets user needs and expectations [48].

As a multidisciplinary project combining health care research and computer science expertise, we integrated health care research methods, such as feasibility studies and RCTs, with HCD to develop and evaluate Carer eSupport. In the following sections, we detail the 3 iterations that shaped the final version of the application. Findings from the first and second iterations have been presented in detail elsewhere [15,49]. This study primarily focuses on the third iteration, with its key findings presented in the Results section. Table 1 provides an overview of all 3 HCD iterations, including the major activities conducted and a summary of their outputs during the design and development of Carer eSupport.

Table . Overview of the human-centered design (HCD) iterations in the development of Carer eSupport.

HCD iterations	Key activities	Outputs
1: Prototype design	• Focus groups • Prototyping • Developed the initial content of Carer eSupport • Evaluation with the expert group of ICs[a]	Interactive prototype and initial content of Carer eSupport
2: Feasibility study	• Developed the initial version of Carer eSupport • Mixed methods evaluation (surveys, application usage logs, interviews) • Extended the UTAUT[b] model with eudaimonic motivation	Confirmation of the feasibility and acceptability of Carer eSupport and suggestions for improvements before the randomized controlled trial
3: Pilot study	• Developed the current version of Carer eSupport • Pilot study evaluation	Confirmation to continue the full-scale randomized controlled trial and improvements of its procedures

[a]ICs: informal caregivers.

[b]UTAUT: Unified Theory of Acceptance and Use of Technology.

Iteration 1: Prototype Design

Understanding ICs' Needs and Specifying Their Requirements

To identify the challenges and needs of ICs, we conducted focus groups with ICs and HCPs. These sessions provided essential insights into user preferences and the functional requirements for the Carer eSupport application. In this iteration, we explored both hedonic and eudaimonic aspects of well-being. Using a deductive thematic analysis, we examined factors such as virtue, autonomy, competence, relatedness, achievements, and engagement derived from the focus group data. Our findings

XSL•FO
RenderX

indicate that pleasure, as a component of hedonic motivation, is not applicable in the caregivers' context—a point we discussed in detail in our previous study [49].

Three main stakeholder groups took part in this phase. ICs shared their experiences and highlighted the complexity of caregiving, stressing the need for support in patient care and their well-being. HCPs, including nurses, doctors, dietitians, and speech therapists, ensured the application was clinically relevant. A caregiver expert group also reviewed the content in several sessions and gave feedback that helped refine and improve the application. This expert group comprised ICs who had more extensive and continuous involvement in the design process, working closely with the research team to review, discuss, and refine both the application structure and its content. Their role differed from that of ICs participating in focus groups or later study phases, such as the feasibility study and the RCT, as they were engaged in repeated sessions and contributed more deeply to design decisions. In recognition of their sustained contribution and level of engagement, members of the expert group were financially compensated for their participation.

Prototype Development

Building on these insights, we developed an interactive prototype of Carer eSupport using Figma, a collaborative tool for designing and evaluating responsive prototypes [50]. The prototype was created to address the identified needs and preferences of ICs. It included key functions such as information on HNC treatment, resources to support ICs' well-being, a web-based discussion forum, and an online meeting space. The project team reviewed all the functionalities collaboratively to ensure they aligned with user requirements.

Prototype Evaluation

The caregiver expert group preliminarily evaluated the prototype. To collect their feedback, 2 meetings were organized, 1 online and 1 in person. During the online meeting, 2 health care researchers and the HCI researcher presented the initial prototype, along with early versions of the portal content, to the expert group.

The second meeting was a full-day, in-person session held at Uppsala University. During the first 2 sessions, the group was introduced to the application's video content, including demonstrations, lectures, and presentations. These sessions enabled discussion of the relevance and usefulness of the materials for caregiving. In a subsequent session, the prototype's main functions were demonstrated, which supported more detailed feedback on usability and navigation.

The expert group reported overall satisfaction with the interface but also identified areas for improvement in both content and functionality. Their suggestions were incorporated into the design, and the prototype was refined before proceeding to the feasibility study in the second iteration of the HCD process.

Iteration 2: Feasibility Study

Development of the Initial Version of Carer eSupport

Based on feedback from the first iteration, the initial functional version of Carer eSupport was developed and structured around 3 major categories. The first category provides educational content on HNC diagnosis, treatment, and management. All the content was provided through text and video to facilitate diverse user needs. The second category provides critical interventions and practical advice on managing symptoms, side effects, and caregiving challenges, focusing on areas such as oral health care, nutrition, and fatigue. The third category, Self-Care for ICs themselves, provides resources to support ICs' well-being by addressing issues such as everyday life support, sleep difficulties, tiredness, and a diary for finding balance. All content was curated in collaboration with HCPs and researchers specialized in HNC to ensure clinical accuracy and relevance.

Evaluation of Initial Version Carer eSupport

A feasibility study using a mixed methods approach was conducted in accordance with the Medical Research Council's guidelines for developing and evaluating complex interventions [51]. Quantitative data, such as survey responses and application log usage, were combined with qualitative insights from semistructured interviews [44]. The evaluation confirmed the application's relevance, acceptability, and usability, with recommendations for further refinement. Positive feedback from ICs also highlighted the application's potential to support their caregiving responsibilities and to improve their well-being. We also extended the UTAUT model in the feasibility study by integrating the eudaimonic motivation construct and proposed an extended UTAUT framework for analyzing the application's impact on ICs' well-being.

Iteration 3: Pilot Study

In the third iteration of the HCD process, building on insights from previous iterations, the current version of Carer eSupport was developed and is currently being evaluated through an ongoing RCT [44]. Widely considered a gold standard for examining the effectiveness of interventions in health care and behavioral research [52], RCTs use a rigorous methodology where study participants are randomly assigned to either an intervention group or a control group. The random allocation of participants ensures an unbiased comparison by minimizing the influence of confounding variables and helps researchers establish a causal relationship between the intervention and its outcomes [53]. Before proceeding to a full-scale RCT, the last step was to conduct a pilot study to evaluate feasibility, refine procedures, and assess the application's potential for large-scale implementation. Although this paper outlines the entire HCD process for designing and developing Carer eSupport, it primarily focuses on the outcomes of the third iteration, the pilot study.

Procedures

In this section, we present the procedures for the RCT and pilot study conducted in the third iteration of the HCD process. Patients aged 18 years or older with HNC scheduled to receive radiotherapy with or without surgical or oncology medical treatment were identified and screened by nurses from ear, nose, and throat or oncology and radiotherapy clinics. Eligible patients were approached by a clinical or research nurse, in person or by telephone, and were provided with detailed information about the study. Consent was obtained to contact an IC nominated by

XSL•FO
RenderX

the patient and access information from the medical record about the HNC diagnosis, treatment, and outcome.

Once the patient provided informed consent, the identified IC was contacted using their preferred method (in person, by telephone, or by email) and was provided comprehensive information about the study before obtaining their oral and written consent. Enrolled ICs were then provided with credentials to complete an internet-administered baseline assessment. These credentials included a web address, a unique user ID, a temporary password, and a detailed user guide with screenshots to facilitate navigation. After completing the baseline assessment, ICs were randomly assigned to 1 of 2 participant groups: (1) the Carer eSupport plus standard care group or (2) the standard care-only group. ICs in the Carer eSupport plus standard care group received additional login credentials to engage with the application. The data collection methods, including the primary outcome (preparedness for caregiving), secondary outcomes, and assessment points, are presented in detail in the study protocol [44].

Pilot Study Analysis

A pilot study was conducted to confirm the feasibility of the study procedures and the Carer eSupport application. Pilot studies in RCTs are essential for gauging the likely success of the main trial and, if needed, refining its design [54]. Thirty ICs participated in this phase. Feasibility was analyzed using predefined progression criteria to determine whether the study could proceed to a full-scale RCT. These criteria were as follows: at least 45% of eligible ICs providing written consent, at least 50% actively using the Carer eSupport application, and an attrition rate below 30% one month after randomization. Meeting these benchmarks would allow the study to proceed to a full-scale RCT; otherwise, recruitment would be discontinued. If the criteria were not met, ICs assigned to the Carer eSupport plus standard care group would retain access to the application for the entire 18-week intervention period. However, no further follow-up assessments would be conducted. Please refer to the study protocol for more detailed insights into the study methods and procedures [44].

In addition, qualitative data were collected and analyzed during the pilot study. One week after the login details were distributed, ICs who had not logged in were contacted by telephone to check whether they required assistance. Two researchers conducted these calls and took detailed notes, which were discussed jointly to ensure a shared understanding of the content. One researcher then analyzed the notes using thematic analysis following Braun and Clarke's framework [55]. The analysis involved familiarization with the data through repeated readings, generation of initial codes, and organization of these codes into broader themes. Themes were iteratively reviewed and refined to ensure they were grounded in the data and aligned with the study aims. The final set of themes was discussed and agreed upon within the research team, which strengthened the credibility and trustworthiness of the analysis.

Ethical Considerations

All research procedures in the project were approved by the Swedish Ethical Review Council (Dnr: 2020-04650), and the feasibility evaluation is also registered on ClinicalTrials.gov (NCT05028452). The feasibility study was registered at ClinicalTrials.gov (NCT05028452), and the RCT is also registered at ClinicalTrials.gov (NCT06307418).

Informed consent was also obtained from patients, their ICs, and HCPs before their enrollment in the study. Before participating, ICs received a written information letter outlining the study's purpose and data collection procedures and signed an informed consent form. In addition, a verbal explanation was provided prior to the focus groups and interviews to ensure participants had a clear understanding of the study. No compensation was provided to the participants in this study.

Safeguarding anonymity and privacy was a key ethical priority. Participants were informed that no identifiable personal information would be disclosed in any publications or reports. All interviews were audio- and video-recorded and transcribed verbatim. The recordings were stored securely in Uppsala University's database and were accessible only to authorized members of the research team.

Results

Overview

This section first gives an overview of the current version of the Carer eSupport application, developed during the third iteration of the HCD process. We present the application's main modules and explain how the eudaimonic perspective is integrated into them. It serves as a practical example of how a web-based application is developed using an iterative HCD process, with a special focus on eudaimonic motivation to enhance ICs' well-being and improve their preparedness for caregiving. We then present the findings from the pilot study, which indicate that the predefined progression criteria were achieved, supporting the continuation to a full-scale RCT. Finally, we present the initial challenges ICs faced during the pilot study phase when using the application and describe the strategies implemented to improve its usability and accessibility.

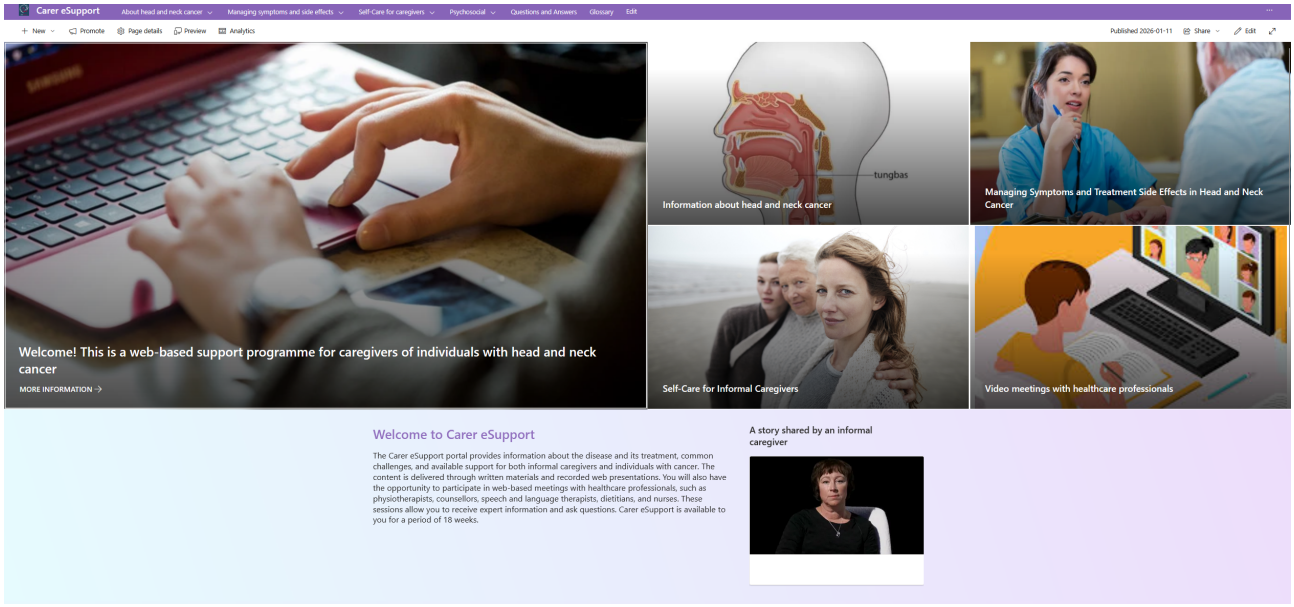
Carer eSupport Application

Overview

The main page of Carer eSupport is organized into 5 modules: Information About HNC, Managing Symptoms and Side Effects of HNC, Self-Care for ICs, Psychosocial Support for ICs, and Interactive Video Meetings with HCPs and ICs (see Figure 1). In addition to these modules, the application features a Frequently Asked Questions section. This section addresses common queries related to HNC, caregiving, and the broader health care system. A glossary is provided to clarify medical terms and expressions, which helps ICs understand complex information with greater ease. Most of the content is provided in multiple formats, including lectures and written text, allowing ICs to choose the format that is easier for them. The application was initially developed in Swedish; however, the main page and module pages were explicitly translated into English for this paper to support a broader audience. To protect participants' privacy, all screenshots were anonymized by blurring any faces that could lead to identification.

XSL•FO
RenderX

Figure 1. The main page of Carer eSupport.



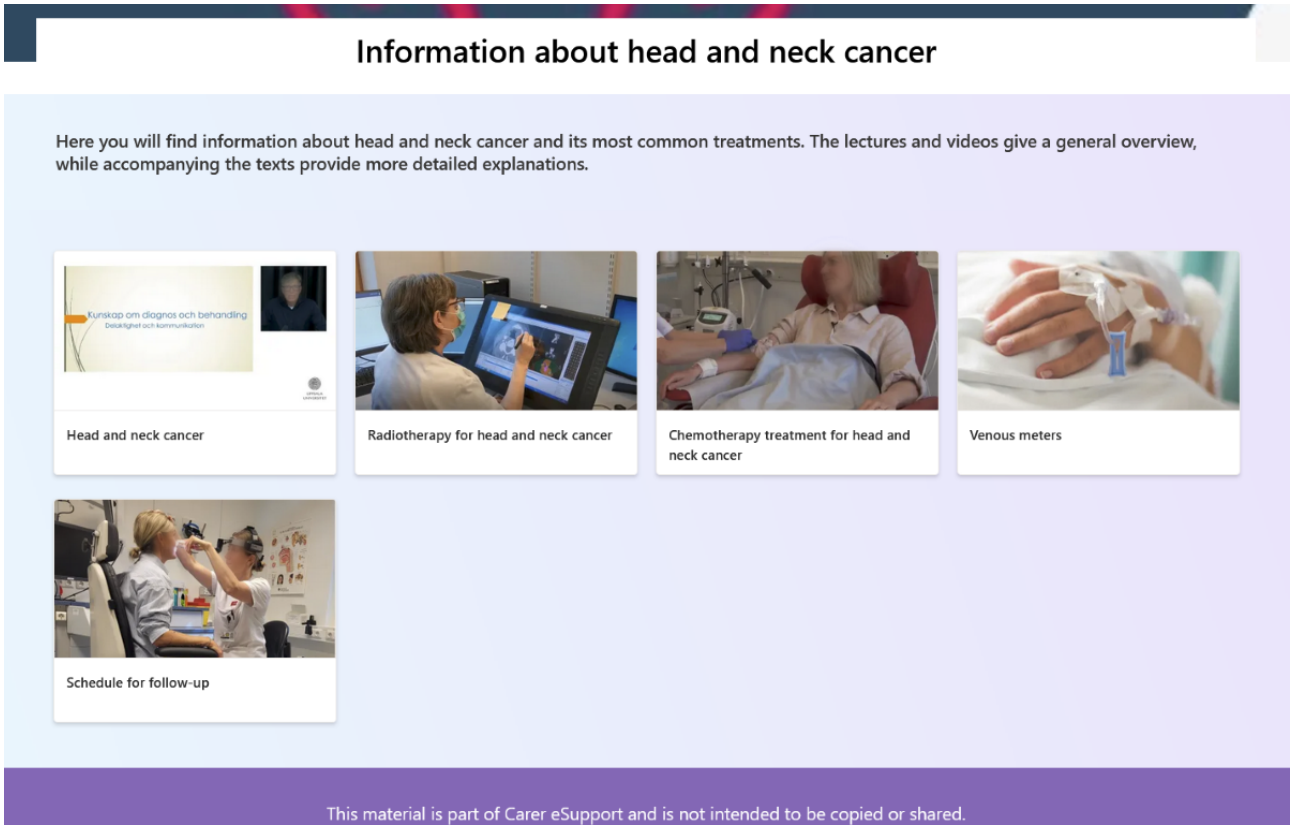
Information About HNC

This module, as presented in Figure 2, provides a comprehensive resource on HNC and its common treatments to enhance the preparedness of ICs. It is designed to empower ICs by offering precise and detailed information on HNC and its treatment options, thereby improving their ability to manage the consequences of the disease and its treatment. The module addresses different aspects of HNC care and is anticipated to enhance a sense of competence and help ICs make autonomous and informed decisions aligned with a eudaimonic perspective.

The module is organized into 5 sections. The first section offers an overview of HNC, presenting essential information about the disease, its risk factors, and its typical progression. The following sections focus on treatment-specific details. One section provides a detailed explanation of radiation therapy, outlining its protocols, potential side effects, and strategies for managing these effects. Another section in this module is dedicated to chemotherapy, describing treatment regimens and common adverse reactions and offering practical advice for supporting patients during this modality. Moreover, the module includes guidance on using and managing venous catheters and presents the best practices to ensure patient safety. The final section, a follow-up schedule, explains the purpose and importance of each posttreatment visit and equips ICs with the knowledge necessary to support patients effectively during recovery.

XSL•FO
RenderX

Figure 2. Module for information about head and neck cancer.



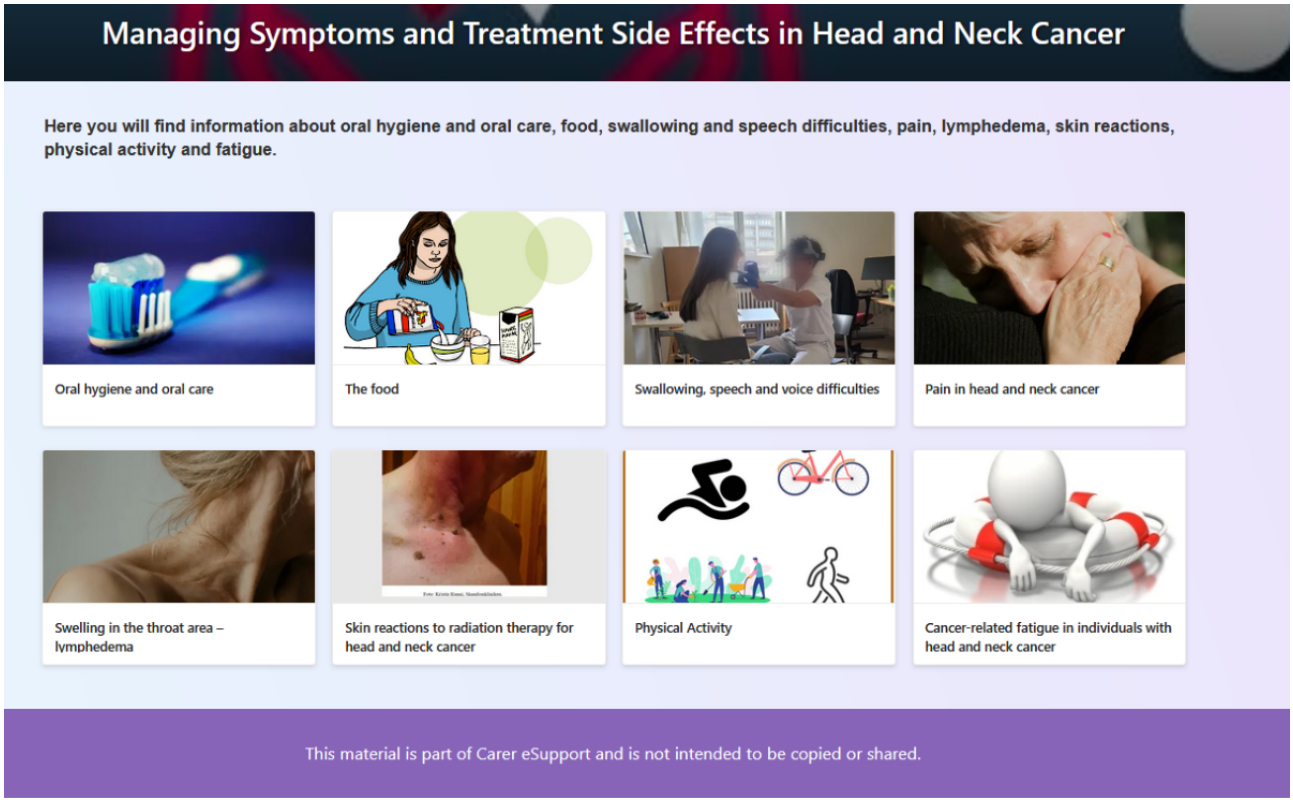
Managing Symptoms of HNC and Side Effects of Treatments

This module is designed to better prepare ICs for the challenges of HNC by providing detailed guidance on managing the symptoms and side effects commonly experienced by patients (as presented in Figure 3). Structured into 8 sections, the module addresses key areas such as oral hygiene, providing detailed recommendations for maintaining oral health, especially after reconstructive surgery and radiotherapy. It includes guidance on nutrition, with a focus on dietary adjustments during treatment, along with resources like a recipe bank and information on feeding tubes.

Support is also offered for issues related to swallowing, speech, and voice, presenting practical strategies to manage these difficulties. The module includes targeted advice for pain relief and approaches to managing lymphedema, including recommendations for reducing swelling during or after radiation therapy. Strategies to mitigate skin reactions caused by radiation are outlined as well. Furthermore, it provides tailored approaches to maintain an active lifestyle that supports overall well-being. It offers insights and practical strategies to manage cancer-related fatigue, ensuring that ICs can support patients and themselves throughout treatment.

XSL•FO
RenderX

Figure 3. Module for managing symptoms and side effects.



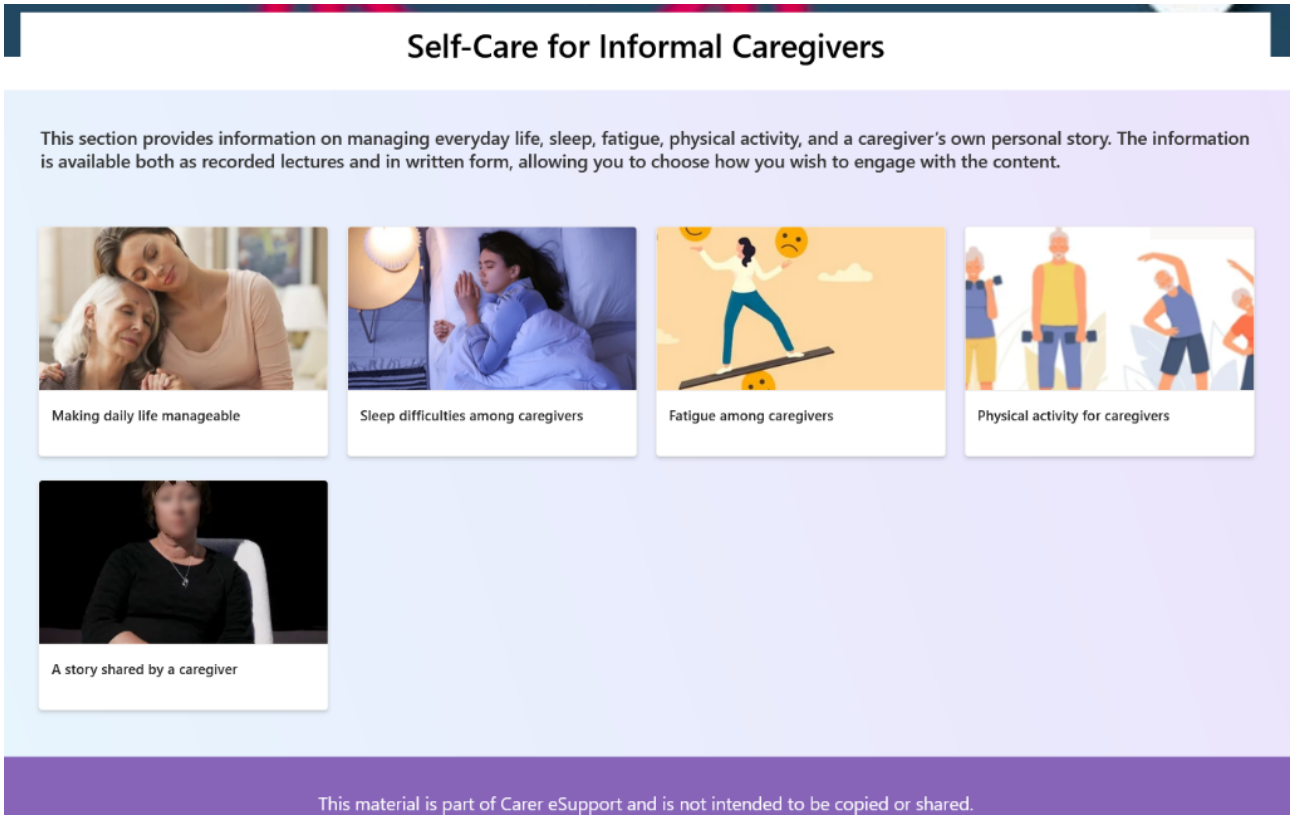
Self-Care for ICs

This module, as shown in Figure 4, supports the physical and psychological well-being of ICs by addressing everyday challenges such as managing routine tasks, sleep difficulties, fatigue, and the need for physical activity. The module is divided into 4 sections. It first provides strategies for adapting to life changes following an HNC diagnosis and then offers recommendations for overcoming common sleep issues. It also tackles exhaustion by introducing practical tools like a diary and a daily activity log to help balance responsibilities and self-care. It also emphasizes the benefits of maintaining an active lifestyle through regular physical activity.

Additionally, a personal story from an IC offers valuable insight into the emotional and physical challenges of caregiving, which may create empathy and a more profound understanding among ICs. The module promotes a eudaimonic perspective by strengthening ICs' sense of competence and personal growth. Providing ICs with practical knowledge and strategies helps them feel more confident and capable, which may contribute to their well-being and resilience.

XSL•FO
RenderX

Figure 4. Module for informal caregivers' self-care.



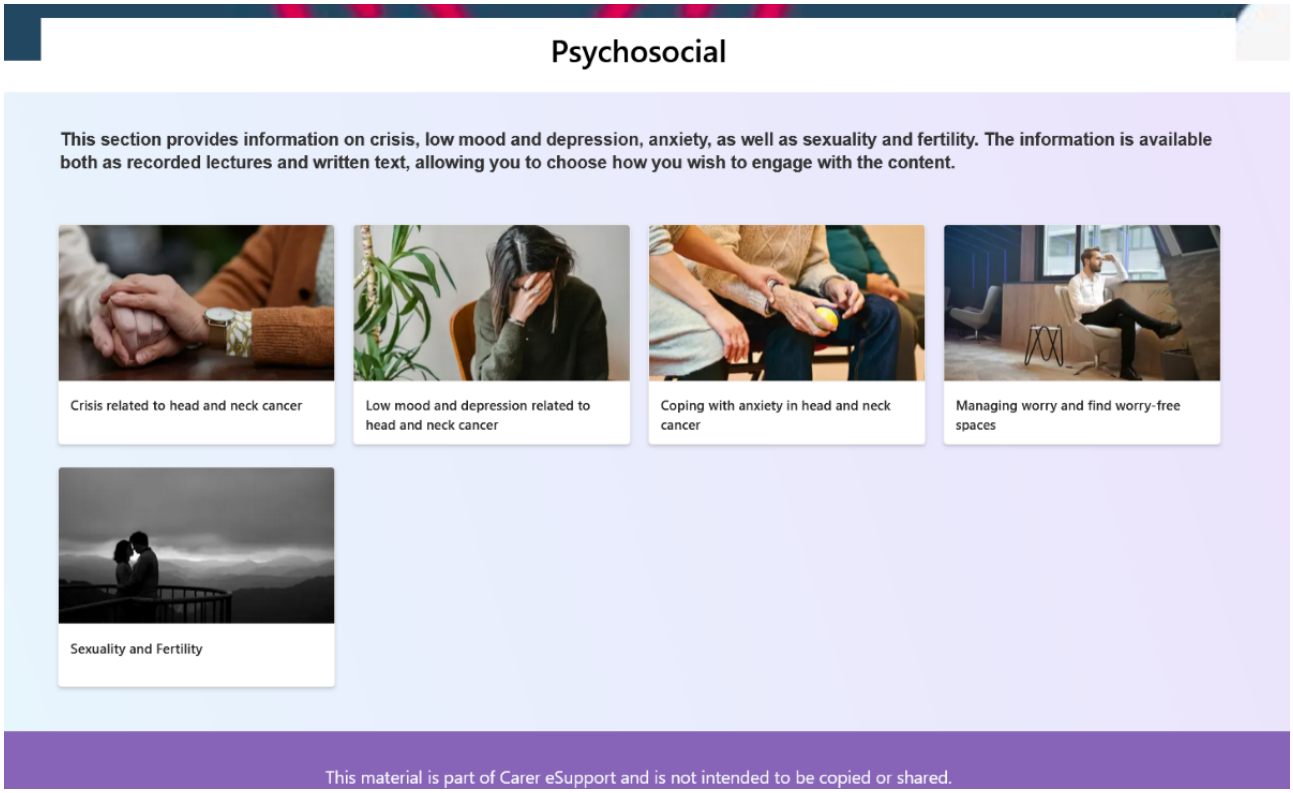
Psychosocial Support

The Psychosocial Support module, shown in Figure 5, is designed to enhance the emotional and psychological well-being of both ICs and patients by offering practical strategies for managing emotional health. This module is divided into 5 sections. It first addresses crisis reactions by providing coping strategies for the immediate psychological challenges during diagnosis and treatment. It then explores sadness and depression, offering guidance on recognizing and managing these emotions effectively. The following sections focus on anxiety, presenting practical strategies for managing worry and stress while also offering advice to create calm environments and concluding with an examination of the impact of cancer on sexuality and fertility, providing support for those affected by these issues.

From a eudaimonic perspective, each section is intentionally crafted to promote personal growth and to effectively prepare ICs to navigate the psychological challenges that accompany significant life changes. Beyond aiding emotional regulation, the module aims to empower ICs with a deeper sense of meaning and control in their roles. By reframing their experiences as opportunities for personal development, this approach may enhance resilience and improve their ability to cope with the demands of caregiving.

XSL•FO
RenderX

Figure 5. Module for psychosocial support.



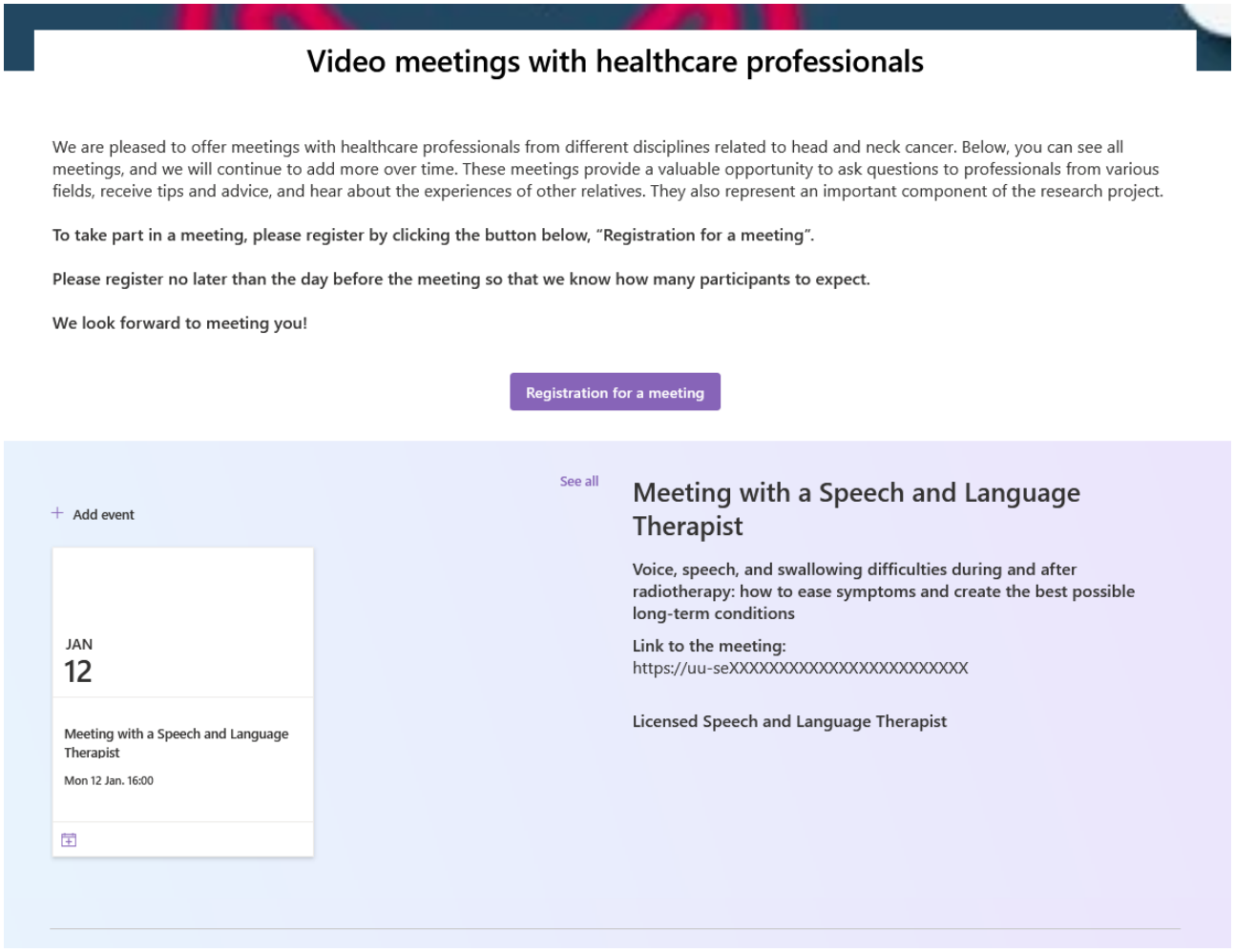
Video Meetings With HCPs

This module allows participants to meet with HCPs from various fields related to HNC (Figure 6). These sessions allow ICs to ask questions, receive expert advice, and connect with others facing similar challenges. The module offers flexible participation options by allowing ICs to join meetings with or without video and remain anonymous if preferred, ensuring better accessibility and alignment with a eudaimonic perspective. This flexibility caters to diverse comfort levels with technology and fosters a safe space for open engagement.

Recognizing that meaningful community relationships are central to eudaimonic well-being, the design of these virtual sessions encourages ICs to share their experiences, ask questions, and engage in group discussions. Such interactions provide practical guidance while nurturing deep interpersonal connections that enhance emotional support and validation. Occasionally, sessions include short presentations or lectures, with details provided in advance on the application. ICs are encouraged to attend multiple meetings to reinforce learning and maintain access to relevant information. Additionally, they can email questions to HCPs anytime, receiving responses from nurses or other specialists, further strengthening their support system throughout caregiving.

XSL•FO
RenderX

Figure 6. Module for video meetings with health care professionals.

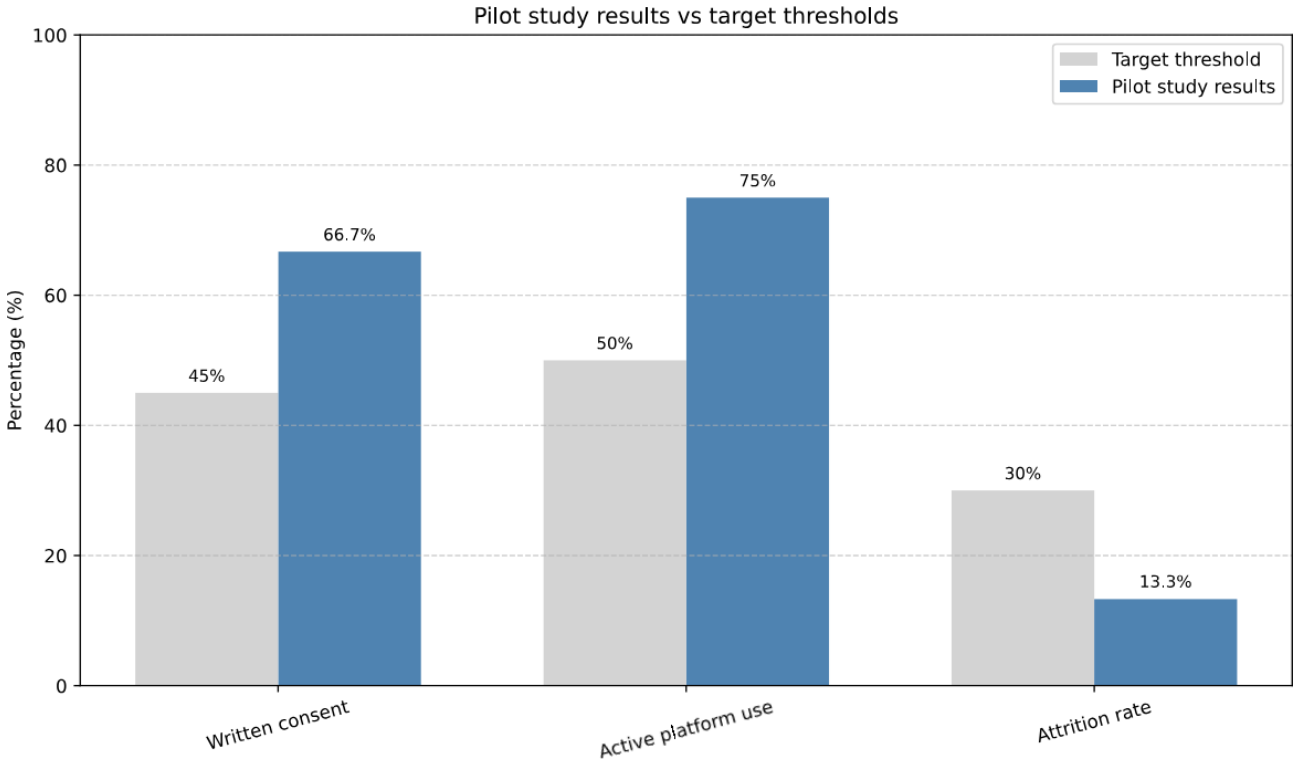


Result From the Pilot Study

The results from the pilot study were satisfactory and met the established progression criteria with a good margin (Figure 7). Among the 30 participants, 20 (66.7%) ICs and 17 (56.7%) patients provided written consent. A total of 12 ICs were randomized to receive access to the application, and 9 out of these 12 (75%) successfully accessed the application. Additionally, 4 out of 30 (13.3%) participants discontinued their study participation.

XSL•FO
RenderX

Figure 7. Pilot study results.



Challenges and Suggestions to Improve Usability and Accessibility of Carer eSupport

Overview

One week after sending the login details, we contacted participants who had not yet logged in to determine whether they required assistance. In this section, we summarize feedback from ICs regarding their experiences with Carer eSupport. Key challenges included time constraints due to the demanding nature of caregiving, communication barriers, login and authentication difficulties, limited IT knowledge, and inadequate access to proper IT infrastructure. To address these challenges and improve the application's usability and accessibility, we implemented a range of solutions, as detailed below.

Demanding Nature of Caregiving

Due to their ongoing responsibilities, many ICs struggled to find the time and energy to learn new technology. Some of them noted that when their relatives were recently diagnosed with HNC and undergoing treatment, they had little opportunity to engage with the application. Follow-up contacts acted as gentle reminders, prompting them to log in once their situations stabilized. For example, one participant explained that although they could not access the application initially due to pressing concerns, a subsequent reminder enabled them to log in after their circumstances had calmed.

Communication Problems

Communication barriers with ICs also hindered their engagement with the application. Some ICs reported not receiving the initial email with their login details or finding it in their spam folder, which delayed access. To resolve this, we implemented a multichannel communication strategy by sending login details via email and SMS text messaging, making follow-up phone calls to clarify any issues, and offering real-time troubleshooting. It reduced delays in accessing the application and helped build trust with the users.

Login and Authentication Problems

Several ICs experienced difficulties logging into the application. Since Carer eSupport is built on Microsoft SharePoint, users with existing Microsoft accounts sometimes encounter conflicts, leading to errors when the wrong account is used for authentication. In addition, the 2-factor authentication process proved confusing for many users. To address these challenges, we updated the login instructions and tutorials and produced a video tutorial demonstrating the correct login procedure step by step. These improvements reduced errors to some extent for the users, and more users accessed the application without help from the support team.

Access to IT Infrastructure

Some participants faced challenges due to inadequate access to appropriate devices for using Carer eSupport. A few lacked a private computer and relied on work computers or smartphones, which sometimes imposed restrictions that prevented them from logging in. One participant, who relied solely on a smartphone, repeatedly had difficulty entering their username and password, resulting in failed login attempts and further complicating access. As we were unable to provide additional IT infrastructure during the study, we addressed these challenges through extensive telephone-based support. With this support, ICs were eventually able to access the portal. However, access would likely have been more straightforward if a computer or tablet had been available.

XSL•FO
RenderX

Limited IT Knowledge

The limited IT knowledge of ICs emerged as a barrier, particularly among older ICs or those with minimal technical experience. Navigating a new digital application can be overwhelming for ICs who are not tech-savvy. To mitigate this problem, we provided them with telephone support and developed clear and concise tutorials, demonstrating the login process and easy access to the application.

Discussion

Principal Findings

This study introduced an integrated HCD process that incorporated eudaimonic perspectives to develop the Carer eSupport application to improve ICs' preparedness for caregiving and enhance their well-being. Our findings demonstrate that IT applications like Carer eSupport can be effectively developed through a multidisciplinary approach that integrates health care science and HCI methodologies. As a web-based application, Carer eSupport required complementary insights from HCI, mainly through an iterative HCD process, to ensure usability and user acceptance. This integration bridges conventional health care research with HCI, developing evidence-based and user-centered technologies, as supported by recent research [56]. To our knowledge, the Carer eSupport project is the first to use this strategy to support ICs of individuals with HNC.

A well-known challenge in HCI, and one central to this work, is balancing personalization with generalization [27,57]. Technologies that reflect ICs' own situations are often perceived as more useful and engaging. For this reason, the HCD approach used to develop Carer eSupport focused on understanding ICs' specific needs and caregiving contexts related to HNC. At the same time, practical constraints related to time, resources, and technical feasibility limit the degree of personalization that can be achieved [28,45]. HCI research recognizes that users may differ in their abilities, preferences, values, and life circumstances. HCD processes, therefore, begin with methods such as interviews and focus groups to identify human needs, which are then translated into requirements, personas, and scenarios. However, it is not feasible to design a separate system for each IC. Instead, design efforts focus on groups of ICs that share key characteristics while allowing variation within those groups. This tension was evident during the design of Carer eSupport. ICs preferred content reflecting their own HNC caregiving experiences, while HCPs emphasized the need for practical, sustainable solutions. The final design represents a pragmatic compromise, offering HNC-specific content at the user-group level while remaining relevant to a broad range of caregiving situations.

In this study, we presented the major modules and functions of the Carer eSupport application, explicitly incorporating the factors that may enhance the eudaimonic motivation of ICs. This was a challenging task due to the existing research's broad and often vague definitions of eudaimonia. Previous research also highlighted that eudaimonic well-being is inherently subjective and can vary significantly from one individual to another [49,58]. For future research, it is essential to contextualize eudaimonic perspectives according to the user group and the contexts in which they interact with technology [17,43,49]. Otherwise, as noted in previous studies, eudaimonia risks becoming a buzzword without precise, practical implementation, despite its importance in technology adoption [27,41].

Our pilot study yielded promising results, with a 66.7% consent rate, a 75% successful login rate among randomized participants, and a low attrition rate of 13.3%. These outcomes met established progression criteria and confirmed the application's readiness for a full-scale RCT. The study also provided important insights into usability challenges. ICs reported limited time due to demanding caregiving responsibilities, delays in receiving login details, difficulties with authentication, restricted access to suitable IT infrastructure, and limited digital literacy.

We provided multichannel communication via email, SMS text messaging, and follow-up phone calls to overcome these challenges. We also enhanced login instructions, created video tutorials, and provided additional telephone support. All these measures improved the application's overall usability and accessibility, especially for first-time access to the application. The lessons learned from this study will be valuable in a broader context, contributing to the development and evaluation of complex interventions and to the refinement of RCT procedures for future web-based interventions like Carer eSupport.

As an important first step in the HCD process, understanding the user's context and the use of technology is crucial for the smooth implementation of IT systems [33,59]. During the evaluation phase of an HCD process, real-world challenges often emerge that were not apparent at the start. For example, while assisting two ICs in logging into an application via smartphone, we discovered that they were attempting to juggle multiple tasks on a single device. The ICs were on a phone call and, at the same time, opened their email to look up their username and password. Afterward, they tried to access the application in another tab on the same smartphone. This multitasking led to repeated errors when entering their login credentials. The 2-factor authentication process, which requires ICs to retrieve and enter a code on the same smartphone, further complicates their access. When the support team asked whether they had access to an alternative device, such as a tablet or computer, the ICs said they had no other device available. After several attempts and extensive assistance from the support team, the ICs finally managed to log in to the application; however, it was not an easy process for them. Such cases highlight that while understanding user context is emphasized early in the HCD process, vital insights often emerge during evaluation. Such experiences, as highlighted in previous research, underscore the importance of iterative design improvements to ensure systems are more effectively aligned with users' real-life circumstances [59,60].

Our integration of the HCD process and health care science, enriched by a eudaimonic perspective, offers initial guidelines for creating a framework that can be applied beyond the caregiving context. For technology designers and HCI researchers, especially those focusing on vulnerable groups, this study can inform the design and development of digital

XSL•FO
RenderX

solutions that are both evidence-based and empathetic. However, it is worth noting that the study is limited to the pilot study. A complete evaluation of Carer eSupport can only be concluded once the full-scale RCT is completed. We look forward to exploring the full RCT outcomes, particularly regarding improvements in ICs' preparedness for caregiving and their well-being.

Conclusion

This paper presented the design and development of a web-based application that integrates eudaimonic perspectives within an iterative HCD process to prepare ICs for caregiving and improve their well-being. The pilot study results confirmed the application's viability for an RCT, as evidenced by high consent, login, and retention rates. However, it highlighted significant user challenges, such as time constraints and technology-related barriers. Integrating the eudaimonic perspective into the application design also proved challenging due to the subjective and nuanced nature of the elements of eudaimonic motivation. Given the limited participant pool in this study, a more comprehensive evaluation is necessary to assess the relevance, effectiveness, and usability of Carer eSupport. This evaluation is currently underway in our ongoing RCT. Results from this study indicate that personalized, interactive web-based applications like Carer eSupport hold considerable potential to support ICs of individuals with HNC.

Acknowledgments

We are thankful to the informal caregivers for sharing their valuable perspectives. Their active participation contributed to the design and development of Carer eSupport and the study procedures. We also extend our gratitude to the health care professionals, as their insights and support were of great help in shaping Carer eSupport. All text in this article was originally written by the authors. However, an AI tool (ChatGPT, version 5.2) was used only for language refinement and grammatical editing, without contributing to the content, analysis, or conclusions.

Funding

The entire project is funded by the Swedish Research Council (grant 2019-01231), the Swedish Cancer Society (grant 20 1014 PjF), the Swedish state through the ALF agreement between the Swedish government and the county councils (Uppsala County: ALF-941900), and the Oncology Department Foundation's Research Fund in Uppsala. The funding bodies had no involvement in the study design, data collection, data analysis, interpretation of data, or manuscript preparation.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Ahmad A, Mozelius P. Human-computer interaction for older adults - a literature review on technology acceptance of eHealth systems. J Eng Res Sci 2022 Apr;1(4):119-126. [doi: 10.55708/js0104014]
2. Wheatley MC. Adapting health systems for an aging world: policy innovations and global strategies. Prem J Public Health 2024;24:362. [doi: 10.70389/PJPH.100010]
3. Langegård U, Cajander Å, Carlsson M, et al. Internet-based support for informal caregivers to individuals with head and neck cancer (Carer eSupport): a study protocol for the development and feasibility testing of a complex online intervention. BMJ Open 2022 May 27;12(5):e057442. [doi: 10.1136/bmjopen-2021-057442] [Medline: 35623759]
4. Kent EE, Dionne-Odom JN. Population-based profile of mental health and support service need among family caregivers of adults with cancer. J Oncol Pract 2019 Feb;15(2):e122-e131. [doi: 10.1200/JOP.18.00522] [Medline: 30496020]
5. Hudson P, Payne S. The future of family caregiving: research, social policy and clinical practice. In: Family Carers in Palliative Care: A Guide for Health and Social Care Professionals: Oxford University Press; 2009:277-303 URL: https://pure.qub.ac.uk/en/publications/the-future-of-family-caregivingresearch-social-policy-and-clinica [accessed 2026-03-16]
6. Gérain P, Zech E. Informal caregiver burnout? Development of a theoretical framework to understand the impact of caregiving. Front Psychol 2019;10:1748. [doi: 10.3389/fpsyg.2019.01748] [Medline: 31428015]
7. Tang C, Chen Y, Cheng K, Ngo V, Mattison JE. Awareness and handoffs in home care: coordination among informal caregivers. Behav Inf Technol 2018 Jan 2;37(1):66-86. [doi: 10.1080/0144929X.2017.1405073]
8. Adams E, Boulton M, Watson E. The information needs of partners and family members of cancer patients: a systematic literature review. Patient Educ Couns 2009 Nov;77(2):179-186. [doi: 10.1016/j.pec.2009.03.027] [Medline: 19406609]
9. Karampela M, Porat T, Mylonopoulou V. Needs of head and neck cancer patients and stakeholders during rehabilitation. Presented at: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2019; Trento, Italy p. 415-421. [doi: 10.1145/3329189.3329236]

XSL•FO
RenderX

10. Schuit AS, Holtmaat K, Hooghiemstra N, et al. Efficacy and cost-utility of the eHealth self-management application "Oncokompas", helping partners of patients with incurable cancer to identify their unmet supportive care needs and to take actions to meet their needs: a study protocol of a randomized controlled trial. Trials 2020 Jan 31;21(1):124. [doi: 10.1186/s13063-019-4037-5] [Medline: 32005280]
11. Wang T, Mazanec SR, Voss JG. Needs of informal caregivers of patients with head and neck cancer: a systematic review. Oncol Nurs Forum 2021 Jan 4;48(1):11-29. [doi: 10.1188/21.ONF.11-29] [Medline: 33337440]
12. Slev VN, Mistiaen P, Pasman HRW, Verdonck-de Leeuw IM, van Uden-Kraan CF, Francke AL. Effects of eHealth for patients and informal caregivers confronted with cancer: a meta-review. Int J Med Inform 2016 Mar;87:54-67. [doi: 10.1016/j.ijmedinf.2015.12.013] [Medline: 26806712]
13. Heins M, Schellevis F, Rijken M, Donker G, van der Hoek L, Korevaar J. Partners of cancer patients consult their GPs significantly more often with both somatic and psychosocial problems. Scand J Prim Health Care 2013 Dec;31(4):203-208. [doi: 10.3109/02813432.2013.861153] [Medline: 24299045]
14. Applebaum AJ, Breitbart W. Care for the cancer caregiver: a systematic review. Palliat Support Care 2013 Jun;11(3):231-252. [doi: 10.1017/S1478951512000594] [Medline: 23046977]
15. Langegård U, Cajander Å, Ahmad A, et al. Understanding the challenges and need for support of informal caregivers to individuals with head and neck cancer - a basis for developing internet-based support. Eur J Oncol Nurs 2023 Jun;64:102347. [doi: 10.1016/j.ejon.2023.102347] [Medline: 37290161]
16. Köhle N, Drossaert CH, Oosterik S, et al. Needs and preferences of partners of cancer patients regarding a web-based psychological intervention: a qualitative study. JMIR Cancer 2015 Dec 29;1(2):e13. [doi: 10.2196/cancer.4631] [Medline: 28410157]
17. Ahmad A, Cajander Å, Johansson B, Tiblom Ehrsson Y, Langegård U. Designing for human well-being: a case study with informal caregivers of individuals with cancer. In: Challenges of Trustable AI and Added-Value on Health: IOS Press; 2022:214-218. [doi: 10.3233/SHTI220440]
18. Plöthner M, Schmidt K, de Jong L, Zeidler J, Damm K. Needs and preferences of informal caregivers regarding outpatient care for the elderly: a systematic literature review. BMC Geriatr 2019 Mar 13;19(1):82. [doi: 10.1186/s12877-019-1068-4] [Medline: 30866827]
19. Mazanec SR, Reichlin D, Gittleman H, Daly BJ. Perceived needs, preparedness, and emotional distress of male caregivers of postsurgical women with gynecologic cancer. Oncol Nurs Forum 2018 Mar 1;45(2):197-205. [doi: 10.1188/18.ONF.197-205] [Medline: 29466348]
20. Freund M, Carey M, Dilworth S, et al. Effectiveness of information and communications technology interventions for stroke survivors and their support people: a systematic review. Disabil Rehabil 2022 Aug;44(17):4563-4578. [doi: 10.1080/09638288.2021.1913245] [Medline: 33905279]
21. Young R, Willis E, Cameron G, Geana M. "Willing but unwilling": attitudinal barriers to adoption of home-based health information technology among older adults. Health Informatics J 2014 Jun;20(2):127-135. [doi: 10.1177/1460458213486906] [Medline: 24056750]
22. Ahmad A, Mozelius P, Ahlin K. Factors influencing acceptance of technology-enhanced speech and language relearning for stroke survivors - a systematic review. Presented at: ICSLT 2021; Jun 10-12, 2021. [doi: 10.1145/3477282.3477285]
23. Maguire M. Methods to support human-centred design. Int J Hum Comput Stud 2001 Oct;55(4):587-634. [doi: 10.1006/ijhc.2001.0503]
24. Cajander Å, Lárusdóttir MK, Lind T, Nauwerck G. Walking in the jungle with a machete: ICT leaders' perspectives on user-centred systems design. Behav Inf Technol 2022 Apr 26;41(6):1230-1244. [doi: 10.1080/0144929X.2020.1864776]
25. Gulliksen J, Göransson B, Boivie I, Blomkvist S, Persson J, Cajander Å. Key principles for user-centred systems design. Behav Inf Technol 2003 Nov;22(6):397-409. [doi: 10.1080/01449290310001624329]
26. Bevilacqua R, Bailoni T, Maranesi E, et al. Framing the human-centered artificial intelligence concepts and methods: scoping review. JMIR Hum Factors 2025 May 28;12:e67350. [doi: 10.2196/67350] [Medline: 40435517]
27. Stephanidis C, Salvendy G, Antona M, et al. Seven HCI grand challenges. Int J Hum–Comput Interact 2019 Aug 27;35(14):1229-1269. [doi: 10.1080/10447318.2019.1619259]
28. van Velsen L, Ludden G, Grünloh C. The limitations of user-and human-centered design in an eHealth context and how to move beyond them. J Med Internet Res 2022 Oct 5;24(10):e37341. [doi: 10.2196/37341] [Medline: 36197718]
29. Türkay S, Lin A, Johnson D, Formosa J. Self-determination theory approach to understanding the impact of videogames on wellbeing during COVID-19 restrictions. Behav Inf Technol 2023 Aug 18;42(11):1720-1739. [doi: 10.1080/0144929X.2022.2094832]
30. Desmet PMA, Pohlmeyer AE. Positive design: an introduction to design for subjective well-being. Int J Des 2013;7(3) [FREE Full text]
31. Grundgeiger T, Hurtienne J, Happel O. Why and how to approach user experience in safety-critical domains: the example of health care. Hum Factors 2021 Aug;63(5):821-832. [doi: 10.1177/0018720819887575] [Medline: 31914323]
32. Peters D, Calvo RA, Ryan RM. Designing for motivation, engagement and wellbeing in digital experience. Front Psychol 2018;9:797. [doi: 10.3389/fpsyg.2018.00797] [Medline: 29892246]

XSL•FO
RenderX

33. Szalma JL. On the application of motivation theory to human factors/ergonomics: motivational design principles for human-technology interaction. Hum Factors 2014 Dec;56(8):1453-1471. [doi: 10.1177/0018720814553471] [Medline: 25509824]
34. Zhang P. Toward a positive design theory: principles for designing motivating information and communication technology. In: Designing Information and Organizations with a Pasitive Lens: Elsevier; 2008:45-74 URL: https://www.researchgate.net/publication/257308474_Toward_a_Positive_Design_Theory_Principles_for_Designing_Motivating_Information_and_Communication_Technology [accessed 2026-03-16]
35. Kiefer RA. An integrative review of the concept of well-being. Holist Nurs Pract 2008;22(5):244-252. [doi: 10.1097/01.HNP.0000334915.16186.b2] [Medline: 18758272]
36. Deci EL, Ryan RM. Hedonia, eudaimonia, and well-being: an introduction. J Happiness Stud 2008 Jan;9(1):1-11. [doi: 10.1007/s10902-006-9018-1]
37. Huta V, Pelletier LG, Baxter D, Thompson A. How eudaimonic and hedonic motives relate to the well-being of close others. J Posit Psychol 2012 Sep;7(5):399-404. [doi: 10.1080/17439760.2012.705318]
38. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology1. MIS Q 2012 Mar 1;36(1):157-178. [doi: 10.2307/41410412]
39. Kuriakose M, Nagasubramaniyan G. Beyond pleasure, desire for meaningful consumption and peacefulness from digital entertainment platforms; extending UTAUT2 model with eudemonic motivation and tranquility. Int J Hum–Comput Interact 2025 Jan 2;41(1):792-806. [doi: 10.1080/10447318.2024.2305479]
40. Seaborn K, Pennefather P, Fels DI. Eudaimonia and hedonia in the design and evaluation of a cooperative game for psychosocial well-being. Hum–Comput Interact 2020 Jul 3;35(4):289-337. [doi: 10.1080/07370024.2018.1555481]
41. Sequeiros H, Oliveira T, Thomas MA. The impact of IoT smart home services on psychological well-being. Inf Syst Front 2022 Jun;24(3):1009-1026. [doi: 10.1007/s10796-021-10118-8]
42. Woźniak PW, Hak M, Kotova E, et al. Quantifying meaningful interaction: developing the eudaimonic technology experience scale. Presented at: DIS '23: Proceedings of the 2023 ACM Designing Interactive Systems Conference; Jul 10-14, 2023; Pittsburgh, PA, USA p. 1904-1914. [doi: 10.1145/3563657.3596063]
43. Jörs JM, De Luca EW. Through eudaimonia's lenses: eudaimonic well-being variables as guiding principles for AI design. Presented at: Designing (with) AI for Wellbeing (CHI 2024 Workshop); May 11-16, 2024 URL: https://www.researchgate.net/publication/380938771_Through_Eudaimonia's_Lenses_EudAImonic_Well-Being_Variables_As_Guiding_Principles_For_AI_Design [accessed 2026-03-16]
44. Johansson B, Cajander Å, Ahmad A, et al. The effect of internet-administered support (Carer eSupport) on preparedness for caregiving in informal caregivers of patients with head and neck cancer compared with support as usual: a study protocol for a randomized controlled trial. BMC Cancer 2024 Apr 18;24(1):494. [doi: 10.1186/s12885-024-12273-y] [Medline: 38637744]
45. Campese C, Amaral DC, Mascarenhas J. Restating the meaning of UCD And HCD for a new world of design theories. Interact Comput 2020 Jan 13;32(1):33-51. [doi: 10.1093/iwc/iwaa003]
46. Cristancho-Lacroix V, Moulin F, Wrobel J, et al. A web-based program for informal caregivers of persons with Alzheimer's disease: an iterative user-centered design. JMIR Res Protoc 2014 Sep 15;3(3):e46. [doi: 10.2196/resprot.3607] [Medline: 25263541]
47. Babione JN, Ocampo W, Haubrich S, et al. Human-centred design processes for clinical decision support: a pulmonary embolism case study. Int J Med Inform 2020 Oct;142:104196. [doi: 10.1016/j.ijmedinf.2020.104196] [Medline: 32947115]
48. ISO 9241-221:2023 – ergonomics of human-system interaction — part 221: human-centred design process assessment model. ISO. URL: https://www.iso.org/standard/81384.html [accessed 2025-02-20]
49. Ahmad A, Premanandan S, Langegård U, et al. Positive design framework for carer eSupport: qualitative study to support informal caregivers of patients with head and neck cancer in Sweden. JMIR Cancer 2023 May 30;9(1):e45748. [doi: 10.2196/45748] [Medline: 37252765]
50. Figma. URL: https://www.figma.com/ [accessed 2023-05-29]
51. Craig P, Dieppe P, Macintyre S, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655. [doi: 10.1136/bmj.a1655] [Medline: 18824488]
52. Hariton E, Locascio JJ. Randomised controlled trials – the gold standard for effectiveness research. BJOG 2018 Dec;125(13):1716-1716. [doi: 10.1111/1471-0528.15199] [Medline: 29916205]
53. Bhide A, Shah PS, Acharya G. A simplified guide to randomized controlled trials. Acta Obstet Gynecol Scand 2018 Apr;97(4):380-387. [doi: 10.1111/aogs.13309] [Medline: 29377058]
54. Avery KNL, Williamson PR, Gamble C, et al. Informing efficient randomised controlled trials: exploration of challenges in developing progression criteria for internal pilot studies. BMJ Open 2017 Feb 17;7(2):e013537. [doi: 10.1136/bmjopen-2016-013537] [Medline: 28213598]

XSL•FO
RenderX

55. Braun V, Clarke V. Thematic analysis. In: APA Handbook of Research Methods in Psychology: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological: American Psychological Association; 2023:65-81. [doi: 10.1037/0000319-004]
56. Kauppinen K, Keikhosrokiani P, Khan S. Human-centered design and benefit realization management in digital health care solution development: protocol for a systematic review. JMIR Res Protoc 2024 May 21;13:e56125. [doi: 10.2196/56125] [Medline: 38772023]
57. Tunis R, Fleischmann KR, Smith ADR. Designing for personalization in personal informatics: barriers and pragmatic approaches from the perspectives of designers, developers, and product managers. Presented at: DIS '24: Proceedings of the 2024 ACM Designing Interactive Systems Conference; Jul 1-5, 2024; Copenhagen, Denmark p. 584-596. [doi: 10.1145/3643834.3661622]
58. Joers J, De Luca EW. Examining the eudaimonic experience: differences between well-being orientations regarding motivation, engagement, and basic psychological needs in interaction with AI. Presented at: Proceedings of the 58th Hawaii International Conference on System Sciences; Jan 7-10, 2025. [doi: 10.24251/HICSS.2025.718]
59. Göttgens I, Oertelt-Prigione S. The application of human-centered design approaches in health research and innovation: a narrative review of current practices. JMIR Mhealth Uhealth 2021 Dec 6;9(12):e28102. [doi: 10.2196/28102] [Medline: 34874893]
60. Ahmad A, Brodén K, Mozelius P, Sodhro AH. Speech and language relearning for stroke patients-understanding user needs for technology enhancement. Presented at: eTELEMED 2021: The Thirteenth International Conference on eHealth, Telemedicine, and Social Medicine; Jul 18-22, 2021 URL: https://www.researchgate.net/publication/353346407_Speech_and_Language_Relearning_for_Stroke_Patients-Understanding_User_Needs_for_Technology_Enhancement [accessed 2026-03-16]

Abbreviations

HCD: human-centered design
HCI: human-computer interaction
HCP: health care professional
HNC: head and neck cancer
IC: informal caregiver
RCT: randomized controlled trial
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by A Kushniruk; submitted 05.Aug.2025; peer-reviewed by E Lee, V Bartali; revised version received 21.Feb.2026; accepted 28.Feb.2026; published 08.Apr.2026.

Please cite as:
Ahmad A, Cajander Å, Langegård U, Johansson B, Henriksson A, Pettersson M, Ahmad W, Tiblom Ehrsson Y
The Development of a Web-Based Application to Support Informal Caregivers of Individuals With Head and Neck Cancer: Human-Centered Design Approach
JMIR Hum Factors 2026;13:e81896
URL: https://humanfactors.jmir.org/2026/1/e81896
doi:10.2196/81896

© Awais Ahmad, Åsa Cajander, Ulrica Langegård, Birgitta Johansson, Anna Henriksson, Mona Pettersson, Waqas Ahmad, Ylva Tiblom Ehrsson. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 8.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Improving Usability of the Interrupting Prolonged Sitting With Activity Virtual Teacher Training Modules: Case Study

Rebecca E Hasson[1], PhD; Tiffany Tam[2], BS; Lucy Myers[2]; Anna Schwartz[1], MPH; Alissa Li[1], BS

[1]School of Kinesiology, University of Michigan, 830 North University Avenue, 4220 SKB, Ann Arbor, MI, United States

[2]School of Information, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Rebecca E Hasson, PhD
School of Kinesiology, University of Michigan, 830 North University Avenue, 4220 SKB, Ann Arbor, MI, United States

Abstract

Background: High-quality professional development can help teachers build the skills and confidence needed to implement evidence-based practices, such as classroom physical activity breaks. While in-person training is often preferred, virtual asynchronous training offers a flexible alternative for teachers. However, its effectiveness may be limited by design and usability challenges.

Objective: The aim of this study was to conduct a usability assessment of the Interrupting Prolonged Sitting with Activity (InPACT) virtual teacher training modules, using a human-centered design (HCD) approach to align the training with end-user preferences and needs.

Methods: The InPACT professional development program includes nine modules delivered through an online platform (Qualtrics XM). A usability assessment was conducted using (1) structured online surveys with elementary school teachers who had completed the modules, (2) a heuristic evaluation based on the Jakob Nielsen 10 usability heuristics, and (3) a competitive analysis of three learning management systems (Moodle, Teachable, and Thinkific) to identify platform strengths, limitations, and insights. Findings and recommendations were compiled to inform module improvements.

Results: Eleven teachers completed the survey. They reported that the modules were easy to use, the content was informative and relevant, and they valued the interactive and practical components. Areas for improvement included enhancing content engagement and increasing technical flexibility. The heuristic evaluation identified 14 usability issues across nine of the Nielsen heuristics (eg, navigation difficulties and text-heavy pages). The competitive analysis highlighted features that enhance user experience, such as progress tracking, estimated completion times, interactive elements, and feedback on quiz answers.

Conclusions: Usability assessments grounded in an HCD approach can enhance virtual training for educators, improving the uptake and implementation of evidence-based practices, such as classroom physical activity breaks. Five overarching recommendations emerged as follows: (1) removing video time constraints, (2) implementing accurate progress bars, (3) incorporating active learning or retention activities, (4) creating consistent and actionable end pages, and (5) ensuring consistency in titles and references to physical materials. Building on preliminary positive data from revised modules, future research should evaluate the impact of HCD revisions on teacher self-efficacy, training completion, and fidelity of program implementation.

(JMIR Hum Factors 2026;13:e83427) doi:10.2196/83427

KEYWORDS

user experience; website; implementation science; human-centered design; HCD; web-based; interview; heuristic; competitive analysis; classroom physical activity; teachers

Introduction

Physical activity plays a vital role in promoting child and adolescent health by improving cognition, bone health, fitness, and overall health, while also reducing the risk of anxiety and depression [1]. However, most children and adolescents in the United States do not meet the Physical Activity Guidelines for Americans, which recommend at least 60 minutes of moderate-to-vigorous physical activity daily to attain the greatest health benefits [1]. National surveillance data show that only 42% of children aged 6 to 11 years meet this recommendation [2], with participation declining sharply during adolescence [3] and rates being even lower for youth residing in resource-limited communities [4]. This shortfall increases the risk of obesity and related chronic conditions [5,6], highlighting the need to improve youth physical activity engagement.

Schools offer a critical and equitable setting to help close the physical activity gap among children and adolescents [7,8]. Among school-based strategies, classroom-based physical activity interventions have emerged as effective and feasible

XSL•FO
RenderX

approaches for increasing students' daily activity levels, particularly in resource-limited settings [9,10]. These interventions provide structured opportunities for all children to engage in regular exercise by integrating short bouts of movement into the school day. Compared to physical education and after-school sports programs, they are less resource-intensive and easier to implement, making them well-suited for schools with limited staff, space, or funding [11,12].

Recognized as an evidence-based practice by the Community Preventive Services Task Force in 2021, classroom physical activity interventions typically involve teachers incorporating 1 to 5 activity breaks per day, each lasting 4 to 10 minutes [13]. Core implementation components include dedicated time, adequate classroom space, and access to age-appropriate videos and exercise resources. A systematic review and meta-analysis by the Task Force, which included 22 studies (11 using direct observation), found that physical activity breaks significantly increased students' activity levels during school hours and improved time-on-task by a median of 26.2 minutes [10]. These findings provide strong evidence that classroom physical activity interventions can simultaneously promote physical activity and enhance classroom engagement.

Despite the well-established benefits, national data show that fewer than half of US elementary teachers incorporate these breaks regularly [14,15], underscoring the need to better understand barriers and facilitators to their implementation. While many teachers recognize these benefits—such as increased student focus, enjoyment, and a more positive classroom climate—they often experience a lack of confidence, uncertainty about how to integrate activity without disrupting instruction, and the belief that physical activity is not aligned with academic goals [16-21]. Key individual-level factors that influence implementation include teachers' knowledge, attitudes, self-efficacy, and perceptions of social norms [22-24]. Educators who feel confident in their ability to use physical activity, value its role in learning, and perceive support from peers and administrators are more likely to use activity breaks consistently [25-27]. Strengthening these individual factors is critical to expanding the reach and sustainability of classroom physical activity.

To overcome these barriers—particularly teachers' lack of confidence and uncertainty about how to integrate activity without disrupting learning—high-quality professional development is essential [28]. For physical activity integration to become routine, teachers need concrete, classroom-ready strategies that align with existing instructional practices [29]. The Interrupting Prolonged Sitting with Activity (InPACT) intervention addresses this need by providing a skills-based professional development program designed to support direct translation to classroom practice [30]. Through structured training, InPACT equips educators with the tools and strategies needed to integrate physical activity in ways that are feasible within daily classroom routines.

Grounded in the Guskey [28] model of teacher change, which emphasizes practice as a precursor to belief change, the InPACT approach prioritizes the adoption of new classroom behaviors before targeting shifts in attitudes or perceptions. As teachers observe improvements in student focus, classroom climate, and learning, these practice-based successes reinforce the feasibility and value of integrating movement into the school day, thereby supporting subsequent changes in beliefs and acceptability.

The educational strategies embedded within the InPACT training were intentionally selected to reflect adult learning principles and evidence-based models of effective teacher professional development [31,32]. These strategies include practical, classroom-ready examples, active skill-building, structured reflection, and explicit connections between training content and teachers' daily instructional practices. The training was co-created with the University of Michigan School of Education's Instructional and Program Design Coordinator, ensuring alignment with evidence-based pedagogy and responsiveness to teacher-identified needs [30].

The InPACT professional development program consists of 9 interactive training modules that can be delivered either in-person or asynchronously via an online learning platform (Qualtrics XM). In person, the modules are delivered by a trained InPACT professional using a Microsoft PowerPoint presentation, with interactive opportunities for hands-on activities, modeling of classroom activity breaks, and group discussions. In-person InPACT training sessions typically last 6 to 8 hours. The virtual format consists of asynchronous modules combining text, videos, images, and interactive exercises, with each module taking approximately 45 to 60 minutes to complete. Teachers are encouraged to complete all modules within a 4- to 6-week period to support continuity and application to classroom practice. Participants engage in guided self-reflection activities and complete premodule and postmodule assessments to track learning progress. Table 1 provides a description of each training module.

XSL•FO
RenderX

Table . Summary of the Interrupting Prolonged Sitting with Activity (InPACT) virtual teacher training modules.

Module	Title	Description
1	Evidence base for classroom physical activity	Introduces the evidence supporting classroom physical activity and its benefits for health, well-being, and academic performance.
2	Overview of InPACT	Explains how to incorporate short movement breaks (InPACT) into the school day.
3	Optimizing classroom layouts	Guides teachers in creating safe and functional spaces for movement in the classroom.
4	Classroom management during activity breaks	Focuses on maintaining structure and engagement during physical activity breaks.
5	Integrating physical activity with academic content	Helps teachers align movement with learning goals to support academic achievement.
6	Selecting exercise videos and assessing intensity	Covers selecting appropriate videos and assessing student intensity levels to meet physical activity goals.
7	Motivational techniques and gamification	Introduces strategies like gamification to enhance and sustain student engagement.
8	Inclusive adaptation of physical activity	Provides strategies to adapt activities to ensure accessibility and inclusion for all students.
9	Safety guidelines	Emphasizes safety considerations and offers best practices for injury prevention and safety planning.

After developing this content-rich program, we sought to understand teachers' preferences for how these modules could best be delivered. Findings from a group of general education teachers (n=35, teaching experience ranging from 2 to 27 years; mean 14.8, SD 7.4) highlighted the importance of offering flexible training options for classroom physical activity implementation. While nearly two-thirds of teachers expressed a preference for in-person professional development, a substantial portion—approximately one-third—preferred a virtual, asynchronous training option. These teachers cited competing demands, stress, and logistical challenges (such as preparing substitute teacher plans) as barriers to in-person participation (Friday et al, PhD, unpublished data, 2021). Providing autonomy in training modality—allowing educators to select either virtual or in-person formats—could therefore enhance participation and reduce perceived burden, particularly if training is integrated into existing professional development sessions.

Offering virtual training may increase accessibility for teachers facing time constraints and logistical challenges. However, it also introduces a new set of considerations—particularly regarding the usability of the digital learning environment. Ensuring that these online modules are accessible, understandable, and engaging is essential to addressing the very barriers—such as stress and overload—that teachers have identified.

Teachers' ability to benefit from online modules depends on the usability and design quality of the digital learning environment [33]. Usability is defined as "a quality attribute assessing how easy user interfaces are to use, focusing on 5 core components: learnability, efficiency, memorability, error handling, and satisfaction" [34]. Research shows that poorly designed websites can hinder engagement, leading to frustration and high dropout rates [35]. Conversely, user-friendly platforms can promote greater interaction and retention [33]. To address this, we conducted a usability assessment of the InPACT virtual teacher training modules to evaluate how accessible, understandable, and engaging they are for classroom educators. Ultimately, improving the usability of the InPACT virtual modules can empower more educators to confidently and sustainably implement physical activity breaks, regardless of the training format they choose.

Methods

Overview

This study used a mixed methods approach grounded in a human-centered design (HCD) framework to evaluate the usability and delivery of the InPACT virtual teacher training modules, with an emphasis on understanding the end-user needs, constraints, and contexts prior to large-scale implementation, consistent with established HCD models for digital health interventions [36]. Within the HCD lifecycle, this study focused on formative inspiration and ideation stages to inform iterative refinement of the training modules, operationalized through complementary methods, including a structured online survey of teachers, a heuristic usability evaluation, and a competitive analysis of learning management systems (LMSs). HCD served as the overarching, iterative, and participatory framework, integrating principles from user-centered design, which prioritizes usability and functionality from the end-user perspective, and user experience (UX) design, which focuses on optimizing the quality of users' interactions with a system. Together, these principles informed design decisions related to the structure, usability, and delivery of the virtual modules.

XSL•FO
RenderX

Later stages of HCD, including implementation and outcome evaluation, were beyond the scope of this study and are the focus of ongoing and future work.

Recruitment

Elementary school teachers who had completed at least one InPACT virtual module were recruited for this study. Principals in Michigan schools currently implementing InPACT provided the contact information of teachers who participated in the training. Eligibility was determined through these principal-provided contact lists. Eligible teachers were invited via email to participate in the research, with a link to a structured online survey. Participants could opt out at any time.

Structured Online Survey

A structured online survey was administered to evaluate teachers' experiences with the InPACT intervention. Developed by the research team, the survey included both closed-ended and open-ended items to capture quantitative indicators of module use and qualitative feedback on UX. The first section gathered demographic information, such as the grade level taught and years of teaching experience. The second section assessed teachers' experiences with the InPACT modules, including challenges encountered, clarity of module content, perceived strengths and weaknesses, preferred format of the program guide (ie, the written version of the modules; physical book vs digital pdf), use of supplementary links, suggested improvements, and frequency of classroom activity implementation following training.

All participants completed the same set of structured questions, designed to take approximately 15 to 20 minutes. Responses were collected anonymously and analyzed by the research team. Participants were offered the chance to enter a US $25 gift card raffle as an incentive for completing the survey. The full survey instrument, including all closed-ended and open-ended items, is provided in Multimedia Appendix 1.

Heuristic Evaluation

A heuristic evaluation was conducted to assess the usability of the InPACT virtual teacher training modules. The evaluation team consisted of 5 evaluators with diverse academic backgrounds and varying levels of expertise in UX design. UX design is defined as "the process of designing digital or physical products with a focus on optimizing users' overall experience, including usability, accessibility, and satisfaction" [37]. Some evaluators had completed formal coursework in UX design, while others contributed complementary perspectives, including spatial and visual design and applied analytics. All evaluators had prior hands-on experience conducting usability evaluations through Michigan Open UX, the University of Michigan's student organization focused on applied UX research and practice.

Heuristic evaluation is a cost-effective usability inspection method in which evaluators assess an interface against established usability principles, or heuristics, to identify potential design issues [38,39]. This evaluation was guided by the Jakob Nielsen 10 usability heuristics [40], which are grounded in cognitive psychology and human-computer interaction principles. Each evaluator independently completed the virtual training modules and documented any usability issues encountered. Identified issues were mapped to the relevant heuristic and assigned a severity rating ranging from 0 ("no usability problem") to 4 ("usability catastrophe").

Competitive Analysis

A competitive analysis was conducted using a FigJam board (Figma, Inc) to generate strategic insights into potential alternative LMS platforms that could support the scalability and long-term delivery of the InPACT training program. As the program continues to expand, evaluating LMS options beyond its current Qualtrics-based delivery is essential for optimizing UX and infrastructure sustainability.

Three LMS platforms were selected for evaluation: Moodle (version 5.0), Teachable (version 3.3.0), and Thinkific (version 1.17). These platforms were selected based on their relevance to professional education delivery, scalability potential, and cost considerations. Each platform was evaluated through hands-on exploration from both the lecturer and learner perspectives. From the lecturer's perspective, evaluators examined features related to course creation, content uploading, learner analytics, and collaboration among multiple contributors. From the learner's perspective, the evaluation focused on progress tracking, feedback mechanisms, navigation, and alignment with usability heuristics such as visibility of system status (eg, progress bars and activity feedback) and user control and freedom (eg, back and exit buttons). Additional features relevant to both perspectives—including media integration, retention activities, and search or filtering capabilities—were also assessed.

The competitive analysis aimed to (1) identify LMS platforms that may better support InPACT's future delivery needs and (2) inform potential enhancements to the current system by identifying effective design features used in other platforms. The analysis was conducted iteratively, recognizing the evolving nature of LMS technologies and prioritizing transferable insights over the selection of a single replacement platform.

Data Analysis

Closed-ended survey items on the online survey (eg, yes or no responses and categorical frequency ratings) were summarized using descriptive statistics to characterize teacher demographics, module use, and perceived frequency of classroom activity implementation. Open-ended survey responses (eg, "What are some positives of the modules? Negatives?", "If there is something you could change about the modules, what would it be?", and "Is there anything else you'd like to add?") generated free-text responses that were treated as qualitative data and analyzed using thematic analysis. All free-text responses were independently reviewed by members of the evaluation team to identify recurring patterns related to usability, content clarity, engagement, and technical functionality. An iterative process was used to group similar responses, refine theme definitions, and reach consensus on final themes. Representative quotes were selected to illustrate each theme and provide contextual depth to the findings. To protect participant confidentiality, pseudonyms were used for all quotes presented.

XSL•FO
RenderX

Data from the heuristic evaluation were analyzed by compiling all usability issues identified by individual evaluators and categorizing them according to the Jakob Nielsen 10 usability heuristics. Severity scores were determined through group discussion and consensus and were used to prioritize design recommendations.

Competitive analysis data were analyzed through a structured cross-platform comparison of LMS features relevant to usability and instructional design, including course creation, UX, and interactivity. Cross-platform patterns were synthesized to identify design features that could inform the optimization of the InPACT teacher training modules.

Overall, data analysis integrated descriptive quantitative summaries with qualitative thematic analysis and usability-focused evaluative techniques.

Ethical Considerations

This study was approved by the institutional review board of the University of Michigan (HUM00192745). Prior to each interview, all participants provided verbal informed consent. Each participant was assigned a unique numerical identifier, and pseudonymized transcripts and personal data were stored separately on a secure Childhood Disparities Research Laboratory U-Drive, accessible only to authorized research team members. Participants received a US $25 gift card as compensation for their time.

Results

Structured Online Survey

A convenience sample of 11 schoolteachers (grade levels: kindergarten through 5th grade; teaching experience range: 3 - 28 years; mean 18.0, SD 8.9) completed the structured online survey. Overall, teachers reported minimal technical or content-related barriers when completing the InPACT teacher training modules. Only 1 teacher reported experiencing technical issues, and no teachers reported confusion with the module content.

Use of supplemental materials varied across respondents. Approximately half of the teachers reported using a written program guide (PDF or physical version of the modules) alongside the online training, whereas the remaining teachers relied exclusively on the digital modules. Engagement with embedded supplementary links was mixed: 4 teachers reported frequent use, while others reported occasional use (n=5) or no use (n=2), often noting that the core module content was sufficient without additional resources.

Following the completion of the training, all teachers reported incorporating activity breaks into their classrooms, with most reporting doing so very often (n=8) and others somewhat often (n=3). Taken together, these findings indicate that the InPACT teacher training modules were clear, functional, and feasible to implement, with few reported barriers and flexible options for engaging with written materials and supplemental content.

Textbox 1 presents the themes derived from teachers' open-ended survey responses, along with illustrative quotes drawn directly from participants' free-text comments. Five major themes emerged from the thematic analysis of open-ended survey responses. The first theme concerned ease of use and accessibility. Most teachers reported that the modules were easy to navigate and understand, both in terms of the platform interface and the clarity of the content. The flexibility to complete the modules at one's own pace was particularly appreciated, as many teachers noted they were unable to finish a module in a single sitting.

XSL•FO
RenderX

Textbox 1. Themes emerging from thematic analysis of open-ended responses in the structured online survey. Pseudonyms are used for all quotes presented.

Theme 1: ease of use and accessibility

- "I thought they were very easy and useful." [P5, Janice Marquez]
- "Easy access." [P7, Jose Jackson]
- "They were easy to follow and understand." [P11, Lucy Johnson]

Theme 2: content utility

- "I've been using what I learned for years now." [P12, Lauren Smith]
- "It had a lot of useful information." [P3, Isabel Kim]
- "I have been using many activity breaks throughout my day and noticing significant improvement in student's engagement, focus, but mostly happiness." [P12, Lauren Smith]

Theme 3: interactive and applied learning

- "Examples that got us up and moving as we completed the modules." [P4, Farrah Bandow]
- "More interactive activities/links." [P3, Isabel Kim]

Theme 4: engagement and presentation

- "It was not presented in a very engaging way." [P3, Isabel Kim]
- "I would be more concise and shorten them up." [P13, Rhonda Barrales]
- "The video was hard to read." [P2, Emily Canton]

Theme 5: technical flexibility

- "You had to wait before the module let you move forward." [P11, Lucy Johnson]
- "The delay for the videos was sometimes longer than necessary so you had to wait before the module let you move forward." [P10, Sarah Lee]

The second theme focused on the perceived utility of the content. Content utility is defined as how effectively content fulfills a user's specific need, goal, or task, resulting in user satisfaction [41]. Participants described the modules as informative and relevant, highlighting their value as a resource for both learning new strategies and reinforcing prior knowledge. Specific content areas, such as classroom management techniques and safety considerations, were frequently cited as especially useful.

The third theme pertained to the interactive and applied nature of the learning experience. Teachers expressed appreciation for the modules' interactive components and noted the practical application of the material in their classroom settings. They valued being able to convey information to students and implement the strategies directly in their teaching practice.

The fourth theme addressed engagement and presentation. While participants acknowledged the usefulness of the content, some indicated that the modules were overly dense and lacked engaging delivery. Suggestions for improvement included making the modules more concise and enhancing the presentation style to increase user engagement.

Finally, the fifth theme related to technical flexibility. A few participants reported technical issues, such as difficulties with video readability and excessive delays during video progression. To improve the overall UX, teachers recommended shortening the content and providing greater control over video playback, including options to skip or speed up sections as needed.

Heuristic Evaluation

Table 2 presents the results of the heuristic evaluation conducted on the InPACT teacher training modules. A total of 14 usability issues across 9 of the Nielsen 10 heuristics were identified. For the heuristic of help and documentation, the training received a score of 2, indicating minor usability problems. For 6 heuristics—visibility of system status, match between the system and the real world, consistency and standards, recognition rather than recall, aesthetic and minimalist design, and help users recognize, diagnose, and recover from errors—the training received a score of 3, indicating major usability problems. Finally, the heuristics of user control and freedom and error prevention were rated with a severity score of 4, representing usability catastrophes. The heuristic of flexibility and efficiency of use—which emphasizes allowing both novice and expert users to interact effectively with a system, for example, by providing shortcuts or customizable features that make frequent actions faster—was not evaluated, as novice end users were not included in this evaluation.

XSL•FO
RenderX

Table . Heuristic evaluation conducted on Interrupting Prolonged Sitting with Activity (InPACT) modules.

Heuristic	Violation	Recommendation	Severity
1. Visibility of system status	• The survey completion bar does not reflect the user's progress, and there is no indication that progress is automatically saved.	• Add a clear message that progress is autosaved and improve the survey bar accuracy.	3
2. Match between the system and the real world	• The training modules can be seen to be more of a self-evaluation, rather than a training module, as there are "level of agreement" questions premodule and postmodule. • Users would expect assessments, interactive elements, resources, and so on, when completing training modules.	• Introduce interactive case studies that allow users to apply what they learned to a specific scenario. • Add interactive assessments and resources for users to complete and/or learn more about as they complete the modules.	3
3. User control and freedom	• Users cannot go back to previous screens without rewatching the entire video. • No exit button for users to leave the module or return to a main menu (constraint with Qualtrics XM).	• Allow users to navigate back without forcing them to rewatch completed videos. • Add a clear exit or home button for better navigation.	4
4. Consistency and standards	• Inconsistency in page titles (eg, "Do" and "Read" appear on some pages but not others).	• Standardize headings across all modules for clarity.	3
5. Error prevention	• No clear instruction that videos have a time Boolean, leading to confusion. Additionally, even if users fully watch the videos on a different video speed, they cannot advance to the next page.	• Remove the time Boolean. • Incorporate JavaScript to reflect the completion of videos. • Inform users that they cannot go back without having to watch the video again.	4
6. Recognition rather than recall	• Some pages contain large blocks of text without clear sectioning, making content hard to recall.	• Break up text with headings, bullet points, and images.	3
7. Flexibility and efficiency of use	• N/A[a]	• N/A	N/A
8. Aesthetic and minimalist design	• Some pages are very text-heavy, making content overwhelming. • The final page is labeled as a "Summary of Responses" but does not clearly indicate next steps.	• Incorporate images, videos, and better spacing to improve readability. • Reword the title and include a clear call to action on what to do next.	3
9. Help users recognize, diagnose, and recover from errors	• The form for entering an email address does not specify what a "valid email" format is. • Users do not receive feedback if they have not completed a necessary step (eg, watching a video fully).	• Provide clear instructions on acceptable email formats. • Display an error message explaining why the next button is disabled.	3

XSL•FO
RenderX

Heuristic	Violation	Recommendation	Severity
10. Help and documentation	• It is unclear if teachers need the program guidebook to complete the training. • Check-in does not provide contact links for additional help.	• Clearly state whether the book is required or optional. • Include links to resources and a contact person for support.	2

[a]N/A: not applicable.

Competitive Analysis

Table 3 highlights the strengths and limitations across the three evaluated LMS platforms. Moodle demonstrated strong capabilities for course creation and content organization, including support for multimedia integration, multiple courses per module, and clear navigation features, though it lacked consistent progress indicators, time estimates, and a dedicated "next" button. Teachable offered a responsive design across devices, collaborative course-building features, analytics, and interactive activities, but some aspects of content presentation were confusing, and students could bypass quizzes. Thinkific provided intuitive course creation, clear progress tracking, interactive quizzes with feedback, and helpful media labeling, but the course title occupied excessive screen space, limiting navigation efficiency. Collectively, these findings highlight the relative advantages and trade-offs of each platform, offering guidance for selecting or refining LMS solutions to optimize usability, learner engagement, and module scalability for the InPACT teacher training program.

XSL•FO
RenderX

Table . Competitive analysis, where course creation reflects how easily content can be developed and adapted; user experience captures usability, navigation, and accessibility; and interactivity reflects opportunities for engagement, practice, and feedback that support learning and implementation.

LMS[a] name	Course creation	User experience	Interactivity
Moodle	• (+[b]) Allows for the inclusion of slideshows, images, and videos within course content. • (+) Ability to create multiple courses for each training module for less information overload.	• (+) The user can filter and search through courses. • (+) Tracking progress bar shows the user how far into the module they are. • (+) There are organization and categorization of lessons (dropdowns for sublessons). • (+) Indication of completed section (green dot). • (+) There can be multiple pages within main sections to have less information presented all at once. • (+) Clearly marked "back" and "exit" buttons, able to send the user back very easily and without confusion. • (–[c]) Difficult to see progress on course, for some modules, no indications if a section is "completed" or not. • (–) No time estimates. • (–) No "next" button, user must navigate on their own with little to no information. • (–) No progress on how many subsections are in each module, only using a percentage for progress.	• (+) Moodle has learning retention activities, such as quizzes.
Teachable	• (+) Similar process to Qualtrics XM, where the platform allows for blocks to add or remove items. • (+) Ability to see users (students) and other analytics. • (+) Collaborative features; multiple people can work on a Teachable course at once. • (+) Option to bulk upload files (resources), videos, documents, and audio to create a lesson. • (+) AI[d] tools to include a section summary.	• (+) Responsive design, so users can complete the modules on their laptop, tablet, and phones. • (+) Students can see their progress and sections in the course. • (–) Content section is a bit confusing with the circle icon.	• (+) Interactive activities, such as quizzes and clicking on resources. • (–) Students can complete a section without taking the quiz.
Thinkific	• (+) Can easily select categories to create modules. • (+) Ability to change settings on quiz. For example, requiring a passing grade or importing questions. • (+) Easy to navigate course creation.	• (+) Search bar within course allows easy navigation for the user. • (+) Progress bar with numbers with the option to expand on sections. • (+) Informs users about the type of media in each section (eg, video and quiz) and provides an estimated time to complete it. • (+) Provides feedback on quiz when a user answers (correct or incorrect). • (–) The title of the course takes up approximately 1/3 of the navigation section, and there is no way to minimize it.	• (+) Interactive activities, such as quizzes and videos.

XSL•FO
RenderX

[a]LMS: learning management system.

[b]The plus sign (+) denotes observed strengths.

[c]The minus sign (–) denotes observed weaknesses.

[d]AI: artificial intelligence.

Discussion

Principal Results

Given the barriers teachers face in integrating physical activity into classroom routines—particularly related to confidence, stress, and limited time—the purpose of this study was to apply a HCD approach to evaluate and enhance the usability of the InPACT virtual teacher training modules. Using mixed methods—including a structured online survey, heuristic evaluation, and competitive analysis—we assessed teachers' training experiences and identified key usability challenges and opportunities for improvement. Results from the structured survey revealed strong support for the modules' content and structure, including perceived usefulness, flexibility, and applicability. However, teachers also noted that the modules could benefit from improved pacing, greater interactivity, and enhanced video controls. The heuristic evaluation uncovered 14 usability issues across 9 of the Nielsen 10 heuristics, including multiple major usability problems and a severity score of 4 ("usability catastrophe") for user control and freedom. These findings pointed to critical design flaws that may hinder user engagement and reduce the likelihood of training completion. The competitive analysis compared three LMS platforms (Moodle, Teachable, and Thinkific) and highlighted key platform features—such as course creation flexibility, learner feedback mechanisms, and collaborative design tools—that could inform future iterations of the InPACT platform. Taken together, these results suggest that while the InPACT training modules are content-rich and well-aligned with teacher needs, several design and delivery improvements are needed to enhance usability, improve training uptake, and support the sustainable adoption of classroom physical activity practices.

Although this study focused primarily on individual-level factors related to teacher engagement with the training modules, the implementation of classroom-based physical activity interventions is influenced by multiple, interacting levels of context. At the inner setting level, factors such as administrative support, school culture, time constraints, and competing instructional priorities can directly shape teachers' ability to apply training content in daily practice [23,42]. At the outer setting level, district policies, professional development requirements, and broader accountability pressures may further enable or constrain implementation [43,44]. Prior work in low-resource school settings has shown that even well-designed, evidence-based interventions may experience reduced fidelity or sustainment when organizational and system-level conditions are not aligned with implementation demands [44-46]. As demonstrated by Hasson et al [47], addressing these multilevel influences through leadership engagement, district partnerships, and adaptive implementation strategies is critical for translating individual-level training gains into sustained practice. Future efforts to scale teacher training modules should therefore incorporate complementary strategies that target school- and district-level barriers alongside individual professional development.

Comparison With Prior Work

Prior research underscores the critical role of professional development in the successful implementation of school-based health programs [48]. Teachers are more likely to adopt and sustain new interventions when they receive training that is accessible, practical, and tailored to their classroom needs [49]. In the context of physical activity programming, effective professional development has been linked to improvements in teacher confidence, fidelity of program delivery, and student engagement [42,50,51]. Accordingly, the InPACT virtual teacher training modules were designed to serve as a central implementation strategy [52].

However, usability challenges identified in this evaluation suggest that the training modules fell short of best practices for online professional development. Using the Nielsen usability heuristic framework, the evaluation revealed 3 primary issues: lack of user control and freedom, inconsistency and lack of standards, and insufficient recognition rather than recall. Teachers described frustration with not being able to pause or skip segments of the training videos, which reflects poor user control and freedom. This limitation is especially problematic for time-constrained educators [47], who may prefer to pace their training across multiple sittings. Inconsistencies in video labeling and navigation—such as unclear module titles and nonuniform layouts—violated established design standards, making it harder for users to locate content, which can contribute to cognitive overload [53]. Finally, the training platform required users to remember previously viewed content with limited visual cues to support recall. This reflects insufficient recognition rather than recall, a heuristic that emphasizes reducing memory load by making options, actions, and information visible so that users can recognize what to do instead of relying on memory. Without visual scaffolding, such as clear navigation menus or progress indicators, the cognitive burden increases, and learnability suffers [54].

These usability shortcomings contrast with best practices from prior studies, which demonstrate that professional development resources should be easy to navigate, chunked into digestible segments, and provide clear cues to help users orient themselves within the training [40]. Addressing these issues is critical to improving the training experience for teachers and, in turn, increasing the implementation fidelity of classroom physical activity programming in schools.

Recommendations for Enhancement of the InPACT Virtual Teacher Training Modules

A major usability barrier identified during the evaluation was teacher frustration with video replay requirements, particularly when navigating backward in a module. To address this issue,

XSL•FO
RenderX

the evaluation team recommended removing the time-based Boolean constraint on embedded videos. Specifically, they proposed using JavaScript event listeners integrated with the YouTube application programming interface to track video play or pause actions and determine when a sufficient portion of the video has been viewed. This would allow users to resume or advance through content without having to rewatch full segments. In cases where this solution was not feasible within the current Qualtrics platform, the team recommended displaying clear notifications indicating that videos must be viewed in full before proceeding.

Additional enhancements were proposed to improve time management and overall UX. Teachers noted that the existing page-based progress bar was misleading, making it difficult to anticipate the time commitment. To improve clarity, the evaluation team proposed a content-based progress bar with labeled topic nodes, or alternatively, the inclusion of manual page indicators and estimated completion times.

To promote engagement and retention, the team suggested reducing dense text and incorporating visual aids and interactive components such as collapsible content sections, quizzes with feedback, real-world case studies, and downloadable action plans. These enhancements were aimed at strengthening active learning and implementation fidelity. Navigation improvements were also prioritized. Teachers expressed difficulty tracking progress across modules, so a standardized end-of-module message with clear next steps (eg, “To continue to the next module, click here”) was recommended. Inconsistencies in how modules referenced the companion program guide led to confusion, prompting a recommendation to standardize these references and clearly indicate when guide components are optional.

To implement these enhancements and support scalability, we partnered with Michigan Virtual, a statewide leader in accessible professional learning [54]. Michigan Virtual committed to redesigning and developing the revised modules using their Brightspace LMS, ensuring compliance with Web Content Accessibility Guidelines and Americans with Disabilities Act standards. Their instructional designers identified necessary content updates, helped revise training materials, and provided technical support. Michigan Virtual also offered to host the updated modules on its Professional Learning Portal, manage State Continuing Education Clock Hours processing, and share educator completion data with the InPACT team. This partnership is expected to ensure a high-quality, accessible, and scalable professional development experience for educators statewide.

Preliminary data from the revised InPACT teacher training modules hosted on Michigan Virtual indicate strong early uptake, engagement, and performance during the first eight months following release. As of April 7, 2026, 2061 educators have enrolled in the training with 1138 participants earning certificates of completion, reflecting substantially greater reach and completion volume than the prior Qualtrics-hosted training (35 teachers over a 3-year period). Knowledge acquisition was assessed using standardized postmodule quizzes embedded within each training module. Across these module quizzes, participants demonstrated high mastery of content, with an average score of 90.6% (SD 2.7%), and required relatively few attempts to achieve passing scores (mean 1.5, SD 0.2 attempts per module). Collectively, these preliminary findings suggest that the revised training—coupled with a more accessible delivery platform—has markedly improved scalability, engagement, and learning outcomes.

Partnerships played a central role in guiding the enhancements to the InPACT teacher training modules. Engaging a diverse set of collaborators—including classroom teachers, instructional designers, accessibility specialists, and statewide professional learning organizations—was essential for identifying usability challenges, prioritizing improvements, and ensuring that module revisions were both practical and scalable. These partnerships enabled the co-development of solutions that enhanced navigation, interactivity, and accessibility, while ensuring alignment with accessibility and disability standards. For teams with limited budgets, leveraging existing relationships, focusing on high-impact usability improvements, and incorporating low-cost accessibility strategies—such as consistent page structure, clear labeling, and visual scaffolding—can help maximize impact. Additionally, integrating clear data visualization and feedback mechanisms, alongside ongoing communication and iterative feedback with end users, is recommended to promote engagement, learning retention, and sustainable implementation of virtual professional development programs.

Strengths and Limitations

This study has several notable strengths. First, we used a mixed methods usability evaluation that incorporated both end-user feedback and expert recommendations to guide improvements to the InPACT teacher training modules. Second, the evaluation was conducted by a multidisciplinary team with expertise in instructional design, educational technology, and human-computer interaction. Finally, our collaborative and HCD approach fostered partnerships between teachers, evaluators, and implementation support organizations, enabling the co-development of a more accessible, engaging, and scalable professional learning experience for educators across Michigan.

Limitations of this study also warrant consideration. First, teacher feedback was based on self-report and retrospective recall, which may be subject to recall bias or social desirability bias and may not fully capture real-time usability challenges encountered during module completion. Although self-report data are commonly used in early-stage usability and implementation research to efficiently gather end-user perspectives, these methods may overlook nuanced interaction patterns, moment-to-moment frustration points, or navigation errors that occur during actual use [55]. Future usability evaluations that integrate observational and system-level data with teacher self-reports will be critical for identifying real-time barriers and optimizing training engagement and completion.

Second, the participant sample consisted primarily of teachers who were already implementing the InPACT program. While this was appropriate for the study’s aim of evaluating the usability of existing training modules, these participants may have had greater familiarity with the program content and

XSL•FO
RenderX

underlying concepts than educators new to InPACT. Future evaluations should include educators unfamiliar with InPACT to examine initial usability, onboarding experiences, and barriers faced by first-time users.

Third, this study did not assess whether the usability improvements applied in the revised modules translated into changes in teacher self-efficacy, adoption of classroom activities, or fidelity of implementation. Future research is needed to determine whether usability-informed redesigns of the InPACT teacher training modules lead to measurable improvements in teacher self-efficacy, training completion, adoption, and implementation fidelity.

Conclusions

Most teachers in the United States do not implement classroom physical activity breaks with the frequency or fidelity recommended to improve student health and academic outcomes [14,15]. Virtual training modules offer a scalable approach to build teacher capacity, but their effectiveness depends not only on high-quality content but also on user-centered design and usability. In this study, a combination of heuristic evaluation, competitive analysis, and end-user surveys was used to identify and address barriers to teacher engagement with the InPACT virtual teacher training modules. Key recommendations focused on improving navigation, video accessibility, and visual appeal to reduce user frustration and enhance the learning experience. Applying an HCD approach enabled collaboration between evaluators, developers, and end users to optimize virtual training for broader adoption. Future work should examine whether these HCD-informed revisions lead to measurable improvements in teacher self-efficacy, training completion, classroom physical activity program adoption, implementation fidelity, and student outcomes.

Acknowledgments

The authors would like to thank Rongyu Shen, Sanvika Inturi, Shannon Lee, and the University of Michigan School of Information client engagement team for their helpful assistance throughout the research process. The authors are also grateful to the study participants for their involvement. Finally, the authors appreciate the Michigan Virtual team for making subsequent improvements. ChatGPT (GPT-3.5/4.0) was used to edit the grammar and sentence structure of the manuscript. The results are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

Funding

The Michigan Health Endowment Fund provided funding for this study (grant G-2404-154026).

Data Availability

Data are available upon request by sending an email to the corresponding author.

Authors' Contributions

REH wrote the manuscript and oversaw data collection for the project. AS and AL developed the Qualtrics virtual training modules and implemented updates. TT and LM conducted the teacher interviews, heuristic evaluation, and competitive analysis. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Structured online survey instrument.
[DOCX File, 20 KB - humanfactors_v13i1e83427_app1.docx]

References

1. Piercy KL, Troiano RP, Ballard RM, et al. The physical activity guidelines for Americans. JAMA 2018 Nov 20;320(19):2020-2028. [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
2. Carlson JA, Moon M, Steel C, et al. Results from the 2024 United States report card on physical activity for children and youth. J Phys Act Health 2025;23(3):293-300. [doi: 10.1123/jpah.2024-0872] [Medline: 41352335]
3. Nader PR, Bradley RH, Houts RM, McRitchie SL, O'Brien M. Moderate-to-vigorous physical activity from ages 9 to 15 years. JAMA 2008 Jul 16;300(3):295-305. [doi: 10.1001/jama.300.3.295] [Medline: 18632544]
4. Richmond TK, Hayward RA, Gahagan S, Field AE, Heisler M. Can school income and racial/ethnic composition explain the racial/ethnic disparity in adolescent physical activity participation? Pediatrics 2006 Jun;117(6):2158-2166. [doi: 10.1542/peds.2005-1920] [Medline: 16740860]
5. Lieb DC, Snow RE, DeBoer MD. Socioeconomic factors in the development of childhood obesity and diabetes. Clin Sports Med 2009 Jul;28(3):349-378. [doi: 10.1016/j.csm.2009.02.004] [Medline: 19505621]

XSL•FO
RenderX

6. Wyszyńska J, Ring-Dimitriou S, Thivel D, et al. Physical activity in the prevention of childhood obesity: the position of the European childhood obesity group and the European academy of pediatrics. Front Pediatr 2020;8:535705. [doi: 10.3389/fped.2020.535705] [Medline: 33224905]
7. Committee on Physical Activity and Physical Education in the School Environment, Food and Nutrition Board, Institute of Medicine. In: Kohl HW IIICook HD, editors. Educating the Student Body: Taking Physical Activity and Physical Education to School: National Academies Press; 2013. [doi: 10.17226/18314] [Medline: 24851299]
8. Pulling Kuhn A, Stoepker P, Dauenhauer B, Carson RL. A systematic review of multi-component Comprehensive School Physical Activity Program (CSPAP) interventions. Am J Health Promot 2021 Nov;35(8):1129-1149. [doi: 10.1177/08901171211013281] [Medline: 33955278]
9. Bassett DR, Fitzhugh EC, Heath GW, et al. Estimated energy expenditures for school-based policies and active living. Am J Prev Med 2013 Feb;44(2):108-113. [doi: 10.1016/j.amepre.2012.10.017] [Medline: 23332325]
10. Masini A, Marini S, Gori D, Leoni E, Rochira A, Dallolio L. Evaluation of school-based interventions of active breaks in primary schools: a systematic review and meta-analysis. J Sci Med Sport 2020 Apr;23(4):377-384. [doi: 10.1016/j.jsams.2019.10.008] [Medline: 31722840]
11. Hoff DL, Mitchell SN. Should our students pay to play extracurricular activities? Educ Digest 2007;72(6):27-34 [FREE Full text]
12. Zdroik J, Veliz P. The influence of pay-to-play fees on participation in interscholastic sports: a school-level analysis of Michigan's public schools. J Phys Act Health 2016 Dec;13(12):1317-1324. [doi: 10.1123/jpah.2016-0099] [Medline: 27617620]
13. Physical activity: classroom-based physically active lesson interventions. : Community Preventive Services Task Force (CPSTF) & Centers for Disease Control and Prevention (U.S); 2021 URL: https://stacks.cdc.gov/view/cdc/168558 [accessed 2026-03-27]
14. Densley B, Calvert HG, Boedeker P, Turner L. Implementation of physical activity in US elementary schools: the role of administrative support, financial resources, and champions. Int J Environ Res Public Health 2021 Apr 23;18(9):4476. [doi: 10.3390/ijerph18094476] [Medline: 33922470]
15. Turner L, Chaloupka FJ. Reach and implementation of physical activity breaks and active lessons in elementary school classrooms. Health Educ Behav 2017 Jun;44(3):370-375. [doi: 10.1177/10901981166677714] [Medline: 27608645]
16. Dinkel D, Schaffer C, Snyder K, Lee JM. They just need to move: teachers' perception of classroom physical activity breaks. Teach Teach Educ. 2017 Apr;63:186-195. [doi: 10.1016/j.tate.2016.12.020]
17. Huberty J, Dinkel D, Coleman J, Beighle A, Apenteng B. The role of schools in children's physical activity participation: staff perceptions. Health Educ Res 2012 Dec;27(6):986-995. [doi: 10.1093/her/cys071] [Medline: 22717940]
18. McLoughlin GM, Calvert HG, Turner L. Individual and contextual factors associated with classroom teachers' intentions to implement classroom physical activity. Int J Environ Res Public Health 2023 Feb 18;20(4):3646. [doi: 10.3390/ijerph20043646] [Medline: 36834340]
19. McMullen J, Kulinna P, Cothran D. Chapter 5 physical activity opportunities during the school day: classroom teachers' perceptions of using activity breaks in the classroom. J Teach Phys Educ 2014;33(4):511-527. [doi: 10.1123/jtpe.2014-0062]
20. Michael RD, Webster CA, Egan CA, et al. Facilitators and barriers to movement integration in elementary classrooms: a systematic review. Res Q Exerc Sport 2019 Jun;90(2):151-162. [doi: 10.1080/02701367.2019.1571675] [Medline: 30794089]
21. Webster CA, Zarrett N, Cook BS, Egan C, Nesbitt D, Weaver RG. Movement integration in elementary classrooms: teacher perceptions and implications for program planning. Eval Program Plann 2017 Apr;61:134-143. [doi: 10.1016/j.evalprogplan.2016.12.011] [Medline: 28068554]
22. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012 Apr 24;7:37. [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
23. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 7;4:50. [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
24. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42. [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
25. Bartholomew JB, Jowers EM. Physically active academic lessons in elementary children. Prev Med 2011 Jun;52 Suppl 1(Suppl):S51-S54. [doi: 10.1016/j.ypmed.2011.01.017] [Medline: 21281672]
26. Carlson JA, Engelberg JK, Cain KL, et al. Contextual factors related to implementation of classroom physical activity breaks. Transl Behav Med 2017 Sep;7(3):581-592. [doi: 10.1007/s13142-017-0509-x] [Medline: 28589532]
27. Walker TJ, Craig DW, Robertson MC, Szeszulski J, Fernandez ME. The relation between individual-level factors and the implementation of classroom-based physical activity approaches among elementary school teachers. Transl Behav Med 2021 Apr 7;11(3):745-753. [doi: 10.1093/tbm/ibaa133] [Medline: 33598700]
28. Guskey TR. Professional development and teacher change. Teach Teach (Theory Pract) 2002 Aug;8(3):381-391. [doi: 10.1080/135406002100000512]
29. Fullan M, Miles M. Getting reform right: what works and what doesn't. Phi Delta Kappan 1992;73(10):745-752 [FREE Full text]

XSL•FO
RenderX

30. Beemer LR, Ajibewa TA, O'Sullivan MP, et al. Feasibility of the InPACT intervention to enhance movement and learning in the classroom. Transl J Am Coll Sports Med 2018;3(18):136-151. [doi: 10.1249/TJX.0000000000000069]
31. Darling-Hammond L, Hyler ME, Gardner M. Effective teacher professional development. : Learning Policy Institute (LPI); 2017 URL: https://files.eric.ed.gov/fulltext/ED606743.pdf [accessed 2026-03-27]
32. Knowles MS, Holton EFI, Swanson RA, Robinson PA. The Adult Learner: The Definitive Classic in Adult Education and Human Resource Development, 9th edition: Routledge; 2020.
33. Garett R, Chiu J, Zhang L, Young SD. A literature review: website design and user engagement. Online J Commun Media Technol 2016 Jul;6(3):1-14. [Medline: 27499833]
34. Nielsen J. Usability Engineering: Academic Press; 1993.
35. The impact of website design on customer acquisition and retention. AIContentfy. 2023. URL: https://aicontentfy.com/en/blog/impact-of-website-design-on-customer-acquisition-and-retention [accessed 2024-06-28]
36. Norman DA, Draaper SW. User Centered System Design New Perspectives on Human-Computer Interaction: Lawrence Erlbaum Associates, Inc; 1986. [doi: 10.1201/9780367807320]
37. Kehoe B. User experience (UX) design glossary: improving your customer experience with a good UX. Qualtrics. 2024. URL: https://www.qualtrics.com/articles/customer-experience/ux-design-glossary/ [accessed 2026-04-07]
38. Nielsen J. Heuristic evaluation. In: Nielsen J, Mack RL, editors. Usability Inspection Methods: Joh Wiley & Sons; 1994:25-62.
39. Nielsen J, Molich R. Heuristic evaluation of user interfaces. Presented at: CHI '90: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; Apr 1-5, 1990; Seattle, Washington, United States p. 249-256. [doi: 10.1145/97243.97281]
40. Nielsen J. 10 usability heuristics for user interface design. Nielsen Norman Group. 1994. URL: https://www.nngroup.com/articles/ten-usability-heuristics/ [accessed 2025-08-05]
41. Garg V. Content utility in content strategy and design. Medium. 2022. URL: https://medium.com/@vingar/content-utility-in-content-strategy-and-design-428e8fbf044d [accessed 2026-04-07]
42. Turner L, Calvert HG, Carlson JA. Supporting teachers' implementation of classroom-based physical activity. Transl J Am Coll Sports Med 2019 Sep 1;4(17):165-172. [doi: 10.1249/TJX.0000000000000098]
43. Friday PJ, Beemer LR, Martindale D, et al. A novel policy alignment and enhancement process to improve sustainment of school-based physical activity programming. Int J Environ Res Public Health 2023 Jan 18;20(3):1791. [doi: 10.3390/ijerph20031791] [Medline: 36767157]
44. Hasson RE, Eisman AB, Wassmann A, et al. Aligning organizational priorities and system policies to support implementation scale-up of a tailored classroom-based physical activity intervention in low-resource schools. J Sch Health 2023 Jun;93(6):464-474. [doi: 10.1111/josh.13321] [Medline: 36918350]
45. Eisman AB, Martin J, Hasson RE, Kilbourne AM. Resource realities: exploring the foundations of successful implementation in school-based drug prevention. J Sch Health 2024 Dec;94(12):1185-1195. [doi: 10.1111/josh.13539] [Medline: 39721940]
46. McLeod BD, Sutherland KS, Conroy MA, et al. Study protocol: multi-level determinants of implementation and sustainment in the education sector. J Emot Behav Disord 2023 Mar;31(1):27-40. [doi: 10.1177/10634266221090160] [Medline: 36874907]
47. Hasson RE, Beemer LR, Eisman AB, Friday P. Closing the gap between classroom-based physical activity intervention adoption and fidelity in low-resource schools. Kinesiol Rev 2023;12(1):36-46. [doi: 10.1123/kr.2022-0041]
48. Otten C, Nash R, Patterson K. Professional development in health education for primary school teachers: a systematised review of the literature. Prof Dev Educ 2024;50(5):809-831. [doi: 10.1080/19415257.2022.2038233]
49. Manal Q, Cobo C, Wilichowski T, Patil A. Effective teacher professional development using technology: technology-based strategies from across the globe to enhance teaching practices—a guidance note. : World Bank; 2022 URL: https://documents1.worldbank.org/curated/en/099835106172233833/pdf/P1742520d4de390610b784047359e0cefc7.pdf [accessed 2026-03-27]
50. Calvert HG, Lane HG, Bejarano CM, et al. An evaluation of the coverage of theoretically based implementation factors in disseminated classroom physical activity programs. Transl Behav Med 2020 Oct 8;10(4):959-969. [doi: 10.1093/tbm/iby134] [Medline: 30590851]
51. Stewart G, Webster CA, Weaver RG, et al. Evaluation of a classroom movement integration training delivered in a low socioeconomic school district. Eval Program Plann 2019 Apr;73:187-194. [doi: 10.1016/j.evalprogplan.2018.12.010] [Medline: 30682534]
52. Hasson RE, Beemer LR, Ajibewa TA, Eisman AB. Adapting the InPACT intervention to enhance implementation fidelity and flexibility. Prev Sci 2021 Apr;22(3):324-333. [doi: 10.1007/s11121-020-01199-z] [Medline: 33453045]
53. Skulmowski A, Xu KM. Understanding cognitive load in digital and online learning: a new perspective on extraneous cognitive load. Educ Psychol Rev 2022 Mar;34(1):171-196. [doi: 10.1007/s10648-021-09624-7]
54. Cuccolo K, DeBruler K. A look back at 3 years of Michigan virtual research. Michigan Virtual. 2024. URL: https://michiganvirtual.org/research/publications/a-look-back-at-3-years-of-michigan-virtual-research/ [accessed 2025-08-05]
55. Sauro J. Can users self-report usability problems? MeasuringU. 2010. URL: https://measuringu.com/self-reporting [accessed 2026-04-07]

XSL•FO
RenderX

Abbreviations

HCD: human-centered design
InPACT: Interrupting Prolonged Sitting with Activity
LMS: learning management system
UX: user experience

Edited by A Bamgboje-Ayodele; submitted 03.Sep.2025; peer-reviewed by C Provvidenza, H Li; revised version received 10.Feb.2026; accepted 10.Feb.2026; published 21.Apr.2026.

Please cite as:
Hasson RE, Tam T, Myers L, Schwartz A, Li A
Improving Usability of the Interrupting Prolonged Sitting With Activity Virtual Teacher Training Modules: Case Study
JMIR Hum Factors 2026;13:e83427
URL: https://humanfactors.jmir.org/2026/1/e83427
doi:10.2196/83427

© Rebecca E Hasson, Tiffany Tam, Lucy Myers, Anna Schwartz, Alissa Li. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 21.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Personalizing Mobile Apps for Health Behavioral Change According to Personality: Cross-Sectional Validation of a Preference Matrix

Laetitia Gosetto[1,2], PhD; Gilles Falquet[1], PhD; Frederic Ehrler[3], PhD

[1]Research Institute for Statistics and Information Science, Geneva School of Economics and Management, University of Geneva, 40 Boulevard du Pont-d'Arve, Geneva, Switzerland

[2]Division of Medical Information Sciences, University Hospital of Geneva, Geneva, Switzerland

[3]Information Systems Directorate, University Hospital of Geneva, Geneva, Switzerland

Corresponding Author:
Laetitia Gosetto, PhD
Research Institute for Statistics and Information Science, Geneva School of Economics and Management, University of Geneva, 40 Boulevard du Pont-d'Arve, Geneva, Switzerland

Abstract

Background: Mobile health (mHealth) apps are increasingly used to support healthy lifestyle behaviors through features such as health tracking and personalized reminders. Personalized messaging, tailored to users' profiles, has been shown to improve engagement and retention in health-related contexts. Prior research has linked personality traits, based on the Big Five model, to preferences for specific app mechanisms, leading to the development of a preference matrix for personalizing mHealth apps. This matrix comprises 15 mechanisms derived from behavior change techniques and gamification elements, intended to guide developers in optimizing engagement according to user profiles.

Objective: This study aimed to validate this preference matrix by examining whether the associations between mechanisms and Big Five personality traits reported in the literature align with user preferences observed in an experimental setting.

Methods: A cross-sectional study was conducted using an online survey that collected demographic data, mHealth app usage, and personality traits. Participants were presented with mockups illustrating 15 mechanisms and were asked to select their preferred options. Logistic regression and ordinal logistic regression analyses were performed to examine associations between personality traits, mechanism selection, and motivation scores. All analyses were adjusted using the Bonferroni correction to account for multiple comparisons.

Results: A total of 214 participants completed the survey (mean age 29.42, SD 10.41 y; n=118, 55.1% women; n=89, 41.6% men; n=5, 2% identifying as other; and n=2, 1% nonrespondents). Higher conscientiousness significantly increased the likelihood of selecting the collection mechanism (eg, collecting badges or points; odds ratio [OR] 1.87, 95% CI 1.27-2.75). For competition (eg, competing with other users), conscientiousness (OR 3.22, 95% CI 1.73-6.00) and agreeableness (OR 1.93, 95% CI 1.08-3.45) were significant predictors. Preferences for rewards (eg, virtual incentives such as points or virtual currency) were associated with conscientiousness (OR 2.36, 95% CI 1.53-3.63) and neuroticism (OR 1.97, 95% CI 1.36-2.86). Additionally, 4 mechanisms—self-monitoring, progression, challenge, and quest—were selected by more than half of the participants, independent of personality traits.

Conclusions: The findings partially validate the proposed preference matrix. Conscientiousness consistently emerged as a key predictor of preference across multiple mechanisms, highlighting its central role in engagement with gamified mHealth features. While some mechanisms appear to have universal appeal, others show personality-specific preferences, underscoring the value of combining baseline mechanisms with targeted personalization strategies in mHealth app design.

International Registered Report Identifier (IRRID): RR2-10.2196/38603

(JMIR Hum Factors 2026;13:e78939) doi:10.2196/78939

KEYWORDS

mobile health; mHealth; behavior change technique; personalization; personality; Big Five

XSL•FO
RenderX

Introduction

Background

Many studies have demonstrated the role of healthy lifestyle behaviors in increasing life expectancy. Specifically, the adoption of a healthy diet, maintenance of a healthy weight, cessation of smoking, consumption of alcohol in moderation, and regular exercise are associated with significant reductions in mortality. The researchers discovered that the adoption of each healthy lifestyle behavior contributes to an improved quality of life and enhanced longevity [1].

The number of health apps designed to assist individuals in adopting healthier behaviors has continued to grow steadily, with new apps reaching the market every year. In 2018, there were over 35,000 health apps available for download [2]. The advent of smartphone apps has created new opportunities for engaging in health-promoting behaviors. These apps offer immediate access to health-related information, medication reminders, and tools for monitoring progress, which can collectively support healthier lifestyle adoption [3].

Several scales have been developed to assess the quality of mobile health (mHealth) apps, such as the Mobile App Rating Scale [4] and the App Behavior Change Scale [5]. A key feature shared across these frameworks is the inclusion of personalization as a critical quality dimension. Personalization is widely acknowledged as essential in the design of behavioral interventions. For example, 1 study has demonstrated that messages tailored to the user's specific characteristics are more likely to be read, recalled, and retained in memory, and are perceived as more personally relevant than untailored messages. Tailored messages also attract greater attention, stimulate more discussion, and are shared more frequently compared to nontailored messages [6].

The Preference Matrix

Personalization of mobile apps can be done according to the user's profile. In previous studies [7], we identified preference relations between personality traits and specific mechanisms (components that can be integrated into an mHealth app) based on a scoping review of the literature. These findings led to the development of a preference matrix indicating the types of mechanisms preferred according to a user profile. The preference matrix can be used as a framework for the personalization of mobile apps with the aim of encouraging behavioral change. When designing a mobile app intended to support health-related behavior change, designers can draw on this preference matrix to identify the mechanisms most likely to be effective for their target users. For instance, if an individual is classified as extraverted according to the Big Five personality traits, mechanisms that facilitate social comparison or collaboration between users may be particularly relevant [8].

The preference matrix includes 15 mechanisms associated with the Big Five personality traits, which are presented in detail in the table found in Multimedia Appendix 1. For personality profiling, we used one of the most widely applied classification systems, namely the Big Five personality model [8-16].

Personality Measure: The Big Five

One of the most widely used instruments for the assessment of personality is the Big Five model. The Big Five model describes an individual's personality based on 5 distinct traits, as summarized in Table 1.

This model has the advantage of being widely adopted in academic research. Between 1990 and 1994, the number of publications on this topic was approximately 400, while between 2005 and 2009, this figure rose to approximately 1600 [17].

The other advantage of this model is its cross-cultural universality. The Big Five model has been shown to be universal across human populations [18]. Indeed, the authors translated their personality questionnaire, the Revised NEO Personality Inventory (NEO-PI-R), into 6 languages belonging to the same language family as English (German, Portuguese, Hebrew, Chinese, Korean, and Japanese), and the Big Five factors consistently emerged across these versions.

In a subsequent study, researchers examined whether the Big Five personality dimensions, as measured by the NEO-PI-R, were universal across 50 cultures spanning American, European, Arabic, Asian, and African regions [19]. The NEO-PI-R was translated into the local language of each participating country. The participants were primarily students from the country in which the test was administered. Factor analysis demonstrated that the American structure of the NEO-PI-R self-report questionnaire (Form S) could be replicated in all cultures, with 94.4% of factors replicated. Moreover, the general factor structure remained recognizable across all cultural contexts included in the study.

Table . Definition of the Big Five factors of personality.

Traits	The 5 traits represent the tendency to... [20]
Neuroticism	Worried, nervous, emotional, anxious, maladjusted, and hypochondriac
Extraversion	Sociable, active, talkative, open to others, optimistic, fun-loving, and affectionate
Openness	Curious, eclectic, creative, original, imaginative, and nonconformist
Agreeableness	Compassionate, easy-going, trusting, helpful, indulgent, credulous, and honest
Conscientiousness	Organized, reliable, hard-working, disciplined, punctual, meticulous, careful, ambitious, and persevering

XSL•FO
RenderX

Selection of the 15 Mechanisms

Based on the hypothesis that individuals with different profiles exhibit distinct preferences regarding components of mHealth apps designed to facilitate behavior change, a review of the literature identified 15 mechanisms associated with user profiles [21]. In our context, the term "mechanism" is used to denote the entirety of components that can be incorporated within an mHealth app with the objective of promoting behavioral change. We classified the mechanisms into 2 categories: the mechanisms linked to behavior change techniques and the mechanisms linked to game elements. For further details and definition of mechanisms, see Table 2.

Table . List of mechanisms with their definitions.

Mechanism	Definition
Behavior changes techniques	
Prompt and cues	Usually, a message delivered to the user to prompt or recall a behavior at a specified time, with the app or user defining when the message should be sent [22]
Demonstration of the behavior	Enables users to observe the cause-and-effect linkage of their behavior, such as seeing a simulation of their bodies after a diet [23]
Self-monitoring	Users can track their behaviors, providing information on both past and current activities [23]
Punishment	Virtually penalizes the user for not performing the desired behavior or reaching their goal [10]
Social comparison	An individual's perceptions of the prevailing beliefs and behaviors within a social group
Social support	Enables communication between users, such as through chat or sharing activities with other users [24]
Game elements	
Progression	Users can track their progression with steps through the system's purpose over time, visualized with mechanisms like stars or flags along a path [23]
Competition	Users can compete to accomplish the desired behavior [23]
Cooperation	Users collaborate to achieve a shared objective [23]
Collection	Allows users to gather virtual objects
Rewards	Virtual rewards offered to users for engaging in the target behavior [23]
Quest	Users can enter or define the objectives targeted for the activity they will perform [22]
Challenge	Presents various situations that require effort from the user to be completed [24] (eg, accomplishing 3 hours of physical activity per week)
Avatar	Allows users to share their data in the system without revealing their name [24]
App mechanism	
Customization	In contrast to personalization—which involves adjusting automatically the system to the user—customization refers to the user's ability to modify the content or functionalities of the mobile app according to their own preferences [23]. This approach enables users to actively tailor the system based on users' choices.

Preferences Relations Between Big Five Traits and Mechanisms

The preference relations between the mechanisms and the Big Five traits identified in the literature enable us to construct a preference matrix summarized in Multimedia Appendix 1. Each cell of the matrix indicates the number of studies in the review reporting a preference relation between a given user profile and a specific mechanism [21]. Most of the preference relations were positive (53/75, 71%), indicating that a particular personality profile tends to prefer a given mechanism (represented by a plus sign in Multimedia Appendix 1). Negative relations, representing an aversion toward a mechanism (represented by a minus sign in Multimedia Appendix 1), accounted for 12% (9/75) of all potential relations. It is important to note that the relation matrix remains incomplete, as only 73% (55/75) of all potential relations were identified in the literature review.

This study aims to corroborate or enrich this preference matrix by experimentally identifying new preference or nonpreference relations between mechanisms and personality traits.

XSL•FO
RenderX

Methods

Ethical Considerations

This study was approved by the University of Geneva Research Ethics Committee (CUREG_2021-04-38). This section is based on our previously published protocol study [25]. Participants were required to complete a consent form that detailed the purpose of the study, the procedure to be followed, and their right to withdraw from the study at any time. Confirmation that they had read, understood, and agreed to the terms of the consent form was mandatory before accessing the questionnaire and was required for the investigators to use the submitted data for research purposes. Participation was voluntary, responses were collected anonymously, and only deidentified data were used for research purposes in accordance with applicable data protection regulations. Participants did not receive any financial compensation for their participation.

Study Design

We conducted a cross-sectional study to address our research objectives. Participants completed an online questionnaire following the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) guidelines (Checklist 1) [26].

Outcomes

The primary outcome was the identification of preferred mechanisms according to the user profile.

Study Population

The study targeted French-speaking adults aged 18 years and older. Recruitment was conducted both within the University of Geneva community and through social media platforms (eg, Facebook and Twitter, subsequently rebranded X), enabling participation from a broader population beyond university affiliates. The survey was open to anyone meeting the language and age criteria.

Sample Size Calculation

We calculated the required sample size using a multiple regression power analysis in R (R Foundation for Statistical Computing). The parameters used were the number of predictors (u=3), effect size (f^2=0.07), significance level (α=.05), power of 0.9, and estimated variance of 202.403. These estimates were based on the hypothesis that individuals with higher levels of agreeableness, as measured by the Big Five personality traits, show a preference for social networks [11,24].

To approximate the variance, we used data from a previous study [27] that assessed Big Five personality traits in relation to user preferences for social network posters. Specifically, we relied on the variance in Big Five trait ratings for altruistic participants (n=46) based on their average scores for a blood donation poster, which was evaluated on a 0-to-100 scale. Based on these parameters, we determined that a minimum sample size of 206 participants was required.

Procedure

Participants completed an online questionnaire developed by the investigators using Qualtrics software (Qualtrics; Multimedia Appendix 2). First, participants were required to complete a consent form that detailed the purpose of the study, the procedure to be followed, and their right to withdraw from the study at any time. Participants were then asked to provide demographic information. To proceed to the remainder of the survey, they were required to confirm that they were aged 18 years or older. Eligible participants subsequently completed two main components of the questionnaire: (1) a standardized personality assessment measuring their Big Five profile, and (2) a task in which they reviewed 15 mechanisms, selected their 5 preferred ones, and rated the extent to which each selected mechanism would motivate them to engage in physical activity on a scale from 0 to 100.

Measures and Measurement

Demographic Questions

The participants were asked to provide information regarding their gender, age, occupation, and level of education.

Profile Assessment

To assess the personality of the participants, we used the French version of the Big Five Inventory-10 (BFI-10-Fr) scale, translated and validated by Courtois [26]. The internal reliability of the Big Five Inventory-10 (BFI-10) is reported to be relatively low, with Cronbach α coefficients ranging from 0.37 to 0.83. However, this limitation is expected, as Cronbach α is not well-suited for very short scales, particularly those with only 2 items per dimension [26]. The instrument consists of 10 items, with 2 items pertaining to each of the 5 Big Five dimensions. Participants were requested to indicate on a 5-point Likert scale whether they approve or disapprove of the statements pertaining to themselves. For instance, respondents indicated whether they perceived themselves as reserved or as someone who is easily anxious. Scores for each personality dimension were computed by summing the two relevant items, with reverse scoring applied where appropriate.

This scale was selected because its factorial structure mirrors that of the full BFI-10-Fr, providing a valid measure of personality while minimizing participant burden [26]. Given the number of instruments included in the overall protocol, using a short, validated scale was deemed necessary to reduce fatigue and prevent attrition associated with excessively long questionnaires.

Choice of Mechanisms

Each of the 15 gamification mechanisms was represented by a mockup accompanied by a short textual description (Multimedia Appendix 2). To minimize potential bias related to visual design preferences, all mockups were created using a deliberately simple and neutral layout. The interface used only black-and-white tones and standardized icons to ensure clarity and comparability across mechanisms (Multimedia Appendix 2). The 15 mockups were displayed in a randomized order to control for possible primacy or recency effects. Participants were instructed to carefully review each screen and its corresponding description before making their choices. After viewing all 15 mechanisms, they were asked to select the five they personally found most motivating.

XSL•FO
RenderX

Explanation of Choice

For each of the 5 mechanisms selected in the previous step, participants were asked to indicate the degree to which that specific mechanism would motivate them to adopt healthier behaviors, using a continuous scale ranging from 0 (not at all motivating) to 100 (extremely motivating). After rating each of their 5 chosen mechanisms, participants were also invited to provide an open-ended written explanation describing the reasons behind their selections. All mechanisms that were not included among the participants' 5 preferred choices were automatically assigned a score of zero.

Analysis

Preference Matrix Validation

To compare the experimental results with the preference patterns reported in the literature, three matrices were generated.

The first matrix (M1) summarizes the preference relationships identified in the scoping review previously conducted [21]. Each cell represents the strength of evidence supporting a link between a Big Five personality trait and a gamification mechanism. The numerical value indicates the number of studies reporting a positive association (eg, 1=one study, 2=two studies, 3=three studies, and so on). A value of –1 denotes an aversion relationship, 0 indicates that both preference and aversion were observed, and null corresponds to an absence of a reported relationship.

The second matrix (M2) was derived from the empirical data collected in this study. Each cell represents a discriminative value computed for the combination of a Big Five trait (column) and a gamification mechanism (row). This value reflects the difference between the average personality scores of participants who selected a given mechanism and those who did not, as calculated using the following formula: $C_{m,b}=round(\sum_{p=1}^{n}1_{p,m}\cdot S_{p,b}\sum_{p=1}^{n}1_{p,m}-\sum_{p=1}^{n}0_{p,m}\cdot S_{p,b}\sum_{p=1}^{n}0_{p,m})$ (1)

where $1_{p,m}=1$ if participant p selected mechanism m, $0_{p,m}=1$ if participant p did not select mechanism m, $S_{p,\beta}$ is the score of participant p on the Big Five dimension $_{\beta}$, and n is the total number of participants.

To facilitate interpretation, discriminative values were categorized based on their magnitude: 0 - 1 (minimal selection of the mechanism by that trait), 2 (moderate selection), 3 (high selection), and 4+ (very high selection).

The third matrix (M3) represents the difference between M1 and M2, allowing direct visual and quantitative comparison between literature-based expectations and experimental results. Positive values indicate stronger associations in the current data than previously reported, whereas negative values indicate weaker or opposing trends. Cells with "null" values in M1 were excluded from the difference calculations.

Statistical Analysis

A logistic regression analysis was conducted to examine the relationship between the mechanisms and the Big Five personality dimensions. The purpose of this analysis was to determine whether participants' scores on the BFI-10-Fr significantly predicted the likelihood of selecting a given mechanism. Accordingly, logistic regression analysis was conducted for each mechanism to ascertain whether the BFI-10-Fr scores were predictive of the mechanism selection.

In addition, an ordinal logistic regression analysis was performed using the motivation scores assigned to the selected mechanisms as dependent variables and the BFI-10-Fr scores as predictors. A distinct model was computed for each mechanism to evaluate whether personality scores predicted the level of motivation attributed to that mechanism.

To reduce the risk of type I errors, the Bonferroni correction was applied across all regression analyses. This adjustment was implemented to account for the multiple comparisons conducted across mechanisms and personality dimensions, thereby minimizing the likelihood of false positive results.

Results

Demographics Data

A total of 214 participants completed the questionnaire, including 118 (55.1%) women, 89 (41.6%) men, 5 (2%) individuals identified as "other," and 2 (1%) participants who did not report their gender. The mean age of the sample was 29.42 (SD 10.41) years. Additional demographic characteristics are presented in Table 3.

XSL•FO
RenderX

Table . Demographics data. Participants were recruited both within the University of Geneva community and through social media platforms, allowing for the inclusion of individuals beyond the academic environment (N=214).

Characteristics	Values
Age (y), mean (SD)	29.42 (10.41)
Gender (N=214), n (%)	
Women	89 (41.6)
Men	118 (55.1)
Others	7 (3.3)
Education level (n=211[a]), n (%)	
Mandatory education	53 (24.8)
Bachelor's degree	63 (29.4)
Master's degree	80 (37.4)
Doctorate	15 (7)
Smartphone use (N=214), n (%)	
Not comfortable	10 (4.7)
Comfortable	204 (95.3)
Already used mHealth[b], n (%)	135 (63.1)

[a]Education level: 3 missing responses (n=211).

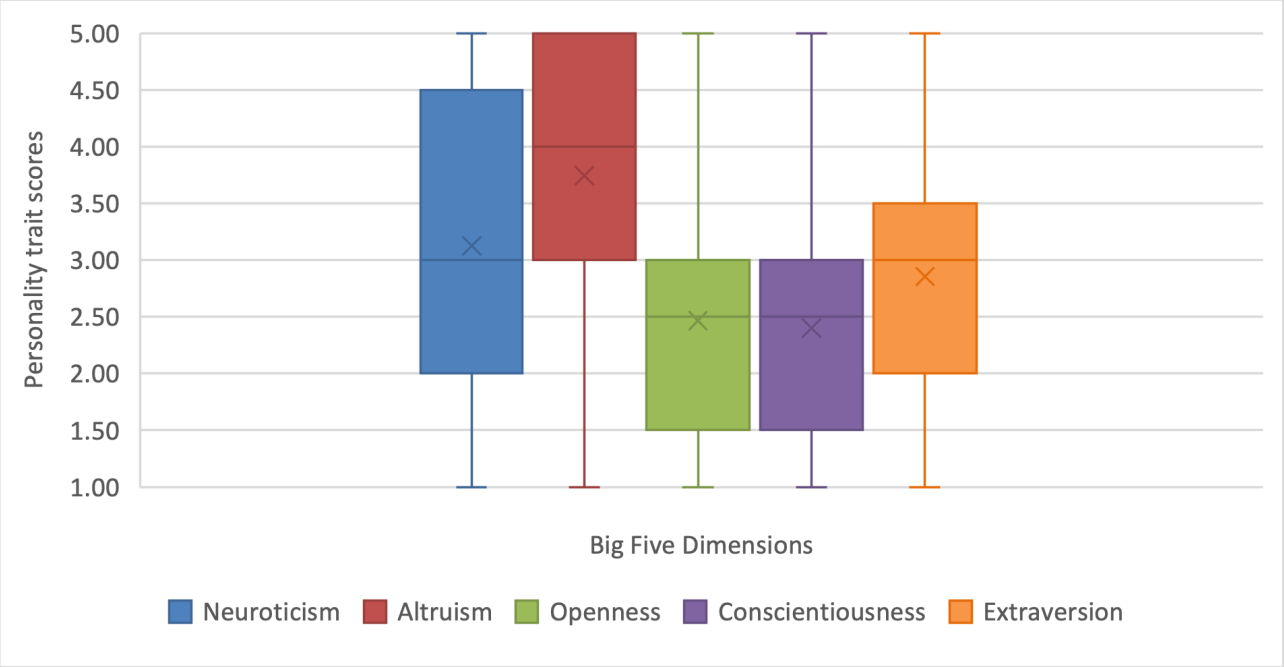
[b]mHealth: mobile health.

Characteristics of the Participants on the Big Five

Personality trait scores were measured on a 1 to 5 scale. The distribution of the sample across the Big Five dimensions is presented in Figure 1, where the line depicts the median and the cross indicates the mean.

Overall, participants in this study tended to score higher on agreeableness (mean 3.71, SD 1.31), followed by neuroticism (mean 3.12, SD 1.23) and extraversion (mean 2.86, SD 1.17).

Figure 1. Boxplot of the representativeness of the participants on the Big Five (N=214).



XSL•FO
RenderX

Characteristics of Participants' Selections for Each Mechanism

A total of 4 mechanisms were selected by more than half of the participants. These mechanisms were self-monitoring, progression, challenge, and quest (Table 4).

Table . Selection frequency of each mechanism and mean Big Five trait scores among participants (N=214).

Mechanisms	People who selected the mechanism, n (%)	Neuroticism, mean (SD)	Agreeableness, mean (SD)	Openness, mean (SD)	Conscientiousness, mean (SD)	Extraversion, mean (SD)
Self-monitoring	149 (69.6)	3.1 (1.2)	3.8 (1.3)	2.5 (1.0)	2.3 (1.0)	2.9 (1.1)
Progression	127 (59.3)	3.1 (1.2)	3.7 (1.3)	2.6 (1.0)	2.3 (0.9)	2.9 (1.2)
Challenge	112 (52.3)	3.2 (1.6)	3.7 (1.4)	2.4 (1.0)	2.4 (1.0)	2.9 (1.2)
Quests	111(51.9)	3.0 (1.2)	3.8 (1.3)	2.6 (1.0)	2.3 (0.9)	2.9 (1.2)
Cooperation	70 (32.7)	3.1 (1.2)	3.8 (1.3)	2.3 (1.0)	2.3 (0.9)	3.0 (1.3)
Demonstration of the behavior	61 (28.5)	3.2 (1.2)	3.6 (1.3)	2.3 (0.9)	2.3 (1.0)	2.7 (1.2)
Prompt and cues	53 (24.8)	3.3 (1.3)	3.6 (1.3)	2.6 (1.1)	2.4 (0.9)	2.8 (1.2)
Rewards	47 (2%)	3.4 (1.4)	3.6 (1.3)	2.5 (1.0)	2.8 (1.0)	2.8 (1.3)
Social comparison	44 (20.6)	3.2 (1.2)	3.7 (1.3)	2.7 (1.0)	2.6 (0.9)	2.8 (0.9)
Collectible	41 (19.2)	2.9 (1.3)	3.5 (1.6)	2.5 (1.0)	2.7 (1.0)	3.0 (1.2)
Avatar	29 (13.6)	3.2 (1.3)	3.8 (1.5)	2.6 (0.9)	2.5 (0.8)	3 (1.3)
Competition	29 (13.6)	2.8 (1.3)	3.9 (1.4)	2.4 (0.9)	2.8 (1.1)	2.7 (1.2)
Social support	20 (9.3)	2.9 (1.2)	4.0 (1.0)	2.5 (1.0)	2.4 (1.1)	2.9 (1.1)
Punition	8 (3.7)	3.3 (1.5)	3.5 (1.3)	2.2 (0.7)	2.3 (1.0)	3.1 (1.2)

Comparison Between This Study's Results and the Preference Matrix

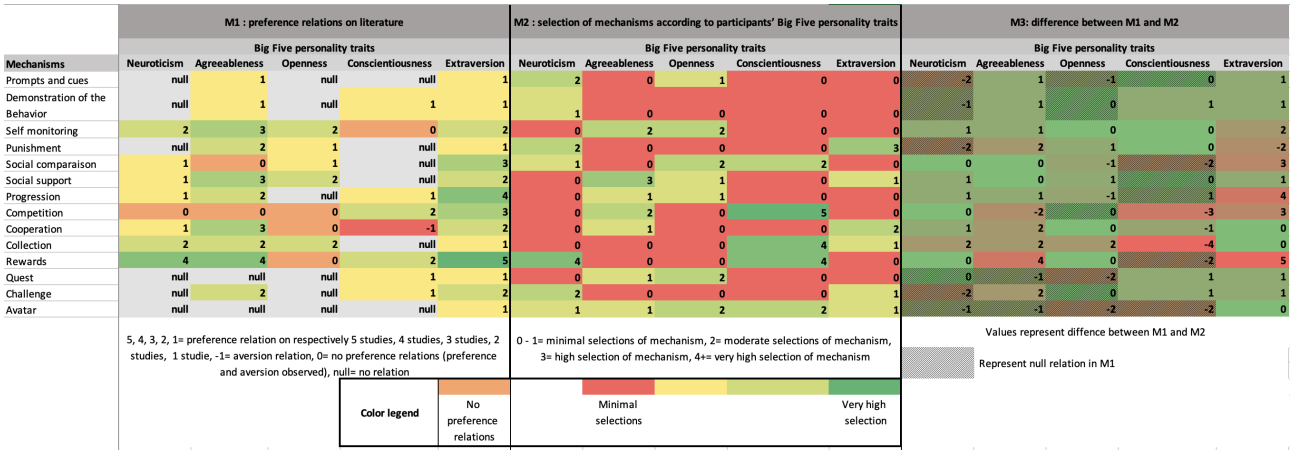
The initial matrix (M1) summarizes the preference relations identified in the literature between personality traits and mechanisms. A positive relation indicates that individuals with a given Big Five profile tend to prefer a particular mechanism, whereas a negative relation (aversion) reflects a lower interest or dislike for that mechanism among individuals with that profile. Mixed relations correspond to cases where both preference and aversion were reported across studies, while missing relations denote the absence of evidence. Overall, M1 presents 60% (42/70) positive preference relations, 1% (1/70) aversion relations, 10% (7/70) mixed relations, and 29% (20/70) missing relations (see M1 in Figure 2).

The second matrix (M2) represents the empirical data collected in this study. For each selection mechanism (eg, collection, challenge, or reward), a discriminative score was calculated. This score indicates how strongly a given personality trait differentiates between participants who selected a mechanism and those who did not. M2 shows that 5% (4/70) of mechanisms had a very high discriminative score (>4), 3% (2/70) high (3), 17% (12/70) moderate (2), and 74% (52/70) low (0 - 1; see M2 in Figure 2).

The second matrix (M2) represents the empirical data collected in this study. For each selection mechanism—defined as one of the gamified motivational strategies participants could choose (eg, collection, challenge, or reward)—a discriminative score was calculated. This score indicates how strongly a given personality trait differentiates between participants who selected a mechanism and those who did not. M2 shows that 5% (4/70) of mechanisms had a very high discriminative score (>4), 3% (2/70) high (3), 17% (12/70) moderate (2), and 74% (52/70) low (0 - 1; see M2 in Figure 2).

XSL•FO
RenderX

Figure 2. Matrices for the representation of preference relations in literature (M1), the selection mechanisms used by participants based on their Big Five traits scores (M2), and the difference between M1 and M2 (M3).



Preferred Mechanisms and Big Five: Statistical Analysis

A separate logistic regression analysis was performed for each mechanism to determine whether the personality scores predicted the likelihood of selecting that mechanism. In addition, an ordinal logistic regression was conducted using the motivation ratings assigned to the selected mechanisms as the dependent variables and the BFI-10-Fr scores as the predictors. Across analyses, 3 mechanisms yielded statistically significant models.

Collection Mechanism

For the collection mechanism, defined as a feature that allows users to gather virtual objects, the regression model was statistically significant (χ^2_5=13.07; n=207; P<.05), indicating that the model successfully distinguished between participants who selected the collection mechanism and those who did not. The model explained between 6% (Cox and Snell R^2) and 10% (Nagelkerke R^2) of the variance in mechanism selection and correctly classified 83.1% of cases. As shown in Table 5, only 1 predictor made a unique statistically significant contribution: conscientiousness with an odds ratio (OR) of 1.87 (95% CI 1.27-2.75). This indicates that participants scoring higher on conscientiousness were over 1.87 times more likely to select the collection mechanism compared with those scoring lower on this trait, after controlling for all other predictors in the model.

Table . Logistic regression predicting the likelihood of selecting the collection mechanism.

Personality trait	B	SE	Wald chi-square (*df*)	*P* value	OR[a] (95% CI)
Openness	–0.019	0.193	0.010 (1)	.92	0.981 (0.672-1.432)
Conscientiousness	0.625	0.198	9.987 (1)	.002	1.868 (1.268-2.752)
Extraversion	0.035	0.168	0.043 (1)	.84	1.035 (0.745-1.439)
Agreeableness	–0.165	0.146	1.279 (1)	.26	0.848 (0.637-1.129)
Neuroticism	–0.191	0.163	1.376 (1)	.24	0.826 (0.600-1.137)
Constant	–2.089	1.093	3.654 (1)	.06	0.124 (0.015-1.054)

[a]OR: odds ratio.

Competition Mechanism

The logistic regression for the competition mechanism, a feature emphasizing performance comparison or rivalry between users, was statistically significant (χ^2_5=31.49; n=200; P<.01). This result indicates that the model successfully distinguished between participants who selected this mechanism and those who did not, based on their Big Five personality scores. The regression explained between 15% (Cox and Snell R^2) and 33% (Nagelkerke R^2) of the variance in the selection of the competition mechanism and correctly classified 91.5% of cases. As shown in Table 6, 4 personality dimensions made a unique statistically significant contribution to the model: neuroticism, conscientiousness, extraversion, and agreeableness. Conscientiousness emerged as the strongest positive predictor (OR 3.22, 95% CI 1.73-6.00), followed by agreeableness (OR 1.93, 95% CI 1.08-3.45). In contrast, neuroticism (OR 0.43, 95% CI 0.25-0.74) and extraversion (OR 0.52, 95% CI 0.30-0.89) were negatively associated with selecting the competition mechanism, indicating that individuals with higher levels of these traits were substantially less likely to choose this feature.

XSL•FO
RenderX

Table . Logistic regression predicting the likelihood of selecting the competition mechanism.

Personality trait	B	SE	Wald chi-square (*df*)	*P* value	OR[a] (95% CI)
Openness	–0.195	0.286	0.466 (1)	.50	0.823 (0.469-1.441)
Conscientiousness	1.169	0.318	13.532 (1)	<.001	3.220 (1.727-6.003)
Extraversion	–0.660	0.280	5.574 (1)	.02	0.517 (0.299-0.894)
Agreeableness	0.660	0.296	4.969 (1)	.03	1.934 (1.083-3.454)
Neuroticism	–0.854	0.279	9.364 (1)	.002	0.426 (0.246-0.736)
Constant	–3.680	1.798	4.187 (1)	.04	0.025 (0.001-0.856)

[a]OR: odds ratio.

An ordinal logistic regression was conducted for each mechanism to determine whether participants' Big Five personality scores predicted their motivation ratings for the competition mechanism. The model for the competition mechanism was statistically significant (χ^2_5=11.16; N=214; *P*=.05), indicating that the predictors reliably differentiated participants' motivation levels for engaging in mHealth apps through competition-based features. The model explained between 5% (Cox and Snell R^2) and 8% (Nagelkerke R^2) of the variance in the motivation scores. As shown in Table 7, 2 personality traits made unique and statistically significant contributions to the model: conscientiousness and neuroticism. Higher conscientiousness was associated with greater motivation to use competition-based features (OR 1.63, 95% CI 0.08-0.89), whereas higher neuroticism was associated with lower motivation (OR 0.71, 95% CI 0.70-0.01).

Table . Logistic regression predicting the likelihood of the score of the competition mechanism.

Personality trait	B	SE	Wald chi-square (*df*)	*P* value	OR[a] (95% CI)
Openness	–0.072	0.207	0.121 (1)	.73	0.931 (–0.478 to 0.553)
Conscientiousness	0.487	0.206	5.586 (1)	.02	1.628 (0.083 to 0.892)
Extraversion	–0.272	0.191	2.028 (1)	.15	0.762 (–0.647 to 0.102)
Agreeableness	0.206	0.177	1.357 (1)	.24	1.229 (–0.141 to 0.553)
Neuroticism	–0.345	0.179	3.705 (1)	.05	0.708 (–0.696 to 0.006)

[a]OR: odds ratio.

Reward Mechanism

The logistic regression for the reward mechanism, a feature offering virtual incentives for engaging in the target behavior, was statistically significant (χ^2_5=35.29; n=204; *P*<.001). This indicates that the model successfully distinguished participants who selected this mechanism from those who did not, based on their Big Five personality traits. The model explained between 16% (Cox and Snell R^2) and 27% (Nagelkerke R^2) of the variance and correctly classified 81.4% of cases. As shown in Table 8, 2 personality traits made unique statistically significant contributions to the model: conscientiousness and neuroticism. Participants scoring higher on conscientiousness were 2.36 times more likely to select the reward mechanism, while those scoring higher on neuroticism were 1.97 times more likely to choose it, controlling for all other predictors.

Table . Logistic regression predicting likelihood of selecting the reward mechanism.

Personality trait	B	SE	Wald chi-square (*df*)	*P* value	OR[a] (95% CI)
Openness	0.060	0.194	0.095 (1)	.76	1.061 (0.726-1.551)
Conscientiousness	0.858	0.220	15.262 (1)	<.001	2.360 (1.534-3.630)
Extraversion	–0.220	0.187	1.392 (1)	.24	0.802 (0.557-1.157)
Agreeableness	0.068	0.161	0.177 (1)	.67	1.070 (0.781-1.466)
Neuroticism	0.679	0.190	12.696 (1)	<.001	1.971 (1.357-2.864)
Constant	–5.919	1.327	19.897 (1)	<.001	0.003 (0.0002-0.036)

[a]OR: odds ratio.

The logistic ordinal regression was statistically significant (χ^2_5=14.36; N=214; *P*=.01), indicating that the model successfully differentiated participants according to their motivation scores for engaging with this mechanism. The model explained between 6% (Cox and Snell R^2) and 9% (Nagelkerke R^2) of the variance. As shown in Table 9, 2 personality traits made a unique statistically significant contribution to the model: neuroticism and conscientiousness. Conscientiousness was the

XSL•FO
RenderX

strongest predictor (OR 1.66, 95% CI 0.16-0.85), followed by neuroticism (OR 1.38, 95% CI 0.03-0.61), indicating that higher levels of these traits were associated with greater self-reported motivation to use the rewards mechanism.

Table . Logistic regression predicting the likelihood of the score of the reward mechanism.

Personality trait	B	SE	Wald chi-square (*df*)	*P* value	OR[a] (95% CI)
Openness	0.025	0.170	0.023 (1)	.88	1.026 (–0.307 to 0.358)
Conscientiousness	0.504	0.176	8.199 (1)	.004	1.66 (0.159 to 0.849)
Extraversion	–0.031	0.156	0.040 (1)	.84	0.969 (–0.336 to 0.274)
Agreeableness	–0.083	0.132	0.389 (1)	.53	0.921 (–0.342 to 0.177)
Neuroticism	0.320	0.149	4.606 (1)	.03	1.378 (0.028 to 0.61)

[a]OR: odds ratio.

Discussion

Principal Findings

The objective of this study was to validate the preference matrix derived from a previous literature review by examining whether similar relations could be observed empirically in our sample.

Four mechanisms were selected by more than half of the participants: self-monitoring (n=149), progression (n=127), challenge (n=112), and quest (n=111). No significant associations were found between these mechanisms and the Big Five personality dimensions. However, when comparing these findings with the validated preference matrix (M1), a more nuanced interpretation emerges. According to the matrix (M1), self-monitoring is generally appreciated across all Big Five profiles, with the exception of one study reporting lower preference among highly conscientious individuals. The progression mechanism also appears to be universally preferred, reinforcing its broad motivational relevance. By contrast, the preference matrix (M1) suggests that the quest mechanism is predominantly preferred by conscientious and extraverted individuals, whereas the challenge mechanism tends to be favored by users scoring high in agreeableness, conscientiousness, and extraversion. Taken together, these results indicate that certain gamification mechanisms, particularly self-monitoring and progression, possess strong, cross-profile motivational appeal. These mechanisms should, therefore, be considered foundational elements in the design of mHealth apps, providing a core layer of user engagement upon which more personalized, profile-specific interventions may be developed.

Comparison Between M1 and M2

The comparison between matrices M1 and M2 highlights several developments in the identification of potential relationships between mechanisms and Big Five personality traits. In M2, new associations emerge that were not identified in M1, particularly between neuroticism and mechanisms such as prompts and cues, demonstration of behavior, punishment, quest, challenge, and avatar. Similarly, openness shows newly emerging association with prompts and cues, progression, quest, and challenge, while conscientiousness displays novel relations with social comparison and avatar, as well as a particularly strong association with the collection mechanism in M2. These findings reflect a refined and evolving understanding of how specific personality traits may align with distinct mechanisms when examined in an experimental context.

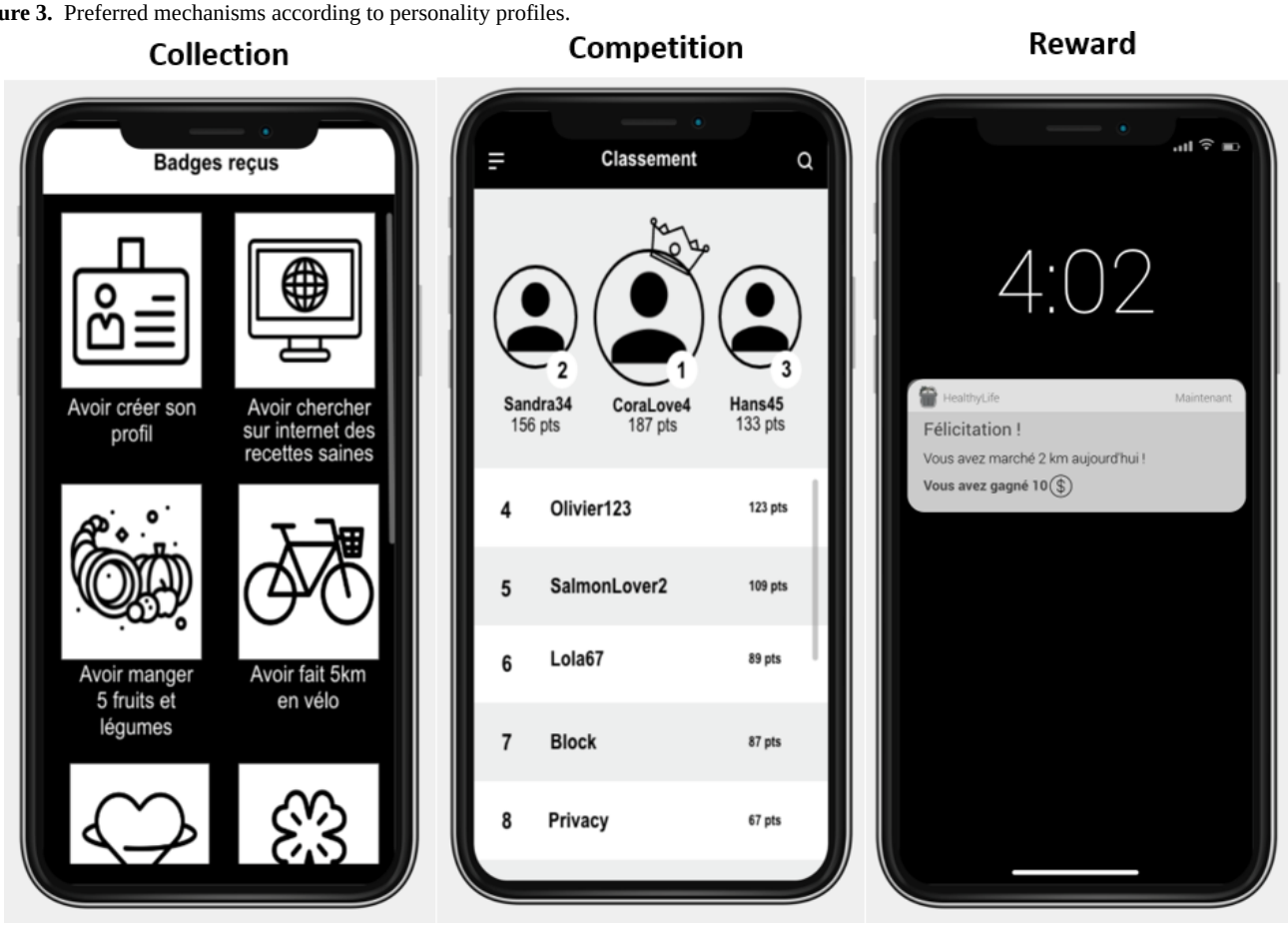
In addition, whereas M1 did not show a clear consensus regarding the relationships involving agreeableness and openness, M2 reveals emerging associations, notably between agreeableness and competition, and between openness and rewards. Conversely, several relationships previously identified in M1 do not reappear in M2. For example, extraversion was associated in M1 with a wide range of mechanisms, including prompts and cues, demonstration of behavior, self-monitoring, punishment, social comparison, social support, progression, competition, cooperation, collection, rewards, and quest, but these associations were not observed in M2. A similar reduction in associations is observed for agreeableness, which in M1 was linked to prompts and cues, demonstration of behavior, punishment, collection, challenge, and especially rewards. Despite these discrepancies, statistically significant results in this study were identified for only 3 mechanisms: collection, competition, and rewards. Of the 5 significant relationships observed, 3 were already present in the preference matrix M1 (conscientiousness-competition, conscientiousness-rewards, and neuroticism-rewards). Two relationships, those involving neuroticism-competition and agreeableness-competition, had shown inconsistent evidence in M1, with both preference and nonpreference relations reported in the literature.

This study contributes to refining the preference matrix in two ways. First, it enriches the matrix by identifying a new significant preference relation between conscientiousness and the collection mechanism. Second, in cases where the literature reports incongruent findings, our results help clarify these relationships by supporting a nonpreference relation for neuroticism and agreeableness with competition.

In the following sections, we examine in greater detail the 3 mechanisms for which significant regressions were obtained. Figure 3 presents the mockups used in the online survey to illustrate the 3 mechanisms preferred by participants according to their personality traits. These visual representations supported participants' understanding of each mechanism and facilitated more reliable self-selection during the experimental procedure.

XSL•FO
RenderX

Figure 3. Preferred mechanisms according to personality profiles.



Collection

In this study, the collection mechanism refers to the accumulation of virtual items or badges over time. Collecting has been identified as a motivational strategy that can support engagement in health-related behaviors, such as maintaining physical activity [28]. Our results show that participants scoring higher on conscientiousness were 1.87 times more likely to select the collection mechanism than those with lower levels of this trait.

These findings are largely consistent with the existing literature, which reports a general preference for collection-based mechanisms across most Big Five personality traits in the preference matrix M1 [9,24,29-31]. Notably, however, none of the previously reviewed studies identified a specific preference for collection among highly conscientious individuals. As such, the present result constitutes a novel contribution to the preference matrix by establishing a new, trait-specific association between conscientiousness and the collection mechanism.

Competition

In this study, the competition mechanism is defined as enabling users to engage in competitive interaction with others in order to encourage the target behavior [23].

According to the preference matrix M1, competition is associated with all 5 Big Five personality traits. However, our results reveal a more differentiated pattern. Individuals scoring high in conscientiousness (OR 3.22, 95% CI 1.73-6.00) and agreeableness (OR 1.93, 95% CI 1.08-3.45) were significantly more likely to select competition as a motivating mechanism for maintaining their health compared with participants scoring lower on these traits. In contrast, participants with higher levels of extraversion (OR 0.43, 95% CI 0.25-0.74) and neuroticism (OR 0.52, 95% CI 0.30-0.89) were significantly less likely to select this mechanism.

Consistent with these findings, the ordinal regression analysis showed that participants with higher conscientiousness scores rated the competition mechanism as more motivating (OR 1.63, 95% CI 0.08-0.89), whereas those with higher neuroticism scores assigned lower motivation ratings (OR 0.71, 95% CI 0.70-0.01).

These divergent results suggest a controversial or polarized preference for competitive mechanisms. Qualitative feedback indicated that some participants favored individual-based mechanisms that emphasize personal effort rather than social comparison or rivalry. This interpretation aligns with previous research, which reports similar reservations toward competitive features in health-related apps, particularly when users prefer self-directed rather than socially driven motivation [10].

> *No comparison with others because it stresses me out if I don't win. I'm more motivated by the idea of doing myself good.* [P14]

> *I choose for myself, not for others, so there's no competition or comparison with others.* [P39]

XSL•FO
RenderX

> *I've excluded those that involve other people because I prefer autonomy.* [P129]
>
> *I no longer have anything to prove to anyone but myself.* [P134]
>
> *The individuality aspect. In my opinion, when you do something, you do it for yourself and not for others.* [P112]
>
> *I think the principle of fitness should be personal and not a competition with others.* [P140]

While other participants said they found competition more motivating:

> *I find competition very enjoyable, it pushes me to excel*. [P44]
>
> *Comparing myself to others pushes me to do better.* [P176]
>
> *I'm very competitive and I like it when people show me that I've made progress. It encourages me to keep going.* [P192]

The preference matrix (M1) further indicates that individuals scoring high in agreeableness and conscientiousness tend to prefer competition [8,10,32]. However, it also reports a preference relation with extraversion [9,10,24] and neuroticism [33], whereas our results show that individuals high in these 2 traits selected the competition mechanism less frequently. This discrepancy is noteworthy, as the preference matrix (M1) is based on evidence from three studies reporting a positive association between extraversion and competition, which contrasts with the nonpreference pattern observed in our data. In addition, several studies examining competitiveness more broadly have documented a general attraction to competitive contexts among extraverted individuals, further emphasizing the unexpected nature of our findings [34,35]. One possible explanation lies in the heterogeneity of the extraversion construct itself: preferences for competitive environments appear to vary depending on specific subfacets of extraversion [36]. Individuals scoring higher on friendliness ("make friends easily" and "often feel uncomfortable around others") or cheerfulness ("have a lot of fun" and "am not easily amused") tend to show reduced interest in competitive dynamics. In contrast, those characterized by higher assertiveness ("take control of things" and "wait for others to lead the way") or excitement-seeking ("love action" and "dislike loud music") display stronger preferences for competition and achievement-oriented contexts [36]. These nuances suggest that the link between extraversion and preference for competitive mechanisms is not uniform and may be facet-dependent, an aspect that warrants deeper investigation in future research.

By contrast, the preference matrix (M1) also reports a nonpreference relation for individuals with higher levels of neuroticism profiles [32], a pattern that is corroborated by the results of our experiment.

Rewards

The reward mechanism involves the provision of virtual incentives to users for completing a target behavior [23]. It differs from the collection mechanism in that rewards are not necessarily accumulated as part of a predefined or structured set of items.

In our study, individuals with higher levels of neuroticism (OR 1.97, 95% CI 1.36-2.86) and conscientiousness (OR 2.36, 95% CI 1.53-3.63) were significantly more likely to select the reward mechanism. In addition, these participants assigned higher motivation scores to this mechanism ($OR_{Conscientiousness}$1.66, 95% CI 0.16-0.85; $OR_{Neuroticism}$1.38, 95% CI 0.03-0.61).

These findings are consistent with the preference matrix (M1), which reports established preference relations between the reward mechanisms and both conscientiousness and neuroticism across multiple studies [24,30,32,33,37,38].

Big Five Dimensions

Notably, conscientiousness emerged as a significant predictor across all 3 mechanisms identified in this study. In contrast, no significant associations were observed for openness, despite prior literature suggesting preferences for certain mechanisms, such as collection and competition, among individuals high in this trait.

Similarly, no significant preference results were identified for extraversion, even though the preference matrix derived from the literature (M1) reports associations between extraversion and multiple mechanisms. A comprehensive summary of the experimental findings alongside the preference matrix is presented in Table 10.

XSL•FO
RenderX

Table . Preferred and nonpreferred mechanisms for each Big Five dimension.

Big Five dimension	Mechanisms preferred on this study	Mechanisms nonpreferred on this study	Mechanisms preferred on preference matrix (M1)	Mechanisms nonpreferred on preference matrix (M1)
Openness to experience	—[a]	—	Self-monitoring, punishment, social comparison, social support, competition, cooperation, collection, rewards, and customization	Rewards, competition, and cooperation
Agreeableness	*Competition*[b,c]	—	Prompts and cues, demonstration of the behavior, self-monitoring, punishment, social comparison, social support, progression, competition[b], cooperation, collection, rewards, challenge, and customization	Social comparison and *competition*
Conscientiousness	Collection, competition[b], and rewards[b]	—	Demonstration of the behavior, self-monitoring, progression, competition[b], rewards[b], quest, and challenge	Self-monitoring, cooperation, and customization
Extraversion	—	*Competition*	Prompts and cues, demonstration of the behavior, self-monitoring, punishment, social comparison, social support, progression, *competition*, cooperation, collection, rewards, quest, challenge, avatar, and customization	Customization
Neuroticism	Rewards[b]	*Competition*[b]	Self-monitoring, social comparison, social support, progression, *competition*, cooperation, collection, rewards[b], and customization	Competition[b]

[a]Not applicable.

[b]Indicates findings consistent with the literature

[c]*Italicized text* indicates findings inconsistent with the literature.

Applying the Preference Matrix in Practice

Implementing the preference matrix involves 2 main steps: identifying the user profile and tailoring the app content accordingly. User profiling traditionally relies on validated personality questionnaires (eg, BFI-10 and NEO-PI-R); however, these instruments can be time-consuming and may discourage user participation. As an alternative, automated profiling approaches based on smartphone usage patterns [39,40] or social media data [41-43] have shown promising results, although privacy concerns remain a major limitation. Such approaches require access to personal data, sometimes including sensitive inputs such as audio or video, which many users are reluctant to share [44]. Moreover, extensive data collection may increase perceptions of intrusiveness or manipulation, potentially undermining user trust and the perceived validity of behavior-based personalization [44]. To address these concerns, we recommend making personalization optional, offering users the choice to complete a brief, validated personality questionnaire if they opt in, and implementing strict ethical safeguards to ensure responsible data use. In particular, personality-related data should be collected solely for the purpose of enabling individualized personalization within the app, should not be repurposed for any secondary uses or shared with third parties, and should be handled with full transparency. Users must be clearly informed about what information is being collected, how it will be processed, and for what specific purpose (to tailor the app's features to their personality profile), thereby promoting informed consent, user autonomy, and trust.

Once the profile is established, app content can be adapted by selecting mechanisms aligned with the user's dominant personality traits, as defined by the preference matrix. For example, a user scoring high in conscientiousness might be shown mechanisms emphasizing rewards, competition, and collection, while competitive features could be intentionally avoided for a more extraverted user. This automated adaptation enhances user engagement by aligning motivational strategies with intrinsic preferences derived from psychological profiling.

However, to ensure that personalization remains user-centered and flexible, the system should also allow users to manually adjust or refine these automatically suggested mechanisms. After reviewing the proposed configuration, users can confirm, modify, or reject specific mechanisms based on their subjective preferences or contextual needs. This hybrid approach, combining data-driven profiling with user autonomy, maximizes

XSL•FO
RenderX

both the accuracy and acceptability of personalization. It also mitigates potential mismatches between predicted and actual motivation, ensuring that the final app configuration genuinely reflects the user's individual preferences.

Limitations

The requirement for participants to select exactly 5 mechanisms out of 15 constitutes a methodological limitation. Some participants may have selected mechanisms they did not find particularly motivating to reach the required number, while others may have wished to select more than 5 mechanisms but were constrained by the study design. This forced-choice approach may, therefore, have influenced the distribution of selected mechanisms and attenuated individual preference differences.

In addition, our study sample was not fully representative of the general population, as it included a high proportion of university students, who are typically younger ($mean_{age}$ 29.42, SD_{age} 10.41). Such demographic homogeneity may restrict the extent to which these findings can be generalized to other age groups or populations. Nevertheless, prior research on digital behavior changes and web-based health interventions suggests that age does not substantially moderate the effects of these interventions on health-related behaviors [45,46]. Systematic reviews have similarly reported that demographic variables such as age, sex, and education tend to have weak or inconsistent moderating influences on the effectiveness of these techniques [45]. Moreover, individuals most inclined to engage with health apps are generally younger and possess higher levels of education and income [47], which aligns closely with the characteristics of our sample and supports the external validity of our findings. Future research should, however, aim to include older adults and individuals less familiar with digital technologies to examine whether similar patterns of preference and engagement are observed in these groups. This population is particularly relevant to investigate, as older adults also represent an active and growing user base of mHealth apps, often using them for health monitoring, disease prevention, or physical activity tracking. Recent evidence from a comprehensive scoping review underscores the growing importance of this demographic and highlights that addressing key barriers, such as evolving capability requirements, cost, privacy concerns, and age-related design assumptions, can substantially enhance meaningful engagement. It also emphasized the importance of empowering older adults through learner-centered, need-based digital education [48]. Building on these insights, future studies should explore how mHealth apps can be designed to support sustained digital engagement among older adults and to better capture the diversity of their experiences and values in technology use.

Another limitation concerns the personality assessment instrument used in this study. We used the BFI-10-Fr, an ultra-short version of the Big Five Inventory, to minimize participant burden given the overall length of the online questionnaire. Although the BFI-10-Fr provides a practical and time-efficient measure, previous research has noted that its internal reliability can be relatively low due to the limited number of items per trait [49]. Nevertheless, according to the original validation study by Rammstedt and John [49], the BFI-10 demonstrates satisfactory test–retest stability and acceptable construct validity, supporting its suitability for large-scale or time-constrained surveys. Future studies should, however, consider replicating these analyses using a more comprehensive personality inventory, such as the Big Five Inventory-44 or NEO-PI-R, to ensure higher internal consistency and further validate the robustness of the associations observed.

We also observe that, despite our sample size calculation, it appears that it did not allow for proper discrimination, as we obtained significant results for only 3 out of 14 mechanisms. Therefore, this calculation should be reconsidered.

It was not feasible to assess the customization mechanism, given its expansive scope and the consequent challenges in representing it in the format of a conventional mockup screen.

Although the regression analyses yielded statistically significant results for several mechanisms, the explanatory power of the models remained relatively modest. For instance, the proportion of variance explained ranged from 6% to 10% for the collection mechanism, and from 5% to 9% for the ordinal regressions predicting motivation scores. Even for the most predictive models, such as competition (15% - 33%) and Rewards (16% - 27%), a substantial portion of the variance remained unexplained. This indicates that while personality traits, particularly conscientiousness and neuroticism, appear to play a meaningful role in the selection of gamification mechanisms, other unmeasured variables likely contribute to these motivational choices. Therefore, the practical significance of these findings should be interpreted with caution: the Big Five dimensions alone may not fully account for users' preferences, which are likely shaped by situational, contextual, or motivational factors beyond personality. Future studies could aim to include such variables, or use alternative modeling approaches (eg, interaction terms or structural models) to capture more of the underlying variance in mechanism selection.

Conclusions

The entire preference matrix could not be validated with experimental data. Significant relations were identified for three mechanisms only: collection, competition, and rewards. Most of these preference relations are consistent with those reported in the literature-based preference matrix (M1), apart from the preference relation for the collection mechanism and conscientiousness and the nonpreference relation for competition regarding the extraversion trait, which is significant in this study but, according to the preference matrix M1, is a preference relation. The neuroticism trait is significant for competition as nonpreferred, yet the preference matrix studies demonstrated both a preference and a nonpreference relation for this trait. Considering these findings, it can be concluded that a nonpreference relation is corroborated. It would be interesting to replicate this study with a larger, less student-dominated panel to achieve better representation and more significant results on mechanism preference relations by Big Five traits.

XSL•FO
RenderX

Acknowledgments

This manuscript was developed with the assistance of ChatGPT (OpenAI, GPT-5 model), which was used to improve the clarity, conciseness, and scientific phrasing of the English text. The authors reviewed, verified, and are fully responsible for all content and interpretations.

Funding

This work has been cofunded by the European Union and the Swiss State Secretariat for Education, Research and Innovation, in the frame of the SHIELD (Strategic Health Initiatives for Effective Disease Prevention; grant 101156751) project. Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the Swiss State Secretariat for Education, Research and Innovation. Neither the European Union nor the granting authorities can be held responsible for them.

Data Availability

The datasets generated during the study are available from the corresponding author upon reasonable request.

Authors' Contributions

LG conceived the study with the involvement and advice of FE and GF. All authors were involved in writing, reading, and approving the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Relation between Big Five traits and mechanisms.
[DOCX File, 17 KB - humanfactors_v13i1e78939_app1.docx]

Multimedia Appendix 2
Print version of the online questionnaire.
[DOCX File, 1458 KB - humanfactors_v13i1e78939_app2.docx]

Checklist 1
CHERRIES checklist.
[PDF File, 95 KB - humanfactors_v13i1e78939_app3.pdf]

References

1. Wingard DL, Berkman LF, Brand RJ. A multivariate analysis of health-related practices: a nine-year mortality follow-up of the Alameda County Study. Am J Epidemiol 1982 Nov;116(5):765-775. [doi: 10.1093/oxfordjournals.aje.a113466] [Medline: 7148802]
2. Health tech mobile apps: landscape overview. : Deep Knowledge Ventures; 2018 URL: http://analytics.dkv.global/data/pdf/Health-Tech-Mobile-Apps-Analytical-Report.pdf [accessed 2022-04-08]
3. Ernsting C, Dombrowski SU, Oedekoven M, et al. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Apr 5;19(4):e101. [doi: 10.2196/jmir.6838] [Medline: 28381394]
4. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
5. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: a systematic review. J Telemed Telecare 2018 Jan;24(1):22-30. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
6. Skinner CS, Campbell MK, Rimer BK, Curry S, Prochaska JO. How effective is tailored print communication? Ann Behav Med 1999;21(4):290-298. [doi: 10.1007/BF02895960] [Medline: 10721435]
7. Gosetto L, Ehrler F, Falquet G. Personalization dimensions for mHealth to improve behavior change: a scoping review. Stud Health Technol Inform 2020 Nov 23;275:77-81. [doi: 10.3233/SHTI200698] [Medline: 33227744]
8. Halko S, Kientz JA. Personality and persuasive technology: an exploratory study on health-promoting mobile applications. Presented at: 5th International Conference on PERSUASIVE Technology (PERSUASIVE 2010); Jun 7-10, 2010; Copenhagen, Denmark p. 150-161. [doi: 10.1007/978-3-642-13226-1_16]

XSL•FO
RenderX

9. Jia Y, Xu B, Karanam Y, Voida S. Personality-targeted gamification: a survey study on personality traits and motivational affordances. Presented at: CHI'16: CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, CA p. 2001-2013. [doi: 10.1145/2858036.2858515]
10. Orji R, Nacke LE, Di Marco C. Towards personality-driven persuasive health games and gamified systems. 2017 Presented at: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems; May 6-11, 2017; Denver, Colorado, USA p. 1015-1027. [doi: 10.1145/3025453.3025577]
11. Hallifax S, Serna A, Marty JC, Lavoué G, Lavoué E. Factors to consider for tailored gamification. 2019 Presented at: Proceedings of the Annual Symposium on Computer-Human Interaction in PLAY (CHI PLAY); Oct 14-16, 2019; Barcelona, Spain p. 559-572. [doi: 10.1145/3311350.3347167]
12. Lipnevich AA, Gjicali K, Asil M, Smith JK. Development of a measure of receptivity to instructional feedback and examination of its links to personality. Pers Individ Dif 2021 Feb;169:110086. [doi: 10.1016/j.paid.2020.110086]
13. Heinström J. Five personality dimensions and their influence on information behavior. Inf Res 2003 Oct;9(1) [FREE Full text]
14. Johnson D, Wyeth P, Sweetser P, Gardner J. Personality, genre and videogame play experience. 2012 Sep 4 Presented at: 4th International Conference on FUN and GAMES (FUN & GAMES 2012); Sep 4-6, 2012. [doi: 10.1145/2367616.2367633]
15. Ferro LS, Walz SP, Greuter S. Towards personalised, gamified systems: an investigation into game design, personality and player typologies. Presented at: IE'2013: The 9th Australasian Conference on Interactive Entertainment; Sep 30 to Oct 1, 2013; Melbourne, Victoria, Australia p. 1-6. [doi: 10.1145/2513002.2513024]
16. Belmon LS, Middelweerd A, Te Velde SJ, Brug J. Dutch young adults ratings of behavior change techniques applied in mobile phone apps to promote physical activity: a cross-sectional survey. JMIR mHealth uHealth 2015 Nov 12;3(4):e103. [doi: 10.2196/mhealth.4383] [Medline: 26563744]
17. John OP, Naumann LP, Soto CJ. Paradigm shift to the integrative big five trait taxonomy: history, measurement, and conceptual issues. In: Handbook of Personality: Theory and Research, 3rd edition: Guilford Press; 2008:114-158 URL: https://www.sciepub.com/reference/247568 [accessed 2026-03-23]
18. Costa PT, McCrae RR. Stability and change in personality assessment: the revised NEO Personality Inventory in the year 2000. J Pers Assess 1997 Feb;68(1):86-94. [doi: 10.1207/s15327752jpa6801_7] [Medline: 9018844]
19. McCrae RR, Terracciano A, Personality Profiles of Cultures Project. Universal features of personality traits from the observer's perspective: data from 50 cultures. J Pers Soc Psychol 2005 Mar;88(3):547-561. [doi: 10.1037/0022-3514.88.3.547] [Medline: 15740445]
20. Costa P, McCrae R. Revised NEO Personality Inventory (NEO PI-R) and NEP Five-Factor Inventory (NEO-FFI): Professional Manual: Psychological Assessment Resources; 1992. URL: https://books.google.co.in/books?id=H_jXtgAACAAJ [accessed 2026-03-10]
21. Gosetto L, Falquet G, Ehrler F. Personalizing mobile applications for health based on user profiles: a preference matrix from a scoping review. PLOS Digit Health 2025 Aug;4(8):e0000978. [doi: 10.1371/journal.pdig.0000978] [Medline: 40828800]
22. Villalobos-Zúñiga G, Cherubini M. Apps that motivate: a taxonomy of app features based on self-determination theory. Int J Hum Comput Stud 2020 Aug;140:102449. [doi: 10.1016/j.ijhcs.2020.102449]
23. Orji R, Mandryk RL, Vassileva J. Improving the efficacy of games for change using personalization models. ACM Trans Comput-Hum Interact 2017 Oct 31;24(5):1-22. [doi: 10.1145/3119929]
24. Klock ACT, Gasparini I, Pimenta MS, Hamari J. Tailored gamification: a review of literature. Int J Hum Comput Stud 2020 Dec;144:102495. [doi: 10.1016/j.ijhcs.2020.102495]
25. Gosetto L, Pittavino M, Falquet G, Ehrler F. Personalization of mobile apps for health behavior change: protocol for a cross-sectional study. JMIR Res Protoc 2023 Jan 5;12:e38603. [doi: 10.2196/38603] [Medline: 36602850]
26. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34. [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
27. Gosetto L. Impact de la présentation d'affiches personnalisées en fonction du big-five sur l'intention de don du sang [Impact of personalized posters based on the Big Five on blood donation intention] [Master's Thesis] [Article in French]. : University of Geneva; 2018 URL: https://archive-ouverte.unige.ch/unige:110083 [accessed 2022-08-09]
28. Werbach K, Hunter D. For the Win: How Game Thinking Can Revolutionize Your Business: Wharton Digital Press; 2012. URL: https://books.google.co.in/books/about/For_the_Win.html?hl=es&id=aL2tzgEACAAJ&redir_esc=y [accessed 2026-03-23]
29. Tondello GF, Nacke LE. Validation of user preferences and effects of personalized gamification on task performance. Front Comput Sci 2020 Aug 6;2. [doi: 10.3389/fcomp.2020.00029]
30. Codish D, Ravid G. Personality based gamification: how different personalities percive gamification. Presented at: Proceedings of the 22nd European Conference on Information Systems (ECIS 2014); Jun 9, 2014 to Jun 11, 2022 URL: https://aisel.aisnet.org/ecis2014/proceedings/track12/10/ [accessed 2026-03-10]
31. Tondello GF, Mora A, Nacke LE. Elements of gameful design emerging from user preferences. 2017 Presented at: CHI PLAY 2017 - Proceedings of the Annual Symposium on Computer-Human Interaction in PLAY Association for Computing Machinery; Oct 15-18, 2017; Amsterdam, The Netherlands p. 129-140. [doi: 10.1145/3116595.3116627]

XSL•FO
RenderX

32. Ndulue C, Oyebode O, Iyer RS, Ganesh A, Ahmed SI, Orji R. Personality-targeted persuasive gamified systems: exploring the impact of application domain on the effectiveness of behaviour change strategies. User Model User Adapt Interact 2022;32(1-2):165-214. [doi: 10.1007/s11257-022-09319-w] [Medline: 35281337]
33. Anagnostopoulou E, Magoutas B, Bothos E, Schrammel J, Orji R, Mentzas G. Exploring the links between persuasion, personality and mobility types in personalized mobility applications. 2017 Presented at: 12th International Conference on PERSUASIVE Technology (PERSUASIVE 2017); Apr 4-6, 2017; Amsterdam, The Netherlands p. 107-118. [doi: 10.1007/978-3-319-55134-0_9]
34. Fong M, Zhao K, Smillie LD. Personality and competitiveness: extraversion, agreeableness, and their aspects, predict self-reported competitiveness and competitive bidding in experimental auctions. Pers Individ Dif 2021 Feb;169:109907. [doi: 10.1016/j.paid.2020.109907]
35. Ross SR, Rausch MK, Canada KE. Competition and cooperation in the five-factor model: individual differences in achievement orientation. J Psychol 2003 Jul;137(4):323-337. [doi: 10.1080/00223980309600617] [Medline: 12943183]
36. Fletcher TD, Nusbaum DN. Trait competitiveness as a composite variable: linkages with facets of the big-five. Pers Individ Dif 2008 Sep;45(4):312-317. [doi: 10.1016/j.paid.2008.04.020]
37. Alqahtani F, Meier S, Orji R. Personality-based approach for tailoring persuasive mental health applications. User Model User-Adap Inter 2022 Jul;32(3):253-295. [doi: 10.1007/s11257-021-09289-5]
38. Orji R. Exploring the persuasiveness of behavior change support strategies and possible gender differences. 2014 Presented at: Proceedings of the Second International Workshop on Behavior Change Support Systemsco-located with the 9th International Conference on PERSUASIVE Technology (PERSUASIVE 2014); May 22, 2014; Padua, Italy p. 41-57 URL: https://ceur-ws.org/Vol-1153/Paper_5.pdf [accessed 2026-03-10]
39. Chittaranjan G, Blom J, Gatica-Perez D. Who's who with big-five: analyzing and classifying personality traits with smartphones. Presented at: 15th Annual International Symposium on Wearable Computers; Jun 12-15, 2011. [doi: 10.1109/ISWC.2011.29]
40. Staiano J, Lepri B, Aharony N, Pianesi F, Sebe N, Pentland A. Friends don't lie: inferring personality traits from social network structure. 2012 Presented at: Proceedings of the 14th ACM International Conference on Ubiquitous Computing (UbiComp '12); Sep 5-8, 2012; Pittsburgh, PA p. 321-330. [doi: 10.1145/2370216.2370266]
41. Azucar D, Marengo D, Settanni M. Predicting the Big 5 personality traits from digital footprints on social media: a meta-analysis. Pers Individ Dif 2018 Apr;124:150-159. [doi: 10.1016/j.paid.2017.12.018]
42. Bai S, Zhu T, Cheng L. Big-five personality prediction based on user behaviors at social network sites. arXiv. Preprint posted online on Apr 21, 2012. [doi: 10.48550/arXiv.1204.4809]
43. Souri A, Hosseinpour S, Rahmani AM. Personality classification based on profiles of social networks' users and the five-factor model of personality. Hum Cent Comput Inf Sci 2018 Dec;8(1):24. [doi: 10.1186/s13673-018-0147-4]
44. Kim S, Thakur A, Kim J. Understanding users' perception towards automated personality detection with group-specific behavioral data. Presented at: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems; Apr 25-30, 2020. [doi: 10.1145/3313831.3376250]
45. Lustria MLA, Noar SM, Cortese J, Van Stee SK, Glueckauf RL, Lee J. A meta-analysis of web-delivered tailored health behavior change interventions. J Health Commun 2013;18(9):1039-1069. [doi: 10.1080/10810730.2013.768727] [Medline: 23750972]
46. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
47. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR mHealth uHealth 2015 Nov 4;3(4):e101. [doi: 10.2196/mhealth.4924] [Medline: 26537656]
48. Kebede AS, Ozolins LL, Holst H, Galvin K. Digital engagement of older adults: scoping review. J Med Internet Res 2022 Dec 7;24(12):e40192. [doi: 10.2196/40192] [Medline: 36477006]
49. Rammstedt B, John OP. Measuring personality in one minute or less: a 10-item short version of the Big Five Inventory in English and German. J Res Pers 2007 Feb;41(1):203-212. [doi: 10.1016/j.jrp.2006.02.001]

Abbreviations

BFI-10: Big Five Inventory-10
BFI-10-Fr: Big Five Inventory-10, French version
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
mHealth: mobile health
NEO-PI-R: Revised NEO Personality Inventory
OR: odds ratio

XSL•FO
RenderX

Edited by A Kushniruk; submitted 16.Jun.2025; peer-reviewed by A Adepoju, R Wiraguna; revised version received 05.Jan.2026; accepted 13.Jan.2026; published 22.Apr.2026.

Please cite as:
Gosetto L, Falquet G, Ehrler F
Personalizing Mobile Apps for Health Behavioral Change According to Personality: Cross-Sectional Validation of a Preference Matrix
JMIR Hum Factors 2026;13:e78939
URL: https://humanfactors.jmir.org/2026/1/e78939
doi:10.2196/78939

© Laetitia Gosetto, Gilles Falquet, Frederic Ehrler. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 22.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

On the Design of a Sign Language Corpus of Medical Terms for Automatic Translation Systems: Mixed Methods Approach

Milena Soriano Marcolino[1], MD, MSc, PhD; Luiza Marinho Motta Santa Rosa[2], MD; Elidea Lucia Almeida Bernardino[3], MSc, PhD; Mario Fernando Montenegro Campos[4], MSc, PhD

[1]Medical School and Telehealth Center, Universidade Federal de Minas Gerais, 190 Professor Alfredo Balena Ave, Belo Horizonte, Brazil

[2]Faculdade de Ciências Médicas de Minas Gerais, Belo Horizonte, Brazil

[3]Faculty of Arts & Sciences, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

[4]Department of Computer Science, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Corresponding Author:
Milena Soriano Marcolino, MD, MSc, PhD
Medical School and Telehealth Center, Universidade Federal de Minas Gerais, 190 Professor Alfredo Balena Ave, Belo Horizonte, Brazil

Abstract

Background: Hearing loss is a global health issue affecting millions and creating significant communication barriers, particularly in accessing health care services. These barriers can lead to complications and iatrogenic events, emphasizing the need for assistive technologies that enhance communication efficiency.

Objective: This study aimed to develop a corpus of medical terms for the "Captar-Libras" project, designed to improve communication between health care professionals and deaf patients through a bidirectional sign language system.

Methods: This study used the Delphi method to obtain consensus on key terms for a sign language translation system in health care emergency consultations. Initially, a questionnaire with common emergency questions was developed and distributed to health care professionals. The collected data were analyzed by a team of experts and adapted to Brazilian Sign Language (Língua Brasileira de Sinais [Libras]). Simulated clinical scenarios were then created to validate the system and ensure the accuracy of the vocabulary in the medical context.

Results: Among the 16 participants, most were physicians (n=14, 87.5%) with experience in emergency care, and half had previously treated patients with hearing loss in emergency settings. The questions evaluated received high average importance scores, particularly those related to initial symptoms and pain intensity. Some suggestions for adjustments were made, with two wording modifications significantly improving clarity regarding smoking and alcohol use. Additional suggestions to enhance the medical interview were also proposed. This study aimed to identify essential questions for emergency consultations with deaf patients, focusing on developing a corpus for Libras recognition system. The findings emphasize the importance of effective communication and highlight the challenges of translating medical terms into Libras. To address these complexities, a multidisciplinary team used the Delphi method to ensure linguistic and cultural accuracy. Additionally, the study reinforces the need for clear, structured medical queries to improve accessibility in emergency care. As a next step, system validation through simulated scenarios will be conducted. Despite certain limitations, this research lays a solid foundation for advancing sign language recognition in medical settings.

Conclusions: This study represents a significant methodological step toward improving communication between health care professionals and deaf individuals in emergency medical settings. Rather than proposing a universal solution, the study presents a structured and participatory approach for developing a corpus of medical terms in Libras, with interdisciplinary validation. The process included the involvement of deaf sign language experts during the translation and linguistic adaptation phase, ensuring that the corpus reflects authentic usage and articulation in Libras.

(JMIR Hum Factors 2026;13:e72789) doi:10.2196/72789

KEYWORDS

deafness; sign language; assistive technology; emergency health services; Delphi technique

Introduction

Hearing loss is a significant global health issue, affecting millions worldwide. According to the World Health Organization (WHO), more than 430 million people globally experience some degree of disabling hearing loss, a number projected to rise to 700 million by 2050 [1]. Approximately 10 million people in Brazil are deaf, 2.6 million of whom have

XSL•FO
RenderX

severe functional difficulties [2]. For these individuals, sign language is not just a communication tool but a vital aspect of their cultural identity, enabling interactions based on gestures and facial expressions.

Despite the critical role of sign language in the lives of deaf individuals, significant communication barriers persist, particularly in accessing health care. These barriers are a reflection of the broader challenges faced by people with disabilities, who often encounter services that are not equipped to meet their specific needs [3]. In the context of health care, these challenges can have profound implications. The lack of effective communication between health care providers and deaf patients can lead to misdiagnosis, inappropriate treatments, and overall poorer health outcomes [4].

The situation is even more challenging in emergency settings, where the inability to establish timely and effective communication can directly impact the health and safety of deaf individuals. Delays in diagnosis, treatment, and even basic information exchange can result in worse prognosis, higher incidence of complications, and iatrogenic events [4,5]. This underscores the urgent need for improved accessibility and preparedness within health care systems to address the needs of deaf individuals.

To address some of these challenges, the project "Captar-Libras: Video Communication System for the deaf applied to medical pre-care" from the Universidade Federal de Minas Gerais (UFMG) aims to overcome this obstacle by developing a prototype for the recognition and synthesis system for Brazilian Sign Language (Língua brasileira de sinais [Libras]), with the recognition of signs and generation of gestures and expressions with a photo-realistic avatar.

For successful implementation and future applicability, it was first necessary to select the most relevant terms and questions most commonly used in medical consultations. Freitas [6] defines "corpus" as a collection of text documents compiled according to a specific objective that can be processed by machines. Thus, Corpus Linguistics plays an important role in enabling the development of new assistive technologies, especially in the construction of databases for the study of languages [7].

However, in a systematic review conducted by our group, we found that no studies have been published regarding the development of such a corpus of questions and possible answers to be used in translation systems [8]. Therefore, this study aims to construct a linguistic corpus specifically tailored for emergency medical care, serving as the foundational resource for the development of real-time sign language translation technologies to improve health care accessibility for the deaf community. We hypothesize that the development of a medical corpus based on commonly used terms and questions in emergency consultations will provide the necessary linguistic foundation for effective and accurate sign language translation systems.

Methods

Study Design

This study used a qualitative, exploratory, and consensus-based Delphi method within a mixed methods framework. It constitutes a subanalysis of the "Captar-Libras Project"—a multidisciplinary initiative focused on developing a bidirectional sign language recognition and translation system for primary medical consultations. Recognizing the paucity of literature on sign language translation systems in health care, the Delphi method was selected for term selection. Initially conceptualized in the 1950s, this research strategy effectively elicits expert consensus on subjects where scientific evidence is insufficient or divergent [9].

Step 1: Initial Question Formulation

The process began by identifying the most frequent questions encountered in emergency care settings. This was based on the extensive experience of the lead researcher (MSM), who had more than 12 years of experience in emergency care. Drawing from this expertise, a preliminary set of questions was developed to reflect the most common and critical inquiries health care professionals need to communicate with patients during emergencies.

Step 2: Questionnaire Development

Using the initial set of questions devised in Step 1 as a foundation, a detailed questionnaire was created. The questionnaire included items developed through a Delphi approach to identify and reach consensus on the most relevant medical terms and questions for emergency care interactions with deaf patients, using Likert scale items. It also included open-ended questions to gather qualitative feedback. This mixed methods approach ensured that the final selection of terms and questions would meet the practical needs of a consultation at the emergency care service. Then, the questions were organized and clustered to provide the respondents with a natural flow and facilitate the data collection process, which was carefully planned to minimize the time and burden of answering the questions and to optimize communication between our team and the participants.

Step 3: Questionnaire Content

The questionnaire consisted of 6 questions regarding the professional profile, 21 Likert scale questions, and 2 open-ended questions. Besides the questions defined and established in Step 1, there were also questions designed to evaluate the participants' experience with emergency care and interacting with deaf patients (Multimedia Appendix 1). One of the questions assessed the professionals' workplace, with options: "hospital"—large health establishments with an organized flow of care, "emergency care unit"—medium-complexity services with open access whose objective is to provide initial assistance in urgencies or emergencies, and "pre-hospital care"—which involves the first care and transportation to hospitals.

Step 4: Data Collection

The questionnaire was created in Google Forms (Google LLC) for easy access and availability, and the link was sent via

XSL•FO
RenderX

WhatsApp (Meta Platforms, Inc), a widely popular instant messaging app in Brazil. Before answering, the participants received a brief overview of the "Captar-Libras Project" and an informed consent form.

Expert Selection

Experts were recruited among physicians and nurses currently working or with prior professional experience in emergency care services. This criterion was adopted because familiarity with emergency consultations is essential for identifying the most relevant questions and medical terms to be included in the corpus.

Step 5: Data Analysis

After the data collection period ended, responses were analyzed to identify consensus on the most critical terms and the need to include additional questions for the sign language recognition corpus. Consensus was defined a priori as at least 80% of participants rating an item as 4 or 5 on the 5-point Likert scale. Nevertheless, all qualitative suggestions provided in the open-ended questions were also considered. The team of linguists and Libras specialists evaluated these contributions to determine their applicability to sign language translation, ensuring linguistic and cultural adequacy. Once the final corpus was consolidated in Brazilian Portuguese, adjustments were carried out for translation and adaptation into Libras.

Quantitative data from the Likert scale questions were analyzed using descriptive statistics. Specifically, the proportion of participants selecting each response option was calculated for each question. No inferential statistical tests were conducted, as the primary aim was to identify consensus and assess the relative importance of terms for inclusion in the sign language corpus.

While the Delphi process established the selection and prioritization of questions, the response options were constructed based on the expertise of the lead author (MSM) to reflect realistic scenarios and ensure practical applicability. This approach allowed the incorporation of clinically accurate and contextually appropriate responses, which were subsequently adapted into Libras during the translation and validation process.

Step 6: Adaptation Process for Sign Language

The sign language team carefully analyzed the corpus in Portuguese for translation and cultural and linguistic adaptation. A direct word-for-word translation was not feasible due to the structural and semantic specificities of sign languages. Therefore, it was essential to make adaptations and adjustments in the corpus to ensure that the corpus reflected how deaf users articulated specific medical terms. The initial adaptation team included 2 deaf individuals, a sign language linguist (university professor), and a graduate student with health-related experience, and 2 hearing professionals (a sign language linguist fluent in Libras and a sign language interpreter). The team worked collaboratively to identify terminological discrepancies between spoken Portuguese and Libras, considering both linguistic and contextual aspects.

The adaptation process followed an iterative cycle:

1. Initial translation and restructuring: the team identified questions that were too complex for direct translation and restructured them into shorter, sequential items (eg, the cluster "If the complaint is pain: Where is the pain? What is the pain like? Does it feel like a tightness or weight, or burning, or stabbing, or a shock?" was expanded into multiple questions regarding location, type, and sensation). Some concepts required additional explanation or modification; for example, (1) "Pain in the middle of the belly" was clarified by medical staff for accurate articulation. (2) The term "trauma" lacks a direct sign, so translations were elaborated to convey "physical trauma" or "psychological trauma" depending on context. (3) "Penetrating trauma" was translated into a descriptive sign sequence ("material group cut also long thin object stick") to ensure comprehension. (4) Aggression was translated using the most comprehensive and contextually appropriate signs, reflecting various scenarios of physical aggression. Decisions were guided by the principle that the translated signs should reflect the intended meaning, be understandable by deaf informants, and align with authentic Libras usage. This approach is consistent with previous work highlighting that effective communication in health care requires corpora that are functionally and contextually realistic, thereby improving the accuracy and usability of sign language recognition systems [10].
2. Peer validation: each item was translated to Libras by a deaf undergraduate student and recorded as individual videos. To validate comprehension, the videos were presented to a small group of deaf graduates and to a deaf university professor who had not seen the questionnaire in Portuguese. Their feedback focused on clarity, naturalness, and cultural adequacy.
3. Revisions: terms or expressions flagged as unclear were reformulated. For example, grammatical suggestions that had no functional equivalent in Libras were discarded, while ambiguous medical terms were replaced with more visually descriptive alternatives.
4. Final validation: this cycle of translation, feedback, and revision was repeated until mutual comprehension was reached across both deaf linguists and health professionals.

During the project, team composition evolved (the deaf professor left for personal reasons and was replaced by another deaf linguistics professor, while new members, including a physiotherapist and 2 undergraduate students, joined), but the iterative approach remained constant. Throughout the adaptation process, all feedback from deaf linguists, health professionals, and volunteer reviewers was carefully considered. Suggestions that enhanced clarity, comprehension, or cultural accuracy were incorporated directly into the corpus, while minor issues irrelevant to sign language translation (eg, Portuguese punctuation) were documented but not applied. This iterative approach ensured that each input led to specific refinements, resulting in a final corpus that accurately reflected the intended medical terminology in Libras and was understandable by the target population. Overall, the adaptation phase spanned approximately 3 months, encompassing reevaluation, translation, and video recording, culminating in the production of a definitive video corpus. The final version served as a

XSL•FO
RenderX

"teleprompter" guide for deaf volunteers, who reproduced the validated signs in a controlled setup composed of multiple cameras, chromakey background, and illumination. The data collection started in January 2024 and was finalized in May 2025.

Step 7: Development of Hypothetical Medical Care Situations

Hypothetical clinical scenarios were developed by the aforementioned researcher (MSM), who has clinical expertise in emergency care, based on common situations encountered in emergency settings. Each scenario was designed to simulate realistic interactions between health care professionals and deaf patients, including typical questions about symptoms, pain, and treatment history. This step was fundamental to ensure adequate understanding by the systems team, enabling the translation into sign language process to be carried out without losing critical information or meaning.

The evaluation of the sign language recognition system, which is being developed, will focus on three aspects:

1. Comprehension by the target users: deaf volunteers will assess whether the terms and questions are clearly understood and reflect realistic emergency situations.
2. Accuracy of translation: the system's ability to correctly interpret and reproduce the intended medical terms in Libras will be checked against the original Portuguese questionnaire.
3. User experience and system performance: participants will provide feedback on ease of use, clarity of interaction, and any technical issues such as system crashes or delays, which will inform iterative improvements to the interface and functionality.

Feedback from participants will be systematically recorded, and necessary adjustments will be made iteratively to ensure the clarity, accuracy, and usability of the system in emergency care contexts.

Ethical Considerations

This study was approved by the UFMG Ethics Review Board, under protocol number CAAE 71623323.8.0000.5149, in accordance with national and international research ethics guidelines. All participants, including health care professionals and sign language experts, provided informed consent prior to participation in the Delphi process, translation, and validation activities. In addition to obtaining informed consent, all procedures were designed to respect cultural and linguistic nuances within the deaf community. The participants' autonomy, comfort, and comprehension were prioritized during data collection. Data collected through questionnaires were anonymized to ensure confidentiality. Video recordings were securely stored, accessible only to the research team, and were used exclusively for the purposes of corpus development. Participants and the team members who appeared in video recordings of sign language production signed specific consent forms authorizing the use of their images for research purposes and for inclusion as multimedia material accompanying this publication. No other personally identifiable information is disclosed in this paper or in the supplementary material. Participants who were deaf and took part in the video recordings received financial compensation to cover transportation costs to the UFMG, where the recordings took place.

Results

Profession and Field of Activity

Among the 16 participants who answered the questionnaire, 12/16 (75%) were currently working in emergency and urgent care services, which aligns with the focus of this study. Among these 12 participants, 11/12 (91.67%) were physicians and 1/12 (8.33%) was a nurse. Considering all participants (N=16), 3/16 (18.75%) participants had worked in the area for less than 5 years, 4/16 (25%) for 5 to 10 years, and 5/16 (31.25%) for more than 10 years. Regarding the workplace, 7/16 (43.75%) participants reported working in hospitals, 2/16 (12.5%) in emergency care units, 1/16 (6.25%) in prehospital care, and 2/16 (12.5%) worked simultaneously in hospitals and emergency care units.

Experience in Treating a Hearing-Impaired Patient in an Emergency

Among the 12 participants working in emergency and urgent care services, 8 (66.66%) had already treated hearing-impaired patients in emergencies. Regarding the frequency of these occurrences, 2/12 (16.67%) professionals had treated 2 patients in this scenario, 4/12 (33.33%) had treated between 2 and 5, and 2/12 (16.67%) had treated more than 5.

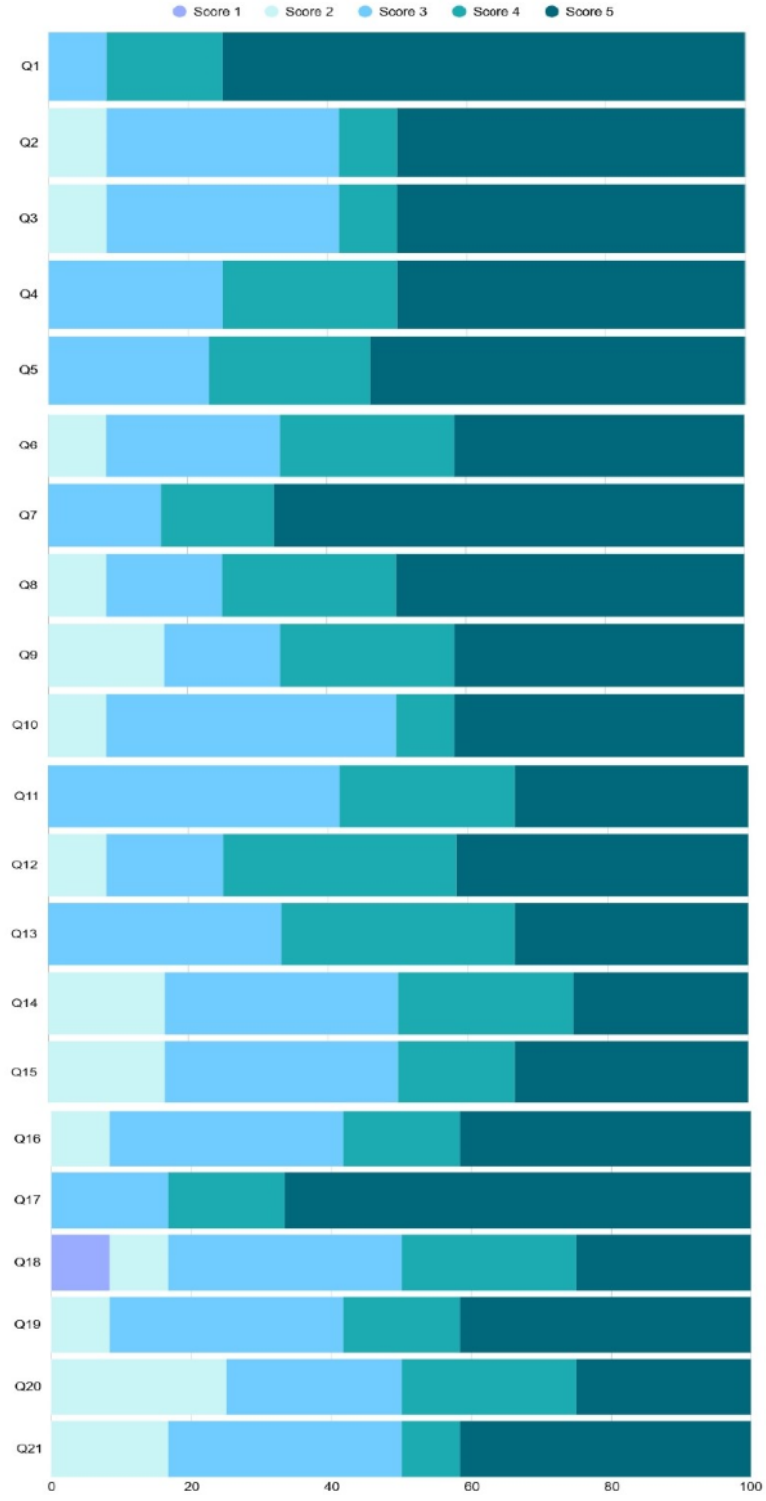
Level of Relevance of the Proposed Questions

Figure 1 and Multimedia Appendix 2 summarize the general results. Most questions were considered important, with medians predominantly between 4 and 5. Notably, questions addressing the primary complaint (eg, "What are you feeling that made you seek the hospital?") and the patient's medical history (eg, "Are you allergic to any medication?") were consistently rated as highly relevant.

A Likert scale ranging from 1 to 5 was used, with the following correspondence: 1=not important at all; 2=of little importance; 3=of average importance; 4=very important; 5=absolutely essential.

XSL•FO
RenderX

Figure 1. Graphical representation of the importance score assigned to each question by health professionals with emergency care experience.



Opinion on the Wording of the Questions and Contact Availability

Six participants suggested 10 recommendations for improvement of the already proposed questions (Textbox 1). Most suggestions involved minor grammatical adjustments, which were deemed nonrelevant for sign language translation by the team of linguistics specialists. Questions regarding smoking and alcohol consumption were revised to better capture the patients' habits, reflecting these contributions. The other 6 participants indicated no need for changes.

XSL•FO
RenderX

Textbox 1. Questions included in the final version for the corpus development.

1. If the complaint is pain: was it 1 episode of pain, or more episodes? If more episodes, when was the first episode? How did the pain evolve? Did it increase in intensity? Did it increase in duration?
2. Are you feeling nauseous or vomiting? Do you have diarrhea? If so, what is it like? Is there blood, mucus, or pus? Do you have stomach pain?
3. Have you had any surgery? If so, which one?

General Suggestions and Proposals for New Questions

Five participants provided other general suggestions or proposed new questions:

1. Incorporating the Manchester Protocol for prioritizing patient care
2. Organizing the assessment in a flowchart format, based on the main complaint and the Manchester protocol
3. Inserting a question regarding the duration of the pain and its evolution (whether it has improved or worsened over time)
4. Inserting a question regarding other known allergies and surgical history
5. Inserting questions regarding gastrointestinal complaints (“Are you feeling nauseous or vomiting? Do you have diarrhea? Do you have stomach pain?”)

Although the inclusion of the Manchester Protocol and organizing in a flowchart assessment would be enriching, these suggestions exceeded the scope of this study. In the case of the Manchester Protocol, information about waiting time depending on the severity of the patient’s pain is relevant for health care professionals. For patients, this is not a piece of information that needs to be translated into sign language. Regarding the other suggestions to insert new questions, the changes were made (Table 1).

XSL•FO
RenderX

Table . Initial questions with their respective modifications and sign language videos.

Proposed question	Version after Delphi method	Final version after team correction
1. What are you feeling that made you seek the hospital?	No suggestions made	What are you feeling that made you seek the hospital?
2. When did it start?	No suggestions made	When did it start?
3. How did it start? What were you doing when it started?	How did it start? What were you doing the moment it started?	How did it start? What were you doing when it started?
4. How long has it lasted?	No suggestions made	How long has it lasted?
5. Do you feel anything else, besides this symptom?	No suggestions made	Do you feel anything else, besides this symptom?
6. Is there any factor that makes it worse, or better?	Is there any factor that makes it worse or better?	Is there any factor that makes it worse or better?
7. If the complaint is chest pain or shortness of breath: does it get worse when you make a physical effort? Climbing a hill, walking on a flat surface, or taking a shower? Or does it appear when you are at rest, without any effort? Has it improved with any medication?	No suggestions made	If the complaint is chest pain or shortness of breath: does it get worse when you make a physical effort? Climbing a hill, walking on a flat surface, or taking a shower? Or does it appear when you are at rest, without any effort? Has it improved with any medication?
8. If the complaint is pain: Where is the pain? What is the pain like? Does it feel like a tightness or weight, or burning, or stabbing, or a shock?	No suggestions made	If the complaint is pain: Where is the pain? What is the pain like? Does it feel like a tightness or weight, or burning, or stabbing, or a shock?
9. On a scale of 1 to 10, with 1 being very weak pain and 10 being unbearable pain, what is the intensity of the pain?	No suggestions made	On a scale of 1 to 10, with 1 being very weak pain and 10 being unbearable pain, what is the intensity of the pain?
10. Does it stay in one fixed place, or does it spread to another location?	Does it stay in one fixed place or does it spread to another location?	Does it stay in one fixed place, or does it spread to another location?
11. Does it get worse when you press? Does it get worse when you change your body position?	Does it get worse when you press or compress it? Does it get worse when you change your body position (or move your body)?	Does it get worse when you press? Does it get worse when you change your body position?
12. If you complain of fever or chills, do you have sneezing or nasal discharge?	No suggestions made	If you complain of fever or chills, do you have sneezing or nasal discharge?
13. Do you have a cough? If so, what is the cough like, productive or dry? What color is the phlegm?	Do you have a cough? If so, what is the cough like, productive or dry? What is the phlegm's color?	Do you have a cough? If so, what is the cough like, productive or dry? What color is the phlegm?
14. If you don't sneeze or cough, do you feel a burning sensation when urinating?	If you don't sneeze or cough, does it burn when urinating?	If you don't sneeze or cough, do you feel a burning sensation when urinating?
15. If you complained of fever, chills, cough, shortness of breath, or if you said you had sneezing or nasal discharge: have you been vaccinated against COVID-19? If so, how many doses? Have you been vaccinated against the flu?	If you complained of fever, chills, cough, or shortness of breath, or if you said you had sneezing or nasal discharge: were you vaccinated against COVID-19? If so, how many doses? Have you been vaccinated against the flu?	If you complained of fever, chills, cough, shortness of breath, or if you said you had sneezing or nasal discharge: have you been vaccinated against COVID-19? If so, how many doses? Have you been vaccinated against the flu? Did you bring your vaccination card?
16. Are you being treated for any disease? If so, which one?	No suggestions made	Are you being treated for any disease? If so, which one?
17. Are you allergic to any medication? If so, which one?	No suggestions made	Do you have any allergies? Are you allergic to any medication? If so, which one?
18. Did you bring a medical report?	Did you bring a medical document?	Did you bring a medical report?
19. Did you bring a prescription? If not, are you taking any medication?	No suggestions made	Did you bring a prescription? If not, are you taking any medication?
20. Do you smoke?	Do you smoke or have you ever smoked? If so, for how long and how many cigarettes per day?	Do you smoke? If so, for how long and how many cigarettes per day?
21. Do you drink alcohol? If so, how much?	Do you drink alcohol? If so, which drinks and how much?	Do you drink alcohol? If so, which drinks and how much?

Besides the listed changes, some newly proposed questions were included and are shown in Textbox 1.

The final version of the questions and their respective videos in sign language (Libras) is shown in Multimedia Appendix 3.

XSL•FO
RenderX

Hypothetical Medical Care Situations

Fourteen hypothetical scenarios were developed as a corpus enhancement tool in order to test its applicability and comprehensiveness (Multimedia Appendix 4). The medical conditions were selected according to their prevalence in emergency services, in a way that the main diagnosis involved could be broadly represented. Therefore, the scenarios included common infections, inflammatory conditions such as acute abdomen, migraine, respiratory decompensation, cardiovascular, and thromboembolic events. In each of them, a clinical description of the patient's probable complaints and expected physical examination data was made. This stage was essential for the corpus's validation process.

Discussion

Main Findings

This study demonstrated the feasibility and methodological value of using a Delphi technique to develop a clinically relevant corpus of medical terms. In this analysis, the process focused on emergency care and involved health care professionals to identify and refine expressions considered essential for patient-provider communication. The subsequent linguistic adaptation phase engaged deaf and hearing sign language experts to ensure that the selected content was accurate, natural, and culturally appropriate in Libras. The final corpus includes questions and expressions organized to reflect the logical sequence of emergency medical anamnesis, thus bridging clinical communication needs with the linguistic and cultural norms of the deaf community. These findings highlight the potential of interdisciplinary, consensus-based approaches to improve accessibility and guide the development of sign language resources and technologies in health care. They also underscore the complexities involved in developing a bidirectional translation system that takes into account the linguistic and cultural specificities of the deaf community.

The need for this study arose from a larger project, whose objective is to develop a prototype for recognizing Brazilian sign language in the context of medical settings [11]. This area is still largely underexplored, with a growing demand for development based on emerging technological innovations and the use of novel artificial intelligence models.

Comparison With Prior Work

While the development of translation systems for sign languages has been explored in prior research, and some studies have described the creation of sign language corpora for health care–related applications, none have used a structured consensus method such as the Delphi technique. For example, Camgöz et al [12] developed a Turkish Sign Language Recognition Corpus that included content from health, finance, and daily life domains. Although the authors consulted linguists, members of the deaf community, and domain specialists, they did not report how the final corpus content was selected. Similarly, Wille et al [13] described the development of a Flemish Sign Language corpus based on descriptive grammatical research, including elicited and spontaneous narratives, conversations, and on-topic interviews. However, these corpora were primarily linguistically oriented, with no focus on health care, and did not involve systematic validation by health care professionals. Roelofsen et al [14] reported collecting a set of phrases commonly used during the diagnosis and treatment of COVID-19, based on consultations with health care professionals at the Amsterdam University Medical Center as well as on direct experience (one of the authors is a medical doctor). They also referenced phrases from the SignTranslate system. However, the authors were not transparent about how the most important sentences were selected or ranked for inclusion in the corpus [14].

In a previous systematic review of sign language translation systems, the corpus development process was not even mentioned [15]. Our recent systematic review identified 23 studies involving sign language recognition systems developed and tested to be applied in health care settings [8]. These studies often emphasize technical aspects, such as gesture recognition, but provide little detail on how the linguistic content, particularly in specialized domains such as health care, is selected or validated. The vast majority were in the development or testing stage and used corpora based on isolated terms, often defined by researchers or adapted from dictionaries, with minimal contextual adaptation. Very few studies involved health care professionals or sign language experts in corpus development, and none used structured consensus methods such as the Delphi technique. Furthermore, most systems lacked integration of key communicative features such as facial expressions and bidirectionality, both essential for health care communication. To our knowledge, no prior study has detailed a participatory, interdisciplinary approach to constructing a medically oriented sign language corpus using both clinical and linguistic validation. Our study addresses this gap, contributing a novel methodology tailored to emergency health care settings in Libras.

The lack of literature on this specific process led us to create approaches that seemed to be most adequate. After much investigation and planning, the Delphi method was selected [9], since it could assist in directing and supporting the process of choosing terms, thus ensuring its reproducibility. With the collection and analysis of the responses, the process of adaptation to sign language was carried out in collaboration with the linguistics team.

Relevance of Proposed Questions

The high relevance ratings for most of the proposed questions (with medians predominantly between 4 and 5 on the Likert scale) reflect the importance of accurate, detailed, and efficient communication in emergency care. Specifically, questions related to the patient's primary complaint (eg, "What are you feeling that made you seek the hospital?") and medical history (eg, "Are you allergic to any medication?") were consistently rated as critical. This aligns with the standard practices in emergency medical care, where understanding the nature and timeline of symptoms is essential for accurate diagnosis and appropriate treatment [16].

The consistency in the importance assigned to these questions indicates that health care professionals value information that directly impacts clinical decision-making. This also reinforces the necessity for such questions to be clearly and accurately

XSL•FO
RenderX

translatable into sign language, ensuring that deaf patients can provide this vital information without delay or confusion.

Challenges in Sign Language Translation

One of the main challenges identified in this study was the translation of medical terms and questions into Libras. As sign languages are linguistically distinct from spoken languages, direct translation of phrases often leads to ambiguities or misinterpretations [17]. The need for careful adaptation, particularly for complex medical queries (eg, questions regarding the nature of pain or its progression), was emphasized by the research team. In many cases, questions had to be broken down into simpler, shorter queries, a process that aligns with previous findings in the literature regarding the importance of adapting questions for clarity in sign language. This process, while time-consuming, is crucial to ensure that the information exchanged during an emergency consultation is accurate and comprehensive.

When translating into Libras, 2 questions were more challenging, for different reasons. The first of them, question 9, "On a scale of 1 to 10, with 1 being very weak pain and 10 being unbearable pain, what is the intensity of the pain?" It has a more complex grammatical structure, and it was not possible to break it into 2 or more distinct questions, as we did with question 8, for example. As a translation strategy, we transformed the question into a comment topic. In other words, we start the sentence with the topic "pain," followed by the question-comment "on a scale from 1 to 10, with intensity 1 being a weak pain, intensity 5 being a medium pain, and 10 being a very strong pain, what is the intensity of the pain?" The addition of "average pain" gave the question a better idea of the intended gradation.

In the second question 10, "Does it stay in one fixed place, or does it spread to another location?," the challenge was to define a pain location without interfering with the original question. This is because "[to] stay in one fixed place" makes no sense in Libras if it is not defined in the signer's body what that place is. The translation choice of the deaf woman who recorded the question was "pain, for example [points to a place on her own body], stays in the same place or spreads [performs the sign [to spread] starting from the chosen location on her body for the pain and spreading to adjacent places]."

These translation strategies were discussed with the translation team, tested, and evaluated until we reached a consensus on what would be the best way to translate them without interfering with the desired meaning, but taking into account how the deaf person would clearly understand.

Additionally, it became evident that translation is not merely a matter of substituting words; it involves a deep understanding of both the linguistic structure of sign language and the cultural context in which it is used [17].

The involvement of a multidisciplinary team, including deaf linguists, health professionals, and sign language interpreters, was essential to ensure the validity and appropriateness of the translations. The medical professionals on the team were essential in the initial process of suggesting sentences, which, after evaluation by the group of health professional experts, were submitted to the linguistics team for careful evaluation. This framework ensured that all necessary adjustments were made to guarantee that the final product was scientifically and grammatically correct, both in the written and signed versions.

Surely much can be learned from the advances in natural language processing; however, there are many new dimensions yet to be explored when it comes to sign language understanding and synthesis. If, on the one hand, spoken languages embed significant information such as rhythm, tone, and pauses, on the other hand, it does not require delving into finer details such as facial expressions, as is the case with sign languages. When describing the creation of a text-to-sign translation method, Bertin-Lemée et al [18] mention the need to preserve the linguistic structures of sign language during the translation process and point it out as an inherent challenge to the process.

This stage involved making the necessary adjustments to the sentences to ensure their correct meaning in both written and signed forms. All suggestions were carefully analyzed, with several related to the positioning of commas and grammatical structure, without significantly interfering with the message conveyed or the intent of the questions. It is important to note that those unfamiliar with sign language may not realize that punctuation, such as commas, does not hold the same relevance in this context. The structure of sign language differs significantly from spoken languages, and such differences are often overlooked by society at large [17]. Only 2 suggestions were deemed significant, leading to adjustments in questions 20 and 21. This highlights the importance of having knowledgeable researchers with extensive technical expertise to ensure a proper critical analysis and adaptation of the content.

Implications for Practice

This project has significant implications for improving health care accessibility for deaf individuals, especially in emergency care settings. By developing a system that enables real-time communication in Libras, the "Captar-Libras" Project aims to reduce the critical communication barriers that deaf patients face when seeking urgent medical care. It also contributes to the broader movement toward greater inclusion and accessibility in health care systems.

One of the key outcomes of this study is the development of a clear, structured corpus of terms that can be directly applied not only in the sign language recognition system developed by our group but also in systems created by other research teams. By contributing to the existing body of literature, this work aims to contribute to the development of future systems for medical communication in sign language. The next steps will focus on validating the system through hypothetical medical care scenarios, ensuring that the signs synthesized by the system are accurate and comprehensible for deaf users.

Limitations

Similar to spoken languages, sign languages are inherently regional and culturally specific. Our work focused on Brazilian Sign Language (Libras), which limits the generalizability of the corpus to sign languages of other countries. While this limitation is intrinsic to any language-based system, future adaptations will require cultural and linguistic tailoring to adapt it to other

XSL•FO
RenderX

contexts, thus ensuring broader applicability. However, the primary contribution of this study lies in the methodological approach: a structured, interdisciplinary process for corpus development tailored to health care contexts, rather than in proposing an unabridged compendium of signs.

Second, the number of health care professionals participating in the Delphi process was relatively small. Although their responses were consistent and included a range of clinical experiences, a larger and more diverse sample, including professionals from different regions, more sign language users, and a broader range of professional roles and experiences, could further strengthen the representativeness and validity of the selected terms. Future iterations should aim to broaden participation, possibly involving deaf health care users as well.

Third, the list of terms and questions was based on hypothetical medical emergencies and predefined scenarios. While this approach provided structure and facilitated consensus, it may not fully capture the emotional complexity and spontaneity of real-life interactions in emergency care. Future studies should validate the corpus using real-world consultations or simulations, allowing for refinement based on authentic communication dynamics.

Despite these limitations, we believe that this study proposes a solid foundation for future corpus development efforts. The use of a transparent, interdisciplinary, and consensus-based approach represents a significant methodological advancement in the field of sign language translation systems for health care.

Potential Scalability and Future Directions

Although this study focused on developing a Brazilian Sign Language (Libras) corpus for emergency care, the methodology and participatory approach used may provide a model for adaptation to other sign languages and cultural contexts. Key elements such as expert consensus on terminology, iterative translation with linguistic validation, and the incorporation of hypothetical clinical scenarios could be applied in different health care settings, allowing for the creation of contextually and culturally appropriate corpora.

Building on this initial effort, future research should focus on validating the proposed corpus in real or simulated health care interactions to assess its communicative adequacy in practice. Further expansions may include broadening the development of corpora for other care contexts, such as primary care or mental health. Additional Delphi rounds involving deaf patients themselves could enrich the corpus with user-centered insights. Moreover, adapting the proposed methodology to other sign languages and national health systems may support the development of culturally relevant corpora worldwide. Finally, the integration of this corpus into translation tools and health professional training programs could enhance accessibility and equity in health care communication.

Conclusions

This study represents a significant methodological step toward improving communication between health care professionals and deaf individuals in emergency medical settings. Rather than proposing a universal solution, the study presents a structured and participatory approach for developing a corpus of medical terms in Libras, with interdisciplinary validation. The process included the involvement of deaf sign language experts during the translation and linguistic adaptation phase, ensuring that the corpus reflects authentic usage and articulation in Libras.

The findings underscore the complexity of translating medical terminology into sign language and highlight the importance of integrating both health care and linguistic expertise. While this corpus provides a solid foundation, future validation in real-world or simulated scenarios remains essential to assess its practical utility.

Acknowledgments

The authors thank the researchers Raniere Cordeiro, Raiane Barros, Felipe Barros, Isabel Franco de Carvalho, Natália Afonso, and Michelle Murta, who are deaf Libras signers, and who helped in the corpus discussion and adaptation.

Funding

This study was funded in part by Financiadora de Estudos e Projetos (FINEP, grant 01.21.0097.00) and SignumWeb Comunicação Inclusiva Ltda, as part of the Project “Captar-Libras: Sistema de Comunicação por vídeos para surdos aplicado ao pré-atendimento médico." MSM was supported in part by the Institute for Health Assessment and Translation for Chronic and Neglected Diseases of High Relevance (Instituto para Avaliação e Translação em Saúde para Doenças Crônicas e Negligenciadas de Alta RElevância [IATS CARE])/National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq; grant number 408659/2024-6) and CNPq (Bolsa de Produtividade em Pesquisa 311742/2025-4). MFMC was supported in part by CNPq (Bolsa de Produtividade em Pesquisa 316841/2021-8). The funding sources had no role in the design of this study and will not have any role during its execution, analyses, interpretation of the data, or decision.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

MSM developed the project methodology and selected the most common questions in emergencies. MFMC provided technical supervision of the video collection and storage process. LMMSR developed hypothetical medical scenarios with MSM's supervision. ELAB guided the sign language adaptation process. All authors drafted the manuscript and approved its final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questions included in the questionnaire.
[DOCX File, 15 KB - humanfactors_v13i1e72789_app1.docx]

Multimedia Appendix 2
Importance scores assigned by participants with previous emergency services experience to each question proposed for deaf individuals' health care.
[DOCX File, 18 KB - humanfactors_v13i1e72789_app2.docx]

Multimedia Appendix 3
Final version of the questions and their respective videos in sign language.
[DOCX File, 19 KB - humanfactors_v13i1e72789_app3.docx]

Multimedia Appendix 4
Hypothetical medical care situations.
[DOCX File, 18 KB - humanfactors_v13i1e72789_app4.docx]

References

1. Deafness and hearing loss. World Health Organization. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss [accessed 2026-03-18]
2. IBGE census. Brazilian Institute of Geography and Statistics. 2022. URL: https://censo2022.ibge.gov.br [accessed 2026-03-18]
3. Middleton A, Niruban A, Girling G, Myint PK. Communicating in a healthcare setting with people who have hearing loss. BMJ 2010 Sep 29;341(3):c4672. [doi: 10.1136/bmj.c4672] [Medline: 20880905]
4. Ruiz-Arias N, Barría-Pailaquilén RM. Experiences of hospital health professionals in the care of deaf patients. Rev Salud Publica (Bogota) 2023 Dec;25(6):109176. [doi: 10.15446/rsap.V25n6.109176] [Medline: 40099108]
5. Emond A, Ridd M, Sutherland H, Allsop L, Alexander A, Kyle J. The current health of the signing Deaf community in the UK compared with the general population: a cross-sectional study. BMJ Open 2015 Jan 25;5(1):e006668. [doi: 10.1136/bmjopen-2014-006668] [Medline: 25619200]
6. Freitas C. Corpus, computational linguistics and digital humanities. In: Leite M, Gabriel CT, Leite M, Gabriel CT, editors. Linguagem, Discurso, Pesquisa & Educação: DP&A Editora; 2015:18-46.
7. Sardinha TB. Corpus linguistics: history and issues. Delta (São Paulo) 2000;16(2):323-367. [doi: 10.1590/S0102-44502000000200005]
8. Marcolino MS, Oliveira LD, Valle LR, et al. Innovations in deaf health care communication: systematic review of sign language recognition systems. J Med Internet Res 2026 Apr 9;28:e70417. [doi: 10.2196/70417] [Medline: 41955542]
9. Hsu C, Sandford BA. The Delphi technique: making sense of consensus. Pract Assess Res Eval 2007;12(1):10. [doi: 10.7275/pdz9-th90]
10. Rivas Velarde M, Izquierdo Martinez LC, Dalal J, et al. Video remote sign language interpreting in health communication for deaf people: protocol for a randomized controlled trial. JMIR Res Protoc 2024 Dec 2;13:e64590. [doi: 10.2196/64590] [Medline: 39622021]
11. Marcolino MS, Oliveira LFR, Valle LR, et al. Sign language recognition system for deaf patients: protocol for a systematic review. JMIR Res Protoc 2025 Jan 23;14:e55427. [doi: 10.2196/55427] [Medline: 39847417]
12. Camgöz NC, Kındıroğlu AA, Karabüklü S, Kelepir M, Özsoy AS, Akarun L. BosphorusSign: a Turkish sign language recognition corpus in health and finance domains. Presented at: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16); May 23-28, 2016; Portorož, Slovenia p. 1383-1388 URL: https://aclanthology.org/L16-1220.pdf [accessed 2026-03-19]
13. Wille B, Beukeleers I, Van Herreweghe M, Vermeerbergen M. Big things often have small beginnings: a review on the development, use and value of small and big corpora for Flemish sign language linguistic research. Front Psychol 2021;12:779479. [doi: 10.3389/fpsyg.2021.779479] [Medline: 35069360]

XSL•FO
RenderX

14. Roelofsen F, Esselink L, Mende-Gillings S, et al. Sign language translation in a healthcare setting. Presented at: Proceedings of the Translation and Interpreting Technology ONline (TRITON) Conference; Jul 5-7, 2021; Shoumen, Bulgaria p. 110-124. [doi: 10.26615/978-954-452-071-7_013]
15. Boggaram A, Boggaram A, Sharma A, Ramanujan AS, Bharathi R. Sign language translation systems: a systematic literature review. Int J Softw Sci Comput Intell 2022 Jan 1;14:1-33. [doi: 10.4018/IJSSCI.311448]
16. Schell CO, Khalid K, Wharton-Smith A, et al. Essential emergency and critical care: a consensus among global clinical experts. BMJ Glob Health 2021 Sep;6(9):e006585. [doi: 10.1136/bmjgh-2021-006585] [Medline: 34548380]
17. Rodrigues CH. The interpretation for brazilian sign language: modality effects and inferential processes [doctorate]. : Faculty of Arts and Sciences, Universidade Federal de Minas Gerais; 2013 URL: https://repositorio.ufmg.br/server/api/core/bitstreams/92ee9d46-dc41-41f0-9e25-76756afe6cee/content [accessed 2026-04-19]
18. Bertin-Lemée É, Braffort A, Challant C, Danet C, Filhol M. Example-based machine translation from text to a hierarchical representation of sign language. Presented at: Proceedings of the 24th Annual Conference of the European Association for Machine Translation; Jun 12-15, 2023 URL: https://aclanthology.org/2023.eamt-1.3/ [accessed 2026-03-19]

Abbreviations

Libras: Língua Brasileira de Sinais
UFMG: Universidade Federal de Minas Gerais
WHO: World Health Organization

Edited by A Kushniruk; submitted 19.Feb.2025; peer-reviewed by P Escudeiro, R Tripathi; revised version received 09.Sep.2025; accepted 16.Oct.2025; published 29.Apr.2026.

Please cite as:
Marcolino MS, Rosa LMMS, Bernardino ELA, Campos MFM
On the Design of a Sign Language Corpus of Medical Terms for Automatic Translation Systems: Mixed Methods Approach
JMIR Hum Factors 2026;13:e72789
URL: https://humanfactors.jmir.org/2026/1/e72789
doi:10.2196/72789

© Milena Soriano Marcolino, Luiza Marinho Motta Santa Rosa, Elidea Lucia Almeida Bernardino, Mario Fernando Montenegro Campos. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 29.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Development, Implementation, and Preliminary Results of a Novel Social Network for Rare Disease Communities: Mixed Methods Study

Lara Fendrich[1,2], BSc, MD; Henriette Högl[3], MD; Ralf Schmidt[4]; David Bascom[5], MDes; Armin Berger[6], BSc, MSc, PhD; Max Lübbering[6], BSc, MSc, PhD; Gundula Ernst[7], PD, PhD; Justus Schumacher[1]; Sabrina Hölzer[8]; Lorenz Grigull[1], MBA, MME, MD, Prof Dr Med

[1]Center for Rare Diseases, University Hospital Bonn, Venusberg-Campus 1, Geb. 13 (BMZ), Bonn, Germany

[2]Pediatric Oncology, Hematology and Immunology, Olgahospital, Klinikum Stuttgart, Stuttgart, Germany

[3]Kindernetzwerk e.V, Aschaffenburg, Germany

[4]digitalUX, Uplengen, Germany

[5]99 Grad, Wiesbaden, Germany

[6]Fraunhofer Institute for Intelligent Analysis and Information Systems, Bonn, Germany

[7]Hannover Medical School, Hanover, Germany

[8]Ernst Abbe University of Applied Sciences Jena, Jena, Germany

Corresponding Author:
Lorenz Grigull, MBA, MME, MD, Prof Dr Med
Center for Rare Diseases, University Hospital Bonn, Venusberg-Campus 1, Geb. 13 (BMZ), Bonn, Germany

Abstract

Background: Rare diseases affect approximately 20 million Europeans, presenting unique challenges such as delayed diagnoses, limited therapies, and significant personal and financial burden. While resilience-supporting factors such as peer support are available and artificial intelligence–based diagnostic tools are being developed further, there is a lack of a dedicated online social network connecting patients, caregivers, relatives, and experts. This study presents the development and preliminary findings of Unrare.me, a novel social network designed to provide a secure space for experts and individuals affected by rare and chronic diseases (diagnosed and undiagnosed).

Objective: This study aimed to design, develop, and evaluate a social networking platform tailored to the needs of different stakeholders of the rare disease community, facilitating interaction, knowledge exchange, and emotional support while prioritizing data security.

Methods: This multidisciplinary, multicenter initiative brought together patient groups, health care professionals, psychologists, and web design experts. A literature review assessed existing networking approaches in the rare disease community. Structured interviews and user journey mapping defined user needs and essential app features. Iterative prototyping and stakeholder discussions informed the final design, which was developed into a functional app launched in December 2023 on major platforms. A survey conducted four months post-launch evaluated user feedback. Data security was prioritized throughout development.

Results: A total of 270 users (approximately 1 in 7 users at the time) participated in the evaluation. Most of them (n=221, 81.9%) registered to connect with others in similar situations, whereas 56.7% (n=153) sought expert input and 44.4% (n=120) looked for disease-related information. The app received positive ratings for usability (mean 6.12, SD 1.03; out of 7), accessibility (mean 5.59, SD 1.22), and design (mean 5.84, SD 1.12), as well as overall impression (mean grade of 2.24, SD 0.90 on a scale from 1-6, with 1 being the best score). Data security was highly rated (mean 5.58, SD 1.15). The app's ontology was suitable for 77% (n=208) of the participants, enabling them to find their diagnosis, and 60.7% (n=164) of users found at least one match. Matching preferences centered on shared diagnosis (mean 82.5, SD 25.1 on a visual analog scale from 0 to 100), symptoms (mean 74.2, SD 25.8), and everyday experiences (mean 69.6, SD 29.5). Overall, users welcomed the opportunity to network with each other securely and highlighted areas for further improvement, such as enhanced matching features and group chat options.

Conclusions: Unrare.me has generated significant interest and engagement within the German rare disease community, serving as a valuable tool for peer support, knowledge sharing, and expert identification. Current challenges include optimizing user acquisition and refining matching algorithms. Planned features include group chats, expert interaction, and gamification elements. Unrare.me illustrates the potential of tailored digital solutions to address unmet needs in the rare disease community.

(JMIR Hum Factors 2026;13:e80230) doi:10.2196/80230

XSL•FO
RenderX

KEYWORDS

rare disease; social network; mutual support; app; artificial intelligence; AI

Introduction

Rare and chronic diseases pose unique challenges for those affected in terms of both health care and social well-being. Rare diseases are defined as conditions affecting fewer than 1 in 2000 individuals; worldwide, over 300 million people live with one of the 6000 to 8000 known rare diseases. In the European Union alone, these diseases impact approximately 20 million individuals along with their caregivers [1].

One of the first challenges for people with rare diseases is receiving an accurate diagnosis. Due to the low prevalence of these conditions, health care professionals often lack the resources to recognize and diagnose them efficiently. On average, a time frame of 5 years from the first symptom to diagnosis is accepted as typical for rare diseases [2]. During this period, patients may undergo misdiagnoses, unnecessary treatments, and worsening symptoms [3]. This diagnostic delay increases the emotional and psychological burden on patients and their families and caregivers, leading to higher rates of anxiety and depression, feelings of isolation, and helplessness. This is particularly pronounced in children, placing even greater demands on caregivers to balance emotional and practical responsibilities [4]. Compounding these emotional challenges is the financial strain on both individuals and health care systems as the complexity of diagnosis and treatment leads to much higher health care costs than those for common diseases. This burden is exacerbated by the excessive cost of orphan drugs and indirect costs such as lost productivity due to illness or caregiving responsibilities [5].

Beyond those challenges, people with rare diseases and their caregivers also struggle with a lack of access to information and support and connecting with others who share similar experiences [6]. Traditionally, health care systems are ill-equipped to provide specialized knowledge or resources, and patients frequently report difficulties in finding experts or communities who understand their condition. The scarcity of research and information contributes to this challenge, leaving patients and caregivers to rely on fragmented information from various sources. The rarity of their conditions means that opportunities for peer support are limited, whether through local health care providers, community groups, or even large-scale social networks. As a result, many patients turn to online communities and digital platforms as a means of finding solidarity, sharing experiences, and accessing critical information. Studies have shown that such online communities can be powerful tools for providing emotional and social support, helping patients manage both the physical and psychological aspects of their condition [7]. The mutual benefit of networking is illustrated by European activities such as Share4Rare to connect experts to improve knowledge and foster therapy [8-11]. Similarly, the necessity of including a community-based stakeholder group of those affected by rare pediatric diseases was recently illustrated by Berrios et al [12].

Many individuals with rare diseases are organized within common online social networks such as Facebook or WhatsApp, whereas individuals might also use Instagram and TikTok for engagement and information exchange. Individuals actively seek emotional support, disease-related information, and peer connection in the online world [13].

However, common online social networks are rarely designed to meet the needs of this population. General platforms may lack the privacy, specialized and trustworthy information, and support structures required for navigating rare and chronic diseases [14,15]. In addition, these social networks are less beneficial for those who do not know their diagnosis as they lack specific questionnaires or search methods for this population.

In the past years, the work on diagnostic support systems demonstrated the potential of computational matching approaches. Shen et al [16,17] demonstrated that undiagnosed individuals benefit from symptom-based matching, particularly in clinical decision support settings. Yates et al [18] showed that matching patient symptoms with the Human Phenotype Ontology allowed for the ranking of diagnoses. The open-source tool PatientMatcher also allows for phenotype-driven matching of undiagnosed individuals [19]. These previous works focus on diagnostic help for medical professionals but do not directly address affected people and their needs.

Given the unmet needs of individuals with rare and chronic diseases and their relatives and our experiences in mathematical classifying processes, we first prototyped a platform (called RarePairs) for individuals with rare diseases [20]. This prototype evaluated the hypothesis that a detailed questionnaire (focusing on daily activities and past medical history) could assist in matching individuals with different rare diseases. In the pilot project, a dataset of 700 anonymized questionnaires collected among individuals with different rare diseases was used; nearest neighbor was used for the classifying procedures. Briefly, the results indicated that a set of 53 questions might suit to “match” individuals with selected diseases according to their answer pattern [20]. This prototype was designed only to demonstrate the feasibility of questionnaire-based matching and did not include a publicly available platform, real-time user interaction, data security concepts, or real-world evaluation. This manuscript details the evolution of the prototypical idea (RarePairs) into a fully operational and secure app designed for diverse user groups (eg, patients, caregivers, experts, and health networkers) with ontology-based and machine learning (ML)–supported matching and real-world use and evaluation.

Specifically, this work contributes (1) a user-centered requirement analysis involving patients, caregivers, and professionals; (2) the development and iterative validation of an ontology-based and ML-supported matching logic for real-world use; (3) the establishment of a robust data security and privacy framework tailored to sensitive health data; and (4) an implementation approach designed to be feasible, scalable, and appropriate for health care–related applications.

XSL•FO
RenderX

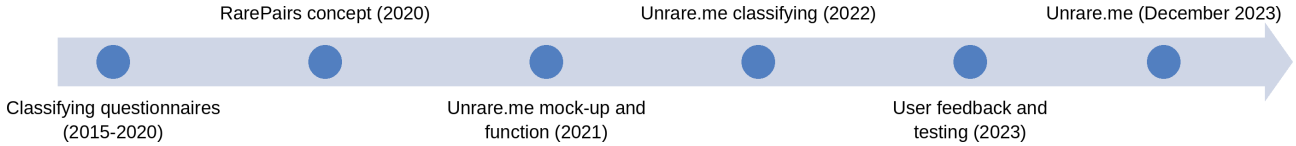
In essence, Unrare.me should function as a specialized networking platform for experts in and individuals with rare or undiagnosed health conditions, providing a vital digital space for support and information sharing.

Methods

Study Overview

The development of Unrare.me followed a multicenter, mixed methods approach. The process was led by a multiprofessional team comprising medical doctors from diverse specialties, including a focus on rare diseases; psychologists; representatives from patient advocacy groups; gaming experts; programmers; data analysts; and designers. The timeline is shown in Figure 1.

Figure 1. Timeline of the Unrare.me development process.



User Needs and Requirements

To assess the needs and preferences of potential users, we conducted 15 semistructured interviews targeting individuals impacted by rare diseases and their caregivers early after launching Unrare.me. The interviews served as initial orientation and gathered information in this new thematic field. The aim was to identify commonalities and typical features. Data analysis was based on the approach by Meuser and Nagel [21], consisting of paraphrasing important passages, creating headings for paraphrased text, and conducting thematic comparison.

App Development

Through a series of workshops using collaborative tools (eg, Miro boards), the interview findings were discussed and assessed for their relevance to the app's functional design. It became evident that there is no singular "need" for individuals with rare diseases; rather, needs vary over time, following the trajectory of the disease. For instance, immediately following diagnosis, users have a strong demand for information, whereas as the condition progresses, needs shift toward disease-specific support, such as finding experts for managing particular complications. To illustrate user needs, we employed the persona method and developed corresponding screen mock-ups (Figures 2-4).

XSL•FO
RenderX

Figure 2. User journey illustrated in mock-ups during app development. This mock-up user, alisa192, has been searching for a diagnosis for 4 years. She is 35 years old and female. In her short profile, she states that she has experienced persistent dizziness for 5 years, accompanied by numbness in her hands and other symptoms. She has gone through an odyssey of consultations with various specialists—unfortunately still without a clear diagnosis.

XSL•FO
RenderX

Figure 3. User journey illustrated in mock-ups during app development. In this screen, we see the dashboard of peterpiper62. He can see that his profile is 40% complete and that another user, yana21, has asked him a question ("Did your doctor run a full blood panel back then?"). He also has 5 additional unanswered questions. At the bottom of the screen, new relevant matches are displayed for him. At the very bottom, he can navigate between different app functions (eg, chat and matches).

XSL•FO
RenderX

Figure 4. User journey illustrated in mock-ups during app development. Here, the user can see a list of items in their profile that have already been completed and where additional information can or should be added. This user has already filled in the basics (eg, age and gender) and has written their short profile. However, they have not yet fully entered their symptoms. To integrate these diverse and shifting user needs into the app and its mock-ups, we used various user stories (persona method) at an early stage. This approach involved brainstorming based on fictional app users (persona concept) to understand how they might interact with the app, their potential expectations, and ways to address these in the app's design.



Matching Logic

To enable user-to-user connections ("matching"), it was necessary to collect essential information regarding participants' medical and personal histories. To ensure the compatibility of these data with the matching algorithm, a structured approach to organizing users' medical information was required. Therefore, suitable medical ontologies that would support standardized data representation were analyzed [22,23]. The review identified several relevant resources, including the Orphanet database and its derived ontology Orphanet Rare Disease Ontology, the Human Phenotype Ontology, Online

XSL•FO
RenderX

Mendelian Inheritance in Man, the International Classification of Inherited Metabolic Disorders, and Mondo Disease Ontology, as well as 2 classification systems specifically addressing autoinflammatory disorders and immunodeficiencies [24,25].

The matching logic underwent comprehensive discussion in the context of user stories, as well as considering the medical and patient group expertise present within the team. A particular focus was placed on defining the characteristics that would constitute optimal matching. This included exploring parameters such as shared diagnoses, symptomatology, disease progression stages, and lifestyle factors that might impact mutual support or therapeutic benefit.

Concurrently, the structure of the databases was assessed based on the merged ontologies to ensure that they could accommodate the complexity of matching criteria relevant to individuals with rare diseases and chronic conditions. This assessment involved a rigorous validation process in which potential matching algorithms were tested against automatically generated datasets to simulate real-world user scenarios. This testing aimed to refine the algorithms by evaluating their performance in achieving accurate and meaningful matches, ensuring that users relate to others who have comparable experiences, informational needs, or therapeutic requirements.

The integration of these steps provided a robust foundation for creating a responsive matching system that adapts to the varied and evolving needs of users grounded in both clinical insights and practical patient experiences. Beyond ontology-based similarity, user feedback was used to iteratively refine matching. The underlying ML-based optimization is described in detail elsewhere [26,27].

Evaluation Methods

Every user was offered to take part in a feedback survey through a banner on the start page of the app four month after launch. We used SoSci Survey (SoSci Survey GmbH), a secure cloud service for conducting online surveys, to collect data between December 11, 2023, and March 12, 2024.

In addition to demographic data, we collected feedback concerning usability, design, content, use, expectations, and overall impression. The survey used validated questionnaires for evaluating websites, which were slightly adapted for evaluating the app. It comprised 7 usability items from the Usability Metric for User Experience Lite [28] and the Perceived Website Usability–German [29]; 4 items on design from the Visual Aesthetics of Websites Inventory–Short Form [30]; 10 items from the Website–Clarity, Likeability, Informativeness, and Credibility [31]; 2 items measuring intention to revisit the website [32]; and 3 self-constructed items measuring fulfillment of expectations (1="Not at all"; 7="Absolutely"), trust in data security (1="Not at all"; 7="Absolutely"), and an overall rating (1="Very good"; 6="Very bad"). In addition, users were able to provide open feedback.

Participants were also asked whether they had found a match and, if so, to rate it on a visual analog scale from 0 to 100. In addition, they should rate the importance of 5 criteria for matching: same diagnosis, same symptoms, same everyday experiences, reason for app registration, age, and gender. Multiple selection was possible.

User statistics such as number of users, user group, and behavior of these user groups were analyzed separately and are not part of this paper. This evaluation was designed as a formative pilot implementation study to gather early user feedback on usability, perceived value, and initial matching experience. The survey was not powered for inferential group comparisons and, therefore, focuses on descriptive results.

Ethical Considerations

To take data privacy and ethical considerations into account, the app was developed in close collaboration with the senior data protection officer at University Hospital Bonn and in consultation with the local ethics committee. Due to the specific nature and purpose of the project—focused on app development—a formal ethics vote was deemed unnecessary.

All individuals who participated in the semistructured interviews were informed in advance about the purpose and use of the findings and provided explicit informed consent. All app users were informed about the data privacy terms, and participation in the user survey was voluntary and anonymous. According to § 2 (1) of the Statutes of the Ethics Committee of the Medical Faculty of the University of Bonn [33] and § 15 (1) of the Professional Code of Conduct for Physicians of the North Rhine Medical Association [34], no ethics committee approval was required for this study, as it did not constitute a clinical trial of a medicinal product or a clinical investigation of a medical device.

Results

User Needs and Requirements

Participants for the semistructured interviews included representatives from patient organizations (n=3), health care professionals (n=4), patients (n=4), and relatives (n=4). The ages of the participants ranged from 20 to 62 years, and their self-reported technical proficiency was rated as "average."

All participants reported currently using platforms such as Facebook and/or WhatsApp for peer communication, often focusing on specific aspects of their condition (eg, transplantation or augmentative and alternative communication). All participants also emphasized the importance of credibility and professionalism in the desired platform. Both were associated with the design and easy and failure-free management of the app. Data safety was not a predominant concern among participants. Participants expressed clear nonnegotiable features of the new platform. These included avoidance of commercial content (3/15, 20%); prevention of fake profiles and trolls (6/15, 40%); and minimal collection of personal data, limited strictly to essential information (5/15, 33%).

Concerns were also raised about unreliable information, populist influencers, the influence of the "Pharma industry," and negative interactions that might exacerbate psychological distress. To mitigate this, one-third of the participants (5/15, 33%) expressed a desire for active moderation to ensure a supportive environment. Additionally, participants highlighted the

XSL•FO
RenderX

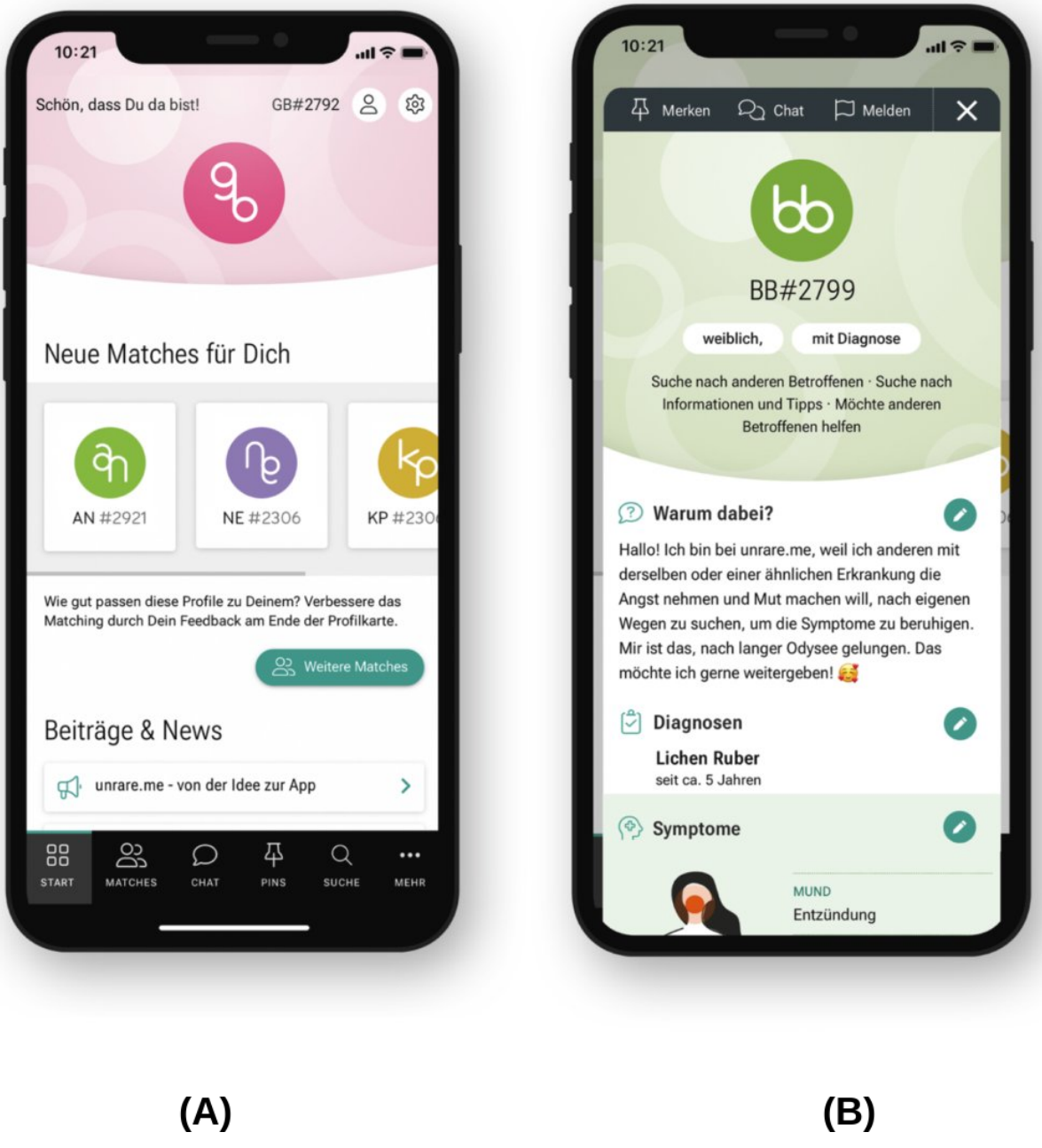
importance of encouraging positivity and sharing uplifting news to foster a sense of hope and encouragement.

Regarding platform functionality, participants expressed concerns about low user numbers potentially hindering fast and meaningful interactions. As a result, they emphasized the importance of customizable settings for profiles, visibility, and chat search functions. The opportunity to interact with researchers and experts was universally welcomed; however, health care professionals voiced apprehensions about potential overextension of their responsibilities beyond working hours. They suggested implementing "offline" modes to ensure boundaries between professional and personal time.

App Development

On the basis of the mock-ups shown in the methods section, we developed a preliminary data structure and designed the core screens of the app (Figures 5 and 6) and the brand design (Figure 7).

Figure 5. (A) The dashboard of user GB#2792. New matches are displayed here, along with new posts and news items relevant to the user. (B) The profile of BB#2799, a female patient with a diagnosis. Her profile states that she is looking for others affected by the condition as well as information and advice. She also wishes to support other patients. In her short profile, she explains that she joined Unrare.me to encourage others living with the disease and share her experiences. Under "Diagnoses," it is noted that she has been living with lichen ruber moniliformis for 5 years. At the bottom, partially visible, is the symptom body map, where she has indicated inflammation in the mouth.



XSL•FO
RenderX

Figure 6. The final app interface as displayed to a user registered in the patient role. It provides an overview of the different screens: the dashboard, the post page, the symptom map, an example of another user's profile, and feedback on the user's own profile completeness.



Figure 7. Brand design of the app: colors and typeface.



During the development of Unrare.me, the primary goal was to balance performance, stability, and cost efficiency. The back end needed to be lightweight, reliable, and inexpensive to operate while remaining easy to maintain—both by the Unrare.me team and potential external contributors.

We implemented a cost-efficient, maintainable cross-platform architecture to support iOS and Android users from a shared codebase and enable iterative updates. The technical stack was selected to ensure scalability, stability, and low operational overhead.

A robust data security framework was developed in collaboration with local data security experts. Key components of this framework included the hosting of servers in Germany to ensure compliance with local data protection regulations and

XSL•FO
RenderX

implementing an Alphatar concept to avoid collecting personal information of patients and their relatives. We also identified the most important data protection issues to be addressed:

- Using (end-to-end) encryption technologies for data safety
- Following the principle of data minimization (eg, Alphatar concept, refer to Figure 8)
- Establishing a secure end-to-end encrypted chat function
- Promoting transparency through data minimization and clear explanations of data collection purposes
- Providing detailed adjustment possibilities for personal and medical data (Figures 9 and 10), permanently and irrecoverably deleting all user data upon account deletion.

The app became available in popular app stores for Apple and Android in December 2023. During the first 6 months of operation of the online platform, no major technical issues were recorded. Users had the possibility to provide feedback on technical issues via email and specific support buttons inside the app. An increasing number of users necessitated technical adjustments: computational power and server capacities had to be optimized to ensure fast matching and the stable functionality of the app.

XSL•FO
RenderX

Figure 8. The Alphatar concept of Unrare.me. Users are anonymous using a 2-letter chiffre and a color during private chats. In this screen, the user is encouraged to enter their initials and choose a color for the avatar.

XSL•FO
RenderX

Figure 9. The screen shows information on how users can adjust the visibility of their profile, including personal information and symptoms (“Other users can see my user profile”).

XSL•FO
RenderX

Figure 10. On this screen, users can choose whether other users may contact them and, if so, which user groups (eg, other patients, physicians, and research centers).

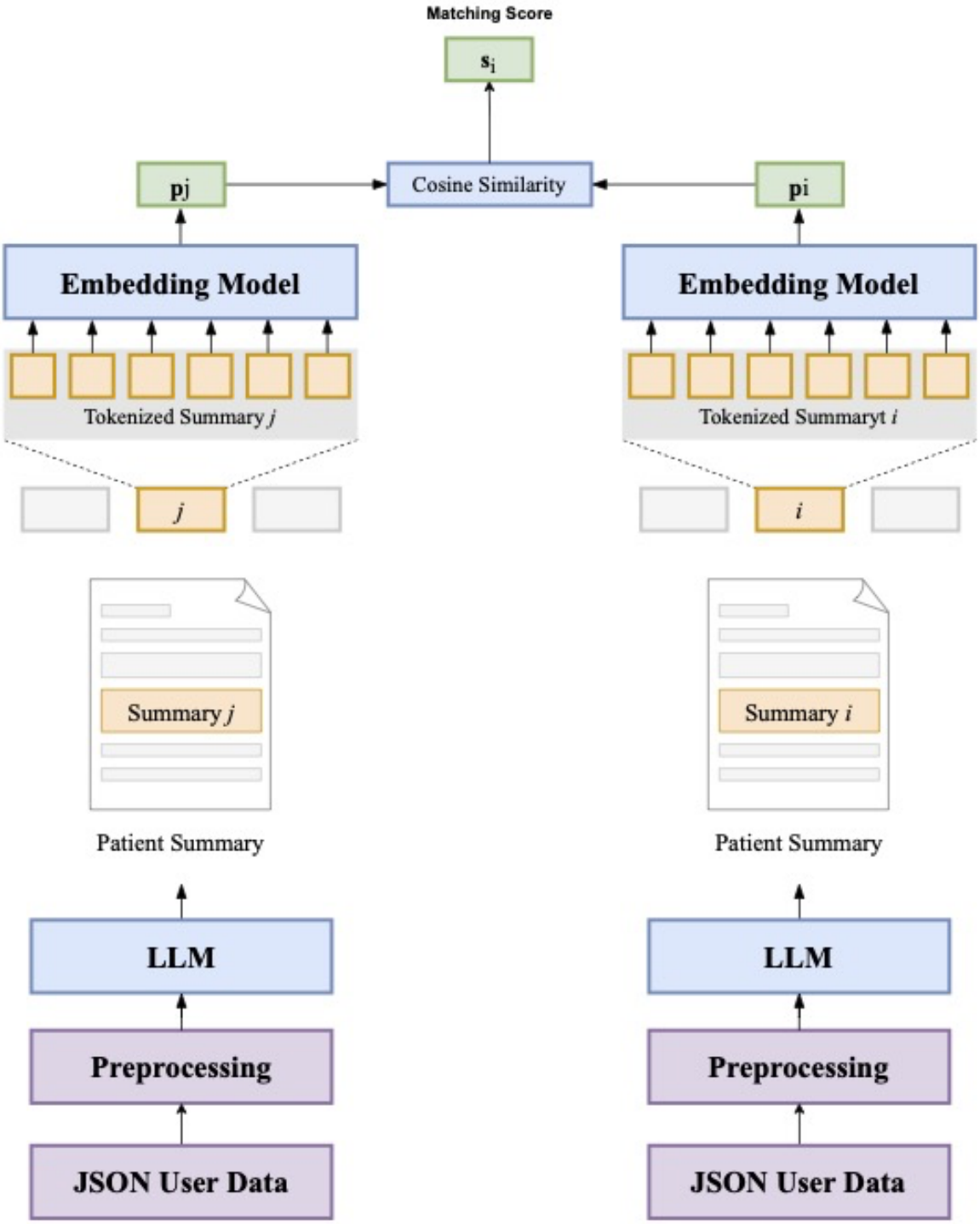
XSL•FO
RenderX

Matching Approach

For the Unrare.me matching architecture, the different open-source ontologies evaluated were translated, merged, weighted, and integrated into the back-end structure. The matching algorithm was primarily based on the similarity of diagnoses but also considered symptoms, age, and gender. A user feedback system enabled users to rate each matching suggestion. This information was then used by the algorithm (Figure 11) to further refine the matching process beyond ontology-based criteria [35].

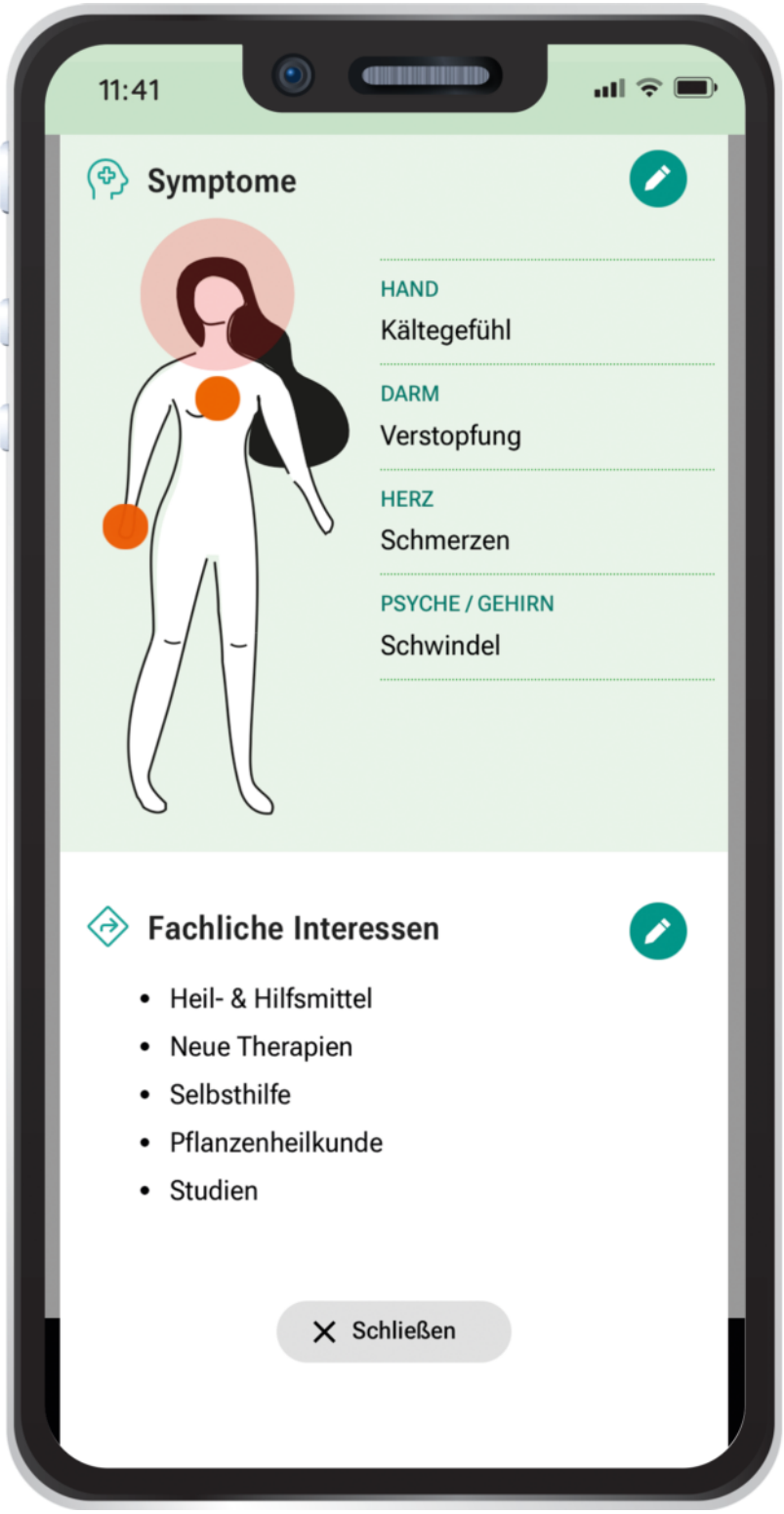
Therefore, we fully designed and implemented a symptom finder that can be used by laypeople (Figure 12).

Figure 11. Schematic visualization of the user data augmentation approach. LLM: large language model.



XSL•FO
RenderX

Figure 12. A body compass and a symptom checker enable users to enter relevant clinical information themselves. On this screen, the user has entered symptoms such as feeling cold, constipation, pain, and dizziness, as well as topics of interest such as assistive devices, new therapies, self-help, herbal medicine, and clinical studies. The technical algorithm approach is described by Berger et al [26,27].



Evaluation Study

Overview

To obtain user feedback on the first version of the app, a feedback questionnaire was made available to users during the first 3 months after the launch. More detailed information about user groups and their behavior will be published in a separate paper.

Accordingly, the findings (Table 1) should be interpreted as early implementation insights rather than evidence of effectiveness.

XSL•FO
RenderX

Table . Results of the user experience evaluation measured using the Usability Metric for User Experience Lite; Perceived Website Usability–German; Visual Aesthetics of Websites Inventory–Short Form; and Website–Clarity, Likeability, Informativeness, and Credibility. Items translated from German.

Questionnaire and items	Score (1-7), mean (SD)
Usability (in general)	5.81 (0.29)
The app's capabilities meet my requirements.	5.45 (1.42)
The app is easy to use.	6.12 (1.03)
The way of function of the app is easy to understand.	6.13 (1.03)
I can easily understand the structure of the app.	5.93 (1.17)
It is easy to navigate the app.	5.91 (1.16)
I can quickly access the content I was looking for.	5.53 (1.29)
The app meets accessibility standards.	5.59 (1.22)
Content (total)	5.51 (1.23)
The app sparks my interest.	6.23 (1.01)
The content is of high quality.	5.27 (1.28)
The text provides the most important information concisely.	5.45 (1.25)
The app's content is engaging.	5.17 (1.29)
The language used is common and easy to understand.	5.99 (1.11)
The content is presented in a clear and illustrative manner.	5.63 (2.24)
The app is useful.	5.50 (1.35)
I can trust the app.	5.46 (1.21)
The app is informative.	5.24 (1.28)
I enjoy using this app.	5.11 (1.33)
Design (total)	5.59 (1.20)
Everything in the app fits together cohesively.	5.36 (1.29)
The design is pleasantly varied.	5.51 (1.24)
The color scheme is attractive.	5.67 (1.16)
The design is professional.	5.84 (1.12)

Sample

A total of 15.8% (270/1708) of the app's user base at the time took part in the survey. Most participants (240/270, 88.9%) were female. Patients constituted 65.2% (176/270) of the sample, followed by relatives (79/270, 29.3%), health care professionals (13/270, 4.8%), and health network professionals (2/270, 0.7%). Most of the respondents (221/270, 81.9%) stated that they had registered on the app to find other people in the same situation, and 56.7% (153/270) of the users were looking for experts on their diagnosis or symptoms. Another 44.4% (120/270) were looking for information on their condition. Furthermore, 32.2% (87/270) registered out of curiosity, and 23% (62/270) stated that they wanted to find the correct diagnosis. This pattern suggests that Unrare.me was primarily used as a peer support tool, while expert access and information seeking represented important secondary motivations.

Usability, Design, Content, and Satisfaction of User Demands

Overall, the self-constructed items "I will use the app again" and "I would recommend the app to others" received high levels of agreement, with mean scores of 6.05 (SD 1.08) and 5.99 (SD 1.18), respectively. The item "The app met my expectations and requirements" showed an average level of agreement (mean 4.74, SD 1.37). Participants also rated the question about perceived data security on the app with an average score of 5.58 (SD 1.15).

In addition, the item asking users to give an overall rating for the app (1=highest rating; 6=lowest rating) yielded an average of 2.24 (SD 0.90).

High usability and trust ratings indicate that privacy-oriented design does not necessarily compromise user experience in rare disease communities.

In general, the participants' open-text responses reflected both positive feedback and suggestions for improvement. On the one hand, the app was described as "a great idea," with users valuing the possibility to connect with others "sharing the same destiny." On the other hand, participants requested enhanced matching features (eg, regional search) and the addition of group chats.

During registration, 77% (208/270) of the participants found their diagnosis on the app. However, 17% (46/270) only found

XSL•FO
RenderX

their diagnosis partially, and others (22/270, 8.1%) only found it under a different name.

The main goal of obtaining at least one match in the first 3 months was achieved by 60.7% (164/270) of the participants. On a visual analog scale from 0 to 100, the mean satisfaction with this match was 42.9 (SD 28.2; 1=worst match; 100=best match). According to the participants, the most relevant factors for matching were having the same diagnosis (mean 82.5, SD 25.1), symptoms (mean 74.2, SD 25.8), and everyday experiences (mean 69.6, SD 29.5). The age (mean 27.8, SD 31.3) and gender (mean 11.8, SD 23.5) of matching partners were rated as less relevant.

Discussion

Principal Findings

Networking allows people with health issues to learn from others' experiences [36]. This is particularly evident in the widespread success of platforms such as Facebook for patient groups [37]. However, there is a significant lack of appropriate digital networking spaces for health-related topics, especially for individuals with less common diseases [38]. Although there are isolated efforts to establish noncommercial digital networks within the European Union, such as Share4Rare, these initiatives have so far had little visible impact on service provision [12]. As a result, many people turn to commercial social media platforms such as Facebook. This brings with it the danger of information overload and makes it difficult to distinguish between valid personal experiences or scientific information and advertising or hearsay [37].

Recognizing the value of shared experiences as a resource for individuals with chronic and rare diseases, we developed and launched the social network Unrare.me in December 2023. The app provides a platform for networking with other affected individuals and health care professionals, enabling the exchange of information and experiences in a data-secure environment.

The goals of Unrare.me are in line with the analyses by Smith et al [39], who identified these factors as resources for improving health. Similarly, as early as 2010, the World Health Organization promoted the impact of peer support in managing complex health challenges in the context of diabetes, including "horizontal and vertical networking among users," "providing access to relevant information," "promoting self-help and peer support," "alleviating the burden on individuals and their families," and "leveraging the collective knowledge of users (lay expertise and collective intelligence)" [40].

The results of our evaluation underline the successful implementation process of Unrare.me. Overall, users found the app easy to use and interesting, and its design was considered professional. Participants' high rating (mean 5.99, SD 1.18 on a 7-point scale) for the intention to recommend the app indicates strong user motivation to share the app with others, suggesting a generally positive user experience. However, given its innovative approach, the development and implementation of Unrare.me faced several challenges. First, effective marketing is crucial to attract a critical mass of users, yet the resources for marketing in this primarily scientific project were limited. The question of what constitutes a sufficient user base for a rare disease network remains. Due to the diversity of rare diseases, it is still uncertain whether—or to what extent—users can genuinely learn from each other when their conditions or diseases differ significantly. This problem is evident in the evaluation results. The evaluation was conducted at an early stage, before many users had registered. Although three-quarters of users (208/270, 77%) were offered a match, this was not always an ideal fit due to the limited size of the database, as reflected in the slightly negative assessment of the matching. The rating would probably be higher if the evaluation were conducted at this point given the considerable efforts that have been made to improve the app, especially the matching algorithm.

Second, ensuring optimal data privacy is a persistent challenge. Users need assurance that their personal and health information is secure. The balance between providing a useful service and maintaining stringent data privacy standards is delicate and ongoing []. The Unrare.me initiative addressed this issue by using both an extremely transparent information approach and an information-conserving strategy. Another concern is the participation of minors in social networks. Specific measures need to be implemented to protect younger users while allowing them to benefit from the network [41].

Third, it is crucial to ensure the long-term sustainability of Unrare.me. Planned future enhancements to the app include additional interaction features such as group chats, structured feedback mechanisms, bulletin boards, and current news updates. These features aim to foster a more vivid, engaging, and supportive community. However, this also involves exploring viable business models and alternative funding sources so that the platform can be maintained and expanded without relying on a traditional commercial approach.

Finally, reducing barriers to access is essential. This involves enhancing accessibility features, expanding the platform's international reach, and incorporating multilingual support to cater to a diverse user base.

Consequently, the Unrare.me team is currently planning the further development of the platform, including technical aspects, funding, collaboration, matching optimization, and new features for the app. The continuous positive feedback from users serves as positive reinforcement and illustrates the urgent need for a network such as Unrare.me, which is expected to grow with the support of strong stakeholders. Specific next steps include the establishment of an Unrare.me registry structure as well as "meet the expert" offerings within the platform.

Limitations

An evaluation of Unrare.me was carried out alongside app development using a multimodal approach. However, monitoring was only possible for the first 3 months after app launch. Therefore, the results may only be of limited significance. As previously mentioned, the small user pool for matching in this early phase is a limitation. However, technical problems in the early stages may also have distorted the results. Given the many improvements made since then, it can be assumed that the evaluation would be more positive today.

XSL•FO
RenderX

However, on the other hand, only 1 in 7 users took part in the survey. Therefore, distortions caused by particularly committed individuals cannot be ruled out.

The capacity for conducting interviews was also limited at the beginning. A total of 15 qualitative interviews were conducted. To reflect a wide range of needs and concerns, the interviewees were carefully selected based on factors such as age, gender, digital literacy, and migration background. Nevertheless, the interviews did not cover all interests. To take these limitations into account, a re-evaluation of the app is planned.

In conclusion, while Unrare.me addresses a significant gap in the digital networking space for individuals with chronic and rare diseases, several areas require ongoing attention and development. Effective marketing, robust data privacy measures, sustainable funding models, enhanced user interaction, and improved accessibility are key to the platform's success. By continuing to evolve and adapt, Unrare.me has the potential to become a vital resource for the rare disease community, facilitating meaningful connections and support.

Acknowledgments

The authors would like to thank the Kindernetzwerk e.V. and all involved self-help groups for their valuable support that significantly enriched this study. Dr Claudio Cinquemani (CureRare), Dr David Katzer (University Hospital Bonn), and Eric Seitz and his team provided helpful feedback on how to improve the app. Achim Flender is acknowledged for helpful feedback on data safety aspects within Unrare.me. Dr Werner Lechner and Professor Frank Klawonn are gratefully acknowledged for valuable support and fruitful discussions during the early phase of the project. The authors are also grateful to their first real-life testers for their insightful feedback and encouragement. Professor Dr. med. Lars Pape and Dr. med. Metin Cetiner from the NEOCYST initiative are acknowledged for their excellent collaboration. Professor Nadine Scholten provided valuable insights on the perspective of patients with chronic diseases, and Professor Dagmar Wachten and team (Institute of Innate Immunity) connected the Unrare.me initiative to the Bardet-Biedl syndrome community, which was of immense help. The authors used GPT-5.1 (OpenAI) for language and grammar improvement.

Funding

This research was supported by the German Federal Ministry of Health (*Bundesministerium für Gesundheit*) under grant ZMI5-2521SBG004.

Data Availability

Data are available on request from the authors.

Conflicts of Interest

None declared.

References

1. Sestini S, Paneghetti L, Lampe C, et al. Social and medical needs of rare metabolic patients: results from a MetabERN survey. Orphanet J Rare Dis 2021 Aug 3;16(1):336. [doi: 10.1186/s13023-021-01948-5] [Medline: 34344397]
2. Namsrai T, Parkinson A, Chalmers A, et al. Diagnostic delay of myositis: an integrated systematic review. Orphanet J Rare Dis 2022 Nov 21;17(1):420. [doi: 10.1186/s13023-022-02570-9] [Medline: 36411487]
3. Grigull L, Mehmecke S, Rother AK, et al. Common pre-diagnostic features in individuals with different rare diseases represent a key for diagnostic support with computerized pattern recognition? PLoS One 2019;14(10):e0222637. [doi: 10.1371/journal.pone.0222637] [Medline: 31600214]
4. Manalel JA, Sumrall S, Davidson H, Grewal M, Granovetter MA, Koehly LM. Stress, coping, and positive aspects of caregiving among caregivers of children with rare disease. Psychol Health 2024;39(2):216-232. [doi: 10.1080/08870446.2022.2057494] [Medline: 35620936]
5. Sequeira AR, Mentzakis E, Archangelidi O, Paolucci F. The economic and health impact of rare diseases: a meta-analysis. Health Policy Technol 2021 Mar;10(1):32-44. [doi: 10.1016/j.hlpt.2021.02.002]
6. Inhestern L, Otto R, Brandt M, et al. Patient experiences of interprofessional collaboration and intersectoral communication in rare disease healthcare in Germany - a mixed-methods study. Orphanet J Rare Dis 2024 May 13;19(1):197. [doi: 10.1186/s13023-024-03207-9] [Medline: 38741100]
7. Long JC, Best S, Nic Giolla Easpaig B, et al. Needs of people with rare diseases that can be supported by electronic resources: a scoping review. BMJ Open 2022 Sep 1;12(9):e060394. [doi: 10.1136/bmjopen-2021-060394] [Medline: 36581982]
8. Breaking the access deadlock to leave no one behind. EURORDIS. 2017. URL: https://www.eurordis.org/publications/breaking-the-access-deadlock-to-leave-no-one-behind/ [accessed 2026-04-21]
9. Radu R, Hernández-Ortega S, Borrega O, et al. Global collaborative social network (Share4Rare) to promote citizen science in rare disease research: platform development study. JMIR Form Res 2021 Mar 29;5(3):e22695. [doi: 10.2196/22695] [Medline: 33779572]

XSL•FO
RenderX

10. Reinhard C, Bachoud-Lévi AC, Bäumer T, et al. The European Reference Network for rare neurological diseases. Front Neurol 2021;11:616569. [doi: 10.3389/fneur.2020.616569] [Medline: 33519696]
11. Brodszki N, Frazer-Abel A, Grumach AS, et al. European Society for Immunodeficiencies (ESID) and European Reference Network on Rare Primary Immunodeficiency, Autoinflammatory and Autoimmune Diseases (ERN RITA) complement guideline: deficiencies, diagnosis, and management. J Clin Immunol 2020 May;40(4):576-591. [doi: 10.1007/s10875-020-00754-1] [Medline: 32064578]
12. Berrios C, McBeth M, Bradley-Ewing A, et al. Developing a community-led rare disease ELSI research agenda. Orphanet J Rare Dis 2024 Jan 22;19:23. [doi: 10.1186/s13023-023-02986-x]
13. Bi Q, Shen L, Evans R, et al. Determining the topic evolution and sentiment polarity for Albinism in a Chinese online health community: machine learning and social network analysis. JMIR Med Inform 2020 May 29;8(5):e17813. [doi: 10.2196/17813] [Medline: 32469320]
14. Ashtari S, Taylor AD. The internet knows more than my physician: qualitative interview study of people with rare diseases and how they use online support groups. J Med Internet Res 2022 Aug 25;24(8):e39172. [doi: 10.2196/39172] [Medline: 36006679]
15. Mcmullan J, Lohfeld L, McKnight AJ. Needs of informal caregivers of people with a rare disease: a rapid review of the literature. BMJ Open 2022 Dec 12;12(12):e063263. [doi: 10.1136/bmjopen-2022-063263] [Medline: 36523233]
16. Shen F, Liu S, Wang Y, Wang L, Afzal N, Liu H. Leveraging collaborative filtering to accelerate rare disease diagnosis. AMIA Annu Symp Proc 2018;2017:1554-1563. [Medline: 29854225]
17. Shen F, Liu S, Wang Y, Wen A, Wang L, Liu H. Utilization of electronic medical records and biomedical literature to support the diagnosis of rare diseases using data fusion and collaborative filtering approaches. JMIR Med Inform 2018 Oct 10;6(4):e11301. [doi: 10.2196/11301] [Medline: 30305261]
18. Yates J, Gutiérrez-Sacristán A, Jouhet V, et al. Finding commonalities in rare diseases through the undiagnosed diseases network. J Am Med Inform Assoc 2021 Jul 30;28(8):1694-1702. [doi: 10.1093/jamia/ocab050] [Medline: 34009343]
19. Rasi C, Nilsson D, Magnusson M, et al. PatientMatcher: a customizable python-based open-source tool for matching undiagnosed rare disease patients via the matchmaker exchange network. Hum Mutat 2022 Jun;43(6):708-716. [doi: 10.1002/humu.24358] [Medline: 35192731]
20. Kühnle L, Mücke U, Lechner WM, Klawonn F, Grigull L. Development of a social network for people without a diagnosis (RarePairs): evaluation study. J Med Internet Res 2020 Sep 29;22(9):e21849. [doi: 10.2196/21849] [Medline: 32990634]
21. Meuser M, Nagel U. Das experteninterview — konzeptionelle grundlagen und methodische anlage. In: Pickel S, Pickel G, Lauth HJ, Jahn D, editors. Methoden Der Vergleichenden Politik- Und Sozialwissenschaft: VS Verlag für Sozialwissenschaften; 2009. [doi: 10.1007/978-3-531-91826-6_23]
22. Ferreira CR, Rahman S, Keller M, Zschocke J, ICIMD Advisory Group. An international classification of inherited metabolic disorders (ICIMD). J Inherit Metab Dis 2021 Jan;44(1):164-177. [doi: 10.1002/jimd.12348] [Medline: 33340416]
23. Robinson PN, Graessner H. Rare-disease data standards [Article in German]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2022 Nov;65(11):1126-1132. [doi: 10.1007/s00103-022-03591-2] [Medline: 36149471]
24. Haimel M, Pazmandi J, Heredia RJ, et al. Curation and expansion of Human Phenotype Ontology for defined groups of inborn errors of immunity. J Allergy Clin Immunol 2022 Jan;149(1):369-378. [doi: 10.1016/j.jaci.2021.04.033] [Medline: 33991581]
25. Maassen W, Legger G, Kul Cinar O, et al. Curation and expansion of the Human Phenotype Ontology for systemic autoinflammatory diseases improves phenotype-driven disease-matching. Front Immunol 2023;14:1215869. [doi: 10.3389/fimmu.2023.1215869] [Medline: 37781402]
26. Berger A, Berghaus D, Bashir AH, et al. Advancing personalized medicine: a scalable LLM-based recommender system for patient matching. Presented at: 2024 IEEE International Conference on Big Data (BigData); Dec 15-18, 2024. [doi: 10.1109/BigData62323.2024.10825567]
27. Berger A, Lagones TA, Grigull L, et al. Tackling data sparsity and combinatorial challenges in rare disease matching with medical informed machine learning. Presented at: 2024 IEEE International Conference on Big Data (BigData); Dec 15-18, 2024. [doi: 10.1109/BigData62323.2024.10825404]
28. Lewis JR, Utesch BS, Maher DE. UMUX-LITE: when there's no time for the SUS. Presented at: CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; Apr 27 to May 2, 2013. [doi: 10.1145/2470654.2481287]
29. Flavián C, Guinalíu M, Gurrea R. The role played by perceived usability, satisfaction and consumer trust on website loyalty. Inf Manag 2006 Jan;43(1):1-14. [doi: 10.5555/1140892.1709369]
30. Moshagen M, Thielsch MT. A short version of the visual aesthetics of websites inventory. Behav Inf Technol 2013 Dec;32(12):1305-1311. [doi: 10.1080/0144929X.2012.694910]
31. Thielsch MT, Hirschfeld G. Facets of website content. Hum Comput Interact 2018;34(4):279-327. [doi: 10.1080/07370024.2017.1421954]
32. Moshagen M, Thielsch MT. Facets of visual aesthetics. Int J Hum Comput Stud 2010 Oct;68(10):689-709. [doi: 10.1016/j.ijhcs.2010.05.006]

XSL•FO
RenderX

33. Medical Faculty of the University of Bonn. Statutes of the Ethics Committee of the Medical Faculty of the Rheinische Friedrich-Wilhelms-Universität Bonn. Amended and restated version of 29 April 2024 [Article in German]. Rheinische Friedrich-Wilhelms-Universität Bonn. 2024. URL: https://ethik.meb.uni-bonn.de/satzung/ [accessed 2026-05-06]
34. Professional Code of Conduct for Physicians of the North Rhine Medical Association. Version of 14 November 1998, as amended on 22 March 2025 (entered into force on 23 July 2025) [Report in German]. : North Rhine Medical Association; 2025. URL: https://www.aekno.de/fileadmin/user_upload/aekno/downloads/2025/berufsordnung-2025.pdf [accessed 2026-05-06]
35. Zhu Q, Nguyen DT, Alyea G, Hanson K, Sid E, Pariser A. Phenotypically similar rare disease identification from an integrative knowledge graph for data harmonization: preliminary study. JMIR Med Inform 2020 Oct 2;8(10):e18395. [doi: 10.2196/18395] [Medline: 33006565]
36. Castro AR, Chougui K, Bilodeau C, Tsimicalis A. Exploring the views of osteogenesis imperfecta caregivers on internet-based technologies: qualitative descriptive study. J Med Internet Res 2019 Dec 18;21(12):e15924. [doi: 10.2196/15924] [Medline: 31850851]
37. Titgemeyer SC, Schaaf CP. Facebook support groups for rare pediatric diseases: quantitative analysis. JMIR Pediatr Parent 2020 Nov 19;3(2):e21694. [doi: 10.2196/21694] [Medline: 33211019]
38. Cardão C, Barros L, Francisco R, Silva D, Ferreira VR. Experiences of parents with children with congenital disorders of glycosylation: what can we learn from them? Disabil Health J 2021 Jul;14(3):101065. [doi: 10.1016/j.dhjo.2021.101065] [Medline: 33531289]
39. Smith J, Damm K, Hover G, Chien J. Lessons from an experiential approach to patient community engagement in rare disease. Clin Ther 2021 Feb;43(2):421-429. [doi: 10.1016/j.clinthera.2020.12.002] [Medline: 33375998]
40. Boothroyd RI, Fisher EB. Peers for progress: promoting peer support for health around the world. Fam Pract 2010 Jun;27 Suppl 1:i62-i68. [doi: 10.1093/fampra/cmq017] [Medline: 20483801]
41. Hausmann JS, Weitzman ER. The promises and perils of social media for pediatric rheumatology. Rheum Dis Clin North Am 2022 Feb;48(1):233-243. [doi: 10.1016/j.rdc.2021.09.005] [Medline: 34798949]

Abbreviations

ML: machine learning

Edited by A Kushniruk; submitted 07.Jul.2025; peer-reviewed by E Nielsen, E Opoku-Agyemang; revised version received 25.Jan.2026; accepted 09.Feb.2026; published 15.May.2026.

Please cite as:
Fendrich L, Högl H, Schmidt R, Bascom D, Berger A, Lübbering M, Ernst G, Schumacher J, Hölzer S, Grigull L
Development, Implementation, and Preliminary Results of a Novel Social Network for Rare Disease Communities: Mixed Methods Study
JMIR Hum Factors 2026;13:e80230
URL: https://humanfactors.jmir.org/2026/1/e80230
doi:10.2196/80230

©Lara Fendrich, Henriette Högl, Ralf Schmidt, David Bascom, Armin Berger, Max Lübbering, Gundula Ernst, Justus Schumacher, Sabrina Hölzer, Lorenz Grigull. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 15.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Mobile App (Tpro) for Symptom Management in Patients With Deep Vein Thrombosis Based on Patient-Reported Outcomes: Design and Development Using an Iterative Convergent Mixed Methods Approach

Qiaodan Lu[1], MS; Yafei Liu[1], BS; Manna Shao[2], BS; Lei Wang[1], MS

[1]Department of Vascular Surgery, Chinese Academy of Medical Sciences and Peking Union Medical College, Peking Union Medical College Hospital, No. 1, Shuaifuyuan, Dongcheng District, Beijing, China

[2]Department of Nursing, Chinese Academy of Medical Sciences and Peking Union Medical College, Peking Union Medical College Hospital, Beijing, China

Corresponding Author:
Lei Wang, MS
Department of Vascular Surgery, Chinese Academy of Medical Sciences and Peking Union Medical College, Peking Union Medical College Hospital, No. 1, Shuaifuyuan, Dongcheng District, Beijing, China

Abstract

Background: Deep vein thrombosis (DVT) is a significant global health issue, often associated with a high symptom burden and reduced quality of life, especially after discharge. Traditional symptom management models are typically passive, clinician-centered, and lack real-time monitoring and feedback, which can lead to delayed interventions and poor patient engagement. While mobile health (mHealth) interventions offer a promising alternative, they require rigorous usability testing to ensure both efficacy and adoption.

Objective: This study aimed to design and develop Tpro (developed by the Department of Vascular Surgery, Peking Union Medical College Hospital), a patient-reported outcome (PRO)–based mobile app for patients with DVT. Using an iterative convergent mixed methods design, the app seeks to enable proactive symptom monitoring, health education, clinician-patient interaction, and peer support, thereby optimizing its usability, functionality, and alignment with patient needs.

Methods: The development followed an iterative convergent mixed methods design, comprising predevelopment and iterative optimization phases. Initial functions were informed by qualitative interviews with 14 patients with DVT. Over 4 iterative cycles, qualitative feedback and quantitative usability data (including task completion rates and User Interface Usability Questionnaire [UIUQ] scores) were concurrently collected, analyzed, and integrated via joint displays to guide refinements in content, interface, and system architecture until usability benchmarks were met.

Results: The final Tpro app encompasses 4 core modules, including gamified symptom reporting, multimodal health education, clinician-patient communication, and a peer support community. Iterative testing identified and resolved key usability issues. The final prototype demonstrated high usability, achieving an excellent UIUQ score (mean 89, SD 13) and a task completion rate of 92% (12/13), indicating high user acceptance and operational reliability.

Conclusions: Applying the iterative convergent mixed methods approach enabled the systematic and user-centered development of Tpro. This methodology effectively integrated diverse stakeholder feedback into a functional and engaging patient-reported outcome–based app, ready for subsequent efficacy trials. This approach offers a replicable model for developing digital health tools in complex clinical contexts.

(JMIR Hum Factors 2026;13:e92738) doi:10.2196/92738

KEYWORDS

deep vein thrombosis; patient-reported outcomes; mobile app; mixed methods research; iterative design

Introduction

Deep vein thrombosis (DVT) has emerged as a significant contributor to cardiovascular disease-related mortality on a global scale [1]. Individuals afflicted with DVT frequently experience a substantial array of acute and chronic symptoms, including persistent pain, swelling, functional impairment, and psychological distress, leading to a diminished quality of life and subsequently escalating the overall economic burden on health care systems worldwide [2].

Effective symptom management is crucial for predicting adverse prognostic events, preventing complications, optimizing

XSL•FO
RenderX

treatment outcomes, improving overall health, and enhancing quality of life in patients with DVT [3]. However, this traditional symptom management model operates on a passive-reactive and clinician-centric framework, predominantly reliant on infrequent routine ward rounds or scheduled in-person hospital consultations [4]. This approach inherently lacks mechanisms for active, real-time monitoring of patients' symptoms, particularly those occurring between visits or in out-of-hospital settings, and fails to adequately incorporate the patient's subjective experience and perspective. Consequently, health care providers are often unable to promptly identify worsening symptoms or address emerging patients' needs, leading to significant delays in intervention and a critical disconnect between the patient's actual clinical status and the health care system's responsiveness.

The integration of patient-reported outcomes (PRO) into symptom management exemplifies a patient-centered approach, fully aligning with core principles of patient-centered medical care. This model empowers patients to autonomously monitor symptoms, alleviate symptom burden, and enhance quality of life [5]. While widely applied in clinical trials, drug approval, and tumor prognosis, this model has yet to be integrated into symptom management for patients with DVT [6,7].

Mobile health (mHealth) apps offer a promising platform to facilitate PRO collection and support self-management, enabling patients to engage in disease management beyond clinical settings [8]. Nevertheless, many mHealth solutions face poor adoption due to insufficient usability and a lack of end-user involvement during development [9]. Currently, there is no mHealth app based on PRO that has been developed with a rigorous user-centered design approach specifically tailored to address the symptom management needs of patients with DVT.

To tackle these obstacles, this study used an iterative convergent mixed methods design [10], a robust methodology that concurrently gathers and evaluates qualitative and quantitative data throughout various development phases, to create Tpro, an mHealth app centered on PRO specifically for individuals with DVT. The app is structured to facilitate proactive symptom monitoring, deliver health education, enhance clinician-patient communication, and provide peer support. In practice, patients are guided through a daily workflow and after receiving a customizable reminder, they complete a symptom report, during which they can message their clinician. Upon submission, tailored health education content is automatically unlocked. A knowledge library and a peer community remain available for self-directed use at any time. By merging the PRO framework with iterative user-centered development, Tpro strives to enhance usability, functional significance, and alignment with patient requirements, thereby promoting a more adaptive and patient-involved approach to managing DVT symptoms.

Methods

Overview of Development Methodology

The development of Tpro used an iterative convergent mixed methods design as formalized by Alwashmi et al [10]. This approach is characterized by the concurrent collection and integration of qualitative and quantitative data across cyclical iterations, continuing until predefined usability endpoints are achieved [10]. In this study, the process was structured across 2 interlinked phases, namely a predevelopment phase, which defined core functions and wireframes based on prior qualitative inquiry and clinical consensus, and an iterative optimization phase consisting of 4 consecutive cycles of prototype testing and refinement. Within each cycle, qualitative feedback and quantitative usability metrics were systematically integrated through joint displays and team interpretation sessions, directly informing real-time design modifications. This evidence-driven, user-centered process continued until design saturation was achieved, that is, until no new critical usability issues emerged and key performance metrics stabilized.

Predevelopment Phase

Defining Core Functions Through Qualitative Inquiry

The primary function of Tpro was derived from qualitative interviews conducted in the initial phase of the study with 14 patients with DVT engaged in 6-month out-of-hospital self-management [11]. The findings suggest that enhancing health education can enhance patients' understanding, improve doctor-patient communication, facilitate monitoring, and establish a support system to enhance the quality of out-of-hospital patient with DVT management [11].

Expanding on these results, the researchers, leveraging their substantial clinical expertise in managing patients with DVT (each investigator with at least 5 years of experience), reached a consensus. As a result, the primary roles of Tpro were identified as symptom monitoring, health education, facilitating medical-patient interaction, and fostering peer support among patients.

Wireframing and Content Development

Wireframes, shown as line diagrams and geometric placeholders, define spatial hierarchy, modular component organization, and relationships among interactive elements. They explicitly convey an app's basic architectural topology and provide a visual framework for design and development teams. Wireframes are used for concept communication, user-flow optimization, and the establishment of the structural foundation during the initial development stage [12]. The researchers generated a wireframe diagram of Tpro based on its 4 primary functions. As shown in Multimedia Appendix 1, the wireframe illustrates the initial layout and user flow among the symptom monitoring, education, communication, and community modules.

Following the definition of the app's core functions and structure, the app was named "Tpro," a term derived from combining the target condition ("Thrombosis") with the core methodology ("Patient-Reported Outcomes"). The multimedia content was subsequently developed through a systematic, evidence-based process. This included original anatomical illustrations (eg, diagrams of human vasculature) created by medical illustrators; educational texts and instructional videos, adapted from existing patient education materials and produced in collaboration with vascular nursing specialists, with all materials undergoing health literacy review [13]; a knowledge question bank clinically validated by experts; and gamified

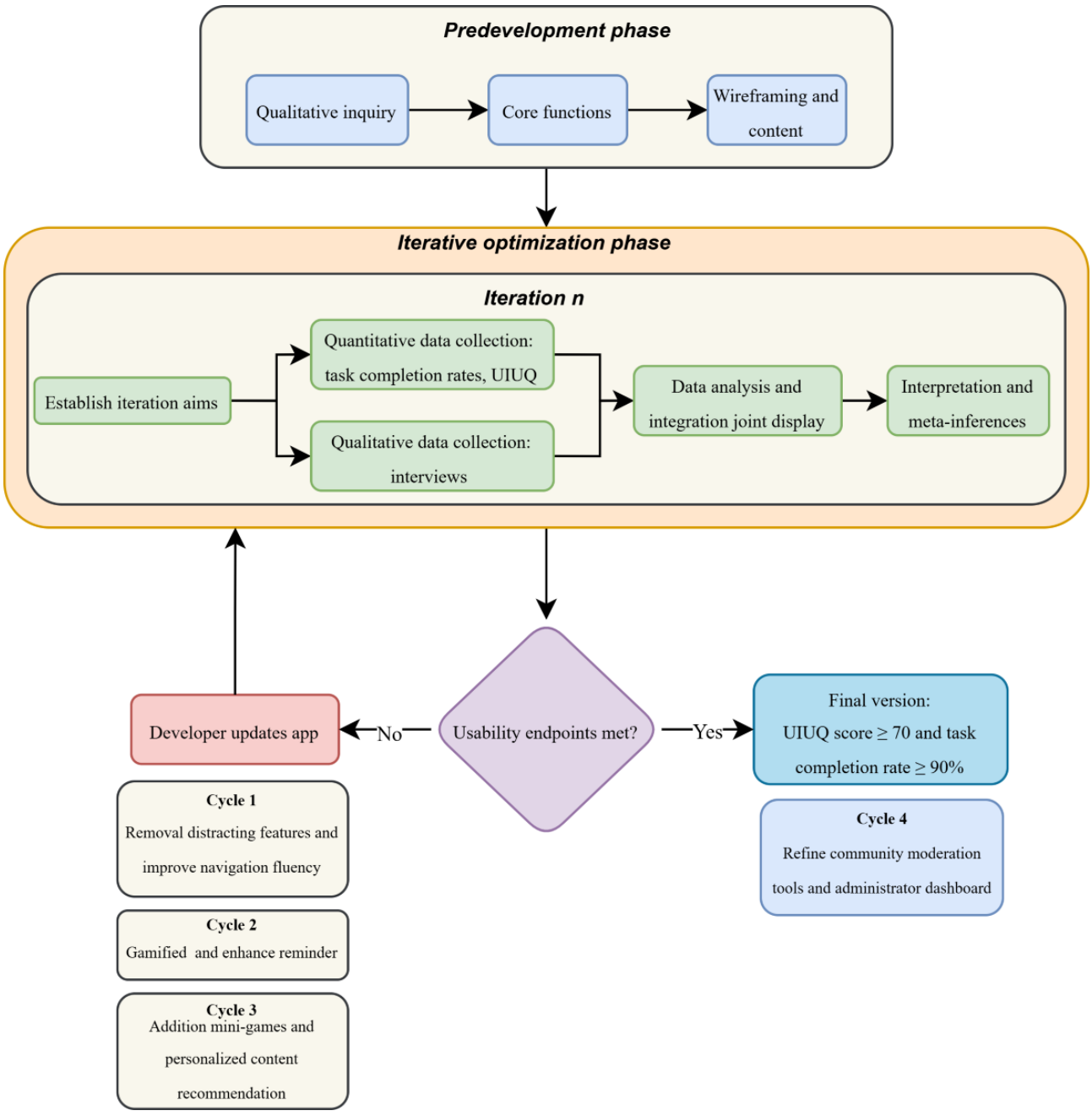
XSL•FO
RenderX

learning scripts, designed based on behavior change theory [14] and integrated with the core text and image materials.

Iterative Optimization Phase

We conducted 4 complete iterative cycles of testing and refinement between November 2024 and March 2025. Each cycle followed the structured process outlined in Figure 1 and incorporated the integration procedures detailed below. Representative screenshots of key interface updates from each cycle (Cycle 1-Cycle 4) are provided in Multimedia Appendix 2.

Figure 1. Flowchart of the 4-cycle iterative convergent mixed methods optimization process for Tpro development. UIUQ: User Interface Usability Questionnaire.



Integration in the Research Questions Dimension

The study used 3 types of research questions that guided each iteration:

(1) Mixed methods aim: to illustrate, explore, and measure how to improve the usability and functionality of Tpro for DVT symptom management

(2) Quantitative research questions: focused on measuring effectiveness (task completion rates) and satisfaction (User Interface Usability Questionnaire [UIUQ] scores) [15]. Task completion rates were defined as the proportion of participants who successfully completed both the symptom reporting process and at least one health education activity (knowledge reading or quiz completion) within the Tpro app. The UIUQ developed by Chinese scholar Liu consists of 20 items scored on a 0 - 5 scale, yielding a total score of 0 - 100 points. Higher scores

XSL•FO
RenderX

indicate better interface usability. Based on the total score: below 50 points represents unacceptable usability, 50 - 70 points indicates marginally acceptable usability, and above 70 points reflects satisfactory usability.

(3) Qualitative research questions: centered on understanding how and why users interacted with Tpro as they did, their experiences with specific features, and suggestions for improvement. Table 1 summarizes the data collection methods and integration timing used in this study.

Table . Data collection methods and integration timing. This table summarizes the mixed methods data collection strategy, specifies the types of data collected, and outlines how they were integrated concurrently to inform real-time design decisions during each development cycle.

Data type	Methods	Collection timing	Integration approach
Quantitative	UIUQ[a], task completion rates	During and immediately after app use	Concurrent with qualitative data collection
Qualitative	Semistructured interviews	During app use and in postuse debriefing	Immediate feedback into quantitative understanding

[a]UIUQ: User Interface Usability Questionnaire.

Integration in the Data Collection Dimension

We implemented 3 integration strategies during data collection: (1) matching: qualitative questions were intentionally matched to quantitative constructs. For example, when low task completion rates were observed quantitatively, we asked, "I noticed you hesitated when completing the task, can you tell me more about what was unclear?" (2) Diffracting: qualitative questions explored aspects beyond quantitative measures, such as emotional responses to the gamified interface, trust in medical recommendations, and preferences for community interaction styles. And (3) expanding: when quantitative and qualitative findings diverged, we expanded questioning to understand both perspectives. For instance, when users reported high satisfaction qualitatively but gave medium UIUQ scores quantitatively, we explored this discrepancy specifically.

Collection of Qualitative Data and Researcher Reflexivity

Qualitative data within each cycle were collected primarily via semistructured interviews, conducted by 2 researchers (MS and YL). Both interviewers held clinical research backgrounds and had completed formal training in qualitative interview techniques. They had no prior clinical or therapeutic relationship with the patient participants, nor had they engaged in any discussions with participants regarding the Tpro app or DVT management prior to the study, which was explicitly explained at the outset to minimize social desirability bias and encourage candid feedback. Researchers maintained reflexive journals to document and bracket their own assumptions prior to and following interviews. Data collection proceeded within each cycle until informational saturation was attained on the usability themes specific to that cycle's prototype, which was determined when subsequent interviews yielded no new insights related to the core evaluation objectives.

Integration in the Data Analysis Dimension

We used a combined independent and interactive analysis strategy:

(1) Interactive analysis: during early iterations (Cycles 1 - 2), researchers analyzed qualitative and quantitative findings in real-time as data were collected, allowing immediate exploration of emerging issues

(2) Independent intramethod analysis: in later iterations (Cycles 3 - 4), qualitative data underwent inductive thematic analysis and quantitative data were analyzed statistically prior to integration. Specifically, for thematic analysis, 2 researchers (QL and LW) independently coded transcripts from initial cycles to develop a preliminary codebook. Discrepancies were resolved through discussion until consensus was reached, thereby enhancing interpretive reliability. The refined codebook was then applied to subsequent cycles, with ongoing peer debriefing within the research team to ensure analytical rigor. Representative quotations illustrating key themes are provided in Multimedia Appendix 3.

(3) Data integration procedures: we used joint displays to visually represent the integration of qualitative and quantitative findings (see Table 2 for example). The fit of the 2 data types was examined for confirmation, complementarity, expansion, or discordance. Discordant findings were resolved through additional data collection or reanalysis.

XSL•FO
RenderX

Table . Joint display from Cycle 1 showing data integration and resulting design decisions. This table provides a concrete example of the convergent mixed methods analysis, showing how specific usability metrics were interpreted alongside direct user feedback to identify root causes and translate them into actionable design modifications.

Quantitative findings	Qualitative findings	Integration insights	Design decisions
Around 52% (11/21) of users failed to complete symptom reporting or knowledge learning first attempt	• "I got distracted by the chat function on main screen and forgot to report symptoms" (Patient 03) • "app is super convenient, but I don't really have time to learn about thrombosis. I just check it out from time to time" (Patient 10) • "The knowledge base still has some room for improvement. There are so many topics I want to learn about! It would be great to have more educational mini-games" (Patient 11, Patient 08)	Core functionality was being undermined by secondary features and cognitive load	• Removed chat function from main screen to dedicated section • Added interactive mini-games and expanded knowledge content • Enhanced personalized reminder system based on usage patterns
UIUQ[a] score for efficiency (86/100)	• "I wasn't sure if my report went through-no confirmation" (Patient 07) • "The navigation feels jumpy between sections" (Patient 05)	Lack of feedback and inconsistent navigation patterns created uncertainty and frustration	• Added clear submission confirmations • Standardized navigation patterns across all modules • Improved transition animations

[a]UIUQ: User Interface Usability Questionnaire.

Integration in the Interpretation Dimension

During each iteration cycle, the research team conducted interpretation sessions where quantitative and qualitative findings were merged to draw metainferences that considered both types of data. These sessions followed a structured process:

(1) Data triangulation: comparing quantitative metrics with qualitative themes

(2) Pattern identification: looking for convergent, complementary, or divergent findings

(3) Priority setting: determining which issues to address in the next iteration based on severity and frequency

(4) Solution generation: brainstorming specific design modifications

Participant Recruitment

Participant recruitment used a combination of convenience and purposive sampling strategies to engage key stakeholder groups. Patient participants were recruited via convenience sampling from the vascular surgery outpatient clinic of Peking Union Medical College Hospital. Inclusion criteria for patients required a confirmed DVT diagnosis, proficiency in smartphone use, and expressed willingness to engage in multiple testing cycles. Exclusion criteria encompassed significant cognitive impairment, visual disabilities preventing effective smartphone interaction, or concurrent participation in other digital health studies. Clinical experts, including vascular surgeons and specialized nurses, were purposively recruited from the Department of Vascular Surgery at Peking Union Medical College Hospital based on their documented clinical expertise and experience in DVT management. Similarly, nonmedical technical evaluators with specialized backgrounds in usability engineering, software development, or human-computer interaction were purposively selected from the collaborating technical team involved in this research project.

Sample Size Justification

The sample size for iterative usability testing was determined a priori based on established principles of user-centered design and the iterative convergent design framework adopted for this study. Foundational usability research indicates that testing with 5-12 users per iteration is sufficient to identify the majority of interface problems [16]. Aligned with this framework [10], the primary endpoint of the iterative process was not statistical power but the achievement of thematic and design saturation, the point at which additional testing cycles yield no new critical usability themes, barriers, or actionable design feedback. Recruitment and iterative testing continued until this saturation criterion was operationally confirmed, which was assessed through the stabilization of quantitative usability metrics (task completion rate and UIUQ score) in consecutive cycles.

Ethical Considerations

The study received approval from the Institutional Review Board of Peking Union Medical College Hospital (approval number: I-24PJ1204). Written informed consent was obtained from all participants prior to their enrollment in the study. We attest that the privacy and confidentiality of all research subjects' data and identities were strictly maintained throughout the study.

Development Technical Specifications

Tpro is constructed using the Flutter (Google LLC) framework, enabling simultaneous app development for Android, iOS, Web, and desktop (Windows, macOS, and Linux), thereby

XSL•FO
RenderX

streamlining code writing across platforms, shortening project timelines, and reducing costs. Embedded pages are built on a Java-based low-code platform. The Tpro app is compatible with Android 5 and higher, as well as iOS 11 and newer versions. Initially developed in Chinese to align with regional preferences, Tpro will be offered in English and other languages in the future.

During the iterative testing cycles, the Tpro package was installed on participants' personal devices via a secure, authenticated download link, or was made accessible on preconfigured study devices where necessary. Access required individual user accounts verified by the clinical research team. This deployment strategy ensured controlled distribution and facilitated rapid iterative updates during the optimization cycles.

Data Security and Privacy Protection

The Tpro system was designed with a multilayered security and privacy architecture. All data transmission was protected by TLS encryption, and sensitive data were encrypted at rest. The app backend was deployed on a secure, internal server to ensure physical and administrative data control. Access to patient information was strictly governed by role-based permissions within the clinician admin portal, and all accesses were logged. In the peer community, user identities were anonymized. These measures were designed to comply with relevant data protection regulations and were reviewed as part of the ethical approval process.

Results

Participant Characteristics and Iterative Testing Process

The iterative development of Tpro was conducted over 4 cycles spanning a 4-month period, engaging a total of 21 participants. The patient cohort consisted of 12 individuals with confirmed DVT (8 males and 4 females), with an age range of 25-72 years, all recruited from the vascular surgery outpatient clinic. Their educational backgrounds ranged from elementary school to master's degree, and they presented with diverse DVT locations (including iliofemoral, calf, and femoral veins) and varying symptom severities. The multidisciplinary evaluation team comprised 6 clinical experts: 2 vascular surgeons (mean 10.5, SD 6.4 years of specialization) and 4 specialized nurses (mean 6, SD 1.4 years of experience). Additionally, the team included 3 nonmedical technical evaluators specializing in usability research and software engineering (mean 8, SD 3.6 years of relevant experience).

Engagement across the iterative cycles was structured to balance depth of longitudinal feedback with the introduction of fresh perspectives, continuing until thematic and data saturation were achieved. All 6 clinical experts and all 3 technical evaluators participated in all 4 cycles, providing consistent expert and technical oversight. Among the 12 patients, 4 completed all 4 cycles, forming a core longitudinal feedback group, while the remaining 8 patients participated in 2-3 cycles, ensuring that new user perspectives were incorporated at each stage. Consequently, a total of 13 participants (4 patients, 6 clinical experts, and 3 technical evaluators) completed all 4 testing cycles, thereby providing the continuous and in-depth feedback necessary to reach this saturation endpoint. The comprehensive demographic, clinical, professional, and cycle-by-cycle participation details for each individual are provided in Multimedia Appendix 4.

Evolution of Usability Metrics

The systematic application of the iterative convergent approach yielded substantial improvements across all usability metrics. The UIUQ score demonstrated progression from an initial score of 86 in the first cycle to a final score of 89 in the fourth cycle. Task completion rates showed consistent improvement, rising from 46% (6/13) in initial testing to 92% (12/13) in the final evaluation (Table 3).

Table . Usability metrics evolution across 4 iterative cycles (n=13). This table presents the progression of key quantitative usability indicators, including the overall User Interface Usability Questionnaire (UIUQ) score and the task completion rate, demonstrating the systematic improvement achieved through the iterative refinement process.

Metric	Cycle 1	Cycle 2	Cycle 3	Cycle 4
Overall UIUQ[a] score, mean (SD)	86 (6)	87 (6)	88 (8)	89 (13)
Task completion rate, n (%)	6 (46)	10 (77)	11 (85)	12 (92)

[a]UIUQ: User Interface Usability Questionnaire.

Key Refinements Through Integrated Data Analysis

The integrated analysis throughout the development cycles revealed a consistent pattern of strong initial usability with progressive optimization in user engagement. From the first cycle, quantitative data showed excellent UIUQ scores (mean 86, SD 6), while qualitative feedback confirmed the interface was perceived as "intuitive and pleasant." However, the convergent analysis uncovered a critical gap between this high usability perception and actual engagement metrics, with only 46% (6/13) task completion rate initially. This paradox was elucidated through user reports of forgetting to complete tasks despite appreciating the design, leading to the insight that excellent usability alone was insufficient to ensure consistent engagement. Consequently, we implemented a gamified task sequence with progressive rewards and streamlined the workflow between symptom reporting and educational content.

During the second cycle, as task completion rates rose substantially to 77% (10/13) while maintaining high UIUQ scores (mean 87, SD 6), qualitative feedback indicated that the improved task flow made it easier for users to remember subsequent actions. This convergence of metrics confirmed that structural enhancements had successfully translated the strong

XSL•FO
RenderX

usability foundation into improved engagement. Building on this success, we enhanced the reminder system with contextual prompts and personalized notification timing. In the subsequent cycles 3‐4, further improvements in task completion (12/13, 92%) with stable UIUQ scores were reinforced by qualitative reports of increased enjoyment from educational games and anticipation of weekly content discoveries. This integrated understanding highlighted the importance of continuous content innovation, prompting the introduction of rotating educational mini-games and personalized knowledge recommendations based on symptom patterns to sustain long-term engagement.

Final App Architecture

The iterative development process culminated in a refined app comprising well-integrated functional modules that collectively support comprehensive DVT management. Figure 2 shows the final integrated architecture including the user onboarding, gamified reporting map, dynamic education ecosystem, and communication channels. The patient registration system evolved into a streamlined onboarding process where users complete demographic and clinical profiles after administrative verification, capturing essential information including DVT location, existing symptoms, and relevant medical history through an intuitive form-based interface.

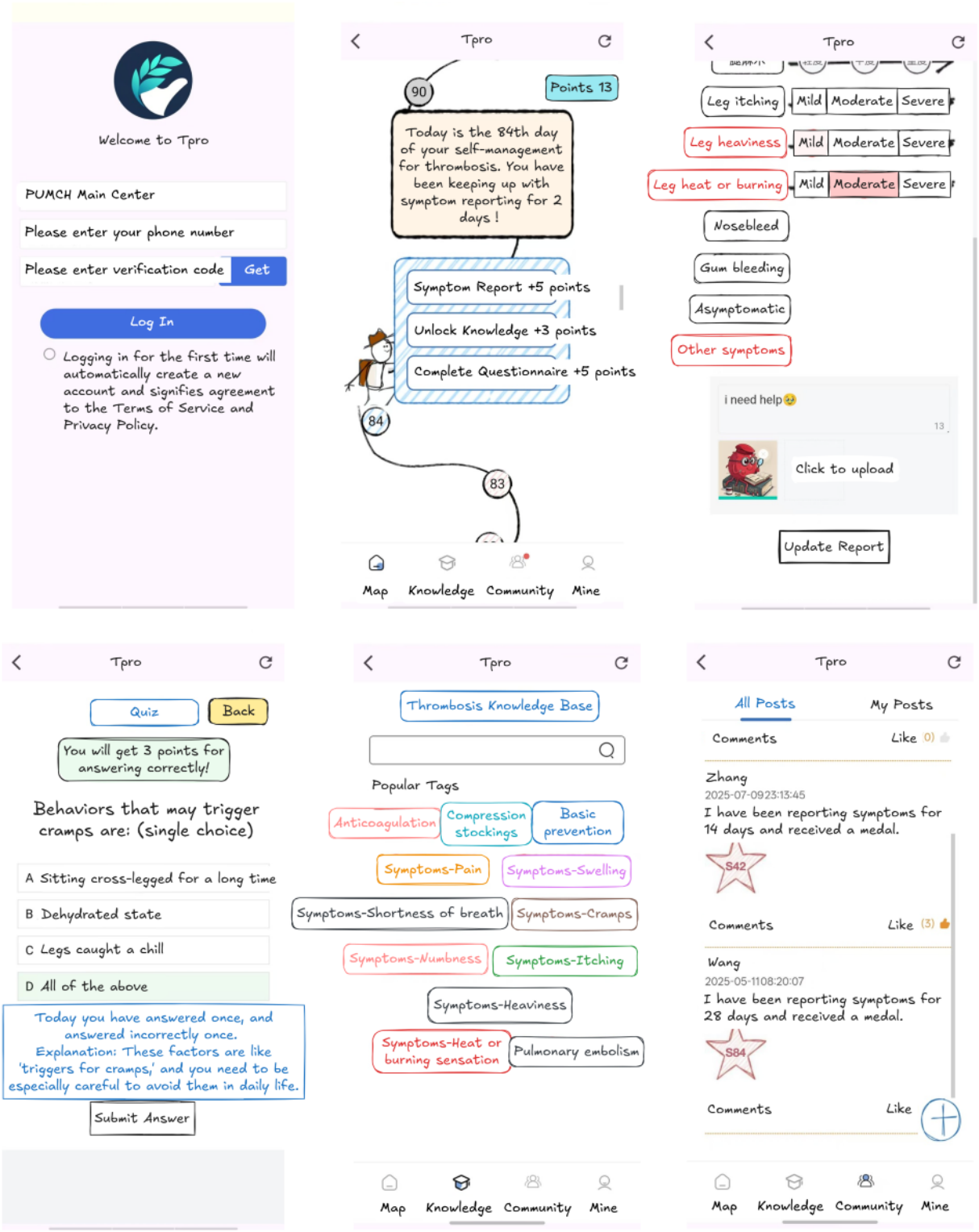
Central to the app is the gamified symptom monitoring module. Patients are guided through a structured daily workflow, initiated by an adaptive reminder system that sends a notification at a default time of 5:00 PM, which can be customized to individual preferences. Upon entering the app, patients complete their symptom report through an “Exploration Map” interface, which transforms daily symptom reporting into an engaging, game-like activity. A carefully designed incentive mechanism awards points for consistent reporting, which accumulate toward priority scheduling for in-person follow-up appointments. For clinical support, the system features an integrated clinician-patient communication channel that enables patients to submit detailed queries with text and photographic evidence directly through the symptom reporting interface. This module operates with a sophisticated triage protocol where acute symptoms such as dyspnea trigger immediate alerts for urgent clinical response, while chronic concerns receive guaranteed feedback within a 48-hour framework.

The educational component underwent significant transformation from static information delivery to a dynamic learning ecosystem. Upon submission of each symptom report, health education content specifically matched to the symptoms reported that day is automatically unlocked and delivered through multiple formats, including illustrated materials, instructional videos, interactive quizzes, and gamified learning activities, with each educational interaction further contributing to the user’s point accumulation. At the final testing stage, the module contained 3 instructional videos, 53 illustrated materials, 292 interactive knowledge quiz items, and 45 gamified learning activities, which together constituted a content library to support rotated delivery. Based on user feedback collected during the testing phase indicating a need for continuously refreshed educational content, the clinical team will review user needs on a monthly basis and develop and add new materials accordingly. This content rotation and expansion strategy is designed to sustain long-term user engagement beyond the testing phase. For patients with a strong interest in learning, the comprehensive knowledge library remains accessible for self-directed exploration at any time, independent of the daily reporting workflow.

Complementing these core functions, the app incorporates a dedicated peer support community where patients share treatment experiences and disease management insights through text and image posts at any time. This social platform fosters mutual support through interactive features including likes and comments, while maintaining a therapeutic environment through active moderation by health care professionals.

XSL•FO
RenderX

Figure 2. Final integrated system architecture of Tpro (English translations). The architecture comprises 4 interconnected modules developed through iterative refinement: a gamified "Exploration Map" for symptom reporting, a dynamic educational ecosystem that unlocks content based on user actions, a tiered clinician-patient communication channel with acute symptom alerts, and a professionally moderated peer support community. This design successfully integrates a high-usability interface with patient-centered functional requirements for comprehensive deep vein thrombosis (DVT) management.



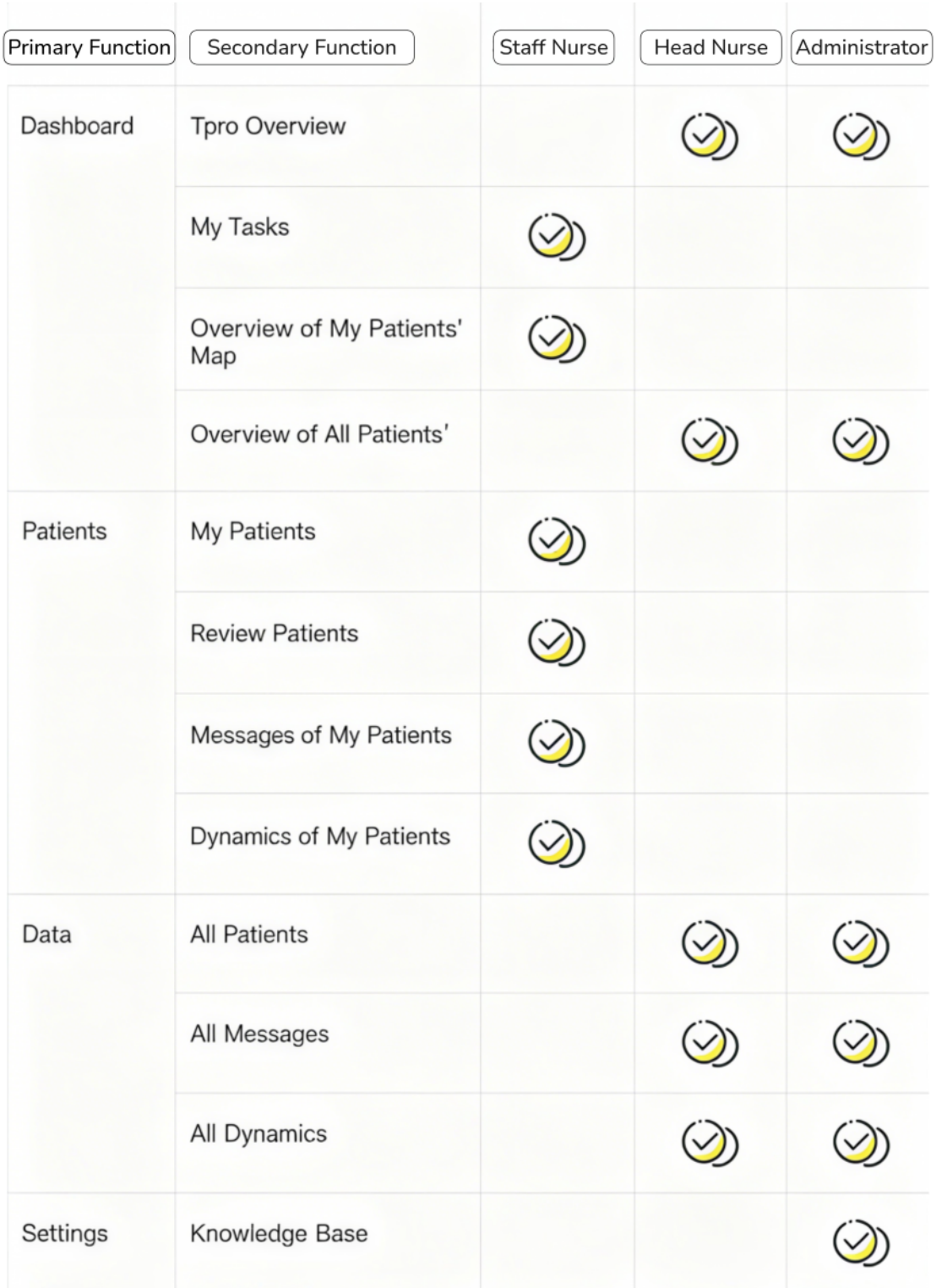
Administrative System Development

The backend administrative portal similarly evolved through the iterative refinement process, developing enhanced role-based permissions with granular access controls, automated flagging systems for urgent patient communications, advanced analytics dashboards for tracking patient engagement patterns, dynamic content management capabilities for educational material updates, and integrated moderation tools for community management. This comprehensive administrative infrastructure supported efficient oversight while maintaining flexibility for institutional customization. Figure 3 presents the admin portal

XSL•FO
RenderX

interface with its analytics dashboard, content management, and moderation tools.

Figure 3. Overview of the backend administrative portal for Tpro. The portal provides role-based access controls, an analytics dashboard for tracking patient engagement, dynamic content management tools, and community moderation features. This comprehensive administrative infrastructure enables efficient patient oversight and supports institutional customization.



Final Usability Outcomes

The final evaluation demonstrated outstanding usability outcomes across all metrics. The system achieved an excellent UIUQ score with a mean of 89 (SD 13) in the fourth cycle, and the task completion rate, defined as completing both symptom reporting and at least one health education activity, reached 92% (12/13), confirming high operational reliability and indicating that nearly all patients proceeded from symptom reporting to the automatically unlocked educational content, which was delivered in randomly presented formats, including illustrated materials, videos, quizzes, and games. Regarding

XSL•FO
RenderX

clinician-patient communication, 12 patients generated a total of 6 inquiries across the testing period, all submitted through the symptom reporting interface. In the peer support community, 4 posts were shared across the 12 patients, while several others browsed content without actively posting. Notably, 83% (10/12) of patient participants expressed willingness to continue using Tpro and recommend it to peers, and clinical doctors and nurses consistently provided highly positive evaluations regarding the system's clinical utility and workflow integration.

These convergent findings from multiple stakeholder perspectives substantiate that Tpro has evolved into a clinically valuable, highly usable, and widely accepted digital health intervention, positioning it as a promising tool for optimizing out-of-hospital management of DVT.

Discussion

Principal Findings

This study successfully demonstrates the application of an iterative convergent mixed methods design in developing Tpro, a comprehensive mHealth solution for DVT symptom management. The systematic integration of quantitative usability metrics with qualitative user feedback across multiple development cycles enabled the transformation of an initial prototype into a refined digital health intervention. The final app not only achieved excellent usability, as evidenced by a high UIUQ score (mean 89, SD 13) and task completion rate (12/13, 92%), but also received endorsement from clinical experts. More importantly, this methodology facilitated a deeper understanding of the complex interplay between technical functionality, user engagement, and clinical utility that would likely remain obscured through traditional single-method approaches.

Comparison With Prior Work

From a design philosophy perspective, the construction of Tpro based on PRO represents an essential reshaping of the medical paradigm. Traditional clinical research has predominantly focused on objective indicators from the medical perspective. For instance, Tritschler et al's [17] scoping review on clinical outcomes of venous thromboembolism interventions in nonpregnant adults identified recurrence, major bleeding, and mortality as the most frequently reported outcome measures. However, these metrics often fail to accurately capture patients' real-life challenges, such as prolonged immobilization, anxiety, and depression induced by anticoagulant therapy, or the inconvenience of daily activities caused by wearing compression stockings [18,19]. In contrast, PRO directly reflects patients' physical, psychological, and social functional experiences [20]. By collecting first-hand information through PRO, Tpro bridges the critical gap in physician-patient communication, precisely identifies patient needs, and enables targeted subsequent interventions.

The design of each core module in Tpro was meticulously informed by this PRO-driven philosophy and the specific needs uncovered in our preliminary research [11]. The symptom monitoring module uses an interactive exploration map format, while the daily reporting feature adopts a gamified approach akin to unlocking new levels. This innovative design strategically leverages game-based mechanics to transform traditionally monotonous health management tasks into engaging daily activities, fostering patient motivation through an integrated incentive system. Underlying this framework is the application of behavior change theory, which posits that patients are significantly more likely to adhere to healthy behaviors when they perceive autonomy, competence, and relatedness [14]. The multimodal health education component incorporates graphics, videos, interactive quizzes, and gamified elements to accommodate diverse learning preferences, thereby translating complex medical knowledge into easy-to-understand and easy-to-use resources [21,22].

Furthermore, Tpro actively fosters social support networks, a critical component often neglected in traditional care. The integrated clinician-patient communication channel provides a "lifeline" for acute concerns and establishes reliable, timely feedback for chronic issues, offering patients psychological security [23]. The peer support community, moderated by health care professionals, facilitates emotional resonance and the sharing of practical lived experiences, which can enhance disease-related resilience and mitigate feelings of isolation [24].

Methodological Contributions

This study advances digital health development methodology by addressing documented limitations in conventional approaches, particularly the high abandonment rates of health apps due to poor usability [9]. Our iterative convergent mixed methods framework enabled proactive identification and resolution of usability barriers through continuous refinement based on rich, contextualized user understanding. The successful translation of user needs into design features, particularly the gamified symptom reporting and tiered communication system, aligns with evidence supporting behaviorally-informed design in digital health [9,25,26]. Beyond confirming what works, our approach systematically leverages quantitative-qualitative integration to reveal why and how features succeed across diverse users.

Strengths

This study boasts several notable strengths. First, the adoption of an iterative convergent mixed methods design represents a rigorous and transparent methodology for mHealth development. This approach allowed for the simultaneous collection and integration of rich qualitative insights with quantitative usability metrics, enabling us to not only identify what usability issues existed but, more importantly, to understand why they occurred and how to effectively address them. Second, the development process was deeply rooted in a patient-centered philosophy, from the initial qualitative inquiry that defined the core functions to the iterative testing with end users (both patients and clinicians). This ensured that the final Tpro app was not only highly usable but also directly addressed the authentic needs and challenges faced by individuals managing DVT in their daily lives. Finally, the study delivers a fully functional and rigorously refined digital health intervention with demonstrated high usability and acceptance, which is now primed for subsequent efficacy testing, thereby providing a tangible and significant contribution to the field.

XSL•FO
RenderX

Limitations and Challenges

Several limitations should be considered when interpreting the findings of this study. First, the generalizability of the results is constrained by the sample characteristics. All participants, including the 12 patient participants, were recruited from a single, high-resource academic medical center. Although the sample size and setting were appropriate for achieving usability saturation within the iterative convergent design framework, which was the primary aim of this development phase, they may limit the direct transferability of the findings to other contexts, such as community hospitals or populations with differing socioeconomic backgrounds, health literacy, or digital access. Second, individuals without smartphones or with significant visual or cognitive impairments were excluded, meaning accessibility barriers for these vulnerable groups remain unaddressed. Third, this version of Tpro is available only in Chinese, which necessitates thorough cultural and linguistic adaptation for broader global implementation. Fourth, engagement with the peer support community was relatively modest during the testing phase, with only 4 posts generated across the 12 patient participants. This may reflect the limited testing duration, the small user group, or the acute nature of DVT, where patients may prioritize symptom management and education over social interaction in the initial phase of self-management. The community was designed as one component of a multifaceted intervention, and the observed pattern of intermittent, need-based use may represent an appropriate level of engagement for this population. Finally, while this study demonstrates high usability and acceptance, the definitive clinical efficacy of Tpro on outcomes such as symptom burden, quality of life, and DVT recurrence requires validation through a future randomized controlled trial (RCT).

Security Considerations

The integration of foundational data security and privacy protections, including end-to-end encryption, internal server deployment, and role-based access controls, was essential for Tpro as an mHealth tool handling sensitive patient information. These safeguards were designed to protect confidentiality, build user trust, and meet core data protection requirements. For future implementations in other settings, such controls must be adapted to local regulatory frameworks. A focus on privacy-by-design addressed a key barrier to patient adoption and supported sustainable clinical integration of digital health tools.

Implications for Practice and Research

This study presents significant implications for both mHealth development and future research. For developers and health care organizations, the iterative convergent mixed methods approach offers a replicable framework for creating evidence-based digital health solutions that effectively balance technical sophistication with user-centered design principles. The specific design strategies developed through this process, particularly the gamification elements and tiered communication system, provide valuable reference models for other chronic disease management apps. For researchers, this methodology demonstrates how rigorous mixed methods can substantially enhance intervention development quality and impact.

Future research should prioritize a multicenter RCT to establish the clinical efficacy of Tpro. A proposed design for such a trial includes the following key components:

(1) Design and participants: a pragmatic, 2-arm, parallel-group RCT comparing Tpro plus usual care versus usual care alone. We plan to recruit approximately 200‐250 patients with confirmed DVT from multiple clinical sites, stratified by factors such as initial clot location and risk profile.

(2) Intervention: the intervention group will receive access to the Tpro app alongside standard clinical management. Usual care will follow current local guidelines without the app.

(3) Primary outcomes: the primary endpoints will focus on patient-centered outcomes, including symptom burden and health-related quality of life, assessed at 3 and 6 months postenrollment.

(4) Secondary outcomes: key secondary outcomes will include objectively confirmed DVT recurrence rates, health care use, adherence to anticoagulant therapy, and sustained engagement metrics within the app.

(5) Analysis: the primary analysis will follow the intention-to-treat principle. This definitive efficacy trial will be essential to translate the demonstrated usability of Tpro into evidence for clinical and patient-centered benefit. Beyond this RCT, subsequent research directions remain important, integrating artificial intelligence for personalized symptom management; validating the system across diverse health care systems and cultural contexts to enhance generalizability; exploring strategies to foster more active peer interaction within the community module; and conducting longitudinal studies to examine the relationship between sustained engagement and long-term health outcomes.

Conclusions

The iterative convergent mixed methods design enabled the successful development of Tpro, a PRO-based mobile app for DVT symptom management, yielding a highly usable and clinically relevant digital health intervention. Through 4 iterative cycles integrating mixed methods research, we resolved critical usability issues while gaining profound insights into patient engagement patterns. The resulting app provides comprehensive symptom monitoring, education, communication, and peer support within an evidence-based framework, now ready for formal efficacy testing. This study validates a rigorous methodology for developing digital health solutions that effectively balance technical sophistication with human-centered design, offering significant potential to enhance out-of-hospital DVT care while establishing a replicable approach for creating meaningful solutions for patients and clinicians.

XSL•FO
RenderX

Acknowledgments

We acknowledge the software engineers involved in the design process of the Tpro app. We used the generative AI tool DeepSeek solely during the language polishing and proofreading stages. After using it, we conducted the necessary review and modifications to the content.

Funding

The authors disclose receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the CAMS innovation Fund for Medical Sciences (CIFMS) (2023-I2M-C&T-B-034).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: QL, YL, MS, LW. Methodology: QL, LW. Data curation/data collection: MS, YL. Formal analysis: QL. Investigation: MS, YL. Writing-original draft: QL. Writing-review & editing: QL, YL, MS, LW.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Wireframe diagram of the Tpro app (English translations). This wireframe was generated during the predevelopment phase to establish the initial structural layout of the 4 core modules. It illustrates the spatial hierarchy, modular organization, and the intended user flow among symptom monitoring, health education, clinician-patient communication, and peer support community, providing the foundational visual scaffold for subsequent iterative refinement.
[PNG File, 306 KB - humanfactors_v13i1e92738_app1.png]

Multimedia Appendix 2
Representative screenshots of the Tpro app across 4 iterative testing cycles (English translations).
[PNG File, 2389 KB - humanfactors_v13i1e92738_app2.png]

Multimedia Appendix 3
Representative quotations from qualitative interviews across iterative cycles. This material supported the qualitative findings presented in the manuscript. It provided authentic participant quotes that illustrated key themes related to usability, engagement, educational content, and communication, enhancing the transparency and richness of the qualitative analysis.
[DOCX File, 15 KB - humanfactors_v13i1e92738_app3.docx]

Multimedia Appendix 4
Characteristics of individual participants in the iterative testing. This table provided deidentified, participant-level data for the 21 individuals involved in the study, including patients, clinical experts, and technical evaluators. It contained demographics, professional details, and specific cycle-by-cycle participation.
[DOCX File, 21 KB - humanfactors_v13i1e92738_app4.docx]

References

1. Malhotra K, Bawa A, Goyal K, Wander GS. Global impact of deep vein thrombosis awareness month: challenges and future recommendations. Eur Heart J 2022 Sep 21;43(36):3379-3381. [doi: 10.1093/eurheartj/ehac252] [Medline: 35574723]
2. Gwozdz AM, de Jong CMM, Fialho LS, et al. Development of an international standard set of outcome measures for patients with venous thromboembolism: an International Consortium for Health Outcomes Measurement consensus recommendation. Lancet Haematol 2022 Sep;9(9):e698-e706. [doi: 10.1016/S2352-3026(22)00215-0] [Medline: 36055334]
3. Abuzied Y, Deeb A, AlAnizy L, Al-Amer R, AlSheef M. Improving venous thromboembolism prophylaxis through service integration, policy enhancement, and health informatics. Glob J Qual Saf Healthc 2024 Feb;7(1):22-27. [doi: 10.36401/JQSH-23-16] [Medline: 38406656]
4. Dai W, Feng W, Zhang Y, et al. Patient-reported outcome-based symptom management versus usual care after lung cancer surgery: a multicenter randomized controlled trial. J Clin Oncol 2022 Mar 20;40(9):988-996. [doi: 10.1200/JCO.21.01344] [Medline: 34995100]

XSL•FO
RenderX

5. Snyder DJ, Zilinyi RS, Cohen DJ, Parikh SA, Sethi SS. Patient-reported outcomes in venous thromboembolism: a systematic review of the literature, current challenges, and ways forward. J Am Heart Assoc 2023 Dec 5;12(23):e032146. [doi: 10.1161/JAHA.123.032146] [Medline: 38014656]
6. Cracchiolo JR, Arafat W, Atreja A, et al. Getting ready for real‐world use of electronic patient‐reported outcomes (ePROs) for patients with cancer: a national comprehensive cancer network ePRO workgroup paper. Cancer 2023 Aug 15;129(16):2441-2449. [doi: 10.1002/cncr.34844] [Medline: 37224181]
7. Boulanger MC, Petrillo LA, Temel JS. Listen to the patient: integrating patient-reported outcomes into clinical care. J Natl Cancer Inst 2023 Dec 6;115(12):1451-1453. [doi: 10.1093/jnci/djad174] [Medline: 37740689]
8. Lee YC, Lu F, Colls J, et al. Outcomes of a mobile app to monitor patient-reported outcomes in rheumatoid arthritis: a randomized controlled trial. Arthritis Rheumatol 2021 Aug;73(8):1421-1429. [doi: 10.1002/art.41686] [Medline: 33559338]
9. Lowe C, Hanuman Sing H, Browne M, Alwashmi MF, Marsh W, Morrissey D. Usability testing of a digital assessment routing tool: protocol for an iterative convergent mixed methods study. JMIR Res Protoc 2021 May 18;10(5):e27205. [doi: 10.2196/27205] [Medline: 34003135]
10. Alwashmi MF, Hawboldt J, Davis E, Fetters MD. The iterative convergent design for mobile health usability testing: mixed methods approach. JMIR Mhealth Uhealth 2019 Apr 26;7(4):e11656. [doi: 10.2196/11656] [Medline: 31025951]
11. Xu X, Wang Y, Liu X, Wang L. Qualitative study on poor compliance with self-management behaviors in patients with deep vein thrombosis of lower extremities. Chin J Mod Nurs 2020;26(20):2785-2788. [doi: 10.3760/cma.j.cn115682-20191104-03985]
12. Salimzadeh Z, Damanabi S, Ferdousi R, Shaafi S, Kalankesh LR. A mobile app (IDoThis) for multiple sclerosis self-management: development and initial evaluation. BMC Med Inform Decis Mak 2022 Dec 13;22(1):328. [doi: 10.1186/s12911-022-02078-z] [Medline: 36514043]
13. Zheng Y, Wang L, Chen Y. The Wonderful Journey of Blood Vessels: Popular Science Press; 2021:105-124.
14. Gillison FB, Rouse P, Standage M, Sebire SJ, Ryan RM. A meta-analysis of techniques to promote motivation for health behaviour change from a self-determination theory perspective. Health Psychol Rev 2019 Mar;13(1):110-130. [doi: 10.1080/17437199.2018.1534071] [Medline: 30295176]
15. Liu K. Research on usability of human-computer interaction interface based on user experience metric. : Hefei University of Technology; 2018 URL: https://link.oversea.cnki.net/doi/10.27101/d.cnki.ghfgu.2018.000149 [accessed 2026-06-29]
16. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. 1993 Presented at: The SIGCHI conference; Amsterdam, The Netherlands p. 206-213 URL: http://portal.acm.org/citation.cfm?doid=169059 [accessed 2026-06-16] [doi: 10.1145/169059.169166]
17. Tritschler T, Cusano E, Langlois N, et al. Identification of outcomes in clinical studies of interventions for venous thromboembolism in non-pregnant adults: a scoping review. J Thromb Haemost 2022 Oct;20(10):2313-2322. [doi: 10.1111/jth.15787] [Medline: 35717670]
18. Genge L, Krala A, Tritschler T, et al. Evaluation of patients' experience and related qualitative outcomes in venous thromboembolism: a scoping review. J Thromb Haemost 2022 Oct;20(10):2323-2341. [doi: 10.1111/jth.15788] [Medline: 35730241]
19. Kankam HKN, Lim CS, Fiorentino F, Davies AH, Gohel MS. A summation analysis of compliance and complications of compression hosiery for patients with chronic venous disease or post-thrombotic syndrome. Eur J Vasc Endovasc Surg 2018 Mar;55(3):406-416. [doi: 10.1016/j.ejvs.2017.11.025] [Medline: 29329662]
20. de Jong CMM, de Wit K, Black SA, et al. Use of patient-reported outcome measures in patients with venous thromboembolism: communication from the ISTH SSC subcommittee on predictive and diagnostic variables in thrombotic disease. J Thromb Haemost 2023 Oct;21(10):2953-2962. [doi: 10.1016/j.jtha.2023.06.023] [Medline: 37394119]
21. Morgado M, Botelho J, Machado V, Mendes JJ, Adesope O, Proença L. Full title: video-based approaches in health education: a systematic review and meta-analysis. Sci Rep 2024 Oct 10;14(1):23651. [doi: 10.1038/s41598-024-73671-7] [Medline: 39384592]
22. Yıldız M, Yildiz M, Kayacık AD. Rising gamification in health education: a bibliometric study. Nurse Educ Pract 2024 Jul;78:103993. [doi: 10.1016/j.nepr.2024.103993] [Medline: 38788617]
23. Dai Y, Min H, Sun L, Wang X, Zhu C, Gu C. Assessing women's and health professionals' views on developing a midwifery-led mobile health app intervention in pregnancy: a descriptive qualitative study. J Adv Nurs 2024 Oct;80(10):4259-4271. [doi: 10.1111/jan.16086] [Medline: 38332497]
24. Johansson V, Islind AS, Lindroth T, Angenete E, Gellerstedt M. Online communities as a driver for patient empowerment: systematic review. J Med Internet Res 2021 Feb 9;23(2):e19910. [doi: 10.2196/19910] [Medline: 33560233]
25. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012 May 8;14(3):e70. [doi: 10.2196/jmir.2058] [Medline: 22564332]
26. Edwards EA, Caton H, Lumsden J, et al. Creating a theoretically grounded, gamified health app: lessons from developing the Cigbreak smoking cessation mobile phone game. JMIR Serious Games 2018 Nov 29;6(4):e10252. [doi: 10.2196/10252] [Medline: 30497994]

XSL•FO
RenderX

Abbreviations

DVT: deep vein thrombosis
mHealth: mobile health
PRO: patient-reported outcome
RCT: randomized controlled trial
UIUQ: User Interface Usability Questionnaire

Edited by D Gooch; submitted 03.Feb.2026; peer-reviewed by M Bass, R Pryss; revised version received 01.Jun.2026; accepted 02.Jun.2026; published 03.Jul.2026.

Please cite as:
Lu Q, Liu Y, Shao M, Wang L
A Mobile App (Tpro) for Symptom Management in Patients With Deep Vein Thrombosis Based on Patient-Reported Outcomes: Design and Development Using an Iterative Convergent Mixed Methods Approach
JMIR Hum Factors 2026;13:e92738
URL: https://humanfactors.jmir.org/2026/1/e92738
doi:10.2196/92738

© Qiaodan Lu, Yafei Liu, Manna Shao, Lei Wang. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 3.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

An Online Epilepsy Self-Management Tool for Patients With Epilepsy: A Formative Usability Study

Katarzyna Czerniak[1], MLA, MPH, PhD; Ross Shegog[1], PhD; Refugio Sepulveda[2], PhD; Robert Addy[1], PhD; Youngran Kim[3], PhD; Sahiti Myneni[4], PhD; Alejandra Garcia-Quintana[1], MPH, DDS; Kimberly Martin[5], LVN; David M Labiner[6], PhD

[1]Center for Health Promotion & Prevention Research, School of Public Health, The University of Texas Health Science Center at Houston, University Center Tower, Suite #2672-2, 7000 Fannin St., Houston, TX, United States

[2]Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, United States

[3]Center for Health Care Data, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, United States

[4]Department of Clinical and Health Informatics, D. Bradley McWilliams School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, United States

[5]Epilepsy Foundation Central South Texas, San Antonio, TX, United States

[6]Department of Neurology, University of Arizona, Tucson, AZ, United States

Corresponding Author:
Katarzyna Czerniak, MLA, MPH, PhD
Center for Health Promotion & Prevention Research, School of Public Health, The University of Texas Health Science Center at Houston, University Center Tower, Suite #2672-2, 7000 Fannin St., Houston, TX, United States

Abstract

Background: Epilepsy is a serious chronic neurological condition with no permanent cure. Continual self-management is important to mitigate seizure frequency and optimize quality of life in people with epilepsy who have greater disparities in accessing epilepsy care. The Management Information & Decision Support Epilepsy Tool (MINDSET [UTHealth, University of Arizona, and Radiant Digital]) was developed to enhance accessibility to epilepsy self-management (ESM) assessment and treatment. The purpose of this formative usability pilot study was to assess the user experience and functionality of MINDSET 2.0, an enhanced cross-platform online version of MINDSET, among a sample of patients with epilepsy prior to feasibility testing within neurology clinic settings.

Methods: MINDSET 2.0 comprised an updated cross-platform architecture for easier accessibility and added quality of life, cognitive function, and social determinants assessments. User experience and functionality were assessed in January 2022. Six patients with epilepsy in Texas (n=4) and Arizona (n=2) participated in individual online usability sessions, completing a sociodemographic survey, accessing all components of MINDSET, and then completing usability rating scales and an exit interview. Logical inconsistencies in embedded algorithms were examined for the usability sample and in user case challenges to ensure functional fidelity.

Results: Patients reported low adherence to ESM behaviors in each of the management domains. More than 80% of patients agreed that MINDSET 2.0 was acceptable, easy to use, likable, credible, of appropriate duration, and motivationally appealing. Patients agreed that the program helped them think about and manage their epilepsy more carefully, and that it improved decision-making between them and their health care providers (100%). Patients provided lower ratings (≤50%) and reported the greatest number of difficulties with their understanding of how to select goals and strategies, and develop an action plan due to constraints of item response options leading to user confusion. An inconsistency in algorithm branched logic was identified that related to translating depression scores into recommendations for depression self-management programming.

Conclusions: The results replicated usability findings from earlier versions of MINDSET but also catalyzed adjustments to user survey response options and algorithm repair. The value of the formative user experience functionality assessment was demonstrated to ensure a high-fidelity program prior to feasibility testing in neurology clinic settings.

(JMIR Hum Factors 2026;13:e79540) doi:10.2196/79540

KEYWORDS

internet; mHealth; mobile health; digital health; eHealth; epilepsy; epilepsy self-management; usability

XSL•FO
RenderX

Introduction

Epilepsy is a neurological condition, defined by repeated seizures of varied onset and type that affects more than 50 million people globally [1]. Approximately 2.9 million adults in the United States have active epilepsy [2,3]. People with epilepsy exhibit seizures that can manifest with temporary loss of awareness or consciousness; disturbed movement; altered vision, hearing, and taste; reduced mental function; anxiety and depression; and lower quality of life [4,5]. Epilepsy self-management (ESM) is important to enhance the self-efficacy of people with epilepsy in managing seizures and to optimize their quality of life [6]. ESM comprises a range of behaviors that include (1) medication management (eg, adherence to prescribed medication and clinical visit regimens), (2) seizure management (eg, preparation for, and response to, seizure episodes), (3) lifestyle management (eg, altering behaviors to avoid seizures and/or adverse consequences of seizures), and shared decision-making between the patient and their health care provider (HCP) [7,8]. ESM is associated with increased management self-efficacy and decision-making by people with epilepsy and their families, fewer breakthrough seizures, lower rates of hospitalization, and decreased risk for injury and premature death [8].

The COVID-19 pandemic disrupted traditional in-person clinic visits due to social distancing requirements and promoted digital solutions for patient treatment, such as telemedicine [9]. Digital decision support can assist HCPs to tailor patient care and increase patient awareness, knowledge, skills, and reach [10]. The Management Information & Decision Support Epilepsy Tool (MINDSET) was developed in 2015 to enhance patients with epilepsy and HCP assessment of ESM and has demonstrated improvement in adherence of patients with epilepsy to ESM [11-14]. An enhanced cross-platform web-based version of MINDSET, MINDSET 2.0, has been developed to enable improved function, phone-based accessibility, and added assessment and feedback on patient quality of life, cognitive difficulties, and social determinants of health [15-17]. User experience testing is a necessary antecedent of successful implementation of digital decision support to ensure it meets the specific needs of the priority population prior to deployment in the field [18]. The purpose of this study was to assess the user experience and functionality of MINDSET 2.0 with a sample of patients with epilepsy prior to deployment for feasibility testing within neurology clinic settings.

Methods

The Intervention

MINDSET is a theoretically and empirically based digital ESM program, originally accessible on desktop or tablet computers in the clinic [11,19]. Patients choose their preferred language (English or Spanish) and input data (“My Epilepsy”) about their epilepsy condition (overall subjective health rating, seizure frequency in the last 30 days, seizure symptoms, antiseizure drug [ASD] prescription, and days of missed doses in the last 2 weeks) and their adherence to ESM behaviors within each of the 3 ESM domains (seizure, medication, and lifestyle management) using the Epilepsy Self-management Survey (ESMS) [20-22]. They complete the Neurological Disorders Depression Inventory in Epilepsy (NDDI-E) to assess depression [23] and the Liverpool Adverse Events Scale to assess ASD side effects [24-26] (Table 1). Patients select up to 3 ESM goals (one each for seizure, medication, and lifestyle management) to improve low adherent ESM behaviors (“My Goals”), choose strategies to accomplish these goals, rate their self-efficacy to achieve their selected goals, and list barriers to meet these goals [20,22]. The goals are designed to increase patient adherence to low adherent behaviors. A tailored goal-based action plan (AP; “My Action Plan”) is dynamically generated and is printed to review and discuss with the HCP during the clinic encounter (Figures 1 and 2). Page 1 of the AP comprises a summary of the patient’s epilepsy history and current condition, assessment of ESM behaviors and depression (Figure 2). Pages 2 to 4 summarize the selected goal, strategies, self-efficacy, and barriers for seizure, medication, and lifestyle management, respectively.

XSL•FO
RenderX

Table . Measures in MINDSET[a] and user experience testing.

	Description
MINDSET measures (original version)	
Behavior	
Epilepsy Self-Management Scale[b]	A 38-item survey with 5-point Likert scale response format assessing frequency of epilepsy self-management practice in 5 management domains: seizure, medication, lifestyle, safety, and information ("I plan ahead so that I do not run out of seizure medication," "I do things I enjoy to help manage stress," etc). Each item is rated from 1 (never) to 5 (always) with negative behaviors reverse scored. Cronbach α: 0.81 - 0.84 [20,27-30].
Neurological Disorders Depression Inventory for Epilepsy	A 6-item survey with 4-point Likert response format assessing frequency of depressive symptoms within the last 2 weeks (range is 1 "never" to 4 "always or often"). Final score is sum of all responses (range 6 - 24). Scores ≥15 indicate depressive symptoms (specificity: 90%, sensitivity: 81%, and positive predictive value of 0.62 for detecting major depressive disorder [23]). Cronbach α: 0.85, test-retest reliability: 0.78 [31,32].
Clinical outcomes	
Seizure frequency	Frequency of seizures experienced in the past 4 weeks. People with epilepsy input a number in an open textbox format.
Adverse Drug Effects Scale[c]	Assesses medication side effects experienced during the previous 4 weeks from a list of 19 adverse effects. The original uses a 4-point Likert scale response set: 1 (never a problem) to 4 (always a problem). Score range: 19 to 76. Cronbach α: 0.90 (baseline) and 0.91 (3-month follow-up) [24-26].
Missed doses item	Assesses how many doses of anti-seizure medications patients missed in the last 2 weeks. People with epilepsy can select any number between 1 and 20 as well as nonnumeric answers which are "more than 20," "don't know," and "Not currently on any seizure medications."
MINDSET 2.0 measures (enhanced version): MINDSET 2.0 contains all the measures from the original version listed above plus the additional measures listed below	
Behavior	
QOLIE-10[d] Survey[e]	Assesses how often epilepsy has affected the patient's health and daily activities in the last 4 weeks. Scale contains 10 epilepsy-specific items from 3 domains: epilepsy effects (memory, physical, and mental effects of medication), mental health (energy, depression, and general quality of life), and role functioning (seizure worry, work, driving, and social limits). Likert responses vary and include: 1 (all of the time) to 6 (none of the time), 1 (a great deal) to 5 (not at all), and 1 (not at all bothersome) to 5 (extremely bothersome). Test-retest reliability and internal consistency are high [33].
QOLIE-31[f] Survey (single item)[g]	A single item that assesses overall perceived quality of life using a rating scale of 0 (worst possible quality of life) to 10 (best possible quality of life). This item is analogous to a general health question that often "front ends" a clinic visit enquiry. A score is calculated by converting the raw numeric values to 0 - 100 point scores. Higher scores reflect better overall quality of life [34].
QOLIE-31 cognitive subscale	This 6-item self-report subscale assesses how often patients experience epilepsy-related cognitive issues in their daily lives. Response selections vary by item and include 1 (all of the time) to 6 (none of the time), 1 (a great deal) to 4 (not at all), and 1 (not at all bothersome) to 5 (extremely bothersome). A score is calculated by converting the raw numeric values to 0 - 100 point scores with higher scores indicating better cognitive health [34].
Health Leads social determinants inventory[h]	An 8-item social determinants inventory screening tool to assess if patients are having difficulties in 8 areas of their daily lives. People with epilepsy respond Yes or No to each of the 8 items [35].
User experience testing measures	

XSL•FO
RenderX

	Description
MINDSET measures (original version)	
Demographic survey	A 26-item demographics survey that collects information about gender, age, race, ethnicity, residency in the United States, education, marital status, income, employment status, health care status, seizure history, quality of life, computer and smartphone access, and language preference [36].
MINDSET Usability Rating Scale[i]	An adapted 36-item rating inventory to assess usability parameters of acceptability, likability, credibility, motivational appeal, the content, duration to complete the program, usefulness, perceived impact on epilepsy, ease of use, and understandability. Seven items were added to cover new content in MINDSET 2.0. This rating inventory assessed specific functions in MINDSET and enabled comparison with results from previous MINDSET studies [12,37].
Feasibility exit interview[j]	A 30-minute structured exit interview conducted after survey data was collected. A trained facilitator asked 6 questions, with additional prompts as needed, to assess the patient's experience and their perceptions of the strengths, weaknesses, utility, and appeal of MINDSET 2.0, and to elicit recommendations for enhancement [12,37].

[a]MINDSET: Management Information & Decision Support Epilepsy Tool.

[b]MINDSET patients are considered adherent if the positive behaviors occur "usually" or "always." Epilepsy self-management behaviors are flagged by the program as nonadherent if patients report "never," "rarely," or "sometimes" for a positive behavior. This is reversed for negative behaviors.

[c]In MINDSET, the response set was adjusted to a check box selection of any side effects and 3 additional adverse effects were included informed by neurologist review. All reported medication side effects are listed on the action plan produced at the end of the program, which helps health care providers identify patients needing treatment adjustments during clinic visits. Adverse drug effects choices: unsteadiness, tiredness, restlessness, aggression, nervousness, hair loss, skin changes or rash, blurred vision, upset stomach, concentration difficulty, mouth and gum problems, shaky hands, weight gain, dizziness, sleepiness, depression, memory problems, and disturbed sleep.

[d]QOLIE-10: Quality of Life in Epilepsy Inventory-10.

[e]Items with opposite response sets are reverse scored so that positive responses are lower numbers and negative responses higher. The total score is calculated by averaging the items answered. Total scores range between 0 and 6, with lower scores indicating better quality of life.

[f]QOLIE-31: Quality of Life in Epilepsy Inventory-31.

[g]The item is designed for assessing Quality of Life but doesn't itself trigger intervention recommendations. All those who use MINDSET receive a recommendation to Program for Active Consumer Engagement in Self-Management regardless of their answer choices to ensure that everyone has access to more intense intervention should they one day need it.

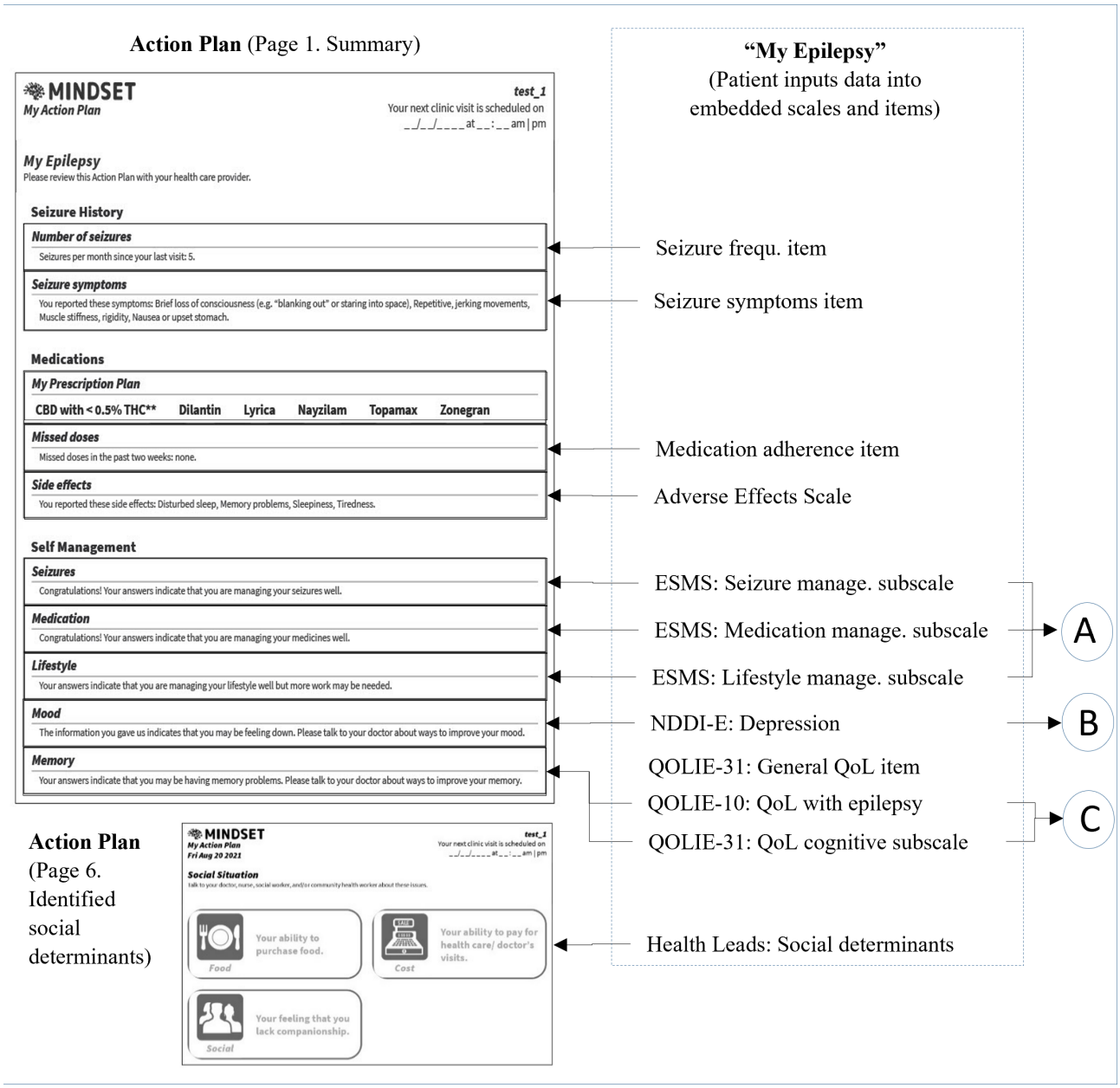
[h]Health Leader assesses food insecurity, housing instability, financial resource strain, transportation challenges, social isolation, lack of childcare, healthcare illiteracy, and utility needs.

[i]The 36-item inventory provides statements (Table 2) about various aspects of MINDSET 2.0 and the action plan for which the patient provides a rating using binary response options including "yes/no," "helpful/not helpful," "can be trusted/can't be trusted." The patient also rates the structure and design (eg, colors, font size, and screen layout) using a 5-point Likert scale (dislike a lot (1) to like a lot (5)), and ease of use using 4-point Likert scales (not easy (1) to very easy (4)).

[j]Questions included: "Describe your overall experience with MINDSET 2.0; What did you like best or least about MINDSET 2.0 and why?; Was there anything that made you feel upset, embarrassed, or uncomfortable?; How relevant or relatable was the information presented to you?; What would you have liked to see but did not see?; What would you tell other patients about MINDSET 2.0?

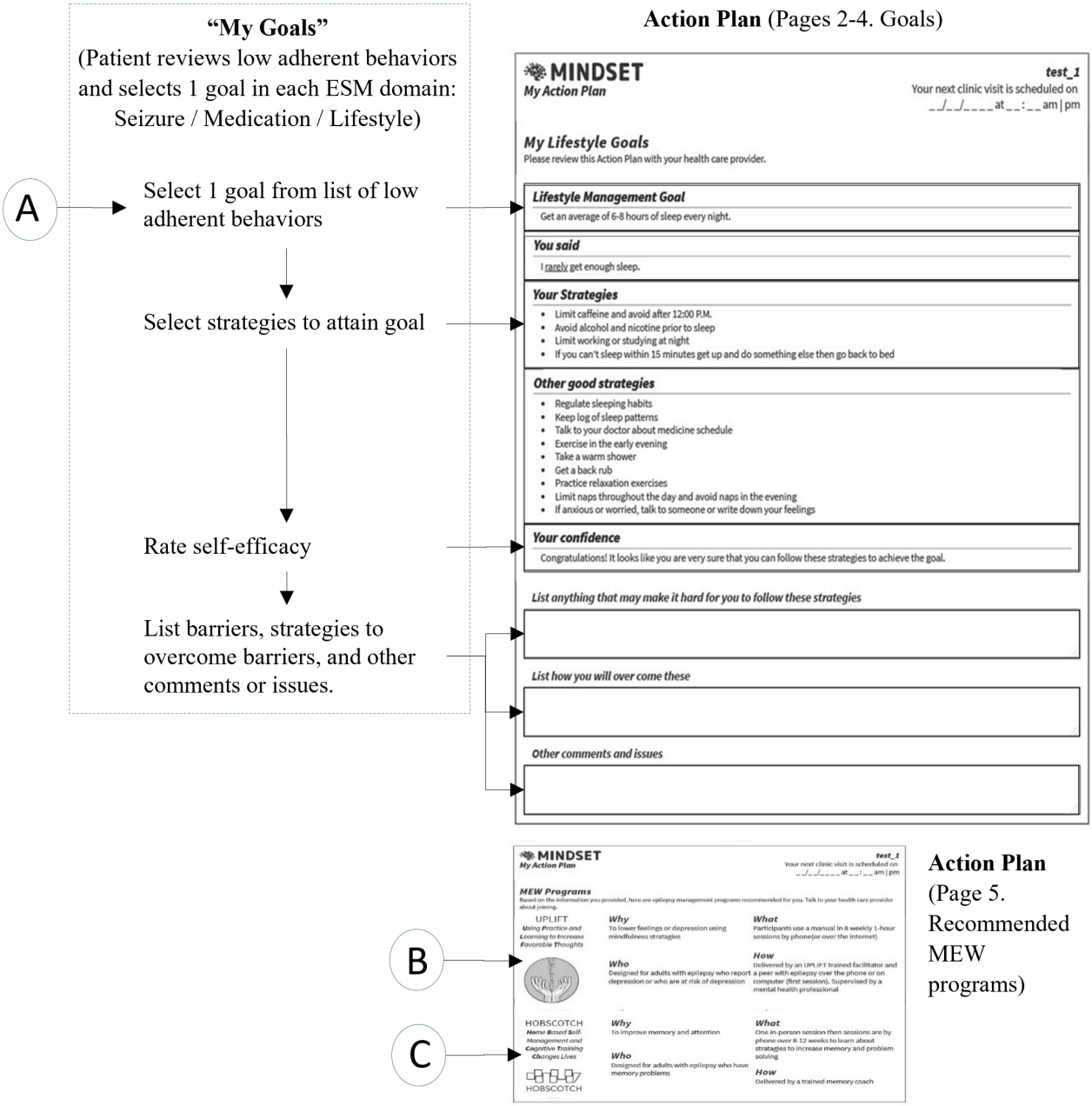
XSL•FO
RenderX

Figure 1. Decision logic for action plan tailoring part 1 (pages 1 and 6). ESMS: Epilepsy Self-management Survey; NDDI-E: Neurological Disorders Depression Inventory in Epilepsy; QoL: Quality of Life; QOLIE-10: Quality of Life in Epilepsy Inventory-10; QOLIE-31: Quality of Life in Epilepsy Inventory-31.



XSL•FO
RenderX

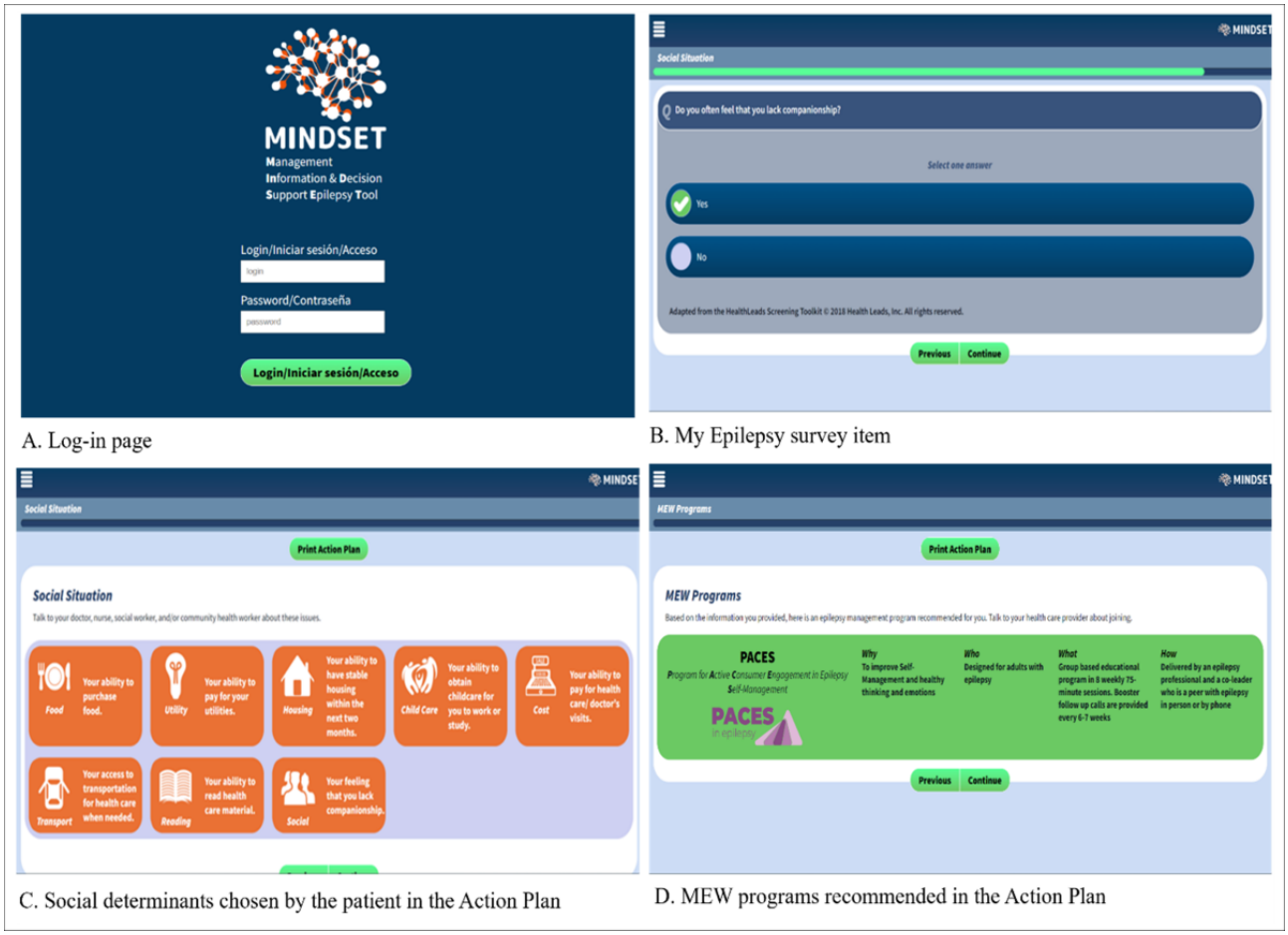
Figure 2. Decision logic for action plan part 2 (pages 2-5). ESM: epilepsy self-management; MEW: Managing Epilepsy Well.



MINDSET has been enhanced to MINDSET 2.0 with added features (Table 1) that include (1) accessibility on any device with an internet connection, (2) assessment of depression using the NDDI-E where a score of ≥15 triggers a summary message in the AP to discuss mood and a recommendation that the patient may benefit from the Using Practice and Learning to Increase Favorable Thoughts (UPLIFT) program designed to train patients with epilepsy to manage comorbid depression [38], and (3) assessment of quality of life using the Quality of Life in Epilepsy Inventory-10 (QOLIE-10) [33] and cognitive (memory) difficulties using the Quality of Life in Epilepsy Inventory-31 (QOLIE-31) cognitive subscale [34]. A QOLIE-31 cognitive subscale score of ≤70 and/or the QOLIE-10 memory item can trigger a recommendation that the patient may benefit from the Home-Based Self-Management and Cognitive Training Changes Lives (HOBSCOTCH) program designed to train patients with epilepsy to manage comorbid memory problems (Figure 2, AP page 5) [39]. Additional features included (4) a default recommendation for the Program for Active Consumer Engagement in Self-Management (PACES) as a foundational ESM training program given its suitability for all patients with epilepsy. UPLIFT, HOBSCOTCH, and PACES are available in English and Spanish. Lastly, MINDSET 2.0 added an (5) assessment of social determinants using the Health Leads social determinant inventory [35] to cue discussion about possible social service needs (Figure 3). The HCP can link the patient to recommended epilepsy programs and connect them with community resources as needed.

XSL•FO
RenderX

Figure 3. Screen captures of MINDSET 2.0. MEW: Managing Epilepsy Well.



User Experience Testing

In-house alpha testing by the research team was conducted to ensure that the content was accurate and consistent with current clinic practice, that spelling and grammar were error-free, that inputted data corresponded to the backend database, and that the Spanish translation was correct. User experience testing occurred in January 2022. A convenience sample of 7 patients with epilepsy was recruited from the Epilepsy Foundation Central & South Texas (EFCST) clinic in San Antonio, Texas (n=5) and the Banner Clinic in Tucson, Arizona (n=2). Inclusion criteria were that patients be 18 years or older; have a diagnosis of epilepsy; and have access to a computer, phone, or tablet with a working internet connection. MINDSET was designed for patients with epilepsy who are able to self-manage their epilepsy, so neurologists determined exclusion of patients with neurological, medical, motor disorders, or learning problems as well as psychiatric and/or behavioral problems that would excessively inhibit their ability to use MINDSET 2.0, complete surveys, or practice ESM activities.

Clinic staff contacted eligible patients via phone and introduced those who verbally consented to the research team coordinator who arranged a testing session. Sessions were conducted on Zoom (Zoom Communications, Inc), a familiar platform to most participants, due to clinical mandates for COVID-19–related social distancing. In each session, participants completed a consent form, a 26-item sociodemographic survey, accessed all components of MINDSET (eg, the screening tool, patient profile, recommendations, and AP), and then completed an adapted 36-item usability rating scale (Multimedia Appendix 1) and a 30-minute structured exit interview (Table 1). The research team coordinators guided participants through the session and noted any issues in a problem log. The coordinators comprised a doctoral candidate in Behavioral Sciences with 6 years of experience in qualitative research in Texas and an assistant research professor at a state college of public health with an adjunct appointment at a department of neurology in Arizona. A bilingual master's in public health (MPH) student took notes and served as a Spanish translator for patients electing to complete MINDSET 2.0 and study surveys in Spanish. Participants were allowed to complete the study with assistance from caregivers if needed. Participants were compensated for their time with Amazon e-gift cards valued at US $60 at the end of the session. User experience data were assessed with descriptive statistics using Stata 18 (StataCorp LLC) [40]. A priori criteria user agreement of at least 70% was set as a usability benchmark. Features that fell below this were identified for modification. Exit interviews were recorded, and transcripts were assessed and collated to qualify each usability parameter.

Ethical Considerations

The study was approved by the institutional review boards of the University of Texas Health Science Center at Houston (HSC-SPH-19 - 0893) and the University of Arizona (2005674587). Informed consent was obtained from all participants who were given the option to withdraw from the

XSL•FO
RenderX

study at any time. All data collected in the study were deidentified to maintain patient privacy and confidentiality.

Functionality Testing

The functionality assessment comprised (1) noting "bugs" or suboptimal performance in a program log and (2) reviewing algorithm logic fidelity from the user experience testing data, whereby inputted participant data were compared to AP output and assessed against an a priori benchmark for success of 100% correspondence. Staff manually calculated participant scores (eg, NDDI-E depression scores and QOLIE-31 cognitive scores) to predict expected recommendations and compared these to the actual recommendations generated on the AP to test functionality of the algorithms. The assessment also included (3) challenging the MINDSET 2.0 algorithm function with an array of user profile test cases ranging from a "worst case" ESM patient profile to a "best case" and variations between these. A "worst case" patient answered all questions negatively and a "best case" positively. User test cases were assigned unique ESM behavior profiles for seizure, medication, and lifestyle management and varied depression, Quality of Life (QoL), and social determinant indices. Each tester was assigned a predetermined set of responses to ensure that all item answers were inputted at least once. Testers selected the corresponding goal for each assigned at-risk item in the profile to ensure all goals were selected at least once. APs were then reviewed to determine that messages and goals were logically consistent with the data input. Similar to the user experience algorithm review, an a priori benchmark for success was set at 100% correspondence between data input and AP output.

Results

Participant Demographics, Epilepsy, and Self-Management Characteristics

Six participants completed MINDSET. One participant withdrew without completing the AP citing epilepsy cognitive impairment. Participants had a mean age of 37.3 (SD 15) years, were equally male and female (n=3, 50%), Hispanic or Latino (n=6, 100%), with a mean US residency of 23 (SD 13.2) years, mainly with high school (n=3, 50%) or 1 - 3 years of college (n=2, 33.3%) education, single (n=4, 66.7%), with income below US $50K (n=5, 83.3%), and were either health insured (n=3, 50%) or not health insured (n=3, 50%; Table 2). Most reported having access to a smartphone and/or a computer (n=5, ≥83.3%) and preferred English or mainly English for reading, speaking at home, and thinking (n=5, 83.3%).

XSL•FO
RenderX

Table . Participant self-reported demographic characteristics (N=6).

Construct	Value
Age (years), mean (SD); median (IQR); range	37.3 (15); 39.5 (21-50); 19-55
Sex, n (%)	
Female	3 (50.0)
Male	3 (50.0)
Ethnicity	
Hispanic	6 (100)
Race	
White	6 (100)
Residency in United States (years), mean (SD); median (IQR); range	23.7 (13.2); 20 (13-35); 10 - 44
Education level, n (%)	
GED[a] (HS[b] graduate)	3 (50.0)
College 1 - 3 years	2 (33.3)
College ≥4 years	1 (16.7)
Marital status, n (%)	
Single	4 (66.7)
Married or living with other	2 (33.3)
Income level (US $), n (%)	
< $10,000	2 (33.3)
$10,000 - $49,000	3 (50.0)
≥$50,000	1 (16.7)
Employment status, n (%)	
Employed for wages	2 (33.3)
Not working (unable, no work, and student)	4 (66.7)
Health care insurance, n (%)	
Employer	1 (16.7)
Public	1 (16.7)
Medicaid	1 (16.7)
No coverage	3 (50.0)
Technology access, n (%)	
Computer	5 (83.3)
Smartphone	6 (100)
Language preferred with friends, n (%)	
English	4 (66.7)
English >Spanish	1 (16.7)
Language preferred (speaking at home), n (%)	
English	3 (50.0)
English >Spanish	2 (33.4)
Spanish	1 (16.7)
Language preference (thinking), n (%)	
English	5 (83.3)
English >Spanish	1 (16.7)

XSL•FO
RenderX

[a]GED: General Educational Development.

[b]HS: high school.

Participants rated their QoL as mostly "very good" (n=3, 50%; Table 3). Their first ever seizure occurred between ages 2 and 15 years. They had active epilepsy, reporting seizures in the last 12 months (n=6), and in the last 30 days (n=3). They mostly reported 5 to 6 seizures in the last month and 4 seizures since their last clinic visit. They reported adherence to their ASD dosing schedule, on average missing a dose on one day in the last 2 weeks. They reported 3 to 8 side effects. The sample mean NDDI-E score was 12.8 (SD 5.6) with 2 participants scoring ≥15. The mean QOLIE-31 item score was 85 (SD 15.2) and the mean QOLIE-31 cognitive subscale score was 51.6 (SD 33.3), indicating poor QoL. However, the mean QOLIE-10 score was 2.5 (SD 1.02), indicating good QoL in relation to patients' epilepsy. The greatest adherence to ESM was for medicine management behaviors (mean 4.5, SD 0.70), and the lowest adherence was for lifestyle management behaviors (mean 3.9, SD 0.41) and information management behaviors (mean 3.5, SD 0.63; Table 3).

XSL•FO
RenderX

Table . Participant self-reported clinical characteristics (N=6).

Construct and category	Value
Quality of Life[a]	
Quality of Life (last 30 days)	
Excellent	1 (16.7)
Very good	3 (50.0)
Fair	2 (33.3)
Epilepsy duration (years), mean (SD); median (IQR); range	
Age at first seizure	8.5 (5.3); 7.5 (5 - 14); 2 - 15
Age at diagnosis	9.8 (6.8); 8.5 (5 - 15); 2 - 20
Seizures[a]	
Seizure frequency (yes or no)	
Last 12 months	6 (100)
Last 30 days	3 (50)
Seizure frequency (number of seizures in 30 days), mean (SD); median (IQR); range	13.3 (14.4); 5 (5-30); 5 - 30
Seizures[b], mean (SD); median (IQR); range	
Seizure frequency since last clinic visit	6.3 (7.6); 4 (0 - 10); 0 - 20
Medication, mean (SD); median (IQR); range	
Number of days in the last 2 weeks ASD[c] medications missed	0.4 (0.54); 0 (0 - 1); 0 - 1
Number of ASD side effects (last 2 wk)	5.7 (0.3); 6.5 (3-8); 1 - 9
Depression, mean (SD); median (IQR); range; number of participants with depression (score ≥15)	
Mood (NDDI-E[d])[e]	12.8 (5.7); 12 (8-19); 6 - 20; 2
Cognitive problems, mean (SD); median (IQR); range	
QOLIE-31 cognitive subscale[f]	51.6 (33.3); 47.3 (22.2 - 79.4); 13.3 - 100
Quality of Life[b], mean score (SD); median (IQR); range	
QOLIE-31[g]	85 (15.2); 85 (80-100); 60 - 100
QOLIE-10[h]	2.5 (1.02); 2.3 (1.7 - 3); 1.5 - 4.3
Behavior	
Epilepsy self-management[i], mean score (SD); median (IQR)	
Seizure	4.4 (0.60); 4.6 (4-5)
Medicine	4.5 (0.70); 4.7 (4.3 - 5)
Lifestyle	3.9 (0.41); 3.8 (3.6 - 4.2)
Safety	4.2 (0.73); 4.1 (3.8 - 5)
Information	3.5 (0.63); 3.41(3-4)

[a]Collected using the demographics survey.

[b]Collected using the data collection portion of the MINDSET program.

[c]ASD: antiseizure drug.

[d]NDDI-E: Neurological Disorders Depression Inventory in Epilepsy.

[e]Responses, Likert scale—never: 1 to always or often: 4; symptoms, maximum score 24, score ≥15 is symptomatic.

[f]Cognitive subscale (items 12, 15, 16, 17, 18, and 26) from QOLIE-31 scale. Combined score of ≥70 is obtained if the memory problem is “little” or “none” of the time.

[g]QoLIE-31 item 1. Higher scores reflect better quality of life; lower ones, worse quality of life. Possible range is 0-100.

XSL•FO
RenderX

[h]QOLIE-10 scale. Total scores range between 0 and 6, with lower scores indicating better quality of life.

[i]Self-management range of mean item scores from a 5-point Likert response set with 1-Never, 2-Rarely, 3-Sometimes, 4-Most of the time, 5-Always

User Experience

Participants took approximately 30 - 60 minutes to complete MINDSET, up to 30 minutes to complete the survey and ratings scale, and up to 30 minutes to complete the exit interview. Participants mostly selected the English language version (n=5, 83.3%). More than 80% of participants agreed that MINDSET 2.0 was acceptable, likable, and motivationally appealing (that they would use it again during future visits and recommend it to others) and indicated it to be "exciting," "colorful," and "relevant" (Table 4). They rated it as credible (n=6, 100%) and of appropriate duration (n=5, 83.3%), and that, in its entirety, it was "easy to use" and "simple" (n=6, 100%), and that the words used were understandable (n=6, 100%), requiring no help to answer survey items (n=5, 83.3%). They commented favorably on its comprehensiveness, stating that it was "thorough" and "in-depth," and appreciated the degree of detail in the tailored AP. They rated it as providing needed (n=5, 83.3%) and helpful (n=6, 100%) information, recommending ESM programs appropriate to their needs (n=5, 83.3%), and that it helped them think about and manage their epilepsy more carefully, and that it would improve decision-making with their HCPs (n=6, 100%). They rated it as more useful than other programs they had used in the clinic (n=5, 83.3%) and described it as an effective and "innovative" evaluation tool that introduced them to new things and assisted them to understand broader lifestyle factors associated with seizure reduction such as sleep, stress, diet, and substance use. Participant recommendations for enhancement of MINDSET content were confined to suggestions that MINDSET "assess family dynamics" due to the reliance of many patients with epilepsy on family members to assist with seizure and medication management, and provide support to mitigate stigma and financial strain [7,13]. Requests were also made that MINDSET 2.0 flag suicidal ideation based on participant experience.

XSL•FO
RenderX

Table . Patient agreement on usability parameters.

Usability parameters and items	Agreement, n (%)
Acceptability	
Entire program	6 (100)
Likeability[a] (interface design)	
Colors used in this program	5 (83.3)
Buttons and sliders used in this program	6 (100)
Size of the letters	5 (83.3)
Device used	5 (83.3)
Layout of the screen	6 (100)
Motivational appeal	
I would recommend MS+ to other patients: Yes	6 (100)
I would use MS+ again in a clinic visit: Yes	6 (100)
Credibility	
I think the information I got from MS+ can be trusted	6 (100)
Duration	
Time to use was	5 (83.3)
Just right “Too Long”	1 (16.7)
Usefulness	
I think the information I got from MS+ was helpful	6 (100)
MS+ gave me most of the information that I needed	6 (83.3)
The MEW[b] Network programs that MS+ suggested is appropriate for my needs	5 (83.3)
Having MS+ suggest MEW[b] Network programs was helpful	5 (83.3)
MS+ is as useful or more useful than other seizure programs that I have used[c]	1 (100)
Perceived impact	
I think the information provided...:	
Helped me to think carefully about my epilepsy	6 (100)
Will help me talk to my doctor or nurse about my epilepsy	6 (100)
Will help me manage my epilepsy better	6 (100)
Easy to use	
It was easy to[d]: Enter your responses into MS	6 (100)
I think MS was easy to uses	6 (100)
Understandability	
I did not need help to answer questions	5 (83.3)
I knew and understood most of the words	6 (100)
Action plan development online	
Understand instructions for “My Goals”	0 (0)
Understand the action plan	2 (33.3)
It was easy to[d] choose a goal to work on	3 (50)
Choose your strategies	3 (50)
Action plan development	
Understanding the PDF version of the action plan	3 (50)
Understanding the PDF action plan layout	0 (0)

XSL•FO
RenderX

Usability parameters and items	Agreement, n (%)
Understanding the PDF action plan content	2 (33.3)
Downloading PDF version of action plan	1 (16.7)

[a]Likert scale: like a lot (1)–dislike a lot (4); Like = “Likes a lot” and “like.”

[b]MEW: Managing Epilepsy Well.

[c]Respondents who have not used another program for comparison (n=5).

[d]Likert scale: very easy–not easy; easy = ”somewhat easy” and “very easy.”

Item Formatting

Patients reported lower agreement with the ease of selecting goals and strategies (n=3, 50%) and understanding and using the AP (n=2, ≤33%). The MINDSET 2.0 AP occasionally provided feedback and goal options that seemed to the participants to be misaligned or incongruous to their daily ESM. Participants expressed some frustration and confusion as to why certain behaviors seemed incorrectly flagged as low adherent. The interviews revealed that many patients had difficulty providing responses to the questions due to a restricted 4-point Likert scale. For example, to the item “I use safety precautions in the kitchen,” a patient in Texas responded “never,” because she is “never in the kitchen.” Given this rationale, the more accurate response would have been “not applicable” rather than a response indicative of nonadherence. A patient in Arizona responded, “all the time” to the statement “I go swimming alone,” their rationale being that they do not need to “hold on to anyone or anything” when going to swim rather than being unaccompanied. Patients’ misinterpretation of the meaning and purpose of the questions suggested problems with understandability at the item level. Despite published findings of empirically sound psychometrics for the ESM survey, this feedback indicated the need to incorporate “not applicable” response options into items to increase response validity. In interviews, most reported difficulties were in the development and understanding of the AP. One participant reported that it seemed “repetitive” and “arduous,” while another expressed difficulty in understanding how to use it, inclusive of selecting goals and strategies, interpreting the output, and AP download.

Patient Profiling and Goal Selection

MINDSET 2.0 exhibited no connectivity or bandwidth lags or data dropout, and back-end data collection was error-free. ESM was positively skewed and highest for medicine, seizure, and safety management subdomains (Table 3). The number of low adherence behaviors per participant ranged from 1 or 2 for seizure management, 2 to 6 for medication management, and 2 to 15 for lifestyle management (Table 5). Most patients (n=4, 66.7%) reported high adherence to medicine management behaviors, which was compatible with the low frequency of missed ASDs. Selected goals were “always keeping enough medicine” to avoid running out and “using a pill container or pill box” for ASDs (Table 5, column 2). All patients were flagged as having at least one low adherent lifestyle behavior. Selected goals focused on stress and relaxation techniques, getting an average of 6 - 8 hours of sleep (n=2), carrying an epilepsy ID, forming social connections, and eating a healthy diet. The goals were unique to each participant apart from “getting enough sleep,” which was selected by 2 patients. For any chosen goal, patients selected between 0 and 7 strategies (Table 5, Column 3). The majority selected one strategy from the list of strategies provided. Two patients listed potential barriers to meeting their goals, both for seizure management, that related to the need to get a primary care doctor or neurologist, and health insurance (Table 5, column 3). No participant self-efficacy ratings were downloaded in the AP. This was traced to an error in coding logic failure that was converting categorical Likert scale scores into binary scores.

XSL•FO
RenderX

Table . Flagged ESM[a] behaviors, participant-selected goals, and strategies[b].

Specific participants	Low adherent (flagged) ESM behaviors[c]	Goal[d]	Strategies[e] and barriers[f]
Seizure management (I Rarely...)[g]		(Always...)	
1	• Call my doctor if I am having more seizures than usual.	Contact your doctor or nurse when you are experiencing increased seizures.	• Other ["Applying for health insurance to have a primary doctor"] • Barrier ["Don't have primary doctor or health insurance"]
2	• Call my doctor if I am having more seizures than usual. • Have blood tests done when the doctor orders them.	Same as above.	• Keep a log of seizures • Barrier ["…to find a neuro that I can work with…"]
6	• Call my doctor if I am having more seizures than usual.	Same as above.	• None ["Memorizando"]
Medication management (I Rarely...)[h]		(Always...)	
2	• Spread out the time between doses when my seizure medication is running out • Plan ahead so that I do not run out of my seizure medicine • Take my seizure medicine at the same time each day • Use a pill container or pill box for my seizure medicines • Have a way to remind myself to take my seizure medicine • Forget to take my seizure medicine[i]	Keep enough medicine.	• Ask for a refill of medicines at least 2 weeks before • If away from home for a period of time, make a plan so that you don't run out of medicine
4	• Plan ahead so that I do not run out of my seizure medicine • Use a pill container or pill box for my seizure medicines	Use a pill container or pill box.	• Use a pill box
Lifestyle management (I Rarely...)		(Always...)	
1	• Have ways to help relax to reduce chances of having a seizure • Eat regular, healthy meals • Keep a record of seizures • Talk to or chat online with other people who have epilepsy • Stay out late at night[i] • Climb on high stools or chairs or ladders[i]	Reduce your stress and practice relaxation techniques.	• Prioritize what is important in your life

XSL•FO
RenderX

Specific participants	Low adherent (flagged) ESM behaviors[c]	Goal[d]	Strategies[e] and barriers[f]
2	• Make sure I get enough sleep • Do things that I enjoy to help manage stress • Eat regular, healthy meals • Keep a record of seizures • Use safety precautions when in the kitchen • Keep track of the side effects of seizure medicine • Check with my doctor before taking other medicines. • Go swimming alone[i] • Climb on high stools or chairs or ladders[i] • Get enough exercise • Avoid driving until seizure free for the minimum amount of time required in my state • Wear a bracelet or necklace, or carry information stating that I have seizures/epilepsy • Tell doctor or nurse when I think I'm having unexpected side effects from seizure meds • Use power tools such as electric saws, electric hedge trimmers, or electric knives without an automatic shutoff[i]	Get an average of 6 - 8 hours of sleep every night.	• Regulate sleeping habits • Keep log of sleep patterns • Exercise in the early evening • Practice relaxation exercises
3	• Do things that I enjoy to help manage stress • Wear a bracelet or necklace, or carry information stating that I have seizures or epilepsy • Carry a cell phone to call someone if I need help • Use safety precautions when in the kitchen • Keep a record of seizures • Keep track of side effects of seizure meds	Always carry an ID stating your epilepsy.	• None selected
4	• Make sure I get enough sleep • Get enough exercise • Wear a bracelet or necklace, or carry information stating that I have seizures or epilepsy • Talk to or chat online with other people who have epilepsy • Stay out late at night[i] • Climb on high stools or chairs or ladders[i]	Get an average of 6 - 8 hours of sleep every night.	• Regulate sleeping habits • Keep log of sleep patterns • Limit working or studying at night • Take a warm shower • Limit naps throughout day and avoid naps in the evening • If anxious or worried, talk to someone or write down feelings • If can't sleep within 15 minutes get up and do something else then go back to bed.
5	• Avoid driving until seizure free for the minimum amount of time required in my state • Talk to or chat online with other people who have epilepsy	Find people to talk to or chat online with who also have epilepsy.	• Join an epilepsy support group (local or community groups or online)
6		Take control of your diet.	• Eat a well-balanced diet

XSL•FO
RenderX

Specific participants	Low adherent (flagged) ESM behaviors[c]	Goal[d]	Strategies[e] and barriers[f]
	• Have ways to help relax to reduce chances of having a seizure • Avoid driving until seizure free for the minimum amount of time required in my state • Keep track of side effects of seizure meds • Talk to or chat online with other people who have epilepsy • Get enough exercise • Keep a record of seizures • Keep the water temperature in my home low enough so that it would not burn me if I had a seizure in the shower • Wear a bracelet or necklace, or carry information stating that I have seizures or epilepsy • Do things that I enjoy to help manage stress • Eat regular, healthy meals		

[a]ESM: epilepsy self-management.

[b]Self-efficacy was not collected because the Likert scale response was converted to a binary response. This was repaired subsequent to this study.

[c]Low adherent epilepsy self-management behaviors are defined as having a frequency of "never," "rarely," or "sometimes" on a 5-point Likert scale.

[d]Selected by the participant.

[e]When an "other" category was selected, the patient could provide a written response in the text box provided.

[f]Patients have the option to list barriers that impede their ability to meet goals as well as list strategies used to overcome barriers in an open-ended data entry box if desired.

[g]Participants 3, 4, and 5 had no flagged behaviors.

[h]Participants 1, 3, 5, and 6 had no flagged behaviors.

[i]For negative behaviors, answer selections and scores were reversed.

ESM Program Recommendation Logic

All participants reported having no past use or experience with any of the Managing Epilepsy Well (MEW) Network programs (UPLIFT, HOBSCOTCH, or PACES). In accordance with the MINDSET 2.0 design, all patients were recommended to attend the PACES epilepsy training program, which was the default recommendation. The HOBSCOTCH cognitive skills training program recommendations for patients flagged as having cognitive difficulties (n=5) were also consistent with the MINDSET logic model (Table 6). However, correspondence between predicted and actual recommendations for the UPLIFT program was less than 100% for 2 participants (Table 6, participants #3 and #6). These 2 participants were flagged as having depressive symptoms (NDDI-E score >15) but were not recommended by MINDSET 2.0 to attend the UPLIFT depression program despite no prior history with this program. This represented a failure of the algorithm to translate NDDI-E depression scores into a recommendation to remediate for depression and was traced to incorrect algorithm decision logic.

XSL•FO
RenderX

Table . ESM[a] program recommendation algorithm functionality.

Data input ("my epilepsy" survey)										Data output (action plan)						Correspondence of predicted to actual (%)[b]
Participant[c]	ESM Behavior Flagged[d]			Depression[e]	Memory[f]		Previous Exposure to P/U/H[g]			MEW[h] Program Recommendation[i]						
	S	M	L	NDDI-E[j]	QOLIE-31[k] cognitive scale	QOLIE-10[l] item	P	U	H	PACES[m]		UPLIFT[n]		HOB[o]		
										Pred.	AP[p]	Pred.	AP	Pred.	AP	
1	✓	—	✓	—	—	✓	No	No	No	✓	✓	—	—	✓	✓	100
2	✓	✓	✓	—	✓	✓	No	No	No	✓	✓	—	—	✓	✓	100
3	—	—	✓	✓	✓	✓	No	No	No	✓	✓	✓	—	✓	✓	66.6[q]
4	—	✓	✓	—	✓	✓	No	No	No	✓	✓	—	—	✓	✓	100
5	—	—	✓	—	—	—	No	No	No	✓	✓	—	—	—	—	100
6	✓	—	✓	✓	✓	✓	No	No	No	✓	✓	✓	—	✓	✓	66.6[q]

[a]ESM: epilepsy self-management.

[b]Correspondence of predicted to actual recommendations. The benchmark is 100%.

[c]Represents a specific patient with epilepsy participant.

[d]ESM domains (S=Seizure; M=Medicine; L=Lifestyle). A domain is flagged with a checkmark if a behavior within that domain has low adherence. Goal options are generated for each low adherent behavior for the patient to select during the action plan development

[e]Depression is indicated on the action plan if a patient obtains a summed score of 15 or more on the Neurological Disorders Depression Inventory in Epilepsy scale.

[f]Cognitive issues are indicated on the action plan if a patient obtains a score of 70% or lower on the Quality of Life in Epilepsy Inventory-31 cognitive subscale OR answers 3, 4, or 5 (extremely bothersome) on the Quality of Life in Epilepsy Inventory-10 item about memory OR both.

[g]Past exposure to Managing Epilepsy Well Network programs (P=PACES; U=UPLIFT; H=HOBSCOTCH) is a criterion contributing to determine if a patient receives a recommendation to one or more of these programs. Patients self-report any past exposure as Yes (Y), No (N), or Don't Know (DK) for each of the 3 Managing Epilepsy Well programs.

[h]MEW: Managing Epilepsy Well.

[i]Predicted (pred.) and actual (AP) recommendations for Managing Epilepsy Well Network programs are shown (✓=recommended; - =not recommended).

[j]NDDI-E: Neurological Disorders Depression Inventory in Epilepsy.

[k]QOLIE-31: Quality of Life in Epilepsy Inventory-31.

[l]QOLIE-10: Quality of Life in Epilepsy Inventory-10.

[m]PACES: Program for Active Consumer Engagement in Self-Management

[n]UPLIFT: Using Practice and Learning to Increase Favorable Thoughts.

[o]HOBSCOTCH: Home-Based Self-Management and Cognitive Training Changes Lives.

[p]AP: action plan.

[q]Participants #3 and #6 had noncorresponding results where depression scores should have caused a recommendation for UPLIFT in the action plan.

Social Determinant Feedback

Four patients reported at least one social determinant that may have hindered their epilepsy management, mainly food insecurity (n=3) and financial resource strain (n=3). Other listed determinants were housing instability (n=2), low health literacy (n=2), transportation challenges (n=1), utility needs (n=1), and social isolation (n=1).

Discussion

Principal Findings

The purpose of this study was to assess the usability of MINDSET 2.0 prior to deployment for feasibility testing in neurology clinics that serve patients with epilepsy. The 6 testers, who were people with a doctor's diagnosis of epilepsy, were

XSL•FO
RenderX

experienced in epilepsy management and forthcoming about recommending improvements to MINDSET 2.0 to enhance its comprehensiveness and relevance. Consistent with the health disparities often cited for patients with epilepsy at the national level, the participants reported limited insurance coverage, education, and employment, and the most reported social determinants were food insecurity and financial strain that can impede their ability to manage their epilepsy. The usability findings were largely consistent with previous studies of MINDSET in that the patients agreed that MINDSET 2.0 was acceptable, likable, credible, easy to use, and useful for decision-making and ESM, meeting a priori criteria of 70% agreement [12,37]. Patients reported less overall adherence to, and competency in, information management (monitoring and recording seizures and symptoms, and communicating about epilepsy) and lifestyle management (eg, ensuring adequate exercise, sleep, and strategies to relieve stress) compared to other more medically oriented ESM behaviors of managing medication, safety, and seizures. This was consistent with prior studies on ESM among patients with epilepsy [12,37].

MINDSET 2.0 has added features to screen for comorbidities of depression symptoms and subjective cognitive difficulties and to recommend training programs of UPLIFT and/or HOBSCOTCH, respectively. Patients with epilepsy have a greater prevalence of psychiatric comorbidities than the general population and are at greater risk of being undertreated for depression, which is consistently demonstrated to be inversely associated with successful ESM [41,42]. Consistent with previous studies, more than 33% (n=2) of MINDSET participants reported symptoms of depression. A question item regarding suicidal ideation is also embedded in the NDDI-E survey. Adding an alert in the AP specific to suicidality was suggested that will further highlight this issue. Specific reference to this item has since been included in the neurologist MINDSET 2.0 training.

People with epilepsy have an increased risk for impaired memory, executive function, attention, and social skills in comparison to people without epilepsy [43,44]. Cognitive function can be further compromised by early onset of epilepsy, frequent recurrent seizures, and polytherapy [43]. Low QOLIE-31 cognitive subscale scores (≤70%) were consistent with the prevalence of cognitive impairment of 60% and 70% for patients with epilepsy in general [43,45]. Patients with epilepsy often have cognitive impairments that can compromise their use of MINDSET 2.0. The positive ratings for many usability parameters are encouraging and consistent with previous studies of ESM interventions that report outcomes of increased awareness, feelings of autonomy, ESM adherence, and a sense of hope [13,17] and significant improvement in cognitive health and depression. Previous studies have also reported that between 30% and 50% of patients with epilepsy will underestimate or overestimate their cognitive performance, so objective assessments, in conjunction with caregiver reports as well as mood and anxiety measures, are required to establish truly comprehensive and accurate cognitive assessments [46,47]. MINDSET assists HCPs in identifying and referring those patients who require further objective neuropsychological assessment but does not provide a direct referral pathway for neuropsychological testing nor does it replace neuropsychological testing.

The words used in MINDSET were understandable, suggesting an acceptable reading level. Participants appreciated the program's comprehensiveness while also indicating it was of appropriate duration, appraisals that are typically inversely associated. However, participants had difficulty when responding to and interpreting items that they perceived as incongruous to their experience. The "forced" selection resulted in inaccurate responses. The addition of a "not applicable" response option to previously published survey items and associated adjustment of the decision algorithm logic will help to increase response validity and increase the salience of AP recommendations. While not a majority opinion, the difficulties in AP development reinforced the advantage of involving caregivers in using MINDSET and generating the AP, and of ensuring HCPs reinforce it as well as instruct on its use.

Strengths and Limitations

A large sample size for usability testing is not necessarily required since statistical significance is not a determining factor to establish usability problems, but a limited sample does reduce generalizability to the broader epilepsy population [48]. MINDSET 2.0 offers the potential for increased accessibility and reach. Paradoxically, there is a "digital divide" that may naturally limit reach to some patients. The study's requirement of access to internet-accessible devices introduces a selection bias toward higher socioeconomic status and greater digital literacy. Further, the system supports Spanish, but testing in Spanish was limited (n=1), and thus, no conclusions about user experiences with the Spanish translation can be drawn, but additional testing of the Spanish version is indicated. Patients with more severe forms of epilepsy that impair their ability to use MINDSET were also excluded from this study, further limiting generalizability to the broader population of people with epilepsy. Therefore, future testing of MINDSET among more representative clinical groups is needed.

MINDSET elicits self-reported responses from patients with epilepsy and is therefore subject to associated recall bias. Consistent with their long-term experience with epilepsy, participants were adherent to prescribed ASD dosing and medication management behaviors. However, self-reported medication adherence has low sensitivity for unintentional nonadherence because people with epilepsy tend to overestimate their self-reported medication adherence [49,50] and unintentionally fail to notice suboptimal ASD adherence [51].

MINDSET 2.0 was developed to be used independently, and patients indicated that they could complete it unaided, but the study supported the benefits of enabling assistance from caregivers and/or clinic staff, especially accounting for patients who have comorbid cognitive challenges. Caregiver and family support is an important associated factor, so content addressing this was suggested for inclusion since the concept of "familias" is particularly salient to the Hispanic culture [7]. Inclusion of caregivers, as well as clinic staff, can improve the validity of data collection, goal selection, and AP management. Studies on the application of MINDSET 2.0 in conjunction with caregivers and community health workers are planned.

XSL•FO
RenderX

Conclusion

This study indicates that MINDSET 2.0 may be a useful program for addressing low adherent ESM behaviors, providing tailored management intervention based on patient-mediated goals and strategy selection, screening for depression and memory function, identification of salient social determinants, and cuing linkage of patients to local resources designed to alleviate social needs and to evidence-based ESM epilepsy programs. Findings from this formative usability study replicated those from earlier versions of MINDSET but also highlighted the need for flexible protocols to enable caregiver support and lead to the initiation of enhancements to mitigate inaccurate algorithm decision logic and item response formatting prior to feasibility testing in neurology clinics [12,37].

Acknowledgments

We thank all participants who volunteered their time to complete the study as well as the neurologists and staff members at the Epilepsy Foundation Central and South Texas and University of Arizona Banner Neurology Clinics who helped identify and recruit patients. Generative AI tools were not used at any stage in the preparation of this manuscript.

Funding

This research was funded by a CDC Special Interest Project Grant (SIP 19-003, grant number 6U48DP006413-01-01). In addition, the publication was supported with funding from the University of Texas Health Science Center-Houston (UTHealth Houston).

Data Availability

The datasets generated and analyzed during this study are available on reasonable request. Please contact author RS.

Authors' Contributions

RS, RSEP, RA, and DML conceptualized the study and measurement. KM, RSEP, and KC led participant recruitment. KC, RSEP, and AG-Q facilitated the usability sessions. Data transcription and analyses was conducted by KC toward doctoral dissertation research. KC and RS prepared the manuscript draft, and all authors contributed to critical review and revision to develop the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
MINDSET 2.0 usability instrument.
[DOCX File, 1031 KB - humanfactors_v13i1e79540_app1.docx]

References

1. Beghi E, Giussani G, Nichols E. Global, regional, and national burden of epilepsy, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2019 Apr;18(4):357-375. [doi: 10.1016/S1474-4422(18)30454-X] [Medline: 30773428]
2. Kobau R, Luncheon C, Greenlund K. Active epilepsy prevalence among U.S. adults is 1.1% and differs by educational level—National Health Interview Survey, United States, 2021. Epilepsy Behav 2023 May;142:109180. [doi: 10.1016/j.yebeh.2023.109180]
3. Thijs RD, Surges R, O'Brien TJ, Sander JW. Epilepsy in adults. Lancet 2019 Feb;393(10172):689-701. [doi: 10.1016/S0140-6736(18)32596-0]
4. DeGiorgio CM, Curtis A, Carapetian A, Hovsepian D, Krishnadasan A, Markovic D. Why are epilepsy mortality rates rising in the United States? A population-based multiple cause-of-death study. BMJ Open 2020 Aug 24;10(8):e035767. [doi: 10.1136/bmjopen-2019-035767] [Medline: 32839157]
5. Beghi E. The epidemiology of epilepsy. Neuroepidemiology 2020;54(2):185-191. [doi: 10.1159/000503831] [Medline: 31852003]
6. Huang Y, Zhang M, Liu T. Effects of self-management education on quality of life in patients with epilepsy: a systematic review and meta-analysis. Epilepsy Behav 2025 Jul;168:110393. [doi: 10.1016/j.yebeh.2025.110393]
7. Chong J, Drake K, Atkinson PB, Ouellette E, Labiner DM. Social and family characteristics of Hispanics with epilepsy. Seizure 2012 Jan;21(1):12-16. [doi: 10.1016/j.seizure.2011.08.008] [Medline: 21900025]
8. Luedke MW, Blalock DV, Goldstein KM, et al. Self-management of epilepsy: a systematic review. Ann Intern Med 2019 Jul 16;171(2):117-126. [doi: 10.7326/M19-0458] [Medline: 31261386]
9. Miller WR, Von Gaudecker J, Tanner A, Buelow JM. Epilepsy self-management during a pandemic: experiences of people with epilepsy. Epilepsy Behav 2020 Oct;111:107238. [doi: 10.1016/j.yebeh.2020.107238] [Medline: 32593874]

XSL•FO
RenderX

10. Shegog R, Braverman L, Hixson JD. Digital and technological opportunities in epilepsy: toward a digital ecosystem for enhanced epilepsy management. Epilepsy Behav 2020 Jan;102:106663. [doi: 10.1016/j.yebeh.2019.106663]
11. Shegog R, Begley CE. Clinic-based mobile health decision support to enhance adult epilepsy self-management: an intervention mapping approach. Front Public Health 2017;5(256):256. [doi: 10.3389/fpubh.2017.00256] [Medline: 29043247]
12. Begley C, Shegog R, Harding A, Goldsmith C, Hope O, Newmark M. Longitudinal feasibility of MINDSET: a clinic decision aid for epilepsy self-management. Epilepsy Behav 2015 Mar;44:143-150. [doi: 10.1016/j.yebeh.2014.12.031]
13. Sepulveda R, Chong J, Shegog R, et al. Experiences of using the MINDSET self-management mobile health app among Hispanic patients: results of a qualitative study. Epilepsy Behav 2024 Apr;153:109702. [doi: 10.1016/j.yebeh.2024.109702] [Medline: 38412570]
14. Shegog R, Begley C, Chong J, et al. MINDSET: clinic-based decision support demonstrates longitudinal efficacy for increased epilepsy self-management adherence among Spanish speaking patients. Epilepsy Behav 2020 Dec;113:107552. [doi: 10.1016/j.yebeh.2020.107552] [Medline: 33242775]
15. Shegog R, Sepulveda R, Czerniak K, et al. MINDSETPLUS: the “management and information decision support epilepsy tool” to promote assessment, goal-based skills training, and service linkage for people with epilepsy. In: Addy R, editor. Presented at: 15th International Conference on ICT, Society and Human Beings (ICT 2022), the 19th International Conference Web Based Communities and Social Media (WBCSM 2022) and 14th International Conference on e-Health (EH 2022); Jul 19-22, 2022. [doi: 10.33965/ICT_WBC_EH2022_202204C036]
16. Sepulveda R SR, Czerniak K, Addy R, Martin K, Wali S, Torres M. In: Labiner DM, editor. Longitudinal Efficacy Study of MINDSET 20: A Web-Based Clinic Decision-Support Tool for Epilepsy Self-Management (Visit 1 Baseline Findings): American Epilepsy Society; 2024. URL: https://aesnet.org/abstractslisting/title-longitudinal-efficacy-study-of-mindset-2-0-a-web-based-clinic-decision-support-tool-for-epilepsy-self-management-visit-1-baseline-findings [accessed 2026-07-30]
17. Sepulveda R, Sepulveda R, Shegog R, et al. Advancing epilepsy care and self-management for Latino Spanish speaking patients through the use of a mobile web-based application (MINDSET 2.0). Presented at: APHA 2023 Annual Meeting and Expo; Nov 12-15, 2023 URL: https://apha.confex.com/apha/2023/meetingapi.cgi/Paper/525697?filename=2023_Abstract525697.html&template=Word [accessed 2026-07-22]
18. Broekhuis M, van Velsen L, Hermens H. Assessing usability of eHealth technology: a comparison of usability benchmarking instruments. Int J Med Inform 2019 Aug;128:24-31. [doi: 10.1016/j.ijmedinf.2019.05.001] [Medline: 31160008]
19. Bartholomew LK, Parcel GS, Kok G. Intervention mapping: a process for developing theory- and evidence-based health education programs. Health Educ Behav 1998 Oct;25(5):545-563. [doi: 10.1177/109019819802500502] [Medline: 9768376]
20. Dilorio C, Faherty B, Manteuffel B. The development and testing of an instrument to measure self-efficacy in individuals with epilepsy. J Neurosci Nurs 1992 Feb;24(1):9-13. [doi: 10.1097/01376517-199202000-00004] [Medline: 1532014]
21. Dilorio C, Faherty B. Epilepsy self‐management: partial replication and extension. Res Nurs Health 1994 Jun;17(3):167-174. [doi: 10.1002/nur.4770170304]
22. Dilorio C, Faherty B, Manteuffel B, Hoeffer B, Hilbert GA. Self-efficacy and social support in self-management of epilepsy. West J Nurs Res 1992 Jun;14(3):292-303. [doi: 10.1177/019394599201400303] [Medline: 1615646]
23. Kim DH, Kim YS, Yang TW, Kwon OY. Optimal cutoff score of the Neurological Disorders Depression Inventory for Epilepsy (NDDI-E) for detecting major depressive disorder: a meta-analysis. Epilepsy Behav 2019 Mar;92:61-70. [doi: 10.1016/j.yebeh.2018.12.006]
24. Gilliam FG, Fessler AJ, Baker G, Vahle V, Carter J, Attarian H. Systematic screening allows reduction of adverse antiepileptic drug effects: a randomized trial. Neurology 2004 Jan 13;62(1):23-27. [doi: 10.1212/wnl.62.1.23] [Medline: 14718691]
25. Perucca P, Carter J, Vahle V, Gilliam FG. Adverse antiepileptic drug effects: toward a clinically and neurobiologically relevant taxonomy. Neurology 2009 Apr 7;72(14):1223-1229. [doi: 10.1212/01.wnl.0000345667.45642.61] [Medline: 19349601]
26. Panelli RJ, Kilpatrick C, Moore SM, Matkovic Z, D’Souza WJ, O’Brien TJ. The liverpool adverse events profile: relation to AED use and mood. Epilepsia 2007 Mar;48(3):456-463. [doi: 10.1111/j.1528-1167.2006.00956.x] [Medline: 17284301]
27. DiIorio C, Shafer PO, Letz R, et al. Project EASE: a study to test a psychosocial model of epilepsy medication management. Epilepsy Behav 2004 Dec;5(6):926-936. [doi: 10.1016/j.yebeh.2004.08.011]
28. DiIorio C, Faherty B, Manteuffel B. Epilepsy self-management: partial replication and extension. Res Nurs Health 1994 Jun;17(3):167-174. [doi: 10.1002/nur.4770170304] [Medline: 8184128]
29. DiIorio C, Osborne Shafer P, Letz R, et al. The association of stigma with self-management and perceptions of health care among adults with epilepsy. Epilepsy Behav 2003 Jun;4(3):259-267. [doi: 10.1016/s1525-5050(03)00103-3] [Medline: 12791327]
30. DiIorio C, Shafer PO, Letz R, et al. Behavioral, social, and affective factors associated with self-efficacy for self-management among people with epilepsy. Epilepsy Behav 2006 Aug;9(1):158-163. [doi: 10.1016/j.yebeh.2006.05.001] [Medline: 16798100]
31. Gilliam FG, Barry JJ, Hermann BP, Meador KJ, Vahle V, Kanner AM. Rapid detection of major depression in epilepsy: a multicentre study. Lancet Neurol 2006 May;5(5):399-405. [doi: 10.1016/S1474-4422(06)70415-X] [Medline: 16632310]

XSL•FO
RenderX

32. Friedman DE, Kung DH, Laowattana S, Kass JS, Hrachovy RA, Levin HS. Identifying depression in epilepsy in a busy clinical setting is enhanced with systematic screening. Seizure 2009 Jul;18(6):429-433. [doi: 10.1016/j.seizure.2009.03.001] [Medline: 19409813]
33. Cramer JA, Perrine K, Devinsky O, Meador K. A brief questionnaire to screen for quality of life in epilepsy: the QOLIE-10. Epilepsia 1996 Jun;37(6):577-582. [doi: 10.1111/j.1528-1157.1996.tb00612.x] [Medline: 8641236]
34. Cramer JA, Perrine K, Devinsky O, Bryant-Comstock L, Meador K, Hermann B. Development and cross-cultural translation of a 31-item quality of life in epilepsy inventory. Epilepsia 1998 Jan;39(1):81-88. [doi: 10.1111/j.1528-1157.1998.tb01278.x] [Medline: 9578017]
35. The Health Leads screening toolkit. Health Leads. 2022. URL: https://healthleadsusa.org/communications-center/resources/the-health-leads-screening-toolkit [accessed 2026-07-07]
36. Shegog R, Bartholomew LK, Sockrider MM, et al. Computer-based decision support for pediatric asthma management: description and feasibility of the Stop Asthma Clinical System. Health Informatics J 2006 Dec;12(4):259-273. [doi: 10.1177/1460458206069761] [Medline: 17092998]
37. Shegog R, Begley CE, Harding A, et al. Description and feasibility of MINDSET: a clinic decision aid for epilepsy self-management. Epilepsy Behav 2013 Dec;29(3):527-536. [doi: 10.1016/j.yebeh.2013.09.023]
38. Quarells RC, Spruill TM, Escoffery C, et al. Depression self-management in people with epilepsy: adapting project UPLIFT for underserved populations. Epilepsy Behav 2019 Oct;99:106422. [doi: 10.1016/j.yebeh.2019.07.023]
39. Caller TA, Ferguson RJ, Roth RM, et al. A cognitive behavioral intervention (HOBSCOTCH) improves quality of life and attention in epilepsy. Epilepsy Behav 2016 Apr;57:111-117. [doi: 10.1016/j.yebeh.2016.01.024]
40. StataCorp LLC. 2021. URL: https://www.stata.com/ [accessed 2026-07-22]
41. O'Kula SS, Briggs FBS, Brownrigg B, et al. Depression and suicidality among Hispanics with epilepsy: findings from the Managing Epilepsy Well (MEW) Network integrated database. Epilepsy Behav 2021 Dec;125:108388. [doi: 10.1016/j.yebeh.2021.108388]
42. Sajatovic M, Johnson EK, Fraser RT, et al. Self-management for adults with epilepsy: aggregate Managing Epilepsy Well Network findings on depressive symptoms. Epilepsia 2019 Sep;60(9):1921-1931. [doi: 10.1111/epi.16322] [Medline: 31486072]
43. Novak A, Vizjak K, Rakusa M. Cognitive impairment in people with epilepsy. J Clin Med 2022 Jan 5;11(1):267. [doi: 10.3390/jcm11010267] [Medline: 35012007]
44. Motamedi G, Meador K. Epilepsy and cognition. Epilepsy Behav 2003 Oct;4 Suppl 2:S25-S38. [doi: 10.1016/j.yebeh.2003.07.004] [Medline: 14527481]
45. Helmstaedter C, Witt JA. Epilepsy and cognition - a bidirectional relationship? Seizure 2017 Jul;49:83-89. [doi: 10.1016/j.seizure.2017.02.017] [Medline: 28284559]
46. Hohmann L, Berger J, Kastell SU, Holtkamp M. Domain-specific relationships of subjective and objective cognition in epilepsy. Epilepsia 2023 Jul;64(7):1887-1899. [doi: 10.1111/epi.17624] [Medline: 37083235]
47. Hohmann L, Jipping JN, Oltmanns F, Holtkamp M. Changes in overlap of subjective and objective cognition over time in epilepsy surgery candidates. Seizure-Eur J Epilep 2024 Jul;119:36-43. [doi: 10.1016/j.seizure.2024.05.003]
48. Inal Y, Wake JD, Guribye F, Nordgreen T. Usability evaluations of mobile mental health technologies: systematic review. J Med Internet Res 2020 Jan 6;22(1):e15337. [doi: 10.2196/15337] [Medline: 31904579]
49. Buelow JM, Smith MC. Medication management by the person with epilepsy: perception versus reality. Epilepsy Behav 2004 Jun;5(3):401-406. [doi: 10.1016/j.yebeh.2004.02.002]
50. Stirratt MJ, Dunbar-Jacob J, Crane HM, et al. Self-report measures of medication adherence behavior: recommendations on optimal use. Transl Behav Med 2015 Dec;5(4):470-482. [doi: 10.1007/s13142-015-0315-2] [Medline: 26622919]
51. Margolis SA, Gonzalez JS, Spindell J, Mohamadpour M, Grant AC, Nakhutina L. Assessment of medication management capacity in a predominantly African American and Caribbean American sample of adults with intractable epilepsy. Epilepsy Behav 2018 Nov;88:308-314. [doi: 10.1016/j.yebeh.2018.09.022]

Abbreviations

AP: action plan
ASD: anti-seizure drug
EFCST: Epilepsy Foundation Central & South Texas
ESM: epilepsy self-management
ESMS: Epilepsy Self-management Survey
HCP: health care provider
HOBSCOTCH: Home-Based Self-Management and Cognitive Training Changes Lives
MEW: Managing Epilepsy Well
MINDSET: Management Information & Decision Support Epilepsy Tool
MPH: Masters in Public Health
NDDI-E: Neurological Disorders Depression Inventory in Epilepsy

XSL•FO
RenderX

PACES: Program for Active Consumer Engagement in Self-Management
QoL: Quality of Life
QOLIE-10: Quality of Life in Epilepsy Inventory-10
QOLIE-31: Quality of Life in Epilepsy Inventory-31
UPLIFT: Using Practice and Learning to Increase Favorable Thoughts

Edited by A Kushniruk; submitted 24.Jun.2025; peer-reviewed by R Colebunders, SA Margolis; revised version received 27.Mar.2026; accepted 29.May.2026; published 05.Aug.2026.

Please cite as:
Czerniak K, Shegog R, Sepulveda R, Addy R, Kim Y, Myneni S, Garcia-Quintana A, Martin K, Labiner DM
An Online Epilepsy Self-Management Tool for Patients With Epilepsy: A Formative Usability Study
JMIR Hum Factors 2026;13:e79540
URL: https://humanfactors.jmir.org/2026/1/e79540
doi:10.2196/79540

© Katarzyna Czerniak, Ross Shegog, Refugio Sepulveda, Robert Addy, Youngran Kim, Sahiti Myneni, Alejandra Garcia-Quintana, Kimberly Martin, David M Labiner. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 5.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Designing a Trauma-Informed Safety by Design Framework for Nonconsensual Intimate Image Removal Platforms: A 2-Phase Sequential Exploratory Study

Ji-yeon Lee, PhD; Yejun Koh, BA

Department of Counseling-UX Psychology, Hankuk University of Foreign Studies, 107 Imun-ro, Seoul, Republic of Korea

Corresponding Author:
Ji-yeon Lee, PhD
Department of Counseling-UX Psychology, Hankuk University of Foreign Studies, 107 Imun-ro, Seoul, Republic of Korea

Abstract

Background: Removal platforms for nonconsensual intimate images (NCIIs) are essential, as they have become the recourse for victims of such abuse; yet their interfaces are not designed around the cognitive and emotional realities of trauma. Safety by Design directs platforms to protect vulnerable users but not how; trauma-informed care defines what that protection requires, while usability heuristics keep interfaces usable but are not themselves trauma-sensitive. However, no framework integrates them into design guidance that is both usable and protective for NCII removal platforms.

Objective: This study develops trauma-informed design, a cohesive framework that integrates trauma-informed care with established usability heuristics under Safety by Design, so that NCII removal platforms protect trauma-affected users while remaining realistically usable; a preliminary, domain-informed evaluation provides initial supporting evidence.

Methods: We used a 2-phase sequential exploratory design, drawing on a globally accessible NCII removal platform as an illustrative case. Phase 1 developed trauma-differentiated personas and mapped each persona's user journey to surface pain points that varied by level of trauma. Phase 2 analyzed a focus group of 8 female Counseling-UX (User Experience) Psychology students who had completed coursework in user psychological safety with consensual qualitative research methodology. Participants assessed the plausibility of the personas and the proposed framework with respect to perceived emotional safety and perceived usability.

Results: Mapping each persona's journey surfaced pain points that differed by level of trauma, which were synthesized into a 6-guideline trauma-informed design framework: emotional scaffolding, process transparency, self-pacing mechanisms, institutional trust signals, context-aware support, and grounding tools. In phase 2, domain-informed evaluators judged the personas plausible, and the framework directions emotionally safer, and potentially more usable.

Conclusions: Integrating trauma principles into user experience evaluation can help identify and address psychological barriers in NCII removal platforms. The framework provides the preliminary, domain-informed foundation that trauma-sensitive design of such platforms requires. Because the framework was not tested with NCII victim-survivors or users actively seeking image-removal support, findings represent preliminary plausibility and acceptability evidence rather than ecological validation or real-world effectiveness.

(JMIR Hum Factors 2026;13:e94438) doi:10.2196/94438

KEYWORDS

nonconsensual intimate images; persona; Safety by Design; trauma-informed design; trauma; online safety; consent

Introduction

Background

Nonconsensual intimate image (NCII) removal platforms are an essential recourse for victims of NCII abuse. Yet despite real advances in content hashing and proactive detection [1], most removal workflows were built around takedown logistics rather than the psychological realities of trauma, leaving help-seeking more burdensome and distressing than it needs to be [2,3]. Because users must navigate these interfaces while in acute distress, the burden falls on the very people the system aims to protect, thereby contravening Safety by Design's principle that safety should never rest solely on the user [4]. These concerns can deter victim-survivors from seeking help altogether [5]. The problem is therefore not merely technical or regulatory but clinical: it requires understanding how trauma reshapes the cognitive and emotional capacities that digital interaction demands.

Empirical accounts of how victim-survivors experience reporting and removal systems reinforce this concern. Studies of technology-facilitated abuse document that survivors frequently

XSL•FO
RenderX

encounter fragmented, effortful, and opaque help-seeking pathways, and that maintaining safety online imposes a substantial and sustained burden [3]. Research on image-based sexual abuse describes the compounding distress victim-survivors face when navigating institutional and platform responses that are difficult to access, slow, or inadequately explained [6]. Controlled comparison of NCII reporting interfaces across 45 platforms shows that the interaction style of the report itself produces measurable trade-offs in clarity, efficiency, and elicited distress [7]. Survivor-centered work developing digital tools for gender-based violence has further shown that survivor-facing systems often fail not because the underlying technology is inadequate but because they are not designed around victim-survivors' psychological and contextual realities [2]. Taken together, this evidence indicates that the limitations of current NCII removal platforms are experienced by users at the point of help-seeking and justify a design approach grounded explicitly in trauma psychology.

Safety by Design [4], trauma-informed care (TIC) [8], and usability heuristics [9] each address part of what NCII removal platforms require, but none integrates the others into interaction-level guidance, a gap this work addresses. Recent NCII-specific scholarship has mapped how harms arise across sociotechnical layers and where interventions might sit [10], yet no existing work integrates clinical trauma theory, usability evaluation, and platform-level safety engineering into an actionable implementation framework for NCII removal platforms.

The insufficiency of any single tradition is clearest at the points where NCII removal platforms can fall short for the users they aim to serve: the gaps tend to surface at the interaction level, when a victim-survivor must initiate and sustain a multistep reporting process while managing acute distress. Removal workflows implicitly assume a trauma-free user who can tolerate ambiguity, accept opaque data handling, and execute irreversible actions under stress. In contrast, the population served is better characterized by reduced perceived control, dissociative disengagement when retriggered, and fragile reengagement after interruption [11,12]. Such states may manifest in predictable ways: dissociation may interrupt a multistep process, hyperarousal may prompt premature abandonment, shame may inhibit the disclosure some steps require, and learned helplessness may impede reengagement after a process breaks [12]. These patterns are consistent with the clinical conditions that characterize this population, grounded in the neurobiology of interpersonal trauma [13]: the interface tends to assume a user who can negotiate such moments, whereas the trauma-affected user's more likely response may be to withdraw.

To address this gap, we develop and preliminarily evaluate a trauma-informed design (TID) framework for NCII removal platforms through a 2-phase sequential exploratory design. The aim is not to evaluate any single service but to provide a detailed framework, grounded in Safety by Design, for supporting victim-survivors of NCII abuse across removal platforms. The term "trauma-informed design" is not new: related approaches have been advanced for digital systems, including trauma-informed computing [1] and TID heuristics for a social service website [14], building on TIC [8]. Our contribution is not the term but its operationalization, which provides concrete, evaluable design guidance for NCII removal platforms, grounded in Safety by Design. Specifically, we apply trauma-differentiated personas to a representative NCII removal platform to identify barriers invisible to any single framework and subsequently assess candidate design responses with domain-informed student evaluators. By integrating clinical trauma knowledge into user experience (UX) evaluation, this study addresses design barriers shaped by trauma-altered cognition, affect, and behavior among NCII removal platform users. The contributions are 3-fold: (1) a diagnostic methodology bridging TIC, Safety by Design, and usability evaluation to identify psychological design gaps invisible to any one approach; (2) 6 preliminary TID guidelines; and (3) preliminary evidence that domain-informed student evaluators perceived the personas and several design responses as plausible, emotionally safer, and potentially more usable.

Research Questions

The research questions (RQs) are as follows:

- RQ1. When trauma-differentiated personas are applied to the user journey, what trauma-related barriers emerge that conventional usability heuristics alone may not fully capture?
- RQ2. Under Safety by Design, how can TIC and usability heuristics principles be integrated into design guidance that addresses these barriers?
- RQ3. How do domain-informed evaluators judge the plausibility of the trauma-differentiated personas and the resulting design directions with respect to perceived emotional safety and perceived usability?

Methods

Conceptual Framework

Three established traditions each address part of this problem at a different level, and the present framework integrates them. Safety by Design sets the overarching mandate: platforms, not users, must bear the burden of protection (the why); TIC supplies the clinical content of that protection (the what); and usability heuristics govern how the interaction is delivered (the how). We take each in turn and then identify the gap that remains.

Safety by Design

Safety by Design embeds user safety into digital products from conception rather than retrofitting protections after harm [4,15]. Australia's eSafety Commissioner established the framework around 3 overarching principles: Service Provider Responsibility (safety burdens never fall solely on users), User Empowerment (protective defaults with meaningful user controls), and Transparency and Accountability (published enforcement data and accessible policies). The Trust and Safety Professional Association [16] operationalizes these through implementation guidance including proactive risk assessment, friction features (interstitials, pause points, and content warnings), and holistic safety across the full product life cycle.

XSL•FO
RenderX

TIC: Clinical Foundations

TIC originated in behavioral health as an organizational framework that reframes the service encounter from "What is wrong with you?" to "What happened to you?" [8,15]. Following the Substance Abuse and Mental Health Services Administration framework [8], trauma-informed practice is organized around 6 principles: Safety; Trustworthiness and Transparency; Peer Support; Collaboration and Mutuality; Empowerment, Voice, and Choice; and Cultural, Historical, and Gender Issues. These principles, rather than recovery-oriented treatment, anchor and are operationalized by the present framework. The epidemiological basis is substantial: approximately 90% of US adults have experienced at least 1 traumatic event [17], and World Health Organization data indicate 70.4% lifetime trauma exposure across 24 countries [18]. Adverse childhood experiences alone generate an estimated US $581 billion in annual costs across Europe and North America [19].

For digital platforms serving NCII victim-survivors, the assumption of a nontraumatized user is empirically untenable [1]. This is where TIC's clinical specificity becomes essential for design: trauma does not merely cause distress; it structurally alters the cognitive and emotional architecture that digital interaction presupposes. Posttraumatic stress disorder (PTSD) disrupts executive functions including planning, sustained attention, and working memory [11], precisely the capacities that complex multistep digital workflows demand. Trauma erodes institutional trust [12], making conventional credibility signals insufficient. It produces hypervigilance toward perceived threats, meaning that ambiguous interface elements are processed as dangerous rather than neutral. These are not usability problems in the conventional sense; they are clinical phenomena requiring clinical understanding. TIC's Physical Environment domain provides the conceptual bridge: if physical environments must embody TIC through architecture and spatial design, digital environments must do the same through visual design, information architecture, and interaction flows.

A consistent principle across the sexual violence recovery literature is that disempowerment and loss of control are central to the injury, and that restoring a victim-survivor's sense of control and agency is therefore central to recovery [12,20-22]. This reframes perceived control as more than a clinical outcome: it is a design-relevant construct. Interactions that compel a survivor to surrender control (eg, opaque eligibility decisions, irreversible actions, and unexplained delays) risk reenacting the disempowerment of the original harm, whereas interfaces that return choice, pacing, and transparency can reinforce the recovery process itself. TID accordingly treats the restoration of control and agency as the through line linking clinical trauma care to interaction design, and the design guidelines developed here can be read as its concrete instantiations.

The convergence with TIC is significant: Service Provider Responsibility, User Empowerment, and Transparency parallel TIC's institutional accountability, its Empowerment, Voice, and Choice principle, and Trustworthiness. The 2 operate at different levels; however, Safety by Design governs how platforms and policies are structured, whereas TIC governs how interactions feel and what they presuppose about the user's psychological state. NCII removal platform design therefore requires both structural safety and psychological specificity.

Usability Heuristics: Convergence With and Limits Beyond TIC

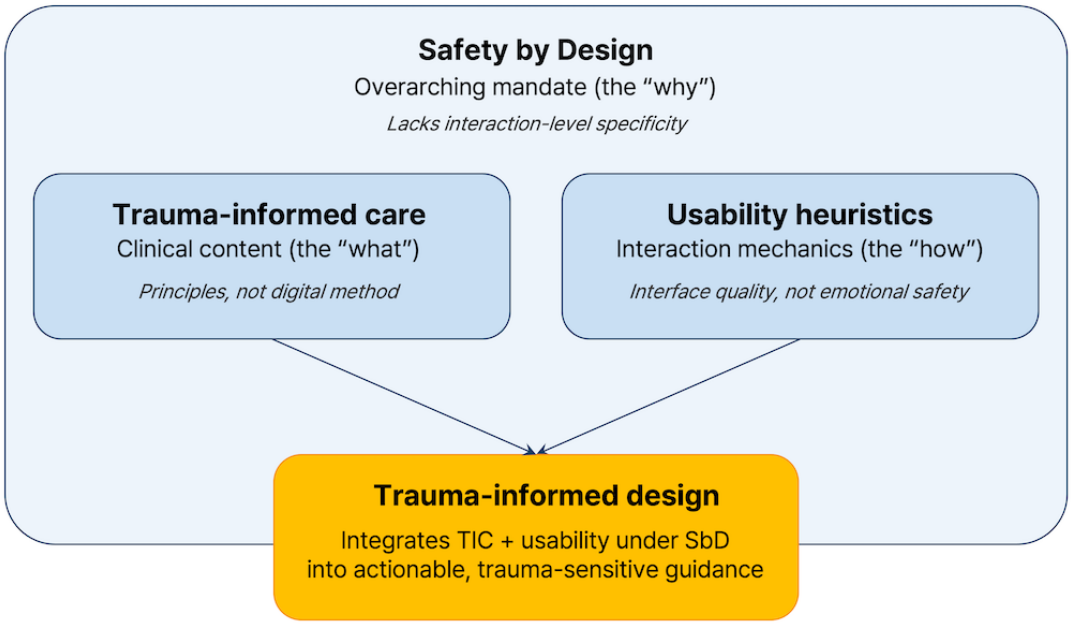
Usability heuristics map onto TIC with notable directness [9]: system status visibility corresponds to Trustworthiness, user control to Empowerment, consistency to Safety through predictability, and recognition over recall to Safety through cognitive load reduction [14]. This convergence suggests that good usability and TID share common ground: both prioritize clarity, control, and predictability. However, the convergence has critical limits. Heuristics assess whether interfaces are learnable and efficient for a general user population. They cannot assess whether an interface is emotionally safe, whether it accommodates the specific cognitive deficits produced by PTSD, or whether it replicates relational dynamics of powerlessness (eg, forced workflows without exit, absence of opt-out mechanisms, and institutional opacity that mirrors abuser control) [23]. For NCII removal platforms where users may be in acute crisis, cognitive load reduction is not an enhancement but a prerequisite. This gap between what heuristics detect and what trauma-informed evaluation requires defines the conceptual space TID occupies.

The Remaining Gap

Trauma-informed computing maps trauma principles onto design but warns that design research can resemble psychotherapy without safeguards [1]; like a subsequent scoping review [5], it remains at the principle-mapping stage. Design-justice work further shows that safety cannot be imposed through expert assumption alone but must be grounded in victim-survivors' lived realities [2]. The gap is thus 3-fold: TIC supplies principles without a digital methodology, usability heuristics supply evaluation tools without emotional safety assessment, and Safety by Design supplies structural mandates without psychological specificity. To date, no existing work integrates all 3 into an operationalized framework for NCII removal platforms. Figure 1 illustrates the 3 traditions, their limits, and how TID bridges them.

XSL•FO
RenderX

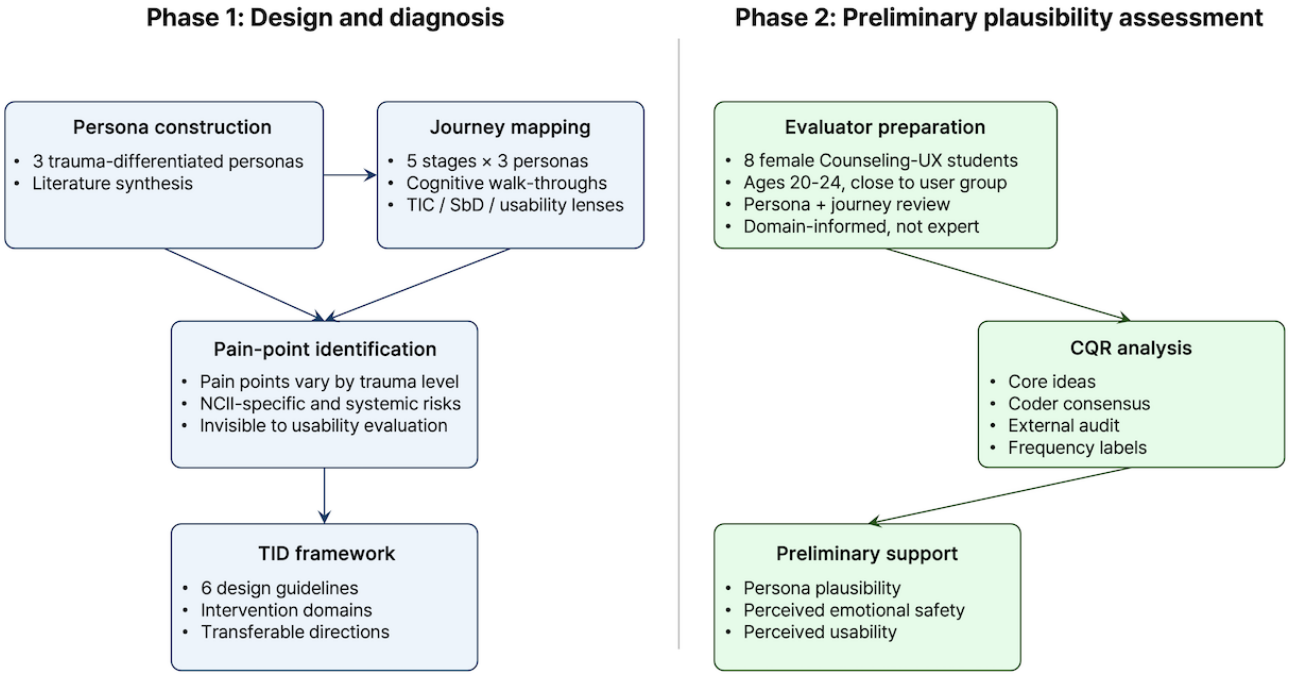
Figure 1. Conceptual framework for trauma-informed design integrating trauma-informed care, Safety by Design, and usability heuristics. SbD: Safety by Design; TIC: trauma-informed care.



Research Design

This study used a 2-phase sequential exploratory design [24]. Phase 1 (Design and Diagnosis) constructed trauma-differentiated personas and applied them via User Journey Mapping to identify TIC and Safety by Design misalignments in an existing NCII removal platform. Phase 2 (Domain-Informed Evaluation) evaluated persona plausibility, perceived emotional safety, and perceived usability of TID-informed design response materials through consensual qualitative research (CQR) [25]. The study was conducted at Hankuk University of Foreign Studies in Seoul, South Korea, from March to August 2025, as a preliminary framework development and plausibility evaluation study. Figure 2 illustrates the overall design.

Figure 2. Research design overview. CQR: consensual qualitative research; NCII: nonconsensual intimate image; SbD: Safety by Design; TID: trauma-informed design; UX: user experience.



A globally accessible NCII removal platform was selected as the case based on its content-hashing technology, international user base, and public accessibility [26]. Platform-identifying details are anonymized to protect ongoing operations.

XSL•FO
RenderX

Persona Methodology and User Journey Mapping

The persona methodology [27] creates fictional yet research-grounded user profiles to anchor design decisions in human needs rather than abstract specifications. Personas range along a continuum from assumption-based to data-driven [28]. Our personas were constructed from an evidence synthesis of published findings on NCII victimization, trauma psychology, and adolescent digital risk, rather than from direct survivor interviews, reflecting a deliberate ethical decision to avoid premature survivor involvement in an early-stage framework study [1,29]. Acknowledging the stereotyping risks inherent in personas constructed without direct user participation [28,30], we used 2 mitigations: grounding every persona characteristic in published empirical findings and subjecting personas to preliminary plausibility assessment through CQR with domain-informed student evaluators.

User journey mapping [31] visualizes emotional, cognitive, and behavioral arcs across interaction touchpoints. We adapted it from a generative design tool into a diagnostic instrument: rather than designing new interactions, we used trauma-differentiated personas to surface where existing interactions violate TIC, Safety by Design, and usability principles. This constitutes a specialized trauma-informed variant of the product risk assessments recommended in Safety by Design practice [16], enriched by clinical understanding of how trauma shapes each journey stage.

Phase 1: Persona Construction and Journey Mapping

The personas were developed by the authors through an iterative, evidence-grounded process. Candidate user vulnerabilities, help-seeking contexts, and platform interaction risks were generated and checked against evidence on NCII victimization [1,6], adolescent digital risk [32,33], and intimate partner violence [3,23]. Dimensions that recurred across these sources (trauma severity, perceived control, institutional trust, developmental status, help-seeking autonomy, and capacity to sustain a multistep digital task) differentiated the personas; perceived control was treated as central because it predicts help-seeking engagement and tolerance for uncertainty [34]. The personas were reviewed against Herman's [12] trauma continuum to ensure clinically meaningful rather than merely demographic differentiation, yielding 3 profiles presented in the Results section (Table 1): Sara (low-moderate trauma), Rina (severe trauma with complex PTSD indicators), and Minji (adolescent acute distress with developmental vulnerability).

Table . Trauma-differentiated personas.

Dimension	Sara	Rina	Minji
Age (sex)	28 years (female)	32 years (female)	16 years (female)
Context	Ex-partner shared images post-breakup	Abusive partner distributed images during prolonged IPV[a]	Peer coercion; fear of image leak
Trauma level	Low-moderate	Severe (complex PTSD[b] indicators)	Acute distress and developmental vulnerability
Key drivers	Distrust; desire for clarity	Shame, hypervigilance, and learned helplessness	Fear, confusion, and isolation
Primary goal	Remove images via transparent process	Regain control; requires emotional safety	Understand eligibility; age-appropriate support
TIC[c] principles	Trustworthiness and empowerment	Safety, empowerment, and collaboration	Trustworthiness, safety, and empowerment
Safety by Design	Transparency and provider responsibility	User empowerment and friction features	Provider responsibility and age-specific risk

[a]IPV: intimate partner violence.

[b]PTSD: posttraumatic stress disorder.

[c]TIC: trauma-informed care.

Each persona was applied to 5 interaction stages (Awareness, Consideration, Decision, Action, and Postaction) [31], using a structured matrix documenting anticipated needs, actions, barriers, proposed design responses, and the TIC, Safety by Design, and usability principles implicated. At each stage, the authors recorded both what a standard usability evaluation (Nielsen's heuristics) would flag and what clinical trauma knowledge (TIC and Safety by Design) would flag, surfacing features that can pass conventional heuristic evaluation while failing trauma-informed evaluation. The authors completed the matrix independently and reconciled differences through structured discussion; claims that could not be reconciled were narrowed or excluded. The resulting journey maps appear in the Results section (Table 2).

XSL•FO
RenderX

Table . Trauma-differentiated user journey maps.

Persona and stage	Needs	Action	Current barrier	TID[a]-informed response	TIC[b]	Safety by design
Sara (low-moderate)						
Awareness	Find credible platform	Searches removal	Difficult to locate via search	Visible entry points; onboarding	Trustworthiness	Transparency
Consideration	Understand function	Reads home page cautiously	Governance and data handling unclear	Plain-language data explanations	Trustworthiness	Transparency
Decision	Reassurance; error protection	Attempts case creation	No preview; fears irreversibility	Gentle prompts; preview before submission	Empowerment	Safety
Action	Privacy reassurance	Uploads content hash	Upload triggers shame; no support	Calming messages; grounding tools	Safety	Protective defaults
Postaction	Affirmation; status visibility	Waits for confirmation	Insufficient status visibility	Affirming follow-up; status tracking	Trustworthiness	Accountability
Rina (severe)						
Awareness	Welcoming language	Hesitant search; delays	No welcoming or survivor voice	Affirming, survivor-centered entry	Safety, empowerment	Proactive
Consideration	Non-triggering content	Briefly visits platform	Elements not validated as safe	Validated non-triggering UI[c]	Safety	Protective defaults
Decision	Control; ability to pause	Attempts case but leaves	Rigid; no save; no support	Grounding techniques; save draft	Empowerment	Friction features
Action	Pause without losing progress	Reattempts upload	No support; solitary completion	Pacing indicators; helplines	Safety, collaboration	Empowerment
Postaction	Long-term support pathways	Seeks external support	No pathways to long-term support	Connect to NGOs[d], peer networks	Collaboration	Holistic
Minji (adolescent)						
Awareness	Age-appropriate guidance	Searches due to fear	Unclear guidance for minors	Messaging for minors; youth links	Trustworthiness	Age-specific risk
Consideration	Verify eligibility	Reads cautiously	FAQ–UI[c] conflict	Simplified FAQ[e]; consistent messaging	Trustworthiness	Transparency
Decision	Clarity on access	Looks for eligibility	No minor-specific jurisdiction info	Age-specific legal guidance	Empowerment	Provider responsibility
Action	Support if restricted	Attempts case creation	Block; no guidance or support	Clear alternative steps; reassurance	Collaboration	Empowerment
Postaction	Next-step guidance	Redirected to generic list	Generic list; no minor help	Visible helplines; tailored follow-up	Empowerment	Proactive

[a]TID: trauma-informed design.

[b]TIC: trauma-informed care.

[c]UI: user interface.

[d]NGOs: nongovernmental organizations.

[e]FAQ: frequently asked questions.

Journey findings were translated into low-fidelity TID-informed interface examples for use as phase 2 evaluation stimuli, each reviewed through scenario-based cognitive walk-throughs [35]; these examples were not treated as final design solutions. Crosscutting barrier themes were then derived by grouping recurrent persona-stage barriers and cross-walking each against the TIC principle, Safety by Design principle, and usability heuristic it implicated. The 6 TID guidelines were synthesized from these themes by consolidating barriers with shared design implications into broader intervention targets (eg, opaque data handling, unclear steps, and uncertain timelines into process transparency). Each of the 5 crosscutting barriers was mapped to 1 or more guidelines (Table 3), and every guideline was anchored to the TIC principle and Safety by Design mandate it operationalizes; guideline derivation and labeling followed the

XSL•FO
RenderX

same independent-coding-then-consensus procedure used for journey mapping.

Table . TID[a] guideline framework: derivation and evaluated design responses[b].

Guideline (domain)	Barrier addressed	Persona(s) most affected	TIC[c]	Safety by design	Usability-heuristic overlap	Design response	Interpretive status (phase 2)
G1: Emotional scaffolding (visual and emotional tone)	Emotionally triggering content	Rina, and Sara	Safety	Service-provider responsibility	Aesthetic and minimalist design	Reduce threat cues; supportive language	Perceived as more comforting
G2: Process transparency (language and framing)	Opaque data handling; ambiguous eligibility language	Sara, and Rina	Trustworthiness and transparency	Transparency	System status visibility	Make process, data use, and timelines explicit	Perceived as trust-supporting
G3: Self-pacing mechanisms (self-pacing and grounding)	Multi-step workflows without exit	All	Empowerment, voice, and choice	User empowerment	User control; help documentation	Enable pause, save, and return without penalty	Perceived as reducing pressure
G4: Institutional trust signals (information architecture)	Low institutional credibility	All	Trustworthiness and transparency	Service-provider responsibility	(No direct equivalent)	Show recognizable affiliations and security cues	Consistently perceived as trust-supporting
G5:Context-aware support (information architecture)	Absent tailored or contextual support	Minji, and Sara	Cultural, historical, and gender issues	Service-provider responsibility	Consistency; system–-world match	Tailor guidance to age, jurisdiction, and need	Identified as relevant; requires survivor-centered testing
G6: Grounding tools (self-pacing and grounding)	Absent emotional support	All	Peer support	User empowerment	User control; help documentation	Provide grounding and immediate support options	Perceived as usable; not behavioral evidence

[a]TID: trauma-informed design.

[b]G1-G6 denote the 6 TID guidelines. Columns map each guideline to the barrier it addresses, the personas most affected, its TIC principle and Safety by Design mandate, and the overlapping usability heuristic [9]; the final 2 columns give the candidate design response and its phase 2 interpretive status. Design responses were evaluated as plausibility and acceptability stimuli by domain-informed student evaluators, not tested behaviorally with nonconsensual intimate image victim-survivors or active help seekers.

[c]TIC: trauma-informed care.

Phase 2: Domain-Informed Plausibility Evaluation

Eight female university students (aged 20 - 24 years) majoring in Counseling-UX Psychology at a Korean university were recruited through purposive sampling as domain-informed student evaluators. Participants were selected for their uncommon dual training in counseling psychology and UX: few individuals, and particularly established specialists, command both, yet this combination is what allows an evaluator to read an interface through both the usability-heuristics and the TIC lenses the framework integrates. As young women in the age range most affected by NCII and image-based sexual abuse, they were also demographically closer to the affected user population than older specialists would be, which supported empathic appraisal of the personas' experiences. They were not recruited as credentialed experts, NCII survivors, representative end users, or individuals actively seeking image-removal support. Because trauma-informed digital safety design is an emerging interdisciplinary area and formally established specialist roles are limited in the Korean context, this group was used for preliminary plausibility assessment rather than expert validation. The 8 evaluators were distinct from the persona development team, preserving the independence of the appraisal. Prior to the focus group, participants received a structured briefing on the study purpose, the NCII removal platform context, the 3 persona summaries, a brief explanation of the TID framework and its 6 guidelines, and the distinction between evaluating plausibility and reporting personal trauma experience. The 1-hour semistructured focus group was conducted in a private university setting by a researcher trained in Counseling-UX Psychology. The focus group guide covered four blocks: (1) initial platform navigation and spontaneous reactions; (2) evaluation of persona and journey plausibility; (3) side-by-side comparison of the original interface and TID-informed interface examples instantiating the 6 guidelines, including whether the design responses cohere with the framework; and (4) reflection on emotional safety and willingness to use. The focus group guide is provided in Multimedia Appendix 1. Audio was transcribed verbatim in Korean with representative quotes translated into English by a bilingual researcher and verified by a second.

CQR was selected for its emphasis on researcher consensus through team-based analysis and its demonstrated suitability for exploratory research with small, homogeneous samples [25,36]. Analysis proceeded through 5 steps. First, the 2 coders independently reviewed the transcript and segmented participant responses into meaning units. The 2 coders were a master's-level

XSL•FO
RenderX

practitioner in counseling psychology and the first author (JL), a licensed counseling psychologist with expertise in trauma, digital mental health, and human-computer interaction; assumptions and potential biases were documented prior to analysis following CQR protocol [25] and trauma-informed research guidelines [1]. Second, each analyst abstracted the essential meaning of each excerpt into preliminary core ideas, preserving the participant's intended meaning. Third, the analysis team met to compare core ideas, discuss discrepancies, and reach consensus through iterative discussion. Disagreements were resolved by returning to the transcript and evaluating which interpretation was most strongly supported by the raw data. Fourth, agreed-upon core ideas were clustered into thematic domains and categories through cross-analysis across participants. Fifth, an external auditor, who is a counseling psychology faculty member specializing in digital mental health and was not involved in data collection, reviewed the thematic domain structure, category labels, and representative excerpts, in line with established qualitative trustworthiness criteria [37]. Auditor feedback was discussed by the analysis team, and revisions were made when the audit identified overinterpretation, unclear category boundaries, or insufficient evidentiary support. Frequency labels followed CQR conventions: general (7 - 8 of 8 participants), typical (4-6), and variant (2-3).

Ethical Considerations

This study was reviewed and approved by the Institutional Review Board of Hankuk University of Foreign Studies (approval number HUFS-2507 - 009). It was conducted as a preliminary, minimal-risk UX and plausibility evaluation with adult student evaluators; participants were not recruited as NCII victim-survivors, were not asked to disclose personal NCII victimization or trauma histories, and evaluated summarized persona, journey map, and interface-example materials rather than reporting personal experience. Written informed consent was obtained before participation; participants were advised in advance that the evaluation materials addressed NCII and could be distressing, that they could decline any question or pause or withdraw at any time without penalty, and that psychological support was available, if needed. Privacy and confidentiality were protected by deidentifying transcripts, reporting summarized findings rather than identifiable raw statements, and limiting data access to the research team. Safeguards included advance briefing, unrestricted withdrawal rights, on-site counselor availability during and after the session, and summarized rather than raw presentation of potentially triggering NCII content [1]. No compensation was provided.

Results

Phase 1: Personas, Journey Maps, and Barrier Identification

Phase 1 produced 3 trauma-differentiated personas, summarized in Table 1. Each persona encodes a distinct configuration of trauma severity, perceived control, institutional trust, and capacity to sustain a multistep digital task under stress. Applying each persona across the 5 interaction stages produced the journey maps presented in Table 2, which document, for each stage, anticipated needs, likely actions, the current barrier, a TID-informed response, and the corresponding TIC and Safety by Design principles.

For Sara (low-moderate trauma), absent procedural clarity and opaque data governance violated Trustworthiness and Safety by Design's Transparency principle [23,38,39]. Although a general population user might tolerate ambiguous data-handling information as a minor frustration, Sara's trauma-eroded institutional trust transforms opacity from a usability irritation into a barrier to engagement. Specifically, she cannot proceed without understanding exactly how her sensitive data will be handled, by whom, and under what legal protections. For Rina (severe trauma), the platform's rigid, linear workflow, which lacked any save-progress functionality, forced a complete restart after trigger-induced abandonment. Given that PTSD specifically disrupts executive functions including planning, sustained attention, and working memory [11], this rigid workflow created a functionally inaccessible environment for users in exactly the crisis state the platform was designed to serve. The misalignment was dual: TIC's Empowerment, Voice, and Choice principle requires that users maintain control over the pace and direction of their engagement, while Safety by Design's friction feature guidance recommends pause points and save states precisely to prevent forced completion designs [16]. For Minji (adolescent), ambiguous eligibility criteria violated Trustworthiness: a minor navigating a confusing legal landscape needs unambiguous guidance, not contradictory information across frequently asked questions and interface elements. Generic referral on restriction (ie, redirecting a distressed minor to an undifferentiated list of organizations without age-specific guidance) violated both Safety and Safety by Design's age-specific risk assessment requirement [40,41].

Five crosscutting barrier themes were synthesized across all 3 journey maps (Table 3): ambiguous eligibility language, emotionally triggering content, absent emotional support mechanisms, opaque data handling, and low institutional credibility. Each theme represents a convergent pattern across clinical, platform-safety, and usability lenses rather than a single-source observation; the full derivation appears in Table 3.

Proposed TID Framework

The 6-guideline TID framework (Table 3) integrates TIC and Safety by Design principles across 4 intervention domains: Visual and Emotional Tone, Language and Framing, Information Architecture, and Self-Pacing and Grounding. The framework is specified at the level of design principles and the psychological targets they address; the concrete interface features named under each guideline are illustrative instantiations used here as evaluation stimuli, rather than prescribed or finalized designs for any particular service.

Guideline 1 (Emotional Scaffolding): calming color palettes, nontriggering imagery, and empathetic language implemented as platform defaults rather than user-selected options. This guideline targets trauma-related hyperarousal and the appraisal of otherwise neutral interface cues as threatening [12,13]. Guideline 2 (process transparency): plain-language explanations of data handling, process steps, and expected timelines at every interaction stage. It addresses the hypervigilant threat monitoring

XSL•FO
RenderX

associated with eroded institutional trust while reducing uncertainty that taxes trauma-compromised working memory [11,12]. Guideline 3 (Self-Pacing Mechanisms): save-as-draft functionality, visible progress indicators, and explicit pause points that validate the user's need to step away. It counters learned helplessness and the PTSD-related executive-function deficits, such as planning, sustained attention, and working memory, which cause rigid, multistep workflows to fail for trauma-affected users [11,12]. Guideline 4 (Institutional Trust Signals): prominent display of recognized organizational affiliations, government endorsements, and security certifications. It responds to trauma-eroded institutional trust, for which conventional credibility signals are often insufficient [12]. Guideline 5 (Context-Aware Support): curated resources tailored to user characteristics, including age-specific guidance, jurisdiction-relevant legal information, and culturally appropriate referrals. To operationalize the cultural, historical, and gender issues principle, this guideline further specifies gender- and culturally responsive content, which includes language that explicitly counters victim-blaming, a concern especially salient in the Korean sociocultural context. For instance, displaying a message such as "What do you do if someone is threatening to share your intimate images?" on the main page could potentially place the burden on victims who are already navigating a difficult reporting process. This framing may inadvertently suggest that resolving the situation falls primarily on the victim, which might feel less validating for those seeking support. To avoid this, the guideline mandates that the guidance never implicitly attributes responsibility to the victim-survivor [2,6]. Guideline 6 (Grounding Tools): accessible helpline links, supportive messaging during high-anxiety interaction points, and mobile-first crisis design reflecting the platforms through which victim-survivors most commonly seek help [5]. Consistent with the Peer Support principle, it also offers optional connection to vetted survivor and peer-support networks to address the isolation that frequently accompanies NCII victimization [42]; its grounding components target acute hyperarousal that can exceed a user's window of tolerance at high-distress steps [12,13]. To prevent specific identification of the platform, visual interface details are omitted; instead, the final columns of Table 3 summarize the candidate design responses used as phase 2 evaluation stimuli.

Phase 2: Domain-Informed Evaluation Results

Overview

CQR analysis yielded 5 thematic domains and 14 categories (Table 4). Following CQR conventions, frequency labels are used descriptively to indicate how many participants endorsed a category within this specific evaluator sample. Given the small and homogeneous sample, "General" should not be interpreted as population-level prevalence or generalizability. Rather, it indicates that a theme was consistently observed across the 8 domain-informed student evaluators in this preliminary study.

XSL•FO
RenderX

Table . CQR[a] results: thematic domains and categories[b].

Thematic domain and category	Core idea	Frequency	Representative quote
Initial perceptions			
Unclear site purpose	Users could not discern site purpose	General	"I had no idea what this site was for."
Visual distrust	Design felt untrustworthy	Typical	"The color scheme made it feel cheap."
Cognitive overload	Excessive information caused confusion	Typical	"So much information, I didn't know what to click."
Persona credibility			
Psychological plausibility	Trauma-informed behavior felt realistic	General	"Of course they'd want to avoid repeating steps."
Behavioral journey fit	Reactions matched experience	Typical	"Wanting to double-check safety—that resonated."
Need for diversity	More diverse personas needed	Variant	"Maybe add a male victim persona."
UX[c] feedback			
Trust-building features	Logos created confidence	General	"The official logo made me trust it more."
Comforting elements	Language supported ease	Typical	"'Find support' felt much less intimidating."
Improved flow	Save feature helped	Typical	"Saving progress made a huge difference."
Emotional safety			
Trust based on prior knowledge	Trust needed context	General	"Without background info, I wouldn't trust it."
Fear of fake sites	Impersonation fears	Typical	"I'd question if this was real."
Institutional endorsement	Official affiliation needed	Typical	"Endorsed by a government agency."
Systemic barriers			
Search invisibility	Not found in search	General	"I searched and it doesn't show up."
Desire for human support	Real-time help wanted	Variant	"A chat option mid-process would help."

[a]CQR: consensual qualitative research.

[b]General=7-8 of 8 participants, typical=4-6, and variant=2-3.

[c]UX: user experience.

Domain 1: Initial Perceptions

The platform's purpose was not readily apparent to participants (general). Aesthetic elements, particularly color choices and layout density, elicited active distrust rather than merely failing to inspire confidence (typical). Information density created cognitive paralysis, with participants reporting inability to identify relevant next steps (typical).

Domain 2: Persona Credibility

Personas were received as psychologically plausible, particularly regarding trauma-shaped behaviors: repeated safety checking before commitment, hesitation at irreversible steps, and need for institutional legitimacy verification (general). Journey narratives were emotionally resonant (typical), with participants reporting that they could envision users exhibiting the personas' described hesitations and retreat patterns. The variant-level concern about exclusively female representation warrants attention in future iterations. General-level endorsement provides preliminary support for the plausibility of the diagnostic methodology within this evaluator group, but it does not establish ecological validity with NCII victim-survivors. This finding partially addresses the methodological concern that personas constructed without direct user research risk stereotyping [28,30], while also underscoring the need for subsequent survivor-centered evaluation. Furthermore, domain-informed plausibility rated by participants trained in the same field presents a lower bar than it might appear; establishing external validity ultimately requires evaluators outside this disciplinary overlap.

Domain 3: UX Design Feedback

Institutional logos transformed perceptions from "potentially fraudulent" to "officially endorsed," with participants describing

XSL•FO
RenderX

a categorical rather than incremental shift in willingness to engage (general), supporting guideline 4. Language changes that replaced "Create a case" with "Find support" were perceived as substantially more comforting and less bureaucratic (typical), supporting guideline 1. The save-as-draft function received positive responses (typical), with multiple participants noting that the original rigid workflow would be prohibitive during emotional distress, providing preliminary support for the clinical rationale underlying guideline 3. This convergence between clinical prediction and evaluator feedback suggests that mental health expertise may improve early-stage safety design, although behavioral effectiveness remains untested.

Domain 4: Emotional Safety

Trust was conditional on external endorsement: participants required a trusted intermediary (counselor, teacher, and government website) to vouch for the platform before engaging, regardless of interface quality (general). Fraudulent website fears persisted even after viewing framework-informed interface examples with improved trust signals (typical).

Domain 5: Systemic Barriers

Search engine invisibility was the most prominent systemic concern (general): participants could not locate the platform through standard search queries, rendering interface improvements irrelevant for users who never arrive. Desire for real-time human support was noted (variant), consistent with TIC's Collaboration and Mutuality principle. These findings identify structural prerequisites beyond interface design, including search engine optimization strategy, institutional referral partnerships, and integrated support service connections, consistent with the argument by Chen et al [1] that trauma-informed technology requires systemic organizational transformation, not merely interface-level changes, and with Safety by Design's Holistic principle requiring safety considerations across the full product life cycle [16].

Discussion

Principal Findings

This study developed and preliminarily evaluated a TID framework for NCII removal platforms. Applying trauma-differentiated personas to an existing platform (RQ1) surfaced 5 crosscutting barriers, including ambiguous eligibility language, emotionally triggering content, absent emotional support, opaque data handling, and low institutional credibility, which conventional usability evaluation alone may not fully capture. Jointly operationalizing TIC, Safety by Design, and usability heuristics (RQ2) produced 6 TID guidelines. In domain-informed evaluation (RQ3), evaluators judged the personas plausible and viewed design responses featuring institutional trust signals, self-pacing, and supportive language as emotionally safer and potentially more usable, while emphasizing that survivor-centered validation remains necessary. The study's primary contribution is to operationalize the Safety by Design mandate for NCII removal platforms: pairing TIC, as clinical content, with usability heuristics, as interaction mechanics, to turn a broad protective obligation into concrete, actionable design guidance, thereby helping to close the gap between safety policy frameworks and interface-level guidance.

TIC as a Foundational Framework for Digital Safety Design

The central finding is that the barriers that arise on NCII removal platforms are rooted in the psychological consequences of trauma, and that TIC provides the theoretical apparatus necessary to diagnose them. The rigid, save-progress-free workflow in Rina's journey map illustrates this: in usability terms a minor violation of Nielsen's [9] user-control heuristic, but through a TIC lens a reproduction of learned helplessness, as effort expended and then arbitrarily erased mirrors the powerlessness Herman [12] identified as central to prolonged interpersonal trauma, inadvertently replicating how abusers exploit technology to maintain control [23]. Likewise, ambiguous data-handling policies (a transparency gap in usability terms) activate hypervigilant threat monitoring in users whose institutional trust is compromised [3], and information-dense interfaces (an aesthetic-minimalism violation) exceed the diminished executive function capacity documented in PTSD [11]. This pattern (the same feature receiving qualitatively different diagnoses depending on whether TIC principles are applied) illustrates why TIC is foundational to this diagnostic approach. Prior work mapped heuristic-TIC convergences, demonstrating that the 2 traditions share substantial common ground [14].

Consistent with this interpretation, evaluators' difficulty discerning the platform's purpose and their distrust of information-dense screens map onto the trauma-compromised attention, working memory, and executive function documented in PTSD populations [11], such that information-dense interfaces function as inaccessible rather than merely suboptimal environments. Likewise, the conditioning of trust on external endorsement reflects trauma-eroded institutional trust [12]: for trauma-affected users, procedural transparency is necessary but insufficient without validation from trusted institutions, implying that referral partnerships with counseling services, government agencies, and NGOs may matter more for initial engagement than interface improvements alone.

A Diagnostic Methodology Bridging 3 Traditions

The combination of trauma-differentiated personas with TIC-principled journey mapping provides a methodology that may surface concerns invisible to any single tradition. General-level persona endorsement provides preliminary support for the approach within this domain-informed student evaluator group and demonstrates the unreliability of persona construction without clinical grounding [30]. Centering victim-survivor experiences remains important for digital tool development in the gender-based violence context [2]. The present work does not substitute for such survivor involvement; our contribution is complementary: TID provides a diagnostic method for an earlier phase when direct survivor involvement may not yet be ethically appropriate, producing evidence-grounded foundations that subsequent survivor-centered evaluation can refine and validate. The persona methodology serves as the integration mechanism, whereby personas embody clinical trauma profiles [12,13] in a format that UX practitioners can apply through

XSL•FO
RenderX

journey mapping and that safety engineers can evaluate against risk frameworks [16]. Because clinicians, UX practitioners, and safety engineers each work with diagnostic tools the others lack, surfacing the full range of barriers requires integrating the 3 perspectives rather than relying on any single tradition.

Implications for Policy and Platform Design

These findings also reframe the purpose of TID. Its value lies less in serving as an ethics checklist or a compliance layer added to an existing system than in functioning as psychologically realistic interaction design that anticipates the predictable states of a help-seeking population (reduced control, avoidance, dissociation, and the anticipated cost of disclosure or refusal) and specifies how an interface should respond when a user's capacity to negotiate, object, or reengage is compromised. On this view, the features evaluators endorsed most strongly (institutional trust signals, self-pacing mechanisms, and grounding supports) function not as courtesies but as the system's relational response, situating TID within the interaction model rather than in supplementary policy. This framing extends the logic of survivor-centered conversational systems [43,44] from stand-alone agents to the removal workflow itself.

TID also continues a cumulative design lineage (from Privacy and Security by Design to Safety by Design) that progressively widens the categories of user vulnerability a system is expected to anticipate. As the Digital Services Act [45] and Online Safety Act [46] translate Safety by Design into regulatory obligation, TID offers a candidate interaction-level method for rendering those obligations as user-facing criteria, although transferability should be claimed cautiously. Building on calls for empirically evaluable trauma-informed frameworks [5], we propose a standardized TID heuristic spanning interface-level criteria (emotional tone, cognitive load, and pacing) and systemic criteria (referral pathways and institutional partnerships), applied alongside conventional heuristic evaluation and with clinical expertise embedded in requirements definition rather than added as post hoc consultation.

Limitations and Future Directions

Several limitations warrant acknowledgment. First, the framework was not evaluated with NCII victim-survivors or with users actively seeking image-removal support. This decision was ethically motivated, given the sensitivity of NCII, the possibility of retraumatization, and the need to establish preliminary procedural safety before involving survivors. However, it constrains the ecological validity of the present findings. Future research should evaluate the framework with NCII victim-survivors through ethically robust, survivor-centered methods [2,47]. Second, phase 2 involved a single focus group with 8 female Counseling-UX Psychology university student evaluators. CQR frequency labels provide useful within-sample descriptors, but they do not establish saturation across populations or prevalence estimates. The small, homogeneous, domain-informed student sample may respond differently from victim-survivors, users without counseling-UX training, users with lower digital literacy, or users in acute crisis. Third, all personas and participants were female. Although this reflects the strongly gendered distribution of NCII abuse [6], in nationwide data, women are nearly twice as likely as men to be targeted by nonconsensual intimate imagery [48], and the overwhelming majority of sexual deepfakes is nonconsensual and almost exclusively depicts women [49]. Nonetheless, the variant-level diversity concern raised by participants underscores the need to broaden future persona sets across gender, sexuality, age, disability status, and cultural contexts. Fourth, the framework was developed through a single NCII-focused case and should not be treated as fully transferable to all trauma-related digital services; some findings are likely NCII-specific (eg, fear of image circulation and shame around uploading intimate material), whereas elements such as procedural transparency, user-controlled pacing, and grounding support may transfer to adjacent digital safety contexts. Finally, the dual-lens method is, by design, theory-driven: it applies TIC as an analytic lens to surface barriers that usability evaluation alone tends to leave unexamined. Its contribution is accordingly generative (identifying and specifying candidate barriers and design responses) rather than confirmatory, and establishing that a trauma-informed lens yields outcomes superior to conventional evaluation would require controlled comparative testing, which we identify as a priority for subsequent work.

Conclusions

Emotional safety is a design requirement, not an enhancement: for platforms meant to help people already harmed, designing for the user who cannot readily advocate for themselves is what separates a tool that protects from one that may reproduce the original loss of control. Jointly operationalized through a clinically informed, persona-based diagnostic method, TIC, Safety by Design, and usability principles surfaced psychological gaps that no single method detects alone; the 6 guidelines are an actionable but preliminary starting point, and the personas and journey maps show how Safety by Design can be applied concretely rather than as aspiration. The framework now requires survivor-centered, cross-cultural, and behavioral evaluation before claims of effectiveness or broad transferability. Embedding these principles in removal platforms may enhance both their usability and their adoption, increasing the likelihood that the support these systems are designed to provide reaches the victim-survivors who seek it.

Acknowledgments

The authors would like to thank Sumin Kim, Yeonju Jeon, and the other students who participated in the user experience course and contributed valuable insights to this study. The authors also thank Eunseo Kim, MA, a counseling psychologist, for serving as an independent second coder in the consensual qualitative research analysis. The authors declare the use of generative AI (Gemini) for grammar and language improvement during manuscript preparation. It was not used to generate intellectual content of the study. Responsibility for the final manuscript lies entirely with the authors.

XSL•FO
RenderX

Funding

This research was supported by the Hankuk University of Foreign Studies Research Fund. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2025S1A5C3A01008166).

Data Availability

The datasets analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: JL
Methodology: JL (lead), YK (supporting)
Resources: JL
Data collection: YK
Data analysis: JL
Plausibility evaluation: JL
Writing – original draft: JL (lead), YK (supporting)
Writing – review & editing: JL (lead), YK (supporting)
Supervision: JL

Conflicts of Interest

None declared.

Multimedia Appendix 1
Focus group guide.
[DOCX File, 20 KB - humanfactors_v13i1e94438_app1.docx]

References

1. Chen JX, McDonald A, Zou Y, et al. Trauma-informed computing: towards safer technology experiences for all. Presented at: CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; Apr 30 to May 5, 2022; New Orleans, LA p. 1-20. [doi: 10.1145/3491102.3517475]
2. Henry N, Witt A, Vasil S. A 'design justice' approach to developing digital tools for addressing gender-based violence: exploring the possibilities and limits of feminist chatbots. Inf Commun Soc 2025 Aug 18;28(11):1884-1907. [doi: 10.1080/1369118X.2024.2363900]
3. Matthews T, O'Leary K, Turner A, et al. Stories from survivors: privacy and security practices when coping with intimate partner abuse. Presented at: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems; May 6-11, 2017; Denver, CO p. 2189-2201. [doi: 10.1145/3025453.3025875]
4. Safety by design puts user safety and rights at the centre of the design and development of online products and services. eSafety Commissioner. URL: https://www.esafety.gov.au/industry/safety-by-design [accessed 2026-02-23]
5. Eggleston M, Jones EP, Khan N, et al. A scoping review of trauma-informed care principles applied in design and technology. Digit Health 2025;11:20552076251360925. [doi: 10.1177/20552076251360925] [Medline: 40735549]
6. Henry N, McGlynn C, Flynn A, Johnson K, Powell A, Scott AJ. Image-Based Sexual Abuse: A Study on the Causes and Consequences of Non-Consensual Nude or Sexual Imagery: Routledge; 2020. [doi: 10.4324/9781351135153]
7. De Angeli A, Falduti M, Menendez-Blanco M, Tessaris S. Reporting non-consensual pornography: clarity, efficiency and distress. Multimed Tools Appl 2023 Apr;82(9):12829-12858. [doi: 10.1007/s11042-022-14291-z]
8. Trauma-informed care in behavioral health services (Treatment Improvement Protocol Series 57). : HHS Publication No. (SMA) 13-4801. Substance Abuse and Mental Health Services Administration, Center for Substance Abuse Treatment; 2014 URL: https://www.refugees.org/wp-content/uploads/2022/02/sma14-4816.pdf [accessed 2026-07-18]
9. Nielsen J. Enhancing the explanatory power of usability heuristics. Presented at: CHI '94: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; Apr 24-28, 1994; Boston, MA p. 152-158. [doi: 10.1145/191666.191729]
10. Qiwei L, McDonald A, Haimson OL, Schoenebeck S, Gilbert E. The sociotechnical stack: opportunities for social computing research in non-consensual intimate media. Proc ACM Hum Comput Interact 2024;8:1-21. [doi: 10.1145/3686914]
11. Samuelson KW, Neylan TC, Metzler TJ, et al. Neuropsychological functioning in posttraumatic stress disorder and alcohol abuse. Neuropsychology 2006 Nov;20(6):716-726. [doi: 10.1037/0894-4105.20.6.716] [Medline: 17100516]
12. Herman JL. Trauma and Recovery: Basic Books; 1992. URL: https://ia803207.us.archive.org/14/items/radfem-books/Trauma%20and%20Recovery_%20The%20Afterm%20-%20Judith%20L.%20Herman.pdf [accessed 2026-07-14]

XSL•FO
RenderX

13. Zaleski KL, Johnson DK, Klein JT. Grounding Judith Herman's trauma theory within interpersonal neuroscience and evidence-based practice modalities for trauma treatment. Smith Coll Stud Soc Work 2016 Oct;86(4):377-393. [doi: 10.1080/00377317.2016.1222110]
14. Eggleston M, Noel LA. Trauma-informed design: leveraging usability heuristics on a social services website. J User Exp 2024;19(3):123-138 [FREE Full text]
15. Fallot RD, Harris M. Creating cultures of trauma-informed care (CCTIC): a self-assessment and planning protocol. Community Connections. 2009. URL: https://aeoworks.org/wp-content/uploads/2020/05/Creating-Cultures-of-Trauma-Informed-Care.pdf [accessed 2026-06-18]
16. Trust and Safety Professional Association. Safety by design: a framework for trust and safety professionals. 2024. URL: https://www.tspa.org/curriculum/ts-fundamentals/safety-by-design/ [accessed 2026-02-23]
17. Kilpatrick DG, Resnick HS, Milanak ME, Miller MW, Keyes KM, Friedman MJ. National estimates of exposure to traumatic events and PTSD prevalence using DSM-IV and DSM-5 criteria. J Trauma Stress 2013 Oct;26(5):537-547. [doi: 10.1002/jts.21848] [Medline: 24151000]
18. Benjet C, Bromet E, Karam EG, et al. The epidemiology of traumatic event exposure worldwide: results from the World Mental Health Survey Consortium. Psychol Med 2016 Jan;46(2):327-343. [doi: 10.1017/S0033291715001981] [Medline: 26511595]
19. Bellis MA, Hughes K, Ford K, Ramos Rodriguez G, Sethi D, Passmore J. Life course health consequences and associated annual costs of adverse childhood experiences across Europe and North America: a systematic review and meta-analysis. Lancet Public Health 2019 Oct;4(10):e517-e528. [doi: 10.1016/S2468-2667(19)30145-8] [Medline: 31492648]
20. Frazier PA. Perceived control and distress following sexual assault: a longitudinal test of a new model. J Pers Soc Psychol 2003 Jun;84(6):1257-1269. [doi: 10.1037/0022-3514.84.6.1257] [Medline: 12793588]
21. Ullman SE, Peter-Hagene L. Social reactions to sexual assault disclosure, coping, perceived control and PTSD symptoms in sexual assault victims. J Community Psychol 2014 May 1;42(4):495-508. [doi: 10.1002/jcop.21624] [Medline: 24910478]
22. Campbell R, Dworkin E, Cabral G. An ecological model of the impact of sexual assault on women's mental health. Trauma Violence Abuse 2009 Jul;10(3):225-246. [doi: 10.1177/1524838009334456] [Medline: 19433406]
23. Freed D, Palmer J, Minchala D. "A Stalker's Paradise": how intimate partner abusers exploit technology. Presented at: CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems; Apr 21-26, 2018; Montréal, QC p. 1-13. [doi: 10.1145/3173574.3174241]
24. Creswell JW, Plano Clark VL. Designing and Conducting Mixed Methods Research, 3rd edition: SAGE; 2018.
25. Hill CE, Knox S, Thompson BJ, Williams EN, Hess SA, Ladany N. Consensual qualitative research: an update. J Couns Psychol 2005;52(2):196-205. [doi: 10.1037/0022-0167.52.2.196]
26. Stake RE. The Art of Case Study Research: SAGE Publication; 1995. URL: https://uk.sagepub.com/en-gb/eur/the-art-of-case-study-research/book4954#description [accessed 2026-07-14]
27. Cooper A. The Inmates Are Running the Asylum: Why High-Tech Products Drive Us Crazy and How to Restore the Sanity: Sams Publishing; 1999. URL: https://archive.org/details/inmatesarerunnin00coop [accessed 2026-07-14]
28. Nielsen L. Personas—User Focused Design, 2nd edition: Springer; 2019. [doi: 10.1007/978-1-4471-7427-1]
29. Lazar J, Feng JH, Hochheiser H. Research Methods in Human-Computer Interaction, 2nd edition: Morgan Kaufmann; 2017. URL: https://soclibrary.futa.edu.ng/books/Research%20Methods%20in%20Human%20Computer%20Interaction%20by%20Jonathan%20Lazar,%20Jinjuan%20Feng%20and%20Harry%20Hochheiser%20(Auth.)%20(z-lib.org).pdf [accessed 2026-07-14]
30. Chapman CN, Milham RP. The personas' new clothes: methodological and practical arguments against a popular method. Proc Hum Fact Ergon Soc Annu Meet 2006 Oct;50(5):634-636. [doi: 10.1177/154193120605000503]
31. Howard T. Journey mapping: a brief overview. Commun Des Q Rev 2014 May 1;2(3):10-13. [doi: 10.1145/2644448.2644451]
32. Carcelén‐García S, Márquez OC, Vilches MD. Online risk behaviours among adolescents: identifying areas of digital vulnerability. Child Soc 2026 May;40(3):474-488. [doi: 10.1111/chso.12998]
33. Orben A, Meier A, Dalgleish T, Blakemore SJ. Mechanisms linking social media use to adolescent mental health vulnerability. Nat Rev Psychol 2024;3(6):407-423. [doi: 10.1038/s44159-024-00307-y]
34. Lee JY. Silenced by non-consensual deepfakes? Perceived threat and online self-censorship among female users: mediating roles of perceived control, belief in a just world, and passive SNS use. Telemat Inform Rep 2026 Mar;21:100279. [doi: 10.1016/j.teler.2025.100279]
35. Nielsen J, Molich R. Heuristic evaluation of user interfaces. 1990 Presented at: CHI '90: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; Apr 1-5, 1990; Seattle, WA p. 249-256. [doi: 10.1145/97243.97281]
36. Hill CE, Thompson BJ, Williams EN. A guide to conducting consensual qualitative research. Couns Psychol 1997 Oct;25(4):517-572. [doi: 10.1177/0011000097254001]
37. Lincoln YS, Guba EG. Naturalistic Inquiry: SAGE Publication; 1985.
38. Ananny M, Crawford K. Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. New Media Soc 2018 Mar;20(3):973-989. [doi: 10.1177/1461444816676645]
39. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? Soc Sci Med 2007 May;64(9):1853-1862. [doi: 10.1016/j.socscimed.2007.01.012] [Medline: 17328998]

XSL•FO
RenderX

40. Badillo-Urquiola K, Chouhan C, Chancellor S, De Choudhary M, Wisniewski P. Beyond parental control: designing adolescent online safety apps using value sensitive design. J Adolesc Res 2020 Jan;35(1):147-175. [doi: 10.1177/0743558419884692]
41. Wisniewski P, Xu H, Rosson MB, Carroll JM. Parents just don't understand: why teens do not talk to parents about their online risk experiences. Presented at: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing; Feb 25 to Mar 1, 2017; Portand, OR p. 523-540. [doi: 10.1145/2998181.2998236]
42. Knipschild R, Covers M, Bicanic IAE. From digital harm to recovery: a multidisciplinary framework for first aid after online sexual abuse. Eur J Psychotraumatol 2025 Dec;16(1):2465083. [doi: 10.1080/20008066.2025.2465083] [Medline: 40012502]
43. Maeng W, Lee J. Designing and evaluating a chatbot for survivors of image-based sexual abuse. 2022 Presented at: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; Apr 29 to May 5, 2022; New Orleans, LA p. 1-21. [doi: 10.1145/3491102.3517629]
44. Park H, Lee J. Designing a conversational agent for sexual assault survivors: defining burden of self-disclosure and envisioning survivor-centered solutions. Presented at: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems; May 8-13, 2021; Yokohama, Japan p. 1-17. [doi: 10.1145/3411764.3445133]
45. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a single market for digital services and amending directive 2000/31/EC (Digital Services Act). European Union. URL: https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng [accessed 2026-07-14]
46. Online Safety Act 2023. Legislation UK. 2023. URL: https://www.legislation.gov.uk/ukpga/2023/50 [accessed 2026-02-23]
47. Eaton AA, Krieger MA, Siegel JA, Miller AM. Victim‐survivors' proposed solutions to addressing image‐based sexual abuse in the U.S.: legal, corporate, educational, technological, and cultural approaches. Anal Soc Iss Public Policy 2024 Aug;24(2):307-334. [doi: 10.1111/asap.12395]
48. Eaton AA, Jacobs H, Ruvalcaba Y. 2017 Nationwide online study of nonconsensual porn victimization and perpetration: a summary report. Cyber Civil Rights Initiative. 2017. URL: https://www.cybercivilrights.org/wp-content/uploads/2017/06/CCRI-2017-Research-Report.pdf [accessed 2026-06-18]
49. Ajder H, Patrini G, Cavalli F, Cullen L. The state of deepfakes: landscape, threats, and impact. : Deeptrace; 2019 Sep URL: https://regmedia.co.uk/2019/10/08/deepfake_report.pdf [accessed 2026-07-14]

Abbreviations

CQR: consensual qualitative research
NCII: nonconsensual intimate image
PTSD: posttraumatic stress disorder
RQ: research question
TIC: trauma-informed care
TID: trauma-informed design
UX: user experience

Edited by S Law; submitted 01.Mar.2026; peer-reviewed by AU Zaidi, M Gawlik-Kobylinska; revised version received 13.Jun.2026; accepted 16.Jun.2026; published 10.Aug.2026.

Please cite as:
Lee JY, Koh Y
Designing a Trauma-Informed Safety by Design Framework for Nonconsensual Intimate Image Removal Platforms: A 2-Phase Sequential Exploratory Study
JMIR Hum Factors 2026;13:e94438
URL: https://humanfactors.jmir.org/2026/1/e94438
doi:10.2196/94438

© Ji-yeon Lee, Yejun Koh. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 10.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Optimizing Usability of Digital Health Interventions for Nondigitally Native Adults: A Framework-Based Approach to Develop Just-in-Time Adaptive Interventions (JITAIs) and Other Adaptations

Abby L Cheng[1], MD, MSc; Christine Y Gou[1], MD; Adriana Martin[1], JD, MSc; Sarah M Hartz[2], MD, PhD; Joanna Abraham[3], PhD

[1]Division of Musculoskeletal Physical Medicine and Rehabilitation, Department of Orthopaedic Surgery, Washington University in St. Louis, Campus Box MSC 8233-0004-05, 660 S Euclid Ave, St. Louis, MO, United States

[2]Department of Psychiatry, Washington University in St. Louis, St. Louis, MO, United States

[3]Department of Anesthesiology and Institute for Informatics, Washington University in St. Louis, St. Louis, MO, United States

Corresponding Author:
Abby L Cheng, MD, MSc
Division of Musculoskeletal Physical Medicine and Rehabilitation, Department of Orthopaedic Surgery, Washington University in St. Louis, Campus Box MSC 8233-0004-05, 660 S Euclid Ave, St. Louis, MO, United States

Abstract

Background: Health-related technology use among nondigitally native adults is becoming widespread. Nevertheless, digital health interventions are typically not designed with the unique usability needs and preferences of this population in mind. Furthermore, most of these middle-aged and older adults are managing multiple medical conditions, which can also impact their preferences related to digital health interventions. A prime opportunity to address multimorbidity in nondigital natives using a digital health intervention is the intersection between mental health and chronic pain.

Objective: The goal of this study was to use established frameworks and end-user feedback to identify actionable, usability-related features that can be incorporated into existing digital health interventions and that are preferred among nondigitally native adults. We aimed to identify general usability adaptations, as well as just-in-time adaptive interventions (JITAIs).

Methods: We conducted a qualitative usability study of a convenience sample using a hybrid inductive-deductive content analysis to evaluate an existing mental health app (Wysa for Chronic Pain). Participants were 45 years or older; reported at least moderate symptoms of depression or anxiety (Patient Health Questionnaire -9 ≥10 or Generalized Anxiety Disorder -7 score ≥10); and endorsed having pain at least most days in the past 3 months. The Framework for Reporting Adaptations and Modifications to Evidence-based Implementation Strategies was used to identify potential usability-related adaptations for the target population. Development of usability-related JITAIs was guided by Nahum-Shani's pragmatic framework for JITAI development and the Behavioral Integration Technology model.

Results: Forty-two participants completed usability testing (mean age 57, SD 8 years; women n=32, 76%). Participants identified numerous opportunities to optimize their user experience, primarily by ensuring the app clearly describes how all its features are intended to be used and by minimizing navigation burden. Specifically, participants requested clear, step-by-step orientation and navigation instructions, rather than a brief onboarding experience that relies on user-led exploration and familiarity with conventional app symbols. Participants also recommended a customized experience based on their unique usage patterns. The most common JITAI opportunity identified was to strategically reduce user notifications in order to reduce the risk of notification fatigue and subsequent complete disengagement with the app.

Conclusions: Identifying actionable opportunities to improve usability and prompt engagement with "the right tool at the right time for the right person" holds promise to improve the effectiveness of digital health interventions across the age span. The usability-related refinement opportunities identified in this study are generalizable to other digital health interventions that are relevant to older users who are likely to be managing multimorbidity, are not digital natives, and are more likely than younger users to have physical or cognitive challenges. The integrated framework-based process described in this article can also serve as a model for optimizing other existing digital health interventions.

(JMIR Hum Factors 2026;13:e87365) doi:10.2196/87365

KEYWORDS

mental health; depression; anxiety; chronic pain; mHealth; usability; user experience; satisfaction; digital health; digital mental health intervention; just-in-time adaptive intervention; middle-age; older adult

XSL•FO
RenderX

Introduction

Most digital health interventions are not equally "accessible" to everyone [1]. Specifically, they are typically not designed with the usability needs and preferences of nondigital natives (born before 1980) in mind [2,3]. Technology use is rapidly growing across the age span [4-6], but compared to younger users who grew up "digitally native" with widespread access to technology and for whom many interventions were initially designed, middle-aged and older adults tend to have different habits and characteristics. For instance, they tend to check their mobile devices less frequently, and most do not instinctively search for a technological solution to a problem. In addition, this older population often experiences age-related changes in visual acuity and eye tracking patterns, and some have age-related dexterity or cognitive deficits [4,7-10]. Furthermore, the prevalence of multimorbidity increases with age [11]. As a result, nondigitally native adults tend to prefer different screen layouts than younger technology users; they may benefit from more "push" notifications, and they are more often interested in interventions that address multimorbidity [4,5].

A growing body of evidence is available to guide developers in understanding the usability preferences of nondigitally native adults [4,5,12], but a gap still exists in translating these general preferences into real interface changes that are implemented within existing digital health interventions. Furthermore, just-in-time adaptive interventions (JITAIs) are a relatively new and persistently underused method that aim to deliver the "right intervention at the right time to the right person," and they have the potential to reduce the need for users to initiate interaction with a digital health intervention or to navigate through the intervention independently [13]. We hypothesize that JITAIs may be particularly impactful to improve usability for nondigitally native adults, but they are not yet widely applied to publicly available digital health interventions [14]. Guiding frameworks have also been developed in recent years that can facilitate optimal reach and effectiveness of digital health interventions, but they, too, have not been routinely used in intervention development [15]. Applying these frameworks and JITAI capabilities to adapt existing interventions for key target populations could rapidly improve the accessibility and subsequent effectiveness of publicly available digital health interventions.

A prime opportunity to address multimorbidity in nondigitally native adults using a digital health intervention is at the intersection between mental health and chronic pain [16-21]. Digital mental health interventions are now widely available to address depression and anxiety symptoms [22-27], in large part because they can be effective and improve access to mental health support for people who otherwise may be limited by barriers such as stigma, cost, or geography [27-32]. Nevertheless, chronic pain commonly coexists with symptoms of depression and anxiety, and it reduces the effectiveness of mental health treatments unless pain is simultaneously addressed [20,33-38]. Furthermore, the prevalence of chronic pain increases with age [20,35].

The goal of this study was to use established frameworks and end-user feedback to identify usability preferences among nondigitally native adults for digital health interventions. We specifically focused on nondigitally native adults with chronic pain and coexisting symptoms of depression or anxiety, and we aimed to obtain feedback on participants' preferred general usability adaptations, as well as usability-related JITAIs. Our motivating scientific premise is that incorporating usability-related adaptations will improve overall engagement and subsequent effectiveness of digital health interventions, and evaluation of this potential relationship will be assessed in future work.

Methods

Study Design

This was a formative, qualitative usability study in which participants from the target population of interest completed usability testing of an existing mental health app. Usability testing occurred in 2024, and data analysis occurred in 2024 and 2025. Of note, additional feedback from study participants regarding contextual determinants of engagement with mental health apps was reported separately [12].

This study was performed via an academic-industry partnership. Academic investigators had full independence in designing the study, recruiting participants, conducting the usability testing, performing data analysis, and reporting study results. The industry partner provided access to the mental health app under investigation and created mock-ups of refined app screens and workflows for the iterative usability testing process. The academic investigators shared the study findings with the industry partner so that the industry partner could refine its mental health app accordingly.

Ethical Considerations

Ethical approval was obtained by the Washington University Institutional Review Board via a waiver of written informed consent prior to participant recruitment (Institutional Review Board ID 202311024). Participant data were maintained in secure databases and were deidentified before sharing outside the institutional review board–approved study team. Participants received a US $30 gift card for participation.

Participants or Users

Participants were at least 45 years old, reported at least moderate depression or anxiety symptoms (Patient Health Questionnaire-9 ≥ 10, Generalized Anxiety Disorder-7 score ≥ 10, or both), and endorsed chronic pain (defined as pain on most days or every day in the past three months) [20,39-41].

Exclusion criteria included endorsement of an active mental health crisis that warranted escalation of care (eg, active suicidal ideation and psychosis), cognitive impairment that would interfere with communication with the researcher or meaningful engagement with the mental health app (assessed informally by the research coordinator asking the potential participant to restate the study purpose, activities, and risks, benefits, and alternatives of study participation during the informed consent process), and residing outside the United States.

XSL•FO
RenderX

Recruitment

Participants were a self-selected convenience sample recruited from multiple sources including a chronic musculoskeletal pain clinic at a US academic medical center, email, and social media advertisement via the medical center's volunteer research participant registry, and targeted advertisement in a free online health community (The Mighty). Eligibility screening was performed using a self-reported questionnaire hosted by REDCap (Research Electronic Database Capture; Vanderbilt University), followed by verification by the research coordinator [42,43]. Recruitment was closed when usability testing by additional participants no longer yielded new themes within the convenience sample during the iterative analysis process [44].

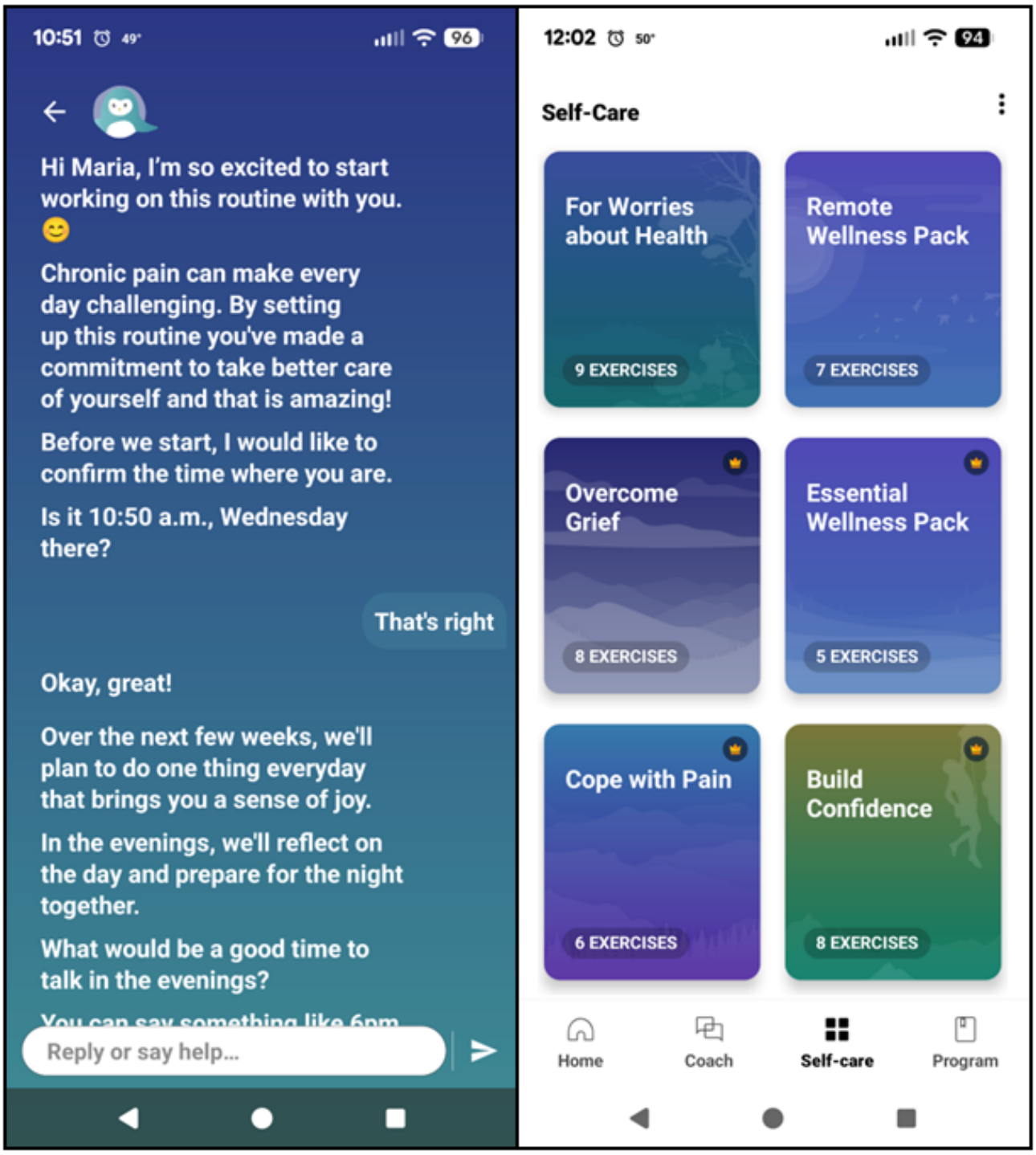
Intervention Assessed in Usability Testing

The mental health app assessed in usability testing, Wysa for Chronic Pain, aims to address mental health and chronic pain through a multifaceted approach by facilitating behavioral activation, pain acceptance, and improved sleep quality [45-48]. It guides users through evidence-based cognitive behavioral therapy, mindfulness-based interventions, and sleep support. The app has used a rule-based conversational agent with natural language understanding to deliver therapeutic content since 2016. In 2024, it augmented this approach with the addition of large language models at specific nodal points. Large language model–based capabilities were integrated into check-in conversations to improve the chatbot's ability to understand context-specific content over multiple interactions and respond in a more personalized manner to the user. However, risk-based and crisis conversations have remained completely rule-based and written by a clinician. Although the parent Wysa company offers multiple commercially available products, Wysa for Chronic Pain is not currently commercially available.

The therapeutic tools within Wysa for Chronic Pain are designed to be delivered via a combination of a structured curriculum and optional user-led exploration. By default, the app sends push notifications to users twice daily. The morning notification prompts an activity scheduling and planning task, and the evening notification prompts an activity tracking and mood reflection task and then offers a sleep meditation (Figure 1). Additional therapeutic content such as cognitive restructuring, thought recording, journaling, anxiety management, etc is delivered in a weekly cadence over 8 weeks. Each week, users receive a push notification that directs them to a report that details their app engagement thus far. Accessing the report also launches a congratulatory message for their engagement and unlocks access to new therapeutic tools as a reward for their engagement. Users can also independently navigate to therapeutic tools at any time, which are organized into "tool packs." As a safety feature, the app also includes an "SOS" safety plan feature, which lists crisis hotlines and allows users to enter and store self-reported sources of hope and "safe people or places" to reference in moments of crisis.

XSL•FO
RenderX

Figure 1. Screenshots of Wysa for Chronic Pain.



Usability Testing Process

Participants answered sociodemographic questions via an electronic REDCap questionnaire for descriptive purposes. Usability tests were conducted and recorded using audio and visual conferencing via a secure Zoom (Zoom Video Communications, Inc.) platform, each lasting approximately 45 - 60 minutes. After participants downloaded Wysa for Chronic Pain to their mobile device, they were asked to complete the app's onboarding process and complete several tasks within the app, all while performing "think aloud" testing, during which they described their thoughts and experiences in real-time (Multimedia Appendix 1: interview guide). During this process, the facilitator also took note of participants' hesitations and inaccuracies in navigating the app, and the facilitator requested feedback on potential JITAIs that could be incorporated into the app. JITAI options and their details were iteratively refined throughout the usability testing process via a rapid iterative testing process. Based on feedback from the first 21 participants, mock-ups of revised app screens and workflows were created and shared with the next 13 participants using Figma (Dev Mode) and Twine (Chris Klimas) software. Additional minor refinements were made to the mock-ups for review by the final 8 participants. However, the evaluation tasks and the app's core functionalities remained consistent for all participants.

XSL•FO
RenderX

Analysis Frameworks

The Framework for Reporting Adaptations and Modifications to Evidence-based Implementation Strategies (FRAME-IS) was used to identify and report potential usability-related adaptations needed for our target user population (Table 1) (Figure 2) [49]. FRAME-IS builds from the Framework for Reporting Adaptations and Modifications-Enhanced and is organized into modules to systematically document modifications to implementation strategies [50]. We elected to use the FRAME-IS rather than the Framework for Reporting Adaptations and Modifications-Enhanced because our goal was to adapt the delivery of therapeutic content within the digital health intervention (ie, usability and engagement optimizations as an implementation strategy), rather than adapt the core therapeutic content itself (ie, cognitive behavioral therapy, mindfulness exercises, etc as the evidence-based intervention).

XSL•FO
RenderX

Table . Application of the Framework for Reporting Adaptations and Modifications to Evidence-based Implementation Strategies, which guided proposed adaptations to improve usability of a mental health app for nondigitally native adults who have chronic pain.[a]

FRAME-IS[b] module or sub-component	Application of FRAME-IS to target intervention and population
Module 1	
The intervention being implemented is:	• Therapeutic tools within the Wysa for Chronic Pain digital platform
The implementation strategy being modified is:	• Improving usability of the intervention's platform
The modifications being made are:	• General adaptations to the intervention's interface • Incorporation of JITAIs[c]
The reasons for the modifications are:	• To lower the user's navigation burden and frustration through the intervention's interface • To increase the likelihood that the user is aware of, and engages with, relevant therapeutic tools
Module 2	
What is modified?	• Format-related contextual modifications (See Tables 2 and 3)
Module 3	
What is the nature of the content, evaluation, or training modification?	• Refining packaging of intervention content
Optional: what is the relationship to core elements?	• Fidelity-consistent: core elements of the intervention are preserved
Module 4	
What are the goals?	• Increase the usability and acceptability of the intervention to the target user group • Improve intervention fidelity and engagement such that users engage with therapeutic tools as intended (ie, improve retention) • Decrease disparities in intervention delivery (ie, improve uptake and retention by nondigitally native adults)
What is the level of the rationale for the modifications?	• Recipient (end-user)
Module 5	
When is the modification initiated?	• Scale-up phase (ie, after the intervention was developed but before it is made available to the general public)
Is the modification planned?	• Planned, proactive
Module 6	
Who participates in the decision to modify?	• Recipients (end-users), researchers, intervention developers
Optional: who makes the ultimate decision:	• Intervention developers (ie, based on feasibility and alignment with other intervention features)
Module 7	
How widespread is the modification?	• All recipients (end-users) of the intervention

[a]The left column outlines the elements of the FRAME-IS, and the right column outlines the application of the framework to this study's target intervention and population..

[b]FRAME-IS: Framework for Reporting Adaptations and Modifications to Evidence-based Implementation Strategies

[c]JITAIs: just-in-time adaptive interventions

XSL•FO
RenderX

Table . General usability adaptations proposed to improve usability of a mental health app for nondigitally native adults who have chronic pain.

Name of adaptation	Description
Onboarding conversation	Include a thorough, text-based description of how users are intended to engage with the app (eg, timing, frequency, content, goals of engagement). Also explain terms used by the app (eg, "roadmap," "journal," etc).
Onboarding tour	Include a dynamic visual tour of how to navigate the app's interface, including audio explanation and text captions.
Therapeutic approach	Explain the app's therapeutic approach during the onboarding experience. For example, explain "how an app can help with pain" by differentiating pain interference from pain severity and by stating the goals of reducing pain interference and improving pain acceptance.
Search bar	Enable users to rapidly search the library of therapeutic tools based on a tool's name, category, or other feature.
Clarify "SOS"	Appropriately label an app's crisis tools so users do not assume that pushing a button labeled "SOS" will automatically dial 911 (but rather, will direct to creation and reference of user-reported sources of hope and "safe people/places," list of crisis hotlines, etc)
Personalized "flare-up kit"	Create easy reference to a personalized plan for symptom flares (such as what would be developed by the user during a standard cognitive behavioral therapy session). For example, pin a "preparing for pain flares" plan near existing tool packs, which lists users' preferred nonpharmacologic pain management strategies (eg, ice and heat modalities, distraction techniques, mental relaxation techniques, etc).
Pinned toolkit	Maintain a group of in-app therapeutic tools that are useful for acute needs (eg, pain flares) and are always easily accessible from the app home screen (eg, mindfulness exercises, gentle stretches, reminder of user-reported preferred pain management strategies in the "flare-up kit").
Experience customization	Customize and recommend therapeutic tools based on users' responses during onboarding and subsequent conversations within the app (eg, facilitated by generative artificial intelligence).
Other challenges to address	When prompting users to report challenges they are living with, include options for loneliness, health concerns, and finances (which impact overall well-being and should then be incorporated into customizing the user's app experience).
Response to tools within app	Track users' self-reported response to each therapeutic tool, at least after first use. Use this data for future experience customization.
Weekly symptom assessment	Assess symptoms weekly with a brief measure (eg, PHQ-4)[a] [51], to facilitate (app- and user-led) pattern recognition between symptoms and other events (eg, to facilitate future app experience customization and for users to reference while journaling).
Weekly report	Relate symptom changes (eg, mood and pain) to app engagement (eg, which tools, how often), and allow users to add other external events to the report which they believe may have impacted their symptoms (eg, work, family events).
Preparing for hard days	Before closing the "weekly report" experience, ask users, "Do you expect any extra challenges this week?" If so, inquire about the timing and type of challenge. Offer to schedule a personalized tool in preparation for the challenge (eg, grounding exercise before a person with agoraphobia leaves home, meditation exercise for an anticipated activity-related pain flare, etc).
Tip of the day or week	As part of the daily check-in or weekly report, highlight a feature or tool that has not yet been explored by the user or that was rated highly but has not been used recently.
Curriculum change	After a user completes the app's structured curriculum, ask if the user prefers to continue structured check-ins versus transition to only user-led engagement with tools.

[a]PHQ-4: Patient Health Questionnaire-4.

XSL•FO
RenderX

Table . Usability-related JITAIs proposed to improve usability of a mental health app for nondigitally native adults who have chronic pain.

JITAI[a] name	Purpose	Tailoring variables	Intervention options	Proximal outcome(s)
Onboarding	Increase onboarding completion rates	If user downloaded app but did not complete onboarding within intended period	Then, send a notification that encourages the user to finish onboarding.	• Completion of onboarding
Short disengaged	Motivate a user to engage	If no app engagement for short period (eg, 24 h)	Then, send a notification that directly opens to a relevant tool.	• Any engagement[b] • Content engagement[c]
Long disengaged	Respect that users' preferred or needed frequency of engagement may fluctuate based on symptom severity and life events	If no app engagement for a prolonged period (eg, 7 d)	Then, send an "FYI/opt[d] back in"[e] notification, and stop all notifications for a brief period (eg, 3 d). After the brief period, send a notification that directly opens to a relevant tool.	• Any engagement[b] • Content engagement[c]
AM check-ins only	Personalize timing of content delivery to users' demonstrated preferences	If no engagement with PM check-ins but continued engagement with AM check-ins	Then, send an "FYI/opt back in"[e] notification, stop daily "PM check-in" notifications, and move the "activity tracking and mood reflection" task to the AM check-ins.	• Engagement with the "activity tracking and mood reflection" task compared to a preceding time interval • Change in content engagement[c] compared to a preceding time interval
PM check-ins only	Personalize timing of content delivery to users' demonstrated preferences	If no engagement with AM check-ins but continued engagement with PM check-ins	Then, send an "FYI/opt back in"[e] notification, stop daily "AM check-in" notifications, and move the "activity scheduling and planning" task to the PM check-ins.	• Engagement with the "activity scheduling and planning" task compared to a preceding time interval • Change in content engagement[c] compared to a preceding time interval
Sporadic engagement only	Personalize timing of content delivery to users' demonstrated preferences	If no engagement with AM or PM check-ins but continued user-led engagement at other times	Then, send an "FYI/opt back in"[e] notification, and stop daily "AM check-in" and "PM check-in" notifications. Instead, each day, send a single personalized notification that directly opens to a relevant tool.	• Change in content engagement[c] compared to a preceding time interval
Symptom health flare[f]	Offer increased support when increased need is demonstrated	If symptom score (eg, PHQ-4[g] [51]) worsens compared to the preceding score and the user's current push notification frequency is less than maximal frequency	Then, send an "FYI/opt out"[e] notification, and increase notification frequency back to maximal frequency (eg, twice daily).	• Change in content engagement[c] compared to a preceding time interval
Accessibility	Reduce communication barriers	If user's typo frequency interferes with effective communication with the app or if user's typing speed is slower than a designated threshold	Then, send a notification that describes accessibility options (eg, increase font size, speech-to-text, text-to-speech) and offer to direct the user to their mobile device's accessibility settings	• Change in content engagement[c] compared to a preceding time interval

[a]JITAI: just-in-time adaptive interventions

[b]"Any engagement" is operationalized as opening the app during the specified target engagement time interval. This is a binary outcome.

[c]"Content engagement" is operationalized as the total number of tools, conversations, and check-ins the user engaged with during the specified target engagement time interval. This is a continuous outcome.

[d]FYI: for your information.

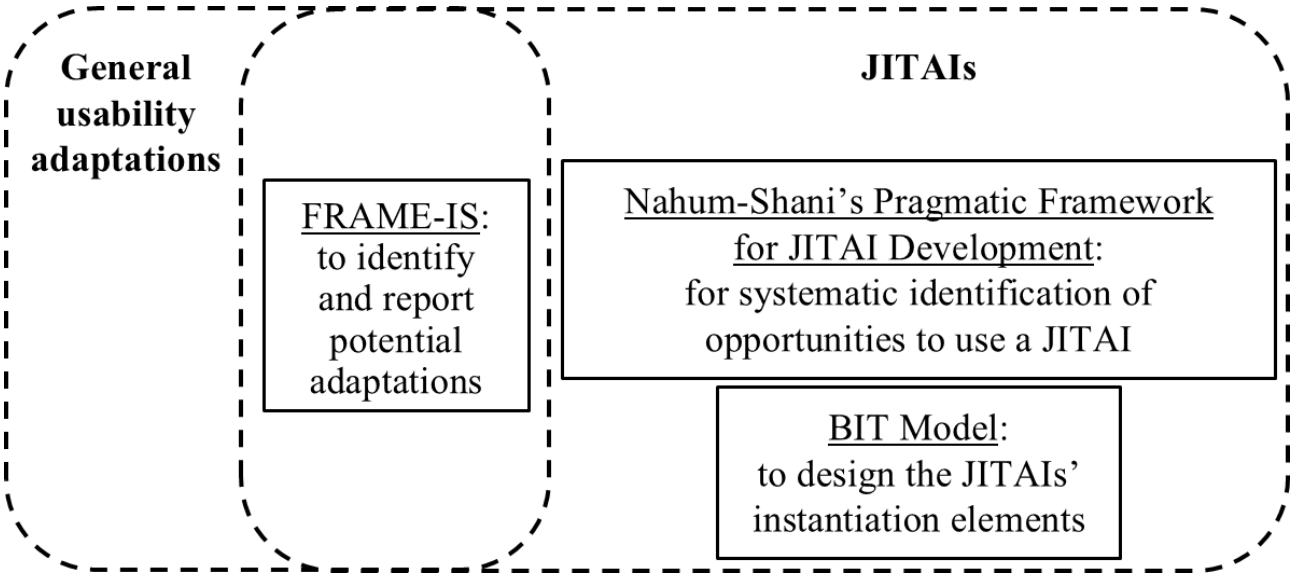
XSL•FO
RenderX

[e]If the "opt back in" or "opt out" option is selected, the rest of the intervention is not delivered.

[f]The "symptom flare" JITAI over-rides any engagement-related JITAIs which would otherwise reduce the frequency of push notifications to the user.

[g]PHQ-4: Patient Health Questionnaire-4.

Figure 2. Integration of established frameworks to guide identification and development of usability-related adaptations. FRAME-IS: Framework for Reporting Adaptations and Modifications to Evidence-based Implementation Strategies; JITAI: just-in-time adaptive intervention; BIT: Behavioral Integration Technology



Specifically for developing usability-related JITAIs, we applied Nahum-Shani's pragmatic framework for JITAI development and the Behavioral Integration Technology (BIT) model [52,53]. Nahum-Shani's pragmatic framework facilitated the systematic identification of opportunities for intervention using a JITAI [52]. That is, the framework was used to operationalize intended temporal progression of events through the app, which led to identification of promising opportunities for intervention with a JITAI. The BIT model facilitates effective translation of behavioral interventions into digital delivery systems. Specifically, we used the BIT model to design the JITAIs' instantiation elements with intentional incorporation of behavior change strategies, implementation science principles, and user interface design principles that appeal to our target population [54].

Analysis

Using a framework-informed hybrid inductive-deductive content analysis approach [55], iterative versions of general usability adaptations and usability-related JITAIs were developed and refined throughout the usability testing process based on feedback from participants and mapping to the analysis frameworks. As usability tests were completed, audio recordings of the usability tests were professionally transcribed (Rev.com), cleaned, and chunked (CYG). Three members of the study team (ALC, CYG, AM) collaboratively conducted open coding of the first 5 transcripts using an inductive, data-driven approach to identify concepts emerging from the data. Through discussion and consensus, they developed an initial codebook, which was then applied independently to pilot code the next 3 transcripts. Interrater reliability for these three transcripts was approximately 90%, with fewer than 5 coding discrepancies across coders. Discrepancies were resolved through consensus, and the codebook was refined accordingly. As usability tests were conducted with additional participants, the finalized codebook was applied using an independent open coding approach to these additional transcripts, each of which was coded by a single team member (CYG or AM), with periodic cross-checks between coders to ensure coding consistency. Coding was managed using NVivo 15 (Lumivero). Three team members (ALC, CYG, and AM) collaboratively and iteratively grouped related codes into preliminary themes that reflected participants' usability needs and suggested app adaptations, including JITAIs. This process was continued until analysis of additional transcripts no longer yielded new themes. Then, we conducted a deductive analysis by examining the inductively derived themes in relation to the FRAME-IS framework, Nahum-Shani's pragmatic framework for JITAI development, and the BIT model. The 3 team members (ALC, CYG, and AM) reviewed and discussed the themes to determine their alignment with these guiding frameworks, refine interpretation where appropriate, and translate findings into a final set of proposed general usability adaptations and usability-related JITAIs.

Reflexivity Statement

Usability testing was facilitated by a senior female research coordinator who has clinical research experience working with older adults and in the fields of mental health and orthopedics (AM). Participants' only contact with the facilitator was related to study recruitment, consent, and participation. The other 2 research team members who contributed to codebook development included a female physician scientist with a clinical background in physical medicine and rehabilitation (ALC), as well as a female physician trainee in physical medicine and rehabilitation residency (CYG). The clinical and musculoskeletal background of these team members may have influenced data collection and interpretation such that physical barriers and related adaptations were disproportionately considered over

XSL•FO
RenderX

solutions that considered mental health and systems-level factors. We worked to balance this potential bias by incorporating input from a clinician scientist with mental health clinical expertise (SH) and a health information technology expert (JA).

Rigor and Reproducibility

Informal member checking was performed by the facilitator (AM), who paraphrased and restated participants' feedback during the usability testing process to confirm that participants' intent and insights were accurately understood. Additionally, preliminary themes that emerged from early interviews were shared with later participants to assess for perceived validity. Participants generally affirmed the face validity of these themes.

Results

Of the 62 people who completed the screening questionnaire and were determined to be eligible, 42 users participated in usability testing (mean age 57, SD 8 years; women: n=32, 76%; mean chronic pain duration 16, SD 12 years) (Table 4). Reasons for nonparticipation included (1) not responding to invitations to schedule an appointment (or missing a scheduled appointment) (n=4), and (2) achievement of thematic saturation prior to their participation (n=16).

XSL•FO
RenderX

Table . Characteristics of usability testing participants (N=42).

Characteristics	Values
Age (years), median (IQR; range)	55 (51-61; 45-76)
Sex[a], n (%)	
Female	32 (76)
Male	10 (24)
Race, n (%)	
White	30 (71)
Black or African American	9 (21)
Other	1 (2)
More than one race	1 (2)
Unknown or prefer not to answer	1 (2)
Ethnicity, n (%)	
Hispanic or Latino	3 (7)
Not Hispanic or Latino	36 (86)
Unknown or prefer not to answer	3 (7)
Employment status, n (%)	
Working full-time	11 (26)
Working part-time	4 (10)
Disabled	13 (31)
Retired	10 (24)
Unemployed	2 (5)
Prefer not to answer	2 (5)
Frequency of internet use[b], n (%)	
Most of the day	12 (29)
Multiple times a day	28 (67)
At least once a day	1 (2)
Less than once a day	1 (2)
High-impact chronic pain present[c], n (%)	33 (80)
BPI Pain Interference score[d], median (IQR; range)	7.1 (6 - 8.1; 2.4 - 9)
PHQ-9[e] score (years), median (IQR; range)	14 (12-17; 7-22)
GAD-7[f] score (years), median (IQR; range)	12 (9-15; 1-21)

[a]Sex and gender identity were concordant for all participants.

[b]Self-reported internet use was captured by the Pew Digital Savviness Classifier [56].

[c]"High impact chronic pain" is defined as pain that limits life or work activities on most or every day over the past 3 months [20].

[d]The Brief Pain Inventory (BPI) Pain Interference score ranges from 0 to 10 [57]. Higher scores indicate greater pain interference.

[e]The Patient Health Questionnaire (PHQ)-9 is a 9-item screening measure for depressive symptoms [39]. Scores range from 0 to 27. Higher scores suggest more depressive symptoms.

[f]The Generalized Anxiety Disorder (GAD)-7 is a 7-item screening measure for anxiety symptoms [40]. Scores range from 0 to 21. Higher scores suggest more anxiety symptoms.

We identified general usability adaptations, as well as usability-related JITAIs, that users perceived would enhance their app experience. Rapid, nearly universal convergence was achieved related to suggestions for general adaptations to provide a more thorough onboarding experience to the app (Table 2). Specifically, users requested clear, step-by-step navigation instructions, rather than a brief onboarding experience that relied on user-led exploration of the app and familiarity with conventional app symbols (eg, three vertical

XSL•FO
RenderX

ellipsis dots indicating "more options"). A 49-year-old female stated:

> *For those who may not be as familiar – and you're going to start getting a lot of those much older users, especially in this kind of a world with health concerns – I don't know that they would know what that [icon] means without a tour of the app first.*

Users also appreciated a customized app experience based on their unique challenges and app usage patterns, although specific desired customizations were somewhat varied, and no clear patterns emerged related to demographic characteristics or recruitment source, based on impressionistic observation.

Consistent with Nahum-Shani's pragmatic framework [52], the majority of participants favored that some of the usability-related JITAI opportunities be designed to strategically reduce notifications to the user in order to reduce the likelihood of a negative impact such as notification fatigue and subsequent complete disengagement with the app (Table 5). As one 60-year-old female participant said:

> *If I've not engaged in a week and after three days of nothing bothering me – no notifications – I think that a notification out of the blue would probably kick me back into gear. Give me that little kick start I need.*

When a notification is delivered, application of the BIT model [53] facilitated identification of the notification's instantiation components that were most appealing to users (Table 6).

Table . Application of Nahum-Shani's pragmatic framework which guided development of usability-related JITAIs to improve usability of a mental health app for nondigitally native adults who have chronic pain.

Component of Nahum-Shani's pragmatic framework	Application to target intervention and population
Defining the problem	
Whom are you trying to help?	• Nondigitally native adults, especially those who have chronic pain that coexists with symptoms of depression, anxiety, or both. This target population may be less likely than younger adults to readily and proficiently navigate through an app independently and find tools of interest.
What is the desired distal outcome of the JITAI([c]s)?	• Improve depression and anxiety symptoms (primary outcomes) • Improve behavioral activation, pain acceptance, and sleep quality (proposed mediators of primary outcomes)
What is the temporal progression of key factors towards the distal outcome?[b]	• Morning engagement: for behavioral activation (eg, activity scheduling and planning) • Evening engagement: for behavioral activation (eg, activity tracking, mood reflection), sleep optimization (eg, meditation) • Weekly engagement: for reflection and introduction of new cognitive behavioral skills (eg, cognitive reframing, thought recording) • Sustained engagement (8 wk): for growth, maintenance, and eventual independence with cognitive behavioral and pain self-management skills
What are contender proximal outcomes?	• Tool-specific engagement within the app
Defining just-in-time in the context of the identified problem	
What factors mark a state of vulnerability or opportunity?	• Lack of engagement with the app (as intended or as previously engaged)
What possible intervention options can affect the proximal outcomes?	• A notification or nudge that is noticed by, and appeals to, the user
What factors mark a state of unreceptivity to the selected intervention options?	• Already engaging with the app as intended or more frequently • Not responding to existing notifications or nudges
Formulating the adaptation strategy:	
What are the tailoring variables?	• User's current pattern of engagement
For each possible level of the tailoring variable, which intervention option is likely to have the desired effect on the proximal outcome?	• See Table 3. Additional research is needed to identify optimal cut-points for each tailoring variable.
What plausible decision rules can be generated to operationalize effective adaptation?	• See Table 3. Additional research is needed to identify optimal cut-points for each tailoring variable.

[a]JITAI: just-in-time adaptive interventions.

[b]The temporal progression described reflects how users are *intended* to engage with the app.

XSL•FO
RenderX

Table . Application of the BIT[a] model which guided development of usability-related JITAIs[b] to improve usability of a mental health app for nondigitally native adults who have chronic pain.

BIT component	Application to target intervention and population
Theoretical	
Why: aims	
Clinical aims	Improve depression and anxiety symptoms
Usage aims	Increase engagement with evidence-based therapeutic tools
How (conceptual): behavior change strategies	
Clinical strategies	Improve behavioral activation (eg, through goal setting), pain acceptance (eg, through pain neuroscience education and mindfulness exercises), and sleep quality (eg, through sleep hygiene education and meditation exercises)
Usage strategies	Direct user through a structured educational curriculum and on-demand access to therapeutic tools
Instantiation	
What: elements	Notifications that are delivered with personalized timing and content based on passively and actively collected data
How (technical): characteristics	Each push notification will feature a visual and audio alert, as permitted by the user's self-selected mobile device settings. The notifications direct the user to relevant therapeutic tools, as identified from the user's prior responses (eg, challenges expressed) and app engagement patterns. The content of the notifications is encouraging but not overly cheerful, with judicious use of emojis.
When: workflow	Timing of delivery of the push notifications is determined by event-based rules. However, precise timing relative to occurrence of the event(s) may be adaptive based on the user's engagement history and other characteristics.

[a]BIT: behavioral intervention technology

[b]JITAI: just-in-time adaptive intervention.

We identified 8 potential usability-related JITAIs to consider for build-out and formal effectiveness testing (Table 3). These JITAIs were developed to address the competing priorities of (1) providing predictable support to the user, and (2) acknowledging that pain and mood symptoms (and hence, a user's need for support) can fluctuate. Users reported varying preferences regarding the frequency of push notifications and indications to change the push notification frequency. Therefore, our JITAIs that propose to change the frequency of future push notifications all include an "opt back in" or "opt out" option so that users maintain autonomy over their app experience. As one 48-year-old male described:

> *Some people, if they're in a certain mental state, may need a little bit more persistence, whereas other people may need more space. I feel like me personally – if I didn't engage for a while and then had a [several-day break from notifications], it would be like, 'Hey, we're still here if you need us.' I think that would work for me personally, but I think it would all depend on the individual. If the app gave you an option or the ability to change the [notification frequency] or suggest how you want to be taken care of, yeah, I think that would be a good idea.*

The precise text of each JITAI notification is intended to be engaging and varied over time, even if the same JITAI is triggered more than once. As relevant and possible, the JITAI notifications should leverage information that was previously reported by the user in order to personalize the content and resonate with the user (eg, suggesting tools based on prior challenges reported, tools that were previously rated as helpful, and tools that have not been used yet but are expected to be helpful based on the user's reported challenges and prior tools used). When selected, most of the JITAI notifications are intended to directly open the app to a relevant tool for the user.

Discussion

Principal Results

In this study, we identified adaptations to an existing mental health app that better meet the stated usability preferences and needs of nondigitally native adults who have chronic pain and coexisting symptoms of depression or anxiety. Study participants specifically identified several opportunities to optimize their user experience by ensuring that the digital health intervention clearly describes how all its features are intended to be used and by minimizing navigation burden. For instance, users proposed general usability adaptations for a more robust orientation experience during app onboarding and for ongoing personalized engagement prompts using usability-related JITAIs.

XSL•FO
RenderX

Comparison With Prior Work

Based on prior literature and feedback from participants in this study [15,58], when identifying decision rules for our proposed JITAIs, we aimed to balance the use of passively collected app engagement data and user-directed triggers in order to respect users' dual desires for reduced navigation burden and also maintenance of autonomy and privacy. In related work [12], the majority of users in this target population expressed hesitation about allowing a digital mental health app to collect and use other data, such as physical location that is passively collected from their mobile device or other wearable sensors. Furthermore, older adults and those with multimorbidity are less likely to continuously carry a mobile device or use a wearable sensor, which would limit the impact of JITAIs that rely on datapoints from these sources [59-62]. As a result, the JITAIs developed in this study were intentionally designed to operate using only engagement-related and user-reported data that is collected within the app itself.

From a practical standpoint, the proposed JITAIs identified in this study represent general concepts that were perceived favorably by users. Determining the effectiveness of the proposed JITAIs and identifying ideal specific cut-points for each JITAI's tailoring variables (eg, number of hours or days of no app engagement before delivering an engagement-related JITAI) requires additional investigation. Although preliminary cutoff points for proposed JITAIs can be developed based on qualitative evidence from a usability study such as this, recommended best practices suggest follow-up investigation using complex analytical techniques to refine evidence-based (and potentially adaptive) cutoff points [15]. Implementation of potential JITAIs also depends on feasibility within the digital health platform, as well as confirmation that the final decision rules and cutoff points of the JITAIs do not conflict with one another or with other elements of the intervention's features or user experience. We plan to refine the details of these JITAIs using a series of microrandomized trials and additional qualitative user feedback [41].

In addition to leveraging existing evidence-based frameworks, we also leveraged an academic-industry partnership to expedite the incorporation of evidence-based principles into a digital health intervention with potential for rapid dissemination to the general public [63]. Specifically, the commercial app company facilitated participant access to the intervention and iterative mock-ups for the usability testing. The academic researchers led the usability testing itself and applied evidence-based frameworks to guide data collection and analysis. After the identified adaptations are fully built and subsequently tested for effectiveness, they can rapidly be rolled out to several thousand daily users in a much faster time scale than would be possible if an academic group developed a digital health intervention de novo. Our process to identify implementable usability-related adaptations of an existing mental health app can serve as a proof-of-concept method for relatively rapid refinement of other existing digital health interventions in order to better serve specific target populations, especially for apps that already have robust evidence-based content but could benefit from adaptations for specific subpopulations.

A major strength of this study is the integrated framework-based approach to identify usability-related opportunities for a target user population with unique needs. This approach can serve as a model for other studies, and the usability-related principles that were identified in this study are broadly generalizable for other digital health interventions that are intended for nondigitally native adults with multimorbidity. The robust sample size is also a strength relative to other digital health usability studies [64,65].

Limitations

Some limitations are inherent to the study design used. Observer bias (ie, the Hawthorne effect) may have resulted in participants reporting more confidence or demonstrating greater tenacity with technical features than they otherwise would in a real-world setting. Additionally, the iterative nature of our usability testing resulted in some heterogeneity in participant experiences. While participants were not exposed to identical interfaces, this approach is consistent with user-centered and formative usability evaluation methods, in which iterative refinement is prioritized to rapidly identify and address usability barriers [66]. Furthermore, details of some of the usability-related adaptations and JITAIs that were identified in this study are predominantly applicable to the specific mental health app that was tested. Our proposed adaptations also reflect participants' stated preferences, but quantitative assessment and data-driven refinement was outside the scope of this study. Finally, user preferences among nondigitally native adults may evolve over time as the population's overall experience with technology continues to grow.

Given the demographics of our sample, our study findings are likely most generalizable to middle-aged adults in the United States who are relatively digitally proficient and already use the internet routinely throughout the day. As a result, findings from this study may not fully capture the usability needs and other barriers faced by older adults with low digital literacy, which remains an important and underserved segment of the target user population. Interpretation of generalizability is also limited by the self-selected nature of the convenience sample and the lack of a standardized cognitive screening instrument, especially since the prevalence of mild cognitive impairment increases with age.

Conclusions

Identifying implementable opportunities to improve usability and prompt meaningful engagement at relevant timepoints holds promise to improve the effectiveness of digital health interventions because intervention attrition is a well-known challenge that is suspected to reduce the effectiveness of digital health interventions on a population level [67-70]. In this study, we used an integrated framework-based approach and user feedback to identify usability-related optimizations to an existing mental health app that are preferred among relatively digitally proficient but nondigitally native adults who have chronic pain and coexisting symptoms of depression or anxiety. Users identified general usability-related adaptations and usability-related JITAIs which were appealing for their potential to (1) provide clarity with regard to how to navigate the digital health intervention, and (2) reduce the user's burden of

XSL•FO
RenderX

independent navigation within the intervention. These usability-related refinement opportunities are generalizable to other digital health interventions intended for older users who are likely to be managing multimorbidity, are not digital natives, and are more likely to have physical or cognitive impairments than younger users (eg, medication reminder apps, daily safety monitoring apps, etc). The integrated framework-based approach taken in this study can also serve as a model for refinement of digital health interventions with other target populations, as well.

Acknowledgments

The authors thank Dr Patricia Areán for her contributions to the study conception. Generative artificial intelligence (AI) was not used in any portion of conducting this study or generating this manuscript.

Funding

This study was funded by the National Institute of Mental Health (grant R01MH131989; principal investigator ALC). The funder did not play a role in data collection, management, analysis, or interpretation of the data; writing of the report; or the decision to submit the report for publication. Use of Washington University's instance of REDCap for this study was supported by the Siteman Cancer Center's National Cancer Institute Cancer Center Support Grant P30CA091842 and by the Washington University Institute of Clinical and Translational Sciences Grant UL1TR002345 from the National Center for Advancing Translational Sciences.

Data Availability

The dataset generated and analyzed during this study is available in the National Institute of Mental Health (NIMH) Data Archive (NDA).

Authors' Contributions

Conceptualization: ALC, JA
Methodology: ALC, CYG, SMH, JA
Study investigation: AM
Data management: AM
Analysis: ALC, CYG, AM
Analysis interpretation: SMH, JA
Validation: CYG, AM
Writing: ALC, CYG, AM
Editing: SMH, JA
Supervision: ALC, JA
Funding acquisition: ALC

Conflicts of Interest

The academic study team was independent and distinct from the employees, developers, and owners of the Wysa for Chronic Pain intervention. Wysa conceived of and developed the original Wysa for Chronic Pain intervention and provided access to the Wysa for Chronic Pain intervention by the study team and study participants. Based on participant feedback obtained by the study team as part of grant R01MH131989, Wysa created intervention mock-ups to be used in the iterative usability testing of this study. Chaitali Sinha is the chief clinical research and development officer at Wysa, and she is a co-investigator on grant R01MH131989. Wysa was not involved in the development of the study design, conduct of the usability testing, data analysis, final interpretation of the data, or final decision to submit study findings for peer-reviewed publication.

Multimedia Appendix 1
Interview guide.
[DOCX File, 28 KB - humanfactors_v13i1e87365_app1.docx]

References

1. Gallois CB, Machado MM, de Oliveira CDS, Xavier ACM, Dreher CB, Manfro GG. Is digital psychiatry really for all? A cross-sectional analysis from two randomized clinical trials. Trends Psychiatry Psychother 2025;47:e20240826. [doi: 10.47626/2237-6089-2024-0826] [Medline: 39052861]
2. Schumacher RM. Gran got tech: inclusivity and older adults. J User Exp 2023;18(2):62-67 [FREE Full text]
3. Prensky M. Digital natives, digital immigrants part 1. On Horizon 2001 Sep;9(5):1-6. [doi: 10.1108/10748120110424816]
4. Fortuna KL, Torous J, Depp CA, et al. A future research agenda for digital geriatric mental healthcare. Am J Geriatr Psychiatry 2019 Nov;27(11):1277-1285. [doi: 10.1016/j.jagp.2019.05.013] [Medline: 31196619]

XSL•FO
RenderX

5. Wang S, Bolling K, Mao W, et al. Technology to support aging in place: older adults' perspectives. Healthcare (Basel) ;7(2):60. [doi: 10.3390/healthcare7020060] [Medline: 30974780]
6. Perrin A. Mobile technology and home broadband 2021. : Pew Research Center; 2021 URL: https://www.pewresearch.org/internet/2021/06/03/mobile-technology-and-home-broadband-2021/ [accessed 2026-07-24]
7. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: a literature based framework (MOLD-US). Int J Med Inform 2018 Jun;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
8. Wagner N, Hassanein K, Head M. The impact of age on website usability. Comput Hum Behav 2014 Aug;37(Aug):270-282. [doi: 10.1016/j.chb.2014.05.003]
9. Romano Bergstrom JC, Olmsted-Hawala EL, Jans ME. Age-related differences in eye tracking and usability performance: website usability for older adults. Int J Hum Comput Interact 2013 Aug 3;29(8):541-548. [doi: 10.1080/10447318.2012.728493]
10. Americans check their phones 96 times a day. : Asurion URL: https://www.asurion.com/press-releases/americans-check-their-phones-96-times-a-day/ [accessed 2026-07-24]
11. Bartels SJ, DiMilia PR, Fortuna KL, Naslund JA. Integrated care for older adults with serious mental illness and medical comorbidity: evidence-based models and future research directions. Psychiatr Clin North Am 2018 Mar;41(1):153-164. [doi: 10.1016/j.psc.2017.10.012] [Medline: 29412843]
12. Cheng AL, Gou CY, Martin A, Hartz SN, Abraham J. Design preferences for mental health apps among nondigitally native adults with chronic pain: qualitative analysis. JMIR Mhealth Uhealth (forthcoming) 2026. [doi: 10.2196/87358]
13. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
14. Teepe GW, Da Fonseca A, Kleim B, et al. Just-in-time adaptive mechanisms of popular mobile apps for individuals with depression: systematic app search and literature review. J Med Internet Res 2021 Sep 28;23(9):e29412. [doi: 10.2196/29412] [Medline: 34309569]
15. van Genugten CR, Thong MSY, van Ballegooijen W, et al. Beyond the current state of just-in-time adaptive interventions in mental health: a qualitative systematic review. Front Digit Health 2025;7(1460167):1460167. [doi: 10.3389/fdgth.2025.1460167] [Medline: 39935463]
16. DeVeaugh-Geiss AM, West SL, Miller WC, Sleath B, Gaynes BN, Kroenke K. The adverse effects of comorbid pain on depression outcomes in primary care patients: results from the ARTIST trial. Pain Med 2010 May;11(5):732-741. [doi: 10.1111/j.1526-4637.2010.00830.x] [Medline: 20353408]
17. Roughan WH, Campos AI, García-Marín LM, et al. Comorbid chronic pain and depression: shared risk factors and differential antidepressant effectiveness. Front Psychiatry 2021;12:643609. [doi: 10.3389/fpsyt.2021.643609] [Medline: 33912086]
18. Gerrits MMJG, van Oppen P, Leone SS, van Marwijk HWJ, van der Horst HE, Penninx BW. Pain, not chronic disease, is associated with the recurrence of depressive and anxiety disorders. BMC Psychiatry 2014 Jun 25;14:187. [doi: 10.1186/1471-244X-14-187] [Medline: 24965597]
19. Gerrits MMJG, Vogelzangs N, van Oppen P, van Marwijk HWJ, van der Horst H, Penninx BWJH. Impact of pain on the course of depressive and anxiety disorders. Pain 2012 Feb;153(2):429-436. [doi: 10.1016/j.pain.2011.11.001] [Medline: 22154919]
20. Dahlhamer J, Lucas J, Zelaya C, et al. Prevalence of chronic pain and high-impact chronic pain among adults - United States, 2016. MMWR Morb Mortal Wkly Rep 2018 Sep 14;67(36):1001-1006. [doi: 10.15585/mmwr.mm6736a2] [Medline: 30212442]
21. Bhattarai P, Phillips JL. The role of digital health technologies in management of pain in older people: an integrative review. Arch Gerontol Geriatr 2017;68:14-24. [doi: 10.1016/j.archger.2016.08.008] [Medline: 27584871]
22. Carlo AD, Hosseini Ghomi R, Renn BN, Strong MA, Areán PA. Assessment of real-world use of behavioral health mobile applications by a novel stickiness metric. JAMA Netw Open 2020 Aug 3;3(8):e2011978. [doi: 10.1001/jamanetworkopen.2020.11978] [Medline: 32744628]
23. Carlo AD, Hosseini Ghomi R, Renn BN, Areán PA. By the numbers: ratings and utilization of behavioral health mobile applications. NPJ Digit Med 2019;2(54):54. [doi: 10.1038/s41746-019-0129-6] [Medline: 31304400]
24. Torous J, Bucci S, Bell IH, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct;20(3):318-335. [doi: 10.1002/wps.20883] [Medline: 34505369]
25. Torous J, Linardon J, Goldberg SB, et al. The evolving field of digital mental health: current evidence and implementation issues for smartphone apps, generative artificial intelligence, and virtual reality. World Psychiatry 2025 Jun;24(2):156-174. [doi: 10.1002/wps.21299] [Medline: 40371757]
26. Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M. Perceptions and opinions of patients about mental health chatbots: scoping review. J Med Internet Res 2021 Jan 13;23(1):e17828. [doi: 10.2196/17828] [Medline: 33439133]

XSL•FO
RenderX

27. Mohr DC, Kwasny MJ, Meyerhoff J, Graham AK, Lattie EG. The effect of depression and anxiety symptom severity on clinical outcomes and app use in digital mental health treatments: meta-regression of three trials. Behav Res Ther 2021 Dec;147:103972. [doi: 10.1016/j.brat.2021.103972] [Medline: 34600398]
28. Arean PA, Hallgren KA, Jordan JT, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J Med Internet Res 2016 Dec 20;18(12):e330. [doi: 10.2196/jmir.6482] [Medline: 27998876]
29. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
30. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295. [doi: 10.2196/jmir.7740] [Medline: 28838887]
31. Mojtabai R, Olfson M, Sampson NA, et al. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. Psychol Med 2011 Aug;41(8):1751-1761. [doi: 10.1017/S0033291710002291] [Medline: 21134315]
32. Schure M, McCrory B, Tuchscherer Franklin K, Greist J, Weissman RS. Twelve-month follow-up to a fully automated internet-based cognitive behavior therapy intervention for rural adults with depression symptoms: single-arm longitudinal study. J Med Internet Res 2020 Oct 2;22(10):e21336. [doi: 10.2196/21336] [Medline: 33006561]
33. Gureje O, Von Korff M, Simon GE, Gater R. Persistent pain and well-being: a World Health Organization study in primary care. JAMA 1998 Jul 8;280(2):147-151. [doi: 10.1001/jama.280.2.147] [Medline: 9669787]
34. Ruoff GE. Depression in the patient with chronic pain. J Fam Pract 1996 Dec;43(6 Suppl):S25-S33. [Medline: 8969710]
35. Zelaya CE, Dahlhamer JM, Lucas JW, Connor EM. Chronic pain and high-impact chronic pain among U.S. adults, 2019. NCHS Data Brief 2020 Nov(390):1-8. [Medline: 33151145]
36. Sharpe L, McDonald S, Correia H, et al. Pain severity predicts depressive symptoms over and above individual illnesses and multimorbidity in older adults. BMC Psychiatry 2017 May 4;17(1):166. [doi: 10.1186/s12888-017-1334-y] [Medline: 28472936]
37. Sheng J, Liu S, Wang Y, Cui R, Zhang X. The link between depression and chronic pain: neural mechanisms in the brain. Neural Plast 2017;2017:9724371. [doi: 10.1155/2017/9724371] [Medline: 28706741]
38. Kamper SJ, Apeldoorn AT, Chiarotto A, et al. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain: cochrane systematic review and meta-analysis. BMJ 2015 Feb 18;350:h444. [doi: 10.1136/bmj.h444] [Medline: 25694111]
39. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
40. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
41. Cheng AL, Abraham J, Hartz SM, Laber EB, Miller JP. Evaluating and optimizing just-in-time adaptive interventions in a digital mental health intervention (Wysa for Chronic Pain) for middle-aged and older adults with chronic pain: protocol for a series of randomized trials. JMIR Res Protoc 2025 Sep 17;14:e77532. [doi: 10.2196/77532] [Medline: 40961490]
42. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
43. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
44. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
45. Cheng AL, Agarwal M, Armbrecht MA, Abraham J, Calfee RP, Goss CW. Behavioral mechanisms that mediate mental and physical health improvements in people with chronic pain who receive a digital health intervention: prospective cohort pilot study. JMIR Form Res 2023 Nov 17;7:e51422. [doi: 10.2196/51422] [Medline: 37976097]
46. Leo AJ, Schuelke MJ, Hunt DM, et al. A digital mental health intervention in an orthopedic setting for patients with symptoms of depression and/or anxiety: feasibility prospective cohort study. JMIR Form Res 2022 Feb 21;6(2):e34889. [doi: 10.2196/34889] [Medline: 35039278]
47. Leo AJ, Schuelke MJ, Hunt DM, Miller JP, Areán PA, Cheng AL. Digital mental health intervention plus usual care compared with usual care only and usual care plus in-person psychological counseling for orthopedic patients with symptoms of depression or anxiety: cohort study. JMIR Form Res 2022 May 4;6(5):e36203. [doi: 10.2196/36203] [Medline: 35507387]
48. Sinha C, Cheng AL, Kadaba M. Adherence and engagement with a cognitive behavioral therapy-based conversational agent (Wysa for Chronic Pain) among adults with chronic pain: survival analysis. JMIR Form Res 2022 May 23;6(5):e37302. [doi: 10.2196/37302] [Medline: 35526201]
49. Miller CJ, Barnett ML, Baumann AA, Gutner CA, Wiltsey-Stirman S. The FRAME-IS: a framework for documenting modifications to implementation strategies in healthcare. Implement Sci 2021 Apr 7;16(1):36. [doi: 10.1186/s13012-021-01105-3] [Medline: 33827716]
50. Wiltsey Stirman S, Baumann AA, Miller CJ. The FRAME: an expanded framework for reporting adaptations and modifications to evidence-based interventions. Implement Sci 2019 Jun 6;14(1):58. [doi: 10.1186/s13012-019-0898-y] [Medline: 31171014]

XSL•FO
RenderX

51. Kroenke K, Spitzer RL, Williams JBW, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. Psychosomatics 2009;50(6):613-621. [doi: 10.1176/appi.psy.50.6.613] [Medline: 19996233]
52. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: a pragmatic framework. Health Psychol 2015 Dec;34S:1209-1219. [doi: 10.1037/hea0000306] [Medline: 26651462]
53. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 5;16(6):e146. [doi: 10.2196/jmir.3077] [Medline: 24905070]
54. Zingg A, Singh T, Myneni S. Towards digestible digital health solutions: application of a health literacy inclusive development framework for peripartum depression management. AMIA Annu Symp Proc 2021;2021:1274-1283. [Medline: 35308913]
55. Harrison Ginsberg K, Babbott K, Serlachius A. Exploring participants' experiences of digital health interventions with qualitative methods: guidance for researchers. J Med Internet Res 2024 Nov 28;26:e62761. [doi: 10.2196/62761] [Medline: 39607999]
56. Colbenson GA, Cook DA, Stephenson CR. Learner engagement, teaching effectiveness, and digital proficiency in in-person versus livestream continuing medical education. J Contin Educ Health Prof 2026;46(2):88-94. [doi: 10.1097/CEH.0000000000000622] [Medline: 41143745]
57. Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. Ann Acad Med Singap 1994 Mar;23(2):129-138. [Medline: 8080219]
58. Fouyaxis J, Bidargaddi N, Du W, Looi JCL, Lipschitz J. Critical design decisions and user demographics in enhancing real-time digital mental health interventions: a systematic review. Digit Health 2024;10:20552076241306782. [doi: 10.1177/20552076241306782] [Medline: 39687526]
59. Rising CJ, Jensen RE, Moser RP, Oh A. Characterizing the US population by patterns of mobile health use for health and behavioral tracking: analysis of the National Cancer Institute's Health Information National Trends Survey data. J Med Internet Res 2020 May 14;22(5):e16299. [doi: 10.2196/16299] [Medline: 32406865]
60. Jaana M, Paré G. Comparison of mobile health technology use for self-tracking between older adults and the general adult population in Canada: cross-sectional survey. JMIR Mhealth Uhealth 2020 Nov 27;8(11):e24718. [doi: 10.2196/24718] [Medline: 33104517]
61. Paré G, Leaver C, Bourget C. Diffusion of the digital health self-tracking movement in Canada: results of a national survey. J Med Internet Res 2018 May 2;20(5):e177. [doi: 10.2196/jmir.9388] [Medline: 29720359]
62. Strain T, Wijndaele K, Brage S. Physical activity surveillance through smartphone apps and wearable trackers: examining the UK potential for nationally representative sampling. JMIR Mhealth Uhealth 2019 Jan 29;7(1):e11898. [doi: 10.2196/11898] [Medline: 30694198]
63. Hall SS, Hastings O, Prentice KM, et al. Principles of industry-academic partnerships informed by digital mental health collaboration: mixed methods study. JMIR Ment Health 2025 Sep 10;12:e77439. [doi: 10.2196/77439] [Medline: 40929728]
64. Sin J, Woodham LA, Henderson C, Williams E, Sesé Hernández A, Gillard S. Usability evaluation of an eHealth intervention for family carers of individuals affected by psychosis: a mixed-method study. Digit Health 2019;5:2055207619871148. [doi: 10.1177/2055207619871148] [Medline: 31489205]
65. Stonbraker S, Cho H, Hermosi G, Pichon A, Schnall R. Usability testing of a mHealth app to support self-management of HIV-associated non-AIDS related symptoms. Stud Health Technol Inform 2018;250:106-110. [Medline: 29857399]
66. Marcilly R, Monkman H, Pelayo S, Lesselroth BJ. Usability evaluation ecological validity: is more always better? Healthcare (Basel) 2024 Jul 16;12(14):1417. [doi: 10.3390/healthcare12141417] [Medline: 39057560]
67. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11. [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
68. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the wild: using novel data to examine the reach, use, and impact of PTSD coach. JMIR Ment Health 2015;2(1):e7. [doi: 10.2196/mental.3935] [Medline: 26543913]
69. Laing BY, Mangione CM, Tseng CH, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Ann Intern Med 2014 Nov 18;161(10 Suppl):S5-12. [doi: 10.7326/M13-3005] [Medline: 25402403]
70. Y Wang A, Vereschagin M, G Richardson C, et al. Examining the effects of engagement with an app-based mental health intervention: a secondary analysis of a randomized control trial with treatment non-compliance. Int J Ment Health Syst 2025 Oct 9;19(1):30. [doi: 10.1186/s13033-025-00688-4] [Medline: 41068963]

Abbreviations:

BIT: behavioral integration technology
FRAME-IS: Framework for Reporting Adaptations and Modifications to Evidence-based Implementation Strategies
JITAI: just-in-time adaptive intervention
REDCap: Research Electronic Database Capture

XSL•FO
RenderX

Edited by S Law; submitted 07.Nov.2025; peer-reviewed by F Bruni, SS Funghetto; revised version received 15.Jul.2026; accepted 15.Jul.2026; published 11.Aug.2026.

Please cite as:
Cheng AL, Gou CY, Martin A, Hartz SM, Abraham J
Optimizing Usability of Digital Health Interventions for Nondigitally Native Adults: A Framework-Based Approach to Develop Just-in-Time Adaptive Interventions (JITAIs) and Other Adaptations
JMIR Hum Factors 2026;13:e87365
URL: https://humanfactors.jmir.org/2026/1/e87365
doi:10.2196/87365

© Abby L Cheng, Christine Y Gou, Adriana Martin, Sarah M Hartz, Joanna Abraham. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 11.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Personalized Intelligent Chatbot Based on AI-Generated Content Assists Memoir Writing for Older Adults With Cognitive Impairment: Mixed Methods Study

Yibo Meng[1*], MSc; Yuan Que[2*], BSc; Zhe Yan[3*], BSc; Bingyi Liu[4*], MSc; Zixin Wang[5*], MSc; Mandi Yang[6], BA; Huidi Lu[7], PhD

[1]Tsinghua University, Beijing, China
[2]Southwest Jiaotong University, Chongqing, China
[3]The Chinese University of Hong Kong, Shenzhen, Shenzhen, China
[4]University of Michigan, Ann Arbor, Ann Arbor, MI, United States
[5]University of Pennsylvania, Philadelphia, PA, United States
[6]Nankai University, Tianjin, China
[7]Said Business School, University of Oxford, Park End Street, Oxford, England, United Kingdom
[*]these authors contributed equally

Corresponding Author:
Huidi Lu, PhD
Said Business School, University of Oxford, Park End Street, Oxford, England, United Kingdom

Abstract

Background: Older adults with cognitive impairment often face significant challenges in memoir writing, including memory fragmentation, emotional loneliness, and speech and language disorders. Although AI-generated content (AIGC) technologies such as GPT-3.5 show potential in creative tasks, they often lack the personalization and adaptability required for users with dementia. Generic AIGC tools often fail to address the heterogeneous cognitive and emotional needs of this population.

Objective: This study aimed to design and evaluate, as a proof-of-concept, a personalized AIGC-powered chatbot to assist older adults with cognitive impairment in memoir writing and emotional support.

Methods: We developed a multimethod collaborative design framework integrating Kansei Engineering, Quality Function Deployment, Axiomatic Design, and the Technique for Order of Preference by Similarity to Ideal Solution decision model. The system dynamically adapts interaction strategies based on users' Mini-Mental State Examination (MMSE) scores, using a clinical threshold of 20 to distinguish mild (20-26) from moderate-to-severe (<20) impairment. In a single-session, nonrandomized, matched-pair evaluation using minimization-based allocation with an active control condition, performance was assessed via usability testing (System Usability Scale), affect assessment (Positive and Negative Affect Schedule), and blinded psychiatrist-rated memoir quality among 20 participants (10 per arm).

Results: The experimental group showed significantly greater improvement than the active control group in positive affect, negative affect, and psychiatrist-rated memoir quality (all *Ps*≤.03 by matched-pair analysis; all comparisons remained significant after Benjamini-Hochberg correction), with moderate-to-large effect sizes (Cohen *d*=0.85 - 1.03 for affect outcomes; rank-biserial *r*=1.00 for memoir quality). Within the sample (MMSE range 11 - 23), participants in the lower MMSE tier (<20; n=12) appeared to benefit more from AI-driven narrative generation, while those in the higher tier (20-23; n=8) responded better to keyword-based prompting; these subgroup observations are descriptive, given the small cell sizes. The mean System Usability Scale total score of 92.2 (SD 5.1) exceeded the acceptability threshold and fell in the excellent range, though single-session exposure and potential acquiescence bias warrant cautious interpretation.

Conclusions: These findings provide preliminary, hypothesis-generating evidence that a personalized AIGC-based chatbot may support short-term affective and narrative outcomes among older adults with cognitive impairment. The adaptive, multimodal design shows promise for human-AI collaboration in memoir writing and older adult care contexts, but larger, adequately powered randomized controlled trials with verified active control fidelity, multisession follow-up, and content source–differentiated outcome scoring are needed before clinical conclusions can be drawn.

(JMIR Hum Factors 2026;13:e83428) doi:10.2196/83428

KEYWORDS

AI-generated content; older adults with dementia; memoir writing; multimethod design; conversational robot; cognitive impairment

XSL•FO
RenderX

Introduction

Dementia affects over 50 million individuals globally, impairing memory, cognition, language, and emotional processing, and this number is projected to grow significantly with the aging population [1]. Among nonpharmacological interventions, memoir writing has been shown to support cognitive function and emotional healing by preserving personal narratives and enhancing identity [2,3]. AI-driven approaches to structuring fragmented life narratives have shown early technical promise in organizing and preserving personal stories for therapeutic use [4]. However, traditional self-narration methods impose high cognitive demands, particularly for individuals with language difficulties and fragmented memory [5]. Emerging AI-generated content (AIGC) technologies such as GPT-3.5 offer new opportunities for content generation with minimal user input [5,6]; yet, most current systems are not designed to adapt to the cognitive and emotional needs of people with dementia.

By 2050, the global prevalence of dementia is projected to exceed 152 million [1]. Memoir writing, as a fundamental nonpharmacological intervention, has been increasingly validated for its dual benefits in cognitive preservation and emotional healing [2,3]. Nonetheless, the cognitive impediments (eg, memory fragmentation) and emotional requirements (eg, loneliness) of this demographic are predominantly unaddressed by current technologies [5,7]. Generic AIGC tools do not possess the necessary adaptability for dementia care [7]. This research gap highlights the urgent need for personalized, cognition-adaptive AIGC solutions. In addition to cognitive support, memoir writing functions as a means of preserving narrative identity and dignity [8,9]. Bitenc [9] emphasized that coherent life narratives play a crucial role in preserving selfhood in individuals with dementia, reinforcing the emphasis on both cognitive and emotional support in the chatbot design.

Recent studies in human-computer interaction (HCI) emphasize the role of asymmetric collaboration [10,11], where AI systems adapt dynamically to users with varying cognitive abilities. Assistive technologies can act as "ability balancers," helping patients with dementia engage meaningfully despite limitations [12-14]. Yet, most existing tools focus on functional or informational support and do not address the emotional and narrative dimensions central to memoir writing. AIGC tools like Replika have been explored for emotional companionship and mental health applications [15,16], but they often lack medical adaptability and rigorous ethical safeguards required in dementia care [17]. Existing dementia-focused chatbots have primarily targeted family caregivers rather than patients themselves [18], leaving a gap in direct patient-facing conversational support. Traditional dementia-assistive technologies emphasize personalization and usability [7,19], but seldom provide real-time narrative generation or emotion-sensitive feedback.

To address these limitations, we propose a personalized AIGC-powered chatbot that dynamically adjusts its interaction based on user cognitive states measured by Mini-Mental State Examination (MMSE) scores [20] and responds to emotional needs using Kansei Engineering (KE) principles [21]. The system integrates methods from mechanical engineering, packaging design, and HCI, aiming to support memoir writing, improve emotional well-being, and enable adaptive, human-centered AI experiences for older adults with cognitive impairment. This interdisciplinary design adheres to contemporary evidence on nonpharmacological interventions in dementia, which have demonstrated efficacy in improving emotional well-being [22,23], and addresses the demand for emotion-centric and cognition-adaptive assistive technologies in HCI and gerontechnology research [24,25].

This paper makes 3 contributions. First, we present a cognition-adaptive AIGC chatbot designed specifically to support memoir writing among older adults with cognitive impairment. The system combines MMSE-based interaction control with a structured life-event memory system and caregiver-verified memory anchors. Second, we describe a reproducible multimethod design process integrating KE, Quality Function Deployment (QFD), Axiomatic Design (AD), and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) to translate affective and cognitive needs into system specifications. Third, we report a preliminary matched-pair evaluation involving 20 participants, using a minimization-based allocation procedure to examine short-term changes in affect, psychiatrist-rated memoir quality, and perceived usability. The study is intended as a proof-of-concept evaluation of the system's feasibility and potential rather than evidence of clinical efficacy. Additional design details are provided in Multimedia Appendix 1.

Methods

Research Objectives

This study aimed to design and evaluate a personalized AIGC-based chatbot to support older individuals with cognitive impairment in memoir writing and emotional support. The system dynamically adapts its interaction mode based on users' MMSE scores, offering keyword prompts for those with mild impairment and active narrative generation for those with severe cognitive decline. The chatbot identifies and responds to emotional needs such as loneliness and the desire for achievement through KE principles. Following system design, a user evaluation study assessed effectiveness.

System Design

Design Framework

To develop a personalized AIGC system supporting memoir writing for older individuals with dementia, we adopted a multistage collaborative design framework integrating KE, QFD, AD, and the TOPSIS decision model.

In the first stage, KE was applied to identify users' emotional and psychological needs, drawing on the dementia-care literature on memory loss, caregiver burden, and identity disruption to derive core affective themes such as "frustration from forgetting" and "the desire for recognition." In the second stage, QFD mapped these emotional needs to quantifiable technical specifications, such as ensuring a natural language processing

XSL•FO
RenderX

response delay of less than 2 seconds and incorporating a comforting tone in audio feedback. In the third stage, AD decomposed the system into 3 functionally independent but complementary modules: memory triggering, narrative generation, and emotional feedback. Finally, the TOPSIS method evaluated and selected among 3 intervention strategies (low, medium, and high AI proactivity), with emotional needs weighted more heavily than technical complexity to prioritize emotionally resonant interactions.

This multimethod framework is supported by existing literature. Li and Li [26] demonstrated that combining KE, QFD, and AD can improve user satisfaction in medical product design for older adults, providing empirical support for this multimethod approach. The emotional needs mapping in the QFD phase aligns with the finding of Raghunath et al [24] that translating user needs into structured design specifications can meaningfully reduce cognitive demands on users. The emphasis on emotional needs in the TOPSIS model is further supported by ethical research underscoring emotional well-being as a fundamental outcome in dementia care [23].

System Implementation

The system uses GPT-3.5 as its core AIGC engine, integrated within a companion device intended for home use. The device's main controller processor coordinates a Wi-Fi module for wireless connectivity; ultrasonic sensors and an indicating light for environmental sensing and status feedback; a speech recognition chip linked to a microphone for voice input; a speech decoding and playback module linked to a horn for audio output and to a USB flash drive, TF card, and flash memory for local audio storage; and a TC117 driving chip powered jointly by a battery and an SY8101B power module. Interaction is controlled through prompt engineering: high-level prompt templates define the system's role, task objectives, and interaction norms, with placeholders reserved for cognitive state variables. A typical template reads: "You are a memoir construction assistant for older adults. Based on the user's cognitive level (MMSE score: [X]), guide them to elaborate on life events [Event List] through simple, step-by-step questions. Ensure your responses are concise, empathetic, and aligned with the user's cognitive capacity; avoid complex sentence structures or ambiguous prompts."

Adaptive control logic maps MMSE scores to interaction strategies. The MMSE threshold of 20 follows established clinical practice: scores of 20 to 26 indicate mild cognitive impairment, and scores below 20 indicate moderate-to-severe impairment per standard neuropsychological guidelines [20]. When scores fall within 20 to 26 (mild cognitive impairment), the system prioritizes open-ended guided prompts. When scores fall below 20 (moderate impairment), the system switches to closed-ended, yes or no, or multiple-choice prompts to reduce cognitive load. Generation parameters are dynamically adjusted across three dimensions: (1) prompt granularity, with step-by-step prompts triggered for scores of 20 or below; (2) response length, limited to 50 to 80 words for scores below 20 and 100 to 150 words for scores of 26 or above; and (3) memory retrieval cues, with more sensory and contextual cues incorporated for users with lower scores. In the present sample (MMSE range 11‐23), all participants fell within the below 20 or 20‐26 tiers; the 26-or-above response-length strategy was not triggered.

Recorded life events serve as foundational memory anchors. Initial acquisition follows a 2-step process: trained researchers conduct semistructured interviews of 30 to 45 minutes, and family members or caregivers complete a supplementary questionnaire. Storage uses a structured relational database organized hierarchically to support efficient retrieval. Cross-session updates follow 3 rules: supplemental update of newly mentioned events, semantic deduplication (cosine similarity threshold ≥0.8), and emotional tagging based on subsequent narratives. Feng et al [27] confirmed that emotional tagging enhances memory cue efficacy.

The system combines voice guidance with a visual timeline interface. The voice module delivers preloaded audio responses—recorded and stored locally on the device rather than generated via cloud text-to-speech. Emotional alignment is achieved through a rule-based mapping layer: the adaptive control logic classifies each interaction turn into 1 of 5 predefined emotional states (encouragement, gentle redirection, empathetic acknowledgment, memory affirmation, and session closure) based on MMSE score, topic category, and detected response latency, and selects the corresponding preloaded audio clip. The visual timeline interface presents life events in a hierarchical, chronologically organized format, reducing barriers for users with language and cognitive decline.

Evaluation

Participants

Participants were recruited through online platforms (WeChat, Xiaohongshu, Bilibili, and Baidu) and in-person community outreach in Hebei, Henan, Tianjin, and Shandong provinces in China. Inclusion criteria: aged 65 to 80 years; clinically confirmed diagnosis of cognitive impairment by a neurologist or psychiatrist based on *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5) criteria [28]; basic verbal communication abilities; and voluntary participation with written informed consent from participant and legal representative. Exclusion criteria: poor health requiring continuous medical care, severe psychiatric or behavioral symptoms, and severe uncorrectable sensory impairments.

Psychiatrists were recruited as independent evaluators. Inclusion criteria: valid clinical license in psychiatry, active practice in frontline psychiatric settings, at least 5 years of independent clinical experience, and clinical or research experience in geriatric psychiatry or neurocognitive disorders. Psychiatrists were excluded for conflicts of interest, personal relationships with participants, or inability to maintain blinding. All psychiatrists completed standardized rater training before formal scoring.

Ethical Considerations

This study received ethics approval from the institutional review board of Southwest University (SW20250115). All procedures complied with the Declaration of Helsinki. Each participant received about US $15 as compensation. The use of generative

XSL•FO
RenderX

AI in memoir construction for individuals with reduced decision-making capacity raises ethical considerations beyond standard research ethics protocols. Four issues were addressed in the study design. First, informed consent explicitly covered AIGC: participants and their legal representatives were informed that the system may generate or paraphrase narrative text based on caregiver-verified memory anchors, and that this content would be incorporated into the memoir transcript. Second, output transparency was built into the caregiver verification workflow: caregiver-verified memory anchors were flagged in the system database, allowing a postsession distinction between content spoken directly by the participant, content paraphrased by the system, and content generated from verified anchors. Third, data governance: memoir-related inputs processed via the GPT-3.5 API were limited to anonymized life-event descriptors; no personally identifiable information was transmitted to the API, and all locally stored data are subject to deletion upon participant request. Fourth, caregiver verification was required before any memory anchor was added to the database, ensuring that AI-generated narrative content remained grounded in the participant's actual experience rather than system-inferred assumptions [17,23].

Procedures

Researchers collected demographic and clinical information including age, sex, diagnosis, medical history, MMSE score, and years of education. Participants completed the Technology Use Attitudes and Proficiency Questionnaire and the Positive and Negative Affect Schedule (PANAS) and narrated a recent personal story for 5 minutes. Audio recordings were transcribed with consent and rated by psychiatrists using predefined criteria (see Multimedia Appendix 2 for MMSE).

Factor matching minimized baseline imbalance on 4 variables: MMSE total score, baseline memoir quality, years of education, and technology use attitudes and proficiency. Pairs were formed only after all 20 participants had completed baseline assessment. A researcher not otherwise involved in the subsequent group-assignment step reviewed the baseline profiles of all enrolled participants across these same 4 variables and manually grouped them into 10 pairs of participants judged to be most similar in overall baseline profile; this pairing was completed independently of, and prior to, any group assignment. Group allocation used a minimization procedure rather than simple randomization: for each of these 10 pairs, a researcher not otherwise involved in participant recruitment or assessment determined, in real time, which assignment of the 2 paired participants would minimize cumulative imbalance between the intervention and control groups across the 4 matching variables, and assigned the participants accordingly. This was a nonrandomized allocation procedure: assignment was determined by the minimization algorithm's real-time balancing calculation rather than by chance (eg, a random-number generator or coin flip), and no random element was used at any stage of group assignment.

All participants completed a 10-minute oral memoir narration in the same controlled environment. The intervention group received assistance from the AIGC system, while the control group received an active control condition matched to the intervention group in session duration, number of prompting turns, frequency of audio feedback, device novelty, and researcher behavior, with only the AI-generated personalization and memory anchor–driven narrative removed. The entire process was audio-recorded and transcribed with participant consent. After the intervention, all participants completed the PANAS again. For the intervention group, AI-paraphrased content and AIGC were excluded from the transcript prior to rating, using the content-source flags recorded in the system database, so that only content spoken directly by the participant was retained for evaluation. Psychiatrists blinded to group assignment independently rated these participant-only transcripts; control-group transcripts were unaffected by this procedure, as they contained no AIGC. Intervention group participants also completed the System Usability Scale (SUS) and were interviewed about their experience (see Multimedia Appendix 3 for SUS).

All 10 intervention group participants completed a semistructured, one-on-one interview lasting approximately 15 to 20 minutes, using a guide (Multimedia Appendix 4) covering 5 domains: usability experience, emotional experience, perceived authenticity of AIGC, cognitive load, and willingness to continue use. Each domain began with a broad, nonleading question, followed by neutral, nondirective probes (eg, "could you say more about that?") to elicit elaboration without suggesting a valenced response. Interviews were audio-recorded with consent. Because the interview sample coincided with the full intervention arm, no independent saturation criterion was applied.

Data Analysis

Quantitative Analysis

Affective measures were assessed using the PANAS [29], which contains two 10-item subscales measuring positive affect (PA) and negative affect (NA) on a 5-point scale (1=very slightly or not at all; 5=extremely). Both subscale scores range from 10 to 50.

We assessed the consistency of the physicians' ratings, which serves as the foundation for the subsequent data analysis. Interrater reliability was good to excellent: composite memoir quality intraclass correlation coefficient (ICC)=0.82, narrative coherence ICC=0.83, richness of detail ICC=0.81, and personal uniqueness ICC=0.85.

Memoir quality was evaluated by psychiatrists across 3 dimensions: narrative coherence, richness of detail, and personal uniqueness, on a 5-point scale (1=very poor; 5=excellent). All psychiatrists rated transcripts in a fully blinded manner.

System usability was assessed using the simplified SUS [30], in which odd-numbered items are positively worded and even-numbered items are negatively worded; even-numbered items were reverse-scored prior to analysis, and the total score was computed as the sum of all adjusted item scores multiplied by 2.5, yielding a scale of 0 to 100. Technology use baseline was measured using a Technology Use Experience and Attitudes Questionnaire scored 0 to 40.

XSL•FO
RenderX

Analyses proceeded in four steps: (1) baseline comparability, (2) normality testing of within-group pre- and postdifference scores and of matched-pair between-group difference scores using the Shapiro-Wilk test, (3) within-group pre- and postcomparisons, and (4) between-group comparisons conducted on matched-pair differences, treating each matched pair as the unit of analysis. Because participants were allocated within matched pairs using a minimization procedure, the primary between-group analysis used paired *t* tests or Wilcoxon signed rank tests where normality of the matched-pair differences was violated. Parametric paired *t* tests were applied where normality held; Wilcoxon signed rank tests were applied elsewhere. All tests were 2-tailed (α=.05). Effect sizes are reported as Cohen *d* (normally distributed differences) or rank-biserial correlation *r* (nonnormal distributions).

To account for multiple comparisons across the 3 prespecified coprimary outcomes (PA, NA, and psychiatrist-rated memoir quality), we applied Benjamini-Hochberg correction (Q=0.05) to the 6 hypothesis tests involving the experimental condition (3 within-group and 3 between-group matched-pair comparisons). Baseline equivalence tests were excluded, as these test for nondifference and are not part of the confirmatory hypothesis family.

All quantitative analyses were conducted in R (version 4.3.2; R Foundation for Statistical Computing).

Qualitative Analysis

Audio recordings of the interviews were transcribed verbatim after participants gave consent, and researchers then checked the transcripts against the recordings and manually corrected obvious speech recognition errors and dialect expressions to ensure that the transcripts accurately reflected participants' original words. Transcripts were deidentified before analysis, retaining only participant identification numbers and no personally identifiable information, and all coding was carried out in NVivo software (Lumivero) to facilitate code management and retrieval.

The analysis used reflexive thematic analysis. Coding was carried out independently by 2 researchers, each of whom read through all transcripts and assigned initial codes, without aiming in advance for agreement between their coding results. After independent coding was completed, the 2 researchers compared their coding results through discussion, examined each disagreement in turn, reinterpreted the material in light of its original context, and reached consensus through discussion or retained legitimate alternative interpretations, an approach consistent with the methodological position of reflexive thematic analysis, which treats coding as an interpretive and subjective process of knowledge production. Building on the coding, related codes were grouped and merged to progressively develop higher-level themes and subthemes, a thematic map was produced to show the hierarchy and relationships among themes, and a complete audit trail was maintained to document how the coding evolved from initial codes to final themes. Throughout the analysis, the emerging themes were continually checked against the original interview material to avoid overinterpretation detached from the data itself, and the recordings were revisited when necessary to verify tone and context.

Results

Overview

This study recruited 20 participants aged 67 - 80 years. Baseline MMSE scores spanned the range of mild to moderate cognitive decline, and there was individual variation in baseline memoir quality (as assessed by a psychiatrist) as well as in attitudes toward and proficiency in technology use. Participants had 3 - 12 years of education, reflecting the characteristics of an older population with diverse educational backgrounds. Detailed participant information is provided in Multimedia Appendix 5. There were no participant dropouts throughout the entire experimental process.

Quantitative Results

Baseline comparability between the experimental (n=10) and control (n=10) groups was evaluated using paired *t* tests and standardized mean differences (SMDs). No statistically significant differences were found on any baseline variable (all P>.05), and all SMDs were negligible (|SMD|≤0.20), confirming successful group equivalence prior to the intervention (Table 1).

XSL•FO
RenderX

Table . Baseline demographic and clinical characteristics of experimental (n=10) and active control (n=10) participants in a single-session, nonrandomized, matched-pair evaluation of an AIGC[a]-based memoir-writing chatbot among older adults aged 67 - 80 years with clinically diagnosed cognitive impairment, recruited from Hebei, Henan, Tianjin, and Shandong provinces, China.

Variable	Experimental, mean (SD)	Control, mean (SD)	SMD[b]	*t* test[c] (*df*=9)	*P* value
MMSE[d] total score	17.60 (3.84)	17.80 (3.79)	−0.052	−1.000	.34
Psychiatrist-rated baseline memoir quality	2.77 (0.63)	2.85 (0.54)	−0.128	−0.605	.56
Technology attitude and proficiency	18.10 (4.07)	18.20 (4.24)	−0.024	−0.287	.78
Education (years)	6.60 (2.76)	6.10 (2.23)	+0.199	1.861	.10
Age (years)	72.70 (3.74)	73.10 (3.90)	−0.105	−1.500	.17

[a]AIGC: AI-generated content.

[b]SMD: standardized mean difference.

[c]Paired *t* test.

[d]MMSE: Mini-Mental State Examination.

As shown in Table 2, normality of difference scores was assessed using the Shapiro-Wilk test, separately for within-group pre- and postdifferences and for the matched-pair between-group differences that form the basis of the primary analysis. Within-group differences were normally distributed only for PA in the experimental group (W=0.866; P=.09); all remaining within-group differences violated the normality assumption (Ps<.05). Among the matched-pair between-group differences, PA (W=0.909; P=.28) was normally distributed and was analyzed with a paired *t* test; NA (W=0.843; P=.048) and psychiatrist-rated quality (W=0.833; P=.04) violated normality and were analyzed with the Wilcoxon signed rank test.

Table . Shapiro-Wilk tests of normality for within-group and matched-pair between-group difference scores in positive affect, negative affect, and psychiatrist-rated memoir quality, among 20 older adults with cognitive impairment participating in a single-session, nonrandomized, matched-pair evaluation of an AIGC[a]-based memoir-writing chatbot in China.

Difference tested	Group or level	*W*	*P* value	Normal (P≥.05)	Test selected
Positive affect	Within-group, experimental	0.866	.09	Yes	Paired *t* test
Positive affect	Within-group, control	0.774	.007	No	Wilcoxon
Negative affect	Within-group, experimental	0.768	.006	No	Wilcoxon
Negative affect	Within-group, control	0.731	.002	No	Wilcoxon
Psychiatrist quality	Within-group, experimental	0.802	.02	No	Wilcoxon
Psychiatrist quality	Within-group, control	0.794	.01	No	Wilcoxon
Positive affect	Between-group, matched-pair difference	0.909	.28	Yes	Paired *t* test
Negative affect	Between-group, matched-pair difference	0.843	.048	No	Wilcoxon
Psychiatrist quality	Between-group, matched-pair difference	0.833	.04	No	Wilcoxon

[a]AIGC: AI-generated content.

Within-group changes were examined separately for each group; all tests were 2-tailed (α=.05). In the experimental group, PA increased significantly from pretest (mean 23.50, SD 4.58) to posttest (mean 25.30, SD 4.74; Δ=1.80, 95% CI 0.55-3.05; t_9=3.25; P=.01; Cohen d=1.03); the control group showed no significant change (Δ=−0.20; W=10.0; P=.54; r=−0.29). NA decreased significantly in the experimental group from pretest (mean 22.10, SD 5.22) to posttest (mean 18.40, SD 3.69; Δ=−3.70, 95% CI −7.19 to −0.21; W=3.0; P=.01; r=−0.89); the control group showed no significant change (Δ=0.50; W=0.0; P=.09; r=1.00). Psychiatrist-rated memoir quality, assessed from participant-generated content only, improved significantly in the experimental group from pretest (mean 2.775, SD 0.63) to posttest (mean 2.95, SD 0.73; Δ=0.175, 95% CI 0.054-0.296;

XSL•FO
RenderX

W=0.0; P=.03; r=1.00); the control group showed no significant change (Δ=−0.05; W=2.5; P=.42; r=−0.50). Full within-group results are presented in Table 3. Additionally, Multimedia Appendix 6 presents separate ratings for AIGC.

Table . Within-group pre- to postintervention changes in positive affect, negative affect, and psychiatrist-rated memoir quality for the experimental (AIGC[a] chatbot-assisted) and active control groups, in a one-time experimental evaluation among 20 older adults with cognitive impairment in China.

Outcome and group	Preintervention, mean (SD)	Postintervention, mean (SD)	Δ (95% CI)	Test statistic	*P* value	Effect size
Positive affect						
Experimental	23.50 (4.58)	25.30 (4.74)	1.80 (0.55 to 3.05)	t_9=3.25	.01	d=1.03
Control	23.10 (4.25)	22.90 (3.96)	−0.20 (−1.14 to 0.74)	W=10.0	.54	r=−0.29
Negative affect						
Experimental	22.10 (5.22)	18.40 (3.69)	−3.70 (−7.19 to −0.21)	W=3.0	.01	r=−0.89
Control	25.30 (5.21)	25.80 (4.71)	0.50 (−0.01 to 1.01)	W=0.0	.09	r=1.00
Psychiatrist-rated quality						
Experimental	2.77 (0.63)	2.95 (0.73)	0.175 (0.054 to 0.296)	W=0.0	.03	r=1.00
Control	2.85 (0.54)	2.80 (0.57)	−0.05 (−0.16 to 0.06)	W=2.5	.42	r=−0.50

[a]AIGC: AI-generated content.

Between-group comparisons were conducted on matched-pair differences rather than independent-samples tests, because participants had been allocated within matched pairs, and each pair therefore constitutes the appropriate unit of comparison. Paired t tests were used for PA and psychiatrist-rated memoir quality, and the Wilcoxon signed rank test was used for NA, based on the normality of the matched-pair differences (Table 2). The experimental group showed significantly larger improvements than the control group across all 3 outcomes: PA (matched-pair Δ=2.00, 95% CI 0.31-3.69; t_9=2.68; P=.03; Cohen d=0.85), NA (matched-pair Δ=−4.20, 95% CI −7.59 to −0.81; W=2.0; P=.008; rank-biserial r=−0.93), and psychiatrist-rated memoir quality, assessed from participant-generated content only (matched-pair Δ=0.225, 95% CI 0.093-0.357; W=0; P=.02; rank-biserial r=1.00; Table 4).

Table . Matched-pair between-group comparison of change scores in positive affect, negative affect, and psychiatrist-rated memoir quality between the AIGC[a] chatbot-assisted experimental group and the active control group, among 10 matched pairs of older adults with cognitive impairment in a one-time experimental evaluation in China.

Outcome	Matched-pair Δ, mean (SD)	95% CI	Test statistic	*P* value	Effect size
Positive affect	2.00 (2.36)	0.31 to 3.69	t_9=2.68	.03	d=0.85
Negative affect	−4.20 (4.73)	−7.59 to −0.81	W=2.0	.008	r=−0.93
Psychiatrist-rated quality	0.225 (0.184)	0.093 to 0.357	W=0	.02	r=1.00

[a]AIGC: AI-generated content.

To account for multiple comparisons across the 3 prespecified coprimary outcomes (PA, NA, and psychiatrist-rated memoir quality), we applied Benjamini-Hochberg correction (Q=0.05) to the 6 hypothesis tests involving the experimental condition (3 within-group and 3 between-group matched-pair comparisons). Baseline equivalence tests in Table 1 were excluded, as these test for nondifference and are not part of the confirmatory hypothesis family. All 6 tests remained significant after correction (q≤0.031; Table 5). The 3 within-group control-condition comparisons, reported descriptively, remained nonsignificant (P≥.089).

XSL•FO
RenderX

Table . Benjamini-Hochberg (BH) false discovery rate correction applied to the 6 confirmatory hypothesis tests (3 within-group and 3 matched-pair between-group comparisons) for positive affect, negative affect, and psychiatrist-rated memoir quality, among 20 older adults with cognitive impairment participating in a single-session, nonrandomized, matched-pair evaluation of an AIGC[a]-based memoir-writing chatbot in China.

Test	Raw *P* value	Rank	BH-adjusted *q*	Significant ($q<0.05$)
Between-group, negative affect	.008	1	.024	Yes
Within-group, positive affect	.01	2	.024	Yes
Within-group, negative affect	.01	3	.024	Yes
Between-group, psychiatrist quality	.02	4	.024	Yes
Between-group, positive affect	.03	5	.030	Yes
Within-group, psychiatrist quality	.03	6	.031	Yes

[a]AIGC: AI-generated content.

System usability was assessed using the simplified SUS [30], in which odd-numbered items are positively worded and even-numbered items are negatively worded; even-numbered items were reverse-scored prior to analysis. The mean SUS total score among intervention group participants was 92.2 (SD 5.1; range 82.5 - 97.5), which exceeds the widely used acceptability threshold of 68 and falls within the "excellent" range (>85) for dementia-assistive technologies [31,32]. Individual scores ranged from 82.5 (Participant 10) to 97.5 (Participant 6 and Participant 7), with no participant scoring below the acceptable threshold. Subscale analysis following Orgeta et al [22] yielded a mean effectiveness score (items 1, 3, 5, 7, and 9) of 4.80 (SD 0.19) and a mean simplicity score (items 2, 6, 8, and 10) of 4.47 (SD 0.40), both on a 1 - 5 scale, indicating consistently high perceived effectiveness and adequate simplicity across participants.

The experimental group showed greater improvement than the active control group across PA, NA, and psychiatrist-rated memoir quality. Because the control condition was matched to the intervention condition in session duration, number of prompting turns, frequency of audio feedback, device exposure, and researcher behavior, the observed differences are less likely to be explained solely by interaction time or device novelty. However, the fidelity of the manually delivered control condition was not independently assessed, and the study was not designed to determine which specific component of the AIGC intervention—personalized prompting, memory-anchor retrieval, narrative generation, or their combination—accounted for the observed differences. The findings should therefore be interpreted as preliminary evidence concerning the integrated intervention rather than as proof of the effectiveness of any single system component.

Subgroup patterns offer additional insight. In the present sample, 12 of 20 participants (experimental: n=6; control: n=6) had MMSE scores below 20 (range 11 - 18), triggering the closed-ended, high-proactivity interaction mode; the remaining 8 (experimental: n=4; control: n=4) had MMSE scores of 20 - 23, triggering the open-ended, keyword-based mode. Participants in the lower MMSE tier appeared to benefit more from AI-driven narrative generation, while those in the higher tier responded better to keyword-based prompting. This is consistent with cognitive load theory [33]: when residual capacity is limited, proactive narrative scaffolding reduces the compositional burden and allows users to engage with their memories without being overwhelmed. Users with greater preserved capacity may find fully automated narrative generation less engaging, though the mechanism remains speculative. These patterns tentatively suggest that cognitive severity may shape the appropriate level of AI proactivity, and that a uniform interaction mode would be inadequate for this population. Given the small *n* within each tier (6 and 4 per group), these observations are descriptive and should not be interpreted as subgroup-level statistical conclusions.

Qualitative Results

First, most participants reported some initial trial and error during the first 1 or 2 conversational turns within the session, but that they became more familiar with the system as the number of turns increased. (Throughout this section, references to "uses" or "times" refer to individual conversational turns or exchanges within the single, approximately 10-minute session, not separate sessions on different occasions.) This pattern of initial difficulty followed by increasing fluency recurred throughout the interviews, particularly with regard to the timing and triggering of voice commands. Many participants mentioned that during their first few turns they were unsure "when they should speak and when they should wait for the system to finish speaking," and this uncertainty clearly diminished as the session progressed.

> *At the beginning I did not really know how to use it, but fortunately the system was simple, and after using it a few times I got the hang of it . Later on I did not really get stuck anymore. (The references to "few times" by the participants here refer to multiple rounds of dialogue, rather than the conduct of multiple experiments.)* [Participant 2]

XSL•FO
RenderX

Second, in terms of emotional experience, participants' descriptions showed a fairly consistent positive tendency. Most described the act of "telling their story" itself as a process of being taken seriously and being listened to. Some participants compared this with their everyday communication with family and friends, feeling that the system "does not interrupt" and "does not get impatient," which they saw as setting it apart from interpersonal communication.

> *It was a lot of fun. I think it gave me a fairly good window that made it convenient for me to express myself. How should I put it, it lets me talk, and I mean that in a meaningful sense, like confiding in someone, it made that more convenient. It was very patient, it did not interrupt me, and it let me finish speaking slowly.* [Participant 1]

Next is the issue of autonomy. This was the theme discussed in the greatest depth in the interviews, and the one on which participants' attitudes diverged most clearly. Most participants felt that even though the system organized and rewrote their spoken content, the resulting story still "belonged to them," because the core facts and emotions came from themselves and the system only played a supporting role in "polishing" or "recording" it.

> *I think the story is my own. As for the final generated result, when I look at most of it, I feel it matches what I wanted to express, and there is not really a problem.* [Participant 3]

However, a small number of participants (such as Participant 6 and Participant 8) expressed more nuanced reservations, noting that the wording the system produced after rewriting sometimes differed from their usual way of expressing themselves. They hoped to have the opportunity to check or revise the rewritten portions rather than simply accepting the final text passively.

> *Even though on the whole I feel most of it is fitting, to be honest some of the details still feel a bit off to me. For example that table, the one it generated, is not entirely how I wanted it to look. But it is true that if you wanted it to be completely accurate, that would probably take a lot of effort.* [Participant 6]

Regarding the system's design of adjusting its prompting style according to cognitive state (open-ended prompts vs stepwise closed-ended prompts), participant feedback showed a divergent trend related to their self-reported cognitive burden. Among participants with lower cognitive burden (corresponding to the MMSE stratum of 20 to 23), some felt that stepwise prompting was "a bit wordy" and preferred to speak directly and freely, feeling that their train of thought was interrupted.

> *I feel it could be more concise. Sometimes it says things that feel unnecessary, and I do not think those are really needed.* [Participant 7]

Participants with higher cognitive burden (corresponding to an MMSE score below 20), by contrast, were more inclined to report that stepwise prompting "was a huge help," reducing the pressure of organizing their language and easing the anxiety of "not knowing where to start."

> *I think it was extremely useful, it helped me a lot. Otherwise, having to do everything myself would really have been difficult.* [Participant 4]

This divergence corroborates the original intent behind this study's MMSE stratified design, namely, that greater cognitive burden calls for more structured guidance. It also suggests that future versions could allow users to actively switch the density of prompting during use rather than having it entirely preset by the system, so as to accommodate the desire for autonomous expression among participants with milder impairment. It should be noted that because this study is only a preliminary formative study with a small sample size, it is difficult to support adequate subgroup analysis, and the description above should be treated only as an indication of a trend.

With regard to willingness to continue using the system, all participants expressed a positive attitude, but this positive attitude often came with specific conditions. For example, whether the system could remember what had previously been said (to avoid the fatigue of repeating the same content) and whether interest would wane once the novelty faded.

> *It is easy to use, a lot of fun, and interesting, but I am not sure whether it can keep a record. I hope it can keep recording continuously and save all of it permanently.* [Participant 10]

One participant, Participant 8, linked the willingness to continue using the system to a deeper motivation, that of "leaving something behind for later generations," which goes beyond a purely instrumental evaluation of the tool and positions the system as a medium for recording and passing things on rather than simply a companion or entertainment tool.

> *After I am gone, this will count as a kind of digital photo album too, and that is a good thing.* [Participant 8]

Discussion

Principal Findings

Improvements in PA and NA, as measured by PANAS, suggest that the system may have supported emotional well-being during the memoir-writing session. This aligns with the KE component of the design framework, which translated affective needs (eg, the desire for recognition and anxiety about forgetting) into features such as empathetic audio feedback and sensory memory cues. Pappadà et al [7] noted that dementia-assistive technologies tend to prioritize cognitive function while underserving emotional needs; these results suggest that emotional engineering principles may contribute meaningful affective benefits.

The empirical findings, together with the design process, suggest 3 implications for future dementia-assistive AIGC systems. First, AI proactivity should adapt to cognitive capacity: more severe impairment calls for greater AI narrative initiative, while milder impairment is better served by user-driven prompting that preserves authorial agency, or the user's sense that the memoir remains their own story. Second, memoir tools should support affective safety alongside memory retrieval. Third, long-term personalization requires a structured life-event

XSL•FO
RenderX

memory system with caregiver verification mechanisms and ethical safeguards to ensure that AIGC remains grounded in the user's actual experience. The multimethod design framework (KE, QFD, AD, and TOPSIS) remains relatively underexplored in dementia-assistive technology [21] and may offer a reproducible starting point for similar work.

Comparison to Prior Work

This study extends prior work on AI-mediated reminiscence and narrative support for older adults [5,12,27]. More broadly, multimodal AI-driven platforms have also shown promise in supporting expressive and creative interventions for home-based older adult care, such as interactive art therapy systems [34], reinforcing the potential of AI-based creative engagement to benefit older populations beyond memoir writing specifically. Wang et al [12] demonstrated that AI-driven multimodal photo albums support personalized reminiscence in cognitively intact older adults; our study extends this to individuals with diagnosed cognitive impairment, using MMSE scores to drive adaptive interaction control. Feng et al [27] found that contextual and multimodal stimuli enhanced engagement in people with dementia during human-robot interaction; our results suggest that similar multimodal principles may be effective in chatbot-mediated memoir writing. Chang et al [35] further demonstrated that reinforcement learning-based AI can deliver personalized adaptive support in dementia care contexts, reinforcing the value of dynamic AI adaptation for this population.

The present findings also converge with recent longitudinal evidence. Kang et al (2025) [36] found that regular AI care calls over 7 months significantly reduced depression scores and improved memory function in 80 community-dwelling individuals with dementia. That study found that sex and education level moderated intervention outcomes, paralleling our subgroup finding that cognitive severity shapes response to AI interaction and reinforcing the view that individual characteristics meaningfully affect how people with dementia benefit from AI-based support.

The findings also speak to HCI discussions of asymmetric human-AI collaboration [10,37], in which AI systems dynamically compensate for differential user capabilities. Here, narrative initiative is inversely calibrated to residual cognitive capacity, so the AI assumes greater responsibility where the user needs it most. This contrasts with many prior dementia-assistive systems that rely on relatively fixed interaction modes [7]. Whether similar mechanisms generalize to other cognitive assessment instruments remains a question for future work.

Limitations

However, this study also has some limitations. First, the sample size (n=20; 10 per arm) is small. Effect sizes derived from small samples are inherently unstable, and the Benjamini-Hochberg correction applied to the 6 confirmatory hypothesis tests should be read with caution: because within- and between-group tests were derived from the same 10 participants per arm rather than independent samples, the positive regression–dependence assumption underlying the procedure was plausible but not formally verified, and several outcomes assessed with Wilcoxon signed rank tests at n=9‐10 yield P and q values that take on a limited set of discrete values. The correction should be read as a conservative response to multiplicity concerns rather than a substitute for replication in a larger, adequately powered, multisite trial.

Second, this study used a nonrandomized design: group allocation used a minimization procedure rather than simple or block randomization. This method was selected specifically to ensure equivalence on 4 baseline variables, given the small overall sample size, and is a recognized alternative to simple randomization in small trials. However, it differs from simple randomization in that assignment for later-entered pairs is partly determined by the accumulated composition of previously assigned groups rather than by chance alone. As a result, unmeasured or unbalanced covariates could, in principle, be associated with allocation in ways that pure randomization is designed to prevent, and a person with knowledge of the algorithm and running group totals could in principle anticipate upcoming assignments. Larger-scale replications should use simple or stratified block randomization with formal allocation concealment to strengthen causal inference.

Third, because participants were allocated within matched pairs, all confirmatory hypothesis tests were analyzed using paired methods (paired t tests or Wilcoxon signed rank tests on matched-pair differences) rather than independent-samples tests, consistent with the matched design. This is a more conservative and design-appropriate approach than treating change scores as independent observations; all 6 confirmatory comparisons remained statistically significant under this analysis and after Benjamini-Hochberg correction, with moderate-to-large effect sizes (Cohen d=0.85‐1.03; rank-biserial r=−0.93 to 1.00).

Fourth, this was a one-time experimental evaluation without follow-up assessment. This has 2 consequences. It precludes conclusions about the durability of affective and narrative gains; whether benefits persist, accumulate with repeated use, or diminish with habituation is unknown. It also means the high SUS scores (mean 92.2, SD 5.1) should be interpreted cautiously: single exposure to a novel system can temporarily inflate satisfaction ratings, and older adults with cognitive impairment are additionally prone to acquiescence bias, meaning the tendency to agree with survey items regardless of content, particularly under limited cognitive resources [38]. Holden [30] noted similar interpretive challenges applying the SUS in this population. Longitudinal, multisession usability assessment is needed to obtain a more reliable estimate of system acceptability.

Fifth, participants were recruited exclusively from 4 Chinese provinces, with a mean education level of approximately 6 years and corresponding technology-use attitudes. System performance may differ in populations with higher digital literacy, different cultural or linguistic backgrounds, or distinct dementia etiologies, and these findings should not be generalized beyond this context without further testing.

Sixth, the system's core AIGC engine relies on GPT-3.5. GPT-3.5 represented a relatively advanced, readily available commercial large language model at the time this system was

XSL•FO
RenderX

developed, and its selection was reasonable within that context. However, given the pace at which generative AI models have since evolved, findings obtained with this specific model may not generalize to systems built on newer architectures. Replication with current-generation models is needed, alongside systematic safety evaluation under diverse real-world conditions, which has not yet been conducted for either version.

Seventh, the control condition requires clarification and cautious interpretation. As described in the Procedures section, the control group received an active control matched to the intervention group on session duration, number of prompting turns, frequency of audio feedback, device novelty, and researcher behavior, with only the AI-generated personalization and memory anchor–driven narrative content removed. This design was intended to isolate the specific contribution of AI-generated scaffolding rather than general attention or novelty effects. However, no formal fidelity check (eg, independent rating of researcher adherence to the matched protocol) was conducted to confirm that the manually delivered control condition in fact replicated the intervention's interaction structure and pacing. Residual, unmeasured differences in perceived responsiveness or engagement between human-delivered and AI-delivered support therefore cannot be fully excluded, and this should be addressed in future trials through structured fidelity monitoring or a non-AI digital comparator.

Eighth, psychiatrist-rated memoir quality was assessed from transcripts with AI-paraphrased content and AIGC excluded, so that scores reflect participant-generated narrative content specifically rather than the system's language capabilities.

Conclusions

AI-assisted memoir construction for people with reduced decision-making capacity raises important questions about authorship, narrative authenticity, and data governance. As described in the Ethical Considerations section, this study addressed these concerns through explicit consent for AIGC, caregiver verification of generated narratives, anonymized API data transmission, and postsession transparency about system outputs. Future deployments should extend these safeguards to longitudinal use by formalizing policies for data retention, access, revision, and deletion [17,23].

Future studies should evaluate the system across multiple sessions and with stronger comparison conditions, such as structured human prompting or non-AI digital memoir tools, to better isolate the contribution of AI-generated scaffolding. A larger and more diverse randomized controlled trial would also enable subgroup analyses by dementia etiology, severity, education, and cultural context. Finally, because the current MMSE-based adaptation logic remains relatively coarse, future versions could incorporate real-time behavioral signals, such as response latency, lexical complexity, and speech-detected emotional tone, to support more fine-grained within-session adaptation [39]. Together, these steps would help move the system from a promising single-session prototype toward a clinically useful and ethically accountable long-term reminiscence support tool.

Acknowledgments

The authors thank all the older participants and physicians for their outstanding contributions to this study. Generative AI (GPT-3.5) was part of our design; however, it was not used for research or writing the paper.

Funding

This research was supported by the British Academy/Leverhulme Small Research Grant (SRG2526/260306).

Authors' Contributions

Conceptualization: YM, YQ, ZY, BL, ZW
Data curation: YM, YQ, ZY, BL, ZW, MY
Formal analysis: YM, YQ, ZY, BL, ZW, MY
Funding acquisition: YM, HL
Investigation: YM, YQ, ZY, BL, ZW, MY
Methodology: YM, YQ, ZY, BL, ZW, MY
Project administration: YM
Resources: YM, HL
Supervision: YM, HL
Validation: YM, YQ, ZY, BL, ZW
Visualization: YM, YQ, ZY, BL, ZW
Writing – original draft: YM, YQ, ZY, BL, ZW, MY
Writing – review & editing: YM, HL

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Hardware prototype.
[DOCX File, 1871 KB - humanfactors_v13i1e83428_app1.docx]

Multimedia Appendix 2
Items of the Technology Use Experience and Attitudes Questionnaire.
[DOCX File, 16 KB - humanfactors_v13i1e83428_app2.docx]

Multimedia Appendix 3
Items of the Simplified System Usability Scale.
[DOCX File, 15 KB - humanfactors_v13i1e83428_app3.docx]

Multimedia Appendix 4
Semistructured interview guide.
[DOCX File, 16 KB - humanfactors_v13i1e83428_app4.docx]

Multimedia Appendix 5
Baseline demographic and clinical characteristics of 20 individually matched participants.
[DOCX File, 15 KB - humanfactors_v13i1e83428_app5.docx]

Multimedia Appendix 6
Sensitivity comparison of psychiatrist-rated memoir quality scored from human-only versus AI-inclusive transcript content among 10 experimental-group participants.
[DOCX File, 14 KB - humanfactors_v13i1e83428_app6.docx]

References

1. Nichols E, Vollset SE, Steinmetz JD, et al. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. Lancet Public Health 2022 Feb;7(2):e105-e125. [doi: 10.1016/S2468-2667(21)00249-8]
2. Ryan EB, Bannister KA, Anas AP. The dementia narrative: writing to reclaim social identity. J Aging Stud 2009 Aug;23(3):145-157. [doi: 10.1016/j.jaging.2007.12.018]
3. Zawadzki P, Adamczyk AK. To remember, or not to remember? Potential impact of memory modification on narrative identity, personal agency, mental health, and well-being. Bioethics 2021 Nov;35(9):891-899. [doi: 10.1111/bioe.12926] [Medline: 34427951]
4. Gui F, Yang M, Jin L, Qian J. AI-enhanced automatic life story structuring for reminiscence therapy in older adults: technical feasibility study. JMIR Aging 2026 Apr 6;9:e83122. [doi: 10.2196/83122] [Medline: 41941365]
5. Maddali HT, Dixon E, Pradhan A, Lazar A. Investigating the potential of artificial intelligence powered interfaces to support different types of memory for people with dementia. Presented at: CHI '22: CHI Conference on Human Factors in Computing Systems; Apr 29 to May 5, 2022. [doi: 10.1145/3491101.3519858]
6. Lahat A, Sharif K, Zoabi N, et al. Assessing generative pretrained transformers (GPT) in clinical decision-making: comparative analysis of GPT-3.5 and GPT-4. J Med Internet Res 2024 Jun 27;26:e54571. [doi: 10.2196/54571] [Medline: 38935937]
7. Pappadà A, Chattat R, Chirico I, Valente M, Ottoboni G. Assistive technologies in dementia care: an updated analysis of the literature. Front Psychol 2021;12:644587. [doi: 10.3389/fpsyg.2021.644587] [Medline: 33841281]
8. Heersmink R. Preserving narrative identity for dementia patients: embodiment, active environments, and distributed memory. Neuroethics 2022 Apr;15(1):8. [doi: 10.1007/s12152-022-09479-x]
9. Bitenc R. Dementia Narratives in Contemporary Literature, Life Writing, and Film [Doctoral thesis]. : Durham University; 2017 URL: https://etheses.durham.ac.uk/id/eprint/12157/ [accessed 2026-08-17]
10. Bréhault V, Dubois E, Prouzeau A, Serrano M. A systematic literature review to characterize asymmetric interaction in collaborative systems. Presented at: CHI '25: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems; Apr 26 to May 1, 2025. [doi: 10.1145/3706598.3713129]
11. Dritsas E, Trigka M, Vonitsanos G, Mylonas P. Human-computer interaction in the big data era: systems, cognition, and interactive intelligence. Presented at: CHIGreece '25: Proceedings of the 3rd International Conference of the ACM Greek SIGCHI Chapter; Sep 24-26, 2025. [doi: 10.1145/3749012.3749070]
12. Wang X, Li J, Liang T, et al. Promoting personalized reminiscence among cognitively intact older adults through an AI-driven interactive multimodal photo album: development and usability study. JMIR Aging 2024 Jan 23;7:e49415. [doi: 10.2196/49415] [Medline: 38261365]

XSL•FO
RenderX

13. Wolfe BH, Oh YJ, Choung H, et al. Caregiving artificial intelligence chatbot for older adults and their preferences, well-being, and social connectivity: mixed-method study. J Med Internet Res 2025 Mar 13;27:e65776. [doi: 10.2196/65776] [Medline: 40080043]
14. Hasan WU, Zaman KT, Wang X, Li J, Xie B, Tao C. Empowering Alzheimer's caregivers with conversational AI: a novel approach for enhanced communication and personalized support. NPJ Biomed Innov 2024 Dec 4;1(1):3. [doi: 10.1038/s44385-024-00004-8] [Medline: 42032173]
15. Wang X, Pang CC, Hui P. 'My Dataset of Love': a preliminary mixed-method exploration of human-AI romantic relationships. Proc ACM Hum-Comput Interact 2025 Oct 18;9(7):1-34. [doi: 10.1145/3757532] [Medline: 42023015]
16. Laestadius L, Bishop A, Gonzalez M, Illenčík D, Campos-Castillo C. Too human and not human enough: a grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. New Media Soc 2024 Oct;26(10):5923-5941. [doi: 10.1177/14614448221142007]
17. Wang T, Zhang Y, Qi S, Zhao R, Xia Z, Weng J. Security and privacy on generative data in AIGC: a survey. ACM Comput Surv 2025 Apr 30;57(4):1-34. [doi: 10.1145/3703626]
18. Cheng ST, Ng PHF, Wakui T. The PDC30 chatbot-development of a psychoeducational resource on dementia caregiving among family caregivers: mixed methods acceptability study. JMIR Aging 2025 Jan 6;8:e63715. [doi: 10.2196/63715] [Medline: 39761547]
19. Miguel Cruz A, Daum C, Comeau A, et al. Acceptance, adoption, and usability of information and communication technologies for people living with dementia and their care partners: a systematic review. Disabil Rehabil Assist Technol 2023 May;18(4):443-457. [doi: 10.1080/17483107.2020.1864671] [Medline: 33378627]
20. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
21. Cao S, Yang F, Yang W. A framework for applying kansei engineering principles in the design of small household appliances. In: Marcus A, Rosenzweig E, Soares MM, editors. Design, User Experience, and Usability: Springer Nature Switzerland; 2024:16-27. [doi: 10.1007/978-3-031-61362-3_2]
22. Orgeta V, Qazi A, Spector AE, Orrell M. Psychological treatments for depression and anxiety in dementia and mild cognitive impairment. Cochrane Database Syst Rev 2015(1):CD009125. [doi: 10.1002/14651858.CD009125.pub2]
23. Zhang Q, Zhang R, Xiong Y, Sui Y, Tong C, Lin FH. Generative AI mental health chatbots as therapeutic tools: systematic review and meta-analysis of their role in reducing mental health issues. J Med Internet Res 2025 Dec 16;27:e78238. [doi: 10.2196/78238] [Medline: 41401240]
24. Raghunath N, Dahmen J, Brown K, Cook D, Schmitter-Edgecombe M. Creating a digital memory notebook application for individuals with mild cognitive impairment to support everyday functioning. Disabil Rehabil Assist Technol 2020 May;15(4):421-431. [doi: 10.1080/17483107.2019.1587017] [Medline: 30907223]
25. Chen LH, Liu YC. Interface design for products for users with advanced age and cognitive impairment. Int J Environ Res Public Health 2022 Feb 21;19(4):2466. [doi: 10.3390/ijerph19042466] [Medline: 35206653]
26. Li X, Li H. Age-appropriate design of domestic intelligent medical products: an example of smart blood glucose detector for the elderly with AHP-QFD Joint KE. Heliyon 2024 Mar;10(5):e27387. [doi: 10.1016/j.heliyon.2024.e27387]
27. Feng Y, Perugia G, Yu S, Barakova EI, Hu J, Rauterberg GWM. Context-enhanced human-robot interaction: exploring the role of system interactivity and multimodal stimuli on the engagement of people with dementia. Int J Soc Robot 2022 Apr;14(3):807-826. [doi: 10.1007/s12369-021-00823-4]
28. Sachdev PS, Blacker D, Blazer DG, et al. Classifying neurocognitive disorders: the DSM-5 approach. Nat Rev Neurol 2014 Nov;10(11):634-642. [doi: 10.1038/nrneurol.2014.181] [Medline: 25266297]
29. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037/0022-3514.54.6.1063]
30. Holden RJ. A Simplified System Usability Scale (SUS) for cognitively impaired and older adults. Proc Int Symp Hum Factors Ergon Health Care 2020 Sep;9(1):180-182. [doi: 10.1177/2327857920091021]
31. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum-Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
32. Chien SY, Zaslavsky O, Berridge C. Technology usability for people living with dementia: concept analysis. JMIR Aging 2024 Jul 3;7(1):e51987. [doi: 10.2196/51987] [Medline: 38959053]
33. Sweller J. Cognitive load during problem solving: effects on learning. Cogn Sci 1988 Apr;12(2):257-285. [doi: 10.1207/s15516709cog1202_4]
34. Sang D, Miao L, Wu Q. A smart community interactive art therapy platform based on multimodal computer graphics and resilient artificial intelligence for home-based elderly care. Sci Rep 2025 Dec 3;15(1):43057. [doi: 10.1038/s41598-025-27129-z] [Medline: 41339409]
35. Chang WT, Wang S, Kramer S, et al. Human-centered AI for dementia care: using reinforcement learning for personalized interventions support in eating and drinking scenarios. Presented at: HHAI 2024: Hybrid Human AI Systems for the Social Good; Jun 10-14, 2024. [doi: 10.3233/FAIA240185]
36. Kang S, Lee Y, Kim N, et al. Beneficial effect of artificial intelligence care call on memory and depression in community dwelling individuals with dementia. Sci Rep 2025;15(1):27116. [doi: 10.1038/s41598-025-12895-7]

XSL•FO
RenderX

37. Gomez C, Cho SM, Ke S, Huang CM, Unberath M. Human-AI collaboration is not very collaborative yet: a taxonomy of interaction patterns in AI-assisted decision making from a systematic review. Front Comput Sci 2024;6:1521066. [doi: 10.3389/fcomp.2024.1521066]
38. Rabbitt P, Abson V. "Lost and found": some logical and methodological limitations of self-report questionnaires as tools to study cognitive ageing. Br J Psychol 1990 Feb;81 (Pt 1)(1):1-16. [doi: 10.1111/j.2044-8295.1990.tb02342.x] [Medline: 2322763]
39. Kosch T, Karolus J, Zagermann J, Reiterer H, Schmidt A, Woźniak PW. A survey on measuring cognitive workload in human-computer interaction. ACM Comput Surv 2023 Dec 31;55(13s):1-39. [doi: 10.1145/3582272]

Abbreviations

AD: Axiomatic Design
AIGC: AI-generated content
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
HCI: human-computer interaction
ICC: intraclass correlation coefficient
KE: Kansei Engineering
MMSE: Mini-Mental State Examination
NA: negative affect
PA: positive affect
PANAS: Positive and Negative Affect Schedule
QFD: Quality Function Deployment
SMD: standardized mean difference
SUS: System Usability Scale
TOPSIS: Technique for Order of Preference by Similarity to Ideal Solution

Edited by J Sarvestan; submitted 02.Sep.2025; peer-reviewed by J Zhong, M Li; revised version received 16.Jul.2026; accepted 17.Jul.2026; published 25.Aug.2026.

Please cite as:
Meng Y, Que Y, Yan Z, Liu B, Wang Z, Yang M, Lu H
Personalized Intelligent Chatbot Based on AI-Generated Content Assists Memoir Writing for Older Adults With Cognitive Impairment: Mixed Methods Study
JMIR Hum Factors 2026;13:e83428
URL: https://humanfactors.jmir.org/2026/1/e83428
doi:10.2196/83428

© Yibo Meng, Yuan Que, Zhe Yan, Bingyi Liu, Zixin Wang, Mandi Yang, Huidi Lu. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 25.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

User-Centered Virtual Clinic Services for Remote, Rural, and Underserved Sub-Saharan Africa: Development and Usability Study

Abby Blocker[1], BSc; Ahmed Biyabani[2], SB, MS, PhD; Joyce Mwangama[3], BSc, MSc, PhD; Mohammed Ishaaq Datay[4], MMed, MBChB, DPhil, FCP (SA); Bessie Malila[1,5], BSc, MSc, PhD

[1]Division of Biomedical Engineering, University of Cape Town, 7th Floor, Anatomy Building, 1 Anzio Road, Observatory, Cape Town, South Africa

[2]Carnegie Mellon University Africa, Kigali, Rwanda

[3]Department of Electrical Engineering, University of Cape Town, Cape Town, South Africa

[4]Primary Health Care Directorate, University of Cape Town, Cape Town, South Africa

[5]Department of Electrical and Electronic Engineering Science, University of Johannesburg, Johannesburg, South Africa

Corresponding Author:
Abby Blocker, BSc
Division of Biomedical Engineering, University of Cape Town, 7th Floor, Anatomy Building, 1 Anzio Road, Observatory, Cape Town, South Africa

Abstract

Background: Virtual clinics allow doctors to connect to patients in difficult-to-reach locations. While this can result in improved access to care, implementing existing virtual clinics in these locations is difficult due to their contextual constraints. To address these challenges, a virtual clinic system for remote, rural, and underserved areas was developed based on user-centered design principles. To complete the user-centered design cycle, the developed system was then implemented and evaluated in the target contexts.

Objective: This study aimed to conduct a pilot study with the developed user-centered virtual clinic system to evaluate its performance in resource-limited settings. Patient, doctor, and nurse feedback was collected to understand how the virtual clinic system and implementation strategy might be improved before long-term implementation.

Methods: A pilot study was conducted at 2 health care facilities in South Africa—1 underserved public primary health clinic in the city of Cape Town, and 1 remote occupational health clinic in the Northern Cape Province. Doctor and nurse dyads participated in using the system to consult with real patients in each clinic. Patients received both virtual and physical examinations by the same doctor, and these 2 consultations were compared with one another. Surveys were conducted to gather patient feedback pre– and post–virtual clinic consult. Observations of system usage were collected, and doctor and nurse participant interviews were conducted at the conclusion of each day. Qualitative data were thematically analyzed to understand barriers and facilitators of virtual consultations, as well as patient satisfaction.

Results: A total of 12 patient consultations were conducted using the virtual clinic system across 2 health care facilities. Two doctors and 2 nurses participated in the study as users. Doctor and nurse confidence in using the system increased over time. Doctors were confident that the virtual consultation format could be used to diagnose patients remotely. Key indicators of patient satisfaction were being included in consultation communication and understanding the benefits virtual care could offer. Potential barriers to virtual consultations were infrastructure offered by the implementation environment and medical device limitations.

Conclusions: The results from this pilot study indicate that virtual clinic consultation is possible in low-resource settings using the developed virtual clinic system. This system can support patients in remote, rural, and underserved areas to receive certain health care services from a doctor without the doctor having to be physically present in the facility. The results encourage further scaling of the system to support long-term implementation in low-resource health clinics in South Africa and other sub-Saharan African countries.

(JMIR Hum Factors 2026;13:e84272) doi:10.2196/84272

KEYWORDS

telemedicine; remote consultation; digital health; user-centered design; South Africa; Africa South of the Sahara; rural health; resource-limited settings; medically underserved area; primary health care

XSL•FO
RenderX

Introduction

Background

Virtual clinics are a type of digital health solution that connect health care professionals to patients and other staff through digital means [1]. Implementation of virtual clinics is associated with many advantages: better quality and continuity of care [2-5], greater efficiency leading to reduced wait times [6,7], and reduced costs for both patients and health systems [8,9]. Due to the COVID-19 pandemic, the number of virtual clinics has grown [10], and implementations have been realized globally with a range of different applications.

Virtual Clinics in Sub-Saharan Africa

While virtual clinics have become common in many urban and developed areas, they have not yet reached many remote, rural, and underserved areas. This is because these areas present contextual challenges, including a lack of telecommunication infrastructure [11,12], limited access to electricity [13], and low levels of digital literacy [14]. Current virtual clinic services have not been adequately designed for these constraints, contributing to a lack of adoption in these areas. Furthermore, populations in remote, rural, and underserved areas often lack proper health facilities and skilled health professionals [15]. Virtual clinics have the potential to greatly improve access to care in these areas [16].

National programs in sub-Saharan Africa have attempted to implement digital health policy and services with mixed results. A national telemedicine program was implemented in Cabo Verde to connect doctors across the 9 inhabited islands of the country [16]. There were several technical challenges with this implementation, including a cyberattack that affected service provision, extensive financial investment, and the inability to maintain the hardware. Evidence collected in a review of telemedicine in sub-Saharan Africa corroborates these challenges [17]. Ogundaini and Mlitwa [18] discuss how digital health initiatives across sub-Saharan Africa require improved capacity building, knowledge sharing, and resources to be successful. South Africa, Kenya, and Nigeria have emerged as leaders on the continent for telemedicine implementations; however, digital literacy limitations, regulatory gaps, and infrastructure constraints were mentioned as limiting factors to success [19]. Implementation recommendations from the World Health Organization [16] and the International Telecommunication Union [20] provide some guidance for addressing these challenges, but alignment between recommendation and existing digital health system design still persists.

User-Centered Design for Virtual Clinics

User-centered design (UCD) is a methodology that may be applied during the design and development process to align the needs of the context with the digital health system [21-25]. UCD focuses on user needs to iteratively develop solutions [26,27]. There are 4 steps in the UCD process, which are (1) specify context of use, (2) specify requirements, (3) create design solutions, and (4) evaluate designs. Incorporation of UCD can lead to improved user interfaces and functionalities within the system. By incorporating user feedback early in the design process, development time decreases and user acceptance increases [28,29]. Several virtual clinics that have implemented UCD methods have shown improved user experience and system uptake [30,31].

Previous work outlined the application of UCD methods to develop a virtual clinic system for sub-Saharan Africa [32]. In brief, a situational analysis was conducted to understand the challenges associated with health care provision and digital health in austere environments in sub-Saharan Africa and South Africa. This addressed step 1 of the UCD process. From this situational analysis, system requirements were developed, addressing UCD step 2. The virtual clinic was designed according to these requirements, completing UCD step 3. Initial usability testing was conducted with nurses and doctors in Cape Town, completing UCD step 4. This completed the initial UCD iteration. Another UCD iteration was then begun by using the feedback from the usability study to further improve the next prototype.

The virtual clinic was designed for provider-to-provider synchronous telemedicine. In this model, a patient would visit their regular health clinic in a remote, rural, or underserved location. It is assumed that there would be limited or no doctors available to provide services in person. If the nurse or community health worker is unsure how to diagnose or treat a patient, they would typically refer the patient to a district or tertiary health facility or ask them to wait for the doctor to visit the facility. Instead, using the virtual clinic system, the nurse can call an available doctor who can advise on how to proceed with the patient remotely. The patient would either be adequately treated via the virtual consultation, or the doctor can confirm that the patient should be referred. This reduces unnecessary referrals, which is an advantage for both the patient and the health care system.

This design acknowledges the needs of low-resource settings in several ways. First, the provider-to-provider model reduces the onus on the patient that is typically associated with virtual care. The patient does not need to purchase any devices or have proper infrastructure such as electricity or internet in their home to participate in virtual consultations. The nurse is the user who facilitates the digital connection for the patient. Additionally, the system and infrastructure can be concentrated at a central point (the existing clinic facility) and shared among many patients, addressing the potential lack of resources for health care in these areas. Finally, it allows patients to have access to a doctor. Many patients do not fulfill referral appointments for various reasons, including cost, lack of transportation, long distances with poor health, and lack of child or older adult care. By implementing the virtual clinic system, barriers to care are reduced, and patients have an opportunity to receive better quality health care.

Research Objectives

In this paper, we present results of the evaluation of the virtual clinic system that was developed using UCD methods in an effort to address access to care in remote, rural, and underserved areas in sub-Saharan Africa. The improved system was evaluated

XSL•FO
RenderX

during this study in real-world contexts, completing step 4 and thus the second iteration of the UCD process.

The UCD method emphasizes usability and contextually relevant design, so this evaluation was focused on understanding how the virtual clinic system performed in real-world environments and how the users experienced the system. Specifically, the objectives were to understand how the virtual clinic performed in a real-world setting and to identify potential barriers that could be encountered during a scaled-up implementation, which answer the following three research questions: (1) What barriers do users encounter when delivering primary-level care using the virtual clinic? (2) What user behaviors are commonly present in successful virtual clinic consultations? (3) How satisfied are patients with receiving primary-level care using the virtual clinic? The following sections present the methods used for answering these research questions, the results of the data collection, and the recommendations for future implementation, as well as the limitations of this study.

Methods

Theoretical Background

There is extensive literature documenting the evaluation of telemedicine systems, and these evaluations vary significantly in their objectives. Ekeland et al [33] provide a review of methodologies for evaluating telemedicine systems and highlight the difficulties in aligning these methodologies. They highlight that telemedicine exists at the intersection of medical, technological, and social sciences, and it can be difficult to reconcile evaluation methods that satisfy each of these fields. The differences can be summarized as positivism versus naturalism. Positivism uses objective observation and measurement to develop generalizable laws, while naturalism focuses on understanding phenomena within their natural context, often through qualitative methods, and acknowledges the influence of various factors on human behavior. Both approaches can yield valuable insights, with positivist evaluations such as randomized controlled trials being important for establishing clinical and generalized outcomes and naturalist evaluations establishing context-relevant information and human experience and behaviors. They highlight that the context of the telemedicine implementation and objectives of the evaluation should dictate the appropriate approach.

In regions where telemedicine is still evolving, such as sub-Saharan Africa and South Africa, the strict design of positivist evaluations may not always be appropriate. Randomized controlled trials may not account for rapid developments in technology, leading to results that may not be relevant by the time the study concludes [34]. Furthermore, in contexts where telemedicine has not yet been implemented, gathering human experience and feedback via a naturalistic approach can be advantageous to understand the cultural and contextual effects of telemedicine implementations [35]. This also allows for refinement of implementation techniques that are context-based.

Given that the developed virtual clinic system is still at an early technology readiness level, and telemedicine within remote, rural, and underserved areas in South Africa has not yet been widely attempted, the evaluation approach used in this study leans naturalist rather than positivist. However, elements of both are incorporated to strengthen the study design and outcomes. Therefore, a mixed methods feasibility study was used to evaluate the virtual clinic.

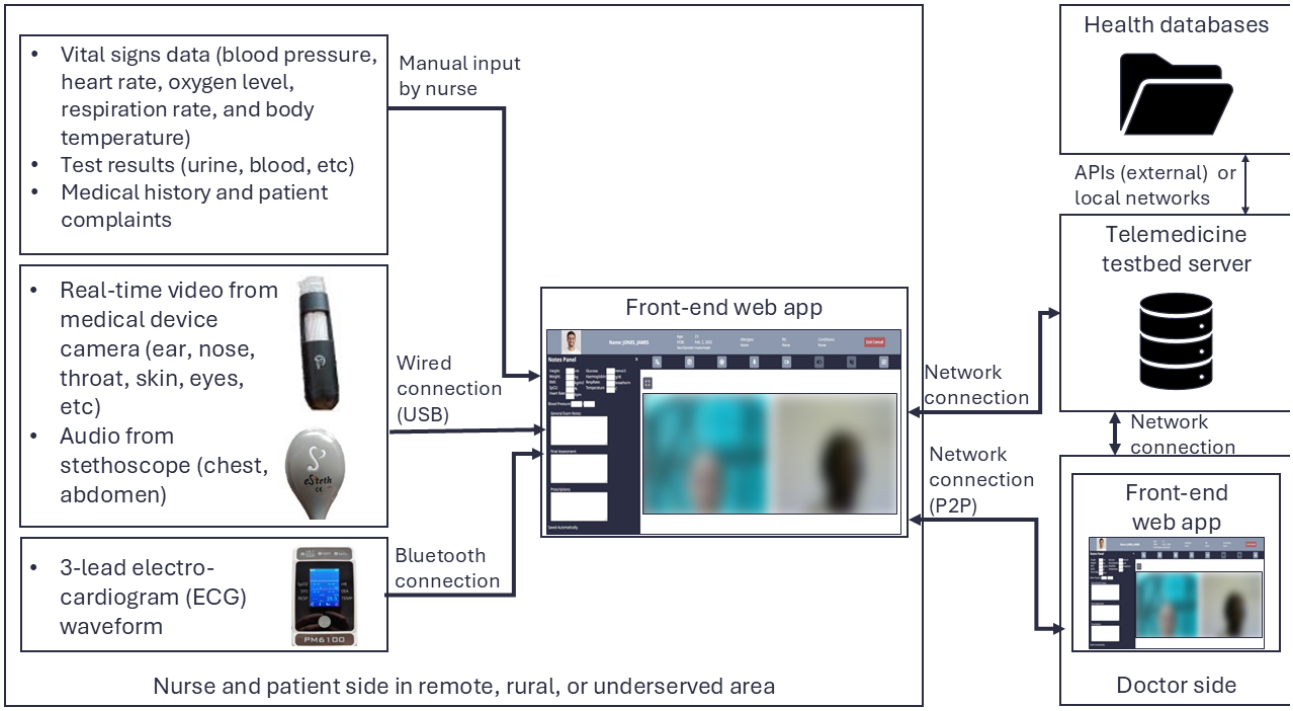
Virtual Clinic System

The developed virtual clinic system was used for virtual consultations during this study. The system consists of a web app with integrated medical devices, as described in Figure 1. It includes three main subsystems: (1) the consultation platform, which facilitates real-time communication and transmits real-time health data; (2) the electronic health record, which houses patient health data and consultation data from consultations hosted on the virtual clinic platform; and (3) a help desk to support user training and troubleshooting while using the system. Each of the subsystems was designed with user input and consideration of the contexts for which the system was intended.

There are three main medical devices that were integrated with the web app: (1) a digital stethoscope, which collects real-time auscultation sounds; (2) a medical camera, which provides general-purpose high-quality magnified video of the patient and has an otoscope adapter to provide ear, nose, and throat video; and finally (3) a vital signs monitor that can transmit 3-lead electrocardiogram readings via Bluetooth. The medical camera, which provides magnified medical images, is distinct from the webcam, which provides a real-time video of the nurse and patient interactions to the doctor. Images, recordings, and notes are captured during virtual consultations and stored in the electronic health record subsystem, so that this information can be referenced in future consultations. The devices used in this study can be substituted with similar devices. This allows clinics the ability to interchange devices depending on their budget, medical needs, and market availability in the region of implementation. For example, if the digital stethoscope used in the implementation in South Africa is not sold in other countries, implementers could replace it with a similar device, which may also be cheaper. This is in contrast to many existing telemedicine systems, in which medical devices and hardware are linked and cannot be interchanged. This design improves feasibility in the target contexts of remote, rural, and underserved areas.

XSL•FO
RenderX

Figure 1. Virtual clinic platform architecture.



A help desk is also integrated into the consultation platform. The help desk contains extensive training material for users, including video demonstrations and written guides on how to use the medical devices and software. There are also 4 case studies developed as training material, which users can go through as independent practice with the system. This setup allows users to become comfortable with the tools on their own time, rather than assuming that training time will be provided during implementation.

Study Sites and Participant Recruitment

Two study sites were used for virtual clinic implementation: Mitchell's Plain Community Health Centre in Cape Town and the Occupational Health Clinic at the Klerefontein Base for the South African Radio Astronomy Observatory (SARAO) Square Kilometre Array (SKA) project in Northern Cape Province. These sites are hereafter referred to as sites 1 and 2.

Site 1 is a public health facility in Mitchell's Plain township in Cape Town. It is estimated that the clinic serves approximately 1.2-million patients, including Mitchell's Plain and surrounding areas, with approximately 46,000 patients seen per month [36]. Despite being located in the urban center of Cape Town, there are still difficulties involved in getting doctors to the facility. Safety is a concern in Mitchell's Plain; the area has the highest reported assault and sexual assault rates in South Africa and is in the top 10 areas for robbery-related crimes [37]. This can make it difficult for doctors to get in and out of the facility. Therefore, virtual clinics offer a potential solution for remote clinical consultation without requiring doctors to physically visit the clinic. The study was conducted over 2 consecutive weekend days at site 1.

Site 2 is an occupational health clinic that serves approximately 200 employees at the SARAO SKA site. The nearest town is Carnarvon, which is approximately 16 km away. The clinic has an on-site nurse who provides day-to-day care. In addition, an occupational health nurse and doctor fly to the site once every 3 months for 1 day. During these visits, medical consultations are conducted for approximately 5 hours. The on-site nurse arranges for the patients who are in most need of the doctor to come to the clinic during these site visits. Alternatively, the workers may visit the community health clinic or the district hospital in Carnarvon, where there are an estimated 3 doctors who provide services between these sites. These facilities can also provide some trauma care in case of an emergency, although more serious cases of trauma must be transported to the tertiary hospital in De Aar, which is 200 km from Carnarvon. At site 2, the study was conducted over 1 day, during a site visit by the occupational health doctor.

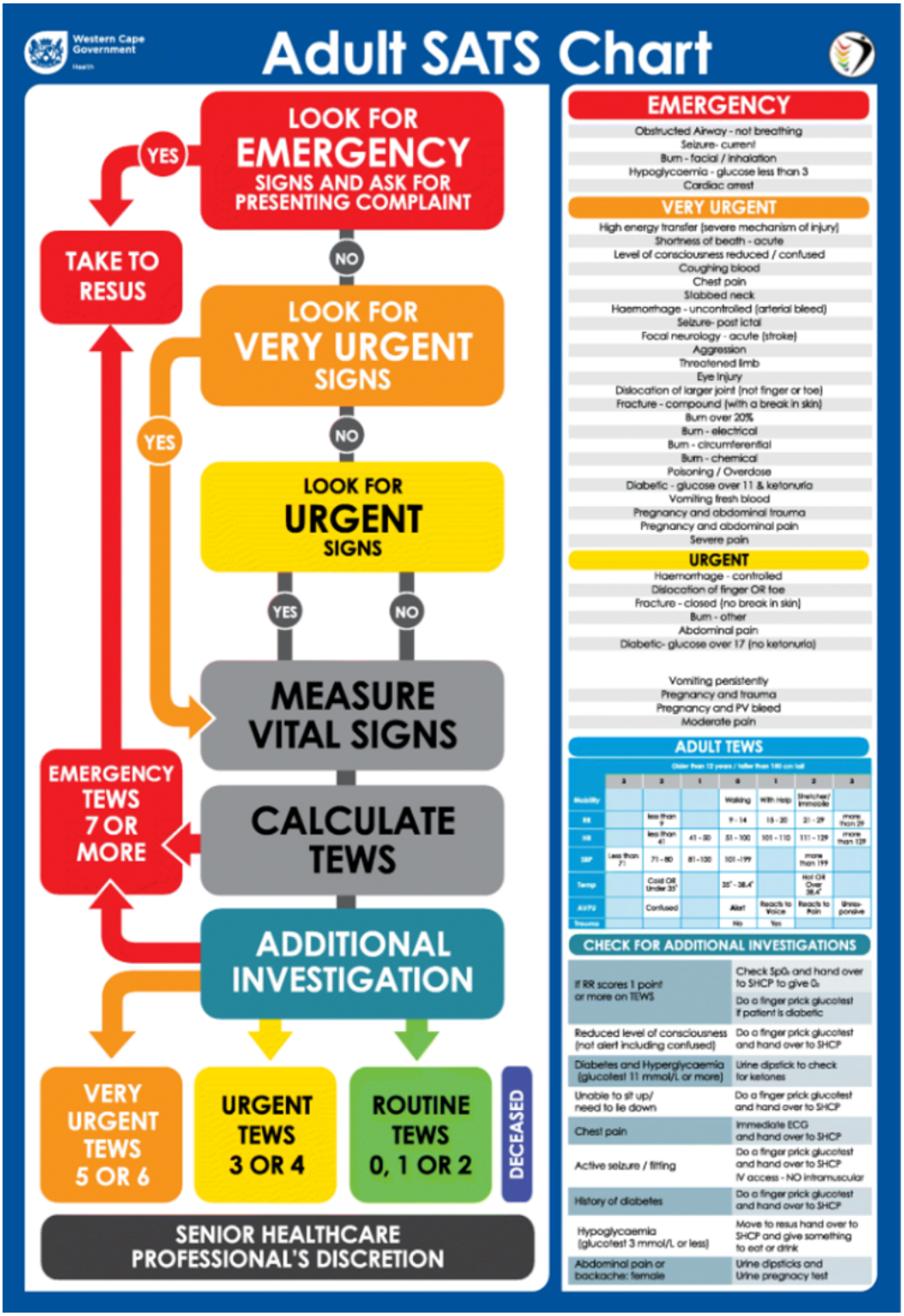
Patients were recruited from each clinic using a convenience sampling method. Inclusion criteria for patients were adults who were fluent in English. Patients were also required to be triaged as level green according to the South African Triaging Scale protocol, shown in Figure 2 [38]. Level green is the lowest risk patient, with a routine-reported Triage Early Warning Score of <3. At site 1, patients underwent triaging after entering the clinic facility. At site 2, the on-site nurse recruited patients who fit the inclusion criteria in advance of the study day. These patients were scheduled for consultation times on the study day. Exclusion criteria were patient cases in which the main complaint required a procedure, such as a catheter replacement, as these cases did not involve a diagnosis and therefore did not fit the study design.

Doctors and nurses were recruited using a snowball sampling method. Participants recruited for the doctor role were required to have a valid Bachelor of Medicine and Bachelor of Surgery qualification and a current registration with the Health Professions Council of South Africa. Participants recruited for the nurse role were required to have a National Qualifications Framework level 7 nursing degree or diploma and current

XSL•FO
RenderX

registration with the South African Nursing Council. The inclusion criteria for both doctors and nurses were that they should have no prior clinical experience using real-time virtual consultation systems.

Figure 2. The SATS flowchart for adults (adapted from the SATS training manual) [34]. SATS: South African Triaging Scale.



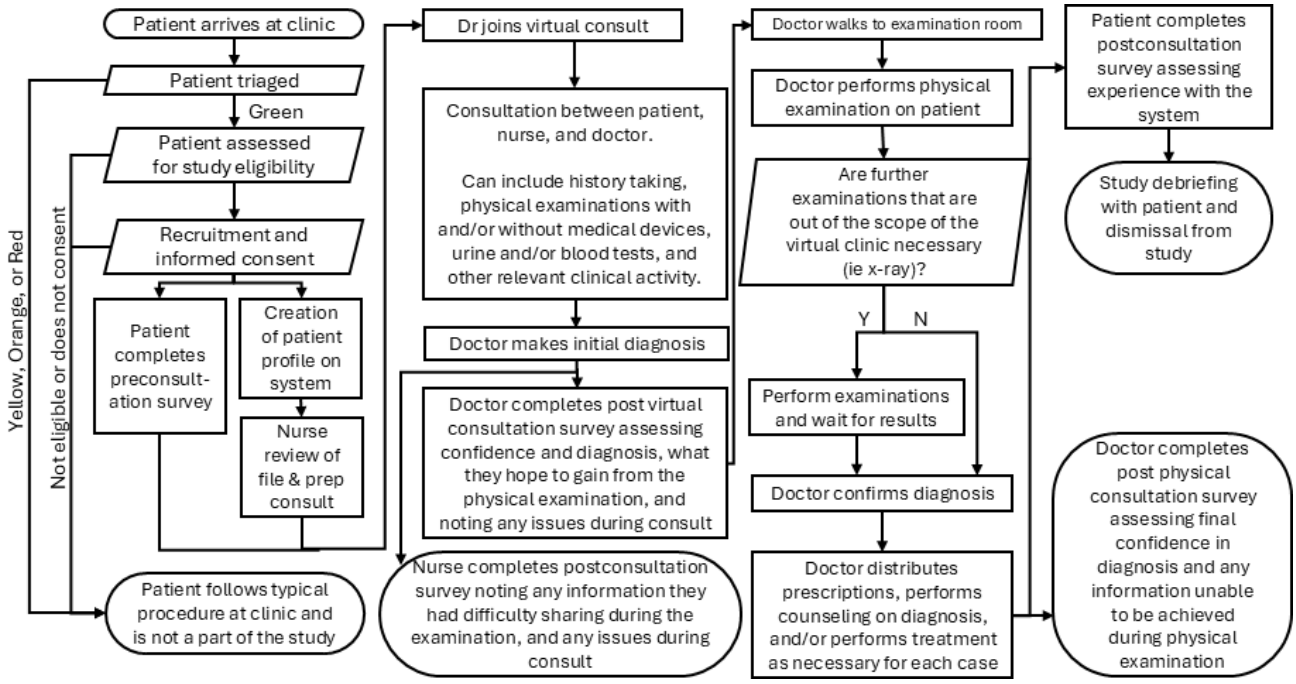
Study Design and Overview

To evaluate the developed virtual clinic system, a mixed methods, cross-sectional, pre-experimental pilot feasibility study was conducted. The virtual clinic system was implemented in 2 field clinic settings with patient, doctor, and nurse participants. A detailed depiction of the consultation study procedure is shown in Figure 3. The study commenced by training the doctor and nurse participants on the virtual clinic system. For doctors, the training took place the morning of their first study day for approximately 30 minutes. A dedicated training session was done for the nurses a day before the system evaluation at the clinic. This was because the nurses needed to familiarize

XSL•FO
RenderX

themselves with the medical devices, whereas the doctors only viewed the data collected from the devices, which did not require extensive training. Consultation sessions began in the clinic after the doctor and nurse completed their training on the system. During the system evaluation phase, patients participated in a virtual consultation facilitated by the nurse, with a doctor present virtually. The doctor was located in the same health facility as the nurse and the patient, but in a separate room, so they could not hear or see the examination other than through the virtual clinic system. The nurse then facilitated the conversation between the doctor and the patient, provided the medical history, and performed physical examinations as instructed by the doctor. After the virtual examination was finished, the doctor then entered the examination room and performed a physical examination to compare virtual and physical examination results. The study day concluded by conducting dyadic interviews with the doctor and the nurse about their experience with the system.

Figure 3. Study procedure used during the virtual clinic field pilot study.



Data Collection and Analysis

Both quantitative and qualitative data were collected during the study. A total of 5 different structured, self-report questionnaires were administered, which included 2 patient questionnaires, 2 doctor questionnaires, and 1 nurse questionnaire. A mixture of quantitative (scale rankings or binary) and qualitative (open answer) questions was included. The questionnaires are provided in Multimedia Appendix 1. Clinical notes were recorded during each section by the doctors and nurses in the virtual clinic platform. Simple observational data were collected by the researcher, which included the duration of each consult, examinations performed, and challenges encountered. Dyadic unstructured interviews were conducted with the nurse and the doctor at the conclusion of each study day to gather self-reported user feedback regarding the virtual clinic system.

Regarding research question 2 (What user behaviors are commonly present in successful virtual clinic consultations?), a successful virtual clinic consultation was defined as a consultation conducted using the virtual clinic where the doctor came to a diagnosis during the virtual consultation and reported a confidence level equal to that reported during the physical consultation. Regarding question 3, patient satisfaction was defined as a combination of “quality of care, similarity to face-to-face encounter, perception of the interaction” [39], willingness to use the service again, and willingness to recommend the service. Two standardized tools were incorporated within the patient and doctor questionnaires: the Telehealth Satisfaction Survey (TeSS) and a 5-point Likert scale diagnostic confidence ranking, respectively. TeSS was used to understand patient satisfaction during virtual consults. This questionnaire was developed by Yip et al [39] and has since been used to evaluate rural patients’ satisfaction with telemedicine in several studies [40,41]. To define successful consultations, a 5-point Likert scale was used to assess doctor confidence in virtual and physical diagnosis. This method has previously been used to rank confidence in diagnosis across many domains [42,43]. Although more detailed surveys have been developed for assessing diagnostic confidence, such as the Primary Care Confidence Scale [44], due to the short time span between virtual and physical consultations it was important to use a quick and simple questionnaire. Thus, the Likert scale confidence ranking was used.

The data collected were transcribed and cleaned as necessary to remove identifying information. Qualitative data were analyzed using an adapted practical thematic analysis method for health service research proposed by Saunders et al [45]. There are 3 steps of the method. The first is to read and annotate the data to understand the breadth and depth of the results. Then, the data are coded with descriptive titles that report detailed themes. After this, the low-level titles are synthesized to identify broad themes from the results. This results in 3 levels of

XSL•FO
RenderX

thematic codes. An iterative procedure, conducted by multiple reviewers, was taken to ensure data saturation. NVivo (Lumivero) was used to complete thematic analysis. Quantitative data from surveys are reported but not statistically analyzed due to the small sample size and nature of the study design [46].

Ethical Considerations

To maintain participant confidentiality, all data collected from participants were deidentified. To ensure that patient participants received the proper quality of care, patients were seen by a doctor both physically and virtually. Only the lowest risk patients were recruited for the study. Patients received a R100 grocery voucher as compensation for their time. The nurse and doctor participants were compensated at a standard hourly rate based on their qualifications for their time conducting consultations. There were no foreseen risks with participation in the study for any of the participant groups. The medical devices that were used for virtual examinations were all approved by the Federal Drug Administration and the Conformité Européenne for use in clinical diagnostics. This study was approved by the University of Cape Town Human Research Ethics Committee, reference number HREC322/2024. The relevant organizational bodies, including the Western Cape Department of Health (study number WC_202411_029), Mitchell's Plain Community Health Centre, and SARAO SKA management, also provided approval for the study.

Results

Cases and Consultations

Two doctor-nurse dyads participated in the virtual clinic visits, with 1 team per site. The doctor and nurse conducting visits at site 1 had not met before data collection. The doctor and nurse at site 2 were familiar with each other and had worked together previously. A total of 12 patients were seen during the study, with 9 at site 1 and 3 at site 2. This participant recruitment size is in line with similar research studies [47]. The average patient age was 42 years, with a range of 23 - 73 years. Female patients accounted for 66.6% of the population for both sites. An overview of the participants and cases is shown in Table 1. For patient participants, the presenting complaints are listed along with a brief description of the final assessment made by the doctor.

The full results of the thematic analysis are provided in Multimedia Appendix 2. The following 6 sections provide descriptions of each of the 6 themes identified and their supporting codes and evidence.

Table . Overview of patient participants and brief description of cases.

Site	Day	Age, years	Sex	Presenting complaint	Final assessment
1	1	73	Male	Pain in back and buttocks	Skin lesion
1	1	45	Female	Abdominal pain	Suspected TB[a]
1	1	44	Female	Headache and dizziness	Hypertension medication nonadherence
1	1	28	Female	Headache and cough	Flu
1	1	51	Male	Chest pain with cough	Pneumonia
1	2	43	Female	Back pain	Suspected TB
1	2	38	Male	Pain in nose and face	Gingivitis and abscessed teeth
1	2	23	Female	Abdominal pain and nausea	Bladder infection
1	2	55	Female	Body aches	Sciatica flare-up
2	3	30	Male	Raised bump on eye	Eye nodule + hypertension
2	3	32	Female	Abdominal pain	Gastritis and acid reflux
2	3	36	Female	Abdominal pain during certain activities	Potential bladder prolapse, referral to gynecology

[a]TB: tuberculosis.

Characteristics of Successful Virtual Clinic Consultations

First, experience was an enabler of confidence and success for new users. For both doctor-nurse dyads, confidence was reported to grow over time as more consultations were conducted. Doctor's confidence in diagnosis from virtual and physical examinations over time is reported in Figure 4, where it is demonstrated that confidence grew over time for the doctor at site 1. The confidence between consultations became equal starting at the end of day 1, and the pattern continued into day 2, demonstrating sustained confidence. In addition to doctor's confidence, the time spent verifying the outcomes of the virtual examination during the physical examination shortened over

XSL•FO
RenderX

time, as shown in Figure 5. This aligns with both doctors' questionnaire responses, which state that as consultations progressed, they were using only the physical examination to verify the findings they had made during the virtual examination.

Figure 4. Doctor-reported Likert-scale rankings of confidence in final assessment between virtual and physical consultations for each patient.

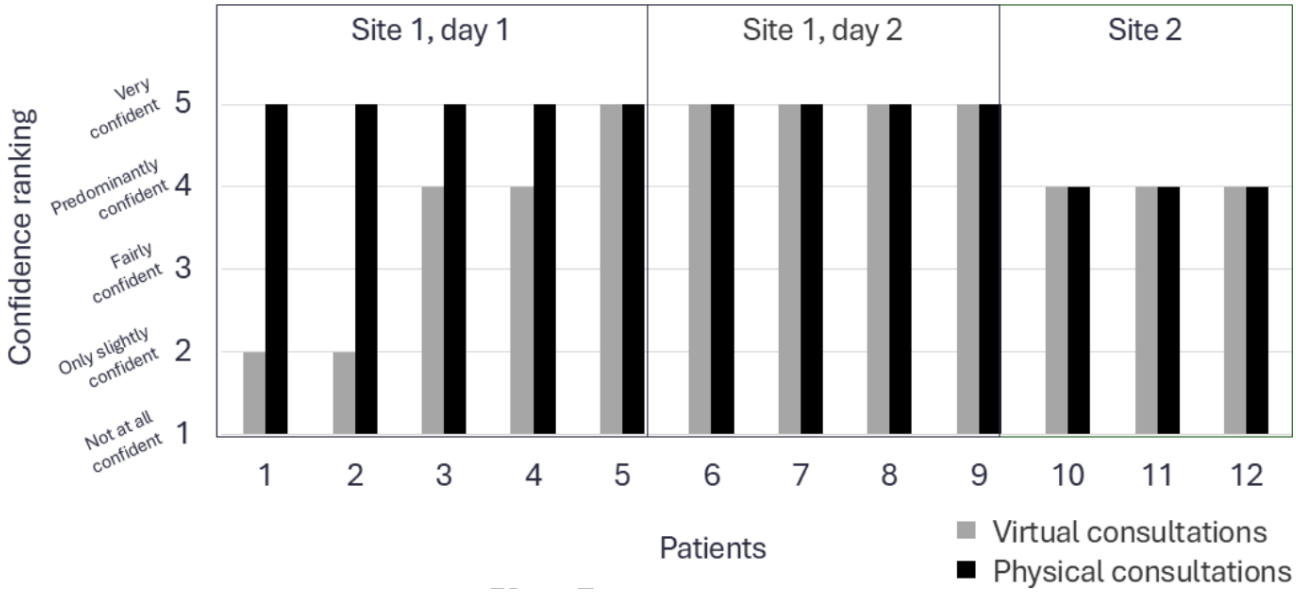
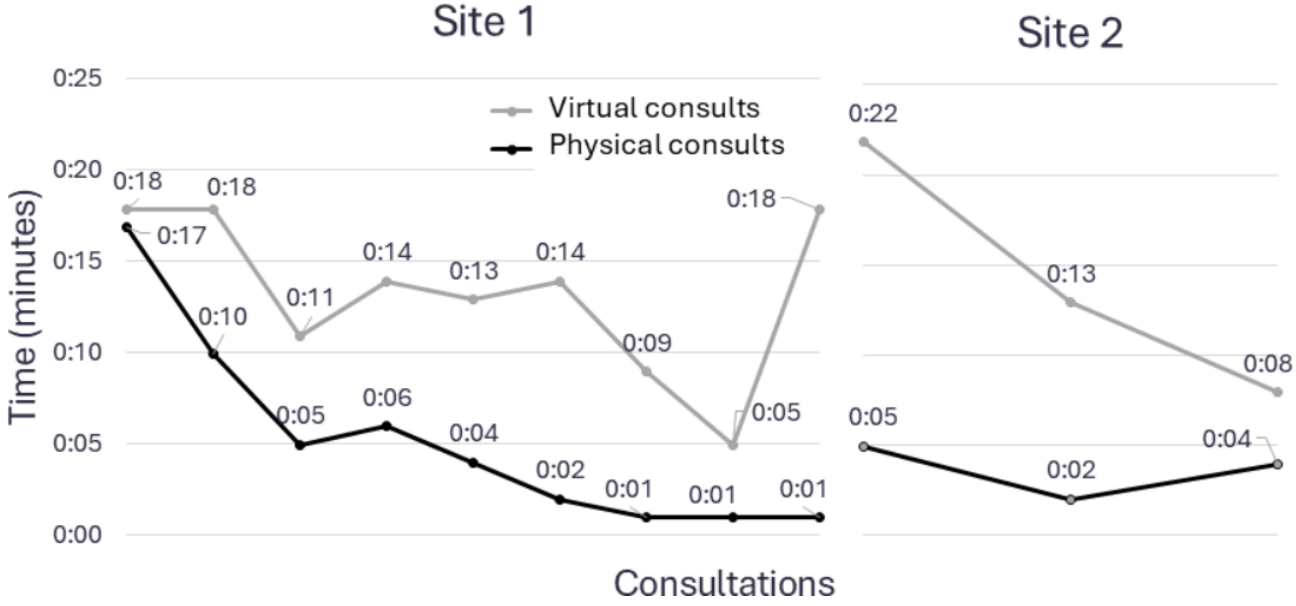


Figure 5. Duration of virtual and physical consultations for each site.



The nurse at site 1 had the following to say about her experience with the virtual clinic on the first day:

> *It was not as bad as I thought. I really thought it was going to be hard to collect the information, [and] the patients would be uncomfortable with not knowing exactly who they're speaking to. And to patients, what they also like is they like to be touched. So I was quite uncertain about it. And then, the more we got to do it, I got more comfortable with the system.*

Both doctors mentioned the importance of being able to exclude emergency diagnoses during virtual consults. During several consultations for both doctors, emergency diagnoses could be excluded through a combination of history taking and physical examinations conducted by the nurse. For the second case on day 1, the doctor noted that she was able to exclude signs of meningitis through the nurse's physical examination reports. The doctor at site 2 also reported that he was able to exclude emergency diagnoses for all 3 patients during the virtual consults.

The study design itself may have also contributed to building confidence in virtual diagnosis. Although not initially intended to do so, comparison between virtual and physical examinations allowed doctors to build trust in the technology and the nurse they were working with. The doctor at site 1 stated the following regarding her experience after the first day of using the system:

> *Initially I was very sceptical because we rely a lot on clinical examination to make a diagnosis. Especially as South African clinicians, because we don't always have special investigations, so to touch, to feel, to smell, yeah, it does make a difference. So I was*

XSL•FO
RenderX

> *sceptical. But then as we progressed, it got easier. And then, what [the nurse] found and what I found was more or less the same thing, when we examined.*

She also had the following to say regarding the exercise of comparing virtual and physical diagnoses:

> *Because like with [the patient] who needed an X-ray, which you need an in-house consultation, I'm kind of like would that actually have made a difference to my treatment in the end? Because based on what we found in the clinical examination, the X-ray just confirmed what we already decided, so I think it is beneficial just to rule out any really bad things. But, I think we did still get to the right diagnosis through the first virtual consultation.*

As consultations continued, confidence in virtual consults built among the doctors and nurses. The exercise of verification of diagnoses between virtual and physical formats is theorized to have contributed to this confidence-building, as it allowed for the establishment of trust between the nurse and the doctor. The ability to exclude emergency diagnoses is also theorized to have contributed to building confidence. The confidence demonstrated by the conclusion of the study days motivates for proof that successful virtual clinic consultations can be conducted in remote, rural, and underserved areas with the virtual clinic system.

Clear, descriptive communication between providers facilitated successful consults. An observed element of successful consultations was clear communication. This is discussed in the previous section with reference to both patients and providers, but provider-to-provider communication is also extremely important in the virtual setting. During consultations, both doctor and nurse pairs were able to conduct physical examinations over the virtual clinic platform, with the nurses providing detailed descriptions and the doctor directing the examination. Physical examinations on several areas of the body were conducted in this fashion, including the head, neck, abdomen, back, and buttocks. The doctor at site 1 noted the importance of communication in these examinations with the following:

> *I think you just you need to be very good at explaining things. It helps with [the nurse] because she already asks the questions and does the things, but if you don't have someone that's as experienced, you would just need to be able to relay that information. Like this is the particular examination that I want, [and] this is how you conduct that examination.*

These descriptions can also be important when complementing the information obtained from the medical devices. The nurse at site 1 noted during one of the examinations:

> *Unclear otoscopic images made it difficult for the doctor to see clearly, and therefore ask[ed] further investigations/assessment questions.*

Both doctors noted that they liked working with a provider-to-provider virtual clinic model rather than a patient-to-provider model. The doctor at site 2 noted that having a professional nurse on the ground who has medical experience is helpful when conducting the physical examinations and could also build trust and confidence in the patient receiving the service. The site 1 team corroborated this comment with the following:

> *Nurse: I'm just thinking if someone was at home and you don't have a nurse on the other end, it would be difficult.*
>
> *Doctor: It would be very difficult. Yeah. So what makes this flawless is having an experienced sister, because if it is a patient at home and I'm just gathering information by speaking, I can't touch them. There is no one there that can do it. I can't listen to the chest so I could see that guy who looks fine. Maybe just the green sputum is the only thing that would make you think infection, but it would make it difficult.*

Finally, the usefulness of this communication was highlighted by the nurse and the doctor at site 1. They stated that the ability to verify diagnoses and referrals between providers would be beneficial to building health care workers' experience and confidence working in remote, rural, and underserved areas. When asked whether this service would be helpful to the nurse at site 1 during her previous work in the Eastern Cape, she said:

> *Extremely. I remember when I was a nurse, the only thing we had was the group on WhatsApp with all of us, and then we helped each other. And that we would have one doctor that was nice enough to accept our patients and he would be like, OK, have you done this? Have you done that? And then eventually the doctor would be like, OK, you can refer. So yeah.*

The doctor at site 1 followed by stating:

> *And even for junior doctors [this would be helpful]. Yeah, because you are running a hospital on your own. And you don't have enough experience. And then you need to try and get to a specialist at the other hospital, who doesn't necessarily have time.*

Therefore, by providing a platform for establishing communication between providers, detailed accounts of physical examinations can be shared to improve health care delivery.

Patient Satisfaction With Virtual Care

Patient satisfaction was demonstrated when patients understood the benefits of virtual consults and were included in communication. Overall, patients lacked experience with virtual clinics but reported positive experiences after participating in a virtual consult. Results from preconsult and postconsult patient surveys are reported in Table 2. On the preconsult survey, only 1 patient participant answered yes to having previously used telemedicine services. On the postconsult surveys, patients indicated satisfaction with the system, with the majority of TeSS responses falling in the "good" or "excellent" categories. All patients reported that they would use the service again and would recommend the service to another person.

XSL•FO
RenderX

Table . Summary of the patient responses to pre- and postconsult surveys.

Survey item	Response					
	Yes, n (%)	No, n (%)	Excellent, n (%)	Good, n (%)	Poor, n (%)	Fair, n (%)
Preconsult patient survey: have you consulted a doctor or a nurse through phone or video before?	1 (8.3)	11 (91.7)	—[a]	—	—	—
Postconsult patient survey statement. How satisfied were you with:						
The voice quality of the equipment?	—	—	6 (50)	5 (41.7)	0 (0)	6 (50)
The visual quality of the equipment?	—	—	6 (50)	5 (41.7)	1 (8.3)	6 (50)
Your personal comfort in being seen using the virtual clinic system?	—	—	8 (66.7)	4 (33.3)	0 (0)	8 (66.7)
The length of time with the doctor you saw?[b]	—	—	5 (45.5)	5 (45.5)	1 (8.3)	0 (0)
The explanation of your treatment by the virtual clinic doctor?	—	—	9 (75)	3 (25)	0 (0)	0 (0)
The thoroughness, carefulness, and skillfulness of the virtual clinic doctor?	—	—	9 (75)	2 (16.7)	1 (8.3)	0 (0)
The courtesy, respect, sensitivity, and friendliness of the virtual clinic doctor?	—	—	10 (83.3)	2 (16.7)	0 (0)	0 (0)
How well the virtual clinic doctor respected your privacy?	—	—	7 (58.3)	4 (33.3)	1 (8.3)	0 (0)
How well the staff answered your questions about the equipment?[b]	—	—	5 (45.5)	5 (45.5)	1 (8.3)	0 (0)
Your overall treatment experience using the virtual clinic?	—	—	7 (58.3)	5 (41.7)	0 (0)	0 (0)
Would you recommend the virtual clinic service to another person?	12 (100)	0 (0)	—	—	—	—
Would you use the virtual clinic service again?	12 (100)	0 (0)	—	—	—	—

[a]Not applicable.

[b]Questions that were not answered by 1 respondent each.

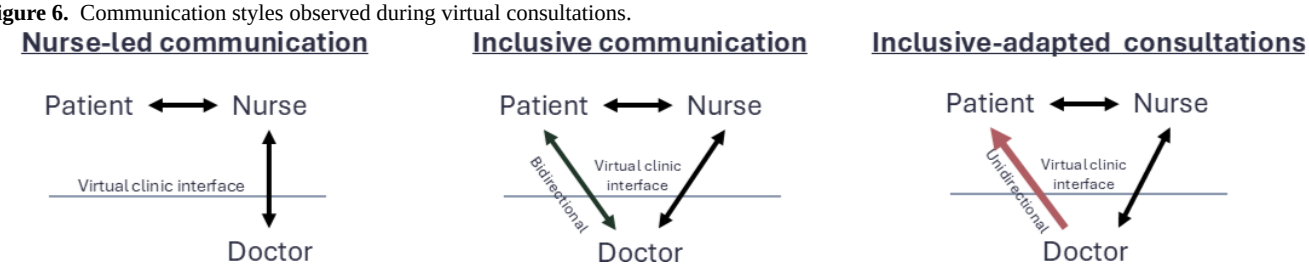
XSL•FO
RenderX

The nurse at site 1 had the following to say about one of the patient's demonstrated behaviors during a virtual consultation:

> *At the beginning, there was the lady that came, the one who just had the headache. She was a bit unsure at the beginning. She was very edgy, but then once she realized that, actually, I'm going to see a doctor, then I think she started relaxing.*

An observation during consultations was that the model of communication between the doctor, nurse, and patient was not always consistent. The observed models of communication are demonstrated in Figure 6. The optimal mode of communication was when the patient, nurse, and doctor all spoke directly to one another, labeled as inclusive communication. This also involved using a video view where both the patient and the nurse were included. This was important for the doctor to see nonverbal communications, such as body language or shaking of the head "yes" or "no" to answer questions. In one case, the doctor was able to determine that the patient was nasally congested by speaking directly to the patient and hearing their voice. However, in some cases, inclusive communication was not used. Particularly, these were in earlier consultations for site 1, but in some cases, this style of communication was also not feasible. Two other models of communication were thus observed, and these were labeled nurse-led and inclusive-adapted. In nurse-led communication, the nurse fully facilitated communication between the doctor and the patient. This was necessary during consultation 1 on the first day, when the patient was hard of hearing and the doctor's comments had to be repeated to him by the nurse. In inclusive-adapted communication, the doctor spoke directly to the patient, but the patient would respond through the nurse. Whether the patient intended to speak to the doctor directly was difficult to assess; in some cases, the patient was soft-spoken, and in other cases, their body language indicated that they were speaking to the nurse. In both the nurse-led and inclusive-adapted models, there was increased repetition of information that could have contributed to longer consultation durations. However, these styles of communication may be more comfortable for certain patients and thus should not be discounted. Rather, nurses and doctors should be informed about these styles of communication and be directed to make a conscientious choice that balances patient comfort and communication efficiency.

Figure 6. Communication styles observed during virtual consultations.



Additionally, patients who participated in the study may have received higher-quality care than they would have received in the traditional clinic setting. This was mentioned as particularly true for the green-triaged patients at site 1, who often face extremely long wait times as the day progresses, when more emergency cases take priority. In fact, 2 patients were recruited for the study on day 1, approximately 1 hour after the clinic opened. Neither patient met the inclusion criteria and was thus not included in the study. At the conclusion of the study day on day 1, approximately 5 hours after the nonincluded patients had entered the clinic, these patients were still in the queue and had not yet been seen. The long wait times are corroborated by the doctor and the nurse at site 1, who have previous experience working in other public clinics across the country.

> *Doctor: [The virtual clinic] actually makes a big difference, because we have a lot of clinics [that] don't have doctors and then that patient gets referred to a facility with a doctor. And it takes them weeks or months to get a date to see the doctor.*
>
> *Nurse: It's actually hard to get a date, especially in the rural areas. I'm not looking just at the Western Cape, but thinking Eastern Cape.*

While the quality-of-care impact is stark for the patients recruited in this study because they were green, or low priority, patients, it is recognized that this may not be the case for all virtual clinic implementations. However, what can be acknowledged is that patients were satisfied to receive care from a real doctor and nurse, with a physical examination component, and that these elements could be provided by the virtual clinic service.

Barriers to Virtual Clinic Care Delivery

Barriers caused by infrastructure and the implementation environment were able to be addressed adequately during the study. They did, however, threaten the success of the virtual consultations. Therefore, proper consideration should be given to these barriers to ensure that they have been addressed adequately.

On day 1 of the site 1 implementation, several technical issues were encountered. As this was the first implementation of the virtual clinic in a real-world setting, this was not surprising. The major challenge was establishing a quality network connection. The doctor was connected to the virtual clinic app using a tablet, which was receiving internet connection from a mobile phone hotspot that was connected to a local mobile network provider. The nurse's computer was connected directly to the mobile network, using a USB SIM modem. Throughout day 1, there were intermittent disconnections along with video and audio lag, which disrupted virtual consults. The mobile phone hotspot was suspected to be contributing to the poor connection, and various SIM cards, phones, and settings were adjusted throughout the day to try to improve this. It was also attempted to use USB tethering to hotspot the mobile network, and with this, the network quality improved. However, this

XSL•FO
RenderX

caused the battery of the tablet to drain quickly, and the tablet died, interrupting a virtual consultation. With these challenges, virtual consults could still take place, but the duration and confidence associated with these consults were low. These challenges were highlighted in a written response by the doctor referencing a consult with network issues:

> *Video lag. Network issues. Online consultation tends to take long due to lag issues (network).*

The nurse also noted the potential risks with loss of network connection during one of the consultations:

> *Only 1 difficulty was encountered, network glitching. Resulted in the doctor repeating suggestion/questions, or possibly missing some information from my side to the doctor or from the doctor to my side.*

There were several actions taken between day 1 and day 2 at site 1 to improve the quality of the virtual clinic consultations, and these were carried over into the site 2 study day. The laptop webcam was substituted by a high-quality HD1080p camera, and a USB SIM modem was used to provide mobile network to a laptop on the doctor's side, instead of a tablet. The pair of headphones that the doctor had been using was adjusted to provide better sound quality. These changes had a drastic impact on the success of virtual consultations. There were no observed or mentioned network challenges on day 2 at site 1, despite using the same network providers and the same location. This indicates that the network challenges experienced on the first day were due to the hardware used and not due to the service provider.

Some contextual factors of the clinic environment caused brief challenges in communication, particularly due to the background noise. At site 1, there were intercom announcements that interrupted communication, and at site 2, there were construction noises and a heater which caused background noise. However, the doctor at site 2 noted that the background noise was present but not disturbing when using the headphones.

Another environmental factor that affected virtual consultations was the arrangement of the examination space. At site 1, the examination bed had to be positioned approximately 4 meters from the laptop and camera, while at site 2, the examination bed could be positioned only 1 meter away. The arrangement at site 1 caused challenges during the demonstration of physical examinations that were not present during site 2 consultations. The nurse at site 1 reported:

> *Since the laptop is situated on the desk and the bed is further away, when doing an abdominal exam, it would be nice if the video camera would be closer so the doctor can see what areas are being palpated and what the abdomen looks like, direct where to palpate, and what kind of palpations/assessments to do.*

It is important to note that if these environmental factors are not considered, it threatens the ability of the system to streamline existing clinical workflows. The doctor at site 1 mentions this, saying:

> *So if I was in a busy setting, kind of having to communicate via another person to speak to the patient, there is time that that takes, whereas I wouldn't even have sat and asked as many questions as I usually do. So you would need to have a flawless system where there's no lags in between all of the [consults]. Yeah, it would not be as relaxed as it is now. Because if I had to see 30 patients in this time, [or] five who took so long?*

As such, pilot studies such as this are extremely important in identifying these concerns and mitigating them through future design and recommendations. Another barrier encountered during virtual clinic consultations was that medical device data were not always suitable for conveying certain types of health information. During consultations, the medical devices were used with the virtual clinic platform to provide real-time health data. An overview of how the medical devices were used for each consult is provided in Table 3. The vital signs monitor was not included in the table, as no consult required electrocardiogram usage. During initial consults at each site, the vital signs monitor was used for vital signs examination. However, at site 1, the triaging report data were used from consult 3 onward to capture patients' vital signs. At site 2, the nurse preferred to use the clinic's devices to take measurements, potentially due to the familiarity that they provided.

XSL•FO
RenderX

Table . Medical device usage during virtual consults.

Site	Day	Medical device usage (yes/no) and description of usage	
		Medical camera	Stethoscope
1	1	No	No
1	1	Yes: pupil reactivity to light	No
1	1	Yes: pupil reactivity to light, inside ear, and throat	No
1	1	Yes: inside throat	Yes: chest examination
1	1	No	Yes: chest examination
1	2	No	Yes: chest examination
1	2	Yes: inside mouth and throat	No
1	2	No	No
1	2	No	No
2	3	Yes: eyelid examination	Yes: chest examination
2	3	No	Yes: abdominal examination
2	3	No	No

The medical devices were not always able to properly convey what the nurse saw or heard on the ground. The stethoscope was reported to provide variable sound quality. When comparing the performance of the digital stethoscope with her own stethoscope, the doctor at site 1 noted:

> *Digital stethoscope note: audio very loud can over-amplify sound which could lead to overinterpretation of pathology. Low frequencies are very loud, high frequencies not audible.*

For the first patient at site 2, the doctor could not hear the breath or heart sounds from the digital stethoscope, as can be seen from the following note:

> *Stethoscope sound quality was not sufficient to assess heart sounds.*

However, the sounds could be heard for the subsequent 2 patients at site 2, and the doctor reported high sound quality for these patients. The doctor at site 2 determined that this could be due to the faintness of the breath and heart sounds in the first patient, which the nurse could hear but he could not. Therefore, while the stethoscope demonstrated satisfactory performance for some patients, this was not consistent between patients. Furthermore, the examination camera was more useful in sharing dermatological images rather than otoscopic images. The nurse at site 1 provided the following comment:

> *During this consultation it was a bit challenging to demonstrate clearly using the otoscope, to show the mouth clearly. For the light (otoscope) I tried removing the black head and focuser but the light wasn't able to clearly show the focused part of the mouth/throat and show the gingivitis/swelling and infection in the gums and teeth. After removing the black head for the clear head and focuser, it still wasn't easy to see the mouth/gums/teeth. It seems the colour of the light (light blue light) doesn't work well, we tried using the light from the ward's otoscope and it was more yellow/orange and the viewing/imaging was much better in that light.*

So, to alleviate the challenges from the medical camera, the device was used in combination with the light provided from the clinic otoscope. In addition, it was mentioned several times that the medical camera could not show a clear view of the tympanic membrane in the ear. This was a similar comment received during usability testing, which motivated a redesign of the otoscope to allow for a narrower tip to navigate the ear canal. However, this redesign was not able to change the design of the device itself, which may need to be adjusted to improve this capability.

Although the devices used were Federal Drug Administration–and Conformité Européenne–approved, they did not end up providing the quality necessary during all examinations. The evidence provided supports user-guided feedback on how the devices may be improved. These conclusions also motivate further development of open medical devices that are not tied to proprietary software, as the market options for these types of devices are limited. Development and improvement of these devices for use in systems such as the developed virtual clinic will be advantageous to providing better quality virtual care.

Finally, there were limitations of the virtual clinic system in providing primary health-level services. While most consults conducted using the virtual clinic were successful, there were some limitations of the system that should be mentioned. Some challenges simply cannot be addressed, given the current technical capabilities of virtual consultation. For example, the nurse at site 1 highlights one piece of information she struggled to share with the doctor during a consult:

> *The warm of the patient's hands, that had to be described extensively.*

At the current state, it would not be feasible to add a device that can convey this information with certainty; however, there could be additional training or research in description or examination

XSL•FO
RenderX

that could assist with conveying the necessary information to the doctor. Ultimately, it was not within the scope of this study to determine what cases are not appropriate for virtual care and which are best suited, but this is recommended for further research.

Discussion

Principal Findings

The results from this study inform recommendations for digital health systems for remote, rural, and underserved areas. These recommendations are summarized in Table 4. There were still some areas for improvement in the system design, as demonstrated by clinical implementation feedback, but most of the user feedback centered around implementation feedback rather than system design feedback. This transition demonstrates how UCD-based iteration proved effective in taking user feedback to develop an improved virtual clinic system that addressed user needs [28].

Table . Summary of virtual clinic recommendations.

Result	Supporting evidence	Recommendation
Experience is an enabler of confidence and success for new users.	• As consults progressed, doctors reported less reliance on the follow-up physical examinations to diagnose. • Physical consult time reduced as consults progressed. • Doctors' confidence in virtual diagnosis increased as consults progressed.	• Nurse and doctor users should participate in training and hands-on practice with the system.
Patients' satisfaction is demonstrated when patients understand the benefits of virtual consults and are included in communication.	• Patients reported high satisfaction with the virtual clinic system postconsultation. • Wait times were reduced for patient participants in comparison with other patients in the clinics. • Three communication styles emerged, which had different effects on patient involvement during the virtual consult.	• Nurse and doctor users should be educated on the different communication styles for virtual consults and apply them based on the patient context. • Patients should be educated on the potential positive impacts of virtual consults to encourage adoption.
Clear, descriptive communication between providers facilitates successful consults.	• Nurses used verbal descriptions to convey the patient's condition to the doctor during virtual consultations. • Doctors could confidently diagnose patients based on verbal description supplemented with medical device data.	• Nurse and doctor users should be trained together on using descriptive verbal communication during virtual consults. • Nurse and doctor users should practice virtual and physical consults together to build trust in one another's skills.
Barriers caused by infrastructure and the implementation environment can be addressed adequately but must not be ignored; otherwise, success could be threatened.	• Changing the network access hardware for the virtual clinic app correlated with better quality of service and an increase in doctors' confidence in virtual diagnosis. • Nurse and doctor users commented positively on the higher-quality camera implemented on the second day of study.	• High-quality periphery devices (cameras, headphones, and network access devices) should be used with the virtual clinic app to enable virtual consult success. • The physical clinic setup should support virtual consults, including positioning the examination bed and the patient near the camera and microphone.
Medical device data may not be suitable for conveying certain types of health information.	• The stethoscope device did not convey soft or shallow breath and heart sounds adequately for the remote doctor to hear. • The medical camera could not show mouth, teeth, and ear images adequately for the remote doctor to see.	• Existing medical devices should be improved based on the user's feedback provided in this study. • There should be more open medical devices for open software apps, such as the virtual clinic. • Nurses should be trained on how to supplement medical device data with verbal descriptions.
There are limitations of the virtual clinic system in providing primary health-level services.	• The virtual clinic system could not convey data regarding the warmth of a patient's hands, which the nurse deemed important for diagnosis.	• The constraints of the virtual clinic system should be further defined based on additional studies. • The virtual clinic system should consider the health data requirements of the implementation context and the feasibility (cost and market availability) of the technology used to convey these health data.

XSL•FO
RenderX

The user feedback provided during clinical implementation aligns with the Capabilities, Opportunities, Motivation, and Behavior (COM-B) framework for behavioral change [48]. COM-B states that users need the capacity to engage, a conducive sociophysical environment, and the drive to perform a behavior. In this context, capability refers to the user's ability to interact with the system, which is supported by help desk materials and exposure to the system. Opportunities can involve the infrastructure used to support the implementation. Motivation included the intrinsic desire of the health care worker to provide proper care to their patients [49], which was enabled by the virtual clinic system. This result has been reported in the development of other digital health solutions that used UCD principles [50,51]. Doctor and nurse participants highlighted the need for a solution in the public health care system and expressed a greater tendency for behavior change after they felt more capable of using the system. Working in teams was an enabler of success for both users, as trust was built and effective communication styles were established. This aligned with similar findings of virtual clinic implementations [52]. The threats of the physical environment and technology that were experienced, such as the network issues on day 1 of implementation, did have a negative effect on the provider's experience with the system, aligning with the opportunities aspect of COM-B and the consensus in literature that technological functionality has an effect on provider telemedicine satisfaction [53]. Additionally, there was a demonstrated improvement from the previous results of usability testing [32], where users indicated that they would like more training with the system. This aligns with the capability aspect of the COM-B framework, as users were provided an opportunity to build their confidence with the virtual clinic system before consulting with patients [54]. This included hardware and software components. The help desk subsystem, including the hands-on training that was incorporated, provided for this aspect of motivation. This design also supported potential threats identified, such as a lack of generational digital literacy [55]. Motivation was still clearly existent, as the providers also understood the needs of the health care system and the benefits that the virtual clinic system could provide. Therefore, when the technological challenges were addressed, all factors of the COM-B framework were addressed, and participants had a greater likelihood of wanting to implement the virtual clinic system. At site 2, these challenges had been addressed, and therefore quicker uptake of the virtual clinic system was demonstrated.

The results of this evaluation serve as a proof of concept that the system can work in real-world environments and motivate for further research and implementation based on these recommendations. Not only does this information provide insights into the virtual clinic, but it also provides insights into the landscape and readiness of South Africa for digital health implementations. The information, such as patient satisfaction with virtual clinic services, provider opinions toward virtual clinic consultations, quality of service of various African and South African networks, and hardware considerations for better design, motivates for implementation of virtual clinic and digital health solutions in these areas.

Conclusions

The results of this study support the acceptability and feasibility of the developed virtual clinic system in a real-world setting. Furthermore, the results of this study provide evidence of virtual clinics in remote, rural, and underserved areas of South Africa. This research highlighted that provider experience and confidence with using the digital health system was an enabler of successful virtual clinic consultations. Additionally, clear communication between the health care worker in the remote, rural, or underserved area and the remotely located doctor resulted in quicker consultations. Patient feedback emphasized that when providers directly communicate with patients during virtual consultations, and when the patient understands the benefits of using virtual clinic consultations, patients have improved experience. Finally, barriers that are discussed in similar implementations, including limited infrastructure [11-13] and contextual challenges [14,15], were also observed during this study. However, these contextual challenges could be overcome with appropriate design methodologies such as UCD. This included selection of medical devices that align with the needs of the context and the limitations of the environment.

The current evidence of interventions in remote, rural, and underserved areas is limited; therefore, these results contribute not only to the development of this system but also to the development of other digital health solutions. The nature of the feedback and observations provides insights into the motivation and behaviors of doctors, nurses, and patients using virtual clinic services in low-resource South African contexts. The recommendations from the Results section provide a framework of how other digital health developers might consider implementation in these contexts, aligning with the monitoring and evaluation phase of the World Health Organization Telemedicine Implementation Guidelines [16].

There were some methodological and contextual limitations of this study. There was a lack of additional reviewers and coders used during thematic analysis. While this is important to reduce bias, other measures, such as reporting direct quotes and performing iterative thematic analysis, were introduced to reduce bias. Considering the patient population recruited, there may have been bias in patient satisfaction with the system, particularly at site 1. This is because patients who were recruited were green-triaged patients, who are the lowest priority patients at the clinic. Often, these patients face the longest wait times as the other, higher priority patients take precedence over them. Some green-triaged patients do not see a doctor at all when entering the clinic, particularly on a weekend when clinic staff is limited. They may also be limited to seeing a nurse only, and not a doctor at the clinic, depending on their complaint. Their participation in the study allowed them to see a doctor almost immediately after being triaged at the clinic. As such, they may have indicated satisfaction due to the speed and quality of care they were provided by their participation in the study. Given this, further validation of patient satisfaction results during long-term virtual clinic implementation is needed. Finally, it is suggested that additional studies incorporate a direct comparison between diagnosis during virtual and diagnosis during physical examinations. This allows for a better comparison between the 2 interventions.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank the clinical staff and facilities who supported this research, including Dr Roland Kroukamp at Mitchell's Plain Community Health Centre and Mark Johnson at South African Radio Astronomy Observatory Square Kilometre Array. Generative AI was not used in the creation of this manuscript.

Funding

This document has been produced with the financial assistance of the European Union and cofunding from the Carnegie, Foundation of New York (grant DCIPANAF/2020/420 - 028), through the African Research Initiative for Scientific Excellence (ARISE), pilot program. ARISE is implemented by the African Academy of Sciences with support from the European Commission and the African Union Commission (EUC). The statements made and views expressed are solely the responsibility of the authors. The research is also supported by Telkom South Africa, via the Telkom Centre of Excellence in Broadband Networks at the University of Cape Town.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

ABl contributed to conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, visualization, and writing – original draft. JM participated in supervision and writing – review & editing. MID participated in supervision, methodology, and writing – review & editing. ABi participated in supervision and writing – review & editing. BM contributed to conceptualization, project administration, funding acquisition, resources, and writing – review & editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study questionnaires.
[DOCX File, 17 KB - humanfactors_v13i1e84272_app1.docx]

Multimedia Appendix 2
Thematic analysis results.
[DOCX File, 22 KB - humanfactors_v13i1e84272_app2.docx]

References

1. Vas V, North S, Rua T, et al. Delivering outpatient virtual clinics during the COVID-19 pandemic: early evaluation of clinicians' experiences. BMJ Open Qual 2022 Jan;11(1):e001313. [doi: 10.1136/bmjoq-2020-001313] [Medline: 34996810]
2. Gray LC, Edirippulige S, Smith AC, et al. Telehealth for nursing homes: the utilization of specialist services for residential care. J Telemed Telecare 2012 Apr;18(3):142-146. [doi: 10.1258/jtt.2012.SFT105]
3. Hobson EV, Baird WO, Partridge R, et al. The TiM system: developing a novel telehealth service to improve access to specialist care in motor neurone disease using user-centered design. Amyotroph Lateral Scler Frontotemporal Degener 2018 Aug;19(5-6):351-361. [doi: 10.1080/21678421.2018.1440408] [Medline: 29451026]
4. Ren X, Zhai Y, Song X, Wang Z, Dou D, Li Y. The application of mobile telehealth system to facilitate patient information presentation and case discussion. Telemed J E Health 2020 Jun;26(6):725-733. [doi: 10.1089/tmj.2020.0084] [Medline: 32298208]
5. Atasoy H, Greenwood BN, McCullough JS. The digitization of patient care: a review of the effects of electronic health records on health care quality and utilization. Annu Rev Public Health 2019 Apr 1;40:487-500. [doi: 10.1146/annurev-publhealth-040218-044206] [Medline: 30566385]
6. Fairchild RM, Ferng-Kuo SF, Laws S, Rahmouni H, Hardesty D. Telehealth decreases rural emergency department wait times for behavioral health patients in a group of critical access hospitals. Telemed J E Health 2019 Dec;25(12):1154-1164. [doi: 10.1089/tmj.2018.0227] [Medline: 30735100]
7. Raza T, Joshi M, Schapira RM, Agha Z. Pulmonary telemedicine--a model to access the subspecialist services in underserved rural areas. Int J Med Inform 2009 Jan;78(1):53-59. [doi: 10.1016/j.ijmedinf.2008.07.010] [Medline: 18809352]
8. Nord G, Rising KL, Band RA, Carr BG, Hollander JE. On-demand synchronous audio video telemedicine visits are cost effective. Am J Emerg Med 2019 May;37(5):890-894. [doi: 10.1016/j.ajem.2018.08.017] [Medline: 30100333]

XSL•FO
RenderX

9. Sayani S, Muzammil M, Saleh K, Muqeet A, Zaidi F, Shaikh T. Addressing cost and time barriers in chronic disease management through telemedicine: an exploratory research in select low- and middle-income countries. Ther Adv Chronic Dis 2019 Jan;10. [doi: 10.1177/2040622319891587]
10. Bajgain B, Rabi S, Ahmed S, Kiryanova V, Fairie P, Santana MJ. Patient-reported experiences and outcomes of virtual care during COVID-19: a systematic review. J Patient Rep Outcomes 2023 Dec 1;7(1):126. [doi: 10.1186/s41687-023-00659-8] [Medline: 38038800]
11. Global fiber development index: 2020, a global index comparing fiber development on a country-by-country basis. Omdia. 2020. URL: https://worldbroadbandassociation.com/wp-content/uploads/2021/08/FDI-White-Paper-Final_151020.pdf [accessed 2025-07-06]
12. Chiaraviglio L, Blefari-Melazzi N, Liu W, et al. Bringing 5G into rural and low-income areas: is it feasible? IEEE Comm Stand Mag 2017;1(3):50-57. [doi: 10.1109/MCOMSTD.2017.1700023]
13. Casati P, Moner-Girona M, Khaleel SI, Szabo S, Nhamo G. Clean energy access as an enabler for social development: A multidimensional analysis for Sub-Saharan Africa. Energy Sustainable Dev 2023 Feb;72:114-126. [doi: 10.1016/j.esd.2022.12.003]
14. Bodie GD, Dutta MJ. Understanding health literacy for strategic health marketing: eHealth literacy, health disparities, and the digital divide. Health Mark Q 2008 Jul 2;25(1-2):175-203. [doi: 10.1080/07359680802126301]
15. Scheil-Adlung X. Global evidence on inequities in rural health protection: new data on rural deficits in health coverage for 174 countries. International Labour Organization. 2015. URL: https://www.ilo.org/secsoc/information-resources/publications-and-tools/Workingpapers/WCMS_383890/lang--en/index.htm [accessed 2025-07-08]
16. Consolidated telemedicine implementation guide. : World Health Organization; 2022 URL: https://iris.who.int/bitstream/handle/10665/364221/9789240059184-eng.pdf?sequence=1 [accessed 2026-08-18]
17. Dodoo JE, Al-Samarraie H, Alsswey A. The development of telemedicine programs in Sub-Saharan Africa: progress and associated challenges. Health Technol 2022 Jan;12(1):33-46. [doi: 10.1007/s12553-021-00626-7]
18. Ogundaini OO, Mlitwa NBW. A situational review of national Digital Health strategy implementation in sub-Saharan Africa. J Transdiscipl Res S Afr 2024;20(1):a1476. [doi: 10.4102/td.v20i1.1476]
19. Agbeyangi AO, Lukose JM. Telemedicine adoption and prospects in Sub-Sahara Africa: a systematic review with a focus on South Africa, Kenya, and Nigeria. Health Care (Don Mills) 2025;13(7):762. [doi: 10.3390/healthcare13070762]
20. World Health Organization. National eHealth Strategy Toolkit: World Health Organization; 2012. URL: https://iris.who.int/bitstream/handle/10665/75211/9789241548465_eng.pdf?sequence=1&isAllowed=y [accessed 2025-08-11]
21. Klaassen B, van Beijnum BJF, Hermens HJ. Usability in telemedicine systems—a literature survey. Int J Med Inform 2016 Sep;93:57-69. [doi: 10.1016/j.ijmedinf.2016.06.004] [Medline: 27435948]
22. Hearn J, Wali S, Birungi P, et al. A digital self-care intervention for Ugandan patients with heart failure and their clinicians: user-centred design and usability study. Digit Health 2022;8:20552076221129064. [doi: 10.1177/20552076221129064] [Medline: 36185389]
23. Dominguez-Rodriguez A, De La Rosa-Gómez A. A perspective on how user-centered design could improve the impact of self-applied psychological interventions in low- or middle-income countries in Latin America. Front Digit Health 2022;4:866155. [doi: 10.3389/fdgth.2022.866155] [Medline: 35721795]
24. Bekele R, Groher I, Sametinger J, et al. User-centered design in developing countries: a case study of a sustainable intercultural healthcare platform in Ethiopia. Presented at: 2019 IEEE/ACM Symposium on Software Engineering in Africa (SEiA); May 28, 2019; Montreal, QC, Canada p. 11-15. [doi: 10.1109/SEiA.2019.00010]
25. Sobrinho A, da Silva LD, Perkusich A, Pinheiro ME, Cunha P. Design and evaluation of a mobile application to assist the self-monitoring of the chronic kidney disease in developing countries. BMC Med Inform Decis Mak 2018 Dec;18(1). [doi: 10.1186/s12911-018-0587-9]
26. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Encyclopedia of Human-Computer Interaction: Sage Publications; 2004, Vol. 37:445-456.
27. US Department of Health and Human Services. The Research-Based Web Design & Usability Guidelines: US Government Printing Office; 2006. URL: https://www.usability.gov/sites/default/files/documents/guidelines_book.pdf [accessed 2025-07-08]
28. Smaradottir B, Gerdes M, Martinez S, Fensli R. The EU-project United4Health: user-centred design of an information system for a Norwegian telemedicine service. J Telemed Telecare 2016 Oct;22(7):422-429. [doi: 10.1177/1357633X15615048] [Medline: 26541347]
29. Calvillo-Arbizu J, Roa-Romero LM, Estudillo-Valderrama MA, et al. User-centred design for developing e-Health system for renal patients at home (AppNephro). Int J Med Inform 2019 May;125:47-54. [doi: 10.1016/j.ijmedinf.2019.02.007] [Medline: 30914180]
30. Verma N, Lehmann H, Alam AA, Yazdi Y, Acharya S. Development of a digital assistant to support teleconsultations between remote physicians and frontline health workers in India: user-centered design approach. JMIR Hum Factors 2023 Feb 2;10:e25361. [doi: 10.2196/25361] [Medline: 36729578]
31. Sanches LMP, Harris MR, Abbott PA, Novaes MA, Lopes M. Collaborative software development for a Brazilian telehealth program. Stud Health Technol Inform 2014;201:211-218. [Medline: 24943546]

XSL•FO
RenderX

32. Blocker AB, Datay MI, Mwangama J, Malila B. Development of a telemedicine virtual clinic system for remote, rural, and underserved areas using user-centered design methods. Digit Health 2024 Jan;10. [doi: 10.1177/20552076241256752]
33. Ekeland AG, Bowes A, Flottorp S. Methodologies for assessing telemedicine: a systematic review of reviews. Int J Med Inform 2012 Jan;81(1):1-11. [doi: 10.1016/j.ijmedinf.2011.10.009] [Medline: 22104370]
34. Maeder AJ, Wilson LS. Chapter 25: evaluating telehealth interventions. In: Handbook of eHealth Evaluation: An Evidence-Based Approach: University of Victoria; 2017.
35. Farag J. Cross-cultural telemedicine: sub-Saharan Africa. Intersect Stanford J Sci Technol Soc 2015;9(1) [FREE Full text]
36. About us: Mitchell's Plain Community Health Centre. Western Cape Department of Health. 2023. URL: https://d7.westerncape.gov.za/mitchellsplain?toc_page=1 [accessed 2025-07-08]
37. Crime statistics: crime situation in republic of South Africa twelve (12) months (April to March 2019-20). : South African Police Service URL: https://www.saps.gov.za/services/april_to_march_2019_20_presentation.pdf [accessed 2025-07-08]
38. Western Cape Department of Health. The South African Triage Scale (SATS): training manual 2012. : Western Cape Government; 2012 URL: https://emssa.org.za/wp-content/uploads/2011/04/SATS-Manual-A5-LR-spreads.pdf [accessed 2025-07-08]
39. Yip MP, Chang AM, Chan J, MacKenzie AE. Development of the Telemedicine Satisfaction Questionnaire to evaluate patient satisfaction with telemedicine: a preliminary study. J Telemed Telecare 2003;9(1):46-50. [doi: 10.1258/135763303321159693] [Medline: 12641893]
40. Moreno L, Peck JL. Nurse practitioner–led telehealth to improve outpatient pediatric tracheostomy management in South Texas. J Pediatr Health Care 2020;34(3):246-255. [doi: 10.1016/j.pedhc.2019.11.008] [Medline: 32059818]
41. Talarico I. The use of telehealth to increase mental health services access and promote medication adherence in rural locations. J Am Assoc Nurse Pract 2021;33(11):1074-1079. [doi: 10.1097/JXX.0000000000000495]
42. Kits A, De Luca F, Kolloch J, et al. One-minute multi-contrast echo planar brain MRI in ischemic stroke: a retrospective observational study of diagnostic performance. J Magn Reson Imaging 2021 Oct;54(4):1088-1095. [doi: 10.1002/jmri.27641] [Medline: 33942426]
43. Lawonn K, Luz M, Preim B, Hansen C. Illustrative visualization of vascular models for static 2D representations. In: Navab N, Hornegger J, Wells W, Frangi A, editors. Medical Image Computing and Computer-Assisted Intervention: Springer, Cham; 2015:9350. [doi: 10.1007/978-3-319-24571-3_48]
44. Shavit R, Kushnir T, Nudelman Y, Weisman A, Springer S. Confidence in providing primary care to patients with low back pain among physiotherapists. BMC Prim Care 2025 May 6;26(1):147. [doi: 10.1186/s12875-025-02850-w] [Medline: 40329185]
45. Saunders CH, Sierpe A, von Plessen C, et al. Practical thematic analysis: a guide for multidisciplinary health services research teams engaging in qualitative analysis. BMJ 2023;381:e074256. [doi: 10.1136/bmj-2022-074256]
46. Teresi JA, Yu X, Stewart AL, Hays RD. Guidelines for designing and evaluating feasibility pilot studies. Med Care 2022 Jan 1;60(1):95-103. [doi: 10.1097/MLR.0000000000001664] [Medline: 34812790]
47. Babaei N, Zamanzadeh V, Valizadeh L, Lotfi M, Avazeh M. The implementation requirements of a virtual care program for patients with chronic wounds: a qualitative study. Int Wound J 2025 Jul;22(7):e70719. [doi: 10.1111/iwj.70719] [Medline: 40603262]
48. West R, Michie S. A brief introduction to the COM-b model of behaviour and the PRIME theory of motivation. Qeios. Preprint posted online on Apr 9, 2020. [doi: 10.32388/WW04E6.2]
49. Berdud M, Cabasés JM, Nieto J. Incentives and intrinsic motivation in healthcare. Gac Sanit 2016;30(6):408-414. [doi: 10.1016/j.gaceta.2016.04.013] [Medline: 27321370]
50. Ikwunne T, Hederman L, Wall PJ. DECENT: a sociotechnical approach for developing mobile health apps in underserved settings. Digit Health 2023;9:20552076231203595. [doi: 10.1177/20552076231203595] [Medline: 37786402]
51. Szinay D, Perski O, Jones A, Chadborn T, Brown J, Naughton F. Perceptions of factors influencing engagement with health and well-being apps in the United Kingdom: qualitative interview study. JMIR Mhealth Uhealth 2021 Dec 16;9(12):e29098. [doi: 10.2196/29098] [Medline: 34927597]
52. Tennankore K, Jones J, Miller A, et al. Transforming healthcare delivery: a descriptive study of a novel provider-to-provider virtual care platform. Front Public Health 2023;11:1284566. [doi: 10.3389/fpubh.2023.1284566] [Medline: 38155888]
53. Totten AM, Womack DM, Griffin JC, et al. Telehealth-guided provider-to-provider communication to improve rural health: a systematic review. J Telemed Telecare 2024 Sep;30(8):1209-1229. [doi: 10.1177/1357633X221139892] [Medline: 36567431]
54. Seston EM, Willis SC, Fenton C, et al. Implementation of behaviour change training in practice amongst pharmacy professionals in primary care settings: Analysis using the COM-B model. Res Social Adm Pharm 2023 Aug;19(8):1184-1192. [doi: 10.1016/j.sapharm.2023.04.123] [Medline: 37121797]
55. Papp-Zipernovszky O, Horváth MD, Schulz PJ, Csabai M. Generation gaps in digital health literacy and their impact on health information seeking behavior and health empowerment in Hungary. Front Public Health 2021;9:635943. [doi: 10.3389/fpubh.2021.635943] [Medline: 34055714]

XSL•FO
RenderX

Abbreviations

COM-B: Capabilities, Opportunities, Motivation, and Behavior
SARAO: South African Radio Astronomy Observatory
SKA: Square Kilometre Array
TeSS: Telehealth Satisfaction Survey
UCD: user-centered design

Edited by A Kushniruk; submitted 17.Sep.2025; peer-reviewed by NABILA HARAHAP; revised version received 28.Apr.2026; accepted 22.Jul.2026; published 01.Sep.2026.

Please cite as:
Blocker A, Biyabani A, Mwangama J, Datay MI, Malila B
User-Centered Virtual Clinic Services for Remote, Rural, and Underserved Sub-Saharan Africa: Development and Usability Study
JMIR Hum Factors 2026;13:e84272
URL: https://humanfactors.jmir.org/2026/1/e84272
doi:10.2196/84272

© Abby Blocker, Ahmed Biyabani, Joyce Mwangama, Mohammed Ishaaq Datay, Bessie Malila. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 1.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mental Health Monitoring in Esports Athletes Using an mHealth App: Requirements Analysis and Usability Study

Leona Stolberg[1], BA; Daniel Hieber[1,2,3], MSc; Peter Kuhn[1], PhD; Michael Hebel[4], MA; Johannes Schobel[1,4], PhD

[1]DigiHealth Institute, Neu-Ulm University of Applied Sciences, Wileystraße 1, Neu-Ulm, Bavaria, Germany

[2]Institute of Medical Data Science, University Hospital Würzburg, Würzburg, Bavaria, Germany

[3]Institute of Neuropathology, Ulm University Hospital, Ulm, Baden-Wuerttemberg, Germany

[4]HNU Game Institute, Neu-Ulm University of Applied Sciences, Neu-Ulm, Bavaria, Germany

Corresponding Author:
Johannes Schobel, PhD
DigiHealth Institute, Neu-Ulm University of Applied Sciences, Wileystraße 1, Neu-Ulm, Bavaria, Germany

Abstract

Background: Esports athletes face substantial psychological stressors comparable to those experienced by traditional athletes, yet mental health tools tailored to their specific needs remain scarce.

Objective: This study addresses this critical gap by introducing MindAthlete, a prototype for a user-centered mobile health (mHealth) app intended to enhance mental health monitoring and support among esports athletes.

Methods: A 3-stage development process was used over 5 months. First, structured qualitative interviews were conducted online via Google Meet with 7 experienced (sports) psychologists recruited through snowball sampling within professional esports networks. The interviews explored current mental health monitoring practices and key feature requirements for MindAthlete. Data were analyzed using Mayring's approach to qualitative content analysis. Subsequently, a 32-screen high-fidelity prototype was developed in Figma based on the identified requirements. In the final stage, usability testing was conducted with 25 semiprofessional and professional esports athletes recruited via convenience and snowball sampling. Participants interacted with the Figma prototype and completed an online questionnaire administered via LimeSurvey, which included demographic items, the validated 10-item System Usability Scale (SUS) rated on a 5-point Likert scale, and open-ended questions on navigation, aesthetics, and overall functionality. Inferential analyses included Mann-Whitney *U* tests, Kruskal-Wallis tests, and Spearman correlations. Open-ended responses were analyzed using Mayring's content analysis.

Results: Expert interviews identified 8 core features for mental health monitoring in esports, clustered around 3 themes: continuous self-monitoring, structured assessment, and intervention and insight. Athlete usability testing yielded a median SUS score of 77.5 (IQR 67.5-87.5), classifying usability as "good" and exceeding a benchmark median of 68.3 derived from comparable mHealth apps. No significant differences were found by sex, age, years of professional experience, or prior mHealth usage. Qualitative feedback highlighted strengths including intuitive navigation and appealing aesthetics, alongside areas for improvement such as shorter animations and clearer initial onboarding. However, findings must be interpreted in light of the sample's demographic homogeneity, which limits generalizability to the broader esports population.

Conclusions: This research provides a stakeholder-driven, empirically validated blueprint for esports-specific mental health monitoring via mHealth. Future work will focus on technical implementation, integration of wearable sensor data, and broader evaluations with larger, more diverse samples to confirm generalizability and enhance usability.

(JMIR Hum Factors 2026;13:e90140) doi:10.2196/90140

KEYWORDS

mental health; mobile health; mHealth; usability evaluation; requirements analysis; esports; gamification; user-centered design

Introduction

Esports is defined as playing video games regulated through ranking systems and official leagues via electronic devices, encompassing a spectrum from amateur players to elite professionals [1]. While esports participation broadly covers various levels, this work focuses specifically on competitive esports athletes operating at semiprofessional and professional tiers. As participation and competitiveness have grown, understanding the health profile of esports players has become a central focus. Recent systematic and scoping reviews indicate that esports participants do not inherently exhibit poorer baseline mental health than nonparticipants [2], although high-volume participation is associated with specific physical, lifestyle, and psychological vulnerabilities [3,4]. Crucially, psychological states and health risks differ across competitive tiers.

XSL•FO
RenderX

Nonprofessional and amateur players primarily face lifestyle-related risks, such as sedentary behavior, elevated body fat, and gaming disorder tendencies [2]. In contrast, semiprofessional and professional athletes, who operate under structured team environments, contractual expectations, and tight competition schedules, face intense domain-specific stressors, including performance expectations, team conflicts, and demanding tournament schedules [4-7], which predict a high prevalence of mental health challenges [8,9]. Among competitive players, these multidimensional demands frequently result in elevated stress, anxiety, depression, burnout, sleep disturbances, and night-eating tendencies [10-13]. These conditions highlight an ethical imperative to address mental health concerns proactively in esports, as untreated issues can compromise immediate performance and long-term psychological well-being [7,14]. However, the esports industry still lacks sufficient mental health support systems [15]. While players report positive attitudes toward psychological support [16], hiring dedicated staff is costly (especially for smaller organizations), and continuous manual monitoring is time-consuming. Mobile health (mHealth) apps may offer a scalable alternative to streamline screening and ongoing monitoring [17-19].

Although mHealth tools have demonstrated efficacy in symptom tracking for broad clinical conditions such as stress, depression, and posttraumatic stress disorder [20-23], they target general psychopathology rather than performance-related distress and contextual triggers unique to competitive athletes. Furthermore, such generic tools frequently suffer from poor usability and low long-term engagement [24,25]. In sports specifically, mHealth apps have demonstrated feasibility for mood tracking, automated risk detection, and psychological skills training, though tool quality and evidence bases vary widely [17,26]. Blended models combining in-person and digital components have been proposed to enhance athlete accessibility and real-time clinical decision-making [27,28]. However, these sport-specific tools are typically embedded within colocated, traditional training environments, and their design assumptions around physical load, coaching structures, and recovery processes often do not align with the screen-bound workflows of esports team structures, limiting their transferability. The esports ecosystem itself has seen a rise in performance-enhancement tools lacking scientific validation [1], and players may also engage with generic mindfulness apps or game-embedded focus aids. However, these tools are designed to support general relaxation or productivity rather than to screen for or monitor mental health risk, offer no structured pathway to practitioner oversight, and do not capture the specific performance, organizational, and social stressors that distinguish this population [5,7], limiting their value as tools for proactive mental health support.

Given these limitations, we developed MindAthlete—a novel, noncommercial mHealth prototype designed specifically for mental health screening and monitoring in esports athletes. It is shaped through direct consultation with practicing esports psychologists and esports players, ensuring that each design decision reflects real-world clinical needs rather than assumed requirements.

To guide this development process, 2 research questions were formulated:

Research question 1: What features are essential for mental health monitoring in esports athletes according to (sports) psychologists?

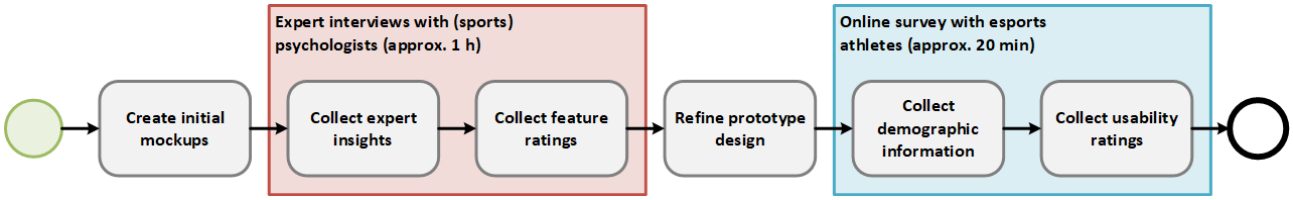
Research question 2: How do esports athletes rate the usability of the prototype, and what do they identify as strengths, areas for improvement, and additional feature needs?

Methods

Study Design

In this study, we followed a 3-stage development process (Figure 1). This process was guided by the design thinking methodology, which emphasizes iterative, user-centered development and encourages early stakeholder involvement to ensure the solution fits real-world needs [29]. While user-centered design typically involves end users at every stage, our approach used a phased stakeholder engagement strategy. Given that MindAthlete is designed for a dual-user ecosystem (athletes for data entry and specialized practitioners for clinical monitoring), we prioritized the "expert-user" during the initial define and ideate phases. This choice was driven by the clinical necessity to ensure that the app features were scientifically valid and actionable for practitioners, who serve as the primary stakeholders for data interpretation and intervention. The first stage involved systematically gathering functional and nonfunctional requirements through structured interviews with domain experts. In the second stage, we created a design prototype to visualize the user interface and intended workflow. In the final stage, we conducted usability testing with semiprofessional and professional esports athletes to empirically evaluate MindAthlete's usability and collect targeted feedback for further refinement.

Figure 1. Overall study design. Red: expert interviews with (sports) psychologists; blue: online survey with esports athletes.



XSL•FO
RenderX

Expert Interviews

The recruitment phase for the expert interviews followed a snowball sampling approach, as the population of practitioners specialized in esports psychology and mental health is extremely small globally. This scarcity is reflected in the broader industry reality, where dedicated sports psychology support remains absent across much of the professional landscape [15]. Leveraging initial contacts within a professional network via email, additional suitable experts were identified through recommendations, ensuring that participants had highly relevant and specialized expertise. Snowball sampling was particularly appropriate given the niche nature of the esports psychology field. Inclusion criteria required participants to hold an academic degree in psychology or sports psychology and to have at least 3 years of professional experience. Additionally, eligible participants had to be actively involved in the esports industry as practitioners, either working with athletes competing in higher divisions of the European Regional Leagues or providing guidance to professional esports players in international top-tier competitions. Finally, participants had to be proficient in German or English. These criteria were selected to ensure that only experts with a solid academic background and direct practical experience in supporting esports athletes' mental health contributed to the study.

After the experts gave their informed consent, an online meeting was scheduled. The interviews followed a predefined interview guide to ensure a standardized procedure (Multimedia Appendix 1) and were conducted by the first author (LS), who possessed domain expertise in esports player performance and formal training in qualitative research methodologies. At the beginning of the interview, experts were asked about the necessity and current practices of mental health monitoring in esports. Then, they were introduced to mHealth apps and asked about the required features for MindAthlete. Eight potential features were preselected based on a literature review of mHealth apps for mental health and the evaluation of leading apps in this domain. This helped to identify core functionalities commonly recommended or implemented for psychological self-monitoring. These features were presented to the experts via screen sharing and explained in detail. Experts were then invited to provide feedback regarding the potential benefit of these features. Additionally, they could suggest additional features, which were then verbally refined in close collaboration. Experts were then asked to rank both the preselected and their own suggested features based on perceived importance and usefulness. While all experts ranked the 8 preselected features, the total number of ranked items varied because experts could also rank their own suggested features. However, the scale always remained fixed at 8 points for the highest-ranked item down to 1 point for the eighth. The 8 highest-ranked features across all interviews were selected for implementation. A point-based system was introduced, assigning 8 points to the highest-ranked feature, 7 to the second-ranked feature, and continuing in this manner until the eighth-ranked feature, which received 1 point. In cases of equal point totals, features were ranked by evaluating their specificity to esports mental health vs broader health contexts and prioritizing the feature with more immediate relevance to professional esports. Finally, experts were asked about their preferences regarding the user interface, platform accessibility, and data privacy concerns.

All interviews were recorded and automatically transcribed using the aTrain software [30] and manually checked for accuracy. Data analysis involved both deductive and inductive content analysis, following Mayring's approach [31]. Two researchers were involved in the qualitative analysis. The first author (LS) conducted the initial coding. Deductive categories were derived from the interview guide and covered the perceived importance of mental health tracking, past and present monitoring methods, their sufficiency, comprehensive metrics for mental health assessment, feedback on each of the 8 literature-based candidate features, further expert-suggested features, interface preferences including colors and platform, data security concerns, feature ratings, and reasons for those ratings. The feature-related deductive categories were each coded along two dimensions: expert opinions on the feature and specific suggestions or requirements for its implementation. To capture unanticipated themes, inductive categories were derived directly from the transcripts following Mayring's inductive category formation procedure. The criterion of selection was defined to include any statement addressing contextual factors, organizational conditions, or operational requirements not covered by the predefined feature matrix. Text segments meeting this criterion were open-coded, grouped by thematic similarity through iterative reading, and subsumed into broader categories (cooperation with support staff, the perceived purpose of the application, user engagement strategies, and challenges in application adoption). A senior supervisor (JS) reviewed the full code structure to ensure consistency and resolve discrepancies, providing feedback on category definitions and relevance (Multimedia Appendix 1). Following the coding process, categories related to feature ratings and expert rationale were used to calculate the point-based feature ranking system, while interface and contextual categories informed nonfunctional design decisions for the prototype.

Design Prototype Development

The insights and user requirements identified through expert interviews were systematically aggregated, evaluated for coherence, and directly integrated into a high-fidelity design prototype using the prototyping software Figma [32]. A total of 32 screens were designed, representing the primary user flow from onboarding to feature interaction, including feedback elements and animations. The 8 highest-ranked features, as determined by experts, were implemented. Decisions to include, modify, or exclude certain feedback were guided by their alignment with expert consensus and evidence, and their coherence with established best practices identified in the literature. Additional feedback was retained if (1) the suggestion was voiced by at least 4 of the 7 experts, a threshold chosen to reflect clear majority consensus, or (2) the expert provided a concrete scientific or clinical rationale. Feedback that did not meet these criteria was archived for future iterations.

Low-fidelity mockups were iteratively refined into a high-fidelity prototype by the main author (LS) and the supervisor (JS), directly informed by the synthesis of the expert interviews. Choosing a high-fidelity prototype at this stage was

XSL•FO
RenderX

justified by the need to closely simulate real user interactions, enabling an accurate assessment of usability. All screens and interaction flows were designed following user-centered design principles, drawing on Nielsen's usability heuristics [33], Norman's design principles [34], and empirical guidelines from health usability research [35], to ensure consistent layouts, clear navigation, and immediate feedback.

Applying these principles, the resulting prototype comprises the key features depicted in Figure 2. The Home Screen (Figure 2A) provides an overview of sessions, daily tips, and gamification status, with a sidebar for navigation. The Mood Tracker (Figure 2B) uses a slider and context tags to log current or daily mood. The Dashboard (Figure 2C) presents interactive graphs that users can filter by year, month, or week. Sport psychological questionnaires (Figure 2D) also include a weekly check-in for lifestyle data. The selection of specific validated instruments is ongoing and will be finalized in the next development phase. Instruments of interest are the World Health Organization-Five Well-Being Index (WHO-5 [well-being]), Generalized Anxiety Disorder-7 (GAD-7 [anxiety]), Patient Health Questionnaire-9 (PHQ-9 [depression]), Perceived Stress Scale (PSS [stress]), or Athlete Psychological Strain Questionnaire (APSQ [athlete-specific psychological strain]). Grounded in self-determination theory [36], gamification mechanics were incorporated to enhance engagement: each completed survey or tracked entry triggers a sky lantern animation (Figure 2E) that floats upward to symbolize progress and award points, ranks, and trophies, directly mirroring the reward systems native to competitive esports environments. This prototype formed the basis for subsequent usability testing.

Figure 2. Screenshots of the Figma prototype. (A) Home Screen displaying the next session with the (sports) psychologist, daily tips, and gamification elements. (B) Mood Tracker allowing players to select their mood and add descriptive words. (C) Dashboard presenting results from the mood and stress trackers as well as questionnaires. (D) Example of a sport psychological questionnaire. (E) Sky Lantern animation representing the gamification theme and the metaphorical sense of letting go.



Athlete Usability Testing

Participants were recruited through a combination of convenience and snowball sampling. Leveraging personal and professional contacts within the esports industry, initial invitations were distributed directly to esports athletes, teams, and affiliated organizations. Additional invitations were sent via professional esports networks to maximize outreach across different teams and regions using email and Discord, a primary communication platform within esports communities. Inclusion criteria required that participants be at least 18 years old and proficient in English, and actively competing in esports at a semiprofessional or professional level. Specifically, professional status was defined as athletes competing in top-tier continental leagues under full-time contracts, while semiprofessional status included athletes in regulated national or regional leagues (eg, Division 1 or 2). The participants in the latter category also operated under professional conditions, characterized by fixed salaries, official team affiliations, and commitment to training and competition. All participants provided their informed consent before participating.

Participants were invited to test the MindAthlete design prototype and complete an online questionnaire administered through LimeSurvey [37]. Usability was assessed using the validated 10-item SUS questionnaire, rated on a 5-point Likert scale [38]. The SUS was chosen due to its proven reliability and widespread acceptance in usability research. A target sample size of 20 to 30 participants was deemed sufficient based on recommendations for usability testing with the SUS [39]. Additionally, open-ended questions were included to solicit detailed feedback on navigation, aesthetics, and overall functionality of the prototype. The questionnaire also collected demographic and contextual information, including age, sex, nationality, esports experience (total and paid years), primary income status, and prior mHealth usage. All items were mandatory for completion, and the athletes were able to change their answers through a back button.

Data analysis consisted of three steps. First, descriptive statistics were computed to summarize the data. To determine the appropriate measures of central tendency, normality was assessed using the Shapiro-Wilk test, and outliers were examined using the IQR method. Second, inferential analyses were conducted to examine prespecified group differences and

XSL•FO
RenderX

relationships. Specifically, SUS scores were compared across sex groups (female, male, and other) and age groups (18-21, 22-25, 26-29, and 30-33 y) using the Mann-Whitney *U* test and Kruskal-Wallis test, respectively. Associations between SUS scores and continuous variables (years of professional experience and paid years of experience) were assessed using Spearman correlations. Additionally, group comparisons were performed for main income status (esports as primary income: yes/no), prior mHealth usage (yes/no), and nationality. A Bonferroni correction was applied to control the family-wise error rate across all inferential tests. Each obtained *P* value was multiplied by the total number of comparisons performed (k=7) and compared against the nominal significance threshold of α=0.05. Finally, open-ended responses were analyzed qualitatively using Mayring's content analysis approach [31] by the primary coder (LS) and reviewed by (JS), following the 2-tiered deductive-inductive procedure described previously. Top-level deductive domains were anchored directly to the 3 open-ended survey prompts: positive feedback, areas for improvement, and suggestions for additional features. Within each deductive domain, inductive subcategories were derived directly from the response text through line-by-line coding and iterative thematic grouping. For positive feedback, inductive analysis yielded 3 categories: perceived learnability and low cognitive load, aesthetic acceptance, and perceived safety through help-seeking visibility. Areas for improvement produced 5 categories: gamification fatigue, onboarding deficit, navigation ambiguity, information density, and individual usability concerns. Feature suggestions yielded 4 categories: behavioral prompting and habit formation, intelligent clinical integration, privacy-controlled communication, and holistic well-being expansion.

Researcher Characteristics and Reflexivity

The research team possessed an interdisciplinary background combining expertise in game development, esports, and health informatics. The first author previously worked as a project manager within the esports industry. This prior insider knowledge aided in the contextual interpretation of data across both studies. The second author, serving as the supervisor, is an experienced researcher in digital health who reviewed all coding decisions to mitigate potential bias. These characteristics were considered throughout the research process, particularly during data analysis, where iterative discussions between both researchers aimed to balance insider familiarity with critical distance.

Ethical Considerations

All participants were asked to give their informed consent after reading information about the purpose of the project, the investigators, the length of time, data protection, and the questionnaire domains. To safeguard participant information, all data were treated confidentially. Expert interview recordings were pseudonymized immediately after transcription, with participants assigned neutral identifiers and the original recordings subsequently deleted. Athlete questionnaire data were collected anonymously, with no personally identifiable information retained. Ethical aspects were reviewed and approved by the Joint Ethics Committee of the Universities of Applied Sciences of Bavaria (GEHBa) in accordance with current scientific best-practice guidelines (GEHBa-202403-V-170).

Results

The following section presents the results of both studies, highlighting key findings and outcomes relevant to the research questions.

Expert Interview Results

Structured qualitative interviews were conducted with 7 (sports) psychologists via Google Meet (Table 1). The participants represented 5 different nationalities and included sports psychologists (4/7, 57%) and general psychologists (3/7, 43%). The aim was to explore their perspectives on mental health monitoring in esports and to identify key requirements for MindAthlete.

Table . Information on interview partners and interview details.

Expert ID	Sex	Degree	Language	Duration, min
Expert 1	Female	Sports Psychology	German	108
Expert 2	Male	Sports Psychology	German	100
Expert 3	Female	Psychology	German	56
Expert 4	Female	Sports Psychology	English	95
Expert 5	Female	Psychology	German	50
Expert 6	Female	Psychology	German	81
Expert 7	Male	Sports Psychology	German	53

Current Mental Health Monitoring Practices and Gaps

Experts affirmed the need for systematic monitoring given athletes' vulnerability to mental health issues (perceived importance of mental health tracking). Current practice, relying on questionnaires, passive observation, and informal check-ins (past and present monitoring methods), is limited by subjectivity, declining engagement, and uncertain self-report honesty, as Expert 7 noted: "I'm just a bit dependent on it: do they tell me the truth?" (sufficiency of these methods). Comprehensive monitoring should integrate psychological, behavioral, emotional, and physical dimensions, with mood as an early indicator and an esports-adapted context (comprehensive metrics for mental health assessment). This is reflected in the prototype's

XSL•FO
RenderX

combined use of validated assessment tools, lifestyle-focused context input, and behavioral self-report tasks that pair quantifiable ratings with contextual, esports-specific detail, directly addressing experts' concern that passive observation alone relies too heavily on subjective interpretation, while still capturing an esports-adapted picture that generic clinical tools do not offer.

Feature Requirements

Experts evaluated 8 candidate features (feedback on literature-based candidate features), most strongly endorsing the mood and stress trackers and a dashboard providing trend views for both athletes and practitioners, as Expert 3 put it: "You need a dashboard where I can glance at trends, spreadsheets simply don't cut it." Sport psychological questionnaires and mental health risk screening, both grounded in validated measures, were also well-supported. Stress relief techniques and help-seeking options were valued as around-the-clock resources, given esports' irregular schedules. Experts also proposed features beyond the preselected set (expert-suggested features), most notably a weekly check-in capturing lifestyle contexts, such as sleep, social contacts, and fitness, alongside push notifications, journaling, goal-setting, and a buddy system, of which only the weekly check-in ranked among the top 8. The ratings (feature ratings and reasons for ratings) converged strongly across psychologists and sports psychologists, with data collection features rated highest, as Expert 4 summarized: "All the trackers are the most important bit for me...But what is the point of having trackers if you do not have a dashboard?" This convergence highlights a shared need to turn raw data into actionable clinical insight, not merely to collect it. The 8 highest-ranked features (Table 2) directly informed the final feature set, clustering into continuous self-monitoring (mood tracker, stress tracker, and weekly check-in), structured assessment (sport psychological questionnaires and mental health risk screening), and intervention and insight (stress relief techniques, help-seeking options, and dashboard), reflecting a balance between mental health monitoring and on-demand support. Help-seeking options were prioritized over messages in a tie-break, as there are other already existing channels. Psychoeducation, while supported, raised feasibility concerns and did not rank in the final 8.

Table . Top 8 features for MindAthlete and their total points.

Feature	Points
Mood tracker	48
Stress tracker	43
Dashboard	27
Sport psychological questionnaires	21
Stress relief techniques	20
Mental health risk screening	19
Weekly check-in	13
Help-seeking options	11

Design and Implementation Considerations

Two deductive categories addressed practical design requirements. Experts favored nonclinical, gaming-familiar aesthetics, including a dark mode, a gaming-inspired color scheme, and consistently preferred desktop access for practitioners' data review alongside mobile access for athletes (interface preferences including colors and platform). Both the dark mode and color scheme were incorporated into the mobile prototype, and the platform preference directly shaped MindAthlete's dual-platform design: a mobile app for athletes and a web-based practitioner dashboard planned for future development. Experts also emphasized data security given the sensitivity of mental health data, recommending anonymized identifiers, biometric or PIN-based access, and transparent communication about data use (data security concerns), as Expert 4 noted: "Kind of considering the ethics of this, like where is the information going? Who has access to information?" These measures have not yet been implemented, having been deprioritized in favor of the top 8 core features and initial engagement mechanics.

Broader Application Context

Experts envisioned MindAthlete as a support ecosystem that reduces practitioners' administrative burden through preanalyzed data (purpose of the application), as Expert 1 put it: "I should not be looking at the data myself, it needs to land on my desk already analyzed." They also anticipated adoption barriers, including practitioners' limited review time, athletes' limited emotional literacy and tendency toward socially desirable responding, and a self-reinforcing risk of declining engagement if athletes perceive no tangible benefit from their input (challenges in application adoption). A related category emphasized the need for shared understanding among all staff, including ways for nonspecialist staff such as coaches to contribute their own observations about an athlete (cooperation with staff). Finally, experts recommended sustaining engagement through positive reinforcement, gamification, and minimal interaction friction (user engagement strategies). These insights were reflected in several design decisions. To address athletes' limited emotional literacy, the mood and stress trackers were designed with selectable response options, lowering the barrier to reporting. Experts' engagement recommendations were directly operationalized through the prototype's gamification features, confirming the relevance of the

XSL•FO
RenderX

point-based mechanic and rank system. Their suggestions regarding staff cooperation have not yet been implemented. Together with the broader vision and adoption risks described above, these considerations extend beyond the current prototype's scope and inform priorities for future development.

These expert insights directly address research question 1 by identifying and prioritizing the core features considered most relevant for mental health monitoring in esports.

Athlete Usability Testing Results

A total of 134 esports athletes, identified as competitors in 1 of 5 different video games, were contacted for this study. This cohort primarily reflected the male-dominated demographic typical of the professional esports scene. After excluding 3 incomplete responses, the final sample comprised 25 participants. The sample was predominantly female (n=17, 68%) and heavily concentrated in the 22 to 25 age bracket (n=14, 56%). The cohort consisted almost entirely of League of Legends players (n=24, 96%), with an average professional experience of 4.8 (SD 3.2) years, and 17 of 25 (68%) relying on esports as their primary income source. Notably, 22 of 25 (88%) reported no prior experience with mHealth apps. Complete participant demographics are detailed in Table 3.

Table . Participant demographics (N=25).

Characteristic	Participants, n (%)
Age group (y)	
18 - 21	2 (8)
22 - 25	14 (56)
26 - 29	8 (32)
30 - 33	1 (4)
Sex	
Female	17 (68)
Male	7 (28)
Other	1 (4)
Nationality	
Germany	12 (48)
France	4 (16)
Great Britain	3 (12)
Spain	2 (8)
Other	4 (16)
Primary game	
League of Legends	24 (96)
EA Sports FC	1 (4)
Esports as main source of income	
Yes	17 (68)
No	8 (32)
Prior mHealth app usage	
Yes	3 (12)
No	22 (88)

The distribution of SUS scores significantly deviated from normality (Shapiro-Wilk *W* 0.915, N=25; *P*=.04), showing moderate left-skewness (skewness −0.81). Therefore, the median is reported as the primary measure of central tendency: median 77.5 (IQR 67.5 - 87.5), range 42.5 - 92.5, classifying as "good" usability according to established SUS benchmarks [40]. Analysis using the IQR method did not identify outliers (Table 4).

XSL•FO
RenderX

Table . Descriptive summary of System Usability Scale (SUS) scores overall and by sex and age.

Group	n (%)	Median (IQR); range
Total	25 (100)	77.5 (67.5 - 87.5); 42.5 - 92.5
Sex		
Female	17 (68)	75.0 (65.0 - 82.5); 42.5 - 92.5
Male	7 (28)	82.5 (73.8 - 86.3); 47.5 - 92.5
Other[a]	1 (4)	90.0[b]
Age (y)		
18-21[a]	2 (8)	80.0[b]
22 - 25	14 (56)	73.8 (61.3 - 79.4); 42.5 - 90.0
26 - 29	8 (32)	86.3 (79.4 - 88.1); 47.5 - 92.5
30-33[a]	1 (4)	82.5[b]

[a]SD is not reported for n<2. Data are left-skewed; the median (IQR) is reported as the primary measure.

[b]IQR and range are not reported for n<3.

Sex differences in SUS scores (female: n=17, 68%; male: n=7, 28%; other: n=1, 4%) were evaluated using the Mann-Whitney *U* test for female and male participants only. Females had a median SUS of 75.0 (IQR 65.0 - 82.5), and males had 82.5 (IQR 73.8 - 86.3). The Mann-Whitney test revealed no significant difference (U=74.0, z=−0.89, adjusted P>.99, r=0.19). Comparing age groups, the 22 to 25 (n=14, 56%) group showed a median SUS of 73.8 (IQR 61.3 - 79.4), and the 26 to 29 (n=8, 32%) group had 86.3 (IQR 79.4 - 88.1). Initially, the Mann-Whitney *U* test indicated a difference (U=26.5, z=−2.06, P=.047, r=0.43), but after Bonferroni correction, the adjusted P=.33, rendering the difference nonsignificant. Age groups with fewer than 5 participants (18 - 21 y: n=2, 8%; 30 - 33 y: n=1, 4%) were reported descriptively only. Spearman correlations between SUS scores and years of professional experience (ρ=0.14, adjusted P>.99) as well as paid experience (ρ=0.12, adjusted P>.99) were nonsignificant. A Mann-Whitney test comparing SUS by main income status (yes vs no) showed no significant difference (U=62.5, z=−0.29, adjusted P>.99, r=0.06). Similarly, prior mHealth usage yielded no significant difference (U=27.0, z=−0.47, adjusted P>.99, r=0.10). A Kruskal-Wallis test for nationality differences was nonsignificant (H[7]=6.29, adjusted P>.99), and small group sizes in some nationalities precluded further inferential analysis.

Table 5 summarizes the qualitative feedback provided by participants in response to the open-ended questions and is grouped into 3 domains: positive feedback, areas for improvement, and suggested features.

XSL•FO
RenderX

Table . Athlete qualitative feedback categories from open-ended responses (N=25).

Domain and category	Participants, n (%)	Description
Positive feedback		
Perceived learnability and low cognitive load	13 (52)	Interface experienced as immediately accessible without prior instruction
Aesthetic acceptance	7 (28)	Visual design perceived as engaging rather than clinical
Perceived safety through help-seeking visibility	5 (20)	Prominent emergency support options contributed to a sense of psychological safety
Areas for improvement		
Gamification fatigue	5 (20)	Lantern and trophy animations perceived as disruptive with repeated use
Onboarding deficit	3 (12)	Absence of a first-launch guide created uncertainty despite an intuitive interface
Navigation ambiguity	2 (8)	Bottom-menu icons lacked sufficient labeling
Information density	2 (8)	Home screen perceived as visually crowded
Individual usability concerns	3 (12)	Isolated issues, including a mobile scrolling bug, session-scheduling stress, and ranking-system confusion
Suggestions for additional features		
Behavioral prompting and habit formation	4 (16)	Reminders and alarms requested to sustain daily tracking behavior
Intelligent clinical integration	3 (12)	Desire for automated follow-up recommendations linking data to professional support
Privacy-controlled communication	2 (8)	Need for a private reflection space and a separate staff communication channel
Holistic well-being expansion[a]	1 (4)	Nature-sound relaxation, gamified daily-streak tracker, sleep and nutrition tracking, weekly summary dashboard

[a]Each suggestion was raised by a single participant and did not meet the threshold for a distinct category.

Positive feedback centered on ease of use, most prominently in the perceived learnability and low cognitive load category (13/25, 52%), reflecting athletes' experience of the interface as immediately accessible without prior instruction. As P3 remarked, "The basic level of the UI made it feel very easy to understand and use when needing to enter information, making conveying what I felt into the application be done confidently." Athletes also perceived the visual design as engaging rather than clinical and reported that prominent help-seeking options contributed to a sense of psychological safety. Together, these categories suggest that the design achieved both functional usability and the emotional accessibility necessary for engaging with sensitive self-monitoring content.

Critical feedback concentrated on motivational elements, initial orientation, and visual clarity rather than core functionality. The most frequently cited concern was gamification fatigue (5/25, 20%): while gamification elements were implemented to enhance engagement, the animations were perceived as disruptive with repeated use, as P1 noted, "the speed at which the lantern and trophy animations are played...feels like it could become annoying." Onboarding and interface clarity issues were also raised. These findings suggest that MindAthlete's core design foundations are sound, with targeted refinements needed in motivational calibration and onboarding.

Suggested features extended beyond aesthetic preferences toward a more integrated vision of the app. The most common request was for behavioral prompts such as reminders and alarms to sustain daily tracking (4/25, 16%), followed by a desire for automated follow-up recommendations that link daily data to professional support, reinforcing the dual-user ecosystem underlying the prototype's design.

These findings provide a direct answer to research question 2, demonstrating that the prototype is perceived as user-friendly while also highlighting concrete areas for refinement and future development.

Discussion

The following section discusses the key findings of this study, reflects on methodological limitations, and outlines implications for future research and development.

Interpretation of Key Findings

Experts' unanimous recognition of mental health tracking as a core professional responsibility reflects esports athletes' high vulnerability to burnout and anxiety, as well as related stressors such as performance pressure and demanding schedules [5,6,10-12], with untreated issues compromising both immediate

XSL•FO
RenderX

performance and long-term well-being [5,14]. This is also consistent with documented pathways from esports-specific stressors to mental ill health [8], including burnout [12] and night-eating [13]. Furthermore, experts' emphasis on comprehensive mood tracking and dashboard visualization mirrors findings that early mood deterioration predicts severe episodes [41] and that continuous data visualization supports clinical decision-making [25,42], while future sensor integration enriches self-reported psychological data [43]. Feature ratings indicate an underlying rationale: partial deployment (eg, providing tracking features without an integrated dashboard) may limit their practical value. This suggests that MindAthlete's features may be most useful when considered together, rather than as fully independent components. Inductive categories highlight engagement and organizational integration as central challenges. The fear of a self-reinforcing cycle in which absent feedback leads to indifferent athlete input mirrors persistent engagement drop-off documented across mHealth self-reporting tools more broadly [24,25], underscoring that closing the feedback loop is essential for sustained use. Similarly, organizational integration may be as important as technical functionality, consistent with evidence that workflow integration, institutional culture, and staff engagement are among the most prevalent determinants of mHealth adoption, often outweighing technological considerations [44].

MindAthlete's usability score exceeded the benchmark median of 68.3 for mHealth apps, excluding physical activity applications [45], aligning with comparable athlete-facing tools ranging from 72.89 to 78.5 [46-48], despite being an unpolished prototype. Athletes' qualitative feedback suggested that MindAthlete achieved emotional accessibility alongside functional usability, a distinction SUS scores alone cannot capture, particularly relevant given evidence that stigma and embarrassment limit engagement with mental health apps [49]. This is consistent with evidence that esports players welcome personalized, trustworthy support [16], suggesting MindAthlete may address these expectations. Furthermore, their requests for behavioral prompts and privacy controls reinforce health informatics research highlighting the necessity of trust and social scaffolding for long-term engagement [50].

Traditional sports mHealth tools often focus on physical workload, psychological skills training, or mood tracking, but are not designed for esports-specific stressors and rhythms [17,26], while generic mental health apps rely on nonspecialized screening and face persistent engagement challenges [23-25]. By grounding assessment categories, terminology, and gamification mechanics directly in esports culture, MindAthlete represents an initial step toward adapting mHealth principles to the specific performance context of esports.

Potential Implications for Practice

These findings suggest several potential applications for MindAthlete beyond the current prototype stage, though each would require further development and validation. For organizations without dedicated psychological staff, the always-on trackers and screening features may offer accessible first-line support, and experts noted that a future practitioner dashboard could help nonspecialist staff recognize concerning trends. In organizations with existing psychological support, such a dashboard could complement rather than replace informal monitoring practices. More broadly, MindAthlete's approach is consistent with blended models that combine digital self-monitoring with in-person or remote psychological support, which have been associated with improved continuity of care [27,28], though realizing this potential would require careful attention to data ownership and escalation protocols in future iterations.

Limitations

Several limitations merit acknowledgment. First, relying primarily on experts during requirements gathering risked overlooking athlete-led innovations from lived experience. Experts were used exclusively at this stage because they could ground candidate features in established clinical and sport psychology practice, and we mitigated the resulting bias by validating and extending these features through athlete usability testing and open-ended qualitative feedback (research question 2). Only 7 experts were recruited, though strong thematic convergence suggests key requirements were captured despite the absence of full saturation. Snowball sampling may have introduced homogeneity among participating experts, and no formal intercoder reliability statistics were calculated, though a senior supervisor reviewed the coding structure to resolve discrepancies. Additionally, implementation constraints limited the prototype to 8 features, though experts could still propose novel features beyond this set to capture insights outside our literature-based preselection. Suggestions varied considerably across experts, but one novel feature was requested multiple times and ultimately incorporated. While this open approach increased data variability, it enriched the evaluation with previously unconsidered practitioner perspectives.

Regarding the usability testing, additional limitations exist that must be considered. Most notably, 68% of participants were female, a marked mismatch with the male-dominated professional esports industry. Although 134 predominantly male athletes were contacted, the final sample reflects strong self-selection toward female participation, consistent with evidence that men are less likely to engage in routine health-related help-seeking and survey participation [51,52]. Because engaging with a mental health application can itself be understood as help-seeking behavior, usability scores and open-ended feedback should be interpreted with caution. The sample was also highly homogeneous, concentrated in League of Legends (96%) and Germany (48%), reflecting the regional concentration of the European esports ecosystem and our recruitment network within it. Regional differences in working hours, training structures, and health-seeking norms—for instance, the considerably longer and less regulated training hours reported among professional esports athletes in China [53]—may limit the generalizability of usability perceptions to less structured competitive environments.

Relatedly, the cognitive and physiological demands of League of Legends as a Multiplayer Online Battle Arena (MOBA) title differ from those of other genres, with First-Person Shooter titles eliciting greater cardiovascular stress than MOBA titles [54], and MOBA gameplay emphasizing strategic planning and

XSL•FO
RenderX

coordination over the reaction-time demands of First-Person Shooter and Real-Time Strategy titles [1]. These genre-specific profiles are relevant to the feature set, as stressors such as team communication breakdowns are structurally absent in individual titles such as fighting games or sim racing, and League of Legends' team size may have provided athletes with more immediate social support, potentially reducing perceived need for MindAthlete.

Finally, only athletes aged 18 and older were included, and evaluations were based on a high-fidelity prototype rather than a final product. Motivation loss, notification fatigue, and long-term engagement could not be assessed, and positive usability scores may not reflect sustained real-world use. Cultural or prior experience factors may also have influenced the interpretation of SUS items.

Conclusions

This study addressed the need for efficient mental health monitoring in esports by defining practitioner-driven requirements through expert interviews, translating them into an interactive prototype, and evaluating its usability with esports athletes. By grounding the prototype in esports-specific expert and athlete input, this study offers a practical foundation for future mental health tools tailored to the competitive esports context, rather than adapted from generic or traditional sports solutions. MindAthlete's median SUS score of 77.5 demonstrates good usability for this niche audience. However, these findings must be interpreted in light of the sample's demographic homogeneity. Future work should begin with the technical implementation of the mobile app and a web-based practitioner dashboard, the development of a validated crisis intervention protocol, and the establishment of robust data privacy and security mechanisms, developed together with (sports) psychologists, who are also being consulted on instrument selection and escalation management. Athlete feedback from this study should inform the next design iteration. Moreover, further studies should be conducted with larger sample sizes. Specifically, future research must deliberately target cross-cultural and longitudinal studies, focusing on non-European regions, other competitive game titles, and more male-dominated cohorts to explore whether usability and feature preferences differ, and to assess long-term engagement. Finally, integrating passive data collection via wearables could further enable objective monitoring and remote intervention, though this would require careful management of data overload, false positives, and privacy concerns.

Acknowledgments

The authors thank the esports athletes for their participation and candid feedback, and the (sports) psychologists whose expertise was essential in shaping MindAthlete's feature set.

Parts of the manuscript were refined using generative AI tools: Gemini (Google DeepMind) and ChatGPT (OpenAI) were used for translation and grammatical editing, while Claude (Anthropic) was used for text condensation and language editing (all accessed in September 2026). The authors have reviewed and take full responsibility for all final content.

Funding

The authors declared no financial support was received for this work.

Data Availability

The raw data supporting the conclusions of this paper will be available upon request.

Authors' Contributions

Conceptualization: LS
Data curation: LS
Methodology: LS, JS
Project administration: JS
Validation: LS
Visualization: LS
Writing – original draft: LS, DH, JS
Writing – review & editing: LS, DH, PK, MH, JS

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guideline and codebook.
[DOCX File, 35 KB - humanfactors_v13i1e90140_app1.docx]

References

XSL•FO
RenderX

1. Pedraza-Ramirez I, Musculus L, Raab M, Laborde S. Setting the scientific stage for esports psychology: a systematic review. Int Rev Sport Exerc Psychol 2020 Jan 1;13(1):319-352. [doi: 10.1080/1750984X.2020.1723122]
2. Tang D, Liu J, Gou P, et al. Association between esports participation and health: a systematic review and meta-analysis. Sports Med Open 2026 Jul 29;12(1):108. [doi: 10.1186/s40798-026-01083-0] [Medline: 42525214]
3. Monteiro Pereira A, Costa JA, Verhagen E, Figueiredo P, Brito J. Associations between esports participation and health: a scoping review. Sports Med 2022 Sep;52(9):2039-2060. [doi: 10.1007/s40279-022-01684-1] [Medline: 35486374]
4. Sanz-Matesanz M, Gea-García GM, Martínez-Aranda LM. Physical and psychological factors related to player's health and performance in esports: a scoping review. Comput Human Behav 2023 Jun;143:107698. [doi: 10.1016/j.chb.2023.107698]
5. Leis O, Lautenbach F, Birch PD, Elbe AM. Stressors, associated responses, and coping strategies in professional esports players: a qualitative study. Int J Esports 2022;1(1) [FREE Full text]
6. Smith MJ, Birch PDJ, Bright D. Identifying stressors and coping strategies of elite esports competitors. Int J Gaming Comput-Mediat Simul 2019;11(2):22-39. [doi: 10.4018/IJGCMS.2019040102]
7. Leis O, Lautenbach F. Stress management in esports. In: Sharpe BT, Poulus DR, editors. The Psychology of Esports Performance, 1st edition: Routledge; 2026:135-157. [doi: 10.4324/9781003663676-10]
8. Smith M, Sharpe B, Arumuham A, Birch P. Examining the predictors of mental ill health in esport competitors. Healthcare (Basel) 2022 Mar 26;10(4):626. [doi: 10.3390/healthcare10040626] [Medline: 35455804]
9. Kegelaers J, Mairesse O, Van Ruysevelt L, et al. The mental health status of esports athletes. J Electron Gaming Esports 2025;3(1). [doi: 10.1123/jege.2025-0017]
10. Lee S, Bonnar D, Roane B, et al. Sleep characteristics and mood of professional esports athletes: a multi-national study. Int J Environ Res Public Health 2021 Jan 14;18(2):664. [doi: 10.3390/ijerph18020664] [Medline: 33466788]
11. Pereira AM, Teques P, Verhagen E, Gouttebarge V, Figueiredo P, Brito J. Mental health symptoms in electronic football players. BMJ Open Sport Exerc Med 2021;7(4):e001149. [doi: 10.1136/bmjsem-2021-001149] [Medline: 34691761]
12. Ahn H, Kim I. An exploratory study on the conceptualization of burnout among the professional esports athletes: focused on League of Legends Champions Korea League. Healthcare (Basel) 2024 May 31;12(11):1127. [doi: 10.3390/healthcare12111127] [Medline: 38891202]
13. Arslan S, Atan RM, Sahin N, Ergul Y. Evaluation of night eating syndrome and food addiction in esports players. Eur J Nutr 2024 Aug;63(5):1695-1704. [doi: 10.1007/s00394-024-03368-0] [Medline: 38520524]
14. Kegelaers J, Trotter MG, Watson M, et al. Promoting mental health in esports. Front Psychol 2024;15:1342220. [doi: 10.3389/fpsyg.2024.1342220] [Medline: 38312393]
15. Schary DP, Jenny SE, Koshy A. Leveling up esports health: current status and call to action. Int J Esports 2022;1(1) [FREE Full text]
16. Leis O, Meichsner N, Swettenham LD. Exploring esport players' perspectives on sport psychology support: an interview study. Team Perform Manag 2025 Nov 11;31(5-6):347-363. [doi: 10.1108/TPM-07-2024-0080]
17. Balcombe L, De Leo D. Psychological screening and tracking of athletes and digital mental health solutions in a hybrid model of care: mini review. JMIR Form Res 2020 Dec 14;4(12):e22755. [doi: 10.2196/22755] [Medline: 33271497]
18. Funnell EL, Spadaro B, Martin-Key N, Metcalfe T, Bahn S. mHealth solutions for mental health screening and diagnosis: a review of app user perspectives using sentiment and thematic analysis. Front Psychiatry 2022;13:857304. [doi: 10.3389/fpsyt.2022.857304] [Medline: 35573342]
19. Price M, Yuen EK, Goetter EM, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436. [doi: 10.1002/cpp.1855] [Medline: 23918764]
20. O'Rourke T, Vogel C, John D, et al. The impact of coping styles and gender on situational coping: an ecological momentary assessment study with the mHealth application TrackYourStress. Front Psychol 2022;13:913125. [doi: 10.3389/fpsyg.2022.913125] [Medline: 35795429]
21. Hashemi B, Ali S, Awaad R, Soudi L, Housel L, Sosebee SJ. Facilitating mental health screening of war-torn populations using mobile applications. Soc Psychiatry Psychiatr Epidemiol 2017 Jan;52(1):27-33. [doi: 10.1007/s00127-016-1303-7] [Medline: 27815623]
22. Kuhn E, Greene C, Hoffman J, et al. Preliminary evaluation of PTSD Coach, a smartphone app for post-traumatic stress symptoms. Mil Med 2014 Jan;179(1):12-18. [doi: 10.7205/MILMED-D-13-00271] [Medline: 24402979]
23. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res 2018 Dec;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006] [Medline: 30347316]
24. Polillo A, Kozloff N, Dixon LB. Mobile "mHealth" interventions in mental health care. Psychiatr Serv 2021 Nov 1;72(11):1363-1364. [doi: 10.1176/appi.ps.721103] [Medline: 34734753]
25. Bidargaddi N, Schrader G, Klasnja P, Licinio J, Murphy S. Designing m-Health interventions for precision mental health support. Transl Psychiatry 2020 Jul 7;10(1):222. [doi: 10.1038/s41398-020-00895-2] [Medline: 32636358]
26. Bonetti R, Rod B, Sabourin C, Hauw D. Quality of mobile apps for psychological skills training in sport: A MARS-based study. Computers in Human Behavior Reports 2025 May;18:100635. [doi: 10.1016/j.chbr.2025.100635]
27. Chan SR, Torous J, Hinton L, Yellowlees P. Mobile tele-mental health: increasing applications and a move to hybrid models of care. Healthcare (Basel) 2014 May 6;2(2):220-233. [doi: 10.3390/healthcare2020220] [Medline: 27429272]

XSL•FO
RenderX

28. Valentine L, McEnery C, Bell I, et al. Blended digital and face-to-face care for first-episode psychosis treatment in young people: qualitative study. JMIR Ment Health 2020 Jul 28;7(7):e18990. [doi: 10.2196/18990] [Medline: 32720904]
29. Brown T. Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation: Harper Business; 2009.
30. Haberl A, Fleiß J, Kowald D, Thalmann S. Take the aTrain. Introducing an interface for the accessible transcription of interviews. J Behav Exp Finance 2024 Mar;41:100891. [doi: 10.1016/j.jbef.2024.100891]
31. Mayring P. Qualitative content analysis. Forum Qual Sozialforsch 2000;1(2):159-176. [doi: 10.17169/fqs-1.2.1089]
32. Figma. URL: https://www.figma.com [accessed 2026-02-04]
33. Nielsen J. Enhancing the explanatory power of usability heuristics. In: Olson JS, Dumais S, Olson JS, editors. CHI '94: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Association for Computing Machinery; 1994:152-158. [doi: 10.1145/191666.191729]
34. Norman DA, editor. User-Centered System Design: New Perspectives on Human-Computer Interaction, 1st edition: CRC Press; 1986. [doi: 10.1201/9780367807320]
35. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
36. Deci EL, Ryan RM. The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 2000 Oct;11(4):227-268. [doi: 10.1207/S15327965PLI1104_01]
37. LimeSurvey. URL: http://www.limesurvey.org [accessed 2026-04-02]
38. Brooke J. SUS—a quick and dirty usability scale. In: Jordan PW, McClelland IL, Weerdmeester BA, Thomas B, editors. Usability Evaluation in Industry: Taylor & Francis; 1996:189-194.
39. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003 Aug;35(3):379-383. [doi: 10.3758/bf03195514] [Medline: 14587545]
40. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
41. Costello EJ. Early detection and prevention of mental health problems: developmental epidemiology and systems of support. J Clin Child Adolesc Psychol 2016;45(6):710-717. [doi: 10.1080/15374416.2016.1236728] [Medline: 27858462]
42. Schwab JD, Schobel J, Werle SD, et al. Perspective on mHealth concepts to ensure users' empowerment—from adverse event tracking for COVID-19 vaccinations to oncological treatment. IEEE Access 2021;9:83863-83875. [doi: 10.1109/ACCESS.2021.3087315]
43. Karthan M, Martin R, Holl F, et al. Enhancing mHealth data collection applications with sensing capabilities. Front Public Health 2022;10:926234. [doi: 10.3389/fpubh.2022.926234] [Medline: 36187627]
44. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR mHealth uHealth 2020 Feb 20;8(2):e15935. [doi: 10.2196/15935] [Medline: 32130167]
45. Hyzy M, Bond R, Mulvenna M, et al. System usability scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022 Aug 18;10(8):e37290. [doi: 10.2196/37290] [Medline: 35980732]
46. Bae SW, Laccetti A, Ozolcer M, Zhang T. PXAI-coach: designing and evaluating a person-centric explainable AI dashboard for athlete health monitoring and coaching. Proc Int Conf Act Behav Comput 2025:1-13. [doi: 10.1109/ABC64332.2025.11118561]
47. Santos-Gago JM, Ramos-Merino M, Alvarez-Sabucedo JA, Santos-González I. Towards a personalised recommender platform for sportswomen. In: Rocha A, Adeli H, Reis L, Costanzo S, editors. New Knowledge in Information Systems and Technologies: Springer; 2019:504-514. [doi: 10.1007/978-3-030-16181-1_48]
48. Rahman MM, Morshed MB, Rashid N, Chandrashekhara S, Kuang J. MindfulBuddy: extracting comprehensive breathing biomarkers for breathing exercise biofeedback using earbud motion sensors. IEEE Internet Things J 2025;12(12):20420-20434. [doi: 10.1109/JIOT.2025.3545471]
49. Garrido S, Oliver E, Chmiel A, Doran B, Boydell K. Encouraging help-seeking and engagement in a mental health app: what young people want. Front Digit Health 2022;4:1045765. [doi: 10.3389/fdgth.2022.1045765] [Medline: 36620186]
50. Murnane EL, Walker TG, Tench B, Voida S, Snyder J. Personal informatics in interpersonal contexts: towards the design of technology that supports the social ecologies of long-term mental health management. Proc ACM Hum-Comput Interact 2018;2(CSCW):127. [doi: 10.1145/3274396]
51. Susukida R, Mojtabai R, Mendelson T. Sex differences in help seeking for mood and anxiety disorders in the National Comorbidity Survey-Replication. Depress Anxiety 2015 Nov;32(11):853-860. [doi: 10.1002/da.22366] [Medline: 25903117]
52. Wendt D, Shafer K. Gender and attitudes about mental health help seeking: results from national data. Health Soc Work 2016 Feb;41(1):e20-e28. [doi: 10.1093/hsw/hlv089]
53. Pang Z, Su L, Zhang Y. Daily physical activity, coffee and energy drink consumption, and sleep patterns among Chinese elite professional esports athletes: a case study of Zhejiang Regans Gaming. Front Public Health 2025;13:1557533. [doi: 10.3389/fpubh.2025.1557533] [Medline: 40078766]
54. Sousa A, Ahmad SL, Hassan T, et al. Physiological and cognitive functions following a discrete session of competitive esports gaming. Front Psychol 2020;11:1030. [doi: 10.3389/fpsyg.2020.01030] [Medline: 32547452]

XSL•FO
RenderX

Abbreviations

APSQ: Athlete Psychological Strain Questionnaire
GAD-7: Generalized Anxiety Disorder-7
mHealth: mobile health
MOBA: Multiplayer Online Battle Arena
PHQ-9: Patient Health Questionnaire-9
PSS: Perceived Stress Scale
SUS: System Usability Scale
WHO-5: World Health Organization-Five Well-Being Index

Edited by S Law; submitted 22.Dec.2025; peer-reviewed by B Sharpe, D Tang; revised version received 07.Sep.2026; accepted 10.Sep.2026; published 30.Sep.2026.

Please cite as:
Stolberg L, Hieber D, Kuhn P, Hebel M, Schobel J
Mental Health Monitoring in Esports Athletes Using an mHealth App: Requirements Analysis and Usability Study
JMIR Hum Factors 2026;13:e90140
URL: https://humanfactors.jmir.org/2026/1/e90140
doi:10.2196/90140

© Leona Stolberg, Daniel Hieber, Peter Kuhn, Michael Hebel, Johannes Schobel. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 30.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Co-Designing a Digital Platform for the Management of Modifiable Risk Factors and Comorbidities in Patients With Atrial Fibrillation: Mixed Methods Usability Study

Sofia Klavebäck[1,2], MD; Frieder Braunschweig[1,2], MD, PhD; Matthias Lidin[1,3*], RN, PhD; Emma Svennberg[1,2*], MD, PhD

[1]Department of Cardiology, Heart and Vascular Center, Karolinska University Hospital, Stockholm, Sweden

[2]Department of Medicine Huddinge, Unit of Integrative Cardiovascular Research (ICCA), Karolinska Institutet, Stockholm, Sweden

[3]Department of Medicine, Karolinska Institutet, Solna, Stockholm, Sweden

[*]these authors contributed equally

Corresponding Author:
Sofia Klavebäck, MD
Department of Medicine Huddinge
Unit of Integrative Cardiovascular Research (ICCA)
Karolinska Institutet
H7, Medicin Huddinge
ICCA, Braunschweig
Stockholm, 17177
Sweden
Phone: 46 8 123 80000
Email: sofia.klaveback@ki.se

Abstract

Background: Clinical guidelines recommend integrated, person-centered care with control of modifiable risk factors and coexisting conditions in patients with atrial fibrillation (AF), but many receive insufficient risk factor management. Digital health technologies may provide valuable support in addressing this gap.

Objective: Our aim was to evaluate the usability of a co-designed digital platform for supporting person-centered management of modifiable risk factors in individuals with AF.

Methods: This mixed methods usability study was conducted at a tertiary university hospital in Sweden between March and April 2025. Eligible participants with AF were recruited to evaluate the co-designed digital platform using the System Usability Scale (SUS) and semistructured interviews analyzed with inductive manifest content analysis.

Results: A total of 22 patients were included (mean age 67.7, SD 7.6 y); 32% (7/22) were female, the mean BMI was 27.7 (SD 3.6) kg/m^2, 65% (14/22) had paroxysmal AF, 36% (8/22) had persistent AF, and the median AF duration was 26 (IQR 13.3-72.8) months. Relevant comorbidities were hypertension (17/22, 77%), heart failure (8/22, 36%), diabetes mellitus type 2 (3/22, 14%), and ischemic heart disease (4/22, 18%). Usability was rated high, with a median SUS score of 78.8 (IQR 63.8-86.9), indicating above-average user acceptance. Participants' requirements were summarized into 4 main categories and 10 subcategories. First, they valued a *clear layout*, *simple design*, and *easy navigation*. Second, they appreciated *positive content*, which is *informative*, *inclusive*, and *motivating*. Third, they requested *personalized information* on *different aspects* and provided on *different levels*. Fourth, they desired *individualized medical recommendations* that are *personalized* and *flexible* but *open to individual choice*.

Conclusions: Participants found the usability acceptable and valued specifically a *clear layout*, *positive content*, *personalized information*, and *individualized medical recommendations*.

(JMIR Hum Factors 2026;13:e95810) doi:10.2196/95810

KEYWORDS

atrial fibrillation; risk factors; mHealth; mobile health; person-centered care; digital health

XSL•FO
RenderX

Introduction

The incidence of atrial fibrillation (AF) increases with age and affects more than 3% of adults in a population-based Swedish study, making it the most common cardiac arrhythmia [1]. Patients can experience paroxysmal or persistent AF, and for symptomatic patients, guidelines recommend restoring sinus rhythm, which can be achieved by medical treatment, electrical cardioversion, or more permanently via catheter ablation [2].

AF is associated with nonmodifiable risk factors such as increasing age, height, and male sex [3]. However, it is also associated with modifiable and dynamic risk factors related to comorbidities including hypertension, hyperglycemia, obesity, obstructive sleep apnea, and mental health conditions, but also to unhealthy lifestyle choices including risk consumption of alcohol and physical inactivity [3]. Over half of AF cases in middle-aged adults can be attributed to inadequate control of modifiable risk factors [4]. Several studies have shown that better risk factor management is associated with lower risk of incident AF [5-7], sinus rhythm maintenance [8], and improved outcomes after catheter ablation [9]. Therefore, guidelines specifically recommend integrated management of modifiable risk factors for all persons with AF, especially those undergoing rhythm interventions [3]. Furthermore, the benefits of person-centered care have recently been emphasized, including multidisciplinary, coordinated, and tailored solutions focusing on shared decision-making [10].

Digital health interventions may facilitate person-centered care by supporting self-management and risk factor modification. Although such interventions have shown promising results in cardiovascular disease and other chronic conditions [11,12], evidence in AF is still limited. We therefore co-designed a digital platform to support person-centered management of modifiable risk factors in patients with AF where we now aim to evaluate usability.

Specifically, we aimed to investigate how patients with AF evaluate design, function, and user experience of a digital platform, to identify how to further optimize it according to the needs of the patients, the end users.

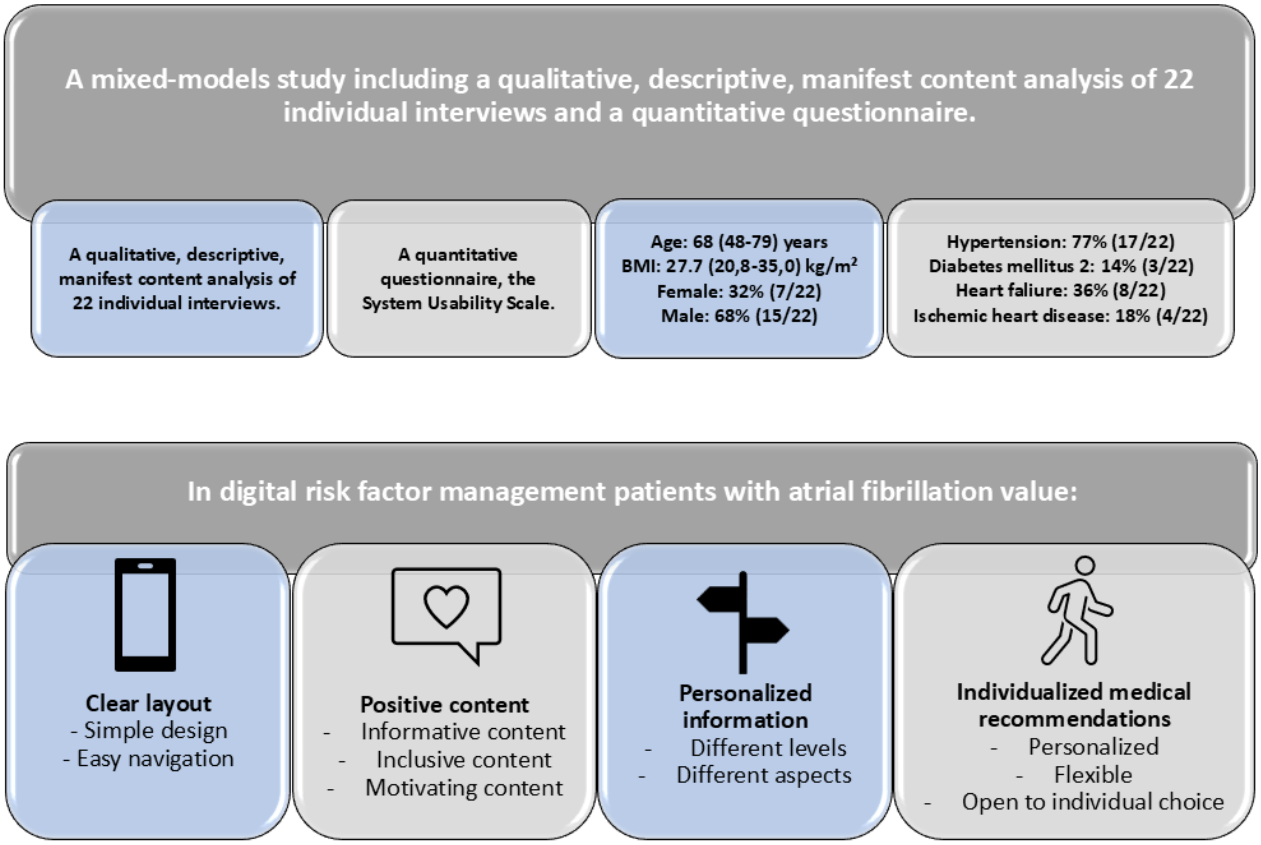
Methods

Description of the Co-Design Process

We co-designed a digital platform with the aim of improving support for patients with AF in the management of modifiable risk factors (Figure 1). As a first step, we conducted a study where we identified unmet needs with regard to modifiable risk factors among individuals with AF and further explored the specifics of the desired functions and usability of such a platform [13]. Specifically, participants with AF requested *person-centered information*, support in *managing healthy lifestyle habits*, and more *regular communication* with caregivers [13]. There was a general desire for person-centered care, defined as a partnership where interdisciplinary health care professionals actively listen to and involve the patient in forming a health plan together [14-17]. A prototype was developed based on these findings, tested with the same patient group, and subsequently adjusted according to their feedback, resulting in the co-designed digital prototype evaluated in this study. In this next step of the co-design, we aim to more broadly evaluate the usability of the prototype of our digital platform for supporting person-centered management of modifiable risk factors in individuals with AF.

XSL•FO
RenderX

Figure 1. Graphical abstract.



Description of the Digital Platform

The platform can be used on a smartphone or computer, where patients waiting for catheter ablation for AF enter their medical information, including data on comorbidities and medications, and answer questionnaires about symptoms and lifestyle. Based on their input, an algorithm identifies which factors are scored outside of targets recommended by the European Society of Cardiology guidelines for AF. For example, regarding comorbidities, if a patient enters a BMI>27 kg/m^2 or blood pressure>130/80 mm/Hg, the algorithm identifies these possible targets to optimize based on guidelines. Questionnaires are used in a similar fashion to identify responses that might merit additional attention, for example, individuals who do not reach recommended targets of healthy eating habits or physical activity. These areas are then presented as suggestions to the user to improve. However, the choice of which recommendations to work with remains with the users. Based on the users' choice, a week-long sprint is initiated, providing advice on their areas of improvement. These have been developed with a nurse, dietitian, physician, and physiotherapist and are targeted at each area of improvement with a new task each week. For example, the advice can include weight management support for elevated BMI, physical activity recommendations specific for patients with AF, and healthy recipes to improve eating habits. Users can also follow the progress of measurable values such as blood pressure and BMI over time. Examples of these interventions are illustrated in Figure 2 and Figure 3. Development is purely research-based, and there are no commercial interests involved.

XSL•FO
RenderX

Figure 2. Illustration of the co-design process.

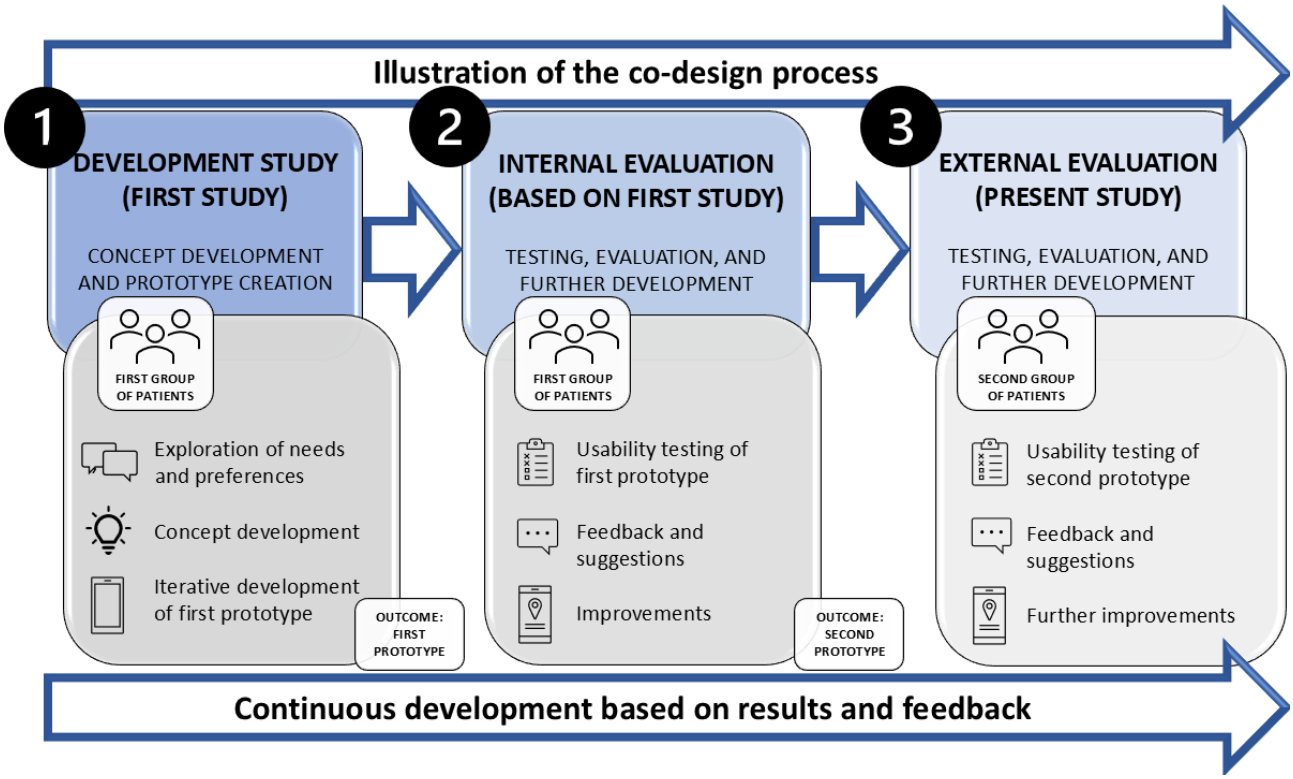
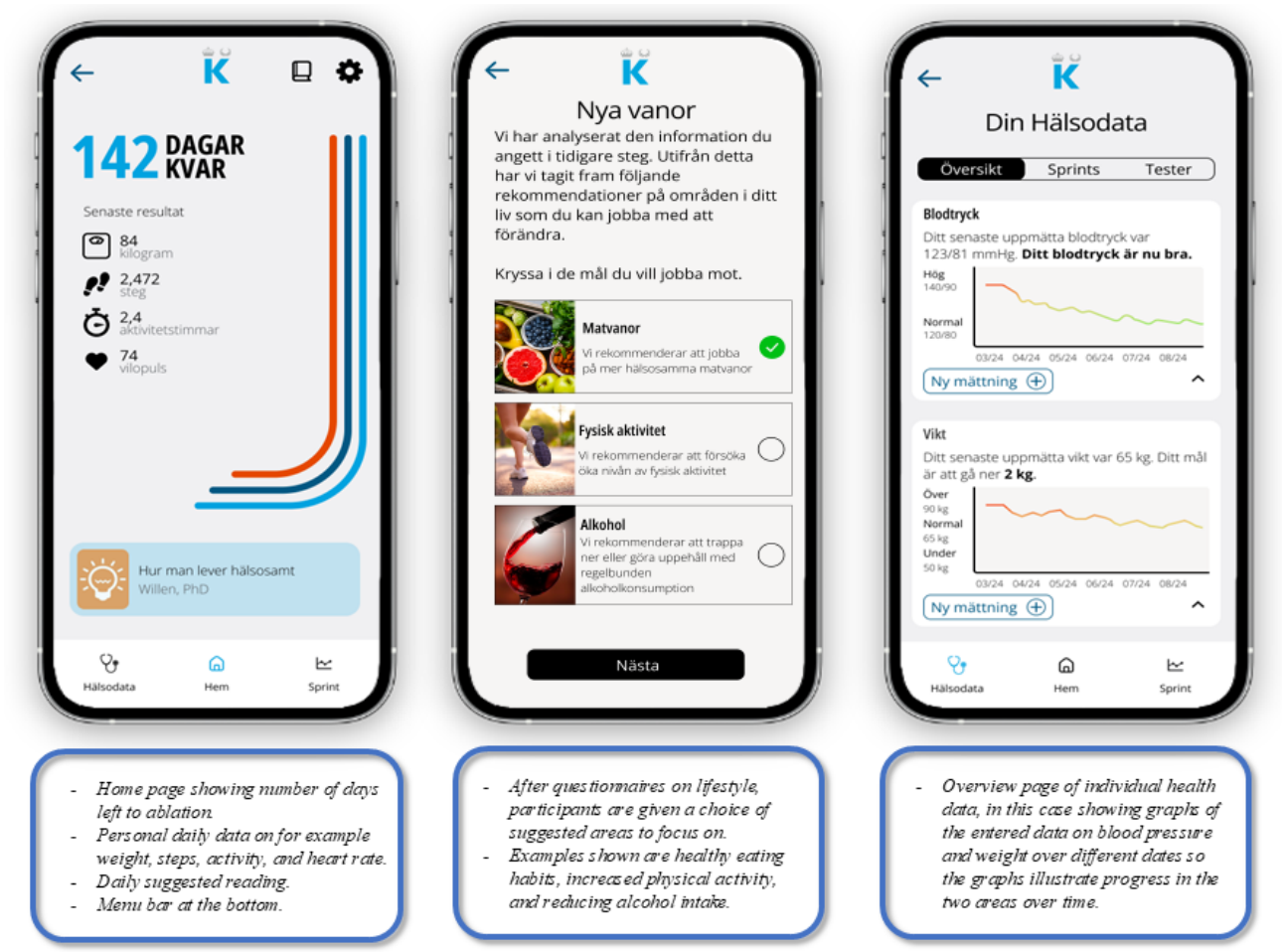


Figure 3. Illustration exemplifying our digital platform.



- *Home page showing number of days left to ablation.*
- *Personal daily data on for example weight, steps, activity, and heart rate.*
- *Daily suggested reading.*
- *Menu bar at the bottom.*

- *After questionnaires on lifestyle, participants are given a choice of suggested areas to focus on.*
- *Examples shown are healthy eating habits, increased physical activity, and reducing alcohol intake.*

- *Overview page of individual health data, in this case showing graphs of the entered data on blood pressure and weight over different dates so the graphs illustrate progress in the two areas over time.*

XSL•FO
RenderX

Study Design

This was a mixed methods usability study including quantitative components to evaluate and rate our co-designed tool, as well as a qualitative component to further explore patients' needs and perceptions and identify areas of improvement. In this study, the initial phase of the platform was tested, including signing up, answering questionnaires, and going through the suggested areas of improvement, and provided recommendations for feedback. First, a widely used quantitative scoring questionnaire for the usability of digital tools was used. Second, a qualitative, descriptive study was conducted with an inductive approach, based on semistructured individual interviews that were analyzed by manifest content analysis [18,19].

Recruitment

Eligibility included patients with (1) a confirmed diagnosis of AF who were (2) admitted to the elective Cardiology ward at Karolinska University Hospital (Stockholm, Sweden), specifically for this condition during the interview period of 2 weeks (March 31, 2025, to April 11, 2025). The sole exclusion criterion was not being able to understand or speak Swedish. Eligible patients were identified among those admitted to the cardiology ward during the study period. They were approached in person by the first author (SK), who is a medical doctor and PhD student, and she had no prior relationship with any of the study participants. All participants first received oral and written information about the study; those who accepted participation were then asked to provide written informed consent. Signed forms were collected before data collection.

Data Collection

Immediately after informed consent, first author SK collected baseline data from patients and medical records on age, sex, weight, BMI, medical conditions, as well as previous and ongoing treatments and medications. After this, SK left the room so that patients could begin the usability test alone with a study-provided smartphone with the prototype on the screen. They were suggested to take 1-2 hours to do this by themselves in their single-patient rooms at the cardiology ward but were able to call on the interviewer sooner if ready with the usability test or to request more time if needed. Most patients took about an hour for the usability test, and none demanded more time than 2 hours.

The testing procedure consisted of a structured usability test. Detailed instructions were given in writing to all participants (for exact translated instructions, refer to Multimedia Appendix 1) and included the following steps: creating a user, filling in questionnaires, choosing one of the recommended target areas suggested by the algorithm, and then going through the given recommendations within their suggested area, including, for example, requested measurements, activities, recipes, etc.

Immediately after testing, patients were first asked to answer a System Usability Scale (SUS) Questionnaire [20] (Multimedia Appendix 2). It is a widely used and well-known scoring system for the usability of digital tools consisting of 10 statements where participants answer each one on a Likert scale from 1 to 5 between "strongly disagree" and "strongly agree" [21]. Results are calculated to a total score between 0 and 100, where a higher score reflects a better perceived usability. Above 68 points is considered acceptable usability [21]. Second, directly after filling in the questionnaire, a semistructured individual interview was conducted by author SK, using a structured interview guide (Multimedia Appendix 3) covering topics of design, function, usability, and user experience.

Data Analysis

Quantitative SUS data were analyzed using descriptive statistics, median (IQR). The interviews were recorded (audio), and a contracted transcription service was used to transcribe all interviews verbatim. The qualitative, descriptive, manifest content analysis with an inductive approach was performed according to Graneheim et al [18], Graneheim and Lundman [19], and Krippendorff [22]. The analysis was inductive, and categories emerged from the interview material. The transcripts were read several times and relevant meaning units were identified, condensed, and coded. Codes with similar content were then grouped into subcategories and broader categories. Examples of meaning units, codes, and categories are shown in Table 1. The analysis was conducted after completion of all interviews and included reading the transcribed interviews several times, identification of initial meaning units, condensation (reduction), labeling with codes, and creation of initial categories and subcategories. Wording was completely emergent from the interviews. Analysis was first done separately by author SK and author ML, who is a specialist nurse in cardiology and PhD and experienced in qualitative methods. The process was then further discussed with the author ES and FB, who are both electrophysiologists in the field of AF. They were presented with both raw and already coded material, and the developing categories and discussions were held to strengthen trustworthiness. Where disagreement, discussions were held until consensus was reached among all authors.

XSL•FO
RenderX

Table 1. Examples of meaning units and codes related categories.

Meaning units	Codes	Categories
…do not complicate it with many colors and arrows…	Simple design	Clear layout
…as easy and clear as possible to that you cannot go wrong …	Easy navigation	Clear layout
…inform without frightening…	Informative	Positive content
…non-binary, other…	Inclusive	Positive content
…it doesn't have to be boring to be healthy…	Motivating	Positive content
…atrial fibrillation, what it is…	Different aspects	Personalized information
… with more for those who are extra interested…	Different levels	Personalized information
…depending on the patient…	Personalized	Individualized medical recommendations
…if one doesn't want to, then one can just skip…	Flexible	Individualized medical recommendations
…one should be able to choose…	Open to individual choice	Individualized medical recommendations

Ethical Considerations

The study protocol adhered to the Declaration of Helsinki and was approved by the Regional Ethical Review Board (Dnr 2025-01228-02). All participants were informed of the premises of the study beforehand, participation was voluntary, and they could at any time choose to discontinue their involvement in the study without having to state a reason for it. Data were transcribed anonymously and stored in a designated locked room at the Cardiology Research Department. A detailed COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist is provided as a supplement (Multimedia Appendix 4).

Results

Participant Characteristics and Interviews

In total, 30 participants were eligible and approached. Out of 30 participants, 22 were included and 8 declined participation with the main reasons stated being lack of time and/or interest. The interviews of the 22 participants lasted, on average, 24.3 (SD 9.7) minutes each and resulted in a total interview time of 535.2 minutes with material for analysis. Baseline characteristics are presented in Table 2.

XSL•FO
RenderX

Table 2. Participant characteristics (N=22).

Characteristics	Participants
Demographics	
Age (y)	
Mean (SD)	67.7 (SD 7.6)
Median (IQR; range)	69 (62.25-72.75; 48-79)
Sex	
Male, n (%)	15 (68.2)
Female, n (%)	7 (31.8)
Persistent, n (%)	8 (36.4)
Paroxysmal, n (%)	14 (63.6)
Previous direct current cardioversions	
Mean (SD)	2.6 (4.2)
Median (IQR; range)	2 (0-3; 0-20)
Previous ablations	
Mean (SD)	1.4 (0.6)
Median (IQR; range)	1 (1-2; 1-3)
Duration of atrial fibrillation (months)	
Mean (SD)	49 (55.3)
Median (IQR; range)	26 (13.3-72.8; 6-216)
Risk factors	
BMI (kg/m²)	
Mean (SD)	27.7 (SD 3.6)
Median (IQR; range)	27.9 (26.2-30.2; 20.8-35.0)
CHADS-VA[a] score	
Mean (SD)	2 (1.8)
Median (IQR; range)	1.5 (1-3; 0-8)
Hypertension, n (%)	17 (77.3)
Diabetes, n (%)	3 (13.6)
Heart failure, n (%)	8 (36.4)
Ejection fraction %	
Mean (SD)	49.8 (6.8)
Median (IQR; range)	50 (50-55; 30-60)
Ischemic heart disease, n (%)	4 (18.2)
Currently smoking, n (%)	0 (0)
Ceased smoking, n (%)	13 (59.1)
Consumes alcohol, n (%)	9 (40.9)
Alcohol standard glasses per week	
Mean (SD)	2.4 (4.2)
Median (IQR; range)	0 (0-1; 0-18)
Treatment, n (%)	
Betablockers	19 (86.4)
Angiotensin-receptor neprilysin inhibitors / Angiotensin-converting-enzyme inhibitors	16 (72.7)
Mineralocorticoid receptor antagonists	7 (31.8)

XSL•FO
RenderX

Characteristics	Participants
Calcium inhibitors	3 (13.6)
Statins	15 (68)
Antiarrhythmics	6 (27.3)
Non-vitamin K oral anticoagulants	21 (95.5)
Warfarin	1 (4.5)

[a]CHADS-VA: risk score including congestive heart failure, hypertension, age, diabetes mellitus, stroke, transient ischemic attack, thromboembolism, and vascular disease.

Quantitative Results – System Usability Scale

The SUS questionnaire resulted in a median score of 78.8 (IQR 63.8-86.9) points, presented in Table 3.

Table 3. Results: System Usability Scale scores by item and total.

System Usability Scale Score	Median (IQR)
I think that I would like to use this system frequently	4 (3.5-5)
I found the system unnecessarily complex	1 (1-2)
I thought the system was easy to use	4 (4-5)
I think that I would need the support of a technical person to be able to use this system	1 (1-2)
I found the various functions in this system were well integrated	4 (3-5)
I thought there was too much inconsistency in this system	2 (1-3)
I would imagine that most people would learn to use this system very quickly	4 (3-5)
I found the system very cumbersome to use	2 (1-2.5)
I felt very confident using the system	4 (3-5)
I needed to learn a lot of things before I could get going with this system	2 (1-3)
Total score	78.8 (63.8-86.9)

Responses were relatively consistent across the individual SUS items, with median scores ranging from 1 (IQR 1-2) to 4 (IQR 4-5) on the 0-5 scale. One participant indicated responses between 2 response options on the 1-5 Likert scale; these responses were recorded as the corresponding midpoint values (eg, 1.5 or 3.5). In exploratory subgroup analyses, the median SUS score was 85.0 (IQR 75.0-95.0) among female participants and 72.5 (IQR 60.0-83.8) among male participants. Participants aged <70 years had a median SUS score of 80.0 (IQR 71.3-86.3), compared with 76.3 (IQR 49.4-86.9) among those aged ≥70 years. These analyses were descriptive, and no statistical testing was performed. The 2 participants with the highest SUS scores gave all items the best possible rating. However, the 2 participants with the lowest SUS scores also expressed during the subsequent qualitative interviews that they found the platform unnecessarily complex and cumbersome and that they would need to learn more before using it.

Qualitative Results—Categories

Overview

Some patients made initial comments regarding the developed digital tool:

> *…It is so good that you are working on this type of tool.*
>
> *…I would absolutely want to use this.*
>
> *…This app, I think it can be meaningful, because the wait is long and you have a lot of questions.*

The qualitative analysis of interviews revealed four main categories and 10 subcategories related to participants' preferences and perceptions, and these are summarized in Textbox 1.

XSL•FO
RenderX

Textbox 1. Results: categories and subcategories.

Clear layout

- Simple design
- Easy navigation

Positive content

- Informative
- Inclusive
- Motivating

Personalized information

- Different aspects
- Different levels

Individualized medical recommendations

- Personalized
- Flexible
- Individual choice

First, they value a *clear layout* with *simple design* and *easy navigation*. Second, they appreciate *positive content*, which is *informative*, *inclusive*, and *motivating*. Third, they request *personalized information* on *different aspects* and provided on *different levels*. Fourth, they request *individualized medical recommendations* that are *personalized* and *flexible* but *open to individual choice*.

Clear Layout

The first category, a *clear layout*, was consistently emphasized throughout the interviews and included the subcategories *simple design* and *easy navigation*.

Participants expressed that it was a digital tool with a *simple design*, which is more likely to be understood and used by users of different ages and different backgrounds. One participant noted that:

> *I like that it is simple and not too many different colors, because often diagrams and such can become too much with different colors and a little difficult to read. I like this simplicity, and the text is a good size.*

Frequently brought up was that using different colors, texts, and figures can become cluttered and be frustrating to users. One individual summarized that:

> *It is good, it is simple, it is easy to read. Do not complicate it with many colors and arrows, it is not necessary. I think that most appreciate simplicity.*

Participants agreed that a simple design, without unnecessary complexity, is essential in a digital tool they would see themselves using over time.

One issue repeatedly raised by participants was the importance of *easy navigation* as an important determinant to further improve satisfaction with the digital tool. As one individual suggested:

> *You should be able to go back in an easy way, a go-back button, you should try to make this intuitive.*

It was clear amongst participants that ease of navigation is essential to facilitate and avoid unnecessary frustration for users.

Positive Content

The second category, the desire for *positive content*, recurred throughout interviews and included the subcategories *informative*, *inclusive*, and *motivating* content.

Central for participants was that the content provided should focus on being *informative*. It was agreed that informative content is appreciated, but that it is not received well if presented as too demanding or alarming. As 1 participant pointed out about our information:

> *It is an important balance, so that you do not scare people. Inform without frightening.*

Information also needs to be adapted to the intended users, as 1 participant mentioned:

> *Not in medical language.*

This highlights how the overall tone of the content needs to be informative, yet accessible to be received positively by users.

Participants also valued our content that was perceived as *inclusive* to individuals regardless of age, gender, economic, and social situation. One participant noted and appreciated:

> *Do you have non-binary or other? Yes, you do.*

Some brought up that the recommendations provided on healthy eating habits need to also be inclusive of individuals with different preferences and lifestyles, for example, regarding eating habits to include also:

> *What is gluten-free and vegetarian.*

The findings underscore the importance of framing information in a way respecting diversity.

XSL•FO
RenderX

It was agreed that content needs to be *motivating* to the intended users. Many mentioned how motivating content is perceived as more respectful, and that motivation is much more likely to have an effect as compared with shaming or lecturing users. One participant emphasized that it is essential to encourage and inspire users and convey:

> *That it should be joyful as well, it does not have to be boring to be healthy.*

This demonstrates how essential it is to present information in a way that is motivating and respectful to users to empower them.

Personalized Information

The third category, *personalized information*, emerged and included the subcategories of personalizing information on *different aspects* and providing it on *different levels*.

Participants expressed a desire for personalized information on *different aspects* tailored to their individual needs. Some participants wanted basic information on, for example:

> *What it is that you do during an ablation. And concerning atrial fibrillation, what it is.*

Others requested in-depth information on associated topics such as causes, risk factors, genetics, mechanisms, medication, interactions, exercise, psychological effects, and effects on sexual function. Some suggested having links or icons leading to more detailed information.

This highlights the need for a wide variety of information on different aspects but personalized to support both users with limited prior knowledge and those seeking more detailed understanding.

The importance of information at *different levels* was another need that came up consistently in all the interviews. Some participants felt that they wanted only general information, while others felt that they had already had a lot of information on AF.

Some participants wanted much more and in-depth information provided, such as:

> *You could maybe press a button for references, for the best studies and meta-analyses.*

Thus, having access to information on different levels is essential.

Individualized Medical Recommendations

The fourth central category was the desire for *individualized medical recommendations*, which included the subcategories of *personalized* and *flexible* but *open to individual choice*.

An important feature requested by participants was that medical recommendations should be *personalized*. Participants brought up the heterogeneity in the group of individuals with AF, with a large range of knowledge, engagement, motivation, and different lifestyles. It was mentioned how frustrating it is to get recommendations to stop smoking and reduce alcohol consumption to individuals who stopped smoking and drinking years ago. Some felt frustrated getting recommendations on physical activity that they could not perform due to other diagnoses and disabilities. There was general agreement that the recommendations provided should be personalized.

Participants also expressed a desire for *flexibility* in the medical recommendations provided. There should be no shame in some not being able to perform a certain recommendation, and as 1 participant suggested, they should be able to choose digitally from recommendations they felt were appropriate, such as:

> *Today I choose… click… a 20-minute walk.*

This underlines the importance of being flexible in recommendations so as not to lose user interest and involvement.

Also central in the interviews was the importance of recommendations being open to *individual choice*. Most participants requested some level of individual choice in different terms, and many pointed out how each patient is an individual with their own preferences:

> *People react to information in different ways, and what one may think is good, another might feel this is not something I want.*

They emphasized the desire not to feel mastered but to feel respected as adults and to feel like they had a choice and would be supported regardless. This reflects a preference for personalized, nonprescriptive guidance that accommodates individual needs and circumstances.

Discussion

Principal Results

In this study, participants reported overall satisfaction with our developed digital tool. The SUS score is a validated measure of perceived usability [23]. Our digital tool achieved a median SUS score of 78.8 (IQR 63.8-86.9). The fact that the 2 lowest scorers found it to be unnecessarily complex and cumbersome, and that they would need to learn more before using it, aligned with the findings in the qualitative analysis, which revealed that persons with AF particularly valued a *clear layout, positive content, personalized information*, and *individualized medical recommendations*. The quantitative and qualitative findings complemented one another, with the SUS indicating overall acceptable usability while the interviews identified specific design features that influenced participants' experiences and highlighted priorities for further development.

The layout emerged as central to determining our participants' satisfaction with the digital tool. Regardless of the content, people may be less likely to use a digital tool if considered too difficult and complex to use. For example, a previous meta-analysis on 114 digital health apps found that apps with a higher SUS score were more likely to be adopted and used consistently [21]. The same meta-analysis aimed to investigate whether the SUS score could be used for assessing usability and found it to be reliable and that a mean score above 68 was acceptable [21]. In our interviews, participants appreciated simple layouts, and this was strengthened by the overall median SUS score of 78.8 (IQR 63.8-86.9), which confirmed the acceptability to users. Furthermore, it is important to consider the demographics of the intended users, who, in the case of individuals with AF, tend to be slightly older; in our study, the

XSL•FO
RenderX

age range was 48-79 years. A systematic review of 132 studies, for a population older than 60 years of age, demonstrated that simplified interfaces, clear navigation, and readable fonts were especially important to this group [24]. Because of the age range of users, it is essential to clarify and simplify layout, design, and navigation in health apps.

Maintaining an informative, inclusive, and motivating tone throughout the content provided was crucial to the participants in this study. This is in line with findings in our previous study where participants requested information conveyed in an appropriate manner without being condescending or moralizing [13]. The participants valued communication that was inspiring and motivating [13]. This theme has been recognized by other authors, for example, in a review of digital weight management interventions where regular and encouraging feedback was appreciated by participants requesting that the tone should not be too generic or impersonal [25]. In a review on lifestyle management intervention, it was identified that customized and personalized feedback was one of the features responsible for successful implementation of digital interventions [26], highlighting the need for positive and motivating content.

Personalized information was central to our participants, which relates to the concept of person-centered care, where the patient is an active part in jointly planning individualized care [14]. Similarly, in our prior study, participants specifically desired person-centered and customized information [13]. In both studies, they wished for information about many different aspects of AF, such as healthy lifestyle, eating habits, physical activity, alcohol consumption habits, stress, and medication [13]. Participants in both studies also requested information to be readily available but at different levels, so as to fit each individual's backgrounds, prior knowledge, and preferences [13]. Another study evaluated the "INFORM-AF" program, a co-designed digital AF educational program developed by Ferguson et al and delivered to participants' smartphones via the QStream platform, and they reported an excellent SUS score of 84 [27,28]. This was higher than in our study, demonstrating that the present digital platform has room for improvement in terms of usability.

Regarding the medical recommendations given, personalization and flexibility emerged as central to users. In cardiovascular medicine, there is a need for a shift toward multidisciplinary, coordinated, and tailored solutions focusing on shared decision-making [10]. To achieve person-centered care, an essential component is "respecting the patient's information needs and preferences" [29], which could mean integrating tailored information with shared decision-making. Tailored information was also requested by participants in our previous study [13]. Furthermore, a review of digital weight management interventions concluded that personalization is an essential component for the success of the intervention [25]. Additionally, a systematic review of the literature on mobile health (mHealth) apps for managing AF found that an app for tailored patient education rated positively on all aspects [30]. The mAFA app, developed by Guo et al [31], includes clinical decision-support tools with educational materials, patient involvement strategies with self-protocols and structured follow-up [31]. It was found to be easy, user-friendly, and helpful, and resulted in increased knowledge of AF, improved drug adherence and satisfaction, and quality of life scores [31]. It was also associated with a lower rate of the composite outcome (ischemic stroke, systemic thromboembolism, death, and rehospitalization), and lower rates of rehospitalization [32]. These findings highlight the potential of digital health intervention to improve outcomes, which warrants further research.

The digital platform used in this trial has been co-designed with a group of patients with AF who were initially interviewed for requests and requirements of a digital platform for lifestyle and risk factor management. [13] The same individuals were then asked to test the platform further and provide their feedback on further development. In our current trial, we have included patients who were not part of the original development to test the digital platform. From this trial, we will set out to further develop the platform, focusing particularly on navigation, simplicity, layout, and personalization. The co-design process shaped the platform by turning patient feedback into design changes, especially a simpler interface, clearer navigation, and less visual clutter. Patients also pushed for more control and less directive advice, which led to a choice-based system with layered information at different levels. These decisions likely improved usability and acceptability in the findings by making the tool easier to use and more aligned with how patients want to engage. Many recommendations for managing comorbidities and healthy lifestyle habits for patients with AF are similar to those for other patient groups. However, this study highlights that the unpredictable nature of AF, the wide range of functional impact, and its association with psychological and physical stress and reduced quality of life require flexible, individualized support for patients with AF.

Trustworthiness of a qualitative study can be evaluated through credibility, transferability, dependability, confirmability, and authenticity [18,19]. This study demonstrates high credibility, as the participants had relevant experience with AF and risk factors. Transferability is good as baseline characteristics are representative of individuals with AF undergoing intervention. There were no current smokers in the study population, which is consistent with the low prevalence of smoking in Sweden. However, this may limit the generalizability of the findings to individuals who smoke. To maintain dependability, confirmability, and authenticity, coding was done independently, and then ongoing discussions were held throughout the analysis process until consensus was reached. In this study, we support the conclusion by providing citations supporting the results and analysis.

Limitations

Although the results are likely generalizable to persons with AF as reflected by the baseline characteristics (Table 2), a main limitation is that all participants were recruited from a single Swedish tertiary hospital with a large AF ablation service, which may reduce generalizability. It is worth noting that they were hospitalized specifically for AF, and this might create a selection bias of patients who have a slightly better understanding of AF in general. Although the SUS score is widely used and recommended for scoring perceived usability [23], it has inherent limitations, as it yields a single overall usability score

XSL•FO
RenderX

without identifying domain-specific problems or offering qualitative insights, since it does not allow participants to elaborate on their responses. To address this, we also performed a qualitative analysis using a structured interview guide designed with open-ended questions to encourage participants to speak freely, to allow for a person-centered exploration of themes, perceptions, and meaning, and thereby reduce the risk of bias. Furthermore, the results were independently coded by 2 of the authors (SK and ML), and any disagreement was discussed multidisciplinary until consensus was reached among all authors. Nevertheless, no repeat interviews, field notes, or member checking were performed, and some risk of subjective interpretation and potential bias remains and cannot be fully excluded

As generative AI becomes more widely available, disease-specific platforms might complement these tools by providing individualized, guideline-based support integrated with clinical care. These findings will guide further development of digital tools for better risk factor management for individuals with AF undergoing intervention. Future testing is planned to be performed at home in a real-world setting, and we also aim to include other AF groups than those hospitalized. In addition, we will also aim to make sure we have a representative sample of older patients, and individuals with lower digital literacy. Future research should assess the effectiveness of digital tools in promoting better management of risk factors in AF, improved outcomes after AF intervention, and better utilization of resources.

Conclusions

Persons with AF desire support in managing modifiable risk factors and find it likely that a digital tool could be used to facilitate this. In our digital tool for risk factor management, participants valued specifically a *clear layout*, *positive content*, *personalized information*, and *individualized medical recommendations*. This feedback will be used to guide development and guide work to develop digitally aided and person-centered management of modifiable risk factors for persons with AF, hopefully improving outcomes and reducing the cost of care.

Acknowledgments

We are grateful to all the patients who participated in the study and provided valuable insights to further develop care for patients with AF.

Data Availability

Not available due to ethical regulations.

Funding

This work was partly funded by Innovationsfonden Region Stockholm, a research grant from the Swedish Research Council Vetenskapsrådet (DNR 2022-01466). ES is funded by a grant from Region Stockholm (Clinical Researcher).

Generative AI Use

The authors declare that generative AI was not used in the preparation of this manuscript.

Authors' Contributions

Conceptualization: SK, FB, ML, ES (equal)
Data curation: SK, ML (equal)
Formal analysis: SK, ML (lead), FB, ES (supporting)
Funding acquisition: ES
Investigation: SK (lead), ML (supporting)
Methodology: ML (lead), SK (supporting)
Project administration: SK (lead), FB, ML, ES (supporting)
Resources: FB, ML, ES (equal)
Supervision: FB, ML, ES (equal)
Validation: ML (lead), FB, ES (supporting)
Visualization: SK (lead), FB, ML, ES (supporting)
Writing - original draft: SK (lead), FB, ML, ES (supporting)
Writing - review & editing: SK (lead), FB, ML, ES (supporting)

Conflicts of Interest

SK has none to declare. FB has received speaker/consultancy fees from Biotronik, Medtronic, Abbott, Boston Scientific, Boehringer, Novartis, Pfizer, Orion. ML has received speaker/consultancy fees from Pfizer, Eli Lilly, Novo Nordisk, AbbVie, Sanofi, Bayer MSD, UCB Pharma. ES has received speaker/consultancy fees from Bayer, Bristol-Myers Squibb-Pfizer, Boehringer- Ingelheim, Johnson & Johnson, Merck Sharp & Dohme.

XSL•FO
RenderX

Multimedia Appendix 1
Usability test.
[DOCX File , 19 KB - humanfactors_v13i1e95810_app1.docx]

Multimedia Appendix 2
System Usability Scale.
[DOCX File , 23 KB - humanfactors_v13i1e95810_app2.docx]

Multimedia Appendix 3
Interview guide.
[DOCX File , 22 KB - humanfactors_v13i1e95810_app3.docx]

Multimedia Appendix 4
COREQ (Consolidated Criteria for Reporting Qualitative Research).
[DOCX File , 21 KB - humanfactors_v13i1e95810_app4.docx]

References

1. Björck S, Palaszewski B, Friberg L, Bergfeldt L. Atrial fibrillation, stroke risk, and warfarin therapy revisited: a population-based study. Stroke 2013 Nov;44(11):3103-3108. [doi: 10.1161/STROKEAHA.113.002329] [Medline: 23982711]
2. Van Gelder IC, Rienstra M, Bunting KV, Casado-Arroyo R, Caso V, Crijns HJGM, ESC Scientific Document Group. 2024 ESC Guidelines for the management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). Eur Heart J 2024 Sep 29;45(36):3314-3414 [FREE Full text] [doi: 10.1093/eurheartj/ehae176] [Medline: 39210723]
3. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, ESC Scientific Document Group. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Eur Heart J 2021 Mar 01;42(5):373-498 [FREE Full text] [doi: 10.1093/eurheartj/ehaa612] [Medline: 32860505]
4. Huxley RR, Lopez FL, Folsom AR, Agarwal SK, Loehr LR, Soliman EZ, et al. Absolute and attributable risks of atrial fibrillation in relation to optimal and borderline risk factors: the Atherosclerosis Risk in Communities (ARIC) study. Circulation 2011 Apr 12;123(14):1501-1508 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.110.009035] [Medline: 21444879]
5. Okin PM, Hille DA, Larstorp ACK, Wachtell K, Kjeldsen SE, Dahlöf B, et al. Effect of lower on-treatment systolic blood pressure on the risk of atrial fibrillation in hypertensive patients. Hypertension 2015 Aug;66(2):368-373. [doi: 10.1161/HYPERTENSIONAHA.115.05728] [Medline: 26056336]
6. Wong CX, Sullivan T, Sun MT, Mahajan R, Pathak RK, Middeldorp M, et al. Obesity and the Risk of Incident, Post-Operative, and Post-Ablation Atrial Fibrillation: A Meta-Analysis of 626,603 Individuals in 51 Studies. JACC Clin Electrophysiol 2015 Jun;1(3):139-152 [FREE Full text] [doi: 10.1016/j.jacep.2015.04.004] [Medline: 29759357]
7. Pathak RK, Middeldorp ME, Meredith M, Mehta AB, Mahajan R, Wong CX, et al. Long-Term Effect of Goal-Directed Weight Management in an Atrial Fibrillation Cohort: A Long-Term Follow-Up Study (LEGACY). J Am Coll Cardiol 2015 May 26;65(20):2159-2169 [FREE Full text] [doi: 10.1016/j.jacc.2015.03.002] [Medline: 25792361]
8. Rienstra M, Hobbelt A, Alings M, Tijssen J, Smit M, Brügemann J, RACE 3 Investigators. Targeted therapy of underlying conditions improves sinus rhythm maintenance in patients with persistent atrial fibrillation: results of the RACE 3 trial. Eur Heart J 2018 Aug 21;39(32):2987-2996. [doi: 10.1093/eurheartj/ehx739] [Medline: 29401239]
9. Pathak RK, Middeldorp ME, Lau DH, Mehta AB, Mahajan R, Twomey D, et al. Aggressive risk factor reduction study for atrial fibrillation and implications for the outcome of ablation: the ARREST-AF cohort study. J Am Coll Cardiol 2014 Dec 02;64(21):2222-2231 [FREE Full text] [doi: 10.1016/j.jacc.2014.09.028] [Medline: 25456757]
10. Ski CF, Cartledge S, Foldager D, Thompson DR, Fredericks S, Ekman I, et al. Integrated care in cardiovascular disease: a statement of the Association of Cardiovascular Nursing and Allied Professions of the European Society of Cardiology. Eur J Cardiovasc Nurs 2023 Jul 19;22(5):e39-e46 [FREE Full text] [doi: 10.1093/eurjcn/zvad009] [Medline: 36617217]
11. Qi Y, Mohamad E, Azlan AA, Zhang C, Ma Y, Wu A. Digital Health Solutions for Cardiovascular Disease Prevention: Systematic Review. J Med Internet Res 2025 Jan 23;27:e64981 [FREE Full text] [doi: 10.2196/64981] [Medline: 39847411]
12. Eaton CK, McWilliams E, Yablon D, Kesim I, Ge R, Mirus K, et al. Cross-Cutting mHealth Behavior Change Techniques to Support Treatment Adherence and Self-Management of Complex Medical Conditions: Systematic Review. JMIR Mhealth Uhealth 2024 May 01;12:e49024 [FREE Full text] [doi: 10.2196/49024] [Medline: 38717433]

XSL•FO
RenderX

13. Klavebäck S, Svennberg E, Nymark C, Braunschweig F, Lidin M. Management of modifiable risk factors and comorbidities in atrial fibrillation: suggestions for improvement from a patient perspective. Eur J Cardiovasc Nurs 2024 Mar 12;23(2):169-175 [FREE Full text] [doi: 10.1093/eurjcn/zvad055] [Medline: 37249076]
14. University of Gothenburg, Centre for Person-centred Care (GPCC). About person-centred care. 2026. URL: https://www.gu.se/en/gpcc/about-person-centred-care [accessed 2026-09-23]
15. Ekman I, Swedberg K, Taft C, Lindseth A, Norberg A, Brink E, et al. Person-centered care--ready for prime time. Eur J Cardiovasc Nurs 2011 Dec;10(4):248-251 [FREE Full text] [doi: 10.1016/j.ejcnurse.2011.06.008] [Medline: 21764386]
16. Olsson LE, Jakobsson Ung E, Swedberg K, Ekman I. Efficacy of person-centred care as an intervention in controlled trials - a systematic review. J Clin Nurs 2013 Mar;22(3-4):456-465. [doi: 10.1111/jocn.12039] [Medline: 23231540]
17. Jennings C, Astin F. A multidisciplinary approach to prevention. Eur J Prev Cardiol 2017 Jun;24(3_suppl):77-87 [FREE Full text] [doi: 10.1177/2047487317709118] [Medline: 28618913]
18. Graneheim UH, Lindgren B, Lundman B. Methodological challenges in qualitative content analysis: A discussion paper. Nurse Educ Today 2017 Sep;56:29-34. [doi: 10.1016/j.nedt.2017.06.002] [Medline: 28651100]
19. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Mar;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
20. Brooke J. SUS: a quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester B, McClelland IL, editors. Usability Evaluation in Industry. London: Taylor & Francis; 1996:189-194.
21. Hyzy M, Bond R, Mulvenna M, Bai L, Dix A, Leigh S, et al. System Usability Scale Benchmarking for Digital Health Apps: Meta-analysis. JMIR Mhealth Uhealth 2022 Aug 18;10(8):e37290 [FREE Full text] [doi: 10.2196/37290] [Medline: 35980732]
22. Krippendorff K. Content Analysis - An Introduction to Its Methodology. Thousands Oak,CA: Sage Publication; 2013.
23. Lewis J. The System Usability Scale: Past, Present, and Future. International Journal of Human–Computer Interaction 2018 Mar 30;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
24. Amouzadeh E, Dianat I, Faradmal J, Babamiri M. Optimizing mobile app design for older adults: systematic review of age-friendly design. Aging Clin Exp Res 2025 Aug 14;37(1):248. [doi: 10.1007/s40520-025-03157-7] [Medline: 40804492]
25. Sharpe EE, Karasouli E, Meyer C. Examining Factors of Engagement With Digital Interventions for Weight Management: Rapid Review. JMIR Res Protoc 2017 Oct 23;6(10):e205 [FREE Full text] [doi: 10.2196/resprot.6059] [Medline: 29061557]
26. Chatterjee A, Prinz A, Gerdes M, Martinez S. Digital Interventions on Healthy Lifestyle Management: Systematic Review. J Med Internet Res 2021 Nov 17;23(11):e26931 [FREE Full text] [doi: 10.2196/26931] [Medline: 34787575]
27. Ferguson C, Allida S, William S, Downie A, Wilson T, John R, et al. INFORM-AF II-Phase 1 pilot randomised controlled trial evaluating the effect of an atrial fibrillation digital education programme: study protocol. J Clin Nurs 2024 Oct;33(10):4138-4147. [doi: 10.1111/jocn.17343] [Medline: 38923162]
28. Ferguson C, Shaikh F, Downie A, Wilson T, Griffin K, Allida S, et al. A pilot study of a digital education program (INFORM-AF) for patients living with atrial fibrillation. BMC Digit Health 2024 Dec 05;2(1). [doi: 10.1186/s44247-024-00137-1]
29. Scholl I, Zill JM, Härter M, Dirmaier J. An integrative model of patient-centeredness - a systematic review and concept analysis. PLoS One 2014;9(9):e107828 [FREE Full text] [doi: 10.1371/journal.pone.0107828] [Medline: 25229640]
30. Lane DA, McMahon N, Gibson J, Weldon JC, Farkowski MM, Lenarczyk R, et al. Mobile health applications for managing atrial fibrillation for healthcare professionals and patients: a systematic review. Europace 2020 Aug 27:1567-1578 [FREE Full text] [doi: 10.1093/europace/euaa269] [Medline: 32853369]
31. Guo Y, Chen Y, Lane DA, Liu L, Wang Y, Lip GYH. Mobile Health Technology for Atrial Fibrillation Management Integrating Decision Support, Education, and Patient Involvement: mAF App Trial. Am J Med 2017 Dec;130(12):1388-1396.e6 [FREE Full text] [doi: 10.1016/j.amjmed.2017.07.003] [Medline: 28847546]
32. Guo Y, Lane DA, Wang L, Zhang H, Wang H, Zhang W, mAF-App II Trial Investigators. Mobile Health Technology to Improve Care for Patients With Atrial Fibrillation. J Am Coll Cardiol 2020 Apr 07;75(13):1523-1534 [FREE Full text] [doi: 10.1016/j.jacc.2020.01.052] [Medline: 32241367]

Abbreviations

AF: atrial fibrillation
COREQ: Consolidated Criteria for Reporting Qualitative Research
mHealth: mobile health
SUS: System Usability Scale

XSL•FO
RenderX

Edited by S Law; submitted 23.Mar.2026; peer-reviewed by JY Chyou, EL Sandberg, JY Lee, C Arum; comments to author 19.Jun.2026; revised version received 21.Sep.2026; accepted 21.Sep.2026; published 22.Oct.2026.

Please cite as:
Klavebäck S, Braunschweig F, Lidin M, Svennberg E
Co-Designing a Digital Platform for the Management of Modifiable Risk Factors and Comorbidities in Patients With Atrial Fibrillation: Mixed Methods Usability Study
JMIR Hum Factors 2026;13:e95810
URL: https://humanfactors.jmir.org/2026/1/e95810
doi:10.2196/95810
PMID:

©Sofia Klavebäck, Frieder Braunschweig, Matthias Lidin, Emma Svennberg. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 22.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Engagement-Supportive Mechanisms in Stress-Focused Digital Mental Health Interventions With Adaptive, Conversational, and Context-Aware Systems: Narrative Review

Munesu Maminimini[1,2*], BSc, MSc; Sofia Asteria Piffer[2*], BSc; Silvia Gabrielli[1,2*], MSc, PhD

[1]Department of General Psychology, University of Padua, Via Luigi Luzzatti, 4, Padova, Veneto, Italy

[2]Center for Digital Health & Wellbeing, Fondazione Bruno Kessler, Trento, Trentino-Alto Adige, Italy

[*]all authors contributed equally

Corresponding Author:
Munesu Maminimini, BSc, MSc
Department of General Psychology, University of Padua, Via Luigi Luzzatti, 4, Padova, Veneto, Italy

Abstract

Background: Stress is a widespread public health concern associated with cardiovascular disease, depression, anxiety, and reduced quality of life. Digital mental health interventions (DMHIs), delivered through smartphones, web platforms, and wearables, offer scalable support for stress management but often struggle to sustain user engagement. Evidence describing how adaptive, conversational, and context-aware mechanisms function together within stress-focused DMHIs remains fragmented.

Objective: This review aims to synthesize evidence on engagement-supportive mechanisms in stress-focused DMHIs, with attention to personalization; biometric feedback; just-in-time adaptive interventions (JITAIs); gamification; AI, including large language models (LLMs); and methodological approaches to studying engagement. It also examines methodological, motivational, and experiential dimensions of engagement.

Methods: This narrative review was informed by a structured search of PubMed, IEEE Xplore, and Scopus for peer-reviewed studies published between January 2018 and April 2025, followed by an evidence update through August 2026. Eligible studies described stress-focused DMHIs or related affective support interventions reporting at least 1 engagement-supportive mechanism. Randomized controlled trials, qualitative investigations, simulation or modeling studies, and implementation studies were included. Data were extracted using a structured form capturing study characteristics, engagement mechanisms, outcomes, and alignment with the review domains. Descriptive mapping summarized study types and mechanism frequencies, while a narrative synthesis integrated behavioral, experiential, methodological, and theory-informed findings.

Results: Thirty-two studies met the inclusion criteria. Randomized or microrandomized controlled trials were the most common design (n=17), followed by qualitative investigations (n=6), design or implementation studies (n=3), mixed methods studies (n=3), simulation or modeling studies (n=2), and a quasi-experimental study (n=1). Dynamic adaptation was identified in 14 studies, conversational support in 10, static tailoring in 9, social or co-design mechanisms in 9, context-aware prompting in 7, gamified reinforcement in 5, biometric-responsive adaptation in 4, and narrative immersion in 2. These mechanisms frequently overlapped within layered adaptive architectures rather than functioning as isolated engagement features. AI-enabled systems included conversational agents, LLM-supported systems, recommender systems, reinforcement learning, and computational modeling approaches. Outcomes included behavioral indicators such as adherence, retention, and prompt receptivity, alongside experiential indicators such as perceived relevance, trust, emotional support, and immersion. Evidence strength varied substantially; many AI-enabled systems contributed feasibility, acceptability, or optimization-oriented evidence rather than definitive comparative effectiveness evidence.

Conclusions: Stress-focused DMHIs are increasingly adopting adaptive, conversational, and context-aware engagement-supportive mechanisms, but the evidence base remains heterogeneous and methodologically uneven. Dynamic adaptation, conversational support, and static tailoring were most frequently represented, whereas biometric-responsive adaptation, narrative immersion, and deeply integrated gamified systems were less common. AI-enabled systems may support contextual responsiveness and personalized interaction but require transparent adaptive logic, privacy safeguards, and human-centered evaluation. Future work should prioritize longitudinal, mixed methods research capable of distinguishing behavioral engagement, experiential engagement, feasibility, acceptability, and sustained stress-related outcomes.

(JMIR Hum Factors 2026;13:e90855) doi:10.2196/90855

XSL•FO
RenderX

KEYWORDS

digital mental health interventions; stress management; user engagement; just-in-time adaptive interventions; conversational agents; artificial intelligence

Introduction

Stress is a pervasive psychological and physiological response to perceived internal or external demands, and when it is prolonged or maladaptively regulated, it may contribute to psychological distress and stress-related disorders including anxiety, depression, sleep disturbance, and poorer quality of life [1,2]. Digital mental health interventions (DMHIs), delivered via mobile apps, web platforms, and wearables, can provide scalable, low-threshold support through psychoeducation, monitoring, and behavioral strategies. Although meta-analyses and recent randomized controlled trial (RCT) evidence indicate that smartphone-based and app-based interventions can reduce stress or anxiety symptoms [1,3], their long-term impact and effectiveness are often limited by poor engagement and high dropout rates [4].

Engagement in DMHIs is a multidimensional construct encompassing behavioral, cognitive, emotional, experiential, and social dimensions [4-7]. Behavioral engagement is commonly reflected through indicators such as uptake, adherence, frequency and duration of use, and retention and completion rates, while cognitive and experiential engagement relate more closely to perceived relevance, motivational investment, emotional connection, immersion, and trust in the intervention experience [2,5]. Social or participatory dimensions may additionally emerge through co-design processes, peer interaction, clinician support, or blended care approaches [8,9]. Engagement-supportive mechanisms refer to the intervention design or delivery processes intended to influence 1 or more of these dimensions.

The “law of attrition” describes the declining DMHI use and loss to follow-up over time [4]. Barriers including low perceived relevance, burdensome data entry, and inadequate feedback may contribute to disengagement [2]. Reviews have consequently emphasized user-centered design and adaptive technologies as possible means of supporting continued engagement [5,10]. To improve user engagement and retention, recent DMHI research has explored personalization, biometric feedback, gamification, just-in-time adaptive interventions (JITAIs), and AI-enabled conversational or recommendation systems.

Personalization may use demographic characteristics, preferences, psychological profiles, or behavioral data to increase perceived relevance and autonomy [11-14]. Biometric and ecological momentary assessment (EMA) data can inform JITAIs that seek to provide support when it is needed and users are receptive [2,15,16]. Gamification uses elements such as challenges, feedback, points, or narratives to encourage persistence, although poorly aligned elements may be ineffective or counterproductive [17-20]. Conversational and other AI-enabled systems may provide responsive psychoeducation and support at scale while raising concerns about safety, privacy, transparency, and perceived authenticity [21-27]. Together, these approaches address the relevance, timing, and experience of digital support and may overlap within the same intervention.

Although individual mechanisms have been examined, evidence describing how they are implemented, combined, and assessed in stress-focused DMHIs remains fragmented. Previous work has mapped broad engagement barriers [2,6,28], examined personalization, or JITAI elements [5,29-31] and examined chatbots, gamification, and recommender systems [11,17,18,21,32]. Recent syntheses continue to identify inconsistent engagement definitions, heterogeneous measures, and limited longitudinal evidence [33-36]. What remains less clear is how several mechanisms operate together; how static tailoring differs from dynamic adaptivity; and what controlled, qualitative, implementation, and computational studies each contribute to understanding engagement.

Accordingly, this narrative review aims to synthesize and critically evaluate how stress-focused DMHIs conceptualize and operationalize engagement; how personalization, biometric feedback, JITAI timing, gamification, and AI-enabled approaches are used to support it; and how different methodological approaches contribute to understanding their implementation and effects.

Methods

Design

This narrative review was informed by structured systematic search, screening, and extraction procedures. A narrative synthesis was selected because the evidence spans clinical trials, qualitative inquiry, implementation research, and computational studies and because the review examines how engagement mechanisms are implemented and interpreted rather than estimating a pooled effect. A completed PRISMA-P (Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols) checklist is provided in Checklist 1.

Search Strategy

The original search covered PubMed, IEEE Xplore, and Scopus from January 2018 through April 2025. Search concepts combined digital mental health, stress or distress, engagement, and adaptive mechanisms, including personalization, biometric feedback, EMA, JITAIs, gamification, conversational agents, LLMs, and reinforcement learning. Well-being terminology was also considered because preventative DMHIs may describe emotional support or coping without using stress reduction as their primary label. A further update covering May 2025 through August 2026 was conducted using the same concept blocks for the focused database searches and a supplementary journal search of recent peer-reviewed work. The review domains and illustrative search concepts are summarized in Table 1, and the full database-specific strings, dates, and the supplementary search description are reported in Multimedia Appendix 1.

XSL•FO
RenderX

Table . Review domains and illustrative search concepts.

Review domain	Search query
Engagement-supportive mechanisms and behavioral or experiential engagement outcomes	(“digital health” OR “mHealth” OR “eHealth”) AND (“engagement” OR “adherence” OR “retention”) AND (“stress” OR “chronic diseases”)
Personalization, biometric feedback, JITAI[a] timing, and gamification	(“personalisation” OR “tailoring” OR “adaptive intervention” OR “biometric data” OR “heart rate” OR “EMA” OR “ecological momentary assessment” OR “JITAI” OR “just-in-time” OR “gamification”) AND (“engagement” OR “adherence” OR “retention”) AND (“digital health” OR “mHealth” OR “eHealth”) AND (“stress”)
AI-enabled adaptive and engagement-oriented personalization	(“artificial intelligence” OR “machine learning” OR “adaptive systems” OR “LLM” OR “large language models”) AND (“digital health” OR “personalised intervention”)
Methodological contributions to engagement-informed DMHI[b] design	(“self-research” OR “autoethnography”) AND (“digital health” OR “engagement”)

[a]JITAI: just-in-time adaptive intervention.

[b]DMHI: digital mental health intervention.

Eligibility Criteria

Studies were included if they (1) described a DMHI aiming primarily to reduce stress or improve related affective outcomes within preventative mental health contexts; (2) reported at least 1 engagement-supportive mechanism, such as personalization, biometric feedback, JITAI, gamification, or AI or chatbot functionality, implicitly or explicitly; (3) presented empirical, qualitative, implementation, simulation, modeling, or system design evidence relevant to engagement; (4) were peer-reviewed; and (5) were published in English. Protocols, reviews, dissertations, nondigital interventions, and non–mental health interventions were excluded. A study was also eligible if it addressed 1 or more engagement-supportive mechanisms and met the broader stress-focused criteria; it did not need to include every mechanism or use the term “engagement” explicitly. Commercial applications, app store–only evaluations, and other gray literature were out of scope unless evaluated in a peer-reviewed study meeting these criteria. Two technical preprints from the original corpus were retained as a narrow exception because they addressed emerging AI and reinforcement learning architectures [25,37]; they were interpreted as design or computational evidence rather than clinical effectiveness evidence. No new preprints were added, and the evidence update did not broaden the eligibility criteria.

Study Selection and Data Extraction

MM and SAP independently screened titles and abstracts and then assessed potentially eligible full texts in the original review. Disagreements were resolved through discussion with SG until consensus was reached. MM and SAP extracted data from studies identified in the original search using the structured 17-item form described below. For studies identified in the search update, MM conducted the initial screening and data extraction using the unchanged criteria, and SG verified the draft eligibility decisions, extracted data, and interpretations.

A structured 17-item form captured publication details, population, setting, intervention modality, duration, follow-up, study design, engagement definitions and measures, engagement-supportive mechanisms, AI or sensing components, key findings, and design implications. The same form was applied to the papers identified in the search update. Eligibility and mechanism coding were kept separate so that methodological studies could be included in the corpus without being misclassified as intervention-level engagement mechanisms.

Synthesis Approach

The synthesis proceeded in 2 stages. First, descriptive mapping characterized study design, population, modality, duration, and follow-up. Second, overlapping mechanisms were coded as static tailoring, dynamic adaptation, conversational support, context-aware prompting, biometric-responsive adaptation, gamified reinforcement, narrative immersion, or social and co-design support. In this review, AI encompassed conversational agents and LLM-based systems; recommender systems; contextual bandits; reinforcement learning, rule-based, or fuzzy logic approaches; and simulation or modeling used for assessment, timing, content selection, or interaction. Higher-level labels such as JITAI or AI were classified according to their functional role in sensing, decision-making, or delivery.

Findings were synthesized across behavioral, cognitive, experiential, relational, and methodological evidence. Behavioral indicators were distinguished from subjective or interaction-oriented engagement processes, and associations were interpreted descriptively because interventions frequently combined mechanisms and rarely isolated their independent effects.

Self-determination theory (SDT) informed the interpretation of motivational and experiential engagement. Meaningful choice may support autonomy, appropriate feedback and achievable challenges may support competence, and socially responsive interaction may support relatedness [38,39]. Work on immersive digital environments provided further context on need satisfaction, motivation, enjoyment, and continued participation [40]. These were theory-informed interpretations and included studies were not assumed to have explicitly adopted or tested SDT.

XSL•FO
RenderX

Ethical Considerations

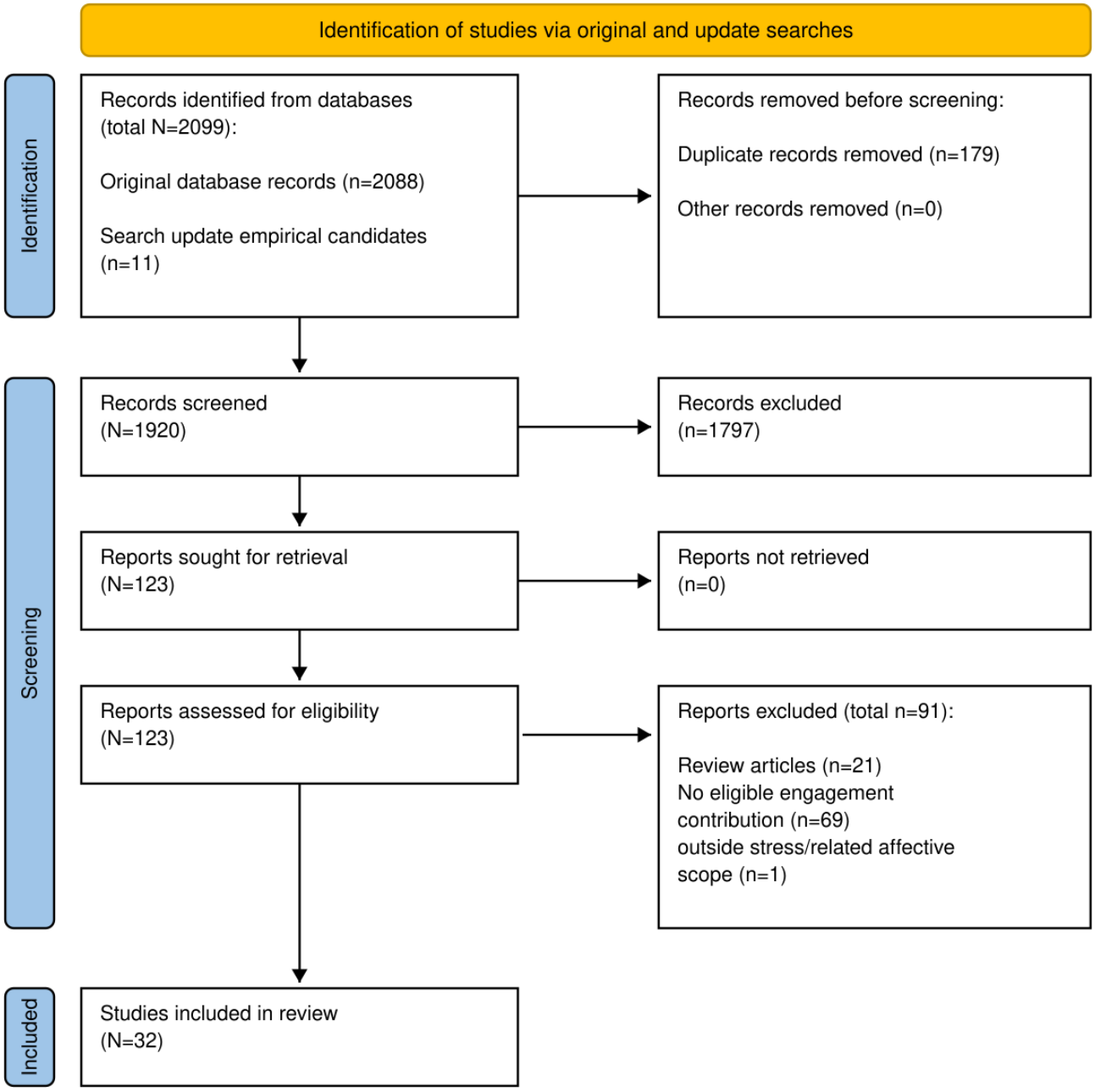
This review analyzed published reports only and did not recruit participants, intervene in care, or collect or process identifiable personal data. Ethics approval was therefore not obtained. This decision was consistent with the University of Padua ethics self-assessment and JMIR guidance [41,42]. Informed consent and participant compensation were not applicable.

Results

Study Selection

Figure 1 summarizes the selection process. The original search identified 2088 records; 179 duplicates were removed, and 1909 records were screened. After 1797 title or abstract exclusions, 112 full-text reports were assessed, and 22 studies met the original criteria. Eleven unique peer-reviewed candidate empirical papers identified during the evidence update were screened against the same criteria [43-53]. Ten studies met all criteria and were included in the review corpus [43-48,50-53]. The insomnia component trial [49] fell outside the stress or related affective eligibility boundary and was retained only as external interpretive evidence. Four reviews and meta-analyses [33-36] also informed interpretation but were not counted. The final review corpus comprised 32 studies.

Figure 1. Study selection across the original and update searches. Original database search covered January 2018 to April 2025 (n=2088), and an update covering May 2025 to August 2026 was conducted (11 unique, peer-reviewed empirical candidates). Four reviews and meta-analyses used only as contextual evidence were not counted in this empirical selection flow.



XSL•FO
RenderX

Overview of Included Studies

The 32 included studies were published between 2021 and 2026 [8,9,12-14,24-27,37,43-48,50-65]. Of these, 30 appeared from 2024 onward, while 1 was published in 2021 [59] and 1 in 2023 [65], reflecting the rapid recent growth of adaptive, context-aware, and AI-enabled DMHI research. Dissemination spanned clinical, digital health, human-computer interaction (HCI), and computing journals and conferences, illustrating the multidisciplinary intersection of mental health, behavioral science, and technology.

Most empirical studies were conducted in high-income settings in North America, Western Europe, Australia, New Zealand, East Asia, and Israel. Culturally adapted interventions in Jordan provided evidence from an otherwise underrepresented regional context [57,63]. Target populations included adolescents, young adults, and students with stress-, anxiety-, or mood-related needs; caregivers and parents; employees and health care trainees exposed to occupational stress; people awaiting psychotherapy; users of youth mental health platforms or mental health chatbots; breast cancer survivors; and people with chronic physical illnesses, including chronic obstructive pulmonary disease [8,9,12-14,24-27,37,43-48,50-65].

Intervention modalities were correspondingly diverse. They included mobile apps with monitoring or guided coping strategies [13,14,43]; web-based and blended programs [8,12,47,52]; wearable-integrated systems using physiological or stress detection to trigger support [37,56]; narrative or gamified interventions [51,55,59,65]; chatbot- and LLM-based conversational systems [24-27,46,53,61,63]; context-aware JITAIs [37,43,44,48,56,64,65]; and simulation, modeling, or adaptive trial frameworks relevant to decision-making and optimization [45,50,52,54,62]. Across these modalities, engagement was examined through behavioral use and adherence, perceived relevance and burden, emotional or relational experience, contextual receptivity, and methodological optimization. Modalities and engagement-supportive mechanisms frequently overlapped within the same intervention.

Study characteristics and design distribution are summarized in Table 2. Randomized or microrandomized controlled trials were the most common primary design, accounting for 53% (17/32) of the studies [12,13,24,26,44-46,48,50-52,57,59,61,63-65]. Qualitative investigations accounted for 19% (6/32), including interview-based, experiential, diary-based, and reflective designs [9,14,27,53,56,58]. Design or implementation studies accounted for 9% (3/32) [8,25,37]. Mixed methods studies also accounted for 9% (3/32) [47,55,60]. Simulation or modeling studies represented 6% (2/32) [54,62], while 1 quasi-experimental study accounted for 3% (1/32) [43]. These mutually exclusive categories reflect each study's primary methodological contribution, as some controlled trials also incorporated mixed methods components. Overall, 81% (26/32) of the included papers reported empirical studies involving conventional participant samples [8,12-14,24,26,27,43-48,50-53,55,57-61,63-65], while 19% (6/32) focused on design, reflexive inquiry, simulation, modeling, or reinforcement learning approaches without a conventional participant sample [9,25,37,54,56,62].

XSL•FO
RenderX

Table . General characteristics of included studies.

Study	Title	Study type	Duration	Follow-up
Weintraub et al [12]	Effects of a multifamily app-enhanced cognitive-behavioral therapy delivered via telehealth for parents of adolescents with mood or psychotic spectrum disorders	RCT[a]; mixed methods	9-week treatment; 3-month posttreatment follow-up	F [b]
Sirbu et al [13]	Comparative efficacy of PsyPills and OCAT mobile psychological interventions in reducing depressive, anxiety and stress symptoms: a blinded randomized clinical trial	Blinded RCT	10-day intervention; 1-month follow-up	F
Wang et al [54]	HealthQ: unveiling questioning capabilities of LLM[c] chains in healthcare conversations	Simulation study	N/A[d]	N/A
Heinz et al [24]	Randomized trial of a generative AI chatbot for mental health treatment	RCT	4-week Therabot intervention; follow-up at 8 weeks	F
Guo et al [25]	SouLLMate: an adaptive LLM-driven system for advanced mental health support and assessment, based on a systematic application survey	Survey and design	N/A	N/A
Sadprasid et al [55]	Leveraging idle games to incentivize intermittent and frequent practice of deep breathing	Mixed methods	4-week at-home deployment; 2 weeks per app condition	N/A
Neupane et al [56]	Wearable meets LLM for stress management: a duoethnographic study integrating wearable-triggered stressors and LLM chatbots for personalized interventions	Duoethnography study	22-day duoethnographic use period	N/A
Alvarez-Jimenez et al [8]	A national evaluation of a multi-modal, blended, digital intervention integrated within Australian youth mental health services	Implementation study	32-month implementation; 12-week clinical support allocation; engagement tracked to 24 weeks	F
Miller et al [9]	Facilitating co-design in healthcare: a collaborative autoethnography of personalities, power, and competing priorities	Autoethnography design	N/A	N/A
Dardas et al [57]	Validating a digital depression prevention program for adolescents in Jordan: cultural adaptation and user testing in a randomized controlled trial	RCT; mixed methods	10-session/module intervention; exact intervention duration not clearly reported	N/A

XSL•FO
RenderX

Study	Title	Study type	Duration	Follow-up
Rutherford et al [58]	Experiences of home - based pulmonary rehabilitation with mHealth and - based pulmonary rehabilitation in people with chronic obstructive pulmonary disease: a qualitative study	Qualitative study	8-week pulmonary rehabilitation program	N/A
Schwarz et al [59]	Does dynamic tailoring of a narrative-driven exergame result in higher user engagement among adolescents? Results from a cluster-randomized controlled trial	Cluster RCT	4-week game-based intervention period	N/A
Slade et al [60]	Design guidelines for improving mobile sensing data collection: prospective mixed methods study	Mixed methods	30-day mobile sensing user study	Poststudy/exit survey only[e]
Valentine et al [14]	A personalized, transdiagnostic smartphone app (Mello) targeting repetitive negative thinking for depression and anxiety: qualitative analysis of young people's experience	Qualitative study	6-week Mello intervention period	F
Campellone et al [26]	Safety and user experience of a generative artificial intelligence digital mental health intervention: exploratory randomized controlled trial	Exploratory RCT	2-week exploratory randomized trial	N/A
Tong et al [61]	Effectiveness of topic-based chatbots on mental health self-care and mental well-being: randomized controlled trial	RCT	10-day chatbot intervention; 1-month follow-up	F
Ying et al [62]	Mental health assessment model for college students using circular q-rung orthopair fuzzy Muirhead means and MULTIMOORA method	Computational modeling	N/A	N/A
de Graaff et al [63]	Evaluation of a guided chatbot intervention for young people in Jordan: feasibility randomized controlled trial	Feasibility RCT	8-week STARS[f] chatbot intervention	N/A
Coppens et al [64]	Exploring the added effect of three recommender system techniques in mobile health interventions for physical activity: a longitudinal randomized controlled trial	Longitudinal RCT	8-week app use; 6-month follow-up questionnaire	F
Siddals et al [27]	"It happened to be the perfect thing": experiences of generative AI chatbots for mental health	Qualitative study	N/A	N/A
Wang and Fang [65]	Effectiveness of gamified team competition as mHealth intervention for medical interns: a cluster micro-randomized trial	Micro-RCT	≈12-week intervention	N/A

XSL•FO
RenderX

Study	Title	Study type	Duration	Follow-up
Aqajari et al [37]	Enhancing performance and user engagement in everyday stress monitoring: a context-aware active reinforcement learning approach	Reinforcement learning study	N/A	N/A
Kor et al [43]	Developing and testing a brief mindfulness just-in-time adaptive intervention to reduce stress among caregivers of people with dementia: quasi-experimental study	Quasi-experimental study	18-day intervention	N/A
Elmer et al [44]	A social support just-in-time adaptive intervention for individuals with depressive symptoms: feasibility study with a microrandomized trial design	Feasibility micro-RCT, mixed methods	18-day intervention	N/A
Newby et al [45]	Brief digital interventions for psychological distress: an AI-enhanced response-adaptive randomized clinical trial	Response-adaptive RCT	2-week onboarding plus 2-week intervention	N/A
Shoshani et al [46]	Efficacy of a conversational AI agent for psychiatric symptoms and digital therapeutic alliance: a randomized clinical trial	3-arm RCT	12-week intervention; 3-month follow-up	F
Valentine et al [47]	Sustained engagement with a digital youth mental health platform: a mixed-methods study	Convergent mixed methods study	Real-world platform use; no conventional fixed duration reported	N/A
Dai et al [48]	A design framework for context-aware digital micro-interventions: integrating emotion recognition and just-in-time adaptive intervention for workplace stress management	Design framework and 2-arm RCT	6-week RCT	N/A
Chow et al [50]	Digital mental health treatment and symptoms of depression and anxiety in breast cancer survivors: a randomized clinical trial	Sequential multiple assignment randomized trial	8-week intervention; 6- and 12-month follow-up	F
Barry et al [51]	A digital behavioral health game for adolescent mental health: a prespecified secondary analysis of a randomized clinical trial	Prespecified secondary analysis of an RCT	6-week intervention; 12-month follow-up	F
Rouvere et al [52]	Personalization strategies for increasing engagement with digital mental health resources: sequential multiple assignment randomized trial	Sequential multiple assignment randomized trial	Single-session real-world website experiment	N/A

XSL•FO
RenderX

Study	Title	Study type	Duration	Follow-up
Xu et al [53]	The digital therapeutic alliance with mental health chatbots: diary study and thematic analysis	Longitudinal diary study and reflexive thematic analysis	4-week user study	N/A

[a]RCT: randomized controlled trial.

[b]F refers to an assessment or engagement observation beyond the immediate intervention or primary postassessment period.

[c]LLM: large language model.

[d]N/A denotes studies where duration or follow-up could not be meaningfully defined, such as simulation, modeling, design, or autoethnographic studies, or where no follow-up was reported.

[e]"Poststudy/exit survey only" indicates assessment at the end of the study period rather than longer-term follow-up.

[f]STARS: Scalable Technology for Adolescents and Youth to Reduce Stress.

Engagement-Supportive Mechanisms

Table 3 summarizes the presence or absence of the primary engagement-supportive mechanisms identified across the 32 included studies. Studies were classified using overlapping mechanism categories rather than mutually exclusive intervention types because many stress-focused DMHIs combined adaptive, conversational, contextual, motivational, or social components within the same intervention architecture.

Personalization reflected 2 conceptually distinct approaches. Static tailoring appeared in 28% (9/32) of the studies and used predefined characteristics, preferences, symptom profiles, cultural factors, or user-provided information to select or modify support [8,12,13,43,44,52,57,63,64]. Dynamic adaptation appeared in 44% (14/32) and used accumulating behavioral, contextual, physiological, or conversational information to modify intervention timing, content, allocation, or interaction over time [14,24-26,37,43,44,46,48,50,54,56,59,64]. Some interventions combined both approaches [43,44,64], illustrating that static and dynamic personalization are related but not interchangeable. Personalization did not uniformly improve engagement. In Rouvere et al [52], additional tailoring questions increased disengagement, while the effects of tailoring differed according to the information used.

Conversational support appeared in 31% (10/32) of the studies [24-27,46,53,54,56,61,63]. These interventions used chatbot-based or guided dialogue to provide psychoeducation, emotional responsiveness, reflective interaction, or contextually relevant support. Conversational studies reported acceptability and user experience outcomes as well as relational dimensions such as therapeutic alliance and perceived bonding. Shoshani et al [46] found that perceived therapeutic alliance was associated with engagement and symptom improvement, while Xu et al [53] identified conversational control; input fit; caring responses; useful advice; colloquial language; and a private, nonjudgmental setting as factors shaping users' bonds with mental health chatbots.

Social or co-design mechanisms appeared in 28% (9/32) of the studies [8,9,44,47,50,51,58,63,65]. These included participatory design, blended care, culturally responsive guidance, peer or team interaction, support network activation, and selective human coaching. Valentine et al [47] associated sustained engagement with a youth mental health platform with validation, normalization, belonging, and the practical application of therapeutic insights. Chow et al [50] reported that adding human coaching increased the number of days of app use among participants whose use was low during the first intervention week.

Context-aware prompting appeared in 22% (7/32) of the studies [14,37,43,44,48,56,64]. These systems adjusted intervention delivery according to behavioral, contextual, physiological, or user-reported information. Receptivity and intervention timing were examined directly in several context-aware studies. Kor et al [43] used accumulating receptivity data to inform mindfulness prompt delivery, Elmer et al [44] found that prompts triggered by a stated need for support were rated as better timed and more helpful than prompts triggered by distress thresholds, and Dai et al [48] reported higher adherence with context-aware delivery than fixed delivery.

Biometric-responsive adaptation appeared in 13% (4/32) of the studies [37,48,56,59]. Physiological or emotion recognition data generally functioned as inputs to contextual assessment, prompting, or adaptive delivery rather than as independent engagement mechanisms. Gamified reinforcement appeared in 16% (5/32) of the studies [12,51,55,59,65], while narrative immersion appeared in 6% (2/32) of the studies [51,59]. Their implementation and findings varied: an idle breathing game increased practice frequency [55]; dynamic tailoring increased narrative presence without improving measured immersion, while play time and completed sessions were lower in the tailored condition [59]; and team competition produced early affective benefits that weakened over time [65]. In Barry et al [51], a co-designed narrative game reduced depressive symptoms relative to an active control over 12 months but did not improve anxiety; mechanism-specific engagement or narrative resonance outcomes were not reported.

These frequencies should be interpreted cautiously. The review intentionally foregrounded adaptive, technologically mediated, and engagement-supportive interventions; therefore, the distribution within this selected corpus should not be treated as an estimate of mechanism prevalence across all DMHIs. In addition, 3 studies [45,60,62] were not assigned a mechanism (Table 3). Slade et al [60] examined sensing compliance, user control, and data collection burden; Ying et al [62] contributed a computational mental health assessment model; and Newby et al [45] used response-adaptive allocation across trial participants rather than an intervention-level engagement mechanism. These studies remained eligible because they

XSL•FO
RenderX

contributed methodological or computational evidence relevant to engagement and adaptive intervention design.

XSL•FO
RenderX

Table . Presence (✓) or absence (—) of engagement-supportive mechanisms in each included study.[a]

Study	Static tailoring	Dynamic adaptation	Conversational support	Context-aware prompting	Biometric-responsive adaptation	Gamified reinforcement	Narrative immersion	Social/co-design
Weintraub et al [12]	✓	—	—	—	—	✓	—	—
Sirbu et al [13]	✓	—	—	—	—	—	—	—
Wang et al [54]	—	✓	✓	—	—	—	—	—
Heinz et al [24]	—	✓	✓	—	—	—	—	—
Guo et al [25]	—	✓	✓	—	—	—	—	—
Sadprasid et al [55]	—	—	—	—	—	✓	—	—
Neupane et al [56]	—	✓	✓	✓	✓	—	—	—
Alvarez-Jimenez et al [8]	✓	—	—	—	—	—	—	✓
Miller et al [9]	—	—	—	—	—	—	—	✓
Dardas et al [57]	✓	—	—	—	—	—	—	—
Rutherford et al [58]	—	—	—	—	—	—	—	✓
Schwarz et al [59]	—	✓	—	—	✓	✓	✓	—
Slade et al [60]	—	—	—	—	—	—	—	—
Valentine et al [14]	—	✓	—	✓	—	—	—	—
Campellone et al [26]	—	✓	✓	—	—	—	—	—
Tong et al [61]	—	—	✓	—	—	—	—	—
Ying et al [62]	—	—	—	—	—	—	—	—
de Graaff et al [63]	✓	—	✓	—	—	—	—	✓
Coppens et al [64]	✓	✓	—	✓	—	—	—	—
Siddals et al [27]	—	—	✓	—	—	—	—	—
Wang (Jitao) et al [65]	—	—	—	—	—	✓	—	✓
Aqajari et al [37]	—	✓	—	✓	✓	—	—	—
Kor et al [43]	✓	✓	—	✓	—	—	—	—
Elmer et al [44]	✓	✓	—	✓	—	—	—	✓
Newby et al [45]	—	—	—	—	—	—	—	—
Shoshani et al [46]	—	✓	✓	—	—	—	—	—

XSL•FO
RenderX

Study	Static tailoring	Dynamic adaptation	Conversational support	Context-aware prompting	Biometric-responsive adaptation	Gamified reinforcement	Narrative immersion	Social/co-design
Valentine et al [47]	—	—	—	—	—	—	—	✓
Dai et al [48]	—	✓	—	✓	✓	—	—	—
Chow et al [50]	—	✓	—	—	—	—	—	✓
Barry et al [51]	—	—	—	—	—	✓	✓	✓
Rouvere et al [52]	✓	—	—	—	—	—	—	—
Xu et al [53]	—	—	✓	—	—	—	—	—

[a]Since many interventions combined adaptive and interdependent mechanisms, the categories presented in this table should be interpreted as overlapping analytical dimensions rather than mutually exclusive intervention types.

Co-Occurrence of Mechanisms

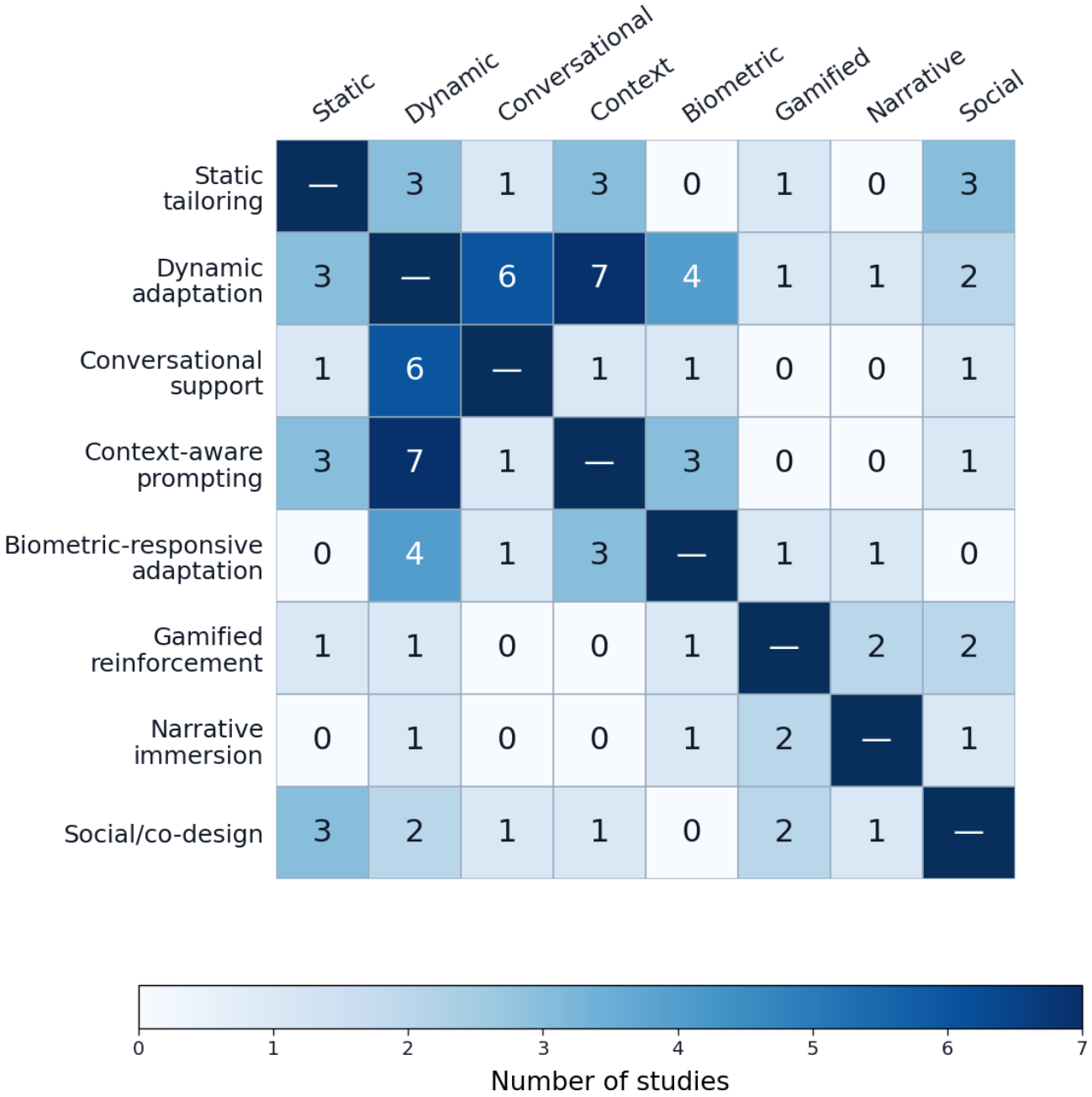
Figure 2 illustrates the co-occurrence of engagement-supportive mechanisms across the included studies. Dynamic adaptation most frequently co-occurred with context-aware prompting in 7 studies [14,37,43,44,48,56,64] and with conversational support in 6 studies [24-26,46,54,56]. These combinations reflected systems in which behavioral, contextual, or conversational information informed decisions about the timing, content, or form of support.

Biometric-responsive adaptation co-occurred with dynamic delivery in all 4 studies in which it appeared [37,48,56,59] and with context-aware prompting in 3 studies [37,48,56]. This pattern reflects the role of physiological or emotion recognition data as sensing inputs within broader adaptive systems. Context-aware prompting similarly tended to operate alongside personalization, recommendation logic, conversational support, or reinforcement learning rather than as an isolated feature.

Gamified and narrative mechanisms remained comparatively separate from AI-enabled conversational architectures, although Schwarz et al [59] combined dynamic narrative tailoring, biometric responsiveness, and gamified reinforcement. Social mechanisms spanned co-design, blended care, culturally guided support, team competition, support network activation, human coaching, and relational experiences of validation or belonging [8,9,44,47,50,51,58,63,65]. Overall, the co-occurrence patterns show that engagement support was commonly implemented through interacting sensing, decision, and delivery components. These descriptive relationships do not establish that combining more mechanisms necessarily produced stronger engagement.

XSL•FO
RenderX

Figure 2. Co-occurrence of engagement strategies across the studies. Values show the number of studies in which each pair of mechanisms co-occurred. Mechanisms are overlapping analytical dimensions and not mutually exclusive intervention categories. Biometric: biometric-responsive adaptation; Context: context-aware prompting; Conversational: conversational support; Dynamic: dynamic adaptation; Gamified: gamified reinforcement; Narrative: narrative immersion; Social: social/co-design; Static: static tailoring.



AI, Adaptivity, and Methodological Contribution

AI-enabled systems were represented across controlled trials, qualitative studies, simulation and modeling research, and adaptive system design. AI served several distinguishable functions: direct conversational interaction [24-27,46,53,61,63]; within-person adaptation of timing, content, or prompting [37,43,44,48,56,64]; response-adaptive allocation across trial participants [45]; and back-end assessment, simulation, or modeling [25,37,54,62]. Therefore, AI functioned both as an intervention component and as methodological infrastructure for developing or evaluating adaptive systems.

This distinction was particularly important for Newby et al [45]. Its contextual bandit modified allocation probabilities across sequential participants according to distress severity, while the brief interventions received by individual participants remained standardized. The study consequently contributed evidence about AI-enhanced trial methodology rather than dynamically personalized intervention delivery. By contrast, within-person adaptive studies used user-level information to change when or how support was delivered [37,43,44,48,56,64].

The type of evidence also differed across AI implementations. Controlled conversational studies primarily reported completion, acceptability, therapeutic alliance, safety, user experience, or symptom-related outcomes [24,26,46,61,63]. Qualitative and diary studies examined emotional responsiveness, trust, perceived support, conversational fit, and concerns about accuracy or authenticity [27,53,56]. Simulation, modeling, and

XSL•FO
RenderX

reinforcement learning studies examined prompt construction, adaptive assessment, recommendation logic, sensing burden, or optimization before or alongside real-world deployment [25,37,45,54,62].

Across the corpus, AI and algorithmic adaptation served distinguishable conversational, within-person adaptive, response-adaptive, and modeling functions. Shoshani et al [46] connected conversational alliance with engagement; Dai et al [48] compared context-aware delivery with fixed delivery; and Newby et al [45] demonstrated the methodological distinction between adapting trial allocation and adapting intervention delivery. Nevertheless, most studies evaluated AI as part of a multicomponent system, and few isolated its independent contribution to sustained engagement.

Patterns in Engagement Feature Adoption by Methodology

Patterns of mechanism adoption varied across study designs. Static tailoring was prominent in controlled, culturally adapted, and real-world personalization studies [12,13,43,44,52,57,63,64]. Dynamic adaptation appeared more often in AI-enabled, recommender-based, sensing-supported, or contextually responsive systems [14,24-26,37,43,44,46,48,50,54,56,59,64]. Context-aware prompting was concentrated in JITAI, recommendation, and reinforcement learning studies [14,37,43,44,48,56,64], while biometric-responsive adaptation remained limited to sensing-enabled adaptive interventions [37,48,56,59].

Conversational support was represented in both controlled chatbot evaluations and qualitative studies of relational experience [24-27,46,53,54,56,61,63]. Gamified reinforcement appeared primarily in controlled or mixed methods evaluations [12,51,55,59,65], whereas social and co-design mechanisms were particularly visible in qualitative, participatory, implementation, and adaptive support studies [8,9,44,47,50,51,58,63,65].

These methodological patterns affected what could be learned about engagement. Controlled and microrandomized studies contributed behavioral indicators, receptivity data, and symptom-related outcomes; qualitative, diary-based, participatory, and reflective studies revealed trust, burden, cultural fit, power dynamics, and relational experience; and implementation, simulation, and modeling studies clarified scalability, decision logic, and optimization. The resulting evidence types were complementary but not directly interchangeable.

Study Durations and Follow-Up

Intervention and observation periods varied substantially. Thirteen studies involved intervention or user study periods of approximately 4 weeks or less [13,24,26,43-45,52,53,55,56,59-61]. These included brief app and chatbot trials, a 4-week exergame, a 4-week gamified breathing deployment, an 18-day JITAI, a 22-day wearable LLM study, a 30-day sensing study, and a single-session personalization experiment. Seven studies lasted between 5 and 8 weeks [14,48,50,51,58,63,64], and 3 studies involved 9- to 12-week intervention or support periods [12,46,65]. Alvarez-Jimenez et al [8] was embedded in a longer implementation and tracked engagement for up to 24 weeks.

Eight studies did not report a conventional fixed duration or involved simulation, system design, computational modeling, autoethnography, qualitative accounts of previous use, or real-world platform engagement [9,25,27,37,47,54,57,62]. These studies should not be interpreted as fixed-duration interventions.

Follow-up or assessment beyond the immediate intervention or primary postassessment period was reported in 31% (10/32) of the studies [8,12-14,24,46,50,51,61,64]. Follow-up periods included 1 month [13,61], 8 weeks after a 4-week intervention [24], approximately 3 months [12,46], 6 months [64], engagement tracking for up to 24 weeks [8], and 6- or 12-month assessments in studies of adaptive coaching and narrative game-based support [50,51]. Most studies therefore provided limited evidence about sustained engagement beyond the immediate intervention period.

Structured Comparison of Findings

Tables 4 and 5 provide complementary summaries of the engagement-supportive mechanisms, AI-enabled systems, methodological approaches, and key findings represented across the included studies. Table 4 presents the engagement-supportive mechanisms identified in each study, the associated findings, and the resulting design implications. The mechanisms represent intervention characteristics, interaction processes, or adaptive functions intended to support engagement; they should not be treated as engagement outcomes themselves. Reported outcomes included behavioral indicators such as uptake, adherence, completion, retention, and receptivity; experiential and relational dimensions such as relevance, burden, immersion, trust, alliance, and emotional support; and qualitative or methodological findings relevant to intervention design.

Across Table 4, controlled studies most commonly contributed behavioral and symptom-related findings, while qualitative and participatory studies foregrounded the contextual, experiential, and relational dimensions of engagement. Adaptive timing, personalization, and conversational responsiveness were frequently associated with greater relevance, acceptability, adherence, or perceived support within individual study contexts [14,24,43,44,46,48,50,52,64]. These findings do not demonstrate that any single mechanism independently produced the reported outcome because most interventions combined several components. Gamified and narrative interventions also produced variable patterns across practice frequency, usage, immersion, affective outcomes, and longer-term symptoms [51,55,59,65].

XSL•FO
RenderX

Table . Engagement-supportive mechanisms, key findings, and design implications.

Study	Primary engagement-supportive mechanism(s)	Key findings (feature focus)	DMHI[a] design implications
Weintraub et al [12]	• Static user-tailored personalization • CBT[b]-based self-guided support	App-enhanced intervention lacked effective engagement features; parents showed lower practice than standard CBT	Tailor digital features (reminders and gamification) to engagement needs of adolescents and their caregivers
Sirbu et al [13]	• Static user-tailored personalization • Self-management support	Daily cognitive techniques (PsyP-ills) with personalization improved acceptability and reduced stress	Brief structured interventions with personalized content enhance engagement
Wang et al [54]	• Dynamic adaptation • Conversational/LLM[c] prompt support	Chain-of-thought prompts generated empathetic, context-aware questions	Use adaptive LLM prompts for personalized dialogue
Heinz et al [24]	• Conversational support • Emotional responsiveness • Adaptive dialogue interaction	Generative AI chatbot provided real-time, empathic interactions	Conversational AI can deliver personalized JITAI[d]-like support
Guo et al [25]	• Conversational support • Reflective self-monitoring	Adaptive LLM chains improved reliability and safety of mental health dialogues	Implement multistage LLM frameworks for safe, adaptive conversations
Sadprasid et al [55]	• Gamified behavioral reinforcement • Breathing practice support	Gamified idle breathing increased practice frequency	Gamification can incentivize self-regulation practice
Neupane et al [56]	• Context-aware adaptive prompting • Biometric-responsive adaptation • Conversational support	Wearable-triggered prompts and LLM chatbots created context-aware support	Combine biometric triggers with chatbots for JITAI
Alvarez-Jimenez et al [8]	• Social/community support • Blended therapeutic engagement	Blended self-help tools and social networking supported varied engagement	Integrate multiple engagement strategies and clinician monitoring
Miller et al [9]	• Co-design participation • User-centered interaction design	User engagement shaped by co-design processes	Engagement features should be co-designed with users
Dardas et al [57]	• Static user-tailored personalization • Culturally adaptive support • Motivational engagement	Culturally adapted content and motivational interviewing increased adherence	Cultural tailoring and personalization support engagement
Rutherford et al [58]	• Flexible delivery • Social motivation • Experiential engagement	Home-based flexibility vs center-based social motivation	Blend flexibility with social engagement features
Schwarz et al [59]	• Narrative immersion • Gamified behavioral reinforcement • Motivational tailoring	Dynamic tailoring increased narrative presence but did not improve measured immersion; play time and session completion were lower	Assess narrative presence, immersion, and behavioral use separately, alongside sensing burden and usability
Slade et al [60]	• User control over sensing • Low-burden data collection	User fatigue from frequent sensing; valued autonomy	Allow users to control biometric data collection
Valentine et al [14]	• Dynamic adaptive personalization • Emotional responsiveness	Participants described check-ins as supporting reflection and self-awareness; low mood and feeling “stuck” hindered use	Allow customizable prompts and self-guided use; address low mood and motivation barriers

XSL•FO
RenderX

Study	Primary engagement-supportive mechanism(s)	Key findings (feature focus)	DMHI[a] design implications
Campellone et al [26]	• Conversational support • Emotional responsiveness • Psychoeducational interaction	Generative responses achieved higher rates of user-confirmed accurate reflections (98% vs 69%); satisfaction and bond ratings were similar	Generative AI requires safety guardrails to balance empathy and accuracy
Tong et al [61]	• Conversational support • Reflective dialogue	The chatbot program improved short-term self-care intentions and mental health literacy; sustained comparative benefits were limited	Offer modular content tailored to user needs
Ying et al [62]	• Computational adaptive assessment modeling • Personalized mental health state evaluation	Proposed fuzzy logic assessment framework; no engagement features studied	Back-end assessment models should inform personalized interventions
de Graaff et al [63]	• Static tailoring • Guided conversational support • Culturally adapted/social support	Culturally adapted STARS[e] was feasible and acceptable; lesson and support call completion remained limited	Combine guidance with culturally sensitive content
Coppens et al [64]	• Context-aware adaptive prompting • Personalized behavioral support	Contextual bandit recommendations increased message relevance and motivation	Use adaptive recommender systems to personalize prompts
Siddals et al [27]	• Conversational support • Emotional responsiveness • Relational engagement	Users appreciated empathetic AI but questioned accuracy	Transparent AI explanations help manage expectations
Wang and Fang [65]	• Gamified behavioral reinforcement • Team-based social motivation	Team-based gamification boosted affect early, but effects waned	Introduce periodic social challenges to maintain engagement
Aqajari et al [37]	• RL[f] optimization • Dynamic adaptive personalization • Biometric-responsive adaptation	RL reduced prompts while maintaining detection accuracy	Leverage RL to optimize biometric-triggered JITAI
Kor et al [43]	• Static and adaptive JITAI timing • Receptivity-informed prompting	In 120 caregivers, retention was 100%; 92.5% found the app easy to use, and 81.7% found it helpful; stress, depression, and caregiver burden decreased from preintervention to postintervention, without a control group	Receptivity models can reduce timing mismatch, but comparative trials are needed before attributing outcome change to adaptation
Elmer et al [44]	• Personalized and need-triggered social-support JITAI	EMA[g] completion was 85.37%; prompts triggered by stated support need were rated as better timed and more helpful than distress threshold triggers; engagement was moderate and heterogeneous	Trigger rules should incorporate expressed support need and practical barriers and not distress thresholds alone
Newby et al [45]	• AI-enabled response-adaptive trial allocation	A contextual bandit allocated 1282 students to brief interventions by distress severity; physical activity, mindfulness, and sleep hygiene effects differed across severity strata; engagement and usability were secondary outcomes	Separate adaptation of trial allocation from adaptation of intervention delivery when interpreting engagement mechanisms
Shoshani et al [46]	• Conversational AI • Therapeutic alliance • Dynamic dialogue		Alliance, warmth, and competence should be measured alongside usage; developer conflicts and long-term safety require scrutiny

XSL•FO
RenderX

Study	Primary engagement-supportive mechanism(s)	Key findings (feature focus)	DMHI[a] design implications
		In a 995-participant RCT[h], the AI intervention improved several symptom and well-being outcomes; perceived alliance was associated with engagement (β=.31); engagement was associated with symptom change (β=−.58), with an indirect alliance-to-symptom association through engagement (β=−.18)	
Valentine et al [47]	• Relational and social engagement • Practical utility	Among 36 young people, validation, normalization, belonging, and applying therapeutic insights in daily life were associated with sustained platform engagement	Design for relational value and low-effort practical use; address overload, low motivation, and limited reminder support
Dai et al [48]	• Context-aware JITAI • Emotion recognition • Biometric-responsive adaptation	In a 6-week RCT of 120 office workers, adaptive delivery produced higher adherence than fixed scheduling (78% vs 52%) and greater stress reduction; context congruence partly mediated effects	Validate sensing accuracy and decision rules while minimizing burden and protecting privacy
Chow et al [50]	• Early-use monitoring • Adaptive escalation to human coaching	In a SMART[i] of 313 breast cancer survivors, the digital treatment improved anxiety but not depression relative to psychoeducation; among participants with low use in week 1, coaching increased the number of days of app use	Use early behavioral signals to target scarce human support, and distinguish engagement effects from symptom effects
Barry et al [51]	• Co-designed game • Narrative decision-making • Peer scenarios	Among 532 adolescents, the 6-week game reduced depressive symptoms relative to an active control over 12 months (adjusted difference at 12 months −1.12, 95% CI −1.85 to −0.38) but did not improve anxiety; beliefs about psychological services improved at 6 weeks	Gamified narrative may support durable benefit when embedded in co-design and tested against an active control; mechanism-specific engagement data remain important
Rouvere et al [52]	• Static tailoring of messages and resources	In a large real-world SMART, extra tailoring questions increased disengagement (OR[j] 1.14); demographic tailoring reduced disengagement (OR 0.90) and increased featured-resource clicks (OR 1.90), whereas need or next-step tailoring did not reduce disengagement	Account for the interaction cost of collecting tailoring data; validate each tailoring variable rather than assuming that greater personalization is better
Xu et al [53]	• Conversational relationship • Flexible input • Privacy and nonjudgmental support	In a 4-week diary study of 26 adults using Woebot and Wysa, 18 reported a bond or light bond; 6 themes linked bonding to conversational control, input fit, caring responses, useful advice, colloquial language, and a private nonjudgmental space	Design for preference fit and relational continuity; offer flexible input and test alliance without equating anthropomorphic rapport with clinical safety

[a]DMHI: digital mental health intervention.

[b]CBT: cognitive behavioral therapy.

[c]LLM: large language model.

[d]JITAI: just-in-time adaptive intervention.

[e]STARS: Scalable Technology for Adolescents and Youth to Reduce Stress.

[f]RL: reinforcement learning.

[g]EMA: ecological momentary assessment.

[h]RCT: randomized controlled trial.

[i]SMART: sequential multiple assignment randomized trial.

XSL•FO
RenderX

[j]OR: odds ratio.

Table 5 focuses specifically on AI-enabled systems and methodological contributions. It distinguishes studies in which AI functioned as enabling infrastructure (eg, assessment modeling, recommendation, reinforcement learning, response-adaptive allocation, or simulation) from studies in which conversational or adaptive systems directly shaped users' experiences through dialogue, contextual delivery, emotional support, or personalized interaction. The table also shows how controlled, qualitative, implementation, and computational methods answered different questions concerning efficacy, feasibility, acceptability, safety, relational experience, system logic, and optimization.

Taken together, Tables 4 and 5 indicate that AI was most directly relevant to engagement when embedded within a broader engagement-supportive process than when presented as a standalone feature. They should not be interpreted as rankings of effectiveness because the studies used heterogeneous designs, outcomes, and observation periods and because most did not isolate the effects of individual mechanisms.

XSL•FO
RenderX

Table . AI-enabled engagement-supportive systems, methodological contributions, and implications for adaptive DMHI[a] design.

Study	AI component	Key findings (AI focus)	Methodological contributions	Implications for research
Weintraub et al [12]	No AI	No AI component was used; engagement difficulties were reported in the app-enhanced condition	Compared app-enhanced vs standard CBT[b] and revealed differential engagement	Mixed methods clarify which features support parental engagement
Sirbu et al [13]	No AI	PsyPills used structured REBT[c] principles without AI	Measured symptom changes and acceptability in a brief intervention	Blinded trials test the efficacy of structured digital interventions
Wang et al [54]	Simulation of LLM[d] prompts	Chain-of-thought LLM prompts produced empathetic questioning	Evaluated LLM prompting configurations without end user data	Simulation guides design choices before deployment
Heinz et al [24]	Generative AI chatbot	AI chatbot users experienced greater mental health improvements and high completion	Measured mental health outcomes and adherence to chatbot intervention	Controlled trials provide evidence for AI chatbot efficacy but need long-term follow-up
Guo et al [25]	LLM survey	Adaptive LLM chains improved reliability in mental health support	Combined descriptive app survey with architectural development and benchmarking	Surveys inform system design and highlight safety concerns
Sadprasid et al [55]	No AI	Gamified breathing interaction increased short-term engagement and practice frequency through repetitive behavioral reinforcement	Randomized comparison and qualitative feedback captured gamification effects	Mixed methods link usage data with user perceptions of gamification
Neupane et al [56]	Wearable and LLM	Wearable-triggered prompts and LLM chatbots offered contextual support but lacked depth	Reflective narratives illuminated lived experiences of wearable-triggered interventions	Autoethnography uncovers contextual and emotional dimensions often missed by trials
Alvarez-Jimenez et al [8]	No AI	Engagement reliant on blended human-digital support	Real-world evaluation with descriptive statistics and surveys	Implementation research highlights scalability and service integration challenges
Miller et al [9]	No AI	Engagement shaped by co-design and not AI	Analyzed co-design processes and power dynamics	Qualitative inquiry elucidates barriers to participatory engagement
Dardas et al [57]	No AI	Cultural adaptation and motivational interviewing drove engagement	Combined quantitative outcomes with user feedback on culturally adapted content	Mixed methods ensure that cultural tailoring is evaluated for efficacy and acceptability
Rutherford et al [58]	No AI	Flexibility and social motivation were key	Semistructured interviews captured experiences of home vs center programs	Qualitative methods reveal experiential factors impacting engagement
Schwarz et al [59]	No AI	Dynamic tailoring increased narrative presence but did not improve measured immersion; play time and session completion were lower	Tested dynamic vs static exergames across schools	Cluster trials evaluate gamified DHIs[e] in real-world settings
Slade et al [60]	No AI	User control over sensing increased trust	Linked sensor compliance data with thematic interviews	Integrating quantitative sensing and qualitative insights guides user-centered data collection
Valentine et al [14]	No AI	Users valued self-guided flexibility and reflective check-ins; low mood and motivation could hinder use	Provided insights into user experiences with a transdiagnostic app	Qualitative evaluations refine personalized content and support features

XSL•FO
RenderX

Study	AI component	Key findings (AI focus)	Methodological contributions	Implications for research
Campellone et al [26]	Generative AI vs rule-based	Generative responses achieved higher rates of user-confirmed accurate reflections (98% vs 69%); satisfaction and bond ratings were similar	Compared generative vs rule-based chatbots on reflection accuracy, satisfaction, bond, and safety	Early-phase trials evaluate AI safety and user satisfaction before scaling
Tong et al [61]	Rule-based chatbots	The chatbot program improved short-term self-care intentions and mental health literacy; sustained comparative benefits were limited	A 2-arm RCT[f] compared a 10-topic chatbot program with a waitlist control	Factorial RCTs could assess the relative benefit of different modules
Ying et al [62]	Modeling	Computational assessment framework demonstrated how adaptive modeling architectures may support future personalized intervention systems, although no direct engagement outcomes were evaluated	Used fuzzy logic and multicriteria decision-making to build assessment model	Model-based studies lay the groundwork for adaptive systems but need empirical validation
de Graaff et al [63]	Guided chatbot	Culturally adapted STARS[g] was feasible and acceptable; lesson and support call completion remained limited	Assessed acceptability and retention of a guided chatbot in Jordan	Feasibility trials evaluate cultural adaptations and inform larger RCTs
Coppens et al [64]	Recommender system	Adaptive recommendation logic increased perceived relevance and motivational responsiveness of intervention prompts	Multiweek study compared recommender algorithms	Longitudinal trials gauge sustained engagement and algorithm performance
Siddals et al [27]	Generative AI experiences	Users valued empathic AI but raised concerns about accuracy	Interviews captured user experiences with generative AI chatbots	Qualitative feedback informs ethical guidelines and user education
Wang and Fang [65]	No AI (gamification)	Team-based gamification improved affect without AI	Microrandomized design assessed immediate affective responses to gamified competition	Microrandomized trials identify short-term effects and inform optimal timing[h]
Aqajari et al [37]	RL[i]	Adaptive RL architecture reduced prompting burden while maintaining contextual responsiveness	Offline simulation and online deployment evaluated RL for stress monitoring	Iterative evaluation demonstrates RL applicability in DHIs
Kor et al [43]	Machine learning receptivity models	Static, random, and participant-updated adaptive models selected prompt timing	Within-person delivery phases linked accumulated receptivity data to timing decisions	Compare adaptive and nonadaptive schedules in controlled designs and report model calibration
Elmer et al [44]	Personalized threshold and need trigger logic	Support need triggers were more acceptable than distress variable thresholds	A microrandomized design compared 4 trigger strategies using EMA[j] and qualitative data	Model practical opportunity and user-declared need and not distress alone
Newby et al [45]	Contextual multiarmed bandit	The algorithm updated allocation probabilities across sequential minitrials by distress severity	Response-adaptive allocation improved trial efficiency but reduced power for some contrasts	Report clearly whether AI adapts research allocation, intervention delivery, or both
Shoshani et al [46]	Conversational AI agent	Therapeutic alliance was associated with engagement and symptom improvement	A 3-arm RCT compared AI, group therapy, and waiting list with 3-month follow-up	Replicate independently and examine safety, alliance durability, and developer conflicts
Valentine et al [47]	No AI	Sustained use was associated with relational experience and practical application	Usage data were integrated with interviews comparing high and low engagers	Mixed methods can explain why similar usage levels may reflect different experiences

XSL•FO
RenderX

Study	AI component	Key findings (AI focus)	Methodological contributions	Implications for research
Dai et al [48]	Multimodal emotion recognition and context-aware decision rules	Adaptive microinterventions improved adherence relative to fixed scheduling	Delphi development, prototype testing, and a 2-arm RCT linked design decisions to outcomes	Externally validate sensing and mediation results before wider workplace deployment
Chow et al [50]	No AI; SMART[k] decision rule using week 1 app use	Human coaching increased the number of days of app use among early low engagers	Sequential randomization tested an adaptive support pathway within an effectiveness trial	Predefine decision points and thresholds and report both engagement and clinical outcomes
Barry et al [51]	No AI	The co-designed narrative game produced a small long-term reduction in depressive symptoms but no anxiety effect	An active control and 12-month follow-up strengthened inference about durability	Measure which game and narrative interactions drive engagement and benefit
Rouvere et al [52]	No AI; rule-based tailoring	Tailoring effects depended on the information used, and additional questions themselves increased disengagement	A large real-world sequential experiment separated data collection burden from tailoring effects	Evaluate the net engagement benefit of personalization after accounting for interaction burden
Xu et al [53]	Commercial conversational agents; AI component not experimentally isolated	Relational engagement depended on fit between user preferences and conversational style, input mode, and perceived support	Longitudinal diaries, screenshots, and interviews captured alliance formation over time	Combine behavioral telemetry with repeated qualitative and alliance measures

[a]DMHI: digital mental health intervention.

[b]CBT: cognitive behavioral therapy.

[c]REBT: rational emotive behavior therapy.

[d]LLM: large language model.

[e]DHI: digital health intervention.

[f]RCT: randomized controlled trial.

[g]STARS: Scalable Technology for Adolescents and Youth to Reduce Stress.

[h]Microrandomized trials function as methodological frameworks for evaluating adaptive intervention delivery rather than engagement features in themselves.

[i]RL: reinforcement learning.

[j]EMA: ecological momentary assessment.

[k]SMART: sequential multiple assignment randomized trial.

Discussion

Principal Findings

This narrative review synthesized engagement-supportive mechanisms across 32 heterogeneous studies [8,9,12-14,24-27,37,43-48,50-65]. Dynamic adaptation, conversational interaction, static tailoring, and social or co-design support were the most frequently represented mechanisms and were commonly combined with contextual or sensing components, although this pattern partly reflects the search focus.

Beyond these frequencies, 3 broader patterns concerned the fit between users' needs, the system's decisions, and the support delivered. Firstly, engagement-supportive mechanisms generally operated as layered systems in which sensing or user input informed a decision process and support was delivered through conversational, social, gamified, or other interfaces. Secondly, personalization separated into relatively static tailoring based on predefined characteristics and dynamic adaptation responsive to changing user information or context. Lastly, engagement was represented through behavioral, experiential, relational, and contextual evidence, alongside contributions to system optimization.

Interpretation and Comparison With Existing Literature

Across the context-aware and JITAI studies, the findings suggest that adaptive decision rules should account for users' needs and receptivity rather than simply increasing the frequency or complexity of prompting. Adaptive recommendation, reinforcement learning, and wearable-triggered conversational systems illustrated potential to improve message relevance and reduce prompting burden, alongside limitations involving contextual interpretation and conversational depth [37,56,64]. Receptivity-informed timing and context-aware delivery were associated with acceptability or adherence [43,48], while prompts triggered by a stated support need were rated as better timed and more helpful than distress-triggered prompts [44]. However, the quasi-experimental timing study did not isolate the contribution of the timing strategy [43]. Considered together with the wider JITAI literature [29-31,66], these findings suggest

XSL•FO
RenderX

that decision rules must represent both vulnerability and opportunity: a detectable stress state does not necessarily mean that a person is able or willing to act.

Biometric responsiveness was less common and generally functioned as enabling infrastructure rather than as a standalone engagement strategy. Physiological, behavioral, or affective signals contributed to stress detection, contextual timing, or adaptive delivery [37,48,56,59], but frequent mobile sensing could create fatigue, and users valued control over data collection [60]. Its potential to make support more contextually relevant while reducing unnecessary prompts must therefore be balanced against device and monitoring burden, privacy, and inference accuracy.

Personalization showed potential benefits alongside interaction costs. Tailored, culturally adapted, and guided interventions were associated with greater perceived relevance or acceptability, reduced stress, or improved emotional regulation [13,14,57,63,64], although digital augmentation did not consistently overcome engagement difficulties [12]. Demographic tailoring reduced disengagement, whereas tailoring based on perceived need or intended next steps did not, and collecting extra tailoring data increased disengagement [52]. Complementary evidence outside the stress-focused corpus showed that interface optimization and reminders improved engagement, whereas adaptive human support improved satisfaction but not engagement [49]. Collectively, these findings suggest that the value of personalization depends on both the information used and the effort required to provide it, qualifying the assumption that more complex adaptation necessarily outperforms simpler improvements in interface design, guidance, or reminders.

Conversational systems provided another pathway to engagement by creating responsive and potentially relational forms of interaction. Alongside favorable acceptability or mental health outcomes [24,63], higher rates of user-confirmed accurate reflections with generative responses [26] and short-term gains in self-care intentions and mental health literacy with topic-based chatbots [61] support assessing conversational responsiveness and self-care learning separately. User experience and alliance findings linked relational value to immediacy and availability [27]; conversational control; input suited to users' preferred expression; caring responses; and a private, nonjudgmental setting [53], alongside cultural adaptation and human guidance [63]. Perceived alliance was associated with engagement, including an indirect path through engagement to symptom change [46]. These associative and qualitative findings do not establish a causal relational mechanism, and concerns about accuracy and authenticity [27] caution against equating rapport with clinical reliability or safety.

Relational engagement was not limited to interaction with chatbots. Clinical or peer support, social motivation, validation, normalization, belonging, and the practical application of therapeutic insights contributed to engagement [8,47,63,65]. Home-based delivery offered flexibility, whereas center-based programs provided social motivation [58], suggesting that delivery choices should account for practical circumstances and preferences for social interaction. Limited support availability and time constraints [44], cognitive overload and low motivation [47], and increased number of days of app use following targeted coaching among participants with low initial use [50] suggest that human support should address the particular barriers preventing participation. Relational issues also arose during intervention development: co-design revealed how power and competing priorities influenced whose perspectives informed the intervention [9].

Gamification produced variable findings depending on the form and purpose of the game elements. Gamified breathing increased practice frequency [55], dynamic tailoring increased narrative presence without improving measured immersion, while play time and completed sessions were lower in the tailored condition [59], and team competition produced early affective benefits that weakened over time [65]. A co-designed narrative game was associated with lower depressive symptoms than an active control at 12 months, but anxiety did not improve, and engagement quality or narrative resonance were not measured [51]. Considered alongside work on motivationally aligned gamification [19,20], these contrasts indicate that practice, narrative presence, immersion, and affect require separate assessment and that rewards or greater usage alone do not establish meaningful or sustained engagement [7]. The limited narrative evidence [51,59] does not yet establish whether immersion produces sustained use or whether immersion or sustained use mediates mental health outcomes.

AI-related evidence likewise differed in function and maturity. LLM prompting, architectural benchmarking, fuzzy logic assessment, and reinforcement learning studies contributed evidence about system design, assessment, and optimization [25,37,54,62] but did not necessarily demonstrate sustained engagement in everyday use. Response-adaptive allocation concerned research allocation across participants and not within-person adaptation of the standardized interventions they received [45]. The corpus consequently provides evidence of feasibility and experiential potential but limited evidence that AI itself independently improves sustained engagement. AI should therefore be understood according to its functional role within the intervention or research design rather than being treated as a single engagement mechanism.

These findings align with broader digital health and HCI literature showing that engagement cannot be reduced to usage metrics alone [2,5-7,33-36]. Behavioral activity, emotional investment, perceived usefulness, and relational value may diverge across users and contexts. Frameworks emphasize interactions among motivational, immersive, and social processes [7,38-40], but the corpus provided limited evidence on which combinations support meaningful or sustained engagement. Emotional safety, therapeutic credibility, burden, privacy, and contextual inference [28,67,68] may help explain why advances in sensing, adaptive delivery, and AI-enabled support have not been matched by evaluation of users' experiences and continued participation over time.

SDT as an Interpretive Lens

SDT helps connect these mechanism-level findings without treating technology as intrinsically motivating. Autonomy may

XSL•FO
RenderX

be supported when culturally or individually tailored content increases relevance, users retain meaningful control, and prompts fit their preferences or circumstances [12-14,43,44,52,57,64]. It may be undermined when additional questions, opaque inference, excessive sensing, or poorly timed notifications increase burden [43,44,52,60]. Competence may be supported through achievable practice, feedback, graduated challenge, and opportunities to apply skills, including well-integrated gamified or narrative experiences [51,55,59,65]. Relatedness is reflected in conversational alliance, nonjudgmental support, belonging, co-design, blended care, and selective human coaching [8,9,24,26,27,44,46,47,50,53,58,63]. These are theory-informed alignments, not evidence that SDT was directly tested across the corpus. Their explanatory value lies in distinguishing mechanisms that merely increase exposure from designs that may support volitional, competent, and socially connected participation.

Implications for Design and Evaluation

Designers should specify what information an adaptive system collects, how that information affects its decisions, and how the resulting support is delivered. Receptivity and declared support needs should be evaluated alongside stress detection, and sensing and tailoring burden should be minimized. Users should be able to understand, control, or contest consequential recommendations. For conversational systems, this also requires attention to accuracy, privacy, explainability, perceived authenticity, emotional dependency, and safe escalation to human support [22,23,26,27,33,34,46,53]. Independent replication is important where intervention developers or commercially affiliated researchers participate in evaluations.

Evaluation should combine behavioral telemetry with experiential, relational, and implementation outcomes matched to the intended engagement pathway. Relevant outcomes include alliance in conversational systems, belonging, validation, and practical use in social platforms and repetition, competence, immersion, or social motivation in gamified systems. Mixed methods designs can connect use or sensing compliance with users' accounts of burden, relevance, and perceived support [44,47,55,57,60,69]. Qualitative, participatory, autoethnographic, and duoethnographic approaches can reveal power, emotional friction, cultural fit, and trust that usage data may not capture [9,14,27,56,58,70-72]. Future studies should measure psychological need support and motivational quality explicitly if SDT is presented as a design rationale.

Controlled and longitudinal studies can estimate behavioral and clinical effects, with microrandomized trials examining proximal timing effects and factorial trials testing component effects. Sequential multiple assignment randomized trials address adaptive escalation, while response-adaptive designs concern allocation efficiency. Multiphase evaluation can connect modeling, feasibility testing, and controlled comparison with real-world implementation [73], where Proctor taxonomy and the RE-AIM (Reach, Effectiveness, Adoption, Implementation, and Maintenance) framework address acceptability, appropriateness, feasibility, adoption, and maintenance [74,75]. Consistent reporting of engagement definitions, intervention exposure, and adaptive decision rules, together with longer follow-up in routine settings, would improve comparison without reducing engagement to a single universal metric.

Limitations

This is a narrative review informed by structured procedures and not an exhaustive systematic review or meta-analysis. The searches were restricted to English-language records and selected databases, and the mechanism-focused terminology may have missed relevant interventions. Excluding app store–only evaluations, commercial deployments, and other gray literature may also underrepresent engagement practices used in widely adopted consumer DMHIs [2,11].

Study designs, populations, and engagement measures were heterogeneous, and several interventions combined mechanisms. Mechanism counts should therefore not be interpreted as comparative effectiveness estimates. The narrative synthesis also does not establish that engagement mechanisms caused the reported behavioral or clinical outcomes when those mechanisms were not experimentally isolated.

Most intervention periods were brief; only 10 studies reported assessment beyond the immediate intervention or primary postassessment period, and the evidence remains concentrated in high-income settings. These limitations constrain conclusions about sustained engagement and generalizability. Although the studies from Jordan [57,63] contribute important culturally adapted evidence, the limited representation of lower-resource settings means that assumptions about connectivity, digital literacy, intervention burden, social support, and culturally appropriate engagement cannot be generalized confidently.

Two technical preprints were retained as narrow exceptions because of their emerging architectural contributions [25,37], but they were not treated as clinical effectiveness evidence. The insomnia trial [49] and 4 reviews and meta-analyses [33-36] were used only as external support and were not included in the study tables, mechanism counts, or denominators.

Conclusions

Engagement in stress-focused DMHIs can be understood as the fit among users' needs and opportunities, the system's sensing and decision logic, and the way support is delivered. Adaptive timing, conversational interaction, relational support, motivational design, and meaningful tailoring can contribute to that fit, but none should be assumed effective merely because it is technically sophisticated. The distinction between static tailoring and dynamic adaptation and between behavioral use and experiential or relational engagement should remain explicit in both design and evaluation.

The broader requirement is to build systems that are useful without being intrusive, responsive without being opaque, and scalable without weakening human oversight. Inclusive co-design, independent evaluation, and longitudinal mixed methods evidence remain priorities for translating technological promise into meaningful, sustained, and ethically responsible engagement in stress-management DMHIs.

XSL•FO
RenderX

Generative AI Use

ChatGPT (GPT-5; OpenAI) was used in the preparation of this manuscript in a limited capacity to improve linguistic editing and phrasing. All scientific content, study design, data collection, analysis, interpretation, and final decisions remain the sole responsibility of the human authors.

Funding

This research received no external funding.

Data Availability

All data analyzed in this review were derived from published sources cited in the manuscript. Study-level characteristics, mechanism coding, and synthesized findings are presented in Tables 2-5. No participant-level dataset was generated.

Authors' Contributions

Conceptualization: MM (lead), SG (equal)
Data curation: MM (lead), SAP (supporting)
Formal analysis: MM
Investigation: MM (lead), SAP (supporting)
Methodology: MM (lead), SAP (supporting)
Supervision: SG
Validation: SAP (lead), SG (equal)
Visualization: MM
Writing – original draft: MM
Writing – review & editing: SAP (lead), SG (equal)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Database search strategies and research question operationalization.
[DOCX File, 8 KB - humanfactors_v13i1e90855_app1.docx]

Checklist 1
PRISMA-P checklist.
[DOCX File, 15 KB - humanfactors_v13i1e90855_app2.docx]

References

1. Firth J, Torous J, Nicholas J, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298. [doi: 10.1002/wps.20472] [Medline: 28941113]
2. Borghouts J, Eikey E, Mark G, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387. [doi: 10.2196/24387] [Medline: 33759801]
3. Bress JN, Falk A, Schier MM, et al. Efficacy of a mobile app-based intervention for young adults with anxiety disorders: a randomized clinical trial. JAMA Netw Open 2024 Aug 1;7(8):e2428372. [doi: 10.1001/jamanetworkopen.2024.28372] [Medline: 39163044]
4. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11. [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
5. Chaudhari S, Ghanvatkar S, Kankanhalli A. Personalization of intervention timing for physical activity: scoping review. JMIR Mhealth Uhealth 2022 Feb 28;10(2):e31327. [doi: 10.2196/31327] [Medline: 35225811]
6. Lipschitz JM, Van Boxtel R, Torous J, et al. Digital mental health interventions for depression: scoping review of user engagement. J Med Internet Res 2022 Oct 14;24(10):e39204. [doi: 10.2196/39204] [Medline: 36240001]
7. Doherty K, Doherty G. Engagement in HCI: conception, theory and measurement. ACM Comput Surv 2018;51(5):99. [doi: 10.1145/3234149]
8. Alvarez-Jimenez M, Nicholas J, Valentine L, et al. A national evaluation of a multi-modal, blended, digital intervention integrated within Australian youth mental health services. Acta Psychiatr Scand 2025 Mar;151(3):317-331. [doi: 10.1111/acps.13751] [Medline: 39260824]

XSL•FO
RenderX

9. Miller E, Hayter C, Manchester H. Facilitating co-design in healthcare: a collaborative autoethnography of personalities, power, and competing priorities. CoDesign 2026 Jan 2;22(1):129-143. [doi: 10.1080/15710882.2025.2460037]
10. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
11. Wu A, Scult MA, Barnes ED, Betancourt JA, Falk A, Gunning FM. Smartphone apps for depression and anxiety: a systematic review and meta-analysis of techniques to increase engagement. NPJ Digit Med 2021 Feb 11;4(1):20. [doi: 10.1038/s41746-021-00386-8] [Medline: 33574573]
12. Weintraub MJ, Ichinose MC, Zinberg JL, et al. Effects of a multifamily app-enhanced cognitive-behavioral therapy delivered via telehealth for parents of adolescents with mood or psychotic spectrum disorders. Behav Ther 2025 Sep;56(5):892-906. [doi: 10.1016/j.beth.2025.02.002] [Medline: 40849116]
13. Sîrbu V, David OA, Sanchez-Lopez A, Blanco I. Comparative efficacy of PsyPills and OCAT mobile psychological interventions in reducing depressive, anxiety and stress symptoms: a blinded randomized clinical trial. J Affect Disord 2025 Jan 15;369:945-953. [doi: 10.1016/j.jad.2024.10.052] [Medline: 39426508]
14. Valentine L, Arnold C, Nicholas J, et al. A personalized, transdiagnostic smartphone app (Mello) targeting repetitive negative thinking for depression and anxiety: qualitative analysis of young people's experience. J Med Internet Res 2024 Nov 27;26:e63732. [doi: 10.2196/63732] [Medline: 39602793]
15. Castro Ribeiro T, Sobregrau Sangrà P, García Pagès E, et al. Assessing effectiveness of heart rate variability biofeedback to mitigate mental health symptoms: a pilot study. Front Physiol 2023;14:1147260. [doi: 10.3389/fphys.2023.1147260] [Medline: 37234414]
16. Goessl VC, Curtiss JE, Hofmann SG. The effect of heart rate variability biofeedback training on stress and anxiety: a meta-analysis. Psychol Med 2017 Nov;47(15):2578-2586. [doi: 10.1017/S0033291717001003] [Medline: 28478782]
17. Litvin S, Saunders R, Maier MA, Lüttke S. Gamification as an approach to improve resilience and reduce attrition in mobile mental health interventions: a randomized controlled trial. PLoS One 2020;15(9):e0237220. [doi: 10.1371/journal.pone.0237220] [Medline: 32877425]
18. Litvin S, Saunders R, Jefferies P, Seely H, Pössel P, Lüttke S. The impact of a gamified mobile mental health app (eQuoo) on resilience and mental health in a student population: large-scale randomized controlled trial. JMIR Ment Health 2023 Jul 21;10:e47285. [doi: 10.2196/47285] [Medline: 37477955]
19. Tondello GF, Wehbe RR, Diamond L, Busch M, Marczewski A, Nacke LE. The gamification user types hexad scale. In: CHI PLAY '16: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in PLAY: Association for Computing Machinery; 2016:229-243. [doi: 10.1145/2967934.2968082]
20. Tondello GF, Orji R, Nacke LE. Recommender systems for personalized gamification. In: Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization: Association for Computing Machinery; 2017. [doi: 10.1145/3099023.3099114]
21. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. J Med Internet Res 2020 Jul 13;22(7):e16021. [doi: 10.2196/16021] [Medline: 32673216]
22. Sui A, Sui W, Liu S, Rhodes R. Ethical considerations for the use of consumer wearables in health research. Digit Health 2023;9:20552076231153740. [doi: 10.1177/20552076231153740] [Medline: 36756643]
23. De Freitas J, Cohen IG. The health risks of generative AI-based wellness apps. Nat Med 2024 May;30(5):1269-1275. [doi: 10.1038/s41591-024-02943-6] [Medline: 38684859]
24. Heinz MV, Mackin DM, Trudeau BM, et al. Randomized trial of a generative AI chatbot for mental health treatment. NEJM AI 2025 Mar 27;2(4):AIoa2400802. [doi: 10.1056/AIoa2400802]
25. Guo Q, Tang J, Sun W, Tang H, Shang Y, Wang W. SouLLMate: an adaptive LLM-driven system for advanced mental health support and assessment, based on a systematic application survey. arXiv. Preprint posted online on Oct 6, 2024. [doi: 10.48550/arXiv.2410.11859]
26. Campellone TR, Flom M, Montgomery RM, et al. Safety and user experience of a generative artificial intelligence digital mental health intervention: exploratory randomized controlled trial. J Med Internet Res 2025 May 23;27:e67365. [doi: 10.2196/67365] [Medline: 40408143]
27. Siddals S, Torous J, Coxon A. "It happened to be the perfect thing": experiences of generative AI chatbots for mental health. NPJ Ment Health Res 2024 Oct 27;3(1):48. [doi: 10.1038/s44184-024-00097-4] [Medline: 39465310]
28. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022 Apr 26;24(4):e35120. [doi: 10.2196/35120] [Medline: 35471414]
29. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
30. Suh J, Howe E, Lewis R, et al. Toward tailoring just-in-time adaptive intervention systems for workplace stress reduction: exploratory analysis of intervention implementation. JMIR Ment Health 2024 Sep 12;11:e48974. [doi: 10.2196/48974] [Medline: 39264703]

XSL•FO
RenderX

31. van Genugten CR, Thong MSY, van Ballegooijen W, et al. Beyond the current state of just-in-time adaptive interventions in mental health: a qualitative systematic review. Front Digit Health 2025;7:1460167. [doi: 10.3389/fdgth.2025.1460167] [Medline: 39935463]
32. Valentine L, D'Alfonso S, Lederman R. Recommender systems for mental health apps: advantages and ethical challenges. AI & Soc 2023 Aug;38(4):1627-1638. [doi: 10.1007/s00146-021-01322-w]
33. Smith KA, Ward T, Lambe S, et al. Engagement and attrition in digital mental health: current challenges and potential solutions. NPJ Digit Med 2025 Jul 2;8(1):398. [doi: 10.1038/s41746-025-01778-w] [Medline: 40604240]
34. Torous J, Linardon J, Goldberg SB, et al. The evolving field of digital mental health: current evidence and implementation issues for smartphone apps, generative artificial intelligence, and virtual reality. World Psychiatry 2025 Jun;24(2):156-174. [doi: 10.1002/wps.21299] [Medline: 40371757]
35. Liu C, Torous J, Fuller-Tyszkiewicz M, et al. Uptake, adherence, and attrition in clinical trials of depression and anxiety apps: a systematic review and meta-analysis. JAMA Psychiatry 2026 Jan 1;83(1):43-50. [doi: 10.1001/jamapsychiatry.2025.3439] [Medline: 41259035]
36. Zainal NH, Wang V, Garthwaite B, Curtiss JE. What factors are related to engagement with digital mental health interventions (DMHIs)? A meta-analysis of 117 trials. Health Psychol Rev 2026 Mar;20(1):151-171. [doi: 10.1080/17437199.2025.2547610] [Medline: 40857362]
37. Aqajari SAH, Wang Z, Tazarv A, et al. Enhancing performance and user engagement in everyday stress monitoring: a context-aware active reinforcement learning approach. ACM Trans Comput Healthc 2026. [doi: 10.1145/3837063]
38. Hindman J, Zugai J, Foster K, Raeburn T. Self-determination theory as a framework for research and design of digital applications for nurses' well-being. J Adv Nurs 2026 Jan;82(1):897-906. [doi: 10.1111/jan.16895] [Medline: 40098227]
39. Zhang Z, Sun J, Peters D, Calvo RA. Beyond wellbeing apps: co-designing immersive, embodied, and collective digital wellbeing interventions for healthcare professionals. In: DIS'25: Proceedings of the 2025 ACM Designing Interactive Systems Conference: Association for Computing Machinery; 2025. [doi: 10.1145/3715336.3735681]
40. Ijaz K, Ahmadpour N, Wang Y, Calvo RA. Player experience of needs satisfaction (PENS) in an immersive virtual reality exercise platform describes motivation and enjoyment. Int J Hum-Comput Interact 2020 Aug 8;36(13):1195-1204. [doi: 10.1080/10447318.2020.1726107]
41. STARS@UNIPD funding programme—2021: guide for applicants, annex 2 ethics issues table. : University of Padua; 2021 URL: https://wwwassets.unipd.it/sites/default/files/pdf_imported_files/STARS_2021_GFA_TEMPLATES_UPDATE_3.pdf [accessed 2026-10-21]
42. (For authors) institutional research board/research ethics board and informed consent. JMIR Publications. 2025. URL: https://support.jmir.org/hc/en-us/articles/360048970851--for-authors-Institutional-Research-Board-Research-Ethics-Board-and-Informed-Consent [accessed 2026-10-21]
43. Kor PPK, Tsang APL, Cheung DSK, et al. Developing and testing a brief mindfulness just-in-time adaptive intervention to reduce stress among caregivers of people with dementia: quasi-experimental study. JMIR Aging 2026 Jun 5;9:e87316. [doi: 10.2196/87316] [Medline: 42247635]
44. Elmer T, Wolf M, Snippe E, Scholz U. A social support just-in-time adaptive intervention for individuals with depressive symptoms: feasibility study with a microrandomized trial design. JMIR Ment Health 2025 Aug 26;12:e74103. [doi: 10.2196/74103] [Medline: 40857095]
45. Newby J, Gupta S, Hoon L, et al. Brief digital interventions for psychological distress: an AI-enhanced response-adaptive randomized clinical trial. JAMA Netw Open 2025 Oct 1;8(10):e2540502. [doi: 10.1001/jamanetworkopen.2025.40502] [Medline: 41171275]
46. Shoshani A, Gurfinkel B, Kor A, et al. Efficacy of a conversational AI agent for psychiatric symptoms and digital therapeutic alliance: a randomized clinical trial. JAMA Netw Open 2026 Apr 1;9(4):e266713. [doi: 10.1001/jamanetworkopen.2026.6713] [Medline: 41979879]
47. Valentine L, Nicholas J, Sorenson R, et al. Sustained engagement with a digital youth mental health platform: a mixed-methods study. Internet Interv 2026 Mar;43:100899. [doi: 10.1016/j.invent.2025.100899] [Medline: 41551031]
48. Dai G, Li X, Ying F, Yao C. A design framework for context-aware digital micro-interventions: integrating emotion recognition and just-in-time adaptive intervention for workplace stress management. IEEE Trans Affective Comput :1-23. [doi: 10.1109/TAFFC.2026.3711335]
49. Hentati A, Hentati Isacsson N, Rosén A, et al. Effects of intervention design on engagement and outcomes in digital self-help for insomnia - factorial RCT. NPJ Digit Med 2025 Jul 8;8(1):416. [doi: 10.1038/s41746-025-01839-0] [Medline: 40628992]
50. Chow PI, You W, Shaffer KM, et al. Digital mental health treatment and symptoms of depression and anxiety in breast cancer survivors: a randomized clinical trial. JAMA Netw Open 2026 Jul 1;9(7):e2623871. [doi: 10.1001/jamanetworkopen.2026.23871] [Medline: 42475097]
51. Barry CM, Boomer T, Haile K, Xie H, Fiellin LE. A digital behavioral health game for adolescent mental health: a prespecified secondary analysis of a randomized clinical trial. JAMA Netw Open 2026 Aug 3;9(8):e2629173. [doi: 10.1001/jamanetworkopen.2026.29173] [Medline: 42606863]

XSL•FO
RenderX

52. Rouvere J, Griffith Fillipo IR, Romanelli M, et al. Personalization strategies for increasing engagement with digital mental health resources: sequential multiple assignment randomized trial. JMIR Ment Health 2025 Nov 4;12:e73188. [doi: 10.2196/73188] [Medline: 41187311]
53. Xu Z, Lee YC, Stasiak K, Warren J, Lottridge D. The digital therapeutic alliance with mental health chatbots: diary study and thematic analysis. JMIR Ment Health 2025 Oct 10;12:e76642. [doi: 10.2196/76642] [Medline: 41072011]
54. Wang Z, Li H, Huang D, Kim HS, Shin CW, Rahmani AM. HealthQ: unveiling questioning capabilities of LLM chains in healthcare conversations. Smart Health 2025 Jun;36:100570. [doi: 10.1016/j.smhl.2025.100570]
55. Sadprasid B, Mei A, Mariakakis A, Bateman S, Chevalier F. Leveraging idle games to incentivize intermittent and frequent practice of deep breathing. In: CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems: Association for Computing Machinery; 2024:1-17. [doi: 10.1145/3613904.3642430]
56. Neupane S, Dongre P, Gracanin D, Kumar S. Wearable meets LLM for stress management: a duoethnographic study integrating wearable-triggered stressors and LLM chatbots for personalized interventions. : Association for Computing Machinery; 2025 Apr 26 Presented at: CHI EA '24: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems. [doi: 10.1145/3706599.3720197]
57. Dardas LA, Al-Leimon O, Gladstone T, Dabbas AA, Alammouri I, Van Voorhees B. Validating a digital depression prevention program for adolescents in Jordan: cultural adaptation and user testing in a randomized controlled trial. Front Psychiatry 2025;16:1529006. [doi: 10.3389/fpsyt.2025.1529006] [Medline: 40012714]
58. Rutherford H, Dale M, Wootton S, et al. Experiences of home-based pulmonary rehabilitation with mhealth and centre-based pulmonary rehabilitation in people with chronic obstructive pulmonary disease: a qualitative study. Health Expect 2025 Apr;28(2):e70181. [doi: 10.1111/hex.70181] [Medline: 40019189]
59. Schwarz A, Cardon G, Chastin S, et al. Does dynamic tailoring of a narrative-driven exergame result in higher user engagement among adolescents? results from a cluster-randomized controlled trial. Int J Environ Res Public Health 2021 Jul 12;18(14):7444. [doi: 10.3390/ijerph18147444] [Medline: 34299895]
60. Slade C, Benzo RM, Washington P. Design guidelines for improving mobile sensing data collection: prospective mixed methods study. J Med Internet Res 2024 Nov 18;26:e55694. [doi: 10.2196/55694] [Medline: 39556828]
61. Tong ACY, Wong KTY, Chung WWT, Mak WWS. Effectiveness of topic-based chatbots on mental health self-care and mental well-being: randomized controlled trial. J Med Internet Res 2025 Apr 30;27:e70436. [doi: 10.2196/70436] [Medline: 40306635]
62. Ying G. Mental health assessment model for college students using circular q-rung orthopair fuzzy Muirhead means and MULTIMOORA method. IEEE Access 2025;13:35433-35451. [doi: 10.1109/ACCESS.2025.3542804]
63. de Graaff AM, Habashneh R, Fanatseh S, et al. Evaluation of a guided chatbot intervention for young people in Jordan: feasibility randomized controlled trial. JMIR Ment Health 2025 Feb 5;12:e63515. [doi: 10.2196/63515] [Medline: 39908538]
64. Coppens I, De Pessemier T, Martens L. Exploring the added effect of three recommender system techniques in mobile health interventions for physical activity: a longitudinal randomized controlled trial. User Model User-Adap Inter 2024 Nov;34(5):1835-1890. [doi: 10.1007/s11257-024-09407-z]
65. Wang J, Fang Y, Frank E, et al. Effectiveness of gamified team competition as mHealth intervention for medical interns: a cluster micro-randomized trial. NPJ Digit Med 2023 Jan 11;6(1):4. [doi: 10.1038/s41746-022-00746-y] [Medline: 36631665]
66. Mishra V, Künzler F, Kramer JN, Fleisch E, Kowatsch T, Kotz D. Detecting receptivity for mHealth interventions in the natural environment. Proc ACM Interact Mob Wearable Ubiquitous Technol 2021 Jun;5(2):1-24. [doi: 10.1145/3463492] [Medline: 34926979]
67. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567. [doi: 10.2196/14567] [Medline: 31573916]
68. Jardine J, Nadal C, Robinson S, Enrique A, Hanratty M, Doherty G. Between rhetoric and reality: real-world barriers to uptake and early engagement in digital mental health interventions. ACM Trans Comput-Hum Interact 2024 Apr 30;31(2):1-59. [doi: 10.1145/3635472]
69. Creswell JW, Plano Clark VL. Designing and Conducting Mixed Methods Research: SAGE Publications; 2017. URL: https://www.sagepub.com/shop/buy-a-book/designing-and-conducting-mixed-methods-research-3-241842 [accessed 2026-10-23]
70. Ellis C, Adams TE, Bochner AP. Autoethnography: an overview. Hist Soc Res 2011;36(4):273-290. [doi: 10.17169/fqs-12.1.1589]
71. Kaltenhauser A, Stefanidi E, Schöning J. Playing with perspectives and unveiling the autoethnographic kaleidoscope in HCI – a literature review of autoethnographies. In: CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems: Association for Computing Machinery; 2024:1-20. [doi: 10.1145/3613904.3642355]
72. Ezimora I, Lundberg T, Miars D, et al. Reflections of foster youth engaging in the co-design of digital mental health technology: duoethnography study. JMIR Form Res 2025 Jan 20;9:e53231. [doi: 10.2196/53231] [Medline: 39832159]
73. Bowen DJ, Kreuter M, Spring B, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457. [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]

XSL•FO
RenderX

74. Proctor E, Silmere H, Raghavan R, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76. [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
75. Glasgow RE, Harden SM, Gaglio B, et al. RE-AIM planning and evaluation framework: adapting to new science and practice with a 20-year review. Front Public Health 2019;7:64. [doi: 10.3389/fpubh.2019.00064] [Medline: 30984733]

Abbreviations

CBT: cognitive behavioral therapy
DHI: digital health intervention
DMHI: digital mental health intervention
EMA: ecological momentary assessment
HCI: human-computer interaction
JITAI: just-in-time adaptive intervention
LLM: large language model
PRISMA-P: Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols
RCT: randomized controlled trial
RE-AIM: Reach, Effectiveness, Adoption, Implementation, Maintenance
SDT: self-determination theory

Edited by S Law; submitted 05.Jan.2026; peer-reviewed by H Rathore, K Williams, Z Zhang; revised version received 06.Oct.2026; accepted 06.Oct.2026; published 02.Nov.2026.

Please cite as:
Maminimini M, Piffer SA, Gabrielli S
Engagement-Supportive Mechanisms in Stress-Focused Digital Mental Health Interventions With Adaptive, Conversational, and Context-Aware Systems: Narrative Review
JMIR Hum Factors 2026;13:e90855
URL: https://humanfactors.jmir.org/2026/1/e90855
doi:10.2196/90855

© Munesu Maminimini, Sofia Asteria Piffer, Silvia Gabrielli. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 2.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile Apps Supporting Medication Adherence in Substance Use Disorder Treatment: Narrative Review

Jane J Kim[1*], MSc; Lua Boschman[2*], BA; Pouya Azar[3], MD; Borna Noureddin[4], PhD

[1]Department of Psychiatry, University of British Columbia, 2775 Laurel Street, Vancouver, BC, Canada

[2]School of Computing and Academic Studies, British Columbia Institute of Technology, Vancouver, BC, Canada

[3]Department of Psychiatry, Vancouver General Hospital, Vancouver, BC, Canada

[4]Faculty, Bachelor of Technology Program, British Columbia Institute of Technology, Vancouver, BC, Canada

[*]these authors contributed equally

Corresponding Author:
Jane J Kim, MSc
Department of Psychiatry, University of British Columbia, 2775 Laurel Street, Vancouver, BC, Canada

Abstract

Background: Substance use disorders (SUDs) and related public health harms have risen significantly in Canada over the past decade. While existing medications for addiction treatment have shown efficacy, poor treatment adherence remains a challenge. With growing rates of smartphone adoption, mobile apps represent an opportunity for digital interventions that can support medication adherence.

Objective: This study provides a narrative review of current app-based interventions supporting medication adherence for individuals with SUDs and analyzes the common features for advantages and limitations.

Methods: A literature search was conducted on the MEDLINE, APA PsycInfo, CINAHL, PubMed, Scopus, and Web of Science Core Collection databases from inception to October 2026 using the combination of keywords related to *substance use*, *medication adherence*, and *mobile app*. Selected papers focused on specific apps to support medication adherence for SUDs. Common technological features, potential advantages, and limitations of the identified technology apps were extracted and analyzed.

Results: This review examined 22 articles. A total of 7 feature types were identified: reminders or reinforcers (eg, contingency management), patient-health care professional communication, information, community boards to connect with support or other anonymous patients, self-monitoring or goal tracking with patient self-report, external tracking (eg, location or electronic medication dispensing), and gamification. The most frequently occurring features across the apps were reminder messages (14/22, 64%), self-assessment and goal-tracking tools (n=14, 64%), and patient-provider communication (n=13, 59%). Across the studies, interpretation of efficacy was limited, as many studies were not designed to assess effectiveness. They also examined small, geographically and demographically homogeneous samples with substantial reliance on self-reported outcomes and incomplete data. In contrast, feasibility and acceptability were consistently high, with generally positive usability ratings and participant engagement across interventions.

Conclusions: This narrative review summarizes the technical features and various advantages and limitations of apps designed to support medication adherence for SUDs in the literature. This information provides a useful tool to guide the development and selection of future apps designed to improve medication adherence outcomes. Future studies are needed to determine which specific app features and implementation strategies are most effective in supporting adherence for specific populations and medications.

(JMIR Hum Factors 2026;13:e104420) doi:10.2196/104420

KEYWORDS

mHealth; mobile health; substance use disorders; mobile app; adherence; treatment

Introduction

Substance use disorders (SUDs) and related public health harms, such as hospitalization, health care system strains, and deaths, have risen substantially both in Canada and globally during the last decade. In Canada, this has been reflected in an estimated CAD $13.4 billion (CAD $1=US $0.71 as of November 9, 2026) in annual direct health care costs and more than 190,000 annual acute hospital stays in 2023 [1]. While multiple substance types contribute to this burden, a class of particular concern is opioids and their attendant risk for fatal overdose, with a rapid rise of 2900 opioid-related Canadian deaths in 2016 to 8049 in 2023 [2]. Pharmacological management has demonstrated safety and efficacy for many SUDs, but poor adherence remains a limiting

XSL•FO
RenderX

factor behind low retention rates observed relative to patients with other conditions [3-5]. *Adherence* refers to the extent to which patients follow prescribed medication regimens as directed by a health care provider. Examples of poor adherence in the context of SUD care can include skipping doses, taking doses that are higher or lower than prescribed, or discontinuation of treatment [6].

In recent years, the rapid proliferation of mobile technologies and increasing rates of smartphone ownership have offered a new opportunity to address the challenge of medication adherence [7]. Given the growing ubiquity of smartphones, the potential of mobile apps to promote adherence by providing point-of-care support and improving convenience to patients with SUDs has been identified [8,9]. Despite these opportunities, only a small number of mobile apps have been developed for this specific purpose, which are additionally limited to a handful of features such as text alerts and delivering monetary rewards [10-12]. Determining the most effective features and functionalities for these apps remains uncertain. A recent review has outlined technology-based medication adherence and monitoring interventions used in psychiatric disorders as a whole, but not specific to SUDs [13]. A compiled summary and assessment of the features of currently available mobile apps targeting medication adherence in SUDs is needed to better understand the emerging evidence base and to inform the design and development of future interventions. The aims of this narrative review are to (1) analyze existing literature on mobile apps targeting medication adherence in addiction and the implemented features for user engagement and (2) identify their benefits and limitations for future work.

Methods

Literature Search

A narrative literature review of existing literature was conducted using the MEDLINE, APA PsycInfo, CINAHL, PubMed, Scopus, and Web of Science Core Collection databases from inception to October 2026, which provide broad coverage of biomedical literature relevant to mobile apps and SUDs. This design allows for a comprehensive synthesis and interpretation of existing literature to gain foundational understanding and identify key insights [13] and was chosen due to the heterogeneity and evolving practices in this area of mobile apps. The key search terms included “substance use,” “misuse,” “abuse,” “dependence,” “addiction,” “drug,” “alcohol,” “cocaine,” “cannabis,” “marijuana,” “heroin,” “opioid,” “opiate,” “amphetamine,” “stimulant,” “mobile application,” “mobile health,” “smartphone application” “telemedicine,” “mHealth,” “eHealth,” “digital health,” “medication adherence,” “medication taking,” “treatment adherence,” “medication compliance,” and “treatment compliance.” The appropriate search terms were selected through a preliminary search of a variety of terms on Google Scholar. When conducting the full search, terms were combined using logical operators “OR” and “AND” and were restricted to searches within titles and abstracts only.

Eligibility Criteria

Studies were included in this narrative review if they described primary research on mobile apps for SUD medication adherence. All study methods were included. Only articles published in English with their full text available were included. Studies were excluded if they did not focus on apps available via mobile devices such as smartphones. Apps not developed for patients with SUDs were also excluded. Study selection was performed manually and independently by 2 reviewers using this set of eligibility criteria. The methodological quality of the included studies was assessed using the Mixed Methods Appraisal Tool [14] (Multimedia Appendix 1).

Data Extraction and Information Synthesis

Data were extracted independently by the authors using an a priori created data extraction form designed for this study, including the target and test populations, major characteristics of each app, patient satisfaction, and efficacy. Extracted data from the final pool of articles were summarized qualitatively and categorized according to features. To categorize app features, we used an inductive approach of three steps: (1) open coding of implementation features in each article; (2) consolidation of codes into higher-order categories using constant comparison and identifying recurring functionalities; and (3) grouping of categories into broader feature categories. After independent categorization by 2 authors, discrepancies were resolved through discussion, and unresolved cases were adjudicated by a senior reviewer through rechecking of the original papers. Thematic analysis was also conducted to identify barriers and facilitators to the different app characteristics. The authors reviewed articles and extracted verbatim text segments detailing barriers and facilitators to app feature uses and their effectiveness associated with medication adherence. Extracted quotes were compared across studies and grouped by similarity to generate overarching themes. Where possible, category labels were retained from the original articles to preserve the authors’ intended conceptualization of the features. Individual features could contribute to multiple themes when they reflected more than 1 relevant concept. All identified codes and themes were discussed with authors, and a consensus was reached.

Results

Study Selection and Characteristics

Of the 22 reviewed articles, the majority (n=18, 82%) were published between March 2020 and May 2025. The identified studies were published primarily in the United States (n=10, 45%), followed by England (n=6, 27%), Canada (n=4, 18%), Ireland (n=1, 4%), and Switzerland (n=1, 4%). A total of 9 (41%) studies assessed or planned to assess for efficacy in the form of randomized trials as a single-arm or parallel-group design. All trials were open label due to the nature of the intervention. Furthermore, 13 studies were nonexperimental, assessing for feasibility or initial acceptability and usability through a pilot trial or qualitative interviews with patients and clinicians (Table 1).

XSL•FO
RenderX

Table . A description of included mobile apps[a].

App name	Target population	Year	Country	Trial type	App aims	Trial population	Results	Limitations
MySafeRx	Individuals with OUD[b]	2018	England	Efficacy	Improve daily medication adherence, reduce medication diversion, enhance treatment outcomes, and avoid patient burdens	12 patients aged 18 - 39 y with OUD and prescribed buprenorphine	72% medication adherence, average 78.8 usability score, overall satisfaction in last week 6.5 ± 0.7 (maximum 7)	Small sample size, treatment dropout, discrepancy between self-report and supervised administration
ART-CHESS	Individuals with HIV and SUDs[c]	2024	Canada	Efficacy	Improve medication adherence to ART[d]	208 participants aged >18 y, receiving medical care for HIV infection, with a history of SUDs	Linked to reduced alcohol and stimulant use and greater medication adherence	Small sample size, missing survey data
A-CHESS	Individuals with SUDs	2016	United States	Efficacy	Improve abstinence from illicit opioid use, medication adherence and engagement with other forms of treatment	414 patients with OUD of at least moderate severity and on methadone, injectable naltrexone, or buprenorphine	Associated with greater abstinence from opioids but not among those using illicit drugs or experiencing withdrawal	Homogeneous sample, unequal sample sizes for different medication types, unavailable dosage information
Not specified	Individuals with OUD	2021	Ireland	Efficacy	Promote buprenorphine treatment adherence	41 adults with OUD and interest in buprenorphine treatment	Incentivized participants more likely to enroll in treatment (71.4%) over nonincentivized (30%), but few adhered to treatment and no between-group differences found in positive urine samples	Low treatment adherence, self-report data, homogeneous patient population
TIES	Individuals with heroin addiction	2020	England	Efficacy	Improve adherence to methadone treatment	10 adults prescribed oral methadone and owning a mobile phone	Found to be feasible and well-liked by patients and health care staff	Low recruitment rates of target population
WEconnect	Individuals with OUD	2024	United States	Efficacy	Provide contingency management and medication adherence	600 patients receiving MOUD[e]	Fewer days of opioid use and longer treatment retention found with app	Nonrandomized participants, homogeneous population, less than 1/5 of sample used app

XSL•FO
RenderX

App name	Target population	Year	Country	Trial type	App aims	Trial population	Results	Limitations
Pillsy	Individuals with OUD	2023	Canada	Feasibility and acceptability	Improve medication adherence	20 participants aged 18 years or older and receiving opioid substitution treatment	Received positive feedback for features that reinforce taking medications (>80%)	Small sample size, homogeneous sample, participants did not use app themselves
Mobile Contingency Management	Individuals with OUD	2025	United States	Feasibility and acceptability	Improve OAT[f] adherence and encourage heroin abstinence	40 participants receiving OAT who continue to use heroin	N/A[g]	Small sample size
Not specified	Individuals with past or current illicit drug use	2017	England	Feasibility and acceptability	Improve retention in care and medication adherence	19 participants older than 18 and with an HIV viral load greater than 1000 copies/mL	Intervention components, mHealth[h] and peer navigation, found to be acceptable and responsive	Small sample size
HOPE	Individuals with OUD	2021	Canada	Feasibility and acceptability	Improve OUD treatment adherence	25 participants with OUD	Found to be feasible with high usability scores	Small sample size
HOPE	Individuals with OUD	2022	England	Feasibility and acceptability	Improve OUD treatment adherence	25 participants at least 18 years old and diagnosed with moderate or severe OUD	Most participants retained in care at 6 months and active on app	Small sample size, limited ability to track retention, no urine drug screen comparisons
APP+	MSM[i] with HIV and StUD[j]	2019	United States	Feasibility and acceptability	Achieve optimal Anti-Retroviral Therapy adherence	90 MSM living with HIV who use stimulants	Higher self-reported ART[d] adherence in intervention group (89%) vs control (77.2%) in 30 days at 4-month time point	Small sample size, not all participants in intervention arm provide self-reports
DynamiCare	Individuals with SUD	2021	United States	Feasibility and acceptability	Improve treatment adherence through contingency management	61 patients with AUD[k] of at least moderate severity and alcohol as primary substance	66% of participants exhibited high utilization and adherence, found to be feasible to use in community treatment programs; higher treatment retention rates found with app than without	Small and homogeneous sample, missing satisfaction data, low and unequal participation rates in interviews

XSL•FO
RenderX

App name	Target population	Year	Country	Trial type	App aims	Trial population	Results	Limitations
DynamiCare	Individuals with OUD	2022	Switzerland	Feasibility and acceptability	Promote adherence to buprenorphine treatment	20 participants with OUD and history of stimulant use	High adherence found throughout the study, with decreasing trend over time; all 20 participants retained throughout study	Small sample with no baseline data collected; 40% classified as not high risk for treatment dropout
PROCare Recovery	Individuals with OUD	2022	Canada	Feasibility and acceptability	Improve treatment adherence through contingency management	10 patients receiving buprenorphine for OUD	High perceived usability and acceptability	Small and homogenous sample
DIDE	Individuals with SUD	2024	England	Feasibility and acceptability	Improve treatment adherence	24 adult patients with any type of addiction	Good acceptability rates, 70.8% met primary end point and 29.2% attrition rate	Small sample, single center design, short study duration
Not specified	Individuals with OUD	2020	England	Feasibility and acceptability	Improve buprenorphine adherence	Participants aged 18 years or older, within their first 28 days of either starting or restarting prescribed sublingual buprenorphine treatment	N/A	Small sample size
MIAPP	Individuals with OUD	2025	United States	Feasibility and acceptability	Improve buprenorphine adherence	20 participants aged 18 years or older, prescribed buprenorphine, and use of methamphetamine within the past 30 days	N/A	Small sample, single center design
Pillsy	Individuals with OUD	2024	United States	Efficacy	Improve buprenorphine adherence	41 participants prescribed buprenorphine	Medication adherence was nonsignificantly higher in the intervention group (84%) compared to control (75%)	Small and homogenous sample
emocha	Individuals with OUD	2021	United States	Feasibility and acceptability	Improve medication adherence and substance use	40 participants aged 18 to 65 years and prescribed buprenorphine	Most participants (71%) “very satisfied” with application	Pilot study, small sample size

XSL•FO
RenderX

App name	Target population	Year	Country	Trial type	App aims	Trial population	Results	Limitations
ReSET-O	Individuals with OUD	2026	United States	Efficacy	Improve MOUD adherence	48 participants on MOUD (28 on buprenorphine and 20 on methadone)	18.9% rate difference for 24-wk MOUD adherence in the intervention group compared to the control group	Small sample size, underpowered
Recovery Connect	Individuals with OUD	2025	United States	Efficacy	Improve MOUD adherence and treatment continuance	Patients in outpatient MOUD clinics	Patients who downloaded and linked with mental health professional had 81.3% likelihood of retention, compared with 72% (P<.001) among those not linked	No external comparator, short-term outcomes (30 d)

[a]Information presented is restricted to details available in the included papers. Results are N/A for protocol papers.

[b]OUD: opioid use disorder.

[c]SUD: substance use disorder.

[d]ART: antiretroviral therapy.

[e]MOUD: medications for opioid use disorder.

[f]OAT: opioid agonist therapy

[g]N/A: not applicable.

[h]mHealth: mobile health.

[i]MSM: men who have sex with men.

[j]StUD: stimulant use disorder

[k]AUD: alcohol use disorder.

Of 22 studies, 16 (73%) focused on participants with OUD, 3 (14%) on SUD, and 3 (14%) on participants with HIV and concurrent substance use. The most common treatments studied were medications for opioid use disorder (n=17, 77%) and antiretroviral therapy for HIV (n=3, 14%).

Major Themes for App Features and Characteristics

Overview

The identified app features were categorized into 7 major groups based on their technical design and adherence supporting functions: reminders or reinforcers (eg, contingency management), patient-provider communication, information, community boards to connect with support or other anonymous patients, self-monitoring or goal tracking with patient self-report, external tracking (eg, location, video monitoring, or electronic medication dispensing), and gamification (Table 2).

XSL•FO
RenderX

Table . A summary of app features and their description.

App features	Description
Reinforcers	Positive monetary and nonmonetary reinforcers used as rewards for consistent engagement in treatment; monetary rewards included prepaid or gift cards and cell phone data and minutes, often tied to video verifications, and nonmonetary reinforcers included visual streaks and completion bars, timely reminders, and messages of praise, congratulations, or goals
Patient-provider communication	Facilitating patient interactions with health care providers to support treatment engagement; support personnel included individuals with lived experience, counselors, clinicians, and recovery coaches; patients and providers can privately or securely message through the app interface. Optional adjunct support meetings provided daily or weekly
Information	Access to psychoeducation through online learning modules or informational links to institution-specific patient portals and local community or support groups, clinic phone numbers, medication and treatment information
Community boards	Discussion forums for patients to interact with other users and provide emotional, informational, and instrumental support, often moderated by the development team or anonymous
Self-monitoring	Interface and pages for patients to observe and track daily goals, behaviors, and symptoms (eg, craving, anxiety, sleep, and motivation) as measures of data collection and pattern awareness; visualized with charts, graphs, or tailored feedback upon completion
External tracking	Optional location permissions for periodic user tracking and local recommendations around potential high-risk sites such as bars
Gamification	Inclusion of game-like elements in app for additional motivation and satisfaction; examples can include storytelling, collecting rewards or accomplishments, or leaderboards

The following sections outline the common app features gathered from the existing literature. Some features were counted under multiple headings. An example is progress visualizations (eg, a bar graphic showing progress toward a treatment goal), which could be considered as all of the following: a reinforcer, an instance of goal tracking, and an example of gamification. Feature categorizations were dependent on their description in the original text and interpretation of the authors.

Reinforcers

Of 22 studies, 11 (50%) used monetary reinforcers to reward patients for consistently engaging in treatment [9-11,15-21], with 5 of these specifically using contingency management. Contingency management is a behavioral intervention wherein people receive rewards for achieving their treatment and recovery goals [22]. Monetary reinforcers were tied to video submissions of medication adherence in 6 of these apps. In addition, 1 study found the majority (78.9%) of their participants to be amenable to taking a selfie video while taking medication doses [18]. These videos were encrypted and transmitted to a secure web portal, where they were reviewed by staff to verify and deliver incentives. These rewards were typically delivered via prepaid cards with restrictions on eligible spending locations such as liquor stores. Other rewards included gift cards and cell phone data and minutes.

Apps also made use of nonmonetary reinforcers. Of 17 apps, 7 (41%) used progress visualization (eg, completion bars and streak numbers). A total of 5 studies (29%) made explicit mention of messages of praise or congratulations. Moreover, 11 apps (65%) sent treatment reminders delivered as push notifications or SMS text messages. Reminders were one of the most consistent features across apps, varying from simple fixed daily treatment reminders to random test alerts and intelligent reminders based on user data. A total of 3 apps (18%) contained options for users to record their personal motivations, goals, or reasons for gratitude. Users were occasionally prompted to re-engage with this content. Furthermore, 1 app displayed uplifting quotations in its emergency support system [23]. Users suggested further opportunities for improvement, such as displaying self-selected motivations, compatibility with external calendar apps, and messages worded to preserve privacy and minimize disclosure of treatment purpose.

Patient-Provider Communication

Of 22 studies, 13 (59%) used the app to improve patient-provider communication [10,12,17,19,21,23-30]. Design and functionality components included private and secure messaging, session scheduling, and facilitated phone and video calls with preapproved clinicians and peer support coaches. In addition, 1 app sent automated alerts to counselors so they could intervene promptly through tailored messages when relapse risk or treatment disengagement was detected [24]. A total of 6 studies out of 13 (46%) described a dashboard for the clinical team, with features such as patient history, medication adherence, and substance use reports, and the ability to release codes to open an electronic pill dispenser.

XSL•FO
RenderX

Information

A total of 11 apps (50%) included information on medication treatment, clinician contact information, and local community and social support resources [9,19-21,23,24,26,28,29,31,32]. The information was presented in various ways, including cards appearing on the home page, educational learning modules, contact information, links to external resources, frequently asked questions, and videos. Treatment content included evidence-based therapeutic interventions and resources, such as cognitive behavioral therapy skills and medications for opioid use disorder treatment information. Furthermore, 1 study provided participants with the opportunity to earn rewards for accessing resources in their psychoeducational library as a method of helping educate, support, and encourage patients [9].

Community Boards to Connect With Support or Other Anonymous Patients

A total of 5 apps (23%) had a community board for patients to communicate with each other, with the intention of fostering an exchange of support among patients [23-25,27,28]. Participants could communicate with each other on topics of their choice, with monitored discussions to ensure appropriate use. In addition, 1 app allowed participants to join online support meetings facilitated by certified peers, improving access to support spaces. Furthermore, 1 study found their community board to be used by 36% of their cohort, with an overall average for the full cohort of 0.9 posts per participant during their first month of app use [28]. Posts on the community board reporting substance use were also used by 1 study to gauge each participant's risk threshold for relapse and alert a clinician or trigger further support if needed.

Self-Monitoring or Goal Tracking With Patient Self-Report

A total of 14 apps (64%) included self-assessment and goal-tracking tools [9,11,16,18-21,23-25,27-29,31]. These encompassed a wide range of behaviors, including medication adherence, substance use, readiness to change, craving, mood, sleep, activity, and encouraging experiences. Participants of 1 study were encouraged to set daily goals of their own choosing, including both substance use and self-care (eg, attending a treatment session), which in turn served as treatment goals that were financially reinforced upon completion [16]. A total of 4 apps used surveys related to drug use or emotional states as a method of prevention against relapse by offering immediate support and linkage to needed services upon concerning responses [9,20,27,29]. Patients were prompted to complete questionnaires of multiple choices independently or through an optional conversation with a chatbot in 1 app. In addition, 1 study found more than half of the patients (58.3%) to complete 2 or more questionnaires each day they used the app; however, a small percentage also reported experiencing too strong cravings to fill the questionnaires or self-stigma when recording drug use [27]. Participants recommended push notifications and reminders to complete questionnaires. Furthermore, 7 apps (32%) also included visualized assessments, such as graphs, streak records, or progress bars of how many doses each patient reported taking during the last or current week.

External Tracking

A total of 4 apps (18%) featured location tracking: 1 to detect the patients' presence at therapy appointments, 2 to detect their presence at a distributing pharmacy, and 1 to detect the patient's presence at locations that the patient had identified as high risk [10,11,15,24]. For each patient that approached a high-risk location, the app initiated a previously defined recovery process, starting with a beep, a vibration, and then a list of preapproved contacts and options for support. GPS was also used as a method of finding nearby support groups or meetings. Participants had the option to turn off the location tracker if perceived to be too burdensome or invasive. Furthermore, 3 apps worked in conjunction with smart packaging: displaying the code for an electronic pill dispenser and a companion app for a Bluetooth-enabled pill bottle cap that tracks adherence.

Gamification

Gamification refers to the incorporation of game-like elements (eg, points, leaderboards, levels, and loot boxes) into nongame contexts [33]. A total of 7 apps (32%) had basic game-like elements, including virtual points and streak systems with bonuses and resets. Only 1 app made use of a storyline-based feature to enhance engagement and make the content more relatable to patients. This was an optional Choose-Your-Own-Adventure–style narrative that contained medication adherence tips. Of 34 participants who downloaded the app during the intervention period, 22 (62%) made use of this feature [31]. Another app that did not include gamification features received feedback from their participants wishing for more gamified elements to be included in future iterations [27].

Discussion

Principal Findings

This narrative review provides an overview of the technical features and their advantages and limitations in mobile apps targeting medication adherence for SUDs reported in the literature. A total of 22 studies and 7 feature categories were identified, with the most common features implemented being reminder messages (n=14), self-assessment and goal-tracking tools (n=14), and patient-provider communication (n=13). Where efficacy outcomes were reported, results were not statistically significant but directionally positive. In contrast, feasibility and acceptability outcomes were consistently strong across studies. The apps generally demonstrated good acceptability rates, with high usability scores and positive qualitative feedback from participants.

One prominent feature of apps was contingency management, which has been observed to be an effective intervention in OUD treatment [34]. Mobile apps carry the potential to streamline contingency management through remote behavior verification and reward delivery, which could make its implementation more affordable and scalable [35]. Among the reviewed apps, both monetary rewards and video verification were found to be acceptable to users. There was, however, substantial heterogeneity in design and reward amount, suggesting a need for further research to identify the most effective reward structures. This review also found video-based medication

XSL•FO
RenderX

treatment verification delivered via mobile apps to offer a low-barrier alternative for patients who may have difficulty reaching clinics, including rural populations [36]. On the other hand, the multistep nature of video verification can be procedurally complex, increasing patient burden through technical difficulties, such as poor video quality or trouble filming or uploading videos [37]. Another risk would involve patients forgetting to self-record as they ingest their medication, which may lead to inaccurate reports of medication adherence. Successful implementation would require streamlined verification processes and adequate user training to ensure that the varying degrees of technological restrictions per feature are minimized as much as possible. Despite such potential limitations, studies have consistently reported high feasibility and acceptability of video verifications, which may improve their future adoption and integration of medication adherence monitoring technologies in clinical settings. As apps scale, the need for strong encryption and privacy measures will also increase [38]. Similar to this are location tracking features, which provide users with the opportunity to locate and track participation at local resources, clinics, and support groups. One app also demonstrated a use case for tracking the patient's presence at self-identified risk locations and then initiating a sequence of intervention prompts, such as vibrations and a list of contacts. However, privacy remains a major factor influencing the adoption of location tracking and other remote monitoring technologies, with each patient's privacy concerns varying from person to person [39]. As with video verification, attention to security and privacy, including options for users to disable tracking features, would be important for future app development.

Other features with positive responses from patients included improved access to support resources and peer support personnel. Implementation of greater clinician or peer support personnel engagement facilitated via apps could be achieved in part through improved training but would represent a resource challenge in any scaled intervention. Self-monitoring and goal-tracking features were also commonly implemented across apps to empower and engage users in their own treatment, as well as to serve as a safety measure for situations and environments with a high risk for relapse. Several apps included the ability to send data directly to clinicians or to automated systems when concerning substance use patterns were detected, allowing for timely interventions. However, evidence of efficacy remains low, and some users suggested that tracking triggers or cravings could be discouraging or triggering in themselves. To account for the variability in user preference and personalization of their app experience, a recent systematic review has highlighted the need for balance between the use of passively collected app engagement data and user-directed triggers [40]. Studies of other apps also received suggestions on the specific phrasing of questions, which may have different connotations for different populations and would require greater tailoring.

Some less common features included information, community boards, and gamification. There was very limited reporting on the efficacy of medical adherence information, even though several teams had developed learning modules, searchable databases, informational videos, and, in 1 case, a story-based game that delivered adherence tips. More research is required to discover the most effective means of delivering information and which types of information have the greatest impact on adherence. Community boards were consistently perceived as valuable by participants, particularly for social and emotional support. However, actual use was often low due to a lack of board activity, as a result of small sample sizes. A more effective implementation of this feature would likely require a larger user base and active support from a team of moderators. Additionally, patients responded positively to gamified features that tracked their treatment streaks and records of achievement. Through the lens of behavior change theory, gamification may be seen as a potential means of allowing users to gain control and actively engage with the intervention [41]. While high levels of use and patient preferences suggest a potential role for these features in supporting user engagement, gamification was found to be underused and represents a potential area for future research in SUD interventions [38].

These findings can furthermore be interpreted through the fit between individuals, task, and technology (FITT) framework, which proposes that technology adoption depends on the fit between individual users, the technology, and the clinical tasks and processes [42]. From this perspective, the potential complexities behind verification procedures may reflect a poor task-technology fit, while privacy concerns may affect the individual-technology fit and users' willingness to engage with the intervention. Future work should analyze the facilitators and barriers of their local implementation and include features that can positively affect the alignment between the technology with the user and tasks, respectively, including but not limited to external motivation and training of users, easily accessible user support, and reduction of technical workload (eg, appropriate number of steps for verification and frequencies of reminders).

Comparison With Prior Work

One important observation of our review is that all apps differed in the number of features implemented and that the combined impact of features was often not considered in usability or acceptability studies. Studies have shown that technological features are a fundamental aspect of the success of health care apps [40]. Despite the commonness of all features described in this and other similar reviews, each app was found to include a relatively low number of features for medication adherence, with a range of 2 to 6 features per app. It may be possible that the overlap of multiple features, and thus the option for users to choose their own features to their liking and progress, may be an enabler of increased engagement with the app. A potential option could be to match the number and content of features with the level of engagement observed from users, with more intensive features, such as contingency management and location tracking, provided to patients with lower rates of use, compared to less and more passive features, such as community boards, provided to patients with demonstrated higher levels of engagement. Given the added complexity and challenge of retaining patients in SUD treatment, future studies should explore how to stratify findings of efficacy and patient satisfaction by the number and functionality of features.

XSL•FO
RenderX

Limitations

This narrative review has some limitations. Our database selection and search strategies may not have been extensive enough to capture all published literature, along with the inclusion of studies only published in English. Furthermore, we did not attempt to download any of the apps and interact with them, which means that some design elements, such as messages of congratulation upon adherence or progress visualization, may have been present but not explicitly described. Across the studies, evidence of efficacy was limited as many studies were pilots or feasibility trials with small sample sizes and demographically homogeneous user populations, restricting generalizability. Future studies will need to take into consideration the diversity of SUDs and the settings in which research is conducted. Despite these limitations, this study found an overall lack of dedicated apps for SUDs relative to need and offers insights into the optimal design and features for future apps designed for this population. As the adoption and development of mobile-based interventions continues to increase, understanding their key characteristics will be vital to creating solutions catered to user needs.

Conclusions

This narrative review provides an overview of the technical features found in apps designed to support medication adherence for SUD treatment reported in the literature, along with the advantages and limitations of the features. A total of 7 main categories of app features were identified, with the most common being reminder messages, self-assessment and goal-tracking tools, and patient-provider communication. Further research is needed to determine which specific app features and implementation strategies are most effective in supporting medication adherence. These findings may inform the design and implementation of future mobile app interventions, supporting their broader adoption and effectiveness.

Generative AI Use

The authors attest that there was no use of generative artificial intelligence (AI) technology in the generation of text, figures, or other informational content of this manuscript.

Generative AI Use

The authors attest that there was no use of generative artificial intelligence (AI) technology in the generation of text, figures, or other informational content of this manuscript.

Funding

This work was supported by the NSERC Discovery Horizons grant and the VGH & UBC Hospital Foundation.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information file.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information file.

Authors' Contributions

Conceptualization: BN (lead), PA (equal)
Data curation: JK
Formal analysis: LB (lead), JK (equal)
Funding acquisition: BN
Investigation: BN
Methodology: JK
Project administration: JK (lead), LB (supporting)
Resources: PA
Supervision: AB
Validation: AB
Visualization: EF (lead), AB (supporting)
Writing – original draft: AB (lead), CD (supporting)
Writing – review & editing: AB (lead), CD (supporting), EF (supporting)
Conceptualization: BN (lead), PA (equal)
Data curation: JK
Formal analysis: LB (lead), JK (equal)
Funding acquisition: BN
Investigation: BN
Methodology: JK
Project administration: JK (lead), LB (supporting)
Resources: PA

XSL•FO
RenderX

Supervision: AB
Validation: AB
Visualization: EF (lead), AB (supporting)
Writing – original draft: AB (lead), CD (supporting)
Writing – review & editing: AB (lead), CD (supporting), EF (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Critical appraisal of quantitative studies.
[DOCX File, 2632 KB - humanfactors_v13i1e104420_app1.docx]

References

1. Substance use trends across Canadian jurisdictions, 2008–2024: the prevalence interactive chart. Canadian Substance Use Costs and Harms. 2023. URL: https://csuch.ca/explore-the-data/data-charts/ [accessed 2026-11-07]
2. Opioid- and stimulant-related harms in Canada. : Public Health Agency of Canada; 2026 Sep URL: https://health-infobase.canada.ca/substance-related-harms/opioids-stimulants/ [accessed 2026-11-09]
3. Mattick RP, Breen C, Kimber J, Davoli M. Buprenorphine maintenance versus placebo or methadone maintenance for opioid dependence. Cochrane Database Syst Rev 2014 Feb 6;2014(2):CD002207. [doi: 10.1002/14651858.CD002207.pub4] [Medline: 24500948]
4. Mattick RP, Breen C, Kimber J, Davoli M. Methadone maintenance therapy versus no opioid replacement therapy for opioid dependence. Cochrane Database Syst Rev 2009 Jul 8;2009(3):CD002209. [doi: 10.1002/14651858.CD002209.pub2] [Medline: 19588333]
5. Larance B, Degenhardt L, Lintzeris N, Winstock A, Mattick R. Definitions related to the use of pharmaceutical opioids: extramedical use, diversion, non-adherence and aberrant medication-related behaviours. Drug Alcohol Rev 2011 May;30(3):236-245. [doi: 10.1111/j.1465-3362.2010.00283.x] [Medline: 21545553]
6. Sordo L, Barrio G, Bravo MJ, et al. Mortality risk during and after opioid substitution treatment: systematic review and meta-analysis of cohort studies. BMJ 2017 Apr 26;357:j1550. [doi: 10.1136/bmj.j1550] [Medline: 28446428]
7. Ud Din Arshad QK, Kashif AU, Quershi IM. A review on the evolution of cellular technologies. In: 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST - 2019): Institute of Electrical and Electronics Engineers; 2019:989-993. [doi: 10.1109/IBCAST.2019.8667173]
8. Marsch LA, Campbell A, Campbell C, et al. The application of digital health to the assessment and treatment of substance use disorders: the past, current, and future role of the National Drug Abuse Treatment Clinical Trials Network. J Subst Abuse Treat 2020 Mar;112S:4-11. [doi: 10.1016/j.jsat.2020.02.005] [Medline: 32220409]
9. Proctor SL. Rewarding recovery: the time is now for contingency management for opioid use disorder. Ann Med 2022 Dec;54(1):1178-1187. [doi: 10.1080/07853890.2022.2068805] [Medline: 35471190]
10. DeFulio A, Rzeszutek MJ, Furgeson J, Ryan S, Rezania S. A smartphone-smartcard platform for contingency management in an inner-city substance use disorder outpatient program. J Subst Abuse Treat 2021 Jan;120:108188. [doi: 10.1016/j.jsat.2020.108188] [Medline: 33298295]
11. Hammond AS, Sweeney MM, Chikosi TU, Stitzer ML. Digital delivery of a contingency management intervention for substance use disorder: a feasibility study with DynamiCare Health. J Subst Abuse Treat 2021 Jul;126:108425. [doi: 10.1016/j.jsat.2021.108425] [Medline: 34116816]
12. Schuman-Olivier Z, Borodovsky JT, Steinkamp J, et al. MySafeRx: a mobile technology platform integrating motivational coaching, adherence monitoring, and electronic pill dispensing for enhancing buprenorphine/naloxone adherence during opioid use disorder treatment: a pilot study. Addict Sci Clin Pract 2018 Sep 24;13(1):21. [doi: 10.1186/s13722-018-0122-4] [Medline: 30249279]
13. Steinkamp JM, Goldblatt N, Borodovsky JT, et al. Technological interventions for medication adherence in adult mental health and substance use disorders: a systematic review. JMIR Ment Health 2019 Mar 12;6(3):e12493. [doi: 10.2196/12493] [Medline: 30860493]
14. Hong QN, Fàbregues S, Bartlett G, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Educ Inf 2018 Nov 1;34(4):285-291. [doi: 10.3233/EFI-180221]
15. Holtyn AF, Toegel F, Novak MD, Leoutsakos JM, Fingerhood M, Silverman K. Remotely delivered incentives to promote buprenorphine treatment engagement in out-of-treatment adults with opioid use disorder. Drug Alcohol Depend 2021 Aug 1;225:108786. [doi: 10.1016/j.drugalcdep.2021.108786] [Medline: 34087746]
16. Marino EN, Karns-Wright T, Perez MC, Potter JS. Smartphone app–based contingency management and opioid use disorder treatment outcomes. JAMA Netw Open 2024 Dec 2;7(12):e2448405. [doi: 10.1001/jamanetworkopen.2024.48405] [Medline: 39621343]

XSL•FO
RenderX

17. Metrebian N, Carr E, Goldsmith K, et al. Mobile telephone delivered contingency management for encouraging adherence to supervised methadone consumption: feasibility study for an RCT of clinical and cost-effectiveness (TIES). Pilot Feasibility Stud 2021 Jan 7;7(1):14. [doi: 10.1186/s40814-020-00761-4] [Medline: 33407950]
18. Smith CL, Keever A, Bowden T, et al. Patient feedback on a mobile medication adherence app for buprenorphine and naloxone: closed and open-ended survey on feasibility and acceptability. JMIR Form Res 2023 Apr 19;7:e40437. [doi: 10.2196/40437] [Medline: 37074780]
19. Westergaard RP, Genz A, Panico K, et al. Acceptability of a mobile health intervention to enhance HIV care coordination for patients with substance use disorders. Addict Sci Clin Pract 2017 Apr 26;12(1):11. [doi: 10.1186/s13722-017-0076-y] [Medline: 28441962]
20. Kawasaki S, Lehman E, Kong L, Spangler A, Aydingolo N, Nunes E. A digital therapeutic mobile app, ReSET-O, as adjunct to medication treatment for opioid use disorder: a randomized clinical trial at an opioid treatment program. J Addict Med 2026 Mar 16. [doi: 10.1097/ADM.0000000000001682] [Medline: 41833956]
21. Bhatraju EP, Kennedy DN, Gojic AJ, et al. mHealth Incentivized Adherence Plus Patient Navigation (MIAPP): protocol for a pilot randomized controlled trial to improve linkage and retention on buprenorphine for hospitalized patients with methamphetamine use and opioid use disorder. Addict Sci Clin Pract 2025 Jan 29;20(1):6. [doi: 10.1186/s13722-025-00538-1] [Medline: 39881397]
22. Stitzer ML, Vandrey R. Contingency management: utility in the treatment of drug abuse disorders. Clin Pharmacol Ther 2008 Apr;83(4):644-647. [doi: 10.1038/sj.clpt.6100508] [Medline: 18305456]
23. Hodges J, Waselewski M, Harrington W, et al. Six-month outcomes of the HOPE smartphone application designed to support treatment with medications for opioid use disorder and piloted during an early statewide COVID-19 lockdown. Addict Sci Clin Pract 2022 Mar 7;17(1):16. [doi: 10.1186/s13722-022-00296-4] [Medline: 35255965]
24. Gustafson DH Sr, Landucci G, McTavish F, et al. The effect of bundling medication-assisted treatment for opioid addiction with mHealth: study protocol for a randomized clinical trial. Trials 2016 Dec 12;17(1):592. [doi: 10.1186/s13063-016-1726-1] [Medline: 27955689]
25. Mi RZ, Yang EF, Tahk A, et al. mHealth engagement for antiretroviral medication adherence among people with HIV and substance use disorders: observational study. J Med Internet Res 2024 Dec 20;26:e57774. [doi: 10.2196/57774] [Medline: 39705693]
26. Schramm ZA, Leroux BG, Radick AC, et al. Video directly observed therapy intervention using a mobile health application among opioid use disorder patients receiving office-based buprenorphine treatment: protocol for a pilot randomized controlled trial. Addict Sci Clin Pract 2020 Jul 31;15(1):30. [doi: 10.1186/s13722-020-00203-9] [Medline: 32736660]
27. Stocker A, Navarro N, Schmitt L, et al. Acceptability of “DIDE”, a mobile application designed at facilitating care adherence of patients with substance use disorder. Addict Sci Clin Pract 2024 Oct 15;19(1):72. [doi: 10.1186/s13722-024-00500-7] [Medline: 39407238]
28. Waselewski ME, Flickinger TE, Canan C, et al. A mobile health app to support patients receiving medication-assisted treatment for opioid use disorder: development and feasibility study. JMIR Form Res 2021 Feb 23;5(2):e24561. [doi: 10.2196/24561] [Medline: 33620324]
29. Palacios JE, Sherrick R, Janssen T, et al. Implementation of a mobile digital tool supporting medication for opioid use disorder treatment improves retention: stepped-wedge cluster randomized controlled trial. J Med Internet Res 2025 Dec 22;27:e83346. [doi: 10.2196/83346] [Medline: 41429011]
30. McPherson SM, Smith CL, Hall L, et al. Mobile Medication Adherence Platform for Buprenorphine (MAP4BUP): a phase I feasibility, usability and efficacy pilot randomized clinical trial. Drug Alcohol Depend 2024 Mar 1;256:111099. [doi: 10.1016/j.drugalcdep.2024.111099] [Medline: 38306822]
31. Horvath KJ, Lammert S, MacLehose RF, Danh T, Baker JV, Carrico AW. A pilot study of a mobile app to support HIV antiretroviral therapy adherence among men who have sex with men who use stimulants. AIDS Behav 2019 Nov;23(11):3184-3198. [doi: 10.1007/s10461-019-02597-3] [Medline: 31309348]
32. Godersky ME, Klein JW, Merrill JO, et al. Acceptability and feasibility of a mobile health application for video directly observed therapy of buprenorphine for opioid use disorders in an office-based setting. J Addict Med 2020;14(4):319-325. [doi: 10.1097/ADM.0000000000000608] [Medline: 31972762]
33. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013 Dec 12;1(1):e3. [doi: 10.2196/games.3139] [Medline: 25658754]
34. Bolívar HA, Klemperer EM, Coleman SR, DeSarno M, Skelly JM, Higgins ST. Contingency management for patients receiving medication for opioid use disorder: a systematic review and meta-analysis. JAMA Psychiatry 2021 Oct 1;78(10):1092-1102. [doi: 10.1001/jamapsychiatry.2021.1969] [Medline: 34347030]
35. Litvin EB, Abrantes AM, Brown RA. Computer and mobile technology-based interventions for substance use disorders: an organizing framework. Addict Behav 2013 Mar;38(3):1747-1756. [doi: 10.1016/j.addbeh.2012.09.003] [Medline: 23254225]
36. Benavides-Vaello S, Strode A, Sheeran BC. Using technology in the delivery of mental health and substance abuse treatment in rural communities: a review. J Behav Health Serv Res 2013 Jan;40(1):111-120. [doi: 10.1007/s11414-012-9299-6] [Medline: 23093443]

XSL•FO
RenderX

37. Mason M, Cho Y, Rayo J, Gong Y, Harris M, Jiang Y. Technologies for medication adherence monitoring and technology assessment criteria: narrative review. JMIR Mhealth Uhealth 2022 Mar 10;10(3):e35157. [doi: 10.2196/35157] [Medline: 35266873]
38. Martin K, Shilton K. Putting mobile application privacy in context: an empirical study of user privacy expectations for mobile devices. Inf Soc 2016 Apr 13;32(3):200-216. [doi: 10.1080/01972243.2016.1153012]
39. Sun L, Yang B, Kindt E, Chu J. Privacy barriers in health monitoring: scoping review. JMIR Nurs 2024 May 9;7(1):e53592. [doi: 10.2196/53592] [Medline: 38723253]
40. Yeung AW, Torkamani A, Butte AJ, et al. The promise of digital healthcare technologies. Front Public Health 2023 Sep 26;11:1196596. [doi: 10.3389/fpubh.2023.1196596] [Medline: 37822534]
41. Kaveladze BT, Young SD, Schueller SM. Antifragile behavior change through digital health behavior change interventions. JMIR Form Res 2022 Jun 3;6(6):e32571. [doi: 10.2196/32571] [Medline: 35657665]
42. Ammenwerth E, Iller C, Mahler C. IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study. BMC Med Inform Decis Mak 2006 Jan 9;6(1):3. [doi: 10.1186/1472-6947-6-3] [Medline: 16401336]

Edited by C Jacob; submitted 19.Jun.2026; peer-reviewed by M Litt, VP Jover; comments to author 23.Aug.2026; revised version received 05.Oct.2026; accepted 14.Oct.2026; published 19.Nov.2026.

Please cite as:
Kim JJ, Boschman L, Azar P, Noureddin B
Mobile Apps Supporting Medication Adherence in Substance Use Disorder Treatment: Narrative Review
JMIR Hum Factors 2026;13:e104420
URL: https://humanfactors.jmir.org/2026/1/e104420
doi:10.2196/104420

© Jane J Kim, Lua Boschman, Pouya Azar, Borna Noureddin. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 19.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Persuasive Systems Design Features of Smartphone Apps for Psychosis: Systematic Review

Theresa Taylor, ClinPsy; Jasmin Peat, MSc; David Kell, MSc; Shadi Daryan, MSc; Pamela Jacobsen, PhD

Department of Psychology, University of Bath, Claverton Down, Bath, United Kingdom

Corresponding Author:
Pamela Jacobsen, PhD
Department of Psychology, University of Bath, Claverton Down, Bath, United Kingdom

Abstract

Background: It is unclear why some smartphone apps designed for people with psychosis are engaging, while others are not. One possible explanation is the apps' persuasive features and the operationalization and implementation of these features.

Objective: This systematic review set out to quantify and describe the persuasive features used in smartphone apps for psychosis, investigate whether there was any association between persuasive features and attrition or adherence rates. and document the quality of the included apps.

Methods: We searched electronic databases PsycINFO, PubMed, and Google Scholar for eligible papers published between the years of 2013 and 2025. Hand searches of reference lists were completed. Apps were selected if they were designed for people with psychosis and there were published empirical studies investigating the apps. Two reviewers from the review team (TT, JP, DK, and SD) independently screened papers and extracted data on adherence and attrition, as well as coded papers for evidence of persuasive features according to the persuasive systems design (PSD) model. Available data were synthesized descriptively and narratively. We attempted to access apps via app stores or by correspondence with the research team.

Results: We found 22 apps for psychosis, with 30 associated published papers. The persuasive features were as follows: personalization (18 apps); reminders (15 apps); suggestions (11 apps); tunneling and self-monitoring (10 apps); reduction (9 apps); liking (8 apps); social role (6 apps); rehearsal, praise, and similarity (5 apps); rewards (4 apps); simulation, real-world feel, and social learning (3 apps); surface credibility and normative influence (2 apps); and trustworthiness, social comparison, and social facilitation (1 app). Expertise, authority, third-party endorsements, verifiability, cooperation, competition, and recognition were present in zero apps. Features in the categories of primary task support and dialogue support were well represented, while social support and system credibility support were underused. It was found that there was no association between the number of persuasive features and attrition; an association between persuasive features and adherence could not be assessed. The quality of the apps could not be judged due to 20 of the 22 apps being inaccessible either through the research papers' authors or through app stores.

Conclusions: Our findings indicate that in psychosis apps there is potential to include a broader range of persuasive features, which might maximize engagement. Psychosis apps may benefit from incorporating more features that leverage the persuasive impact of having users interact (social support) and incorporating features that emphasize system credibility and trustworthiness. Further studies could determine whether an increase in the number of persuasive features will impact app engagement and which features are most impactful in this context.

(JMIR Hum Factors 2026;13:e81101) doi:10.2196/81101

KEYWORDS

persuasive systems design; psychosis; schizophrenia; human computer interaction; smartphone app; mhealth; app

Introduction

Background

There is a well-recognized global gap between the need for and provision of services to prevent, identify, and treat mental health conditions [1]. People who experience psychosis are among those with the highest need for support [2]; thus, interest and research into digital treatments for this condition have grown over the last decade. A recent meta-analysis of randomized controlled trials (RCTs) assessing apps designed for symptom reduction in patients with schizophrenia found statistically significant symptom reduction across positive (standard mean difference [SMD]=−0.205, 95% CI −0.388 to −0.022) and negative symptoms (SMD=−0.406, 95% CI −0.791 to −0.020) [3]. However, adherence to smartphone apps for mental health is notoriously low in both trial and real-world conditions. For example, a systematic review of 93 popular publicly available mental health and well-being apps found that only 3.3% (IQR 6.2%) of users were still accessing the app after 30 days [4].

XSL•FO
RenderX

Specifically relating to mental health apps, a meta-analysis of 70 RCTs found that adherence consistently declined with time [5]. One potential contributor to disengagement is insufficient attention to system design and how it affects users' experience of apps [6]. Increasing researcher and app designer understanding of the role the technology is playing in digital interventions has the potential to shift patterns of user disengagement [7].

Prior Work

One area of system design that can be applied to app development is persuasive systems design (PSD). The PSD is defined as the use of computer-mediated products that aim to influence or change people's cognition and behaviors [8]. A framework has been developed that transforms persuasive principles into software requirements and system features, namely primary task support, dialogue support, system credibility support, and social support [9]. For example, an app praising a user after they complete a session is a system feature of the PSD principle "dialogue support," which pertains to the feedback an interactive system provides to its users to help them move toward their goal.

It has been hypothesized that persuasive features will increase alliance, adherence, and engagement, in turn leading to better treatment outcomes. Kelders et al [10] conducted a systematic review of web-based health interventions and found persuasive features accounted for a significant difference in adherence scores. However, more recent research on apps has produced mixed results. A review of 70 apps for health and wellness available to the public found a positive correlation between apps' ranks of popularity (based on ratings and installation) and the apps' persuasive features [11]. In contrast, Alqahtani et al [12] found no link between the number of persuasive features in mental health apps and efficacy or engagement. Interestingly, a recent meta-analysis looking at efficacy, engagement, and persuasive design in smartphone apps for depression and anxiety found persuasive features were linked to increased effect sizes; however, an increased number of persuasive features was also linked to rising attrition [13]. Thus, the relationship between persuasive design, efficacy, and engagement is multifaceted and opaque. It may be that for some client groups, persuasive design or over-application of persuasive design has the opposite of the intended impact, causing users to disengage. This requires further investigation.

There is evidence that motivation type [12] and psychological and demographic characteristics [14] impact mobile health users' proclivity to persuasive design, meaning the characteristics of the user may be relevant to understanding how they relate to an app's features. People with psychosis may have unique needs. For example, a review on the implementation of digital interventions with people with psychosis and bipolar disorder found that this client group may have cognitive impairments that make engagement with complex interventions challenging [15]. A related difficulty for people with serious mental illness is difficulties in social functioning [16,17]. Furthermore, in human-to-human psychosis interventions, the quality of the therapeutic alliance has been found to be associated with treatment outcome; thus, the quality of this client group's interaction with the therapeutic agent (app) may be paramount to the intervention's success [18]. Thus, understanding how apps are being designed for people with psychosis and the links between system design and app adherence could provide valuable information for improving these interventions going forward.

Goal of This Study

There is a gap in the literature for a review that looks in a granular manner at persuasive features within mobile apps designed for people with psychosis. This review aimed to answer the following questions: (1) what PSD features do apps aimed at people with psychosis have? (2) Which PSD features of smartphone apps aimed at people with psychosis are associated with app adherence and attrition rates? (3) What is the quality of apps aimed at people with psychosis?

Methods

Overview

This study's protocol was preregistered on the Open Science Framework [19]. The data associated with this review are also available at the same location on the Open Science Framework. The electronic databases searched were PsycINFO, PubMed, and Google Scholar. Due to the fast-evolving nature of technology, database searches were limited from the year 2013 onward and searched in November 2023 and March 2025, respectively. Query strings were (smartphone* or "mobile phone" or "cell phone") AND ("app" or "apps" or "application" or "applications") AND ("schizophrenia" or "schizo" or "psychosis" or "psychotic"). To identify articles not found in the original searches, we reviewed the reference lists of included papers for relevant apps. The authors also used prior knowledge of apps to identify additional inclusions and linked these with their associated papers. Where papers could not be accessed, study authors were contacted.

Study Eligibility and Screening

The "unit of interest" in this review was the smartphone apps themselves, as we were aware that multiple papers may be published relating to various aspects of the same app (eg, pilot studies, clinical trials, and qualitative papers). We linked together multiple papers where they related to the same app and included those most relevant for gathering the information for this paper.

Inclusion criteria for eligible apps were as follows:

1. The app has associated with it a minimum of one peer-reviewed journal article reporting empirical data.
2. The app is aimed at people with psychosis.
3. The study sample has been confirmed to have psychosis via a prior diagnosis, psychometric measures of symptoms clinically relevant to psychosis, or a clinical interview.
4. App features were sufficiently described in the associated publications, provided by the authors upon request, or the intervention was accessible to the review team.

Exclusion criteria were as follows:

XSL•FO
RenderX

1. The app targets medication adherence, another condition (eg, cannabis addiction, smoking, or diabetes), or is aimed at supplying the user information (eg, on their legal rights).
2. The app is aimed only at monitoring or accessing treatment records.
3. Other digital interventions, for example, computer-based interventions, text, email, online counseling, and videoconferencing.
4. Apps that require other digital devices (eg, virtual reality [VR] headsets).
5. Apps targeting families of people with psychosis or the care team.

Covidence was used for all stages of review, including screening and data extraction. After duplicate records were excluded, 2 reviewers (TT, DK, SD, and JP) independently screened each paper at both title and abstract and full-text review. All screeners completed training, and a consensus check on a small selection of papers was done to ensure consistency between screeners. The initial agreement between screeners was 81% at title and abstract review and 70% at full-text review. Where there were conflicting decisions between screeners, consensus was reached by discussion and with input from the senior author (PJ) when needed.

Data Extraction

Relevant data from included studies were extracted in duplicate independently by 2 reviewers (TT or JP and DK or SD) using a predetermined data collection tool. Oinas-Kukkonen and Harjumaa [9], which lays out descriptions of persuasive features and examples of how the features can be defined, and their implementation was used as a coding guide for determining the persuasive features. Prior to extraction, reviewers piloted the data extraction tool and discussed discrepancies. During extraction, regular meetings allowed discussion of the coding guide and resolution of discrepancies. The review team extracted information on the apps, including the intervention name, a brief description of the app, treatment target, whether the app was automated (used without a human intervention component alongside) or blended (used with an accompanying human therapist), intended frequency of intervention, and intended duration of intervention. Blended interventions with human support may inherently include persuasive elements outside the app interface, which was out of scope for this review.

From the most recent empirical paper on the app, information describing the study was extracted (study design, location, sample size, participant mean age and diagnoses, attrition during trial, and available adherence data). Missing data were reported in the write-up. “Interrater reliability was assessed across all 22 included apps and 28 PSD features (616 coding decisions). Cohen Kappa was κ=0.95 (98.4% observed agreement).”

Quality Assessment

The reviewers had planned to assess the quality of the apps using the Mobile Application Rating Scale [20], a quality assessment tool for health apps with published psychometrics [21]. The Apple and Android app stores were searched, and the corresponding authors were contacted to request access to the apps. However, only 2 of the 22 included apps could be accessed; thus, assessing app quality could not be done meaningfully, and this was not completed.

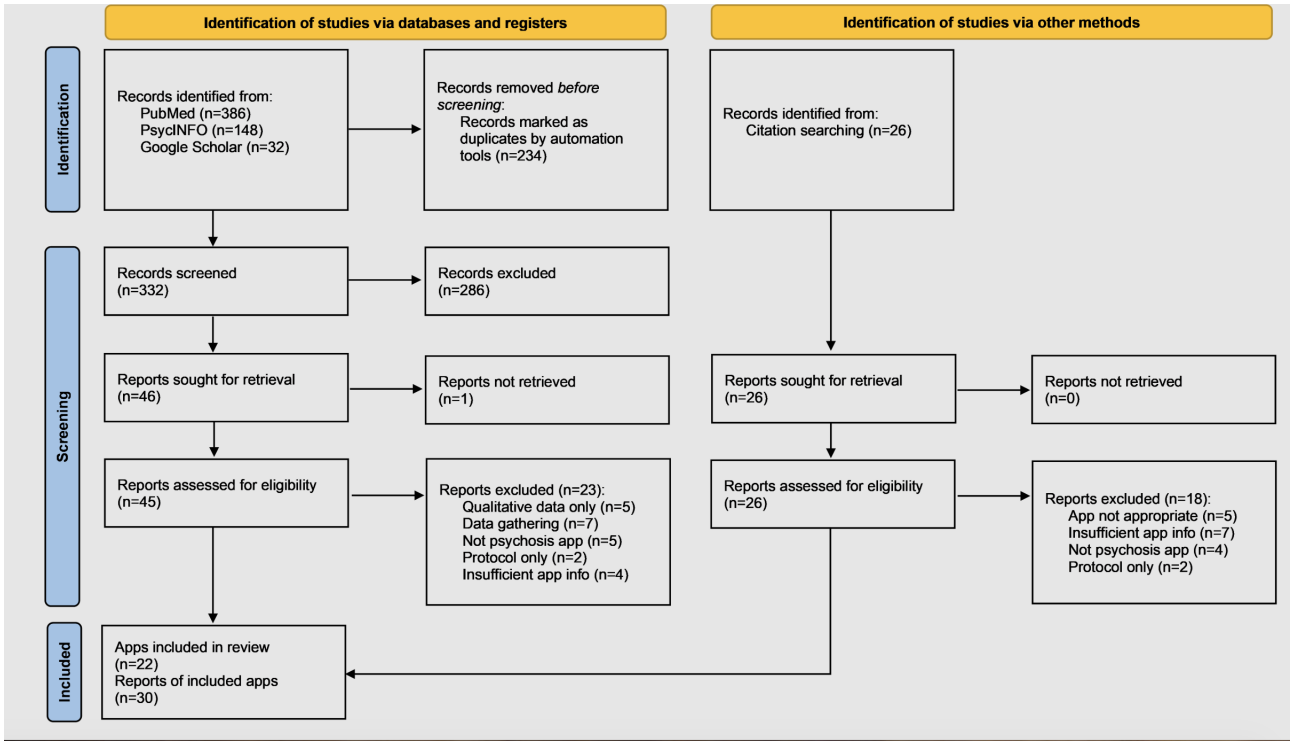
Results

Overview

The search identified 566 studies, including 234 duplicates. After title, abstract, and full-text screening, 30 [22-51] eligible papers reporting on 22 apps were included. The screening process is summarized in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Checklist 1) diagram (Figure 1). App names, treatment targets, and whether the app is automated or blended are detailed (Table 1). Further information on the apps, associated outcome study characteristics, and attrition and adherence rates is provided (Multimedia Appendix 1) and a matrix of persuasive features of each app (Multimedia Appendix 2). Across 22 apps, 21 categories of persuasive features were identified. Primary task support and dialogue support were the most common, with social support and system credibility support coded considerably less often.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram of study selection and screening process.



XSL•FO
RenderX

Table . Table depicting the included app's names, their treatment target, and whether the apps were an automated or blended intervention.

App name	Treatment target	Automated or blended
A4i (MEMOTEXT Corporation) [22,23]	Medication adherence, personal recovery, and psychiatric symptomatology	Automated
Actissist (University of Manchester) [24]	CBT[a]-informed early psychosis intervention	Automated
ChillTime (Psyxnovation Inc) [25]	Emotional regulation in dual diagnosis psychosis and substance use disorder	Automated
CBT2go (Center for Technology and Behavioral Health; Dartmouth College) [26]	CBT for negative symptoms. Aims to reduce the severity of defeatist performance attitudes	Blended (group therapy+app)
Connect+ (University of Manchester) [27]	Loneliness	Automated
Focus (Northwestern University and University of Illinois at Chicago) [28,29]	Sleep, mood, social, meds, and voices	Automated
Focus-AV (Northwestern University) [30]	Sleep, mood, social functioning, medication adherence, and voices	Automated
Grasp (New York State Psychiatric Institute and Columbia University) [31]	Social cognition and social functioning	Blended (group therapy+app)
IMPACHS (King's College London) [32,33]	Reducing psychotic and depressive symptom severity	Blended
MASS (University of California Los Angeles) [34]	Social skills training	Automated
Moneo (University of Cambridge) [35,36]	Cognitive training/cognitive rehabilitation	Automated
My Journey (Dartmouth College) [37,38]	Develop self-management skills, achieve self-determined recovery goals, and avoid future relapses	Blended
MCI-S (University Medical Center Hamburg-Eppendorf) [39]	Metacognitive beliefs, psychotic symptoms, and social functioning	Automated and blended versions
PEAR004 (Pear Therapeutics) [40]	Positive and negative symptoms of schizophrenia	Automated
PRIME (University of California San Francisco) [41]	Motivation, cognition, and negative symptoms	Blended (app+ coaches)
Savvy (University of Groningen) [42,43]	Voice hearing	Blended
Sleep app (University of Oxford) [44]	Sleep	Blended
SlowMo (King's College London) [45,46]	Paranoia	Blended
SMART app (University of Melbourne) [47]	Improve social functioning and psychiatric symptoms	Automated
TechCare (King's College London) [48,49]	Positive and negative symptoms of schizophrenia	Automated
TemStem (University Medical Center Utrecht) [50]	Voice hearing, emotionality, and vividness	Automated
WeCOPE (University of Hong Kong) [51]	Improvement in schizophrenia symptoms and personal and social functioning.	Automated

[a]CBT: cognitive behavioral therapy.

Research Question 1: PSD Elements

Twenty-one categories of persuasive features were found across the 22 apps [22-51]. The most coded areas of persuasive features were primary task support and dialogue support, with social support and system credibility support being coded significantly less frequently. The maximum number of PSD elements in an app was 11, while the least was 2 (Multimedia Appendix 2).

Primary Task Support

The primary task support category includes features that support the user to carry out their primary tasks in the app [9].

Personalization

Personalization, defined as the app noting user inputs and adjusting content accordingly, was the most common primary task support feature, present in 18 of 22 reviewed apps [22,24-26,28,30,33,34,37,40-42,44,46,47,49-51]. Personalization encompassed tailored goals [37,41,51], steps toward goals [22,26,34], encouragement [26,44,50] (eg, in the TemStem [University Medical Center Utrecht] app, if the user assessed themselves as feeling "powerless," an avatar would give them personalized support such as "You are strong!" [50]), using the user's name [46], customizable skill toolboxes [22,33,40],

XSL•FO
RenderX

ecological momentary assessment informing personalized coping strategies [24,28,42,47,49]. The ChillTime (Psyxnovation Inc) app allowed users to give feedback on the personalized coping strategies offered and used artificial intelligence to improve future suggestions [25]. Some apps provided multiple ways to interact with content, such as video or audio [25,30,33], while the WeCOPE (University of Hong Kong) app allowed the user to choose either a male or female voice in audio recordings [51]. Other examples of personalization included customizing the app's looks [24] and the app triggering a personalized crisis plan [49], and multiple apps allowed users to personalize notification frequency [26,28,37,40]. Hornstein et al [52] suggest 4 personalization levels (intervention content, content order, and guidance/communication level) and 4 underlying mechanisms (user choice, provider choice, decision rules, and machine learning). Most reviewed apps are personalized at the intervention content level, using user choice or decision rules.

Tunneling

Tunneling was found in 10 reviewed apps [22,24,26,27,31,33,34,39,41,50]. Tunneling means the app content changes progressively with use, providing various opportunities for the user to be persuaded to make behavioral changes. Observed tunneling strategies included content that changed in a stepwise fashion toward goals [22,26,34,39,41] and progressively more challenging modules [27,41,50].

Self-Monitoring

Self-monitoring, a feature enabling users to track symptoms or goal progress, was found in 10 apps [22,24,26,27,33,34,37,41,42,50]. Capturing data and feeding it back to users can encourage behavior change. Several apps provided graphs for self-monitoring [22,24,33,37,41,50], and the IMPACHS app linked self-monitoring with customizable action plans [33].

Reduction

Reduction, which simplifies complex behavior into simple tasks, was found in 9 apps [25,27,30,31,33,34,44,50,51]. This feature was implemented through simplified language [25,33], standardized content for familiarity [25], breaking tasks into components [31,33,34], and ensuring tasks were quick to complete [25,27,44].

Simulation

Simulation, a feature allowing users to observe cause-and-effect relationships (eg, increased physical activity and weight loss), was found in 3 apps. In all instances, an actor demonstrated a skill and its contribution to achieving the user's goal [27,30,34].

Rehearsal

Rehearsal provides a way for users to rehearse a behavior in the app that they then use in daily life and was present in 6 apps [26,28,30,31,42,46].

Tailoring

Tailoring means there are different user groups that are shown different versions of the app, which was featured in 2 apps [34,44].

Dialogue Support

Dialogue support incorporates the features a system uses to increase interactivity between the app and the user.

Reminders

Reminders are notifications that remind the user of their target behavior, and these reminders were the most widely used dialogue support feature (present in 15 of 22 apps) [22,24-27,31,34,36,37,40,42,44,46,47,49]. Besides encouraging app access, reminders were sent to promote medication adherence [22,36,37,47], symptom assessment [47,49], health care appointments [22], daily activities like hygiene [47], goal-focused behavior change [26,37,44], and coping strategies such as positive self-talk [42]. While the SlowMo (King's College London) app made notifications optional [46], most apps sent multiple daily notifications without specifying if users could disable them.

Suggestion

Suggestion, a feature where apps recommend skills or behaviors to users, was found in 11 apps [22,24-28,33,34,40,41,47]. Most apps stated that various cognitive and behavioral strategies were encouraged without detailing these [22,24,28,33,40,41]. However, some apps clearly defined their suggestions, such as CBT2Go's (Dartmouth College) behavioral experiments to test beliefs (eg, "Try asking someone to go for a short walk") [26] and MASS's (University of California Los Angeles) step-by-step actions toward goals like making a new friend [34]. Apps also varied in applying suggestions; for example, ChillTime suggests random strategies to encourage trying new things [25], while the SMART (University of Melbourne) app matches suggestions to the user's mood [47].

Liking

Liking, the app's visual appeal to the user, was considered and implemented in 8 reviewed apps [24,25,27,28,33,37,41,46]. Strategies for creating visually attractive apps included using multiple images [24], visual simplicity [25], color [37], user customization of visuals [24,46], and familiar visual formats, such as Connect+'s (University of Manchester) Instagram (Meta Inc)-like design [27].

Social Role

The system adopted a social role in 6 reviewed apps [22,26,33,36,37,41], primarily related to user interactions with health care professionals. This included connecting users to health care providers [22], collecting information to facilitate deeper conversations between users and health care professionals [26,33,36,37], and arranging appointments with health care providers [36]. Only the PRIME (University of California, San Francisco) app actively encouraged user-to-user interaction [41] by generating daily prompts like "Share a silly selfie!" or "Write down a list of things you are grateful for" to encourage participants to share responses.

Similarity

Similarity, in which the system meaningfully reminds users of themselves (eg, by describing relatable scenarios or using familiar language), was reported in 5 apps [24,27,34,39,44].

XSL•FO
RenderX

Some apps used videos depicting others in comparable situations progressing toward their goals [24,27,34], such as Actisist's (University of Manchester) video story "William walks away from weed" about a person with psychosis quitting cannabis [24]. Connect+ aimed to imitate youth language for increased relatability [27]. Several apps valued input from people with lived experience to create these features [24,27,34,44].

Praise

Praise, a form of positive persuasive system feedback, was found in 5 reviewed apps [26,34,41,44,50]. The MASS app used praise to increase anticipatory pleasure for future social experiences in accordance with the temporal experience of pleasure model in schizophrenia [34]. CBT2Go praised high confidence ratings for goal-related tasks while challenging low ratings [26].

Rewards

Rewards are a tactic to reinforce target behaviors and were present in 4 apps [26,27,44,50]. For most apps, rewards were given in the form of points, stickers, or stars; however, the CBT2Go app allowed a selfie to be taken after an achievement [26].

System Credibility Support

System credibility support, an umbrella term for the ways a system promotes the perception of its own credibility and therefore increases its persuasiveness, was poorly represented in the reviewed articles. Assessing apps in this area was challenging, as most publications lacked details, and only 2 apps, Actisist and ChillTime [24,25], out of 22 apps, were directly accessible. These 2 apps met the surface credibility threshold (competent look and feel). A real-world feel, where a system provides information about its content creators to build credibility, was found in 3 apps [24,25,30]. Two apps provided details of involved clinicians, such as names and work addresses [24,25], while 2 showed videos of clinicians demonstrating skills or explaining content rationale [24,30]. One app's qualitative data showed evidence of trustworthiness, with users calling it "honest" [27]. No evidence could be found of third-party endorsements (the system provides endorsements from other sources), authority (the system refers to people in the role of authority), expertise (provides information showing knowledge, experience, and competence), or verifiability (the system provides means to verify the accuracy of the program via outside sources) in the apps or their associated papers.

Social Support

Social support encompasses how a system harnesses social influence to motivate users, such as enabling users to observe others' behaviors or compare their own behaviors to others. By leveraging the persuasive power of social interaction, a system can motivate individuals to achieve their target behavior. Social support was detailed in 4 apps [27,34,39,41]. Social learning, where users observe others performing target tasks or behaviors through the system, was present in 3 apps [27,34,39]. One app used social facilitation, allowing users to discern that others are performing the behavior alongside them, while 2 engaged normative influence, using peer pressure to encourage target behaviors [22,41]. A4i (MEMOTEXT Corporation) included a moderated, anonymized platform for users to share strategies and tips [22]. PRIME used social comparison, facilitating performance comparisons between users, and automatically generated accomplishment posts for completed challenges, providing opportunities for social reinforcement [41]. No examples of competition, cooperation, or recognition were found in the examined apps.

Research Question 2: Which PSD Features of Smartphone Apps Aimed at People With Psychosis Are Associated With App Adherence and Attrition Rates?

The number of persuasive features and attrition percentage for each app showed no evidence of an association (Pearson r=0.07, 95% CI −0.4 to 0.5). However, this finding assumes equal weight for each PSD feature, which is likely unjustifiable. Certain features may be more strongly associated with attrition than others, if any relationship exists, but these data are insufficient for this level of analysis. Inconsistent adherence reporting across apps prevented comparisons regarding adherence-persuasive feature links.

Research Question 3: What Is the Quality of Apps Aimed at People With Psychosis?

App stores were searched, and corresponding authors were contacted to access all apps. However, only 2 apps were accessible, making a quality assessment impossible, as it would have required opening and using the apps to evaluate each criteria point.

Discussion

Principal Results

This review assessed PSD features in mobile apps for people with psychosis, aiming to describe the features, links to adherence and attrition, and app quality. The categories of primary task and dialogue support were well-represented, while social support and system credibility support were underused, potentially impacting app efficacy and engagement, as both social connectivity (social support) and a user having ways to judge a system's credibility (system credibility support) could impact motivation with this client group. No link was found between persuasive features and attrition; the relationship with adherence and app quality couldn't be assessed.

Comparison With Prior Work

No previous study has systematically documented the PSD features present in apps for psychosis. The finding that personalization, self-monitoring, and reminders were among the most common strategies is consistent with previous findings on physical activity, mental health, and health apps [11,12,53]. However, the existence of these features does not guarantee a positive impact on users; how these features are applied is crucial to their utility. For example, reminders were a prominent persuasive feature in the psychosis apps reviewed; however, most apps did not give users a choice in notification frequency. One of the postulates of persuasive systems is that the system should aim to be unobtrusive, and inappropriate or ill-timed reminders may have a reverse of the desired impact [54]. Persuasion depends on symbolic strategies to trigger emotions

XSL•FO
RenderX

[55], and a blunt, unskilled approach such as sending countless reminders to use the app per day can provoke a negative emotional response, which is counterproductive. Similarly, self-monitoring is considered a fundamental feature of several evidence-based psychological therapies, including cognitive behavior therapy [56], dialectical behavior therapy [57], mindfulness [58], emotion-focused therapy [59], and acceptance and commitment therapy [60]. Thus, it is logical that psychosis apps would utilize self-monitoring to help users track their feelings, thoughts, and behaviors. However, the apps reviewed in this study limited data gathering to manual input, prompted by several daily notifications. Manual entry of personal data can be time-consuming and tedious [61] and may be a barrier to people with serious mental health difficulties. Thus, in the context of psychosis, it is unclear if self-monitoring promoted engagement in the manner intended or had a reverse to desired impact due to the frequency and high cost to the user of multiple manual inputs per day.

Research indicates that psychological and demographic variables influence susceptibility to persuasion [14]; thus, in planning these interventions, these variables should be considered. For example, in this review, reduction (simplifying tasks) and suggestion (suggesting actions to users) were among the top strategies used. However, in a review of mental health apps, these 2 strategies were used in fewer than 5 of 103 apps [12]. Thus, these strategies, which facilitate ease of app use, may be more widely implemented and necessary in a psychosis sample due to cognitive impairment and disability associated with this client group [15-17].

This study also found that apps relied heavily on the primary task support and dialogue support categories of persuasive design, with social support and system credibility support underrepresented. This means that attempts to be persuasive are unbalanced; there is emphasis on helping users to complete tasks and the interactiveness of the app, while techniques to make the app more trustworthy and credible to users and methods to leverage the impact of social forces to support behavior change are ignored. Social support strategies have been found to be more widely implemented in physical than in mental health apps [62]. However, social support is an important strategy for users who experience mental health issues because it is common to feel isolated or stigmatized, and digital interventions could benefit from intervening at the level of the social rather than limiting treatment targets to internal drivers for poor mental health. Lack of system credibility support is common across multiple types of apps [12,62]. A postulate of persuasive systems is that they should always be open to the user and reveal designer bias behind the system both to protect fairness and increase persuasiveness. We argue that system credibility support is fundamental in this time, where technology is developing at a rapid rate and users cannot be expected to understand a system's functions without guidance. Users need to be assured of not only the effectiveness, reliability, and evidence base of the app's content, but also that their data will be protected (privacy). Furthermore, without this in place, it is difficult to advance other features such as passive monitoring if users are not fully assured of the app's credibility.

Our findings suggest that the number of persuasive features used in psychosis apps is not linked to attrition rates. However, this is not necessarily an indicator that the persuasive features were not assisting the user to achieve their purpose with the app. Some authors propose that disengagement is not a barrier to efficacy if a user has achieved their goals with the digital intervention at the point of disengagement [63]. For example, a recent meta-analysis looking at efficacy, engagement, and persuasive design in smartphone applications for depression and anxiety found increased persuasive features were linked to increased effect sizes and increased attrition rates [13]. However, there may be other factors explaining the disconnect between increased persuasive features and engagement in this review; for example, there is a lack of evidence on which persuasive features are the most effective at increasing adherence and reducing dropout. Therefore, it is likely that totaling persuasive features is insufficient. Using one appropriate feature is preferable to multiple inappropriate features, too many features from one persuasive category, or a combination of appropriate and inappropriate features [64]. Furthermore, although PSD is commonly used to inform app development, the model does not contain a guideline on operationalizing the features [9]. Consequently, the same persuasive feature can be shaped differently in different interventions. We found some apps with highly creative implementation of PSD, while others executed persuasive features in a minimal manner. This mirrors the results of Sporrel et al [65], who examined 29 physical activity apps and found heterogeneity in how the features were executed, which appeared to influence the effectiveness of the intervention. These diverse implementations mean that interventions with the same persuasive feature might evoke different user responses, rendering it difficult to draw a conclusion about the impact of a specific persuasive feature at the theoretical level. Furthermore, technological advancements mean that some apps used features that did not fit within the PSD framework [53]. Thus, the lack of clarity on how many persuasive features are optimal, the lack of guidance on the operationalization of features, and the lack of acknowledgment of emerging features highlight the need to extend the PSD model. Furthermore, extension of the PSD model could include standardized checklists or algorithms for the implementation of persuasive features as guidance to app developers and reviewers.

This study's findings provide possible guidance to app developers. The finding that, although reminders are a prominent feature, little choice of their frequency was offered indicates that app engagement may benefit from offering users choice when providing notifications. This review found that self-monitoring was a common feature; however, given the functional impairments this patient group frequently faces, passive monitoring may lower the burden on the user. To implement these more advanced features, user trust is crucial. Thus, given this review's finding that system credibility features were not common, developers could benefit from including trust-building features such as highlighting the team, evidence, endorsements, backend and data storage of the app system, and real-world institutions behind the app product. This client group faces social isolation, and this review found there was little focus on social support features; thus, developers can consider integrating peer-to-peer interactions or content that gives an

XSL•FO
RenderX

indirect indication of peers and others using the app (eg, leaderboards) into apps for psychosis.

Limitations

This review has several methodological strengths, including a preregistered protocol, a systematic search strategy, and 2 independent raters throughout study selection and data extraction. However, several limitations should be noted.

The most significant constraint on interpretation is that only 2 of the 22 apps could be tested directly, rather than through reading their described features in their associated research papers [24,25]. This means the review may not fully capture the persuasive features present in each app, particularly those that are less tangible or harder to articulate—such as system credibility support—which researchers may be less likely to describe explicitly. The reported prevalence of feature categories should therefore be interpreted with this in mind, as it may reflect limitations in how apps are described in the literature as much as the features themselves.

Identifying persuasive features is also an inherently interpretive process; review teams may conceptualize and operationalize features differently. Relatedly, this review's feature-counting approach assumes equivalent value across different persuasive features, which may not hold in practice; however, a more nuanced weighting of features was beyond the scope of this analysis.

The absence of a formal risk of bias assessment is a further limitation. Although a quality appraisal of the apps was planned, it was not feasible given their inaccessibility. Additionally, attrition and adherence data were either unreported or reported inconsistently across studies, preventing any meaningful analysis of the relationship between persuasive features and these outcomes. Together, these constraints limit conclusions about feature effectiveness, and further empirical work is needed.

Finally, while non-English articles were not explicitly excluded, studies would only have been identified if their abstracts were available in English, meaning relevant research published in other languages may have been missed.

Conclusions

Our findings indicate that in psychosis apps there is an uneven use of persuasive features, with most apps relying on primary task support and dialogue support features, while social support and system credibility support remain underrepresented. This may be impacting the efficacy and engagement of these interventions. Psychosis apps may benefit from incorporating more features that leverage the persuasive impact of having users interact (social support). There is a human instinct to respond to social forces, and thus incorporating these domains into psychosis apps may allow developers to engage nonengagers. Incorporating features that emphasize system credibility and trustworthiness to the user is also crucial. As technology advances, there is an increasing need for designer openness about their positioning, bias, and how data are stored and used, and further technological advances can only be widely applied if users have confidence in the systems that gather their data. Future research could explore expanding the PSD model, as this review indicates it no longer encompasses all features apps are using, and although the underlying principles may remain sound, the way the model quantifies and describes features is becoming outdated with rapid technological advancements. Furthermore, there is a lack of clarity on how many persuasive features are optimal, how to operationalize features, and how persuasive categories work together. It may be that certain features are more effective in this context than others, and emphasis should not be on increasing numbers of persuasive features but on those most likely to influence adherence, attrition, and efficacy. In future research, a positive deviance approach could be used to examine the features of the most engaging and effective apps, while correlational or experimental research designs could explore the impact of different features on adherence and attrition. Qualitative interviews with users may shed light on what features are experienced as most impactful by users and contribute to understanding their emotional valence and possible impacts on effectiveness.

Acknowledgments

A machine tool (Claude Opus 3) was used to suggest language to improve within the manuscript, for example, grammar, and spell check on content written by the research team. Responsibility for the final manuscript lies entirely with the authors. Generative AI tools are not listed as authors and do not bear responsibility for the final outcomes. There were no funders or sponsors for this research.

Funding

The authors declared no financial support was received for this work.

Data Availability

The datasets generated or analyzed during this study are available in the Open Science Framework repository [19].

Authors' Contributions

TT conceptualized this study and developed its methodology with supervision and oversight from PJ. TT conducted study searches and performed all study screening as well as data extraction. SD, DK, and JP did the independent second screening of studies

XSL•FO
RenderX

and independently extracted all data. PJ provided research supervision. TT wrote the original draft manuscript, and PJ, SD, DK, and JP reviewed and edited the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of apps included in the review.
[DOCX File, 41 KB - humanfactors_v13i1e81101_app1.docx]

Multimedia Appendix 2
Table of persuasive systems design (PSD) features present in each app.
[DOCX File, 35 KB - humanfactors_v13i1e81101_app2.docx]

Checklist 1
PRISMA checklist.
[PDF File, 71 KB - humanfactors_v13i1e81101_app3.pdf]

References

1. Keynejad R, Spagnolo J, Thornicroft G. WHO mental health gap action programme (mhGAP) intervention guide: updated systematic review on evidence and impact. Evid Based Ment Health 2021 Apr 26;24(3):124-130. [doi: 10.1136/ebmental-2021-300254] [Medline: 33903119]
2. Ben-Zeev D, Buck B, Kopelovich S, Meller S. A technology-assisted life of recovery from psychosis. NPJ Schizophr 2019 Sep 18;5(1):15. [doi: 10.1038/s41537-019-0083-y] [Medline: 31534139]
3. Kim SK, Lee M, Jeong H, Jang YM. Effectiveness of mobile applications for patients with severe mental illness: a meta-analysis of randomized controlled trials. Jpn J Nurs Sci 2022 Jul;19(3):e12476. [doi: 10.1111/jjns.12476] [Medline: 35174976]
4. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567. [doi: 10.2196/14567] [Medline: 31573916]
5. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
6. Balcombe L, De Leo D. Human-computer interaction in digital mental health. Informatics (MDPI) 2022 Mar;9(1):14. [doi: 10.3390/informatics9010014]
7. Gega L, Jankovic D, Saramago P, et al. Digital interventions in mental health: evidence syntheses and economic modelling. Health Technol Assess 2022 Jan;26(1):1-182. [doi: 10.3310/RCTI6942] [Medline: 35048909]
8. Fogg BJ. Persuasive technology: using computers to change what we think and do. Ubiquity 2002 Dec 1;2002:2. [doi: 10.1145/764008.763957]
9. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009 Mar 1;24(1). [doi: 10.17705/1CAIS.02428]
10. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen J. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152. [doi: 10.2196/jmir.2104] [Medline: 23151820]
11. AlSlaity A, Suruliraj B, Oyebode O, Fowles J, Steeves D, Orji R. Mobile applications for health and wellness: a systematic review. Proc ACM Hum-Comput Interact 2022 Jun 14;6(EICS):1-29. [doi: 10.1145/3534525]
12. Alqahtani F, Al Khalifah G, Oyebode O, Orji R. Apps for mental health: an evaluation of behavior change strategies and recommendations for future development. Front Artif Intell 2019;2:30. [doi: 10.3389/frai.2019.00030] [Medline: 33733119]
13. Wu A, Scult MA, Barnes ED, Betancourt JA, Falk A, Gunning FM. Smartphone apps for depression and anxiety: a systematic review and meta-analysis of techniques to increase engagement. NPJ Digit Med 2021 Feb 11;4(1):20. [doi: 10.1038/s41746-021-00386-8] [Medline: 33574573]
14. McGowan A, Sittig S, Bourrie D, Benton R, Iyengar S. The intersection of persuasive system design and personalization in mobile health: statistical evaluation. JMIR Mhealth Uhealth 2022 Sep 14;10(9):e40576. [doi: 10.2196/40576] [Medline: 36103226]
15. Aref-Adib G, McCloud T, Ross J, et al. Factors affecting implementation of digital health interventions for people with psychosis or bipolar disorder, and their family and friends: a systematic review. Lancet Psychiatry 2019 Mar;6(3):257-266. [doi: 10.1016/S2215-0366(18)30302-X] [Medline: 30522979]
16. Couture SM, Penn DL, Roberts DL. The functional significance of social cognition in schizophrenia: a review. Schizophr Bull 2006 Oct;32 Suppl 1(Suppl 1):S44-S63. [doi: 10.1093/schbul/sbl029] [Medline: 16916889]

XSL•FO
RenderX

17. Fett AKJ, Viechtbauer W, Dominguez MDG, Penn DL, van Os J, Krabbendam L. The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: a meta-analysis. Neurosci Biobehav Rev 2011 Jan;35(3):573-588. [doi: 10.1016/j.neubiorev.2010.07.001] [Medline: 20620163]
18. Goldsmith LP, Lewis SW, Dunn G, Bentall RP. Psychological treatments for early psychosis can be beneficial or harmful, depending on the therapeutic alliance: an instrumental variable analysis. Psychol Med 2015 Aug;45(11):2365-2373. [doi: 10.1017/S003329171500032X] [Medline: 25805118]
19. Taylor T. Persuasive features of smartphone applications for psychosis: a systematic review. OSF. 2024. URL: https://osf.io/f6mz4 [accessed 2025-04-21]
20. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
21. Terhorst Y, Philippi P, Sander LB, et al. Validation of the Mobile Application Rating Scale (MARS). PLOS ONE 2020;15(11):e0241480. [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
22. Kidd SA, Feldcamp L, Adler A, et al. Feasibility and outcomes of a multi-function mobile health approach for the schizophrenia spectrum: App4Independence (A4i). PLOS ONE 2019;14(7):e0219491. [doi: 10.1371/journal.pone.0219491] [Medline: 31306439]
23. Kidd SA, D'Arcey J, Tackaberry-Giddens L, et al. App for independence: a feasibility randomized controlled trial of a digital health tool for schizophrenia spectrum disorders. Schizophr Res 2025 Jan;275:52-61. [doi: 10.1016/j.schres.2024.11.011] [Medline: 39657429]
24. Bucci S, Barrowclough C, Ainsworth J, et al. Actissist: proof-of-concept trial of a theory-driven digital intervention for psychosis. Schizophr Bull 2018 Aug 20;44(5):1070-1080. [doi: 10.1093/schbul/sby032] [Medline: 29566206]
25. Pennou A, Lecomte T, Potvin S, et al. A mobile health app (ChillTime) promoting emotion regulation in dual disorders: acceptability and feasibility pilot study. JMIR Form Res 2023 Jan 27;7(1):e37293. [doi: 10.2196/37293] [Medline: 36705963]
26. Granholm E, Holden J, Dwyer K, Mikhael T, Link P, Depp C. Mobile-assisted cognitive behavioral therapy for negative symptoms: open single-arm trial with schizophrenia patients. JMIR Ment Health 2020 Dec 1;7(12):e24406. [doi: 10.2196/24406] [Medline: 33258792]
27. Lim MH, Gleeson JFM, Rodebaugh TL, et al. A pilot digital intervention targeting loneliness in young people with psychosis. Soc Psychiatry Psychiatr Epidemiol 2020 Jul;55(7):877-889. [doi: 10.1007/s00127-019-01681-2] [Medline: 30874828]
28. Achtyes ED, Ben-Zeev D, Luo Z, et al. Off-hours use of a smartphone intervention to extend support for individuals with schizophrenia spectrum disorders recently discharged from a psychiatric hospital. Schizophr Res 2019 Apr;206:200-208. [doi: 10.1016/j.schres.2018.11.026] [Medline: 30551981]
29. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull 2014 Nov;40(6):1244-1253. [doi: 10.1093/schbul/sbu033] [Medline: 24609454]
30. Ben-Zeev D, Brian RM, Aschbrenner KA, Jonathan G, Steingard S. Video-based mobile health interventions for people with schizophrenia: Bringing the "pocket therapist" to life. Psychiatr Rehabil J 2018 Mar;41(1):39-45. [doi: 10.1037/prj0000197] [Medline: 27295133]
31. Sedgwick O, Hardy A, Greer B, Newbery K, Cella M. "I wanted to do more of the homework!"-feasibility and acceptability of blending app-based homework with group therapy for social cognition in psychosis. J Clin Psychol 2021 Dec;77(12):2701-2724. [doi: 10.1002/jclp.23193] [Medline: 34101177]
32. Austin SF, Jansen JE, Petersen CJ, Jensen R, Simonsen E. Mobile app integration into dialectical behavior therapy for persons with borderline personality disorder: qualitative and quantitative study. JMIR Ment Health 2020 Jun 11;7(6):e14913. [doi: 10.2196/14913] [Medline: 32525488]
33. von Malachowski A, Schlier B, Austin SF, et al. IMPACHS: feasibility and acceptability of an m-health solution integrated into routine clinical treatment for psychosis. Schizophr Res 2022 Feb;240:150-152. [doi: 10.1016/j.schres.2021.12.011] [Medline: 35026600]
34. Fulford D, Gard DE, Mueser KT, et al. Preliminary outcomes of an ecological momentary intervention for social functioning in schizophrenia: pre-post study of the motivation and skills support app. JMIR Ment Health 2021 Jun 15;8(6):e27475. [doi: 10.2196/27475] [Medline: 34128812]
35. Krzystanek M, Borkowski M, Skałacka K, Krysta K. A telemedicine platform to improve clinical parameters in paranoid schizophrenia patients: results of a one-year randomized study. Schizophr Res 2019 Feb;204:389-396. [doi: 10.1016/j.schres.2018.08.016] [Medline: 30154027]
36. Krzystanek M, Krysta K, Borkowski M, et al. The effect of smartphone-based cognitive training on the functional/cognitive markers of schizophrenia: a one-year randomized study. J Clin Med 2020 Nov 16;9(11):3681. [doi: 10.3390/jcm9113681] [Medline: 33207811]
37. Steare T, O'Hanlon P, Eskinazi M, et al. Smartphone-delivered self-management for first-episode psychosis: the ARIES feasibility randomised controlled trial. BMJ Open 2020 Aug 26;10(8):e034927. [doi: 10.1136/bmjopen-2019-034927] [Medline: 32847902]

XSL•FO
RenderX

38. Steare T, Giorgalli M, Free K, et al. A qualitative study of stakeholder views on the use of a digital app for supported self-management in early intervention services for psychosis. BMC Psychiatry 2021 Jun 19;21(1):311. [doi: 10.1186/s12888-021-03317-9] [Medline: 34147075]
39. Han M, Lee K, Kim M, Heo Y, Choi H. Effects of a metacognitive smartphone intervention with weekly mentoring sessions for individuals with schizophrenia: a quasi-experimental study. J Psychosoc Nurs Ment Health Serv 2023 Feb;61(2):27-37. [doi: 10.3928/02793695-20220706-01] [Medline: 35858205]
40. Ghaemi SN, Sverdlov O, van Dam J, Campellone T, Gerwien R. A smartphone-based intervention as an adjunct to standard-of-care treatment for schizophrenia: randomized controlled trial. JMIR Form Res 2022 Mar 28;6(3):e29154. [doi: 10.2196/29154] [Medline: 35343910]
41. Fisher M, Etter K, Murray A, et al. The effects of remote cognitive training combined with a mobile app intervention on psychosis: double-blind randomized controlled trial. J Med Internet Res 2023 Nov 13;25(1):e48634. [doi: 10.2196/48634] [Medline: 37955951]
42. Bell IH, Rossell SL, Farhall J, et al. Pilot randomised controlled trial of a brief coping-focused intervention for hearing voices blended with smartphone-based ecological momentary assessment and intervention (SAVVy): feasibility, acceptability and preliminary clinical outcomes. Schizophr Res 2020 Feb;216:479-487. [doi: 10.1016/j.schres.2019.10.026] [Medline: 31812327]
43. Bell IH, Fielding-Smith SF, Hayward M, et al. Smartphone-based ecological momentary assessment and intervention in a blended coping-focused therapy for distressing voices: development and case illustration. Internet Interv 2018 Dec;14:18-25. [doi: 10.1016/j.invent.2018.11.001] [Medline: 30510910]
44. Taylor KM, Bradley J, Cella M. A novel smartphone-based intervention targeting sleep difficulties in individuals experiencing psychosis: a feasibility and acceptability evaluation. Psychol Psychother 2022 Sep;95(3):717-737. [doi: 10.1111/papt.12395] [Medline: 35481697]
45. Garety PA, Ward T, Freeman D, et al. SlowMo, a digital therapy targeting reasoning in paranoia, versus treatment as usual in the treatment of people who fear harm from others: study protocol for a randomised controlled trial. Trials 2017 Nov 2;18(1):510. [doi: 10.1186/s13063-017-2242-7] [Medline: 29096681]
46. Garety P, Ward T, Emsley R, et al. Digitally supported CBT to reduce paranoia and improve reasoning for people with schizophrenia-spectrum psychosis: the SlowMo RCT. Efficacy Mech Eval 2021;8(11):1-90. [doi: 10.3310/eme08110] [Medline: 34398537]
47. Hanssen E, Balvert S, Oorschot M, et al. An ecological momentary intervention incorporating personalised feedback to improve symptoms and social functioning in schizophrenia spectrum disorders. Psychiatry Res 2020 Feb;284:112695. [doi: 10.1016/j.psychres.2019.112695] [Medline: 31831201]
48. Husain N, Gire N, Kelly J, et al. TechCare: mobile assessment and therapy for psychosis - an intervention for clients in the early intervention service: a feasibility study protocol. SAGE Open Med 2016;4:2050312116669613. [doi: 10.1177/2050312116669613] [Medline: 27790373]
49. Gire N, Caton N, McKeown M, et al. 'Care co-ordinator in my pocket': a feasibility study of mobile assessment and therapy for psychosis (TechCare). BMJ Open 2021 Nov 16;11(11):e046755. [doi: 10.1136/bmjopen-2020-046755] [Medline: 34785541]
50. Jongeneel A, Libedinsky I, Reinbergen A, et al. Momentary effects of Temstem, an app for voice-hearing individuals: Results from naturalistic data from 1048 users. Internet Interv 2022 Dec;30:100580. [doi: 10.1016/j.invent.2022.100580] [Medline: 36277314]
51. de Almeida RS, Couto A, Marques A, Queirós C, Martins C. Mobile application for self-management in schizophrenia: a pilot study. J Technol Hum Serv 2018 Oct 2;36(4):179-190. [doi: 10.1080/15228835.2018.1483859]
52. Hornstein S, Zantvoort K, Lueken U, Funk B, Hilbert K. Personalization strategies in digital mental health interventions: a systematic review and conceptual framework for depressive symptoms. Front Digit Health 2023;5:1170002. [doi: 10.3389/fdgth.2023.1170002] [Medline: 37283721]
53. Aldenaini N, Alqahtani F, Orji R, Sampalli S. Trends in persuasive technologies for physical activity and sedentary behavior: a systematic review. Front Artif Intell 2020;3(7):7. [doi: 10.3389/frai.2020.00007] [Medline: 33733127]
54. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 1;3(1):e7. [doi: 10.2196/mental.4984] [Medline: 26932350]
55. Miller W, Rollnick S. Motivational interviewing: preparing people for change, 2nd ed. J Healthc Qual 2003 May;25(3):46. [doi: 10.1097/01445442-200305000-00013]
56. Cohen JS, Edmunds JM, Brodman DM, Benjamin CL, Kendall PC. Using self-monitoring: implementation of collaborative empiricism in cognitive-behavioral therapy. Cogn Behav Pract 2013 Nov;20(4):419-428. [doi: 10.1016/j.cbpra.2012.06.002]
57. Feldman G, Harley R, Kerrigan M, Jacobo M, Fava M. Change in emotional processing during a dialectical behavior therapy-based skills group for major depressive disorder. Behav Res Ther 2009 Apr;47(4):316-321. [doi: 10.1016/j.brat.2009.01.005] [Medline: 19232571]
58. Erisman SM, Roemer L. A preliminary investigation of the effects of experimentally induced mindfulness on emotional responding to film clips. Emotion 2010 Feb;10(1):72-82. [doi: 10.1037/a0017162] [Medline: 20141304]

XSL•FO
RenderX

59. Paivio SC. Essential processes in emotion-focused therapy. Psychotherapy (Chic) 2013 Sep;50(3):341-345. [doi: 10.1037/a0032810] [Medline: 24000848]
60. Hayes SC, Strosahl KD, Wilson KG. Acceptance and Commitment Therapy: The Process and Practice of Mindful Change: The Guilford Press; 2012.
61. Orji R, Lomotey R, Oyibo K, Orji F, Blustein J, Shahid S. Tracking feels oppressive and "punishy": exploring the costs and benefits of self-monitoring for health and wellness. Digit Health 2018;4:2055207618797554. [doi: 10.1177/2055207618797554] [Medline: 30202544]
62. Matthews J, Win KT, Oinas-Kukkonen H, Freeman M. Persuasive technology in mobile applications promoting physical activity: a systematic review. J Med Syst 2016 Mar;40(3):72. [doi: 10.1007/s10916-015-0425-x] [Medline: 26748792]
63. O'Brien HL, Roll I, Kampen A, Davoudi N. Rethinking (Dis)engagement in human-computer interaction. Comput Human Behav 2022 Mar;128:107109. [doi: 10.1016/j.chb.2021.107109]
64. Orji R, Mandryk RL, Vassileva J. Improving the efficacy of games for change using personalization models. ACM Trans Comput-Hum Interact 2017 Oct 31;24(5):1-22. [doi: 10.1145/3119929]
65. Sporrel K, Nibbeling N, Wang S, Ettema D, Simons M. Unraveling mobile health exercise interventions for adults: scoping review on the implementations and designs of persuasive strategies. JMIR Mhealth Uhealth 2021 Jan 18;9(1):e16282. [doi: 10.2196/16282] [Medline: 33459598]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSD: persuasive systems design
RCT: randomized controlled trial
SMD: standard mean difference

Edited by J Sarvestan; submitted 22.Jul.2025; peer-reviewed by J Wong, S Narayan, X Liang; revised version received 30.Mar.2026; accepted 31.Mar.2026; published 07.May.2026.

Please cite as:
Taylor T, Peat J, Kell D, Daryan S, Jacobsen P
Persuasive Systems Design Features of Smartphone Apps for Psychosis: Systematic Review
JMIR Hum Factors 2026;13:e81101
URL: https://humanfactors.jmir.org/2026/1/e81101
doi:10.2196/81101

© Theresa Taylor, Jasmin Peat, David Kell, Shadi Daryan, Pamela Jacobsen. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 7.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Soma Design for Digital Mental Health and Well-Being Interventions: Scoping Review

Sachiyo Ito-Jaeger[1,2], PhD; Aysegul Humeyra Kafadar[3], PhD; Steve Benford[2,4], PhD; Elvira Perez Vallejos[1,2,4], PhD

[1]Mental Health & Clinical Neuroscience, School of Medicine, University of Nottingham, Innovation Park, Triumph Road, Nottingham, United Kingdom

[2]National Institute for Health and Care Research (NIHR) Nottingham Biomedical Research Centre, Nottingham, United Kingdom

[3]The Institute for Lifecourse Development, University of Greenwich, London, United Kingdom

[4]School of Computer Science, University of Nottingham, Nottingham, United Kingdom

Corresponding Author:
Sachiyo Ito-Jaeger, PhD
Mental Health & Clinical Neuroscience, School of Medicine, University of Nottingham, Innovation Park, Triumph Road, Nottingham, United Kingdom

Abstract

Background: Digital interventions for mental health and well-being are increasingly moving beyond screen-based applications toward more embodied approaches, necessitating design methodologies that emphasize bodily experiences. Soma design offers a distinctive interaction design approach that integrates bodily awareness with aesthetic appreciation, viewing the mind and body as an inseparable whole.

Objective: This scoping review aims to map and analyze the emerging applications of soma design within digital mental health and well-being (DMHW) interventions, offering a comprehensive overview of this holistic design methodology for researchers and practitioners.

Methods: This review was conducted in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines. Studies were included if they used soma design to develop DMHW intervention.

Results: Nine papers were eligible for inclusion. The interventions varied in their stage of development: 5 were conceptual design concepts created by adolescents, while 6 were prototypes or experiential artifacts developed by researchers and/or participants. All interventions incorporated soma awareness exercises, with Feldenkrais lessons being the most commonly used. Toolkits, such as Soma Bits and the Menarche toolkit, supported the design of 2 interventions. Soma design methods benefited both designers and users: designers used embodied practices to inform interaction design, while users reported increased bodily awareness, full-body engagement, emotional comfort, and relaxation.

Conclusions: Soma design represents a valuable approach for developing embodied, user-sensitive DMHW interventions. It offers a participatory, holistic co-design methodology that can meaningfully engage end users. However, many interventions identified in this review remain in early stages of development and lack systematic evaluation. Advancing the field will require interdisciplinary collaboration among mental health professionals, human-computer interaction researchers, clinicians, industry partners, and individuals with lived experience. These partnerships are essential for co-designing, testing, and implementing interventions that are both effective and scalable, ultimately extending the reach and impact of soma design in digital mental health contexts.

(JMIR Hum Factors 2026;13:e79400) doi:10.2196/79400

KEYWORDS

scoping review; soma design; digital mental health; well-being; human-computer interaction; embodiment

Introduction

Background

Overview

The use of digital interventions to support mental health and well-being (MHW) is rapidly expanding [1]. These interventions harness technology to deliver accessible, scalable, and often cost-effective solutions for a wide range of psychological conditions. To date, the most studied digital mental health tools are screen-based, typically delivered through websites and mobile apps [2-4].

However, there is growing interest in more embodied digital interventions, those that engage the body through movement, sensation, and physical interaction. Technologies, such as virtual reality (VR) therapy, socially assistive robots, and wearable

XSL•FO
RenderX

devices, are increasingly being explored for their potential to create more immersive and physically engaging therapeutic experiences [5-7]. As the field moves toward more embodied approaches, design methodologies that center bodily experience become increasingly relevant. One such methodology that has gained traction in human-computer interaction (HCI) is soma design, which offers a holistic design stance on embodied interaction that could inform the development of future digital mental health and well-being (DMHW) interventions.

Embodiment in MHW

Embodiment refers to the concept that our mental and emotional experiences are deeply intertwined with our bodily sensations and movements [8]. Embodiment interventions are based on the idea that changes in bodily movements or postures can generate or enhance emotional experiences and activate related cognitions [9]. Neuroscientific evidence supports this bidirectional relationship, showing that bodily states, such as facial expressions, hand contractions, and posture, can influence both emotional experience and cognitive functioning [10].

A recent literature review highlights the therapeutic benefits of various embodied approaches in treating mental health symptoms [8]. For example, Dance Movement Therapy reduced depression while it improved well-being and overall health [11-13]. Similarly, relaxation training, such as mindful deep breathing, has demonstrated efficacy in alleviating symptoms of anxiety disorders [14].

Digital mental health interventions are also increasingly incorporating embodied technologies. One prominent example is VR exposure–based treatment, which immerses patients in controlled, interactive environments that facilitate gradual exposure to feared stimuli. This method has demonstrated efficacy in treating conditions such as post-traumatic stress disorder and eating disorders, offering patients the ability to tailor exposure scenarios and modulate their intensity [15-18].

In addition, socially assistive robots have shown potential in supporting MHW. A recent systematic review highlights their promise, though it also underscores limitations in the current body of research, including a narrow scope, limited generalizability, and methodological weaknesses, pointing to the need for more rigorous studies [6].

Wearable technologies are also emerging as effective tools in clinical contexts. For instance, a wrist-worn device that delivers rhythmic electrical stimulation to the median nerve has been shown to significantly reduce the frequency and intensity of tics in individuals with Tourette syndrome, illustrating the therapeutic potential of neuromodulation through wearable interfaces [7].

Overall, the evidence suggests that embodiment practices, both analog and digital, offer valuable contributions to MHW interventions, warranting further exploration and integration into DMHW frameworks.

Soma Design

Soma design and embodiment are closely intertwined concepts. Soma design is a distinctive approach within HCI that foregrounds the role of the body in interaction design, integrating bodily experience with aesthetic and sensory engagement. Rooted in the philosophical framework of *somaesthetics*, developed by Richard Shusterman, soma design draws on the concept of the *soma*, the living, sentient body as experienced from within, and *aesthetics*, the cultivation of sensory awareness and appreciation [19]. This framework emphasizes the integration of body, movement, thought, and emotion, placing embodied subjectivity at the center of design [20,21].

Building upon Dewey's notion that aesthetic experience arises through active engagement rather than being an inherent property of objects [22], Shusterman's *somaesthetics* encourages a deepened awareness of sensory experience. Soma design translates these philosophical insights into a practical design methodology that challenges traditional mind-body dualisms. It embraces a holistic view of human experience, recognizing that cognition, emotion, and perception are inseparably linked to bodily states and movements [23,24].

This approach invites designers to create technologies that enrich users' lived, embodied experiences. It does so by incorporating the designer's own bodily awareness and the physical-digital materials they engage with as integral components of the design process [24-26]. This involves paying attention to how users feel, move, and perceive their interactions with technology and creating designs that enhance these experiences in a harmonious and integrated manner. This integration ensures that the resulting interactive systems/artifacts can effectively influence and enhance the end users' and the designers' sensual appreciation and movement experiences [26].

Central to this methodology is the cultivation of somatic awareness through practices, such as body scans [27], Feldenkrais exercises [28], Alexander technique [29], yoga, and other movement-based activities [30-32] (Table 1). These practices help designers attune to their own bodily sensations and develop a first-person understanding of embodied experience, which then informs the creation of interactive systems for others [23,25]. This shift from verbal or cognitive ideation to movement-based exploration enables designers to "do" rather than "talk," anchoring their process in the subtleties of sensation and presence [21,33].

XSL•FO
RenderX

Table . Definitions of soma awareness exercises.

Soma awareness exercises	Definitions
Alexander technique	A mindfulness practice and somatic modality that helps people become consciously aware of their habitual movement patterns and posture, with the goal of releasing inefficient or harmful physical habits that create tension and pain
Body scans	A type of mindfulness-based stress reduction exercise in which practitioners systematically direct focused attention to bodily sensations throughout different parts of the body in sequence
Feldenkrais exercises	A somatic movement practice developed by Moshe Feldenkrais that focuses on increasing awareness of habitual movement patterns, through slow, mindful movements guided by a practitioner, aiming to help people discover more efficient ways of moving by reducing unnecessary effort while achieving optimal results. The method helps individuals rediscover alternative movement patterns they may have forgotten, particularly beneficial for those experiencing pain or movement difficulties
Other bodily practices	Other bodily practices include various introspective physical activities and experiences, such as movement exploration, touch awareness, and horseback riding that designers engage with to inform and ground their design work

Philosopher and dancer Maxine Sheets-Johnstone further enriches this perspective by emphasizing the moving, living body as central to how we inhabit and make sense of social, physical, and temporal spaces [34]. Soma design aligns with this view, encouraging designers to engage with the body not as an object, but as a dynamic, expressive medium of experience.

To deepen bodily awareness and disrupt habitual perception, soma design often employs Soma Bits, technologies embedded in soft, body-worn materials that use heat, vibration, or inflation to evoke unfamiliar sensations and prompt reflection (see Table 2) [35,36]. These tools help both designers and users become attuned to the nuances of the embodied interactions.

Table . Definitions of toolkits.

Toolkits	Definitions
Soma Bits	Soma Bits is a design toolkit consisting of foam shapes covered in elastic cotton textile that can be placed on different body parts, combined with various actuators (heat pads, vibration motors, and pneumatic shape-changing materials) that can be inserted into shapes pockets. The toolkit serves as a living, growing library intended to help designers explore and ideate with smart materials by providing the immediate, direct experience of potential soma designers through different combinations of shapes and actuators on the body.
Menarche Bits	Menarche Bits is a customizable prototyping kit of soft, silicone, body-worn interfaces that respond to pressure through air-actuated chambers. Designed for young adolescents experiencing menarche, it includes both preprogrammed components with haptic feedback and modular pieces that can be interconnected, allowing users to create personalized technologies that help them better understand and trust their menstruating bodies.

Soma design has gained traction in HCI over the past decade. However, its application in DMHW remains underexplored. It is also unclear whether existing Soma-based interventions primarily support general well-being or target specific mental health conditions. Digital mental health is defined as “the application of digital technologies in mental health care which can be used for many purposes, including MHW promotion and prevention, well-being maintenance/self-care, early intervention, or for treating specific mental illness…” [37] While well-being can encompass physical, mental, social, and environmental dimensions [38], in this review, we focus specifically on mental well-being.

Aims and Research Questions

This scoping review aimed to explore soma design applications in DMHW interventions.

To guide this review, we formulated the following research questions:

- Main question: What DMHW interventions have been designed using soma design methods?
- Subquestions:

1. What soma awareness exercises and other embodied activities have been employed in these interventions?
2. Who participated in the design process?
3. Do these interventions primarily engage general well-being or target specific mental health conditions?

XSL•FO
RenderX

4. What evaluation approaches were used to assess these interventions?
5. What outcomes were reported?

Methods

Study Design

The scoping review was conducted to map the existing literature in this research area [39,40], following the 5-stage framework: (1) identifying the research question, (2) identifying relevant studies, (3) selecting studies, (4) charting the data, and (5) collating, summarizing, and reporting the results [39] and the principles of the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines (Checklist 1) [41].

Search Strategy and Data Sources

Two of the authors (SI-J and AHK) conducted initial literature searches relevant to the research question using databases, including Google Scholar. These initial searches were exploratory and aimed to identify relevant literature and refine the search strategy. Based on the relevant literature found during the initial searches, the search terms for the scoping review were discussed and selected during meetings among the authors. The search terms were combined using Boolean operators as follows:

((“somaesthetic*” OR “soma design”) AND (“health” OR “wellbeing” OR “wellness” OR “illness” OR “care” OR “therap*” OR “affect*” OR “emotion*” OR “feeling*”) AND (“digital mental health” OR “DMH” OR “digital technolog*” OR “technolog*” OR “digital health” OR “digital” OR “mHealth” OR “telehealth” OR “telemedicine” OR “electronic health” OR “mobile apps” OR “mobile health” OR “online based” OR “mobile*” OR “eHealth*” OR “electronic mental health” OR “e-mental health” OR “online-based*” OR “internet-based*” OR “web-based*” OR “computer-based”))

Three databases, PubMed, IEEE Explore, and ACM Digital Library, were searched on March 21, 2024, and again on January 18, 2026, and all results were imported into Rayyan [42] for screening. These databases were selected to capture the most relevant literature across health and medical research (PubMed), engineering and technology (IEEE Xplore), and HCI and design (ACM Digital Library). To enhance the comprehensiveness of the review, we additionally searched Google Scholar (first 50 records) and conducted backwards citation tracking.

Study Selection and Data Extraction

Studies were eligible for inclusion in this review if they met the following inclusion criteria:

1. Concept: digital mental health or well-being interventions designed using soma design methods and explicitly described as targeting or supporting mental health or well-being
2. Context: studies published in the English language, in any country

Studies were excluded from this review if they met any of the following exclusion criteria: review papers, an invitation to a workshop, papers that did not specifically state that the interventions targeted or supported mental health or well-being, interventions that did not involve digital technology, or papers that did not include sufficient details about the intervention (eg, soma design methods used, soma awareness exercises employed, or the purpose of the intervention).

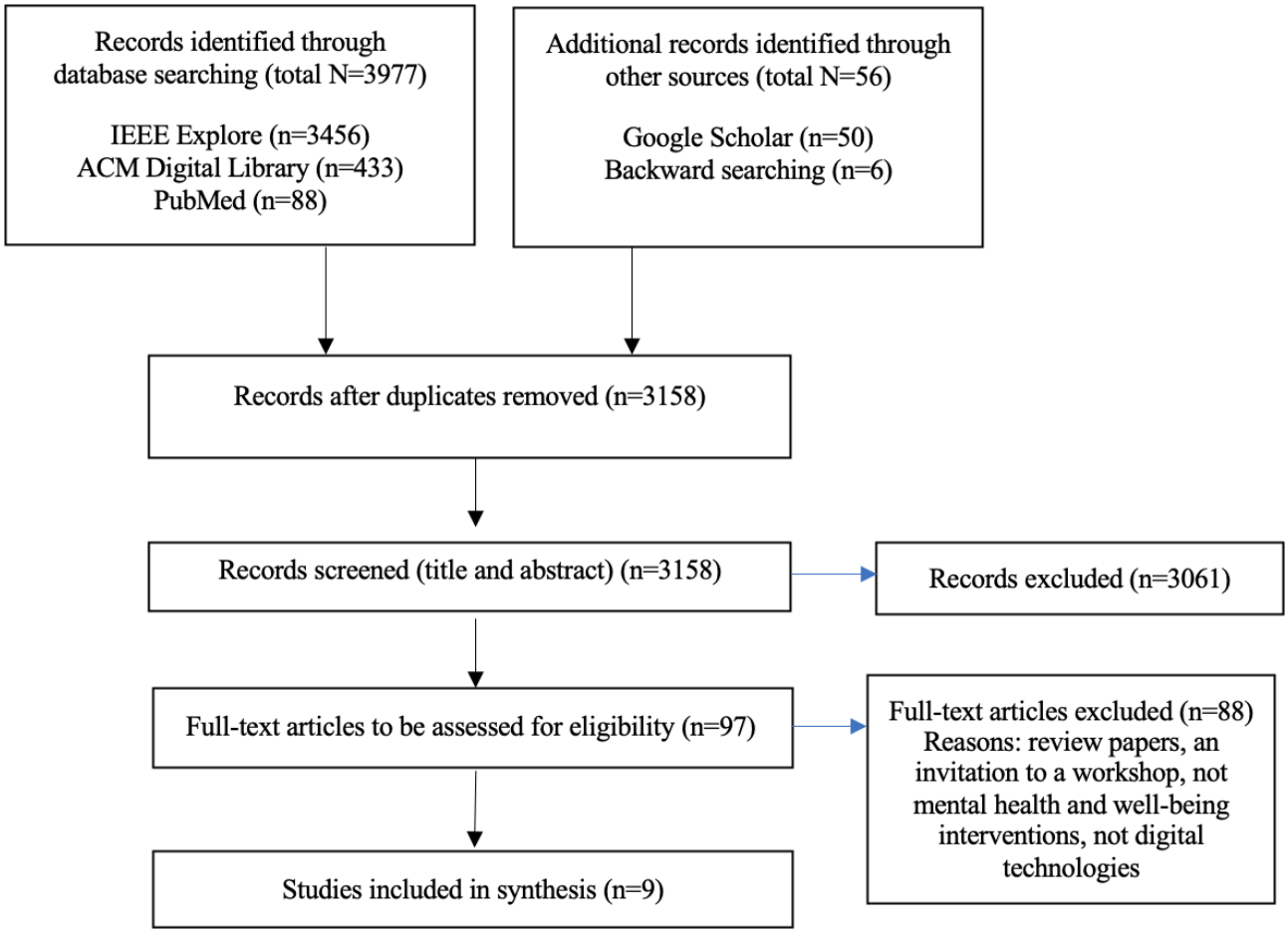
Results

Selection of Sources of Evidence

In total, 3977 articles were identified from the electronic database searches (3456 IEEE Explore, 433 ACM Digital Library, and 88 PubMed), and an additional 56 records were identified through other sources. Following title and abstract screening, 97 articles were selected for full-text review. Two authors (SI-J and AHK) independently reviewed the articles and resolved disagreements through discussion until consensus was reached. Finally, 9 papers met the eligibility criteria and were included in this review (see Figure 1 for the PRISMA-ScR flow diagram).

XSL•FO
RenderX

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) flow diagram of the study selection process.



Several papers described distinct stages or facets of prototype development, while others presented multiple prototypes, design concepts, or experiential artifacts. In total, 11 such items were included in the analysis.

Data Extraction

Data extraction was conducted independently by 2 of the authors (SI-J and AHK). Extracted information included lead author, year of publication, country of study, digital mental health intervention(s), stage(s) of the product development, MHW framing and timing (ie, whether the importance of MHW was prespecified by designers [predesign] or emerged through collaboration with prospective users and whether mental health conditions or well-being were focused), designers and/or participants, sample size (n), age range, soma awareness exercises and other activities, documentation methods, analytical approaches, and primary outcomes or conclusions (see Tables 3 and 4).

XSL•FO
RenderX

Table . Characteristics of the sources of evidence and outcomes.

Author, year of publication, and country	Digital mental health intervention(s), stages of the product development, MHW[a] framing and timing	Designers/ participants and soma awareness exercises and other activities
Ezer et al [43], 2024, Israel	Somaesthetic Meditation Wearable, prototype, predesign (well-being)	Participants, n=5 (varying experience in meditation) 1. A 10-minute meditation session, followed by an interview about their awareness of various body locations. 2. Researchers (experienced meditators) explored various low-fi wearable designs with participants. Participants, n=4 1. Researchers tested low-fi wearable designs with 4 participants. Participants, n=3 1. Researchers tested the next design iteration with 3 participants during meditation. Eastern medicine expert 1. Researchers designed the final prototype. 2. Safety and health consultation with Eastern medicine expert. Participants, n=20, age range: 24-59 years 1. 10-minute warm-up. 2. 10-minute meditation: Half of the participants meditated with the wearable and half without. 3. 2-4 minutes reflection sessions. 4. 10-minute meditation session: Half of the participants meditated with the wearable and half without. 5. A postexperience interview.
Höök et al [20], 2016; Stahl et al [35,44], 2016, 2022, Sweden	Soma Mat and Breathing Light, prototypes, predesign (well-being)	Researchers 1. Feldenkrais-lessons once a week for 2 years. 2. Researchers touched, felt, interacted with the materials and created the first prototype. Participants, n=7 1. Seven participants tried the prototypes, leading to a range of refinements of the prototype. Participants, n=22, tried the prototype in a 1-hour study: 1. Body map. 2. 30 minutes doing a body scan in the prototypes. 3. Body map. 4. Unstructured interview. Participants in 4 households for 3 months: 1. Use the prototypes 3-5 times a week. 2. Semistructured interview.

XSL•FO
RenderX

Author, year of publication, and country	Digital mental health intervention(s), stages of the product development, MHW[a] framing and timing	Designers/ participants and soma awareness exercises and other activities
Jung et al [45], 2021, Sweden	Deep Touch Pressure (DTP) garment, experiential artifact, predesign (well-being)	Workshops: Researchers, n=5, age range: 23-56 years Four collaborative soma design workshops (3 h each) exploring breathing-based interactions with shape-changing pressure feedback: 1. Feldenkrais practice. 2. Soma Bits. 3. Body maps. 4. Used the drawings to initiate a discussion into how their bodily sensations had changed throughout the session. Design sessions: Researcher, n=1, age=23 years 1. The researcher created a restrictive garment which contains the shape-changing pads, pressing against both the upper and lower back. 2. For 3 weeks, the researcher conducted 1-2 sessions daily for 30-60 minutes each.
La Delfa et al [46], 2020, Australia	Drone Chi, prototype, predesign (well-being)	The principal designer and a Tai Chi master (co-design) 1. The principal designer took regular Tai Chi lessons over 6 months. 2. Consulted the “somatic connoisseurship” of the Tai Chi master who gave the lessons, by involving him in a 2-hour co-design session and engaging with the drones together. Participants: n=32 (13 males, 19 females) 1. Participants tried using the prototype.
Montoya et al [47,48], 2023, Australia	Fluito, prototype, predesign (well-being)	Researchers 1. The main designer interacted with water in different settings over 1 month. 2. The main designer explored different technologies in water settings. 3. Researchers conducted, recorded, and documented three 30-minute sessions over 3 weeks (1 per wk) in their laboratory-installed flotation tank. 4. Slowstorming to combine interactive technology with a flotation tank. Participants: n=13 (6 females, 7 males), age ranged from 22 to 65 years (mean 31.8) 1. Float for 10 minutes in the Fluito to get used to the flotation in the tank, 2. Put on the VR[b] headset in the tank. 3. Semistructured interview 30-45 minutes.
Søndergaad et al [49], 2021, Sweden	Menstrual technologies 5 design concepts by the adolescents 1. *Värme Plagget* (heat garment) 2. *The Calm-inator* 3. *U-formen (The U-shape)* 4. *Korven (The Sausage)* 5. *Transformer* Design concepts, predesigned (well-being: emotional aspects), emerged through collaboration: (anxious and depressive feelings)	

XSL•FO
RenderX

Author, year of publication, and country	Digital mental health intervention(s), stages of the product development, MHW[a] framing and timing	Designers/ participants and soma awareness exercises and other activities
		Researchers: age ranged from 29 to 38 years from Northern European and South American countries: 1. Recalled and re-enacted their menarche experiences. Increased their bodily awareness of their current bodies in sport activities. Participants: adolescents who do sport and who recently started to menstruate, n=7, age ranged from 16 to 18 years. 4-hour workshops: 1. Drew their menstrual cycles. 2. Wrote down two menstrual memories. 3. Body map. 4. Feldenkrais-inspired lesson to create awareness of breathing and pelvis movements. 5. Body map. 6. Designed a body-worn technology that engages with their unique experience of their menstrual cycle, using Menarche Bits. 7. The participants recorded a short video of themselves presenting their design concepts on their own smartphones.

[a]MHW: mental health and well-being.

[b]VR: virtual reality.

XSL•FO
RenderX

Table . Characteristics of the sources of evidence and outcomes (continued).

Author, year of publication, and documentation methods	Analytical approaches	Primary outcomes/conclusions
Ezer et al [43] (2024)		
• Interview	• N/A[a]	• Low-fi wearable designs • Participants prefer the circular shapes and to explore various body locations • The need to support easy placements on a variety of body parts • No risk • Recommended starting with the lower abdomen or lower back
• Semistructured interview (questions adapted from the State Mindfulness Scale)	• Thematic coding process	• The warmth provided emotional comfort
Höök et al [20] (2016); Stahl et al [35,44] (2016, 2022)		
• Body map before and after the session • Unstructured interview	• N/A	• Overall, the experience was very positive, and participants expressed that they wanted to have these at home as a relaxation tool
• All interviews were transcribed verbatim	• Storytelling	• Through using the prototypes, the users learned a new experience, which gave them an alternative way of being in the world
Jung et al [45] (2021)		
Workshop: • Notes • Pictures • Videos • Soma body sheets/body maps	• N/A	• These sessions enabled the authors to narrow down the physical schematics that would facilitate meaningful interactions. They included a set of suitable pad shapes/sizes and bodily placements that could be embedded into a wearable interactive system, later finalized by one of the researchers • Four experiential esthetic qualities were developed • Calm and relaxed feelings
La Delfa et al [46] (2020)		
• Embodied sketching (use the body to act out design ideas in situ, capturing them on media like video as they unfold in time)	• Applied the method in reflective encounter with the materials	• N/A
• Video recording • Body sheets after the sessions • Semistructured interviews	• Thematic analysis	• Initially the hands were demanding the most attention. By the second session, most participants were moving more fluidly, coordinating the whole body to produce smooth moving hands • Feeling relaxed
Montoya et al [47,48] (2023)		
• Body maps, written descriptions, and video recording before, after, and where possible during these experiences	• N/A	• N/A

XSL•FO
RenderX

Author, year of publication, and documentation methods	Analytical approaches	Primary outcomes/conclusions
• Semistructured interview	• Inductive thematic analysis	• Four themes, including "Fluito as a facilitator for relaxation." All participants reported relaxation due to additional reasons relating to the interactive technology.
Søndergaad et al [49] (2021) Researchers: • Video recording • Notes • Photos Participants: • Body maps • A-lab sheets • Phone video recording	• Transcribed the participants' videos and the key moments from the video recording • Wrote thick descriptions	• The participants designed five technologies that respond to their experiences of menstrual cycles in sport. The design concepts demonstrate interactions with breathing, touch, modularity, and rhythms on intimate placements on the body, including the pelvis bone and chest • Throughout the workshop, the participants' accounts of menstrual cramps and depressive and anxious feelings went from linguistic to embodied.

[a]N/A: not applicable.

Stages of the Product Development

The products described across the included papers varied considerably in their stage of development. Some were fully developed prototypes created by researchers and evaluated by participants, while others represented earlier phases of the design process, such as design concepts or experiential artifacts. *Design concepts* refer to bodily informed, experience-driven ideas for potential future products. *Experiential artifacts* are intentionally designed systems that elicit specific experiences based on the properties of a chosen digital material, serving as exploratory tools for design teams rather than goal-oriented, deployable products [50]. Throughout this paper, we use the terms, *prototype, design concept, or experiential artifact*, to accurately reflect the development stage of each product. When discussing collectively, we use the term *interventions* for clarity and readability.

Eleven DMHW interventions created using soma design methods were found. The *Soma Mat* is a body-sized mat, which directs the user's attention by providing heat feedback to different parts of their body while they follow the instructions of a prerecorded Feldenkrais lesson [20,35,44]. The authors recommend that the *Soma Mat* should preferably be used with another intervention, the *Breathing Light* [20,35,44]. The *Breathing Light* is a lamp that measures the user's breathing, creating an ambient light that dims in cadence with their breathing. The user is recommended to lie on the *Soma Mat* while placing the *Breathing Light* above their face to provide them a private space.

The *deep touch pressure (DTP) garment* is an experiential artifact created to help increase breathing awareness and appreciation through interaction with breathing-based deep pressure [45]. *Fluito* has a combination of technology (a flotation tank, a VR headset, a heart rate sensor, and a pneumatic system), which creates different experiences for the user, including relaxation [47,48]. *Drone Chi* is inspired by the martial art of Tai Chi, providing a close-range human-drone experience by the user either following or leading the drone [46]. The *Somaesthetic meditation wearable* is a novel targeted warmth wearable intended to be used during meditation [43].

Adolescent athletes developed 5 innovative menstrual technology concepts to address menstrual cramps and premenstrual depressive and anxious feelings [49]. Using Menarche Bits [49], participants designed technologies that incorporate touch, breathing rhythms, and modularity in intimate body placements (see Table 2). *Värme Plagget* (heat garment) is a wearable garment that provides heat to alleviate menstrual pain in the lower abdomen. *The Calm-inator* is a blouse equipped with air pumps that massage the chest area, helping to increase breathing awareness. *U-formen (The U-shape)* consists of 2 heat cushions that warm the lower abdomen and pumps that create pulsating sensations on the hip bones. *Korven (The Sausage)* is a portable heat cushion to relieve menstrual pain and bloating before and during menstruation. *Transformer* consists of heat and pumping machines, which can be placed in the middle of the stomach to ease menstrual pain.

MHW Framing and Timing

Across most interventions, MHW considerations were prespecified by the designers, typically framed around goals, such as relaxation or calmness, rather than clinical treatment. In one study, the relevance to MHW emerged during collaboration with prospective users, although emotional dimensions had already been implied by the authors [49]. Overall, all included interventions focused on general well-being rather than addressing specific mental health conditions.

XSL•FO
RenderX

Designers

Five of the interventions were designed primarily by researchers. At the time of the papers' publication, 1 intervention (*DTP garment*) had been developed solely by the research team [45], whereas 4 interventions (*The Soma Mat*, *The Breathing Light*, *Drone Chi*, and *Fluito*) were first designed by researchers and subsequently tried by participants, who provided experiential feedback [20,46,47].

The *Somaesthetic Meditation Wearable* was designed by participants with varying experience in meditation [43]. The 5 menstrual technologies were designed by adolescents who played sports [49]. Specialist expertise was integrated into 2 studies: one study involved a "somatic connoisseur," a consultant specializing in a particular bodily practice (in this case, a Tai Chi master) who contributed to a co-design session [46], whereas the other consulted an expert in Eastern medicine during the design process [43].

Soma Awareness Exercises and Toolkits

Soma Awareness Exercises

The soma awareness exercises practiced most frequently during the design processes were Feldenkrais lessons. Researchers practiced Feldenkrais lessons before designing 3 of the interventions: *Soma Mat*, *Breathing Light*, *and DTP garment* [20,45], while adolescents practiced them before creating the 5 design concepts of the menstrual technologies [49]. The types of Feldenkrais lessons varied depending on the focus of each project. Other soma awareness exercises tried by the designers included Tai Chi, meditation, interacting with water, and engaging in sports [43,46,47,49].

Toolkits

The *Soma Bits Toolkit* and the *Menarche Bits* were specifically designed and created for soma design practice [26,51]. The *Soma Bits* were used to design the *DTP garment* [45], whereas the *Menarche Bits* were used by adolescents to design the 5 menstrual technologies [49] (see Table 2).

Documentation Methods

In addition to taking notes during workshops and conducting participant interviews, researchers employed a range of strategies to document the soma experiences encountered during the design process. To capture how designers or participants felt in their bodies, each person filled out the body maps [52] before and after bodily activities in 8 studies (Table 5). These body maps were filled out by researchers in the development of 2 interventions (*DTP garment* and *Fluito*) [45,47], by adolescents during the co-design of 5 menstrual technologies [49], and by participants to report their experiences with 3 of the interventions: *Soma Mat*, *Breathing Light*, and *Drone Chi* [20,35,44,46].

Table . Definitions of body maps and experiential qualities.

Documentation	Definitions
Body maps	Body maps are tools for body awareness consisting of preprinted body silhouettes on paper that can be annotated and drawn upon. They serve as a method to document and reflect on one's lived bodily experience, allowing individuals to externalize and track changes in body awareness, particularly before and after somatic practices, such as Feldenkrais. By sharing these visual representations, body maps help develop a more refined vocabulary for describing bodily experiences while fostering trust and comfort among participants.
Experiential qualities (XQs)	XQs are characteristics embedded in a design and felt during user interaction, serving as guiding principles throughout the design process while allowing creative freedom in translating somatic experiences into multiple design ideas and materials.

Researchers also frequently used video recordings to document key moments during the design process and interviews. For 3 interventions, such as *DTP garment*, *Drone Chi*, and *Fluito*, researchers recorded soma design workshops to reflect on their own experiences [45-47]. Additional recordings captured participants using the *Drone Chi* intervention [46] and adolescents engaging in the design of menstrual technologies [49]. In one study, participants were asked to record short videos on their phones to present their design concepts (menstrual technologies) [49].

Analytical Approaches

In all studies, designers engaged in first-person, subjective reflection to foreground experiential insights, facilitate iterative exploration, and foster reflective practice throughout the design process. To analyze the data collected from participants, a range of qualitative methods was employed. Thematic analysis was conducted to analyze semistructured interviews with participants for 3 of the interventions, such as *Somaesthetic Meditation Wearable*, *Drone Chi*, and *Fluito*, to identify recurring patterns and themes [43,46,47]. For the *Soma Mat* and *Breathing Light* interventions, a storytelling approach was adopted to capture and interpret individual participant experiences [35]. In many of the studies, the use of video data, which captured the moments from the design process or participant interviews, was not explicitly detailed. However, in the case of *Drone Chi*, video content was integrated into the interviews and analyzed thematically [46]. For the menstrual technologies project, researchers produced thick descriptions of key workshop moments to support interpretive analysis [49].

Outcomes

For Designers Developing the Interventions

Designers reported a range of outcomes from applying soma design methods in the development of 7 interventions. In the

XSL•FO
RenderX

case of the *DTP garment*, engagement in soma-based activities, such as Feldenkrais practice and body mapping, enabled the designers (ie, researchers) to identify physical schematics that would support meaningful interaction. These practices also helped uncover key experiential qualities [53] (see Table 5) that informed the design of the intervention [23,45].

Similarly, adolescent designers who participated in a soma design workshop to develop menstrual technologies experienced a shift in how they articulated their symptoms. Their descriptions of menstrual cramps, depression, and anxiety evolved from linguistic expressions to embodied understandings. Through this process, they began to perceive their bodies and minds not as separate entities, but as interconnected aspects of their lived experience. Their resulting design proposals aimed to alleviate menstrual discomfort by incorporating pain relief and breathing techniques to address both physical and emotional symptoms [49].

For Participants Experiencing the Interventions

Five of the 11 interventions developed using soma design were evaluated by participants who had not been involved in their creation. Overall, these participants reported positive and meaningful experiences when engaging with the interventions. For instance, users of *Fluito* described an increased awareness of their breathing, which contributed to both mental and physical relaxation [47]. Similarly, participants interacting with *Drone Chi* demonstrated a gradual shift in their engagement. Initially focused on hand movements, they began to move more fluidly over time, incorporating their entire bodies to produce smooth, coordinated gestures. Analysis of body maps, video recordings, and interviews revealed that *Drone Chi* supported a range of mental states, including meditative experiences characterized by present-moment focus and relaxation [46]. Participants using *Somaesthetic Meditation Wearable* during meditation reported that the warmth provided emotional comfort [43].

Participants who used the *Soma Mat* and *Breathing Light* for 30-minute sessions also reported positive experiences and expressed a desire to use these tools at home for relaxation. Notably, after 3 months of home use, participants reported being able to self-regulate and calm themselves during stressful moments, even without the physical presence of the prototypes, suggesting a lasting impact on their embodied coping strategies [35].

Discussion

Current Evidence for Soma Design in DMHW

This scoping review mapped the current landscape of DMHW interventions developed using soma design methods. All included studies reported positive outcomes associated with soma design. Among the 5 interventions tested by participants, users reported increased bodily awareness, full-body engagement, enhanced emotional comfort, and relaxation. Notably, 2 of the interventions demonstrated lasting effects, with participants applying the skills they had developed, such as self-regulation and relaxation, beyond the duration of the intervention sessions. Designers also reported significant benefits from using soma design. These included the identification of experiential qualities that informed the design process and a shift in how adolescent designers perceived their bodies, not as separate from the mind, but as integrated aspects of their whole being. These findings suggest that soma design can offer meaningful value for both designers and users.

However, the balance of evidence is inconclusive due to the heterogeneity in the stage of design development (ranging from conceptual designs to functional prototypes), the identity of the designers (eg, researchers vs adolescents), the extent of end-user involvement, and the analytical approaches employed. In the following section, we explore both the opportunities and challenges in advancing soma design within the DMHW field.

Opportunities and Challenges in Advancing Soma Design for DMHW

This scoping review highlighted the value of mind-body-focused approaches within DMHW, drawing attention to the often-overlooked role of bodily awareness in design processes. While much of DMHW has traditionally centered on cognitive and screen-based interventions, soma design underscores the importance of attending to the body as a site of both experience and intervention [54]. It adopts a first-person, subjective stance in which the designer's own bodily experience becomes a source of insight and knowledge production [33]. Grounded in the principles of Research through Design, soma design values reflective practice, iterative making, and somatically grounded engagement with materials and technologies [21]. These epistemological foundations not only distinguish the soma design methodologically but also enable the development of more nuanced, embodied, and user-sensitive DMHW interventions.

In doing so, soma design may offer a bridge between 2 emerging trends in health care: digital mental health and digital physical health. While these domains have often evolved in parallel, one focusing on psychological well-being through digital therapeutics, the other on physiological monitoring and movement-based interventions, soma design invites a more integrated perspective. It encourages the design of technologies that do not treat the mind and body as separate domains but instead recognize their continuous interplay [24]. This integration is particularly relevant in conditions where mental and physical symptoms co-occur [55]. By fostering embodied awareness and sensory attunement, soma design opens up new possibilities for interventions that are not only technologically innovative but also deeply attuned to the lived, felt experience of users. In this way, it contributes to a more holistic vision of digital health, one that supports the whole person, rather than isolating mental or physical aspects of well-being.

This integrative approach aligns with a broader movement within HCI and DMHW toward more body-centric technologies and interaction paradigms. Similar to HCI, the DMHW field is increasingly moving beyond traditional screen-based or VR-focused interventions and embracing emerging modalities, such as wearables, soft robotics, and other forms of embodied interaction that integrate seamlessly with the body [54]. As such, soma design contributes to ongoing conversations about the future of DMHW, suggesting that embodied, somatically

XSL•FO
RenderX

informed design practices may play a critical role in shaping more holistic and responsive mental health technologies.

This holistic vision is operationalized through a range of embodied design techniques that lie at the heart of soma design practice. These include soma awareness exercises, body mapping exercises, and the use of Soma Bits, technologies that deliver unfamiliar sensations through heat, vibration, or inflation to disrupt habitual perception and foster deeper bodily engagement [54]. These methods serve as generative design strategies, enabling participants and designers to access and articulate subtle aspects of embodied experience. In doing so, they help ground the design process in lived bodily sensation, reinforcing soma design's commitment to creating technologies that are attuned to the full spectrum of human experience.

Cocreation and Stakeholder Engagement

While soma design offers a compelling framework for creating more holistic and embodied DMHW interventions, realizing its full potential depends on meaningful engagement with the people that these technologies are intended to support. However, the studies reviewed revealed varying degrees of user participation. Only 2 studies involved end users from the onset, resulting in the development of 1 prototype and 5 distinct design concepts. In contrast, others engaged participants only during later testing phases.

In mental health research, cocreation is widely recognized as a cornerstone of effective intervention design. It involves engaging end users as active stakeholders throughout the design, development, implementation, and evaluation of digital interventions [56]. This approach aligns with the principle of responsible research and innovation, which emphasizes the importance of creating technologies that are ethically sound, socially desirable, and aligned with users' values and needs [56].

Participatory soma design holds significant potential as an embodied, holistic coproduction approach. While cocreation has gained traction in mental health research, it has predominantly focused on cognitive-based interventions, often through patient and public involvement workshops or focus groups. Participatory soma design methods present new opportunities for mental health research by offering a more embodied and experiential approach.

This is particularly important when designing for populations that differ from the researchers themselves. For example, Søndergaard et al [49] demonstrated the value of involving adolescents in the design of interventions targeted at their age group. Similarly, including individuals with lived experience of neurodiversity or mental health conditions is essential to ensure that interventions are both relevant and effective.

However, the time- and resource-intensive nature of soma design can make sustained user involvement challenging. Ideally, end users should be engaged from the early stages of design. Where this is not feasible, researchers should still involve users in later stages of refinement to ensure that interventions align with their embodied needs and preferences. Adapted or phased models of participation may help balance feasibility with meaningful engagement.

Addressing these challenges requires not only thoughtful engagement strategies but also broader collaboration across disciplines and stakeholders to support the development and implementation of effective soma design interventions.

Collaborations to Create Effective Mental Health Interventions for the Users

While soma design offers a compelling framework for embodied DMHW interventions, its broader impact relies on the ability to translate rich, experiential practices into solutions that are scalable and accessible. Despite its potential to support MHW in diverse contexts, the field currently lacks systematic evaluation frameworks to assess the effectiveness of these interventions. Much of the existing research remains exploratory, limiting its integration into DMHW practices that require rigorous validation and replicability.

Additionally, this review found a notable distribution between mental health conditions and well-being: no included intervention explicitly targeted a specific mental health condition (eg, anxiety and depression). Instead, current applications predominantly focus on well-being outcomes (eg, relaxation and calm). Addressing this gap is essential for understanding the therapeutic potential of soma design and guiding its future applications in both clinical and everyday settings. The interventions currently available may inform the development of future condition–targeted digital mental health interventions.

Furthermore, the interventions reviewed in this study varied widely in their stage of development, ranging from early-stage design concepts to advanced prototypes undergoing iterative testing. In many cases, the primary focus was on exploring and refining the methodological aspects of soma design rather than producing interventions intended for broader deployment. While this has contributed valuable insights into embodied design practices, it has also limited the immediate applicability of soma interventions in real-world mental health contexts. Nevertheless, examples, such as *Soma Mat* and *Breathing Light* [20], demonstrated the potential of soma design to deliver meaningful, positive experiences to a broader audience. These examples illustrate how embodied interaction can support well-being through sensory engagement and bodily awareness.

To fully realize this potential, interdisciplinary collaboration is essential. Bringing together mental health professionals, HCI researchers, industrial partners, and individuals with lived experience can foster the co-design, evaluation, and implementation of interventions that are both effective and scalable. Such partnerships ensure that interventions are grounded in robust design methodologies while remaining responsive to diverse user needs and real-world constraints. In particular, deepening collaboration with clinical partners and mental health researchers will be crucial for moving soma design beyond exploratory prototypes toward clinically relevant, deployable DMHW solutions.

Strengths and Limitations of This Review

This paper presents the first scoping review to systematically map existing research on DMHW interventions developed using soma design methods. To our knowledge, it is also the first to introduce soma design to the digital mental health research

XSL•FO
RenderX

community in a structured and comprehensive manner. A rigorous literature search was conducted across 3 major databases, supplemented by Google Scholar screening and backward citation tracking to ensure comprehensive coverage. Study selection and data extraction were performed in duplicate to enhance reliability. Through this process, this review offers a valuable synthesis of current knowledge and highlights the potential of soma design to contribute novel, embodied approaches to MHW research.

Despite these strengths, several limitations should be acknowledged. First, the review included only studies published in English, which may limit the cultural and geographical diversity of the findings. This narrow scope may overlook relevant work published in other languages. Second, the included studies varied widely in terms of methodological detail and reporting quality, which may affect the consistency and comparability of findings. Finally, while this review maps the current landscape of soma design in digital mental health, it does not assess the clinical effectiveness of soma design interventions, as no included studies have conducted formal evaluations.

Future research should address these limitations by involving more diverse populations of designers and participants, including individuals from underrepresented cultural and linguistic backgrounds as well as those with lived experience, such as people with mental health conditions or neurodivergence. Broadening both the methodological and demographic scope of soma design research will be essential for realizing its full potential and ensuring its relevance and effectiveness in digital mental health contexts.

Conclusions

This scoping review provides the first comprehensive synthesis of DMHW interventions developed using soma design methods. The findings suggest that soma design offers a distinctive, embodied approach that can enrich both the design process and user experience by fostering bodily awareness, relaxation, and meaningful engagement.

At the same time, the review identifies important gaps in current literature, particularly around evaluation practices, user involvement, and the diversity of design contexts. Addressing these gaps will be essential for advancing soma design from an exploratory methodology to a widely applicable framework for DMHW innovation.

Looking forward, soma design presents a valuable opportunity to reimagine how DMHW interventions are conceptualized and cocreated. By fostering deeper collaboration among mental health researchers, HCI practitioners, clinicians, industry partners, and individuals with lived experience, future work can support the development of more inclusive, effective, and embodied mental health technologies. In doing so, soma design may help shape a more holistic and human-centered future for digital mental health.

Funding

This study was funded by the National Institute for Health and Care Research (NIHR) Nottingham Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. The authors also gratefully acknowledge the support of the Engineering and Physical Sciences Research Council through the Turing AI Fellowship on Somabotics—Creatively Embodying Artificial Intelligence (grant EP/Z534808/1).

Conflicts of Interest

None declared.

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 86 KB - humanfactors_v13i1e79400_app1.docx]

References

1. Global strategy on digital health 2020-2025. : World Health Organization; 2021 URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2026-04-08]
2. De Witte NAJ, Joris S, Van Assche E, Van Daele T. Technological and digital interventions for mental health and wellbeing: an overview of systematic reviews. Front Digit Health 2021;3:754337. [doi: 10.3389/fdgth.2021.754337] [Medline: 35005695]
3. Chen T, Ou J, Li G, Luo H. Promoting mental health in children and adolescents through digital technology: a systematic review and meta-analysis. Front Psychol 2024;15:1356554. [doi: 10.3389/fpsyg.2024.1356554] [Medline: 38533221]
4. Zhang X, Qiao X, Peng K, Gao S, Hao Y. Digital behavior change interventions to reduce sedentary behavior and promote physical activity in adults with diabetes: a systematic review and meta-analysis of randomized controlled trials. Int J Behav Med 2024 Dec;31(6):959-973. [doi: 10.1007/s12529-023-10188-9] [Medline: 37391571]
5. Cushnan J, McCafferty P, Best P. Clinicians' perspectives of immersive tools in clinical mental health settings: a systematic scoping review. BMC Health Serv Res 2024 Sep 18;24(1):1091. [doi: 10.1186/s12913-024-11481-3] [Medline: 39294727]

XSL•FO
RenderX

6. Scoglio AAJ, Reilly ED, Gorman JA, Drebing CE. Use of social robots in mental health and well-being research: systematic review. J Med Internet Res 2019;21(7):e13322. [doi: 10.2196/13322]
7. Maiquez BM, Smith C, Dyke K, et al. A double-blind, sham-controlled, trial of home-administered rhythmic 10-Hz median nerve stimulation for the reduction of tics, and suppression of the urge-to-tic, in individuals with Tourette syndrome and chronic tic disorder. J Neuropsychol 2023 Sep;17(3):540-563. [doi: 10.1111/jnp.12313] [Medline: 37133932]
8. Fugate JMB, Macrine SL, Hernandez-Cuevas EM. Therapeutic potential of embodied cognition for clinical psychotherapies: from theory to practice. Cogn Ther Res 2024 Aug;48(4):574-598. [doi: 10.1007/s10608-024-10468-y]
9. Weineck F, Research MME. Embodiment research – building bridges to evidence-based clinical practice: a practitioner's guide. In: Hauke G, Kritikos A, editors. Embodiment in Psychotherapy: Springer International Publishing; 2018:113-126. [doi: 10.1007/978-3-319-92889-0_9]
10. Price TF, Peterson CK, Harmon-Jones E. The emotive neuroscience of embodiment. Motiv Emot 2012 Mar;36(1):27-37. [doi: 10.1007/s11031-011-9258-1]
11. Hyvönen K, Pylvänäinen P, Muotka J, Lappalainen R. The effects of dance movement therapy in the treatment of depression: a multicenter, randomized controlled trial in Finland. Front Psychol 2020;11:1687. [doi: 10.3389/fpsyg.2020.01687] [Medline: 32903394]
12. Koch SC, Wirtz G, Harter C, et al. Embodied self in trauma and self-harm: a pilot study of effects of Flamenco therapy on traumatized inpatients. J Loss Trauma 2019 Aug 18;24(5-6):441-459. [doi: 10.1080/15325024.2018.1507472]
13. Koch SC. Indications and contraindications in dance movement therapy: learning from practitioners' experience. GMS J Arts Ther 2020;2. [doi: 10.3205/jat000006]
14. Francesco P, Mauro MG, Gianluca C, Enrico M. The efficacy of relaxation training in treating anxiety. Int J Behav Consult Ther 2010;5(3-4):264-269. [doi: 10.1037/h0100887]
15. Riva G. Virtual reality in psychotherapy: review. Cyberpsychol Behav 2005 Jun;8(3):220-230. [doi: 10.1089/cpb.2005.8.220]
16. Riva G. Virtual reality and telepresence. Science 2007 Nov 23;318(5854):1240-1242. [doi: 10.1126/science.318.5854.1240d] [Medline: 18033867]
17. Riva G, Wiederhold BK, Mantovani F. Neuroscience of virtual reality: from virtual exposure to embodied medicine. Cyberpsychol Behav Soc Netw 2019 Jan;22(1):82-96. [doi: 10.1089/cyber.2017.29099.gri] [Medline: 30183347]
18. Freeman D, Reeve S, Robinson A, et al. Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychol Med 2017 Oct;47(14):2393-2400. [doi: 10.1017/S003329171700040X] [Medline: 28325167]
19. Shusterman R. Thinking Through the Body: Essays in Somaesthetics: Cambridge University Press; 2012. [doi: 10.1017/CBO9781139094030]
20. Höök K, Jonsson MP, Ståhl A, Mercurio J. Somaesthetic appreciation design. Presented at: CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, CA p. 3131-3142. [doi: 10.1145/2858036.2858583]
21. Höök K, Benford S, Tennent P, et al. Unpacking non-dualistic design: the soma design case. ACM Trans Comput-Hum Interact 2021 Dec 31;28(6):1-36. [doi: 10.1145/3462448]
22. Dewey J. Art as Experience: Penguin Publishing Group; 2005.
23. Ståhl A, Tsaknaki V, Balaam M. Validity and rigour in soma design-sketching with the soma. ACM Trans Comput-Hum Interact 2021 Dec 31;28(6):1-36. [doi: 10.1145/3470132]
24. Höök K. Designing with the Body: Somaesthetic Interaction Design: The MIT Press; 2018. [doi: 10.7551/mitpress/11481.001.0001]
25. Neustaedter C, Sengers P. Autobiographical design in HCI research: designing and learning through use-it-yourself. Presented at: DIS '12: Proceedings of the Designing Interactive Systems Conference; Jun 11-15, 2012; Newcastle Upon Tyne, United Kingdom p. 514-523. [doi: 10.1145/2317956.2318034]
26. Windlin C, Ståhl A, Sanches P, et al. Soma Bits - mediating technology to orchestrate bodily experiences. Presented at: Proceedings of the 4th Biennial Research Through Design Conference 2019; Mar 19-22, 2019 URL: https://zenodo.org/records/3248842 [accessed 2026-04-08]
27. Kabat-Zinn J. Full Catastrophe Living: Using the Wisdom of Your Body and Mind to Face Stress, Pain, and Illness: Bantam; 2005.
28. Feldenkrais M. Awareness Through Movement: Health Exercises for Personal Growth: Harper and Row; 1972.
29. Shusterman R. Body Consciousness: A Philosophy of Mindfulness and Somaesthetics: Cambridge University Press; 2008. [doi: 10.1017/CBO9780511802829]
30. Hummels C, Overbeeke KCJ, Klooster S. Move to get moved: a search for methods, tools and knowledge to design for expressive and rich movement-based interaction. Pers Ubiquit Comput 2007 Oct 16;11(8):677-690. [doi: 10.1007/s00779-006-0135-y]
31. Schiphorst T. Soft(n): toward a somaesthetics of touch. Presented at: CHI EA '09: CHI '09 Extended Abstracts on Human Factors in Computing Systems; Apr 4-9, 2009; Boston, MA p. 2427-2438. [doi: 10.1145/1520340.1520345]
32. Höök K. Transferring qualities from horseback riding to design. Presented at: NordiCHI '10: Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries; Oct 16-20, 2010. [doi: 10.1145/1868914.1868943]

XSL•FO
RenderX

33. Borg A, Núñez Pacheco C, Karlsson A, Höök K. Designerly ways of knowing – Soma Design meets industrial design thinking. Presented at: HttF '24: Proceedings of the Halfway to the Future Symposium; Oct 21-23, 2024. [doi: 10.1145/3686169.3686174]
34. Sheets-Johnstone M. The Primary of Movement: John Benjamins Publishing; 2011. [doi: 10.1075/aicr.82]
35. Ståhl A, Balaam M, Comber R, Sanches P, Höök K. Making new worlds – transformative becomings with soma design. Presented at: CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; Apr 29 to May 5, 2022. [doi: 10.1145/3491102.3502018]
36. Windlin C, Höök K, Laaksolahti J. Sketching Soma Bits. Presented at: DIS '22: Proceedings of the 2022 ACM Designing Interactive Systems Conference; Jun 13-17, 2022. [doi: 10.1145/3532106.3533510]
37. Bond RR, Mulvenna MD, Potts C, O'Neill S, Ennis E, Torous J. Digital transformation of mental health services. Npj Ment Health Res 2023 Aug 22;2(1):13. [doi: 10.1038/s44184-023-00033-y] [Medline: 38609479]
38. Kiefer RA. An integrative review of the concept of well-being. Holist Nurs Pract 2008;22(5):244-252. [doi: 10.1097/01.HNP.0000334915.16186.b2] [Medline: 18758272]
39. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
40. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5(1):69. [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
41. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
42. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. Syst Rev 2016;5(1):210. [doi: 10.1186/s13643-016-0384-4]
43. Ezer TS, Giron J, Erel H, Zuckerman O. Somaesthetic meditation wearable: exploring the effect of targeted warmth technology on meditators' experiences. Presented at: CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems; May 11-16, 2024. [doi: 10.1145/3613904.3642557]
44. Ståhl A, Jonsson M, Mercurio J, Karlsson A, Höök K, Banka Johnson EC. The soma mat and breathing light. Presented at: CHI EA '16: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems; May 7-12, 2016; San Jose, CA p. 305-308. [doi: 10.1145/2851581.2889464]
45. Jung A, Alfaras M, Karpashevich P, Primett W, Höök K. Exploring awareness of breathing through deep touch pressure. Presented at: CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems; May 8-13, 2021. [doi: 10.1145/3411764.3445533]
46. La Delfa J, Baytas MA, Patibanda R, Ngari H, Khot RA. Drone Chi: somaesthetic human-drone interaction. Presented at: CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems; Apr 25-30, 2020; Honolulu, HI p. 1-13. [doi: 10.1145/3313831.3376786]
47. Montoya MF, Ji Y, Wee R, et al. Fluito: towards understanding the design of playful water experiences through an extended reality floatation tank system. Proc ACM Hum-Comput Interact 2023 Sep 29;7(CHI PLAY):948-975. [doi: 10.1145/3611056]
48. Montoya MF, Ji Y, Pell SJ, Mueller F. Demonstrating fluito: a playful floatation tank experience. Presented at: CHI EA '23: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems; Apr 23-28, 2023. [doi: 10.1145/3544549.3583908]
49. Søndergaard MLJ, Ciolfi Felice M, Balaam M. Designing menstrual technologies with adolescents. Presented at: CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems; May 8-13, 2021. [doi: 10.1145/3411764.3445471]
50. Sundström P, Vaara E, Solsona J, et al. Experiential artifacts as a design method for somaesthetic service development. Presented at: RDURP '11: Proceedings of the 2011 ACM symposium on The role of design in UbiComp research & practice; Sep 18, 2011; Beijing, China p. 33-36. [doi: 10.1145/2030031.2030041]
51. Søndergaard MLJ, Kilic Afsar O, Ciolfi Felice M, Campo Woytuk N, Balaam M. Designing with intimate materials and movements. Presented at: DIS '20: Proceedings of the 2020 ACM Designing Interactive Systems Conference; Jul 6-10, 2020; Eindhoven, Netherlands p. 587-600. [doi: 10.1145/3357236.3395592]
52. Loke L, Khut GP, Kocaballi AB. Bodily experience and imagination: designing ritual interactions for participatory live-art contexts. Presented at: DIS '12: Proceedings of the Designing Interactive Systems Conference; Jun 11-15, 2012; Newcastle Upon Tyne, United Kingdom p. 779-788. [doi: 10.1145/2317956.2318073]
53. Löwgren J. Pliability as an experiential quality: exploring the aesthetics of interaction design. Artifact 2007 Oct;1(2):85-95. [doi: 10.1080/17493460600976165]
54. Höök K, Ståhl A, Jonsson M, Mercurio J, Karlsson A, Johnson ECB. Cover story somaesthetic design. interactions 2015 Jun 25;22(4):26-33. [doi: 10.1145/2770888]
55. Kind S, Otis JD. The interaction between chronic pain and PTSD. Curr Pain Headache Rep 2019 Nov 28;23(12):91. [doi: 10.1007/s11916-019-0828-3] [Medline: 31781875]
56. Jirotka M, Grimpe B, Stahl B, Eden G, Hartswood M. Responsible research and innovation in the digital age. Commun ACM 2017 Apr 24;60(5):62-68. [doi: 10.1145/3064940]

XSL•FO
RenderX

Abbreviations

DMHW: digital mental health and well-being
DTP: deep touch pressure
HCI: human-computer interaction
MHW: mental health and well-being
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
VR: virtual reality

Edited by A Kushniruk; submitted 30.Jun.2025; peer-reviewed by J Donia, J Karpova; revised version received 23.Mar.2026; accepted 25.Mar.2026; published 11.May.2026.

Please cite as:
Ito-Jaeger S, Kafadar AH, Benford S, Perez Vallejos E
Soma Design for Digital Mental Health and Well-Being Interventions: Scoping Review
JMIR Hum Factors 2026;13:e79400
URL: https://humanfactors.jmir.org/2026/1/e79400
doi:10.2196/79400

© Sachiyo Ito-Jaeger, Aysegul Humeyra Kafadar, Steve Benford, Elvira Perez Vallejos. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 11.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Engaging Older Adults With Neurocognitive Disorders in Digital Health Technologies: Scoping Review

Sié Mathieu Aymar Romaric Da[1,2], MSc; Maxime Sasseville[1,2], PhD; Marie-Soleil Hardy[1,2], PhD; Idrissa Beogo[3], PhD; Amédé Gogovor[4], PhD; Samira Amil[2,5], MD, RD; Achille R Yameogo[1,2], MSc; Florian Naye[6], PhD; Farzaneh Yousefi[1], PhD; Frédéric Bergeron[7], MSI; Anik Giguère[2,4], PhD; Annie LeBlanc[2,4], PhD; James Plaisimond[1,2], MSc; Carole Rivard-Lacroix[8]; Marie-Pierre Gagnon[1,2], PhD; Chloé Cachinho[1], MSc

[1]Faculty of Nursing Sciences, Université Laval, Pavillon Ferdinand-Vandry, local 1426, 1050, Medicine Avenue, Quebec, QC, Canada

[2]VITAM, Research Center for Sustainable Health, Integrated University Health and Social Services Center (CIUSSS) of the Capitale-Nationale, Laval University, Québec, QC, Canada

[3]School of Nursing, University of Ottawa, Ottawa, ON, Canada

[4]Faculty of Medicine, Université Laval, Québec, QC, Canada

[5]School of Nutrition, Université Laval, Québec, QC, Canada

[6]Faculty of Health Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada

[7]Library – Consulting Services Directorate, Université Laval, Québec, QC, Canada

[8]Institute for Aging and Social Participation of Seniors, Université Laval, Québec, QC, Canada

Corresponding Author:
Marie-Pierre Gagnon, PhD
Faculty of Nursing Sciences, Université Laval, Pavillon Ferdinand-Vandry, local 1426, 1050, Medicine Avenue, Quebec, QC, Canada

Abstract

Background: Population aging is associated with a growing prevalence of neurocognitive disorders among adults aged 65 years and older. Digital health technologies offer promising opportunities to support cognitive health and well-being in this population. However, their effectiveness largely depends on users' level of engagement. Despite the recognized importance of engagement in digital health, limited evidence exists on how engagement is conceptualized, measured, and related to intervention outcomes among older adults living with neurocognitive disorders.

Objective: This scoping review aimed to describe how engagement with digital health technologies among older adults with neurocognitive disorders is conceptualized and measured, examine the relationship between engagement and the effectiveness of digital health interventions, and identify factors that facilitate or hinder engagement.

Methods: A scoping review was conducted following the Joanna Briggs Institute methodological guidance and reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist. A comprehensive search strategy, developed in collaboration with an information specialist, was applied to MEDLINE, Embase, CINAHL, Web of Science, and Google Scholar, without date restrictions. Empirical studies involving adults aged 65 years and older living with neurocognitive disorders and using digital health technologies were included. Study selection and data extraction were performed independently by at least 2 reviewers, and the results were synthesized narratively.

Results: Of the 1665 records identified after duplicate removal, 2 studies met the inclusion criteria. One study examined computerized cognitive stimulation and cognitive engagement programs among community-dwelling older adults with mild neurocognitive disorders, whereas the other explored the use of a personalized digital reminiscence application in long-term care settings among individuals with major neurocognitive disorders. No study used a validated instrument to directly measure engagement. However, observable indicators and markers related to the behavioral, cognitive, and affective components of engagement were reported. Both studies also documented concurrent cognitive or psychosocial outcomes. Factors facilitating engagement included professional support, content personalization, and involvement of informal caregivers, whereas limiting factors included cognitive fluctuations, fatigue, technical complexity, and reliance on external support.

Conclusions: This scoping review highlights a significant gap in the literature regarding the explicit conceptualization and standardized measurement of engagement with digital health technologies among older adults living with neurocognitive disorders. The findings underscore the need to develop and apply multidimensional, context-sensitive engagement measurement tools tailored to this population to better understand and optimize digital health interventions.

(JMIR Hum Factors 2026;13:e70157) doi:10.2196/70157

XSL•FO
RenderX

KEYWORDS

engagement; older adults; neurocognitive disorders; digital health technologies; eHealth

Introduction

Background

Population aging is a rapidly accelerating global phenomenon [1]. Between 2015 and 2050, the proportion of older adults worldwide is expected to nearly double, increasing from 12% to 22% [1,2]. Aging, defined as the progressive decline of physiological functions necessary for survival and reproduction [3], is frequently associated with an increased prevalence of neurocognitive disorders [4].

Neurocognitive disorders refer to an alteration in one or more cognitive functions, regardless of their underlying mechanism, etiology, or reversibility [5]. Such impairments may result from neurological, psychiatric, pharmacological, or other causes [5] and can affect various cognitive domains, including attention, memory, executive function, visuospatial abilities, and language [5]. In clinical and diagnostic literature, neurocognitive disorders are generally classified into 2 main categories according to the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-5*) [6]: mild neurocognitive disorder, which corresponds to the concept historically referred to as mild cognitive impairment (MCI), and major neurocognitive disorder, which largely corresponds to what is described as dementia in the International Classification of Diseases, still widely used in the international literature [6,7].

Major neurocognitive disorder is characterized by a significant, acquired, and progressive decline in cognitive abilities that are severe enough to interfere with independence in activities of daily living [5]. In contrast, MCI involves objectively measurable cognitive decline with relatively preserved functional autonomy [5,8].

Although most individuals living with neurocognitive disorders are aged 65 years and older, cognitive disorders can also affect younger populations [9]. In this scoping review, the term “older adults” refers to individuals aged 65 years and older, in accordance with commonly used criteria by statistical and public health organizations in Quebec and Canada [10-12].

Globally, the prevalence of MCI among older adults is estimated to range from 15.6% to 23.7%, depending on community or institutional settings [13]. In addition, approximately 55 million people worldwide are currently living with major neurocognitive disorders, nearly 73% of whom are older adults [14]. This number continues to rise rapidly, with nearly 10 million new cases diagnosed each year [14]. Major neurocognitive disorder encompasses several clinical entities, including Alzheimer disease, which accounts for 60% to 70% of cases, and cerebrovascular diseases, and less common forms, such as frontotemporal degeneration, dementia with Lewy bodies, and mixed neurocognitive disorders [9,14].

In Canada, in 2025, it is estimated that nearly 771,939 individuals may live with major neurocognitive disorders, a number projected to reach approximately 1 million by 2030 and 1.7 million by 2050 [15]. The consequences of neurocognitive disorders, whether mild or major, are substantial for older adults themselves as well as for their informal caregivers (family members and friends) and formal caregivers (health and social care professionals). For example, informal caregivers in Canada provide more than 580 million hours of care annually to older adults living with neurocognitive disorders, equivalent to approximately 290,000 full-time jobs [15]. The associated costs of care and support exceed CAD $10 billion per year [9].

In response to this growing burden, strategies aimed at preserving cognitive health in older adults have become a public health priority. Two key mechanisms have been identified in the literature: cognitive reserve and vascular reserve [16]. Cognitive reserve refers to the brain’s capacity to recruit alternative neural networks and strengthen synaptic connections to compensate for neuropathological damage, thereby maintaining cognitive performance despite underlying brain changes [16,17]. Vascular reserve, in turn, reflects the ability of the cerebral vascular system to maintain adequate oxygen and nutrient delivery despite vascular insults or hemodynamic variations, contributing to overall brain resilience [16,17]. Together, these complementary mechanisms may delay the clinical expression of neurocognitive disorders.

Several protective factors that may influence these mechanisms have been identified, including physical activity, social participation, and cognitive engagement [16,18-20]. These factors can be supported through various interventions, including the use of digital health technologies.

Digital health technologies encompass stand-alone software applications as well as integrated systems and devices accessible via computers, smartphones, tablets, wearable devices, or environmental sensors [21,22]. Among older adults, these technologies, often referred to as gerontechnologies, are used to support multiple aspects of health care and health promotion, such as cognitive stimulation, physical activity monitoring, and communication with family members and health care professionals [4,23].

However, for digital health technologies to achieve their intended benefits, user engagement remains a central determinant. Engagement with digital health technologies is commonly conceptualized as a multidimensional construct comprising 3 interrelated components: behavioral, cognitive, and affective [24-27]. The behavioral component refers to patterns of technology use, including frequency, duration, intensity, and depth of use [26]. The cognitive component encompasses constructs, such as attention and interest directed toward the technology [24,25]. Attention reflects the allocation of cognitive resources to specific information at the expense of competing stimuli [25], whereas interest represents a more enduring preference for certain activities, promoting voluntary engagement and the use of specific cognitive strategies [27]. Finally, the affective component relates to emotional responses elicited by the technology, including hedonic valence (pleasure or displeasure) and emotional activation [27].

XSL•FO
RenderX

Within digital health interventions, technologies function as stimuli designed to reinforce use, attention, interest, and affective responses to support behavior change [28]. Nevertheless, the literature reveals considerable heterogeneity in how engagement with digital health technologies is defined and measured. Engagement is frequently conflated or used interchangeably with related constructs, such as use, adherence, acceptability, satisfaction, feasibility, adoption, or retention [27,29]. In this scoping review, engagement is examined according to the 3-component conceptualization (behavioral, cognitive, and affective) to ensure theoretical consistency [24-27].

Although the effectiveness of digital health technologies for individuals living with neurocognitive disorders has been widely documented [20], knowledge regarding engagement with these technologies among older adults with neurocognitive disorders remains fragmented. Yet, the effectiveness of digital interventions largely depends on their adoption, sustained use, and integration into daily life, all of which require sufficient user engagement [30]. Engagement is therefore considered a key factor in explaining why certain digital health technologies are beneficial for some older adults but not for others [31].

In practice, some older adults may have access to digital health technologies and possess the necessary skills to use them, yet choose not to engage, often due to issues related to design, content, or functionality [26,32]. Others may initially engage but discontinue use over time. For example, a systematic review examining attrition and adherence in mobile mental health interventions among older adults reported attrition rates of up to 30%, with particularly low adherence rates ranging from 2% to 10% [33]. Factors associated with disengagement included dissatisfaction with perceived progress, time and effort demands, and limited usability [30].

Given the limited and heterogeneous evidence on engagement with digital health technologies among older adults living with neurocognitive disorders, there is a clear need to map existing knowledge. A comprehensive understanding of how engagement is conceptualized, assessed, and related to intervention effectiveness is essential to support the development, adaptation, and implementation of more relevant and inclusive digital health interventions for the prevention and management of neurocognitive disorders in older adults.

Study Objectives and Research Questions

The objective of this scoping review was to describe how engagement with digital health technologies among older adults living with neurocognitive disorders is conceptualized and assessed in the scientific literature. It also aimed to explore factors that facilitate or hinder engagement and to examine the relationship between engagement and the effectiveness of digital health interventions.

The review addressed the following research questions:

1. How is engagement conceptualized and operationalized in studies examining digital health technologies for older adults living with neurocognitive disorders?
2. What measures are used to assess the level of engagement of older adults living with neurocognitive disorders in digital health technologies?
3. What factors facilitate or hinder engagement with digital health technologies among older adults living with neurocognitive disorders?
4. What relationships are reported between engagement levels and the effectiveness of digital health interventions?

Methods

Study Design

This study adopted a scoping review methodology to map the existing body of knowledge, clarify conceptual boundaries, and identify gaps in the literature regarding engagement with digital health technologies among older adults living with neurocognitive disorders [34]. Scoping reviews are particularly appropriate when concepts are heterogeneous, variably defined, and examined using diverse methodological approaches [34].

We conducted the review following the Joanna Briggs Institute (JBI) methodological guidance for scoping reviews [35]. Reporting followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist (Checklist 1) [36] to ensure transparency and reproducibility.

Eligibility Criteria

Eligibility criteria were defined a priori using the Participants (Population), Concept, and Context framework recommended by the JBI [35].

Population

Studies involving older adults aged 65 years and older living with neurocognitive disorders, whether mild or major, were considered eligible. Given the variability in how age was reported across studies, exceptions were allowed. Specifically, studies were included when the mean or median age of participants was greater than 65 years [37]. Furthermore, several conceptual frameworks and recent evidence syntheses indicate that the involvement of caregivers, whether formal or informal, is a key determinant of digital engagement among individuals living with a neurocognitive disorder. This is largely due to their role in content personalization, support for technology use, and technological mediation between the user and the digital tool [38-40]. Accordingly, although caregivers were not considered the primary population of interest in this review, their data were included when they directly informed the engagement of older adults living with a neurocognitive disorder in the use of digital health technologies.

Concept

The central concept of this review was engagement with digital health technologies. No restrictions were imposed on how engagement was defined in the included studies, in order to capture the diversity of conceptualizations present in the literature. Digital health technologies included stand-alone software applications as well as integrated systems and devices accessible via platforms, such as computers, smartphones, tablets, wearable devices, and environmental sensors [21,22].

XSL•FO
RenderX

Context

Studies conducted in any setting (eg, health care facilities, long-term care settings, residential facilities for older adults, or community-based environments) were eligible for inclusion, with no geographical restrictions.

Exclusion Criteria

Publications that did not report original empirical data were excluded. These included knowledge syntheses, such as systematic reviews and narrative reviews, as well as commentaries, editorials, and study protocols. Conference abstracts, theses, dissertations, and gray literature were also excluded.

Search Strategy

The search strategy (Multimedia Appendix 1) was developed in collaboration with an experienced health sciences librarian specialized in knowledge synthesis (FB). It combined free-text terms and controlled vocabulary related to the concepts of *engagement*, *digital health*, *older adults*, and *neurocognitive disorders*, including relevant synonyms and spelling variants.

An initial comprehensive search was conducted on April 4, 2024, and updated on October 9, 2025. The databases searched included MEDLINE (Ovid), Embase, CINAHL, Web of Science, and Google Scholar. Considering the known limitations of search engines and based on preliminary testing, we elected to perform a series of structured subsearches [41]. To maximize the effectiveness of the relevance-ranking algorithm, each subsearch was restricted to the top 20 results. In total, this strategy produced 160 records, including those retrieved during the update. Only studies published in English or French were included. No other restrictions were applied. Finally, backward citation searching was performed on all included studies to identify further relevant publications.

Study Selection

All records retrieved from the databases were imported into Covidence (Veritas Health Innovation), a web-based collaborative software designed to manage systematic and scoping reviews [42]. A total of 780 duplicates were eliminated using the automated function of Covidence, and 5 duplicates were removed manually.

Study selection was conducted in two stages: (1) title and abstract screening and (2) full-text review. At each stage, eligibility criteria were applied independently and blindly by several reviewers (SMARD, FN, SA, ARY, FY, JP, and CC). Discrepancies were resolved through discussion and consensus, and when necessary, by consultation with the senior investigator (MPG).

A PRISMA 2020 flow diagram was used to document the identification, screening, eligibility assessment, and inclusion of studies, as well as reasons for exclusion at the full-text review stage [43].

Data Extraction

A standardized data extraction form was developed by the research team using Microsoft Excel. Data extraction was performed independently by several reviewers (SMARD, ARY, FN, and SA) and subsequently validated by the senior investigator (MPG). Discrepancies were resolved through consensus among the reviewers involved or, when needed, with the support of a third team member.

The following data were extracted: (1) study and intervention characteristics, such as title, authors, year of publication, country, study design, eligibility criteria, care setting, target population, type of digital health technology, duration, and reported effectiveness; and (2) engagement-related variables, such as conceptualization of engagement, operationalization of engagement, engagement indicators (behavioral, cognitive, and affective), and measurement tools used.

In addition, variables derived from the PROGRESS-Plus framework for health equity were extracted [32]. This framework includes 8 core dimensions (place of residence, race/ethnicity, occupation, gender, religion, education, socioeconomic status, and social capital) and an additional "Plus" component intended to capture other characteristics that may influence health inequities. The use of this framework allowed for an examination of the extent to which digital health technologies were accessible and adapted to potentially vulnerable subgroups, including older adults living with neurocognitive disorders [44].

Data Analysis

A descriptive analysis was conducted to synthesize the characteristics of the included studies, which were presented in a tabular form.

To synthesize findings on engagement with digital health technologies, we used a 3-dimensional conceptualization of engagement (behavioral, cognitive, and affective) as a descriptive analytical tool [24-27]. This framework served solely to organize and categorize elements reported in the included studies, without assessing the relevance, validity, or psychometric properties of the conceptual model. This approach aligns fully with the purpose of a scoping review, which aims to map concepts, definitions, and operationalization modes of engagement in the existing literature [34].

Variables reported in the included studies were recoded a posteriori by the research team, under the supervision of the principal investigator (MPG), into engagement indicators corresponding to the 3 components, to ensure a theoretically coherent interpretation of findings derived from heterogeneous study designs.

We did not assess the methodological quality of the included studies, in accordance with the JBI guidance for scoping reviews [35]. The objective of this review was to map existing evidence rather than evaluate effectiveness or draw conclusions based on the level of evidence.

Results

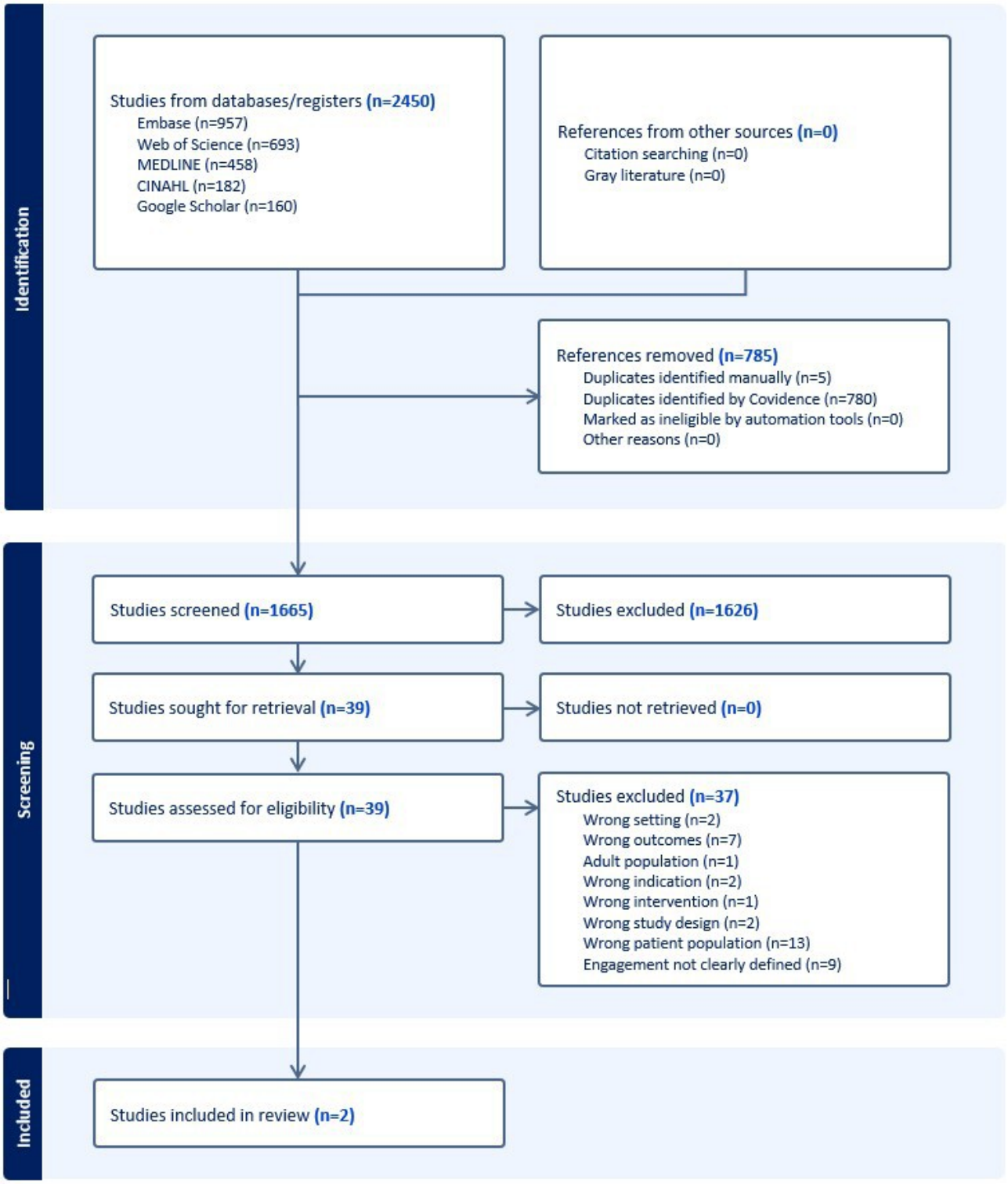
Study Selection

The search strategy identified a total of 1665 records after the removal of duplicates (Figure 1). During the first screening stage, based on titles and abstracts, 37 publications were retained for full-text review. Following full-text assessment, 2

XSL•FO
RenderX

publications [45,46], reporting on 2 distinct studies, met all eligibility criteria and were included in the scoping review.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flowchart for study inclusion [43].



Study Characteristics

The first study [45] assessed the feasibility and acceptability of 2 distinct digital interventions among older adults living with MCI. It compared a computerized cognitive stimulation (CCS) program with a computerized cognitive engagement (CCE) program. The CCS program aimed to stimulate multiple cognitive domains through computerized cognitive exercises combined with structured social interactions among participants. In contrast, the CCE program focused on learning to use a digital tablet, combined with the stimulation of social interactions within the group.

This study used a single-blind randomized controlled trial design. Participants were community-dwelling adults aged 60 years and older with MCI, recruited from a memory clinic in France. A total of 20 participants were randomly and equally

XSL•FO
RenderX

allocated to the CCS and CCE groups. The mean age was 78.2 (SD 7.0) years in the CCE group and 75.2 (SD 6.4) years in the CCS group. Both groups participated in weekly group sessions lasting 90 minutes over a 3-month period. Sessions were facilitated by a trained neuropsychologist who was blinded to the outcome assessments.

The second study [46] explored strategies to enhance engagement among older adults living with dementia in a long-term care setting using a digital reminiscence application called *Memory Keeper*. This application, developed by the study authors, aimed to support well-being and engagement among individuals with dementia by stimulating personal memories through personalized multimedia content.

The authors examined barriers and facilitators to using the application, perceived benefits of its integration into long-term care, and its potential use within existing care practices. Participants included 3 residents (2 men and 1 woman) aged 76 to 87 years with major neurocognitive disorders, 6 informal caregivers (family members or close relatives), and 1 formal caregiver (the facility's lifestyle coordinator). Participants were recruited using purposive sampling.

Informal caregivers contributed to the creation of personalized content (photographs and image libraries) integrated into the application and were trained in its use. Among the residents with major neurocognitive disorders, 1 had a diagnosis of dementia with Lewy bodies, 1 had Alzheimer disease, and 1 had dementia of an unspecified type. Although the study design was not explicitly stated in the *Methods* section, data collection and analysis procedures were qualitative in nature (see Table 1 for detailed study characteristics).

Table 1 presents the main characteristics of the included studies, including country and year of publication, study population, type of neurocognitive disorders, digital health intervention, study design, duration, and setting.

Table . Characteristics of included studies.

Study and year	Country	Population	Type of neurocognitive disorders	Intervention	Design	Duration	Setting
Djabelkhir et al [45] (2017)	France	20 community-dwelling older adults; mean age >75 years; CCE[a]: 78.2 (SD 7) years; CCS:[b] 75.2 (SD 6.4) years	MCI[c]	CCS versus CCE	Randomized controlled trial, single-blind	3 months, weekly 90-min sessions	Community
McAllister et al [46] (2020)	Australia	3 residents (ages 76 - 87 y); 6 informal caregivers; n=1 formal caregiver	Dementia: Alzheimer, Lewy body, and unspecified	Digital reminiscence application (Memory Keeper)	Qualitative methods	Duration not specified (real-world exploration)	Long-term care facility

[a]CCE: computerized cognitive engagement.

[b]CCS: computerized cognitive stimulation.

[c]MCI: mild cognitive impairment.

Conceptualization and Operationalization of Engagement With Digital Health Technologies

Conceptualization of Engagement

Both included studies [45,46] conceptualized engagement as a sustained interaction between older adults living with neurocognitive disorders and digital health technologies, mediated by digital stimuli and embedded within a social dimension.

More specifically, the study involving older adults with MCI [45] conceptualized engagement through participation in cognitively stimulating group-based activities combining computerized cognitive exercises and structured social interactions. In contrast, the study conducted among individuals with dementia in long-term care [46] conceptualized engagement as observable reactions and dyadic interactions elicited by personalized digital stimuli (reminiscence content) within a caregiving context.

Operationalization of Engagement

Engagement with digital health technologies was operationalized heterogeneously across the included studies [45,46]. However, the analytical recoding conducted in this scoping review enabled the alignment of all reported indicators with the 3-component conceptualization of engagement (behavioral, cognitive, and affective) [24-27], which was used as the analytical framework.

It is important to note that this 3-component analytical framework was applied a posteriori to heterogeneous data for analytical purposes rather than evaluative assessment, in accordance with the objectives of a scoping review.

At the behavioral level, engagement was documented through indicators, such as session attendance and attrition [45]. In the long-term care context, behavioral engagement was reflected by the frequency and duration of visits and by observable interaction-related behaviors during application use, including dancing or tapping movements [46]. The cognitive component of engagement was operationalized using standardized

XSL•FO
RenderX

neuropsychological tests, including the Trail Making Test Part A (TMT-A, attentional speed), Trail Making Test Part B (TMT-B, error reduction), free recall tasks (RL/RI-16), and the visuospatial memory test [45]. In another study, cognitive engagement was reflected by qualitative markers, such as sustained attention to the application and spontaneous recall of names or details associated with personalized reminiscence content [46]. Finally, the affective component of engagement was operationalized through indicators, such as motivation, self-esteem, and technology acceptance [45]. In the long-term care context, affective engagement was reflected by observable emotional markers, including smiling, singing, positive facial expressions, absence of negative reactions, and visits perceived as more pleasant by caregivers [46].

Measurement of Engagement

This section describes the approaches used to measure engagement as reported in the included studies and highlights the diversity of methods used to discuss capture engagement with digital health technologies among older adults living with neurocognitive disorders. It also specifies the participant groups to which the different measures were applied, in order to reflect study-specific data collection modalities, as well as the engagement components documented in the existing literature.

None of the included studies [45,46] used a validated instrument specifically designed to directly measure engagement with digital health technologies. However, in a study by Djabelkhir et al [45], a combination of validated instruments was administered to participants in the CCE and CCS groups to assess selected indicators associated with engagement. Executive functions were measured using the TMT-A and TMT-B. Episodic memory was assessed using the RL/RI-16 free and cued recall test as well as the visuospatial memory test from the cognitive efficiency profile. Global self-esteem was measured using the Rosenberg Self-Esteem Scale, and acceptance of information and communication technologies was assessed using a technology acceptance questionnaire. In the other study [46], engagement among residents living with dementia was documented using observable qualitative markers, primarily collected during interactions with the application, based on observations reported by caregivers and staff.

From an analytical perspective, the reported measures were subsequently examined according to the behavioral, cognitive, and affective components of engagement, while considering the participant groups associated with each type of measure.

The behavioral component was assessed using participation- and use-related measures collected from participants in the CCE and CCS groups in the study by Djabelkhir et al [45], and from residents in the study by McAllister et al [46], including indicators related to attendance, attrition, and the frequency and duration of interactions, supplemented by observations of interaction-related behaviors.

The cognitive component was documented through cognitive performance assessments administered to participants in the CCE and CCS groups, as well as through qualitative markers reflecting attention and memory processes mobilized by residents during technology use.

Finally, the affective component was captured using psychosocial indicators collected from participants in the CCE and CCS groups and through observational markers documenting emotional responses and residents' perceived experiences during interactions, as reported by caregivers.

Together, these approaches allowed engagement to be examined across behavioral, cognitive, and affective components (Table 2).

XSL•FO
RenderX

Table . Indicators and markers by engagement component.

Engagement component and type of measure	Participant groups	Instruments/markers used
Behavioral		
Participation measures	Participants in the CCE[a] and CCS[b] groups (study [45])	Attendance at activities and attrition
Use-related measures	Residents living with dementia (study [46])	Frequency and duration of interactions with the application
Behavioral observation	Residents, observations reported by caregivers and staff (study [46])	Interaction-related behaviors with the technology (eg, movements and gestures)
Cognitive		
Cognitive performance assessments	Participants in the CCE and CCS groups (study [45])	Trail Making Test (parts A and B), free and cued recall (RL/RI-16), and VST[c]
Qualitative cognitive markers	Residents, observations by caregivers and staff (study [46])	Attention to the application and engagement of memory processes during interactions
Affective		
Psychosocial indicators	Participants in the CCE and CCS groups (study [45])	Motivation, self-esteem (Rosenberg Self-Esteem Scale), and technology acceptance
Observational affective markers	Residents, observations reported by caregivers (study [46])	Emotional responses and perceived experience during interactions

[a]CCE*:* computerized cognitive engagement.

[b]CCS*:* computerized cognitive stimulation.

[c]VST: visuospatial memory test.

Relationship Between Engagement and the Effectiveness of Digital Health Interventions

Across the included studies, different analytical approaches were used to present the findings. A study by Djabelkhir et al [45] adopted a quantitative approach, relying on pre-post statistical analyses to examine cognitive and psychosocial data collected from participants in the CCE and CCS groups; for selected indicators, effect sizes were also reported to characterize the magnitude of the observed changes. In contrast, a study by McAllister et al [46] used a qualitative descriptive approach and used a framework analysis to structure and synthesize observational data derived from interactions between residents living with dementia and the digital application, *Memory Keeper*.

In this scoping review, these analytical methods are reported for descriptive purposes only, to contextualize how the findings were generated, without inferring any causal relationship between engagement and the reported outcomes.

Both included studies [45,46] provided data that enabled the concurrent examination of engagement manifestations and cognitive or psychosocial outcomes associated with digital health interventions among older adults living with neurocognitive disorders, across different contexts and levels of severity.

In the study involving older adults with MCI [45], sustained engagement with digital health interventions was observed alongside improvements in cognitive performance. These improvements included gains on the TMT-A and TMT-B, a medium effect size for free recall, and a trend toward improvement in visuospatial memory.

In the long-term care study involving older adults with dementia [46], engagement was not assessed using standardized measures. However, qualitative observations indicated that active participation by residents, characterized by sustained attention, recall of personal details, and increased interactions with caregivers, coincided with positive experiences during the use of the digital reminiscence application.

At the affective level, both studies reported convergent indicators [45,46]. High motivation, improved self-esteem, and increased technology acceptance were documented in a study by Djabelkhir et al [45], whereas the other study reported positive emotional expressions, interactions perceived as more pleasant, and an absence of negative reactions during the application use [46].

Facilitating and Limiting Factors of Engagement With Digital Health Technologies

Overview

The facilitating and limiting factors of engagement with digital health technologies identified in the included studies are presented in Table 3. These factors were identified transversally based on the data reported in the studies [45,46] and interpreted in light of the 3-component conceptualization of engagement (behavioral, cognitive, and affective), which was used as the analytical framework in this scoping review.

XSL•FO
RenderX

Table . Facilitators and barriers to engagement.

Factor type and reported factors	Engagement components involved	Studies
Facilitating		
Structured and professionally supervised activities, integration of technologies into existing contexts (group sessions and family visits), and intuitive interfaces	Behavioral	[45,46]
Guided learning, task adaptation, and personalization of digital content	Cognitive	[45,46]
Motivation, sense of personal value, and positive technology acceptance	Affective	[45]
Limiting		
Time constraints, reliance on caregiver or staff availability, and perceived activity duration	Behavioral	[45,46]
Fatigue, drowsiness, and fluctuations in cognitive state and attention	Cognitive	[46]
Loss of personal relevance of content, sedated states, and emotional disengagement	Affective	[46]

Facilitating Factors of Engagement

Several facilitating factors of engagement were reported across the included studies, although their nature varied according to the intervention context and the population studied.

At the behavioral level, engagement was facilitated by participation in structured and professionally supervised activities as well as by the integration of technologies into existing use contexts, such as group sessions in community settings [45] or family visits and interactions with staff in long-term care settings [46]. Ease of use and the intuitive nature of digital interfaces were also reported as factors facilitating interaction with technologies [45,46].

At the cognitive level, facilitating factors included guided learning in the use of technologies and the adaptation of tasks or digital content to participants' abilities [45]. The personalization of digital content, particularly in the form of reminiscence materials, was also described as supporting attention and cognitive engagement, especially among individuals living with dementia [46].

Finally, at the affective level, facilitating factors, such as sustained motivation to participate, a sense of personal value, and positive acceptance of digital technologies, were reported [45]. These elements were primarily described in the study conducted among older adults living with MCI.

Limiting Factors of Engagement

The included studies also identified several limiting factors of engagement with digital health technologies. At the behavioral level, organizational constraints, such as limited time available for technology use or reliance on the availability of caregivers or staff to initiate use, were reported as barriers to engagement [46]. In some cases, the perceived duration of certain activities was also identified as a source of frustration [45].

At the cognitive level, fluctuations in the cognitive state, including fatigue, drowsiness, or variations in attention, were described as factors limiting engagement, particularly among individuals living with dementia [46]. These factors could restrict the ability to maintain sustained interaction with the technologies.

Finally, at the affective level, limiting factors were reported when digital content lost its personal relevance or meaning for participants [46]. Sedated states or transient emotional disengagement were also mentioned as elements likely to limit the experience perceived as positive during technology use [46].

Facilitating and limiting factors are presented in Table 3.

Discussion

This scoping review aimed to examine how engagement with digital health technologies is conceptualized, operationalized, and measured among older adults living with mild or major neurocognitive disorders and to explore its relationship with intervention effectiveness.

Principal Findings

The findings highlight a major gap in the scientific literature regarding engagement with digital health technologies among older adults living with neurocognitive disorders. Despite a comprehensive search strategy across multiple databases, only 2 empirical studies [45,46] met the inclusion criteria among the 1665 records identified after the removal of duplicates. These findings suggest that empirical research explicitly addressing engagement in older adults living with neurocognitive disorders remains limited. The results show that although engagement was neither explicitly defined nor directly measured in the included studies, the reported data encompassed all dimensions commonly associated with engagement in digital health, as described in the literature. These manifestations included high attendance and low attrition, signs of sustained attention and cognitive mobilization, and positive affective responses during technology use. In addition, this review identified several factors that facilitated or hindered engagement, which were related to

XSL•FO
RenderX

technological characteristics, individual factors, and contextual conditions.

Although the studies reported engagement-related indicators alongside cognitive or psychosocial outcomes, the absence of standardized and explicit engagement measures precludes any causal interpretation. These findings highlight the need to better conceptualize, operationalize, and measure engagement with digital health technologies among older adults living with neurocognitive disorders, while accounting for individual, social, and contextual determinants.

Conceptualization and Operationalization of Engagement

The findings of this review indicate that engagement with digital health technologies among older adults living with neurocognitive disorders remains variably and often implicitly conceptualized. This observation is consistent with previous literature showing that engagement is frequently conflated with related constructs, such as use, adherence, acceptability, or retention, rather than being treated as a distinct concept [29,31].

In the included studies, engagement was implicitly conceptualized as sustained interaction between the individual and digital stimuli, embedded within a social context (group-based or dyadic). This implicit conceptualization aligns with dominant theoretical frameworks in digital health, which describe engagement as a multidimensional process comprising behavioral, cognitive, and affective components [22,24-27].

Although the 3-component conceptualization of engagement was not explicitly adopted by the included studies, the results indicate that the observed operationalization, reconstructed a posteriori in this review, covered all 3 components. This suggests that among older adults living with neurocognitive disorders, engagement extends beyond mere technical use and encompasses central cognitive and affective dimensions.

Recent methodological work emphasizes the importance of adapting engagement measurement to the context of use and study objectives, acknowledging that available tools variably capture different engagement components [25,26]. A key methodological contribution of this review lies in the application of a 3-component analytical framework to harmonize findings derived from heterogeneous study designs. This approach highlights that, in older adults living with mild or major neurocognitive disorders, observable affective responses (eg, smiling and singing) may be as informative as traditional behavioral usage metrics [24,26,28].

Measurement of Engagement

None of the included studies used a validated instrument specifically designed to measure engagement with digital health technologies in older adults living with neurocognitive disorders. This limitation mirrors broader challenges in the digital health literature, including the lack of methodological consensus and the predominance of indirect indicators [22,27].

In long-term care contexts, the use of validated observational tools, such as those developed to assess engagement among individuals with neurocognitive disorders [28], could allow for a more nuanced assessment of experiential engagement, particularly among individuals with major neurocognitive disorders. The absence of standardized engagement measures limits comparability across studies and constrains the ability to formally examine engagement as a mechanism of action in digital interventions.

Facilitating and Limiting Factors: The Central Role of Context and Human Support

Facilitating factors identified in this review included content personalization, ease of use, professional supervision, and involvement of informal caregivers. These findings are consistent with user-centered models emphasizing alignment between individual capacities, technological design, and contextual conditions [24,47].

Conversely, several limiting factors were identified, including technical complexity, loss of personal relevance of content, fatigue, drowsiness, fluctuations in cognitive state, and organizational constraints in care settings. Reliance on external support to initiate technology use emerged as a particularly important barrier, underscoring that engagement cannot be considered independently of the social and institutional environments in which technologies are deployed.

Engagement and Effectiveness of Digital Health Interventions

Although the number of included studies was limited, the available data allow for a cautious discussion of the potential link between engagement and intervention effectiveness. The studies reported concurrent engagement manifestations and cognitive or psychosocial outcomes, without enabling causal or directional conclusions. This limitation is consistent with the broader digital health literature, in which engagement is often assumed to be a key mechanism but is rarely empirically tested as such [22,30].

These observations align with a wider body of evidence suggesting that well-designed and appropriately adapted digital interventions may yield modest to moderate cognitive benefits among individuals living with neurocognitive disorders, despite substantial methodological heterogeneity [48,49]. However, the included studies were not designed to formally test engagement as a mediator or moderator, and any association between engagement and effectiveness must therefore be interpreted cautiously.

Strengths and Limitations

This scoping review has several strengths. First, a major strength lies in its theoretically informed analytical approach. Engagement-related data were recoded a posteriori using a 3-component conceptualization of engagement (behavioral, cognitive, and affective), which served as an analytical framework to harmonize heterogeneous data without aiming to validate or test this conceptual framework. This approach is particularly relevant in a research field characterized by substantial conceptual and methodological variability. Second, the review explicitly incorporated a health equity perspective using the PROGRESS-Plus framework. Although such information was rarely reported in the included studies, its systematic consideration allowed for the identification of an

XSL•FO
RenderX

important gap in the literature and strengthened the scope of the evidence mapping.

Despite these strengths, several limitations should be acknowledged. First, the available evidence is constrained by small sample sizes, the underrepresentation of individuals living with major neurocognitive disorders, heterogeneous study designs, and the absence of standardized instruments specifically developed to measure engagement with digital health technologies. These factors limit the robustness and comparability of the findings. Moreover, engagement was most often inferred or reconstructed a posteriori rather than explicitly defined and systematically measured as a primary construct. Second, considerations related to health equity were largely absent from the included studies. Variables encompassed within the PROGRESS-Plus framework were infrequently reported, limiting the examination of how social, demographic, and contextual factors may influence engagement and restricting the generalizability of the findings to diverse populations of older adults. Third, the limited number of eligible studies and their predominantly exploratory nature reflect the early stage of development of this field of research. Although this limits the scope of conclusions that can be drawn, it also underscores the appropriateness of adopting a scoping review approach to map existing evidence, identify gaps, and inform future research directions.

Future Directions

The findings of this review have implications for both practice and research. The identified facilitating factors, particularly content personalization, professional supervision, and caregiver involvement, suggest that digital health technologies for older adults living with mild or major neurocognitive disorders should be developed and implemented in close alignment with users' cognitive capacities, care contexts, and available support networks.

The findings also highlight the need to develop validated, multidimensional engagement measurement tools tailored to specific contexts of use and to conduct larger-scale studies that integrate equity considerations.

Conclusion

This scoping review highlights a substantial gap between the theoretical recognition of engagement as a key determinant of the effectiveness of digital health technologies and its empirical operationalization among older adults living with neurocognitive disorders. It shows that engagement with digital health technologies in this population remains an emerging research area, characterized by significant methodological challenges and considerable potential for development.

There is a critical need to develop contextualized, rigorous, and population-sensitive engagement measurement tools as well as to ensure the systematic and explicit integration of the behavioral, cognitive, and affective dimensions of engagement. Addressing these priorities is essential to strengthen the design, evaluation, and implementation of digital health interventions that are more relevant and better tailored to supporting cognitive health among older adults living with neurocognitive disorders.

Acknowledgments

The authors would like to thank all individuals who were consulted or who participated in this study. They did not use any generative artificial intelligence tools at any stage of the preparation of this manuscript.

Funding

This project was funded by the Canadian Institutes of Health Research (Operating Grant: Knowledge Synthesis and Mobilization, Brain Health, and Reducing the Risk of Age-Related Neurocognitive Disorders).

Data Availability

Data related to the search strategy of this study are provided in Multimedia Appendix 1. All data generated or analyzed during this study (including the study flowchart) are included in Multimedia Appendices 1 and 2. Databases created during the data extraction process will be made available upon reasonable request.

Authors' Contributions

Funding, acquisition, and supervision: MPG
Design: SMARD, MS, MPG
Methodology: SMARD, FB, MPG
Study selection and data extraction: SMARD, SA, ARY, FN, FY, CC
Drafting of the manuscript: SMARD
Final drafting and editing of the manuscript: all authors

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.

XSL•FO
RenderX

[DOCX File, 33 KB - humanfactors_v13i1e70157_app1.docx]

Multimedia Appendix 2
Research strategy.
[DOCX File, 27 KB - humanfactors_v13i1e70157_app2.docx]

Checklist 1
PRISMA-ScR checklist.
[PDF File, 509 KB - humanfactors_v13i1e70157_app3.pdf]

References

1. Koh WQ, Whelan S, Heins P, Casey D, Toomey E, Dröes RM. The usability and impact of a low-cost pet robot for older adults and people with dementia: qualitative content analysis of user experiences and perceptions on consumer websites. JMIR Aging 2022 Feb 22;5(1):e29224. [doi: 10.2196/29224] [Medline: 35191844]
2. Ageing and health. World Health Organization. 2025 Oct 1. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2026-04-03]
3. Liberale L, Badimon L, Montecucco F, Lüscher TF, Libby P, Camici GG. Inflammation, aging, and cardiovascular disease: JACC review topic of the week. J Am Coll Cardiol 2022 Mar;79(8):837-847. [doi: 10.1016/j.jacc.2021.12.017] [Medline: 35210039]
4. Tsai YIP, Beh J, Ganderton C, Pranata A. Digital interventions for healthy ageing and cognitive health in older adults: a systematic review of mixed method studies and meta-analysis. BMC Geriatr 2024 Mar 4;24(1):217. [doi: 10.1186/s12877-023-04617-3] [Medline: 38438870]
5. Parcours de soins des patients présentant un trouble neurocognitif associée à la maladie d'Alzheimer ou à une maladie apparentéetroubles cognitifs et troubles neurocognitifs [Article in French]. : Haute Autorité de santé; 2018 URL: https://www.has-sante.fr/upload/docs/application/pdf/2018-05/fiche_1_troubles_cognitifs_et_trouble_neurocognitifs.pdf [accessed 2024-04-27]
6. Diagnostic and Statistical Manual of Mental Disorders: DSM-5: American Psychiatric Association; 2013. URL: https://psychiatryonline.org/doi/epub/10.1176/appi.books.9780890425596 [accessed 2026-04-03]
7. International Statistical Classification of Diseases and Related Health Problems (10th revision). : World Health Organization; 2019 URL: https://icd.who.int/browse10/Content/statichtml/ICD10Volume2_en_2019.pdf [accessed 2026-04-03]
8. Gauthier S, Besner J, Gélinas I, Morais JA, Rosa-Neto P, Webster C. Détection, diagnostic et prise en charge des troubles neurocognitifs majeurs: un guide pratique pour les professionnels de la santé [Report in French]. : Université McGill; 2025 URL: https://www.uspalz.com/sites/default/files/2025-03/Livret%20sur%20troubles%20neurocognitifs%20pour%20e%CC%81tudiants%20version%2021%20mars%202025.pdf [accessed 2026-04-03]
9. Que sont les troubles neurocognitifs? [Website in French]. Société Alzheimer du Canada. 2024. URL: https://alzheimer.ca/fr/au-sujet-des-troubles-neurocognitifs/que-sont-les-troubles-neurocognitifs [accessed 2026-04-03]
10. Effectifs et proportions des personnes de 65 ans et plus par région administrative [Website in French]. Institut de la statistique du Québec. 2024. URL: https://statistique.quebec.ca/vitrine/vieillissement/themes/population/65-ans-et-plus-region-administrative [accessed 2026-04-03]
11. Défis associés au vieillissement de la population [Article in French]. Institut National de Santé Publique du Québec. URL: https://www.inspq.qc.ca/sante-personnes-ainees/defis [accessed 2026-04-15]
12. Le quotidien [Website in French]. Statistique Canada. 2022. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/220427/dq220427a-fra.htm [accessed 2026-04-03]
13. Bai W, Chen P, Cai H, et al. Worldwide prevalence of mild cognitive impairment among community dwellers aged 50 years and older: a meta-analysis and systematic review of epidemiology studies. Age Ageing 2022 Aug 2;51(8):afac173. [doi: 10.1093/ageing/afac173] [Medline: 35977150]
14. Démence [Website in French]. Organisation mondiale de la Santé. 2025. URL: https://www.who.int/fr/news-room/fact-sheets/detail/dementia [accessed 2026-04-03]
15. Les chiffres sur les troubles neurocognitifs au canada [Website in French]. Société Alzheimer du Canada. 2024. URL: https://alzheimer.ca/fr/au-sujet-des-troubles-neurocognitifs/que-sont-les-troubles-neurocognitifs/chiffres [accessed 2026-04-03]
16. Maheu C. La santé cognitive des personnes aînées: pourquoi et comment la préserver [Website in French]. Institut national de santé publique du Québec. 2021. URL: https://www.inspq.qc.ca/publications/2786 [accessed 2026-04-03]
17. La santé cognitive [Report in French]. : Cisss du Bas-Saint-Laurent; 2025 URL: https://www.cisss-bsl.gouv.qc.ca/sites/default/files/document_dinformation_sante_cognitive.pdf [accessed 2026-04-03]
18. National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Board on Health Sciences Policy; Committee on Preventing Dementia and Cognitive Impairment. In: Downey A, Stroud C, Landis S, Leshner AI,

XSL•FO
RenderX

editors. Preventing Cognitive Decline and Dementia: A Way Forward: National Academies Press; 2017. URL: https://pubmed.ncbi.nlm.nih.gov/28650595/ [accessed 2026-04-03]
19. Piolatto M, Bianchi F, Rota M, Marengoni A, Akbaritabar A, Squazzoni F. The effect of social relationships on cognitive decline in older adults: an updated systematic review and meta-analysis of longitudinal cohort studies. BMC Public Health 2022 Feb 11;22(1):278. [doi: 10.1186/s12889-022-12567-5] [Medline: 35148704]
20. Risk reduction of cognitive decline and dementia: WHO guidelines. : World Health Organization; 2019 URL: https://iris.who.int/server/api/core/bitstreams/8fc586d7-b864-42eb-9274-95462496b328/content [accessed 2026-04-03]
21. Notice: Health Canada's approach to digital health technologies. Government of Canada. 2018. URL: https://www.canada.ca/en/health-canada/services/drugs-health-products/medical-devices/activities/announcements/notice-digital-health-technologies.html [accessed 2026-04-03]
22. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
23. Dequanter S, Gagnon MP, Ndiaye MA, et al. The effectiveness of e-health solutions for aging with cognitive impairment: a systematic review. Gerontologist 2021 Sep 13;61(7):e373-e394. [doi: 10.1093/geront/gnaa065] [Medline: 32525977]
24. Kelders SM, van Zyl LE, Ludden GDS. The concept and components of engagement in different domains applied to eHealth: a systematic scoping review. Front Psychol 2020;11:926. [doi: 10.3389/fpsyg.2020.00926] [Medline: 32536888]
25. Milne-Ives M, Homer S, Andrade J, Meinert E. The conceptualisation and measurement of engagement in digital health. Internet Interv 2024 Jun;36:100735. [doi: 10.1016/j.invent.2024.100735] [Medline: 38558760]
26. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
27. Short CE, DeSmet A, Woods C, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292. [doi: 10.2196/jmir.9397] [Medline: 30446482]
28. Cohen-Mansfield J, Dakheel-Ali M, Marx MS. Engagement in persons with dementia: the concept and its measurement. Am J Geriatr Psychiatry 2009 Apr;17(4):299-307. [doi: 10.1097/JGP.0b013e31818f3a52] [Medline: 19307858]
29. Pham Q, Graham G, Carrion C, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941. [doi: 10.2196/11941] [Medline: 30664463]
30. Kebede AS, Ozolins LL, Holst H, Galvin K. Digital engagement of older adults: scoping review. J Med Internet Res 2022 Dec 7;24(12):e40192. [doi: 10.2196/40192] [Medline: 36477006]
31. Kelders SM, Kip H. Development and initial validation of a scale to measure engagement with eHealth technologies. Presented at: CHI EA '19: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland, UK p. 1-6. [doi: 10.1145/3290607.3312917]
32. Woolley KE, Bright D, Ayres T, Morgan F, Little K, Davies AR. Mapping inequities in digital health technology within the World Health Organization's European region using PROGRESS PLUS: scoping review. J Med Internet Res 2023 Apr 28;25:e44181. [doi: 10.2196/44181] [Medline: 37115613]
33. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
34. Davis K, Drey N, Gould D. What are scoping studies? A review of the nursing literature. Int J Nurs Stud 2009 Oct;46(10):1386-1400. [doi: 10.1016/j.ijnurstu.2009.02.010] [Medline: 19328488]
35. Peters MDJ, Marnie C, Tricco AC, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020 Oct;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
36. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
37. Martinez YV, Renom-Guiteras A, Reeves D, et al. A set of systematic reviews to help reduce inappropriate prescribing to older people: study protocol. BMC Geriatr 2017 Oct 16;17(Suppl 1):231. [doi: 10.1186/s12877-017-0570-9] [Medline: 29047332]
38. Ditton A, Alodan H, Fox C, Evans S, Cross J. Exploring the effectiveness and experiences of people living with dementia interacting with digital interventions: a mixed methods systematic review. Dementia (London) 2025 Apr;24(3):506-551. [doi: 10.1177/14713012241302371] [Medline: 39604136]
39. Evans S, Harrison M, Morgan N, Kilkenny C, Morton T. Community makers: report on developing an online toolkit for supporting people with dementia to connect during the pandemic and beyond. WWOP 2022 Apr 28;26(2):140-150. [doi: 10.1108/WWOP-01-2022-0002]
40. Kaser AN, Mikula CM, Kiselica AM. Technology assistance in dementia (Tech-AiD): a framework for care in the digital age. J Health Serv Psychol 2024 Feb;50(1):37-46. [doi: 10.1007/s42843-024-00101-7] [Medline: 40018263]
41. Bramer WM. Variation in number of hits for complex searches in Google Scholar. J Med Libr Assoc 2016 Apr;104(2):143-145. [doi: 10.3163/1536-5050.104.2.009] [Medline: 27076802]

XSL•FO
RenderX

42. Covidence. University of Tennessee Health Science Center Library. 2024. URL: https://libguides.uthsc.edu/az/covidence [accessed 2026-04-21]
43. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
44. Karran EL, Cashin AG, Barker T, et al. Using PROGRESS-Plus to identify current approaches to the collection and reporting of equity-relevant data: a scoping review. J Clin Epidemiol 2023 Nov;163:70-78. [doi: 10.1016/j.jclinepi.2023.09.017] [Medline: 37802205]
45. Djabelkhir L, Wu YH, Vidal JS, et al. Computerized cognitive stimulation and engagement programs in older adults with mild cognitive impairment: comparing feasibility, acceptability, and cognitive and psychosocial effects. Clin Interv Aging 2017;12:1967-1975. [doi: 10.2147/CIA.S145769] [Medline: 29200836]
46. McAllister M, Dayton J, Oprescu F, Katsikitis M, Jones CM. Memory Keeper: a prototype digital application to improve engagement with people with dementia in long-term care (innovative practice). Dementia (London) 2020 May;19(4):1287-1298. [doi: 10.1177/1471301217737872] [Medline: 29096546]
47. Czaja SJ, Charness N, Fisk AD, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352. [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
48. Chan ATC, Ip RTF, Tran JYS, Chan JYC, Tsoi KKF. Computerized cognitive training for memory functions in mild cognitive impairment or dementia: a systematic review and meta-analysis. NPJ Digit Med 2024 Jan 3;7(1):1. [doi: 10.1038/s41746-023-00987-5] [Medline: 38172429]
49. Zhang H, Huntley J, Bhome R, et al. Effect of computerised cognitive training on cognitive outcomes in mild cognitive impairment: a systematic review and meta-analysis. BMJ Open 2019 Aug;9(8):e027062. [doi: 10.1136/bmjopen-2018-027062]

Abbreviations

CCE: computerized cognitive engagement
CCS: computerized cognitive stimulation
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
JBI: Joanna Briggs Institute
MCI: mild cognitive impairment
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
TMT: Trail Making Test

Edited by A Kushniruk, Y Gong; submitted 16.Dec.2024; peer-reviewed by A Ditton, B Ye, D Steijger, H Gains; revised version received 26.Feb.2026; accepted 07.Mar.2026; published 17.Jul.2026.

Please cite as:
Da SMAR, Sasseville M, Hardy MS, Beogo I, Gogovor A, Amil S, Yameogo AR, Naye F, Yousefi F, Bergeron F, Giguère A, LeBlanc A, Plaisimond J, Rivard-Lacroix C, Gagnon MP, Cachinho C
Engaging Older Adults With Neurocognitive Disorders in Digital Health Technologies: Scoping Review
JMIR Hum Factors 2026;13:e70157
URL: https://humanfactors.jmir.org/2026/1/e70157
doi:10.2196/70157

© Sié Mathieu Aymar Romaric Da, Maxime Sasseville, Marie-Soleil Hardy, Idrissa Beogo, Amédé Gogovor, Samira Amil, Achille R Yameogo, Florian Naye, Farzaneh Yousefi, Frédéric Bergeron, Anik Giguère, Annie LeBlanc, James Plaisimond, Carole Rivard-Lacroix, Marie-Pierre Gagnon, Chloé Cachinho. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 17.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Physical and Psychological Effects of Smartphone App–Based Walking Interventions in Community-Dwelling Older Adults: Systematic Review and Behavior Change Technique–Informed Analysis

Hiroki Abe[1*], MSN; Tomohiro Isinuki[1*], MSN; Rinako Ono[2*], BA; Naoki Kanaya[2*], BAgrSc; Masaya Tanno[1*], MD, PhD; Michiyo Hirano[1*], PhD

[1]Department of Nursing, School of Health Sciences, Sapporo Medical University, South 1 West 17, Chuo-ku, Sapporo, Japan

[2]Scholarly Communication Center Library, Sapporo Medical University, Sapporo, Japan

[*]all authors contributed equally

Corresponding Author:
Hiroki Abe, MSN
Department of Nursing, School of Health Sciences, Sapporo Medical University, South 1 West 17, Chuo-ku, Sapporo, Japan

Abstract

Background: With the global increase in population aging, promoting walking as a health behavior to maintain and enhance well-being among older adults has become increasingly important. In recent years, advances in information and communication technology and mobile health have supported the development of health interventions delivered through smartphone apps. However, no review to date has included psychological aspects such as motivation or intention to walk, and the behavior change techniques (BCTs) embedded in smartphone apps that effectively promote walking remain unclear.

Objective: This study had 2 primary aims: (1) to evaluate the effects of smartphone app use on walking behavior and intention to walk among community-dwelling older adults and (2) to identify the specific BCTs delivered through these apps that may promote walking behavior and intention effectively.

Methods: Eligible studies were those published in English or Japanese between March 1, 2015, and February 28, 2025, that focused on community-dwelling older adults, implemented smartphone app–based interventions, and reported walking-related outcomes. A systematic search strategy was designed using keywords such as “older,” “smartphone,” and “walking.” Risk of bias was evaluated using the Study Quality Assessment Tools. The features of the apps described in the selected studies were examined to identify the BCTs they used, as categorized by the BCT Taxonomy.

Results: Of the 296 studies initially retrieved, 8 met the inclusion criteria. These studies varied in terms of participant characteristics, intervention duration, app features, and outcomes, and most were pilot studies. While several apps were designed specifically to increase walking, others included features that facilitated social interaction among users. In addition, 2 studies also reported improved motivation to walk. Apps that were associated with statistically significant improvements in walking behavior frequently used BCTs from the following clusters: (1) goals and planning, (2) feedback and monitoring, and (4) shaping knowledge. Notably, 5 BCTs were not incorporated into any of the reviewed apps.

Conclusions: Although smartphone apps have the potential to improve walking behavior and intention among community-dwelling older adults, the current body of evidence remains limited. Apps that deliver walking-related knowledge, facilitate goal setting, and support behavioral monitoring appear especially effective and may strengthen walking behavior and intention in this population.

Trial Registration: PROSPERO CRD420251007296; https://www.crd.york.ac.uk/PROSPERO/view/CRD420251007296

(JMIR Hum Factors 2026;13:e78042) doi:10.2196/78042

KEYWORDS

ICT; information and communication technology; smartphone; apps; walking; BCT; behavior change technique; older adults

Introduction

In 2024, Japan officially became a super-aged nation, leading a global demographic shift [1,2]. In aging societies such as ours, it is imperative to maintain and improve older adults’ health, especially given the growing burden of chronic diseases and the resulting pressure on health care systems [3]. Walking has been shown to reduce mortality among older adults [4-6] and is consequently recommended as a key health behavior [7,8]. Moreover, walking is one of the most widely adopted and

XSL•FO
RenderX

preferred forms of physical activity among older adults [9]. Accordingly, it is particularly important to identify and systematize interventions that promote walking as a strategy for maintaining and enhancing health.

Techniques used to support behavior change have been systematically classified in the Behavior Change Technique (BCT) Taxonomy, which provides a useful framework for analyzing and developing health interventions [10]. Details regarding the BCT Taxonomy, comprising 93 individual techniques organized into 16 clusters, are provided in Multimedia Appendix 1 [11]. Previous studies have demonstrated the effectiveness of BCT-based interventions in promoting physical activity among older adults [12]. Moreover, interventions targeting older populations have already been categorized using the BCT Taxonomy, emphasizing its use in designing and implementing behavioral programs [13]. However, to the best of our knowledge, the specific BCTs that effectively promote walking have not yet been systematically investigated. Therefore, analyzing existing intervention studies on walking promotion through the lens of the BCT Taxonomy and systematically extracting effective intervention components as BCTs is warranted.

In recent years, an increasing number of intervention studies have used smartphone apps in studies involving older adults within the framework of mobile health (mHealth) [14]. Previous reviews have focused on outcomes such as sedentary behavior [14] and physical performance [15], yielding comprehensive insights into specific mHealth features effective in promoting walking among older adults [16]. However, few studies have reported on the psychological effects of mHealth use in this population [17], and to the best of our knowledge, no systematic review has addressed psychological outcomes such as the intention to walk. Existing studies have also underscored the need to address this limitation [16]. Furthermore, current research remains insufficient in elucidating the psychological changes and underlying mechanisms associated with interventions targeting older adults [18,19]. Therefore, it is essential to comprehensively organize both the physical and psychological outcomes related to walking and to determine which smartphone app-based interventions effectively enhance walking behavior and intention among older adults. Moreover, the rapid advancement of information and communication technology (ICT), including the integration of large language models and generative artificial intelligence in health care [20], necessitates continuous knowledge updating and reorganization to maintain relevance and applicability.

Building on the above, the aim of this study was 2-fold: first, to clarify the effects of smartphone app use on walking behavior and intention to walk among community-dwelling older adults; and second, to identify the BCTs delivered through these apps that may effectively promote walking behavior and intention. This study addressed the following research questions:

1. Does the use of smartphone apps improve walking behavior and intention to walk among older adults?
2. Which BCTs, incorporated into smartphone apps, are effective in promoting walking among older adults?

Methods

Research Design

This study used a systematic review approach to investigate whether the use of smartphone apps promotes walking behavior among community-dwelling older adults. The review protocol was registered with PROSPERO (International Prospective Register of Systematic Reviews; CRD420251007296) before the commencement of the database search. The PRISMA checklist is available in Checklist 1.

Search Strategy and Inclusion Criteria

Keywords were identified based on the review questions, and a systematic search was conducted across 5 databases in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. The databases were selected for the following reasons: PubMed, a leading database in medical and health sciences, provides extensive coverage and supports searches using Medical Subject Headings. CINAHL (via EBSCOhost) was included due to its specialization in nursing and allied health literature, which is often underrepresented in traditional medical databases. The Cochrane Database of Systematic Reviews and the Cochrane Central Register of Controlled Trials, both accessed via Wiley Online Library, were selected for their focus on systematic reviews and clinical trials. Finally, Ichushi-Web was included for its comprehensive coverage of Japanese-language literature in medicine, nursing, and health sciences.

The search strategy was developed using a combination of Medical Subject Headings and relevant free-text terms. To reflect the rapidly evolving landscape of eHealth and mHealth research, priority was given to studies using smartphone apps. Consistency in search procedures was ensured to accurately capture studies targeting community-dwelling older adults. To enhance the comprehensiveness of the review, no restrictions were applied to study design during the initial screening phase. The search was conducted on March 19, 2025, and the Boolean operators used are summarized in Table 1. The complete list of search terms is provided in Multimedia Appendix 2. Studies were included if they met the following eligibility criteria:

- Studies were peer-reviewed original research articles.
- Participants were older adults. Given that the definition of older adults varies across countries and studies, older adults were defined as individuals aged 65 years or older, or those explicitly referred to as older adults in the original study, to accommodate regional variation (eg, adults aged ≥60 years in China).
- Studies targeting institutionalized or hospitalized older adults were excluded.
- The intervention used smartphone apps to promote walking. This included apps designed specifically to increase walking, as well as apps with other primary purposes that incorporated walking-promotion features.
- No restrictions were placed on the duration of the intervention.
- Studies in which participants used the app in their daily lives were included, whereas studies conducted in controlled

XSL•FO
RenderX

laboratory settings, such as step counter accuracy tests, were excluded.
- Studies were required to report changes in walking-related indicators such as step count, walking frequency, walking distance, walking awareness, or achievement of step-based goals. Studies focusing exclusively on physical function tests, such as the Timed Up and Go Test or the 5-meter walk test, were excluded, as this review emphasized habitual walking behavior and associated intentions.
- Eligible study designs included experimental studies, quasi-experimental studies, and randomized controlled trials (RCTs).
- Studies published in English or Japanese between March 1, 2015, and February 28, 2025, were included.

XSL•FO
RenderX

Table . Search strings used in academic databases.

Database	Search string
PubMed	(“aged”[MeSH[a] Terms] OR “elderly”[Title/Abstract] OR “older adult*“[Title/Abstract] OR “older people”[Title/Abstract] OR “older person*“[Title/Abstract] OR “older men”[Title/Abstract] OR “older women”[Title/Abstract] OR “old age”[Title/Abstract] OR “senior citizen*“[Title/Abstract]) AND (“independent living”[MeSH Terms] OR “independent living”[Title/Abstract] OR “community dwell*“[Title/Abstract] OR “community living”[Title/Abstract] OR “community setting”[Title/Abstract] OR “aging in place”[Title/Abstract] OR “aging in place”[Title/Abstract] OR “age in place”[Title/Abstract] OR “non-institutionalized”[Title/Abstract] OR “living at home”[Title/Abstract] OR “healthy”[Title/Abstract] OR “rural”[Title/Abstract]) AND (“smartphone”[MeSH Terms] OR “smartphone*“[Title/Abstract]) AND (“mobile applications”[MeSH Terms] OR “application*“[Title/Abstract] OR “app”[Title/Abstract] OR “mHealth”[Title/Abstract] OR “video games”[MeSH Terms] OR “video game*“[Title/Abstract]) AND (“pedometer”[Title/Abstract] OR “accelerometer”[Title/Abstract] OR “step*“[Title/Abstract] OR “stroll”[Title/Abstract] OR “activ*“[Title/Abstract] OR “walking”[MeSH Terms] OR “walk*“[Title/Abstract] OR “gait”[Title/Abstract] OR “gait analysis”[MeSH Terms] OR “exercise”[MeSH Terms] OR “exerci*"[Title/Abstract])
CINAHL	((MH[b] aged+) OR (TI[c] elderly OR AB[d] elderly) OR (TI “older adult*” OR AB “older adult*“) OR (TI “older people” OR AB “older people”) OR (TI “older person*” OR AB “older person*“) OR (TI “older men” OR AB “older men”) OR (TI “older women” OR AB “older women”) OR (TI “old age” OR AB “old age”) OR (TI “senior citizen*” OR AB “senior citizen*“)) AND ((MH “independent living+“) OR (TI “independent living” OR AB “independent living”) OR (TI “community dwell*” OR AB “community dwell*“) OR (TI “community living” OR AB “community living”) OR (TI “community setting” OR AB “community setting”) OR (TI “aging in place” OR AB “aging in place”) OR (TI “aging in place” OR AB “aging in place”) OR (TI “age in place” OR AB “age in place”) OR (TI non-institutionalized OR AB non-institutionalized) OR (TI “living at home” OR AB “living at home”) OR (TI healthy OR AB healthy) OR (TI rural OR AB rural)) AND ((MH smartphone+) OR (TI smartphone* OR AB smartphone*)) AND ((MH “mobile applications+“) OR (TI application* OR AB application*) OR (TI app OR AB app) OR (TI mHealth OR AB mHealth) OR (MH “video games+“) OR (TI “video game*” OR AB “video game*”)) AND ((TI pedometer OR AB pedometer) OR (TI accelerometer OR AB accelerometer) OR (TI step* OR AB step*) OR (TI stroll OR AB stroll) OR (TI activ* OR AB activ*) OR (MH walking+) OR (TI walk* OR AB walk*) OR (TI Gait OR AB Gait) OR (MH “gait analysis+") OR (MH exercise+) OR (TI exerci* OR AB exerci*))
Cochrane Database of Systematic Reviews	(“aged”[MeSH] OR “elderly”[Title/Abstract] OR “older NEXT adult*“[Title/Abstract] OR “older people”[Title/Abstract] OR “older NEXT person*“[Title/Abstract] OR “older men”[Title/Abstract] OR “older women”[Title/Abstract] OR “old age”[Title/Abstract] OR " senior NEXT citizen*“[Title/Abstract]) AND (“independent living”[MeSH] OR “independent living”[Title/Abstract] OR " community NEXT dwell*“[Title/Abstract] OR “community living”[Title/Abstract] OR “community setting”[Title/Abstract] OR “aging in place”[Title/Abstract] OR “aging in place”[Title/Abstract] OR “age in place”[Title/Abstract] OR “non-institutionalized”[Title/Abstract] OR “living at home”[Title/Abstract] OR “healthy”[Title/Abstract] OR “rural”[Title/Abstract]) AND (“smartphone”[MeSH] OR “smartphone*“[Title/Abstract]) AND (“mobile applications”[MeSH] OR “application*“[Title/Abstract] OR “app”[Title/Abstract] OR “mHealth”[Title/Abstract] OR “video games”[MeSH] OR " video NEXT game*“[Title/Abstract]) AND (“pedometer”[Title/Abstract] OR “accelerometer”[Title/Abstract] OR “step*“[Title/Abstract] OR “stroll”[Title/Abstract] OR “activ*“[Title/Abstract] OR “walking”[MeSH] OR “walk*“[Title/Abstract] OR “Gait”[Title/Abstract] OR “gait analysis”[MeSH] OR “exercise”[MeSH] OR “exerci*"[Title/Abstract])

XSL•FO
RenderX

Database	Search string
Cochrane Central Register of Controlled Trials	(“aged”[MeSH] OR “elderly”[Title/Abstract] OR “older NEXT adult*“[Title/Abstract] OR “older people”[Title/Abstract] OR “older NEXT person*“[Title/Abstract] OR “older men”[Title/Abstract] OR “older women”[Title/Abstract] OR “old age”[Title/Abstract] OR " senior NEXT citizen*“[Title/Abstract]) AND (“independent living”[MeSH] OR “independent living”[Title/Abstract] OR " community NEXT dwell*“[Title/Abstract] OR “community living”[Title/Abstract] OR “community setting”[Title/Abstract] OR “aging in place”[Title/Abstract] OR “aging in place”[Title/Abstract] OR “age in place”[Title/Abstract] OR “non-institutionalized”[Title/Abstract] OR “living at home”[Title/Abstract] OR “healthy”[Title/Abstract] OR “rural”[Title/Abstract]) AND (Keywords: “smartphone”[MeSH] OR “smartphone*“[Title/Abstract]) AND (“mobile applications”[MeSH] OR “application*“[Title/Abstract] OR “app”[Title/Abstract] OR “mHealth”[Title/Abstract] OR “video games”[MeSH] OR " video NEXT game*“[Title/Abstract]) AND (“pedometer”[Title/Abstract] OR “accelerometer”[Title/Abstract] OR “step*“[Title/Abstract] OR “stroll”[Title/Abstract] OR “activ*“[Title/Abstract] OR “walking”[MeSH] OR “walk*“[Title/Abstract] OR “Gait”[Title/Abstract] OR “gait analysis”[MeSH] OR “exercise”[MeSH] OR “exerci*"[Title/Abstract])
医中誌Web	(高齢者/TH or 高齢者評価/TH or 高齢/TA or 老年/TA or 老齢/TA or 老人/TA) and (自立生活/TH or 自立/TA or 在住/TA or 居住/TA or 住民/TA or 在宅/TA or 健常/TA or 独居/TA or 地域高齢者/TA) and (スマートフォン/TH or スマートフォン/TA or 携帯電話/TH or 携帯電話/TA) and モバイルアプリケーション/TH or アプリ/TA or (@[モバイルアプリケーション]/TH and @[スマートフォン]/TH) and (歩行/TH or 歩行運動/TH or 歩行分析/TH or 歩数/TA or 歩行/TA or 散歩/TA or 万歩計/TA or “歩数計"/TH)

[a]MeSH: Medical Subject Headings.

[b]MH: controlled terms.

[c]TI: title.

[d]AB: abstract.

Study Selection

Prior to the study selection, all authors discussed the inclusion and exclusion criteria based on the Population, Intervention, Comparison, and Outcome framework and reached a consensus. Study selection was conducted in 2 stages by three authors (HA, TI, and MH). The initial screening was conducted using Rayyan. Two authors (HA and TI) independently assessed the titles and abstracts according to predefined eligibility criteria. When a disagreement occurred, a third author (MH) served as the adjudicator. During the second stage, the same two authors independently reviewed the full texts of studies identified as eligible, and any conflicts were resolved through discussion with the third author (MH). To ensure consistency in study selection and data extraction, interrater agreement at each stage was assessed using Cohen kappa (κ) statistic [21].

Data Extraction

The following information was extracted from each of the included study: study identifiers (eg, author names and publication y), sample characteristics (eg, participant demographics, sample size, and health status), intervention duration, smartphone app details, and reported outcomes. All data were systematically organized and presented in tabular form.

BCT Classification

First, the lead author conducted an in-depth review of all 8 studies and generated a list of app functions and components. The second author subsequently cross-checked this list against the original articles to ensure accuracy. Finally, all authors reviewed the app functions and purposes described in the eligible studies and classified the corresponding BCTs accordingly. Classification was conducted using the BCT Taxonomy v1, which organizes 93 techniques into 16 clusters [11]. To visualize the relationship between the included studies and the identified BCTs, we created charts at the cluster level. In addition, to closely examine the mechanisms of behavior change, we mapped individual app functions to specific BCTs whenever possible. Any discrepancies were resolved through discussion until full consensus was reached.

Risk of Bias Assessment

To evaluate the risk of bias in the included studies, the National Institutes of Health Study Quality Assessment Tool was used. This tool is designed to assess the quality of observational cohort and cross-sectional studies involving human participants [22]. The tool comprises either 12 or 14 items, depending on the study design. The proportion of items rated “yes” was calculated for each study to determine overall methodological quality and potential risk of bias. Based on these percentages, categorical quality ratings were defined as follows: high (>70%), moderate (70% - 50%), and low (≤49%).

XSL•FO
RenderX

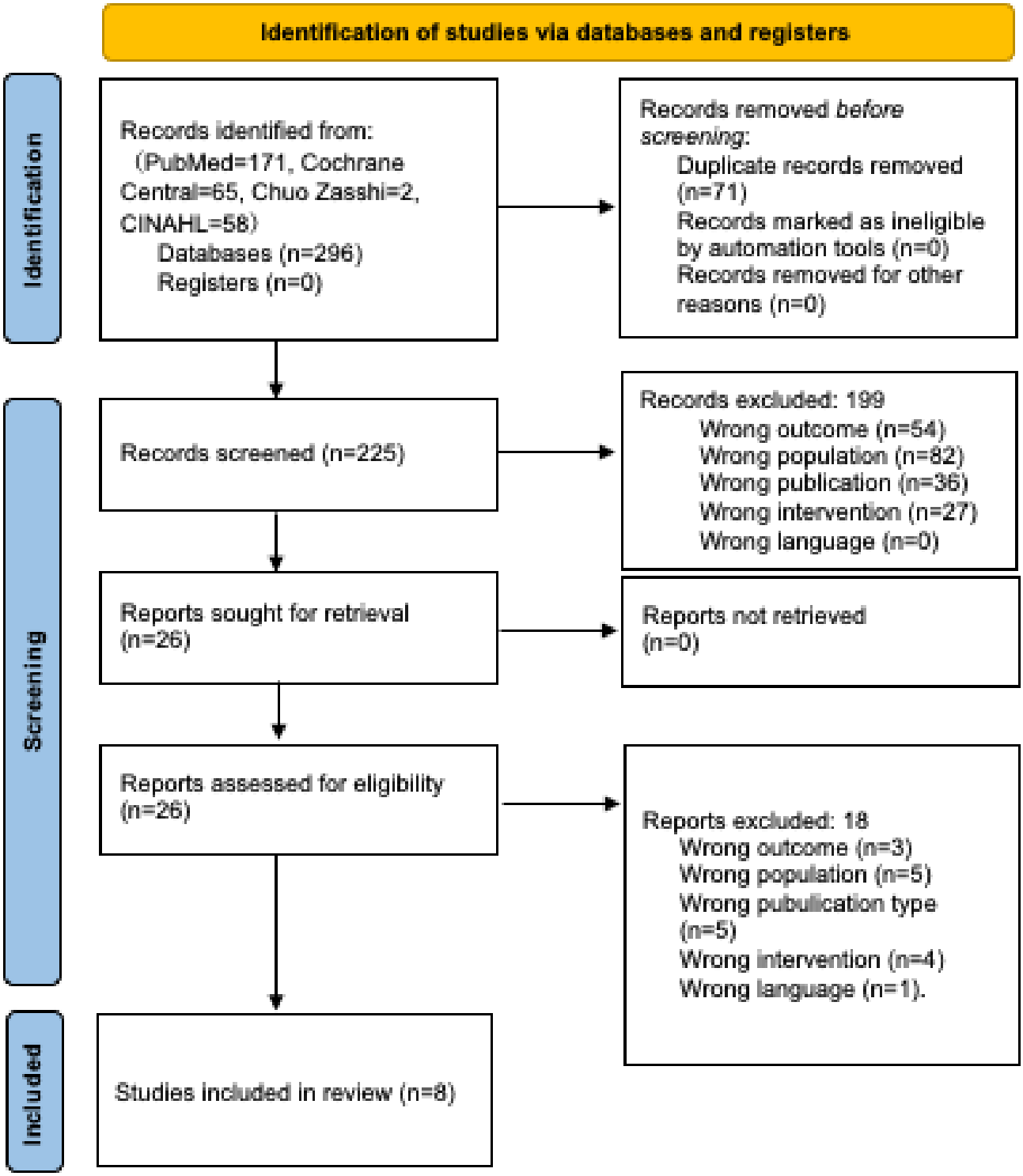
Results

Overview

Figure 1 illustrates the screening and selection process for this systematic review. A total of 296 records were identified through 4 database searches: PubMed (n=171), Cochrane Central (n=65), Igaku Chuo Zasshi (n=2), and CINAHL (n=58). A total of 71 duplicate records were collected. During the first screening of titles and abstracts, 4 conflicts between the authors were noted and resolved, and 199 articles were excluded, based on the inclusion criteria. The reasons for exclusion were as follows: incorrect outcomes (n=54), ineligible population (n=82), nonqualifying publication type (n=36), and irrelevant intervention (n=27). In the second screening stage, the full texts of 26 articles were reviewed without any reviewer conflict. Of these, 18 were excluded for the following reasons: incorrect outcome (n=3), ineligible population (n=5), nonqualifying publication type (n=5), irrelevant intervention (n=4), and non-English/Japanese language (n=1), resulting in 8 studies that met all inclusion criteria and were included in the final analysis. During the second screening stage, studies were excluded if the intervention involved smartphone apps for older adults but lacked walking-related outcome measures. For example, studies that assessed broader indicators such as overall physical activity [23] were excluded as having the wrong outcome. An interrater reliability analysis showed a high level of agreement between the authors. During the initial screening, the interrater agreement rate was 98.2%, with a Cohen kappa (κ) of 0.925, indicating almost perfect agreement. In the second screening, agreement reached 100%, with a κ value of 1.0, reflecting perfect concordance.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart reporting the literature search strategy.



Study Characteristics

Table 2 presents an overview of the intervention study outcomes for smartphone apps promoting walking among community-dwelling older adults, along with their associated BCTs. The mean age (SD) of participants in the included studies ranged from 59.8 (6.4) to 78.0 (9.0) years. All participants were ambulatory, although several studies included older adults diagnosed with frailty, prefrailty, or heart failure. The duration of the interventions ranged from 2 weeks (excluding the empirical evaluation study by Zhong and Rau [24]) to 1 year. In terms of study design, none of the included studies used a full RCT. The studies comprised 4 controlled intervention studies, 2 observational cohort or cross-sectional studies, and 2 uncontrolled pre-post studies. The results of the risk-of-bias assessment for each study are presented in Multimedia Appendix 3.

XSL•FO
RenderX

Table . Overview of intervention study outcomes on smartphone apps promoting walking among community-dwelling older adults and their associated BCT[a].

Authors	Population and term	Used smartphone app	Findings	Types of BCT[b]	QA[c]

Fong et al [25] (2016)	• Older adults capable of walking
	  • IG[d]: 97 participants (mean age 60.1, SD 5.5 y)
	  • CG[e]: 54 participants (mean age 65.3, SD 8.7 y)
	• 2 wk
		Pedometer Lite, Pedometer++, and WalkLogger are existing step-counting apps that enable users to track and manage their daily steps	Willingness to walk (qualitative results)	2	4/14 low

Paul et al [26] (2017)	• Healthy older adults: 16 participants (mean 71.1, SD 5.2 y) • 6 wk	STARFISH is a smartphone app that promotes physical activity among older adults by tracking steps and providing animated feedback. Users set goals and receive weekly updates. Used in groups of 4, the app fosters motivation through shared progress and group-based rewards	Improving the number of steps	1, 2, 3, 6, 8, 10, and 12	8/12 middle
Matsuoka et al [27] (2024)	• Older adults shopping at AEON Malls in Japan (mean 69.9, SD 4.5 y) • 1 y	The AEON Mall app, developed for customers of the AEON supermarket chain in Japan, uses GPS to track when users are inside a mall and records their daily step counts. If users exceed 1000 steps in a day, they receive a digital coupon for a lottery to win up to 500 yen (US $3.23) in points. The app encourages physical activity among the general public and may help restore decreased step counts following the COVID-19 pandemic, offering potential benefits for public health.	Improving the number of steps (statistical analysis was not conducted)	2 and 10	10/14 high

XSL•FO
RenderX

Authors	Population and term	Used smartphone app	Findings	Types of BCT[b]	QA[c]

Chan et al [28] (2022)	• Individuals capable of using a smartphone and walking independently for at least 15 min
	  • IG: 70 participants (mean age 59.8, SD 6.4 y)
	  • CG: 69 participants (mean age 59.8, SD 6.8 y)
	• 12 wk
		ZTExApp is a smartphone app developed to promote physical activity in patients with coronary heart disease by encouraging simple, equipment-free exercises that can be integrated into daily routines. The app offers 4 key features: educational content on ZTEx and its benefits, exercise examples based on 6 fitness components, self-assessment tools like a 30-sec chair stand test, and an electronic diary for goal setting and activity tracking	Improving walking time per week (P<.05)	1, 2, and 4	10/14
					high

Kawaguchi et al [29] (2024)	• This study involved community-dwelling older adults who used smartphones and LINE, spoke Japanese, and were not certified under Japan's LTCI[f] system. Participants were divided into an IG (n=85, mean age 69.9) and a CG (n=94, mean age 70.1) • 12 wk	The ESP[g] app is designed to promote social participation among older adults by tracking outing frequency, locations visited, walking distance, and step count using GPS and pedometer data. It classifies users into 4 participation levels—Beginner, Skilled, Master, and Expert—and displays them in a ranking format. The app also provides weekly articles on the importance of social participation. These metrics serve as indicators of both social engagement and daily physical activity.	Improving the number of steps	2, 3, 4, 5, 7, 8, and 10	12/14 high
Ohta et al, [30] (2024)	• Thirty older adults 65 y and above who were classified as frail or prefrail • 6 wk		Improving the number of steps (P<.05)	1, 2, 3, 4, 6, 8, and 10	8/12 high

XSL•FO
RenderX

Authors	Population and term	Used smartphone app	Findings	Types of BCT[b]	QA[c]
		Online Kayoinoba is a care prevention app designed to promote physical activity and social interaction among older adults. It features goal setting based on steps and exercise time, municipality-specific exercise videos, meal logging with nutritional balance visualization, social functions like chat and photo sharing, cognitive games to prevent decline, and health check tools. A point system rewards app use, with user activity rankings displayed at national, prefectural, municipal, and community levels.			
Zhong et al [24] (2020)	• Independent, ambulatory older adults (mean age 69.8, SD 7.0 y) • Each participant performed 3 types of 40-m indoor walking tests (ST[h], DT[i], and FW[j]) only once. Each walking trial was preceded by a 30-second intervention.	Pocket Gait is a smartphone app developed by the authors to evaluate and manage walking quality in the daily lives of older adults. It is designed as a tool to support the early detection of gait abnormalities and fall risk. The app uses the smartphone's built-in accelerometer to measure 5 gait parameters: step frequency, acceleration RMS[k], gait regularity, symmetry, and variability. Walking data are collected in real time and presented graphically for user interpretation.	Willingness to walk (qualitative results)	2	6/14 middle

Blomqvist et al [31] (2025)	• Physically inactive adults diagnosed with heart failure and equipped with a home-based self-care support device called Optilogg.
	  • IG: 10 participants (mean age 78, SD 9 y)
	  • CG: 10 participants (mean age 77, SD 5 y)
	• 12 wk
			Improving the number of steps	1, 2, 1, 5, 7, and 15	6/14 middle

XSL•FO
RenderX

Authors	Population and term	Used smartphone app	Findings	Types of BCT[b]	QA[c]
		Activity Coach is an mHealth app developed to promote physical activity and reduce sedentary behavior in patients with heart failure. As an extension of the Optilogg self-management platform, it offers features like activity logging in 10-minute intervals, weekly goal setting, educational messages, and visual feedback. Users can define and track their own physical activities, such as walking or housework. The app provides daily educational messages during the first week to build beliefs in the benefits of activity, boost self-efficacy, and emphasize personal goal importance. Progress is displayed visually to support ongoing motivation and behavior change.			

[a]BCT: behavior change technique.

[b]The numbers of the types of BCT refer to the following meanings: 1=goals and planning; 2=feedback and monitoring; 3=social support; 4=shaping knowledge; 5=natural consequences; 6=comparison of behavior; 7=associations; 8=repetition and substitution; 9=comparison of outcomes; 10=reward and threat; 11=regulation; 12=antecedents; 13=identity; 14=scheduled consequences; 15=self-belief; and 16=covert learning.

[c]QA: quality assessment.

[d]IG: intervention group.

[e]CG: control group.

[f]LTCI: long-term care insurance.

[g]ESP: encouragement of social participation.

[h]ST: single task.

[i]DT: dual task.

[j]FW: fast walk.

[k]RMS: root mean square.

Among the 6 studies that reported physical outcomes [26-31], 2 demonstrated statistically significant effects of app use on walking behavior [28,30]. Four studies included comparison groups; however, none used a full RCT design. Among the 8 studies, 4 were identified as pilot studies [26,28,30,31] and only 2 of these [28,31] reported predetermined sample sizes sufficient for statistical analysis. Among the 6 studies reporting physical outcomes, 4 conducted statistical analyses [27-29,31]; however, 2 of these did not reach their target sample sizes [29,31].

According to the risk-of-bias assessment, 4 studies were rated as high quality [27-30], 3 as medium quality [24,26,31], and 1 as low quality [25].

App Features

Of the 8 studies, 1 used a commercially available app, while the remaining 7 used apps developed by the researchers. Six of the 8 apps were specifically designed to promote physical activity by monitoring step counts [24-28,31]. In contrast, 2 apps [29,30] were designed to promote not only physical activity but also social engagement. Six apps were intended for individual use [25-29,31,32], while 2 apps [26,30] incorporated group-based features that allowed users to interact on a shared platform.

Basic step-counting apps such as Pedometer Lite, Pedometer++, and WalkLogger [25] enable users to record and manage daily step counts. STARFISH [26] promotes increased physical activity among older adults by combining step tracking with animated feedback and group-oriented goal setting and reward mechanisms. The AEON Mall app [27] uses GPS to detect user visits to shopping malls and provides digital coupons when users exceed 1000 steps per day, thereby contributing to public

XSL•FO
RenderX

health initiatives aimed at restoring physical activity levels that declined following the COVID-19 pandemic.

ZTExApp [28], designed for patients with coronary heart disease, promotes simple, equipment-free exercises and provides educational content, self-assessment tools, and an activity diary. The ESP app [29], aimed at supporting social participation among older adults, records users' outings, step counts, and visited locations, classifies them into 4 participation levels, and provides weekly educational articles. Online Kayoinoba [30] integrates step goals, exercise videos, dietary tracking, social features, cognitive games, and health checks and uses a point-based system with multilevel rankings to enhance user engagement.

Among the more advanced monitoring tools, Pocket Gait [24] uses smartphone accelerometers to measure walking frequency, root mean square acceleration, regularity, symmetry, and variability, thereby supporting early detection of gait impairments. Activity Coach [31], developed for patients with heart failure, includes activity logs, weekly goal setting, educational messages, and visual feedback to reduce sedentary behavior and strengthen self-efficacy for maintaining physical activity.

The Effects of Walking Behavior and Intention to Walk by Using a Smartphone App Among Older Adults

Among the included studies, 5 assessed step count as a walking-related outcome, while 1 evaluated weekly walking time. Of these, 1 study reported a statistically significant increase in step count, and another observed a significant improvement in weekly walking time. Outcomes related to intention to walk were reported narratively in 2 studies, both of which indicated positive effects on walking motivation.

BCT Classification

Table 3 presents the relationship between the apps used in the reviewed studies and the corresponding BCT clusters. Based on the functional features of the apps, the number of BCTs identified per app ranged from 2 to 13 of the 93 techniques defined in the BCT Taxonomy. The smartphone apps incorporated BCTs from 1 to 7 of the 16 clusters defined in the BCT Taxonomy. All apps used techniques from cluster 2: feedback and monitoring. Improvements in walking motivation were also associated with BCTs from cluster 2: feedback and monitoring. Additionally, 4 apps included techniques from clusters 1: goals and planning, and cluster 10: reward and threat. Apps that demonstrated statistically significant effects on walking behavior consistently used techniques from clusters 1, 2, and 4: shaping knowledge. Conversely, 5 BCT clusters were not used in any of the reviewed apps: cluster 9: comparison of outcomes, 11: regulation, 13: identity, 14: scheduled consequences, and 16: covert learning.

XSL•FO
RenderX

Table . Relationship between apps used by research studies and BCT[a] clusters[b].

	Goals and planning	Feedback and monitoring	Social support	Shaping knowledge	Natural consequences	Comparison of behavior	Associations	Repetition and substitution	Comparison of outcomes	Reward and threat	Regulation	Antecedents	Identity	Scheduled consequences	Self-belief	Covert learning
Fong et al [25] (2016)	—[c]	□[d]	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Paul et al [26] (2017)	○[e]	○[e]	○[e]	—	—	○[e]	—	—	—	○[e]	—	○[e]	—	—	—	—
Zhong et al [24] (2020)	—	□[d]	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Chan et al [28] (2022)	●[f]	●[f]	—	●[f]	—	—	—	—	—	—	—	—	—	—	—	—
Matsuoka et al [27] (2024)	—	○[e]	—	—	—	—	—	—	—	○[e]	—	—	—	—	—	—
Ohta et al [30] (2024)	●[f]	●[f]	●[f]	●[f]	—	●[f]	—	●[f]	—	●[f]	—	—	—	—	—	—
Kawaguchi et al [29] (2024)	—	○[e]	○[e]	○[e]	○[e]	—	○[e]	○[e]	—	○[e]	—	—	—	—	—	—
Blomqvist et al [31] (2025)	○[e]	○[e]	—	—	○[e]	—	○[e]	—	—	—	—	—	—	—	○[e]	—
Total number of BCT types[g]	4	8	3	3	2	2	2	2	0	4	0	1	0	0	1	0

[a]BCT: behavior change technique.

[b]The above table shows the relationship between each study and the BCT clusters.

[c]Not applicable.

[d]The study demonstrated qualitative effects.

[e]The effect was not statistically significant.

[f]A statistically significant effect.

[g]The last row shows the total number of occurrences in each cluster.

Discussion

Overview

This systematic review investigated the effectiveness of smartphone apps on walking behavior and the intention to walk among community-dwelling older adults, based on 8 existing studies. Although the number of included studies was smaller than in comparable reviews (eg, Llopis et al [16]), this review distinctively focused on community-dwelling older adults, excluding those residing in institutions or hospitals, and explicitly evaluated both walking behavior and walking intention. Furthermore, by classifying the BCTs embedded in app functions, the review identified app-based intervention components that may effectively promote walking.

XSL•FO
RenderX

Does the Use of Smartphone Apps Improve Walking Behavior and Intention to Walk Among Older Adults?

The use of smartphone apps demonstrated positive effects on both walking behavior and walking intention [24,25,28,30]. Among the included studies, 6 measured physical outcomes (eg, step count or walking duration), while 2 assessed psychological outcomes such as walking intention. Although both studies reporting psychological outcomes relied on narrative data, indicating the need for further quantitative examination, the findings suggest that smartphone apps may have the potential to improve both physical and psychological outcomes related to walking.

The 2 studies that assessed psychological outcomes used apps that included step-monitoring functions. Previous research has demonstrated that activity tracker use can effectively increase step counts [33], suggesting that enabling older adults to self-monitor their walking behavior may also exert a positive influence on their walking intention.

Most smartphone app-based intervention studies targeting older adults have been pilot studies [34], a trend that is consistent with observations reported in previous systematic reviews [16]. Given this trend and the limited number of eligible studies, the conclusions of this review should be interpreted with caution. Nevertheless, some studies reported statistically significant improvements in walking behavior among community-dwelling older adults following app use, and narrative findings suggested psychological benefits as well [24,25].

Which BCTs Incorporated Into Smartphone Apps Are Effective in Promoting Walking Among Older Adults?

Among the studies included in this review, 2 smartphone app–based interventions [28,30] demonstrated statistically significant improvements in walking-related outcomes. Both of these apps implemented the following BCT clusters: (1) goal setting and planning, (2) feedback and monitoring, and (3) shaping knowledge. This overlap holds theoretical significance for the design of smartphone apps aimed at promoting walking among older adults. Gilchrist et al [13] demonstrated that interpersonal physical activity programs incorporating goal setting and self-monitoring effectively enhance physical activity levels, which aligns with the findings of this review.

Although the BCT Taxonomy provides a comprehensive framework for behavior change, this review suggests that, in the context of walking promotion among older adults, providing relevant knowledge, facilitating goal setting and planning, and enabling behavior monitoring may be especially effective. It is also important that apps designed for older adults maintain simplicity and minimize cognitive burden [35]. Because the number of implemented BCTs is not necessarily associated with increased physical activity [36], overly complex or multifunctional apps may impose unnecessary burden and hinder sustained use. Therefore, designing apps that focus on the specific BCTs shown to be effective is likely to lead to more desirable outcomes.

According to the BCT Taxonomy, all apps reviewed in this study incorporated cluster 2: feedback and monitoring, and positive effects were observed on both walking behavior and walking intention. These apps provided feedback related to walking behavior, such as step counts and activity levels. Smartphones are highly portable and are well-suited for real-time monitoring of physical activity [37], and health platforms such as Apple's HealthKit [38] and Android's Google Fit [39] allow developers to efficiently integrate feedback and monitoring functions via available application programming interfaces [40]. This not only facilitates the implementation of BCT-aligned features but also reduces development costs, offering advantages from both behavioral science and engineering perspectives.

In contrast, 5 of the 16 BCT clusters were not implemented in any of the reviewed apps. The absence of these techniques may reflect technological constraints inherent in app-based interventions. For instance, BCT 13.1 (identification of self as a role model) necessitates a high degree of self-reflection or interaction with professionals, making it challenging to implement in apps structured around predefined conditional logic [40]. Many of the reviewed apps were designed to deliver standardized, rule-based responses to user behavior, thereby limiting their capacity to offer context-sensitive or highly personalized feedback.

According to the scoping review by Panicker et al [41], smartphone apps were the most frequently used tools in interventions using wearable devices to promote physical activity, although wrist-worn and pendant-type devices were also used. Selecting hardware and software that supports user engagement and integrating them when appropriate may facilitate the implementation of a broader range of BCTs. Additionally, an RCT conducted by Lunde et al [42], which was excluded from this review due to population mismatch, reported a significant increase in step count when professional support was delivered via a smartphone app. This finding suggests that integrating smartphone apps with professional guidance may enable more personalized interventions and generate synergistic effects. Furthermore, with the rapid advancement of generative artificial intelligence in health care [20], large language models can now be integrated into mHealth interventions via application programming interfaces, thereby enabling more personalized and adaptive interventions [43]. This development may expand the capacity of mHealth systems to support a broader range of BCTs in the future.

Limitations

This study has several limitations that warrant caution when interpreting the findings. First, none of the included studies were full RCTs, so the review lacks high-level evidence on the effectiveness of smartphone apps for promoting walking among older adults. Second, many of the reviewed studies featured research designs that may be considered suboptimal, likely reflecting the challenge of keeping pace with the rapid evolution of ICT. This limitation could have resulted in the underestimation of the health effects of ICT-based interventions, potentially biasing the results of this review. Third, the operational definition of "older adults" varied across studies, introducing heterogeneity and potential bias. Moreover, older adults who are able to use ICT tend to be relatively younger and of higher sociodemographic status, reflecting a selection

XSL•FO
RenderX

bias. This inherent selection bias should be considered when evaluating the generalizability of our findings. Fourth, this review included only 8 studies, which is fewer than in reviews encompassing older adults with more diverse sociodemographic backgrounds [28]. Because this review focused exclusively on community-dwelling older adults rather than those residing in institutions or hospitals, its findings may not generalize to populations with different backgrounds. Finally, this type of research has inherent limitations that should be acknowledged, such as publication bias (the tendency for statistically significant studies to be published preferentially) [44] and selection bias (the possibility that eligible studies were not captured in the search) [45].

Conclusion

In this study, we conducted a systematic review of 8 studies examining the effects of smartphone app use on walking behavior among community-dwelling older adults. Several studies reported improvements in step counts and intention to walk. Despite the limited number of eligible studies, the findings suggest that apps incorporating features to support walking-related knowledge acquisition, goal setting and planning, and behavioral monitoring may be especially effective in promoting walking behavior and intention among older adults.

Acknowledgments

No generative artificial intelligence was used at any stage of the research conduct or manuscript preparation.

Funding

This study was supported by JSPS KAKENHI (grant JP23K27908 and JP25K14166).

Data Availability

Data are supplied in supporting files available for download along with the published manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details of the behavior change techniques taxonomy.
[DOCX File, 19 KB - humanfactors_v13i1e78042_app1.docx]

Multimedia Appendix 2
Literature search strategy.
[DOCX File, 28 KB - humanfactors_v13i1e78042_app2.docx]

Multimedia Appendix 3
Study quality assessment tools.
[PDF File, 50 KB - humanfactors_v13i1e78042_app3.pdf]

Checklist 1
PRISMA 2020 checklist.
[DOCX File, 276 KB - humanfactors_v13i1e78042_app4.docx]

References

1. Current population estimates as of October 1, 2024. Statistics Bureau of Japan. URL: https://www.stat.go.jp/english/data/jinsui/2024np/index.html [accessed 2025-06-14]
2. Ageing and health. World Health Organization. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2025-06-14]
3. Merchant RA, Morley JE, Izquierdo M. Editorial: exercise, aging and frailty—guidelines for increasing function. J Nutr Health Aging 2021;25(4):405-409. [doi: 10.1007/s12603-021-1590-x] [Medline: 33786554]
4. Zhao W, Ukawa S, Kawamura T, et al. Health benefits of daily walking on mortality among younger-elderly men with or without major critical diseases in the New Integrated Suburban Seniority Investigation Project: a prospective cohort study. J Epidemiol 2015;25(10):609-616. [doi: 10.2188/jea.JE20140190] [Medline: 26155815]
5. Shah S, Chen Y, Owen A, et al. Walking for transport and all-cause mortality: a prospective cohort study of Australian community-dwelling older adults. BMJ Public Health 2024 Jun;2(1):e000709. [doi: 10.1136/bmjph-2023-000709] [Medline: 40018171]

XSL•FO
RenderX

6. Patel AV, Hildebrand JS, Leach CR, et al. Walking in relation to mortality in a large prospective cohort of older U.S. adults. Am J Prev Med 2018 Jan;54(1):10-19. [doi: 10.1016/j.amepre.2017.08.019] [Medline: 29056372]
7. Grimmer M, Riener R, Walsh CJ, Seyfarth A. Mobility related physical and functional losses due to aging and disease: a motivation for lower limb exoskeletons. J Neuroeng Rehabil 2019 Jan 3;16(1):2. [doi: 10.1186/s12984-018-0458-8] [Medline: 30606194]
8. Ferrucci L, Cooper R, Shardell M, Simonsick EM, Schrack JA, Kuh D. Age-related change in mobility: perspectives from life course epidemiology and geroscience. J Gerontol A Biol Sci Med Sci 2016 Sep;71(9):1184-1194. [doi: 10.1093/gerona/glw043] [Medline: 26975983]
9. Leung KM, Ou KL, Chung PK, Thøgersen-Ntoumani C. Older adults' perceptions toward walking: a qualitative study using a social-ecological model. Int J Environ Res Public Health 2021 Jul 20;18(14):7686. [doi: 10.3390/ijerph18147686] [Medline: 34300136]
10. Ahmed S, Lazo Green K, McGarrigle L, Money A, Pendleton N, Todd C. Interventions based on behavior change techniques to encourage physical activity or decrease sedentary behavior in community-dwelling adults aged 50-70: systematic review with intervention component analysis. J Aging Phys Act 2024 Aug 1;32(4):554-577. [doi: 10.1123/japa.2023-0140] [Medline: 38663855]
11. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
12. Arnautovska U, O'callaghan F, Hamilton K. Behaviour change techniques to facilitate physical activity in older adults: what and how. Ageing Soc 2018 Dec;38(12):2590-2616. [doi: 10.1017/S0144686X17001027]
13. Gilchrist H, Oliveira JS, Kwok WS, et al. Use of behavior change techniques in physical activity programs and services for older adults: findings from a rapid review. Ann Behav Med 2024 Feb 10;58(3):216-226. [doi: 10.1093/abm/kaad074] [Medline: 38300788]
14. Chen S, Wang C, Ko A, et al. Effectiveness of mobile health interventions for reducing sitting time in older adults: systematic review and meta-analysis. J Med Internet Res 2025 May 8;27:e60889. [doi: 10.2196/60889] [Medline: 40340833]
15. Kim H, Kim G, Kim Y, Ha J. The effects of ICT-based interventions on physical mobility of older adults: a systematic literature review and meta-analysis. Int J Clin Pract 2023;2023:5779711. [doi: 10.1155/2023/5779711] [Medline: 38020536]
16. Llopis P, Oviedo-Trespalacios O, Alonso F, Faus M, Useche SA. Walking apps for active aging? A literature and multi-source review of their impact on older adult users' mobility and wellbeing. Sustain Fut 2025 Dec;10:100982. [doi: 10.1016/j.sftr.2025.100982]
17. Yerrakalva D, Yerrakalva D, Hajna S, Griffin S. Effects of mobile health app interventions on sedentary time, physical activity, and fitness in older adults: systematic review and meta-analysis. J Med Internet Res 2019 Nov 28;21(11):e14343. [doi: 10.2196/14343] [Medline: 31778121]
18. Stockwell S, Schofield P, Fisher A, et al. Digital behavior change interventions to promote physical activity and/or reduce sedentary behavior in older adults: a systematic review and meta-analysis. Exp Gerontol 2019 Jun;120:68-87. [doi: 10.1016/j.exger.2019.02.020] [Medline: 30836130]
19. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232. [doi: 10.2196/jmir.7126] [Medline: 28663162]
20. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun;6(2):94-98. [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
21. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282. [doi: 10.11613/BM.2012.031] [Medline: 23092060]
22. Study quality assessment tools. National Heart Lung and Blood Institute (NHLBI). 2021. URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [accessed 2025-06-14]
23. Castro PC, Romano LB, Frohlich D, et al. Tailoring digital apps to support active ageing in a low income community. PLoS One 2020;15(12):e0242192. [doi: 10.1371/journal.pone.0242192] [Medline: 33301455]
24. Zhong R, Rau PLP. A mobile phone-based gait assessment app for the elderly: development and evaluation. JMIR Mhealth Uhealth 2020 Feb 29;8(2):e14453. [doi: 10.2196/14453] [Medline: 32452821]
25. Fong SSM, Ng SSM, Cheng YTY, et al. Comparison between smartphone pedometer applications and traditional pedometers for improving physical activity and body mass index in community-dwelling older adults. J Phys Ther Sci 2016 May;28(5):1651-1656. [doi: 10.1589/jpts.28.1651] [Medline: 27313391]
26. Paul L, Brewster S, Wyke S, et al. Increasing physical activity in older adults using STARFISH, an interactive smartphone application (app); a pilot study. J Rehabil Assist Technol Eng 2017;4:2055668317696236. [doi: 10.1177/2055668317696236] [Medline: 31186925]
27. Matsuoka Y, Yoshida H, Hanazato M. A smartphone-based shopping mall walking program and daily walking steps. JAMA Netw Open 2024 Jan 2;7(1):e2353957. [doi: 10.1001/jamanetworkopen.2023.53957] [Medline: 38289599]

XSL•FO
RenderX

28. Chan NPT, Lai AYK, Choy HK, et al. Feasibility and potential effectiveness of a smartphone zero-time exercise intervention for promoting physical activity and fitness in patients with coronary heart disease: a pilot randomized controlled trial. Front Public Health 2022;10:865712. [doi: 10.3389/fpubh.2022.865712] [Medline: 35910893]
29. Kawaguchi K, Nakagomi A, Ide K, Kondo K. Effects of a mobile app to promote social participation on older adults: randomized controlled trial. J Med Internet Res 2024 Sep 30;26:e64196. [doi: 10.2196/64196] [Medline: 39348180]
30. Ohta T, Osuka Y, Shida T, et al. Feasibility, acceptability, and potential efficacy of a mobile health application for community-dwelling older adults with frailty and pre-frailty: a pilot study. Nutrients 2024 Apr 16;16(8):1181. [doi: 10.3390/nu16081181] [Medline: 38674872]
31. Blomqvist A, Bäck M, Klompstra L, Strömberg A, Jaarsma T. Testing the recruitment frequency, implementation fidelity, and feasibility of outcomes of the Heart Failure Activity Coach study (HEALTHY): pilot randomized controlled trial. JMIR Form Res 2025 Jan 8;9:e62910. [doi: 10.2196/62910] [Medline: 39778202]
32. Turrisi TB, Bittel KM, West AB, et al. Seasons, weather, and device-measured movement behaviors: a scoping review from 2006 to 2020. Int J Behav Nutr Phys Act 2021 Feb 4;18(1):24. [doi: 10.1186/s12966-021-01091-1] [Medline: 33541375]
33. Ferguson T, Olds T, Curtis R, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022 Aug;4(8):e615-e626. [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
34. Wilson SA, Byrne P, Rodgers SE, Maden M. A systematic review of smartphone and tablet use by older adults with and without cognitive impairment. Innov Aging 2022;6(2):igac002. [doi: 10.1093/geroni/igac002] [Medline: 35243008]
35. Gomez-Hernandez M, Ferre X, Moral C, Villalba-Mora E. Design guidelines of mobile apps for older adults: systematic review and thematic analysis. JMIR Mhealth Uhealth 2023 Sep 21;11:e43186. [doi: 10.2196/43186] [Medline: 37733401]
36. Laranjo L, Ding D, Heleno B, et al. Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. Br J Sports Med 2021 Apr;55(8):422-432. [doi: 10.1136/bjsports-2020-102892] [Medline: 33355160]
37. Seifert A, Schlomann A, Rietz C, Schelling HR. The use of mobile devices for physical activity tracking in older adults' everyday life. Digit Health 2017;3:2055207617740088. [doi: 10.1177/2055207617740088] [Medline: 29942617]
38. HealthKit documentation. Apple Developer. URL: https://developer.apple.com/documentation/healthkit [accessed 2025-06-14]
39. Google Fit for developers [Website in Japanese]. Google Fit. URL: https://developers.google.com/fit?hl=ja [accessed 2025-06-14]
40. Machorro-Cano I, Olmedo-Aguirre JO, Alor-Hernández G, Rodríguez-Mazahua L, Sánchez-Morales LN, Pérez-Castro N. Cloud-based platforms for health monitoring: a review. Informatics (MDPI) 2024;11(1):2. [doi: 10.3390/informatics11010002]
41. Panicker RM, Chandrasekaran B. "Wearables on vogue": a scoping review on wearables on physical activity and sedentary behavior during COVID-19 pandemic. Sport Sci Health 2022;18(3):641-657. [doi: 10.1007/s11332-021-00885-x] [Medline: 35018193]
42. Lunde P, Bye A, Grimsmo J, et al. Effects of individualized follow-up with an app postcardiac rehabilitation: five-year follow-up of a randomized controlled trial. J Med Internet Res 2025 Feb 13;27:e60256. [doi: 10.2196/60256] [Medline: 39946716]
43. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida AJ, et al. Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: current scenario and challenges. J Med Internet Res 2023 May 4;25:e44030. [doi: 10.2196/44030] [Medline: 37140973]
44. Dwan K, Gamble C, Williamson PR, Kirkham JJ, Reporting Bias Group. Systematic review of the empirical evidence of study publication bias and outcome reporting bias: an updated review. PLOS One 2013;8(7):e66844. [doi: 10.1371/journal.pone.0066844] [Medline: 23861749]
45. Drucker AM, Fleming P, Chan AW. Research techniques made simple: assessing risk of bias in systematic reviews. J Invest Dermatol 2016 Nov;136(11):e109-e114. [doi: 10.1016/j.jid.2016.08.021] [Medline: 27772550]

Abbreviations

BCT: behavior change technique
ICT: information and communication technology
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial

XSL•FO
RenderX

Edited by A Kushniruk; submitted 03.Jul.2025; peer-reviewed by F Alonso, S Assander; revised version received 06.Dec.2025; accepted 02.Jan.2026; published 12.Feb.2026.

Please cite as:
Abe H, Isinuki T, Ono R, Kanaya N, Tanno M, Hirano M
Physical and Psychological Effects of Smartphone App–Based Walking Interventions in Community-Dwelling Older Adults: Systematic Review and Behavior Change Technique–Informed Analysis
JMIR Hum Factors 2026;13:e78042
URL: https://humanfactors.jmir.org/2026/1/e78042
doi:10.2196/78042

© Hiroki Abe, Tomohiro Isinuki, Rinako Ono, Naoki Kanaya, Masaya Tanno, Michiyo Hirano. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 12.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

How and Why Telehealth Interventions Improve Self-Care Among Vulnerable Groups of Patients With Heart Failure: Scoping Review and Rapid Realist Synthesis

Saleema Allana[1], RN, PhD; Colleen Norris[2], RN, PhD; Armish Hussain[3], BScN, RN, MNSc; Nada Alaidarous[1], RN, PhD candidate; Raza Muhammad Haider[1], BSc, MMASc; Alexander M Clark[4], RN, PhD

[1]Faculty of Health Sciences, School of Nursing, Western University, 1151 Richmond St, London, ON, Canada

[2]Faculty of Health Sciences, University of Alberta, Edmonton, AB, Canada

[3]Queen's University, Kingston, ON, Canada

[4]Faculty of Health Disciplines, Athabasca University, Athabasca, AB, Canada

Corresponding Author:
Saleema Allana, RN, PhD
Faculty of Health Sciences, School of Nursing, Western University, 1151 Richmond St, London, ON, Canada

Abstract

Background: Heart failure (HF) is a prevalent condition among older adults in Canada, often leading to reduced quality of life and frequent hospitalizations. HF disease management interventions, particularly those delivered through telehealth, aim to improve care by fostering self-care and reducing readmissions. However, disparities in access to and use of HF telehealth services persist among vulnerable populations.

Objective: This study aimed to present the findings of a scoping review and a rapid realist synthesis of HF telehealth interventions for vulnerable groups of patients with HF. This review is underpinned by the metatheory of critical realism and intersectionality theory.

Methods: A rapid realist synthesis of the retrieved literature was undertaken to explore the underlying mechanisms and contexts that make HF telehealth interventions work or not work for marginalized groups of patients with HF.

Results: The realist review findings indicated that vulnerable patients require simple interventions. The findings also suggested that for effective use of telehealth and remote monitoring services, these patients require simplified training that could increase their confidence. The review findings further demonstrated that involving patients' family members in the delivery of telehealth interventions ensures success.

Conclusions: Future research with vulnerable populations should be underpinned by the critical realism and intersectionality theory and should apply the principles of intersectionality at all stages of the research process, including evaluation and analysis. This review also urges HF practitioners to apply the principles of intersectionality and health equity in clinical practice, such that interventions are simple and personalized, involve family members, include an in-person component, provide training for patients and health professionals, and integrate telemonitoring data.

(JMIR Hum Factors 2026;13:e78859) doi:10.2196/78859

KEYWORDS

telehealth; heart failure; marginalized patients; intersectionality; rapid realist review

Introduction

Background

Heart failure (HF) is a widespread health condition affecting many older adults in Canada. HF can significantly impair quality of life and is becoming increasingly common as populations age globally [1]. Heart failure disease management interventions (HFDMIs) aim to address these challenges by fostering self-care, enhancing well-being, and reducing recurrent hospitalizations and subsequent health care expenditure. These programs achieve this through patient education, ongoing follow-up, and improving care coordination, which in turn reduces emergency visits and promotes outpatient care [2,3].

HFDMIs encompass a wide range of interventions, including in-person interventions in hospitals, communities, or clinics, as well as remote interventions delivered through telehealth technologies, such as telephone, email, internet, text messages, telemonitoring, mobile health, video consultation, and wireless monitoring systems [4-9]. Despite its potential to improve chronic disease management, telehealth interventions are complex and multifaceted [10]. While telehealth technologies were introduced into HF management in 1998 [5], their

XSL•FO
RenderX

complexity as a health delivery model becomes evident when considering the various components and interactions involved and when compared with traditional face-to-face interventions [11]. Effective implementation of telehealth interventions requires harmonizing clinical protocols, leadership support, and technological factors, along with ensuring that patient and health care providers are confident in using technology through education and training [11]. As such, understanding the specific contexts and mechanisms that make telehealth interventions effective in managing HF becomes crucial.

This study details the findings of a rapid realist synthesis, which investigated the use of telehealth solutions for vulnerable patients with HF. The synthesis is framed within the approach of critical realism (CR) [12], aiming to explore the underlying mechanisms and contexts that contribute to the significance of telehealth interventions in improving HF management among vulnerable populations. Additionally, an intersectionality lens was applied to examine how the intersecting social identities of vulnerable groups with HF may affect their access to and use of telehealth services [13].

Intersectionality, Vulnerability, and Disparities—The Complex Interplay

Intersectionality highlights how different social identities, such as race, gender, age, culture, sexual orientation, socioeconomic status, and geographic location, combine to shape individual experiences [14-16]. These overlapping factors can either grant privilege or deepen inequalities, particularly in health care [14-16]. In the context of HF care, intersectionality plays a vital role in determining access to resources, medical treatment, and self-care support [14]. Patients from marginalized groups, including racialized minority groups and gender minority groups, rural residents, older adults, and people with disabilities, often struggle to receive adequate health care and telehealth services [17]. These disparities stem from systemic barriers such as economic hardship, geographic isolation, and implicit biases within health care systems [17]. The unfortunate reality is that those who require the most care often face the greatest barriers to accessing it, perpetuating health care inequities. These intersecting social identities may act as contextual factors or mechanisms that intensify disparities among marginalized populations and contribute to poorer health outcomes.

HF Telehealth Interventions—The Inherent Complexity

The use of telehealth has increased during the COVID-19 pandemic [18]. Nonetheless, telehealth adoption varies among demographic groups, with factors such as age, income, race, ethnicity, health literacy, digital skills, and language barriers playing significant roles. These disparities stem from challenges at multiple levels, including the patients themselves, health care systems, telehealth platforms, and overarching policies [18]. Recognizing and addressing these disparities is crucial to ensure fairness in telehealth services, optimize their impact on health care delivery, and ultimately promote telehealth equity [18].

The complexity of HFDMIs lies in the multifaceted nature of the condition, requiring innovative approaches to optimize patient outcomes. The adoption of telehealth interventions for vulnerable patients with HF is complicated by the overlapping social identities that can create barriers to accessing and using telehealth services. Although telehealth became widely accessible during the COVID-19 pandemic [19], these at-risk groups have still faced considerable barriers in making full use of telehealth options [20].

Current understandings of telehealth interventions for HF in vulnerable populations are hindered by 2 issues. First, there is a significant gap in research specifically addressing telehealth solutions tailored to marginalized patients with HF with intersecting identities, limiting insights into their effectiveness [21]. Second, addressing the complexities of telehealth in HF care for disadvantaged groups calls for advanced research methodologies, such as realist methods, that can unravel the contextual elements driving these challenges [22].

Rapid Realist Synthesis—Exploring Complexity

The use of theory-based research approaches to evaluate telehealth interventions has been urged in the literature [23]. There is increasing consensus that telehealth interventions should be evaluated using complexity-driven approaches [23]. In relation to such interventions, these approaches often bring complexity into various aspects, including the theoretical underpinnings of the research, the type of research questions posed, and the types of methods used. In recent years, the research review and synthesis field has seen an expansion in the variety of approaches used [24]. Among these approaches, realist synthesis has gained importance. Unlike traditional reviews that assess the effectiveness of interventions, realist synthesis aims to comprehend and analyze the underlying mechanisms that contribute to the success or failure of an intervention, offering explanatory insights rather than simple evaluations of its effectiveness [24,25]. Rapid realist syntheses are philosophically and theoretically underpinned by CR.

The CR Paradigm

CR focuses on uncovering the underlying mechanisms that generate outcomes, making it an explanatory rather than purely descriptive approach. It is particularly suited for addressing "why" questions about social and health phenomena. Within intervention research, this perspective shifts attention from simply assessing effectiveness to examining the reasons behind its effects. In other words, it asks what makes an intervention work in practice. This focus is not only practically valuable but also grounded in CR's ontological stance [12]. This study used a rapid realist synthesis approach that is underpinned by CR and intersectionality theory, which offers an in-depth understanding of how intersecting vulnerabilities, intervention components, and contextual health care factors interact to shape outcomes in HF telehealth delivery.

Methods

The search and screening methodology of this review has already been described in a previously published scoping review and intersectionality-based analysis titled "A scoping review and intersectionality-based analysis of heart failure telehealth interventions for vulnerable populations" [26]. The search strategy, screening process, and data extraction were identical

XSL•FO
RenderX

to this previous scoping review; full methodological details are previously published in the abovementioned scoping review paper [26]. However, the research questions and the subsequent analysis of data were completely different across the 2 papers. The previous paper reported the findings of a quantitative intersectionality-based analysis of the included studies [26]. However, this paper reports the findings of a rapid realist synthesis, which was undertaken to identify the contexts, mechanisms, and outcomes that make HF telehealth interventions work or not for vulnerable groups of patients with HF. This realist synthesis was important for understanding which HF telehealth interventions work for which vulnerable populations and why.

This scoping review was undertaken following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist (Checklist 1). To summarize the search strategy, a systematic search was conducted across MEDLINE, CINAHL, Scopus, and the Cochrane Central Register of Controlled Trials, with additional gray literature sourced via ProQuest Dissertations and Theses Global. Search terms included variations of “heart failure” and “telehealth,” combined with terms representing intersecting identities (eg, race, gender, income, and rurality) to capture studies involving marginalized populations. The screening process used a 2-step, independent, dual-reviewer approach using Covidence (Veritas Health Innovation Ltd), with conflicts resolved by a third reviewer. Studies were eligible if they examined telehealth HFDMIs involving marginalized patients with HF and/or their caregivers. Only full, English-language papers with primary data on the development, implementation, or evaluation of such interventions were included. Outcomes of interest included readmissions, mortality, quality of life, and self-care. Qualitative and mixed methods studies capturing patient, caregiver, or health care professionals’ experiences with telehealth were also considered. Data extraction was guided by a standardized form and included study characteristics, intervention details, and target populations. Study quality was appraised using the Mixed Methods Appraisal Tool.

This rapid realist synthesis was conducted following the Realist and Meta-narrative Evidence Syntheses: Evolving Standards (RAMESES) guidelines for realist reviews, using a theory-driven approach informed by CR and intersectionality theory [27]. Our aim was to uncover how and why HF telehealth interventions work, or do not work, for vulnerable populations by identifying contexts, mechanisms, and their outcomes from published literature.

Unlike stakeholder-engaged realist reviews that involve co-development of program theories through consultation, this rapid realist synthesis is document driven. We did not engage stakeholders in the formulation or refinement of program theories as it was a rapid realist review constrained by time and resources. Instead, the contexts and mechanisms were identified through analysis of peer-reviewed studies that were already mapped in a previously published scoping review [26].

We specifically searched for contexts and mechanisms, identified from the outcomes of the studies, in which HF telehealth interventions worked and did not work. Every study was thoroughly reviewed, analyzed, and summarized to identify these factors. To identify the underlying mechanisms and contexts from the included studies, retroduction was used. Retroduction is a method of analysis in which a pertinent observation or effect is first identified, and then, the researcher moves backward to understand the underlying mechanism or the causal reasoning [28]. Context- and mechanism-related data extracted from the included studies were synthesized by identifying the reoccurring themes across studies.

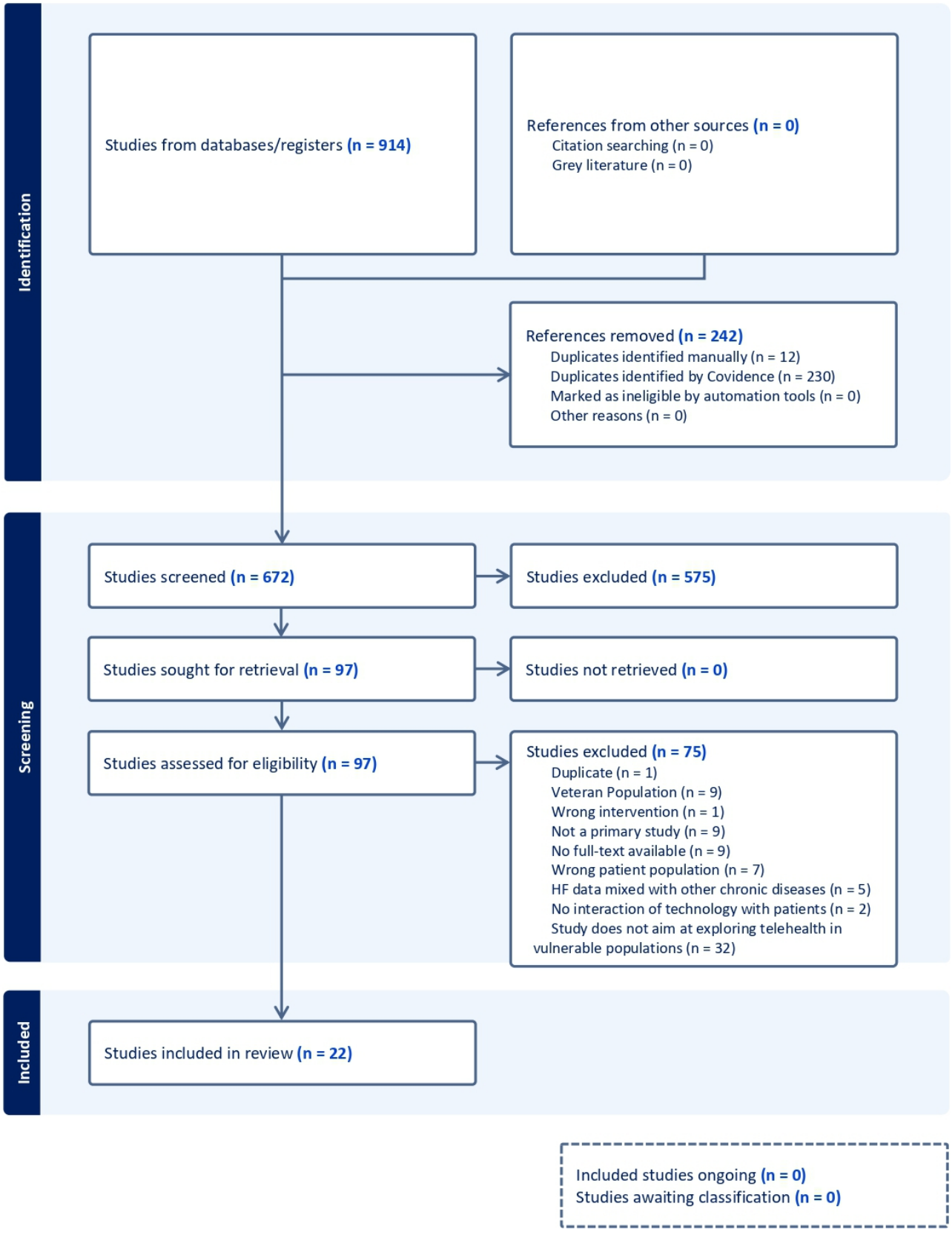
Results

Total Number of Studies Screened and Included

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram for this scoping review is shown in Figure 1, illustrating the number of studies included, excluded, and screened at each stage of the screening process (Checklist 1).

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram. HF: heart failure.



Characteristics of the Included Studies

As assessed using the Mixed Methods Appraisal Tool, the studies included in this review were generally of high quality. A diverse range of methodologies and a relatively narrower geographical scope were identified among the 22 studies included in the review. The methodologies varied widely, with 8 (36.4%) studies using randomized controlled trials [29-36]. Other designs included 2 (9.1%) pretest-posttest studies [37,38], 3 (13.6%) cross-sectional or quantitative surveys [39-41], and 3 (13.6%) retrospective data collection and analysis [42-44]. The review also identified 2 (9.1%) feasibility trials [6,45], 2

XSL•FO
RenderX

(9.1%) qualitative studies [46,47], and 1 (4.55%) mixed methods study [48]. Notably, 1 (4.5%) study focused solely on developing and implementing a telehealth program for patients with HF [49]. The studies primarily examined vulnerable groups, including Black [6,29-32,35,42,44,46,47], Hispanic [30,31,33,35,39,42,47], rural [32,34,36,38,40,49], underserved [30,38,41,43], low-income [6,29,37,43-45,47], and low-education patients with HF [6,29,32,44,48]. Telehealth interventions varied, including telephone support [6,29,32,34,36,48], remote monitoring [31,38-40,42,43,45,46,48], video consultations [37,44,49], and combinations of remote monitoring and video consultations [30,47]. Other approaches included telephone case management [33] and a remote monitoring with symptom response system [35]. Among the included studies, 19 (86.4%) were conducted in the United States, indicating that the data predominantly originated from the United States. A total of 10 contexts and mechanisms were identified, of which 7 were contexts and 3 were mechanisms (Table 1).

XSL•FO
RenderX

Table . Characteristics of the included studies.

Study	Study design	Vulnerable population studied	Type of telehealth intervention
Bakitas et al [6]	Feasibility trial	African Americans, rural and urban population, and low educational level	Telephonic nurse coach sessions and monthly calls
Bakitas et al [29]	Randomized clinical trial	African Americans, rural and urban population, and low educational level	Telephonic nurse coach sessions and monthly calls
Newell et al [49]	Telehealth intervention design and implementation	Rural population	Video consultations
Pekmezaris et al [30]	Randomized controlled trial	Underserved Blacks and Hispanics	Remote monitoring and video consultations
Pekmezaris et al [47]	Qualitative study	Lower-income Blacks and Hispanics	Remote monitoring and video consultations
Dang et al [31]	Randomized controlled trial	Blacks and Hispanics	Mobile phone–based telemonitoring
Dionne-Odom et al [32]	Randomized clinical trial	African Americans, rural and urban population, and low educational level	Telephonic nurse coach sessions and monthly calls
Rosen et al [37]	Pretest and posttest design	Low-income patients	Video consultations
Riegel et al [33]	Randomized controlled trial	Hispanics of Mexican origin	Telephone case management, including decision support software and telephone support
Turchioe et al [39]	Cross-sectional feasibility study	English- and Spanish-speaking older adults	Mobile app to report symptoms
Heiney et al [46]	Qualitative study	African Americans	Mobile app for weight monitoring, health messages, and journaling
Jaana and Sherrard [40]	Cross-sectional study	Rural population	Remote monitoring via mobile phone
Bakhshi et al [45]	Feasibility trial	Low-income patients	Remote monitoring
Liu [41]	Quantitative survey method	Medically underserved urban population	Telemonitoring
Srisuk et al [34]	Randomized controlled trial	Rural population	Telephone support
Cabán [42]	Retrospective quantitative study	Hispanics, non-Hispanic Blacks, and non-Hispanic Whites	Home telemonitoring
Soran et al [35]	Randomized controlled trial	Older women, African Americans, and Hispanics	Telemonitoring and telephone-based symptom response system
Riley et al [38]	Pretest-posttest design	Underserved patients in rural communities	Remote monitoring
Krum et al [36]	Randomized controlled trial	Rural and remote communities of Australia	Telephone support
Lefler et al [48]	Mixed methods study	Low educational level	Remote monitoring and telephone support
Davis et al [43]	Retrospective cohort design	Underserved patients	Remote monitoring
Sammour et al [44]	Retrospective study design	African Americans, low-income, and low educational level	Video and telephone consultations

Contexts That Made HF Telehealth Interventions Work

Family Members' Involvement in Telehealth or Remote Monitoring Programs

The involvement of family members in training and implementing telehealth or remote monitoring programs was found to be crucial for the success of interventions. Studies by Riley et al [38] and Pekmazaris et al [47] demonstrated the positive impact of including family members in training and support. A study conducted in rural Thailand exclusively tested a family-based telehealth intervention and found it to be highly successful [34]. Similarly, Bakitas et al [6,47] and Dionne-Odom et al [32] tested palliative care interventions for patients with

XSL•FO
RenderX

HF that involved the active participation of family caregivers and demonstrated success.

Training of Health Professionals

The importance of training patients, family caregivers, and health professionals for telehealth programs has been emphasized in the review findings [6,31,35,37]. Bakitas et al [6] provided 20 hours of training to nurse coaches, focusing on effective communication and patient management through telehealth intervention. Dang et al [31] strongly recommended staff training based on their experience with telehealth interventions involving Black and Hispanic patients with HF. A different study observed that primary care physicians did not respond appropriately to clinical deterioration in remotely monitored patients with HF [35]. Accordingly, Soran et al [35] recommended training physicians and other health professionals involved in telehealth and remote monitoring interventions to ensure prompt and appropriate actions. In an interesting telehealth intervention by Rosen et al [37], social workers played a primary role in delivering the intervention. They received extensive training in health-related aspects of HF care while contributing their expertise in social aspects of care.

Significance of an In-Person Component Along With the Telehealth or Remote Monitoring Program

One of the significant findings of this review was that, in many studies, participants appreciated or requested at least 1 in-person meeting or interaction with their health care providers along with a telehealth or remote monitoring component. Those interventions that were a combination of both in-person and telehealth components were also found to be relatively more effective [6,32,34,37,43,47]. The in-person components, in most cases, were meant to build rapport with patients and support training or education. In the study by Pekmazaris et al [47], the Community Advisory Board, consisting of low-income Black and Hispanic patients, recommended arranging the first in-person meeting between patients with HF and the telemonitoring nurse. Comparing their study findings with other studies in this area of research, Soran et al [35] strongly suggested including a home visit, even if it was only possible once.

One Size Does Not Fit All—Personalized Tailoring of Telehealth Interventions

In this review, the importance of personalized and tailored interventions for vulnerable groups of patients with HF has been highlighted through several studies [31,33-35,39,43,46]. Riegel et al [33] conducted a trial of telephone case management with Hispanic patients with HF of Mexican origin, customizing the intervention to align with their cultural needs. Bilingual and bicultural health professionals, along with the integration of cultural values, trust, family inclusion, and problem-solving, were key components of this tailored intervention. Another example was the HF telehealth intervention by Srisuk et al [34] for rural Thai patients, which involved family caregivers at each stage of the intervention and incorporated cultural elements such as translated materials and culturally relevant content. Pekmazaris et al [47] used a community-based participatory research approach, involving Black and Hispanic patients with HF in developing a user-centered intervention through a Community Advisory Board. Patient feedback was used to modify the intervention delivery and equipment. Davis et al [43] recommended personalized, consistent feedback for patients with HF by telehealth providers (physicians, nurses, and allied health professionals involved in the delivery of telehealth services) based on their findings, while Dang et al [31] allowed patients to choose their preferred time and language for daily questions, emphasizing the importance of personalization. Heiney et al [46] involved African American patients in developing a mobile health intervention and ensured that the content was culturally relevant to their values. Pekmazaris et al [47] also emphasized the unique characteristics of vulnerable patients with HF, such as younger age and higher comorbidity rates, necessitating tailored telehealth programs considering cultural values, beliefs, and language.

Contextual Realities of Health Care Systems

Telehealth interventions are performed within the broader health care system; therefore, these are bound to be affected by the contextual realities of health care systems. In a study by Lefler et al [48], participants alluded to the unavailability of health professionals to respond to them or to unsatisfactory responses by them, when patients needed guidance on managing their symptoms. This could potentially lead to clinical deterioration and could result in the very outcomes that we are trying to prevent through the use of telehealth interventions. Dang et al [31] suggested using more health professionals and a variety of them (nurses, doctors, and dieticians) to be better able to support patients with HF via telehealth interventions.

Another very important contextual reality has been discussed by Soran et al [35]. They indicated the fact that a typical patient with HF is not seen by a cardiologist or a chronic care team; rather, such patients often present to primary care physicians in community settings. This holds true for most vulnerable patients with HF. However, most disease management programs, including telehealth interventions, are run in large academic centers [35]. To reach vulnerable patient populations, HF telehealth interventions need to be tailored accordingly and must be offered in community settings via primary care physicians' offices.

Soran et al [35] also alluded to another important fact that is part of the context. They indicated that when telehealth interventions are not integrated into workflow practices, it becomes difficult to run these interventions successfully. Riley et al [38] also made important recommendations in this regard. They suggested that to be useful, telemonitoring data must be incorporated in the care team's workflow, such that health care providers (physicians, nurses, and allied health professionals) know when and how to intervene based on the data received from patients via telemonitoring [38].

Equipment-Related Issues

As part of this review, some studies have highlighted the equipment-related issues that interfered with the delivery of telehealth interventions. Most often, these issues were highlighted by the patients with HF themselves. Some participants described setting up the equipment as a tough task,

XSL•FO
RenderX

such as leveling the weighing scale, setting up the iPad, etc [48]. In another study by Dang et al [31], Black and Hispanic patients with HF suggested having larger font sizes on screen, accurate translation of equipment names in Spanish, slowing down the speed of verbal instructions provided during telemonitoring, and decreasing the use of medical jargon.

Technical Issues

As telehealth and remote monitoring interventions largely rely on internet connectivity for virtual consultations and data transmission, internet connectivity problems emerged as one of the biggest technical issues that hampered the delivery of telehealth and remote monitoring interventions [38,47]. Riley et al [38] reported working proactively to ensure adequate network stability required for the transmission of patient data. Another important technical issue was discussed by Dang et al [31] based on their experience with web browser messaging. They suggested that text messages are not only more user-friendly than web browser messages but also require less data and therefore pose fewer connectivity challenges.

Mechanisms That Made HF Telehealth Interventions Work

Simplified Interventions Work vs Too Complex

Simplified interventions have been found to be more effective for HF telehealth interventions targeting vulnerable patient populations, considering their low levels of education, health literacy, and digital literacy [38,45,48]. Riley et al [38] emphasized the complexity of remote monitoring interventions for rural patients, while Lefler et al [48] highlighted the cognitive changes and poor health literacy among patients with HF. Bakhshi et al [45] advocated for simplicity in telehealth interventions, ensuring the automatic transmission of weight data for ease of remote monitoring.

Simplified Patient Training Increases Confidence for the Use of Telehealth and Remote Monitoring Equipment

Simplified patient training plays a crucial role in increasing confidence and proficiency in using telehealth and remote monitoring equipment [6,41,42,47]. Cabán [42] provided basic training to racialized groups, teaching them essential skills and electronic communication with health care providers (physicians, nurses, and allied health professionals). Dang et al [31] offered one-on-one instruction and visual manuals for participants with low educational attainment, while another study emphasized the need for more visual training materials [6]. Additionally, lower-income and racialized minority groups and gender minority groups with HF suggested increasing training time and offering in-person sessions to enhance the effectiveness of the telehealth program for these patients [47].

Sense of Security and Increased Comfort With Managing Symptoms Reinforces the Use of Telemonitoring

An increased sense of security and comfort in managing symptoms was reported as positive outcomes of telemonitoring interventions, reinforcing their use [41,43,48]. For instance, in a study by Lefler et al [48], a theme emerged, “watching over me,” whereby participants described a sense of confidence that emerged with telemonitoring. This confidence came from the feeling that someone experienced and knowledgeable in health care was watching over them, leading them to use telemonitoring more effectively. Participants in another study discussed a similar sense of security [43]. Another important theme in the Lefler et al [48] study was the increased comfort in managing HF symptoms and heightened awareness of one’s health experienced by patients who used telemonitoring regularly. This comfort and awareness reinforced their use of telemonitoring daily. Liu [41] suggested that patients with HF who are confident and proficient in using technology and who perceive it as easy to use are more likely to view telehealth interventions as useful.

Discussion

Principal Findings

As per the findings of the previous paper, there was a lack of integration of the intersectionality-based principles in the body of literature on HF telehealth use among vulnerable populations [26]. To build upon these findings, we wanted to explore which of these interventions worked or did not work for these vulnerable populations and how and why they worked or did not work. Therefore, a realist synthesis approach was used to build on the findings of the initial review. This rapid realist synthesis outlined important contextual factors, including family involvement, health professional training, in-person components, personalized tailoring, contextual realities and systems factors, equipment, and technical issues. The synthesis also highlighted mechanisms that can improve HF telehealth programs for vulnerable patient populations. It recognized the effectiveness of simplified interventions, including patient training, which can lead to positive outcomes such as (1) increased patient confidence and proficiency in using telehealth technologies and (2) enhanced sense of security and comfort in managing symptoms and overall health.

To further strengthen the analysis, we considered the application of intersectionality in designing, implementing, and evaluating HF telehealth programs for vulnerable populations. The review highlights the need for increased research attention to health inequities experienced by these populations, particularly in designing interventions across different countries and vulnerable or marginalized groups. This aligns with other research highlighting the importance of using various research methods, including qualitative, interpretive, or realist approaches, to comprehensively understand complexities and mechanisms of these interventions [12].

The analysis in this review indicates that while studies have focused on specific populations such as African Americans, Hispanics, and rural low-income groups, there is a need to include other vulnerable populations, including other racialized minority groups and gender minority groups, immigrants and refugees, women with intersecting social identities, and Indigenous peoples. Additionally, the review highlights the effectiveness of various telehealth interventions such as telephone support, video consultations, remote monitoring, and case management, emphasizing the importance of tailoring these interventions to meet the specific needs of vulnerable patients with HF.

XSL•FO
RenderX

Access and use inequities in telehealth services for vulnerable patients are significant, given their low educational, health literacy, and digital literacy levels [50]. Simplified interventions and training are essential to enhance their confidence and proficiency in using telehealth and remote monitoring equipment [50]. Family involvement, particularly in racialized minority groups and rural populations, is crucial for success, considering the cultural significance of family in patient care [33,34].

The review also highlights the importance of health professionals' training in delivering effective telehealth interventions [51]. Ensuring access to HF telehealth programs for vulnerable patients, particularly within primary care clinics in community settings, is essential to maximize their benefits [12].

In-person interactions and the personalization of interventions are valued by patients, emphasizing the significance of building rapport and tailoring interventions to cultural, linguistic, and contextual needs [52]. Integrating telemonitoring data into the care team's workflow enables prompt action and decision-making in case of clinical deterioration [12].

Communication-related issues, such as language translations, avoiding medical jargon, and using simple and understandable language, are crucial for effectively using telehealth interventions by racialized minority groups and individuals with lower health literacy [12]. The review also highlights the positive impact of telemonitoring interventions on patients' confidence, sense of security, and comfort in managing HF symptoms, reinforcing their use [12].

Strengths and Limitations

This review is notable for its originality, being the first to conduct a rapid realist synthesis in this research area, and its systematic search in multiple and multidisciplinary databases. However, it is limited by its reliance on published articles in English, the exclusion of gray literature, and the inability to collect primary data from key stakeholders, which could have enhanced the findings and theory generation.

Recommendations

The review findings offer valuable insights for improving HF telehealth interventions in clinical practice and research. Key practice-based recommendations include simplifying telehealth interventions for patients with lower education, health literacy, and digital literacy; providing tailored training; involving family members for support; training health care professionals; incorporating in-person components; personalizing interventions; offering telehealth in community settings; integrating telemonitoring data; building trust and acceptance for self-management; and using clear language and appropriate translations for racialized minority groups.

Conclusions

In conclusion, this scoping review highlights the need for greater incorporation of appropriate philosophical and theoretical foundations in research on HF telehealth interventions for vulnerable populations. While intersectionality principles have been applied in problem identification and intervention development, their application in evaluation stages needs to be improved. In addressing this, future research should adopt CR and intersectionality theory and apply intersectionality principles throughout the research process. The review emphasizes the importance of HF practitioners applying intersectionality and health equity principles in clinical practice by implementing simple, personalized interventions involving family members; including an in-person component; training patients and health professionals; and integrating telemonitoring data into care workflows. By fully implementing these findings, HF telehealth interventions can significantly improve care for vulnerable patients, addressing their specific needs; enhancing access, acceptance, and use of telehealth; and ultimately reducing health inequities.

Funding

This work was supported by the Canadian Institutes of Health Research (CIHR) and Alberta Innovates funding.

Data Availability

Data sharing is not applicable to this paper as no datasets were generated or analyzed during this study.

Conflicts of Interest

None declared.

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 86 KB - humanfactors_v13i1e78859_app1.docx]

References

1. Piepoli MF, Adamo M, Barison A, et al. Preventing heart failure: a position paper of the Heart Failure Association in collaboration with the European Association of Preventive Cardiology. Eur J Prev Cardiol 2022 Feb 19;29(1):275-300. [doi: 10.1093/eurjpc/zwab147] [Medline: 35083829]

XSL•FO
RenderX

2. Jaarsma T, Hill L, Bayes-Genis A, et al. Self-care of heart failure patients: practical management recommendations from the Heart Failure Association of the European Society of Cardiology. Eur J Heart Fail 2021 Jan;23(1):157-174. [doi: 10.1002/ejhf.2008] [Medline: 32945600]
3. Zhao Q, Chen C, Zhang J, Ye Y, Fan X. Effects of self-management interventions on heart failure: systematic review and meta-analysis of randomized controlled trials. Int J Nurs Stud 2020 Oct;110:103689. [doi: 10.1016/j.ijnurstu.2020.103689] [Medline: 32679402]
4. Savard LA, Thompson DR, Clark AM. A meta-review of evidence on heart failure disease management programs: the challenges of describing and synthesizing evidence on complex interventions. Trials 2011 Aug 16;12:194. [doi: 10.1186/1745-6215-12-194] [Medline: 21846340]
5. Kotb A, Cameron C, Hsieh S, Wells G. Comparative effectiveness of different forms of telemedicine for individuals with heart failure (HF): a systematic review and network meta-analysis. PLoS ONE 2015;10(2):e0118681. [doi: 10.1371/journal.pone.0118681] [Medline: 25714962]
6. Bakitas M, Dionne-Odom JN, Pamboukian SV, et al. Engaging patients and families to create a feasible clinical trial integrating palliative and heart failure care: results of the ENABLE CHF-PC pilot clinical trial. BMC Palliat Care 2017 Aug 31;16(1):45. [doi: 10.1186/s12904-017-0226-8] [Medline: 28859648]
7. Athilingam P, Jenkins B. Mobile phone apps to support heart failure self-care management: integrative review. JMIR Cardio 2018 May 2;2(1):e10057. [doi: 10.2196/10057] [Medline: 31758762]
8. Gallagher J, James S, Keane C, et al. Heart failure virtual consultation: bridging the gap of heart failure care in the community - a mixed-methods evaluation. ESC Heart Fail 2017 Aug;4(3):252-258. [doi: 10.1002/ehf2.12163] [Medline: 28772044]
9. Chowdhury ME, Khandakar A, Alzoubi K, et al. Real-time smart-digital stethoscope system for heart diseases monitoring. Sensors (Basel) 2019 Jun 20;19(12):2781. [doi: 10.3390/s19122781] [Medline: 31226869]
10. Kvedar J, Coye MJ, Everett W. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Aff (Millwood) 2014 Feb;33(2):194-199. [doi: 10.1377/hlthaff.2013.0992] [Medline: 24493760]
11. Alami H, Fortin JP, Gagnon MP, Pollender H, Têtu B, Tanguay F. The challenges of a complex and innovative telehealth project: a qualitative evaluation of the eastern Quebec Telepathology Network. Int J Health Policy Manag 2017;7(5):421-432. [doi: 10.15171/ijhpm.2017.106]
12. Allana S, Clark A. Applying meta-theory to qualitative and mixed-methods research: a discussion of critical realism and heart failure disease management interventions research. Int J Qual Methods 2018;17(1). [doi: 10.1177/1609406918790042]
13. Allana S, Thompson DR, Ski CF, Clark AM. Intersectionality in heart failure self-care: ignorance is not an option. J Cardiovasc Nurs 2020;35(3):231-233. [doi: 10.1097/JCN.0000000000000661] [Medline: 32282518]
14. Allana S, Ski CF, Thompson DR, Clark AM. Intersectionality and heart failure: what clinicians and researchers should know and do. Curr Opin Support Palliat Care 2021 Jun 1;15(2):141-146. [doi: 10.1097/SPC.0000000000000547] [Medline: 33905386]
15. Al-Faham H, Davis AM, Ernst R. Intersectionality: from theory to practice. Annu Rev Law Soc Sci 2019 Oct 13;15(1):247-265. [doi: 10.1146/annurev-lawsocsci-101518-042942]
16. Ghasemi E, Rajabi F, Majdzadeh R. From theory to practice of designing for diversity: applying intersectionality to improve HIV testing uptake. PLoS ONE 2025;20(10):e0335311. [doi: 10.1371/journal.pone.0335311] [Medline: 41144340]
17. Morris A, Shah KS, Enciso JS, et al. The impact of health care disparities on patients with heart failure. J Card Fail 2022 Jul;28(7):1169-1184. [doi: 10.1016/j.cardfail.2022.04.008] [Medline: 35595161]
18. Dixit RA, Ratwani RM, Bishop JA, et al. The impact of expanded telehealth availability on primary care utilization. NPJ Digit Med 2022 Sep 9;5(1):141. [doi: 10.1038/s41746-022-00685-8] [Medline: 36085158]
19. Bhatia RS, Chu C, Pang A, Tadrous M, Stamenova V, Cram P. Virtual care use before and during the COVID-19 pandemic: a repeated cross-sectional study. CMAJ Open 2021;9(1):E107-E114. [doi: 10.9778/cmajo.20200311] [Medline: 33597307]
20. Shaw J, Brewer LC, Veinot T. Recommendations for health equity and virtual care arising from the COVID-19 pandemic: narrative review. JMIR Form Res 2021 Apr 5;5(4):e23233. [doi: 10.2196/23233] [Medline: 33739931]
21. Allana S, Ski CF, Thompson DR, Clark AM. Bringing intersectionality to cardiovascular health research in Canada. CJC Open 2021 Dec;3(12 Suppl):S4-S8. [doi: 10.1016/j.cjco.2021.08.016] [Medline: 34993427]
22. Allana SM. Using intersectionality theory and critical realism for exploring heart failure telehealth interventions for vulnerable patient populations: a scoping review, intersectionality-based analysis, and a rapid realist review [Dissertation]. : University of Alberta; 2022 URL: https://ualberta.scholaris.ca/items/d798750d-fc1b-4a55-9161-7ef3bf8d27e5 [accessed 2025-05-25]
23. Schlief M, Saunders KRK, Appleton R, et al. Synthesis of the evidence on what works for whom in telemental health: rapid realist review. Interact J Med Res 2022 Sep 29;11(2):e38239. [doi: 10.2196/38239] [Medline: 35767691]
24. Brunton G, Oliver S, Thomas J. Innovations in framework synthesis as a systematic review method. Res Synth Methods 2020 May;11(3):316-330. [doi: 10.1002/jrsm.1399] [Medline: 32090479]
25. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review--a new method of systematic review designed for complex policy interventions. J Health Serv Res Policy 2005 Jul;10 Suppl 1:21-34. [doi: 10.1258/1355819054308530] [Medline: 16053581]

XSL•FO
RenderX

26. Allana S, Norris C, Hussain A, Clark A. A scoping review and intersectionality-based analysis of heart failure telehealth interventions for vulnerable populations. J Adv Nurs 2023 Nov;79(11):4097-4111. [doi: 10.1111/jan.15756] [Medline: 37409794]
27. Wong G, Greenhalgh T, Westhorp G, Buckingham J, Pawson R. RAMESES publication standards: realist syntheses. BMC Med 2013 Jan 29;11(1):21. [doi: 10.1186/1741-7015-11-21] [Medline: 23360677]
28. Mukumbang FC, Kabongo EM, Eastwood JG. Examining the application of retroductive theorizing in realist-informed studies. Int J Qual Methods 2021 Jan;20. [doi: 10.1177/16094069211053516]
29. Bakitas MA, Dionne-Odom JN, Ejem DB, et al. Effect of an early palliative care telehealth intervention vs usual care on patients with heart failure: the ENABLE CHF-PC randomized clinical trial. JAMA Intern Med 2020 Sep 1;180(9):1203-1213. [doi: 10.1001/jamainternmed.2020.2861] [Medline: 32730613]
30. Pekmezaris R, Nouryan CN, Schwartz R, et al. A randomized controlled trial comparing telehealth self-management to standard outpatient management in underserved black and Hispanic patients living with heart failure. Telemed J E Health 2019 Oct;25(10):917-925. [doi: 10.1089/tmj.2018.0219] [Medline: 30418101]
31. Dang S, Karanam C, Gómez-Orozco C, Gómez-Marín O. Mobile phone intervention for heart failure in a minority urban county hospital population: usability and patient perspectives. Telemed J E Health 2017 Jul;23(7):544-554. [doi: 10.1089/tmj.2016.0224] [Medline: 28051761]
32. Dionne-Odom JN, Ejem DB, Wells R, et al. Effects of a telehealth early palliative care intervention for family caregivers of persons with advanced heart failure: the ENABLE CHF-PC randomized clinical trial. JAMA Netw Open 2020 Apr 1;3(4):e202583. [doi: 10.1001/jamanetworkopen.2020.2583] [Medline: 32282044]
33. Riegel B, Carlson B, Glaser D, Romero T. Randomized controlled trial of telephone case management in Hispanics of Mexican origin with heart failure. J Card Fail 2006 Apr;12(3):211-219. [doi: 10.1016/j.cardfail.2006.01.005] [Medline: 16624687]
34. Srisuk N, Cameron J, Ski CF, Thompson DR. Randomized controlled trial of family-based education for patients with heart failure and their carers. J Adv Nurs 2017 Apr;73(4):857-870. [doi: 10.1111/jan.13192] [Medline: 27779768]
35. Soran OZ, Piña IL, Lamas GA, et al. A randomized clinical trial of the clinical effects of enhanced heart failure monitoring using a computer-based telephonic monitoring system in older minorities and women. J Card Fail 2008 Nov;14(9):711-717. [doi: 10.1016/j.cardfail.2008.06.448] [Medline: 18995174]
36. Krum H, Forbes A, Yallop J, et al. Telephone support to rural and remote patients with heart failure: the Chronic Heart Failure Assessment by Telephone (CHAT) study. Cardiovasc Ther 2013 Aug;31(4):230-237. [doi: 10.1111/1755-5922.12009] [Medline: 23061492]
37. Rosen D, Berrios-Thomas S, Engel RJ. Increasing self-knowledge: utilizing tele-coaching for patients with congestive heart failure. Soc Work Health Care 2016 Oct;55(9):711-719. [doi: 10.1080/00981389.2016.1191581] [Medline: 27348634]
38. Riley WT, Keberlein P, Sorenson G, et al. Program evaluation of remote heart failure monitoring: healthcare utilization analysis in a rural regional medical center. Telemed J E Health 2015 Mar;21(3):157-162. [doi: 10.1089/tmj.2014.0093] [Medline: 25025239]
39. Reading Turchioe M, Grossman LV, Baik D, et al. Older adults can successfully monitor symptoms using an inclusively designed mobile application. J Am Geriatr Soc 2020 Jun;68(6):1313-1318. [doi: 10.1111/jgs.16403] [Medline: 32157679]
40. Jaana M, Sherrard H. Rural-urban comparison of telehome monitoring for patients with chronic heart failure. Telemed J E Health 2019 Feb;25(2):101-108. [doi: 10.1089/tmj.2017.0303] [Medline: 29847242]
41. Liu J. Pervasive telemonitoring for patients living with chronic heart failure: a quantitative study of telemedicine acceptance [Dissertation]. : Capella University; 2009 URL: https://www.proquest.com/docview/305166705?pq-origsite=gscholar&fromopenview=true&sourcetype=Dissertations%20&%20Theses [accessed 2026-01-28]
42. Cabán PL. The use of home telemonitoring for heart failure management among Hispanics, non-Hispanic Blacks, and non-Hispanic Whites. Home Healthc Now 2019;37(6):345-349. [doi: 10.1097/NHH.0000000000000798] [Medline: 31688469]
43. Davis C, Bender M, Smith T, Broad J. Feasibility and acute care utilization outcomes of a post-acute transitional telemonitoring program for underserved chronic disease patients. Telemed J E Health 2015 Sep;21(9):705-713. [doi: 10.1089/tmj.2014.0181] [Medline: 25955129]
44. Sammour Y, Spertus JA, Shatla I, et al. Comparison of video and telephone visits in outpatients with heart failure. Am J Cardiol 2021 Nov 1;158:153-156. [doi: 10.1016/j.amjcard.2021.08.008] [Medline: 34470705]
45. Bakhshi S, Li X, Semenov N, et al. Congestive heart failure home monitoring pilot study in urban Denver. Annu Int Conf IEEE Eng Med Biol Soc 2011;2011:3150-3153. [doi: 10.1109/IEMBS.2011.6090859] [Medline: 22255008]
46. Heiney SP, Donevant SB, Arp Adams S, Parker PD, Chen H, Levkoff S. A smartphone app for self-management of heart failure in older African Americans: feasibility and usability study. JMIR Aging 2020 Apr 3;3(1):e17142. [doi: 10.2196/17142] [Medline: 32242822]
47. Pekmezaris R, Schwartz RM, Taylor TN, et al. A qualitative analysis to optimize a telemonitoring intervention for heart failure patients from disparity communities. BMC Med Inform Decis Mak 2016 Jun 24;16:75. [doi: 10.1186/s12911-016-0300-9] [Medline: 27343060]

XSL•FO
RenderX

48. Lefler LL, Rhoads SJ, Harris M, et al. Evaluating the use of mobile health technology in older adults with heart failure: mixed-methods study. JMIR Aging 2018 Dec 4;1(2):e12178. [doi: 10.2196/12178] [Medline: 31518257]
49. Newell MC, Strauss CE, Freier T, et al. Design and initial results of the Minneapolis Heart Institute teleheart program. Circ Cardiovasc Qual Outcomes 2017 Oct;10(10):e003904. [doi: 10.1161/CIRCOUTCOMES.117.003904] [Medline: 29021334]
50. Lyles CR, Sarkar U. Health literacy, vulnerable patients, and health information technology use: where do we go from here? J Gen Intern Med 2015 Mar;30(3):271-272. [doi: 10.1007/s11606-014-3166-5] [Medline: 25588688]
51. Butzner M, Cuffee Y. Telehealth interventions and outcomes across rural communities in the United States: narrative review. J Med Internet Res 2021 Aug 26;23(8):e29575. [doi: 10.2196/29575] [Medline: 34435965]
52. Howe LC, Leibowitz KA, Crum AJ. When your doctor "gets it" and "gets you": the critical role of competence and warmth in the patient-provider interaction. Front Psychiatry 2019;10:475. [doi: 10.3389/fpsyt.2019.00475] [Medline: 31333518]

Abbreviations

CR: critical realism
HF: heart failure
HFDMI: heart failure disease management intervention
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
RAMESES: Realist and Meta-narrative Evidence Syntheses: Evolving Standards

Edited by A Kushniruk; submitted 10.Jun.2025; peer-reviewed by L Laverty, P Dunn; revised version received 18.Sep.2025; accepted 30.Nov.2025; published 23.Feb.2026.

Please cite as:
Allana S, Norris C, Hussain A, Alaidarous N, Haider RM, Clark AM
How and Why Telehealth Interventions Improve Self-Care Among Vulnerable Groups of Patients With Heart Failure: Scoping Review and Rapid Realist Synthesis
JMIR Hum Factors 2026;13:e78859
URL: https://humanfactors.jmir.org/2026/1/e78859
doi:10.2196/78859

© Saleema Allana, Colleen Norris, Armish Hussain, Nada Alaidarous, Raza Muhammad Haider, Alexander M Clark. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 23.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Machine Learning for the Analysis of Healthy Lifestyle Data: Scoping Review and Guidelines

Tony Estrella[1,2], MSc; Lluis Capdevila[1,2], PhD; Carla Alfonso[1,2], PhD; Josep-Maria Losilla[1,3], PhD

[1]Sport Research Institute, Universitat Autònoma de Barcelona, Bellaterra, Spain

[2]Department of Basic Psychology, Universitat Autònoma de Barcelona, Edifici N, Planta 1, Cerdanyola del Vallès, Barcelona, Spain

[3]Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Barcelona, Spain

Corresponding Author:
Tony Estrella, MSc
Sport Research Institute, Universitat Autònoma de Barcelona, Bellaterra, Spain

Abstract

Background: Advances in data science and technology have transformed lifestyle research by enabling the integration of multimodal information and the generation of large-scale datasets. Despite the growing interest in machine learning (ML) within health behavior research, significant methodological gaps remain.

Objective: The study aims to systematically review the applications of supervised ML algorithms in the analysis of healthy lifestyle data, with a particular focus on the methodological approaches used. The specific objectives are to explore the types and sources of data used for health outcomes, examine the ML processes used, including explainable artificial intelligence (XAI) methods, and review the software tools used. Additionally, this review aims to provide practical guidelines to enhance the quality and transparency of future ML research in health.

Methods: Following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) recommendations, the search was conducted across PubMed, PsycINFO, and Web of Science, yielding 65 studies that met the inclusion criteria.

Results: Most studies (48/65, 74%) integrated multidomain data from physical activity, diet, sleep, and stress. Data sources were split between self-acquired data (33/65, 51%) and health repositories (32/65, 49%). Single-item measurements were common, particularly for physical activity, diet, and sleep. Although 40 of 65 studies used a multimodel approach, random forest was the most frequently applied algorithm. To improve explainability, 22 of 65 (33.84%) studies incorporated specific XAI methods, with 21 using Shapley Additive Explanation values and 1 using local interpretable model-agnostic explanations. R (R Core Team) and Python (Python Software Foundation) were the most widely used software tools, with variation in the libraries used.

Conclusions: This review highlights methodological gaps in the application of supervised ML to healthy lifestyle data. The ML workflow should span from data acquisition to explainability, using iterative steps to improve methodological rigor. Although multidomain data collection enhances the understanding of health issues related to lifestyle, representativeness remains limited due to methodological shortcomings in data acquisition. While random forest was the most commonly used algorithm, a multimodel approach is recommended for a comprehensive comparison. Lifestyle components consistently ranked among the top features in studies integrating XAI. Incorporating XAI methods into the ML pipeline can support personalized interventions, provided data collection is accurate. The R metapackage (tidymodels; Max Kuhn and Hadley Wickham) facilitates process evaluation through unified syntax, improving replicability. Methodological and reporting guidelines and a checklist are provided to enhance transparency and replicability in multidisciplinary ML research.

International Registered Report Identifier (IRRID): RR2-10.37766/inplasy2023.3.0065

(JMIR Hum Factors 2026;13:e78648) doi:10.2196/78648

KEYWORDS

machine learning; artificial intelligence; healthy lifestyle; physical activity; diet; sleep; stress; review; data analysis; XAI; explainable artificial intelligence

Introduction

There is a growing interest in understanding the effects of synergistic relationships among lifestyle behaviors and their effect on health outcomes [1,2]. Traditionally, healthy lifestyle (HL) research has primarily focused on physical activity and diet. However, recent studies increasingly include sleep and stress management as critical components of lifestyle [3,4]. For instance, stress has been shown to negatively influence physical activity, sleep, and dietary habits [5], which in turn have an

XSL•FO
RenderX

overall impact on health and well-being. This multidimensional perspective has gained attention in public health under the concept of lifestyle medicine, which incorporates physical activity, diet, sleep, and stress management as cost-effective interventions to prevent noncommunicable diseases, such as cardiovascular and metabolic diseases [6-8].

Technological advances, including wearable devices and lifelogging processes, have significantly enhanced the capability to collect multimodal, high-frequency, and ecological lifestyle data [9,10]. This wealth of data provides valuable contextual information and insights for researchers and users [11]. However, the vast amount and complexity of behavioral and physiological data expose significant analytical challenges. Traditional statistical models often struggle with the high dimensionality, heterogeneity, and nonlinearity typical of lifestyle studies. Recent progress in computational power and artificial intelligence (AI), particularly machine learning (ML), has contributed to addressing these limitations [12].

ML models are capable of analyzing complex data types and generating insights and knowledge to improve decision-making [13,14]. Furthermore, ML algorithms can flexibly handle nonlinear relationships among features and outcomes. While the boundary between classical statistics and ML is not clear, ML algorithms are recognized for their flexible data-driven approach, avoiding the imposition of a predetermined relational structure between variables [15-17]. Additionally, prioritizing algorithms that maximize generalizability to new data, often referred to as scalability in the big data context, is crucial to face new health challenges [18,19]. These characteristics make ML analysis a suitable methodology for predictive modeling and feature extraction in health-related lifestyle research.

ML models are broadly classified into supervised learning (SL) and unsupervised learning (UL). In SL, the model is trained with labeled data, where each observation has an associated response measurement, to predict known outcomes such as disease risk or behavioral adherence [19]. The goal of SL is to fit a model that can predict the response when applied to new data. When the response value is continuous, this is known as a "regression problem"; when the response is categorical, it is known as a "classification problem." In contrast, in UL models, the goal is to discover patterns rather than predict outcomes, since there is no associated response to the input, and the model seeks relationships and similarities between observations. In the health domain, where diagnosis and detection are key focuses, SL, and particularly classification tasks, are more prevalent due to their ability to evaluate these predictions [20,21]. Clinical applications of SL include triage systems, prognosis prediction, and disease classification using rapid testing [22]. Consequently, SL methods are standard in epidemiology to enhance clinical decisions based on input-output relationships [23]. Since prediction and explainability are central concerns in health research, this scoping review focuses specifically on SL methods.

Despite the growing attention to ML in health behavior research, there remain significant methodological gaps. Prior reviews have focused primarily on outcome effectiveness or AI chatbot interventions, often providing limited detail about the ML process involved [24]. A recent scoping review on ML methods used in health promotion and behavioral change found that the main interventions studied are those related to physical activity, while other crucial aspects of HL were overlooked, revealing an imbalance in this literature [25]. Similarly, Lai et al [26] reviewed the applications of large language models in exercise recommendations and physical activity, highlighting methodological limitations associated with these AI models. In sum, these studies underscore the need for a more comprehensive review to include a holistic concept of HL. Furthermore, methodological details such as data preprocessing, model evaluation, and explainability are often underreported, hindering transparency, reproducibility, and interdisciplinary collaboration.

To address the lack of explainability, explainable artificial intelligence (XAI) has emerged, which focuses on understanding AI algorithms and making them more transparent. XAI aims to provide human-understandable explanations for the decisions made by ML models [27]. In HL research, XAI can be directed to identify the set of behaviors that significantly influence health, thereby enhancing transparency and trust in AI. It is important to distinguish between interpretability and explainability in the AI context. While interpretability refers to understanding the influence of each feature in the original model, explainability involves deriving actionable human insights from the model's predictions [28]. Interpretability enables AI developers to delve into the model's decision-making to comprehend how algorithms reach their decisions. Conversely, explainability refers to the process for creating common meaning from model decisions and therefore provides human-readable explanations [28]. Therefore, reporting the explainability method used in ML projects is crucial not only to enhance the decision-making process of the end user but also to understand how lifestyle factors interact with health outcomes.

Therefore, this study aims to systematically review the applications of supervised ML algorithms in analyzing HL data, with a specific focus on the methodological aspects used in these studies, rather than their results. The specific objectives are to explore (1) the specific lifestyle data used in health outcome studies; (2) the sources and types of data subjected to analysis; (3) the characteristics of the ML models, including XAI methods; and (4) the programs and libraries used for ML implementation. Additionally, based on the findings of this scoping review, we aim to provide practical guidelines to enhance the quality and transparency of future ML research in lifestyle science. A scoping review is the type of systematized review (ie, systematic, transparent, and replicable) most appropriate for addressing these objectives [29].

Method

Overview

To maximize the reporting quality of this scoping review, we followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) recommendations [30] (checklist provided in Checklist 1). The protocol for this scoping review was registered with the International Platform of Registered Systematic Review and

XSL•FO
RenderX

Meta-Analysis Protocols (INPLASY) [31]. All data generated in this review are provided in Multimedia Appendix 1 and are accessible in the institutional repository [32].

Search Strategy

In this scoping review, we searched for primary studies in the 3 principal health databases: PubMed (National Center for Biotechnology Information), PsycINFO (ProQuest), and Web of Science (Clarivate). The search was restricted to medical and psychological databases to capture studies directly relevant to health outcomes. Consequently, studies primarily published in engineering or computer science, which may focus on algorithm development or sensor-based data processing, were not included. The search strategy followed the PRESS (Peer Review of Electronic Search Strategies) [33] and PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses–Search extension) guidelines [34] and consisted of 2 groups of search terms referring to (1) HL and (2) ML. We also added a third group of terms preceded by the Boolean operator "NOT" to improve the specificity of the search strategy. This scoping review adopts a health-focused perspective, in which HL is treated as a multidimensional construct rather than the sum of isolated behaviors [6,35,36]. Therefore, the umbrella term HL was combined using the operator "OR" with an interaction block including (1) physical activity, (2) diet, (3) sleep, and (4) stress. This block is aligned with the multiple health behavior change and lifestyle medicine frameworks, in which the interaction between behaviors is a central construct [37,38].

The search strategy was adapted to the specific syntax of each database (Table S1 in Multimedia Appendix 1). The search was conducted on October 10, 2025, with language restrictions (English and Spanish) but without limitations on publication years.

Study Selection

Inclusion and Exclusion Criteria

Studies were included in or excluded from the review according to the following criteria provided in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria:

- Used supervised machine learning (ML) models for analyzing lifestyle data.
- Analyzed lifestyle behaviors as either inputs or outputs of the ML models.
- Used data from real individuals (not simulations).
- Published in English or Spanish.

Exclusion criteria:

- Focused on unsupervised learning (UL) without connection to supervised learning (SL) modeling.
- Focused on mathematical formulation or guidelines for implementing ML models in health.
- Used simulated data or aimed to develop a chatbot or app based on ML.
- Primarily addressed substance abuse, such as alcohol intake or smoking cessation.
- Focused exclusively on classical statistical regression algorithms, such as linear or logistic regression, which were not considered ML on their own in this review.

Justification of Exclusions

UL algorithms were excluded because they do not have an associated response to inputs, thereby lacking performance evaluation. Classical statistical regression algorithms, such as linear or logistic regression, were not considered in this review. While the boundary between classical statistics and ML is not clear, ML algorithms are recognized for their flexible data-driven approach, avoiding the imposition of a predetermined relational structure between variables [15-17]. Additionally, prioritizing algorithms that maximize generalizability to new data, often referred to as scalability in the big data context, is crucial to address new health challenges [18,19]. Consequently, studies focusing exclusively on this type of statistical algorithm were excluded. However, we acknowledge that the use of a model ensemble approach allows for the inclusion of these statistical algorithms to assess the performance of different algorithms during the evaluation step. Studies on substance abuse disorders were excluded as they involve distinct behavioral and neurobiological mechanisms that differ substantially from the domains of physical activity, diet, sleep, and stress, which are the core components of HL behaviors as defined in this review. In addition, substance abuse is categorized within the risk avoidance cluster, which is conceptually distinct from the other 4 behaviors examined in this review [3]. This distinction is well established in multiple health behavior theories, which differentiate behaviors that enhance health from those that reduce risk through avoidance. Therefore, its exclusion preserves the applicability of results to a multiple health behavior framework, as the selected behaviors are interrelated through shared psychological resources [2,39].

Two reviewers (TE and CA) independently screened titles and abstracts in the first phase and full texts in the second phase. Discrepancies were resolved by consensus, with the participation of a third reviewer (JML) when necessary. Agreement between reviewers during the selection process was analyzed by calculating Cohen κ.

XSL•FO
RenderX

Data Management

Mendeley was used as reference management software; the results of the search strategy were entered, and duplicates were merged or removed. An ad hoc checklist was used to extract information from the included papers. The checklist was divided into 5 sections:

- General information: authors, title, year, and country of affiliation.
- Methodological data: type of study, aim, year of data collection, form of data acquisition, sample, and countries represented in the data.
- Study variables: health issues, lifestyle features, and the model's input and output variables.
- Software: statistical programming language, libraries, and packages.
- Model aspects: type of problem, stages of ML analysis, ML methods, model evaluation, evaluation metrics, and XAI methods.

Strategy of Data Synthesis

The review was presented as a narrative synthesis, and the information was summarized in tables and figures. The information extracted from the studies was divided into 3 blocks: type of data, ML process, and software. For data extraction, we focused on lifestyle components, health outcomes, data sources, acquisition methods, and data typology. Regarding the ML process, we focused on the whole process, consisting of preprocessing, modeling, validation, evaluation, and XAI methods. For model evaluation, only the procedures used for final performance assessment were extracted. When studies reported cross-validation, we classified it as the final performance estimation method unless authors explicitly stated its use for hyperparameter optimization. To identify the top-ranked lifestyle components in the XAI analysis, we systematically examined the figures and tables reported in each study. In this review, the term ranked refers to features that were highlighted by the XAI algorithm. In contrast, unranked indicates that appeared in the model but were not reported in the XAI visualization, while not available denotes features that were not included in the ML model. Software used in each study was also recorded.

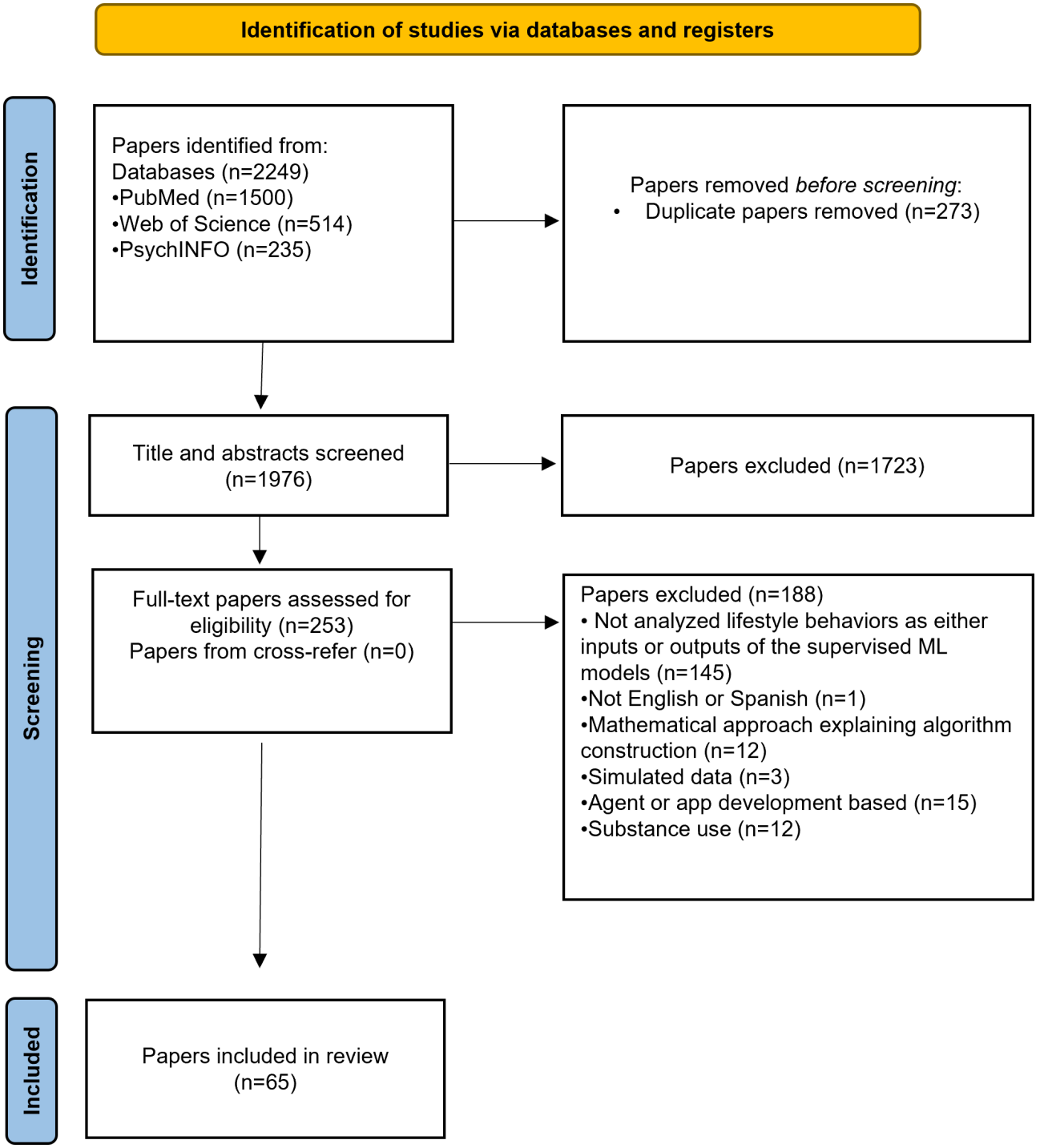
Results

Overview

A total of 2249 papers were retrieved from the databases, and 65 studies met the eligibility criteria and were included in this scoping review (refer to Figure 1). There was very good agreement between reviewers during the selection process: 96% (n=52; κ=0.84, 95% CI 0.61-1.0) in the title and abstract screening and 94% (n=35; κ=0.88, 95% CI 0.71-1.0) in the full-text screening.

XSL•FO
RenderX

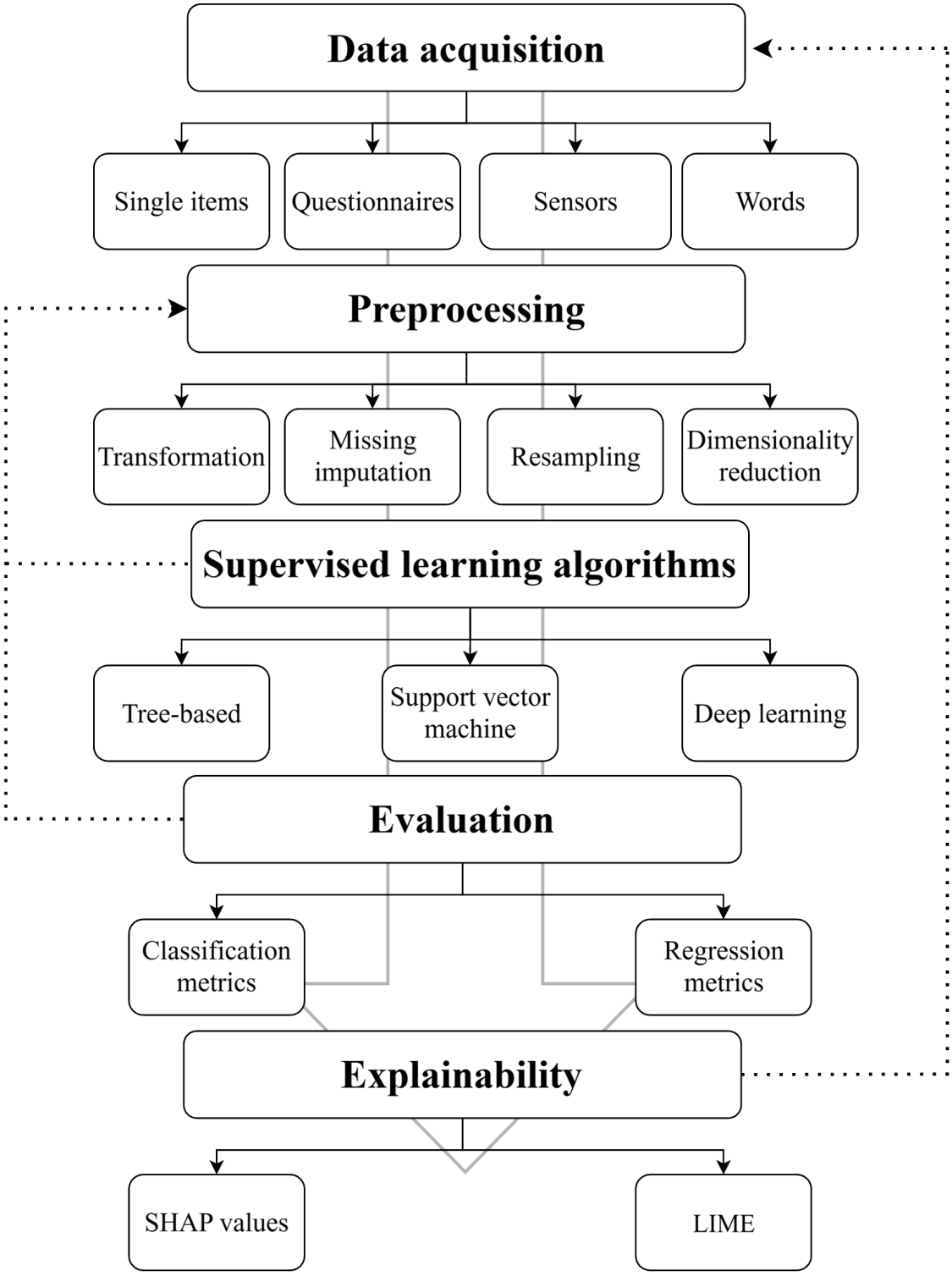
Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the scientific literature search and selection. ML: machine learning.



From this point forward, the "Results" section is structured following the ML workflow depicted in Figure 2, which illustrates the 5 key steps in the ML pipeline. The process begins with data acquisition, followed by preprocessing to prepare the data. Then, SL algorithms are applied and evaluated to determine their effectiveness. Finally, explainability techniques are used to understand the models. The dashed lines indicate that modeling, evaluation, and explainability can improve earlier stages, making the process iterative. Each stage of the process corresponds to a subsection. Finally, we examined the software used throughout the entire process in the included studies.

XSL•FO
RenderX

Figure 2. Overview of the machine learning workflow, spanning from data acquisition to explainability. The dashed lines represent iterative feedback loops within the process. LIME: local interpretable model-agnostic explanations; SHAP: Shapley Additive Explanation.

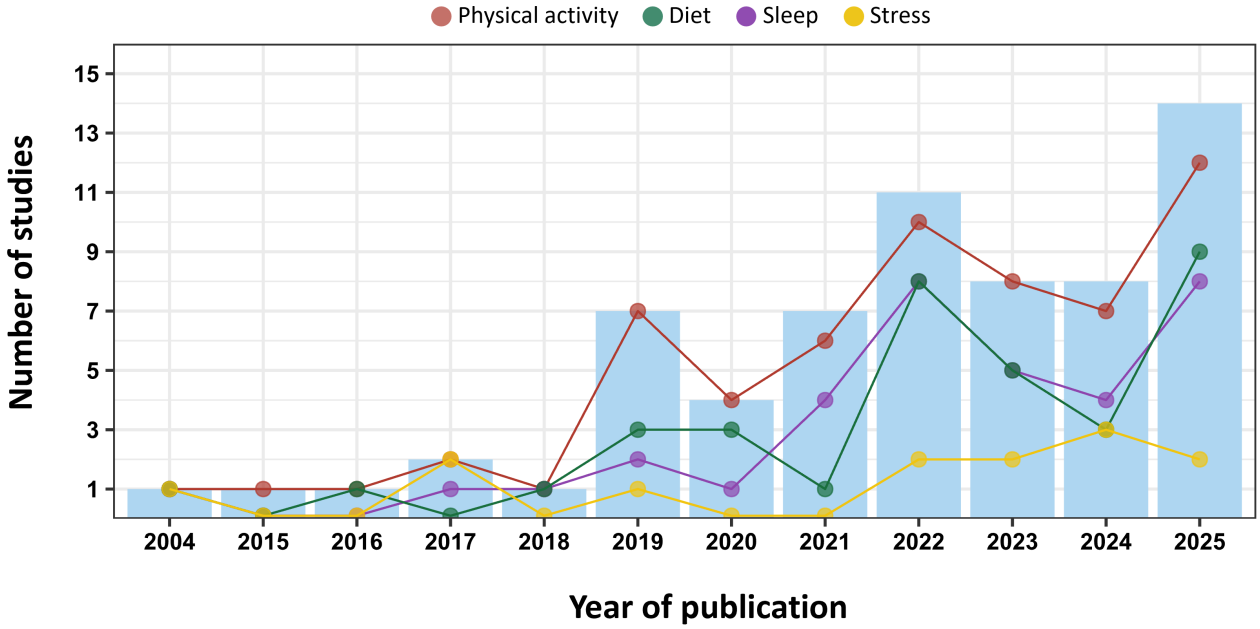


Data Acquisition: Collection Modes, Data Typology, Lifestyle Variables, and Health Outcomes

The 65 papers included in this review were published between 2004 and 2025, with 57 (87.7%) published since 2019. Figure 3 shows the annual productivity output stratified by lifestyle components. The studies were carried out in several geographical regions across 4 continents (Table S2 and Figure S1 in Multimedia Appendix 1). The mean sample size was 29,905.14 participants, with the smallest study including 8 participants and the largest including 470,778 participants.

XSL•FO
RenderX

Figure 3. Bar graph showing productivity by publication year, with each line and dot representing a lifestyle component.



Four lifestyle domains were identified in the review: physical activity, diet, sleep, and stress. Most studies (48/65, 74%) integrated data from more than 1 lifestyle domain (refer to Table 1). The most studied component was physical activity, featured in 61 of 65 studies. Diet and sleep appeared in 33 of 65 (51%) and 34 of 65 (52%) studies, respectively, while stress appeared in 15 of 65 (23%) studies (refer to Figure 3). Only 17 of 65 (26%) studies focused exclusively on a single domain (15 on physical activity [40-54] and 2 on sleep [55,56]).

Regarding data sources, 33 of 65 (51%) studies relied on self-acquired data, while 32 of 65 (49%) studies used either private or public health datasets, such as UK Biobank [57]. Among the studies using self-acquired data, the mean sample size was 10,503.41 participants. Six studies focused exclusively on females [46,47,58-61], and 27 studies reported a female proportion ranging from 41% to 89.4%. Studies using health repositories exhibited a greater mean sample size of 80,406.38 participants. Nineteen studies reported a female proportion ranging from 41.33% to 70%, 12 did not report sex distribution, and 1 study focused exclusively on females.

Figure 4 summarizes the different data collection modes used for each of the 4 lifestyle domains identified (Table S3 in Multimedia Appendix 1 provides a detailed description of each measure). Single-item measures were used in 40 of 65 (61.5%) studies assessing physical activity, 18 of 65 (27.7%) studies assessing diet, and 22 of 65 (33.8%) studies assessing sleep. Despite the heterogeneity of these items, distinct categories emerged within each lifestyle domain.

For physical activity, the emerging categories included engagement in physical activities [47,50-52,54,58,62-66], intensity [10,42,49,57,67-71], frequency [43,44,60,72-86], and environmental factors [87,88]. Within the diet domain, categories included frequency of consumption [58,60,66,72,84], types of products [75,76,80,82], environmental factors [87], and consumption habits [59,62,65,68,70,71,86,89]. Regarding sleep, categories included sleep duration [57,60,62,66-68,70,72,77-79,81,82,84,85,87], perceived sleep quality [63,89], and sleep-related problems [69,80]. Finally, within the stress domain, the emerging category focused primarily on stress level [10,57,79,82,86,89,90].

Questionnaires were used to collect physical activity data in 13 studies. The standardized questionnaires included the Global Physical Activity Questionnaire (GPAQ) [53,72,91], the International Physical Activity Questionnaire (IPAQ) [92], the Lifetime Total Physical Activity Questionnaire (exercise and sport subscale) [93], the Nutritional and Social Healthy Habits (NutSo-HH) scale [94], the Indian Migration Study Physical Activity Questionnaire (IMS-PAQ) [95], physical fitness test [41], the Short Questionnaire to Assess Health-Enhancing Physical Activity (SQUASH) [73,74], the physical activity scale from the Active Living Index [88], the Pregnancy Physical Activity Questionnaire [61], and the Physical Activity Scale for the Elderly [96]. For diet assessment, 12 studies used the Food Frequency Questionnaire (FFQ) [57,73,74,83,88,91,92,95], the PrimeScreen questionnaire [61], the Mini Nutritional Assessment [97], the NutSo-HH scale [94], and a nonstandardized questionnaire consisting of items from different questionnaires [93]. The standardized questionnaires used to measure sleep were the Pittsburgh Sleep Quality Index [10,56,61,86,91,92,96,97], the NutSo-HH scale [94], the Munich Chronotype questionnaire and Sleep Disturbance Scale for Children [64], and the Epworth Sleepiness Scale [96]. To measure stress, the stress subscale of the Depression Anxiety Stress Scale (DASS) [92,93], the INTERHEART stress questionnaire [59], the Psychosocial Well-being Index-Short Form [91], the Perceived Stress Scale [61,64,88], and the Profile of Mood States [81] were used. Regarding data collection through sensors, most studies used wearable devices. One study used a smartphone to obtain points of interest related to physical activity and diet [98], and 1 sleep study used polysomnography [55]. Finally, 2 studies used words related to physical activity and diet, 1 derived from Google Trends [99] and the other from Twitter (Twitter, Inc) [100].

XSL•FO
RenderX

Concerning the modeled inputs, 56 of 65 (86.1%) studies used multimodal data. The input modalities were lifestyle (60/65, 92.3%), sociodemographic (49/65, 75.3%), clinical (29/65, 44.6%), anthropometric (14/65, 21.5%), psychological (20/65, 30.7%), physical (3/65, 4.62%), environmental (10/65, 15.38%), physiological (3/65, 4.62%), and behavioral (2/65, 3.07%). The model outcomes included lifestyle domains in 14 (22% studies; 5 physical activity [40,44-46,90], 5 sleep [10,48,86,97,101], 3 diet [62,66,94], and 1 stress [61]) and other health outcomes in 51 (75% studies; with mental health, cancer, cardiovascular diseases, and diabetes being the most frequent categories; refer to Table 1).

Cross-sectional data were acquired in 36 (55%) studies [42-44,51-53,56-58,60-62,65,66,71-74,77-80,84,85,89,92,94,95,97], longitudinal data in 18 (28%) [10,41,50,54,55,59,64,67-69,81,82,96] studies, time-series data in 7 (11%) [45,46,48,75,76,101,102] studies, combined longitudinal and time-series data in 1 study [90], textual data in 2 studies [99,100], and combined cross-sectional and geographical data in 1 study [103].

XSL•FO
RenderX

Table . Summary of the machine learning (ML) workflow from data acquisition to explainability in the included studies.

Study	Physical activity	Diet	Sleep	Stress	Health outcome	Preprocess	ML[a] algorithm	Model evaluation	XAI[b]
Abdul Rahman et al [72]	Questionnaire (standardized) and single items (frequency)	Single items (frequency)	Single items (sleep hours)	—	Mental health	Missing imputation, Resampling, and dimensionality reduction	RF[c], ANN[d], NB[e], and KNN[f]	Hold-out test set	—
Afrash et al [58]	Single items (engagement)	Single items (frequency)	—	—	Cancer	Transformation, missing imputation, and dimensionality reduction	DT[g], MLPNN[h], RBFNN[i], FNN[j], PNN[k], and KNN	10-fold cross-validation (final evaluation)	—
Ai et al [93]	Questionnaire (standardized) and single items (frequency)	Questionnaire (nonstandardized)	—	Questionnaire (standardized)	Alzheimer disease	Transformation, missing imputation, and dimensionality reduction	RF and SVM[l]	Nested cross-validation	—
Allen [87]	Single items (environment)	Single items (environment)	Single items (sleep hours)	—	Obesity	Missing imputation and dimensionality reduction	RF and DT	2-fold cross-validation (final evaluation)	LIME[m]
Alshuraf et al [59]	Sensor (wearable)	Single items (habits)	—	Questionnaire (standardized)	Cardiovascular disease	Transformation, missing imputation, and dimensionality reduction	RF, DT, KNN, and NB	Leave-one-out cross-validation (LOOCV)	—
Birk et al [95]	Questionnaire (standardized)	Questionnaire (standardized)	—	—	Diabetes	Resampling and dimensionality reduction	RF	Hold-out test set	—
Bôto et al [62]	Single items (engagement)	Single items (habits)	Single items (sleep hours)	—	Lifestyle (diet)	Transformation and dimensionality reduction	DT	Not reported	—
Butkevičiūtė et al [40]	Sensor (wearable)	—	—	—	Lifestyle (physical activity)	Transformation	RF	5-fold cross-validation (final evaluation)	—
Cai et al [41]	Questionnaire (standardized)	—	—	—	Successful aging	Transformation and dimensionality reduction	RF, GBM[n], and ANN	10-fold cross-validation (final evaluation)	—
Cheung et al [90]	Sensor (wearable)	—	—	Single items (stress level)	Lifestyle (physical activity)	Dimensionality reduction	RF and DT	Not reported	—
Chiang and Dey [102]	Sensor (wearable)	—	Sensor (wearable)	—	Blood pressure	Transformation, missing imputation, and dimensionality reduction	RF, GBM, MLPNN, LSTM-RNN[o], and SVM	5-fold cross-validation (final evaluation) and online weighted-resampling	—

XSL•FO
RenderX

Study	Physical activity	Diet	Sleep	Stress	Health outcome	Preprocess	ML[a] algorithm	Model evaluation	XAI[b]
Cortés-Ibañez et al [73]	Questionnaire (standardized) and single items (frequency)	Questionnaire (standardized)	—	—	Cancer	Transformation, missing imputation, resampling, and dimensionality reduction	RF and SVM	Hold-out test set	—
Cortés-Ibañez et al [74]	Questionnaire (standardized) and single items (frequency)	Questionnaire (standardized)	—	—	Cancer	Transformation, resampling, and dimensionality reduction	RF, GBM, and SVM	5-fold cross-validation (final evaluation)	—
Dianati-Nasab et al [47]	Single items (engagement)	—	—	—	Cancer	Missing imputation and dimensionality reduction	RF, DT, XGBoost[p], and ANN	10-fold cross-validation (final evaluation)	—
Faruqui et al [75]	Single items (frequency)	Single items (type of products)	—	—	Diabetes	Transformation and missing imputation	LSTM-RNN, ANN, and KNN	Hold-out test set	—
Gu et al [60]	Single items (frequency)	Single items (frequency)	Single items (sleep hours)	—	Infertility risk in women	Dimensionality reduction	RF, DT, BoostDT[q], LightGBM[r], and AdaBoost[s]	Hold-out test set	SHAP[t] values
Guthrie et al [76]	Single items (frequency)	Single items (type of products)	—	—	Cardiometabolic disease	—	RF	Leave-one-out cross-validation (LOOCV)	SHAP values
Hu et al [77]	Single items (frequency)	—	Single items (sleep hours)	—	Cardiovascular disease	Missing imputation and dimensionality reduction	RF and BART[u]	Not reported	—
Hu et al [78]	Single items (frequency)	—	Single items (sleep hours)	—	Cardiovascular disease	Missing imputation and dimensionality reduction	RF, XGBoost, and BART	5-fold cross-validation (final evaluation)	—
Huang et al [63]	Single items (engagement)	—	Single items (sleep quality)	—	Cognitive function	Missing imputation and resampling	RF, BoostDT, XGBoost, and LSTM-RNN	Hold-out test set	SHAP values
Jin and Halili [67]	Single items (intensity)	—	Single items (sleep hours)	—	Mental health	Transformation, missing imputation, resampling, and dimensionality reduction	RF, DT, XGBoost, LightGBM, CatBoost[v], Bagging[w], HistGBM[x], SVM, and MLPNN	Hold-out test set	SHAP values
Kim et al [91]	Questionnaire (standardized)	Questionnaire (standardized)	Questionnaire (standardized) and Single items (Sleep hours)	Questionnaire (standardized)	Quality of life	Transformation, Resampling	RF, DT, XGBoost, SVM, NB, KNN	6-fold cross-validation (final evaluation)	SHAP values

XSL•FO
RenderX

Study	Physical activity	Diet	Sleep	Stress	Health outcome	Preprocess	ML[a] algorithm	Model evaluation	XAI[b]
Kimura et al [104]	Sensor (wearable)	—	Sensor (wearable)	—	Alzheimer disease	Dimensionality reduction	SVM	5-fold cross-validation (final evaluation)	—
Kiss et al [64]	Single items (engagement)	—	Questionnaire (standardized) and single items (sleep hours)	Questionnaire (standardized)	Mental health	Transformation, missing imputation, and dimensionality reduction	XGBoost	Nested cross-validation	SHAP values
Li and Song [50]	Single items (engagement)	—	—	—	Cognitive function	Transformation	CNN[y], Transformer, LSTM-RNN, GRU-Attention[z], WaveNet, and RNN[aa]	10-fold cross-validation (final evaluation)	SHAP values
Lim et al [42]	Single items (intensity)	—	—	—	Osteoarthritis	Transformation, missing imputation, resampling, and dimensionality reduction	FFNN[ab]	Hold-out test set	—
Lim et al [10]	Single items (intensity)	—	Questionnaire (standardized) and single items (sleep quality)	Single items (Stress level)	Lifestyle (sleep)	Transformation, missing imputation, and dimensionality reduction	RF and DT	Hold-out test set	—
Lin et al [51]	Single items (engagement)	—	—	—	Loneliness	Transformation, Missing imputation, Dimensionality reduction	RF, DT, SVM, MLP, and KNN	10-fold cross-validation (final evaluation)	SHAP values
Liu et al [53]	Questionnaire (standardized)	—	—	—	Cardiovascular disease	Transformation	RSF[ac]	Hold-out test set	—
Luo et al [79]	Single items (frequency)	—	Single items (sleep hours)	Single items (stress level)	Social network addiction risk	Transformation and dimensionality reduction	RF	Hold-out test set	—
Luo et al [57]	Single items (intensity)	Questionnaire (standardized)	Single items (sleep hours)	Single items (stress level)	Chronic kidney disease	Missing imputation and dimensionality reduction	GBM	Hold-out test set	—
Luo et al [68]	Single items (intensity)	Single items (habits)	Single items (sleep hours)	—	Frailty	Missing imputation	XGBoost	10-fold cross-validation (final evaluation)	SHAP values
Majcherek et al [80]	Single items (frequency)	Single items (type of products)	Single items (sleep problems)	—	Mental health	Missing imputation	XGBoost	Not reported	SHAP values

XSL•FO
RenderX

Study	Physical activity	Diet	Sleep	Stress	Health outcome	Preprocess	ML[a] algorithm	Model evaluation	XAI[b]
Majcherek et al [65]	Single items (engagement)	Single items (habits)	—	—	Diabetes	Resampling	RF, DT, AdaBoost, CatBoost, HistGBM, LightGBM, XGBoost, KNN, NB, and Nearest Centroid	Hold-out test set	SHAP values
Matta et al [48]	Sensor (wearable)	—	—	—	Lifestyle (sleep)	Transformation	MLP	Hold-out test set	—
Moon and Woo [89]	—	Single items (habits)	Single items (sleep quality)	Single items (stress level)	Mental health	Transformation, missing imputation, resampling, and dimensionality reduction	RF and ANN	Not reported	—
Morris et al [88]	Questionnaire (standardized) and single items (environment)	Questionnaire (standardized) and single items (environment)	—	Questionnaire (standardized)	Cardiovascular disease	Missing imputation	RF and ANN	10-fold cross-validation (final evaluation)	SHAP values
Mousavi et al [66]	Single items (engagement)	Single items (frequency)	Single items (sleep hours)	—	Lifestyle (diet)	Dimensionality reduction	FFNN	Hold-out test set	—
Mun and Geng [81]	Single items (frequency)	—	Single items (sleep hours)	Questionnaire (standardized) and sensor (wearable)	Fatigue	Transformation, missing imputation, and dimensionality reduction	RF	10-fold cross-validation (final evaluation)	—
Nichols et al [61]	Questionnaire (standardized)	Questionnaire (standardized)	Questionnaire (standardized)	Questionnaire (standardized)	Lifestyle (stress)	Transformation, missing imputation, resampling, and dimensionality reduction	SVM	Hold-out test set	—
Oladeji et al [99]	Words (Google Trends)	Words (Google Trends)	—	—	Obesity	Dimensionality reduction	RF, GBM, and SVM	Out-of-sample	—
Park et al [49]	Single items (intensity)	—	—	—	Adverse health event	Dimensionality reduction	XGBoost	Hold-out test set	—
Park and Edington [82]	Single items (frequency)	Single items (type of products)	Single items (sleep hours)	Single items (stress level)	Diabetes	Missing imputation and resampling	MLPNN	Hold-out test set	—
Park [83]	Single items (frequency)	Questionnaire (standardized)	—	—	Visceral fat	Transformation, missing imputation, and dimensionality reduction	RF, XGBoost, and ANN	Hold-out test set	SHAP values
Pereira et al [92]	Questionnaire (standardized)	Questionnaire (standardized)	Questionnaire (standardized)	Questionnaire (standardized)	Mental health	Transformation and missing imputation	RF, XGBoost, and SVM	10-fold cross-validation (final evaluation)	—

XSL•FO
RenderX

Study	Physical activity	Diet	Sleep	Stress	Health outcome	Preprocess	ML[a] algorithm	Model evaluation	XAI[b]
Puterman et al [69]	Single items (intensity)	—	Single items (sleep problems)	—	Mortality	Transformation, missing imputation, and dimensionality reduction	RSF	Hold-out test set	—
Qasrawi et al [70]	Single items (intensity)	Single items (habits)	Single items (sleep hours)	—	Mental health	Missing imputation and dimensionality reduction	RF, DT, XGBoost, SVM, ANN, KNN	10-fold cross-validation (final evaluation)	—
Recenti et al [43]	Single items (frequency)	—	—	—	Lifestyle (physical activity)	Missing imputation and resampling	RF, GBM, and AdaBoost	10-fold cross-validation (final evaluation)	—
Recenti et al [44]	Single items (frequency)	—	—	—	Diabetes	Missing imputation and resampling	RF, GBM, and AdaBoost	10-fold cross-validation (final evaluation)	—
Ren et al [52]	Single items (engagement)	—	—	—	Cognitive function	Missing imputation, Resampling	RF, XGBoost, SVM	Hold-out test set	SHAP values
Ruiz et al [54]	Single items (engagement)	—	—	—	Depression	Not reported	DT	Not reported	—
Sandri et al [94]	Questionnaire (standardized)	Questionnaire (standardized)	Questionnaire (standardized)	—	Lifestyle (diet)	Transformation and resampling	RF, DT, XGBoost, CatBoost, HistGBM, and FFNN	Hold-out test set	SHAP values
Sathyanarayana et al [101]	Sensor (wearable)	—	Sensor (wearable)	—	Lifestyle (sleep)	Missing imputation	MLPNN, CNN, SETRNN[ad], and LSTM-RNN	Hold-out test set	—
Shi et al [84]	Single items (frequency)	Single items (frequency)	Single items (sleep hours)	—	Osteoporosis	Transformation, missing imputation, resampling, and dimensionality reduction	RF, DT, SVM, and KNN	Hold-out test set	SHAP values
Staudenmayer et al [45]	Sensor (wearable)	—	—	—	Lifestyle (physical activity)	Transformation	RF, DT, ANN, and SVM	Leave-one-out cross-validation (LOOCV)	—
Stemmer et al [100]	Words (Twitter)	Words (Twitter)	—	—	Inflammatory bowel disease	—	RF, GBM, AdaBoost, and SVM	Hold-out test set	—
Su et al [56]	—	—	Questionnaire (standardized)	—	Resilience	Dimensionality reduction	RF, DT, and XGBoost	Not reported	SHAP values
Wallace et al [96]	Questionnaire (standardized)	—	Questionnaire (standardized) and single items (sleep hours)	—	Mortality	—	RSF	Not reported	—

XSL•FO
RenderX

Study	Physical activity	Diet	Sleep	Stress	Health outcome	Preprocess	ML[a] algorithm	Model evaluation	XAI[b]
Wallace et al [55]	—	—	Single items (sleep hours) and sensor (polysomnography)	—	Mortality	Missing imputation and dimensionality reduction	RF	External dataset	—
Wang et al [97]	—	Questionnaire (standardized)	Questionnaire (standardized)	—	Lifestyle (sleep)	Missing imputation, resampling, and dimensionality reduction	GBM, LightGBM, SVM, MLPNN, and KNN	10-fold cross-validation (final evaluation)	SHAP values
Xin and Ren [85]	Single items (frequency)	—	Single items (Sleep hours)	—	Mental health	Dimensionality reduction	RF	Hold-out test set	SHAP values
Zhang et al [86]	Single items (frequency)	Single items (habits)	Questionnaire (standardized)	Single items (stress level)	Lifestyle (sleep)	Resampling	RF, DT, XGBoost, SVM, ANN, and KNN	External dataset	SHAP values
Zhou et al [46]	Sensor (wearable)	—	—	—	Lifestyle (physical activity)	Transformation	SVM	Out-of-sample	—
Zhou et al [98]	Sensor (phone)	Sensor (phone)	—	—	Obesity	Missing imputation and dimensionality reduction	RF, GRF[ae], and ANN	10-fold cross-validation (final evaluation)	—

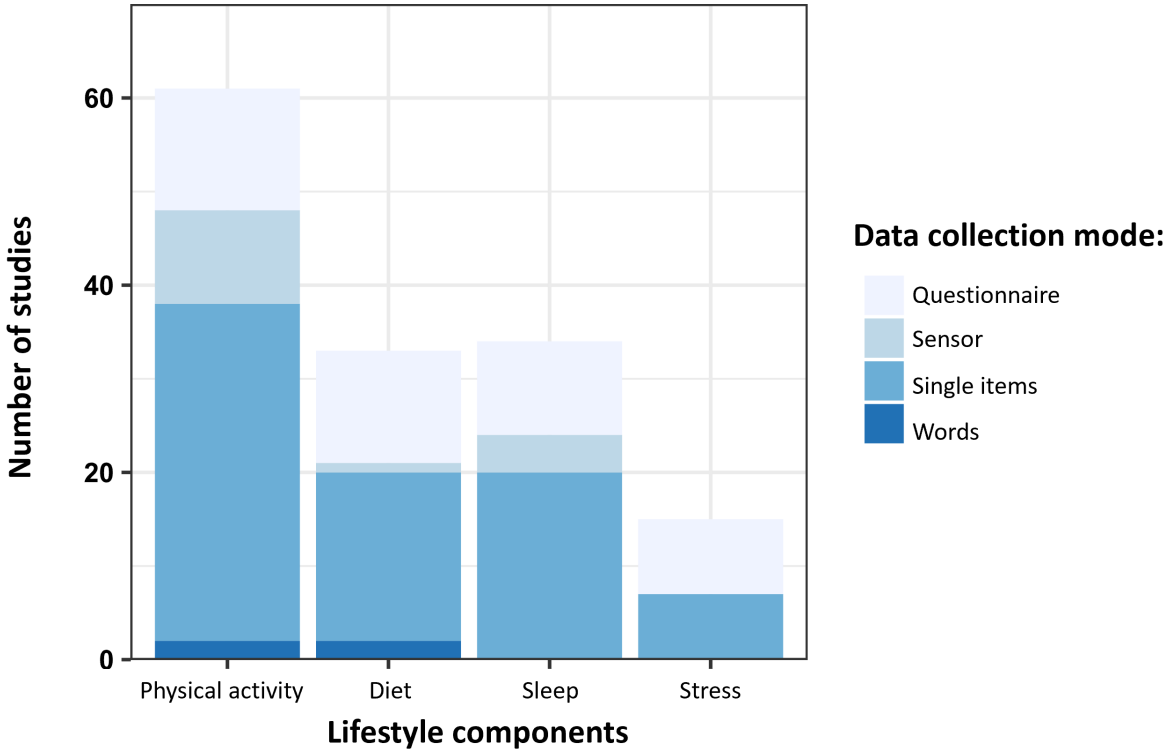
XSL•FO
RenderX

Study	Physical activity	Diet	Sleep	Stress	Health outcome	Preprocess	ML[a] algorithm	Model evaluation	XAI[b]
Zhou et al [71]	Single items (intensity)	Single items (habits)	—	—	Psoriasis	Resampling	XGBoost	Hold-out test set	SHAP values

[a]ML: machine learning.
[b]XAI: explainable artificial intelligence.
[c]RF: random forest.
[d]ANN: artificial neural network.
[e]NB: naive Bayes.
[f]KNN: k-nearest neighbor.
[g]DT: decision tree.
[h]MLPNN: multilayer perceptron neural network.
[i]RBFNN: radial basis function neural network.
[j]FNN: fuzzy neural network.
[k]PNN: probabilistic neural network.
[l]SVM: support vector machine.
[m]LIME: local interpretable model-agnostic explanations.
[n]GBM: gradient boosting.
[o]LST-RNN: long short-term memory recurrent neural network.
[p]XGBoost: extreme gradient boosting.
[q]BoostDT: boost decision tree.
[r]LightGBM: light gradient boosting machine.
[s]AdaBoost: adaptive boosting.
[t]SHAP: Shapley Additive Explanations.
[u]BART: Bayesian additive regression trees.
[v]CatBoost:categorical boosting.
[w]Bagging: bootstrap aggregating.
[x]HistGBM: histogram-based gradient boosting machine.
[y]CNN: convolutional neural network.
[z]GRU-Attention: gated recurrent unit with attention.
[aa]RNN: recurrent neural network.
[ab]FFNN: feed-forward neural network.
[ac]RSF: random survival forest.
[ad]SETRNN: simple Elman-type recurrent neural network.
[ae]GRF: generalized random forest.

XSL•FO
RenderX

Figure 4. Stacked bar chart summarizing the data acquisition methodology for each lifestyle component.



Preprocessing

The preprocessing phase was divided into variable transformation, missing imputation, resampling, and dimensionality reduction. At least one of these preprocessing phases was reported by 59 (90.76%) studies (refer to Table 1 and Table S4 in Multimedia Appendix 1 for more details).

There were 26 (40%) studies that reported normalization or other arithmetic or statistical transformations of variables before the modeling phase [40,41,45,48,50,51,53,58,59,62,64,67,69,73-75,79,81,83,84,89,91-94,102]. Six (9.23%) studies recoded categorical variables into quantitative variables, 5 used one-hot encoding [10,53,64,67,102], and 1 used principal component analysis with quantile transformer scaler [42].

Missing data imputation was reported in 37 (56.92%) studies. Twelve papers simply removed cases with missing data [10,42-44,61,64,77,78,80,93,98,101], while others applied cutoff percentages for missing values (eg, 10% [87], 30 % [73], or >50% [58,72]), and 1 study removed observations with missing values in the output [102]. Techniques included single imputation (mean, median, or mode) [51,52,58,59,70,83,88,92], multiple imputation by chained equations [57,63,72,73], k-nearest neighbor [89,102], regression-based algorithm [57,81], random forest (RF)–based multiple imputation [55,69,84], the MissForest algorithm [67,68], imputation based on peers with similar health profile group [82], imputation using training data [104], and replacement of missing values with the last available data [75].

Resampling techniques were reported in 21 (32.30%) papers. Eighteen studies balanced the datasets using methods such as using the minority class as a reference, undersampling the majority class [42,61], or the synthetic minority oversampling technique (SMOTE) [43,44,52,60,63,67,71,72,84,86,89,91,94,95,97]. One paper compared the results of SMOTE against the adaptive synthetic algorithm [65]. Finally, 1 study stabilized variations in underrepresented outcome classes using bootstrap resampling [82]. Regarding cancer studies where cases were fewer than controls, 2 different strategies were applied to the same dataset: sample-size equalization by randomly grouping cancer-free participants based on the number of cancer survivors [74], while another study matched cases and controls by sex, age, and education level, then selected a random sample resulting in 50% cases and 50% controls [73].

Dimensionality reduction was used in 38 (58.46%) studies using 3 approaches. The first approach involved assessing the relationship between features and outcomes by removing redundant information [55,61,64,70,83,85,87,93,98,99]. Other methods included factor analysis [10] and principal component analysis [42,59,69]. The second approach optimized models to achieve lower prediction error [58,66,104]. The third approach involved automatic selection of predictors during model training [41,49,57,62,72-74,77,78,81,90,95,102]. For more information, refer to Table S4 in Multimedia Appendix 1.

SL Models

Models Overview

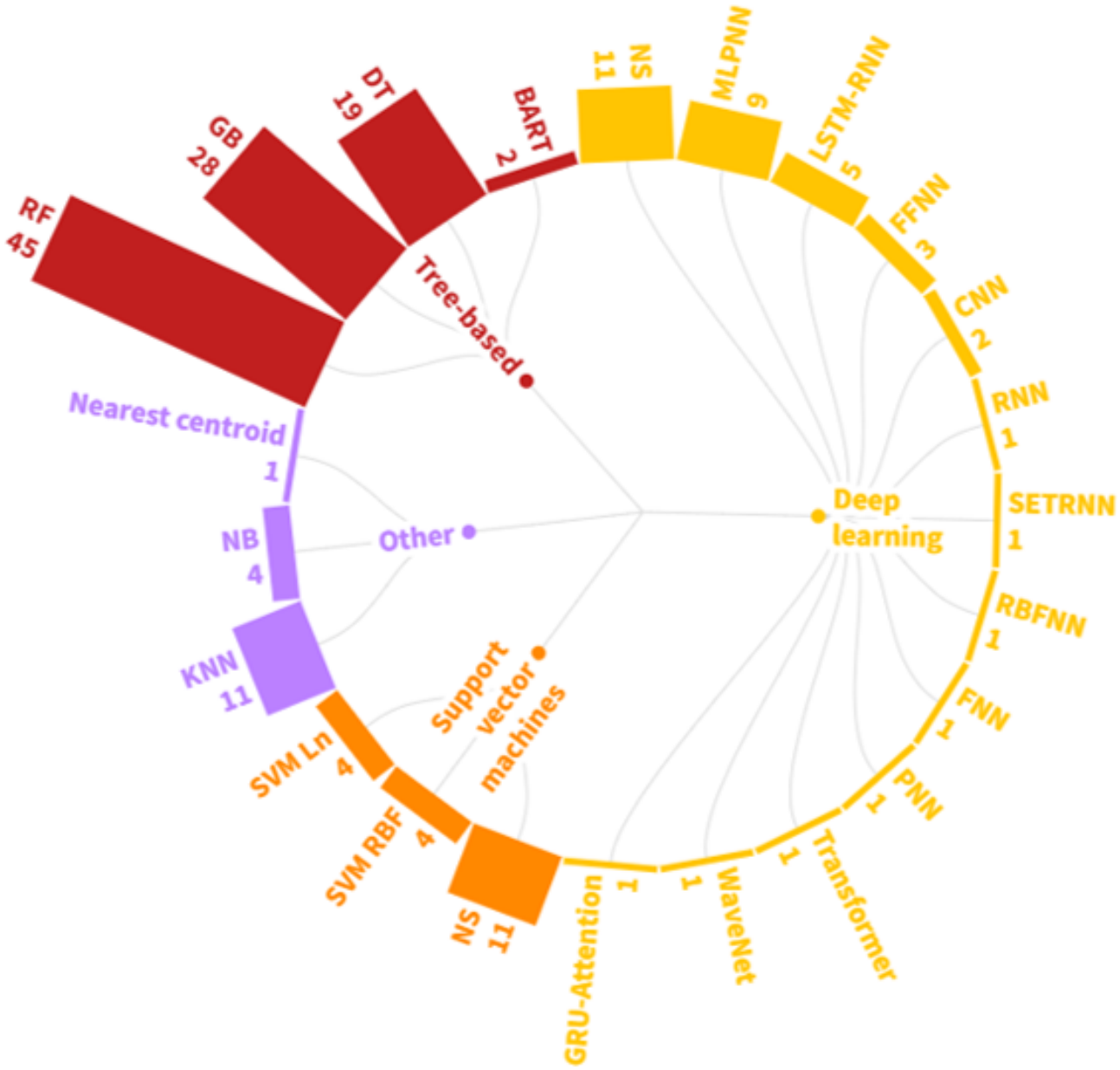
In this section, the core components of ML models are described, beginning with problem formulation and algorithm families, followed by evaluation components. Depending on the purpose of the ML analysis, papers were grouped as classification or regression when the objective was prediction, and as feature selection when the goal was explanation [19]. Most studies (46/65, 70.77%) focused on classification, 9

XSL•FO
RenderX

(13.83%) on regression, 2 (3.1%) on both classification and regression, and 8 (12.30%) on feature selection.

Six families of algorithms emerged from the studies: tree-based, deep learning, support vector machines, k-nearest neighbors, naïve Bayes, and nearest centroid. The approach adopted across 40 studies was a multimodel approach, with 29 of them incorporating algorithms from diverse algorithmic families. Figure 5 provides a comprehensive taxonomy of the specific algorithms implemented. The following subsections describe the application of specific algorithms within the 3 most used families (tree-based, deep learning, and support vector machines) in relation to the type of data used.

Figure 5. Machine learning families and algorithms taxonomy. BART: Bayesian additive regression trees; CNN: convolutional neural network; DT: decision tree; FFNN: feed-forward neural network; FNN: fuzzy neural network; GB: gradient boosting; GRU-attention: gated recurrent unit with attention; KNN: k-nearest neighbor; LSTM-RNN: long short-term memory recurrent neural network; MLPNN: multilayer perceptron neural network; NB: naive Bayes; NS: not specified; PNN: probabilistic neural network; RBFNN: radial basis function neural network; RF: random forest; SETRNN: simple Elman-type recurrent neural network; SVM Ln: support vector machine with linear kernel; SVM RBF: support vector machine with a radial basis function.



Tree-Based Algorithms

Tree-based algorithms were applied in 55 (84.61%) studies, covering all data types. RF was used in 45 (69.23%) out of 55 studies. Specifically, RF was implemented in 27 cross-sectional studies [40,43,44,47,51-53,56,60,65,70,72-74,77-79,84-87,89,91-95], 11 longitudinal studies [10,41,55,59,63,67,69,81,83,88,96], 3 time-series studies [45,76,102], 2 textual studies [99,100], 1 study with both cross-sectional and geographical data [98], and 1 study with both longitudinal and time-series data [90].

Different versions of the gradient boosting algorithm were performed in 28 studies, including gradient boosting machines, extreme gradient boosting, adaptive boosting, and light gradient boosting machine

XSL•FO
RenderX

[41,43,44,47,49,52,56,57,60,63-65,67,68,70,71,74,78,80,83,86,91,92,94,97,99,100,102]. Finally, decision tree algorithms were implemented in 19 studies [10,45,47,51,54,56,58-60,62,65,67,70,84,86,87,90,91,94], and Bayesian additive regression trees in 2 studies [77,78] (Table 1; Figure 5).

Deep Learning Algorithms

Neural networks (NNs) are considered the cornerstone of deep learning algorithms. Various NN architectures were applied in 24 (36.92%) of the reviewed studies. A multilayer perceptron neural network was used in cross-sectional [51,58,97], longitudinal [67,82], and time-series data [48,101,102]. Long short-term memory recurrent neural network was applied to longitudinal [50,63] and time-series data [75,101,102]. Feed-forward neural networks were used for cross-sectional data [42,66,94]. Convolutional neural networks were used to analyze time-series [101] and cross-sectional data [50]. Simple Elman-type recurrent neural networks were applied in a time-series study [101]. Radial basis function neural networks, fuzzy neural networks (FNNs), and probabilistic neural networks were used in a cross-sectional study [58]. In addition, Transformer, gated recurrent unit with attention, WaveNet, and RNNs were used to analyze longitudinal data [50]. In contrast, 11 studies did not specify the artificial neural network architecture used [41,45,47,70,72,75,83,86,88,89,98] (Table 1; Figure 5).

Support Vector Machine Algorithms

Support vector machine (SVM) algorithms were used in 19 (27.08%) studies, applied across various data types. SVM was implemented in 11 cross-sectional studies, 2 longitudinal studies, 3 time-series studies, and 2 studies that analyzed textual data. Configurations included support vector machine with a radial basis function [61,70,102,104] or a support vector machine with linear kernel [51,73,93,100]. Eleven studies did not report the type of kernel used [45,46,52,67,74,84,86,91,92,97,99] (Table 1; Figure 5).

Evaluation

Final model evaluation procedures were explicitly reported in 57 (87.69%) reviewed studies, with the hold-out test set being the most applied strategy. These included a hold-out test set [10,42,48,49,52,53,57,60,61,63,65-67,69,71-73,75,79,82-85,94,95,100,101], k-fold cross-validation (final evaluation) [40,41,43,44,47,50,51,58,68,70,74,78,81,87,88,91,92,97,98,104], nested cross-validation [64,93], leave-one-out cross-validation (for small datasets with n<150) [45,59,76]. Two studies used external datasets for model performance assessment [55,86]. For time-series data, 2 studies divided the dataset based on the time of acquisition, keeping an out-of-sample dataset for model evaluation [46,99], while 1 study [102] used 5-fold cross-validation in an offline setting followed by an online weighted resampling methodology to address drift.

Among the 11 studies that addressed regression problems, the reported evaluation metrics included mean absolute error [44,75,87,102], mean squared error [44,68,75], root-mean-square error [44,45,68,78,98,99,102], mean absolute percentage error [102], and coefficient of determination (R^2) [44,68,79,81,98,99].

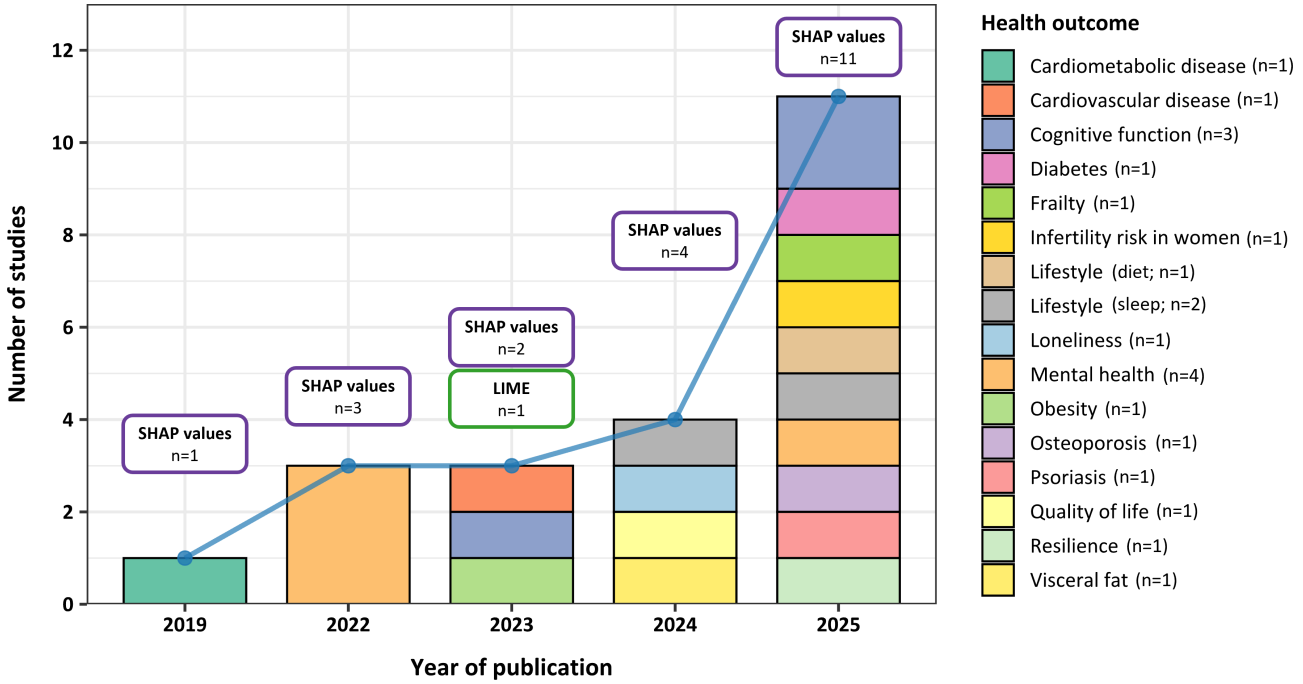
For classification problems, reported evaluation metrics included specificity and sensitivity (recall) [41-44,47,48,58,63,66,72,76,82,85,86,91,93,95,104], precision and recall [40-43,48,61,70,80,86,100,101], the confusion matrix [42,48,80], error rate as the proportion of misclassified observations (1 – accuracy) [48,58,72,90], Cohen κ [58,72], F_1-score [41,43,49,59,70,80,86,91,100,101,104], and model training time [58]. The most frequently used metrics were accuracy [10,41-45,47-49,58,59,61,63,66,70,72,80,82,83,85,86,88,91,93,101] and area under the receiver operating characteristic curve [41-43,46-49,58,59,63,70,72-74,76,83,85,86,91,93,95,100,101,104] (Table S5 in Multimedia Appendix 1). Additionally, 1 epidemiologic study used the Brier score to assess cardiovascular mortality [53].

Explainability

To enhance explainability, 22 (33.84%) studies implemented specific XAI methods to clarify the contribution of each predictor and the direction of the relationship with the outcome. Of these, 21 studies used Shapley Additive Explanations (SHAP) values, while 1 used local interpretable model-agnostic explanations (LIME). As shown in Figure 6, XAI methods, particularly SHAP values, have been applied across different health domains, showing a remarkable increase in use in 2025. The most common visualization methods were beeswarm and bar plots, with the contribution of each feature ranked along the y-axis, placing the most important feature at the top of the plot. The number of features plotted ranges from 6 to 50, with 20 being the most common (refer to Table 2).

XSL•FO
RenderX

Figure 6. Bar plot depicting the application of explainable artificial intelligence (XAI) algorithms in different years per health outcome (N=22). LIME: local interpretable model-agnostic explanations; SHAP: Shapley Additive Explanations.



XSL•FO
RenderX

Table . Summary of explainable artificial intelligence use by health outcome and healthy lifestyle components.

Study	Health outcome	Total number of features (features ranked)	Healthy lifestyle components (category, position, features ranked)	XAI[a] (visualization)
Allen [4]	Obesity	64 (10)	Physical activity (physical inactivity, 1, 10); diet (food insecurity, unranked, n/a); sleep (sleep hours, unranked, n/a); stress (n/ac[b], n/a[c], n/a).	LIME[d] (waterfall plot)
Gu et al [16]	Infertility risk in women	39 (10)	Physical activity (physical activity health score, unranked, n/a); diet (diet health score, 10, 10); sleep (sleep health score, unranked, n/a); stress (n/ac, n/a, n/a).	SHAP[e] values (beeswarm plot, bar plot, and dependency plot)
Guthrie et al [17]	Cardiometabolic disease	13 (13)	Physical activity (minutes of physical activity, 8, 13); diet (plant-based meal, 6, 13); sleep (n/ac, n/a, n/a); stress (n/ac, n/a, n/a).	SHAP values (beeswarm plot)
Huang et al [20]	Cognitive function	20 (20)	Physical activity (exercise, 10, 20); diet (n/ac, n/a, n/a); sleep (sleep quality, 8, 20); stress (n/ac, n/a, n/a).	SHAP values (beeswarm plot and bar plot)
Jin and Halili [21]	Mental health	21 (21)	Physical activity (intensity, unranked, n/a); diet (n/ac, n/a, n/a); sleep (sleep hours, 1, 21); stress (n/ac, n/a, n/a).	SHAP values (bar plot)
Kim et al [22]	Quality of life	21 (20)	Physical activity (transport-related, 4, 20; physical activity score, 7, 20); diet (eating index, 15, 20); sleep (sleep quality, 2, 20; sleep hours, 6, 20); stress (stress level, 1, 20).	SHAP values (beeswarm plot)
Kiss et al [24]	Mental health			SHAP values (bar plot)
		Positive affect: 231 (20)	Physical activity (doing outdoor activities, 3, 20; duration of sitting, 6, 20; frequency of walking, 7, 20); diet (n/ac, n/a, n/a); sleep (sleep disorder, 8, 20); stress (coping strategies, 9, 20).	
		Perceived stress: 228 (20)	Physical activity (engagement, 12, 20); diet (n/ac, n/a, n/a); sleep (sleep hours, 7, 20); stress (coping strategies, 5, 20).	
		Anxiety: 228 (20)	Physical activity (engagement, unranked, n/a); diet (n/ac, n/a, n/a); sleep (sleep hours, 15, 20); stress (coping strategies, 10, 20).	
		Depressive symptoms: 240 (20)	Physical activity (engagement, 14, 20); diet (n/ac, n/a, n/a); sleep (sleep hours, 16, 20); stress (coping strategies, 9, 20).	

XSL•FO
RenderX

Study	Health outcome	Total number of features (features ranked)	Healthy lifestyle components (category, position, features ranked)	XAI[a] (visualization)
Li and Song [25]	Cognitive function	20 (20)	Physical activity (sport social capital index, 3, 20); diet (n/ac, n/a, n/a); sleep (n/ac, n/a, n/a); stress (n/ac, n/a, n/a).	SHAP values (beeswarm plot, heat map, temporal analysis, and dependency plot)
Lin et al [28]	Loneliness	15 (15)	Physical activity (exercise, 7, 15); diet (n/ac, n/a, n/a); sleep (n/ac, n/a, n/a); stress (n/ac, n/a, n/a).	SHAP values (beeswarm plot)
Luo et al [33]	Frailty			SHAP values (beeswarm plot and dependency plot)
		US cohort: 121 (20)	Physical activity (play sports and exercise, 8, 20; moderate physical activity, 16, 20); diet (n/ac, n/a, n/a); sleep (sleep problems, 4, 20); stress (n/ac, n/a, n/a).	
		UK cohort: 125 (20)	Physical activity (vigorous physical activity, 4, 20; moderate physical activity, 5, 20); diet (fruit consumption, 19, 20); sleep (sleep problems, 2, 20; sleep duration, 6, 20); stress (n/ac, n/a, n/a).	
		China cohort: 94 (20)	Physical activity (n/ac, n/a, n/a); diet (n/ac, n/a, n/a); sleep (sleep problems, 5, 20; sleep duration, 8, 20); stress (n/ac, n/a, n/a).	
Majcherek et al [34]	Mental health	26 (24)	Physical activity (sport, 6, 24; walking time, 11, 24); diet (vegetable portion, 13, 24; fruit portion, 14, 24); sleep (n/ac, n/a, n/a); stress (n/ac, n/a, n/a).	SHAP values (dependency plot)
Majcherek et al [35]	Diabetes	22 (6)	Physical activity (regular physical activity, unranked, n/a); diet (habits, unranked, n/a); sleep (n/ac, n/a, n/a); stress (n/ac, n/a, n/a).	SHAP values (dependency plot)
Morris et al [38]	Cardiovascular disease	50 (50)	Physical activity (availability of outdoor activities, 5, 50); diet (available favorable food stores, 14, 50); sleep (n/ac, n/a, n/a); stress (global stress, 18, 50).	SHAP values (bar plot)
Park et al [43]	Visceral fat	32 (20)	Physical activity (frequency, unranked, n/a); diet (high rice consumption, 2, 20; Asian-style balanced diet, 6, 20); sleep (n/ac, n/a, n/a); stress (n/ac, n/a, n/a).	SHAP values (beeswarm plot and bar plot)
Ren et al [51]	Cognitive function	39 (20)	Physical activity (exercise, 16, 20); diet (n/ac, n/a, n/a); sleep (n/ac, n/a, n/a); stress (n/ac, n/a, n/a).	SHAP values (beeswarm plot, bar plot, and force plot)
Sandri et al [53]	Lifestyle (diet)			SHAP values (beeswarm plot)

XSL•FO
RenderX

Study	Health outcome	Total number of features (features ranked)	Healthy lifestyle components (category, position, features ranked)	XAI[a] (visualization)
		Mediterranean diet: 41 (20)	Physical activity (sport, 8, 20); diet (fish consumption, 1, 20); sleep (sleep quality, unranked, n/a); stress (n/ac, n/a, n/a).	
		Intermittent fasting: 41 (20)	Physical activity (sport, 3, 20); diet (fish consumption, 1, 20); sleep (sleep quality, 14, 20); stress (n/ac, n/a, n/a).	
		Vegan diet: 41 (20)	Physical activity (sport, 5, 20); diet (fish consumption, 1, 20); sleep (sleep quality, 9, 20); stress (n/ac, n/a, n/a).	
		Vegetarian diet: 41 (20)	Physical activity (sport, 5, 20); diet (fish consumption, 1, 20); sleep (sleep quality, unranked, n/a); stress (n/ac, n/a, n/a).	
Shi et al [55]	Osteoporosis	45 (20)	Physical activity (physical activity health score, 16, 20); diet (diet health score, unranked, n/a); sleep (sleep health score, 5, 20); stress (n/ac, n/a, n/a).	SHAP values (beeswarm plot, waterfall plot, and force plot)
Su et al [58]	Resilience	12 (4)	Physical activity (n/ac, n/a, n/a); diet (n/ac, n/a, n/a); sleep (sleep disturbance, 3, 4); stress (n/ac, n/a, n/a).	SHAP values (beeswarm plot, bar plot, and dependency plot)
Wang et al [60]	Lifestyle (sleep)	7 (7)	Physical activity (n/ac, n/a, n/a); diet (nutritional status, 3, 7); sleep (outcome, n/a, n/a); stress (n/ac, n/a, n/a).	SHAP values (beeswarm plot, bar plot, and waterfall plot)
Xin and Ren [62]	Mental health			SHAP values (beeswarm plot and bar plot)
		Rural older adults: 55 (20)	Physical activity (exercise, 19, 20); diet (n/ac, n/a, n/a); sleep (sleep hours, 13, 20); stress (n/ac, n/a, n/a).	
		Urban older adults: 55 (16)	Physical activity (exercise, unranked, n/a); diet (n/ac, n/a, n/a); sleep (sleep hours, 6, 16); stress (n/ac, n/a, n/a).	
Zhang et al [63]	Lifestyle (sleep)	20 (10)	Physical activity (sedentary time, 5, 10); diet (vegetable consumption, 3, 10); sleep (outcome, n/a, n/a); stress (stress score, 1, 10).	SHAP values (beeswarm plot and dependency plot)
Zhou et al [63]	Psoriasis	150 (20)	Physical activity (intensity, 17, 20); diet (habits, unranked, n/a); sleep (n/ac, n/a, n/a); stress (n/ac, n/a, n/a).	SHAP values (beeswarm plot)

[a]XAI: explainable artificial intelligence.

[b]n/ac: not applicable.

[c]n/a: not available.

[d]LIME: local interpretable model-agnostic explanations.

[e]SHAP: Shapley Additive Explanations.

XSL•FO
RenderX

Because the nonlifestyle features included in each model vary substantially across studies, the ranking reported in Table 2 should be interpreted in relative terms. When a lifestyle behavior appears among the top-ranked features, this indicates that it contributed more strongly to the model than nonlifestyle variables included in the analysis. Conversely, lifestyle components appearing in lower positions acted as secondary predictors.

In a study on cardiometabolic disease [76], the ML solution was first explained at the individual participant level to provide specific behavioral feedback, and then at the group level to reveal the ranking of features for succeeding in behavioral changes. In both models, physical activity and diet variables were among the top contributors. However, in another study [65], neither regular physical activity nor diet habits were among the top 6 variables for predicting diabetes in adults.

In 4 mental health studies [64,67,80,85], SHAP values were used to rank the contribution of each feature. In 3 of the 4 studies, lifestyle variables were among the top contributors. In [80], physical activity and fruit and vegetable consumption; in [64], sleep variables were the top predictors of stress in young adolescents during the COVID-19 pandemic; and in [85], sleep duration was identified as important for predicting depression. However, in [67], the intensity of physical activity was unranked because it was excluded from the XAI analysis during a feature selection step prior to modeling. In the same study, sleep hours were the most important variable for predicting depression among adults. Regarding mental health studies, psychological resilience was assessed among medical students [56], with sleep disturbance being a key factor affecting their resilience.

Three studies focusing on older adults in China had cognitive function as their outcome [50,52,63]. In these studies, physical activity variables were ranked as top predictors. Additionally, in 1 of these studies [63], sleep quality was ranked as the eighth top feature out of 20. Furthermore, when loneliness was assessed among older adults from China [51], exercise was ranked as the seventh variable out of 15.

In a study predicting quality of life [91], stress, sleep quality, and physical activity emerged as the strongest predictors, with the eating index appearing in the top 15 variables.

Two studies focused on sleep as the specific outcome. In [86], the SHAP values ranked stress score, vegetable consumption, and sedentary time among the top 5 variables for predicting sleep disturbance. In [97], nutritional status was the third most important variable for predicting the risk of sleep disorders in older adults.

In [88], “favorable” food stores and global stress were identified as the top variables for predicting incidence of cardiovascular disease, with the availability of outdoor activities ranking in the top 5. In a longitudinal study investigating the association of diet with long-term reduction in waist circumference, SHAP values highlighted the importance of high-quality components in reducing visceral fat [83]. This study also measured exercise with a single item of frequency, which was not included in the top ranking. Regarding a diet study [94], the adoption of different diets in the Spanish population was assessed, with fish consumption positioned as the most important variable for all diets studied (Mediterranean, intermittent fasting, vegan, and vegetarian). In the same study, practicing sport was ranked among the top variables across the different diets. However, sleep quality was only included in the ranking for intermittent fasting and the vegan diet.

Physical inactivity emerged as the most important feature in explaining county-level obesity using LIME [87]. In this study, the food environment and insufficient sleep, both measured as single items, were not included as top predictors of obesity prevalence.

In a cross-national study assessing frailty [68], sleep variables were among the 20 most important variables across all the cohorts studied. Physical activity was included in the US and UK cohorts and was also ranked in the top 20 variables. In the UK cohort, fruit consumption was the 19th out of 20 key contributors to frailty.

One study on osteoporosis [84] used the Life’s Essential 8 scores for physical activity, diet, and sleep. Sleep and physical activity scores were among the 20 most important variables, but the diet health score was unranked. Similarly, the Life’s Essential 8 scores were used to determine infertility risk in women [60], with the diet score being the only one ranked among the top 10 variables.

Finally, 1 study combined lifestyle factors with metabolites associated with psoriasis [71], with physical activity intensity ranked among the 20 key factors out of 150 variables, whereas dietary habits were not ranked.

Software to Implement ML Models

Of the included papers in the review, 23 (35.38%) used R software (R Core Team) [105] for data analysis (Table 3). The R packages used were data.table for data manipulation, tidyverse [106] as a general package for data science, Multivariate Imputation by Chained Equations [107] for missing data imputation, missMDA for performing multiple imputation with principal component analysis, FactoMineR for exploratory data analysis and principal component analysis, Boruta [108] for feature selection through a wrapper algorithm, caret (Classification and Regression Training) [109] for creating models, randomForest for RF analysis, randomForestSCR for RF for survival, regression, and classification analysis, rpart for recursive partitioning and regression trees, xgboost for extreme gradient boosting, bartMachine for Bayesian additive regression trees, kernlab and e1071 for support vector machines, survival for survival analysis, lime for local interpretable model-agnostic explanations, and finally, SuperLearner [110] to choose the optimal learner for a given prediction problem with a k-fold cross-validation algorithm.

XSL•FO
RenderX

Table . Softwares used in the reviewed studies to perform machine learning (ML) algorithms.

Software used	Number of studies	Study references
R (R Core Team)	n=23	[41,45-47,49,54,55,69,72-74,77,78,80,87,90,93,95-97,99,101,111]
Python (Python Software Foundation)	n=24	[42,50,52,53,56,58,61,64,65,67,70,76,79,81,83,84,86,88,89,91,98,100,102,104]
R and Python	n=4	[57,63,68,71]
SPSS (IBM Corp)	n=2	[62,92]
MATLAB (The MathWorks Inc)	n=2	[48,66]
KNIME [112]	n=2	[43,44]

In contrast, 24 (36.92%) studies developed the models in Python (Python Software Foundation), using the following libraries: Scikit-learn, used in all studies for predictive data analysis; pandas, for manipulating tabular data; NumPy, for mathematical functions; Keras and TensorFlow, for implementing deep learning; lightGBM, for performing light gradient boosting machine; SHAP, to explain ML solutions; creme, for online ML; Bayesian optimization, as a global optimization package to find the maximum value of an unknown function in as few iterations as possible; imbalanced-learn, to combine either undersample or oversample methods; and TextBlob, emoji, nltk, and profanity, for processing and analyzing textual data.

Finally, 4 studies used both Python and R, and 6 studies used other software programs such as SPSS, MATLAB, and KNIME. Eight papers did not report the software used [10,40,59,60,75,82,85,94].

Discussion

Overview

This scoping review of 65 studies provides the current state of the application of supervised ML algorithms for the analysis of lifestyle data. The increase in studies in this field since 2019 indicates that it is a noteworthy area of study. The diversity in the sample origin, alongside the accessibility to new AI tools and novel methods for monitoring health outcomes (eg, wearables), denotes global attention to lifestyle. This section addresses the methodological shortcomings found in the reviewed studies.

About Data Acquisition

In relation to lifestyle data, we found that most studies adopted a multidomain approach, integrating more than just a single component. This strategy enhances and facilitates a more comprehensive understanding of health problems related to the 4 lifestyle domains considered in this review: physical activity, diet, sleep, and stress. The distribution of lifestyle domains identified here was similar to that reported in a previous scoping meta-review [113], although we observe that sleep has gained prominence in recent years, now reaching a level comparable to diet. Although these results highlight the growing recognition of the interrelated nature of lifestyle behaviors, the imbalance in the distribution of these factors limits the capacity of current studies to fully model and understand the interaction among the 4 lifestyle components and their combined effects on health.

Concerning the data acquisition process, over half of the studies acquired their own data. This acquisition process implies control over variables and reduces the time required for cleaning [114]. Interestingly, both self-acquired datasets and those sourced from private or public health repositories demonstrated gender parity in the analyzed datasets. However, we detected a major limitation in this part of the process regarding the data acquisition methodology. In most studies, data were collected through single items, such as regular physical activity (response “yes” or “no”) [58] or usual time of waking up and going to bed [62], resulting in low representativeness of the construct being measured. The result of this acquisition method is high heterogeneity in measures, which hinders their generalizability. Therefore, the quality of data must be one of the challenges to be addressed, and specifically, the consistency in measures [115,116].

Nevertheless, the current accessibility and precision of health sensors such as wearables [117] and the Internet of Things [118] may contribute to transferability and actionability in the population [119]. The growth in technology allows the integration of different data forms as well as more objective measures of lifestyle, substantially reducing the impact of retrospective bias by tracking real-time data in an ecological situation [120]. Therefore, merging questionnaires and sensor data may be the key to identifying relationships between lifestyle measurements and personalizing interventions or changes in specific behaviors. This integration would include physiological, psychological, and behavioral factors, which are the most common analysis types in the ML community to extract clinical insights [121].

About Characteristics of ML Models

Regarding the analysis of ML, 2 different approaches emerged in the reviewed studies: 1 focused on prediction through classification and regression problems, and the other focused on interpretability through feature selection. The first is already an acknowledged approach, while the second typically constitutes an important component of the ML process, specifically during the preprocessing stage. However, feature selection studies do not use model evaluation metrics, which can limit their statistical validity and the generalizability of results. Remarkably, the family of ML algorithms most closely related to feature selection is tree-based because it provides indices of the importance of each variable. Although most papers in this scoping review combined different families of algorithms and compared their results, the most common model family was tree-based, which was applied for each data typology identified.

XSL•FO
RenderX

Specifically, RF is the most used algorithm, which may be due to its robustness in handling missing values, the consideration of complex interaction in the data [122], and its lower sensitivity to variable scales [123]. Despite the benefits of RF, the underusage of DL algorithms represents a critical missed opportunity for robustly analyzing complex and multimodal data. In this review, DL algorithms were underused for lifestyle data. This result may reflect a gap in expertise or access to computational resources among lifestyle researchers, potentially limiting the application of more complex models. With ongoing advances in computational power and algorithmic efficiency, it is expected that the use of DL algorithms will become more widespread in the near future [124].

Regarding the preprocessing stage, most studies detailed some phases of the process, but there is no consensus on the description of this stage of ML. Variable transformation is a crucial step for certain algorithms, particularly for SVM and specific architectures of DL that exhibit sensitivity to the raw form of the variables. In this review, 12 out of 19 studies that performed the SVM algorithm, and 13 out of 26 studies focused on DL, reported variable transformation. Additionally, it is worth noting that these algorithms cannot handle missing values, requiring imputation before the modeling phase. Among the SVM studies, 11 out of 19 reported techniques for missing imputation, and 17 out of 26 DL studies explicitly addressed this. In contrast, tree-based algorithms are less sensitive to variable scales and missing values, yet incorporating these feature engineering steps could enhance model performance [123]. It is noteworthy that the preprocessing steps, specifically how missing values are addressed, have been identified as a potential concern for transparency. This procedural aspect could introduce sampling biases, thereby influencing the generalizability and comprehension of the dataset context [125].

Resampling techniques, aiming to balance the dataset, are commonly implemented in classification problems. SMOTE has been the most widely used technique in this review, especially because it achieves better results than a simple undersampling of the majority class. In the health domain, imbalanced datasets are common, and SMOTE oversamples the minority class with synthetic examples and randomly undersamples the majority class to balance the dataset [126].

Finally, dimensionality reduction enables capturing the most relevant information for the outcome while eliminating noise and redundant information. In this review, dimensionality reduction was the most frequent preprocessing step, appearing in 38 studies. Notably, not only could SVM and DL models benefit from removing irrelevant predictors, but also in tree-based algorithms, dimensionality reduction minimizes model complexity, resource consumption, and data acquisition costs [123].

The division of the original dataset is an essential step for assessing the performance of the ML solution. In this review, 8 studies did not report how they split their data to assess the model, which denotes a lack of generalization of their results. This omission represents a common issue in ML research that should be carefully addressed to minimize bias [127]. Train and test division, also known as “hold out,” is a method with considerable variability due to the use of a unique random data distribution [127]. Therefore, other methods might be more suitable. For example, leave-one-out cross-validation, which trains the model on “n – 1” observations and makes predictions on the remaining one. Although effective for small datasets, it is computationally intensive with large datasets [128]. K-fold cross-validation involves randomly dividing the original dataset into k groups. K-fold cross-validation not only offers computational advantage over leave-one-out, but also gives more accurate estimations due to the bias-variance trade-off [128]. In time-series data, only 1 paper [102] introduced a different form of data split, considering the dependencies of the entire series. Ideally, this type of data should be treated with a method called rolling forecast origin resampling, which estimates the model with historical data and evaluates it with the most recent data [129]. In other words, the training set should ideally comprise observations that occurred before those in the test set; however, this method was not found in this review.

Regarding the evaluation of ML models, the choice of metrics depends on the nature of the problem, whether regression or classification. Specific evaluation metrics tailored to each problem are crucial for correct evaluation, aligning with the priorities and needs of each field. For instance, in medical studies where the cost of treatment in terms of health is high, it becomes crucial to identify true patients over false positives. In contrast, if the treatment has minimal side effects and has demonstrated benefits, sensitivity might not be as important as specificity. It is worth highlighting that in the field of data science, precision and recall are more commonly used, whereas in medical fields, specificity and sensitivity are more prevalent [18]. These differences may cause misunderstandings between the 2 domains.

About Explainability Methods

Regarding model explainability, 22 studies incorporated a dedicated step in the ML process for explainability. SHAP values and LIME were the only XAI methods applied to lifestyle data, and these 2 methods were the most common in a recent systematic review on XAI methods [130]. XAI-related studies in this review were published since 2019, with a notable increase in the number of publications in 2025, comprising half of the papers. This exponential growth was also found in [27,28], where the trend in published papers occurred between 2016 and 2022. Thus, our review demonstrates this exponential distribution in the health and behavioral sciences, where XAI methods are gaining prominence.

Although tree-based algorithms, especially decision trees, are known for facilitating interpretation, SHAP values can be applied to any type of model [131]. The adoption of XAI in lifestyle studies remains low (33.84%). One possible explanation for this is that explainability algorithms are often not integrated into a standard ML pipeline, thereby increasing the technical complexity of the workflow. However, some efforts are being made by R and Python developers to incorporate XAI algorithms into pipelines using libraries such as H2O [132].

In this review, lifestyle components (physical activity, diet, sleep, and stress) consistently appeared among the top-ranked features in models using XAI techniques, highlighting their

XSL•FO
RenderX

substantive contribution relative to nonlifestyle variables. These findings align with prior research emphasizing the integration of diverse lifestyle components [133,134]. However, the level of interpretability achieved also depends on the quality of data used in the models, which in some studies did not meet expected standards. For instance, in a study where the focus was on obesity, the diet component was not among the top-ranked features [87]. In this county-level study, a food environment index was measured as a single item, potentially inadequately representing the diet component of lifestyle.

Therefore, integrating XAI methods into the ML process could enable tailored interventions based on model results, provided that measures are collected accurately. Furthermore, the adoption of XAI algorithms contributes to increased trust and verification of the fairness of the models. This approach can also facilitate the translation of findings to stakeholders and health systems, thereby enhancing transparency, promoting the adoption of models in society, and supporting informed decision-making [27].

About the Software for Implementing ML Models

Competition between Python and R for ML software dominance in data science is currently intense. Both Python and R are freely distributed, object-oriented software with large and active communities. Python, as a programming language, offers specific implementations through libraries tailored for statistical analysis, including ML and DL. In contrast, R is a statistical software that integrates fundamental statistics into its base functionalities. While Python requires libraries for each stage of analysis, its well-established libraries streamline the process. On the other hand, R faces challenges due to its heterogeneous libraries, which hinder replicability and require expertise in varying syntax across packages. To address this concern, the meta-package tidymodels (Max Kuhn and Hadley Wickham) resolves these issues by integrating all necessary packages for each ML step, using a unified syntax. Additionally, tidymodels integrates user-friendly interfaces and promotes good methodological practice, thereby preventing user errors [135]. Conversely, Python presents a preferable environment for DL with the TensorFlow and PyTorch frameworks. In this regard, the possibility of developing ML projects on powerful computational cloud-based platforms, such as Google Colaboratory (also referred to as Google Colab) [136], offers Python a remarkable advantage over R local environments.

Methodological and Reporting Guidelines and Checklist

Based on the review's results and to enhance transparency and replicability in multidisciplinary sciences [137], we provide comprehensive methodological and reporting guidelines and a checklist for ML projects. Although various studies have proposed guidelines and checklists [138,139], the rapid expansion of ML algorithms in health domains necessitates iterative evaluation to incorporate new steps into the ML research workflow. The guidelines and checklist (Checklist 2) are based on the 5 stages of the ML workflow, as depicted in Figure 2, with added software tools.

Data Acquisition

The integration of multidomain data enhances the comprehension of real-world problems. Using appropriate methods to collect data ensures representativeness. We recommend the use of standardized questionnaires and validated sensors. Regarding health repositories, we recommend providing information about data characteristics such as gender distribution, sample size, and variable descriptions [140].

Preprocessing

Reporting the preprocessing methods used in the data analysis is particularly crucial for ensuring replicability. While preprocessing contributes to improving data quality, different preprocessing methods can lead to different results. We propose the following recommendations for each preprocessing step, although not all steps need to be performed in every ML project.

Transformation

Categorical data should be encoded using methods such as one-hot encoding and dummy variables. Continuous data should be transformed using normalization or rescaling of features with different units to ensure algorithm performance, particularly for those sensitive to the raw form of variables [123].

Missing Imputation

Some algorithms cannot handle missing data and require imputation before modeling. Depending on the number of observations and the data distribution, imputation with the mean, median, or mode is typical. For time-series data, imputation with the last or next observation is preferred, though rolling statistics imputation or interpolation may offer better solutions [141].

Resampling

Imbalanced datasets can bias models, resulting in poor performance on underrepresented classes [142]. While SMOTE is an effective technique for handling imbalanced datasets, it is not without its limitations. However, when the class imbalance ratio is extremely high, SMOTE can potentially bias the model performance by overfitting the minority class. This issue is particularly pronounced in datasets containing noise, as synthetic observations may replicate these artifacts. To mitigate these challenges, recent studies have proposed tree algorithms, which have shown effectiveness in handling class imbalance [143]. Furthermore, due to their robustness, the use of tree-based algorithms is increasingly recommended when working with class imbalance [144].

Dimensionality Reduction

Removing noise from the dataset and retaining features directly related to the outcome can enhance both data acquisition and modeling efficiency. Removing correlated features is particularly beneficial [123].

Modeling

Our guidelines focused on SL algorithms for classification and regression problems. The choice of algorithms depends on the measurement of the outcome. We recommend using multiple algorithms to compare results and select the best fit, given that there is significant variability across problems [145].

XSL•FO
RenderX

Additionally, comparing different families of algorithms is also advisable, as some improvements exist within the same family. When it comes to replicability and transparency, reporting algorithm hyperparameters is crucial in ML problems, as different configurations can yield varying results [146].

Evaluation

Avoiding overfitting requires dividing the original dataset appropriately. This step is fundamental in ML implementation and should be considered in every study to ensure that the extracted insights are reliable and generalizable. This division depends on data typology; we identified 3 different typologies relevant at this stage, defined by whether time is an implicit factor in the data acquisition process.

Cross-Sectional

Data are collected at a single point in time, with no temporal dependencies. Leave-one-out cross-validation is effective for small datasets (n<150) typically seen in life sciences; the computational complexity increases [147]. K-fold cross-validation is recommended for large datasets to balance the bias-variance trade-off [128]. Cross-validation with bootstrap resampling can also be used to evaluate the performance of the models and estimate CIs of performance metrics [148].

Longitudinal

In longitudinal studies, researchers collect repeated measures, potentially with dependencies between observations and high correlation that can bias the model [135]. Data should be divided by grouping individual participants' information. The methods used are similar to those for cross-sectional data, but with consideration of partitioning.

Time Series

Time series data are a sequence of data points in chronological order. Rolling forecast origin resampling is suitable for this data [129]. The training set should include observations occurring before those in the test set.

Evaluation Metrics

Appropriate metrics should be selected based on the type of ML problem (regression or classification) [149] and the characteristics of the study field [18]. It is also recommended to compute evaluation metrics repeatedly across cross-validation samples and to apply nonparametric tests, such as the Wilcoxon signed-rank test and the Friedman test, to assess model performance [149]. For imbalanced datasets, performance metrics such as balanced accuracy and F_1-score are recommended. For example, in an imbalanced dataset, a model may achieve high accuracy by predominantly predicting the majority class. In contrast, balanced accuracy assigns equal weight to each class regardless of its frequency, providing a more informative evaluation of model performance [150].

Explainability

Explainability Overview

Reporting the method of explainability used in the ML projects is essential. It is important to distinguish between interpretability and explainability. While explainability refers to understanding the effect of each feature on the original model, interpretability involves deriving actionable insights from the model's prediction. Although tree-based methods include importance metrics, they do not indicate the direction of relationships with the outcome. Incorporating explainable resources such as SHAP values [151] or LIME [152] enhances the interpretability of the results, providing both actionability and transparency, and transforming black boxes into glass box models. The H2O package in R offers XAI algorithms that are compatible with the framework tidymodels, enabling a unified workflow for modeling and explainability. For a more detailed taxonomy of XAI packages in R, we refer the reader to [153].

Additionally, the choice of appropriate visualization methods and the number of features displayed are crucial for ensuring comprehensive results and supporting fine-grained decision-making. Regarding the number of features, we recommend visualizing at least the top 20 features whenever possible, as this allows a broader understanding of the contribution of the most important variables. Furthermore, future studies should report all model features used in the ML analysis, not only the top-ranked ones, to ensure full transparency and allow readers to verify which HL components were excluded or had low impact in the model. Based on SHAP values, we propose the following XAI visualization techniques [94,154]:

Beeswarm Plot

This is the most common summary plot, where features are ranked on the y-axis from most to least contribution by their mean absolute SHAP value. The x-axis represents the SHAP values, which express the change in the log-odds (or model output), resulting in a positive or negative contribution for a specific observation. Each dot represents an observation in the dataset, and the color is indicative of the original value for that observation. Higher values are displayed as red and lower values as blue. The vertical dotted line represents the zero SHAP contribution. Contributions to the right assign a positive effect, and those to the left assign a negative effect.

Dependency Plots

Scatter plots that show the effect a single feature has on the model's predictions. Each point represents an instance's feature value and its corresponding SHAP value. This plot provides a detailed and clear explanation of the direction and magnitude of the relationship (whether linear or nonlinear) between the feature and the outcome.

Bar Plot

This plot provides information on the global feature importance. Features are ordered on the y-axis from the highest to the lowest average contribution to the outcome.

Waterfall Plot

This plot is a local explanation that illustrates the contribution of each feature in transforming the expected value E[f(x)] into the final prediction f(x) for a specific instance. Each row represents a feature's contribution; positive contributions are red, while negative contributions are blue.

XSL•FO
RenderX

Force Plot

This plot is a local explanation that shows the effect and the direction of the most impactful features for a given observation.

Software

Python and R were the most widely used software in the review. However, the variety of libraries can complicate the process. To address this, the metapackage "tidymodels" provides unified syntax, enhancing replicability [135]. For DL, Python offers a more powerful environment thanks to PyTorch and TensorFlow frameworks and the possibility of developing analyses on cloud services such as Google Colab. Furthermore, we encourage researchers to make their code publicly available in open repositories such as GitHub.

Limitations

At the level of the review itself, some limitations emerge. Although time-series data were frequently used in the analysis of lifestyle behaviors, the primary focus of this review was on the methodological framework of supervised ML, rather than on specific time-series modeling approaches. Future research should specifically address this gap by examining the current use of time-series models applied to wearable data to better capture human behavior. Moreover, the exclusion of UL algorithms limits the scope of ML algorithms covered in this review. For example, cluster analysis has been used to classify children according to their eating behaviors and identify features related to obesity [155], with findings suggesting that interventions such as reducing eating speed may help prevent childhood obesity. This scoping review focuses on supervised ML approaches, but future research could examine the current application of UL in lifestyle data. Finally, a further limitation of this review is that the search strategy prioritized studies using the umbrella term "healthy lifestyle." While this approach captured a representative sample of research explicitly conceptualized as HL, it may have excluded multidimensional studies examining combinations of physical activity, diet, sleep, or stress that did not use this terminology. Although an interaction block was incorporated to address the interconnectedness among these domains, we acknowledge that some multidomain research may not have been retrieved due to the specificity of the search string.

Conclusion

This review has identified several limitations within the studies reviewed that need to be addressed. First, ensuring data quality remains a significant challenge that must be addressed by carefully selecting data acquisition methods to build reliable and robust models. Second, the evaluation process is crucial for preventing overfitting, and using hold-out cross-validation can lead to high variance partitioning. Therefore, it is recommended to implement k-fold cross-validation at various stages, such as during validation; for time-series data, rolling forecast origin resampling is recommended [129].

In conclusion, this scoping review provides a comprehensive analysis of lifestyle using ML models and serves as a guideline for future research. While the relationship between lifestyle and health is well-established, ongoing efforts are needed to refine how we measure lifestyle to create robust models. It is essential to focus not only on model performance but also on data representativeness, which is closely related to the granularity established during data collection. Although RF algorithms are prominent in lifestyle data analysis, it is recommended to compare their performance with other algorithms within and across families. Future research should also incorporate SHAP values to enhance interpretability within the ML workflow. Additionally, the *tidymodels* metapackage (R software) with H2O for XAI can assist researchers in evaluating process quality with unified syntax, thereby contributing to replicability.

Acknowledgments

This work was supported by the Catalonian Government through a FI-SDUR grant (ref 2023FISDU_00217).

Funding

This work was supported by grants PID2019-107473RB-C21 and PID2022-141403NB-I00, funded by MCIN/AEI/10.13039/501100011033/FEDER, UE, and 2021SGR-00806, funded by the Catalonian Government. Additional funding was provided by the Catalonian Government through an FI-SDUR grant (ref 2023FISDU_00217).

Data Availability

The datasets generated or analyzed during this study are available in the

CORA Repositori de Dades de Recerca repository [31].

Authors' Contributions

TE contributed to conceptualization, validation, formal analysis, investigation, data curation, writing of the original draft, and visualization. LC contributed to conceptualization, writing (review and editing), supervision, project administration, and funding acquisition. CA contributed to validation and writing (review and editing). JML contributed to conceptualization, methodology, writing (review and editing), supervision, project administration, and funding acquisition.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Extended data and detailed analyses of the included studies.
[DOCX File, 5266 KB - humanfactors_v13i1e78648_app1.docx]

Checklist 1
PRISMA checklist.
[PDF File, 178 KB - humanfactors_v13i1e78648_app2.pdf]

Checklist 2
Reporting checklist for machine learning analysis in healthy lifestyle data.
[DOCX File, 17 KB - humanfactors_v13i1e78648_app3.docx]

References

1. Furihata R, Konno C, Suzuki M, et al. Unhealthy lifestyle factors and depressive symptoms: a Japanese general adult population survey. J Affect Disord 2018 Jul;234:156-161. [doi: 10.1016/j.jad.2018.02.093] [Medline: 29529548]
2. Nudelman G, Shiloh S. Connectionism and behavioral clusters: differential patterns in predicting expectations to engage in health behaviors. Ann Behav Med 2018 Sep 13;52(10):890-901. [doi: 10.1093/abm/kax063] [Medline: 30212846]
3. Nudelman G, Shiloh S. Mapping health behaviors: constructing and validating a common-sense taxonomy of health behaviors. Soc Sci Med 2015 Dec;146:1-10. [doi: 10.1016/j.socscimed.2015.10.004] [Medline: 26473449]
4. Kaminsky LA, German C, Imboden M, Ozemek C, Peterman JE, Brubaker PH. The importance of healthy lifestyle behaviors in the prevention of cardiovascular disease. Prog Cardiovasc Dis 2022 Jan;70:8-15. [doi: 10.1016/j.pcad.2021.12.001]
5. Braun C, Foreyt JP, Johnston CA. Stress: a core lifestyle issue. Am J Lifestyle Med 2016;10(4):235-238. [doi: 10.1177/1559827616642400] [Medline: 30202277]
6. Wong VWH, Ho FYY, Shi NK, Sarris J, Chung KF, Yeung WF. Lifestyle medicine for depression: a meta-analysis of randomized controlled trials. J Affect Disord 2021 Apr;284:203-216. [doi: 10.1016/j.jad.2021.02.012]
7. Sagner M, Katz D, Egger G, et al. Lifestyle medicine potential for reversing a world of chronic disease epidemics: from cell to community. Int J Clin Pract 2014 Nov;68(11):1289-1292. [doi: 10.1111/ijcp.12509]
8. Cerf ME. Healthy lifestyles and noncommunicable diseases: nutrition, the life‐course, and health promotion. Lifestyle Medicine 2021 Apr;2(2):e31 [FREE Full text] [doi: 10.1002/lim2.31]
9. Gurrin C, Smeaton AF, Doherty AR. LifeLogging: personal big data. Foundations and Trends® in Information Retrieval 2014 Jun 16;8(1):1-125. [doi: 10.1561/1500000033]
10. Lim J, Jeong CY, Lim JM, et al. Assessing sleep quality using mobile EMAs: opportunities, practical consideration, and challenges. IEEE Access 2022;10:2063-2076. [doi: 10.1109/ACCESS.2021.3140074]
11. Kline A, Wang H, Li Y, et al. Multimodal machine learning in precision health: a scoping review. NPJ Digit Med 2022 Nov 7;5(1):171. [doi: 10.1038/s41746-022-00712-8] [Medline: 36344814]
12. Beam AL, Kohane IS. Big data and machine learning in health care. JAMA 2018 Apr 3;319(13):1317-1318. [doi: 10.1001/jama.2017.18391] [Medline: 29532063]
13. Maleki F, Ovens K, Najafian K, Forghani B, Reinhold C, Forghani R. Overview of machine learning part 1. Neuroimaging Clin N Am 2020 Nov;30(4):e17-e32. [doi: 10.1016/j.nic.2020.08.007]
14. Secinaro S, Calandra D, Secinaro A, Muthurangu V, Biancone P. The role of artificial intelligence in healthcare: a structured literature review. BMC Med Inform Decis Mak 2021 Apr 10;21(1):125. [doi: 10.1186/s12911-021-01488-9] [Medline: 33836752]
15. Breiman L. Statistical modeling: the two cultures (with comments and a rejoinder by the author). Statist Sci 2001 Aug 1;16(3). [doi: 10.1214/ss/1009213726]
16. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to machine learning, neural networks, and deep learning. Transl Vis Sci Technol 2020 Feb 27;9(2):14. [doi: 10.1167/tvst.9.2.14] [Medline: 32704420]
17. Efron B. Prediction, estimation, and attribution. Int Statistical Rev 2020 Dec;88(S1) [FREE Full text] [doi: 10.1111/insr.12409]
18. Bruce PC, Bruce A, Gedeck P. Estadística Práctica Para Ciencia de Datos Con R y Python: Más de 50 Conceptos Esenciales, 2nd edition: Marcombo; 2022.
19. James G, Witten D, Hastie T, Learning TRS. An Introduction to Statistical Learning: Springer; 2021:15-57. [doi: 10.1007/978-1-0716-1418-1_2]
20. Saqib K, Khan AF, Butt ZA. Machine learning methods for predicting postpartum depression: scoping review. JMIR Ment Health 2021 Nov 24;8(11):e29838. [doi: 10.2196/29838] [Medline: 34822337]

XSL•FO
RenderX

21. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. Psychol Med 2019 Jul;49(9):1426-1448. [doi: 10.1017/S0033291719000151] [Medline: 30744717]
22. Ono S, Goto T. Introduction to supervised machine learning in clinical epidemiology. Ann Clin Epidemiol 2022;4(3):63-71. [doi: 10.37737/ace.22009] [Medline: 38504945]
23. Bi Q, Goodman KE, Kaminsky J, Lessler J. What is machine learning? A primer for the epidemiologist. Am J Epidemiol 2019 Dec 31;188(12):2222-2239. [doi: 10.1093/aje/kwz189] [Medline: 31509183]
24. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence-based chatbots for promoting health behavioral changes: systematic review. J Med Internet Res 2023 Feb 24;25:e40789. [doi: 10.2196/40789] [Medline: 36826990]
25. Goh YS, Ow Yong JQY, Chee BQH, Kuek JHL, Ho CSH. Machine learning in health promotion and behavioral change: scoping review. J Med Internet Res 2022 Jun 2;24(6):e35831. [doi: 10.2196/35831] [Medline: 35653177]
26. Lai X, Chen J, Lai Y, et al. Using large language models to enhance exercise recommendations and physical activity in clinical and healthy populations: scoping review. JMIR Med Inform 2025 May 27;13:e59309. [doi: 10.2196/59309] [Medline: 40424584]
27. Mersha M, Wood J, AlShami A, Kalita J, Lam K. Explainable artificial intelligence: a survey of needs, techniques, applications, and future direction. arXiv. Preprint posted online on Aug 30, 2024. [doi: 10.48550/ARXIV.2409.00265]
28. Ali S, Abuhmed T, El-Sappagh S, et al. Explainable artificial intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence. Information Fusion 2023 Nov;99:101805. [doi: 10.1016/j.inffus.2023.101805]
29. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143. [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
30. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
31. Estrella T, Alfonso C, Capdevila L, Losilla JM. Machine learning for the analysis of healthy lifestyle data: a scoping review protocol. Universitat Autònoma de Barcelona. 2023 Mar. URL: https://portalrecerca.uab.cat/en/publications/machine-learning-for-the-analysis-of-healthy-lifestyle-data-a-sco/ [accessed 2026-02-20]
32. Estrella T, Capdevila L, Alfonso C, Losilla JM. Replication data for machine learning for the analysis of healthy lifestyle data: a scoping review. : CORA.Repositori de Dades de Recerca; 2024. [doi: 10.34810/DATA1088]
33. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 guideline statement. J Clin Epidemiol 2016 Jul;75:40-46. [doi: 10.1016/j.jclinepi.2016.01.021] [Medline: 27005575]
34. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA Statement for reporting literature searches in systematic reviews. Syst Rev 2021 Jan 26;10(1):39. [doi: 10.1186/s13643-020-01542-z] [Medline: 33499930]
35. Loef M, Walach H. The combined effects of healthy lifestyle behaviors on all cause mortality: a systematic review and meta-analysis. Prev Med 2012 Sep;55(3):163-170. [doi: 10.1016/j.ypmed.2012.06.017] [Medline: 22735042]
36. Yu J, Feng Q, Kim JH, Zhu Y. Combined effect of healthy lifestyle factors and risks of colorectal adenoma, colorectal cancer, and colorectal cancer mortality: systematic review and meta-analysis. Front Oncol 2022 Jul 22;12:827019. [doi: 10.3389/fonc.2022.827019]
37. Lippman D, Stump M, Veazey E, et al. Foundations of lifestyle medicine and its evolution. Mayo Clin Proc Innov Qual Outcomes 2024 Feb;8(1):97-111. [doi: 10.1016/j.mayocpiqo.2023.11.004] [Medline: 38304165]
38. Sayburn A. Lifestyle medicine: a new medical specialty? BMJ 2018 Oct 25:k4442. [doi: 10.1136/bmj.k4442]
39. Fleig L, Ngo J, Roman B, et al. Beyond single behaviour theory: adding cross‐behaviour cognitions to the health action process approach. British J Health Psychol 2015 Nov;20(4):824-841 [FREE Full text] [doi: 10.1111/bjhp.12144]
40. Butkevičiūtė E, Bikulčienė L, Žvironienė A. Physiological state evaluation in working environment using expert system and random forest machine learning algorithm. Healthcare (Basel) 2023 Jan 11;11(2):220. [doi: 10.3390/healthcare11020220] [Medline: 36673588]
41. Cai T, Long J, Kuang J, You F, Zou T, Wu L. Applying machine learning methods to develop a successful aging maintenance prediction model based on physical fitness tests. Geriatrics Gerontology Int 2020 Jun;20(6):637-642. [doi: 10.1111/ggi.13926]
42. Lim J, Kim J, Cheon S. A deep neural network-based method for early detection of osteoarthritis using statistical data. Int J Environ Res Public Health 2019 Apr 10;16(7):1281. [doi: 10.3390/ijerph16071281] [Medline: 30974803]
43. Recenti M, Ricciardi C, Edmunds KJ, et al. Healthy aging within an image: using muscle radiodensitometry and lifestyle factors to predict diabetes and hypertension. IEEE J Biomed Health Inform 2021 Jun;25(6):2103-2112. [doi: 10.1109/JBHI.2020.3044158] [Medline: 33306475]
44. Recenti M, Ricciardi C, Edmunds K, Jacob D, Gambacorta M, Gargiulo P. Testing soft tissue radiodensity parameters interplay with age and self-reported physical activity. Eur J Transl Myol 2021 Jul 12;31(3):9929. [doi: 10.4081/ejtm.2021.9929] [Medline: 34251162]
45. Staudenmayer J, He S, Hickey A, Sasaki J, Freedson P. Methods to estimate aspects of physical activity and sedentary behavior from high-frequency wrist accelerometer measurements. J Appl Physiol 2015 Aug 15;119(4):396-403. [doi: 10.1152/japplphysiol.00026.2015]

XSL•FO
RenderX

46. Zhou M, Fukuoka Y, Goldberg K, Vittinghoff E, Aswani A. Applying machine learning to predict future adherence to physical activity programs. BMC Med Inform Decis Mak 2019 Aug 22;19(1):169. [doi: 10.1186/s12911-019-0890-0] [Medline: 31438926]
47. Dianati-Nasab M, Salimifard K, Mohammadi R, et al. Machine learning algorithms to uncover risk factors of breast cancer: insights from a large case-control study. Front Oncol 2023;13:1276232. [doi: 10.3389/fonc.2023.1276232] [Medline: 38425674]
48. Matta SC, Sankari Z, Rihana S. Heart rate variability analysis using neural network models for automatic detection of lifestyle activities. Biomed SIGNAL Process CONTROL 2018 Apr;42:145-157. [doi: 10.1016/j.bspc.2018.01.016]
49. Park H, Jung SY, Han MK, et al. Lowering barriers to health risk assessments in promoting personalized health management. J Pers Med 2024 Mar 18;14(3):316. [doi: 10.3390/jpm14030316] [Medline: 38541058]
50. Li Y, Song H. The association between sports social capital and cognitive health: a longitudinal study of middle-aged and elderly adults in China. SSM Popul Health 2025 Jun;30:101778. [doi: 10.1016/j.ssmph.2025.101778] [Medline: 40212736]
51. Lin Y, Li C, Wang X, Li H. Development of a machine learning-based risk assessment model for loneliness among elderly Chinese: a cross-sectional study based on Chinese longitudinal healthy longevity survey. BMC Geriatr 2024 Nov 14;24(1). [doi: 10.1186/s12877-024-05443-x]
52. Ren H, Zheng Y, Li C, et al. Using machine learning to predict cognitive decline in older adults from the Chinese longitudinal healthy longevity survey: model development and validation study. JMIR Aging 2025 Apr 30;8:e67437. [doi: 10.2196/67437] [Medline: 40305830]
53. Liu X, Luo Z, Jing F, et al. Estimating cardiovascular mortality in patients with hypertension using machine learning: the role of depression classification based on lifestyle and physical activity. J Psychosom Res 2025 Feb;189:112030. [doi: 10.1016/j.jpsychores.2024.112030] [Medline: 39752763]
54. Amú Ruiz FA, Gonzalez Bustamante DA, Ortiz González K, Sandoval KM. Árbol de clasificación para la identificación de síntomas asociados a la depresión en estudiantes de una universidad pública (Classification tree for the identification of symptoms associated with depression in students of a public university). Retos 2024;52:104-114. [doi: 10.47197/retos.v52.100138]
55. Wallace ML, Coleman TS, Mentch LK, et al. Physiological sleep measures predict time to 15‐year mortality in community adults: application of a novel machine learning framework. J Sleep Res 2021 Dec;30(6):e13386. [doi: 10.1111/jsr.13386] [Medline: 33991144]
56. Su W, Jia H, Chang W, et al. The secrets of medical students' psychological resilience: a dual perspective of machine learning and path analysis. Int J Med Inform 2026 Jan;205:106111. [doi: 10.1016/j.ijmedinf.2025.106111]
57. Luo W, Gong L, Chen X, et al. Lifestyle and chronic kidney disease: a machine learning modeling study. Front Nutr 2022;9:918576. [doi: 10.3389/fnut.2022.918576] [Medline: 35938107]
58. Afrash MR, Bayani A, Shanbehzadeh M, Bahadori M, Kazemi-Arpanahi H. Developing the breast cancer risk prediction system using hybrid machine learning algorithms. J Educ Health Promot 2022;11:272. [doi: 10.4103/jehp.jehp_42_22] [Medline: 36325225]
59. Alshurafa N, Sideris C, Pourhomayoun M, Kalantarian H, Sarrafzadeh M, Eastwood JA. Remote health monitoring outcome success prediction using baseline and first month intervention data. IEEE J Biomed Health Inform 2017 Mar;21(2):507-514. [doi: 10.1109/JBHI.2016.2518673] [Medline: 26780823]
60. Gu X, Li Q, Wang X. Using Life's Essential 8 and heavy metal exposure to determine infertility risk in American women: a machine learning prediction model based on the SHAP method. Front Endocrinol (Lausanne) 2025;16:1586828. [doi: 10.3389/fendo.2025.1586828] [Medline: 40687585]
61. Nichols ES, Pathak HS, Bgeginski R, et al. Machine learning-based predictive modeling of resilience to stressors in pregnant women during COVID-19: a prospective cohort study. PLoS ONE 2022;17(8):e0272862. [doi: 10.1371/journal.pone.0272862] [Medline: 35951588]
62. Bôto JM, Marreiros A, Diogo P, et al. Health behaviours as predictors of the Mediterranean diet adherence: a decision tree approach. Public Health Nutr 2022 Jul;25(7):1864-1876. [doi: 10.1017/S1368980021003293] [Medline: 34369348]
63. Huang Y, Huang Z, Yang Q, et al. Predicting mild cognitive impairment among Chinese older adults: a longitudinal study based on long short-term memory networks and machine learning. Front AGING Neurosci 2023;15. [doi: 10.3389/fnagi.2023.1283243]
64. Kiss O, Alzueta E, Yuksel D, et al. The pandemic's toll on young adolescents: prevention and intervention targets to preserve their mental health. J Adolesc Health 2022 Mar;70(3):387-395. [doi: 10.1016/j.jadohealth.2021.11.023] [Medline: 35090817]
65. Majcherek D, Ciesielski A, Sobczak P. AI-driven analysis of diabetes risk determinants in U.S. adults: exploring disease prevalence and health factors. PLoS ONE 2025;20(9):e0328655. [doi: 10.1371/journal.pone.0328655] [Medline: 40901823]
66. Mousavi H, Karandish M, Jamshidnezhad A, Hadianfard AM. Determining the effective factors in predicting diet adherence using an intelligent model. Sci Rep 2022 Jul 19;12(1):12340. [doi: 10.1038/s41598-022-16680-8] [Medline: 35853992]
67. Jin T, Halili A. Predicting the risk of depression in older adults with disability using machine learning: an analysis based on CHARLS data. Front Artif Intell 2025;8:1624171. [doi: 10.3389/frai.2025.1624171] [Medline: 40673213]

XSL•FO
RenderX

68. Luo Y, Guo M, Zhang Q. Cross-national analysis of social determinants of frailty among middle-aged and older adults: a machine learning study in the USA, England, and China. Humanit Soc Sci Commun 2025 May 30;12(1). [doi: 10.1057/s41599-025-05088-0]
69. Puterman E, Weiss J, Hives BA, et al. Predicting mortality from 57 economic, behavioral, social, and psychological factors. Proc Natl Acad Sci USA 2020 Jul 14;117(28):16273-16282. [doi: 10.1073/pnas.1918455117]
70. Qasrawi R, Vicuna Polo S, Abu Khader R, et al. Machine learning techniques for identifying mental health risk factor associated with schoolchildren cognitive ability living in politically violent environments. Front PSYCHIATRY 2023;14:1071622. [doi: 10.3389/fpsyt.2023.1071622] [Medline: 37304448]
71. Zhou X, Wang Y, Chang Q, et al. Lifestyle-associated serum metabolites profiling in relation to risk of late-onset psoriasis. J Eur Acad Dermatol Venereol 2025 Sep. [doi: 10.2139/ssrn.5085567] [Medline: 40956048]
72. Abdul Rahman H, Kwicklis M, Ottom M, et al. Machine learning-based prediction of mental well-being using health behavior data from university students. Bioengineering (Basel) 2023 May 10;10(5):575. [doi: 10.3390/bioengineering10050575] [Medline: 37237644]
73. Cortés-Ibañez FO, Nagaraj SB, Cornelissen L, et al. Prediction of incident cancers in the lifelines population-based cohort. Cancers (Basel) 2021 Apr 28;13(9):2133. [doi: 10.3390/cancers13092133] [Medline: 33925159]
74. Cortés-Ibañez FO, Belur Nagaraj S, Cornelissen L, Sidorenkov G, de Bock GH. A classification approach for cancer survivors from those cancer-free, based on health behaviors: analysis of the lifelines cohort. CANCERS (Basel) 2021 May 12;13(10):2335. [doi: 10.3390/cancers13102335] [Medline: 34066093]
75. Faruqui SHA, Du Y, Meka R, et al. Development of a deep learning model for dynamic forecasting of blood glucose level for type 2 diabetes mellitus: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2019 Nov 1;7(11):e14452. [doi: 10.2196/14452] [Medline: 31682586]
76. Guthrie NL, Carpenter J, Edwards KL, et al. Emergence of digital biomarkers to predict and modify treatment efficacy: machine learning study. BMJ Open 2019 Jul 23;9(7):e030710. [doi: 10.1136/bmjopen-2019-030710] [Medline: 31337662]
77. Hu L, Liu B, Li Y. Ranking sociodemographic, health behavior, prevention, and environmental factors in predicting neighborhood cardiovascular health: a Bayesian machine learning approach. Prev Med 2020 Dec;141:106240. [doi: 10.1016/j.ypmed.2020.106240] [Medline: 32860821]
78. Hu L, Liu B, Ji J, Li Y. Tree-based machine learning to identify and understand major determinants for stroke at the neighborhood level. J Am Heart Assoc 2020 Nov 17;9(22):e016745. [doi: 10.1161/JAHA.120.016745] [Medline: 33140687]
79. Luo L, Yuan J, Xu C, et al. Mental health issues and 24-Hour movement guidelines–based intervention strategies for university students with high-risk social network addiction: cross-sectional study using a machine learning approach. J Med Internet Res 2025 Jun 13;27:e72260. [doi: 10.2196/72260] [Medline: 40512996]
80. Majcherek D, Kowalski AM, Lewandowska MS. Lifestyle, demographic and socio-economic determinants of mental health disorders of employees in the European countries. Int J Environ Res Public Health 2022 Sep 21;19(19):11913. [doi: 10.3390/ijerph191911913] [Medline: 36231214]
81. Mun EY, Geng F. Predicting post-experiment fatigue among healthy young adults: random forest regression analysis. Psychol Test Assess Model 2019 Nov 8;61(4):471-493. [Medline: 32038903]
82. Park J, Edington DW. Application of a prediction model for identification of individuals at diabetic risk. Methods Inf Med 2004;43(3):273-281. [Medline: 15227557]
83. Park S. Association of a high healthy eating index diet with long-term visceral fat loss in a large longitudinal study. Nutrients 2024 Feb;16(4):534. [doi: 10.3390/nu16040534]
84. Shi H, Fang Y, Ma X. Application of machine learning algorithms in osteoporosis analysis based on cardiovascular health assessed by life's essential 8: a cross-sectional study. J Health Popul Nutr 2025 May 29;44(1). [doi: 10.1186/s41043-025-00941-z]
85. Xin Y, Ren X. Predicting depression among rural and urban disabled elderly in China using a random forest classifier. BMC Psychiatry 2022 Feb 15;22(1):118. [doi: 10.1186/s12888-022-03742-4] [Medline: 35168579]
86. Zhang L, Zhao S, Yang Z, Zheng H, Lei M. An artificial intelligence platform to stratify the risk of experiencing sleep disturbance in university students after analyzing psychological health, lifestyle, and sports: a multicenter externally validated study. Psychol Res Behav Manag 2024;17:1057-1071. [doi: 10.2147/PRBM.S448698] [Medline: 38505352]
87. Allen B. An interpretable machine learning model of cross-sectional U.S. county-level obesity prevalence using explainable artificial intelligence. PLOS ONE 2023;18(10):e0292341. [doi: 10.1371/journal.pone.0292341] [Medline: 37796874]
88. Morris MC, Moradi H, Aslani M, et al. Predicting incident cardiovascular disease among African-American adults: a deep learning approach to evaluate social determinants of health in the Jackson heart study. PLOS ONE 2023;18(11):e0294050. [doi: 10.1371/journal.pone.0294050] [Medline: 37948388]
89. Moon Y, Woo H. Key risk factors of generalized anxiety disorder in adolescents: machine learning study. Front Public Health 2024;12:1504739. [doi: 10.3389/fpubh.2024.1504739] [Medline: 39839408]
90. Cheung YK, Hsueh PYS, Qian M, et al. Are nomothetic or ideographic approaches superior in predicting daily exercise behaviors? METHODS Inf Med 2017;56(6):452-460. [doi: 10.3414/ME16-02-0051] [Medline: 29582914]

XSL•FO
RenderX

91. Kim J, Jeong K, Lee S, Baek Y. Machine-learning model predicting quality of life using multifaceted lifestyles in middle-aged South Korean adults: a cross-sectional study. BMC Public Health 2024 Jan 11;24(1):159. [doi: 10.1186/s12889-023-17457-y] [Medline: 38212741]
92. Pereira MG, Santos M, Magalhães R, Rodrigues C, Araújo O, Durães D. Burnout risk profiles in psychology students: an exploratory study with machine learning. Behav Sci (Basel) 2025 Apr 9;15(4):505. [doi: 10.3390/bs15040505] [Medline: 40282126]
93. Ai M, Morris TP, Zhang J, et al. Resting-state MRI functional connectivity as a neural correlate of multidomain lifestyle adherence in older adults at risk for Alzheimer's disease. Sci Rep 2023 May 9;13(1):7487. [doi: 10.1038/s41598-023-32714-1] [Medline: 37160915]
94. Sandri E, Cerdá Olmedo G, Piredda M, Werner LU, Dentamaro V. Explanatory AI predicts the diet adopted based on nutritional and lifestyle habits in the Spanish population. Eur J Investig Health Psychol Educ 2025 Jan 24;15(2):11. [doi: 10.3390/ejihpe15020011] [Medline: 39997075]
95. Birk N, Matsuzaki M, Fung TT, et al. Exploration of machine learning and statistical techniques in development of a low-cost screening method featuring the global diet quality score for detecting prediabetes in rural India. J Nutr 2021 Oct 23;151(12 Suppl 2):110S-118S. [doi: 10.1093/jn/nxab281] [Medline: 34689190]
96. Wallace ML, Buysse DJ, Redline S, et al. Multidimensional sleep and mortality in older adults: a machine-learning comparison with other risk factors. J Gerontol A Biol Sci Med Sci 2019 Nov 13;74(12):1903-1909. [doi: 10.1093/gerona/glz044] [Medline: 30778527]
97. Wang X, Zhang D, Lu L, et al. Development and validation of an explainable machine learning model for predicting the risk of sleep disorders in older adults with multimorbidity: a cross-sectional study. Front Public Health 2025;13:1619406. [doi: 10.3389/fpubh.2025.1619406] [Medline: 40860561]
98. Zhou RZ, Hu Y, Tirabassi JN, Ma Y, Xu Z. Deriving neighborhood-level diet and physical activity measurements from anonymized mobile phone location data for enhancing obesity estimation. Int J Health Geogr 2022 Dec 30;21(1):22. [doi: 10.1186/s12942-022-00321-4] [Medline: 36585658]
99. Oladeji O, Zhang C, Moradi T, et al. Monitoring information-seeking patterns and obesity prevalence in Africa with internet search data: observational study. JMIR Public Health Surveill 2021 Apr 29;7(4):e24348. [doi: 10.2196/24348] [Medline: 33913815]
100. Stemmer M, Parmet Y, Ravid G. Identifying patients with inflammatory bowel disease on Twitter and learning from their personal experience: retrospective cohort study. J Med Internet Res 2022 Aug 2;24(8):e29186. [doi: 10.2196/29186] [Medline: 35917151]
101. Sathyanarayana A, Joty S, Fernandez-Luque L, et al. Sleep quality prediction from wearable data using deep learning. JMIR Mhealth Uhealth 2016 Nov 4;4(4):e125. [doi: 10.2196/mhealth.6562] [Medline: 27815231]
102. Chiang PH, Dey S. Offline and online learning techniques for personalized blood pressure prediction and health behavior recommendations. IEEE Access 2019;7:130854-130864. [doi: 10.1109/ACCESS.2019.2939218]
103. Zhong J, Liu W, Niu B, Lin X, Deng Y. Role of built environments on physical activity and health promotion: a review and policy insights. Front Public Health 2022;10:950348. [doi: 10.3389/fpubh.2022.950348] [Medline: 35910910]
104. Kimura N, Aota T, Aso Y, et al. Predicting positron emission tomography brain amyloid positivity using interpretable machine learning models with wearable sensor data and lifestyle factors. Alz Res THERAPY 2023;15(1). [doi: 10.1186/s13195-023-01363-x]
105. The R project for statistical computing. : R Foundation for Statistical Computing; 2022. URL: https://www.R-project.org/ [accessed 2026-02-20]
106. Wickham H, Averick M, Bryan J, et al. Welcome to the Tidyverse. JOSS 2019 Nov 21;4(43):1686. [doi: 10.21105/joss.01686]
107. Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Soft 2011;45(3). [doi: 10.18637/jss.v045.i03]
108. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. J Stat Soft 2010;36(11). [doi: 10.18637/jss.v036.i11]
109. Kuhn M. Building predictive models in R using the caret package. J Stat Soft 2008;28(5). [doi: 10.18637/jss.v028.i05]
110. van der Laan MJ, Polley EC, Hubbard AE. Super learner. Stat Appl Genet Mol Biol 2007;6(1). [doi: 10.2202/1544-6115.1309] [Medline: 17910531]
111. Lin S(. The "loneliness epidemic", intersecting risk factors and relations to mental health help-seeking: a population-based study during COVID-19 lockdown in Canada. J Affect Disord 2023 Jan;320:7-17. [doi: 10.1016/j.jad.2022.08.131]
112. Berthold MR, Cebron N, Dill F, et al. KNIME - the Konstanz information miner: version 2.0 and beyond. SIGKDD Explor Newsl 2009 Nov 16;11(1):26-31. [doi: 10.1145/1656274.1656280]
113. Castro R, Ribeiro-Alves M, Oliveira C, et al. What are we measuring when we evaluate digital interventions for improving lifestyle? A scoping meta-review. Front Public Health 2021;9:735624. [doi: 10.3389/fpubh.2021.735624] [Medline: 35047469]
114. Pasquetto IV, Borgman CL, Wofford MF. Uses and reuses of scientific data: the data creators' advantage. HDSR 2019;1(2) [FREE Full text] [doi: 10.1162/99608f92.fc14bf2d]
115. Aldoseri A, Al-Khalifa KN, Hamouda AM. Re-thinking data strategy and integration for artificial intelligence: concepts, opportunities, and challenges. Appl Sci (Basel) 2023 Jun 13;13(12):7082. [doi: 10.3390/app13127082]

XSL•FO
RenderX

116. Kim KH, Oh SW, Ko SJ, Lee KH, Choi W, Choi IY. Healthcare data quality assessment for improving the quality of the Korea Biobank Network. In: Meri A, editor. PLoS ONE 2023;18(11):e0294554. [doi: 10.1371/journal.pone.0294554] [Medline: 37983215]
117. Iqbal SMA, Mahgoub I, Du E, Leavitt MA, Asghar W. Advances in healthcare wearable devices. npj Flex Electron 2021;5(1):9. [doi: 10.1038/s41528-021-00107-x]
118. Mohanta B, Das P, Patnaik S. Healthcare 5.0: a paradigm shift in digital healthcare system using artificial intelligence, IOT and 5G communication. 2019 Presented at: 2019 International Conference on Applied Machine Learning (ICAML); May 25-26, 2019; Bhubaneswar, India p. 191-196. [doi: 10.1109/ICAML48257.2019.00044]
119. Kamel Boulos MN, Koh K. Smart city lifestyle sensing, big data, geo-analytics and intelligence for smarter public health decision-making in overweight, obesity and type 2 diabetes prevention: the research we should be doing. Int J Health Geogr 2021 Mar 3;20(1):12. [doi: 10.1186/s12942-021-00266-0] [Medline: 33658039]
120. Perez-Pozuelo I, Zhai B, Palotti J, et al. The future of sleep health: a data-driven revolution in sleep science and medicine. NPJ Digit Med 2020;3(1):42. [doi: 10.1038/s41746-020-0244-4] [Medline: 32219183]
121. Reinertsen E, Clifford GD. A review of physiological and behavioral monitoring with digital sensors for neuropsychiatric illnesses. Physiol Meas 2018 May 15;39(5):05TR01. [doi: 10.1088/1361-6579/aabf64] [Medline: 29671754]
122. Breiman L. Random forests. Mach Learn 2001 Oct;45(1):5-32. [doi: 10.1023/A:1010933404324]
123. Kuhn M, Johnson K. Feature Engineering and Selection: A Practical Approach for Predictive Models: CRC Press, Taylor & Francis Group; 2020.
124. Thompson NC, Ge S, Manso GF. The importance of (exponentially more) computing power. arXiv. Preprint posted online on Jun 28, 2022. [doi: 10.48550/ARXIV.2206.14007]
125. Arora A, Alderman JE, Palmer J, et al. The value of standards for health datasets in artificial intelligence-based applications. Nat Med 2023 Nov;29(11):2929-2938. [doi: 10.1038/s41591-023-02608-w]
126. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. jair 2002 Jun 1;16:321-357. [doi: 10.1613/jair.953]
127. Aliferis C, Overfitting SG. Underfitting and general model overconfidence and under-performance pitfalls and best practices in machine learning and AI. In: Simon GJ, Aliferis C, editors. Artificial Intelligence and Machine Learning in Health Care and Medical Sciences: Springer International Publishing; 2024:477-524. [doi: 10.1007/978-3-031-39355-6_10]
128. James G, Witten D, Hastie T, Methods TRR. Resampling Methods An Introduction to Statistical Learning: Springer; 2021:197-223. [doi: 10.1007/978-1-0716-1418-1_5]
129. Hyndman RJ, Athanasopoulos G. Forecasting: Principles and Practice, 2nd edition: OTexts; 2018. URL: https://otexts.org/fpp2/ [accessed 2026-02-20]
130. Saarela M, Podgorelec V. Recent applications of explainable AI (XAI): a systematic literature review. Appl Sci (Basel) 2024 Oct 2;14(19):8884. [doi: 10.3390/app14198884]
131. Rodríguez-Pérez R, Bajorath J. Interpretation of compound activity predictions from complex machine learning models using local approximations and Shapley values. J Med Chem 2020 Aug 27;63(16):8761-8777. [doi: 10.1021/acs.jmedchem.9b01101] [Medline: 31512867]
132. Overview. H2O.ai. 2024. URL: https://docs.h2o.ai/h2o/latest-stable/h2o-docs/index.html [accessed 2026-02-20]
133. Hosker DK, Elkins RM, Potter MP. Promoting mental health and wellness in youth through physical activity, nutrition, and sleep. Child Adolesc Psychiatr Clin N Am 2019 Apr;28(2):171-193. [doi: 10.1016/j.chc.2018.11.010] [Medline: 30832951]
134. Kris-Etherton PM, Sapp PA, Riley TM, Davis KM, Hart T, Lawler O. The dynamic interplay of healthy lifestyle behaviors for cardiovascular health. Curr Atheroscler Rep 2022 Dec;24(12):969-980. [doi: 10.1007/s11883-022-01068-w] [Medline: 36422788]
135. Kuhn M, Silge J. Tidy Modeling with R: A Framework for Modeling in the Tidyverse, 1st edition: O'Reilly; 2022.
136. Colaboratory BEG. Building Machine Learning and Deep Learning Models on Google Cloud Platform: Apress; 2019:59-64. [doi: 10.1007/978-1-4842-4470-8_7]
137. Han S, Olonisakin TF, Pribis JP, et al. A checklist is associated with increased quality of reporting preclinical biomedical research: a systematic review. In: Boltze J, editor. PLoS ONE 2017;12(9):e0183591. [doi: 10.1371/journal.pone.0183591]
138. Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res 2016 Dec 16;18(12):e323. [doi: 10.2196/jmir.5870] [Medline: 27986644]
139. Al-Zaiti SS, Alghwiri AA, Hu X, et al. A clinician's guide to understanding and critically appraising machine learning studies: a checklist for Ruling Out Bias Using Standard Tools in Machine Learning (ROBUST-ML). Eur Heart J Digit Health 2022 Jun;3(2):125-140. [doi: 10.1093/ehjdh/ztac016] [Medline: 36713011]
140. Stevens GA, Alkema L, Black RE, et al. Guidelines for accurate and transparent health estimates reporting: the GATHER statement. The Lancet 2016 Dec;388(10062):e19-e23. [doi: 10.1016/S0140-6736(16)30388-9]
141. Moritz S, Sardá A, Bartz-Beielstein T, Zaefferer M, Stork J. Comparison of different methods for univariate time series imputation in r. arXiv. Preprint posted online on Oct 13, 2015. [doi: 10.48550/ARXIV.1510.03924]

XSL•FO
RenderX

142. Mohsen F, Al-Absi HRH, Yousri NA, El Hajj N, Shah Z. A scoping review of artificial intelligence-based methods for diabetes risk prediction. NPJ Digit Med 2023 Oct 25;6(1):197. [doi: 10.1038/s41746-023-00933-5] [Medline: 37880301]
143. Safi SK, Gul S. An enhanced tree ensemble for classification in the presence of extreme class imbalance. Mathematics 2024 Oct 16;12(20):3243. [doi: 10.3390/math12203243]
144. Velarde G, Weichert M, Deshmunkh A, et al. Tree boosting methods for balanced and imbalanced classification and their robustness over time in risk assessment. Intelligent Systems with Applications 2024 Jun;22:200354. [doi: 10.1016/j.iswa.2024.200354]
145. Caruana R, Niculescu-Mizil A. An empirical comparison of supervised learning algorithms. 2006 Presented at: Proceedings of the 23rd international conference on Machine learning - ICML '06; Jul 24-28, 2006; Pittsburgh, PA p. 161-168 URL: http://portal.acm.org/citation.cfm?doid=1143844 [accessed 2026-02-20] [doi: 10.1145/1143844.1143865]
146. Arnold C, Biedebach L, Küpfer A, Neunhoeffer M. The role of hyperparameters in machine learning models and how to tune them. PSRM 2024 Oct;12(4):841-848. [doi: 10.1017/psrm.2023.61]
147. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. Statist Surv. Preprint posted online on 2009. [doi: 10.1214/09-SS054]
148. Tsamardinos I, Greasidou E, Borboudakis G. Bootstrapping the out-of-sample predictions for efficient and accurate cross-validation. Mach Learn 2018 Dec;107(12):1895-1922. [doi: 10.1007/s10994-018-5714-4]
149. Rainio O, Teuho J, Klén R. Evaluation metrics and statistical tests for machine learning. Sci Rep 2024 Mar 13;14(1):6086. [doi: 10.1038/s41598-024-56706-x]
150. Grandini M, Bagli E, Visani G. Metrics for multi-class classification: an overview. arXiv. Preprint posted online on Aug 13, 2020. [doi: 10.48550/ARXIV.2008.05756]
151. Lundberg S, Lee SI. A unified approach to interpreting model predictions. arXiv. Preprint posted online on May 22, 2017 URL: http://arxiv.org/abs/1705.07874 [accessed 2024-04-22]
152. Ribeiro MT, Singh S, Guestrin C. Model-agnostic interpretability of machine learning. arXiv. Preprint posted online on Jun 16, 2016. [doi: 10.48550/ARXIV.1606.05386]
153. Maksymiuk S, Gosiewska A, Biecek P. Landscape of R packages for explainable artificial intelligence. arXiv. Preprint posted online on Sep 24, 2020. [doi: 10.48550/ARXIV.2009.13248]
154. Ponce‐Bobadilla AV, Schmitt V, Maier CS, Mensing S, Stodtmann S. Practical guide to SHAP analysis: explaining supervised machine learning model predictions in drug development. Clinical Translational Sci 2024 Nov;17(11) [FREE Full text] [doi: 10.1111/cts.70056]
155. Lim H, Lee H. Eating habits and lifestyle factors related to childhood obesity among children aged 5-6 years: cluster analysis of panel survey data in Korea. JMIR Public Health Surveill 2024 Apr 5;10:e51581. [doi: 10.2196/51581] [Medline: 38578687]

Abbrevations

AI: artificial intelligence
DASS: Depression Anxiety Stress Scale
FFQ: Food Frequency Questionnaire
FNN: fuzzy neural network
GPAQ: Global Physical Activity Questionnaire
HL: healthy lifestyle
IMS-PAQ: Indian Migration Study Physical Activity Questionnaire
INPLASY: International Platform of Registered Systematic Review and Meta-Analysis Protocols
IPAQ: International Physical Activity Questionnaire
LIME: local interpretable model-agnostic explanations
ML: machine learning
NN: neural network
NutSo-HH: Nutritional and Social Healthy Habits Scale
PRESS: Peer Review of Electronic Search Strategies
PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses–Search extension
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RF: random forest
SHAP: Shapley Additive Explanation
SL: supervised learning
SMOTE: synthetic minority oversampling technique
SQUASH: Short Questionnaire to Assess Health-Enhancing Physical Activity
SVM: support vector machine
UL: unsupervised learning

XSL•FO
RenderX

XAI: explainable artificial intelligence

Edited by A Kushniruk; submitted 06.Jun.2025; peer-reviewed by A Adepoju, VG Yogeshappa; revised version received 06.Dec.2025; accepted 14.Dec.2025; published 27.Feb.2026.

Please cite as:
Estrella T, Capdevila L, Alfonso C, Losilla JM
Machine Learning for the Analysis of Healthy Lifestyle Data: Scoping Review and Guidelines
JMIR Hum Factors 2026;13:e78648
URL: https://humanfactors.jmir.org/2026/1/e78648
doi:10.2196/78648

© Tony Estrella, Lluis Capdevila, Carla Alfonso, Josep-Maria Losilla. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 27.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Predicting Electronic Health Record Usability: Scoping Review of Adoption Models, Metrics, and Future Directions

Dillon Chrimes[1], BSc, MSc, PhD; Alex Thomo[2], BSc, MSc, PhD; Mu-Hsing (Alex) Kuo[1], BSc, MSc, PhD; Elizabeth M Borycki[1], RN, HBScN, MN, PhD; Andre Kushniruk[1], BSc, MSc, PhD

[1]School of Health Information Science, Faculty of Health, University of Victoria, PO Box 1700 STN CSC, Victoria, BC, Canada

[2]Computer Science, Faculty of Engineering & Computer Science, University of Victoria, Victoria, BC, Canada

Corresponding Author:
Dillon Chrimes, BSc, MSc, PhD
School of Health Information Science, Faculty of Health, University of Victoria, PO Box 1700 STN CSC, Victoria, BC, Canada

Abstract

Background: Electronic health records (EHRs) play an essential role in modern health care, enabling data sharing and improving patient safety; however, even though vendors must adhere to International Organization for Standardization–related usability standards for EHR certification, persistent usability issues continue to undermine efficiency, contribute to clinician burden, and increase the risk of preventable errors.

Objective: This scoping review synthesizes existing research on EHR adoption and usability, emphasizing theoretical models, measurement approaches, factors, and analytic methods used to assess or predict usability. We identify gaps and opportunities for integrating predictive analytics and artificial intelligence (AI) to advance research and improve the usability of EHRs.

Methods: Following Joanna Briggs Institute and PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines, we systematically searched MEDLINE, Web of Science, IEEE Xplore, and Scopus library databases for studies published between January 1, 2009, and April 9, 2025. Inclusion criteria focused on empirical research using predictive methods or models related to EHR usability. Data were charted and synthesized thematically.

Results: Of the 2323 screened papers, 47 studies met inclusion criteria. Most research examined or predicted EHR adoption (not usability) using dominant frameworks, such as the technology acceptance model, unified theory of acceptance and use of technology, and the information system success model, which comprised usability. Factors related to usability—particularly perceived usefulness, perceived ease of use, effort expectancy, and facilitating conditions via the EHR adoption models—appeared frequently. Regression-based methods and structural equation modeling were the most common analytic techniques. No studies applied predictive modeling or AI to predict EHR usability.

Conclusions: The focus of this study on the prediction of EHR usability and adoption for the past 15 years is distinctive in the literature. It extends prior usability reviews (mostly focusing on adoption, not prediction of usability). Predictive modeling for EHR usability remains underdeveloped throughout 2009 to 2025. Dominant frameworks in the EHR literature continue to prioritize predicting adoption over operational usability. These models rely heavily on self-reported, cross-sectional measures captured at a single postimplementation time point, embedding systematic bias and obscuring longitudinal usability dynamics. Despite the application of increasingly sophisticated predictive techniques—primarily variants of regression and structural equation modeling—usability has remained analytically subordinate to adoption and acceptance constructs for more than 15 years. As a result, widely used models, such as the technology acceptance model and unified theory of acceptance and use of technology, position usability merely as an antecedent to intention or use, rather than as an independent, system-level property that can be empirically measured, modeled, and predicted. Therefore, there is substantial opportunity to integrate predictive analytics, AI, and longitudinal usability measures to build dynamic models.

(JMIR Hum Factors 2026;13:e86076) doi:10.2196/86076

KEYWORDS

artificial intelligence; electronic health records; perceived usefulness; predictive analytics; technology acceptance model; usability

Introduction

Electronic health records (EHRs) are central to modern health care delivery, enabling real-time access to patient information, supporting continuity of care, and enhancing clinical decision-making. However, persistent usability challenges continue to increase cognitive workload, reduce user satisfaction (US), and contribute to patient safety risks [1-5]. Usability can be defined as the extent to which a system can be used efficiently, effectively, and satisfactorily by specified users in

XSL•FO
RenderX

a specified context (International Organization for Standardization [ISO] 9241‐11) [6]. In contrast, adoption refers to the extent to which users accept and implement technology [7]. Usability can directly influence the uptake of the system; however, it is not the same as adoption.

Studies show that inadequate usability contributes to data entry errors, inefficient workflows, increased cognitive burden, and user frustration, thereby limiting the potential benefits of EHRs [8,9]. While EHR vendors adhere to certification processes, these often fail to capture human factors that shape system utility and usefulness [10,11]. Reliable measures of the usability of EHRs—particularly those that can predict performance before and after go-live—remain scarce, and even fewer studies have examined usability over extended durations post implementation of ≥1 year [12].

Over the past 2 decades, theoretical frameworks, such as technology acceptance model (TAM), unified theory of acceptance and use of technology (UTAUT), and the DeLone and McLean (D&M) information system success (ISS) model, have been widely applied to understand EHR adoption and acceptance. Core determinants in the constructs of these theoretical frameworks, such as perceived usefulness (PU) and perceived ease of use (PEOU), consistently predict user intentions and behaviors [13,14]. However, these adoption models often rely on cross-sectional surveys to collect data to run the model and do not comprise any ISO or industry standards, offering limited insight into how the EHR system performs, as usability evolves over time. Hill et al [7] concluded that despite the widespread adoption of EHR systems, a very small number of research papers analyzed the use, impact, and weaknesses of existing search functions for clinical tasks. Moreover, EHR usability changes dynamically with user experience and across the software development life cycle [15,16], which EHR adoption frameworks do not capture.

Many EHR interfaces remain cluttered, nonintuitive, and difficult to navigate, requiring ongoing training [1]. Such design limitations increase cognitive workload, prolong documentation time, and contribute to alert fatigue and workflow interruptions [17-19]. Physicians spend nearly half of their workday interacting with EHRs [20]. Alert fatigue, task switching, and interoperability failures further reduce usability standards and compromise care coordination [21,22].

A range of methods has been employed to assess usability. Heuristic evaluations allow for comparisons of EHRs against usability principles [23,24], while clinician-based testing highlights workflow inefficiencies and inadequate training [25]. Surveys reveal widespread dissatisfaction, with 70% of physicians citing usability concerns [26]. PU indicators include improved performance, productivity, and effectiveness [27], whereas PEOU indicators include learnability, clarity, flexibility, and ease of becoming skillful [28]. The D&M ISS model extends these perspectives with constructs such as system quality (SQ) and information quality (IQ) [29-31]. However, these models typically lack task-level analyses, despite evidence that inefficient task flows waste time, increase errors, and reduce physician satisfaction, altering the behavior, intent, and acceptance of adopting the EHR workflow [32-35].

Audit logs and system metadata provide opportunities to quantify usability through measures such as task completion time, error rates, system usability scale (SUS) scores, National Aeronautics and Space Administration-Task Load Index (NASA-TLX) scores for cognitive load, clickstream and keystroke logs, US surveys, and eye-tracking data [24,36,37]. Yet these metrics are not captured by EHR adoption models and are rarely used to predict specific usability problems or outcomes.

Emerging approaches in data mining, human-computer interaction (HCI), and artificial intelligence (AI) offer new opportunities for predictive modeling of usability. Techniques such as logistic regression, decision trees, random forests, gradient boosting (eg, XGBoost), support vector machines, and neural networks—including recurrent neural networks, convolutional neural networks, and long short-term memory networks—have been applied to analyze system logs, classify sessions, and model usability [38-40]. Predictors such as interaction complexity (eg, clicks per task), alert fatigue, workload peaks, demographic variables (eg, age and digital literacy), and prior EHR training have also been identified [19,21,25,31,37]. Integrating such factors into dynamic feedback loops could enable continuous monitoring and improvement for safety-enhanced design with vendors and end users in clinical settings. However, the extent to which these approaches have been systematically applied in EHR usability research remains to be fully studied.

Given the large datasets (eg, Office of the Coordinators Certified Health Product List dataset of usability testing results for EHR certification [41] that had been collecting data on EHR usability from 2014 until the present), this knowledge could be useful for predicting and anticipating potential usability problems with specific EHRs. Additionally, the data could be mined and used for supporting, selecting, and procuring usable EHRs for health care organizations that match their contexts, as well as for predicting and anticipating usability issues that may occur over time.

To address these gaps, we conducted a scoping review to examine any existing research that predicts the usability of EHRs with special emphasis on postimplementation. The main research objective was to examine and identify gaps in the literature on the prediction of EHR usability that could follow how theoretical frameworks, usability metrics, and analytic methods have been applied. By doing so, we aimed to critique existing models and highlight opportunities for predictive factors in EHR usability research and any further development of dynamic approaches to predict usability. These insights into the usability of EHRs over time and across systems are of value to health care organizations, system designers, EHR vendors, government certification bodies, and clinicians.

Methods

Design and Framework

Methodology and Data

This scoping review began on February 10, 2023. Our review followed the Joanna Briggs Institute (JBI) scoping review

XSL•FO
RenderX

methodology and adhered to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) reporting guidelines (Checklist 1) [42]. The sequential steps defined by Arksey and O'Malley [43] were applied: (1) identifying the research question, (2) identifying relevant studies, (3) selecting studies for inclusion, (4) charting the data, and (5) collating, summarizing, and reporting the results.

Four electronic databases were searched: MEDLINE (Ovid), Web of Science, IEEE Xplore, and Scopus. The MEDLINE database was selected because it comprises operational research on EHRs in clinical settings, nursing, health care, and related fields from 1946 to the present. Similarly, the Web of Science database was selected because it contains research on EHR usability in clinical settings, HCI journals, implementation science, and human factors research on EHRs across sciences, social sciences, arts, and humanities. The IEEE Xplore database was selected because it contains papers about HCI with EHRs and technical considerations of implemented EHRs related to usability across computer science, engineering, and electronics. Finally, the Scopus database was selected because it contains human factor research across scientific and technical fields (health, physical, and social science) plus arts and humanities.

Step 1: Identifying the Research Question

The research questions were as follows:

1. What is the state of the literature relevant to the prediction of EHR usability?
2. What studies have been conducted that predict or model EHR usability?
3. What factors are used to predict EHR usability?
4. Which methodologies have been used to evaluate or predict EHR usability?

Step 2: Protocol Registration, Screening Reliability, and Data Charting Procedures

This scoping review was conducted in accordance with the JBI methodology and reported using PRISMA-ScR guidelines. A formal review protocol defining the research questions, eligibility criteria, search strategy, and data charting framework was developed a priori and finalized prior to database screening. To ensure screening reliability, 3 independent reviewers (DC, EMB, and AK) conducted both title/abstract and full-text screening, with conflicts resolved through structured consensus meetings. Interreviewer agreement was monitored throughout the screening process, and discrepancies were discussed until full agreement was achieved, thereby strengthening internal validity. The reviewers AW, EMB, AK, AT, and DC have published and worked with EHRs, and all have experience and expertise as content experts and as methodological experts.

Data charting followed a structured extraction framework that systematically captured study design, theoretical models, analytic methods, usability metrics, predictors, and outcomes. Charting was conducted iteratively and audited across multiple cycles to ensure internal consistency and completeness. Inclusion and exclusion criteria were defined a priori with a clear rationale grounded in the study objective of identifying predictive—not purely descriptive—approaches to EHR usability, thereby justifying the exclusion of studies focused solely on organizational acceptance, vendor reports, non-EHR systems, or clinical outcome prediction without a usability construct. These safeguards ensured methodological transparency, reproducibility, and commitment to JBI and PRISMA-ScR standards.

Step 3: Identifying Relevant Studies

Eligibility criteria were defined using inclusion and exclusion rules that determine whether a study was relevant to the research question (Table 1). Inclusion and exclusion criteria were established before the first screening was carried out. The criteria specified study characteristics, such as publication timeframe, focus on usability, predictive methods, and EHR/electronic medical record (EMR) context for inclusion.

XSL•FO
RenderX

Table . Inclusion and exclusion criteria in the search strategy.

Domain	Inclusion criteria	Exclusion criteria
Modeling frameworks	• Adoption- or sociotechnical-related factors linked to usability • Regression analyses or modeling approaches aimed at predicting usability • Analytic approaches and measurable usability in EHR[a] adoption frameworks	• Papers describing organizational-level acceptance only • Papers linking usability to safety only • Health information technology (HIT) only
Electronic records	• EMR[b] and EHR, or broader health information system contexts • EMR and EHR adoption	• Dental systems, medication systems, and pharmacy • Personal health record • General information systems, eye care, and radiology • Surveillance • Laboratory information systems
EHR adoption and usability	• Technology acceptance or sociotechnical models relevant to usability • User-centered design, tasks, social-technical • User acceptance, patient safety • User satisfaction, human-computer interaction • Safety-enhanced design • Cognitive task load	• Usability survey only • Usability of other systems • User experience only
Informatics and data used in the model	• Usability data modeled • User experience data modeled • EHR adoption	• EHR data, Public Health Informatics • Patient data and EHR phenotyping • Medical conditions • Health information exchange • Clinical data
Artificial intelligence	• Data mining, regression, confirmatory analysis, hierarchical regression, and exploratory factor analysis • Text mining • Algorithm • Machine learning and deep learning	• Pattern recognition and large language models
Context	• Prediction or analysis EHR or EMR	• Clinical or solely clinical context
Time frame	• From January 2009 to April 9, 2025	• Published before 2009
Study design	• Usability research studies • Qualitative and quantitative analyses • Cognitive load with the EHR usability model • Usability data analysis of EHRs	• No literature reviews, commentary, or opinion • Papers without full text • EHR data analysis for medical outcomes • No technical reports • No position or opinion viewpoints • No vendor-based reports

[a]EHRs: electronic health records.

[b]EMR: electronic medical record.

These inclusion criteria were defined to capture studies using quantitative, qualitative, or mixed methods approach that assessed, modeled, or predicted aspects of EHR usability, EHR adoption, or sociotechnical factors primarily. The exclusion criteria removed research papers that were not related to EHRs, studies focused on usability but not prediction, and excluded literature reviews, opinion papers, or technical reports. All papers included were in English only.

Search Strategy

A comprehensive search was conducted across 4 databases—MEDLINE (Ovid), Web of Science, IEEE Xplore, and Scopus—to identify studies evaluating or predicting EHR usability. All searches were limited to English-language publications from January 1, 2009, to April 9, 2025. Keyword strategies incorporated terms related to usability, prediction or analytics, and EHRs.

XSL•FO
RenderX

The search strategy included combinations of controlled vocabulary terms and keywords related to usability ("usability," "user experience," "human factors," "human-computer interaction"), prediction ("predict*," "regression," "data mining," "machine learning"), and EHR ("EHR," "electronic health record," "EMR").

Search strategies for each of the 4 databases were as follows:

1. IEEE database ((((usability OR user OR "human factors" OR "human-computer interaction")))) AND (("predict*" OR "classif*" OR forecasting OR "data mining" OR "text mining" OR "regression*" OR "confirmatory factor analy*") AND (EHR OR "electronic health record*" OR EMR OR "electronic medical record*")).
2. Web of Science (usability or user or "human factors" or "human-computer interaction") AND (predict* or classif* or forecasting or "data mining" or "text mining" or "regression*" or "confirmatory factor analy*") AND (EHR or "electronic health record*" or EMR or "electronic medical record*").
3. MEDLINE (Ovid) ((usability or user or "human factors" or "human-computer interaction")).tw,kf. AND (("predict*" OR "classif*" OR "data mining" OR "regression*" OR "confirmatory factor analy*" OR algorithm*)).tw,kf. AND (EHR or "electronic health record*" or EMR or "electronic medical record*").tw,kf.
4. Scopus (TITLE-ABS ((usability OR user OR "human factors" OR "human-computer interaction")) AND TITLE-ABS (("predict*" OR "classif*" OR "data mining" OR "regression*" OR "confirmatory factor analy*")) AND TITLE-ABS (ehr OR "electronic health record*" OR emr OR "electronic medical record*")) AND (LIMIT-TO (LANGUAGE, "English")) AND (LIMIT-TO (SRCTYPE, "j") OR LIMIT-TO (SRCTYPE , "p")).

Screening, Data Extraction, and Analysis

Step 4: Study Selection

Title and abstract screening were completed by 3 reviewers, namely EMB, AW, and DC. Each reviewer independently screened the titles and abstracts of papers, with conflicts resolved by consensus during regular review meetings via Covidence. Title and abstract screening established the study section as a quick review of papers that the reviewers deemed included or excluded based on the inclusion and exclusion criteria. For papers that met the eligibility criteria, the papers were analyzed using a conventional content analysis approach to review the papers, chart the data, and identify themes.

For papers that did not meet the criteria, the reviewers noted this. Once the first screening was completed, the second full-text screening was then conducted to determine final eligibility. Studies were excluded during screening if they did not address usability or EHR-related usability via EHR adoption models; did not include predictive or analytic methods; did not directly examine EMR/EHR usability; lacked empirical methods or analytic rigor, or had the wrong study design.

The screening workflow ensured that only studies with explicit relevance to usability measurement, modeling, or prediction were retained.

Data extraction was carried out using a structured template. Extracted variables included study design, country and year, models or frameworks used, adoption determinants, usability-related metrics or instruments, analytic methods (eg, regression, SEM), main predictors examined, and key findings. Extraction fields were defined a priori to ensure consistency and comprehensiveness across studies.

Steps 5 and 6: Charting, Collating, and Reporting with AI Method Utilization

Study characteristics were manually charted by the research team. These characteristics included country of study, publication year, keywords, study design, analytic method, and predictive or explanatory techniques.

Extracted variables and categorizations were cross-checked using ChatGPT (GPT-5) to identify inconsistencies, ensure terminology alignment, and cross-check study characteristics. It should be noted that GPT-5 was used after manually updating the critique of the 47 journal papers (2009-2025) in Table S1 in Multimedia Appendix 1. It did not reproduce that table. For each of the papers selected for full review, GPT-5 was prompted with the instruction "summarize the study design, models, factors, significant findings and overall critique" to cross-check the content of the table's columns of models, factors, and critique. The response from GPT was mostly consistent, with only a few discrepancies found between GPT-5 and the human researcher. Any discrepancies were manually reviewed and updated thereafter. GPT-5 was only used for checking human processes in the review and critique of the papers found. GPT-5 was also used to obtain frequency counts for categories of predictive techniques in the papers. This was also cross-checked with frequency counting by the human researcher, with any discrepancies in the categories and the counts resolved through further analysis by the human researcher.

Ethical Considerations

This study was reviewed by the University of Victoria Human Research Ethics Board (UVic HREB) Chair and determined to be exempt from formal human research ethics review because the research does not involve human participants and is limited to the analysis of existing, publicly available documents and/or datasets. The determination was made in accordance with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS2 2022) and the University of Victoria Board of Governors Research Policy RH8100 and University of Victoria Board of Governors Regulations for Research Involving Humans RH8105. The UVic Human Research Ethics Chair concluded that the activities described in this study fall outside the scope of research requiring institutional ethics review because they do not involve human participants, human biological materials, or identifiable private information. The work is limited to the analysis of publicly available documentation and datasets.

Results

Study Selection and Characteristics

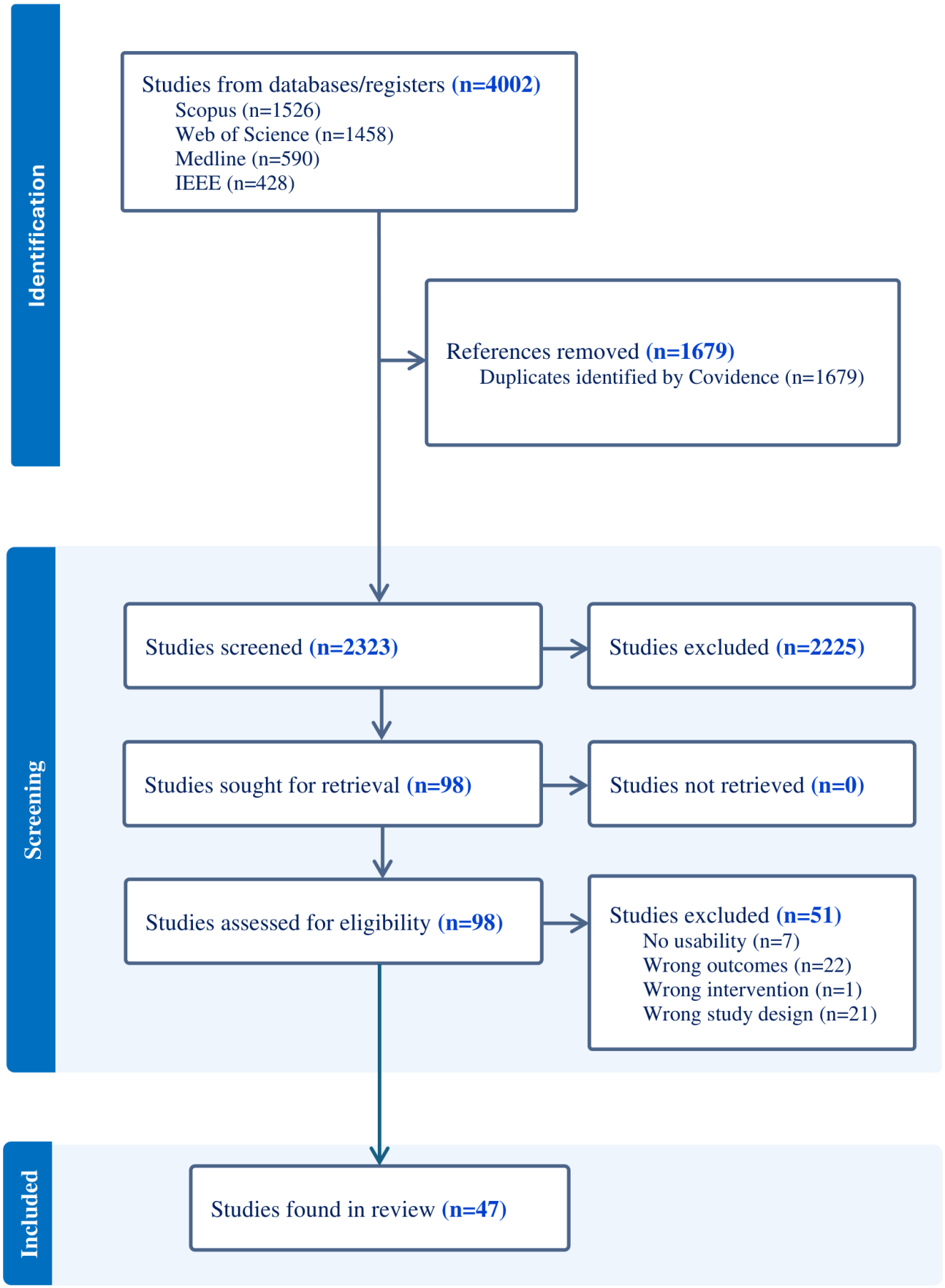
The included studies were drawn from 4 major databases: Scopus (n=1526), Web of Science (n=1458), MEDLINE

XSL•FO
RenderX

(n=590), and IEEE (n=428). Database searches yielded 428 papers in IEEE Xplore, 1458 in Web of Science, 590 in MEDLINE (Ovid), and 1526 in Scopus. After the removal of duplicates (n=1679), a total of 2323 records underwent title and abstract review in the first screening (Figure 1). A total of 98 full-text papers were assessed during the full-text screening. In the full paper screening, 51 records were excluded because of wrong outcomes (n=22), wrong study design (n=21), studies that did not address usability (n=7), and wrong intervention (n=1). Thus, a total of 47 studies met the full screening eligibility and were retained for analysis.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.



From 2323 records screened, 47 studies met the inclusion criteria. The majority were cross-sectional surveys that applied TAM, UTAUT, or ISS frameworks to evaluate EHR adoption and US. Common predictors included PU, PEOU, effort expectancy (EE), and facilitating conditions (FCs).

XSL•FO
RenderX

Cross-sectional designs were the most common (31/47), followed by survey-based studies (26/47) (Multimedia Appendix 2). Almost all studies were conducted postimplementation. Data collection methods most frequently included direct data capture (36/47) and questionnaires (31/47). Regression analysis was the most common modeling approach (26/47).

Study Characteristics, Metrics, and Data Sources

Across the 47 included studies, EHR adoption was most commonly predicted with usability measures that were predominantly subjective, relying on surveys and self-reported satisfaction. Objective indicators—such as task completion time, error rates, and audit log data—were rarely incorporated, and no study combined task-based HCI data with predictive analytics for longitudinal monitoring.

The reliance on survey-based methods, with limited integration of task-level or objective usability metrics, was a consistent finding. Acronym use across the studies (Table S1 in Multimedia Appendix 1) reflects a synthesis of 47 studies from 26 countries (2009 - 2025), largely dominated by TAM, UTAUT, and D&M ISS frameworks. Analytic approaches included regression, SEM, and in a small number of studies, machine learning or decision trees. While nearly every study examined adoption or acceptance, very few explicitly focused on prediction even with the EHR adoption constructs with significant determinants.

Model Applications

TAM and UTAUT frameworks in many studies were evaluated using regression and SEM. Some studies used confirmatory factor analysis (CFA) and exploratory factor analysis to validate constructs. A small number integrated machine learning or neural networks with SEM, but no study applied AI to predict EHR usability.

Regression was the most commonly applied predictive analytic method, followed by SEM, which was the most frequently applied analytic methods, supplemented by ANOVA, multivariate ANOVA, CFA, and exploratory factor analysis in 13 studies [18,44-55].

Common Determinants and Constructs

Table S1 in Multimedia Appendix 1 presents 26 reported adoption-related determinants and model constructs such as TAM, tripolar model of technology acceptance, theory of planned behavior, UTAUT, and Systems Engineering Initiative for Patient Safety (SEIPS) that had usability aspects of EHR/EMR that included (1) PU, PEOU, performance expectancy (PE), user behavior (UB), and US; (2) behavioral intention (BI), EE, FCs, and job relevance; (3) IQ, IT support, organizational culture, SQ, and service quality (SRQ); (4) broader usability frameworks and scales, such as the NASA-TLX, National Usability–Focused Health Information System Scale (NuHISS), and SUS.

Frequency of Metrics

Metrics were grouped into nine categories: (1) attitudes and social aspects, (2) behavioral and intention-related factors, (3) IQ, (4) organizational and environmental factors, (5) PU, (6) SRQ, (7) SQ, (8) US and perceived usability, and (9) work efficiency and productivity.

FC, defined as the degree to which an individual believes that an organization and its technical infrastructure exist to support the use of the system [56], was the most frequently measured determinant construct, or predictor variable, followed by PEOU, PU, and attitude. In total, more than 50 unique usability-related metrics were reported across the 15-year span of literature.

Temporal and Geographic Distribution

The studies spanned 16 years from 2009 to 2025, with 12 published before 2015 (Multimedia Appendix 3). Study duration was, on average, less than 1 year and ranged from 2 months to 14 years. No clear temporal trend was observed, though extending the screening into 2023 to 2025 revealed an increasing number of studies. However, databases searched in 2025 only included papers published in that year up to April 9.

Geographically, the 47 studies originated from 26 countries (Multimedia Appendix 4). The United States (10 studies) and Finland (5 studies) contributed the largest number of studies found. North American studies more often employed the D&M ISS model compared to other countries, while nursing-focused studies frequently adopted the UTAUT framework.

Keywords and Clusters

Most frequent keywords across studies were EHR, usability, TAM, and UTAUT (Multimedia Appendix 5).

Across all studies published between 2009 and 2025, no study directly predicted EHR usability using computational or longitudinal modeling techniques. Most research examined adoption outcomes, satisfaction ratings, or intention-to-use constructs rather than usability performance or workflow-based usability metrics. Psychometric determinants (eg, PU, PEOU, EE, and FC) were frequently evaluated across adoption frameworks.

Regression and SEM techniques predominated, supporting cross-sectional inferences but offering limited value for forecasting usability over time. Although FCs often emerged as a significant predictor of intention-to-use EHRs, they were rarely examined within dynamic contexts, task-based assessments, or sequential workflows that reflect actual clinical use.

There were 95 predictive techniques of analysis across 12 technique categories found, with many studies using multiple techniques (Multimedia Appendix 6). Regression-based methods dominated (40/95) [15,17,18,27,44,45,47-49,51,55,57-73] (Multimedia Appendix 7) followed by SEM (26/95) [50-55,74-77]. Furthermore, there were several regression and SEM variants in the studies found.

Only a limited number of studies incorporated AI-related techniques, but these were used for prediction of EHR adoption and not for usability. Almarzouqi et al [78] used neural networks with SEM to predict adoption, Alsyouf and Ishak [79] used decision trees with heuristic evaluation and UTAUT, and Sachdeva et al [74] used decision trees to discover behavior patterns. Cluster analysis was used to predict general BI [79] and evaluate EHR use in primary care providers [70]. Four studies used principal component analysis for dimensionality reduction and feature extraction [44,45,70,74].

XSL•FO
RenderX

Discussion

Key Findings

Across 47 studies, regression and SEM dominated, indicating minimal methodological progression toward predictive, automated, or real-time usability modeling.

This scoping review found that EHR adoption models dominated studies identified using usability and prediction-related keywords. However, adoption is not synonymous with usability, despite often being treated as interchangeable in the literature. As noted by Shachak et al [80], models such as TAM and UTAUT reduce complex sociotechnical properties of health information technology to individual perceptions and intentions, privileging acceptance over system-level performance and postimplementation use. Consequently, usability was assumed through adoption intent rather than directly measured or predicted.

The predominance of EHR adoption models in the reviewed studies underscores the need to address postimplementation system success. EHRs are complex sociotechnical systems in which usability emerges from interactions among interface design, configuration, organizational context, and user roles, making post hoc evaluation insufficient for meaningful intervention.

The continued use of TAM, UTAUT, and D&M ISS models highlights an enduring emphasis on explaining acceptance and satisfaction rather than modeling or forecasting usability performance in clinical practice. Developing predictive models could shift usability research from reactive post hoc assessment to proactive identification of inefficiencies, workflow burdens, and potential contributors to clinician burnout.

This scoping review highlights that although the TAM, UTAUT, and D&M ISS provide useful foundations for understanding adoption, they offer limited ability to capture evolving usability patterns with task complexity or changes in user interaction over time. Poor usability (clunky interface and workflow disruption) hinders adoption and satisfaction, causing burnout, whereas good usability (intuitive design and efficient workflows) drives adoption and better patient outcomes, even if a system is technically adopted. Thus, an EHR can be widely adopted but still have terrible usability that leads to frustration and workarounds. Therefore, predicting usability adds value to adoption and more insight as the system becomes operationalized.

The predictive focus distinguishes this study. Zhang and Walji [12] conceptualized EHR usability as a balance between intrinsic complexity (reflecting the work domain's inherent demands) and extrinsic difficulty (stemming from interface and workflow design). Their TURF framework—task, user, representation, and function—offers a roadmap linking usability determinants such as usefulness, satisfaction, and efficiency. Despite its potential, none of the included studies applied TURF to predictive modeling of EHR usability. EHR metadata, including audit or task logs, could serve as the foundation for predictive models of usability and clinician workload or burnout [34]. Significant predictors, such as PEOU and PU, highlight opportunities for AI-driven approaches that integrate adoption constructs with task-level metadata to advance predictive usability research.

Across the 16-year period reviewed (2009 - 2025), very few studies attempted to predict EHR usability directly, regardless of model sophistication. The facilitating conditions, such as the predictor variable of EHR adoption, appeared to be significant and impacted the intent to use EHRs. The FC determinant is defined as "the degree to which an individual believes that an organization and its technical infrastructure exist to support the use of the system" [56]. This psychometric determinant was deemed to be a useful metric for predicting adoption.

FCs have a direct positive effect on the intention-to-use, but this effect often diminishes after initial system use, suggesting its relevance may shift toward sociotechnical factors affecting learnability and decision-making. That is, the FCs would have linkages to sociotechnical aspects related to learnability, which is one of the factors in usability as defined by Nielsen [81]. The implications for understanding HCI and clinical decision-making suggest that survey-based constructs may insufficiently capture deeper usability determinants rooted in workflow, cognitive processes, and task structure, as mentioned by Kushniruk et al [82]. The review showed that several countries represented in their use of EHR adoption models over the past 15 years to find success or failure in the actual or intended use of EHRs postimplementation.

None of the reviewed models predicted EHR usability. In addition, no models incorporated detailed task-level inputs (eg, time-on-task, error rates, and interaction complexity) into predictions of intended or actual system use. Instead, many studies focused on PEOU or PU as attributes of usability and not tasks. Nevertheless, these attributes are important to establish (eg, easy to learn, efficient use, easy to remember, few errors, and subjectively pleasing) as recognized core usability attributes in the concept of usability engineering in the 1990s as defined by Nielsen [81]. Therefore, this result from the literature review makes it clear that predicting EHR adoption based on BI and actual use is still under investigation, with very limited modeling related to EHR usability. Importantly, the models used in the papers did not state that categories or aspects of usability and their usefulness could be dynamically utilized postimplementation.

Logistic regression or hierarchical regression via SEM with many predictive variables and hypotheses was studied in several of the papers. However, these approaches generally reflect relationships among perceptions rather than true indicators of usability performance. Aspects of task completion or failure appear directly related to usability, yet they were rarely used to establish predictive usability metrics. Furthermore, SEM in the studies [48,50-54,59,64,75,76,78,83-86] showed complex networks of hypotheses linked through latent constructs. This highlights the need for more evidence connecting specific usability issues (eg, coded problems from video-based testing [82]) with operational real-world EHR use, which takes time, and predictive models of EHR usability could reduce the burden of large and extensive usability testing.

XSL•FO
RenderX

Interpretation of Findings

The literature indicates an increasing sophistication in applying TAM, UTAUT, and D&M ISS constructs by researchers, often through SEM. Many predictive analytics applied within these models reflect the complexity of usability-related factors. Researchers frequently tested 10 to 15 hypotheses of TAM, UTAUT, or ISS in the analysis, demonstrating rigorous statistical approaches. However, the extensive statistical testing often highlighted variability in the acceptance or intention to use the system postimplementation rather than convergence toward models capable of predicting usability. A major trend was the use of TAM or UTAUT with SEM to evaluate multiple factors—even though many of these factors could, in principle, support predictive usability modeling, for example, by linking task-based observations with survey responses.

SEM's prominence reflects its origins in psychology and social sciences, where latent constructs such as attitudes and intentions have long been modeled through survey-based items. The Task–Technology Fit framework has also been used in SEM to assess alignment between tasks and IT [87]. CFA was commonly employed to evaluate usability parameters within hierarchical regression models [48]. However, the reviewed studies modeled adoption, satisfaction, or BI, which are important and should be done in unison with usability testing to evaluate its efficiency with UB and acceptance.

Global Distribution and Framework Use

The 47 studies, representing 26 countries, indicate global interest in EHR adoption research. Regional differences were minimal, although the D&M ISS model was used more frequently in North America. In 2024, Rihayha and Ismaili [52] proposed a modified human–organization–technology fit model for resource-limited settings, integrating SQ, IQ, and SRQ via SEM. Hyppönen et al [48] proposed adoption predictors, including user friendliness, perceived benefits, technical problems, and cooperation metrics, applying NuHISS in Finland. Welchen et al [88] applied NuHISS in Brazil through a star-schema confirmatory analysis. Even though NuHISS is very detailed, it lacks objective usability measures such as error rates and time-on-task. The use of NuHISS across multiple countries was a major finding, though no multicountry comparative studies were identified.

TAM and Its Determinants

Studies have updated versions of the TAM [16,50,57,58,69,76,79,89], TAM2 [66,84], and TAM3 (a hybrid of TAM and UTAUT) [63,66,78,86]. Important TAM usability determinants included PEOU and PU [16,44,48,50,54,57,58,60,63,64,66-68,70,78,83,85,89,90]. PU and PEOU can significantly predict health care professionals' intentions to use health information systems [13,14,58,73]. Additionally, subjective norms, relevance, and computer anxiety predict intentions beyond PU and PEOU [87,88]. One study showed that PEOU was influenced by computer self-efficacy, external control perceptions, anxiety, and enjoyment [55]. Chi and Ku [61] showed that voluntariness affected determinants of actual use, with EE increasing in importance as voluntariness decreased. Despite their importance, PU and PEOU may be undervalued in their significance for predicting BIs, especially when contextualized within newer constructs.

UTAUT Framework and Critique

UTAUT consistently demonstrated stronger predictive performance than TAM. Shiferaw and Mehari [53] reported that UTAUT explains up to 70 % of the variance in the intention-to-use technology. However, Bagozzi [91] critiqued the model in that it coheres the "many splinters of knowledge" to explain decision-making and its subsequent extensions, stating "UTAUT is a well-meaning and thoughtful presentation," but "it presents a model with 41 independent variables for predicting intentions and at least 8 independent variables for predicting behavior." Furthermore, Li et al [92] argued that UTAUT's reliance on moderators complicates model interpretation and may artificially inflate R^2 values. They also criticized construct labeling practices that combine disparate items into a single psychometric measure.

Nevertheless, UTAUT adds self-efficacy, social influence, and FCs, replacing TAM's PU and PEOU with PE and EE. For example, one research paper found that self-efficacy was a significant predictor variable in comparing urban, rural, and remote settings in the Philippines [49], indicating that one's ability to succeed or accomplish a task significantly and positively increased the intention to use an EHR in that setting. FC is a direct and commonly significant determinant of UB in the UTAUT framework with many subfactors [91-95].

EHR Usability and Related Frameworks

Several tools and frameworks have been developed to assess EHR usability. A few papers use the SUS and NASA-TLX [69,79]. However, this review identified no studies that incorporated ISO-based frameworks [96] as evaluation or predictive factors for EHR usability or adoption.

There was mention of the Health Information Technology Usability Evaluation Scale. This scale is specifically designed for evaluating the usability of health IT applications, including EHRs. It includes subscales for measuring PU, PEOU, and user control, among others [37]. Additionally, Carayon et al [97] proposed the SEIPS model, which provides a framework for understanding the relationships between technology, work processes, and patient outcomes in health care settings. This review identified a study by Alsyouf and Ishak [79] that used SEIPS to predict and evaluate the usability of EHR systems by identifying potential mismatches between the technology and the users' workflows.

The regression model by Butler and Johnson [60] did not predict EHR usability but rather productivity, which may be related to usability. Butler and Johnson [60] used an estimation framework for a productivity index based on physician characteristics, EMR functions, and vendors. Another study by Sulley and Ndanga [71] predicted US among physicians and different EHR system types or vendors and found that the EHR system type was significant. In addition, this study suggests that some EHR vendors can have higher satisfaction levels, which could be linked to the Butler and Johnson estimation of productivity using EHR. Moreover, Yin et al [73] used a logistic regression model and found that participants who are more likely to

XSL•FO
RenderX

consider EHR portals as valuable tools were those who found EHR portals useful for information.

Therefore, a regression model of productivity with usability factors and EHR components across vendors could be an alternative for predicting usability instead of adding external usability variables to more common EHR adoption frameworks, such as constructs of TAM, UTAUT, and D&M ISS models. Furthermore, cognitive aspects, such as the theory of reasoned action, theory of planned behavior, and a user's motivation to use IT, are influenced by the system features and capabilities in TAM, UTAUT, or D&M ISS model constructs. Further investigation is needed to predict EHR usability based on productive HCIs, the incorporation of aspects of EHR adoption frameworks, and other related models, such as Health Information Technology Usability Evaluation Scale, human factors engineering, SEIPS, and many more.

Future Work

Emerging technologies, such as AI, machine learning, and advanced data mining, offer opportunities to extend these models and utilize large datasets. There is almost no research using AI with TAM, UTAUT, and D&M ISS.

The EHR adoption models initially emerged several decades ago. AI algorithms could add feedback loops or feedforward recursive networks to adjust to different data over time, for example, sequences of tasks, usability, and actual use of EHRs under different scenarios. Aspects of EHR usability could be added to the TAM and UTAUT models with training and test datasets via machine learning potentially integrated within SEM and CFA frameworks. Machine-learning techniques have been applied to the usability of health information systems [98].

Integrating audit log data, clickstream analyses, and HCI measures can enable predictive modeling of usability performance, informing adaptive design and proactive system improvements. There is evidence of increasing use of EHR audit log files as evidence for the macrostructure of EHR work completed [99]. Furthermore, there is a viewpoint that there are emerging domains for measuring health care delivery with EHR metadata: (1) team structure and dynamics, (2) workflows, and (3) cognitive environment [100]. These domains contain modified index or scoring, continuity index, comprehensiveness index, conformity score, undivided attention, and task switching [100].

Furthermore, Rule et al [34] stated that metadata from EHR from audit logs can be used to predict physician burnout, which could be considered a usability metric. Moreover, even if there are more recent studies that utilize metadata from EHR system logs, a major limitation is the absence of explicit task-level variables via large datasets to predict usability over time. There is, nevertheless, capacity to include such task measures in SEM-based models as determinants of PE, self-efficacy, EE, and traditional TAM constructs such as PU and PEOU.

Therefore, researchers should consider combining established adoption frameworks with continuous usability monitoring and predictive analytics. Predictive approaches—leveraging design attributes, certification data, usage patterns, and organizational indicators—could enable early identification of usability risks, when redesign and mitigation remain feasible. Furthermore, predicting EHR usability could be used for analyzing large datasets now being collected on EHR usability, anticipating usability problems with specific systems and supporting selection and procurement decisions for acquiring more usable EHRs. The work in this paper could form the groundwork for establishing future efforts in analyzing and predicting usability. Future work could develop feedback loops that link UB, system tasks, and usability outcomes, providing actionable insights for EHR vendors and health care organizations.

Conclusion

No study in this review developed a predictive model for EHR usability postimplementation only, and very few studies used AI to predict adoption. To advance both EHR adoption and EHR usability sciences, researchers could bridge gaps between static adoption models and dynamic human factors engineering by integrating AI, audit log analysis, and real-time usability testing.

Acknowledgments

The authors declare the use of generative AI in the research and writing process. According to the GAIDeT (Generative AI Delegation Taxonomy), the following tasks were delegated to GAI tools: quality assessment. The GAI tool used was ChatGPT-5. Responsibility for the final manuscript lies entirely with the authors. GAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Funding

The authors declared no financial support was received for this work.

Conflicts of Interest

AK is Editor-in-Chief of *JMIR Human Factors* and EMB is Editor-in-Chief of *JMIR Nursing*. AK and EMB were not involved in the journal's review process or editorial decision-making for this paper. All remaining authors declare no conflict of interest.

Multimedia Appendix 1

Table S1. Critique of 47 journal articles (2009-2025) reviewed and based on analytical techniques, metrics, and predictive findings and factors critique.

XSL•FO
RenderX

[DOCX File, 61 KB - humanfactors_v13i1e86076_app1.docx]

Multimedia Appendix 2
Frequency of usability metric.
[PNG File, 44 KB - humanfactors_v13i1e86076_app2.png]

Multimedia Appendix 3
Frequency of studies per year from the literature review.
[PNG File, 43 KB - humanfactors_v13i1e86076_app3.png]

Multimedia Appendix 4
Frequency of studies per country from the literature review.
[PNG File, 98 KB - humanfactors_v13i1e86076_app4.png]

Multimedia Appendix 5
Most frequent keywords in the research articles from the literature review.
[PNG File, 70 KB - humanfactors_v13i1e86076_app5.png]

Multimedia Appendix 6
Predictive analytic techniques of the 47 studies of the literature categorized into 12 data mining categories: classification and machine learning, correlation and association, dimensionality reduction and factor methods, measurement and validation methods, model evaluation and fit testing, qualitative and mixed methods approaches, regression methods, rank-based test for group comparison, resampling and robustness methods, structural equation modeling and latent constructs, and time series analysis. Many studies had more than 1 predictive analytic technique related to data mining.
[XLSX File, 42 KB - humanfactors_v13i1e86076_app6.xlsx]

Multimedia Appendix 7
Number of different techniques of predictive analytics across selected literature. The initially derived list of techniques and respective categories is from Multimedia Appendix 2. SEM: structural equation modeling.
[PNG File, 62 KB - humanfactors_v13i1e86076_app7.png]

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 33 KB - humanfactors_v13i1e86076_app8.docx]

References

1. Cimino JJ. Putting the “why” in “EHR”: capturing and coding clinical cognition. J Am Med Inform Assoc 2019 Nov 1;26(11):1379-1384. [doi: 10.1093/jamia/ocz125] [Medline: 31407781]
2. Hammond WE, Bent B, West VL. Goodbye electronic health record? Stud Health Technol Inform 2022 Aug 31;298:107-111. [doi: 10.3233/SHTI220917] [Medline: 36073466]
3. Aaron S, McEvoy DS, Ray S, Hickman TTT, Wright A. Cranky comments: detecting clinical decision support malfunctions through free-text override reasons. J Am Med Inform Assoc 2019 Jan 1;26(1):37-43. [doi: 10.1093/jamia/ocy139] [Medline: 30590557]
4. Johnson KB, Neuss MJ, Detmer DE. Electronic health records and clinician burnout: a story of three eras. J Am Med Inform Assoc 2021 Apr 23;28(5):967-973. [doi: 10.1093/jamia/ocaa274] [Medline: 33367815]
5. Ellsworth MA, Dziadzko M, O’Horo JC, Farrell AM, Zhang J, Herasevich V. An appraisal of published usability evaluations of electronic health records via systematic review. J Am Med Inform Assoc 2017 Jan;24(1):218-226. [doi: 10.1093/jamia/ocw046] [Medline: 27107451]
6. ISO 9241-11 ergonomics of human-system interaction — part 11: usability: definitions and concepts. International Organization for Standardization (ISO). URL: https://www.iso.org/standard/63500.html [accessed 2026-03-16]
7. Hill JR, Visweswaran S, Ning X, Schleyer TK. Use, impact, weaknesses, and advanced features of search functions for clinical use in electronic health records: a scoping review. Appl Clin Inform 2021 May;12(3):417-428. [doi: 10.1055/s-0041-1730033] [Medline: 34261171]
8. Khairat S, Coleman C, Newlin T, et al. A mixed-methods evaluation framework for electronic health records usability studies. J Biomed Inform 2019 Jun;94:103175. [doi: 10.1016/j.jbi.2019.103175] [Medline: 30981897]
9. Carayon P, Hoonakker P. Human factors and usability for health information technology: old and new challenges. Yearb Med Inform 2019 Aug;28(1):71-77. [doi: 10.1055/s-0039-1677907] [Medline: 31419818]

XSL•FO
RenderX

10. Gomes KM, Ratwani RM. Evaluating improvements and shortcomings in clinician satisfaction with electronic health record usability. JAMA Netw Open 2019 Dec 2;2(12):e1916651. [doi: 10.1001/jamanetworkopen.2019.16651] [Medline: 31834390]
11. Ratwani RM, Reider J, Singh H. A decade of health information technology usability challenges and the path forward. JAMA 2019 Feb 26;321(8):743-744. [doi: 10.1001/jama.2019.0161] [Medline: 30715133]
12. Zhang J, Walji MF. TURF: toward a unified framework of EHR usability. J Biomed Inform 2011 Dec;44(6):1056-1067. [doi: 10.1016/j.jbi.2011.08.005] [Medline: 21867774]
13. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep 1;13(3):319-340. [doi: 10.2307/249008]
14. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172. [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
15. Dimitrovski T, Bath PA, Ketikidis P, Lazuras L. Factors affecting general practitioners' readiness to accept and use an electronic health record system in the Republic of North Macedonia: a national survey of general practitioners. JMIR Med Inform 2021 Apr 5;9(4):e21109. [doi: 10.2196/21109] [Medline: 33818399]
16. Gagnon MP, Ghandour EK, Talla PK, et al. Electronic health record acceptance by physicians: testing an integrated theoretical model. J Biomed Inform 2014 Apr;48:17-27. [doi: 10.1016/j.jbi.2013.10.010] [Medline: 24184678]
17. Beasley JW, Wetterneck TB, Temte J, et al. Information chaos in primary care: implications for physician performance and patient safety. J Am Board Fam Med 2011;24(6):745-751. [doi: 10.3122/jabfm.2011.06.100255] [Medline: 22086819]
18. Downing NL, Bates DW, Longhurst CA. Physician burnout in the electronic health record era: are we ignoring the real cause? Ann Intern Med 2018 Jul 3;169(1):50-51. [doi: 10.7326/M18-0139] [Medline: 29801050]
19. Sinsky C, Colligan L, Li L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. Ann Intern Med 2016 Dec 6;165(11):753-760. [doi: 10.7326/M16-0961] [Medline: 27595430]
20. Tipping MD, Forth VE, Magill DB, Englert K, Williams MV. Systematic review of time studies evaluating physicians in the hospital setting. J Hosp Med 2010;5(6):353-359. [doi: 10.1002/jhm.647] [Medline: 20803675]
21. Ancker JS, Edwards A, Nosal S, et al. Effects of workload, work complexity, and repeated alerts on alert fatigue in a Clinical Decision Support System. BMC Med Inform Decis Mak 2017 Apr 10;17(1):36. [doi: 10.1186/s12911-017-0430-8] [Medline: 28395667]
22. Cash JJ. Alert fatigue. Am J Health Syst Pharm 2009 Dec 1;66(23):2098-2101. [doi: 10.2146/ajhp090181] [Medline: 19923309]
23. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004 Feb;37(1):56-76. [doi: 10.1016/j.jbi.2004.01.003] [Medline: 15016386]
24. Zahabi M, Kaber DB, Swangnetr M. Usability and safety in electronic medical records interface design: a review of recent literature and guideline formulation. Hum Factors 2015 Aug;57(5):805-834. [doi: 10.1177/0018720815576827] [Medline: 25850118]
25. Ratwani RM, Savage E, Will A, et al. A usability and safety analysis of electronic health records: a multi-center study. JAMA 2018 Sep 1;25(9):1197-1201. [doi: 10.1093/jamia/ocy088]
26. Slight SP, Berner ES, Galanter W, et al. Meaningful use of electronic health records: experiences from the field and future opportunities. JMIR Med Inform 2015 Sep 18;3(3):e30. [doi: 10.2196/medinform.4457] [Medline: 26385598]
27. Tubaishat A. Perceived usefulness and perceived ease of use of electronic health records among nurses: application of technology acceptance model. Inform Health Soc Care 2018 Dec;43(4):379-389. [doi: 10.1080/17538157.2017.1363761] [Medline: 28920708]
28. Lee S. Effect of electronic medical record quality on nurses' perceived usefulness and ease of use. Comput Inform Nurs 2022 Aug 1;40(8):562-570. [doi: 10.1097/CIN.0000000000000845] [Medline: 35120366]
29. DeLone WH, McLean ER. Information systems success: the quest for the dependent variable. Inf Syst Res 1992 Mar;3(1):60-95. [doi: 10.1287/isre.3.1.60]
30. Melnick ER, Ong SY, Fong A, et al. Characterizing physician EHR use with vendor derived data: a feasibility study and cross-sectional analysis. J Am Med Inform Assoc 2021 Jul 14;28(7):1383-1392. [doi: 10.1093/jamia/ocab011] [Medline: 33822970]
31. Lau F, Hagens S, Muttitt S. A proposed benefits evaluation framework for health information systems in Canada. Healthc Q 2007;10(1):112-116. [Medline: 17326376]
32. Hundt AS, Adams JA, Carayon P. A collaborative usability evaluation (CUE) model for health IT design and implementation. International Journal of Human–Computer Interaction 2017 Apr 3;33(4):287-297. [doi: 10.1080/10447318.2016.1263430]
33. Eschenroeder HC, Manzione LC, Adler-Milstein J, et al. Associations of physician burnout with organizational electronic health record support and after-hours charting. J Am Med Inform Assoc 2021 Apr 23;28(5):960-966. [doi: 10.1093/jamia/ocab053] [Medline: 33880534]
34. Rule A, Kannampallil T, Hribar MR, et al. Guidance for reporting analyses of metadata on electronic health record use. J Am Med Inform Assoc 2024 Feb 16;31(3):784-789. [doi: 10.1093/jamia/ocad254] [Medline: 38123497]
35. Rule A, Melnick ER, Apathy NC. Using event logs to observe interactions with electronic health records: an updated scoping review shows increasing use of vendor-derived measures. J Am Med Inform Assoc 2022 Dec 13;30(1):144-154. [doi: 10.1093/jamia/ocac177] [Medline: 36173361]

XSL•FO
RenderX

36. Shah M, De Arrigunaga S, Forman LS, West M, Rowe SG, Mishuris RG. Cumulated time to chart closure: a novel electronic health record-derived metric associated with clinician burnout. JAMIA Open 2024 Apr;7(1):ooae009. [doi: 10.1093/jamiaopen/ooae009] [Medline: 38333109]
37. Yen PY, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012;19(3):413-422. [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
38. Shickel B, Tighe PJ, Bihorac A, Rashidi PDE. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE J Biomed Health Inform 2018 Sep;22(5):1589-1604. [doi: 10.1109/JBHI.2017.2767063] [Medline: 29989977]
39. Kotsiantis SB. Supervised machine learning: a review of classification techniques. Informatica 2007;31(3):249-268 [FREE Full text]
40. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. Nat Med 2019 Jan;25(1):24-29. [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]
41. Certified health IT product list (CHPL). US Department of Health and Human Services. URL: https://chpl.healthit.gov [accessed 2026-03-31]
42. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
43. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
44. Akwaowo CD, Sabi HM, Ekpenyong N, et al. Adoption of electronic medical records in developing countries-a multi-state study of the Nigerian healthcare system. Front Digit Health 2022;4:1017231. [doi: 10.3389/fdgth.2022.1017231] [Medline: 36479191]
45. Alsohime F, Temsah MH, Al-Eyadhy A, et al. Satisfaction and perceived usefulness with newly-implemented electronic health records system among pediatricians at a university hospital. Comput Methods Programs Biomed 2019 Feb;169:51-57. [doi: 10.1016/j.cmpb.2018.12.026] [Medline: 30638591]
46. Bossen C, Jensen LG, Udsen FW. Evaluation of a comprehensive EHR based on the DeLone and McLean model for IS success: approach, results, and success factors. Int J Med Inform 2013 Oct;82(10):940-953. [doi: 10.1016/j.ijmedinf.2013.05.010] [Medline: 23827768]
47. Campione J, Liu H. Perceptions of hospital electronic health record (EHR) training, support, and patient safety by staff position and tenure. BMC Health Serv Res 2024 Aug 20;24(1):955. [doi: 10.1186/s12913-024-11322-3] [Medline: 39164672]
48. Hyppönen H, Kaipio J, Heponiemi T, et al. Developing the National Usability-Focused Health Information System Scale for physicians: validation study. J Med Internet Res 2019 May 16;21(5):e12875. [doi: 10.2196/12875] [Medline: 31099336]
49. De Mesa RYH, Galingana CLT, Tan-Lim CSC, et al. Facing the digital frontier: exploring user acceptance of electronic health records in an urban, rural and remote setting in the Philippines. BMJ Open Qual 2024 Apr 18;13(2):e002621. [doi: 10.1136/bmjoq-2023-002621] [Medline: 38637020]
50. Mijin N, Jang H, Choi B, Khongorzul G. Attitude toward the use of electronic medical record systems: exploring moderating effects of self-image. Inf Dev 2019 Jan;35(1):67-79. [doi: 10.1177/0266666917729730]
51. Napitupulu TA, Patria SHJ. Factors that determine electronic medical records users satisfaction: a case of Indonesia. J Theor Appl Inf Technol 2013;58(3):499-505 [FREE Full text]
52. Rhayha R, Alaoui Ismaili A. Development and validation of an instrument to evaluate the perspective of using the electronic health record in a hospital setting. BMC Med Inform Decis Mak 2024 Oct 8;24(1):291. [doi: 10.1186/s12911-024-02675-0] [Medline: 39379909]
53. Shiferaw KB, Mehari EA. Modeling predictors of acceptance and use of electronic medical record system in a resource limited setting: using modified UTAUT model. Informatics in Medicine Unlocked 2019;17:100182. [doi: 10.1016/j.imu.2019.100182]
54. Turan AH, Koç T. Health information technology adoption and acceptance of Turkish physicians-a model proposal and empirical assessment. Health Informatics J 2022;28(2):1-23. [doi: 10.1177/14604582221096041] [Medline: 35636431]
55. Vainiomäki S, Aalto AM, Lääveri T, et al. Better usability and technical stability could lead to better work-related well-being among physicians. Appl Clin Inform 2017 Oct;8(4):1057-1067. [doi: 10.4338/ACI-2017-06-RA-0094] [Medline: 29241245]
56. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view1. MIS Q 2003 Sep 1;27(3):425-478. [doi: 10.2307/30036540]
57. Abdekhoda M, Ahmadi M, Dehnad A, Hosseini AF. Information technology acceptance in health information management. Methods Inf Med 2014;53(1):14-20. [doi: 10.3414/ME13-01-0079] [Medline: 24248003]
58. Addo K, Agyepong PK. Evaluating the health information system implementation and utilization in healthcare delivery. Health Informatics J 2024;30(4):14604582241304705. [doi: 10.1177/14604582241304705] [Medline: 39621578]
59. Beglaryan M, Petrosyan V, Bunker E. Development of a tripolar model of technology acceptance: hospital-based physicians' perspective on EHR. Int J Med Inform 2017 Jun;102:50-61. [doi: 10.1016/j.ijmedinf.2017.02.013] [Medline: 28495348]
60. Butler RJ, Johnson WG. Rating the digital help: electronic medical records, software providers, and physicians. Int J Health Econ Manag 2016 Sep;16(3):269-283. [doi: 10.1007/s10754-016-9190-8] [Medline: 27878677]

XSL•FO
RenderX

61. Chiu TML, Ku BPS. Moderating effects of voluntariness on the actual use of electronic health records for allied health professionals. JMIR Med Inform 2015 Feb 10;3(1):e7. [doi: 10.2196/medinform.2548] [Medline: 25720417]
62. Dubale AT, Mengestie ND, Tilahun B, Walle AD. User satisfaction of using electronic medical record system and its associated factors among healthcare professionals in Ethiopia: a cross-sectional study. Biomed Res Int 2023;2023(1):4148211. [doi: 10.1155/2023/4148211] [Medline: 37101688]
63. Ebnehoseini Z, Tara M, Tabesh H, Dindar FH, Hasibian S. Understanding key factors affecting on hospital electronic health record (EHR) adoption. J Family Med Prim Care 2020 Aug;9(8):4348-4352. [doi: 10.4103/jfmpc.jfmpc_109_20] [Medline: 33110858]
64. Iqbal U, Ho CH, Li YCJ, Nguyen PA, Jian WS, Wen HC. The relationship between usage intention and adoption of electronic health records at primary care clinics. Comput Methods Programs Biomed 2013 Dec;112(3):731-737. [doi: 10.1016/j.cmpb.2013.09.001] [Medline: 24091088]
65. Kaihlanen AM, Gluschkoff K, Hyppönen H, et al. The associations of electronic health record usability and user age with stress and cognitive failures among Finnish registered nurses: cross-sectional study. JMIR Med Inform 2020;8(11):876-885. [doi: 10.2196/23623] [Medline: 33206050]
66. Ketikidis P, Dimitrovski T, Lazuras L, Bath PA. Acceptance of health information technology in health professionals: an application of the revised technology acceptance model. Health Informatics J 2012 Jun;18(2):124-134. [doi: 10.1177/1460458211435425] [Medline: 22733680]
67. Kutney-Lee A, Sloane DM, Bowles KH, Burns LR, Aiken LH. Electronic health record adoption and nurse reports of usability and quality of care: the role of work environment. Appl Clin Inform 2019 Jan;10(1):129-139. [doi: 10.1055/s-0039-1678551] [Medline: 30786302]
68. Kutney-Lee A, Brooks Carthon M, Sloane DM, Bowles KH, McHugh MD, Aiken LH. Electronic health record usability: associations with nurse and patient outcomes in hospitals. Med Care 2021 Jul 1;59(7):625-631. [doi: 10.1097/MLR.0000000000001536] [Medline: 33797506]
69. Dunn Lopez K, Chin CL, Leitão Azevedo RF, et al. Electronic health record usability and workload changes over time for provider and nursing staff following transition to new EHR. Appl Ergon 2021 May;93:103359. [doi: 10.1016/j.apergo.2021.103359] [Medline: 33556884]
70. Messeri P, Khan S, Millery M, et al. An information systems model of the determinants of electronic health record use. Appl Clin Inform 2013;4(2):185-200. [doi: 10.4338/ACI-2013-01-RA-0005] [Medline: 23874357]
71. Sulley S, Ndanga M. Influence of EHR system type and physician demographics on satisfaction levels in health information technology. Discov Public Health 2025;22(1):119. [doi: 10.1186/s12982-025-00507-8]
72. Tilahun B, Fritz F. Modeling antecedents of electronic medical record system implementation success in low-resource setting hospitals. BMC Med Inform Decis Mak 2015 Aug 1;15(1):61. [doi: 10.1186/s12911-015-0192-0] [Medline: 26231051]
73. Yin R, Neyens DM, Law KE. Factors influencing users' perceived value of electronic health record patient portals. Presented at: Proceedings of the Human Factors and Ergonomics Society Annual Meeting was held as a virtual meeting; Oct 5-9, 2020 URL: https://journals.sagepub.com/doi/10.1177/1071181320641377 [accessed 2026-03-16] [doi: 10.1177/1071181320641377]
74. Sachdeva S, Madaan A, Bhalla S. Discovery of patterns to improve usability of electronic health record systems. J Inf Process 2012;20(1):327-339. [doi: 10.2197/ipsjjip.20.327]
75. Sharifian R, Askarian F, Nematolahi M, Farhadi P. Factors influencing nurses' acceptance of hospital information systems in Iran: application of the unified theory of acceptance and use of technology. HIM J 2014 Oct;43(3):23-28. [doi: 10.1177/183335831404300303]
76. Sintonen S, Mäkelä K, Miettinen R. User acceptance of electronic health records: a post-implementation study. Int J Healthc Technol Manag 2015;15(2):162. [doi: 10.1504/IJHTM.2015.074556]
77. Tilahun B, Fritz F. Comprehensive evaluation of electronic medical record system use and user satisfaction at five low-resource setting hospitals in Ethiopia. JMIR Med Inform 2015 May 25;3(2):e22. [doi: 10.2196/medinform.4106] [Medline: 26007237]
78. Almarzouqi A, Aburayya A, Salloum SA. Determinants predicting the electronic medical record adoption in healthcare: a SEM-Artificial Neural Network approach. PLOS One 2022;17(8):e0272735. [doi: 10.1371/journal.pone.0272735] [Medline: 35972979]
79. Alsyouf A, Ishak AK. Understanding EHRs continuance intention to use from the perspectives of UTAUT: practice environment moderating effect and top management support as predictor variables. Int J Electron Healthc 2018;10(1/2):24-59. [doi: 10.1504/IJEH.2018.092175]
80. Shachak A, Kuziemsky C, Petersen C. Beyond TAM and UTAUT: future directions for HIT implementation research. J Biomed Inform 2019 Dec;100:103315. [doi: 10.1016/j.jbi.2019.103315] [Medline: 31629923]
81. Nielsen J. Usability Engineering: Morgan Kaufmann Publisher; 1993. URL: https://www.sciencedirect.com/book/monograph/9780125184069/usability-engineering [accessed 2026-03-16]
82. Kushniruk AW, Monkman H, Kitson N, Borycki EM. Development of a video coding scheme for understanding human-computer interaction and clinical decision making. Stud Health Technol Inform 2019 Aug 9;265:80-85. [doi: 10.3233/SHTI190142] [Medline: 31431581]

XSL•FO
RenderX

83. Sayyah Gilani M, Iranmanesh M, Nikbin D, Zailani S. EMR continuance usage intention of healthcare professionals. Inform Health Soc Care 2017 Mar;42(2):153-165. [doi: 10.3109/17538157.2016.1160245] [Medline: 27100821]
84. Hadji B, Martin G, Dupuis I, Campoy E, Degoulet P. 14 years longitudinal evaluation of clinical information systems acceptance: the HEGP case. Int J Med Inform 2016 Feb;86:20-29. [doi: 10.1016/j.ijmedinf.2015.11.016] [Medline: 26725691]
85. Morton ME, Wiedenbeck S. EHR acceptance factors in ambulatory care: a survey of physician perceptions. Perspect Health Inf Manag 2010 Jan 1;7(Winter):1c. [Medline: 20697466]
86. Mwogosi A, Kibusi S. Effectiveness of EHR systems on decision support in primary healthcare: a technology acceptance model 3 perspective. J Health Organ Manag 2025 Apr 10;39(3):310-333. [doi: 10.1108/JHOM-07-2024-0296]
87. Goodhue DL, Thompson RL. Task-technology fit and individual performance. MIS Q 1995 Jun 1;19(2):213-236. [doi: 10.2307/249689]
88. Welchen V, Matte J, Giacomello CP, Dalle Molle F, Camargo ME. Usability perception of the health information systems in Brazil: the view of hospital health professionals on the electronic health record. RAUSP Manag J 2022 Jul 12;57(3):264-279. [doi: 10.1108/RAUSP-02-2021-0023]
89. Tamadur S, Akyürek Ç, Al-Shdayfat NM, Alsaqqa HH. Assessment of factors influencing healthcare provider's acceptance of electronic health record in Jordanian hospitals. J Health Manag 2024. [doi: 10.1177/09720634241278863]
90. Sicotte C, Paré G, Moreault MP, Lemay A, Valiquette L, Barkun J. Replacing an inpatient electronic medical record. Lessons learned from user satisfaction with the former system. Methods Inf Med 2009;48(1):92-100. [doi: 10.3414/ME0557] [Medline: 19151889]
91. Bagozzi R. The legacy of the technology acceptance model and a proposal for a paradigm shift. J Assoc Inf Syst 2007;8(4). [doi: 10.17705/1jais.00122]
92. Li H, Gupta A, Zhang J, Sarathy R. Examining the decision to use standalone personal health record systems as a trust-enabled fair social contract. Decis Support Syst 2014 Jan;57:376-386. [doi: 10.1016/j.dss.2012.10.043]
93. Kalavani A, Kazerani M, Shekofteh M. Acceptance of evidence based medicine (EBM) databases by Iranian medical residents using unified theory of acceptance and use of technology (UTAUT). Health Policy Technol 2018 Sep;7(3):287-292. [doi: 10.1016/j.hlpt.2018.06.005]
94. Im I, Hong S, Kang MS. An international comparison of technology adoption: testing the UTAUT model. Inf Manag 2011 Jan;48(1):1-8. [doi: 10.1016/j.im.2010.09.001]
95. Borycki EM, Kushniruk AW. Towards an integrative cognitive-socio-technical approach in health informatics: analyzing technology-induced error involving health information systems to improve patient safety. Open Med Inform J 2010 Sep 15;4:181-187. [doi: 10.2174/1874431101004010181] [Medline: 21594010]
96. Ergonomics of human-system interaction—Part 210: human-centred design for interactive systems. International Organization for Standardization. 2019. URL: https://www.iso.org/standard/77520.html [accessed 2026-03-16]
97. Carayon P, Schoofs Hundt A, Karsh BT, et al. Work system design for patient safety: the SEIPS model. Qual Saf Health Care 2006 Dec;15 Suppl 1(Suppl 1):i50-i58. [doi: 10.1136/qshc.2005.015842] [Medline: 17142610]
98. Kim S, Lou SS, Baratta LR, Kannampallil T. Classifying clinical work settings using EHR audit logs: a machine learning approach. AM J Manag Care 2023 Jan 1;29(1):e24-e30. [doi: 10.37765/ajmc.2023.89310]
99. Lou SS, Liu H, Harford D, Lu C, Kannampallil T. Characterizing the macrostructure of electronic health record work using raw audit logs: an unsupervised action embeddings approach. J Am Med Inform Assoc 2023 Feb 16;30(3):539-544. [doi: 10.1093/jamia/ocac239] [Medline: 36478460]
100. Tawfik D, Rule A, Alexanian A, et al. Emerging domains for measuring health care delivery with electronic health record metadata. J Med Internet Res 2025 Mar 6;27:e64721. [doi: 10.2196/64721] [Medline: 40053814]

Abbreviations

AI: artificial intelligence
BI: behavioral intention
CFA: confirmatory factor analysis
EE: effort expectancy
EHR: electronic health record
EMR: electronic medical record
FC: facilitating condition
HCI: human-computer interaction
IQ: information quality
ISO: International Organization for Standardization
ISS: DeLone and McLean information system success
JBI: Joanna Briggs Institute
NASA-TLX: National Aeronautics and Space Administration-Task Load Index
NuHISS: National Usability-Focused Health Information System Scale
PE: performance expectancy

XSL•FO
RenderX

PEOU: perceived ease of use
PRISMA-ScR : Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
PU: perceived usefulness
SEIPS: Systems Engineering Initiative for Patient Safety
SEM: structural equation modeling
SQ: system quality
SRQ: service quality
SUS: system usability scale
TAM: technology acceptance model
TMTA: Tripolar Model of Technology Acceptance
TURF: task, user, representation, and function
UB: user behavior
US: user satisfaction
UTAUT: unified theory of acceptance and use of technology

Edited by J Sarvestan; submitted 17.Oct.2025; peer-reviewed by S Mohanadas, S Choemprayong; revised version received 25.Feb.2026; accepted 26.Feb.2026; published 08.Apr.2026.

Please cite as:
Chrimes D, Thomo A, Kuo MH(, Borycki EM, Kushniruk A
Predicting Electronic Health Record Usability: Scoping Review of Adoption Models, Metrics, and Future Directions
JMIR Hum Factors 2026;13:e86076
URL: https://humanfactors.jmir.org/2026/1/e86076
doi:10.2196/86076

© Dillon Chrimes, Alex Thomo, Mu-Hsing (Alex) Kuo, Elizabeth Borycki, Andre Kushniruk. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 8.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Psychometric Properties of Digital Divide Measurement Instruments in Health Contexts: Systematic Review

Noelia Navas-Echazarreta[1], PhD; Elena Chover-Sierra[2,3], PhD; César Rubio-Tarazón[2], BSN; Raquel María Martínez-Pascual[2], BSN; Pablo Del Pozo-Herce[1,4], PhD; Jesús Martínez-Tofé[1], PhD; Raúl Juárez-Vela[1], PhD; Antonio Martínez-Sabater[1,2,5], PhD

[1]Research Group GRUPAC, Pre-departmental Biomedical Sciences and Health Specialties Unit, Faculty of Health Sciences, University of La Rioja, C/ Duquesa de la Victoria, 88, Logroño, Spain

[2]Nursing Care and Education Research Group (GRIECE), GIUV2019-456, Nursing Department, University of Valencia, Valencia, Spain

[3]Internal Medicine Department, Consorci Hospital General Universitari de Valencia, Valencia, Spain

[4]Research Group on Innovation in Health Care and Nursing Education (INcUidE), University of UNIE, Madrid, Spain

[5]Care Research Group (INCLIVA), University Clinical Hospital of Valencia, Valencia, Spain

Corresponding Author:
Noelia Navas-Echazarreta, PhD
Research Group GRUPAC, Pre-departmental Biomedical Sciences and Health Specialties Unit, Faculty of Health Sciences, University of La Rioja, C/ Duquesa de la Victoria, 88, Logroño, Spain

Abstract

Background: The digital divide constitutes a growing form of social inequality, influencing access to, use of, and benefit from digital health technologies. Accurate assessment of this phenomenon requires psychometrically sound measurement instruments, especially in health care contexts.

Objective: This systematic review aimed to identify and evaluate the psychometric properties of instruments developed to measure the digital divide, with a particular focus on their application in health-related settings.

Methods: A systematic review was conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 and COSMIN (Consensus-Based Standards for the Selection of Health Measurement Instruments) guidelines. Literature searches were performed in PubMed, Scopus, Dialnet, and Web of Science for studies published between 2015 and 2025. Inclusion criteria required studies to report a psychometric evaluation of instruments measuring the digital divide.

Results: Four studies met the inclusion criteria (ranging from 875 to 1337 participants). The identified scales primarily measured digital skills and capital. All studies reported adequate internal consistency, with Cronbach α coefficients ranging from 0.70 to 0.94. Structural validity was confirmed through factor analysis in all studies. However, data on temporal stability and responsiveness were absent.

Conclusions: Existing instruments show potential for assessing digital inequality in health, but further psychometric validation and cultural adaptation are required. Nursing professionals are encouraged to lead efforts in developing and validating context-sensitive instruments to promote digital health equity.

Trial Registration: PROSPERO CRD420251068347; https://www.crd.york.ac.uk/PROSPERO/view/CRD420251068347

(JMIR Hum Factors 2026;13:e82005) doi:10.2196/82005

KEYWORDS

digital divide; psychometric properties; measurement instruments; digital health; health inequalities; systematic review; nursing

Introduction

Digital transformation has profoundly altered the way in which individuals interact with information, communication, and self-management of their health [1]. However, this evolution has not been uniform across all citizens. While some have benefited from these new digital tools, others have been left behind. This inequality is known as the digital divide, a multidimensional phenomenon encompassing disparities in access, use, digital skills, and benefits derived from technology [2].

This multidimensional concept involves not only inequalities in access to devices or connectivity (first-level divide) but also limitations in the competencies required to use technologies effectively (second-level divide) and disparities in the actual benefits obtained, such as improvements in health outcomes (third-level divide) [2,3]. Consequently, the digital divide acts as an emerging social determinant, disproportionately affecting

XSL•FO
RenderX

older adults, individuals with low educational attainment, socioeconomically disadvantaged groups, and those with limited digital literacy [4,5].

Within the health care sector, the digital divide compromises access to digital services such as telemedicine, mobile health apps, self-management tools, and eHealth resources. This situation hinders the implementation of technology-based interventions and limits the effectiveness of public policies aimed at promoting equity [6-8]. These limitations particularly impact older adults, individuals with lower educational levels, those in vulnerable socioeconomic situations, and individuals with cognitive impairments [5]. This digital gap critically undermines the efficacy of telehealth interventions as patients lacking necessary digital skills or capital cannot effectively engage with virtual care models, thereby exacerbating existing health inequities [4,5,7].

In this regard, the COVID-19 pandemic exacerbated exclusion when a substantial portion of health care services transitioned to virtual environments. These difficulties in health care delivery were particularly evident among groups with low digital literacy or communication disabilities [6].

Despite the growing interest in this phenomenon, the measurement of the digital divide remains a challenge. Although scales and instruments assessing digital literacy or technological competencies exist, their specific psychometric validity, comprehensive structural approach, and cross-cultural applicability are not always ensured [7,8]. This situation undermines comparability across studies, limits accurate assessment of disparities, and hampers the design of evidence-based interventions [9].

Research conducted by Kayser et al [5] and Mota [10] relied on instruments focused on assessing digital literacy, attitudes, or knowledge regarding technology; however, these tools fail to adequately address the structural dimensions of the digital divide. Moreover, in several studies, such as that by Jongebloed et al [7], questionnaires lacking proper psychometric validation were used, thereby compromising the reliability of findings and restricting their utility in future research. Consequently, the guidelines proposed by Beaton et al [8] and the COSMIN (Consensus-Based Standards for the Selection of Health Measurement Instruments) checklist developed by Mokkink et al [9] are essential for the robust evaluation of measurement properties.

The availability of psychometrically sound instruments adapted to diverse sociocultural contexts is crucial to quantify the digital divide [5], understand its implications for health, and guide effective strategies for its reduction, particularly within nursing and public health, which play key roles in fostering digital equity.

In this context, this study aimed to conduct a systematic review of scientific literature to identify and analyze the psychometric properties of instruments designed to measure the digital divide, with special emphasis on their applicability in health care settings and their potential use in Spanish-speaking populations.

Methods

Study Design

This study was designed as a systematic review of scientific literature conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines and the COSMIN framework [9,11] with the objective of comprehensively analyzing the scales used to measure the digital divide, with a particular focus on the health care context. The review protocol was registered in the PROSPERO database (registration number CRD420251068347).

The review focused on the structural and methodological dimensions of the instruments, as well as on their key psychometric properties, encompassing reliability, validity in all its forms (content, construct, convergent, discriminant, and predictive or concurrent), and structural fit assessed through exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), in addition to the Rasch item response model [12].

The rigorous application of the PRISMA and COSMIN guidelines ensured process transparency, minimized methodological bias, and guaranteed quality in the selection and analysis of the relevant scientific literature [9,11].

Search Strategy

The search strategy was designed to ensure a broad and exhaustive retrieval of scientific literature related to instruments measuring the digital divide.

A systematic search was conducted across 4 scientifically recognized databases—Scopus, Dialnet, PubMed, and Web of Science—aimed at covering both English-language literature and studies conducted in Spanish-speaking contexts. The databases were searched with the time window set from 2015 to 2025. The final search across all databases was executed on January 15, 2025.

Boolean search strings were constructed by combining controlled and free-text terms primarily in English. These terms were organized into 3 blocks interconnected using the Boolean operators “AND,” “OR,” and “NOT.” The first block referred to terms related to the digital divide (“digital divide” and “digital inequality”), the second referred to instruments or related tools (“scale,” “instrument,” and “tool”), and the third referred to the measurement of psychometric properties (“psychometric properties,” “validity,” and “measurement properties”).

Document retrieval and bibliographic management were performed using the RefWorks software (Clarivate Analytics). This information is summarized in Table 1, which details the databases searched, the specific Boolean strings applied, the total number of documents identified, and the number of documents ultimately selected for this research.

XSL•FO
RenderX

Table . The search strategy.

Database and search string	Documents retrieved, n	Documents selected, n
Scopus		
(“digital divide” or “digital inequality” or “digital gap”) and (“scale” or “instrument” or “measure” or “tool”) and (“psychometric Properties” or “validity” or “reliability” or “internal consistency” or “measurement properties”)	68	3
(“digital divide” or “digital inequality” or “digital gap”) and (“scale” Or “instrument”) and (“validity” or “reliability”)	35	1
Dialnet		
(“digital divide” or “digital inequality” or “digital gap”) and (“scale” or “instrument” or “measure” or “tool”) and (“psychometric Properties” or “validity” or “reliability” or “internal consistency” or “measurement properties”)	11	0
(“brecha digital” or “desigualdad digital”) and (“instrumento” or “escala” or “cuestionario”) and (“validez” or “fiabilidad” or “propiedades psicométricas”)	9	0
PubMed		
(“digital divide” or “digital inequality” or “digital gap”) and (“scale” or “instrument” or “measure” or “tool”) and (“psychometric Properties” or “validity” or “reliability” or “internal consistency” or “measurement properties”)	14	1
WOS[a]		
(“digital divide” or “digital inequality” or “digital gap”) and (“scale” or “instrument” or “measure” or “tool”) and (“psychometric Properties” or “validity” or “reliability” or “internal consistency” or “measurement properties”)	55	3
(“digital divide” or “digital gap”) and (“scale” or “measure”) and (“validity” or “reliability” or “psychometric properties”)	32	3

[a]WOS: Web of Science.

Selection Criteria

The selection of studies included in this review was conducted through a careful, coherent, and progressive process. Studies published in English or Spanish, available in full text, and published within the time frame of 2015 to 2025 were included, coinciding with the increased research on the digital divide over the last decade, particularly following the acceleration of digitalization triggered by the COVID-19 pandemic.

Eligible articles were required to focus on the development, validation, adaptation, or application of structured instruments designed to measure the digital divide and provide data on their psychometric properties, including reliability, validity, internal consistency, or factor analysis.

Articles that did not directly address the measurement of the digital divide through psychometric instruments, as well as narrative reviews, opinion papers, and similar works, were excluded. Studies focusing exclusively on digital literacy without addressing digital inequalities or lacking psychometric data were also excluded. Selection was based on these inclusion and exclusion criteria following a detailed review of titles, abstracts, and full texts of articles meeting the eligibility requirements.

Effect Measures

The psychometric properties of the identified scales were assessed using the COSMIN guidelines [9]. Effect measures included indicators of reliability (Cronbach α coefficient, ω coefficient, Kuder-Richardson Formula 20, and intraclass correlation coefficient [ICC]); content, construct, convergent, discriminant, and predictive or concurrent validity; factor analysis (EFA and CFA); and estimates derived from the Rasch item response model [12].

Unlike classical test theory, which focuses on total scores and sample-dependent statistics, item response theory (IRT) and Rasch modeling allow for the evaluation of individual item properties independent of the specific sample. Consequently, Rasch analysis is preferred for its ability to verify measurement invariance and transform ordinal data into linear interval

XSL•FO
RenderX

measures, offering a more rigorous validation than classical test theory alone [9,12]. Each property was qualitatively classified according to the criteria by Terwee et al [13] as sufficient (+), insufficient (−), or indeterminate (?). This categorization facilitated the assessment of the methodological robustness of the instruments included in the review and supported the narrative synthesis of the findings [13].

Study Selection Process

Studies were included if they were relevant to the research question, met the inclusion criteria, and demonstrated adequate methodological quality. Two reviewers (AM-S and CR-T) independently screened the titles and abstracts of retrieved documents to identify potentially eligible studies. The full texts of these studies were then independently assessed by 2 other reviewers (CR-T and NN-E). Discrepancies were resolved through discussion, with a third reviewer (AM-S) consulted when necessary. Interrater agreement was measured using the ICC.

Study Selection and Data Extraction

Evaluated aspects included risk of bias, methodological design, and reporting quality. The PRISMA and COSMIN frameworks were applied to assess study quality. Data were identified, verified, and extracted independently by 2 authors (CR-T and NN-E). Extracted variables included author, year, country, instrument name, target population, number of items, reported psychometric properties, and applied statistical methods [9,11].

Data Summarization

Results were synthesized narratively and organized into descriptive tables detailing the main characteristics, psychometric properties, and scientific quality of the identified instruments. Findings were grouped according to property type and corresponding evaluations.

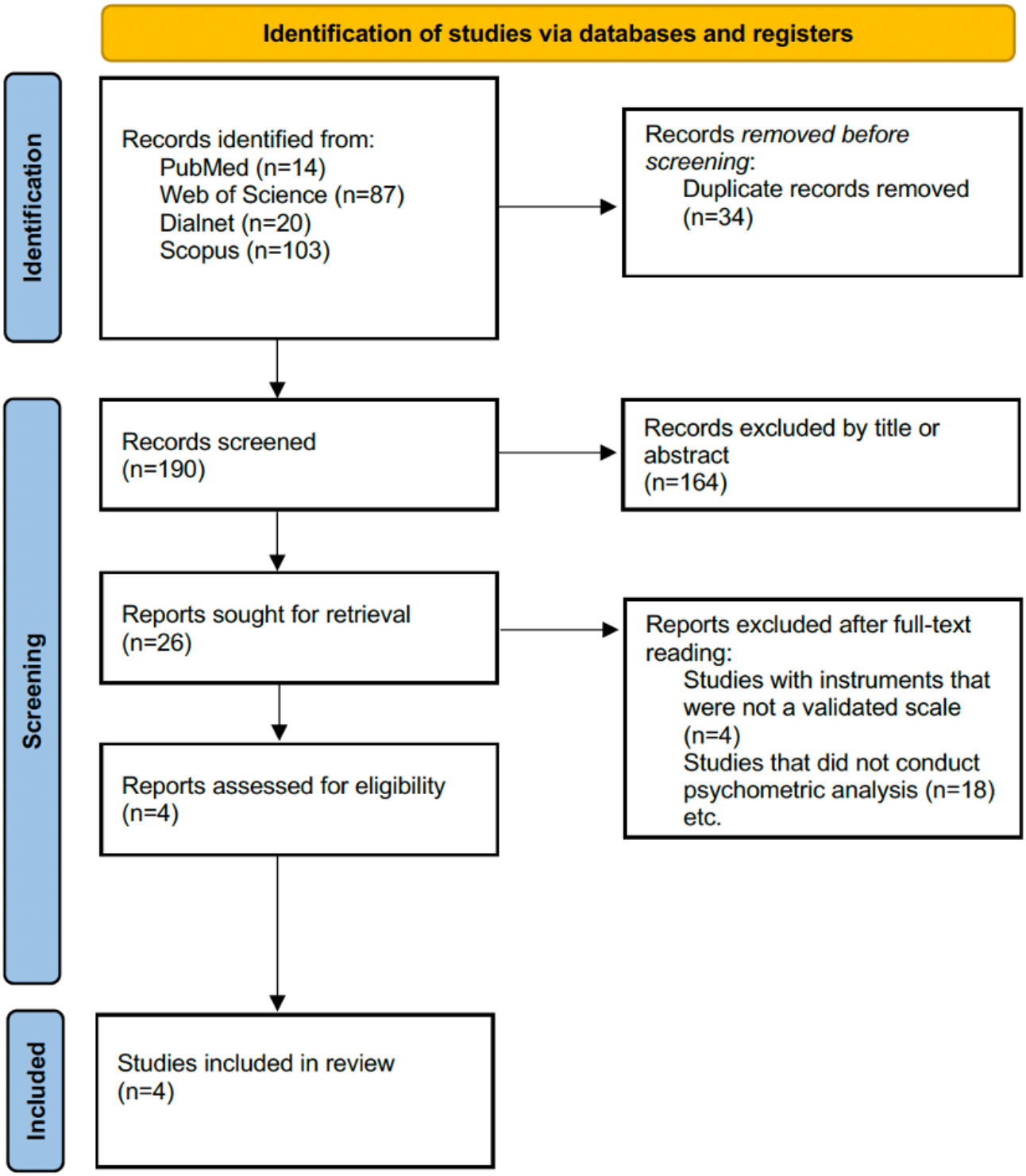
Results

Overview

Through a systematic bibliographic search across scientific databases, 224 records were initially identified. Of these 224 records, after removing 34 (15.2%) duplicates using RefWorks, 190 (84.8%) remained. Title and abstract screening resulted in the exclusion of 86.3% (164/190) of the remaining studies for not meeting the inclusion criteria. Consequently, 26 studies were selected for full-text review, of which 22 (88.5%) were excluded due to lack of instruments, absence of psychometric analysis, or limited thematic relevance. Ultimately, 4 studies met the inclusion criteria and were included in the systematic review, as illustrated in the PRISMA flow diagram (Figure 1). The ICC between reviewers was 0.80 (95% CI 0.68 - 0.88).

XSL•FO
RenderX

Figure 1. Selection process flow diagram.



Studies Included

The 4 studies analyzed were published between 2015 and 2023. These investigations were conducted in the United Kingdom [14], the Netherlands [2,14], the United States [15], and Italy [16]. Two studies focused on samples of students or young individuals [2,15], whereas the remaining 2 administered questionnaires to representative adult populations [14,16]. The sample sizes in all included studies were adequate according to COSMIN standards (N>100), with respondent-to-item ratios ranging from 35 to 127, ensuring robust conditions for the factor analyses performed [2,14-16]. All included studies used a cross-sectional design.

Identified Scales and Instruments

The 4 instruments evaluated different dimensions of the digital divide, as shown in Table 2. Each instrument was developed by the original authors and culturally adapted to the country where it was applied. The tools measured aspects such as digital competencies, digital capital, technological adaptability, and functional digital skills, with particular emphasis on the

XSL•FO
RenderX

psychometric properties of the measurement tools used [2,14-16].

Table . Studies included in the systematic review.

Study	Study type	Measuring instrument	Language	Objective	Population	Cross-cultural adaptation
van Deursen et al [14], 2015	Cross-sectional study	ISS[a]	English	Develop a tested instrument for measuring internet skills	Adults from the UK[b] and the Netherlands (N=1337)	No
van Deursen et al [2], 2022	Cross-sectional study	IoTSS[c]	English	Develop and validate a multidimensional scale of digital skills	Young Dutch people—15-25 y (N=875)	No
Puckett [15], 2022	Cross-sectional study	Digital Adaptability Scale	English	Measure digital adaptive capacity as a dimension of digital inequality	University students in the United States (N=897)	No
Addeo et al [16], 2023	Cross-sectional study	Digital Capital Index	Italian	Validate a composite index of digital capital in the original population	Italian adults (N=1016)	No

[a]ISS: Internet Skills Scale.

[b]UK: United Kingdom.

[c]IoTSS: Internet of Things Skills Scale.

The instruments were based on diverse theoretical frameworks, ranging from models focused on the collection and storage of digital resources [16] to self-assessment scales of performance in technological contexts [15] and European regulatory frameworks on fair competition [2]. All were administered through online surveys and underwent psychometric analysis.

Psychometric Properties Assessed

Table 3 presents the analysis of the psychometric properties of the instruments identified in the 4 studies included in this review. Regarding factor analysis, Puckett [15] applied a 1-stage EFA, whereas van Deursen et al [14], Addeo et al [16], and van Deursen et al [2] conducted both EFA and CFA to verify the internal structure of their scales.

Table . Psychometric properties of the measurement instruments included in the review.

Study	Measuring instrument	Items, n	Reliability, Cronbach α	Factor analysis	Reported validity
van Deursen et al [14], 2015	ISS[a]	35	0.83-0.94	EFA[b]+CFA[c]	Content and construct validity
van Deursen et al [2], 2022	IoTSS[d]	25	0.89-0.92	EFA+CFA	Construct validity: convergent and discriminant
Puckett [15], 2022	Digital Adaptability Scale	12	0.77	EFA	Construct validity: convergent and discriminant
Addeo et al [16], 2023	Digital Capital Index	8	0.701-0.946	EFA	Content and construct validity

[a]ISS: Internet Skills Scale.

[b]EFA: exploratory factor analysis.

[c]CFA: confirmatory factor analysis.

[d]IoTSS: Internet of Things Skills Scale.

All studies reported Cronbach α values exceeding 0.70, indicating at least adequate internal consistency. This suggests that the items in the questionnaires consistently measured the same construct. However, it should be noted that none of the studies reported data on temporal stability (test-retest reliability) or interrater reliability, which limits the assessment of the instruments' consistency over time [2,14-16].

Validity analysis demonstrated that the instruments accurately measured the intended concepts. van Deursen et al [14] and Addeo et al [16] assessed both construct and content validity, whereas Puckett [15] and van Deursen et al [2] examined convergent validity (similarity to previous tests) and discriminant validity (ensuring that the test does not measure unrelated concepts).

XSL•FO
RenderX

In reviewing the instruments used to assess digital skills and thereby measure the digital divide, 4 scales with robust psychometric evidence were identified in the included studies. The Digital Adaptability Scale [15] and the Digital Capital Index [16] were developed as new instruments, whereas the Internet of Things Skills Scale (IoTSS) [2] sought to validate its structure based on the preexisting Internet Skills Scale (ISS) [14] and additional items related to the Internet of Things.

The internal consistency reliability of the instruments was assessed using the Cronbach α coefficient. The interpretative thresholds were as follows: excellent (Cronbach α≥0.90), good (Cronbach α=0.80 - 0.89), acceptable (Cronbach α=0.70 - 0.79), moderate (Cronbach α=0.60 - 0.69), poor (Cronbach α=0.50 - 0.59), and unacceptable (Cronbach α<0.50) [17]. These thresholds align with established psychometric standards for the evaluation of internal consistency in health measurement instruments.

The factor structure of the instruments was analyzed using 2 primary approaches. First, EFA was used in studies where no a priori theoretical structure was assumed. This method is typically used during the development of new instruments or when the dimensionality of the construct is unknown. Regarding CFA, it was used in studies with a predefined hypothesis about the number of factors and the item-factor relationships based on existing theory or prior empirical evidence. Some studies applied a combined approach, initially exploring the underlying structure and subsequently confirming it in a separate analysis or sample [9].

The ISS [14], consisting of 35 items, demonstrated high internal consistency (Cronbach α=0.83 - 0.94) and underwent both EFA and CFA, providing evidence of content and construct validity. Subsequently, the IoTSS [2], comprising 25 items with similarly elevated reliability coefficients (Cronbach α=0.89 - 0.92), was validated through EFA and CFA, with construct validity specifically supported by convergent and discriminant evidence. Similarly, the Digital Adaptability Scale [15], consisting of 12 items and yielding a Cronbach α of 0.77, was validated via EFA, demonstrating construct validity through its convergent and discriminant relationships with other variables. Finally, the Digital Capital Index [16], which includes 8 items, exhibited a wide range in internal consistency (Cronbach α=0.701 - 0.946) as a result of a 2-stage EFA and provided evidence of both content and construct validity. Collectively, these instruments exhibit adequate levels of reliability and validity, making them valuable tools for measuring digital competencies in various contexts.

Methodological Quality Assessment

The risk of bias of the included studies was assessed using the COSMIN Risk of Bias tool, which evaluates the methodological rigor in the design, development, and application of measurement instruments (Table 4). Overall, the studies appropriately defined the construct to be measured and justified its theoretical relevance. However, recurring weaknesses were observed in the preliminary phases, particularly regarding the design of pilot-testing and the evaluation of item comprehensibility from the perspective of the target population.

Table . Assessment of the level of bias in tool design according to the COSMIN (Consensus-Based Standards for the Selection of Health Measurement Instruments) Risk of Bias tool.

Study	PROM[a] (design)	PROM relevance and comprehensiveness	Pilot test design	Comprehensibility of the pilot test	Comprehensiveness of the pilot test	Final assessment (the lowest qualitative score obtained)
van Deursen et al [14], 2015	Very good	Very good	Very good	Very good	Very good	Very good
van Deursen et al [2], 2022	Very good	Very good	Very good	Very good	Very good	Very good
Puckett [15], 2022	Adequate	Adequate	Doubtful	Doubtful	Doubtful	Doubtful
Addeo et al [16], 2023	Adequate	Adequate	Doubtful	Doubtful	Doubtful	Doubtful

[a]PROM: patient-reported outcome measure.

The studies by van Deursen et al [14] and van Deursen et al [2] received a “very good” final rating, having systematically met the COSMIN criteria, including a well-designed and adequately reported pilot phase. In contrast, the studies by Puckett [15] and Addeo et al [16] showed limitations in the implementation and reporting of comprehensibility testing and cognitive interviews, leading to reduced methodological transparency and a final rating of “adequate.”

The quality of the psychometric properties of the instruments was assessed according to the criteria proposed by Terwee et al [13], classifying each property as sufficient (+), insufficient (–), or indeterminate (?; Table 5). Regarding reliability, all 4 studies reported adequate levels, with Cronbach α coefficients exceeding 0.70, indicating sufficient internal consistency (+). Structural validity was also rated positively in all cases as EFA and/or CFA were applied, thus meeting the established standards (+).

XSL•FO
RenderX

Table . Evaluation of psychometric properties.

Study	Reliability	Structural validity	Content validity	Convergent validity	Discriminant validity
van Deursen et al [14], 2015	+[a]	+	+	?[b]	?
van Deursen et al [2], 2022	+	+	+	+	+
Puckett [15], 2022	+	+	+	+	+
Addeo et al [16], 2023	+	+	+	?	?

[a]Sufficient.

[b]Indeterminate.

In terms of content validity, all 4 instruments were grounded in solid theoretical frameworks and developed based on prior literature or expert review, leading to a sufficient rating (+). With respect to convergent and discriminant validity, only the studies by van Deursen et al [2] and Puckett [15] provided clear empirical evidence supporting these aspects of construct validity and, therefore, were rated as sufficient (+). In contrast, the studies by van Deursen et al [14] and Addeo et al [16] did not report enough information to determine the presence of these forms of validity and, thus, were rated as indeterminate (?).

Finally, none of the reviewed instruments used advanced measurement models such as Rasch analysis or IRT, which represents a limitation in terms of more rigorous psychometric evaluation.

Summary of the Most Relevant Results

The 4 studies contributed recent, culturally contextualized instruments for measuring various dimensions of the digital divide. Notably, they reported adequate reliability levels and validated their structures through factor analysis. However, aspects such as convergent and discriminant validity, as well as the use of advanced measurement models, were less frequently addressed.

From a methodological perspective, all studies defined their theoretical content, but limitations were noted in the studies by Puckett [15] and Addeo et al [16] regarding the implementation of pilot-testing and the detailed evaluation of item comprehensibility, which may affect instrument validity in certain contexts [2,16].

Despite these limitations, the validated instruments—particularly the ISS by van Deursen et al [14] and the IoTSS by van Deursen et al [2]—represent valuable tools for analyzing digital inequality in social and health care settings. These tools may be further strengthened in future research through broader validation efforts.

Discussion

Principal Findings

This study is framed within the growing interest in assessing the digital divide arising from unequal access to, use of, and benefit from health technologies, particularly following the rapid digital transformation of recent years accelerated by the COVID-19 pandemic. The digital divide has been widely recognized as a social determinant that conditions equitable access to health care in increasingly digitized environments. This inequality has driven the development of measurement instruments for various contexts, along with the assessment of their corresponding psychometric properties [2,18].

Although the digital divide has 3 levels, as described in the Introduction section, the instruments identified in this review predominantly map onto the second level (digital skills and use) and third level (digital capital and tangible outcomes). Specifically, the scales by van Deursen et al [14] and Puckett [15] operationalize the transition from technical competencies (second level) to social and economic benefits (third level). This focus is crucial for health contexts as it moves beyond mere physical access (first level) to measure effective engagement with digital health resources. The review findings reveal measurement instruments for the digital divide that demonstrate adequate reliability levels and validated structures through factor analysis. However, methodological limitations persist in some studies, such as those by Puckett [15] and Addeo et al [16], including the absence of pilot-testing, inadequate evaluation of item comprehensibility, and a lack of complementary analyses—such as Rasch modeling—that could strengthen scale validity and robustness [2,16]. These gaps complicate cross-study comparisons and limit applicability across diverse contexts.

A critical issue identified is the limited cross-cultural adaptation of the analyzed instruments. None of the included studies [2,14-16] carried out formal translation or linguistic validation processes for use beyond their original context, which restricts their cultural applicability and relevance to Spanish-speaking populations. Nevertheless, the ISS [14] and IoTSS [2] evaluate general competencies, facilitating potential cross-cultural adaptation.

The IoTSS by van Deursen et al [2] stands out as one of the instruments with the strongest psychometric properties and high applicability among young populations, with potential utility in Spanish-speaking educational and health care settings. Its use could promote international result comparisons and inform the design of more context-specific interventions [2].

Adapting measurement tools for vulnerable populations represents a pressing need in public health and clinical research. Sociodemographic factors such as advanced age, migrant status, or rural residence influence access to health care services, health literacy, and psychosocial well-being [19,20]. For instance, research conducted in China by Tang et al [19] demonstrated

XSL•FO
RenderX

that the digital divide negatively impacts quality of life and mental health among older adults in rural areas, primarily due to limited information seeking skills. This finding underscores the importance of culturally adapting measurement instruments by incorporating variables that reflect the social, technological, and educational conditions of diverse population groups.

Despite progress in the design of health-related measurement instruments, there is limited integration of validated scales in cross-sectional studies involving vulnerable populations. Evidence shows that phenomena such as the "grey digital divide" described by Mubarak and Suomi [21] significantly affect older adults, who face technological and social barriers that hinder access to essential services, including health care. The scarcity of cross-sectional studies using validated scales restricts the generation of comparable data and limits the planning of evidence-based interventions [21].

Nursing, due to its central role in health promotion and patient education, should lead cross-cultural validation and adaptation processes for measurement instruments in vulnerable settings. This leadership aligns with nursing practice, characterized by a holistic approach and close ties to community environments. Nurses also possess the necessary competencies to ensure that measurement tools are not only linguistically translated but also adapted to the cultural, social, and educational specificities of target populations, taking into account literacy levels, access to IT, and social support networks [19].

The COVID-19 pandemic critically highlighted structural inequalities, particularly those related to the digital divide and its impact on vulnerable groups. In education, Golden et al [22] documented how the transition to online learning amplified preexisting disparities among racialized and socioeconomically disadvantaged youth, affecting their academic performance and access to mental health resources. Similar inequities are observed in digital health care access, impacting older adults, rural residents, and migrants.

Furthermore, Meka'a et al [23] demonstrated that digital technology use significantly increased the likelihood of youth employment in a Cameroonian population, indicating that information and communications technologies can play a protective role in socioeconomic well-being. However, the lack of internet access or digital devices creates structural gaps that are difficult to overcome [23,24]. Therefore, measurement scale adaptation must consider these sociocultural variables to ensure valid and clinically useful results [22].

The limited integration of measurement instruments into intervention studies or public policies compromises the evaluation of strategies aimed at reducing the digital health divide. Without valid, reliable, and culturally adapted tools, there is a risk of perpetuating—or even exacerbating—inequities in access to digital health technologies, undermining the effectiveness of health care digital transformation [19,21]. Recent research [19,21,25] highlights that insufficient technology access and low digital literacy deepen social and health inequalities. Moreover, without the systematic application of culturally adapted scales, it becomes difficult to identify the actual barriers faced by vulnerable populations and design policies that effectively promote health equity..

Limitations

The predominant use of cross-sectional designs in the included studies limited the assessment of temporal stability and responsiveness. Longitudinal studies are needed to evaluate how digital divide metrics evolve over time. The limited number of fully validated instruments also restricts the availability of evidence regarding the cross-cultural validity of these scales rather than the generalizability of the systematic review findings themselves [8,9]. Additionally, the limited number of studies conducted across diverse demographic or cultural contexts restricts the assessment of broader scale applicability.

Implications for Nursing Practice

Nurses should not only apply validated instruments but also actively engage in their cultural adaptation, ensuring that assessment tools are understandable and relevant for the most vulnerable social groups. For rural populations, older adults, or communities with low digital literacy, it is essential that instruments are accessible and tailored to the cognitive, technological, and sociocultural capacities of participants, thereby addressing the impact of digital exclusion on access to health care and well-being.

In clinical practice, these instruments can serve as screening tools during nursing admission assessments. By identifying patients with low digital capital or skills, nurses can tailor education strategies (opting for nondigital alternatives or providing guided assistance for telehealth use), thereby preventing exclusion from care pathways.

Future Research

From a practical perspective, the analyzed instruments provide a useful foundation for developing new public policies, digital literacy programs, and interventions focused on digital equity. However, their applicability in real-world settings remains challenging due to the need for further strengthening of psychometric properties, particularly sensitivity to change, external validity, and cultural adequacy.

Conclusions

The instruments identified in this review represent a significant advancement in the structured measurement of the digital divide, showing adequate internal reliability and factorial validation. Nonetheless, a critical analysis of their psychometric properties reveals common methodological limitations such as insufficient assessment of convergent and discriminant validity, lack of pilot-testing, and absence of the use of advanced analytical models such as Rasch or IRT.

Moreover, none of the instruments have undergone cross-cultural adaptation for use outside their original linguistic contexts. Despite these limitations, the reviewed tools hold substantial potential for use in health care, particularly if their validation and adaptation processes are strengthened.

Finally, there is a clear need to expand research in this area in terms of both the number of studies and their methodological quality to develop robust instruments capable of accurately and contextually assessing the digital health divide. Future studies should prioritize validation in vulnerable groups, use advanced analytical models, and incorporate cross-cultural adaptation and

XSL•FO
RenderX

uniform scale validation to enable a more precise, inclusive, and comparable evaluation of the digital divide in health.

Funding

The authors declared no financial support was received for this work.

Conflicts of Interest

None declared.

Checklist 1
PRISMA 2020 checklist.
[DOCX File, 36 KB - humanfactors_v13i1e82005_app1.docx]

References

1. Vidal-Alaball J, Alarcon Belmonte I, Panadés Zafra R, Escalé-Besa A, Acezat Oliva J, Saperas Perez C. Approach to digital transformation in healthcare to reduce the digital divide [Article in Spanish]. Aten Primaria 2023 Sep;55(9):102626. [doi: 10.1016/j.aprim.2023.102626] [Medline: 37267831]
2. van Deursen A, van der Zeeuw A, de Boer P, Jansen G, van Rompay T. Development and validation of the Internet of Things Skills Scale (IoTSS). Inf Commun Soc 2022 Oct 3;25(13):1883-1899. [doi: 10.1080/1369118X.2021.1900320]
3. La brecha digital y la administración digital en España: conocimiento clave sobre las competencias digitales de la población para la promoción de la inclusión digital en relación con la administración digital 2023. Fundació Ferrer i Guàrdia. URL: https://www.ferrerguardia.org/web/content/32722?access_token=237f1fe3-d344-40ae-8dd1-8ef4ed7fd2bd&unique=1cf3cac3ec68f3e25467eaefbac4af3445b8bd08 [accessed 2026-04-10]
4. Arroyo Menéndez M, Díaz Velázquez E. Las tecnologías digitales en el ámbito de la salud: brechas sociales, accesibilidad y despersonalización. Teknokult Rev Cult Digit Mov Soc 2021;18(2):95-101. [doi: 10.5209/tekn.75516]
5. Kayser L, Karnoe A, Furstrand D, et al. A multidimensional tool based on the eHealth Literacy Framework: development and initial validity testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36. [doi: 10.2196/jmir.8371] [Medline: 29434011]
6. García-Martín J, García-Sánchez JN. The digital divide of know-how and use of digital technologies in higher education: the case of a college in Latin America in the COVID-19 era. Int J Environ Res Public Health 2022 Mar 12;19(6):3358. [doi: 10.3390/ijerph19063358] [Medline: 35329046]
7. Jongebloed H, Anderson K, Winter N, et al. The digital divide in rural and regional communities: a survey on the use of digital health technology and implications for supporting technology use. BMC Res Notes 2024 Mar 28;17(1):90. [doi: 10.1186/s13104-024-06687-x] [Medline: 38549176]
8. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000 Dec 15;25(24):3186-3191. [doi: 10.1097/00007632-200012150-00014] [Medline: 11124735]
9. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010 May;19(4):539-549. [doi: 10.1007/s11136-010-9606-8] [Medline: 20169472]
10. Mota FP. Attitudes towards internet use: development and proposal of a measurement scale. Rev Gest Tecnol 2020;20(3):216-236. [doi: 10.20397/2177-6652/2020.v20i3.1541]
11. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
12. Prieto G. Uso del Modelo de Rasch de facetas múltiples para analizar test con ítems de respuesta abierta. Rev Iberoam Diagn Eval Aval Psicol 2025 Jun;76(2):13. [doi: 10.21865/RIDEP76.2.02]
13. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]
14. van Deursen A, Helsper EJ, Eynon R. Development and validation of the Internet Skills Scale (ISS). Inf Commun Soc 2015 Aug 25;19(6):804-823. [doi: 10.1080/1369118X.2015.1078834]
15. Puckett C. Digital adaptability: a new measure for digital inequality research. Soc Sci Comput Rev 2022 Jun;40(3):641-662. [doi: 10.1177/0894439320926087]
16. Addeo F, D'Auria V, Delli Paoli A, Punziano G, Ragnedda M, Ruiu ML. Measuring digital capital in Italy. Front Sociol 2023;8:1144657. [doi: 10.3389/fsoc.2023.1144657] [Medline: 37274604]
17. George D, Mallery P. SPSS for Windows step by step: a simple guide and reference, 11.0 update. Internet Archive. 2003. URL: https://archive.org/details/spssforwindowsst00geor [accessed 2026-04-11]

XSL•FO
RenderX

18. Oh SS, Kim KA, Kim M, Oh J, Chu SH, Choi J. Measurement of digital literacy among older adults: systematic review. J Med Internet Res 2021 Feb 3;23(2):e26145. [doi: 10.2196/26145] [Medline: 33533727]
19. Tang Y, Li Q, Wu Y. The impact of the digital divide on rural older people's mental quality of life: a conditional process analysis. Heliyon 2024 Aug;10(17):e37109. [doi: 10.1016/j.heliyon.2024.e37109] [Medline: 39281457]
20. Navas-Echazarreta N, Satústegui-Dordá PJ, Rodríguez-Velasco FJ, et al. Media health literacy in Spanish nursing students: a multicenter cross-sectional study. Nurs Rep 2024 Sep 22;14(3):2565-2579. [doi: 10.3390/nursrep14030189] [Medline: 39330743]
21. Mubarak F, Suomi R. Elderly forgotten? Digital exclusion in the information age and the rising grey digital divide. Inquiry 2022;59:469580221096272. [doi: 10.1177/00469580221096272] [Medline: 35471138]
22. Golden AR, Srisarajivakul EN, Hasselle AJ, Pfund RA, Knox J. What was a gap is now a chasm: remote schooling, the digital divide, and educational inequities resulting from the COVID-19 pandemic. Curr Opin Psychol 2023 Aug;52:101632. [doi: 10.1016/j.copsyc.2023.101632] [Medline: 37437380]
23. Meka'a CB, Nouffeussie AC, Noufelie R, Timba GT. Use of ICTs: what effect on the quality of youth employment in Cameroon? Center for Economic Studies and ifo Institute. 2023. URL: https://www.econstor.eu/bitstream/10419/282555/1/cesifo1_wp10867.pdf [accessed 2026-04-11]
24. Bibi S, Zada H, Awan T, Wong WK, Khan N. Evaluating the moderating role of governance on the relationships between social inclusion, ICT infrastructure, and financial inclusion. Heliyon 2024 Jul 15;10(13):e33711. [doi: 10.1016/j.heliyon.2024.e33711] [Medline: 39040300]
25. Njiepue Nouffeussie AC, Meka'A CB, Noufelie R, Ndouyang Balguessam B. Use of ICT: what effect on youth access to employment in Cameroon? Heliyon 2024 Oct 30;10(21):e39967. [doi: 10.1016/j.heliyon.2024.e39967] [Medline: 39559198]

Abbreviations

CFA: confirmatory factor analysis
COSMIN: Consensus-Based Standards for the Selection of Health Measurement Instruments
EFA: exploratory factor analysis
ICC: intraclass correlation coefficient
IoTSS: Internet of Things Skills Scale
IRT: item response theory
ISS: Internet Skills Scale
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by A Kushniruk, E Borycki; submitted 07.Aug.2025; peer-reviewed by MA Zoubi, M Pankhurst; revised version received 11.Dec.2025; accepted 14.Dec.2025; published 22.Apr.2026.

Please cite as:
Navas-Echazarreta N, Chover-Sierra E, Rubio-Tarazón C, Martínez-Pascual RM, Del Pozo-Herce P, Martínez-Tofé J, Juárez-Vela R, Martínez-Sabater A
Psychometric Properties of Digital Divide Measurement Instruments in Health Contexts: Systematic Review
JMIR Hum Factors 2026;13:e82005
URL: https://humanfactors.jmir.org/2026/1/e82005
doi:10.2196/82005

© Noelia Navas-Echazarreta, Elena Chover-Sierra, César Rubio-Tarazón, Raquel María Martínez-Pascual, Pablo Del Pozo-Herce, Jesús Martínez-Tofé, Raúl Juárez-Vela, Antonio Martínez-Sabater. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 22.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

User-Centered Design for Digital Patient-Navigation Tools in Oncology: Scoping Review

Saba Kheirinejad[1], PhD; Brianna M White[1], DrPH, MPH; Parnian Kheirkhah Rahimabad[1], MD, PhD; Janet A Zink[1], PhD, MS; Soheil Hashtarkhani[1], PhD; Fekede Asefa Kumsa[1], PhD; Rezaur Rashid[1], PhD; Lokesh Chinthala[1], PhD; Christopher L Brett[2], MD; Robert L Davis[1], MD, MPH; David L Schwartz[3,4], MD; Arash Shaban-Nejad[1], PhD, MPH

[1]Center for Biomedical Informatics, Department of Pediatrics, College of Medicine, University of Tennessee Health Science Center, Memphis, TN, United States

[2]Graduate School of Medicine, University of Tennessee at Knoxville, Knoxville, TN, United States

[3]Departments of Radiation Oncology, College of Medicine, University of Tennessee Health Science Center, Memphis, TN, United States

[4]Departments of Preventive Medicine, College of Medicine, University of Tennessee Health Science Center, Memphis, TN, United States

Corresponding Author:
Arash Shaban-Nejad, PhD, MPH
Center for Biomedical Informatics, Department of Pediatrics, College of Medicine
University of Tennessee Health Science Center
50 N Dunlap St
Memphis, TN, 38103
United States
Phone: 1 9012875863
Email: ashabann@uthsc.edu

Abstract

Background: Navigation programs for patients with cancer improve access and continuity of care, yet their digital transformation is often limited by poor usability and inadequate uptake. Applying user-centered and human-centered design (UCD/HCD) principles may close this gap, but the extent to which such design methods are used and evaluated in oncology navigation tools remains unclear.

Objective: This scoping review identifies how UCD/HCD principles have been, and should be, applied in developing and implementing digital health tools for navigation for patients with cancer.

Methods: A scoping review was conducted following PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) and Joanna Briggs Institute guidance. A total of 7 databases (PubMed/MEDLINE, Scopus, IEEE Xplore, Web of Science, Embase, ACM Digital Library, and CINAHL) were searched for English-language articles published between January 2015 and July 2025. Eligible studies reported original, peer-reviewed research on digital or mobile health interventions linked to cancer navigation and documented at least 1 UCD/HCD activity. Two reviewers independently screened records and charted data on context, target users, functions, tool modality, design phase, methods, and outcomes. Findings were synthesized descriptively and thematically.

Results: A total of 36 studies met the inclusion criteria. Findings were organized into 4 domains: study characteristics, navigation functions and digital modalities, design processes and methods, and UCD/HCD application. Iterative prototyping and usability testing were the most common, while participatory design and implementation evaluation were underused.

Conclusions: UCD/HCD approaches enhance usability and patient relevance of digital cancer navigation tools. However, their application remains limited across cancer types, regions, and functions. Broader stakeholder participation and evaluation beyond usability are needed to strengthen coordination, equity, and sustainability in cancer care.

(JMIR Hum Factors 2026;13:e87686) doi:10.2196/87686

KEYWORDS

digital health; HCD; human-centered design; mHealth; mobile health; navigation for patient with cancer; UCD; user-centered design

XSL•FO
RenderX

Introduction

Cancer remains a leading health challenge, with an estimated 35 million new cases to be diagnosed globally by 2050 [1]. Given ever-aging and evolving populations, the intricate nature of cancer care can pose significant challenges to patients and those caring for them [2]. Patient navigation strategies have emerged to eliminate barriers and improve access to care [3]. Community-based patient navigation programs provide personalized support to patients, families, and caregivers across the cancer care continuum [3,4].

Navigation programs improve care engagement, enhance adherence and treatment completion, reduce diagnostic delays, support survivorship care, and improve overall patient quality of life and satisfaction. Moreover, these programs can help patients overcome language barriers, schedule care appointments, and respond to social needs (ie, lack of transportation, childcare, etc). Patient navigation programs also empower patients by helping them better manage their own medical information and promote active participation in their care, including the power to make informed decisions [4,5].

Despite these benefits, patient navigation program design and delivery vary widely and lack standardization. This complicates multidisciplinary care coordination and introduces additional challenges to already complex health care systems, including financial sustainability concerns that may present an unmanageable burden for many care centers [5]. Addressing these issues requires a comprehensive understanding of the sociotechnical aspects of current technologies and workflows. To optimize the reach and effectiveness of patient navigation programs, there is growing interest in leveraging digital health technologies, including mobile health (mHealth) solutions.

Digital health broadly refers to the use of information and communication technologies to support health care delivery, including web-based platforms, mobile apps, telehealth systems, remote monitoring tools, and data-driven decision support technologies [6]. In oncology, digital health tools supporting patient navigation span multiple modalities, such as web portals, mobile apps, hybrid systems, and, increasingly, artificial intelligence–enabled or conversational tools [7]. Defined by the World Health Organization as “medical and public health practice supported by mobile devices,” mHealth encompasses a wide range of tools, including mobile apps, wearables, sensors, and communication platforms designed to improve health outcomes and promote patient-centered care [8]. In oncology settings, mHealth interventions have been used to deliver real-time patient alerts, streamline manual administrative tasks, and improve communication between patients and care teams [9]. Accordingly, this review adopts a broad digital health lens, situating mHealth within a wider ecosystem of technologies used to support navigation for patients with cancer.

Qualitative studies have found that oncology navigators particularly value digital features that reduce administrative workload and facilitate timely intervention (ie, automatic appointment reminder systems, real-time patient messaging, etc). The findings emphasize the importance of care coordination for patients with cancer undergoing chemotherapy and the critical role of technology in improving this process by enhancing communication and access to information for patients [9]. However, many digital tools fail to simultaneously meet the needs of patients and care teams due to poor design, low usability, and a lack of user engagement. Patients often disengage from apps that are not intuitive, while care navigators may find systems incompatible with clinical workflows. To address these issues, there is increasing recognition of the importance of user-centered design (UCD) and human-centered design (HCD).

UCD is a design philosophy that focuses on placing the end user at the center of the design process [10]. In some contexts, the terms HCD and UCD are used interchangeably. However, HCD is often used to emphasize a broader consideration of impacts on various stakeholders, beyond just the direct users. A UCD/HCD approach, which focuses on understanding user needs, engaging stakeholders, and creating tailored solutions [11], can help address these complexities and improve care delivery [12]. These approaches prioritize the needs, preferences, and lived experiences of end users (ie, patients, caregivers, and providers) through iterative, participatory design processes that foster usability, engagement, and real-world applicability [13,14].

Specifically, cancer care is a complex and emotionally demanding process that involves not only patients but also caregivers and clinicians. Patients often face multiple treatments, fluctuating symptoms, and interactions with numerous health care professionals across different systems. Caregivers experience emotional stress and logistical burdens, while clinicians manage heavy workloads and complex decision-making demands. These challenges can lead to physical strain, emotional distress, cognitive fatigue, and variable energy levels, all of which influence how each group engages with digital tools. Therefore, designing tools that are intuitive, accessible, and emotionally supportive is essential for all stakeholders, as this can enhance usability, encourage sustained engagement, and foster coordinated cancer care.

Given the potential for UCD and HCD to improve their effectiveness, there is a need to systematically map the literature on how these design principles have been applied in this context. This scoping review aims to identify and characterize how UCD and HCD approaches have been applied across the development and implementation of digital health and mHealth tools to support navigation for patients with cancer. Rather than assessing effectiveness, the review examines how UCD/HCD has been operationalized across four interrelated dimensions: (1) study characteristics, (2) navigational functions and digital modalities, (3) design processes and methods, and (4) UCD/HCD application. These dimensions were selected because design choices influence not only usability, but also stakeholder inclusion, workflow integration, scalability, and equity. By evaluating current practices against these dimensions, this review seeks to clarify where UCD/HCD efforts have been concentrated, where gaps remain, and how future digital navigation tools might better align with the complex realities of cancer care.

XSL•FO
RenderX

Methods

Overview

This review was guided by the Joanna Briggs Institute methodology [15] for scoping reviews and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines (Multimedia Appendix 1) [16]. Ethical approval was not applicable to this study, as it involved secondary analysis of previously published, peer-reviewed literature and did not include human participants.

Search Strategy

A comprehensive literature search was undertaken to map how UCD and HCD principles have been used when developing or implementing digital and mHealth tools for navigation for patients with cancer. The search strategy was developed iteratively by the study team; no formal librarian consultation was conducted. The review incorporated peer-reviewed studies that were identified from 7 electronic databases (PubMed/MEDLINE, Scopus, IEEE Xplore, Web of Science Core Collection, Embase [Ovid], ACM Digital Library, and CINAHL [EBSCO]) and were published between January 1, 2015, and July 30, 2025. The search was conducted on July 30, 2025, and the strategy combined the following three concept blocks joined by the Boolean operator AND: (1) cancer navigation terminology ("cancer navigation," "patient navigation," "care coordination," "care transitions," "cancer care pathway," "cancer journey support"), (2) digital and mHealth terminology ("digital health," "mobile health," "mHealth," "eHealth," "mobile application," "wearable," "remote monitoring," "telehealth," "teleoncology," "health technology"), and (3) UCD/HCD terminology ("user-centered design," "human-centered design," "UCD," "HCD," "participatory design," "person-centered care" combined with design, "usability testing"). Database-specific controlled vocabulary (eg, Medical Subject Headings and Emtree) was incorporated where available, and search strings were adapted to the syntax of each platform.

Full, database-specific search strategies and dates of execution are provided in Multimedia Appendix 2.

Eligibility Criteria

Studies were included for review if they met all of the following criteria: (1) published between January 1, 2015, and July 30, 2025, (2) described the development or implementation of 1 or more cancer care navigation-related functions, including care coordination, access facilitation, communication between patients and care teams, appointment management, psychosocial support, financial or logistical assistance, or guidance across the cancer care continuum, (3) reported at least 1 UCD or HCD component (ie, needs assessment, co-design workshop, iterative prototyping, and formal usability testing), (4) involved patients with cancer of any age, their caregivers, or navigation personnel, (5) were original, peer-reviewed research articles or design case studies, and (6) were available in the English language. Studies were excluded if they evaluated oncology digital tools unrelated to navigation, lacked any description of UCD/HCD processes, examined nononcology populations, or appeared as conference abstracts, dissertations, protocols, commentaries, or other gray literature.

Study Selection

All search results were imported into Covidence [17] for screening. Duplicates were removed automatically and confirmed manually. The screening process was conducted in two sequential stages: (1) a primary title and abstract screening followed by full-text review of potentially eligible studies and (2) a secondary bidirectional snowballing process, which involved recursively pursuing relevant references cited in the retrieved literature (backward snowballing) as well as identifying studies that cited the included articles (forward snowballing).

Two reviewers (SK and BMW) independently screened titles and abstracts to assess relevance to the eligibility criteria; disagreements were resolved through discussion, with a third reviewer available for arbitration if needed. Studies meeting the inclusion criteria were advanced to the full-text review stage, where the same dual-review procedure was applied.

Definitions and Conceptual Framework

In this review, digital health refers broadly to the use of information and communication technologies to support health care delivery, including web-based platforms, mobile apps, telehealth systems, remote monitoring tools, and data-driven decision support technologies. mHealth represents a subset of digital health technologies supported by mobile devices.

UCD and HCD are design approaches that emphasize the active involvement of end users and stakeholders throughout the design process to ensure that technologies align with user needs, values, and contexts. In this review, UCD/HCD applications were examined across multiple design phases, including discovery, design, prototyping, evaluation, and implementation.

Design methods refer to specific activities or techniques used to engage users or stakeholders, such as interviews, focus groups, co-design workshops, or usability testing. These design methods may be applied at different design phases and are not mutually exclusive.

Participatory methods are defined as design approaches that involve users or stakeholders as active contributors to the design process rather than solely as subjects of evaluation.

Implementation, as used in this review, refers to the deployment and integration of digital navigation tools within real-world care settings, including considerations related to feasibility, workflow integration, training, and sustainability.

Data Extraction

Data were extracted using a pilot charting form in Microsoft Excel. Two reviewers (SK and BMW) independently recorded bibliographic details (author, year, country, and journal), study design and methods, participant characteristics (cancer type, age range, and navigator role), digital tool attributes (platform, functionality, and stage of deployment), UCD/HCD activities undertaken, navigation tasks supported, and reported outcomes (usability measures, patient engagement, navigator workload,

XSL•FO
RenderX

and clinical indicators). Discrepancies were reconciled by consensus.

Data Analysis

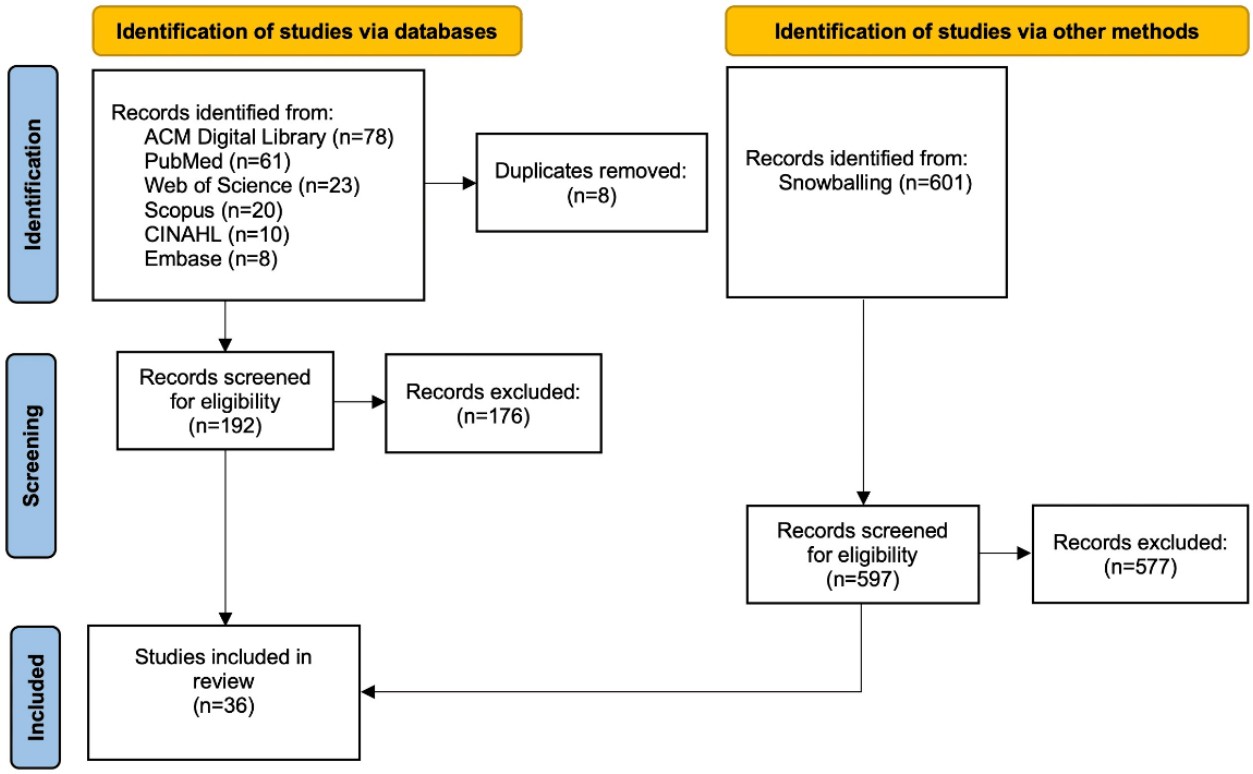
Thematic analysis was conducted through a qualitative synthesis of the studies presented in Table S1 in Multimedia Appendix 2. Each article was reviewed in full to identify key concepts, use cases, and reported outcomes relating to the application of UCD/HCD in digital cancer navigation. An inductive approach was used to identify patterns emerging directly from the data, which were then organized into the following four overarching domains reflecting the clinical and methodological dimensions of UCD/HCD: (1) study characteristics, (2) navigation functions and digital modalities, (3) design processes and methods, and (4) conceptualizations of cancer navigation through UCD/HCD. Domains were refined iteratively until they provided a coherent, nonoverlapping structure for grouping the findings. Alongside the qualitative synthesis, quantitative frequencies (publication year, geography, cancer type, and digital modality) were summarized in descriptive tables and figures to map the landscape of evidence.

Results

Overview

A total of 36 studies met the inclusion criteria for review, as shown in Figure 1. Of the 192 records screened for inclusion through the primary title and abstract review, 176 records were excluded for failure to meet the eligibility criteria. In total, 16 studies met the inclusion criteria and were retained for qualitative synthesis. Bidirectional snowballing of these records resulted in the inclusion of 20 additional articles.

Figure 1. PRISMA-ScR flow diagram of literature search and screening.



Study characteristics are summarized in Table S1 in Multimedia Appendix 2. Below, findings are organized around 4 domains that emerged from qualitative synthesis of the extraction matrix.

Table 1 presents a high-level synthesis of the thematic findings from the included studies, summarizing reported UCD/HCD design processes, functional scope, and digital modalities, and evaluation approaches and impacts in oncology patient navigation tools. The table highlights commonly used methods, dominant functional focuses, and key gaps identified across the literature. Descriptive study characteristics, such as cancer types, geographic distribution, and study settings, are provided separately in the text and Table S1 in Multimedia Appendix 2 to contextualize the evidence base.

XSL•FO
RenderX

Table 1. Summary of key findings on user-centered design (UCD)/human-centered design (HCD) for digital patient navigation in oncology.

Domain	Key themes	Key findings
Design processes (UCD/HCD phases, design methods, and stakeholder engagement)	• Iterative prototyping and usability testing were most common. • Discovery and implementation phases rarely included. • Interviews and surveys were most frequently used, few participatory or co-design workshops. • Engagement was largely patient-only; minimal involvement of clinicians, navigators, or caregivers.	• UCD applied primarily to midcycle for usability evaluation. • Participatory methods underused, limiting innovation and sustainability.
Functional scope of tools (navigation purpose and digital modality)	• Web portals and mobile apps were overwhelmingly used; focused primarily on education and communication. • Limited coverage of financial, logistical, or psychosocial barriers. • Few integrated or multimodal systems.	• Tools improved patient knowledge and engagement. • Narrow functional range; informational focus was most common. • Opportunities to integrate financial, transportation, and emotional-support features remain.
Evaluation and impact (navigation outcomes and author reflections)	• Usability and satisfaction were the most used metrics. • Few studies assessed navigation effectiveness, coordination, or clinical outcomes. • Equity, timeliness, and continuity rarely evaluated. • Some studies conceptually broadened navigation toward empowerment and communication.	• Tools perceived as usable but with limited real-world evidence of improved care continuity. • Shift needed from acceptability to system-level effectiveness. • UCD reframes navigation as multidimensional (communication, coordination, and empowerment).

Study Characteristics

Across the 36 included studies, most focused on all adult cancers (n=16, 44.4%) or specifically on breast cancer (n=8, 22.2%). Fewer addressed pediatric populations (n=3, 8.3%), lung (n=3, 8.3%), or prostate cancer (n=2, 5.6%), while single-study contexts included brain, cervical, chronic myeloid leukemia, colorectal, esophageal, head and neck, melanoma and renal cell carcinoma, and multiple myeloma (n=1, 2.8% each). This distribution reflects a strong emphasis in the current literature on adult and breast cancer contexts, with limited representation of other cancer types and life stages.

Regarding target user groups, most digital tools were designed for patients (33/36, 91.7%) and clinicians or providers (25/36, 69.4%), while caregivers (6/36, 16.7%) and patient navigators (2/36, 5.6%) were considered in just 2 studies. This distribution highlights a strong emphasis on patient-facing tools in the current literature, with comparatively limited attention to caregivers and patient navigators. For example, OncoKompas [13], an app codeveloped with cancer survivors, enables users to monitor their quality of life through patient-reported outcomes and receive automated, personalized feedback and supportive care recommendations, illustrating how tools are often designed directly with patients in mind to provide personalized self-management support.

Geographically, nearly half of the studies were conducted in the United States (17/36, 47.2%), followed by the Netherlands (7/36, 19.4%) and Canada (3/36, 8.3%). Fewer studies were conducted in Africa, Germany, and Sweden (2/36, 5.6% each), and from Australia, Iran, Norway, and Taiwan (1/36, 2.8% each). The predominance of studies from North American and European settings underscores a need for more geographically diverse applications of UCD/HCD approaches to enhance their generalizability and global relevance in cancer navigation.

Navigation Functions and Digital Modalities

Navigation Purpose

According to our findings, navigation functions were most frequently conceptualized as education and information provision (23/36, 63.9%). Closely following were overall care system navigation or access to care and patient-to-provider communication (21/36, 58.3% each), along with care coordination (20/36, 55.6%). Less frequently, studies reported functions related to emotional or psychosocial support (11/36, 30.6%) and appointment scheduling and tracking (7/36, 19.4%). Additionally, functions included patient-to-patient communication (3/36, 8.3%), symptom management (2/36, 5.6%), and financial or insurance navigation, medication support, and transportation or logistics (1/36, 2.8% each).

These results show a predominant focus on education, system navigation, and communication functions, with limited attention to supportive services, financial or logistical navigation, and peer interaction. For example, Mendu et al [18] created and tested a culturally tailored virtual patient educator for Hispanic women to improve cervical cancer and human papillomavirus education using an iterative, UCD. The *CMyLife* platform presented by Ector et al [14] was codeveloped to help patients interpret their own test results and take action across their care journey, reframing navigation as patient empowerment rather than simple wayfinding. Similarly, digital personal health network systems supported chemotherapy patients by enhancing communication with providers and improving care coordination, while usability studies identified needed refinements in data entry and connectivity [9,19]. Moreover, interventions such as *iNNOVBC* presented by Mendes-Santos et al [20] highlighted the underexplored potential of navigation tools to provide psychological support, positioning digital navigation to address care gaps in survivorship.

XSL•FO
RenderX

Type of Digital Tools

Web platforms were the most common type of proposed digital tools, reported in 18 (50%) studies, closely followed by mobile apps (17/36, 47.2%). A smaller proportion presented only a conceptual design without testing a specific tool (3/36, 8.3%), while software prototypes and telehealth or video tools were each reported in a single study (1/36, 2.8%).

This distribution suggests that research has largely concentrated on familiar digital formats (ie, web platforms and mobile apps), with comparatively limited integration of newer modalities such as conversational agents or artificial intelligence–driven chatbots, which were categorized as digital health modalities within this review. For example, the *MijnAVL* [21] platform was iteratively developed with survivor input as a patient-facing web portal to enhance usability and provide accessible educational content. Similarly, the *StressProffen* [22] platform illustrates the use of mobile apps to deliver psychosocial support. Broader approaches also emerged, such as a human-centered mobile and web application (*mPCL*) [23] in Tanzania, which extended functionality beyond patient-facing use to support coordinated palliative care among patients, caregivers, and local health workers.

UCD/HCD Phase and Methods Used

Across studies included, UCD/HCD approaches were reported at multiple phases of the design cycle, including discovery, design and prototyping, testing, and implementation. Rather than following a linear or standardized sequence, most studies reported applying UCD/HCD selectively at 1 or more phases, with considerable variability in timing and depth.

In this review, discovery refers to early-stage activities aimed at understanding user needs, care contexts, and system requirements through methods such as interviews, focus groups, and needs assessments. Design and prototyping encompass the iterative creation of low- to high-fidelity solutions, including wireframes, mock-ups, and functional prototypes, often developed through co-design with stakeholders. Testing includes usability evaluations and pilot studies conducted to assess feasibility, acceptability, and workflow fit in controlled or limited real-world settings. Implementation indicates the deployment and integration of a digital navigation tool into routine clinical workflows or operational care settings, where real users engage with the system as part of care delivery. Distinguishing these phases allowed us to examine the maturity of UCD/HCD application across studies and identify gaps in extending user-centered practices into real-world deployment.

Discovery-phase activities, such as needs assessments, contextual inquiry, and early exploration of user workflows, were reported in fewer than half of the studies (17/36, 47.2%). These approaches were typically used to identify barriers to care coordination, information gaps, or workflow challenges faced by patients, caregivers, or clinicians. As examples in early discovery, Carr et al [12] examined barriers in the workflows of patient navigators supporting patients with breast cancer and explored how an mHealth app could help improve their care delivery and coordination, Ankrah et al [24] interviewed childhood cancer survivors and caregivers to explore survivorship experiences and offered sociotechnical design insights to better support young survivors, and Lau and Loiselle [25] examined oncology nurses' views of an evidence-based eHealth tool, highlighting that incorporating their perspectives makes eHealth tools valuable for patient care and nursing practice.

Design and prototyping phases were more frequently represented (19/36, 52.7%), often involving the development of functional prototypes and working systems. Studies such as *CMyLife* [14] and *iNNOVBC* [20] illustrate how patient-centered platforms were developed and iteratively refined through user feedback, leading to improvements in navigation, personalization, and care coordination features. These examples highlight the use of UCD/HCD primarily to refine near-functional systems and support real-world implementation.

Evaluation and testing activities were the most consistently reported component of UCD/HCD application. Testing was used in 63.8% (23/36) of studies, reflecting a strong emphasis on evaluating and refining tools once developed. As an example in usability testing, the *ASyMS* platform developed by Moradian et al [26], a remote monitoring system, underwent think-aloud usability testing with patients to refine symptom reporting functions before moving toward broader clinical use.

Implementation was the least represented phase (7/36, 19.4%), indicating that relatively few studies extended user-centered approaches into deployment, integration, or evaluation of digital navigation tools within routine clinical workflows. Studies such as Mittal et al [27] and Nabelsi and Plouffe [28] demonstrated that user feedback during implementation revealed system-level challenges, including training needs and resource constraints. These findings underscore the importance of extending UCD/HCD beyond development and evaluation into real-world deployment contexts.

Overall, UCD/HCD approaches were mostly used in prototyping and evaluation phases, with less frequent attention to early discovery and downstream implementation in cancer navigation. Their potential for shaping problem definition and ensuring sustainable implementation remains less frequently explored in available literature.

A variety of UCD/HCD design methods were reported across the included studies, with 80% (29/36) of the studies reporting the use of more than 1 method across the design process. Interviews were the most frequently used approach (25/36, 69.4%), followed by usability testing (16/36, 44.4%). End-user feedback sessions and surveys were reported in 22.2% (8/36) of studies, while focus groups appeared in 13.9% (5/36). Less commonly reported design methods included content analysis, scenarios, stakeholder consultations, and thematic analysis (2/36, 5.6% each). Participatory design sessions (2/36, 5.6%) and co-design workshops (1/36, 2.8%) were comparatively rare. Other approaches, such as direct observation, field trials, iterative design, journey mapping, pilot-testing, and system usage data, each appeared only once (1/36, 2.8% each).

Several cases illustrate this methodological skew toward consultative rather than collaborative techniques. *OncoKompa* [13] and *MijnAVL* [21] primarily drew on survivor interviews

XSL•FO
RenderX

and usability testing to refine prototypes, while *ASyMS* [26] relied on structured usability evaluations, including think-aloud protocols, to improve device navigation. These approaches ensured tools were aligned with patient needs but often stopped short of deeper participatory design. By contrast, studies such as *StressProffen* [22] integrated stakeholder perspectives more systematically, combining evidence-based principles with design thinking methods to produce a culturally sensitive psychosocial intervention. Likewise, the *mPCL* [23] project in Tanzania demonstrated the value of participatory methods by actively involving patients, caregivers, and local health workers in shaping the platform, ensuring contextual relevance and cultural appropriateness.

In summary, researcher-led design methods that capture user perspectives were predominant across the included studies, while collaborative and participatory approaches were less common. Only a small number of studies reported using co-design workshops, journey mapping, or multistakeholder sessions. This indicates that participatory techniques aimed at involving users and stakeholders throughout the design process remain limited in current cancer navigation research.

Conceptualizations of Cancer Navigation Through UCD/HCD

Overview

Across included studies, UCD/HCD approaches were most often applied as practical design strategies rather than overarching frameworks for navigation. Based on our findings from currently available literature, their use has been concentrated to three main functions: (1) evaluating usability and satisfaction outcomes as markers of navigation success, (2) centering patient perspectives as the primary driver of design decisions, and (3) using iterative cycles of prototyping and testing to refine tools. For example, O'Malley et al [29] developed and tested the *e-EXCELS* eHealth tool for cancer survivorship follow-up care.

Usability testing revealed concerns about patient burden, tool fatigue, timing of introduction, relevance, and privacy, underscoring the importance of using a UCD process.

Less frequently, studies broadened participation to include caregivers, clinicians, or navigators, or leveraged design methods to reconceptualize navigation itself as a process of information exchange, communication, and coordination.

In summary, UCD/HCD approaches in cancer navigation were mainly used to ensure that patient-facing tools are usable and acceptable. Fewer studies applied these approaches to address system-level outcomes, involve multiple stakeholders, or support implementation in real-world settings.

Outcomes as Navigation Constructs

Across included studies, outcomes most often reflected usability (28/36, 77.8%) and user satisfaction (26/36, 72.2%), with these measures commonly serving as proxies for the overall success of cancer care navigational tools. Navigation effectiveness was reported less frequently (17/36, 47.2%), while patient engagement appeared in just over one-third of studies (13/36, 36.1%).

Broader outcomes such as clinical use (9/36, 25%), design or strategy insights (8/36, 22.2%), and health professional engagement (6/36, 16.7%) were less assessed. For example, Ho et al [23] evaluated a digital case management platform for community-based palliative care in sub-Saharan Africa by gathering clinicians' perspectives on its impact on care delivery, showing that it improved access, enabled remote symptom management, and supported more efficient patient prioritization. Even more rarely, studies addressed feasibility and implementation logistics (5/36, 13.9%), collaboration and coordination (3/36, 8.3%), or educational value (2/36, 5.6%). For example, Nikkhah et al [30] applied role theory to understand how parents act as caregiving teams in children's hospitals, emphasizing the need for design features that support coordination and shared responsibilities. As another example, Womack et al [31] engaged patients in conceptualizing clinic processes and digital health tools aimed at raising awareness and supporting education on integrative oncology services.

These findings reveal a strong emphasis on usability and user satisfaction in cancer navigation tools, illustrating how UCD/HCD has been applied primarily to ensure tools are acceptable and easy to use, rather than to measure their long-term effectiveness or system-level impact. For instance, *StressProffen* [22] and *Enable* [32] demonstrated high usability and satisfaction among survivors, underscoring the success of iterative, user-centered refinements. Similarly, *ASyMS* [26] usability testing revealed strong patient motivation to engage with symptom monitoring, even as participants called for a more attractive design and expanded features. Ahmadi et al [33] developed a mobile app for self-care management in Iranian women with breast cancer, and usability testing confirmed that its content and functions were appropriate and satisfactory. Moving beyond usability, tools such as *OncoKompas* [13] and *mPCL* [23] illustrate a shift toward evaluating navigation effectiveness and feasibility in real-world contexts, whether through tailoring survivorship resources or coordinating palliative care across diverse settings.

Overall, the reviewed studies primarily measured outcomes related to usability and user satisfaction, with fewer examining broader aspects such as navigation effectiveness, coordination, or sustainability. Only a limited number of studies assessed system-level or implementation-related outcomes, such as timeliness, collaboration, or equitable access. Toward this direction, Geerts et al [34] developed and evaluated *MM E-Coach*, a patient-centered eHealth application designed to support patients with multiple myeloma and health care professionals by improving treatment management, usability, and communication through an iterative, user- and stakeholder-driven design process. They further initiated a randomized clinical trial to evaluate its clinical effectiveness.

Similarly, Morse et al [35] designed, developed, and tested mPCL, a human-centered mobile and web application that supports outpatient palliative symptom assessment, care coordination, and pain control among Tanzanian patients with cancer by connecting patients, caregivers, local health workers, and palliative care specialists. Their findings emphasized the need for further evaluation to assess the tool's effectiveness and sustainability, particularly in remote and resource-limited

XSL•FO
RenderX

settings. Mittal et al [27] also designed and deployed *SCOPE*, a platform for technology-enhanced collaborative care combining a patient-facing mobile app with a provider-facing registry. They deployed *SCOPE* for 24 months for actual care in 6 cancer clinics.

Levels of Stakeholder Engagement in UCD/HCD

Across included studies, stakeholder engagement in UCD/HCD processes varied in scope and depth. Most studies engaged patients as the primary voice in the design process, often through semistructured interviews and surveys. This strong emphasis on patient perspectives reflects the value of empathy and coproduction, which are central principles of UCD/HCD as shown by Hou et al [36] and Womack et al [31]. Schadewaldt et al [37] also suggested that applying evidence and stakeholder input through a UCD/HCD approach can inform implementation strategies that reduce the risk of digital health failure.

In line with these findings, Glaser et al [38] evaluated the *CAPABLE* prototype for immunotherapy patients, showing that involving target users improved perceived usefulness and revealed critical symptom–reporting issues that would have gone unnoticed without their participation. Tools such as *OncoKompas* [13] and *MijnAVL* [21] exemplify this approach, with cancer survivors directly involved in iterative prototype testing to refine content, navigation features, and overall functionality. *CMyLife* presented by Ector et al [14], developed for individuals with chronic myeloid leukemia, equipped patients with tools to track and interpret their results while aligning design decisions with survivor priorities identified during participatory sessions. Similarly, *SmartSurvivor* by Baseman et al [39] explored mHealth options for survivorship care, explicitly embedding patient perspectives into the functionality and usability of the tool.

Although most included studies centered on participation of patients with cancer in UCD/HCD processes, a smaller but important subset extended participation to include health care providers, clinicians, caregivers, and navigators [40,41]. These examples demonstrate how human-centered approaches can extend beyond individual patients' needs to reflect the complex, multiactor systems that shape cancer care. For example, Salako et al [42] brought together researchers, clinicians, and technology developers to review remote symptom monitoring for palliative cancer care in Africa. They described 3 approaches from different countries, highlighting delivery models, outcomes, and implementation experiences. The study concluded that codeveloping and adapting these tools with local stakeholders is feasible, with user-centered insights guiding future scale-up.

Nabelsi and Plouffe [28] engaged nurse navigators and interfacility teams to refine a web-based coordination platform designed to improve oncology and nononcologic thoracic surgery care pathways. Similarly, the *mPCL* project by Ho et al [23] incorporated local health workers in Tanzania to co-design a mobile platform for palliative care, ensuring cultural and contextual relevance in a low-resource setting. Aronoff-Spencer et al [43] also described a human-centered participatory design approach that engaged patients, caregivers, and community stakeholders to support rural patients experiencing distress during cancer treatment. Patel et al [44] interviewed 20 caregivers of children with cancer to explore potential design solutions. They identified 4 caregiving coordination technology archetypes at diagnosis and showed how caregivers' preferences evolved over time.

Across the reviewed studies, inclusion of diverse stakeholders such as clinicians, caregivers, and navigators was less common compared to patient-only approaches. Studies that incorporated multiple stakeholder groups identified additional factors affecting navigation design, including workflow alignment, caregiver responsibilities, and infrastructure limitations. Overall, most studies maintained a primary focus on patient perspectives, with fewer addressing the broader system-level context of care delivery, such as organizational workflows, care coordination processes, technological infrastructure, and institutional constraints that shape how navigation services are implemented and sustained.

Iteration and Responsiveness

Across included studies, UCD/HCD was frequently operationalized as an iterative cycle in which prototypes were tested and refined in response to user feedback. This responsiveness ensured that navigation tools were grounded in real-world needs and experiences.

For example, tools such as *StressProffen* [22] and *CMyLife* [14] demonstrated how usability testing can identify challenges with data entry and workflow guidance, which directly informed revisions to interface design and content presentation to improve navigation pathways. Likewise, iterative development of *MijnAVL* presented by Kuijpers et al [21] led to usability improvements, including the addition of FAQs and hyperlinks to improve patient experience. Similarly, an iterative development process of a cancer prevention web application by Carrera et al [45] and the prevention-focused web application described by Anders et al [32] incorporated repeated cycles of feedback that shaped tutorials and interactive features, supporting its potential to enhance risk understanding and healthy behavior.

Redefining Navigation Through Design

Beyond creating tools for discrete functions, several studies applied UCD/HCD approaches to reconceptualize what "navigation" means in cancer care. Rather than limiting navigation to scheduling or wayfinding, studies framed it as a broader process of information exchange, communication, and care coordination.

For example, Basch et al [46] leveraged patient-reported outcomes to enhance communication between patients and clinicians, transforming symptom monitoring into a navigation mechanism for timely intervention and treatment adjustment. Similarly, Donawa et al [47] co-designed a mobile tool for distress management in rural communities, positioning navigation as an integrated support system that addressed emotional, informational, and logistical barriers simultaneously. *Interaktor*, presented by Langius-Eklöf et al [48], also broadened navigation by increasing patients' sense of security and self-reflection, thereby embedding psychosocial support directly into routine symptom monitoring.

XSL•FO
RenderX

Discussion

Principal Findings

This scoping review synthesizes how UCD and HCD principles have been applied in the development and implementation of digital cancer navigation, while also identifying gaps that limit broader impact. Overall, UCD/HCD was most often applied to patient-facing tools during prototyping and usability evaluation, with comparatively less attention to early discovery phases and downstream implementation in routine care settings. These efforts appear to have improved alignment between navigational tools and patient needs; however, fewer studies addressed system-level barriers, equity, cancer-specific navigational needs, and global applicability. Future research and development should consider deliberately extending UCD/HCD across a wider variety of cancers, diverse geographic contexts, multistakeholder design processes, and outcomes that measure more than usability. To support interpretive clarity, we distinguish between findings derived directly from the reviewed studies and forward-looking implications intended to inform future research, design practice, and implementation.

Findings demonstrate that UCD/HCD in cancer navigation has been applied most extensively to adult cancers in high-income settings. Pediatric and rare cancers were comparatively underrepresented, as were studies conducted in low- and middle-income settings, despite the substantial coordination demands often present in these contexts. Addressing these gaps may be important for improving the generalizability and equity relevance of digital navigation research. As an illustrative scenario, in a pediatric oncology ward in sub-Saharan Africa, co-design workshops with caregivers could reveal the daily burdens of medication access and transportation [23]. Embedding those insights into a navigation app might prioritize transportation support (eg, ride-share scheduling or community volunteer coordination) alongside education, addressing barriers that may influence treatment continuity. Without attention to these barriers, digital navigation may risk reproducing existing inequities rather than mitigating them.

At the stakeholder level, most tools were developed directly with patient input, demonstrating the strength of UCD/HCD in capturing and centering lived experience. However, caregivers, navigators, and clinicians were rarely included, which may leave workflow and coordination barriers underexplored. Caregivers and patient navigators can play pivotal roles in access, treatment timeliness, and continuity of cancer care, while clinician engagement at each iterative stage of UCD/HCD is critical for integration into workflows and sustained adoption [2,30]. Limiting participatory design primarily to patients could risk overlooking system-level and workforce perspectives needed for real-world implementation. To address this gap, design processes could prioritize multistakeholder participation. For instance (as an illustrative example), in developing a survivorship platform for rural patients, engaging nurse navigators could identify how symptom alerts fit within clinical workflows, while caregiver voices could flag gaps in home-based support. This approach may shift tools from stand-alone patient tools toward integrated systems designed to bolster continuity of care.

Current tools focus heavily on education and information provision, with limited attention to psychosocial, financial, or logistical barriers. These functions may improve self-efficacy and patient activation [49], but they may be insufficient for addressing the psychosocial, financial, or logistical barriers that many patients face [50]. While UCD/HCD approaches have supported the development of education and communication functions responsive to patient needs, future applications could also extend to the broader spectrum of navigation services that shape equity, access, and continuity of care. This pattern reflects how UCD/HCD has been applied primarily to informational and communicative aspects of navigation, while more complex or resource-intensive functions remain underexplored.

Strengths and Limitations

This review offers one of the first comprehensive maps of how UCD/HCD principles have been applied to digital and mHealth tools for navigation for patients with cancer. A comprehensive search strategy was applied across 7 major databases, combining controlled vocabulary and free-text terms to capture diverse design approaches and navigation functions. However, several limitations should be noted. Despite the breadth of databases searched, the review was limited to English-language publications between 2015 and 2025, potentially excluding relevant studies in other languages or earlier foundational work. While the dual-review process enhanced rigor, inclusion criteria required explicit mention of UCD or HCD, which may have excluded studies using participatory approaches under different terminology. Although descriptive and thematic analysis revealed important patterns, no formal quality appraisal was conducted, in keeping with scoping review methodology. Given this, the strength of evidence underlying individual findings could not be assessed. Moreover, most studies were conducted in high-income countries and focused on adult cancer populations, limiting generalizability to low- and middle-income settings or underrepresented cancer types.

Future Directions

Overall, the review's comprehensive scope provides a foundation for future research. The implications outlined below are forward-looking and intended to inform future research and practice; several examples are illustrative and were not directly evaluated in the reviewed studies. Subsequent systematic reviews should incorporate quality appraisal frameworks, cross-cultural analyses, and longitudinal evaluations to further examine the effectiveness, equity impact, and sustainability of UCD/HCD in cancer navigation.

Moving forward, developers should also consider expanding navigation functions to directly address psychosocial, financial, and logistical barriers identified in the literature. For example, a survivorship tool could embed a financial navigation module where patients track insurance claims and receive cost-of-living support resources, similar to the Wheeler et al [51] model. Likewise, a peritreatment app might integrate transportation (ie, ride-share) scheduling to reduce missed appointments in gynecologic oncology. Designing and implementing these

XSL•FO
RenderX

functions may require moving beyond informational features toward more complex domains with potential system-level impact.

Most reviewed studies relied on traditional UCD/HCD design methods (ie, interviews, surveys, and usability testing applied midcycle) with limited reporting of upstream discovery or downstream implementation. Discovery and routine care implementation were rarely discussed, which may limit both innovation and sustainability. This pattern reflects broader digital health literature, where many interventions are tested in laboratory- or pilot-type settings, very few report robust real-world deployment, maintenance, or adoption over time [52]. Overreliance on consultative design methods risks capturing only surface-level usability preferences while overlooking systemic challenges such as workflow misalignments, caregiver burdens, or sociotechnical barriers [53]. These observations and gaps carry important public health policy implications, as insufficient integration of UCD/HCD across the design-to-implementation continuum can undermine investments in digital oncology infrastructure, limit scalability, and hinder equitable access to effective navigation technologies. Health systems and policymakers should prioritize frameworks that embed UCD/HCD into digital health standards and implementation incentives.

Additionally, a shift toward embedding participatory and iterative design methods from start to finish is needed. Journey mapping at the discovery stage could uncover bottlenecks in care transitions (ie, from surgery to rehabilitation), guiding more effective tool design. Co-design workshops and multistakeholder field trials during implementation could then test how tools integrate into routine care, adapt across cancer contexts, and endure beyond pilots into real-world applications [54-57].

Correspondingly, outcomes were dominated by usability and satisfaction, while far fewer studies assessed navigation effectiveness, care coordination, or clinical use. Moreover, even fewer included measures of equity, timeliness, or continuity of care. This narrow scope risks producing tools that are usable but not transformative for access, coordination, or sustainability. Evaluation frameworks must expand to capture system-level outcomes. Pairing implementation outcomes with theory-informed, causally grounded evaluation approaches can help assess not only acceptability, but whether proposed mechanisms (eg, improved coordination leading to fewer delays) are supported by observed data rather than usability proxies alone [58].

One example could include comprehensive trials to assess whether navigation apps reduce treatment delays across racial groups, improve communication between community hospitals and tertiary centers, or decrease financial burden for patients and families. Embedding these outcomes would help ensure UCD/HCD advances are judged by their tangible contribution to health equity and system performance, and not just whether users like the interface.

Finally, UCD/HCD approaches may contribute to broadening the conceptualization of navigation from wayfinding toward communication, coordination, and empowerment. Tools such as *StressProffen* [22] and *CMyLife* [14] illustrate how iterative design can expand navigation beyond tracking to include psychosocial support and relational care. For practice, this shift underscores the importance of embedding patient perspectives alongside those of clinicians, navigators, and caregivers, ensuring tools align with real-world workflows. Scaling this orientation across cancer contexts and diverse health systems could transform navigation from a patient app to a core function of oncology care delivery. For research, future studies should move beyond single-site, short-term evaluations toward longitudinal and multicountry designs that capture not only usability but also system-level outcomes. By expanding participatory approaches and testing tools in diverse contexts, UCD/HCD has the potential to transform navigation into a durable, equity-sensitive component of cancer care delivery.

Conclusions

UCD and HCD approaches have demonstrated value in shaping digital cancer navigation tools that are usable, acceptable, and responsive to patient needs; however, their current application remains narrow, concentrated in adult cancers, high-income settings, and informational functions. While these contributions are notable, relatively few studies address the broader coordination, equity, and implementation challenges that define navigation in real-world oncology settings.

The evidence from this review suggests that UCD/HCD has been most impactful when applied iteratively, embedding patient perspectives to refine tools that improve engagement, communication, and self-management. At the same time, caregiver, navigator, and clinician voices remain underrepresented, participatory methods are underused, and outcomes largely privilege usability over system-level effectiveness. These patterns may limit both the scalability and sustainability of current tools. To more fully realize the full potential of UCD/HCD in cancer navigation, future research and practice should extend beyond interface refinement toward the development of equitable, context-sensitive systems. This includes deliberate inclusion of a wide variety of cancers and low- and middle-income settings, broader stakeholder engagement across patients, caregivers, and providers, and systematic use of participatory and co-design methods that span discovery through implementation.

Equally, evaluation frameworks should extend beyond usability to capture coordination, timeliness, continuity, and equity as essential markers of navigation success. These findings underscore that UCD/HCD should both refine cancer navigation tools and redefine what navigation means within oncology. An integrated process of empowerment, communication, and coordination across diverse care contexts is warranted. By embracing inclusivity, methodological innovation, and system-level evaluation, the next generation of digital navigational tools can move beyond promising prototypes to sustainable solutions that improve outcomes and reduce disparities in cancer care delivery.

XSL•FO
RenderX

Funding

This study was supported by a grant from the State of Tennessee Department of Health.

Data Availability

Data sharing is not applicable to this article as no new data were created or analyzed in this study. All data extracted and analyzed in this review are available in Multimedia Appendix 2, which provides bibliographic and methodological details of the included studies.

Authors' Contributions

Conceptualization: SK, BMW
Drafting of manuscript: SK, BMW, AS-N
Review and editing of manuscript: SK, BMW, PKR, JAZ, SH, FAK, RR, CLB, RLD, DLS, AS-N
Supervision: DLS, AS-N
Funding acquisition: DLS, AS-N

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR checklist.
[DOCX File , 84 KB - humanfactors_v13i1e87686_app1.docx]

Multimedia Appendix 2
Summary of Included Studies and Full Search Strategies.
[DOCX File , 582 KB - humanfactors_v13i1e87686_app2.docx]

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74(3):229-263 [FREE Full text] [doi: 10.3322/caac.21834] [Medline: 38572751]
2. Chen M, Wu VS, Falk D, Cheatham C, Cullen J, Hoehn R. Patient navigation in cancer treatment: a systematic review. Curr Oncol Rep 2024;26(5):504-537 [FREE Full text] [doi: 10.1007/s11912-024-01514-9] [Medline: 38581470]
3. Freeman HP, Rodriguez RL. History and principles of patient navigation. Cancer 2011;117(15 Suppl):3539-3542 [FREE Full text] [doi: 10.1002/cncr.26262] [Medline: 21780088]
4. Sleiman M, Yockel M, Fleischmann A, Silber E, Liu M, Young O, et al. The role of peer support and patient navigation for empowerment in breast cancer survivors: implications for community cancer control. J Psychosoc Oncol Res Pract 2024;6(2):128. [doi: 10.1097/or9.0000000000000128] [Medline: 39463701]
5. Neal CD, Weaver DT, Raphel TJ, Lietz AP, Flores EJ, Percac-Lima S, et al. Patient navigation to improve cancer screening in underserved populations: reported experiences, opportunities, and challenges. J Am Coll Radiol 2018;15(11):1565-1572. [doi: 10.1016/j.jacr.2018.03.001] [Medline: 29685346]
6. Maita KC, Maniaci MJ, Haider CR, Avila FR, Torres-Guzman RA, Borna S, et al. The impact of digital health solutions on bridging the health care gap in rural areas: a scoping review. Perm J 2024;28(3):130-143 [FREE Full text] [doi: 10.7812/TPP/23.134] [Medline: 39135461]
7. Monjaras-Avila C, Blakaj D, Hu K, Yamoah K, Gamez E, Comas A, et al. Integrating digital health into oncology: a comprehensive review. Preprints.org. Preprint posted online on January 7, 2026 2026. [doi: 10.20944/preprints202601.0502.v1]
8. Kheirinejad S, Visuri A, Suryanarayana SA, Hosio S. Exploring mHealth applications for self-management of chronic low back pain: a survey of features and benefits. Heliyon 2023;9(6):e16586 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e16586] [Medline: 37346357]
9. Ngo V, Matsumoto CG, Joseph JG, Bell JF, Bold RJ, Davis A, et al. The personal health network mobile app for chemotherapy care coordination: qualitative evaluation of a randomized clinical trial. JMIR Mhealth Uhealth 2020;8(5):e16527 [FREE Full text] [doi: 10.2196/16527] [Medline: 32452814]
10. Anderson NS, Norman DA, Draper SW. User Centered System Design: New Perspectives on Human-Computer Interaction. Hillsdale, NJ: Lawrence Erlbaum Associates; 1986.

XSL•FO
RenderX

11. Melles M, Albayrak A, Goossens R. Innovating health care: key characteristics of human-centered design. Int J Qual Health Care 2021;33(Supplement_1):37-44 [FREE Full text] [doi: 10.1093/intqhc/mzaa127] [Medline: 33068104]
12. Carr AL, Vinod N, Farha P, Lu T, Farha MJ, Graves KD. An exploration of facilitators and barriers to patient navigator core functions with breast cancer patients: implications for the development of a human-centered mHealth app. PEC Innov 2023;3:100226 [FREE Full text] [doi: 10.1016/j.pecinn.2023.100226] [Medline: 37859864]
13. Duman-Lubberding S, van Uden-Kraan CF, Jansen F, Witte BI, van der Velden LA, Lacko M, et al. Feasibility of an eHealth application "OncoKompas" to improve personalized survivorship cancer care. Support Care Cancer 2016;24(5):2163-2171 [FREE Full text] [doi: 10.1007/s00520-015-3004-2] [Medline: 26563178]
14. Ector GI, Westerweel PE, Hermens RP, Braspenning KA, Heeren BC, Vinck OM, et al. The development of a web-based, patient-centered intervention for patients with chronic myeloid leukemia (CMyLife): design thinking development approach. J Med Internet Res 2020;22(5):e15895 [FREE Full text] [doi: 10.2196/15895] [Medline: 32412424]
15. Santos WMD, Secoli SR, Püschel V. The Joanna Briggs Institute approach for systematic reviews. Rev Lat Am Enfermagem 2018;26:e3074 [FREE Full text] [doi: 10.1590/1518-8345.2885.3074] [Medline: 30462787]
16. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
17. Covidence systematic review software. Veritas Health Innovation. URL: https://www.covidence.org/ [accessed 2026-04-08]
18. Mendu S, Boukhechba M, Gordon J, Datta D, Molina E, Arroyo G, et al. Design of a culturally-informed virtual human for educating Hispanic women about cervical cancer. Int Conf Pervasive Comput Technol Healthc 2018;2018:360-366 [FREE Full text] [doi: 10.1145/3240925.3240968] [Medline: 30555731]
19. Kim K, Bell J, Bold R. A personal health network for chemotherapy care coordinationvaluation of usability among patients. Stud Health Technol Inform 2016;225:232-236. [doi: 10.3233/978-1-61499-658-3-232]
20. Mendes-Santos C, Nunes F, Weiderpass E, Santana R, Andersson G. Development and evaluation of the usefulness, usability, and feasibility of iNNOV breast cancer: mixed methods study. JMIR Cancer 2022;8(1):e33550 [FREE Full text] [doi: 10.2196/33550] [Medline: 35166682]
21. Kuijpers W, Groen WG, Oldenburg HS, Wouters MW, Aaronson NK, van Harten WH. Development of MijnAVL, an interactive portal to empower breast and lung cancer survivors: an iterative, multi-stakeholder approach. JMIR Res Protoc 2015;4(1):e14 [FREE Full text] [doi: 10.2196/resprot.3796] [Medline: 25614924]
22. Børøsund E, Mirkovic J, Clark MM, Ehlers SL, Andrykowski MA, Bergland A, et al. A stress management app intervention for cancer survivors: design, development, and usability testing. JMIR Form Res 2018;2(2):e19 [FREE Full text] [doi: 10.2196/formative.9954] [Medline: 30684438]
23. Ho YX, Morse RS, Lambden K, Mushi BP, Ngoma M, Mahuna H, et al. How a digital case management platform affects community-based palliative care of Sub-Saharan African cancer patients: clinician-users' perspectives. Appl Clin Inform 2022;13(5):1092-1099 [FREE Full text] [doi: 10.1055/s-0042-1758223] [Medline: 36384234]
24. Ankrah E, Bhattacharya A, Donjuan L. When worlds collide: boundary management of adolescent and young adult childhood cancer survivors and caregivers. 2022 Presented at: CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; April 29; New Orleans, LA, USA p. 1-16. [doi: 10.1145/3491102.3517544]
25. Lau GJ, Loiselle CG. E-health tools in oncology nursing: perceptions of nurses and contributions to patient care and advanced practice. Can Oncol Nurs J 2018;28(2):118-124 [FREE Full text] [doi: 10.5737/23688076282118124] [Medline: 31148816]
26. Moradian S, Krzyzanowska MK, Maguire R, Morita PP, Kukreti V, Avery J, et al. Usability evaluation of a mobile phone-based system for remote monitoring and management of chemotherapy-related side effects in cancer patients: mixed-methods study. JMIR Cancer 2018;4(2):e10932 [FREE Full text] [doi: 10.2196/10932] [Medline: 30578238]
27. Mittal A, Jones T, Karkar R, Suh J, Williams S, Zheng Y, et al. SCOPE: examining technology-enhanced collaborative care management of depression in the cancer setting. Proc ACM Hum Comput Interact 2025;9(2):CSCW162. [doi: 10.1145/3711060] [Medline: 40519326]
28. Nabelsi V, Plouffe V. Enhancing care coordination in oncology and nononcology thoracic surgery care pathways through a digital health solution: mixed methods study. JMIR Form Res 2024;8:e60222 [FREE Full text] [doi: 10.2196/60222] [Medline: 39591606]
29. O'Malley DM, Davis SN, Devine KA, Sullivan B, Bator A, Clemow L, et al. Development and usability testing of the e-EXCELS tool to guide cancer survivorship follow-up care. Psychooncology 2020;29(1):123-131 [FREE Full text] [doi: 10.1002/pon.5222] [Medline: 31626397]
30. Nikkhah S, Rode AU, Mittal P. "I feel like I need to split myself in half": using role theory to design for parents as caregiving teams in the children's hospital. 2022 Presented at: CSCW'22 Companion: Companion Computer Supported Cooperative Work and Social Computing; November 8-22; Virtual Event, Taiwan p. 115-120. [doi: 10.1145/3500868.3559466]
31. Womack DM, Kennedy R, Chamberlin SR, Rademacher AL, Sliney CD. Patients' lived experiences and recommendations for enhanced awareness and use of integrative oncology services in cancer care. Patient Educ Couns 2022;105(7):2557-2561 [FREE Full text] [doi: 10.1016/j.pec.2021.11.018] [Medline: 34865887]

XSL•FO
RenderX

32. Anders C, Moorthy P, Svensson L, Müller J, Heinze O, Knaup P, et al. Usability and user experience of an mHealth app for therapy support of patients with breast cancer: mixed methods study using eye tracking. JMIR Hum Factors 2024;11:e50926 [FREE Full text] [doi: 10.2196/50926] [Medline: 38441959]
33. Ahmadi M, Shahrokhi SN, Khavaninzadeh M, Alipour J. Development of a mobile-based self-care application for patients with breast cancer-related lymphedema in Iran. Appl Clin Inform 2022;13(5):935-948 [FREE Full text] [doi: 10.1055/s-0042-1757295] [Medline: 36198310]
34. Geerts P, Eijsink J, Moser A, Ter Horst P, Boersma C, Postma M. Rationale and development of an e-health application to deliver patient-centered care during treatment for recently diagnosed multiple myeloma patients: pilot study of the MM E-coach. Pilot Feasibility Stud 2023;9(1):85 [FREE Full text] [doi: 10.1186/s40814-023-01307-0] [Medline: 37210584]
35. Morse RS, Lambden K, Quinn E, Ngoma T, Mushi B, Ho YX, et al. A mobile app to improve symptom control and information exchange among specialists and local health workers treating Tanzanian cancer patients: human-centered design approach. JMIR Cancer 2021;7(1):e24062 [FREE Full text] [doi: 10.2196/24062] [Medline: 33755022]
36. Hou IC, Lan MF, Shen SH, Tsai PY, Chang KJ, Tai HC, et al. The development of a mobile health app for breast cancer self-management support in Taiwan: design thinking approach. JMIR Mhealth Uhealth 2020;8(4):e15780 [FREE Full text] [doi: 10.2196/15780] [Medline: 32352390]
37. Schadewaldt V, O'Brien T, Kalla M, Krishnasamy M, Burns K, Bray SC, et al. Development of an evidence-informed implementation strategy for a digital supportive care platform for brain tumour patients, their carers and healthcare professionals. Digit Health 2025;11:20552076251316713 [FREE Full text] [doi: 10.1177/20552076251316713] [Medline: 39949847]
38. Glaser SLC, Fraterman I, van Brummelen N, Tibollo V, Del Campo LM, Mallo H, et al. Usability and usefulness of a symptom management coaching system for patients with cancer treated with immune checkpoint inhibitors: comparative mixed methods study. JMIR Form Res 2025;9:e57659 [FREE Full text] [doi: 10.2196/57659] [Medline: 39847771]
39. Baseman J, Revere D, Baldwin LM. A mobile breast cancer survivorship care app: pilot study. JMIR Cancer 2017;3(2):e14 [FREE Full text] [doi: 10.2196/cancer.8192] [Medline: 28951383]
40. Timmerman JG, Tönis TM, Dekker-van Weering MGH, Stuiver MM, Wouters MWJM, van Harten WH, et al. Co-creation of an ICT-supported cancer rehabilitation application for resected lung cancer survivors: design and evaluation. BMC Health Serv Res 2016;16:155 [FREE Full text] [doi: 10.1186/s12913-016-1385-7] [Medline: 27121869]
41. Iott B, Caverly T, Fishstrom A, King D, Meng G, Flynn A. Clinician perspectives on the user experience, configuration, and scope of use of a patient reported outcomes (PRO) dashboard. 2019 Presented at: PervasiveHealth'19: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 20-23; Trento Italy p. 21-30. [doi: 10.1145/3329189.3329198]
42. Salako O, Enyi A, Miesfeldt S, Kabukye JK, Ngoma M, Namisango E, et al. Remote symptom monitoring to enhance the delivery of palliative cancer care in low-resource settings: emerging approaches from Africa. Int J Environ Res Public Health 2023;20(24):7190 [FREE Full text] [doi: 10.3390/ijerph20247190] [Medline: 38131741]
43. Aronoff-Spencer E, McComsey MY, Chih M, Hubenko A, Baker C, Kim J, et al. Designing a framework for remote cancer care through community co-design: participatory development study. J Med Internet Res 2022;24(4):e29492 [FREE Full text] [doi: 10.2196/29492] [Medline: 35412457]
44. Patel S, Nikkhah S, Nadkarni S. Designing for resilient caregiving coordination journeys: prioritizing features across families and phases. 2024 Presented at: CSCW Companion '24: Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing; 2024 November 9-13; San Jose Costa Rica p. 525-532. [doi: 10.1145/3678884.3681901]
45. Carrera P, Elias O, Zhang R. User-centered design and usability evaluation of a cancer prevention web application: an iterative approach in Germany. medRxiv. Preprint posted online on April 30, 2025 2025. [doi: 10.1101/2025.04.29.25326679]
46. Basch E, Stover AM, Schrag D, Chung A, Jansen J, Henson S, et al. Clinical utility and user perceptions of a digital system for electronic patient-reported symptom monitoring during routine cancer care: findings from the PRO-TECT trial. JCO Clin Cancer Inform 2020;4:947-957 [FREE Full text] [doi: 10.1200/CCI.20.00081] [Medline: 33112661]
47. Donawa A, Powell C, Wang R, Chih MY, Patel R, Zinner R, et al. Designing survey-based mobile interfaces for rural patients with cancer using Apple's ResearchKit and CareKit: usability study. JMIR Form Res 2024;8:e57801 [FREE Full text] [doi: 10.2196/57801] [Medline: 39326043]
48. Langius-Eklöf A, Christiansen M, Lindström V, Blomberg K, Hälleberg Nyman M, Wengström Y, et al. Adherence to report and patient perception of an interactive app for managing symptoms during radiotherapy for prostate cancer: descriptive study of logged and interview data. JMIR Cancer 2017;3(2):e18 [FREE Full text] [doi: 10.2196/cancer.7599] [Medline: 29089290]
49. Tuominen L, Leino-Kilpi H, Poraharju J, Cabutto D, Carrion C, Lehtiö L, et al. Interactive digital tools to support empowerment of people with cancer: a systematic literature review. Support Care Cancer 2024;32(6):396 [FREE Full text] [doi: 10.1007/s00520-024-08545-9] [Medline: 38816629]
50. Dau H, Kassam F, Payne BA, Miller H, Ogilvie G. Digital technology as a tool to provide social support to individuals with cancer in low- and middle-income countries: a scoping review. PLOS Digit Health 2025;4(4):e0000609. [doi: 10.1371/journal.pdig.0000609] [Medline: 40257979]

XSL•FO
RenderX

51. Wheeler SB, Manning ML, Gellin M, Padilla N, Spees LP, Biddell CB, et al. Impact of a comprehensive financial navigation intervention to reduce cancer-related financial toxicity. J Natl Compr Canc Netw 2024;22(8):557-562. [doi: 10.6004/jnccn.2024.7030] [Medline: 39413831]
52. Duffy A, Christie GJ, Moreno S. The challenges toward real-world implementation of digital health design approaches: narrative review. JMIR Hum Factors 2022;9(3):e35693 [FREE Full text] [doi: 10.2196/35693] [Medline: 36083628]
53. Busolo D, Woodgate R. Palliative care experiences of adult cancer patients from ethnocultural groups: a qualitative systematic review protocol. JBI Database System Rev Implement Rep 2015;13(1):99-111. [doi: 10.11124/jbisrir-2015-1809] [Medline: 26447011]
54. Voorheis P, Wong JV, Lazarevic N, Imtiaz B, Bhuiya A, Steele Gray C. Using journey mapping and service blueprinting to design digital health behavior change innovations: a scoping review. J Diabetes Sci Technol 2025;19(5):1196-1214. [doi: 10.1177/19322968251334396] [Medline: 40338057]
55. Messina A, Annoni AM, Amati R, Bano B, Franscella G, Albanese E, et al. Participatory methods in designing digital health interventions for informal caregivers of people with dementia. A systematic review. Internet Interv 2025;39:100799 [FREE Full text] [doi: 10.1016/j.invent.2024.100799] [Medline: 40161473]
56. Henni SH, Maurud S, Fuglerud KS, Moen A. The experiences, needs and barriers of people with impairments related to usability and accessibility of digital health solutions, levels of involvement in the design process and strategies for participatory and universal design: a scoping review. BMC Public Health 2022;22(1):35 [FREE Full text] [doi: 10.1186/s12889-021-12393-1] [Medline: 34991540]
57. Jallow F, Orbegozo AM, Caruhapoma MDC. Design workshops: participatory co-design and iterative adaptation of cervical cancer screening implementation plans in Peru. Cancer Epidemiol Biomarkers Prev 2023;32((6_Suppl)):28. [doi: 10.1158/1538-7755.asgcr23-abstract-28]
58. Brenas JH, Shaban-Nejad A. Health intervention evaluation using semantic explainability and causal reasoning. IEEE Access 2020;8:9942-9952. [doi: 10.1109/access.2020.2964802]

Abbreviations

HCD: human-centered design
mHealth: mobile health
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
UCD: user-centered design

Edited by A Kushniruk; submitted 12.Nov.2025; peer-reviewed by C MuÑoz Olivar, L Cong; comments to author 19.Jan.2026; revised version received 20.Feb.2026; accepted 31.Mar.2026; published 29.Apr.2026.

Please cite as:
Kheirinejad S, White BM, Kheirkhah Rahimabad P, Zink JA, Hashtarkhani S, Kumsa FA, Rashid R, Chinthala L, Brett CL, Davis RL, Schwartz DL, Shaban-Nejad A
User-Centered Design for Digital Patient-Navigation Tools in Oncology: Scoping Review
JMIR Hum Factors 2026;13:e87686
URL: https://humanfactors.jmir.org/2026/1/e87686
doi:10.2196/87686
PMID:

©Saba Kheirinejad, Brianna M White, Parnian Kheirkhah Rahimabad, Janet A Zink, Soheil Hashtarkhani, Fekede Asefa Kumsa, Rezaur Rashid, Lokesh Chinthala, Christopher L Brett, Robert L Davis, David L Schwartz, Arash Shaban-Nejad. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 29.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Process Evaluation of Digital Mental Health Interventions for Psychosis: Scoping Review With Framework Synthesis

Chloe Hampshire[1], MSc; Charlotte Dack[1], PhD; Shadi Daryan[1], MSc; Carolina Fialho[2], MSc; Rayan Taher[2], PhD; Ashley-Louise Teale[1], DClinPsy; Jenny Yiend[2], PhD; Pamela Jacobsen[1], DClinPsy, PhD

[1]Department of Psychology, University of Bath, Bath, England, United Kingdom

[2]Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, England, United Kingdom

Corresponding Author:
Pamela Jacobsen, DClinPsy, PhD
Department of Psychology
University of Bath
Claverton Down
Bath, England, BA2 7AY
United Kingdom
Phone: 44 (0) 1225 385091
Email: pcj25@bath.ac.uk

Abstract

Background: Implementing digital mental health interventions (DMHI) for those with psychosis is a persistent challenge. A process evaluation, or studies conducted alongside trials, is one research method that may address this issue. However, a synthesis of process evaluation data in this area is missing.

Objective: This study aimed to understand what is known about context, implementation, and mechanisms of impact by synthesizing process evaluation data from trials evaluating DMHIs used by people with psychosis.

Methods: A scoping review using a 2-phase search strategy underpinned by the Medical Research Council (MRC) process evaluation framework was conducted. Database searches of Cochrane Central Register of Controlled Trials and PsycInfo in 2024 and 2025 first identified an index sample of peer-reviewed trials predominantly conducted in the United Kingdom (≥50% of samples from the United Kingdom in multicountry studies). Next, papers linked to the index sample were retrieved and included if they reported process evaluation data as operationalized in the MRC framework. Two authors independently screened references, extracted summary data, and assessed the quality of index trials. One author qualitatively synthesized process evaluation data using a deductive framework synthesis approach using the MRC framework. Findings were triangulated with senior authors and presented as a narrative synthesis.

Results: Searches identified 14 DMHIs and 45 papers reporting process evaluation data, though only 2 were labeled as such. Qualitative syntheses of process evaluation data generated five themes aligned with the MRC framework: (1) enhancing fit and supporting delivery (implementation strategies); (2) DMHI implementation varied across users, staff, and delivery settings (implementation outcomes); (3) helping users to respond in more helpful ways (mechanisms); (4) addressing perceived and actual implementation factors (context); and (5) limited impact of user characteristics on DMHI outcomes (context).

Conclusions: There is preliminary evidence that DMHIs can be delivered to people experiencing psychosis within trial settings, although use varied between individuals. Future implementation efforts may benefit from addressing contextual factors influencing DMHI use, including users' treatment needs and preferences, everyday demands, and staff availability for blended interventions. Future research could evaluate implementation strategies, validate how and for whom DMHIs work, and embed process evaluation in trials.

Trial Registration: PROSPERO CRD42024439117; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024439117

(JMIR Hum Factors 2026;13:e88032) doi:10.2196/88032

KEYWORDS

digital mental health interventions; framework synthesis; process evaluation; psychosis; randomized controlled trial; scoping review

XSL•FO
RenderX

Introduction

Psychosis is characterized by strongly held unusual beliefs and often distressing sensory experiences, such as hearing voices [1], and can cause difficulties in many aspects of life, including interpersonal relationships, reduced quality of life, and increased risk of early death due to poorer health outcomes [2]. In the United Kingdom, clinical practice guidelines recommend psychological interventions, such as cognitive behavioral therapy or family interventions, for individuals with psychosis [1]. However, access to these therapies remains below recommended levels [3,4].

Given these persistent access challenges, attention has increasingly turned to digital mental health interventions (DMHIs), or technology-based treatment designed to improve users' mental health [5], as a potential way to expand psychological support. Notably, a 2024 early value assessment by the National Institute for Health and Care Excellence in the United Kingdom highlighted DMHIs could be a promising treatment mode for those with psychosis [6]. Although evidence suggests DMHIs can reduce mental health symptoms experienced by those with psychosis [7,8], implementing these interventions in practice remains a persistent challenge [9-11].

To help bridge this DMHI evidence-practice gap, an international commission from Lancet Psychiatry in 2024 recommended adopting methods that embed implementation work throughout the research process [12]. One such research method is a process evaluation, or studies that are often embedded in randomized controlled trials (RCTs) that explore how, for whom, and where interventions work [13]. The most recent process evaluation guidance was published by the Medical Research Council (MRC) in the United Kingdom in 2015 [13]. Process evaluation in the MRC framework explores 3 outcomes (Textbox 1) [13].

Textbox 1. Medical Research Council process evaluation components [13].

- Context: external factors that influence how an intervention is implemented or functions.
- Implementation: the quantity and quality of, and processes required for, intervention delivery.
- Mechanisms of impact: how an intervention produces an effect.

Some studies have used process evaluation to evaluate DMHIs used by those with psychosis [14-17]. Existing evidence highlights that digital treatments may work by empowering users or facilitating self-reflection [14-17], and has identified some barriers to implementation, including user-level (eg, competing life demands) [17] and service-level (eg, staff turnover) [16] factors. However, no review has synthesized process evaluation findings for DMHIs used by those with psychosis. Further, process evaluation is often underused [18] or inconsistently labeled as such [19,20], meaning relevant data may be difficult to identify. Consequently, the breadth, nature, and gaps in the available evidence base are currently unclear.

A scoping review is therefore warranted to understand what is currently known in this area and to identify key evidence gaps to inform future research, evaluation, and implementation efforts. In line with methodological recommendations [21,22], our review focuses on research conducted in the United Kingdom, as differences in health care systems and psychosis care pathways can limit comparisons between countries. Our research question was, "What is currently known about the context, implementation, and mechanisms of impact from RCTs predominantly conducted in the United Kingdom used by people with psychosis?"

Methods

Design

This review followed the Joanna Briggs Institute methodology for scoping reviews [23] and is reported in accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [24] and Enhancing Transparency in the Reporting of Syntheses of Qualitative evidence [25] checklists (Tables S1 and S2 in Multimedia Appendix 1). The protocol was preregistered and can be accessed at PROSPERO (International Prospective Register of Systematic Reviews; CRD42024439117). No amendments were made to the protocol.

Search Strategy and Eligibility Criteria

A 2-phase search strategy, supported by Covidence software, modeled on a prior review of process evaluation [19], was used to first identify relevant RCTs, then associated papers (Figure 1 adapted from French et al [19]). This approach helped address inconsistent use of the term "process evaluation" [19,20] and is recommended for retrieving such studies [20].

Phase 1 identified DMHIs used by people with psychosis that were evaluated in an RCT (index RCTs) using preplanned database searches of PsycInfo and Cochrane Central Register of Controlled Trials on January 31, 2024, and October 22, 2025 (Table S1 in Multimedia Appendix 2). Studies included in phase 1 met the following requirements:

- Population: participants aged >14 years with psychosis established by either (1) a diagnosis of a psychotic disorder using a validated tool, or (2) a score on a standardized psychosis measure (indicating psychotic symptoms are the primary presenting difficulty), or (3) receipt of care from a specialist psychosis service.
- Intervention: digitally delivered, targeting a mental health difficulty.
- Studies: the main outcome paper (the first peer-reviewed paper publishing the prespecified primary outcome data) of peer-reviewed predominantly RCTs based in the United Kingdom (≥50% of samples in the United Kingdom in multicountry studies with comparable health care systems).

Phase 2 systematically identified peer-reviewed publications associated with index RCTs that reported process evaluation

XSL•FO
RenderX

data. Following Cochrane guidance [26], process evaluations were identified using the 2015 MRC framework [13], irrespective of how the studies were labeled. Definitions and coding rules are provided in Table 1. Studies in phase 2 met the following requirements: reporting at least 1 process evaluation outcome from the 2015 MRC framework (context, implementation, or mechanisms of impact) [13] using qualitative, quantitative, or mixed methods, and being clearly linked to the main RCT (via trial registration or program grant).

In phases 1 and 2, masked reviewers (CH, SD, CF, and ALT) independently screened search results in Covidence using eligibility criteria on titles and abstracts, then the full text. Agreement between reviewers was calculated using Cohen κ for both screening stages in Covidence, and discrepancies in inclusion were resolved through discussion, with senior team members (PJ and CD) consulted where appropriate.

Figure 1. Method overview. MRC: Medical Research Council; RCT: randomized controlled trial.

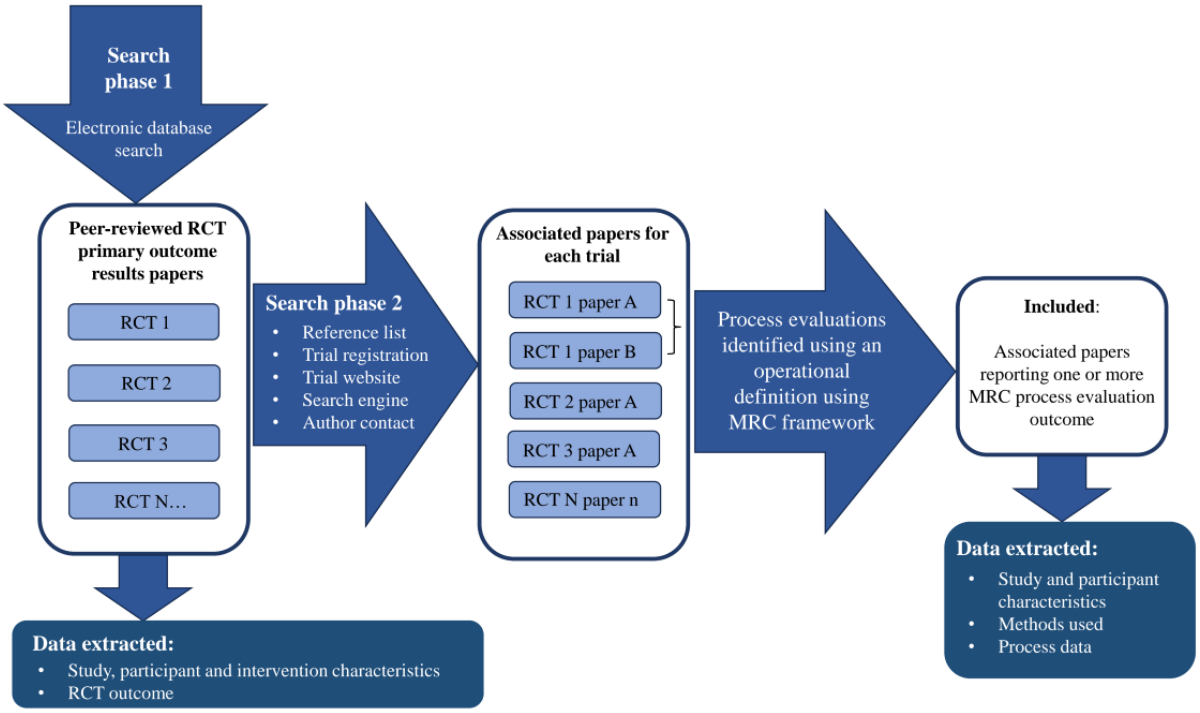


Table 1. Coding framework operationalizing process evaluation components [13].

Component and definition	Coding rules
Context	
• External factors that influence intervention implementation or functioning in a trial.	• Code data about factors that are external to the intervention and its components. • For intervention implementation, code data that influences how the intervention delivery or components are carried out. • For intervention functioning, code data that influences the purpose or desired effect of interventions.
Implementation	
• How much and/or how well an intervention is used and/or delivered in a trial (implementation outcomes) [27].	• For implementation outcomes, code data about any of the following: who participates in, uses, or delivers the intervention, how much the intervention is implemented, and the degree to which the intervention is implemented as intended.
• How delivery is achieved in a trial (implementation strategies) [28].	• For implementation strategies, code data about the actions taken by researchers to enable or support delivery in a trial.
Mechanisms of impact	
• How interventions bring about change in a trial.	• Code data about the intermediate structures between the intervention and the outcomes that produce an effect. Data can include intended and unintended mechanisms.

XSL•FO
RenderX

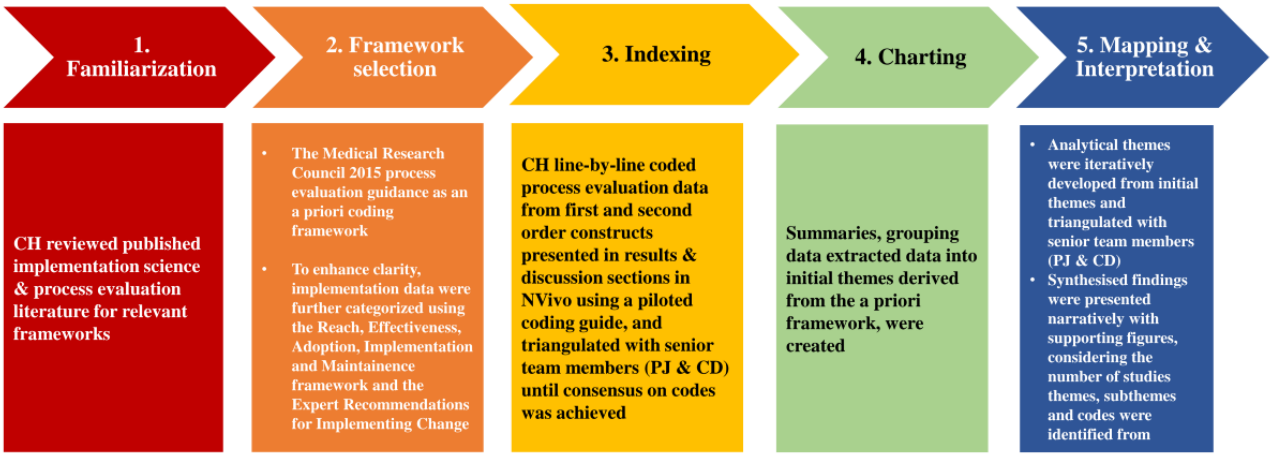
Data Extraction and Critical Appraisal

Relevant data were electronically extracted in Covidence to reduce bias. Masked reviewers (CH, SD, and ALT) extracted summary characteristics from index RCTs (study, participant, and intervention characteristics and RCT outcome [positive/negative]) and process evaluation studies (study and participant characteristics and methods used). CH extracted process evaluation data using a predefined coding guide (Table 1). Missing data were not sought from the authors. The methodological quality of associated publications was not assessed because a validated tool for process evaluations has not been developed [20], and inclusion of all studies in qualitative syntheses is recommended [29]. Index RCTs were independently assessed by masked reviewers (CH and RT) using the Cochrane risk-of-bias 2 tool, evaluating five domains: (1) randomization procedures, (2) deviations from intended interventions, (3) missing outcome data, (4) measurement of the outcome, and (5) selection of the reported result [30]. As scoping reviews do not typically exclude studies based on methodological quality, these assessments were conducted to provide contextual information on the included evidence rather than to inform study inclusion.

Data Synthesis

We used a constructivist approach [31] with a critical realist perspective [32]. A convergent data-based synthesis design, supported by NVivo (Lumivero), was used to qualitatively synthesize process evaluation data deductively [33]. Framework synthesis, using the 2015 MRC process evaluation framework [13], guided analysis in five stages: (1) familiarization, (2) framework selection, (3) indexing, (4) charting, (5) mapping and interpretation (Figure 2) [34]. Process evaluation data were indexed using the coding framework outlined in Table 1. To enhance the interpretability of findings, data coded within the implementation domain in the MRC framework were categorized into implementation outcomes using relevant domains from the Reach, Effectiveness, Adoption, Implementation, and Maintenance framework, which includes 5 outcomes relevant to evaluating implementation [27]. Data were also coded as implementation strategies (actions taken to support delivery) from the Expert Recommendations for Implementing Change (ERIC), which includes 73 defined implementation strategies that have been ranked by implementation experts [28]. These are established implementation science frameworks that are commonly used to structure and evaluate implementation processes and outcomes in health care research [35,36], and, therefore, organized data using terminology consistent with the implementation science field. CH was the primary analyst, iterated stages as necessary, and triangulated findings with senior researchers (PJ and CD; Figure 2). Results are presented as a narrative synthesis. To enhance synthesis clarity, we applied a frequency criterion adapted from Graham-Rowe and colleagues [37], whereby the number of studies associated with themes was considered in the narrative synthesis. The coding tree is reported in Multimedia Appendix 2.

Figure 2. Framework synthesis overview.



Reflexivity Statement

The authors have experience in intervention delivery and implementation and DMHI projects. The team includes a PhD candidate completing a doctorate on the implementation of a DMHI for paranoia (CH), clinical psychologists with experience delivering interventions to service users experiencing psychosis (PJ and ALT), an expert in psychosis research who has conceptualized a DMHI for paranoia included in our review (CBM-pa; JY) [38], researchers involved in evaluating an adapted version of this DMHI in an RCT (CH, SD, CF, ALT, and RT) [39], and a senior researcher with experience implementing health interventions (CD). We recognize that our professional and research backgrounds and interest in DMHI implementation may have influenced analyses. Our combined clinical and research experience may have provided a nuanced understanding of RCT implementation processes and mechanisms of impact. Direct engagement with DMHI participants provided insight into end user perspectives, which may have informed attention to the relational and practical aspects of implementation. Involvement of senior authors not directly engaged with DMHI users (PJ, CD, and JY) provided an external perspective that may have counterbalanced potential biases. Bias was also mitigated in quality assessment and interpretation, as JY was not involved in the quality rating of CBM-pa or the analysis of intervention findings.

XSL•FO
RenderX

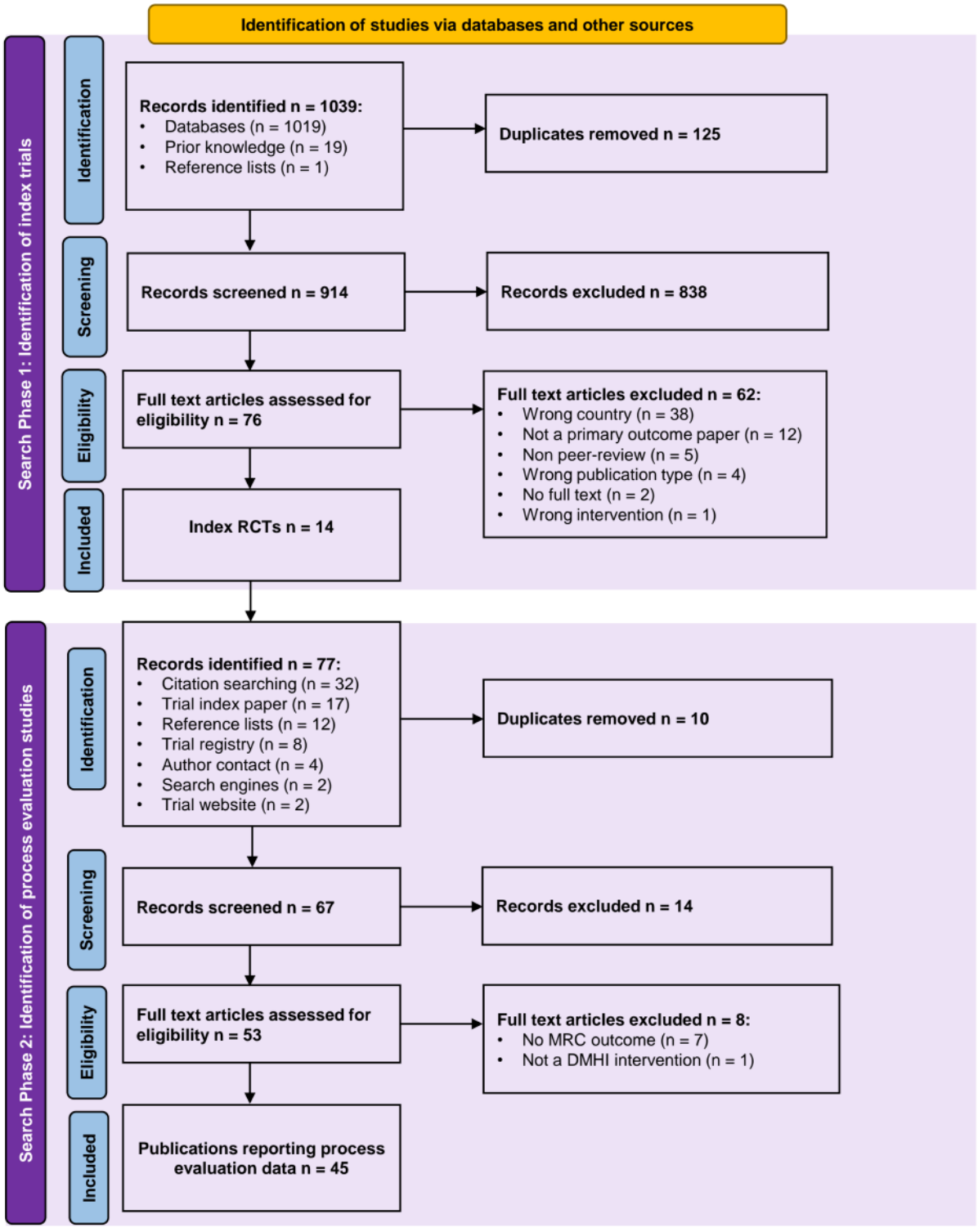
Results

Overview

The search results are summarized in Figure 3. Phase 1 initially identified 1039 papers, 914 of which were excluded after deduplication, primarily because they were protocol papers, were conducted outside of the United Kingdom, or were not primary outcome papers. In phase 1, a total of 14 index RCTs were included. Agreement in phase 1 was moderate to substantial for title/abstract screening (k=0.55-0.72), and substantial to perfect (k=0.62-1) at full text screening. Phase 2 screened 77 papers for process evaluation data (14 were index RCTs included in phase 1, and 63 papers were systematically identified in phase 2). From this, 10 duplicates and 22 papers (mostly protocol papers and papers not reporting process evaluation data) were excluded. In phase 2, a total of 45 papers reporting process evaluation data were included (13 index RCTs and 32 associated papers). Agreement in phase 2 was perfect for title/abstract screening (k=0.87-1) and substantial at full text (k=0.75).

XSL•FO
RenderX

Figure 3. Search flow diagram. DMHI: digital mental health intervention; MRC: Medical Research Council; RCT: randomized controlled trial.



Index RCT Methodological Quality

Quality ratings of index RCTs are included in Table S2 in Multimedia Appendix 2. Most disagreements arose in domain 2 (deviations from the intended interventions) and were resolved by discussion between raters (CH and RT). Half (7/14, 50%) RCTs showed some risk of bias, primarily due to bias arising from deviations from intended interventions or outcome measurement [40-44]. The remaining 7 RCTs had low risk of bias [38,45-47].

Index RCT Summary Characteristics

Summary characteristics of index RCTs are reported in full in Table S3 in Multimedia Appendix 2, and intervention characteristics are presented in Table 2. Most DMHIs targeted psychosis specific symptoms (12/14, 86%) and were delivered in a participant-chosen or any setting (9/14, 64%). Many (8/14,

XSL•FO
RenderX

57%) RCTs were conducted from a London-based university [38,44-47] or the University of Oxford [40-42]. One RCT was conducted outside of England in Scotland and Australia, but was included because less than half of the participants were Australian (24/73, 29%) [48]. Based on available data, intervention users were mostly White (656/1023, 64%), male (653/1023, 64%), had a mean age of 38.6 (SD 1.5) years, and a primary diagnosis of a psychotic disorder (890/1023, 87%).

Table 2. Intervention characteristics (N=14).

Intervention	Index RCT[a] reference	Intervention target		Delivery setting			Delivery mode		
		Psychosis specific symptom	General mental health	Any setting/participant choice	Clinical setting/study site	Not stated	Virtual reality	Computer/web	Smartphone
Actissist	Bucci et al (2018) [43]	✔		✔					✔
Actissist 2.0	Bucci et al (2024) [49]	✔		✔					✔
AVATAR	Craig et al (2018) [45]	✔			✔			✔	
AVATAR2	Garety et al 2024 [50]	✔		✔	✔			✔	
CBM-pa	Yiend et al (2023) [38]		✔		✔			✔	
EMPOWER	Gumley et al (2022) [48]	✔		✔					✔
__[b]	Freeman et al (2016) [40]	✔			✔		✔		
gameChange	Freeman et al (2022) [41]		✔	✔			✔		
—	Leff et al (2013) [44]	✔				✔		✔	
MUSE	Dodgson et al (2025) [51]	✔		✔				✔	
MUSE-FEP	Dudley et al (2024) [52]	✔		✔				✔	
SloMo	Garety et al (2021) [47]	✔		✔				✔	✔
THRIVE	Freeman et al (2023) [42]	✔		✔			✔		
V-NeST	Cella et al (2022) [46]	✔				✔	✔		

[a]RCT: randomized controlled trial.

[b]None given.

Studies Reporting Process Data Summary Characteristics

Summary characteristics of studies reporting process evaluation data are reported in full in Table S4 in Multimedia Appendix 2 and summarized in Table 3. Most studies were published between 2021 and 2023 (24/45, 53%), quantitative (19/45, 42%), used trial data (31/45, 68%), evaluated the study intervention (39/45, 87%), and did not describe the project as a process evaluation (43/45, 96%).

XSL•FO
RenderX

Table 3. Characteristics of studies reporting process evaluation data (N=45).

Process evaluation characteristics	Value, n (%)[a]
Publication year	
2013-2016	2 (4)
2017-2020	11 (24)
2021-2023	24 (53)
2024-2025	8 (18)
Study design	
Quantitative	19 (42)
Qualitative	16 (35)
Mixed methods	10 (22)
Process scope	
Study intervention	39 (87)
Intervention implementation	9 (20)
Trial procedures	9 (20)
Process evaluation data source	
Trial data	31 (68)
Additional stakeholder feedback	18 (40)
Author description	8 (18)
Labeled as process evaluation	
Yes	2 (4)
No	43 (96)

[a]Some papers fall into more than one category, so categories may add to more than the overall total (N=45).

There was a range of participants and contributors to studies reporting process evaluation data (Table S4 in Multimedia Appendix 2), with trial participants comprising the majority (1856/2155, 86%) across 37 (82%) studies. Other contributors included health care professionals (195/2155, 9%) and individuals with lived experience (184/2155, 9%). However, missing sociodemographic data in 17 (37%) studies limited the ability to accurately summarize this information.

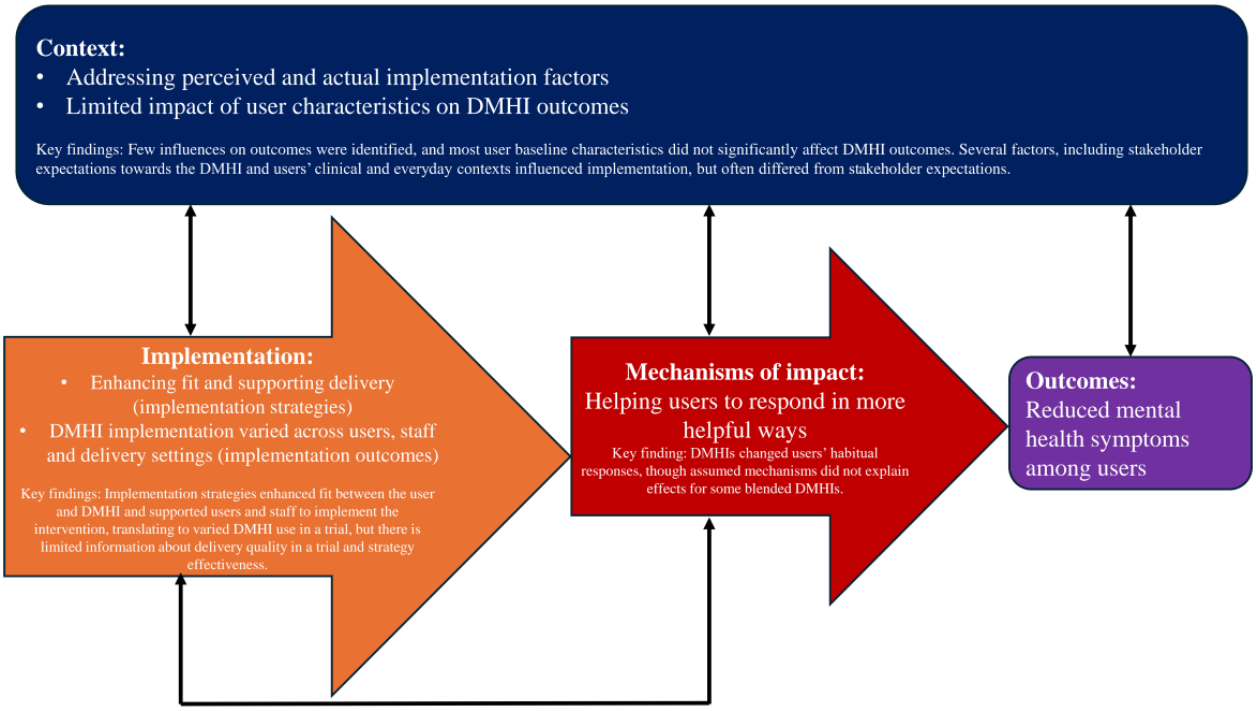
Framework Synthesis Findings

Overview

Framework synthesis of process evaluation data generated 5 themes (Figure 4 [13]), all of which were coded within the components deductively derived from the MRC framework (context, implementation, and mechanisms of impact) [13].

XSL•FO
RenderX

Figure 4. Framework synthesis themes summarized using the Medical Research Council process evaluation diagram. DMHI: digital mental health intervention.



Implementation

Overview

Most findings related to implementation, defined in this review as (1) the actions taken by trial teams to support DMHI delivery in an RCT (implementation strategies) [28] and (2) how much and by whom, a DMHI was used and delivered within an RCT, including if an intervention was delivered as planned (implementation outcomes) [27] (38/45, 84% studies; 13/13, 100% DMHIs). Two themes, mapped to the MRC framework, were created: (1) enhancing fit and supporting delivery (implementation strategies) and (2) DMHI implementation varied across users, staff, and delivery settings (implementation outcomes).

Enhancing Fit and Supporting Delivery (Implementation Strategies)

Implementation strategies were identified in 34 (76%) studies across all 13 (100%) DMHIs. These data, identified from stakeholder feedback, trial data, and author interpretation, were summarized in two subthemes: (1) strategies to improve fit and (2) strategies to support use and delivery.

Strategies to Improve Fit

Most trial teams (12/13, 92%) used implementation strategies that aimed to improve the fit between the DMHI and the user. This included the following implementation strategies from the ERIC: adapt and tailor to context, engage consumers, and change infrastructure [28]. Most trial teams (10/12, 83%) designed DMHIs that could be adapted to the user. This included allowing users to personalize virtual reality– (VR), web-, or app-based intervention content in 7 (58%) DMHIs [41,43,49,52-61], (2) offering on-demand access to 4 (33%) interventions, primarily via apps [43,44,49,53,55,59-62], and enabling multisetting delivery for 3 (25%) blended DMHIs [41,42,49,63]. These features could improve fit with users' needs, schedules, or delivery preferences. Stakeholders consistently valued DMHI flexibility [43,55,58,61] or requested it when lacking [17,62,64]. Half of the trial teams (6/12, 50%) also involved patients and potential consumers in the implementation effort, using co-design or user-centered design methods intended to make the DMHI more relevant to end users [49,60,64-69]. Three of these trial teams (Actissist 2.0, AVATAR2, and SloMo) attributed engagement to these approaches [47,49,55,69,70]. Finally, 3 RCTs (Actissist, Actissist 2.0, and MUSE) addressed potential infrastructural barriers by enabling access to the technology needed for the DMHI. This included delivering a blended DMHI using existing National Health Service (NHS) technology [71] and providing study phones, which were used by 70% [49] and 87% [43] of participants.

Strategies to Support Use and Delivery

Trial teams (11/12, 92%) also used implementation strategies that enhanced DMHI delivery and use by assisting intervention users and staff responsible for delivering the intervention. This included the following implementation strategies from the ERIC: training and educating stakeholders and providing interactive assistance [28]. Most strategies focused on supporting users in the RCT. Many (9/12, 75%) teams reported providing interactive assistance to users, primarily via staff involvement, though 2 DMHIs (SloMo and gameChange) also incorporated digitized human support [55,58,65]. In 2 DMHIs (CBM-pa and Actissist), trained researchers provided professional support specifically for the trial [53,64]. Across DMHIs, staff offered practical and technical assistance to facilitate DMHI use [17,46,53,58-60,62,65,72-74]. For blended DMHIs, users frequently highlighted the importance of positive relationships with the delivery staff. Positive relationships were developed

XSL•FO
RenderX

through delivery staff's nonspecific therapeutic skills, including active listening, empathy [17,55,58,60,62,73], and encouragement [58,62,73]. User feedback on interpersonal support was mixed: some users described assistance as central to engaging with the DMHI [58,62,73], while others viewed it as less important or preferred using the intervention independently [46,58,60,61,64]. No clear patterns of user preference emerged according to intervention features. Fewer trial teams reported implementation strategies aimed at supporting delivery staff. A total of 4 (33%) teams reported training and educating delivery staff on the DMHI prior to the RCT [51,55,65,75].

DMHI Implementation Varied Across Users, Staff, and Delivery Settings (Implementation Outcomes)

Implementation outcomes data were identified in 21 (47%) studies across all 13 (100%) DMHIs. Most implementation outcome data focused on individual-level implementation, particularly intervention reach and DMHI use, whereas less information was available about implementation by delivery staff or services.

User-level implementation outcomes are primarily related to intervention reach (the extent to which intended users participated in the intervention) and DMHI use within RCTs (13/13, 100% DMHIs). Reach was often operationalized as participation rates and reported by all but 1 team [49]. Across studies, the proportion of eligible individuals who participated in trials varied substantially, ranging from 27% [46] to 100% [38,41,42,45]. No clear patterns emerged regarding variation in participation rates according to DMHI characteristics or trial team.

Information about who was considered suitable to participate in DMHI trials was often incompletely reported. Several teams stated that participants were excluded because they did not meet eligibility criteria, without specifying which criteria were unmet [43,46,47,52]. Where reasons for exclusion were reported, data most commonly indicated participants were those who were currently experiencing the mental health difficulty targeted by the DMHI [41,42,45], and not receiving concurrent psychological therapy [38,41,45,52] or considered high risk, such as being acutely unwell, hospitalized, or experiencing suicidal intent [42,45]. Just 4 (31%) trials teams reported whether participants were representative of the intended target population, including those experiencing psychosis [38,47,50] or local service populations [51]. Where discussed, representativeness was primarily assessed using basic demographic characteristics. Trial samples were commonly described as predominantly White, male, or unemployed, with these characteristics used to justify representativeness compared to target populations [38,47,50,51].

Quantitative trial data indicated that DMHI use varied among users of both blended (DMHIs are delivered in a session and guided by a member of staff, 9/13, 69%) and self-administered (DMHIs are used independently, 4/13, 31%) delivery models. One intervention (SloMo) used both models, allowing in-session and independent DMHI use [47]. Session attendance for blended DMHIs ranged from 0% to 100% [41,42,45,46,50-52]. For self-administered DMHIs, use was operationalized using adherence criteria, with the proportion of users meeting this ranging between 65.4% [63] and 91% [48]. Two RCTs (Actissist and Actissist 2.0) reported detailed usage data showing that app engagement declined by 67%-67.5% over 12-week intervention periods, with users accessing the app at varied times of day [53,54].

Implementation outcomes relating to DMHI delivery staff and services were less frequently reported (9/13, 69% trial teams). A total of 8 (62%) trial teams reported the adoption (implementation by settings or delivery staff) of DMHIs in an RCT. However, these studies rarely reported formal adoption metrics, such as the proportion of settings or clinicians willing to deliver the DMHI. Instead, adoption was typically inferred indirectly from descriptions of delivery staff and their level of expertise (7/8, 88% trial teams). Across these studies, the representativeness of delivery staff appeared variable, although reporting was generally limited to professional role descriptions. Overall, 4 (58%) blended DMHIs were delivered by therapists who often had expertise in psychological therapies for psychosis [44,45,55]. In contrast, 3 (42%) DMHIs were delivered by staff without specialist psychotherapy roles [51,52,65], such as assistant psychologists or care coordinators. Further, a few trial teams reported how successfully delivery staff implemented the DMHI in an RCT. Only 3 (23%) of these teams assessed therapist adherence and skill in delivering AVATAR1 and MUSE therapies, reporting high fidelity to therapy manuals [45,51,52]. Another trial (CBM-pa) noted successful delivery without supporting data [38]. Finally, only 2 trial teams reported adoption at the service level. EMPOWER recruited and retained all 8 participating community mental health services throughout the trial [48], whereas the MUSE-FEP reported substantial site attrition, with 7 of 15 recruiting sites withdrawing during the trial due to the COVID-19 pandemic [52].

Context

Overview

All interventions reported contextual findings, or external factors that influenced DMHI implementation or functioning in RCTs (33/45, 73% studies; 13/13, 100% DMHIs). Contextual findings were explored quantitatively using moderation analyses (10/33, 30% studies) [41,42,47,48,50,53,54,63,76,77], and qualitatively using stakeholder consultation (11/33, 33% studies) [16,17,46,58,60,61,64,73,78,79], and reasons for trial participation or withdrawal (12/33, 36% studies) [38,43,45,46,49-52,55,58,75,80]. Two themes that aligned with the MRC framework components were created: (1) addressing perceived and actual implementation factors and (2) the limited impact of user characteristics on DMHI outcomes [13].

Addressing Perceived and Actual Implementation Factors

Overview

Contextual data about implementation (factors that influenced how a DMHI was delivered in RCTs) were identified in 26 (58%) studies across 12 (92%) DMHIs. Three subthemes were created from stakeholder feedback [16,17,46,55,57,58,60-62,64,73,78,79], trial data [43-46,48,50,52-55,63,71,75,80], and author description [59]: (1) stakeholder misconceptions about user-related barriers, (2)

XSL•FO
RenderX

service-level barriers: perceptions and realities, and (3) hope and help: positive expectations facilitate implementation.

Stakeholder Misconceptions About User-Related Barriers

Health care professionals and some service users held generalized concerns that psychosis-related and technology-related factors would hinder DMHI implementation. Staff in pre–COVID-19 studies anticipated that users' limited technology access and digital skills would act as key implementation barriers [78,79]. However, process evaluation data contradicted these assumptions: users engaged with a range of DMHI modalities, including computer-based, smartphone-based, and VR-based interventions, regardless of digital confidence [17,46,55,60,64]. Only SloMo reported higher engagement among more technologically confident users before the RCT [63]. Concerns about users' technology-related paranoia or illness severity were also overgeneralized [57,78,79] as not all users had experienced technology-related paranoia [62,73], and some used DMHIs despite concerns [60,67,73].

Actual implementation barriers included different psychosis symptoms (auditory hallucinations [44,59,62] and negative symptoms [46]), everyday life and situational factors outside of therapy (eg, users having other time commitments such as working or studying [17,43,53,73]), and DMHIs not meeting users' treatment needs or preferences (eg, lack of interest in the DMHI or RCT [41-43,45-48,51,52,70] or feeling that the DMHI was not suited to them [17,43,45,55,57,75]). Demographic factors had a limited influence on use in 3 (25%) DMHIs: older age [53,54], White ethnicity [54], and higher baseline smartphone confidence [63] were associated with higher engagement. A number of other factors were generally nonsignificant (gender, education, employment, and prior smartphone ownership) [53,54,63].

Service-Level Barriers: Perceptions and Realities

Stakeholders (service users and staff) often perceived psychosis-specialist services as underresourced for DMHI delivery, citing staff availability, physical space, and technology constraints [57,78,79]. These concerns were particularly noted for interventions requiring clinical involvement (EMPOWER and gameChange) [78,79], the latter with VR-specific equipment and spatial needs [45]. Evidence from 2 RCTs (EMPOWER and MUSE) indicated staff resource constraints linked to high workloads and turnover hindered recruitment [16], staff training [51], and data collection [48]. In contrast, staff perceptions for Actissist did not align with its delivery model [43,78], as the smartphone-based intervention required neither the assumed NHS systems nor staff support for delivery [43].

Hope and Help: Positive Expectations Facilitate Implementation

Positive expectations among users and health care professionals facilitated DMHI use. Most users tried the DMHI expecting personal benefit from use [17,58,60,62,73,79,80], while some users were motivated by helping others [62,80] or trying something different within their treatment [46,62]. Staff reasons for wanting to deliver a DMHI, though less frequently reported, aligned with user expectations as some anticipated benefits for their clients [57,78].

Limited Impact of User Characteristics on DMHI Outcomes

A few contextual factors influenced DMHI functioning. Only 2 (15%) DMHIs (gameChange and MUSE) identified 3 moderators, reporting enhanced effects among participants with higher baseline illness severity [76], earlier psychosis symptom onset [50], and those who felt valued from using the intervention [77]. Across 5 (38%) DMHIs, most user baseline characteristics did not affect outcomes; no consistent patterns emerged, though gender was not found to moderate effects [41,42,47,50,52].

Mechanisms of Impact: Helping Users to Respond in More Helpful Ways

Mechanisms of impact data, or how DMHIs produced change in an RCT, were identified in most (10/13, 77%) DMHIs across 19 (42%) studies. Evidence predominantly came from qualitative sources, including DMHI user feedback (8/19, 42% studies [17,43,58,60-62,64,73]), intervention content (4/19, 21% studies [55,72,74,75]), and author commentary (2/19, 10% studies [51,59]), while only 4 studies quantitatively tested hypothesized mechanisms [38,41,42,47].

Across the 10 (77%) interventions, DMHIs appeared to help users respond to difficult thoughts or situations in more helpful ways by changing their habitual thinking (eg, belief inflexibility, beliefs about voices, or interpretation biases [17,38,41,45,47,55,60,62,64,72,75]) or behaviors (eg, reducing defense behaviors or increasing assertive behaviors [41,42,45,55,58,59,61,72-74]). Many (7/13, 54%) interventions were theoretically informed and targeted putative maintaining mechanisms [17,38,41-43,47,49]. Changes were achieved by replacing unhelpful responses (eg, negative interpretations of everyday situations or avoidance of everyday social situations that were perceived to be unsafe) with more helpful responses (eg, seeking alternative, less personally threatening explanations [38,47,55,58,60,62,64], or entering challenging situations [41,58]). For some users, changing habitual responses was achieved by becoming aware they were responding in unhelpful ways [55,58,60,61,64], understanding why they responded in unhelpful ways [55,72,75], or practicing responding in new ways [55,58,60,62].

However, assumed mechanisms did not consistently explain outcomes in 4 DMHIs with a blended delivery model delivered via an app [47,48] or VR [41,42]. In 3 of these DMHIs, targeted mechanisms (safety beliefs [41,42], defense behaviors [42], or a reasoning bias [47]) did not mediate intervention effects, even when symptom improvement occurred [41,47]. In another DMHI, EMPOWER, user interviews revealed that not all users experienced changes to the targeted mechanism (relapse appraisal) [17].

Discussion

Principal Findings

Overview

We created 5 themes after synthesizing process evaluation data from 45 peer-reviewed papers evaluating 13 DMHIs used by those experiencing psychosis in RCTs predominantly based in the United Kingdom. These themes aligned with deductive

XSL•FO
RenderX

MRC framework domains: implementation, context, and mechanisms of impact [13].

Implementation

Overall, implementation findings indicated that research in trials primarily focused on the implementation at the level of the individual user, with substantially less information available regarding delivery staff and services. Current DMHI implementation research for psychosis has therefore prioritized understanding user engagement, while providing limited insight into whether services can realistically deliver DMHIs.

Implementation data suggested it may be feasible to deliver DMHIs to people experiencing psychosis in a trial. Across studies, users engaged with DMHIs in a trial setting. Where considered, trial participants were often described as broadly representative of populations with psychosis or local services [38,47,50,71], suggesting that DMHI delivery may be feasible among groups currently receiving psychosis care. However, representativeness was inconsistently assessed and typically based on basic demographic characteristics. Further, while DMHI use in a trial varied between individuals, this finding is consistent with previous reviews [10,81,82] and attendance patterns observed in psychological therapies for psychosis more broadly [83]. Consistency in variation may suggest that some level of disengagement could be expected when implementing DMHIs and, importantly, may not indicate implementation failure unique or issues inherent in this delivery mode. Variation in use persisted across blended and self-delivered DMHI using different technology, suggesting that DMHI use may be shaped by external, nontherapy related factors, including fluctuating needs, user autonomy in care, or unmet treatment preferences, all factors identified in this (see the "Context" section below) and other work [84].

Implementation strategies similarly primarily focused on addressing potential barriers at the user level, including intervention accessibility, usability, and user support needs. Trial teams often aimed to improve the fit between the DMHI and the user through intervention flexibility, stakeholder involvement, tailored support, and improving access to necessary technology. These approaches align with broader recommendations both in the DMHI [11,85] and implementation science [28] literature. Consequently, although implementation science frameworks were rarely explicitly used, trial teams were often already undertaking implementation-focused work. However, greater use of implementation science frameworks may help structure, evaluate, and consistently report these efforts and enable a better understanding of how DMHIs can be successfully implemented. Stakeholder feedback highlighted some of these strategies, including ensuring the intervention content, delivery approach, and amount of professional support provided [43,55,58,61], as potential design priorities for future work. However, despite routine use of promising implementation strategies, no studies evaluated whether these strategies improved implementation outcomes, and variation in DMHI use persisted across interventions. Further, despite DMHIs often being discussed as scalable or resource-efficient interventions [86], implementation costs were not reported in the included studies. Our review challenges assumptions that DMHIs are inherently low-resource interventions because many implementation strategies likely require substantial resources, including money (eg, payment for specialist workers and end users in co-design processes) and time (eg, developing and delivering training). Future work clarifying these "hidden" costs, in addition to evaluating implementation strategy effectiveness, would better establish how best to implement DMHIs for psychosis.

In contrast, evidence relating to delivery staff and services remained limited across implementation strategies and outcomes. Available evidence primarily focused on the delivery staff, with less attention given to wider service or organizational implementation requirements. Although all included DMHIs included professional support in delivery, only 3 (23%) trial teams evaluated delivery staff competency [45,51,52], despite staff being important determinants of successful delivery [10]. Few studies reported how staff were trained to deliver DMHIs, limiting understanding of the workforce support required to deliver these interventions consistently and to a high standard. However, blended DMHIs appeared deliverable by staff with differing levels of expertise, ranging from highly specialized therapists to broader mental health workforces in the United Kingdom. This suggests that workforce requirements and implementation demands may differ depending on the specific DMHI. Beyond delivery staff, implementation within health care services is also shaped by organizational factors, including available technology, delivery space, and financial resources [10]. These factors were rarely evaluated in included studies, limiting understanding of what resources are needed to implement DMHIs both in and after trials.

Context

Most studies (33/45, 73%) reported contextual findings, primarily relating to DMHI implementation, whereas few examined the influence of context on DMHI outcomes. Only 2 (15%) DMHIs identified moderators, including some illness factors and feeling valued from DMHI use [50,76]. Across 5 (38%) DMHIs, most user baseline characteristics did not significantly affect treatment outcomes [41,42,47,50,52]. While this could suggest that included DMHIs may work broadly across different user profiles, conclusions about subgroup differences remain tentative due to the small number of studies examining moderators and variability in examined moderators. Notably, the absence of gender moderation across several DMHIs contrasts with previous reviews [10,82], potentially reflecting differences in intervention type or populations. These inconsistencies highlight the need for more nuanced research to determine for whom DMHIs are most effective.

A key finding was the role of stakeholder DMHI appraisals. First, appraisals of the intervention shaped implementation, including stakeholder appraisals about DMHI relevance, fit with treatment needs, and expected benefits [20,48,52,64]. This emphasizes the importance of supporting user choice and aligning DMHIs with individual preferences (see the "Implementation" section above). Second, stakeholder misconceptions about user- and service-related barriers were common. Pretrial assumptions often overestimated difficulties due to users' limited digital literacy, device availability, or

XSL•FO
RenderX

technology-related paranoia [78,79], yet process evaluation data showed users engaged with a range of DMHI modalities regardless of prior digital experience [17,46,55,60,64], and concerns regarding technology-related paranoia were overgeneralized [62,73]. Only 1 intervention (SloMo) demonstrated higher engagement among more digitally confident users [63], suggesting that technological familiarity may enhance, but is not essential for, use. Service-level constraints were also highlighted. Service staff often cited limited resources as barriers [78], though evidence linking resource limitations to implementation was limited but highlighted staff availability as a relevant factor [16,48]. Further, staff perceptions did not always match actual intervention demands [78]. Given staff are important determinants of DMHI implementation [11,87], addressing misconceptions about DMHI users and delivery requirements may be critical. However, most studies predated COVID-19 [78,79], meaning attitudes toward DMHIs and user capability may have changed after the pandemic forced a move to almost entirely remote services in the United Kingdom due to 2 years of prolonged lockdowns [88].

User-reported barriers differed from assumed factors, including other psychosis-related symptoms (auditory hallucinations [59,62] and negative symptoms [46]) and life commitments outside of therapy [17,53]. These findings underscore the importance of addressing real, not assumed, contextual factors; as DMHI use is shaped by both clinical and everyday contexts, this may further reinforce the potential value of flexible intervention design to optimize use (see the “Implementation” section above).

Mechanisms of Impact

Evidence for DMHI mechanisms remains limited. Process evaluation data suggested DMHIs reduced users’ mental health symptoms by supporting more adaptive cognitive and behavioral responses to challenging situations. This mechanism was observed across different intervention modes and targets, though the specific mechanisms varied and were often not clearly delineated as such, which impeded review syntheses. However, evidence for these mechanisms is tentative and is currently insufficiently tested and validated. Most insights were derived from qualitative sources (user feedback or intervention content), rather than recommended quantitative hypothesis-driven approaches [89], which were reported in only 4 (10%) studies [38,41,42,47]. This limited the ability to confirm whether these processes causally explained symptom change. Further, hypothesized mechanisms were not consistently activated. Some blended DMHIs demonstrated effects through assumed mechanisms, such as belief inflexibility or defense behaviors [38,64], while others reported symptom improvements that were not mediated by hypothesized mechanisms (safety beliefs, reasoning, and safety behaviors) [41,42,47], suggesting that DMHIs may work through alternative or nonspecific processes. For interventions delivered via blended delivery models, intrapersonal factors, such as a therapeutic alliance, which has been identified as a driver of change when treating psychosis [90], may play an unrecognized role in driving change. Further, variability in mechanistic responses, such as inconsistent changes to reduced fear of relapse among EMPOWER users [17], highlights the likelihood that mechanisms function differently across users and contexts (see the “Context” section above).

Comparison to Prior Work

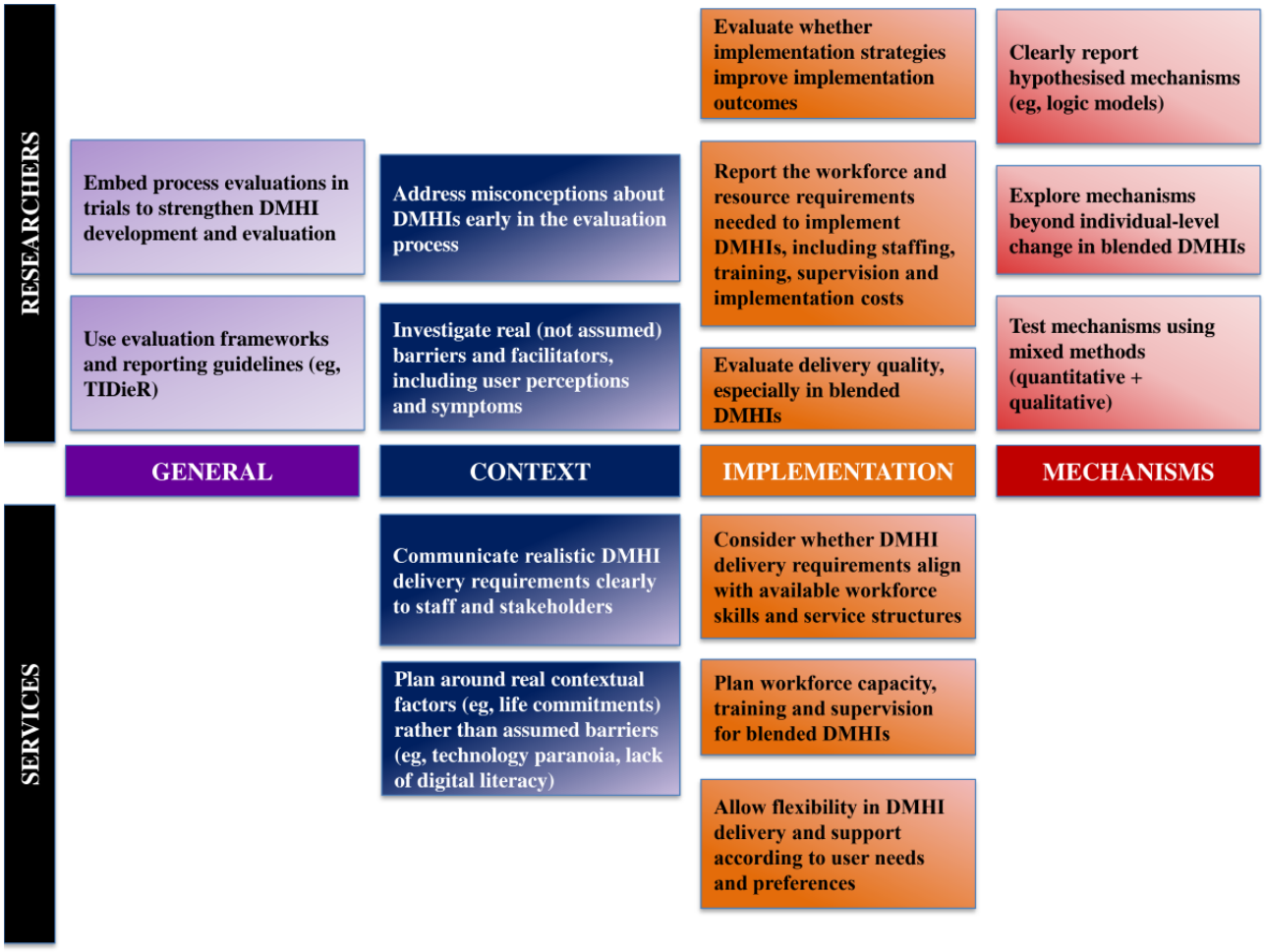
Unlike previous reviews focusing on the effectiveness of DMHIs for psychosis [7,8,91], this synthesis of process evaluation data extends the understanding of implementation science in this area. Further, our review extends prior syntheses of process evaluation in other health domains, such as physical health interventions [18,92]. Consistent with broader findings that process evaluation remains underused across research fields [18,92] and that implementation science remains underapplied in DMHI research [93], we found limited use of process evaluation, with only 1 trial explicitly adopting this approach [16,17]. Although most DMHI research remains service-based [9,10], our findings show that both user-related and service-related factors can shape engagement and outcomes, and may provide a starting point for identifying priorities for future research and implementation efforts in and outside of services.

Future Directions

Our review highlights several priorities to strengthen DMHI evaluation and implementation (Figure 5). For researchers, key steps include embedding process evaluation in trials using established frameworks and reporting guidelines, further clarification of how DMHIs reduce user mental health symptoms, and routine assessment of the delivery quality of blended DMHIs. Future research could evaluate which implementation strategies improve implementation outcomes. Further work is needed to clarify the workforce, organizational, and resource requirements needed to deliver DMHIs consistently, including staff training, supervision, and implementation costs. For services, priorities include clearly communicating DMHI requirements to staff, addressing real, not assumed, contextual factors, and adopting flexible delivery approaches.

XSL•FO
RenderX

Figure 5. Recommendations for researchers and services. DMHI: digital mental health intervention; TIDieR: Template for Intervention Description and Replication.



Review Strengths and Limitations

This review has several strengths. To our knowledge, it is the first review to synthesize process evaluation data in this field. Broad inclusion criteria, aligned with relevant guidance [20], enabled us to capture relevant studies irrespective of authors' use of the term "process evaluation." The lack of date and method restrictions and updated searches ensured comprehensive and up-to-date coverage. Rigorous methods, including review preregistration, double, masked screening, and data extraction and use of an established process evaluation framework for synthesis, helped minimize bias. Finally, unlike other scoping reviews [94], our use of established implementation science frameworks [13,27,28] helped ensure our findings were appropriately documented, which may help inform the broader field.

Several limitations should be noted. As in another review [92], most studies did not explicitly use process evaluation terminology, and data were primarily synthesized by a single author (CH), which may have introduced subjectivity despite the use of an established framework. The framework was operationalized at the theme level to enhance usability, possibly limiting analytical depth. Use of the frequency-based criterion from Graham-Rowe and colleagues [37] in the narrative synthesis may underemphasize less commonly reported findings. Extracting process evaluation data from only the results and discussion sections may have excluded relevant information from the methods sections. Many trial teams (10/13, 77%) described implementation processes (eg, staff training and device provision) in methods sections without supporting data, which were not captured in our syntheses. Finally, 1 included DMHI was implemented in the United Kingdom and Australia [48], which may have introduced variability in country contexts. While we excluded data explicitly linked to Australian contributors to mitigate this risk, some aggregated data may reflect mixed contexts.

Conclusions

There is preliminary evidence that DMHIs can be delivered to people experiencing psychosis in a trial. Where assessed, interventions were generally delivered as intended and often reached participants considered representative of existing psychosis services. DMHIs may reduce symptoms by modifying users' habitual behavioral and cognitive responses, yet mechanistic pathways require further validation. Implementation strategies aimed to improve the fit between DMHIs and users through flexible delivery and stakeholder involvement, and support DMHI use and delivery in trials through professional support and staff training. While these approaches appear promising, engagement with DMHIs varied between users, and the effectiveness of implementation strategies remains largely untested. Contextual factors often differed from stakeholder

XSL•FO
RenderX

expectations, relating to users' everyday factors outside of therapy, treatment expectations, or symptoms, rather than assumed technology-related issues, highlighting the need to address misconceptions and clearly communicate delivery requirements. Process evaluation remains underused; embedding it in future trials is a priority and could clarify mechanisms, guide implementation, and support the translation of effective DMHIs from trials to practice.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Funding

CH is supported by a fully funded University Research Studentship Award from the University of Bath. The funder had no role in the design, conduct, analysis, interpretation, or reporting of this review. Any opinion expressed in the paper does not necessarily reflect the view of the funder. At the time of this work JY was part-funded by the NIHR Biomedical Research Centre hosted at South London and Maudsley NHS Foundation Trust in partnership with King's College London and the Medical Research Council (MRC) Biomedical Catalyst: Developmental Pathway Funding Scheme (DPFS), Reference: MR/V027484/1.

Authors' Contributions

Conceptualization: CH
Methodology: CH, CD, PJ
Data curation: CH
Funding acquisition: CH (lead) PJ (support)
Project administration: CH (lead), PJ, (support), CD (support)
Validation: CH (lead), SD, (support) CF (support), ALT (support), RT (support)
Visualization: CH
Formal analysis: CH (lead), CD (support), PJ (support)
Investigation: CH, SD, CF, ALT, RT
Supervision: PJ (joint), CD (joint), JY (support)
Writing—original draft: CH
Writing—review and editing: CD, SD, CF, PJ, ALT, RT, JY

Conflicts of Interest

JY is the chief investigator on one of the included interventions included in the review (CBM-pa). However, JY was not involved in the quality rating of this index paper or the analysis of intervention findings. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Completed review reporting checklists.
[DOCX File , 53 KB - humanfactors_v13i1e88032_app1.docx]

Multimedia Appendix 2
Review supplementary materials.
[DOCX File , 351 KB - humanfactors_v13i1e88032_app2.docx]

References

1. Kuipers E, Yesufu-Udechuku A, Taylor C, Kendall T. Management of psychosis and schizophrenia in adults: summary of updated NICE guidance. BMJ 2014;348:g1173. [doi: 10.1136/bmj.g1173] [Medline: 24523363]
2. Khanna S. The impact of psychosis on mental health and quality of life of an individual. In: Psychosis-From Cell to Society. London: IntechOpen; 2025.
3. Ince P, Haddock G, Tai S. A systematic review of the implementation of recommended psychological interventions for schizophrenia: rates, barriers, and improvement strategies. Psychol Psychother 2016;89(3):324-350. [doi: 10.1111/papt.12084] [Medline: 26537838]
4. Colling C, Evans L, Broadbent M, Chandran D, Craig TJ, Kolliakou A, et al. Identification of the delivery of cognitive behavioural therapy for psychosis (CBTp) using a cross-sectional sample from electronic health records and open-text information in a large UK-based mental health case register. BMJ Open 2017;7(7):e015297 [FREE Full text] [doi: 10.1136/bmjopen-2016-015297] [Medline: 28716789]
5. Meyerhoff J, Kornfield R, Lattie EG, Knapp AA, Kruzan KP, Jacobs M, et al. From formative design to service-ready therapeutic: a pragmatic approach to designing digital mental health interventions across domains. Internet Interv 2023;34:100677 [FREE Full text] [doi: 10.1016/j.invent.2023.100677] [Medline: 37808416]

XSL•FO
RenderX

6. Digital health technologies to help manage symptoms of psychosis and prevent relapse in adults and young people: early value assessment. NICE. URL: https://www.nice.org.uk/guidance/hte17 [accessed 2025-02-06]
7. Clarke S, Hanna D, Mulholland C, Shannon C, Urquhart C. A systematic review and meta-analysis of digital health technologies effects on psychotic symptoms in adults with psychosis. Psychosis 2019;11(4):362-373. [doi: 10.1080/17522439.2019.1632376]
8. Morales-Pillado C, Fernández-Castilla B, Sánchez-Gutiérrez T, González-Fraile E, Barbeito S, Calvo A. Efficacy of technology-based interventions in psychosis: a systematic review and network meta-analysis. Psychol Med 2023;53(13):6304-6315 [FREE Full text] [doi: 10.1017/S0033291722003610] [Medline: 36472150]
9. Aref-Adib G, McCloud T, Ross J, O'Hanlon P, Appleton V, Rowe S, et al. Factors affecting implementation of digital health interventions for people with psychosis or bipolar disorder, and their family and friends: a systematic review. Lancet Psychiatry 2019;6(3):257-266. [doi: 10.1016/S2215-0366(18)30302-X] [Medline: 30522979]
10. Eisner E, Faulkner S, Allan S, Ball H, Di Basilio D, Nicholas J, et al. Barriers and facilitators of user engagement with digital mental health interventions for people with psychosis or bipolar disorder: systematic review and best-fit framework synthesis. JMIR Ment Health 2025;12:e65246 [FREE Full text] [doi: 10.2196/65246] [Medline: 39832352]
11. Torous J, Linardon J, Goldberg SB, Sun S, Bell I, Nicholas J, et al. The evolving field of digital mental health: current evidence and implementation issues for smartphone apps, generative artificial intelligence, and virtual reality. World Psychiatry 2025;24(2):156-174 [FREE Full text] [doi: 10.1002/wps.21299] [Medline: 40371757]
12. McGinty EE, Alegria M, Beidas RS, Braithwaite J, Kola L, Leslie DL, et al. The Lancet Psychiatry Commission: transforming mental health implementation research. Lancet Psychiatry 2024;11(5):368-396. [doi: 10.1016/S2215-0366(24)00040-3] [Medline: 38552663]
13. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Ccouncil guidance. BMJ 2015;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
14. Omer S, Golden E, Priebe S. Exploring the mechanisms of a patient-centred assessment with a solution focused approach (DIALOG+) in the community treatment of patients with psychosis: a process evaluation within a cluster-randomised controlled trial. PLoS One 2016;11(2):e0148415 [FREE Full text] [doi: 10.1371/journal.pone.0148415] [Medline: 26859388]
15. Pemovska T, Jovanović N, Radojičić T, Arënliu A, Džubur-Kulenović A, Novotni A, et al. Protocol for a process evaluation of a cluster randomised controlled trial to improve psychosocial treatment of patients with psychotic spectrum disorders: the IMPULSE trial. Glob Psychiatry Arch 2021;4(1):20-30. [doi: 10.52095/gp.2020.1407]
16. Allan S, McLeod H, Bradstreet S, Bell I, Whitehill H, Wilson-Kay A, et al. Perspectives of trial staff on the barriers to recruitment in a digital intervention for psychosis and how to work around them: qualitative study within a trial. JMIR Hum Factors 2021;8(1):e24055 [FREE Full text] [doi: 10.2196/24055] [Medline: 33666555]
17. Allan S, Beedie S, McLeod HJ, Farhall J, Gleeson J, Bradstreet S, et al. Using EMPOWER in daily life: a qualitative investigation of implementation experiences. BMC Psychiatry 2023;23(1):597 [FREE Full text] [doi: 10.1186/s12888-023-05096-x] [Medline: 37592231]
18. Liu Y, Wadey CA, Barker AR, Williams CA. Process evaluation of school-based high-intensity interval training interventions for children and adolescents: a systematic review and meta-analysis of randomized controlled trials. BMC Public Health 2024;24(1):348 [FREE Full text] [doi: 10.1186/s12889-024-17786-6] [Medline: 38308213]
19. French C, Pinnock H, Forbes G, Skene I, Taylor SJC. Process evaluation within pragmatic randomised controlled trials: what is it, why is it done, and can we find it?-a systematic review. Trials 2020;21(1):916 [FREE Full text] [doi: 10.1186/s13063-020-04762-9] [Medline: 33168067]
20. Cargo M, Harris J, Pantoja T, Booth A, Harden A, Hannes K, et al. Cochrane Qualitative and Implementation Methods Group guidance series-paper 4: methods for assessing evidence on intervention implementation. J Clin Epidemiol 2018;97:59-69. [doi: 10.1016/j.jclinepi.2017.11.028] [Medline: 29223325]
21. Booth A, Moore G, Flemming K, Garside R, Rollins N, Tunçalp Ö, et al. Taking account of context in systematic reviews and guidelines considering a complexity perspective. BMJ Glob Health 2019;4(Suppl 1):e000840 [FREE Full text] [doi: 10.1136/bmjgh-2018-000840] [Medline: 30775011]
22. Booth A, Noyes J, Flemming K, Moore G, Tunçalp Ö, Shakibazadeh E. Formulating questions to explore complex interventions within qualitative evidence synthesis. BMJ Glob Health 2019;4(Suppl 1):e001107 [FREE Full text] [doi: 10.1136/bmjgh-2018-001107] [Medline: 30775019]
23. Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
24. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
25. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. BMC Med Res Methodol 2012;12:181 [FREE Full text] [doi: 10.1186/1471-2288-12-181] [Medline: 23185978]

XSL•FO
RenderX

26. Flemming K, Booth A, Hannes K, Cargo M, Noyes J. Cochrane Qualitative and Implementation Methods Group guidance series-paper 6: reporting guidelines for qualitative, implementation, and process evaluation evidence syntheses. J Clin Epidemiol 2018;97:79-85 [FREE Full text] [doi: 10.1016/j.jclinepi.2017.10.022] [Medline: 29222060]
27. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
28. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015;10:21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
29. Carroll C, Booth A. Quality assessment of qualitative evidence for systematic review and synthesis: is it meaningful, and if so, how should it be performed? Res Synth Methods 2015;6(2):149-154 [FREE Full text] [doi: 10.1002/jrsm.1128] [Medline: 26099483]
30. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
31. MacLeod A, Burm S, Mann K. Constructivism: learning theories and approaches to research. In: Researching Medical Education, 2nd Edition. New Jersey: John Wiley & Sons, Ltd; 2022:25-40.
32. Lawani A. Critical realism: what you should know and how to apply it. Qual Res J 2020;21(3):320-333. [doi: 10.1108/qrj-08-2020-0101]
33. Hong QN, Pluye P, Bujold M, Wassef M. Convergent and sequential synthesis designs: implications for conducting and reporting systematic reviews of qualitative and quantitative evidence. Syst Rev 2017;6(1):61 [FREE Full text] [doi: 10.1186/s13643-017-0454-2] [Medline: 28335799]
34. Booth A, Caroll C. Framework synthesis. In: Noyes J, Harden A, editors. Cochrane-Campbell Handbook for Qualitative Evidence Synthesis. London, UK: Cochrane; 2023.
35. Ashcraft LE, Goodrich DE, Hero J, Phares A, Bachrach RL, Quinn DA, et al. A systematic review of experimentally tested implementation strategies across health and human service settings: evidence from 2010-2022. Implement Sci 2024;19(1):43 [FREE Full text] [doi: 10.1186/s13012-024-01369-5] [Medline: 38915102]
36. Glasgow R, Battaglia C, McCreight M, Ayele R, Maw A, Fort M, et al. Use of the reach, effectiveness, adoption, implementation, and maintenance (RE-AIM) framework to guide iterative adaptations: applications, lessons learned, and future directions. Front Health Serv 2022;2:959565 [FREE Full text] [doi: 10.3389/frhs.2022.959565] [Medline: 36925843]
37. Graham-Rowe E, Lorencatto F, Lawrenson JG, Burr JM, Grimshaw JM, Ivers NM, et al. Barriers to and enablers of diabetic retinopathy screening attendance: a systematic review of published and grey literature. Diabet Med 2018;35(10):1308-1319. [doi: 10.1111/dme.13686] [Medline: 29790594]
38. Yiend J, Lam CLM, Schmidt N, Crane B, Heslin M, Kabir T, et al. Cognitive bias modification for paranoia (CBM-pa): a randomised controlled feasibility study in patients with distressing paranoid beliefs. Psychol Med 2023;53(10):4614-4626 [FREE Full text] [doi: 10.1017/S0033291722001520] [Medline: 35699135]
39. Yiend J, Taher R, Fialho C, Hampshire C, Hsu C, Kabir T, MPIT, Avegen. Assessing the efficacy and safety of STOP (successful treatment for paranoia)-an app-based cognitive bias modification therapy for paranoia: a randomised clinical trial protocol. Trials 2024;25(1):806 [FREE Full text] [doi: 10.1186/s13063-024-08570-3] [Medline: 39623444]
40. Freeman D, Bradley J, Antley A, Bourke E, DeWeever N, Evans N, et al. Virtual reality in the treatment of persecutory delusions: randomised controlled experimental study testing how to reduce delusional conviction. Br J Psychiatry 2016;209(1):62-67 [FREE Full text] [doi: 10.1192/bjp.bp.115.176438] [Medline: 27151071]
41. Freeman D, Lambe S, Kabir T, Petit A, Rosebrock L, Yu L, gameChange Trial Group. Automated virtual reality therapy to treat agoraphobic avoidance and distress in patients with psychosis (gameChange): a multicentre, parallel-group, single-blind, randomised, controlled trial in England with mediation and moderation analyses. Lancet Psychiatry 2022;9(5):375-388 [FREE Full text] [doi: 10.1016/S2215-0366(22)00060-8] [Medline: 35395204]
42. Freeman D, Lister R, Waite F, Galal U, Yu L, Lambe S, et al. Automated virtual reality cognitive therapy versus virtual reality mental relaxation therapy for the treatment of persistent persecutory delusions in patients with psychosis (THRIVE): a parallel-group, single-blind, randomised controlled trial in England with mediation analyses. Lancet Psychiatry 2023;10(11):836-847 [FREE Full text] [doi: 10.1016/S2215-0366(23)00257-2] [Medline: 37742702]
43. Bucci S, Barrowclough C, Ainsworth J, Machin M, Morris R, Berry K, et al. Actissist: proof-of-concept trial of a theory-driven digital intervention for psychosis. Schizophr Bull 2018;44(5):1070-1080 [FREE Full text] [doi: 10.1093/schbul/sby032] [Medline: 29566206]
44. Leff J, Williams G, Huckvale MA, Arbuthnot M, Leff AP. Computer-assisted therapy for medication-resistant auditory hallucinations: proof-of-concept study. Br J Psychiatry 2013;202:428-433. [doi: 10.1192/bjp.bp.112.124883] [Medline: 23429202]
45. Craig TK, Rus-Calafell M, Ward T, Leff JP, Huckvale M, Howarth E, et al. AVATAR therapy for auditory verbal hallucinations in people with psychosis: a single-blind, randomised controlled trial. Lancet Psychiatry 2018;5(1):31-40 [FREE Full text] [doi: 10.1016/S2215-0366(17)30427-3] [Medline: 29175276]

XSL•FO
RenderX

46. Cella M, Tomlin P, Robotham D, Green P, Griffiths H, Stahl D, et al. Virtual reality therapy for the negative symptoms of schizophrenia (V-NeST): a pilot randomised feasibility trial. Schizophr Res 2022;248:50-57 [FREE Full text] [doi: 10.1016/j.schres.2022.07.013] [Medline: 35939920]
47. Garety P, Ward T, Emsley R, Greenwood K, Freeman D, Fowler D, et al. Effects of slowMo, a blended digital therapy targeting reasoning, on paranoia among people with psychosis: a randomized clinical trial. JAMA Psychiatry 2021;78(7):714-725 [FREE Full text] [doi: 10.1001/jamapsychiatry.2021.0326] [Medline: 33825827]
48. Gumley AI, Bradstreet S, Ainsworth J, Allan S, Alvarez-Jimenez M, Aucott L, et al. The EMPOWER blended digital intervention for relapse prevention in schizophrenia: a feasibility cluster randomised controlled trial in Scotland and Australia. Lancet Psychiatry 2022;9(6):477-486 [FREE Full text] [doi: 10.1016/S2215-0366(22)00103-1] [Medline: 35569503]
49. Bucci S, Berry N, Ainsworth J, Berry K, Edge D, Eisner E, et al. Effects of Actissist, a digital health intervention for early psychosis: a randomized clinical trial. Psychiatry Res 2024;339:116025 [FREE Full text] [doi: 10.1016/j.psychres.2024.116025] [Medline: 38870774]
50. Garety PA, Edwards CJ, Jafari H, Emsley R, Huckvale M, Rus-Calafell M, et al. Digital AVATAR therapy for distressing voices in psychosis: the phase 2/3 AVATAR2 trial. Nat Med 2024;30(12):3658-3668. [doi: 10.1038/s41591-024-03252-8] [Medline: 39468363]
51. Dodgson G, Singh A, Barclay N, Birkett L, Boyle C, Brandon T, et al. Managing unusual sensory experiences in at-risk mental state for psychosis in England: a parallel group, single-blind, randomised controlled feasibility trial. Psychiatry Res 2025;351:116564 [FREE Full text] [doi: 10.1016/j.psychres.2025.116564] [Medline: 40494058]
52. Dudley R, Dodgson G, Common S, Ogundimu E, Liley J, O'Grady L, et al. Effects of a novel, brief psychological therapy (managing unusual sensory experiences) for hallucinations in first episode psychosis (MUSE FEP): findings from an exploratory randomised controlled trial. J Psychiatr Res 2024;174:289-296 [FREE Full text] [doi: 10.1016/j.jpsychires.2024.04.031] [Medline: 38678686]
53. Eisner E, Berry N, Morris R, Emsley R, Haddock G, Machin M, et al. Exploring engagement with the CBT-informed Actissist smartphone application for early psychosis. J Ment Health 2023;32(3):643-654. [doi: 10.1080/09638237.2023.2182429] [Medline: 36850040]
54. Hassan L, Eisner E, Berry K, Emsley R, Ainsworth J, Lewis S, et al. User engagement in a randomised controlled trial for a digital health intervention for early psychosis (Actissist 2.0 trial). Psychiatry Res 2023;329:115536 [FREE Full text] [doi: 10.1016/j.psychres.2023.115536] [Medline: 37857132]
55. Ward T, Hardy A, Holm R, Collett N, Rus-Calafell M, Sacadura C, et al. SlowMo therapy, a new digital blended therapy for fear of harm from others: an account of therapy personalisation within a targeted intervention. Psychol Psychother 2022;95(2):423-446 [FREE Full text] [doi: 10.1111/papt.12377] [Medline: 35019210]
56. Rus-Calafell M, Ward T, Zhang XC, Edwards CJ, Garety P, Craig T. The role of sense of voice presence and anxiety reduction in AVATAR therapy. J Clin Med 2020;9(9):2748 [FREE Full text] [doi: 10.3390/jcm9092748] [Medline: 32854387]
57. Brown P, Waite F, Lambe S, Jones J, Jenner L, Diamond R, et al. Automated virtual reality cognitive therapy (GameChange) in inpatient psychiatric wards: qualitative study of staff and patient views using an implementation framework. JMIR Form Res 2022;6(4):e34225 [FREE Full text] [doi: 10.2196/34225] [Medline: 35412462]
58. Bond J, Kenny A, Pinfold V, Couperthwaite L, GameChange Lived Experience Advisory Panel, Kabir T, et al. A safe place to learn: peer research qualitative investigation of GameChange virtual reality therapy. JMIR Serious Games 2023;11:e38065 [FREE Full text] [doi: 10.2196/38065] [Medline: 36645707]
59. Leff J, Williams G, Huckvale M, Arbuthnot M, Leff AP. Avatar therapy for persecutory auditory hallucinations: what is it and how does it work? Psychosis 2014;6(2):166-176 [FREE Full text] [doi: 10.1080/17522439.2013.773457] [Medline: 24999369]
60. Hardy A, Wojdecka A, West J, Matthews E, Golby C, Ward T, et al. How inclusive, user-centered design research can improve psychological therapies for psychosis: development of SlowMo. JMIR Ment Health 2018;5(4):e11222 [FREE Full text] [doi: 10.2196/11222] [Medline: 30518514]
61. Bucci S, Zhang X, Di Basilio D, Richardson C, Berry N, Berry K, et al. "It's like having a friend in your pocket." The service user experience of the Actissist digital health intervention for early psychosis: a qualitative study. BMC Psychiatry 2025;25(1):682 [FREE Full text] [doi: 10.1186/s12888-025-07071-0] [Medline: 40624605]
62. Greenwood KE, Gurnani M, Ward T, Vogel E, Vella C, McGourty A, SlowMo Patient, Public Involvement (PPI) team. The service user experience of SlowMo therapy: a co-produced thematic analysis of service users' subjective experience. Psychol Psychother 2022;95(3):680-700 [FREE Full text] [doi: 10.1111/papt.12393] [Medline: 35445520]
63. Hardy A, Ward T, Emsley R, Greenwood K, Freeman D, Fowler D, et al. Bridging the digital divide in psychological therapies: observational study of engagement with the SlowMo mobile app for paranoia in psychosis. JMIR Hum Factors 2022;9(3):e29725 [FREE Full text] [doi: 10.2196/29725] [Medline: 35776506]
64. Leung CJ, Fosuaah A, Frerichs J, Heslin M, Kabir T, Lee TMC, et al. A qualitative study of the acceptability of cognitive bias modification for paranoia (CBM-pa) in patients with psychosis. BMC Psychiatry 2019;19(1):225 [FREE Full text] [doi: 10.1186/s12888-019-2215-3] [Medline: 31337373]

XSL•FO
RenderX

65. Lambe S, Knight I, Kabir T, West J, Patel R, Lister R, et al. Developing an automated VR cognitive treatment for psychosis: GameChange VR therapy. J Behav Cogn Ther 2020;30(1):33-40. [doi: 10.1016/j.jbct.2019.12.001]
66. Berry N, Machin M, Ainsworth J, Berry K, Edge D, Haddock G, et al. Developing a theory-informed smartphone app for early psychosis: learning points from a multidisciplinary collaboration. Front Psychiatry 2020;11:602861 [FREE Full text] [doi: 10.3389/fpsyt.2020.602861] [Medline: 33362612]
67. Greenwood K, Robertson S, Vogel E, Vella C, Ward T, McGourty A, SlowMo Patient Public Involvement (PPI) team. The impact of patient and public involvement in the SlowMo study: reflections on peer innovation. Health Expect 2022;25(1):191-202 [FREE Full text] [doi: 10.1111/hex.13362] [Medline: 34585482]
68. Knight I, West J, Matthews E, Kabir T, Lambe S, Waite F, et al. Participatory design to create a VR therapy for psychosis. Des Health 2021;5(1):98-119. [doi: 10.1080/24735132.2021.1885889]
69. Owrid O, Richardson L, Allan S, Grant A, Gogan S, Hamilton N, et al. "There's no us vs. them, it's just us": a creative approach to centring lived experience within the AVATAR2 trial. BMC Psychiatry 2024;24(1):807 [FREE Full text] [doi: 10.1186/s12888-024-06268-z] [Medline: 39543539]
70. Hardy A, Taylor KM, Grant A, Christie L, Walsh L, Gant T, et al. Co-designing technology to improve psychological therapy for psychosis: SloMo, a blended digital therapy for fear of harm from others. Schizophr Res 2024;274:526-534 [FREE Full text] [doi: 10.1016/j.schres.2024.11.004] [Medline: 39579660]
71. Dudley R, O'Grady L, Cioroboiu N, Bates E, Gibbs C, Dodgson G, et al. Digitally delivered treatment for unusual sensory experiences for people with psychosis: a real-world service evaluation study. Behav Cogn Psychother 2025;53(3):253-263. [doi: 10.1017/S1352465825100969] [Medline: 40926495]
72. O'Brien C, Rus-Calafell M, Craig TK, Garety P, Ward T, Lister R, et al. Relating behaviours and therapeutic actions during AVATAR therapy dialogue: an observational study. Br J Clin Psychol 2021;60(4):443-462. [doi: 10.1111/bjc.12296] [Medline: 33949726]
73. Rus-Calafell M, Ehrbar N, Ward T, Edwards C, Huckvale M, Walke J, et al. Participants' experiences of AVATAR therapy for distressing voices: a thematic qualitative evaluation. BMC Psychiatry 2022;22(1):356 [FREE Full text] [doi: 10.1186/s12888-022-04010-1] [Medline: 35610590]
74. Deamer F, Richardson E, Ward T, Garety P, Craig TKJ, Rus-Calafell M, et al. Hypothetical active voicing in AVATAR therapy. J Appl Linguist Prof Pract 2021;18(1):34-65. [doi: 10.1558/jalpp.24121]
75. Ward T, Rus-Calafell M, Ramadhan Z, Soumelidou O, Fornells-Ambrojo M, Garety P, et al. AVATAR therapy for distressing voices: a comprehensive account of therapeutic targets. Schizophr Bull 2020;46(5):1038-1044 [FREE Full text] [doi: 10.1093/schbul/sbaa061] [Medline: 32372082]
76. Freeman D, Lambe S, Galal U, Yu L, Kabir T, Petit A, et al. Agoraphobic avoidance in patients with psychosis: severity and response to automated VR therapy in a secondary analysis of a randomised controlled clinical trial. Schizophr Res 2022;250:50-59 [FREE Full text] [doi: 10.1016/j.schres.2022.10.008] [Medline: 36343472]
77. Freeman D, Rosebrock L, Waite F, Loe BS, Kabir T, Petit A, et al. Virtual reality (VR) therapy for patients with psychosis: satisfaction and side effects. Psychol Med 2023;53(10):4373-4384 [FREE Full text] [doi: 10.1017/S0033291722001167] [Medline: 35477837]
78. Bucci S, Berry N, Morris R, Berry K, Haddock G, Lewis S, et al. "They Are Not Hard-to-Reach Clients. We Have Just Got Hard-to-Reach Services." Staff views of digital health tools in specialist mental health services. Front Psychiatry 2019;10:344 [FREE Full text] [doi: 10.3389/fpsyt.2019.00344] [Medline: 31133906]
79. Allan S, Bradstreet S, McLeod H, Farhall J, Lambrou M, Gleeson J, EMPOWER Group, et al. Developing a hypothetical implementation framework of expectations for monitoring early signs of psychosis relapse using a mobile app: qualitative study. J Med Internet Res 2019;21(10):e14366 [FREE Full text] [doi: 10.2196/14366] [Medline: 31651400]
80. Hall J, Rus-Calafell M, Omari-Asor L, Ward T, Emsley R, Garety P, et al. Assessing the subjective experience of participating in a clinical trial (AVATAR). Psychiatry Res 2018;263:82-87 [FREE Full text] [doi: 10.1016/j.psychres.2018.02.026] [Medline: 29502043]
81. Lipschitz JM, Pike CK, Hogan TP, Murphy SA, Burdick KE. The engagement problem: a review of engagement with digital mental health interventions and recommendations for a path forward. Curr Treat Options Psychiatry 2023;10(3):119-135 [FREE Full text] [doi: 10.1007/s40501-023-00297-3] [Medline: 38390026]
82. Killikelly C, He Z, Reeder C, Wykes T. Improving adherence to web-based and mobile technologies for people with psychosis: systematic review of new potential predictors of adherence. JMIR Mhealth Uhealth 2017;5(7):e94 [FREE Full text] [doi: 10.2196/mhealth.7088] [Medline: 28729235]
83. Fahy L, Lee L, Newlove L, Wood L, O'Driscoll C. Barriers and facilitators to engagement in psychological therapy in first episode psychosis: a meta-ethnography and qualitative comparative analysis. Psychol Psychother 2025;98(2):232-255. [doi: 10.1111/papt.12576] [Medline: 39898491]
84. Ferrari M, MacDonald K, Sabetti J, Cowan T, Iyer SN. Developing a spectrum model of engagement in services for first episode psychosis: beyond attendance. Front Psychiatry 2024;15:1429135 [FREE Full text] [doi: 10.3389/fpsyt.2024.1429135] [Medline: 39421063]

XSL•FO
RenderX

85. Smith KA, Hardy A, Vinnikova A, Blease C, Milligan L, Hidalgo-Mazzei D, et al. Digital mental health for schizophrenia and other severe mental illnesses: an international consensus on current challenges and potential solutions. JMIR Ment Health 2024;11:e57155 [FREE Full text] [doi: 10.2196/57155] [Medline: 38717799]
86. Bradley G, Rehackova L, Devereaux K, Bruce TA, Nunn V, Gilfellon L, et al. Classifying the features of digital mental health interventions to inform the development of a patient decision aid. PLOS Digit Health 2025;4(3):e0000752 [FREE Full text] [doi: 10.1371/journal.pdig.0000752] [Medline: 40138606]
87. Strudwick G, Kassam I, Torous J, Patenaude S. Building the infrastructure for sustainable digital mental health: it is "Prime Time" for implementation science. JMIR Ment Health 2025;12:e78791 [FREE Full text] [doi: 10.2196/78791] [Medline: 40690592]
88. Molfenter T, Heitkamp T, Murphy AA, Tapscott S, Behlman S, Cody OJ. Use of telehealth in mental health (MH) services during and after COVID-19. Community Ment Health J 2021;57(7):1244-1251 [FREE Full text] [doi: 10.1007/s10597-021-00861-2] [Medline: 34165695]
89. Cuijpers P, Reijnders M, Huibers MJH. The role of common factors in psychotherapy outcomes. Annu Rev Clin Psychol 2019;15:207-231. [doi: 10.1146/annurev-clinpsy-050718-095424] [Medline: 30550721]
90. Faith LA, Hillis-Mascia JD, Wiesepape CN. How does individual psychotherapy promote recovery for persons with psychosis? A systematic review of qualitative studies to understand the patient's experience. Behav Sci (Basel) 2024;14(6):460 [FREE Full text] [doi: 10.3390/bs14060460] [Medline: 38920792]
91. Berry N, Lobban F, Emsley R, Bucci S. Acceptability of interventions delivered online and through mobile phones for people who experience severe mental health problems: a systematic review. J Med Internet Res 2016;18(5):e121 [FREE Full text] [doi: 10.2196/jmir.5250] [Medline: 27245693]
92. Ma J, Lander N, Eyre ELJ, Barnett LM, Essiet IA, Duncan MJ. It's not just what you do but the way you do it: a systematic review of process evaluation of interventions to improve gross motor competence. Sports Med 2021;51(12):2547-2569. [doi: 10.1007/s40279-021-01519-5] [Medline: 34339044]
93. Ball H, Eisner E, Nicholas J, Wilson P, Bucci S. How theories, models, and frameworks have been used to implement digital health interventions in services for people with severe mental health problems: a scoping review. BMC Public Health 2025;25(1):1023 [FREE Full text] [doi: 10.1186/s12889-025-22189-2] [Medline: 40098003]
94. Chen JI, Roth B, Dobscha SK, Lowery JC. Implementation strategies in suicide prevention: a scoping review. Implement Sci 2024;19(1):20 [FREE Full text] [doi: 10.1186/s13012-024-01350-2] [Medline: 38409000]

Abbreviations

DMHI: digital mental health intervention
ERIC: Expert Recommendations for Implementing Change
MRC: Medical Research Council
NHS: National Health Service
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
VR: virtual reality

Edited by A Kushniruk; submitted 18.Nov.2025; peer-reviewed by L Martinengo; comments to author 04.May.2026; revised version received 02.Jun.2026; accepted 23.Jun.2026; published 23.Jul.2026.

Please cite as:
Hampshire C, Dack C, Daryan S, Fialho C, Taher R, Teale AL, Yiend J, Jacobsen P
Process Evaluation of Digital Mental Health Interventions for Psychosis: Scoping Review With Framework Synthesis
JMIR Hum Factors 2026;13:e88032
URL: https://humanfactors.jmir.org/2026/1/e88032
doi:10.2196/88032
PMID:

©Chloe Hampshire, Charlotte Dack, Shadi Daryan, Carolina Fialho, Rayan Taher, Ashley-Louise Teale, Jenny Yiend, Pamela Jacobsen. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 23.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR

XSL•FO
RenderX

Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Usability and User Experience Assessment of Health Care Conversational Agents Using Validated Subjective Instruments: Systematic Review and Comparative Analysis

João Pavão[1], Prof Dr; Rute Bastardo[2], Prof Dr; Anabela Gonçalves Silva[3], Prof Dr; Nelson Pacheco Rocha[4], Prof Dr

[1]Institute for Systems and Computer Engineering, Technology and Science—INESC-TEC, Science and Technology School, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal

[2]UNIDCOM, Science and Technology School, University of Trás-os-Montes and Alto Douro, Vila Real, Portugal

[3]Center for Health Technology and Services Research, School of Health Sciences, University of Aveiro, Aveiro, Portugal

[4]Institute of Electronics and Informatics Engineering of Aveiro, Department of Medical Sciences, University of Aveiro, Aveiro, Portugal

Corresponding Author:
Nelson Pacheco Rocha, Prof Dr
Institute of Electronics and Informatics Engineering of Aveiro
Department of Medical Sciences
University of Aveiro
Campus Universitário de Santiago
Aveiro, 3810-193
Portugal
Phone: 351 234 370 200
Email: npr@ua.pt

Abstract

Background: Digital applications based on health care conversational agents (HCAs) are increasingly being developed to support health care provision. The usability and user experience of these solutions are critical determinants of their acceptability and, consequently, their impact on health-related outcomes.

Objective: This systematic review aims to synthesize current evidence on the use of valid and reliable subjective instruments for assessing the usability and user experience of HCAs and to examine whether assessment outcomes vary according to their technical characteristics.

Methods: A systematic search was conducted in PubMed, Web of Science, and Scopus from inception to February 2026. Studies were included if they used subjective instruments to evaluate the usability or user experience of HCAs.

Results: A total of 127 studies met the inclusion criteria. The studies examined 3 categories of HCAs—text-based, voice-based, and embodied—applied to patient care, health education and prevention, health data collection, and support for daily activities among older adults. The System Usability Scale (SUS) was the most frequently used assessment instrument. Comparative analysis of SUS scores indicated higher usability ratings for text-based HCAs relative to voice-based and embodied systems. However, SUS and other subjective instruments used in the included studies may not fully capture key dimensions of usability and user experience of HCAs. Additionally, substantial heterogeneity was observed in assessment methodologies across studies.

Conclusions: Comparative analysis suggested that text-based HCAs were associated with significantly higher SUS scores than voice-based and embodied HCAs. However, this finding should be interpreted with caution given the substantial heterogeneity across the included studies in health care application domains, study designs, evaluation contexts, participant populations, and HCAs' implementation and use characteristics, as well as the limitations of the SUS in evaluating the usability of modern HCAs. The variability in assessment approaches underscores the need for standardized protocols and the development of more context-specific evaluation frameworks to enhance methodological consistency and comparability across studies.

(JMIR Hum Factors 2026;13:e102911) doi:10.2196/102911

KEYWORDS

digital health; health care provision; usability; user experience; usability assessment; user experience assessment; conversational agents; older adults

XSL•FO
RenderX

Introduction

Since the development of ELIZA in 1966 [1], a rule-based system developed to emulate text-based interactions with a psychotherapist, advancements in information technology have driven the widespread adoption of health care conversational agents (HCAs) [2,3], which are applications designed to support interactive, multiturn communication through written or spoken language [4]. Consumer solutions such as Siri, Google Now, Cortana, and Alexa are well-known examples of agents designed to help users find information and accomplish tasks [5].

Modern HCAs extend well beyond simple rule-based input-output interactions. Advances in large language models have facilitated their evolution from narrowly defined, task-oriented tools into open-domain systems capable of retaining conversational context, following topics across multiple turns, and sustaining coherent, continuous dialogue for more natural interactions [6,7]. In addition, AI techniques have enabled more advanced dialogue management, facilitating the exhibition of human-like behaviors, the recognition of users' emotions and intentions, and the use of various information sources for reasoning and decision-making [8-10].

Building on these advancements, researchers have explored anthropomorphized interfaces, leading to the development of embodied agents, or virtual humans [11,12]. Unlike traditional agents, which rely solely on text or voice, embodied agents combine verbal communication with nonverbal cues (eg, posture, gestures, eye movements, or facial expressions) to foster more lifelike interactions [13-15]. This embodiment enhances users' perception of social presence and engagement, making interactions feel more intuitive and emotionally rich [16,17]. Moreover, the introduction of virtual reality technologies has paved the way for increasingly sophisticated agents designed to promote user engagement through immersive experiences [18].

HCAs are widely used in health care [19], reflecting the widespread adoption of similar agents across other human-centered sectors, including customer service [20], entertainment [21], and education [22].

From a health care research perspective, the assessment of HCAs should consider health-related outcomes, such as symptom reduction, disease knowledge, disease self-management, and quality of life [8]. This assessment should be conducted in the context of continuous use of the applications and preferably supported by randomized controlled trials. However, before conducting randomized controlled trials, it is important to safeguard aspects that might affect health-related outcomes. This means that preliminary assessments (ie, feasibility studies) should be conducted [23] to evaluate the technical characteristics of HCAs or their quality in terms of usability and user experience and, consequently, their potential to promote user engagement among the target users [8,24].

System usability comprises several components (ie, learnability, efficiency, memorability, errors, and satisfaction) and relates to the ease of using the system interface [25]. The rise of new technologies, devices, and interaction methods has expanded the concept of usability into the broader notion of user experience, which encompasses the cognitive, affective, social, and physical aspects of interaction [26-28].

There are 3 main methods for evaluating usability and user experience, namely, inspection, testing, and inquiry [29]. The inspection method consists of a systematic analysis of interaction mechanisms carried out by experts according to specific guidelines, while the testing and inquiry methods are based on the collection of usage data [29]. The testing method involves observing users while they perform tasks with a given system to collect primarily quantitative data (eg, the number of completed tasks) that can provide empirical evidence on how to improve the interaction mechanisms [29]. By contrast, the inquiry method involves collecting qualitative data from users, namely through interviews or subjective measurement instruments such as questionnaires or scales [29]. Examples of commonly used instruments include the Post-Study System Usability Questionnaire (PSSUQ) [30] and the System Usability Scale (SUS) [31]. These instruments have had their measurement properties assessed, particularly validity (ie, how effectively usability or user experience is measured) and reliability (ie, the consistency of measurements when repeated under comparable conditions) [30-33].

In this context, this review aims to synthesize the current evidence on usability and user experience assessments of HCAs using valid and reliable subjective instruments, to identify the purposes and categories of HCAs being assessed, and to determine, through a comparative analysis, how this typification affects their usability and user experience.

Several reviews [34-41] have already focused on different aspects of HCAs (eg, the use of AI or large language models), including usability and user experience assessments (eg, [42-45]); however, these reviews did not perform comparative analyses of usability and user experience across different categories of HCAs. Such analyses are important because different characteristics of HCAs affect how users interact with them (eg, by shaping communication style, pacing, and information processing), as well as cognitive load and perceptions of trust and privacy. Therefore, this review may provide insights to inform better design decisions and the development of standardized evaluation frameworks, enabling researchers and practitioners to benchmark usability and user experience across different HCAs.

Methods

Study Protocol Registration and Scope

The research protocol for this systematic review was registered in the Open Science Framework and followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; see Multimedia Appendix 1) guidelines [46]. The research scope (Textbox 1) was structured using the Population, Concept, and Context framework [47].

XSL•FO
RenderX

Textbox 1. Research scope.

1. Population

Adults (≥18 years) managing their health conditions using digital health applications.

2. Concept

Usability and user experience assessments of health care conversational agents designed for interactive, multiturn communication through written or spoken language to support the management of users' health conditions, supported by validated subjective instruments (ie, questionnaires or scales used to systematically collect users' self-reported evaluations of usability or user experience).

3. Context

Community or institutional care settings.

Literature Search

Scopus, Web of Science, and PubMed were selected as the online databases for retrieving studies included in this review. The search query included keywords representing the following 3 main components:

- HCAs: conversational agent, relational agent, virtual human, virtual agent, virtual assistant, virtual companion, virtual coach, chatbot, enhanced communication agent, embodied communication agent, ECA, and avatar.
- Health care: health care, health care, care, patient, digital health, electronic health, e-health, eHealth, older adults, and elderly.
- Usability or user experience assessment: usability, user experience, acceptance, evaluation, assessment, and test.

The keywords within each component were linked using the Boolean operator OR, whereas the expressions derived from different components were linked using the Boolean operator AND. As an example, Figure 1 presents the search query prepared for the Scopus database.

Figure 1. Search query for the Scopus database.

```
TITLE-ABS-KEY(("conversational agent" OR "relational agent" OR "virtual human" OR "virtual agent" OR "virtual assistant" OR "virtual companion" OR "virtual coach" OR chatbot OR "enhanced communication agent" OR "embodied communication agent" OR ECA OR avatar) AND (patient OR "digital health" OR "electronic health" OR "e-health" OR ehealth OR healthcare OR "health care" OR care OR "older adults" OR elderly) AND (usability OR "user experience" OR acceptance OR evaluation OR assessment OR test)).
```

Inclusion and Exclusion Criteria

To be included in the review, studies had to meet the eligibility criteria presented in Textbox 2.

Textbox 2. Inclusion and exclusion criteria.

1. Inclusion criteria

- Empirical studies published in peer-reviewed scientific journals or conference proceedings.
- Adults (≥18 years).
- Studies evaluating health care conversational agents (HCAs) to support the management of users' health conditions.
- Studies performing usability or user experience assessments of HCAs based on validated subjective instruments.

2. Exclusion criteria

- Review articles, opinion pieces, letters, and editorials.
- Children or adolescents.
- Studies evaluating the integration of HCAs into clinician-oriented digital health applications or applications focused on domains other than health care.
- Studies assessing usability or user experience without using validated subjective instruments or not focusing on this type of assessment.
- Preliminary versions of studies for which a more mature version was included in the review (eg, when a study was published as both a conference paper and later a journal article, the conference version was excluded).
- Studies published in languages other than English.
- Studies for which the full text was not available.

XSL•FO
RenderX

Identification Process

All references retrieved from the databases were imported into a Microsoft Excel spreadsheet. The selection process then followed the general guidelines of the PRISMA flowchart [46]: (1) exclusion of duplicates, (2) title and abstract screening, and (3) full-text screening. Additionally, a backward search was performed based on the reference lists of the articles included for full-text screening.

All stages were performed manually by 2 reviewers (JP and NPR) without the use of automation tools. Disagreements between the reviewers were discussed and resolved collaboratively.

Quality Assessment

The methodological quality assessment of the reviewed studies was conducted using a subset of the Critical Assessment of Usability Studies Scale (CAUSS) [48]. As this scale evaluates both testing and inquiry methods, the reviewers (AGS and NPR) selected 5 items related to assessments using inquiry methods: (1) representativeness of the participants, (2) proximity to the real-world context of use, (3) adequacy of the number of participants, (4) representativeness of the functionalities assessed, and (5) continuous and prolonged use of the app.

The excluded items comprised (1) items related to the validation of measurement instruments, as the eligibility criteria for this review explicitly excluded studies that did not use validated instruments; (2) items related to testing methods involving the direct observation of participants while they performed tasks to collect objective measures, such as error rates or mean task completion time; and (3) the item related to the triangulation of testing and inquiry methods, as testing methods were not considered in this review.

As only a subset of the CAUSS was applied, the resulting scores should not be interpreted as reflecting the overall quality of the usability studies, but rather the quality of the use of subjective measurement instruments.

Data Extraction

The following data were extracted:

- Demographic characteristics, namely, the author(s), nationality of the authors' affiliations, type of publication, and year of publication.
- Purposes of the assessed HCAs.
- Technical characteristics of the HCAs and supporting technologies (eg, AI or virtual reality).
- Types of experimental study designs (eg, cross-sectional or longitudinal studies).
- Duration of the longitudinal studies.
- Number and characteristics of the study participants.
- Recruitment and distribution of the participants.
- Validated subjective instruments assessing usability or user experience.
- Results of the use of the validated subjective assessment instruments (ie, scores and SDs).

Data were independently extracted from all included studies by 2 reviewers (JP and RB) using a customized Microsoft Excel form. Discrepancies were resolved through discussion and consensus.

Data Synthesis

Results were synthesized into summary tables, figures, and descriptive text according to the following domains: (1) demographic characteristics (ie, publication types and publication years); (2) purposes of HCAs; (3) technical characteristics of HCAs; (4) study experimental characteristics (ie, experimental design, number and characteristics of the participants, and measurement instruments); and (5) results of usability and user experience assessments.

A quantitative analysis was performed only for studies using the SUS to assess usability. The number of studies using other instruments was insufficient for quantitative analysis. Studies using the SUS were subdivided based on the categories of HCAs being assessed. To compare SUS scores and the number of participants across the categories of HCAs, the Kruskal-Wallis test was used because the data did not follow a normal distribution (Kolmogorov-Smirnov test, $P<.05$). Medians and IQRs were provided for each group. Additionally, a subgroup analysis comparing SUS scores between studies involving older adults and those involving adults was performed for all studies and within each category of HCAs using the Mann-Whitney U test. The Bonferroni correction for multiple tests was applied. The significance level was set at $P<.05$. Analyses were performed using SPSS.

In addition, the effect size r was manually calculated as the standardized test statistic from the Mann-Whitney U test or the Kruskal-Wallis test (z) divided by the square root of the total sample size (N) and interpreted as small (≥0.1), medium (≥0.3), or large (≥0.5) [49]. For the Kruskal-Wallis test, the ordinal eta-squared (η^2H) was calculated as $\eta^2H=(H - K + 1)/(N - K)$, where H is the test statistic, N is the sample size, and K is the number of groups. It was interpreted as small ($\eta^2H \geq 0.01$), medium ($\eta^2H \geq 0.06$), or large ($\eta^2H \geq 0.14$) [49].

Given the substantial heterogeneity among the included studies—including differences in health care application domains, study designs, evaluation contexts, participant populations, and HCAs' implementation and use characteristics—the quantitative comparison of SUS scores was intended as an exploratory synthesis of patterns across studies rather than as evidence of causal differences in usability among HCA categories. Moreover, the SUS may not adequately capture all relevant aspects of the usability of the evaluated HCAs, such as conversational quality. Consequently, the comparative analyses should be interpreted as identifying associations at the study level rather than establishing the intrinsic usability of specific HCA categories.

Results

Identification of Studies

The online search was performed in February 2026, and 8639 references were retrieved. As presented in Figure 2, after the exclusion of 3391 duplicates, the titles and abstracts of 5248 references were analyzed, and 5094 were excluded according

XSL•FO
RenderX

to the eligibility criteria. Therefore, 154 references were included for full-text analysis. During the full-text analysis, 30 references [50-79] were excluded because they did not meet the eligibility criteria (eg, studies that did not apply valid and reliable instruments to measure usability or user experience, or studies focused on the assessment of clinician-oriented applications). Additionally, the backward search identified 3 references, resulting in 127 references [80-206] being included in the review.

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart.

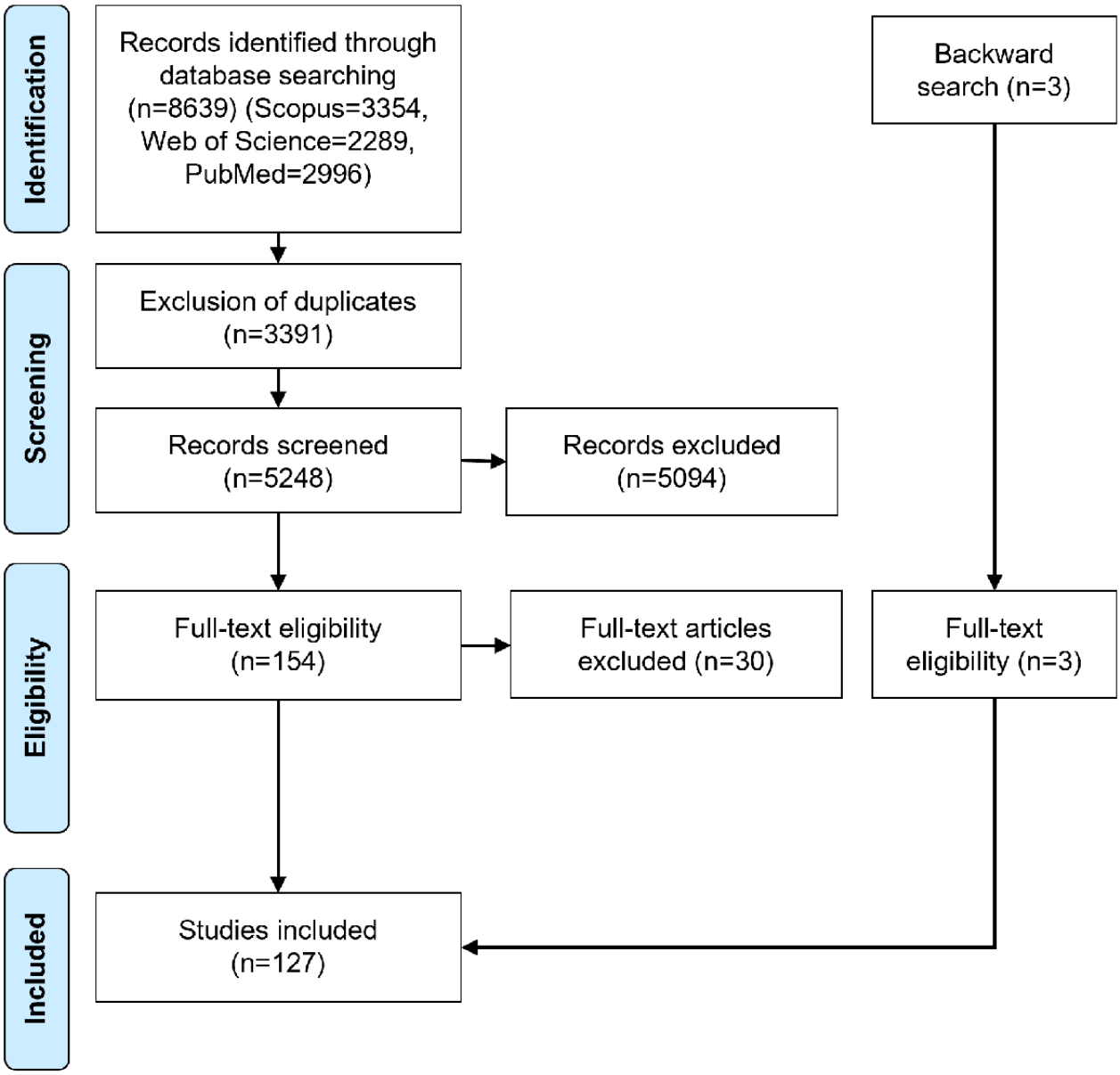
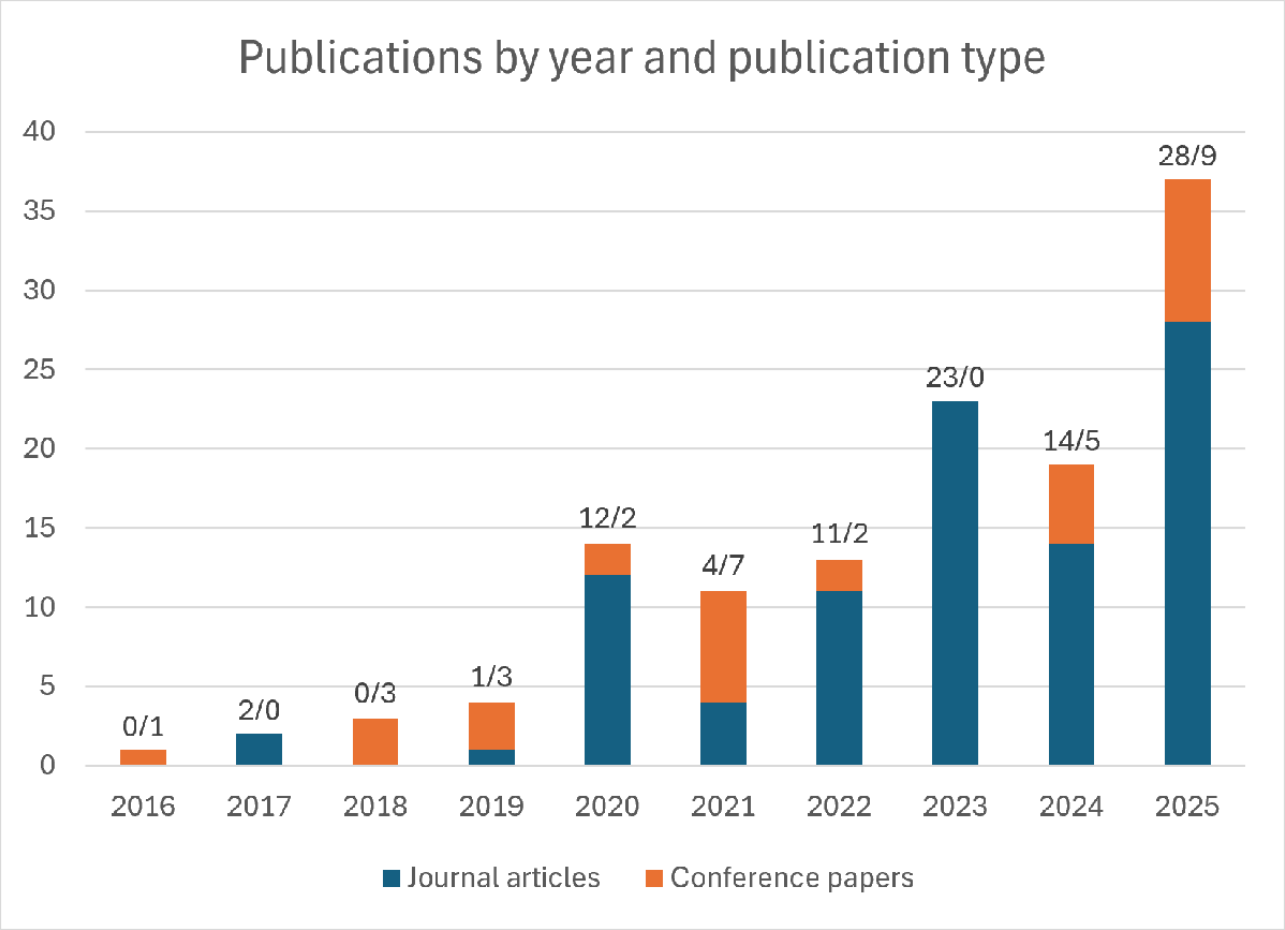


Figure 3 presents the distribution of publication years by publication type (ie, journal articles and conference proceedings).

XSL•FO
RenderX

Figure 3. Publications by year and type.



Quality Assessment Results

Most CAUSS criteria considered for the methodological quality assessment were met by a majority of studies: (1) representativeness of the participants (113/127, 89%); (2) proximity to the real-world context of use (105/127, 82.7%); (3) adequacy of the number of participants (90/127, 70.9%); and (4) representativeness of the functionalities assessed (96/127, 75.6%). The exception was the criterion related to the continuous use of the application, which was met by only 44/127 (34.6%) studies. In turn, 28/127 (22%) studies met all 5 criteria.

Purposes of the HCAs

In terms of the purposes of the HCAs described in the included articles, 4 major categories were identified:

- Patient care: 73 studies [81,82,85,87,89,92,95-97,104,106,107,111,113-115,119,121,123,124,127,128,130,132,133,135,136,138,140,141,144-151,153-157,159,162,165-170,172-175,177,179-184,187-189,191,194,197,198,203-206].
- Health education and prevention: 33 studies [86,90,93,103,109,110,112,116-118,122,125,129,131,134,137,142,143,158,160,163,164,171,176,185,192,193,195,196,199-202].
- Health data collection: 15 studies [88,91,98,99,101,102,105,108,120,126,139,152,161,178,186].
- Daily activities of older adults: 6 studies [80,83,84,94,100,190].

Concerning patient care, almost half of the respective articles (ie, 34 [81,82,85,92,95,97,103,104,119,121,123,124,132,147,149,150,153,155,165-170,172,173,179,181,184,188,191,204-206]) targeted mental health disorders (eg, depression, anxiety, stress, posttraumatic stress disorder, eating disorders, and suicidal risk), while the remaining 39 [87,89,96,106,107,111,114,115,127,128,130,133,135,136,138,140,141,144-146,148,151,154,156,157,159,162,174,175,177,180,182,183,187,189,194,197,198,203] articles targeted a range of other conditions, including cancer [138,145,182,183,197], chronic kidney disease [144], chronic pain [115,162,198], COVID-19 [135], delirium [146], diabetes [89,96,111,115,140,194], hemophilia [136], heart-related diseases [128,130,133,154,159,174,187,203], hydrocephalus [177], intestinal ostomy [148], neurodegenerative diseases [107,141,157,175], psoriasis [127], sickle cell disease [114], traumatic injuries [106,151,180], and vestibular disorders [189].

Looking at health education and prevention, beginning with health education, some articles [117,131,137,158,160,171,176,185,193,195,196,202] focused on general health education topics, while others focused on specific topics, such as (1) combating misinformation about COVID-19 [109,113,134], (2) endoscopy explanation [164], (3) hematologic malignancies [110], (4) hereditary breast and ovarian cancer [201], (5) perinatal information for parents [143], and (6) vaccine pharmacovigilance [142]. In the studies that focused on health prevention, the assessed HCAs were designed to promote behavior change (eg, physical activity and weight maintenance) [86,93,112,116,118,125,129,199,200], support vaccination (ie, COVID-19 [122] and human papillomavirus [192]), and provide balance training [90,163].

XSL•FO
RenderX

In terms of health data collection, 11 [88,98,101,105,108,120,126,139,152,178,186] studies assessed HCAs for collecting patient-reported outcome measures, 3 [91,99,161] focused on collecting family health history, and 1 [102] described an application designed to collect data for the evaluation of health care services and quality of care.

Finally, 6 [80,83,84,94,100,190] studies assessed HCAs designed to promote the independence and autonomy of older adults by assisting them in managing activities of daily living, including problem-solving [100] and cooking [190].

Technical Characteristics of HCAs

Three categories of HCAs were identified: (1) text-based HCAs, (2) voice-based HCAs, and (3) embodied HCAs.

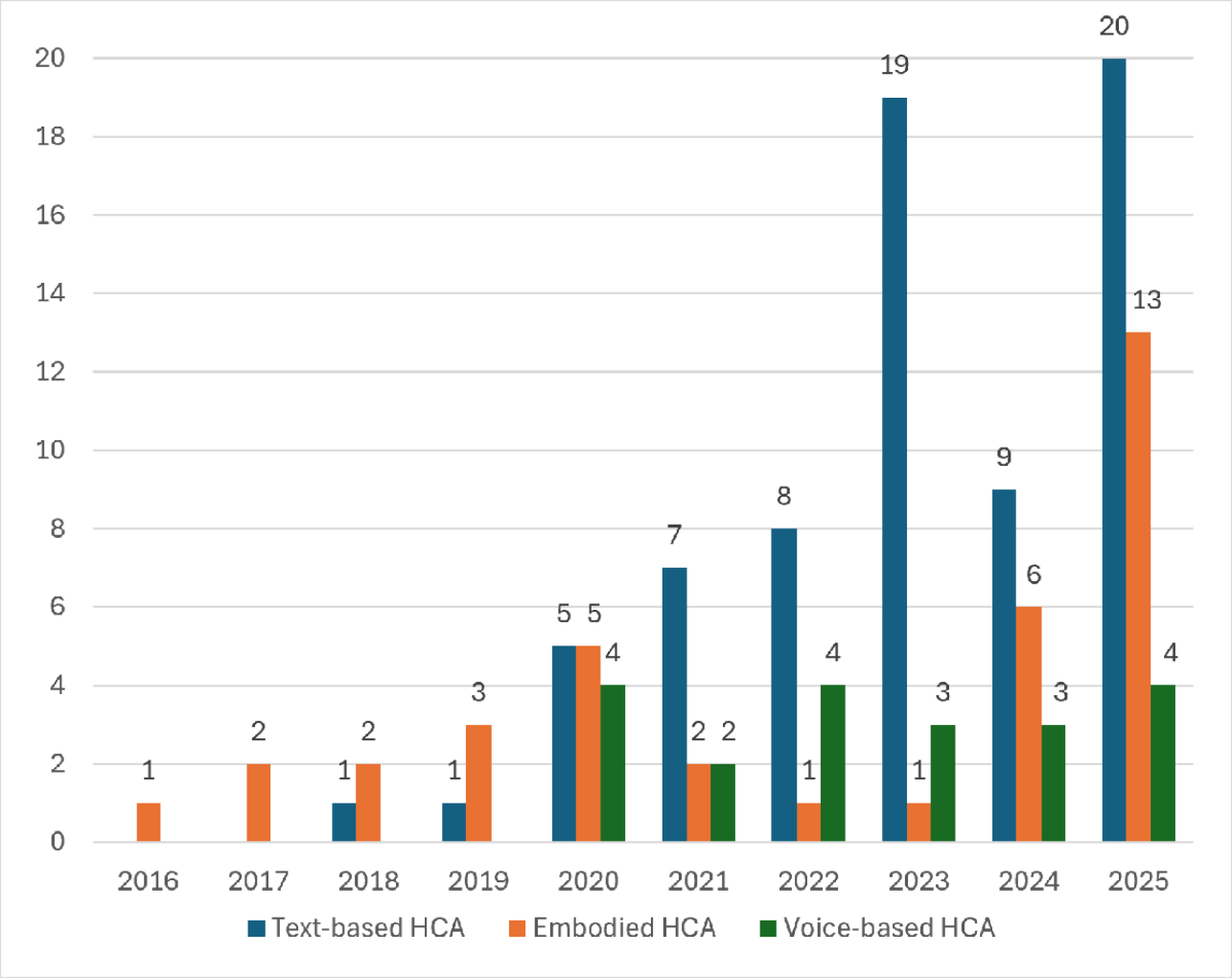
Five studies [87,98,123,153,169] assessed 2 categories of HCAs: (1) voice-based and embodied [87], (2) text- and voice-based [98,123], and (3) text-based and embodied [153,169]. The remaining studies focused on a single category, according to the following distribution:

- Text-based HCAs: 69 studies [85,86,91,99,100,102,103,105,106,109,110,112-117,119,120,122,125,126,129-134,137-145,147-150,154,157-159,161,164,166,168,170,177,184-187,191,194-206].
- Voice-based HCAs: 18 studies [93,97,101,104,111,118,124,127,135,136,146,155,165,167,172,183,190,192].
- Embodied HCAs: 35 studies [80-84,88-90,92,94-96,107,108,121,128,151,152,156,160,162,163,171,173-176,178-182,188,189,193].

Some studies [91,99,100,102,105,108,109,112,124,126,137,138,147,152,155,161,163,173,178], although maintaining the HCA categories, established different comparisons, namely, (1) HCAs versus regular computer-based questionnaires [91,99,102,105,126,161]; (2) similar applications with different functionalities [108,109,112,137,152,155,173,178] or different intervention strategies [100,124,147,163]; and (3) HCAs versus usual care [138].

Figure 4 illustrates the distribution of studies across publication years by HCA category. According to the figure, the number of studies assessing voice-based HCAs has remained approximately constant, whereas the number of studies assessing text-based and embodied HCAs has increased over the past 3 years. Table 1 presents the relationship between HCA categories and their purposes.

Figure 4. Distribution of studies across publication years, categorized by types of health care conversational agents (HCAs).



XSL•FO
RenderX

Table 1. Relation of the categories of HCAsa and the purposes of the applications.

Purposes	Text-based HCAs	Voice-based HCAs	Embodied HCAs
Patient care—mental health	[85,103,119,123,132,147,149,150,153,166,168-170,184,191,204-206]	[97,104,123,124,155,165,167,172]	[81,82,92,95,121,153,169,173,179,181,188]
Patient care—other	[87,106,114,115,130,133,138,140,141,144,145,148,154,157,159,177,187,194,197,198,203]	[111,127,135,136,146,183]	[87,89,96,107,128,151,156,162,175,180,189]
Health care education and prevention	[86,112,113,116,122,125,129,164,185,195,196,199-202]	[93,118,192]	[90,163,171,193]
Health data collection	[91,98,99,102,105,120,126,139,161,186]	[98,101]	[88,108,152,178]
Older adult daily activities	[100]	[190]	[80,83,84,94]

[a]HCA: health care conversational agent.

Eleven of the embodied HCAs [87,90,92,108,163,171,174,181,182,188,189] were implemented using virtual reality technologies. Of these, 7 [92,108,163,174,181,182,189] were implemented with fully immersive virtual reality technologies, while the remaining 4 used augmented reality [90,171] and mixed reality [87,188] technologies.

Moreover, some studies [80,83,84,89,93,94,96,103,109,110,129,134,155,156,172,179-181] implemented algorithms to recognize users' emotional states. The main reasons for emotion recognition were to enable the HCAs to provide appropriate empathetic responses [80,83,84,89,93,94,96,109,110,129,134,155,156,179] or to suggest specific activities for emotion regulation [155,172,180,181].

Natural language processing based on AI frameworks (eg, Google Dialogflow, Microsoft Language Understanding Intelligent Service, Rasa, or Oscova) was used in several studies to improve dialogue flow [85,86,101,110,112,117,120,124,136,145,167,168,178,179,185,191,195-197,200,202]. In addition, large language models were used in some of the more recent studies [179,185,191,195,196,202].

AI techniques were implemented not only to improve dialogue flow, but also for other purposes, namely, (1) text-based emotion recognition [103,155]; (2) facial emotion recognition [172]; (3) definition of contextualized instructions [171]; (4) avatar generation [176]; (5) prediction of diabetes self-management [194]; and (6) improvement of the consistency and reliability of the information provided [204]. However, most studies did not report evidence of the efficacy of these techniques.

Studies' Experimental Characteristics

Experimental Design

A total of 43 studies [83,84,96,100,104,107,110,112,115,116,118,121,122,124,127-130,132,138,140,141,144,145,149,151,154,160,162,165,170,173,175,177,183,184,197,199-201,203-205] used a longitudinal design, among which 15 studies [100,116,121,124,138,140,141,145,162,173,175,177,184,204,205] randomized participants. The remaining 84 studies used a cross-sectional design.

Cross-sectional studies were conducted in controlled settings, including research or academic settings [81,86,87,89-91,93,97-99,101,103,105,106,108,109,111,113,114,117,120,123,125,126,131,133-137,139,142,143,148,152,153,155-159,161,163,166-169,171,172,174,176,178-181,185-196,198,202], outpatient clinical settings [82,85,88,92,95,102,119,164], hospital clinical settings [146,147,150,182,206], and residential care facilities for older adults [80,94]. In the longitudinal studies, participants used the applications as part of their daily routines, although in some cases data collection was performed in controlled settings. The duration of the longitudinal studies varied from 1 week to 24 weeks: (1) 1 week, 6 [104,110,122,124,129,201] studies; (2) 2 weeks, 8 [100,121,132,151,173,183,197,204] studies; (3) 3 weeks, 1 [140] study; (4) 4 weeks, 7 [96,107,115,128,154,203,205] studies; (5) 5 weeks, 1 [199] study; (6) 6 weeks, 4 [112,160,162,165] studies; (7) 8 weeks, 2 [116,145] studies; (8) 9 weeks, 1 [170] study; (9) 12 weeks, 12 [83,84,118,127,133,138,141,144,149,175,184,200] studies; and (10) 24 weeks, 1 [177] study.

Number and Characteristics of the Participants

Studies included participants from specific age groups (ie, adults or older adults) and health status categories (ie, healthy participants or patients): (1) healthy adults (ie, participants between 18 and 65 years of age) [85-87,92,98,99,101,102,104,105,108-113,117,120-122,124-126,131,132,134,137,142,143,149,152,153,155,158,160,161,163,164,166-168,171-173,177-181,185,186,189,191,192,195,196,199,200,202,204,205]; (2) healthy older adults (ie, participants older than 65 years) [83,84,88,90,93,94,100,115,116,118,123,129,139,141,146,190]; (3) adult patients [81,82,89,95,97,114,119,127,128,130,133,135,136,138,140,144,145,147,150,151,154,156,169,170,176,183,184,187,194,197,198,201,203]; (4) older adult patients [162,165,175,182,206]; (5) healthy adults and adult patients [103,106,148,157,159,174,193]; (6) healthy adults and older adults [91,188]; and (7) healthy adults and older adult patients [80,96,107].

In terms of patients, the following conditions were identified: mental health disorders [81,82,85,92,95,97,103,104,119,121,123,124,132,147,149,150,153,155,165-170,172,173,179,181,184,188,191,204-206], cancer [138,145,182,183,197], heart disease [154], chronic kidney disease [144], chronic pain [105,162,198], COVID-19 [95], delirium [146], dementia [107], diabetes [89,96,111,115,140,194], heart failure [130,187], hydrocephalus [177], hypertension [133,203], intestinal ostomy [148], mild cognitive impairment [141], Parkinson disease [157,175], psoriasis [127], sickle cell disease [114], stroke [128,159,174], traumatic brain injury [151], traumatic injuries [106,180], vestibular disorders [189], and other unspecified chronic diseases [87,156].

XSL•FO
RenderX

Concerning the number of participants, the studies included a total of 7040 participants, of whom 875 were older adults. However, the sample sizes varied considerably across studies. Specifically, 56 studies [82,83,87,91,92,94,95,98-102,107,109,112-116,121,122,124,126,130,134-142,144,145,149,152,153,158,160,161,167,169,170,173,177,180,181,191,194,195,197,200,203,204,206] included more than 30 participants, and 18 of these studies [82,95,100,112,126,139,142,149,158,160,161,170,173,180,181,194,200,204] enrolled more than 100 participants. The remaining 71 studies included 30 or fewer participants.

Validated Subjective Assessment Instruments

Three categories of validated subjective instruments were identified:

- Usability assessment instruments: (1) PSSUQ [30], (2) SUS [31], (3) Chatbot Usability Questionnaire (CUQ) [207], (4) Computer System Usability Questionnaire (CSUQ) [208], (5) mHealth App Usability Questionnaire (MAUQ) [209], (6) Usability Metric for User Experience—Lite Version (UMUX-Lite) [210], and (7) Usefulness, Satisfaction, and Ease of Use Questionnaire (USE) [211].
- User experience assessment instruments: (1) User Experience Questionnaire (UEQ) [212] and (2) User Experience Questionnaire—Short Version (UEQ-S) [213].
- Acceptability assessment instrument: Acceptability E-Scale (AES) [214], which includes a subscale for assessing usability.

Nine studies [86,105,117,119,145,146,174,175,191] applied more than 1 assessment instrument: (1) UEQ and SUS [86,105,117,145,174,175,191]; (2) SUS and USE [119]; and (3) CUQ and SUS [146]. See Table 2 for the validated subjective assessment instruments.

Table 2. Validated subjective assessment instruments.

Instruments	Studies	Total, n
System Usability Scale	[80,81,83-90,92,96-98,100,105-109,111-123,126-130,132,133,135,136,140-142,144-148,150,151,153-157,160-168,170-172,174-177,179,180,182,184,186,190-202]	91
User Experience Questionnaire	[86,101-103,105,110,117,143,145,174,175,178,181,189,191]	15
Chatbot Usability Questionnaire	[125,134,138,146,203,204]	6
Computer System Usability Questionnaire	[91,99,131,139,205]	5
Acceptability E-Scale	[82,95,149,169,206]	5
mHealth App Usability Questionnaire	[159,173,185,187]	4
Usefulness, Satisfaction, and Ease of Use Questionnaire	[94,104,119,183]	4
User Experience Questionnaire—Short Version	[93,152,158]	3
Usability Metric for User Experience—Lite Version	[124,188]	2
Post-Study System Usability Questionnaire	[137]	1

Results of Usability and User Experience Assessments

The usability and user experience scores were generally high. The PSSUQ was applied in only 1 [137] study, which reported a total score of 6.0 (above average, ≥6). Additionally, Table 3 summarizes the ranges of usability and user experience scores for the remaining instruments, based on the studies that reported these scores.

Table 3. Range of the reported usability and user experience scores according to subjective assessment instruments and the standardized scores for each instrument indicating good usability.

Instruments	Score range	Standardized scores indicating good usability or user experience
System Usability Scale	51.4-93.8	≥70.0
Chatbot Usability Questionnaire	40.6-85.9	≥70.0
Computer System Usability Questionnaire	4.6-5.8	≥4.5
Acceptability E-Scale	22.0-25.4	≥24.0
mHealth App Usability Questionnaire	4.8-6.7	≥5.0
Usefulness, Satisfaction, and Ease of Use Questionnaire	4.5-5.7	≥5.0
Usability Metric for User Experience—Lite Version	82.4-86.1	≥70.0
User Experience Questionnaire and User Experience Questionnaire—Short Version	0.4-1.7	≥1.0

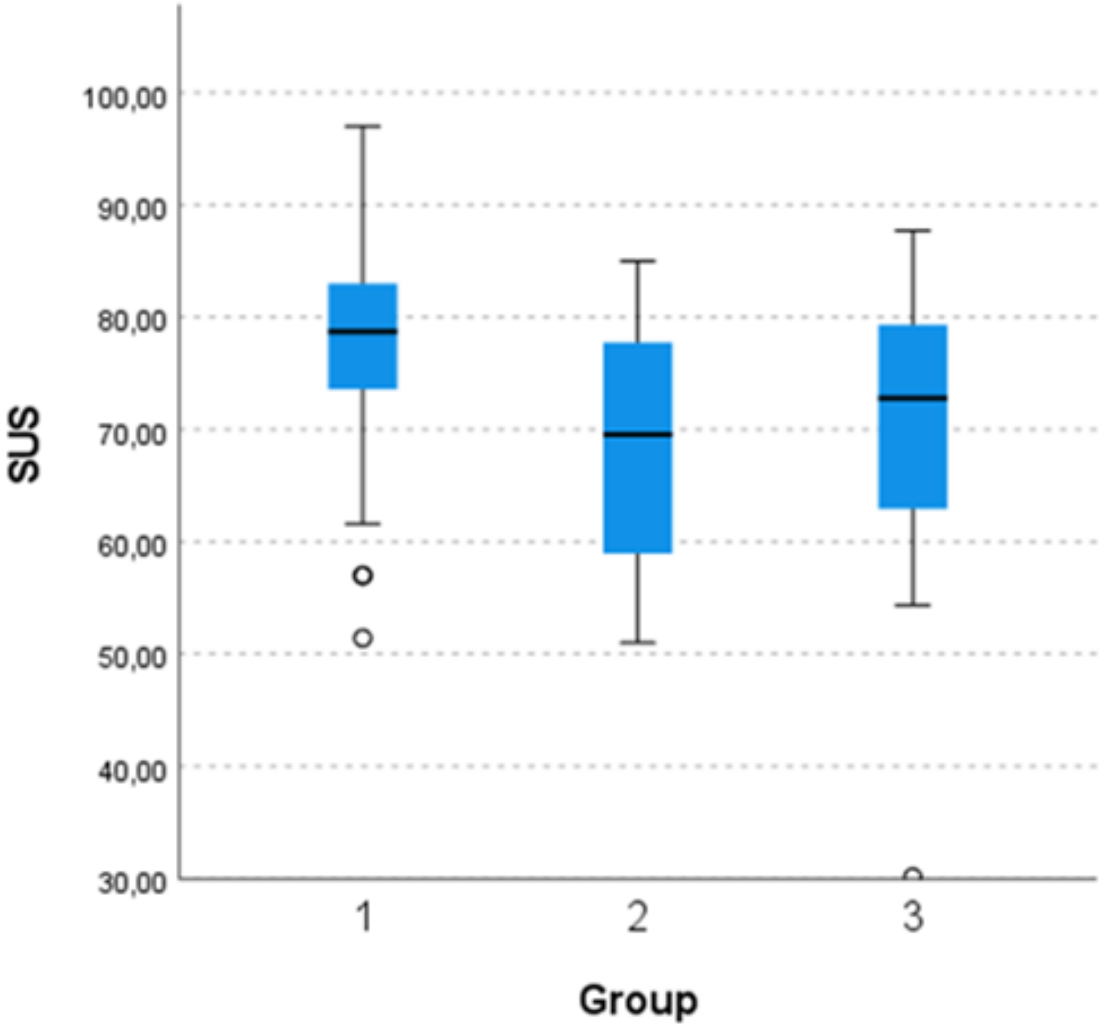
XSL•FO
RenderX

Comparative Analysis of the SUS Results

The following comparative analysis should be interpreted as exploratory because it synthesizes SUS scores reported by independent studies rather than direct comparisons conducted under common experimental conditions. Consequently, the observed differences among HCA categories may be influenced not only by the interaction modality itself but also by variations in health care application domains, study designs, evaluation contexts, participant populations, and HCA implementation and use characteristics. In addition, although the SUS is a well-established measure of perceived usability, it may not adequately capture important aspects of HCAs, such as conversational quality, which is influenced by factors including turn-taking, the ability to sustain coherent and continuous dialogue, and the naturalness of interactions.

Considering the studies that applied the SUS and reported usability scores, a significant difference among groups was found when comparing usability scores across the 3 HCA categories (P=.001; η^2H=0.11; Figure 5, Table 4). Pairwise comparisons showed significant differences between text-based and embodied HCAs (P=.02) and between text- and voice-based HCAs (P=.006). No significant difference was found between voice-based and embodied HCAs (P>.99). When comparing the number of participants across the 3 HCA categories, no significant difference was found (P=.18; η^2H=0.016).

Figure 5. Median and IQR for System Usability Scale (SUS) scores (group 1, text-based HCAs; group 2, voice-based HCAs; and group 3, embodied HCAs). HCA: health care conversational agent.



XSL•FO
RenderX

Table 4. Comparative analysis of SUS[a] scores across the 3 categories of HCAs[b].

Category	Number of studies, n	Number of participants, median (IQR[c])	SUS scores, median (IQR[c])	SUS range
Text-based HCAs	58	30.00 (11.50-52.75)	78.70 (73.43-83.06)	51.40-97.00
Voice-based HCAs	18	22.00 (13.0-34.25)	69.55 (58.90-78.47)	51.00-85.00
Embodied HCAs	26	18.00 (11.00-26.25)	72.79 (62.29-79.38)	30.13-87.70

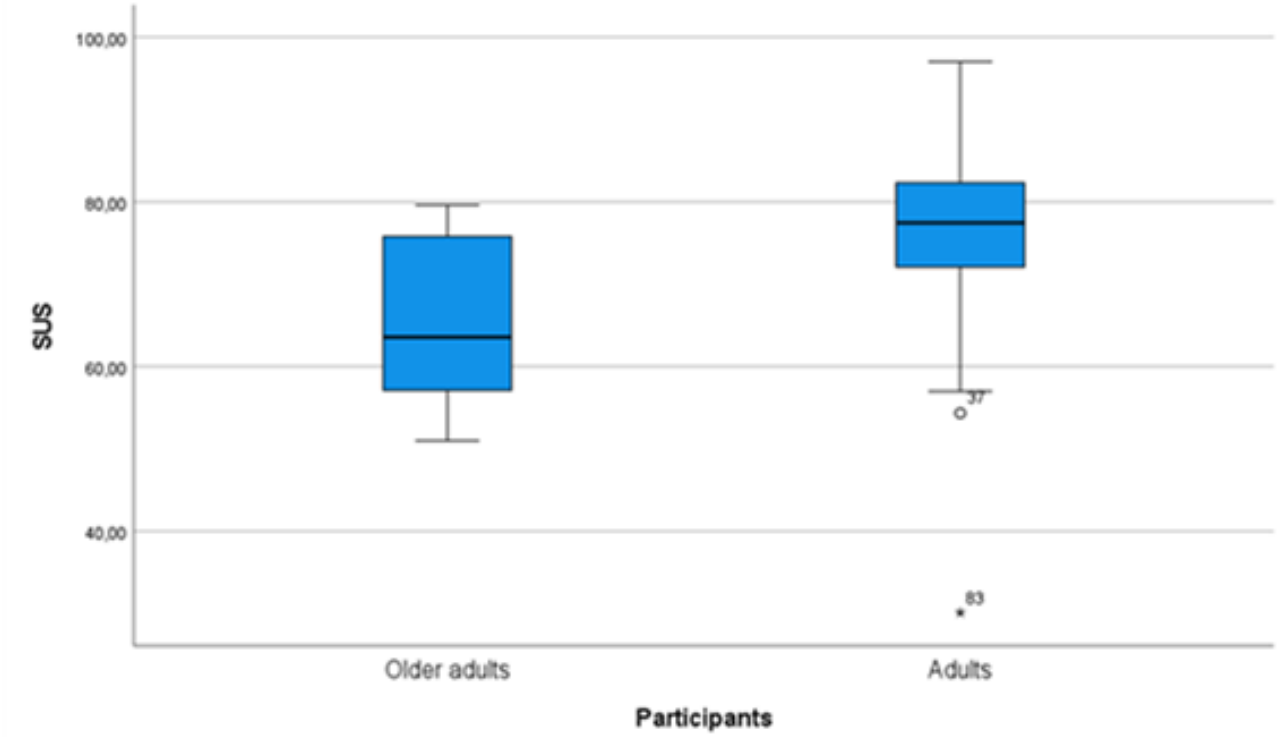
[a]SUS: System Usability Scale.

[b]HCA: health care conversational agent.

[c]25th-75th percentiles.

A significant difference in SUS scores was found when comparing studies involving adults with those involving older adults in the overall sample of studies (P<.001; r=0.39; Figure 6, Table 5). Similarly, a significant difference in SUS scores between studies involving older adults and those involving adults was found for text-based HCAs (P=.009; r=0.33), indicating lower SUS scores in studies involving older adults. No significant between-group difference was found for voice-based HCAs (P=.10; r=0.38) or embodied HCAs (P=.20; r=0.25).

Figure 6. Median and IQR for System Usability Scale (SUS) scores in studies using adults and older adults.



XSL•FO
RenderX

Table 5. Comparative analysis of SUS[a] scores across age range (adults vs older adults).

Category and age group	Number of studies, n	SUS, median (IQR[b])
All HCAs[c]		
Adults	82	77.45 (72.06-82.43)
Older adults	19	63.56 (57.07-76.19)
Text-based HCAs		
Adults	51	79.50 (75.00-83.80)
Older adults	6	64.05 (55.58-78.28)
Voice-based HCAs		
Adults	13	72.10 (59.97-81.17)
Older adults	5	59.69 (52.17-72.00)
Embodied HCAs		
Adults	18	74.41 (67.36-81.08)
Older adults	8	63.75 (57.65-77.78)

[a]SUS: System Usability Scale.

[b]25th-75th percentiles.

[c]HCA: health care conversational agent.

Discussion

This review provides an overview of the current evidence on the usability and user experience of HCAs assessed using validated subjective instruments. As HCAs become increasingly widespread, a growing number of individuals interact with them daily, and it is anticipated that these agents will play an important role in future interactive systems. Therefore, it is crucial to investigate the assessment of their usability and user experience, particularly in the context of emerging human-machine interaction, because well-designed assessments contribute to the overall improvement of interactions. Although well-designed interactions do not necessarily mean that digital health applications are effective, their efficacy and effectiveness may be limited by usability and user experience constraints [215].

Patient care emerged as the predominant application domain for HCAs, particularly in mental health. Other reviews [8,216] have already identified the importance of patient care, while the efficacy of HCAs in supporting mental health disorders has been the subject of several systematic reviews (eg, [217,218]) and a meta-analysis (eg, [219]). Among the other application domains addressed, HCAs may support health education and prevention by optimizing the delivery of understandable, personalized health education at scale and by providing interactive and engaging information across platforms (eg, mobile, web, or smart devices) [220,221]. Similarly, health data collection has already been identified as a core use case of HCAs [216,222], namely, in terms of automating diagnostic interviews and longitudinal monitoring of patient-reported outcomes (eg, anxiety or depression scales) to track symptoms and adherence to treatments over time. Furthermore, the use of HCAs designed to promote the independence and autonomy of older adults by assisting them in managing activities of daily living and self-management of health conditions is in line with current interest in developing digital applications to support the care and well-being of older adults [223,224]. Another relevant point is that a high proportion of the studies included older adults, who accounted for 875 of the 7040 (12.43%) participants.

Although text-based HCAs were the most frequently evaluated, all 3 identified categories were applied across different health care domains. This suggests that developers select interaction modalities according to technological feasibility and user requirements rather than the clinical purpose itself.

The increasing integration of virtual reality with embodied HCAs reflects growing interest in immersive human-computer interaction. However, the relatively limited number of studies indicates that these technologies remain at an early stage of adoption in health care, likely due to higher development costs, hardware requirements, and implementation complexity.

Although AI was increasingly incorporated into HCAs, most of these implementations appear to remain at a relatively immature stage, which is consistent with other studies demonstrating the difficulties of translating AI implementations into clinical workflows [225-227].

SUS clearly dominated the assessment of HCAs, despite having been developed for generic interactive systems. SUS is widely used for usability evaluation of different types of digital health applications [215,228-230] because it is an easy and quick tool with a scoring system that has been translated into multiple languages [228]. However, it was not originally developed to assess complex multimodal interaction mechanisms and does not capture the holistic user experience because it focuses only on usability aspects such as ease of use, satisfaction, and learnability.

Although several alternative instruments were identified, CUQ was the only instrument specifically appropriate for assessing HCAs, even though it does not fully capture the multimodal

XSL•FO
RenderX

and AI-driven characteristics of modern HCAs. The absence of recently developed instruments [46], such as the Conversational Agents Scale [231], the Conversational Agent's Usage Scale [232], the User Experience Evaluation of Conversational Agents [233], or the Conversational Agent Scale for User Experience [234], further highlights the need for greater methodological standardization in future evaluations. Moreover, a relevant number of studies were excluded from this review because usability and user experience assessments were based on questionnaires consisting of Likert scale questions developed by the researchers. However, these questionnaires were not validated instruments because they had not undergone psychometric analysis, which is frequently a lengthy process that can be particularly challenging for researchers who lack experience in developing and validating subjective instruments.

Usability and user experience assessments were conducted in both controlled (ie, research or academic settings, outpatient clinical settings, hospital clinical settings, and institutional residences for older adults) and uncontrolled (eg, participants' homes) environments. Controlled experiments generally focus on intensive sessions to capture immediate usability or user experience feedback. These experiments enable consistent data collection, but they may constrain the observation of more natural interaction patterns. By contrast, longitudinal studies conducted in uncontrolled settings allow participants to engage with HCAs more organically, integrating their use into daily routines. Therefore, uncontrolled environments provide authentic user experiences and valuable insights, as well as functional challenges, including connectivity issues, reflecting authentic usage scenarios. Moreover, many longitudinal studies strategically combined uncontrolled environments with data collection in controlled environments.

The participants comprised both patients and healthy adults or older adults. However, the reviewed studies often did not adequately justify their choice of sample sizes. Small samples can undermine the reliability of results by increasing the risk of sampling bias and reducing statistical power. Consequently, the findings may not capture the full variability of the population, thereby restricting their generalizability.

The quality assessment of the components related to the application of subjective measurement instruments highlighted methodological limitations, including insufficient evaluation of prolonged real-world use and inadequate justification of sample sizes, the 2 CAUSS items with lower percentages. These shortcomings reduce confidence in the ecological validity and generalizability of the reported usability findings.

Looking at the assessment results, the usability and user experience of HCAs tend to be above average; however, the results indicate differences among the various HCA categories. The comparison of usability scores across the 3 HCA modalities revealed statistically significant differences, with pairwise comparisons indicating that text-based HCAs achieved significantly higher usability scores than both voice-based and embodied HCAs, whereas no significant differences were observed between voice-based and embodied HCAs. The estimated eta-squared (η^2H=0.11) indicates a moderate effect size, suggesting a moderate association between interaction modality and users' perceptions of usability. No significant differences were found in the number of participants across the 3 HCA categories, indicating that the studies included broadly comparable sample sizes. Consequently, systematic differences in study size are unlikely to fully explain the observed differences in usability.

Several factors may contribute to the higher usability observed for text-based HCAs. Although the reviewed studies did not directly investigate the underlying mechanisms, one possible explanation is users' greater familiarity with text-based communication, particularly through messaging applications. Moreover, text interfaces provide a persistent conversation history that enables users to review previous exchanges, potentially increasing their sense of control and reducing memory demands during interaction. In addition, text-based interactions allow users to engage with HCAs at their own pace, giving them more time to read, reflect on, and formulate responses. By contrast, voice-based HCAs may be affected by speech recognition inaccuracies, environmental noise, and privacy concerns, whereas embodied HCAs introduce additional visual and social interaction elements that may increase interaction complexity without necessarily improving perceived usability.

The results also suggest that older adults tend to perceive HCAs as less usable than adults. This may be associated with older adults generally having lower digital literacy, as well as age-related accessibility needs. For example, age-related sensory, cognitive, and motor changes, together with the higher prevalence of multimorbidity among older adults, may affect their ability to interact effectively with HCAs. Nevertheless, the available evidence from the included studies does not allow this explanation to be confirmed.

Interestingly, the difference between adults and older adults was statistically significant only for text-based HCAs, despite this modality achieving the highest usability scores overall. A potential explanation is that text-based HCAs are currently more widespread than voice-based or embodied HCAs, which may lead to greater differences in familiarity and prior experience between age groups.

Although no statistically significant age-related differences were observed for voice-based or embodied HCAs, the estimated effect sizes were of comparable magnitude across modalities. Therefore, the absence of statistical significance may reflect the smaller number of available studies and limited statistical power rather than provide evidence that age has no influence on perceived usability in these interaction modalities.

These results should be interpreted with caution. The included studies exhibited substantial heterogeneity in terms of application domains (patient care, health education and prevention, health data collection, and support for daily activities), study designs (eg, cross-sectional or longitudinal), evaluation contexts (controlled versus real-world settings), target populations, underlying technologies (eg, AI, large language models, or virtual reality), task complexity, and duration of HCA use (single-session vs prolonged use). Consequently, the observed differences in SUS scores may have been influenced by factors beyond the interaction modality itself. Furthermore,

XSL•FO
RenderX

the SUS assesses perceived usability rather than objective performance or long-term acceptance, and therefore high SUS scores do not necessarily translate into sustained engagement or continued use. Future studies should investigate which specific design characteristics contribute most to usability across different HCA modalities while accounting for user characteristics and contextual factors.

Despite the use of extensive search strategies to locate relevant studies, some articles may have been overlooked due to variations in terminology. Furthermore, the exclusion of non-English publications and gray literature may have limited the overall representativeness of the findings and introduced language and publication bias, as studies published in other languages or unpublished studies with null or negative findings may have been underrepresented.

The wide variety of assessment instruments used meant that it was generally not possible to identify enough studies to enable a comparative analysis of usability or user experience outcomes by instrument across different HCA categories. The only exception was SUS; however, as it was developed in the 1990s, it may not adequately capture relevant aspects of the usability of modern HCAs.

Finally, this review focused solely on the usability and user experience of HCAs and did not address their efficacy or effectiveness.

In conclusion, this review included 127 studies applying validated subjective instruments to assess the usability and user experience of 3 HCA categories: text-based, voice-based, and embodied HCAs. In terms of health care applications, HCAs aimed to support patient care, health education and prevention, health data collection, and daily activities of older adults.

The usability and user experience of HCAs were assessed in cross-sectional and longitudinal studies using the following instruments: SUS, UEQ, UEQ-S, CUQ, AES, MAUQ, UMUX-Lite, USE, CSUQ, and PSSUQ.

This review highlights that, except for CUQ, the identified subjective assessment instruments were not specifically developed to evaluate the usability and user experience of HCAs. Therefore, the findings suggest that such evaluations may be limited by the inadequacy of these instruments.

Considering the comparative analysis of SUS scores, text-based HCAs tend to show higher usability scores than voice-based and embodied HCAs, and older adults tend to perceive HCAs as less usable than adults. However, these findings represent associations observed across the included studies rather than direct evidence that one HCA category is inherently more usable than another, because the studies differ substantially in their health care applications, design, evaluation contexts, participant populations, and implementation and use characteristics. Moreover, a relevant number of studies used small sample sizes, which reduce statistical power, may fail to capture population variability, limit the generalizability of the results, and increase the risk of misleading conclusions. Most HCAs were also assessed in controlled environments at a single point in time, without considering the impact of continuous and long-term use in real-world settings.

Considering the substantial differences in how evaluations were designed and conducted across studies, especially regarding data collection methods, participant selection, testing conditions, and reporting of results, there is a need for standardized protocols and established guidelines to measure the usability and user experience of HCAs. Such standardization would improve methodological consistency and enhance the comparability of findings across studies, enabling researchers and practitioners to benchmark usability and user experience.

Funding

This research received no external funding.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.
[PDF File (Adobe PDF File), 502 KB - humanfactors_v13i1e102911_app1.pdf]

References

1. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. Commun ACM 1966 Jan;9(1):36-45. [doi: 10.1145/365153.365168]
2. Caldarini G, Jaf S, McGarry K. A literature survey of recent advances in chatbots. Information 2022 Jan 15;13(1):41. [doi: 10.3390/info13010041]
3. Leschanowsky A, Rech S, Popp B, Bäckström T. Evaluating privacy, security, and trust perceptions in conversational AI:a systematic review. Computers in Human Behavior 2024 Oct;159:108344. [doi: 10.1016/j.chb.2024.108344]
4. McTear M. The rise of the conversational interface: a new kid on the block? Cham, Switzerland: Springer; 2017 Presented at: International Workshop on the Future and Emerging Trends in Language Technology; 30 November to 2 December 2016; Seville, Spain p. 38-49. [doi: 10.1007/978-3-319-69365-1_3]

XSL•FO
RenderX

5. McTear M, Callejas Z, Griol D. Future directions. In: McTear M, Callejas Z, Griol D, editors. The Conversational Interface. Cham, Switzerland: Springer; 2016:403-418.
6. Abbasian M, Azimi I, Rahmani A, Jain R. Conversational health agents: a personalized large language model-powered agent framework. JAMIA Open 2025 Aug;8(4):ooaf067 [FREE Full text] [doi: 10.1093/jamiaopen/ooaf067] [Medline: 40620478]
7. Wang J, Ma B, Nan Y, He Y, Sun H, Liu C, et al. Improving matching models with contextual attention for multi-turn response selection in retrieval-based chatbots. IEEE Trans Netw Sci Eng 2025 May;12(3):1497-1509. [doi: 10.1109/tnse.2025.3532663]
8. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
9. Liu B. Sentiment Analysis: Mining opinions, Sentiments, and Emotions. Cambridge, UK: Cambridge University Press; 2020.
10. Shah M, Kavathiya H. Exploring conversational AI. In: Al-Marzouqi A, Salloum S, Al-Saidat M, Aburayya A, Gupta B, editors. Artificial Intelligence in Education: The Power and Dangers of ChatGPT in the Classroom. Cham, Switzerland: Springer; 2024:511-526.
11. Cassell J, Bickmore T, Billinghurst M, Campbell L, Chang K, Vilhjálmsson H, et al. Embodiment in conversational interfaces: Rea. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 1999 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 15-20, 1999; Pittsburgh, PA p. 520-527. [doi: 10.1145/302979.303150]
12. Swartout W, Gratch J, Hill JR, Hovy E, Marsella S, Rickel J, et al. Toward virtual humans. The USC Institute for Creative Technologies. 2006. URL: https://people.ict.usc.edu/~gratch/AI-mag06.pdf [accessed 2026-07-27]
13. Wolfert P, Robinson N, Belpaeme T. A review of evaluation practices of gesture generation in embodied conversational agents. IEEE Trans Human-Mach Syst 2022 Jun;52(3):379-389. [doi: 10.1109/THMS.2022.3149173]
14. Dai X, Liu Z, Liu T, Zuo G, Xu J, Shi C, et al. Modelling conversational agent with empathy mechanism. Cognitive Systems Research 2024 Mar;84:101206. [doi: 10.1016/j.cogsys.2023.101206]
15. Nguyen TTN, Dam QT, Tran DT, Lee J. A survey on generative nonverbal facial behavior for highly realistic embodied agents. Intel Serv Robotics 2026 Jan 28;19(2):27. [doi: 10.1007/s11370-025-00674-2]
16. Sajjadi P, Hoffmann L, Cimiano P, Kopp S. A personality-based emotional model for embodied conversational agents: effects on perceived social presence and game experience of users. Entertainment Computing 2019 Dec;32:100313. [doi: 10.1016/j.entcom.2019.100313]
17. Kroczek LO, May A, Hettenkofer S, Ruider A, Ludwig B, Mühlberger A. The influence of persona and conversational task on social interactions with a LLM-controlled embodied conversational agent. Computers in Human Behavior 2025 Nov;172:108759. [doi: 10.1016/j.chb.2025.108759]
18. Yang F, Acevedo P, Guo S, Choi M, Mousas C. Embodied conversational agents in extended reality: a systematic review. IEEE Access 2025;13:79805-79824. [doi: 10.1109/access.2025.3566698]
19. Huynh AL, Roy TJ, Jackson KN, Lee AG, Liaw W, Hossain MM. Applications of artificial intelligence-based conversational agents in healthcare: a systematic umbrella review. Int J Med Inform 2026 Mar 01;207:106204 [FREE Full text] [doi: 10.1016/j.ijmedinf.2025.106204] [Medline: 41337874]
20. Bălan C. Chatbots and voice assistants: digital transformers of the company–customer interface—a systematic review of the business research literature. JTAER 2023 May 18;18(2):995-1019. [doi: 10.3390/jtaer18020051]
21. Xu Z, Liu F, Xia G, Wang S, Duan Y, Yu L, et al. Immersive HCI for intangible cultural heritage in tourism contexts: a narrative review of design and evaluation. Sustainability 2025 Dec 23;18(1):153. [doi: 10.3390/su18010153]
22. Yusuf H, Money A, Daylamani-Zad D. Pedagogical AI conversational agents in higher education: a conceptual framework and survey of the state of the art. Education Tech Research Dev 2025 Jan 15;73(2):815-874. [doi: 10.1007/s11423-025-10447-4]
23. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337(sep29 1):a1655-a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
24. Fu L, Burns R, Xie Y, Shen J, Zhe S, Estabrooks P, et al. The development and use of AI chatbots for health behavior change: scoping review. J Med Internet Res 2026 Jan 28;28:e79677 [FREE Full text] [doi: 10.2196/79677] [Medline: 41604667]
25. Nielsen J. The usability engineering life cycle. Computer 1992 Mar;25(3):12-22. [doi: 10.1109/2.121503]
26. Hassan H, Galal-Edeen G. From usability to user experience. In: Proceedings of the International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS). New York, NY: IEEE; 2017 Presented at: International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS); November 24-26, 2017; Okinawa, Japan p. 216-222.
27. International Organization for Standardization (ISO). ISO 9241-210-2019: Ergonomics of human-system interaction—Part 210: Human-centred design for interactive systems. ISO. Geneve, Switzerland: International Organization for Standardization; 2019. URL: https://www.iso.org/standard/77520.html [accessed 2026-07-22]

XSL•FO
RenderX

28. Ntoa S. Usability and user experience evaluation in intelligent environments: a review and reappraisal. International Journal of Human–Computer Interaction 2024 Sep 12;41(5):1-30. [doi: 10.1080/10447318.2024.2394724]
29. Martins A, Queirós A, Silva A, Rocha N. Usability evaluation methods: a systematic review. In: Saeed S, Bajwa IS, Mahmood Z, editors. Human Factors in Software Development and Design. New York, NY: IGI Global; 2015:250-273.
30. Lewis JR. Psychometric evaluation of the Post-Study System Usability Questionnaire: The PSSUQ. Proceedings of the Human Factors Society Annual Meeting 1992;36(16):1259-1260. [doi: 10.1177/154193129203601617]
31. Brooke J. SUS-A quick and dirty usability scale. In: Jordan P, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation in Industry. Boca Raton, FL: CRC Press; 1996:4-7.
32. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. International Journal of Human–Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
33. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. Journal of Usability Studies 2009;4(3):114-123 [FREE Full text]
34. Branda F, Stella M, Ceccarelli C, Cabitza F, Ceccarelli G, Maruotti A, et al. The role of AI-based chatbots in public health emergencies: a narrative review. Future Internet 2025 Mar 26;17(4):145. [doi: 10.3390/fi17040145]
35. Orrù L, Mannarini S. The role of artificial intelligence in clinical psychology: how AI and NLP systems are reshaping psychological interventions. A systematic review. Clin Psychol Psychother 2026 Feb 25;33(2):e70242. [doi: 10.1002/cpp.70242] [Medline: 41738110]
36. Mahajan P, Kadam S, Kachwala H, Tapikar N, Namde V, Poonawala E. Reimagining mental health support: the role of chatbots in bridging gaps and raising ethical questions. Couns and Psychother Res 2026 Feb 15;26(1):e70095. [doi: 10.1002/capr.70095]
37. Cesar Abrantes P, Netto AV, Kazuo Takahata A. Healthbots for conducting clinical screening and remote monitoring with patient mood assessment: a scoping review. Int J Med Inform 2026 Mar 01;207:106186. [doi: 10.1016/j.ijmedinf.2025.106186] [Medline: 41308276]
38. Fiske S, Cody JL, Shen M, Choi J. AI conversational agents in older adults with chronic disease: a scoping review. Geriatr Nurs 2026 Mar;68:103856. [doi: 10.1016/j.gerinurse.2026.103856] [Medline: 41576576]
39. Dong P, Zhang S, Zheng X, Yuan W, Wu L, Chen Y. Artificial intelligence robots for mental health applications: a scoping review. Psychiatry Res 2026 Jun;360:117059. [doi: 10.1016/j.psychres.2026.117059] [Medline: 41819775]
40. Jalali S, You Q, Xu V, Chen L, Zini J, Zamorano T, et al. The use of artificial intelligence for personalized treatment in psychiatry. Curr Psychiatry Rep 2025 Dec 29;28(1):7. [doi: 10.1007/s11920-025-01656-y] [Medline: 41457119]
41. Tahir H, Özdemir MK, Alhajj R. The role of LLM-powered chatbots in assisting elderly people: systematic review. Netw Model Anal Health Inform Bioinforma 2025 Dec 27;15(1):26. [doi: 10.1007/s13721-025-00698-9]
42. Curtis RG, Bartel B, Ferguson T, Blake HT, Northcott C, Virgara R, et al. Improving user experience of virtual health assistants: scoping review. J Med Internet Res 2021 Dec 21;23(12):e31737 [FREE Full text] [doi: 10.2196/31737] [Medline: 34931997]
43. Bastardo R, Pavão J, Rocha N. User-centred usability evaluation of embodied communication agents to support older adults: a scoping review. Cham, Switzerland: Springer; 2022 Presented at: International Conference on Information Technology Systems; February 9-11, 2022; San Carlos, Costa Rica p. 509-518. [doi: 10.1007/978-3-030-96293-7_42]
44. Li K, Wu S, Zhang Y, Zhu B, Qi Z, Hou S, et al. The usability and experience of artificial intelligence-based conversational agents in health education for cancer patients: a scoping review. J Clin Nurs 2025 Jul 14:1-11 (forthcoming). [doi: 10.1111/jocn.70020] [Medline: 40654175]
45. Faruk LID, Babakerkhell MD, Mongkolnam P, Chongsuphajaisiddhi V, Funilkul S, Pal D. A review of subjective scales measuring the user experience of voice assistants. IEEE Access 2024;12:14893-14917. [doi: 10.1109/access.2024.3358423]
46. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8(5):336-341 [FREE Full text] [doi: 10.1016/j.ijsu.2010.02.007] [Medline: 20171303]
47. Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020 Oct;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
48. Silva AG, Simões P, Santos R, Queirós A, Rocha NP, Rodrigues M. A scale to assess the methodological quality of studies assessing usability of electronic health products and services: Delphi study followed by validity and reliability testing. J Med Internet Res 2019 Nov 15;21(11):e14829 [FREE Full text] [doi: 10.2196/14829] [Medline: 31730036]
49. Fiel PF. Effect sizes for nonparametric tests. Biochemia Medica 2026;36(1):5-16. [doi: 10.11613/bm.2026.010101]
50. Mendu S, Boukhechba M, Gordon J, Datta D, Molina E, Arroyo G. Design of a culturally-informed virtual human for educating hispanic women about cervical cancer. In: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare. New York, NY: Association for Computing Machinery; 2018 Presented at: 12th EAI International Conference on Pervasive Computing Technologies for Healthcare; 21-24 May 2018; New York, NY p. 330-336. [doi: 10.1145/3240925.3240968]
51. Sun O, Chen J, Magrabi F. Using voice-activated conversational interfaces for reporting patient safety incidents: a technical feasibility and pilot usability study. Stud Health Technol Inform 2018;252:139-144. [doi: 10.3233/978-1-61499-890-7-139]

XSL•FO
RenderX

52. Booth A, van der Krogt M, Buizer A, Steenbrink F, Harlaar J. The validity and usability of an eight marker model for avatar-based biofeedback gait training. Clin Biomech (Bristol) 2019 Dec;70:146-152. [doi: 10.1016/j.clinbiomech.2019.08.013] [Medline: 31499394]
53. Huff JE, Mack N, Cummings R, Womack K, Gosha K, Gilbert J. Evaluating the usability of pervasive conversational user interfaces for virtual mentoring. Cham, Switzerland: Springer; 2019 Presented at: International Conference on Human-Computer Interaction; July 26-31, 2019; Orlando, FL. [doi: 10.1007/978-3-030-21817-1_7]
54. Ter Stal S, Broekhuis M, van Velsen L, Hermens H, Tabak M. Embodied conversational agent appearance for health assessment of older adults: explorative study. JMIR Hum Factors 2020 Sep 04;7(3):e19987 [FREE Full text] [doi: 10.2196/19987] [Medline: 32886068]
55. Roman MK, Bellei EA, Biduski D, Pasqualotti A, De Araujo CDSR, De Marchi ACB. "Hey assistant, how can I become a donor?" The case of a conversational agent designed to engage people in blood donation. J Biomed Inform 2020 Jul;107:103461 [FREE Full text] [doi: 10.1016/j.jbi.2020.103461] [Medline: 32504669]
56. Luengo-Polo J, Conde-Caballero D, Rivero-Jiménez B, Ballesteros-Yáñez I, Castillo-Sarmiento CA, Mariano-Juárez L. Rationale and methods of evaluation for ACHO, a new virtual assistant to improve therapeutic adherence in rural elderly populations: a user-driven living lab. Int J Environ Res Public Health 2021 Jul 26;18(15):7904 [FREE Full text] [doi: 10.3390/ijerph18157904] [Medline: 34360194]
57. de Pennington N, Mole G, Lim E, Milne-Ives M, Normando E, Xue K, et al. Safety and acceptability of a natural language artificial intelligence assistant to deliver clinical follow-up to cataract surgery patients: proposal. JMIR Res Protoc 2021 Jul 28;10(7):e27227 [FREE Full text] [doi: 10.2196/27227] [Medline: 34319248]
58. Bassi G, Donadello I, Gabrielli S, Salcuni S, Giuliano C, Forti S. Early development of a virtual coach for healthy coping interventions in type 2 diabetes mellitus: validation study. JMIR Form Res 2022 Feb 11;6(2):e27500 [FREE Full text] [doi: 10.2196/27500] [Medline: 35147505]
59. Islam A, Chaudhry B. Early usability evaluation of a relational agent for the COVID-19 pandemic. In: Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology. New York, NY: Association for Computing Machinery; 2022 Presented at: 35th Annual ACM Symposium on User Interface Software and Technology; 29 October to 2 November 2022; Bend, OR p. 1-3. [doi: 10.1145/3526114.3558702]
60. Nikou S, Chang M. Learning by building chatbot: a system usability study and teachers' views about the educational uses of chatbots. Cham, Switzerland: Springer; 2023 Presented at: International Conference on Intelligent Tutoring Systems; June 2-5, 2023; Corfu, Greece p. 342-351. [doi: 10.1007/978-3-031-32883-1_31]
61. Li J, Cesar P. Social virtual reality (VR) applications and user experiences. In: Valenzise G, Alain M, Zerman E, Ozcinar C, editors. Immersive Video Technologies. Cambridge, MA: Academic Press; 2023:609-648.
62. García AS, Fernández-Sotos P, Vicente-Querol MA, Sánchez-Reolid R, Rodriguez-Jimenez R, Fernández-Caballero A. Co-design of avatars to embody auditory hallucinations of patients with schizophrenia. Virtual Reality 2021 Jul 08;27(1):217-232. [doi: 10.1007/s10055-021-00558-7]
63. Blasco J, Díaz-Díaz B, Igual-Camacho C, Pérez-Maletzki J, Hernández-Guilén D, Roig-Casasús S. Effectiveness of using a chatbot to promote adherence to home physiotherapy after total knee replacement, rationale and design of a randomized clinical trial. BMC Musculoskelet Disord 2023 Jun 15;24(1):491. [doi: 10.1186/s12891-023-06607-3]
64. Lent HC, Ortner VK, Karmisholt KE, Wiegell SR, Nissen CV, Omland SH, et al. A chat about actinic keratosis: examining capabilities and user experience of ChatGPT as a digital health technology in dermato‐oncology. JEADV Clinical Practice 2023 Oct 27;3(1):258-265. [doi: 10.1002/jvc2.263]
65. Knobel SEJ, Oberson R, Räber J, Schütz N, Egloff N, Botros A, et al. Evaluation of a new mobile virtual reality setup to alter pain perception: pilot development and usability study in healthy participants. JMIR Serious Games 2024 Dec 11;12:e52340-e52340 [FREE Full text] [doi: 10.2196/52340] [Medline: 39661868]
66. Murawski A, Ramirez-Zohfeld V, Mell J, Tschoe M, Schierer A, Olvera C, et al. NegotiAge: development and pilot testing of an artificial intelligence-based family caregiver negotiation program. J Am Geriatr Soc 2024 Apr 13;72(4):1112-1121. [doi: 10.1111/jgs.18775] [Medline: 38217356]
67. Macedo P, Madeira RN, Santos PA, Mota P, Alves B, Pereira CM. A conversational agent for empowering people with Parkinson's disease in exercising through motivation and support. Applied Sciences 2024 Dec 30;15(1):223. [doi: 10.3390/app15010223]
68. Holderried F, Stegemann-Philipps C, Herschbach L, Moldt J, Nevins A, Griewatz J, et al. A generative pretrained transformer (GPT)-powered chatbot as a simulated patient to practice history taking: prospective, mixed methods study. JMIR Med Educ 2024 Jan 16;10:e53961 [FREE Full text] [doi: 10.2196/53961] [Medline: 38227363]
69. Ahmmed A, Butts E, Naeiji K, Thiamwong L, Daher S. System usability and technology acceptance of a geriatric embodied virtual human simulation in augmented reality. In: Proceedings of the 2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). New York, USA: IEEE; 2024 Presented at: 2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR); October 21-25, 2024; Bellevue, WA p. 594-603. [doi: 10.1109/ISMAR62088.2024.00074]
70. Arreola W, Rivas JJ, Castrejón L, Sucar LE, Rivas JJ, Rivas JJ, et al. Affective embodied agent for patient assistance in virtual rehabilitation. IEEE Trans Affective Comput 2025 Oct;16(4):3110-3121. [doi: 10.1109/taffc.2025.3569676]

XSL•FO
RenderX

71. Ajana K, Everard G, Lejeune T, Edwards MG. An immersive virtual reality serious game set for the clinical assessment of spatial attention impairments: effects of avatars on perspective? Technol Health Care 2025 Jul 17;33(4):1626-1644. [doi: 10.1177/09287329241296380] [Medline: 39973878]
72. Tasleem U, Qamar T. Enhancing interview skills through virtual reality–a usability study. In: Proceedings 2025 International Conference on Emerging Technologies in Electronics, Computing, and Communication (ICETECC). New York, NY: IEEE; 2025 Presented at: 2025 International Conference on Emerging Technologies in Electronics, Computing, and Communication (ICETECC); June 23-26, 2025; Brest, France p. 1-5. [doi: 10.1109/icetecc65365.2025.11070266]
73. Laverde N, Grévisse C, Jaramillo S, Manrique R. Integrating large language model-based agents into a virtual patient chatbot for clinical anamnesis training. Comput Struct Biotechnol J 2025;27:2481-2491. [doi: 10.1016/j.csbj.2025.05.025] [Medline: 40547455]
74. Liu M, Su Y. Mediguide: reducing cognitive load and improving elderly patients' healthcare experience. In: Proceedings of the Extended Abstracts of the 2025 CHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2025 Presented at: 2025 CHI Conference on Human Factors in Computing Systems; April 26 to May 1, 2025; Yokohama, Japan p. 1-7. [doi: 10.1145/3706599.3706701]
75. Gomes de Siqueira A, Gehling GM, Subramanian A, Le A, Garg R, Islam AR, et al. Methods to modernize a multimedia, web-based reproductive health education intervention for individuals with sickle cell disease or trait using virtual human narration and user-centered design. Health Informatics J 2025 Nov 21;31(4):14604582251397311 [FREE Full text] [doi: 10.1177/14604582251397311] [Medline: 41270235]
76. Suri A, Sindwani G. Mothers' Assistant for Labor analgesia (MALA): a novel artificial intelligence interactive avatar for patient education in obstetric anesthesia. Int J Obstet Anesth 2025 Nov;64:104762. [doi: 10.1016/j.ijoa.2025.104762] [Medline: 40915118]
77. Valério MDP, Boschi SRMDS, Ribeiro DL, Fernandes da Silva AR, Moura LDA, Alexandre GD, et al. Proposal of a serious game for dynamic balance training using a force platform: a pilot study. Games Health J 2025 Aug 01;14(4):312-320. [doi: 10.1089/g4h.2023.0194] [Medline: 40126159]
78. Lehocki F, Dudasko S, Vrins A, Tirpakova V, Discantini I, Putekova S, et al. To physically embody or not? A comparison of virtual vs. physical robots as exercise coaches for older adults. In: Proceedings of the 34th International Conference on Robot and Human Interactive Communication (RO-MAN). New York, NY: IEEE; 2025 Presented at: 34th International Conference on Robot and Human Interactive Communication (RO-MAN); August 25-29, 2025; Eindhoven, The Netherlands p. 2340-2345. [doi: 10.1109/RO-MAN63969.2025.11217653]
79. Swallow V, Horsman J, Mazlan E, Campbell F, Zaidi R, Julian M, et al. DigiBete, a novel chatbot to support transition to adult care of young people/young adults with type 1 diabetes mellitus: outcomes from a prospective, multimethod, nonrandomized feasibility and acceptability study. JMIR Diabetes 2025 Jul 23;10:e74032 [FREE Full text] [doi: 10.2196/74032] [Medline: 40699892]
80. Hanke S, Sandner E, Kadyrov S, Stainer-Hochgatterer A. Daily life support at home through a virtual support partner. In: Proceedings of the 2nd IET International Conference on Technologies for Active and Assisted Living (TechAAL 2016). London, UK: IET; 2016 Presented at: 2nd IET International Conference on Technologies for Active and Assisted Living (TechAAL 2016); October 24-25, 2016; London, UK p. 1-7. [doi: 10.1049/ic.2016.0058]
81. Tielman ML, Neerincx MA, Bidarra R, Kybartas B, Brinkman W. A therapy system for post-traumatic stress disorder using a virtual agent and virtual storytelling to reconstruct traumatic memories. J Med Syst 2017 Aug 11;41(8):125 [FREE Full text] [doi: 10.1007/s10916-017-0771-y] [Medline: 28699083]
82. Philip P, Micoulaud-Franchi J, Sagaspe P, Sevin ED, Olive J, Bioulac S, et al. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. Sci Rep 2017 Feb 16;7(1):42656 [FREE Full text] [doi: 10.1038/srep42656] [Medline: 28205601]
83. Tsiourti C, Quintas J, Ben-Moussa M, Hanke S, Nijdam N, Konstantas D. The CaMeLi framework—a multimodal virtual companion for older adults. Cham, Switzerland: Springer; 2018 Presented at: IntelliSys 2016: SAI Intelligent Systems Conference; 21-22 September 2016; London, UK p. 196-217. [doi: 10.1007/978-3-319-69266-1_10]
84. Tsiourti C, Ben MM, Quintas J, Loke B, Jochem I, Lopes J, et al. A virtual assistive companion for older adults: design implications for a real-world application. Cham, Switzerland: Springer; 2018 Presented at: IntelliSys 2016: SAI Intelligent Systems Conference; September 21-22, 2016; London, UK p. 1014-1033. [doi: 10.1007/978-3-319-56994-9_69]
85. Cameron G, Cameron D, Megaw G, Bond R, Mulvenna M, O'Neill S. Assessing the usability of a chatbot for mental health care. Cham, Switzerland: Springer; 2019 Presented at: International Conference on Internet Science; October 24-26, 2018; St. Petersburg, Russia p. 121-132. [doi: 10.1007/978-3-030-17705-8_11]
86. Holmes S, Moorhead A, Bond R, Zheng H, Coates V, McTear M. Usability testing of a healthcare chatbot: can we use conventional methods to assess conversational user interfaces? In: Proceedings of the 31st European Conference on Cognitive Ergonomics. New York, NY: Association for Computing Machinery; 2019 Presented at: 31st European Conference on Cognitive Ergonomics; September 10-13, 2019; Belfast, Northern Ireland, UK p. 207-214. [doi: 10.1145/3335082.3335094]
87. Kim K, Norouzi N, Losekamp T, Bruder G, Anderson M, Welch G. Effects of patient care assistant embodiment and computer mediation on user experience. In: Proceedings of the IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR). New York, NY: IEEE; 2019 Presented at: IEEE International Conference on Artificial

XSL•FO
RenderX

Intelligence and Virtual Reality (AIVR); December 9-11, 2019; San Diego, CA p. 17-177. [doi: 10.1109/AIVR46125.2019.00013]
88. Sun R, Aldunate RG, Paramathayalan VR, Ratnam R, Jain S, Morrow DG, et al. Preliminary evaluation of a self-guided fall risk assessment tool for older adults. Arch Gerontol Geriatr 2019 May;82:94-99. [doi: 10.1016/j.archger.2019.01.022] [Medline: 30735851]
89. Faddoul G, Chatterjee S. The virtual diabetician: a prototype for a virtual avatar for diabetes treatment using persuasion through storytelling. In: Proceedings of the 25th Americas Conference on Information Systems, AMCIS 2019. Atlanta, GA: Association for Information Systems; 2019 Presented at: 25th Americas Conference on Information Systems, AMCIS 2019; August 15-17, 2019; Cancún, Mexico p. 1-10.
90. Mostajeran F, Steinicke F, Ariza NO, Gatsios D, Fotiadis D. Augmented reality for older adultsxploring acceptability of virtual coaches for home-based balance training in an aging population. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. New York, USA: ACM; 2020 Presented at: 2020 CHI Conference on Human Factors in Computing Systems; April 25-30, 2020; Honolulu, HI p. 1-12. [doi: 10.1145/3313831.3376565]
91. Ponathil A, Ozkan F, Bertrand J, Agnisarman S, Narasimha S, Welch B, et al. An empirical study investigating the user acceptance of a virtual conversational agent interface for family health history collection among the geriatric population. Health Informatics J 2020 Dec 16;26(4):2946-2966 [FREE Full text] [doi: 10.1177/1460458220955104] [Medline: 32938275]
92. Miloff A, Carlbring P, Hamilton W, Andersson G, Reuterskiöld L, Lindner P. Measuring alliance toward embodied virtual therapists in the era of automated treatments with the Virtual Therapist Alliance Scale (VTAS): development and psychometric evaluation. J Med Internet Res 2020 Mar 24;22(3):e16660 [FREE Full text] [doi: 10.2196/16660] [Medline: 32207690]
93. El KM, Angelini L, Lalanne D, Abou KO, Mugellini E. Multimodal conversational agent for older adults' behavioral change. In: Proceedings of the 2020 International Conference on Multimodal Interaction. New York, NY: Association for Computing Machinery; 2020 Presented at: 2020 International Conference on Multimodal Interaction; October 25-29, 2020; Utrecht, The Netherlands p. 270-274. [doi: 10.1145/3395035.3425315]
94. Jegundo AL, Dantas C, Quintas J, Dutra J, Almeida AL, Caravau H, et al. Perceived usefulness, satisfaction, ease of use and potential of a virtual companion to support the care provision for older adults. Technologies 2020 Jul 25;8(3):42. [doi: 10.3390/technologies8030042]
95. Philip P, Dupuy L, Auriacombe M, Serre F, de Sevin E, Sauteraud A, et al. Trust and acceptance of a virtual psychiatric interview between embodied conversational agents and outpatients. NPJ Digit Med 2020 Jan 07;3(1):2 [FREE Full text] [doi: 10.1038/s41746-019-0213-y] [Medline: 31934646]
96. Balsa J, Félix I, Cláudio AP, Carmo MB, Silva ICE, Guerreiro A, et al. Usability of an intelligent virtual assistant for promoting behavior change and self-care in older people with type 2 diabetes. J Med Syst 2020 Jun 13;44(7):130. [doi: 10.1007/s10916-020-01583-w] [Medline: 32533367]
97. Arem H, Scott R, Greenberg D, Kaltman R, Lieberman D, Lewin D. Assessing breast cancer survivors' perceptions of using voice-activated technology to address insomnia: feasibility study featuring focus groups and in-depth interviews. JMIR Cancer 2020 May 26;6(1):e15859 [FREE Full text] [doi: 10.2196/15859] [Medline: 32348274]
98. Chen J, Lyell D, Laranjo L, Magrabi F. Effect of speech recognition on problem solving and recall in consumer digital health tasks: controlled laboratory experiment. J Med Internet Res 2020 Jun 01;22(6):e14827 [FREE Full text] [doi: 10.2196/14827] [Medline: 32442129]
99. Ponathil A, Ozkan F, Welch B, Bertrand J, Chalil Madathil K. Family health history collected by virtual conversational agents: an empirical study to investigate the efficacy of this approach. J Genet Couns 2020 Dec 03;29(6):1081-1092. [doi: 10.1002/jgc4.1239] [Medline: 32125052]
100. Bennion MR, Hardy GE, Moore RK, Kellett S, Millings A. Usability, acceptability, and effectiveness of web-based conversational agents to facilitate problem solving in older adults: controlled study. J Med Internet Res 2020 May 27;22(5):e16794 [FREE Full text] [doi: 10.2196/16794] [Medline: 32384055]
101. Rehman UU, Chang DJ, Jung Y, Akhtar U, Razzaq MA, Lee S. Medical instructed real-time assistant for patient with glaucoma and diabetic conditions. Applied Sciences 2020 Mar 25;10(7):2216. [doi: 10.3390/app10072216]
102. Te Pas ME, Rutten WGMM, Bouwman RA, Buise MP. User experience of a chatbot questionnaire versus a regular computer questionnaire: prospective comparative study. JMIR Med Inform 2020 Dec 07;8(12):e21982 [FREE Full text] [doi: 10.2196/21982] [Medline: 33284125]
103. Denecke K, Vaaheesan S, Arulnathan A. A mental health chatbot for regulating emotions (SERMO) - concept and usability test. IEEE Trans Emerg Topics Comput 2021 Jul 1;9(3):1170-1182. [doi: 10.1109/tetc.2020.2974478]
104. Chung K, Cho HY, Park JY. A chatbot for perinatal women's and partners' obstetric and mental health care: development and usability evaluation study. JMIR Med Inform 2021 Mar 03;9(3):e18607 [FREE Full text] [doi: 10.2196/18607] [Medline: 33656442]
105. Iftikhar A, Bond RR, McGilligan V, Leslie SJ, Rjoob K, Knoery C, et al. Comparing single-page, multipage, and conversational digital forms in health care: usability study. JMIR Hum Factors 2021 May 26;8(2):e25787 [FREE Full text] [doi: 10.2196/25787] [Medline: 34037531]

XSL•FO
RenderX

106. da Silva Lima Roque G, Roque de Souza R, Araújo do Nascimento JW, de Campos Filho AS, de Melo Queiroz SR, Ramos Vieira Santos IC. Content validation and usability of a chatbot of guidelines for wound dressing. Int J Med Inform 2021 Jul;151:104473. [doi: 10.1016/j.ijmedinf.2021.104473] [Medline: 33964703]
107. Stara V, Vera B, Bolliger D, Rossi L, Felici E, Di Rosa M, et al. Usability and acceptance of the embodied conversational agent Anne by people with dementia and their caregivers: exploratory study in home environment settings. JMIR Mhealth Uhealth 2021 Jun 25;9(6):e25891 [FREE Full text] [doi: 10.2196/25891] [Medline: 34170256]
108. Yu K, Gorbachev G, Eck U, Pankratz F, Navab N, Roth D. Avatars for teleconsultation: effects of avatar embodiment techniques on user perception in 3D asymmetric telepresence. IEEE Trans Vis Comput Graph 2021 Nov;27(11):4129-4139. [doi: 10.1109/TVCG.2021.3106480] [Medline: 34449373]
109. El HW, El BA, Herbert C, Abdennadher S. Chase away the virus: a character-based chatbot for COVID-19. In: Proceedings of the 9th International Conference on Serious Games and Applications for Health (SeGAH). New York, NY: IEEE; 2021 Presented at: 9th International Conference on Serious Games and Applications for Health (SeGAH); August 4-6, 2021; Online p. 1-8. [doi: 10.1109/SEGAH52098.2021.9551895]
110. Chatzimina M, Papadaki H, Pontikoglou C, Koumakis L, Marias K, Tsiknakis M. Designing a conversational agent for patients with hematologic malignancies: usability and usefulness study. In: Proceedings of the 2021 EMBS International Conference on Biomedical and Health Informatics (BHI). New York, NY: IEEE; 2021 Presented at: 2021 EMBS International Conference on Biomedical and Health Informatics (BHI); July 27-30, 2021; Athens, Greece p. 1-4. [doi: 10.1109/bhi50953.2021.9508587]
111. Nguyen TT, Sim K, Kuen AT, O'Donnell RR, Lim ST, Wang W, et al. Designing AI-based conversational agent for diabetes care in a multilingual context. In: Proceedings of the Twenty-Fifth Pacific Asia Conference on Information Systems. 2021 Presented at: Twenty-Fifth Pacific Asia Conference on Information Systems; July 12-14, 2021; Dubai, UAE p. 1-14.
112. To QG, Green C, Vandelanotte C. Feasibility, usability, and effectiveness of a machine learning-based physical activity chatbot: quasi-experimental study. JMIR Mhealth Uhealth 2021 Nov 26;9(11):e28577 [FREE Full text] [doi: 10.2196/28577] [Medline: 34842552]
113. Roque G, Cavalcanti A, Nascimento J, Souza R, Queiroz S. BotCovid: development and evaluation of a chatbot to combat misinformation about COVID-19 in Brazil. In: Proceedings of the 2021 International Conference on Systems, Man, and Cybernetics (SMC). New York, NY: IEEE; 2021 Presented at: 2021 International Conference on Systems, Man, and Cybernetics (SMC); October 17-20, 2021; Online p. 2506-2511. [doi: 10.1109/smc52423.2021.9658693]
114. Issom D, Hardy-Dessources M, Romana M, Hartvigsen G, Lovis C. Toward a conversational agent to support the self-management of adults and young adults with sickle cell disease: usability and usefulness study. Front Digit Health 2021 Jan 29;3:600333. [doi: 10.3389/FDGTH.2021.600333]
115. Hurmuz MZ, Jansen-Kosterink SM, Beinema T, Fischer K, Op den Akker H, Hermens HJ. Evaluation of a virtual coaching system eHealth intervention: a mixed methods observational cohort study in the Netherlands. Internet Interv 2022 Mar;27:100501 [FREE Full text] [doi: 10.1016/j.invent.2022.100501] [Medline: 35198411]
116. Kramer LL, van Velsen L, Clark JL, Mulder BC, de Vet E. Use and effect of embodied conversational agents for improving eating behavior and decreasing loneliness among community-dwelling older adults: randomized controlled trial. JMIR Form Res 2022 Apr 11;6(4):e33974 [FREE Full text] [doi: 10.2196/33974] [Medline: 35404255]
117. Minutolo A, Damiano E, De Pietro G, Fujita H, Esposito M. A conversational agent for querying Italian Patient Information Leaflets and improving health literacy. Comput Biol Med 2022 Feb;141:105004. [doi: 10.1016/j.compbiomed.2021.105004] [Medline: 34774337]
118. Jansons P, Dalla Via J, Daly R, Fyfe J, Gvozdenko E, Scott D. Delivery of home-based exercise interventions in older adults facilitated by Amazon Alexa: a 12-week feasibility trial. J Nutr Health Aging 2022 Jan;26(1):96-102 [FREE Full text] [doi: 10.1007/s12603-021-1717-0] [Medline: 35067710]
119. Shah J, DePietro B, D'Adamo L, Firebaugh M, Laing O, Fowler LA, et al. Development and usability testing of a chatbot to promote mental health services use among individuals with eating disorders following screening. Int J Eat Disord 2022 Sep 18;55(9):1229-1244 [FREE Full text] [doi: 10.1002/eat.23798] [Medline: 36056648]
120. Afrizal S, Hakiem N, Permanasari A, Albab H, Sanjaya G, Lazuardi L. A user-centered design of natural language processing for maternal monitoring chatbot system. In: Proceedings of the 2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS). New York, NY: IEEE; 2022 Presented at: 2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS); November 16-17, 2022; Jakarta, Indonesia p. 244-248. [doi: 10.1109/ICIMCIS56303.2022.10017517]
121. Six SG, Byrne KA, Aly H, Harris MW. The effect of mental health app customization on depressive symptoms in college students: randomized controlled trial. JMIR Ment Health 2022 Aug 09;9(8):e39516 [FREE Full text] [doi: 10.2196/39516] [Medline: 35943788]
122. Luk TT, Lui JHT, Wang MP. Efficacy, usability, and acceptability of a chatbot for promoting COVID-19 vaccination in unvaccinated or booster-hesitant young adults: pre-post pilot study. J Med Internet Res 2022 Oct 04;24(10):e39063 [FREE Full text] [doi: 10.2196/39063] [Medline: 36179132]
123. Striegl J, Gotthardt M, Loitsch C, Weber G. Investigating the usability of voice assistant-based CBT for age-related depression. In: Proceedings of the International Conference on Computers Helping People With Special Needs. Cham,

XSL•FO
RenderX

Switzerland: Springer; 2022 Presented at: International Conference on Computers Helping People With Special Needs; July 11-15, 2022; Lecco, Italy p. 432-441. [doi: 10.1007/978-3-031-08648-9_50]
124. He Y, Yang L, Zhu X, Wu B, Zhang S, Qian C, et al. Mental health chatbot for young adults with depressive symptoms during the COVID-19 pandemic: single-blind, three-arm randomized controlled trial. J Med Internet Res 2022 Nov 21;24(11):e40719 [FREE Full text] [doi: 10.2196/40719] [Medline: 36355633]
125. Larbi D, Denecke K, Gabarron E. Usability testing of a social media chatbot for increasing physical activity behavior. J Pers Med 2022 May 20;12(5):828 [FREE Full text] [doi: 10.3390/jpm12050828] [Medline: 35629252]
126. Soni H, Ivanova J, Wilczewski H, Bailey A, Ong T, Narma A, et al. Virtual conversational agents versus online forms: patient experience and preferences for health data collection. Front Digit Health 2022 Oct 13;4:954069 [FREE Full text] [doi: 10.3389/fdgth.2022.954069] [Medline: 36310920]
127. Roca S, Almenara M, Gilaberte Y, Gracia-Cazaña T, Morales Callaghan AM, Murciano D, et al. When virtual assistants meet teledermatology: validation of a virtual assistant to improve the quality of life of psoriatic patients. Int J Environ Res Public Health 2022 Nov 05;19(21):14527 [FREE Full text] [doi: 10.3390/ijerph192114527] [Medline: 36361408]
128. Richards D, Miranda Maciel PS, Janssen H. The co-design of an embodied conversational agent to help stroke survivors manage their recovery. Robotics 2023 Aug 22;12(5):120. [doi: 10.3390/robotics12050120]
129. Martinho D, Crista V, Carneiro J, Matsui K, Corchado JM, Marreiros G. Effects of a gamified agent-based system for personalized elderly care: pilot usability study. JMIR Serious Games 2023 Nov 23;11:e48063 [FREE Full text] [doi: 10.2196/48063] [Medline: 37995116]
130. Gingele AJ, Amin H, Vaassen A, Schnur I, Pearl C, Brunner-La Rocca H, et al. Integrating avatar technology into a telemedicine application in heart failure patients: a pilot study. Wien Klin Wochenschr 2023 Dec 02;135(23-24):680-684 [FREE Full text] [doi: 10.1007/s00508-022-02150-8] [Medline: 36732377]
131. Wanberg LJ, Kim A, Vogel RI, Sadak KT, Teoh D. Usability and satisfaction testing of game-based learning avatar-navigated mobile (GLAm), an app for cervical cancer screening: mixed methods study. JMIR Form Res 2023 Aug 08;7:e45541 [FREE Full text] [doi: 10.2196/45541] [Medline: 37552527]
132. Wrightson-Hester A, Anderson G, Dunstan J, McEvoy PM, Sutton CJ, Myers B, et al. An artificial therapist (Manage Your Life Online) to support the mental health of youth: co-design and case series. JMIR Hum Factors 2023 Jul 21;10:e46849 [FREE Full text] [doi: 10.2196/46849] [Medline: 37477969]
133. Griffin AC, Khairat S, Bailey SC, Chung AE. A chatbot for hypertension self-management support: user-centered design, development, and usability testing. JAMIA Open 2023 Oct;6(3):ooad073 [FREE Full text] [doi: 10.1093/jamiaopen/ooad073] [Medline: 37693367]
134. Powell L, Nour R, Sleibi R, Al Suwaidi H, Zary N. Democratizing the development of chatbots to improve public health: feasibility study of COVID-19 misinformation. JMIR Hum Factors 2023 Dec 28;10:e43120 [FREE Full text] [doi: 10.2196/43120] [Medline: 37290040]
135. Islam A, Chaudhry BM. Design validation of a relational agent by COVID-19 patients: mixed methods study. JMIR Hum Factors 2023 Jun 08;10:e42740 [FREE Full text] [doi: 10.2196/42740] [Medline: 36350760]
136. Babington-Ashaye A, de Moerloose P, Diop S, Geissbuhler A. Design, development and usability of an educational AI chatbot for people with haemophilia in Senegal. Haemophilia 2023 Jul 22;29(4):1063-1073. [doi: 10.1111/hae.14815] [Medline: 37347648]
137. Biro J, Linder C, Neyens D. The effects of a health care chatbot's complexity and persona on user trust, perceived usability, and effectiveness: mixed methods study. JMIR Hum Factors 2023 Feb 01;10:e41017 [FREE Full text] [doi: 10.2196/41017] [Medline: 36724004]
138. Tawfik E, Ghallab E, Moustafa A. A nurse versus a chatbot the effect of an empowerment program on chemotherapy-related side effects and the self-care behaviors of women living with breast cancer: a randomized controlled trial. BMC Nurs 2023 Apr 06;22(1):102 [FREE Full text] [doi: 10.1186/s12912-023-01243-7] [Medline: 37024875]
139. Wilczewski H, Soni H, Ivanova J, Ong T, Barrera JF, Bunnell BE, et al. Older adults' experience with virtual conversational agents for health data collection. Front Digit Health 2023;5:1125926 [FREE Full text] [doi: 10.3389/fdgth.2023.1125926] [Medline: 37006821]
140. Bruijnes M, Kesteloo M, Brinkman W. Reducing social diabetes distress with a conversational agent support system: a three-week technology feasibility evaluation. Front Digit Health 2023 Jun 13;5:1149374 [FREE Full text] [doi: 10.3389/fdgth.2023.1149374] [Medline: 37383944]
141. Bevilacqua R, Felici E, Cucchieri G, Amabili G, Margaritini A, Franceschetti C, et al. Results of the Italian RESILIEN-T pilot study: a mobile health tool to support older people with mild cognitive impairment. J Clin Med 2023 Sep 22;12(19):6129 [FREE Full text] [doi: 10.3390/jcm12196129] [Medline: 37834775]
142. de SCA, do NJ, Roque G, de SR, de MQS, Correa J. A virtual assistant in vaccine pharmacovigilance: content and usability validation. CIN: Computers, Informatics, Nursing 2023;41(7):482-490. [doi: 10.1097/cin.0000000000000978]
143. Chua JYX, Choolani M, Chee CYI, Yi H, Chan YH, Lalor JG, et al. 'Parentbot - a digital healthcare assistant (PDA)': a mobile application-based perinatal intervention for parents: development study. Patient Educ Couns 2023 Sep;114:107805. [doi: 10.1016/j.pec.2023.107805] [Medline: 37245443]

XSL•FO
RenderX

144. Chen N, Huang C, Fan C, Lu L, Lin F, Liao J, et al. User evaluation of a chat-based instant messaging support health education program for patients with chronic kidney disease: preliminary findings of a formative study. JMIR Form Res 2023 Sep 19;7:e45484 [FREE Full text] [doi: 10.2196/45484] [Medline: 37725429]
145. Albino de Queiroz D, Silva Passarello R, Veloso de Moura Fé V, Rossini A, Folchini da Silveira E, Aparecida Isquierdo Fonseca de Queiroz E, et al. A wearable chatbot-based model for monitoring colorectal cancer patients in the active phase of treatment. Healthcare Analytics 2023 Dec;4:100257. [doi: 10.1016/j.health.2023.100257]
146. Alonso-Mencía J, Castro-Rodríguez M, Herrero-Pinilla B, Alonso-Weber JM, Rodríguez-Mañas L, Pérez-Rodríguez R. ADELA: a conversational virtual assistant to prevent delirium in hospitalized older persons. J Supercomput 2023 May 09;79(15):17670-17690. [doi: 10.1007/s11227-023-05352-7]
147. Larkin C, Tulu B, Djamasbi S, Garner R, Varzgani F, Siddique M, et al. Comparing the acceptability and quality of intervention modalities for suicidality in the emergency department: randomized feasibility trial. JMIR Ment Health 2023 Oct 24;10:e49783 [FREE Full text] [doi: 10.2196/49783] [Medline: 37874619]
148. Morato J, do Nascimento JWA, Roque G, de Souza RR, Santos ICRV. Development, validation, and usability of the chatbot ESTOMABOT to promote self-care of people with intestinal ostomy. Comput Inform Nurs 2023 Dec 01;41(12):1037-1045. [doi: 10.1097/CIN.0000000000001075] [Medline: 37725781]
149. Sanchez-Ortuno MM, Pecune F, Coelho J, Micoulaud-Franchi JA, Salles N, Auriacombe M, et al. Predictors of users' adherence to a fully automated digital intervention to manage insomnia complaints. J Am Med Inform Assoc 2023 Nov 17;30(12):1934-1942. [doi: 10.1093/jamia/ocad163] [Medline: 37672004]
150. Larkin C, Djamasbi S, Boudreaux ED, Varzgani F, Garner R, Siddique M, et al. ReachCare mobile apps for patients experiencing suicidality in the emergency department: development and usability testing using mixed methods. JMIR Form Res 2023 Jan 27;7:e41422 [FREE Full text] [doi: 10.2196/41422] [Medline: 36705961]
151. Hocking J, Maeder A, Powers D, Perimal-Lewis L, Dodd B, Lange B. Mixed methods, single case design, feasibility trial of a motivational conversational agent for rehabilitation for adults with traumatic brain injury. Clin Rehabil 2024 Mar 06;38(3):322-336 [FREE Full text] [doi: 10.1177/02692155231216615] [Medline: 38058144]
152. Ashrafi N, Vona F, Hinzmann S, Graf P, Harnisch P, Voigt-Antons J. Single vs dual: influence of the number of displays on user experience within virtually embodied conversational systems. In: Proceedings of the 30th ACM Symposium on Virtual Reality Software and Technology. New York, NY: Association for Computing Machinery; 2024 Presented at: 30th ACM Symposium on Virtual Reality Software and Technology; October 9-11, 2024; Trier, Germany p. 1-2. [doi: 10.1145/3641825.3689700]
153. Thunström AO, Carlsen HK, Ali L, Larson T, Hellström A, Steingrimsson S. Usability comparison among healthy participants of an anthropomorphic digital human and a text-based chatbot as a responder to questions on mental health: randomized controlled trial. JMIR Hum Factors 2024 Apr 29;11:e54581 [FREE Full text] [doi: 10.2196/54581] [Medline: 38683664]
154. Martins A, Velez Lapão L, Nunes IL, Paula Giordano A, Semedo H, Vital C, et al. A conversational agent for enhanced self-management after cardiothoracic surgery. Int J Med Inform 2024 Dec;192:105640 [FREE Full text] [doi: 10.1016/j.ijmedinf.2024.105640] [Medline: 39321492]
155. Striegl J, Richter JW, Grossmann L, Bråstad B, Gotthardt M, Rück C, et al. Deep learning-based dimensional emotion recognition for conversational agent-based cognitive behavioral therapy. PeerJ Comput Sci 2024;10:e2104. [doi: 10.7717/peerj-cs.2104] [Medline: 38983201]
156. Schlieter H, Gand K, Weimann TG, Sandner E, Kreiner K, Thoma S, et al. Designing virtual coaching solutions. Bus Inf Syst Eng 2024 May 29;66(3):377-400. [doi: 10.1007/s12599-024-00871-w]
157. Alves B, Mota PR, Sineiro D, Carmo R, Santos P, Macedo P, et al. MoveONParkinson: developing a personalized motivational solution for Parkinson's disease management. Front Public Health 2024 Aug 19;12:1420171 [FREE Full text] [doi: 10.3389/fpubh.2024.1420171] [Medline: 39224558]
158. Ali SH, Rahman F, Kuwar A, Khanna T, Nayak A, Sharma P, et al. Rapid, tailored dietary and health education through a social media chatbot microintervention: development and usability study with practical recommendations. JMIR Form Res 2024 Dec 09;8:e52032 [FREE Full text] [doi: 10.2196/52032] [Medline: 39652870]
159. Devittori G, Akeddar M, Retevoi A, Schneider F, Cvetkova V, Dinacci D. Towards RehabCoach: design and preliminary evaluation of a conversational agent sup-porting unsupervised therapy after stroke. In: Proceedings of the 10th RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob). New York, NY: IEEE; 2024 Presented at: 10th RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob); September 1-4, 2024; Heidelberg, Germany p. 569-574. [doi: 10.1109/BioRob60516.2024.10719881]
160. Nkabane-Nkholongo E, Mpata-Mokgatle M, Jack BW, Julce C, Bickmore T. Usability and acceptability of a conversational agent health education app (Nthabi) for young women in Lesotho: quantitative study. JMIR Hum Factors 2024 Mar 12;11:e52048 [FREE Full text] [doi: 10.2196/52048] [Medline: 38470460]
161. Nguyen MH, Sedoc J, Taylor CO. Usability, engagement, and report usefulness of chatbot-based family health history data collection: mixed methods analysis. J Med Internet Res 2024 Sep 30;26:e55164 [FREE Full text] [doi: 10.2196/55164] [Medline: 39348188]

XSL•FO
RenderX

162. Reilly ED, Kelly MM, Grigorian HL, Waring ME, Quigley KS, Hogan TP, et al. Virtual coach-guided online acceptance and commitment therapy for chronic pain: pilot feasibility randomized controlled trial. JMIR Form Res 2024 Nov 08;8:e56437 [FREE Full text] [doi: 10.2196/56437] [Medline: 39514264]
163. Segear S, Chheang V, Baron L, Li J, Kim K, Barmaki RL. Visual feedback and guided balance training in an immersive virtual reality environment for lower extremity rehabilitation. Comput Graph 2024 Apr;119:103880. [doi: 10.1016/j.cag.2024.01.007] [Medline: 38645661]
164. Morita Y, Wong Z. Designing and adopting a video-based LINE chatbot system for endoscopy explanation in a real-world hospital: a mixed method approach. In: Proceedings of the 2024 Australasian Computer Science Week. New York, NY: Association for Computing Machinery; 2024 Presented at: 2024 Australasian Computer Science Week; January 29 to February 2, 2024; Sydney, Australia p. 87-91. [doi: 10.1145/3641142.3641161]
165. Song S, Seo Y, Hwang S, Kim H, Kim J. Digital phenotyping of geriatric depression using a community-based digital mental health monitoring platform for socially vulnerable older adults and their community caregivers: 6-week living lab single-arm pilot study. JMIR Mhealth Uhealth 2024 Jun 17;12:e55842 [FREE Full text] [doi: 10.2196/55842] [Medline: 38885033]
166. Wan Ab.Rahman WN, Abdul Hamid NM. Rule-based chatbot for early self-depression indication: a promising approach. Int J Inform Visualization 2024 Nov 24;8(3-2):1625. [doi: 10.62527/joiv.8.3-2.1628]
167. De la Puente G, Silva A, Felix R. Development of a chatbot powered by artificial intelligence to diagnose and improve stress and anxiety levels in university students. In: Proceedings of the XXXI International Conference on Electronics, Electrical Engineering and Computing (IN-TERCON). New York, NY: IEEE; 2024 Presented at: XXXI International Conference on Electronics, Electrical Engineering and Computing (IN-TERCON); November 6-8, 2024; Lima, Peru p. 1-8. [doi: 10.1109/intercon63140.2024.10833503]
168. Abilkaiyrkyzy A, Laamarti F, Hamdi M, Saddik AE. Dialogue system for early mental illness detection: toward a digital twin solution. IEEE Access 2024;12:2007-2024. [doi: 10.1109/access.2023.3348783]
169. Menard C, Coelho J, De SE, Levavasseur Y, Philip P, Pecune F. Quantitative and qualitative evaluation of the acceptability of an embodied conversational agent for insomnia screening. In: Proceedings of the 24th ACM International Conference on Intelligent Virtual Agents. New York, NY: Association for Computing Machinery; 2024 Presented at: 24th ACM International Conference on Intelligent Virtual Agents; September 16-19, 2024; Glasgow, Scotland, UK p. 1-4. [doi: 10.1145/3652988.3673940]
170. Papiernik P, Dzula S, Zimanyi M, Millgate E, Bouazzaoui M, Buttimer J, et al. Acceptability of a conversational agent-led digital program for anxiety: mixed methods study of user perspectives. JMIR Hum Factors 2025 Nov 04;12:e76377 [FREE Full text] [doi: 10.2196/76377] [Medline: 40824528]
171. Shi J, Jain R, Chi S, Doh H, Chi H, Quinn A, et al. Caring-AI: towards authoring context-aware augmented reality instruction through generative artificial intelligence. In: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2025 Presented at: 2025 CHI Conference on Human Factors in Computing Systems; April 26 to May 1, 2025; Yokohama, Japan p. 1-23. [doi: 10.1145/3706598.3713348]
172. Mendes C, Pereira R, Frazao L, Ribeiro J, Rodrigues N, Costa N. Emotionally intelligent customizable conversational agent for elderly care: development and impact of Chatto. In: Proceedings of the 11th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion. New York, NY: Association for Computing Machinery; 2025 Presented at: 11th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion; November 13-15, 2024; Abu Dhabi, UAE p. 208-214. [doi: 10.1145/3696593.3696619]
173. Six S, Schlesener E, Hill V, Babu SV, Byrne K. Impact of conversational and animation features of a mental health app virtual agent on depressive symptoms and user experience among college students: randomized controlled trial. JMIR Ment Health 2025 Apr 11;12:e67381-e67381 [FREE Full text] [doi: 10.2196/67381] [Medline: 40215483]
174. Lazem H, Harris D, Hall A, Mansoubi M, Pontes RG, de Mello Monteiro CB, et al. Validity, safety, usability, and user experience of virtual reality gamified home-based exercises in stroke. Clin Rehabil 2025 Nov 02;39(11):1527-1540 [FREE Full text] [doi: 10.1177/02692155251371435] [Medline: 40891724]
175. Del Pino R, de Echevarría AO, Díez-Cirarda M, Ustarroz-Aguirre I, Caprino M, Liu J, et al. Virtual coach and telerehabilitation for Parkinson´s disease patients: vCare system. J Public Health (Berl) 2023 Nov 13;33(7):1583-1596. [doi: 10.1007/s10389-023-02082-1]
176. Haider SA, Prabha S, Gomez-Cabello CA, Genovese A, Collaco B, Wood N, et al. Artificial intelligence physician avatars for patient education: a pilot study. J Clin Med 2025 Dec 04;14(23):8595 [FREE Full text] [doi: 10.3390/jcm14238595] [Medline: 41375898]
177. Düzgün MV, İşler A, Kazan MS. Effect of avatar-based education program in hydrocephalus on ventriculoperitoneal shunt complications and parents' knowledge and care skills: multicenter randomized controlled trials. Pediatr Neurol 2025 Aug;169:131-139. [doi: 10.1016/j.pediatrneurol.2025.05.019] [Medline: 40505427]
178. Ashrafi N, Vona F, Hinzmann S, Henning J, Vergari M, Warsinke M. Size matters: the impact of avatar size on user experience in healthcare applications. In: Proceedings of the 17th International Conference on Quality of Multimedia

XSL•FO
RenderX

Experience (QoMEX). New York, NY: IEEE; 2025 Presented at: 17th International Conference on Quality of Multimedia Experience (QoMEX); September 29 to October 3, 2025; Madrid, Spain p. 1-7. [doi: 10.1109/QoMEX65720.2025.11219980]
179. Pinto G, De SJ, Rosa R, Rodriguez D. Embodied multimodal chatbot for mental health support in web-based 3D environments. In: Proceedings of the 2025 International Conference on Software, Telecommunications and Computer Networks (SoftCOM). New York, NY: IEEE; 2025 Presented at: 2025 International Conference on Software, Telecommunications and Computer Networks (SoftCOM); September 18-20, 2025; Split, Croatia p. 1-6. [doi: 10.23919/softcom66362.2025.11197343]
180. Hopman K, Richards D, Norberg MM. A staged approach to the development of an embodied conversational agent to support wellbeing after injury. Behaviour & Information Technology 2025 Mar 20;45(8):1490-1514. [doi: 10.1080/0144929x.2025.2477757]
181. Elliott TC, Yang Y, Knibbe J, Henry JD, Baghaei N. Avatar customization and embodiment in virtual reality self-compassion therapy for depressive symptoms: three-part mixed methods study. JMIR Form Res 2025 Oct 02;9:e71004-e71004 [FREE Full text] [doi: 10.2196/71004] [Medline: 41037663]
182. Hernandez R, Nisar H, Kesavadas, McGee MC, Gerstner GJ, Martinez A, et al. Assessing safety and feasibility of virtual reality intervention in patients with lung cancer: a pilot study. Support Care Cancer 2025 Mar 25;33(4):318. [doi: 10.1007/s00520-025-09338-4] [Medline: 40131484]
183. Groninger H, Arem H, Ayangma L, Gong L, Zhou E, Greenberg D. Development of a voice-activated virtual assistant to improve insomnia among young adult cancer survivors: mixed methods feasibility and acceptability study. JMIR Form Res 2025 Mar 10;9:e64869 [FREE Full text] [doi: 10.2196/64869] [Medline: 40063947]
184. Sharp G, Dwyer B, Randhawa A, McGrath I, Hu H. The effectiveness of a chatbot single-session intervention for people on waitlists for eating disorder treatment: randomized controlled trial. J Med Internet Res 2025 May 21;27:e70874 [FREE Full text] [doi: 10.2196/70874] [Medline: 40397933]
185. Fetrati H, Chan G, Orji R. Leveraging generative and rule-based models for persuasive STI education: a multi-chatbot mobile application. In: Proceedings of the 7th ACM Conference on Conversational User Interfaces. New York, NY: Association for Computing Machinery; 2025 Presented at: 7th ACM Conference on Conversational User Interfaces; July 8-10, 2025; Waterloo, ON, Canada p. 1-9. [doi: 10.1145/3719160.3737619]
186. Isaacs K, Shifflett A, Patel K, Karpisek L, Cui Y, Lawental M, et al. Women empowered to connect with addiction resources and engage in evidence-based treatment (WE-CARE)-an mHealth application for the universal screening of alcohol, substance use, depression, and anxiety: usability and feasibility study. JMIR Form Res 2025 Feb 07;9:e62915 [FREE Full text] [doi: 10.2196/62915] [Medline: 39918861]
187. Sivakumar B, Ricupero M, Mahajan A, Jefferson K, Wenger J, Code J, et al. A mobile app intervention to support nutrition education for heart failure management: co-design, development and user-testing. BMC Nutr 2025 Jul 12;11(1):139. [doi: 10.1186/s40795-025-01117-y] [Medline: 40652288]
188. Zhou D, Globa A. Social connectedness in older adults: the role of MR and haptic interaction in remote digital storytelling. In: Proceedings of the 37th Australian Conference on Human-Computer Interaction. New York, NY: Association for Computing Machinery; 2025 Presented at: 37th Australian Conference on Human-Computer Interaction; November 29 to December 3, 2025; Sydney, Australia p. 807-818. [doi: 10.1145/3764687.3769932]
189. Kuswik H, Poletaykina A, Rittmann A, Kruse L, Steinicke F. Instruct me! Comparing virtual agents and static picture instructions in therapeutic virtual reality exercises. In: Proceedings of the 2025 ACM Symposium on Spatial User Interaction. New York, NY: Association for Computing Machinery; 2025 Presented at: 2025 ACM Symposium on Spatial User Interaction; November 10-11, 2025; Montreal, QC, Canada p. 1-10. [doi: 10.1145/3694907.3765934]
190. Carneiro P, Águia I, Palricas D, Rocha A, Teixeira A, Almeida N. Kitchen assistant for active ageing at home. In: Proceedings of the 11th International Conference on Software Development Technologies for Enhancing Accessibility Fighting Info-exclusion. New York, NY: Ass; 2025 Presented at: 11th International Conference on Software Development Technologies for Enhancing Accessibility Fighting Info-exclusion; November 13-15, 2024; Abu Dhabi, UAE p. 193-202. [doi: 10.1145/3696593.3696624]
191. Rosteck N, Striegl J, Loitsch C. Bridging the treatment gap: a novel LLM-driven system for scalable initial patient assessments in mental healthcare. In: Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2025 Presented at: CHI Conference on Human Factors in Computing Systems; April 26 to May 1, 2025; Yokohama, Japan p. 1-8. [doi: 10.1145/3706599.3720043]
192. Tang L, Weng C, Hou S, Savas L, Neumann A, Moore N. Constructing a conversation-first dialogue model for comprehensive, evidence-based counseling for HPV vaccination for young adults. In: Proceedings of the International Conference on Human-Computer Interaction. Cham, Switzerland: Springer; 2025 Presented at: International Conference on Human-Computer Interaction; June 22-27, 2025; Gothenburg, Sweden p. 72-84. [doi: 10.1007/978-3-031-93502-2_5]
193. Ejem D, Bakitas M, Durant RW, Parker TN, Oppong KD, Esterson J, et al. Exploring the acceptability and feasibility of a self-directed approach to identifying health priorities in a sample of southern older African American Adults with multiple chronic conditions. J Racial Ethn Health Disparities 2026 Aug 23;13(4):2940-2957. [doi: 10.1007/s40615-025-02469-8] [Medline: 40410637]

XSL•FO
RenderX

194. Begum TUS, Devi RR, Haridas D, Bacanin N, Jovicic MD, Nikolic B. My diabetes care: an AI-based mobile app with conversational agent for type 2 diabetes self-management. Sci Rep 2025 Nov 17;15(1):40228 [FREE Full text] [doi: 10.1038/s41598-025-24172-8] [Medline: 41249394]
195. Suffoletto B, Clark DB, Lee C, Mason M, Schultz J, Szeto I, et al. Development and preliminary testing of a secure large language model-based chatbot for brief alcohol counseling in young adults. Drug Alcohol Depend 2025 Jul 01;272:112697. [doi: 10.1016/j.drugalcdep.2025.112697] [Medline: 40334327]
196. Khairat S, Geracitano J, Mendis K. Acceptability of academic large language model for patients seeking health information. Studies in Health Technology and Informatics 2025;328:31. [doi: 10.3233/shti250667]
197. Leung YW, So J, Sidhu A, Asokan V, Gancarz M, Gajjar VB, et al. The extent to which artificial intelligence can help fulfill metastatic breast cancer patient healthcare needs: a mixed-methods study. Curr Oncol 2025 Mar 02;32(3):145 [FREE Full text] [doi: 10.3390/curroncol32030145] [Medline: 40136349]
198. Parry M, Huang T, Clarke H, Bjørnnes AK, Harvey P, Parente L, et al. Development and systematic evaluation of a progressive web application for women with cardiac pain: usability study. JMIR Hum Factors 2025 Apr 17;12:e57583 [FREE Full text] [doi: 10.2196/57583] [Medline: 40245401]
199. Daniels K, Vonck S, Robijns J, Quadflieg K, Bergs J, Spooren A, et al. Exploring the feasibility of a 5-week mHealth intervention to enhance physical activity and an active, healthy lifestyle in community-dwelling older adults: mixed methods study. JMIR Aging 2025 Jan 27;8:e63348 [FREE Full text] [doi: 10.2196/63348] [Medline: 39869906]
200. Vandelanotte C, Maher C, Hodgetts D, Imam T, Rashid M, To Q, et al. Impact of iterative development and beta-testing on the usability and acceptability of a novel just-in-time adaptive digital physical activity intervention. J Phys Act Health 2025 Oct 01;22(10):1315-1321 [FREE Full text] [doi: 10.1123/jpah.2025-0107] [Medline: 40467054]
201. Wolff D, Kupka T, Reichert C, Ammon N, Oeltze-Jafra S, Vajen B. Personalized support in hereditary breast and ovarian cancer after genetic counseling by the chatbot-based GENIE mobile app: proof-of-concept wizard of oz study. JMIR Form Res 2025 Jun 05;9:e69115-e69115 [FREE Full text] [doi: 10.2196/69115] [Medline: 40471748]
202. Chisty S, Miami A, Noor J. Smart Sheba: enhancing elderly user experience with LLM-enabled chatbots and user-centered design. In: Proceedings of the 13th International Conference on Information & Communication Technologies and Development. New York, NY: Association for Computing Machinery; 2025 Presented at: 13th International Conference on Information & Communication Technologies and Devel-opment; December 9-11, 2024; Nairobi, Kenya p. 69-93. [doi: 10.1145/3700794.3700803]
203. Antia SE, Ugwu CN, Ghodka V, Chori BS, Nazir MS, Odili CA, et al. Healthy Heart Assistant, a WhatsApp-based generative pretrained transformer technology, for self-care in hypertensive patients. Mayo Clinic Proceedings: Digital Health 2025 Sep;3(3):100243. [doi: 10.1016/j.mcpdig.2025.100243]
204. Vowels LM, Vowels MJ, Sweeney SK, Hatch SG, Darwiche J. The efficacy, feasibility, and technical outcomes of a GPT-4o-based chatbot Amanda for relationship support: a randomized controlled trial. PLOS Ment Health 2025 Sep 24;2(9):e0000411 [FREE Full text] [doi: 10.1371/journal.pmen.0000411] [Medline: 41662052]
205. Yslado-Méndez R, Escobar-Agreda S, Villarreal-Zegarra D, Trejo Flores WM, Sánchez-Broncano JD, Vilela-Estrada AL, et al. Effectiveness, usability, and satisfaction of a self-administered digital intervention for reducing depression, anxiety, and stress in a university community in the Andean region of Peru: randomized controlled trial. JMIR Form Res 2025 Oct 15;9:e71465-e71465 [FREE Full text] [doi: 10.2196/71465] [Medline: 41092084]
206. Chanteclair A, Lartigau M, Salles N, Coelho J, Pécune F, Philip P, et al. Assessing the acceptability of a sleep-targeted digital intervention among geriatric inpatients: a preliminary study. Digital Health 2025 Jan 29;11:1-10. [doi: 10.1177/20552076241293935]
207. Borsci S, Schmettow M, Malizia A, Chamberlain A, van der Velde F. A confirmatory factorial analysis of the Chatbot Usability Scale: a multilanguage validation. Pers Ubiquit Comput 2022 Aug 04;27(2):317-330. [doi: 10.1007/s00779-022-01690-0]
208. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. International Journal of Human-Computer Interaction 1995 Jan;7(1):57-78. [doi: 10.1080/10447319509526110]
209. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
210. Lewis J, Utesch B, Maher D. UMUX-LITE: when there's no time for the SUS. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 27 to May 2, 2013; Paris, France p. 2099-2102. [doi: 10.1145/2470654.2481287]
211. Lund A. Measuring usability with the USE questionnaire. Usability Interface 2001;8(2):3-6 [FREE Full text]
212. Rauschenberger M, Schrepp M, Perez-Cota M, Olschner S, Thomaschewski J. Efficient measurement of the user experience of interactive products. how to use the User Experience Questionnaire (UEQ). Example: Spanish language version. IJIMAI 2013;2(1):39-45. [doi: 10.9781/ijimai.2013.215]
213. Mitchell-Box K, Braun KL. Fathers' thoughts on breastfeeding and implications for a theory-based intervention. J Obstet Gynecol Neonatal Nurs 2012;41(6):E41-E50. [doi: 10.1111/j.1552-6909.2012.01399.x] [Medline: 22861175]

XSL•FO
RenderX

214. Tariman JD, Berry DL, Halpenny B, Wolpin S, Schepp K. Validation and testing of the Acceptability E-scale for web-based patient-reported outcomes in cancer care. Appl Nurs Res 2011 Feb;24(1):53-58 [FREE Full text] [doi: 10.1016/j.apnr.2009.04.003] [Medline: 20974066]
215. Galavi Z, Montazeri M, Khajouei R. Which criteria are important in usability evaluation of mHealth applications: an umbrella review. BMC Med Inform Decis Mak 2024 Nov 29;24(1):365 [FREE Full text] [doi: 10.1186/s12911-024-02738-2] [Medline: 39614283]
216. Griffin A, Xing Z, Khairat S, Wang Y, Bailey S, Arguello J, et al. Conversational agents for chronic disease self-management: a systematic review. In: Proceedings of the AMIA Annual Symposium. Washington, DC: AMIA; 2021 Presented at: AMIA Annual Symposium; October 30 to November 3, 2021; San Diego, CA p. 504-513.
217. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
218. Gaffney H, Mansell W, Tai S. Conversational agents in the treatment of mental health problems: mixed-method systematic review. JMIR Ment Health 2019 Oct 18;6(10):e14166 [FREE Full text] [doi: 10.2196/14166] [Medline: 31628789]
219. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. J Med Internet Res 2020 Jul 13;22(7):e16021 [FREE Full text] [doi: 10.2196/16021] [Medline: 32673216]
220. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]
221. Singh B, Olds T, Brinsley J, Dumuid D, Virgara R, Matricciani L, et al. Systematic review and meta-analysis of the effectiveness of chatbots on lifestyle behaviours. NPJ Digit Med 2023 Jun 23;6(1):118 [FREE Full text] [doi: 10.1038/s41746-023-00856-1] [Medline: 37353578]
222. Geoghegan L, Scarborough A, Wormald JCR, Harrison CJ, Collins D, Gardiner M, et al. Automated conversational agents for post-intervention follow-up: a systematic review. BJS Open 2021 Jul 06;5(4):zrab070 [FREE Full text] [doi: 10.1093/bjsopen/zrab070] [Medline: 34323916]
223. Alruwaili MM, Shaban M, Elsayed Ramadan OM. Digital health interventions for promoting healthy aging: a systematic review of adoption patterns, efficacy, and user experience. Sustainability 2023 Dec 02;15(23):16503. [doi: 10.3390/su152316503]
224. Liang M, Cui J, Fan X, Zhang J, Liu X, Liu D. The effect of digital health interventions in older adults with frailty: a systematic review and meta-analysis. Int J Nurs Stud Adv 2026 Jun;10:100470 [FREE Full text] [doi: 10.1016/j.ijnsa.2025.100470] [Medline: 41551860]
225. Ahmed MI, Spooner B, Isherwood J, Lane M, Orrock E, Dennison A. A Systematic Review of the Barriers to the Implementation of Artificial Intelligence in Healthcare. Cureus 2023 Oct;15(10):e46454 [FREE Full text] [doi: 10.7759/cureus.46454] [Medline: 37927664]
226. Rahamtalla BM, Medani IE, Abdelhag ME, Eltigani SA, Rajan SK, Falgy E, et al. The AI-powered healthcare ecosystem: bridging the chasm between technical validation and systemic integration—a systematic review. Future Internet 2025 Nov 29;17(12):550. [doi: 10.3390/fi17120550]
227. Artsi Y, Sorin V, Glicksberg BS, Korfiatis P, Nadkarni GN, Klang E. Large language models in real-world clinical workflows: a systematic review of applications and implementation. Front Digit Health 2025 Sep 30;7:1659134 [FREE Full text] [doi: 10.3389/fdgth.2025.1659134] [Medline: 41098649]
228. Lim PC, Lim YL, Rajah R, Zainal H. Usability questionnaire for standalone or interactive mobile health applications: a systematic review. BMC Digit Health 2025 Apr 01;3(1):11. [doi: 10.1186/s44247-025-00150-y]
229. Sharma S, Kumar BA. A systematic review of user-based usability testing practices in self-care mHealth apps. Digit Health 2025 Aug 29;11:20552076251374184 [FREE Full text] [doi: 10.1177/20552076251374184] [Medline: 40893169]
230. Wang Q, Liu J, Zhou L, Tian J, Chen X, Zhang W, et al. Usability evaluation of mHealth apps for elderly individuals: a scoping review. BMC Med Inform Decis Mak 2022 Dec 02;22(1):317 [FREE Full text] [doi: 10.1186/s12911-022-02064-5] [Medline: 36461017]
231. Wechsung I, Weiss B, Kühnel C, Ehrenbrink P, Möller S. Development and validation of the conversational agents scale (CAS). In: Proceedings of the INTERSPEECH. Grenoble, France: ISCA; 2013 Presented at: INTERSPEECH; August 25-29, 2013; Lyon, France p. 1106-1110. [doi: 10.21437/interspeech.2013-298]
232. Israfilzade K. Conversational marketing as a framework for interaction with the customer: development validation of the Conversational Agent's Usage Scale. Journal of Life Economics 2021 Oct 31;8(4):533-546. [doi: 10.15637/jlecon.8.4.12]
233. Guerino G, Silva W, Coleti T, Valentim N. Assessing a technology for usability and user experience evaluation of conversational systems: an exploratory study. In: Proceedings of the ICEIS 2021. Setúbal, Portugal: Scitepress; 2021 Presented at: ICEIS 2021; Online; April 26-28, 2021 p. 463-473. [doi: 10.5220/0010450204630473]
234. Faruk LID, Pal D, Funilkul S, Perumal T, Mongkolnam P. Introducing CASUX: a standardized scale for measuring the user experience of artificial intelligence based conversational agents. International Journal of Human–Computer Interaction 2024 Jun 03;41(9):5274-5298. [doi: 10.1080/10447318.2024.2359206]

XSL•FO
RenderX

Abbreviations

AES: Acceptability E-Scale
CAUSS: Critical Assessment of Usability Studies Scale
CSUQ: Computer System Usability Questionnaire
CUQ: Chatbot Usability Questionnaire
HCA: health care conversational agent
MAUQ: mHealth App Usability Questionnaire
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSSUQ: Post-Study System Usability Questionnaire
SUS: System Usability Scale
UEQ: User Experience Questionnaire
UEQ-S: User Experience Questionnaire—Short Version
UMUX-Lite: Usability Metric for User Experience—Lite Version
USE: Usefulness, Satisfaction, and Ease of Use Questionnaire

Edited by A Parush; submitted 29.May.2026; peer-reviewed by M Chatzimina, L Kettle; comments to author 30.Jun.2026; revised version received 18.Jul.2026; accepted 20.Jul.2026; published 30.Jul.2026.

Please cite as:
Pavão J, Bastardo R, Silva AG, Rocha NP
Usability and User Experience Assessment of Health Care Conversational Agents Using Validated Subjective Instruments: Systematic Review and Comparative Analysis
JMIR Hum Factors 2026;13:e102911
URL: https://humanfactors.jmir.org/2026/1/e102911
doi:10.2196/102911
PMID:42530974

©João Pavão, Rute Bastardo, Anabela Gonçalves Silva, Nelson Pacheco Rocha. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 30.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effect of Pharmacy Student Peer Supervision on the Accuracy of Admission Medication Reconciliation: Prospective Pre-Post Observational Study

Basma Ammor[1], PharmaD; Morgane Masse[1,2], PharmaD, PhD; Anne Toulemonde[1], PharmaD; Laurine Cadart[3], MPH; Jean-Baptiste Beuscart[4], MD, PhD; Marc Lambert[5], MD, PhD; Pascal Odou[1,2], PharmaD, PhD; Bertrand Décaudin[1,2], PharmaD, PhD

[1]CHU Lille, Institut de Pharmacie, Lille, France

[2]Univ Lille, CHU Lille, ULR 7365-GRITA-Groupe de Recherche sur les formes Injectables et les Technologies Associées, Lille, France

[3]CHU Lille, Service de Biostatistiques, Lille, France

[4]Univ Lille, CHU Lille, ULR 2694 - METRICS : Évaluation des Technologies de Santé et des Pratiques Médicales, Lille, France

[5]Service de médecine interne, centre national de référence maladies systémiques et auto-immunes rares (sclérodermie systémique), CHU de Lille, UFR de médecine, Université de Lille, Lille, France

Corresponding Author:
Bertrand Décaudin, PharmaD, PhD
CHU Lille, Institut de Pharmacie, Lille, France

Abstract

Background: Although medication reconciliation is known to reduce the frequency of medication errors, its practical implementation can be challenging in several respects. In our institution, pharmacy students perform medication reconciliations at admission under the supervision of a pharmacist or pharmacy resident.

Objective: The objective of the study was to evaluate the impact of peer supervision (ie, the supervision by a pharmacy student of a medication reconciliation performed by another pharmacy student) on the accuracy and efficiency of admission medication reconciliations.

Methods: A prospective, single-center, observational study was conducted in 2 clinical departments at Lille University Medical Center (Lille, France). Initially, organizational procedures were defined and a checklist for reconciliation supervision was developed. A baseline (reference) period without peer supervision was compared with an implementation period with peer supervision.

Results: A total of 317 medication reconciliations were conducted: 102 (32.2%) without supervision and 215 (67.8%) with supervision by a pharmacy student. Peer supervision reduced the pharmacist time required for this task by 52%; the mean time decreased from 23 (SD 11) minutes to 11 (SD 6) minutes. Furthermore, peer supervision was associated with a decrease in the number of errors made by students (from 1.5 to 0.9 per reconciliation) and detected by pharmacists during reconciliation validation.

Conclusions: Student peer validation appears to be an innovative, strategic method for optimizing medication reconciliations, freeing up pharmacist time, and leveraging the skills of pharmacy students.

(JMIR Hum Factors 2026;13:e77486) doi:10.2196/77486

KEYWORDS

clinical pharmacy; admission medication reconciliation; peer validation; pharmacy students; supervision

Introduction

Admission medication reconciliation plays a crucial role in the prevention and interception of medication errors [1,2]. It ensures the accurate and comprehensive transfer of information about patients' home medications among health care professionals [3]. Admission medication reconciliation also addresses other objectives, including continuity of treatment, diagnosis of drug-related iatrogenesis, and establishment of a reliable reference dataset for subsequent medication reviews.

Patient admission is a critical moment for medication safety: the risk of prescription errors is 70% higher on admission than during the rest of the hospital stay [4]. To address this problem, admission medication reconciliation is crucial for the improvement of patient safety and care quality. This practice is now essential in modern health care. With a view to maximizing the benefits of admission medication reconciliation and reducing medication-related errors, our institution (Lille University Medical Center, Lille, France) is now considering

XSL•FO
RenderX

the expansion of this reconciliation technique from high-risk patients only to a broader range of patients [5].

The implementation of clinical pharmacy activities requires changes in organizational practices as a function of available human resources. In our institution, about 50 pharmacy students are assigned to clinical departments. Some of these students conducted medication reconciliations by preparing "best possible medication histories" (BPMHs) [6]. These histories were then compared with admission prescriptions under the supervision of a pharmacist or pharmacy resident [7]. All the pharmacy students attended a standardized program of practical and knowledge-based training sessions developed by 4 of our clinical pharmacists. Each student was then qualified by the supervising pharmacist of the clinical service using a validated checklist of criteria.

Medication reconciliation is increasingly recognized as a critical component of patient safety and is now integrated into care pathways in many health care systems. In France, its implementation is included as a certification criterion by the French National Authority for Health (*Haute Autorité de Santé*, HAS), particularly for high-risk patient groups, such as those in geriatrics and oncology. Despite its recognized importance, medication reconciliation remains challenging in hospital settings, primarily due to the significant time investment required from pharmacists, which competes with their other responsibilities. Its successful implementation depends on both the commitment of health care professionals and the resources allocated, factors that are often limited in practice [8,9]. There are various strategies for optimizing hospital resources, especially when reconciliation involves a number of different health care stakeholders (eg, pharmacy technicians and nurses) [10,11]. The BPMHs drawn up by pharmacy students and pharmacy technicians as part of the medication reconciliation process are accurate and efficient, reduce costs, and provide support for other health care professionals [12]. However, some errors persist, due to the students' limited clinical experience and the relatively short duration of their placements—typically up to 3 months and sometimes on a part-time basis. Therefore, oversight by experienced pharmacists remains essential to maintain the quality and safety of the reconciliation process, given that these individuals are still in training. In our acute geriatric units (AGUs) and internal medicine departments (IMDs), medication reconciliation is particularly time- and resource-intensive; hence, we are exploring organizational changes that can address constraints on pharmacist time. One of these changes might be the introduction of student supervisors as a means of supporting pharmacy residents and pharmacists.

We hypothesized that incorporating peer review, a process commonly used in scientific research for critical evaluation [13], into medication reconciliation could both optimize pharmacist time and maintain high-quality practices. Additionally, supervision and peer review may provide pharmacy students with valuable opportunities to develop essential professional skills, including effective communication, priority management, teamwork, and cohesion.

The objective of this study was to evaluate the impact of a peer supervision process among pharmacy students on the accuracy and efficiency of admission medication reconciliations.

Methods

Study Design

This was a prospective, single-center, observational study conducted at Lille University Medical Center in 2 phases in an AGU and an IMD. In the initial, qualitative phase, we developed an organizational model. In the second phase, we evaluated this organizational model by collecting data on admission medication reconciliations for patients aged 65 years and older.

Medication reconciliation was performed in accordance with the recommendations of Haute Autorité de Santé, which defines four essential steps in its implementation: (1) collection of information on medications currently taken or intended to be taken by the patient, (2) development of a BPMH synthesizing this information, (3) validation of the BPMH, and (4) sharing and use of the BPMH within the care team [10]. The medication reconciliation process was conducted under the responsibility of each clinical pharmacist who was a member of the multidisciplinary team involved in the patient's care. This process did not include medication prescribing by the pharmacist but was based on communication and collaboration with the physician responsible for the patient's follow-up.

To ensure that reconciliations were comprehensively reviewed by supervisors, we considered various tools for guiding the students. Discussions between the initiating investigators thus led to the development of a checklist for the student supervisor. The supervisory checklist was developed through a multistep, consensus-building process [14] (Multimedia Appendices 1-4). First, 2 investigators created a draft version of the checklist and shared it with the reference pharmacists involved in medication reconciliation. The latter evaluated each item, provided feedback, and suggested improvements. The checklist was revised accordingly and sent back to the reference pharmacists for further assessment. This iterative process continued until a consensus on the checklist's content was reached.

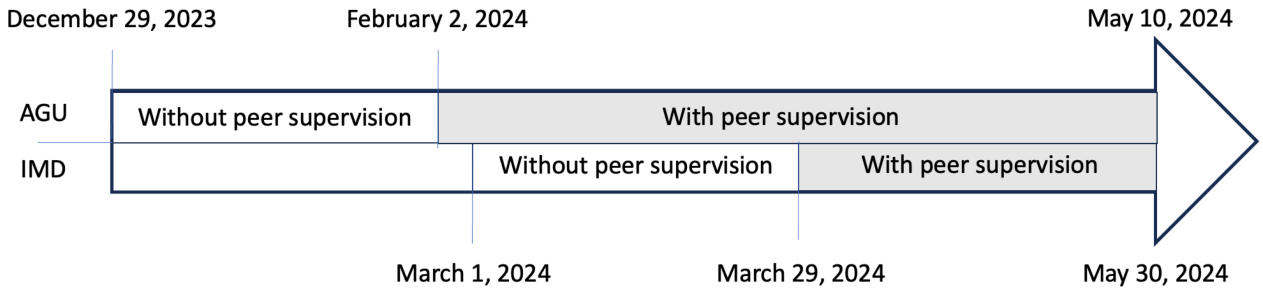
The study was conducted from December 2023 to May 2024 in an AGU and an IMD at Lille University Medical Center. The evaluation period comprised a baseline phase from December 2023 to February 2024 (during which medication reconciliation data were collected without supervision and without a checklist) and an intervention phase. Although consensus on the final checklist was achieved in April 2024, an initial version was used for supervision in February and March 2024. The evaluation involved a reorganization that allowed trained volunteer students to review, modify, and validate their peers' admission medication reconciliations in collaboration with pharmacists and pharmacy residents. The BPMHs and the medical prescriptions at admission were recorded in our medical center's electronic patient record system (Sillage, version 22.1; MipihSIB).

For each department, the evaluation was carried out over 2 periods (Figure 1). The first period served as a baseline: admission medication reconciliations performed by pharmacy

XSL•FO
RenderX

students were supervised solely by pharmacy residents or pharmacists. The second (intervention) period corresponded to student supervision of admission medication reconciliations conducted by other pharmacy students. The supervising students were trained in 2 stages. Initially, a clinical pharmacist with postgraduate specialty training and 7 years of experience or a final-year pharmacy resident in clinical pharmacy supervised an admission medication reconciliation and demonstrated the method to the student. Once the student had learned the method, they supervised a reconciliation by another student but under the observation of the clinical pharmacist or the final-year pharmacy resident in clinical pharmacy. If the pharmacist validated the student's skills, the latter was allowed to supervise another student's admission medication reconciliations alone.

Figure 1. Study timeline by department. AGU: acute geriatric unit; IMD: internal medicine department.



Variables

The study data were collected by 2 investigators (a clinical pharmacist in the IMD and a final-year pharmacy resident in the AGU) using a standardized collection form. For each reconciliation, the following data were recorded. First, we recorded the total time taken by the student to complete the medication reconciliation [15]. In both phases of the study, this duration included the time spent collecting and checking the patient's medication information for a BPMH, comparing the BPMH with the medicines prescribed at admission, and identifying and rectifying any discrepancies [7]. The BPMH integrates multiple information sources, prioritized according to their reliability, including primary objective data (outpatient pharmacy dispensing records and electronic prescriptions), interviews with community pharmacists, telephone contact with the primary care physician, prior medical documentation of treatment changes, patient or caregiver interviews, and hospital medical records. Patient-reported discrepancies that cannot be independently verified are flagged. Reconciliation accuracy is assessed by comparing the BPMH with clinical judgment and current prescriptions: clinically justified discrepancies (eg, temporary withholding of anticoagulant therapy due to bleeding risk) are considered appropriate when clearly documented and communicated, whereas unintentional discrepancies (such as omissions, duplications, or dosing errors) are classified as medication errors requiring resolution. Second, we recorded the time taken by the pharmacist to validate the medication reconciliation (after the latter had been supervised by a student, if applicable), up until the discussion with the prescriber. Third, we recorded the errors identified and corrected by the student supervisor and/or pharmacist, including the number, type, and medications involved. An "error" was defined as an omission, unjustified addition, or misinterpretation of an element in the student's observation that contradicted the instructions received during the training or a specific item of the validation checklist. All identified errors were analyzed. Fourth, we recorded the total number of medications per BPMH. Fifth, we recorded the adverse drug event risk score developed by Trivalle et al [16], which has been validated in patients aged 65 years and older in France. The score is derived from 3 risk factors: polypharmacy (≥7 medications), antipsychotic treatment, and recent anticoagulant exposure within the previous 3 months. Sixth, we recorded the initials of the student supervisor and the validating pharmacist.

The number of errors identified and corrected by pharmacists, as well as the time required for pharmacists to validate medication reconciliations, was assessed in 2 independent samples before and after the implementation of student supervision. This approach enabled us to estimate the impact of student oversight on the quality and efficiency of admission medication reconciliations performed by peers. No threshold for an acceptable number of errors was defined; all errors were recorded, whether identified by the student supervisor or the supervising pharmacist, without distinguishing between the 2 sources.

Statistical Analysis

The data were entered into computer spreadsheets (Microsoft Excel). Qualitative variables were described as the frequency (percentage). Quantitative variables were described as the mean (SD) or, when the data were not normally distributed, the median (IQR). The normality of distribution was assessed both graphically and with the Shapiro-Wilk test.

The 2 periods (baseline and intervention) were compared using 2-tailed Student *t* test or, when the data were not normally distributed, the Mann-Whitney *U* test. The comparison included all medication reconciliations performed without peer supervision (baseline phase) and all peer-supervised medication reconciliations (intervention phase). Reconciliations performed by student supervisors were not included because they did not involve supervision, as only 1 supervisor was available per week.

The number of errors before and after the introduction of student supervision was analyzed using a generalized linear regression model for count data (negative binomial distribution with log link function), adjusted for confounding factors (the number of medications per BPMH, the adverse drug event prediction score, and time taken by the student to perform the reconciliation). An adjusted rate ratio (RR) with 95% CI was reported as a

XSL•FO
RenderX

measure of association. The time taken by the pharmacist to correct and validate the reconciliation before and after the introduction of student supervision was compared by applying the same method.

Factors associated with the number of errors found in each reconciliation by the pharmacist were identified in a generalized linear regression model for count data (negative binomial distribution with a log link function). Again, the RR (95% CI) was calculated as a measure of association. Factors associated with the time taken by the pharmacist to correct and validate the reconciliation were identified by applying the same method.

The association between the number of errors and the time taken by the pharmacist was assessed by calculating the Spearman correlation coefficient (ρ).

The threshold for statistical significance was set at $P<.05$. All statistical analyses were performed using SAS software (version 9.4; SAS Institute Inc).

Ethical Considerations

No personal data (medical or otherwise) of the patients involved in the reconciliations were collected. In line with French legislation, neither informed consent nor approval by an institutional review board was required or sought. The study database was registered with the data protection officer of Lille University Hospital under the terms of the MR-004 French national reference methodology [17] on the processing of personal data for study, evaluation, or research purposes not involving human subjects.

Results

Overview

The evaluation phase involved various pharmacists and pharmacy students. The first part involved 2 pharmacists, and the second part involved 15 students (including n=8, 53.3% supervisors), 3 pharmacists, and 2 pharmacy residents. Data on 317 admission medication reconciliations were collected: 102 (32.2%) reconciliations were not peer supervised (n=31, 9.8% in the AGU and n=71, 22.4% in the IMD), and 215 (67.8%) were peer supervised (n=191, 60.3% in the AGU and n=24, 7.6% in the IMD). A rotation system had been set up in the AGU: of the 4 pharmacy students involved, 1 supervised the reconciliations for 2 weeks at a time, and the remaining 3 performed the reconciliations. In the IMD, a student supervised the other 2 students.

The Medication Profile of the Patients Included in the Study

The mean number of medications per BPMH was higher during the student-supervised period than during the nonsupervised period (9.6, SD 3.8 vs 8.6, SD 4 medications, respectively), whereas the mean total time needed for reconciliation was shorter (138, SD 42 min vs 174, SD 96 min, respectively; Table 1). The adverse drug event risk score was similar in the 2 periods.

Table . Comparison of the reference and implementation phases regarding patient medication profiles and the duration of reconciliations performed by students.

	Without peer supervision (n=102)	With peer supervision (n=215)	*P* value
Adverse drug event risk score, median (IQR)	3.0 (0-6)	4.0 (1-6)	.10
Total number of medications in the best possible medication history, mean (SD)	8.6 (4)	9.6 (3.8)	.047
Total time taken by the student to complete the reconciliation (min), mean (SD)	174 (96)	138 (42)	<.001

Factors Influencing the Time Taken by the Pharmacist to Validate a Reconciliation

The time taken by the pharmacist was significantly associated with the number of medications per BPMH (RR=1.03, 95% CI 1.01 - 1.04; $P=.001$). However, this time was not influenced by the adverse drug event prediction score (RR=1.02, 95% CI 0.99 - 1.04; $P=.05$) or the time taken by the student to perform the reconciliation (RR=1.02, 95% CI 0.96 - 1.08; $P=.06$).

Factors Influencing the Number of Errors Corrected by the Pharmacist

The number of errors corrected by the pharmacist was significantly associated with the number of medications per BPMH (RR=1.09, 95% CI 1.05 - 1.12; $P<.001$), the time taken by the student to perform a reconciliation (RR=1.19, 95% CI 1.06 - 1.33; $P=.004$), and the adverse drug event prediction (RR=1.10, 95% CI 1.05 - 1.14; $P<.001$).

Impact of Peer Supervision

Our results showed that the number of errors was smaller and the total pharmacist time was shorter during the student-supervised period than during the nonsupervised period (Table 2). The mean time taken by the pharmacist to rectify errors in reconciliation was significantly shorter during the student-supervised period (mean 11, SD 6 minutes) than during the nonsupervised period (mean 23, SD 11 minutes); this corresponded to a time saving of 53% (RR=0.47, 95% CI 0.42 - 0.52; $P<.001$). Moreover, the number of errors corrected by the pharmacist fell from 1.5 to 0.9 (a 36% decrease; RR=0.64, 95% CI 0.49 - 0.83; $P<.001$). For the study period as a whole (N=317 admission medication reconciliations), we observed a significant linear correlation between the time taken by the

XSL•FO
RenderX

pharmacist and the number of errors rectified (ρ=0.45, 95% CI 0.38 - 0.55; *P*<.001).

Table . Impact of peer supervision on the number of errors corrected and the pharmacist validation time.

	Without peer supervision (n=102), mean (SD)	With peer supervision (n=215), mean (SD)	Rate ratio (95% CI)	*P* value
Time taken to validate the reconciliation (minutes)	22.5 (10.9)	11.4 (6.08)	0.47 (0.42 - 0.52)	<.001
Number of errors corrected	1.5 (1.51)	0.9 (1.24)	0.64 (0.49 - 0.83)	<.001

Analysis of Errors

We identified 143 errors corrected by student supervisors or pharmacists (Multimedia Appendix 5). Not all errors were described in detail. The most frequent error (accounting for n=25, 17.5% of all errors) was the omission of the precise administration frequency for as-needed (ie, on-demand) medications. The second most frequent error (n=13, 9.1%) was the misinterpretation of differences between the patient's medication history and the admission prescriptions.

The analysis of reconciliation errors also enabled us to identify the most frequently involved drug classes: other analgesics and antipyretics were involved in 28 (19.6%) errors, followed by antihypertensives (n=18, 12.6%; Multimedia Appendix 6).

Discussion

Principal Findings

Peer validation of admission medication reconciliations by pharmacy students proved to be effective in optimizing the pharmacists' working time; the average pharmacist time per reconciliation decreased from 23 minutes to 11 minutes, and the average number of errors made by the student conducting the reconciliation fell from 1.5 to 0.9. However, these reconciliation time savings must be balanced against the time required to train the student supervisors, that is, 2 to 4 half-days. We observed a significant linear correlation between the time spent by pharmacists and the number of errors rectified (*P*<.001), which confirmed the validity of our data.

In France, the current regulations do not specify how admission medication reconciliations should be validated. Here, we chose to have all student-conducted and student-supervised reconciliations validated by a pharmacist or a resident to ensure the quality of the work performed. Accordingly, the pharmacy students entered their reconciliation reports into our medical center's electronic patient record system so that the validated reconciliations could be consulted by the prescribers. According to the French Society of Clinical Pharmacy (Société française de pharmacie clinique), the pharmacist is the guarantor of the reconciliation process [18] and the validation of certain steps in this process [15].

Peer evaluation in pharmacy education is a valuable pedagogical tool that helps students develop key skills such as communication, teamwork, priority management, and critical thinking [19]. It fosters accountability, self-reflection, empathy, and cultural sensitivity within a supportive, student-centered environment. Peer learning also enhances satisfaction for both students and supervisors, promotes lifelong learning, and is cost-effective [20]. Although commonly used in research and initial training, peer evaluation is less documented in hospital pharmacy settings [21-23]. However, studies show that peer mentoring improves academic success [24] and practical skills. Peer supervision encourages open feedback [25], motivation, and confidence by creating a less hierarchical and collaborative learning atmosphere [26]. However, successful implementation relies on a culture of transparency and mutual trust. Peer supervision may also improve the quality of patient care and can help pharmacy students to acquire skills that will be essential in their careers.

With a view to improving the student supervision of admission medication reconciliations, we developed a checklist. This type of tool is known to be effective because it ensures that no details are overlooked, saves time (by providing a written reference instead of relying on memory), and enhances overall organization. Checklists are widely used in the medical field in general and operating rooms in particular; the cross-checking of essential criteria before, during, and after each operation helps increase levels of safety and efficiency [27].

It has been reported that the empowerment of students with medication reconciliation tasks positively influences their perception of future responsibilities and helps them develop field-specific, general, and organizational skills [28]. Active participation enhances the students' perceived level of autonomy [29]. In a questionnaire, the pharmacists stated that they favored student supervision of admission medication reconciliations under certain conditions, such as those with specialty-specific issues and medication switching. Student involvement allows health care institutions to free up pharmacist time and reinforces trust in medication histories [30]. Trust—based on training, evaluations, and qualifications—is crucial for assigning tasks and peer supervision. Key concepts guiding supervising students are professionalism, self-awareness, and communication [31]. Training in medication reconciliation is essential and improves students' comfort, confidence, and skills [32]. Supervisors require specific training and experience to assess their peers' work accurately. Learning programs enable students to contribute to continuity of care activities and are crucial in pharmacy faculties for enhancing preparation for practice, adding value to patient care, and benefiting the health care system [33].

Although these positive outcomes were evident, the student supervisors' impact on admission medication reconciliations might have been underestimated. During the reference phase (ie, with nonsupervised reconciliations), the pharmacy students were well integrated into the team and were actively engaged in discussions with prescribers before validation. This high level

XSL•FO
RenderX

of integration might have reduced the number of errors (relative to what might be expected with novice students) and thus might have partly masked the true benefits of supervision. Furthermore, reconciliations performed by student supervisors were not included in the evaluation phase because they did not involve supervision, which explains the lower number of reconciliations. Other limitations of our study should be acknowledged, including an imbalance in recruitment between the 2 departments and the single-center nature of the study, which warrants confirmation of our findings in other organizational contexts.

This research could be extended by an evaluation of the system in other health care departments, a comparison of the clinical pharmacists' activity indicators with and without student pharmacist supervision, and the measurement of levels of satisfaction among students, pharmacists, and prescribers [34].

Analysis of errors in our geriatric population showed that first-line analgesics were most frequently involved, with the main error being students' failure to specify the dosing frequency for as-needed medications. It remains unclear whether these errors were due to knowledge gaps or simple oversight, as we did not distinguish between errors made under supervision by more or less experienced supervisors. Further evaluation of pharmacist interventions and the context of these errors would help clarify whether they stem from carelessness on the part of the student or a lack of professional practice-based knowledge.

Supervision develops autonomy and can be viewed as an entrustable professional activity, that is, a specific task that a student can perform independently once competent [35]. This approach balances autonomy against safety, especially in areas such as geriatrics, where there is a high risk of adverse drug events and potentially unfamiliar medication switches. Supervision creates a training pathway toward autonomous practice, although senior verification is still required to ensure care quality and patient safety. According to the literature on medication review analysis, the integration of a checklist into an artificial intelligence process might enhance efficiency, reduce errors, and improve traceability [36]. However, final validation by a pharmacist is still essential for maintaining rigorous standards and high-quality outcomes amid ongoing technological progress.

Conclusions

A new peer supervision method for admission medication reconciliations by pharmacy students optimized pharmacists' working time and was successfully integrated into our IMDs and AGUs. This validation process helped identify and correct potential errors or inconsistencies before reconciliations were finalized. Peer supervision reduced errors through the use of a standardized procedure and specific training. By decreasing the number of errors to be corrected, this approach reduced the time that pharmacists spent validating reconciliations. This new organizational model had a positive impact by improving pharmacists' time management and availability for other activities, while also providing students with substantial educational benefits in medication reconciliation.

Acknowledgments

The authors would like to thank the pharmacy students and pharmacists who participated in the development and evaluation of the organizational model.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this paper.

Data Availability

All data generated or analyzed during this study are supplied in supporting files available for download along with the published manuscript.

Authors' Contributions

Investigation: BA
Formal analysis: BA, LC
Methodology: BA, MM, AT, LC, BD
Supervision: MM, AT, JBB, ML, PO, BD
Project administration: BD
Resources: BD
Validation: BD
Writing–original draft: BA
Writing–review and editing: BA, MM, AT, JBB, ML, PO, BD

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Development of the organizational model.
[DOCX File, 19 KB - humanfactors_v13i1e77486_app1.docx]

Multimedia Appendix 2
Responses from pharmacists (first questionnaire; Q1) and pharmacy students (second questionnaire; Q2).
[DOCX File, 17 KB - humanfactors_v13i1e77486_app2.docx]

Multimedia Appendix 3
Votes on validation in various reconciliation scenarios.
[DOCX File, 18 KB - humanfactors_v13i1e77486_app3.docx]

Multimedia Appendix 4
Student supervisor's checklist.
[DOCX File, 21 KB - humanfactors_v13i1e77486_app4.docx]

Multimedia Appendix 5
Type of errors found in reconciliations supervised by students.
[DOCX File, 16 KB - humanfactors_v13i1e77486_app5.docx]

Multimedia Appendix 6
Drug classes involved in errors in reconciliations supervised by students.
[DOCX File, 15 KB - humanfactors_v13i1e77486_app6.docx]

References

1. Phatak A, Prusi R, Ward B, et al. Impact of pharmacist involvement in the transitional care of high-risk patients through medication reconciliation, medication education, and postdischarge call-backs (IPITCH Study). J Hosp Med 2016 Jan;11(1):39-44. [doi: 10.1002/jhm.2493] [Medline: 26434752]
2. Schnipper JL, Kirwin JL, Cotugno MC, et al. Role of pharmacist counseling in preventing adverse drug events after hospitalization. Arch Intern Med 2006 Mar 13;166(5):565-571. [doi: 10.1001/archinte.166.5.565] [Medline: 16534045]
3. Allenet B, Juste M, Mouchoux C, et al. De la dispensation au plan pharmaceutique personnalisé: vers un modèle intégratif de pharmacie clinique. Pharm Hosp Clin 2019 Mar;54(1):56-63. [doi: 10.1016/j.phclin.2018.12.003]
4. Dornan T, Ashcroft D, Heathfield H, et al. An in depth investigation into causes of prescribing errors by foundation trainees in relation to their medical education. EQUIP study. : General Medical Council; 2009 URL: https://www.gmc-uk.org/cdn/documents/final-report-prevalence-and-causes-of-prescribing-errors_pdf-28935150.pdf [accessed 2026-02-20]
5. Belda-Rustarazo S, Cantero-Hinojosa J, Salmeron-García A, González-García L, Cabeza-Barrera J, Galvez J. Medication reconciliation at admission and discharge: an analysis of prevalence and associated risk factors. Int J Clin Pract 2015 Nov;69(11):1268-1274. [doi: 10.1111/ijcp.12701] [Medline: 26202091]
6. Best possible medication history. NSW Government. URL: https://www.cec.health.nsw.gov.au/keep-patients-safe/medication-safety/cmm/bpmh [accessed 2025-04-30]
7. Medication reconciliation. NSW Government. URL: https://www.cec.health.nsw.gov.au/keep-patients-safe/medication-safety/cmm/mec-rec [accessed 2025-04-30]
8. La conciliation médicamenteuse: enquête sur son déploiement national. French Ministry of Health. 2024. URL: https://sante.gouv.fr/soins-et-maladies/qualite-des-soins-et-pratiques/qualite/article/la-conciliation-medicamenteuse-enquete-sur-son-deploiement-national [accessed 2025-04-30]
9. Baranyai A, Vaniet F, Coussemacq M, Cousein E. La conciliation médicamenteuse. Société Française de Médecine d'Urgence. URL: https://www.sfmu.org/upload/70_formation/02_eformation/02_congres/Urgences/urgences2013/donnees/communications/resume/posters/CP364.pdf [accessed 2025-04-30]
10. Conciliation des traitements médicamenteux – prévenir les erreurs. Haute Autorité de Santé. 2019. URL: https://www.has-sante.fr/jcms/pprd_2974294/fr/conciliation-des-traitements-medicamenteux-prevenir-les-erreurs [accessed 2025-04-30]
11. Duwez M, Valette A, Foroni L, Allenet B. Implication du préparateur en pharmacie hospitalière dans le déploiement de la conciliation des traitements médicamenteux en France: représentations et engagement des préparateurs. Ann Pharm Fr 2019 Mar;77(2):168-177. [doi: 10.1016/j.pharma.2018.11.003]
12. Champion HM, Loosen JA, Kennelty KA. Pharmacy students and pharmacy technicians in medication reconciliation: a review of the current literature. J Pharm Pract 2019 Apr;32(2):207-218. [doi: 10.1177/0897190017738916] [Medline: 29105575]

XSL•FO
RenderX

13. Ross-Hellauer T. What is open peer review? A systematic review. F1000Res 2017 Apr 27;6:588. [doi: 10.12688/f1000research.11369.2] [Medline: 28580134]
14. Makhmutov R. The Delphi method at a glance. Pflege 2021 Aug;34(4):221. [doi: 10.1024/1012-5302/a000812] [Medline: 34292075]
15. Le Maoût M, Satori D, Chaouch AA, Corsin L, Dalle-Pécal M. Intégration de la conciliation médicamenteuse dans le parcours de soins du patient insuffisant cardiaque âgé. Actual Pharm 2024 Feb;63(633):35-39. [doi: 10.1016/j.actpha.2023.12.009]
16. Trivalle C, Burlaud A, Ducimetière P, IMEPAG Group. Risk factors for adverse drug events in hospitalized elderly patients: a geriatric score. Eur Geriatr Med 2011 Oct;2(5):284-289. [doi: 10.1016/j.eurger.2011.07.002]
17. Délibération n° 2018-155 du 3 mai 2018 portant homologation de la méthodologie de référence relative aux traitements de données à caractère personnel mis en œuvre dans le cadre des recherches n'impliquant pas la personne humaine, des études et évaluations dans le domaine de la santé (MR-004). Légifrance. 2018. URL: https://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000037187528 [accessed 2026-02-13]
18. SFPC mémo conciliation des traitements médicamenteux. Société Française de Pharmacie Clinique. 2015. URL: https://sfpc.eu/wp-content/uploads/2019/08/FicheMemoConciliationMedicamenteuseSFPC.pdf [accessed 2025-04-30]
19. Plott AJ, McIntosh T, O'Ferral H, Bennett MC, Taylor S. Impact of early direct patient care introductory pharmacy practice experiences on student pharmacists' career exploration. Curr Pharm Teach Learn 2021 Feb;13(2):164-168. [doi: 10.1016/j.cptl.2020.09.004] [Medline: 33454074]
20. Dyar A, Henriksson P, Stenfors T, Lachmann H, Kiessling A. Differences in supervision on peer learning wards: a pilot survey of the supervisor's perspective. Adv Med Educ Pract 2024 Feb 2;15:85-96. [doi: 10.2147/AMEP.S439968] [Medline: 38327849]
21. Stark JE, Cole JL, Barnes LM, Chapman A, Costner M. Implementation of a novel learning experience in scientific writing, publishing, and peer review into a first year pharmacy practice residency. J Am Pharm Assoc (2003) 2022;62(3):840-844. [doi: 10.1016/j.japh.2021.11.014] [Medline: 34887188]
22. Le S, Howard ML. Peer evaluation of teaching programs within pharmacy education: a review of the literature. Curr Pharm Teach Learn 2023 Nov;15(11):984-992. [doi: 10.1016/j.cptl.2023.09.009] [Medline: 37734970]
23. Morbitzer KA, McLaughlin JE, Henson B, Fassett KT, DiVall MV. Current state of and opportunities for enhancing peer evaluation practices across the pharmacy academy. Am J Pharm Educ 2022 Nov;86(8):ajpe8887. [doi: 10.5688/ajpe8887] [Medline: 34987069]
24. Asal NJ, Provisor EM, Ryu R. Peer mentoring in pharmacy programs: recommendations for implementation based on a review of available literature. Curr Pharm Teach Learn 2024 Jun;16(6):496-502. [doi: 10.1016/j.cptl.2024.03.014] [Medline: 38692946]
25. #PeerRevWk19 is almost here!. X (formerly Twitter). 2019. URL: https://x.com/Editage/status/1172019578760642560 [accessed 2025-04-30]
26. Mills F, Swift SJ. What can be gained through peer supervision? Educ Child Psychol 2015;32(3):105-118. [doi: 10.53841/bpsecp.2015.32.3.105]
27. Les check-lists pour la sécurité du patient. Haute Autorité de Santé. 2024. URL: https://www.has-sante.fr/jcms/c_1518984/fr/les-check-lists-pour-la-securite-du-patient [accessed 2025-04-30]
28. Pittenger AL, Chapman SA, Frail CK, Moon JY, Undeberg MR, Orzoff JH. Entrustable professional activities for pharmacy practice. Am J Pharm Educ 2016 May 25;80(4):57. [doi: 10.5688/ajpe80457] [Medline: 27293224]
29. Eudaley ST, Brooks SP, Jones MJ, Franks AS, Dabbs WS, Chamberlin SM. Evaluation of student-perceived growth in entrustable professional activities after involvement in a transitions-of-care process within an adult medicine advanced pharmacy practice experience. Curr Pharm Teach Learn 2022 Feb;14(2):193-199. [doi: 10.1016/j.cptl.2021.11.032] [Medline: 35190161]
30. Perrier Q, Vitale E, Tanty A, et al. Formation pratiques des étudiants en pharmacie aux soins pharmaceutiques hospitaliers. Le Pharm Clin 2024 Jun;59(2):e97. [doi: 10.1016/j.phacli.2024.04.154]
31. Siraj HH, Salam A, Azmina H, et al. Self awareness and reflective skills in the promotion of personal and professional development of future medical professionals. Educ Med J 2013;5(4):e29-e33. [doi: 10.5959/eimj.v5i4.171]
32. Newsom LC, Dupree LH, Thurston MM, Vivian Liao T, Nwaesei AS. A scoping review of student pharmacist-led transitions-of-care initiatives. Am J Pharm Educ 2023 Jun;87(6):100001. [doi: 10.1016/j.ajpe.2023.02.001] [Medline: 37316136]
33. Yailian AL, Reallon E, Paillet C, Janoly-Dumenil A. Analyse des pratiques de conciliation médicamenteuse auprès des prescripteurs et étudiants en pharmacie. Le Pharm Clin 2023 Mar;58(1):32-39. [doi: 10.1016/j.phacli.2022.06.002]
34. Jacquet M, Maillart C, Georges F, Poumay M. De la satisfaction à la performance: dépasser les indicateurs de satisfaction pour évaluer et réguler efficacement. Educ Form 2014;e-302:109-120 [FREE Full text]
35. Haines ST, Pittenger AL, Stolte SK, et al. Core entrustable professional activities for new pharmacy graduates. Am J Pharm Educ 2017 Feb 25;81(1):S2. [doi: 10.5688/ajpe811S2] [Medline: 28289312]

XSL•FO
RenderX

36. Long J, Yuan MJ, Poonawala R. An observational study to evaluate the usability and intent to adopt an artificial intelligence-powered medication reconciliation tool. Interact J Med Res 2016 May 16;5(2):e14. [doi: 10.2196/ijmr.5462] [Medline: 27185210]

Abbreviations

AGU: acute geriatric unit
BPMH: best possible medication histories
IMD: internal medicine department
RR: rate ratio

Edited by A Kushniruk; submitted 16.May.2025; peer-reviewed by D Ajayi, H Hosono, I Iqbal, RM Kruer, Y Wang; revised version received 26.Dec.2025; accepted 19.Jan.2026; published 09.Mar.2026.

Please cite as:
Ammor B, Masse M, Toulemonde A, Cadart L, Beuscart JB, Lambert M, Odou P, Décaudin B
Effect of Pharmacy Student Peer Supervision on the Accuracy of Admission Medication Reconciliation: Prospective Pre-Post Observational Study
JMIR Hum Factors 2026;13:e77486
URL: https://humanfactors.jmir.org/2026/1/e77486
doi:10.2196/77486

© Basma Ammor, Morgane Masse, Anne Toulemonde, Laurine Cadart, Jean-Baptiste Beuscart, Marc Lambert, Pascal Odou, Bertrand Décaudin. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 9.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Health Information Technology–Related Loss of Central Surveillance Data in a Heart Intensive Care Unit: Multi-Framework Case Report

Md Shafiqur Rahman Jabin[1,2], PhD

[1]Department of Medicine and Optomtery, Linnaeus University, Universitetsplatsen 1, Kalmar, Sweden

[2]Faculty of Health and Social Care, University of Bradford, Bradford, England, United Kingdom

Corresponding Author:
Md Shafiqur Rahman Jabin, PhD
Department of Medicine and Optomtery, Linnaeus University, Universitetsplatsen 1, Kalmar, Sweden

Abstract

Background: Centralized electronic surveillance systems are widely used in intensive care settings to support continuous physiological monitoring and patient safety. Failures in health information technology (HIT) infrastructure can disrupt workflows, reduce situational awareness, and create latent risks for serious harm. Understanding such events requires analytic approaches that go beyond single-classification frameworks.

Objective: This study aimed to classify and analyze an HIT-related incident that involved loss of central surveillance data in a heart intensive care unit using multiple complementary patient safety and human factors frameworks.

Methods: This study is a qualitative case report analysis of an incident in which a central surveillance system intermittently lost server connectivity, resulting in unavailability and loss of monitoring data. The narrative was derived from an incident report and supporting documentation and was translated and linguistically adapted for publication. The incident was independently classified using 5 frameworks: the International Classification for Patient Safety (ICPS), the Health Information Technology Classification System (HIT-CS), Systems Engineering Initiative for Patient Safety (SEIPS) 2.0, the sociotechnical model by Sittig and Singh, and the Human Factors Analysis and Classification System for health care (HFACS-Healthcare). Findings were synthesized across frameworks.

Results: All 5 frameworks characterized the event as an HIT-driven system failure involving information unavailability, delayed detection, and multi-patient impact. The HIT-CS identified a technical failure in system availability and recovery. The ICPS classified the event as a documentation or information incident with potential for severe harm. The SEIPS 2.0 and sociotechnical models highlighted disruptions to monitoring tasks and the organization's reliance on IT intervention. The HFACS-Healthcare attributed the event primarily to organizational influences and preconditions for unsafe acts, with no frontline unsafe acts identified. Convergence across frameworks emphasized system-level contributors.

Conclusions: HIT-related monitoring failures in high-acuity settings are best understood as sociotechnical system events rather than isolated technical faults or individual errors. A multi-framework approach provided complementary insights into detection, recovery, and governance vulnerabilities, supporting improved learning and resilience in clinical surveillance systems.

(JMIR Hum Factors 2026;13:e92560) doi:10.2196/92560

KEYWORDS

safety incident analysis; clinical information systems; monitoring system downtime; system availability failure; technology-induced error; incident classification; resilience engineering; safety-critical systems; system recovery processes; intensive care informatics

Introduction

Health information technology (HIT) is increasingly embedded in high-acuity clinical environments, where system reliability is critical for patient safety. Centralized electronic surveillance systems are widely used in intensive and cardiac care units to support continuous physiological monitoring, early detection of deterioration, and retrospective review of patient data [1,2]. While such systems offer substantial safety benefits, failures in HIT infrastructure can disrupt clinical workflows, obscure patient status, and create latent risks for severe harm [3,4]. Therefore, understanding and learning from HIT-related incidents remains a priority for patient safety research [5,6].

Incident reporting systems capture a wide range of adverse events and near-misses, but single-classification approaches often fail to fully capture the complexity of HIT-related failures [6,7]. To address this limitation, several complementary classification frameworks have been developed, each emphasizing different aspects of safety incidents. The World

XSL•FO
RenderX

Health Organization's International Classification for Patient Safety (ICPS) provides a standardized, high-level taxonomy for categorizing incident types, patient outcomes, and contributing factors, supporting comparability across settings and jurisdictions [8]. However, the ICPS does not explicitly focus on the mechanisms underlying HIT failures [9].

To more precisely characterize HIT-related safety events, Magrabi, Jabin, and colleagues developed the Health Information Technology Classification System (HIT-CS), which focuses on information flow, technical failure modes, and outcome manifestations specific to HIT [6,10,11]. The HIT-CS enables detailed identification of where and how HIT failures arise, such as system unavailability, data loss, or delayed detection, and has been widely used in analyses of technology-induced errors [12-14]. Previous research has highlighted limitations in existing incident reporting structures for HIT-related events and emphasized the need for refined classification systems to support meaningful learning and governance [15]. Strengthening classification frameworks is particularly important in high-acuity environments where system-level failures may have multi-patient implications.

In parallel, human factors and systems engineering frameworks offer additional perspectives on how technology interacts with people, tasks, and organizations. The Systems Engineering Initiative for Patient Safety (SEIPS) 2.0 model conceptualizes patient safety as an emergent property of the work system, encompassing interactions among technology, tasks, people, organizational structures, environments, and processes [16]. SEIPS supports the analysis of how HIT failures disrupt clinical work and contribute to unsafe conditions rather than focusing solely on technical faults [17].

Similarly, the sociotechnical model proposed by Sittig and Singh [18,19] emphasizes that HIT safety incidents arise from interdependencies among technical infrastructure, clinical content, human-computer interfaces, workflows, organizational policies, external pressures, and system monitoring. This framework is particularly useful for identifying latent conditions and design vulnerabilities that may not be apparent when incidents are examined through a single disciplinary lens [20].

Finally, human factors classification systems such as the Human Factors Analysis and Classification System (HFACS) for health care (HFACS-Healthcare) provide a structured approach to examine human, supervisory, and organizational contributors to safety events [21]. HFACS-Healthcare allows differentiation between frontline actions, preconditions for unsafe acts, supervisory factors, and organizational influences and is especially valuable for demonstrating when incidents are driven by system-level factors rather than individual error [22,23].

This case report describes an HIT-related incident involving loss of central patient surveillance data in a heart intensive care unit. To provide a comprehensive and methodologically rigorous analysis, the incident was classified using 5 complementary frameworks: ICPS, HIT-CS, SEIPS 2.0, the Sittig and Singh sociotechnical model, and HFACS-Healthcare. By applying multiple classification systems to a single incident, this study aimed to illustrate how different frameworks contribute distinct and complementary insights into HIT-related patient safety risks and to support more comprehensive learning from HIT failures.

Methods

Study Design

This study is a qualitative case report that describes and analyzes an HIT-related patient safety incident that occurred in a heart intensive care unit. A single-incident case design was chosen to enable an in-depth examination of technical, organizational, and human factors contributing to the event. Case reports are well suited for studying rare or complex HIT failures and for generating transferable safety insights when experimental or quantitative approaches are not feasible [1,3,12,24].

Ethical Considerations

This study involved secondary analysis of a fully deidentified incident report obtained from a national incident reporting repository. The data contained no patient identifiers or confidential information. Ethical advice was obtained from the Ethical Advisory Board in southeast Sweden (Dnr 934 - 2023). Based on the retrospective analysis of anonymized quality and safety data, formal ethical review and informed consent were not required under applicable regulations (the Swedish Ethical Review Act 2003:460) [25].

Data Source and Incident Selection

The incident analyzed in this study was obtained from a national incident reporting repository that captures medical device–related and HIT-related events reported by health care professionals and clinical engineering staff. The incident reports in the repository typically include structured fields and free-text narratives describing the incident, the results of any internal investigation, and measures taken in response [26-28]. A single incident was selected for analysis on the basis of the following criteria:

- involvement of a medical device with embedded software,
- manifestation of the incident during active clinical care,
- relevance to HIT-mediated information loss, and
- availability of sufficient narrative detail to support qualitative analysis.

The incident report included 3 core narrative components: an incident description, a summary of investigation findings, and a summary of measures. These components formed the primary data for analysis (Textbox 1). The incident description was translated and linguistically adapted from the original incident report to improve clarity and readability, without altering the factual content.

XSL•FO
RenderX

Textbox 1. Description of a health information technology–related incident involving loss of central patient surveillance data in a heart intensive care unit, including incident description, investigation findings, and implemented or proposed safety measures.

Incident description

On January 18, 2026, the central surveillance system in a heart intensive care unit lost connection with its server on 2 occasions, around 11 AM and 1 PM. During these periods, real-time patient monitoring data were unavailable and not stored. The issue was not detected immediately and became apparent at 4 PM when staff attempted to review recorded surveillance data. A trained superuser attempted to restart the system using previously distributed instructions, but the restart command did not respond, and the system interface subsequently froze. The IT department was contacted and performed a restart of the main computer.

Summary of investigation

The investigation identified a failure in the health information technology infrastructure, specifically the loss of server connectivity affecting the central surveillance system. The incident was compounded by delayed detection, as there were no real-time alerts indicating loss of data capture or system disconnection. Restart procedures that relied on local user action were ineffective, and the system entered an unresponsive state requiring IT intervention. The event highlighted dependencies between clinical monitoring, server availability, and recovery procedures, as well as limitations in system resilience and monitoring transparency.

In a heart intensive care unit, continuous centralized surveillance supports early detection of life-threatening arrhythmias, hemodynamic instability, and acute clinical deterioration. Therefore, loss of real-time and stored monitoring data creates a risk that critical physiological changes may go unnoticed or recognition may be delayed, particularly in settings where centralized monitoring complements bedside observation. Although no confirmed patient harm was identified in this case, the interruption of surveillance constituted a significant patient safety vulnerability.

Summary of measures

Immediate measures included restarting the central surveillance system with assistance from the IT department to restore functionality. Planned and recommended measures focused on improving system reliability and patient safety, including implementation of automated alerts for loss of server connectivity, clarification and testing of restart and recovery procedures, improved collaboration between clinical staff and IT services, and evaluation of redundancy or fail-safe mechanisms to prevent loss of monitoring data during system downtime.

Data Preparation and Deidentification

The original incident report was written in a language other than English and contained informal clinical and technical expressions typical of voluntary incident reporting. The report was translated into English using a direct linguistic translation approach focused on preserving semantic accuracy. No interpretive synthesis or restructuring of factual content was performed. Following translation, minor language editing was undertaken to improve clarity, including grammatical corrections and standardization of terminology, without altering the original meaning or sequence of events. The translated version was reviewed by a clinical engineer experienced in medical device incident management to ensure technical accuracy and contextual fidelity.

To ensure confidentiality and compliance with ethical and publication standards, all potentially identifying information, including names of individuals, institutions, manufacturers, locations, dates, and device identifiers, was removed prior to analysis. The final narrative was reviewed to confirm that it remained fully deidentified while retaining sufficient detail for analytical purposes [29].

Analytical Approach

To provide a comprehensive and systems-oriented analysis, the incident was classified using five complementary patient safety and human factors frameworks: (1) ICPS, (2) HIT-CS, (3) the SEIPS 2.0 model, (4) the sociotechnical model for HIT safety proposed by Sittig and Singh, and (5) HFACS-Healthcare.

Each framework was applied independently to the same incident narrative to capture different dimensions of the event, including incident type and outcome (ICPS), HIT-specific failure mechanisms (HIT-CS), work system interactions (SEIPS 2.0), sociotechnical dependencies (Sittig and Singh sociotechnical model), and human and organizational contributors (HFACS-Healthcare).

International Classification for Patient Safety

The ICPS was used to classify the incident at a high level, including incident type, detection, patient outcome, and degree of harm. It provides an internationally standardized taxonomy for patient safety incidents and supports comparison across settings and reporting systems [8]. The framework was applied descriptively to characterize the incident without attributing causality.

Health Information Technology Classification System

The HIT-CS, developed by Magrabi, Jabin, and colleagues, was used to classify the HIT-specific characteristics of the incident, including the primary issue type, secondary technical issues, outcome manifestation, workflow location, and impact scale [6,10,11]. The HIT-CS focuses on information flow and technical failure modes and is particularly suited for identifying technology-induced risks and system unavailability events [30,31].

SEIPS 2.0 Work System Analysis

The SEIPS 2.0 framework was applied to examine how the incident affected and emerged from interactions among technology, tasks, people, organizational structures, environment, and processes [16]. This analysis emphasized that the surveillance system failure disrupted continuous monitoring and retrospective data review and that recovery depended on organizational coordination with IT services.

Sociotechnical Classification

The sociotechnical model proposed by Sittig and Singh [18,19] was used to analyze interdependencies among technical infrastructure, clinical content, human-computer interfaces,

XSL•FO
RenderX

workflows, organizational policies, and system monitoring. This framework was applied to identify latent conditions such as lack of system status visibility and centralized architecture that contributed to delayed detection and multi-patient impact [4,18,20].

HFACS-Healthcare Classification

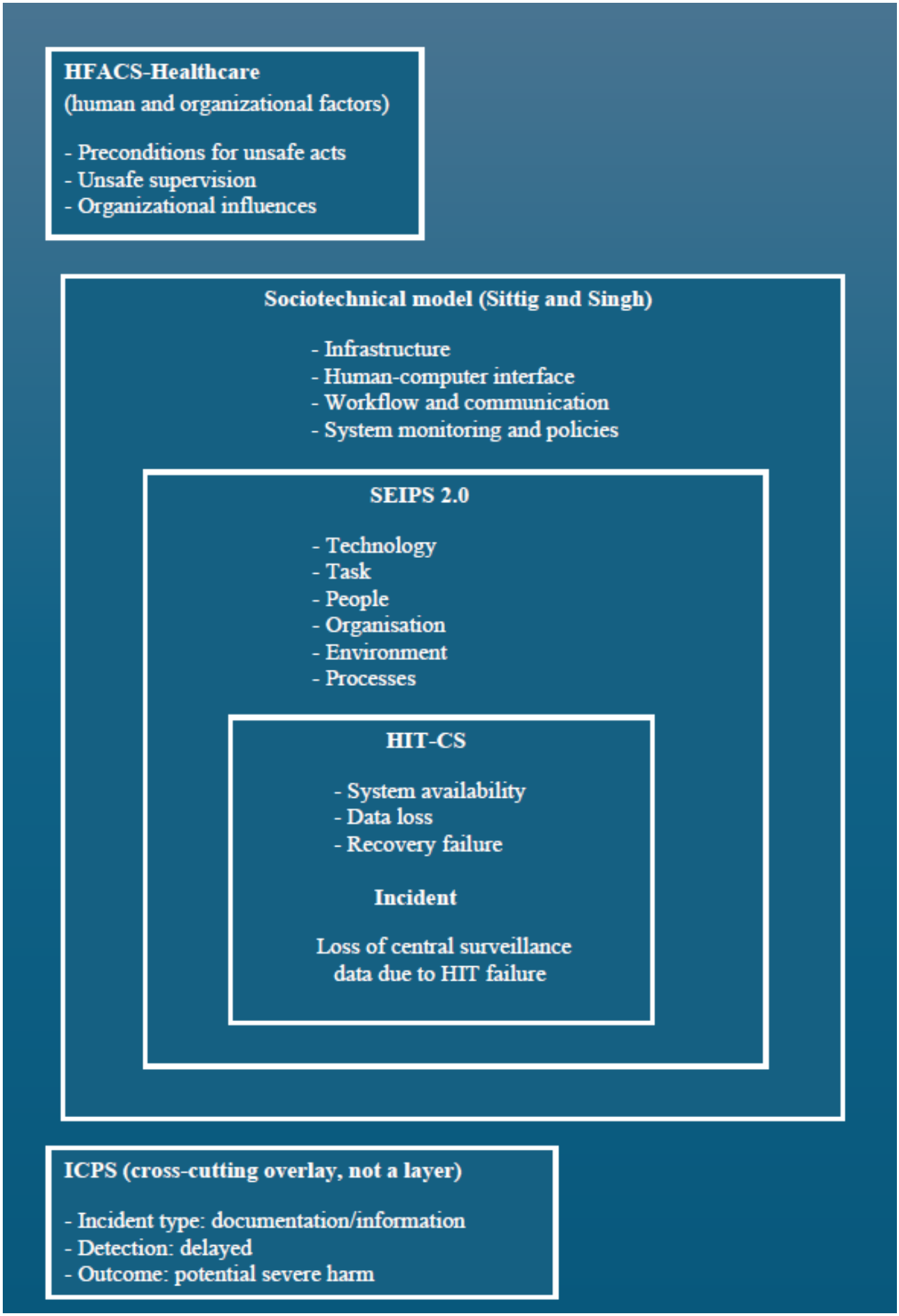
HFACS-Healthcare was applied to examine human and organizational contributors to the incident across 4 levels: unsafe acts, preconditions for unsafe acts, unsafe supervision, and organizational influences [21,32]. This framework was used to distinguish between frontline performance and system-level factors and to assess whether human error contributed to the incident. HFACS-Healthcare was applied analytically rather than punitively, consistent with its use in health care safety research [22,23,33].

Figure 1 illustrates the conceptual relationship among the 5 classification frameworks applied in this study and their complementary roles in characterizing the incident.

XSL•FO
RenderX

Figure 1. Multi-framework conceptual mapping of the HIT-related incident. HFACS, Human Factors Analysis and Classification System; HIT, health information technology; HIT-CS, Health Information Technology Classification System; ICPS, International Classification for Patient Safety; SEIPS, Systems Engineering Initiative for Patient Safety.

XSL•FO
RenderX



XSL•FO
RenderX

Results

Overview of Incident Classification

The incident was analyzed using 5 complementary classification frameworks to capture its patient safety, technical, sociotechnical, and human factors dimensions. This section summarizes the classification results according to each framework. Together, these analyses demonstrated that the incident was primarily driven by HIT system failures and latent organizational and design factors rather than by frontline human error.

ICPS Classification

Using the ICPS, the incident was primarily classified as a documentation/information incident, with unavailable and lost patient monitoring data as the central safety concern. A secondary classification under medical device/equipment reflected the role of a centralized electronic surveillance system in clinical monitoring. Detection was classified as delayed, as the system failure was identified retrospectively during data review rather than at the time of occurrence. No confirmed patient harm was identified; however, the incident was classified as having potential for severe harm due to the risk of missed or delayed recognition of life-threatening cardiac events. Table 1 presents the detailed ICPS classification, including incident type, detection, degree of harm, and contributing factors.

Table . Classification of the incident according to the International Classification for Patient Safety (ICPS).

ICPS domain and classification	Description applied to this incident
Incident type (primary)	
Documentation/information	Unavailable and lost patient monitoring data due to a loss of connection between the central surveillance system and the server
Incident type (secondary)	
Medical device/equipment (health information technology)	Failure of a central electronic surveillance system supporting physiological monitoring
Patient outcome	
No harm (outcome unknown)	No confirmed patient injury was identified; patient impact could not be fully assessed due to missing data.
Degree of harm	
No harm/potential harm	Potential for severe harm due to the risk of missed or delayed detection of life-threatening cardiac events
Detection	
Delayed detection	The incident was identified retrospectively when staff attempted to review stored monitoring data.
Contributing factors/hazards	
Organizational/system factors	Lack of automated alerts for system disconnection; reliance on a centralized monitoring infrastructure
Technical factors	Server connectivity failure; ineffective restart function; system freeze requiring IT intervention
Mitigating factors	
Human factors	Presence of a trained superuser who attempted recovery procedures
Ameliorating actions	
System-level response	Restart of the main system by the IT department and review of system recovery procedures

No confirmed patient harm was identified; however, the incident was classified as having potential for severe harm due to the risk of missed or delayed recognition of life-threatening cardiac events. In high-acuity cardiac care, even short periods of absent surveillance may increase the risk of undetected arrhythmias or delayed response to clinical deterioration.

HIT-CS Classification

Using the HIT-CS, the incident was classified as a technical HIT failure that affected system availability and connectivity. The primary issue type was classified as a software/hardware (technical) failure, specifically loss of system availability due to server disconnection. A secondary technical issue involved a system recovery failure: user-initiated restart commands did not function, and the system became unresponsive. The outcome manifestation was classified as information unavailable/data loss, reflecting both an interruption of real-time monitoring and a permanent loss of stored surveillance data.

The HIT-CS analysis also highlighted the workflow location of the failure within continuous physiological surveillance and retrospective data review, as well as a multi-patient scale

XSL•FO
RenderX

characteristic, given that the central surveillance system supported simultaneous monitoring of multiple patients. Similar multi-patient propagation patterns following HIT failures have been reported in Swedish incident analyses, where system-level disruptions affected care management across multiple patients simultaneously [30]. Table 2 summarizes the HIT-CS classification dimensions and their application to the incident.

Table . Classification of the incident using the Health Information Technology Classification System (HIT-CS) framework.

HIT-CS dimension	HIT-CS category	Specific subtype (per HIT-CS)	How it applies to this incident
Primary issue type	Software/hardware (technical)	System availability/connectivity failure	Loss of connection between the central surveillance system and the server caused intermittent unavailability of real-time and stored patient monitoring data.
Secondary technical issue	Software/hardware (technical)	System recovery/restart failure	Restart commands initiated via the system interface did not execute, and the system entered an unresponsive (frozen) state requiring IT intervention.
Outcome manifestation	Information (machine/output)	Information unavailable/data loss	Patient monitoring data were not displayed in real time or stored, resulting in permanent gaps in surveillance records.
Optional contributing factor	Information (use/human)	Delayed detection/delayed monitoring	The system failure was not immediately apparent and was detected retrospectively when staff attempted to review recorded data several hours later.
Workflow location (contextual)	HIT-mediated[a] clinical monitoring workflow	Continuous physiological surveillance/data review	The failure occurred at the stage where the central surveillance system transmits, displays, and archives physiological data.
Scale characteristic	HIT system behavior	Multi-patient impact	The central surveillance system supports the simultaneous monitoring of multiple patients; therefore, the loss of functionality affected all monitored patients during downtime.

[a]HIT: health information technology.

SEIPS 2.0 Classification

Using the SEIPS 2.0 framework, the incident was characterized as a disruption of the clinical work system arising from interactions among technology, tasks, people, organization, environment, and processes. The technology domain captured the central surveillance system's loss of reliability. The task analysis showed that neither continuous monitoring nor retrospective data review could be performed as intended. The people domain indicated appropriate user actions, with staff following established restart instructions without success.

At the organizational level, system recovery depended on escalation to IT services, highlighting the reliance on centralized technical support. The environment domain reflected dependency on a single central surveillance workspace, while the processes domain identified delayed incident detection and reactive response as key characteristics. The outcome domain captured the potential for patient harm due to unavailable monitoring data. The SEIPS 2.0 classification is presented in Table 3.

XSL•FO
RenderX

Table . Classification of the incident according to Systems Engineering Initiative for Patient Safety (SEIPS) 2.0.

SEIPS 2.0 domain	Classification	How it applies to this incident
Technology and tools	Central surveillance system reliability	Loss of server connectivity and failure of system restart functionality rendered the monitoring system unavailable.
Tasks	Continuous physiological monitoring and data review	Real-time surveillance and retrospective review of patient monitoring data could not be performed as intended.
People	Clinical staff and superusers	Staff appropriately followed established restart instructions but were unable to recover system functionality.
Organization	IT support and escalation pathways	Resolution depended on the IT department's intervention, highlighting the reliance on centralized technical support.
Environment	Central surveillance workspace	The monitoring environment depended on a single central system, amplifying the impact of system failure.
Processes	Incident detection and response	The lack of automated alerts led to delayed detection and a reactive response after a retrospective data review.
Outcomes	Potential patient harm	Risk of missed or delayed detection of life-threatening cardiac events due to unavailable monitoring data

Sociotechnical Classification (Sittig and Singh Model)

The sociotechnical analysis identified multiple interacting system components contributing to the incident. Hardware and software infrastructure failures resulted in system unavailability and data loss, while deficiencies in the human-computer interface limited users' ability to understand the system state or recover functionality. The workflow and communication dimensions revealed delayed awareness and escalation due to the absence of system status alerts.

At the organizational level, recovery procedures depended on IT intervention, and the system lacked redundancy or fail-safe mechanisms to maintain monitoring during outages. The system measurement and monitoring dimension highlighted the absence of automated alerts for loss of connectivity or for data capture failures. These interacting sociotechnical factors created latent conditions that amplified the impact of the technical failure and increased patient safety risk. Table 4 presents a comprehensive sociotechnical classification of the incident.

XSL•FO
RenderX

Table . Sociotechnical classification of the incident (Sittig and Singh model).

Sociotechnical dimension	Classification	How it applies to this incident
Hardware and software computing infrastructure	System connectivity and availability failure	Loss of connection between the central surveillance system and the server caused intermittent system unavailability and data loss.
Clinical content	Availability and completeness of monitoring data	Physiological monitoring data were unavailable in real time and not stored, resulting in gaps in clinical information.
Human-computer interface	Lack of system feedback and controllability	Restart commands produced no visible response, and the system interface became unresponsive, preventing user-led recovery.
People	Roles, training, and system use	A trained superuser followed established restart instructions appropriately but was unable to restore functionality.
Workflow and communication	Monitoring, review, and escalation processes	The absence of real-time alerts delayed detection; escalation occurred only after a retrospective review of missing data.
Internal organizational policies and procedures	System recovery and downtime procedures	Recovery depended on contacting IT services, indicating limited local recovery capability and unclear resilience to downtime.
External rules, regulations, and pressures	Patient safety and monitoring requirements	Continuous cardiac monitoring is a regulatory and clinical safety expectation; system failure undermined this requirement.
System measurement and monitoring	System health visibility and alerting	No automated alerts indicated loss of server connectivity or data capture failure, contributing to delayed detection.
System configuration and upgrades	Configuration resilience	Centralized system architecture lacked redundancy or fail-safe mechanisms to maintain monitoring during server outages.
Outcomes and consequences	Potential patient harm	Risk of missed or delayed detection of life-threatening cardiac events affecting multiple patients simultaneously.

HFACS-Healthcare Classification

The HFACS-Healthcare analysis demonstrated that the incident was not associated with unsafe acts by frontline staff. At the level of preconditions for unsafe acts, technological environment factors such as poor system feedback and centralized system dependency reduced situational awareness and limited recovery options. At the unsafe supervision level, reliance on manual detection and the absence of real-time system oversight contributed to the delayed recognition of the failure.

At the organizational level, resource management and system design decisions, including limited redundancy and reliance on centralized IT support for recovery, were identified as key contributors. Overall, the HFACS-Healthcare analysis indicated that the incident arose primarily from latent system and organizational factors rather than from individual human error. The HFACS-Healthcare classification is summarized in Table 5.

XSL•FO
RenderX

Table . Classification of the incident according to the Human Factors Analysis and Classification System (HFACS)-Healthcare.

HFACS-Healthcare level and category	Classification	How it applies to this incident
Level 1: unsafe acts		
Errors	No unsafe acts identified	Clinical staff and the superuser acted appropriately and followed established restart procedures; no misuse or deviation was identified.
Level 2: preconditions for unsafe acts		
Technological environment	Poor system feedback and controllability	The surveillance system provided no feedback when restart commands were issued and became unresponsive, limiting users' ability to understand or manage the system state.
Physical/technological environment	Centralized system dependency	A single central surveillance system supported multiple patients, thereby increasing the impact when it failed.
Cognitive factors	Reduced situational awareness	Staff were unaware of the system failure until retrospective review, indicating diminished awareness of system status rather than clinical oversight.
Level 3: unsafe supervision		
Inadequate supervision	Insufficient monitoring of system health	There was no active oversight mechanism or real-time alerting to detect loss of server connectivity or data capture.
Planned inappropriate operations	Reliance on manual detection	Continued operation without automated system status alerts assumed that failures would be noticed through routine use.
Level 4: organizational influences		
Resource management	Limited redundancy and resilience	The surveillance architecture lacked redundancy or fail-safe mechanisms to maintain monitoring during server outages.
Organizational climate	Separation between the clinical and IT domains	Recovery depended on IT intervention, reflecting organizational boundaries that limited frontline recovery capacity.
Organizational processes	Downtime and recovery procedures	Restart instructions existed but were insufficient for nonresponsive system states, requiring escalation beyond clinical control.

Cross-Framework Synthesis

Across all 5 classification frameworks, the incident was consistently characterized as an HIT-driven system failure with delayed detection, multi-patient impact, and potential for severe harm. While each framework emphasized different aspects of the incident, all converged on the absence of frontline human error and the presence of latent technical, organizational, and design vulnerabilities. The combined use of the ICPS, HIT-CS, SEIPS 2.0, sociotechnical, and HFACS-Healthcare frameworks provided a comprehensive and complementary understanding of the incident and its patient safety implications. To facilitate comparison across frameworks and to synthesize key findings, Table 6 presents an alignment of major incident characteristics as identified by the ICPS, HIT-CS, SEIPS 2.0, Sittig and Singh sociotechnical model, and HFACS-Healthcare. Despite differences in terminology and analytical focus, the frameworks showed strong convergence in identifying the incident as a technology-driven failure with delayed detection, multi-patient impact, and predominant system-level and organizational contributors rather than frontline human error.

XSL•FO
RenderX

Table . Cross-framework alignment of key findings from the classification of a health information technology–related incident using the ICPS[a], HIT-CS[b], SEIPS 2.0[c], Sittig and Singh sociotechnical model, and HFACS[d]-Healthcare, highlighting convergent and complementary perspectives on incident nature, detection, human contribution, organizational factors, and impact.

Analytic theme	ICPS	HIT-CS	SEIPS 2.0	Sociotechnical (Sittig and Singh) model	HFACS-Healthcare
Primary incident nature	Documentation/information	System availability failure	Technology domain	Infrastructure failure	Preconditions (technological environment)
Detection	Delayed detection	Delayed detection	Process failure	Lack of system monitoring	Unsafe supervision
Human contribution	Not specified	Appropriate use	People domain	Human-computer interface	No unsafe acts
Organizational factors	Contributing factors	Workflow location	Organization	Policies and procedures	Organizational influences
Scale of impact	Potential severe harm	Multi-patient impact	Outcome	Centralized architecture	Resource management
Recovery	Mitigation actions	Recovery failure	Processes	IT dependency	Supervisory dependence

[a]ICPS: International Classification for Patient Safety.

[b]HIT-CS: Health Information Technology Classification System.

[c]SEIPS: Systems Engineering Initiative for Patient Safety.

[d]HFACS: Human Factors Analysis and Classification System.

Discussion

Principal Findings

This case report demonstrates that an HIT failure in a high-acuity clinical setting can pose a substantial patient safety risk despite appropriate frontline actions. Across all 5 applied classification frameworks—that is, ICPS, HIT-CS, SEIPS 2.0, the Sittig and Singh sociotechnical model, and HFACS-Healthcare—the incident was consistently characterized as a system-driven failure involving loss of monitoring information, delayed detection, and multi-patient impact. The absence of frontline human error identified across frameworks aligns with prior research showing that many HIT-related safety events arise from latent system conditions rather than individual mistakes [3,4,11,19].

Importantly, the incident should not be interpreted as a benign technical malfunction. In high-acuity cardiac settings, centralized surveillance serves as a safety layer to detect arrhythmias and sudden physiological deterioration. Interruption of this layer introduces a window of vulnerability during which time-critical events may be missed or responses delayed, underscoring the clinical seriousness of HIT availability failures. Incident report analyses have demonstrated that software-related challenges frequently reflect latent system conditions rather than isolated user errors, reinforcing the importance of sociotechnical interpretation of HIT failures [34].

The alignment of findings across frameworks (Table 6) illustrates how different taxonomies emphasize complementary dimensions of the same event. The ICPS provided a standardized patient safety classification [7,8], the HIT-CS enabled precise characterization of HIT-specific failure mechanisms [9,10], and the SEIPS, sociotechnical, and HFACS-Healthcare frameworks highlighted how interactions among technology, workflows, supervision, and organizational context shaped the emergence and impact of incidents [13,15,19].

Methodological Contribution

This study contributes methodologically by demonstrating the value of multi-framework incident classification. Prior work has noted that reliance on a single taxonomy may obscure important aspects of HIT-related risk [6,11]. In this case, the HIT-CS clarified technical failure modes; the SEIPS and sociotechnical models contextualized the disruption to clinical work; and the HFACS-Healthcare provided a structured explanation of how organizational and supervisory factors contributed to the disruption, without attributing blame to frontline staff [18,19].

The integrative alignment table (Table 6) supports synthesis across frameworks and reduces interpretive fragmentation. Similar alignment approaches have been recommended to strengthen learning from complex safety events and to improve translation of analytic findings into practice [22,24].

Technological Failures as System-Level Safety Events

Failure of centralized HIT should be interpreted through the lens of systemic safety governance, rather than isolated technical faults. Complex care, such as extracorporeal treatments, requires structured communication, protocols, and organizational support to mitigate clinical risk. Nalesso et al [35] emphasize the importance of proactive and reactive safety management in critical care settings, including multidisciplinary risk assessment, procedural checklists, and safety culture tools to reduce clinical incidents in high-risk technologies.

High-risk technologies in critical care demand governance approaches that extend beyond device functionality to include organizational coordination, communication, and monitoring infrastructure. In complex therapies such as extracorporeal blood

XSL•FO
RenderX

purification, safety governance includes structured protocols, feedback loops, and systematic risk analysis to prevent latent failures that could lead to patient harm [18,35]. Digital incident reporting systems themselves require quality governance to ensure meaningful learning from HIT failures. Prior work has demonstrated variability in documentation quality and analytic depth within digital reporting systems, underscoring the need for structured governance frameworks [28].

Remote Monitoring and Alerting: Benefits and Limitations

Remote monitoring and device control can improve patient care and staff safety, but such systems must be interpreted in context with cognitive load and alert management. Garzotto et al [36] describe how remote control of medical devices reduced risk and unnecessary interventions in the intensive care unit during the COVID-19 pandemic, highlighting operational benefits and the potential for remote surveillance to support safety in critical settings [37].

Remote monitoring and control of medical devices can reduce direct exposure risks and streamline clinical workflows, particularly during infectious outbreaks. Garzotto et al [36] reported that remote device control decreased unnecessary interventions and supported patient safety in intensive care environments [37,38]. However, as with all automated systems, trade-offs such as increased alert burden and clinician cognitive load must be considered to ensure that alerting improves situational awareness without inducing fatigue [37,38].

Telemedicine and Remote Surveillance Concepts

Although not specific to intensive care unit central surveillance, telemedicine and remote monitoring paradigms offer conceptual bridges for understanding distributed monitoring and alerting across care settings. Ricci and Ronco [39] describe telemedicine support for peritoneal dialysis, illustrating how continuous remote surveillance of treatment data provides clinical oversight and improves outcomes in chronic care. These principles parallel the reliability of centralized surveillance; both involve automated data capture, transmission, and interpretation to support timely clinical decisions [35,40,41].

Concepts from telemedicine and continuous remote monitoring provide relevant frameworks for ensuring the reliability of centralized surveillance. Remote monitoring of chronic therapies such as peritoneal dialysis involves continuous data transmission and clinician oversight, improving the detection of therapy deviations and enabling timely intervention [35,41,42].

IT and Safety in Critical Care Therapies

Early literature emphasizes that IT in critical care therapies (such as continuous renal replacement therapy) impacts patient safety, monitoring, and documentation. Ricci and Ronco [39] underscored that IT contributes to practice variation, patient assessment, monitoring, and documentation in critical care, highlighting the potential for technology to influence safety outcomes when integrated with governance and quality systems [4].

IT implementation in critical care therapies can enhance patient assessment, documentation, and therapeutic monitoring, but it also introduces new failure modes that must be governed within broader safety and quality infrastructures [4,37,39].

Implications for Clinical Practice

Early Detection of HIT Failures

Delayed detection emerged as a dominant theme across all classification approaches. Similar delays in recognizing HIT failures have been described in prior studies, particularly in settings that rely on retrospective review rather than real-time system monitoring [11,28]. These findings underscore the importance of implementing automated system status alerts and monitoring dashboards to support timely detection of connectivity loss, data capture failure, or system degradation [16,26].

From a practical perspective, health care organizations should consider the following:

- Deployment of automated system status alerts distinct from physiological alarms
- Real-time monitoring dashboards accessible to both clinical units and IT services
- Clear escalation pathways triggered by connectivity or data integrity failures

However, implementation of automated alerts must be carefully designed to avoid unintended consequences such as alert fatigue and increased cognitive load among clinical staff. Excessive or nonspecific alerts may desensitize users and reduce responsiveness, particularly in intensive care environments already characterized by high alarm burden [1,2]. Therefore, system status alerts for surveillance failures should be prioritized, clearly distinguishable from routine physiological alarms, and integrated into existing workflows to balance early detection with usability and clinician attention.

Managing Operational Vulnerabilities in HIT Environments

Similar operational disruptions following software modifications, including security patching, have been documented in health care settings, illustrating how even routine IT maintenance may introduce unintended workflow vulnerabilities [27]. This reinforces the need for structured change management processes and proactive risk assessment before implementing system updates in high-acuity environments.

Evidence from quality improvement interventions in radiology demonstrates that structured system redesign and multidisciplinary engagement are more effective than isolated procedural adjustments, further supporting system-level corrective strategies for HIT-related failures [43,44]. Accordingly, mitigation efforts should emphasize the following:

- Multidisciplinary planning for HIT upgrades and maintenance
- Prospective hazard analysis before system changes
- Integration of IT governance with clinical safety leadership

XSL•FO
RenderX

Recovery Capacity and Local Resilience

The case also highlights limitations in frontline recovery capacity. Despite appropriate use by trained superusers, recovery required IT intervention, reflecting a lack of graduated recovery mechanisms at the clinical unit level. Human factors research emphasizes that resilient systems should support adaptation and recovery close to the point of failure, rather than relying solely on centralized escalation [22,23].

Designing HIT systems that provide meaningful feedback during recovery attempts and clearly guide escalation pathways may reduce downtime and associated patient risk. Practical measures may include the following:

- Transparent system status indicators during restart attempts
- Tiered recovery protocols with defined thresholds for escalation
- Simulation-based training for downtime scenarios

Governance, Redundancy, and Architectural Risk

From an organizational perspective, the incident illustrates the safety implications of centralized HIT architectures without redundancy. Central surveillance systems offer efficiency but can amplify harm when failures occur, a phenomenon previously described in analyses of large-scale HIT incidents [26,28]. Explicit assessment of single points of failure and investment in redundancy or fail-safe monitoring strategies are therefore critical in intensive care environments.

Broader analyses of digital technology implementation during crisis conditions have demonstrated that technological reliability, equity, and organizational readiness are critical determinants of safe digital transformation in health care systems [45,46]. These findings support the need for the following:

- System-level governance structures overseeing HIT safety
- Regular auditing of architectural vulnerabilities
- Investment in redundancy for high-acuity monitoring systems
- Integration of clinical, technical, and leadership oversight

Translation of Multi-Framework Findings Into Action

Building on the findings from the multi-framework analysis and the implications outlined above, a set of corrective and preventive strategies were devised to mitigate the risk of similar HIT-related surveillance failures in high-acuity care settings (see Textbox 2). These strategies were developed by synthesizing insights from the ICPS, HIT-CS, SEIPS 2.0, the sociotechnical model of Sittig and Singh, and HFACS-Healthcare, with an emphasis on system-level resilience, early detection, recovery capacity, and organizational coordination rather than on individual performance.

XSL•FO
RenderX

Textbox 2. Preventive and corrective strategies to mitigate health information technology–related central surveillance failures.

System reliability and infrastructure resilience

- Establish redundancy for central surveillance systems, including failover servers or mirrored systems, to reduce the impact of single points of failure in high-acuity settings.
- Implement automatic buffering or local storage of monitoring data to prevent data loss during temporary server connectivity disruptions.
- Ensure routine stress testing and resilience testing of surveillance systems under peak load and failure scenarios.

System monitoring, detection, and alerting

- Implement real-time system health monitoring with automated alerts for loss of server connectivity, data transmission failure, or interruption of data storage.
- Ensure that system status indicators are clearly visible to clinical users, indicating whether monitoring data are actively being captured and stored.
- Configure alerts to notify both clinical units and IT services simultaneously to enable parallel awareness and response.

Recovery procedures and contingency planning

- Develop and validate clear, stepwise recovery procedures for surveillance system failures, including criteria for escalation to IT support.
- Design restart and recovery functions that provide explicit user feedback (eg, confirmation messages or error notifications) to avoid ambiguity during recovery attempts.
- Formalize fallback monitoring workflows during downtime, such as increased bedside monitoring or alternative documentation processes, to maintain patient surveillance.

Human factors and training

- Provide regular training for clinical staff and designated superusers on recognition of abnormal system behavior, limits of local recovery actions, and early escalation pathways.
- Conduct interdisciplinary simulation or tabletop exercises involving clinical staff and IT personnel to rehearse detection, response, and recovery from HIT failures.
- Reinforce a nonpunitive reporting culture to encourage early reporting of HIT anomalies and near-misses.

Organizational governance and coordination

- Clearly define roles and responsibilities between clinical units, IT services, and vendors for surveillance system maintenance, monitoring, and incident response.
- Integrate HIT failure scenarios into organizational risk management and patient safety governance structures.
- Ensure that timely communication mechanisms are in place to inform all affected clinical staff of unexpected system downtimes and recovery status.

System design and integration

- Design surveillance systems to align with clinical workflows, minimizing reliance on retrospective data review for detection of system failures.
- Ensure interoperability between bedside monitoring devices and central surveillance systems to support continuity of monitoring during partial system failures.
- Incorporate human-centered design principles to improve system transparency, controllability, and usability in safety-critical situations.

The proposed strategies are grouped into key sociotechnical domains to facilitate practical implementation and learning.

Strengths and Limitations of the Study

A key strength of this study is the systematic application of 5 established classification frameworks to a single HIT-related incident, enabling a comprehensive and triangulated analysis. This approach aligns with calls for richer sociotechnical analyses of HIT safety events [15,16]. Additionally, focusing on an incident with high potential severity but no confirmed harm supports proactive safety learning, consistent with modern patient safety principles [5,23].

Several limitations should be acknowledged. The analysis is based on a single incident report, limiting generalizability and precluding causal inference. Patient outcomes could not be fully assessed due to missing monitoring data, which constrained the classification of harm under ICPS [7]. Furthermore, while the frameworks used are well established, their application involves interpretive judgment; alternative analysts might emphasize different dimensions. Nevertheless, the strong convergence of findings across frameworks suggests robust analytical results.

In addition, the reconstruction of the event relied primarily on the narrative documented in the incident report; it is possible that relevant contextual factors were not recorded and therefore could not be captured within the analytic frameworks applied.

XSL•FO
RenderX

Implications for Future Research

Future research should examine whether similar patterns of delayed detection, limited recovery capacity, and organizational dependency are observed across larger samples of HIT-related incidents using comparable multi-framework approaches [11,28]. Comparative studies evaluating how different classification systems influence safety recommendations could further guide methodological choices in HIT safety research.

There is also a need for empirical evaluation of design and governance interventions, such as real-time HIT health monitoring, redundancy strategies, and improved IT-clinical coordination models, to assess their effectiveness in reducing risk in high-acuity settings [16,26,29]. Finally, integrating resilience-oriented concepts such as Safety-II into HIT safety analysis may offer additional insights into how systems can better support adaptive clinical work under failure conditions [22,23].

Conclusions

This case report demonstrates that HIT-related safety incidents in intensive care settings are best understood as emergent properties of complex sociotechnical systems rather than isolated technical malfunctions or individual errors. The combined application of ICPS, HIT-CS, SEIPS 2.0, the sociotechnical model, and HFACS-Healthcare revealed convergent system-level vulnerabilities involving delayed detection, limited recovery capacity, and architectural and organizational design factors. These findings underscore the importance of governance structures, redundancy planning, and human-centered alerting strategies in high-acuity monitoring environments. Integrative, multi-framework analysis may strengthen learning from HIT failures and support safer design, implementation, and resilience of clinical surveillance systems.

Acknowledgments

The author thanks the health care professionals and medical technology staff for their contributions to incident reporting and investigation, which made this analysis possible. The author especially acknowledges Dr Abdullah Hassoun for his assistance in reviewing and verifying the translated incident narrative. The author also recognizes the important role of national incident reporting systems in facilitating learning from health information technology–related safety events.

Artificial intelligence–based tools were used to assist with language refinement and manuscript organization. The author reviewed and verified all content and takes full responsibility for the accuracy and integrity of the work.

Funding

No specific funding was received for this study.

Data Availability

The data analyzed in this study consist of a deidentified incident report obtained from a national incident reporting repository. For ethical and confidentiality reasons, the underlying data are not publicly available. Reasonable requests for access to the deidentified data may be considered by the corresponding author, subject to applicable data governance and institutional policies.

Authors' Contributions

MSRJ developed the protocol, obtained ethical approval, and drafted the manuscript. MSRJ performed data analysis and devised the set of strategies. MSRJ reviewed and edited the manuscript and approved the final version.

Conflicts of Interest

None declared.

References

1. Cvach M. Monitor alarm fatigue: an integrative review. Biomed Instrum Technol 2012;46(4):268-277. [doi: 10.2345/0899-8205-46.4.268] [Medline: 22839984]
2. Drew BJ, Harris P, Zègre-Hemsey JK, et al. Insights into the problem of alarm fatigue with physiologic monitor devices: a comprehensive observational study of consecutive intensive care unit patients. PLoS One 2014;9(10):e110274. [doi: 10.1371/journal.pone.0110274] [Medline: 25338067]
3. Magrabi F, Ong MS, Runciman W, Coiera E. Using FDA reports to inform a classification for health information technology safety problems. J Am Med Inform Assoc 2012;19(1):45-53. [doi: 10.1136/amiajnl-2011-000369] [Medline: 21903979]
4. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc 2004;11(2):104-112. [doi: 10.1197/jamia.M1471] [Medline: 14633936]
5. Bates DW, Gawande AA. Improving safety with information technology. N Engl J Med 2003 Jun 19;348(25):2526-2534. [doi: 10.1056/NEJMsa020847] [Medline: 12815139]

XSL•FO
RenderX

6. Jabin MSR. Identifying and characterising problems arising from interactions between medical imaging and health information technology as a basis for improvements in practice [Dissertation]. : University of South Australia; 2019.
7. Runciman WB, Williamson JAH, Deakin A, Benveniste KA, Bannon K, Hibbert PD. An integrated framework for safety, quality and risk management: an information and incident management system based on a universal patient safety classification. Qual Saf Health Care 2006;15(Suppl 1):i82-i90. [doi: 10.1136/qshc.2005.017467] [Medline: 17142615]
8. Conceptual framework for the international classification for patient safety: final technical report. : World Health Organization; 2010 URL: https://iris.who.int/handle/10665/70882 [accessed 2025-01-15]
9. McElroy LM, Woods DM, Yanes AF, et al. Applying the WHO conceptual framework for the International Classification for Patient Safety to a surgical population. Int J Qual Health Care 2016;28(2):166-174. [doi: 10.1093/intqhc/mzw001] [Medline: 26803539]
10. Magrabi F, Baker M, Sinha I, et al. Clinical safety of England's national programme for IT: a retrospective analysis of all reported safety events 2005-2011. Int J Med Inform 2015 Mar;84(3):198-206. [doi: 10.1016/j.ijmedinf.2014.12.003] [Medline: 25617015]
11. Jabin MSR, Magrabi F, Hibbert P, Schultz T, Bessen T, Runciman W. Identifying and characterizing system issues of health information technology in medical imaging as a basis for recommendations. In: 2019 IEEE International Conference on Imaging Systems and Techniques (IST): IEEE; 2019:1-6. [doi: 10.1109/IST48021.2019.9010426]
12. Coiera E, Magrabi F. Information system safety. In: Coiera E, editor. Guide to Health Informatics, 3rd edition: CRC Press; 2015:195-220. [doi: 10.1201/b13617]
13. Pan D, Nilsson E, Jabin MSR. A review of incidents related to health information technology in Swedish healthcare to characterise system issues as a basis for improvement in clinical practice. Health Informatics J 2024;30(3):14604582241270742. [doi: 10.1177/14604582241270742] [Medline: 39116887]
14. Jabin MSR, Magrabi F, Hibbert P, Schultz T, Runciman W. Identifying and classifying incidents related to health information technology in medical imaging as a basis for improvements in practice. In: 2019 IEEE International Conference on Imaging Systems and Techniques (IST): IEEE; 2019. URL: https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=8976373 [accessed 2026-04-01] [doi: 10.1109/IST48021.2019.9010109]
15. Jabin MSR. The need for a refined classification system and national incident reporting system for health information technology-related incidents. Front Digit Health 2024;6:1422396. [doi: 10.3389/fdgth.2024.1422396] [Medline: 39131183]
16. Holden RJ, Carayon P, Gurses AP, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 2013;56(11):1669-1686. [doi: 10.1080/00140139.2013.838643] [Medline: 24088063]
17. Carayon P, Wooldridge A, Hoonakker P, Hundt AS, Kelly MM. SEIPS 3.0: Human-centered design of the patient journey for patient safety. Appl Ergon 2020 Apr;84:103033. [doi: 10.1016/j.apergo.2019.103033] [Medline: 31987516]
18. Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Qual Saf Health Care 2010 Oct;19 Suppl 3(Suppl 3):i68-i74. [doi: 10.1136/qshc.2010.042085] [Medline: 20959322]
19. Singh H, Sittig DF. Measuring and improving patient safety through health information technology: the Health IT Safety Framework. BMJ Qual Saf 2016 Apr;25(4):226-232. [doi: 10.1136/bmjqs-2015-004486] [Medline: 26369894]
20. Ash JS, Sittig DF, Dykstra RH, Guappone KP, Carpenter JD, Seshadri V. Categorizing the unintended sociotechnical consequences of computerized provider order entry. Int J Med Inform 2007 Jun;76(Suppl 1):S21-S27. [doi: 10.1016/j.ijmedinf.2006.05.017] [Medline: 16793330]
21. Holden RJ, Scanlon MC, Patel NR, et al. A human factors framework and study of the effect of nursing workload on patient safety and employee quality of working life. BMJ Qual Saf 2011 Jan;20(1):15-24. [doi: 10.1136/bmjqs.2008.028381] [Medline: 21228071]
22. Diller T, Helmrich G, Dunning S, Cox S, Buchanan A, Shappell S. The Human Factors Analysis Classification System (HFACS) applied to health care. Am J Med Qual 2014;29(3):181-190. [doi: 10.1177/1062860613491623] [Medline: 23814026]
23. Shappell S, Detwiler C, Holcomb K, Hackworth C, Boquet A, Wiegmann DA. Human error and commercial aviation accidents: an analysis using the human factors analysis and classification system. Hum Factors 2007 Apr;49(2):227-242. [doi: 10.1518/001872007X312469] [Medline: 17447665]
24. Coiera E. Why e-health is so hard. Med J Aust 2013 Mar 4;198(4):178-179. [doi: 10.5694/mja13.10101] [Medline: 23451947]
25. What the act says. Etikprövningsmyndigheten. 2003. URL: https://etikprovningsmyndigheten.se/en/what-the-act-says [accessed 2026-04-02]
26. Jabin MSR, Wepa D, Hassoun A. A case report of system configuration issue in medical imaging due to system upgrade-changes in hardware and software. Front Digit Health 2024;6:1371761. [doi: 10.3389/fdgth.2024.1371761] [Medline: 39347445]
27. Jabin MSR. Operational disruption in healthcare associated with software functionality issue due to software security patching: a case report. Front Digit Health 2024;6:1367431. [doi: 10.3389/fdgth.2024.1367431] [Medline: 38550716]
28. Jabin MSR, Steen M, Wepa D, Bergman P. Assessing the healthcare quality issues for digital incident reporting in Sweden: Incident reports analysis. Digit Health 2023;9:20552076231174307. [doi: 10.1177/20552076231174307] [Medline: 37188073]

XSL•FO
RenderX

29. Negash B, Katz A, Neilson CJ, et al. De-identification of free text data containing personal health information: a scoping review of reviews. Int J Popul Data Sci 2023;8(1):2153. [doi: 10.23889/ijpds.v8i1.2153] [Medline: 38414537]
30. Jabin MSR, Pan D, Nilsson E. Characterizing healthcare incidents in Sweden related to health information technology affecting care management of multiple patients. Health Informatics J 2022;28(2):14604582221105440. [doi: 10.1177/14604582221105440] [Medline: 35762538]
31. Jabin MSR, Pan D, Nilsson E. Characterizing patient details-related challenges from health information technology-related incident reports from Swedish healthcare. Front Digit Health 2024;6:1260521. [doi: 10.3389/fdgth.2024.1260521] [Medline: 38380372]
32. Roggow BJ, Zarei E. Human Factors Analysis and Classification System (HFACS): development, tiers, nanocodes, application, and adaptations. In: Safety Causation Analysis in Sociotechnical Systems: Advanced Models and Techniques: Springer; 2024:151-180. [doi: 10.1007/978-3-031-62470-4_7]
33. Jalali M, Dehghan H, Habibi E, Khakzad N. Application of "Human Factor Analysis and Classification System" (HFACS) model to the prevention of medical errors and adverse events: a systematic review. Int J Prev Med 2023;14:127. [doi: 10.4103/ijpvm.ijpvm_332_22] [Medline: 38264566]
34. Jabin MSR, Pan D. Software-related challenges in Swedish healthcare through the lens of incident reports: a desktop study. Digit Health 2023;9:20552076231203600. [doi: 10.1177/20552076231203600] [Medline: 37744748]
35. Nalesso F, Garzotto F, Martello T, et al. The patient safety in extracorporeal blood purification treatments of critical patients. Front Nephrol 2022;2:871480. [doi: 10.3389/fneph.2022.871480] [Medline: 37675020]
36. Garzotto F, Comoretto RI, Ostermann M, et al. Preventing infectious diseases in intensive care unit by medical devices remote control: lessons from COVID-19. J Crit Care 2021 Feb;61:119-124. [doi: 10.1016/j.jcrc.2020.10.014] [Medline: 33157307]
37. Magrabi F, Ong MS, Runciman W, Coiera E. Patient safety problems associated with heathcare information technology: an analysis of adverse events reported to the US Food and Drug Administration. AMIA Annu Symp Proc 2011;2011:853-857. [Medline: 22195143]
38. Tripodi A, Moia M. The accuracy of the International Normalized Ratio and the American College of Chest Physicians recommendations on the use of vitamin K to reverse over‐anticoagulation. J Thromb Haemost 2012 Oct;10(10):2207-2208. [doi: 10.1111/j.1538-7836.2012.04879.x] [Medline: 22882756]
39. Ricci Z, Ronco C. Information technology for CRRT and dose delivery calculator. Contrib Nephrol 2007;156:197-202. [doi: 10.1159/000102084] [Medline: 17464128]
40. Claggett J, Petter S, Joshi A, Ponzio T, Kirkendall E. An infrastructure framework for remote patient monitoring interventions and research. J Med Internet Res 2024 May 30;26:e51234. [doi: 10.2196/51234] [Medline: 38815263]
41. Jansen AJS, Peters GM, Kooij L, Doggen CJM, van Harten WH. Device based monitoring in digital care and its impact on hospital service use. NPJ Digit Med 2025 Jan 8;8(1):16. [doi: 10.1038/s41746-024-01427-8] [Medline: 39779761]
42. Safavi KC, Driscoll W, Wiener-Kronish JP. Remote surveillance technologies: realizing the aim of right patient, right data, right time. Anesth Analg 2019 Sep;129(3):726-734. [doi: 10.1213/ANE.0000000000003948] [Medline: 31425213]
43. Jabin MSR, Schultz T, Mandel C, et al. A mixed-methods systematic review of the effectiveness and experiences of quality improvement interventions in radiology. J Patient Saf 2022;18(1):e97-e107. [doi: 10.1097/PTS.0000000000000709] [Medline: 32433438]
44. Jabin SR, Schultz T, Hibbert P, Mandel C, Runciman W. Effectiveness of quality improvement interventions for patient safety in radiology: a systematic review protocol. JBI Database System Rev Implement Rep 2016 Sep;14(9):65-78. [doi: 10.11124/JBISRIR-2016-003078] [Medline: 27755318]
45. Wepa D. Digital disparities: tech solutions for indigenous communities. Open Access Gov 2025;46(1):26-27. [doi: 10.56367/OAG-046-11905]
46. Wepa D, Thomas S, Jabin MSR. The experience and impact of digital technologies on Indigenous populations in New Zealand during the COVID-19 pandemic and Cyclone Gabrielle: The Kaupapa Māori methodology. JMIR Aging 2025 Oct 28;8:e73974. [doi: 10.2196/73974] [Medline: 41150869]

Abbreviations

HFACS: Human Factors Analysis and Classification System
HFACS-Healthcare: HFACS for health care
HIT: health information technology
HIT-CS: Health Information Technology Classification System
ICPS: International Classification for Patient Safety
SEIPS: Systems Engineering Initiative for Patient Safety

XSL•FO
RenderX

Edited by A Kushniruk; submitted 31.Jan.2026; peer-reviewed by C Recsky, F Nalesso; revised version received 01.Mar.2026; accepted 10.Mar.2026; published 09.Apr.2026.

Please cite as:
Jabin MSR
Health Information Technology–Related Loss of Central Surveillance Data in a Heart Intensive Care Unit: Multi-Framework Case Report
JMIR Hum Factors 2026;13:e92560
URL: https://humanfactors.jmir.org/2026/1/e92560
doi:10.2196/92560

© Md Shafiqur Rahman Jabin. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 9.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Dynamic Balance Control and Postural Adaptation in Human-Robot Collaborative Manipulation: Within-Subject Experimental Study

Alessia de Nobile[1], MSc; Daniele Bibbo[1], PhD; Simone Ranaldi[1], PhD; Maurizio Schmid[1], PhD; Giovanni Corvini[2], PhD; Silvia Conforto[1], PhD

[1]Department of Industrial, Electronic and Mechanical Engineering, Roma Tre University, Via Vito Volterra, 62 - Corpo B, Roma, Italy

[2]Biorobotics Group, Spanish National Research Council (CSIC) Center for Automation and Robotics (CAR) CSIC-UPM, Madrid, Spain

Corresponding Author:
Daniele Bibbo, PhD
Department of Industrial, Electronic and Mechanical Engineering, Roma Tre University, Via Vito Volterra, 62 - Corpo B, Roma, Italy

Abstract

Background: The integration of robots into industrial settings has rapidly advanced, aiming to reduce human involvement in demanding tasks while improving overall efficiency. As collaborative robots (cobots) become more prevalent, assessing the physical strain during joint tasks is essential to promote long-term well-being in the workplace.

Objective: This study aimed to investigate how human-robot collaboration influences workers' postural control and musculoskeletal load during manipulation tasks performed in parallel.

Methods: Fourteen healthy male participants performed manipulation tasks under 3 conditions: without robotic assistance, with a cobot providing load support (Robot Free [RF]) and a cobot constrained to horizontal movement (Robot Plane [RP]). Center of pressure trajectories were computed, and nonlinear recurrence quantification analysis indicators (recurrence rate [REC], determinism [DET], and their ratio) were calculated in the anteroposterior, mediolateral, and anteroposterior-mediolateral planes

Results: Statistical analysis showed greater postural sway in robot-assisted conditions compared to Free. Mean distance increased from 1.7 (SD 0.6) cm in Free to 2.4 (SD 0.6) cm in RF (P<.001) and 2.3 (SD 0.6) cm in RP (P<.001). Mean velocity increased from 2.9 (SD 0.9) cm/s in Free to 4.3 (SD 1.4) cm/s in RF and RP. Confidence ellipse area increased from 7.6 (SD 4.1) cm^2 in Free to 24.9 (SD 14.2) cm^2 in RF and 23.1 (SD 13.4) cm^2 in RP. Sway area increased from 1.5 (SD 0.7) cm^2/s in Free to 2.9 (SD 1.2) cm^2/s in RF and RP. Nonlinear metrics revealed lower recurrence rates in robot-assisted conditions, decreasing from 0.31 (SD 0.08) in Free to 0.2 (SD 0.08) in RF and 0.2 (SD 0.04) in RP in the anteroposterior-mediolateral plane (P<.001), from 0.33 (SD 0.08) in Free to 0.28 (SD 0.07) in RF (P=.02) and 0.16 (SD 0.03) in RP (P=.007) in the mediolateral direction, and from 0.36 (SD 0.07) in Free to 0.3 (SD 0.06) in RF (P=.009) and 0.26 (SD 0.03) in RP (P<.001) in the anteroposterior direction. Determinism remained stable (values close to 1), leading to higher determinism-to-recurrence ratios for robot-assisted conditions, increasing from 3.41 (SD 0.87) in Free to 5.41 (SD 1.69) in RF and 5.51 (SD 1.11) in RP in the anteroposterior-mediolateral plane (P<.001), from 3.11 (SD 0.63) in Free to 3.64 (SD 0.72) in RF (P=.02) and 3.92 (SD 0.48) in RP (P=.007) in the mediolateral direction, and from 2.82 (SD 0.48) in Free to 3.47 (SD 0.57) in RF (P=.009) and 3.46 (SD 0.45) in RP (P<.001) in the anteroposterior direction. No significant differences were found between the robot-assisted conditions.

Conclusions: Interaction increases postural sway, indicating reduced stability and higher physical demand. This could reflect impaired balance or adaptation. Nonlinear analysis reveals that postural control remains structured. Results also suggest that the mere presence of the cobot is the primary driver of these postural changes.

(JMIR Hum Factors 2026;13:e79930) doi:10.2196/79930

KEYWORDS

human-robot collaboration; center of pressure; recurrence quantification analysis; biomechanical risk; ergonomics; HRC; COP; RQA

Introduction

Industry 4.0 has brought transformative changes to manufacturing, emphasizing the integration of industrial robots to minimize workers' involvement in repetitive tasks [1]. Robots are now widely used in industrial settings to assist or replace workers in physically demanding tasks, offering benefits such as lower error rates, higher productivity, lower costs, and

XSL•FO
RenderX

improved product quality. Recent advancements in robotics have also facilitated closer interactions between humans and robots, giving rise to human-robot collaboration (HRC), which has become a key focus in robotics research [2-4]. In HRC, humans and robots work side by side toward shared goals, engaging in continuous interactions where they adapt and respond to each other's actions [5]. This synergy enhances efficiency and productivity, as human abilities complement robotic precision, strength, and endurance [6]. However, while HRC helps both workers and companies, its impact on physical and cognitive well-being requires careful assessment.

Numerous studies have explored how HRC affects workers, focusing on various aspects such as ergonomics, cognitive workload, safety, and overall well-being. Evaluating physical and mental demands in collaborative settings has been highlighted as a key factor [7], as ergonomic risk factors and comfort directly impact workers' long-term health [8,9]. Although collaborative robots (cobots) aim to reduce physical strain by assisting workers in repetitive or strenuous tasks, they do not completely prevent work-related musculoskeletal disorders (WMSDs), which may arise from improper postures or behaviors [10]. WMSDs are a major concern in industries with manual labor and repetitive motions, potentially leading to chronic pain, reduced productivity, and long-term health problems [11,12]. When work tasks are performed in quasi-static conditions, direct measures of posture could reveal insights into the stability and balance of workers and can thus be used to improve ergonomics during HRC [13]. Poor or sustained postures can cause musculoskeletal imbalances, increased strain on joints and muscles, and altered movement patterns over time [14]. Posturography data could help evaluate musculoskeletal load, identify ergonomic risk factors, and develop strategies to mitigate them.

Traditionally, sensory-motor and cognitive mechanisms involved in balance control are studied using conventional center of pressure (COP) measures. COP represents the point where the resultant ground reaction force is applied during standing or movement, and its displacement over time provides valuable insights into postural stability and balance. Standard COP metrics include mean amplitude and mean distance (reflecting the extent of the postural sway), mean velocity (representing the speed of postural adjustments), and confidence ellipse area and sway area (characterizing the distribution of the sway) [15]. These parameters are crucial for understanding how the body maintains equilibrium, particularly during tasks that involve dynamic movement or changes in posture. Additionally, variations in COP data can help identify deviations from normal posture, which may indicate potential risks for musculoskeletal discomfort or injury.

However, some researchers have argued that traditional analyses of the spatial dynamics of postural activity are insufficient to fully understand the mechanisms underlying postural control and, in particular, cannot provide a complete understanding of stability and instability in the control of posture [16,17]. These measures offer an overall view of postural sway and may be linked to the adopted muscular strategy, but they do not directly refer to the adopted control strategy or to the attentional demand allocated to the control of balance. Specifically, these traditional parameters primarily focus on linear characteristics of postural control and may not fully capture the complexity of human balance regulation. This limitation suggests that nonlinear analysis could complement traditional measures of the extent and distribution of postural sway, offering a more comprehensive perspective on postural dynamics [18]. Deepening understanding of the mode of control and its underlying mechanisms has encouraged researchers to call for more powerful and relevant techniques of analysis [19].

Among the indicators recently introduced to capture distinct aspects of the nonlinear temporal dynamics of time series data, recurrence quantification analysis (RQA) has been hypothesized as a viable method to capture aspects linked to balance maintenance [20-26]. RQA describes the temporal structure of a dynamic system by quantifying the number and duration of repeated states [27]. Crucially, RQA does not merely describe the magnitude or variability of movement, as traditional COP metrics do, but reveals how postural adjustments evolve. This includes capturing aspects such as the repetitiveness of the displacement trajectory, the complexity of the control strategy, and the system's predictability or adaptability [18], also in conditions where balance maintenance is challenged by the presence of repetitive supra-postural tasks [28]. Some of the most commonly used RQA parameters are the relative amount of recurrence and determinism [18].

Hence, traditional COP parameters offer a general overview of gesture characteristics, while RQA provides complementary insight by capturing the temporal structure and regularity of postural adjustments. Accordingly, this study combines both types of metrics to assess postural control and biomechanical risk during the execution of manipulatory tasks performed with and without robotic assistance. The experimental task followed a structured sequence consisting of rest, object grasping, object manipulation, object release, and rest. For this analysis, we focused only on the manipulation phase, when the participant actively interacted with the object, while grasping, release, and rest periods were excluded. This approach allows for a specific assessment of postural control precisely during sustained object handling. Unlike previous studies [13,29], which focused on sequential task execution with a cobot, this work examines a parallel task configuration, where the human and the cobot interact simultaneously, sharing the same space and concurring on the same task [30], thus introducing fundamentally different interaction dynamics. Accordingly, the goal of this work is twofold: first, to investigate whether the assistance of the cobot influences postural control and musculoskeletal load compared to performing the task alone; and second, to explore whether varying the mode of robotic collaboration within parallel tasks elicits distinct biomechanical responses. Ultimately, this approach aims to facilitate real-time monitoring of workers' physical strain and support the development of ergonomic strategies to reduce WMSDs and enhance workplace safety.

Methods

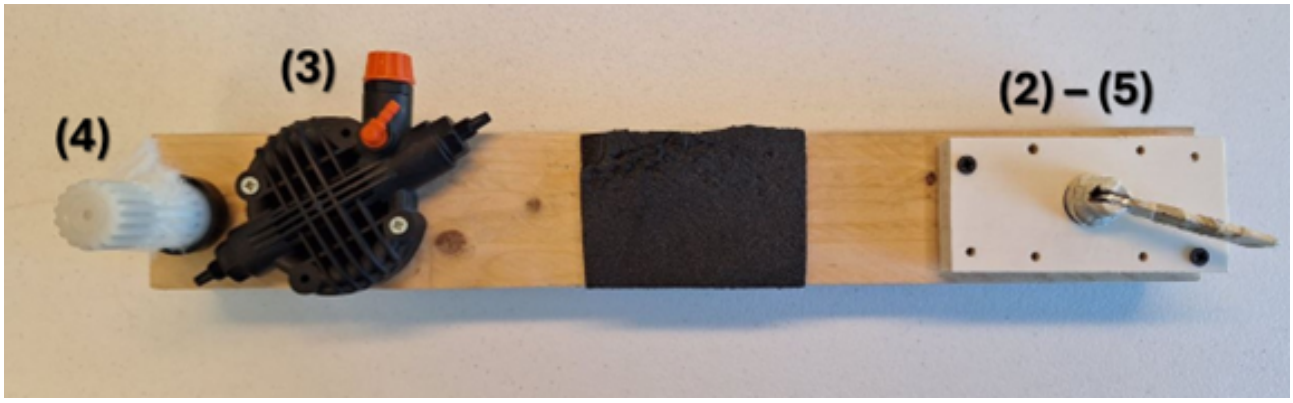
Experimental Setup

The experimental setup was designed using a 2 × 1 m table on which a cobot (Franka Emika Panda, Munich, Germany) was

XSL•FO
RenderX

positioned opposite the participant, who stood in an upright posture, to simulate a collaborative workbench environment. Participants were instructed to cyclically perform a manipulation task using a custom-designed tool (Figure 1), specifically developed to simulate common and functionally relevant manipulation activities. The tool is a wooden beam equipped with 4 interactive mechanisms: a white knob, a rotating component with a high-positioned orange knob, a black padded section designed for grip, and a white lever mounted on a rectangular structure.

Figure 1. Object designed to be manipulated in the task. The numbered labels identify the items to be manipulated at each corresponding step of the task sequence.



The manipulation task consisted of the following steps: (1) take the object, (2) open the lever on the right by rotating it 90° counterclockwise, (3) rotate the high-positioned orange knob until it reaches the end of its stroke, (4) unscrew the white knob on the left for 2 full turns, (5) close the lever that was opened at step 2, and (6) return the object to its original position. After completing this sequence, participants rested with their arms at their sides. Following a brief pause, they independently decided when to initiate the next cycle of manipulation. The experiment included three conditions:

1. Free: the participant performed the task independently. The object was positioned on an elevated support on the right side of the table for easy access. Using their right hand, the participant grasped the object and executed a predefined sequence of movements. The task was considered complete once the object was returned to its initial position.
2. Robot Free (RF): the cobot assisted the participant by securely holding the object at its central part, simulating a "gravity compensation" condition, thus minimizing the load the participant was required to support. In this condition, the cobot could move freely in response to the participant's needs. The task began with the cobot holding the object, after which the participant approached and performed the required sequence of movements. The task ended when the participant guided the cobot back to its starting point on their right.
3. Robot Plane (RP): similar to the RF condition, the cobot securely held the object at its central part. However, in this condition, the robot's movement was constrained to a virtual horizontal plane, locking the object's vertical movement and allowing free movement in other directions. The task shared the same starting and ending configurations as the RF condition. The height of the simulated plane was kept fixed for all experiments, regardless of the participant.

The inclusion of 2 distinct robot-assisted conditions was motivated by the hypothesis that varying the mode of collaboration could elicit different effects on human motor behavior and postural control. Specifically, the RP condition implements a gravity-compensation scenario in which the object is virtually weightless, and its vertical position is fixed. This constraint results in a different interaction with the cobot, while in the RF condition, participants need to continuously support and stabilize the cobot with 1 hand while performing the task with the other. In the RP condition, they can rely on the cobot's fixed height, potentially leading to the adoption of different motor strategies. Moreover, the inability to move the robot vertically may require participants to adapt their posture or reposition themselves to comfortably perform specific movements. This distinction between RF and RP thus allows the researchers to examine how different kinematic constraints and interaction modes with the cobot influence postural control and task execution. An illustrative photo of an individual performing the task under robot-assisted conditions is shown in Figure 2.

Participants were free to move naturally in all conditions, with the only requirement being the use of their right hand to grasp the object. Each condition was repeated 3 times, with each trial consisting of 5 repetitions of the manipulation task. The trials were conducted in a randomized order to minimize potential learning effects and ensure natural task execution.

XSL•FO
RenderX

Figure 2. Example of a manipulation task with robotic assistance.



Data Acquisition and Processing

Fourteen healthy right-handed male participants took part in the study (age: mean 26, SD 3 years; height: mean 177, SD 6 cm; and weight: mean 71, SD 7 kg). Participants were recruited based on their nonfamiliarity with the task and the absence of any known musculoskeletal, neurological, or visual impairments. The study was approved by the local ethical committee of Roma Tre University.

Two 6-component force platforms (BTS P-6000 and BTS Bioengineering) were positioned under the participants' feet to record ground reaction forces and to derive the COP coordinates. Force data were collected at a sampling frequency of 500 Hz. Each platform was positioned under 1 foot. Additionally, a marker-based optoelectronic system (BTS SMART-DX 6000 and BTS Bioengineering) was used to segment and identify each task repetition, with passive markers placed at key anatomical landmarks. These kinematic data were acquired using a sampling frequency of 250 Hz. Force platform data were then resampled at the same rate.

Data Processing and Analysis

Trajectories of the COP for each of the lower limbs (ie, COP-right and COP-left) were extracted from the force platform data. The total COP was then obtained as the weighted average between COP-right and COP-left, and its displacement was defined by both the anteroposterior and the mediolateral coordinates with respect to a reference system centered between the 2 force plates. Cycles of the movement were identified using motion capture data: segmentation was made by detecting the time instant when the marker on the right wrist reached the maximum rightward excursion in the mediolateral direction, indicating the completion of a repetition. Only manipulation phases were extracted for the analysis, as the focus was specifically on these active interactions, excluding the other phases (ie, rest, object pick-up, and object release).

XSL•FO
RenderX

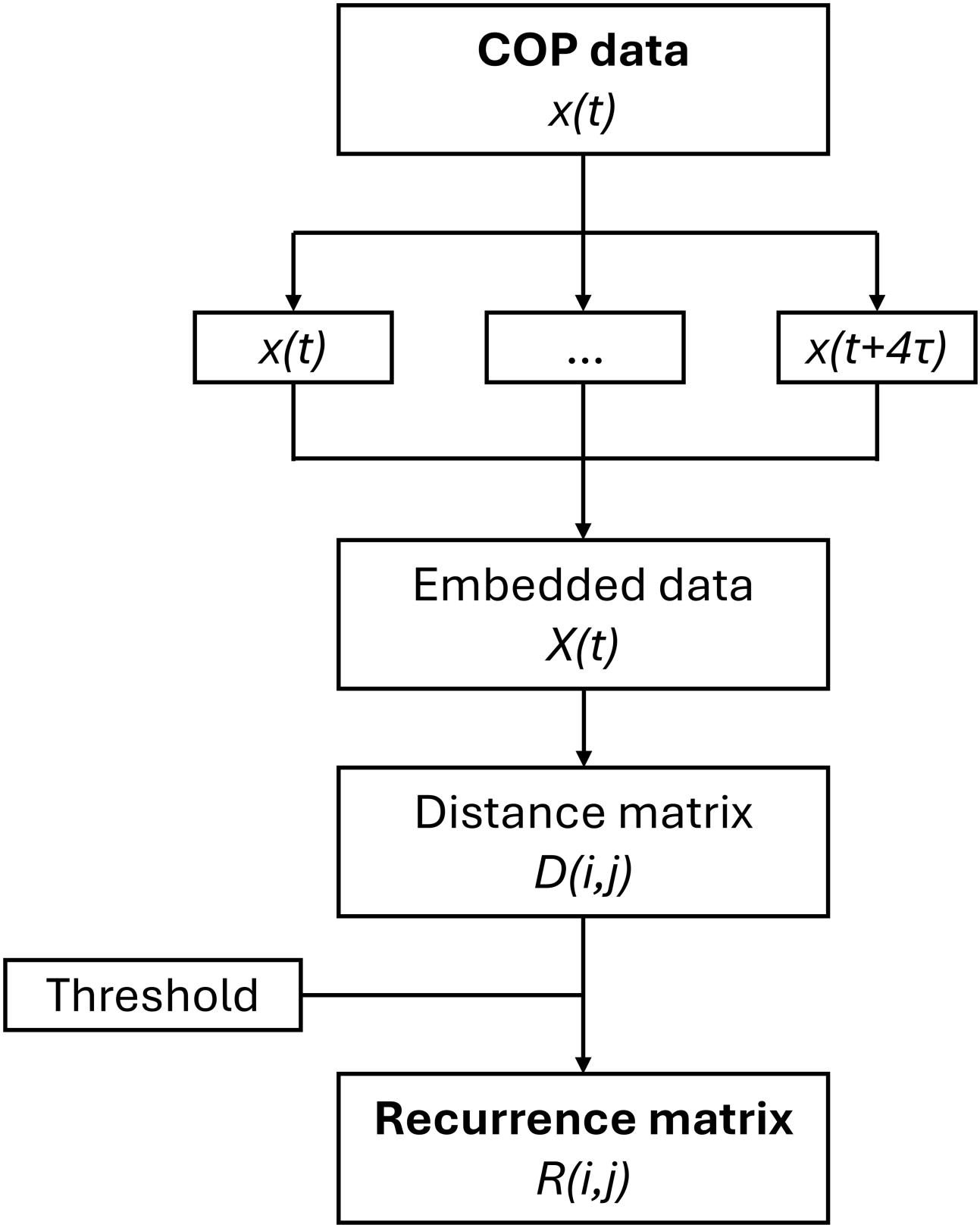
Four COP-derived sway parameters were selected to evaluate postural stability and quantify balance changes: mean distance, mean velocity, 95% confidence ellipse area, and sway area. These metrics were chosen based on their ability to capture distinct yet complementary aspects of postural control and stability. Mean distance quantifies the extent of postural displacement, reflecting the range of movement and how far the COP deviates from a reference point. Mean velocity measures the speed of postural displacement, indicating how quickly the individual adjusts their posture. The ellipse area captures the spread of the COP, providing insights into postural control by assessing variability and dispersion. This metric was preferred over the confidence circle area, as the elliptical estimate accounts for the directionality of sway—a relevant feature when COP trajectories are not uniformly distributed but follow a predominant path [13]. The sway area provides a measure of the extent of postural instability during task execution. For each cycle of the task repetition in each experimental condition, these parameters were calculated as defined in [30]. Mean values were then computed and subjected to inferential statistical analysis to assess potential differences across conditions.

In addition to these conventional analyses, an RQA was performed to explore the underlying dynamical patterns of postural sway. These measures provide insight into the temporal organization, predictability, and adaptability of postural control—aspects that are especially relevant in dynamic and interactive tasks—and are not influenced by participants' anthropometric characteristics. To investigate dynamics on a cycle-based time reference, the time delay was set to the average length of the cycle evaluated for each trial. The processing steps adopted for extracting the recurrence maps, similar to those used in [29], are reported in Figure 3.

RQA was carried out on the radial component of the COP, as well as on the anteroposterior and mediolateral components alone. Specifically, 3 parameters were calculated to analyze the dynamics of postural sway [29]. The recurrence rate (REC), also known as the density of recurrence points in the recurrence plot [18], quantifies the percentage of recurrent states in the COP trajectory, reflecting the overall stability of postural sway. Determinism indicates the proportion of those recurrence points that form structured diagonal lines, which is associated with the predictability and regularity of sway patterns. The ratio, defined as the ratio between determinism and REC, represents the extent to which the recurring patterns are deterministic in nature.

XSL•FO
RenderX

Figure 3. Flowchart for the embedding of center of pressure (COP) data and the extraction of the recurrence matrix.



Statistical Analysis

Normality of sway parameters and RQA indicators was assessed using the Shapiro-Wilk test. Since at least 1 of the distributions of each parameter was not normal, nonparametric statistical analyses were applied. To evaluate differences in both traditional and RQA-based parameters across the 3 experimental conditions, a Friedman test was performed. When a significant main effect was found, pairwise comparisons were conducted using the Wilcoxon rank-sum test. The significance level was set to 0.05, 0.01, and 0.001 (indicated by *, **, and ***, respectively). Effect sizes were calculated to quantify the magnitude of observed differences and are reported as correlation coefficients (*r*), with thresholds of 0.1 - 0.3,

XSL•FO
RenderX

0.3 - 0.5, and >0.5 denoting small, medium, and large effects, respectively.

Ethical Considerations

Before participation, all participants provided informed consent, and the study was approved by the Ethical Commission of Roma Tre University (BRIC2022-CoRoMan approval r.01-24/06/2025). Participants did not receive any compensation for the participation, and all data was anonymized for maintaining privacy and confidentiality of the information.

Results

Linear Metrics

Figure 4 shows the COP trajectories of 3 representative participants in the 3 experimental conditions.

XSL•FO
RenderX

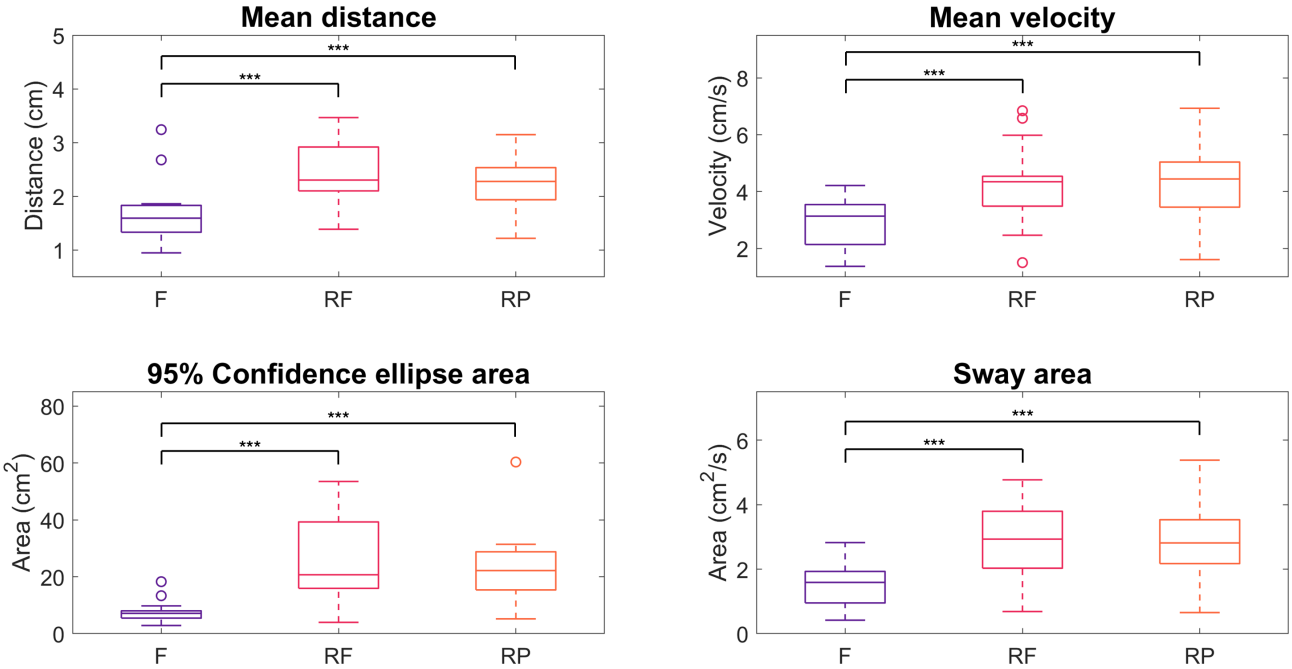
Figure 4. Center of pressure (COP) trajectories of 3 representative participants across the 3 experimental conditions: (A) Free, (B) Robot Free, and (C) Robot Plane. Each panel displays individual COP traces, with colors indicating different task repetitions. AP: anteroposterior; ML: mediolateral.

XSL•FO
RenderX

In the Free condition (Figure 4A), COP trajectories exhibit a compact pattern for all considered participants, extending roughly 5 cm in both mediolateral and anteroposterior directions. Paths remain well-contained within a small spatial envelope, showing limited variability across cycles. In the RF condition (Figure 4B), COP trajectories display a broader pattern compared to those observed in the Free condition (Figure 4A), particularly along the anteroposterior axis. The range of displacement remains around 5 cm in the mediolateral direction but increases to approximately 15 cm in the anteroposterior direction. Intrasubject variability across cycles remains limited. Similar to what is observed in the RF condition (Figure 4B), COP trajectories in the RP condition (Figure 4C) exhibit a broader pattern compared to those in the Free condition (Figure 4A), particularly along the anteroposterior axis. Displacement in the mediolateral direction remains around 5 cm, while in the anteroposterior direction, it increases to approximately 10 cm. Intrasubject variability across cycles remains generally limited. These patterns were observed across all participants, with limited intersubject variability, demonstrating the consistency of the findings. Refer to multimedia appendices for the complete set of COP traces of all participants and all experimental conditions (refer to Multimedia Appendix 1 for Free, Multimedia Appendix 2 for RF, and Multimedia Appendix 3 for RP).

The qualitative differences observed in COP trajectories are supported by the statistical analysis of sway parameters, as illustrated in Figure 5. The results reveal significant differences between the Free condition and both robot-assisted conditions, with the latter exhibiting consistently higher values across all parameters. In contrast, no significant differences were found between the 2 robot-assisted conditions for any of the parameters analyzed. Mean distance increased from 1.7 (SD 0.6) cm in Free to 2.4 (SD 0.6) cm in RF (P<.001; r=0.86) and 2.3 (SD 0.65) cm in RP (P<.001; r=0.86). Mean velocity increased from 2.9 (SD 0.9) cm/s in Free to 4.3 (SD 1.4) cm/s in RF and RP (P<.001; r=0.88). Confidence ellipse area increased from 7.6 (SD 4.1) cm^2 in Free to 24.9 (SD 14.2) cm^2 in RF (P<.001; r=0.88) and 23.1 (SD 13.4) cm^2 in RP (P<.001; r=0.88). Sway area increased from 1.5 (SD 0.7) cm^2/s in Free to 2.9 (SD 1.2 cm^2/s) in RF and RP (P<.001; r=0.88).

Figure 5. Distribution of sway parameters, differentiated by experimental conditions. "***" denotes P<.001 as obtained from inferential statistics. F: Free; RF: Robot Free; RP: Robot Plane.

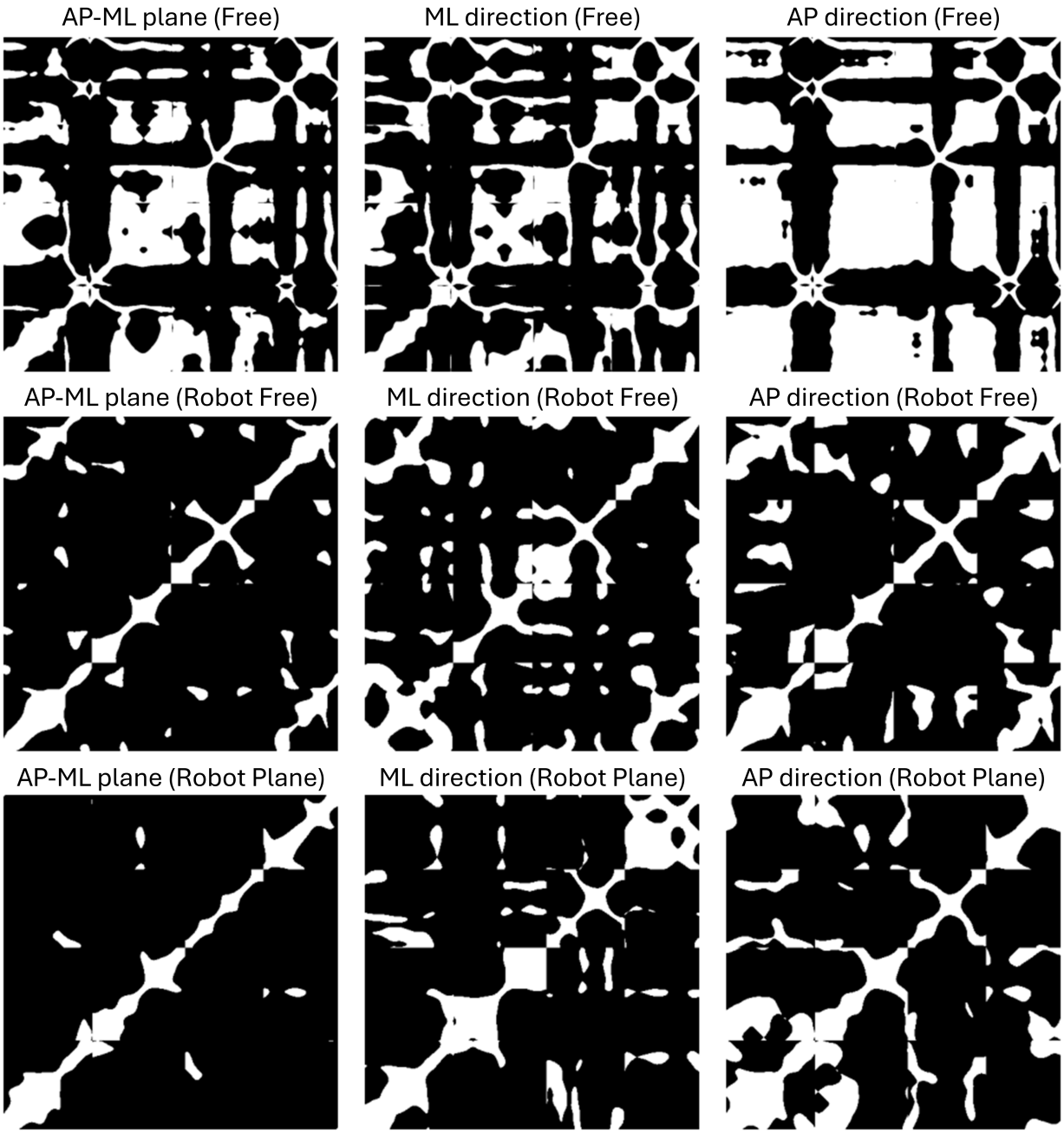


Nonlinear Metrics

An example of the RQA recurrence maps obtained from a subset of the cycles relative to 1 exemplary participant is reported in Figure 6. A consistent pattern emerges across both robot-assisted conditions, marked by a greater density of points along the principal diagonal line in the recurrence maps compared to the Free condition. The Free condition exhibits a more pronounced checkered structure, reflecting the prevalence of recurrent quasi-stationary states.

XSL•FO
RenderX

Figure 6. Sample recurrence maps, relative to 1 exemplary participant, in different trials and for different directions. AP: anteroposterior; ML: mediolateral.

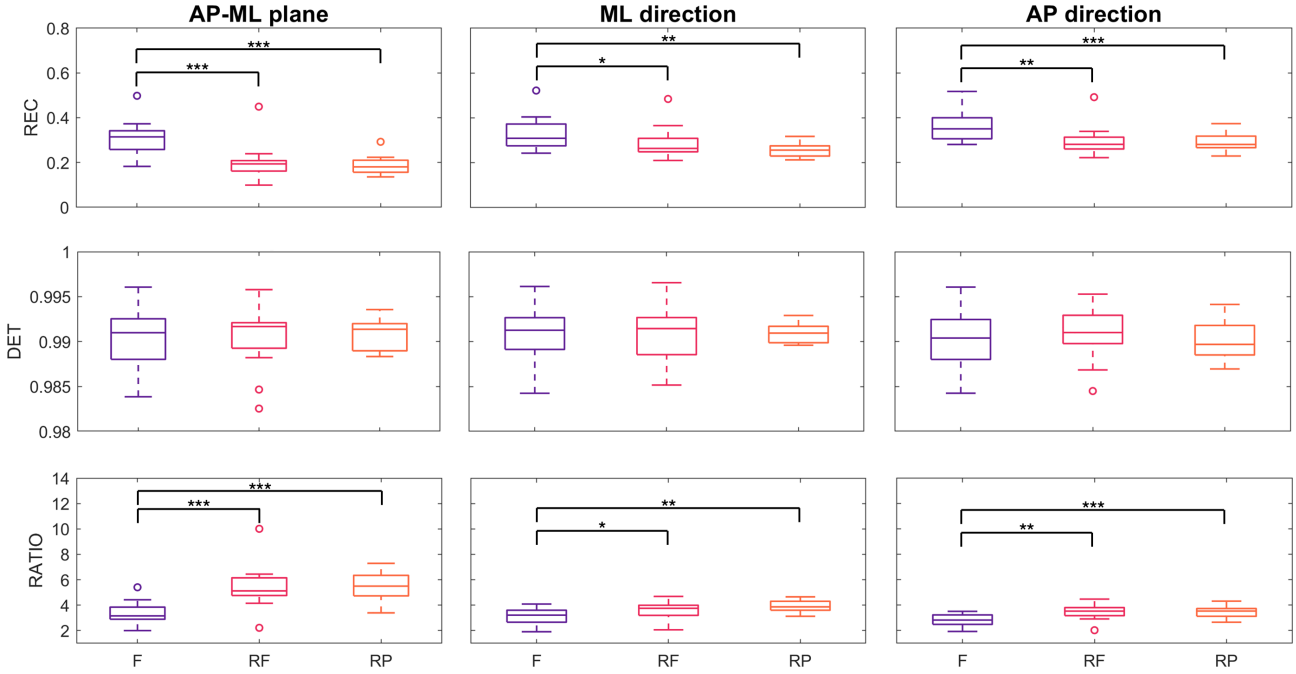


These qualitative and specific observations are supported by the quantitative results shown in Figure 7, which presents boxplots of the RQA-based parameters for all participants. Specifically, the REC is significantly lower in both robot-assisted conditions. REC decreased from 0.31 (SD 0.08) in Free to 0.2 (SD 0.08) in RF (P<.001; r=0.88) and 0.2 (SD 0.04) in RP (P<.001; r=0.88) in the anteroposterior-mediolateral plane; it decreased from 0.33 (SD 0.08) in Free to 0.28 (SD 0.07) in RF (P=.02; r=0.65) and 0.26 (SD 0.03) in RP (P=.007; r=0.75) in the mediolateral direction; it decreased from 0.36 (SD 0.07) in Free to 0.3 (SD 0.06) in RF (P=.009; r=0.75) and 0.26 (SD 0.03) in RP (P<.001; r=0.86) in the anteroposterior direction. On the other hand, determinism remains largely unchanged. Consequently, the ratio between determinism and REC is significantly higher during robotic assistance, reflecting a shift toward more periodic and structured postural control strategies when the cobot is engaged. This ratio increased from 3.41 (SD 0.87) in Free to 5.41 (SD 1.69) in RF (P<.001; r=0.88) and 5.51 (SD 1.11) in RP (P<.001; r=0.88) in the anteroposterior-mediolateral plane; it increased from 3.11 (SD 0.63) in Free to 3.64 (SD 0.72) in RF (P=.02; r=0.6) and 3.92 (SD 0.48) in RP (P=.007; r=0.75) in the mediolateral direction; it increased from 2.82 (SD 0.48) in Free to 3.47 (SD 0.57) in RF (P=.009; r=0.75) and 3.46 (SD 0.45) in RP (P<.001; r=0.86) in the anteroposterior direction. It is worth noting that the determinism values observed here are higher than those reported in [28], which can be attributed to our use of a

XSL•FO
RenderX

lower-dimensional embedding space and different threshold settings in the RQA procedure.

Figure 7. Distribution of recurrence quantification analysis (RQA) parameters, differentiated by experimental conditions. "*," "**," and "***" denote $P<.05$, $P<.01$, and $P<.001$, respectively, as obtained from inferential statistics. AP: anteroposterior; DET: determinism; F: Free; ML: mediolateral; RATIO; REC: recurrence rate; RF: Robot Free; RP: Robot Plane.



Discussion

Principal Findings

The analysis of the COP trajectories shows a greater extent along the anterior-posterior direction in robot-assisted conditions with respect to the Free condition, possibly indicating a higher number of oscillations of the body when the cobot intervenes in the task execution. This could suggest that participants tend to deviate from their postural stability, potentially compromising the overall efficiency and coordination of the kinematic chain during task execution. The absence of an evident difference between COP paths in the RF and RP conditions may indicate that the mode of collaboration with the cobot does not substantially alter the way robotic assistance affects postural stability.

Suggestions resulting from these qualitative observations align with quantitative results from the analysis of traditional COP parameters. Indeed, higher values of all parameters were found in robot-assisted conditions compared to the Free condition, suggesting altered postural dynamics in response to robotic assistance, regardless of the mode of collaboration with the cobot.

Higher mean distance may result from robotic constraints altering postural stability and strategy. Since mean distance is commonly interpreted as an indicator of postural control effectiveness [31], higher values are typically associated with a reduced ability to maintain a stable upright position [15]. Larger COP excursions have also been linked to reduced postural control [32] and balance impairments [33-35].

Higher mean velocity may suggest increased postural control effort and adjustments, likely due to imposed constraints from robotic assistance. Faster COP movements are often interpreted as indicators of increased demand for active postural control [31] or postural instability [32,36], assuming that the mean distance remains constant. However, in our case, the mean distance also increases under collaborative conditions. This combined increase complicates the interpretation; rather than indicating instability alone, it may reflect an active and purposeful postural adaptation to the new collaborative context.

A larger confidence ellipse area usually indicates higher COP exploratory behavior in postural sway, which may represent an adaptive response to the altered task dynamics under robotic assistance. This measure is in some cases associated with reduced balance control [37] and difficulty in maintaining stability during upright standing [38].

A greater sway area reflects increased movement, possibly in response to the higher postural control demands imposed by robot interaction. Since sway area is a hybrid measure capturing both the interaction between postural control activity and stability level [31], larger values are usually interpreted as a sign of decreased stability [39].

Contrary to our initial hypothesis, the analysis revealed no significant differences between the RF and RP conditions across all the parameters. While we expected that constraining the robot's movement might lead to changes in postural dynamics, the results suggest otherwise: the mere presence of robotic support, regardless of the collaboration mode, plays a dominant role in shaping participant responses. Indeed, the assistance provided by the cobot in both conditions may be sufficient to modulate the demands of the task, leading to similar adaptations.

Overall, COP tends to travel more (higher mean velocity and, to a certain extent, sway area) and further (mean distance and

XSL•FO
RenderX

ellipse area) when robotic assistance is provided. In general, larger COP excursions are associated with reduced balance control, indicating a decreased ability to remain in a smaller (and typically more stable) area during task repetitions [13]. This may suggest that the cobot influences task execution, which is particularly relevant given that increased COP excursions are associated with the development of WMSDs [40]. From this perspective, the Free condition appears to allow the most effective postural control and imposes the lowest associated risk, as evidenced by more stable and restricted sway patterns.

These findings might be the result of an alteration of the body schema, an integrated and dynamic internal representation of the body and its surrounding space [41], continuously shaped and updated through sensory and social interactions [42]. Efficient movement in space requires both localizing objects in extrapersonal space and maintaining a constantly updated status of body shape and posture [43]. The body schema reflects how the body interacts with its environment, and it is highly plastic, adapting to novel tools or altered body representations [44]. Here, the cobot holding the object may have challenged this schema, leading to difficulty in incorporating the cobot interaction and thereby potentially explaining the increased postural sway. Indeed, recent works may suggest that interactions with robots and virtual agents can disrupt or extend embodiment, requiring new sensorimotor adaptations [45-50].

Nonlinear results further support this interpretation. A lower REC in robot-assisted conditions may suggest a more variable or exploratory balance strategy, as postural states repeat less frequently. This indicates a shift away from stable, habitual postural control patterns toward more dynamic adjustments, potentially reflecting the fact that the cobot is not yet fully integrated into the participant's body schema. In this context, participants may need to continuously recalibrate their posture in response to the cobot's actions, rather than relying on familiar strategies. This interpretation aligns with the idea that technological artifacts reshape bodily experience and require new sensorimotor adaptations [43,48,51,52]. Notably, REC values are comparable across the 2 robot-assisted conditions, suggesting that the mode of interaction does not further alter the degree of postural control. In other words, the presence of the cobot itself introduces a fundamental shift in balance strategy, but the specifics of the interaction mode may have a limited additional effect on this aspect of postural control.

Comparable determinism values across all conditions indicate that, despite the increased postural sway and lower recurrence observed in the robot-assisted conditions, the postural responses remain structured rather than random. This suggests that participants are still using organized control strategies, even if these strategies are more adaptive and less repetitive. This reflects a balance between flexibility and structure: while the presence of the cobot necessitates ongoing adjustments, participants appear to adopt consistent patterns, rather than reacting in a chaotic or unpredictable manner. This might be related to Black's critique of the "natural interface" assumption [51]: the robot-object system does not immediately integrate as a seamless extension of the body, but does not cause complete instability either. Although the cobot may provide sufficient task stability, its presence is not yet fully integrated into the body schema, requiring continuous movement adjustments while maintaining a coordinated response. However, confirming this interpretation would require long-term longitudinal studies to assess whether such adaptations persist, diminish, or evolve.

Importantly, as observed with conventional COP measures, no significant differences were found between the 2 robot-assisted conditions across all RQA-based parameters, further indicating that the specific mode of parallel collaboration did not substantially affect overall postural outcomes. Thus, RQA findings suggest that interaction with the cobot challenges postural control without introducing randomness in balance control. The presence of the cobot appears to prevent full postural automatization, requiring individuals to engage in ongoing adjustments rather than relying on familiar movement patterns.

These results are also consistent with broader research on tool use, which demonstrates that tools can alter both perceptual and motor processes in healthy individuals as well as those with brain damage [41,43,52,53]. Effective tool use involves more than simple physical manipulation; it requires recognizing the object as a tool, knowing how to interact with it, and understanding the outcomes of its use [54,55]. These processes may be difficult to fully integrate into the body schema, particularly when new dynamics are introduced that the body has not yet adapted to, such as those posed by a cobot.

Overall, our results suggest that robotic assistance affects postural control and may challenge the body schema. However, interpretations in terms of motor learning, tool use, or body schema adaptation remain speculative at this stage. The observed changes in linear and nonlinear metrics indicate postural adjustments, yet it is unclear whether these reflect short-term responses or longer-term learning. Further research is needed to clarify how these adaptations develop over time.

Limitations

Although this study offers meaningful evidence regarding the effects of cobots on human postural control, several limitations should be considered when interpreting the results.

First, the research was conducted in a controlled laboratory setting using a standardized manipulation sequence, a simulated workbench, and a custom tool. The task sequence was carefully designed to closely resemble typical industrial manipulation operations and to capture key aspects of HRC, such as coordination with a robotic partner, repetitive handling, and precise posture adjustments in this context. While lab-based tasks cannot fully replicate the complexity and variability of real industrial scenarios, our design sought to approximate these conditions to ensure relevance to typical HRC environments. Future studies should incorporate diverse object types, task sequences, and interaction dynamics to further strengthen ecological validity and assess robustness across broader industrial contexts.

Second, the participant sample consisted of 14 healthy young adult males, and this limits generalizability to the broader industrial workforce. Different populations may respond differently in HRC contexts: older adults or individuals with musculoskeletal impairments might exhibit greater postural

XSL•FO
RenderX

sway and slower movements, while sex-related differences in strength, flexibility, and neuromuscular coordination could alter movement strategies. These factors suggest that our findings capture only part of the variability present in industrial settings. At the same time, the large effect sizes observed in our analyses indicate that the identified effects are robust and not solely sample dependent. Moreover, when considering a broader population sample, we expect that, even though sway parameters may be affected by anthropometric characteristics, nonlinear indicators would reasonably maintain similar variations. Future research should, however, include participants of different ages, sexes, and health conditions to improve outcomes.

Moreover, the study focused on short-term cobot interactions, leaving open questions about the long-term effects of repeated exposure, which might influence postural stability and increase the risk of developing musculoskeletal disorders. Investigating whether initial adaptations persist, diminish, or intensify over time, by incorporating more extended sessions, is therefore essential to clarify how postural control reacts or adapts to long-term exposure to HRC.

Finally, although RQA measures are computationally robust and have undergone rigorous validation, their novelty makes interpretation complex and context dependent [35]. As noted in the “Discussion” section, attributing the observed alterations in postural dynamics to body schema changes remains hypothetical and requires complementary methods (eg, kinematics, electromyography, and questionnaires on perceived effort) and longitudinal studies to support the evidence and strengthen causal inferences. This more comprehensive approach would reinforce the ergonomic perspective and offer a more precise understanding of both physical and mental demands in HRC scenarios.

Conclusions

This study investigated how HRC influences postural control and musculoskeletal load during dynamic manipulation tasks performed in upright stance. Postural control was evaluated using traditional COP-derived linear metrics and nonlinear RQA indicators. Findings show that interaction with the cobot increases COP excursions and velocity, indicating higher postural demands and reduced stability compared to performing the task alone [56]. These effects likely stem from task-specific challenges, movement constraints imposed by the cobot, and adaptive neuromuscular system responses. RQA analysis revealed lower RECs but stable determinism, suggesting that postural adjustments become more variable yet remain structured, reflecting adaptive rather than random responses. No significant differences were observed between the 2 robot-assisted conditions, implying that the cobot’s presence, rather than its movement constraints, drives these postural changes.

Overall, these findings indicate that HRC tasks may challenge the body schema and require continuous sensorimotor adaptation, potentially influencing both task execution and balance strategies. From an ergonomic perspective, the results highlight that cobotic tasks can affect whole-body postural organization, even when local task execution remains unchanged. This supports the inclusion of postural control and postural dynamics as complementary indicators in ergonomic assessments of HRC, alongside conventional biomechanical measures. Such insights are critical for ergonomic cobot design, emphasizing the importance of offering assistance that supports, rather than disrupts, natural postural control. In industrial contexts, cobots have the potential to enhance safety, productivity, and ergonomics; however, their impact on postural stability must be carefully considered. Increased COP excursions have been linked to a higher risk of musculoskeletal disorders [13]; poorly designed cobot interactions could thus inadvertently increase physical strain. Conversely, when properly integrated, cobots could reduce the physical demands of repetitive tasks, enabling safer and more ergonomic working conditions. Optimizing cobot design and interaction strategies is therefore essential to ensure that cobots complement human capabilities, support natural movement, minimize physical strain, and promote sustainable, safer human-centered work environments.

Funding

This work was partially supported by the INAIL projects BRIC2019-BRISK and BRIC2022-CoRoMan (CUP: F83C22002780001). The activities of MS, SR, and SC were also supported by the Project ECS0000024 Rome Technopole (CUP: F83B22000040006).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

DB, MS, and SC contributed to conceptualization. DB and SR contributed to the development of the methodology. AdN and GC contributed to software development. DB, GC, and MS performed validation. AdN, DB, and SR conducted the formal analysis. AdN, DB, and SR curated the data. AdN and DB wrote the original draft of the manuscript. AdN, DB, GC, MS, SR, and SC reviewed and edited the manuscript. DB performed the visualization. DB and SC supervised the study. MS and SC administered the project and acquired funding.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Center of pressure (COP) trajectories for each participant during the Free experimental condition. Each color is associated with each task repetition.
[DOCX File, 149 KB - humanfactors_v13i1e79930_app1.docx]

Multimedia Appendix 2
Center of pressure (COP) trajectories for each participant during the Robot Free experimental condition. Each color is associated with each task repetition.
[DOCX File, 184 KB - humanfactors_v13i1e79930_app2.docx]

Multimedia Appendix 3
Center of pressure (COP) trajectories for each participant during the Robot Plane experimental condition. Each color is associated with each task repetition.
[DOCX File, 184 KB - humanfactors_v13i1e79930_app3.docx]

References

1. Inkulu AK, Bahubalendruni MVAR, Dara A, K. S. Challenges and opportunities in human robot collaboration context of Industry 4.0 - a state of the art review. IR 2022 Feb 11;49(2):226-239. [doi: 10.1108/IR-04-2021-0077]
2. Ajoudani A, Albrecht P, Bianchi M, et al. Smart collaborative systems for enabling flexible and ergonomic work practices. IEEE Robot Automat Mag 2020;27(2):169-176. [doi: 10.1109/MRA.2020.2985344]
3. Ajoudani A, Zanchettin AM, Ivaldi S, Albu-Schäffer A, Kosuge K, Khatib O. Progress and prospects of the human-robot collaboration. Auton Robot 2018 Jun;42(5):957-975. [doi: 10.1007/s10514-017-9677-2]
4. Matheson E, Minto R, Zampieri EGG, Faccio M, Rosati G. Human–robot collaboration in manufacturing applications: a review. Robotics 2019;8(4):100. [doi: 10.3390/robotics8040100]
5. Gervasi R, Mastrogiacomo L, Franceschini F. A conceptual framework to evaluate human-robot collaboration. Int J Adv Manuf Technol 2020 May;108(3):841-865. [doi: 10.1007/s00170-020-05363-1]
6. Siciliano B, Khatib O, editors. Springer Handbook of Robotics: Springer; 2008. URL: https://www.academia.edu/26665877/Handbook_Springer_of_Robotics [accessed 2024-03-25]
7. de Nobile A, Bibbo D, Russo M, Conforto S. A focus on quantitative methods to assess human factors in collaborative robotics. Int J Ind Ergon 2024 Nov;104:103663. [doi: 10.1016/j.ergon.2024.103663]
8. Apraiz A, Lasa G, Mazmela M. Evaluation of user experience in human–robot interaction: a systematic literature review. Int J of Soc Robotics 2023 Feb;15(2):187-210. [doi: 10.1007/s12369-022-00957-z]
9. Cardoso A, Colim A, Bicho E, Braga AC, Menozzi M, Arezes P. Ergonomics and human factors as a requirement to implement safer collaborative robotic workstations: a literature review. Safety 2021;7(4):71. [doi: 10.3390/safety7040071]
10. Giallanza A, La Scalia G, Micale R, La Fata CM. Occupational health and safety issues in human-robot collaboration: state of the art and open challenges. Saf Sci 2024 Jan;169:106313. [doi: 10.1016/j.ssci.2023.106313]
11. Lorenzini M, Lagomarsino M, Fortini L, Gholami S, Ajoudani A. Ergonomic human-robot collaboration in industry: a review. Front Robot AI 2022;9:813907. [doi: 10.3389/frobt.2022.813907] [Medline: 36743294]
12. Peternel L, Fang C, Tsagarakis N, Ajoudani A. A selective muscle fatigue management approach to ergonomic human-robot co-manipulation. Robot Comput Integr Manuf 2019 Aug;58:69-79. [doi: 10.1016/j.rcim.2019.01.013]
13. Bibbo D, Corvini G, Schmid M, Ranaldi S, Conforto S. The impact of human-robot collaboration levels on postural stability during working tasks performed while standing: experimental study. JMIR Hum Factors 2025 Feb 27;12:e64892. [doi: 10.2196/64892] [Medline: 40014403]
14. Smith MJ, Karsh BT, Moro FBP. A review of research on interventions to control musculoskeletal disorders. In: Work-Related Musculoskeletal Disorders: Report, Workshop Summary, and Workshop Papers: National Academies Press (US); 1999. URL: https://www.ncbi.nlm.nih.gov/books/NBK230864 [accessed 2025-01-06]
15. Ramdani S, Tallon G, Bernard PL, Blain H. Recurrence quantification analysis of human postural fluctuations in older fallers and non-fallers. Ann Biomed Eng 2013 Aug;41(8):1713-1725. [doi: 10.1007/s10439-013-0790-x] [Medline: 23536111]
16. Lacour M, Bernard-Demanze L, Dumitrescu M. Posture control, aging, and attention resources: models and posture-analysis methods. Neurophysiol Clin 2008 Dec;38(6):411-421. [doi: 10.1016/j.neucli.2008.09.005] [Medline: 19026961]
17. Stergiou N, Decker LM. Human movement variability, nonlinear dynamics, and pathology: is there a connection? Hum Mov Sci 2011 Oct;30(5):869-888. [doi: 10.1016/j.humov.2011.06.002] [Medline: 21802756]
18. Błażkiewicz M, Hadamus A, Borkowski R. Recurrence quantification analysis as a form of postural control assessment: a systematic review. Appl Sci (Basel) 2023;13(9):5587. [doi: 10.3390/app13095587]
19. Franco C, Fleury A, Diot B, Vuillerme N. Applying entropy to human center of foot pressure data to assess attention investment in balance control. Presented at: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Jul 18-21, 2018; Honolulu, HI p. 5586-5589. [doi: 10.1109/EMBC.2018.8513533]

XSL•FO
RenderX

20. Fernandez-Cervantes E, Montesinos L, Gonzalez-Nucamendi A, Pecchia L. Recurrence quantification analysis of center of pressure trajectories for balance and fall-risk assessment in young and older adults. IEEE Trans Neural Syst Rehabil Eng 2023;31:926-935. [doi: 10.1109/TNSRE.2023.3236454] [Medline: 37018724]
21. Li J, Li K, Wei N. Recurrence quantification analysis of age-related changes in postural fluctuations during quiet standing on a slope. Presented at: 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI); Oct 13-15, 2018; Beijing, China p. 1-5. [doi: 10.1109/CISP-BMEI.2018.8633264]
22. Mazaheri M, Negahban H, Salavati M, Sanjari MA, Parnianpour M. Reliability of recurrence quantification analysis measures of the center of pressure during standing in individuals with musculoskeletal disorders. Med Eng Phys 2010 Sep;32(7):808-812. [doi: 10.1016/j.medengphy.2010.04.019] [Medline: 20570205]
23. Negahban H, Sanjari MA, Karimi M, Parnianpour M. Complexity and variability of the center of pressure time series during quiet standing in patients with knee osteoarthritis. Clin Biomech (Bristol) 2016 Feb;32:280-285. [doi: 10.1016/j.clinbiomech.2015.10.011] [Medline: 26588884]
24. Negahban H, Salavati M, Mazaheri M, Sanjari MA, Hadian MR, Parnianpour M. Non-linear dynamical features of center of pressure extracted by recurrence quantification analysis in people with unilateral anterior cruciate ligament injury. Gait Posture 2010 Apr;31(4):450-455. [doi: 10.1016/j.gaitpost.2010.01.020] [Medline: 20163962]
25. Riley MA, Balasubramaniam R, Turvey MT. Recurrence quantification analysis of postural fluctuations. Gait Posture 1999 Mar;9(1):65-78. [doi: 10.1016/s0966-6362(98)00044-7] [Medline: 10575072]
26. Seigle B, Ramdani S, Bernard PL. Dynamical structure of center of pressure fluctuations in elderly people. Gait Posture 2009 Aug;30(2):223-226. [doi: 10.1016/j.gaitpost.2009.05.005] [Medline: 19493680]
27. van Emmerik REA, Ducharme SW, Amado AC, Hamill J. Comparing dynamical systems concepts and techniques for biomechanical analysis. J Sport Health Sci 2016 Mar;5(1):3-13. [doi: 10.1016/j.jshs.2016.01.013] [Medline: 30356938]
28. Haddad JM, Van Emmerik REA, Wheat JS, Hamill J. Developmental changes in the dynamical structure of postural sway during a precision fitting task. Exp Brain Res 2008 Oct;190(4):431-441. [doi: 10.1007/s00221-008-1483-9] [Medline: 18629482]
29. Ranaldi S, Bibbo D, Corvini G, Schmid M, Conforto S. Modalities of sequential human robot collaboration trigger different modifications of trunk oscillations. Front Neurorobot 2023;17:1183164. [doi: 10.3389/fnbot.2023.1183164] [Medline: 37425334]
30. Kolbeinsson A, Lagerstedt E, Lindblom J. Foundation for a classification of collaboration levels for human-robot cooperation in manufacturing. Production & Manufacturing Research 2019 Jan 1;7(1):448-471. [doi: 10.1080/21693277.2019.1645628]
31. Prieto TE, Myklebust JB, Hoffmann RG, Lovett EG, Myklebust BM. Measures of postural steadiness: differences between healthy young and elderly adults. IEEE Trans Biomed Eng 1996 Sep;43(9):956-966. [doi: 10.1109/10.532130] [Medline: 9214811]
32. Ruhe A, Fejer R, Walker B. Center of pressure excursion as a measure of balance performance in patients with non-specific low back pain compared to healthy controls: a systematic review of the literature. Eur Spine J 2011 Mar;20(3):358-368. [doi: 10.1007/s00586-010-1543-2] [Medline: 20721676]
33. SHELDON JH. The effect of age on the control of sway. Gerontol Clin (Basel) 1963;5(3):129-138. [doi: 10.1159/000244784] [Medline: 13977074]
34. Baloh RW, Jacobson KM, Enrietto JA, Corona S, Honrubia V. Balance disorders in older persons: quantification with posturography. Otolaryngol Head Neck Surg 1998 Jul;119(1):89-92. [doi: 10.1016/S0194-5998(98)70177-9] [Medline: 9674519]
35. Langeard A, Milot E, Quarck G, Stoffregen T. The complexity of center of pressure positions during quiet stance and its relationship to cognition, aging, and falls. J Gerontol A Biol Sci Med Sci 2024 Oct 1;79(10):glae063. [doi: 10.1093/gerona/glae063] [Medline: 38394379]
36. Tipton N, Alderink G, Rhodes S. Approximate entropy and velocity of center of pressure to determine postural stability: a pilot study. Appl Sci (Basel) 2023;13(16):9259. [doi: 10.3390/app13169259]
37. Park JH, Mancini M, Carlson-Kuhta P, Nutt JG, Horak FB. Quantifying effects of age on balance and gait with inertial sensors in community-dwelling healthy adults. Exp Gerontol 2016 Dec 1;85:48-58. [doi: 10.1016/j.exger.2016.09.018] [Medline: 27666186]
38. Kushiro K, Goto F. Effect of masticating chewing gum on postural stability during upright standing. Neurosci Lett 2011 Jan 7;487(2):196-198. [doi: 10.1016/j.neulet.2010.10.021] [Medline: 20959136]
39. Karpinski M, Kizilova N. Posturographic study of human body sway before and after a work day for diagnosis of tiredness. Presented at: Biomdlore 2016; Oct 20-22, 2016; Vilnius, Lithuania p. 76-79. [doi: 10.3846/biomdlore.2016.20]
40. Gómez-Galán M, Pérez-Alonso J, Callejón-Ferre Á, López-Martínez J. Musculoskeletal disorders: OWAS review. Ind Health 2017 Aug 8;55(4):314-337. [doi: 10.2486/indhealth.2016-0191] [Medline: 28484144]
41. Holmes NP, Spence C. The body schema and the multisensory representation(s) of peripersonal space. Cogn Process 2004 Jun;5(2):94-105. [doi: 10.1007/s10339-004-0013-3] [Medline: 16467906]
42. Assaiante C, Barlaam F, Cignetti F, Vaugoyeau M. Body schema building during childhood and adolescence: a neurosensory approach. Neurophysiol Clin 2014 Jan;44(1):3-12. [doi: 10.1016/j.neucli.2013.10.125] [Medline: 24502900]

XSL•FO
RenderX

43. Maravita A, Iriki A. Tools for the body (schema). Trends Cogn Sci 2004 Feb;8(2):79-86. [doi: 10.1016/j.tics.2003.12.008] [Medline: 15588812]
44. Hoffmann M, Marques H, Arieta A, Sumioka H, Lungarella M, Pfeifer R. Body schema in robotics: a review. IEEE Trans Auton Mental Dev 2010;2(4):304-324. [doi: 10.1109/TAMD.2010.2086454]
45. Aymerich-Franch L, Fosch-Villaronga E. A self-guiding tool to conduct research with embodiment technologies responsibly. Front Robot AI 2020;7:22. [doi: 10.3389/frobt.2020.00022] [Medline: 33501191]
46. Rognini G, Blanke O. Cognetics: robotic interfaces for the conscious mind. Trends Cogn Sci 2016 Mar;20(3):162-164. [doi: 10.1016/j.tics.2015.12.002] [Medline: 26750512]
47. Xu J, Sun G, Xu J, Wang C. Bodioid: philosophical reflections on the hybrid of bodies and artifacts toward posthuman. AI & Soc 2025 Oct;40(7):5151-5165. [doi: 10.1007/s00146-025-02279-w]
48. Zigrino A, Zivi P, Ferlazzo F, Sdoia S. Body schema plasticity of the arm: a systematic review of the methods and tasks. Front Psychol 2025;16:1458409. [doi: 10.3389/fpsyg.2025.1458409] [Medline: 40667400]
49. Beckerle P, Castellini C, Lenggenhager B. Robotic interfaces for cognitive psychology and embodiment research: a research roadmap. Wiley Interdiscip Rev Cogn Sci 2019 Mar;10(2):e1486. [doi: 10.1002/wcs.1486] [Medline: 30485732]
50. Rudnicki S. The body, technology and translation: mapping the complexity of online embodiment. Sociol Res Online 2017 May;22(2):35-47. [doi: 10.5153/sro.4322]
51. Black D. Where bodies end and artefacts begin: tools, machines and interfaces. Body Soc 2014;20(1):31-60. [doi: 10.1177/1357034X13506946]
52. Cardinali L, Frassinetti F, Brozzoli C, Urquizar C, Roy AC, Farnè A. Tool-use induces morphological updating of the body schema. Curr Biol 2009 Jun 23;19(12):R478-R479. [doi: 10.1016/j.cub.2009.05.009] [Medline: 19549491]
53. Baccarini M, Martel M, Cardinali L, Sillan O, Farnè A, Roy AC. Tool use imagery triggers tool incorporation in the body schema. Front Psychol 2014;5:492. [doi: 10.3389/fpsyg.2014.00492] [Medline: 24910624]
54. Baber C, Parekh M, Cengiz TG. Tool use as distributed cognition: how tools help, hinder and define manual skill. Front Psychol 2014;5:116. [doi: 10.3389/fpsyg.2014.00116] [Medline: 24605103]
55. de Vignemont F. Body schema and body image--pros and cons. Neuropsychologia 2010 Feb;48(3):669-680. [doi: 10.1016/j.neuropsychologia.2009.09.022] [Medline: 19786038]
56. Raymakers JA, Samson MM, Verhaar HJJ. The assessment of body sway and the choice of the stability parameter(s). Gait Posture 2005 Jan;21(1):48-58. [doi: 10.1016/j.gaitpost.2003.11.006] [Medline: 15536033]

Abbreviations

Cobot: collaborative robot
COP: center of pressure
HRC: human-robot collaboration
REC: recurrence rate
RF: Robot Free
RP: Robot Plane
RQA: recurrence quantification analysis
WMSD: work-related musculoskeletal disorder

Edited by A Parush; submitted 01.Jul.2025; peer-reviewed by A Panahi, DE Patil; revised version received 17.Dec.2025; accepted 18.Jan.2026; published 23.Apr.2026.

Please cite as:
de Nobile A, Bibbo D, Ranaldi S, Schmid M, Corvini G, Conforto S
Dynamic Balance Control and Postural Adaptation in Human-Robot Collaborative Manipulation: Within-Subject Experimental Study
JMIR Hum Factors 2026;13:e79930
URL: https://humanfactors.jmir.org/2026/1/e79930
doi:10.2196/79930

© Alessia de Nobile, Daniele Bibbo, Simone Ranaldi, Maurizio Schmid, Giovanni Corvini, Silvia Conforto. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 23.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Prevalence of Work-Related Pain or Discomfort Among Urologists in the State of Florida: Results From the Florida Urologic Society Task Force on Ergonomic Challenges Experienced by Its Members

Justin David[1], MD; Austin Hill[2], MD; Jared Schommer[1], MD, PhD; Eric Regele[1], MD; Colleen Ball[1], BS; Sanoj Punnen[3], MD; Joseph Costa[4], MD; Adam Ball[5], MD; Jamin Brahmbhatt[6], MD; David D Thiel[1], MD; Raymond Pak[1], MD; Ram A Pathak[1], MD

[1]Mayo Clinic in Florida, 4500 San Pablo Rd S, Jacksonville, FL, United States

[2]Mayo Clinic, Rochester, MN, United States

[3]Desai Sethi Urology Institute, Miami, FL, United States

[4]University of Florida Health- Jacksonville, Jacksonville, FL, United States

[5]Gulf Stream Urology Associates, Port St. Lucie, FL, United States

[6]Orlando Health, Orlando, FL, United States

Corresponding Author:
Justin David, MD
Mayo Clinic in Florida, 4500 San Pablo Rd S, Jacksonville, FL, United States

Abstract

Background: Pain and work-related musculoskeletal disorders are commonly seen in surgeons, significantly impacting quality of life and burnout. A questionnaire-based study was conducted to further investigate the nature and etiology of work-related pain among urologists in the state of Florida.

Objective: This study aimed to quantify the number of urologists who reported work-related musculoskeletal disorders >25% of the time.

Methods: The Florida Urologic Society Task Force developed a survey based on the Nordic Musculoskeletal Questionnaire, with additional input from Cornell's ergonomic studies. The Mayo Clinic Survey Research Center conducted the survey and distributed it to 504 members of the Florida Urologic Society in 2023.

Results: The total response rate was 18.6% (94/504). The primary outcome (number of urologists who reported pain >25% of the time) was 45.3% (34/75). In total, 32.4% (22/68) of the respondents reported pain associated with endoscopic surgery >25% of the time, 40.0% (14/35) reported pain for major open cases, 20.6% (13/63) reported pain for minor open cases, and 22.7% (5/22) reported pain for robotic cases. In total, 68.8% (53/77) of the respondents attributed their work-related pain to uncomfortable operating positions, and 29.9% (23/77) chose to ignore their pain.

Conclusions: In this contemporaneous population of Florida urologic surgeons, almost half of the respondents describe having work-related pain >25% of the time. The data show that major open surgery had the highest rate of pain, followed closely by endoscopic surgery. Over 70% of the urologists in Florida are interested in official ergonomics training, which, if developed, may lead to increased productivity and better emotional, personal, and interpersonal well-being.

(JMIR Hum Factors 2026;13:e88848) doi:10.2196/88848

KEYWORDS

ergonomics; robotics; laparoscopy; endoscopy; urology; pain

Introduction

In recent years, ergonomics has gained considerable interest due to the increasing conscientiousness for physical and mental well-being [1-4], including in the field of urology. In the era of open pelvic surgery, complaints of work-related musculoskeletal disorders (WRMDs) were very common, present in more than 66% of surgeons [5,6]. Despite the invention of minimally invasive techniques such as laparoscopic or endoscopic surgery, WRMDs are just as common [5]. A study reported that urologists who regularly perform endoscopic procedures are at a 3 times higher risk for developing a chronic musculoskeletal disorder than those who do not [2].

However, even with attempts at improving ergonomics, such as robotic surgery, WRMDs are still present. A survey of robotic surgeons showed that pain is very prevalent, with the neck (73.7%), shoulder (52.6%), and lower back (42.1%) being the

XSL•FO
RenderX

most common areas [1]. An astounding 70% of urologists performing robotic, open, or endoscopic cases have used nonsteroidal anti-inflammatory drugs to cope with musculoskeletal pain while operating [3]. The type of injury likely differs. For example, WRMDs associated with major open surgery may be more related to the trunk and neck, while robotic surgeons have higher hand and/or wrist injuries [1,5,6]. Both open and robotic surgery have seen improvements with respect to ergonomics. Prismatic loupes and passive neck exoskeletons are recent improvements for open surgery that have been shown to decrease pain [7,8]. Robotic consoles continue to improve, such as the da Vinci 5 allowing for straighter back and neck posture and lighter hand controls with a rubber grip to improve finger or wrist discomfort [9].

Robotics and endoscopy are a significant part of many urologists' surgical schedules; however, 80% report no formal ergonomic training in the operating room [3]. This research indicates that there is a high prevalence of work-related pain experienced by surgeons and that minimal ergonomic education has taken place to prevent such problems.

In conjunction with the Florida Urologic Society (FUS), we sent a survey to urologists in the state of Florida to gather more contemporary information regarding the prevalence of work-related pain and discomfort to better understand the role ergonomics plays among Florida urologists.

Methods

Overview

The FUS funded and conducted this study in partnership with the Mayo Clinic Survey Research Center. A survey was developed based on the Nordic Musculoskeletal Questionnaire with additional input from Cornell's ergonomic studies, which is provided in Multimedia Appendix 1. This survey contained 24 multiple choice questions. There are 10 questions used to delineate the etiology, frequency, and treatment of pain. The next 8 questions inquire about the types of surgeries performed and the pain associated with them. A total of 2 questions assess if surgeons received ergonomic training or if they have interest in training. The final 4 questions consist of the surgeons' demographics and years spent in practice. The Mayo Clinic Survey Research Center distributed the survey to members of the FUS consisting of 504 practicing urologists in 2023.

The primary outcome was the number of urologists who reported pain >25% of the time during surgery. Secondary outcomes included demographics and the type of procedure, length of the procedure, cause of discomfort, specific muscles affected, and methods attempted to minimize discomfort. Inclusion criteria included members of the FUS in 2023, and there were no exclusion criteria. Data collection was performed using Research Electronic Data Capture (REDCap; Vanderbilt University) software, and the number of each response out of the total number of responders for each question was recorded. The total number of respondents was identified for each question, and the number of answers for each option was added. Sample size is dependent on response rate.

Ethical Considerations

This study was exempt from institutional review board approval.

Results

Demographics

In total, 18.6% (94/504) of the urologists within the FUS returned the survey. Not all participants completed every question.

Demographic data are presented in Table 1. In total, 46.8% (44/94) of the respondents worked in private practice, 30.9% (n=29) worked at a teaching-based hospital with resident assistance, 9.6% (n=9) worked at a teaching-based hospital without resident assistance, and 12.8% (n=12) worked at a community hospital.

Table . The type of urology practice was selected in the survey, with percentages of each answer provided in the table.

Type of urology practice	Values, n (%)
Private practice	44 (47)
Teaching-based hospital with resident assistance	29 (31)
Community hospital	23 (13)
Teaching-based hospital without resident assistance	9 (9)

In total, 76.6% (72/94) of the respondents completed the question regarding their experience level. Of the 72 respondents, 5.6% (n=4) were residents, 12.5% (n=9) had been in practice less than 5 years, 30.6% (n=22) had been in practice for 6 to 20 years, and 51.3% (n=37) have been practicing for more than 20 years.

In total, 75.5% (71/94) of the respondents completed the question regarding their gender identity, height, and weight. Of the 71 respondents, 84.5% (n=60) were male, 11.2% (n=8) were female, 2.8% (n=2) chose another description, and 1.4% (n=1) chose not to answer. Respondents could also choose their height and weight in a range. The most common height was between 5'9" to 6'0" (n=27, 38.0%) followed by 5'5" to 5'8" (n=25, 35.2%). The most common weight range was 151 to 200 lbs (n=47, 66.2%).

Types of Surgeries Performed: Endoscopic (Question 9)

In total, 79.7% (75/94) of the respondents completed the questions regarding type of surgery performed. Of the 75 respondents, 33.3% (n=25) performed between 1 and 5 endoscopic cases weekly, while 57.3% (n=43) performed >5 weekly.

XSL•FO
RenderX

Types of Surgeries Performed: Open (Question 11 and 13)

In total, 45.3% (34/75) of the respondents performed 1 to 4 major open cases, defined as cases taking more than 3 hours, while 53.3% (n=40) performed 0 weekly, and 1.3% (n=1) performed >4 cases. In total, 66.7% (n=50) of the respondents performed 1 to 4 minor open cases, 14.6% (n=11) performed 0, and 17.3% (n=13) performed >4 cases.

Types of Surgeries Performed: Robotic (Question 15)

In total, 70.7% (53/75) of the respondents performed 0 robotic cases weekly, while 22.7% (n=17) performed 1% to 3%, and 6.7% (n=5) performed >3 robotic cases per week.

Etiology of Pain or Discomfort

In total, 88.3% (83/94) of the respondents completed the question regarding types of WRMDs. Of the 83 respondents, 90.4% (n=75) reported pain at any point in the last week.

The primary outcome was pain or discomfort greater than 25% of the time during surgery. In total, 79.8% (75/94) of the respondents answered questions regarding the frequency of pain depending on the type of surgery. Of the 75 respondents, 45.3% (n=34) reported pain or discomfort greater than 25% of the time during surgery. Years in practice was not a significant predictor of pain, as 53.8% (7/13) of the urologists in practice for 5 years or less reported pain greater than 25% of the time during surgery, while 45.9% (17/37) of the urologists in practice for more than 20 years reported pain greater than 25% of the time during surgery.

In total, 32.8% (22/68) of the urologists who perform endoscopic surgery reported pain or discomfort for at least 25% of the time. In total, 32.0% (8/25) of the urologists who perform less than 5 endoscopic cases per week reported pain, while 32.6% (14/43) who perform greater than 5 cases reported pain. In total, 40.0% (14/35) of the urologists who perform major open surgery reported pain, with 41.8% (14/34) who perform 1 to 4 cases per week reported pain and 0% (0/1) of urologists who perform more than 4 major open cases per week. In total, 20.6% (13/63) of the urologists who perform minor open cases reported pain, with 14.0% (14/50) in those who perform 1 to 4 cases per week and 46.2% (6/13) in those who perform more than 4 cases. In total, 22.7% (5/22) of the urologists who perform robotic cases reported pain or discomfort at least 25% of the time, with 11.8% (2/17) in those who perform 1 to 3 cases per week and 60.0% (3/5) in those who perform more than 3 robotic cases per week.

We then analyzed the cause of discomfort (Table 2). In total, 18.2% (14/77) of the respondents reported no work-related discomfort, while 68.8% (n=53) of the respondents attributed their work-related pain and discomfort to an uncomfortable operating position, 51.9% (n=40) to extensive standing, 35.1% (n=27) to equipment hindrance, 19.5% (n=15) to heavy tools, 15.6% (n=12) to moving patients, and 10.4% (n=8) to another reason not listed on the survey.

Table . The causes of work-related pain, with the percentage of each response detailed in the right-hand column.

Cause of work-related pain	Values, n (%)
No work-related pain/discomfort	14 (18)
Uncomfortable operating position	53 (69)
Extensive standing	40 (52)
Equipment hindrance	27 (35)
Heavy tools	15 (20)
Moving patients	12 (16)
Other	8 (10)

Surgeon-Directed Measures to Reduce Pain or Discomfort

In total, 61.0% (47/77) of the respondents changed positions, 37.7% (n=29) used a chair for seated surgery, 35.1% (n=27) implemented wearing supportive footwear or stockings, and 31.2% (n=24) took microbreaks. Furthermore, 29.9% (n=23) of the respondents chose to ignore their intraoperative discomfort, 37.7% (n=29) managed their pain with over-the-counter medications, 14.3% (n=11) pursued medical attention, 14.3% (n=11) decreased surgical workload, and 10.4% (n=8) took time away from the operating room. In total, 6.49% (n=5) of the respondents chose to change to robotic surgery to attempt to alleviate pain.

The survey also explored the types of neuromusculoskeletal disorders formally diagnosed by the surgeon's medical care team, which revealed that 17.1% (13/76) of the respondents had cervical disc disease, 7.9% (n=6) reported wrist or forearm tendonitis, and 28.9% (n=22) reported back pain secondary to lumbar disc disease. In total, 46.6% (35/75) of the respondents stated that their sleep was affected by their work-related discomfort, 45.3% (n=34) stated that their work-related pain affected their surgical posture, 22.7% (n=17) reported deleterious effects on their mobility, and 21.3% (n=16) reported decreased stamina.

We then studied pursued medical intervention. In total, 35.1% (26/74) of the respondents pursued massage therapy, 27.0% (n=20) required diagnostic studies, 24.3% (n=18) required prescription medications, 20.3% (n=15) required referral to a specialist, 18.9% (n=13) used a brace or support device in the operating room, and 16.2% (n=12) even needed surgical intervention (Table 3).

XSL•FO
RenderX

Finally, 75% (54/72) of respondents have never received formal ergonomic training, and 72% (n=52) would be interested in completing such training.

Table . Methods used to minimize operative discomfort. Respondents could select multiple options.

Methods to minimize operative discomfort	Values, n (%)
Seek medical help	11 (14)
Changing position	47 (61)
Take a break	24 (31)
Go slower	3 (4)
Adjust some aspect of the surgical field	26 (34)
Change height by using step	23 (30)
Use chair for seated surgery	29 (38)
Changing instruments	4 (5)
Switch to robotic surgery	5 (6)
Consider different operative approach	5 (6)
Reduce case load	11 (14)
Footwear or support stockings	27 (35)
Surgical floor mats	16 (21)
Time away from the operating room	8 (10)
Over-the-counter medication	29 (38)
Ignore it	23 (30)
Other	8 (10)

Discussion

Principal Findings

The results of the present survey showed the prevalence of WRMDs in Florida urologists estimated at 45.3%, which is consistent with the literature estimating around 47% to 62% prevalence for urologists, who are at risk due to the high use of minimally invasive surgical modalities such as endoscopy, laparoscopy, and robotics [2]. Our definition of WRMD included those who experienced pain at least 25% of the time during surgery, while comparative studies have less restrictive criteria.

Our survey not only provided more information concerning the prevalence of WRMDs and lack of ergonomic education within the field of urology but also provided valuable details about which types of cases are causing injury, the circumstances during those cases causing discomfort, methods being used to combat this issue, and how these injuries are affecting physician performance. In total, 80% (75/94) of the respondents described having pain in the workplace, primarily due to an uncomfortable operating position or extensive standing.

Over half of the survey respondents reported being in practice for a minimum of 20 years, raising the possibility of selection bias. Further review of this cohort of surgeons revealed that almost half of the respondents performed open surgery weekly, which likely influenced the nature of WRMDs. While the number of years in practice did not influence prevalence of pain, the number of surgeries performed per week, especially for minor open cases and robotics, may influence the prevalence of WRMDs.

An intentional byproduct of robotic surgery was to lessen the ergonomic demand of surgeons. However, WRMDs associated with robotic surgery are still prevalent, but the affected body part may differ. In our study, the prevalence of WRMDs was higher with major open (14/35, 40.0%) versus robotic surgery (5/22, 22.7%)

Our survey found that over 1 in 4 urologists sought medical care for their pain, which contributes substantially to health care costs as the median estimated cost per claim for WRMDs is over US $20,000 [10]. While many urologists attempted to treat their pain with over-the-counter medications or other modalities, almost a third of respondents ignored their pain. Ignoring the pain is an alarming finding, as it often coexists with physician burnout [11].

A significant takeaway is the disparity between infrequency, effectiveness, and substantial interest in ergonomic training. Our study showed that 75% (54/72) of the urologists had no formal training, and 72% (52/72) would be interested in such education. A study evaluating the effectiveness of ergonomic training showed that 100% (32/32) of those who received such instruction found it helpful, 88% (28/32) changed their practice as a result, and 74% (24/32) noticed a decrease in strain after training [1]. This particular study also showed that a small time commitment to ergonomic training can dramatically decrease the incidence of strain and injury from performing robotic procedures by greater than 70% [1]. Because of the effectiveness of ergonomics training, its implementation into practice may

XSL•FO
RenderX

improve the pain and discomfort experienced during urologic surgery. If actual training sessions are not possible, general advice regarding acceptable intraoperative modifications to prevent injury should be practiced more routinely. Some tips include modifying table height to decrease back strain, maintaining a neutral position throughout a procedure, and taking 2-minute pauses every half hour or as needed to move, stretch, and relax during cases and stretching between cases [12-15]. Regarding robotics specifically, it has been associated with surgeon discomfort especially in the neck and trunk [16]. Becoming familiar with the robot's ergonomic settings has been shown to lower the rate of physical symptoms during and after procedures by not allowing greater than 20 degrees of neck flexion, avoiding excess knee flexion, and using the clutch often to restore proper and comfortable posture frequently [15].

There are limitations to this study. Any questionnaire-based study is associated with selection bias, as participants who complete the study may be more likely to have physical ailments or WRMDs. In total, 18.6% (94/504) of the recipients completed the survey, which is near the response rate of similar studies assessing surgical WRMDs [17-19]. Despite the consistent response rate with other studies, the low response rate may not truly represent the burden of WRMDs. A significant number of respondents practiced >20 years, risking the potential of selection bias. This study is also limited by convenience sampling as only urologists from Florida were included and respondents may have been affected by their accessibility to the survey. Another limitation is the inability to account for other confounding variables outside a surgeon's career, such as preexisting conditions or injuries that could contribute to pain. Additionally, many urologists perform several different types of surgeries, but it may be difficult to attribute their pain to a particular type. Our survey could not account for the type of practice a urologist has, as surgeons in an academic setting may have resident substitutes for parts of the surgery.

The goal of this survey was to explore the prevalence and etiology of pain in urologists. With further knowledge about the etiology of urologic surgery–related pain in this contemporary cohort, we hope that changes can be implemented to minimize the pain and improve work-related satisfaction, comfort, and surgeon longevity.

Conclusions

Surgeons are burdened by WRMDs far too frequently. Our survey found that nearly half of the urologists who answered the survey reported pain for more than 25% of the time during surgery, and more than one-fourth have sought medical attention for this pain. In addition, we found a higher rate of pain during surgery for major open surgery compared to robotic surgery. In total, 75% (54/72) of the respondents have received no formal ergonomic training, which could potentially decrease pain, health care costs, and burnout. Although ergonomic changes have continued to be implemented, such as with advanced robotic consoles, work-related pain is still prevalent.

Funding

The authors declared no financial support was received for this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey distributed to members of the Florida Urologic Society to assess prevalence and characteristics of intraoperative discomfort.
[DOCX File, 34 KB - humanfactors_v13i1e88848_app1.docx]

References

1. Franasiak J, Craven R, Mosaly P, Gehrig PA. Feasibility and acceptance of a robotic surgery ergonomic training program. JSLS 2014;18(4):e2014.00166. [doi: 10.4293/JSLS.2014.00166] [Medline: 25489213]
2. Tjiam IM, Goossens RH, Schout BM, et al. Ergonomics in endourology and laparoscopy: an overview of musculoskeletal problems in urology. J Endourol 2014 May;28(5):605-611. [doi: 10.1089/end.2013.0654] [Medline: 24417180]
3. Kerr RR. Practicing urology takes physical toll. Urol Times J 2019;47(8).
4. Stucky CC, Cromwell KD, Voss RK, et al. Surgeon symptoms, strain, and selections: systematic review and meta-analysis of surgical ergonomics. Ann Med Surg (Lond) 2018;27:1-8. [doi: 10.1016/j.amsu.2017.12.013] [Medline: 29511535]
5. Catanzarite T, Tan-Kim J, Whitcomb EL, Menefee S. Ergonomics in surgery: a review. Female Pelvic Med Reconstr Surg 2018;24(1):1-12. [doi: 10.1097/SPV.0000000000000456] [Medline: 28914699]
6. Fan X, Forsman M, Yang L, Lind CM, Kjellman M. Surgeons' physical workload in open surgery versus robot-assisted surgery and nonsurgical tasks. Surg Endosc 2022 Nov;36(11):8178-8194. [doi: 10.1007/s00464-022-09256-0] [Medline: 35589973]
7. Manning C, Jorgensen M. The price of pain: workers compensation costs for musculoskeletal claims in the state of Kansas, 2014-2022. J Occup Environ Med 2024 Mar 1;66(3):252-262. [doi: 10.1097/JOM.0000000000003034] [Medline: 38234103]
8. Pillado E, Li RD, Chia MC, et al. Reported pain at work is a risk factor for vascular surgery trainee burnout. J Vasc Surg 2024 May;79(5):1217-1223. [doi: 10.1016/j.jvs.2024.01.003] [Medline: 38215953]

XSL•FO
RenderX

9. Fan X, Yang L, Young N, Kaner I, Kjellman M, Forsman M. Ergonomics and performance of using prismatic loupes in simulated surgical tasks among surgeons - a randomized controlled, cross-over trial. Front Public Health 2024;11:1257365. [doi: 10.3389/fpubh.2023.1257365] [Medline: 38264242]
10. Kim JY, Norasi H, Cassivi SD, Black DM, Hallbeck MS. Use of an intraoperative head, neck, and back support exoskeleton on surgeons' pain and posture. Ann Surg 2026 Feb 1;283(2):184-191. [doi: 10.1097/SLA.0000000000006723] [Medline: 40201977]
11. Covas Moschovas M, Saikali S, Gamal A, et al. First impressions of the new da Vinci 5 robotic platform and experience in performing robot-assisted radical prostatectomy. Eur Urol Open Sci 2024 Sep 12;69:1-4. [doi: 10.1016/j.euros.2024.08.014] [Medline: 39314913]
12. Gabrielson AT, Clifton MM, Pavlovich CP, et al. Surgical ergonomics for urologists: a practical guide. Nat Rev Urol 2021 Mar;18(3):160-169. [doi: 10.1038/s41585-020-00414-4] [Medline: 33432182]
13. Park AE, Zahiri HR, Hallbeck MS, et al. Intraoperative "micro breaks" with targeted stretching enhance surgeon physical function and mental focus: a multicenter cohort study. Ann Surg 2017 Feb;265(2):340-346. [doi: 10.1097/SLA.0000000000001665] [Medline: 28059962]
14. Betsch D, Gjerde H, Lewis D, Tresidder R, Gupta RR. Ergonomics in the operating room: it doesn't hurt to think about it, but it may hurt not to!. Can J Ophthalmol 2020 Jun;55(3 Suppl 1):17-21. [doi: 10.1016/j.jcjo.2020.04.004] [Medline: 32448408]
15. Lloyd GL, Chung AS, Steinberg S, Sawyer M, Williams DH, Overbey D. Is your career hurting you? The ergonomic consequences of surgery in 701 urologists worldwide. J Endourol 2019 Dec;33(12):1037-1042. [doi: 10.1089/end.2019.0150] [Medline: 31187638]
16. Van't Hullenaar CD, Hermans B, Broeders I. Ergonomic assessment of the da Vinci console in robot-assisted surgery. Innov Surg Sci 2017;2(2):97-104. [doi: 10.1515/iss-2017-0007] [Medline: 31579742]
17. Norasi H, Kim J, Hallbeck MS, et al. Surgeons' dominant surgical modality: impacts on wellbeing, burnout, and interventions for neuromusculoskeletal disorders. Am J Surg 2024 Nov;237:115769. [doi: 10.1016/j.amjsurg.2024.05.010] [Medline: 38796376]
18. Norasi H, Hallbeck MS, Elli EF, Tollefson MK, Harold KL, Pak R. Impact of preferred surgical modality on surgeon wellness: a survey of workload, physical pain/discomfort, and neuromusculoskeletal disorders. Surg Endosc 2023 Dec;37(12):9244-9254. [doi: 10.1007/s00464-023-10485-0] [Medline: 37872425]
19. Healy KA, Pak RW, Cleary RC, Colon-Herdman A, Bagley DH. Hand problems among endourologists. J Endourol 2011 Dec;25(12):1915-1920. [doi: 10.1089/end.2011.0128] [Medline: 22004843]

Abbreviations

FUS: Florida Urologic Society
WRMD: work-related musculoskeletal disorder

Edited by E Borycki; submitted 02.Dec.2025; peer-reviewed by D Menychtas, RK Supramanian; revised version received 11.Mar.2026; accepted 07.Apr.2026; published 17.Jun.2026.

Please cite as:
David J, Hill A, Schommer J, Regele E, Ball C, Punnen S, Costa J, Ball A, Brahmbhatt J, Thiel DD, Pak R, Pathak RA
Prevalence of Work-Related Pain or Discomfort Among Urologists in the State of Florida: Results From the Florida Urologic Society Task Force on Ergonomic Challenges Experienced by Its Members
JMIR Hum Factors 2026;13:e88848
URL: https://humanfactors.jmir.org/2026/1/e88848
doi:10.2196/88848

© Justin David, Austin Hill, Jared Schommer, Eric Regele, Colleen Ball, Sanoj Punnen, Joseph Costa, Adam Ball, Jamin Brahmbhatt, David D Thiel, Raymond Pak, Ram A Pathak. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 17.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

C8 Health, a Platform for the Implementation of Best Practices: Survey-Based Usability Study

Faria Nisar[1*], MD; Nicolas Mario D Alessandro[1*], MD; Jessica Suratkal[2*], MPH; Ido Zamberg[3], MD, MEHP; Samantha Elen Pope[1], BA; Ali Ali[1], DO; Luis Etienne Tollinche[1,4,5*], MD, MPH

[1]Department of Anesthesiology and Pain Management, School of Medicine, The MetroHealth System, 2500 MetroHealth Drive, Cleveland, OH, United States

[2]Case Western University School of Medicine, Cleveland, OH, United States

[3]School of Education, Johns Hopkins University, Baltimore, MD, United States

[4]Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH, United States

[5]Outcomes Research Consortium, Cleveland, OH, United States

[*]these authors contributed equally

Corresponding Author:
Faria Nisar, MD
Department of Anesthesiology and Pain Management, School of Medicine, The MetroHealth System, 2500 MetroHealth Drive, Cleveland, OH, United States

Abstract

Background: Mobile health (mHealth) apps are increasingly integrated into clinical workflows to support decision-making and adherence to best practices. Usability is a critical determinant of adoption, engagement, and long-term use.

Objective: This study aimed to evaluate the usability of the C8 Health platform deployed in the Department of Anesthesiology at MetroHealth in 2024.

Methods: A quality improvement initiative was conducted using the mHealth App Usability Questionnaire (MAUQ). A total of 142 anesthesiology clinicians participated by completing the questionnaire. This study was reported in accordance with the SQUIRE (Standards for Quality Improvement Reporting Excellence) guidelines for quality improvement reporting.

Results: The overall MAUQ scores indicated high usability, with mean scores greater than 5.5 across core items (overall mean MAUQ score 5.73, SD 0.81). These findings suggest strong user satisfaction and positive engagement with the C8 Health platform among anesthesiology clinicians.

Conclusions: The C8 Health platform demonstrated high usability in an anesthesiology setting. These results support its integration as a clinical decision support tool to enhance workflow efficiency and adherence to best practices.

(JMIR Hum Factors 2026;13:e83547) doi:10.2196/83547

KEYWORDS

C8; mobile health; mHealth; mHealth App Usability Questionnaire; MAUQ; usability; quality initiative; anesthesiology

Introduction

Mobile health (mHealth) technologies have become integral to modern health care, offering “low-cost and high accessibility” solutions that empower clinicians and patients alike and help streamline clinical workflows [1,2]. However, a significant barrier to adoption lies in usability. Poor app design, data entry burden, and frustrating interfaces often cause users to abandon these tools. To address this, the mHealth App Usability Questionnaire (MAUQ) was developed as a validated, reliable instrument tailored specifically for evaluating usability in health care apps, covering ease of use, interface satisfaction, and perceived usefulness [3].

In June 2024, the MetroHealth System introduced C8 Health, an mHealth app designed to help health care institutions implement their best practices effectively through the centralization of clinical protocols, checklists, and workflows for care teams. Though early feedback has been promising [4-6], systematic usability evaluation among end users has not yet been conducted.

This study aimed to fill that gap by assessing C8 Health’s usability among MetroHealth anesthesia clinicians using the MAUQ. Evaluating usability at this stage is critical not only to refine the app’s user interface and functional workflow but also to understand how such tools might support scalable, protocol-driven perioperative care [7]. The results are expected to guide both iterative improvements to C8 Health and inform

XSL•FO
RenderX

frameworks for introducing similar mHealth solutions in hospital settings.

Methods

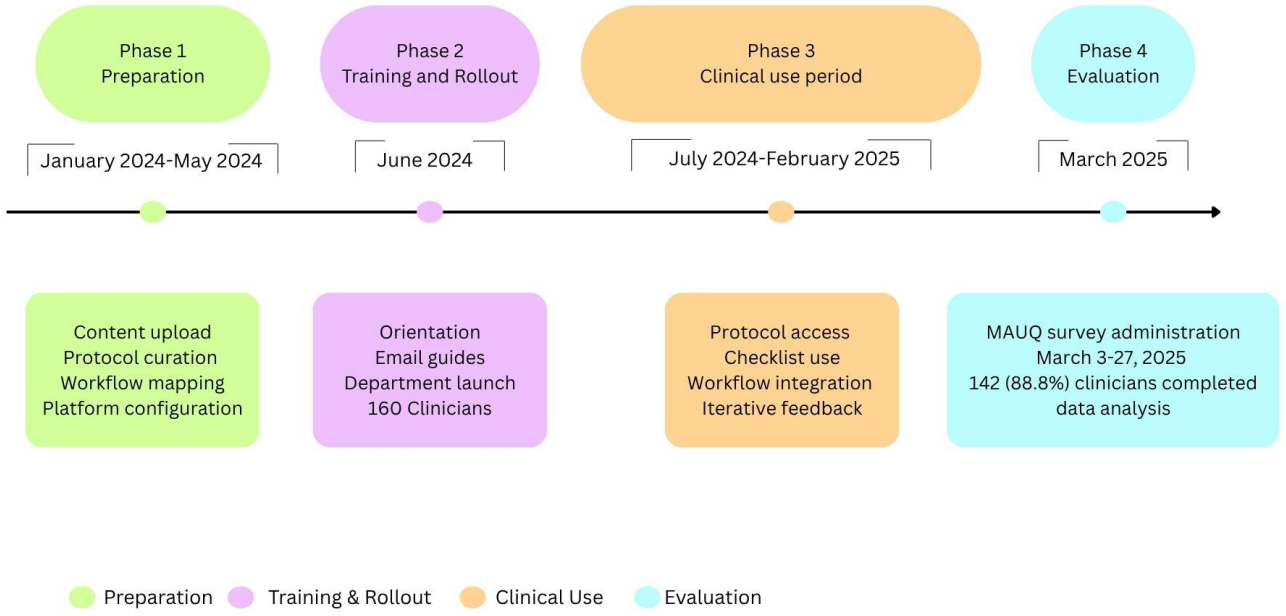
Study Design and Setting

This was a cross-sectional survey study conducted in the Department of Anesthesiology at the MetroHealth System, a tertiary academic medical center in Cleveland, Ohio, United States. The study evaluated the usability of C8 Health, a mHealth app introduced in June 2024 to support clinical workflow efficiency and adherence to best practices. This study was reported in accordance with the SQUIRE (Standards for Quality Improvement Reporting Excellence) guidelines for quality improvement reporting.

Participants and Recruitment

All anesthesiology clinicians employed at the MetroHealth System during the study period were invited to participate. Inclusion criteria included active clinical practice in the department and familiarity with the C8 Health platform. Participation was voluntary and anonymous. Recruitment was conducted via institutional email invitations, and nonresponders received 2 to 3 reminder emails per week from research faculty to maximize response rates (Figure 1).

Figure 1. Implementation timeline of the C8 Health platform in the Department of Anesthesiology and Pain Management, the MetroHealth System, Cleveland, Ohio, United States. MAUQ: mHealth App Usability Questionnaire.



Data Collection

The survey was administered online using SurveyMonkey (SurveyMonkey Inc) between March 3 and March 27, 2025. Survey responses were deidentified to preserve confidentiality. No personally identifiable information was collected.

Measures

Usability was assessed using the validated MAUQ, which evaluates 3 domains: ease of use, system information arrangement, and usefulness in health care delivery. The questionnaire uses a 7-point Likert scale, with higher scores indicating greater usability. Open-ended questions were also included to capture qualitative feedback regarding functionality, interface design, and overall user experience.

Outcomes

The primary outcome was the overall usability score derived from the MAUQ. Secondary outcomes included subscale scores for usability domains and thematic analysis of open-ended responses to identify strengths and areas for improvement.

Statistical Analysis

Descriptive statistics were used to summarize the survey responses. Continuous variables (eg, MAUQ scores) were reported as means with SDs. Subscale analyses were performed to compare usability domains. Open-ended comments were analyzed thematically to identify recurring usability issues and user recommendations.

Ethical Considerations

The study received institutional review board exemption (STUDY00000729) from the MetroHealth System, as it was classified as a quality improvement initiative involving minimal risk to participants. As the study was granted an institutional review board exemption, informed consent was waived. All data were handled in accordance with institutional privacy and confidentiality standards, and no compensation was provided to participants.

Results

The survey was distributed to all 160 clinicians, including faculty physicians, residents, and departmental staff, yielding a response rate of 88.8% (n=142).

Overall, results showed high usability ratings. On a 7-point Likert scale, the app received weighted averages above 5.5 for most core items. Clinicians strongly agreed that the app was easy to use (122/142, 86%; mean 5.73), easy to learn (124/142,

XSL•FO
RenderX

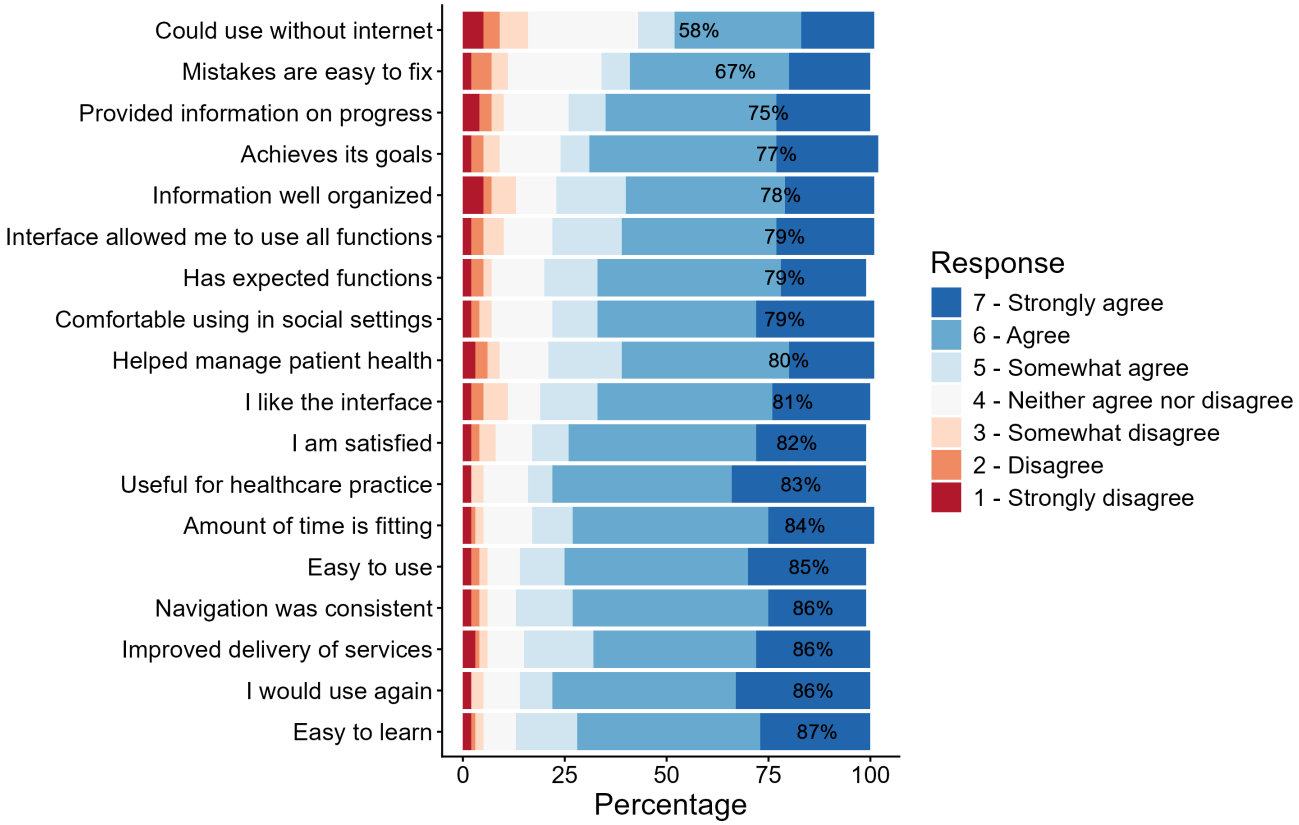
87%; mean 5.78), and supported consistent navigation (122/142, 86%; mean 5.70). Additionally, participants (112/142, 79%) valued the design and were able to access all key functions (mean 5.50). Because of the straightforward interface, 84% (119/142) of respondents felt the time required to use the app was appropriate.

Surveyed clinicians also found the information distributed by C8 Health beneficial, as 79% (112/142) of users appreciated the app's serviceability. The app was perceived as valuable in practice, with strong agreement that it improved health care service access (mean 5.66) and supported effective patient health management (mean 5.54, SD).

Although largely positive, the survey highlighted areas for improvement. Offline utility received a weighted average of 4.95, with only 58% (82/142) of users finding this component satisfactory. Additionally, error recovery received a lower average of 5.28 (SD), yet 67% (95/142) of participants felt satisfied with this function.

However, overall satisfaction remained high (mean 5.72), affirming the app's acceptance and perceived benefit in a clinical setting. Respondents strongly agreed that they would use C8 Health again (mean 5.87). These findings (Figure 2) validate the C8 Health app's design and implementation as a useful, user-friendly digital tool for anesthesiology professionals. Ongoing iterative improvements, particularly focused on offline functionality and error handling, are recommended to maximize utility and engagement.

Figure 2. mHealth App Usability Questionnaire survey results with percentages representing the summation of "agree" and "strongly agree" responses.



Discussion

Principal Findings

In this usability study, of the 160 clinicians from the Department of Anesthesiology at the MetroHealth System, 142 (88.8%) completed the MAUQ questionnaire to assess C8 Health, an mHealth app that centralizes clinical protocols and checklists to enhance clinical workflow and adherence to best practices. The MAUQ is a validated tool to assess the usability of the mHealth platform targeted for health care providers. Overall, the scores showed high usability and satisfaction, with weighted averages above 5.5 on a 7-point Likert scale for key aspects such as ease of use, accessibility of functions, design appreciation, and perceived value in practice.

Beyond global usability ratings, individual MAUQ items provide insight into how C8 Health supported clinical practice. A total of 86% (122/142) of respondents agreed or strongly agreed that the platform improved the delivery of health care services, and 83% (118/142) found it useful for their health care practice, reflecting its integration into day-to-day clinical workflow. Notably, 80% (114/142) agreed that it helped them manage patient health, suggesting that clinicians perceived the platform as more than an informational repository but rather as an active support tool in patient care decisions. The high agreement on information organization (111/142, 78%) and availability of expected functions (112/142, 79%) further suggested that clinicians could reliably locate protocols and checklists when needed at the point of care, a key requirement in time-sensitive perioperative environments such as the operating room.

XSL•FO
RenderX

About C8 Health

C8 Health is a platform for the implementation of best practices that centralizes medical knowledge to streamline health care operations and improve patient care. It disseminates essential information such as clinical protocols, policies, procedures, guidelines, checklists, and workflows, giving care teams fast access to standardized, evidence-based resources [7]. Apps such as C8 Health have the potential to offer benefits in health care by improving access to standardized clinical knowledge with the goal of reducing care variability and informing clinical decision-making. By unifying resources and enabling seamless knowledge sharing across departments and organizations, these tools may support more consistent, accurate, and timely care delivery. Whether this translates to direct improvements in patient outcomes, including reductions in medical errors or mortality rates, remains to be established through controlled outcome studies.

Current State of Best Practices Implementation in Health Care

Adherence to established clinical protocols significantly enhances patient outcomes by reducing complications, shortening hospital stays, and lowering health care costs. For instance, following the Enhanced Recovery After Surgery (ERAS) protocol in breast reconstruction has been shown to decrease postoperative complications and reduce hospital length of stay [8]. Similarly, compliance with anesthesia-related ERAS measures, such as multimodal analgesia and opioid-sparing strategies, has been associated with significantly shorter hospital stays in colorectal and gynecologic surgeries [9,10]. Furthermore, the use of ventilators and infection prevention bundles in intensive care unit settings has been shown to decrease mortality and ventilator-associated pneumonia rates when properly implemented [11,12]. Nonadherence to anesthesia guidelines in pain interventions has also been linked to severe complications, emphasizing the risks of deviation [13]. Collectively, these findings reinforce that rigorous protocol compliance in surgical and anesthetic care settings leads to safer patient outcomes, with patients subsequently facing fewer complications and lower out-of-pocket expenses related to preventable hospitalizations or errors.

Current adherence to medical protocols such as the ERAS protocol remains suboptimal despite widespread implementation. A recent large-scale study found that the actual compliance rate with the ERAS protocol was only 32.3%, with a potential compliance rate of 53.2% if patients without postoperative medical issues were included [14]. These findings are consistent with broader literature, which highlights that adherence to medical protocols is often challenged by clinical complexity and organizational factors [15]. Thus, while protocols such as ERAS have proven benefits, real-world adherence is influenced by a complex interplay of patient, health care provider, and system-level factors [14,15].

Systems for implementation of best practices, such as C8 Health, have demonstrated the potential to improve adherence to medical protocols by providing clinicians with timely, evidence-based recommendations and reminders integrated into clinical workflows. Multiple studies indicate that similar interventions are associated with small to moderate improvements in the delivery of guideline-recommended care, with average absolute increases in process adherence of approximately 5.8% and best-case improvements reaching up to 8.5% [16]. Features such as treatment recommendations and automated flagging of best-practice alerts are particularly effective in supporting guideline adherence and quality assurance in patient care [17]. These systems can also enhance diagnostic accuracy, support monitoring, and facilitate preventive screening, all of which contribute to better protocol compliance [17]. Overall, while the extent of improvement varies, systems for implementation of best practices consistently contribute to better adherence to clinical protocols across diverse health care settings, indicating a substantial functional role of apps such as C8 Health in improving patient outcomes [16-18].

From the health care provider's perspective, these digital tools may enhance the experience of delivering care by potentially reducing cognitive burden, supporting confidence in clinical decisions, and encouraging a culture of collaboration and transparency. Features such as natural language search, real-time updates, and integrated authoring tools help clinicians spend less time searching for information and more time focused on patient care. This contributes to reduced burnout and greater job satisfaction, both critical factors in an era marked by workforce shortages and growing demands on health care professionals [19].

Assessing C8 Health

Considering the benefits of apps such as C8 Health to patients, health care providers, and health care institutions, one of the first steps in determining the feasibility of implementing this app is assessing usability. Using the MAUQ, the C8 Health app received a high overall MAUQ score among anesthesiologists at MetroHealth, with most core items receiving >5.5 on a 7-point Likert scale. Overall satisfaction with the app was high, with users agreeing that the app was easy to use with an agreeable interface. Notably, the app was perceived as valuable in practice, as users agreed that it improved health care service access and supported effective patient health management. These positive findings have important future implications: as the app was well perceived by anesthesiologists, broader adoption and integration into clinical workflows are increasingly possible, potentially supporting streamlined communication and increased efficiency, though improvements in patient outcomes remain hypothetical at this stage and require validation through prospective studies with objective outcome measures across health care institutions.

Strengths and Limitations

Strengths of this study include that usability was assessed using the MAUQ, a validated and widely used instrument in mHealth research. The use of this tool increases both the reliability of our measurements and the comparability of findings with those from other studies in the field. The evaluation was conducted solely at a single postimplementation time point. The absence of preimplementation data precludes direct before-and-after comparisons of workflow efficiency, protocol adherence, or platform use, which would have provided stronger causal evidence. Furthermore, although voluntary participation in the questionnaire introduces the possibility of selection bias, this

XSL•FO
RenderX

risk is mitigated by the exceptionally high response rate (142/160, 88.8%), which strengthens confidence in the representativeness of the data.

This study did not collect objective operational or clinical outcome data (eg, protocol adherence rates, time-to-information retrieval, or patient-level outcomes). Consequently, the improvements in workflow efficiency and care quality discussed herein are based on user perception rather than measured performance indicators.

One significant limitation of this study is that the evaluation was conducted at a single center and lacked a control group, both of which may restrict the generalizability of the findings. In addition, limitations of the C8 Health platform itself were identified. These include limited offline functionality, suboptimal error recovery, and a lack of integration with existing electronic health record systems. Such constraints may reduce workflow efficiency and continuity of care in practical settings. However, these issues represent identifiable opportunities for improvement to be addressed in future system development.

Future Directions

In future research, quantifying improvements in patient care and institutional productivity following implementation of the app may assist in valuing the app in the context of outcomes and efficiency metrics. There is also a need for multicenter and longitudinal studies to assess sustained use. Future research should explore the integration of C8 Health with electronic health record systems to create a more comprehensive clinical decision support ecosystem. This integration could enable automated protocol selection based on patient characteristics, streamlined documentation workflows, and enhanced tracking of clinical outcomes. Future studies should also incorporate prospective pre-post designs with baseline data collection before implementation to better isolate the platform's effect on clinical workflow and protocol adherence.

Overall, the strong usability and perceived value of the C8 Health app pave the way for it to become a transformative tool in perioperative care, setting a new standard for technology-driven support in anesthesiology and beyond.

Conclusions

These findings indicate that the C8 platform demonstrates strong perceived usability and holds significant promise for supporting perioperative clinical workflows, as reported by end users. Ongoing development based on user feedback may further improve its effectiveness and extend its applicability to other clinical areas.

Acknowledgments

The authors would like to acknowledge the Department of Anesthesiology at the MetroHealth System for its support in the distribution of the survey and for facilitating clinician participation. The authors thank the clinicians who generously contributed their time and feedback, which made this study possible. The authors further extend their sincere gratitude to Tomer Cramer for his careful review of the manuscript and for his insightful suggestions, which significantly enhanced its quality.

Funding

This project was supported in part by the Clinical and Translational Science Collaborative (CTSC) of Cleveland, which is funded by the National Institutes of Health (NIH), National Center for Advancing Translational Sciences (NCATS), Clinical and Translational Science Award (CTSA) grant UL1TR002548. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Data Availability

The survey data generated during this study are not publicly available due to institutional privacy restrictions but are available from the corresponding author upon reasonable request.

Authors' Contributions

Resources: SEP
Methodology: IZ
Project administration: LET
Software: IZ
Supervision: LET, FN
Visualization: IZ, LET
Writing—original draft: FN
Writing—review and editing: FN, NMDA, JS, IZ, AA, LET

Conflicts of Interest

LET is a consultant and member of the advisory board at C8 Health Inc. LET is a grant recipient through the Merck Investigator Studies Program (MISP) to fund a clinical trial at Memorial Sloan Kettering Cancer Center (NCT03808077). LET serves as a consultant and advisor for Merck & Co Pharmaceutical Company and GE Healthcare. LET receives a stipend for his role as an examiner with the American Board of Anesthesiology. LET is an expert testimony witness and receives compensation for this

XSL•FO
RenderX

role. LET is a grant recipient from the National Institutes of Health to conduct a clinical trial. This material is based upon work supported by the National Institutes of Health under award number 20-00-00-1005210. LET is a board member of the MetroHealth Foundation. LET is a board member of the Harvard Club of Northeast Ohio Alumni Board. LET serves on the medical advisory board of C8 Health and is president of the Case Western Reserve University School of Medicine Alumni Board. IZ is the cofounder and chief medical officer at C8 Health. All other authors declare no other conflicts of interest.

References

1. Cho H, Yen PY, Dowding D, Merrill JA, Schnall R. A multi-level usability evaluation of mobile health applications: a case study. J Biomed Inform 2018 Oct;86:79-89. [doi: 10.1016/j.jbi.2018.08.012] [Medline: 30145317]
2. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida AJ, et al. Quality, usability, and effectiveness of mHealth Apps and the role of artificial intelligence: current scenario and challenges. J Med Internet Res 2023 May 4;25:e44030. [doi: 10.2196/44030] [Medline: 37140973]
3. Zhou L, Bao J, Setiawan IM, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500. [doi: 10.2196/11500] [Medline: 30973342]
4. Zamberg I, Schiffer E, Stoermann-Chopard C. Novice and advanced learners' satisfaction and perceptions of an e-learning renal semiology module during the COVID-19 pandemic: mixed methods study. JMIR Med Educ 2021 Jun 28;7(2):e29216. [doi: 10.2196/29216] [Medline: 34048357]
5. Windisch O, Zamberg I, Zanella MC, et al. Using mHealth to increase the reach of local guidance to health professionals as part of an institutional response plan to the COVID-19 outbreak: usage analysis study. JMIR Mhealth Uhealth 2020 Aug 19;8(8):e20025. [doi: 10.2196/20025] [Medline: 32749996]
6. Zamberg I, Manzano S, Posfay-Barbe K, Windisch O, Agoritsas T, Schiffer E. A mobile health platform to disseminate validated institutional measurements during the COVID-19 outbreak: utilization-focused evaluation study. JMIR Public Health Surveill 2020;6(2):e18668. [doi: 10.2196/18668] [Medline: 32250958]
7. C8 Health. URL: https://c8health.com/ [accessed 2025-09-03]
8. Gruber AG, Zandomenico JG, Calegari HG, Schuelter-Trevisol F. Adherence to the Enhanced Recovery After Surgery (ERAS) protocol in breast reconstruction plastic surgery. RBCP 2023;38(3):e0805. [doi: 10.5935/2177-1235.2023RBCP0805-EN]
9. Hayek J, Zorrilla-Vaca A, Meyer LA, et al. Patient outcomes and adherence to an enhanced recovery pathway for open gynecologic surgery: a 6-year single-center experience. Int J Gynecol Cancer 2022 Nov 7;32(11):1443-1449. [doi: 10.1136/ijgc-2022-003840] [Medline: 36202425]
10. Grant MC, Pio Roda CM, Canner JK, et al. The impact of anesthesia-influenced process measure compliance on length of stay: results from an enhanced recovery after surgery for colorectal surgery cohort. Anesth Analg 2019 Jan;128(1):68-74. [doi: 10.1213/ANE.0000000000003458] [Medline: 29782405]
11. Ferreira CR, de Souza DF, Cunha TM, et al. The effectiveness of a bundle in the prevention of ventilator-associated pneumonia. Braz J Infect Dis 2016;20(3):267-271. [doi: 10.1016/j.bjid.2016.03.004] [Medline: 27102778]
12. Rajandra A, Yunos NM, Teo CH, et al. Incidence, compliance, and risk factor associated with central line-associated bloodstream infection (CLABSI) in intensive care unit (ICU) patients: a multicenter study in an upper middle-income country. Antibiotics (Basel) 2025 Mar 7;14(3):271. [doi: 10.3390/antibiotics14030271] [Medline: 40149082]
13. El-Tallawy SN, Pergolizzi JV, Albasher MJ, et al. Non-adherence to clinical practice guidelines in regional anesthesia and pain interventions: insights from a comprehensive case series review. Cureus 2024 Dec;16(12):e75847. [doi: 10.7759/cureus.75847] [Medline: 39822475]
14. Park SH, Kang SH, Lee SJ, et al. Actual compliance rate of Enhanced Recovery After Surgery protocol in laparoscopic distal gastrectomy. J Minim Invasive Surg 2021 Dec 15;24(4):184-190. [doi: 10.7602/jmis.2021.24.4.184] [Medline: 35602855]
15. Grimshaw JM, Shirran L, Thomas R, et al. Changing provider behavior: an overview of systematic reviews of interventions. Med Care 2001 Aug;39(8 Suppl 2):II2-II45. [Medline: 11583120]
16. Kwan JL, Lo L, Ferguson J, et al. Computerised clinical decision support systems and absolute improvements in care: meta-analysis of controlled clinical trials. BMJ 2020 Sep 17;370:m3216. [doi: 10.1136/bmj.m3216] [Medline: 32943437]
17. Grechuta K, Shokouh P, Alhussein A, et al. Benefits of clinical decision support systems for the management of noncommunicable chronic diseases: targeted literature review. Interact J Med Res 2024 Nov 27;13:e58036. [doi: 10.2196/58036] [Medline: 39602213]
18. Sunjaya AP, Ansari S, Jenkins CR. A systematic review on the effectiveness and impact of clinical decision support systems for breathlessness. NPJ Prim Care Respir Med 2022 Aug 20;32(1):29. [doi: 10.1038/s41533-022-00291-x] [Medline: 35987745]
19. Shanafelt TD, Dyrbye LN, West CP. Addressing physician burnout: the way forward. JAMA 2017 Mar 7;317(9):901-902. [doi: 10.1001/jama.2017.0076] [Medline: 28196201]

XSL•FO
RenderX

Abbreviations

ERAS: Enhanced Recovery After Surgery
MAUQ: mHealth App Usability Questionnaire
mHealth: mobile health
SQUIRE: Standards for Quality Improvement Reporting Excellence

Edited by A Kushniruk; submitted 04.Sep.2025; peer-reviewed by KH Lin; revised version received 21.Apr.2026; accepted 04.May.2026; published 05.Jun.2026.

Please cite as:
Nisar F, D Alessandro NM, Suratkal J, Zamberg I, Pope SE, Ali A, Tollinche LE
C8 Health, a Platform for the Implementation of Best Practices: Survey-Based Usability Study
JMIR Hum Factors 2026;13:e83547
URL: https://humanfactors.jmir.org/2026/1/e83547
doi:10.2196/83547

© Faria Nisar, Nicolas Mario D Alessandro, Jessica Suratkal, Ido Zamberg, Samantha Elen Pope, Ali Ali, Luis Etienne Tollinche. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 5.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Validation of the Taiwan Chinese Version of the Assistive Technology Usability Questionnaire for People With Neurological Diseases for Wearable Robotic Exoskeletons: Usability Study

Hander Wang[1*], MD; Jian-Jia Huang[2,3*], PhD; Yu-Cheng Pei[1,2,3], MD, PhD; Huey-Wen Liang[4,5], MD, PhD

[1]School of Medicine, Chang Gung University, Taouyuan, Taiwan

[2]Department of Physical Medicine and Rehabilitation, Chang Gung Memorial Hospital, No. 5, Fuxing St., Guishan District, Taouyuan, Taiwan

[3]Master of Science Degree Program in Innovation for Smart Medicine, Chang Gung University, Taouyuan, Taiwan

[4]Department of Physical Medicine and Rehabilitation, National Taiwan University College of Medicine, Taipei City, Taiwan

[5]Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, Taipei City, Taiwan

[*]these authors contributed equally

Corresponding Author:
Yu-Cheng Pei, MD, PhD
School of Medicine, Chang Gung University, Taouyuan, Taiwan

Abstract

Background: Assessing usability is important given the increasing use of technology in rehabilitation. While wearable robotic exoskeletons (WREs) are commonly incorporated into neurological rehabilitation in both clinical and community settings, there is a need to examine the applicability of different usability instruments in this context. The Assistive Technology Usability Questionnaire for People with Neurological Diseases (NATU Quest) was developed to evaluate the usability of assistive technology in individuals with neurological conditions. However, its reliability and validity have been established only for assistive devices such as wheelchairs and canes, restricting its generalizability.

Objective: This study aimed to cross-culturally adapt a Taiwan Chinese version of the NATU Quest and to preliminarily evaluate its applicability, reliability, and validity for assessing WRE use among patients who had a stroke as an expanding application context.

Methods: Cross-cultural adaptation was achieved through a structured translation from Spanish to Taiwan Chinese, followed by content validation by an expert panel of 10 raters. Following the adaptation process, psychometric testing of the NATU Quest was administered to 20 patients who had a stroke following the completion of a comprehensive course of robot-assisted gait training facilitated by a WRE. The internal consistency of the NATU Quest was evaluated using the Cronbach α coefficient. The reliability of the test-retest procedure was evaluated using the intraclass correlation coefficient (ICC). Convergent validity was ascertained by examining the correlation between the NATU Quest scores and the System Usability Scale (SUS).

Results: The culturally adapted NATU Quest showed acceptable content validity, with item-level content validity index (CVI) values ranging from 0.80 to 1.00 and scale-level CVI values of 0.93 for clarity and 0.97 for relevance. Cronbach α was 0.943, indicating high internal consistency. Of the 20 participants who completed all 6 rehabilitation sessions, 13 participants completed repeat testing. Preliminary test-retest reliability was acceptable (ICC=0.823; P<.001; SE of measurement=0.31). NATU Quest scores were strongly correlated with SUS scores (Spearman ρ=0.818; P<.001), providing preliminary support for convergent validity.

Conclusions: The adapted NATU Quest demonstrated promising preliminary psychometric properties when applied to WRE use among individuals who had a stroke. Further studies with larger sample sizes are warranted to examine additional psychometric properties such as construct validity and responsiveness.

(JMIR Hum Factors 2026;13:e89556) doi:10.2196/89556

KEYWORDS

usability; rehabilitation; exoskeletons; stroke; robot-assisted gait training; questionnaires; wearable robotic exoskeletons; assistive technology; psychometrics

XSL•FO
RenderX

Introduction

Wearable robotic exoskeletons (WREs) have been widely used for robot-assisted gait training (RAGT) to restore walking ability in individuals with neurological or mobility impairments for many years [1]. They offer high-intensity and task-specific training and have demonstrated benefits for stroke, spinal cord injury, Parkinson disease, and cerebral palsy [2-5]. Moreover, WREs are increasingly transitioning beyond clinical and laboratory settings toward widespread adoption in homes, communities, or even industrial settings, reflecting a broader application for home-based rehabilitation and assistive devices [6-8].

While the application of WREs has become increasingly popular in rehabilitation, usability assessment for their effective adoption and clinical integration has not yet been comprehensively addressed. According to the International Organization for Standardization, usability is defined as "the effectiveness, efficiency, and satisfaction with which users achieve specified goals in specific contexts of use" [9]. It is an important feature of an assisted technology in rehabilitation related to user comfort, feedback mechanisms, difficulties in donning and doffing, compliance mismatch, and environmental conditions [10-15]. The World Health Organization also recommends that usability assessment should be the initial step in evaluating any digital health intervention [16]. However, a systematic review by Koumpouros [17] indicated that 64% of studies relied on either custom-developed or nonvalidated instruments to assess the usability of assistive robotic technologies. The lack of a comprehensive psychometric evaluation of these instruments will likely raise doubts about reliability and validity, limiting comparability between studies.

Usability evaluation of WREs requires consideration of the elements that contribute to the interaction between the device and the user. These elements include aspects such as comfort, ease of donning and doffing, individual adaptability, perceived safety, and functional support [17]. However, there is limited research on usability evaluation instruments for WREs. Up to 67% of the measures used for subjective assessment of rehabilitation and assistive robot devices are custom-made and lack validation [17]. Among the standardized measures, two are most commonly used, that is, the System Usability Scale (SUS) [18] and the Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST 2.0) [19,20]. The SUS was originally developed for website usability. Despite its broad application to various products, it focused more on the software design and users' interaction with systems [21] without addressing the physical, task-dependent, and dynamic human-device interactions that are central to rehabilitation robotics [22]. The QUEST 2.0 demonstrates a stronger alignment with our intended focus on device-related usability; however, its inclusion of service-related domains, such as delivery, repairs, and follow-up services, limits its relevance in clinical scenarios where patients who had a stroke predominantly use WREs within hospital-based rehabilitation settings [19]. Recently, an Assistive Technology Usability Questionnaire for People with Neurological Diseases (NATU Quest) was developed and addresses the paucity of assessment tools covering specific domains relating to assistive technology, such as adaptability to user needs, perceived safety, contextual fit, and user experience [23]. These domains may represent a contextually appropriate measure of immediate user-device interaction in WRE-assisted gait training; however, their feasibility remains to be examined.

The NATU Quest is a self-reported questionnaire containing 10 items selected through 2 rounds of a Delphi study based on the opinion of 69 experts. It is scored on a 6-point Likert scale, with item scores ranging from 0 to 5. The total questionnaire score is calculated by summing the scores of valid responses to items 1 to 10 and dividing by the number of valid items, resulting in a final score ranging from 0 to 5, with higher scores indicating greater usability. The initial validation study assessed the usability of conventional assistive tools (eg, wheelchairs and walking aids) in patients with neurological impairments and demonstrated good internal consistency (Cronbach α=0.895), high test-retest reliability (intraclass correlation coefficient [ICC]=0.869), and construct validity (Spearman ρ=0.756 with 8 items of the QUEST 2.0) [23]. In a recent study, Bosch-Barceló et al [24] further applied the NATU Quest on a gamified virtual reality treadmill intervention used in rehabilitation for patients with Parkinson disease, expanding its application to different technology.

Our study had 2-fold goals. First, we would adapt a Taiwan Chinese version of the questionnaire following an established guideline to ensure the cross-culture equivalence of the questionnaire. Second, we examined whether the NATU Quest can be applied in a group of patients undergoing RAGT, a representative example of wearable robotic rehabilitation technology, and will preliminarily evaluate its psychometric properties. We hypothesized that the adapted questionnaire had acceptable reliability and validity. The results could provide pilot evidence to ascertain its suitability for usability assessments of WREs in future studies.

Methods

Cross-Cultural Adaptation

The adaptation of the Spanish version of the 10-item NATU Quest questionnaire was conducted according to the guidelines proposed in the study by Guillemin et al [25] after we obtained the permission from the original authors.

It was first forward translated into Taiwan Chinese by 2 native Taiwan Chinese translators proficient in Spanish without medical backgrounds. The first draft was concluded by a consensus-based expert panel consisting of a rehabilitation physician, an occupational therapist, and one of the translators to review both translations and resolve discrepancies. Back translation of the first draft into Spanish was conducted by an independent native Spanish translator fluent in both Taiwan Chinese and Spanish, who was blinded to the original NATU Quest. The expert panel conducted a second review to compare the back-translated version to the original Spanish NATU Quest and resolve discrepancies to ensure semantic, idiomatic, experiential, and conceptual equivalence. During this process, the panel was assisted by one of the original authors (Maria

XSL•FO
RenderX

Masbernat-Almenara) through personal communication to clarify minor ambiguities. Finally, the content validity index (CVI) of the finalized version was tested by a panel of 10 physical therapists and occupational therapists who are all familiar with the use of WREs. Each item was assessed based on clarity and relevance related to the WREs using a 4-point Likert scale [26]. The results were used to compute 2 indices: item-level content validity index (I-CVI) and scale-level content validity index (S-CVI) according to equations (1) and (2). The minimum threshold of the I-CVI and S-CVI was 0.78 and 0.8, respectively [26].

$$\text{I-CVI}=\frac{\text{Number of rating 3 or 4}}{\text{Total number of experts}} \quad (1)$$

$$\text{S-CVI}=\frac{\sum \text{I-CVI}}{\text{Total number of items}} \quad (2)$$

Study Design and Participants

This was an extended study of a prospective and pilot case-control study aiming at investigating the therapeutic effects of RAGT using a user-initiated powered WRE in patients in the early stages of recovery after stroke [27]. All RAGT sessions used the same WRE, Keeogo Exoskeleton (Wistron Medical Technology; Figure 1). It is an electrically powered robotic device that provides knee power assistance while preserving passive hip motion [28,29]. Participants were patients who had a stroke in the early subacute stage who could maintain basic postural stability and demonstrated sufficient lower-limb motor capacity to safely engage in user-initiated robotic gait training. Individuals with substantial cognitive impairment, severe musculoskeletal or cardiopulmonary conditions, insufficient joint range or muscle tone, or dermatological or vascular problems preventing safe device use were excluded [27]. Given the limited availability of clinically eligible participants for inpatient exoskeleton training, a pragmatic pilot sample was recruited.

XSL•FO
RenderX

Figure 1. A patient who had a stroke undergoing Keeogo exoskeleton robotic gait training.



Ethical Considerations

The study was approved by the institutional review board (IRB) of the Chang Gung Medical Foundation (IRB approval number 202200822B0, 2022) and adhered to the Declaration of Helsinki. Before enrollment, all participants were informed of the study purpose, procedures, and potential risks and benefits, and written informed consent was obtained from each participant. Participants' privacy and confidentiality were protected throughout the study. All research data were deidentified and stored securely, and only authorized members of the research team had access to the study data. No financial compensation was provided to participants.

Study Procedures

Each participant completed 6 sessions of the RAGT (40 minutes per session, 1 session per day, 2 - 3 sessions per week) and was invited to answer the NATU Quest and SUS after the final session (T1). The NATU Quest was repeated within 1 week after the final session (T2) for the evaluation of test-retest reliability [30].

XSL•FO
RenderX

The SUS is a standardized 10-item questionnaire used to assess perceived usability across different technologies. Each item is rated on a 5-point Likert scale, and the total score ranges from 0 to 100, with higher scores indicating higher perceived usability [18]. The SUS has been extensively studied and has demonstrated well-established reliability and validity across diverse technologies [21]. It has also been used for usability assessment in robotic rehabilitation and assistive robotic devices, including upper- and lower-limb exoskeletons, end effectors, and sensor-based systems [31-38]. Therefore, the SUS was used to assess convergent validity and to facilitate comparisons across studies. A validated Chinese version of SUS was used for the current study [39].

Statistical Analysis

Demographic data were analyzed using nonparametric and descriptive statistics for a small sample size. The psychometric properties of the adapted questionnaire were analyzed as follows:

1. The proportion of participants who achieved the lowest or highest possible score was computed. More than 15% of participants having the maximal or minimal score was defined as having a ceiling and floor effect, respectively.
2. The internal reliability of the scale at T1 was evaluated by calculating Cronbach α and Cronbach α if an item was deleted. Values of Cronbach α between 0.70 and 0.80 were considered good internal consistency, whereas values above 0.80 were considered very good [40]. Item-total statistics were computed to evaluate the corrected item-total correlation for each item, with values ≥0.3 considered acceptable and values ≥0.5 indicating strong homogeneity [41].
3. Test-retest reliability was evaluated with an ICC and their 95% CIs for scores obtained at T1 and T2. The ICC estimates were based on a single rating, absolute-agreement, and 2-way mixed-effects model, and interpreted as poor (<0.40), good (0.40 - 0.75), and excellent (>0.75) [42].
4. The SE of measurement (SEM) was calculated to reflect the variability in the scores of individual items in the questionnaire using the following formula, where $SD_{baseline}$ is the SD of the scores from the first measurement (T1) and ICC is the test-retest reliability between the T1 and the second measurement (T2):
   SEM=SDbaseline×1-ICC
   Subsequently, the minimal detectable change (MDC) was calculated to determine the minimum change in score that can be interpreted as a real difference, beyond measurement error, at a 95% confidence level with a *z* score of 1.96.
   MDC=1.96×SEM×2
   MDC was also expressed as a percentage of a parameter's grand mean (MDC%) by dividing the MDC by the mean scores of the NATU Quest at T1. An MDC% of <30% is considered acceptable [43].
1. Convergent validity was tested by computing the Spearman rank correlation coefficient (ρ) between the NATU Quest and SUS scores, and the size of the correlation coefficient was interpreted as being strong (>0.75), moderate (approximately 0.50-0.75), fair (approximately 0.25-0.50), and no relationship (approximately 0.00-0.25) [44].

All statistical analyses were performed using SPSS software (version 26.0; IBM Corp). The level of statistical significance was defined as $P<.05$.

Results

The NATU Quest was translated from the original Spanish version, and there was generally no dispute to form the consensus, except for the Q5 ("Me siento seguro usando/llevando...//...es seguro en su utilización"). Two versions of translation: "I feel that I can use the exoskeleton robot very safely" and "I feel that the exoskeleton robot is safe to use" were consolidated into "I feel safe when wearing the exoskeleton robot." The finalized Taiwan Chinese version of the NATU Quest was submitted to the original author of the NATU Quest for a final approval.

The I-CVI for all the items regarding the clarity and relevance of the adapted version exceeded the threshold of 0.78, while the S-CVI, calculated using the averaging method, was 0.93 for clarity and 0.97 for relevance. Therefore, the adapted version was used for further testing without modifications.

Twenty participants were enrolled in the validation study, most of whom were first-time survivors of ischemic stroke in the early subacute phase following a stroke (Table 1). They completed the 6 RAGT within an average of 20.3 (SD 4.5) days, and no adverse events were reported during the experimental process.

XSL•FO
RenderX

Table . Demographic characteristics of the participants (N=20).

Variables	Results
Age (years), mean (SD)	55.8 (11.4)
Sex, male, n (%)	15 (75)
Educational level, n (%)	
Middle school and below	4 (20)
High school or vocational school	8 (40)
College or above	8 (40)
Stroke-related characteristics	
Affected side, n (%)	
Left	12 (60)
Right	8 (40)
Stroke type, n (%)	
Ischemic	14 (70)
Hemorrhagic	6 (30)
First-time stroke, n (%)	18 (90)
Stroke duration in days, median (IQR)	37 (22 - 62)

The median scores of each item of the NATU Quest ranged from 4 to 5, and the overall scores were 4.4 and 4.1 on the two occasions (Table 2). No participants reported floor scores on the NATU Quest, whereas 2 (10%) participants reported ceiling scores in T1, below the 15% criterion for a ceiling effect. Meanwhile, none reported a ceiling or floor score with the SUS.

Table . Descriptive data and internal consistency measures of the Assistive Technology Usability Questionnaire for People with Neurological Diseases (NATU Quest) and System Usability Scale.

Components	Score, range	Values, median (IQR)	Item-total correlation	Cronbach α (if an item was deleted)
NATU Quest				
Q1. Effectiveness	3-5	5 (4-5)	0.791	0.937
Q2. Comfort	2-5	4 (4-5)	0.785	0.936
Q3. Adaptability or adjustability	2-5	4 (4-5)	0.768	0.936
Q4. Donning or doffing	1-5	4 (4-5)	0.736	0.940
Q5. Perceived safety	3-5	4.5 (4-5)	0.680	0.940
Q6. Functionality	2-5	4 (4-5)	0.804	0.935
Q7. Ergonomics or fit to body	3-5	4 (3.75 - 5)	0.829	0.934
Q8. Ease of use	2-5	4.5 (3.75 - 5)	0.868	0.932
Q9. Ease of learning	2-5	4 (4-5)	0.685	0.940
Q10. Satisfaction (overall)	3-5	5 (4-5)	0.780	0.936
Total scores				
T1	2.3-5	4.4 (3.9 - 4.7)	—[a]	—
T2	3.2-4.9	4.1 (3.4 - 4.6)	—	—
System Usability Scale				
T1	45-82.5	66.3 (60.6 - 75.0)	—	—

[a]Not applicable.

XSL•FO
RenderX

The Cronbach α of the NATU Quest at T1 was 0.943, and the Cronbach α if item deleted was between 0.932 and 0.940 (Table 2), lower than the Cronbach α of the overall scale. The item-total correlations were all above 0.5, with the highest for Q8 (ease of use: 0.868) and the lowest for Q9 (ease of learning: 0.685). This finding suggested that all items contributed to the internal consistency of the scale in this sample.

Thirteen participants completed the NATU Quest on 2 occasions, with a mean interval of 3.5 (SD 1.5) days, ranging from 2 to 6 days. There were no significant differences in demographic characteristics between participants who completed the survey at T2 and those who did not, including age (P=.70), sex (P=.61), stroke duration (P=.23), modified Rankin Scale (P=.59), and SUS scores (P=.88) by the Mann-Whitney U test. The ICC was 0.823 (95% CI 0.533 - 0.942; P<.001), indicating an excellent test-retest reliability. The SEM, MDC, and MDC% were 0.31, 0.86, and 20.1%, respectively. The collective results supported good reliability of the adapted NATU Quest.

The Spearman ρ coefficient between the NATU Quest and SUS scores was 0.818 (P<.001) in all 20 participants, indicating a strong correlation.

Discussion

Principal Findings

We adapted a Taiwan Chinese version of the NATU Quest to facilitate future cross-cultural comparisons in the usability assessment of assistive technology. We also conducted a preliminary evaluation of its psychometric properties in patients who had a stroke using WREs for RAGT, a rapidly evolving field. The results demonstrated acceptable CVI scores, test-retest reliability, and convergent validity when referenced with the SUS. By following the guidelines for the adaptation process, we ensured the contextual clarity of the adapted questionnaire. The reliability and convergent validity of the adapted NATU Quest support its potential use in the usability assessment of WREs. However, the findings should be interpreted with caution because of the modest sample size.

The 10 items of the NATU Quest were derived using the Delphi method, and the original validation study included wheelchairs, walking aids, splints, and treadmills, without specific focus on WREs. Similar to the SUS and QUEST 2.0, it is not a device-specific instrument. However, these validated instruments offer the advantages of requiring fewer resources and enabling comparability across studies [45]. Our results showed a high CVI from 10 therapists who were familiar with this WRE and confirmed the face validity. They generally align with the essential features of the medical device as emphasized by the guidelines established by the International Electrotechnical Commission and the US Food and Drug Administration guidelines [46,47]. However, caution should be exercised, as the use of a generic tool may fail to capture certain risks and usability dimensions specific to WREs [48]. WREs are available in many forms and multiple purposes, such as rehabilitation, assistive mobility, industrial and military use, sports and function, and personal mobility [49]. The WRE we tested, Keeogo, has been used in stroke, multiple sclerosis, and older adults with morbidities [27,50,51]. Its intended use includes not just for rehabilitation but also personal use. We did not attempt to add device-specific items during the adaptation process, as this could compromise the construct validity of the original questionnaire. Therefore, future users should be aware of the potentially limited coverage of certain device-specific concerns, including but not limited to biomechanical misalignment, falls, and soft-tissue pressure–related injuries.

Several psychometric properties tested in this study aligned closely with those reported for the original NATU Quest [23]. The median scores of the participants for Keeogo are between 4.1 and 4.4, around 82% to 88% of the highest scores, implying a generally positive perception of the device. Test-retest reliability (ICC=0.823) also paralleled the previously reported excellent temporal stability [23]. However, up to 10% of the participants reported ceiling scores at T1, and 2 items had median scores of 5, raising concerns regarding ceiling effects and restricted score variability in this specific clinical context. In addition, the internal consistency (Cronbach α=0.943) was higher than that reported in a previous study involving various devices and patients with diverse diagnoses [23]. Although the Cronbach α values if an item was deleted were mostly below 0.94, potential item overlap or redundancy should be examined in future studies with a larger sample size.

A strong correlation between the NATU Quest and SUS scores (Spearman ρ=0.818) supported convergent validity. Both the SUS and QUEST 2.0 have been used to assess the usability of rehabilitation robots [52,53], and QUEST 2.0 demonstrated greater overlap with the components evaluated by the NATU Quest than the SUS (Table 3). However, we selected the SUS over the QUEST 2.0 to test convergent validity because up to 4 items in the QUEST 2.0 were service-related, including service delivery, repairs, and follow-up services, and were therefore not applicable to the present hospital-based inpatient RAGT context. In contrast, the SUS is widely used and may facilitate comparisons across studies. Nonetheless, the convergent validity with respect to the SUS should be interpreted cautiously because of the limited overlap between questionnaire items. Future studies assessing the convergent validity of usability questionnaires for WRE may benefit from considering instruments in addition to, or instead of, the SUS.

XSL•FO
RenderX

Table . Components of different usability questionnaires.

Components	NATU Quest[a]	SUS[b]	QUEST 2.0[c]
Effectiveness	✓	✓	✓
Comfort	✓		✓
Adaptability or adjustability	✓		✓
Donning or doffing	✓		✓
Perceived safety	✓		✓
Functionality	✓		✓
Ergonomics or fit to body	✓		
Ease of use	✓	✓	✓
Ease of learning	✓	✓	
Satisfaction (overall)	✓	✓	✓
Device size or weight			✓
Durability			✓
Service or repair			✓

[a]NATU Quest: Assistive Technology Usability Questionnaire for People with Neurological Diseases.

[b]SUS: System Usability Scale.

[c]QUEST 2.0: Quebec User Evaluation of Satisfaction with Assistive Technology.

Limitations

This study has several limitations. First, the small sample size and the absence of longitudinal follow-up prevented the evaluation of essential psychometric properties, including construct validity, discriminant validity, and responsiveness. These properties should be examined in future studies before the adapted NATU Quest can be more broadly recommended for WRE usability assessment. Second, the inclusion of only patients who had a stroke receiving a single type of powered lower-limb WRE limited the generalizability of the findings, as WREs are designed for diverse patient populations and clinical applications. Future studies should evaluate the applicability of the NATU Quest across different neurological and musculoskeletal conditions and a broader range of WRE systems. Finally, the NATU Quest was originally developed as a general assistive technology usability instrument and may not fully capture WRE-specific concerns, such as biomechanical alignment, fall-related risks, or soft-tissue pressure–related issues. Given the wide variety of WREs currently available on the market, developing device-specific usability instruments for each system would be time-consuming. Therefore, the NATU Quest may provide a practical general usability measure, although complementary WRE-specific safety and biomechanical assessment items may still be required.

Conclusions

The present study adapted a Taiwan Chinese version of the NATU Quest and preliminarily evaluated its psychometric properties in patients who had a stroke undergoing RAGT with 1 powered lower-limb WRE. The findings provided preliminary evidence supporting its content validity, internal consistency, test-retest reliability, and convergent validity within this specific clinical context. Given that no device-specific instrument is currently available for assessing WRE usability, the NATU Quest may provide a useful general usability measure for patients who had a stroke receiving RAGT with a powered lower-limb WRE. However, given the safety-critical and biomechanical nature of user-device interactions in WREs, the NATU Quest may not fully capture these aspects of usability. Further studies are needed to evaluate its comprehensive psychometric properties, including construct validity, discriminant validity, and responsiveness, before the adapted NATU Quest can be broadly recommended for usability assessment of WREs across different devices, patient populations, and clinical settings.

Acknowledgments

The authors would like to express their sincere gratitude to Dr Maria Masbernat-Almenara from the University of Lleida for authorizing the translation of the original Assistive Technology Usability Questionnaire for People with Neurological Diseases and assisting in the finalization of the adapted questionnaire.

XSL•FO
RenderX

Funding

This work was supported by the Ministry of Science and Technology (MOST; grant MOST 111-WHK-051-0674) and the National Science and Technology Council (NSTC; grant NSTC113-2314-B182A-092-MY2) for study design and execution and by Chang Gung Medical Foundation (grant CMRPG3N0013) for manpower support.

Data Availability

The Taiwan Chinese version of the NATU Quest is available in Multimedia Appendix 1. Due to the involvement of human participants and institutional ethics restrictions, the data are not publicly available outside the journal submission. Additional information may be made available from the corresponding authors upon reasonable request and with approval from the institutional review board.

Authors' Contributions

HWL contributed to the conceptualization and methodology of the study, led the cross-cultural adaptation of the NATU Quest into Taiwan Chinese, conducted content validity index evaluation, performed formal analysis, acquired funding, and critically reviewed and edited the manuscript. YCP contributed to formal analysis and funding acquisition, conducted the test-retest reliability analysis, and provided supervision and critical review of the manuscript. HW was responsible for project administration, data collection, and data handling; conducted the investigation; drafted the original manuscript; and prepared the visualizations. JJH contributed to project administration, validation of the study, and drafting of the original manuscript. HW and JJH contributed equally to this work as first authors. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The Taiwan Chinese version of the Assistive Technology Usability Questionnaire for People with Neurological Diseases.
[DOCX File, 33 KB - humanfactors_v13i1e89556_app1.docx]

References

1. Cha JM, Hong J, Yoo J, Rha DW. Wearable robots for rehabilitation and assistance of gait: a narrative review. Ann Rehabil Med 2025 Aug;49(4):187-195. [doi: 10.5535/arm.250093] [Medline: 40819657]
2. Fortunati M, Febbi M, Negro M, Gennaro F, D'Antona G, Crisafulli O. Lower-limb exoskeletons for gait training in Parkinson's disease: the state of the art and future perspectives. Healthcare (Basel) 2024 Aug 16;12(16):1636. [doi: 10.3390/healthcare12161636] [Medline: 39201194]
3. Hunt M, Everaert L, Brown M, Muraru L, Hatzidimitriadou E, Desloovere K. Effectiveness of robotic exoskeletons for improving gait in children with cerebral palsy: a systematic review. Gait Posture 2022 Oct;98:343-354. [doi: 10.1016/j.gaitpost.2022.09.082] [Medline: 36306544]
4. Tan K, Koyama S, Sakurai H, Teranishi T, Kanada Y, Tanabe S. Wearable robotic exoskeleton for gait reconstruction in patients with spinal cord injury: a literature review. J Orthop Translat 2021;28:55-64. [doi: 10.1016/j.jot.2021.01.001] [Medline: 33717982]
5. Yang J, Zhu Y, Li H, Wang K, Li D, Qi Q. Effect of robotic exoskeleton training on lower limb function, activity and participation in stroke patients: a systematic review and meta-analysis of randomized controlled trials. Front Neurol 2024;15:1453781. [doi: 10.3389/fneur.2024.1453781] [Medline: 39193147]
6. Claeys R, Embrechts E, Bourazeri A, et al. Understanding the perspectives of older adults and physiotherapists on home-based lower-limb exoskeletons. Wearable Technol 2025;6:e31. [doi: 10.1017/wtc.2025.10015] [Medline: 40692807]
7. Phan TQ, Alghamdi M, Egwim M, Lee SW. Development and validation of Customizable, Adaptive SPring-actuated Hand EXoskeleton (CASPHEX) designed for home-based training of finger movements post-stroke. IEEE Trans Neural Syst Rehabil Eng 2025;33:4018-4030. [doi: 10.1109/TNSRE.2025.3615078] [Medline: 41004369]
8. Hofstetter S, Müller P, Behr D, et al. Evaluation of the HAL® lumbar type exoskeleton in long-term care: protocol for a mixed-methods feasibility study. Front Digit Health 2025;7:1591866. [doi: 10.3389/fdgth.2025.1591866] [Medline: 40901407]
9. ISO 9241-11:2018: Ergonomics of Human-System Interaction: Part 11: Usability: Definitions and Concepts, 2nd edition: International Organization for Standardization; 2018.
10. Baig MM, GholamHosseini H, Moqeem AA, Mirza F, Lindén M. A systematic review of wearable patient monitoring systems - current challenges and opportunities for clinical adoption. J Med Syst 2017 Jul;41(7):115. [doi: 10.1007/s10916-017-0760-1] [Medline: 28631139]

XSL•FO
RenderX

11. Tsertsidis A, Kolkowska E, Hedström K. Factors influencing seniors' acceptance of technology for ageing in place in the post-implementation stage: a literature review. Int J Med Inform 2019 Sep;129:324-333. [doi: 10.1016/j.ijmedinf.2019.06.027] [Medline: 31445274]
12. Nasr A, Inkol K, McPhee J. Safety in wearable robotic exoskeletons: design, control, and testing guidelines. J Mech Robot 2025 May;17(5):050801. [doi: 10.1115/1.4066900]
13. Wang H, Tao D, Yu N, Qu X. Understanding consumer acceptance of healthcare wearable devices: an integrated model of UTAUT and TTF. Int J Med Inform 2020 Jul;139:104156. [doi: 10.1016/j.ijmedinf.2020.104156] [Medline: 32387819]
14. Attig C, Franke T. Abandonment of personal quantification: a review and empirical study investigating reasons for wearable activity tracking attrition. Comput Hum Behav 2020 Jan;102:223-237. [doi: 10.1016/j.chb.2019.08.025]
15. Ravneberg B. Usability and abandonment of assistive technology. J Assist Technol 2012 Nov 30;6(4):259-269. [doi: 10.1108/17549451211285753]
16. Global strategy on digital health 2020-2025. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/9789240020924 [accessed 2026-07-07]
17. Koumpouros Y. A systematic review on existing measures for the subjective assessment of rehabilitation and assistive robot devices. J Healthc Eng 2016;2016:1048964. [doi: 10.1155/2016/1048964] [Medline: 27196802]
18. Brooke J. SUS: a "quick and dirty" usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation in Industry: CRC Press; 1996:189-194. [doi: 10.1201/9781498710411]
19. Demers L, Weiss-Lambrou R, Ska B. The Quebec User Evaluation of Satisfaction with assistive Technology (QUEST 2.0): an overview and recent progress. Technol Disabil 2002;14(3):101-105. [doi: 10.3233/TAD-2002-14304]
20. Demers L, Weiss-Lambrou R, Ska B. Development of the Quebec User Evaluation of Satisfaction with assistive Technology (QUEST). Assist Technol 1996;8(1):3-13. [doi: 10.1080/10400435.1996.10132268] [Medline: 10159726]
21. Lewis JR. The System Usability Scale: past, present, and future. Int J Hum Comput Interact 2018;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
22. Vianello L, Short M, Manczurowsky J, et al. Robot-mediated physical human-human interaction in neurorehabilitation: a position paper. arXiv. Preprint posted online on Jul 23, 2025. [doi: 10.48550/arXiv.2507.17561]
23. Masbernat-Almenara M, Rubi-Carnacea F, Opisso E, Duarte-Oller E, Medina-Casanovas J, Valenzuela-Pascual F. Developing an assistive technology usability questionnaire for people with neurological diseases. PLoS One 2023;18(1):e0281197. [doi: 10.1371/journal.pone.0281197] [Medline: 36719908]
24. Bosch-Barceló P, Masbernat-Almenara M, Martínez-Navarro O, Tersa-Miralles C, Pakarinen A, Fernández-Lago H. A gamified virtual environment intervention for gait rehabilitation in Parkinson's disease: co-creation and feasibility study. J Neuroeng Rehabil 2024 Jun 24;21(1):107. [doi: 10.1186/s12984-024-01399-6] [Medline: 38915103]
25. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. J Clin Epidemiol 1993 Dec;46(12):1417-1432. [doi: 10.1016/0895-4356(93)90142-n] [Medline: 8263569]
26. Haynes SN, Richard DC, Kubany ES. Content validity in psychological assessment: a functional approach to concepts and methods. Psychol Assess 1995;7(3):238-247. [doi: 10.1037/1040-3590.7.3.238]
27. Huang JJ, Chang SC, Lin LC, Cheng CH, Chang YH, Pei YC. Therapeutic effects of powered exoskeletal robot-assisted gait training in inpatients in the early stage after stroke: a pilot case-controlled study. J Neuroeng Rehabil 2024 Nov 26;21(1):206. [doi: 10.1186/s12984-024-01510-x] [Medline: 39593164]
28. Mcleod JC, Ward SJ, Hicks AL. Evaluation of the Keeogo™ dermoskeleton. Disabil Rehabil Assist Technol 2019 Jul;14(5):503-512. [doi: 10.1080/17483107.2017.1396624] [Medline: 29092649]
29. Huang JJ, Chang SC, Cheng CH, Wan T, Pei YC. Gait analysis in powered exoskeleton-assisted walking in patients with stroke: a case series cohort. In: 2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC): IEEE; 2023. [doi: 10.1109/APSIPAASC58517.2023.10317529]
30. Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. J Clin Epidemiol 2003 Aug;56(8):730-735. [doi: 10.1016/s0895-4356(03)00084-2] [Medline: 12954464]
31. Basla C, Hungerbühler I, Meyer JT, Wolf P, Riener R, Xiloyannis M. Usability of an exosuit in domestic and community environments. J Neuroeng Rehabil 2022 Dec 1;19(1):131. [doi: 10.1186/s12984-022-01103-6] [Medline: 36457037]
32. Guillén-Climent S, Garzo A, Muñoz-Alcaraz MN, et al. A usability study in patients with stroke using MERLIN, a robotic system based on serious games for upper limb rehabilitation in the home setting. J Neuroeng Rehabil 2021 Feb 23;18(1):41. [doi: 10.1186/s12984-021-00837-z] [Medline: 33622344]
33. Luger T, Bär M, Seibt R, Rieger MA, Steinhilber B. Using a back exoskeleton during industrial and functional tasks-effects on muscle activity, posture, performance, usability, and wearer discomfort in a laboratory trial. Hum Factors 2023 Feb;65(1):5-21. [doi: 10.1177/00187208211007267] [Medline: 33861139]
34. Maggio MG, Luca A, D'Agate C, Italia M, Calabrò RS, Nicoletti A. Feasibility and usability of a non-immersive virtual reality tele-cognitive app in cognitive rehabilitation of patients affected by Parkinson's disease. Psychogeriatrics 2022 Nov;22(6):775-779. [doi: 10.1111/psyg.12880] [Medline: 36319267]

XSL•FO
RenderX

35. Ranzani R, Albrecht M, Haarman CJ, et al. Design, characterization and preliminary usability testing of a portable robot for unsupervised therapy of hand function. Front Mech Eng 2023;8:2022. [doi: 10.3389/fmech.2022.1075795]
36. Rojo A, Castrillo A, López C, et al. PedaleoVR: usability study of a virtual reality application for cycling exercise in patients with lower limb disorders and elderly people. PLoS One 2023;18(2):e0280743. [doi: 10.1371/journal.pone.0280743] [Medline: 36812248]
37. Semprini M, Lencioni T, Hinterlang W, et al. User-centered design and development of TWIN-Acta: a novel control suite of the TWIN lower limb exoskeleton for the rehabilitation of persons post-stroke. Front Neurosci 2022;16:915707. [doi: 10.3389/fnins.2022.915707] [Medline: 36507352]
38. Yurkewich A, Ortega S, Sanchez J, Wang RH, Burdet E. Integrating hand exoskeletons into goal-oriented clinic and home stroke and spinal cord injury rehabilitation. J Rehabil Assist Technol Eng 2022;9:20556683221130970. [doi: 10.1177/20556683221130970] [Medline: 36212185]
39. Wang Y, Lei T, Liu X. Chinese System Usability Scale: translation, revision, psychological measurement. Int J Hum Comput Interact 2020;36(10):953-963. [doi: 10.1080/10447318.2019.1700644]
40. Nunnally JC, Bernstein IH. Psychometric Theory, 3rd edition: McGraw Hill; 1994.
41. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011 Jun 27;2:53-55. [doi: 10.5116/ijme.4dfb.8dfd] [Medline: 28029643]
42. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]
43. Huang SL, Hsieh CL, Wu RM, Tai CH, Lin CH, Lu WS. Minimal detectable change of the timed "Up & Go" test and the Dynamic Gait Index in people with Parkinson disease. Phys Ther 2011 Jan;91(1):114-121. [doi: 10.2522/ptj.20090126] [Medline: 20947672]
44. Fleiss JL. Design and Analysis of Clinical Experiments: John Wiley & Sons; 2011.
45. Ranganathan P, Caduff C. Designing and validating a research questionnaire - part 1. Perspect Clin Res 2023;14(3):152-155. [doi: 10.4103/picr.picr_140_23] [Medline: 37554243]
46. IEC 62366-1:2015: Medical Devices: Part 1: Application of Usability Engineering to Medical Devices, 1st edition: International Organization for Standardization; 2015.
47. Applying human factors and usability engineering to medical devices: guidance for industry and Food and Drug Administration staff. U.S. Food and Drug Administration. 2016. URL: https://www.fda.gov/media/80481/download [accessed 2026-07-07]
48. Massardi S, Pinto-Fernandez D, Babič J, et al. Relevance of hazards in exoskeleton applications: a survey-based enquiry. J Neuroeng Rehabil 2023 May 31;20(1):68. [doi: 10.1186/s12984-023-01191-y] [Medline: 37259115]
49. Foroutannia A, Mohammadian M. A comprehensive survey of lower limb assistive exoskeleton robots: models, dynamics, mechanics, and control. Robot Auton Syst 2026 Jan;195:105232. [doi: 10.1016/j.robot.2025.105232]
50. Yeh ST, Hsieh YS. Effectiveness of Keeogo exoskeleton-assisted rehabilitation on functional mobility in older adults with multimorbidity: a single-arm pilot study. Int J Gen Med 2025;18:6923-6932. [doi: 10.2147/IJGM.S541280] [Medline: 41244949]
51. McGibbon CA, Sexton A, Jayaraman A, et al. Evaluation of the Keeogo exoskeleton for assisting ambulatory activities in people with multiple sclerosis: an open-label, randomized, cross-over trial. J Neuroeng Rehabil 2018 Dec 12;15(1):117. [doi: 10.1186/s12984-018-0468-6] [Medline: 30541585]
52. Mol M, van Schaik A, Dozeman E, et al. Dimensionality of the System Usability Scale among professionals using internet-based interventions for depression: a confirmatory factor analysis. BMC Psychiatry 2020 May 12;20(1):218. [doi: 10.1186/s12888-020-02627-8] [Medline: 32398111]
53. Nicora G, Santangelo G, Marzero G, Dagliati A, Aprile I, Quaglini S. Exploring usability standards in robotic rehabilitation: a meta-analysis of system usability scale scores. IEEE Int Conf Rehabil Robot 2025;2025:476-482. [doi: 10.1109/ICORR66766.2025.11063026] [Medline: 40644156]

Abbreviations

CVI: content validity index
I-CVI: item-level content validity index
ICC: intraclass correlation coefficient
IRB: institutional review board
MDC: minimal detectable change
NATU Quest: Assistive Technology Usability Questionnaire for People with Neurological Diseases
QUEST 2.0: Quebec User Evaluation of Satisfaction with Assistive Technology
RAGT: robot-assisted gait training
S-CVI: scale-level content validity index
SEM: SE of measurement
SUS: System Usability Scale

XSL•FO
RenderX

WRE: wearable robotic exoskeleton

Edited by A Parush; submitted 15.Dec.2025; peer-reviewed by T Basmaji; revised version received 08.Jun.2026; accepted 11.Jun.2026; published 20.Jul.2026.

Please cite as:
Wang H, Huang JJ, Pei YC, Liang HW
Validation of the Taiwan Chinese Version of the Assistive Technology Usability Questionnaire for People With Neurological Diseases for Wearable Robotic Exoskeletons: Usability Study
JMIR Hum Factors 2026;13:e89556
URL: https://humanfactors.jmir.org/2026/1/e89556
doi:10.2196/89556

© Hander Wang, Jian-Jia Huang, Yu-Cheng Pei, Huey-Wen Liang. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 20.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Utility of Augmented Reality Glasses With Waveguide Optics and Facial Recognition Technology for Patient Verification in a Simulated Outpatient Setting: Physician Questionnaire Survey Study

Ryo Andy Ogasawara[1], MD; Soichiro Yoshida[1], MD, PhD; Kentaro Okamoto[2], MD, PhD; Eiki Nagaoka[3], MD, PhD; Hirotaka Sagawa[4], MD; Kazutaka Sumita[4], MD, PhD; Hajime Tanaka[1], MD, PhD; Satoshi Shiraga[5], MS; Tomohiro Amemiya[6], PhD; Yasuhisa Fujii[1], MD, PhD

[1]Department of Urology, Institute of Science Tokyo, 1-5-45 Yushima, Bunkyo-ku, Tokyo, Japan

[2]Department of Pediatric Surgery, Institute of Science Tokyo, Bunkyo-ku, Tokyo, Japan

[3]Department of Cardiovascular Surgery, Institute of Science Tokyo, Bunkyo-ku, Tokyo, Japan

[4]Department of Endovascular Surgery, Institute of Science Tokyo, Bunkyo-ku, Tokyo, Japan

[5]Cellid, Inc, Minato-ku, Tokyo, Japan

[6]Department of Electrical and Electronic Engineering, Institute of Science Tokyo, Meguro-ku, Tokyo, Japan

Corresponding Author:
Soichiro Yoshida, MD, PhD
Department of Urology, Institute of Science Tokyo, 1-5-45 Yushima, Bunkyo-ku, Tokyo, Japan

Abstract

Background: Augmented reality (AR) glasses with waveguide optics and facial recognition may enhance patient identification and clinical workflows. However, their usability among physicians has not yet been evaluated.

Objective: The objective of this study was to assess the usability and acceptability of prototype AR glasses with facial recognition in a simulated outpatient setting.

Methods: In May 2025, 14 urologists from a single institution tested prototype AR glasses developed by Cellid Inc. The device incorporates waveguide-based optical displays, a miniature camera, and deep learning–based facial recognition software. Physicians identified preregistered mock patients, with demographic and clinical data displayed on the right lens. After the trial, participants completed an anonymous questionnaire on usability, comfort, visibility, and self-reported symptoms using 3- or 5-point Likert scales.

Results: Eleven participants (79%; 95% CI 52% - 92%) found the device helpful in performing clinical procedures, and 12 (86%; 95% CI 60% - 96%) believed it improved patient safety. Most participants adapted quickly to wearing the device, and 79% (95% CI 52% - 92%) expressed a willingness to use it in future practice. Four of 13 respondents (31%; 95% CI 13% - 58%) selected “somewhat tired” for eye fatigue, and only 1 participant experienced minor and transient physical symptoms including headache and nausea. Three of 13 respondents (23%; 95% CI 8% - 50%) rated the glasses as slightly heavy, and among the 5 participants who wore prescription glasses, 4 (80%; 95% CI 38% - 96%) reported difficulty wearing the prototype over their glasses. Surrounding-environment visibility and perceived facial recognition responsiveness were generally favorable; however, 8 of 13 respondents (62%; 95% CI 36% - 82%) rated the displayed patient information as slightly difficult to see.

Conclusions: In this pilot study, AR glasses with facial recognition were well accepted by physicians. This technology has demonstrated perceived clinical utility and has the potential to assist medical professionals, though further refinement and larger clinical evaluations are needed.

(JMIR Hum Factors 2026;13:e95664) doi:10.2196/95664

KEYWORDS

augmented reality; automated facial recognition; smart glasses; data display; patient care

Introduction

Augmented reality (AR) is a technology that overlays digital content onto the real world in real time, thereby enhancing human perception of reality. AR technologies are advancing rapidly in many fields, and the use of AR in clinical settings is highly anticipated. Previous studies have proposed various clinical applications of AR and virtual reality, such as surgical

XSL•FO
RenderX

navigation and educational uses for medical students and residents [1-5]. The use of wearable devices in clinical settings requires hardware development and application implementation specifically designed for clinical use, which differs from consumer applications.

Patient misidentification is a critical issue in health care settings. Since patient identification errors can lead to serious medical malpractice, medical staff have to exercise the utmost caution when identifying patients. The accuracy of facial recognition technology is rapidly improving, and its application to patient identification shows great promise. By incorporating the latest AR and facial recognition technologies, health care workers will be able to identify patients more easily and accurately than is currently possible.

A previous study reported favorable overall usability of AR glasses equipped with facial recognition functionality among nurses working in an ophthalmic ward [6]. However, no previous studies have investigated the usability of AR glasses equipped with facial recognition functionality among doctors.

The purpose of this study was to evaluate the clinical usability of AR glasses that are equipped with facial recognition technology developed by Cellid Inc and that use a waveguide-based optical system. This evaluation was conducted through a questionnaire survey targeting physicians in a simulated outpatient scenario. The study was not designed to estimate face-matching accuracy or to determine effects on misidentification, patient safety, or clinical workflow.

Methods

Study Design

This study was conducted in May 2025 to evaluate the usability of AR glasses equipped with facial recognition technology in a simulated clinical setting. Urological physicians were invited to wear the AR device and experience its patient identification function using mock patients preregistered in the system. The pilot study used a convenience sample of urologists because urology outpatient practice was the initial formative use case at the investigators' institution. Participants were tasked with identifying preregistered mock patients using the facial recognition capabilities of the AR device.

Recruitment and Participants

A total of 14 physicians, including 9 urology specialists and 5 residents, were recruited from the Department of Urology at the Institute of Science Tokyo. All participants used the AR glasses under standardized conditions using mock patient data.

Device and System Description

The AR glasses used in this study were a see-through type with waveguide optics and weighed 58 g. The device included a miniature camera (Figure 1) mounted on the bridge and required a wired connection to a smartphone (ROG Phone 7, ASUS), weighing 239 g, via a 20 g cable, resulting in a total carry weight of 317 g. Since the prototype AR glasses require external power and do not support wireless data transmission, they must be physically tethered to a smartphone running facial recognition software.

XSL•FO
RenderX

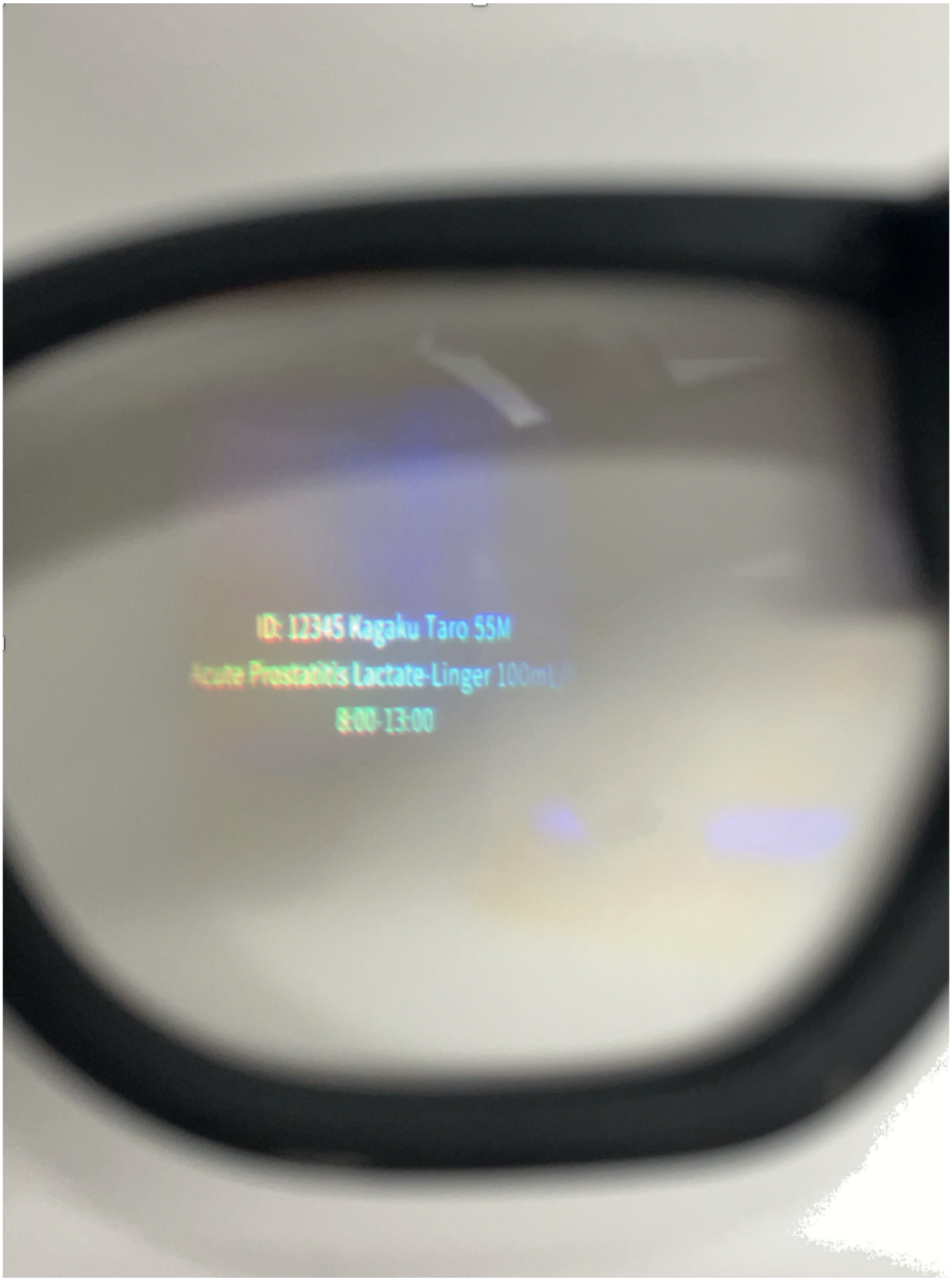
Figure 1. Photograph of a doctor wearing augmented reality glasses. A diffraction grating is installed on the right lens (framed red arrowhead), and a built-in miniature camera is located on the bridge (red arrow).



The facial recognition system was developed by Cellid Inc and used Amazon Rekognition (Amazon Web Services) for deep learning–based image recognition. Upon recognition of a registered face, patient information—including name, age, diagnosis, and allergies—was displayed in real time on the right lens via the waveguide optical system (Figure 2). This design enabled simultaneous visualization of the real-world environment and patient data. During the evaluated task, the display operated as an automatic, 1-way information presentation; direct manipulation of holographic content was not evaluated. Since the pilot study was designed as a user-experience demonstration rather than a technical validation study, recognition attempts, false matches, false rejections, repeated attempts, and software or network failures were not prospectively logged. Consequently, sensitivity, specificity, error rates, recognition success rates, and processing latency cannot be calculated from this study.

XSL•FO
RenderX

Figure 2. Patient information displayed on the augmented reality glasses lens.



Data Collection

After using the AR glasses, participants completed an anonymous questionnaire evaluating the glasses' usability and potential adverse effects. The investigator-developed questionnaire consisted of 15 items and an optional free-text field. The questionnaire was designed to assess the integrated prototype experience rather than facial recognition accuracy alone. Six domains were evaluated: perceived clinical utility (items 1 - 2), adaptation and intended adoption (items 3 - 4 and 10), wearing-related symptoms (items 5 - 9), weight and fit (item 11 and free-text comments), visual usability and task interference (items 12 - 14), and perceived facial recognition responsiveness (item 15). Items 1 - 11 and 15 used 5-category response scales, while items 12 - 14 used 3-category scales. A nonresponse was treated as missing data rather than a scale category. The questionnaire was not psychometrically validated,

XSL•FO
RenderX

and no composite usability score was calculated. The complete questionnaire is provided in Multimedia Appendix 1.

Statistical Analysis

Questionnaire data were analyzed using descriptive statistics. Categorical variables, including responses on usability, comfort, visibility, and adverse effects, are presented as frequencies and percentages. Ordinal variables collected via 3- or 5-point Likert scales are treated as categorical and summarized using counts and percentages for each response category. Continuous variables, such as participants' years of clinical experience, are reported as median and range. For all reported categorical outcomes, post hoc Wilson score 95% CIs were calculated to show sampling uncertainty. No hypothesis testing, subgroup comparisons, or adjustment for multiplicity was performed because of the small exploratory sample.

Ethical Considerations

This study was approved by the institutional review board of the Institute of Science Tokyo (approval number: I2024-070). All participants provided written informed consent prior to participation. This study used mock patient data to avoid the risks associated with handling real patient information, thereby minimizing ethical concerns related to privacy and data protection.

Results

Principal Findings

Fourteen urology specialists or residents with a median of 13 years of experience as physicians (range 3 - 24 years; IQR 3.75-17.75 years) were enrolled in this study. Thirteen (93%) participants were male and 1 (7%) was female. Five (36%) participants wore prescription glasses. All participants used the facial recognition function and completed the anonymous questionnaire. The results of the questionnaire are shown in Table 1.

XSL•FO
RenderX

Table . Results of questions.

Domain	Questions	Responses (N=14), n
Perceived clinical utility	Did the use of the AR[a] glasses assist you in performing medical procedures?	Not at all, 0 Slightly unhelpful, 0 Neutral, 3 Slightly helpful, 5 Helpful, 6
Perceived clinical utility	Did the use of the AR glasses help improve safety in clinical practice?	Not at all, 0 Slightly unhelpful, 0 Neutral, 2 Slightly helpful, 3 Helpful, 9
Adaptation and adoption	Did it take time for you to get used to wearing the AR glasses?	Not at all, 6 Minimal, 2 Neutral, 1 Somewhat, 4 Very much, 1
Adaptation and adoption	Do you wish to use AR glasses in future medical procedures?	Not at all, 0 Unlikely, 0 Neutral, 3 Somewhat, 4 Strongly, 7
Self-reported symptoms	To what extent did you experience eye fatigue from using the AR glasses?	Not at all, 5 Minimal, 2 Neutral, 2 Somewhat tired, 4 Very tired, 0 No response, 1
Self-reported symptoms	Did you experience physical fatigue while using the AR glasses?	Not at all, 4 Minimal, 8 Neutral, 1 Somewhat tired, 1 Very tired, 0
Self-reported symptoms	Did you experience any headaches while using the AR glasses?	Not at all, 7 Minimal, 6 Neutral, 0 Somewhat, 1 Significantly, 0
Self-reported symptoms	Did you experience any dizziness while using the AR glasses?	Not at all, 7 Minimal, 6 Neutral, 1 Somewhat, 0 Significantly, 0
Self-reported symptoms	Did you experience any nausea while using the AR glasses?	Not at all, 7 Minimal, 6 Neutral, 0 Somewhat, 1 Significantly, 0

XSL•FO
RenderX

Domain	Questions	Responses (N=14), n
Adaptation and adoption	Would you recommend the use of AR glasses to other medical professionals?	Not at all, 1 Unlikely, 0 Neutral, 3 Somewhat, 7 Strongly, 3
Weight and fit	What did you think about the weight of the AR glasses?	Very light, 3 Slightly light, 1 Just right, 6 Slightly heavy, 3 Very heavy, 0 No response, 1
Visual usability	How was the visibility of patient information through the AR display?	Very hard to see, 0 Slightly hard to see, 8 Easy to see, 5 No response, 1
Visual usability	How was the visibility of the surrounding environment?	Very hard to see, 0 Slightly hard to see, 3 Easy to see, 10 No response, 1
Task interference	To what extent did you find the AR glasses interfered with clinical tasks?	Very obstructive, 0 Slightly obstructive, 3 Not obstructive, 10 No response, 1
Perceived responsiveness	How did you find the responsiveness of the facial recognition function?	Too slow, 0 Slightly slow, 2 Just right, 10 Slightly fast, 1 Too fast, 0 No response, 1

[a]AR: augmented reality.

Perceived Clinical Utility and Future Adoption

For perceived assistance with clinical procedures, 5 of the 14 participants (36%) rated the device as “slightly helpful,” and 6 (43%) rated it as “helpful.” In total, 11 participants selected “slightly helpful” or “helpful” (79%; Wilson 95% CI 52% - 92%); the remaining 3 selected “neutral.” Similarly, for perceived support of clinical safety, 12 selected “slightly helpful” or “helpful” (86%; 95% CI 60% - 96%), with 3 (21%) selecting “slightly helpful” and 9 (64%) selecting “helpful”; 2 (14%; 95% CI 4% - 39%) selected “neutral.” Eleven of 14 participants expressed willingness to use the glasses in future practice (79%; 95% CI 52% - 92%), and 10 of 14 would recommend the prototype to other professionals (71%; 95% CI 45% - 88%).

Wearability and Adaptation

of the 14 participants, 8 (57%; 95% CI 33% - 79%) reported that no or minimal time was needed to adapt to the glasses, 1 (7%; 95% CI 1% - 30%) selected “neutral,” 4 (29%; 95% CI 12% - 55%) selected “somewhat,” and 1 (7%; 95% CI 1% - 30%) selected “very much.” Among 13 respondents to the weight item, 3 (23%; 95% CI 8% - 50%) rated the glasses as very light, 1 (8%; 95% CI 1% - 33%) as slightly light, 6 (46%; 95% CI 23% - 71%) as just right, and 3 (23%; 95% CI 8% - 50%) as slightly heavy; 1 participant did not answer. In free-text comments, 4 of 5 participants (80%; 95% CI 38% - 96%) using prescription glasses mentioned difficulty wearing the prototype over their prescription glasses.

Visual Usability, Task Interference, and Perceived Responsiveness

Patient information legibility was an important concern. Among 13 respondents, 8 (62%; 95% CI 36% - 82%) rated the displayed patient information as slightly hard to see, while 5 (38%; 95% CI 18% - 65%) rated it as easy to see; 1 participant did not answer. By contrast, 10 of 13 respondents (77%; 95% CI 50% - 92%) rated the surrounding environment as easy to see, and 3 (23%; 95% CI 8% - 50%) rated it as slightly hard to see. Ten of 13 (77%; 95% CI 50% - 92%) considered the device nonobstructive and 3 (23%; 95% CI 8% - 50%) as slightly obstructive. Perceived facial recognition responsiveness was rated as just right by 10 of 13 respondents (77%; 95% CI 50% - 92%), slightly slow by 2 (15%; 95% CI 4% - 42%), and

XSL•FO
RenderX

slightly fast by 1 (8%; 95% CI 1% - 33%); 1 participant did not answer. These ratings do not constitute an objective latency or accuracy assessment.

Self-Reported Symptoms

For eye fatigue, 5 of 13 respondents (38%; 95% CI 18% - 65%) selected "not at all," 2 (15%; 95% CI 4% - 42%) selected "minimal," 2 (15%; 95% CI 4% - 42%) selected "neutral," and 4 (31%; 95% CI 13% - 58%) selected "somewhat tired"; none selected "very tired," and 1 participant did not answer. Mild physical fatigue, headaches, and nausea were reported by a single participant. However, the reported symptoms were confined to the trial period and resolved spontaneously. No participant selected the highest-severity category for physical fatigue, headache, dizziness, or nausea.

Discussion

Principal Results

This study reported generally favorable reactions from doctors to the use of facial recognition–equipped AR glasses in a simulated clinical setting. Most participants considered the AR glasses to be useful both for performing medical procedures and in enhancing clinical safety.

Huang et al [6] reported favorable overall usability of an AR-AI assistant among 15 ophthalmic day-ward nurses, and Lin et al [7] evaluated an intelligent verification tool for ophthalmic surgical information among surgical nurses. To the best of our knowledge, this is the first study to evaluate the usability of AR glasses equipped with facial recognition functionality among physicians.

While no serious self-reported symptoms were observed in association with wearing the AR glasses, it is noteworthy that their use may induce mild symptoms, including eye fatigue, physical fatigue, headache, and nausea. Moreover, the AR glasses used in this study have room for technical improvement. Although the device used waveguide optics that enabled a more compact design compared to other AR display technologies [8] and achieved weight reduction through wired connection to a smartphone, 21% of participants rated it as "slightly heavy." There is also strong anticipation of the development of wireless devices in the future. While this study only evaluated the facial recognition functionality of the AR glasses and limited the display to basic patient information, the device could potentially display a wider range of information, such as CT scans and real-time vital signs, during procedures. This could support doctors not only in outpatient care but also during surgery and other medical interventions. Further research is needed to evaluate the utility of AR glasses equipped with enhanced functions in clinical settings. Furthermore, AR glasses have potential applications for health care professionals beyond physicians. Our previous study [9] demonstrated that when scrub nurses used see-through head-mounted displays, it facilitated their understanding of surgical procedures and improved the efficiency of instrument handling. Additionally, since nurses and medical administrative staff are also involved in patient identification, implementing facial recognition technology could potentially benefit all medical staff. However, facial recognition in this setting should be regarded as an adjunct rather than a replacement for established patient identification procedures. Wristbands, barcodes or QR codes, and confirmation using multiple identifiers have the advantage of being embedded in established workflows. A near-eye system may offer hands-free access to contextual information and reduce the need to look away from the patient, but face matching can be affected by lighting, pose, occlusion, masks, image quality, database composition, demographic performance variation, and the selected decision threshold. A false match may also create automation bias. Therefore, practical clinical use requires a fail-safe system for unrecognized faces, as well as independent confirmation using standard identification methods.

Limitations

This study was limited by a small sample size. Consequently, the 95% CIs for the survey responses are notably wide (eg, 52% - 92% for 11 of 14 participants), which underscores the substantial uncertainty in our point estimates. Although the prototype used a cloud-based facial recognition service, this study did not evaluate the data routing, retention, deletion, access control, or regulatory requirements necessary for clinical deployment. As this was a pilot usability study, precise device use duration, detailed facial data registration procedures, and database size were not recorded. The usability of the AR glasses was assessed subjectively via a questionnaire lacking objective indicators. Similarly, participants' prior experience with AR technologies and familiarity with wearable devices were not recorded, which might have influenced individual adaptation times and usability ratings. Furthermore, the potential risk of cognitive overload associated with processing simultaneous real-world visual inputs and AR-displayed information was not evaluated in this study. Industry provision of the prototype and the participation of a company executive as an author may have introduced expectancy bias despite anonymous responses. Finally, no real patient clinical data were used; the displayed demographic and clinical information was fictitious. Because this study was conducted in a simulated setting rather than actual clinical practice, future research is needed to evaluate the system's acceptability from the patients' perspective.

Conclusions

We found overall favorable responses from doctors regarding the use of AR glasses using waveguide optics with facial recognition functionality in a simulated clinical setting. This technology has the potential to assist medical professionals, providing a safeguard against patient misidentification.

Acknowledgments

We gratefully acknowledge Cellid Inc for supplying the prototype augmented reality glasses used in this study.

XSL•FO
RenderX

During preparation and revision of this manuscript, Gemini (Google) and ChatGPT (OpenAI) were used to assist English-language editing and organization of the revised text. Generative AI was not used to generate study data or perform the statistical calculations. All authors reviewed and revised the resulting text and take full responsibility for the final content.

Funding

This work was supported by Japan Science and Technology Agency CREST, grant number JPMJCR24R1.

Data Availability

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: SY
Formal analysis: RAO
Funding acquisition: SY
Investigation: RAO
Methodology: SY
Project administration: SY, TA
Resources: SS
Supervision: YF
Writing – original draft: RAO
Writing – review & editing: SY, KO, EN, HS, SK, HT

Conflicts of Interest

SS is the CEO of Cellid Inc. The company supplied the prototype as in-kind support. Cellid Inc had no role in the study design, data collection, data analysis, interpretation of data, or manuscript preparation. The other authors declare no conflicts of interest.

Multimedia Appendix 1
Survey questionnaire on augmented reality glasses' usability and potential adverse effects.
[DOCX File, 14 KB - humanfactors_v13i1e95664_app1.docx]

References

1. Ayoub A, Pulijala Y. The application of virtual reality and augmented reality in oral & maxillofacial surgery. BMC Oral Health 2019 Nov 8;19(1):238. [doi: 10.1186/s12903-019-0937-8] [Medline: 31703708]
2. Taghian A, Abo-Zahhad M, Sayed MS, Abd El-Malek AH. Virtual and augmented reality in biomedical engineering. Biomed Eng Online 2023 Jul 31;22(1):76. [doi: 10.1186/s12938-023-01138-3] [Medline: 37525193]
3. Daher M, Ghanimeh J, Otayek J, Ghoul A, Bizdikian AJ, El Abiad R. Augmented reality and shoulder replacement: a state-of-the-art review article. JSES Rev Rep Tech 2023 Aug;3(3):274-278. [doi: 10.1016/j.xrrt.2023.01.008] [Medline: 37588507]
4. Cannizzaro D, Zaed I, Safa A, et al. Augmented reality in neurosurgery, state of art and future projections. A systematic review. Front Surg 2022;9:864792. [doi: 10.3389/fsurg.2022.864792] [Medline: 35360432]
5. Yamazaki A, Ito T, Sugimoto M, et al. Patient-specific virtual and mixed reality for immersive, experiential anatomy education and for surgical planning in temporal bone surgery. Auris Nasus Larynx 2021 Dec;48(6):1081-1091. [doi: 10.1016/j.anl.2021.03.009] [Medline: 34059399]
6. Huang C, Sheng Y, Lian H, et al. AR-AI assisted ophthalmic nursing: preliminary usability study in clinical settings. Digit Health 2024;10:20552076241269470. [doi: 10.1177/20552076241269470] [Medline: 39257872]
7. Lin H, Huang X, Sheng Y, et al. Intelligent verification tool for surgical information of ophthalmic patients: a study based on artificial intelligence technology. J Patient Saf 2025 Mar 1;21(2):62-68. [doi: 10.1097/PTS.0000000000001295] [Medline: 39432546]
8. Koulieris GA, Akşit K, Stengel M, Mantiuk RK, Mania K, Richardt C. Near‐eye display and tracking technologies for virtual and augmented reality. Comput Graph Forum 2019 May;38(2):493-519. [doi: 10.1111/cgf.13654]
9. Yoshida S, Sasaki A, Sato C, et al. A novel approach to surgical instructions for scrub nurses by using see-through-type head-mounted display. Comput Inform Nurs 2015 Aug;33(8):335-338. [doi: 10.1097/CIN.0000000000000153] [Medline: 26018576]

Abbreviations

AR: augmented reality

XSL•FO
RenderX

Edited by A Kushniruk; submitted 18.Mar.2026; peer-reviewed by C Corciova, K Akşit, P O'Donnell; revised version received 08.Aug.2026; accepted 08.Sep.2026; published 02.Oct.2026.

Please cite as:
Ogasawara RA, Yoshida S, Okamoto K, Nagaoka E, Sagawa H, Sumita K, Tanaka H, Shiraga S, Amemiya T, Fujii Y
Utility of Augmented Reality Glasses With Waveguide Optics and Facial Recognition Technology for Patient Verification in a Simulated Outpatient Setting: Physician Questionnaire Survey Study
JMIR Hum Factors 2026;13:e95664
URL: https://humanfactors.jmir.org/2026/1/e95664
doi:10.2196/95664

© Ryo Andy Ogasawara, Soichiro Yoshida, Kentaro Okamoto, Eiki Nagaoka, Hirotaka Sagawa, Kazutaka Sumita, Hajime Tanaka, Satoshi Shiraga, Tomohiro Amemiya, Yasuhisa Fujii. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 2.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Feasibility and Preferences to Adopt mHealth-Based Interventions for HIV Prevention Among High-Risk Groups: Cross-Sectional Study

Faizur Rehman[1], PharmD; Muhammad Wasay Shahid[1], PharmD; Mehran Riaz[1], PharmD; Malik Muhammad Umair[2], MBBS, MPH; Farah Azhar[3], PharmD, PhD; Adeel Siddiqui[4], PharmD, MPhil; Ali Ahmed[5], PharmD, PhD

[1]Department of Pharmacy Practice, Riphah Institute of Pharmaceutical Sciences, Riphah International University, Islamabad, Pakistan

[2]Common Management Unit, National Institute of Health, Islamabad, Pakistan

[3]Department of Pharmaceutics, Riphah Institute of Pharmaceutical Sciences, Riphah International University, Islamabad, Pakistan

[4]Department of Pharmacy, Shaukat Khanum Memorial Cancer Hospital and Research Center, Lahore, Pakistan

[5]Division of Infectious Diseases and Global Public Health, School of Medicine, University of California, San Diego, La Jolla, CA, United States

Corresponding Author:
Ali Ahmed, PharmD, PhD
Division of Infectious Diseases and Global Public Health, School of Medicine, University of California, San Diego, La Jolla, CA, United States

Abstract

Background: HIV disproportionately affects men who have sex with men (MSM), transgender individuals, and people who inject drugs, with 70.1% of cases in Pakistan linked to these groups, aggravated by stigma and legal constraints. Mobile health (mHealth) interventions offer the potential to enhance HIV prevention, yet their acceptability remains underexplored.

Objective: This study aimed to assess the feasibility, willingness to use, and preferences for mHealth-based interventions designed to improve HIV knowledge, testing uptake, and risk behavior reduction among MSM, transgender individuals, and people who inject drugs in Pakistan.

Methods: A cross-sectional survey was conducted from February 2025 to May 2025 with 210 participants, including MSM and transgender individuals, recruited through respondent-driven sampling in collaboration with community-based organizations and nongovernmental organizations, while people who inject drugs were recruited from the rehabilitation centers of trusted community-based organizations and nongovernmental organizations, to ensure privacy and reduce safety concerns. Participants completed a self-administered questionnaire covering demographics, HIV knowledge, access to and use of mobile technology, acceptability and preferences for mHealth interventions, and HIV-related risk behaviors.

Results: The majority of participants owned smartphones (161/210, 76.7%) and had daily internet access (115/210, 54.8%), spending an average of 3.41 (SD 3.10) hours online daily. High willingness (182/210, 86.7%) to use a clinic-integrated mHealth app for HIV prevention was reported. Preferred app-delivered features included educational resources (85/210, 40.5%), daily medication reminders (108/210, 51.4%), weekly HIV prevention information (110/210, 52.4%), e-consultation (158/210, 75.2%), and mental health information (134/210, 63.8%). Smartphone ownership (odds ratio 4.14, 95% CI 1.41 - 12.21; *P*=.009) and prior experience with an mHealth app (odds ratio 2.97, 95% CI 1.00 - 8.81; *P*=.05) were significantly associated with willingness to adopt an mHealth app. While 80.5% (169/210) reported some knowledge of HIV, only 17.1% (29/169) rated their knowledge high, and 46.1% (97/210) were unaware of their HIV status. In terms of vulnerability, participants reported that injected drug use (122/210, 58.1%), experiences of violence (87/210, 41.4%), and police detention (128/210, 61%) were prevalent. In terms of risky sexual behaviors, nearly half engaged in transactional sex (104/210, 49.5%) or had multiple partners (101/210, 48.1%), while consistent condom use remained low (90/210, 42.9%).

Conclusions: There is high acceptability and considerable potential for mHealth interventions to enhance HIV prevention efforts among MSM, transgender individuals, and people who inject drugs in Pakistan. Tailored, discreet, and comprehensive mHealth platforms that include educational content, medication reminders, e-consultation, and mental health support are warranted. Addressing criminalization and ensuring user privacy will be crucial for the successful design and implementation of these interventions. Future research should focus on implementation research to assess the real-world uptake of these findings.

(JMIR Hum Factors 2026;13:e81111) doi:10.2196/81111

XSL•FO
RenderX

KEYWORDS

mobile health; mHealth; mHealth apps; smartphone; mobile phone; HIV; HIV prevention; smartphone apps; health technology; key populations; marginalized communities; digital health

Introduction

Despite remarkable advancements in HIV care and prevention, the acquisition of HIV remains a significant public health issue globally [1]. As of 2023, approximately 39.9 million people were living with HIV worldwide, including 1.3 million who were newly diagnosed, around 0.45 million of these in Southeast Asia [2]. The epidemic still disproportionately impacts key populations: the HIV prevalence was 11 times higher among gay men and other men who have sex with men, 7 times higher among people who inject drugs, and 14 times higher among transgender individuals, in comparison to the general adult population [3]. Discrimination, stigmatization, and punitive laws targeting key populations are costing lives and preventing the world from achieving the agreed global 95-95-95 AIDS target: 95% know their HIV status, 95% diagnosed with HIV are on antiretroviral therapy (ART), and 95% of those on ART are virally suppressed. Structural barriers, including poverty, limited health care access, and legal and human rights challenges, raise the likelihood of HIV infection 35-fold for people who inject drugs, 30-fold for female sex workers, and 14-fold for transgender women compared to adults in the general population [4]. Globally, studies confirm that HIV incidence remains high among sex workers, especially in countries that criminalize sex work compared to their counterparts in countries where sex work is wholly or partially legalized [5].

Mobile health (mHealth), defined as the use of mobile technologies for health services, has emerged as a promising tool for improving global health care access and delivery, improving health outcomes, and promoting preventive health behaviors [1,6]. mHealth, particularly smartphone apps and messaging, holds great promise for HIV prevention, especially when linked to HIV testing and pre-exposure prophylaxis (PrEP). The World Health Organization (WHO) and the Joint United Nations Program on HIV/AIDS have endorsed the integration of mHealth in HIV care strategies due to its reach across low-, middle-, and high-income countries [7]. For example, a study from China found that men who have sex with men who received customized short-message interventions reported significantly lower rates of risky sexual activities like multiple sex partners and condomless anal sex than those who did not receive the intervention [8].

Pakistan, the world's fifth-most populous country, has an estimated 0.33 million people living with HIV. According to the latest data, only 23% of people living with HIV know their HIV status; among them, 74% are on ART [9]. In the absence of recent national viral suppression data, estimates can be drawn from the 2023 Joint United Nations Program on HIV/AIDS report, which indicated that 34,000 individuals on ART have achieved viral suppression [2]. When compared with ART enrollment figures reported by the country's National AIDS Control Program, this suggests an approximate viral load of 63% among people living with HIV on ART in Pakistan [9]. Pakistan continues to face significant funding constraints in its HIV/AIDS response. The reallocation of HIV/AIDS-related resources to address the COVID-19 pandemic has further strained an already under-resourced health system, leading to disruptions in essential services for sexual and gender diverse populations [10]. Notably, in 2025, the HIV/AIDS component implemented by the National AIDS Control Program in collaboration with the United Nations Development Program (UNDP) Pakistan has experienced a reduction exceeding US $4 million [11]. Pakistan is still falling short of achieving the previous 90-90-90 global target set for 2020, and remains far from the current 2025 goals [12,13]. The number of new HIV infections in Pakistan increased from 14,000 in 2010 to 25,000 in 2020 [14]. AIDS-related deaths in Pakistan surged from 1400 in 2010 to 8000 in 2020 [3]. The HIV epidemic in Pakistan has shown evidence of shifting from isolated key populations to spreading into broader sexual networks. In Pakistan, 70.1% of HIV cases belong to key groups: men who have sex with men (22%), people who inject drugs (38.4%), female sex workers (2.2%), and transgender individuals (7.5%) [15]. Legal provisions such as Section 377 of the Pakistan Penal Code further exacerbate stigma and limit access to essential HIV services [16]. A total of 190 million cellular mobile connections were active as early as 2025, equivalent to 75.2% of the total population, and over 45.7% were using the internet [17]. Recent studies in other Asian and low- and middle-income countries have shown strong willingness among men who have sex with men, transgender individuals, and people who inject drugs to engage with mHealth platforms for HIV-related services [1,6,18].

Despite this, data on the feasibility and acceptability of mHealth solutions in Pakistan remain sparse, particularly among the most at-risk communities. A few digital platforms, such as "Sehat Dost" UNDP with support from the Global Fund and "HIV/AIDS Pakistan" by the Association of People Living with HIV, offer HIV information, self-assessment tools, and self-testing kits [19]. However, these apps lack integration, user-centered features, and functionality typical of comprehensive mHealth solutions. To address this, this study explores the willingness, preferences, and feasibility of adopting mHealth-based interventions among men who have sex with men, transgender individuals, and people who inject drugs in Pakistan. It aims to generate evidence for designing effective, user-centered digital tools that have the potential to improve HIV knowledge, increase testing uptake, and reduce high-risk behaviors in marginalized communities.

Methods

Study Design and Setting

A cross-sectional study was conducted from February 2025 to May 2025, using the quantitative survey-based method to assess the willingness, preferences, and feasibility of mHealth interventions for HIV prevention among the key populations in Pakistan, with data collected through collaboration with

XSL•FO
RenderX

community-based organizations (CBOs) and nongovernment organizations (NGOs). The STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guideline (Checklist 1) was followed to conduct this study.

Sample Size

The minimum sample size for estimating the proportion willing to use mHealth was first calculated with the single-proportion formula (95% CI, 5% precision, and anticipated prevalence=50%), giving 384 individuals. Because the accessible sampling frame in our sites was approximately 500 individuals, we applied a finite-population correction [20], resulting in 218 individuals. Anticipating minor design clustering across the 3 key population strata, a design effect of 1.1 was used, yielding approximately 240 individuals. However, the final achieved sample size was limited to 210 due to practical constraints such as limited recruitment time and resource limitations. Furthermore, many individuals from marginalized populations were reluctant to participate because of fears related to stigma, confidentiality, and social consequences. Although the final sample size was slightly lower than the initial target, it does not meaningfully compromise the overall interpretability of our findings. A post hoc power check indicates that with a sample of 210 participants, the study still retains acceptable precision for estimating willingness proportions with approximately ±6% to 7% precision at a 95% CI [21,22].

Definitions of Key Terms

Key Population

Groups who, due to specific higher-risk behaviors, are at increased risk of HIV, irrespective of the epidemic type or local context. Also, they often have legal and social issues related to their behaviors that increase their vulnerability to HIV. According to WHO guidelines, key populations relevant to this study include: (1) men who have sex with men, (2) people who inject drugs, and (3) transgender individuals [23].

Men Who Have Sex With Men

This group encompasses male individuals who engage in sexual activities with other male individuals, regardless of their sexual orientation or identity. The term is used to focus on behavior rather than identity, recognizing that not all men who have sex with men identify as gay or bisexual [24].

People Who Inject Drugs

Individuals who inject psychotropic (or psychoactive) substances for nonmedical purposes through intravenous, intramuscular, subcutaneous, or other injectable routes. People who inject drugs are at heightened risk of HIV due to practices such as sharing contaminated injection equipment. This definition also applies to people who inject other nonmedical substances.

Transgender Individuals

An umbrella term for individuals whose gender identity and expression do not conform to the norms and expectations traditionally associated with the sex assigned to them at birth; it includes people who are transsexual, transgender, or otherwise gender nonconforming. Transgender people may self-identify as transgender, female, male, a transgender woman, or a transgender man [25].

Study Participants and Procedures

Inclusion criteria were (1) being aged 18 years or older; (2) self-identified as men who have sex with men, transgender individuals, or people who inject drugs; (3) resident of Pakistan; and (4) being able to understand Urdu or English.

Other key populations, such as female sex workers and people in prisons, were not included due to legal and ethical considerations. Although multiple groups are criminalized under Section 377, female sex workers in Pakistan are subjected to stronger policies, social stigma, and the absence of secure community-based networks, which create substantial barriers to safe, confidential, and ethical recruitment [16,26]. These factors present significant barriers to safe recruitment and ethical engagement in research contexts involving these groups.

Respondents were recruited using respondent-driven sampling (RDS), which is a network-based sampling method commonly used to recruit the hard-to-reach population, and combined with peer-driven recruitment [27]. Initial respondent-driven sampling participants, also called seed participants, were selected upon recommendations of trusted NGOs and CBOs for men who have sex with men and transgender individuals. We selected a sample of 10 seed participants from each group to ensure diversity. Each seed who completed the interviewer-administered questionnaire received 5 recruitment coupons to invite their peers. In turn, the subsequent participants were also given 5 coupons to recruit more peers. Each coupon card contained a unique RDS number that allowed us to trace the peer recruitment chain and related study information. Coupon management software tracked, distributed, and redeemed coupons during the study, and a standard numbering system was implemented to monitor the recruiter-recruit relationship. Due to ethical and security concerns, people who inject drugs were recruited through convenience sampling from rehabilitation centers and harm reduction centers operated by trusted NGOs and CBOs. The participants were only recruited if they were stable, conscious, and could provide informed consent. This approach was adopted to ensure privacy, reduce participant risk, and allow for more secure and supportive recruitment and data collection. Data collection was conducted in the private room of the NGO or CBO to ensure participants' trust and privacy. Individuals could choose whether to come during the working hours of the NGO or CBO or set an appointment by phone. Individuals who presented with a valid coupon underwent initial eligibility screening. If eligible, an Android tablet was provided to complete the survey questionnaire, which was hosted on Google Forms in a private room. Participants were allowed to self-identify their gender from a list of inclusive options or choose not to disclose, in line with ethical guidelines for gender identity data collection [28]. Each participant completed the questionnaire in approximately 20 minutes, while study staff waited outside to ensure privacy.

Ethical Considerations

The ethical approval for this study was obtained from the research and ethics committee of Riphah Institute of

XSL•FO
RenderX

Pharmaceutical Sciences, Riphah International University (reference: REC-RIPS/RARE/2025/16); all ethical guidelines for research involving human participants were followed. Before participating, all participants were verbally informed about the purpose of the study, the nature of their involvement, potential risks and benefits, and their rights as research participants. Informed written consent was secured from all participants before data collection. Moreover, all the participants were informed that participation in this study was voluntary and that they could withdraw at any time. All participants were assured that their data would remain confidential, and they were informed that their data would be used for publication purposes. Participants were invited to self-identify their gender using inclusive options, with a choice to opt out, ensuring respect for individual autonomy and alignment with ethical standards for gender identity data collection. All the study procedures adopted complied with the Principles of the Declaration of Helsinki, Good Clinical Practices, and the applicable laws and regulations of research involving human subjects in Pakistan. The SAGER (Sex and Gender Equity in Research) guidelines were followed throughout the study and can be found in Checklist 2. Participants received PKR 500 (US $2.77) for study participation and an additional PKR 200 (US $1.11) for each peer successfully recruited to the study, up to a maximum of 5 peers.

Measures

Overview

After reviewing the literature, the questionnaire for this study was adopted from a validated questionnaire used in a similar mHealth study in Malaysia [27]. After carefully reviewing the questionnaire and discussing it with other team members and the principal investigator, we made some modifications. Some questions were removed from the questionnaire, keeping in mind the cultural and ethical standards of the country. The questionnaire was subsequently translated into Urdu, and HIV experts in Pakistan, along with the principal investigator, evaluated its content validity. The main scale achieved a Cronbach α of 0.75, indicating acceptable reliability for a formative exploratory study. The questionnaire used for the study can be found in Multimedia Appendix 1.

Participant Characteristics

We collected participant characteristics: type (men who have sex with men, transgender individuals, or people who inject drugs), sexual orientation, education, employment status, and ethnicity. We did not take provincial and residential information to ensure confidentiality and privacy.

Knowledge About HIV Prevention

In the first section, participants were asked, “Do you know about HIV care and prevention methods?” Those who responded “Yes” to the previous question were asked to rate their HIV knowledge on a scale from 1 (very low) to 5 (very high).

Access to and Frequency of Use of Communication Technology

Participants were asked if they had their own smartphone and daily internet access, and the number of hours spent on the internet each week. Participants’ use of smartphones for various internet-based activities (including social networking, sending or receiving emails, using websites, searching for health-related information, or using health-related apps) was assessed using a 5-point Likert scale (ranging from 1=“never” to 5=“daily”). The participants were then asked if they had used any health-related mHealth apps previously, and a follow-up on their experience using the mHealth app was assessed using the Likert scale (ranging from 1=“very dissatisfied” to 5=“very satisfied.”

Acceptability of mHealth App

Participants were asked about their willingness to use a clinic-integrated mHealth app for HIV prevention, and their willingness to use 5 mHealth-related features: medication reminders, HIV prevention information, e-consultation with the doctor, mental health information, and engagement in a virtual mental health group. The study also evaluated respondents’ preferred frequency (daily, weekly, or monthly) and mode of mHealth delivery (such as phone calls, SMS text messaging, or mobile apps). Participants were asked, “What feature would you consider most important in a clinic-integrated mHealth app for HIV prevention?,” followed by a question: “On a scale of 1 (not at all feasible) to 5 (extremely feasible), how feasible do you think it is to integrate mobile apps into an HIV prevention program?”

Risky Sexual and Drug-Related Behavior

Participants were asked information about their sexual behavior, including recent (past 6 mo) engagement in anal sex, sexual relationships with multiple partners, consistent condom use, and transactional sexual relationships. Additionally, participants were asked about injectable drug use at any point in their lives. Participants were asked if they had ever tested for HIV and their HIV status. Participants were then asked if they had ever been diagnosed with other sexually transmitted infections (STIs), and if they had ever used PrEP or postexposure prophylaxis.

Data Analysis

We calculated the descriptive statistics for categorical variables, such as frequency and percentage, and mean and SD for the continuous variables. Cross-tabulation and chi-square tests were used to measure the association between the categorical variables. Bivariate and multivariate logistic regression analyses were conducted to identify the correlates of mHealth acceptance. Results were reported as odds ratio (OR) with 95% CI. Analyses were performed in SPSS (Statistical Package for the Social Sciences; version 27; IBM Corp) and RStudio (2025.05.0+496), an integrated development environment for R (version 4.5.0; R Foundation for Statistical Computing).

Results

Participants’ Characteristics

A total of 210 participants were included in the study (mean age of 29.55, SD 4.82 y). Of the 210 participants, 83 (39.5%) were people who inject drugs, 72 (34.3%) were transgender individuals, and 55 (26.2%) were men who have sex with men. Regarding sexual orientation, 84 (40%) identified as heterosexual, while 72 (34.3%) identified as homosexual. Most

XSL•FO
RenderX

respondents were college graduates (n=88, 41.9%), while only a few reported having completed higher education (n=25, 11.9%). Employment status showed that half of the participants were unemployed (n=105, 50%), and a large proportion identified themselves as Muslims (n=159, 75.7%; Table 1).

Table . Demographic characteristics of the participants (N=210).

Demographic characteristics	Value, n (%)
Key population	
Men who have sex with men	55 (26.2)
Transgender individuals	72 (34.3)
People who inject drugs	83 (39.5)
Sexual orientation	
Homosexual	72 (34.3)
Heterosexual	84 (40)
Bisexual	53 (25.2)
Education	
Less than high school	53 (25.2)
High school	88 (41.9)
College or associate	44 (21)
Higher education	25 (11.9)
Employment status	
Unemployed	105 (50)
Part-time employed	67 (31.9)
Full-time employed	32 (15.2)
Ethnicity or race	
Muslim	159 (75.7)
Non-Muslim	51 (24.3)
Knowledge about HIV and its transmission	
Yes	169 (80.5)
No	41 (19.5)
HIV knowledge rating	
Low	88 (52.1)
Moderate	54 (31.9)
High	29 (17.1)

Overall, 169 out of 210 (80.5%) participants reported having some knowledge of HIV and its transmission, while 29 out of 210 (17.1%) demonstrated a high level of HIV-related knowledge. Men who have sex with men reported the highest proportion of strong HIV knowledge (23/55, 41.8%), followed by transgender individuals (18/72, 25%) and people who inject drugs (13/83, 15.6%).

Access to and Frequency of Use of Mobile Technology

Of the 210 participants, 161 (76.7%) had smartphones, and 115 (54.8%) had daily access to the internet. The most common activities that participants used the internet on their smartphones for were online social networking apps like Facebook, WhatsApp, or TikTok; on average, they spent 3.41 (SD 3.10) hours per day on them. A total of 77 out of 210 (36.7%) participants used mobile apps several times a week, and 74 out of 210 (35.2%) used them daily. Additionally, 127 out of 210 (60.5%) of the participants stated that they had used an mHealth app, and 94 out of 210 (44.7%) were satisfied with their experience with the mHealth app. A significant association was found between the education levels and previous mHealth app experiences (χ^2_4=24.08; P<.001). Among those who reported using the app (127/210, 60.5%), 40 out of 127 (31.4%) were transgender individuals, 27 out of 127 (21.2%) were men who have sex with men, and 32 out of 127 (25.1%) had a high school level education, suggesting a positive association between education and mHealth app engagement.

Acceptability of the mHealth App

Of the 210 participants, 182 (86.7%) expressed a willingness to use a clinic-integrated mHealth app for HIV prevention (Table 2). More than half (n=108, 51.5%) indicated interest in a daily

XSL•FO
RenderX

medication reminder feature, while 83 (39.5%) reported no interest in this function. Among those interested in medication reminders, 89 of 130 (68.5%) preferred to receive notifications through the app.

XSL•FO
RenderX

Table . Acceptability of clinic-integrated mobile health app for HIV prevention (N=210).

Variables	Value, n (%)
Clinic-integrated mobile app for HIV prevention	
No	28 (13.3)
Yes	182 (86.7)
Frequency of medication reminder	
Daily	108 (51.4)
Monthly	3 (1.4)
Never	83 (39.5)
Weekly	16 (7.6)
Preferred mechanism of getting reminders	
App notifications	89 (68.5)
Phone calls	17 (13.1)
SMS text messaging	21 (16.2)
Websites	3 (2.3)
Frequency of HIV prevention information	
Daily	21 (10)
Monthly	49 (23.3)
Never	30 (14.3)
Weekly	110 (52.4)
Preferred mechanism of HIV prevention information	
Apps	77 (43)
Phone calls	56 (31.3)
SMS text messaging	39 (21.8)
Websites	7 (3.9)
E-consultation	
No	52 (24.8)
Yes	158 (75.2)
Preferred mechanism of e-consultation	
Apps	75 (48.1)
Phone calls	68 (43.6)
SMS text messaging	8 (5.1)
Websites	5 (3.2)
Mental health information	
No	76 (36.2)
Yes	134 (63.8)
Preferred mechanism for mental health information	
Apps	73 (52.9)
Phone calls	40 (29)
SMS text messaging	19 (13.8)
Websites	6 (4.3)
Most important feature	
Educational resources	85 (40.5)
HIV self-testing information	42 (20.2)

XSL•FO
RenderX

Variables	Value, n (%)
Secure messaging with health care providers	28 (13.5)
Appointment reminders	15 (7.5)

Regarding the dissemination of HIV prevention information, 110 out of 210 (52.4%) respondents preferred receiving updates on a weekly basis, followed by 49 out of 210 (23.3%) who favored monthly updates. Among these, 77 out of 180 (42.7%) indicated a preference for receiving such information via the app. Additionally, two-thirds of the participants (158/210, 75.2%) expressed interest in e-consultation services, with preferences split between app-based consultations (75/156, 48.1%) and phone calls (68/156, 43.6%). Furthermore, 134 out of 210 (63.8%) participants showed interest in receiving mental health information, with 73 out of 138 (52.9%) preferring to access this content through the app. Most participants (85/210, 40.5%) indicated that educational resources was the most important feature, followed by information on HIV self-testing (42/210, 20.2%) and secure messaging with health care providers (28/210, 13.5%). Furthermore, a total of 75 out of 210 (35.7%) participants reported moderate feasibility, and 61 out of 210 (29%) reported high feasibility for an mHealth app for HIV prevention.

There was a positive association between smartphone ownership, daily internet access, and interest in mHealth for HIV prevention; 99.1% (114/115) of them were willing to use an mHealth app for HIV prevention. A chi-square test indicated a statistically significant association between smartphone use, daily internet access, and willingness to use an mHealth app (χ^2_2=71.19; P<.001). The likelihood-ratio chi-square test confirmed this association (χ^2_2=63.35; P<.001). The Cramer V was 0.41, indicating a moderately strong association between internet access, smartphone ownership, and willingness to use a clinic-integrated mHealth app.

HIV Testing, Prevention, and Risk Behaviors

Out of 210 participants, 141 (67.1%) participants had been tested for HIV, but only 111 of 208 (52.9%) respondents knew their current HIV status. HIV prevention services uptake was limited, with 105 (50%) reporting no use; PrEP (47/88, 53.4%) was the most common. Self-reported awareness of STI status was moderate, with 92 out of 210 (43.8%) reporting that they knew their previous STI status. Participants reported a substantial level of vulnerability: 87 out of 210 (41.4%) had experienced violence, 128 out of 210 (61%) had been detained by the police, and 122 out of 210 (58.1%) used injectable drugs. Risky sexual behaviors were frequent, 104 out of 210 (49.5%) engaged in transactional sex, and 101 out of 210 (48.1%) reported multiple partners, while consistent condom use remained low (90/210, 42.9%), though slightly higher among men who have sex with men and transgender individuals (Table 3). Cross-tabulation analysis showed differences in patterns across groups. PrEP users reported higher use of a health-related mobile app. When grouped by sexual orientation, bisexual participants reported the highest levels of risky sexual behavior, followed by homosexual participants, while heterosexual participants reported the lowest levels. Participants who used any HIV prevention service tended to report moderately risky sexual behaviors, whereas those who did not use prevention services were more likely to report the highest levels of risk.

XSL•FO
RenderX

Table . HIV testing, prevention, and risk behavior (N=210).

Variables	Value, n (%)
HIV testing	
Yes	141 (67.1)
No	69 (32.9)
Frequency of HIV testing	
Regularly (every year)	52 (36.4)
Occasionally (every few years)	42 (29.4)
Rarely	49 (34.3)
HIV status	
Negative	111 (52.9)
Do not know	97 (46.1)
HIV prevention services	
Yes	105 (50)
No	105 (50)
Prevention services	
PrEP[a]	47 (53.4)
PEP[b]	3 (3.4)
Event-driven PrEP	20 (22.7)
Safe needle practice	14 (15.9)
Previously diagnosed with STI[c]	
Yes	36 (17.1)
No	56 (26.7)
Do not know	118 (56.2)
PrEP use	
Yes	90 (42.8)
No	120 (57.2)
Violence in the past	
Yes	87 (41.4)
No	123 (58.6)
Detained by the police	
Yes	128 (61)
No	82 (39)
Depressive symptoms	
Yes	129 (61.4)
No	81 (38.6)
Injected drugs	
Yes	122 (58.1)
No	88 (41.9)
Anal sex in past 6 months	
Yes	98 (46.7)
No	112 (53.7)
Multiple sex partners in the past 6 months	

XSL•FO
RenderX

Variables	Value, n (%)
Yes	101 (48.1)
No	109 (51.9)
Transactional sex in the past 6 months	
Yes	104 (49.5)
No	106 (50.5)
Condoms in sexual encounters in the past 6 months	
Yes	90 (42.9)
No	120 (57.1)

[a]PrEP: pre-exposure prophylaxis.

[b]PEP: postexposure prophylaxis.

[c]STI: sexually transmitted infection.

Correlates of mHealth App Acceptance

Part-time workers and students were significantly less likely to express a willingness to adopt a clinic-integrated mHealth app for HIV prevention compared to unemployed participants. Specifically, part-time workers had 70% lower odds (OR 0.30, 95% CI 0.12 - 0.77; *P*=.01), and students had 87% lower odds (OR 0.13, 95% CI 0.02 - 0.98; *P*=.047) of being willing to use the app. In contrast, participants who owned a smartphone were over 4 times more likely to be willing to adopt the mHealth app (OR 4.14, 95% CI 1.41 - 12.21; *P*=.009). Moreover, individuals with prior experience using a mHealth app had nearly 3 times greater odds of willingness to adopt the app (OR 2.97, 95% CI 1.00 - 8.81; *P*=.05) compared to those with no prior mHealth exposure (Table 4).

Table . Correlates of mHealth app acceptance.

Variables	OR[a] (95% CI)	*P* value
Intercept	0.54 (0.25 - 1.21)	.14
Key populations (vs people who inject drugs)		
Men who have sex with men	1.38 (0.43 - 4.49)	.57
Transgender	1.46 (0.58 - 3.69)	.43
Education (vs low education)		
Higher education	2.33 (0.85 - 6.43)	.10
Employment (vs unemployed)		
Full time	0.57 (0.09 - 3.73)	.56
Part time	0.30 (0.12 - 0.77)	.01
Student	0.13 (0.02 - 0.98)	.047
HIV knowledge (vs no)		
Yes	2.05 (0.78 - 5.39)	.15
Smartphone ownership (vs no)		
Yes	4.14 (1.41 - 12.21)	.009
Daily internet access (vs no)		
Yes	1.07 (0.36 - 3.17)	.91
mHealth[b] app experience (vs no)		
Yes	2.97 (1.00 - 8.81)	.05

[a]OR: odds ratio.

[b]mHealth: mobile health.

XSL•FO
RenderX

Discussion

Principal Findings

This study is the first in Pakistan to explore the feasibility and acceptability of mHealth interventions for HIV prevention among key populations, including men who have sex with men, transgender individuals, and people who inject drugs. The findings reveal a significant unmet need for accessible HIV prevention services, evidenced by high-risk behaviors and low engagement with existing methods. However, the widespread use of smartphones and social networking platforms presents a promising avenue for digital health solutions. Participants showed strong interest in a clinic-integrated mHealth app, favoring features like educational content, medication reminders, and e-consultation services.

Our findings identify several key features that prospective users desire in an mHealth intervention for HIV prevention. Specifically, participants expressed strong interest in receiving daily medication reminders, weekly prevention information, access to e-consultation with health care providers, and mental health resources. These preferences align with evidence from India and Nepal, where men who have sex with men and transgender individuals similarly valued medication and appointment reminders, HIV prevention information, and teleconsultation services. In the Nepali study, men who have sex with men also emphasized additional needs such as screening and monitoring illicit drug use, chemsex behavior, and ordering HIV-related supplies, reflecting broader expectations for comprehensive digital support. However, one notable difference is that most participants in Nepal preferred monthly messages and reminders, which contradicts our finding of strong interest in daily reminders and weekly prevention content. Preferences reported in India were more consistent with our results, as participants supported daily medication reminders and weekly HIV prevention information [1,29,30]. In our study, consistent condom use in the past 6 months was reported by 63.6% of men who have sex with men and 62.5% of transgender participants. These rates are considerably higher than those reported in Malaysia, where only 19% of men who have sex with men reported condom use, and moderately higher than in Bangladesh, where 48.6% of men who have sex with men reported consistent use [27,31]. However, these findings were comparable with those of Indonesia, where 76.97% reported protected sex in the past 6 months [32]. Limited access to HIV prevention information, weak social and health care systems, and criminalization of same-sex behaviors and cultural taboos in Muslim-majority settings such as Malaysia, Indonesia, and Bangladesh intensify privacy concerns and restrict access to condoms and prevention strategies. These constraints likely contribute to inconsistent condom use and reduced engagement with HIV prevention programs in these contexts [31,33,34]. mHealth programs have shown promise in improving condom use and mental health. In one example, a self-paced web-based intervention for Chinese men who have sex with men dating app users significantly reduced condomless anal sex and increased participants' use of self-efficacy and positive attitudes towards condoms [35]. More broadly, digital interventions in the settings of low- and middle-income countries have generally been found to enhance sexual and reproductive health knowledge and attitudes, which support safer behavior [36]. Our study found that participants who reported using HIV prevention services such as PrEP were often involved in moderately high-risk activities, and this finding aligns with existing research showing that PrEP use can sometimes be associated with an increase in risky sexual behaviors and the incidence of bacterial STIs [37,38]. This may reflect risk compensation or that individuals with higher baseline risk are more likely to seek PrEP. However, this association is not universal; studies in South-East Asia showed that PrEP use does not always translate into risk compensation, and educational awareness may lead to less involvement in high-risk behavior [39,40]. These findings suggest that integrating HIV prevention, reproductive health information, and risk-reduction counseling into mHealth platforms may strengthen consistent condom use and reduce risky sexual practices.

Also, high rates of violence (41.4%) and police detention (61%) were observed among men who have sex with men and people who inject drugs. Research in the Middle East and North Africa, including countries such as Pakistan, Iran, Lebanon, and Ghana, highlighted a high level of violence, stigma, and discrimination among sexual and gender minorities, with police often implicated in discriminatory practices [41]. These legal and social pressures create significant barriers to accessing care, a finding that is consistent with qualitative research on the lived experiences of people with HIV in Pakistan who face pervasive stigma and structural challenges [13]. Studies from Malaysia and Iran reported that high stigma and criminalization towards men who have sex with men and transgender individuals increased the privacy concerns and limited the open health communication and emphasized mHealth as a discreet option [6,42]. Similarly, a cross-sectional survey conducted among Malaysian men who have sex with men reported that although 74.9% of participants expressed willingness to use a hypothetical HIV prevention app, concerns related to privacy and confidentiality significantly limited acceptability [43]. A Malaysian focus group study among men who have sex with men emphasized that encrypted messages, discreet notifications, and neutral app design are crucial to avoid accidental disclosure of HIV-related content [44].

HIV testing is the first step in reaching those at increased risk of HIV acquisition. Our study found that 63.7% of respondents rarely undergo HIV testing, highlighting the need for frequent screening. This is especially critical in Pakistan, where documented HIV outbreaks have been linked to gaps in testing and prevention among key populations [14]. mHealth could play a vital role in assessing risks and reminding and potentially motivating the marginalized community to be tested regularly through interactive quizzes and tailored push notifications [45,46]. A randomized control trial in Vietnam showed that mHealth improved knowledge and PrEP adherence through interactive quizzes and tailored push notifications [47]. As such, innovative technology can accelerate engagement in HIV prevention strategies and facilitate the PrEP uptake and HIV self-testing, especially in areas where these services are underused [48]. Our results highlight that there is considerable interest in specific mHealth strategies, such as receiving

XSL•FO
RenderX

information related to HIV prevention, e-consultation, medication reminders, and mental health information. Given the lack of public dialogue and stigma, it is not surprising that the Pakistani marginalized population, like their counterparts [49,50], particularly China [51], Vietnam [52], and Malaysia [27], expressed interest. Our study reported low to moderate HIV knowledge, and most of the participants were unaware of their HIV status and STIs, which is confirmed by another study that reported a lack of awareness and low literacy rate among people living with HIV [53], highlighting the need for effective interventions to boost awareness and education. Several studies have reported that mHealth interventions can be helpful in enhancing HIV awareness and education, such as studies from India, Nepal, and Malaysia, demonstrated greater improvement in HIV knowledge awareness, HIV testing, PrEP uptake, and behavioral outcomes among marginalized communities [18,54,55]. Similarly, a cohort study in Indonesia showed that a mobile app increased HIV knowledge from 20% to 60% among men who have sex with men, from 22% to 57% among transgender individuals, and from 49% to 74% among people who inject drugs. Additionally, HIV testing uptake rose by 31% among men who have sex with men, 49% among transgender individuals, and 26% among people who inject drugs [50]. Our study found high feasibility for a clinic-integrated mHealth app that aligns with a systematic review that found that mHealth was both feasible and acceptable in reducing HIV incidence and enhancing care engagement in gay, bisexual, and other men who have sex with men communities [56]. 60.5% of the respondents reported that they had used clinic-integrated mHealth, and some of them had experienced using the smartphone for searching health-related information, which reflects the digital revolution that is especially explosive in Asia, and the fast-paced growth of technology and use within the community [27]. However, it is crucial to contextualize these findings regarding people who inject drugs, as our sample was recruited from rehabilitation centers. Among people who inject drugs in rehabilitation and treatment programs, mHealth interventions show high feasibility and willingness, with around 97% owning mobile phones and strong engagement in text-based HIV prevention and adherence support [57]. In contrast, people who inject drugs outside treatment face major barriers, including low phone ownership, unstable housing, device loss or sale, and limited literacy, reducing mHealth uptake [58]. For example, in Tanzania, only 32% of people who inject drugs owned phones, and just 6.4% had smartphones [59]. Evidence from Malaysia similarly shows higher mobile phone and retention among people who inject drugs in harm reduction programs compared to those outside care [60]. These findings indicate that mHealth works well for stabilized people who inject drugs in treatment but requires tailored strategies to reach broader community populations. Rapid advancements in mobile technology and app development are creating new opportunities to incorporate mHealth into existing HIV prevention services in the region. A mHealth-based app could serve as an additional tool to assist marginalized communities with HIV prevention or care needs between clinic visits, guiding them to needed services, potentially enhancing clinical care, and supporting them through screening and recommendations, and providing different modes of accessing information and prevention commodities. However, the high willingness to use reported in our study may not always translate into sustained use. Studies showed that real-world factors such as the affordability of mobile data, variable digital literacy, especially among marginalized populations, strongly impact users' ability to navigate and benefit from these apps, with digital health literacy often requiring skills beyond general health literacy [61,62]. Mistrust of technology-based health services, rooted in historical health care disparities and concerns about privacy, further reduces long-term acceptability and use [63]. Addressing these barriers through community health worker support, culturally sensitive design, and policies promoting digital literacy and affordable connectivity is critical to translating initial interest into sustained mHealth app use.

Limitations

Despite valuable insights, this study has several limitations. First, the sample size was limited to 210 participants due to resource constraints and challenges in accessing the hard-to-reach population. A larger, more diverse sample, especially including men who have sex with men and people who inject drugs from underdeveloped areas, could improve generalizability. Moreover, interpretation of the regression analysis should be approached with caution, as several predictors, including students, HIV knowledge, and prior mHealth experience, showed wide confidence intervals likely due to small subgroup sizes. Second, key populations like sex workers and prisoners were excluded due to legal and ethical constraints, notably the criminalization of sex work under Islamic and national laws. Third, reliance on self-reported data for HIV or STI status, sexual behavior, and drug use may introduce social desirability bias, especially given the criminalization of same-sex behavior and drug abuse under Section 377 [16] and the Narcotic Control Act [64]. While self-administered, anonymous questionnaires helped reduce this bias, future studies could ensure privacy using branching logic and adaptive questioning. Fourth, because RDS weights were not applied, estimates for men who have sex with men and transgender participants reflect sample characteristics rather than population-level estimates. This may limit generalizability and affect direct comparability with the convenience sample of people who inject drugs. Furthermore, recruiting people who inject drugs via convenience sampling from rehabilitation centers may overestimate mHealth feasibility compared to the broader, nontreated populations like people who inject drugs, who likely face greater structural and digital access barriers. Findings should therefore be interpreted with these sampling differences in mind. Finally, while participants expressed a high willingness to use mHealth, stated interest may not always translate to actual usage. Future implementation studies are needed to assess real-world uptake. Although this study focused on HIV-negative individuals, mHealth also shows potential for people living with HIV, particularly in supporting ART adherence and care management [6,26,65]. Future qualitative studies involving both people living with HIV and health care providers are recommended.

XSL•FO
RenderX

Practical Implications

Overview

This study offers important implications for public health programs and digital health policy in Pakistan. High smartphone ownership (76.7%) and internet access (54.8%) among the key population support the early development of the pilot mHealth intervention, with potential funding support from international donors. However, high willingness to use may not always lead to actual use, and long-term engagement may be affected by mobile data costs, varying digital literacy, and mistrust of technology-based health services [61,62]. Digital platforms offer scalable, resilient health care delivery, crucial during disruptions like those faced by HIV-positive individuals during the COVID-19 pandemic [66,67].

Short-Term Goals

A pilot study should be implemented to test the actual feasibility, usability, and retention in use of mHealth interventions. Given widespread digital engagement, tailored interventions such as apps or SMS text messaging–based systems with multilingual options (eg, Urdu and Punjabi) and gamified education can serve as cost-effective, discreet platforms for HIV education, medication reminders, e-consultation, and mental health resources. Considering the high rate of depressive symptoms (61.4%) and injectable drug use (58.1%), Ahmed et al [13] measured health-related quality of life among people living with HIV and also reported that 63.1% of participants were highly depressed or anxious. Apps should also ensure anonymity, encrypted interfaces, and offline accessibility, especially in a legally restrictive environment where same-sex behavior and drug use are criminalized. Low rates of HIV testing and limited knowledge (16.9%) highlight the need for mHealth tools to promote self-testing, PrEP education, and risk assessment. Push notifications and app-based tools may encourage testing and PrEP uptake, especially among high-risk groups.

Long-Term Goals

After the pilot stage and successful implementation of mHealth among the key population, the focus must shift toward integrating people living with HIV and other correlated programs, such as tuberculosis. This requires training health care providers to integrate these tools into ART centers and national prevention strategies. Retention in care remains a major challenge in Pakistan, especially among mobile communities who are at high risk of being lost to follow-up [68]. The app should be connected with the national database, and services like ART medication delivery through an anonymous way should be incorporated to facilitate these groups and engage them in care. To address structural barriers like violence (41.4%) and police detention (61.6%), mHealth interventions should be supported by community-based organizations that can offer legal and economic assistance. Such an app could incorporate a one-touch emergency support button connecting users to a trusted CBO focal person, a brief and low-literacy legal rights guide in Urdu, a directory of verified safe spaces such as partner clinics or drop-in centers, and secure in-app messaging for confidential safety or rights-related advice [69,70].

Conclusions

This study aimed to assess the feasibility, willingness, and preferences for mHealth interventions among men who have sex with men, transgender individuals, and people who inject drugs in Pakistan. Our findings demonstrate high acceptability and readiness to engage with a future mHealth solution tailored to their needs. Such platforms, yet to be developed, could overcome key barriers, including stigma, low HIV knowledge, and irregular testing, by offering discreet, accessible, and user-centered services. These should include features like HIV education, medication reminders, e-consultation, and mental health support. Policymakers, in collaboration with CBOs, should develop encrypted, culturally sensitive platforms with e-consultation and mental health support aimed at enhancing reach and mitigating HIV risk. Future research, including qualitative studies, is needed to optimize app design for these populations and people living with HIV.

Acknowledgments

The author would like to thank all the participants who participated in the design, data collection, data analysis, manuscript writing, and completion of this survey. The authors affirmed that the study adhered to SAGER (Sex and Gender Equity in Research) guidelines, and sex and gender considerations were integrated into the study design, data collection, analysis, and reporting. The authors acknowledged the use of the artificial intelligence–powered tools ChatGPT (OpenAI) and Grammarly to refine and enhance the manuscript's readability. After refinement and readability enhancement, 2 reviewers thoroughly reviewed the manuscript to ensure accuracy and scholarly rigor.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article. However, the authors are thankful to the University of California, San Diego, for providing the article processing fee.

Data Availability

Data will be available upon reasonable request by contacting the corresponding author.

XSL•FO
RenderX

Authors' Contributions

All authors contributed to the review and approval of the final manuscript. FR led all aspects of the study, including conceptualizing the research, designing the questionnaire, collecting and analyzing the data, and writing the manuscript. AA contributed to conceptual development and questionnaire design. The first draft of the questionnaire and manuscripts was shared with AA, who provided feedback. FR addressed all comments and revisions before sharing with other authors. The other authors (MWS, MR, FA, MMU, and AS) reviewed the draft version, provided their feedback and recommendations, and FR addressed their comments through revision. The process continued until all authors reached consensus on the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sample questionnaire.
[DOCX File, 46 KB - humanfactors_v13i1e81111_app1.docx]

Checklist 1
STROBE checklist.
[PDF File, 179 KB - humanfactors_v13i1e81111_app2.pdf]

Checklist 2
SAGER checklist.
[DOCX File, 35 KB - humanfactors_v13i1e81111_app3.docx]

References

1. Gautam K, Paudel K, Ahmed A, et al. High interest in the use of mHealth platform for HIV prevention among men who have sex with men in Nepal. J Community Health 2024 Aug;49(4):575-587. [doi: 10.1007/s10900-024-01324-x] [Medline: 38281283]
2. Global HIV & AIDS statistics fact sheet 2023. Joint United Nations Programme on HIV/AIDS. 2023. URL: https://www.unaids.org/en/resources/fact-sheet [accessed 2025-06-12]
3. Executive summary—the path that ends AIDS: UNAIDS global AIDS update 2023. Joint United Nations Programme on HIV/AIDS. 2023. URL: https://www.unaids.org/en/resources/documents/2023/global-aids-update-2023-executive-summary [accessed 2025-05-24]
4. Dangerous inequalities: World AIDS Day report 2022. Joint United Nations Programme on HIV/AIDS. 2022. URL: https://www.unaids.org/en/resources/documents/2022/dangerous-inequalities [accessed 2025-05-17]
5. Lyons CE, Schwartz SR, Murray SM, et al. The role of sex work laws and stigmas in increasing HIV risks among sex workers. Nat Commun 2020 Feb 18;11(1):773. [doi: 10.1038/s41467-020-14593-6] [Medline: 32071298]
6. Krishnan A, Weikum D, Cravero C, Kamarulzaman A, Altice FL. Assessing mobile technology use and mHealth acceptance among HIV-positive men who have sex with men and transgender women in Malaysia. PLoS ONE 2021;16(3):e0248705. [doi: 10.1371/journal.pone.0248705] [Medline: 33755693]
7. Chandler R, Guillaume D, Francis S, et al. "I care about sex, I care about my health": a mixed-methods pre-test of a HIV prevention mobile health app for Black women in the southern United States. PLoS One 2023;18(10):e0289884. [doi: 10.1371/journal.pone.0289884] [Medline: 37851669]
8. Huang H, Xu Z, Ge Q, et al. The impact of customized short message service on high-risk behaviors among MSM in China, a randomized controlled trial study. AIDS Behav 2023 Aug;27(8):2720-2730. [doi: 10.1007/s10461-023-03995-4] [Medline: 36705771]
9. National AIDS control programme (NACP). Common Management Unit. 2025. URL: https://www.cmu.gov.pk/nacp-national-aids-control-programme [accessed 2025-05-17]
10. Banik S, Khan S, Jami H, Sivasubramanian M, Dhakal M, Wilson E. Transforming unequal gender relations in india and beyond. In: Social Determinants of Sexual Health Among Sexual and Gender Diverse People in South Asia: Lessons Learned from India: Springer; 2023:327-352. [doi: 10.1007/978-981-99-4086-8_21]
11. Bhatti MW. Global fund cuts aid to Pakistan despite rising HIV, TB, malaria. Geo News. 2025. URL: https://www.geo.tv/latest/613509-global-fund-cuts-pakistans-health-grants-as [accessed 2025-07-13]
12. National AIDS control programme of Pakistan. Ministry of National Health Services Regulation and Coordination, Government of Pakistan. 2022. URL: https://www.nih.org.pk/public/national-aids-control-programme [accessed 2025-05-17]
13. Ahmed A, Saqlain M, Bashir N, et al. Health-related quality of life and its predictors among adults living with HIV/AIDS and receiving antiretroviral therapy in Pakistan. Qual Life Res 2021 Jun;30(6):1653-1664. [doi: 10.1007/s11136-021-02771-y] [Medline: 33582967]

XSL•FO
RenderX

14. Ahmed A, Hashmi FK, Khan GM. HIV outbreaks in Pakistan. Lancet HIV 2019 Jul;6(7):e418. [doi: 10.1016/S2352-3018(19)30179-1] [Medline: 31204244]
15. Raza HA, Raja MHR, Khakwani MM, Jamil B. Pakistan's HIV high-risk populations: critical appraisal of failure to curtail spread beyond key populations. IJID Reg 2024 Jun;11:100364. [doi: 10.1016/j.ijregi.2024.100364] [Medline: 38660580]
16. Khan FA. Khwaja sira: "transgender" activism and transnationality in Pakistan. In: Wadle SS, editor. South Asia in the World: An Introduction: Routledge; 2014:170-184 URL: https://www.researchgate.net/publication/280533731_Khwaja_Sira_Transgender_Activism_and_Transnationality_in_Pakistan [accessed 2026-02-05]
17. Kemp S. Digital 2025: Pakistan. DataReportal. 2025. URL: https://datareportal.com/reports/digital-2025-pakistan [accessed 2025-05-19]
18. Shrestha R, Altice FL, Khati A, et al. Clinic-integrated smartphone app (JomPrEP) to improve uptake of HIV testing and pre-exposure prophylaxis among men who have sex with men in Malaysia: mixed methods evaluation of usability and acceptability. JMIR Mhealth Uhealth 2023 Feb 16;11:e44468. [doi: 10.2196/44468] [Medline: 36795465]
19. Hayat MO. Sehat Dost: digital solutions to HIV-related discrimination. United Nations Development Programme. 2024. URL: https://www.undp.org/pakistan/stories/sehat-dost-digital-solutions-hiv-related-discrimination [accessed 2025-05-17]
20. Egbuchulem KI. The basics of sample size estimation: an editor's view. Ann Ib Postgrad Med 2023 Jun;21(1):5-10. [Medline: 37528816]
21. Althubaiti A. Sample size determination: a practical guide for health researchers. J Gen Fam Med 2022 Mar;24(2):72-78. [doi: 10.1002/jgf2.600] [Medline: 36909790]
22. Naing L, Nordin RB, Abdul Rahman H, Naing YT. Sample size calculation for prevalence studies using Scalex and ScalaR calculators. BMC Med Res Methodol 2022 Jul 30;22(1):209. [doi: 10.1186/s12874-022-01694-7] [Medline: 35907796]
23. Global health sector strategy on HIV/AIDS 2011–2015. : World Health Organization; 2011 URL: https://iris.who.int/server/api/core/bitstreams/0918c708-f163-4886-8adc-461d67a72fb2/content [accessed 2025-05-18]
24. Differentiated service delivery for key populations: men who have sex with men, sex workers, transgender people, people who inject drugs and prisoners and other people living in closed settings: a background review. Differentiated Service Delivery. 2016. URL: https://www.differentiatedservicedelivery.org/wp-content/uploads/2022/05/DSD-Key-populations-review.pdf [accessed 2025-05-21]
25. Consolidated guidelines on HIV prevention, diagnosis, treatment and care for key populations—2016 update. : World Health Organization; 2016 URL: https://iris.who.int/server/api/core/bitstreams/918a4a9a-5116-49d7-9074-e7bcf295d0b6/content [accessed 2026-02-05]
26. Irfan A. Finding the gaps in stories of sex workers in Pakistan. New Line Magazine. 2023. URL: https://newlinesmag.com/spotlight/finding-the-gaps-in-stories-of-sex-workers-in-pakistan [accessed 2024-06-19]
27. Shrestha R, Maviglia F, Altice FL, et al. Mobile health technology use and the acceptability of an mHealth platform for HIV prevention among men who have sex with men in Malaysia: cross-sectional respondent-driven sampling survey. J Med Internet Res 2022 Jul 25;24(7):e36917. [doi: 10.2196/36917] [Medline: 35877172]
28. Heidari S, Babor TF, De Castro P, Tort S, Curno M. Publisher correction: Sex and Gender Equity in Research: rationale for the SAGER guidelines and recommended use. Res Integr Peer Rev 2024 Dec 20;9(1):15. [doi: 10.1186/s41073-024-00155-7] [Medline: 39707570]
29. Shrestha R, Altice FL, DiDomizio E, Sibilio B, Ranjit YS, Copenhaver MM. Feasibility and acceptability of an mHealth-based approach as an HIV prevention strategy among people who use drugs on pre-exposure prophylaxis. Patient Prefer Adherence 2020;14(null):107-118. [doi: 10.2147/PPA.S236794] [Medline: 32021122]
30. Rawat S, Wilkerson JM, Lawler SM, et al. Recommendations for the development of a mobile HIV prevention intervention for men who have sex with men and Hijras in Mumbai: qualitative study. JMIR Public Health Surveill 2018 May 3;4(2):e46. [doi: 10.2196/publichealth.9088] [Medline: 29724705]
31. Reza MM, Rana AM, Azim T, et al. Changes in condom use among males who have sex with males (MSM): measuring the effect of HIV prevention programme in Dhaka city. PLoS One 2020;15(7):e0236557. [doi: 10.1371/journal.pone.0236557] [Medline: 32706840]
32. Harjana NPA, Lubis DS, Fauk NK, Gesesew HA, Ward PR, Januraga PP. Health information-seeking behaviour, intimate partner violence, and sexual risk-taking among MSM in Bali, Indonesia. J Men's Health 2024;20(6):25. [doi: 10.22514/jomh.2024.087]
33. Abdul Manaf R, Dickson N, Lovell S, Ibrahim F. Consistent condom use and its predictors among female sexual partners of people who inject drugs in Klang Valley, Malaysia. BMC Public Health 2019 Nov 7;19(1):1473. [doi: 10.1186/s12889-019-7855-1] [Medline: 31699061]
34. Fauk NK, Kustanti CY, Liana DS, Indriyawati N, Crutzen R, Mwanri L. Perceptions of determinants of condom use behaviors among male clients of female sex workers in Indonesia: a qualitative inquiry. Am J Mens Health 2018 Jul;12(4):666-675. [doi: 10.1177/1557988318756132] [Medline: 29468942]
35. Choi EPH, Wu C, Choi KWY, et al. Ehealth interactive intervention in promoting safer sex among men who have sex with men. NPJ Digit Med 2024 Nov 9;7(1):313. [doi: 10.1038/s41746-024-01313-3] [Medline: 39516336]

XSL•FO
RenderX

36. Isaacs N, Ntinga X, Keetsi T, et al. Are mHealth interventions effective in improving the uptake of sexual and reproductive health services among adolescents? A scoping review. Int J Environ Res Public Health 2024 Jan 31;21(2):165. [doi: 10.3390/ijerph21020165] [Medline: 38397656]
37. Hart TA, Noor SW, Berlin GW, et al. Pre-exposure prophylaxis and bacterial sexually transmitted infections (STIs) among gay and bisexual men. Sex Transm Infect 2023 May;99(3):167-172. [doi: 10.1136/sextrans-2021-055381] [Medline: 35701145]
38. Yan X, Jia Z, Zhang B. Evaluating the risk compensation of HIV/AIDS prevention measures. Lancet Infect Dis 2022 Apr;22(4):447-448. [doi: 10.1016/S1473-3099(22)00151-7] [Medline: 35338865]
39. Baek YJ, Lee Y, Lee JA, et al. Sexual risk compensation and retention in PrEP care in Korea: an HIV PrEP demonstration study. J Korean Med Sci 2025 Jun 9;40(22):e102. [doi: 10.3346/jkms.2025.40.e102] [Medline: 40491083]
40. To KW, Lee SS. HIV pre-exposure prophylaxis in South East Asia: a focused review on present situation. Int J Infect Dis 2018 Dec;77:113-117. [doi: 10.1016/j.ijid.2018.10.027] [Medline: 30395980]
41. Abboud S, Veldhuis C, Ballout S, Nadeem F, Nyhan K, Hughes T. Sexual and gender minority health in the Middle East and North Africa region: a scoping review. Int J Nurs Stud Adv 2022 Dec;4:100085. [doi: 10.1016/j.ijnsa.2022.100085] [Medline: 38745607]
42. Ameli V, Haberer J, Sabin L, et al. Tailored mHealth intervention for improving treatment adherence for people living with HIV in Iran (HamRaah): protocol for a feasibility study and randomised pilot trial with a nested realist evaluation. BMJ Open 2021 Jun 23;11(6):e042296. [doi: 10.1136/bmjopen-2020-042296] [Medline: 34162631]
43. Shrestha R, Fisher C, Wickersham JA, et al. Privacy and confidentiality concerns related to the use of mHealth apps for HIV prevention efforts among Malaysian men who have sex with men: cross-sectional survey study. JMIR Form Res 2021 Dec 16;5(12):e28311. [doi: 10.2196/28311] [Medline: 34924355]
44. Khati A, Wickersham JA, Rosen AO, et al. Ethical issues in the use of smartphone apps for HIV prevention in Malaysia: focus group study with men who have sex with men. JMIR Form Res 2022 Dec 23;6(12):e42939. [doi: 10.2196/42939] [Medline: 36563046]
45. Sanabria G, Scherr T, Garofalo R, et al. Usability evaluation of the mLab app for improving home HIV testing behaviors in youth at risk of HIV infection. AIDS Educ Prev 2021 Aug;33(4):312-324. [doi: 10.1521/aeap.2021.33.4.312] [Medline: 34370566]
46. Biello KB, Horvitz C, Mullin S, et al. HIV self-testing and STI self-collection via mobile apps: experiences from two pilot randomized controlled trials of young men who have sex with men. Mhealth 2021;7:26. [doi: 10.21037/mhealth-20-70] [Medline: 33898595]
47. Tran BX, Bui TM, Do AL, et al. Efficacy of a mobile phone–based intervention on health behaviors and HIV/AIDS treatment management: randomized controlled trial. J Med Internet Res 2023 Apr 27;25:e43432. [doi: 10.2196/43432] [Medline: 37104001]
48. Biello KB, Hill-Rorie J, Valente PK, et al. Development and evaluation of a mobile app designed to increase HIV testing and pre-exposure prophylaxis use among young men who have sex with men in the United States: open pilot trial. J Med Internet Res 2021 Mar 24;23(3):e25107. [doi: 10.2196/25107] [Medline: 33759792]
49. Lu X, Gao P, Wang X, et al. User preferences for an mHealth approach to support HIV self-testing and linkage to HIV prevention or care services for MSM in China. AIDS Educ Prev 2021 Dec;33(6):534-550. [doi: 10.1521/aeap.2021.33.6.534] [Medline: 34874759]
50. Garg PR, Uppal L, Mehra S, Mehra D. Mobile health app for self-learning on HIV prevention knowledge and services among a young Indonesian key population: cohort study. JMIR Mhealth Uhealth 2020 Sep 8;8(9):e17646. [doi: 10.2196/17646] [Medline: 32896831]
51. Marley G, Fu G, Zhang Y, et al. Willingness of Chinese men who have sex with men to use smartphone-based electronic readers for HIV self-testing: web-based cross-sectional study. J Med Internet Res 2021 Nov 19;23(11):e26480. [doi: 10.2196/26480] [Medline: 34806988]
52. Trang K, Le LX, Brown CA, et al. Feasibility, acceptability, and design of a mobile ecological momentary assessment for high-risk men who have sex with men in Hanoi, Vietnam: qualitative study. JMIR Form Res 2022 Jan 27;6(1):e30360. [doi: 10.2196/30360] [Medline: 35084340]
53. Aizaz M, Abbas FA, Abbas A, Tabassum S, Obeagu EI. Alarming rise in HIV cases in Pakistan: challenges and future recommendations at hand. Health Sci Rep 2023 Aug;6(8):e1450. [doi: 10.1002/hsr2.1450] [Medline: 37520460]
54. Ayer R, Poudel KC, Kikuchi K, Ghimire M, Shibanuma A, Jimba M. Nurse-led mobile phone voice call reminder and on-time antiretroviral pills pick-up in Nepal: a randomized controlled trial. AIDS Behav 2021 Jun;25(6):1923-1934. [doi: 10.1007/s10461-020-03122-7] [Medline: 33389377]
55. Patel VV, Rawat S, Dange A, Lelutiu-Weinberger C, Golub SA. An internet-based, peer-delivered messaging intervention for HIV testing and condom use among men who have sex with men in India (CHALO!): pilot randomized comparative trial. JMIR Public Health Surveill 2020 Apr 16;6(2):e16494. [doi: 10.2196/16494] [Medline: 32297875]
56. Nelson KM, Perry NS, Horvath KJ, Smith LR. A systematic review of mHealth interventions for HIV prevention and treatment among gay, bisexual, and other men who have sex with men. Transl Behav Med 2020 Oct 12;10(5):1211-1220. [doi: 10.1093/tbm/ibaa007] [Medline: 33044531]

XSL•FO
RenderX

57. Chandrasekaran S, Kyaw NTT, Harries AD, et al. Enrolment and retention of people who inject drugs in the Needle & Syringe Exchange Programme in Malaysia. Public Health Action 2017 Jun 21;7(2):155-160. [doi: 10.5588/pha.17.0003] [Medline: 28695090]
58. Pang J, Danaee M, Balasingam Kasinather V, Des Jarlais D, Kamarulzaman A, Mohd Salleh NA. Current drug use patterns and HIV and HCV prevalence among people who inject drugs in suburban areas of Malaysia. J Int AIDS Soc 2025 May;28(5):e26420. [doi: 10.1002/jia2.26420] [Medline: 40285368]
59. Munishi C, Ndumwa HP, Massawe JE, et al. Acceptability of using mobile Health (mHealth) as an intervention tool for people with drug use disorders in Tanga, Tanzania. PLOS Digit Health 2023 Sep;2(9):e0000257. [doi: 10.1371/journal.pdig.0000257] [Medline: 37756266]
60. Kumar D, Hemmige V, Kallen MA, Giordano TP, Arya M. Mobile phones may not bridge the digital divide: a look at mobile phone literacy in an underserved patient population. Cureus 2019 Feb 20;11(2):e4104. [doi: 10.7759/cureus.4104] [Medline: 31057998]
61. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. Int J Cardiol 2019 Oct 1;292:280-282. [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]
62. Kruse C, Betancourt J, Ortiz S, Valdes Luna SM, Bamrah IK, Segovia N. Barriers to the use of mobile health in improving health outcomes in developing countries: systematic review. J Med Internet Res 2019 Oct 9;21(10):e13263. [doi: 10.2196/13263] [Medline: 31593543]
63. Giebel GD, Speckemeier C, Abels C, et al. Problems and barriers related to the use of digital health applications: scoping review. J Med Internet Res 2023 May 12;25:e43808. [doi: 10.2196/43808] [Medline: 37171838]
64. Control of Narcotic Substances Act, 1997 (Act no.XXV of 1997). : Government of Pakistan; 1997 URL: https://www.unodc.org/cld/uploads/res/document/pak/control-of-narcotic-substances-act-xxv_html/CONTROL_OF_NARCOTIC_SUBSTANCES_ACT_pakistan.pdf [accessed 2026-02-21]
65. Escobar-Viera C, Zhou Z, Morano JP, et al. The Florida Mobile Health Adherence Project for people living with HIV (FL-mAPP): longitudinal assessment of feasibility, acceptability, and clinical outcomes. JMIR Mhealth Uhealth 2020 Jan 8;8(1):e14557. [doi: 10.2196/14557] [Medline: 31913127]
66. Raza A, Ullah I, Tahir MJ, Jabbar A, Ahmed A. Coronavirus disease 2019 (COVID-19) is a healthcare dilemma for human immunodeficiency virus (HIV)-positive individuals in Pakistan. Infect Control Hosp Epidemiol 2022 Nov;43(11):1735-1736. [doi: 10.1017/ice.2021.376] [Medline: 34412723]
67. Shah SA, Altaf A, Vermund SH. Challenges of antiretroviral treatment for human immunodeficiency virus infection in Pakistan. J Pak Med Assoc 2021 Mar;71(3):788-790. [Medline: 34057921]
68. Ahmed A, Dujaili JA, Jabeen M, et al. Barriers and enablers for adherence to antiretroviral therapy among people living with HIV/AIDS in the era of COVID-19: a qualitative study from Pakistan. Front Pharmacol 2022;12:807446. [doi: 10.3389/fphar.2021.807446] [Medline: 35153763]
69. Terlikbayeva A, Primbetova S, Gatanaga OS, et al. UMAI-WINGS: Evaluating the effectiveness of implementing mHealth intimate partner violence prevention intervention in reducing intimate partner violence among women from key affected populations in Kazakhstan using a community-based approach. Behav Sci (Basel) 2025 May 9;15(5):641. [doi: 10.3390/bs15050641] [Medline: 40426420]
70. Lalla-Edward ST, Mashabane N, Stewart-Isherwood L, et al. Implementation of an mHealth app to promote engagement during HIV care and viral load suppression in Johannesburg, South Africa (iThemba Life): pilot technical feasibility and acceptability study. JMIR Form Res 2022 Feb 2;6(2):e26033. [doi: 10.2196/26033] [Medline: 35107427]

Abbreviations

ART: antiretroviral therapy
CBO: community-based organization
mHealth: mobile health
NGO: nongovernmental organization
OR: odds ratio
PrEP: pre-exposure prophylaxis
RDS: respondent-driven sampling
SAGER: Sex and Gender Equity Research
SPSS: Statistical Package for the Social Sciences
STI: sexually transmitted infection
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
UNDP: United Nations Development Programme
WHO: World Health Organization

XSL•FO
RenderX

Edited by A Kushniruk; submitted 22.Jul.2025; peer-reviewed by IU Rehman, MD Butt; revised version received 09.Dec.2025; accepted 19.Jan.2026; published 12.Mar.2026.

Please cite as:
Rehman F, Shahid MW, Riaz M, Umair MM, Azhar F, Siddiqui A, Ahmed A
Feasibility and Preferences to Adopt mHealth-Based Interventions for HIV Prevention Among High-Risk Groups: Cross-Sectional Study
JMIR Hum Factors 2026;13:e81111
URL: https://humanfactors.jmir.org/2026/1/e81111
doi:10.2196/81111

© Faizur Rehman, Muhammad Wasay Shahid, Mehran Riaz, Malik Muhammad Umair, Farah Azhar, Adeel Siddiqui, Ali Ahmed. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 12.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluation of a Contactless Sleep Monitoring Device for Sleep Stage Detection at Home in a Healthy Population: Prospective Study in Free-Living Conditions

Marie-Ange Stefanos[1,2,3], MSc; Guillaume de Laboulaye[3], MSc; David Campo[3], PhD; Martin de Gourcuff[3], MSc; Pierre Escourrou[3,4], MD, PhD; Boris Matrot[1], PhD; Anna Sigridur Islind[2], PhD; Pierre Alexis Geoffroy[1,5,6], MD, PhD

[1]Inserm, F-75019, NeuroDiderot, SleepCmd, Université Paris Cité, Paris, France

[2]Department of Computer Science, Reykjavík University, Reykjavik, Iceland

[3]Withings, 2 rue Maurice Hartmann, Issy-les-Moulineaux, France

[4]Centre Interdisciplinaire du Sommeil, Paris, France

[5]Département de Psychiatrie et D'addictologie, Hopital Bichat - Claude Bernard, F-75018, AP-HP, GHU Paris Nord, DMU Neurosciences, Paris, France

[6]GHU Paris - Psychiatry and Neurosciences, Centre ChronoS, Paris, France

Corresponding Author:
Marie-Ange Stefanos, MSc
Inserm, F-75019, NeuroDiderot, SleepCmd, Université Paris Cité, Paris, France

Abstract

Background: Sleep is essential for overall health and well-being, but assessing sleep architecture is often costly and time-consuming, relying primarily on polysomnography (PSG). While wrist-worn wearables offer alternatives, they face limitations regarding user compliance, such as battery charging and physical discomfort. Nearable devices address these burdens, but they regularly lack rigorous validation, especially in real-world settings.

Objective: This study evaluates the accuracy and reliability of the Withings Sleep Analyzer (WSA), a contactless sleep monitoring device, compared to PSG in a home setting using a large and diverse cohort of healthy individuals.

Methods: A total of 117 healthy volunteers (69 women; mean 39.9, SD 11.4 years), prospectively recruited from the general population, underwent home-based PSG and simultaneous WSA recording. The study was conducted under free-living conditions, without constraints on substance intake, prebedtime activity, or forced sleep schedules. The main outcomes were the device's performance in sleep-wake distinction and sleep stage identification using accuracy, kappa, sensitivity, specificity, and the mean absolute error of sleep measures on the entire population and demographic, clinical, and environmental subgroups.

Results: WSA demonstrates high sensitivity (93%, 95% CI 92%-94%) for sleep detection and moderate sensitivity (73%, 95% CI 69%-77%) for wakefulness, achieving an overall accuracy of 87% (95% CI 86%-87%) for sleep-wake distinction. The device showed consistent performance across various demographic subgroups, including different age, BMI, mattress, and sleep arrangements (with or without bed partner) categories. Challenges were noted in accurately classifying specific sleep stages, particularly in distinguishing between light and deep sleep, with a mean accuracy of 63% (95% CI 62%-65%) and a Cohen κ of 0.49 (95% CI 0.47-0.51). The WSA tended to overestimate total sleep time (20 min, 95% CI 10 min to 31 min) and light sleep (1 h 21 min, 95% CI 1 h 8 min to 1 h 36 min) while underestimating rapid eye movement (−15 min, 95% CI −23 min to −8 min) and deep sleep (−46 min, 95% CI −59 min to −34 min) durations. Disagreements between expert reviewers were mirrored in part by the WSA's misclassifications. Participants reported altered perceived sleep quality during the night with the PSG, suggesting discomfort during sleep.

Conclusions: Being contactless and placed under the mattress, the WSA offers a promising approach to long-term sleep monitoring in natural home environments. It shows competitive performance in sleep-wake and sleep stage identification compared to other consumer devices. Progress in wearable and nearable devices is necessary to enhance their accuracy to better support the monitoring of populations with impaired sleep, although limited by an imperfect gold standard. This work also emphasizes the importance of using large, diverse, and challenging datasets, as well as the need for a standardized methodology for accurate sleep stage classification.

(JMIR Hum Factors 2026;13:e77033) doi:10.2196/77033

KEYWORDS

health; informatics; nearable; wearable; sleep stage; home polysomnography; performance assessment; healthy population; sleep monitoring device

XSL•FO
RenderX

Introduction

Sleep is an essential component of human health and plays an important role in daily functioning and overall well-being [1]. Poor sleep quality can significantly impact multiple aspects of life, including physical health, social interactions, and emotional regulation [2,3]. Moreover, poor sleep quality can diminish alertness and cognitive function, increasing the risk of accidents. Furthermore, inadequate sleep has been linked to numerous health issues, such as metabolic disorders, cardiovascular disease, neurocognitive impairments, and mental health challenges [4]. In severe cases, poor sleep quality can lead to critical outcomes, including an increased risk of suicide [5]. Despite the recognized importance of sleep, sleep-related difficulties remain common in the general population [6]. Approximately 30% of adults experience at least one symptom of nocturnal insomnia on a chronic basis [6], over 15% report sleepiness that affects daily activities [7], and more than 10% experience obstructive sleep apnea [8].

Polysomnography (PSG), considered the gold standard for identifying the micro and macro sleep architecture, records a variety of signals (electroencephalogram [EEG], electrooculogram [EOG], electromyogram [EMG], electrocardiogram [ECG], pulse oximetry, air flow, and breathing effort [9]). As the hypnogram helps physicians identify abnormal sleep patterns, quantify sleep fragmentation, and analyze sleep stage distribution, it supports the diagnosis of various sleep disorders, such as sleep-related breathing disorders, narcolepsy, hypersomnia, and disturbing sleep-related behaviors [10].

PSG has limitations. It often involves extensive setup time and significant costs, making it less feasible for routine monitoring [11]. It is scored manually, which is a labor-intensive process that can lead to long waiting times and has been shown to be prone to human error [12]. Additionally, since it is typically conducted over just a few nights, it fails to capture the full picture of an individual's circadian rhythm and night-to-night variability [13,14]. In most cases, PSG is conducted in a sleep lab with numerous sensors that can disrupt natural sleep patterns [15]. For all these reasons, PSG results in data that may not accurately represent a person's typical sleep in their own home [16].

Ambulatory PSG attempts to address these limitations. Lab studies support its accuracy [17-19]. However, in an uncontrolled home setting, sleep metrics may differ; for example, total sleep time (TST) can increase, potentially due to the absence of lab-induced constraints on waking time [20].

The need for more accessible and cost-effective solutions for multinight sleep monitoring remains, driven by the desire to enhance our understanding of sleep quality and facilitate the identification of potential sleep disorders [14,21]. Home-based sleep monitoring devices can enable continuous tracking in the comfort of one's own home, while providing data that reflect real-life conditions more accurately than laboratory-based examinations [22]. This can provide a more realistic picture of an individual's sleep patterns and helps to address the limitations of traditional sleep studies [23].

To address the challenges of cost and complexity in multinight assessments, wearable sleep devices are commonly used to detect sleep and wake states. These devices increasingly combine actigraphy, which tracks body movements [16], with photoplethysmography (PPG) sensors to increase the accuracy of sleep measures estimation [23]. These devices still show two limitations. First, they struggle to accurately differentiate between wakefulness and sleep when individuals remain still while awake [24]. Second, wearable devices require users to maintain proper charging and wearing, leading to frequent data loss and inconsistent performance [25]. In contrast, nearables, which are not in direct contact with the individual, eliminate the need for charging or wearing a device. Nearables for measuring sleep are typically positioned under the mattress or at the bedside and require minimal intervention after setup.

Withings Sleep Analyzer (WSA) is a nearable device designed to provide a noninvasive alternative for monitoring sleep stages over consecutive nights. Despite the proliferation of wearable and nearable sleep-tracking devices, there is a notable lack of rigorous evaluation of the algorithms, leading to concerns about data accuracy and reliability [26,27]. To date, the WSA performance to detect apnea has been evaluated in laboratory conditions [28-32], for which it is European Conformity (Conformité Européenne) marked and Food and Drug Administration–cleared. In addition, it has been assessed for nocturnal sleep periods detection [33] and cardiorespiratory monitoring [34]. However, validation of its sleep stage detection algorithm remains limited. This study, therefore, aims to assess the WSA sleep stage identification algorithm in a real-world home setting in a cohort of 117 healthy adult participants.

There are several important methodological issues addressed in this work. Existing research in sleep-tracking technologies highlights several persistent problems and limitations. Many studies, such as those on the WatchPAT [35] and Somnofy [36] devices, suggest potential utility but often lack comprehensive validation across diverse populations and real-world settings. Evaluations of consumer-grade devices, like those conducted by Chinoy et al [37] and Kainec et al [38], are frequently limited by small sample sizes and strict exclusion criteria (young individuals with controlled caffeine, alcohol, screen time, and sleep regularity), which limit the generalizability of their findings. These studies typically occur in controlled environments, such as sleep labs, which do not accurately reflect regular user conditions at home.

Moreover, the assessment of device accuracy in these studies is often complicated by two distinct factors. First, the absence of synchronized epoch-by-epoch data in some reports limits performance evaluation to night-scale sleep measures, such as TST, precluding the calculation of detailed classification metrics (eg, accuracy, sensitivity, and specificity). Second, the practice of using lights-off and lights-on times to define the temporal analysis window, which excludes all data outside this interval, impairs the reliability of the computed error for metrics like sleep onset latency. This method indeed assumes observable in-bed time is equivalent to the intention to initiate sleep, which cannot be passively identified in free-living conditions [39].

XSL•FO
RenderX

Recent publications emphasize the importance of assessing these devices under everyday conditions, aligning with our work (eg, [40]). Other research emphasizes the need for standardized evaluation protocols and methodological consistency across studies, reinforcing the approach taken in our research [41-43]. These studies collectively highlight the critical need for standardized and rigorous evaluation methods in sleep monitoring technology. Benchmarking efforts of devices are also being addressed, as recent studies manifest [37,38,44-47].

Our study contributes to the growing literature on sleep tracking by assessing the WSA's performance in a home setting, on a large and diverse sample of healthy individuals, in free-living conditions, that is, without constraints about substance intake, activity before bed, or forced sleeping period, providing insights into the device's accuracy and reliability in real-world conditions. The primary goal is to evaluate WSA's performance using classification metrics and mean absolute error (MAE) on standard sleep metrics [48]. Additionally, we compare performance across subgroups based on demographic and clinical characteristics, such as sex, age, BMI, bed partner presence, mattress type, but also Pittsburgh Sleep Quality Index (PSQI) score [49], mattress, and thickness.

By addressing these gaps, our study enhances the understanding of WSA's capabilities and contributes to the development of more reliable and user-friendly sleep-tracking devices.

Methods

Ethical Considerations

Data were collected during a clinical study approved by the French Institutional Review Board (approval number: 2018-A03129-46) and conducted by Withings from 2020 to 2022. The inclusion process for the study was conducted prospectively by a third party, a general-purpose survey company owning listings of self-declared individuals from the general population. Participants were contacted based on the inclusion and exclusion criteria. In particular, the contacted persons had to have no preexisting medical conditions that would necessitate a PSG examination, and inclusion criteria aimed to ensure a balanced distribution of BMI within each age group. The cohort consisted of a consecutive series of 194 volunteers who provided informed consent. To guarantee data privacy, a unique, randomized anonymization code was assigned to each participant (used for both WSA and PSG data correspondence). Withings only had access to these pseudonymized datasets and never had access to direct identifiers. No compensation from Withings was provided to the participants.

Dataset

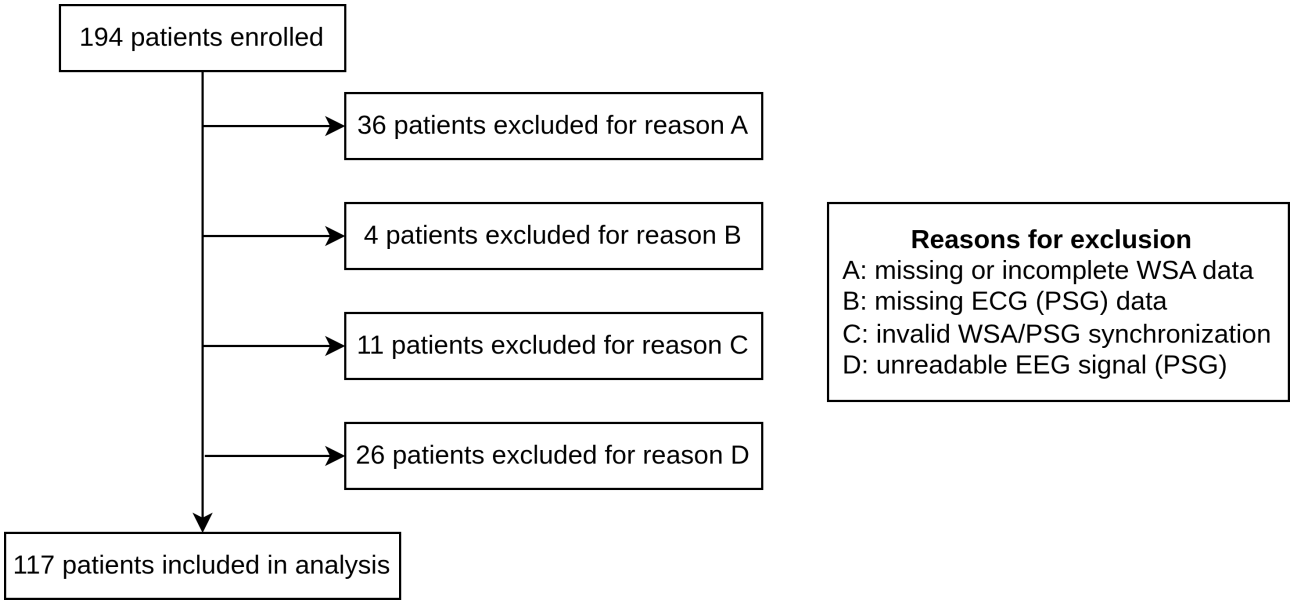
Participants

A total of 194 participants slept for the first night at home with WSA and the CID-LXe portable PSG (CIDELEC), both installed by a technician the evening before the recording. The night after that, the participants slept at home with the WSA, without the PSG. Each participant was asked to answer once to the PSQI questionnaire upon enrollment. After each night, participants were asked to complete a survey addressing the following: their sleep quality for that night, a comparison of that night's sleep quality to their usual sleep, whether they took sleeping pills, and any sleep disturbances related to the sensors.

As illustrated in the flowchart (Figure 1), several sources of data loss have led to the inclusion of 117 participants in this analysis. This trial was notably affected by data collection issues due to a dedicated firmware meant to collect high-frequency sensor data (reason for exclusion A). Unlike the commercial version which transmits aggregated data, the devices in this study were programmed to upload raw, high-frequency sensor signals, which introduced connectivity issues detailed in the Limitations section. We performed Little's missing completely at random (MCAR) test, which did not reject the null hypothesis of MCAR data due to reasons A, C, and D (P>.05). While a statistically significant result was observed for reason B (P=.04), this subgroup comprised only 4 participants and was therefore unlikely to represent a source of systematic bias (see Section A of Multimedia Appendix 1 for detailed results).

XSL•FO
RenderX

Figure 1. Study inclusions flowchart. ECG: electrocardiogram; EEG: electroencephalogram; PSG: polysomnography; WSA: Withings Sleep Analyzer.



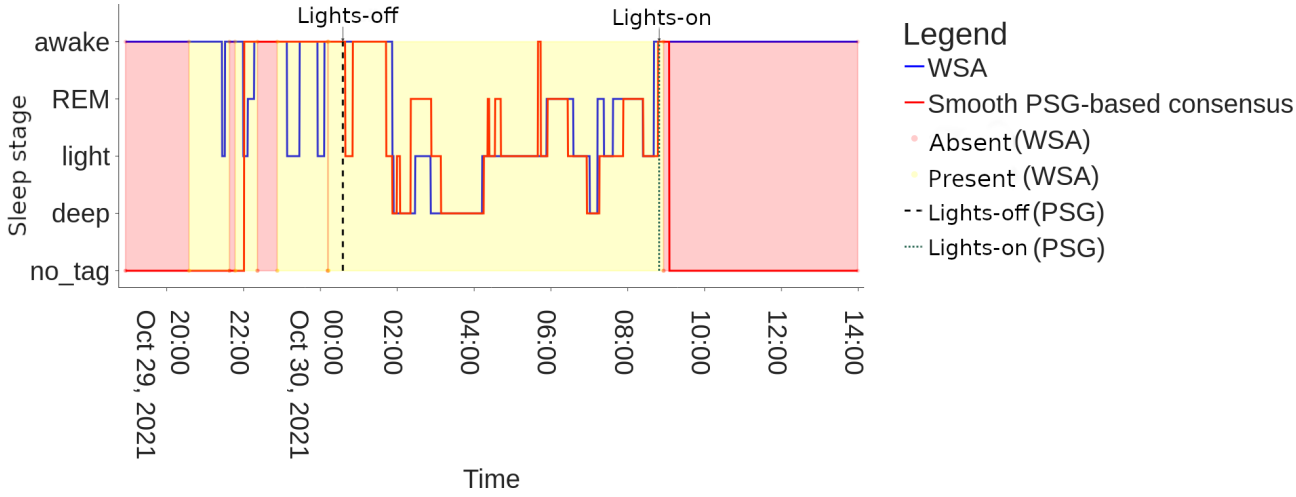
Procedure and Study Design

WSA sampled the raw data from the pressure sensors at 250 Hz and preprocessed data from the microphone at 8 Hz. The reference device was CID-LXe PSG (CIDELEC) that recorded the following signals: 3 EEG, 1 ECG, 2 EOG, 1 submental EMG, and ambient light. Based on experience from past investigations, the digital oximeter and the nasal cannula are among the most sleep-disruptive sensors during PSG, even in a home setting. To minimize this sleep disruption and obtain the most natural possible sleep, PSG was therefore done at home with only EEG, EMG, ECG, and ambient light sensors. As sleep stage annotations from expert reviewers are based on EEG and light sensors only [50], the gold standard reference quality is not affected by the removal of the oximeter and the nasal cannula.

Data Processing

WSA is designed to identify sleep stages over 6-minute windows. As a result, this analysis focuses on evaluating WSA's performance in detecting stages longer than 6 minutes. It should be kept in mind that sleep stages shorter than this are ignored by the device in this study. In addition, WSA does not distinguish stages N1 and N2, so for the aim of comparison, the consensus sleep stages N1 and N2 obtained from the PSG were combined into the "light" category. Rapid eye movement (REM) and awake stages of the WSA and the consensus are directly comparable, and N3 is compared with the stage "deep sleep" from the WSA. The resulting stages are illustrated in Figure 2.

Figure 2. Example of hypnogram (participant 20211029_1) showing WSA sleep stages and the reference (PSG consensus). PSG: polysomnography; REM: rapid eye movement (sleep stage); WSA: Withings Sleep Analyzer.



Two independent certified sleep technologists annotated the PSG signals in 30-second epochs, classifying sleep stages as N1, N2, N3, REM, and awake, following the American Academy of Sleep Medicine (AASM) scoring manual [48]. A third professional reviewed the 2 initial readings and established a consensus for reference data. Sleep stage comparisons were performed over windows of 30 seconds. To evaluate WSA's performance in detecting sleep stages longer than 6 minutes, consensus data were smoothed by selecting for each 6-minute window the stage with the highest occurrence of 30-second epochs. For more details about the procedure, please refer to Section A of Multimedia Appendix 1.

XSL•FO
RenderX

In addition to standard sleep quality measures calculated over the night, such as TST and the duration of each stage, evaluating stage classification required synchronizing epoch-by-epoch data from both devices over time. Synchronization of the clocks of the devices was obtained by maximizing the correlation of the breathing signals of the two devices [51]. To verify the robustness of our breathing-based synchronization, we assessed the impact of temporal shifts (up to ±90 seconds) on classification performance [39,52]. This sensitivity analysis revealed negligible variations, with the SD for sensitivity and specificity across all sleep stages remaining below 0.005. Details of the procedure can be found in Section A of Multimedia Appendix 1.

An example hypnogram is presented in Figure 2, showing the smoothed, PSG-based consensus in red and WSA sleep stages in blue. WSA indicates bed absence/presence, marked respectively by a red and yellow background, information which is absent from the PSG data. The consensus annotations use instead the events lights-on/off detected by the ambient light sensor of the PSG as proxies for bed-in and bed-out events. For further information on challenges arising from using lights-off/on timestamps as proxies for bed presence, please refer to Section A of Multimedia Appendix 1.

Performance Evaluation Metrics and Tools

Standard sleep quality measures used for this analysis and based on the AASM scoring manual [48] are the following: total sleep time, time spent in each sleep stage (TIREM, TILight, TIDeep, and TIWake), proportion of the time spent in each sleep stage compared to total sleep time (PIREM, PILight, PIDeep, and PIWake), and number of episodes in each sleep stage (NEREM, NELight, NEDeep, and NEWake) [50]. Each of these measures has been computed by each device (PSG and WSA) and then compared using the MAE and a Bland-Altman analysis, described in Section B of Multimedia Appendix 1.

Epoch-by-epoch agreement is assessed using standard classification metrics. Subgroup analyses were conducted to determine whether WSA performance is affected by demographic, clinical, and environmental factors. Details regarding the evaluation metrics and subgroup analyses are available in Sections B and C of Multimedia Appendix 1.

To compare the median MAE between different groups (categories of a given metadata as listed in Table S3 in Multimedia Appendix 1), a Kruskal-Wallis test has been performed on the MAE obtained on all the metadata described in Table S3 in Multimedia Appendix 1 for each of the following sleep quality measures (TST, TIREM, TILight, TIDeep, PIREM, PILight, and PIDeep) and for the number of episodes in each sleep stage.

To compare the MAE variance between different groups, a Brown-Forsythe test has been performed on the same data for the same sleep quality measures. The Holm-Bonferroni correction is used after both statistical tests to address the issue of multiple comparisons.

Results

Patient Characteristics

Demographic and clinical data are summarized in Table 1 for the population included in the analyses.

XSL•FO
RenderX

Table . Demographic characteristics and clinical data for the entire population (N=117).

Characteristics	Values
Demographics	
Age (years)	
Mean (SD)	39.9 (11.4)
Min-max	19 - 69
Height (cm)	
Mean (SD)	171.6 (8.3)
Min-max	153 - 194
Weight (kg)	
Mean (SD)	70.4 (12.0)
Min-max	49 - 103
BMI (kg/m²)	
Mean (SD)	23.9 (3.3)
Min-max	17.9 - 30.9
Sex, n (%)	
Male	48 (41)
Female	69 (59)
Skin type[a], n (%)	
Skin type 1	85 (75)
Skin type 2	21 (18)
Skin type 3	8 (7)
Known pathologies, n (%)	
Asthma	5 (4)
Arterial hypertension	3 (<3)
Bipolar disorder	1 (<1)
Ulcerative colitis	1 (<1)
Ankylosing spondylitis	1 (<1)
Pathologies, n (%)	
Cardiovascular[b]	4 (4)
Neurological[c]	0 (0)
Digestive[d]	0 (0)
Metabolic[e]	2 (2)
Psychiatric[f]	0 (0)
Other	6 (5)
PSQI[g] questionnaire [49]	
PSQI score	
Mean (SD)	6.0 (2.9)
Min-max	0 - 15
Component 1: subjective sleep quality	
Mean (SD)	1.2 (0.7)
Min-max	0 - 3
Component 2: sleep latency	

XSL•FO
RenderX

Characteristics	Values
Mean (SD)	1.3 (0.9)
Min-max	0 - 3
Component 3: sleep duration	
Mean (SD)	0.8 (0.8)
Min-max	0 - 2
Component 4: habitual sleep efficiency	
Mean (SD)	0.4 (0.7)
Min-max	0 - 3
Component 5: sleep disturbances	
Mean (SD)	1.3 (0.5)
Min-max	0 - 3
Component 6: use of sleeping medications	
Mean (SD)	0.2 (0.6)
Min-max	0 - 3
Component 7: daytime dysfunction	
Mean (SD)	0.9 (0.8)
Min-max	0 - 3
Medications, n (%)	
Antihypertensive	3 (3)
Antidepressant	2 (2)
Hypnotics	1 (1)
Corticoid	1 (1)
Antiepileptic	1 (1)
Antipsychotic	1 (1)
Mattress type, n (%)	
Foam	41 (35)
Spring	29 (25)
Latex	24 (21)
Memory foam	12 (10)
Other	8 (7)
Unknown	3 (3)
Bed partner, n (%)	
Yes	64 (55)
No	52 (45)
Sensors disturbance (due to PSG[h] device), n (%)	
Yes	46 (39)

XSL•FO
RenderX

Characteristics	Values
No	71 (61)

[a]Skin type is based on the reduced Fitzpatrick scale.

[b]Arterial hypertension, atrial fibrillation, or myocardial infarction.

[c]Parkinson, stroke, adenoma, manic-depression, or epilepsy.

[d]Hemolytic anemia due to glutathione reductase deficiency, ulcer, sleeve, or diverticula.

[e]Diabetes, hypothyroidism, dyslipidemia, or obesity/overweight.

[f]Depression, insomnia, or anxiety.

[g]PSQI: Pittsburgh Sleep Quality Index.

[h]PSG: polysomnography.

A total of 117, mostly healthy, participants were included in the analysis. The cohort comprised 59% (n=69) women and 41% (n=48) men, with a mean age of 39.9 (SD 11.4) years, ranging from 19 to 69 years. The average BMI was 23.9 (SD 3.3) kg/m².

The average sleep quality in this population is considered mildly impaired, with a mean PSQI score of 6 out of 21. The SD of 2.9 highlights significant variability.

The most common mattress types were foam (n=41, 35%), spring (n=29, 25%), and latex (n=24, 21%). Around 55% (n=64) of participants reported having a bed partner. Sensor disturbances were reported by 46 (39%) participants due to the PSG device, while no disturbances were reported regarding the WSA nearable.

Sleep Stages Classification

Figure 3 presents the confusion matrix containing the epochs across all nights, and Table 2 gathers sleep stages classification results on the whole dataset.

Figure 3. Sleep stage classification confusion matrix (N=117). (Left) Raw confusion matrix displaying total epoch counts for WSA sleep stage detection versus PSG ground truth. (Right) Row-normalized confusion matrix (expressed in percentages) illustrating the classification sensitivity and misclassification rates for each true sleep stage. PSG: polysomnography; REM: rapid eye movement (sleep stage); WSA: Withings Sleep Analyzer.

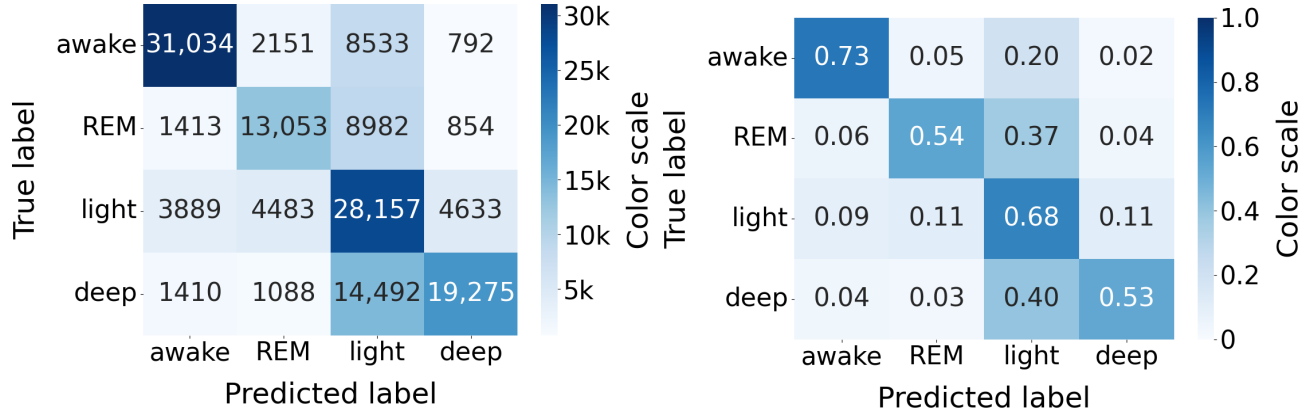


Table . Summary of sleep stage classification results over nights (N=117) (computed as the averages of all nights).

Mean of results per night	Sleep stages								Sleep-wake			
	Kappa	Accuracy	Se[a] REM[b]	Se light	Se deep	Sp[c] REM	Sp light	Sp deep	Kappa	Accuracy	Se sleep	Sp sleep
Mean	0.49	0.63	0.53	0.68	0.56	0.93	0.69	0.94	0.65	0.87	0.93	0.73
95% CI	0.47-0.51	0.62-0.65	0.48-0.57	0.65-0.71	0.51-0.60	0.92-0.94	0.67-0.71	0.93-0.95	0.62-0.69	0.86-0.89	0.92-0.94	0.69-0.77
SD	0.12	0.09	0.24	0.16	0.25	0.05	0.11	0.06	0.19	0.07	0.06	0.22

[a]Se: sensitivity.

[b]REM: rapid eye movement (sleep stage).

[c]Sp: specificity.

Sensitivity and Specificity

The WSA sensitivity for detecting sleep was 93% (95% CI 92%-94%; N=117), meaning it correctly identifies when a person is asleep 93% of the time. The mean specificity was 73% (95% CI 69%-77%), indicating it accurately detects when a person is awake 73% of the time.

XSL•FO
RenderX

The performance in accurately classifying each individual sleep stage against the other three is more variable, with sensitivities of 53% (95% CI 48%-57%), 68% (95% CI 65%-71%), 56% (95% CI 51%-60%), and 73% (95% CI 69%-77%) for REM, light sleep, deep sleep, and wake stage, respectively. The corresponding specificities were 93% (95% CI 92%-94%), 69% (95% CI 67%-71%), 94% (95% CI 93%-95%), and 93% (95% CI 92%-94%), respectively.

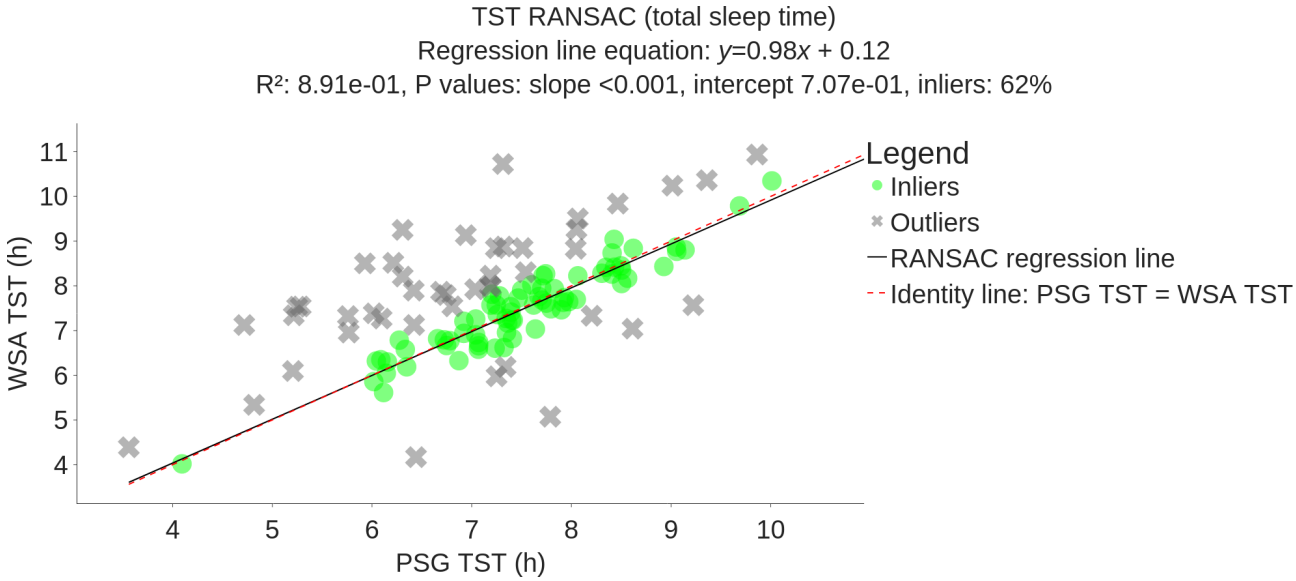
Cohen κ and Accuracy

Please note that, contrary to sensitivity and specificity, Cohen κ and accuracy depend on the relative sizes (prevalences) of each class. In this study's dataset, the WSA mean Cohen κ was 0.65 (95% CI 62%-69%) for distinguishing between sleep and wake stages, and 0.49 (95% CI 47%-51%) for differentiating among the 4 stages (N=117). WSA accuracy to distinguish between sleep and wake was 87% (95% CI 86%-89%). The mean accuracy for the 4 stages was 63% (95% CI 62%-65%).

Assessment of Sleep Quantity Measures

Figure 4 shows the TST estimated by the WSA versus the PSG. The slope of the linear regression is 0.98 (P<.001) and the R^2 value is 0.89, with 38% of outliers, indicating a moderate linear relationship with PSG.

Figure 4. RANSAC regression plot between WSA and PSG total sleep time (black line) compared to the identity line (red dashed line) on 117 participants (R^2=0.89, slope P value P < .001, inliers: 62%). PSG: polysomnography; RANSAC: random sample consensus algorithm; TST: total sleep time; WSA: Withings Sleep Analyzer.



WSA deep sleep was not correlated to time spent in deep sleep (slope estimate of 0.05, P=.7, Figure S4 in Multimedia Appendix 1). The WSA estimates of light sleep and REM were moderately correlated with PSG values, with respective slopes of the regression lines of 0.4 (P<.001, Figure S5 in Multimedia Appendix 1) and 0.65 (P<.001, Figure S6 in Multimedia Appendix 1).

Precision of the WSA estimates is measured with the MAE. The MAE values of total sleep time (equal to the MAE of wake duration because the sum of TST and wake duration is the recording time), REM duration, light sleep duration, and deep sleep duration are 44 min, 35 min, 1 h 33 min, and 1 h 4 min, respectively, with average PSG values of 7 h 15 min (3 h 03 min for awake duration), 1 h 43 min, 2 h 57 min, and 2 h 35 min, respectively. Corresponding 95% CIs are reported in Table 3. The average number of episodes in each sleep stage and the average number of stage changes can be found in Table 4. The WSA tends to underestimate the number of episodes for each sleep stage and is less sensitive to transitions between epochs.

XSL•FO
RenderX

Table . Summary of Bland-Altman results obtained on sleep-related metrics between WSA[a] and PSG[b] estimations (N=117).

Sleep quality measures	PSG value, mean (SD)	MAE[c] (95% CI)	Bias, WSA – PSG (95% CI)
TST[d]	7 h 15 min (1 h 14 min)	44 min (36 min to 52 min)	20 min (10 min to 31 min)
TIREM[e]	1 h 43 min (45 min)	35 min (30 min to 40 min)	–15 min (–23 min to –8 min)
TILight[f]	2 h 57 min (1 h 12 min)	1 h 32 min (1 h 21 min to 1 h 44 min)	1 h 21 min (1 h 8 min to 1 h 36 min)
TIDeep[g]	2 h 35 min (54 min)	1 h 4 min (55 min to 1 h 13 min)	–46 min (–59 min to –34 min)
TIWake[h]	3 h 3 min (1 h 19 min)	44 min (36 min to 52 min)	–20 min (–31 min to –10 min)
PIREM[i]	17% (8.8%)	5.7% (4.9% to 6.5%)	–2.4% (–3.6% to –1.1%)
PILight[j]	29% (14%)	15% (13% to 17%)	13% (11% to 15%)
PIDeep[k]	25% (11%)	10% (8.8% to 12%)	–7.4% (–9.4% to –5.4%)
PIWake[l]	29% (11%)	7.1% (5.8% to 8.4%)	–3.3% (–5.0% to –1.6%)

[a]WSA: Withings Sleep Analyzer.
[b]PSG: polysomnography.
[c]MAE: mean absolute error.
[d]TST: total sleep time.
[e]TIREM: time spent in rapid eye movement sleep stage.
[f]TILight: time spent in light sleep stage.
[g]TIDeep: time spent in deep sleep stage.
[h]TIWake: time spent in wake stage.
[i]PIREM: proportion of the night in rapid eye movement sleep stage.
[j]PILight: proportion of the night in light sleep stage.
[k]PIDeep: proportion of the night in deep sleep stage.
[l]PIWake: proportion of the night in wake stage.

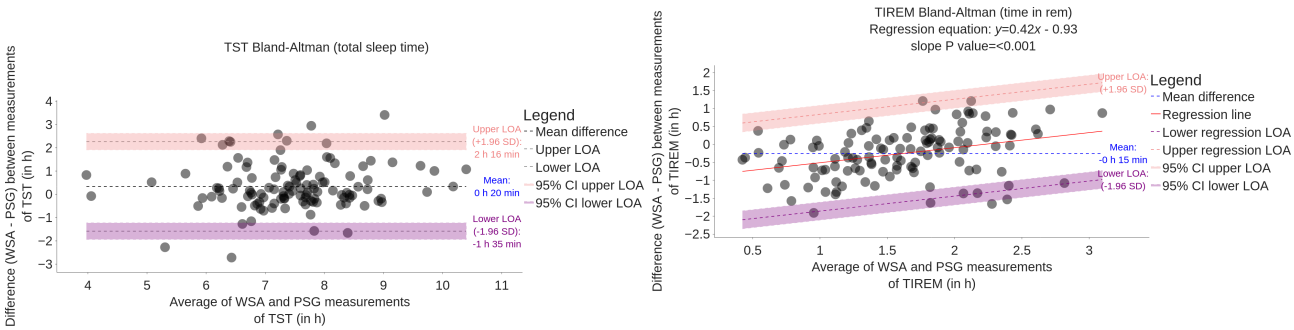
Table . Comparison of the number of sleep stage episodes and stage changes between WSA[a] and PSG[b] (N=117).

Sleep quality measure	Episodes[c] of REM[d]	Episodes[c] of light	Episodes[c] of deep	Episodes[c] of wake	Stage changes
WSA, mean (95% CI)	4.1 (3.8-4.5)	8.9 (8.4-9.3)	4.4 (4.0-4.7)	4.4 (4.1-4.8)	8.9 (8.4-9.3)
PSG, mean (95% CI)	5.0 (4.7-5.3)	13.2 (12.7-13.8)	7.1 (6.7-7.6)	4.9 (4.5-5.4)	13.2 (12.7-13.8)

[a]WSA: Withings Sleep Analyzer.
[b]PSG: polysomnography.
[c]An episode is a sequence of consecutive epochs in the same sleep stage.
[d]REM: rapid eye movement (sleep stage).

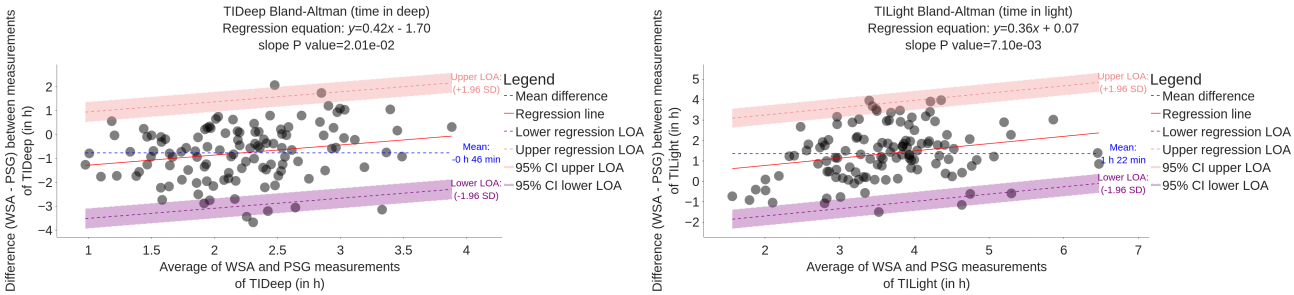
Additional scatter plots and regression analyses of these metrics can be found in Section D of Multimedia Appendix 1. The Bland-Altman plots are presented in Figures 5 and 6. Biases are summarized in Table 3.

Figure 5. Bland-Altman plot comparing WSA and PSG estimation of total sleep time and time in REM on 117 participants. LOA: limits of agreement; PSG: polysomnography; REM: rapid eye movement (sleep stage); TIREM: time spent in rapid eye movement sleep stage; TST: total sleep time; WSA: Withings Sleep Analyzer.



XSL•FO
RenderX

Figure 6. Bland-Altman plot comparing WSA and PSG estimation of time in deep and time in light (N=117). LOA: limits of agreement; PSG: polysomnography; TIDeep: time spent in deep sleep stage; TILight: time spent in light sleep stage; WSA: Withings Sleep Analyzer.



In Figure 5, the Bland-Altman plot for TST does not reveal any proportional bias with the duration of the night. The mean error is 20 minutes. The wide limits of agreement show that there is a substantial amount of variability between PSG and WSA. By contrast, the average error in REM, light, and deep sleep durations depends on the duration of the phase, therefore exhibiting proportional bias. Light sleep is *on average* overestimated for any duration, deep sleep is *on average* underestimated for any duration, and REM is *on average* underestimated for short durations and overestimated for long durations.

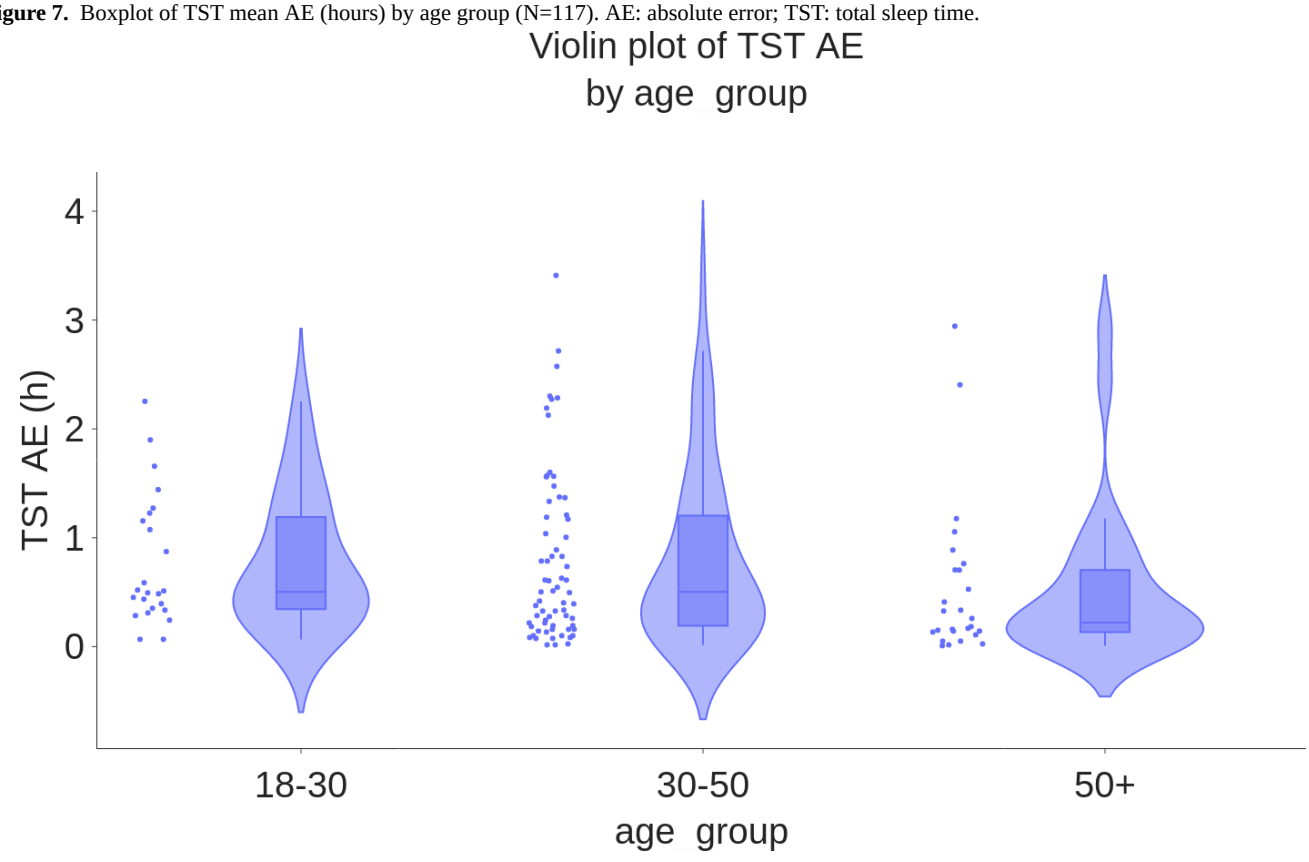
The mean error (bias) in light sleep duration is important, with 1 h 21 min. Biases in REM and deep sleep are −15 min and −46 min, respectively. The mean error in TST is therefore 20 min.

Subgroup Analyses

No significant differences were found in the median or variance of the error across most groups for all the quantitative sleep variables of Table 3. This includes comparisons based on gender, BMI categories, age groups, sleeping arrangements, PSQI scores, self-reported sleep quality, mattress thickness, and night awakenings. However, significant differences in median total sleep time and wakefulness proportion were observed within the PSQI sleep duration component subgroup. Median MAE of TST is respectively 21 min and 1 h 11 min for good and bad sleep duration subgroups, and median MAE of wakefulness proportion is respectively 3% and 12% for good and bad sleep duration subgroups (N=117). This finding suggests that distinguishing sleep from wake is a harder task for the WSA when the participant has subjective short sleep duration (<6 h), highlighting an area for further investigation. For a comprehensive overview of the statistical tests and detailed results, please refer to Multimedia Appendix 2.

As an illustration, Figures 7-9 show the boxplots of TST MAE of WSA compared to PSG, stratified by age groups, BMI ranges, and whether the participant was alone in bed. They suggest that the WSA performance did not depend on these factors. Boxplots for TST MAE for the other subgroups can be found in Section D of Multimedia Appendix 1.

Figure 7. Boxplot of TST mean AE (hours) by age group (N=117). AE: absolute error; TST: total sleep time.



XSL•FO
RenderX

Figure 8. Boxplot of total sleep time mean AE (hours) by BMI group (N=117). AE: absolute error; TST: total sleep time.

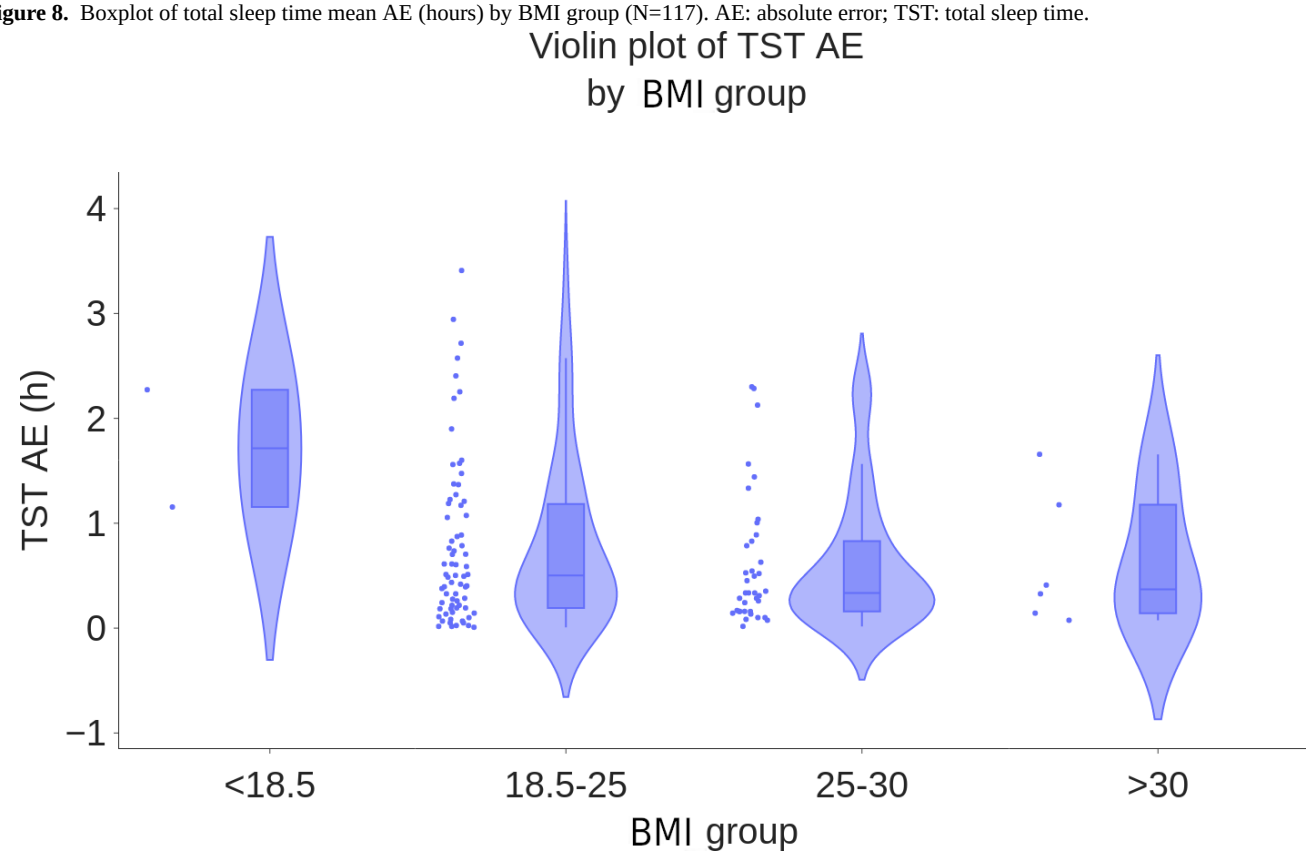


Figure 9. Boxplot of total sleep time mean AE (hours) by *is couple* group (N=117). No: participants who slept alone the night of the recording; yes: participants who had a sleep partner the night of the recording. AE: absolute error; TST: total sleep time.



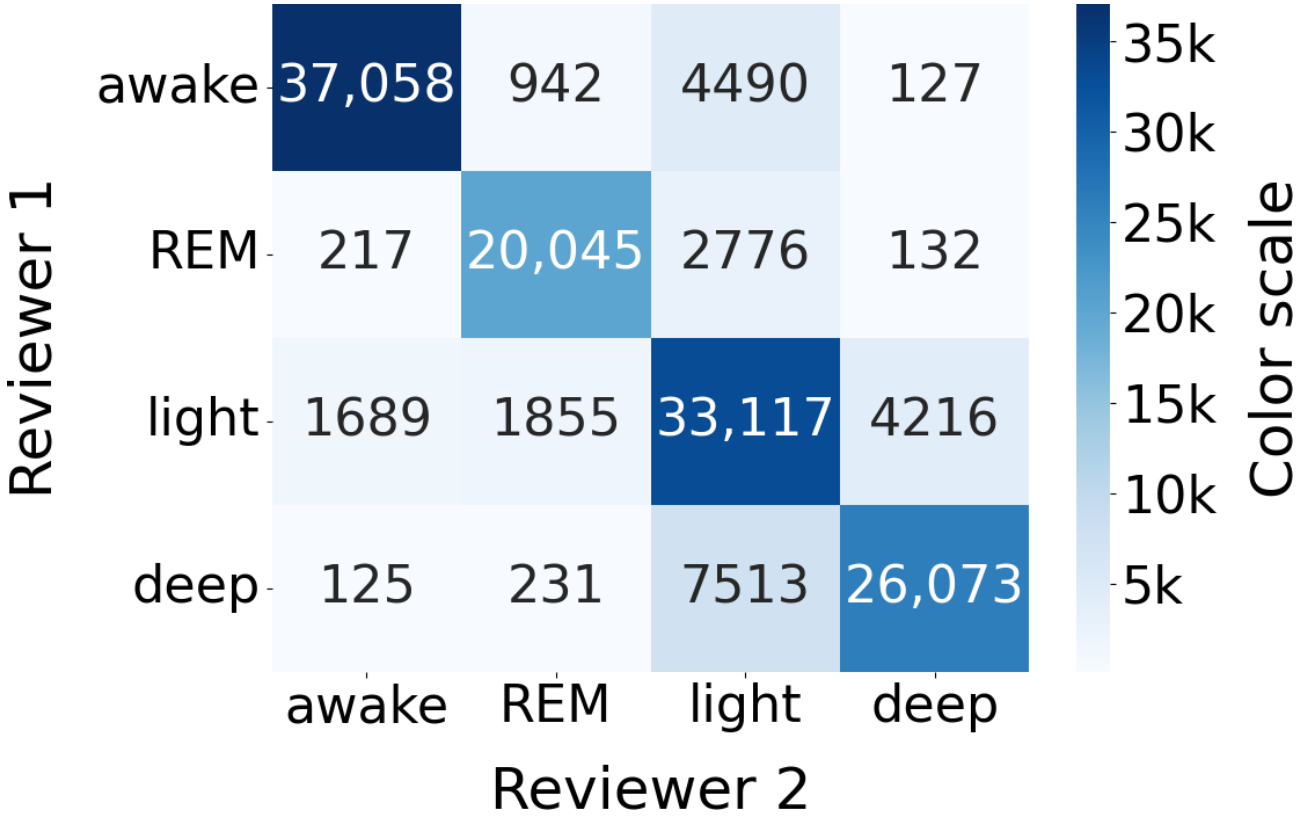
XSL•FO
RenderX

Agreement Between Reviewers

Each 30-second epoch of the PSG data was independently annotated by two reviewers. The mean κ value for sleep stage determination between reviewers before reaching consensus is 0.71 (SD 0.11), closely aligning with literature findings (κ=0.76 in [53]), while it is 0.81 (SD 0.17) for sleep-wake classification.

In the analysis of reviewer agreement on sleep stage classifications, the primary disagreements occur between light and deep stages, as well as between light sleep and awake/REM stages (Figure 10). This indicates that even expert reviewers, considered the gold standard, face challenges in accurately distinguishing these stages due to their inherent similarities and subtle physiological differences. This inherent difficulty in distinguishing light, deep, awake, and REM sleep stages by human experts is reflected in WSA, which was trained on PSG data labeled using a similar protocol. It is likely that the imprecision of the gold standard is passed on to the device.

Figure 10. Sleep stage classification confusion matrix between the first two reviewers' annotations (N=117). REM: rapid eye movement (sleep stage).



Discrepancies in Performance: Epochs With Reviewers' Agreement vs Disagreement

Given the discrepancies encountered in PSG annotation among reviewers, it is pertinent to investigate whether the WSA performed better on epochs where reviewers reached consensus compared to those where disagreements were present. To perform this analysis, the dataset was divided into two groups of epochs: those where reviewers reached agreement and those where disagreements occurred. Separate performance analyses were then conducted on these two complementary groups.

Results are summarized in Table 5. In the "agreement" column (resp. "disagreement" column), only epochs where both reviewers agreed (resp. disagreed) were included, while epochs with disagreement (resp. agreement) were excluded. Figure 11 shows the confusion matrices obtained for both sets.

XSL•FO
RenderX

Table . Summary of performance in reviewers' agreement and disagreement subgroups (N=117).

WSA[a] vs PSG[b]	Agreement	Disagreement	*z* score
Sleep stages			
Epochs, n (%)	113,364 (79)	30,875 (21)	—[c]
Kappa, mean (95% CI)	0.52 (0.50 - 0.55)	0.27 (0.24 - 0.30)	78.06
Accuracy, mean% (95% CI)	0.66 (0.64 - 0.68)	0.53 (0.50 - 0.55)	42.00
Se[d] REM[e], mean (95% CI)	0.55 (0.50 - 0.59)	0.44 (0.39 - 0.49)	34.32
Se light, mean (95% CI)	0.7 (0.67 - 0.72)	0.64 (0.60 - 0.67)	20.16
Se deep, mean (95% CI)	0.56 (0.51 - 0.60)	0.37 (0.32 - 0.43)	68.80
Se wake, mean (95% CI)	0.72 (0.67 - 0.76)	0.51 (0.44 - 0.58)	69.85
Sp[f] REM, mean (95% CI)	0.94 (0.93 - 0.95)	0.92 (0.90 - 0.93)	12.70
Sp light, mean (95% CI)	0.72 (0.69 - 0.74)	0.52 (0.48 - 0.56)	66.63
Sp deep, mean (95% CI)	0.95 (0.94 - 0.96)	0.92 (0.90 - 0.94)	20.25
Sp wake, mean (95% CI)	0.94 (0.93 - 0.95)	0.92 (0.90 - 0.93)	12.70
Sleep-wake			
Epochs, n (%)	133,447 (93)	10,792 (7)	—
Kappa, mean (95% CI)	0.64 (0.59 - 0.68)	0.33 (0.26 - 0.38)	63.72
Accuracy, mean% (95% CI)	0.88 (0.86 - 0.89)	0.76 (0.71 - 0.78)	35.77
Se wake, mean (95% CI)	0.74 (0.67 - 0.76)	0.52 (0.46 - 0.59)	49.15
Se sleep, mean (95% CI)	0.94 (0.93 - 0.95)	0.86 (0.83 - 0.89)	32.20

[a]WSA: Withings Sleep Analyzer.

[b]PSG: polysomnography.

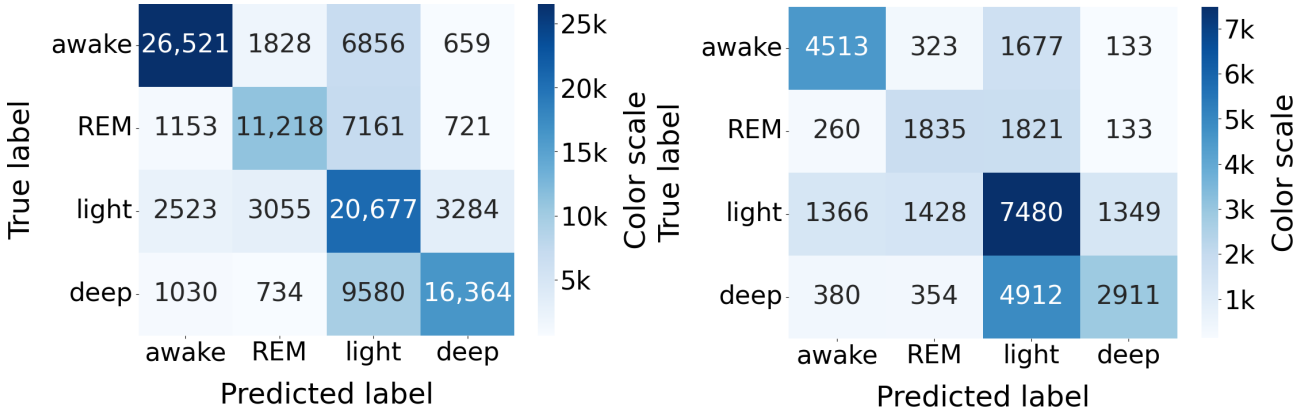
[c]Not applicable.

[d]Se: sensitivity.

[e]REM: rapid eye movement (sleep stage).

[f]Sp: specificity.

Figure 11. Sleep stage classification confusion matrices (N=117) on sets with agreement (left) and disagreement (right); WSA sleep stage detection vs PSG ground truth. PSG: polysomnography; REM: rapid eye movement (sleep stage); WSA: Withings Sleep Analyzer.



We used a *z* test for proportions to test for differences between sensitivity, specificity, accuracy, and kappa. All the results are statistically significant, with WSA performing better on epochs where the two reviewers agreed. The progression is particularly significant for the sensitivity in deep sleep, reaching 0.56, compared to 0.37 when the reviewers disagreed, as in the latter case, WSA predominantly classified deep sleep stages as light sleep. The specificity for light sleep was also starkly improved from 0.52 to 0.72. As a result, the Cohen κ is almost double when reviewers agreed (0.64 vs 0.33 for S/W classification, and 0.52 vs 0.27 for 4-stage classification).

Sleep Quality With vs Without PSG

Since this analysis does not involve the WSA or the PSG data, but only the participants' answers to the questionnaires, it was

XSL•FO
RenderX

conducted on the 186 participants for whom reported sleep quality data were available. Results are given in Table 6.

Table . Comparison of reported sleep quality and sensor disturbance with and without PSG[a] (n=186). In both cases, the participant used a WSA[b] and two ScanWatch devices.

Qualitative sleep quality measures	First night with PSG	Second night without PSG
Sleep quality, n (%)		
Very poor	10 (5.4)	3 (1.6)
Poor	65 (34.9)	32 (17.2)
Quite good	75 (40.3)	71 (38.2)
Good	32 (17.2)	58 (31.2)
Very good	4 (2.2)	22 (11.8)
Sleep quality compared to usual sleep quality, n (%)		
Much worse	28 (15.1)	6 (3.2)
Worse	69 (37.1)	32 (17.2)
Same	77 (41.4)	111 (59.7)
Better	10 (5.4)	31 (16.7)
Much better	2 (1.1)	6 (3.2)
Sensor disturbance		
Device that caused sleep disturbances	PSG	WSA
Yes, n (%)	77 (41.4)	0 (0)
No, n (%)	109 (58.6)	0 (0)

[a]PSG: polysomnography.

[b]WSA: Withings Sleep Analyzer.

Despite the removal of the nasal cannula and the blood oximeter from our home PSG test, the sleep test was associated with significant discomfort in 41% (77/186) of the participants during the night of the PSG home test, compared to 0% (0/186) in the subsequent night with the WSA only. Fifty-two percent (97/186) (resp. 20%, 38/186) of the participants reported a degraded sleep quality on the night of the PSG (resp. without the PSG).

Figure 12 illustrates the changes in sleep quality among participants over 2 consecutive nights. Participants generally experienced better sleep without the PSG sensors. Notable improvements were observed in 52% (97/186) of cases, with transitions from fairly good to good (26/186, 14%), and from poor to fairly good (31/186, 17%), good (12/186, 6%), or very good (10/186, 5%). Additionally, some participants reported significant improvements from very poor to very good sleep, and others from poor to fairly good. These improvements may be influenced by several factors, including the removal of PSG sensors, which can be intrusive and uncomfortable, potentially disrupting sleep. Furthermore, the awareness of being recorded might alter participants' sleep behavior, either positively or negatively. The use of unfamiliar devices could also play a role, as participants may initially experience discomfort or anxiety, which diminishes over time. External conditions such as noise and room temperature, which were not controlled in this study, could also significantly impact sleep quality. Therefore, while the data suggest improvements, it is crucial to consider these potential confounding factors when interpreting the results.

XSL•FO
RenderX

Figure 12. Comparison of reported sleep quality with and without PSG (N=186). Green: improved; yellow: identical; red: deteriorated; black: first night with PSG; grey: second night without PSG. In both cases, the participant used a WSA and two ScanWatch devices. PSG: polysomnography; WSA: Withings Sleep Analyzer.

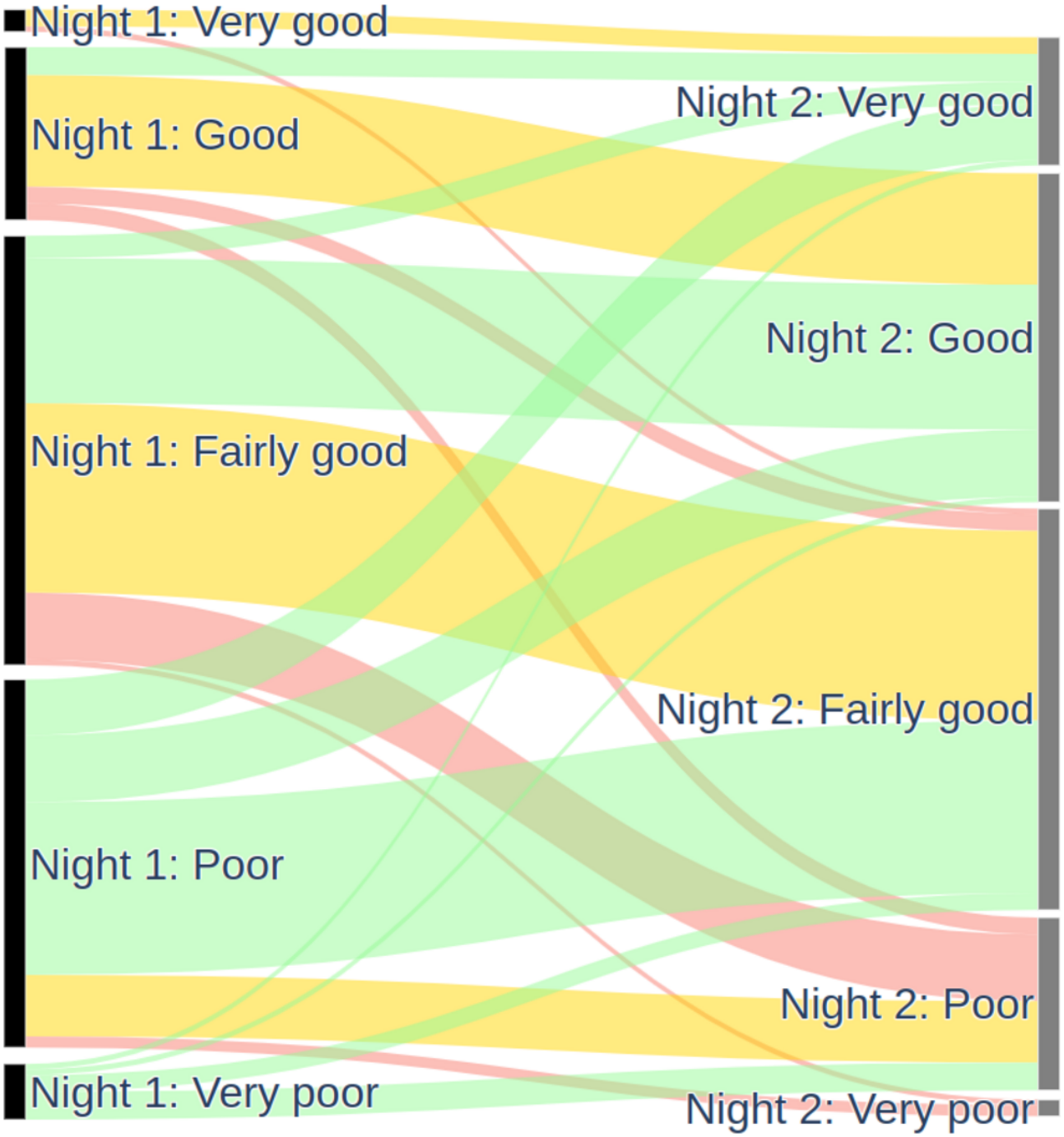


Table 6 highlights that participants slept as well with WSA as they typically do. On the second night, 59.7% (111/186) of participants reported the same sleep quality, 16.7% (31/186) reported better sleep, and 3.2% (6/186) reported much better sleep compared to their usual experience. This suggests that Withings sleep monitoring devices do not negatively impact sleep quality and may even enhance it for some users.

Discussion

Main Findings

The evaluation of the WSA for sleep stage detection in a cohort of 117 participants in a home setting revealed several key insights. The device demonstrated a high sensitivity of 93% for detecting sleep, underscoring its strong capability to accurately identify sleep periods. However, the specificity was moderate at 73%. The duration of the recording before actual sleep onset and after wake-up significantly affects observed performance of wake detection, with shorter periods inflating specificity, accuracy, and Cohen κ. This occurs due to a shortening of quiet wake periods hard to distinguish from sleep (reading or screen time in bed, for instance), leading to fewer false positives while true negatives remain relatively constant. Conversely, overly long (eg, 24-hour) recordings, dominated by easily identifiable diurnal wakefulness, also tend to inflate specificity. In both cases, the performance of the evaluated device is overestimated.

XSL•FO
RenderX

The most clinically relevant and challenging scenario for performance evaluation involves analyzing data encompassing the entire in-bed period, including calm wakefulness before sleep onset and after awakening. While AASM guidelines suggest using lights-off and lights-on as proxies for night boundaries, this approach truncates crucial pre– and post–sleep wake data, therefore introducing a bias in TST and classification performance endpoints. To mitigate this, we used all available PSG annotations, irrespective of lights-off and lights-on times, to provide a more comprehensive and unbiased assessment of performance across the full in-bed duration.

The classification of sleep stages (into wake, REM, light, and deep phases) is more challenging. The confusion matrix (Figure 3) shows that 40% of deep sleep phases, 37% of REM sleep, and 20% of wake phases were misclassified as light sleep by the WSA. As a result, the corresponding sensitivities were rather low (0.56, 0.53, and 0.73, respectively). Thirty-two percent of light sleep stages were misclassified, in similar proportions, into REM, deep sleep, and wake. In consequence, (1) the REM and deep sleep total durations were underestimated (by approximately 30% and 15%), while light sleep duration was overestimated by almost 48% (biases in Table 3), and (2) the specificities for wake, REM, and deep sleep are very high (0.93, 0.94, and 0.94, respectively), while the specificity for light sleep is moderate (0.69).

We showed that these misclassifications of WSA mirror the disagreements between PSG reviewers, who mainly disagree on light versus all the other sleep stages (Figure 11). WSA classification was indeed significantly better on epochs where reviewers agreed (79% of the epochs) than on epochs where they disagreed. WSA's sleep stage classification uses an algorithm trained by supervised learning on PSG data annotated by technician experts, so it is expected that WSA performance is limited by the interreader variability.

Consistent with existing validation studies [33,44], we also report accuracy and Cohen κ. We caution that these performance metrics are strongly dependent on the prevalence of the classes (here, wake, sleep, REM, deep, and light), which creates difficulties in comparing results between studies. In particular, for sleep studies, two factors modify the prevalence of the classes significantly. First, there is no standardized definition of the beginning of the night. Second, the choice of the reference device has some importance, causing more or less sleep disturbances.

With these caveats in mind, the WSA achieved an overall accuracy of 87% for sleep-wake distinction, which is promising for a contactless home monitoring device. Despite this, the accuracy for classifying specific sleep stages was moderate, with REM sleep at 53%, light sleep at 68%, and deep sleep at 56%. The overall accuracy for sleep stage classification was 63%, with a Cohen κ of 0.49.

Notably, the WSA's performance was consistent across various demographic subgroups, including different age and BMI categories. Importantly for a device under the mattress, there was no difference between individuals who slept alone and those with a bed partner, nor between different types or thicknesses of mattresses. Other groups analyzed included sex, PSQI scores, self-reported sleep quality, and whether individuals woke up during the night, all of which also showed no significant differences in MAE, except for wake proportion by subjective sleep duration group. This consistency suggests the device may be useful for a wide range of participants.

Comparison With Previous Research

Differences in Studies' Protocols

To contextualize our findings, we used published validation studies of other consumer sleep trackers. We advise interpreting these comparisons with caution, as study protocols vary significantly (eg, in-lab vs at-home settings, population characteristics, and data handling procedures). We compare our results on WSA with influential studies that have contributed to the understanding of consumer sleep-tracking technology. Studies [29], [37], and [33], respectively, evaluated 7, 5, and 6 consumer devices against PSG, mainly wearable devices, revealing significant variability in their accuracy for sleep stage detection. Studies [33,54-60] compared a total of 8 contact-free devices and 1 wearable against PSG. These studies provide a valuable context for evaluating our device's performance, illustrating both the advancements and challenges in the field of consumer sleep technology. All the data used for this comparison can be found in Multimedia Appendix 3.

We compared the performance metrics reported by other studies on their respective datasets to the performance we obtained on our dataset, without reproducing their results or running their device algorithms on our data. When a result was not explicitly provided in the study report, we calculated it from the confusion matrix if it was available.

Consistent with recent validations [39], our study population encompasses a diverse range of cofactors of sleep quality, presenting a wide range of age and BMI, as was the case in only 2 of the 11 studies used for the comparison [55,59]. Four studies [37,38,44,58] included only young adults, and one study focused on older adults [33].

The conditions of our study are not controlled since participants stayed at home, as in only one other study [54] but on a much smaller scale (n=20). Our dataset is therefore richer regarding important environmental cofactors of sleep quality, including sleeping arrangements (alone or with a partner), mattress types, ambient temperature, lighting and sound, prebedtime activities (eg, no restriction on caffeine or alcohol consumption, on screen time, no forced sleep regularity), and activities during time in bed (eg, no forced turning off of the lights).

In addition, some studies use strict exclusion criteria also after data collection [38] to ensure data quality and reliability, such as excluding individual nights with a total recording time of less than 7 hours and 50 minutes [37].

Thus, our real-world home setting captures the natural variability in sleep patterns and environmental factors. While this enhances the generalizability of our results, reflecting typical user experiences, it also presents a less favorable scenario for the device, posing additional challenges for accurate sleep stage detection and analysis (for instance, activities in bed like reading or screen time can be confounded by the device as sleep).

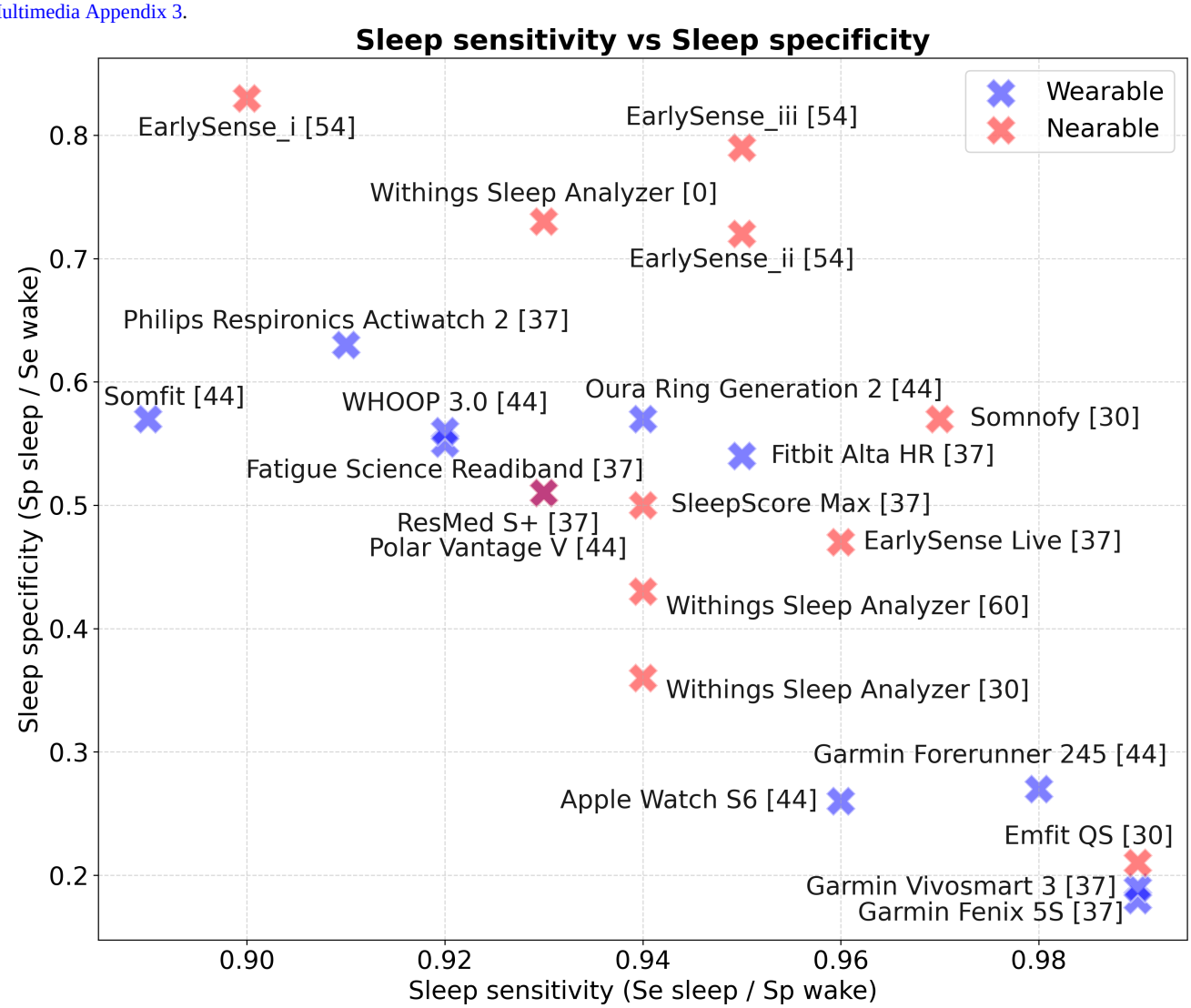
XSL•FO
RenderX

Consequently, our findings provide a more realistic evaluation of the device's performance in everyday conditions.

Sleep-Wake Classification

The choice of a sleep monitoring device requires being informed about the trade-off between sleep and wake identification. This is illustrated in Figure 13. Devices with high sleep sensitivity (≥0.98) exhibit low specificity (<0.30), as observed with the Garmin Forerunner 245, Emfit QS [33], Garmin Vivosmart 3, and Garmin Fenix S5. Conversely, devices achieving higher specificity (≥0.70), such as the EarlySense [54] and WSA, tend to show slightly lower sensitivity (0.90 - 0.95).

Figure 13. Overview of sleep-wake reported sensitivities from varying validation studies. Indication [0] denotes this study. Note: Direct comparison is limited by differences in study protocols and populations [33,37,44,54,60]. The exact data points corresponding to the figure are fully detailed in Multimedia Appendix 3.



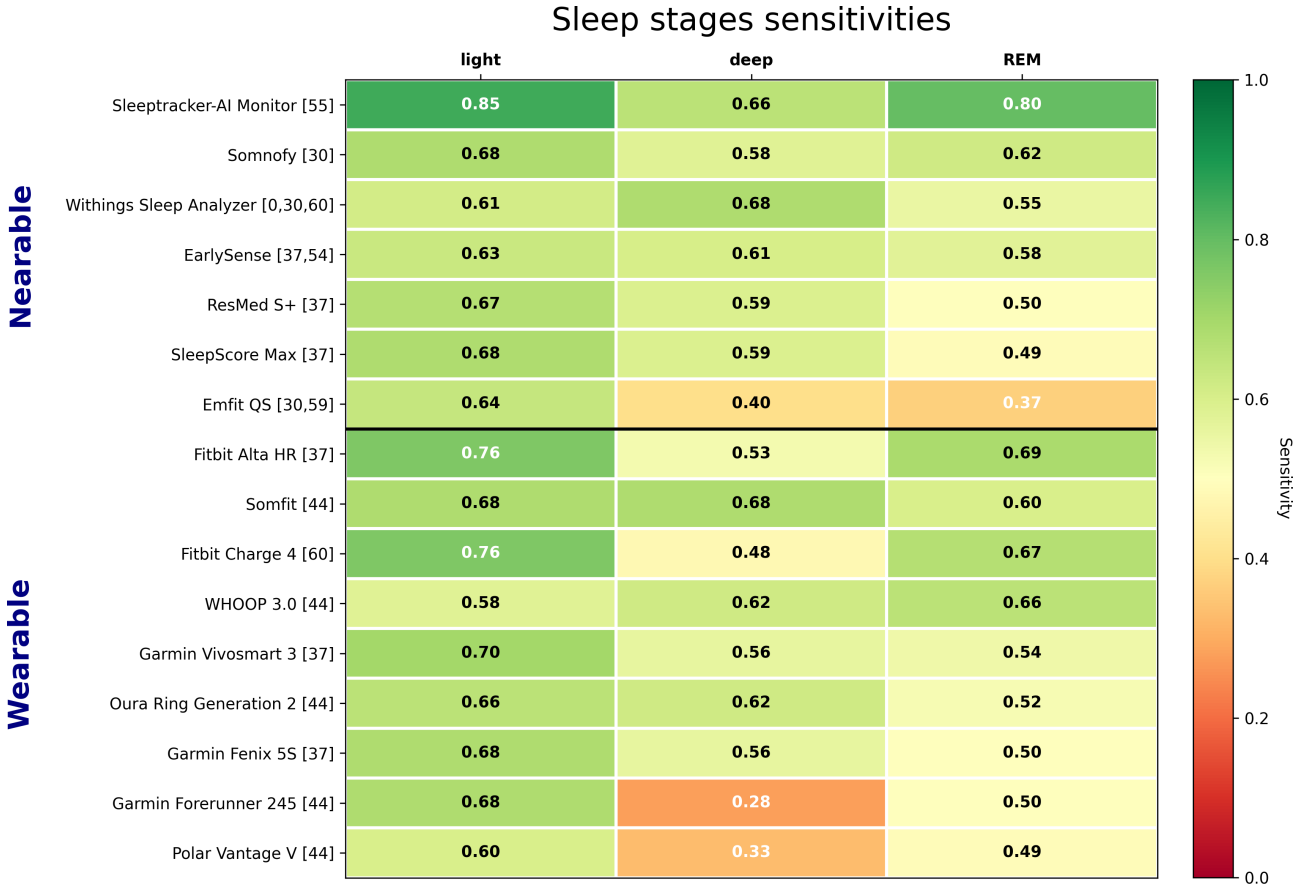
As explained above, the proportion of wake stages in the dataset before and after sleep onset can significantly impact the sleep/wake endpoints. These studies vary in the proportions of wake epochs, which refer to the total time spent awake divided by the total recording duration of the entire dataset. Specifically, these proportions ranged from 8% to 14% in [37], 15% in [54], ranged from 20% to 33% in [44], and from 34% to 42% in [33], compared to 30% in our dataset. This information was unfortunately not available for most of the studies evaluating nearable devices [55-60]. Lower proportions of wake data can inflate overall accuracies and kappas if sleep epoch classification dominates, as noted in [61]. In datasets with a low proportion of wake epochs, overall accuracy and kappa can be misleadingly high due to the dominance of sleep epoch classification, which is better classified than wake epochs as indicated in the previous paragraph.

Sleep Stages Classification

The trade-off observed in sleep-wake classification is even more pronounced in sleep stage analysis, where no single device consistently outperforms the others across all stages (Figure 14). The range of sensitivities is particularly large for REM, from 0.37 to 0.80, and quite large for deep sleep (0.28 to 0.68), even after excluding wearables (0.40 to 0.68). The sensitivities for light sleep detection are more homogeneous, with 12 sensitivities between 0.6 and 0.7. The WSA is in the top tier of devices for deep sleep classification, but midrange for light sleep and REM.

XSL•FO
RenderX

Figure 14. Overview of sleep stage sensitivities from varying validation studies. Indication [0] denotes this study. Note: Direct comparison is limited by differences in study protocols and populations [33,37,44,54,55,59,60].



The confusion matrices show that, for all devices, the highest proportion of misclassification for each stage is wrong light sleep detection, which can be explained by the inherent difficulty in distinguishing light from the other sleep stages by human experts, as shown in the Results section.

Device Comparison Conclusion

This analysis highlights the necessity of considering a comprehensive set of metrics to accurately evaluate the capabilities of each device. Relying on isolated overall metrics such as accuracy or kappa is insufficient to provide a fair assessment [62,63]. Every device exhibits relative strengths and weaknesses, and no single device clearly outperforms the others across all tasks. The WSA specifically demonstrates robust performance in sleep-wake distinction, while the precise identification of individual sleep stages remains a more challenging task.

Methodological Aspects and Strengths of the Study

The rapid adoption of new sleep technologies in consumer and clinical settings demands rigorous performance evaluation [64]. To improve the reliability and comparability of sleep research, we describe our methodology in detail and with transparency, an approach that is supported by recent guidelines advocating for uniform analysis pipelines [41,42,44]. Our study was designed to resolve methodological issues frequently found in the literature.

The study's strength lies first in its diverse sample of over 100 participants, who are for the most part healthy and with diverse demographic characteristics. Additionally, the inclusion of relevant factors that may negatively impact the device's performance, like the presence of a bed partner and variations in mattress type and thickness, further enriches the dataset, allowing for a comprehensive evaluation of the device's robustness across different real-life conditions in a healthy population. This allowed us to conduct meaningful subgroup analyses.

A second strength of the study is the method of PSG analysis. Unlike [44,55-58], which opted for a single reviewer, or two reviewers without consensus as in [37], our study is aligned with [33,38] by implementing a double review process with adjudication by a third reviewer in case of disagreement. This consensus methodology is preferable since it enhances the reliability of the sleep stage classification. In particular, we have shown that reviewers often err by misclassifying light sleep stages, a finding consistent with previous research reporting that the interrater reliabilities for light and deep stages were moderate [61]. A reliable assessment of devices' performance in sleep stage classification therefore requires a strong methodology to establish the ground truth.

Third, with a view to limit the participants' discomfort, we chose to perform PSG at home with a limited number of sensors, without the nasal cannula and oximeter. This setup reduces the sleep disturbances typically associated with lab-based PSG, allowing for a more natural assessment of sleep patterns. This limited set of sensors is sufficient for the evaluation of TST and sleep stage classification, with the trade-off that sleep onset,

XSL•FO
RenderX

time in bed, and wakefulness after sleep onset cannot be assessed. Yet, significant sleep disturbances were reported by the participant during the PSG test, underscoring the challenges of completely undisturbed monitoring even in home environments. Despite this, the study offers a realistic evaluation of the WSA's capabilities under typical sleeping conditions.

Fourth, we aimed to give a comprehensive report of the device's performance. Collecting epoch-by-epoch data enables a granular examination of sleep patterns, the consistency of measurement errors over time, and thus enables an accurate assessment of the reliability and utility of sleep monitoring devices in real-world settings. We emphasized the need for such comprehensive reporting and the elements of the protocol that may bias the performance estimates. Thus, transparency requires a presentation of the results including all the performance endpoints, each presenting a different facet of the device's performance [41-43].

In conclusion, given these methodological strengths, our study provides valuable insights into the WSA sleep stage algorithm's performance in real-world settings.

Limitations and Future Work

Despite our efforts for a solid methodology, the study faced several challenges.

Challenges in Defining Bedtime and Wake Time in Sleep Studies

A first challenge is created by the lack of a standardized definition of the start time and end time of a sleep study. As per AASM guidelines [50], bedtime and wake time are defined as the "lights-off" and "lights-on," respectively. This is not adapted to studies in real-world settings, where participants often go to bed and engage in activities before attempting to fall asleep. In everyday life, individuals may not turn off the lights immediately upon going to bed, nor do they always intend to sleep right away. Similarly, in the morning, people may not turn on the lights immediately upon awakening, and their actual wake time may not coincide with the "lights-on" reference. This discrepancy is compounded by the lack of a position sensor, which makes it difficult to accurately determine when a person intends to sleep. In other studies [37], start and end times of sleep are determined by the participants themselves, as suggested in standard sleep diaries [65], but the precision of this method is poor.

This lack of standardization is detrimental to the assessment of nearable and wearable devices intended to be used in home settings. Errors in sleep estimation, where wakefulness is mistakenly classified as sleep, are more frequent at the beginning and end of the night. The reason is that many devices, including the WSA, rely heavily on movement detection to distinguish wake from sleep. Thus, wake periods when the participant is immobile and calm in bed, such as when reading or scrolling with a low heart rate, are frequently mistaken as sleep by the device.

Additionally, the practice of activating the PSG immediately before sleep and deactivating it early prevents the gathering of comprehensive data on a participant's sleep habits. This includes behaviors such as spending extended periods in bed without the intention to sleep. To address this limitation, future studies could extend the duration of PSG recordings to capture activities that occur between the time a person gets into bed and falls asleep, as well as between waking and getting out of bed. This approach would provide a more complete picture of sleep behaviors and enable a thorough evaluation of sleep stage classification algorithms in real-world settings.

Challenges in Data Collection From a Nearable

A second challenge in studies with connected devices is data loss. Our study is no exception since nearly 19% of the data were lost, primarily due to the wireless transfer by WSA of a large amount of raw sensor data (250 samples per second). Note that this high rate is not representative of the intended use of the device, which, in its commercial version, only transfers processed data at a low frequency. To mitigate the risk of data loss, Wi-Fi routers paired with the WSA were installed at the participant's home. The routers experienced occasional connectivity problems. Data loss was further exacerbated by issues such as accidental unplugging of the device and improper reconnection without calibration. These technical challenges underscore the difficulties of gathering high-quality, high-frequency data in uncontrolled settings.

Another factor impacted the accuracy of sleep stage classification; despite receiving instructions from a technician, participants with large beds did not always sleep above the sensor but on the other side of the bed. In such cases, the WSA is not operated optimally, which may have restricted its performance.

Future Work

To validate the utility of longitudinal passive data collection, it is essential to demonstrate that errors in sleep quality measures are consistent over time for a given user.

In addition, our study population only includes healthy participants. There is no data on the performance of the device in special populations that might benefit the most from a longitudinal follow-up.

Looking ahead, the contactless devices hold significant potential for a wide range of research and clinical applications [14,64,66-70]. Their capability to monitor sleep in natural environments [68] underscores their versatility and relevance in advancing our understanding of sleep health and its implications.

Conclusions

In this study, we aimed to (1) assess the performance of a contactless device for sleep stage identification against the gold standard PSG, (2) benchmark it against other comparable devices, and (3) better understand the limitations of such devices. Our paper contributes to those, while also laying out important methodological issues.

This study contributes to the growing body of evidence supporting the utility of nearables for longitudinal sleep monitoring. The WSA's contactless, undermattress design facilitates unobtrusive data collection in free-living conditions

XSL•FO
RenderX

and demonstrates sleep-wake detection performance comparable to established consumer devices. While precise sleep stage classification remains a technical challenge, our results highlight the importance of testing on large, diverse, challenging datasets to capture real-world variability. This research underscores the need for standardized performance assessment methodologies to enable reliable comparisons across devices and studies, offering insights for future device design and guidance for utilization, advancing the goal of improved sleep monitoring for the general population [71].

Acknowledgments

We would like to acknowledge Paul Edouard and Sébastien Ducrot, both employees of Withings, for their crucial support and technical expertise in the offline run of the algorithm. We are also grateful to two anonymous reviewers and the journal editor for their suggestions and comments, which contributed to improving this manuscript. We declare the use of generative AI in the writing process; according to the GAIDeT taxonomy (2025), proofreading was delegated to the Gemini Pro 2.5 tool under full human supervision.

Funding

This study received funding from Withings.

Data Availability

The dataset analyzed during this study is not publicly available due to concerns for privacy, ethical, and legal reasons, but queries for data access can be directed to the corresponding author at privacy@withings.com.

Authors' Contributions

Conceptualization: MAS
Methodology: MAS (lead), DC (lead), MDG (equal), GDL (equal)
Software: MAS
Formal analysis: MAS
Investigation: MAS
Data Curation: MAS
Visualization: MAS
Writing – original draft preparation: MAS
Supervision: DC (lead), MDG (lead), GDL (lead), PAG (equal), ASI (equal)
Writing – review & editing: DC (lead), MDG (equal), GDL (equal), PAG (supporting), ASI (supporting)

Conflicts of Interest

This study received funding from Withings, the manufacturer of the device evaluated in this study. MAS is a PhD student affiliated with Université Paris Cité, Reykjavík University, and Withings. MDG, GDL, and DC are employees of Withings. PE is a consultant for Withings. Withings employees contributed to all the stages of this study, from conceptualization to manuscript writing. No other competing interests.

Multimedia Appendix 1
Details on methods and algorithm performance per sleep stage.
[DOCX File, 3319 KB - humanfactors_v13i1e77033_app1.docx]

Multimedia Appendix 2
Results of Kruskal-Wallis and Brown-Forsythe statistical tests followed by Holm-Bonferroni correction.
[XLSX File, 38 KB - humanfactors_v13i1e77033_app2.xlsx]

Multimedia Appendix 3
Comparison with previous research.
[XLSX File, 65 KB - humanfactors_v13i1e77033_app3.xlsx]

References

1. Steptoe A, O'Donnell K, Marmot M, Wardle J. Positive affect, psychological well-being, and good sleep. J Psychosom Res 2008 Apr;64(4):409-415. [doi: 10.1016/j.jpsychores.2007.11.008]
2. O'Leary K, Bylsma LM, Rottenberg J. Why might poor sleep quality lead to depression? A role for emotion regulation. Cogn Emot 2017 Dec;31(8):1698-1706. [doi: 10.1080/02699931.2016.1247035] [Medline: 27807996]

XSL•FO
RenderX

3. Lallukka T, Sivertsen B, Kronholm E, Bin YS, Øverland S, Glozier N. Association of sleep duration and sleep quality with the physical, social, and emotional functioning among Australian adults. Sleep Health 2018 Apr;4(2):194-200. [doi: 10.1016/j.sleh.2017.11.006]
4. Grandner MA. Sleep, health, and society. Sleep Med Clin 2022 Jun;17(2):117-139. [doi: 10.1016/j.jsmc.2022.03.001] [Medline: 35659068]
5. Geoffroy PA, Hoertel N, Etain B, et al. Insomnia and hypersomnia in major depressive episode: prevalence, sociodemographic characteristics and psychiatric comorbidity in a population-based study. J Affect Disord 2018 Jan;226:132-141. [doi: 10.1016/j.jad.2017.09.032]
6. Morin CM, Jarrin DC. Epidemiology of insomnia: prevalence, course, risk factors, and public health burden. Sleep Med Clin 2013;8(3). [doi: 10.1016/j.jsmc.2013.05.002] [Medline: 35659072]
7. Young TB. Epidemiology of daytime sleepiness: definitions, symptomatology, and prevalence. J Clin Psychiatry 2004;65 Suppl 16(Suppl 16):12-16. [Medline: 15575799]
8. Senaratna CV, Perret JL, Lodge CJ, et al. Prevalence of obstructive sleep apnea in the general population: a systematic review. Sleep Med Rev 2017 Aug;34:70-81. [doi: 10.1016/j.smrv.2016.07.002]
9. Rundo JV, Downey RI. Polysomnography. Handb Clin Neurol 2019;160:381-392. [doi: 10.1016/B978-0-444-64032-1.00025-4] [Medline: 31277862]
10. Rajaratnam SMW, Barger LK, Lockley SW, et al. Sleep disorders, health, and safety in police officers. JAMA 2011 Dec 21;306(23):2567-2578. [doi: 10.1001/jama.2011.1851]
11. Bloch KE. Polysomnography: a systematic review. Technol Health Care 1997 Oct;5(4):285-305. [doi: 10.3233/THC-1997-5403] [Medline: 9429270]
12. Keenan SA. Polysomnographic technique: an overview. In: Chokroverty S, editor. Sleep Disorders Medicine: Basic Science, Technical Considerations, and Clinical Aspects: Butterworth-Heinemann; 1994:79-94. [doi: 10.1016/B978-0-7506-9002-7.50011-2]
13. Kang SG, Kang JM, Ko KP, Park SC, Mariani S, Weng J. Validity of a commercial wearable sleep tracker in adult insomnia disorder patients and good sleepers. J Psychosom Res 2017 Jun;97:38-44. [doi: 10.1016/j.jpsychores.2017.03.009]
14. Lechat B, Scott H, Manners J, et al. Multi-night measurement for diagnosis and simplified monitoring of obstructive sleep apnoea. Sleep Med Rev 2023 Dec;72:101843. [doi: 10.1016/j.smrv.2023.101843]
15. Toussaint M, Luthringer R, Schaltenbrand N, et al. First-night effect in normal subjects and psychiatric inpatients. Sleep 1995 Jul;18(6):463-469. [doi: 10.1093/sleep/18.6.463] [Medline: 7481418]
16. Smith MT, McCrae CS, Cheung J, et al. Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: an American Academy of Sleep Medicine systematic review, meta-analysis, and GRADE assessment. J Clin Sleep Med 2018 Jul 15;14(07):1209-1230. [doi: 10.5664/jcsm.7228]
17. Kapur VK, Auckley DH, Chowdhuri S, et al. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American Academy of Sleep Medicine clinical practice guideline. J Clin Sleep Med 2017 Mar 15;13(03):479-504. [doi: 10.5664/jcsm.6506]
18. Braun M, Stockhoff M, Tijssen M, Dietz-Terjung S, Coughlin S, Schöbel C. A systematic review on the technical feasibility of home-polysomnography for diagnosis of sleep disorders in adults. Curr Sleep Medicine Rep 2024 Jun 1;10(2):276-288. [doi: 10.1007/s40675-024-00301-z]
19. Mykytyn IJ, Sajkov D, Neill AM, Mc Evoy RD. Portable computerized polysomnography in attended and unattended settings. Chest 1999 Jan;115(1):114-122. [doi: 10.1378/chest.115.1.114]
20. Andrade L, Paiva T. Ambulatory versus laboratory polysomnography in obstructive sleep apnea: comparative assessment of quality, clinical efficacy, treatment compliance, and quality of life. J Clin Sleep Med 2018 Aug 15;14(8):1323-1331. [doi: 10.5664/jcsm.7264] [Medline: 30092885]
21. Van de Water ATM, Holmes A, Hurley DA. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography – a systematic review. J Sleep Res 2011 Mar;20(1 Pt 2):183-200. [doi: 10.1111/j.1365-2869.2009.00814.x] [Medline: 20374444]
22. Pan Q, Brulin D, Campo E. Current status and future challenges of sleep monitoring systems: systematic review. JMIR Biomed Eng 2020 Aug 26;5(1):e20921. [doi: 10.2196/20921]
23. Vitazkova D, Kosnacova H, Turonova D, et al. Transforming sleep monitoring: review of wearable and remote devices advancing home polysomnography and their role in predicting neurological disorders. Biosensors (Basel) 2025 Feb 17;15(2):117. [doi: 10.3390/bios15020117] [Medline: 39997019]
24. Sivertsen B, Omvik S, Havik OE, et al. A comparison of actigraphy and polysomnography in older adults treated for chronic primary insomnia. Sleep 2006 Oct;29(10):1353-1358. [doi: 10.1093/sleep/29.10.1353]
25. Paquet J, Kawinska A, Carrier J. Wake detection capacity of actigraphy during sleep. Sleep 2007 Oct;30(10):1362-1369. [doi: 10.1093/sleep/30.10.1362] [Medline: 17969470]
26. de Zambotti M, Goldstein C, Cook J, et al. State of the science and recommendations for using wearable technology in sleep and circadian research. Sleep 2024 Apr 12;47(4):38149978. [doi: 10.1093/sleep/zsad325]
27. Kolla BP, Mansukhani S, Mansukhani MP. Consumer sleep tracking devices: a review of mechanisms, validity and utility. Expert Rev Med Devices 2016 May;13(5):497-506. [doi: 10.1586/17434440.2016.1171708] [Medline: 27043070]

XSL•FO
RenderX

28. Edouard P, Campo D, Bartet P, et al. Validation of the Withings Sleep Analyzer, an under-the-mattress device for the detection of moderate-severe sleep apnea syndrome. J Clin Sleep Med 2021 Jun;17(6):1217-1227. [doi: 10.5664/jcsm.9168]
29. Lechat B, Naik G, Reynolds A, et al. Multinight prevalence, variability, and diagnostic misclassification of obstructive sleep apnea. Am J Respir Crit Care Med 2022 Mar 1;205(5):563-569. [doi: 10.1164/rccm.202107-1761OC]
30. G Ravindran KK, della Monica C, Atzori G, et al. Reliable contactless monitoring of heart rate, breathing rate, and breathing disturbance during sleep in aging: digital health technology evaluation study.. JMIR Mhealth Uhealth 2024;12:e53643. [doi: 10.2196/53643]
31. Jagielski JT, Bibi N, Gay PC, et al. Evaluating an under-mattress sleep monitor compared to a peripheral arterial tonometry home sleep apnea test device in the diagnosis of obstructive sleep apnea. Sleep Breath 2023 Aug;27(4):1433-1441. [doi: 10.1007/s11325-022-02751-7]
32. Huyett P, Yu PK, Kent D, et al. Comparison of home sleep devices and sleep study testing in hypoglossal nerve stimulation patients. Laryngoscope 2025 Jan;135(1):469-477. [doi: 10.1002/lary.31710] [Medline: 39162306]
33. Ravindran KKG, della Monica C, Atzori G, et al. Contactless and longitudinal monitoring of nocturnal sleep and daytime naps in older men and women: a digital health technology evaluation study. Sleep 2023 Oct 11;46(10):zsad194. [doi: 10.1093/sleep/zsad194]
34. Yang RY, Bendjoudi A, Buard N, Boutouyrie P. Pneumatic sensor for cardiorespiratory monitoring during sleep. Biomed Phys Eng Express 2019;5(5):055014. [doi: 10.1088/2057-1976/ab3ac9]
35. Hedner J, White DP, Malhotra A, et al. Sleep staging based on autonomic signals: a multi-center validation study. J Clin Sleep Med 2011 Jun 15;7(3):301-306. [doi: 10.5664/JCSM.1078] [Medline: 21677901]
36. Toften S, Pallesen S, Hrozanova M, Moen F, Grønli J. Validation of sleep stage classification using non-contact radar technology and machine learning (Somnofy®). Sleep Med 2020 Nov;75:54-61. [doi: 10.1016/j.sleep.2020.02.022]
37. Chinoy ED, Cuellar JA, Huwa KE, et al. Performance of seven consumer sleep-tracking devices compared with polysomnography. Sleep 2021 May 14;44(5):zsaa291. [doi: 10.1093/sleep/zsaa291] [Medline: 33378539]
38. Kainec KA, Caccavaro J, Barnes M, Hoff C, Berlin A, Spencer RMC. Evaluating accuracy in five commercial sleep-tracking devices compared to research-grade actigraphy and polysomnography. Sensors (Basel) 2024;24(2):635. [doi: 10.3390/s24020635]
39. Nelson BW, Saeb S, Barman P, et al. Performance evaluation of the verily numetric watch sleep suite for digital sleep assessment against in-lab polysomnography. J Sleep Res 2025 Oct;34(5):e70036. [doi: 10.1111/jsr.70036] [Medline: 40170300]
40. Willoughby AR, Golkashani HA, Ghorbani S, et al. Performance of wearable sleep trackers during nocturnal sleep and periods of simulated real-world smartphone use. Sleep Health 2024 Jun;10(3):356-368. [doi: 10.1016/j.sleh.2024.02.007] [Medline: 38570223]
41. Benedetti D, Menghini L, Vallat R, et al. Call to action: an open-source pipeline for standardized performance evaluation of sleep-tracking technology. Sleep 2023 Feb 8;46(2):zsac304. [doi: 10.1093/sleep/zsac304] [Medline: 36611112]
42. Menghini L, Cellini N, Goldstone A, Baker FC, de Zambotti M. A standardized framework for testing the performance of sleep-tracking technology: step-by-step guidelines and open-source code. Sleep 2021 Feb 12;44(2):zsaa170. [doi: 10.1093/sleep/zsaa170] [Medline: 32882005]
43. Chee MWL, Baumert M, Scott H, et al. World Sleep Society recommendations for the use of wearable consumer health trackers that monitor sleep. Sleep Med 2025 Jul;131:106506. [doi: 10.1016/j.sleep.2025.106506]
44. Miller DJ, Sargent C, Roach GD. A validation of six wearable devices for estimating sleep, heart rate and heart rate variability in healthy adults. Sensors (Basel) 2022 Aug 22;22(16):6317. [doi: 10.3390/s22166317] [Medline: 36016077]
45. Birrer V, Elgendi M, Lambercy O, Menon C. Evaluating reliability in wearable devices for sleep staging. NPJ Digit Med 2024 Mar 18;7(1):74. [doi: 10.1038/s41746-024-01016-9] [Medline: 38499793]
46. Yuan H, Hill EA, Kyle SD, Doherty A. A systematic review of the performance of actigraphy in measuring sleep stages. J Sleep Res 2024 Oct;33(5):e14143. [doi: 10.1111/jsr.14143] [Medline: 38384163]
47. Imtiaz SA. A systematic review of sensing technologies for wearable sleep staging. Sensors (Basel) 2021 Jan;21(5):1562. [doi: 10.3390/s21051562]
48. Berry RB. Fundamentals of Sleep Medicine: Elsevier; 2012.
49. Buysse DJ, Reynolds CF III, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
50. Berry RB, Brooks R, Gamaldo CE, Harding SM, Marcus CL, Vaughn BV. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications: American Academy of Sleep Medicine; 2015. URL: https://aasm.org/resources/pdf/scoring-manual-preface.pdf [accessed 2026-02-14]
51. Schmidt M, Schumann A, Müller J, Bär KJ, Rose G. ECG derived respiration: comparison of time-domain approaches and application to altered breathing patterns of patients with schizophrenia. Physiol Meas 2017 Apr 1;38(4):601-615. [doi: 10.1088/1361-6579/aa5feb]

XSL•FO
RenderX

52. de Zambotti M, Goldstone A, Claudatos S, Colrain IM, Baker FC. A validation study of Fitbit Charge 2TM compared with polysomnography in adults. Chronobiol Int 2018 Apr;35(4):465-476. [doi: 10.1080/07420528.2017.1413578] [Medline: 29235907]
53. Lee YJ, Lee JY, Cho JH, Choi JH. Interrater reliability of sleep stage scoring: a meta-analysis. J Clin Sleep Med 2022 Jan;18(1):193-202. [doi: 10.5664/jcsm.9538]
54. Tal A, Shinar Z, Shaki D, Codish S, Goldbart A. Validation of contact-free sleep monitoring device with comparison to polysomnography.. J Clin Sleep Med 2017 Mar 15;13(3):517-522. [doi: 10.5664/jcsm.6514] [Medline: 27998378]
55. Ding F, Cotton-Clay A, Fava L, et al. Polysomnographic validation of an under-mattress monitoring device in estimating sleep architecture and obstructive sleep apnea in adults. Sleep Med 2022 Aug;96:20-27. [doi: 10.1016/j.sleep.2022.04.010]
56. Ellender CM, Zahir SF, Meaklim H, Joyce R, Cunnington D, Swieca J. Prospective cohort study to evaluate the accuracy of sleep measurement by consumer-grade smart devices compared with polysomnography in a sleep disorders population. BMJ Open 2021 Nov;11:e044015. [doi: 10.1136/bmjopen-2020-044015]
57. Hsiou DA, Gao C, Matlock RC, Scullin MK. Validation of a nonwearable device in healthy adults with normal and short sleep durations. J Clin Sleep Med 2022 Mar 1;18(3):751-757. [doi: 10.5664/jcsm.9700] [Medline: 34608858]
58. Tuominen J, Peltola K, Saaresranta T, Valli K. Sleep parameter assessment accuracy of a consumer home sleep monitoring ballistocardiograph beddit sleep tracker: a validation study.. J Clin Sleep Med 2019 Mar 15;15(3):483-487. [doi: 10.5664/jcsm.7682] [Medline: 30853052]
59. Kholghi M, Szollosi I, Hollamby M, Bradford D, Zhang Q. A validation study of a ballistocardiograph sleep tracker against polysomnography. J Clin Sleep Med 2022 Apr 1;18(4):1203-1210. [doi: 10.5664/jcsm.9754] [Medline: 34705630]
60. Manners J, Kemps E, Lechat B, Catcheside P, Eckert DJ, Scott H. Performance evaluation of an under-mattress sleep sensor versus polysomnography in > 400 nights with healthy and unhealthy sleep. J Sleep Res 2025 Dec;34(6):e14480. [doi: 10.1111/jsr.14480] [Medline: 40017337]
61. Kelly JM, Strecker RE, Bianchi MT. Recent developments in home sleep-monitoring devices. ISRN Neurol 2012 Oct 14;2012:1-10. [doi: 10.5402/2012/768794]
62. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. J Clin Epidemiol 1993 May;46(5):423-429. [doi: 10.1016/0895-4356(93)90018-v] [Medline: 8501467]
63. Gjørup T. The kappa coefficient and the prevalence of a diagnosis. Methods Inf Med 1988 Oct;27(4):184-186. [doi: 10.1055/s-0038-1635539] [Medline: 3226333]
64. Pellegrini M, Lannin NA, Mychasiuk R, Graco M, Kramer SF, Giummarra MJ. Measuring sleep quality in the hospital environment with wearable and non-wearable devices in adults with stroke undergoing inpatient rehabilitation.. Int J Environ Res Public Health 2023 Feb 23;20(5):3984. [doi: 10.3390/ijerph20053984] [Medline: 36900995]
65. Carney CE, Buysse DJ, Ancoli-Israel S, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. Sleep 2012 Feb;35(2):287-302. [doi: 10.5665/sleep.1642]
66. Lechat B, Naik G, Appleton S, et al. Multi-night assessment of snoring is associated with hypertension independent of OSA severity [Abstract]. Am J Respir Crit Care Med 2023;207:A6200. [doi: 10.1164/ajrccm-conference.2023.207.1_MeetingAbstracts.A6200]
67. de Havenon A, Falcone G, Rivier C, et al. Impact of sleep quality and physical activity on blood pressure variability. PLoS ONE 2024;19(4):e0301631. [doi: 10.1371/journal.pone.0301631]
68. Scott H, Lechat B, Guyett A, et al. Sleep irregularity is associated with hypertension: findings from over 2 million nights with a large global population sample. Hypertension 2023 May;80(5):1117-1126. [doi: 10.1161/HYPERTENSIONAHA.122.20513]
69. Scott H, Naik G, Lechat B, et al. Are we getting enough sleep? Frequent irregular sleep found in an analysis of over 11 million nights of objective in-home sleep data. Sleep Health 2024 Feb;10(1):91-97. [doi: 10.1016/j.sleh.2023.10.016]
70. Chapple C, Bliwise D, Maislisch L, Roitmann E, Burtea T. Night-time voids, level of bother and sleep characteristics in a non-patient population of wearable devices users. Int J Clin Pract 2020 Jul;74(7):e13495. [doi: 10.1111/ijcp.13495] [Medline: 32100396]
71. Chaudhry FF, Danieletto M, Golden E, et al. Sleep in the natural environment: a pilot study. Sensors (Basel) 2020;20(5):1378. [doi: 10.3390/s20051378]

Abbreviations

AASM: American Academy of Sleep Medicine
ECG: electrocardiogram
EEG: electroencephalogram
EMG: electromyogram
EOG: electrooculogram
MAE: mean absolute error
MCAR: missing completely at random
NEDeep: number of episodes in deep sleep stage

XSL•FO
RenderX

NELight: number of episodes in light sleep stage
NEREM: number of episodes in REM sleep stage
NEWake: number of episodes in wake stage
PIDeep: proportion of the night in deep sleep stage
PILight: proportion of the night in light sleep stage
PIREM: proportion of the night in REM sleep stage
PIWake: proportion of the night in wake stage
PSG: polysomnography
PSQI: Pittsburgh Sleep Quality Index
REM: rapid eye movement (sleep stage)
TIDeep: time spent in deep sleep stage
TILight: time spent in light sleep stage
TIREM: time spent in REM sleep stage
TIWake: time spent in wake stage
TST: total sleep time
WSA: Withings Sleep Analyzer

Edited by J Waterson; submitted 06.May.2025; peer-reviewed by BW Nelson, FA Etindele Sosso, L Pinilla; revised version received 10.Jan.2026; accepted 26.Jan.2026; published 02.Apr.2026.

Please cite as:
Stefanos MA, de Laboulaye G, Campo D, de Gourcuff M, Escourrou P, Matrot B, Islind AS, Geoffroy PA
Evaluation of a Contactless Sleep Monitoring Device for Sleep Stage Detection at Home in a Healthy Population: Prospective Study in Free-Living Conditions
JMIR Hum Factors 2026;13:e77033
URL: https://humanfactors.jmir.org/2026/1/e77033
doi:10.2196/77033

© Marie-Ange Stefanos, Guillaume de Laboulaye, David Campo, Martin de Gourcuff, Pierre Escourrou, Boris Matrot, Anna Sigridur Islind, Pierre Alexis Geoffroy. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 2.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Designing for Knowledge Generalization in Medical Device Instructions Among Health Care Professionals: Qualitative Interview and Observation Study

Qing (Nancy) Xia[1], MSc; Jeremy Opie[1,2], PhD; Tom Brookshaw[3], MSc; Eloy Erro Barbarin[3], MSc; Ann Blandford[1,2], PhD, CEng; Clare Selden[3], PhD

[1]UCL Interaction Centre, University College London, 66-72 Gower Street, London, United Kingdom

[2]Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS), University College London, London, United Kingdom

[3]Institute for Liver and Digestive Health, University College London, London, United Kingdom

Corresponding Author:
Qing (Nancy) Xia, MSc
UCL Interaction Centre, University College London, 66-72 Gower Street, London, United Kingdom

Abstract

Background: Instructional documents are crucial for safely operating medical devices. However, few studies have explicitly considered designing instructions for health care professionals (HCPs). Psychological research suggests that advanced learners with prior expertise in a domain are better able to generalize their existing knowledge to new areas compared to those with little prior expertise, resulting in different informational needs.

Objective: We aimed to understand how HCPs learn from and use their existing expertise when interacting with instructions for use to familiarize themselves with a novel medical device. This would allow us to explore how instructional documents could be designed to better accommodate HCPs' needs as experienced learners with prior knowledge in the domain.

Methods: We conducted our studies centered around a novel liver support system, the HepatiCan and its current instructions for use. We conducted 3 user studies, first using semistructured interviews and think-aloud protocols to understand HCPs' expectations for the process of setting up the HepatiCan and how training for the use of medical devices typically takes place. We followed this up with an observational study to understand HCPs' actual experience in practice and corresponding instructional design needs.

Results: Our results showed that HCPs' prior expertise allowed them to prioritize key areas for attention, but also led them to make assumptions and potentially skim and miss important information. Visual representations and communication were preferred, as well as designs that supported memorization, as this was an essential function of using instruction manuals.

Conclusions: Developers must be aware of what is considered "common" and potentially ignored knowledge within a specific domain in order to clearly emphasize important, safety-relevant information. We also recommend developers focus on using images with attached appendices where possible. This allows advanced users to process information more efficiently without compromising the needs of less familiar users for greater instructional details.

(JMIR Hum Factors 2026;13:e82405) doi:10.2196/82405

KEYWORDS

instructional design; interview; think-aloud; observation; health care professionals; human factors engineering; usability studies; medical devices

Introduction

Medical devices play a crucial role in the care and treatment of patients. Ensuring the safety and usability of such devices is therefore essential [1], particularly in mitigating human errors. Human errors account for nearly 70% of all errors leading to preventable adverse medical events [2], such as patient injury or death [3].

One of the primary contributors to medical error is a lack of consideration for human factors and, consequently, poor design that fails to account for user needs [4-6]. Human factors, as defined by the Association for the Advancement of Medical Instrumentation, is "the application of knowledge about human capabilities (physical, sensory, emotional, and intellectual) and limitations, to the design and development of tools, devices, systems, environments, and organizations" [7]. Developing a strong knowledge of users' needs and capabilities is therefore essential to appropriately cater for and support users in engaging in safe practices, particularly in the context of a medical device.

XSL•FO
RenderX

A key resource for ensuring the usability of a medical device is its accompanying documentation and manuals, that is, instructions for use (IFU). In this paper, the term IFU is broadly used to encompass any document containing information or guidelines that could inform users about how a task or procedure should be executed under normal conditions [8]. This includes standard operating procedures and commercial user manuals. As the functionality of novel medical technologies increases, however, so too does the complexity of user operation and the likelihood of errors during initial setup [9]. IFUs therefore serve as an essential tool for supporting user learning with novel medical devices and should be subject to the same level of rigor and assessment for usability, user comprehension, and learning, as any other component of the system [10]. Indeed, there have been recent demands for greater transparency and clarification for how safety standards and evaluations should be implemented during the development of a novel medical device [11,12]. Our study adds to the growing collection of studies that address these calls to present detailed usability evaluations across a range of medical devices [5,13-16].

IFUs typically communicate information via both text-based and visual representations. Within the field of instructional design, cognitive load theory is commonly applied to help designers explain which types of media and forms of communication may be most beneficial for different types of users in information comprehension and processing [17-19]. Cognitive load theory suggests that individuals' internal cognitive resources, such as working memory, are finite, and tasks such as learning and processing information place demands, or "cognitive loads," on these resources [18,20]. Higher cognitive loads can limit an individual's capacity to attend to important information and are thus more likely to result in lower task performance [21].

The extent to which cognitive load is induced in an individual depends on external factors, such as the design of interaction materials, as well as individual factors, such as an individual's prior knowledge. This is why it is important to study health care professionals (HCPs) and evaluate their needs in instructional design as distinct from that of layman groups or patients, as HCPs differ significantly in prior domain experience and related medical device knowledge. Previous research suggests that individuals with prior domain experience perform better using procedural, action-oriented, and concise instructions (eg, image only), but worse when using highly detailed, explanatory instructions that are more useful for novices. This phenomenon is known as the "expertise reversal" effect [18,22], and has been observed across many domains and task types ranging from literary interpretation to accounting [18,23,24]. In addition, experienced individuals can encounter negative knowledge transfer [25]. This means they may struggle to disengage from well-established routines and adapt to new knowledge, such as differences in interface designs between similar products [26,27]. Previous research has considered the role of prior knowledge in patients' use of medical devices [28] and the impact of cognitive workload on HCPs' preferences for receiving patient feedback [29]. However, there remains a lack of insight into how HCPs themselves navigate unfamiliar medical devices, especially how their extensive training and prior expertise shape their instructional design preferences.

Our study aims to address the issue of evaluating instructional design principles for HCPs [4,30-32]. First, medical device developers often encounter challenges in accessing the perspectives of HCPs during development [33], so it is important for research to provide a clear and practical description of HCPs' needs. Second, despite the importance of designing devices suited to user needs, there is little advice currently available in regulatory guidelines about specific considerations that are relevant for different user groups (eg, designing for patients or nonmedical staff, compared to trained HCPs) [7,8]. While existing literature has highlighted the importance of supporting HCPs' training and understanding the challenges these individuals face when navigating complex technologies [34,35], there has been little consideration for the specific characteristics of HCPs which differentiate their needs from other user groups in human factors.

The goal of our research is therefore to investigate how challenges of learning and knowledge transfer may reflect in HCPs' behaviors in an applied learning setting, and to consider the types of designs which may help to mitigate these issues based on users' feedback and performance (ie, accuracy or errors made) when following instructions. In the context of medical devices, errors or issues in performance could lead to delays in patient treatment or potentially harmful impacts to patients during treatment [9]. Our research aim is therefore to understand how HCPs learn from, and use their existing expertise when interacting with IFUs to familiarize themselves with a novel medical device.

To address this issue, we use a human-centered approach following the guidelines of the International Electrotechnical Commission [8] to conduct a series of usability studies involving interviews, think-aloud protocols, and observations of HCPs interacting with the IFU for a novel medical device known as the HepatiCan. By conducting these studies, we contribute further insight into the information processing needs of HCPs as experienced learners, how such needs manifest in terms of user behaviors, and provide implications for designing instructions for HCPs.

Methods

Case Study: The HepatiCan

HepatiCan is an extracorporeal liver support system that uses cell therapy to treat patients with liver failure. It serves to temporarily replace the functions of the liver, allowing patients' livers to repair and regenerate. HepatiCan consists of 3 circuits of tubing that are used to circulate blood for treatment. As part of this assembly, sections of circuitry tubing must be connected together using Luer connections and installed into the device's workstation.

HepatiCan served as an ideal case study for our research purposes because it shares many common features with existing extracorporeal medical devices, such as dialysis machines and apheresis machines, so there was room for an expert in extracorporeal treatment to generalize their knowledge in

XSL•FO
RenderX

assembling HepatiCan; and because, as a device still within early developmental stages at the time of the study, there were no existing users whose experience would have otherwise confounded our exploration of "novice" interactions with the IFU. Our paper focuses exclusively on the instructions for assembly for HepatiCan, as this is the first stage of device use before active operation.

Though our focus on the HepatiCan as a case study necessarily means that our sample and corresponding findings are oriented on the experiences of practitioners in extracorporeal treatment, we have striven in this paper to present the needs and patterns that may be applicable to HCPs of varying backgrounds, rather than focusing on specific comments or details which may only be applicable to HepatiCan or the field of extracorporeal treatment.

Study Design

Part of the original study design aimed to evaluate the usability of the HepatiCan, but in this paper, we present only the information related to the current research aims regarding HCPs' prior knowledge and instructional needs. The overall study consisted of 2 separate phases, each designed to address a specific research goal. We applied a qualitative descriptive design to capture users' self-reported experiences in general medical device training as well as their reactions and interactions with the IFU provided in our study, as well as a pragmatic research paradigm to generate actionable insights and recommendations for designers. The first phase consisted of a semistructured interview to understand how HCPs use IFUs in their current workflows and as part of their learning. This helped to contextualize the role of IFUs for HCPs within their wider responsibilities. In the second phase, participants engaged directly with the HepatiCan IFU through a think-aloud and observational study. The think-aloud study had participants read and react to both the IFU document and videos of the IFU procedures carried out by a developer, capturing participants' perceptions of "work-as-imagined" (WAI) based on simply reading the IFU [36,37]. The observational study had participants interact directly with the HepatiCan and attempt assembly themselves while following the IFU. This allowed us to capture participants' real-world experiences of following the IFU, or "work-as-done" (WAD) [36,37]. While the arrangement for participants to read through the IFU before interacting with the HepatiCan mirrored how training interventions take place for practitioners, as reflected in the interviews, the goal of this study was not to examine how practitioners respond to training for a novel medical device, but instead the challenges they face in interpreting the IFU specifically. As such, study procedures were designed to place exclusive focus on how participants react to and interact with the IFU.

Participants

Participants were recruited to this study via invitation from the HepatiCan project leader, as they knew relevant experts with transferable knowledge. Participants had to fulfill 2 criteria to be considered eligible. First, as we were interested in studying "experienced" learners, it was necessary to recruit individuals with existing expertise in areas that were relevant or highly generalizable to the knowledge required in the assembly of HepatiCan's device components. We therefore looked for participants who had experience in related types of extracorporeal treatment (eg, dialysis), experience or knowledge in hepatology, and/or experience in similar types of extracorporeal systems and medical devices (eg, apheresis machines). This was essential, as HepatiCan is both an extracorporeal device and one that provides hepatological treatments, and thus, in implementation, is intended to be operated by practitioners who work with extracorporeal devices (such as apheresis machines) and those working in hepatological treatments. However, as this is a novel medical device, there are currently no existing user groups. Second, it was essential for us to recruit participants without any prior knowledge or familiarity with the process of assembling the HepatiCan to more accurately replicate the authentic experience of a target user encountering and following the IFU as they would in a real-world context.

In total, 6 participants were invited to take part in the research, though only 3 participants took part in all 3 user studies (Table 1). This limited number was due to the challenges of recruiting participants in the corresponding specialism who had availability during the scheduled timeframe for conducting the study. For the observational study, we required participants to travel to the researcher's location to engage with HepatiCan, which was a unique prototype and could not be readily transported. Participants were compensated for their time and travel expenses.

XSL•FO
RenderX

Table . Participants' backgrounds and participation across all 3 user studies.

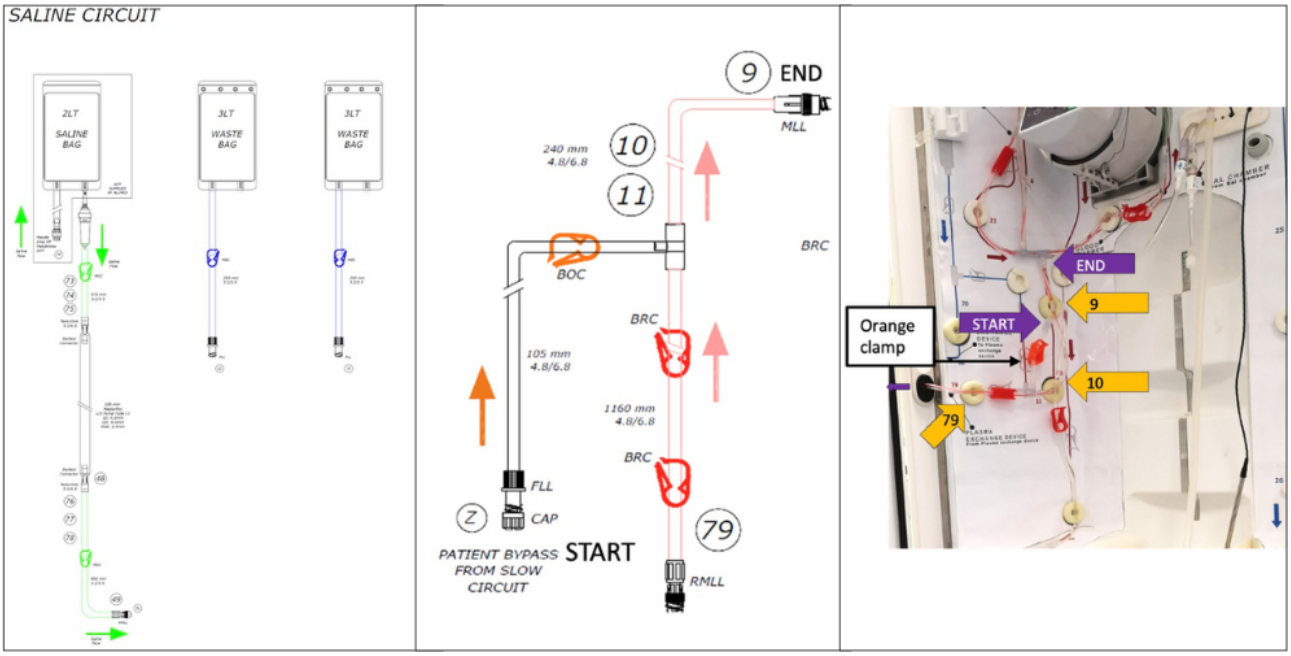
Participant ID	Study participation	Background and experience
P1	• Think-aloud • Observation	PhD graduate in cryobiology
P2	• Interview • Think-aloud	Training consultant for apheresis machines
P3	• Interview • Think-aloud • Observation	Researcher and clinical consultant specializing in hepatology and nephrology
P4	• Interview • Think-aloud • Observation	Perfusion specialist and clinical operations manager
P5	• Interview • Think-aloud	Perfusion specialist and clinical research manager
P6	• Observation	Clinical apheresis consultant and researcher

These practical considerations limited the outreach and participant sample of our study, which is a common challenge in health care–centered human factors research. However, our use of qualitative methods ensured that we were able to gather rich information and insights from each participant, and our careful recruitment criteria ensured that our findings reflected the experiences of participants who could be considered knowledgeable learners for the context in question.

Materials

Our study focused on the IFU for the assembly of the HepatiCan, which was used in both the think-aloud and observational study. This document contained a combination of text and labeled images, with instructions divided across the 3 main circuits of the HepatiCan. Three different types of images were used overall—full circuit diagrams (Figure 1A), circuit section diagrams (Figure 1B), and annotated photos (Figure 1C). Full circuit diagrams depicted labeled engineer schematics of the entirety of HepatiCan's tubing and inner circuitry. The circuit section diagrams consisted of segments from the full circuit diagram, but zoomed in to show only the relevant circuit section and were presented at the beginning of the corresponding instructions for assembling a particular section. The circuit section diagrams were accompanied by annotated photos of the particular circuit section when assembled to show how the circuit should appear when correctly installed, initially with unmarked colored arrows to denote positions of components or tubing holders. Photos were also used to supplement text instructions to show a consecutive series of steps within a relevant procedure (eg, to demonstrate how 2 locking Luers should be connected).

Figure 1. Example of a full circuit diagram schematic (left), a circuit section diagram (right), and an annotated photo for a circuit section.



XSL•FO
RenderX

For the think-aloud, participants were additionally shown video clips of a HepatiCan developer conducting the procedures outlined on the IFU. The videos were silent with the text from the IFUs overlayed as subtitles showing one of the developers assembling the corresponding section. Each clip corresponded to one section of the IFU. There were 4 videos in total, each video corresponding to sections of the IFU relating to assembly of 4 distinct sections of the HepatiCan. The total length of the videos was 12 minutes and 18 seconds. Participants viewed around 10 - 30 seconds of the video at a time, representing each step on the IFU.

For the observational study, participants were invited to the laboratory and were asked to assemble the HepatiCan according to the IFU. The materials given to participants included (1) an empty HepatiCan workstation, with the component tubing required for assembly packed in a separate box and labeled as they would be once the device is in clinical use; (2) a printed, color copy of the IFU on single-sided loose pages; and (3) 3 printed and laminated quick reference guides, each on one side of A4, containing the full tubing circuit diagrams for the workstation.

Data Collection

Both the semistructured interview and think-aloud study were conducted in a single session online using Microsoft Teams. During the interview, the first author used an interview guide to discuss participants' main priorities when learning from IFUs, their perceptions of how their prior expertise informs their approach to navigating a novel medical device, and how IFUs are generally used during work or training. These discussions were focused on participants' own experiences and did not involve the HepatiCan or its related documents at this stage. Interviews took around 30 minutes to complete.

Following the interview, participants conducted the think-aloud study. In this study, participants read through the IFU and watched step-by-step video clips of the described assembly process carried out by a HepatiCan developer. Participants verbalized their feedback on the IFU and how well the writing conveyed and aligned with their expectations when watching the video. Feedback on the presentation and layout of the IFU was then implemented following the think-aloud study to improve usability in preparation for the second phase. The think-aloud study took around 1.5 - 2 hours to complete.

The observational study was then conducted 2 to 3 weeks after the think-aloud study. Participants were presented with a physical copy of the IFU and were asked to follow the instructions to assemble the HepatiCan device in person. The study was overseen by the first and second authors; 2 developers, 1 in-person and 1 online, were also present throughout the observation. Participants were asked to follow the IFU as they would naturally and were free to ask questions to one of the developers present if they felt it necessary. The in-person developer was told to answer questions and help if asked by the participants, but otherwise not to explicitly intervene except if there was a risk of serious harm or damage. They were also told to identify instances of errors to the researcher (but not to the participant) so that any mistakes which were not noticed or explicitly addressed by the participant could be identified. This was because direct intervention and corrections from developers may otherwise mask the issues and challenges users experience due to instructions from the IFU. The first author observed and recorded participants' performance in the assembly. Performance was evaluated based on the time taken to complete the assembly procedure, as well as challenges encountered by users, such as errors, hesitation, confusion, and whether they interacted in unexpected ways with the system or required intervention from the developer. The average completion time for the assembly task was around 2 hours across all 4 participants.

Following the observation, the first author conducted a 15-minute follow-up interview exploring (1) overall use experience; (2) preferences and positive interactions with the IFU; (3) challenges and negative interactions with the IFU; (4) use of quick reference materials; (5) addressing key user needs of learnability, ensuring accuracy, and preserving sterility; and (6) suggestions for improvements in the IFU.

Analysis

In the first phase of the study, the interview transcripts were generated automatically from Microsoft Teams recordings and were manually corrected by the first author. The interview transcripts were then analyzed using inductive thematic analysis to ensure that the themes identified were reflective of participants' responses [38]. Descriptive codes were initially generated to provide an overview of the transcript data related to the research aims. These codes were then grouped based on relatedness and summarized by higher-level themes.

For the second phase of the study (ie, the think-aloud and observational studies), a spreadsheet was used with columns to represent each section of the IFU evaluated, and each column represented an individual participant's feedback and queries for each of the respective IFU sections. For the observation, notes were documented and classified using predefined categories to highlight challenges users encountered with using the IFU during assembly. The categories included comments or behaviors related to (1) how participants used and interacted with the IFU; (2) the assembly procedures and how participants conducted each stage of assembly; (3) participants' experience physically interacting with the hardware, that is, tubing or the workstation itself; (4) errors made by participants while conducting the task (defined as any deviations from the IFU procedures and including errors which were later corrected by the user themselves or the developer); (5) any comments verbalized during the task; and (6) questions and clarifications directed to developers during the assembly process. The follow-up interview was recorded using the first author's phone and manually transcribed.

During analysis of the data collected in the second phase, descriptive codes informed by the literature were used with a specific focus on topics relevant to the current research aims. This included (1) comments evaluating information appropriateness and relevance to their specific expertise level; (2) comments assessing the advantages and drawbacks of different types of information and presentation formats (eg, visual and text-based); and (3) observed interaction behaviors with the IFU which correspond either to participants' own descriptions and reported issues with the IFU, or are in line with

XSL•FO
RenderX

previous literature documenting challenges experienced users face with specific types of IFUs [17,22].

Our findings are validated through methodological triangulation [39], as different research methods used to explore the same research aim reveal corresponding evidence to support our conclusions. However, there were also instances where participants' think-aloud reflections and imagined assessment of the IFU differed from their observed behavior and reflections after real-world interactions with the IFU. We report these differences in the "Results" section. In cases of discrepancies, we prioritize recommendations for instructional design based on participants' observed behaviors rather than their initial think-aloud reactions. This is because individuals' ability to successfully carry out the necessary procedures had greater implications for risk assessment and patient safety in practice compared to their self-reported perceptions and preferences to specific aspects of the IFU materials.

Ethical Considerations

The University College London Interaction Centre Research Ethics Committee approved both studies under project number UCLIC/1819/006/BlandfordProgramme-Ethics. The interview and think-aloud activity in Study 1 was conducted and recorded online using Microsoft Teams, which meets the standards of the General Data Protection Regulation and is managed by University College London. Recordings and transcripts were initially stored on OneDrive before being transferred to an encrypted USB drive. All participants were informed of the multistage nature of the study in the information sheet and consent form prior to participation and were given the option to partake in whichever stage they were available for. Participants were paid £35 (US $47.19) for participating in the online study (interview and think-aloud), and £65 (US $87.65) for the in-person session, as well as compensation for travel costs. Transcripts were pseudonymized, and all transcripts and data were stored in an encrypted USB drive.

Results

Overview

In the following section, we summarize and report the main findings related to how HCPs use their existing knowledge when interacting with IFUs, and what their corresponding needs may be. We divide the results into 2 sections. The first section reports findings from the semistructured interview study, providing the background and context for understanding how HCPs use IFUs for medical devices in their typical work context. The second section combines results from the think-aloud study and observation. This highlights the key themes identified in how participants initially reacted to the IFU document, both when reading and when applying the document in practice.

Interview Study

IFUs as Initial Learning Tools to Be Discarded With Expertise

The IFU was an essential part of the users' initial learning process when interacting with a novel medical device. Participants described how IFUs were useful for developing their early conceptual understanding of the device and helping to fill in any gaps in their existing knowledge, such as unfamiliar components within the system. P4 and P5 emphasized that an initial read-through of the IFU was an important part of the process before instructions were implemented and followed in practice.

> *The operator's manual is a fundamental tool because you need to go through the operator's manual before you even start dealing with a new device. [...] Typically when we train on the first time [...] you should have the operator's manual next to you, which you read before, and then you just follow the manual again in practical terms.* [P5]

Three participants mentioned that the main desired outcome of using the IFU was to achieve a point of familiarization where they no longer needed to use the IFU for assembly or operation. Participants emphasized that this level of expertise was achieved either through intentional practice using the IFU or through a natural result of repetitions over the course of engaging in well-established work routines.

> *I have the trainee go through it again and again and again, so that it [...] becomes muscle memory.* [P4]
>
> *These people are doing the treatments day in, day out. [...] They're doing it frequently, that's the key [to memorisation]. If you only do it twice a year, you forget.* [P3]

Users Generalize Previous Knowledge to Anticipate Potential Problems

As the IFU provides a preview into the medical device system, use of the IFU allows users to efficiently integrate novel information into their existing knowledge about similar types of devices. When asked, all interviewed participants demonstrated extensive expertise and were able to identify several common patterns across extracorporeal systems, including similar components, terminologies, and procedures. These common features allowed participants to identify areas where previous knowledge from one system could be applied to another unfamiliar system with shared features.

> *The knowledge base I gained from [named extracorporeal system] I think is transferable to multiple devices. So the principles are generally the same.* [P4]

Notably, participants' prior experience with similar systems allowed them to anticipate and monitor for commonly occurring problems. This enabled them to formulate generalized assumptions across devices based on shared features and to efficiently prioritize attention toward specific components, assembly procedures, or common errors, even if they have not used a device previously.

> *If it's a peristaltic pump or if it's a centrifugal pump, then you expect different things from it. And I think it's important because you are pre-emptively looking for potential weaknesses or potential sources of error when you encounter the device for the first time.* [P5]

XSL•FO
RenderX

Beyond the problem of patient safety, consolidated knowledge and familiarity when using a new medical device were also essential because participants often had to fend against external stressors in the environment that were likely to otherwise disrupt their performance or concentration. These stressors included issues such as tension, time pressure, and distractions from colleagues and patients, which may lead users to forget key information that may contribute to errors.

The likelihood of distractions, paired with poor or misleading design choices, can therefore mean that users are liable to make mistakes if they are not consolidated enough in their knowledge.

> *You have the pressure of having it ready for the therapy. You have different timings that come into play with transplants.* [P4]

Consolidation of knowledge is important, as decisions made based on the design of the device may be vulnerable to assumptions, and users are therefore liable to make mistakes.

> *So you might think 'OK, [a tubing line] comes from the patient, go to the first point, then the second point, then the third point, fourth point, and then the dialysis [...] but that's not how it works. You actually have to put the line through the blood pump first [...] There is a clue [...] but if somebody's forgotten about that, the line basically looks the same. So it is possible to put it in the wrong way round.* [P3]

Focus on Task Completion May Cause Some Users to Ignore IFUs Completely

Though the utility of the IFU and the information it delivers is widely acknowledged, P2 and P4 both noted, based on their experience observing other HCPs in training, that some users may omit the IFU entirely in favor of working directly on the device. This was largely seen as undesirable, with the use of the IFU generally viewed as an essential tool for reducing the likelihood of user errors.

> *Some people don't like to follow instructions and [...] they'll do it [set up] in an illogical order [...] As long as it's taught properly [...] they generally don't make mistakes.* [P2]

To counter the tendency to omit the IFU at the beginning of training, P4 described how they would develop information delivered in short form to reduce the perceived time and attention required to use the IFU properly, and encourage further engagement from users.

> *What I've noticed when I've given training is people tend not to read operator's manuals and so we created a quick reference guide, so it's an easy to follow, 'do this then do this, then do this.'* [P4]

Summary of Interview Findings

Overall, participants' accounts emphasize that the IFU was an essential tool for initial learning. The IFU supported users in having a clear conceptual understanding of the medical device and to avoid errors particularly during the stages of assembly and operation. However, the role of the IFU was also seen to be supplementary, one that should ultimately be discarded due to convenience and the demands of the work environment. As a result of these external influences, HCPs aimed to consolidate their knowledge enough to be able to perform device operations and assembly correctly without reference to the manual. In practice, a potential issue associated with this is that individuals may ignore the IFU from the outset in order to accelerate this learning process and to try and familiarize themselves directly with the device without supplementary support. This may be because participants ultimately build up a general knowledge of similar devices, which allows them to anticipate problems.

Think-Aloud and Observational Studies

Detailed Information Is Initially Useful, but Counterproductive in Practice

Participants' opinions on the level of detail required for useful instructions shifted from the think-aloud study and the observational study. One recurring example where this debate occurred in the IFU was during a protocol that outlined the assembly process for 2 Luer connections. Luers were connecting points that resembled turn screws, which had to be placed together and turned in a certain direction to ensure that the tubing was securely connected to prevent leakage. Participants described this to be a standard procedure for the assembly of extracorporeal devices, but acknowledged that individuals lacking in expertise may not be familiar with the terms or the process of connecting 2 together. The IFU features an extensive description of how to connect Luers together at each step where Luer connection is required.

During the think-aloud study, participants were generally appreciative of the details and explanatory information incorporated in the Luer protocols, despite their existing familiarity with the steps. P2 and P4 also felt that the details provided were beneficial for minimizing confusion and errors for less experienced individuals than themselves.

> *This is all very obvious for me. I have experience with transducers [...] So it's actually really well-written.[...] For me it's obvious, but for somebody that's not used to pumps or for somebody who's not used to this pump...* [P4]

We did note, however, that P1, who had less practical experience with extracorporeal medical devices, had more difficulty interpreting the text during the think-aloud. He was able to clarify the meaning upon watching the video resources.

> *It's the first part that I'm struggling to visualise what's going on... [...] I don't know. Is it too much writing? [After seeing video] It doesn't seem as complicated as [the text] is making it out to be. I think the fact that it's so wordy so closely might make it seem more complicated than it is.* [P1]

During the observational study, however, feedback on the same Luer connection protocols was mainly negative, with the consensus being that such protocols were unnecessary due to the assumed experience of the target users. P6 found that the articulation of the text often did not clearly correspond to how he would have performed the connection protocol according to his own expertise, and that this would cause him to restart

XSL•FO
RenderX

connection processes in order to check that his interpretations aligned with his experience. When reflecting on the importance of these instructional texts, he stated:

> *A perfusionist [...] does not need to be told 'how to connect a luer lock' [...] Yes, it does need to be there if you pulled somebody off the street and said 'follow this' [...] but you shouldn't need to tell a healthcare professional.* [P6]

Three participants reflected during the observation that the most important information to convey to a reader was the fact that there was only one possible direction that the Luers could turn for a successful connection. In most cases, participants' expertise allowed them to deduce this fact despite this information not being clearly emphasized.

Participants suggested that one way of mitigating this issue was to collate potentially extraneous information to be collected in a separate section of the document (eg, a form of glossary) so that users who wanted the extra information could refer back through the document. Participants emphasized, however, that references to different sections of the document should be clearly identified with page numbers and section numbers so that the information could be easily found.

> *And if you reference other sections of the of the manual [...] I need to know right now where is it? I cannot spend time trying to find it, or going through page by page.* [P5]

HCPs May Prioritize Their Own Routines Over Information in IFUs

During our interviews, participants described how HCPs may neglect to follow IFUs during training, despite the importance of the document. We similarly identified this phenomenon during our own observations. During this part of the study, we found that participants may choose not to follow the IFU to prioritize their own specific routines or familiarity with procedures in related devices.

We noted this again in the Luer connection process, which all participants reported to be familiar with prior to the study. Due to the complexity and the perception that the details of the protocol were unnecessary, we found that users, when performing the outlined Luer connection steps in the IFU, often preferred to carry out the procedures based on their own experience, rather than cross-referencing or double-checking with the IFU. Often, users would prefer to deviate from the IFU or even skip large sections of the IFU. P1, P3, and P6 were often seen carrying out the actions instructed in the IFU (eg, "connect the pressure transducer"), completing the action, only then to find more detailed information about the action that they had just performed and how it should be performed. Often, participants would simply assume that the information in the IFU is redundant or repetitive without using further close scrutiny and skim or even skip sections of the IFU to search for the next step. This process of skimming or skipping occasionally resulted in errors and other relevant steps being skipped as well.

In addition to the above behavior, P3 engaged in additional practices that were not mentioned in the IFU—namely, closing all the clamps on the circuitry tubing once they were assembled. Though this did not have a negative impact on the overall assembly process or the outcome of the assembly or operation of the device, it was nevertheless considered a deviation from the IFU as this was not recommended within the text. When asked about this in the interview, P3 clarified that this practice was encouraged in their specific institution when working with similar extracorporeal devices such as dialysis machines. They carried such practices over and applied them to HepatiCan based on their own judgment of appropriateness and relevance, rather than the information conveyed in the IFU.

> *It's to do with the dialysis equipment, so I'm very scared of air getting into the system. [...] We always make sure that the clamps are closed initially. The manufacturer may not do that, but we would normally clamp everything [...].* [P3]

Images Alone Are Sufficient for Conveying Key Information

Visual representations played a key role in supporting user understanding, both in the think-aloud and in the observation. During the think-aloud, all participants reported using the circuit section diagrams and photos to help them get a sense of the appearance and layout of HepatiCan. However, different images appeared to serve different purposes. Generally, participants described using the circuit section diagrams and their labels for understanding the system flow and identifying specific components and connections. Participants would also refer to the full circuit diagram to understand the relative position of the circuit section they were currently reading about. Photos were frequently used for understanding the relative positions, orientation, state, and appearance of components.

> *I tend to use the line diagrams for referring to the specific component, and the photos I would say are to back up what I'm thinking. So I look at the diagrams – 'that makes sense', and then I look at the photos to see how it's applied.* [P4]

Images played a similar role in practice as participants performed the assembly task in the observational study. Participants were observed frequently setting pages with the relevant section images aside to reference while handling and orienting tubing. Participants often tried to physically manipulate the tubing, using the marked numbers on tubing sections and components and attempting to match them to position numbers marked on the HepatiCan workstation. This behavior aligned with the procedures participants reported mentally visualizing in the assembly procedure in the semistructured interview, when they were just reading the IFU. Participants would frequently switch between reference images and the real-world equipment, such as the tubing circuits and the workstation.

> *Well, I pull the tubing out, identify which end was which end based on the schematic [i.e., circuit section diagram], then I looked at the photo, and identified which end was which end on the Workstation, then I looked at the Workstation, and got started because I knew that was the start point.* [P1]

XSL•FO
RenderX

This study also highlighted the importance of the tubing map on the interior of the HepatiCan workstation wall, which was often used in conjunction with the photos and circuit diagrams. Three participants were also observed relying primarily on the images in the IFU during assembly and only referring to the text to double-check the procedural instructions once components were in place. P6 even suggested creating an instruction manual that was predominantly image-based and using less text, stating that this would improve clarity and reduce the ambiguity of text-based instructions.

> *I found [the wall-map] to actually be quite helpful. But it was obvious, having studied first of all the picture in the [IFU], and then going to it.* [P3]

> *Using that [IFU], I would like to see more detailed pictures and fewer words. It always makes life easier if there are pictures there showing exactly where something should be connected. Then you can remove a lot of the sentences that start creating ambiguity.* [P6]

Indeed, instances where the images of the IFU did not correspond exactly to the state of the HepatiCan device were immediately detected by some participants and led to confusion and hesitation regarding which section they should be working on as they tried to interpret the discrepancy. These events were predominantly caused by instances in the IFU, which used images from a different section of the circuitry to serve as an example to illustrate a general procedure.

Other discrepancies or ambiguities in the images, which sometimes caused participants to ask for clarification, included a lack of scaling of circuit diagrams to real-life tubing, which meant participants sometimes had difficulty identifying corresponding physical features in the real-life tubing and thus the correct orientation of the tubing.

Summary of the Think-Aloud and Observational Studies' Findings

Participants' interactions with the IFU revealed that, while detailed information was considered useful and necessary for novices, it can also be counterproductive for experienced practitioners. Our observations revealed that, for our sample, such detail was often unnecessary, difficult to interpret, and disrupted their workflow in practice. It also led some participants to default to tasks according to their prior routines, sometimes ignoring the IFU altogether. Images, however, consistently played an important role in supporting user interpretation. Most participants responded positively to the use of these images and were even allowed to follow the IFU accurately by depending on images alone. Nevertheless, we also acknowledge that there may be limits to the generalizability of these findings beyond assembly tasks and the context of extracorporeal devices, which we reflect on further in the "Discussion" section below.

Discussion

Principal Findings

In our study, we identify the key role that IFUs play in allowing experienced learners, such as HCPs, to familiarize themselves with the knowledge necessary to operate a complex medical device. We further establish how HCPs use their existing knowledge to prioritize their attention and time for certain tasks or to prepare for specific risks. However, information that is considered too "obvious" or redundant may negatively impact the performance of experienced users by condoning skipping behaviors that can result in users missing relevant information, such as safety information or confusing the sequence of procedures. Both our observations and participants' feedback further suggest that visual instructions alone were largely sufficient for HCPs to identify how to proceed with navigating unfamiliar assembly procedures.

In the following sections, we explore the main design implications for HCP-oriented IFUs, namely, the need for memorization, use of prior knowledge, impact of redundant information, and the benefit of visual information, which we summarize in Table 2. Additionally, we present future directions for research and limitations encountered.

XSL•FO
RenderX

Table . Design implications for HCP[a]-oriented IFUs[b].

Design implication	Summary
Support memorization	• Recognize that users' end goals are to memorize and reduce the need to reference the IFU where possible • Design of medical devices should be intuitive, and allow users to work out procedures with guidance even if IFUs are not available • Using clear labels, color coding, mapping, numbering, etc within the device itself to support user learnability
Identify similarities with other devices and distinguish "novel" information	• Designers should be familiar with common medical devices used by the target group and recognize that HCP users are not learning to use a novel medical device in isolation, but draw from previous knowledge of related devices for common procedures and to anticipate challenges • Instructional design should provide clear cues delineating "common knowledge" and "novel information," so users can more easily identify relevant information
Reduce redundant information	• Avoid redundancies and overrepetition in text and content so users are not encouraged to legitimize behaviors such as skipping, which increases the risk of confusing procedural sequences, delaying learning, and missing relevant information
Recognize the importance of visual information for experienced users	• Provide clear visual aids, diagrams, photos, and pictures of procedure sequences to support understanding, clarity, and ease of referencing during usage • Supports user goals to minimize reliance on IFUs by reducing the need for slower, text-based information processing for HCPs

[a]HCP: health care professional.

[b]IFU: instruction for use.

Support Memorization

We found that participants relied on instruction documents predominantly for memorizing procedures and for engaging in deliberate practice that was key for developing both their skills and knowledge. One of the benefits of memorization may be linked to speed and convenience, which is often important for HCPs, as they must contend with a variety of competing tasks in their work [40]. Indeed, checking reference materials can slow down performance. In the observational study, participants were often observed actively attempting to minimize interactions with the IFU, with behaviors such as relying only or predominantly on images rather than reading the text; attempting to work out their next steps based on their prior knowledge, intuition, or on the tubing map in the workstation interior; and skipping quickly through steps without referring to other sections in the IFU even when instructed to do so in the text. These findings indicate that, where possible, both the system and instructional designs should prioritize learnability, notably memorization.

Our findings further highlight the need for device designs to be as intuitive as possible, in order to reduce cognitive load and the need for memorization in the first place. Participants' positive feedback to the system design of HepatiCan demonstrates potentially useful strategies—for example, clear color coding and number coding, along with a wall map of the circuitry layout—which can minimize the need for user learning. Many participants in the observation were able to simply match up components to the correct location during assembly without directly referring to the IFU, which meant they were able to quickly assemble the device. This can increase the likelihood with which individuals carry out correct procedures in cases where they may fail to recall the relevant knowledge.

Identify Similarities With Other Devices and Distinguish "Novel" Information

Findings from the semistructured interview study suggest that participants were able to apply their existing knowledge across different extracorporeal systems to assist with prioritizing and anticipating potential issues. This is in line with cognitive load theory, outlined above, which suggests that experienced individuals have well-established cognitive processes that can facilitate the processing, memorization, and learning of information more quickly if it relates to the experienced domain, even when the information was originally new or previously unknown to the user [19].

Participants were able to refer to an extensive knowledge base of common patterns across the different medical devices they worked with, including technical knowledge of system components and procedures. This knowledge allowed them to anticipate potential problems and minimize risk by being selective about where and how they directed their attention. In the observational study, participants were also able to use their existing knowledge of Luer connections to deduce the correct procedures without reading the protocol and provided advice on the information that would be most relevant for a user to know. Previous research on the behaviors of expert individuals found that they were also able to fixate more quickly on

XSL•FO
RenderX

task-relevant information because they were more quickly able to discern where they should allocate their attention compared to novices [41-43]. The results from our project consolidate these findings and identify how selective attention strategies translate in real-world contexts to support expert users when learning new domain-related information.

It would therefore be beneficial for developers of complex medical systems to be aware of what components are considered "standard" or "common knowledge" within their field, and what components may be completely novel to users and would require extra guidance. Developers could then present all technical components of a medical system and group them according to what is standard and what is novel. This could not only help users understand how the system operates overall [44], but also allow experienced users to quickly identify the information that is relevant to them.

Reduce Redundant Information

In line with the predictions of the "expertise reversal effect" [24], we observed that more detailed information impeded, rather than supported, performance when users were already familiar with the domain. A key example of this was the repetition of descriptions for recurring protocols, such as Luer connections, which, despite positive feedback during the think-aloud study, were noted to be cumbersome to read and interpret in practice. During feedback in the think-aloud, participants indicated an appreciation for details, but in practice, they were observed engaging in undesirable behaviors such as skimming or even skipping steps in the IFU entirely. In the case of our particular assembly task, as we outline in the "Limitations" section, the consequences of skipping the Luer protocol information were not severe, and errors were usually easily identifiable and recoverable. However, it does highlight the differences between WAI and WAD and provides key insights into some of the conditions that may result in undesirable deviations from WAI.

Importantly, these differences emphasize that it is important for instructional designers not to assume practitioner engagement with IFU. Though in our case, the repetition of the Luer protocol information meant participants were justified in skipping over this information, what was undesirable was the fact that this, in cases, led to users accidentally skipping over genuinely novel or unfamiliar information. Indeed, experts who lapse into automatic or routine behaviors are more vulnerable to errors [26,27,45], and so it is important that designers are thoughtful and careful about repeating information that may appear to legitimize skipping behaviors in practitioners. It should also be acknowledged that participants themselves still indicated an appreciation for extra details and the need to include them to support less experienced individuals. Overall, there were 3 main situations where participants clearly saw the need for additional information (1) to highlight exceptions to the rule, (2) to identify crucial areas for attention, and (3) to inform and clarify genuinely unfamiliar or novel procedures.

Appropriate information delivery in IFUs must therefore strike a balance between providing information that is relevant to users, without giving so much detail that it disrupts comprehension. One way to achieve this balance may be, as some participants suggested in the interview, to provide references that tell users where to find additional information, if deemed necessary. References could either redirect users to previous pages or lead them to a specific glossary or appendix for relevant information. The advantage of this approach is that it can reduce the cognitive load of experienced users, allowing them to focus on information more directly related to their specific task needs, without running the risk of assuming user expertise by excluding information that may be beneficial for less experienced users.

Recognize the Importance of Visual Information for Experienced Users

Visual information was identified as critical for supporting HCPs to complete tasks. This was made explicit during the in-person studies, with some participants completing the assembly tasks by relying predominantly on images. This supports the findings of Kalyuga et al [17], who ran a study that demonstrated how experts could achieve a similar level of performance using only images, compared to novices using both image and text. This finding suggests that not all information, and certainly not all textual information, is required in IFUs for experienced users such as HCPs to achieve adequate performance and to feel confident about completing a task. This further concurs with the existing literature on IFUs for medical devices, highlighting the effectiveness of visual aids for comprehension [46,47].

However, we acknowledge that in many instances, reliance on visual materials alone can be limited beyond assembly tasks, and a degree of reliance on text-based or alternative media of instructional materials is needed to ensure clarity, especially when communicating the rationale behind different procedures the user may be unfamiliar with. Nevertheless, even in text-based materials, we suggest that visual cues in formatting, such as the use of colors, bolding, or formatting (eg, via bullet points for emphasis), could still be important in supporting experienced users to quickly identify relevant information.

One interesting observation within our study was that participants appeared to prefer images that corresponded more closely to the real-life context of the workstation. Abstract representations, such as the circuit section diagrams, were also useful for understanding the orientation and features of a specific tubing section, although participants ultimately preferred to use photos and the wall map to complete the final assembly. This may be because it was easier for participants to "translate" the positions and features in the photos and wall map to their assembly context, meaning less effort was required for interpretation [48].

Limitations

We identify 3 key limitations of this study. The first is the relatively limited number of participants recruited, which impacts the representativeness of the sample and the extent to which our findings may be generalized to other practitioners in the field. This was predominantly due to difficulties in accessing the relevant target end users, a prevalent issue for clinical research with medical practitioners [49]. We navigate this by using a variety of in-depth self-report measures and an

XSL•FO
RenderX

observational study to collect rich, detailed data to inform our findings.

Another issue as a result of the small sample was that 3 out of the 4 participants studied in the observation took part in the think-aloud study 3 weeks prior as well. Thus, by the time of the observation, there may be some degree of familiarity and anticipation of the procedures with the HepatiCan and the IFU, which may affect how participants interact with the device. However, we argue that as there were at least 2 to 3 weeks between the think-aloud and observation, and no participant had seen the IFU prior to the think-aloud, it was unlikely that any of the participants had become sufficiently familiar with the IFU or the procedures for assembling the HepatiCan to have a significant influence on their assembly performance. Nevertheless, a possible avenue for future work could be to more explicitly explore and manipulate features of the IFU on 2 different groups of naïve users, and to compare performance and feedback within each group in order to establish a causal relationship between designs and desired outcomes.

Finally, we recognize that our recommendations for the contents of IFU design do not consider how external factors such as regulations, legal considerations, and risk management may also impact IFU content, as this can differ widely across medical devices, countries, and institutions. Nevertheless, the findings of our study and recommendations highlight the conditions needed for an "ideal" IFU design where the goal is to optimize learning, efficiency in performance, and minimize the effects of negative knowledge transfer for HCPs. Instructional designers and developers should therefore apply individual discretion and evaluation of their own contexts when considering how to balance these recommendations with other legal and regulatory requirements for IFUs.

Future Work

Previous works that have explored user needs for instructional designs (eg, [46,50]) typically explore user performance for smaller-scale medical devices, such as asthma inhalers, or nonmedical contexts. Our study examines user responses to instructional designs for a much more time-consuming and complex medical device assembly task. However, we recognize that there may be limitations in the generalizability of our conclusions to other types of device operation tasks. For example, from a procedural perspective, the assembly of the circuits within the medical device involved many repetitions of similar, simple-to-execute behaviors (ie, connecting tubing through Luers and attaching the tubing to the device), which may explain why users skipped through materials or did not feel the need to pay close attention to the IFU, as mistakes made during these procedures were often easy to notice, address, and were relatively inconsequential at this phase of operation.

Future works could therefore focus more on exploring the issue of attention for HCPs' use of instructions, for example, by testing whether HCPs are more likely to pay close attention to tasks where execution is more complex, and/or where there are greater consequences for patient safety and error recovery, even if the information is already familiar to them. One example of a potentially relevant procedure to study this is device priming, which involves running saline solution through a device in preparation for operation. Priming is a complex task, requiring not only knowledge of the placement and positions of key components (eg, blood warmer and filters), but also domain knowledge regarding the procedures and purpose of priming. This includes knowledge about best practices, monitoring of the priming process, awareness of the status of different components and measures such as pump speed, and the ability to intervene and troubleshoot problems. In such cases, the increase in task complexity and need for monitoring and troubleshooting may mean HCPs are more likely to remain engaged with the IFU. Future work could therefore use the approaches outlined in these studies to explore whether task complexity and higher stakes improve user attention and adherence to the materials.

In addition, we suggest that a focus on providing users with visual cues in instructional design, both for the instructions themselves and for the purpose of locating important information, can be useful. For example, in cases where recommendations in the IFU may differ from participants' own experiences and practices, experienced users may benefit from clearly visually outlined, emphasized, and easy-to-scan bullet points, colors, and presentations explaining developers' justifications for their recommendations, so users could quickly locate and compare practices and motivations underlying different practices. However, further research could explore whether such cues are more appropriate or useful for different types of information and task needs. If different types of information could be identified in order of importance and relevance for users, then they could be explicitly identified and extracted during the design of IFUs.

Another potential area for exploration is to consider how the effects of prior expertise may result in challenges or issues over the course of long-term learning and training. Our user studies provide only a cross-sectional examination of user interactions with the IFU in the early stages of learning, but longitudinal methods using quantitative metrics of performance may be able to provide further insights and comparisons between experienced and nonexperienced users in the formation of specific habits, changes in mental workload, and changes in performance over the course of learning [51]. In the case of the HepatiCan, a direct comparison between HCPs and patient users was not feasible, as the device was intended solely for clinical use. However, we believe this line of research can be particularly fruitful in relation to medical devices designed for both clinical and patient users [13], so that future developers may be better able to clarify the differences in instructional needs for both user groups.

Conclusions

This paper offers important insights into the little-researched field of instructional design for HCPs. Using a combination of interviews and observational methods, we identified that users prioritized memorization and efficiency in task completion. We further identified evidence that the "expertise reversal effect" applies to HCPs in the context of IFU use, and that highly detailed explanations and instructions regarding information already known to the individual can prompt undesirable behaviors such as skimming or skipping. We suggest that developers, in designing IFUs, should focus on identifying

XSL•FO
RenderX

common knowledge constructs and terms within a target user base, so that appropriate emphasis can be placed on information more directly relevant to user needs. We further recommend that developers should emphasize conveying information through visual media such as images and diagrams, which not only communicate key positional information but are also succinct and quick for users to process. Through these findings, we aim to provide guidance for developing medical device instructions that are more intuitive, easy to learn, and easy to apply, to support HCPs better in performing their roles.

Acknowledgments

We would like to acknowledge the support of the research team that developed HepatiCan and the participants in this study.

Funding

This research was funded in whole, or in part, by the Wellcome Trust (203145/Z/16/Z), EPSRC (grant number NS/A000050/1), UCL-HEIF (KEI2020-05-03) award 156780, NIHR (II-LA-0417-20002), and The Liver Group Charity (571348).

Conflicts of Interest

None declared.

References

1. Bitkina O, Kim HK, Park J. Usability and user experience of medical devices: an overview of the current state, analysis methodologies, and future challenges. Int J Ind Ergon 2020 Mar;76:102932. [doi: 10.1016/j.ergon.2020.102932]
2. Kohn LT, Corrigan JM, Donaldson MS. To Err Is Human: Building a Safer Health System: National Academy Press; 2000. URL: https://www.ncbi.nlm.nih.gov/books/NBK225182 [accessed 2026-02-15] [doi: 10.17226/9728]
3. Hegde V. Role of human factors / usability engineering in medical device design. 2013 Presented at: 2013 Annual Reliability and Maintainability Symposium (RAMS); Jan 28-31, 2013; Orlando, FL p. 1-5. [doi: 10.1109/RAMS.2013.6517650]
4. Branaghan RJ. Human factors in medical device design. Crit Care Nurs Clin North Am 2018 Jun;30(2):225-236. [doi: 10.1016/j.cnc.2018.02.005]
5. Iacovides I, Blandford A, Cox A, Back J. How external and internal resources influence user action: the case of infusion devices. Cogn Tech Work 2016 Nov;18(4):793-805. [doi: 10.1007/s10111-016-0392-0]
6. Ward JR, Clarkson PJ. An analysis of medical device-related errors: prevalence and possible solutions. J Med Eng Technol 2004 Jan;28(1):2-21 [FREE Full text] [doi: 10.1080/03091900310001237 47]
7. HE75:2025 human factors engineering - design of medical devices. : ANSI/AAMI; 2025 URL: https://webstore.ansi.org/preview-pages/AAMI/preview_ANSI+AAMI+HE75-2009+(R2018).pdf [accessed 2026-02-27]
8. IEC 62366-1:2015. medical devices—part 1: application of usability engineering to medical devices. ISO. 2015. URL: https://www.iso.org/standard/63179.html [accessed 2026-02-15]
9. Geissler N, Byrnes T, Lauer W, et al. Patient safety related to the use of medical devices: a review and investigation of the current status in the medical device industry. Biomed Tech (Berl) 2013 Feb;58(1):67-78. [doi: 10.1515/bmt-2012-0040] [Medline: 23314500]
10. Opie J, McCann E, Qureshi H, et al. An academician's approach to the application of human factors standards: a case study on a liver support system. Hum Factors Healthcare 2024 Jun;5:100070. [doi: 10.1016/j.hfh.2024.100070]
11. Bowers S. Doctors demand to see evidence on safety of medical devices approved in Europe. BMJ 2018 Nov 30;363:k5105. [doi: 10.1136/bmj.k5105] [Medline: 30504128]
12. Feinmann J. Medical devices: new EU regulations won't guarantee safe design, doctors warn. BMJ 2019 Jun 27;365:l4446. [doi: 10.1136/bmj.l4446] [Medline: 31248858]
13. Candidori S, Dozio N, Osouli K, et al. Improving maternal safety: usability and performance assessment of a new medical device for the treatment of postpartum haemorrhage. Appl Ergon 2024 May;117:104223. [doi: 10.1016/j.apergo.2023.104223] [Medline: 38219376]
14. Coldewey B, Diruf A, Röhrig R, Lipprandt M. Causes of use errors in ventilation devices - systematic review. Appl Ergon 2022 Jan;98:103544. [doi: 10.1016/j.apergo.2021.103544] [Medline: 34461508]
15. Lyons I, Blandford A. Safer healthcare at home: detecting, correcting and learning from incidents involving infusion devices. Appl Ergon 2018 Feb;67:104-114. [doi: 10.1016/j.apergo.2017.09.010] [Medline: 29122181]
16. Rajkomar A, Farrington K, Mayer A, Walker D, Blandford A. Patients' and carers' experiences of interacting with home haemodialysis technology: implications for quality and safety. BMC Nephrol 2014 Dec 11;15:195. [doi: 10.1186/1471-2369-15-195] [Medline: 25495826]
17. Kalyuga S, Chandler P, Sweller J. Levels of expertise and instructional design. Hum Factors 1998 Mar;40(1):1-17. [doi: 10.1518/001872098779480587]
18. Sweller J, van Merriënboer JJG, Paas F. Cognitive architecture and instructional design: 20 years later. Educ Psychol Rev 2019 Jun;31(2):261-292. [doi: 10.1007/s10648-019-09465-5]

XSL•FO
RenderX

19. Sweller J, van Merrienboer JJG, Paas F. Cognitive architecture and instructional design. Educ Psychol Rev 1998 Sep;10(3):251-296. [doi: 10.1023/A:1022193728205]
20. Baddeley A. Working memory. Science 1992 Jan 31;255(5044):556-559. [doi: 10.1126/science.1736359] [Medline: 1736359]
21. Klepsch M, Seufert T. Understanding instructional design effects by differentiated measurement of intrinsic, extraneous, and germane cognitive load. Instr Sci 2020 Feb;48(1):45-77. [doi: 10.1007/s11251-020-09502-9]
22. Kalyuga S. Expertise reversal effect and its implications for learner-tailored instruction. Educ Psychol Rev 2007 Oct 9;19(4):509-539. [doi: 10.1007/s10648-007-9054-3]
23. Chen O, Kalyuga S, Sweller J. The expertise reversal effect is a variant of the more general element interactivity effect. Educ Psychol Rev 2017 Jun;29(2):393-405. [doi: 10.1007/s10648-016-9359-1]
24. Kalyuga S, Renkl A. Expertise reversal effect and its instructional implications: introduction to the special issue. Instr Sci 2010 May;38(3):209-215. [doi: 10.1007/s11251-009-9102-0]
25. Brewer WF. Bartlett's concept of the schema and its impact on theories of knowledge representation in contemporary cognitive psychology. In: Saito A, editor. Bartlett, Culture and Cognition: Psychology Press; 2000:69-89.
26. Besnard D, Cacitti L. Interface changes causing accidents. An empirical study of negative transfer. Int J Hum Comput Stud 2005 Jan;62(1):105-125. [doi: 10.1016/j.ijhcs.2004.08.002]
27. Finstad K. Analogical problem solving in casual and experienced users: when interface consistency leads to inappropriate transfer. Hum–Comput Interact 2008 Dec 5;23(4):381-405. [doi: 10.1080/07370020802532734]
28. Chaniaud N, Métayer N, Megalakaki O, Loup-Escande E. Effect of prior health knowledge on the usability of two home medical devices: usability study. JMIR Mhealth Uhealth 2020 Sep 21;8(9):e17983. [doi: 10.2196/17983] [Medline: 32955454]
29. Rego C, Montague E. The impact of feedback modalities and the influence of cognitive load on interpersonal communication in nonclinical settings: experimental study design. JMIR Hum Factors 2023 Oct 5;10:e49675. [doi: 10.2196/49675] [Medline: 37796596]
30. Bailey R. Applying design principles to instructional materials. In: Applied Human Factors in Medical Device Design: Elsevier; 2019:139-162. [doi: 10.1016/B978-0-12-816163-0.00009-8]
31. Kalsher MJ, Wogalter MS. Warnings and instructions: design factors for medical devices and systems. Proc Hum Factors Ergon Soc Ann Meet 2000 Jul;44(28):537-540. [doi: 10.1177/154193120004402833]
32. Wogalter MS, Conzola VC, Smith-Jackson TL. Research-based guidelines for warning design and evaluation. Appl Ergon 2002 May;33(3):219-230 [FREE Full text] [doi: 10.1016/s0003-6870(02)00009-1] [Medline: 12164506]
33. Privitera MB, Evans M, Southee D. Human factors in the design of medical devices - approaches to meeting international standards in the European Union and USA. Appl Ergon 2017 Mar;59(Pt A):251-263. [doi: 10.1016/j.apergo.2016.08.034] [Medline: 27890135]
34. Sowan AK, Gomez TM, Tarriela AF, Reed CC, Paper BM. Changes in default alarm settings and standard in-service are insufficient to improve alarm fatigue in an intensive care unit: a pilot project. JMIR Hum Factors 2016 Jan 11;3(1):e1. [doi: 10.2196/humanfactors.5098] [Medline: 27036170]
35. Sowan AK, Tarriela AF, Gomez TM, Reed CC, Rapp KM. Nurses' perceptions and practices toward clinical alarms in a transplant cardiac intensive care unit: exploring key issues leading to alarm fatigue. JMIR Hum Factors 2015 Mar 16;2(1):e3. [doi: 10.2196/humanfactors.4196] [Medline: 27025940]
36. Blandford A, Furniss D, Vincent C. Patient safety and interactive medical devices: realigning work as imagined and work as done. Clin Risk 2014 Sep;20(5):107-110. [doi: 10.1177/1356262214556550] [Medline: 25866466]
37. Hollnagel E, Clay-Williams R. Work-as-imagined and work- as-done. In: Implement Sci: Routledge; 2022. [doi: 10.4324/9781003109945-52]
38. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
39. Guion LA, Diehl DC, McDonald D. Triangulation: establishing the validity of qualitative studies. EDIS 2011;2011(8):3. [doi: 10.32473/edis-fy394-2011]
40. Alkahtani M, Aziz T, Ahmad A, Darmoul S. Multitasking in healthcare systems. ProQuest. 2015. URL: https://www.proquest.com/docview/1792022789?pq-origsite=gscholar&fromopenview=true&sourcetype=Scholarly%20Journals [accessed 2026-02-15]
41. Abernethy B. Visual search in sport and ergonomics: its relationship to selective attention and performer expertise. Hum Perform 1988 Dec;1(4):205-235. [doi: 10.1207/s15327043hup0104_1]
42. Gegenfurtner A, Lehtinen E, Säljö R. Expertise differences in the comprehension of visualizations: a meta-analysis of eye-tracking research in professional domains. Educ Psychol Rev 2011 Dec;23(4):523-552. [doi: 10.1007/s10648-011-9174-7]
43. Yu C, Zhong Y, Fricker D. Selective attention in cross-situational statistical learning: evidence from eye tracking. Front Psychol 2012;3:148 [FREE Full text] [doi: 10.3389/fpsyg.2012.00148] [Medline: 22712020]
44. Kieras DE, Bovair S. The role of a mental model in learning to operate a device. Cogn Sci 1984 Sep;8(3):255-273. [doi: 10.1016/S0364-0213(84)80003-8]
45. Toner J, Montero BG, Moran A. The perils of automaticity. Rev Gen Psychol 2015 Dec;19(4):431-442. [doi: 10.1037/gpr0000054]

XSL•FO
RenderX

46. Kools M, van de Wiel MWJ, Ruiter RAC, Kok G. Pictures and text in instructions for medical devices: effects on recall and actual performance. Patient Educ Couns 2006 Dec;64(1-3):104-111. [doi: 10.1016/j.pec.2005.12.003] [Medline: 16472960]
47. Lee K, Nathan-Roberts D. Using visual aids to supplement medical instructions, health education, and medical device instructions. Proc Int Symp Hum Factors Ergon Healthc 2021 Jun;10(1):257-262. [doi: 10.1177/2327857921101190]
48. Tyner BC, Fienup DM. A comparison of video modeling, text-based instruction, and no instruction for creating multiple baseline graphs in Microsoft Excel. J Appl Behav Anal 2015 Sep;48(3):701-706 [FREE Full text] [doi: 10.1002/jaba.223] [Medline: 26173573]
49. Li Y, Oladimeji P, Monroy C, et al. Design of interactive medical devices: feedback and its improvement. 2011 Presented at: 2011 IEEE International Symposium on IT in Medicine and Education; Dec 9-11, 2011. [doi: 10.1109/ITiME.2011.6132022]
50. Irrazabal N, Saux G, Burin D. Procedural multimedia presentations: the effects of working memory and task complexity on instruction time and assembly accuracy. Appl Cogn Psychol 2016 Nov;30(6):1052-1060 [FREE Full text] [doi: 10.1002/acp.3299]
51. Schmettow M, Schnittker R, Schraagen JM. An extended protocol for usability validation of medical devices: research design and reference model. J Biomed Inform 2017 May;69:99-114. [doi: 10.1016/j.jbi.2017.03.010] [Medline: 28336478]

Abbreviations

HCP: health care professional
IFU: instructions for use
WAD: work as done
WAI: work as imagined

Edited by A Parush; submitted 14.Aug.2025; peer-reviewed by D Lyell, O Nguyen; revised version received 07.Jan.2026; accepted 22.Jan.2026; published 06.Mar.2026.

Please cite as:
Xia Q(, Opie J, Brookshaw T, Barbarin EE, Blandford A, Selden C
Designing for Knowledge Generalization in Medical Device Instructions Among Health Care Professionals: Qualitative Interview and Observation Study
JMIR Hum Factors 2026;13:e82405
URL: https://humanfactors.jmir.org/2026/1/e82405
doi:10.2196/82405

© Qing (Nancy) Xia, Jeremy Opie, Tom Brookshaw, Eloy Erro Barbarin, Ann Blandford, Clare Selden. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 6.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Assistive Technologies for Pulse-Based Tele-Practice in Traditional Chinese Medicine: Two-Phase, Dual-Method Feasibility and Pilot Study

Jia Yin Ruan[1,2], PhD; Shucheng Chen[1], PhD; Minru Wu[1], MSc; Shuhan Li[1], MSc; Yuan Shan Ho[1], PhD; Ka Lun Fan[3], PhD; Wing Leung Chow[4], MPhil; Leung Chiu[4], PhD; Wing Fai Yeung[1,5], PhD

[1]School of Nursing, the Hong Kong Polytechnic University, GH506, Hung Hom, Kowloon, Hong Kong SAR, China (Hong Kong)

[2]New York University Rory Meyers College of Nursing, New York, NY, United States

[3]Department of Applied Data Science, Hong Kong Shue Yan University, Hong Kong SAR, China (Hong Kong)

[4]Logistics and Supply Chain MultiTech R&D Centre, Hong Kong SAR, China (Hong Kong)

[5]Research Centre for Chinese Medicine Innovation, Hong Kong SAR, China (Hong Kong)

Corresponding Author:
Wing Fai Yeung, PhD
School of Nursing, the Hong Kong Polytechnic University, GH506, Hung Hom, Kowloon, Hong Kong SAR, China (Hong Kong)

Abstract

Background: Pulse palpation is essential for accurate traditional Chinese medicine diagnosis. However, this tactile-sensory-dependent technique is not feasible in teleconference, leaving traditional Chinese medicine underserved by conventional tele-practice frameworks. To address this, we developed an Assistive Pulse Data Collection (APDC; Logistics and Supply Chain MultiTech R&D Centre) device.

Objective: This study aimed to evaluate feasibility and to preliminarily examine the Chinese medicine practitioners (CMPs) agreement on real-person pulses and machine-generated pulses and collect users' feedback.

Methods: Following World Health Organization guidelines for piloting new medical technologies, a 2-phase, dual-method evaluation was conducted. In the feasibility phase, 10 participants' pulses were recorded using the APDC device. Overall, 5 CMPs evaluated the participants' and machine-generated pulses using a 5-point Likert scale (1="Strongly Disagree" to 5="Strongly Agree"). In the pilot phase, 10 CMPs assessed pulses from 42 participants and refined the regenerated outputs for comparison. Quantitative and qualitative feedback from CMPs and participants was collected.

Results: CMPs evaluated machine-regenerated pulses across 4 parameters: frequency (mean agreement score 4.1, SD 0.6), rhythm (mean 3.8, SD 0.6), width (mean 2.7, SD 0.9), and force (mean 2.4, SD 1), indicating moderate-to-high agreement for frequency or rhythm but lower consensus for width and force. Following device refinements (eg, adjustable armrest, pressure calibration, and pulse algorithms), the pilot phase CMPs' agreement improved for frequency (mean 4.3, SD 0.7) and rhythm (mean 4, SD 0.8), while width (mean 3, SD 0.9) and force (mean 2.8, SD 0.9) remained suboptimal. CMPs reported enhanced clarity (mean 4.3, SD 1) but persistent inefficiency (mean 2.5, SD 0.5) and neutral satisfaction (mean 2.6, SD 0.5). Participants maintained favorable ratings for comfort (mean 4, SD 0.7), easy to use (mean 3.9, SD 0.8), and high satisfaction (mean 4, SD 0.7).

Conclusions: The APDC device is feasible and enables CMPs to take remote pulse feature assessment, but further optimization of pulse force and width is warranted.

Trial Registration: ClinicalTrials.gov NCT06974227; https://clinicaltrials.gov/study/NCT06974227

(JMIR Hum Factors 2026;13:e77740) doi:10.2196/77740

KEYWORDS

traditional Chinese medicine; tele-practice; pulse sensing; pulse regeneration; wireless network

Introduction

The COVID-19 pandemic has profoundly disrupted global health care systems, compelling a rapid shift toward tele-practice to mitigate barriers to in-person care [1,2]. While virtual consultations have proven effective for many medical disciplines, their application faces significant challenges in practices requiring tactile diagnostic methods, such as traditional Chinese medicine (TCM) [3]. A major barrier to tele-practice in TCM was the inability to conduct physical examinations, particularly pulse examinations on patients [4]. In TCM, pulse palpation is a diagnostic cornerstone where practitioners assess

XSL•FO
RenderX

the depth, rhythm, and waveform of the radial artery to identify pathological patterns [5] and treatment formulation [6-8], which is irreplaceable for accurate TCM pattern diagnosis [9]. This tactile-sensory–dependent technique is not feasible in teleconference, leaving TCM underserved by conventional tele-practice frameworks. To bridge this gap, innovative solutions are urgently needed to replicate tactile diagnostic interactions remotely. Emerging research proposes robotic systems capable of capturing, quantifying, and recreating pulse waveforms through haptic interfaces, thereby enabling a remote pulse diagnosis [9]. Such technology could not only address pandemic-induced limitations but also expand access to TCM care in underserved regions, underscoring its transformative potential for both tele-practice and traditional practice.

Various sensor technologies have been used for the noninvasive monitoring of radial artery pulses. Currently, the main types of pulse sensors include pressure sensors, photoelectric sensors, and ultrasonic Doppler sensors. Each type has its distinct mechanisms and characteristics. Pressure sensors directly measure the radial artery pulse using piezoelectric, piezoresistive, or piezocapacitive sensors: piezoelectric sensors offer high repeatability and heat resistance but cannot detect static pressure [10-12]; piezoresistive sensors, known for high sensitivity and mature technology, are the most widely used [13,14]; piezocapacitive sensors provide accurate dynamic measurements but exhibit nonlinear characteristics [15,16]. In contrast, photoelectric sensors detect variations in light reflected from blood vessels, converting them into electrical signals [17,18]. While they are less affected by interference, they require a radiation source that complicates the circuitry. In contrast, ultrasonic Doppler sensors analyze reflected ultrasound waves to derive pulse signals from changes in blood flow speed and vessel volume, offering rich data but at a higher cost than the other types [19,20]. All 3 devices primarily focus on detecting and digitizing pulse data. However, in clinical practice, these digitized pulse readings are not easily interpretable by Chinese medicine practitioners (CMPs). Therefore, a device that can convert the collected pulse data back into a pulse format resembling human pulse features that is suitable for CMPs' assessment is critically important for clinical practice.

To fill this critical gap in TCM tele-practice, we developed an Assistive Pulse Data Collection (APDC) device (Logistics and Supply Chain MultiTech R&D Centre, Hong Kong SAR). The APDC device integrates a high-fidelity pulse sensor to capture radial artery waveforms, digitizes key pulse features (eg, rhythm, frequency, and amplitude), and transmits data wirelessly to a robotic regeneration interface. Unlike conventional telemedicine platforms that are limited to audiovisual communication, this system enables TCM practitioners to remotely palpate and interpret pulse waveforms through haptic feedback, thereby replicating the tactile dimension essential for accurate pattern diagnosis. By bridging the sensory disconnect inherent in virtual consultations, the APDC device has the potential to enhance diagnostic precision and expand access to TCM care in remote areas and pandemic-restricted settings.

Validating such a device is imperative to ensure its accuracy, reliability, and safety for clinical use. It is essential to confirm that the device can accurately capture and reproduce the pulse features, as any discrepancies could lead to misdiagnosis and inappropriate treatment. We, therefore, conduct this feasibility and pilot study to evaluate the feasibility and to preliminarily examine the CMPs' agreement on real-person pulses and machine-generated pulses, and to collect CMPs' and participants' feedback on the device. This dual-method evaluation aligns with World Health Organization guidelines for piloting novel medical technologies, ensuring both technical accuracy and user experience are prioritized [21].

Methods

Study Design

The study consisted of 2 phases, namely the feasibility phase (phase 1) and the pilot phase (phase 2). The reporting of this study adhered to the guidelines of the CONSORT (Consolidated Standards of Reporting Trials) statement: extension to randomized pilot and feasibility trials [22].

Phase 1: Feasibility

This phase aimed to preliminarily assess the feasibility of the APDC device in collecting participants' pulse features and regenerating the pulse for CMPs. CMP's rating of the agreement on the likeliness between participants' and regenerated pulse was also preliminarily assessed. User experiences from both CMPs and participants were collected for the team to refine the setting of the APDC Device. The CMPs and participants were recruited from 20 to 28 November 2024.

Procedure

The study procedure was conducted in the following steps:

1. The participant's right-hand pulse features were recorded using the APDC Device.
2. Five CMPs assessed the participant's pulse and then the pulse regenerated from the APDC Device.
3. The CMPs rated the agreement, using a predefined record form on the likeliness between the participants' pulse and APDC-regenerated pulse, in the respects of frequency, rhythm, wideness, and force (5-point Likert for each domain, from "Strongly disagree" to "Strongly agree").
4. The CMPs' and participants' experiences were assessed using a questionnaire.

Phase 2: Pilot

In the pilot phase, the APDC device was refined by the team using the same procedure as above. Details of the refinement can be found in our previous publication regarding the technical development of the ADPC [23], while this study focused on the users' evaluation by participants and CMPs. A total of 42 participants' pulse features were collected, and 10 CMPs each assessed 10 - 11 participants' pulse (a single participant's pulse was assessed by 2 to 3 CMPs) and APDC-regenerated pulse and rated the agreement on the likelihood. They were asked to provide their feedback. The CMPs and participants were recruited from March 1-14, 2025.

Participants Recruitment

The study included participants who met specific inclusion criteria: individuals aged between 18 and 65 years, with a BMI

XSL•FO
RenderX

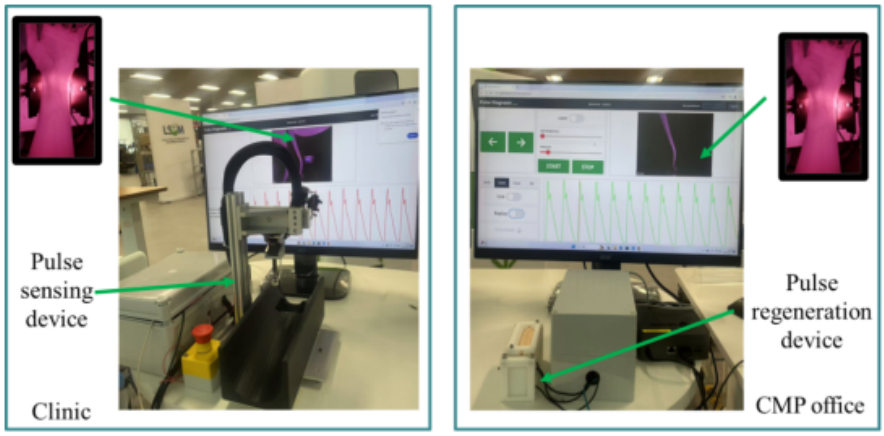
ranging from 18.5 to 30 kg/m², and those willing to provide informed consent. Participants were excluded if they had a dorsally located radial artery.

We recruited registered CMPs with at least 5 years of clinical experience from social media to assess the participants' pulse and the machine-regenerated pulse. Afterward, they completed a predefined record form to evaluate the pulse characteristics.

Tested Device

A prototype of the APDC device (Logistics and Supply Chain MultiTech R&D Centre) was developed using a standard tele-practice infrastructure as an assistive technology for TCM pulse diagnosis (Figure 1). The system comprises 2 primary components: a pulse-sensing device at the clinic and a pulse-regeneration device at the TCM practitioner's office. A pulse collection device was used at the patient's location, featuring a screen to display pulse information. Correspondingly, at the CMP end, a pulse-regeneration device equipped with a screen displayed the transmitted data. Patients can place their arms on a pulse-sensing device. CMP can remotely control the device to adjust its position and precisely determine the optimal location for pulse sensing, as the accuracy of pulse diagnosis is highly dependent on proper positioning. Both modules feature digital interfaces for real-time data visualization, with the sensing unit displaying pulse data acquisition parameters and the regeneration unit showing the transmitted pulse characteristics. The sensing unit incorporates position-adjustable mechanisms that can be remotely controlled by TCM practitioners. This bidirectional setup enables patients to position their arms on the sensing platform, while allowing practitioners to optimize measurement locations remotely.

Figure 1. Overview of the proposed traditional Chinese medicine pulse–based tele-practice system. CMP: Chinese medicine practitioner.



Pulse Sensing Device

Palpation is a professional procedure used in TCM. Practitioners determine the palpation location on the participant's arm and the applied force based on their expertise and experience [24]. The proposed pulse-sensing device features 2 key components: a radial artery viewing device and a dynamic force control, which assist practitioners in accurately determining the palpation location and the appropriate force to apply.

Radial Artery Viewing Device

The APDC system uses infrared technology to illuminate a patient's arm and capture images using an infrared camera, as shown in Figure 2. Unlike the conventional blood vessel viewing device [25], the participants' arm is positioned on a stabilizing device with an infrared light source beneath it and an infrared camera to capture the images shown in Figure 3.

XSL•FO
RenderX

Figure 2. Infrared image of a participant's arm showing the position of radial artery. LED: light-emitting diode.

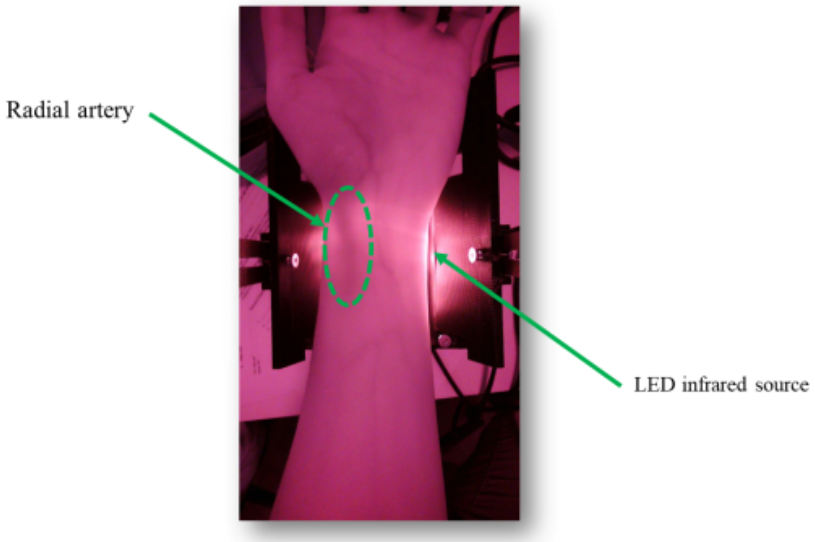
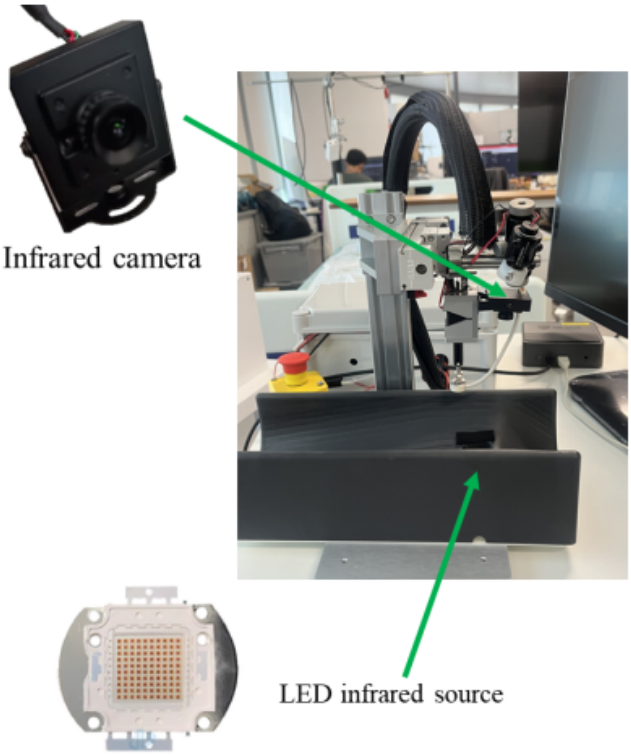


Figure 3. Key components of radial artery viewing device. LED: light-emitting diode.



Pulse-Taking Procedure

The CMPs evaluated the participants' pulse as they would in a clinical setting. The principal investigator assessed each participant's pulse and used a marker pen to mark the radial pulse (cun, according to TCM theory). The technician then assessed the participant's radial pulse, using the principal investigator's mark as a guide, with the aid of infrared light, and began recording the pulse waveform data using the APDC device. The device recorded the participant's pulse for 30 seconds at the wrist (*cun*, *guan*, and chi positions in terms of TCM) individually during the feasibility phase, and simultaneously during the pilot phase, after being upgraded to include 3 probes instead of a single probe based on feedback from CMPs during the feasibility phase.

After the APDC device recorded the pulse, the participant was seated on a chair with their arm resting on a table and their wrist positioned on a pulse pillow. The CMP assessed the pulse of the participant's right or left hand for up to 1 minute each, as

XSL•FO
RenderX

if in clinical practice, and recorded their pulse description on a chart. Subsequently, they assessed the regenerated pulse produced by the APDC device for up to 1 minute. Finally, they rated the agreement between the real participant's pulse and the machine-generated pulse.

Outcome Assessment

The user experience with the APDC device, which comprised a pulse collection device and a pulse re-generation device, was evaluated separately. Quantitative feedback from both participants and CMPs was gathered through a self-developed questionnaire. Qualitative feedback was collected through open-ended questions.

Agreement Between Participants' and Regenerated Pulse

CMPs were instructed to indicate their level of agreement with statements regarding the similarity between the machine-generated pulse and the actual pulse of participants. Responses were recorded using a 5-point Likert scale, ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). The questionnaire evaluated 4 key pulse attributes: frequency, rhythm, wideness, and force. CMPs were asked to assess the extent to which they agreed that the frequency of the machine-regenerated pulse accurately resembled that of a real person's pulse. Similarly, they evaluated the rhythm, wideness, and force of the machine-generated pulse in comparison to the actual pulse. This structured approach allowed for a comprehensive analysis of CMPs' experiences and perceptions, contributing valuable data to the study of machine-regenerated pulse accuracy.

CMPs' Feedback on the Device

A self-developed questionnaire was used to evaluate CMPs' comfort, ease of use, efficiency, and overall satisfaction with the machine. CMPs were asked to indicate their level of agreement with each statement using a 5-point scale. First, CMPs rated their comfort level during the measurement process, with options ranging from "Very uncomfortable" to "Very comfortable." Second, they assessed the ease of understanding the instructions and using the machine, with responses from "Very difficult" to "Very easy." Third, CMPs evaluated the efficiency of the measurement process, indicating their agreement with statements about its quickness and effectiveness, from "Strongly disagree" to "Strongly agree." Finally, they expressed their overall satisfaction with the experience, choosing from "Very dissatisfied" to "Very satisfied."

Additionally, qualitative feedback was collected through open-ended questions in a self-completed questionnaire, seeking insights on "How would you describe the overall experience of using the pulse regenerating device?" and "What were the most noticeable differences in terms of the pulse features between the machine-regenerated pulse and that of the participants?" and "If you could suggest improvements to the machine, what would they be?"

Participants' Feedback on the Device

Participants' feedback on their experience with the pulse measurement process was collected. Participants were asked to indicate their level of agreement with various statements using a 5-point Likert scale. The questionnaire focused on 4 key areas: comfort during the measurement process, ease of understanding and using the machine, the efficiency of the measurement process, and overall satisfaction with the experience. It aimed to assess the user-friendliness and effectiveness of the machine from the Participants' perspective.

Finally, qualitative feedback was gathered through open-ended questions in a self-completed questionnaire. Participants were asked to describe their overall experience with the pulse measuring device, suggest any improvements they would recommend for the machine, and share any additional insights or aspects of their experience that were not addressed in the previous questions. The questions used were: "How would you describe the overall experience of using the pulse-measuring device?" "If you could suggest improvements to the machine, what would they be?" and "Is there anything else you would like to share about your experience or any other aspect that was not covered in the previous questions?"

Statistical Analyses

All analyses were performed using SPSS (version 28.0; IBM Corp). The CMPs' and participants' sociodemographic characteristics, including age and gender, presence of chronic diseases, and the CMPs' rating on the likeliness between participants' pulse features (frequency, rhythm, wideness, and force) and machine-regenerated pulse using descriptive statistics. Both parametric measures (mean [SD]) or nonparametric measures (median [IQR]) were reported accordingly due to the small sample size. Categorical variables (eg, gender and presence of chronic diseases) were expressed as frequencies and percentages (n, %).

Ethical Considerations

The study was reviewed and approved by the Hong Kong Polytechnic University institutional review board (reference number HSEARS20240801005) and registered at ClinicalTrials.gov (Ref. NCT06974227). We recruited participants through social media and mass emails at the University. All participants signed a consent form to confirm their voluntary participation in pulse data collection. As an incentive, each participant received an HK $100 supermarket coupon (equivalent of US $12.81) upon completing either phase 1 or 2. Pulse recordings and associated demographic information were deidentified immediately after collection, and any potentially identifying information was removed or coded to safeguard participant privacy. All data were stored on encrypted, password-protected institutional servers accessible only to authorized research team members, and findings are presented in aggregate form to maintain confidentiality.

Results

Feasibility

During the feasibility phase, 5 CMPs and 10 participants were recruited. The 5 CMPs had a mean age of 36 (SD 6.2) years, 3 (60%) were female, and had a mean clinical experience of 11.2 (SD 6.3) years (Table 1). The 10 participants had a mean age

XSL•FO
RenderX

of 41.1 (SD 15.2) years, 1 female (10%), and a mean BMI of 23.7 (SD 3.2) kg/m^2 (Table 2).

Table . Chinese medicine practitioners' characteristics and rating.

Variables	CMP[a] (feasibility phase; n=5)	CMP (pilot phase; n=10)
Age (years)		
Mean (SD)	36.0 (6.2)	41.5 (5.9)
Median (IQR)	34 (9)	43 (7.5)
Sex, n (%)		
Female	3 (60)	6 (60)
Education level, n (%)		
Tertiary education	0 (0)	1 (10)
Master's degree	4 (80)	5 (50)
Doctoral degree	1 (20)	4 (40)
Clinical experience		
Mean (SD)	11.2 (6.3)	16.0 (5.8)
Median (IQR)	10 (12)	19 (7.5)
Agreement on Machine-regenerated Pulse (5=Totally agreement, 1=Totally disagreement)		
Frequency		
Mean (SD)	4.1 (0.58)	4.3 (0.72)
Median (IQR)	4 (0)	4 (1)
Rhythm		
Mean (SD)	3.8 (0.58)	4.0 (0.82)
Median (IQR)	4 (0)	4 (0.25)
Pulse width		
Mean (SD)	2.7 (0.88)	3.0 (0.94)
Median (IQR)	3 (1)	3 (2)
Pulse force		
Mean (SD)	2.4 (0.95)	2.8 (0.94)
Median (IQR)	2 (1)	3 (1)
Experience with Machine-generated Pulse Measurement (5=Totally agreement, 1=Totally disagreement)		
Comfort level		
Mean (SD)	3 (0.71)	3 (0.82)
Median (IQR)	3 (1)	3 (2)
Clarity of instructions on using the machine		
Mean (SD)	3.8 (1.30)	4.3 (0.95)
Median (IQR)	4 (2.5)	4.5 (1)
Efficiency		
Mean (SD)	2.2 (0.45)	2.5 (0.53)
Median (IQR)	2 (0.5)	2.5 (1)
Overall satisfaction		
Mean (SD)	2.6 (0.55)	2.6 (0.52)
Median (IQR)	3 (1)	3 (1)

[a]CMP: Chinese medicine practitioner.

XSL•FO
RenderX

Table . Participants' characteristics and rating.

Variables	Participants (feasibility phase; n=10)	Participants (pilot phase; n=42)
Age (years)		
Mean (SD)	41.1 (15.2)	41.6 (13.3)
Median (IQR)	31.5 (31)	38.5 (22.75)
Sex, n (%)		
Female	1 (10)	28 (67)
BMI (kg/m^2)		
Mean (SD)	23.66 (3.2)	21.6 (2.1)
Median (IQR)	23.96 (3.76)	21.48 (2.07)
Participants' experience with the device (score range 1 - 5), (5=Totally agreement, 1=Totally disagreement)		
Comfort level		
Mean (SD)	4.1 (0.74)	4.0 (0.66)
Median (IQR)	4 (0.25)	4 (0.25)
Ease of using		
Mean (SD)	4.1 (0.74)	3.9 (0.82)
Median (IQR)	4 (0.25)	4 (0.25)
Efficiency		
Mean (SD)	3.8 (0.92)	3.6 (0.85)
Median (IQR)	4 (1.25)	4 (1)
Overall satisfaction		
Mean (SD)	4.1 (0.74)	4 (0.68)
Median (IQR)	4 (1.25)	4 (1)

CMPs' Rating on Machine-Regenerated Pulse at Feasibility Phase

Five CMPs each evaluated pulses of 10 participants and the participants' machine-generated pulses in terms of frequency, rhythm, width, and force. The results indicated an average agreement score of 4.1 (SD 0.6) for "frequency," suggesting consensus among CMPs that the machine's pulse frequency was consistent with the participants' pulses. The "rhythm" dimension obtained an average score of 3.8 (SD 0.6), demonstrating that CMPs tended to rate this parameter with moderate agreement. "Pulse width" received a lower average score of 2.7 (SD 0.9), reflecting CMPs' neutral to general agreement regarding this characteristic. The lowest average score of 2.4 (SD 1) was observed for "pulse force," indicating a slight leaning toward disagreement among the CMPs.

CMPs' Experience With Machine-Generated Pulse Measurement at Feasibility Phase

CMPs reported their "comfort level" during the measurement process with an average score of 3 (SD 0.7), indicating a neutral perception. The "clarity of instructions on using the machine" received an average score of 3.8 (SD 1.3), demonstrating relatively straightforward operational procedures. Regarding "efficiency," the significantly lower average score of 2.2 (SD 0.5) reflected moderate dissatisfaction among CMPs at the feasibility phase. Overall satisfaction with the measurement system yielded an average score of 2.6 (SD 0.6), suggesting a neutral appraisal of the machine-generated pulse assessments.

CMPs' Qualitative Feedback at the Feasibility Phase

Overall Perception

Performance in Pulse Reproduction

CMPs generally acknowledged that machines can reproduce pulse rate and rhythm with relative accuracy. However, simulations of pulse width and strength were suboptimal, prone to misjudgment, and occasionally exhibited asynchronous performance across the 3 pulse locations (cun, guan, and chi). This is particularly true in complex pulse characteristics during regeneration, inconsistencies in pulse strength variations, and notable deficiencies in detailed feature representation.

Usability Versus Diagnostic Precision

The CMPs deemed the device was user-friendly, although the devices demonstrated compromised overall accuracy, particularly in discerning subtle variations in pulse position intensity.

CMPs' Suggestion for Improvement After the Feasibility Phase

Temporal Continuity and Nuance Preservation

The machine-generated pulse showed discontinuities during pulse cycle transitions (eg, last-to-first beat replay artifacts),

XSL•FO
RenderX

which required mitigation through extended sampling durations (>60 s) to minimize transitional irregularities.

Interactive Visualization Enhancement

Implementation of multimodal feedback systems is proposed, particularly visual documentation of contextual factors (ambient conditions and patient posture) during acquisition phases to compensate for limitations in real-time pulse regeneration fidelity.

Participants' Experience With the Device at the Feasibility Phase

Regarding comfort during the measurement process, participants reported their "comfort level" with an average score of 4.1 (SD 0.7), indicating they felt comfortable. The ease of using the machine was rated at 4.1 (SD 0.7), suggesting the participants found the process easy to follow. For the efficiency of the measurement process, the average score was 3.8 (SD 0.9), reflecting participants' general agreement. Overall satisfaction received a score of 4.1 (SD 0.7), showing participants were satisfied with the machine-generated pulse measurement.

Participants' Qualitative Feedback at the Feasibility Phase

Overall Perception

Comfort

Most users found the device generally comfortable to use without significant discomfort, but it required strict wrist positioning and took some time to adjust.

Efficiency

The participants described the device as fast and efficient with easy-to-understand instructions. However, some participants encountered difficulties in locating accurate pulse positions, leading to prolonged measurement times.

Temperature Perception

The metal components felt colder compared to the warmth of human hands, although this did not cause major discomfort.

Participants' Suggestions for Improvement After the Feasibility Phase

Positioning Accuracy

Participants recommended integrating higher-precision sensors to enable automatic pulse localization, reducing the need for repeated positioning attempts.

Design

The participants also suggested the appearance can be improved. Examples included the exposed internal components and wires housing.

Comfort Optimization

The participants suggested incorporating padded cushions or cotton layers at wrist support areas and considering stabilizing straps to prevent displacement.

Adjustability Features

The participants suggested developing height-adjustable mechanisms for devices or seating to accommodate users' optimal measurement postures.

Pilot

After receiving feedback from the CMPs and participants at the feasibility phase, the team had refined the prototype of the device. The changes include "to increase the comfortability of arm rest from fixed to slidable and lockable when taking pulse," "to enhance initial pressure adjustment to improve the overall experience by adjusting relative pressure to static and dynamic pressure," "to refine dynamic pulse replication algorithms across cun-guan-chi to achieve clinical-grade realism in machine-generated pulse from "no achievable" to "in phase pulse waveforms'." The details of the changes are presented in Table 3.

XSL•FO
RenderX

Table . Changes in the prototype according to Chinese medicine practitioners' and participants' feedback.

Parts	Feedback from CMPs[a] or participants	Initial prototype (feasibility phase)	Final prototype (pilot phase)
Arm rest	Participants suggested the armrest needs to increase its comfortability when taking pulse	Fixed	Slidable and lockable
Pulse sensing technology	CMPs proposed optimizing multisensor arrays in pulse-taking devices to minimize signal artifacts during radial artery data acquisition.	Piezoelectric	Piezoresistive
Sensing parameter	Participants suggested enhancing initial pressure adjustment to improve the overall experience.	Relative pressure	Static and dynamic pressure
Number of pulse sensors	CMPs recommended using instruments equipped with simultaneous 3-finger pulse palpation to collect pulse data.	1	3
Force control technology	CMPs suggested enhancing sensor configurations in pulse-simulating devices to improve the accuracy of pulse waveform replication.	Load cell	Pulse sensor
Silicone skin	CMPs recommended integrating skin-like biomimetic materials into pulse-simulating devices to enhance the fidelity of pulse waveform reproduction	Off the shelf	Custom made
Three-pulse regeneration synchronization	CMPs emphasized refining dynamic pulse replication algorithms across cun-guan-chi to achieve clinical-grade realism in machine-generated pulse	Not achievable, as only one probe was used.	In-phase pulse waveforms are achieved by using 3 probes to collect pulses simultaneously at 3 positions.

[a]CMP: Chinese medicine practitioner.

CMPs and Participants' Characteristics

During the pilot phase, 10 CMPs and 42 participants were recruited. The 10 CMPs had a mean age of 41.5 (SD 5.9) years, 6 (60%) were female, and had a mean clinical experience of 16 (SD 5.8) years. The 42 participants had a mean age of 41.6 (SD 13.3) years, 28 (66.7%) were female, and had a mean BMI of 21.6 (SD 2.1) kg/m^2.

CMPs' Agreement on Machine-Regenerated Pulse at Pilot Phase

Ten CMPs each evaluated 10 or 11 participants' pulses and their machine-generated pulses in terms of frequency, rhythm, wideness, and force. The results indicated that the average agreement score for frequency was 4.3 (SD 0.7), suggesting that CMPs found the machine's pulse frequency to be in agreement. Similarly, the rhythm received an average score of 4 (SD 0.8), indicating that CMPs perceived the rhythm as generally agreed. The wideness of the pulse was rated with an average score of 3 (SD 0.9), reflecting CMPs' views that it had general agreement. Finally, the force of the pulse garnered an average score of 2.8 (SD 0.9), showing that CMPs felt the force reached general agreement.

CMPs Experience With Machine-Generated Pulse Measurement at Pilot Phase

CMPs reported their "comfort level" during the measurement process with an average score of 3 (SD 0.8), indicating that they felt neutral. The "clarity of instructions on using the machine" was rated at 4.3 (SD 1), suggesting that CMPs found the process easily understandable. The "efficiency" of the measurement process was scored at 2.5 (SD 0.5), reflecting CMPs' views that it was less than satisfactory. Overall satisfaction with the experience was rated at 2.6 (SD 0.5), showing that the CMPs held neutral views regarding the machine-generated pulse measurement.

CMPs' Qualitative Feedback at the Feasibility Phase

Overall Experience

User Experience and Comfort

CMPs reported the device's rigid surface texture needed excessive palpation effort, with inaccuracy in the regeneration pulse at sunken (chen 沉), thready (xi細), and deficient (xu 虛) pulse regeneration.

Pulse Accuracy

While capable of replicating pulse frequency, rhythm, and width, the device underperforms in rendering vascular tension, directional force (maishi 脈勢), and pliability. Approximately

XSL•FO
RenderX

more than half of overall pulse morphology can be simulated, although individual variations in cun-guan-chi

Gender Impact

The device demonstrated better pulse reproduction for male participants than female participants, probably due to probe dimensions and positioning protocols, warranting anthropometric optimization through future adjustment.

CMPs' Suggestion for Improvement After the Pilot Phase

Operational Principles and Usability Enhancements

Stratified depth samples of the pulses feature can be collected for CMPs' comparison. The fixed cun-guan-chi spacing requires anthropometric re-engineering for clinical applicability. Noise arising from the motorized component should be reduced.

Automation and Ergonomic Optimization

They suggested developing automated cun-guan-chi localization algorithms and proposed frame expansion for enhanced ergonomics and upgraded depth sensing capabilities. To eliminate "pulsatile extrusion" artifacts, hemodynamic fluidity using algorithmic refinement can be explored. They also suggested implementing triple-finger span adjustability for anatomical adaptability.

Material Science Considerations

Optimize material compliance to mimic cutaneous pliability. Enhance vascular compliance in tubing systems to improve rendering of pulse turgor and tension states.

Participants' Experience With the Device at Pilot Phase

For comfort during the measurement process, participants reported their "comfort level" with an average score of 4 (SD 0.7), indicating that they felt comfortable. The "ease of using" the machine was rated at 3.9 (SD 0.8), suggesting that participants found the process easy to operate. Regarding the "efficiency" of the measurement process, the average score was 3.6 (SD 0.9), reflecting participants' views that it met expectations. Overall satisfaction with the experience was rated at 4 (SD 0.7), showing that the participants were satisfied with the machine-generated pulse measurement.

Participants Qualitative Feedback at Pilot

Overall Perception

Positioning Time Consumption

Prolonged durations during locating the pulse position were noted, requiring multiple attempts to achieve the accurate pulse position due to stringent placement requirements.

Pulse Taking Procedure

Strict postural immobilization requirements posed challenges. Compared to the device, face-to-face CMPs' manual palpation, particularly regarding pressure modulation subtlety, is more comfortable.

User-Driven Optimization Proposals

Implementation of graduated pressure escalation protocols (low initial force with progressive intensification) is advocated to optimize procedural comfort.

Participants' Suggestions for Improvement After the Pilot Phase

Comfort and Efficiency

The device was acknowledged as comfortable in general and operationally efficient. However, the participants suggested the probe interface can be optimized to enhance comfort. They suggested integrating padded cushions at wrist support interfaces and using probes with materials that will not induce cold feeling when touching the skin.

Pulse Localization

Participants reported the technicians faced persistent challenges in achieving accurate localization. Sensor precision upgrades enabling automated localization algorithms were strongly recommended to reduce multiattempt positioning requirements.

Device Appearance

The participants deemed that the appearance of the device could be further improved, and the hand placement space could be expanded.

Discussion

Principal Findings

TCM pulse diagnosis with modern tele-practice technology presents unique challenges in preserving diagnostic authenticity. Our innovative APDC device demonstrates the feasibility of remote pulse assessment through mechanical reproduction. The combination of noninvasive infrared vessel visualization and calibrated force control enables practitioners to locate and assess the cun-guan-chi positions with precision comparable to traditional palpation methods. This technical solution addresses a critical barrier in TCM tele-practice. In this study, the CMPs' and participants' feedback revealed that the APDC device was feasible in collecting pulse data and regenerating the pulse without major discomfort. The CMP deemed it easy to understand the instructions and use the APDC device. However, their rating on efficiency and overall satisfaction with the device per se fell slightly behind. The participants were satisfied with the device in terms of the comfort level and efficiency.

The COVID-19 pandemic highlighted the need for remote TCM diagnostic capabilities [26,27]. Pulse features have been regarded as important information for CMPs to make a proper diagnosis and formulate a treatment plan [7,28]. Our feasibility and pilot study data from the general public demonstrated both the use and limitations of remote pulse diagnosis. While the CMPs tended to agree that the device resembled the pulse feature in frequency and rhythm, they also pointed out the limitation of the device in reproducing the feature in wideness and force. They also reflect their expectation for improvements in the accuracy of reproducing some complex pulse characteristics, such as the depth and directional force of the

XSL•FO
RenderX

pulse. These findings indicate the need for enhanced capabilities in dynamic pulse variation capture.

From the participants' perspective, the APDC device was generally deemed comfortable, easy to use, and efficient. Overall, the participants satisfied the APDC device performance. The participants' suggestions in the pilot phase were that the probe should adjust the physical pressure while sensing the pulse. Our team has replaced the piezoelectric probe in the prototype with a piezoresistive probe, which was used in other similar devices [29,30] and can help reduce discomfort when exerting pressure on the participant's wrist. The participants in both feasibility and pilot phases also suggested increasing the space and adding cushion pads to the hand placement rack. However, it may induce instability of the hand during pulse sensing. Therefore, we remade the rack into a slightly blended shape to fit the wrist posture. Future design can consider using straps to hold the hand position, but it may take longer time and reduce the efficiency of the device. In addition, our study found that the device performed better in male participants than in female participants. The lower agreement scores for pulse force and width observed among female participants may be influenced by the predominance of females in the pilot sample (9/15, 60%) as well as by sex-specific anatomical differences, such as a smaller radial artery diameter or differences in tissue compliance. Accordingly, future refinements incorporating anthropometric optimization are warranted to improve device performance across sexes. Moreover, in response to the qualitative feedback regarding the "cold feeling" and "rigid surface" of the probes, it would be appropriate to further discuss the necessity of adopting biomimetic or softer materials to better mimic human touch, as such material properties may directly influence practitioners' tactile perception and user experience.

From the practitioners' perspective, their overall satisfaction regarding the device did not improve and remained low in phase 2 even after adjustment. It is noteworthy that some key pulse features that are critically determinant in TCM pattern diagnosis were suboptimal, such as pulse width and strength [31]. Another issue was the pulse-taking process by the device that was deemed inefficient by the practitioners. Although we have pen-marked the participants' cun position, the technician needs time to adjust the probes (pulse sensors) to collect the pulse at 3 positions. The practitioners suggested developing automated cun-guan-chi localization algorithms. In addition, while the 3 probes were fixed in positions, this may lead to some inaccuracy in the pulse data collection as the width between cun-guan-chi positions can be varied with the participants' arm length. Therefore, some of the discrepancy between participants' and machine-regenerated pulses can be explained by the variation due to probe placement rather than faulty mechanical regeneration, but we were unable to quantify how much disagreement was attributed to this issue. In the future, the probes should be improved by adjustable in-between width and explore the possibility of automatic locating the pulse position rather than manually.

TCM pulse diagnosis includes diverse variations that reflect complex pathological conditions, and multiple pulse characteristics can coexist [11,32,33]. Given the significantly lower scores observed for pulse force and pulse width, we acknowledge that the APDC device is currently limited to basic hemodynamic monitoring. Consequently, further adjustment and optimization are required to enhance the diagnostic fidelity and clinical applicability of the device. More clinical trials involving diverse patients with well-documented TCM diagnoses would be essential for validating the system's efficacy across various TCM patterns and disease states. Blind tests can be done to have CMPs match the machine-regeneration pulse with the real participant's pulse. Further systematic studies with larger, demographically diverse samples are needed to validate the accuracy of the device in the clinical implementation.

Limitations

This study had some limitations. First, our participants were healthy volunteers recruited from nonclinical settings, without chronic illnesses or any acute or active disease conditions. As a result, the pulse characteristics examined may be less representative than those typically recorded by CMPs in clinical practice. Second, only 4 dimensions of the pulse were measured, and some complex pulse types, such as the tight pulse and soft pulse, require further assessment.

Conclusions

This study demonstrates that an innovative wireless pulse sensing and regeneration system for TCM tele-practice is feasible. This technology enables CMPs to take remote pulse feature assessment. While promising, additional clinical validation is needed to address limitations in capturing complex and dynamic pulse patterns. Future research should focus on enhancing pattern recognition and regeneration capabilities to optimize the device's application in TCM.

Acknowledgments

The authors thank all the participants for their contribution to this study.

Funding

This study received financial support from the Logistics and Supply Chain MultiTech R&D Centre.

Authors' Contributions

WY, LC, and YH designed the study and collected the data. WC, KF, and LC developed the intervention (APDC device). JR and MW analyzed the data. JR, WY, and SC wrote the first draft. All the authors contributed to the articles and approved the submitted version.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Dorsey ER, Okun MS, Bloem BR. Care, convenience, comfort, confidentiality, and contagion: the 5 C's that will shape the future of telemedicine. J Parkinsons Dis 2020;10(3):893-897. [doi: 10.3233/JPD-202109] [Medline: 32538870]
2. Shawwa L. The use of telemedicine in medical education and patient care. Cureus 2023 Apr;15(4):e37766. [doi: 10.7759/cureus.37766] [Medline: 37213963]
3. Chen LLC, Wang SK, Lin TY, Huang LF, Lo MT, Chen CC. A novel telemedicine system to traditional tongue examination for Chinese medical applications. 2021 Presented at: 2021 IEEE International Conference on Digital Health (ICDH); Sep 5-10, 2021; Chicago, IL, USA p. 309-314. [doi: 10.1109/ICDH52753.2021.00055]
4. Wong DCN, Wu Y, Fung H, et al. The provision and perceptions of telemedicine services among traditional Chinese medicine practitioners during COVID-19: a cross-sectional study in Hong Kong, China. Telemed J E Health 2024 Aug;30(8):e2335-e2343. [doi: 10.1089/tmj.2023.0635] [Medline: 38752867]
5. Li P, Longhurst JC, Zaslawski C. The role of the pulse in traditional Chinese medicine. J Acupunct Meridian Stud 2012;5(6):265-274 [FREE Full text]
6. Yeung WF, Chung KF, Ng KY, et al. Prescription of Chinese herbal medicine in pattern-based traditional Chinese medicine treatment for depression: a systematic review. Evid Based Complement Alternat Med 2015;2015:160189. [doi: 10.1155/2015/160189] [Medline: 26180532]
7. Poon MMK, Chung KF, Yeung WF, Yau VHK, Zhang SP. Classification of insomnia using the traditional Chinese medicine system: a systematic review. Evid Based Complement Alternat Med 2012;2012:735078. [doi: 10.1155/2012/735078] [Medline: 22899958]
8. Yeung WF, Chung KF, Poon MMK, et al. Prescription of Chinese herbal medicine and selection of acupoints in pattern-based traditional Chinese medicine treatment for insomnia: a systematic review. Evid Based Complement Alternat Med 2012;2012:902578. [doi: 10.1155/2012/902578] [Medline: 23259001]
9. Wang EX, Zhao X, Chen Y, et al. Design of pulse diagnostic robot for traditional Chinese medicine. 2020 Presented at: 2020 5th International Conference on Robotics and Automation Engineering (ICRAE); Nov 20-22, 2020; Singapore, Singapore p. 87-92. [doi: 10.1109/ICRAE50850.2020.9310822]
10. Yang J, Qie A, Deng J, Hua T, Wu F, Wang X. A novel multi-point pressure sensor for pulse signals acquisition. 2024 Presented at: 2024 IEEE Biomedical Circuits and Systems Conference (BioCAS); Oct 24-26, 2024; Xi'an, China p. 1-5. [doi: 10.1109/BioCAS61083.2024.10798253]
11. Yang J, Wu F, Qie A, Li Q, Wang X, Zhang X. A novel stable and high-fidelity T-structured pulse sensing system for human health assessment. IEEE Sensors J 2024;24(13):21458-21465. [doi: 10.1109/JSEN.2024.3399817] [Medline: 38344259]
12. Nie J, Zhang L, Liu J, Wang Y. Pulse taking by a piezoelectric film sensor via mode energy ratio analysis helps identify pregnancy status. IEEE J Biomed Health Inform 2022 May;26(5):2116-2123. [doi: 10.1109/JBHI.2021.3125707] [Medline: 34748506]
13. Tong-tong Q, Guang-da L, Ting-hang G, Qiu-yue Z, Yu H. Summary of research on digital pulse collection and analysis technology. 2021 Presented at: 2021 36th Youth Academic Annual Conference of Chinese Association of Automation (YAC); May 28-30, 2021; Nanchang, China p. 243-247. [doi: 10.1109/YAC53711.2021.9486512]
14. Pan Y, Fu Z, Yan J, Guo R, Yan H. Developments and prospects of flexible array sensor for pulse diagnostic apparatus. 2017 Presented at: 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET); Mar 22-24, 2017; Chennai p. 2410-2412. [doi: 10.1109/WiSPNET.2017.8300192]
15. Kong KW, Lau WM, Wong KS, et al. A pulse-sensing robotic hand for tactile arterial palpation. 2016 Presented at: 2016 IEEE International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER); Jun 19-22, 2016; Chengdu, China p. 141-145. [doi: 10.1109/CYBER.2016.7574811]
16. Hu CS, Chung YF, Luo CH, Yeh CC, Si X. Pulse differences and 3D pulse mapping in TPNI displacements. Presented at: 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI); Oct 15-17, 2011; Shanghai, China p. 823-827. [doi: 10.1109/BMEI.2011.6098474]
17. Wang D, Zhang D, Lu G. An optimal pulse system design by multichannel sensors fusion. IEEE J Biomed Health Inform 2016 Mar;20(2):450-459. [doi: 10.1109/JBHI.2015.2392132] [Medline: 25608317]
18. Dong K, Chu Y, Tian X, et al. Wearable photoelectric fingertip force sensing system based on blood volume changes without sensory interference. ACS Appl Mater Interfaces 2023 Jul 26;15(29):34578-34587. [doi: 10.1021/acsami.3c06469] [Medline: 37439604]
19. Zhang D, Zhang L, Zhang D, Zheng Y. Wavelet based analysis of doppler ultrasonic wrist-pulse signals. 2008 Presented at: 2008 International Conference on Biomedical Engineering and Informatics; May 27-30, 2008; Sanya, China p. 539-543. [doi: 10.1109/BMEI.2008.326]

XSL•FO
RenderX

20. Zhang J, Niu X, Yang XZ, et al. Design and application of pulse information acquisition and analysis system with dynamic recognition in traditional Chinese medicine. Afr Health Sci 2014 Sep;14(3):743-752. [doi: 10.4314/ahs.v14i3.34] [Medline: 25352897]
21. Priority medical devices list for the COVID-19 response and associated technical specifications: interim guidance, 19 November 2020. World Health Organization. URL: https://apps.who.int/iris/handle/10665/336745 [accessed 2026-02-23]
22. Eldridge SM, Chan CL, Campbell MJ, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot Feasibility Stud 2016;2(1):64. [doi: 10.1186/s40814-016-0105-8] [Medline: 27965879]
23. Leung C. Wireless Radial Arterial Pulse Regeneration System for Traditional Chinese Medicine Tele-Practice: IEEE Robot Autom Lett. URL: https://www.lscm.hk/en/project/database_detail/269/?back=9d00a1401c04be50daa45073dcda172e [accessed 2026-02-25]
24. Dong FH. Precise application of traditional Chinese medicine in minimally-invasive techniques. Zhongguo Gu Shang 2018 Jun 25;31(6):493-496. [doi: 10.3969/j.issn.1003-0034.2018.06.001] [Medline: 29945400]
25. VeinViewer flex. Christie Medical Holdings. URL: https://christiemed.com/products/veinviewer-models/veinviewer-flex/ [accessed 2026-02-23]
26. Chen SC, Yeung WF, Cheng HL, Li MH, Ho YS. Views of Hong Kong Chinese medicine practitioners on the application of the “Chinese Medicine Anti-epidemic Plans” prepared by the Chinese medicine expert group of central authorities: a focus group study. BMC Complement Med Ther 2024 May 4;24(1):184. [doi: 10.1186/s12906-024-04469-3] [Medline: 38704524]
27. Chen SC, Cheng HL, Wang DD, et al. Experience of parents in delivering pediatric tuina to children with symptoms of attention deficit hyperactivity disorder during the COVID-19 pandemic: qualitative findings from focus group interviews. BMC Complement Med Ther 2023 Feb 18;23(1):53. [doi: 10.1186/s12906-023-03891-3] [Medline: 36804041]
28. Ruan J, Chen S, Ho YS, et al. Chinese medicine practitioners’ consensus on traditional Chinese medicine diagnostic patterns, symptoms, and herbal formulas for COVID-19 survivors: a Delphi study. Eur J Integr Med 2024 Feb;66:102339. [doi: 10.1016/j.eujim.2024.102339]
29. Ma Y, Zhao Y, Liu L, et al. Pulse wave measuring via fingertip-like sensor to analyze and identify pulse-rate status. IEEE Sensors J 2023;23(3):3147-3158. [doi: 10.1109/JSEN.2022.3230696] [Medline: 38558829]
30. Huang L, Zhai X, Tian P, et al. Wrist pulse acquisition system based on MEMS piezoresistive sensor array for health information monitoring. IEEE Sensors J 2025;26(3):4615-4624. [doi: 10.1109/JSEN.2025.3559555]
31. Yeung WF, Chung KF, Zhang NLW, et al. Identification of Chinese medicine syndromes in persistent insomnia associated with major depressive disorder: a latent tree analysis. Chin Med 2016;11(1):4. [doi: 10.1186/s13020-016-0076-y] [Medline: 26877762]
32. Ho L, Xu Y, Zhang NL, et al. Quantification of prevalence, clinical characteristics, co-existence, and geographic variations of traditional Chinese medicine diagnostic patterns via latent tree analysis-based differentiation rules among functional dyspepsia patients. Chin Med 2022 Aug 30;17(1):101. [doi: 10.1186/s13020-022-00656-x] [Medline: 36038888]
33. Gadau M, Zhang SP, Yip HY, et al. Pattern differentiation of lateral elbow pain in traditional Chinese medicine: a systematic review. J Altern Complement Med 2016 Nov;22(11):921-935. [doi: 10.1089/acm.2016.0098] [Medline: 27754695]

Abbreviations

APDC: Assistive Pulse Data Collection
CMP: Chinese medicine practitioner
CONSORT: Consolidated Standards of Reporting Trials
TCM: traditional Chinese medicine

Edited by A Kushniruk; submitted 19.May.2025; peer-reviewed by D Marques; revised version received 14.Jan.2026; accepted 09.Feb.2026; published 06.Mar.2026.

Please cite as:
Ruan JY, Chen S, Wu M, Li S, Ho YS, Fan KL, Chow WL, Chiu L, Yeung WF
Assistive Technologies for Pulse-Based Tele-Practice in Traditional Chinese Medicine: Two-Phase, Dual-Method Feasibility and Pilot Study
JMIR Hum Factors 2026;13:e77740
URL: https://humanfactors.jmir.org/2026/1/e77740
doi:10.2196/77740

© Jia Yin Ruan, Shucheng Chen, Minru Wu, Shuhan Li, Yuan Shan Ho, Ka Lun Fan, Wing Leung Chow, Leung Chiu, Wing Fai Yeung. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 6.Mar.2026. This is an open-access

XSL•FO
RenderX

article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Wearable Device Photoplethysmography As a Viable Tool to Longitudinally Monitor Vasoconstriction Biomarkers for Predicting Vaso-Occlusive Crisis in Sickle Cell Disease: Feasibility and Validation Study

Payal Shah[1], MS; Sabrina Sy[2], BS; Mingjing Chen[2], BS; Michael CK Khoo[2], PhD; Thomas D Coates[1], MD; Saranya Veluswamy[1], MD

[1]Division of Hematology, Cancer and Blood Disease Institute, Children's Hospital Los Angeles, University of Southern California, 4650 Sunset Boulevard, Los Angeles, CA, United States

[2]Alfred E. Mann Department of Biomedical Engineering, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Saranya Veluswamy, MD
Division of Hematology, Cancer and Blood Disease Institute, Children's Hospital Los Angeles, University of Southern California, 4650 Sunset Boulevard, Los Angeles, CA, United States

Abstract

Background: Entrapment of sickled red blood cells in the microvasculature leads to sudden painful vaso-occlusive crises (VOCs) in sickle cell disease (SCD). This is potentially triggered by autonomic nervous system–mediated vasoconstriction in the microvasculature. Indeed, vasoconstriction biomarkers derived from a single night of laboratory-based fingertip photoplethysmography (PPG) recording were predictive of a higher frequency of future VOC in SCD. Noninvasive, remote, and longitudinal monitoring of autonomic vasoreactivity will facilitate the development of predictive biomarkers of imminent VOC.

Objective: This study aimed to assess the feasibility and performance of a wearable wristband device to longitudinally monitor nocturnal peripheral autonomic vasoreactivity and to cross-validate the vasoconstriction parameters across the “gold-standard” finger sensor.

Methods: A total of 12 patients with SCD and 6 healthy controls were recruited to wear a wristband device (Biostrap) with a PPG sensor on a nightly basis. For cross-validation studies, 50% (3/6) controls wore both the wristband and a sleep monitoring device (AliceNightOne) with a finger PPG sensor. We quantified autonomic vasoreactivity by processing PPG signals and deriving vasoconstriction parameters—magnitude of vasoconstriction (Mvasoc) and photoplethysmography amplitude coefficient of variation (PPGampCV). We performed a correlation analysis of the vasoconstriction parameters within each device to investigate whether Mvasoc and PPGampCV can be used as surrogate markers of vasoconstriction, and then cross-validated the PPGampCV across the wristband and finger PPG devices.

Results: A total of 131 nocturnal PPG recordings were made with a wristband device (1 - 19 nights per participant; patients with SCD: n=79, 60%; controls: n=52, 40%). A total of 9 nocturnal recordings (3 nights per participant) were made with both wristband and finger sensor devices. Longitudinal continuous PPG recordings were feasible with the wearable device, with significant within-night and night-to-night variability in vasoconstriction parameters, suggesting dynamic changes in autonomic vasoreactivity. Mvasoc and PPGampCV significantly correlated within devices—the maximum overnight correlation was 0.82 (P<.001) for the finger sensor and 0.69 (P<.001) for the wristband sensor, suggesting that PPGampCV can serve as a surrogate for Mvasoc. Cross-validation analysis of PPGampCV across wristband and fingertip sensors showed statistically significant correlations on all 9 nights (overnight correlation coefficient ranging from 0.24 - 0.7), with some nightly segments of PPGampCV showing very strong correlation across devices.

Conclusions: Wearable wristband devices are feasible tools for the collection of continuous PPG measurements and vasoconstriction parameters, which serve as objective markers of autonomic vasoreactivity in users with and without SCD. We have optimized the methods of quantifying vasoconstriction from wearable device PPG signals, and cross-validated them with standardized sensors. These findings enable large-scale, real-time monitoring of autonomic vasoreactivity along with pain outcomes for the development of vasoconstriction parameters as biomarkers imminent VOC in patients with SCD. This biomarker also has the potential to impact other diseases involving autonomic vascular dysregulation.

(JMIR Hum Factors 2026;13:e75465) doi:10.2196/75465

XSL•FO
RenderX

KEYWORDS

sickle cell disease; wearable devices; microvascular blood flow; vasoconstriction; remote monitoring

Introduction

Vaso-occlusive crisis (VOC) due to obstruction of microvascular blood flow by sickled red blood cells is the prime reason for the morbidity and mortality in sickle cell disease (SCD) [1]. Peripheral vasoconstriction in the microvasculature due to dysregulated autonomic nervous system (ANS) responses and the subsequent reduced microvascular blood flow increases red blood cell transit time in the microvasculature, thus potentiating VOC [2-7]. Our recent work on the neurovascular physiology of SCD in humans has established that ANS-mediated decreases in microvascular blood flow, quantified by a novel vasoconstriction biomarker, strongly predict future VOC [8]. This novel vasoconstriction index, derived from the amplitude of finger photoplethysmography (PPG) signals, is a composite measure of the dynamic changes in peripheral blood flow occurring over time and is reflective of sympathetic neural inputs to the microvasculature. A single nocturnal measurement of this index in patients with SCD showed that a higher magnitude of vasoconstriction (Mvasoc) was predictive of increased hospitalizations for VOC in the upcoming years. However, day-to-day variations in autonomic vasoreactivity and their implications for imminent VOC risk remain unknown. All our previous measurements of peripheral vasoreactivity and ANS activity were performed in a well-controlled laboratory environment using standard finger PPG sensors to measure microvascular blood flow. Longitudinal remote monitoring of vasoconstriction events and ANS activity has a high potential for predicting and monitoring individualized pain episodes and is the necessary next step to develop nocturnal vasoconstriction indices as a predictor of imminent VOC.

Current wearable devices use optical PPG sensors to monitor heart rate variability and oxygen saturation. However, the raw PPG signals are often inaccessible before they go through automatic gain and processing, such that key information from the amplitude of the raw PPG signal is lost. Moreover, in the most commonly used commercial wearable devices, the PPG signal is measured from the wrist [9] or at the base of a finger [10,11]. These anatomic locations are not as highly perfused as the fingertip, and the PPG waveforms may be expected to be morphologically different from the fingertip due to differences in local vascular physiology and lower signal-to-noise ratios. On the other hand, wearable device PPG measurements are likely more reliable based on user compliance and practical considerations.

In this study, we established the use of PPG signals from a wrist-worn wearable device for real-time, remote monitoring of autonomic vasoreactivity and measurement of vasoconstriction events during sleep. The Biostrap EVO is a commercially available wrist-worn device (Biostrap LLC) that uses an optical PPG sensor to record standard continuous physiological data [12], including heart rate, computed respiratory activity, oxygen saturation (SpO2), and sleep and activity patterns [13]. Uniquely, our team was also able to access the raw PPG waveforms from the device, which is critical to quantify our key vasoconstriction parameters. The AliceNightOne (ANO) is a Food and Drug Administration–approved home sleep testing device (Koninklijke Philips N.V.), primarily designed to detect sleep apnea [14]. The 3-sensor device with a nasal cannula, finger PPG sensor, and chest belt detects nasal airflow pressure, SpO2, heart rate, respiratory effort, and body positioning. The ANO uses medical-grade “gold-standard” finger PPG probes to measure heart rate and oxygenation. The aims of this study were to

1. Assess the feasibility of wristband wearable devices to remotely monitor nocturnal peripheral autonomic vasoreactivity
2. Optimize the measurement of vasoconstriction parameters from wearable device PPG in patients with SCD and healthy controls.
3. Cross-validate the vasoconstriction parameters derived from a wearable wristband sensor (Biostrap) and a “gold-standard” finger sensor (ANO).

Methods

Ethical Considerations

The study was approved by the Institutional Review Board of the Children’s Hospital, Los Angeles, and all participants provided informed consent or assent to participate in the study. The study was conducted in accordance with the Declaration of Helsinki and the data was de-identified as per the Health Insurance Portability and Accountability Act (HIPAA), and subjects were compensated for their participation.

Study Setup

Patients with SCD and healthy controls underwent autonomic monitoring with a wristband sensor to measure continuous PPG during sleep. Recordings were automatically triggered at sleep onset and paused for 5 minutes every hour to allow syncing of data wirelessly to a phone-based app. The raw PPG waveforms were made accessible from a cloud-based server, exported as .csv files, and selected for analyses based on the quality and completeness of the overnight recording. Raw signals were first screened to determine if there were any segments with significant motion artifacts and were compared alongside accelerometer data also available from the Biostrap cloud server.

For the cross-validation of vasoconstriction parameters, a subset of the control participants wore the ANO sleep device with a finger sensor simultaneously with the Biostrap wristband sensor. Sleepware G3 (version 3.9.7.0; Respironics Inc) provides the capability of storing acquisitions from an ANO sleep device on a cloud-based server and exporting the .edf files (European Data Format) [15]. We converted data from both wristband and finger sensors into .Mat files (MATLAB) to compare them synchronously on 1 platform and derive the physiological parameters of interest.

XSL•FO
RenderX

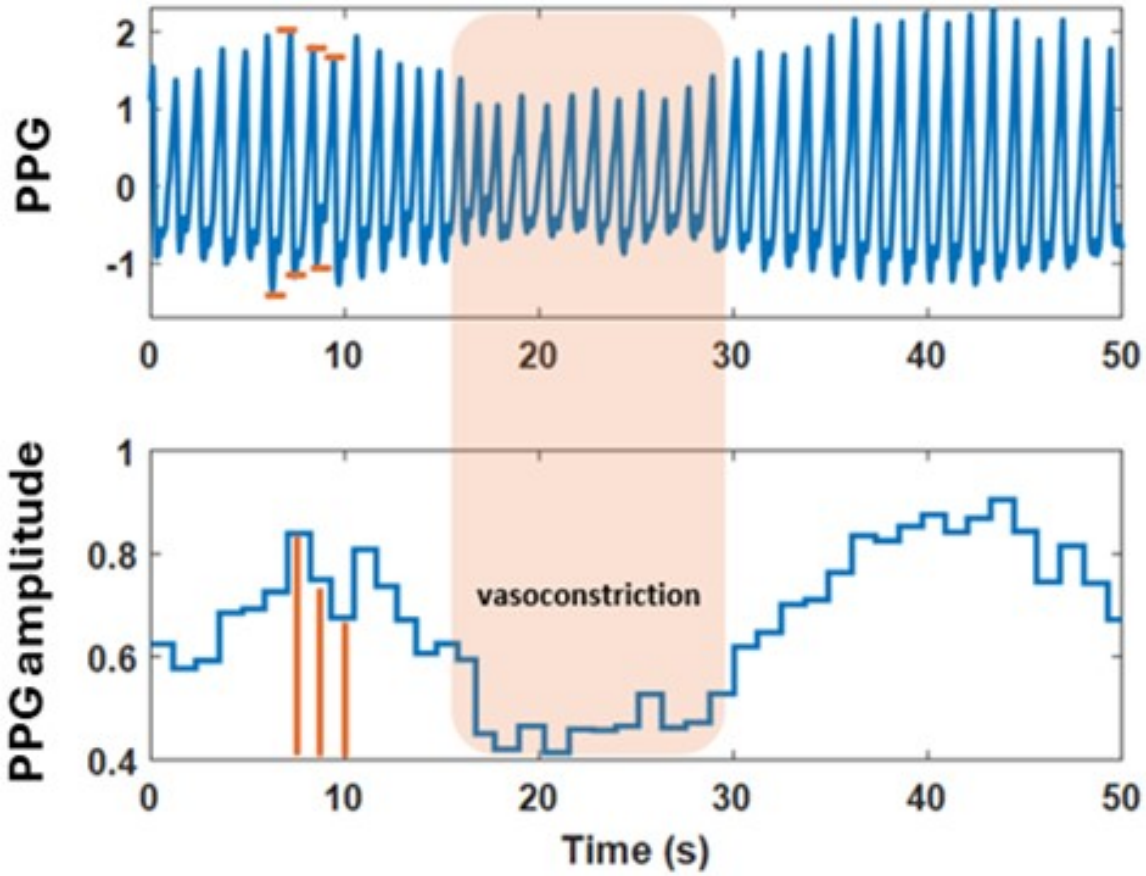
PPG Signal Processing and Derivation of Vasoconstriction Parameters

Signal Processing

The peak-to-trough amplitude of each pulse in the raw PPG signal reflects the pulsatile change in arteriolar blood volume, with a lower amplitude indicating vasoconstriction (Figure 1). The photoplethysmography amplitude (PPGamp), along with the pulse duration, was calculated from both wrist and finger raw PPG signals and divided into consecutive 15-minute segments (Figure 2A). First, segments with clear signal loss were visually identified and manually trimmed or eliminated. Second, outliers were removed by replacing values exceeding the 98th percentile within each segment with the 98th percentile value. Finally, the signal was normalized to its 95th percentile. After linear detrending, the respiratory influence (frequency >0.15 Hz) was filtered out from both wristband and finger PPGamp signals (Figure 2B). These processed "clean" PPGamp signals from each device were used to calculate corresponding vasoconstriction parameters.

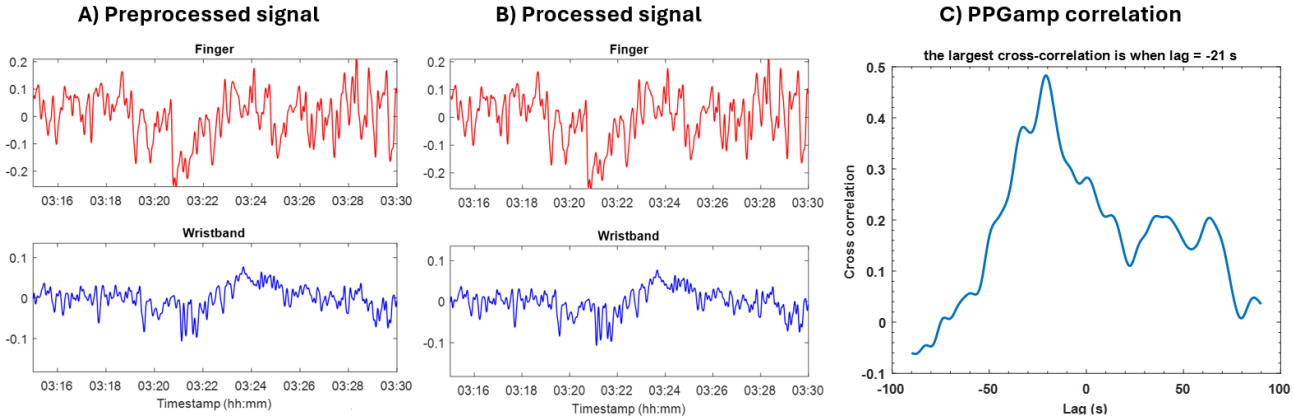
Since each device had its own independent time base, the corresponding finger and wristband PPGamp signals had to be aligned in time in order to facilitate comparison of vasoconstriction parameters across devices. The cross-correlation function is a standard method in time series analysis that is used to assess the degree of linear similarity between 2 signals and has been comprehensively described in the literature [16]. We used this technique to compare and align the PPGamp signals from the 2 different devices. The PPGamp time series from each device was first autocorrelated to ensure that these signals represent physiological measurements rather than random noise. Thereafter, corresponding segments of the finger and wristband PPGamp signals underwent multiple time shifts over the other to locate the point of maximum positive cross-correlation within a +30 or –30 second window, thus synchronizing the PPGamp signals from both devices (Figure 2C).

Figure 1. Photoplethysmography (PPG) amplitude calculated from peak-to-trough of each pulse of raw PPG signal with a decrease in amplitude signifying vasoconstriction.



XSL•FO
RenderX

Figure 2. Representative 15-minute PPGamp segments from finger and wristband PPG sensors. (A) Preprocessed signals depicting finger and wristband PPGamp extracted from raw PPG signals. (B) Processed signals depicting "clean" finger and wristband PPGamp after removal of artifacts and respiratory influence. (C) Cross-correlation of the finger and wristband PPGamp segments. In this example, the maximal correlation occurs at a lag of −21 seconds. PPG: photoplethysmography; PPGamp: photoplethysmography amplitude.

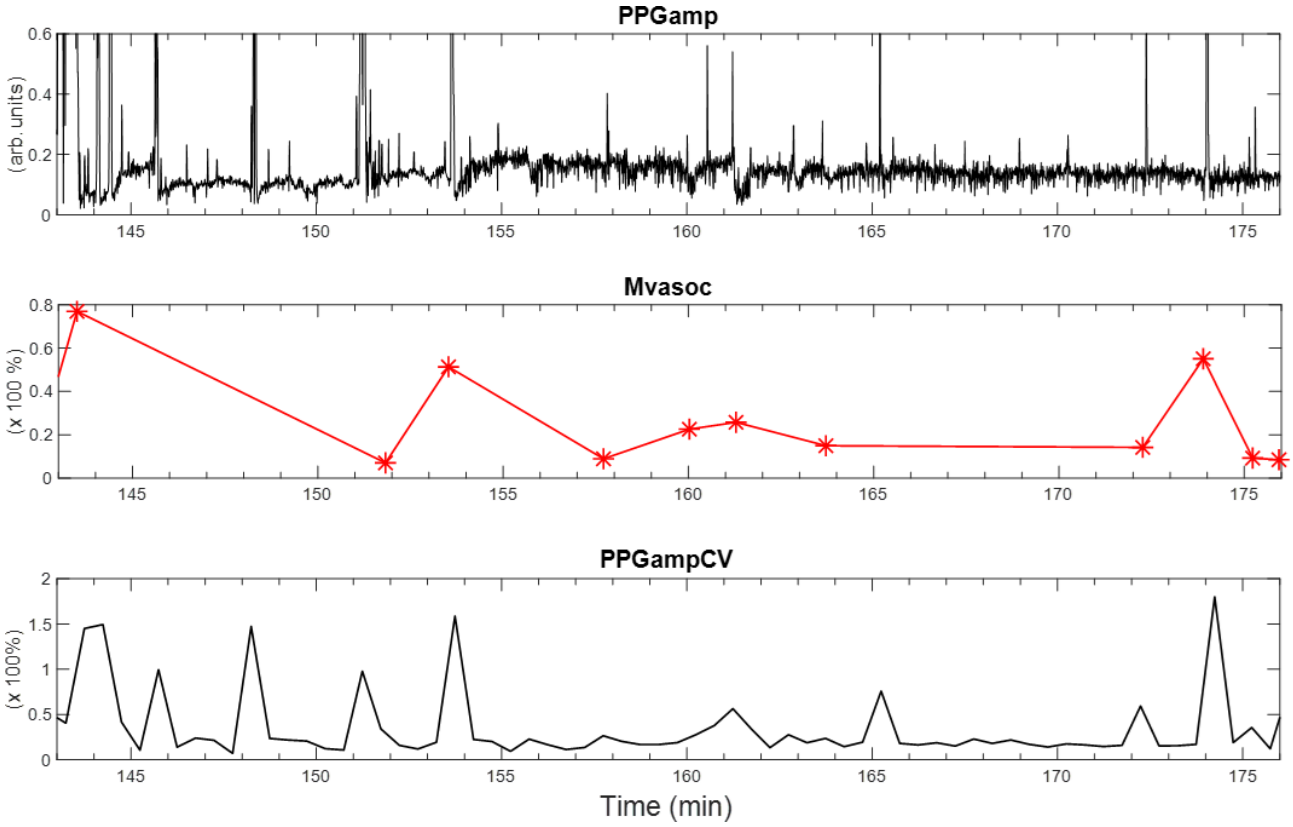


Calculation of Vasoconstriction Parameters: Mvasoc and PPGampCV

We have previously introduced Mvasoc as a unique nocturnal vasoconstriction biomarker that takes into account the frequency, magnitude, and duration of autonomic-mediated spontaneous vasoconstrictions occurring each night [8]. The algorithm was developed to detect and quantify a vasoconstriction event as a significant reduction in PPGamp below the preceding 15-second baseline. Mvasoc is subsequently defined as the ratio of the area of the drop below baseline in PPGamp and the duration of the vasoconstriction, normalized by the baseline PPGamp level (described in detail in the supplement of Chalacheva et al [8]). Because many vasoconstriction events can occur during the night, the median Mvasoc value per night is used as a quantitative index of the user's nocturnal vasoconstriction behavior during a given night [8]. However, when the Mvasoc algorithm is applied to PPGamp signals, the rule-based measure can be sensitive to time-varying changes in signal-to-noise ratio, and thus the time locations and amplitude of the detected vasoconstrictions can vary across PPG devices. To overcome this limitation, we used a second measure to quantify the dynamic variability in beat-to-beat PPGamp. More specifically, the coefficient of variation (CV; CV=SD/mean) of PPGamp within a sliding time window of 5 minutes was calculated every 30 seconds. In initial studies, this parameter (ie, PPGampCV) correlated closely and robustly with Mvasoc, while simultaneously eliminating the need for a rule-based algorithm to detect "significant vasoconstriction" events. Figure 3 is a representation of Mvasoc and PPGampCV parameters from a wristband PPGamp segment.

XSL•FO
RenderX

Figure 3. The first row represents a PPGamp segment from wristband sensor, and the next 2 rows show corresponding Mvasoc and PPGampCV parameters derived from the PPGamp. PPGamp: photoplethysmography amplitude; PPGampCV: photoplethysmography amplitude coefficient of variation; Mvasoc: magnitude of vasoconstriction.



Within-Device Correlation of Mvasoc and PPGampCV Vasoconstriction Parameters

To rigorously assess whether PPGampCV can be used as a surrogate for Mvasoc, we performed a correlation analysis of the 2 parameters as measured by the fingertip sensor in the ANO device and wrist sensor in the Biostrap device and derived the following within-device correlation coefficient (r_{wd}) values:

1. Overnight r_{wd}: represents the correlation between vasoconstriction parameters Mvasoc and PPGampCV for the whole night
2. Peak r_{wd}: represents the segment with the maximum correlation of Mvasoc and PPGampCV out of all the 15-minute segments on a given night

Across-Device Cross-Validation of Vasoconstriction Parameter

To determine if the wrist device gave the same results as the fingertip sensor, we correlated the PPGampCV parameter from the wristband and "gold-standard" finger sensors worn at the same time. The corresponding 15-minute segments from both devices were only considered for cross-correlation if at least 3 vasoconstriction events were detected within each segment. We correlated the PPGampCV between the finger and wrist PPGamp each night and derived the following across-device correlation coefficient (r_{ad}) values:

1. Overnight r_{ad}: represents the correlation of PPGampCV across wrist and finger sensor for all 15-minute segments per night
2. Peak r_{ad}: represents the segment with the maximal correlation of PPGampCV out of all corresponding 15-minute segments per night.

In all the analyses mentioned earlier, *P* values of <.05 suggested that the correlation between signals was statistically significant.

Results

Longitudinal Monitoring of Vasoconstriction Parameters With Wristband Device

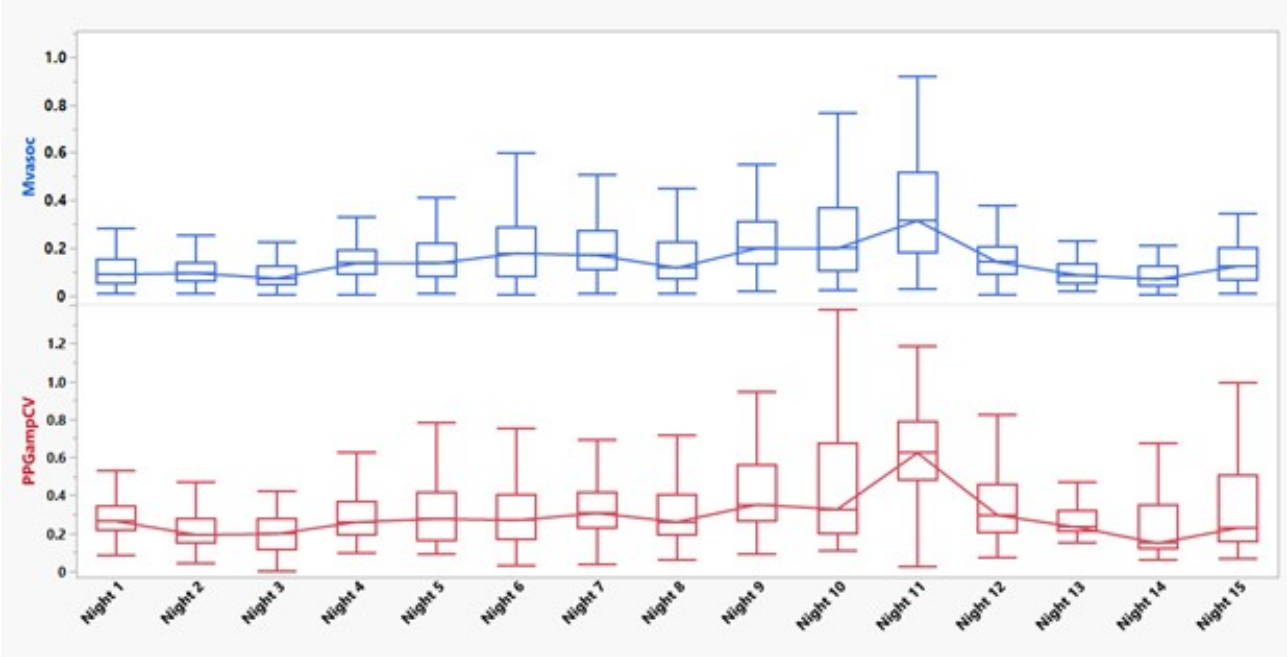
A total of 12 patients with SCD and 6 control participants participated in the study. The mean age of the sample was 25 (range 14 - 40) years. The mean hemoglobin was 8.6 g/dl (SD 1.9 g/dl) and 14.1 g/dl (SD 1.8 g/dl) for patients with SCD and controls, respectively. A total of 10 (83%) of the 12 patients with SCD reported the use of disease-modifying therapy with hydroxyurea at the time of the study. A total of 131 nocturnal PPG recordings were made with wristband devices in all participants. Patients with SCD had a total of 79 (60%) nocturnal recordings (range 1 - 15 nights), and controls had 52 (40%) nocturnal recordings (range 1 - 19 nights). While participants wore the wristband for a variable number of nights due to device availability, there were no reported logistical issues with functioning of the device, and participants reported ease of use.

Longitudinal measurements of Mvasoc and PPGampCV vasoconstriction parameters showed significant variability within each night and from night to night. A representative participant's longitudinal data are depicted in Figure 4 and show that there is a wider range of vasoconstriction activity on nights 10 and

XSL•FO
RenderX

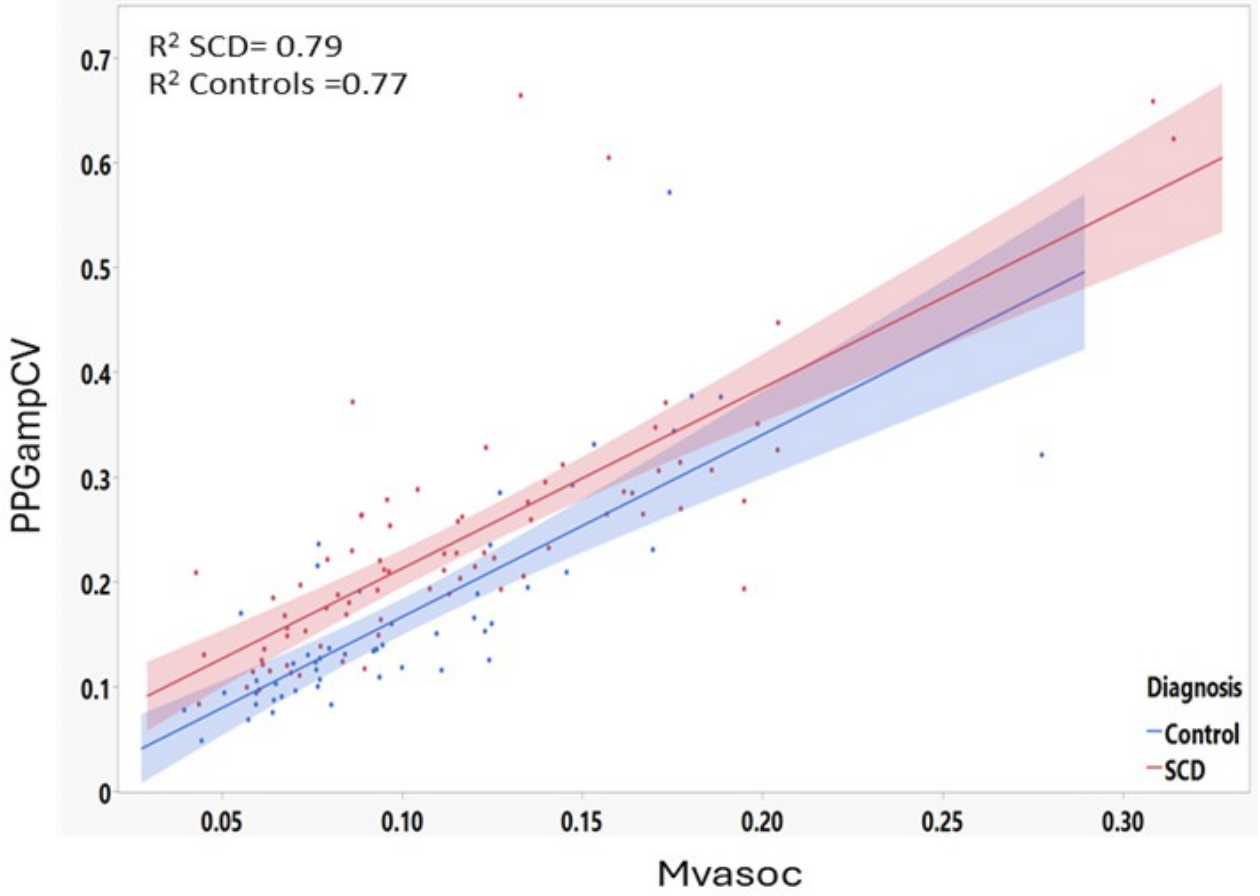
11, with an increase in the median Mvasoc and PPGampCV noted on those nights, reflecting increased autonomic vasomotor activity. On visual inspection, the night-to-night changes in Mvasoc and PPGampCV appeared to closely track each other, suggesting a potential correlation between the 2 parameters. Indeed, the median Mvasoc of each night was closely correlated with the corresponding median PPGampCV both for patients with SCD and controls in all the nightly data (patients with SCD: R^2=0.79; controls: R^2=0.77; Figure 5).

Figure 4. Longitudinal monitoring of vasoconstriction parameters Mvasoc and PPGampCV in a patient with sickle cell disease from the wristband device. The box plot shows the median and IQR per night, and the midline shows the median of these parameters across all nights. PPGampCV: photoplethysmography amplitude coefficient of variation; Mvasoc: magnitude of vasoconstriction.



XSL•FO
RenderX

Figure 5. Correlation of median Mvasoc with corresponding median PPGampCV for all recorded nights from the wristband device among patients with SCD and controls. Mvasoc: magnitude of vasoconstriction; PPGampCV: photoplethysmography amplitude coefficient of variation; SCD: sickle cell disease.



Within-Device Correlation of Vasoconstriction Parameters (Mvasoc and PPGampCV)

While median Mvasoc and PPGampCV track closely in longitudinal nightly recordings, we further analyzed the correlation of these parameters during time segments within individual nightly recordings. Three control participants wore both finger and wristband PPG devices for a total of 9 nights, and each nighttime recording was divided into 15-minute segments. Table 1 shows the number of vasoconstriction (Mvasoc) events detected each night from both the finger sensor and wristband devices. There were a total of 113 viable segments over 9 nights, with an average of 10 and 11 vasoconstriction events detected during each segment from the finger and wristband sensors, respectively.

Table . Number of vasoconstriction events (magnitude of vasoconstriction) detected each night from the finger and wristband sensors.

	Finger sensor, n	Wristband, n
Participant 1		
Night 1 (5 segments)	34	56
Night 2 (8 segments)	90	104
Night 3 (12 segments)	140	152
Participant 2		
Night 1 (20 segments)	186	226
Night 2 (18 segments)	175	195
Night 3 (8 segments)	71	120
Participant 3		
Night 1 (19 segments)	206	214
Night 2 (18 segments)	200	190
Night 3 (5 segments)	40	57

XSL•FO
RenderX

Cross-correlation of Mvasoc and corresponding PPGampCV values were calculated for each segment of these 9 nights. Overnight r_{wd} correlation and segmental peak r_{wd} correlation coefficient data for all 9 nights (Table 2) show that Mvasoc was significantly correlated with PPGampCV each night within both devices and within almost all segments. The highest overnight r_{wd} correlation was 0.82 (P<.001) for the finger sensor and 0.69 (P<.001) for the wristband for all nights. The highest peak r_{wd} correlation was 0.99 (P<.05) in both devices for all nights (113 segments). Figure 6 shows the within-device overnight r_{wd} correlation between Mvasoc and PPGampCV for a representative night.

Table . Cross-correlation between magnitude of vasoconstriction (Mvasoc) and photoplethysmography amplitude coefficient of variation (PPGampCV) using overnight correlation coefficient (overnight r_{wd}) and peak correlation coefficient (peak r_{wd}) within wristband (Biostrap) and finger sensors (AliceNightOne).

Mvasoc versus PPGampCV correlation	Finger sensor	Wristband
Participant 1		
Night 1 (5 segments)		
Overnight r_{wd}[a]	0.82	0.40
Peak r_{wd}[b]	0.90	0.74
Night 2 (8 segments)		
Overnight r_{wd}	0.35	0.47
Peak r_{wd}	0.93	0.78
Night 3 (12 segments)		
Overnight r_{wd}	0.51	0.41
Peak r_{wd}	0.82	0.80
Participant 2		
Night 1 (20 segments)		
Overnight r_{wd}	0.32	0.69
Peak r_{wd}	0.91	0.96
Night 2 (18 segments)		
Overnight r_{wd}	0.29	0.48
Peak r_{wd}	0.97	0.93
Night 3 (8 segments)		
Overnight r_{wd}	0.47	0.34
Peak r_{wd}	0.99	0.92
Participant 3		
Night 1 (19 segments)		
Overnight r_{wd}	0.40	0.49
Peak r_{wd}	0.80	0.99
Night 2 (18 segments)		
Overnight r_{wd}	0.37	0.51
Peak r_{wd}	0.79	0.95
Night 3 (5 segments)		
Overnight r_{wd}	0.33	0.30
Peak r_{wd}	0.95	0.50[c]

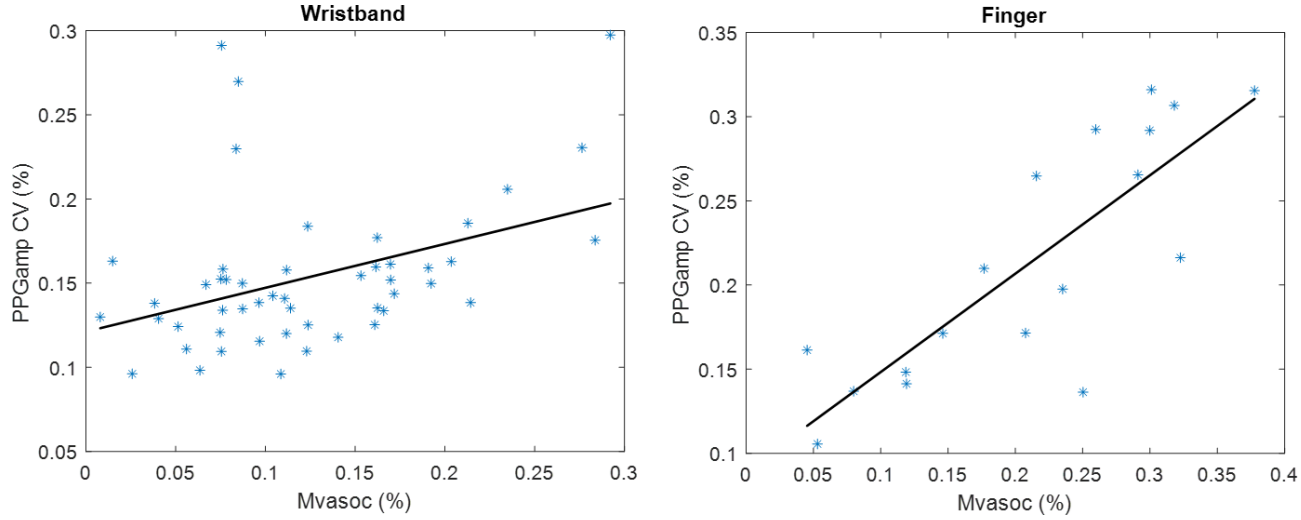
[a]Overnight r_{wd}: correlation coefficient value of vasoconstriction parameters, Mvasoc and PPGampCV, for the whole night; P<.05.

[b]Peak r_{wd}: correlation coefficient value of segment with highest correlation out of all 15-minute segments per night; P<.05.

[c]P>.05.

XSL•FO
RenderX

Figure 6. Example of overnight correlation (r_{wd}) between vasoconstriction parameters, Mvasoc and PPGampCV, for 1 control 1 night in wristband (Biostrap) and finger sensor (AliceNightOne). Panel (a wristband*; r*=0.40; *P*<.005) and AliceNightOne device in panel (b finger sensor; *r*=0.82; *P*<.0001). The significant correlation between Mvasoc and PPGampCV assured us that PPGampCV can be used as a surrogate marker of the magnitude of vasoconstriction. Mvasoc: magnitude of vasoconstriction; PPGampCV: photoplethysmography amplitude coefficient of variation.



Across-Device Cross-Validation of Vasoconstriction Parameters (CV)

Since the PPGampCV serves as a surrogate marker of vasoconstriction and is overall easier to derive from the PPG and compare across devices, we chose the PPGampCV to perform cross-validation of vasoconstriction parameters between the 2 devices. The highest overnight r_{ad} cross-correlation for all nights was 0.70, and the highest peak r_{ad} cross-correlation for all nights was 0.99. Table 3 shows the PPGampCV overnight correlation (r_{ad}) coefficient and segmental peak (r_{ad}) correlation coefficient data for all 9 nights (123 segments). Figure 7 shows an example of correlation analysis of PPGampCV across devices in 1 corresponding 15-minute segment.

Table . Overnight r_{ad} correlation and peak r_{ad} correlation coefficients of photoplethysmography amplitude coefficient of variation between wristband (Biostrap) and finger sensor (AliceNightOne).

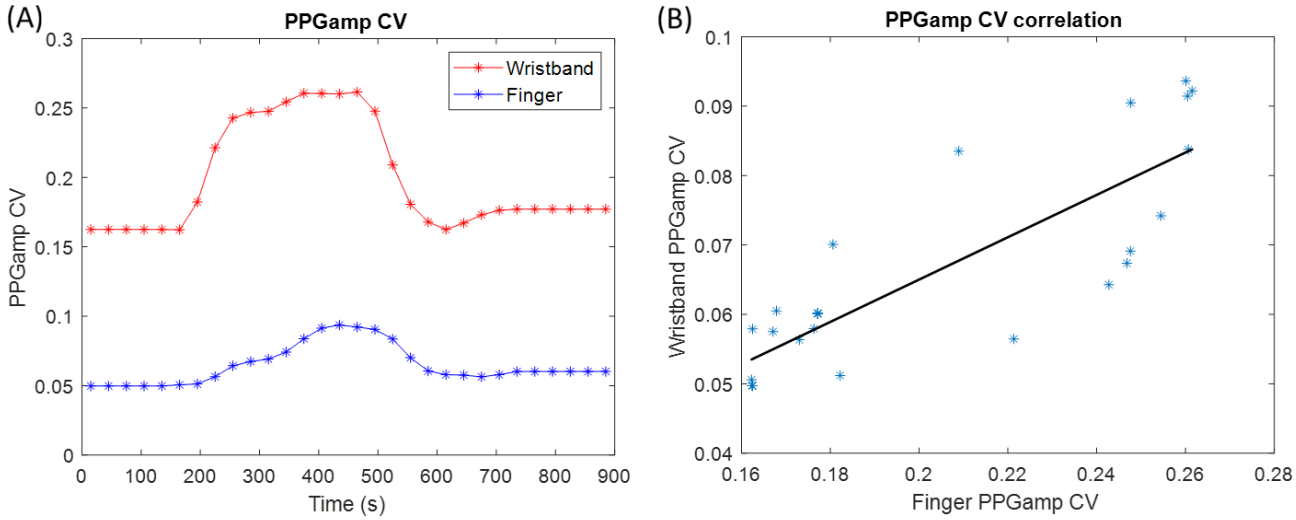
	Overnight r_{ad}[a]	Peak r_{ad}[b]
Participant 1		
Night 1 (5 segments)	0.24	0.76
Night 2 (8 segments)	0.50	0.90
Night 3 (12 segments)	0.70	0.89
Participant 2		
Night 1 (20 segments)	0.49	0.97
Night 2 (18 segments)	0.57	0.99
Night 3 (8 segments)	0.52	0.96
Participant 3		
Night 1 (23 segments)	0.36	0.97
Night 2 (24 segments)	0.62	0.96
Night 3 (5 segments)	0.65	0.56

[a]Overnight r_{ad}: represents the correlation coefficient value for all 15-minute segments per night; *P*<.05.

[b]Peak r_{ad}: represents the correlation coefficient value of segment with highest correlation out of all 15-minute segments per night; *P*<.05.

XSL•FO
RenderX

Figure 7. Correlation of PPGampCV between the wristband and finger device for a representative 15-minute segment of 1 night. The left panel (A) shows the corresponding PPGampCV values from both devices over time, and the right panel (B) shows the correlation of PPGampCV values between the 2 devices (r=0.84; P<.001). PPGampCV: photoplethysmography amplitude coefficient of variation.



Discussion

Principal Findings

The sudden, unpredictable onset of VOC remains the major cause of morbidity and the leading cause of hospitalizations in patients with SCD [17]. While some patients experience an aura or premonition that a VOC is about to occur [18], there are no objective biomarkers that can predict the risk of an imminent VOC and thus open up a therapeutic window to abort its propagation. We have introduced nocturnal vasoconstriction indices of autonomic peripheral vasoreactivity that have the potential to be predictive biomarkers of impending VOC crisis. This study establishes the feasibility of measuring these biomarkers in a real-world setting with a wearable device, such that the vasoconstriction biomarker can be validated in larger populations. To the best of our knowledge, this is the first study that uses the PPG signal from a wearable-based sensor system to quantify microvascular blood flow changes accompanying peripheral vasoconstriction.

We successfully used the Biostrap wearable wristband device to longitudinally monitor raw PPG signals and derive autonomic vasoconstriction parameters. Participants reported ease of use, and no significant barriers were identified in using the device. There was substantial night-to-night variability in the median Mvasoc and PPGampCV parameters within participants, which implies varying levels of sympathetic vasoreactivity. This has potential implications for tracking pain and imminent VOC risk, as SCD physiology dictates that vasoconstriction and resultant decrease in microvascular perfusion increase the likelihood of vaso-occlusion.

Although Mvasoc is a direct quantification of vasoconstriction events from the PPGamp signal, we found that this parameter is more prone to artifacts that can plague wearable device signals. The PPGampCV is a more robust measure of variability of the PPGamp amplitude that closely reflects Mvasoc and is an acceptable surrogate for monitoring vasoconstriction. The ANO with a fingertip PPG sensor is a widely used Food and Drug Administration–approved home sleep device that represents a reduced version of the clinical polysomnography. The vasoconstrictions measured at the fingertip are generally stronger than the corresponding events measured at the wrist, and thus the signal-to-noise ratio of the fingertip signal is higher than that of the wrist signal. Physiologically, this is to be expected since the fingers are highly innervated by alpha-adrenergic fibers and are also highly vascularized along with large numbers of arteriovenous anastomoses, compared to the wrist. This results in much faster and stronger vasoconstrictive responses in the fingers vis-à-vis the wrist. These differences in perfusion also have the potential to impact the cross-correlation values of the PPGampCV signals between the fingertip and the wrist. Using a cutoff of ≥0.5 as an acceptable correlation coefficient, we still see that there is mostly moderate to very strong correlation of PPGampCV between ANO and Biostrap. This shows that the wristband performs comparably to the “gold-standard” fingertip PPG while being much more practical to use. The overall correlation of PPGampCV for each night also remained statistically significant and acceptable between devices, suggesting that the wristband device remains a viable and feasible tool to deploy for longer-term, easy-to-use remote longitudinal monitoring of autonomic vasoreactivity.

This study is also unique in the sense that most wearable devices that purport to provide information about autonomic function focus on heart rate variability and SpO2 changes, but do not monitor changes in PPG amplitude, indicative of peripheral vasoconstriction [9]. Our Mvasoc algorithm and PPGampCV biomarker can be adapted to any wearable device that uses the PPG signal to measure ANS activity and quantify vasoconstriction in users with and without SCD. Promising preliminary data suggest a correlation between the Mvasoc parameter and the intensity of daily pain in patients with SCD. In 10 patients with SCD from this same cohort, we successfully tracked daily pain data using a SMS text message–based red cap survey, along with wristband recordings. Presence of pain and intensity of pain were tracked on a scale of 1 to 10. Of the 68 nocturnal wristband recordings with concurrent pain diary entries, 25 days of pain were reported with a pain intensity

XSL•FO
RenderX

ranging from 3 to 8. While the data were not powered toward the prediction of a VOC crisis given the limited number of nights of autonomic vasoreactivity monitoring, we found that a higher vasoconstriction index (Mvasoc) and a greater parasympathetic withdrawal in the heart rate variability patterns predicted greater intensity of pain the next day on multivariate regression analysis [19]. While these data hint at possible temporal associations between autonomic vasoreactivity patterns and pain, they need to be validated with longer-term longitudinal monitoring of vasoconstriction parameters in larger patient populations with SCD in order to develop it as a predictive tool for imminent VOC.

In conclusion, our pilot data show that the implementation of PPG from wearable sensors has a high potential for the development of an autonomic vasoreactivity biomarker for VOC in SCD. The portability, cost-effectiveness, and low maintenance of these wearable sensors can make them suitable for remote use. Moreover, these vasoconstriction parameters derived from PPG amplitude can serve as easily accessible and readily quantifiable biomarkers of tissue perfusion that could be valuable in other disease states, such as migraine and coronary artery disease, where vasoreactivity plays a role in disease pathology.

Limitations

The small sample size of nocturnal recordings, especially with the ANO device, makes it difficult to generalize the results. About 20% of the PPG signal from the nightly recordings was further eliminated due to loss of signal, primarily from the finger sensor. However, the high number of vasoconstriction events detected within each night of the available data adequately powers the correlation analyses and improves the robustness of the data. Both devices were not synchronized on the same time scale; therefore, we attempted to perform cross-correlation to 15-minute segments within a +30 or –30 second window. This technique can result in segment selection errors and lower correlation values. Future studies with concurrent use of the wristband device during clinical sleep studies with finger PPG sensors can be used to validate these results in a larger population. While we have shared preliminary results on the relationship between Mvasoc and daily pain intensity, the study is primarily focused on the ability to detect vasoconstriction events with wearable devices, and we were unable to derive any clinical inference on any causal relationship with VOC onset, the impact of hemoglobin levels, or the use of hydroxyurea on vasoreactivity.

Funding

This work was supported by grants from the National Institutes of Health; the National Heart, Lung, and Blood Institute (grant U01 HL117718); the Ming Hsieh Institute for Research of Engineering-Medicine for Cancer; and the Saban Research Institute at Children’s Hospital Los Angeles.

Authors' Contributions

PS collected and analyzed the data and wrote the manuscript. SS and MC analyzed the data. MCKK and TDC revised the manuscript. SV designed the study, collected and analyzed data, and revised the manuscript.

Conflicts of Interest

None declared.

References

1. Kato GJ, Piel FB, Reid CD, et al. Sickle cell disease. Nat Rev Dis Primers 2018 Mar 15;4:18010. [doi: 10.1038/nrdp.2018.10] [Medline: 29542687]
2. Coates TD, Chalacheva P, Zeltzer L, Khoo MC. Autonomic nervous system involvement in sickle cell disease. Clin Hemorheol Microcirc 2018;68(2-3):251-262. [doi: 10.3233/CH-189011] [Medline: 29614636]
3. Shah P, Khaleel M, Thuptimdang W, et al. Mental stress causes vasoconstriction in subjects with sickle cell disease and in normal controls. Haematologica 2020 Jan;105(1):83-90. [doi: 10.3324/haematol.2018.211391] [Medline: 30975906]
4. Veluswamy S, Shah P, Khaleel M, et al. Progressive vasoconstriction with sequential thermal stimulation indicates vascular dysautonomia in sickle cell disease. Blood 2020 Sep 3;136(10):1191-1200. [doi: 10.1182/blood.2020005045] [Medline: 32518948]
5. Khaleel M, Puliyel M, Shah P, et al. Individuals with sickle cell disease have a significantly greater vasoconstriction response to thermal pain than controls and have significant vasoconstriction in response to anticipation of pain. Am J Hematol 2017 Nov;92(11):1137-1145. [doi: 10.1002/ajh.24858] [Medline: 28707371]
6. Sangkatumvong S, Khoo MC, Kato R, et al. Peripheral vasoconstriction and abnormal parasympathetic response to sighs and transient hypoxia in sickle cell disease. Am J Respir Crit Care Med 2011 Aug 15;184(4):474-481. [doi: 10.1164/rccm.201103-0537OC] [Medline: 21616995]
7. Sangkatumvong S, Khoo MC, Coates TD. Abnormal cardiac autonomic control in sickle cell disease following transient hypoxia. Annu Int Conf IEEE Eng Med Biol Soc 2008;2008:1996-1999. [doi: 10.1109/IEMBS.2008.4649581] [Medline: 19163084]

XSL•FO
RenderX

8. Chalacheva P, Ji Y, Rosen CL, DeBaun MR, Khoo MC, Coates TD. Nocturnal peripheral vasoconstriction predicts the frequency of severe acute pain episodes in children with sickle cell disease. Am J Hematol 2021 Jan;96(1):60-68. [doi: 10.1002/ajh.26014] [Medline: 33027545]
9. Lubitz SA, Faranesh AZ, Atlas SJ, et al. Rationale and design of a large population study to validate software for the assessment of atrial fibrillation from data acquired by a consumer tracker or smartwatch: the Fitbit heart study. Am Heart J 2021 Aug;238:16-26. [doi: 10.1016/j.ahj.2021.04.003] [Medline: 33865810]
10. Kinnunen H, Rantanen A, Kenttä T, Koskimäki H. Feasible assessment of recovery and cardiovascular health: accuracy of nocturnal HR and HRV assessed via ring PPG in comparison to medical grade ECG. Physiol Meas 2020 May 7;41(4):04NT01. [doi: 10.1088/1361-6579/ab840a] [Medline: 32217820]
11. de Zambotti M, Rosas L, Colrain IM, Baker FC. The sleep of the ring: comparison of the ŌURA sleep tracker against polysomnography. Behav Sleep Med 2019;17(2):124-136. [doi: 10.1080/15402002.2017.1300587] [Medline: 28323455]
12. Mekhael M, Ho C, Noujaim C, et al. Compliance challenges in a longitudinal COVID-19 cohort using wearables for continuous monitoring: observational study. J Med Internet Res 2023 Apr 5;25:e43134. [doi: 10.2196/43134] [Medline: 36763647]
13. Biostrap. URL: https://biostrap.com/ [accessed 2024-03-21]
14. Alice NightOne. Philips. URL: https://www.documents.philips.com/doclib/enc/17329407/Alice_NightOne_Data_Sheet.pdf [accessed 2024-03-21]
15. Kemp B, Värri A, Rosa AC, Nielsen KD, Gade J. A simple format for exchange of digitized polygraphic recordings. Electroencephalogr Clin Neurophysiol 1992 May;82(5):391-393. [doi: 10.1016/0013-4694(92)90009-7]
16. Derrick TR, Thomas JM. Time-series analysis: the cross-correlation function. In: Stergiou N, editor. Innovative Analyses of Human Movement: Human Kinetics Publishers; 2004:189-205.
17. Characteristics of inpatient hospital stays involving sickle cell disease, 2000-2016. Agency for Healthcare Research and Quality. 2019. URL: https://hcup-us.ahrq.gov/reports/statbriefs/sb251-Sickle-Cell-Disease-Stays-2016.jsp [accessed 2024-03-21]
18. Espinoza J, Shah P, Veluswamy S, et al. Aura and mental stress are associated with reports of pain in sickle cell disease-a pilot study using a mobile application. Am J Hematol 2020 Apr;95(4):E101-E103. [doi: 10.1002/ajh.25736] [Medline: 31957078]
19. Veluswamy S, Shah P, Sy S, et al. Nocturnal autonomic and vasoconstriction parameters by wearable wrist photoplethysmography predict intensity of imminent pain in subjects with sickle cell disease. Blood 2023 Nov 2;142(Supplement 1):3875. [doi: 10.1182/blood-2023-187353]

Abbreviations

ANO: AliceNightOne
ANS: autonomic nervous system
Mvasoc: magnitude of vasoconstriction
PPG: photoplethysmography
PPGamp: PPG amplitude
PPGampCV: PPGamp coefficient of variation
SCD: sickle cell disease
SpO2: oxygen saturation
VOC: vaso-occlusive crises

Edited by A Malik; submitted 03.Apr.2025; peer-reviewed by J Azai, NI Shuzan; revised version received 15.Aug.2025; accepted 01.Dec.2025; published 06.Jan.2026.

Please cite as:
Shah P, Sy S, Chen M, Khoo MCK, Coates TD, Veluswamy S
Wearable Device Photoplethysmography As a Viable Tool to Longitudinally Monitor Vasoconstriction Biomarkers for Predicting Vaso-Occlusive Crisis in Sickle Cell Disease: Feasibility and Validation Study
JMIR Hum Factors 2026;13:e75465
URL: https://humanfactors.jmir.org/2026/1/e75465
doi:10.2196/75465

© Payal Shah, Sabrina Sy, Mingjing Chen, Michael CK Khoo, Thomas D Coates, Saranya Veluswamy. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 6.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use,

XSL•FO
RenderX

distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Opportunities for Improved Device Design Based on Central Line Placement Practices: Contextual Inquiry Study

Mary Beth Privitera[1], BSID, MDES, PhD; Sameer Khan[2], MD; Bilal Irfan[3,4], MS; Shayan Ali[5,6]; Cecelia Arredondo[7], MDES; Kyrsten Sanderson[7], MDES; Jordan Bonomo[8], MD

[1]University of Cincinnati, Know Why Design, LLC, Mason, OH, United States

[2]Department of Critical Care Medicine, Hoag Hospital, Newport Beach, CA, United States

[3]Department of Epidemiology, Department of Neurology, University of Michigan Medical School, Ann Arbor, MI, United States

[4]Center for Surgery and Public Health, Brigham & Women's Hospital, Center for Bioethics, Harvard Medical School, Boston, MA, United States

[5]Neuroscience Department, Austin College, Sherman, TX, United States

[6]North Texas Medical Research Institute, Rockwell, TX, United States

[7]School of Design, College of Design, Art, Architecture & Planning, University of Cincinnati, Cincinnati, OH, United States

[8]Section Chief, Critical Care, Department of Emergency Medicine, Neurology, NeuroCritical Care, University of Cincinnati, Cincinnati, OH, United States

Corresponding Author:
Mary Beth Privitera, BSID, MDES, PhD
University of Cincinnati
Know Why Design, LLC
4691 Saddletop Ridge Lane
Mason, OH, 45040
United States
Phone: 1 5132585824
Email: privitmb@ucmail.uc.edu

Abstract

Background: Central venous catheters (CVCs) are indispensable to contemporary critical care, perioperative management, and emergency resuscitation, yet their insertion remains fraught with preventable harm and inefficiency.

Objective: This study aimed to identify all areas of CVC placement that can be improved through device design using human-centered design and qualitative research methods.

Methods: This qualitative study was a contextual inquiry of CVC placement, which included observation alongside brief face-to-face interviews with physicians. It was aimed at providing a depth of understanding using evidence to demonstrate causality. This study was conducted at 3 hospitals in the emergency department, the intensive care unit, and the operating rooms. Where possible and with additional consent, sessions were recorded in video or still photography, or at times both. This study included 19 observations and 24 interviews.

Results: In this study, the approach to CVC insertion was consistent across hospitals and care environments, with moderate variability spanning a few sections, such as suture and dressing use or lack thereof in specific care environments. The described and observed difficulties leave room for improvement in device design. The results of this study indicated that there are 34 discrete steps to placing a CVC line, with most time spent during sterile preparation. As a result of the device or kit design, challenges were observed. These included missing essential materials from kits, difficulty distinguishing between nonsterile and sterile items, challenges with lidocaine ampules, patient claustrophobia from draping, and a lack of user preference for kit contents. Additional challenges included obscured ultrasound views, kinked guidewires, overall procedural untidiness, and considerable waste management issues.

Conclusions: An intuitive kit that aligns with predictable human behavior and eliminates unnecessary multistep detours can reduce novice failure rates, cognitive load, and practice inconsistency, and it could also curb nonrecyclable waste from "backup" kits opened for a single missing item. By reframing CVC systems as sociotechnical solutions rather than static assortments of parts, the same design moves that minimize improvisation and coordination errors for physicians may also reduce dwell time and manipulation events for patients, thereby advancing the core triad of safety, procedural efficacy, and everyday usability. By examining how clinicians place central lines, this study reveals modifiable design flaws that perpetuate risk despite decades of

XSL•FO
RenderX

procedural standardization. Contextual inquiry provides the evidentiary bridge between clinical imperatives to reduce complications and the practical realities of device use. Embedding such investigations at the outset of design and iteratively throughout product life cycles offers a path toward safer, more efficient, and more humane central venous access for both patients and providers.

(JMIR Hum Factors 2026;13:e84621) doi:10.2196/84621

KEYWORDS

contextual inquiry; central line placement; medical device design; patient safety; ease of use

Introduction

Central venous catheters (CVCs) are indispensable to contemporary critical care, perioperative management, and emergency resuscitation, yet their insertion remains fraught with preventable harm and inefficiency. Large contemporary syntheses estimate that approximately 30 in every 1000 patients with a CVC in place for 3 days will experience at least one serious complication, be it arterial cannulation, pneumothorax, central line–associated bloodstream infection (CLABSI), or deep venous thrombosis, whereas catheter malfunction alone occurs at roughly 6 events per 1000 catheter days [1]. Additionally, the insertion process can be a source of complication. CLABSIs still claim thousands of lives each year in the United States and add billions of dollars to health care costs despite decades of prevention bundles and increasingly routine ultrasound guidance intended to curb mechanical mishaps and accelerate successful cannulation [2].

While epidemiological surveillance and randomized trials may have helped refine insertion checklists and sterility protocols, far less attention has been paid to the concrete interactions among users, tools, and environments that shape everyday practice. Human factors frameworks may be an impetus to study actual work as done rather than work as imagined. The US Food and Drug Administration's 2016 guidance on applying human factors in medical devices and the Association for the Advancement of Medical Instrumentation TIR51 standard on contextual inquiry may provide the grounding for early, field-based observation and interview techniques to surface latent use errors, cognitive burdens, and design mismatches before devices reach market or are iterated for safety [3,4].

This study applied a structured contextual inquiry across 3 hospitals and multiple care settings to showcase how central line placement unfolds in real-world settings minute by minute, from presterile preparation to dressing application. By triangulating in situ observation with brief clinician interviews, 34 discrete steps were catalogued, mapping procedure time distributions and documenting recurrent friction points that compromise sterility, ergonomics, and situational awareness. This study aimed to identify all areas of CVC placement that can be improved through device design. From the drapes that obscure patients' faces and guidewires that become kinked to the missing of essential or preferred components, this study highlights the basis for next-generation kit architecture, accessory design, and room layout that align with real user needs in both chaotic and controlled environments.

Methods

Overview

This qualitative study was a contextual inquiry of CVC placement, which included observation alongside brief face-to-face interviews with physicians. It was aimed at providing a depth of understanding using evidence to demonstrate causality. According to Maxwell [5,6], qualitative research is well suited for causal inference as it allows for detailed examination of specific processes and mechanisms in real-world contexts, revealing how and why outcomes occur beyond mere correlations [5,7]. A detailed description of the process, tools, and people involved is provided below. The technique of contextual inquiry is promoted by the Food and Drug Administration human factors guidance (2016) to determine user needs at the start of any design process. The data collected in this study included both observational and interview data, with 3 main areas of focus: the user, the environment, and the tasks as part of the steps and workflow.

This study was conducted at 3 hospitals in the emergency department (ED), intensive care unit (ICU), and operating rooms (ORs). Where possible and with additional consent, sessions were recorded on video or still photography, or at times both. This study included 19 observations and 24 interviews (Table 1). The total number of observations and interviews conducted at each site was determined by the clinical need and patient-clinician consent at the time of the study. In instances in which patient consent was not provided but physicians wanted to participate, the physicians could opt to demonstrate their placement technique using a simulated patient in the care environment.

Table 1. Breakdown of observations and interviews by location.

	Cincinnati, Ohio, n (%)	Wake Forest, North Carolina, n (%)	Sacramento, California, n (%)
Observations (n=19)	9 (47)	7 (37)	3 (16)
Interviews (n=24)	8 (33)	8 (33)	8 (33)

A total of 24 physicians were involved in this study, including emergency medicine, critical care, and anesthesiology specialists. Observations and interviews were conducted in perioperative rooms, ORs, EDs, and ICUs.

XSL•FO
RenderX

Within each hospital, the same brand of kit was used; however, in one instance at Wake Forest, there was a different type of kit used between the OR and the ICU and ED. The kits consisted of other brands and product identification numbers (Table 2).

Table 2. Brand of kits used by location and environment of care.

	Cincinnati, Ohio	Wake Forest, North Carolina	Sacramento, California
Operating room	Edwards Multi-Med CVC[a] 3K20N18141NL	Arrow central venous access kit ASK-21242-PCMH1	Arrow Blue Plus pressure-injectable multilumen CVC kit ASK-45703-PIO
Intensive care unit	Edwards Multi-Med CVC 3K20N18141NL	Arrow pressure-injectable multilumen CVC kit CDC-45703-XP1A	Arrow Blue Plus pressure-injectable multilumen CVC kit ASK-45703-PIO
Emergency department	Edwards Multi-Med CVC 3K20N18141NL	Arrow pressure-injectable multilumen CVC kit CDC-45703-XP1A	Arrow Blue Plus pressure-injectable multilumen CVC kit ASK-45703-PIO

[a]CVC: central venous catheter.

All interviews followed the core principles of contextual inquiry studies in that they were conducted in the users' real-world environment, with the research team establishing a master (physician)-apprentice (researcher) relationship. As the procedure unfolded, and at opportune times, the researcher would share interpretations to uncover observations and deeper insights. All the interviews were guided by the tasks in the CVC placement procedure and focused on the usability of each element. Typical observation or interview experiences ranged from 45 minutes to 1.5 hours, as influenced by clinical responsibilities and the participants' availability.

Ethical Considerations

The study protocol was appropriately reviewed by the University of Cincinnati institutional review board, which determined that it did not constitute human participant research under federal regulations, as the primary focus was on device improvements and quality of care rather than generating generalizable knowledge about individuals. Despite the nonresearch classification, informed consent was required from all participants. As detailed observations were conducted of clinical procedures that could capture sensitive patient information or professional performance details, robust measures were in place to protect privacy and confidentiality. In addition, efforts were made to ensure a diverse participant pool to avoid biased insights that could perpetuate disparities in device design or procedural improvements.

Results

In this study, the average overall time to place a CVC varied across care environments (Table 3). In discussions with clinicians, some postulated that procedure times vary proportionally with the provider's perceived control over the patient and environment (ie, that the more chaotic the environment, the longer the procedure).

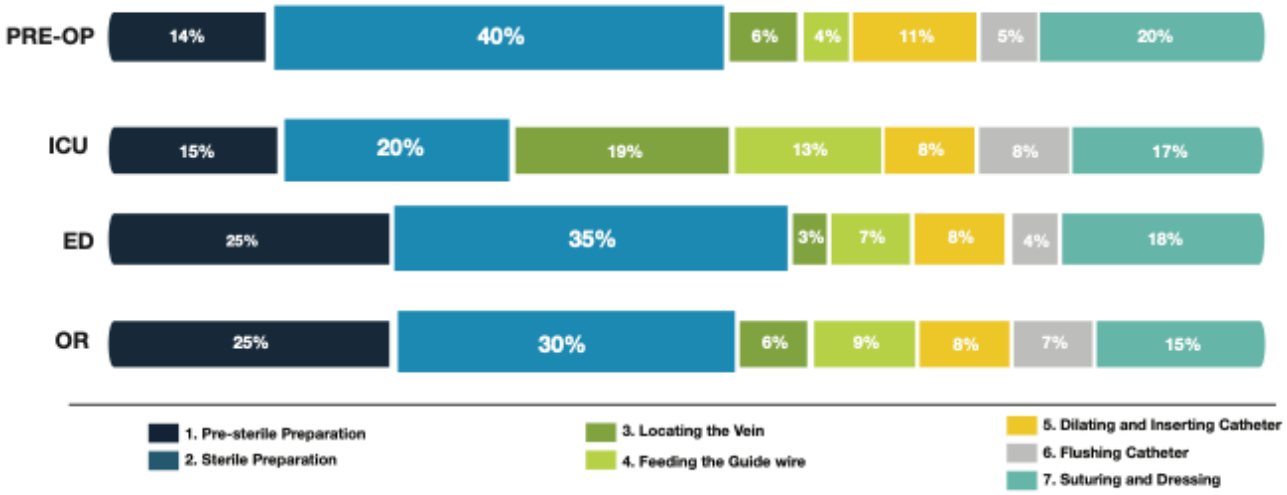
Table 3. Average central venous catheter (CVC) placement timing in different environments of care.

Environment	CVC placement time
Operating room	16 min, 48 s
Preoperative care	20 min, 15 s
Intensive care unit	27 min, 53 s
Emergency department	35 min, 28 s
Overall, mean (SD)	25 min, 6 s (7 min 12 s)

Procedural timing was also broken down into 7 distinct sections within the overall procedure (Figure 1). These included the following: presterile preparation, sterile preparation, vein localization, guidewire insertion, catheter dilation and insertion, catheter flushing, and suturing and dressing. Of note, time from presterile to sterile preparation accounted for more than half of the overall procedure time in all locations except the ICU. The research team noted that the number of staff members participating in setting up for the procedure varied across sites and care environments.

XSL•FO
RenderX

Figure 1. Percentage of time per section of the central venous catheter placement procedure by environment. ED: emergency department; ICU: intensive care unit; OR: operating room.



The patient experience during CVC placement highlights conditions that impact overall safety, efficacy, and usability. Draping a conscious patient can be problematic. During procedure preparation, the provider unpacks sterile drapes from a kit and prepares a sterile field directly over the patient (Figure 2). This extends the sterile field, albeit an unstable one, directly over the patient. As a result of patient movement, the required equipment for the procedure can fall out of the sterile field or become lost. Uncooperative patients require additional procedures, such as intubation, muscle relaxation, and sedation, before CVC placement.

Figure 2. Patient acting as an extended sterile work surface.



In the OR, the patients were intubated and unconscious. However, in the ED and ICU environments, the patients were often conscious and mobile. In all cases, draping covered the patient's face throughout the procedure, creating a potentially claustrophobic and challenging environment (Figure 3). In clinical practice, there is substantial variability in body habitus across the patient population. As a result, the drapes provided in the kits may not fit patients classified as obese. In some cases, the aperture provided by the drape is not large enough to accommodate the insertion location, resulting in the provider improvising by creatively engineering a larger opening.

XSL•FO
RenderX

Figure 3. A patient drape placed over the patient's face in the intensive care unit.



Patient airways require management throughout the procedure. In this study, patients in the ICU were fitted with an oxygen mask. A mask is often placed in case a future procedure is needed to avoid disturbing the sterile field.

Gaining access and positioning can often lead to patient discomfort. Patients are required to hold awkward Trendelenburg positions with their head turned and extended to optimize their anatomy for successful CVC placement (Figure 4). Patients were instructed to maintain position throughout the procedure, although it can be uncomfortable because the plastic portion of the drape may be directly on a patient's skin. For conscious patients, this at times required further anesthesia and posed challenges for physicians, given that, as access was gained, patients would move. This position proves challenging for patients who are obese who may also require additional supplies and more extended periods to complete procedure preparation.

Figure 4. Aperture opening of the drape adhered directly to a patient's face.



The provider experience during CVC placement also underlines conditions that could impact overall safety, efficacy, and usability. Physicians are required to don personal protective equipment in tight spaces without compromising sterility (Figure 5). In this study, the order of preparation varied across care locations. For example, ED physicians sometimes elected to preclean the patient before gowning, which allowed time for the chlorhexidine preparation to dry. In contrast, in the OR, some physicians prepared personal protective equipment in advance.

XSL•FO
RenderX

Figure 5. Donning of personal protective equipment in a tight space.



In the CVC procedure, 34 individual steps require the provider to locate key pieces of equipment and determine where and how to use them while maintaining a sterile workspace. During this multistep process, physicians often improvise, which may place a heavy cognitive burden on them and require coordination. For example, physicians will think through the procedure before beginning, questioning whether they have all the necessary supplies and where to place them to maximize access while maintaining the sterile field and coordinating with other support personnel.

Across all sites and use environments, at least 2 physicians were present. This enabled 1 sterile operator and 1 support person to adjust patient drapes, including using a towel to protect the patient's face; retrieve items that were required but not readily available in the kit or forgotten; and sterilely prepare the ultrasound probe.

Physicians were often in awkward positions due to suboptimal biomechanics resulting from product design, room layout, or both. For example, physicians struggled to fit all the necessary equipment on a single surface due to space constraints. The equipment is often split between multiple work surfaces (ie, patient and table), forcing the provider to turn to access both surfaces and remove their focus from the patient. The space constraints also make it difficult to avoid touching a nonsterile surface during the procedure. Furthermore, flushing with a standard syringe requires an uncomfortable and awkward hand movement needed to achieve aspiration, as the device is designed to deliver a solution. When ultrasound is used for safe access, both hands are required to hold the equipment, which can make aspiration a challenging one-handed maneuver (Figure 6).

XSL•FO
RenderX

Figure 6. One-handed aspiration while holding the ultrasound probe.



Physicians held the syringe and introducer needle at a 45 to 30° angle and subsequently lowered the angle while advancing the needle (Figure 7). To verify position, physicians would draw more blood and assess color and pressure gradient. Some commented that some step was the highest “stress” point in the procedure. One intensivist even remarked that finding the vein was the most stressful part and that, once the wire was in and the needle was out, their stress subsided. At this point, there is a risk of pneumothorax or perforation of the carotid artery. It was observed that maintaining stability and aspirating to visualize venous blood return is ergonomically challenging for physicians. This is further complicated by the ultrasound monitor being placed outside the provider’s field of view (Figure 8).

Figure 7. Access needle angle relative to the ultrasound probe.



XSL•FO
RenderX

Figure 8. Ultrasound monitor position requires physicians to look away from the access site.



Additionally, the pressure that the user applies to the vein with the ultrasound probe can alter central venous anatomy when the device is withdrawn from the skin surface. Physicians were at times compelled to make slow, careful movements, with risk controlled by the user's experience. Rather than an intuitive device design, advanced troubleshooting or error prevention during the procedure may depend on the provider's expertise and dexterity.

While simultaneously troubleshooting and preventing errors, physicians are exposed to sharps and biohazards once access to the vein is achieved. For example, users are exposed to the access needle, where the skin incision scalpel is typically contaminated with the patient's blood within the working area (Figure 9).

Figure 9. Opening the kit midprocedure with the sharps anchored in the kit.



While efforts were made to control and maintain the safety of sharps, the CVC kit design places the cognitive burden, safety, and responsibility on the provider. As vein dilation occurs, the access site often becomes highly disordered and poses a biohazard to health care physicians (Figure 10). Once the vein was dilated, increased blood flow increased the stress and inconvenience of the procedure in that the guidewire itself became a slippery surface. It could become misplaced, requiring additional intervention. This hazardous situation also extends to catheter insertion (Figure 11).

Physicians were required to thread the catheter over the guidewire and adopted a coiling behavior to maintain control and reduce proximal tip movement, which can complicate device alignment (Figure 12).

XSL•FO
RenderX

Figure 10. Advancement of the dilator over the wire in slippery conditions due to venous return.

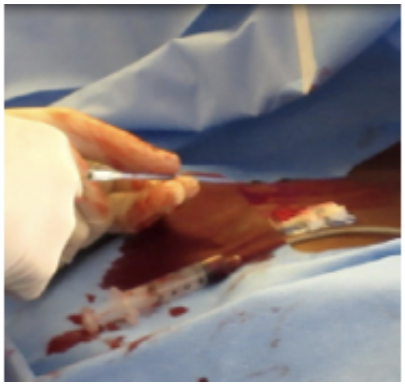


Figure 11. Catheter insertion with exposure to biohazard.



Figure 12. Coiled guidewire enabling control during catheter insertion.



These situations were particularly challenging for novice health care providers, who grappled with each step of the procedure (Figure 13). Consequently, inexperienced health care physicians often have unsuccessful attempts to place CVCs.

The product design of traditional CVC kits was not always entirely optimized for streamlined assembly. Many kits, for example, did not have all the supplies or the preferred supplies for CVC placement. As a result, clinicians often spent additional time gathering the necessary materials (eg, skin preparation, gauze, saline, and lidocaine) as well as the preferred materials because the valves provided in the kits were at times not compatible with other hospital equipment, thus suggesting better packaging so that the risk or delay in the procedure is minimized. Some users preferred vertical valves such as clave connectors (Figure 14). Physicians who chose to use nonstandard or additional supplies often opened a secondary kit sourced from separate locations within the hospital. Items collected may include additional drapes, Mayo stand, lidocaine, sterile flush syringes, chlorhexidine preparation, and ultrasound devices or supplies.

XSL•FO
RenderX

When physicians remove the guidewire, there is no proper receptacle to place it in; it is thrown away with nonsharp materials or hazardous materials. In this study, some physicians did not use sharps holders during the procedure, preferring to leave sharps on the tray (Figure 15), suggesting an unmet need among users. For those who used the temporary sharps holder provided in some trays, it was poorly secured and became top-heavy, often leading to dropped sharps. Additionally, there were no safeguards for the disposal of all sharps.

Figure 13. Novice users have increased exposure to biohazards.



Figure 14. Different forms of valve connectors.



Figure 15. Sharps are placed on top of the kit during the procedure.



There is a considerable amount of nonrecyclable waste and challenges because of the kit design (Figure 16). In this study, the kits at times lacked a clear delineation between the layers that designate items and specify when they should be used. There were excess materials not used in the procedure, which contributed to the overall waste. Physicians often draped the sterile area and manipulated the drape around the patient's face and head, making it easier for the patient to breathe and reducing claustrophobia.

XSL•FO
RenderX

Figure 16. Waste accumulation because of the kit design.



The quality and design of the CVC products' individual elements pose usability challenges, including misalignment with predictable human behavior, heavy reliance on memory, and a lack of visual or tactile references. One challenge in this study was that the guidewire kinked frequently. When this happened, physicians would often try to remove the kink or use the wire regardless. If this guidewire is rendered unusable, the provider may open a new kit to replace the failed one or, in some instances, a missing part (Figure 17).

Figure 17. Three kits opened for the guidewire and dilator due to kinking.



In this study, the dilator was the same color as the draping and could be camouflaged, eventually getting lost in it, causing health care providers to search. Additionally, due to the complex setup, components sometimes fell outside of the sterile field. Another detail that caused some difficulty in the guidewire design was the "J" curve. While it can provide safe intravascular advancement, this curve on the wire can also make the guidewire more difficult to thread into the catheter. Some physicians were observed making what some may describe as counterintuitive workarounds and flipping the guidewire to the straight end to easily advance it through the catheter. Though well intended, this variation in CVC placement may increase the risk of complications.

Furthermore, there is little to no tactile feedback from the guidewire feeder, making it difficult for the user to know whether the wire is advancing. This is further complicated by blood-stained gloves, which make the guidewire markings difficult to see and grasp. The results of these outdated design decisions can create an unnecessary burden on physicians to judge catheter alignment on the guidewire and insertion depth.

Discussion

Principal Findings

In this study, the approach to CVC insertion was consistent across hospitals and care environments, with moderate variability spanning a few sections, such as suture and dressing use or lack thereof in specific care environments. The described and observed difficulties leave room for improvement in device design (Table 4).

XSL•FO
RenderX

Table 4. Summary of use-related issues throughout the procedure.

Procedure section and device used	Problem area
Presterile preparation	
Central venous catheter kit	• It does not include all essential materials needed for central line placement, resulting in unnecessary delays. • It is unable to distinguish nonsterile and sterile items during nonsterile preparation.
Nonsterile syringe	• Health care providers typically empty 3-5 syringes (30-50 mL) of saline solution into the kit basin from nonsterile saline syringes. These were not included in the kit.
Lidocaine	• Because the lidocaine is not sterile, health care providers will either have someone hold it for them, tape it to a Mayo cart, or chase down the ampule while loading the syringe.
Sterile preparation	
Patient drape	• Due to the placement of the full-body drape, patient claustrophobia can occur; it is challenging for the provider to maintain sterility and manage the patient during the procedure.
Locating the vein	
Introducer needle	• Physicians expressed a preference for the 18-gauge introducer needle without a catheter over the 20-gauge needle with an 18-gauge catheter assembly. For patients with obesity, physicians noted that the length (2.5 inches) of the needle was not sufficient. The introducer needles were not echogenic; thus, they were difficult to view in the ultrasound monitor.
Ultrasound	• The ultrasound monitor was not placed within the line of sight of the provider, making it difficult for them to effectively use the tool. • Health care providers do not generally document ultrasound use by capturing an image of the procedure as required by major insurers for reimbursement.
Feeding the guidewire	
Guidewire	• Wire kinking while executing the procedure was problematic, and once deemed unusable, the provider would open another kit only to access an additional wire. • The “J” tip of the wire was difficult to thread and feed. • Physicians had difficulty estimating how much of the wire was inserted. If they were unsure, they took the wire out of the delivery system. Once the wire is removed from the delivery system, it is difficult to reload; thus, once it is out, it typically stays out. Markings on the wire were also not clear to the providers. • Due to the inherent messiness of the procedure, blood is often on the providers’ hands, making it difficult to handle the guidewire delivery system.
Dilating and inserting the catheter	
Dilator	• The dilator often kinks if the provider aggressively pushes. If the dilator is no longer usable, another kit is opened only to access another dilator. • The dilator length (4 inches) is not sufficient for patients with obesity.
Catheter	• Threading the catheter over the wire is a difficult task. Physicians have a hand tremor when performing this task. • It is difficult for the provider to determine whether the tip of the catheter has reached the superior vena cava. Physicians will estimate placement and follow up by x-raying the patient to confirm proper placement.
Flushing the catheter	
Syringe	• Physicians faced challenges with maintaining proper syringe grip during aspiration and flushing. Physicians struggle with the current syringe design to alternate between 2 grips.
Valves	• The T-shaped valves provided in the kit also present challenges when flushing the line after placement. The syringe often slips off the valve when flushing the lines.
Suturing and dressing placement	
Suture loop (feature) on the catheter	• The suture loop anchor and box clamp were not always used. If there is excess catheter outside of the patient’s body, depending on the insertion location and the patient’s anatomy, the provider will use the suture loop anchor and box clamp.

This contextual inquiry reveals that central venous catheterization is not merely a sequence of technically codified steps but also an intricate sociotechnical performance in which device design, environmental constraints, and clinician

XSL•FO
RenderX

improvisation intersect to shape risk. Our mapping of 34 discrete actions, the dominance of preparation time in most settings, and the recurrent need to compensate for missing or poorly engineered components align with contemporary estimates that roughly 3% of patients exposed to a catheter for 3 days sustain a major complication [1]. Large contemporary reviews showcase that CVC is now routine, roughly 8% of hospitalized patients need a CVC, and more than 5 million are placed annually in the United States. Ultrasound-guided puncture should be the default. However, Kehagias et al [8] found that observations of obscured monitors, awkward one-handed aspiration, and improvised wire handling show how the safest technique on paper can still be undermined by poor ergonomics and kit design.

Our findings build on imperatives for field-based inquiry into real work conditions to surface latent design hazards. We observed exactly the kinds of use errors that are of concern: kinking guidewires without tactile feedback, dilators camouflaged against drapes, valves incompatible with existing hospital hardware, and ultrasound displays positioned outside the operator's natural sightline. Such mismatches between device affordances and predictable human behavior may shift cognitive burden to clinicians, who must remember workarounds and coordinate ad hoc assistance while maintaining sterility. A contextual inquiry approach is designed to elicit these mismatches; our study demonstrates its value in an acute, invasive procedure where seconds and millimeters matter. The imperative to integrate these insights into formal design controls is clear.

Several concrete implications for device and kit redesign emerge from our contextual findings and accord with human factors guidance that interventions must fit users' capabilities, workflows, and environments to sustain adoption and fidelity [9-12]. The suggested improvements and justifications are mentioned subsequently.

First, rather than perpetuating one-size-fits-all assortments, kits could be organized and sufficiently complete for the intended procedure and patient mix, thus minimizing secondary searches and redundant openings so that routinely needed sterile flushing media, local anesthetic, echogenic introducer needles of adequate length for patients classified as obese, and valve connectors compatible with local infusion hardware become immediately available; these specific component choices derive from our observations, whereas the general mandate to reduce cognitive and physical load is articulated in human-centered design literature [9,12].

Second, because the procedure advances in a sequential order, CVC kit components should be packaged in the same order to reduce cognitive burden. High-risk elements could incorporate salient visual and tactile affordances such as blood-tolerant depth markings on guidewires, textured feeders that signal advancement, and color-contrasted dilators that cannot visually disappear against drapes, an approach consistent with human factors recommendations to engineer cues that support rapid, accurate action under stress [12].

Third, ultrasound ergonomics require particular attention: practice reviews emphasize that the operator should keep the puncture site, needle, and image within a single line of sight, yet we repeatedly observed monitors positioned laterally or behind the user, encouraging awkward posture and one-handed aspiration [9-11].

Fourth, because maintaining stability of the introducer needle at the insertion site is critical, there is an opportunity for a device that minimizes hand movement or stabilizes the ultrasound probe during vessel access.

Fifth, converging evidence from simulation and ergonomic assessments shows that suboptimal screen and table positioning increases musculoskeletal strain and facilitates needle advancement errors, especially among novices, supporting the design of articulating mounts, probe-holding accessories, and workstation layouts that free the dominant hand and keep the image within the operative field [10-12].

Sixth, additionally, as insertion components are introduced sequentially, combining the components associated with adjacent actions into a single integrated unit would significantly simplify the procedure.

Seventh, as threading the dilator over the guidewire is related to a risk of kinking, a device may reduce or eliminate the need for over-the-wire threading while reducing the risk of guidewire kinking.

Eighth, patient experience is also inseparable from considerations of safety. Drapes that blanket a conscious patient's face created anxiety and impeded airway access, whereas Trendelenburg positioning with head rotation was difficult to sustain, particularly for individuals classified as obese, who then required additional supplies and time. Our observations suggest that rethinking drape architecture to permit facial exposure without compromising sterility and integrating oxygen delivery ports or transparent windows could mitigate claustrophobia and facilitate airway monitoring.

Because CLABSI remains among the costliest hospital-acquired infections on a per-case basis, the marginal gains from such design tweaks may translate into substantial economic and human benefit when scaled across millions of annual insertions [13]. Even in the era of prevention bundles, catheter dwell time remains a potent infection driver; CLABSI rates in one 2-year adult cohort climbed from 4.80 to 8.64 per 1000 catheter days as dwell exceeded 20 days, with multidrug resistant *Acinetobacter baumannii* predominating, mirroring some of our concerns that missing components, repeated kit openings, and ad hoc maintenance steps prolong setup and line life, thereby compounding exposure to contamination [14]. It has been previously emphasized that complication profiles can hinge on site choice, catheter caliber, and positioning—adult data (eg, 3SITES study) link subclavian access to fewer infections but more pneumothoraxes, whereas pediatric series show different risk patterns—and it has been reiterated that neutral rather than exaggerated positioning and meticulous ultrasound use can reduce failed passes and arterial hits, resonating with our field notes on stressful needle advancement angles, off-axis screens, and repeated punctures at the point deemed to be of the highest stress [15]. Other recommendations also somewhat align with our design proposals: select the smallest necessary lumen count;

XSL•FO
RenderX

favor nontunneled catheters for fewer than 3 to 4 weeks and peripherally inserted central catheters when therapy exceeds 6 days; cap prolonged catheterization (approximately 14 days) to curb bacteremia; and mandate real-time ultrasound, chlorhexidine alcohol preparation, and daily site surveillance [16]. These were practices that our contextual inquiry somewhat related to and found were variably executed or actively hindered by kit incompleteness, unclear layer sequencing, and drapes that compromise both sterility and patient comfort. In a simulation of 40 anesthesia providers, patient safety (mean importance score 83.9/100), ease of use (mean score 64.6/100), and reduced clinician error (mean score 61.1/100) topped the selection criteria. A novel CVC system with a sequentially organized tray, enhanced labeling, and a guidewire funnel earned significantly higher scores for satisfaction overall, ease of use, layout, and safety ($P \leq .01$ in all cases) and reduced 5 of 7 common risks (including clinician error and contamination or infection), aligning with the usability deficits (wire kinking, component hunting, and ambiguous tray hierarchy) that our study catalogued [7].

This study also highlights how waste and sharps handling were downstream consequences of kit design. The absence of designated receptacles for used guidewires and scalpels, the lack of sharps holders integrated into trays, and the routine opening of multiple kits for a single missing part create biohazard exposure and substantial nonrecyclable waste. While our qualitative approach did not quantify environmental impact, other health systems work has emphasized the financial and ethical importance of reducing unnecessary disposables; therefore, future prototypes should embed closed-loop sharps to capture and minimize redundant components to support infection prevention and sustainability goals simultaneously. This inference from our data warrants targeted life cycle and cost analyses in subsequent studies.

Viewed through the intersecting lenses of safety, efficacy, and usability for patients, providers, and products, our findings provide the basis for an argument that a device ecosystem should simultaneously consider the precepts mentioned subsequently.

The first precept is to shield patients from infection, vessel injury, and the cascade of “extra” procedures (intubation and deep sedation) by shortening setup and puncture time and by making correct J-wire orientation and tip control essentially foolproof.

The second precept is to protect clinicians from sharps and blood under tension, awkward postures, and protocol drift by embedding ergonomic grips; one-handed aspiration aids; and closed, labeled receptacles that keep contaminated instruments off ad hoc trays.

The third precept is to streamline the product itself with sequential, memory-light assembly; color and texture coding; and built-in safeguards (eg, wire funnels, depth markings visible through bloodied gloves, and lockable sharps wells) that are robust and attuned for real-world conditions such as low light; urgent timelines; moving patients; and gloved, fatigued hands. In effect, an intuitive kit that aligns with predictable human behavior and eliminates unnecessary multistep detours can reduce novice failure rates, cognitive load, and practice inconsistency, just as it could serve to curb nonrecyclable waste from “backup” kits opened for a single missing item.

By reframing CVC systems as sociotechnical solutions rather than static assortments of parts, the same design moves that minimize improvisation and coordination errors for physicians also reduce dwell time and manipulation events for patients, thereby advancing the core triad of safety, procedural efficacy, and everyday usability.

Limitations

Our findings should be interpreted considering several limitations. This study was conducted in 3 US hospitals and included 19 observations and 24 interviews, which constrains generalizability, particularly to resource-limited settings or institutions that use different kit vendors. Although we captured real procedures, observer presence and video recording may have altered behavior, and some demonstrations on simulated patients cannot fully reproduce the stressors of an emergent cannulation. We did not measure clinical outcomes linked to the specific use problems we identified, nor did we quantify cognitive load or musculoskeletal strain. Furthermore, our sample comprised physicians; nurses, technicians, and infection preventionists often influence setup, maintenance, and postplacement care, and their perspectives warrant inclusion in a fuller systems analysis.

Conclusions

Future work should transition from description to intervention. Rapid-cycle prototyping informed by these contextual insights, followed by high-fidelity simulation and usability testing in accordance with relevant guidelines, could generate validated design changes. Eye tracking, motion capture, and workload assessment tools may quantify how redesigned components alter gaze patterns, posture, and error rates. Multicenter trials comparing integrated, human-centered kits with current products could measure effects on insertion time, complication rates, waste generation, and cost. Furthermore, translating these methods to maintenance phases, including line access and dressing changes, could extend human factors benefits beyond placement to the complete catheter life cycle.

By examining how clinicians place central lines, this study reveals modifiable design flaws that perpetuate risk despite decades of procedural standardization. Contextual inquiry provides the evidentiary bridge between clinical imperatives to reduce complications and the practical realities of device use. Embedding such investigations at the outset of design and iteratively throughout product life cycles offers a path toward safer, more efficient, and more humane central venous access for both patients and providers.

XSL•FO
RenderX

Acknowledgments

The authors extend their special thanks to the University of Cincinnati, Department of Emergency Medicine, for their support in making this research possible. Becton Dickinson provided funding support for this research.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to privacy and confidentiality purposes.

Authors' Contributions

The data collection and initial analysis of this contextual inquiry research were led by JB and MBP, with senior direction from KS and CA.

Conflicts of Interest

None declared.

References

1. Teja B, Bosch NA, Diep C, Pereira TV, Mauricio P, Sklar MC, et al. Complication rates of central venous catheters: a systematic review and meta-analysis. JAMA Intern Med 2024 May 01;184(5):474-482. [doi: 10.1001/jamainternmed.2023.8232] [Medline: 38436976]
2. Central Line-associated Bloodstream Infection (CLABSI) basics. Centers for Disease Control and Prevention. 2025 Jun 12. URL: https://www.cdc.gov/clabsi/about/index.html [accessed 2025-07-24]
3. Applying human factors and usability engineering to medical devices. U.S. Food & Drug Administration. 2016 Feb. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/applying-human-factors-and-usability-engineering-medical-devices [accessed 2025-07-24]
4. AAMI TIR51:2014 (R2017): human factors engineering - guidance for contextual inquiry. ANSI. 2014. URL: https://webstore.ansi.org/standards/aami/aamitir512014r2017?srsltid=AfmBOopt6jPd7hRIApqzSnyRv9ZCXa_swEE4FWhr_lpW2e3e-2bm-s1B [accessed 2025-07-24]
5. Maxwell JA. Causal explanation, qualitative research, and scientific inquiry in education. Educ Res 2004 Mar 01;33(2):3-11. [doi: 10.3102/0013189x033002003]
6. Maxwell JA. A Realist Approach for Qualitative Research. Thousand Oaks, CA: SAGE Publications; 2012.
7. Wagner R, Alsbrooks K, Bacon E, O'Brien M, Hoerauf K. Clinician satisfaction and preference for central venous catheter systems promoting patient safety, ease-of-use and reduced clinician error. Expert Rev Med Devices 2023 Jul 04;20(7):607-614 [FREE Full text] [doi: 10.1080/17434440.2023.2219001] [Medline: 37272116]
8. Kehagias E, Galanakis N, Tsetis D. Central venous catheters: which, when and how. Br J Radiol 2023 Nov;96(1151):20220894 [FREE Full text] [doi: 10.1259/bjr.20220894] [Medline: 37191031]
9. Saugel B, Scheeren TW, Teboul JL. Ultrasound-guided central venous catheter placement: a structured review and recommendations for clinical practice. Crit Care 2017 Aug 28;21(1):225 [FREE Full text] [doi: 10.1186/s13054-017-1814-y] [Medline: 28844205]
10. Dzulkafli AH, Abdull Wahab SF, Othman R. Ergonomics risk assessment of musculoskeletal disorder during ultrasound-guided internal jugular venous cannulation. Malays J Med Sci 2024 Oct 08;31(5):196-204. [doi: 10.21315/mjms2024.31.5.13] [Medline: 39416728]
11. Miyazaki A, Fujii A, Kuwabara D, Minoguchi K, Kawakami H, Nakamura K, et al. Enhancing the quality and safety of central venous catheter insertion using projection mapping: a prospective observational simulation study with eye-tracking glasses. BMJ Open 2025 May 24;15(5):e095803 [FREE Full text] [doi: 10.1136/bmjopen-2024-095803] [Medline: 40413056]
12. Edwards Iii GF, Zagarese V, Tulk Jesso S, Jesso M, Harden SM, Parker SH. Designing healthcare for human use: human factors and practical considerations for the translational process. Front Health Serv 2022 Jan 6;2:981450 [FREE Full text] [doi: 10.3389/frhs.2022.981450] [Medline: 36925891]
13. Overall and unit costs of the five most common hospital-acquired infections (HAIs) in the US. One Health Trust. 2013 Nov 8. URL: https://onehealthtrust.org/publications/infographics/overall-and-unit-costs-five-most-common-hospital-acquired-infections-in-us/ [accessed 2026-01-24]
14. Pitiriga V, Bakalis J, Kampos E, Kanellopoulos P, Saroglou G, Tsakris A. Duration of central venous catheter placement and central line-associated bloodstream infections after the adoption of prevention bundles: a two-year retrospective study. Antimicrob Resist Infect Control 2022 Jul 15;11(1):96 [FREE Full text] [doi: 10.1186/s13756-022-01131-w] [Medline: 35841083]
15. Jamshidi R. Central venous catheters: indications, techniques, and complications. Semin Pediatr Surg 2019 Feb;28(1):26-32. [doi: 10.1053/j.sempedsurg.2019.01.005] [Medline: 30824130]

XSL•FO
RenderX

16. Estrada-Orozco K, Cantor-Cruz F, Larrotta-Castillo D, Díaz-Ríos S, Ruiz-Cardozo MA. Central venous catheter insertion and maintenance: evidence-based clinical recommendations. Rev Colomb Obstet Ginecol 2020 Jun 20;71(2):115-162 [FREE Full text] [doi: 10.18597/rcog.3413] [Medline: 32770871]

Abbreviations

CLABSI: central line–associated bloodstream infection
CVC: central venous catheter
ED: emergency department
ICU: intensive care unit
OR: operating room

Edited by A Kushniruk; submitted 22.Sep.2025; peer-reviewed by FNJ Frakking, I Culverhouse, R Marshall; comments to author 11.Nov.2025; revised version received 09.Jan.2026; accepted 19.Jan.2026; published 06.Feb.2026.

Please cite as:
Privitera MB, Khan S, Irfan B, Ali S, Arredondo C, Sanderson K, Bonomo J
Opportunities for Improved Device Design Based on Central Line Placement Practices: Contextual Inquiry Study
JMIR Hum Factors 2026;13:e84621
URL: https://humanfactors.jmir.org/2026/1/e84621
doi:10.2196/84621
PMID:41557444

©Mary Beth Privitera, Sameer Khan, Bilal Irfan, Shayan Ali, Cecelia Arredondo, Kyrsten Sanderson, Jordan Bonomo. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 06.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

An Early-Stage Digital Therapeutic Intervention to Enhance Affective Response During Physical Activity Among Adults With Overweight or Obesity: Benchmark-Driven Formative Testing Study

Rachel Crosley-Lyons[1], MS; Lori Hatzinger[1], MS; Micaela Hewus[1], MPH; Wei-Lin Wang[1], PhD; Delfien Van Dyck[2], PhD; Jimi Huh[1], PhD; Eric Hekler[3,4], PhD; Genevieve F Dunton[1,5], MPH, PhD

[1]Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States
[2]Department of Movement and Sports Sciences, Ghent University, Ghent, Belgium
[3]Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, San Diego, CA, United States
[4]Center for Wireless & Population Health Systems, University of California, San Diego, San Diego, CA, United States
[5]Department of Psychology, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Rachel Crosley-Lyons, MS
Department of Population and Public Health Sciences
Keck School of Medicine
University of Southern California
1845 N Soto Street
Los Angeles, CA, 90032
United States
Phone: 1 3175187670
Email: rclyons@usc.edu

Abstract

Background: Mobile device–enabled interventions known as digital therapeutics (DTx) are increasingly used to prevent chronic disease by targeting psychological and behavioral processes. Individuals' unique experiences while receiving DTx comprise real-world evidence (RWE) for evaluating DTx performance. An emerging strategy for early-stage DTx formative work uses small sample sizes, which facilitate efficient iteration and agile learning, while evaluating performance against descriptive benchmarks defined a priori, therefore minimizing the risk for confirmation bias. This study test benchmarks from the DTx RWE framework to formatively evaluate a novel DTx (the eMOTION study) to enhance affective response (ie, how people feel) during physical activity (PA).

Objective: This study aimed to determine whether the eMOTION DTx met a priori benchmarks for safety (<1% of participants report an adverse event), plausibility (≥51% of participants experience increased enjoyment in PA), usability (eg, ≥51% of participants report adequate usability), sustainability, feasibility (eg, <70% of participants report dissatisfaction), and equity (equity and accessibility are approximately equal across subgroups).

Methods: Participants (N=36; mean age 46, SD 14 years; 20/37, 54% female) underwent stratified random assignment to test one of four DTx versions for 14 days (n=9 each): (1) intensity PA goals, (2) affect PA goals with type and context recommendations, (3) affect PA goals with savoring exercises, and (4) affect PA goals with type, context, and savoring. Participants completed daily intervention sessions, asking them to focus on achieving a target heart rate (intensity) or feeling good (affect) during PA. Smartwatches were used to track PA and answer ecological momentary assessment (EMA) questions about how they felt during PA. Performance toward benchmarks was primarily assessed via official Institutional Review Board reporting channels (safety), interviews (plausibility, accessibility, and usability), and questionnaires (System Usability Scale [usability], Delighted-Terrible Scale [sustainability and feasibility], and equity).

Results: The eMOTION DTx versions exceeded all a priori safety, plausibility, accessibility, usability, sustainability, feasibility, and equity thresholds. For safety, no adverse events were reported. Regarding plausibility, more than half of the participants who received affect goals reported increased PA enjoyment at the end of the study. Moreover, 64%-72% (23-26 out of 36) of participants rated the DTx at or above the standard System Usability Scale cutoff point for acceptable usability. More than 60% (22/36) of

XSL•FO
RenderX

participants reported satisfaction with all DTx components, supporting DTx sustainability and feasibility. Finally, there was evidence for equity, with plausibility and accessibility comparable across sex, race, ethnicity, income, age, BMI, mobility, and physical constraint subgroups.

Conclusions: Since DTx RWE Framework benchmarks for safety, plausibility, accessibility, usability, sustainability, feasibility, and equity were largely met, the eMOTION Study DTx is ready for a full-scale efficacy trial to refine the DTx and optimize efficiency and feasibility. Our approach incorporated transparent decision-making to generate results that are more readily translatable, easily replicable, and reflect current best practices in the field of DTx.

Trial Registration: ClinicalTrials.gov NCT06125964; https://clinicaltrials.gov/study/NCT06125964

(JMIR Hum Factors 2026;13:e71472) doi:10.2196/71472

KEYWORDS

digital therapeutics; real world evidence; formative study; MOST framework; SMART goals; mHealth; physical activity; affective response

Introduction

Digital Therapeutics

Widespread smartphone ownership and expanded wearable device capabilities have steadily facilitated a paradigm shift in health care delivery. The medical system has traditionally used in-person examinations and pharmaceuticals to detect, treat, and manage chronic diseases. However, such diagnostic care is incredibly costly—with global spending projected to reach US $47 trillion by 2030—and inefficient, given that many noncommunicable diseases are potentially avoidable via behavior modification [1]. While clinicians counsel patients about the importance of a healthy lifestyle for mitigating chronic disease risk, compliance with preventive care recommendations remains low. Digital therapeutics (DTx)—mobile device software systems that deliver evidence-based behavioral intervention packages to treat or prevent disease in real-world environments—are poised to complement and improve upon these efforts. For example, DTx can extend preventive care beyond the clinic by facilitating behavior change in daily life and empowering patients to monitor their own progress. High-resolution digital treatment evaluation data collected by DTx serve as real-world evidence (RWE) that can be used for efficient performance assessments during research and development. Even as forecasts show a DTx market compound annual growth rate of 27.2% until 2030 [2], the transition from DTx research and development to regulatory testing and public release remains disjointed and inefficient.

Benchmark-Driven Formative Testing as a Rigorous Approach to Early-Stage DTx Development

DTx involves complex sociotechnical systems with multiple interdependent components—user interfaces, behavioral protocols, technical infrastructure, and human-technology interactions—each of which must function adequately for the intervention to succeed. Failure to establish a robust foundation across all these components before initiating large-scale trials substantially increases the risk of type III error, wherein a study is conducted rigorously but fails to test what it intended to test because the intervention itself was inadequately developed or the control condition was poorly specified.

Traditional approaches to early-stage DTx research face a methodological paradox: formative studies must remain small to enable rapid, resource-efficient learning and iteration; yet, small sample sizes preclude the use of inferential statistics with adequate statistical power. This creates a problematic situation where researchers either (1) inappropriately apply underpowered inferential tests that are likely to produce unreliable results, or (2) rely solely on qualitative interpretation of descriptive results, which introduces substantial risk of confirmation bias, the well-documented human tendency to interpret ambiguous data in ways that confirm preexisting beliefs and expectations.

To address this methodological challenge, leading experts in behavioral intervention science and DTx development have advocated for benchmark-driven formative testing as a rigorous alternative approach. This methodology, articulated in frameworks such as the Obesity-Related Behavioral Intervention Trials (ORBIT) model for behavioral intervention development and Phase II of the DTx RWE Framework, establishes a priori performance thresholds that can be evaluated using purely descriptive statistics while maintaining scientific rigor [3].

The Logic of Benchmark-Driven Formative Testing

Benchmark-driven formative testing addresses 4 fundamental tensions in early-stage intervention research.

First, the need for efficient learning requires small sample sizes that enable rapid iteration cycles and resource-efficient exploration of intervention components, implementation strategies, and potential adaptations for diverse populations. Large trials are essential for establishing generalizability, but they are inappropriate tools for the formative work of intervention refinement.

Second, the inappropriateness of underpowered inferential statistics in small samples is well established. When sample sizes are insufficient for adequate statistical power, *P* values become unreliable indicators of true effects, CIs become too wide to be informative, and the risk of both false positives and false negatives increases substantially. Running inferential statistics on samples of N=30-50 participants typically yields a power of 0.20-0.40 for detecting medium effect sizes—well below the conventional 0.80 standard. In such contexts, null hypothesis significance testing provides an illusion of rigor while actually undermining the validity of conclusions [4,5].

Third, the risk of confirmation bias affects all human judgment, including scientific interpretation. Without prespecified criteria

XSL•FO
RenderX

for success, researchers may unconsciously interpret ambiguous results favorably, seeing "promising trends" or "encouraging patterns" that align with their hypotheses. This risk is particularly acute in small samples where data can appear to support multiple narratives depending on how one frames the analysis.

Fourth, the distinction between statistical and clinical significance is critical but often conflated. A statistically significant finding may have minimal real-world impact, while a clinically meaningful effect may not reach statistical significance in an underpowered study. Effect sizes, minimal clinically important differences (MCIDs), and practical significance must be considered alongside *P* values; yet, traditional null hypothesis testing frameworks prioritize statistical significance over substantive meaningfulness [6].

How Benchmarks Address These Tensions

A priori establishment of benchmarks addresses all 4 tensions simultaneously.

Enabling Small, Efficient Samples

Because benchmarks can be evaluated descriptively (eg, "did 51% of participants report increased PA enjoyment?"). They do not require the large samples necessary for adequate statistical power. This enables formative studies to remain small enough for rapid iteration while still providing clear, interpretable evidence about intervention performance.

Avoiding Inappropriate Statistical Tests

Benchmark evaluation relies on descriptive statistics—percentages, frequencies, and comparisons to predetermined thresholds—rather than inferential tests. This eliminates the problems associated with underpowered hypothesis testing while still providing rigorous evaluation criteria relevant for formative work within small sample sizes.

Minimizing Confirmation Bias

By establishing success criteria before data collection, benchmark approaches create an objective standard against which results must be evaluated. If the benchmark specifies that ≥51% of participants must report increased PA enjoyment, then 45% represents clear failure to meet the criterion, regardless of how "close" it might feel or how researchers might be tempted to interpret it favorably. The threshold is predetermined and nonnegotiable.

Centering Clinical Meaningfulness

Benchmarks are explicitly designed to reflect what would constitute meaningful real-world impact, not merely statistical detectability. For example, in our study here, a 51% threshold for plausibility reflects the expectation that an intervention should benefit the majority of recipients—a clinically meaningful standard that is independent of sample size or statistical power considerations. Benchmarks can incorporate established MCIDs, align with clinically significant thresholds from previous literature, and balance what is both meaningful and plausibly achievable given the current state of the science.

Benchmark Selection and Justification

Rigorous benchmark-driven research requires that thresholds be established through a systematic process that balances clinical meaningfulness with plausible achievability based on previous evidence. Given the novelty of this line of work, our team developed benchmarks that have an initial plausibility of balancing both ambition and likelihood of being observed. While these benchmarks were not explicitly based on previous literature—due to the lack of extant work—they represent a potential baseline that others can improve upon in future proof-of-concept trials.

The process of benchmark development should be transparent and documented, ideally through preregistration, to ensure that thresholds are truly a priori rather than post hoc. In our study, benchmarks were established before data collection and documented in our preregistered analysis plan on Open Science Framework [7].

Interpreting Benchmark Results

Critically, results from benchmark-driven formative studies must not be interpreted as generalizable to broader populations. Meeting benchmarks indicates that an intervention has achieved a sufficient foundation to warrant progression to the next development phase, but it does not establish external validity or population-level efficacy or effectiveness. Those determinations require adequately powered randomized controlled trials with representative samples.

Instead, benchmark-driven formative testing serves a specific purpose in the intervention development pipeline: to establish whether all essential components of an intervention function adequately, whether the intervention can be delivered as intended, whether it demonstrates sufficient signal of benefit to justify further investment, and whether potential implementation barriers can be identified and addressed before committing to large-scale testing. This is analogous to Phase I/II trials in pharmaceutical development, which establish safety, dosing, and preliminary efficacy signals before proceeding to Phase III confirmatory trials.

Meeting formative benchmarks provides evidence that an intervention is "ready" for larger-scale testing—that the foundation is sufficiently solid that a properly powered trial can be conducted without high risk of Type III error. Failing to meet benchmarks indicates that further refinement is needed before proceeding, thereby protecting against the waste of resources on large trials of interventions that are not yet adequately developed.

Rigor in Formative Research

The methodological rigor of formative research is achieved not through the application of inferential statistics to underpowered samples, but through:

1. Systematic benchmark establishment grounded in previous literature and clinical meaningfulness.
2. Prespecification and transparent documentation of all benchmarks before data collection.
3. Comprehensive assessment across multiple domains (safety, plausibility, usability, feasibility, and equity).

XSL•FO
RenderX

4. Objective evaluation of whether benchmarks were met using descriptive statistics.
5. Honest interpretation that acknowledges the formative, nongeneralizable nature of findings.
6. Iterative refinement based on benchmark performance before proceeding to larger trials.

This approach represents a methodologically sound framework for early-stage intervention development that avoids the false precision of underpowered inferential testing while maintaining scientific rigor through prespecified, clinically meaningful evaluation criteria. As Czajkowski and colleagues [3] note in their comprehensive guide to behavioral intervention development, early-stage work should prioritize intervention optimization, rather than hypothesis testing—a fundamentally different scientific objective that requires different methodological tools.

By embracing benchmark-driven formative testing, the field can conduct more rigorous early-stage research that provides a robust foundation for subsequent confirmatory trials. Ultimately, to improve the quality and efficiency of the entire intervention research-to-development pipeline, such early-stage work should systematically and iteratively assess DTx using real-world RWE. The DTx RWE Framework recently proposed by Kim and colleagues [8] incorporates RWE-guided decision-making against benchmarks congruent with existing regulatory standards throughout the DTx lifecycle. The Framework's step-by-step process aims to streamline the research and development-regulatory testing-public release process, facilitate greater transparency, and produce DTx that reflect current best practices and are more readily translatable to real-life settings. This has important public health implications, as emerging DTx are poised to collectively reduce health care burden and expand access to high-quality care via tailoring algorithms that deliver personalized and precise multicomponent treatment [9]. This study applied principles from the DTx RWE Framework in conducting benchmark-driven formative testing of an early-stage DTx intervention to enhance affective response during physical activity (PA) among physically inactive adults with overweight or obesity.

This Study

Overview

Inadequate PA engagement (ie, <150 min/week of moderate intensity physical activity [MPA]) confers substantial risk for a variety of chronic diseases, including cancer [10]. Despite widespread public health initiatives—and many individuals reporting a desire to become more physically active—less than 5% of adults actually meet the ≥150 minute/week MPA threshold [11]. Interventions targeting social-cognitive factors (eg, self-efficacy, intentions, and beliefs) have had limited success, accounting for only 31% of the variability in PA engagement [12-15]. This PA intention-behavior gap suggests that noncognitive factors might be hijacking people's best intentions to engage in PA [16]. According to the theory of hedonic motivation, humans are naturally drawn to activities that are pleasurable and avoid unpleasant activities [17]. Previous affective responses to a given behavior inform the development of affectively charged motivations to pursue or avoid the behavior in the future [18]. People with overweight or obesity are more likely than individuals of a healthy weight to experience pronounced negative affect (eg, discomfort) while engaging in PA, therefore decreasing the likelihood they will consistently engage in PA [19].

Factors that are potentially readily modifiable, such as PA intensity, type, and context, have been shown to be significant predictors of the affective response during PA. For example, most people experience pleasure when engaging in low-intensity PA and displeasure while engaging in vigorous-intensity PA [20], but affective responses to MPA are heterogeneous [21]. At moderate intensities, affective responses might be determined by the type (eg, hiking and dancing) or context (eg, being with others and outdoors) of PA [22-25]. Ultimately, affective responses vary inter- and intrapersonally and impact future PA behavior—but whether they can be experimentally manipulated to promote PA engagement in the real world remains unknown [26,27].

Developing a DTx Solution

Therefore, an existing DTx called MyDayPlan was modified to experimentally manipulate affective mechanisms associated with MPA among physically inactive adults with overweight or obesity. MyDayPlan has previously increased PA engagement in similar populations by leveraging self-regulation strategies posed by the Health Action Process Approach [28]. Specifically, smartphone sessions provide daily PA goals and prompt participants to plan how, where, and when they will accomplish the goal (action planning), identify potential barriers and solutions (coping planning), and monitor their goal progress (self-monitoring) [29]. In its original form, MyDayPlan provided goals focused on sustaining a certain intensity of PA (ie, "intensity-based goals"). The eMOTION Study incorporated affect-based PA goals and enhancements to modify affective mechanisms related to PA into MyDayPlan's existing self-regulatory framework. New affect-based goals asked participants to engage in PA types and contexts that would allow them to experience positive affect (eg, enjoyment) during PA. We hypothesized that participants who received affect-based goals would report more positive affective experiences during PA than participants who received intensity-based goals, which asked participants to attain a target heart rate. Iterative internal testing among research team members was performed to achieve a minimum viable product with a functioning interface, error-free programming, and a reasonable time burden.

Benchmark-Driven Formative Testing

Next, following guidance from Phase II of the DTx RWE framework, SMART (specific, measurable, actionable, realistic, and timely [or time-bound]) benchmarks were defined (Textbox 1). Benchmarks required the DTx to reach safety, plausibility, accessibility, usability, sustainability, feasibility, and equity standards during a formative study before initiating a resource-intensive Comparison Study to test efficacy. Broadly, benchmarks ensured that the DTx was unlikely to cause harm (ie, safety); the affect-based goals were associated with greater PA enjoyment compared to the intensity-based control (ie, plausibility); and the DTx was easy to use (ie, accessibility and usability); and practical for study staff and participants to sustain

XSL•FO
RenderX

over time (ie, feasibility and sustainability). We also explored whether DTx plausibility and accessibility were approximately equal for all participants, regardless of individual-level traits (ie, equity; this benchmark was exploratory and not statistically powered). This paper summarizes results from benchmark-driven formative testing of the eMOTION DTx by measuring performance against these a priori established benchmarks.

Textbox 1. eMOTION Formative Study SMART (specific, measurable, actionable, realistic, and timely [or time-bound]) benchmarks.

Safety:

- Only <1% of Formative Study participants will experience an adverse event.

Plausibility:

- The majority of participants (ie, ≥51%) receiving affect-based goals will report a perceived increase in enjoyment of physical activity at the end of the study.

Accessibility and usability:

- The daily intervention sessions and Fitbit smartwatch features will receive an average score at or above the System Usability Scale cutoff point (≥68).
- The majority of participants (ie, ≥51%) will report that they are able to (1) read the wording of the survey questions, (2) understand the wording of the survey questions, (3) select their answers for the survey questions, and (4) use the exercise settings on the Fitbit Versa smartwatch.
- The majority of participants (ie, ≥51%) will report being able to (1) read and (2) understand the daily intervention sessions on their phone.
- Participants will be given physical activity recommendations that appropriately consider their reported constraints.
- The majority of participants (ie, ≥51%) who remember receiving savoring questions will report being able to understand and follow them.

Sustainability and feasibility:

- The auto-detection algorithm for physical activity will correctly trigger event-contingent ecological momentary assessment (EMA) prompts ≥51% of the time.
- Less than 25% of participants will need to be sent a new Fitbit due to technical issues.
- A total of <70% of participants will report feeling dissatisfied with the Fitbit Versa (1) exercise settings, (2) prompt notifications, and (3) prompt burden.
- In addition, <25% of participants will be sent repeated (ie, >1) reminders to open their Fitbit app and sync their study data.
- Moreover, <70% of participants who remember completing daily intervention sessions will report feeling dissatisfied with the daily intervention sessions.
- Additionally, <70% of participants will report feeling dissatisfied with the Fitbit app.
- In total, <70% of participants who remember receiving physical activity recommendations will report feeling dissatisfied with the physical activity recommendations.
- Furthermore, <70% of participants who remember receiving savoring questions will report feeling dissatisfied with the savoring questions.

Equity:

- Plausibility and accessibility of the eMOTION Intervention will be approximately equal between groups (ie, sex, race, ethnicity, age, BMI, income, mobility, and physical constraints).

Methods

Study Design

A 3-week formative study was used to test DTx RWE Framework benchmarks. Treatment components targeting intensity-based goals and affect-based goals were examined separately in different groups of participants. Stratified random assignment placed participants (N=36) into one of four groups (n=9 each): (1) intensity-based goals, (2) affect-based goals with the type and context enhancement, (3) affect-based goals with the savor enhancement, and (4) affect-based goals with type, context, and savor enhancements. Published recommendations for proof-of-concept and formative studies using similar measures to our own support the use of homogenous samples as small as 5 to 10 subjects per user group [30-32]. Our sample (n=9 per group) precludes comparisons via inferential statistics, but corresponds with current best practices by prioritizing efficient iterations central to early-stage work—and is ultimately still amenable to descriptive statistics to determine whether the DTx meets preliminary benchmarks. Strata within each condition were based on age (ie, <45 vs ≥45 years), sex at birth (ie, male vs female), and BMI classification (ie, those with overweight vs those with obesity) to allow for approximately equal numbers of subgroup participants within each treatment group. The study was triple-blinded so that participants, research staff interacting with participants, and statisticians conducting analyses were all unaware of group

XSL•FO
RenderX

assignment. Methods were preregistered in Open Science Framework [7].

Participants

Rolling recruitment occurred from September 2023 to March 2024 via ResearchMatch, a national health volunteer registry created by academic institutions and supported by the National Institutes of Health as part of the Clinical Translational Science Award program. An online screener assessed participant eligibility. The eligibility criteria are shown in Textbox 2.

Textbox 2. Eligibility criteria.

Inclusion criteria were as follows:

- Ages 18 years and older.
- Residing in the United States.
- BMI ≥ 25 kg/m^2.
- Current engagement in ≤60 minutes of structured physical activity (PA) per week.
- Smartphone ownership.
- Residing in an area with Wi-Fi connectivity.
- Able to speak and read English.
- Interested in starting a PA program.
- Willing to wear a smartwatch every day.
- Able to read the small font on a smartwatch screen.

Exclusion criteria screened out individuals who:

- Had cognitive disabilities precluding participation and/or the ability to provide informed consent.
- Had physical disabilities or medical issues limiting PA engagement.
- Were unable or unwilling to wear a smartwatch.
- Were currently pregnant.

Ethical Considerations

All procedures were conducted remotely, and approval was provided by the University of Southern California Institutional Review Board (ID UP-22-00332). Eligible individuals provided written informed consent to participate in study procedures. Data were not deidentified or anonymized, so a number of protections were implemented to maintain participant privacy and confidentiality. These included the use of pseudonyms in place of names to organize study records, as well as the storage of data on secure databases and encrypted university servers. Participants were compensated up to US $75 for completing all study procedures.

Procedures

Following informed consent, participants completed a baseline questionnaire via REDCap (Research Electronic Data Capture; Vanderbilt University) [33,34]. Onboarding included a 1- to 3-day trial that was “passed” by completing two brief text message surveys within a single day. Participants who “passed” the trial received a Fitbit Versa 3 smartwatch in the mail and attended a 45-minute orientation session to learn about participation activities, finish technological setup, and review a PA guidebook. For the first week of the study, assessment components (eg, Fitbit smartwatch monitoring and check-ins) were introduced gradually. Then, for the following 2 weeks, participants engaged in the full DTx. Participants were asked to wear the smartwatch on their nondominant wrist for ≥23 hours/day. Smartwatch-based ecological momentary assessment (EMA) prompts (ie, “check-ins”) were delivered during PA sessions to measure affective responses to PA. EMA is a methodology involving brief repeated surveys that capture subjective experiences as they unfold in real time [35]. Participants were also instructed to press an exercise button on the smartwatch before starting a planned PA session. Researchers performed remote monitoring of participant compliance and contacted participants if missing data were detected in the study server. Finally, an exit interview and poststudy questionnaire collected feedback relating to the formative study’s SMART benchmarks.

DTx Intervention

The eMOTION DTx was a multicomponent treatment package that was delivered under unsupervised, real-world settings and included the following elements.

PA Guidance and Scheduling

Participants were given a prescription to start at their own pace and comfort level, slowly increasing MPA volume over the course of the study. Because PA was performed asynchronously, a digital guidebook was provided as a reference summarizing strategies to gauge PA intensity and proper warm-up and cool-down procedures. Additionally, on Sundays, participants completed a smartphone-based scheduling survey by indicating the days they intended to engage in PA for that upcoming week.

XSL•FO
RenderX

Daily PA Goal Sessions

On days participants reported they intended to engage in PA, they completed 2 smartphone-based interactive intervention sessions: one in the morning and one in the evening. A text message push notification and up to 4 reminder notifications (once per hour) with a REDCap link alerted participants to complete each session. Morning sessions were scheduled for 6 AM local time. They provided a PA goal for the day ("goal module") that differed according to the participant's random group assignment. Participants then created a plan ("action plan module") to achieve their goal by specifying details, including the time they planned to engage in PA that day. The final module ("coping plan module") asked participants to anticipate barriers and brainstorm potential solutions to reach their goal. Text messages for evening sessions were sent at 7 PM local time and featured a "self-monitoring module" that asked participants to reflect on whether they met their PA goal.

In the intensity-based goal condition, participants were asked to maintain a target heart rate during PA. Target heart rate was based on the approximate age-adjusted maximum heart rate (HRmax; HRmax = 207 – [0.7 × age]), reflecting 55% HRmax for week one and 60% HRmax for week two. In the affect-based goal condition, participants were asked to engage in a type (25% of daily goals) or context (75% of daily goals; alternating places, social situations, and listening contexts) of PA that they found enjoyable. The instructions were generic to encourage self-selection (eg, "Today, your goal is to try a TYPE of PA that makes the experience more enjoyable while you are doing it").

Enhancements

Participants randomized to receive the type and context enhancement were provided tailored recommendations for PA types or contexts most likely to satisfy their unique psychological needs, as measured at baseline (ie, relatedness, social esteem, individual esteem, creativity, mindfulness, learning, challenge, entertainment, escapism, aesthetic appreciation, and morality). Recommendations were generated by an algorithm trained on data from a crowdsourced panel of adults on Amazon Mechanical Turk who were asked to rate a range of PA types and contexts on the extent to which they satisfied psychological needs. Further details on the algorithm are outside the scope of this paper. Recommendations suggested PA types or contexts matched to the participant's most personally important psychological needs categories and eliminated types and contexts from consideration that coincided with any constraints reported at baseline.

Participants who were randomized to receive the SAVOR enhancement engaged in a brief savoring exercise to enhance the positive emotions experienced during their PA session. Immediately following the morning intervention session, these participants received a text message instructing them to click a REDCap link to complete the savoring exercise after they performed their planned PA session. The savoring exercise was comprised of two open-ended questions about the PA (eg, "Reflect on the most rewarding aspect of your PA session today. What made it feel like a success for you? How did this influence your experience?") and surrounding environment (eg, "Let's take a moment to focus on the environment in which you were active today. Can you describe what you saw, felt, heard, and smelled in that space? What aspects of the environment are you grateful for?").

Measures

Overview of Measures

Study variables evaluating DTx performance for SMART benchmarks were collected via questionnaires, videoconference exit interviews, and EMA. During exit interviews, if the participant's initial response to an open-ended question was too vague, the surveyor followed up with neutrally worded clarifying questions. During response coding for analysis, if any response contained conflicting statements and true clarity could not be achieved, the response was always treated conservatively (ie, biased toward the null hypothesis). Measures are briefly summarized by benchmark type below and are elaborated in detail in Multimedia Appendix 1.

Safety

Safety was assessed by calculating the number of adverse events reported via official Institutional Review Board (IRB) channels, which allow participants and study teams to report events. Participants were given IRB contact information for this purpose during informed consent.

Plausibility

Plausibility was primarily assessed with the following exit interview question: "When comparing your experience with PA before and after participation in the eMOTION Study, do you personally feel as if the degree of enjoyment you feel while engaging in PA changed or stayed the same?" Responses were transcribed, and enjoyment was coded as "decreased" (0), "stayed the same" (1), or "increased" (2).

Accessibility and Usability

Daily Intervention Sessions and Fitbit Smartwatch

The poststudy questionnaire included the 10-item System Usability Scale (SUS) to assess the accessibility and usability of the daily intervention sessions and the Fitbit smartwatch [36]. Participants indicated the extent to which they agreed with each statement using a Likert Scale ("strongly disagree" {0} to "strongly agree" {4}). Negative items were reverse-scored, all items summed, and the sum multiplied by 2.5 to yield a total score ranging from 0 to 100. Items displayed acceptable reliability (Cronbach α=0.78 and 0.91 for daily goals and Fitbit smartwatch, respectively). Participants were also asked investigator-created questions in the exit interview about their ability to read and understand the wording of the daily intervention sessions and check-ins; select their answers for the check-ins; and use the exercise buttons on the smartwatch. Responses were transcribed and coded as "unable" (0) or "able" (1).

Tailored Physical Activity Recommendation Enhancement

Multiple measures assessed the accessibility and usability of type and context recommendations. During the exit interview, participants were asked whether they were "generally able to follow the PA [type and context] recommendations" they

XSL•FO
RenderX

received. Responses were transcribed and coded as "unable" (0) or "able" (1). Reported inability was then considered alongside constraints reported at baseline. Specifically, in the baseline questionnaire, a set of questions asked about constraints related to PA type, location, and listening contexts. Participants indicated their ability to engage in 47 different activity types [37] by selecting as many options as applied for each one: "able to do it," "physically unable to do it," "no place to do it," "no equipment to do it," and "I don't know what this is." For some activities, "no equipment to do it" was not considered a constraint because no equipment was required (eg, bodyweight exercises); for others, "I don't know what this is" did not apply because participants were provided with instructional videos for the activity. Next, participants indicated their ability to access 19 different locations and facilities for PA by selecting "I'm able" or "I don't have access or can't get there" for each. Finally, participants selected "yes" or "no" for whether they were able to listen to music, audiobooks, or other media on their smartphone during PA. The total number of possible constraints across PA types and contexts ranged from 0 to 68.

Savoring Enhancement

Participants were asked during the exit interview whether they were able to understand and follow the savoring exercises. Responses were transcribed and coded as "unable" (0) or "able" (1).

Sustainability and Feasibility

Satisfaction With DTx Components

Most benchmarks for sustainability and feasibility were assessed with the Delighted-Terrible Scale [38,39] in the poststudy questionnaire. Participants were asked to indicate how they felt about the smartwatch exercise settings; check-in notifications and burden; daily intervention sessions; smartphone data syncing procedures; and type, context, and/or savor enhancements, as applicable. Multiple choice options for each item were: "delighted" (8), "pleased" (7), "mostly satisfied" (6), "mixed- about equally satisfied and dissatisfied" (5), "mostly dissatisfied" (4), "unhappy" (3), "terrible" (2), "neutral- neither satisfied nor dissatisfied" (1), and "I never thought about it" (0). Response options "mostly dissatisfied," "unhappy," and "terrible" were considered as dissatisfaction with the given DTx component.

Fidelity of PA Auto-Detection Algorithm

An algorithm triggered check-ins when the smartwatch sensors detected PA engagement via sufficiently elevated 10-minute rolling average heart rate (ie, 55% to 60% of age-adjusted HRmax). Heart rate data from smartwatch sensors were considered alongside Fitabase software prompting records, which contained timestamps associated with the timing of smartwatch check-ins.

Research Staff Burden

Some measures assessed whether DTx-associated workloads were sustainable and feasible for research staff to maintain. A REDCap form was used to track smartwatch inventory and document technical difficulties requiring device replacement. Staff also recorded all instances where they had to remind participants to sync their smartphone data due to >48 hours elapsing without data reaching Fitabase servers.

Equity

To explore equity, the proportions of participants recalling increased PA enjoyment after the study, and rating the daily goals and smartwatch at or above the acceptable cutoff point (ie, SUS score ≥68.0), were compared across sex, race, ethnicity, age, BMI, income, mobility, and physical constraints subgroups (subjective comparisons, given a lack of statistical power). Subgroup data were collected with screening and baseline questionnaires. Sex (male or female), race (White or non-White), ethnicity (Hispanic or non-Hispanic), and income (unsure; ≤US $44,999; US $45,000-$84,999; US $85,000-$124,999; or ≥US $125,000) were directly assessed with single items. Age was calculated as the difference between date-of-birth and baseline questionnaire completion date, then split into quartiles (≤37 y, 38-43 y, 44-58 y, and ≥59 y). BMI was calculated using reported height and weight and coded according to established clinical cutoff points (those with overweight ≤29.9, those with obesity ≥30.0). Mobility was assessed with an investigator-developed measure, which asked participants to indicate the degree to which they agreed with 5 different statements (eg, "I have had trouble bending over") using a 10-point sliding scale with the markers "not at all true" (1), "moderately true" (5), and "completely true" (10). Items were summed into a total mobility score (range 5 [least trouble] to 50 [most trouble]), which was then split into quartiles (≤13.5, 13.6-18.5, 18.6-28.5, and ≥28.6). Finally, for physical constraints, activity types designated as "physically unable to do it" by the participant at baseline were tallied to create a total number ranging from 0 to 47, which was split into tertiles (0, 1-4; ≥5).

Statistical Analysis

The formative study was not statistically powered to test hypotheses using inferential statistics. Rather, as is common for proof-of-concept trials, the sample size was chosen to balance monetary and personnel constraints while allowing the DTx to be applied and assessed in the target population. To minimize confirmation bias and reduce the risk of moving forward with a large-scale phase III comparison study prematurely, descriptive statistics were used to compare formative study evaluation data against a priori benchmarks with clear go or no-go criteria [8]. Sensitivity analyses featuring linear and logistic regression models were performed to compare person-level characteristics between the final sample and participants lost to follow-up. All analyses were performed using IBM SPSS Statistics software (version 29).

Results

Sample Characteristics

A total of 46 participants consented and enrolled in the eMOTION Formative Study. Of these, 9 participants left the study during onboarding and before the intervention: 2 withdrew due to personal obligations (ie, surgery and work demands), 1 was unenrolled due to technical difficulties associated with responding to text message EMA on their personal smartphone,

XSL•FO
RenderX

and 6 were lost to follow-up without explanation (ie, study staff were unable to reestablish contact).

Table 1 provides baseline characteristics for the 37 participants who completed the intervention. This baseline sample was 46% female (n=17), 86% non-Hispanic (n=32), and 86% White (n=32). Mean age was 46.2 (SD 13.4) years. A broad range of household incomes was represented (Table 1). One additional participant in the sample was lost to follow-up after completing the intervention but before the poststudy questionnaire and exit interview. Therefore, a final analytic sample of 36 participants who completed the study and provided at least partial outcome data was retained. Sensitivity analyses confirmed that the final sample did not differ from participants lost to follow-up on the basis of their sex (P=.88), age (P=.82), or BMI (P=.98).

The baseline sample comprised 37 participants, while 36 participants provided at least some information for the main outcomes of interest by filling out the poststudy questionnaire and/or completing the exit interview.

Table 1. Participant characteristics assessed at baseline.

Characteristics	Values
Age (years), mean (SD)	46.2 (13.4)
Sex, n (%)	
Male	17 (46)
Female	20 (54)
Ethnicity, n (%)	
Non-Hispanic	32 (86)
Hispanic	5 (14)
Race, n (%)	
White	32 (86)
Black	4 (11)
Asian	1 (3)
Native American	0 (0)
Pacific Islander	0 (0)
Multiracial	0 (0)
Household income (US $), n (%)	
≤44,999	7 (19)
45,000 to 84,999	9 (24)
85,000 to 124,999	8 (22)
≥125,000	10 (27)
Unsure	3 (8)

SMART Benchmarks

Safety

No adverse events were reported during the study. With 0% of participants experiencing an adverse event, the safety benchmark (<1% of participants) was met.

Plausibility

Of the 27 participants who received affect-based goals, 14 (52%) reported during the exit interview that PA enjoyment increased over the course of the study (ie, comparing enjoyment before the study vs after the study), while 13 (48%) said enjoyment stayed the same. Therefore, the plausibility benchmark was just met (ie, ≥51% recalled an overall increase in enjoyment).

Accessibility and Usability

Daily Intervention Sessions and Fitbit Smartwatch

The daily intervention sessions and Fitbit smartwatch received average total SUS scores of 72.4 (SD 17.2) and 73.1 (SD 22.9), respectively. These scores exceeded the prespecified cutoff point in the SMART benchmarks (ie, ≥68). Post hoc calculations additionally determined that 23 (64%) and 26 (72%) participants rated the daily intervention sessions and Fitbit smartwatch components at-or-above the cutoff point (ie, ≥68). The majority of participants also reported being able to read the smartwatch check-ins (n=34, 97%) and daily intervention sessions (n=34, 97%), understand the check-ins (n=32, 91%) and daily sessions (n=34, 97%), select answers for check-ins (n=33, 94%), and use exercise settings on the smartwatch (n=27, 77%), therefore meeting the associated benchmarks.

XSL•FO
RenderX

Type and Context Enhancement

Although 18 participants received type and context recommendations, 3 did not remember receiving them. Table 2 depicts results from the 15 participants who received and remembered type and context recommendations. There was no discernable relationship between the number of activity constraints and reported ability to follow type and context recommendations, suggesting that the benchmark was met (ie, participants were given PA recommendations that appropriately considered their constraints).

Only data from participants who both received and remembered receiving PA type and context recommendations (ie, n=15) were included in the table. An additional 3 participants received, but did not remember receiving, recommendations; 5 participants incorrectly reported that they received type and context (ie, when they did not receive type and context); these data were excluded from the table.

Table 2. Accessibility of type and context recommendations.

PA[a] constraint quartile	Reported being able to follow recommendations, n/N (%)
1	2/5 (40)
2	1/2 (50)
3	2/4 (50)
4	1/4 (25)

[a]PA: physical activity.

SAVOR Enhancement

Although 18 participants received savoring exercises, 4 did not remember receiving them. Of the 14 participants who received and remembered SAVOR, 12 (86%) reported being able to understand and follow the savoring exercises. Therefore, the associated benchmark (≥51% of participants who remember receiving savoring exercises will report being able to understand and follow them) was met.

Sustainability and Feasibility

Satisfaction With DTx Components

The proportion of participants reporting dissatisfaction with a DTx component (ie, Fitbit smartwatch exercise settings, check-in notifications, and check-in burden; daily intervention sessions; Fitbit app; PA type and context recommendations; and savoring exercises) ranged from 0% (n=0) to 39% (n=14; Table 3). Therefore, the associated benchmarks were met.

Table 3. Proportion of participants reporting dissatisfaction with digital therapeutics components.

Benchmark	Result, n (%)[a]
<70% of participants will report feeling dissatisfied with Fitbit features	
Exercise settings	6 (17)
Check-in notifications	12 (33)
Check-in burden	14 (39)
<70% of participants will report feeling dissatisfied with the Fitbit app	3 (8)
<70% of participants who remember an intervention component will report feeling dissatisfied with it	
Daily intervention sessions[b]	3 (9)
PA recommendations[c]	3 (23)
Savoring exercises[d]	0 (0)

[a]N=36 unless otherwise indicated.

[b]n=34 (2 participants did not remember receiving daily intervention sessions).

[c]n=13 (15 remembered receiving type and context recommendations, but 2 did not respond to the satisfaction item).

[d]n=12 (14 remembered receiving SAVOR, but 2 did not respond to the satisfaction item).

Fidelity of PA Auto-Detection Algorithm

All smartwatch check-ins were correctly sent when the rolling 10-minute average heart rate crossed the threshold specified in the algorithm (ie, 55%-60% HRmax). Therefore, the triggering algorithm achieved 100% accuracy and met the threshold (ie, the algorithm triggered smartwatch check-ins correctly ≥51% of the time).

Research Staff Burden

No replacement Fitbit smartwatches were sent out to participants due to technical issues; therefore, the meeting did not meet the associated benchmark (ie, <25% of participants needed a new Fitbit due to technical issues). Additionally, only 6/36 (ie, 17%) participants were sent a single reminder to sync their Fitbit data to their smartphone, and no participants (ie, 0%) required more than one reminder. Therefore, the benchmark (<25% of

XSL•FO
RenderX

participants will be sent repeated [>1] reminders to open their Fitbit app and sync their study data) was met.

Equity

Results for the equity benchmark are summarized in Table 4. Values are the proportion of participants in each subgroup who reported increased PA enjoyment and who scored the daily goals and Fitbit smartwatch at or above the accessibility and usability cutoff point. Cell sizes were too small to use chi-square tests between groups. Subjective comparisons generally suggested equal proportions across group levels, with a few notable exceptions. First, plausibility differed by sex, with 64% (n=9) of women reporting increased enjoyment during the study compared to only 38% (n=5) of men. Additionally, non-White participants reported greater accessibility, usability (4/5 to 5/5, 80%-100% vs 19/31 to 21/31, 61%-68%), and plausibility (3/4, 75% vs 11/23, 48%) compared to White participants. Accessibility and usability were comparable across all income categories except for the highest income group, 90% (9/10) of whom rated the daily goals and Fitbit smartwatch at-or-above the cutoff point (compared to 3/6 and 4/8, 50% to 6/9, 69% in the other groups). Finally, accessibility, usability, and plausibility were lowest among individuals reporting the most physical constraints compared to those reporting fewer than 5 constraints or no constraints, as well as those reporting greater mobility impairment.

In Table 4, proportions and percentages for accessibility and usability reflect participants who scored the daily intervention sessions and Fitbit smartwatch, respectively, at-or-above the System Usability Scale cutoff of 68. Proportions and percentages in the plausibility column are the participants in the affect-based groups who reported increased PA enjoyment over the course of the study. A total of 36 completed the accessibility and usability items in the poststudy questionnaire, and 35 completed the plausibility item in the exit interview. Cell sizes were too small to conduct chi-Square comparisons, so potential group differences were subjectively assessed.

XSL•FO
RenderX

Table 4. Equity benchmark performance.

Characteristics	Accessibility and usability		Plausibility, n/N (%)
	Daily goals, n/N (%)	Fitbit, n/N (%)	
Sex			
Male	11/17 (65)	13/17 (76)	5/13 (38)
Female	12/19 (63)	13/19 (68)	9/14 (64)
Ethnicity			
Non-Hispanic	21/32 (66)	24/32 (75)	13/25 (52)
Hispanic	2/4 (50)	2/4 (50)	1/2 (50)
Race			
White	19/31 (61)	21/31 (68)	11/23 (48)
Non-White	4/5 (80)	5/5 (100)	3/4 (75)
Household income (US $)			
≤44,999	3/6 (50)	3/6 (50)	1/5 (20)
45,000 to 84,999	6/9 (67)	6/9 (67)	7/7 (100)
85,000 to 124,999	4/8 (50)	5/8 (63)	3/7 (43)
≥125,000	9/10 (90)	9/10 (90)	3/5 (60)
Unsure	1/3 (33)	3/3 (100)	0/3 (0)
Age (years)			
≤37	3/9 (33)	6/9 (67)	4/7 (57)
38-43	6/9 (67)	6/9 (67)	2/7 (29)
44-58	8/9 (89)	8/9 (89)	4/5 (80)
≥59	6/9 (67)	6/9 (67)	4/8 (50)
BMI (kg/m^2)			
Overweight	8/12 (67)	10/12 (83)	5/10 (50)
Obese	15/24 (63)	16/24 (67)	9/17 (53)
Mobility score			
≤13.5	6/9 (67)	7/9 (78)	5/7 (71)
13.6-18.5	6/9 (67)	8/9 (89)	3/6 (50)
18.6-28.5	6/9 (67)	7/9 (78)	4/8 (50)
≥28.6	5/9 (56)	4/9 (44)	2/6 (33)
Physical constraints			
No constraints	10/15 (67)	12/15 (80)	5/11 (45)
1-5 constraints	9/12 (75)	9/12 (75)	6/8 (75)
≥5 constraints	4/9 (44)	5/9 (56)	3/8 (38)

Discussion

Principal Findings

The novel eMOTION DTx was systematically designed, developed, and formatively tested according to benchmarks recommended by the DTx RWE Framework. Multiple evidence-based intervention components were integrated to address the problem of inactivity due to low PA enjoyment among adults with overweight or obesity. Daily digitally interactive intervention sessions featured Health Action Process Approach self-regulatory strategies (ie, action planning, coping planning, and self-monitoring) to encourage attainment of PA goals [28]. Affect and Health Behavior Framework mechanisms, including affective responses during PA, were targeted with affect-based PA goals, as well as through two intervention enhancements [40]. The type and context enhancement involved a tailoring algorithm that generated recommendations for specific types and contexts of PA likely to satisfy personally relevant psychological needs. This was informed by a body of evidence demonstrating that the degree of importance ascribed to higher-order psychological needs varies between-persons

XSL•FO
RenderX

[41-44], as well as by our preliminary work suggesting that various PA types and contexts differentially satisfy psychological needs and enhance positive affect during PA [45,46]. The SAVOR enhancement leveraged savoring exercises to intensify the influence of affective responses on PA [47,48]. Savoring is a mindfulness practice that enhances the saliency of positive experiences via attentional deployment and gratitude-making [49,50]. The DTx underwent rigorous internal testing and iterative refinements during phase I to arrive at a minimum viable product for application in phase II.

Principal Results

SMART benchmarks determining success were defined a priori to maximize transparency and preregistered in accordance with open science best practices. Final benchmark results fully supported the safety and plausibility of the eMOTION DTx. Standards for safety and plausibility are considered the most important to uphold, with the Food and Drug Administration's various approval routes all requiring new medical devices to demonstrate sufficient safety and efficacy before public release.

Findings also confirmed eMOTION DTx accessibility and usability. Accessibility and usability were assessed for multidevice (ie, smartphone and Fitbit smartwatch) user interface, study-related functions, and specific intervention content based on previous work, underscoring the importance of these areas for participant engagement and intervention delivery [51-53]. While objective sustainability and feasibility benchmarks were all met, minor real-time iterative adjustments were made to decrease smartwatch check-in burden in response to participant feedback. Early into formative testing, many participants reported receiving a high volume of check-ins throughout the day that were only meant to be received during PA sessions. To increase the specificity of the algorithm (ie, improve differentiation between low-intensity lifestyle activities and structured MPA sessions), the original heart rate threshold (ie, 10-minute rolling average ≥50% HRmax) was increased to 55% HRmax.

In an effort to increase the validity of accessibility, usability, sustainability, and feasibility feedback, participants were asked whether they remembered receiving each of the intervention components (ie, daily intervention sessions, PA recommendations, and savoring exercises). Terminology for these questions was consistent with that used during onboarding and throughout all study materials. Only those who received and remembered receiving a given component were included in the respective analysis. Interestingly, we found that some participants had forgotten the intervention components they received during the study (ie, 2 did not remember receiving daily goal intervention sessions; 3, the tailored PA recommendations; and 4, the savoring exercises). This same general phenomenon has been described by other researchers engaging in complex, multicomponent interventions, in which participants fail to retrospectively identify—or otherwise misremember or misunderstand—critical aspects of the study [54,55]. By comparing objective study records with participant self-reports, we were able to control for this form of information bias [56].

Finally, equity in accessibility, usability, and plausibility was generally observed (via subjective comparisons) across sex, race, ethnicity, income, age, BMI, mobility, and physical constraints subgroups, although notable discrepancies did arise. First, a greater proportion of women reported increased PA enjoyment during the study compared to men. Previous work presents conflicting evidence for whether men or women are more likely to enjoy PA [57,58]. However, our findings might reflect those of others that consistently suggest the presence of sex-based differences in the relevance and plausibility of mHealth interventions for PA. Specifically, women are more likely to use PA-tracking mHealth apps [59], report using PA apps for the purpose of enjoyment [60], and experience more positive health outcomes following PA mHealth interventions compared to men [61]. DTx accessibility, usability, and plausibility also differed by race, but—contrary to what might be expected—favored non-White participants. This result is likely unreliable due to imbalanced cell sizes (ie, 31 White and 5 non-White) and was not necessarily a cause for concern, as the DTx was more accessible, usable, and efficacious for the traditionally disadvantaged group. Racial differences will be further explored using a larger sample in Phase III. More concerning trends were observed for household income, mobility, and physical constraint subgroups. Specifically, the DTx was most accessible and usable for the wealthiest and least efficacious for the lowest-income participants. The DTx was also the least accessible, usable, and least efficacious for participants with the most physical constraints and most limited mobility, respectively. To address these findings, a free YouTube channel was created for participants. The channel included exercise videos using positive, upbeat formats and featuring diverse instructors, contexts, and PA types. Videos required only minimal space with no equipment necessary, and provided optional physical adaptations for each exercise. These changes sought to provide a larger number of enjoyable PA options that were convenient, accessible, and affordable to all participants, regardless of their income or physical abilities.

Limitations

The extant literature contained very few validated measures for smartphone DTx accessibility, usability, sustainability, and feasibility. Therefore, we adapted existing measures from related technologies (eg, online websites and software systems) and created our own when appropriate. Investigator-developed items were highly specific to our particular DTx, potentially limiting generalizability to other DTx setups. Additionally, this phase II study's small sample size precluded the use of inferential statistics; findings were descriptive, therefore, and should be interpreted with caution. Similarly, results for safety, plausibility, accessibility, usability, sustainability, feasibility, and equity of the eMOTION DTx reflect preliminary findings in our small, homogenous sample. Specific subgroups were at times collapsed into more general "affect-based" and "intensity-based" groups, and we were unable to explore whether specific combinations of enhancements drove the main effects. However, the phase III study will feature a sufficiently powered, diverse sample (N=280) to rigorously analyze within- and between-group differences in DTx efficacy and optimize treatment components.

XSL•FO
RenderX

Conclusion

Overall, this study is one of the first to apply benchmarks recommended by the DTx RWE Framework to formatively test a novel digital intervention. By transparently reporting the specific measurement and analytic choices we made to operationalize the DTx RWE Framework, our study hopes to contribute methodology and results that can be replicated. This is an important undertaking, since recent investigations suggest that the inconsistent application of frameworks is hindering the advancement of knowledge in real-world implementation contexts [62]. Future work should aim to apply the DTx RWE Framework in other contexts to determine whether its processes are relevant across diverse health problems, target populations, and DTx formats.

All benchmarks were determined to be met satisfactorily following iterative improvements made to the eMOTION DTx during development. Preliminary evidence, therefore, supports a research transition into DTx RWE Framework phase III, in which a large-scale effectiveness trial will determine whether the eMOTION DTx produces a greater effect (ie, PA enjoyment) among the experimental (compared to control) group. Ultimately, our application of the DTx RWE Framework ensures that the design and development of the eMOTION DTx follow current best practices, increase the likelihood that regulatory standards are met, and facilitate an efficient translation from research to real-world practice.

Acknowledgments

This study was funded by a grant from the National Cancer Institute (R01CA272933). RC-L's effort on this paper was supported by a grant from the National Heart, Lung, and Blood Institute (F31HL176165). The authors attest that no generative artificial intelligence technology was used to generate text, figures, or other informational content in this manuscript.

Funding

This study was funded by a grant from the National Cancer Institute (R01CA272933). Rachel Crosley-Lyons' effort on the current paper was supported by a grant from the National Heart Lung and Blood Institute (F31HL176165).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: RC-L, EH, GFD Methodology: RC-L, EH, GFD Formal analysis: RC-L, WLW Investigation: RC-L, LH, MH, GFD Data curation: RC-L, LH, WLW Software: LH, WLW Resources: GFD Writing – original draft: RC-L, MH, EH Writing – review & editing: RC-L, LH, MH, WLW, DVD, JH, EH, GFD Visualization: RC-L Project administration: MH Supervision: DVD, JH, EH, GFD Funding acquisition: GFD

Conflicts of Interest

None declared.

Multimedia Appendix 1
Benchmark measurement and assessment details.
[DOCX File , 50 KB - humanfactors_v13i1e71472_app1.docx]

References

1. Hacker K. The burden of chronic disease. Mayo Clin Proc Innov Qual Outcomes 2024;8(1):112-119 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2023.08.005] [Medline: 38304166]
2. Grand View. Digital Therapeutics Market Size And Share Report, 2030. Grand View Research; 2024. URL: https://tinyurl.com/2ttfdrte [accessed 2024-09-09]
3. Czajkowski SM, Powell LH, Adler N, Naar-King S, Reynolds KD, Hunter CM, et al. From ideas to efficacy: The ORBIT model for developing behavioral treatments for chronic diseases. Health Psychol 2015;34(10):971-982 [FREE Full text] [doi: 10.1037/hea0000161] [Medline: 25642841]
4. Ioannidis JPA. Why most published research findings are false. PLoS Med 2005;2(8):e124 [FREE Full text] [doi: 10.1371/journal.pmed.0020124] [Medline: 16060722]
5. Ioannidis JPA. Why most discovered true associations are inflated. Epidemiology 2008;19(5):640-648. [doi: 10.1097/EDE.0b013e31818131e7] [Medline: 18633328]
6. Man-Son-Hing M, Laupacis A, O'Rourke K, Molnar FJ, Mahon J, Chan KBY, et al. Determination of the clinical importance of study results. J Gen Intern Med 2002;17(6):469-476 [FREE Full text] [doi: 10.1046/j.1525-1497.2002.11111.x] [Medline: 12133163]

XSL•FO
RenderX

7. Crosley-Lyons R, Dunton G, Hewus M, Hatzinger L, Wang W. Using Real-Time Data Capture to Examine Affective Mechanisms as Mediators of Physical Activity Adherence in Interventions: The eMOTION Formative Study. Published online April 17. 2024. URL: https://osf.io/QTF79 [accessed 2025-12-01]
8. Kim M, Patrick K, Nebeker C, Godino J, Stein S, Klasnja P, et al. The digital therapeutics real-world evidence framework: an approach for guiding evidence-based digital therapeutics design, development, testing, and monitoring. J Med Internet Res 2024;26:e49208 [FREE Full text] [doi: 10.2196/49208] [Medline: 38441954]
9. Dang A, Arora D, Rane P. Role of digital therapeutics and the changing future of healthcare. J Family Med Prim Care 2020;9(5):2207-2213 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_105_20] [Medline: 32754475]
10. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020;54(24):1451-1462 [FREE Full text] [doi: 10.1136/bjsports-2020-102955] [Medline: 33239350]
11. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
12. McEachan RRC, Conner M, Taylor NJ, Lawton RJ. Prospective prediction of health-related behaviours with the Theory of Planned Behaviour: a meta-analysis. Health Psychology Review 2011;5(2):97-144. [doi: 10.1080/17437199.2010.521684]
13. Turner RR, Steed L, Quirk H, Greasley RU, Saxton JM, Taylor SJ, et al. Interventions for promoting habitual exercise in people living with and beyond cancer. Cochrane Database Syst Rev 2018;9(9):CD010192 [FREE Full text] [doi: 10.1002/14651858.CD010192.pub3] [Medline: 30229557]
14. Young MD, Plotnikoff RC, Collins CE, Callister R, Morgan PJ. Social cognitive theory and physical activity: a systematic review and meta-analysis. Obes Rev 2014;15(12):983-995. [doi: 10.1111/obr.12225] [Medline: 25428600]
15. Rebar AL, Rhodes RE, Gardner B. How we are misinterpreting physical activity intention - behavior relations and what to do about it. Int J Behav Nutr Phys Act 2019;16(1):71 [FREE Full text] [doi: 10.1186/s12966-019-0829-y] [Medline: 31438956]
16. Rhodes R, Cox A, Sayar R. What predicts the physical activity intention-behavior gap? a systematic review. Ann Behav Med 2022;56(1):1-20. [doi: 10.1093/abm/kaab044] [Medline: 34231844]
17. Murphy S, Eaves D. Exercising for the pleasure and for the pain of it: the implications of different forms of hedonistic thinking in theories of physical activity behavior. Front Psychol 2016;7:843 [FREE Full text] [doi: 10.3389/fpsyg.2016.00843] [Medline: 27375524]
18. Stults-Kolehmainen MA, Blacutt M, Bartholomew JB, Gilson TA, Ash GI, McKee PC, et al. Motivation states for physical activity and sedentary behavior: desire, urge, wanting, and craving. Front Psychol 2020;11:568390 [FREE Full text] [doi: 10.3389/fpsyg.2020.568390] [Medline: 33240154]
19. Ekkekakis P, Lind E. Exercise does not feel the same when you are overweight: the impact of self-selected and imposed intensity on affect and exertion. Int J Obes (Lond) 2006;30(4):652-660. [doi: 10.1038/sj.ijo.0803052] [Medline: 16130028]
20. Liao Y, Chou CP, Huh J, Leventhal A, Dunton G. Examining acute bi-directional relationships between affect, physical feeling states, and physical activity in free-living situations using electronic ecological momentary assessment. J Behav Med 2017;40(3):445-457 [FREE Full text] [doi: 10.1007/s10865-016-9808-9] [Medline: 27766481]
21. Ekkekakis P, Parfitt G, Petruzzello SJ. The pleasure and displeasure people feel when they exercise at different intensities: decennial update and progress towards a tripartite rationale for exercise intensity prescription. Sports Med 2011;41(8):641-671. [doi: 10.2165/11590680-000000000-00000] [Medline: 21780850]
22. Bartholomew JB, Miller BM. Affective responses to an aerobic dance class: the impact of perceived performance. Res Q Exerc Sport 2002;73(3):301-309. [doi: 10.1080/02701367.2002.10609024] [Medline: 12230337]
23. Carraro A, Paoli A, Gobbi E. Affective response to acute resistance exercise: a comparison among machines and free weights. Sport Sci Health 2018;14(2):283-288. [doi: 10.1007/s11332-018-0427-4]
24. Dunton GF, Liao Y, Intille S, Huh J, Leventhal A. Momentary assessment of contextual influences on affective response during physical activity. Health Psychol 2015;34(12):1145-1153 [FREE Full text] [doi: 10.1037/hea0000223] [Medline: 26053885]
25. Lacharité-Lemieux M, Brunelle JP, Dionne IJ. Adherence to exercise and affective responses: comparison between outdoor and indoor training. Menopause 2015;22(7):731-740. [doi: 10.1097/GME.0000000000000366] [Medline: 25423324]
26. Crosley-Lyons R, Do B, Hewus M, Dunton GF. An ecological momentary assessment study of affectively-charged motivational states and physical activity. Psychol Sport Exerc 2023;67:102423. [doi: 10.1016/j.psychsport.2023.102423] [Medline: 37665876]
27. Dunton GF, Crosley-Lyons R, Rhodes RE. Affective response during real-world physical activity as an intervention mediator. Exerc Sport Sci Rev 2023;51(4):140-149. [doi: 10.1249/JES.0000000000000321] [Medline: 37195062]
28. Degroote L, Paepe AD, Bourdeaudhuij ID, Dyck DV, Crombez G. Effectiveness of the mHealth intervention 'MyDayPlan' to increase physical activity: an aggregated single case approach. Int J Behav Nutr Phys Act 2021;18(1):92 [FREE Full text] [doi: 10.1186/s12966-021-01163-2] [Medline: 34233718]
29. Schwarzer R, Luszczynska A. How to overcome health-compromising behaviors. European Psychologist 2008;13(2):141-151. [doi: 10.1027/1016-9040.13.2.141]

XSL•FO
RenderX

30. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: A systematic review of empirical tests. Soc Sci Med 2022;292:114523 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]
31. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003;35(3):379-383. [doi: 10.3758/bf03195514] [Medline: 14587545]
32. McCombie C, Esponda GM, Schmidt U, Lawrence V. Oualitative diary methods in mental health research: A scoping review and recommendations for research and reporting. European Psychologist 2024;29(1):3-16. [doi: 10.1177/02537176251363817] [Medline: 40799658]
33. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
34. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
35. Stone A, Shiffman S. Ecological momentary assessment (EMA) in behavioral medicine. Ann Behav Med 1994;16(3):199-202. [doi: 10.1093/abm/16.3.199]
36. Brooke J. SUS: A Quick and Dirty Usability Scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation in Industry. 1st ed. Boca Raton, Florida, USA: CRC Press; 1996:189-194.
37. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR, Tudor-Locke C, et al. 2011 compendium of physical activities: a second update of codes and MET values. Med Sci Sports Exerc 2011;43(8):1575-1581. [doi: 10.1249/MSS.0b013e31821ece12] [Medline: 21681120]
38. Westbrook RA. A rating scale for measuring product/service satisfaction. Journal of Marketing 1980;44(4):68-72. [doi: 10.1177/002224298004400410]
39. Andrews FM, Withey SB. Developing measures of perceived life quality: Results from several national surveys. Social Indicators Research 1974;1(1):1-26. [doi: 10.1007/bf00286419]
40. Brand R, Ekkekakis P. Affective–Reflective theory of physical inactivity and exercise. Ger J Exerc Sport Res 2017;48(1):48-58. [doi: 10.1007/s12662-017-0477-9]
41. Deci EL, Ryan RM. The "What" and "Why" of goal pursuits: human needs and the self-determination of behavior. Psychological Inquiry 2000;11(4):227-268. [doi: 10.1207/s15327965pli1104_01]
42. Brock TC, Livingston SD. The Need for Entertainment Scale. In: The Psychology of Entertainment Media: Blurring the Lines between Entertainment and Persuasion. New Jersey: Lawrence Erlbaum Associates Publishers; 2004:255-274.
43. Deci EL, Ryan RM. Self-determination theory: A macrotheory of human motivation, development, and health. Canadian Psychology / Psychologie canadienne 2008;49(3):182-185. [doi: 10.1037/a0012801]
44. Gao W, Liu Z, Li J. How does social presence influence SNS addiction? A belongingness theory perspective. Computers in Human Behavior 2017;77:347-355. [doi: 10.1016/j.chb.2017.09.002]
45. Hatzinger LA, Do B, Hewus M, Crosley-Lyons R, Dunton GF. When physical activity fails to satisfy important psychological needs: Associations with engagement and mental health. J Sports Sci 2025;43(24):3137-3145. [doi: 10.1080/02640414.2025.2574260] [Medline: 41090982]
46. Dunton GF, Do B, Crosley-Lyons R, Naya CH, Hewus M, Kanning M. Assessing basic and higher-level psychological needs satisfied through physical activity. Front Psychol 2023;14:1023556 [FREE Full text] [doi: 10.3389/fpsyg.2023.1023556] [Medline: 36891201]
47. Bryant FB, Veroff J. Savoring: A New Model of Positive Experience. New York: Psychology Press; 2017.
48. Jose PE, Lim BT, Bryant FB. Does savoring increase happiness? A daily diary study. The Journal of Positive Psychology 2012;7(3):176-187. [doi: 10.1080/17439760.2012.671345]
49. Sato I, Jose PE, Conner TS. Savoring mediates the effect of nature on positive affect. International Journal of Wellbeing 2018;8(1):18-33. [doi: 10.5502/ijw.v8i1.621]
50. Bryant FB, Chadwick ED, Kluwe K. Understanding the Processes that Regulate Positive Emotional Experience: Unsolved Problems and Future Directions for Theory and Research on Savoring. International Journal of Wellbeing 2011;1(1) [FREE Full text] [doi: 10.5502/ijw.v1i1.18]
51. Radcliffe E, Lippincott B, Anderson R, Jones M. A pilot evaluation of mHealth app accessibility for three top-rated weight management apps by people with disabilities. Int J Environ Res Public Health 2021;18(7):3669 [FREE Full text] [doi: 10.3390/ijerph18073669] [Medline: 33915917]
52. Slater H, Campbell JM, Stinson JN, Burley MM, Briggs AM. End user and implementer experiences of mHealth technologies for noncommunicable chronic disease management in young adults: systematic review. J Med Internet Res 2017;19(12):e406 [FREE Full text] [doi: 10.2196/jmir.8888] [Medline: 29233804]
53. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth app usability questionnaire (MAUQ): development and validation study. JMIR mHealth and uHealth 2019;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]

XSL•FO
RenderX

54. House A, Bryant L, Russell A. Involvement in and experiences of the intervention and of research participation: results from a mixed-methods study. In: Managing with Learning Disability and Diabetes: OK-Diabetes? a Case-Finding Study and Feasibility Randomised Controlled Trial. Southampton, UK: NIHR Journals Library; 2018.
55. Oxman M, Chesire F, Mugisha M, Ssenyonga R, Nsangi A, Oxman AD, et al. Participants' experiences of potential adverse effects of an intervention to improve critical thinking about health choices: a qualitative cross-trial process evaluation in Kenya, Rwanda and Uganda. BMJ Open 2025;15(10):e104236 [FREE Full text] [doi: 10.1136/bmjopen-2025-104236] [Medline: 41125290]
56. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc 2016 May;9:211-217 [FREE Full text] [doi: 10.2147/JMDH.S104807] [Medline: 27217764]
57. Craft B, Carroll H, Lustyk M. Gender differences in exercise habits and quality of life reports: assessing the moderating effects of reasons for exercise. Int J Lib Arts Soc Sci 2014;2(5):65-76 [FREE Full text] [Medline: 27668243]
58. Larsen S, Mozdoorzoy T, Kristiansen E, Nygaard Falch H, Aune TK, van den Tillaar R. A comparison of motives by gender and age categories for training at norwegian fitness centres. Sports 2021;9(8):113. [doi: 10.3390/sports9080113]
59. Gulec H, Smahel D. Individual and parental factors of adolescents' mHealth app use: nationally representative cross-sectional study. JMIR mHealth and uHealth 2022;10(12):e40340 [FREE Full text] [doi: 10.2196/40340] [Medline: 36525286]
60. Klenk S, Reifegerste D, Renatus R. Gender differences in gratifications from fitness app use and implications for health interventions. Mobile Media & Communication 2017;5(2):178-193. [doi: 10.1177/2050157917691557]
61. Lugones-Sanchez C, Recio-Rodriguez JI, Agudo-Conde C, Repiso-Gento I, G Adalia E, Ramirez-Manent JI, EVIDENT 3 Investigators. Long-term effectiveness of a smartphone app combined with a smart band on weight loss, physical activity, and caloric intake in a population with overweight and obesity (Evident 3 Study): randomized controlled trial. Journal of Medical Internet Research 2022;24(2):e30416 [FREE Full text] [doi: 10.2196/30416] [Medline: 35103609]
62. Damschroder LJ. Clarity out of chaos: use of theory in implementation research. Psychiatry Research 2020;283:112461 [FREE Full text] [doi: 10.1016/j.psychres.2019.06.036] [Medline: 31257020]

Abbreviations

DTx: digital therapeutics
EMA: ecological momentary assessment
HRmax: maximum heart rate
IRB: Institutional Review Board
MCID: minimal clinically important difference
MPA: moderate intensity physical activity
ORBIT: Obesity-Related Behavioral Intervention Trials
PA: physical activity
REDCap: Research Electronic Data Capture
RWE: real-world evidence
SMART: specific, measurable, actionable, realistic, timely (or time-bound)
SUS: System Usability Scale

Edited by A Kushniruk; submitted 19.Jan.2025; peer-reviewed by J Courtney, JH Song; comments to author 11.Sep.2025; revised version received 01.Dec.2025; accepted 02.Jan.2026; published 20.Feb.2026.

Please cite as:
Crosley-Lyons R, Hatzinger L, Hewus M, Wang WL, Van Dyck D, Huh J, Hekler E, Dunton GF
An Early-Stage Digital Therapeutic Intervention to Enhance Affective Response During Physical Activity Among Adults With Overweight or Obesity: Benchmark-Driven Formative Testing Study
JMIR Hum Factors 2026;13:e71472
URL: https://humanfactors.jmir.org/2026/1/e71472
doi:10.2196/71472
PMID:

©Rachel Crosley-Lyons, Lori Hatzinger, Micaela Hewus, Wei-Lin Wang, Delfien Van Dyck, Jimi Huh, Eric Hekler, Genevieve F Dunton. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 20.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Summative Usability Evaluation of an Infusion Pump Through Simulation-Based Testing With Nurses: Mixed Approach Study

Yingjun Wan[1,2], BEng; Yun Zhang[3], MEng; Yunlong Zhao[3], MEng; Han Yu[1,2], PhD; Kaifeng Liu[1,2], PhD

[1]Academy of Medical Engineering and Translational Medicine, Tianjin University, Tianjin, China

[2]State Key Laboratory of Advanced Medical Materials and Devices, Tianjin University, Tianjin, Tianjin, China

[3]Tianjin Key Laboratory of Quality Control and Evaluation Technology for Medical Devices, Tianjin, China

Corresponding Author:
Kaifeng Liu, PhD
Academy of Medical Engineering and Translational Medicine
Tianjin University
#92 Weijin Road, Nankai District
Tianjin, 300072
China
Phone: 86 02283612122
Email: kf_liu@tju.edu.cn

Abstract

Background: Suboptimal design of infusion pumps may lead to usage errors, thereby compromising patient safety. Usability evaluation enables medical device design teams to identify and rectify design-related usability issues in a timely manner. Nevertheless, existing research on infusion pump usability continues to exhibit limitations in aspects such as task design.

Objective: The study aimed to evaluate the usability of an infusion pump (SLGO SP-200 [SLGO Medical Technology Co, Ltd]) through simulation-based testing with nurses in a usability laboratory designed to simulate an intensive care unit.

Methods: A total of 12 registered nurses with experience in using infusion pumps participated in this study. Nurses were asked to perform 12 operational tasks using the infusion pump. The participants were also asked to perform 7 knowledge tasks, where they were required to find relevant information in the user manual. Participants' behavioral measures (task completion time, frequency of manual query, frequency of asking for assistance from researchers, frequencies of operation difficulties, near-misses, and failures), perceptions (perceived ease of use, perceived concentration level required, perceived likelihood of making programming errors, perceived mental workload, satisfaction, and use intention) were collected to evaluate the usability and identify interface design deficiencies of the pump.

Results: The study found that the participants were generally able to complete the tasks. All operational tasks were completed within 3 minutes, and all knowledge tasks were completed within 2 minutes. Our study identified 79 difficult operations, 9 near-miss operations, and 36 operation failures. The causes of the above problems were analyzed. Participants generally found the infusion pump to be user-friendly, requiring a medium level of attention resources, and reported low levels of mental workload and likelihood of making programming errors.

Conclusions: The study results can provide a basis for the design of infusion pumps, help practitioners define the risks of use and the key content of training, and provide an important reference for the design of usability evaluation schemes for medical devices.

(JMIR Hum Factors 2026;13:e86443) doi:10.2196/86443

KEYWORDS

usability; medical device; infusion pump; simulated use; use error

Introduction

Infusion pumps are medical devices that deliver fluids, medications, or nutrients directly into the patient body in a controlled manner [1]. Well-designed infusion pumps of good quality are necessary for providing effective and safe patient care. However, infusion pumps are often involved in incidents and accidents [2-4]. A White Paper published by the US Food and Drug Administration (FDA) in 2010 (White Paper: Infusion Pump Improvement Initiative) indicated that between 2005 and

XSL•FO
RenderX

2009, the FDA received approximately 56,000 adverse event reports related to the use of infusion pumps, and there were 87 infusion pump models recalled due to safety concerns [4]. Other studies have indicated that the number of adverse events or near-miss incidents occurring during the use of medical devices is significantly higher than the data reported officially [5-7].

The poor human-machine interface design of medical devices (eg, unclear information presentation and ambiguous alarms) is one of the main reasons for use errors [5,8-11]. To address this issue, development teams for medical devices like infusion pumps need to incorporate usability engineering, which focuses on safety, efficiency, and user satisfaction during the use of medical devices [12,13]. Medical device regulators worldwide have published various guidelines and standards to integrate usability engineering into the device approval process [14]. In 2001, the American National Standards Institute (ANSI) and the Association for the Advancement of Medical Instrumentation (AAMI) released ANSI/AAMI HE74:2001, which described the usability engineering design process for medical devices; it was one of the earliest standards emphasizing usability [15]. In 2007, the International Electrotechnical Commission (IEC) released IEC 62366, detailing the usability engineering design process for medical devices [16]. This standard was updated in 2015 to further elaborate on the design process, considering risk management. In China, usability engineering for medical devices is still in its early stages. In 2020, the Center for Medical Device Evaluation of the National Medical Products Administration organized the drafting of the guidelines for technical review of human factors design for medical devices (draft for comments). In March 2024, the guideline was officially published, providing standards and regulations that define how usability testing should take place [17].

Usability testing is a crucial process within usability engineering, which consists of formative and summative testing. Formative testing is primarily conducted during the design and development phases of medical devices, while summative testing is conducted after the development phase to ensure that the designed device sufficiently meets user needs [18]. Simulated use testing is a common method of usability testing, where representative users are invited to use the medical device to perform representative tasks in a realistic setting [19-21]. This method allows for the observation and collection of user behavior data, revealing interface design flaws and identifying interaction strengths and opportunities for improvement.

Several previous studies have empirically evaluated the usability of infusion pumps [5,21-23]. For instance, Garmer et al [5] conducted a usability experiment to compare 2 interfaces for an infusion pump. Experimental tasks included setting or adjusting a flow rate or volume to be infused and starting or stopping an infusion. Liu et al [21] performed a heuristic evaluation and simulated use testing to compare the usability level of 4 infusion pumps to inform pump selection. Trbovich et al [22] examined how traditional pump, smart pump, and smart pump with barcode affected nurses' ability to safely administer intravenous medications. Schmettow et al [23] conducted a usability study to compare a novel syringe infusion pump prototype to an existing design. However, existing research still faces a series of limitations. First, previous usability evaluations only considered operational tasks related to pump use, while neglecting knowledge tasks associated with the use of the instruction manual. As an integral component of the system, the clarity, comprehensibility, accessibility, and instructional value of the manual also require evaluation [24]. Second, most evaluation tasks focused solely on core functions, such as infusion programming, with insufficient attention given to alarm handling and accessories installation. Third, current evaluations focused on observing operational difficulties without summarizing the underlying causes and potential risks of these issues. Through linking identified usability issues to potential use-related risks, it can offer a direct pathway from evaluation findings to specific, actionable design improvements aimed at risk mitigation [23].

The importance of usability testing for medical devices is gaining attention among professionals worldwide, highlighting the urgent need for research to establish guidelines for typical usability testing methods, processes, and data analysis. In light of this, this study invited registered nurses to conduct a summative usability test on a new infusion pump, aiming to identify human-machine interface design flaws and provide guidance for the design, evaluation, training, and use of infusion pumps.

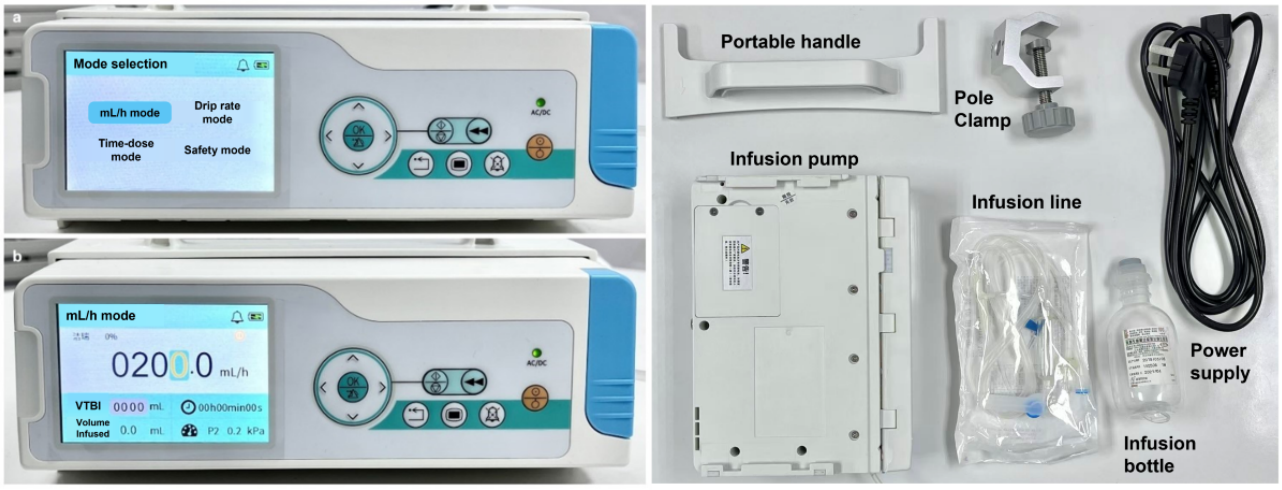
Methods

Design

A summative usability evaluation of an infusion pump (SLGO SP-200, SLGO Medical Technology Co, Ltd; Figure 1) was conducted, where participants were required to complete a series of operational tasks and knowledge tasks in a simulated intensive care unit (ICU). During the evaluation, behavioral data (eg, task completion time, frequency of asking for help, frequency of difficult operation, frequency of near-miss operation, and frequency of operation failure) as well as subjective data (eg, subjective perceptions and interviews) were collected.

XSL•FO
RenderX

Figure 1. The infusion pump being evaluated—main interface (left) and its accessories (right).



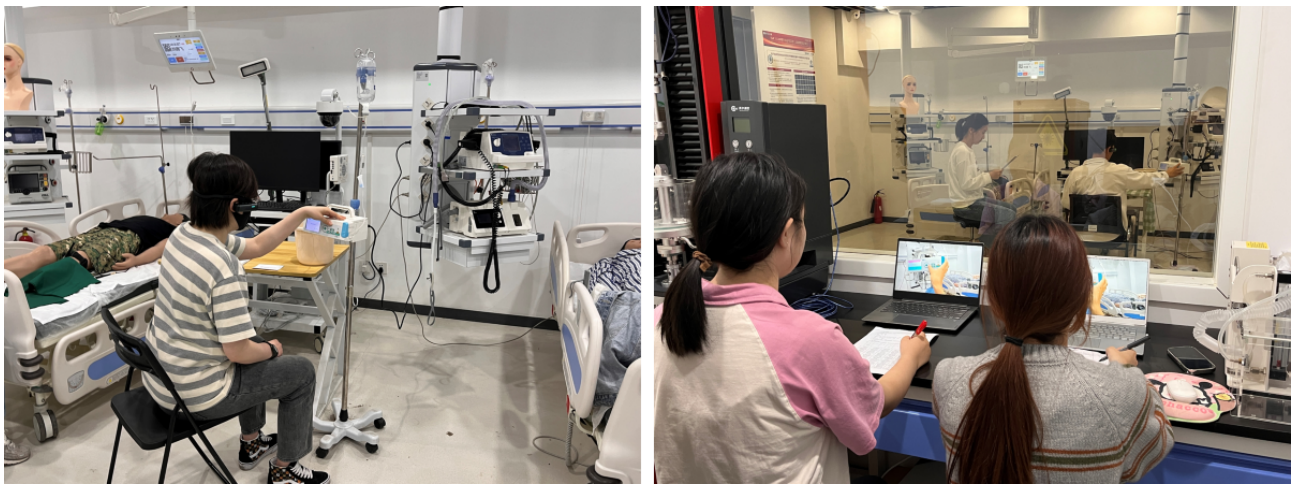
Participants

A total of 12 nurses with experience in using infusion pumps were recruited from 2 hospitals in Tianjin, China. The participants were stratified by clinical unit (specialized or general hospitals), department (ICU, operating room, or medicine), and clinical experience (≤8 y or >8 y) [21]. The participants were aged between 25 and 55 years (mean age 35.3, SD 8.1 y). Previous literature in usability engineering indicated that 80%-90% of usability problems can be detected with 5-8 participants; beyond this point, the incremental yield of new issues diminishes significantly [25,26]. All participants received training on the infusion pump being evaluated before the evaluation.

Environment and Apparatus

The experiment was conducted in a usability laboratory designed to simulate an ICU. The laboratory consists of a testing room and an observation room, separated by a one-way glass. Participants performed experimental tasks in the testing room while wearing a head-mounted camera (Lenovo-Lx918) to continuously record their operations in real time. Observers in the observation room monitored the participant's operations through a computer screen displaying the live feed from the head-mounted camera (Figure 2).

Figure 2. The testing room (left) and the observation room (right) of the usability laboratory.



Task

A total of 12 operational tasks were designed (Table 1), including accessories installation, infusion programming, system configuration, and alarm handling. The 2 alarm handling tasks (tasks 8 and 10) were predesigned as part of the scenario, and the alarms were manually triggered by a research assistant at a planned point during the participant posttask interview. This approach allowed researchers to integrate the alarm response naturally into the ongoing tasks without relying on the participant to inadvertently cause the alert condition. Specifically, the air bubble alarm (task 8) was triggered by the research assistant intentionally introducing a small air bubble into the infusion line. No power connection alarm (task 10) was triggered by the research assistant disconnecting the pump from its power source. Participants were not informed in advance that these alarms would occur, so that we could assess their natural recognition and response to unexpected yet clinically relevant alarms. Additionally, we designed 7 knowledge tasks (Table 2), where participants were required to find relevant information in the user manual.

XSL•FO
RenderX

Table 1. The 12 operational tasks.

Task	Task description	Standard operation steps
1	Please install the pole clamp and portable handle on the infusion pump. Then fix the pump securely on the pole stand.	• Install the pole clamp • Install the portable handle • Fix the pump on the pole stand
2	Please install the power supply and turn on the infusion pump.	• Install the power supply • Turn on the pump
3	Please install the infusion line.	• Push the pump door latch outward and fold down the front panel • Install the infusion line horizontally from right to left and secure it into the guide port • Close the pump door
4	Please adjust the brightness level of the screen to the highest level.	• Press the menu key to enter the system setup interface • Select the key for brightness settings • Use the directional keys to adjust the level
5	Please open the roller clamp on the infusion line. Set the drip rate to 50 drops/min in drip mode. Start infusion after priming.	• Open the roller clamp • Select the drip rate mode on the main interface • Set the drip rate to 50 drops/min • Press the prime and bolus key to purge air from the infusion line • Press the start and stop key to start the infusion
6	Please set the volume to be infused to 58 mL and the infusion duration to 60 min in the time-dose mode, then start the infusion.	• Press the start and stop key to pause the previous infusion • Press the return key to enter to the main interface • Select the time-dose mode • Set the volume to 58 mL and the duration to 60 min • Press the start and stop key to start the infusion
7	Please adjust the flow rate to 254.5 mL/h in the mL/h mode and then start the infusion.	• Press the start and stop key to pause the previous infusion • Press the return key to enter to the main interface • Select the mL/h mode • Set the flow rate to 254.5 mL/h • Press the start and stop key to start the infusion
8	Please address the current alarm (ie, air bubble alarm) and then continue the infusion.	• Press the return key • Press the prime and bolus key until bubbles are discharged from the very end of the infusion line. • Press the start and stop key to restart the infusion.
9	Please set the flow rate to 99 mL/h and the preset volume to 1 mL in safe mode, then start the infusion.	• Press the start and stop key to pause the previous infusion. • Press the return key to enter to the main interface. • Select the safe mode. • Set the flow rate to 99 mL/h and preset volume to 1 mL • Press the start and stop key to start the infusion
10	Please address the current alarm (no power connection).	• Press the start and stop key to pause the previous infusion • Connect the power supply to the pump
11	Please turn off the pump, remove the infusion line, disconnect the power, and take the infusion pump off the stand.	• Turn off the pump • Disconnect the power supply, first the AC power, then the connection to the infusion pump • Remove the infusion line and close the pump door • Take the infusion pump off the stand
12	Please remove the pole clamp and the portable handle from the infusion pump.	• Remove the pole clamp • Remove the portable handle

XSL•FO
RenderX

Table 2. The 7 knowledge tasks.

Task	Question	Answer
1	What is the minimum distance required between the infusion pump and high-frequency surgical equipment?	At least 0.5 m
2	What are the contraindications for the use of the infusion pump?	It is prohibited for blood transfusion, insulin infusion, as well as analgesia, chemotherapy, and epidural anesthesia infusion
3	What are the infusion pump's alarm levels and their corresponding indicator colors?	Red for high priority, yellow for low priority
4	What is the normal operating temperature range for the infusion pump?	+5 °C to +40 °C
5	What is the required distance range between the infusion bottle and the infusion pump?	30 cm to 50 cm
6	Please confirm the meaning of the symbol. 	Confirm and alarm reset key. In the parameter setting interface, it cycles through editable parameters; during an alarm, it performs an alarm reset operation.
7	What is the troubleshooting method for an "Occlusion" alarm on the infusion pump?	Press the start and stop key to pause the infusion. Check if the tubing is kinked, if the roller clamp is closed, or if the needle is blocked. If the issue is due to medication viscosity, the pressure setting can be adjusted.

Procedures

Upon arriving at the laboratory, the researcher first introduced the objective and procedure of the evaluation to the participants. Participants were required to complete 12 operational tasks and verbally report "task completed" after finishing each one. They were instructed to use the infusion pump in the same manner as they would in their daily work. If they encountered difficulties, they could refer to the manual. If they were still unable to solve the problem after referring to the manual, they could seek assistance from the researchers. Participants were also required to complete 7 knowledge tasks, in which they were required to search the manual for relevant information and answer the questions. After completion of all tasks, participants filled out a questionnaire [21,27], which included 6 dimensions—perceived ease of use, perceived mental workload, perceived concentration level required, perceived likelihood of making programming errors, satisfaction, and use intention (Table 3). All items were measured using a 7-point Likert scale (1=strongly disagree and 7=strongly agree). Posttask interviews were held to gather overall user experience and suggestions for pump design. For operational tasks, interview questions included "Do you recall any errors you made while performing the tasks?," "Can you remember any instances where you almost made a mistake but realized it immediately?," "What is your overall impression of using this infusion pump for the first time?," and "Do you have any suggestions for improving this infusion pump?" For knowledge tasks, interview questions included "Did you find it easy to locate the relevant information?," "How would you evaluate the clarity of the corresponding content in the user manual?," and "Do you have any suggestions for improving the content of the user manual?" The duration of the experiment was approximately 60 minutes.

Table 3. Items used to measure participants' perceptions.

Dimensions	Items
Perceived ease of use	• Item 1: It is easy to prime and load a set. • Item 2: It is easy to load and unload a set. • Item 3: It is easy to program a basic infusion. • Item 4: It is easy to program a drug infusion from scratch. • Item 5: It is easy to titrate flow rates. • Item 6: It is easy to turn the pump on and o .
Perceived mental workload	• Item 7: Setting up and programming the infusion pump require a high level of mental workload.
Perceived concentration required	• Item 8: Using this infusion pump requires my high level of concentration.
Perceived likelihood of making programming errors	• Item 9: The design of this infusion pump is complex, making it easy for me to make errors.
Satisfaction	• Item 10: Overall, I am satisfied with this infusion pump.
Use intention	• Item 11: Overall, I am willing to use this infusion pump.

XSL•FO
RenderX

Data Analysis

Data on task completion time, frequency of manual query, frequency of asking for assistance, and frequencies of difficult operations, near-miss operations, and operation failures for each operational task were collected by integrating observers' records and extracting videos. Task completion time was defined as the time taken by the participant to complete a task, measured from when the researcher finished reading the task instructions until the participant verbally reported the task completed. Frequency of manual query referred to the number of times a participant consulted the manual during the task due to encountering difficulties. The frequency of asking for assistance was the number of times a participant sought assistance from the researchers while completing the task. According to the National Medical Products Administration guideline [17], a difficult operation is referred to as task completion that meets requirements but falls short of expectations, such as prolonged operation times or reduced efficiency. This also included tasks completed by chance. Near-miss operation refers to tasks completed under conditions where unacceptable risks were nearly encountered. Operation failure referred to user tasks that did not meet the expectations or were not completed, including instances of exceeding time limits, making mistakes, omissions, or interruptions. Descriptive statistical analyses were conducted on the behavioral and subjective data. The content analysis method was used to analyze the interview data. All interviews were audio-recorded and transcribed verbatim to create a textual dataset for analysis. Furthermore, 2 researchers (YW and KL) independently reviewed transcripts for coding. Any discrepancies were resolved through discussion.

Ethical Considerations

This study has been approved by the Ethics Committee of Tianjin University (TJUE-2024-278). All participants provided written informed consent. Each participant received 200 RMB (approximately US $28.64) as compensation for their time and contribution upon completion of the evaluation. All participant data were kept confidential and anonymized in accordance with the approved ethical protocol.

Results

Operational Task

Tables 4 and 5 present the task completion time and frequency of manual query, asking for assistance, difficult operations, near-miss operations, and operation failures for each operational task. Task 5 (priming the infusion line and setting up an infusion in drip mode) recorded the longest completion time (149.8 s) and was associated with considerable operational challenges. For example, nurses found it difficult to access the infusion mode function and encountered issues with priming the infusion line, which has extended task completion time. Similarly, task 1 (installing the pole clamp, portable handle, and fixing the pump to the stand) and task 6 (setting an infusion in time-dose mode) exhibited a high mean completion time and substantial operational issues. We observed that participants encountered difficulties in installing accessories, which extended the task completion time. Additionally, some participants spent excessive time when using the time-dose mode for infusion setup due to not understanding of the terminology presented on the screen. In contrast, simpler tasks, such as task 2 (installing the power supply and turning on the pump), were completed quickly (26.2 s) with minimal operational issues. Task 3 (installing the infusion line), task 4 (adjusting the brightness level), task 7 (setting an infusion in mL/h mode), task 9 (setting an infusion in safe mode), task 11 (turning off the pump and removing the infusion line from the pump), and task 12 (removing the pole clamp and portable handle from the pump) were completed in moderate times but still involved several operational issues. Alarm-related tasks also showed varied performance. Task 10 (addressing a power connection alarm) was completed relatively quickly with less operational issues than task 8 (addressing an air bubble alarm).

Table 4. Task completion time for each operational task.

Task	Time (s), mean (SD)	Time (s), median (IQR)
1	134.3 (70.3)	142.0 (62.3-192.3)
2	26.2 (5.9)	25.0 (21.5-29.0)
3	46.2 (20.4)	39.5 (33.5-54.5)
4	43.9 (44.8)	22.5 (19.3-51.0)
5	149.8 (130.5)	97.0 (61.5-190.8)
6	134.0 (91.8)	97.0 (57.3-234.8)
7	69.4 (58.7)	49.0 (32.3-73.3)
8	94.4 (36.1)	101.5 (52.8-119.0)
9	44.3 (16.9)	38.5 (30.3-58.5)
10	36.8 (32.0)	25.5 (14.5-42.0)
11	65.0 (14.2)	60.0 (54.8-80.5)
12	47.6 (38.2)	30.0 (22.8-51.0)

XSL•FO
RenderX

Table 5. Frequency of manual query, asking for assistance, difficult operations, near-miss operations, and operation failures for each operational task.

Task	User manual query, n	Asking for assistance, n	Difficult operations, n	Near-miss operations, n	Operation failures, n
1	5	4	1	0	13
2	0	0	0	0	1
3	1	0	2	0	0
4	3	0	14	0	0
5	4	1	16	4	9
6	7	0	3	0	9
7	2	0	8	3	1
8	4	0	8	2	1
9	1	0	7	0	1
10	4	0	5	0	0
11	0	0	11	0	1
12	4	0	4	0	0
Total	35	5	79	9	36

Of the total recorded operations (Table 6), difficult operations were observed most frequently (n=79), with alarm resolution (n=15), screen brightness adjustment (n=12), turning off the pump (n=11), and accessing the infusion mode setting function (n=11) constituting the most prominent challenges. Difficulties in starting (n=8) and pausing an infusion (n=7), along with number adjustment (n=5), were also notable. Near-miss operations, although less frequent (n=9), were primarily characterized by occlusion alarms due to the roller clamp not being opened (n=5) and the use of an incorrect infusion mode (n=2). A total of 36 operation failures were documented. A number of nurses were not able to install the portable handle (n=8) or the pole clamp (n=4). Critical procedural failures involved not priming the infusion line correctly (n=5), not starting the infusion after programming (n=5), and failing to notice or address system alarms (n=7).

XSL•FO
RenderX

Table 6. Type and frequency of difficult operations, near-misses, and failures.

Type	Frequency, n
Difficult operations (n=79)	
Difficult to resolve alarm issues	15
Difficult to adjust the brightness level	12
Difficult to access to the function for infusion mode setting	11
Difficult to turn off the pump	11
Difficult to start the infusion	8
Difficult to pause the infusion	7
Difficult to adjust numbers	5
Difficult to remove the portable handle from the pump	4
Difficult to install the infusion line	2
Difficult to access to the function for brightness level adjustment	2
Difficult to install the pole clamp	1
Difficult to program infusion under the time-dose mode	1
Near-miss operations (n=9)	
Occlusion alarm triggered due to not opening the roller clamp	5
Used wrong mode for infusion	2
Air bubble alarm triggered due to incorrect priming	1
Input wrong number when programming an infusion	1
Operation failures (n=36)	
Installed the portable handle in wrong direction	8
Did not notice the alarm or do not fix the alarm issues	7
Did not prime the infusion line correctly	5
Did not start the infusion after programming	5
Unable to install the pole clamp	4
Installed the power supply in a wrong way	2
Input wrong number when programming an infusion	2
Used wrong infusion mode	1
Unable to install the portable handle	1
Did not open the roller clamp	1

Knowledge Task

Due to incomplete video recordings for 2 participants, the analysis of knowledge task was conducted for only 10 participants. All participants were able to complete the tasks within 2 minutes (Table 7). Among them, the accuracy of 4 knowledge tasks did not reach 100%.

Table 7. Completion time and accuracy rate of the knowledge tasks.

Task	Task completion time (s), mean (SD)	Accuracy rate (%)
1	60.2 (63.3)	100
2	82.4 (52.9)	50
3	98.3 (33.8)	60
4	59.3 (31.4)	70
5	91.6 (49.1)	100
6	35.1 (13.5)	100
7	75.7 (66.8)	90

XSL•FO
RenderX

Subjective Perceptions

The mean scores for perceived ease of use, satisfaction, and use intention were 6.0 (SD 0.7) or above, indicating that participants perceived the pump easy to use, were satisfied with it, and were willing to use it (Table 8). The mean scores for perceived mental workload and perceived likelihood of making programming errors were 2.7 (SD 1.1) and 2.1 (SD 0.9), respectively, suggesting that the participants did not experience significant cognitive load while using the pump. The mean score for perceived concentration level required was 4.2 (SD 1.3), indicating that participants held a neutral stance regarding the need for high levels of concentration when using the infusion pump.

Table 8. Subjective perception results.

Dimension	Mean (SD)
Perceived ease of use	
Overall (items 1-6)	6.3 (0.5)
Priming and loading a set	6.0 (0.9)
Loading and unloading a set	6.3 (0.9)
Programming a basic infusion	6.7 (0.8)
Programming a drug infusion from scratch	6.8 (0.5)
Titrating flow rates	6.2 (0.7)
Turning the pump on and o	6.0 (0.6)
Perceived mental workload	2.7 (1.1)
Perceived concentration level required	4.2 (1.3)
Perceived likelihood of making programming errors	2.1 (0.9)
Satisfaction	6.0 (0.7)
Intention to use	6.0 (0.7)

Interview

Reasons for Difficult Operations, Near-Miss Operations, and Operation Failures

During the installation and disassembly phases, participants encountered difficult operations, near-misses, and failures in key tasks, such as pole clamp installation or removal, portable handle attachment, power supply connection, and infusion line loading. The primary contributing factors were the users' unfamiliarity with accessory installation procedures and the absence of relevant instructional cues on the infusion pump or in its user manual (5/12, 41.7%). For example, as one participant noted,

> *I was unaware that the pole clamp required a pulling action. Even after consulting the diagrams in the instruction manual, I was unable to determine the correct method of operation.* [P7]

In system configuration, difficulties were observed in adjusting screen brightness and powering off the device. Several participants (5/12, 41.7%) reported that the brightness adjustment function was difficult to locate and there was no confirmation feedback after adjustment. For instance, one participant (P1) noted,

> *Following the brightness adjustment, selection of the 'OK' button provided no feedback, neither a return to the previous interface nor a confirmation, to indicate that the new setting had been implemented.* [P1]

The shutdown procedure also deviated from their familiar devices, resulting in prolonged completion times. For instance, one participant said,

> *Based on my prior experience with other infusion pumps, I expected that a long press would typically trigger a shutdown interface.* [P11]

During infusion setup, users experienced difficulties, near-misses, and failures in tasks including line priming, numeric input, mode selection, starting or pausing an infusion, and roller clamp operation. Several participants (4/12, 33.3%) indicated that the infusion mode setting function was not intuitively accessible, and terminologies were not clearly displayed, causing confusion. For example, one participant said,

> *I initially assumed that all mode selection options would be located within the menu system.* [P12]

Operational inconsistencies between this infusion pump and the participants' routinely used devices further contributed to uncertainty in initiating and pausing infusions (1/12, 8.3%). P4 noted a key operational discrepancy:

> *With our conventional infusion pumps, the infusion typically starts immediately upon pressing the confirmation button. However, this specific model requires an additional step of pressing a separate 'start' key to initiate the process.* [P4]

Unclear button labeling also increased the cognitive load during programming (2/12, 16.7%). P11 expressed confusion regarding button functionality, noting:

XSL•FO
RenderX

> *I used the 'OK' button to confirm system status, followed by the 'Pause' button to suspend operation, though I frequently encountered difficulty differentiating between these two controls.* [P11]

Moreover, omissions, such as incomplete priming, not opening the rolling clamp, and incorrect value input, were observed (7/12, 58.3%). For instance, participants said,

> *Initially, I had located the mL/h mode, but subsequently realized that the drip mode was the required setting for this specific task.* [P9]

> *I recognized that the occlusion was caused by not opening the roller clamp.* [P5]

In alarm handling tasks, several nurses (4/12, 33.3%) failed to notice active alarms, while others who did recognize them often could not interpret the alarm messages or determine appropriate actions to resolve the issues. For example, one participant indicated:

> *I observed that it did not affect the infusion process, so I did not address it.* [P5]

P7 reported confusion regarding a system alarm, stating:

> *Initially, I did not understand the meaning of Preset Complete, and consequently could not comprehend the reason for the alarm activation.* [P7]

Suggestions for Improvement of the Pump

Participants generally reported a favorable overall impression of the pump, with most describing it as user-friendly, easy to learn, and featuring a clear interface (11/12, 91.7%). Representative comments included:

> *The overall interface is somewhat better than the infusion pumps I have used before; for instance, the numerical values are more prominent.* [P1]

> *It is quite good that I could start using it without much difficulty on the first try.* [P5]

> *It is mostly similar to other infusion pumps; I should be able to use it by exploring a bit myself.* [P10]

> *The Chinese language is easy to understand, and setting the modes is relatively straightforward.* [P7]

Regarding comfort and weight, participants found the pump generally comfortable to handle (10/12, 83.3%), with comments such as,

> *It is still quite comfortable, and the key sensitivity is good.* [P1]

> *Acceptable, and not very heavy.* [P5]

> *The keys are comfortable, easy to press.* [P2]

Regarding the button design, participants suggested improvements to increase layout spacing, reduce ambiguity, and clarify key functions (6/12, 50%). For example, participants noted,

> *The operation keys are too crowded; making them slightly larger or spacing them further apart might make operation easier.* [P3]

> *The alarm reset key are sometimes confused.* [P4]

> *The start/stop key and the rapid bolus key are too close; I initially mistook the bolus key for the start/stop key.* [P5]

> *The menu and back keys are confusing; I could not find the mode adjustment interface.* [P6]

> *Making the start/stop key more prominent would be better, as starting and stopping are our most frequent actions.* [P9]

> *The meaning of the menu key is not entirely clear.* [P10]

Additionally, participants identified issues related to confirmation feedback, accessories installation, and infusion set loading (5/12, 41.7%). P2 highlighted the lack of feedback during adjustments:

> *Brightness settings and numerical input require some confirmation feedback.* [P2]

P7 expressed concern over functional ambiguity:

> *The OK key and the alarm reset key are combined, which I find confusing; I prefer each key to have a dedicated function.* [P7]

Regarding accessories, P3 raised a safety concern:

> *In clinical practice, infusion pumps are often suspended. The ones we use are one-piece, so we are concerned about whether this clamp pull-ring is secure enough to avoid detachment.* [P3]

Participants noted:

> *The infusion lines on the pumps I used were loaded vertically; I feel gravity might work better that way.* [P6]

> *Vertical loading is more convenient; horizontal loading requires figuring out the direction, which takes extra time.* [P8]

Further suggestions for design enhancement included increasing alarm volume, improving accessory installation, and incorporating touchscreen capability (4/12, 33.3%).

> *The alarm volume seems somewhat low; it is not very salient.* [P5]

> *It would be better if the pole clamp was permanently fixed to the pump. In clinical settings, we rarely need to detach it, and removable parts can get lost.* [P8]

> *Clinical settings have many tangled lines; it would be more convenient if the power cord could be stored attached to the pump after disconnection.* [P4]

P11 proposed using a touchscreen:

> *A touchscreen would be preferable, making numerical input less cumbersome.* [P11]

Suggestions for Improvement of the Manual

Participants generally perceived the user manual as clear and detailed (10/12, 83.3%), while proposing specific enhancements to optimize its utility. First, 2 (16.7%) nurses suggested improving the readability by enlarging the font size and

XSL•FO
RenderX

incorporating colorized interface diagrams to highlight critical information. For instance, participants mentioned,

> *The font size may be somewhat small for senior clinicians.* [P1]
>
> *The section illustrating the pump interface could be presented in color, matching the actual operational display, to facilitate easier use.* [P4]

Second, 1 (8.3%) participant advised consolidating alarm-related content by integrating alarm definitions with corresponding troubleshooting procedures. As P4 explained,

> *Placing the alarm descriptions and resolution methods on the same page would better support clinical workflow, since I typically identify the alarm type and immediately seek the solution.* [P4]

Third, nurses proposed adding a streamlined flowchart at the beginning of the manual to outline key clinical procedures (3/12, 25%). One participant suggested

> *…extracting a 2–3 page quick-reference guide covering essential operations and alarm handling for direct clinical guidance.* [P3]

P8 emphasized the need for "*an initial overview with a simple flowchart before detailed textual explanations*." Finally, 1 (8.3%) participant recommended supplementing certain operational details currently absent from the manual. P10 explicitly noted the absence of guidance on "how to install the pole clamp" as an example of missing critical information.

Remaining Usability Issues and Potential Risks

Table 9 presents the summary of remaining usability issues, potential risks, severity level, and corresponding design recommendations.

Table 9. Remaining usability issues, potential risks, and corresponding design recommendations.

Usability issues	Potential risk	Severity level	Design recommendations
Did not notice the alarm; difficult and unable to resolve alarm issues	Delay in responding to critical alarms	High	Clearly display the cause of the alarm and the steps to resolve it
Used wrong infusion mode	Dosage error	High	Distinctiveness of modes should be improved
Input wrong number when programming an infusion	Dosage error	High	A second confirmation for parameters is required
Difficult to program infusion under the time-dose mode	Inefficiency in programming and potential dosage error	High	Implement step-by-step instruction and clearly explain the meaning of the terminologies on the screen
Did not prime the infusion line correctly	Potential residual air in the infusion line	High	Visibility of pump status need to be clearly displayed; provide clearly instruction for priming
Did not open the roller clamp	Treatment delay	High	Visibility of pump status need to be clearly displayed; remind users to check the openness of the roller clamp before starting an infusion
Did not start the infusion after programming	Treatment delay	High	Visibility of pump status need to be clearly displayed; remind users to start infusion after programming, if not started within a set period, a reminder alarm should be triggered
Difficult to adjust numbers	Inefficiency in programming	High	Clearly indicate the method for rapid value entry
Difficult to start and pause the infusion	Treatment delay or unable to quickly stop the infusion	High	Provide a distinct start and stop key with clear labeling
Difficult to install the infusion line	Treatment delay	High	Provide clear instructions and incorporate error-proof design features
Difficult to find the function for infusion mode setting	Potential infusion in the wrong mode	High	Clear navigation
Difficult to find the function for brightness level adjustment and difficult to adjust the brightness level	Misreading or incorrect operation due to unclear screen display	Middle	Clear navigation; provide confirmation and feedback after settings
Difficult to turn off the pump	Inefficiency	Low	Provide clear instructions
Difficult or unable to install or remove the portable handle or pole clamp or power supply	Unstable or inconvenient device movement; Unstable installation or device falling; Power failure	Low	Provide clear instructions

XSL•FO
RenderX

Discussion

Overview

This study reported on a summative usability testing process for an infusion pump. The overall usability level of the pump was assessed by observing nurses' operating behaviors when using the pump. The study identified difficult operations, near-misses, and failures encountered during tasks, analyzed the causes of these issues through participant interviews, highlighted the design advantages of the infusion pump as well as existing interface design flaws, clarified usage risks, and proposed design improvement suggestions. These findings can provide valuable insights for optimizing the infusion pump design, helping practitioners define use risks and training priorities, and offering important implications for the usability evaluation schemes of medical devices.

Main Findings

Generally, the examined infusion pump effectively assisted participants in completing the tasks. Participants were able to complete all operational tasks within 3 minutes, indicating high overall operational efficiency. For knowledge tasks, although all were completed within 2 minutes, suboptimal accuracy was observed in a portion of tasks. In most cases, the participants completed tasks independently. They generally found the infusion pump to be user-friendly, requiring a medium level of attention resources, and reported low levels of mental workload and likelihood of making programming errors.

However, participants encountered several difficulties when operating the infusion pump, reflecting that there were still some design deficiencies that needed to be addressed. First, the infusion pump should increase the visibility of system status [28]. This issue has also been reported in other studies on infusion pump usability [21,23]. Low visibility of system status may prevent nurses from accurately determining the operational status of the infusion pump, thereby potentially leading to adverse events. Second, function navigation should be improved. Some nurses had difficulty in locating brightness adjustment and infusion mode setting functions. Liu et al [21] also observed that certain functional menus of infusion pumps were difficult to access. This may be attributed to a misalignment between the system interaction flow and user expectancy. A clear and well-structured navigation system can prevent user errors during operation, while excessive navigation options, overly deep menu hierarchies, and ambiguous categorizations can significantly increase users' memory load. Third, the infusion pump interface should be aligned with the mental model of target users, and the operating logic of different infusion pumps should be as consistent as possible. Some nurses reported differences in operational logic compared with pumps they regularly used, resulting in increased difficulty and errors during usage. This constitutes a violation of the consistency principle in human factors design [29]. When nurses encounter an interface that does not align with their established mental models, they are compelled to relearn and reconstruct new cognitive schemas. Fourth, many participants neglected alarms or were unclear about their meanings, which poses significant risks in clinical use. Previous research has indicated that a significant number of nurses failed to comprehend the meaning of infusion pump alarms clearly, leading to untimely or absent responses [21]. This represents a violation of the human factors design principle that systems should support users in recognizing, diagnosing, and recovering from errors. Effective error messages should use plain, unambiguous language, precisely identify the nature of the problem, and offer constructive, actionable guidance toward its resolution. Alarm design can also consider combining visual and auditory signals and avoid masking effects to ensure quick perception in complex clinical environments. Fifth, a few participants made errors in entering infusion parameters or using incorrect infusion modes. Other researchers have reported that confusion regarding dosing units and/or concentrations when using infusion pumps can lead to medication overdoses [22,30]. Input errors pose serious risks to patient safety, and while some infusion pumps include features to detect unsafe infusion rates [31], such errors may not be entirely avoidable through system design [5]. Finally, the clarity and comprehensiveness of the manual should be improved. Participants noted that a combination of images and text might enhance understanding more than using text alone.

Implications

Our study results provide several important implications. For clinicians and educators, the findings pinpoint areas requiring focused training and competency assessment, moving beyond general device familiarity to targeted practice on high-risk, error-prone steps. The problem analysis provided concrete criteria to inform procurement decisions, allowing buyers to compare devices based on specific safety-related usability attributes rather than features alone. For manufacturers, the results offered actionable design improvement recommendations directly linked to mitigating identified use errors, supporting a user-centered design and risk management process as mandated by regulatory frameworks.

Limitations

First, due to constraints related to the duration of the experiment and the availability of participants, we only recruited 12 participants. Given that our primary objective was to identify predominant usability challenges rather than to quantify rare events, this sample size was considered sufficient to capture the majority of issues. Future studies can consider increasing the sample size to ensure statistical power and generalizability. Second, no alternative system interfaces were assessed; therefore, no comparable results were available in our study. This may limit our findings and ideas on system interface improvement. Third, the sequence of tasks was designed to reflect the standard clinical workflow for operating an infusion pump; however, it may induce potential order effects. Future studies can consider randomizing the order of the tasks. Fourth, the impact of training on infusion pump use was not considered. Fifth, the experiment was conducted in a quiet simulated ICU, and nurses' use performance in real clinical scenarios (eg, noise, interruptions, and multitasking) may differ from the testing conditions. Finally, future tests could integrate devices like eye trackers to study the impact of infusion pump interface design on users' visual attention allocation.

XSL•FO
RenderX

Conclusion

This study recruited 12 nurses with infusion pump experience to conduct a simulated usability test of an infusion pump. It not only assessed the current usability of the infusion pump but also identified the root causes of user errors by combining subjective and objective data. This approach can help mitigate potential interaction design issues that could lead to user errors, providing important implications for the design and evaluation of medical devices.

Acknowledgments

The authors extend their grateful acknowledgment to all the participants from Tianjin University Chest Hospital and Tianjin Medical University General Hospital, for their invaluable assistance throughout the research process. KL would like to thank the support of Xiaomi Young Talents Program (Xiaomi Foundation).

Data Availability

Data are available on request from the corresponding author.

Funding

This work was supported by the National Key Research and Development Program of China (2023YFF1203800) and the Open Research Projects (2nd Batch) of the Tianjin Key Laboratory of Quality Control and Evaluation Technology for Medical Devices (TJMDKT-2026T03).

Authors' Contributions

YW conducted the experiment and collected, analyzed data, and drafted the manuscript. KL led the project and cross-checked the data analysis. All authors contributed to the study design and protocol development and participated in the writing and editing of several versions of this manuscript.

Conflicts of Interest

None declared.

References

1. Herrero L, Cano M, Ratwani R, Sánchez L, Sánchez B, Sancibrián R, et al. A review of human factors and infusion pumps: lessons for procurement. Front Digit Health 2025;7:1425409 [FREE Full text] [doi: 10.3389/fdgth.2025.1425409] [Medline: 39981104]
2. Husch M, Sullivan C, Rooney D, Barnard C, Fotis M, Clarke J, et al. Insights from the sharp end of intravenous medication errors: implications for infusion pump technology. Qual Saf Health Care 2005;14(2):80-86 [FREE Full text] [doi: 10.1136/qshc.2004.011957] [Medline: 15805451]
3. Webb RK, Russell WJ, Klepper I, Runciman WB. Equipment failure: an analysis of 2000 incident reports. Anaesth Intensive Care 1993;21(5):673-677. [doi: 10.1177/0310057x9302100533]
4. White paper: Infusion pump improvement initiative. The U.S. Food and Drug Administration. 2010. URL: https://www.fda.gov/medical-devices/infusion-pumps/white-paper-infusion-pump-improvement-initiative [accessed 2026-02-18]
5. Garmer K, Liljegren E, Osvalder AL, Dahlman S. Application of usability testing to the development of medical equipment. Usability testing of a frequently used infusion pump and a new user interface for an infusion pump developed with a human factors approach. International Journal of Industrial Ergonomics 2002;29(3):145-159. [doi: 10.1016/s0169-8141(01)00060-9]
6. Gaba DM, Howard SK, Small SD. Situation awareness in anesthesiology. Hum Factors 1995;37(1):20-31. [doi: 10.1518/001872095779049435] [Medline: 7790008]
7. Gagnon R, Laberge J, Lamsdale A, Histon J, Hudson C, Davies J, et al. A user-centered evaluation of three intravenous infusion pumps. : Sage Publications; 2004 Presented at: Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 2004 Jan 10; Los Angeles, CA p. 1773-1777. [doi: 10.1177/154193120404801531]
8. Graham MJ, Kubose TK, Jordan D, Zhang J, Johnson TR, Patel VL. Heuristic evaluation of infusion pumps: implications for patient safety in intensive care units. Int J Med Inform 2004;73(11-12):771-779. [doi: 10.1016/j.ijmedinf.2004.08.002] [Medline: 15491928]
9. Schnock KO, Dykes PC, Albert J, Ariosto D, Call R, Cameron C, et al. The frequency of intravenous medication administration errors related to smart infusion pumps: a multihospital observational study. BMJ Qual Saf 2017;26(2):131-140. [doi: 10.1136/bmjqs-2015-004465] [Medline: 26908900]
10. Su P, Liu K. Effects of anthropomorphic design on comprehension of self-monitoring test results: integrating evidence of eye-tracking and event-related potential. Displays 2024;82:102664. [doi: 10.1016/j.displa.2024.102664]

XSL•FO
RenderX

11. Liu K, Sun Z, Ren X, Tao D. Effects of information framing cues and age on the comprehension of personal health records for self-care behaviors: an eye-tracking study. J Am Med Inform Assoc 2025;32(7):1174-1185. [doi: 10.1093/jamia/ocaf085] [Medline: 40460029]
12. Infusion pump improvement initiative. Food and Drug Administration. 2017. URL: https://www.fda.gov/medical-devices/infusion-pumps/infusion-pump-improvement-initiative [accessed 2026-02-18]
13. Guidance on applying human factors and usability engineering to medical devices including drug-device combination products in great Britain. Medicines and Healthcare products Regulatory Agency. 2021. URL: https://assets.publishing.service.gov.uk/media/60521d98d3bf7f0455a6e61d/Human-Factors_Medical-Devices_v2.0.pdf [accessed 2026-02-18]
14. Bitkina OV, Kim HK, Park J. Usability and user experience of medical devices: an overview of the current state, analysis methodologies, and future challenges. International Journal of Industrial Ergonomics 2020;76:102932. [doi: 10.1016/j.ergon.2020.102932]
15. Association for the Advancement of Medical Instrumentation. Human Factor Design Process for Medical Devices (ANSI/AAMI HE74:2001). Washington (DC): National Academies Press; 2001.
16. Medical Devices - Application of Usability Engineering to Medical Devices. Switzerland, Geneva: International Electrotechnical Commission; 2007.
17. Guideline for Registration Review of Usability Engineering of Medical Devices. Austin, TX: National Medical Products Administration; 2024.
18. Lewis JR. Usability: lessons learned … and yet to be learned. International Journal of Human-Computer Interaction 2014;30(9):663-684. [doi: 10.1080/10447318.2014.930311]
19. Kim JH, Kim YR, Han MH, Han S, Jeon EH, Hwang EY, et al. A study on the formative usability testing for modular powered wheelchair. Disabil Rehabil Assist Technol 2025;20(2):452-460. [doi: 10.1080/17483107.2024.2372310] [Medline: 38965819]
20. Domanska B, Franchini A, Boroojerdi B, Ashraghi M, Ruzhansky K, Leite MI. Usability of the zilucoplan prefilled syringe for the treatment of generalized myasthenia gravis: insights from a human factors validation study. Expert Rev Med Devices 2024;21(10):977-982 [FREE Full text] [doi: 10.1080/17434440.2024.2403700] [Medline: 39403865]
21. Liu K, Chan FY, Or CK, Sun DTF, Lai WS, So HY. Heuristic evaluation and simulated use testing of infusion pumps to inform pump selection. Int J Med Inform 2019;131:103932. [doi: 10.1016/j.ijmedinf.2019.07.011] [Medline: 31557700]
22. Trbovich PL, Pinkney S, Cafazzo JA, Easty AC. The impact of traditional and smart pump infusion technology on nurse medication administration performance in a simulated inpatient unit. Qual Saf Health Care 2010;19(5):430-434 [FREE Full text] [doi: 10.1136/qshc.2009.032839] [Medline: 20427310]
23. Schmettow M, Schnittker R, Schraagen JM. An extended protocol for usability validation of medical devices: research design and reference model. J Biomed Inform 2017;69:99-114 [FREE Full text] [doi: 10.1016/j.jbi.2017.03.010] [Medline: 28336478]
24. Cherne N, Moses R, Piperato SM, Cheung C. Research: how medical device instructions for use engage users. Biomed Instrum Technol 2020 Jul 01;54(4):258-268. [doi: 10.2345/0899-8205-54.4.258] [Medline: 33171504]
25. Wiklund ME, Kendler J, Strochlic AY. Usability Testing of Medical Devices. USA: CRC Press; 2015.
26. Virzi RA. Refining the test phase of usability evaluation: how many subjects is enough? Hum Factors 1992;34(4):457-468. [doi: 10.1177/001872089203400407]
27. Ginsburg G. Human factors engineering: a tool for medical device evaluation in hospital procurement decision-making. J Biomed Inform 2005;38(3):213-219 [FREE Full text] [doi: 10.1016/j.jbi.2004.11.008] [Medline: 15896694]
28. Ali H, Li H, Lisboa P. Usability analysis and redesign of infusion pump user interface. : SAGE Publications; 2015 Aug 13 Presented at: Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care; 2015 Aug 13; New Delhi, India p. 129-133. [doi: 10.1177/2327857915041027]
29. Nielsen J. Enhancing the explanatory power of usability heuristics. 1994 Presented at: Proceedings of the SIGCHI conference on Human Factors in Computing Systems; 1994 Apr 24; New York, NY, USA p. 152-158. [doi: 10.1145/191666.191729]
30. Bates DW, Vanderveen TW, Seger DL, Yamaga C, Rothschild JM. Variability in intravenous medication practices: implications for medication safety. Jt Comm J Qual Patient Saf 2005;31(4):203-210. [doi: 10.1016/s1553-7250(05)31026-9] [Medline: 15913127]
31. Giuliano KK. Intravenous smart pumps: Usability issues, intravenous medication administration error, and patient safety. Crit Care Nurs Clin North Am 2018;30(2):215-224 [FREE Full text] [doi: 10.1016/j.cnc.2018.02.004] [Medline: 29724440]

Abbreviations

ANSI: American National Standards Institute
AAMI: Association for the Advancement of Medical Instrumentation
FDA: US Food and Drug Administration
ICU: intensive care unit
IEC: International Electrotechnical Commission

XSL•FO
RenderX

Edited by A Parush; submitted 24.Oct.2025; peer-reviewed by H Wang, K Giuliano; comments to author 15.Nov.2025; revised version received 23.Jan.2026; accepted 09.Feb.2026; published 09.Mar.2026.

Please cite as:
Wan Y, Zhang Y, Zhao Y, Yu H, Liu K
A Summative Usability Evaluation of an Infusion Pump Through Simulation-Based Testing With Nurses: Mixed Approach Study
JMIR Hum Factors 2026;13:e86443
URL: https://humanfactors.jmir.org/2026/1/e86443
doi:10.2196/86443
PMID:

©Yingjun Wan, Yun Zhang, Yunlong Zhao, Han Yu, Kaifeng Liu. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 09.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Comprehensive Self-Medication Management System to Reduce Medication Errors Among People Living With Dementia: Mixed Methods Feasibility and Usability Study

Santosh Basapur[1*], MS, PhD; Charles Gellman[2*], MS, MBA; Jamie B Plenge[3*], BS; Amelia Troutman[3*], MA, MS; Erin Yurko[3*], BS; Brandon Woolsey[2*], BS; Justin McClendon[2*], MS, MPH; Lisa Marceau[2*], MPH; Neelum T Aggarwal[3,4*], MD

[1]Department of Family and Preventive Medicine, Rush University Medical Center, Chicago, IL, United States

[2]HiDO Technologies, Folsom, CA, United States

[3]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, United States

[4]Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, United States

[*]all authors contributed equally

Corresponding Author:
Santosh Basapur, MS, PhD
Department of Family and Preventive Medicine
Rush University Medical Center
1700 West Van Buren St
Suite 470
Chicago, IL, 60612
United States
Phone: 1 6302540490
Email: santosh_basapur@rush.edu

Abstract

Background: Medication adherence is a critical challenge for people living with dementia and their caregivers. Standard care relies on appropriate medication management, yet there are few effective options beyond manual pill-counting and caregiver-administered dosing. These methods are prone to errors and impose a significant burden. Technologically enhanced adherence tools include smart caps, reminder apps, and electronic dispensers, which improve tracking but still depend on manual interaction, lack clinical integration, and are often unsuitable for cognitive impairment. The HiDO HomeCare System (HCS) is an artificial intelligence–enabled self-medication device, advancing the field by removing manual pill counting, automating the chain-of-custody, verifying consumption, and logging medication adherence through neuroscience-based logic and real-time monitoring.

Objective: This study evaluated the feasibility, usability, and performance of the HCS for at-home medication management in dyads of people living with dementia and their caregivers. We examined setup, dispensing accuracy, task efficiency, and user satisfaction.

Methods: A pooled analysis usability study was conducted with 35 caregiver-patient dyads at Rush University Medical Center, combining 2 sequential in-clinic cohorts run at different time points with the same protocol. Participants were recruited from the Rush Memory Clinic and Rush Alzheimer's Disease Center data repository. Participant dyads completed device setup, medication dispensing, and simulated medication use with the system's automated logging and dual-camera verification. Dyads repeated dispensing and simulated medication use following automated reminders sent to their mobile device. Dyads were encouraged to repeat dispensing tasks multiple times. Quantitative measures included setup time, time to dispense, success rates, reliability, and System Usability Scale scores. Qualitative measures also captured caregiver perceptions of usability, acceptability, and burden.

Results: All 35 dyads successfully completed at least one dispensing task using the HCS. The average time for the first dispense attempt was 1:41 (SD 1:13) minutes (n=35). The second attempt averaged 1:35 (SD 1:10) minutes (n=21). Attempt 3 averaged 2:03 (SD 1:34) minutes (n=6). The system maintained accuracy across all users, with some variability in timing across age groups. The HCS received an overall mean System Usability Scale score of 70.2 (SD 18.2; n=34), reflecting above-average usability of

XSL•FO
RenderX

the device. Caregivers reported that access to a system like HCS could reduce stress associated with medication administration and recommended improvements to specific design elements.

Conclusions: The HCS demonstrated early feasibility, accuracy, and usability as a self-medication device tailored to people living with dementia. By automating the medication safety chain from delivery through consumption, HCS reduces caregiver workload and enhances patient safety and medication management. These findings support HCS as a viable medication adherence solution, addressing limitations of prior devices and enabling better dementia care. Larger-scale, longitudinal studies are planned to examine clinical outcomes, caregiver burden, and cost-effectiveness within real-world home and community settings.

(JMIR Hum Factors 2026;13:e86828) doi:10.2196/86828

KEYWORDS

dementia; Alzheimer disease; medication adherence; self-medication device; caregiver burden; digital health; robotics; remote monitoring

Introduction

Medication Management Challenges

By 2060, the number of individuals living with Alzheimer disease and Alzheimer disease–related dementias in the United States is projected to double [1]. People living with dementia-related conditions currently have no known cure despite ongoing research, leaving patients primarily reliant on medication-based treatments to manage symptoms and preserve quality of life. People living with dementia have quality of life needs similar to those without dementia, including the ability to complete activities of daily living. Preserving these functions is critical to support independence, enable individuals to remain at home longer, and reduce both caregiver burden and cost of care [2]. Because medication is the primary treatment pathway, the need for appropriate, low-burden medication management is central to maintaining both cognitive and physical health [3]. This can be particularly challenging, as people living with dementia often have multiple chronic conditions such as cardiovascular disease, diabetes, and hypertension [4]. Polypharmacy and cognitive decline are strongly associated with morbidity, higher health care costs, and earlier institutionalization [3,5]. Inadequate management of medications for these conditions accelerates functional decline, increases avoidable hospitalizations, and shortens time at home.

Medication mismanagement is defined as missed doses, incorrect timing, or taking the wrong drug, and is a common and escalating challenge among people living with dementia [3]. As dementia progresses and patients struggle to manage both dementia-related medications and treatments for comorbid conditions, mismanagement leads to disrupted sleep, heightened anxiety, poor nutrition, and increased depression [6,7]. This, in turn, fuels a cycle of decline that often results in institutionalization. These challenges also contribute to caregiver burden, manifesting as stress, depression, and health deterioration that further increase risk for emergency visits and hospitalizations of caregivers, which increases risks to people living with dementia [3,8,9].

Research has established that stabilizing underlying chronic conditions is critical for preserving brain health in people living with dementia, and the most common comorbidities are associated with impaired cerebral blood flow and oxygen delivery, which in turn contribute to neuroinflammation and accelerated decline [10]. Thus, adherence to consistent medication use and ongoing monitoring can impact disease progression and stabilize cognition [11].

Medication mismanagement has far-reaching economic consequences. Hospital stays cost Medicare nearly 3 times more for dementia patients than for other older adults, and these hospitalizations often lead to long-term care facility admissions rather than a return home [1,12]. The average cost of a nursing home stay for dementia care is estimated at US $28,078 per patient every day a patient is prematurely removed from the home, which adds to the billions of dollars per year spent on caring for people living with dementia [13]. Over 11 million individuals provide unpaid care to people living with dementia, resulting in over 18 billion hours of care valued at nearly US $350 billion annually [1,14]. Premature institutionalization due to medication errors represents a modifiable factor that amplifies both direct health care expenditures and the emotional and financial burden on families.

Scientific Premise

Memory relies on distinct systems in the brain. Declarative memory, mediated by the hippocampus and related temporal lobe structures, supports conscious recall and is severely impaired in Alzheimer disease [15]. Procedural (habit) memory is managed in the neostriatum and often remains intact for patients with Alzheimer disease [16]. Preserved habit memory in this population supports unconscious, repetitive learning, allowing people living with dementia to maintain routine tasks such as brushing teeth, dressing, or taking medication, even when declarative memory fails. Recognizing and building on this distinction provides an evidence-based rationale for habit-centered interventions to improve medication management.

The HomeCare System

The HomeCare System (HCS) is an innovative self-medication device (SMD) that builds upon this neuroscience-informed foundation. Current models of SMDs rely heavily on user programming and frequent manual loading, which are contraindicated for patients with dementia. The HCS device reduces common home medication errors by automating safeguards to ensure that the right patient receives the correct medication at the correct dose, limiting manual aspects of medication management. The design reduces reliance on manual activities and leverages habit-forming memory systems through chaining logic, a behavioral method that breaks down complex

XSL•FO
RenderX

routines into manageable, repeatable steps. This approach fosters habit formation and supports independence without requiring explicit recall. Subsequent dispensing and monitoring activities are conducted by the system. The device is designed to reduce errors and relieve caregiver burden by automating the medication safety chain from delivery to consumption. Core functions include (1) encoded medication data into RFID-enabled caps with automated detection of dosage and scheduling for up to 90-day supplies, (2) automated dispensing with chain-of-custody verification and logging, (3) dual-camera FaceID authentication, (4) clinical decision support for early intervention, and (5) customizable alerts and notifications for caregivers and providers delivered via mobile or electronic platforms. Real-time adherence monitoring is coupled with secure video capture of ingestion events, allowing for passive data collection on medication safety and user interactions.

The purpose of this study was to evaluate whether HCS features met the usability needs of patient-caregiver dyads. This was assessed by (1) in-clinic usability testing device set-up time less than 5 minutes, (2) completion of at least 75% of all usability testing tasks, (3) occurrence of no more than one noncritical error per dyad, and (4) mean System Usability Scale (SUS) score greater than 68. The secondary objective was to identify novel platform enhancements that could improve future device use in this population through qualitative feedback. Qualitative feedback was captured through group discussions with patient-caregiver dyads.

Methods

Study Design

HiDO Technologies, in collaboration with Rush University, conducted a human factors usability analysis study to validate the HCS prototype in dyads with people living with dementia and their caregivers. Two sequential in-clinic usability cohorts were run at different time points with the same protocol. Participants were recruited from the Rush Memory Clinic and Rush Alzheimer's Disease Center data repository. For the analysis presented here, data were combined into a pooled dataset, enabling integrated evaluation of usability outcomes and replication of findings. The study was conducted in-clinic and focused on evaluating the system's effectiveness in recognizing and accurately recording medication behaviors among persons with Alzheimer disease and related dementias.

Study Population and Recruitment Process

The study was conducted by the Rush Alzheimer's Disease Center at Rush University Medical Center, Chicago, Illinois. Participants were identified and recruited from 2 sources: the Rush Memory Clinic and the Rush Memory Clinic Data Repository. The Rush Memory Clinic, part of Rush University Medical Center and affiliated with the Rush Alzheimer's Disease Center, is a referral clinic that offers comprehensive clinical evaluations for complaints of memory loss or other cognitive and behavioral changes, and recommendations for managing these conditions. The Rush Memory Clinic Data Repository is a center-maintained institutional review board (IRB)–approved registry that collects and stores information that may help answer future questions about Alzheimer disease and other diseases and conditions associated with memory and aging. Registry participants are given the opportunity to indicate their willingness to be contacted about current and future research studies affiliated with the Rush Alzheimer's Disease Center.

Patients receiving clinical services at the Rush Memory Clinic and meeting preliminary inclusion criteria were identified and provided with information about the study (eg, IRB-approved study flyer) at the time of their clinical appointment or by phone afterwards. Registry participants meeting preliminary inclusion criteria who had indicated willingness to be contacted about future research studies were identified and contacted by phone by research staff to share study information and assess interest. All registry screening and study recruitment were carried out by research assistants at the Rush Alzheimer's Disease Center, specifically trained in cognitive research protocols.

Interested individuals were asked a set of screening questions to confirm eligibility, such as age and availability of a caregiver (age 18 years or older) willing to participate. Individuals with cognitive deficits and/or serious mental health or medical conditions that may have compromised safety or the ability to provide accurate user feedback were excluded.

Eligible and interested individuals were scheduled for an in-clinic visit to obtain informed consent and participate in the usability testing and group discussion session. A study-specific informed consent process was conducted using IRB-approved consent forms. Staff explained study aims, procedures, risks, and assessments, and used teach-back methods to confirm understanding. Written documentation of informed consent was obtained from both dyad members. All data were collected under the direction of the investigators.

Study Procedures

Once consented, dyads participated in usability testing and a group discussion session that included tasks related to the use of the HCS. HiDO staff, our human design expert (SB), and Rush Research Assistant team members gave a live demonstration of the HCS product and answered questions on device setup and medication dispensing prior to testing. There were 2 usability testing steps: setup and use. For the setup tasks, dyads were asked to conduct the HCS device setup, including completing facial recognition tasks and entering a security pin. The patients were required to complete this step, but were informed that if they needed assistance, they could consult their caregiver. Dyads were provided an opportunity to complete the setup up to 3 times. Repeated attempts were optional within the study visit, and lower numbers reflected participant preference rather than dropout after device failure.

Then, the patient was asked to use the HCS device to conduct a medication dispensing activity (vitamin C). The next activity was to successfully simulate "taking of medication" during the usability test, with video capture of this process. Taking medication was performed as a simulation, and no medication was actively taken. During device setup, dyads were asked to perform all device setup and registration steps using a "think-aloud" protocol. In a "think-aloud" protocol, the user completes each step in the process while saying it out loud. Participants were asked to say what came to their minds during

XSL•FO
RenderX

the task. The think-aloud process was recorded on paper data forms by the research assistants during the activity.

Once the device was enabled according to the protocol, patients were asked to complete a second set of predefined device use tasks: (1) recognize the medication reminder on the HCS device and the mobile phone that was in the room with the patient—each dyad had a mobile phone paired to the HCS, (2) unlock the HCS device with facial recognition, and (3) successfully simulate consumption of the dispensed medication (vitamin C). As with setup, this was a simulation, and no medication was taken.

The study team recorded critical and noncritical errors (predefined by the study protocol as developed by the HiDO team and the investigators at Rush) encountered by the participants after the 2-part usability study was completed with each dyad. A critical error was defined as an error that prevented full deployment of the system, preventing medication from being dispensed or dispensed correctly. A noncritical error was defined as an error that diverted the participant from the intended path but did not significantly affect the usability of the system.

Following the setup and use testing, dyads were asked to complete an SUS score with a 0-to-100 rating to measure the perceived usability of the system, as well as a 7-point likeability scale on their preferences related to the HCS. After SUS and likeability ratings were captured, dyads participated in semistructured group discussions. Group discussions included feedback about the device usability, relevance to use in the everyday home environment, enthusiasm, perceptions on setup, medication delivery experience, device ergonomics, alerting, challenges, and suggested improvements. The group discussion was conducted in English in a private room in the Rush Alzheimer's Disease Center conference room by the investigator (SB) with research assistants. The audio of the discussion was recorded and transcribed for analysis. Each session was scheduled for 60 minutes.

Data Collection

Time to set up the device, device task completion rate, number and type of critical and noncritical errors, and SUS scores were quantitatively measured. Setup time was documented in minutes and seconds, and successfully completed tasks were counted and numbered. SUS and likeability scores (reflecting questions targeting overall enthusiasm, setup, medication delivery, device ergonomics, alerting, and challenges) were recorded per dyad. Critical and noncritical errors in device setup and management were recorded, including when to take medication, retrieving the dispensed pill, taking the dispensed pill, and technical interactions with the device. Semistructured interviews were analyzed for targeted feedback about platform features. HCS video and other recorded data from the usability testing were captured and stored as audio files in the Microsoft Teams site of Rush Alzheimer's Disease Center for this study. Recorded think-aloud data and SUS scores were saved as Microsoft Word documents.

Ethical Considerations

The study was approved by the IRB at Rush University Medical Center (protocol 21033001-IRB01). All patient and caregiver participants provided written informed consent. All data used in this study were deidentified. Participants were compensated US $75 per dyad member (US $150 per dyad). The investigators and study staff monitored safety via adverse event reporting defined in the IRB-approved study protocol. Risks were minimal to participants, as they interacted with a device under close supervision by the study team in a clinical environment. No actual medications were administered during the study; vitamin C was dispensed by the platform but not consumed. Video capture was limited to the simulated dispensing session, stored in a secure deidentified form for analysis, and used only for protocol-defined usability assessment.

Statistical Analysis

The primary objective of this human factors usability study was to evaluate functional performance and user interaction with the HCS, not to test a clinical hypothesis. Power calculations were based on standards for usability testing rather than inferential statistical power. U.S. Food and Drug Administration human factors research guidelines suggest that a sample of 15-20 participants per user group is typically sufficient to identify at least 90% of usability issues within a defined task set [17]. Our pooled sample of 35 patient-caregiver dyads exceeded this benchmark and provided sufficient power to assess primary usability endpoints, including task success rate, time-on-task, and SUS scores.

The primary endpoint was the mean SUS score, with a prespecified benchmark of ≥68 indicating above-average usability. Secondary endpoints included setup time (<5 minutes), task completion rate (≥75%), and error rate (≤1 noncritical error per dyad). Given the iterative design of the HCS, 2 independent cohorts were analyzed to confirm the replicability of findings and assess potential improvements between study phases. The pooled analysis was used to ensure adequate power for quantitative comparisons while capturing qualitative insights across both groups.

Qualitative Data Analysis

All qualitative data collected during the usability sessions were analyzed to complement quantitative findings and identify areas for device optimization. Audio recordings from the "think-aloud" usability exercises and semistructured postsession interviews were transcribed verbatim and verified for accuracy by the research team at Rush Alzheimer's Disease Center.

A hybrid deductive-inductive approach was applied. Deductive codes were derived from the study's predefined domains of interest, including setup experience, device navigation, perceived usefulness, and caregiver burden. Inductive codes captured emergent insights related to user experience, accessibility, and design improvements. A trained researcher (SB) independently coded all transcripts. Codes were grouped into categories and overarching themes reflecting participant perceptions of usability, acceptability, and integration within daily routines. Given the formative usability purpose of the study, transcripts were coded by the trained researcher (SB) and discussed with the broader study team, including research assistants, to refine themes. This single-coder approach is a limitation.

XSL•FO
RenderX

A total of 35 transcripts were analyzed, and 547 quotes were coded, yielding 119 open codes and 5 primary themes that captured user experience variability across cohorts and age groups. These included (1) innovation and perceived value; (2) ease of use and learning curve; (3) physical design and environmental fit; (4) cognitive and emotional support for users; and (5) trust, safety, and the role of caregivers.

Results

Evaluation Results

User Statistics

A total of 316 dyads were contacted for study participation. Of those, 125 declined participation, 79 did not meet study eligibility, and 76 remained pending at the time enrollment closed (eg, did not return contact attempts and were still considering participation). A total of 35 dyads (people living with dementia + caregiver = dyad) participated in the study. Demographics and cohort details are defined in Table 1. Cohort 1 included 15 participants, and cohort 2 included 20 participants. People living with dementia were aged 50 years or older and diagnosed with mild cognitive impairment or early-stage dementia (Montreal Cognitive Assessment score 18 to 24, inclusive, or equivalent cognitive assessment) as noted in the chart.

Table 1. Sociodemographic characteristics by cohort and role (people living with dementia and caregivers).

Characteristics	Cohort 1		Cohort 2		Combined totals	
	People living with dementia (n=15)	Caregiver (n=15)	People living with dementia (n=20)	Caregiver (n=20)	People living with dementia (n=35)	Caregiver (n=35)
Age (years), mean (range)	72 (64-86)	64 (34-75)	75 (66-82)	69 (46-81)	74 (64-86)	67 (34-81)
Sex, n (%)						
Female	11 (73)	9 (60)	10 (50)	12 (60)	21 (60)	21 (60)
Male	4 (27)	6 (40)	10 (50)	8 (40)	14 (40)	14 (40)
Race, n (%)						
Black or African American	4 (27)	5 (33)	3 (15)	4 (20)	7 (20)	9 (26)
More than one race, Black or African American, White Asian	1 (7)	0 (0)	0 (0)	0 (0)	1 (3)	0 (0)
Prefer not to answer	0 (0)	0 (0)	1 (5)	0 (0)	1 (3)	0 (0)
White	10 (67)	10 (67)	16 (80)	16 (80)	26 (74)	26 (74)
Level of education, n (%)						
Associate's degree	1 (7)	1 (7)	0 (0)	0 (0)	1 (3)	1 (3)
Bachelor's degree or higher	11 (73)	11 (73)	13 (65)	17 (85)	24 (69)	28 (80)
Primary level education	0 (0)	0 (0)	0 (0)	1 (5)	0 (0)	1 (3)
Secondary level education, GED[a], or equivalent	1 (7)	1 (7)	2 (10)	1 (5)	3 (9)	2 (6)
Some college	2 (13)	2 (13)	4 (20)	1 (5)	6 (17)	3 (9)
Trade/vocational training	0 (0)	0 (0)	1 (5)	0 (0)	1 (3)	0 (0)
Marital status, n (%)						
Divorced	0 (0)	2 (13)	2 (10)	3 (15)	2 (6)	5 (14)
Married	12 (80)	12 (80)	16 (80)	16 (80)	28 (80)	28 (80)
Never married	1 (7)	1 (7)	1 (5)	1 (5)	2 (6)	2 (6)
Widowed	2 (13)	0 (0)	1 (5)	0 (0)	3 (9)	0 (0)

[a]GED: General Education Development.

Across both cohorts, the overall mean age of patients was 74 (range 64-86) years and caregivers averaged 67 (range 34-81) years. In cohort 1, people living with dementia averaged 72 (range 64-86) years and caregivers averaged 64 (range 34-75) years; in cohort 2, people living with dementia averaged 75 (range 66-82) years and caregivers averaged 69 (range 46-81) years. Overall, 60% (21/35) of the people living with dementia were female. By race, 74% (26/35) were White, 20% (7/35) Black or African American, 3% (1/35) reported more than one race, and 3% (1/35) preferred not to answer. The level of education was predominantly higher, with 69% (24/35) holding a bachelor's degree or higher, 17% (6/35) reporting some college, 9% (3/35) reporting secondary level education or General Education Development. Most participants were married

XSL•FO
RenderX

(28/35, 80%), with 6% (2/35) divorced, 6% (2/35) never married, and 9% (3/35) widowed. Patterns were similar across cohorts.

Setup Times

Across the key measures, the overall average initial setup time was 2:59 minutes (n=35). Setup time varied by cohort 1 (3:31 minutes) and cohort 2 (2:35 minutes), with cohort 2 requiring slightly less time than cohort 1 across age groups. See Tables 2 and 3.

Table 2. System usability score, setup, and task attempt completion by phase and age[a].

Cohort and age (years)	Participants, n	SUS[b], mean (SD)	Setup 1 (min:sec), mean (SD)	Setup 2 (min:sec), mean (SD)	Attempt 1 (min:sec), mean (SD)	Attempt 2 (min:sec), mean (SD)
Cohort 1						
64-69	7	83.6 (15.1) (n=7)	02:55 (01:13) (n=7)	N/A[c] (n=0)	00:57 (00:29) (n=7)	01:26 (00:25) (n=2)
70-74	2	73.8 (12.4) (n=2)	03:44 (00:23) (n=2)	N/A (n=0)	03:25 (02:50) (n=2)	03:22 (N/A) (n=1)
75+	6	75.4 (23.9) (n=6)	04:08 (01:51) (n=6)	02:18 (00:04) (n=2)	02:31 (00:58) (n=6)	01:48 (01:37) (n=5)
Overall	15	79 (18.2) (n=15)	03:31 (01:29) (n=15)	02:18 (00:04) (n=2)	01:54 (01:24) (n=15)	01:54 (01:23) (n=8)
Cohort 2						
64-69	4	66.7 (12.6) (n=3)[d]	02:47 (01:39) (n=4)	01:41 (00:24) (n=2)	01:54 (01:44) (n=4)	01:42 (01:50) (n=3)
70-74	4	61.2 (2.5) (n=4)	02:49 (00:30) (n=4)	N/A (n=0)	01:10 (00:30) (n=4)	00:53 (00:16) (n=4)
75+	12	63.1 (18.5) (n=12)	02:26 (00:47) (n=12)	N/A (n=0)	01:32 (00:59) (n=12)	01:32 (00:55) (n=6)
Overall (cohort 2)	20	63.3 (15.2) (n=19)[d]	02:35 (00:55) (n=20)	01:41 (00:24) (n=2)	01:32 (01:04) (n=20)	01:22 (01:02) (n=13)
All (both cohorts)	35	70.2 (18.2) (n=34)[d]	02:59 (01:16) (n=35)	01:59 (00:25) (n=4)	01:41 (01:13) (n=35)	01:35 (01:10) (n=21)

[a]Each cell includes the number of participants for that metric in that subgroup.

[b]SUS: System Usability Scale.

[c]N/A: not applicable.

[d]Indicates fewer SUS observations than the total number of participants per row. One participant had missing SUS scores.

XSL•FO
RenderX

Table 3. Overall system usability and task completion by age group[a].

Age (years)	SUS[b], mean (SD)	Setup 1 (min:sec), mean (SD)	Attempt 1 (min:sec), mean (SD)	Attempt 2 (min:sec), mean (SD)	Time improvement attempt 1>2 (min:sec)	Attempt 3 (min:sec), mean (SD)	Time improvement attempt 2>3 (min:sec)
64-69	78.5 (15.9) (n=10)	02:52 (01:18) (n=11)	01:18 (01:08) (n=11)	01:35 (01:19) (n=5)	+00:18 (n=5)	01:11 (N/A[c]:N/A), (n=1)	–00:24 (n=1)
70-74	65.4 (8.7) (n=6)	03:07 (00:38) (n=6)	01:55 (01:46) (n=6)	01:23 (01:08) (n=5)	–00:32 (n=5)	05:06 (N/A:N/A) (n=1)	+03:43 (n=1)
75+	67.2 (20.6) (n=18)	03:00 (01:27) (n=18)	01:52 (01:04) (n=18)	01:40 (01:13) (n=11)	–00:12 (n=11)	01:30 (00:34) (n=4)	–00:10 (n=4)
Overall	70.2 (18.2) (n=34)	02:59 (01:16) (n=35)	01:41 (01:13) (n=35)	01:35 (01:10) (n=21)	–00:07 (n=21)	02:03 (01:34) (n=6)	+00:28 (n=6)

[a]Each cell includes number of participants for that metric in that subgroup.

[b]SUS: System Usability Scale.

[c]N/A: not applicable.

Medication Reminder and Dispensation of Medicine

All participating dyads (n=35) successfully dispensed medications using the HCS (Tables 2 and 3). Participants were allowed to complete dispensing as many times as they desired, with attempts ranging from 1 to 5 across the sample. Each attempt required participants to respond to a reminder, unlock the device via FaceID, and dispense medications, with time intervals between steps ranging from 5 to 15 minutes from reminder to completion.

Across all 35 participants, the average time for the first dispense attempt was 1:41 minutes (n=35). The second attempt averaged 1:35 minutes (n=21). Attempt 3 was conducted in an average of 2:03 minutes (n=6). The 64-69-year-old age group completed the first attempt faster than the 70-74 or 75+ age groups. Time improvement between attempt 1 and attempt 2 improved overall by 0:07 seconds, with improvement for the 70-74-year-old (–0:32 seconds, n=5) and 75+ age group (–0:12 seconds, n=11), and +0:18 seconds (n=5) for the 64-69-year-old age group. Setup and attempt times between cohort 1 and cohort 2 varied, with overall times slightly faster in cohort 2.

SUS Ratings and Task Performance

The overall mean SUS score across all participants was 70.2 (SD 18.2; n=34) (Tables 2 and 3), reflecting an above-average usability score. Mean SUS scores differed between the two study cohorts. Cohort 1 participants reported a mean SUS score of 79.0 (SD 18.2; n=15), and cohort 2 participants reported a mean score of 63.3 (SD 15.2; n=19). In cohort 1, the 64-69 age group reported the highest score at 83.6 (SD 15.1; n=7), compared to 73.8 (SD 12.4; n=2) for 70-74-year-olds and 75.4 (SD 23.9; n=6) for the 75+ group. In cohort 2, the 64-69 group scored 66.7 (SD 12.6; n=4), the 70-74 group scored 61.2 (SD 2.5; n=4), and the 75+ group scored 63.1 (SD 18.5; n=12).

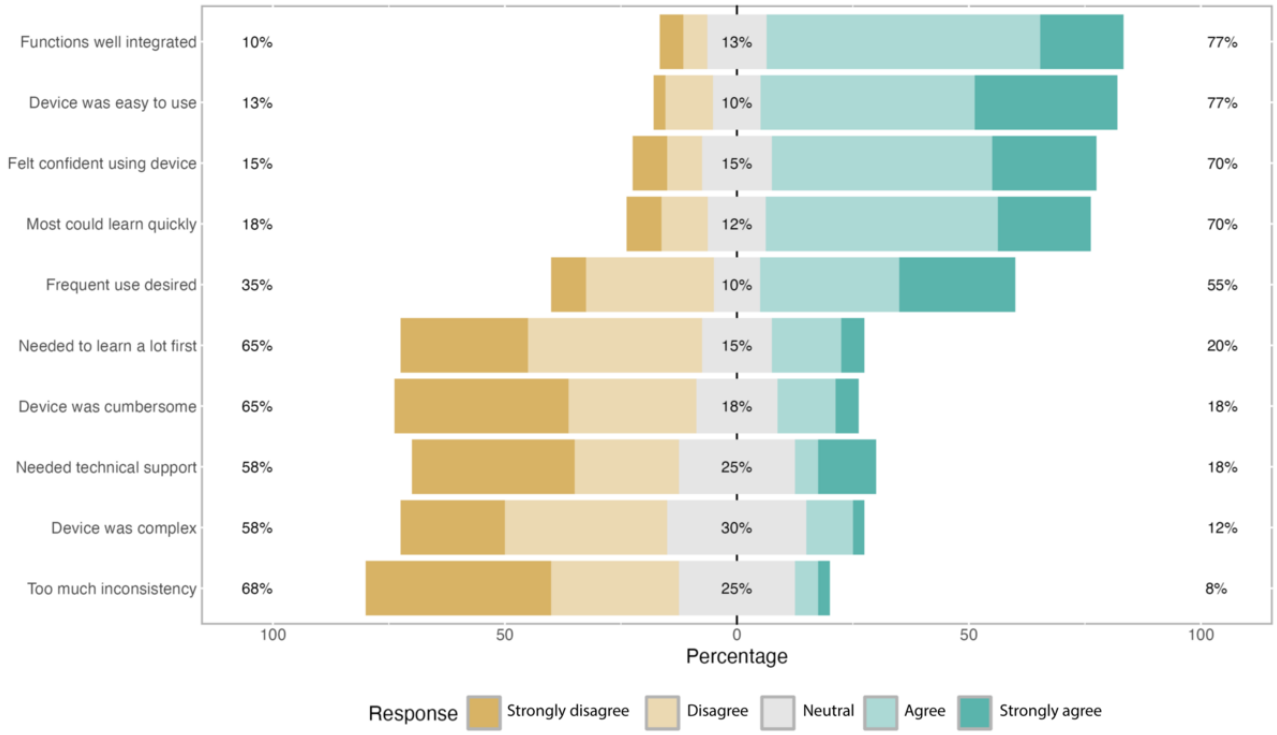
Analysis of the pooled data by age group is summarized in Tables 2 and 3. The 64-69-year-old age group overall reported the highest mean SUS rating at 78.5 (SD 15.9, n=10). This decreased to 65.4 (SD 8.7; n=6) for 70-74-year-olds and was 67.2 (SD 20.6; n=18) for the 75+ year-old cohort. Task performance data demonstrated that the overall mean time for the initial task attempt (attempt 1) was 01:41 (SD 1:13; n=35). The overall mean time for the subsequent attempt (attempt 2) was 01:35 (SD 1:10; n=21), resulting in an average time improvement of 7 seconds (–0:07; n=21). Time improvement varied by age group: the 70-74-year-old age group improved on average by 32 seconds (–0:32; n=5), while the 75+ year-old age group improved by 12 seconds (–0:12; n=11). The 64-69-year-old age group took an average of 18 seconds (0:18; n=5) on their second attempt.

Likert Scale Findings: Usability

Several findings were reported related to usability based on the SUS [17] (Figure 1). Most participants (27/34, 77%) agreed or strongly agreed that the device was easy to use, the device's functions were well integrated (27/34, 77%), and reported confidence using the device (24/34, 70%). Over half of the participants agreed or strongly agreed that they needed technical support to use the device (20/34, 58%), the device was complex (20/34, 58%), and they needed to learn a lot before using the device (23/34, 65%).

XSL•FO
RenderX

Figure 1. System Usability Scale scores on device usability.



Critical and Noncritical Errors

Analyses of participant errors of the pooled sample (n=35) are reported in Table 4. All errors were recorded, and more than one error could be attributed to one participant. Two critical errors were recorded in the pooled sample, both of which occurred in cohort 1 by two unique users. There were 17 noncritical errors recorded by 11 unique participants in the pooled sample. In cohort 1, eight participants made 14 noncritical errors, and in cohort 2, three participants made 3 noncritical errors. The 75+ age group demonstrated the highest frequency of noncritical errors, with 4 participants accounting for 8 total errors. Errors were tracked through the HCS video interface.

Table 4. Critical and noncritical errors by cohort and age group[a].

Cohort and age (years)	Participants, n	Participants with ≥1 critical error[b], n	Total critical errors[c], n	Participants with ≥1 noncritical error[b], n	Total noncritical errors[c], n
Cohort 1					
64-69	7	1	1	3	5
70-74	2	0	0	1	1
75+	6	1	1	4	8
Overall	15	2	2	8	14
Cohort 2					
64-69	4	0	0	1	1
70-74	4	0	0	0	0
75+	12	0	0	2	2
Overall	20	0	0	3	3
Overall (both cohorts)	35	2	2	11	17

[a]The table presents (1) the number of unique participants with at least one critical or noncritical error and (2) the total number of errors recorded for each group. “Overall” rows summarize all age groups within each cohort; the bottom row pools both cohorts. Errors were scored based on predefined definitions and coded in the Participant In-Clinic Usability Case Report Form.

[b]Participants with ≥1 error: number of people who made at least one error of that type.

[c]Total errors: sum of all observed errors, including multiple errors by the same participant.

XSL•FO
RenderX

Qualitative Results

Themes From Qualitative Data

Five primary themes emerged across the 35 dyad interviews which included (1) innovation and perceived value represented as enthusiasm for a technology that could reduce caregiver workload; (2) ease of use and learning curve, or the need for clearer setup instructions and guidance; (3) physical design and the environmental fit, accessibility, and ergonomics including preferences for larger fonts, higher contrast, and adjustable display angles; (4) cognitive and emotional support for users; and (5) trust, safety and the caregiver's role and support for shared use reflecting the desire for connected tools and training resources. Themes are summarized with illustrative quotes from conversations.

Theme 1: Innovation and Perceived Value

Participants viewed the HiDO system as an innovative solution to improve medication adherence and caregiver support. They valued the automation and accountability features, which simplified routines or reduced caregiver burden: "great technology for someone with memory issues." Some participants reflected satisfaction with pillboxes and found the device complex, while others felt it created accountability to take medications.

Theme 2: Ease of Use and Learning Curve

Participants described the device as "self-explanatory" and "easy once set up," yet several admitted to feeling "overwhelmed at first" or "intimidated by all the steps." Participants recommended simplifying the touchscreen interface with larger fonts, slower response time, and multimodal feedback (eg, auditory or visual cues). Most users achieved proficiency after repeated use, with one participant summarizing, "There's a learning curve, but it's simple once you get the hang of it."

Theme 3: Physical Design and Environmental Fit

Participants consistently described the device as bulky and difficult to integrate into home environments. Size, need for continuous power, and aesthetic incongruence were perceived as barriers. Many suggested a smaller or wall-mounted design, enhanced portability, and customizable color or appearance to reduce stigma, with participants noting, "I'd put it in my kitchen if it looked more like an appliance."

Theme 4: Cognitive and Emotional Support for Users

The device's reminder and monitoring capabilities provided reassurance and supported autonomy for some users. Participants valued multimodal alerts and personalized prompts and suggested these could be enhanced. Emotional themes centered on feeling empowered and less dependent on caregivers, with one participant summarizing, "It helps you feel you're still in control of your medications." Usability was seen as optimal for mild cognitive impairment but potentially limited for users with advancing dementia and suggested progressive usability that evolves with cognitive decline.

Theme 5: Trust, Safety, and the Role of Caregivers

Facial recognition and dispensing safeguards enhanced perceived safety, but users also expressed apprehension about malfunction or data misuse. The presence of caregivers was viewed as critical, especially for setup, refilling, and troubleshooting. As one caregiver noted, "You can't rely on the phone ping; you need something that tells me if they didn't take it." Participants highlighted the tension between independence and monitoring.

Discussion

Implication of Our Usability Evaluation

Our study assessed the usability and functionality of the HCS with 35 dyads, including setup and use capability, system usability, and qualitative feedback. Two cohorts with similar participants were included in a pooled analysis. On average, the overall time from reminder recognition to medication dispensation was 1.41 minutes (n=35), which was below the prespecified 5-minute threshold for device setup. Task performance data suggest the device was able to be set up and engaged with by the user population across age groups. There were noted differences by age and attempt, suggesting improvements when making repeated attempts, but the sample size was not designed for subgroup analysis. There were low instances of critical errors across the study, indicating the built-in safeguards were performing as designed. The noncritical errors (greater than or equal to 1) helped identify areas that need further work and modifications in product design, especially onboarding user experience and operating manual design.

The overall SUS ratings exceeded the threshold score of 68, typically considered above average, although scores varied between cohort 1 (79) and cohort 2 (63.3). Within age group variations are important to note for future research, but the sample size of 35 participant dyads was not designed to address subgroup analyses. The difference between the two cohorts may reflect iterative improvements in study implementation rather than inherent usability differences. Cohort 1 participants were the first users to engage in the HCS in this study, and likely had greater staff engagement to ensure the protocol was being followed and to address any unforeseen challenges. This could have contributed to slightly longer completion times in cohort 1. Cohort 2 may have benefited from refinements in user instructions and onboarding. Although both cohorts achieved above-average SUS scores for usability, the lower SUS scores in cohort 2 may also reflect slightly higher staff engagement with the setup and onboarding activities in cohort 1, which could have established a different rapport during the process than in cohort 2. SUS scores also represented valuable insights into usability for in-home SMDs with patients who are cognitively impaired. It adds value to the field that our data demonstrate an SMD like the HCS is reported as easy to learn and use, and participants felt confident using it, but further work is required to build in simple, clear, and precise operating and use instructions to meet the users' needs.

Qualitatively, participants recognized the HCS's potential to improve medication safety and reduce caregiver stress. Adoption was moderated by perceived complexity, physical design, and the degree of human connection integrated into the technology. Findings underscore that usability for populations with cognitive

XSL•FO
RenderX

impairment requires iterative design emphasizing simplicity, adaptability, and trust.

Accounting for the between-cohort differences, findings from this study make a significant contribution to the role of SMDs in supporting the challenges of medication adherence in people with dementia and cognitive impairment. Medication adherence tools have undergone substantial transformation since the 1950s, evolving across 4 generations. Early solutions (Generation 1) relied entirely on manual methods such as pill bottles and organizers, placing the full responsibility on patients and caregivers. Generation 2 introduced digital aids, including smart caps and mobile apps, which provided reminders and basic tracking but remained limited by their dependence on user input. Generation 3 technologies advanced to scheduled dispensers with remote monitoring features, yet they remained insufficient for individuals with cognitive decline, particularly those with Alzheimer disease. Current Generation 4 solutions seek to alleviate user burden by automating manual tasks and incorporating advanced monitoring and intervention features.

Principal Findings

The HCS represents a novel advancement within this fourth generation, extending usability to one of the most vulnerable patient groups—people living with dementia. This population faces unique challenges, including impaired cognitive recall, difficulties with complex device setup, and heavy reliance on caregivers. Findings from this study underscore the potential of the HCS to address these barriers. Participants emphasized the perceived usefulness of an automated medication management system, alleviating reliance on caregivers and reducing burden on patients by accessing habit memory. The automated reminders and facial recognition were noted as valuable for simplifying access, while participants also identified areas for improvement such as device size, display accessibility, and clearer setup support. These insights highlight both the feasibility and the refinement opportunities of a fully automated adherence system in real-world home settings.

Strengths and Limitations

This study has a few important limitations. Recruitment focused on patients who were already being treated in a clinical care setting, which may have limited the diversity of the participants to the available demographics of the Rush Medical Center. While efforts were made to include a diverse sample, the population for both patient and caregiver was two-thirds White and one-third Black, without broader racial and ethnic representation. The participant cohort was more likely to be educated, with most participants holding a bachelor's degree or higher. Because of these factors, our cohort may not represent all older adults taking multiple medications. However, as the primary outcome of this study was to demonstrate the feasibility of HCS to determine usability changes and enhancements in setup and use, a random and representative sample was not feasible or critical to achieve these initial goals. Two important strengths were noted. First, we conducted the study with patient and caregiver dyads in two separate but comparable cohorts, providing a larger sample and a broader dataset for analysis. Second, we incorporated qualitative methods, which enabled dyads to share valuable insights that could improve the HCS device for future implementation.

Conclusions

Demonstrating feasible use in people living with dementia, the HCS has the potential to improve the standard of care for medication adherence more broadly. Unlike prior generations of adherence tools, the HCS integrates features that account for both cognitive and caregiving demands, addressing a critical innovation gap specifically in a patient population with limited options for automated medication dispensing. Addressing usability and functionality of the HCS with patient and caregiver dyads advances the potential for more effective medication management in people living with dementia. Results of this study provide the foundation for future studies to measure medication management compliance, impact on disease progression, prolonged quality of life, and more effective communication between patient, caregiver, and clinical care team.

Acknowledgments

We gratefully acknowledge the contribution of the participants who agreed to participate and share their feedback.

Data Availability

The datasets generated or analyzed during this study are available upon reasonable request and in accordance with data sharing policies and with respect for intellectual property and proprietary data constraints. A request for data can be placed by emailing the corresponding author.

Funding

This study received funding from the National Institutes of Health, National Institute on Aging (grant R43AG077737). The National Institutes of Health did not have any input on the design, development, implementation, and reporting of the findings from the study.

XSL•FO
RenderX

Authors' Contributions

SB, NTA, JBP, AT, EY, JM, BW, and CG participated in the project's design and conduct. SB, NTA, JBP, AT, JM, LM, and CG were involved in the critical writing of the manuscript. SB, NTA, JBP, AT, BW, LM, and CG all participated in editing the manuscript. SB, JM, CG, and NTA also analyzed the data.

Conflicts of Interest

NTA reports being the site principal investigator for the NOURISH trial. SB reports being a coinvestigator on the NOURISH trial. CG is the chief executive officer and cofounder of HiDO Technologies. BW is the chief technology officer and cofounder of HiDO Technologies. Other authors have no competing interests to declare.

References

1. Alzheimer's Association. 2024 Alzheimer's disease facts and figures. Alzheimers Dement 2024;20(5):3708-3821 [FREE Full text] [doi: 10.1002/alz.13809] [Medline: 38689398]
2. Burks HB, des Bordes JK, Chadha R, Holmes HM, Rianon NJ. Quality of life assessment in older adults with dementia: a systematic review. Dement Geriatr Cogn Disord 2021;50(2):103-110. [doi: 10.1159/000515317]
3. Rankin A, Cadogan CA, Patterson SM. Interventions to improve medication management and adherence in people with dementia: a systematic review. Age Ageing 2024;53(2):afad257. [doi: 10.1093/ageing/afad257]
4. Adeyemi AH, Wiredu B, Okobi OE, Nebuwa P, Ezeani EI, Alozie AS. Challenges and strategies in medication management for patients with multiple comorbidities. Cureus 2025;17(6):e85992. [doi: 10.7759/cureus.85992] [Medline: 40662024]
5. Brown SH, Abdelhafiz AH. Institutionalization of older people: prediction and prevention. Aging Health 2011;7(2):187-203. [doi: 10.2217/ahe.10.88]
6. Petrovsky DV, McPhillips MV, Li J, Brody A, Caffeé L, Hodgson NA. Sleep disruption and quality of life in persons with dementia: a state-of-the-art review. Geriatr Nurs 2018;39(6):640-645 [FREE Full text] [doi: 10.1016/j.gerinurse.2018.04.014] [Medline: 29803606]
7. Botto R, Callai N, Cermelli A, Causarano L, Rainero I. Anxiety and depression in Alzheimer's disease: a systematic review of pathogenetic mechanisms and relation to cognitive decline. Neurol Sci 2022 Jul;43(7):4107-4124 [FREE Full text] [doi: 10.1007/s10072-022-06068-x] [Medline: 35461471]
8. Kilmer G, Omura JD, Bouldin ED, Walker J, Spears K, Gore J, et al. Changes in health indicators among caregivers – United States, 2015-2016 to 2021-2022. MMWR Morb Mortal Wkly Rep 2024;73(34):740-746. [doi: 10.15585/mmwr.mm7334a2] [Medline: 39207933]
9. Jaramillo Ramos JJ, Galindo Pupo NM, Mena D, Solis RP, Bedoya Jaramillo JE, Vega Solano M. Cognitive decline in chronic inflammatory conditions: exploring links between systemic inflammation and neurodegeneration. Cureus 2025;17(7):e88397. [doi: 10.7759/cureus.88397] [Medline: 40842786]
10. Religioni U, Barrios-Rodríguez R, Requena P, Borowska M, Ostrowski J. Enhancing therapy adherence: impact on clinical outcomes, healthcare costs, and patient quality of life. Medicina (Kaunas) 2025;61(1):153 [FREE Full text] [doi: 10.3390/medicina61010153] [Medline: 39859135]
11. Daras LC, Feng Z, Wiener JM, Kaganova Y. Medicare expenditures associated with hospital and emergency department use among beneficiaries with dementia. Inquiry 2017;54:46958017696757 [FREE Full text] [doi: 10.1177/0046958017696757] [Medline: 28301976]
12. Nandi A, Counts N, Bröker J, Malik S, Chen S, Han R, et al. Cost of care for Alzheimer's disease and related dementias in the United States: 2016 to 2060. NPJ Aging 2024;10(1):13. [doi: 10.1038/s41514-024-00136-6] [Medline: 38331952]
13. Caregivers of a person with Alzheimer's disease or a related dementia. Centers for Disease Control. 2024 Sep 3. URL: https://www.cdc.gov/caregiving/about/index.html [accessed 2026-04-06]
14. Squire LR. Memory and brain systems: 1969-2009. J Neurosci 2009;29(41):12711-12716 [FREE Full text] [doi: 10.1523/JNEUROSCI.3575-09.2009] [Medline: 19828780]
15. De Wit L, Marsiske M, O'Shea D, Kessels RPC, Kurasz AM, DeFeis B, et al. Procedural learning in individuals with amnestic mild cognitive impairment and Alzheimer's dementia: a systematic review and meta-analysis. Neuropsychol Rev 2021;31(1):103-114 [FREE Full text] [doi: 10.1007/s11065-020-09449-1] [Medline: 32897482]
16. Applying human factors and usability engineering to medical devices: guidance for industry and Food and Drug Administration staff. U.S. Food and Drug Administration. 2016 Feb. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/applying-human-factors-and-usability-engineering-medical-devices [accessed 2026-03-06]
17. Brooke J. SUS: a retrospective. J Usability Stud 2013;8(2):29-40 [FREE Full text]

Abbreviations

HCS: HomeCare System
IRB: institutional review board
SMD: self-medication device
SUS: System Usability Scale

XSL•FO
RenderX

Edited by A Kushniruk; submitted 30.Oct.2025; peer-reviewed by SA Kristina, S Wang; comments to author 19.Mar.2026; revised version received 25.Mar.2026; accepted 27.Mar.2026; published 22.May.2026.

Please cite as:
Basapur S, Gellman C, Plenge JB, Troutman A, Yurko E, Woolsey B, McClendon J, Marceau L, Aggarwal NT
A Comprehensive Self-Medication Management System to Reduce Medication Errors Among People Living With Dementia: Mixed Methods Feasibility and Usability Study
JMIR Hum Factors 2026;13:e86828
URL: https://humanfactors.jmir.org/2026/1/e86828
doi:10.2196/86828
PMID:41921136

©Santosh Basapur, Charles Gellman, Jamie B Plenge, Amelia Troutman, Erin Yurko, Brandon Woolsey, Justin McClendon, Lisa Marceau, Neelum T Aggarwal. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 22.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring Factors Influencing Nursing Task Prioritization for Supportive Information System Design: Qualitative Study With Thematic Analysis

Kodai Iwamoto[1], MSc; Mayumi Toyama[2*], MD, MPH, PhD; Goshiro Yamamoto[3*], PhD; Chang Liu[4], PhD; Kazumasa Kishimoto[1,5], PhD; Yukiko Mori[1,5], MD, PhD; Tomohiro Kuroda[1,5], PhD

[1]Department of Informatics, Graduate School of Informatics, Kyoto University, 54 Shogoin-kawahara-cho, Sakyo-ku, Kyoto, Japan

[2]Department of Health Informatics, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[3]Preemptive Medicine and Lifestyle Related Disease Research Center, Kyoto University Hospital, Kyoto, Japan

[4]Department of Real World Data R&D, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[5]Division of Medical Information Technology and Administration Planning, Kyoto University Hospital, Kyoto, Japan

[*]these authors contributed equally

Corresponding Author:
Kodai Iwamoto, MSc
Department of Informatics, Graduate School of Informatics, Kyoto University, 54 Shogoin-kawahara-cho, Sakyo-ku, Kyoto, Japan

Abstract

Background: Nurses are required to perform multiple tasks concurrently, which leads to multitasking situations, and they have to continuously determine which tasks should be prioritized. This is particularly challenging for novice nurses. Although IT-based systems supporting prioritization have begun to emerge, research on the types of information required when prioritization is processed computationally is scant. Despite the clear need for a supportive information system to assist nursing task prioritization, such systems are not yet sufficiently developed.

Objective: This study aimed to explore the appropriate granularity and structure of information that should be provided to computational systems to support decision-making based on the influencing factors of nursing task prioritization.

Methods: Semistructured interviews were conducted with 10 nurses working in general wards to examine the factors they consider when determining task prioritization during clinical practice. Data were analyzed using an inductive, semantic approach based on a thematic analysis framework.

Results: Three themes and nine categories including (1) medical condition assessment factors (signs of acute physiological changes and indicators of clinical status and conditions), (2) patient-related nursing care factors (physical status, psychological condition, and personal characteristics; care needs during hospitalization; and treatment goals and care preferences), and (3) organizational and operational work factors (temporally structured tasks, requiring collaboration partners for task execution, environmental factors affecting task performance, and institutional- and ward-level policies) were identified.

Conclusions: Analysis of computational tractability of the identified factors indicated that medical condition assessment factors are relatively quantifiable. In contrast, patient-centered care and organizational and operational work factors rely on contextual and experiential judgment, limiting standardization and formalization. Regarding such ambiguous and context-dependent elements, flexible information-processing approaches, such as large language models, in addition to conventional rule-based methods, may be effective. Furthermore, the appropriate level of information granularity should be determined by the nature of the prioritization outputs required in actual nursing practice rather than the degree of abstraction itself.

(JMIR Hum Factors 2026;13:e89940) doi:10.2196/89940

KEYWORDS

nursing task prioritization; computational tractability; semistructured interview; thematic analysis; novice nurses

Introduction

Background

Clinical nurses care for multiple patients while concurrently performing a wide range of patient-specific nursing tasks. Consequently, multiple tasks frequently overlap within the same time frame, a phenomenon commonly defined as multitasking—the concurrent execution of 2 or more tasks [1,2]. Multitasking is considered among the most prevalent stressors in nurses' daily practice [3] and has been reported to occur in

XSL•FO
RenderX

approximately 26% - 50% of nursing activities [1,2,4]. Frequent multitasking is associated with task interruptions [5], which in turn increase cognitive and mental workload, leading to deviations from standard procedures, task errors, and increased risk of burnout [6-10].

Such challenges extend beyond interruptions, persisting throughout nursing care delivery. Particularly, when multiple tasks must be performed concurrently, nurses must determine the tasks that should be prioritized. In Japan, nursing students typically care for only one patient during clinical training, limiting their exposure to multitasking situations. Consequently, novice nurses often struggle to prioritize care after entering clinical practice [11-13]. Although prioritization challenges affect all nurses, they tend to be pronounced in early professional practice.

Nursing prioritization is the ordering of nursing problems by urgency and/or importance [12]. However, in clinical practice, prioritization is not determined by a single criterion; rather, it is influenced by multiple factors, including patient acuity and care intensity, availability of resources, organizational and personal values, nursing theories and models, and nurses' proficiency [12,14]. Additionally, prioritization is influenced by multiple mechanisms, including the Maslow effect (prioritization of life-sustaining and physical care), the outcome effect (immediate visibility of intervention outcomes), the time-cost effect (short or more predictable tasks), the collective adaptation effect (tasks emphasized within the team), and the audit effect (tasks confirmed by physicians or managers) [15]. These findings indicate that nursing prioritization is complex, shaped by multiple, overlapping cognitive and organizational factors rather than a single criterion.

Previous studies have introduced educational interventions using simulations and virtual reality to enhance prioritization skills [16-18]; however, their effectiveness has not been consistently demonstrated. Although real-time role modeling by educators or senior nurses is effective [19], such approaches rely heavily on individual interpretation and experiential learning, limiting their applicability as clinical decision support. In contrast, IT-based approaches, such as rule-based methods that identify relevant factors using logistic regression and construct decision trees for prioritization, have also been proposed [20]. While useful in specific applications such as staff scheduling [21], these approaches face significant challenges owing to implicit, context-dependent conditions that are difficult to formalize mathematically [22]. Even in relatively stable task environments, adapting rule-based methods to real-world clinical practice remains challenging. Although previous studies have attempted to identify and organize detailed nursing elements into rule-based frameworks, the diversity and context-dependence of practice hinder the representation of all relevant factors as explicit conditions. Moreover, only a few studies have systematically examined how factors influencing nursing prioritization can be structured for computational processing, including the level of information granularity and structure of representation that enable such processing [12,14,15,20]. Thus, despite the clear need for a supportive information system to assist nursing task prioritization in clinical practice, such systems remain insufficiently developed. This gap can be addressed by acquiring a clearer understanding of how nurses' prioritization decisions can be represented in a form suitable for computational handling.

Aim

This study aims to explore how nursing prioritization factors can be represented in terms of information granularity and structure when computational processing is assumed. To this end, thematic analysis was used to identify and organize the influencing factors of nurses' prioritization decisions. By elucidating the information structures suitable for computational handling, this study provides foundational insights for developing information systems that support nursing prioritization.

Methods

Study Overview

In this qualitative study, semi-structured interviews and a thematic analysis were conducted. This study was conducted in accordance with the Consolidated Criteria for Reporting Qualitative Research checklist.

Setting and Sampling

Participants were recruited between September and November 2024 from general wards of a national university hospital with a nurse-to-patient ratio of at least 1:7. As factors influencing nursing prioritization may differ by ward type and years of nursing experience, both novice and experienced nurses working in general wards were included. Novice nurses included those with 2 or fewer years of clinical experience, whereas experienced nurses included those with 5 or more years of experience and at least 1 year of service in their current ward. Both wards used the partnership nursing system as their nursing care delivery model. Potential participants were identified through the researchers' professional networks in collaboration with ward nurse managers, and purposive sampling was used. All candidates received an explanation of the study purpose and procedures, including audio recording; those who agreed to participate provided written informed consent.

Data Collection

The interviews were conducted once, face-to-face in Japanese by a single researcher (KI, registered nurse; male) with experience in qualitative research, and took place in the participants' wards. Individual, semistructured interviews were conducted, with primary questions focusing on factors influencing task prioritization in nursing practice; these questions were generated through a pilot study. Follow-up questions were asked as appropriate based on the participants' responses. The order and content of the questions were flexibly adjusted. Multimedia Appendix 1 provides an interview guide. Data collection was concluded when no new categories were generated. The interviewer and participants had previous professional contact through prior research activities. Field notes were taken during the interviews to supplement contextual understanding.

XSL•FO
RenderX

Data Analysis

This study aims to identify the factors influencing nursing task prioritization and elucidate context-dependent, multifaceted patterns of meaning. Therefore, an inductive and semantic approach to thematic analysis, as proposed by Braun and Clarke [23], was used to extract meaning from the data.

The analysis involved the following steps. First, the audio-recorded interviews were transcribed verbatim, and the transcripts were read repeatedly to achieve familiarization with the data. Subsequently, the researcher (KI) independently generated the initial codes inductively by focusing on meaningful units within the data. Similarities among the generated codes were examined, and themes and categories were identified. The identified themes and categories were reviewed collaboratively with another researcher (MT; female), and iterative refinement, including merging and differentiating themes, was conducted until consensus was reached. Finally, each theme and category was clearly defined and named to capture its essential meaning.

To enhance the credibility of the analysis, interpretations were discussed among researchers from different academic backgrounds. To reduce bias from the researcher's professional nursing background, ongoing reflexive discussions were held throughout the analytical process, and guidance was sought from an expert in thematic analysis. To further ensure credibility, transcripts were returned to the participants, and some participants confirmed that the content was accurate. Considering participant burden, feedback on the final analytical results was not requested. All analyses were conducted in Japanese, and the findings were translated into English. The researchers reviewed the translations to ensure accuracy. The data were analyzed using Microsoft Excel (version 2511).

Ethical Considerations

This study was approved by the Kyoto University Graduate School and Faculty of Medicine Ethics Committee (approval 4460). Participation in the study was voluntary. Recorded audio data were processed on devices with appropriate security measures, and participant privacy was protected through pseudonymization during the generation of analysis datasets. All participants were provided with a detailed explanation of the study, and written informed consent was obtained prior to participation. Participants were also informed of their right to withdraw consent at any time. No compensation was provided for participation in the study.

Results

Participant Characteristics

Table 1 presents details of the study participants. The participating wards included cardiology and thoracic surgery; 4 experienced and 6 novice nurses were recruited from the cardiology and thoracic surgery wards, respectively, resulting in 10 interviews. The mean interview length was 52.0 (SD 9.6) minutes. No participants declined to participate or withdrew from the study.

Table . Demographic profile of the participants.

Participants	Groups	Sex	Years of experience		Ward type
			Nursing	Ward	
1	Veteran	Male	9th year	5th year	Cardiology
2	Veteran	Male	15th year	3rd year	Cardiology
3	Veteran	Female	6th year	6th year	Cardiology
4	Veteran	Female	13th year	3rd year	Cardiology
5	Novice	Female	1st year	1st year	Thoracic surgery
6	Novice	Female	2nd year	1st year	Thoracic surgery
7	Novice	Female	2nd year	2nd year	Thoracic surgery
8	Novice	Female	2nd year	2nd year	Thoracic surgery
9	Novice	Female	2nd year	2nd year	Thoracic surgery
10	Novice	Female	2nd year	2nd year	Thoracic surgery

Overall Theme

Overview

Table 2 summarizes the factors influencing task prioritization by the general ward nurses. Three overarching themes including (1) medical condition assessment factors, (2) patient-related nursing care factors, and (3) organizational and operational work factors were identified. These themes comprised 9 categories.

XSL•FO
RenderX

Table . Themes, categories, and codes identified through thematic analysis.

Themes and categories	Codes
Theme A: medical condition assessment factors	
A-1: signs of acute physiological changes	Deterioration of hemodynamic status (eg, onset of arrhythmia and sudden decrease in blood pressure), worsening of respiratory status (eg, sudden decrease in SpO_2)
A-2: indicators of clinical status and conditions	Medical interventions (surgery, medication therapy, medical procedures, and diagnostic tests), laboratory data, vital signs, clinical findings, clinical symptoms, indwelling devices, oxygen therapy, medication therapy, fluid balance, progress notes, changes in condition during hospitalization
Theme B: patient-related nursing care factors	
B-1: physical status, psychological condition, and personal characteristics	Activities of daily living, presence of fall risk, physiological function (ability to expectorate), presence of dementia, presence of agitation, communication ability, personality traits
B-2: care needs during hospitalization	Illness stage, infection presence, intrahospital transfer, assistance with personal grooming, assistance with medication management, response to nurse call systems
B-3: treatment goals and care preferences	Discharge preparation, do not resuscitate orders, understanding of end-of-life preferences
Theme C: organizational and operational work factors	
C-1: temporally structured tasks	Time-specific tasks, non–time-specific tasks, unscheduled tasks, sequence-dependent tasks
C-2: required collaboration partners for task execution	Doctor, nurse, allied health professionals, family
C-3: environmental factors affecting task performance	Physicians' availability, nursing staff resources, temporal working conditions, task complexity, clinical procedures volume, number of assigned patients, location of assigned patients, presence of primary patients
C-4: institutional- and ward-level policies	Within the hospital, within the ward

Theme A: Medical Condition Assessment Factors

Overview

This theme represents factors that nurses consider when evaluating patient acuity and prioritizing tasks based on signs of acute physiological changes and clinical indicators. Nurses assessed and integrated information ranging from life-threatening changes to indicators requiring continuous monitoring to guide prioritization decisions. This theme comprises the following 2 categories.

A-1: Signs of Acute Physiological Changes

This category reflects situations in which nurses detect acute and marked changes in physiological indicators, such as the onset of arrhythmias, sudden decreases in blood pressure, or abrupt drops in oxygen saturation, and recognize these changes as signals of clinical deterioration requiring immediate attention. Such indicators strongly influence nurses' decisions to prioritize specific nursing tasks.

> *If an arrhythmia appears on the ECG or the oxygen saturation drops, I think I would prioritize going to see the patient.* [Participant 1, veteran]

> *If there's a significant drop in blood pressure, I feel that it takes priority over scheduled tasks.* [Participant 6, novice]

A-2: Indicators of Clinical Status and Conditions

This category encompasses information reflecting the patient's current condition, including laboratory data, vital signs, and clinical findings and symptoms, as well as factors related to ongoing medical management, such as medical interventions, indwelling devices, medication therapy, and oxygen therapy. The findings showed that nurses comprehensively integrate these diverse indicators to assess patient stability and care needs and to adjust task prioritization accordingly.

> *Patients who have just undergone surgery, those on the first postoperative day, those on mechanical ventilation, those receiving oxygen therapy, or those with unstable vital signs—I tend to prioritize seeing those patients first.* [Participant 8, novice]

> *Patients with higher severity often have multiple infusions, drains, or chest tubes in place, and sometimes temporary pacing leads as well. I assess those factors because their circulatory status can become quite unstable.* [Participant 2, veteran]

Theme B: Patient-Related Nursing Care Factors

Overview

This theme reflects how nurses prioritize nursing tasks based on patient-specific factors including physical status, psychological condition, and personal characteristics; care needs during hospitalization; and treatment goals and care preferences.

XSL•FO
RenderX

Nurses consider not only patients' medical conditions but also their functional status, psychological state, and understanding of and attitudes toward treatment when prioritizing tasks according to the clinical situation. This theme comprises the following 3 categories.

B-1: Physical Status, Psychological Condition, and Personal Characteristics

This category reflects evaluations based on individual patient characteristics, including physical and physiological functions, such as activities of daily living and the ability to expectorate, as well as psychological characteristics, such as cognitive function, presence of agitation, communication ability, and personality traits. The findings showed that nurses adjusted task prioritization and care approach according to the necessity of care, intervention complexity, and monitoring requirements based on these characteristics.

> *About patients with dementia, there is often a greater need for assistance with ADLs, and in terms of building trust or communication, I tend to go to them first. If I am late, even patients with dementia can feel anxious, stressed, or irritated, so I try to attend to them early.* [Participant 4, veteran]

> *Because they cannot express themselves verbally, I think they should be prioritized in case something urgent happens.* [Participant 6, novice]

B-2: Care Needs During Hospitalization

This category indicates that prioritization was determined based on care needs arising from patients' medical conditions and daily living during hospitalization. Nurses considered clinical assessments such as illness stage and presence of infection, alongside patients' expressed needs related to daily living support, including hospital transfer, personal grooming, and medication management. Furthermore, responses to nurse call systems were prioritized, as they were recognized as directly influencing patients' sense of security and the establishment of trust (Textbox 1).

Textbox 1. Representative quotes related to Category B-2

"In the hyperacute phase, that is the highest priority. For acute and terminal phases, it may depend on the individual's condition. Even among terminal patients, there are those whose condition is relatively stable and those who are much closer to death, and in those cases, they may be prioritized at the same level as hyperacute patients. Patients in a stable phase would come later" [Participant 4, veteran].

"When patients ask to go shopping or to be taken to the post office, I try not to postpone it too much. It's also about the relationship—if you make them wait too long, they become anxious, so I try to be mindful not to keep them waiting for too long" [Participant 3, veteran].

"For nurse call responses, I basically go as soon as it rings, unless there is a very compelling reason not to. In principle, I will respond immediately" [Participant 9, novice].

B-3: Treatment Goals and Care Preferences

This category indicates that factors related to patients' values and decision-making, such as their treatment goals and care preferences, influenced prioritization. Information related to treatment direction, including discharge preparation, do-not-resuscitate orders, and understanding of end-of-life preferences, was shown to be a key factor influencing nurses' decisions regarding care timing and approach (Textbox 2).

Textbox 2. Representative quotes related to Category B-3

"I think we start gathering information early on about how the patient wants to live after discharge, but we also check how they feel at that moment, given their current condition" [Participant 1, veteran].

"For terminal patients, it depends on the severity. If it's really down to just a few days or even that day, then the priority becomes very high. But it also changes depending on whether a DNR has been established or not" [Participant 6, novice].

"For patients receiving best supportive care, if they express how they want to proceed from here, I try to listen to that, and if I can make time during the day, I go to see them" [Participant 7, novice].

Theme C: Organizational and Operational Work Factors

Overview

This theme focuses on organizational and operational conditions under which tasks are performed, highlighting structural and environmental factors influencing planning, execution, and prioritization. Beyond clinical assessments and patient-related factors, nurses adjusted priorities realistically and flexibly by considering temporally structured tasks, required collaboration partners for task execution, environmental factors affecting task performance, and institutional- and ward-level policies. This theme comprises the following 4 categories.

C-1: Temporally Structured Tasks

This category reflects task characteristics with temporal constraints, including tasks with predefined timing or sequence of execution, and those requiring situational adjustment. Nurses differentiated tasks that must be performed at specific times, such as medication administration, flexibly timed tasks, and unscheduled tasks. They prioritized tasks based on overall workflow, while also planning tasks that required a specific execution sequence according to their impact on subsequent activities (Textbox 3).

XSL•FO
RenderX

Textbox 3. Representative quotes related to Category C-1

"I make sure to go at 9 o'clock for patients who have scheduled oral medications, since the administration time is fixed" [Participant 5, novice].

"For tasks that are already scheduled, such as hygiene care or wound care, I adjust my schedule so that they can be completed in the morning" [Participant 3, veteran].

"I often have to hurry to make sure blood glucose measurements are done in time before meals" [Participant 10, novice].

C-2: Required Collaboration Partners for Task Execution

This category reflects the individuals with whom nurses coordinate care as part of their duties, including doctors, other nurses, allied health professionals, and families involved in clinical decision-making or task execution.

> *For procedures, when a physician has a fixed schedule—such as outpatient clinics or surgeries—and there is no flexibility around that timing, we adjust other tasks so that we can carry them out accordingly.* [Participant 4, veteran]
>
> *For admissions, we often do them together as a pair, because there are records to complete and one person may need to explain things while the other documents, so admissions are often handled by two nurses.* [Participant 4, veteran or novice]

C-3: Environmental Factors Affecting Task Performance

This category indicates that the human, physical, and temporal environments surrounding nurses influence task execution and prioritization. Nurses adjusted their work based on human resource constraints such as physician availability, nursing staff resources, and shift schedules, and environmental factors, including clinical procedure volume, task complexity, and patient numbers and locations. These factors resulted in variations in task prioritization depending on the day of the week or time of day. In situations where staffing capacity allowed greater flexibility, tasks that were usually deprioritized were sometimes addressed.

> *On weekends, since there are fewer examinations, the ward tends to be calmer, so we try to focus on things like giving bed baths to patients who require full assistance or shampooing patients who haven't been able to get it done.* [Participant 6, novice]
>
> *When I have fewer assigned patients and a new admission comes in, I try to use that time to provide orientation for the patient, such as explaining the ward routines.* [Participant 10, novice]

C-4: Institutional- and Ward-Level Policies

This category indicates that institutional- and unit-level policies and operational rules influence how nursing tasks are performed, including the timing of care delivery and decisions regarding task prioritization.

> *Visiting hours start at 2:00 p.m., and since that is a hospital-wide rule, we provide guidance or education to family members after they arrive.* [Participant 1, veteran]
>
> *For medications such as immunosuppressants, blood concentration levels are important, so we make sure they are administered at the designated time. We try to administer them as early as possible and confirm adherence—this is something like a ward-level rule.* [Participant 8, novice]

Discussion

Principal Findings

Overview

This study identifies 3 themes and 9 categories influencing nurses' task prioritization in general wards. Theme A represents medical urgency and acuity, theme B reflects patient individuality and values, and theme C encompasses organizational and operational constraints. Consistent with previous studies, the findings suggest that nurses' prioritization is a dynamic process integrating multiple elements according to the situational context [12,14,15]. Regarding differences in information structure across themes, the elements in theme A are relatively amenable to simple classification (eg, binary or categorical) or quantification, whereas aspects of themes B and C rely on situational interpretation and experiential judgment, complicating uniform representation.

Theme A: Medical Condition Assessment Factors

The elements in theme A were largely consistent with previously reported criteria for medical urgency and acuity. Regarding A-1, previous studies have shown signs of acute physiological changes such as dyspnea, chest pain, airway obstruction, stroke, and massive hemorrhage to require immediate, high-priority intervention [14,24-26]. Similarly, the indicators of clinical status and conditions in A-2 were aligned with those reported in earlier research, including medical interventions, laboratory data, vital signs, oxygen therapy, medication therapy, indwelling devices, clinical symptoms, and progress notes [14,15,27]. In contrast, elements such as clinical findings and fluid balance were uniquely identified. These factors may previously have been subsumed under broader concepts such as "physiological needs" or "patient assessment," and therefore not explicitly defined as discrete elements [15,27]. Particularly, fluid balance appears to be a context-dependent factor influenced by the specific clinical setting of a cardiology ward. This finding highlights a key challenge: as elements are identified in greater detail, context-specific judgment criteria increase, potentially complicating comprehensive organization and standardization.

Theme B: Patient-Related Nursing Care Factors

In theme B, judgment elements based on patient individuality and values were identified. Regarding B-1, physical status, psychological condition, and personal characteristics such as activities of daily living, dementia, agitation, and personality

XSL•FO
RenderX

traits have also been reported in previous studies [14,25,28,29]. While some studies emphasize that psychological support should be prioritized [14], others highlight that such care is often left unfinished because of time constraints and difficulties in evaluating outcomes [15]. Although these factors can guide prioritization, their weighting likely depends heavily on individual nurses' values and clinical experience [12]. Furthermore, previous research shows that early-career nurses tend to emphasize psychosocial needs, whereas experienced nurses tend to deprioritize them or consider them the responsibility of other professions [28]. Physiological functions such as the ability to expectorate (ie, clear sputum from the airway) and communication were newly identified in this study and are likely influenced by the thoracic surgery ward study setting. Regarding B-2, care needs during hospitalization, such as the presence of infection and responses to nurse call requests, have been reported in previous studies [15,24]. In contrast, factors including illness stage, intrahospital transfers, assistance with personal grooming, and medication management were newly identified in this study. These elements reflect patient-specific needs and, compared with the medical factors in theme A, may be less likely to become explicit in the prioritization decision-making process. Regarding B-3, treatment goals and care preferences, previous studies show that discharge support and planning are often incomplete [15]. In contrast, this study identifies discharge preparation as a prioritized element. This difference is attributable to the target hospital, where the partnership nursing system has been implemented, suggesting that a care delivery model with relatively sufficient human resources may influence nurses' prioritization decisions. These findings primarily reflect what should be prioritized based on patient-specific considerations.

Theme C: Organizational and Operational Work Factors

In theme C, judgment elements based on organizational and operational constraints were identified. In contrast to theme B, these factors primarily constrain what can be feasibly performed in practice. Regarding C-1, temporally structured tasks were consistent with previous studies, showing that, during physical emergencies, routine tasks are interrupted to provide immediate care [28], corresponding to "unscheduled tasks" in this study. In contrast, few previous studies have clearly distinguished between time-specified, non–time-specified, and sequence-dependent tasks based on time constraints or execution order. Although this represents a highly detailed classification, it reflects the level of granularity required for computational processing. The elements related to C-2, which required collaborators for task execution (doctors, nurses, allied health professionals, and family members), were newly identified in this study. Although these elements may not directly determine task priority, they represent key influencing conditions for task feasibility. Regarding C-3, environmental factors affecting task performance, such as human resources, task difficulty, procedure volume, and number of assigned patients, have also been reported in previous studies [12,14,27-30]. In contrast, factors including physician availability, day of the week, location of assigned patients, and whether the patient was assigned as a primary patient were specific to this study. Differences in the implementation of examinations and rehabilitation depending on the day of the week may influence the structure of prioritization, suggesting that prioritization depends on institutional characteristics. Regarding C-4, institutional- and ward-level policies, previous studies show that ward philosophy and the nursing theories or models adopted influence prioritization decisions [12,15,24,27]. Given that nursing practices are performed within specific environments, context-dependent judgment criteria were identified. The elements identified in C-2, C-3, and C-4 are suggested to serve both mediating and moderating roles in task prioritization. Specifically, they mediate the formation of prioritization judgments while also moderating the strength and direction of these effects. For example, nursing practice is not uniform; the priority assigned to the same task varies across settings, such as hospitals and wards, long-term care facilities, and home care. Furthermore, the range of feasible tasks is constrained by factors such as interprofessional collaboration and staffing conditions, which in turn alter the scope of tasks that can be prioritized. Taken together, these findings suggest that the elements identified in themes B and C can be understood as different dimensions of contextual judgment, encompassing both patient-specific and organizational factors that shape prioritization decisions.

Building on these functional roles, the following discussion further conceptualizes organizational context as a broader mechanism underlying task prioritization. Although this study does not address the attributes of the requester, previous research shows that in nonhospital settings, requests from managerial staff are more likely to be prioritized [12,28]. In contrast, in general hospital wards, nurses primarily assume leadership in task execution. This contrast suggests that the influence of requester attributes is contingent on organizational context, particularly differences in roles, authority structures, and chains of command. These findings suggest that ward characteristics function as a foundational mechanism structuring task prioritization processes. Specifically, such characteristics encompass not only explicit operational rules, such as visiting hours, but also the configuration of stakeholders and their power dynamics. Through their interaction, these factors shape nurses' subjective judgment in practice. Furthermore, as previous studies have shown, nurses tend to prioritize physicians' expectations due to the nature of their roles, indicating that task prioritization is conducted with close attention to established chains of command [26]. Additionally, previous studies indicate that nurses' subjective assessments—such as the time required to complete a task and the predictability of intervention outcomes—as well as value-based considerations, including the prioritization of physician expectations and tasks that emphasize human dignity, also influence prioritization decisions [15,26]. Taken together, these findings suggest that task prioritization is shaped by a complex interplay of organizational, contextual, and individual factors that are difficult to fully formalize, including nursing theories and judgments based on individual values and professional experience.

Characteristics of the Extracted Factors and Their Computational Tractability

This study explores which information structures are amenable to computational processing, based on the elements identified

XSL•FO
RenderX

through thematic analysis, and conceptualizes mechanisms for supporting nursing task prioritization.

Regarding information processing approaches for prioritization, both rule-based methods, such as integer linear programming, and non–rule-based approaches, such as large language models (LLMs), can be considered. ILP enables the derivation of optimal solutions through the explicit formulation of objective functions and constraints; however, it is limited in adequately representing the ambiguous and context-dependent elements inherent in real-world clinical practice. Additionally, its high computational cost and the need for recalculation in dynamically changing environments pose practical challenges for real-world implementation. In contrast, LLMs can process information that includes implicit judgments and ambiguous expressions that are not explicitly defined by rules. This characteristic suggests their potential usefulness in handling the tacit knowledge and context-sensitive decision-making in nursing practice. Nevertheless, LLMs present challenges, including the opacity of their reasoning processes and concerns regarding explainability, safety, and reliability when applied to medical information.

In routine clinical practice, nurses' decision-making is typically grounded in higher-level, abstract categorical understanding, while drawing on more concrete, code-level information. In contrast, from the perspective of computational information processing, even code-level data can be handled in a formalized manner. However, the codes identified in this study do not fully encompass all factors influencing nurses' decisions. Particularly for highly context-dependent elements, explicit structuring remains a challenge. Specifically, the elements within theme A (medical condition assessment factors) are relatively capable of simple classification (eg, binary or categorical) or quantification, whereas aspects of theme B (patient-related nursing care factors) and theme C (organizational and operational work factors) require situational interpretation and experiential judgment. Consequently, these elements cannot be easily represented within a uniform information structure. This structural heterogeneity highlights the difficulty of defining a single, standardized level of granularity or representation for information used in prioritization. Together, the key challenges identified in this study can be summarized as how to address the long-tail structure of decision-making factors and how to determine the appropriate level of granularity for information used in prioritization. As a complementary approach to these challenges, integrating category-level information through retrieval-augmented generation using LLMs may enable the partial incorporation of contextual variability and judgmental nuances that are difficult to capture using conventional rule-based or mathematical optimization methods. Consequently, while defining information at a highly detailed level may limit the comprehensiveness of captured judgment factors, excessive abstraction may render the information too vague for LLMs to interpret in a manner consistent with the intended decision-making process. Therefore, the granularity of information should be determined by the nature of the prioritization outputs required in clinical practice rather than its level of detail.

The focus of this study—prioritization of nursing tasks—can be situated within the framework proposed by Barrett and Jones [31], which conceptualizes 3 types of nurse-artificial intelligence interaction: task augmentation, task automation, and hybrid. Specifically, this study is positioned as an attempt toward task automation [31]. We posit that achieving full automation requires that all elements involved in prioritization decisions be represented in a structured and computable form. However, our findings indicate that the factors underlying task prioritization can be broadly categorized into those that are amenable to quantification and structuring and those that rely on unstructured information such as natural language. This distinction is consistent with the framework proposed by Autor et al [32], in which routine tasks are more susceptible to substitution by computer technologies, whereas nonroutine tasks tend to be complemented rather than replaced. In this context, elements that can be structured may be processed using rule-based approaches, whereas elements that resist structuring may require interpretation through generative artificial intelligence or may be difficult to incorporate into computational processes altogether. These unstructured elements are characterized by their reliance on context-dependent judgment, and their structuring becomes even more challenging when incorporating not only clinical information but also patient preferences and needs. The difficulty of structuring such elements reflects the inherent ambiguity of task characteristics in the nursing domain, suggesting that it is not feasible to fully structure all relevant information. Therefore, the findings of this study highlight the importance of designing and implementing technologies under the assumption that complete structuring is unattainable. Similar to the indispensable role of free-text fields in electronic health records, our results underscore the need for system designs that accommodate both structured and unstructured information.

Limitations

This study has several limitations. First, the codes identified as influencing prioritization may not comprehensively capture all the elements that affect nurses' decision-making. Second, the relative influence of each factor was not quantitatively evaluated; therefore, the relative importance and potential interactions among factors remain unclear. Third, the findings were derived from a specific context and population, and the results may not be generalizable to settings with different ward characteristics, organizational structures, or nursing systems.

Future Directions

Overview

The system envisioned in this study is designed to support the prioritization of nursing tasks and to partially automate the associated information-gathering processes. Such a system may be implemented via a smartphone-based interface that presents the next task to be performed and allows nurses to record task completion through simple interactions. It may be used both before the start of a shift to provide an overview of the day's workload and continuously during the shift to accommodate changes in clinical conditions. The future research required for the development and implementation of this system is outlined below.

XSL•FO
RenderX

Technical Validation of Prioritization Criteria

To translate the findings into rule-based information processing approaches such as ILP, further examination regarding which codes should be incorporated into prioritization systems, the appropriate level of granularity, and how the weighting of each code should be determined is required. Additionally, it is necessary to investigate which levels of abstraction (eg, code or category) and what types of information are most effective when presenting prioritized recommendations to nurses in clinical practice.

Design of Supportive Information System

Previous research shows that nurses tend to undertake nonnursing tasks and that such tendencies are associated with an increased likelihood of nursing tasks being left undone (NTLU) [33]. Moreover, experiences of NTLU have been reported to be strongly associated with job dissatisfaction than performing nonnursing tasks [33]. Importantly, NTLU often involves intrinsically meaningful activities, such as comfort or talk with patients, which are prone to deprioritization under time pressure [28,33]. This phenomenon may be partly explained by the mere-urgency effect, whereby urgent tasks are prioritized over those that are more important but less time-critical, thereby hindering appropriate prioritization [34].

Accordingly, the design of support information systems should not simply aim to mechanistically determine or present task priorities. As suggested by prior research proposing a 7-stage normative framework for prioritizing nursing activities, they should support nurses in making informed decisions while preserving professional values and the core aspects of nursing practice [35]. In other words, instead of immediately reprioritizing nonurgent interruptions, decision support systems should facilitate judgments that consider job satisfaction, nursing care quality, and sustainability of "nursing as a profession."

Conclusions

Although there is a clear need for a supportive information system for nursing task prioritization, such systems remain insufficiently developed. Accordingly, this study explored how factors influencing nursing prioritization can be represented in terms of information granularity and structure.

Thematic analysis was used to identify and organize the factors guiding nurses' prioritization decisions. The results indicate that nurses' prioritization constitutes a multilayered decision-making process comprising three qualitatively distinct themes: (1) objectively observable elements such as medical urgency and acuity, (2) context-dependent factors grounded in patient individuality and values, and (3) environmental factors including time constraints, staffing resources, and organizational policies.

Theme A (medical condition assessment factors) was relatively amenable to quantification. In contrast, themes B (patient-related nursing care factors) and C (organizational and operational work factors) required context-sensitive and experiential judgment. While these elements enable flexible, patient- and institution-specific decision-making, they complicate standardization and formalization. Accordingly, for such ambiguous and context-dependent elements, flexible information-processing approaches, such as LLMs, may be effective in complementing conventional rule-based methods. Furthermore, the appropriate level of information granularity should be determined by the nature of the prioritization outputs required in actual nursing practice, rather than by the degree of abstraction alone.

Funding

This work was supported by the Japan Science and Technology Agency Support for Pioneering Research Initiated by the Next Generation (grant JPMJSP2110) and the next Crossministerial Strategic Innovation Promotion Program on "Integrated Health Care System" (grant JPJ012425).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to ethical and privacy restrictions, as they contain interview data that may include personally identifiable information, but are available from the corresponding author on reasonable request.

Authors' Contributions

KI conceptualized and designed the study, developed the interview guide, conducted the interviews, performed the primary data analysis, and drafted the manuscript. MT provided methodological guidance and contributed to data analysis. KK, CL, GY, and YM contributed to study conceptualization and design, data interpretation, and critical revision of the manuscript. TK supervised the study. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File, 20 KB - humanfactors_v13i1e89940_app1.docx]

XSL•FO
RenderX

References

1. Yen PY, Kellye M, Lopetegui M, et al. Nurses' time allocation and multitasking of nursing activities: a time motion study. AMIA Annu Symp Proc 2018;2018:1137-1146. [Medline: 30815156]
2. Yen PY, Kelley M, Lopetegui M, et al. Understanding and visualizing multitasking and task switching activities: a time motion study to capture nursing workflow. AMIA Annu Symp Proc 2016;2016:1264-1273. [Medline: 28269924]
3. Zhang S, Xiao X, Ai Y, et al. Challenges and coping experiences faced by nursing staff in long-term care facilities in China: a qualitative meta-analysis. Front Public Health 2024;11:1302481. [doi: 10.3389/fpubh.2023.1302481]
4. Kim Y, Lee MJ, Choi M, Cho E, Ryu GW. Exploring nurses' multitasking in clinical settings using a multimethod study. Sci Rep 2023;13(1):5704. [doi: 10.1038/s41598-023-32350-9]
5. Westbrook JI, Duffield C, Li L, Creswick NJ. How much time do nurses have for patients? A longitudinal study quantifying hospital nurses' patterns of task time distribution and interactions with health professionals. BMC Health Serv Res 2011 Nov 24;11(1):319. [doi: 10.1186/1472-6963-11-319] [Medline: 22111656]
6. Westbrook JI, Woods A, Rob MI, Dunsmuir WTM, Day RO. Association of interruptions with an increased risk and severity of medication administration errors. Arch Intern Med 2010 Apr 26;170(8):683-690. [doi: 10.1001/archinternmed.2010.65] [Medline: 20421552]
7. Thomas L, Donohue-Porter P, Stein Fishbein J. Impact of interruptions, distractions, and cognitive load on procedure failures and medication administration errors. J Nurs Care Qual 2017;32(4):309-317. [doi: 10.1097/NCQ.0000000000000256] [Medline: 28448299]
8. Shan Y, Shang J, Yan Y, Ye X. Workflow interruption and nurses' mental workload in electronic health record tasks: an observational study. BMC Nurs 2023 Mar 9;22(1):63. [doi: 10.1186/s12912-023-01209-9] [Medline: 36890555]
9. Diehl E, Rieger S, Letzel S, et al. The relationship between workload and burnout among nurses: the buffering role of personal, social and organisational resources. PLoS One 2021;16(1):e0245798. [doi: 10.1371/journal.pone.0245798] [Medline: 33481918]
10. Pérez-Francisco DH, Duarte-Clíments G, Del Rosario-Melián JM, Gómez-Salgado J, Romero-Martín M, Sánchez-Gómez MB. Influence of workload on primary care nurses' health and burnout, patients' safety, and quality of care: integrative review. Healthcare (Basel) 2020 Jan 3;8(1):12. [doi: 10.3390/healthcare8010012] [Medline: 31947761]
11. Hung SYM, Lam KKS, Wong LM. The transition challenges faced by new graduate nurses in their first year of professional experience. JNHC 2018;5(1). [doi: 10.5176/2345-7198_5.1.19]
12. Hendry C, Walker A. Priority setting in clinical nursing practice: literature review. J Adv Nurs 2004 Aug;47(4):427-436. [doi: 10.1111/j.1365-2648.2004.03120.x] [Medline: 15271162]
13. Gerrish K. Still fumbling along? A comparative study of the newly qualified nurse's perception of the transition from student to qualified nurse. J Adv Nurs 2000 Aug;32(2):473-480. [doi: 10.1046/j.1365-2648.2000.01498.x]
14. Open RN, Ernstmeyer K, Christman E. Prioritization. In: Nursing Management and Professional Concepts: Chippewa Valley Technical College; 2022. URL: https://www.ncbi.nlm.nih.gov/books/NBK598386 [accessed 2026-05-05]
15. Jones TL. What nurses do during time scarcity—and why. J Nurs Adm 2016 Sep;46(9):449-454. [doi: 10.1097/NNA.0000000000000374] [Medline: 27556653]
16. Franklin AE, Thielke L, Gilbert GE, Waller M. TIDES: examining the influence of temporal individual differences on multitasking in educational simulation. Adv Simul (Lond) 2020 Nov 9;5(1):31. [doi: 10.1186/s41077-020-00144-y] [Medline: 33292769]
17. Kaplan B, Ura D. Use of multiple patient simulators to enhance prioritizing and delegating skills for senior nursing students. J Nurs Educ 2010 Jul;49(7):371-377. [doi: 10.3928/01484834-20100331-07] [Medline: 20411859]
18. Chan K, Kor PPK, Liu JYW, Cheung K, Lai T, Kwan RYC. The use of immersive virtual reality training for developing nontechnical skills among nursing students: multimethods study. Asian Pac Isl Nurs J 2024 Jul 10;8:e58818. [doi: 10.2196/58818] [Medline: 38986130]
19. Wang EE, Dyne PL, Du H. Systems-based practice: summary of the 2010 Council of Emergency Medicine Residency Directors Academic Assembly Consensus Workgroup—teaching and evaluating the difficult-to-teach competencies. Acad Emerg Med 2011 Oct;18(Suppl 2):S110-S120. [doi: 10.1111/j.1553-2712.2011.01160.x] [Medline: 21999553]
20. Hirdes JP, Poss JW, Curtin-Telegdi N. The Method for Assigning Priority Levels (MAPLe): a new decision-support system for allocating home care resources. BMC Med 2008 Mar 26;6(1):9. [doi: 10.1186/1741-7015-6-9] [Medline: 18366782]
21. Yilmaz E. A mathematical programming model for scheduling of nurses' labor shifts. J Med Syst 2012 Apr;36(2):491-496. [doi: 10.1007/s10916-010-9494-z] [Medline: 20703701]
22. Drake RG. The nurse rostering problem: from operational research to organizational reality? J Adv Nurs 2014 Apr;70(4):800-810. [doi: 10.1111/jan.12238] [Medline: 23991714]
23. Braun V, Clarke V. Successful Qualitative Research: A Practical Guide for Beginners: SAGE Publications; 2013.
24. Palese A, Bottega M, Cescutti A, et al. Depicting clinical nurses' priority perspectives leading to unfinished nursing care: a pilot Q methodology study. J Nurs Manag 2020 Nov;28(8):2146-2156. [doi: 10.1111/jonm.13036] [Medline: 32335959]
25. Ludlow K, Churruca K, Ellis LA, Mumford V, Braithwaite J. Decisions and dilemmas: the context of prioritization dilemmas and influences on staff members' prioritization decisions in residential aged care. Qual Health Res 2021 Jun;31(7):1306-1318. [doi: 10.1177/1049732321998294] [Medline: 33739185]

XSL•FO
RenderX

26. Bassi E, Jones TL, Chiappinotto S, Tartaglini D, Dal Molin A, Palese A. Educating nurses on prioritization: insights from a mixed-methods study on unfinished nursing care. BMC Nurs 2025 Jul 1;24(1):759. [doi: 10.1186/s12912-025-03345-w] [Medline: 40598238]
27. Kalisch BJ, Xie B. Errors of omission: missed nursing care. West J Nurs Res 2014 Aug;36(7):875-890. [doi: 10.1177/0193945914531859] [Medline: 24782432]
28. Bowers BJ, Lauring C, Jacobson N. How nurses manage time and work in long-term care. J Adv Nurs 2001 Feb;33(4):484-491. [doi: 10.1046/j.1365-2648.2001.01686.x] [Medline: 11251736]
29. Topaz M, Trifilio M, Maloney D, Bar-Bachar O, Bowles KH. Improving patient prioritization during hospital-homecare transition: a pilot study of a clinical decision support tool. Res Nurs Health 2018 Oct;41(5):440-447. [doi: 10.1002/nur.21907] [Medline: 30203417]
30. Slettebø A, Kirkevold M, Andersen B, et al. Clinical prioritizations and contextual constraints in nursing homes—a qualitative study. Scand J Caring Sci 2010 Sep;24(3):533-540. [doi: 10.1111/j.1471-6712.2009.00745.x] [Medline: 20409056]
31. Barrett JJ, Jones CB. Task augmentation, automation, and hybridization in nursing: a conceptual framework for artificial intelligence-integrated care delivery. Nurs Outlook 2025;73(5):102524. [doi: 10.1016/j.outlook.2025.102524] [Medline: 40780090]
32. Autor DH, Levy F, Murnane RJ. The skill content of recent technological change: an empirical exploration. Q J Econ 2003 Nov 1;118(4):1279-1333. [doi: 10.1162/003355303322552801]
33. Bekker M, Coetzee SK, Klopper HC, Ellis SM. Non-nursing tasks, nursing tasks left undone and job satisfaction among professional nurses in South African hospitals. J Nurs Manag 2015 Nov;23(8):1115-1125. [doi: 10.1111/jonm.12261] [Medline: 25345386]
34. Zhu M, Yang Y, Hsee CK. The mere urgency effect. J Consum Res 2018;45(3):673-690. [doi: 10.1093/jcr/ucy008]
35. Patterson ES, Ebright PR, Saleem JJ. Investigating stacking: how do registered nurses prioritize their activities in real-time? Int J Ind Ergon 2011 Jul;41(4):389-393. [doi: 10.1016/j.ergon.2011.01.012]

Abbreviations

ILP: integer linear programming
LLM: large language model
NTLU: nursing task being left undone

Edited by A Kushniruk; submitted 05.Jan.2026; peer-reviewed by E Willis, J Barrett; revised version received 06.Apr.2026; accepted 08.Apr.2026; published 05.Jun.2026.

Please cite as:
Iwamoto K, Toyama M, Yamamoto G, Liu C, Kishimoto K, Mori Y, Kuroda T
Exploring Factors Influencing Nursing Task Prioritization for Supportive Information System Design: Qualitative Study With Thematic Analysis
JMIR Hum Factors 2026;13:e89940
URL: https://humanfactors.jmir.org/2026/1/e89940
doi:10.2196/89940

© Kodai Iwamoto, Mayumi Toyama, Goshiro Yamamoto, Chang Liu, Kazumasa Kishimoto, Yukiko Mori, Tomohiro Kuroda. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 5.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Health Care Professionals' Experience in a Digitally Supported Hospital-at-Home Program: Cross-Sectional Study

Carlos Bibiano Guillén, MD; Eva Maria Arroyo Urea, PhD; Olga Estremera López, MD; Sonia López Lallave, MD; María Consolación Aranda Higuero, NP; Cristina Ubeda García, MS; Silvia Sánchez Sierra, NP; Almudena Calvo Cruz, NP; María Mir Montero, MD

Department of Emergency Medicine, Short Stay Unit and Hospital at Home Unit, Hospital Universitario Infanta Leonor, Av. Gran Vía del Este 80, Madrid, Spain

Corresponding Author:
María Mir Montero, MD
Department of Emergency Medicine, Short Stay Unit and Hospital at Home Unit, Hospital Universitario Infanta Leonor, Av. Gran Vía del Este 80, Madrid, Spain

Abstract

Background: Hospital-at-home (HaH) programs supported by telemedicine and remote monitoring technologies are increasingly being adopted as alternatives to conventional hospitalization. Although clinical outcomes and patient satisfaction have been widely studied, less is known about health care professionals' experiences in established, digitally supported HaH models integrated into routine clinical practice.

Objective: This study aimed to assess health care professionals' perceptions of an established, digitally supported HaH program, focusing on usability, workload, organizational integration, perceived clinical impact, and overall satisfaction.

Methods: We conducted a single-center cross-sectional study in the HaH unit of Infanta Leonor University Hospital (Madrid, Spain) between January and May 2025. All health care professionals assigned to the unit were invited to participate in an anonymous electronic survey composed of 14 Likert-scale items and 1 open-ended question. The items were designed to assess understanding of the program, perceived clinical impact, interaction with technology, workload, and overall evaluation. Nonparametric tests included Mann-Whitney *U* and Kruskal-Wallis tests. *P* values from each set of 14 item-level tests were adjusted separately using the Benjamini-Hochberg procedure.

Results: Of 47 invited professionals, 30 (64%) participated, including physicians (n=13, 43%), nurses (n=10, 33%), and nursing assistants (n=7, 23%). Overall perceptions of the digitally supported HaH model were generally favorable, particularly regarding perceived quality of care, perceived promotion of patient autonomy, and overall satisfaction. The largest exploratory differences across professional roles appeared in items related to the perceived contribution of technology to quality of care (ε^2=0.46, 95% CI 0.17-0.85), recommendation of the model (ε^2=0.45, 95% CI 0.14-0.85), and perceived promotion of patient autonomy (ε^2=0.41, 95% CI 0.15-0.78). No significant differences were identified according to years of professional experience. Optional comments were also provided by 20% (6/30) of participants and were reviewed descriptively, offering contextual examples related to connectivity, platform functionality, and workflow challenges.

Conclusions: Health care professionals generally perceived the digitally supported HaH model positively. Exploratory differences between professional groups varied by item: technology-related perceptions were less favorable among responding nurses, whereas perceived workload was higher among physicians. These findings should be interpreted cautiously, given the small subgroup sizes and the low nurse response rate. Long-term sustainability of HaH programs may benefit from role-specific organizational strategies, targeted training, and better workflow integration.

(JMIR Hum Factors 2026;13:e106615) doi:10.2196/106615

KEYWORDS

hospital-at-home; telemedicine; remote patient monitoring; technology-enabled care; human factors; usability; health care professionals; workload; workflow integration; care delivery models; organizational sustainability; digital health; virtual health

Introduction

Hospital-at-home (HaH) care has become established in recent years as a service model that delivers hospital-level care in patients' homes, reduces complications associated with conventional hospitalization, and eases pressure on hospital capacity [1-3]. Population aging, the rise in chronic diseases, and the need for more sustainable care models have driven the international expansion of these programs [4,5].

XSL•FO
RenderX

In parallel, the development of digital tools has profoundly transformed the organization of hospital-based home care. The incorporation of telemedicine platforms, video consultations, remote monitoring of vital signs, and digital communication systems has expanded the capacity for clinical follow-up beyond the traditional hospital setting [6,7]. These technologies may help clinicians identify patient deterioration earlier; enhance continuity of care; and support a more flexible, patient-centered approach to service delivery [8].

However, introducing digital technologies into HaH programs changes not only how care is delivered but also how professionals communicate, organize their work, and interact with patients [9,10]. Adapting to these digital environments can generate organizational benefits, but it may also introduce new cognitive demands, additional workload, and challenges related to the usability of technological tools and the integration of digital systems into routine clinical practice [11-13].

Existing evidence on HaH has focused primarily on clinical outcomes, cost reduction, and patient satisfaction [14-17]. However, the experience of health care professionals working within highly digital health care models remains less explored, particularly in established programs where telemedicine and remote monitoring are integral to daily operations. Understanding how professionals perceive the usefulness, usability, workload, and care impact of these technologies is essential to ensuring the sustainability and scalability of such programs [18-20].

In this context, the aim of this study was to evaluate health care professionals' experience and perceptions regarding an established, digitally supported HaH model, examining aspects related to usability, organizational integration, workload, clinical communication, and perceived quality of care.

Methods

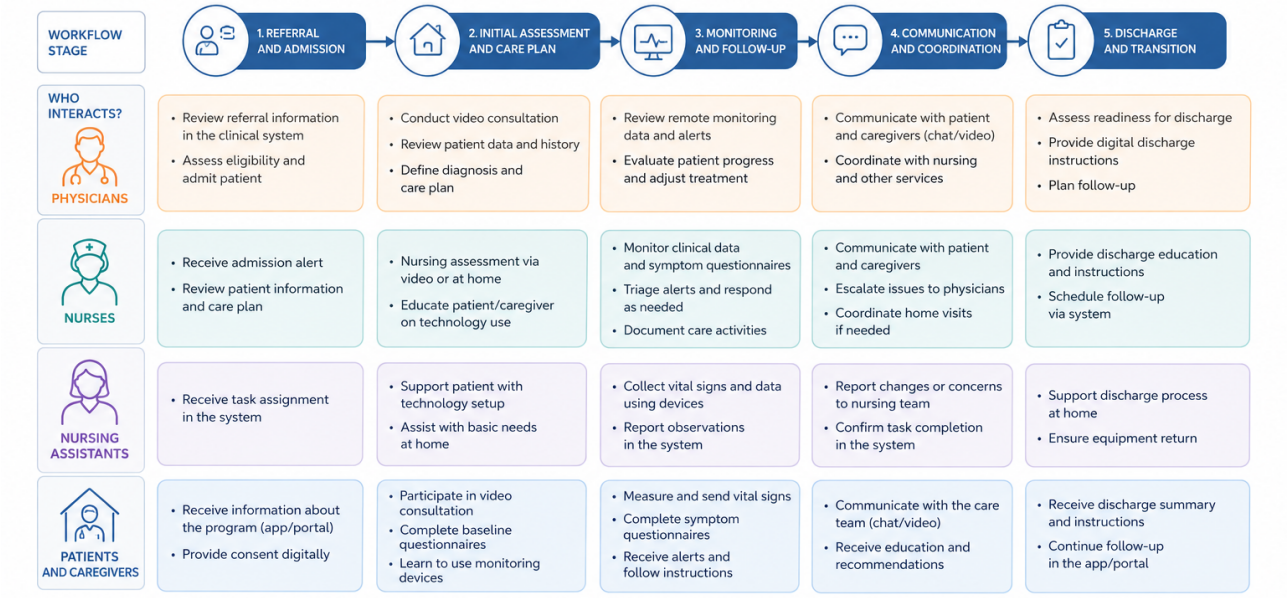
Study Design and Context

We conducted a single-center cross-sectional observational study in the HaH unit of Infanta Leonor University Hospital, Madrid, Spain. The study aimed to evaluate health care professionals' perceptions and experiences regarding an established digitally supported HaH program integrated into routine clinical practice. In this study, the term "established" refers descriptively to a program implemented in 2020 that has maintained sustained operational activity and integrated digital workflows.

The evaluated model combines telemedicine consultations, remote monitoring of vital signs, structured digital symptom questionnaires, and automated clinical alerts. Since its implementation, the program has processed more than 500,000 digital alerts and currently manages approximately 1200 admissions per year. Nearly 60% of these admissions involve adults aged 65 years or older, and more than 75% of follow-up activities are conducted remotely, with in-person visits scheduled according to clinical needs.

To support this level of activity, the HaH unit is staffed by 13 physicians, 25 nurses, and 9 nursing assistants, who provide coordinated multidisciplinary care within a structured digital workflow (Figure 1). The unit manages a broad range of conditions, including respiratory tract infections, urinary tract infections, intra-abdominal infections, hematologic disorders, heart failure, skin and soft tissue infections, and osteoarticular disorders.

Figure 1. Workflow of the hospital-at-home unit, illustrating the roles and responsibilities of each professional within the digitally supported care model.



Participants and Recruitment

Recruitment was conducted between January and May 2025 using an internally distributed QR code linked to an electronic survey. The survey was available throughout the study period, allowing participants to complete it at a time of their choice, either during or outside working hours, with no specific time window during the workday required for completion. All health

XSL•FO
RenderX

care professionals assigned to the HaH unit during the study period were invited to participate (N=47).

Participation was voluntary and anonymous. Before accessing the questionnaire, participants were required to provide electronic informed consent by confirming their willingness to participate and authorizing anonymous data collection.

No additional exclusion criteria were established.

Data Collection

In addition to survey responses, professional background variables were recorded, including professional group, years of experience, and self-reported training status related to the digitally supported HaH model. Training status was categorized as "yes," "partially," or "no" and reflected participants' perceived sufficiency of training rather than objectively verified completion of a training program. A response of "yes" indicated that participants felt they had received, or were currently receiving, sufficient training to use the technological platform. "Partially" indicated that some training had been received but that additional instruction was considered necessary, and "no" indicated that no training had been received.

Survey Instrument

A study-specific questionnaire was developed to explore health care professionals' perceptions of an established digitally supported HaH model. The questionnaire was informed by constructs derived from the technology acceptance model [20,21] and the unified theory of acceptance and use of technology [19]. On the basis of these conceptual frameworks, the research team developed study-specific items tailored to the HaH setting to capture health care professionals' perceptions of the digitally supported HaH model. The questionnaire covered aspects related to understanding of the program, perceived clinical impact, interaction with technology, workload, and overall evaluation of the model.

The instrument was then reviewed by 2 experts with experience in the field, who assessed the relevance and clarity of the items during a joint review meeting. On the basis of their discussion, the wording and content of the items were reviewed and agreed upon by the research team. No formal pilot test was performed before the survey was distributed.

The final instrument included 14 items rated on 5-point Likert-type scales adapted to the nature of each question (Figure S1 in Multimedia Appendix 1). Depending on the item, response anchors reflected levels of agreement, perceived quantity, satisfaction, clarity, or perceived impact. Higher scores consistently indicated more favorable evaluations, except for the workload and stress items, for which higher values denoted greater perceived burden.

Because the questionnaire included multiple related domains, different Likert response anchors, and items with different interpretive directions, no overall Cronbach α was used to support claims of psychometric reliability or validity. The instrument was treated as an exploratory, study-specific questionnaire rather than as a formal psychometric scale.

Separate domain-level reliability estimates were not considered sufficiently informative because the questionnaire included multiple related domains, some of which were represented by only 2 items. The domain structure was therefore used to organize the descriptive presentation of the results rather than to define formally validated subscales. Given the exploratory and practice-oriented nature of the study, the questionnaire was intended to capture health care professionals' perceptions related to clinical workflows rather than function as a formal psychometric scale. Accordingly, no factor analysis was performed.

Additionally, the questionnaire included an optional open-ended question: "Would you like to add any comments or recommendations about your experience with the HaH program using technology?" Responses to this question were reviewed descriptively. No formal qualitative coding or thematic analysis was performed.

Statistical Analysis

Categorical variables were described using absolute frequencies and percentages. Responses to Likert items were summarized using means and SDs to facilitate the overall interpretation of professional perceptions. Medians and IQRs were also calculated because of the ordinal nature of the response scales and are provided in Table S1 in Multimedia Appendix 1.

Given the small sample size, the ordinal nature of the Likert scales, and the presence of skewed distributions with ceiling effects in several items, nonparametric tests were used for group comparisons. Differences according to years of professional experience were analyzed using the Mann-Whitney *U* test, comparing professionals with fewer than 10 years of experience with those with 10 or more years. This dichotomization was selected based on the observed distribution of the sample and to avoid creating smaller subgroups that could limit the stability and interpretability of comparisons. Differences between professional groups were evaluated using the Kruskal-Wallis test.

P values from the Mann-Whitney *U* and Kruskal-Wallis tests were adjusted separately across the 14 items using the Benjamini-Hochberg false discovery rate procedure. The 2 correction families were treated separately because they addressed distinct grouping variables (years of professional experience and professional group), while the same 14 item-level outcomes were evaluated within each family. Within each family, the Benjamini-Hochberg correction was applied across all 14 items because these were the prespecified outcomes evaluated for each grouping variable. This approach preserved the distinction between the 2 grouping comparisons while controlling the false discovery rate across the 14 outcomes within each family. An adjusted *P* value of <.05 was considered statistically significant. Raw *P* values for both sets of comparisons are provided in Table S1 in Multimedia Appendix 1.

For items with Kruskal-Wallis *P* values that remained significant after adjustment, the Dunn multiple-comparisons test was performed to identify pairwise differences. GraphPad Prism (version 8.0.1; GraphPad Software Inc) reported

XSL•FO
RenderX

multiplicity-adjusted *P* values for the pairwise comparisons; no additional multiplicity correction was applied to these adjusted *P* values.

To quantify the magnitude of these group differences, effect sizes were calculated using ε^2 for Kruskal-Wallis analyses and *r* for Mann-Whitney *U* tests, according to the standard methodology for nonparametric group comparisons [22,23]. For the latter, positive values indicated higher scores among professionals with fewer than 10 years of experience, whereas negative values indicated higher scores among those with 10 or more years of experience. The 95% CIs for both effect size measures were estimated using a percentile bootstrap resampling approach with 10,000 iterations.

Statistical analyses were performed using Python (version 3.13.0; Python Software Foundation) and GraphPad Prism (version 8.0.1).

Ethical Considerations

The study was approved by the ethics committee of the Infanta Leonor University Hospital (56301711). Participation was voluntary and anonymous. All participants provided electronic informed consent before completing the questionnaire. Data were treated confidentially and analyzed in accordance with current data protection regulations.

Results

Participant Characteristics

A total of 30 health care professionals from the HaH unit participated in the study, corresponding to 30 (64%) of 47 invited professionals. The sample included physicians (n=13, 43%), nurses (n=10, 33%), and nursing assistants (n=7, 23%). Participation varied across professional groups: all physicians completed the survey (13/13, 100%), followed by nursing assistants (7/9, 78%) and nurses (10/25, 40%). Slightly more than half of the participants (17/30, 57%) had fewer than 10 years of professional experience. Of the 30 participants, 22 (73%) reported sufficient training on the digital platform, 6 (20%) reported partial training, and 2 (7%) reported no training (Table 1).

Table . Characteristics of participating health care professionals (n=30).

Characteristics	Participants, n (%)
Professional group	
Physicians	13 (43)
Nurses	10 (33)
Nursing assistants	7 (23)
Years of experience	
<10 years	17 (57)
≥10 years	13 (43)
Training on the HaH[a] unit's technological platform (completed or ongoing)	
Yes	22 (73)
Partially	6 (20)
No	2 (7)

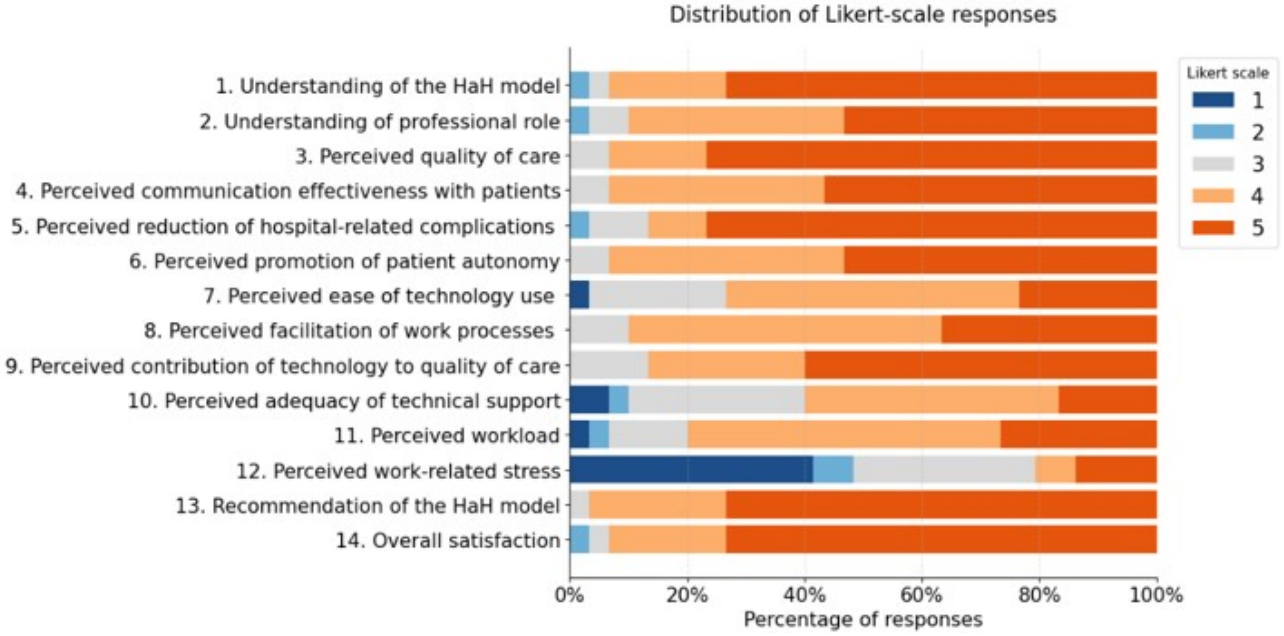
[a]HaH: hospital-at-home.

Survey Results

Overall perceptions of the program were positive (Figure 2). The highest scores were observed in items related to understanding of the program, perceived quality of care, and recommendation of the model. In contrast, perceived stress associated with HaH-related work received the lowest ratings. Greater variability was observed in responses related to ease of technology use, workload, and adequacy of technical support, suggesting more heterogeneous experiences across participants.

XSL•FO
RenderX

Figure 2. Distribution of 5-point Likert scale responses (percentages per item). For workload and stress, higher scores indicate greater perceived burden; for all other items, higher scores indicate more favorable evaluations. HaH: hospital-at-home.



Additionally, no statistically significant differences were identified according to years of professional experience after Benjamini-Hochberg adjustment (Table S2 in Multimedia Appendix 1). However, several exploratory differences across professional groups were observed (Table 2). Dunn post hoc comparisons indicated that significant pairwise differences occurred mainly between physicians and responding nurses, with nurses reporting lower scores in perceived communication effectiveness with patients (item 4), perceived promotion of patient autonomy (item 6), ease of technology use (item 7), facilitation of work processes (item 8), perceived contribution of technology to quality of care (item 9), and recommendation of the HaH model (item 13; Table S3 in Multimedia Appendix 1). Perceived workload (item 11) showed a different pattern, with physicians reporting higher scores, reflecting greater perceived burden.

XSL•FO
RenderX

Table . Mean (SD) scores of the questionnaire items according to professional groups (Likert scale 1 - 5)[a].

Conceptual category and item	Physicians, mean (SD)	Nurses, mean (SD)	Nursing assistants, mean (SD)	BH-adjusted[b] *P* value[c]	Effect size, ε² (95% CI)[d]
Understanding of the program					
Understanding of the HaH[e] model	4.6 (0.7)	4.7 (0.5)	4.6 (1.1)	.82	0 (0 - 0.19)
Understanding of professional role	4.6 (0.7)	4.2 (0.6)	4.3 (1.1)	.31	0.03 (0 - 0.41)
Perceived clinical impact					
Perceived quality of care	4.8 (0.4)	4.4 (0.8)	5 (0)	.19	0.07 (0 - 0.39)
Perceived communication effectiveness with patients	4.7 (0.5)	4.0 (0.7)	4.9 (0.4)	*.03*	0.28 (0.02 - 0.67)
Perceived reduction of hospital-related complications	4.8 (0.6)	4.1 (1.1)	5 (0)	.07	0.17 (0 - 0.57)
Perceived promotion of patient autonomy	4.9 (0.4)	3.9 (0.6)	4.6 (0.5)	*.007*	0.41 (0.15 - 0.78)
Interaction with technology					
Perceived ease of technology use	4.3 (0.6)	3.3 (1.0)	4.0 (0.8)	*.049*	0.20 (0.02 - 0.52)
Perceived facilitation of work processes	4.5 (0.5)	3.8 (0.6)	4.4 (0.5)	*.049*	0.20 (0 - 0.53)
Perceived contribution of technology to quality of care	4.8 (0.4)	3.8 (0.6)	4.7 (0.8)	*.006*	0.46 (0.17 - 0.85)
Perceived adequacy of technical support	4.0 (0.4)	3.1 (1.2)	3.6 (1.4)	.11	0.12 (0 - 0.60)
Workload					
Perceived workload	4.4 (1.1)	3.7 (0.7)	3.6 (0.5)	*.02*	0.31 (0.02 - 0.71)
Perceived work-related stress	2.8 (1.8)	2.2 (1.0)	2.1 (1.2)	.79	0 (0 - 0.25)
Overall evaluation					
Recommendation of the HaH model	4.9 (0.3)	4.2 (0.6)	5 (0)	*.006*	0.45 (0.14 - 0.85)
Overall satisfaction	4.7 (0.9)	4.4 (0.7)	4.9 (0.4)	.21	0.06 (0 - 0.46)

[a]Higher scores indicate higher levels of the construct measured; for workload and stress items, higher scores indicate greater perceived burden.

[b]BH: Benjamini-Hochberg.

[c]Kruskal-Wallis test. *P* values were adjusted using the BH false discovery rate correction across the 14 questionnaire items. BH adjustment was performed using the unrounded raw *P* values. Adjusted *P*<.05 are shown in italics.

[d]Effect sizes (ε²) were reported with 95% CIs and interpreted as small (ε²≥0.01), moderate (ε²≥0.08), or large (ε²≥0.26). The 95% CIs were estimated using percentile bootstrap resampling with 10,000 iterations.

[e]HaH: hospital-at-home.

An exploratory descriptive analysis was also conducted to examine responses to technology-related items according to self-reported training status (Table S4 in Multimedia Appendix 1). Participants reporting sufficient training ("yes") had descriptively higher ratings for ease of technology use and for the perceived contribution of technology to the quality of care, whereas those indicating partial or no training showed lower or more variable scores. Perceived adequacy of technical support displayed a similar descriptive pattern. In contrast, ratings related to the perceived facilitation of work processes appeared relatively consistent across the 3 training categories. Because the no-training group included only 2 participants, these comparisons are descriptive and were not subjected to inferential testing.

Of the 30 participants, 6 (20%) provided optional comments. Given the small number of responses, these comments were reviewed descriptively and are presented as supplementary contextual information (Table S5 in Multimedia Appendix 1). These comments offered contextual examples related to connectivity, platform functionality, and workflow challenges.

Discussion

Principal Findings and Comparison With Prior Work

This study evaluated health care professionals' perceptions and experiences of an established digitally supported HaH program, with particular attention to the interaction between technology and the organization of clinical work in a unit where telemedicine, remote monitoring, and automated alert systems are integrated into routine care workflows.

Overall, professionals reported a favorable perception of the model, particularly regarding perceived clinical usefulness,

XSL•FO
RenderX

perceived quality of care, and overall satisfaction with their experience in the unit. These findings are consistent with previous studies describing high levels of professional acceptance associated with the use of digital tools in health care settings [24,25]. However, unlike many previous investigations focused on pilot implementations or emergency deployment during the COVID-19 pandemic, our study evaluated an established model integrated into everyday clinical practice. This context allows a more realistic assessment of how sustained digital integration is perceived by professionals, while not providing evidence of objective clinical outcomes.

A key exploratory finding was that several differences between professional groups remained statistically significant after adjustment, especially between physicians and responding nurses, in aspects related to interaction with technology, perceived workload, and perceived contribution of technology to quality of care. The direction of these differences varied by item: responding nurses reported lower scores on several technology-related items, whereas physicians reported higher perceived workload. This result aligns with prior research showing that the experience of using digital tools may vary according to clinical role and the type of interaction each professional maintains with the technology throughout the care process [12,26-28]. However, these results should be interpreted with caution given the small subgroup sizes and the potential for nonresponse bias, particularly among nurses, who had the lowest participation rate (10/25, 40%). Because no information was available on the characteristics or perceptions of nonresponding nurses, it remains unclear whether those with more positive or negative experiences were more likely to participate.

In our setting, the observed differences may be related to the organizational structure of the digitally supported HaH model. Nursing staff interact more frequently with the digital platform due to responsibilities related to remote monitoring, alert management, and structured documentation of clinical information. In contrast, physicians use the system in a more targeted manner, primarily for clinical assessment and decision-making. This difference in the intensity and frequency of use may help explain the observed disparities, although the cross-sectional design does not establish causation.

The integration of automated monitoring systems and clinical alerts constitutes another relevant aspect of digitally supported HaH models. Although these tools are intended to facilitate early detection of clinical deterioration, they may also contribute to alert fatigue and increased cognitive workload when alerts are frequent, insufficiently prioritized, or poorly integrated into existing workflows [18,29].

Importantly, despite the observed differences in workflow-related factors, no significant differences were identified regarding overall satisfaction or work-related stress. This finding suggests that health care professionals may perceive the digitally supported HaH model as clinically valuable despite experiencing challenges associated with digital care delivery. Previous studies have shown that professional acceptance depends not only on usability but also on organizational factors and the clinical value that professionals attribute to the system [21,30].

These findings have several practical implications for the implementation of digitally supported HaH programs. First, the implementation of digital tools should be accompanied by organizational adaptation strategies and tailored training according to professional roles. Although formal comparisons could not be conducted due to small subgroup sizes, the exploratory descriptive analysis showed that participants who reported sufficient training had higher descriptive ratings for ease of technology use and perceived adequacy of technical support. In this context, nurse-centered interface design, clearer alert prioritization, simplified documentation workflows, and protected time for remote monitoring may help address the usability and workflow integration needs identified in this study. Second, improving interoperability between digital platforms and clinical information systems may be important for reducing administrative burden and avoiding duplication of tasks.

Furthermore, the availability of fast and accessible technical support could help alleviate part of the workload associated with remote care. Finally, HaH programs should systematically incorporate the evaluation of professional experience as a relevant indicator of organizational quality and model sustainability. Beyond clinical outcomes alone, long-term adoption of HaH programs may depend on how well digital tools fit into everyday clinical workflows.

Limitations

This study has several limitations. First, the sample size was relatively small, and subgroup sizes were particularly limited (13 physicians, 10 nurses, and 7 nursing assistants), which may limit statistical power and reduce generalizability. In particular, the exploratory analysis according to training status was limited by the marked imbalance between categories, particularly the very small no-training group (n=2), and was therefore restricted to descriptive analysis. Second, the low nurse response rate (10/25, 40%) raises the possibility of nonresponse bias. Although the survey was available throughout the study period and could be completed at a time of the participants' choice, workload may still have influenced participation, and nurses with particularly positive or negative experiences may have been more likely to respond.

Third, the questionnaire was study specific and was not formally validated as a psychometric instrument. Although it was informed by established frameworks and reviewed by experts, no psychometric validity or reliability claim is made on the basis of the questionnaire, and no factor analysis was performed.

Fourth, while the digitally supported HaH program has been established since 2020, its operational characterization has been limited to the indicators available for this study. More granular measures, such as episode duration, the number of remote contacts or in-person visits per admission, and alert volume per patient or episode, were not available but could provide additional context on program operations.

Finally, the open-ended component was limited because only a small number of participants provided open-ended responses. As this study was conducted in a single established HaH

XSL•FO
RenderX

program, findings may not be directly transferable to recently implemented programs, programs with different staffing models, or settings using different digital platforms.

Conclusions

Health care professionals in this single-center established HaH program reported generally positive perceptions of the digitally supported model, although exploratory differences in usability, workflow-related perceptions, and perceived workload varied by professional role. Exploratory findings suggest that sustaining digitally supported HaH programs over time may benefit from role-specific support, workflow adaptation, and continued organizational investment. In particular, nurse-centered interface design, alert prioritization, documentation simplification, protected time for remote monitoring, and timely technical support may warrant consideration. Future multicenter studies with larger samples and validated instruments are needed to determine whether such strategies effectively enhance professional experience and workflow integration.

Generative AI Use

Generative AI (Microsoft Copilot) was used to assist with language editing, restructuring, and revision of portions of this manuscript. The authors reviewed all AI-assisted content, verified the scientific content and references, and remain fully responsible for the final manuscript.

Funding

This study was supported by the AI-DOMUS project, funded by the Ministerio para la Transformación Digital y de la Función Pública (Spain) under the Plan de Recuperación, Transformación y Resiliencia—NextGenerationEU, 2024. The funder had no role in the study design, data collection, data analysis, interpretation of results, or manuscript preparation.

Data Availability

The deidentified individual-level survey data generated or analyzed during this study are not publicly available due to institutional data-sharing policies. However, they may be obtained from the corresponding author on reasonable request, subject to institutional approval and applicable data protection requirements.

Authors' Contributions

CBG contributed to conceptualization, methodology, and writing—original draft. EMAU contributed to conceptualization, methodology, and writing—review and editing. OEL and SLL contributed to investigation through the development and drafting of the survey instrument. MCAH and CUG contributed to formal analysis and data curation. SSS and ACC contributed to investigation through data collection. MMM, as the corresponding author, contributed to conceptualization, methodology, writing—review and editing, and project administration. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questionnaire administered to health care professionals participating in the hospital-at-home program.
[DOCX File, 52 KB - humanfactors_v13i1e106615_app1.docx]

References

1. Wong HL, Ong CY, Ngo HJ, Yeo SS, Lee MH. An overview of hospital-at-home versus other models of care. J Gen Fam Med 2024;26(1):19-26. [doi: 10.1002/jgf2.742] [Medline: 39776876]
2. Romero-Ruperto S, Llaneras Artigues J, Mosquera-Brea M, Jiménez-Moreno FX. Hospitalización a domicilio en la asistencia de los pacientes con enfermedad aguda [Hospitalization at home in the assistance for patients with acute pathology] [Article in English, Spanish]. Med Clin (Barc) 2025 Jan;164(2):91-96. [doi: 10.1016/j.medcle.2024.07.022] [Medline: 39271445]
3. Sultani K, Smeulers M, de Vries R, Zonderhuis BM, Nanayakkara PW. Transforming acute care: a scoping review on the effectiveness, safety and implementation challenges of hospital-at-home models. BMJ Open 2025 Aug 8;15(8):e098411. [doi: 10.1136/bmjopen-2024-098411] [Medline: 40780723]
4. Leong MQ, Lim CW, Lai YF. Comparison of hospital-at-home models: a systematic review of reviews. BMJ Open 2021 Jan 29;11(1):e043285. [doi: 10.1136/bmjopen-2020-043285] [Medline: 33514582]
5. Ko SQ, Goh J, Tay YK, et al. Treating acutely ill patients at home: data from Singapore. Ann Acad Med Singap 2022 Jul;51(7):392-399. [doi: 10.47102/annals-acadmedsg.2021465] [Medline: 35906938]
6. Vudathaneni VK, Lanke RB, Mudaliyar MC, Movva KV, Mounika Kalluri L, Boyapati R. The impact of telemedicine and remote patient monitoring on healthcare delivery: a comprehensive evaluation. Cureus 2024 Mar;16(3):e55534. [doi: 10.7759/cureus.55534] [Medline: 38576693]

XSL•FO
RenderX

7. Snoswell CL, Chelberg G, De Guzman KR, et al. The clinical effectiveness of telehealth: a systematic review of meta-analyses from 2010 to 2019. J Telemed Telecare 2023 Oct;29(9):669-684. [doi: 10.1177/1357633X211022907] [Medline: 34184580]
8. Adirim T. Digital health technologies: virtual health and remote patient monitoring. In: Adirim T, editor. Digital Health, AI and Generative AI in Healthcare: Springer; 2025:11-26. [doi: 10.1007/978-3-031-83526-1_2]
9. Isakov TM, Härkönen H, Atkova I, et al. From challenges to opportunities: digital transformation in hospital-at-home care. Int J Med Inform 2024 Dec;192:105644. [doi: 10.1016/j.ijmedinf.2024.105644] [Medline: 39393125]
10. Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. Int J Med Inform 2013 May;82(5):e73-e86. [doi: 10.1016/j.ijmedinf.2012.10.007] [Medline: 23146626]
11. Kaihlanen AM, Laukka E, Nadav J, et al. The effects of digitalisation on health and social care work: a qualitative descriptive study of the perceptions of professionals and managers. BMC Health Serv Res 2023 Jun 30;23(1):714. [doi: 10.1186/s12913-023-09730-y] [Medline: 37386423]
12. Low SY, Ko SQ, Ang IY. Health care providers' experiences and perceptions with telehealth tools in a hospital-at-home program: mixed methods study. JMIR Hum Factors 2025 Apr 17;12:e56860. [doi: 10.2196/56860] [Medline: 40245429]
13. Mohammadnejad F, Freeman S, Klassen-Ross T, Hemingway D, Banner D. Impacts of technology use on the workload of registered nurses: a scoping review. J Rehabil Assist Technol Eng 2023;10:20556683231180189. [doi: 10.1177/20556683231180189] [Medline: 37342268]
14. de Sousa Vale J, Franco AI, Oliveira CV, Araújo I, Sousa D. Hospital at home: an overview of literature. Home Health Care Manag Pract 2020;32(2):118-123. [doi: 10.1177/1084822319880930]
15. Nikmanesh P, Arabloo J, Gorji HA. Dimensions and components of hospital-at-home care: a systematic review. BMC Health Serv Res 2024 Nov 25;24(1):1458. [doi: 10.1186/s12913-024-11970-5] [Medline: 39587580]
16. Shaw A, Moodley M, McSporran K, et al. P18 Safety and effectiveness of an integrated, telehealth-led supported discharge service for Covid-19. Thorax 2021;76:A76. [doi: 10.1136/thorax-2021-BTSabstracts.128]
17. Ko SQ, Chua CM, Koh SH, Lim YW, Shorey S. Experiences of patients and their caregivers admitted to a hospital-at-home program in Singapore: a descriptive qualitative study. J Gen Intern Med 2023 Feb;38(3):691-698. [doi: 10.1007/s11606-022-07765-1] [Medline: 36008593]
18. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367. [doi: 10.2196/jmir.8775] [Medline: 29092808]
19. Venkatesh V, Thong JY, Xu X. Unified theory of acceptance and use of technology: a synthesis and the road ahead. J Assoc Inf Syst 2016;17(5). [doi: 10.17705/1jais.00428]
20. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172. [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
21. Gagnon MP, Orruño E, Asua J, Abdeljelil AB, Emparanza J. Using a modified technology acceptance model to evaluate healthcare professionals' adoption of a new telemonitoring system. Telemed J E Health 2012;18(1):54-59. [doi: 10.1089/tmj.2011.0066] [Medline: 22082108]
22. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: Routledge; 1988. [doi: 10.4324/9780203771587]
23. Fiel Peres F. Effect sizes for nonparametric tests. Biochem Med (Zagreb) 2026 Feb 15;36(1):010101. [doi: 10.11613/BM.2026.010101] [Medline: 41399660]
24. Lampickienė I, Davoody N. Healthcare professionals' experience of performing digital care visits-a scoping review. Life (Basel) 2022 Jun 17;12(6):913. [doi: 10.3390/life12060913] [Medline: 35743944]
25. Tan AJ, Rusli KD, McKenna L, Tan LL, Liaw SY. Telemedicine experiences and perspectives of healthcare providers in long-term care: a scoping review. J Telemed Telecare 2024 Feb;30(2):230-249. [doi: 10.1177/1357633X211049206] [Medline: 34666535]
26. Frennert S, Petersson L, Erlingsdottir G. "More" work for nurses: the ironies of eHealth. BMC Health Serv Res 2023 Apr 27;23(1):411. [doi: 10.1186/s12913-023-09418-3] [Medline: 37106404]
27. Turner K, Bobonis Babilonia M, Naso C, et al. Health care providers' and professionals' experiences with telehealth oncology implementation during the COVID-19 pandemic: a qualitative study. J Med Internet Res 2022 Jan 19;24(1):e29635. [doi: 10.2196/29635] [Medline: 34907900]
28. MacNeill V, Sanders C, Fitzpatrick R, et al. Experiences of front-line health professionals in the delivery of telehealth: a qualitative study. Br J Gen Pract 2014 Jul;64(624):e401-e407. [doi: 10.3399/bjgp14X680485] [Medline: 24982492]
29. Michels EA, Gilbert S, Koval I, Wekenborg MK. Alarm fatigue in healthcare: a scoping review of definitions, influencing factors, and mitigation strategies. BMC Nurs 2025 Jun 20;24(1):664. [doi: 10.1186/s12912-025-03369-2] [Medline: 40542361]
30. Cruz-Cunha MM, Miranda IM, Gonçalves P, editors. Handbook of Research on ICTs and Management Systems for Improving Efficiency in Healthcare and Social Care: IGI Global; 2013. [doi: 10.4018/978-1-4666-3990-4]

XSL•FO
RenderX

Abbreviations

HaH: hospital-at-home

Edited by S Law; submitted 09.Jul.2026; peer-reviewed by WJ He; revised version received 02.Oct.2026; accepted 05.Oct.2026; published 13.Nov.2026.

Please cite as:
Bibiano Guillén C, Arroyo Urea EM, Estremera López O, López Lallave S, Aranda Higuero MC, Ubeda García C, Sánchez Sierra S, Calvo Cruz A, Mir Montero M
Health Care Professionals' Experience in a Digitally Supported Hospital-at-Home Program: Cross-Sectional Study
JMIR Hum Factors 2026;13:e106615
URL: https://humanfactors.jmir.org/2026/1/e106615
doi:10.2196/106615

© Carlos Bibiano Guillén, Eva Maria Arroyo Urea, Olga Estremera López, Sonia López Lallave, María Consolación Aranda Higuero, Cristina Ubeda García, Silvia Sánchez Sierra, Almudena Calvo Cruz, María Mir Montero. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 13.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Contextual Assessments for Chronic Obstructive Pulmonary Disease Transition of Care Bundle Implementation Planning for the Reduce REVISITS Study: Rapid Sequential Explanatory Mixed Methods Approach

Mahima Akula[1], BA; Kim Erwin[2], MDes; Leah Traeger[1], MHA; Hannah Pick[3], MA; Fei Gao[3], MBA; Laura Damschroder[4], MS; Valerie G Press[1], MD, MPH

[1]Department of Medicine, University of Chicago, 5841 S Maryland Ave MC 2007, Chicago, IL, United States

[2]Institute of Design, Illinois Institute of Technology, Chicago, IL, United States

[3]Onda Collective, LLC, Chicago, IL, United States

[4]Implementation Pathways, Chelsea, MI, United States

Corresponding Author:
Valerie G Press, MD, MPH
Department of Medicine, University of Chicago, 5841 S Maryland Ave MC 2007, Chicago, IL, United States

Abstract

Background: Chronic obstructive pulmonary disease (COPD) affects more than 16 million US adults, many of whom experience high rates of acute care revisits (emergency department and hospital) after initial hospitalization. These frequent exacerbations, often due to failures in transitions of care (TOC), lead to lung function decline and premature mortality. While effective interventions exist to reduce readmissions, wide-scale implementation of COPD TOC programs remains limited. The National Institutes of Health–funded Reducing Respiratory Emergency Visits Using Implementation Science Interventions Tailored to Settings (REVISITS) study was designed to address this implementation gap by developing and implementing bundled COPD TOC programs across diverse US hospitals.

Objective: This study aimed to conduct pre-implementation contextual assessments at US hospitals to guide the development of site-specific, evidence-based COPD TOC programs.

Methods: We conducted pre-implementation contextual assessments using a novel semi-structured interview format that integrated the Consolidated Framework for Implementation Research (CFIR) with human-centered design approaches (ethnographic interviewing) to capture real-world experiences of COPD care across inpatient, outpatient, and home settings. We used a sequential explanatory mixed methods design in which pre-interview survey data completed by site leads informed and shaped the subsequent semi-structured interviews. Site leads, clinicians, organizational leaders, patients, and caregivers were interviewed. Interviews explored baseline COPD TOC practices, local resources, opportunities for improvement, as well as participant priorities from a menu of 12 evidence-based interventions (eg, pulmonary rehabilitation, patient navigation, and inhaler teaching). Rapid analysis methods identified intervention priorities across participant groups, along with perceived barriers and facilitators to implementation. Findings were shared with site leads to help guide their development of tailored COPD TOC programs.

Results: Among 194 participants from 21 sites (42 site leads, 29 organizational leaders, 105 clinicians, and 18 patients or caregivers), the highest priority interventions identified during interviews were post–emergency department follow-up visits, education (inhaler technique, disease management, and action plan), and pulmonary rehabilitation. Reported barriers included clinician-level challenges (limited training, staffing, and time), patient-level challenges (social needs and physical burden of COPD), and system-level challenges (lack of standardization, limited resources, and cost). Key facilitators included the presence of dedicated staff and the availability of pre-existing programs or infrastructure. The 3 most commonly chosen interventions for implementation were patient education (eg, inhaler education and COPD action plans), medication reconciliation, and post-discharge care (eg, post-discharge visits and pulmonary rehabilitation).

Conclusions: This study demonstrates how the integration of implementation science and human-centered design approaches can yield valuable insights, beyond what either field could obtain separately, during the pre-implementation phase of COPD TOC program implementation development. Contextual assessments that capture diverse views are instrumental in designing feasible and relevant interventions. Future work will explore how pre-implementation insights relate to post-implementation outcomes across participating sites.

Trial Registration: ClinicalTrials.gov NCT05568043; https://clinicaltrials.gov/study/NCT05568043

XSL•FO
RenderX

(JMIR Hum Factors 2026;13:e82078) doi:10.2196/82078

KEYWORDS

chronic obstructive pulmonary disease; COPD; readmission; revisits; transitions of care

Introduction

Chronic obstructive pulmonary disease (COPD) affects more than 16 million US adults, many of whom experience high rates of acute care revisits (emergency department [ED] and hospital) after initial COPD hospitalizations [1-5]. These frequent exacerbations, often due to failures in transitions of care (TOC), lead to lung function decline and premature mortality [6]. In response, the Centers for Medicare and Medicaid Services launched the Hospital Readmission Reduction Program (HRRP), which includes COPD readmissions as a targeted outcome [7]. Additionally, the Centers for Medicare and Medicaid Services introduced the Bundled Payments for Care Improvement (BPCI) initiative, a voluntary program designed to encourage coordinated care and reduce costs associated with hospital readmissions [7-9]. These policies motivated many US hospitals to pursue interventions aimed at reducing COPD readmissions to avoid financial penalties. However, beyond a few published reports of successful TOC interventions, there is limited clarity regarding what constitutes a successful COPD TOC program [10-14]. To date, no systematic evaluation has been conducted on the types or consistency of care transition interventions used within or across hospitals since the implementation of the HRRP. As a result, wide-scale implementation of COPD TOC programs remains limited by fundamental gaps in effectiveness and implementation data [15-17].

The Reduce Respiratory Emergency Visits Using Implementation Science Interventions Tailored to Settings (REVISITS) study was designed to address these gaps by developing, implementing, and evaluating bundled COPD TOC programs across diverse US hospitals [18]. Before sites launched their tailored COPD TOC programs (comprised of 2‐3 evidence-based interventions), we conducted pre-implementation contextual assessments to identify baseline COPD TOC practices, local resources, and opportunities for improvement. These contextual assessments were intended to inform site-specific program development and promote successful implementation and sustainability.

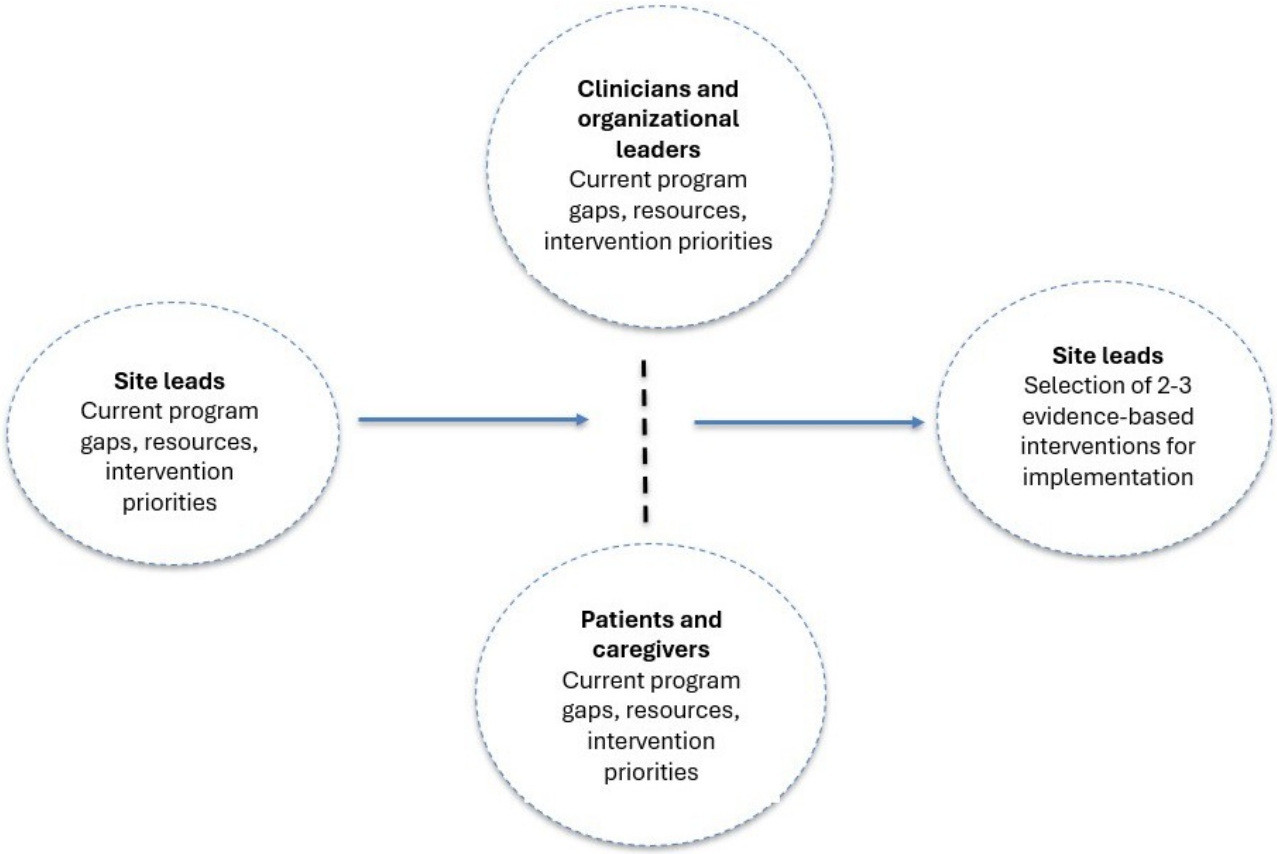
Methods

Study Design and Setting

The Reduce REVISITS study used a sequential explanatory mixed methods design [19]. Site leads first completed a pre-interview survey to identify preliminary priorities, site characteristics, and contextual factors. These survey responses were then used to inform and tailor the subsequent semi-structured interviews, during which participants elaborated on and clarified their survey responses. This sequential approach allowed the qualitative interviews to build on the quantitative survey data, providing deeper insight into site-level priorities and implementation considerations. Using a structured, stepwise discovery process, we conducted pre-implementation contextual assessments at each participating site. These assessments were designed to help sites develop COPD TOC programs tailored to their specific needs and environment. Specifically, the assessments aimed to characterize each site's local context, evaluate readiness for change, and identify desirable evidence-based COPD interventions for inclusion in the site's TOC program (Figure 1).

XSL•FO
RenderX

Figure 1. Pre-implementation contextual assessment approach used in the Reduce Respiratory Emergency Visits Using Implementation Science Interventions Tailored to Settings (REVISITS) study, a sequential explanatory mixed methods study conducted across 21 US hospitals to inform the design of chronic obstructive pulmonary disease transition of care programs. The figure outlines the stepwise discovery process integrating human-centered design and implementation science methods, including pre-interview surveys, multiparticipant interviews, rapid analysis, and site-specific implementation planning.



Hospitals were recruited into this initial phase of the study from the Hospital Medicine Reengineering Network [20] and through the Society of Hospital Medicine [21]. The study aimed to enroll at least 20 hospitals to complete this phase to ensure sufficient power for the randomized trial in the following phase [18].

Each participating hospital designated 2 site leads responsible for leading local study-related COPD TOC intervention implementation efforts. The process began with site leads completing a brief survey (Multimedia Appendix 1), followed by in-depth interviews. During the interviews, site leads described the current state of COPD care at their organization, identified potential barriers and facilitators to implementation, and prioritized interventions for inclusion in their site's TOC program. Site leads then identified additional key participants, including clinicians and organizational leaders, whose input would further inform the COPD TOC program implementation. Site leads were also encouraged to distribute flyers and engage in efforts to recruit patients hospitalized with a COPD exacerbation in the past year and their caregivers to the study.

Data Collection Tools and Procedures

We developed data collection tools by human-centered design (HCD) ethnographic interviewing techniques, such as journey mapping, projective images, and concept ranking [22,23], with implementation science (IS) methodology, specifically constructs from the Consolidated Framework for Implementation Research (CFIR) [24-28].

Qualitative data collection was guided by a pragmatic research paradigm, which emphasizes generating actionable, contextually grounded insights to support real-world implementation planning. This paradigm aligned with the study's objective to identify practical barriers, facilitators, and priorities that could directly inform the design of site-specific COPD TOC programs.

Site leads first completed a pre-interview survey on their site's current practices and resources that informed the patient or caregiver, clinician, and organizational leader interviews. All participants were interviewed using semi-structured guides. During interviews, participants interacted with a structured slide deck to prompt reflection and discussion. For example, patients were presented with 11 projective images and asked to select 2 that represented their experiences with COPD, while clinicians ranked 12 evidence-based interventions by priority. Interviews were conducted by members of the study team trained in qualitative interviewing, including individuals with backgrounds in medicine (VGP), HCD (KE), health services research (MA and LT), behavioral science (HP and FG), and IS (LD). We acknowledge that these professional backgrounds may influence data collection and interpretation. To mitigate bias, interviewers used consistent probes and documented reflections during and after interviews. Each interview was attended by at least 2 members of the study team (MA, LT, KE, HP, and FG). One team member conducted the interview while the second took detailed notes, facilitating cross-validation and supporting subsequent analysis. No interviewers had previous relationships

XSL•FO
RenderX

with any participants. Interviewers introduced themselves by name and study role at the start of each interview and emphasized that participation was voluntary and confidential.

Site leads at each participating hospital assisted with identifying eligible potential participants and provided the study team with contact information for these individuals. Site leads also distributed study flyers to recruit patients hospitalized for a COPD exacerbation in the previous year and their caregivers. The study team used a standardized recruitment protocol across sites, which included initial email or phone outreach, confirmation of interest, and scheduling of interviews. To ensure consistency, all recruitment communications followed a uniform script.

All interviews were conducted virtually via Zoom (Zoom Communications) or a recorded telephone call. For patients and caregivers with limited technology access, study materials were mailed, and phone interviews were offered as an alternative to Zoom. No in-person interviews were conducted.

Data Types and Quantitative or Qualitative Approaches

Using a sequential explanatory mixed methods approach, we collected both qualitative and quantitative data. Quantitative data included structured ratings and rankings (eg, intervention priorities and readiness scores) and were summarized using descriptive statistics (means and medians) to assess inter- and intrasite variation in priorities and context. Qualitative data were analyzed using a qualitative descriptive design to summarize participants' experiences and perspectives. This approach aligns with the pragmatic paradigm and the study's objective to rapidly generate actionable insights rather than interpret lived experience from a philosophical standpoint. We adopted a rapid qualitative analysis approach, which allowed data to be synthesized and reported to site leads in time to inform the development of COPD TOC programs.

Consistent with a sequential explanatory mixed methods design, quantitative survey data were first used to shape and tailor the subsequent qualitative interviews. Integration occurred at multiple points: (1) during data collection, where pre-interview survey responses informed interview probes and areas for follow-up; (2) during analysis, where quantitative intervention rankings were compared with qualitative descriptions of barriers and facilitators; and (3) during interpretation, where findings across data sources were synthesized to develop site-specific contextual needs. This integrated understanding directly informed the site-specific implementation planning discussions.

Rapid Analysis and Summary Development

Interview data from site leads, organizational leaders, clinicians, patients, and caregivers were collected and analyzed by the study team (MA) using a rapid analysis approach. This analysis resulted in the formation of a "priority summary table," which ranked interventions based on the number of participants who identified them as high priorities, and a "resource table," which listed available resources that could support implementation of priority interventions. These materials were compiled to inform intervention planning for each site.

Implementation Planning Discussions and Reports

Post analysis, site-specific findings, and intervention-specific evidence summaries were presented to site leads during facilitated implementation planning discussions led by the principal investigator (VGP). Based on this information, site leads were asked to select 2 - 3 interventions to implement as part of their site's COPD TOC program. Following these discussions, the study team (MA and LT) created tailored "site implementation reports" that site leads could use when planning implementation of their selected interventions. These reports included key takeaways from interview data, such as barriers and facilitators to implementation, priority and resource summary tables, and notes from the implementation planning discussion. Each report also included a "workbook" component to guide sites in planning implementation steps and collecting baseline quantitative data (Multimedia Appendix 2).

Ethical Considerations

This study was reviewed and approved by the University of Chicago Institutional Review Board (IRB; IRB20-0373). All participants provided oral informed consent before participation in semi-structured interviews conducted via Zoom or recorded phone call. Site leads at each participating hospital assisted in identifying eligible stakeholders and shared contact information with the study team, who then contacted potential participants directly.

All research activities were conducted by trained study personnel following a standardized, written protocol. Participant confidentiality was rigorously maintained: all site-level data were fully deidentified, hospital data were stored in REDCap (Research Electronic Data Capture; Vanderbilt University), and interview data were housed on a secure, HIPAA (Health Insurance Portability and Accountability Act)–compliant shared drive managed by the University of Chicago Center for Research Informatics. Study ID numbers were used in place of personal identifiers. Identifiable contact information collected for recruitment and compensation was stored separately from study data and will be destroyed, along with audio files, after study completion.

Participants received compensation approved by the IRB. Patients and caregivers received a US $25 gift card, and clinicians and organizational leaders received a US $10 gift card. Site leads who supported implementation at participating hospitals received an honorarium of US $3500 per year of participation. Raw data will not be shared outside the study investigators, research personnel, and approved external collaborators covered under the University of Chicago IRB.

Results

Site Characteristics

In total, 21 hospital sites were enrolled in the first phase of the Reduce REVISITS study. Approximately two-thirds (13/21) were academic hospitals, and one-third (8/21) were community hospitals. There was considerable geographic diversity among enrolled sites, with the Midwestern (7/21), Northeastern (6/21), and Southern (6/21) regions of the United States roughly equally

XSL•FO
RenderX

represented. However, only 2 sites were enrolled from the Western region (Table 1).

Table . Characteristics of the 21 US hospital sites enrolled in the pre-implementation phase of the Reduce REVISITS[a] study, including hospital type and region.

Characteristic	Sites (n=21), n (%)
Academic	13 (61.9)
Community	8 (38.1)
Northeast	6 (28.5)
Midwest	7 (33.3)
South	6 (28.5)
West	2 (9.5)

[a]REVISITS: Respiratory Emergency Visits Using Implementation Science Interventions Tailored to Settings.

Across the 21 sites, a total of 194 participants were interviewed. Among these participants, 176 were hospital-affiliated individuals (site leads [n=42], organizational leaders [n=29], and clinicians [n=105]), and 18 were patients or caregivers (n=16 and 2, respectively).

Participant Characteristics

Overall, participant demographics varied by participant group (Table 2, Supplementary Table S1 in Multimedia Appendix 3). Hospital-affiliated individuals tended to be younger than patients and caregivers. When obtaining demographic information, we asked participants what gender they currently identify with, and gender distribution varied by participant group. Although side leads were roughly split between males and females (22 females and 20 males), most organizational leaders (21/29) were male, while most clinicians (70/105) and patients and caregivers (13/18) were female. Racial diversity was limited and also varied by participant group. Most notably, there were significant socioeconomic and educational disparities. Most hospital-affiliated individuals (149/176) reported household incomes higher than US $100,000, whereas approximately two-thirds (11/18) of patients and caregivers reported an average household income of less than US $40,000. The majority of site leads, organizational leaders, and clinicians (140/176) held a master's degree or higher, whereas most patients and caregivers (14/18) had not attained a bachelor's degree.

XSL•FO
RenderX

Table . Demographic characteristics of the 194 stakeholders interviewed during pre-implementation contextual assessments for the Reduce REVISITS[a] study, including site leads, clinicians, organizational leaders, and patients or caregivers across 21 US hospitals.

Characteristics	Patients+ caregivers (n=18)	Site leads (n=42)	Organizational leaders (n=29)	Clinicians (n=105)
Age (y)				
Mean (SD)	65.83 (10.62)	42.10 (8.45)	47.61 (9.52)	43.36 (9.25)
Median (IQR)	67.00 (9.5)	40.50 (12.75)	46.00 (9.5)	41.50 (13)
Range (minimum-maximum)	47 (40 - 87)	32 (29 - 61)	36 (33 - 69)	42 (26 - 68)
Declined to answer, n (%)	0 (0)	0 (0)	1 (3.45)	7 (6.67)
Gender identity, n (%)				
Female	13 (68.42)	22 (52.38)	8 (27.59)	70 (66.67)
Male	5 (26.32)	20 (47.62)	21 (72.41)	35 (33.33)
Race, n (%)				
American Indian or Alaskan Native	0 (0)	0 (0)	0 (0)	0 (0)
Asian	0 (0)	11 (26.19)	5 (17.24)	14 (13.33)
Black or African American	5 (26.32)	1 (2.38)	0 (0)	5 (4.76)
Native Hawaiian or Other Pacific Islander	0 (0)	0 (0)	0 (0)	1 (0.95)
White	11 (57.90)	26 (61.90)	23 (79.31)	73 (69.52)
Other	1 (5.26)	3 (7.14)	1 (3.45)	7 (6.67)
Declined to answer	1 (5.26)	1 (2.38)	0 (0)	5 (4.76)
Ethnicity, n (%)				
Hispanic or Latino	0 (0)	1 (2.38)	1 (3.45)	4 (3.81)
Not Hispanic or Latino	18 (100)	39 (92.86)	28 (96.55)	97 (92.38)
Declined to answer	0 (0)	2 (4.76)	0 (0)	4 (3.81)
Average household income (US $), n (%)				
Less than 20,000	9 (47.37)	0 (0)	0 (0)	0 (0)
20,000-40,000	2 (10.53)	0 (0)	0 (0)	0 (0)
40,000-60,000	1 (5.26)	1 (2.38)	0 (0)	1 (0.95)
60,000-80,000	0 (0)	1 (2.38)	0 (0)	4 (3.81)
80,000-100,000	0 (0)	1 (2.38)	2 (6.9)	2 (1.9)
More than 100,000	2 (10.53)	35 (83.33)	26 (89.66)	88 (83.81)
Declined to answer	4 (21.05)	4 (9.52)	1 (3.45)	10 (9.52)
Highest level of education, n (%)				
No formal educational credential	3 (15.79)	0 (0)	0 (0)	0 (0)
High school diploma or equivalent	6 (31.58)	0 (0)	0 (0)	0 (0)
Some college, no degree	4 (21.05)	1 (2.38)	0 (0)	0 (0)
Associate's degree	1 (5.26)	1 (2.38)	0 (0)	10 (9.52)
Bachelor's degree	3 (15.79)	3 (7.14)	1 (3.45)	20 (19.05)
Master's degree	0 (0)	6 (14.29)	3 (10.34)	18 (17.14)
Doctor or professional degree	0 (0)	31 (73.81)	25 (86.21)	57 (54.29)

XSL•FO
RenderX

Characteristics	Patients+ caregivers (n=18)	Site leads (n=42)	Organizational leaders (n=29)	Clinicians (n=105)
Other	1 (5.26)	0 (0)	0 (0)	0 (0)

[a]REVISITS: Respiratory Emergency Visits Using Implementation Science Interventions Tailored to Settings.

Interventions Identified as Priorities by Interview Participants

Across all 194 participants, the most frequently prioritized interventions were post–hospital or ED follow-up visits (90/194, 46%), inhaler education (78/194, 40%), and COPD education (66/194, 34%). Pulmonary rehabilitation was similarly prioritized (66/194, 34%), followed by COPD action plan (61/194, 31%), and patient navigation (60/194, 31%). The least frequently prioritized interventions included remote monitoring and reporting, advanced care planning, and spirometry. High-priority interventions were similar across hospital-affiliated and patient participants, although hospital-affiliated participants more strongly prioritized post–hospital or ED follow-up visits (84/176 hospital-affiliated vs 6/18 patient participants), inhaler education (73/176 hospital-affiliated vs 5/18 patient participants), and COPD action plan (58/176 hospital-affiliated vs 3/18 patient participants).

There was general agreement across participant groups about both the most and least prioritized interventions (Table 3). For example, post–hospital or ED follow-up visit, COPD education, and COPD action plan were frequently selected as top priorities, while remote monitoring and reporting, advanced care planning, and spirometry were less favored. However, there was more participant variation for interventions that were only moderately prioritized. For example, pulmonary rehabilitation was prioritized more often by site leads and patients or caregivers than by organizational leaders and clinicians. The COPD action plan was the top-ranked intervention among clinicians, although it was not equally prioritized by other groups (Table 5).

XSL•FO
RenderX

Table . Evidence-based chronic obstructive pulmonary disease transitions of care interventions selected by site leads at 21 US hospitals during the pre-implementation phase of the Reduce REVISITS[a] study, shown alongside corresponding participant-reported priority counts.

	Frequency, n (%)
Selected interventions (number of sites)	
Inhaler education	20 (95.24)
Post-ED[b] follow-up visit	13 (61.9)
COPD[c] action plan	12 (57.14)
Medication reconciliation	9 (42.86)
Pulmonary rehabilitation	6 (28.57)
Patient navigation	6 (28.57)
Other COPD education	4 (19.05)
Smoking cessation education	4 (19.05)
Advanced care planning	0 (0)
Spirometry	0 (0)
Remote monitoring and reporting	0 (0)
Priority intervention (number of participants)	
Post-ED follow-up visit	90 (46.39)
Inhaler education	78 (40.21)
COPD education	66 (34.02)
Pulmonary rehabilitation	66 (34.02)
COPD action plan	61 (31.44)
Patient navigation	60 (30.93)
Medication reconciliation	57 (29.38)
Smoking cessation education or treatment	53 (27.32)
Remote monitoring and reporting	27 (13.92)
Advanced care planning	15 (7.73)
Spirometry	14 (7.22)

[a]REVISITS: Respiratory Emergency Visits Using Implementation Science Interventions Tailored to Settings.

[b]ED: emergency department.

[c]COPD: chronic obstructive pulmonary disease.

Participant-Identified Barriers and Facilitators

During interviews, site leads, organizational leaders, and clinicians identified a range of barriers and facilitators related to COPD TOC interventions (Table 4). Barriers occurred at the clinician, patient, and systems levels. Clinician-level barriers included a lack of staff, bandwidth, time, and lack of clinician education. Patient-level barriers included social determinants (eg, transportation, socioeconomic status, and technology literacy) as well as physical limitations (eg, poor vision and mobility issues). System-level barriers included a lack of standardization, inadequate resources (eg, placebo inhalers), shifts in care delivery (eg, nebulizer use), and cost-related issues (eg, pulmonary rehabilitation reimbursement and hiring patient navigators).

XSL•FO
RenderX

Table . Qualitative themes describing clinician-level, patient-level, and system-level barriers and facilitators to implementing chronic obstructive pulmonary disease transitions of care interventions, based on multiparticipant interviews conducted at 21 US hospitals during the pre-implementation phase of the Reduce REVISITS[a] study.

TOC[b] interventions	Representative CFIR[c] domains and subdomains with quotes	
	Barriers	Facilitators
Inhaler education	• Inner Setting: Work Infrastructure "We unfortunately don't have those dummy inhalers [to] educate [patients]." -RT[d] • Individuals: Innovation Deliverers "I have no idea how patients get inhaler education. It's not from me. I don't really know how to give that education." -SL[e]	• Inner Setting: Work Infrastructure "We have a respiratory therapist whose job is education, and she is an asset." -SL • Individuals: Innovation Deliverer "Previous to [the pulmonary navigator], I was just going, 'Puff, puff,' to both of them. And he taught me how to use it." -Patient
COPD[f] education	• Inner Setting: Work Infrastructure "It's really hard to have the time to sit with the patient and discuss [education] in the hospital." -Physician • Individuals: Innovation Recipients "My eyes aren't real good right now. So my glasses still ain't letting me read real good." -Patient	• Inner Setting: Work Infrastructure "[The nurse-led education service] get this consult, and then they have a team of nurses that basically meet with the patient and family one-on-one." -RN[g] • Individuals: Innovation Deliverer "The most valuable part [of treatment was] [the pulmonary navigators]'s directions on how to deal with all of it." -Patient
COPD action plan	• Innovation: Innovation Design "Some patients [are] discharged without a clear COPD action plan. Often it should be tucked in as part of their [after visit summary], but they leave with a 30-page document." -Physician • Inner Setting: Work Infrastructure "Just having enough staff to take care of patients is a problem. So whether the [COPD action plan] would be able to happen, I think it probably wouldn't." -Care Coordinator	• Innovation: Innovation Complexity "I feel like we see [the COPD action plan as] the low-hanging fruit one that's relatively easy to do." -Physician • Innovation: Relative Advantage "I have [a COPD action plan] on my refrigerator. [When] you're in the beginning of the yellow, then you might think, 'Oh, okay. It's okay. It'll get better.' But really, that's the starting of it go into the red." -Patient
Smoking cessation and/or treatment	• Inner Setting: Work Infrastructure "Whether they hand them a brochure or sit with the patient and have a 10-minute conversation, it's very dependent on the therapist." -RT • Individuals: Innovation Recipient "I used some of the patches, but they didn't really work because when I came home [from the hospital], my insurance wouldn't pay for them." -Patient	• Inner Setting: Available Resources "There's a smoking cessation clinic. Patients can get nicotine patches or resources to stop smoking." -Physician • Inner Setting: Work Infrastructure Deliverers "I'm trying to free up some full-time equivalents so that I can create some tobacco treatment specialists to go to the bedside." -OL[h]
Medication reconciliation	• Inner Setting: Work Infrastructure "[Medication reconciliation]'s owned by nursing and physicians here just because we simply don't have the pharmacy bandwidth to own that process." -Pharmacist • Outer Setting: Policies and Laws "It can be an hours-long process of trying to get approval for [the inhaler] or finding the right [one] that's on formulary." -SL	• Inner Setting: Information Technology "Our outpatient pharmacy just went online with Epic so that our nursing staff and physicians can see where in the process meds to beds are." -Pharmacist • Individuals: Innovation Deliverers "The transition-of-care pharmacist is extremely helpful. The pharmacists have a keen eye for what somebody's insurance will cover and what they won't." -OL

XSL•FO
RenderX

TOC[b] interventions	Representative CFIR[c] domains and subdomains with quotes	
	Barriers	Facilitators
Advanced care planning	• Inner Setting; Work Infrastructure "Our palliative service is really involved, but they are already over-maxed." -Physician • Individuals: Capability "There comes a point where it would be nice if there were professional people that were telling him this rather than his son telling him that the trajectory is not necessarily positive. I'm reluctant to bring it up with him because it increases his anxiety." -Caregiver	• Inner Setting: Information Technology Infrastructure "And then for the high risk [patients], we get that alert for [a] care conference screening, [which] sometimes [includes] discussing palliative care." -Care Coordinator • Inner Setting: Work Infrastructure "We do have a pretty robust advanced care planning initiative. [We have] people who are specifically hired to do advanced care planning for a good chunk of our inpatients." -SL
Post-ED[i] follow-up	• Inner Setting: Work Infrastructure "[Trying to get patients an appointment is] a very labor-intensive process. One of our biggest frustrations, I would say, as hospitalists trying to discharge patients is that we have no help in making these follow-up appointments." -SL • Outer Setting: Local Conditions "I got a insurance but they don't pay for transportation and stuff like that, I have to catch the bus, get on the train, and run from the train station to [the hospital], and I have to stop and rest because I get short of breath." -Patient	• Inner Setting: Available Resources "We have a transition of care clinic, which is pretty great. If you put in the order, they're good about going bedside to the patient, handing them information related to their discharge diagnosis and then the actual appointment in hand." -Physician • Inner Setting: Work Infrastructure "There's been a very small improvement in [clinician availability] because we've hired two pulmonologists that actually do clinics now." -SL
Pulmonary rehabilitation	• Outer Setting: Financing "There are only two to three pulmonary rehab facilities in [the region]. And it's very difficult for patients to get to, and it is because the Centers for Medicare and Medicaid Services reimbursement for pulmonary rehab is abysmally low." -Physician • Individuals: Innovation Recipients "I think [pulmonary rehab] was very good for me [in the past]. I think would be good for me now too [but] I don't want to ask my son because it's like you're there for two hours and then to travel is another hour." -Patient	• Outer Setting: Partnerships and Connections "To try address [transportation barriers], we are looking at tele pulmonary rehab programs that are currently being offered by external vendors." -Physician • Inner Setting: Information Technology "Pulmonary rehab [is] going to be included in the new COPD discharge order set that's going to roll out later this month." -SL
Patient navigation	• Innovation: Innovation Cost "But unfortunately, we only have one [nurse navigator]. We're hoping that [the system] [will] hire more nurse navigators, but they have not because, unfortunately, they have not proven their return on investment." -OL • Outer Setting: Policies and Laws "Yeah. I wish I had one of these little portable oxygen machine that I could tote on my shoulders. It's all up to my insurance company." -Patient	• Individuals: Innovation Deliverer "So in my role [as a navigator], I do catch a lot when I go to rounds. We do COPD education. If there are discharge needs, then I take care of that. Then [I do] a follow-up phone call at least 48 hours after they're discharged." -RT • Individuals: Innovation Deliverer "[The pulmonary navigator] is the most valuable person I've met [because of] his knowledge and his ability to translate medical terms into laypeople's terms so that I can understand it and was not frightened by any of the approach or the medication." -Patient
Spirometry		

XSL•FO
RenderX

TOC[b] interventions	Representative CFIR[c] domains and subdomains with quotes	
	Barriers	Facilitators
	• Inner Setting: Work Infrastructure "With COVID though, with the lab being shut down for quite a while, the backlog of people who needed pulmonary function testings got so large that we're booking three months out." -Physician • Individuals: Innovation Recipients "But in my opinion, patients are not really aggressive about using spirometry because it's a little aggravating and it's easily forgotten." -Care Coordinator	• Inner Setting: Work Infrastructure "We kind of piloted [an outreach program] on a very small scale. Once we identified our COPD population, [we] just pulled some [patients who] did not have spirometry and reached out [to schedule pulmonary function testings]. We learned that you can cold-call patients, and that patients do respond and come in." -Physician • Inner Setting: IT Infrastructure "[In the portal they ask] patients to fill out a questionnaire, and then based on the results of the questionnaire, it lets them self-schedule to get their spirometry done." -Physician
Remote monitoring and reporting	• Inner Setting: Work Infrastructure "We do have this remote monitoring system but as I mentioned, it's just not utilized. And then there's only so many kits that we have." -SL • Individuals: Innovation Recipients "I think a lot of the elderly patients find the tablet and entering all this to be a little overwhelming. They don't want to deal with it." -Care Coordinator	• Inner Setting: Work Infrastructure "I do know that we started an at-home monitoring system during COVID. Originally, it was designed just to be kind of post-COVID hospitalizations, but then the hospital expanded it to heart failure and COPD." -Physician • Individuals: Innovation Recipients "But I bought myself an oximeter because you can't see the numbers of your oxygen unless you've got a machine to show you the numbers. I'm the guy that wants to know when I can't breathe, what's my number now?" -Patient

[a]REVISITS: Respiratory Emergency Visits Using Implementation Science Interventions Tailored to Settings.

[b]TOC: transitions of care.

[c]CFIR: Consolidated Framework for Implementation Research.

[d]RT: respiratory therapist.

[e]SL: site lead.

[f]COPD: chronic obstructive pulmonary disease.

[g]RN: registered nurse.

[h]OL: organizational leader.

[i]ED: emergency department.

The most prominent facilitators included having a dedicated staff or dedicated team (eg, COPD educator, pulmonary navigator, nurse education team, and TOC pharmacist), existing infrastructure (eg, transition of care clinic, smoking cessation clinic, and electronic health record), and ongoing quality improvement (QI) initiatives and protocols (eg, advanced care planning initiative, virtual pulmonary rehabilitation, and telehealth program).

Patient perspectives often aligned with those of clinicians, especially regarding challenges with follow-up visits and pulmonary rehabilitation. Both groups cited lack of clinician availability and transportation as key barriers. At one site, both groups emphasized the benefit of a dedicated pulmonary navigator. However, there was some disagreement around the utility of printed materials. Some patients and clinicians questioned their usefulness (eg, due to poor eyesight or redundancy), while others endorsed them, particularly the COPD action plan, which some patients kept visible at home and some clinicians described as efficient. These differing perspectives were discussed during intervention selection with site leads.

Interventions Selected for Implementation by Site Leads

Inhaler education was the most frequently selected intervention for implementation, chosen by nearly all sites (20/21). Other commonly selected interventions included posthospital or ED follow-up visits (13/21) and COPD action plan (12/21). Fewer sites chose to implement medication reconciliation (9/21), pulmonary rehabilitation (6/21), patient navigation (6/21), COPD education (4/21), and advanced care planning (4/21). No sites chose advanced care planning, spirometry, and remote monitoring and reporting. While some participants listed spirometry and remote monitoring and reporting as priorities, site leads were discouraged from selecting these interventions as they could not be randomized to virtual or in-person delivery. Site leads were encouraged by the study team to include these

XSL•FO
RenderX

as supplementary interventions if they also selected 2 eligible interventions compatible with the study design.

Overall, intervention selections aligned with the priorities expressed by site participants (Tables 3 and 5). In total, 5 of the 6 interventions ranked as priorities were chosen for implementation. An exception is COPD education, which was ranked high as a priority but was less frequently selected for implementation.

Table . Interventions ranked as high priority by 194 multistakeholder interview participants (site leads, clinicians, organizational leaders, and patients or caregivers) across 21 US hospitals in the pre-implementation phase of the Reduce REVISITS[a] study. Values represent the number and percentage of participants within each participant type who identified the intervention as a high priority.

Intervention	All (n=194), n (%)	Site leads (n=42), n (%)	Organizational leaders (n=29), n (%)	Clinician (n=105), n (%)	Patients+caregivers (n=18), n (%)
Post-ED[b] follow-up visit	90 (46.39)	33 (78.57)	12 (41.38)	39 (37.14)	6 (33.33)
Inhaler education	78 (40.21)	27 (64.29)	7 (24.14)	39 (37.14)	5 (27.78)
COPD[c] education	66 (34.02)	9 (21.43)	10 (34.48)	40 (38.1)	7 (38.88)
Pulmonary rehabilitation	66 (34.02)	24 (57.14)	4 (13.79)	31 (29.52)	7 (38.88)
COPD action plan	61 (31.44)	9 (21.43)	4 (13.79)	45 (42.86)	3 (16.67)
Patient navigation	60 (30.93)	26 (61.9)	8 (27.59)	20 (19.05)	6 (33.33)
Medication reconciliation	57 (29.38)	19 (45.24)	5 (17.24)	32 (30.48)	1 (5.56)
Smoking cessation education or treatment	53 (27.32)	20 (47.62)	4 (13.79)	28 (26.67)	1 (5.56)
Remote monitoring and reporting	27 (13.92)	7 (16.67)	2 (6.9)	17 (16.19)	1 (5.56)
Advanced care planning	15 (7.73)	2 (4.76)	2 (6.9)	10 (9.52)	1 (5.56)
Spirometry	14 (7.22)	0 (0)	0 (0)	14 (13.33)	0 (0)

[a]REVISITS: Respiratory Emergency Visits Using Implementation Science Interventions Tailored to Settings.

[b]ED: emergency department.

[c]COPD: chronic obstructive pulmonary disease.

Discussion

Principal Findings

This study integrated HCD and IS methods to conduct multiparticipant pre-implementation contextual assessments at 21 hospitals. The most frequently prioritized interventions were mostly related to post-discharge care (eg, post-discharge follow-up visits, pulmonary rehabilitation, and post-discharge patient navigation) and patient education (eg, inhaler education, general COPD education, and COPD action plan). Barriers to implementation occurred at the patient level (social and physical challenges), clinician level (resources: time, staffing, and training), and system (lack of standardization and/or resources to support COPD TOC programs). Key facilitators included harnessing available resources and activities, such as having dedicated staff, using existing infrastructure, and partnering with ongoing QI initiatives. Importantly, intervention selections by site leads generally aligned with the priorities identified by other participants at the same site, suggesting strong participant engagement and feasibility.

These findings align with previous studies identifying education, medication management, and structured follow-up as central components of effective COPD TOC programs. This study extends the existing literature by demonstrating the value of integrating HCD with IS methods as a structured, replicable approach for translating contextual data into actionable implementation decisions during the pre-implementation phase. The integration of HCD and IS enabled the identification of both universal priorities and site-specific constraints that directly informed intervention selection and implementation planning. The inclusion of patients and caregivers allowed the identification of accessibility barriers that are often overlooked when relying solely on clinician or administrative perspectives. Additionally, the use of rapid analytic techniques enabled timely translation of qualitative data into site-specific reports, demonstrating a pragmatic approach for supporting real-world implementation planning in multisite QI initiatives.

Variation in intervention prioritization across participant groups highlights the importance of including multiple perspectives in intervention planning. Site leads' reliance on existing staff and infrastructure as facilitators also aligns with IS frameworks emphasizing adaptability, compatibility, and resource fit [24-28]. These findings underscore the value of blending HCD and IS in identifying contextual barriers and facilitators.

This study has several limitations. While there was considerable geographic diversity among sites, there was limited

XSL•FO
RenderX

representation of hospitals in the Western region of the United States compared with the other regions, and community hospitals were slightly less represented compared with academic hospitals; notably, no Veterans Affairs hospitals enrolled. Patient and caregiver recruitment was limited due to reliance on site leads for recruitment and low technology access among patients. Additionally, the use of snowball sampling may have overrepresented clinicians already engaged in QI, and nearly half of the clinician participants were physicians, potentially limiting perspectives from other health professionals. Due to time constraints, such that data analysis was needed to move to the implementation phase, we used a rapid analytic approach to summarize site-specific interview data. We implemented several quality assurance strategies, including standardized note-taking templates, a uniform analytic process, and principal investigator (PI) review of all transcripts, to preserve analytic rigor. However, no formal coding of transcripts was conducted, and the majority of the analysis was performed by a single analyst, which may introduce interpretive bias.

Another important potential source of bias is regarding the COPD TOC interventions. While the study team used rigorous methods to identify evidence-based COPT TOC interventions, including reviewing the literature (PubMed and Google Scholar), reviewing scoping and systematic reviews [29-31], and team-based expertise (eg, PI was lead or senior author on evidence reports) [16,17], it is possible that interventions were missed. However, participants could prioritize any intervention when completing the interviews. In addition, bias related to the intervention chosen may have been introduced during the final site lead interview or discussion. The information regarding summaries of the evidence supporting COPD TOC interventions was presented by the study PI during these implementation planning discussions, and hence, this summary and any additional input from the PI may have influenced intervention selection. For example, inhaler education has been an area of previous research led by the PI, which may have influenced site lead decision-making. However, concern for this limitation is mitigated by the additional data that inhaler education was consistently identified as a top priority across participant groups during the interviews before this session with the PI, and the PI was not present for any of the participant interviews. This suggests that prioritization of inhaler education and other interventions largely reflected the participants' perspectives rather than investigator influence. Additionally, although COPD education was frequently identified as a priority by participants, it was selected less often by site leads. This may reflect the study team's guidance to prioritize specific, measurable education-related interventions or the tendency for site leads to bundle multiple education components together. However, most interventions were chosen by site leads by integrating prespecified priorities in conjunction with site-level feasibility.

Conclusion

Integrating HCD and IS methods during the pre-implementation phase can enhance the relevance and feasibility and possibly the sustainability of hospital-based COPD TOC interventions. Perspectives from patients, clinicians, and organizational leaders help identify both universal priorities and site-specific constraints, enabling tailored program design. These findings illustrate a replicable and scalable model for designing contextually grounded TOC programs in COPD and other chronic conditions. Future work will assess whether these pre-implementation insights predict implementation fidelity, reach, and sustainability as well as patient outcomes across sites in subsequent phases of the Reduce REVISITS study.

Acknowledgments

The authors would like to thank all sites and participants in the Reduce REVISITS (Respiratory Emergency Visits Using Implementation Science Interventions Tailored to Settings) study.

The Reduce REVISITS study team would like to acknowledge and honor Jean Rommes, our dedicated Patient Advocate, who passed away before publication of this manuscript. Jean would have been a co-author had she been able to complete the authorship forms prior to passing. In lieu of co-authorship we wanted to highlight the very important role she played with this study as a member of our study team. Jean provided essential insight into patient experiences and needs during the development and implementation of this research program and dissemination of its findings, and her perspective meaningfully shaped our approach to evaluating and implementing COPD transitions of care.

Funding

This work was supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health (R01HL146644 and K24HL163408). The funder had no involvement in the study design, data collection, analysis, interpretation, or writing of the manuscript.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the data being qualitative but are available from the corresponding author on reasonable request.

Authors' Contributions

MA: Data Collection and Analysis, Writing – Original Draft, Writing – Review & Editing.
KE: Data Collection and Analysis, Writing – Editing.

XSL•FO
RenderX

LT: Data Collection and Analysis, Writing – Editing.
HP: Data Collection, Writing – Editing.
LD: Data Collection and Analysis, Writing – Editing.
VGP: Conceptualization, Methodology, Writing – Review & Editing
The authors attest that no generative artificial intelligence tools were used in the writing, editing, or analysis of this manuscript.

Conflicts of Interest

VGP reports receiving funding from the National Institutes of Health (K24HL163408 and R01HL146644) and AHRQ (R01AS027804). VGP also reports receiving consultation fees from Humana. All other authors have no disclosures. We previously presented these findings as poster presentations at the International American Thoracic Society meeting in May 2025.

Multimedia Appendix 1
Site lead survey.
[DOCX File, 513 KB - humanfactors_v13i1e82078_app1.docx]

Multimedia Appendix 2
Example site implementation plan workbook.
[DOCX File, 93 KB - humanfactors_v13i1e82078_app2.docx]

Multimedia Appendix 3
Participant demographics of hospitalized adults with chronic obstructive pulmonary disease across sites participating in the Reduce REVISITS (Respiratory Emergency Visits Using Implementation Science Interventions Tailored to Settings) study. Values represent aggregated demographic variables provided by sites with available data during the baseline period (n=10,697) and the implementation period (n=2,745). Demographic indicators include age, gender identity, race, ethnicity, insurance coverage, and smoking status. Only sites that submitted demographic data are included.
[DOCX File, 18 KB - humanfactors_v13i1e82078_app3.docx]

References

1. Chronic obstructive pulmonary disease (COPD) includes: chronic bronchitis and emphysema. National Center for Health Statistics. URL: https://www.cdc.gov/nchs/fastats/copd.htm [accessed 2026-01-25]
2. Global initiative for chronic obstructive lung disease. : Global Initiative for Chronic Obstructive Lung Disease - GOLD; 2017.
3. Kumbhare SD, Beiko T, Wilcox SR, Strange C. Characteristics of COPD patients using United States Emergency Care or hospitalization. Chronic Obstr Pulm Dis 2016 Mar 28;3(2):539-548. [doi: 10.15326/jcopdf.3.2.2015.0155] [Medline: 28848878]
4. Hasegawa K, Tsugawa Y, Tsai CL, Brown DFM, Camargo CA. Frequent utilization of the emergency department for acute exacerbation of chronic obstructive pulmonary disease. Respir Res 2014 Apr 10;15(1):40. [doi: 10.1186/1465-9921-15-40] [Medline: 24717062]
5. Yeatts KB, Lippmann SJ, Waller AE, et al. Population-based burden of COPD-related visits in the ED. Chest 2013 Sep;144(3):784-793. [doi: 10.1378/chest.12-1899]
6. Anzueto A. Impact of exacerbations on COPD. Eur Respir Rev 2010 Jun;19(116):113-118. [doi: 10.1183/09059180.00002610] [Medline: 20956179]
7. Hospital Readmissions Reduction Program. Centers for Medicare & Medicaid Services. URL: https://www.cms.gov/medicare/payment/prospective-payment-systems/acute-inpatient-pps/hospital-readmissions-reduction-program-hrrp [accessed 2026-01-25]
8. Bhatt SP, Wells JM, Iyer AS, et al. Results of a Medicare bundled payments for care improvement initiative for chronic obstructive pulmonary disease readmissions. Ann Am Thorac Soc 2017 May;14(5):643-648. [doi: 10.1513/AnnalsATS.201610-775BC] [Medline: 28005410]
9. Waltman A, Konetzka RT, Chia S, et al. Effectiveness of a bundled payments for care improvement program for chronic obstructive pulmonary disease. J Gen Intern Med 2023 Sep;38(12):2662-2670. [doi: 10.1007/s11606-023-08249-6] [Medline: 37340256]
10. Press VG, Arora VM, Trela KC, et al. Effectiveness of interventions to teach metered-dose and Diskus inhaler techniques. A randomized trial. Ann Am Thorac Soc 2016 Jun;13(6):816-824. [doi: 10.1513/AnnalsATS.201509-603OC] [Medline: 26998961]
11. Thomas RM, Locke ER, Woo DM, et al. Inhaler training delivered by internet-based home videoconferencing improves technique and quality of life. Respir Care 2017 Nov;62(11):1412-1422. [doi: 10.4187/respcare.05445] [Medline: 28720676]
12. Locke ER, Thomas RM, Woo DM, et al. Using video telehealth to facilitate inhaler training in rural patients with obstructive lung disease. Telemed J E Health 2019 Mar;25(3):230-236. [doi: 10.1089/tmj.2017.0330] [Medline: 30016216]

XSL•FO
RenderX

13. Mueller SK, Kripalani S, Stein J, et al. A toolkit to disseminate best practices in inpatient medication reconciliation: multi-center medication reconciliation quality improvement study (MARQUIS). Jt Comm J Qual Patient Saf 2013 Aug;39(8):371-382. [doi: 10.1016/s1553-7250(13)39051-5] [Medline: 23991510]
14. Hansen LO, Greenwald JL, Budnitz T, et al. Project BOOST: effectiveness of a multihospital effort to reduce rehospitalization. J Hosp Med 2013 Aug;8(8):421-427. [doi: 10.1002/jhm.2054] [Medline: 23873709]
15. Lindenauer PK, Pekow P, Gao S, Crawford AS, Gutierrez B, Benjamin EM. Quality of care for patients hospitalized for acute exacerbations of chronic obstructive pulmonary disease. Ann Intern Med 2006 Jun 20;144(12):894-903. [doi: 10.7326/0003-4819-144-12-200606200-00006] [Medline: 16785478]
16. Press VG, Au DH, Bourbeau J, et al. Reducing chronic obstructive pulmonary disease hospital readmissions. An official American Thoracic Society Workshop report. Ann Am Thorac Soc 2019 Feb;16(2):161-170. [doi: 10.1513/AnnalsATS.201811-755WS] [Medline: 30707066]
17. Press VG, Myers LC, Feemster LC. Preventing COPD readmissions under the hospital readmissions reduction program: how far have we come? Chest 2021 Mar;159(3):996-1006. [doi: 10.1016/j.chest.2020.10.008] [Medline: 33065106]
18. Lee CT, Traeger L, Akula M, et al. Virtual mentored implementation to improve care transitions in chronic obstructive pulmonary disease: protocol for a pragmatic implementation study. JMIR Res Protoc 2026 Jan 13;15:e82043. [doi: 10.2196/82043] [Medline: 41529255]
19. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res 2013 Dec;48(6 Pt 2):2134-2156. [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
20. Auerbach AD, Patel MS, Metlay JP, et al. The Hospital Medicine Reengineering Network (HOMERuN): a learning organization focused on improving hospital care. Acad Med 2014 Mar;89(3):415-420. [doi: 10.1097/ACM.0000000000000139] [Medline: 24448050]
21. Respiratory Emergency Visits Using Implementation Science Interventions Tailored to Settings (REVISITS study). : SHM: Society of Hospital Medicine URL: https://www.hospitalmedicine.org/clinical-topics/revisits-study/ [accessed 2026-01-25]
22. Holtzblatt K, Beyer H. Contextual Design: Defining Customer-Centered Systems: Morgan Kaufmann (Elsevier); 1997.
23. Erwin K, Krishnan JA. Redesigning healthcare to fit with people. BMJ 2016 Aug 23;354:i4536. [doi: 10.1136/bmj.i4536] [Medline: 27553644]
24. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implementation Sci 2009 Dec;4(1):50. [doi: 10.1186/1748-5908-4-50]
25. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated Consolidated Framework for Implementation Research based on user feedback. Implement Sci 2022 Oct 29;17(1):75. [doi: 10.1186/s13012-022-01245-0] [Medline: 36309746]
26. Keith RE, Crosson JC, O'Malley AS, Cromp D, Taylor EF. Using the Consolidated Framework for Implementation Research (CFIR) to produce actionable findings: a rapid-cycle evaluation approach to improving implementation. Implement Sci 2017 Feb 10;12(1):15. [doi: 10.1186/s13012-017-0550-7] [Medline: 28187747]
27. Serhal E, Arena A, Sockalingam S, Mohri L, Crawford A. Adapting the Consolidated Framework for Implementation Research to create organizational readiness and implementation tools for Project ECHO. J Contin Educ Health Prof 2018;38(2):145-151. [doi: 10.1097/CEH.0000000000000195] [Medline: 29505486]
28. Birken SA, Powell BJ, Presseau J, et al. Combined use of the Consolidated Framework for Implementation Research (CFIR) and the Theoretical Domains Framework (TDF): a systematic review. Implement Sci 2017 Jan 5;12(1):2. [doi: 10.1186/s13012-016-0534-z] [Medline: 28057049]
29. Prieto-Centurion V, Markos MA, Ramey NI, et al. Interventions to reduce rehospitalizations after chronic obstructive pulmonary disease exacerbations. A systematic review. Ann Am Thorac Soc 2014 Mar;11(3):417-424. [doi: 10.1513/AnnalsATS.201308-254OC] [Medline: 24423379]
30. Freedman N. Reducing COPD readmissions: strategies for the pulmonologist to improve outcomes. Chest 2019 Oct;156(4):802-807. [doi: 10.1016/j.chest.2019.06.005] [Medline: 31590710]
31. Sharpe I, Bowman M, Kim A, Srivastava S, Jalink M, Wijeratne DT. Strategies to prevent readmissions to hospital for COPD: a systematic review. COPD 2021 Aug;18(4):456-468. [doi: 10.1080/15412555.2021.1955338] [Medline: 34378468]

ABBREVIATIONS

BPCI: Bundled Payments for Care Improvement
CFIR: Consolidated Framework for Implementation Research
COPD: chronic obstructive pulmonary disease
ED: emergency department
HCD: human-centered design
HIPAA: Health Insurance Portability and Accountability Act
HRRP: Hospital Readmission Reduction Program
IRB: institutional review board

XSL•FO
RenderX

IS: implementation science
PI: principal investigator
QI: quality improvement
REDCap: Research Electronic Data Capture
REVISITS: Respiratory Emergency Visits Using Implementation Science Interventions Tailored to Settings
TOC: transitions of care

Edited by A Kushniruk; submitted 08.Aug.2025; peer-reviewed by O Nguyen, RG Buhr; revised version received 16.Dec.2025; accepted 02.Jan.2026; published 02.Mar.2026.

Please cite as:
Akula M, Erwin K, Traeger L, Pick H, Gao F, Damschroder L, Press VG
Contextual Assessments for Chronic Obstructive Pulmonary Disease Transition of Care Bundle Implementation Planning for the Reduce REVISITS Study: Rapid Sequential Explanatory Mixed Methods Approach
JMIR Hum Factors 2026;13:e82078
URL: https://humanfactors.jmir.org/2026/1/e82078
doi:10.2196/82078

© Mahima Akula, Kim Erwin, Leah Traeger, Hannah Pick, Fei Gao, Laura Damschroder, Valerie G Press. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 2.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Improving the Efficiency of the Intravenous Medicine Preparation Pathway With an Intravenous Workflow Software Solution in Full-Capacity Pharmacy Units at Watford General Hospital: Observational Study and Economic Analysis

Bethany Umpleby[1*], BS; Nick Hex[1*], BA, CIPFA; Teresa Martins[2*], MAPharmT; Husanain Soori[2*], NVQ

[1]York Health Economics Consortium, York, United Kingdom

[2]West Hertfordshire Hospitals NHS Trust, Watford, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Nick Hex, BA, CIPFA
York Health Economics Consortium
Enterprise House, Innovation Way
1 Enterprise House
York, YO105NQ
United Kingdom
Phone: 44 01904323620
Email: nick.hex@york.ac.uk

Abstract

Background: The existing intravenous systemic anticancer therapy (SACT) pathway in pharmacies is operationally inefficient. Manual, paper-based workflows render the system prone to human error, and the need for time-consuming manual verification diverts pharmacy staff time. The introduction of an automated workflow solution for the intravenous SACT pathway could optimize treatment timeliness and improve oncological outcomes for patients, aligning with the National Health Service Long Term Plan for improved cancer care.

Objective: This observational analysis aimed to assess the change in time, cost, and errors following the implementation of the Becton Dickinson (BD) Cato Pharmacy system in an aseptic unit producing intravenous SACT at Watford General Hospital.

Methods: Data on compounding process times were collected manually by pharmacy staff before and after the implementation of the intravenous compounding software (BD Cato). The data were analyzed to estimate annual time savings, opportunity cost savings, and error reduction.

Results: The intravenous compounding software produced a time saving of 18 (SD 9) minutes per drug, equating to 1034 hours saved per year (1034/2591, 39.9% reduction). If this time were repurposed to producing more intravenous SACT, Watford General Hospital could increase production by 66% (2298/3482) annually (2298 additional intravenous SACT). This represents an average cost saving of £11.29 (£1=US $1.273) per drug, equating to an annual opportunity cost saving of £39,246. The intravenous compounding software also decreased observed errors by 86% (43/50), a reduction of 43 errors over 2 months (approximately 258 fewer errors annually). Staff also preferred the intravenous compounding software to the manual system.

Conclusions: Implementing intravenous compounding software can save time, reduce costs, and lower errors in intravenous SACT preparation. This could improve timely treatment access for patients with cancer.

(JMIR Hum Factors 2026;13:e85408) doi:10.2196/85408

KEYWORDS

BD Cato Pharmacy; intravenous; compounding software; anticancer therapy; efficiency; automation

XSL•FO
RenderX

Introduction

The National Health Service (NHS) Long Term Plan includes targets for the NHS in the United Kingdom, including better care for cancer [1]. NHS England made a case for transforming pharmacy aseptic services, including reducing medical waste to reduce NHS costs [2], and reducing medication errors [3]. As medication errors are approximately 5 times more likely to occur for intravenous medications [4] and can cause significant patient harm [5], there is increasing pressure on aseptic units to reduce this error risk [2,6].

In 2022, the British Oncology Pharmacy Association found that 91% of respondents experienced same-day treatment delays due to delays within their pharmacy's service [7]. Delays in preparation within the pharmacy aseptic unit accounted for 69% of these pharmacy-related delays to intravenous systemic anticancer therapy (SACT) administration. Similarly, 76% of respondents had their treatment rescheduled at least once in 6 months [7]. Two main causes resulting in the rescheduling of treatment were known staffing issues (33%) and aseptic unit capacity (35%). This indicates that the current pathway for the preparation of intravenous medicines in full-capacity aseptic units is struggling to manage these demands.

The traditional pathway for the production of intravenous SACT involves a lengthy process, characterized by a manual paper-based workflow, which is slow, susceptible to errors, and lacks traceability. There is also a heavy reliance on manual verification and supervision to ensure a quality-assured process, which requires substantial pharmacy staff time. As the number of patients requiring first-line chemotherapy is estimated to increase each year [8], an alternative automated solution is required to improve efficiency, increase capacity, reduce costs, and reduce the risk of errors [2,9]. Becton Dickinson (BD) has developed a medication workflow software solution, BD Cato, which offers a potential solution to help organizations address these challenges.

A study from 2021 found that using the BD Cato software in East Tallinn Central Hospital, Estonia, reduced compounding time by 35% and doubled medication production capabilities over the course of the study [10]. BD Cato Pharmacy has also been found to detect errors in 7.89% of prepared cancer drugs, which would not have otherwise been detected using traditional compounding methods [11]. Therefore, the BD Cato Pharmacy software has the potential to save time (and potentially save costs) and to reduce errors, thus helping to meet policy targets.

The aim of this economic analysis was to assess the observed change in time, cost, and error rates associated with the implementation of the BD Cato Pharmacy system within an aseptic unit producing intravenous SACT at Watford General Hospital (WGH), England.

Methods

The Intravenous Compounding Software

BD Cato is a medication workflow solution software that standardizes prescription, compounding, and administration processes along the medication pathway.

BD Cato Pharmacy is the module within this workflow related to the compounding process and provides functionality for the gravimetric and volumetric preparation of prescribed intravenous medication, such as chemotherapy, parenteral nutrition, pain medication, and antibiotics. It supports barcode scanning of vials and other components for electronic verification and documentation, automated calculations, step-by-step on-screen workflow guidance, gravimetric preparation using an electronic scale, image capture of preparation steps via a camera, hard stops when deviations have been detected, inventory and remainder management functionality, and production planning and report generation.

Figure 1 demonstrates the pathway for the preparation of intravenous SACT without the use of BD Cato. Figure 2 demonstrates the pathway for the preparation of intravenous SACT with the use of BD Cato.

XSL•FO
RenderX

Figure 1. Current pathway for the preparation of intravenous medications for cancer.

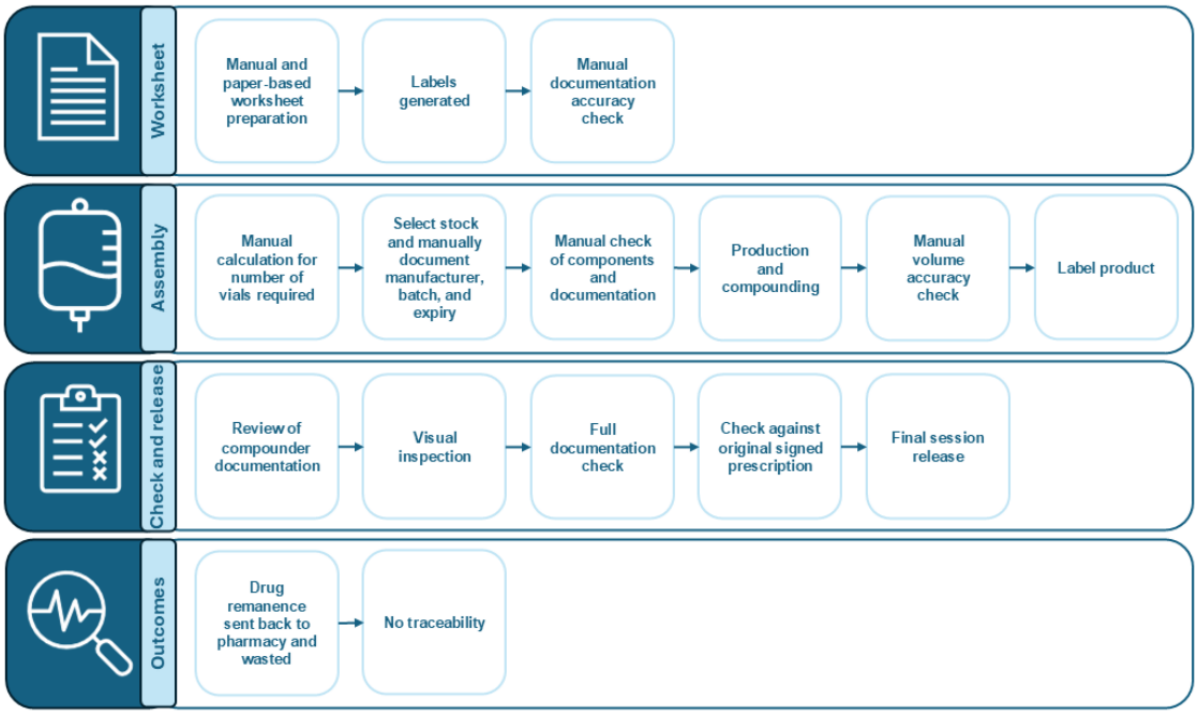
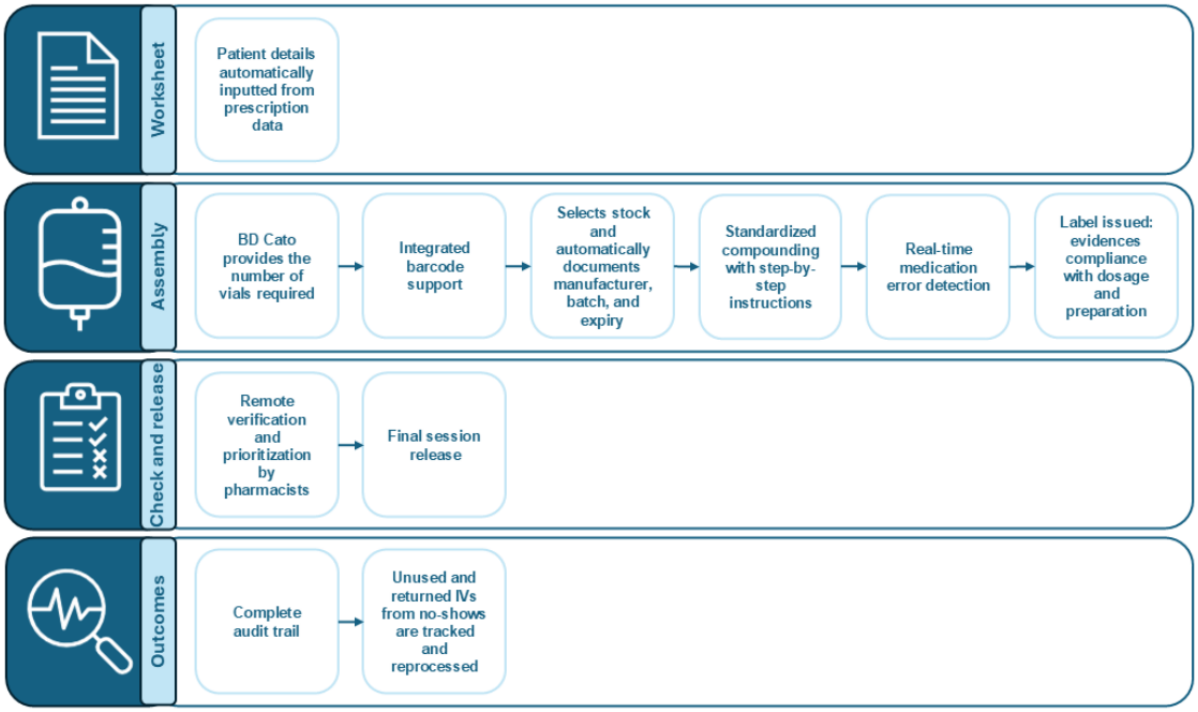


Figure 2. The Becton Dickinson (BD) Cato Pharmacy pathway for the preparation of intravenous (IV) systemic anticancer therapy for cancer in this study.



Study Design

This was a single-center before-and-after observational study. Data collection was carried out in the pharmacy department at WGH, which had decided to adopt the BD Cato system, and was interested in understanding the impact on time, costs, and error rates. The period for data collection before the introduction of the intravenous compounding software was approximately 10 weeks, between March 2021 and May 2021. The period for data collection after the introduction of the intravenous compounding software was approximately 8 weeks, between December 2023 and January 2024.

A data collection sheet was produced in collaboration with a qualified pharmacist at WGH to record the metrics shown in Textbox 1.

Pharmacy staff at WGH were asked to manually complete data collection sheets for every intravenous SACT that was compounded within the period before and after the implementation of the intravenous compounding software.

XSL•FO
RenderX

Before and after the implementation of the intravenous compounding software, manual data collection methods were used, whereby pharmacy staff were asked to record the time it took them to complete different aspects of the manufacturing process on a paper form. These were collated and analyzed by the York Health Economics Consortium.

The time to compound and reconstitute intravenous SACT and the number and types of errors before the study were unknown. This made sample size calculation difficult to determine. A period of 8 to 10 weeks was chosen to allow sufficient numbers of intravenous SACT compounding data to be collected and to be able to accurately measure the time taken for the compounding of different classes of intravenous SACT. It also provided time to record any errors that occurred during the data collection periods.

Textbox 1. Summary of primary metrics and parameters captured during the data collection process.

Metrics recorded in the data collection sheet

- The intravenous systemic anticancer therapy drug name being compounded
- The time taken for each task
  - Worksheet completion
  - Worksheet checking
  - Assembly
  - Assembly checking
  - Compounding
  - Labeling
  - Check and release
- Any identified errors
  - Wrong stability entered
  - Wrong diluent or fluid
  - Wrong volume drawn in compounding
  - Missing assembly item
  - Wrong batch number (assembly)
  - Wrong protocol
  - Wrong vial size chosen
  - Wrong volume entered
  - Wrong dose inputted
  - Wrong dosage form
  - Wrong or no ward
  - Missing or wrong infusion time
  - Other (please state)

Data from the data collection sheets were manually input into Google Sheets to clean and store the data and perform basic data analysis. Once data collection was complete, the data were exported, and analysis was performed in Microsoft Excel.

The primary outcome of the study was the impact of the intravenous compounding software on the time taken to compound intravenous SACT. The secondary outcomes were the impact of the intravenous compounding software on direct costs and the number of errors made during the compounding of intravenous SACT.

In addition to the quantitative analysis, 2 surveys were distributed among pharmacy staff members: 1 before the implementation of the intravenous compounding software and 1 after the implementation of the intravenous compounding software (18 months apart). The objective was to assess staff perception of the BD Cato workflow compared to the previous manual process with regard to safety, accuracy, efficiency, visibility of production status, work satisfaction, and overall preference. It was also used to evaluate staff satisfaction with the introduction of the new system.

Ethical Considerations

This study did not require ethics approval as the analysis was based on a system to automate pharmacy processes and did not involve human participants. Deidentified data on pharmacy processes were retrieved from WGH via staff surveys. West

XSL•FO
RenderX

Hertfordshire Teaching Hospitals NHS Trust provided access and permission to reuse the dataset for this study.

Results

Overview

Preimplementation and postimplementation data for 12 drug types were available from the intravenous compounding software. These drugs are presented in Table 1. Within the preimplementation data, there were an additional 3 drug types that were not present in the postimplementation data: brentuximab, cisplatin, and cytarabine. Within the postimplementation data, there were an additional 5 drug types that were not present in the preimplementation data: bleomycin, carfilzomib, tebentafusp, tocilizumab, and vinblastine. The preimplementation data comprised 408 total preparations, and the postimplementation data comprised 345 total preparations. Average time savings per drug, average time savings per drug per year, and average staff cost savings per drug were calculated. The activities involved in compounding were grouped into production, verification, and check and release. Production consisted of worksheet creation, assembly, compounding, and labeling tasks. Verification consisted of worksheet checking and assembly checking tasks. Check and release consisted of an accuracy and final release check to ensure the medication has been prepared correctly and is suitable for administration.

Table 1. Average time saved.

Drug	Production, n	Verification, n	Check and release, n	Total, n
Average time saved per drug (min), mean (SD)				
All: nondrug-specific	12 (0.24)	5 (0.16)	1 (0.15)	18 (0.55)
Azacitidine	7 (0.42)	5 (0.21)	0 (0.32)	12 (0.94)
Bendamustine	16 (1.23)	6 (0.33)	–1 (0.83)	21 (2.39)
Bortezomib	14 (0.79)	5 (0.31)	3 (0.51)	21 (1.61)
Cyclophosphamide	4 (0.65)	6 (0.37)	0 (0.42)	9 (1.45)
Dacarbazine	6 (1.44)	6 (0.79)	2 (0.97)	14 (3.20)
Doxorubicin	7 (0.62)	4 (0.26)	2 (0.35)	13 (1.22)
Isatuximab	18 (2.37)	6 (1.04)	2 (1.21)	26 (4.62)
Obinutuzumab	9 (0.91)	7 (0.86)	–2 (0.73)	14 (2.49)
Polatuzumab	32 (2.78)	4 (1.22)	1 (1.32)	37 (5.32)
Rituximab (truxima)	10 (0.60)	4 (0.21)	1 (0.39)	14 (1.20)
Vincristine	8 (0.91)	8 (0.77)	1 (0.51)	17 (2.19)
Vincristine P	28 (1.38)	7 (0.54)	2 (1.10)	37 (3.01)
Average time saved per year (h)[a]				
All: nondrug-specific	696	293	45	1034
Azacitidine	69	48	4	121
Bendamustine	67	25	–3	89
Bortezomib	123	40	21	184
Cyclophosphamide	19	29	–2	45
Dacarbazine	8	9	3	20
Doxorubicin	28	15	7	49
Isatuximab	31	11	4	46
Obinutuzumab	22	17	–5	33
Polatuzumab	68	8	2	78
Rituximab (truxima)	68	24	5	97
Vincristine	11	11	1	23
Vincristine P	49	12	3	65

[a]The average time multiplied by the amount of activity (SD could not be calculated).

XSL•FO
RenderX

Time Outcome Results

Time outcomes were examined by calculating the difference between the time taken to complete each activity type in the drug compounding process before and after the implementation of the intravenous compounding software.

Table 1 presents the average time savings per drug compounded in minutes after the introduction of the intravenous compounding software. It presents a nondrug-specific weighted average time saving for each activity group, which included drugs only in the preimplementation phase, drugs only in the postimplementation phase, and drugs included in both data collection phases. It also presents the average time savings for each of the 12 drugs that had both preimplementation and postimplementation data.

Table 1 also presents the average time savings per year in hours. These values were calculated by multiplying the average time saved per drug by the number of each drug compounded per year. This is split into an average nondrug-specific time saving per year and individual time savings per drug per year.

Table 1 shows an average nondrug-specific time saving of 18 (SD 0.55) minutes per drug when using the intravenous compounding software. Time savings were observed for production, verification, and check and release tasks for all drugs, except for check and release for bendamustine and obinutuzumab. However, this did not influence the total average time saved, as all drugs demonstrated a total average time saving. The greatest time saving was observed for production (worksheet, assembly, compounding, and labeling tasks) and for polatuzumab and vincristine P.

Table 1 also shows an average nondrug-specific time saving of 1034 hours per year when using the intravenous compounding software. This represents a 39.9% (1034/2591) reduction in compounding time. This was calculated by multiplying the weighted average time saved per drug (hours) by the 3476 drugs compounded annually by WGH. The greatest time saving for an individual drug was for bortezomib. The individual drug-specific time savings per year were calculated using the number of each drug produced annually at WGH.

Table 2 presents the average time savings per drug per day.

Table 2. Average time saved per drug per day (minutes).

Day	Production, mean (SD)	Verification, mean (SD)	Check and release, mean (SD)	Total, mean (SD)
Monday	16 (0.60)	6 (0.36)	1 (0.33)	24 (1.28)
Tuesday	13 (0.57)	5 (0.20)	0 (0.34)	19 (1.10)
Wednesday	8 (0.57)	7 (0.54)	2 (0.35)	17 (1.45)
Thursday	6 (0.47)	5 (0.18)	0 (0.26)	10 (0.92)
Friday	9 (0.56)	5 (0.21)	1 (0.35)	15 (1.13)

Table 2 shows that when using the intravenous compounding software, there were observed time savings on every day of the week. These time savings could not be weighted as data on the number of drugs compounded each day were not available.

Opportunity Cost Savings

Opportunity cost savings were calculated by multiplying the average time saved by the average staff cost per hour for each task. The average staff cost per hour was based on the staff bands responsible for each task, taken from the Personal Social Services Research Unit [12]. Correspondence with WGH established that production tasks were carried out by staff in bands 3, 4, and 5; verification tasks by staff in bands 5, 6, and 7; and check and release tasks by staff in bands 7 and 8.

The average cost saving per drug and the average cost saving per year are presented in Table 3. This is split into a weighted average nondrug-specific cost saving, which included drugs only in the preimplementation phase, drugs only included in the postimplementation phase, and drugs included in both data collection phases. It also presents individual cost savings per drug for drugs that had both preimplementation and postimplementation data.

XSL•FO
RenderX

Table 3. Average cost saving (£1=US $1.273).

Drug	Production (£)	Verification (£)	Check and release (£)	Total (£)
Average cost saving per drug[a]				
All: nondrug-specific	5.98	4.22	1.10	11.29
Azacitidine	3.50	4.11	0.59	8.20
Bendamustine	8.10	4.99	–1.10	11.99
Bortezomib	7.14	3.85	3.53	14.52
Cyclophosphamide	1.87	4.79	–0.57	6.09
Dacarbazine	2.96	5.33	2.51	10.80
Doxorubicin	3.72	3.45	2.50	9.67
Isatuximab	8.71	5.28	3.28	17.27
Obinutuzumab	4.40	5.71	–3.03	7.08
Polatuzumab	16.04	3.14	1.62	20.80
Rituximab (truxima)	4.94	2.93	0.96	8.83
Vincristine	4.12	6.59	1.34	12.05
Vincristine P	14.13	5.78	2.72	22.62
Average cost saving per year[a]				
All: nondrug-specific	20,773	14,661	3811	39,246
Azacitidine	2052	2412	349	4813
Bendamustine	2009	1238	–273	2973
Bortezomib	3661	1977	1811	7449
Cyclophosphamide	561	1436	–171	1826
Dacarbazine	249	448	211	907
Doxorubicin	821	763	553	2137
Isatuximab	923	559	348	1831
Obinutuzumab	643	833	–442	1033
Polatuzumab	2021	396	205	2621
Rituximab (truxima)	2041	1208	398	3647
Vincristine	329	527	107	964
Vincristine P	1469	601	283	2353

[a]The average time multiplied by the amount of activity (SD could not be calculated).

Table 3 shows a weighted average nondrug-specific total cost saving of £11.29 (£1=US $1.273) per drug. Table 3 also shows a weighted average nondrug-specific total cost saving of £39,246 per year. The use of the intravenous compounding software was observed as cost saving for all production tasks, all verification tasks, and most check and release tasks, excluding bendamustine, cyclophosphamide, and obinutuzumab, which were cost incurring. Nevertheless, this did not have an impact on the total cost savings of these drugs, as all drugs resulted in a total opportunity cost saving. The greatest opportunity cost savings per year were observed for production tasks and for bortezomib.

Error Outcome Results

The reduction in the number of observed errors from the use of the intravenous compounding software was calculated by finding the difference between the number of errors reported before and after its introduction. These results are presented in Table 4. While the intravenous compounding software anticipates potential errors to proactively implement mitigations and prevent process inefficiencies, this analysis focused solely on observed errors that directly resulted in process inefficiencies.

XSL•FO
RenderX

Table 4. Number of errors saved.

Error summary	Before the use of BD[a] Cato Pharmacy software (n=50), n (%)	After the use of BD Cato Pharmacy software (n=7), n (%)	Difference (n=43), n (%)
Wrong stability entered	0 (0)	0 (0)	0 (0)
Wrong diluent or fluid	0 (0)	0 (0)	0 (0)
Wrong volume drawn in compounding	3 (6)	0 (0)	3 (7)
Missing assembly item	27 (54)	1 (14)	26 (60)
Wrong batch number (assembly)	8 (16)	1 (14)	7 (16)
Wrong protocol	1 (2)	0 (0)	1 (2)
Wrong vial size chosen	5 (10)	1 (14)	4 (9)
Wrong volume entered	0 (0)	0 (0)	0 (0)
Wrong dose inputted	0 (0)	0 (0)	0 (0)
Wrong dosage form	0 (0)	0 (0)	0 (0)
Wrong or no ward	1 (2)	0 (0)	1 (2)
Missing or wrong infusion time	0 (0)	0 (0)	0 (0)
Other	5 (10)	4 (57)	1 (2)

[a]BD: Becton Dickinson.

Table 4 shows that the intravenous compounding software decreased the risk of observed errors by 86% (43/50). There was a reduction of 43 observed errors over a 2-month period, which equates to reducing approximately 258 observed errors per year. The greatest reduction in errors was for missing assembly items. The intravenous compounding software reduced errors through barcode scanning, automatic documentation, vial size suggestion, and guided workflows, preventing staff from proceeding to the next step until the current step was correctly completed. There were 9 reported cases of the intravenous compounding software losing signal, which required staff to log out and log back in to continue preparation. While these were not reported as errors, as they did not result in loss of data or require a workflow step to be repeated, they resulted in an increase in time taken to prepare the drug, which is accounted for in the time and opportunity cost saving analysis.

Staff Survey Results

The preimplementation survey had a response rate of 93% (13/14), and the postimplementation survey had a response rate of 100% (8/8). A total of 88% (7/8) of postimplementation respondents preferred to compound a product using the intravenous compounding software and perceived it as being safer, more accurate, and more efficient compared to a manual workflow. In total, 100% (8/8) of respondents believed that the intravenous compounding software facilitates the tracking of sterile products during the compounding process. Regarding work satisfaction, none of the respondents found work dissatisfying using the intravenous compounding software, compared to 8% (1/13) of respondents who found work dissatisfying using the previous manual workflow.

Before receiving training on the use of the intravenous compounding software, preimplementation staff were asked an open-ended question about what they considered to be its potential advantages. The most common response was improvement in process efficiency, with 1 respondent stating that freeing up time would allow more focus on clinical roles related to oncology. The second most common responses were reduction of errors and improved accuracy. Traceability, removing the need for paperwork, waste reduction, and cost savings were also perceived as advantages. One respondent also perceived that the new system would upskill staff in the use of technology.

Staff in the preimplementation survey were also asked an open-ended question on what they would consider the likely challenges or disadvantages of introducing the intravenous compounding software. The most common response was related to overcoming resistance to change and having to adopt new ways of working due to familiarity with old processes and procedures. The second most common response was training requirements, with some respondents acknowledging that it can be a challenge for staff that are lacking in IT skills to adapt to new technology. Other perceived challenges were the need for a downtime plan and new types of errors that could be introduced.

Discussion

Principal Findings

In this observational study, the use of an intravenous compounding software reduced time, costs, and errors during the production of intravenous SACT at WGH. It was estimated that the use of the intravenous compounding software would save 1034 hours of pharmacy staff time per year. If this time were all repurposed toward the production of intravenous SACT, an additional 2298 drugs could be compounded at WGH each year. This is an increase of approximately 66.11% (2298/3476) from the number of intravenous SACT currently being produced by WGH per year (n=3476). If half of the time saved (517 hours)

XSL•FO
RenderX

was repurposed toward the production of intravenous SACT, an additional 1149 drugs could be compounded at WGH each year. It was observed that the use of an intravenous compounding software reduced the time to produce, verify, and check and release intravenous SACT by 39.9% (1034/2591). These results are in line with previous research that also found that the intravenous compounding software produced a 35% reduction in compounding time [10].

On the basis of the time saved and the average staff cost to the NHS, the observed time saved using the intravenous compounding software can be valued at £39,246 per year. This figure can be viewed as an opportunity cost saving, as the time saved could be reinvested into producing more intravenous SACT without incurring additional labor costs. The opportunity cost of an intervention is what is foregone as a consequence of adopting a new intervention [13].

There was some heterogeneity in the observed time savings between different drugs, partly driven by the number of steps in the manual manufacturing process for certain drugs. Bendamustine, in particular, is a complex product to make and typically involves more vials to reconstitute and larger volumes to be measured and then transferred to the infusion bag. As there are more steps in the process, the pharmacist review contains more photos of the steps in the compounding process for the releasing officer to review, and therefore, it would take longer to complete.

The greatest time savings per drug were on a Monday, and the least time savings per drug were on a Thursday. One explanation for this is that different levels of staff may work on different days throughout the week. Another explanation is that the number and types of drugs compounded per day may differ; for example, drugs that are slower to prepare may be compounded on a specific day.

The use of an intravenous compounding software was also shown to decrease the occurrence of errors requiring activity to be redone by 86% (43/50), with the potential to reduce 258 errors per year at WGH. This is in line with previous research that also reported the error-reducing capabilities of BD Cato Pharmacy [11]. It is of interest that there were 0 diluent errors before implementation. This is at odds with a survey carried out by the Institute for Safe Medication Practices, which found that more than half of pharmacists and pharmacist technicians who responded could identify with certainty which drugs, diluents, and volumes were used when verifying compounded sterile preparations [14]. WGH solely used volumetric preparation before implementation, but after implementation, a mix of both volumetric and gravimetric preparations was used, with most products made gravimetrically.

Therefore, the introduction of an intravenous compounding software could contribute toward meeting NHS targets that are currently unmet. For example, the British Oncology Pharmacy Association survey found that one of the main causes of same-day treatment delays was delays within the pharmacy aseptic unit (69%), and one of the main causes of rescheduled treatments was limited aseptic unit capacity (35%) [7]. Therefore, if more intravenous SACT can be compounded more quickly using an intravenous compounding software within aseptic units, same-day treatment delays and rescheduled treatments could be reduced. This could also help to manage the predicted increase in treatments required in the coming years [8].

The implementation of the intravenous compounding software was well received by staff who completed the postimplementation survey. The preimplementation open-ended question on perceived challenges highlights the importance of getting staff on board with the change at an early stage and taking steps to overcome resistance. Sufficient time and resources should also be allocated to training to support the effective use of the system and help overcome the concerns and resistance that staff have.

Limitations

This analysis has several limitations. Because the data were collected manually, there is an increased risk of human error when completing the data collection sheet, as well as potential inconsistencies in how individuals report different data items.

There are also potential confounding variables that may have influenced the findings of the study. Data were collected at 2 different time points, with preimplementation data collected in 2021 and postimplementation data collected in 2023 to 2024. Factors that may have influenced the data collection process between 2021 and 2023 to 2024 include changes in protocols or procedures, changes in staff (including staff competency), and changes in resources other than the intravenous compounding software. Similarly, data were collected over short periods, which might not capture long-term effects or trends. Longitudinal effects or outcomes might differ from these short-term findings.

A similar limitation applies to the staff survey. Preimplementation survey responses were collected in 2021, while postimplementation survey responses were collected between 2022 and 2024. Due to staff turnover, respondents may differ before and after the implementation of the intravenous compounding software. It is also uncertain whether all respondents had experience with compounding using the previous manual process.

The study was not intended to be a return on investment analysis. The time saved has been quantified in financial terms to provide a context for the benefits of automation. Consequently, the investment required for the installation and maintenance of the system was not taken into consideration when savings were calculated. Similarly, no calculation was made of the investment WGH would have required to achieve comparable productivity gains had it retained the previous manual workflow.

In response to the limitations of this study, future research could address the following areas. The economic impact of BD Cato Pharmacy across other aseptic units within the NHS, and for different types of drugs (eg, intravenous antibiotics or nutrition), should be investigated. This would be beneficial to assess the broader applicability of the results. In addition, future research could investigate the cost impact of managing drug remnants and drug waste using BD Cato Pharmacy's remnants functionality.

XSL•FO
RenderX

Conclusions

The aim of this economic analysis was to assess the observed change in time, cost, and error rates associated with the implementation of an intravenous compounding software system within an aseptic unit producing intravenous SACT at WGH. BD Cato Pharmacy has the capability to reduce time, costs, and errors and offers a practical and scalable solution to address key areas within the NHS, such as improving efficiency and increasing capacity. The system's ability to proactively implement mitigations helps to provide necessary transparency for medication management. The system requires some IT infrastructure to operate, including hardware (barcode scanners, label printers, a printer, keyboards, a PC with Windows 10, scales, and visual documentation hardware) and software (SQL server with 4 central processing unit cores and 16 GB memory random-access memory, Microsoft .NET Framework [version 4.8.0], and DirectX [version 9.0]).

Funding

The study was funded by Becton Dickinson.

Data Availability

All data generated or analyzed during this study are included in this published paper. The datasets generated or analyzed are available from the corresponding author upon reasonable request.

Authors' Contributions

The study design was conceptualized by NH and BU. TM and HS collected the data. BU conducted the analysis and drafted the manuscript. All authors contributed to the critical revision of the manuscript for important intellectual content and provided final approval of the version to be published.

Conflicts of Interest

The York Health Economics consortium was paid by Beckton Dickinson to conduct the analysis and draft the manuscript.

References

1. NHS long term plan. National Health Service (NHS) England. 2019. URL: https://webarchive.nationalarchives.gov.uk/ukgwa/20230418155402/https:/www.longtermplan.nhs.uk/publication/nhs-long-term-plan/ [accessed 2025-05-25]
2. Transforming NHS pharmacy aseptic services in England. Department of Health & Social Care. 2020. URL: https://assets.publishing.service.gov.uk/media/5f9afdbdd3bf7f1e405d9a23/aseptic-pharmacy.pdf [accessed 2025-05-29]
3. Elliott RA, Camacho E, Jankovic D, Sculpher MJ, Faria R. Economic analysis of the prevalence and clinical and economic burden of medication error in England. BMJ Qual Saf 2021 Feb 11;30(2):96-105. [doi: 10.1136/bmjqs-2019-010206] [Medline: 32527980]
4. McLeod MC, Barber N, Franklin BD. Methodological variations and their effects on reported medication administration error rates. BMJ Qual Saf 2013 Apr;22(4):278-289. [doi: 10.1136/bmjqs-2012-001330] [Medline: 23322752]
5. ISMP list of high-alert medications in acute care settings. Institute for Safe Medication Practice. 2018. URL: https://www.ismp.org/sites/default/files/attachments/2018-08/highAlert2018-Acute-Final.pdf [accessed 2025-05-29]
6. Medication without harm: policy brief. World Health Organization (WHO). 2023. URL: https://iris.who.int/server/api/core/bitstreams/1eacccb6-838e-4787-bfd9-4bdeb4debfcf/content [accessed 2025-05-29]
7. A national evaluation of capacity in intravenous systemic anti-cancer therapy (IV SACT) preparatory services. British Oncology Pharmacy Association. URL: https://www.bopa.org.uk/resources/national-capacity-ivsact-evaluation/ [accessed 2025-05-29]
8. Wilson BE, Jacob S, Yap ML, Ferlay J, Bray F, Barton MB. Estimates of global chemotherapy demands and corresponding physician workforce requirements for 2018 and 2040: a population-based study. Lancet Oncol 2019 Jun;20(6):769-780. [doi: 10.1016/S1470-2045(19)30163-9] [Medline: 31078462]
9. Batson S, Mitchell SA, Lau D, Canobbio M, de Goede A, Singh I, et al. Automated compounding technology and workflow solutions for the preparation of chemotherapy: a systematic review. Eur J Hosp Pharm 2020 Nov 21;27(6):330-336 [FREE Full text] [doi: 10.1136/ejhpharm-2019-001948] [Medline: 33097615]
10. Meren Ü, Waterson J. Evaluating an automated compounding workflow software for safety and efficiency: implementation study. JMIR Hum Factors 2021 Nov 02;8(4):e29180 [FREE Full text] [doi: 10.2196/29180] [Medline: 34456182]
11. Terkola R, Czejka M, Bérubé J. Evaluation of real-time data obtained from gravimetric preparation of antineoplastic agents shows medication errors with possible critical therapeutic impact: results of a large-scale, multicentre, multinational, retrospective study. J Clin Pharm Ther 2017 Aug;42(4):446-453 [FREE Full text] [doi: 10.1111/jcpt.12529] [Medline: 28429389]

XSL•FO
RenderX

12. Jones KC, Weatherly H, Birch S, Castelli A, Chalkley M, Dargan A, et al. Unit costs of health and social care 2024 manual. Personal Social Services Research Unit (University of Kent) & Centre for Health Economics (University of York). 2025. URL: https://kar.kent.ac.uk/109563/ [accessed 2025-05-29]
13. YHEC's glossary of health economic terms. York Health Economics Consortium. URL: https://yhec.co.uk/resources-and-events/glossary/ [accessed 2025-05-29]
14. ISMP survey on sterile compounding in pharmacies reveals gaps, challenges. ECRI. URL: https://home.ecri.org/blogs/ismp-news/ismp-survey-on-sterile-compounding-in-pharmacies-reveals-gaps-challenges [accessed 2025-05-29]

Abbreviations

BD: Becton Dickinson
NHS: National Health Service
SACT: systemic anticancer therapy
WGH: Watford General Hospital

Edited by A Kushniruk; submitted 07.Oct.2025; peer-reviewed by J Waterson, U Lösch; comments to author 20.Nov.2025; accepted 02.Jan.2026; published 18.Feb.2026.

Please cite as:
Umpleby B, Hex N, Martins T, Soori H
Improving the Efficiency of the Intravenous Medicine Preparation Pathway With an Intravenous Workflow Software Solution in Full-Capacity Pharmacy Units at Watford General Hospital: Observational Study and Economic Analysis
JMIR Hum Factors 2026;13:e85408
URL: https://humanfactors.jmir.org/2026/1/e85408
doi:10.2196/85408
PMID:

©Bethany Umpleby, Nick Hex, Teresa Martins, Husanain Soori. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 18.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Organizational Tensions in the Implementation of Modifiable Off-the-Shelf Technologies in a University Hospital: Qualitative Multimethod Study

Alessia Nowak[1*], MSc; Carolina Gaul[1*], MSc; Elena Hinz[1]; Christine Knoll[1,2], MD, PhD; Begüm Kül[1]; Anja Wels[3], ADN; Julia C Berkmann[3], PhD; Cathrin de Pasquale[4], MSc; Gloria Kremser[5], MA; Daniel Fürstenau[1,6], Prof Dr; Akira-Sebastian Poncette[1], Prof Dr Med; Felix Balzer[1,7], Dr rer nat, Prof Dr Med; Lina Mosch[1,8,9], Dr med

[1]Charité – Universitätsmedizin Berlin, Institute of Medical Informatics, Berlin, Germany

[2]Digital Health-Non-communicable Diseases, University of Potsdam, Hasso Plattner Institute, Potsdam, Brandenburg, Germany

[3]Charité – Universitätsmedizin Berlin, Business Unit Project Management, Berlin, Germany

[4]50Hertz Transmission GmbH, IT Portfolio Management, Digitalization Transformation, Berlin, Germany

[5]Charité – Universitätsmedizin Berlin, Business Unit IT, Back Office, Berlin, Germany

[6]School of Business & Economics, Freie Universität Berlin, Berlin, Germany

[7]Charité – Universitätsmedizin Berlin, Chief Medical Information Office, Berlin, Germany

[8]Charité – Universitätsmedizin Berlin, Department of Anesthesiology and Intensive Care Medicine, Berlin, Germany

[9]Berlin Institute of Health at Charité – Universitätsmedizin Berlin, BIH Biomedical Innovation Academy, BIH Charité Digital Clinician Scientist Program, Berlin, Germany

[*]these authors contributed equally

Corresponding Author:
Lina Mosch, Dr med
Charité – Universitätsmedizin Berlin, Institute of Medical Informatics
Invalidenstraße 90, Ebene 05
Berlin, 10115
Germany
Phone: 49 30 450 5810
Email: lina.mosch@charite.de

Abstract

Background: Hospitals face increasing pressure to accelerate digital transformation. Modifiable off-the-shelf technologies (MOTs) combine standardized products with limited adaptability, offering promising opportunities for rapid digitalization. However, implementing MOTs in complex hospital settings involves multiple barriers, facilitators, and organizational dynamics that require deeper investigation.

Objective: This study identifies barriers and facilitators in the implementation of MOTs in hospitals and explores how organizational dynamics, conceptualized as tensions, emerge throughout this process.

Methods: Guided by a constructivist-interpretivist paradigm, the study followed a collaborative action research approach. A qualitative multimethod design was used, including observations, workshops, and focus groups with clinical and project management staff. Data were analyzed using deductive-inductive qualitative content analysis following Kuckartz's approach.

Results: We analyzed 12 deidentified researcher protocols from 5 observations, 3 workshops, and 4 focus groups. Across these activities, 129 individuals from various wards and departments participated. Barriers and facilitators were clustered into 6 categories representing inhibiting and enabling conditions of MOT implementation. Barriers included product limitations (22.6%, 88/390), misaligned implementation process (4.4%, 17/390), absence of available individuals (3.8%, 15/390), structural challenges (2.8%, 11/390), and resource constraints (2.8%, 11/390). Facilitators included orchestrated implementation process (13.8%, 54/390), product alignment (14.4%, 56/390), effective coordination and communication (9.2%, 36/390), presence of available individuals (6.7%, 26/390), available resources (3.1%, 12/390), and structural assets (1.8%, 7/390). These categories were conceptualized into 3 organizational tensions: limited relative advantage of generic systems, structured participation challenges in complex settings, and constrained engagement despite motivation due to limited resources.

XSL•FO
RenderX

Conclusions: Tensions in MOT implementation reflect organizational misfits that go beyond concrete barriers or facilitators and instead require systemic and agile organizational structures. The analysis introduces the concept of user:ability as an organizational capacity for enabling engagement with MOTs. We derive 3 tailored strategies to mitigate the tensions: ensuring relative advantage through IT integration and learning; empowering user participation through digital transformation and communication; and enabling clinician engagement through dedicated resources and systematic implementation. This study advances prior work by deepening the understanding of implementation determinants into organizational tensions and offering actionable strategies to address them. It also illustrates how collaborative, agile research approaches can illuminate complex organizational dynamics and support hospitals during transformative change.

(JMIR Hum Factors 2026;13:e84841) doi:10.2196/84841

KEYWORDS

health information technology; digital health; organization; qualitative research; implementation; tensions; paradoxes; user involvement

Introduction

A growing body of literature in information systems, health services, and implementation science highlights the potential of digital health technologies (DHTs) to improve health care delivery and outcomes [1-4]. Empirical studies demonstrate that DHTs, such as electronic medication systems, can enhance patient safety, care coordination, and clinical workflow efficiency [5-10].

In Germany, this recognition has been reinforced by major political initiatives. The Krankenhauszukunftsgesetz (Hospital Future Act; KHZG) created a strong regulatory impetus by tying substantial federal and state funding to the adoption of specific digital infrastructures [11]. Intended to improve care quality and efficiency, the KHZG mandates core digital solutions for German hospitals and penalizes nonimplementation with reductions of up to 2% of inpatient reimbursement, thereby enforcing rapid adoption [12,13]. Yet, while the KHZG accelerates digital adoption, DHTs remain subject to stringent safety and performance regulations, which can inadvertently slow innovation, delay responsiveness to emerging user needs, and limit design flexibility [14,15].

In this context, we conceptualize the DHTs examined in this study as modifiable off-the-shelf technologies (MOTs), drawing on modifiable off-the-shelf software or packaged software [16,17]. MOTs are prebuilt, licensed products that allow limited customization but are not developed in-house for a specific organizational use case. In this study, MOTs encompass a range of implementations: from a completed end user–initiated development scaled into the clinic, to system-wide infrastructures, and Software-as-a-Medical-Device, such as digital medication support systems [18], to applications closely integrated with hardware components, such as mobile ward rounds with iPads (Apple Inc). While MOTs promise rapid deployment and regulatory compliance, their limited adaptability often constrains alignment with heterogeneous clinical workflows and evolving organizational needs [19,20]. These dynamics can foster resistance and undermine organizational fit [20,21], particularly in large academic hospitals, where complexity heightens the risk of "unbridled adoption" [22]. In this context, the generic design of MOTs, developed for broad markets rather than the situated needs of specific clinical practices, often reduces usability because it lacks sensitivity to local workflows and limits opportunities for contextualized, implementation-level design [23]. As a result, successful implementation requires a stronger focus on end users during and after implementation, ensuring that users can use the system safely and efficiently through adequate training and sustained support [24].

Implementation science offers structured approaches to analyze these dynamics by examining the conditions under which adoption succeeds or fails. At its core is the examination of barriers and facilitators: barriers are factors that hinder or prevent uptake, while facilitators are enabling conditions within a specific context that promote adoption [25].

Previous research has identified barriers at multiple levels, including technological barriers such as interoperability [26,27] and system complexity [27,28]; individual-level barriers such as limited digital literacy [29] and low acceptance [29,30]; organizational barriers such as the lack of integration of digital technologies into existing workflows [30,31] and limited resources [29,31]; as well as structural barriers, including the lack of stakeholder involvement [26] and insufficient funding [32].

Beyond well-documented barriers, recent research points to deeper organizational dynamics, such as tensions and paradoxes, that are not easily resolved and require ongoing management [33]. Examples include conflicts between the push for standardization and the need for clinical autonomy; the pursuit of digital efficiency and the value of hands-on care; and the promise of technology and the reality of added burden, reflecting organization-technology misfits [33-35]. Building on Strong and Volkoff [19], misfits arise when technology and organizational practices fail to align. They appear in 6 domains: functionality, data, usability, role, control, and organizational culture. Each can surface either as a deficiency (missing features or capabilities) or as an imposition (an inherent technological characteristic needed for seamless integration) [19]. If left unmanaged, organization-technology misfits can impede adoption by fostering resistance, neglect, or misuse of otherwise beneficial technologies [33-35].

To address barriers, tensions, and misfits, implementation science applies evidence-based theoretical models and frameworks [21,36-40]. However, these are often insufficiently applied in practice [41], contributing to a persistent knowledge

XSL•FO
RenderX

translation gap [42]. Action research can help bridge this gap by generating insights that are simultaneously actionable for organizations and generalizable for theory building [43]. Tailored implementation strategies [44] remain essential, as MOT implementation must account for technical and organizational factors alongside competing demands from regulations, limited resources, and staff needs.

Building on the outlined challenges and literature, we posit a set of assumptions that guide our study. As MOTs offer limited opportunities for participation in design, we expect usability and functionality misfits to emerge during implementation. Building on this premise, we suggest that effective MOT implementation requires tailored strategies to address contextual needs during implementation, as the limited modifiability of MOTs restricts contextual adaptation. We further argue that structural characteristics of hospitals can reinforce these misfits and intensify tensions during implementation.

These propositions lead to the following research question: What are the barriers and facilitators in the implementation of MOTs, and what organizational dynamics emerge during this process within complex hospital contexts?

Methods

Study Design

This study followed a qualitative, multimethod, collaborative participatory action research design within a constructivist-interpretivist paradigm. The design aimed to explore how clinical and project management staff experienced the implementation of MOTs in a complex tertiary hospital environment [45,46]. Action research was conducted in iterative Plan-Act-Observe-Reflect cycles [47,48], integrating observations, focus groups, and participatory workshops in a convergent triangulation design. Following the action research paradigm, female implementation researchers with backgrounds in clinical medicine, human factors, and information systems acted as the action researchers, whereas female members of the Chief Medical Information Office and the operational project management unit represented the client system. Participants' accounts were treated as first-order data and interpreted by the researchers to develop second-order concepts [49]. As some team members were involved in coordinating implementation activities, their dual roles not only provided valuable contextual insight but also introduced the possibility of bias. Reflexivity was supported through systematic field notes, joint data collection, and regular team discussions, and analytic distance was strengthened by involving 2 researchers (BK and DF), who had not participated in data collection, in leading the analysis. The study was reported in accordance with the Standards for Reporting Qualitative Research (SRQR; see Multimedia Appendix 1) and the Consolidated Criteria for Reporting Qualitative Research (COREQ; see Multimedia Appendix 2) [50,51].

Before data collection, participants were informed about the study's purpose, the researchers' roles, and the goals of the initiative. No formal relationships were established in advance. Researchers were introduced in line with their collaborative roles as action researchers (AN, CG, EH, and LM) and clients (AW, CdP, JCB, and GK) [47]. Professional backgrounds and interest in implementation processes were disclosed, whereas the research questions were not shared.

Setting

The study was conducted at Charité – Universitätsmedizin Berlin, a large German university hospital employing over 23,000 staff and operating more than 3000 beds across more than 100 clinics and departments [52].

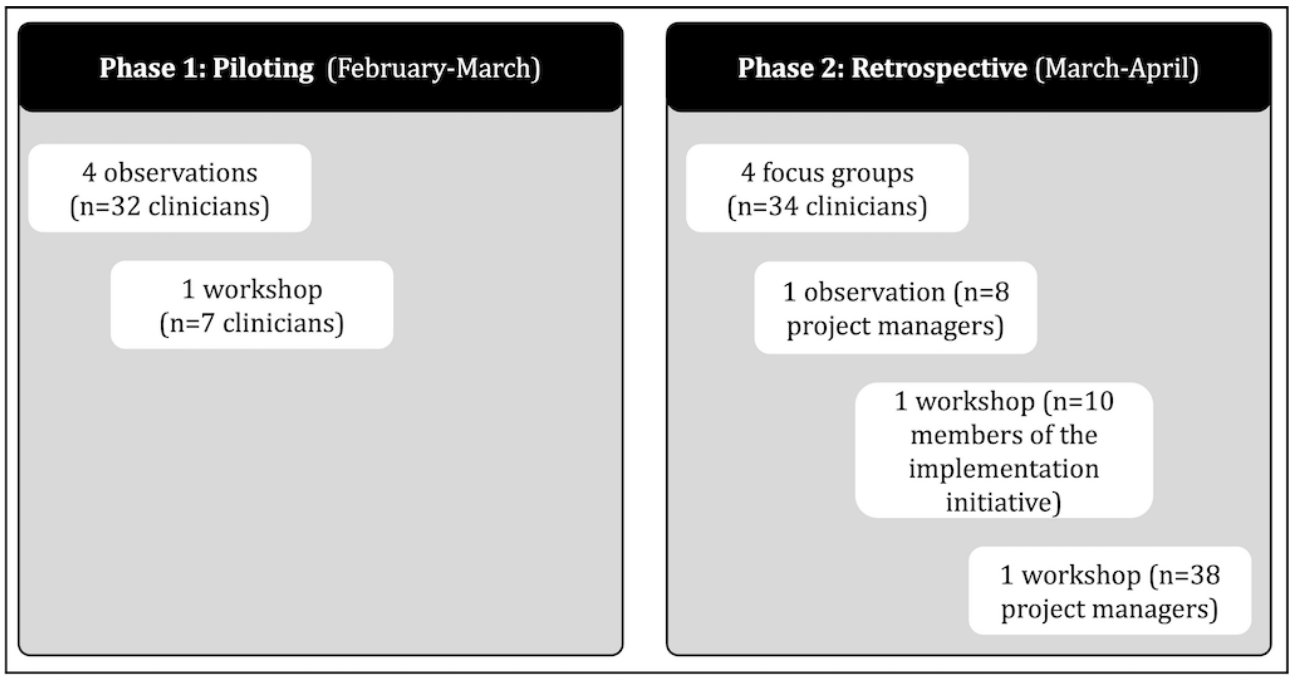
The research was embedded in the hospital's digital transformation initiative funded under the German Hospital Future Act (KHZG), which promotes the implementation of DHTs across German hospitals.

Recruitment and data collection occurred from February to April 2025 across 2 implementation phases: (1) a pilot phase observing pre- and early implementation activities and (2) an evaluation phase retrospectively assessing experiences across wards and departments. Together, these stages provided rich insights into the implementation process (Figure 1). Data were collected in inpatient wards, emergency and outpatient clinics, and administrative project management offices.

Access to the field was established by the initiative's lead via email, drawing on existing professional relationships to promote trust and encourage engagement, as well as through presentations of the initiative in various internal organizational committees. To contextualize these implementation activities, the next section describes the portfolio of digital technologies introduced under the KHZG initiative and examined in this study.

XSL•FO
RenderX

Figure 1. Data collection procedure for a qualitative study on modifiable off-the-shelf technology implementation at a large German academic hospital, illustrating a 2-stage process (pilot and evaluation phases) conducted from February to April 2025 across clinical and administrative settings.



The Implementation Portfolio

The research was conducted within the framework of the Hospital Future Act, a political mandate that allocates targeted funding to accelerate hospitals' digital transformation. Hospitals apply for these funds and must comply with national standards for DHTs to be eligible for support [11]. In line with the requirements of the KHZG and the hospital's overarching digital strategy [53], which emphasizes interoperability, user-centered design, and improved care, a dedicated implementation initiative was launched. Its objective was to introduce 10 preselected MOTs that fulfill KHZG criteria and contribute to the operationalization of the telematics infrastructure, Germany's secure and interoperable digital health network [54]. These technologies, formally approved as part of the hospital's digital transformation program, were implemented across inpatient and outpatient departments to enhance clinical documentation, resource management, device integration, and interprofessional communication. Patient-facing or consumer-oriented applications were not included because they were not part of the active implementation portfolio during the study period and thus did not constitute the exposure of interest. The study aimed to investigate the experiences, practices, and challenges of clinical personnel as active users of KHZG-funded MOTs, rather than to evaluate patient experiences or patient-technology interactions. The portfolio comprised both mandatory and optional solutions (see Table 1). Aligning with the action research approach [47], the initiative was iteratively refined in response to contextual factors, including high workloads and limited resources, as well as ongoing adjustments to evolving KHZG regulations.

XSL•FO
RenderX

Table 1. Overview of mandatory and optional modifiable off-the-shelf technologies included in the implementation portfolio of a digital transformation initiative at a large German university hospital (February-April 2025). Technologies were implemented in inpatient and outpatient clinical areas as part of the KHZG[a]-funded program and represent the digital tools to which clinical staff were exposed during the study.

Technology	Description	Requirement type
Electronic Sick Leave Certificate [55]	Online transmission of sick leave certificates via the telematics infrastructure.	Mandatory
Nationwide electronic health record (elektronische Patientenakte) [56]	Integration and implementation of a unified IT system for medical report writing across the hospital.	Mandatory
ePrescription Implementation [57]	Digital creation and transmission of prescriptions via the telematics infrastructure.	Mandatory
Digital Medication Management (inpatient) [58]	Digital documentation of medication on regular wards and some outpatient clinics; work on emergency and intensive care units in progress.	Mandatory
KIM (Communication in Medicine) – KIM Email [59]	Secure, encrypted email service within the telematics infrastructure for medical communication.	Optional
Operating Room Solutions (Digital Implant Passport) [60][b]	Clinical information system–integrated software module for implant documentation.	Optional
Mobile Ward Round [61]	An iPad app allowing access to critical clinical systems/features to support clinical ward rounds for doctors, nurses, and other staff (Wi-Fi dependent).	Optional
Admission Planning Tool [62][b]	Standardized planning for inpatient admissions.	Optional
Fluid Balance Module [63][b]	Digital documentation of access points, drains, and fluid balances in the clinical information system.	Optional
Digital Whiteboard for Ward Management [64][c]	Digital workflow organization tool for inpatient services in hospital wards and intensive care units.	Optional

[a]KHZG: Krankenhauszukunftsgesetz (Hospital Future Act).

[b]SAP modules were modified by the local IT department.

[c]End user–initiated development.

Participants

Participants were clinical staff (physicians, nurses, and patient coordinators) and nonclinical project management personnel involved in or affected by the digitalization initiative.

Eligibility criteria were (1) employment at Charité in a clinical or project management role and (2) involvement in or exposure to at least one MOT implementation. Recruitment followed a multistage purposeful sampling strategy to ensure maximum variation across disciplines and hierarchical levels [65]:

- Stage 1 (maximum variation sampling): selection of diverse departments (regular ward, emergency, and outpatient settings).
- Stage 2 (emergent purposive sampling): inclusion of additional participants expressing interest during project meetings and presentations.
- Stage 3 (criterion sampling): targeted recruitment of project managers due to their pivotal role in planning and coordination.

Elements of convenience sampling were used when data collection was integrated into existing meeting schedules.

A total of 129 participants were included: 73 clinicians, 46 project managers, and 10 members of the implementation initiative. While participation was generally high, a small number of invited nurses and physicians declined due to time constraints. Exact numbers of refusals and dropouts were not systematically recorded, as participation was integrated into routine clinical and project management activities. Sampling decisions across all stages aimed to capture diverse perspectives from frontline users and institutional implementers. Units of study were 12 researcher protocols: 5 from observations, 4 from focus groups, and 3 from workshops (see Multimedia Appendix 3). While 129 individuals contributed to data collection activities, analytic inferences were drawn from these protocols rather than from individual participants.

Demographic data were not collected because, following Rapoport's [43] view of action research as a collaborative, context-bound process focused on collective practice change, roles and functions in implementation were considered more relevant than individual characteristics, and statistical generalizability was not sought.

Assessments

Primary outcomes were perceived barriers and facilitators to MOT implementation, as well as underlying organizational dynamics that became evident during the process.

Exposures of interest included the characteristics of MOTs (usability, adaptability, and interoperability) and organizational factors (resources, infrastructure, and communication structures).

Potential contextual influences considered during interpretation included differences in departmental resources, professional hierarchies, and workload variation.

All outcomes were derived from qualitative accounts rather than standardized instruments, reflecting participants' lived experiences and professional perspectives.

XSL•FO
RenderX

Data Sources

Overview

An emergent qualitative design [66] guided data collection. Outcomes were assessed using 3 qualitative data sources: focus groups [67], workshops [68], and field observations [69] (see Multimedia Appendix 3). Data collection was conducted across 2 implementation phases and combined real-time engagement with retrospective reflection to capture both situated practices and reflective accounts of barriers, facilitators, and organizational dynamics. Each method contributed distinct perspectives on implementation processes and context-specific determinants.

Field Observations (n=5 Protocols)

Field observations were conducted on-site or online to capture real-time implementation practices, clinical routines, and interactional dynamics within the organizational environment. In phase 1, researchers (AN, CG, LM, AW, JCB, GK, and CdP) conducted 4 departmental observations (each 60-90 minutes; protocols [P] 9-12), focusing on implementation workflows and situated engagement with available MOTs, including how interest in specific solutions emerged or failed to emerge. In phase 2, a 90-minute online observation with project managers provided broader perspectives on cross-departmental coordination processes (P2). Researchers adopted a participant-observer stance, with varying levels of engagement depending on the activity. Observational field notes documented contextual factors, actor interactions, emergent issues, and workflow characteristics [69]. Field observations focused on capturing situated practices and real-time interactions as they occurred in the clinical and organizational context, rather than participants' retrospective accounts or evaluative judgments.

Workshops (n=3 Protocols)

Workshops served as participatory formats to support collective reflection and cocreate implementation strategies. In phase 1, following a ward's expressed interest in an MOT, AN and EH conducted a 2-hour back-casting workshop [70] with senior and assistant physicians, nursing staff, and patient coordinators to envision future workflows and articulate required process changes, starting from a desired future state (P4). In phase 2, a 30-minute workshop with project managers (P6), conducted by CG, AN, and EH, focused on evaluating implementation steps ("do," "don't," "must") and strategies for resource-constrained contexts. A 1-hour internal workshop with implementation initiative members (P7), led by CG and AN, supported reflexive examination of the overall approach. Workshops followed a semistructured but flexible facilitation style to allow participant-led exploration of barriers, facilitators, and implementation needs. They were explicitly future-oriented and designed to support joint sense-making, cocreation, and the development of shared implementation strategies.

Focus Groups (n=4 Protocols)

In phase 2, CG, AN, JCB, CdP, and AW conducted 4 one-hour focus groups with clinicians across diverse professional and hierarchical levels (P1, P3, P5, and P8). Focus groups explored implementation experiences, communication preferences, perceived determinants of adoption, and emergent best practices. Sessions were held on-site or digitally and were researcher-moderated. Focus groups elicited retrospective reflections, complementing the observational insights generated in phase 1. In contrast to workshops and observations, focus groups emphasized retrospective reflection on implementation experiences and facilitated comparison of perspectives across professional roles.

Across all methods, manual note-taking (without audio or video recordings) ensured participant comfort, protected confidentiality, and supported candid dialogue. A structured field note template (see Multimedia Appendix 4) was developed specifically for the study to document setting characteristics, actor interactions, contextual influences, and emerging themes consistently across methods. Only study participants and researchers were present during data collection; no nonparticipating individuals attended. Field notes were primarily compiled by AN and CG and supplemented by LM, CdP, AW, and JCB, enabling iterative refinement of prompts and observational foci in response to field engagement [66].

Study Size

The final study size (n=129 participants across 12 protocols) was determined pragmatically, based on data sufficiency, diversity of perspectives, and information power. This approach aligns with qualitative research conventions rather than statistical power considerations [71]. Breadth and representativeness across roles and departments were prioritized over saturation, given the multimethod, organizational scope of the research.

Data Analysis

All analyses were conducted at the protocol level, with the 12 researcher protocols forming the analytic corpus. Coding frequencies indicate patterns across activities rather than individual-level quantification, consistent with the study's action research design [47].

All protocols were transcribed and analyzed in MAXQDA 2022 (VERBI GmbH) [72] using deductive-inductive qualitative content analysis following Kuckartz and Rädiker [73].

To ensure analytic rigor, all coding activities were conducted by product-independent action researchers who were not part of the client system. AN and CG defined a priori codes from the Consolidated Framework for Implementation Research [21,36], guided by prior knowledge of barriers and facilitators. An independent researcher (BK), who was not involved in data collection, then coded the material in MAXQDA, applying these codes and developing inductive ones specific to the MOT context. BK subsequently assigned these codes to either the first-order category MOT implementation barrier or MOT implementation facilitator, with regular review by AN and CG. AN grouped related codes into 12 second-order categories (6 barriers and 6 facilitators), which were validated in 3 team meetings (AN, CG, and LM). To prevent "data overload," we decided to report only codes with 4 or more coded segments [74]. As Kuckartz and Rädiker [73] do not provide numerical thresholds, these cutoffs were established through analytical reflection to manage data complexity and ensure transparent reporting. This approach aligns with qualitative standards that emphasize conceptual significance over statistical

XSL•FO
RenderX

representativeness. A detailed overview of coding frequencies is provided in Multimedia Appendix 5.

Building on preliminary qualitative findings from the different data collection steps, MOT implementation barriers and facilitators also served as a basis for analyzing organizational tensions and deep-seated dynamics. Together, the 3 data sources provided complementary perspectives on the implementation of MOTs. Field observations captured how available MOTs were encountered in practice and how reasons for adoption or nonadoption emerged in everyday clinical and organizational workflows. Workshops contributed to the structured, future-oriented articulation of implementation pathways, including required process changes and strategies for resource-constrained contexts. Focus groups supported retrospective reflection on individual implementation experiences, preferences, and perceived determinants of adoption. Additionally, findings from the different data collection steps were triangulated to inform the interpretation of recurring patterns and contradictions. Based on access to the coded material, LM and DF, who were not involved in data collection and therefore contributed additional analytical distance, independently examined codes for contradictory demands or competing logics (eg, complexity vs participation, resource constraints vs engagement). Findings were discussed in 3 team meetings with AN and CG.

Integrating findings across methods, perspectives, and study phases, the analysis served as the basis for identifying higher-order organizational tensions. As a result, the final conceptual interpretations and tailored implementation strategies were derived from this integrated, triangulated analytical process.

Triangulation of methods (observations, workshops, and focus groups) and perspectives (clinicians, project managers, and implementation staff) enhanced trustworthiness [75]. Reflexivity was maintained through analytic memos and peer debriefing sessions. Quantitative counting of coded segments (frequencies and percentages) was used descriptively to illustrate emphasis, not to infer significance.

Ethical Considerations

Ethics Approval

The study protocol was reviewed and approved by the Charité – Universitätsmedizin Berlin Ethics Committee (approval number EA4/031/23). As the study involved human participants in the form of clinical and project management staff, institutional ethics oversight was required and obtained.

Informed Consent

Before participation, all individuals received verbal information about the study's aims, procedures, voluntary nature, and their right to withdraw at any time without consequences. Verbal informed consent was obtained before all data collection activities. The ethics committee confirmed that verbal consent was appropriate given the minimal risk, nonclinical nature of the study, and the absence of patient-related data.

Privacy and Confidentiality

To protect participant confidentiality, all field notes and workshop and focus group documentation were deidentified at the point of data collection. No names, professional identification numbers, or other personal identifiers were recorded. Data were stored on secure, password-protected institutional servers accessible only to the research team. Findings are presented exclusively in aggregated or anonymized form.

Compensation

Participants did not receive any financial or material compensation for their involvement. Participation was voluntary and occurred during or adjacent to routine professional activities.

Identifiability in Images or Multimedia Appendices

No identifiable individuals are presented in the manuscript or its multimedia appendices. All illustrative figures and multimedia appendices contain only anonymized documents and deidentified field note excerpts. No visual materials include recognizable individuals; therefore, no individual image-based consent forms were required.

Results

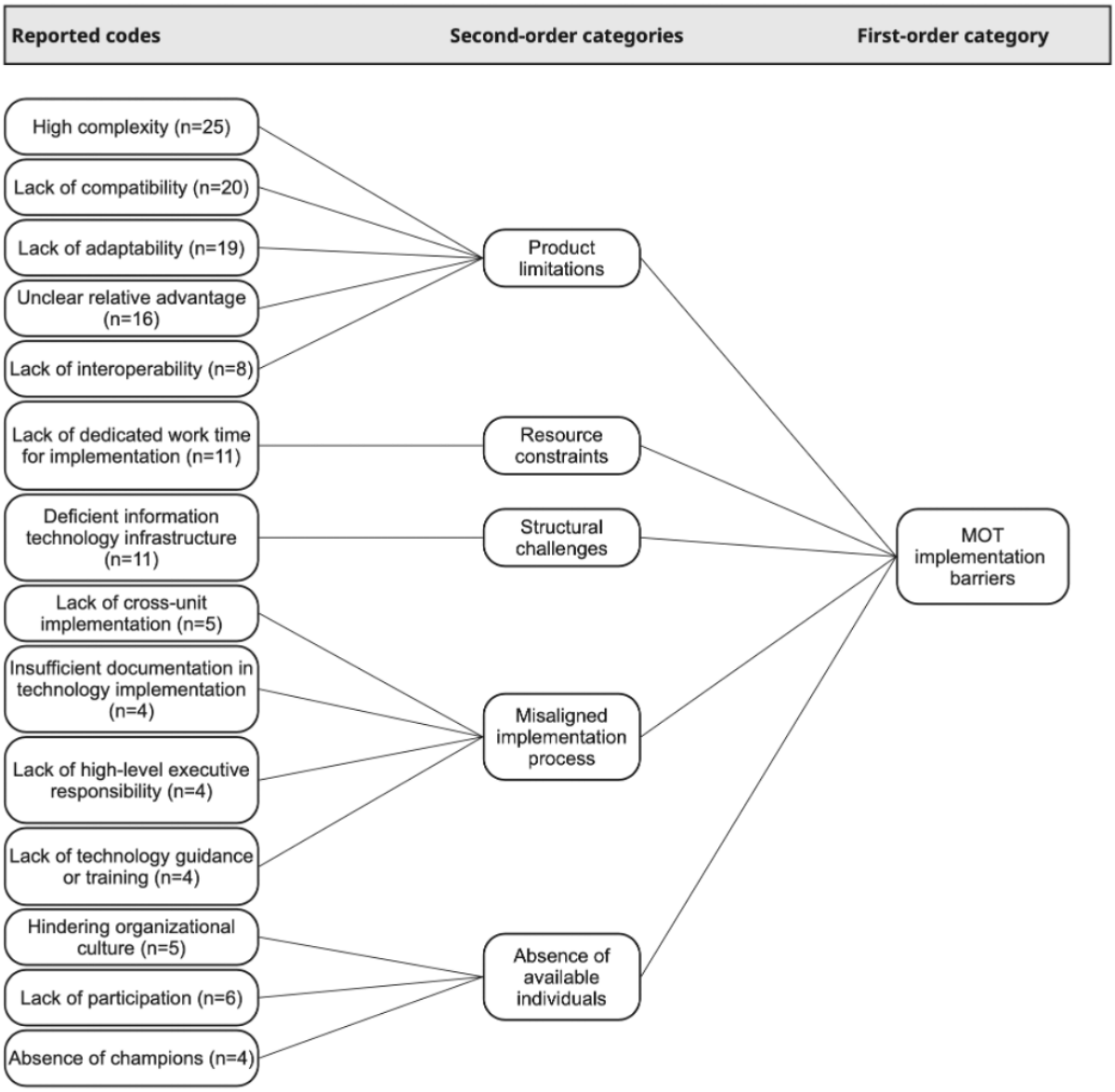
Overview and Analytical Structure

Results are presented in 2 sections: (1) 6 integrated second-order categories combining barriers and facilitators within shared aspects of MOT implementation (see Figures 2 and 3) and (2) organizational dynamics conceptualized as 3 tensions. We defined barriers as factors that hinder or prevent uptake and facilitators as enabling conditions that support uptake [25]. Codes and second-order categories were labeled according to participants' perceptions and our interpretation: barriers were typically described using terms such as lack of, limited, or deficient, whereas facilitators were usually characterized by the perceived presence or availability of a factor.

In total, the coding process identified 390 data segments, comprising 172 barriers and 218 facilitators across the 12 researcher protocols. Frequencies indicate how often themes occurred across protocols and are not intended to represent participant-level prevalence. Multimedia Appendix 5 provides an overview of the distribution of reported and unreported codes, as well as all code frequencies. Representative quotes for all codes can be found in Multimedia Appendix 6. The following sections elaborate on each second-order category, beginning with those related to product characteristics.

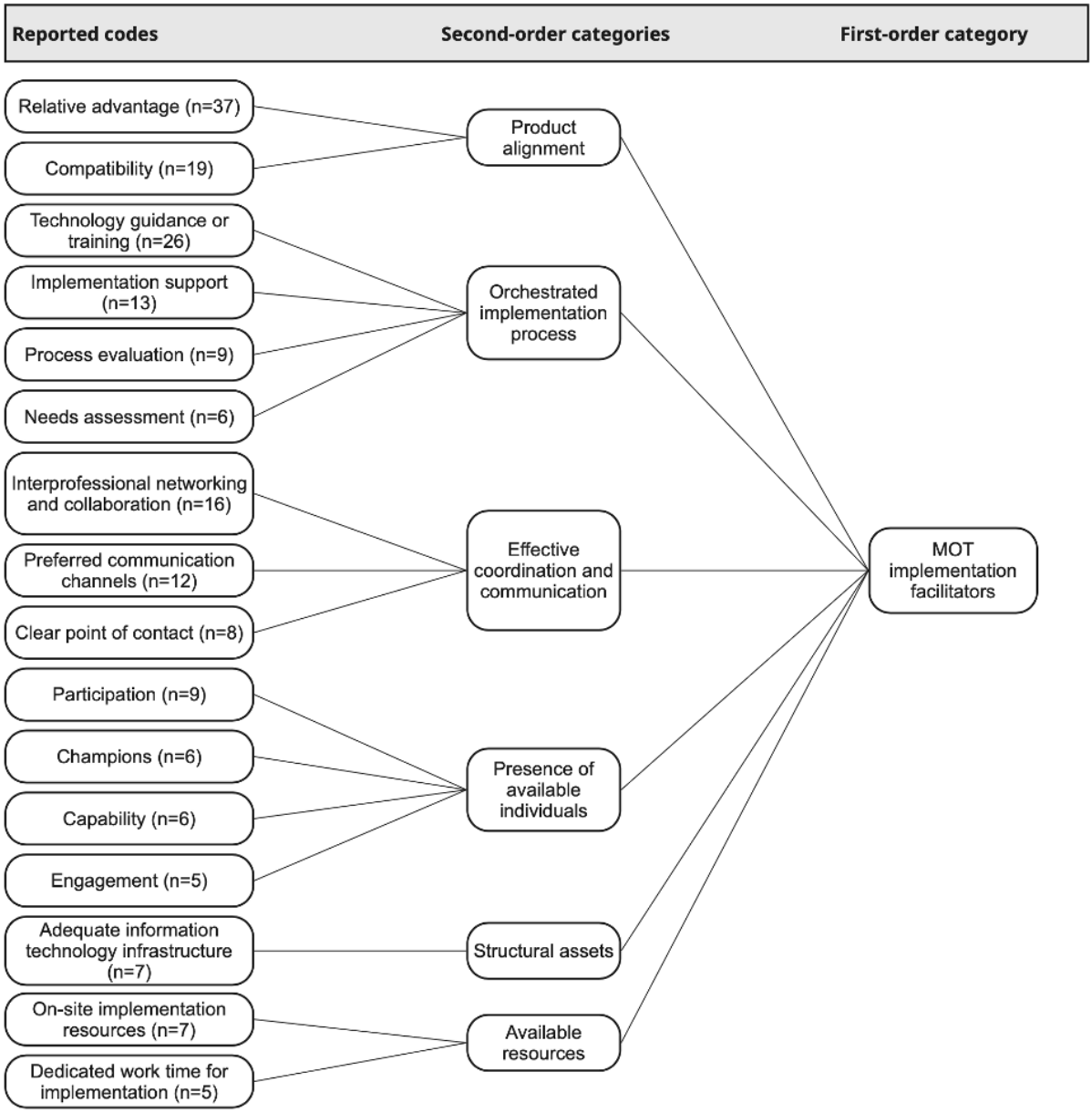
XSL•FO
RenderX

Figure 2. Coding scheme of modifiable off-the-shelf technology (MOT) implementation barriers. Reported codes were aggregated into 5 second-order categories—product limitations, resource constraints, structural challenges, misaligned implementation processes, and absence of available individuals—which together constitute the overarching category of MOT implementation barriers.



XSL•FO
RenderX

Figure 3. Coding scheme of modifiable off-the-shelf technology (MOT) implementation facilitators. Reported codes were grouped into 6 second-order categories—product alignment, available resources, structural assets, an orchestrated implementation process, the presence of available individuals, and effective coordination and communication—which together constitute the overarching category of MOT implementation facilitators.



Product Limitations and Product Alignment

This category captures how the functional, technical, and workflow-related characteristics of MOTs shaped users' experiences.

Barriers arose when MOTs were difficult to operate, poorly aligned with existing routines, or technically constrained. These challenges were reflected in the lack of compatibility (20/172, 11.6%, barriers; 20/390, 5.1%, coded segments), such as when duplicate documentation or parallel paper processes were required. Lack of adaptability (19/172, 11%, barriers; 19/390, 4.9%, coded segments) captured the limited flexibility of MOTs to accommodate diverse wards, workflows, or professional roles. The added value of MOTs was not always evident, as reflected in unclear relative advantage (16/172, 9.3%, barriers; 16/390, 4.1%, coded segments). Difficulties were further reinforced by missing interfaces and fragmented data flows, represented by a lack of interoperability (8/172, 4.7%, barriers; 8/390, 2.1%, coded segments), which required manual bridging and complicated coordination across units. High complexity (25/172, 14.5%, barriers; 25/390, 6.4%, coded segments) referred to multistep logins, unstable Wi-Fi, and slow system responses that conflicted with fast-paced clinical work, as exemplified by a participant describing the mobile ward round:

> *[Mobile Ward Round] You go into the room to see the first patient, document everything, then go to the second patient, then the iPad is off, you have to enter a code and unlock it again, then log back into the HIS – by the time you've logged in, 3-5 minutes have passed and the others are already with the next patient. Then you want to look at the MRI image*

XSL•FO
RenderX

> *again, but the Wi-Fi cuts out. So, the idea is great, but the application isn't* [P9]

Facilitators emerged when MOTs aligned with clinical routines and offered clear benefits. Relative advantage was the most frequently reported facilitator (37/218, 17%, facilitators; 37/390, 9.5%, coded segments), emphasizing efficiency gains and reduced communication effort. Compatibility (19/218, 8.7%, facilitators; 19/390, 4.9%, coded segments) supported adoption when system functions fit existing workflows or professional expectations.

Together, these patterns indicate that MOT uptake depended on the extent to which technical design, usability, and workflow integration cohered with everyday clinical practice.

Resource Constraints and Available Resources

This category captures how staffing levels, protected time, and available support shaped clinicians' ability to engage in MOT implementation activities.

The most prominent barrier was a lack of dedicated work time for implementation, which accounted for 11 out of 172 (6.4%) barriers and 11 out of 390 (2.8%) coded segments. Participants reported that implementation-related tasks had to be completed alongside routine clinical duties, leaving little opportunity for training, experimentation, or supporting colleagues. As P7 noted: "Enthusiasm is there, but the time resource is missing."

Facilitators emerged when additional resources or protected time were available. On-site implementation resources represented 7 out of 218 (3.2%) facilitators and 7 out of 390 (1.8%) coded segments, and dedicated work time for implementation accounted for 5 out of 218 (2.3%) facilitators and 5 out of 390 (1.3%) coded segments. Participants described that responsive local support and focused time windows enabled them to engage in implementation activities without competing clinical responsibilities and to incorporate MOTs more consistently into everyday practice.

Structural Challenges and Structural Assets

This category reflects how participants experienced the adequacy and reliability of the hospital's technical infrastructure during MOT use and implementation.

Barriers were primarily captured in the code-deficient IT infrastructure, which accounted for 11 out of 172 (6.4%) barriers and 11 out of 390 (2.8%) coded segments. Participants described limited device availability, unstable Wi-Fi, slow system performance, and outdated hardware, all of which interfered with routine work and discouraged sustained use of MOT functions. Furthermore, these gaps sometimes prevented MOTs from being implemented in the first place, as required equipment was delayed or unavailable. As one participant noted: "Procurement of technical infrastructure unclear or very slow and bureaucratic via procurement platform" [P3].

By contrast, facilitators emerged when the infrastructure functioned reliably. Adequate IT infrastructure accounted for 7 out of 218 (3.2%) facilitators and 7 out of 390 (1.8%) coded segments. Participants reported smoother logins, stable network access, and consistent availability of clinical information; these conditions supported the routine use of MOTs and reduced the need for workaround practices.

Misaligned Implementation Process and Orchestrated Implementation Process

This category captures how the structure, coordination, and support of implementation activities influenced clinicians' experiences.

Barriers reflected fragmented processes, unclear responsibilities, and insufficient preparation. Lack of cross-unit implementation (5/172, 2.9%, barriers; 5/390, 1.3%, coded segments) and insufficient documentation in technology implementation (4/172, 2.3%, barriers; 4/390, 1.0%, coded segments) highlighted difficulties in understanding which technologies were already in use and how previous rollouts had unfolded. Several participants reported that oversight for implementation had been delegated to local clinical leaders who, due to existing clinical demands, had limited capacity to take on additional coordination tasks, as reflected in the lack of high-level executive responsibility (4/172, 2.3%, barriers; 4/390, 1.0%, coded segments). Reports of lack of technology guidance or training (4/172, 2.3%, barriers; 4/390, 1.0%, coded segments) further contributed to inconsistent processes and limited practical support.

Facilitators emerged when implementation was structured and actively supported. Technology guidance or training was a frequent facilitator in this domain (26/218, 11.9%, facilitators; 26/390, 6.7%, coded segments), alongside implementation support (13/218, 6.0%, facilitators; 13/390, 3.3%, coded segments), process evaluation (9/218, 4.1%, facilitators; 9/390, 2.3%, coded segments), and needs assessment (6/218, 2.8%, facilitators; 6/390, 1.5%, coded segments). Participants emphasized that accessible, context-specific training formats, visible support from the implementation team, and iterative feedback loops helped align expectations and adapt strategies to local workflows: "Planned iterative approach from the start—reminding people, asking about further interaction" [P7].

Absence of Available Individuals and Presence of Available Individuals

This category reflects how the availability, engagement, and roles of clinical staff and members of the implementation team were experienced by participants during MOT adoption.

Barriers were reported when individuals lacked time, recognition, or opportunity to participate meaningfully. Lack of participation accounted for 6 out of 172 (3.5%) barriers and 6 out of 390 (1.5%) coded segments. Participants described that, although interest often existed, competing clinical duties limited involvement. Hindering organizational culture represented 5 out of 172 (2.9%) barriers and 5 out of 390 (1.3%) coded segments, referring to hierarchical structures and strained working climates that reduced openness to change. The absence of champions accounted for 4 out of 172 (2.3%) barriers and 4 out of 390 (1.0%) coded segments, indicating missing local multipliers who could support peers or coordinate with the project team.

XSL•FO
RenderX

By contrast, facilitators emerged when individuals—whether clinicians or implementation team members—were available and engaged. Participation accounted for 9 out of 218 (4.1%) facilitators and 9 out of 390 (2.3%) coded segments, with staff actively initiating or supporting MOT rollout activities. Presence of champions represented 6 out of 218 (2.8%) facilitators and 6 out of 390 (1.5%) coded segments, capturing designated or informal multipliers who guided colleagues, tested functions, or liaised with the implementation team. Capability accounted for 6 out of 218 (2.8%) facilitators and 6 out of 390 (1.5%) coded segments, with participants highlighting digital competence and willingness to learn. Engagement accounted for 5 out of 218 (2.3%) facilitators and 5 out of 390 (1.3%) coded segments, reflected in proactive contributions to problem solving or workflow adaptation. A participant summarized the importance of identifiable and engaged individuals: "What else do we need from the clinics? Clear, consistent contacts who can serve as multipliers" [P6].

Effective Coordination and Communication

This category captures how interprofessional coordination and communication practices shaped the implementation of MOTs.

Although participants occasionally referred to coordination challenges, no communication-related barrier reached the minimum reporting threshold of 4 coded segments used in this study. Facilitators, however, were more frequently described and were reflected in interprofessional networking and collaboration, preferred communication channels, and a clear point of contact.

Interprofessional networking and collaboration accounted for 16 out of 218 (7.3%) facilitators and 16 out of 390 (4.1%) coded segments. Participants emphasized that structured collaboration between clinical professions and the project team supported implementation activities, as illustrated by one clinician who noted the value of "bringing different professional groups together to plan the implementation" [P6]. Preferred communication channels represented 12 out of 218 (5.5%) facilitators and 12 out of 390 (3.1%) coded segments, highlighting the importance of using familiar and accessible communication formats, such as in-person briefings, emails, or departmental meetings. A clear point of contact accounted for 8 out of 218 (3.7%) facilitators and 8 out of 390 (2.1%) coded segments, emphasizing the role of identifiable, reliable contacts for resolving questions and maintaining continuity during implementation.

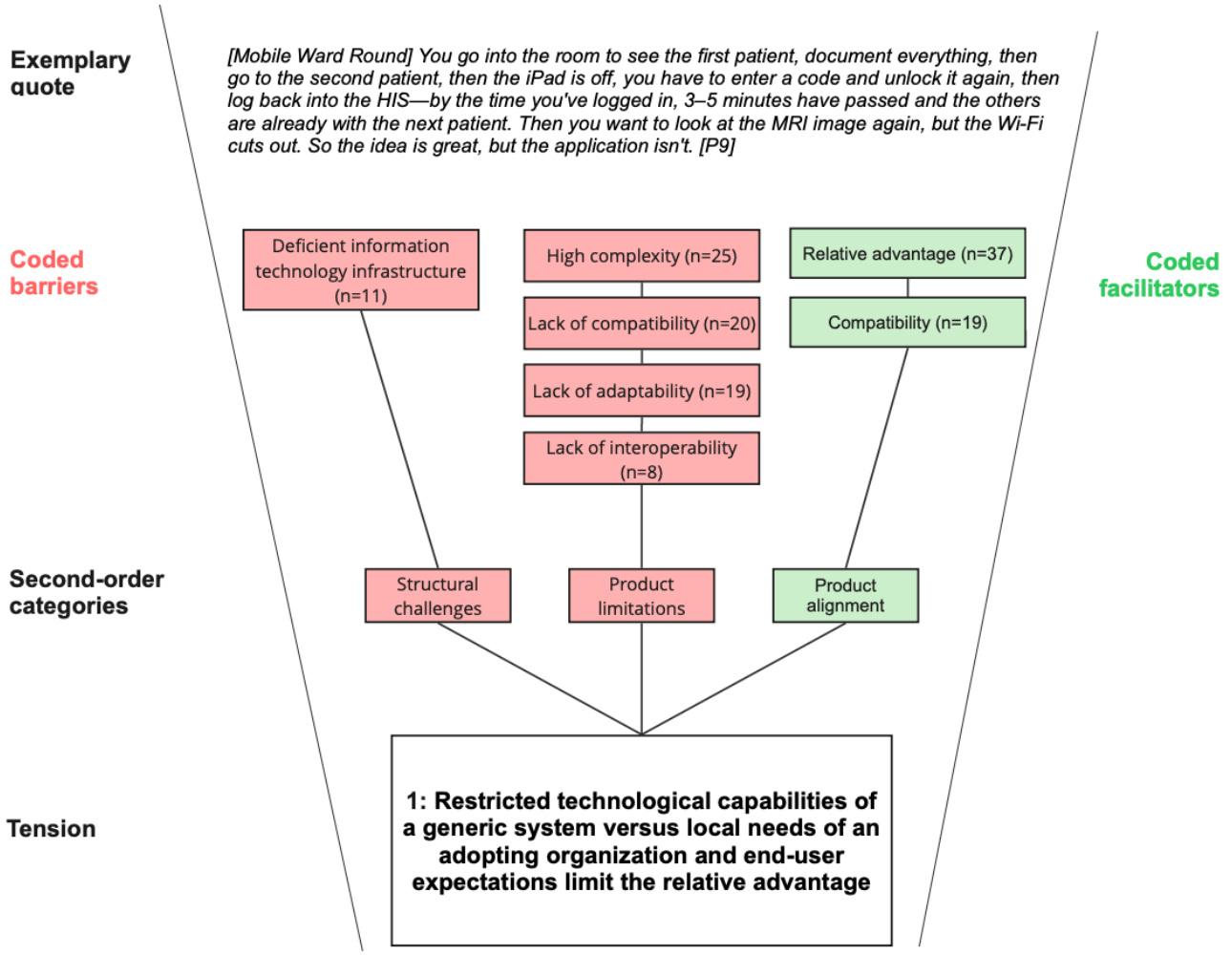
Tensions in MOT Implementation Resulting From Existent and Perceived Barriers and Facilitators

Organizational Tensions in MOT Implementation

Across all categories, our analysis revealed deeper, interrelated patterns beyond isolated barriers or facilitators, pointing to underlying organizational tensions that become apparent during MOT implementation and use. We identified 3 organizational tensions that characterize MOT implementation in complex academic hospital settings (see Figures 4-6).

XSL•FO
RenderX

Figure 4. Coding structure illustrating barriers and facilitators associated with tension 1 (generic system capabilities vs local clinical needs) in the implementation of modifiable off-the-shelf technologies at a large German academic hospital. Reported codes were aggregated into second-order categories that together capture the tension between standardized system design and the requirements of local clinical workflows. Exemplary quotes illustrate the underlying dynamics.



XSL•FO
RenderX

Figure 5. Coding structure illustrating barriers and facilitators associated with tension 2 (organizational complexity vs structured user participation) in the implementation of modifiable off-the-shelf technologies at a large German academic hospital. Reported codes were aggregated into second-order categories that together capture the tension between organizational oversight structures and desired user participation. Exemplary quotes illustrate the underlying dynamics.

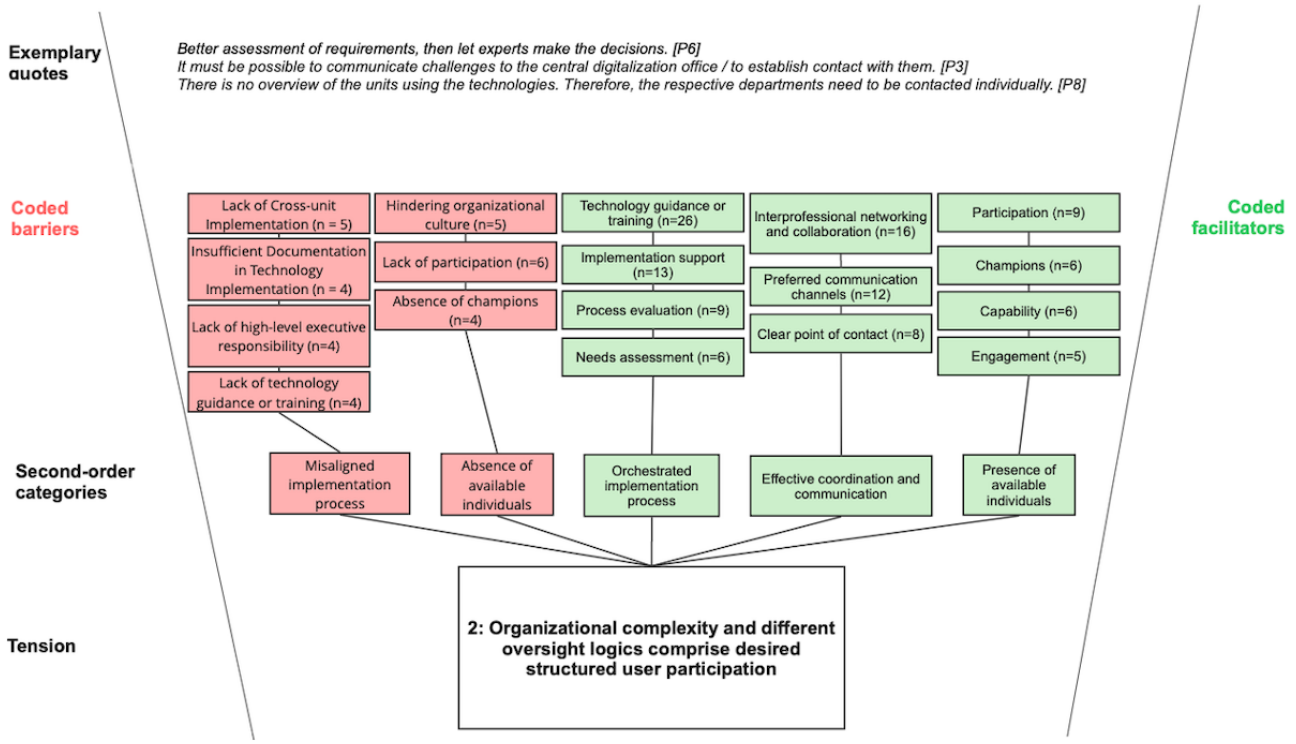
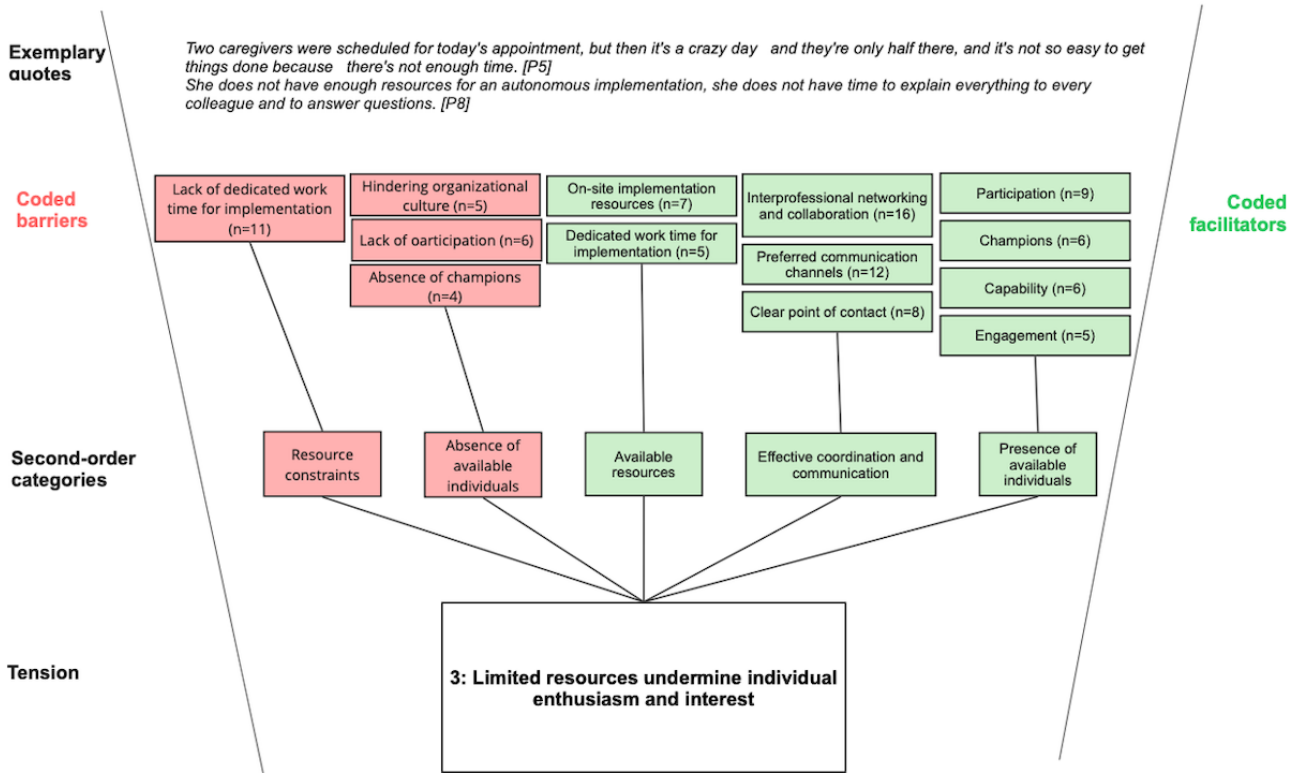


Figure 6. Coding structure illustrating barriers and facilitators associated with tension 3 (limited resources vs individual enthusiasm) in the implementation of modifiable off-the-shelf technologies at a large German academic hospital. Reported codes were aggregated into second-order categories that together capture the tension between constrained organizational resources and the motivation of clinicians and project teams to support implementation. Exemplary quotes illustrate the underlying dynamics.



XSL•FO
RenderX

Tension 1: Restricted Technological Capabilities of a Generic System Versus Local Needs of an Adopting Organization and End User Expectations Limit The Relative Advantage

While users expected a relative advantage from MOT use, product limitations emerged as a major barrier. Lack of interoperability and lack of compatibility led to fragmented data flows and incomplete functional integration, as vendor-driven system designs using preconfigured, standardized interfaces conflicted with the locally required data exchange in everyday care, which is often based on organization-specific terminology and interfaces. This created a tension because standardized system logics prioritized general applicability across diverse settings, whereas local care practices required flexible, context-specific integration of data flows. As a result, users were confronted with constant translation and alignment work to bridge the gap between the generic system architecture and their actual needs.

Closely related, the lack of adaptability and high complexity of MOTs produced usability constraints, forcing workaround practices as MOT use collided with the temporal, relational, and situational logics of clinical practice. This stood in tension with the expectation that digital systems should simplify and streamline clinical work, as MOTs instead introduced additional layers of cognitive and procedural effort. In practice, clinicians had to spend more time navigating complex interfaces or bypassing rigid functions, which diverted attention from patient care.

The emerging technological tension was further exacerbated by structural challenges, another major barrier category, particularly the hospital's deficient IT infrastructure. Ultimately, MOTs were not adopted when their potential efficiency and performance gains were outweighed by these barriers. These included high complexity, limited adaptability, and IT-infrastructural constraints, particularly a lack of interoperability and compatibility. Under such conditions, the relative advantage of MOTs was not apparent to users, which hindered their uptake.

Tension 2: Organizational Complexity and Different Oversight Logics Compromise Desired Structured User Participation

Clinicians emphasized the importance of active involvement before and during MOT implementation through preimplementation needs assessment, implementation support, process evaluation, and the use of preferred communication channels to share information effectively. Likewise, project managers highlighted their need for clear points of contact and designated champions for MOT implementation within departments. While participation was valued and desired by both groups, a hindering organizational culture was reported by clinicians and the implementation team. Clinicians, for example, felt that too much responsibility for MOT implementation was shifted to local clinical leaders, indicating a mismatch between assigned responsibilities and their perceived roles.

Participants perceived misaligned implementation processes—such as lack of technology guidance or training, insufficient documentation in technology implementation, and lack of cross-unit implementation—as barriers to their participation. These challenges may reflect differing oversight logics and governance structures shaped by managerial, economic, and regulatory requirements, such as the KHZG and data protection regulations, which had to be implemented within a short time frame, leaving limited opportunity for thorough planning and preparation. Specifically, rigid, project-bound management, hierarchical structures, and strained working climates reduced openness to participatory formats (eg, needs assessment) and hindered effective coordination and communication. This created a tension between the participatory processes that clinicians valued and the formal structures guiding implementation.

Tension 3: Limited Resources Undermine Individual Enthusiasm and Interest

While some participants expressed enthusiasm and willingness to participate, a major barrier to MOT implementation was resource constraints. Clinicians criticized the lack of dedicated work time they could allocate to implementation and highlighted their need for intensive on-site implementation support from project managers. Project managers, in turn, lacked champions in clinical departments for rollout processes and faced IT staff shortages and tight rollout schedules. This tension arises when clinicians are assigned responsibility as implementation leaders without having the necessary resources or identifying with that role. As a result, hospital managers experience tensions because they must operate under resource constraints while facing a persistent need for dedicated implementation resources.

Furthermore, interprofessional collaboration was recognized as a facilitator of MOT implementation because it was perceived to help align perspectives and create shared ownership. However, in practice, training and support were often organized in monoprofessional silos, meaning that doctors and nurses were trained separately. This reduced opportunities for joint learning and undermined the collaborative momentum required for sustained adoption. These structural and organizational barriers not only slowed implementation processes but also gradually weakened the initial enthusiasm of both clinicians and project managers. This reflects tensions between the demand to be innovative and collaborative and the lack of structures and resources needed to facilitate such collaboration.

Discussion

Principal Findings

This study identified 3 tensions during the implementation of MOTs.

Tension 1 concerned restricted technological capabilities that did not meet local needs and reduced the perceived relative advantage of the MOTs. Key barriers included product limitations, such as limited interoperability, low compatibility, low adaptability, and high complexity, as well as structural challenges that hindered smooth use. Product alignment acted as a facilitator when MOTs fit existing workflows.

XSL•FO
RenderX

Tension 2 arose from organizational complexity and differing oversight logics that compromised structured user participation. Barriers included a misaligned implementation process and the absence of available individuals to coordinate or support activities. Facilitators included an orchestrated implementation process, effective coordination and communication, and the presence of available individuals who could enable cross-unit participation.

Tension 3 reflected limited resources that undermined individual enthusiasm and interest in implementation. Barriers included resource constraints and a recurring absence of available individuals. Facilitators involved available resources, effective coordination and communication, and the presence of available individuals who could help maintain engagement.

The remainder of the "Discussion" section links these tensions and their respective barriers to misfit theory, introduces the concept of user:ability, and examines how 3 tailored strategies that integrate the observed facilitators respond to the organizational tensions in MOT implementation.

Misfits and the Role of User:Ability

Across all 3 tensions, the findings reveal that hospitals operate in a policy environment that seeks to accelerate digitalization while imposing strict timelines and compliance demands. Examples include Germany's KHZG and telematics infrastructure requirements [11,54], EU regulations such as General Data Protection Regulation (GDPR) [76] and the upcoming European Health Data Space [77], and US initiatives like HITECH [78] and the 21st Century Cures Act [79]. In this environment, hospitals often rely on MOTs, even though their standardized architecture restricts adjustments to functions, interfaces, or data flows. Limited involvement of end users during early procurement also contributes to systems that later need upgrades or expensive adjustments, as seen with the electronic health record in Germany [80-82]. These dynamics create fertile ground for organization-technology misfits as described by Strong and Volkoff [19]. Such misfits manifest in fragmented tools, disrupted workflows, higher error risk, and increased resource demands for both clinicians and IT teams [83]. In this study, multiple misfits are reflected in, and help explain, the tensions observed.

Tension 1 is characterized by functionality misfits, data misfits, and usability misfits when systems fail to reflect local workflows or information needs. Consistent with prior research, complex authentication [84], complicated user interface navigation [26], misalignment between systems and workflows [29,85], and limited interoperability [86] contributed to usability, data, and functionality misfits. Together, these factors limited the perceived usefulness of the technologies and thus discouraged adoption while also creating workarounds that made the misfits persistent [29,32,33,87,88]. Structural challenges, such as unreliable Wi-Fi, lack of functional devices [31,84], and difficulties complying with interoperability standards, reinforced these limitations.

Tension 2 reflected role and control misfits. Organizational complexity, unclear responsibilities, and inconsistent oversight made structured participation difficult [35,89]. These patterns mirror earlier studies showing that rigid governance structures and unclear responsibilities impede participation [19,90,91]. In our data, clinicians valued early involvement, whereas project managers emphasized clear contact points and department champions.

Tension 3 reflected role and organizational culture misfits. Clinicians were motivated but lacked time, recognition, and resources to take on implementation responsibilities. Resource constraints limited dedicated work time for both clinicians and project managers, a pattern consistent with the comparatively low IT investment levels in German hospitals [92]. Prior evidence supports that high workload and competing priorities force clinicians to prioritize patient care, limiting their capacity to meaningfully evaluate and adopt MOTs [93-95].

Taken together, these misfit patterns allow a focused reflection on the 3 assumptions from the "Introduction" section. The first assumption, that MOT implementation is shaped by functional and usability misfits, is confirmed but should be broadened. Functionality and usability mattered, but their effects were strengthened by role, control, and cultural misfits, showing that challenges arise from clusters of misfits rather than from product issues alone. The second assumption, that MOT implementation requires capacity-building conditions, is supported but needs refinement. Training, participation opportunities, and protected time are not optional but essential to counteract misfits caused by the limited modifiability of MOTs. However, capacity building alone cannot address structural or cultural misfits when organizations lack available individuals or coordination mechanisms. The third assumption, that structural characteristics intensify misfits, is also supported. Structural conditions not only amplified existing misfits; they also created new roles and cultural misfits by limiting participation, creating inconsistent responsibilities, and reducing available time. Overall, successful MOT implementation requires managing technological, organizational, and cultural misfits together, not only improving product alignment or increasing training.

Based on these findings, we introduce the concept of user:ability. User:ability describes the organizational conditions that enable users to adapt to MOTs effectively. It includes targeted training to clarify relative advantage, structured participation supported by clear communication, and protected time that allows users to engage in implementation tasks. Unlike related concepts such as user participation and user involvement, user:ability does not rely on early co-design, which is usually not possible with MOTs. Instead, it focuses on empowering users during implementation. Strengthening user:ability can help address the 3 tensions by improving perceived relative advantage (resolving tension 1), enabling participation in complex settings (resolving tension 2), and sustaining engagement despite limited resources (resolving tension 3). To operationalize user:ability, the following 3 strategies describe mechanisms to mitigate the tensions identified above, drawing on the facilitators observed in this study and integrating prior evidence.

XSL•FO
RenderX

Strategy 1: Ensuring Relative Advantage Through IT Integration, Product Alignment, and Targeted Learning

Strategy 1 focuses on reducing functionality, data, and usability misfits by strengthening the perceived relative advantage of the technologies through product alignment, IT integration, and targeted learning.

In this study, relative advantage and compatibility emerged as decisive factors for technology adoption, consistent with diffusion of innovations theory and the Consolidated Framework for Implementation Research [21,36,40]. Tailored learning approaches, including training, local implementation support, process evaluation, and microlearning formats, such as short instructional videos and question-and-answer sessions, helped clarify the benefits of MOT adoption and strengthened user confidence and acceptance [31,85,87,96-103]. As MOTs often require repeated upgrades and costly customizations even after rollout, organizations should integrate MOTs into the IT ecosystem with sufficient in-house IT expertise, enabling real-time access and workflow alignment [104].

Strategy 2: Empowering User Participation Through Digital Transformation and Structured Communication

Strategy 2 focuses on reducing role and control misfits by creating structures that enable meaningful participation through clear responsibilities, coordination mechanisms, and consistent communication.

Facilitators such as interprofessional networking and collaboration, clear points of contact, and preferred communication channels promoted knowledge-sharing, joint problem-solving, and shared ownership [31,97,100,101,103]. To strengthen these facilitators, hospitals can align IT and organizational goals through a digital business strategy and improve decision-making and coordination through a participatory process model that promotes aligned leadership, adaptive culture, and effective communication [105,106].

Structured and multimodal communication, including intranet resources, structured training, hybrid meetings, and targeted phone calls, can provide tailored support, clarification, and feedback [107]. These measures help align diverse stakeholders, foster trust, and enable coordination across units, thereby reducing role and control misfits.

Strategy 3: Enabling Clinician Engagement Through Resource Allocation and Systematic Implementation

Strategy 3 focuses on reducing role and cultural misfits by ensuring that individual enthusiasm can translate into sustained engagement through adequate time, support, and recognition.

Hospitals can address these constraints by allocating dedicated resources; providing clear points of contact for clinicians, project teams, and IT support; and structurally embedding clinicians across all phases of implementation [7]. By establishing structured roles such as clinical champions with protected time, for example, 4 hours per week [108], and clear cross-functional points of contact, organizations can sustain engagement, foster ownership, and enhance job satisfaction, supporting effective MOT adoption and evaluation [109-111].

This study contributes to implementation science by moving beyond the common focus on isolated barriers and facilitators. Instead of producing another descriptive list, the analysis integrates the findings into overarching misfits and organizational tensions. This perspective highlights how implementation challenges emerge from structural and relational dynamics rather than from single determinants. The study also introduces the concept of user:ability as an additional contribution. User:ability offers a way to understand how organizations can empower users in settings where technologies allow little adaptation. It shifts the focus from usability to organizational enablement and shows how targeted learning, structured participation, and protected time can help resolve core tensions. This conceptual addition extends existing implementation models by emphasizing the active role organizations play in enabling users to work effectively with MOTs.

While this study focused on barriers to implementing MOTs, the findings should not be interpreted as favoring in-house software development. In-house solutions can involve high development and maintenance costs, limited scalability, and long-term dependence on internal technical expertise and resource commitments [112]. By contrast, off-the-shelf systems benefit from vendor support, regulatory readiness, and standardized functionality, yet implementing them in complex clinical environments frequently requires extensive organizational adaptation to achieve local fit [22,113]. The findings, therefore, do not suggest the superiority of one technological strategy over the other but highlight the need for robust, participatory implementation processes and organizational alignment so that any technology, commercial or locally developed, can be meaningfully integrated into clinical practice.

Limitations

This study has several limitations that may influence interpretation and transferability. First, researchers held dual roles as implementers and evaluators, which is common in action research but may introduce observer bias. This was mitigated through reflexive practices, independent coding, and triangulation. Second, the analysis relies on frequency-based thresholds to structure qualitative data. Although the thresholds were transparently documented and grounded in established guidance for managing data overload, they may restrict the visibility of less frequent but potentially meaningful insights. As the analysis was conducted at the level of researcher protocols, the findings do not allow conclusions about participant-level prevalence or individual differences.

No demographic data were collected, which limits conclusions about individual differences. While this is methodologically acceptable in action research, where the focus is on collective learning and context-specific change [43], it constrains the ability to examine how individual characteristics may have shaped experiences. The findings, therefore, reflect shared group perspectives within the specific organizational context. Participation was voluntary and may have attracted digitally

XSL•FO
RenderX

engaged staff. However, even digitally motivated participants reported substantial barriers, suggesting that the identified tensions reflect systemic conditions rather than the views of a single subgroup. Still, less confident users may be underrepresented.

Some data relied on retrospective accounts, and the short data collection period limited insights into long-term adoption. Moreover, the focus on ready-made solutions may not fully capture upstream influences such as procurement procedures or funding structures. One included technology originated as a local initiative, which only partly aligns with the definition of MOTs but highlights the role of early user involvement.

The study did not include patient-facing technologies and did not collect patient perspectives, limiting insights into how digital transformation affects the care experience. Finally, the study took place in a single, highly resourced university hospital. Transferability is therefore limited, especially to smaller or less digitalized hospitals. Nonetheless, several observed tensions likely resonate across large and complex health care organizations. Future studies should examine these dynamics across hospitals with different structures and digital maturity levels.

Conclusions

This study demonstrates that, in the context of a German university hospital, the successful implementation of MOTs in hospitals depends less on improving usability and more on strengthening user:ability. This reframes digital transformation as an organizational challenge that requires structures that support learning, participation, and engagement when technologies cannot be adapted to local needs.

The 3 strategies derived from the findings provide context-specific guidance for creating such supportive conditions. Their relevance may extend to other large and complex health care institutions, although the implications should be interpreted in light of the single-site setting and the KHZG-funded implementation environment.

More broadly, hospitals must invest in conditions that enable users to work effectively with digital tools. This includes visible benefits for clinicians, coordinated implementation processes for project teams, and strategic resource allocation from hospital leadership.

These insights also extend beyond MOTs. Any digital technology requires alignment between structures, roles, and resources, so the choice between in-house or off-the-shelf solutions is secondary to an organization’s capacity to implement them effectively. Future research should examine how user:ability can be supported across diverse hospital settings to strengthen digital transformation efforts.

Acknowledgments

We gratefully acknowledge the valuable support of Hannah and Lynn during this study. We also extend our sincere thanks to the wards, units, and staff who contributed to data collection, with special appreciation to the nursing management and teams for their commitment and cooperation. GPT-4o (OpenAI) supported improvements to text fluency and grammar. All tool outputs were reviewed and revised by the authors, who take full responsibility for the submitted work.

Data Availability

The datasets generated and analyzed during this study are not publicly available because of data privacy; however, they are available from the corresponding author (LM) upon reasonable request. The qualitative data (field notes, workshop summaries, and focus group materials) generated and analyzed during this study are not publicly available because they contain sensitive information that could compromise participant confidentiality and institutional privacy. This paper includes selected anonymized participant quotations (anchor quotes) that illustrate key themes reported in the findings. No additional data are available for sharing.

Funding

Projects that made the evaluated technologies available for implementation were funded by the European Union through the Hospital Future Act (Krankenhauszukunftsgesetz), by financial support for the provision and maintenance of the telematics infrastructure negotiated between the National Association of Statutory Health Insurance Funds (GKV Spitzenverband) and the German Hospital Federation (Deutsche Krankenhausgesellschaft), and by internal digitalization budgets (regular institutional budgetary resources). Authors from the business units project management and CMIO were funded through the administrative budgets of their respective business units or by the European Union through the Hospital Future Act (Krankenhauszukunftsgesetz). No funder or funding body had any role in the study design, data collection, analysis, interpretation of results, or writing of the manuscript.

Authors' Contributions

Conceptualization: LM, FB, AN, CG
Data curation: BK, AN, CG
Formal analysis: BK, AN, CG, LM
Funding acquisition: FB, JCB

XSL•FO
RenderX

Investigation: AN, CG, AW, CdP, LM, EH, JCB, GK
Methodology: LM, AN, CG
Project administration: LM, CG, AN
Resources: AN, CG, GK, JCB
Supervision: LM, FB, ASP
Validation: AN, CG, BK, LM, EH, FB, CdP, AW, JCB, GK, DF, CK
Visualization: GK, BK, AN, CG, LM, DF
Writing – original draft: AN, CG, CK, EH, JCB, CdP, DF, ASP, GK, LM
Writing – review & editing: All authors

Conflicts of Interest

CG, AN, CK, and EH are researchers at the organization that launched and piloted the evaluated initiative, with no financial interests, no role on the JMIR editorial board, and no involvement in developing the technologies. LM is a participant in the BIH Charité Digital Clinician Scientist Program funded by Charité – Universitätsmedizin Berlin and the Berlin Institute of Health at Charité (BIH), and has received research funding from Masimo Inc and Philips. These contributions are independent of this project. DF is employed as a part-time researcher at the organization that launched and piloted the implementation initiative evaluated in this study; has received research funding from the European Commission, the German Federal Ministry of Research, Technology and Space, the German Federal Ministry of Health, the Berlin University Alliance, and the Einstein Foundation; has received support for attending meetings or travel from the German Academic Exchange Service. These contributions are independent of this project; and has no financial interests in the initiative or related products, has not served on the JMIR editorial board, and was not involved in the development of the evaluated technologies. FB has received research funding from the Hans Böckler Foundation, Berlin University Alliance, Charité Foundation, Joint Federal Committee, European Commission, Fresenius Foundation, the German Federal Ministry of Research, Technology and Space, the German Federal Ministry of Health, the German Research Foundation, Honda Research Foundation, and Philips; and has received honoraria from Pfizer and support for attending meetings or travel from the Association of German Anesthetists, DGAI, the German Society of Anesthesiology and Intensive Care Medicine, Masimo, and the Bavarian Academy of Sciences and Humanities. ASP has received research funding from the German Federal Ministry of Education and Research (BMBF), the European Society of Intensive Care Medicine, and the medical technology companies Masimo and Dräger; serves in an advisory capacity to Philips; and is a co-founder and managing director of INCH Health GmbH. BK declares no financial interests, no involvement in JMIR editorial activities, and no role in developing the evaluated technologies. JCB is employed as a project manager at the organization that launched and piloted the implementation initiative evaluated in this study and has no competing financial interests in the initiative or related products. GK is a digitalization manager within the CMIO at the organization that launched and piloted the initiative and declares no competing financial interests. CdP is a digital manager at the organization that launched and piloted the initiative and declares no competing financial interests. AW is employed as a digital manager and nurse at the organization that launched and piloted the initiative and declares no competing financial interests.

Multimedia Appendix 1
Standards for Reporting Qualitative Research (SRQR) checklist.
[DOCX File , 38 KB - humanfactors_v13i1e84841_app1.docx]

Multimedia Appendix 2
Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist.
[DOCX File , 24 KB - humanfactors_v13i1e84841_app2.docx]

Multimedia Appendix 3
Units of study.
[DOCX File , 24 KB - humanfactors_v13i1e84841_app3.docx]

Multimedia Appendix 4
Observation guide.
[DOCX File , 32 KB - humanfactors_v13i1e84841_app4.docx]

Multimedia Appendix 5
Table of code frequencies.
[DOCX File , 55 KB - humanfactors_v13i1e84841_app5.docx]

Multimedia Appendix 6

XSL•FO
RenderX

Codebook.
[DOCX File , 61 KB - humanfactors_v13i1e84841_app6.docx]

References

1. Bardhan I, Chen H, Karahanna E. Connecting systems, data, and people: a multidisciplinary research roadmap for chronic disease management. MIS Quarterly: Management Information Systems 2020 Mar 01;44(1):185-200. [doi: 10.25300/misq/2020/14644]
2. Sunyaev A, Fürstenau D, Davidson E. Reimagining digital health. Bus Inf Syst Eng 2024 May 27;66(3):249-260. [doi: 10.1007/s12599-024-00870-x]
3. Borges do Nascimento IJ, Abdulazeem HM, Vasanthan LT, Martinez EZ, Zucoloto ML, Østengaard L, et al. The global effect of digital health technologies on health workers' competencies and health workplace: an umbrella review of systematic reviews and lexical-based and sentence-based meta-analysis. Lancet Digit Health 2023 Aug;5(8):e534-e544 [FREE Full text] [doi: 10.1016/S2589-7500(23)00092-4] [Medline: 37507197]
4. Endalamaw A, Zewdie A, Wolka E, Assefa Y. A scoping review of digital health technologies in multimorbidity management: mechanisms, outcomes, challenges, and strategies. BMC Health Serv Res 2025 Mar 15;25(1):382 [FREE Full text] [doi: 10.1186/s12913-025-12548-5] [Medline: 40089752]
5. Campanella P, Lovato E, Marone C, Fallacara L, Mancuso A, Ricciardi W, et al. The impact of electronic health records on healthcare quality: a systematic review and meta-analysis. Eur J Public Health 2016 Feb 30;26(1):60-64. [doi: 10.1093/eurpub/ckv122] [Medline: 26136462]
6. Thompson S, Whitaker J, Kohli R, Jones C. Chronic disease management: how IT and analytics create healthcare value through the temporal displacement of care. MISQ 2020 Jan 01;44(1):227-256. [doi: 10.25300/misq/2020/15085]
7. Kraus S, Schiavone F, Pluzhnikova A, Invernizzi AC. Digital transformation in healthcare: analyzing the current state-of-research. Journal of Business Research 2021 Feb;123:557-567. [doi: 10.1016/j.jbusres.2020.10.030]
8. Insani WN, Zakiyah N, Puspitasari IM, Permana MY, Parmikanti K, Rusyaman E, et al. Digital health technology interventions for improving medication safety: systematic review of economic evaluations. J Med Internet Res 2025 Feb 05;27:e65546 [FREE Full text] [doi: 10.2196/65546] [Medline: 39909404]
9. Woods L, Eden R, Green D, Pearce A, Donovan R, McNeil K, et al. Impact of digital health on the quadruple aims of healthcare: a correlational and longitudinal study (Digimat Study). Int J Med Inform 2024 Sep;189:105528 [FREE Full text] [doi: 10.1016/j.ijmedinf.2024.105528] [Medline: 38935999]
10. Gates P, Hardie R, Raban M, Li L, Westbrook J. How effective are electronic medication systems in reducing medication error rates and associated harm among hospital inpatients? A systematic review and meta-analysis. J Am Med Inform Assoc 2021 Jan 15;28(1):167-176 [FREE Full text] [doi: 10.1093/jamia/ocaa230] [Medline: 33164058]
11. Gesetz für ein Zukunftsprogramm Krankenhäuser (KHZG). Bundesgesundheitsministerium. 2020 Oct 23. URL: https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/Gesetze_und_Verordnungen/GuV/K/bgbl1_S.2208_KHZG_28.10.20.pdf [accessed 2026-02-16]
12. Doelfs G. Streit um enge fristen: kliniken im wettlauf gegen KHZG-Strafzahlungen. kma-online. 2023 May 23. URL: https://www.kma-online.de/aktuelles/it-digital-health/detail/kliniken-im-wettlauf-gegen-khzg-strafzahlungen-49902 [accessed 2025-08-23]
13. Heeser A. Krankenhauszukunftsgesetz: das KHZG und der Hürdenlauf. Wie geht es 2025 weiter? kma-online. 2024 May 17. URL: https://www.kma-online.de/aktuelles/it-digital-health/detail/das-khzg-und-der-huerdenlauf-wie-geht-es-2025-weiter-51987 [accessed 2025-08-23]
14. Svempe L. Exploring impediments imposed by the Medical Device Regulation EU 2017/745 on software as a medical device. JMIR Med Inform 2024 Sep 05;12(1):e58080 [FREE Full text] [doi: 10.2196/58080] [Medline: 39235850]
15. Miclăuş T, Valla V, Koukoura A, Nielsen AA, Dahlerup B, Tsianos G, et al. Impact of design on medical device safety. Ther Innov Regul Sci 2020 Jul 9;54(4):839-849 [FREE Full text] [doi: 10.1007/s43441-019-00022-4] [Medline: 32557299]
16. Mousavidin E, Silva L. Theorizing the configuration of modifiable off-the-shelf software. ITP 2017 Nov 06;30(4):887-909. [doi: 10.1108/itp-07-2014-0140]
17. Lucas H, Walton E, Ginzberg M. Implementing packaged software. MIS Quarterly 1988;12(4):537-549. [doi: 10.2307/249129]
18. Fürstenau D, Gersch M, Schreiter S. Digital therapeutics (DTx). Bus Inf Syst Eng 2023 Apr 06;65(3):349-360. [doi: 10.1007/s12599-023-00804-z]
19. Strong, Volkoff. Understanding organization—enterprise system fit: a path to theorizing the information technology artifact. MIS Quarterly 2010 Dec;34(4):731-756. [doi: 10.2307/25750703]
20. Safi S, Thiessen T, Schmailzl KJ. Acceptance and resistance of new digital technologies in medicine: qualitative study. JMIR Res Protoc 2018 Dec 04;7(12):e11072 [FREE Full text] [doi: 10.2196/11072] [Medline: 30514693]
21. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated Consolidated Framework for Implementation Research based on user feedback. Implement Sci 2022 Oct 29;17(1):75 [FREE Full text] [doi: 10.1186/s13012-022-01245-0] [Medline: 36309746]

XSL•FO
RenderX

22. Marwaha JS, Landman AB, Brat GA, Dunn T, Gordon WJ. Deploying digital health tools within large, complex health systems: key considerations for adoption and implementation. NPJ Digit Med 2022 Jan 27;5(1):13 [FREE Full text] [doi: 10.1038/s41746-022-00557-1] [Medline: 35087160]
23. Li M. An approach to addressing the usability and local relevance of generic enterprise software. Association for Information Systems (eLibrary). 2019. URL: https://aisel.aisnet.org/iris2019/3 [accessed 2026-02-25]
24. Kaipio J, Lääveri T, Hyppönen H, Vainiomäki S, Reponen J, Kushniruk A, et al. Usability problems do not heal by themselves: national survey on physicians' experiences with EHRs in Finland. Int J Med Inform 2017 Jan;97:266-281 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.10.010] [Medline: 27919385]
25. Garcia L, Mendonça G, Benedetti TRB, Borges LJ, Streit IA, Christofoletti M, et al. Barriers and facilitators of domain-specific physical activity: a systematic review of reviews. BMC Public Health 2022 Oct 26;22(1):1964 [FREE Full text] [doi: 10.1186/s12889-022-14385-1] [Medline: 36289461]
26. Zhai Y, Yu Z, Zhang Q, Zhang Y. Barriers and facilitators to implementing a nursing clinical decision support system in a tertiary hospital setting: a qualitative study using the FITT framework. Int J Med Inform 2022 Oct;166:104841. [doi: 10.1016/j.ijmedinf.2022.104841] [Medline: 36027798]
27. Jacob C, Sanchez-Vazquez A, Ivory C. Clinicians' role in the adoption of an oncology decision support app in Europe and its implications for organizational practices: qualitative case study. JMIR mHealth and uHealth 2019 May 03;7(5):e13555 [FREE Full text] [doi: 10.2196/13555] [Medline: 31066710]
28. Giraldo L, Schachner B, Luna D, Benítez S. Exploring nurses' perceptions and expectations toward a BCMA implementation using a mobile app and workstations as a change management strategy. Stud Health Technol Inform 2018;250:134-138. [Medline: 29857405]
29. Seibert K, Domhoff D, Huter K, Krick T, Rothgang H, Wolf-Ostermann K. Application of digital technologies in nursing practice: results of a mixed methods study on nurses' experiences, needs and perspectives. Z Evid Fortbild Qual Gesundhwes 2020 Dec 19;158-159:94-106 [FREE Full text] [doi: 10.1016/j.zefq.2020.10.010] [Medline: 33223491]
30. Kleinpell R, Barden C, Rincon T, McCarthy M, Zapatochny Rufo RJ. Assessing the impact of telemedicine on nursing care in intensive care units. Am J Crit Care 2016 Jan 31;25(1):e14-e20. [doi: 10.4037/ajcc2016808] [Medline: 26724303]
31. Delaforce A, Li J, Grujovski M, Parkinson J, Richards P, Fahy M, et al. Creating an implementation enhancement plan for a digital patient fall prevention platform using the CFIR-ERIC approach: a qualitative study. Int J Environ Res Public Health 2023 Feb 21;20(5):3794 [FREE Full text] [doi: 10.3390/ijerph20053794] [Medline: 36900804]
32. Jacob C, Sanchez-Vazquez A, Ivory C. Factors impacting clinicians' adoption of a clinical photo documentation app and its implications for clinical workflows and quality of care: qualitative case study. JMIR Mhealth Uhealth 2020 Sep 23;8(9):e20203 [FREE Full text] [doi: 10.2196/20203] [Medline: 32965232]
33. Wosny M, Strasser LM, Hastings J. The paradoxes of digital tools in hospitals: qualitative interview study. J Med Internet Res 2024 Jul 15;26:e56095. [doi: 10.2196/56095]
34. Miron-Spektor E, Ingram A, Keller J, Smith WK, Lewis MW. Microfoundations of organizational paradox: the problem is how we think about the problem. AMJ 2018 Feb;61(1):26-45. [doi: 10.5465/amj.2016.0594]
35. Haring M, Freigang F, Amelung V, Gersch M. What can healthcare systems learn from looking at tensions in innovation processes? A systematic literature review. BMC Health Serv Res 2022 Oct 28;22(1):1299 [FREE Full text] [doi: 10.1186/s12913-022-08626-7] [Medline: 36307839]
36. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
37. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. Journal of Medical Internet Research 2017 Nov 01;19(11):e8775 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
38. Glasgow RE, Harden SM, Gaglio B, Rabin B, Smith ML, Porter GC, et al. RE-AIM planning and evaluation framework: adapting to new science and practice with a 20-year review. Front Public Health 2019 Mar 29;7:64 [FREE Full text] [doi: 10.3389/fpubh.2019.00064] [Medline: 30984733]
39. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
40. Rogers E, Singhal A, Quinlan M. Diffusion of Innovations. In: An Integrated Approach to Communication Theory and Research. New York, NY: Routledge; 2014:432-448.
41. Moullin JC, Dickson KS, Stadnick NA, Albers B, Nilsen P, Broder-Fingert S, et al. Ten recommendations for using implementation frameworks in research and practice. Implement Sci Commun 2020 Apr 30;1(1):42 [FREE Full text] [doi: 10.1186/s43058-020-00023-7] [Medline: 32885199]
42. Graham ID, Logan J, Harrison MB, Straus SE, Tetroe J, Caswell W, et al. Lost in knowledge translation: time for a map? Journal of Continuing Education in the Health Professions 2006;26(1):13-24. [doi: 10.1002/chp.47] [Medline: 16557505]
43. Rapoport R. Three dilemmas in action research: with special reference to the Tavistock experience. Human Relations 1970 Dec 01;23(6):499-513. [doi: 10.1177/001872677002300601]

XSL•FO
RenderX

44. Köberlein-Neu J, Söling S. Implementierung von digitalen technologien in der patientenversorgungine frage der passenden strategie(n). In: Technologien und Technologiemanagement im Gesundheitswesen. Wiesbaden, Germany: Springer Fachmedien Wiesbaden; 2024:431-454.
45. Schwandt T. Constructivist, interpretivist approaches to human inquiry. In: Handbook of Qualitative Research: Theories and Issues. Thousand Oaks, CA: Sage; Jan 01, 1994:221-259.
46. Ponterotto JG. Qualitative research in counseling psychology: a primer on research paradigms and philosophy of science. Journal of Counseling Psychology 2005 Apr;52(2):126-136. [doi: 10.1037/0022-0167.52.2.126]
47. Hult M, Lennung S. Towards a definition of action research: a note and bibliography. J Management Studies 1980 May;17(2):241-250. [doi: 10.1111/j.1467-6486.1980.tb00087.x]
48. Kemmis S, McTaggart R, Nixon R. The Action Research Planner: Doing Critical Participatory Action Research. Singapore: Springer Singapore; Nov 12, 2013.
49. Walsham G. Interpretive case studies in IS research: nature and method. European Journal of Information Systems 2017 Dec 19;4(2):74-81. [doi: 10.1057/ejis.1995.9]
50. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. International Journal for Quality in Health Care 2007 Sep 16;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
51. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251. [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
52. Charité. 2024. URL: https://www.charite.de/die_charite/profil/zahlen_fakten [accessed 2025-08-07]
53. Charité strategie 2030. Charité. URL: https://www.charite.de/die_charite/profil/charite_strategie_2030 [accessed 2025-06-10]
54. Bundesamt für Justiz. Sozialgesetzbuch (SGB V) § 306 – telematikinfrastruktur. sozialgesetzbuch SGB. URL: https://www.sozialgesetzbuch-sgb.de/sgbv/306.html [accessed 2025-09-23]
55. eAU. Gematik GmbH. URL: https://www.gematik.de/anwendungen/kim/faq?faqc=47 [accessed 2025-08-29]
56. Nation-wide electronic health record. Gematik GmbH. 2025. URL: https://www.gematik.de/anwendungen/epa-fuer-alle [accessed 2025-08-29]
57. ePrescription. e-Rezept. 2025. URL: https://www.das-e-rezept-fuer-deutschland.de/en [accessed 2025-08-29]
58. ID MEDICS – closed loop: medication support from the hospital information system. ID Berlin. 2025. URL: https://www.id-berlin.de/en/products/emedikation/id-medics/ [accessed 2025-08-29]
59. Kommunikation im medizinwesen. Gematik GmBH. 2025. URL: https://www.gematik.de/anwendungen/kim [accessed 2025-08-29]
60. SAP help portal – SAP for healthcare documentation. SAP. URL: https://help.sap.com/docs/SAP_FOR_HEALTHCARE/7d223867a0ad462a8665d1f9531eb5e9/4dd3306d81a60b63e10000000a42189b.html?version=618.26&q=implantatpass [accessed 2025-08-27]
61. Common MS. eMed: die perfekte mobile Lösung für Cerner i.s.h.med®. Common MS. 2022. URL: https://commonms.com/de/gesundheit/interoperabilitat/ [accessed 2025-08-27]
62. SAP help portal – SAP for healthcare. SA. URL: https://help.sap.com/docs/SAP_FOR_HEALTHCARE/f3a8c1fd3c8748df9bdaeb1736b387f1/4d958010d1b83c46e10000000a42189e-846.html?version=604.74 [accessed 2025-08-27]
63. Flüssigkeitsbilanz – SAP help portal. SAP. URL: https://help.sap.com/docs/SAP_FOR_HEALTHCARE/7d223867a0ad462a8665d1f9531eb5e9/14ecbd1891064031a331e23f5904ac6f.html?version=618.26 [accessed 2025-08-27]
64. Wenn idee auf initiative trifft: der stationsbuddy: Charité. Charité. 2025. URL: https://karriere.charite.de/karrieremagazin/stationsbuddy [accessed 2025-08-27]
65. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Administration and Policy in Mental Health 2015 Sep;42(5):533-544 [FREE Full text] [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
66. Pailthorpe B. Emergent design. In: Matthes J, Davis C, Potter R, editors. The International Encyclopedia of Communication Research Methods. Hoboken, NJ: Wiley; Nov 07, 2017:1-2.
67. Morgan DL. Focus groups. Annu Rev Sociol 1996 Aug 01;22(1):129-152. [doi: 10.1146/annurev.soc.22.1.129]
68. Storvang P, Mortensen B, Clarke A. Using workshops in business research: a framework to diagnose, plan, facilitate and analyze workshops. In: Freytag P, Young L, editors. Collaborative Research Design: Working with Business for Meaningful Findings. Singapore: Springer; 2018:155-174.
69. Mulhall A. In the field: notes on observation in qualitative research. J Adv Nurs 2003 Feb 04;41(3):306-313. [doi: 10.1046/j.1365-2648.2003.02514.x] [Medline: 12581118]
70. Dreborg KH. Essence of backcasting. Futures 1996 Nov;28(9):813-828. [doi: 10.1016/s0016-3287(96)00044-4]
71. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qualitative Health Research 2015 Nov 27;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
72. MAXQDA 2022. VERBI GmbH. URL: https://www.maxqda.com/de/about [accessed 2025-08-07]
73. Kuckartz U, Rädiker S. Qualitative Inhaltsanalyse. Methoden, Praxis, Umsetzung mit Software und künstlicher Intelligenz. 6. Auflage. Weinheim, Germany: Beltz Juventa; 2024.

XSL•FO
RenderX

74. Miles M, Huberman A. Qualitative Data Analysis: An Expanded Sourcebook. Thousand Oaks, CA: Sage Publications; 1994.
75. Carter N, Bryant-Lukosius D, DiCenso A, Blythe J, Neville AJ. The use of triangulation in qualitative research. Oncology Nursing Forum 2014 Sep 01;41(5):545-547. [doi: 10.1188/14.ONF.545-547] [Medline: 25158659]
76. Datenschutz-Grundverordnung (General Data Protection Regulation). European Parliament and of the Council. Brussels, Belgium: European Union; 2016. URL: https://dsgvo-gesetz.de/ [accessed 2025-09-23]
77. European Parliament and of the Council. Regulation (EU) 2025/327 on the European Health Data Space (EHDS). EUR-Lex. Brussels, Belgium: European Union; 2025. URL: https://eur-lex.europa.eu/eli/reg/2025/327/oj/eng [accessed 2025-09-23]
78. Public Law 111-5. Senate and House of Representatives of the United States of America. Washington, DC: US Government Publishing Office; 2009. URL: https://www.govinfo.gov/content/pkg/PLAW-111publ5/pdf/PLAW-111publ5.pdf [accessed 2025-09-23]
79. Public Law-114-255. Senate and House of Representatives of the United States of America. 2016 Dec 13. URL: https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf [accessed 2025-09-23]
80. Tsai CH, Eghdam A, Davoody N, Wright G, Flowerday S, Koch S. Effects of electronic health record implementation and barriers to adoption and use: a scoping review and qualitative analysis of the content. Life (Basel) 2020 Dec 04;10(12):327 [FREE Full text] [doi: 10.3390/life10120327] [Medline: 33291615]
81. Uncertainty surrounds electronic patient record rollout in Germany. The Munich Eye. 2025 Apr 03. URL: https://themunicheye.com/delay-in-electronic-patient-record-rollout-germany-15435 [accessed 2025-08-28]
82. Elektronische patientenakte: ePA für alle. Bundesministerium für Gesundheit. 2025. URL: https://www.bundesgesundheitsministerium.de/epa-vorteile/ [accessed 2025-08-18]
83. Carlos Ferreira J, Elvas LB, Correia R, Mascarenhas M. Enhancing EHR interoperability and security through distributed ledger technology: a review. Healthcare (Basel) 2024 Oct 02;12(19):1967 [FREE Full text] [doi: 10.3390/healthcare12191967] [Medline: 39408147]
84. Duhm J, Fleischmann R, Schmidt S, Hupperts H, Brandt SA. Mobile electronic medical records promote workflow: physicians' perspective from a survey. JMIR Mhealth Uhealth 2016 Jun 06;4(2):e70 [FREE Full text] [doi: 10.2196/mhealth.5464] [Medline: 27268720]
85. Leenen JPL, Rasing HJM, Kalkman CJ, Schoonhoven L, Patijn GA. Process evaluation of a wireless wearable continuous vital signs monitoring intervention in 2 general hospital wards: mixed methods study. JMIR Nurs 2023 May 04;6:e44061 [FREE Full text] [doi: 10.2196/44061] [Medline: 37140977]
86. Torab-Miandoab A, Samad-Soltani T, Jodati A, Rezaei-Hachesu P. Interoperability of heterogeneous health information systems: a systematic literature review. BMC Medical Informatics and Decision Making 2023 Jan 24;23(1):18 [FREE Full text] [doi: 10.1186/s12911-023-02115-5] [Medline: 36694161]
87. Mosch LK, Poncette A, Spies C, Weber-Carstens S, Schieler M, Krampe H, et al. Creation of an evidence-based implementation framework for digital health technology in the intensive care unit: qualitative study. JMIR Form Res 2022 Apr 08;6(4):e22866 [FREE Full text] [doi: 10.2196/22866] [Medline: 35394445]
88. van Offenbeek MAG, Vos JFJ, van den Hooff B, Boonstra A. When workarounds aggravate misfits in the use of electronic health record systems. Information Systems Journal 2023 Oct 13;34(2):293-326. [doi: 10.1111/isj.12478]
89. Hansen S, Baroody AJ. Electronic health records and the logics of care: complementarity and conflict in the U.S. healthcare system. Information Systems Research 2020 Mar 01;31(1):57-75. [doi: 10.1287/isre.2019.0875]
90. Shahzad K, Imran F, Butt A. Digital transformation and changes in organizational structure. Research-Technology Management 2025 Apr 09;68(3):25-40. [doi: 10.1080/08956308.2025.2465706]
91. Verhoef PC, Broekhuizen T, Bart Y, Bhattacharya A, Qi Dong J, Fabian N, et al. Digital transformation: a multidisciplinary reflection and research agenda. Journal of Business Research 2021 Jan 01;122:889-901. [doi: 10.1016/j.jbusres.2019.09.022]
92. Liebe J, Reincke L, Buddrus U. IT-ausgaben in krankenhäusern im internationalen vergleich: Deutschland, Dänemark und die Niederlande. Science to Business GmbH – Hochschule Osnabrück. Osnabrück, Germany: Science to Business GmbH – Hochschule Osnabrück; 2025. URL: https://www.dkgev.de/fileadmin/default/Mediapool/1_DKG/1.7_Presse/1.7.1_Pressemitteilungen/2025/2025-07-30_Anlage_Studie_int._Vergleich_von_IT-Ausgaben_in_Krankenhaeusern.pdf [accessed 2025-08-23]
93. Borges do Nascimento IJ, Abdulazeem H, Vasanthan LT, Martinez EZ, Zucoloto ML, Østengaard L, et al. Barriers and facilitators to utilizing digital health technologies by healthcare professionals. npj Digital Medicine 2023 Sep 18;6(1):161. [doi: 10.1038/s41746-023-00899-4]
94. Zharima C, Mhlanga S, Abdulla S, Goudge J, Griffiths F. What engagement strategies are useful in facilitating the implementation of electronic health records in health care settings? A rapid review of qualitative evidence synthesis using the normalization process theory. Digit Health 2024;10:1-28 [FREE Full text] [doi: 10.1177/20552076241291286] [Medline: 39497787]
95. Grignoli N, Manoni G, Gianini J, Schulz P, Gabutti L, Petrocchi S. Clinical decision fatigue: a systematic and scoping review with meta-synthesis. Fam Med Community Health 2025 Feb 08;13(1):e003033 [FREE Full text] [doi: 10.1136/fmch-2024-003033] [Medline: 39922690]

XSL•FO
RenderX

96. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb;10(1):21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
97. Barchielli C, Marullo C, Bonciani M, Vainieri M. Nurses and the acceptance of innovations in technology-intensive contexts: the need for tailored management strategies. BMC Health Serv Res 2021 Jul 03;21(1):639. [doi: 10.1186/s12913-021-06628-5]
98. Connolly SL, Kuhn E, Possemato K, Torous J. Digital clinics and mobile technology implementation for mental health care. Curr Psychiatry Rep 2021 May 07;23(7):38. [doi: 10.1007/s11920-021-01254-8]
99. Verma N, Mamlin B, Flowers J, Acharya S, Labrique A, Cullen T. OpenMRS as a global good: impact, opportunities, challenges, and lessons learned from fifteen years of implementation. International Journal of Medical Informatics 2021 May;149:104405. [doi: 10.1016/j.ijmedinf.2021.104405]
100. Olesen ML, Rossen S, Jørgensen R, Langballe Udbjørg L, Hansson H. Usefulness of a digitally assisted person-centered care intervention: qualitative study of patients' and nurses' experiences in a long-term perspective. JMIR Nurs 2023 May 18;6:e46673 [FREE Full text] [doi: 10.2196/46673] [Medline: 37200076]
101. Hofstetter S, Zilezinski M, Behr D, Kraft B, Buhtz C, Paulicke D, et al. Integrating digital assistive technologies into care processes: mixed methods study. JMIR Med Educ 2024 Oct 09;10:e54083 [FREE Full text] [doi: 10.2196/54083] [Medline: 39383526]
102. Alotaibi N, Wilson CB, Traynor M. Enhancing digital readiness and capability in healthcare: a systematic review of interventions, barriers, and facilitators. BMC Health Serv Res 2025 Apr 04;25(1):500 [FREE Full text] [doi: 10.1186/s12913-025-12663-3] [Medline: 40186200]
103. Claessens D, Boudewijns EA, Vervloet M, Keijsers LCEM, Gidding-Slok AHM, van Schayck OCP, et al. Barriers and facilitators to the implementation of the Assessment of Burden of Chronic Conditions tool in Dutch primary care: a context analysis. BMJ Open 2025 Jan 20;15(1):e087197 [FREE Full text] [doi: 10.1136/bmjopen-2024-087197] [Medline: 39833000]
104. Mathias R, McCulloch P, Chalkidou A, Gilbert S. Digital health technologies need regulation and reimbursement that enable flexible interactions and groupings. NPJ Digit Med 2024 Jun 18;7(1):148 [FREE Full text] [doi: 10.1038/s41746-024-01147-z] [Medline: 38890404]
105. Hess T, Matt C, Benlian A, Wiesböck F. Options for formulating a digital transformation strategy. In: Strategic Information Management. New York, NY: Routledge; 2020:151-173.
106. Bharadwaj A, El Sawy OA, Pavlou PA, Venkatraman N. Digital business strategy: toward a next generation of insights. MISQ 2013 Feb 2;37(2):471-482. [doi: 10.25300/misq/2013/37:2.3]
107. Adjekum A, Blasimme A, Vayena E. Elements of trust in digital health systems: scoping review. J Med Internet Res 2018 Dec 13;20(12):e11254 [FREE Full text] [doi: 10.2196/11254] [Medline: 30545807]
108. Bunce AE, Gruß I, Davis JV, Cowburn S, Cohen D, Oakley J, et al. Lessons learned about the effective operationalization of champions as an implementation strategy: results from a qualitative process evaluation of a pragmatic trial. Implement Sci 2020 Oct 01;15(1):87 [FREE Full text] [doi: 10.1186/s13012-020-01048-1] [Medline: 32998750]
109. Wieslander L, Bäckström I, Häggström M. Participation in the digital transformation of healthcare: a review of qualitative studies. Int J Health Care Qual Assur 2024 Nov 15;37(3-4):68-84. [doi: 10.1108/IJHCQA-03-2024-0021] [Medline: 39541252]
110. Galazzi A, Fonda F, Chiappinotto S, Justi L, Sønderskov Frydensberg M, Lehmann Boesen R, et al. Recommendations to promote the digital healthcare transformation in the clinical practice: findings from an international consensus development method. BMC Health Serv Res 2025 Jul 04;25(1):929 [FREE Full text] [doi: 10.1186/s12913-025-13079-9] [Medline: 40616118]
111. Hansen A. Rapid adaptation in digital tRansfoRmation: a paRticipatoRy pRocess foR engaging is and Business leadeRs. MIS Quarterly Executive 2011 Dec 4;10(4):175-185 [FREE Full text]
112. Amlung J, Huth H, Cullen T, Sequist T. Modernizing health information technology: lessons from healthcare delivery systems. JAMIA Open 2020 Oct 01;3(3):369-377. [doi: 10.1093/jamiaopen/ooaa027]
113. Takian A, Sheikh A, Barber N. Organizational learning in the implementation and adoption of national electronic health records: case studies of two hospitals participating in the National Programme for Information Technology in England. Health Informatics J 2014 Sep 02;20(3):199-212 [FREE Full text] [doi: 10.1177/1460458213493196] [Medline: 25183608]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
DHT: digital health technology
GDPR: General Data Protection Regulation
KHZG: Krankenhauszukunftsgesetz (Hospital Future Act)
MOT: modifiable off-the-shelf technology
P: protocol
SRQR: Standards for Reporting Qualitative Research

XSL•FO
RenderX

Edited by C LeRouge; submitted 25.Sep.2025; peer-reviewed by R Reid, C Yi; comments to author 30.Oct.2025; accepted 21.Feb.2026; published 13.May.2026.

Please cite as:
Nowak A, Gaul C, Hinz E, Knoll C, Kül B, Wels A, Berkmann JC, de Pasquale C, Kremser G, Fürstenau D, Poncette AS, Balzer F, Mosch L
Organizational Tensions in the Implementation of Modifiable Off-the-Shelf Technologies in a University Hospital: Qualitative Multimethod Study
JMIR Hum Factors 2026;13:e84841
URL: https://humanfactors.jmir.org/2026/1/e84841
doi:10.2196/84841
PMID:

©Alessia Nowak, Carolina Gaul, Elena Hinz, Christine Knoll, Begüm Kül, Anja Wels, Julia C Berkmann, Cathrin de Pasquale, Gloria Kremser, Daniel Fürstenau, Akira-Sebastian Poncette, Felix Balzer, Lina Mosch. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 13.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Cognitive Transaction: Toward a Human Factors Research Agenda for AI in Anesthesia and Perioperative Care

Sheena Warner[1], MS, APN, CRNA; Christopher H Stucky[2], RN, CNOR, CSSM, CNAMB, NEA-BC, PhD; Tamara Haegerich[3], PhD; Young J Yauger[3], CRNA, PhD

[1]Nell Hodgson Woodruff School of Nursing, Emory University, 1520 Clifton Road NE, Atlanta, GA, United States

[2]Brian D Allgood Army Community Hospital, Pyeongtaek, Republic of Korea

[3]American Association of Nurse Anesthesiology Foundation, Rosemont, IL, United States

Corresponding Author:
Sheena Warner, MS, APN, CRNA
Nell Hodgson Woodruff School of Nursing, Emory University, 1520 Clifton Road NE, Atlanta, GA, United States

Abstract

AI is now embedded in the infrastructure of perioperative care. Risk stratification algorithms, hemodynamic prediction tools, and clinical decision support systems are active in operating rooms at major health systems, and their adoption is accelerating. However, the field has studied model performance and organizational implementation while largely bypassing the moment between them: the real-time encounter in which an anesthesia provider must decide, under active case conditions, what to do with an AI-generated output. We term this the cognitive transaction and argue that it is the fundamental unit of perioperative AI implementation. The perioperative environment presents a specific constellation of conditions that existing human-AI interaction research was not designed to address. Continuous real-time decision demands, extreme time compression, high cognitive load, and consequences that unfold in seconds distinguish the operating room from the clinical contexts where most provider-AI interaction research has been conducted. What we know about AI adoption in radiology, oncology, or ambulatory care does not readily translate to this setting. The cognitive moment in anesthesia has its own structure, its own failure modes, and its own research requirements. This paper examines what those requirements are. We analyze how the operating room functions as a pre-existing human-machine cognitive system into which AI is now being inserted, and why the conditions of that system generate predictable vulnerabilities: miscalibrated trust, automation bias, and cognitive friction produced by interfaces optimized for technical accuracy rather than clinical usability. We argue that these failure modes are not incidental but structural and that they will persist regardless of model performance until the provider-AI interaction is itself treated as a research object. We identify 4 priority research domains. The first concerns the structure of provider-AI disagreement and the methods needed to distinguish automation bias from legitimate clinical insight. The second concerns the longitudinal dynamics of trust calibration across repeated clinical encounters rather than single-session experimental designs. The third concerns interface design for high-acuity workflows, specifically what constitutes a usable AI output for a provider managing a patient in real time. The fourth concerns the need for ecologically valid study designs capable of capturing provider reasoning under actual perioperative conditions rather than retrospective or survey-based proxies. The anesthesia and perioperative research community is positioned to lead this work. The clinical specificity, domain knowledge, and professional stake required to design meaningful studies are all present within the field. Evaluating the cognitive transaction under perioperative conditions, not the computational model in isolation, is both a methodological imperative and a patient safety priority.

(JMIR Hum Factors 2026;13:e102683) doi:10.2196/102683

KEYWORDS

cognitive transaction; perioperative artificial intelligence; human–artificial intelligence interaction; human-AI interaction; human factors; clinical decision support; anesthesia informatics

Introduction

This paper articulates a human factors research agenda addressing 3 levels of clinical AI integration. The primary target domain is the perioperative surgical environment, a uniquely complex setting characterized by continuous real-time decision demands, extreme time compression, and high cognitive load [1]. Furthermore, unlike insular environments such as an aviation cockpit, the operating room functions as a highly connected socio-spatial network [2]. Introducing AI into this intricate spatial and social architecture requires a dedicated human factors approach to ensure that new technologies align with actual clinical workflows. The insights derived from this environment function less as a direct template for clinical AI generally than as a stress test: if human-AI collaboration can be made to work under the extreme time compression and cognitive load of an active operating room case, the resulting

XSL•FO
RenderX

principles are more likely to hold with even greater confidence in less demanding clinical AI contexts. Finally, this work offers a theoretical contribution by positioning the "cognitive transaction" (the real-time encounter between algorithmic output and clinical judgment) as the fundamental unit of analysis for safe and effective AI deployment.

The Gap Nobody Is Naming

Perioperative health care is generating more data, more rapidly, than any individual clinician can process. The tools being built to close that gap are arriving faster than the science needed to deploy them safely. AI has entered the perioperative environment not as a future possibility but as present infrastructure. Risk stratification algorithms, early warning systems, and clinical decision support modules are increasingly embedded in the electronic health record platforms that anesthesia providers navigate daily.

The research literature evaluating these tools has expanded accordingly, with a growing body of studies reporting model performance and prospective validation results. A parallel implementation science literature has grown alongside it, examining organizational adoption, clinician behavior change, and system-level barriers to uptake. What neither body of literature has systematically studied is the moment between the two.

An anesthesia provider, mid-case, encounters an AI-generated risk output, such as a predicted probability of hemodynamic instability or a hemorrhage risk score embedded in the obstetric record. The number appears, and the provider must decide in seconds, under high cognitive load and with competing attentional demands, what to do with it. Do they act on it, override it, or adjust their existing assessment? The model has done its work; now the human must do theirs.

That cognitive transaction, the encounter between algorithmic output and clinical judgment in real time, is the fundamental unit of perioperative AI implementation. Remarkably, it is also the unit that the field has studied most inconsistently.

This omission does not stem from a lack of interest. Researchers across clinical AI, human factors engineering, and implementation science have recognized that provider-AI interaction shapes outcomes in ways that model performance alone cannot predict. However, the perioperative environment presents a specific constellation of conditions that makes it categorically different from the clinical contexts where most human-AI interaction research has been conducted. These unique conditions include continuous real-time decision demands, extreme physiologic complexity, and consequences that unfold in seconds rather than hours. The cognitive moment in anesthesia has its own structure, its own failure modes, and its own research requirements. Until those requirements are met, the gap between model performance and clinical value will persist regardless of how good the models are.

Why Perioperative AI Is a Distinctive Case

The perioperative environment requires its own analytical framework, not an adaptation of existing clinical AI research conducted elsewhere. Anesthesia providers do not encounter technology as an occasional resource; rather, they operate within it continuously. A single case requires the simultaneous management of a ventilator, infusion pumps, a physiologic monitoring array, and an anesthesia information management system. Each of these components generates real-time data streams, requires ongoing interpretation, and demands an immediate response when conditions deviate from clinical expectations. This is not a new condition introduced by AI. It is the baseline cognitive architecture of the role.

Cook and Woods [3], writing on adaptation to new technology in the operating room, described this environment as one in which human and machine contributions are already interwoven in ways that cannot be disentangled without fundamentally altering the nature of the work. This baseline matters for AI implementation because it means the operating room is not a neutral surface onto which a decision support tool is introduced. It is an existing human-machine cognitive system, already structured by relationships of trust, workload distribution, and situational awareness, into which an AI layer is being inserted [3,4].

The question is not whether providers can tolerate an additional data source. They tolerate many. The question is what happens to the existing cognitive architecture when one of those data sources is probabilistic, opaque, and carries an implicit authority that physiologic monitors do not.

To an anesthesia provider, a pulse oximeter reading of 88% is self-evident. An AI-generated risk probability of 0.73 is not. A pulse oximeter reading speaks directly to a physiologic state that the provider can verify and act upon. An AI risk probability speaks to a statistical relationship between patient characteristics and an outcome, derived from a training cohort the provider did not choose, using features the provider may not be able to inspect, with a CI that is rarely displayed. Every anesthesia provider knows how to respond to one. Almost none have been trained to critically evaluate the other. Treating these 2 outputs as equivalent in design, training, or implementation is a category error with clinical consequences.

The Human-AI Joint Cognitive Systems (HAIJCS) framework, proposed by Xu and Gao [5] and applied to the obstetric anesthesia context by Warner et al [6], offers a conceptual foundation for understanding this distinction. Under HAIJCS, the human provider and the AI system are understood as collaborative cognitive agents, each contributing to shared situational awareness, task execution, and decision-making under the authority of the human. The relevant question is not only whether the model predicts correctly but also whether the human-AI dyad operating together under actual clinical conditions reaches better decisions than either would reach alone and whether the conditions of the operating room allow that dyad to function as designed.

XSL•FO
RenderX

The Joint Cognitive System (JCS) and its HAIJCS extension model the operating room at the level of the system: a standing architecture of shared situational awareness, trust, and control between human and machine agents [4-6]. The cognitive transaction operates at a different grain. It is not a second name for that system-level relationship; it is the discrete event the relationship depends on, and none of the frameworks it draws on treat that event as an object of study in its own right. Trust calibration, as JCS frames it, describes a general disposition the human holds toward the machine over time. Situation awareness, as modeled by Schulz et al [4] in anesthesia, describes an ongoing process of perceiving, comprehending, and projecting system states. The cognitive transaction is neither the disposition nor the process. It is the specific encounter, this output, this patient, this second, in which trust gets tested and situational awareness gets updated. Distributed cognition makes a related claim at the structural level: cognitive work is located across people, tools, and environment rather than inside any single agent [7]. All 3 describe standing conditions of the joint system. None isolates the moment of integration as a unit that can be observed and measured directly. The cognitive transaction is that unit.

Research from aviation and military operations has established that high-workload, time-pressured environments generate predictable patterns in human-automation interaction, including automation bias, complacency, and degraded monitoring of system states during periods of high task demand [8,9]. Perioperative AI introduces an epistemic challenge that those findings were not built to address. The provider must act on a prediction they cannot verify in real time, about a future state that may never materialize, for a patient who may bear little resemblance to the population from which the model learned. Instruments can be independently cross-checked against observable measurements; risk scores cannot be directly verified in the same way. Translating human factors insights from adjacent domains is a necessary starting point, not a sufficient end point. Perioperative AI implementation requires a research agenda built for the operating room, not borrowed from the cockpit.

What the Current Literature Misses

The clinical AI literature has made substantial progress on model development, producing tools that perform well on retrospective datasets. Reviews of perioperative AI consistently demonstrate improvements in predictive accuracy for outcomes including postoperative complications, intraoperative hypotension, and adverse events across multiple patient populations [10]. This is meaningful work, but model development solves only a fraction of the problem.

A striking pattern has emerged from the broader clinical AI literature: AI systems working independently have begun to outperform clinicians using those same systems as decision support tools [11]. Explanations for this range from automation neglect to a fundamental mismatch between how models communicate uncertainty and how clinicians process it in practice. While the authors acknowledge that this evidence is early and drawn largely from controlled settings, it reveals a critical reality. Giving clinicians access to an AI tool does not automatically produce better decisions, and some have argued that the assistive model itself may be fundamentally flawed as a design paradigm [12]. Regardless, the underlying observation stands. The quality of the human-AI interaction, rather than the model's isolated performance, determines whether the combination is synergistic, additive, or subtractive. This interaction has rarely been studied with the rigor applied to the models themselves.

The implementation science literature has done important work naming this problem. Studies on sepsis decision support tools have found alert fatigue, high override rates, and clinician abandonment of models that appeared predictive on paper [13,14]. These deployment failures are not primarily technical; they are cognitive and interpretive. Clinicians cannot inspect the key features driving a prediction or assess if the training cohort matches the patient in front of them. When a model's output conflicts with clinical assessment, clinicians may lack a clear framework for determining which source of information to trust. The result is either uncritical acceptance, which introduces automation bias, or reflexive dismissal, which undermines the tool's purpose [15].

Reviews have summarized technology and human factors that influence clinician-AI interaction, such as data quality, performance, transparency, expertise, cognitive biases, and trust [16]. Perioperative AI researchers have identified similar concerns [10,17]; however, the specific perioperative environment makes failures involving technology and human factors particularly acute. Three features of the perioperative environment explain why that methodological work is both especially difficult and necessary.

First, deliberation time is not merely compressed but functionally eliminated during critical perioperative events. Other clinical AI contexts deploy decision support in real time, including sepsis alerts, opioid risk prediction, and chronic disease management, but providers in those settings retain a meaningful window for reflection, consultation, and staged response [13,14,18]. A perioperative risk output during an evolving hemodynamic event does not allow for reflection, consultation, or staged response. Cognitive psychology research illustrates that time pressure can reduce the quality of activities, modify risk-taking, change underlying cognitive processing, and increase the relative relevance of different information sources [19]; hence, understanding how time pressure influences the clinician-AI interaction in the context of anesthesia and operative events is critically important. Second, human-AI interaction research has rarely accounted for the profound cognitive load of active case management, where fatigue, divided attention, and the dynamic spatial environment of the surgical team dictate whether an alert registers as a signal or mere noise [4,17]. Third, anesthesia providers lack established norms for interacting with probabilistic AI algorithms. Unlike radiologists, who have worked alongside AI-assisted image review tools for years, anesthesia providers are encountering these tools without a shared professional vocabulary for evaluating, trusting, or overriding them [20,21].

XSL•FO
RenderX

This third feature points to a structural explanation for the literature gap. The clinician-researchers best positioned to study the cognitive moment in perioperative AI, those who understand both the perioperative environment and the methodological tools required to study human-AI interaction, exist at an intersection that the field has not yet organized around. Computational researchers lack deep access to perioperative decision-making processes. Implementation scientists lack the anesthesia-specific domain knowledge to recognize what makes perioperative AI distinctive. Meanwhile, the anesthesia providers who navigate these cognitive demands daily have not traditionally received training in human factors or informatics. The gap is not simply a gap in knowledge. It is a gap in who is asking the question.

Toward a Research Agenda for the Cognitive Moment

Rather than proposing a comprehensive framework, this agenda outlines priority research questions grounded in the distinctive features of the perioperative environment. These domains are drawn primarily from risk-stratification and predictive-alert use cases, which currently represent the AI applications most deeply embedded in perioperative workflows. Real-time hemodynamic prediction tools compress the transaction further, leaving less room for the reflective divergence discussed below. Generative systems, which produce free-text recommendations rather than a single score, introduce additional interpretive demands that the disagreement and interface-design frameworks discussed here do not yet address. Whether the cognitive transaction holds the same structure across these AI types, or requires type-specific variants of each research question, is an open question this agenda raises rather than resolves. The most pressing concern is the structure of provider-AI disagreement: when an anesthesia provider and an AI system arrive at different risk assessments for the same patient, what explains the divergence? Current methodologies, largely adapted from radiology and emergency medicine, focus primarily on agreement metrics rather than investigating these underlying sources of conflict [22]. In the operating room, provider divergence from an AI output may reflect automation bias, but it may equally signal critical clinical insight inaccessible to the model, a patient's affect, a subtle physiological pattern recognized over thousands of cases, or unrecorded institutional variables. A perioperative research agenda must disentangle these sources of divergence rather than collapsing them into a binary agreement metric. Capturing this nuance requires mixed methods designs that pair quantitative performance data with structured qualitative inquiry into provider reasoning, an approach largely absent from the current literature. The critical decision method developed by Klein et al [23] offers an established route to this distinction: a structured retrospective interview technique that reconstructs the cues, prior cases, and contextual signals a provider drew on at the moment of divergence, rather than treating the divergence itself as the end point of analysis. It remains one of the most commonly applied cognitive task analysis techniques in clinical decision-making research, spanning surgical and critical care settings [24].

Trust in an AI system is fluid and not a fixed attribute. However, the current clinical AI literature predominantly evaluates trust through cross-sectional surveys or single-encounter experimental designs, yielding a static snapshot. This fails to capture the most vital question for perioperative implementation: How does trust in a specific AI system develop, erode, and recalibrate over repeated clinical encounters? Anesthesia providers who override a model's output and later find it correct will reasonably become skeptical of the model's future reliability. On the other hand, providers who defer to the model and observe a good outcome may default to uncritical reliance. Longitudinal designs that pair momentary, in-the-moment reliance judgments with case-level outcome tracking, sampled repeatedly across a provider's caseload, would let calibration be measured as a trajectory rather than a single point estimate. The relevant construct is not trust in the abstract but calibrated reliance: whether a provider's confidence in a given output corresponds to its actual reliability on a case-by-case basis.

The perioperative AI literature has given little systematic attention to how model outputs should be communicated to providers operating under the cognitive conditions of active case management. Current tools tend to present a risk score or classification optimized for technical legibility rather than clinical usability, prioritizing mathematical accuracy over the ability of a multitasking clinician to quickly and appropriately integrate the information. Human factors research has long established that the design of information presentation shapes clinical behavior just as much as the information itself. Outputs developed without regard for cognitive load, task context, and provider mental models routinely fail at the point of care, even when the underlying algorithm performs well [4,13]. What constitutes a usable, low-friction AI interface for an anesthesia provider mid-case, in terms of format, timing, level of specificity, and fit with existing workflow, remains unanswered. One example of a low-friction output might surface as a color-coded flag integrated directly into the existing vital sign display rather than a separate probability score requiring independent interpretation. Another might be a brief natural language prompt, such as "rising hemorrhage risk, consider early type and screen," delivered at a natural pause in the case rather than immediately when the model updates. Adjudicating between formats requires usability testing under simulated time pressure, not static comprehension surveys. Solving this for the extreme constraints of the operating room will establish baseline principles for AI usability across all clinical environments. Until this is addressed, interface design will continue to be driven by technical conventions rather than clinical evidence.

Ultimately, the field requires research methodologies capable of capturing not only what anesthesia providers think about AI in the abstract but also how they reason about and integrate these outputs within ecologically valid clinical contexts. This should encompass both perceptual and deliberative processes. Expert anesthesia providers often detect physiological deterioration through pattern recognition that precedes conscious reasoning. Recent work pairing physical simulation with eye tracking has begun to capture this in critical care, finding that clinicians' visual attention did not reliably increase even when an AI recommendation was unsafe [25]. Perioperative research

XSL•FO
RenderX

needs the same instrumentation applied to the compressed conditions of an active case. This necessitates moving beyond retrospective analyses and survey-based designs. Instead, research should use realistic clinical simulations incorporating model outputs and evolving patient data to examine the continuous cognitive process of integration, rather than focusing solely on the final binary decision. We must investigate how providers respond to unexpected algorithmic predictors, how the framing of an output shapes clinical action, and precisely when and why human judgment diverges from a model's recommendation. These are not peripheral questions of user experience. They are the empirical foundation on which safe, deployable perioperative AI must be built.

Why This Matters Now

Perioperative AI is not a future abstraction; it is a present reality. Tools for perioperative risk prediction, hemodynamic management, and postoperative complication forecasting are already embedded in clinical workflows at major health systems, and their adoption is accelerating [10]. The window for shaping how these systems are designed, evaluated, and deployed is open now, and it will not remain open indefinitely. Once these tools become embedded in institutional infrastructure, clinical workflows reorganize around them, and the cost of redesign increases substantially. The decisions being made today about what perioperative AI looks like, what it communicates, and how it interacts with provider judgment will permanently shape the cognitive environment of anesthesia practice.

The research agenda proposed here is not a precondition for building these systems; it is a precondition for building them safely and effectively. Models can be designed and deployed without understanding how providers interact with their outputs. In fact, they already have been. The consequence, documented repeatedly in adjacent clinical AI contexts, is a generation of tools that perform well in evaluation but fail in practice. This occurs not because the underlying science was wrong, but because the implementation ignored the human at the other end of the interface [11,13].

Anesthesia providers are not passive recipients of this technology. They are, by the nature of their role, the final cognitive layer between an algorithmic output and a clinical action. That position is not incidental. It is the precise point at which the entire value proposition of perioperative AI either materializes or collapses. A research program that takes this position seriously, that asks what the provider needs to know, what the interface needs to communicate, and what the conditions are under which human and algorithmic judgment combine to produce outcomes better than either alone, is not a niche methodological interest. It is the central scientific problem of perioperative AI implementation.

The anesthesia and perioperative research community is uniquely positioned to lead this work. The clinical specificity required to design meaningful studies, the domain knowledge required to interpret findings, and the professional stake in getting implementation right are all present within the field. Engaging the anesthesia and perioperative community in the development of AI solutions is consistent with calls for the engagement of other specialties, such as primary care, in research and implementation efforts that affect their patient populations. Such engagement helps ensure that AI applications are developed in ways that address the needs of frontline practitioners and ultimately improve patient outcomes [26]. That leadership, however, must be exercised in partnership. The research priorities identified here sit at the intersection of cognitive psychology, health informatics, implementation science, and clinical practice. No single discipline has the tools, the access, and the domain knowledge to address them alone. Anesthesia providers bring irreplaceable expertise in the perioperative environment, cognitive psychologists bring the methodological frameworks for studying human-AI interaction, informaticists bring the technical infrastructure for capturing and analyzing clinical data, and implementation scientists bring the frameworks for translating findings into practice. In practical terms, each stakeholder has a concrete starting point: anesthesia providers can partner with cognitive psychologists to design the simulation studies this agenda calls for; health system informaticists can instrument existing AI tools to capture override rates and response times rather than waiting for purpose-built research infrastructure; and clinical AI developers can treat these research questions as design requirements to address before deployment, rather than as validation exercises conducted afterward. Underlying all of these is a shared recognition that the cognitive transaction at the intersection of algorithmic output and clinical judgment is itself a discrete research object, distinct from the model, the organization, and the provider in isolation. Naming it as such is the first necessary step toward studying it rigorously and designing perioperative AI systems that are not only accurate but also genuinely useful in the operating room.

Acknowledgments

Claude (Anthropic) was used to assist with refining terminology and formatting, limited to content related to Joint Cognitive Systems thinking and aligned with principles of responsible human-AI interaction. Grammarly was used for grammar checking, clarity improvements, and proofreading, consistent with its intended purpose.

Funding

This research was supported in part by the AANA Foundation (grant 2025-G-8) and partially funded by the National Institute of Nursing Research, National Institutes of Health (NIH; grant T32NR020778; principal investigators: D. Bruner and V. Hertzberg), for trainee Sheena Warner. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

XSL•FO
RenderX

Data Availability

No data were generated or analyzed in this study.

Authors' Contributions

Conceptualization: SW (lead), YJY (supporting)
Funding Acquisition: SW (lead), TH (equal), YJY (equal)
Methodology: SW
Project administration: SW
Supervision: YJY
Validation: YJY
Visualization: SW (lead), CHS (supporting), TH (supporting), YJY (supporting)
Writing – original draft: SW (lead), CHS (supporting), TH (supporting), YJY (supporting)
Writing – review & editing: SW (lead), CHS (supporting), TH (supporting), YJY (supporting)

Conflicts of Interest

None declared.

References

1. Stucky CH, De Jong MJ, Kabo FW. Military surgical team communication: implications for safety. Mil Med 2020 Mar 2;185(3-4):e448-e456. [doi: 10.1093/milmed/usz330] [Medline: 31665390]
2. Kabo FW, Stucky CH, De Jong MJ. Associations of surgical team communication with the layout of physical space: a network analysis of the operating room in a military medical center. HERD 2023 Jul;16(3):134-145. [doi: 10.1177/19375867231159130] [Medline: 36866407]
3. Cook RI, Woods DD. Adapting to new technology in the operating room. Hum Factors 1996 Dec;38(4):593-613. [doi: 10.1518/001872096778827224] [Medline: 8976623]
4. Schulz CM, Endsley MR, Kochs EF, Gelb AW, Wagner KJ. Situation awareness in anesthesia: concept and research. Anesthesiology 2013 Mar;118(3):729-742. [doi: 10.1097/ALN.0b013e318280a40f] [Medline: 23291626]
5. Xu W, Gao Z. Applying HCAI in developing effective human-AI teaming: a perspective from human-AI joint cognitive systems. Interactions 2024 Jan;31(1):32-37. [doi: 10.1145/3635116]
6. Warner S, Xiao R, Wiltse Nicely KL, Chandler R. From joint cognitive systems to human-AI joint cognitive systems: a theory critique and application to obstetric anesthesia risk assessment. Nurs Sci Q 2026 Apr;39(2):117-125. [doi: 10.1177/08943184251414311] [Medline: 41587871]
7. Hollan J, Hutchins E, Kirsh D. Distributed cognition: toward a new foundation for human computer interaction research. ACM Trans Comput Hum Interact 2000 Jun;7(2):174-196. [doi: 10.1145/353485.353487]
8. Prinzel L, Krois P, Ellis K, et al. The adaptable and resilient safety system: the human factor in future in-time aviation safety management systems. In: AIAA SciTech Forum and Exposition, 2024: American Institute of Aeronautics and Astronautics; 2024. [doi: 10.2514/6.2024-1603]
9. Worm A. Joint Tactical Cognitive Systems: modeling, analysis, and performance assessment. Proc Hum Factors Ergon Soc Annu Meet 1998;42(3):315-319. [doi: 10.1177/154193129804200328]
10. Han L, Char DS, Aghaeepour N, Stanford Anesthesia AI Working Group. Artificial intelligence in perioperative care: opportunities and challenges. Anesthesiology 2024 Aug 1;141(2):379-387. [doi: 10.1097/ALN.0000000000005013] [Medline: 38980160]
11. Topol E, Rajpurkar P. When doctors with A.I. are outperformed by A.I. alone: interpreting some surprising results. Ground Truths. 2025. URL: https://erictopol.substack.com/p/when-doctors-with-ai-are-outperformed/ [accessed 2025-02-12]
12. Rajpurkar P, Topol EJ. Beyond assistance: the case for role separation in AI-human radiology workflows. Radiology 2025 Jul;316(1):e250477. [doi: 10.1148/radiol.250477] [Medline: 40728396]
13. Zhang S, Yu J, Xu X, et al. Rethinking human-AI collaboration in complex medical decision making: a case study in sepsis diagnosis. Proc SIGCHI Conf Hum Factor Comput Syst 2024 May;2024:445. [doi: 10.1145/3613904.3642343] [Medline: 38835626]
14. Wong A, Otles E, Donnelly JP, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. JAMA Intern Med 2021 Aug 1;181(8):1065-1070. [doi: 10.1001/jamainternmed.2021.2626] [Medline: 34152373]
15. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc 2012;19(1):121-127. [doi: 10.1136/amiajnl-2011-000089] [Medline: 21685142]
16. Knop M, Weber S, Mueller M, Niehaves B. Human factors and technological characteristics influencing the interaction of medical professionals with artificial intelligence-enabled clinical decision support systems: literature review. JMIR Hum Factors 2022 Mar 24;9(1):e28639. [doi: 10.2196/28639] [Medline: 35323118]

XSL•FO
RenderX

17. Duran HT, Kingeter M, Reale C, Weinger MB, Salwei ME. Decision-making in anesthesiology: will artificial intelligence make intraoperative care safer? Curr Opin Anaesthesiol 2023 Dec 1;36(6):691-697. [doi: 10.1097/ACO.0000000000001318] [Medline: 37865848]
18. Nguyen K, Wilson DL, Diiulio J, et al. Design and development of a machine-learning-driven opioid overdose risk prediction tool integrated in electronic health records in primary care settings. Bioelectron Med 2024 Oct 18;10(1):24. [doi: 10.1186/s42234-024-00156-3] [Medline: 39420438]
19. Maule AJ, Edland AC. The effects of time pressure on human judgement and decision making. In: Crozier R, Ranyard R, Svenson O, editors. Decision Making: Cognitive Models and Explanations: Routledge; 1997:203-218. [doi: 10.4324/9780203444399]
20. Castellino RA. Computer aided detection (CAD): an overview. Cancer Imaging 2005 Aug 23;5(1):17-19. [doi: 10.1102/1470-7330.2005.0018] [Medline: 16154813]
21. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJ. Artificial intelligence in radiology. Nat Rev Cancer 2018 Aug;18(8):500-510. [doi: 10.1038/s41568-018-0016-5] [Medline: 29777175]
22. Tikhomirov L, Semmler C, McCradden M, Searston R, Ghassemi M, Oakden-Rayner L. Medical artificial intelligence for clinicians: the lost cognitive perspective. Lancet Digit Health 2024 Aug;6(8):e589-e594. [doi: 10.1016/S2589-7500(24)00095-5] [Medline: 39059890]
23. Klein GA, Calderwood R, MacGregor D. Critical decision method for eliciting knowledge. IEEE Trans Syst Man Cybern 1989;19(3):462-472. [doi: 10.1109/21.31053]
24. Swaby L, Shu P, Hind D, Sutherland K. The use of cognitive task analysis in clinical and health services research - a systematic review. Pilot Feasibility Stud 2022 Mar 8;8(1):57. [doi: 10.1186/s40814-022-01002-6] [Medline: 35260195]
25. Festor P, Nagendran M, Gordon AC, Faisal AA, Komorowski M. Safety of human-AI cooperative decision-making within intensive care: a physical simulation study. PLOS Digit Health 2025 Feb;4(2):e0000726. [doi: 10.1371/journal.pdig.0000726] [Medline: 39992918]
26. Sarkar U, Bates DW. Using artificial intelligence to improve primary care for patients and clinicians. JAMA Intern Med 2024 Apr 1;184(4):343-344. [doi: 10.1001/jamainternmed.2023.7965] [Medline: 38345801]

Abbreviations

HAIJCS: Human-AI Joint Cognitive Systems
JCS: Joint Cognitive System

Edited by D Kaufman; submitted 27.May.2026; peer-reviewed by ID Moral, L Gupta; revised version received 10.Aug.2026; accepted 19.Aug.2026; published 10.Sep.2026.

Please cite as:
Warner S, Stucky CH, Haegerich T, Yauger YJ
The Cognitive Transaction: Toward a Human Factors Research Agenda for AI in Anesthesia and Perioperative Care
JMIR Hum Factors 2026;13:e102683
URL: https://humanfactors.jmir.org/2026/1/e102683
doi:10.2196/102683

© Sheena Warner, Christopher H Stucky, Tamara Haegerich, Young J Yauger. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 10.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

How to Evaluate the Accuracy of Symptom Checkers and Diagnostic Decision Support Systems: Symptom Checker Accuracy Reporting Framework (SCARF)

Marvin Kopka, PhD, Dr rer medic, MPH, MSc, BSc; Markus A Feufel, Dipl-Ing (FH), MSc, PhD

Division of Ergonomics, Department of Psychology & Ergonomics (IPA), Technische Universität Berlin, Straße des 17. Juni 135, Berlin, Germany

Corresponding Author:
Marvin Kopka, PhD, Dr rer medic, MPH, MSc, BSc
Division of Ergonomics, Department of Psychology & Ergonomics (IPA), Technische Universität Berlin, Straße des 17. Juni 135, Berlin, Germany

Abstract

Symptom checkers are apps and websites that assist medical laypeople in diagnosing their symptoms and determining which course of action to take. When evaluating these tools, previous studies primarily used an approach introduced a decade ago that lacked any type of quality control. Numerous studies have criticized this approach, and several empirical studies have sought to improve specific aspects of evaluations. However, even after a decade, a high-quality methodological framework for standardizing the evaluation of symptom checkers is still lacking. This paper synthesizes empirical studies to outline the Symptom Checker Accuracy Reporting Framework (SCARF) and a corresponding checklist for standardizing evaluations based on representative case selection, an externally and internally valid evaluation design, and metrics that increase cross-study comparability. This approach is supported by several open access resources to facilitate implementation. Ultimately, it should enhance the quality and comparability of future evaluations of online and artificial intelligence (AI)–based symptom checkers, diagnostic decision support systems, and large language models to enable meta-analyses and help stakeholders make more informed decisions.

(JMIR Hum Factors 2026;13:e76168) doi:10.2196/76168

KEYWORDS

symptom checker; symptom assessment applications; evaluation; case vignettes; preclinical; decision support; large language model; data analysis; health technology assessment; impact evaluation; artificial intelligence; AI

Introduction

Symptom checkers (also called “symptom assessment applications,” “online symptom checkers,” or “self-assessment applications”) are websites or mobile apps in which medical laypeople can enter their symptoms. The apps then provide potential diagnoses and “self-triage” advice. Self-triage advice refers to recommendations given in a precare setting to assist users in determining if, how urgently, and in which institution they should seek care. The first study to systematically analyze the accuracy of these apps was conducted in 2015, and their accuracy has been debated ever since [1]. This seminal study evaluated symptom checkers using 45 medical case vignettes (15 emergency care cases in which users would call the national emergency line or go directly to the emergency department, 15 nonemergency cases in which users would seek primary care, and 15 self-care cases in which users would treat symptoms themselves or wait to see if symptoms improve before seeking care) that were taken from various medical resources, including medical education textbooks. The gold standard solution—that is, the most appropriate action for each case—was determined by 2 physicians who independently rated each case and then discussed disagreements. An unrelated researcher entered all cases into the various symptom checkers, and the authors calculated the proportion of cases correctly solved as the main outcome. This procedure has been used in most subsequent studies, sometimes with slight modifications such as adding more vignettes and triage levels, using lay-friendly phrasing of the vignettes, or including large language models as symptom checkers [2-5]. However, most of these studies acknowledged limitations with this approach and called for improved methods. Systematic reviews that attempted to determine the accuracy of symptom checkers across multiple studies quickly reached the consensus that these methods were often of low quality and that cross-study comparability was limited [6-9]. In recent years, some studies have explicitly formalized this criticism, whereas others have proposed solutions to address it, including several of our own [6,10-14].

In this paper, we do not want to add to this criticism; instead, we present the Symptom Checker Accuracy Reporting Framework (SCARF) and a checklist that can (1) be used to conduct high-quality symptom checker evaluation studies and (2) standardize the evaluation procedure to increase cross-study comparability of symptom checkers and large language models. Because self-triage advice is arguably the most useful

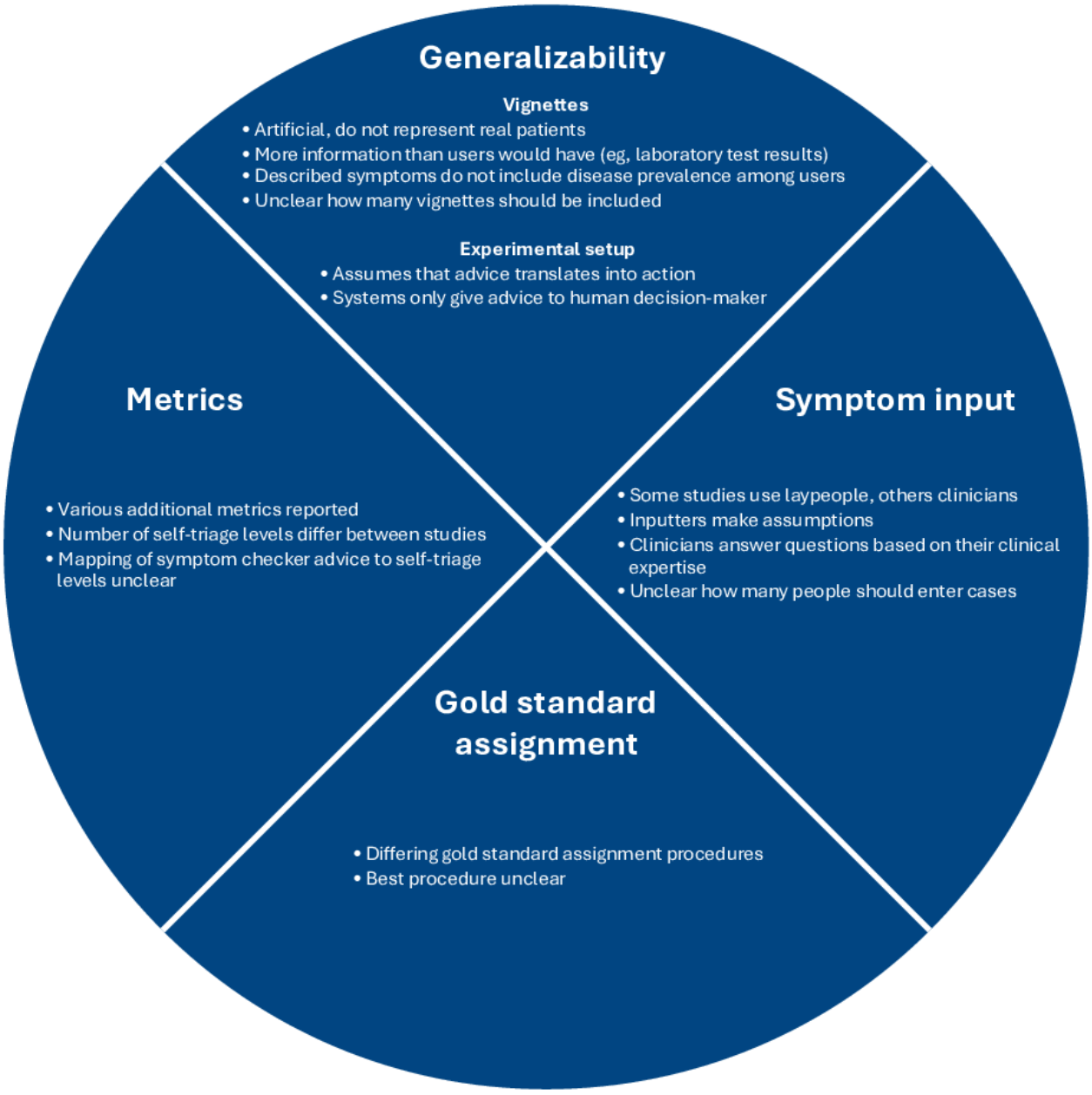
XSL•FO
RenderX

information for medical laypeople, this framework focuses on self-triage accuracy as the main outcome [15].

Limitations and Challenges of Previous Methodologies

Most studies evaluating the triage accuracy of symptom checkers have criticized the existing evaluation approach for being artificial. In particular, the vignettes describe idealized, unambiguous cases, and some include scenarios for which symptom checkers would rarely be consulted (eg, recurrent aphthous stomatitis, which may be unexplainable upon first appearance but is easily recognized once experienced [16]). If the aim is to determine a triage accuracy metric that can be generalized to real-world interactions and scenarios in which symptom checkers are actually used, the inclusion of such cases in evaluations seems questionable. Apart from vignettes, current evaluation approaches have several other shortcomings that we grouped into 4 categories: generalizability, symptom input, gold standard assignment, and metrics (Figure 1). We build on these points to develop our standardized methodological evaluation framework.

Figure 1. Four categories of criticism regarding symptom checker evaluation studies.



The first point concerns the generalizability of the evaluations. This includes both the vignettes and the experimental setup, which, according to ecological validity theory, should resemble real-world use cases and interactions to yield results that can be generalized [17]. Traditional vignettes have been criticized for a lack of representativeness for several reasons. First, they are often derived from medical education textbooks and are therefore artificial, not representing the unspecific concerns for which patients would use a symptom checker [10,14]. Second, these cases are mostly written post hoc by clinicians who have

XSL•FO
RenderX

access to more specialized information (eg, diagnoses, laboratory test results, and clinical examinations) than a patient consulting a symptom checker [10,13,18]. In other words, thus far, the information in existing vignettes does not reflect the types of problems actual users of symptom checkers face, and it is not clear what that information should be. Third, the cases described in the vignettes do not reflect the natural base rates of emergency or nonemergency versus self-care cases among users [10,13,14]. Fourth, there is no consensus on the number of vignettes that should be included in a vignette set or how to ensure their quality [10]. The experimental setup focuses on symptom checker accuracy and has thus been criticized for implicitly assuming that symptom checker advice directly translates into user actions, even though symptom checkers merely provide advice that users may or may not follow [19-21]. This limitation confines research to assessing only the "technical accuracy" of a symptom checker, without addressing its likely real-world impact. To determine whether technical accuracy translates into improved decision-making by users, symptom checkers ultimately need to be evaluated in user studies.

The second point concerns the procedure for inputting symptoms. Typically, a single person—who may or may not have medical expertise—enters the symptoms. Because not all information that a symptom checker might ask for is included in the vignette, the inputter must make assumptions about the case when asked about it. Thus, clinicians tend to rely on their clinical judgment and expertise, whereas laypeople—the actual users of symptom checkers—use various strategies, ranging from guessing to ignoring the questions they are asked [10,11,22]. It is also unclear how many inputters should be involved in the evaluation to yield valid performance estimates [10]. These issues suggest that the final output is highly dependent on the inputter or inputters, which creates an information bias that limits the internal validity of evaluation studies. This information bias is further compounded by the fact that different symptom checkers allow different input modalities (eg, free text, multiple-choice questions, images, or even laboratory results). These variations introduce an inherent comparability limitation, as the same case may be assessed differently depending on the input options of the tool.

The third point relates to the gold standard assignment used to assign the solution to a case vignette. Different studies use varying procedures: some use a single physician, others use multiple physicians, recordings from clinical encounters (such as telephone triage), or sometimes the authors even determine the gold standard solution themselves [10,12]. This variation not only limits the comparability between studies but also raises questions about the accuracy of the assigned gold standard in some cases.

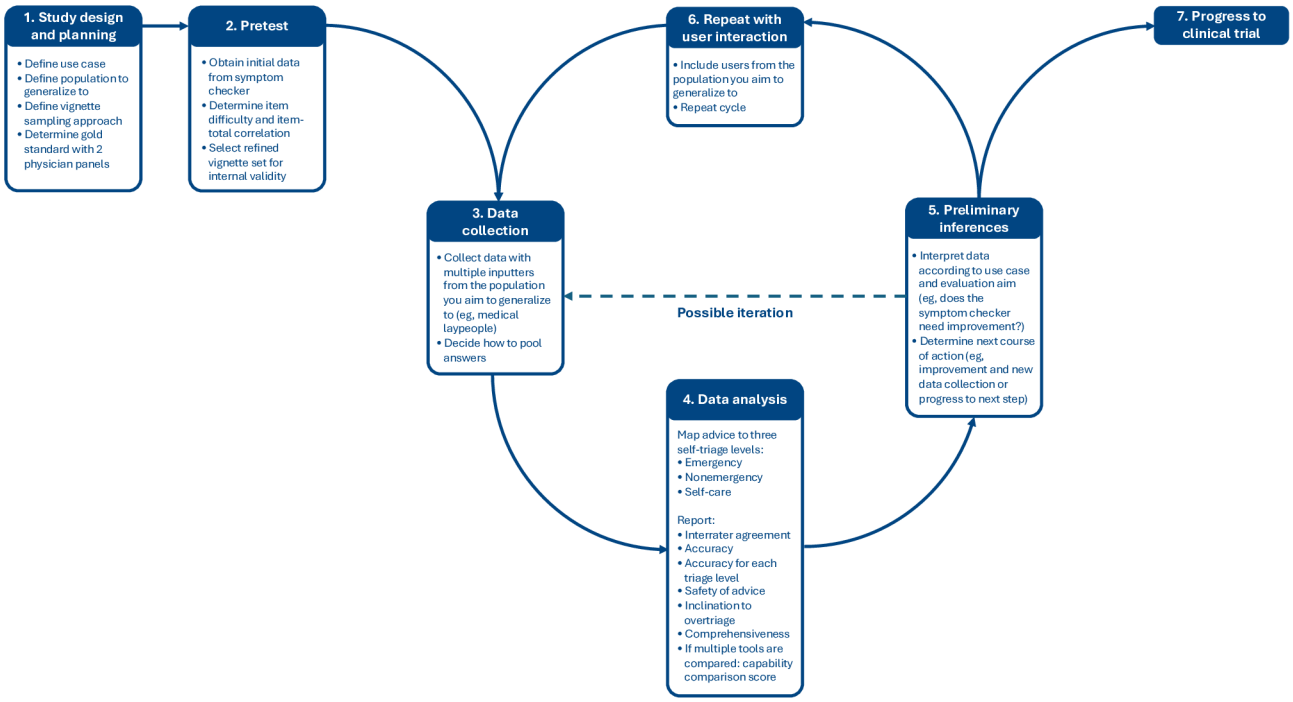
The final point concerns the metrics used to evaluate symptom checkers. Although most studies report triage accuracy as the proportion of cases solved correctly, some also include additional metrics, such as the tendency to overtriage or undertriage and the safety of the advice [14]. The exact self-triage classifications differ between studies as well: for example, Semigran et al [1] used a 3-tiered approach including "emergencies," "nonemergencies," and "self-care cases," whereas Hill et al [2] extended this classification to include "1-day urgent cases." Furthermore, because different symptom checkers use varying classifications as well, it is unclear how their advice should be mapped to the study's triage categories (eg, whether an urgent care clinic is considered emergency or nonemergency care). These issues ultimately limit cross-study comparability.

Framework

To address these points, we developed an evaluation framework by integrating available empirical studies on methodological improvements. This framework can be found in Figure 2. It can be used for preclinical evaluations to identify symptom checkers that are likely to perform well in clinical trials and real-world evaluations. Once identified, the symptom checker should nonetheless undergo testing in a 3-phase clinical trial similar to pharmaceutical trials [23]. Hence, our framework not only standardizes vignette-based symptom checker evaluations but also makes subsequent clinical trials more cost efficient by identifying tools likely to yield positive outcomes. It can be applied both to evaluations across a broad set of cases as well as to those focusing on specific patient groups (eg, patients receiving rheumatology care), by defining the intended use case and population accordingly.

XSL•FO
RenderX

Figure 2. Integrated preclinical evaluation framework (Symptom Checker Accuracy Reporting Framework; SCARF) for evaluating the self-triage accuracy of symptom checkers and artificial intelligence–based tools.



In the beginning (part 1), evaluators should clearly define the use case they intend to examine, such as “self-triage decisions” or “emergency care decisions.” Next, they should specify the target population to which they wish to generalize. For the self-triage use case, this might include symptom checker users deciding on their next course of action. Then, they should define a vignette sampling approach, which ensures that vignettes are representative of real patient cases and accurately reflect the disease or symptom and triage prevalence relevant to the use case. For example, the approach could sample real patient cases stratified according to the prevalence of symptom types entered into symptom checkers. A systematic sampling procedure to do that is available in the RepVig framework, and for the self-triage use case, a representative vignette set is provided in the framework’s validation study [13]. At this stage, researchers should also assign a gold standard solution to each case and define how the possible outputs of a symptom checker are mapped onto these categories. According to a study by El-Osta et al [12], this should involve 2 physician panels that independently rate the cases in focus groups and resolve any disagreement through discussion until consensus is reached.

Next, evaluators should obtain initial data from some symptom checkers to refine the vignette set according to test-theoretical criteria (part 2) to ensure that vignettes are not only externally but also internally valid. This process involves calculating the item-total correlation and excluding any cases with a negative or zero item-total correlation (to ensure that only cases accurately predicting overall performance are included). Additionally, item difficulty for each vignette should be determined, and cases with an item difficulty of zero may be excluded (to ensure that vignettes add meaningful information and are not impossible to solve). However, if these cases can be solved by physicians and are clinically plausible, even items with an item difficulty of zero may be retained in the vignette set to avoid inflating performance estimates. A procedure for this is outlined in one of our previous studies [24]. The size of the final vignette set should ultimately be determined using a power analysis. However, given that entering a large number of vignettes manually may be infeasible and that there is no empirical evidence on optimal set sizes, we pragmatically recommend a minimum of 45 vignettes. This number has proven feasible and has been applied across multiple studies [1,2,13,25], as it can be developed and entered by a single evaluator within a reasonable time frame, while still providing sufficient variation for a statistical analysis.

Using the refined vignette set, data from all symptom checkers can be collected (part 3). Multiple inputters (at least 2, possibly more) should enter every case into each symptom checker and select a “not sure” option in cases of missing information. To minimize inputter variability, inputters should follow a standardized protocol. For instance, Meczner et al [11] instructed inputters to enter only the symptoms explicitly stated in the vignette, allowing synonyms or broader categories but rejecting new information not included in the vignette. Their publication provides entry instructions that can be used in future studies to standardize input procedures. Once the inputters have obtained the data, their results should then be pooled. This can be achieved using several algorithms, but the best approach appears to be a majority vote, that is, the advice most frequently given to all evaluators [11]. For example, if 2 inputters receive the advice to seek emergency care while 1 inputter receives self-care advice, the recommendation should be coded as “emergency”.

In the next step, the data analysis (part 4), evaluators should map the received advice to a multitiered classification system. To increase comparability across studies and health care systems, we suggest using a 3-tiered classification system—“emergency,” “nonemergency,” and “self-care”—to provide a common reference structure. A potential

XSL•FO
RenderX

"1-day-urgent" category could be classified as "nonemergency." At the same time, we acknowledge that some systems use more granular triage categories; therefore, we suggest conducting sensitivity analyses (eg, treating "1-day-urgent" as "emergency" or as its own category, or analyzing the full set of available tiers) to assess the stability of the results. After mapping each recommendation, evaluators should first report the interrater reliability among all inputters to identify the influence of different inputters, followed by a set of metrics: overall accuracy, accuracy for each triage level, safety of advice, inclination to overtriage, and comprehensiveness [14,26]. These metrics were identified through systematic review of previously reported metrics and can increase comparability across different studies [14,26]. If multiple symptom checkers are evaluated simultaneously, we propose additionally reporting the Capability Comparison Score (developed in a previous study) to determine how symptom checkers perform relative to each other [14,26]. To assist researchers in reporting and visualizing these metrics, the R package *symptomcheckR* is available, where the formulas for calculating all metrics are described as well [26].

In the next step, the results should be interpreted according to the defined use case, and the next course of action should be determined (part 5): if developers aim to validate their tools, they may either decide to improve their tool and test it again (by going back to step 3) using the same setting or continue with the evaluation and test the best-performing tools with users in the loop making self-triage decisions (step 6). In this phase, users should be provided with the symptom checker, and the tool's impact, instead of its "technical accuracy," should be assessed in a new evaluation with sufficient statistical power [21]. This step is included because preclinical vignette studies can only benchmark technical accuracy and do not capture whether laypeople actually make better decisions when using a symptom checker. If results of user studies are also promising, the symptom checker can then be tested in a clinical trial with real patients and their symptoms to assess whether the symptom checker advice also translates to improved decisions by users in the real world (step 7).

Open Questions

Our approach leaves several open questions for future research. First, some of the vignettes (such as the vignettes by Semigran et al [1] and our own [13]) do not include additional information for questions that symptom checkers may ask. Although some vignette sets do include additional information, there is no universal way to collect additional information. Future research could develop a method to supplement this missing information—perhaps using a hybrid approach that combines interviews with patients from whom the case vignettes were derived and synthetic artificial intelligence (AI)–generated supplementary data based on these interviews. Second, it remains unclear whether "accuracy" or a "correct" solution should be the main outcome. Perhaps a binary classification of correct versus incorrect in a task like symptom assessment that is associated with high uncertainty may be less relevant than assessing the impact of the advice—specifically, whether it is safe and appropriate for the individual and whether it increases or decreases health care demands. Third, with the introduction of large language models as an alternative to traditional symptom checkers, output variability plays an even greater role. Future research should address how to manage the variability of generated outputs when provided with identical inputs. Fourth, current evaluations do not specifically include atypical presentations. It remains unclear whether case vignettes are only suitable for typical cases or if vignette sets for atypical cases could also be developed. Although the RepVig framework could be used for developing such a vignette set again, assigning a reliable gold standard solution to atypical cases will be challenging [13]. Finally, our approach is highly tailored to a self-triage use case. Although it standardizes most aspects of an evaluation, diagnostic use cases may require additional details (such as clinical plausibility of the vignettes or a procedure to determine whether a diagnosis matches the gold standard) and outcome metrics (such as cumulative diagnostic scores [27]) that are not covered by our approach.

Outlook

The SCARF (and the corresponding checklist in Multimedia Appendix 1 and Multimedia Appendix 2) presented in this paper addresses all previously raised points of criticism and aims to improve the quality and comparability of future symptom checker evaluations. However, we acknowledge that the presented approach is more resource intensive than the traditional approach introduced by Semigran et al [1] and may not be feasible for every evaluation. To aid researchers in integrating these methods into practice, several open resources are available for the presented use case: for example, representative vignettes are openly accessible and free to use [13], a refined vignette set that satisfies test-theoretical criteria is available as well [24], and all metrics can be easily calculated using the open-source symptomcheckR package [26]. We encourage researchers to build on these resources to improve the quality of future evaluations and enhance cross-study comparability.

Conclusions

In this paper, we summarize the limitations and challenges of previous studies evaluating symptom checkers using vignettes. In recent years, several empirical studies have addressed most of these limitations individually, yet a unified methodological and reporting framework integrating these findings was missing. We present a preclinical framework and the corresponding SCARF checklist upon which future vignette-based symptom checker evaluations can build to address generalizability, input variability, gold standard assignment, and metrics, and we highlight several open access resources that evaluators can use. By adopting this approach, researchers can identify well-performing tools for more cost-efficient clinical trials and can significantly increase the quality and comparability of vignette-based symptom checker evaluation studies, thereby enabling reliable evidence syntheses. This can help move closer to assessing and improving the effectiveness of symptom checkers, diagnostic decision support systems, and large language models.

XSL•FO
RenderX

Funding

The authors acknowledge support from the Open Access Publication Fund of Technische Universität Berlin.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Symptom Checker Accuracy Reporting Framework (SCARF) checklist (editable version).
[DOCX File, 22 KB - humanfactors_v13i1e76168_app1.docx]

Multimedia Appendix 2
Symptom Checker Accuracy Reporting Framework (SCARF) checklist (PDF version).
[PDF File, 82 KB - humanfactors_v13i1e76168_app2.pdf]

References

1. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015 Jul 8;351(1–9):h3480. [doi: 10.1136/bmj.h3480] [Medline: 26157077]
2. Hill MG, Sim M, Mills B. The quality of diagnosis and triage advice provided by free online symptom checkers and apps in Australia. Med J Aust 2020 Jun;212(11):514-519. [doi: 10.5694/mja2.50600] [Medline: 32391611]
3. Ceney A, Tolond S, Glowinski A, Marks B, Swift S, Palser T. Accuracy of online symptom checkers and the potential impact on service utilisation. PLoS ONE 2021;16(7):e0254088. [doi: 10.1371/journal.pone.0254088] [Medline: 34265845]
4. Schmieding ML, Kopka M, Schmidt K, Schulz-Niethammer S, Balzer F, Feufel MA. Triage accuracy of symptom checker apps: 5-year follow-up evaluation. J Med Internet Res 2022 May 10;24(5):e31810. [doi: 10.2196/31810] [Medline: 35536633]
5. Ito N, Kadomatsu S, Fujisawa M, et al. The accuracy and potential racial and ethnic biases of GPT-4 in the diagnosis and triage of health conditions: evaluation study. JMIR Med Educ 2023 Nov 2;9:e47532. [doi: 10.2196/47532] [Medline: 37917120]
6. Wallace W, Chan C, Chidambaram S, et al. The diagnostic and triage accuracy of digital and online symptom checker tools: a systematic review. NPJ Digit Med 2022 Aug 17;5(1):118. [doi: 10.1038/s41746-022-00667-w] [Medline: 35977992]
7. Chambers D, Cantrell AJ, Johnson M, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. BMJ Open 2019 Aug 1;9(8):e027743. [doi: 10.1136/bmjopen-2018-027743] [Medline: 31375610]
8. Riboli-Sasco E, El-Osta A, Alaa A, et al. Triage and diagnostic accuracy of online symptom checkers: systematic review. J Med Internet Res 2023 Jun 2;25:e43803. [doi: 10.2196/43803] [Medline: 37266983]
9. Kopka M, von Kalckreuth N, Feufel MA. Accuracy of online symptom assessment applications, large language models, and laypeople for self-triage decisions. NPJ Digit Med 2025 Mar 25;8(1):178. [doi: 10.1038/s41746-025-01566-6] [Medline: 40133390]
10. Painter A, Hayhoe B, Riboli-Sasco E, El-Osta A. Online symptom checkers: recommendations for a vignette-based clinical evaluation standard. J Med Internet Res 2022 Oct 26;24(10):e37408. [doi: 10.2196/37408] [Medline: 36287594]
11. Meczner A, Cohen N, Qureshi A, et al. Controlling inputter variability in vignette studies assessing web-based symptom checkers: evaluation of current practice and recommendations for isolated accuracy metrics. JMIR Form Res 2024 May 31;8:e49907. [doi: 10.2196/49907] [Medline: 38820578]
12. El-Osta A, Webber I, Alaa A, et al. What is the suitability of clinical vignettes in benchmarking the performance of online symptom checkers? An audit study. BMJ Open 2022 Apr 27;12(4):e053566. [doi: 10.1136/bmjopen-2021-053566] [Medline: 35477872]
13. Kopka M, Napierala H, Privoznik M, Sapunova D, Zhang S, Feufel MA. The RepVig framework for designing use-case specific representative vignettes and evaluating triage accuracy of laypeople and symptom assessment applications. Sci Rep 2024 Dec 23;14(1):30614. [doi: 10.1038/s41598-024-83844-z] [Medline: 39715767]
14. Kopka M, Feufel MA, Berner ES, Schmieding ML. How suitable are clinical vignettes for the evaluation of symptom checker apps? A test theoretical perspective. Digit Health 2023;9:20552076231194929. [doi: 10.1177/20552076231194929] [Medline: 37614591]
15. Aboueid S, Meyer S, Wallace JR, Mahajan S, Chaurasia A. Young adults' perspectives on the use of symptom checkers for self-triage and self-diagnosis: qualitative study. JMIR Public Health Surveill 2021 Jan 6;7(1):e22637. [doi: 10.2196/22637] [Medline: 33404515]
16. Baccaglini L, Theriaque DW, Shuster JJ, Serrano G, Lalla RV. Validation of anamnestic diagnostic criteria for recurrent aphthous stomatitis. J Oral Pathology Medicine 2013 Apr;42(4):290-294. [doi: 10.1111/jop.12015]

XSL•FO
RenderX

17. Brunswik E. Representative design and probabilistic theory in a functional psychology. Psychol Rev 1955 May;62(3):193-217. [doi: 10.1037/h0047470] [Medline: 14371898]
18. Yu SWY, Ma A, Tsang VHM, Chung LSW, Leung SC, Leung LP. Triage accuracy of online symptom checkers for accident and emergency department patients. Hong Kong J Emerg Med 2020 Jul;27(4):217-222. [doi: 10.1177/1024907919842486]
19. Verzantvoort NCM, Teunis T, Verheij TJM, van der Velden AW. Self-triage for acute primary care via a smartphone application: practical, safe and efficient? PLoS ONE 2018;13(6):e0199284. [doi: 10.1371/journal.pone.0199284] [Medline: 29944708]
20. Kopka M, Schmieding ML, Rieger T, Roesler E, Balzer F, Feufel MA. Determinants of laypersons' trust in medical decision aids: randomized controlled trial. JMIR Hum Factors 2022 May 3;9(2):e35219. [doi: 10.2196/35219] [Medline: 35503248]
21. Kopka M, Wang SM, Kunz S, Schmid C, Feufel MA. Technology-supported self-triage decision making. NPJ Health Syst 2025 Jan 25;2(1):1-11. [doi: 10.1038/s44401-024-00008-x]
22. Fraser H, Coiera E, Wong D. Safety of patient-facing digital symptom checkers. The Lancet 2018 Nov;392(10161):2263-2264. [doi: 10.1016/S0140-6736(18)32819-8] [Medline: 30413281]
23. You JG, Hernandez-Boussard T, Pfeffer MA, Landman A, Mishuris RG. Clinical trials informed framework for real world clinical implementation and deployment of artificial intelligence applications. NPJ Digit Med 2025 Feb 17;8(1):107. [doi: 10.1038/s41746-025-01506-4] [Medline: 39962232]
24. Kopka M, Feufel MA. Statistical refinement of patient-centered case vignettes for digital health research. Front Digit Health 2024;6:1411924. [doi: 10.3389/fdgth.2024.1411924] [Medline: 39498100]
25. Levine DM, Tuwani R, Kompa B, et al. The diagnostic and triage accuracy of the GPT-3 artificial intelligence model: an observational study. Lancet Digit Health 2024 Aug;6(8):e555-e561. [doi: 10.1016/S2589-7500(24)00097-9] [Medline: 39059888]
26. Kopka M, Feufel MA. Software symptomcheckR: an R package for analyzing and visualizing symptom checker triage performance. BMC Digit Health 2024 Jul 22;2(1):43. [doi: 10.1186/s44247-024-00096-7]
27. Knitza J, Hasanaj R, Beyer J, et al. Comparison of two symptom checkers (ada and symptoma) in the emergency department: randomized, crossover, head-to-head, double-blinded study. J Med Internet Res 2024 Aug 20;26:e56514. [doi: 10.2196/56514] [Medline: 39163594]

Abbreviations

AI: artificial intelligence
SCARF: symptom checker accuracy reporting framework

Edited by A Kushniruk; submitted 17.Apr.2025; peer-reviewed by A Meczner, AJ Wetzel, J Knitza; accepted 20.Nov.2025; published 16.Jan.2026.

Please cite as:
Kopka M, Feufel MA
How to Evaluate the Accuracy of Symptom Checkers and Diagnostic Decision Support Systems: Symptom Checker Accuracy Reporting Framework (SCARF)
JMIR Hum Factors 2026;13:e76168
URL: https://humanfactors.jmir.org/2026/1/e76168
doi:10.2196/76168

© Marvin Kopka, Markus A Feufel. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 16.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Understanding the Limits of Patient Safety Classification Systems for Health Information Technology–Related Incidents

Md Shafiqur Rahman Jabin[1,2], PhD

[1]Department of Medicine and Optometry, Linnaeus University, Kalmar, Sweden

[2]Faculty of Health and Social Care, University of Bradford, Bradford, West Yorkshire, United Kingdom

Corresponding Author:
Md Shafiqur Rahman Jabin, PhD
Department of Medicine and Optometry
Linnaeus University
Universitetsplatsen 1
Kalmar, SE-391 82
Sweden
Phone: 44 7915 673612
Email: mdshafiqur.rahmanjabin@lnu.se

Abstract

Patient safety classification systems are fundamental to surveillance, organizational learning, research, and governance because they enable adverse events and near misses to be organized into standardized categories for comparison and analysis. However, the increasing complexity of health information technology (HIT)–related patient safety incidents challenges the assumptions underpinning conventional classification approaches, as these incidents often emerge from dynamic, distributed, and evolving sociotechnical interactions rather than discrete, time-bounded events. In this Viewpoint, I argue that many of the challenges associated with classifying HIT-related patient safety incidents arise not simply from limitations of individual classification systems but from the inherent representational logic of classification itself. By viewing classification as a knowledge practice rather than merely a technical tool for organizing incident data, I contend that abstraction, boundary-setting, and standardization inevitably simplify complex sociotechnical processes and constrain how safety problems are represented, interpreted, and acted upon. I discuss 4 recurring representational limitations that characterize the application of patient safety classification systems to HIT-related incidents: fragmentation of sociotechnical interactions, loss of temporality and evolving processes, inadequate representation of scale and propagation across systems, and normalization of "use error" through simplified attribution of responsibility. These limitations can contribute to incomplete organizational learning, misaligned safety interventions, and challenges in interpreting and comparing classified patient safety data across health care settings. Rather than arguing against the continued use of patient safety classification systems, I propose that their strengths and limitations should be recognized simultaneously. Classification remains indispensable for surveillance, learning, and governance, but it should be interpreted as one component of a broader sociotechnical understanding of patient safety. Recognizing the representational limits of classification can support more reflexive interpretation of classification-based evidence and encourage complementary approaches that better capture the complexity of HIT-related patient safety.

(JMIR Hum Factors 2026;13:e91783) doi:10.2196/91783

KEYWORDS

sociotechnical systems; knowledge representation; incident reporting; safety science; information infrastructure; organizational learning; sensemaking; complex adaptive systems; data standardization; epistemic limits

Introduction

Classification as Infrastructure in Patient Safety

Classification systems are foundational to patient safety research, surveillance, and institutional learning. By transforming diverse and complex safety incidents into standardized categories, classification enables aggregation, comparison, and analysis across settings and over time. These systems underpin reporting infrastructures, research studies, quality improvement initiatives, and policy conversations, while shaping how patient safety problems are recognized and addressed [1-3].

XSL•FO
RenderX

Diversity of Patient Safety Classification Approaches

Patient safety classification schemes have been developed in a variety of contexts and based on diverse design logics. The International Classification for Patient Safety (ICPS), a World Health Organization (WHO)–created framework, represents a global effort to harmonize safety terminology and incident constructs [3]. Beyond ICPS, there are classification efforts tailored to specific settings or incident types (eg, primary care safety classifications developed through the LINNAEUS [Learning in an International Network About Errors and Understanding Safety] collaboration [4] and medication error taxonomies such as the NCC MERP [National Coordinating Council for Medication Error Reporting and Prevention] taxonomy [3,5,6]). Moreover, structured frameworks, such as human factors taxonomies, provide alternative logics for organizing contributing factors associated with safety events [7-9]. These varied systems share a core commitment: to render unsafe care legible through predefined categories and relationships.

Why Health Information Technology Challenges Classification

The expansion of health information technology (HIT) into clinical work has strained conventional classification logic. HIT-related safety incidents often involve sociotechnical interactions, distributed causality, temporal processes, and emergent behaviors that are not easily reducible to discrete event types and fixed causal categories [10,11]. Although HIT-specific classification efforts (eg, adaptations of Health Information Technology Classification System (HIT-CS) and classifications that integrate technology and device issues) have attempted to capture these phenomena, persistent difficulties in representing HIT-related incidents suggest deeper epistemic limits in how classification systems encode knowledge [1,12].

Although HIT-specific and HIT-adapted classification schemes have been developed to address these challenges, recurring representational difficulties suggest that the problem may not lie solely in the design or implementation of particular frameworks [13]. For example, duplicate medication orders associated with computerized physician order entry (CPOE) systems have been reported as medication-related or HIT-related patient safety incidents, yet such classifications do not fully represent the interacting software, clinical workflow, human, and organizational factors that contributed to their occurrence [14,15].

From Classification Design to Classification Limits

Prior work on HIT-related patient safety has addressed methodological challenges in studying technology-associated incidents and articulated interpretive perspectives on HIT-related risk. Building on this foundation, the present paper shifts attention from how incidents should be classified to what classification systems, by their nature, can and cannot represent. By making these inherent limits explicit, this Viewpoint clarifies the constraints on learning from classified safety data and suggests implications for future research in HIT safety.

Perspective Presented in This Viewpoint

This Viewpoint examines the inherent limits of patient safety classification systems when applied to HIT-related incidents by considering classification as a representational practice rather than simply a technical tool. I argue that 4 recurring representational limitations help explain why patient safety classification systems struggle to adequately represent HIT-related incidents: (1) fragmentation of sociotechnical interactions, (2) loss of temporality and evolving processes, (3) inadequate representation of scale and propagation across systems, and (4) normalization of “use error” through simplified attribution of responsibility. Rather than proposing a new taxonomy or evaluating existing classification systems, this paper examines classification itself as an epistemic practice [16-19]. Importantly, sociotechnical perspectives on HIT safety are well established in the literature, and the present Viewpoint builds on this body of work by examining how classification systems interact with these broader sociotechnical perspectives.

This Viewpoint is intended for patient safety researchers, HIT researchers, patient safety classification developers, health care quality and safety professionals, and policymakers involved in the design, interpretation, and use of patient safety classification systems.

Classification as a Representational Lens: Why Taxonomies Shape Interpretation Rather Than Merely Describe Events

Classification as Knowledge Production

Patient safety classification systems are often presented as neutral tools for organizing information about adverse events and near misses. In practice, however, classification operates as a representational lens that actively shapes how safety problems are understood [1,6,20]. By defining categories and boundaries, classification systems shape what is recognized as an incident, how causality is attributed, and which dimensions of harm are foregrounded or obscured [6,20,21].

The term “system” is used in several ways throughout this paper. Specifically, it may refer to the clinical work system, the HIT system, the incident-reporting system, or the patient safety classification system. Unless otherwise specified, this perspective focuses on classification systems as representational frameworks while recognizing their interaction with broader sociotechnical and reporting systems.

Abstraction and Standardization

At a fundamental level, all classification systems rely on abstraction and standardization. Classification makes complex phenomena easier to analyze by reducing them to predefined units. These units may include event types, contributing factors, or outcomes that can be applied consistently across cases. This abstraction enables aggregation and comparison across cases, supporting patient safety surveillance and learning. At the same time, abstraction necessarily compresses detail, strips away contextual nuance, and stabilizes fluid processes into discrete representational forms [21-23].

XSL•FO
RenderX

Boundary-Setting and Assumptions About Causality

Classification systems also impose boundaries. They delineate what counts as an "incident," what falls within the scope of patient safety, and where responsibility or causation is presumed to reside. These boundaries are embedded in category definitions, hierarchies, and exclusion criteria [21-24].

Classification, Sensemaking, and Governance

Classification influences not only data organization but also sensemaking and governance. Classified data travel across organizational and institutional contexts, informing research syntheses, quality improvement initiatives, and policy decisions [25,26]. As categories are reused and institutionalized, they can stabilize particular narratives about safety problems while marginalizing others. Over time, this can create feedback loops in which certain types of incidents are repeatedly recognized and addressed, while others remain poorly articulated or invisible [25,27,28].

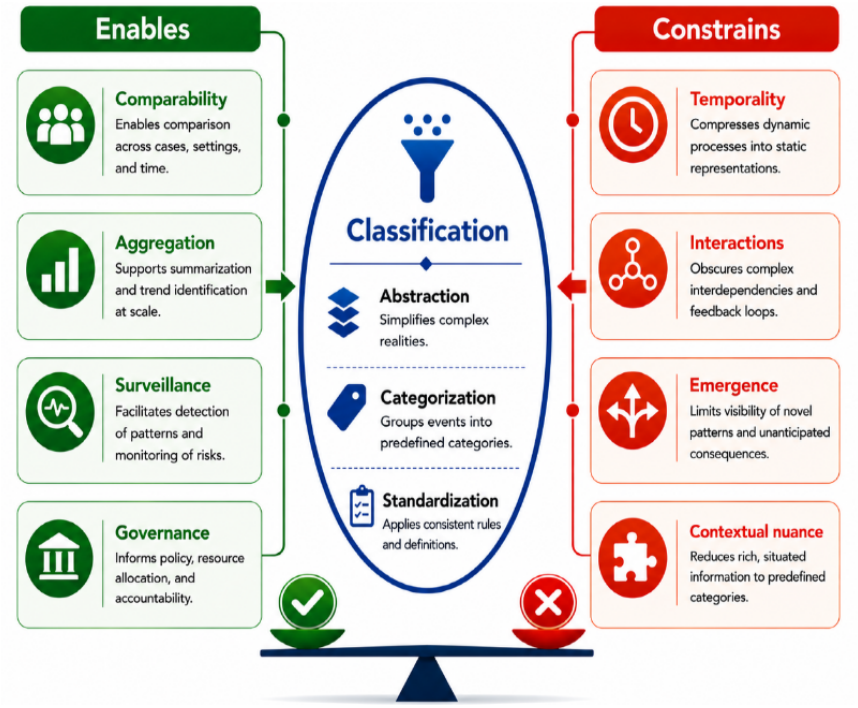
Inherent Limits of Classificatory Work

These characteristics of classification, including abstraction, boundary-setting, and stabilization, are not flaws but constitutive features of classificatory work. As such, they introduce limits that cannot be fully eliminated through refinement or expansion of taxonomies [23,25]. Adding categories, increasing granularity, or tailoring classifications to specific domains may improve representation of particular incident features. However, these refinements do not overcome the underlying logic of classification, which continues to stabilize complex sociotechnical processes into discrete, comparable categories [28]. Instead, they redistribute rather than eliminate representational constraints [27].

In this paper, the term epistemic limits refers to constraints that originate from the fundamental logic of classification itself. These limits differ from design trade-offs, methodological limitations, or pragmatic constraints associated with reporting systems. Design trade-offs involve decisions about how a classification system is constructed or executed, while methodological limitations stem from analytic techniques or data availability. Pragmatic constraints reflect the institutional or operational needs of incident-reporting systems. In contrast, epistemic limits arise from the core features of classification as a knowledge practice, especially abstraction, boundary-setting, and stabilization, which inevitably convert complex, evolving sociotechnical processes into discrete, comparable categories. These limits, therefore, endure across various taxonomies and implementation settings. Figure 1 illustrates the representational trade-offs inherent in patient safety classification. The same structural properties that enable standardized reporting and organizational learning also constrain the representation of complex sociotechnical processes.

Figure 1 illustrates how abstraction, categorization, and standardization simultaneously enable organizational learning while constraining the representation of complex sociotechnical processes.

Figure 1. Representational trade-offs in patient safety classification.



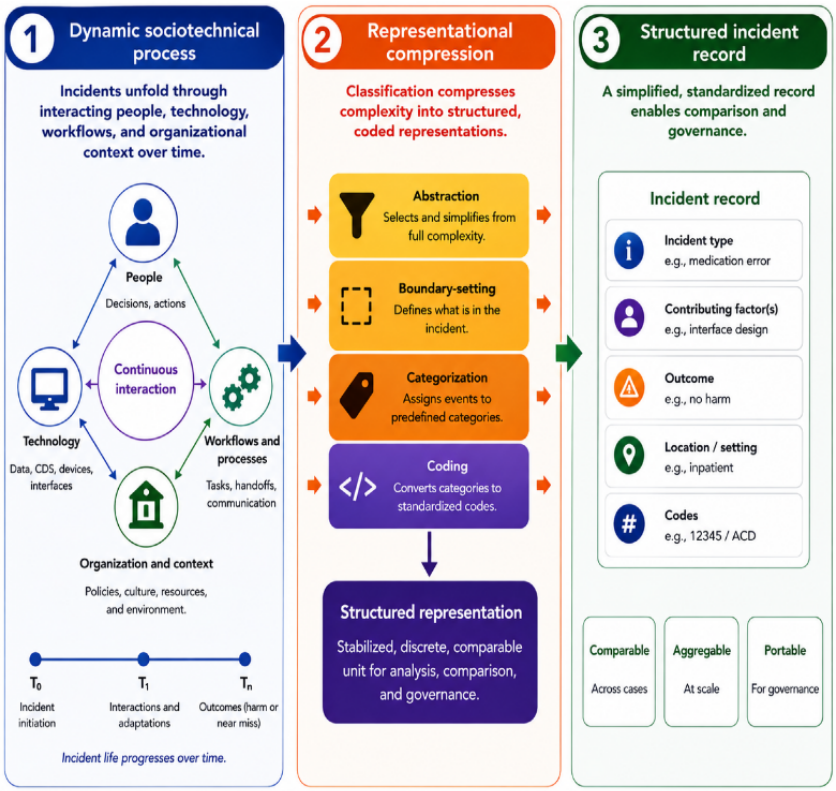
Implications for a Sociotechnical Perspective on Patient Safety

Understanding classification as a representational lens is essential for examining patient safety in complex sociotechnical settings. In such contexts, the tension between the need for standardized representation and the reality of distributed, evolving processes becomes particularly pronounced [18,29,30]. Figure 2 illustrates representational compression in patient safety classification, whereby complex sociotechnical processes are transformed into discrete, standardized incident representations that support recording, categorization, comparison, and analysis.

XSL•FO
RenderX

Taken together, Figure 2 illustrates the core argument of this paper: the difficulties encountered in applying patient safety classification systems to HIT-related incidents arise from fundamental properties of classification itself. By converting evolving sociotechnical processes into discrete, stabilized representations, classification simultaneously enables learning at scale and imposes limits on what can be known, compared, and acted upon.

Figure 2. Representational compression in patient safety classification. CDS: clinical decision support; EHR: electronic health record.



Overview of Patient Safety Classification Approaches Relevant to HIT

Illustrative Classification Traditions and Design Logics

Patient safety classification systems have been developed across diverse institutional, clinical, and analytic contexts. Although they differ in scope, purpose, and intended users, they share common representational commitments, relying on categorization, abstraction, and standardization to render safety incidents legible for analysis, comparison, and organizational learning [20,30,31]. The following examples illustrate how these shared design logics shape the representation of HIT-related safety problems.

Global and General Patient Safety Frameworks

Global patient safety classification efforts aim to establish common conceptual foundations across health care systems. The ICPS was developed to harmonize terminology and incident constructs internationally [23,32,33].

Such frameworks are deliberately generic, prioritizing breadth and interoperability over contextual specificity. Consequently, they typically conceptualize safety incidents as discrete events that can be decomposed into causal elements and outcomes. Although this design supports aggregation and high-level analysis, it also embeds assumptions about temporality, agency, and system boundaries that become consequential when applied to complex sociotechnical phenomena [1,23,32,33].

Reporting-Oriented Classification Systems

Many patient safety classifications are embedded in incident-reporting infrastructures and shaped by practical requirements for data collection, aggregation, and governance. In the United States, the Agency for Healthcare Research and Quality (AHRQ) Common Formats provide standardized categories for reporting adverse events, near misses, and unsafe conditions across care settings [25,34].

Reporting-oriented systems often emphasize clarity, usability, and consistency, favoring predefined categories that frontline reporters can apply. HIT-related issues are typically incorporated as contributing factors or device-related elements rather than as evolving sociotechnical processes [1,25,34].

Domain-Specific and Error-Focused Taxonomies

Some classification systems focus on specific clinical domains or types of harm, such as medication safety. The taxonomy developed by the NCC MERP categorizes medication errors by the stages of the medication-use process and the severity of patient harm [1,5].

Although not specifically designed for HIT, such systems are frequently used for incidents involving electronic prescribing, decision support, and medication administration technologies. In doing so, they often normalize distinctions between “use

XSL•FO
RenderX

error" and system malfunction, reflecting broader classificatory tendencies to localize error within identifiable actions rather than within distributed interactions [3,5].

HIT-Specific and HIT-Adapted Classification Schemes

In response to growing recognition of technology-related risk, HIT-specific classification schemes have been developed to capture incidents involving electronic health records, clinical decision support, and other digital tools. These systems typically expand contributing-factor categories to include software design, interface issues, workflow mismatches, and organizational context. They are often explicitly informed by sociotechnical perspectives [10,35-37].

Despite this specialization, HIT-focused classifications retain core classificatory features shared with broader patient safety taxonomies. Incidents are still represented as bounded events, categorized according to predefined dimensions, and analyzed through stabilized causal structures. Consequently, these systems illuminate how far domain-specific adaptation can go without escaping the underlying logic of classification [10,18,38].

Recently, classification systems have incorporated more detailed categories, mechanisms for linking related events, and structured narrative components to better capture temporal dynamics and sociotechnical interactions [39,40]. Although these developments improve representation, they do not overcome the representational limitations inherent in classification as a mode of organizing patient safety information.

Local and Adapted Classification Approaches

Beyond formal frameworks, many health care organizations and research teams develop local or adapted classification schemes to address perceived gaps in existing systems. These may involve extending categories, reinterpreting definitions, or combining elements from multiple frameworks [6,13,20,41].

The proliferation of adapted schemes underscores both the flexibility and the limits of classification. Efforts to tailor systems to local sociotechnical realities frequently reintroduce familiar representational challenges, suggesting that these challenges are not confined to particular taxonomies but arise from shared classificatory commitments [6,22,35,36].

Table 1 summarizes the illustrative classification design logics discussed in this section and highlights the shared representational commitments that shape the rendering of HIT-related safety problems.

Table 1. Illustrative classification design logics and their representational commitments.

Classification design logic	Primary unit of analysis	Assumptions about causality	Treatment of temporality	Typical handling of HIT[a]
Global patient safety frameworks	Discrete incident	Linear or decomposable	Event-based	Contributing factor
Reporting-oriented systems	Reportable event	Action- or device-linked	Snapshot at report time	Embedded category
Domain-specific taxonomies	Process step or error	Localized failure	Stage-based	Use error or malfunction
HIT-specific schemes	Incident with technical context	Sociotechnical (bounded)	Limited process capture	Expanded factors

[a]HIT: health information technology.

Why Classification Struggles to Represent HIT-Related Incidents

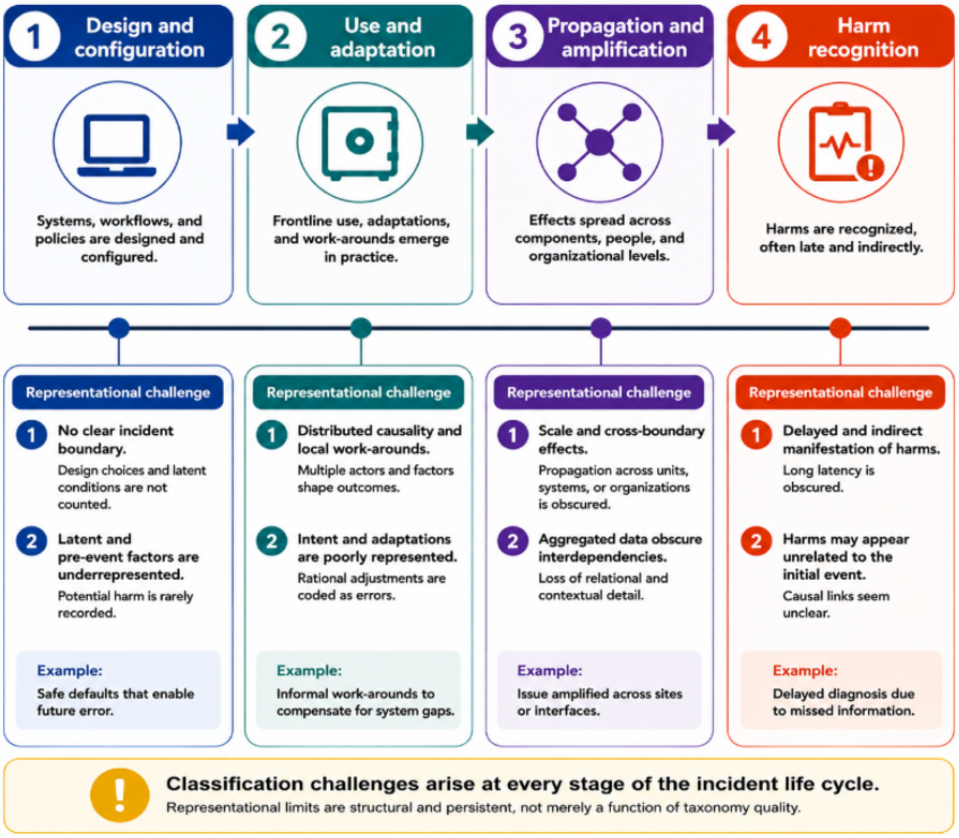
Overview

Patient safety classification systems vary in scope, purpose, and design. However, when applied to HIT-related incidents, they often encounter similar representational difficulties. These difficulties do not arise primarily from isolated design choices or implementation failures. Instead, they reflect recurrent representational tensions that emerge from the structural logic of classification itself [1,18].

To illustrate the representational challenges discussed in this section, consider a simplified example involving a duplicate medication order generated through CPOE [14]. A physician enters a medication order, but due to interface design and workflow constraints, the system allows the same order to be entered twice. The duplicate order is later detected and corrected before it can cause harm. Depending on the classification system used, this incident might be coded as a medication error, a user-entry mistake, a workflow issue, or a software design problem [14,42,43]. Figure 3 illustrates where patient safety classification systems encounter recurring representational challenges across the life cycle of HIT-related incidents, from system design and clinical use to propagation and recognition of harm.

XSL•FO
RenderX

Figure 3. Representational challenges across the life cycle of health information technology–related incidents.



Fragmentation of Sociotechnical Interactions

HIT-related safety incidents typically emerge from interactions among multiple system components, including software functionality, clinical workflows, organizational routines, and professional judgment. Classification systems, however, require incidents to be decomposed into discrete elements, such as event types, contributing factors, or actor roles, to render them categorizable [18,44]. In the CPOE duplicate-order example described previously, the incident could be attributed to user error, interface design, or workflow conditions, depending on the classificatory structure, illustrating how sociotechnical interactions become fragmented across multiple categories [14,42,43].

This decomposition breaks sociotechnical interactions into separate components. Consequently, the relational dynamics through which risk emerges may become less visible. Interdependencies between technology and practice are often redistributed across multiple categories, weakening the visibility of interaction effects. Consequently, the classified representation may suggest a set of parallel contributing factors rather than a coupled sociotechnical process [45,46].

This fragmentation is not an error of representation but a consequence of classificatory abstraction. Classification systems are designed to isolate elements that can be consistently named and compared, even when the phenomena of interest are fundamentally relational. For HIT-related incidents, this leads to representations that understate the role of interaction and overstate the independence of contributing components [18,26,46].

Loss of Temporality and Process

Many HIT-related safety problems unfold over extended periods. Design decisions made during system configuration may interact with evolving patterns of use, work-arounds, and adaptation before contributing to harm that becomes apparent only much later. Classification systems, however, typically represent incidents as temporally bounded events, anchored to a moment of recognition or reporting [15,46,47]. In the duplicate-order example, the classification record may capture the moment the error was detected, while the preceding sequence of interface interactions and workflow adaptations that enabled the duplication remains largely invisible [14,42,43].

This event-based representation collapses trajectories into snapshots. Temporal sequencing, latency, and cumulative effects are difficult to encode within categorical structures that privilege discrete occurrences. Even when temporal fields are included, they rarely capture how risks develop, propagate, and transform over time [47].

The loss of temporality limits classification systems' ability to represent HIT-related incidents as processes rather than outcomes. Consequently, downstream analysis may focus on proximal events while obscuring the longer-term dynamics through which sociotechnical risk accumulates [15,47].

Several patient safety classification systems have attempted to address these issues by introducing mechanisms intended to capture temporal and relational dynamics [48,49]. Although these features improve representation, they do not eliminate the underlying constraint that classification ultimately requires incidents to be stabilized into discrete, comparable categories.

XSL•FO
RenderX

Poor Handling of Scale and Propagation

HIT-related incidents frequently propagate beyond the local context in which they originate. A configuration error, interface design issue, or workflow mismatch may affect multiple users, units, or care settings, with effects that scale unevenly and interact with local practices [46,47]. If similar configurations or interface conditions exist across multiple clinical units, duplicate orders may recur, yet classification systems may record them as independent incidents rather than manifestations of a shared systemic issue [12,50].

Classification systems, however, are typically organized around localized incident units. They privilege representations that tie incidents to specific locations, actors, or moments, making it difficult to capture cross-boundary propagation or system-wide effects. When scale is represented, it is often treated as an attribute of individual incidents rather than as a dynamic property of sociotechnical systems [1,51].

This mismatch between localized incident units and distributed effects constrains the conceptual reach of classification-based data. Propagation is rendered as repetition rather than as a feature of system structure, limiting insight into how HIT-related risks amplify across contexts [46,47].

Normalization of "Use Error"

A recurrent feature of patient safety classification systems is the distinction between technical failure and human or "use" error. In HIT contexts, this distinction often becomes a default mode of representation, particularly when classification categories prioritize observable user actions over design-use coupling [11,52,53].

By treating "use error" as a categorical outcome, classification systems may normalize interpretations that assign responsibility to individual behavior, even when user actions are shaped by interface design, workflow constraints, or organizational pressures. Over time, this can reinforce a conceptual framework that treats sociotechnical misalignment as user deviation rather than as a property of system design [10,11,52].

This normalization effect does not require explicit attribution of blame. It emerges from the repeated use of categories that foreground action over interaction and outcome over process. As such, it reflects a deeper classificatory tendency to individualize causality in contexts where agency is distributed [1,38,51].

Representational Limitations Are Structural, Not System-Specific

Taken together, these breakdowns illustrate that the difficulties classification systems encounter with HIT-related incidents are not confined to particular taxonomies or domains. Fragmentation, loss of temporality, scale mismatch, and normalization of use error recur across classification approaches precisely because they arise from shared classificatory commitments [54,55].

Efforts to address these issues through increased granularity or domain-specific adaptation may shift how representational limitations appear, but they do not eliminate the underlying tension between standardized representation and sociotechnical complexity [1,38,51].

The increasing use of large language models (LLMs) and natural language processing (NLP) for analyzing unstructured patient safety reports does not fundamentally eliminate the epistemic limits discussed in this Viewpoint. Although these technologies may improve the extraction, organization, and classification of information from free-text narratives, they ultimately produce structured representations that remain subject to abstraction, boundary-setting, and representational compression [40,54,56]. Consequently, automated text classification may enhance the efficiency and consistency of incident analysis, but it does not overcome the inherent limits of classification as a knowledge practice. Instead, it may automate and scale the same structural constraints that shape how complex sociotechnical incidents are represented and interpreted.

Table 2 summarizes the 4 recurring representational limitations discussed in this Viewpoint and their implications for the classification of HIT-related incidents.

Table 2. Recurring representational limitations in classifying health information technology–related incidents.

Representational limitations	What is lost or distorted?	Why does classification produce this limit?
Sociotechnical fragmentation	Interactions and dependencies	Forced attribution to components
Loss of temporality	Incident trajectories	Event-based representation
Scale and propagation	System-wide effects	Localized incident units
Normalization of use error	Design-use coupling	Person-centered categories

Consequences for Research and Practice

Overview

The representational constraints described in the previous section have important consequences for how patient safety data are interpreted, compared, and acted upon in both research and practice [1,57].

Classification-based incident data rarely function as the sole basis for safety improvement. Instead, they typically serve as starting points for root cause analysis, sociotechnical assessment, and organizational learning. In HIT, these investigations are often informed by sociotechnical frameworks such as the SAFER (Safety Assurance Factors for Electronic Health Record Resilience) Guides and related evaluation methods for electronic health record safety [18,56,58].

Misleading Learning and Partial Explanations

Classification systems support learning by enabling aggregation and comparison across incidents. However, when HIT-related

XSL•FO
RenderX

safety problems are represented as fragmented, temporally bounded events, the resulting learning may be partial or misleading. Aggregated patterns can obscure the sociotechnical interactions and trajectories through which harm emerges, encouraging interpretations that privilege proximal causes over longer-term or distributed dynamics [1,25,57]. When such interactions are flattened into discrete categories, learning may focus on symptoms rather than on the structural conditions that shape risk [1,25,57].

Misaligned Interventions and Corrective Actions

The way incidents are classified influences not only interpretation but also intervention. When classified representations foreground individual actions or isolated failures, they can channel responses toward training, compliance, or local fixes, even when underlying issues are rooted in system design or organizational structure [30,47].

In HIT contexts, interventions derived from classification-based analyses may therefore focus on user behavior while leaving design-use mismatches or workflow constraints unaddressed. Over time, this can lead to repeated corrective efforts that fail to engage with the sources of sociotechnical risk, reinforcing cycles of reporting without commensurate improvement [30,46].

Comparability Problems Across Studies and Settings

One of the primary motivations for classification is to enable comparability across studies, organizations, and jurisdictions. However, when HIT-related incidents are forced into stabilized categories that inadequately capture sociotechnical complexity, apparent comparability may mask substantive differences in how incidents unfold across contexts [22,33].

Differences in local workflows, system configurations, and organizational practices can lead to divergent interpretations of similar classified categories. Consequently, studies drawing on classified incident data may appear comparable at the level of codes while differing substantially in what those codes represent in practice. This complicates synthesis across studies and limits the interpretability of comparative findings in HIT-related patient safety research [22,23,45].

Consequences as Structural, Not Correctable by Refinement

Taken together, these consequences indicate that the challenges associated with classifying HIT-related incidents are structural rather than contingent. Although greater granularity, additional categories, and domain-specific adaptations may improve representation, they do not eliminate the fundamental trade-offs inherent in classification-based representation [18,22].

Recognizing these consequences is not an argument against classification but a call for interpretive caution. Classification systems remain indispensable for surveillance, aggregation, and governance, yet their outputs should be understood as partial representations shaped by underlying design logics. Without this perspective, classification-based knowledge may be overextended beyond its representational limits [18,22].

Implications for Future Research and Interpretation

Overview

Recognizing the structural limits of classification systems has important implications for how future analyses of HIT-related patient safety are conceptualized and interpreted. These implications are particularly relevant for researchers analyzing patient safety datasets, HIT safety specialists interpreting incident reports, and health care organizations using classification-based data to guide quality improvement or governance activities [1]. Rather than seeking a single, more comprehensive classification framework, future research may benefit from approaches that explicitly acknowledge the partial and situated nature of classificatory representations [25,59]. Classification systems are best viewed as one part of a larger safety analysis ecosystem rather than as complete representations of complex sociotechnical events.

Why No Single Framework Is Sufficient

The recurrent representational constraints discussed in this Viewpoint suggest that no single classification framework can fully capture the complexity of HIT-related safety incidents. Sociotechnical phenomena are characterized by interaction, emergence, and temporal evolution, whereas classification systems necessarily stabilize phenomena into bounded, comparable units. This tension cannot be resolved through additional categories or greater granularity alone [1,51,60]. Consequently, classification-based data should be interpreted as partial representations of complex sociotechnical processes. Although such data remain essential for identifying patterns of harm, prioritizing interventions, and comparing safety performance across settings, they should be interpreted alongside complementary approaches that preserve temporal, organizational, and sociotechnical context.

Efforts to design more inclusive or detailed taxonomies may improve the visibility of certain dimensions of risk, but they simultaneously introduce new boundary problems and interpretive constraints. Consequently, attempts to “solve” classificatory limitations by expanding the framework risk reproducing the same representational trade-offs in different forms. These observations suggest the value of a more reflexive interpretation of classification-based safety data, particularly in sociotechnical contexts [1,51,60]. Recognizing the limits of classification can help analysts interpret incident data more cautiously, particularly when drawing conclusions about causality, system-level risk, or the effectiveness of interventions.

The Need for Reflexive Use of Classification Data

Future patient safety analyses that rely on classified incident data should adopt a reflexive stance toward the categories they use. Rather than treating classifications as transparent representations of underlying events, analysts should consider how category definitions, incident boundaries, and causal structures shape what becomes visible and actionable [18,51]. Without this reflexive orientation, there is a risk that classification outputs will be overinterpreted as direct reflections of sociotechnical reality [26,45,60].

XSL•FO
RenderX

Layered and Complementary Analytic Perspectives

Given the limits of any single representational approach, future research on HIT-related patient safety may benefit from layered perspectives that combine classification-based data with complementary approaches. These may include qualitative inquiry, sociotechnical modeling, process tracing, or narrative analysis, each of which foregrounds dimensions of interaction, temporality, context, and emergence that classification tends to suppress [1,51,61].

Classification systems may remain central for surveillance and aggregation, while complementary approaches are used to explore interaction, temporality, and emergence. Together, these perspectives can support a more nuanced understanding of sociotechnical risk without relying on a single conceptual framework [1,51,61].

From Representational Limits to Analytic Awareness

The implications of this Viewpoint are not methodological prescriptions but a call for more reflexive interpretation of classification-based evidence. By making the representational limits of classification explicit, future research can better align its interpretations and inferences with what classification-based data can reasonably support. This awareness can help prevent overgeneralization, misplaced attribution, and premature conclusions about causality or the effectiveness of interventions [1,51,62].

Treating classification as one analytic lens among many, rather than as a definitive account of safety, may enable more responsible and context-sensitive interpretations of HIT-related patient safety data [1,51,62].

Discussion

Patient safety classification systems have become indispensable for surveillance, organizational learning, research, and governance because they provide a standardized means of organizing complex safety information. However, I argue that their value should not obscure an important limitation: classification is not simply a neutral mechanism for recording events but a representational practice that shapes how safety problems are perceived, interpreted, and addressed [13,51]. This distinction is particularly important for HIT-related patient safety incidents, which emerge through dynamic interactions among technologies, clinical workflows, organizational contexts, and human decision-making rather than as discrete, isolated events [35,36].

From this perspective, the recurring representational limitations discussed in this Viewpoint should not be understood as shortcomings of particular classification systems or reporting infrastructures. Rather, they reflect the inherent logic of classification itself. Processes such as abstraction, boundary-setting, and standardization make comparison and aggregation possible, but they also simplify the complexity of sociotechnical systems. Consequently, aspects of HIT-related incidents, including evolving interactions, temporal dynamics, distributed causation, and contextual influences, may be compressed or excluded during classification. These trade-offs are intrinsic to representational systems and cannot be eliminated solely by developing more detailed or domain-specific taxonomies.

Recognizing these representational limits has important implications for patient safety research and practice. Classification-based datasets are often interpreted as comprehensive accounts of safety events and are subsequently used to inform organizational learning, quality improvement, policy development, and regulatory decision-making. However, when classified data are treated as complete representations of underlying events, they may support only a partial understanding of complex safety problems. This can contribute to interventions that address visible categories of harm while overlooking the sociotechnical conditions that produced them, and it may complicate comparisons across organizations and health care settings where contextual factors differ substantially.

My argument is therefore not that patient safety classification systems should be replaced or that more comprehensive classification schemes will fully resolve these challenges. Greater granularity and HIT-specific refinement can undoubtedly improve representation, but they cannot eliminate the fundamental trade-offs inherent in any attempt to translate complex, evolving practices into stable categories. Instead, classification should be understood as one conceptual perspective among several, with its outputs interpreted alongside complementary approaches that preserve temporal, organizational, and sociotechnical context.

Although this Viewpoint focuses on HIT-related patient safety, the underlying argument extends more broadly to contemporary health care. As digital technologies become increasingly embedded within clinical practice and health systems continue to rely on standardized data infrastructures for surveillance, benchmarking, accountability, and learning, similar representational tensions are likely to emerge wherever complex practices are reduced to standardized forms of knowledge. Appreciating these tensions is therefore essential not only for interpreting patient safety data but also for understanding the broader strengths and limitations of data-driven governance in health care.

The contribution of this Viewpoint is not to advocate for another patient safety classification framework. Rather, it is to clarify the epistemic boundaries within which classification-based patient safety data should be interpreted. By making the limits of classification explicit, I encourage a more reflexive use of classification systems that acknowledges both their indispensable role in surveillance, organizational learning, and governance and their inability to fully capture the complexity of HIT-related patient safety. Such an approach supports a more balanced interpretation of classification-based evidence and reinforces the importance of combining standardized classification with complementary sociotechnical perspectives when investigating, learning from, and responding to digital health safety incidents.

Conclusions

This Viewpoint argues that patient safety classification systems are indispensable for organizing, interpreting, and learning from adverse events and near misses. However, when applied to

XSL•FO
RenderX

HIT-related incidents, they encounter recurring representational limitations that cannot be fully overcome through refinement or domain-specific adaptation. These limitations arise from the inherent properties of classification itself (ie, abstraction, boundary-setting, and stabilization), which transform dynamic sociotechnical processes into stable, comparable categories.

Viewing classification as a representational lens, rather than as a neutral container of safety information, helps explain why fragmentation of sociotechnical interactions, loss of temporality, inadequate representation of scale and propagation, and normalization of use error recur across different classificatory traditions. These limitations are not shortcomings of individual classification systems but predictable consequences of representing complex, evolving sociotechnical phenomena through standardized categories.

The contribution of this Viewpoint is not to advocate for another classification framework but to clarify the epistemic boundaries within which classification-based patient safety data should be interpreted. Recognizing both the strengths and the representational limits of classification supports a more reflexive use of classification-based evidence while reinforcing the indispensable role of classification systems in surveillance, organizational learning, research, and governance.

Acknowledgments

The author would like to thank colleagues and health care professionals whose prior discussions and insights into health information technology–related patient safety informed the development of this Viewpoint. During the preparation of this manuscript, the author used ChatGPT (GPT-5.5; OpenAI) to assist with language refinement, editing, and improving the clarity and readability of the manuscript. The author takes full responsibility for the content of the manuscript and has reviewed and verified all outputs generated by the AI tool.

Data Availability

No new data were generated or analyzed in this study. This Viewpoint paper is informed by previously published studies and the author's prior research. As such, data sharing is not applicable.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' Contributions

The author was solely responsible for the conception, conceptual development, literature review and synthesis, and writing of the manuscript.

Conflicts of Interest

None declared.

References

1. Jabin MSR. Identifying and Characterising Problems Arising From Interactions Between Medical Imaging and Health Information Technology as a Basis for Improvements in Practice [Phd Thesis]. Adelaide, Australia: University of South Australia; 2019.
2. Wyatt KD, Benning TJ, Morgenthaler TI, Arteaga GM. Development of a taxonomy for medication-related patient safety events related to health information technology in pediatrics. Appl Clin Inform 2020 Oct;11(5):714-724 [FREE Full text] [doi: 10.1055/s-0040-1717084] [Medline: 33113568]
3. Brixey J, Johnson TR, Zhang J. Evaluating a medical error taxonomy. Proc AMIA Symp 2002:71-75 [FREE Full text] [Medline: 12463789]
4. Klemp K, Dovey S, Valderas JM, Rohe J, Godycki-Cwirko M, Elliott P, et al. Developing a patient safety incident classification system for primary care. A literature review and Delphi-survey by the LINNEAUS collaboration on patient safety in primary care. Eur J Gen Pract 2015 Sep;21 Suppl(sup1):35-38 [FREE Full text] [doi: 10.3109/13814788.2015.1043723] [Medline: 26339834]
5. Zhu M, Huang Y, Gong Y, Brixey JJ, Zhang J, Turley JP. Mapping concepts in medical error taxonomies. AMIA Annu Symp Proc 2006;2006:1159 [FREE Full text] [Medline: 17238778]
6. Taheri Moghadam S, Sheikhtaheri A, Hooman N. Patient safety classifications, taxonomies and ontologies, part 2: a systematic review on content coverage. J Biomed Inform 2023 Dec;148:104549 [FREE Full text] [doi: 10.1016/j.jbi.2023.104549] [Medline: 37984548]
7. Roggow BJ, Zarei E. Human factors analysis and classification system (HFACS): development, tiers, nanocodes, application, and adaptations. In: Safety Causation Analysis in Sociotechnical Systems: Advanced Models and Techniques. Cham, Switzerland: Springer; Jul 2024:151-180.

XSL•FO
RenderX

8. Holden RJ, Carayon P, Gurses AP, Hoonakker P, Hundt AS, Ozok AA, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 2013;56(11):1669-1686 [FREE Full text] [doi: 10.1080/00140139.2013.838643] [Medline: 24088063]
9. Shappell SA, Wiegmann DA. The human factors analysis and classification system–HFACS. U.S. Department of Transportation, Federal Aviation Administration. 2000. URL: https://commons.erau.edu/publication/737/ [accessed 2026-08-18]
10. Magrabi F, Liaw ST, Arachi D, Runciman W, Coiera E, Kidd MR. Identifying patient safety problems associated with information technology in general practice: an analysis of incident reports. BMJ Qual Saf 2016 Nov;25(11):870-880. [doi: 10.1136/bmjqs-2015-004323] [Medline: 26543068]
11. Magrabi F, Baker M, Sinha I, Ong MS, Harrison S, Kidd MR, et al. Clinical safety of England's national programme for IT: a retrospective analysis of all reported safety events 2005 to 2011. Int J Med Inform 2015 Mar;84(3):198-206. [doi: 10.1016/j.ijmedinf.2014.12.003] [Medline: 25617015]
12. Pan D, Nilsson E, Rahman Jabin MS. A review of incidents related to health information technology in Swedish healthcare to characterise system issues as a basis for improvement in clinical practice. Health Informatics J 2024;30(3):14604582241270742 [FREE Full text] [doi: 10.1177/14604582241270742] [Medline: 39116887]
13. Jabin MS. A qualitative, multi-framework methodology for analysing health information technology-related patient safety incidents. Front Digit Health 2026 May 29;8:1786358 [FREE Full text] [doi: 10.3389/fdgth.2026.1786358] [Medline: 42294047]
14. Horsky J, Kuperman GJ, Patel VL. Comprehensive analysis of a medication dosing error related to CPOE. J Am Med Inform Assoc 2005;12(4):377-382 [FREE Full text] [doi: 10.1197/jamia.M1740] [Medline: 15802485]
15. Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE, et al. Role of computerized physician order entry systems in facilitating medication errors. JAMA 2005 Mar 09;293(10):1197-1203. [doi: 10.1001/jama.293.10.1197] [Medline: 15755942]
16. Carayon P, Schoofs Hundt A, Karsh BT, Gurses AP, Alvarado CJ, Smith M, et al. Work system design for patient safety: the SEIPS model. Qual Saf Health Care 2006 Dec;15 Suppl 1(Suppl 1):i50-i58 [FREE Full text] [doi: 10.1136/qshc.2005.015842] [Medline: 17142610]
17. Carayon P, Wooldridge A, Hoonakker P, Hundt AS, Kelly MM. SEIPS 3.0: human-centered design of the patient journey for patient safety. Appl Ergon 2020 Apr;84:103033 [FREE Full text] [doi: 10.1016/j.apergo.2019.103033] [Medline: 31987516]
18. Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Qual Saf Health Care 2010 Oct;19 Suppl 3(Suppl 3):i68-i74 [FREE Full text] [doi: 10.1136/qshc.2010.042085] [Medline: 20959322]
19. Bowker GC, Star SL. Sorting Things Out: Classification and Its Consequences. Cambridge, MA: MIT Press; 2000.
20. Taheri Moghadam S, Hooman N, Sheikhtaheri A. Patient safety classifications for Health Information Technology (HIT) and medical devices: a review on available systems. Stud Health Technol Inform 2022 May 16;293:153-160. [doi: 10.3233/SHTI220362] [Medline: 35592975]
21. Conceptual framework for the international classification for patient safety: final technical report. World Health Organization. 2009 Jan 15. URL: https://www.who.int/publications/i/item/WHO-IER-PSP-2010.2 [accessed 2026-08-17]
22. Lorbeer ER. Sorting things out: classification and its consequences. Issues Sci Technol Librariansh 2000(25) [FREE Full text] [doi: 10.29173/istl1795]
23. Runciman W, Hibbert P, Thomson R, Van Der Schaaf T, Sherman H, Lewalle P. Towards an International Classification for Patient Safety: key concepts and terms. Int J Qual Health Care 2009 Feb;21(1):18-26 [FREE Full text] [doi: 10.1093/intqhc/mzn057] [Medline: 19147597]
24. Reason J. Human error: models and management. BMJ 2000 Mar 18;320(7237):768-770 [FREE Full text] [doi: 10.1136/bmj.320.7237.768] [Medline: 10720363]
25. Vincent C. Incident reporting and patient safety. BMJ 2007 Jan 13;334(7584):51. [doi: 10.1136/bmj.39071.441609.80] [Medline: 17218667]
26. Braithwaite J, Runciman WB, Merry AF. Towards safer, better healthcare: harnessing the natural properties of complex sociotechnical systems. Qual Saf Health Care 2009 Feb;18(1):37-41 [FREE Full text] [doi: 10.1136/qshc.2007.023317] [Medline: 19204130]
27. Waring JJ. Beyond blame: cultural barriers to medical incident reporting. Soc Sci Med 2005 May;60(9):1927-1935. [doi: 10.1016/j.socscimed.2004.08.055] [Medline: 15743644]
28. Weick KE. Sensemaking in Organizations. Thousand Oaks, CA: SAGE Publications; 1995.
29. Perrow C. Normal Accidents: Living with High Risk Technologies. New York, NY: Basic Books; 1984.
30. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc 2004;11(2):104-112 [FREE Full text] [doi: 10.1197/jamia.M1471] [Medline: 14633936]
31. Runciman B, Merry A, Walton M. Safety and Ethics in Healthcare: A Guide to Getting it Right. London, UK: CRC Press; 2007.

XSL•FO
RenderX

32. Thomson R, Lewalle P, Sherman H, Hibbert P, Runciman W, Castro G. Towards an International Classification for Patient Safety: a Delphi survey. Int J Qual Health Care 2009 Feb;21(1):9-17 [FREE Full text] [doi: 10.1093/intqhc/mzn055] [Medline: 19147596]
33. World Alliance For Patient Safety Drafting Group, Sherman H, Castro G, Fletcher M, World Alliance for Patient Safety, Hatlie M, et al. Towards an International Classification for Patient Safety: the conceptual framework. Int J Qual Health Care 2009 Feb;21(1):2-8 [FREE Full text] [doi: 10.1093/intqhc/mzn054] [Medline: 19147595]
34. Device or medical/surgical supply dashboard. Agency for Healthcare Research and Quality. 2025. URL: https://www.ahrq.gov/npsd/data/dashboard/devices.html [accessed 2026-08-17]
35. Jabin MS. Why health information technology safety problems remain invisible. Front Digit Health 2026 May 15;8:1785141 [FREE Full text] [doi: 10.3389/fdgth.2026.1785141] [Medline: 42221162]
36. Jabin MS. Why digital health fails silently: a sociotechnical theory of health information technology-related risk. Front Digit Health 2026 May 15;8:1785086 [FREE Full text] [doi: 10.3389/fdgth.2026.1785086] [Medline: 42245003]
37. Jabin MS. Health information technology-related loss of central surveillance data in a heart intensive care unit: multi-framework case report. JMIR Hum Factors 2026 Apr 09;13:e92560 [FREE Full text] [doi: 10.2196/92560] [Medline: 41955001]
38. Coiera E. Information system safety. In: Guide to Health Informatics. London, UK: CRC Press; 2015.
39. Gqaleni TM, Mkhize SW, Chironda G. Patient safety incident reporting and learning guidelines implemented by health care professionals in specialized care units: scoping review. J Med Internet Res 2024 Oct 04;26:e48580 [FREE Full text] [doi: 10.2196/48580] [Medline: 39365987]
40. Leistikow I, Mulder S, Vesseur J, Robben P. Learning from incidents in healthcare: the journey, not the arrival, matters. BMJ Qual Saf 2017 Mar;26(3):252-256 [FREE Full text] [doi: 10.1136/bmjqs-2015-004853] [Medline: 27037302]
41. Runciman WB, Williamson JA, Deakin A, Benveniste KA, Bannon K, Hibbert PD. An integrated framework for safety, quality and risk management: an information and incident management system based on a universal patient safety classification. Qual Saf Health Care 2006 Dec;15 Suppl 1(Suppl 1):i82-i90 [FREE Full text] [doi: 10.1136/qshc.2005.017467] [Medline: 17142615]
42. Collin S, Reeves BC, Hendy J, Fulop N, Hutchings A, Priedane E. Implementation of computerised physician order entry (CPOE) and picture archiving and communication systems (PACS) in the NHS: quantitative before and after study. BMJ 2008 Aug 14;337:a939 [FREE Full text] [doi: 10.1136/bmj.a939] [Medline: 18703655]
43. Thomas S, Fylan B, Silcock J, Jabin MS. Medication management using technology-based interventions in older people at home care settings: protocol for a scoping review. JMIR Res Protoc 2025 Oct 03;14:e75200 [FREE Full text] [doi: 10.2196/75200] [Medline: 41042854]
44. Castro GM, Buczkowski L, Hafner JM. The contribution of sociotechnical factors to health information technology-related sentinel events. Jt Comm J Qual Patient Saf 2016 Feb;42(2):70-76. [doi: 10.1016/s1553-7250(16)42008-8] [Medline: 26803035]
45. Aarts J. Towards safe electronic health records: a socio-technical perspective and the need for incident reporting. Health Policy Technol 2012 Mar;1(1):8-15. [doi: 10.1016/j.hlpt.2012.01.008]
46. Harrison MI, Koppel R, Bar-Lev S. Unintended consequences of information technologies in health care--an interactive sociotechnical analysis. J Am Med Inform Assoc 2007;14(5):542-549 [FREE Full text] [doi: 10.1197/jamia.M2384] [Medline: 17600093]
47. Koppel R, Wetterneck T, Telles JL, Karsh BT. Workarounds to barcode medication administration systems: their occurrences, causes, and threats to patient safety. J Am Med Inform Assoc 2008;15(4):408-423 [FREE Full text] [doi: 10.1197/jamia.M2616] [Medline: 18436903]
48. Hibbert PD, Molloy CJ, Hooper TD, Wiles LK, Runciman WB, Lachman P, et al. The application of the Global Trigger Tool: a systematic review. Int J Qual Health Care 2016 Dec 01;28(6):640-649. [doi: 10.1093/intqhc/mzw115] [Medline: 27664822]
49. Pham JC, Girard T, Pronovost PJ. What to do with healthcare incident reporting systems. J Public Health Res 2013 Dec 01;2(3):e27 [FREE Full text] [doi: 10.4081/jphr.2013.e27] [Medline: 25170498]
50. Jabin MS, Wepa D, Hassoun A. A case report of system configuration issue in medical imaging due to system upgrade-changes in hardware and software. Front Digit Health 2024 Sep 2;6:1371761 [FREE Full text] [doi: 10.3389/fdgth.2024.1371761] [Medline: 39347445]
51. Jabin MS. The need for a refined classification system and national incident reporting system for health information technology-related incidents. Front Digit Health 2024 Jul 26;6:1422396. [doi: 10.3389/fdgth.2024.1422396] [Medline: 39131183]
52. Magrabi F, Ong MS, Runciman W, Coiera E. An analysis of computer-related patient safety incidents to inform the development of a classification. J Am Med Inform Assoc 2010;17(6):663-670 [FREE Full text] [doi: 10.1136/jamia.2009.002444] [Medline: 20962128]
53. Sittig DF, Wright A, Coiera E, Magrabi F, Ratwani R, Bates DW, et al. Current challenges in health information technology-related patient safety. Health Informatics J 2020 Mar;26(1):181-189 [FREE Full text] [doi: 10.1177/1460458218814893] [Medline: 30537881]

XSL•FO
RenderX

54. Coiera E, Magrabi F, Talmon J. Engineering technology resilience through informatics safety science. J Am Med Inform Assoc 2017 Mar 01;24(2):244-245 [FREE Full text] [doi: 10.1093/jamia/ocw162] [Medline: 28040683]
55. Magrabi F, Ong MS, Runciman W, Coiera E. Using FDA reports to inform a classification for health information technology safety problems. J Am Med Inform Assoc 2012;19(1):45-53 [FREE Full text] [doi: 10.1136/amiajnl-2011-000369] [Medline: 21903979]
56. Singh H, Sittig DF. Measuring and improving patient safety through health information technology: the Health IT Safety framework. BMJ Qual Saf 2016 Apr;25(4):226-232 [FREE Full text] [doi: 10.1136/bmjqs-2015-004486] [Medline: 26369894]
57. Jones SS, Rudin RS, Perry T, Shekelle PG. Health information technology: an updated systematic review with a focus on meaningful use. Ann Intern Med 2014 Jan 07;160(1):48-54 [FREE Full text] [doi: 10.7326/M13-1531] [Medline: 24573664]
58. SAFER guides. Office of the National Coordinator for Health Information Technology. URL: https://healthit.gov/clinical-quality-and-safety/safer-guides/ [accessed 2026-03-11]
59. Vincent C, Burnett S, Carthey J. Safety measurement and monitoring in healthcare: a framework to guide clinical teams and healthcare organisations in maintaining safety. BMJ Qual Saf 2014 Aug;23(8):670-677 [FREE Full text] [doi: 10.1136/bmjqs-2013-002757] [Medline: 24764136]
60. Capra F, Luisi PL. The Systems View of Life: A Unifying Vision. Cambridge, MA: Cambridge University Press; 2014.
61. Alsaqqa HH. Healthcare organizations management: analyzing characteristics, features and factors, to identify gaps "scoping review". Health Serv Insights 2023 May 03;16:11786329231168130 [FREE Full text] [doi: 10.1177/11786329231168130] [Medline: 37153880]
62. Girard MJ, Lapides J, Roe CM. The fifth discipline: a systems learning model for building high-performing learning organizations. In: Handbook of Human Performance Technology: Principles, Practices, and Potential. Hoboken, NJ: John Wiley & Sons; 2006.

Abbreviations

AHRQ: Agency for Healthcare Research and Quality
CPOE: computerized physician order entry
HIT: health information technology
HIT-CS: Health Information Technology Classification System
ICPS: International Classification for Patient Safety
LINNAEUS: Learning in an International Network About Errors and Understanding Safety
LLM: large language model
NCC MERP: National Coordinating Council for Medication Error Reporting and Prevention
NLP: natural language processing
SAFER: Safety Assurance Factors for Electronic Health Record Resilience
WHO: World Health Organization

Edited by S Law; submitted 20.Jan.2026; peer-reviewed by L Faria, K Fuji; comments to author 10.Mar.2026; revised version received 28.Jul.2026; accepted 13.Aug.2026; published 28.Aug.2026.

Please cite as:
Jabin MSR
Understanding the Limits of Patient Safety Classification Systems for Health Information Technology–Related Incidents
JMIR Hum Factors 2026;13:e91783
URL: https://humanfactors.jmir.org/2026/1/e91783
doi:10.2196/91783
PMID:

©Md Shafiqur Rahman Jabin. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 28.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Viewpoint

Prioritizing Human Factors in the Clinical Translation of Mixed Reality Surgical Guidance

Shayan Bahadori[1], PhD; Peter Buckle[1], PhD

Department of Surgery & Cancer, NIHR HealthTech Research Centre In-vitro Diagnostics, Imperial College London, London, England, United Kingdom

Corresponding Author:
Shayan Bahadori, PhD
Department of Surgery & Cancer
NIHR HealthTech Research Centre In-vitro Diagnostics
Imperial College London
White City Campus
London, England, W12 0BZ
United Kingdom
Phone: 44 207 594 6879
Email: s.bahadori@imperial.ac.uk

Abstract

Mixed reality and augmented reality technologies are increasingly being explored for intraoperative visualization, navigation, and surgical decision support. However, translation into routine surgical practice will depend not only on technical accuracy but also on whether these systems are usable, physically tolerable, and robust in the operating room environment. In this Viewpoint, we highlight 3 human factors domains that warrant greater attention in mixed reality surgical guidance: physical characteristics, dexterity and motor demands, and environmental factors. Published work has repeatedly drawn attention to challenges, including ergonomic burden, limited field of view, headset weight, visual and perceptual constraints, fatigue, and interaction burdens associated with gesture and voice control. These issues are often compounded by environmental conditions, such as noise, glare, crowding, and occlusion in real clinical settings. Although mixed reality may improve spatial awareness and reduce attention shifts in some settings, these benefits remain constrained by persistent usability, workflow, and environmental limitations. Viewed through a broader systems human factors lens, discussion remains concentrated mainly on user-device interaction and selected task-level issues, with comparatively less attention to team, organizational, training, and governance factors. We argue that future human factors research should move beyond technical validation alone and adopt a broader systems perspective that addresses user diversity, sterile and low-burden interaction, environmental resilience, and the wider conditions required for safe clinical integration. Such an approach is essential if mixed reality surgical guidance systems are to become safe, usable, and clinically adoptable.

(JMIR Hum Factors 2026;13:e101995) doi:10.2196/101995

KEYWORDS

human factors; mixed reality; augmented reality; surgery; usability; artificial intelligence; AI

Introduction

Mixed reality and augmented reality technologies have emerged as promising tools for intraoperative visualization, surgical navigation, training, and decision support [1]. Their appeal in surgery lies in the ability to present patient-specific or task-relevant information directly within the user's field of view, with the potential to reduce reliance on external monitors and improve spatial understanding of anatomy [1,2].

Recent reviews have highlighted the potential benefits of mixed reality in the operating room, including gains in efficiency, learning, and anatomical understanding, while also identifying persistent barriers to implementation, such as ergonomic limitations, technical complexity, narrow field of view, battery limitations, and broader usability challenges [3,4]. Reviews focused specifically on optical see-through head-mounted displays in surgery have likewise highlighted recurring problems in perception, ease of use, context adaptation, interaction design, and occlusion [5]. More recent health care–oriented work has emphasized the need for evaluation frameworks that reflect authentic clinical and human visual conditions rather than idealized bench environments [6].

Despite growing technical interest in surgical mixed reality, the human factors conditions required for safe and effective use remain insufficiently addressed in relation to real-world clinical

XSL•FO
RenderX

translation [7]. In this paper, the term "human factors" is used in a broad ergonomics sense to refer to the interactions between people, technologies, tasks, environments, teams, and organizational conditions that influence safety, performance, and implementation in practice [8]. This Viewpoint is intended for researchers, developers, clinicians, and human factors specialists involved in the design, evaluation, and clinical translation of mixed reality surgical guidance systems. Against this broader perspective, we focus on 3 domains that, in our view, represent especially important translational priorities for mixed reality surgical guidance: physical characteristics, dexterity and motor demands, and environmental resilience. These are not presented as the only relevant human factors concerns but as recurring practical conditions that must be addressed if mixed reality systems are to move beyond proof-of-concept use and function reliably in routine operative settings. Mixed reality may offer genuine benefits in spatial awareness and in-field information access, but these advantages are likely to translate reliably into practice only if these practical conditions of use are addressed. Broader team, organizational, training, and governance issues remain essential to safe implementation, but they are treated here primarily as part of the wider systems context rather than as the main detailed focus of this Viewpoint.

Human Factors Framework for Mixed Reality Surgical Guidance

A useful way to frame mixed reality surgical guidance is through a systems human factors lens. In this paper, our framework is informed particularly by Moray's [8] systems approach to ergonomics, which emphasizes that performance and safety are shaped not only by individual users or technologies but also by interacting influences across multiple levels of the wider work system. This orientation is also consistent with broader sociotechnical and health care human factors perspectives that consider how individual, task, team, organizational, and wider system conditions interact in practice [8-10].

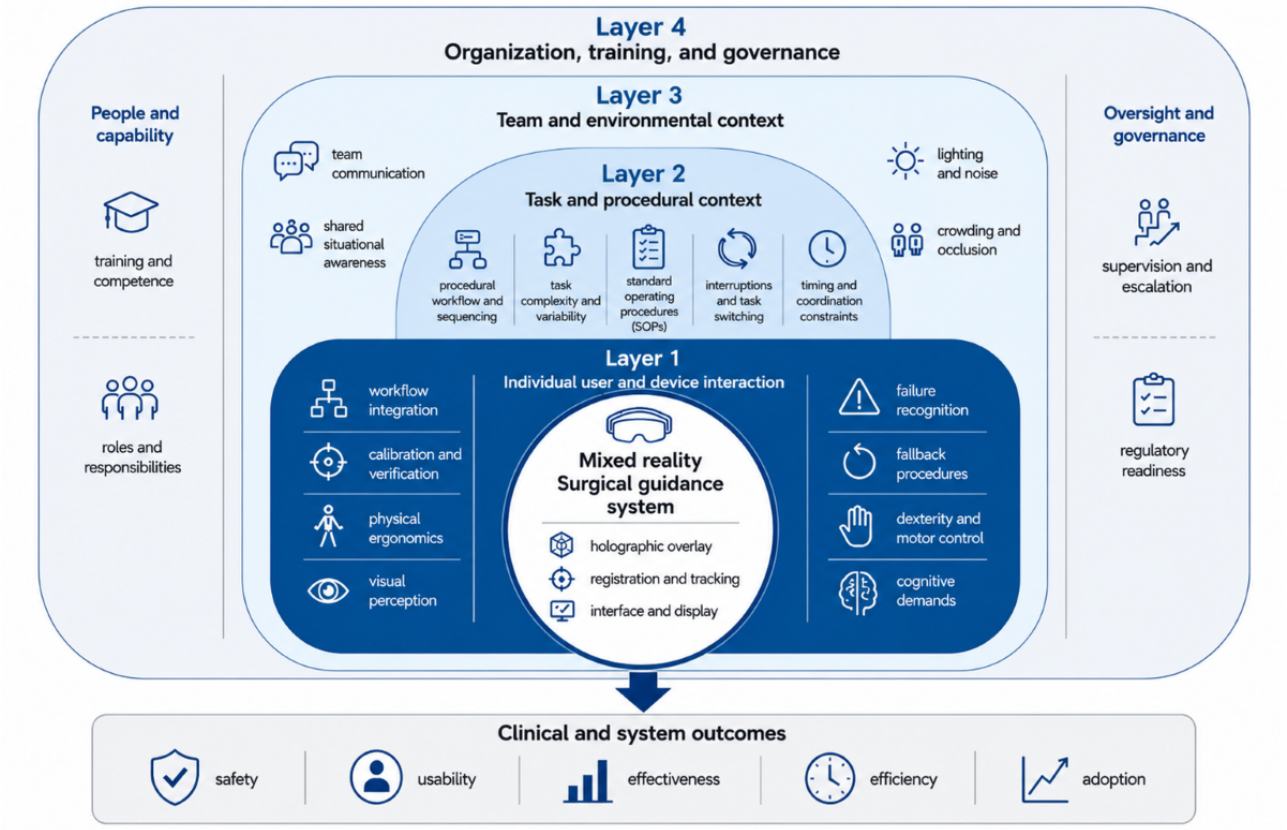
From this perspective, performance is shaped not only by the technology itself but also by interactions across multiple layers of the clinical work system, including individual user-device interaction, task and procedural context, team and environmental context, organization, training, and governance [8-10]. This means that safety, usability, effectiveness, and adoption depend not only on registration accuracy or display performance but also on physical ergonomics, visual perception, dexterity and motor control, workflow integration, calibration and verification, communication, supervision, and the ability to recognize failure and recover safely when the system becomes unreliable.

This framing matters because mixed reality surgical guidance is often discussed primarily in technical terms, whereas its real-world performance depends on how it functions within a layered clinical work system. Figure 1 [8] summarizes this systems human factors perspective and illustrates how factors operating at user, task, team, environmental, and organizational levels may influence clinical and system outcomes. It should therefore be understood not only as a conceptual summary but also as a framework for guiding evaluation. In particular, it highlights how constraints at one level, such as user perception or interaction burden, may be shaped by and have consequences for task demands, team communication, environmental conditions, and organizational arrangements.

The next section examines the 3 domains we consider to be most important (physical characteristics, dexterity and motor demands, and environmental resilience), focusing on how they influence whether mixed reality systems can be comfortably worn, reliably interacted with, and used robustly under operative conditions.

XSL•FO
RenderX

Figure 1. Human factors systems framework for mixed reality surgical guidance, informed by Moray's [8] systems approach to ergonomics and adapted to the clinical work-system context of mixed reality surgical guidance.



Key Human Factors Priorities for Mixed Reality Surgical Guidance

Translational Context for Mixed Reality Surgical Guidance

Published work on mixed reality in surgery spans operative guidance, planning, education, and navigation, with head-mounted devices such as Microsoft HoloLens and Magic Leap featuring prominently [3-5]. Much of this work has centered on navigation-heavy applications, particularly in maxillofacial and neurosurgical settings, where rigid anatomy supports relatively stable registration and validation [5,6]. Many studies remain based on phantom, cadaveric, pilot clinical, or early feasibility designs rather than definitive comparative evaluation [3-5]. From a human factors perspective, attention has focused mainly on user-device interaction, ergonomics, visualization, interface use, and selected workflow issues, while broader team, organizational, training, and governance questions remain less developed. In our view, this imbalance helps explain why mixed reality systems may show promise in early technical studies while remaining incompletely prepared for routine clinical use.

Physical Characteristics as a Translational Priority

Comfort, Weight, and Wearability

Ergonomic burden is one of the most consistently reported physical issues in the mixed reality surgical literature. Magalhães et al [3] identified ergonomic limitations, restricted field of view, and battery autonomy as persistent barriers to operating room adoption. Brockmeyer et al [11] likewise noted that surgeons often rate comfort reasonably well overall but still identify headset weight and size as important shortcomings of current HoloLens-based systems. Similarly, Doughty et al [5] highlighted device obtrusiveness and weight as contributors to both perceptual and ergonomic limitations.

Taken together, these observations suggest that wearability remains a translational issue rather than a minor design inconvenience. Broader evaluation work has also argued that wearability should be assessed under authentic clinical conditions, as laboratory testing may underestimate cumulative strain during prolonged use [6]. This is highly relevant in surgery, where even modest head or cervical load may influence posture, endurance, and willingness to use the system over the duration of a case. In our view, future human factors studies should treat comfort and wearability as core determinants of acceptability rather than as secondary user satisfaction measures.

Field of View and Visual Limitations

Limited field of view is repeatedly described as a key constraint [3,5,12]. Although these platforms may improve information access by placing digital content within the user's line of sight, the restricted display area can limit how much holographic information can be perceived at once and can reduce the practicality of direct overlay guidance [5]. Feasibility studies summarized in the review literature also note that HoloLens devices have a relatively narrow diagonal field of view, even

XSL•FO
RenderX

where users remain broadly satisfied with image quality for targeted tasks [6,13].

From a translational perspective, field of view should not be regarded simply as a hardware specification. It has direct implications for perceptual continuity, attention management, and the extent to which holographic guidance can be integrated into procedural workflow. Future evaluation should therefore examine how field-of-view constraints affect interpretation, task switching, and confidence in guidance during clinically relevant tasks.

Visual Perception and User-Specific Calibration

Physical user differences matter because perception in optical see-through systems is shaped by interpupillary distance, vergence-accommodation conflict, display calibration, and subjective depth interpretation [5,6]. Doughty et al [5] described incorrect interpupillary distance estimation, calibration instability, and depth perception limitations as recurrent contributors to perceptual error in surgical use of optical see-through head-mounted displays. The 2025 augmented reality head-mounted device (AR-HMD) evaluation review emphasized the importance of replicating human visual conditions, including interpupillary distance, pupil size, and observation geometry, when evaluating these devices [6].

These issues are especially important in surgery because a system may appear visually acceptable in generic testing while still introducing subtle perceptual errors that become clinically relevant during alignment, targeting, or spatial judgment. In our view, user-specific calibration should be treated as a core human factors issue rather than a purely technical one, particularly where holographic guidance is expected to support procedural precision.

Fatigue, Dizziness, and Visual Strain

Several reviews note that prolonged headset use may induce headaches, dizziness, motion-related discomfort, or visual fatigue [3,5,11]. The mixed reality operating room review explicitly identified dizziness and reduced field of view among the main ergonomic challenges [3]. These observations suggest that physical user research should extend beyond static comfort ratings and consider cumulative fatigue and tolerance over the duration of a case [3,6].

In a translational context, this matters because surgical mixed reality systems may be acceptable in short feasibility scenarios yet become less tolerable during longer cases or repeated use. Fatigue and visual strain should therefore be treated as key aspects of early evaluation, particularly for systems intended for routine intraoperative use.

Implications

Mixed reality systems cannot assume a one-size-fits-all user model [6,13]. Variation in vision, posture, wearable tolerance, and interaction with eyewear or theater attire is likely to affect safe use yet remains insufficiently addressed in current surgical studies [5,6]. In our view, future evaluation should explicitly incorporate user diversity and test whether systems remain comfortable, interpretable, and usable across realistic ranges of anthropometric and visual variation.

Dexterity and Motor Demands as a Barrier to Usable Interaction

Gesture and Voice Interaction Burden

Gesture and voice are among the most common interaction modes in surgical mixed reality systems [3,4]. The optical see-through review found that gestures and voice were the preferred interaction paradigm in many systems, especially HoloLens-based applications [5]. However, interaction burden remains a recurring problem [4,5]. In feasibility work, voice commands and gesture functions have often been rated less favorably than image quality or perceived overlay utility, indicating that interaction reliability may lag behind visual promise [5].

In our view, this is an important translational warning. Interaction methods may appear acceptable in proof-of-concept testing yet still prove impractical when deployed in time-pressured, sterile, and interruption-prone clinical environments. Human factors evaluation should therefore move beyond asking whether a command can be recognized and instead examine whether interaction remains efficient, intuitive, and low burden in the context of real surgical workflow.

Fine Motor Performance and Latency

Surgical performance depends on precise coordination, and mixed reality systems can interfere with this when interaction design introduces lag, poor input mapping, or repeated corrective actions [4-6]. The AR-HMD evaluation review highlighted that fine motor requirements are particularly critical in health care contexts requiring precision and that latency or poor finger tracking can degrade manipulation reliability [6]. Related work has also suggested that current systems may not yet offer the level of fine motor responsiveness required for delicate clinical use [4,6].

This has clear implications for surgical guidance. In high-precision environments, even minor delays or inconsistencies may alter the user's willingness to rely on the interface, increase hesitation, or encourage workarounds that bypass the system altogether. We therefore suggest that future evaluation should explicitly test the relationship between latency, manual precision, and operator confidence, rather than treating technical responsiveness as a purely engineering metric.

Hand-Eye Coordination and Perceptual-Motor Coupling

The optical see-through review identified hand-eye coordination as a reported human factors issue in some studies, particularly where users had to align tools, physical targets, and holographic guidance simultaneously [5]. Where interface control required users to adjust overlays, invoke commands, or reinterpret a partly registered scene, this increased both motor and cognitive demand [5].

In our view, this challenge is central to the real-world usability of mixed reality guidance. Surgical systems should not assume that visual overlay automatically improves performance. Instead, future studies should examine whether the system supports or disrupts perceptual-motor coupling during actual task execution, particularly in procedures that rely on fine targeting or continuous manual control.

XSL•FO
RenderX

Ease of Use and Workflow Interruption

Dexterity-related challenges are not limited to movement precision alone; they also include the need to interrupt procedural flow to manipulate content, navigate menus, or repeat commands [5]. Consistent with this, published work to date has described workflow disruption arising from the need to manually adapt virtual content during procedures [5,11]. In a sterile setting, any control modality that competes with operative hand use or requires repeated correction may reduce practical utility, even when the display itself appears technically sound [5,13].

This is where the distinction between technical functionality and clinical usability becomes most apparent. A system may function correctly in a narrow technical sense yet still impose enough interaction burden to make it unattractive or unsafe in practice. We therefore argue that workflow interruption should be treated as a primary human factors outcome in future evaluations of mixed reality surgical guidance systems.

Implications

Taken together, mixed reality systems should minimize manual configuration during critical procedural phases and avoid interaction schemes that add fine motor burden or attentional switching. For first-in-human surgical guidance platforms, evaluation should assess not only whether commands are technically successful but also whether interaction remains reliable, safe, and low burden under sterile, time-pressured conditions without impairing procedural control [5,13]. Such interaction demands may also increase cognitive workload by requiring users to divide attention between holographic content, interface management, and ongoing procedural tasks. Future evaluation should therefore consider whether interaction remains manageable in cognitive as well as motor terms under realistic operative conditions. In our view, this is a prerequisite for meaningful clinical translation rather than a later-stage usability refinement.

Environmental Resilience as a Prerequisite for Clinical Adoption

Noise

Operating room noise is a clear challenge for voice-based interaction [13]. Across multiple studies, voice control has failed because of either poor command recognition or background noise [3,11]. This is especially important for surgical environments, where suction, drills, monitors, and team communication are routine features of the setting [3].

In our view, this issue illustrates why apparently attractive interface solutions may prove fragile in practice. Voice control may work well in controlled demonstrations, but its value in the operating room depends on resilience to realistic acoustic conditions. Future human factors studies should therefore test voice-based interaction under typical theater noise rather than under quiet laboratory conditions.

Lighting and Glare

Lighting conditions are repeatedly identified as important to usability and technical performance [5,6]. The 2025 AR-HMD review emphasized that displays may suffer from veiling glare and that lighting conditions should be incorporated explicitly into performance evaluation because they affect both image visibility and tracking reliability [6]. It further described how strong operating room lighting can compromise visibility and the accuracy of holographic matching tasks [6].

These are not merely optical inconveniences. They directly affect whether information can be seen, interpreted, and trusted at the point of care. In our view, future studies should regard lighting variability as a standard part of human factors evaluation, particularly for systems intended to present critical spatial guidance during surgery.

Occlusion and Line-of-Sight Disruption

Occlusion is one of the most consistently cited environmental and technical barriers [5]. In optical see-through systems, occlusion affects both visual presentation and tracking robustness, particularly when marker-based solutions rely on an uninterrupted line of sight [5]. The operating room's crowded and dynamic nature means that staff movement, instruments, smoke, or blood may disrupt tracking or reduce confidence in overlay stability [5].

This is highly relevant for clinical translation because such disruptions are routine rather than exceptional. In our view, systems intended for real intraoperative use must be tested against these ordinary environmental disturbances. Otherwise, there is a risk of overestimating performance in idealized settings and underestimating practical failure modes in the theater.

Real-World Operating Room Complexity

Recent evaluation work has criticized idealized testing environments and stressed the need for ecological validity [4,6]. The 2025 review noted that many current experiments fail to replicate physiological movement, environmental lighting changes, marker interference, and operating room personnel movement [4]. Similarly, the optical see-through review argued that context adaptation remains insufficiently addressed in many systems [4,13].

In our view, this is one of the most important methodological gaps in the field. If mixed reality guidance is to be translated safely into surgery, then testing must move closer to the realities of clinical practice. Human factors evaluation should therefore include realistic environmental complexity as a design condition rather than treat it as a downstream implementation issue.

Implications

Environmental resilience is central to surgical usability [3,13]. A mixed reality system that performs well in static laboratory conditions may still fail clinically if lighting, noise, or crowding degrade display visibility, tracking, or input performance [6]. In practice, these environmental stressors rarely occur in isolation. Noise, glare, movement, crowding, occlusion, and changing lighting conditions may co-occur in the operating room, and their combined effects may shape display visibility, tracking reliability, communication, and user performance more substantially than any single factor considered alone. In our view, future evaluation should treat environmental resilience as a systems property rather than as a set of separate technical stressors.

XSL•FO
RenderX

Implications for Research and Clinical Translation

The central argument of this Viewpoint is that the safe and effective adoption of mixed reality in surgery depends on far more than registration accuracy or visual overlay performance alone. Clinical translation will be shaped by whether these systems can accommodate physical user realities, support reliable low-burden interaction, and remain functional within the environmental complexity of the operating room. These issues also have important implications for patient safety. For systems intended to support intraoperative decision-making, failures in perception, calibration, interaction, communication, or environmental robustness may affect how information is interpreted, whether uncertainty is recognized in time, and how safely clinicians act on device-derived guidance [5,6]. In practice, such failures may contribute to workflow disruption, procedural error, delayed fallback to conventional methods, or misplaced confidence in inaccurate or unstable guidance. Human factors evaluation should therefore be linked not only to usability and adoption but also to safe clinical decision-making and avoidance of patient harm [6,14].

These barriers are not confined to a single domain. Physical tolerability, dexterity and motor demands, and environmental resilience interact in ways that shape overall usability and trust in the system. A device may offer promising visual guidance yet still prove impractical if headset weight causes fatigue, if interaction methods interrupt procedural flow, or if performance degrades under theater lighting, movement, or noise. This shifts attention away from isolated technical optimization and toward the conditions required for reliable use in practice.

This perspective has several immediate implications for future human factors research. Studies of physical characteristics should move beyond generic usability reporting and include explicit consideration of comfort, visual demands, headset fit, fatigue, and compatibility with eyewear, masks, caps, and routine theater practices [3,11]. Studies of dexterity and motor skills should assess not simply whether interaction methods function, but whether they remain efficient, intuitive, and safe during real workflow and under sterile conditions [5,13]. Studies of environmental factors should be conducted in realistic or high-fidelity operative settings, including bright lighting, staff movement, background noise, crowding, and partial occlusion [6].

A systems human factors perspective also draws attention to several underexplored questions [8,9]. These include how system failure or uncertainty is recognized during use, what fallback processes should be triggered when mixed reality guidance becomes unreliable, who assumes responsibility when the system cannot be used as intended, and how shared understanding is maintained when only 1 operator has direct access to the display. This last issue is especially important in the inherently team-based operating room, where asymmetric access to holographic content may affect communication, shared situational awareness, supervision, coordination, and collective decision-making. Where device-derived guidance is visible only to the primary user, important cues may need to be verbally translated or otherwise mediated, increasing coordination burden and altering information flow.

These questions have practical implications for evaluation. Beyond individual usability, future studies should examine how mixed reality systems affect workflow disruption, team communication, shared situational awareness, workload, and failure recognition and recovery under realistic operative conditions. The choice of methods should depend on the procedural context and stage of development, but evaluation should move beyond generic usability reporting alone. Relevant approaches may include workload measures such as the NASA Task Load Index [15], usability instruments such as the System Usability Scale [16], situational awareness assessments, team communication measures, and structured assessment of error recognition and recovery. These practical priorities for evaluation are summarized in Table 1.

Governance and regulatory readiness also require greater attention [8,9,14]. For systems intended to support intraoperative judgment, accountability, clinical responsibility, supervision, escalation, and medicolegal responsibility are not peripheral implementation issues but core conditions of safe adoption. This is particularly important where device-derived guidance may be uncertain, unavailable, or in tension with operative judgment. In such cases, clear arrangements are needed regarding who retains decision-making responsibility, how uncertainty is recognized and communicated, and what governance structures support safe implementation and oversight during early clinical use [10,14].

In our view, human factors are still too often treated as secondary outcomes rather than used to shape system design and translational readiness from the outset [3,5,17]. Published work has also tended to focus more heavily on user-device interaction than on the wider team, organizational, training, and governance issues that are likely to determine real-world implementation. This concern is consistent with broader perspectives in medical AI, which emphasize that emerging AI-enabled technologies should be evaluated with the same rigor as other clinical interventions and implemented through transparent, patient-centered, multidisciplinary frameworks [14]. For mixed reality surgical guidance systems, particularly those intended for first-in-human use, this implies that user-centered design, context-of-use definition, and use-related risk assessment should be treated as core development requirements.

The limiting factor in clinical translation is no longer only whether mixed reality can function in principle but whether it can be integrated safely and practically into real surgical work. This implies that future studies should evaluate mixed reality not only as a display technology but also as a sociotechnical intervention shaped by surgical workflow, environmental complexity, and organizational conditions and guided by the human factors priorities set out in this Viewpoint. A systems perspective further suggests that future research should not be limited to refining user-device interaction alone. Failure recognition, fallback and takeover procedures, responsibility, team communication, shared situational awareness, training, competence, supervision, and governance are all part of the

XSL•FO
RenderX

broader work system in which mixed reality guidance would be used. At present, these higher-level issues appear less well represented in the literature than ergonomic, perceptual, and interface-related concerns. Addressing this imbalance will be essential if mixed reality surgical guidance is to progress from proof-of-concept feasibility to safe and sustainable clinical implementation. Without this shift, mixed reality in surgery may remain technically impressive but operationally fragile, limiting progress toward safe, usable, and clinically sustainable adoption.

Table 1. Priority questions for future human factors research in mixed reality surgical guidance.

Focal domain and priority question	Why it matters
Physical characteristics	
Can the system be worn comfortably and safely for the duration of realistic operative use?	Headset weight, fit, ergonomic burden, and cumulative fatigue may affect tolerance, posture, concentration, and willingness to use the system in practice.
Does the display provide sufficient field of view and perceptual continuity for the intended clinical task?	Restricted field of view and visual limitations may reduce interpretability, disrupt attention management, and constrain practical utility during procedures.
Does the system remain interpretable across realistic user variation, including differences in vision, anatomy, and theater attire?	Mixed reality systems must be usable across diverse clinicians rather than an assumed "average" user.
Dexterity and motor demands	
Can interaction be performed efficiently and reliably without disrupting sterile workflow or procedural control?	Gesture- or voice-based interaction may increase control burden, interrupt workflow, or prove unreliable in time-pressured clinical settings.
Do latency, tracking, or input-mapping limitations impair fine motor performance or operator confidence?	Even small delays or inaccuracies may undermine trust, precision, and usability in high-stakes surgical tasks.
Does use of the system support or interfere with perceptual-motor coordination during clinically relevant tasks?	Systems that appear technically functional may still impose motor or attentional burdens that reduce practical usability.
Environmental resilience	
Does the system remain usable under realistic operating-room lighting, glare, noise, and crowding conditions?	Performance under controlled conditions may overestimate usability and robustness in real operative environments.
How sensitive is the system to occlusion, movement, line-of-sight disruption, or marker interference during procedures?	Environmental disturbance may affect tracking reliability, overlay stability, and confidence in guidance.
Has the system been evaluated under combined environmental stressors rather than isolated idealized conditions?	In practice, lighting, noise, movement, occlusion, and workflow pressures co-occur and may have interacting effects on usability and reliability.

This Viewpoint has limitations. It does not address all human factors domains in equal depth, and some of the safety implications discussed are inferred from human factors principles rather than established through direct comparative clinical evidence. Even so, these issues remain important for technologies intended to support intraoperative decision-making and merit more explicit evaluation in future work.

Conclusions

Physical characteristics, dexterity and motor demands, and environmental conditions are central to whether mixed reality systems can be used effectively in surgery. Although mixed reality shows promise in improving spatial awareness, information access, and attention to the operative field, these potential benefits remain constrained by ergonomic limitations, perceptual challenges, interaction burden, and limited robustness in real operating room settings.

The next phase of development should therefore place human factors evaluation alongside technical validation as a core requirement for successful clinical translation. For mixed reality surgical guidance systems, particularly those intended for patient-specific intraoperative use, future studies should explicitly address physical accommodation of diverse users, low-burden and sterile interaction, reliable performance under realistic environmental conditions, and the broader system conditions required for safe communication, training, supervision, and implementation. Human factors research must therefore move beyond technical accuracy alone and help define how these systems can become safe, usable, and clinically adoptable in practice.

Acknowledgments

Infrastructure support for this research was provided by the NIHR London HealthTech Research Centre in In-Vitro Diagnostics.

Funding

This research received no external funding.

XSL•FO
RenderX

Generative AI Use

AI-assisted tools (eg, ChatGPT version 5.4) were used to support the design and visualization of Figure 1 [8]. All scientific content and interpretation were developed by the authors.

Authors' Contributions

SB contributed to conceptualization, investigation, writing—original draft, writing—review and editing, and visualization. PB contributed to conceptualization, supervision, and writing—review and editing.

Conflicts of Interest

None declared.

References

1. Ke Y, Hu K. Advancing surgical practice with mixed reality: current innovations and future prospects. MedComm Future Med 2025 Dec 23;4(4):e70043. [doi: 10.1002/mef2.70043]
2. Asadi Z, Asadi M, Kazemipour N, Léger É, Kersten-Oertel M. A decade of progress: bringing mixed reality image-guided surgery systems in the operating room. Comput Assist Surg (Abingdon) 2024 Dec;29(1):2355897 [FREE Full text] [doi: 10.1080/24699322.2024.2355897] [Medline: 38794834]
3. Magalhães R, Oliveira A, Terroso D, Vilaça A, Veloso R, Marques A, et al. Mixed reality in the operating room: a systematic review. J Med Syst 2024 Aug 15;48(1):76. [doi: 10.1007/s10916-024-02095-7] [Medline: 39145896]
4. Vervoorn MT, Wulfse M, Van Doormaal TP, Ruurda JP, Van der Kaaij NP, De Heer LM. Mixed reality in modern surgical and interventional practice: narrative review of the literature. JMIR Serious Games 2023 Jan 06;11:e41297 [FREE Full text] [doi: 10.2196/41297] [Medline: 36607711]
5. Doughty M, Ghugre NR, Wright GA. Augmenting performance: a systematic review of optical see-through head-mounted displays in surgery. J Imaging 2022 Jul 20;8(7):203 [FREE Full text] [doi: 10.3390/jimaging8070203] [Medline: 35877647]
6. Zhang P, Wang Z, Wang T, Liu T, Wang J, Gao Y, et al. Evaluating augmented reality head-mounted devices in healthcare: a review of hardware, software, and usability approaches. Med Devices (Auckl) 2025 Aug 22;18:427-445 [FREE Full text] [doi: 10.2147/MDER.S541187] [Medline: 40873666]
7. Lewis CH, Griffin MJ. Human factors consideration in clinical applications of virtual reality. In: Virtual Reality in Neuro-Psycho-Physiology. Amsterdam, The Netherlands: IOS Press; 1997:35-56.
8. Moray N. Culture, politics and ergonomics. Ergonomics 2000 Jul;43(7):858-868. [doi: 10.1080/001401300409062] [Medline: 10929822]
9. Paterson E, Chari S, McCormack L, Sanderson P. Application of a human factors systems approach to healthcare control centres for managing patient flow: a scoping review. J Med Syst 2024 Jun 18;48(1):62 [FREE Full text] [doi: 10.1007/s10916-024-02071-1] [Medline: 38888610]
10. Pickup L, Lang A, Atkinson S, Sharples S. The dichotomy of the application of a systems approach in UK healthcare the challenges and priorities for implementation. Ergonomics 2018 Jan;61(1):15-25. [doi: 10.1080/00140139.2017.1306632] [Medline: 28306384]
11. Brockmeyer P, Wiechens B, Schliephake H. The role of augmented reality in the advancement of minimally invasive surgery procedures: a scoping review. Bioengineering (Basel) 2023 Apr 21;10(4):501 [FREE Full text] [doi: 10.3390/bioengineering10040501] [Medline: 37106688]
12. Dennler C, Bauer DE, Scheibler AG, Spirig J, Götschi T, Fürnstahl P, et al. Augmented reality in the operating room: a clinical feasibility study. BMC Musculoskelet Disord 2021 May 18;22(1):451 [FREE Full text] [doi: 10.1186/s12891-021-04339-w] [Medline: 34006234]
13. Nowak R, Nowak M, Rękas M, Ali MJ. Intraoperative use of mixed reality (MR) in humans across surgical disciplines: a major review. 3D Print Med 2025 Dec 26;11(1):59. [doi: 10.1186/s41205-025-00305-7] [Medline: 41452379]
14. Mohsin Khan M, Shah N, Shaikh N, Thabet A, Alrabayah T, Belkhair S. Towards secure and trusted AI in healthcare: a systematic review of emerging innovations and ethical challenges. Int J Med Inform 2025 Mar;195:105780 [FREE Full text] [doi: 10.1016/j.ijmedinf.2024.105780] [Medline: 39753062]
15. Hart SG, Staveland LW. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv Psychol 1988;52:139-183. [doi: 10.1016/S0166-4115(08)62386-9]
16. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
17. Ferguson A, Newham R, Dunlop E, Preston K, Bennie M. Applications of human factors approaches and methods to the development of teleconsultations in primary care: a systematic scoping review. BMJ Open Qual 2025 Jul 05;14(3):e003023 [FREE Full text] [doi: 10.1136/bmjoq-2024-003023] [Medline: 40617592]

XSL•FO
RenderX

Abbreviations

AR-HMD: augmented reality head-mounted device

Edited by S Law; submitted 21.May.2026; peer-reviewed by F Ferraro, A Yoag, AK Inkulu; comments to author 05.Aug.2026; revised version received 22.Oct.2026; accepted 22.Oct.2026; published 09.Nov.2026.

Please cite as:
Bahadori S, Buckle P
Prioritizing Human Factors in the Clinical Translation of Mixed Reality Surgical Guidance
JMIR Hum Factors 2026;13:e101995
URL: https://humanfactors.jmir.org/2026/1/e101995
doi:10.2196/101995
PMID:

©Shayan Bahadori, Peter Buckle. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 09.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Developing mHealth IT for Older Adult Medication Safety: Remote Participatory Co-Design Using the RAPID Method

Jordan R Hill[1], PhD; Aaron Ganci[2], PhD; Noll L Campbell[3,4], PharmD; Andrew C Pickett[1], PhD; Michelle A Chui[5], PharmD, PhD; Ephrem Abebe[3], PhD; Richard J Holden[1,4], PhD

[1]Department of Health & Wellness Design, School of Public Health - Bloomington, Indiana University, 1025 E 7th St, Bloomington, IN, United States
[2]Department of Visual Communication Design, Herron School of Art and Design, Indiana University, Indianapolis, IN, United States
[3]Department of Pharmacy Practice, College of Pharmacy, Purdue University, West Lafayette, IN, United States
[4]Indiana University Center for Aging Research, Regenstrief Institute, Indianapolis, IN, United States
[5]School of Nursing, University of Wisconsin - Madison, Madison, WI, United States

Corresponding Author:
Jordan R Hill, PhD
Department of Health & Wellness Design, School of Public Health - Bloomington, Indiana University, 1025 E 7th St, Bloomington, IN, United States

Abstract

Background: Participatory co-design is a design approach that involves end users in intervention design and its use in health care applications has become widespread. Traditionally, co-design has been conducted in person in a laboratory-based setting; however, it has recently shifted to being performed remotely. Remote co-design has the potential to overcome some of the limitations of traditional in-person approaches, including expanding a study's geographic reach, recruiting participants from underrepresented groups, reducing power imbalances between researchers and participants, and enhancing engagement through online tools. Given these benefits, further reporting and refinement of remote co-design methods are needed.

Objective: This paper's objective is to present our Remote and Accessible Participatory Intervention Design (RAPID) method and discuss the choices and challenges we encountered adapting participatory co-design for remote use.

Methods: We adapted our previously developed 5-step in-person participatory co-design method for health intervention design. To apply the adapted co-design method, we recruited 2 groups of 5 participants (one of older adult pharmacy patients and the other of pharmacy staff) to design a digital kiosk for use by older adults to promote safe over-the-counter medication purchases in retail pharmacies.

Results: Adaptations made to the co-design process were classified under the following categories: facilitation; collaboration, communication, and sensemaking; accessibility and universality; tangible tools and games; and research compliance. Anecdotally, the remote co-design process took longer when compared to in person due to shorter sessions and between-session refinement, but it allowed for flexible scheduling and makeup sessions when required.

Conclusions: Our RAPID method offers an approach to remote co-design that other teams can implement or adapt to their needs. Our experiences with RAPID identify certain drawbacks to remote co-design; however, these are balanced by advantages in convenience and flexibility.

(JMIR Hum Factors 2026;13:e82366) doi:10.2196/82366

KEYWORDS

co-design; participatory design; remote; health intervention; older adults; digital health; pharmacy

Introduction

Centering patients in health intervention research has become a priority to ensure that interventions align with the patients' needs, values, and priorities [1-6]. Patient-centered interventions are more likely to be adopted and implemented, maximizing their impact [6,7], and there is evidence that they result in improved health outcomes [8]. Participatory co-design methods are a formal way to involve intervention end users in product or service design to directly center user needs (eg, usability, acceptability, and usefulness) [9] including those of patients to design patient-centered interventions [1]. With the rise of the "digital health citizen" [10-12]— empowered patients who are active consumers and generators of their own health data—the demand for patient-centered tools and services will only increase. Co-design with patients and informal caregivers has been increasingly adopted in healthcare systems design to address these demands [13].

Traditional participatory co-design is performed face-to-face, requiring end users to physically travel to a secondary location

XSL•FO
RenderX

to attend activities such as design workshops. Remote approaches to co-design had been considered before but became necessary during the COVID-19 pandemic [14]. Since then, remote co-design has been performed with various populations of interest including children [14-16], racially underrepresented young adults [17], and caregivers of patients with dementia [18,19]. The applications have been similarly heterogeneous from supporting antibullying [16] to helping family caregivers with legal and financial planning [18,19].

While there are challenges to implementing participatory co-design (eg, differences in familiarity with technology and difficulty aligning goals and language of various stakeholders) [20], remote co-design may have advantages over traditional face-to-face approaches. Pioneers of remote co-design methods report recruiting participants from wider geographic areas [15,21] and extending their study's reach to populations sometimes excluded from research [15,17]. The remote format can also facilitate certain user interactions. Bertran et al [21] described how study facilitators and designers were able to disappear "behind the scenes" to attain a more naturalistic perspective of participants interacting with their prototype. Similarly, Fails et al [14] reported that the lack of researchers' physical presence balanced some of the child-adult or facilitator-participant power dynamics. Additionally, some online features such as chats during video calls enabled more participants to engage, respond, and discuss [17]. Some remote co-design methods (eg, surveys and asynchronous feedback) also enable the collection of large amounts of data when compared to small studies conducted in person [22]; this is especially important when considering the big data implications of widespread digital health use and data generation. With the benefits becoming increasingly apparent, it is imperative to report on and improve upon the use of remote participatory co-design methods.

This paper's objective is to present our Remote and Accessible Participatory Intervention Design (RAPID) method. There is a clear need to develop RAPID because, while there have been some methods-focused or instructional papers on remote co-design [14,23,24], we have not found any that specifically focus on engaging older adults in remote co-design. RAPID addresses this gap as it was adapted specifically for older adults and has been used successfully with older adults in other studies [18,19]. There are challenges in engaging older adults in digital research: the digital divide persists with fewer older adults using the internet [25], having access to broadband internet at home [26], and owning a smartphone [27] when compared to other demographics; and older adults are stereotyped as having less technological ability than younger adults [28]. However, older adult technology adoption is growing [27,29]. Nearly 1 in 5 Americans is over the age of 65 years (18%) [30], and this age group experiences more chronic diseases. It is important that digital health interventions are designed centering older adults as they are likely to be widespread users in the future.

In this paper, we discuss the choices and challenges we encountered in adapting participatory co-design for remote use. Building on existing work, we organize our results using the framework outlined in Mallakin et al [23]. We hope this method will be useful for others designing patient-centered interventions for older adults and potentially other populations who may experience challenges with co-design, and that it will encourage further involvement of older adults in digital health intervention design.

Methods

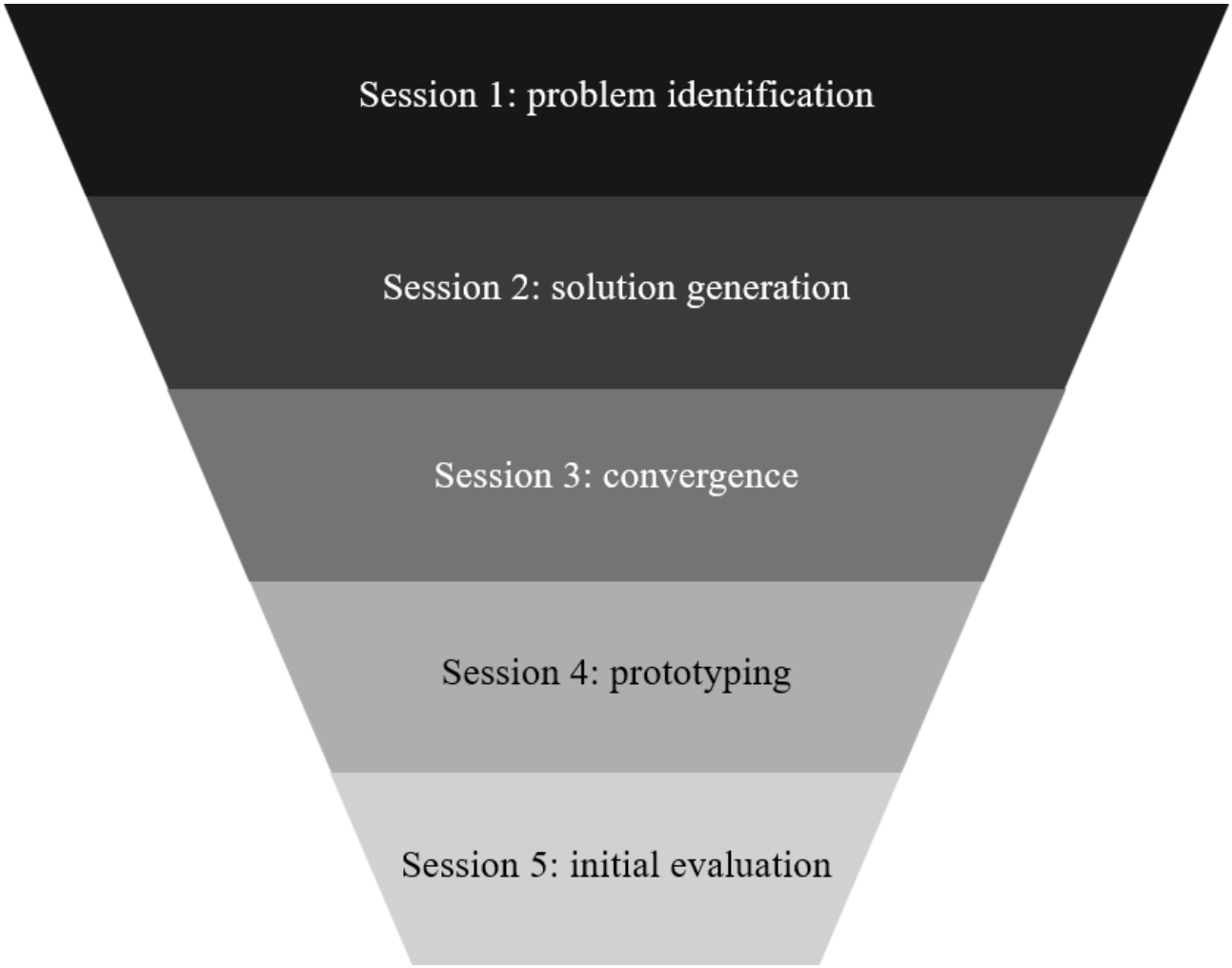
Overview

The RAPID method was originally developed for and applied in a specific study [31]. However, it was also envisioned as a formative method that could be adopted for future studies, further observed in practice, refined, and adapted over time. The initial study, called Senior Station, had the goals of co-designing and testing a technological intervention (a kiosk) to support older adults in making safer over-the-counter (OTC) medication decisions by providing them with information on which medications had ingredients that may not be safe for their brain health. More detail on the intervention designed can be found in a separate publication [31].

To develop RAPID, we adapted our previously developed 5-step in-person participatory co-design method (Figure 1) for remote delivery. This funnel framework was developed for health intervention design in an older adult OTC medication decision-making study [32] and subsequently applied in a broad array of projects and settings [18,19,31,33]. Details on each RAPID session and how we adapted them for remote settings are described in the results section.

XSL•FO
RenderX

Figure 1. Remote and Accessible Participatory Intervention Design (RAPID) co-design stages.



To apply RAPID, we recruited 2 groups of 5 participants each (N=10): one of older adult pharmacy patients and the other of pharmacy staff. All participants were recruited from a large pharmacy chain retailer in the United States.

Older adults, aged 60 years or older, were those who had previously purchased or planned to purchase OTC medication, could communicate in English, and were able to join a video call from a computer, tablet, or smartphone. Potential participants were approached in person by a member of the research team at the pharmacy retailer and were screened for eligibility if they expressed interest in participating. The research team member reviewed a paper study information sheet and obtained verbal consent. Participants provided their contact information for future coordination to attend the remote design sessions.

Pharmacy staff were pharmacists and technicians who had previous experience counseling older adults on the purchase of OTC medications, could communicate in English, and were willing and able to join a video call. They were recruited via the research team's professional contacts at the same retailer from which older adults were recruited and filled out an online survey with their contact information if they were interested in participating. A member of the research team then contacted pharmacy staff, reviewed the consent form, obtained verbal consent for participation, and captured demographic information via a Qualtrics survey.

Each session was facilitated by 3 to 4 members of the research team. An expert in participatory co-design facilitation acted as "master of ceremonies." The "master of ceremonies" was responsible for leading the agenda, guiding the participants through co-design activities, and prompting participants to assess each other's design ideas. In later sessions, the same individual guided participants to review solution prototypes. The second role in each session was an administrator or "problem solver." They ensured participants had the information and resources they needed to participate in the study. This role was critical to help some participants troubleshoot technical issues during the session. The third role was a human-computer interaction expert who observed participants' design ideas through the lens of implementation and usability. Their role during the discussion was mainly asking clarifying questions. Finally, a scribe participated in the sessions to capture participant ideas in real time. In our study, this person was a design student who was also able to sketch prototypes in real time.

The 5 steps of the RAPID method were adapted for remote delivery as shown in Table 1.

The observations in the following sections are based on team self-reflections and a review of notes, correspondence, and other

XSL•FO
RenderX

documents over the course of the project. There was no systematic method in place for the analysis of the results of the RAPID sessions.

Table . Remote and Accessible Participatory Intervention Design (RAPID) procedure for remote co-design.

Step	Procedure	Remote approach
Problem identification	• Introduce the research team and have participants introduce themselves. • Have participants describe their experiences with the challenge or situation of interest (eg, tell a story) • Visually represent these stories using journey maps to identify pain points and targets for intervention • Explicitly introduce the concept or challenge the design seeks to address • Present different choices or scenarios to participants to determine what factors are most important (eg, do cost savings outweigh warnings of risks) and which design features are more desirable	• Videoconferencing (Zoom; Zoom Communications Inc) • Microsoft PowerPoint presentation on shared screen • Turn-taking • Heavy facilitation and experienced facilitator
Solution generation	• Present a problem scenario to participants with an undescribed solution to help them • Guide participants through brainstorming the design of different aspects of the solution, such as size, information communicated, advertisements, features, and symbols. • Have participants roughly sketch their designs on whiteboards and present them to the group. This enables participants to express themselves creatively, provide diverse perspectives, and turn abstract ideas into something tangible. • Take photos or some other record of the drawings	• Videoconferencing (Zoom) • PowerPoint presentation on shared screen • Turn-taking • Heavy facilitation and experienced facilitator • Whiteboards mailed to participants before the session • Photos of drawn designs taken over Zoom or by participants and emailed to the research team
Convergence	• Present low-fidelity, hand-drawn designs to participants based on ideas in previous session. Low-fidelity prototypes make it clearer to participants that design ideas are preliminary and not “set in stone.” • Walk through the sequence of use of design, seeking feedback on each design and making it clear that designs were flexible. • Ask participants to provide more details (eg, specific information and what is prioritized) on designs. • The goal is to converge the multiple designs into one with which to proceed.	• Videoconferencing (Zoom) • PowerPoint presentation on shared screen • Turn-taking • Heavy facilitation and experienced facilitator • Hand-drawn designs on a tablet or computer and shared on screen during session
Prototyping	• Present a higher-fidelity prototype (eg, a clickable wireframe) that is the result of feedback from the previous session. The higher-fidelity prototype made the design feel more real and indicative of the final intervention. • Solicit feedback from participants	• Videoconferencing (Zoom) • PowerPoint presentation on shared screen • Turn taking • Heavy facilitation and experienced facilitator • Prototype in Figma (Figma, Inc.) and presented on screen during session
Initial evaluation	• Present the final design based on the design sessions from both participant groups • Solicit feedback from participants	• Videoconferencing (Zoom) • PowerPoint presentation on shared screen • Turn taking • Heavy facilitation and experienced facilitator • Prototype in Figma and presented on screen during session

XSL•FO
RenderX

Ethical Considerations

This study was approved by the Indiana University Human Research Protection Program (institutional review board number 13200) and co-design sessions occurred from January 2022 to November 2022. Participants were compensated US $50 in gift cards for each session they attended in addition to bonus payments as incentives to avoid participant attrition: US $100 bonus for attending all of the first 3 sessions or US $250 bonus to attend all 5 sessions (for a potential total of US $500 in compensation). A research team member reviewed a paper study information sheet with participants and obtained verbal consent. All participants were able to opt out of the study and terminate their participation at any time. Any transcripts and meeting notes generated during sessions were de-identified to protect confidentiality and privacy. Session photos of participant designs or sketches did not contain any identifiable information.

Results

Here we describe our adaptations, organized retrospectively using Mallakin et al's [23] framework for considerations for remote co-design.

Facilitation

Definition

Facilitation is the process by which designers and researchers "set the scene," then guide participants through the design process to stimulate creativity and generate collaborative design solutions.

Remote Considerations

Remote facilitators may not be aware of the physical space or equipment limitations or assets of the remote co-designers. They cannot directly set up, demonstrate, or manipulate the co-designers' physical design space or equipment and are more constrained in their strategies to establish trust, open dialogue, or rapport such as one-on-one time, informal side conversations, or physical gestures. Remote sessions are typically shorter than in person, which are often day-long or half-day events.

Successful co-design activities rely on open conversation where participants generate prototypes, pitch ideas, provide feedback, and then build on one another's concepts. When online, this process is impeded because the ability to present ideas is limited and the nature of discussions is restricted. For example, it is hard to manage crosstalk during online discussions, which results in participants being more hesitant to contribute out of fear of interrupting another person.

While in-person sessions feel more like a collaborative group conversation, the online session requires significantly more facilitation. For example, in an in-person setting, it would be easy to look at 2 participant prototypes at the same time to compare them. That is much harder online and would require the facilitator to collect images of the prototypes before talking about them.

Our Approach

The co-design facilitator had extensive experience with in-person facilitation and anticipated what interactions could be replicated remotely versus required accommodation. The entire research team, including the facilitator, met extensively before the start of co-design activities to fully conceptualize the design problem and develop cases, scenarios, and prompts to guide participants through the design process. More time was allotted in the first session for rapport building with the research team and among the participants, using turn-taking to ensure everyone had a chance to speak. The facilitator was intentional to lead discussion and call on participants, as a countermeasure to the awkwardness of interrupting a speaker in an online format (due to audio delays, the abruptness of teleconference interruptions, and uncertainty about the source of interruption). Prototyping tools conducive to remote co-design were selected (eg, Adobe XD, Figma, and individual whiteboards). To facilitate co-designers' critiquing of each other, the facilitator guided participants to offer initial feedback then build on each other's initial ideas rather than directly grading designs.

Lessons Learned

The facilitator is important in remote settings because they must be skilled in co-design methods and adjust those methods to suit the setting. Facilitators cannot assume methods that work in person will be as effective online. They should also have strong technology skills for more seamless remote co-design meetings.

Remote settings can feel socially distant, impersonal, or more businesslike compared to in person. A facilitator needs to be intentional about creating a creative and open environment where participants feel comfortable working together. This could include focusing on friendliness by adding a joke to the session or asking pointed questions to get participants to talk about something they mentioned in passing.

Online meeting conventions and limitations (eg, one person speaking at a time, no cross-talk, and limited body language cues) can hamper study interactions. It often falls on the facilitator to smooth the conversation and facilitate group communication.

Facilitator expectations regarding how many design activities can be accomplished in each session need to be adjusted when performing co-design online. There are 2 factors of online co-design that impact a facilitator's time management. First, it is harder to hold participants' attention, thus favoring shorter sessions. Second, technical issues will arise in online sessions. Our team lost time during most sessions waiting for participants to join, troubleshooting technology issues, or clarifying instructions.

Collaboration, Communication, and Sensemaking

Definition

Collaboration, communication, and sensemaking are the processes by which participants work together with the research team to produce co-design.

Remote Considerations

Participants are not co-located in the same physical space (eg, a design studio) and must meet and interact virtually. There is no shared physical whiteboard, flip chart, or other materials to

XSL•FO
RenderX

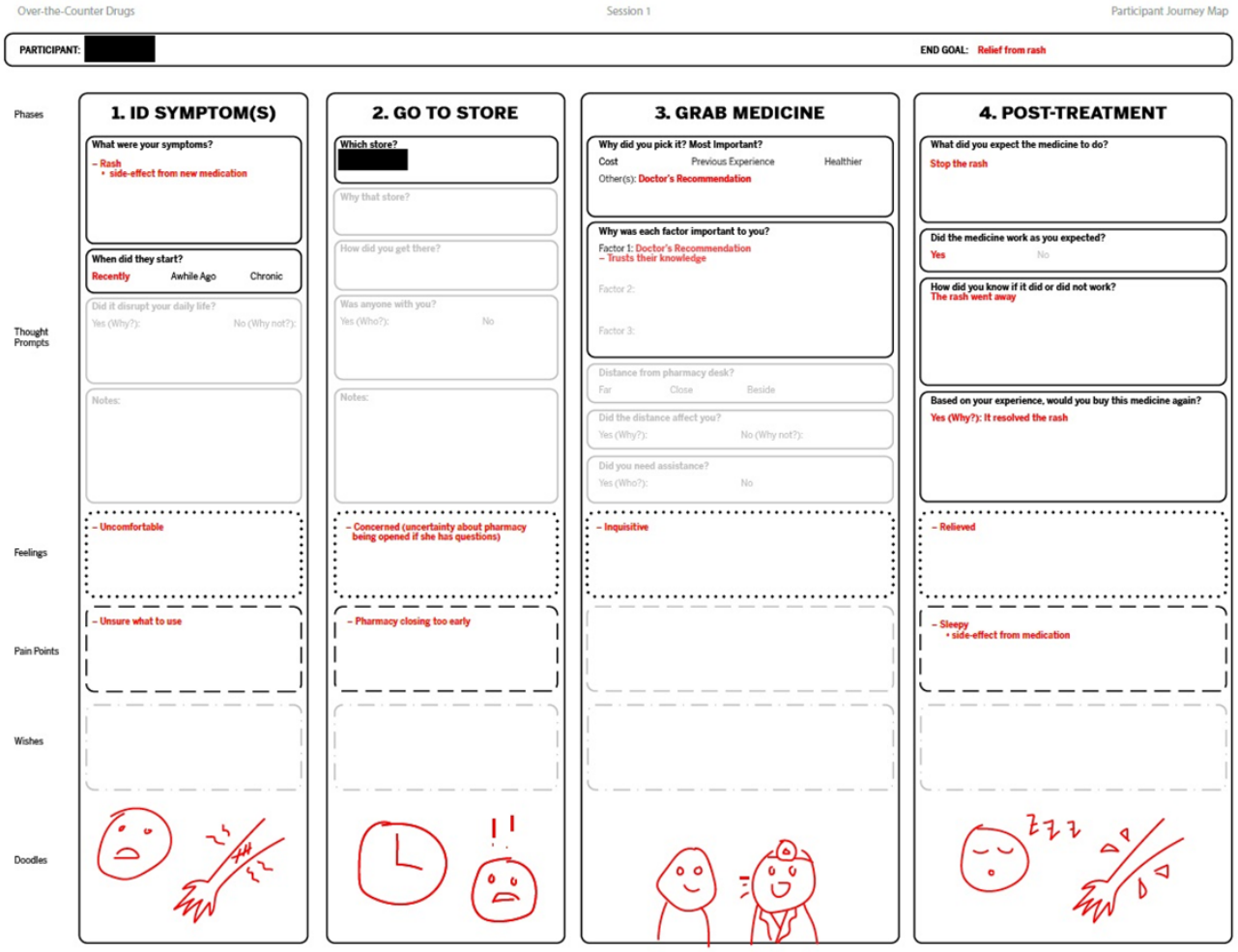
which every participant has real-time access to externalize their thoughts in a manner directly visible to others. Some designs, for example, those that are drawn or constructed on nondigital artifacts (eg, scrap paper) or have 3D elements (eg, gestures), may not be easily shared on a screen or will need to be digitized. Differences in familiarity and comfort with remote technologies may affect how individuals engage in collaboration, communication, and sensemaking.

Our Approach

To host online sessions, we selected Zoom videoconferencing software (Zoom Communications Inc) as it was freely accessible on a range of devices and operating systems, widely used during the pandemic, supported through the research team's institutions, and supported recording for later review and analysis. A design student attended our virtual sessions to take notes so that the facilitator could focus on supporting interaction.

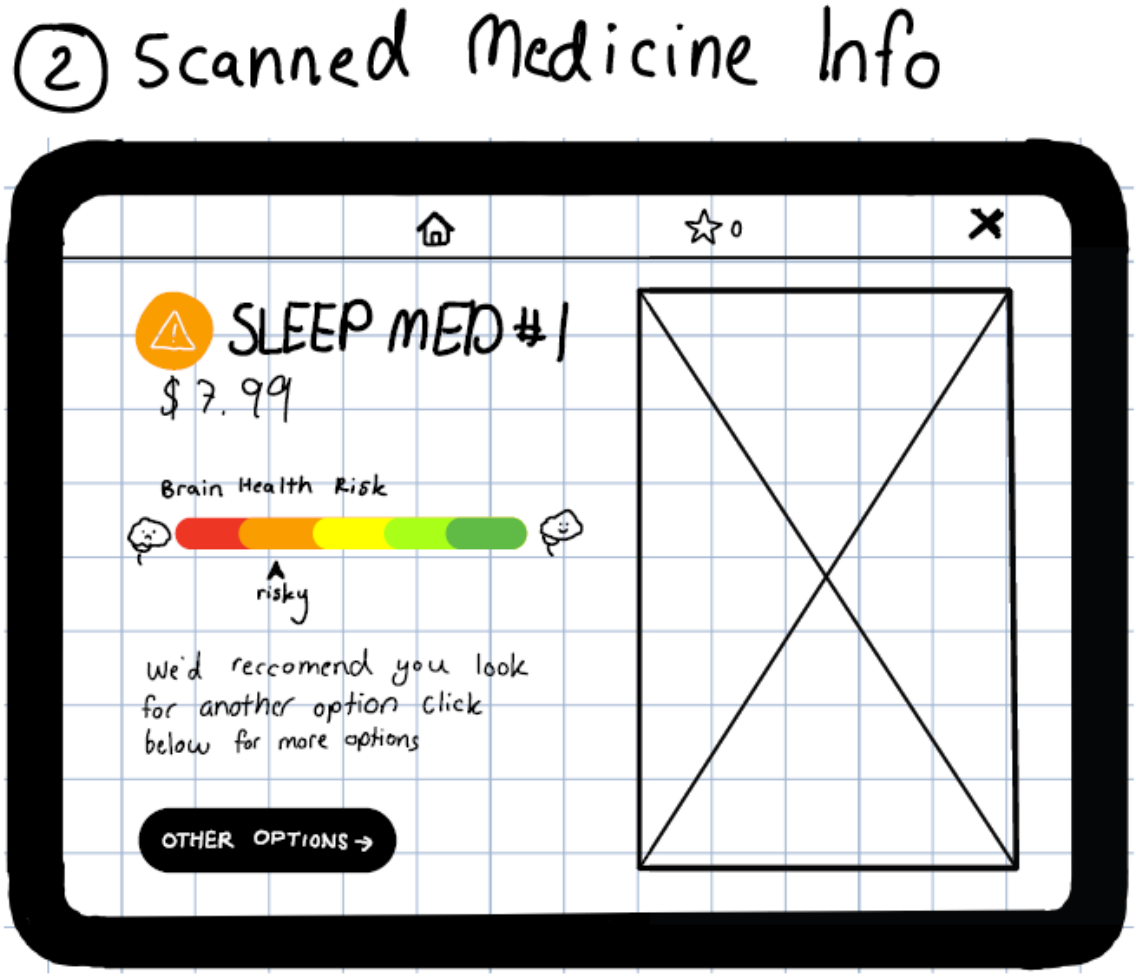
A journey map template was created to help the designer capture notes in a manner that could be easily shared with others (Figure 2). Our designer created simple, low-fidelity wireframes summarizing the designs that could be easily shared online, but despite being digital, looked rudimentary enough to facilitate critiques (Figure 3).

Figure 2. Deidentified journey map notes.



XSL•FO
RenderX

Figure 3. Low-fidelity wireframe.



Lessons Learned

Conversation does not flow as naturally or collectively online as in person, so participants are disinclined to build on each other's ideas, implying they may perceive their roles as developing individual designs rather than a single co-owned design. Participants were frequently distracted in the online environment (phone calls, family members in the home, and so on) and felt more comfortable attending to distractions and multitasking.

The primary physical medium for online co-design, a 2D computer screen, is not as inherently a creative space as a 3D design studio. An atmosphere of creativity is therefore often created through the activities, words, and interactions of co-design rather than the setting. Physical tools and virtual environments could be used to increase creativity.

Participants did nevertheless develop a sense of community and comfort over the course of 5 sessions, despite being in a relatively sterile remote setting.

Accessibility and Universality

Definition

Accessibility and universality refer to the degree to which a wide and inclusive range of individuals can and do participate in co-design.

Remote Considerations

Comfort with technology and access to a stable internet connection are important considerations for remote co-design, especially with certain groups such as older adults. Unexpected problems (eg, technical difficulties) are more likely to occur in remote settings, and not every participant is equally adept at using online interaction features such as typing private or public messages, digital gestures (eg, raised hand) and reactions, electronic whiteboard, screen sharing, audio sharing, muting or unmuting, and hyperlinking. Communication can be hampered or interrupted by having to concurrently deal with technology issues such as typing, audio, video, switching windows, computer slowness, and internet connectivity. Not all participants joined from a desktop or laptop computer. Those who joined using cellphones were hampered by having smaller screens.

XSL•FO
RenderX

Our Approach

One member of the research team spent significant time outside of design sessions working with the participants, including the older adults, to ensure they could join a videoconference and practicing with them to increase comfort. In one case, this required live phone assistance prior to every session. We designated a "problem solver" to provide support for emergent accessibility needs such as troubleshooting technical difficulties and phoning participants who did not join the session on time. To compensate for participants joining from smaller screens (eg, cellphone), the facilitator spent time during the session explaining what was being shown on screen.

Lessons Learned

Including participants who were less comfortable with technology was a benefit, not a downside, of the remote co-design approach, as they provided a novel perspective that was especially important in our attempts to design universally usable technology for older adults. The problem solver was invaluable to the smooth operation of design sessions as they handled unexpected problems without disrupting the session for the facilitator, designer, and participants. The provision of technical support during and outside of sessions requires additional personnel and associated expenses.

Tangible Tools and Games

Definition

Tangible tools and games refer to the tools and games used to encourage participants' creativity and facilitate co-design work.

Remote Considerations

Traditionally, co-design tools and games are hands-on and require physical items. Games are often experienced in a playful, informal, low-risk context, whereas online meetings may be associated with formal, serious, and more sterile functions (eg, work).

Our Approach

Before each session, participants were mailed relevant materials for that session. This included prototyping tools or prototype concepts for their review. We provided personal whiteboards and markers to participants, mailed directly to their homes by an online retailer. Some participants spontaneously did "homework" to create preliminary designs before the session, thus extending creativity beyond the time and space confines of the formal design sessions.

Lessons Learned

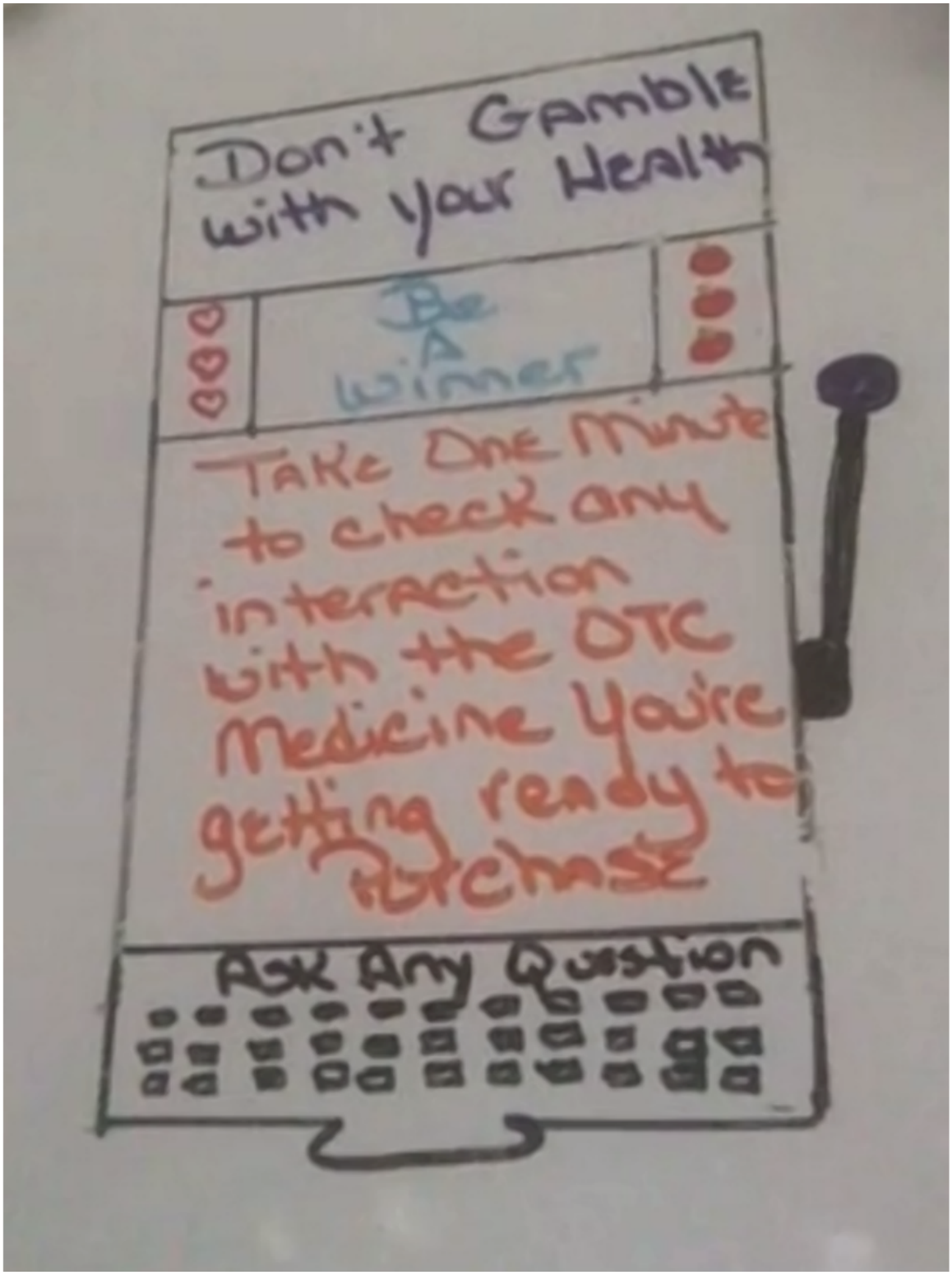
Not all traditional co-design tools and games could be translated to the online setting, and some could not be used. New or adapted tools need to be intentionally and innovatively planned to enable activities such as online sharing of participant-created prototypes. It is more difficult to adjust or introduce new tools or ideas in real-time online, and modifications to set schedules are often not well received or understood in online spaces. As such, remote sessions require additional planning and forethought. The spirit of fun and creativity of in-person games and activities felt constrained in online encounters. Our tools were limited to the software and hardware we chose to use, making it difficult or impossible to introduce ad hoc tools and activities often available in a traditional design studio.

Mailing whiteboards was a low-cost, low-tech option that provided participants the ability to do some design activities without a steep learning curve. Participants were easily able to photograph their whiteboard designs and send them to the research team. Other options were dismissed as too costly, complex to learn, or disruptive to the project timeline.

Participants doing extra work between sessions produced creative designs (Figure 4) and may have increased reflection. It is possible in our project and in others that working between sessions would allow participants to engage others in their household, social network, or community to enrich their designs. In other cases, working independently could allow co-designers the ability to use novel tools such as generative artificial intelligence that could stimulate their creativity and self-efficacy.

XSL•FO
RenderX

Figure 4. Participant design on whiteboard.



Research Compliance

Definition

Research compliance is the degree to which co-design adheres to the applicable ethical principles and requirements

Remote Considerations

When participants are in a public online gathering place, especially one that is being recorded, additional steps are needed to ensure privacy for participants and provide opportunities for participants to privately interact with researchers. While some physical affordances such as private rooms may have online analogs (eg, private breakout rooms), their use may not be as easy.

The physical setting for remote interactions may be less controllable for researchers and co-design participants who may be joining from a shared space such as their home, workplace,

XSL•FO
RenderX

or coffee shop. This impacts both privacy and confidentiality considerations, particularly when using video or audio recording.

Our Approach

We conducted recruitment and informed consenting in person in community pharmacies, rather than during the online sessions, and provided older adult participants with printed study information sheets. Sessions were hosted using institutional videoconferencing accounts (Zoom) for added security. Standard web-based data security approaches applicable to all forms of internet-based research (eg, deidentification of data, and storage of data on secure, approved servers) were used.

Lessons Learned

Participants do join remote co-design sessions from locations such as workplaces, homes, and cars, and those joining from work often turned off their cameras, which made conversations more difficult.

General Lessons

In addition to considerations inspired by the Mallakin et al [23] framework, implementing RAPID produced the following general takeaways.

Drawing on our prior experience with in-person co-design, we observed that the remote co-design process took considerably longer in the remote format. This is likely because RAPID sessions were in shorter 90-minute bursts, and in between sessions, designers spent considerable time refining the designs and preparing materials for the next round. This was advantageous for advancing the designed product but increased the total project duration.

A second general lesson is the need for flexibility and willingness to implement contingencies for the unexpected. For example, we recommend over-recruiting for each group to ensure an adequate number of participants. In both the older adult and pharmacy staff groups, we had fewer than 5 participants participate in all but 1 session. However, a benefit of RAPID was the ease of adding makeup sessions when multiple participants could not attend a prior session.

Third, we performed our remote co-design sequentially with the 2 groups due to recruitment and timing considerations. We believe that it would be beneficial to complete the sessions concurrently to ensure that each group's designs are not incompatible with the other's. For example, the older adult participants wanted a scale to demonstrate a medication's effectiveness versus safety. Upon further investigation and discussion with pharmacy staff, such a scale was infeasible and would not apply to all users. Our research team made some decisions that were not purely end user driven to reconcile differences between the 2 groups' designs.

Participant Feedback

We did not explicitly or systematically collect feedback on the co-design sessions from participants. However, during the last session, participants provided unprompted opinions on their experiences. The feedback was positive and older adult participants expressed their enjoyment in participating:

> *"This was fun! I enjoyed it."* [Female older adult participant, age 79]
>
> *"Thanks for the opportunity to participate."* [Male older adult participant, age 74]
>
> *"Oh my gosh, it was really really fun. I felt so important!"* [Female older adult participant, age 66]

Discussion

Principal Findings

We produced RAPID by adapting our co-design method to be performed remotely and gained valuable learning to guide future remote co-design projects.

The primary lesson learned was that the online setting is intrinsically less creative and more socially distant than traditional, in-person environments. Our experiences of constraints on participant rapport and trust [14,15], engagement [14], and fewer generative side conversations [19] have been reported in other studies. Practically, this emphasizes the importance of having skilled and experienced facilitators (ie, facilitators with strong technology and communication skills, and an in-depth understanding of transforming user needs into design choices), using activities that promote social interaction among participants (eg, ice breakers, jokes, pointed questions, and turn-taking), and anticipating delays and distractions throughout sessions.

In line with other literature, we found that the remote co-design, compared to our past experiences with traditional in-person co-design, had higher participant attrition than expected [16,17,19], required more researcher flexibility [14,15,17], and was slower [14,16]. We attribute our slower process to shorter design sessions, additional between-session activities, and scheduling serial rather than concurrent independent co-design groups (older adult vs pharmacy staff). The restriction on session duration, chosen to manage participant fatigue, may require the most creativity to overcome.

Our experience with participant differences in technology comfort and experience validates recommendations by Jolliff et al [19] and Fails et al [14] to provide technical support outside sessions, during sessions, or both, as we did.

Despite certain disadvantages of remote co-design, there were also benefits that promoted the continued use and refinement of RAPID. The convenience for participants to participate in research from anywhere with a suitable internet connection is the greatest advantage of any remote research methodology and applies to remote co-design. We engaged participants who otherwise may not have participated due to geographical separation, weather, or work or family obligations. This benefit is noted by others [14,15]. We identified an additional financial benefit by avoiding transportation and facility costs.

We found that, without prompting, participants completed extra work between sessions ("homework"). This enhanced design creativity during sessions and comports with the general benefits of presession activities and time for reflection in co-design [15,17,19]. We recommend prospectively incorporating

XSL•FO
RenderX

"homework" into future RAPID applications to help reflect and create without researcher influence or time pressure.

Limitations

Our experiences, lessons learned, and recommendations were limited in several ways, including the fact that the study sample was small, and this was an early-stage, exploratory study. First, we took a pragmatic but unsystematic approach to creating and reflecting on RAPID, disproportionately leveraging our team's expertise and adapting "on the fly." This is to be expected with a method that arose rapidly out of necessity (ie, COVID-19 pandemic restrictions). Indeed, the ability to fail safely and learn quickly is a hallmark of strategies such as Agile [34,35]. However, with time and experience, new approaches such as RAPID and frameworks such as Mallakin et al [23] can be more systematically evaluated and refined. Second, we did not independently measure success of RAPID or compare it to traditional in-person co-design, a necessary future research direction. Explicitly soliciting participant feedback would also be an important future step in assessing the method. Third, despite efforts to include a sufficiently large and heterogeneous set of co-design participants, the first users of RAPID needed to own a computer, tablet, or smartphone with stable internet and sufficient technology comfort to participate in the study. Even if devices with internet access were provided, something we have done in other studies, this may exclude individuals who do not feel comfortable enough with technology.

Conclusions

Remote co-design is increasingly used. Our RAPID method offers an approach to remote co-design that other teams can implement or adapt to their needs. Our experiences with RAPID identify certain drawbacks to remote co-design; however, these are balanced by advantages in convenience and flexibility. Our use of RAPID also offers various strategies to address challenges to remote co-design.

Acknowledgments

The authors would like to thank Rachel Knierman and Nishad Kumbhare for their design support during co-design sessions.

Funding

This project was funded with support from the Indiana Clinical and Translational Sciences Institute, which is funded in part by award UM1TR004402 from the National Institutes of Health, National Center for Advancing Translational Sciences, Clinical and Translational Sciences Award. MAC was also supported by grants from AHRQ (R18 HS024490 and R18HS027737) and from the National Center for Advancing Translational Sciences (UL1TR002373). EA received support through a grant from the National Institute on Minority Health and Health Disparities of the National Institutes of Health (K01MD016167). Sponsor had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors' Contributions

Conceptualization: JRH, NLC, MAC, EA, RJH
Funding acquisition: JRH, MAC, EA, RJH
Investigation: JRH, AG
Methodology: JRH, AG, MAC, EA, RJH
Project administration: JRH, AG
Supervision: AG, EA, RJH
Writing – original draft: JRH, AG
Writing – review & editing: JRH, NLC, ACP, MAC, EA, RJH

Conflicts of Interest

RJH is the founding co-owner of CareNotes LLC. The other authors declare no conflicts.

References

1. Duffy A, Christie GJ, Moreno S. The challenges toward real-world implementation of digital health design approaches: narrative review. JMIR Hum Factors 2022 Sep 9;9(3):e35693. [doi: 10.2196/35693] [Medline: 36083628]
2. Göttgens I, Oertelt-Prigione S. The application of human-centered design approaches in health research and innovation: a narrative review of current practices. JMIR mHealth uHealth 2021 Dec 6;9(12):e28102. [doi: 10.2196/28102] [Medline: 34874893]
3. Ioannidis JPA. Why most clinical research is not useful. PLoS Med 2016 Jun;13(6):e1002049. [doi: 10.1371/journal.pmed.1002049] [Medline: 27328301]
4. Melles M, Albayrak A, Goossens R. Innovating health care: key characteristics of human-centered design. Int J Qual Health Care 2021 Jan 12;33(Supplement_1):37-44. [doi: 10.1093/intqhc/mzaa127] [Medline: 33068104]

XSL•FO
RenderX

5. Mullins CD, Vandigo J, Zheng Z, Wicks P. Patient-centeredness in the design of clinical trials. Value Health 2014 Jun;17(4):471-475. [doi: 10.1016/j.jval.2014.02.012] [Medline: 24969009]
6. Selby JV, Lipstein SH. PCORI at 3 years—progress, lessons, and plans. N Engl J Med 2014 Feb 13;370(7):592-595. [doi: 10.1056/NEJMp1313061] [Medline: 24521104]
7. Alruwaili MM, Shaban M, Elsayed Ramadan OM. Digital health interventions for promoting healthy aging: a systematic review of adoption patterns, efficacy, and user experience. Sustainability 2023;15(23):16503. [doi: 10.3390/su152316503]
8. Yu C, Xian Y, Jing T, et al. More patient-centered care, better healthcare: the association between patient-centered care and healthcare outcomes in inpatients. Front Public Health 2023;11:1148277. [doi: 10.3389/fpubh.2023.1148277] [Medline: 37927879]
9. Sanders L. ON MODELING. An evolving map of design practice and design research. Interactions 2008 Nov;15(6):13-17. [doi: 10.1145/1409040.1409043]
10. Barony Sanchez RH, Bergeron-Drolet LA, Sasseville M, Gagnon MP. Engaging patients and citizens in digital health technology development through the virtual space. Front Med Technol 2022;4:958571. [doi: 10.3389/fmedt.2022.958571] [Medline: 36506474]
11. Petrakaki D, Hilberg E, Waring J. The cultivation of digital health citizenship. Soc Sci Med 2021 Feb;270:113675. [doi: 10.1016/j.socscimed.2021.113675] [Medline: 33434718]
12. Powell J, Newhouse N, Boylan AM, Williams V. Digital health citizens and the future of the NHS. Digit Health 2016;2:2055207616672033. [doi: 10.1177/2055207616672033] [Medline: 29942569]
13. Slattery P, Saeri AK, Bragge P. Research co-design in health: a rapid overview of reviews. Health Res Policy Syst 2020 Feb 11;18(1):17. [doi: 10.1186/s12961-020-0528-9] [Medline: 32046728]
14. Fails JA, Ratakonda DK, Koren N, Elsayed-Ali S, Bonsignore E, Yip J. Pushing boundaries of co-design by going online: lessons learned and reflections from three perspectives. Int J Child Comput Interact 2022 Sep;33:100476. [doi: 10.1016/j.ijcci.2022.100476] [Medline: 36034969]
15. Morris AC, Douch S, Popnikolova T, et al. A framework for remotely enabled co-design with young people: its development and application with neurodiverse children and their caregivers. Front Psychiatry 2024;15:1432620. [doi: 10.3389/fpsyt.2024.1432620] [Medline: 39220185]
16. Sanoubari E, Muñoz Cardona JE, Mahdi H, Young JE, Houston A, Dautenhahn K. Robots, bullies and stories: a remote co-design study with children. 2021 Presented at: IDC '21: Proceedings of the 20th Annual ACM Interaction Design and Children Conference; Jun 24-30, 2021; Athens, Greece p. 592-595. [doi: 10.1145/3459990.3460725]
17. Harrington C, Dillahunt TR. Eliciting tech futures among black young adults: a case study of remote speculative co-design. 2021 Presented at: CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems; May 8-13, 2021; Yokohama, Japan p. 102451. [doi: 10.1145/3411764.3445723]
18. Jolliff A, Dudek A, Zuraw M, et al. Co-design of a financial and legal planning tool for care partners of people living with Alzheimer's disease and related dementias. Innov Aging 2023;7(5):igad046. [doi: 10.1093/geroni/igad046] [Medline: 37360216]
19. Jolliff A, Holden RJ, Valdez R, et al. Investigating the best practices for engagement in remote participatory design: mixed methods analysis of 4 remote studies with family caregivers. J Med Internet Res 2024 Dec 3;26:e60353. [doi: 10.2196/60353] [Medline: 39626228]
20. Duffy A, Boroumandzad N, Sherman AL, Christie G, Riadi I, Moreno S. Examining challenges to co-design digital health interventions with end users: systematic review. J Med Internet Res 2025 Mar 14;27:e50178. [doi: 10.2196/50178] [Medline: 40085834]
21. Bertran FA, Pometko A, Gupta M, Wilcox L, Banerjee R, Isbister K. Designerly tele-experiences: a new approach to remote yet still situated co-design. ACM Trans Comput Hum Interact 2022 Oct 31;29(5):1-30. [doi: 10.1145/3506698]
22. Duffy A, Bennett SE, Yardley L, Moreno S. Crowdsourcing remote co-design towards improving the validity and reliability of mHealth application development – a case study on Sleep Solved: a mHealth app designed virtually with teens. EAI Endorsed Trans Perv Health Tech 2025;11. [doi: 10.4108/eetpht.11.9416]
23. Mallakin M, Dery C, Vaillancourt S, Gupta S, Sellen K. Web-based co-design in health care: considerations for renewed participation. Interact J Med Res 2023 Mar 3;12:e36765. [doi: 10.2196/36765] [Medline: 36595738]
24. Sanders EBN, Shen Y. Direct vs. remote data collection methods in co-design: navigating the evolving landscape. CoDesign 2025:1-15. [doi: 10.1080/15710882.2025.2472172]
25. Perrin A, Atske S. 7% of Americans don't use the internet. Who are they? Pew Research Center. 2021. URL: https://www.pewresearch.org/short-reads/2021/04/02/7-of-americans-dont-use-the-internet-who-are-they/ [accessed 2026-02-10]
26. Internet, broadband fact sheet. Pew Research Center. 2025. URL: https://www.pewresearch.org/internet/fact-sheet/internet-broadband/ [accessed 2026-02-10]
27. Faverio M. Share of those 65 and older who are tech users has grown in the past decade. Pew Research Center. 2022. URL: https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/ [accessed 2026-02-10]
28. Mariano J, Marques S, Ramos MR, et al. Too old for technology? Stereotype threat and technology use by older adults. Behav Inf Technol 2022;41(7):1503-1514. [doi: 10.1080/0144929X.2021.1882577]

XSL•FO
RenderX

29. Kakulla B. Technology adoption continues as older adults integrate tech into daily life. AARP. 2024. URL: https://www.aarp.org/pri/topics/technology/internet-media-devices/2025-technology-trends-older-adults/ [accessed 2026-02-20]
30. Older adults outnumber children in 11 states and nearly half of US counties. United States Census Bureau. 2025. URL: https://www.census.gov/newsroom/press-releases/2025/older-adults-outnumber-children.html [accessed 2026-02-10]
31. Hill JR, Patel H, Gardner B, et al. A digital health kiosk is acceptable, usable, and improves decisions about high-risk over-the-counter medications when used by older adults: results of formative user testing. J Am Pharm Assoc (2003) 2025;65(5):102451. [doi: 10.1016/j.japh.2025.102451] [Medline: 40513943]
32. Reddy A, Lester CA, Stone JA, Holden RJ, Phelan CH, Chui MA. Applying participatory design to a pharmacy system intervention. Res Social Adm Pharm 2019 Nov;15(11):1358-1367. [doi: 10.1016/j.sapharm.2018.11.012] [Medline: 30509852]
33. Werner NE, Campbell NL, Boustani M, Ganci A, Holden RJ. Helping the helpers—a research protocol for user-centered technology to aid caregiver management of medications for people with Alzheimer's disease and related dementias. Res Social Adm Pharm 2022 Sep;18(9):3680-3686. [doi: 10.1016/j.sapharm.2022.03.014] [Medline: 35396151]
34. Boustani MA, van der Marck MA, Adams N, et al. Developing the agile implementation playbook for integrating evidence-based health care services into clinical practice. Acad Med 2019;94(4):556-561. [doi: 10.1097/ACM.0000000000002497]
35. Holden RJ, Boustani MA, Azar J. Agile innovation to transform healthcare: innovating in complex adaptive systems is an everyday process, not a light bulb event. BMJ Innov 2021 Apr;7(2):499-505. [doi: 10.1136/bmjinnov-2020-000574]

Abbreviations:

OTC: over-the-counter
RAPID: Remote and Accessible Participatory Intervention Design

Edited by Y Gong; submitted 13.Aug.2025; peer-reviewed by A Duffy, C Baxter; revised version received 16.Dec.2025; accepted 09.Jan.2026; published 27.Feb.2026.

Please cite as:
Hill JR, Ganci A, Campbell NL, Pickett AC, Chui MA, Abebe E, Holden RJ
Developing mHealth IT for Older Adult Medication Safety: Remote Participatory Co-Design Using the RAPID Method
JMIR Hum Factors 2026;13:e82366
URL: https://humanfactors.jmir.org/2026/1/e82366
doi:10.2196/82366

© Jordan R Hill, Aaron Ganci, Noll L Campbell, Andrew C Pickett, Michelle A Chui, Ephrem Abebe, Richard J Holden. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 27.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Surface Electromyography as a Method for Characterizing Mammogram Discomfort: Cross-Sectional Questionnaire Study of Procedural Stress

Krystyna Gielo-Perczak*, MSE, PhD; Riley McNaboe*, BS; Hugo Posada-Quintero, PhD

Department of Biomedical Engineering, College of Engineering, University of Connecticut, Arthur B Bronwell Bldg, 260 Glenbrook Road, Storrs, CT, United States

*these authors contributed equally

Corresponding Author:
Krystyna Gielo-Perczak, MSE, PhD
Department of Biomedical Engineering, College of Engineering, University of Connecticut, Arthur B Bronwell Bldg, 260 Glenbrook Road, Storrs, CT, United States

Abstract

Background: Mammograms are the most readily used procedures for early breast cancer detection but are notorious for the discomfort they induce in patients. This physiological strain has been validated by many questionnaire-based investigations, some of which indicate that it may discourage and deter women from potentially lifesaving health care. While informative, these subjective measures are highly variable and do not provide an objective perspective regarding the coordinated physiological and ergonomic response required for the procedure.

Objective: A multimuscle surface electromyography (sEMG) methodology is proposed to better understand the machine-patient dynamics and potentially develop objective measures of mammogram-related pain and stress.

Methods: Seven different muscle pairs on the neck, shoulders, and torso were identified as being critical in postural positioning during the industry-standard mammogram procedure. sEMG was recorded during 8 compressions across 2 mammogram simulations for 25 women wearing wireless devices. An illustrative map of muscle activity was created based on a comprehensive 10-metric sEMG analysis that compared baseline recordings with activated states during each compression.

Results: The deltoid demonstrated the highest muscular activation across trials with an increase in the mean of the root mean square activity of up to 436%, while the trapezius upper fibers, infraspinatus, and teres major also showed significant increases in muscle activation, averaging 89% to 155% when compared with rest states. Across metrics, muscle activations were ipsilaterally correlated, with significant differences observed only when the breast was compressed on the same side as the muscle being measured. The serratus anterior and external oblique muscles showed minimal activation for any compression or positioning. No significant differences were found between curved and flat paddle designs. After the breast, patient-reported discomfort localized primarily to the shoulder and neck regions, corroborating the physiological measurements.

Conclusions: The juxtaposition of muscle-specific activity against inactive muscles, along with reported discomfort values, demonstrates that the sEMG methodology accurately captures the patient-machine interaction dynamics with objective, quantifiable precision that complements subjective feedback. By providing real-time physiological data on muscular patterns and biomechanical responses throughout the procedure, this sEMG-based approach offers measurable metrics for evaluating proposed improvements to mammogram equipment design and protocols.

(JMIR Hum Factors 2026;13:e83971) doi:10.2196/83971

KEYWORDS

mammogram; human factors; surface electromyography; procedural stress; discomfort

Introduction

Background

Charles Gross developed the first dedicated unit for obtaining radiographic images of breasts in 1965. Since then, mammogram machines and procedures have undergone continuous improvements in imaging quality, versatility, and accuracy [1]. However, many concerns have arisen, including high rates of false positives, radiation injury, overdiagnosis, and, most notably, procedure-induced pain [2,3]. This pain is commonly accepted by women who endure the taxing requirements of mammograms in hopes of better health [4]. Nonetheless, it is important to try to understand and address mammogram discomfort and related stress both out of compassion for patients and to encourage regular screening.

XSL•FO
RenderX

Understanding the manifestation of pain, both the physical toll it takes and the cognitive stress it induces, is inherently complex. While many researchers have used questionnaires and interviews to better understand mammogram-induced pain in women, a more complete methodology is required to fully examine all the variables that contribute to the discomfort of the procedure. A quantitative measure, such as that provided by surface electromyography (sEMG), can provide an objective perspective on the prevalence and intensity of mammogram-induced discomfort. sEMG is a noninvasive method using sensors on the skin to measure the change in voltage potential that occurs when a muscle is stimulated by the nervous system [5]. The primary waveform acquired is informative, as it reflects muscle activation resulting from the overall recruitment of fibers during contraction. sEMG also enables calculation of a wide range of secondary metrics, each of which can uniquely characterize different signal features as well as the physiological pathway that oversees their activity.

The depth of analysis that sEMG provides allows for a complex, multifaceted mapping of muscle activity, especially when multiple muscles are monitored simultaneously. Analysis of the signal and its related metrics has been used in a wide range of applications including, but not limited to, studies of comfort and ergonomics [6], emotion detection and regulation [7], and diagnostic techniques for muscular diseases [8]. Similarly, sEMG has been used in various approaches to identify and quantify pain and stress [9-11].

As a noninvasive method that can characterize a wide range of human states (eg, muscular demand and discomfort), sEMG is ideal for a comprehensive, quantitative biometric analysis of positioning and the resulting discomfort during mammograms. Such discomfort is realized when the technologist positions the patient into a forced and unnatural posture to ensure proper imaging. The patient stands with feet pointed at the machine and torso parallel to the imaging table while simultaneously rotating the head and grasping the side of the machine with a single arm. While holding this demanding position, patients must hold their breath and remain completely still. A detailed understanding of the muscular response of a patient during this machine-human positioning can provide greater context to fully understand the effect it has on the body. The breast-positioning procedure is extremely demanding on the patients' muscles, ultimately inducing high physiological strain on the patient, a state that is further exacerbated by the intrinsic psychological stress associated with the screening itself.

Building upon our preliminary investigations presented in the studies by Gielo-Perczak et al [12,13], this study presents a comprehensive multimetric analysis of sEMG signals collected from 14 different muscles across 25 women during a traditional mammogram procedure. While previous work provided initial insights into a select range of muscles, the current investigation significantly expands both the scope and the depth of the analysis to evaluate the practicality and efficacy of using sEMG as a tool to understand the patient-machine dynamics and to quantify patient discomfort. The methodology seeks to understand the effects of different breast compressions, the required body positioning, and the overall physiological demand of the procedure. We analyzed the performance of 10 different sEMG metrics across each muscle throughout the baseline phases and stimulated (breast compression) phases. We then compared the difference in each metric response between the 2 phases to identify any significant trends resulting from the force of the paddles and the positioning required for the imaging.

Previous Work

The application of sEMG to understand muscular distress or discomfort and the ergonomics of specific tasks or procedures has been widely explored. This work generally involves naturalistic observation whereby a muscle and an electromyography (EMG) metric are analyzed throughout a given task. Some groups expand upon this approach by including additional muscles or metrics but rarely both. By monitoring the signal during either real or simulated activities, sEMG has been used for diagnostic purposes as well as for targeting specific, localized task characteristics. This technique has helped to identify the physiological impact and strain of specific, observed tasks on participants, such as workers, patients, surgeons, or even musicians.

For example, Steinmetz et al [14] recorded sEMG from the sternocleidomastoid muscle of 2 groups of violinists—those with and those without neck pain. By analyzing the root mean square (RMS) of the signal, the research simultaneously observed the effect of violin playing on the muscle and examined how the presence or absence of pain could be captured with the signal. A similar, single-muscle approach was conducted by de Oliveira and Nadal [15], who analyzed the erector spinae of helicopter pilots to evaluate the effect of required in-flight posture and environmental vibrations on their back muscles. In addition to analyzing the RMS of the signal, they derived the median frequency to quantify muscle fatigue more specifically over time, allowing for a wider analysis. Harnessing a similar multimetric approach, Wijsman et al[16] and collaborators targeted the trapezius upper fibers during a cognitive task representative of work-related stress. They analyzed 9 different sEMG metrics including RMS, mean value of the RMS, median value of the RMS, peak value of the RMS, mean frequency of the signals-derived power spectra, median frequency of the signals-derived power spectra, and trends. Almost all the metrics indicated a significant difference when comparing the response between a given stressed and relaxed state, and some simultaneously exhibited differences between *types* of stressed states. The resulting multimetric visual gave a more profound interpretation as the frequency-based features were analyzed for muscle fatigue over the task period. RMS-based metrics were used to quantify the overall activation, while trends allowed for a synthesized report of RMS activity over time.

In each of these studies, the tasks required not only the targeted muscles but also many other muscles that were not a primary focus. While a multimetric approach like Wijsman's [16] provides a more profound analysis of sEMG from a given muscle, a broader perspective is gained from evaluating entire task-relevant sections of the body. Bhardwaj et al [17] implemented such an approach in a study of truck driver fatigue. Bhardwaj's team collected bilateral sEMG from 3 different posterior muscle groups (trapezius medial, latissimus dorsi

XSL•FO
RenderX

medial, and erector spinae) as drivers operated both semitrucks and simulators. Although they implemented only 1 metric, the additional muscle data provided a more comprehensive view than if only a single muscle was selected. Both Luttmann et al [18] and Niu et al [19] analyzed multiple sEMG metrics to understand the ergonomics or effects of performing surgeries. Luttmann's team used electrical activity (a metric indicative of smoothed and rectified sEMG) and median frequency to characterize sEMG from 3 different posterior muscles over a series of repeated surgeries. With this approach, they were ultimately able to quantify the muscular strain and fatigue. Niu and colleagues [19] observed surgeons' arm muscles during laparoscopic and robotic surgeries to assess biomechanical stress. The analysis focused on integrated electromyography and mean frequency derived from 4 muscles located in the lower and upper arms. These 2 research teams, among others, demonstrate the value of a multimetric, multimuscle approach when implementing sEMG as a quantitative tool.

Often, a single muscle observation is sufficient to satisfy study objectives. However, when the study seeks a more complete illustration of muscular-related physiology (including factors such as posture, pain, fatigue, or stress), additional muscle observations augment the context for interpreting the results. Similarly, using multiple sEMG features in the analysis allows for more profound conclusions. The studies previously mentioned are summarized in Table 1.

XSL•FO
RenderX

Table . Previous studies.

Reference	Participants, n	EMG[a] application	Methodology	Muscles analyzed	Metrics used	Comparison method
Wijsman et al [16] (2013)	30	Cognitive stress	Nine-features analysis of EMG to predict stressed state during a simulated cognitive stress test (VAS[b] and PSS[c])	Trapezius upper fibers	RMS[d], gaps/minute, relative gap time, static/median/peak RMS, mean/median frequency, and trends	% RVC[e]
Steinmetz et al [14] (2016)	55	Task-related muscle function	One-feature analysis of sEMG[f] from violinists with and without neck pain, to understand the impact of playing on pain of the muscle (VAS and NDI[g])	Sternocleidomastoid	RMS	Δ |RMS|
Luttmann et al [18] (1998)	4	Muscular strain	Two-feature analysis of sEMG from urologists over repeated surgeries to analyze muscular strain and fatigue	Trapezius, deltoid, and erector spinae	EA, median frequency	% Maximum EA[h] and Δ |MF|[i]
Bhardwaj et al [17] (2018)	25	Driver fatigue	One-feature analysis to compare efficacy of using EMG to analyze driver fatigue in simulators in comparison with real-life scenarios	Trapezius medial, latissimus dorsi medial, and erector spinae	Mean frequency	Coherency
de Oliveira and Nadal [15] (2004)	12	Occupational muscular fatigue	Two-feature analysis of EMG in comparison with MVC[j] to evaluate the effect of posture and vibration on helicopter pilots' back muscle fatigue while flying	Erector spinae	RMS, median frequency	% MVC
Niu et al [19] (2020)	7	Biomechanical stress	Two-feature analysis of surgeons' lower and upper arms while conducting laparoscopic and robotic surgery to evaluate the difference in biomechanical stress to improve the surgery protocol or ergonomics	Flexor carpi radialis, brachioradialis, biceps brachii, and medial deltoid	iEMG[k], mean frequency	% MVC

[a]EMG: electromyography.

[b]VAS: visual analog scale.

[c]PSS: Perceived Stress Questionnaire.

[d]RMS: root mean square.

[e]RVC: reference voluntary.

[f]sEMG: surface electromyography.

[g]NDI: Neck Disorder Index.

XSL•FO
RenderX

[h]EA: electrical activity.

[i]MF: median frequency.

[j]MVC: maximum voluntary contraction.

[k]iEMG: integrated electromyography.

Methods

Study Design

The proposed sEMG methodology to evaluate patient stress and discomfort during mammograms consists of a multimuscle, multimetric analysis. Researchers applied wearable sensors to 25 participants to simultaneously collect sEMG from 14 different relevant muscles during a simulated mammogram. The recordings collected during the imaging phase provide a comprehensive snapshot of muscular states during compressions. An in-depth analysis of 10 unique sEMG metrics provides important contextual information for the muscular mapping.

Mammogram Positioning and Muscle Selection

The vertical, standing system is one of the most widely used imaging methods for mammograms. While slight variations can be seen between systems depending on the manufacturer, the overall patient-machine interaction remains consistent throughout the field. During the procedure, the patient stands with both feet facing toward the machine and torso pressed against the edge of an imaging table that protrudes from the device. The patient's head is rotated significantly to the left or the right (depending on the breast being imaged), with the side of the face pressed against a rigid shield. The arm opposite to the breast being imaged simultaneously extends toward the machine.

Considering the rigid structures of the machine as well as the stationary placement of the feet, there are several muscles that can be evaluated to capture the restricted trunk torsion in conjunction with the head rotation and arm extension. The serratus anterior and external oblique can be observed to characterize relevant activity in the torso region. The deltoids along with the infraspinatus and teres major groups provide insight into the shoulder and upper arm during arm extension. The sternocleidomastoid and trapezius upper fibers together illustrate neck movements during rotation and extended flexion.

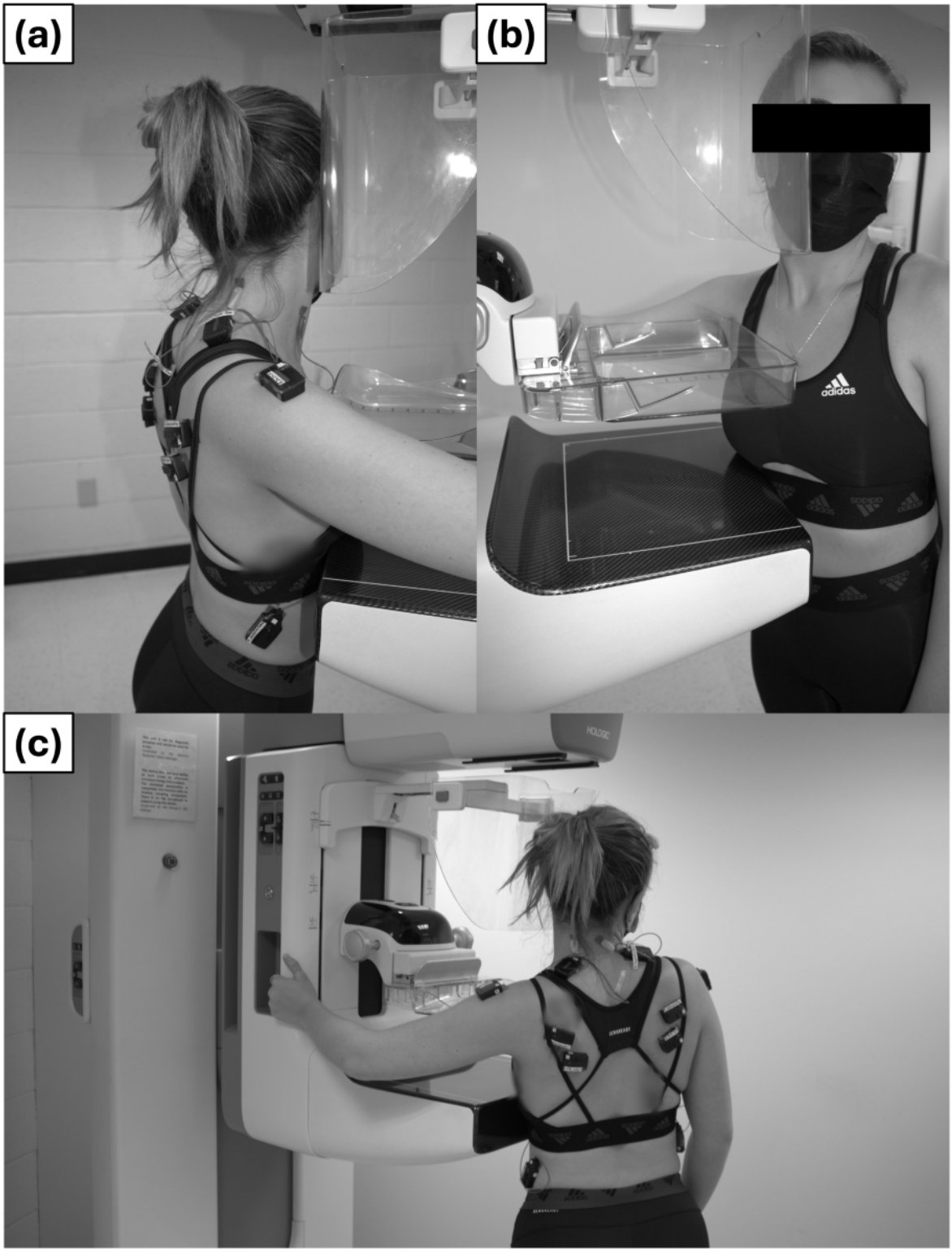
The left and right pairs of these 7 muscles (14 muscles total) were selected for observation in the proposed methodology. These muscles consisted of previously studied muscles [20] as well as a selection of new muscles that targeted areas not yet evaluated. It is important to note that none of the muscles identified are on or in direct contact with the breasts. The breasts themselves have no muscles and consist of lobules and ducts encapsulated by fatty adipose tissue. The proposed sEMG methodology focuses on muscle strain as a result of patient-machine positioning rather than the force applied to the breast itself.

Instrumentation

Surface EMG recordings were taken using wearable Trigno and Avanti sensors from Delsys, Inc. Four different sensor types were used including the Avanti, Avanti Duo, Trigno, and Trigno Mini. The type of sensor used was determined by the characteristics of the corresponding muscle and the restrictions presented by the procedure. The Avanti and Trigno sensors include embedded electrodes and are ideal for direct placement on muscles that exhibit limited movement during the imaging (placed parallel to the muscle fiber pattern on the center medial deltoid, lowest rib on the external abdominal oblique, center infraspinatus, and center teres major). The Mini and Duo systems use external reference electrodes connected by a wire to the measured electrodes, making them more suitable for smaller, isolated muscle areas that show more movement during the procedure (placed parallel to the muscle fiber pattern on the center trapezius upper fibers, center sternocleidomastoid, and the center band of the serratus anterior). Despite differences in electrode configuration, all 4 Delsys sensor models share identical sampling rate, bandwidth, and input range specifications, providing consistent EMG signal characteristics. To account for any minor variability intrinsic to sensor design or placement, all EMG signals were normalized prior to analysis to ensure valid comparisons across sensors and participants. Both the Avanti and Trigno sensor systems use default sampling defined by Delsys (1925.925 Hz). Figure 1A-C shows the localization of some of these sensors.

XSL•FO
RenderX

Figure 1. Example of human-machine positioning for compression and imaging (note: research staff member shown in images; participants were in a hospital gown for the duration of the study). (A) Lateral view of right craniocaudal compression position, (B) front view of right craniocaudal compression, and (C) rear view of left craniocaudal compression.



Simulation

The mammogram procedure was simulated using the 3Dimensions Genius Mammography System from Hologic, Inc. This system (Figure 1) is commonly found in medical offices across the country and is similar to many other brands. Two different paddle designs—a flat, industry standard paddle and a newer Hologic SmartCurve Breast Stabilization Paddle—were used throughout the protocol to observe whether these design modifications produced any response differences [21]. A certified mammography technologist adjusted the height of the machine and positioned the breast for the simulated compressions to ensure proper handling of the patient and to provide a realistic simulation experience.

XSL•FO
RenderX

Ethical Considerations

The University of Connecticut Institutional Review Board for human participant research approved all the study procedures (protocol H20-0146). Participants gave consent after reviewing the subject protocol.

All identifiable information collected from the participants was numerically coded and anonymized.

Compensation of US $100 was given to participants after completing the visit for the study.

Protocol

The study enrolled 25 healthy female volunteers (aged 40 - 67, median 50.8, SD 8 years; mean height 163.7, SD 7 cm; mean weight 78.0, SD 21 kg; and mean BMI 28.9, SD 6 kg/m^2). The sample size was determined based on practical considerations such as recruitment feasibility and clinical workflow constraints, and it aligns with typical cohort sizes in similar procedural studies, which often include 10 to 20 participants [20]. While both men and women are susceptible to breast cancer, only biological female volunteers were selected for this study as breast cancer, which requires repetitive mammography screenings, is significantly more common in this population.

Participants were screened using a preliminary questionnaire that flagged allergies related to any of the study equipment and identified previous injuries or scars that could impact the procedure. Participants were required to have previously completed at least 1 standard mammogram procedure to provide accurate feedback. Additional self-reported information was collected, including previous mammogram visual analog scale (VAS) pain or stress scores from 1 to 10 and a reference level for physical activity derived from grip strength recordings [22].

Once screened, the participant changed into a hospital gown in a private room. The participant's left and right deltoid, infraspinatus, teres major, external oblique, serratus anterior, trapezius upper fibers, and sternocleidomastoid were equipped with an array of wearable, wireless EMG sensors. Prior to the placement of any device, the skin at each electrode location was cleaned with 70% isopropyl alcohol. The placement and orientation of each sensor were kept constant for all participants to aid in signal analysis.

An initial EMG control recording (C1) was taken in a bright, large room (1341 square feet) during which the participant stood still in a relaxed position facing forward. This recording, as well as all subsequent EMG control recordings, lasted 2 minutes. The participant was then guided to the Hologic Mammography system, which was housed in a smaller, quiet, moderately lit room (171 square feet) with an ambient temperature ranging from 26 °C to 27 °C. Another control recording (C2) was taken in this smaller room before starting the simulated mammogram (Table 2).

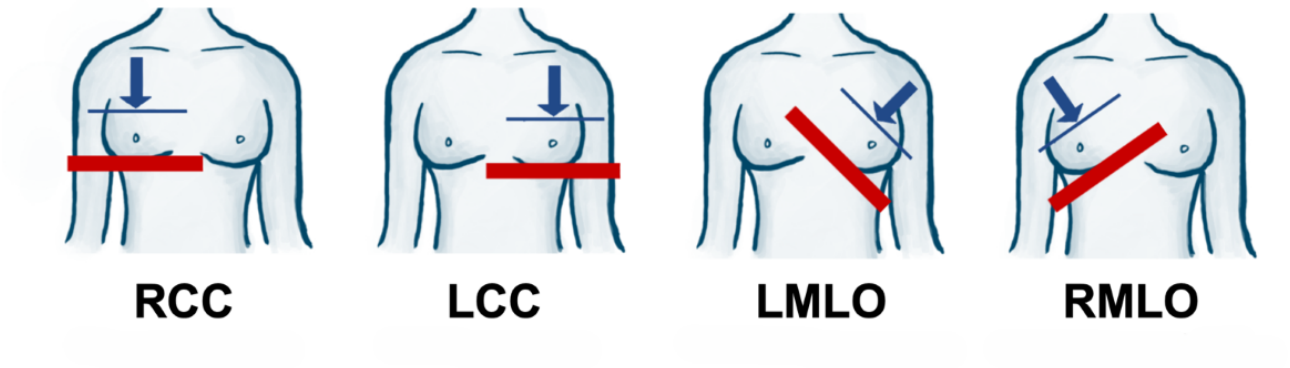
The team then initiated the mammogram simulation. The certified mammography technologist walked the participant through the procedure, which followed the same order of compressions currently used in the health care field. The compressions took the following order: (1) right craniocaudal (RCC) view (the right breast is compressed vertically), (2) left craniocaudal (LCC) view (the left breast is compressed vertically), (3) left mediolateral oblique (LMLO) view (the left breast is compressed laterally), and (4) right mediolateral oblique (RMLO) view (the right breast is compressed laterally). Figure 2 illustrates the 4 compressions. Both CC and MLO views require the participant to grasp the imaging unit for stability. In MLO views, the imaging table extends beneath the axilla, creating a larger body-machine contact surface and placing greater positional demands on the upper arm and shoulder girdle relative to the more vertically oriented CC positioning. Each of the 4 compressions (15 seconds each) was conducted twice, once using the flat traditional paddle (F_) and again with the curved SmartCurve Hologic Paddle (C_). The EMG recording was collected during the simulated imaging portion of the procedure when the breast was fully compressed, and the participant was in the restricted body position holding their breath. The compression force used was less than that which is typically applied in clinical settings. This was done to limit unnecessary participant pain since this work focuses on strain from patient-machine positioning rather than breast compression pain. No X-ray was emitted in this procedure. All other steps followed standard clinical practice. A third control recording (C3) was taken between the first and second sets of 4 compressions when the paddle was being changed. The entire protocol is summarized in Table 2.

XSL•FO
RenderX

Table . . Study protocol.

Compression event	Abbreviations	Duration
Control 1 in large room	C1	2 minutes
Control 2 in examination room	C2	2 minutes
Flat paddle compressions		
Right craniocaudal compression	F_RCC	15 seconds
Left craniocaudal compression	F_LCC	15 seconds
Left mediolateral oblique compression	F_LMLO	15 seconds
Right mediolateral oblique compression	F_RMLO	15 seconds
Control 3 in examination room	C3	2 minutes
Curved paddle compressions		
Right craniocaudal compression	C_RCC	15 seconds
Left craniocaudal compression	C_LCC	15 seconds
Left mediolateral oblique compression	C_LMLO	15 seconds
Right mediolateral oblique compression	C_RMLO	15 seconds

Figure 2. The 4 different positions of the breast for the simulated mammogram—typical for screening procedures. Blue line and arrow indicate paddle and direction of compressive force, respectively. Red line indicates imaging table on which the breast rests while being compressed. LCC: left craniocaudal; LMLO: left mediolateral oblique; RCC: right craniocaudal; RMLO: right mediolateral oblique. Adapted from Gielo-Perczak et al [12,13].



EMG Signal Processing and Data Analysis

The sEMG processing and analysis was completed using MATLAB. The first and last 1.5 seconds were removed from the beginning and the end of each recording to remove fringe artifacts resulting from the misalignment with the start of the recording and the start of the imaging phase. The trimmed segment was then filtered with a fourth-order bandpass filter (20-500 Hz), which accounted for the variability in unique frequencies across the targeted muscles. Each segment for the flat paddle was normalized to the first control of the small room (C2), and the second set of recordings corresponding to the curved paddle was normalized to the second control of the small room (C3).

Ten different metrics were then derived from the cleaned signals to characterize each of the segments. The selection of these metrics was informed by prior literature in which similar measures have proven effective for assessing muscular pain, stress, fatigue, and strain in physically demanding contexts (Table 1). Five time-domain metrics were selected. These include the mean of the root mean square (meanRMS) waveform, standard deviation of the RMS (stdRMS) feature, slope sign change count, zero crossing count, and the waveform length (WFL). Five frequency domain metrics were selected to further analyze the EMG activity during each state of the protocol. These included the mean frequency (meanFreq), median frequency (medianFreq), signal spectral entropy, drop in power (DP), and spectral deformation (SpecDef). To determine these features, the power spectrum of each EMG signal was calculated using Welch's periodogram method with 50% overlap and a 128-sample Blackman window on the filtered data. The average of each metric for a measurement window was used for analysis. Given the static nature of the participants during each measurement period and observed uniformity in sEMG data, limited variation in metrics was expected across each window.

Data Analysis

Three interactions between stages of the protocol were identified as potentially illustrative comparisons. The first interaction illustrates the sEMG activity between the compressive state and its respective control. This finding highlights the sensitivity of the measure to illustrate the change from a relaxed to aroused state while also indicating overall muscle activity (eg, C_RCC

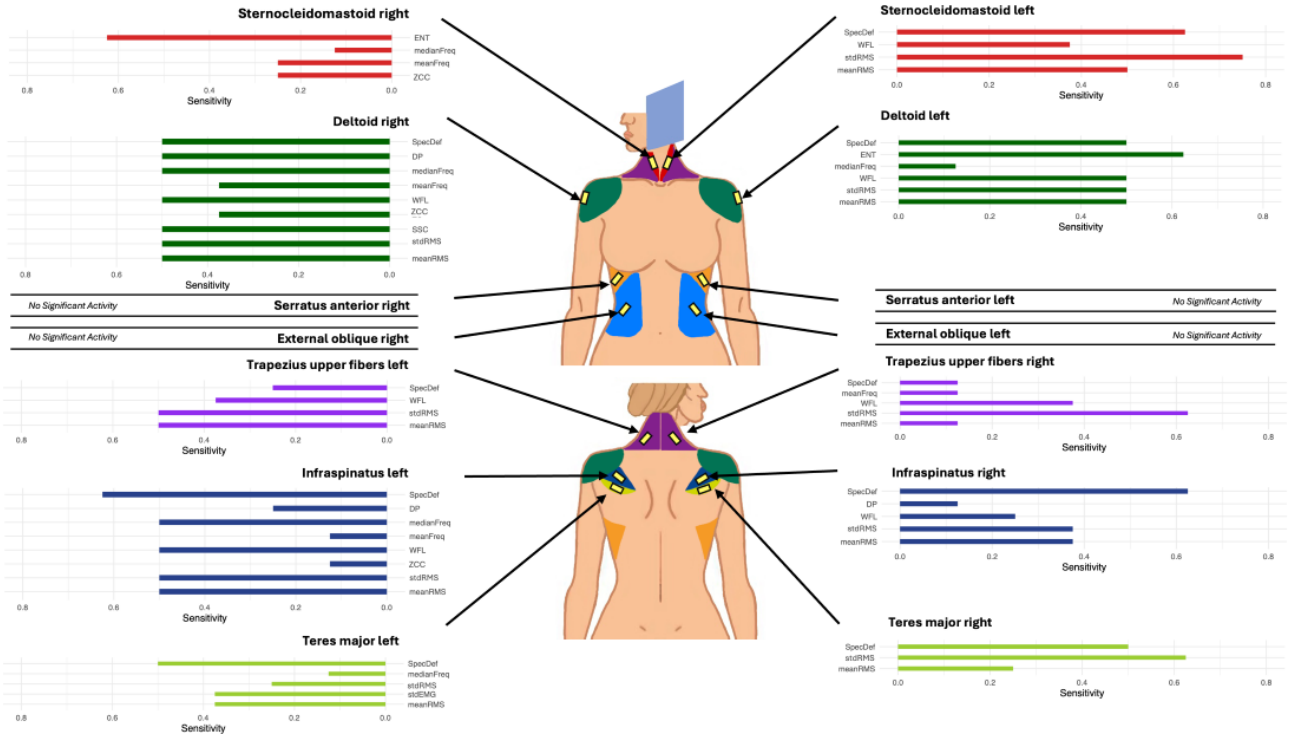
XSL•FO
RenderX

vs C3). The second interaction revealed the sEMG activity between the same compressive state between the 2 different paddle types. This interaction would indicate any difference in muscle activity resulting from a change in the paddle form (eg, F_RCC vs C_RCC). The third interaction revealed the sEMG activity that occurred when the compression angle was changed for a given breast to identify any effect of angle on the signal (eg, F_RCC vs F_RMLO). Normality for each metric was assessed using the Kolmogorov-Smirnov test. Metrics were analyzed using paired *t* tests. Because each metric involved multiple comparisons across compressions, *P* values were adjusted using Bonferroni correction. Bonferroni was selected due to its conservative nature and its suitability for protecting against type 1 errors in settings with many planned comparisons.

Results

The first interaction comparing activity during a compression to a control state was identified as the most informative with respect to the main goals of the proposed methodology. For this comparison, the metrics for a given muscle were evaluated across all 8 compression segments and the 2 small room controls C2 and C3. Statistical difference was noted according to the corrected Bonferroni values. This resulted in a total of 8 comparisons for a given metric and muscle combination. Figure 3 summarizes the percentage of compressions where this comparison was significant. The metrics listed for each muscle are those that detected a significant difference in sEMG activity at least once between relaxed and stimulated states for a given muscle.

Figure 3. Percentage of compressions that showed a statistical difference between control and baseline for each muscle across all metrics. Displayed as a ratio out of the total number of compressions (n=8). Only metrics that showed significance for a given muscle are displayed. DP: drop in power; ENT: spectral entropy; meanFreq: mean frequency; meanRMS: mean of the root mean square; medianFreq: median frequency; SpecDef: spectral deformation; SSC: slope sign change count; stdRMS: standard deviation of the root mean square; WFL: waveform length; ZCC: zero crossing count.



The deltoid and infraspinatus muscles (left and right) demonstrated high relative sensitivity according to the frequency matrix when observing meanRMS, stdRMS, WFL, meanFreq, medianFreq, and SpecDef. The deltoid consistently showed differences in about 50% of cases. Similarly, the trapezius upper fibers and the teres major were also relatively sensitive to compressions when observing meanRMS, stdRMS, WFL, and SpecDef. The serratus anterior and the external oblique did not show any sensitivity to the compressions for any of the metrics explored. Regardless of the other muscles, meanRMS, stdRMS, WFL, mean and medianFreq, and SpecDef all indicated a change in sEMG between the 2 states. It is noted that meanRMS, stdRMS, and WFL are all somewhat correlated in addition to meanFreq and medianFreq. While each is slightly different (as they underscore various characteristics of the signal), the methodology of their derivation lends to shared trends, providing flexibility to researchers in metric selection when pursuing similar methods.

Those muscles that showed both frequent and significant responses across various metrics (all but the serratus and external oblique groups) are included in Table 3. The top 3 most frequently activated metrics in each muscle group are present for both paddle types. Any compression resulting in a significant change in comparison to the controls is noted with an asterisk and red font. Note that only C2 and C3 recordings are included in the table as they were the reference for the flat and curved paddle compression sets, respectively. No significant difference was found between the large room control C1 and C2 or between C1 and C3 in the examination room.

XSL•FO
RenderX

Table . Results for left and right muscle pairs across selected metrics. Most responsive metrics for each muscle pair. Values are presented as mean (SD).

	Control	Right craniocaudal	Left craniocaudal	Left mediolateral oblique	Right mediolateral oblique
Sternocleidomastoid					
Left					
meanRMS[a] (mv)					
Flat	0.9481 (0.021)	1.569 (0.805)	**1.757 (0.861)** [b]	**1.872 (1.18)** [b]	1.399 (0.692)
Curved	0.9331 (0.0383)	1.547 (0.863)	**2.199 (1.12)** [b]	**2.13 (1.39)** [b]	1.429 (0.757)
stdRMS[c] (mv)					
Flat	1 (1.052e-15)	**1.594 (0.707)** [b]	**1.939 (0.91)** [b]	**2.037 (1.26)** [b]	1.703 (0.972)
Curved	1 (7.21e-16)	**1.883 (1.14)** [b]	**2.432 (1.25)** [b]	**2.204 (1.28)** [b]	1.616 (0.896)
ENT[d] (nu)					
Flat	0.7176 (0.016)	0.7088 (0.0138)	0.7071 (0.0125)	0.7116 (0.0204)	0.7115 (0.0169)
Curved	0.7019 (0.0307)	0.7024 (0.0258)	0.7098 (0.0129)	0.7103 (0.017)	0.7076 (0.0194)
Right					
meanRMS (mv)					
Flat	0.9634 (0.007)	1.701 (1.12)	1.097 (0.38)	1.055 (0.253)	1.807 (1.56)
Curved	0.959 (0.00943)	1.444 (0.789)	1.207 (0.49)	1.166 (0.37)	1.939 (1.48)
stdRMS (mv)					
Flat	1 (1.069e-15)	1.838 (1.21)	1.163 (0.385)	1.099 (0.258)	2.068 (1.86)
Curved	1 (6.86e-16)	1.434 (0.661)	1.434 (0.791)	1.278 (0.458)	2.211 (1.63)
ENT (nu)					
Flat	0.7277 (0.011)	**0.7152 (0.0118)** [b]	0.7202 (0.00728)	0.7207 (0.00168)	**0.7084 (0.0172)** [b]
Curved	0.7252 (0.00647)	**0.7054 (0.021)** [b]	**0.7173 (0.00345)** [b]	0.7192 (0.00556)	**0.708 (0.0181)** [b]
Trapezius upper fibers					
Left					
stdRMS (mv)					
Flat	1 (7.96e-16)	**1.493 (0.447)** [b]	**2.158 (1.26)** [b]	**2.35 (1.32)** [b]	**1.51 (0.513)** [b]
Curved	1 (7.05e-16)	**1.883 (1.14)** [b]	**2.432 (1.25)** [b]	**2.204 (1.28)** [b]	1.616 (0.896)
WFL[e] (nu)					
Flat	6986 (893)	8927 (2450)	**15,360 (9420)** [b]	**14,090 (6330)** [b]	**9954 (3190)** [b]
Curved	6255 (1140)	8257 (3340)	9558 (5250)	11,080 (6560)	10,960 (7160)
SpecDef[f] (Ω/Ω[g])					
Flat	2.001 (0.302)	1.874 (0.369)	**1.609 (0.411)** [b]	**1.569 (0.415)** [b]	1.761 (0.441)
Curved	2.039 (0.333)	1.884 (0.396)	1.732 (0.36)	1.805 (0.441)	1.848 (0.389)
Right					
stdRMS (mv)					
Flat	1 (1.09e-15)	**1.448 (0.537)** [b]	1.393 (0.605)	**1.784 (1.01)** [b]	**2.213 (1.49)** [b]
Curved	1 (9.12e-16)	1.434 (0.661)	1.434 (0.791)	1.278 (0.458)	2.211 (1.63)
WFL (nu)					
Flat	6528.9 (1019.34)	**9596 (3550)** [b]	**8996 (2590)** [b]	**10,750 (5180)** [b]	12,060 (7320)

XSL•FO
RenderX

	Control	Right craniocaudal	Left craniocaudal	Left mediolateral oblique	Right mediolateral oblique
Curved	6168 (1610)	8495 (4650)	9464 (6120)	8796 (5050)	11,530 (7670)
SpecDef (Ω/Ω)					
Flat	1.9636 (0.2187)	1.711 (0.419)	1.739 (0.31)	1.749 (0.456)	**1.569 (0.425)** [b]
Curved	1.913 (0.257)	1.787 (0.337)	1.811 (0.319)	1.836 (0.327)	1.779 (0.476)
Deltoid					
Left					
meanRMS (mv)					
Flat	0.9655 (0.00909)	1.374 (0.588)	**4.814 (2.44)** [b]	**5.287 (3.64)** [b]	1.176 (0.416)
Curved	0.9518 (0.0199)	1.073 (0.28)	**4.195 (1.91)** [b]	**4.996 (3.17)** [b]	1.196 (0.517)
WFL (nu)					
Flat	7198 (898)	9563 (4460)	**32,880 (17,800)** [b]	**33,420 (23,800)** [b]	7874 (2980)
Curved	6368 (1250)	7076 (2130)	**27,310 (13,100)** [b]	**36,760 (27,600)** [b]	7723 (3140)
SpecDef (Ω/Ω)					
Flat	1.996 (0.1189)	1.874 (0.212)	**1.187 (0.146)** [b]	**1.234 (0.233)** [b]	1.974 (0.292)
Curved	2.062 (0.187)	2.002 (0.244)	**1.256 (0.16)** [b]	**1.279 (0.282)** [b]	1.884 (0.224)
Right					
meanRMS (mv)					
Flat	0.9639 (0.0118)	**3.884 (2.37)** [b]	1.112 (0.307)	1.237 (0.501)	**3.963 (2.84)** [b]
Curved	0.9631 (0.0102)	**2.913 (2.07)** [b]	1.042 (0.406)	1.135 (0.444)	**4.143 (2.73)** [b]
WFL (nu)					
Flat	4642 (718)	**21,750 (14,300)** [b]	5822 (2470)	5779 (1930)	**22,720 (17,200)** [b]
Curved	4400 (718)	**16,180 (12,700)** [b]	4708 (1760)	5184 (2140)	**22,430 (15,500)** [b]
SpecDef (Ω/Ω)					
Flat	1.773 (0.1295)	**1.019 (0.227)** [b]	1.6 (0.334)	1.683 (0.185)	**1.157 (0.432)** [b]
Curved	1.782 (0.135)	**1.113 (0.277)** [b]	1.6 (0.273)	1.653 (0.189)	**1.133 (0.337)** [b]
Infraspinatus					
Left					
meanRMS (mv)					
Flat	0.9448 (0.0207)	1.175 (0.309)	**2.158 (1.09)** [b]	**1.945 (0.762)** [b]	1.266 (0.433)
Curved	0.9351 (0.0245)	1.194 (0.436)	**1.905 (0.92)** [b]	**2.168 (1.19)** [b]	1.064 (0.389)
WFL (nu)					
Flat	6128 (907)	7298 (1860)	**14,390 (6570)** [b]	**14,910 (9120)** [b]	8023 (2750)
Curved	5673 (1080)	7178 (2580)	**12,350 (5930)** [b]	**15,180 (9300)** [b]	6734 (2570)
SpecDef (Ω/Ω)					
Flat	2.226 (0.266)	2.064 (0.275)	**1.539 (0.305)** [b]	**1.589 (0.355)** [b]	1.95 (0.364)
Curved	2.236 (0.283)	2.046 (0.311)	**1.644 (0.32)** [b]	**1.515 (0.223)** [b]	**1.916 (0.2)** [b]
Right					
meanRMS (mv)					

XSL•FO
RenderX

	Control	Right craniocaudal	Left craniocaudal	Left mediolateral oblique	Right mediolateral oblique
Flat	0.9606 (0.0113)	**1.518 (0.605)** [b]	1.197 (0.319)	**1.275 (0.361)** [b]	**1.698 (0.73)** [b]
Curved	0.9533 (0.0156)	1.239 (0.492)	1.228 (0.551)	0.9943 (0.237)	1.329 (0.596)
WFL (nu)					
Flat	6556 (1230)	**10,040 (4130)** [b]	7817 (2430)	7998 (2470)	**11,500 (5040)** [b]
Curved	5887 (1340)	9090 (4820)	7802 (3780)	7122 (3340)	9552 (5130)
SpecDef (Ω/Ω)					
Flat	2.174 (0.217)	**1.831 (0.272)** [b]	2.07 (0.253)	2.003 (0.282)	**1.731 (0.344)** [b]
Curved	2.135 (0.229)	**1.803 (0.352)** [b]	**1.897 (0.279)** [b]	1.957 (0.246)	**1.798 (0.361)** [b]
Teres major					
Left					
stdRMS (mv)					
Flat	1 (7.3e-16)	1.155 (0.249)	**1.703 (0.679)** [b]	**1.843 (0.796)** [b]	1.181 (0.366)
Curved	1 (6.89e-16)	1.077 (0.254)	1.58 (0.797)	**1.753 (0.934)** [b]	1.118 (0.482)
WFL (nu)					
Flat	6296 (1360)	6793 (1880)	**10,070 (3550)** [b]	**11,010 (5020)** [b]	6921 (2110)
Curved	5627 (1500)	6267 (2050)	8999 (4470)	9262 (4470)	6239 (2640)
SpecDef (Ω/Ω)					
Flat	2.344 (0.267)	2.199 (0.361)	**1.842 (0.3)** [b]	**1.851 (0.36)** [b]	2.131 (0.414)
Curved	2.335 (0.33)	2.17 (0.326)	**1.948 (0.349)** [b]	**1.87 (0.317)** [b]	2.04 (0.409)
Right					
stdRMS (mv)					
Flat	1 (7.3e-16)	**1.5 (0.637)** [b]	**1.362 (0.466)** [b]	**1.445 (0.461)** [b]	**1.896 (1.15)** [b]
Curved	1 (9.21e-16)	1.703 (0.986)	1.187 (0.376)	1.248 (0.52)	**1.824 (0.887)** [b]
WFL (nu)					
Flat	6371 (1110)	8793 (3430)	8071 (2920)	8358 (2730)	10280 (5330)
Curved	5959 (1280)	8853 (4580)	7662 (3450)	7051 (3130)	9616 (5260)
SpecDef (Ω/Ω)					
Flat	2.043 (0.152)	**1.805 (0.264)** [b]	1.84 (0.257)	1.92 (0.185)	**1.746 (0.336)** [b]
Curved	2.032 (0.228)	**1.829 (0.202)** [b]	1.773 (0.268)[b]	1.765 (0.328)	1.765 (0.352)

[a]meanRMS: mean of root mean square.

[b]Bolding indicates significant difference between compressive state and respective control (C2 for flat paddle compressions, C3 for curved paddle compressions) with $P<.05$.

[c]stdRMS: standard deviation of the root mean square.

[d]ENT: spectral entropy.

[e]WFL: waveform length.

[f]SpecDef: spectral deformation.

[g]Ω/Ω: Ohm ratio—unitless.

For the sternocleidomastoid, deltoid, infraspinatus, and teres major muscle groups, significant differences were observed only during compression of the ipsilateral breast (the breast on the same side as the muscle being measured). This pattern is demonstrated by the sensitivity clustering in the LCC and LMLO columns for left-sided muscles and in the RCC and RMLO columns for right-sided muscles.

Curved and flat paddle compressions resulted in similar changes in the metric performance. There was no significant difference found between compressions of the 2 paddle designs. This can

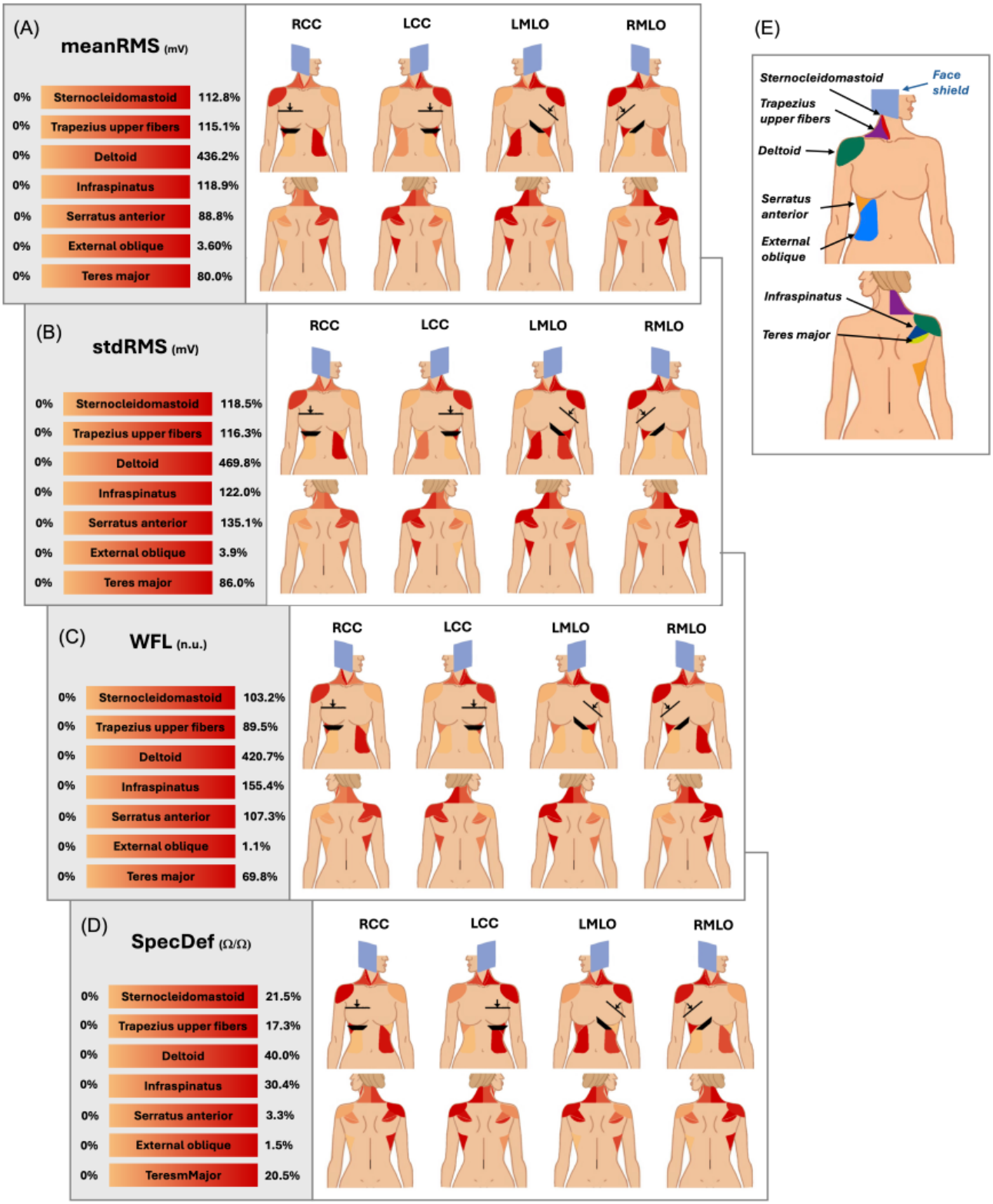
XSL•FO
RenderX

be similarly extrapolated from Table 3 where pairs of rows generally follow the same trend. Parentheses indicate the larger of the 2 mean values for a given metric and compression. It is important to note that in the case of metrics meanFreq, medianFreq, and DP, a decrease in mean was found in response to the provided stimuli as is commonly expected in physiological response to stressors [23].

Figure 4 clarifies the differences by presenting the absolute percentage change relative to the control for the 4 most responsive metrics of the study. Figure 4E provides a summary key of the muscles of interest. Combining the curved and flat compressions into 1 population (RCC, LCC, LMLO, and RMLO), the physiological heatmaps illustrate the change for all the muscles studied. The deltoid had a maximum percentage change of 436.2% when considering meanRMS (Figure 4A and C). The increase is corroborated by the results in Table 3. Similar percentages are seen across stdRMS, WFL, and SpecDef, noting 469.8% - 420.7% and 40.0%, respectively (Figure 4B-D). The teres major showed a maximum increase of 80.0%, 86.0%, 69.8%, and 20.5% for each of the metrics, which are also significant increases. On average, the sternocleidomastoid increased 112.8%, 118.5%, 103.2%, and 155.4%; the trapezius upper fibers 115.1%, 116.3%, 89.5%, and 17.3%; and infraspinatus 118.9%, 122%, 155.4%, and 30.4% for meanRMS, stdRMS, WFL, and SpecDef, respectively. For the external oblique, averages were 3.6%, 3.9%, 1.1%, and 1.5% for meanRMS, stdRMS, WFL, and SpecDef, respectively.

XSL•FO
RenderX

Figure 4. Anatomical heatmap marking absolute percentage change of metrics of interest (A, stdRMS; B, stdRMS; C, WFL; and D, SpecDef) for muscle groups (E) between control and a given compression as participants are positioned against the machine. Curved and flat compression types combined to wholistically represent the 4 mammogram positions. Max% of color bar indicated maximum percentage change considering both left and right components of muscle group. LCC: left craniocaudal; LMLO: left mediolateral; meanRMS: mean of the root mean square; RCC: right craniocaudal; RMLO: right mediolateral; stdRMS: standard deviation of the root mean square; SpecDef: spectral deformation; WFL: waveform length.



Before starting the study, participants were asked to report VAS pain ratings for their previous mammogram experiences. The results were between 1 and 9 and had a mean of 5.6 and an SD of 2.1. They again reported a VAS pain rating after finishing the study with respect to the simulated procedure. The results were between 1 and 5 and had a mean of 2.2 and an SD of 1.2. In the second report, participants were also requested to indicate which locations were the most uncomfortable during the study, if any. As indicated in Table 4, the breast was the most often

XSL•FO
RenderX

reported location for discomfort, followed by the shoulder and the neck. These reports are summarized in Table 4.

Table . Reported VAS[a] pain ratings[b].

Location	Number of reports
Breast	10
Shoulder	7
Neck	2
Leg	3
Feet	2
None	1
Overall VAS pain reported, mean (SD)	2.2 (1.2)

[a]VAS: visual analog scale.

[b]Study sample size (N=25).

Using the same standards but varying the segments, we used statistical comparisons to explore the effects of compression angle as previously discussed. These comparisons demonstrated no notable trends in significance. Similarly, no significance was observed for any metric between the large and small room controls.

Discussion

This study proposed a multimetric analysis of sEMG activity across 14 different muscles as a methodology to analyze women's muscular response during a simulated mammogram. The results allowed for a more complete understanding of the machine-patient dynamics and have the potential to further clarify the manifestation of pain or stress relevant to the procedure. To date, this is one of the most complete and thorough studies of EMG activity during mammogram screening.

To arrive at a nuanced understanding of the muscle's response, 10 EMG-based metrics were used to characterize the signal activity. When compared with related studies, the additional metrics allowed for greater resolution in classifying muscle responses. Out of the 10 metrics, the following 6 frequently indicated differences between the relaxed and stimulated states: meanRMS, stdRMS, medianFreq, meanFreq, WFL, and SpecDef. Out of those 6, meanRMS, stdRMS, medianFreq, and SpecDef showed large differences from one another, with each capturing a distinct characteristic of the signal behavior. In showing greater activation among muscles, these metrics can be reliably used to quantify muscle response throughout the procedure. This reliability is similarly represented in related studies in which greater activation of these metrics was related to greater pain and stress [16,23,24]. Uchiyama et al [20] implemented integrated electromyography, which is a metric that is closely related to meanRMS, and demonstrated similar muscle reactivity when quantifying trapezius, bicep, sternocleidomastoid, and gastrocnemius activation. No prior studies have examined changes in muscle activity trends during mammograms using sEMG recordings and displayed the trends of these 10 metrics. It is important to note that the overall magnitude of percentage change is relative to each muscle so that some muscles may have smaller variations but still represent significant activation.

Analysis of these metrics provides detailed insights into the demanding and stressful positioning required by mammograms. Specifically, the high sensitivity among many of the metrics in the deltoid, teres major, and infraspinatus can be attributed to the strained torsion of the upper torso. Although there is activation relative to the control, the absence of true significance in the external obliques further illustrates this point. The lower torso, including the obliques and abdomen, would normally twist with the upper body to relieve such strain, but mammogram positioning restricts waist and trunk torsion, inducing an uncomfortable, constricted hold. Similarly, in keeping the chest parallel with the imaging table, the patient extends the arm corresponding to the breast being imaged. This position further engages the shoulder or the upper body region and increases the overall demand on the static muscles during imaging. When compared with their baseline, the dominant activation of the deltoid muscles indicates its role as a positioning muscle and a likely source of pain, stress, or fatigue in women during breast compressions. The combined activation of the trapezius upper fibers and the deltoid provides insights into the patient's lower neck rotation and arm positioning due to the interaction with the face shield. The meanRMS and stdRMS of the associated neck muscles corresponding to the breast being compressed indicated differences for most of the respective compressions. This can be attributed to the engagement of the patient's arm with the side of the machine as well as the twisting of the head to accommodate the obtrusive face shield.

While visual observation of the body can begin to qualify the described positioning, the complete sEMG analysis illustrates the intensity of the strain that the procedure induces. There is a significant contrast between the involved muscles of the upper extremity and the limited activity of torso muscles. The torso muscles showed no interactions supposed by the imaging table pressing on the breastbone and rib. The data also corroborate the pain locations reported by participants who noted that, after the breast, the shoulders (deltoid, infraspinatus, and teres major) were the most uncomfortable body region during compressions.

XSL•FO
RenderX

A multimetric approach may lead to a more sensitive model that can quantify complex responses involving multiple muscle contributors of varied sizes and locations. This can be seen when comparing the responses of the trapezius upper fiber, deltoid, infraspinatus, and teres major muscle groups. The meanRMS and stdRMS show high activation among all 4 groups (8 muscles). However, out of these, only the deltoid and the infraspinatus muscle groups (4 muscles) show moderate activation for other metrics including slope sign change count, zero crossing count, meanFreq, medianFreq, and DP. This finding could be indicative of a more intense response on behalf of these muscles in comparison with the others. Similarly, it is possible that the additional metrics capture characteristics unique to each muscle that would otherwise be lost if only meanRMS was observed (one of the most common sEMG metrics applied in practice). The study results illustrated this likelihood in a few cases. One such case considers the mean and median frequency response, metrics that are frequently implemented as measures of fatigue [23]. In the results presented, we see cases in which a significant change of these metrics occurs in the deltoid and the infraspinatus but not in the other muscles. This may indicate that the right and left deltoid and the right and left infraspinatus are exposed to more fatigue-inducing strain arising from the type of positioning, repetitive muscle involvement, or some other factor intrinsic to the procedure.

The robustness and sensitivity of the proposed measurement methodology are further strengthened by evidence that they can distinguish lateral variance in compressions. Figure 3 reveals that the metrics that frequently show sensitivity only do so in roughly half of the total compressions. This lateral bias originates from the idea that the activation of a muscle depends on which breast is being imaged. The breast being imaged in turn dictates the location of the muscular tension in the upper torso and neck muscles, isolating the strain to one side in this group of muscles. For example, in the case of the LMLO, the 5 muscle pairs detected a difference more frequently for left breast compressions (flat paddle 8/8 instances, curved paddle 5/8 instances), whereas the RMLO noted a difference mostly for right breast compressions (flat paddle 6/8 instances, curved paddle 5/8 instances) as seen in Table 3. Similar lateral patterns were observed by Uchiyama et al [20]. While these patterns support the sensitivity of the measurement approach, the ability to detect more subtle differences such as those between paddle types may be limited. A larger sample size would likely be required to resolve smaller effect sizes that distinguish paddle-related influences.

While the study relied on a single overall VAS score, which provided meaningful context by identifying common regions of discomfort consistent with our sEMG findings, this format limited direct statistical correlation. Future studies incorporating VAS ratings after each compression would allow a more precise linkage between subjective and objective indicators for complete biomarker development. It is important to note that the reported pain score during the simulated mammogram was lower than the score the participants assigned to their previous experiences. Human ability to recall pain is surprisingly accurate [25]; therefore, the difference in pain is largely attributed to the reduced compressive force used in the study and is not a result of reporting inaccuracies. Limited but sufficient force was applied during the study so that a constant force reading on the machine was achieved. The participant was in a proper, realistic postural positioning without inducing unnecessary compression. As the study did not take any image of the breast, the compressive force applied did not reach the intensities patients normally experience. The standard force ranges from 100 N to 140 N [26], whereas the forces implemented in the study range from 18.25 N to 71.60 N. It is likely that had our study applied the same degree of pressure required by common procedures, the recorded pain metrics would have been more comparable. It can be realistically hypothesized that the painful stimulus of complete breast compression would likely induce additional muscle tension and a further elevated internal stress response. Consequently, the proposed methodology may detect an even greater biomechanical response via EMG mapping. The procedure also maintained the same order of compressions, flat paddle followed by curved paddle, which may have introduced a procedural bias. Given the back-to-back nature of the compressions, the later compressions could have captured patterns corresponding to extended screening, including muscle fatigue or learning results. The influence of this effect was hopefully limited by comparing the muscle activation during compression with flat and curved paddles to the initial C2 and intermittent C3, respectively.

While the proposed methodology was not able to distinguish between paddle types or compression angles, it was able to accurately capture the positional demands associated with each of the 4 different compressions. Supplemental research is needed to detect the effects of paddle design variations through EMG. Future studies incorporating muscles more directly involved in angular positioning may better capture biomechanical differences between vertical and angled compressions. Additionally, angle-dependent effects may exist but with smaller effect sizes than currently observed, suggesting that a larger sample may be required to achieve sufficient statistical power to detect these effects. It is possible that advanced machine learning approaches could be applied to determine the effects of paddle type and other subtle machine design variations on patient discomfort. Likewise, further decomposing the measurement windows could facilitate the exploration of muscle fatigue trends not only across compressions, as examined in this protocol, but also within each individual imaging period. Such within-compression analyses could reveal shorter-term fatigue patterns that complement the broader longitudinal trends. These complex and potentially hidden relationships underscore the importance of a methodology that integrates multiple metrics across several muscles.

Future work should also expand the study population to improve external validity. Including participants with a broader range of ages, BMIs, physical health, breast densities, and demographic backgrounds would help clarify how these factors interact with muscular demand and discomfort during compression. This expanded population would help ensure that findings generalize across diverse patient groups.

As long as the routine mammogram remains a painful and stressful experience, it is crucial to understand the physiological toll the procedure has on patients. This study applied a robust,

XSL•FO
RenderX

multimetric sEMG analysis across multiple muscles to objectively characterize the physical impacts of the mammogram procedure. While subjective pain reports provide valuable patient perspectives, EMG offers critical advantages by capturing unconscious muscular compensation patterns, precise temporal responses during specific procedure phases, and quantifiable biomechanical data that patients cannot accurately self-report. By identifying the relative simultaneous activation and inactivation of specific muscles, this methodology can highlight the unique, painful demands of the procedure. In the present analysis, the deltoid and infraspinatus muscles emerged as most consistently responsive across the majority of time domain and frequency domain metrics. This finding suggests their strong potential as candidates for sEMG-based biomarkers of procedural strain. Metrics such as RMS, medianFreq, and WFL, in particular, showed robust sensitivity to compression-specific positional demands. These muscle-metric combinations, therefore, represent promising targets for future refinement of objective indicators of discomfort or biomechanical load during mammography. This metric insight, along with other takeaways on sensitive muscles, provides promising results for potentially highlighting a focus point for improvements to the mammogram procedure that would be difficult to isolate through patient feedback alone. By analyzing a wide range of metrics, the methodology can potentially provide objective pain and stress measures such as those developed in similar studies involving the development of electrophysiological biomarkers [27]. The study methodology demonstrated the sensitivity of sEMG to capture mammogram-specific muscular trends during the compressions, revealing human-machine interaction patterns that inform targeted design modifications. These objective physiological data complement subjective feedback by providing quantifiable, reproducible metrics for assessing proposed improvements to the device design and imaging protocols. Future work should confirm reproducibility through larger cohorts, expanded muscle sets, and test-retest reliability. Relating key muscle-metric patterns to patient discomfort will help establish predictive validity and support the development of sEMG-based biomarkers for mammography-induced strain.

Although women may be reluctantly willing to endure mammograms as a necessary evil, there are many good reasons to better understand and improve the procedure. First, there is the potential to enrich the quality of the medical assessment, including both the results and the patient experience. Second, improvements to the patient experience can go a long way in promoting more widespread application of the technology. Furthermore, the comprehensive methodology proposed lays significant groundwork for future research in machine-human dynamics both in and outside of the health care field. With increased human exposure to novel technologies that modify body posture and muscular kinetics, it is crucial to understand the physiological ramifications in order to allow for their safe and effective application.

Acknowledgments

The authors are grateful to all participants for their time and engagement in the study. They thank Hadleigh A. Eastman, Riley Q. McNaboe, Sarah F. Myrick, Raghavi Shanmuganathan, and Alexi Switz for assistance with recruitment and data collection. The authors also acknowledge the mammography equipment assistance provided by Hologic, Inc and technical sensor support provided by Delsys, Inc throughout the study. Similarly, the authors extend their gratitude to the DeLuca Foundation grant for their financial assistance.

Funding

This work was supported by Hologic Inc (grant AG200867) and the DeLuca Foundation (grant 240457).

Authors' Contributions

Conceptualization and design: KG-P
Methodology: KG-P, RM
Investigation: KG-P, RM, HP-Q
Formal analysis: RM, HP-Q
Project administration: KG-P
Resources: KG-P
Supervision: KG-P
Funding acquisition: KG-P
Writing – original draft: RM
Writing – review & editing: KG-P, RM, HP-Q

Conflicts of Interest

None declared.

References

XSL•FO
RenderX

1. Kalaf JM. Mammography: a history of success and scientific enthusiasm. Radiol Bras 2014;47(4):VII-VIII. [doi: 10.1590/0100-3984.2014.47.4e2] [Medline: 25741098]
2. Grimm LJ, Avery CS, Hendrick E, Baker JA. Benefits and risks of mammography screening in women ages 40 to 49 years. J Prim Care Community Health 2022;13:21501327211058322. [doi: 10.1177/21501327211058322] [Medline: 35068237]
3. Baskin-Smith J, Miaskowski C, Dibble SL, Weekes D, Nielsen BB. Perceptions of the mammography experience. Cancer Nurs 1995 Feb;18(1):47-52. [Medline: 7866976]
4. Morris N. When health means suffering: mammograms, pain and compassionate care. Eur J Cancer Care (Engl) 2015 Jul;24(4):483-492. [doi: 10.1111/ecc.12272] [Medline: 25521596]
5. De Luca CJ. The use of surface electromyography in biomechanics. J Appl Biomech 1997;13(2):135-163. [doi: 10.1123/jab.13.2.135]
6. Shafti A, Lazpita BU, Elhage O, Wurdemann HA, Althoefer K. Analysis of comfort and ergonomics for clinical work environments. Annu Int Conf IEEE Eng Med Biol Soc 2016 Aug;2016:1894-1897. [doi: 10.1109/EMBC.2016.7591091] [Medline: 28268697]
7. Kobylińska D, Lewczuk K, Wizła M, et al. Effectiveness of emotion regulation strategies measured by self-report and EMG as a result of strategy used, negative emotion strength and participants' baseline HRV. Sci Rep 2023 Apr 17;13(1):6226. [doi: 10.1038/s41598-023-33032-2] [Medline: 37069211]
8. Kehri V, Ingle R, Awale R, Oimbe S. Techniques of EMG signal analysis and classification of neuromuscular diseases. Presented at: International Conference on Communication and Signal Processing 2016 (ICCASP 2016); Dec 26-27, 2016; Lonere, India p. 481-487. [doi: 10.2991/iccasp-16.2017.71]
9. Westgaard RH, Jensen T, Jansen C. EMG of neck and shoulder muscles: the relationship between muscle activity and muscle pain in occupational settings. In: Kumar S, editor. Electromyography in Ergonomics: Routledge; 1996:227-258. [doi: 10.1201/9780203758670-8] [Medline: 8854985]
10. Oliveira MI, Machado ARP, Chagas VGS, Granado TC, Pereira AA, Andrade AO. On the use of evoked potentials for quantification of pain. Annu Int Conf IEEE Eng Med Biol Soc 2012;2012:1578-1581. [doi: 10.1109/EMBC.2012.6346245]
11. Pourmohammadi S, Maleki A. Stress detection using ECG and EMG signals: a comprehensive study. Comput Methods Programs Biomed 2020 Sep;193:105482. [doi: 10.1016/j.cmpb.2020.105482] [Medline: 32408236]
12. Gielo-Perczak K, McNaboe R, Posada-Quintero H. Evaluation of patient stress during mammograms through surface electromyography analysis. In: Kalra J, editor. Healthcare and Medical Devices AHFE (2024) International Conference AHFE Open Access, Vol 130: AHFE International; 2024. [doi: 10.54941/ahfe1004842]
13. Gielo-Perczak K, McNaboe R, Posada-Quintero H. Using EMG to better understand patient pain and stress during routine mammography. Proc Hum Factors Ergon Soc Annu Meeting 2024 Aug;68(1):542-545. [doi: 10.1177/10711813241260394]
14. Steinmetz A, Claus A, Hodges PW, Jull GA. Neck muscle function in violinists/violists with and without neck pain. Clin Rheumatol 2016 Apr;35(4):1045-1051. [doi: 10.1007/s10067-015-3000-4] [Medline: 26175099]
15. de Oliveira CG, Nadal J. Back muscle EMG of helicopter pilots in flight: effects of fatigue, vibration, and posture. Aviat Space Environ Med 2004 Apr;75(4):317-322. [Medline: 15086121]
16. Wijsman J, Grundlehner B, Penders J, Hermens H. Trapezius muscle EMG as predictor of mental stress. Presented at: WH '10: Wireless Health; Oct 5-7, 2010; San Diego, CA p. 155-163. [doi: 10.1145/1921081.1921100]
17. Bhardwaj R, Parameswaran S, Balasubramanian V. Comparison of driver fatigue trend on simulator and on-road driving based on EMG correlation. Presented at: 2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS); Dec 1-2, 2018; Rupnagar, India p. 94-97. [doi: 10.1109/ICIINFS.2018.8721431]
18. Luttmann A, Sökeland J, Laurig W. Muscular strain and fatigue among urologists during transurethral resections using direct and monitor endoscopy. Eur Urol 1998;34(1):6-13. [doi: 10.1159/000019670] [Medline: 9676407]
19. Niu S, Jin K, Hu Z, Zhang C, Zhao D, Xing Y. Ergonomic evaluation between laparoscopic and robotic surgery based on EMG. Presented at: 2020 IEEE International Conference on Mechatronics and Automation (ICMA); Oct 13-16, 2020; Beijing, China p. 1756-1761. [doi: 10.1109/ICMA49215.2020.9233756]
20. Uchiyama M, Lee Y, Sadakata M, Sayama M, Tsai DY. Measurement of muscle activities for evaluating physical burden and pain during mammography positioning. Tohoku J Exp Med 2012 Sep;228(1):53-58. [doi: 10.1620/tjem.228.53] [Medline: 22976282]
21. 3Dimensions™ Mammography System. Hologic. URL: https://www.hologic.com/hologic-products/breast-health-solutions/3dimensions-mammography-system [accessed 2024-01-10]
22. Kim SH, Kim T, Park JC, Kim YH. Usefulness of hand grip strength to estimate other physical fitness parameters in older adults. Sci Rep 2022;12(1):17496. [doi: 10.1038/s41598-022-22477-6]
23. Allison GT, Fujiwara T. The relationship between EMG median frequency and low frequency band amplitude changes at different levels of muscle capacity. Clin Biomech (Bristol, Avon) 2002 Jul;17(6):464-469. [doi: 10.1016/S0268-0033(02)00033-5]
24. Phinyomark A, Thongpanja S, Hu H, et al. The usefulness of mean and median frequencies in electromyography analysis. In: Naik GR, editor. Computational Intelligence in Electromyography Analysis—A Perspective on Current Applications and Future Challenges: IntechOpen; 2012. [doi: 10.5772/50639]

XSL•FO
RenderX

25. Hunter M, Philips C, Rachman S. Memory for pain. Pain 1979 Feb;6(1):35-46. [doi: 10.1016/0304-3959(79)90138-6] [Medline: 424233]
26. Moshina N, Sagstad S, Sebuødegård S, et al. Breast compression and reported pain during mammographic screening. Radiography (Lond) 2020 May;26(2):133-139. [doi: 10.1016/j.radi.2019.10.003] [Medline: 32052779]
27. Cowen R, Stasiowska MK, Laycock H, Bantel C. Assessing pain objectively: the use of physiological markers. Anaesthesia 2015 Jul;70(7):828-847. [doi: 10.1111/anae.13018] [Medline: 25772783]

Abbreviations

C1: control recording in a large room (1341 square feet) without mammogram machine
C2: control recording in small room (171 square feet) with mammogram machine before the flat paddle compressions
C3: control recording in small room (171 square feet) with mammogram machine after the flat paddle compressions and before the curved paddle compressions
C_: compression with a curved paddle
DP: drop in power
EMG: electromyography
F_: compression with a flat paddle
LCC: left craniocaudal
LMLO: left mediolateral oblique
meanFreq: mean frequency
meanRMS: mean of root mean square
medianFreq: median frequency
RCC: right craniocaudal
RMLO: right mediolateral oblique
RMS: root mean square
sEMG: surface electromyography
SpecDef: spectral deformation
stdRMS: standard deviation of the root mean square
VAS: visual analog scale
WFL: waveform length

Edited by D Gooch; submitted 11.Sep.2025; peer-reviewed by M Radicke, Z Cheng; revised version received 09.Dec.2025; accepted 08.Jan.2026; published 06.Mar.2026.

Please cite as:
Gielo-Perczak K, McNaboe R, Posada-Quintero H
Surface Electromyography as a Method for Characterizing Mammogram Discomfort: Cross-Sectional Questionnaire Study of Procedural Stress
JMIR Hum Factors 2026;13:e83971
URL: https://humanfactors.jmir.org/2026/1/e83971
doi:10.2196/83971

© Krystyna Gielo-Perczak, Riley McNaboe, Hugo Posada-Quintero. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 6.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Development of a Data-Enabled Mixed Method for Designing for Patients From a Human-Centered Design Perspective: Explorative Case Study

Yingtao Sun[1], MSc; Jiwon Jung[1,2,3], PhD

[1]Department of Design, Organisation and Strategy, Faculty of Industrial Design Engineering, Delft University of Technology, Landbergstraat 15, Delft, South Holland, The Netherlands

[2]Department of Surgery, Erasmus University Medical Center, Rotterdam, South Holland, The Netherlands

[3]University College, Korea University, Seoul, Republic of Korea

Corresponding Author:
Jiwon Jung, PhD
Department of Design, Organisation and Strategy, Faculty of Industrial Design Engineering, Delft University of Technology, Landbergstraat 15, Delft, South Holland, The Netherlands

Abstract

Background: In the early stage of human-centered design (HCD), qualitative and generative methods are commonly used to explore patients' contexts and needs, emphasizing active patient involvement to ensure that design insights accurately reflect real experiences and enhance both design effectiveness and patient empowerment. However, certain challenges arise in the early stage of the HCD process, including (1) high vulnerability of patient participants, (2) less diverse and representative patient groups due to recruitment challenges, and (3) insufficient problem framing across diverse patient experiences.

Objective: To address these challenges while embracing the system-level HCD perspective, we propose a data-enabled mixed method combining large-scale patient digital research (module A) and in-depth patient engagement research (module B). This paper explores the feasibility and potential value of this mixed method in addressing the identified challenges through a case study.

Methods: In module A, we analyzed a large-scale dataset of online forum posts and validated the extracted topics with medical experts to create a patient community journey map through cocreation sessions. Guided by these findings, module B involved a diary study using a sensitizing paper prototype and semistructured follow-up interviews with 4 patients.

Results: In module A, 37 topics and 10 upper clusters were summarized from 212,107 online posts, revealing that topics associated with the home context exhibited a higher density of emotional content than those related to the hospital context. Patients placed more emphasis on social and mental health during the follow-up stage than in the diagnosis and treatment phases. This shift reveals a gap in current remote monitoring systems, which focus mainly on physical health. Addressing this identified gap, we developed a prototype for use in the module B diary study involving 4 patients. Patients responded positively to the prototype, noting that remote monitoring incorporating social and mental well-being could help them better understand themselves, enhance self-awareness, and improve communication with their physicians. Overall, this study explores the preliminary value of this mixed method in effectively reframing design problems and deeply contextualizing patient needs.

Conclusions: This study provides preliminary evidence supporting the potential of integrating large-scale digital patient research (module A) with in-depth patient engagement (module B) during the early stages of human-centered health care design. The core strength of this mixed method approach lies in its ability to facilitate problem reframing and cultivate a deeper sense of empathy and understanding of patient vulnerability prior to direct engagement. Simultaneously, it captures both the breadth and depth of patient perspectives, offering evidence-based insights that enhance the overall efficacy of user research. Further research across diverse medical contexts is essential to establish the generalizability of these findings.

(JMIR Hum Factors 2026;13:e87677) doi:10.2196/87677

KEYWORDS

human-centered design; mixed method; health care design; patient-centered design; data-enabled design; patient journey mapping; patient experience; colorectal cancer; remote patient monitoring

XSL•FO
RenderX

Introduction

Background

Human-centered design (HCD) plays an important role in designing for patients in health care in order to understand patient needs [1], improve patient experience [2], and enhance patient safety [3]. The focus is on empowering designers to understand human needs at a systemic level through empathetic communication and interaction with the individuals within the system [4].

The health care design process typically starts with exploring patients' needs using the HCD approach before formulating design solutions [5], specifically in the initial stage, such as the discover and define phases of the Double Diamond Model [6] and the fuzzy front end of the design development process [7]. To understand patient context, qualitative research methods are frequently used in the initial stage of the design process to perceive the context [5]. The specific methods proposed in such an explorative stage include both conventional methods, such as interviews and observations, and other innovative methods, such as generative techniques [8], which enable health care designers to deeply explore patients' needs [9] while also facilitating the examination of the complex interactions among various system components [10].

Central to these methods is involving patients as the end users, serving as a foundational principle for ensuring that insights derived from qualitative research accurately reflect patients' needs and experiences. For instance, Van Smoorenburg et al [11] asked 10 patients with type 2 diabetes mellitus to fill in sensitizing booklets before conducting interviews to prompt them to reflect on their experiences with self-management of diabetes. Researchers illustrate that the involvement of patients can not only increase the effectiveness and success of the design process [12] but also facilitate patients' knowledge-building and relationships [13].

Challenges in Human-Centered Health Care Design

However, implementing HCD and engaging patients in health care settings is more complex than in other domains, primarily due to the intricate nature of health care systems and patients' compromised health states [14,15]. In this paper, we explore the challenges of designing for patients in the initial stage of the design process, including high vulnerability of patient participants, less diverse and representative patient groups due to recruitment challenges, and insufficient problem framing across diverse patient experiences.

High Vulnerability of Patient Participants

Patients are often regarded as a vulnerable demographic, which necessitates careful ethical consideration when involving them in research. For instance, Das and Svanæs [1] noted that discussions about obesity may evoke stigma and shame, requiring designers to plan workshops and participant groupings sensitively to avoid offending others. Furthermore, obtaining medical study approval typically demands a lengthy preparation process, which can be partially attributed to the risk of patients recalling distressing experiences during participation [16]. However, despite the ethical need to protect patients from potential harm, designers are also expected to gain a deep and systematic understanding of their lived experiences to develop meaningful interventions. This requirement poses a notable challenge, as patient participation, despite its recognized value, can introduce burdens that hinder meaningful collaboration [17,18]. Furthermore, designers may not yet be familiar with the patient group in the early design stages, making it challenging to anticipate negative experiences or address such vulnerabilities effectively in advance.

Less Diverse and Representative Patient Groups Due to Recruitment Challenges

In the early exploratory phases of HCD, qualitative research methods such as user observations, interviews, and generative techniques are used to emphasize the necessity of continuous patient involvement throughout the entire design lifecycle, rather than limiting participation to a single stage [8,19]. However, these methods face certain dilemmas related to diversity and representativeness of patient groups, which have been criticized as a manifestation of sampling bias in HCD [17]. Specifically, researchers have attributed this challenge to limited sample sizes, difficulties in engaging representative end users, and insufficient participant diversity during patient recruitment, all of which raise concerns about the generalizability, validity, and reliability of the results [15,16,20]. For example, a study developing an intervention with a participatory design approach has reported difficulties in identifying appropriate patient groups and achieving adequate diversity among participants in terms of age, sex, and disease duration [21].

Insufficient Problem Framing Across Diverse Patient Experiences

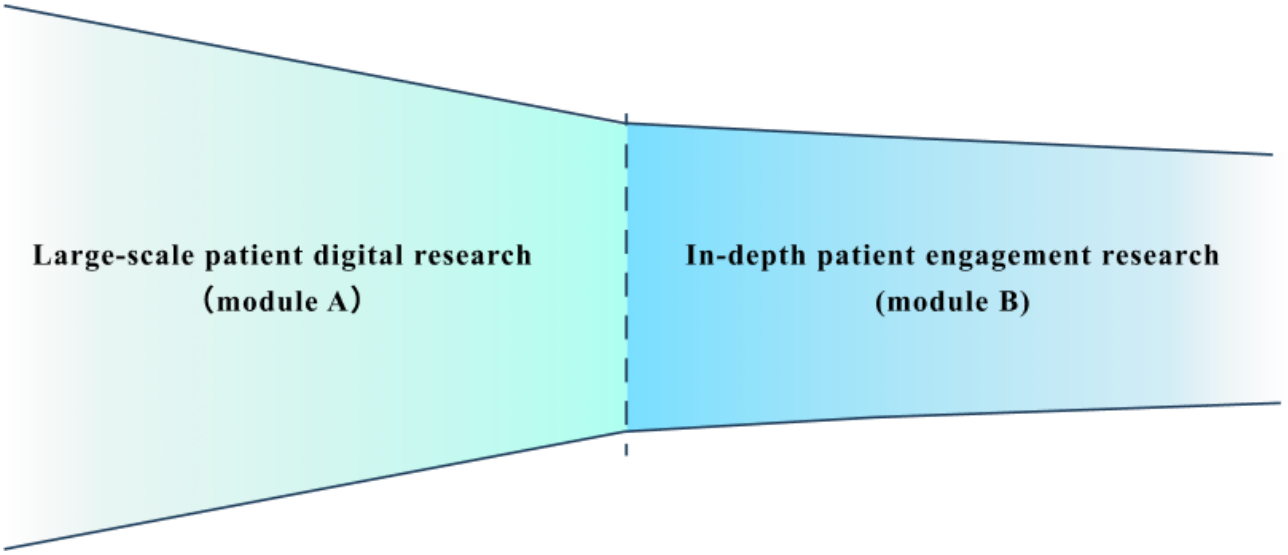
Problem framing is a foundational step in HCD, aiming to ensure that interventions address the right problem [22]. It requires understanding not only users' needs but also their experiences, values, behaviors, and environmental contexts from a system's perspective [8]. Nevertheless, the awareness of problem definition in current HCD practices within health care remains limited, with few studies attempting to reframe the initial problem based on diverse user perspectives [23]. More studies adopt a solution-driven rather than a problem-driven approach, generating numerous ideas while leaving the initial design problem insufficiently defined [5]. Moreover, many studies focus on the general principles of HCD rather than detailing specific processes, tools, or methods [24], thereby marginalizing the attention given to defining problems based on patient experiences. For example, some studies fail to clearly distinguish between the discover and define phases in the Double Diamond model [25]. Problem-driven design is essential for prioritizing interventions based on a deep understanding of the underlying problem [5,26]. When patient experiences are explored only within a preestablished problem scope and narrow aspects, rather than through iterative problem reframing, opportunities for creative and efficient solutions are limited [23].

XSL•FO
RenderX

Advancing the Exploration of Patient Experience With an HCD Perspective: Development of a Data-Enabled Mixed Method

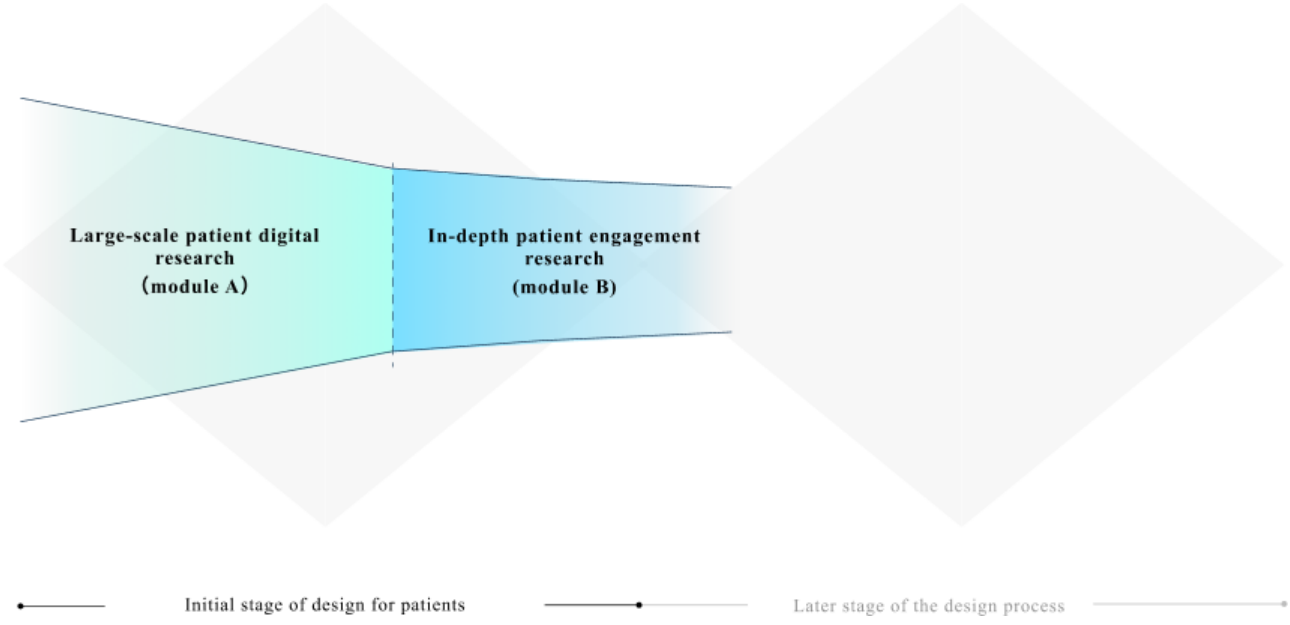
To deal with the challenges of designing for patients and achieve the goal of HCD in the early stage, we propose a data-enabled mixed method as illustrated in Figure 1, which prioritizes large-scale patient digital research (module A), subsequently followed by in-depth patient engagement research (module B). It is important to note that we currently consider this mixed method to be most valuable when adopted during the initial stages of the design process, such as the first diamond of the Double Diamond model [6] and the fuzzy front end of the design development process [7]. Figure 2 illustrates the stage in which this mixed method is applied using the Double Diamond model as an example.

Figure 1. The data-enabled mixed method for designing for patients in the initial stage of the design process.



In large-scale patient digital research, extensive datasets derived from digital patient data are used to generate meaningful insights. The potential data source can be digital patient platforms such as Facebook groups or patient online forums, which are emerging in patient research [27,28]. In this way, designers can indirectly conduct an in-depth analysis of the data with patients' perspectives and empathy for patients' experiences, which further facilitates resonance and the handling of sensitive issues when engaging with patients in the next stage. Patient engagement research in module B refers to conducting conventional qualitative or design research that requires the involvement of patients, such as interviews, diary studies, and focus groups. In other words, we suggest adding large-scale patient digital research to have a generalized understanding before patient engagement in the design process.

Figure 2. The Double Diamond model as an example illustrates the stage in which the data-enabled mixed method is applied.



XSL•FO
RenderX

About This Case

An explorative case study illustrates how this data-enabled mixed method was adopted to explore patient experience during the initial stage. The case provides preliminary indications of this method's potential benefits, including enabling health care designers to acquire rich knowledge that features patients' perspectives, cultivating an empathetic perspective toward the difficult medical context, facilitating designers to define research and design direction, and designing engaging research activities.

The initial goal of the case was to develop remote patient monitoring for patients with colorectal cancer (CRC) in the Netherlands, considering their quality of life during the follow-up phase. Within this context, this case investigates the care path of patients with CRC with a particular focus on understanding their experiences, especially in the follow-up phase, to lay the groundwork for the development of the remote patient-monitoring system. This study brings together a Netherlands-based interdisciplinary team with expertise in design, computer science, and CRC research.

CRC Background and the Theoretical Framework

CRC is the third most common cancer worldwide. It accounts for 10% of the global cancer incidence [29] and is expected to exceed 3 million by 2040 [30]. In the Netherlands, where the case study was conducted, patients who have undergone treatment and surgery for CRC typically enter a 5-year follow-up phase. This phase involves regular meetings between patients and health care professionals to monitor their recovery progress and provide necessary information and support.

During the follow-up phase, patients often experience long-term health issues including physical, social, and mental health. The World Health Organization has highlighted the importance of these 3 aspects in relation to health-related quality of life. These 3 aspects are related to each other and form the basis of a framework called "Three-dimensional Theoretical Framework of Health" (Figure 3) [31]. This framework illustrates that the intersection of the 3 aspects has an equally high impact on the quality of life of patients' health but is often overlooked. It is important to pay attention to all aspects of health of a patient with CRC in the follow-up phase to enhance the patient's quality of life.

XSL•FO
RenderX

Figure 3. Three-dimensional theoretical framework of health [31].

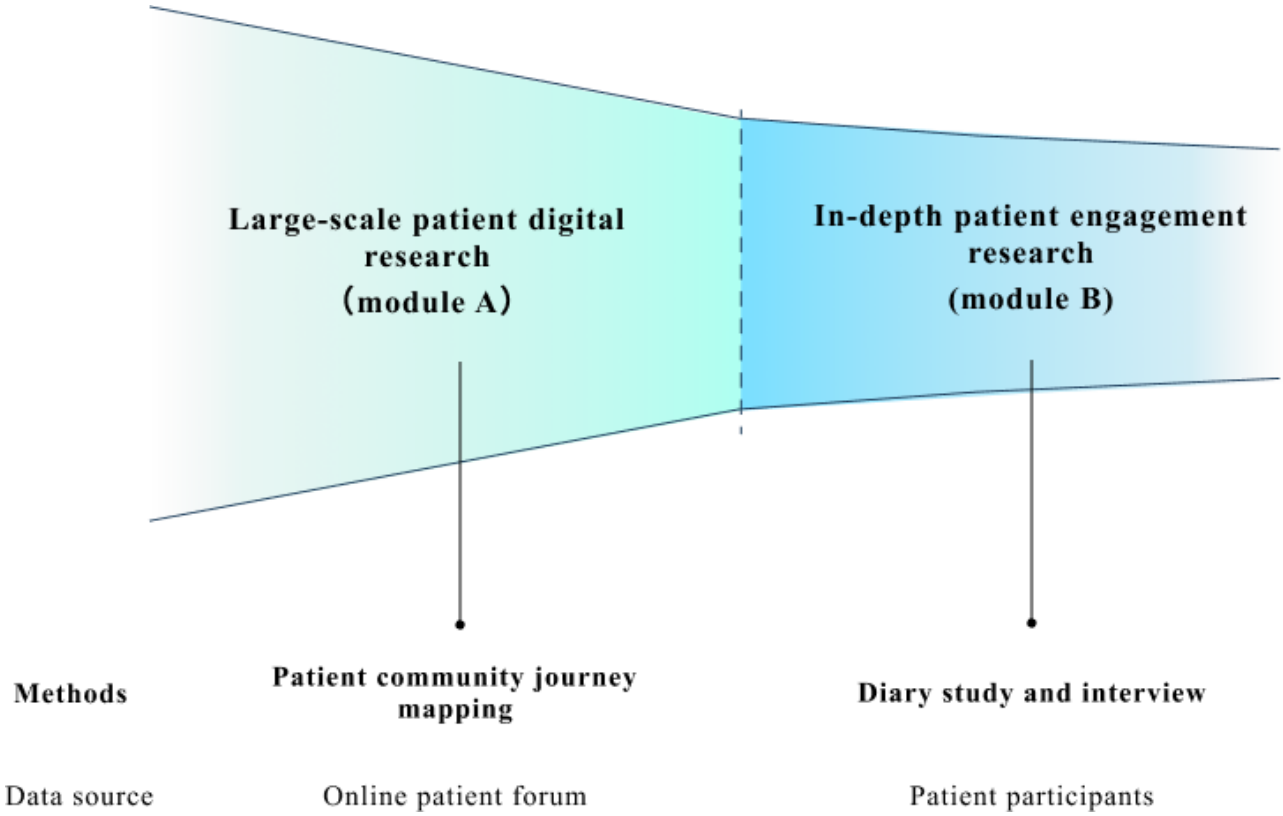
XSL•FO
RenderX

Methods

Study Design

Figure 4 outlines an overview of the methods used in this case. Module A adopts the patient community journey mapping method developed by Jung et al [32] through the collaboration between the designer and a computer scientist. This method applies natural language processing techniques to online patient posts, analyzing the text and visualizing the results as a journey map [32]. The goal of this method is to perceive the experiences of patients with CRC, reframe the research question, and gain an overview of the patient context. Based on the findings and perceptions from module A, we reframed the research question and then conducted a diary study and interviews involving patient participants to dive deep into the context in module B.

Figure 4. The data-enabled mixed method in the colorectal cancer case.



Module A: Patient Community Journey Map

Introduction of Patient Community Journey Map

The conventional development of the patient journey map mainly relies on the data collected from qualitative methods, such as interviews and ethnographic data [33]. Although the qualitative patient journey map reflects a person's entire journey through the health systems, it requires patient involvement for developing the map and restricts the sample size to at most a few dozen patients [32], bringing the 3 challenges previously mentioned to light. In this case, we adopt the data-enabled "patient community journey mapping" method as one of the examples that can address the 3 challenges and the limitations of the qualitative patient journey map. It takes advantage of natural language processing techniques, such as topic modeling, to process online patients' stories and further create a journey map with the results [32].

Data Collection and Analysis

This study uses the same dataset as our earlier work [34], which introduced the data collection and topic modeling that form an earlier version of module A. Building upon that early-stage exploration, we further investigate the combining of these large-scale digital insights (module A) with in-depth qualitative patient engagement (module B). Additionally, we extend the previous analysis of the patient community journey map by introducing the three-dimensional framework of health to contextualize the findings in module A.

To ensure privacy protection, all data were preprocessed by removing personal identifiers prior to analysis. A total of 212,107 forum posts related to CRC were collected from the publicly accessible Cancer Survivors Network and further preprocessed by applying lemmatization and Term Frequency–Inverse Document Frequency by a computer scientist. The computer scientist then used nonnegative matrix factorization for topic modeling. Because quantitative metrics, such as topic coherence and model stability, consistently declined and lacked clear insights, determining the optimal number of topics relied on rigorous human evaluation. Four researchers, including 2 designers, iteratively reviewed models with 40, 45, and 50 topics, ultimately selecting 50 initial topics as they provided the most diverse and distinctive clusters based on keyword and postexamination. Thirteen topics, such as platform use and expressions of gratitude, were then excluded as they were unrelated to patient experiences, leaving 37 topics for the final analysis.

XSL•FO
RenderX

To analyze the 37 topics, 1 designer reviewed the top 50 posts and top 20 keywords per topic. These were selected by the nonnegative matrix factorization algorithm for possessing the highest dominant topic scores and term weights, respectively, calculated using Term Frequency–Inverse Document Frequency metrics. The review of 50 posts was guided by data saturation: recurring narratives typically emerged after analyzing approximately 30 posts, and the additional 20 were reviewed to rigorously confirm and solidify the extracted themes. The 37 topics were then classified into 10 upper clusters.

After the initial interpretation, a 2-step validation process was implemented to ensure the rigor of the interpretation of the themes. First, the initial interpretation of the 37 topics conducted by the designer was reviewed and validated by a senior design researcher. Second, we subsequently organized cocreation sessions with a panel of local domain experts from the Netherlands. During the 90-minute session, the experts used printed 37 topic cards featuring the top 20 keywords and patient quotes to formulate independent interpretations. These interpretations were then compared with the designer's initial descriptions for validation. By actively absorbing the experts' insights and feedback, the team refined the topic descriptions and collaboratively positioned them onto a blank journey map to reflect the 37 topics within the Dutch health care context. This process also ensured that the findings derived from the US-based forum were relevant and contextualized to the Dutch health care system. The expert panel comprised 5 medical professionals: 2 oncological surgeons with 20 and 12 years of experience, respectively; an epidemiologist specializing in patient experience and quality of life with 15 years of expertise; and 2 MD-PhD candidates in surgical oncology with a research focus on CRC. Through these sessions, the expert panel reviewed and validated the identified topics and clusters, ultimately informing the patient community journey map.

After defining the themes, the designer introduced 2 dimensions to classify the topics. The first dimension was to judge whether the topics belonged to a hospital or home context. The second dimension was to decide whether the topics were related to physical, social, or mental health, according to the three-dimensional theoretical framework of health (Figure 3). Additionally, 2 interviews with CRC specialists were carried out to create the journey part of a patient community journey map, visually illustrating patients' experiences and key concerns.

Module B: Diary Study and Follow-Up Interview

Module Overview

In module A, the research question was reframed: rather than exploring how to develop remote monitoring of patients with CRC during the follow-up stage with a broad consideration of quality of life, the focus was narrowed to specifically supporting patients' social and mental health. Carrying this new research question forward, module B dives deep into the context of patients with CRC' by conducting a diary study and interviews on their social and mental health in the follow-up stage when they are at home. The purpose of the study is to understand patients' social and daily activities and explore patients' attitudes toward monitoring their activities and emotions.

Data Collection and Analysis

In module B, we conducted the diary study and interviewed 4 participants in their follow-up stage. In the diary study, the designer created a paper prototype to simulate remote data collection, aiming to capture patients' daily activities and emotional states. The content of the prototype was developed based on the 5W1H (What, Why, Who, When, Where, How) structure. By recording the details (what, where, when, who) and their reflections (why, how), patients capture a complete context of their daily lives. The process concludes with patients drawing an emotional curve for the day to visualize their mood trends.

In addition, to ensure that the diary study was closely grounded in actual patient experiences, the designer synthesized the 37 topics to create a patient-facing Contextual Prompt Card (Figure 5); for example, the eating activities from topic 17 were in the "What" section in the prompt card. Accompanying the 5W1H paper prototype, this card inspired patients and prevented cognitive overload. It functioned purely as a sensitizing tool, featuring open-ended "Other" options to encourage recording individualized experiences.

Four patients with CRC (Table 1) participated in the study through both online and offline recruitment. Participants were recruited via posters, online forums, and Facebook groups. The patients were invited to finish a 3- to 4-day diary study, followed by a 30-minute semistructured interview to share their findings and attitude.

Although serving as a sensitizing paper prototype rather than independent qualitative data for systematic coding, the diary cards revealed intuitive insights that provided core background for the interviews and acted as observational corroboration in the findings. The formal analytical procedure was exclusively applied to the data collected from the 30-minute semistructured follow-up interviews. We used the thematic analysis consisting of 4 iterative phases. First, descriptive coding was applied to segment the interview text and assign unique codes capturing patients' direct feedback and daily reflections. Next, through subcategorization, these codes were grouped into specific categories aligned with the interview questions. During the clustering phase, similar codes within each category were aggregated into narrative topics. Finally, theme determination was conducted to categorize related topics and label the final overarching themes.

XSL•FO
RenderX

Figure 5. The contextual prompt card was used for patients' inspiration during the diary study.

XSL•FO
RenderX

Table . Demographic and characteristics of the 4 patient participants with colorectal cancer involved in the diary study and follow-up interviews (module B).

Participant	Age (years)	Year of diagnosis	Number of finished days for diary study	Online or offline interview
Participant A	61	2014	2	Online
Participant B	53	2021	4	Online
Participant C	45	2019	3	Online
Participant D	50	2021	3	Offline (designer visited patient's home)

Ethical Considerations

The study was approved by the Human Research Ethics Committee of Delft University of Technology (reference number 2596). Participation was entirely voluntary, with participants retaining the right to withdraw their consent at any time. To guarantee confidentiality and privacy, all collected data were fully anonymized. Participants were assigned unique identification codes, and these deidentified records were stored on secure university servers, accessible exclusively to authorized research members.

Results

Module A Results

Overview of Module A Results

Table 2 shows the themes of the 37 topics and 10 upper clusters, and Figure 6 shows the patient community journey map. These topics reflect the broad range of discussions among patients, covering themes from medications and treatments to family life.

XSL•FO
RenderX

Table . Overview of the 10 upper clusters and 37 topics identified from the large-scale online forum posts (module A) of patient with colorectal cancer (adapted from Voigt et al [34], which is published under Creative Commons Attribution 4.0 International License [35]), with topic numbers added for this analysis.

Cluster number, cluster theme, and topic number	Number of posts under the topic	Topic theme
Cluster 1: Experience around medical professionals' opinion		
39	5043	Doubts about treatment opinions from medical professionals
45	4713	Suggestion to look for a second medical opinion
Cluster 2: Understanding treatment including alternatives and adjunctive therapy		
30	7854	Patients share their research about alternative therapy options from websites and articles
10	4759	Making treatment decisions for the future with regard to the best outcome and path
33	8768	Listing type, side effect, regimen, and effectiveness of drugs
14	3610	Sharing experience on using Traditional Chinese Medicine to manage health
21	2854	Sharing experiences and recommendations for supplements and medication
48	3732	Sharing information and experiences regarding clinical trials
Cluster 3: Surgery experience		
25	4979	Sharing experience around radio frequency ablation for the liver
23	4561	Sharing negative experience about resection in liver and lung
Cluster 4: Experience regarding the test results (including being worried and confused)		
15	3873	Sharing negative emotions and experiences to live with CRC[a]: overwhelmed, confused, and scared, especially for the tests and screenings
9	3210	Being worried about upcoming scans and results
18	4293	Sharing their outcomes (clear or not) from scanning and caring about how frequently they scan
49	5048	Describing a stressful experience in a blood test; worry about the numbers in the result
8	13,560	Being worried and confused about odd scan results in liver, lung, and lymph
Cluster 5: Experience with side effects		
42	8341	Sharing their experience on managing the side effects of treatment
37	5122	Negative feelings of hair loss due to cancer treatments
34	6414	Feeling uncomfortable due to the obstipation
32	6275	Experiencing pain and numbness due to neuropathy
Cluster 6: Confusion with insurance		
38	7912	Confusion about insurance coverage
Cluster 7: Experience during recovery phase		
27	5010	Describing negative experience of repeated visits to the hospital

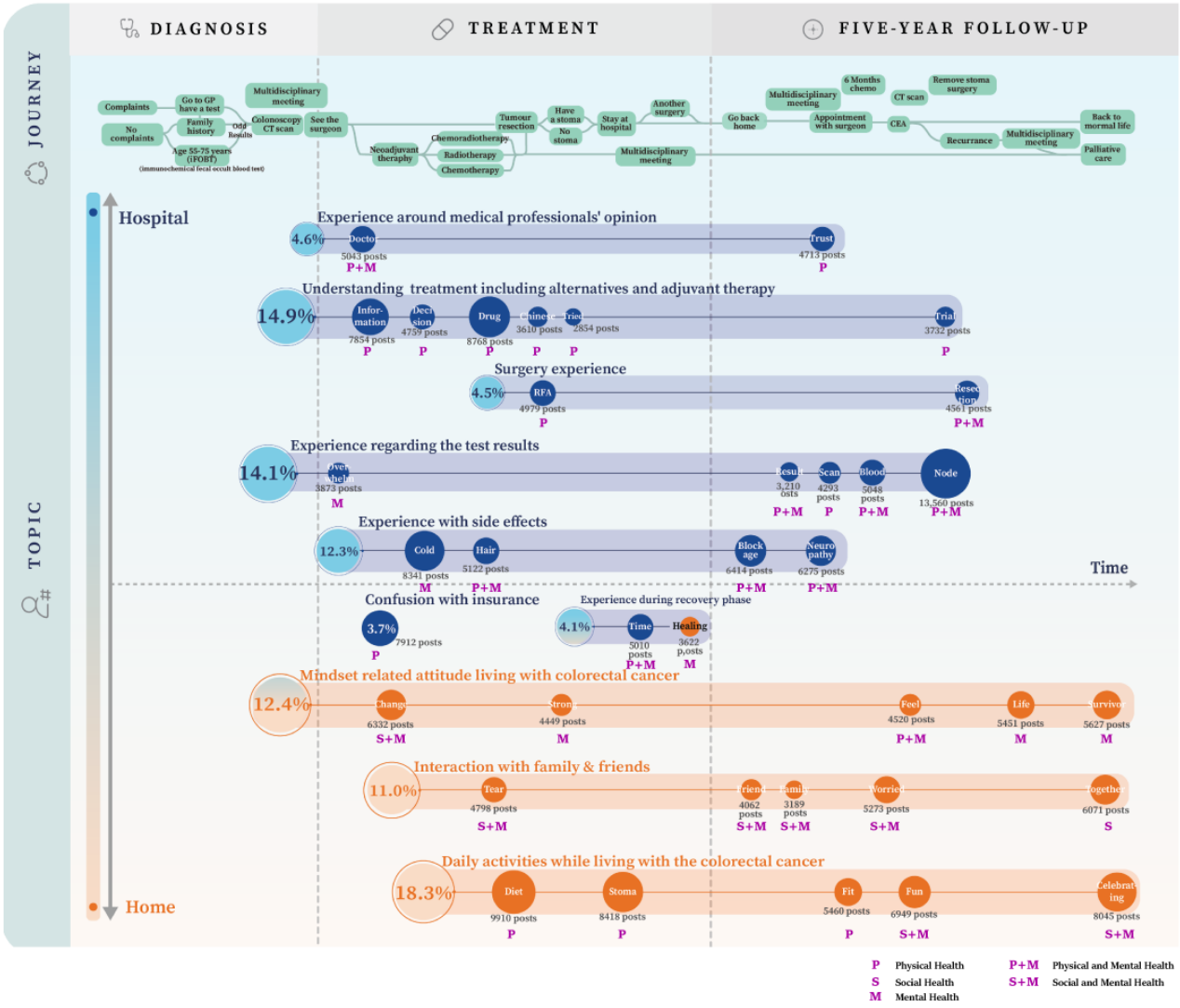
XSL•FO
RenderX

Cluster number, cluster theme, and topic number	Number of posts under the topic	Topic theme
20	3622	Staying positive and making life adjustments to their cancer circumstances during the recovery phase
Cluster 8: Mindset-related attitude living with CRC		
0	6332	Difficulties to adjust and adapt to their lives with CRC
41	4449	Sharing how patients can be resilient and positive
24	4520	Sharing their feelings: sick, tired, weak, and bad
46	5451	Sharing their positive philosophical thoughts about living with cancer
11	5627	Survivors sharing their attitudes toward living with cancer along with survival rate
Cluster 9: Interaction with family and friends		
40	4798	Family members' emotional struggle about having a patient with cancer in their family
47	4062	Experience regarding relationships with friends while having cancer
28	3189	Sharing changes in their family relationship due to cancer journey
43	5273	Being worried about their family members and seeking information on family history (eg, genetic testing)
13	6071	Arguing the importance to spend moments with family members during the cancer journey
Cluster 10: Daily activities while living with the CRC		
17	9910	Sharing suggestions on diets focused on balanced meals and healthy alternatives
22	8418	Suggesting how to take good care of a stoma
35	5460	Sharing experience on being fit again, caring about weight control
31	6949	Sharing ways to stay in a positive mood through planning for distractions
3	8045	Celebrating anniversaries and birthdays for patients as a big milestone of their lives

[a]CRC: colorectal cancer.

XSL•FO
RenderX

Figure 6. Community journey map of patient with colorectal cancer, with 37 topics classified into 2 dimensions: home and hospital dimension, and three-dimensional framework of health (adapted from Voigt et al [34], which is published under Creative Commons Attribution 4.0 International License [35]) by adding the three-dimensional framework of health [31]. CEA: carcinoembryonic antigen; CT: computed tomography; GP: general practitioner; iFOBT: immunochemical fecal occult blood test; RFA: radio frequency ablation.



Some topics highlight common concerns among patients with cancer. For example, cluster 1 revolves around sharing apprehensions about doctors' recommendations and whether seeking a second opinion is necessary. Additionally, under cluster 5, topic 37 indicates the negative feelings of hair loss due to the treatment. Still, some topics go beyond the designer's initial expectations. To give an example, topic 43 reveals patients' concerns about the genetic inheritability of cancer and whether their children should undergo genetic testing.

The designer uncovered new insights through reading the posts. Among others, topic 14 indicates that a significant number of patients use Traditional Chinese Medicine to manage their health, and topic 38 reflects patients' uncertainty about the scope of insurance reimbursement. Additionally, this sensitizing process fostered moments of empathy for the designer. For instance, summarizing topic 40 involved reading numerous posts from family members of patients with cancer, sharing their emotional struggles in coping with a loved one's illness. Similarly, topic 0 revealed the challenges patients face in adapting to life after being diagnosed.

Through the process of immersing in the posts, the designer gradually perceived the diverse contexts in which these topics occur. Some topics are situated in the hospital context, while others unfold entirely at home. Moreover, these topics span across various health-related categories, including physical, mental, and social health related to the three-dimensional theoretical framework of health [31]. Therefore, we introduced 2 dimensions to the 37 topics, classifying them based on whether they belong to the hospital or home context, as well as by the three-dimensional theoretical framework of health. These topics were visualized into the patient community journey map (Figure 6).

Almost all aspects of health are represented on the map when the topics are examined through the lens of the three-dimensional theoretical framework of health. This indicates that patients with CRC face a wide range of health issues, encompassing most of the categories outlined in Figure 3. The only category not represented is "role limitations due to health," corresponding to number 7 in the framework shown in Figure 3, which includes aspects such as sexual functioning.

XSL•FO
RenderX

We assume that this is because topics related to sexual health are highly private, and consequently, patients may be reluctant to discuss them on public social forums. Nevertheless, this does not imply that these issues are nonexistent but still need attention, as several studies highlight significant challenges in the sexual lives of patients with CRC, such as sexual dysfunction [36-39].

Topics under the hospital context are mostly related to physical health, pathological knowledge, medical procedures, and medical systems. Conversely, topics within the home context primarily revolve around daily life, social interactions, and personal reflections. Patients tend to approach discussions within the hospital context in an objective, rational, and focused manner, while discussions within the home context tend to be more emotionally charged and subjective in their descriptions, aligning with our previous research [34].

In addition, a closer reading of the topics reveals a notable trend: patients' mental health is frequently influenced by their physical condition. For example, cluster 5 expresses that side effects often cause a lot of physical pain and daily inconveniences, which can be quite stressful for the patient and further affect mental health. This trend is also observed between social and mental health, usually occurring in the home context. For instance, as illustrated in topic 28, patients may experience strong emotional responses to shifts in their social relationships, largely depending on whether these changes are beneficial to them.

Turning to the right side of Figure 6, namely, in the follow-up stage, there is a significant increase in the number of topics in the home context compared with the number of topics associated with the hospital context. Topics such as sharing experience on being fit again (topic 35) appear in this area, demonstrating that patients pay more attention to life in the follow-up stage. As patients become more accustomed to living with their diagnosis, their acute clinical anxiety typically decreases, allowing them to redirect their energy toward social interactions and daily activities. Therefore, social- and mental-related topics are also emerging actively at this stage. Patients gradually become aware of changes occurring in their surroundings. For example, they notice changes in their relationships with friends (topic 47) and illustrate the importance of spending moments with family members (topic 13). These changes are frequently attributed to the impact of cancer and can also lead to emotional fluctuations in the patient. The patients even began to deeply reflect on life, such as being more resilient in the face of fear and death, and share their thoughts with their peers, as observed in topic 46.

By contrast, during the diagnosis and treatment phase, patients frequently prioritize attending to their physical well-being, such as being worried about upcoming scans and results (topic 9) and numbers in the result of a blood test (topic 49).

Module A Summary

In this case, patient community journey mapping acts as an example to illustrate how large-scale patient research of module A can be conducted. We applied natural language processing techniques to analyze 212,107 posts from patient forums and categorized them into 37 topics. These topics were further classified into 2 dimensions and mapped onto the patient care path.

Since the designer previously lacked knowledge about CRC, this method provides an empathetic and effective starting point, allowing the designer to perceive the context by reading numerous patients' stories and posts. This indirect dialogue not only contributes empathetic and generalized context to the designer but also allows the designer to have adequate preparation for the patient engagement research in module B.

In module A, the designer not only gained insights into patient experiences throughout the entire care journey but also reframed the research question as follows: How can remote patient monitoring be developed for the follow-up stage with a focus on patients' social and mental health? This insight emerges from the patient community journey map, echoing that current remote patient monitoring primarily focuses on physical health aspects, such as tracking vital signs while paying relatively little attention to social and mental health. A gap therefore exists, potentially linked to the limited awareness of patients' social and mental health outside the clinical setting. As a result, it is important to build a remote patient-monitoring system related to social and mental health.

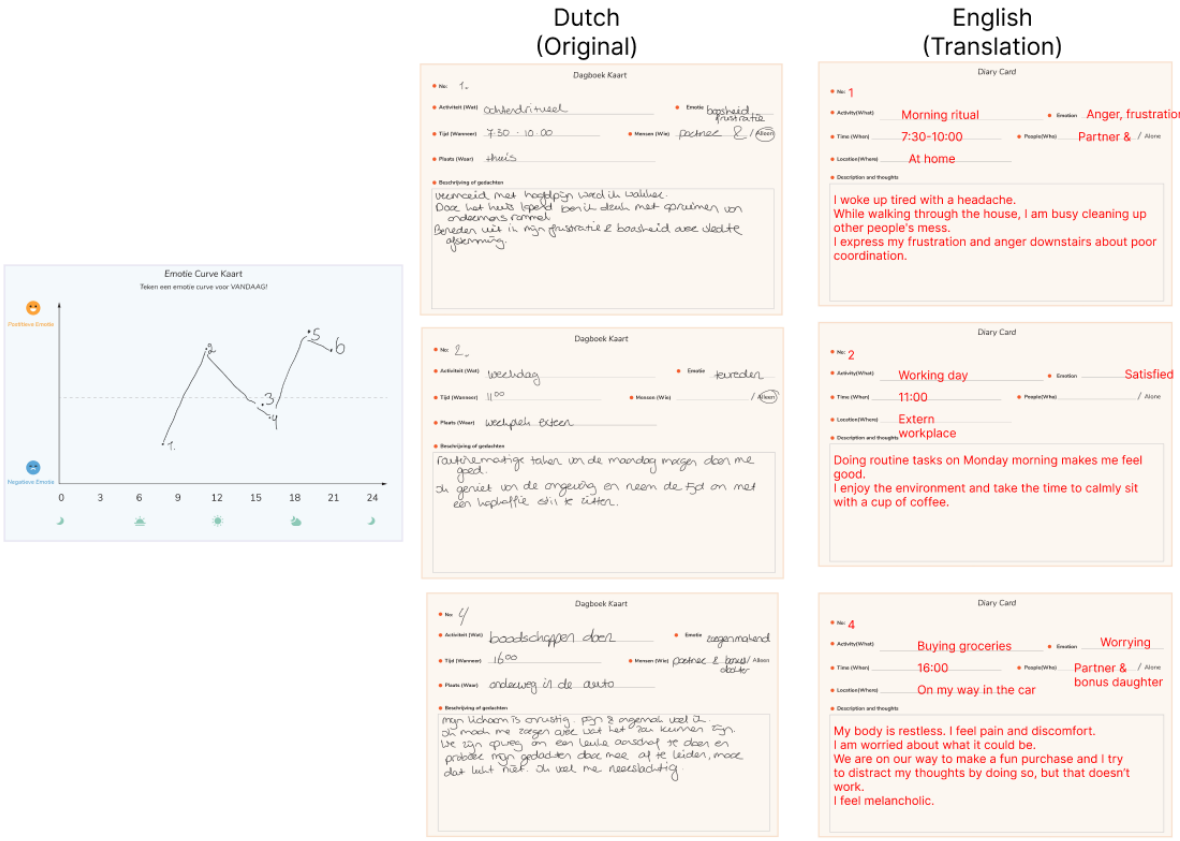
Module B Results

Overview of Module B Results

The designer observed significant fluctuations in most patients' emotions throughout the day (Figure 7). These emotional variations are closely linked to specific events or activities. Specifically, emotions are highly correlated with the 5W1H. When patients describe their emotional fluctuations, they often mention the involvement of individuals or the locations where these emotions were experienced. Being alone can also trigger emotional changes in patients, usually associated with past events or experiences, where memories induce emotional responses.

XSL•FO
RenderX

Figure 7. An example of a completed diary card from a patient with colorectal cancer.



During the study, the designer observed that patients frequently discussed challenges related to their energy levels. Many participants reported experiencing a significant decline in energy following their cancer diagnosis, which restricted their ability to engage in daily activities. For example, participant C expressed that the persistent lack of energy caused considerable frustration:

> *I often feel like I want to do something...but it's very difficult. My body, brain, and mind are not as active anymore...If something takes too much energy, it makes me feel very frustrated.* [Participant C]

Social engagements are also significantly connected to energy levels. Participants reported that their social circles include more patients, providing mutual support. Patients consider this peer support crucial. However, it also brings certain challenges. Due to their reduced energy levels compared with before their cancer diagnosis, maintaining relationships with old friends, especially those living far away, has become increasingly difficult. As participant D mentioned:

> *I'm glad to have friends who are in similar situations as me...and we support each other. However, I also find it difficult to maintain some relationships...I have a good friend who lives a two-hour drive away. In the past, meeting her was easy because driving was not a problem for me, but after the surgery, it has become a bit difficult.* [Participant D]

Regarding their attitudes toward the prototype and tracking progress, the 4 participants have a very positive attitude. Although 1 patient (participant A) did not complete the study as required, she said that she enjoyed the process very much. The difficulty she found was simply due to the timing, as many years had passed since her CRC diagnosis. "If I had been invited to do this study five years ago, I am sure there would have been very much to record," explaining her incomplete entries. This also highlights the necessity of early monitoring, given that patients experience much more severe social and emotional fluctuations initially than they do years after diagnosis. Most participants expressed that the process of recording their activities and emotions had heightened their self-awareness and facilitated a deeper understanding of themselves, as the following quotation illustrates:

> *It (the prototype) can make you think more about the activities and how you feel about it. For me I am shocked to see how I feel.* [Participant B]

This self-awareness enables patients to be aware of what activities can have an impact on their mental health:

> *It was for me a bad week. I wasn't feeling well and then because of it (prototype) I was very aware that my best friends were my biggest triggers. So I understand more on what happened between them and me.* [Participant C]

Some patients have reported that the prototype can influence their way of seeing a day and even subsequent behaviors in a positive way:

XSL•FO
RenderX

> *When I think I had a bad day, I look back at the diary and I think, oh, it wasn't so bad. There were just two bad moments...And what this awareness brings as extra is some influence on me...On the third day of the diary study, I realized I didn't have my mindful moment like before which made me realize that it is so important to me, so I'll pick up again.* [Participant C]

Patients also expressed that they like the prototype more than the questionnaire because it is more personalized and attractive:

> *I prefer this one (prototype) because it's about me.* [Participant A]
>
> *I really like this...The process is like a story that I can look back on from time to time. I am also tracking myself on diet and weight...* [Participant D]

When it comes to sharing their data with stakeholders, patients expressed both cautious and positive attitudes. They felt that sharing information about their daily activities and emotions with doctors could help foster a better understanding of their overall well-being. Patients also perceived that health care providers often paid limited attention to their everyday life conditions and emotional states. As a result, these aspects were frequently managed independently by patients, although they had a substantial impact on their health:

> *I am willing to share with my doctor, especially when I think I need help. Because I also understand that doctors are very busy...it's really nice to share your story with other patients that...I woke up and then I got a leakage again. And there is a lot of information I would share with them.* [Participant D]

Module B Summary

Module B used the diary study and follow-up interviews to study 4 patients, mainly to explore the possibility of remote monitoring of activities and emotional well-being in the follow-up phase. Overall, the patients' attitudes toward the prototype were positive, although there are still several limitations that require further design iterations and refinement. However, the diary study helped us to conduct a preliminary trial of remote patient monitoring and gain a deeper understanding of the patient's daily life after surgery. The interviews helped us understand the patients' attitudes and expectations about remote patient monitoring. It also pointed out the possible impacts of remote patient monitoring on patients and key considerations for future implementations. The results of module B are very inspiring for the design of the monitoring system.

Discussion

Mixed Methods and the Evolving HCD Context

Emerging research shows the benefits of using mixed methods in patient research, including support for investigating complex health care processes and systems and identifying improvement opportunities [40-42], and the patient data from both qualitative and quantitative research are significant for patient-centered improvement [43]. Despite these benefits, the use of mixed methods in the early HCD process remains underdeveloped, lacking clearly defined best practices and implementation approaches.

Meanwhile, a more system-level perspective in the HCD is required in contemporary practice. Recently, HCD has evolved into 5 principles of humanity-centered design that integrate ecosystems and community perspectives more comprehensively than the 4 principles of HCD, which are solving the core issues, focusing on people, taking a systems point of view, and rapid iterations on interventions [44]. Continually testing and refining proposals is therefore necessary to ensure that they truly meet the needs of the people for whom they are intended. A closer look at this brand-new principle reveals: "Design with the community and as much as possible support designs by the community. Professional designers should serve as enablers, facilitators, and resources, aiding community members to meet their concerns [44]." It requires the designers to take actions not only individually but also from a community or even system level, which corresponds with other research from the view of human-computer interaction [45]. Thus, designers need to understand not only the patients themselves but also the environment and systems around them. It implies that research methods should evolve by combining emerging technologies.

Our experience shows that the mixed method of large-scale research and in-depth research can play an essential role in the beginning of the HCD process. The possible implementation of this method is explored by a case study about the development of the remote patient-monitoring system for patients with CRC. It is important to emphasize that this data-enabled mixed method is not intended to replace traditional qualitative approaches (such as standard interview-only or ethnographic methods), nor do we claim it is universally superior. Rather, it serves as a complementary approach that leverages emerging digital technologies to uncover new possibilities within user research. Compared with other mixed methods, the difference is involving large-scale digital research such as big data analysis in module A, followed by patient engagement research.

Connection Between Module A and Module B

The connection between module A and module B relies on combining broad and deep analysis, reframing the problem, and translating data into designer empathy.

First, the 2 modules, where module A's big data results and module B's in-depth insight, worked as a unified process to ensure both breadth and depth. The large-scale analysis in module A established a macro-level understanding of the persistent social and mental health struggles facing patients with CRC, which acted as a research scope for the entire process. This scope was then brought into sharp focus by module B, where the in-depth engagement with the 4 patients provided the concrete lived experiences that humanized the patterns identified in the big data. During the interviews, the participants' words did not simply validate the previous findings but extended them by illustrating how these social and mental challenges manifest in daily life. This connection shows that module A provided the necessary generalizability to scope the research, while module B provided the contextual precision to interpret the findings.

XSL•FO
RenderX

Second, module A facilitated the process of problem reframing. Through the analysis of 212,107 posts, our patient community journey map revealed that patients with CRC in the follow-up stage shifted their focus from clinical recovery to social and mental well-being. By contrasting these large-scale digital insights with existing literature, we identified a critical gap: current remote patient monitoring systems overlook the psychosocial dimensions of home-based recovery. This finding directly dictated the scope of the prototype design, shifting it from simple physiological tracking to a system designed to capture daily activities and emotional nuances.

Finally, module A provided a foundational understanding that informed the design of the engagement prototype in module B. An immersion into the patient experience allowed the designer to develop a rich understanding of the participants' lived realities. By navigating 37 topics and 10 clusters, the designer internalized the emotional landscape of patients with CRC. For instance, when analyzing topic 28, the designer was exposed to the intimate struggles surrounding family relationships and the resulting emotional volatility. This process of deep engagement with patient narratives enabled the designer to move beyond clinical observations and truly resonate with the patients' underlying anxieties and needs. Consequently, this developed empathy acted as the primary design driver: the designer was moved to incorporate an emotion curve tracking task and specific emotional reflection prompts into the diary study. By transforming these internalized emotional insights into practical research materials, the designer ensured that the engagement activities in module B were designed not just to collect data but to resonate with the patients' most pressing social and mental health challenges.

Potential Values of the Data-Enabled Mixed Method

The preliminary findings suggest this data-enabled mixed method has the potential to help health care designers acquire representative insights that reflect patients' perspectives, develop an empathetic understanding to navigate the complexities of medical contexts, guide the definition of research and design directions, and support the creation of engaging research activities. Together, these capabilities provide critical, informed insights that enhance the effectiveness of patient engagement research. In the following sections, we will illustrate these values in detail by addressing the 3 challenges outlined in the introduction.

Addressing the Current HCD Health Care Design Challenges

This case study also provides preliminary evidence that using the data-enabled mixed method has the potential to address the challenges of the current human-centered health care design process involving patients, which are (1) high vulnerability of patient participants, (2) less diverse and representative patient groups due to recruitment challenges, and (3) insufficient problem framing across diverse patient experiences.

The first challenge describes the vulnerability of patient participants. Our data-enabled mixed method tackles this challenge by cultivating designers' contextual sensitivity prior to patient engagement in the study. In our CRC case, the designer identified sensitive topics, necessary caution in words and behavior, and moments that require alertness before involving them. Meanwhile, the diary study and interviews in module B were more considerate of patients' vulnerability. For instance, based on insights from module A that patients with CRC often experience fatigue as topic 24 describes, we provided both online and offline interview study options. As shown in Table 1, 3 out of 4 participants chose the online option. For the participant who opted for an offline interview, the designer visited her home for a face-to-face discussion with her consent. Furthermore, as expected, patients also mentioned the impact of declining energy levels on changes in their daily activities and social circles during the interviews. The designer also gained insights into whether patients express positive or negative emotions when discussing these topics. Particularly when patients presented certain topics with highly negative emotions, the topics were given special attention during the interview process. This enables the designer to avoid language and actions that may cause discomfort or be sensitive to patients, or at the very least, have expectations and preparations for the upcoming situation. The respect for patients aligns with the proposal by Das and Svanæs, referring to "Particular considerations and facilitation is required when including vulnerable user groups in human-centered design activities [1]."

The second challenge, related to less diverse and representative patient groups due to recruitment challenges, can be effectively addressed by the combination of modules A and B. The large dataset of 212,107 posts given in module A contributed to mitigating the lack of representativeness caused by small samples [46]. The 37 topics related to patient experiences during their care path have been identified from these posts, with thousands of posts associated with each topic. Furthermore, we realized that many of these topics align with the ones discussed during the patient engagement studies in module B. For example, participant D illustrates the mutual support shared among patient peers, which also reflects in patient posts of sharing strategies for coping and improving quality of life during the follow-up stage. However, some of the topics found in module A were more detailed and did not show up in module B. For example, topic 43 revealed patients' concerns about the potential hereditary nature of CRC and topic 3 highlighted the significance of anniversaries and birthdays to patients. These findings, which were not evident in the diary study, yet represent intriguing new insights. These representative outcomes provide shared insights from different demographic patient groups and can be applicable to a wider system.

The third challenge concerns the reframing of problems from the patient perspective during the early stages of HCD. The potential of our method in addressing this challenge has been discussed in the previous section (Connection Between Module A and Module B), where we highlighted that conducting large-scale patient research can help reframe research and design problems. If module A had not been conducted and a more solution-driven approach had been adopted instead, the focus might have been on features such as information registration, symptom monitoring, patient education, and personalized advice. Consequently, the social and mental health needs of patients might remain overlooked or fail to emerge as explicit areas of

XSL•FO
RenderX

concern. Our method encourages designers to define research and design questions based on the diverse and in-depth perspectives of patient voices. For instance, designers can easily learn that patients with CRC need to undergo blood tests to monitor treatment effectiveness. However, it was only through topic 49 in module A that the designer realized that the numerical results of these tests can cause emotional instability for a large number of patients. The rich context obtained from the large-scale data can not only identify problems within complex health systems [47] but also contribute to designing desirable experiences [48].

Limitations

Naturally, this data-enabled mixed method has certain limitations. From the perspective of organizations, designers may lack proficiency in handling large-scale data. The data processing process would require the involvement of computer scientists. Although interdisciplinary collaboration is beneficial to the design process, in practice user research in HCD itself has been proven to be time-consuming and costly [49], and collaboration with computer scientists may exacerbate this phenomenon. It is therefore necessary to explore further how to effectively work together with data scientists. While the vastness of the data helps obscure individual identities, ethical and privacy concerns still demand rigorous attention. When collecting raw data, researchers must selectively gather information to ensure complete anonymization.

In addition, a notable limitation of this case study is the selection bias inherent in using English-language online forums, which excludes those with limited digital literacy or non-English proficiency. To mitigate this, we engaged local domain experts from Rotterdam, an area with a high concentration of underserved and diverse patient populations, and further enriched our data through interviews with Dutch-speaking patients in module B. However, due to the small sample size in these engagement phases, our findings may still underrepresent marginalized or offline populations. Future research should incorporate multilingual, multiplatform datasets and larger, more diverse cohorts to enhance generalizability.

Another potential limitation of this case study is that relying on a single designer restricts the capacity for independent cross-validation. While the expert cocreation sessions in module A helped mitigate this risk, future applications should ideally involve a larger interdisciplinary team to independently cross-analyze the modules, thereby further minimizing subjective bias and enhancing overall reliability.

Although the proposed approach was empirically applied to the specific context of CRC, the results indicate preliminary feasibility rather than definitive proof of the method's applicability across diverse health care domains, care paths, or varying team structures. Additionally, the small sample size in module B (n=4) restricts the generalizability of the qualitative findings in this case, precluding a comprehensive evaluation of the prototype's efficacy or definitive design requirements. However, the primary objective of this study was to explore the feasibility and values of the methodological pipeline. Within this scope, the small-scale engagement shed light on preliminary insights into how qualitative interactions can contextualize, ground, and humanize the broad insights derived from the massive data in module A, ultimately addressing the challenges encountered by health care designers in the initial design stage.

Future Research

Building upon the limitations identified in this exploratory case study, future work should prioritize rigorously validating the method's generalizability and establishing best practices. Since our current findings represent preliminary indications based on a CRC case, future comparative studies and multiple case studies across diverse health care domains and different disease care paths are required to establish robust evidence.

Furthermore, to mitigate the selection bias inherent in online datasets, subsequent applications of this data-enabled mixed method should incorporate multilingual datasets from diverse geographical platforms. During the engagement research phase (module B), researchers should also explore the specific contexts of the participants, particularly investigating whether their situations reflect the experiences of marginalized populations, such as non-English speakers or individuals with limited digital access.

In addition to diversifying the data platforms, as longitudinal datasets are increasingly integrated into this process, future studies using this data-enabled mixed method should consider and validate the temporal insights to ensure that they reflect current state-of-the-art practices. To explain this within the context of our case, researchers using longitudinal online datasets must be mindful of the temporal dimension. This caution is necessary because while certain patient experiences, such as psychosocial challenges, often remain consistent over time, specific medical contexts, such as treatment modalities, surgical advancements, or evolving care protocols, can undergo rapid technological shifts.

While this method has initially centered on patients, future research should expand to include other stakeholders, such as caregivers and health care professionals, thereby fostering a more holistic design approach. In parallel with expanding the participant scope, future practices could also explore methodological flexibility. For instance, although we currently recommend conducting module A before module B, researchers might alternate this sequence by using large-scale digital research to validate prior in-depth qualitative findings. Ultimately, to support these methodological advancements, integrate emerging technologies, and mitigate the risk of confirmation bias, continuous interdisciplinary collaboration among health care designers, data scientists, and health care professionals is essential. Such team-based approaches will enable cross-validation of the results in both modules and establish standardized, objective best practices.

Conclusions

This paper highlights the trend of HCD adopting a more system-level perspective, which introduces challenges for health care designers at the early stages of patient-focused design, including (1) high vulnerability of patient participants, (2) less diverse and representative patient groups due to recruitment challenges, and (3) insufficient problem framing across diverse patient experiences. To address these challenges, we propose a

XSL•FO
RenderX

data-enabled mixed method combining large-scale digital patient research (module A) and in-depth engagement research (module B).

We explore this approach through a case study of CRC follow-up care, using patient community journey mapping (module A) and diary studies and interviews (module B) to inform the development of a remote patient-monitoring system.

Based on this specific case study, there are preliminary indications that this method has the potential to support problem reframing and cultivate empathy toward patient vulnerability before conducting engagement research. By capturing the diversity and depth of patient voices within this context, it generated evidence-based insights that appeared to enhance the design of subsequent patient engagement activities. While these initial findings are promising, future research across broader health care contexts is needed to generate generalizable evidence.

Acknowledgments

The authors thank all the patient participants and the computer scientist for their valuable contribution. Google Gemini 3.1 Pro was used for correcting grammatical errors and enhancing readability.

Funding

The authors declared no financial support was received for this work.

Data Availability

The datasets generated and analyzed during module A are available from the corresponding author on reasonable request. The data gathered during module B are not publicly available and cannot be shared due to confidentiality and original ethics approval restrictions.

Authors' Contributions

Conceptualization: YS, JJ
Data curation and formal analysis: YS, JJ
Writing – original draft: YS
Writing – review & editing: YS, JJ
All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Das A, Svanæs D. Human-centred methods in the design of an e-health solution for patients undergoing weight loss treatment. Int J Med Inform 2013 Nov;82(11):1075-1091. [doi: 10.1016/j.ijmedinf.2013.06.008] [Medline: 23886482]
2. Vagal A, Wahab S, Lecky S, et al. Optimizing patient experience using human-centered design. J Am Coll Radiol 2020 May;17(5):668-672. [doi: 10.1016/j.jacr.2019.11.020] [Medline: 31891673]
3. Carayon P, Wooldridge A, Hoonakker P, Hundt AS, Kelly MM. SEIPS 3.0: human-centered design of the patient journey for patient safety. Appl Ergon 2020 Apr;84:103033. [doi: 10.1016/j.apergo.2019.103033] [Medline: 31987516]
4. Giacomin J. What is human centred design? Design J 2014 Dec;17(4):606-623. [doi: 10.2752/175630614X14056185480186]
5. Göttgens I, Oertelt-Prigione S. The application of human-centered design approaches in health research and innovation: a narrative review of current practices. JMIR Mhealth Uhealth 2021 Dec 6;9(12):e28102. [doi: 10.2196/preprints.28102]
6. The double diamond. Design Council. URL: https://www.designcouncil.org.uk/our-resources/the-double-diamond/ [accessed 2024-12-24]
7. Sanders E, Stappers P. Convivial Toolbox: Generative Research for the Front End of Design: Bis; 2012.
8. Melles M, Albayrak A, Goossens R. Innovating health care: Key characteristics of human-centered design. Int J Qual Health Care 2021 Jan 12;33(Supplement_1):37-44. [doi: 10.1093/intqhc/mzaa127] [Medline: 33068104]
9. Palinkas LA, Horwitz SM, Chamberlain P, Hurlburt MS, Landsverk J. Mixed-methods designs in mental health services research: A review. Psychiatr Serv 2011 Mar;62(3):255-263. [doi: 10.1176/ps.62.3.pss6203_0255] [Medline: 21363896]
10. Miller WL, Crabtree BF, Harrison MI, Fennell ML. Integrating mixed methods in health services and delivery system research. Health Serv Res 2013 Dec;48(6 Pt 2):2125-2133. [doi: 10.1111/1475-6773.12123] [Medline: 24279834]
11. van Smoorenburg AN, Hertroijs DFL, Dekkers T, Elissen AMJ, Melles M. Patients' perspective on self-management: type 2 diabetes in daily life. BMC Health Serv Res 2019 Aug 28;19(1):605. [doi: 10.1186/s12913-019-4384-7] [Medline: 31462220]

XSL•FO
RenderX

12. De Vito Dabbs A, Myers BA, Mc Curry KR, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183. [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
13. Woodgate RL, Zurba M, Tennent P. Advancing patient engagement: youth and family participation in health research communities of practice. Res Involv Engagem 2018;4(1):9. [doi: 10.1186/s40900-018-0094-2] [Medline: 29560275]
14. Thies A. On the value of design thinking for innovation in complex contexts: a case from healthcare. IxD&A 2015;27(27):159-171. [doi: 10.55612/s-5002-027-009]
15. Duffy A, Boroumandzad N, Sherman AL, Christie G, Riadi I, Moreno S. Examining challenges to co-design digital health interventions with end users: systematic review. J Med Internet Res 2025 Mar 14;27:e50178. [doi: 10.2196/50178]
16. Groeneveld B, Dekkers T, Boon B, D'Olivo P. Challenges for design researchers in healthcare. Design Health 2018 Jul 3;2(2):305-326. [doi: 10.1080/24735132.2018.1541699]
17. van Velsen L, Ludden G, Grünloh C. The limitations of user-and human-centered design in an eHealth context and how to move beyond them. J Med Internet Res 2022 Oct 5;24(10):e37341. [doi: 10.2196/37341] [Medline: 36197718]
18. Jongsma K, Friesen P. The challenge of demandingness in citizen science and participatory research. Am J Bioeth 2019 Aug;19(8):33-35. [doi: 10.1080/15265161.2019.1619867] [Medline: 31544651]
19. Wang T, Giunti G, Melles M, Goossens R. Digital patient experience: umbrella systematic review. J Med Internet Res ;24(8):e37952. [doi: 10.2196/37952]
20. Voorheis P, Zhao A, Kuluski K, et al. Integrating behavioral science and design thinking to develop mobile health interventions: systematic scoping review. JMIR Mhealth Uhealth 2022 Mar 16;10(3):e35799. [doi: 10.2196/35799] [Medline: 35293871]
21. Noergaard B, Sandvei M, Rottmann N, et al. Development of a web-based health care intervention for patients with heart disease: lessons learned from a participatory design study. JMIR Res Protoc 2017 May 17;6(5):e75. [doi: 10.2196/resprot.7084] [Medline: 28526674]
22. Oshio T, Kupperman J. The problem behind the problem: applying human-centered design to child care in flint. Early Child Educ J 2022;50(8):1373-1382. [doi: 10.1007/s10643-021-01263-5] [Medline: 34566398]
23. Kim EJ, Nam IC, Koo YR. Reframing patient experience approaches and methods to achieve patient-centeredness in healthcare: scoping review. Int J Environ Res Public Health 2022 Jul 27;19(15):9163. [doi: 10.3390/ijerph19159163] [Medline: 35954517]
24. Bazzano AN, Martin J, Hicks E, Faughnan M, Murphy L. Human-centred design in global health: a scoping review of applications and contexts. PLoS One 2017 Nov 1;12(11):e0186744. [doi: 10.1371/journal.pone.0186744]
25. Vial S, Boudhraâ S, Dumont M. Human-centered design approaches in digital mental health interventions: exploratory mapping review. JMIR Ment Health 2022 Jun 7;9(6):e35591. [doi: 10.2196/35591] [Medline: 35671081]
26. Soliman E, Mogefors D, Bergmann JHM. Problem-driven innovation models for emerging technologies. Health Technol 2020 Sep;10(5):1195-1206. [doi: 10.1007/s12553-020-00450-5]
27. Dubbelman J, Ooms J, Havgry L, Simonse L. Communal load sharing of miscarriage experiences: thematic analysis of social media community support. J Med Internet Res 2024 Dec 10;26(1):e56680. [doi: 10.2196/56680] [Medline: 39656509]
28. Pócs D, Adamovits O, Watti J, Kovács R, Kelemen O. Facebook users' interactions, organic reach, and engagement in a smoking cessation intervention: content analysis. J Med Internet Res 2021 Jun 21;23(6):e27853. [doi: 10.2196/27853] [Medline: 34152280]
29. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424. [doi: 10.3322/caac.21492] [Medline: 30207593]
30. Xi Y, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. Transl Oncol 2021 Oct;14(10):101174. [doi: 10.1016/j.tranon.2021.101174] [Medline: 34243011]
31. van Leeuwen M, Husson O, Alberti P, et al. Understanding the quality of life (QOL) issues in survivors of cancer: towards the development of an EORTC QOL cancer survivorship questionnaire. Health Qual Life Outcomes 2018 Jun 4;16(1):114. [doi: 10.1186/s12955-018-0920-0] [Medline: 29866185]
32. Jung J, Kim KH, Peters T, Snelders D, Kleinsmann M. Advancing design approaches through data-driven techniques: Patient community journey mapping using online stories and machine learning. Int J Design 2023;17(2). [doi: 10.57698/v17i2.02]
33. Madathil SC, Lopes AJ, Alfred M. Patient journey mapping: A literature review. : Institute of Industrial and Systems Engineers; 2020 Presented at: IIE Annual Conference Proceedings; May 30 to Jun 2, 2020; Norcross, United States p. 937-942 URL: https://www.proquest.com/docview/2511386708?pq-origsite=gscholar&fromopenview=true [accessed 2026-08-14]
34. Voigt K, Sun Y, Patandin A, et al. A Machine Learning Approach Using Topic Modeling to Identify and Assess Experiences of Patients With Colorectal Cancer: Explorative Study. JMIR Cancer 2025 Jan 27;11:e58834. [doi: 10.2196/58834] [Medline: 39874195]
35. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2026-08-14]

XSL•FO
RenderX

36. Stulz A, Lamore K, Montalescot L, Favez N, Flahault C. Sexual health in colon cancer patients: a systematic review. Psychooncology 2020 Jul;29(7):1095-1104. [doi: 10.1002/pon.5391] [Medline: 32281129]
37. Thyø A, Elfeki H, Laurberg S, Emmertsen KJ. Female sexual problems after treatment for colorectal cancer—a population-based study. Colorectal Dis 2019 Oct;21(10):1130-1139. [doi: 10.1111/codi.14710] [Medline: 31095852]
38. Traa MJ, De Vries J, Roukema JA, Den Oudsten BL. Sexual (dys)function and the quality of sexual life in patients with colorectal cancer: a systematic review. Ann Oncol 2012 Jan;23(1):19-27. [doi: 10.1093/annonc/mdr133] [Medline: 21508174]
39. Traa MJ, De Vries J, Roukema JA, Rutten HJT, Den Oudsten BL. The sexual health care needs after colorectal cancer: the view of patients, partners, and health care professionals. Support Care Cancer 2014 Mar;22(3):763-772. [doi: 10.1007/s00520-013-2032-z] [Medline: 24240645]
40. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res 2013 Dec;48(6 Pt 2):2134-2156. [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
41. Philpot LM, Khokhar BA, DeZutter MA, et al. Creation of a patient-centered journey map to improve the patient experience: a mixed methods approach. Mayo Clin Proc Innov Qual Outcomes 2019 Dec;3(4):466-475. [doi: 10.1016/j.mayocpiqo.2019.07.004] [Medline: 31993565]
42. Sheard L, Marsh C, Mills T, et al. Using patient experience data to develop a patient experience toolkit to improve hospital care: a mixed-methods study. Health Serv Deliv Res 2019 Oct 17;7(36):1-104. [doi: 10.3310/hsdr07360]
43. Luxford K, Safran DG, Delbanco T. Promoting patient-centered care: a qualitative study of facilitators and barriers in healthcare organizations with a reputation for improving the patient experience. Int J Qual Health Care 2011 Oct;23(5):510-515. [doi: 10.1093/intqhc/mzr024] [Medline: 21586433]
44. Norman DA. Design for a Better World: Meaningful, Sustainable, Humanity Centered: MIT Press; 2023. URL: http://ebookcentral.proquest.com/lib/delft/detail.action?docID=29673034 [accessed 2024-05-07]
45. Jung J, Kleinsmann M, Snelders D. A vision for design in the era of collective computing. J Eng Design 2022 Apr 3;33(4):305-342. [doi: 10.1080/09544828.2022.2050682]
46. Mayer-Schönberger V, Ingelsson E. Big data and medicine: a big deal? J Intern Med 2018 May;283(5):418-429. [doi: 10.1111/joim.12721] [Medline: 29235174]
47. May CR, Johnson M, Finch T. Implementation, context and complexity. Implement Sci 2016 Oct 19;11(1):141. [doi: 10.1186/s13012-016-0506-3] [Medline: 27756414]
48. Small S, Sexton J, Priestley D. Leveraging data-driven insights to improve the health experience for patients starting and staying on treatment. J Patient Exp 2025;12:23743735251381444. [doi: 10.1177/23743735251381444] [Medline: 40994896]
49. Goodman-Deane J, Langdon P, Clarkson J. Key influences on the user-centred design process. J Eng Design 2010 Jun;21(2-3):345-373. [doi: 10.1080/09544820903364912]

Abbreviations

CRC: colorectal cancer
HCD: human-centered design

Edited by A Kushniruk, S Law; submitted 24.Nov.2025; peer-reviewed by M Jani, S Matsuda; revised version received 17.May.2026; accepted 09.Jul.2026; published 19.Aug.2026.

Please cite as:
Sun Y, Jung J
Development of a Data-Enabled Mixed Method for Designing for Patients From a Human-Centered Design Perspective: Explorative Case Study
JMIR Hum Factors 2026;13:e87677
URL: https://humanfactors.jmir.org/2026/1/e87677
doi:10.2196/87677

© Yingtao Sun, Jiwon Jung. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 19.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

What Users Say About Reimbursable Digital Therapeutics in Germany: Large-Scale App Store Review Analysis Using a Large Language Model

Benjamin Kinast[1], MA; Henrik Rohde[1], MSc; Björn Schreiweis[1], Prof Dr; Hannes Ulrich[1,2], Dr rer nat

[1]Institute for Medical Informatics and Artificial Intelligence, Kiel University and University Hospital Schleswig-Holstein, Kaistrasse 101, Kiel, Germany
[2]Department of Medical Informatics, Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany

Corresponding Author:
Benjamin Kinast, MA
Institute for Medical Informatics and Artificial Intelligence, Kiel University and University Hospital Schleswig-Holstein, Kaistrasse 101, Kiel, Germany

Abstract

Background: In 2019, Germany introduced a unique regulatory framework for digital therapeutics (DTx), termed digital health applications (DiGAs) in Germany, with the goal of integrating evidence-based DTx into statutory health care. DTx are eligible for reimbursement by statutory health insurance if manufacturers demonstrate positive health care effects, such as improved health status or health literacy, in a controlled study setting. Although regulatory evaluation primarily relies on manufacturer-conducted studies, these studies do not fully capture how users experience and use DiGAs in everyday life.

Objective: The study aims to systematically characterize user experiences with reimbursable DiGAs by analyzing the sentiment and thematic content in publicly available app store reviews at scale to identify recurring patterns of positive and negative feedback across regulatory-relevant quality dimensions.

Methods: All DiGAs with publicly available mobile app reviews were identified via the official German Federal Institute for Drugs and Medical Devices (BfArM) directory. User reviews were extracted from the German Apple App Store and Google Play Store using a tailored Python script. Reviews were then processed using GPT-4o, which extracted up to 5 core statements per review and assigned each a sentiment label (positive, neutral, or negative) and one of ten predefined thematic categories to each statement: content, technology, cost and reimbursement, login and registration, prescription and approval, user experience and design, support, tracking and documentation, effectiveness, and overall impression, all derived from the regulatory and quality criteria applicable to DiGA certification. Classification quality was assessed through manual validation, in which each of the 3 authors independently reviewed one-third of all model outputs.

Results: In total, 44 mobile DiGAs were included and analyzed. After data extraction and cleansing, the final dataset comprised 4328 reviews containing at least one interpretable statement, resulting in 9439 interpretable statements. A systematic validation of the automated classification demonstrated exceptionally high model performance, with 99% accuracy for sentiment classification (F_1-scores of 1.00 for positive and 0.99 for negative categories) and 95% accuracy for category classification (average F_1-score of 0.95). While the categories overall impression (1233/1431, 86.2% positive) and effectiveness (1440/1609, 89.5% positive) received particularly positive feedback, users commented most negatively on the login and registration process (263/280, 93.9% negative) and technology-related aspects (563/668, 84.3% negative).

Conclusions: User feedback on reimbursable DiGAs is predominantly positive, particularly regarding perceived effectiveness, overall impression, and content. However, recurring criticism of login and registration processes, technical reliability, and prescription and approval procedures reveals persistent barriers to access and use. As many of these aspects are prerequisites rather than peripheral convenience issues, they should be treated as essential implementation requirements. Analyzed at scale with large language model-based methods, publicly available app store reviews can complement formal evaluation by providing user-centered real-world evidence for postmarket monitoring and regulatory quality improvement.

(JMIR Hum Factors 2026;13:e92415) doi:10.2196/92415

KEYWORDS

mobile health; mHealth; eHealth; apps; digital therapeutics; large language model; LLM; applications

XSL•FO
RenderX

Introduction

In 2019, Germany introduced a unique regulatory framework for digital therapeutics (DTx) known as digital health applications (DiGAs), with the goal of integrating evidence-based DTx into routine care [1,2]. Under the Digital Healthcare Act (Digitale-Versorgung-Gesetz) [3], DiGAs can be prescribed by physicians and reimbursed by statutory health insurance if they are listed in an official directory maintained by the German Federal Institute for Drugs and Medical Devices (BfArM) [4]. The directory covers DTx across a range of indications, such as mental and behavioral disorders, musculoskeletal disorders, metabolic disorders, and disorders of the nervous system. This initiative aims to improve access to innovative health care technologies while ensuring their safety, quality, and effectiveness.

To qualify for inclusion, DTx must meet clinical, legal, and technical requirements, including demonstrable benefits to patient care or care structures as defined in section 139e of the German Social Code, Book V (SGB V) [3]. These so-called positive health care effects refer to either a medical benefit or to patient-relevant improvements in structures and processes. A medical benefit is defined as an effect such as an improvement in health status, a reduction of disease duration, prolonged survival, or increased quality of life. Patient-relevant structural or procedural improvements may include better coordination of the care process, easier access to health care services, or the promotion of health literacy [5]. To facilitate timely integration into routine care, the law also additionally introduced a fast-track procedure that allows provisional listing for up to 12 months while manufacturers gather evidence of positive health care effects [6,7].

Although regulatory evaluation rests primarily on manufacturer-conducted clinical studies [8,9], such studies are not designed to capture the patient-centered dimensions that determine whether a DiGA is adopted and used in everyday life [10]. For the German DiGA context specifically, it has been argued that acceptance, perceived usefulness, and usability are decisive for sustained use, yet genuine user-centeredness is not yet firmly embedded in DiGA development [11]. Usability, acceptability, and the practical feasibility of integrating an app into daily routines are key determinants of sustained engagement, yet they are only partially addressed during development and approval. To date, the actual use and adherence to a DTx have been identified as a critical but undersupported determinant of effectiveness, for which clinical practice still lacks established assessment tools [12]. Low uptake and prescription rates further underline that factors beyond clinical efficacy shape whether DiGAs reach and benefit patients [13]. As several studies emphasize, user reviews of mobile health (mHealth) apps can reveal valuable insights into real-world usability and recurring quality concerns. All these dimensions are directly relevant to the DiGA approval framework [14-17]. Analyzing user-generated feedback can therefore offer a scalable, real-world complement to clinical evidence and may surface practical barriers to adoption and sustained use.

Importantly, the regulatory framework itself recognizes the relevance of patient perspectives. Section 139e SGB V requires that the outcomes of digital health apps include patient-reported health status during usage as part of mandatory success monitoring [3]. This emphasis is mirrored in the DiGA guideline (DiGA-Leitfaden), which explicitly permits manufacturers to demonstrate positive health care effects using comparative quantitative methods drawn from the social or behavioral sciences, provided validated instruments are used [18]. This institutional acknowledgment underscores that patient experience has become a regulatory concern in the DiGA framework. App store reviews provide spontaneous, large-scale, real-world feedback across the entire DiGA landscape in a way that standardized instruments do not. While they cannot replace validated patient-reported outcome instruments, their systematic analysis can potentially reveal how DiGAs are experienced in everyday use.

To date, however, patient feedback on German DiGAs has received limited systematic attention. Uncovska et al [19] analyzed app store reviews of mHealth apps in Germany, comparing regulated DiGAs with nonregulated consumer apps using BERTopic modeling and sentiment analysis. Their work showed that DiGAs received more favorable contemporary ratings than nonregulated apps. Positive themes included customer service, personalization, and ease of use, while software bugs and cumbersome registration processes were frequent complaints across both app types. Their analysis, however, covered only 15 DiGAs and relied on data-driven topic categories that do not map onto the regulatory dimensions against which DiGAs are formally assessed. Furthermore, it was framed primarily as a comparison between regulated and nonregulated apps rather than a focused account of the user experience with DiGAs themselves.

Internationally, app store reviews have been used to examine user experience with health apps in various contexts, ranging from consumer health apps to regulated prescription DTx in the United States [20,21]. This leaves a clear gap, as to our knowledge, there is no systematic, large-scale account of how users evaluate the full range of listed DiGAs along the dimensions that are central to their regulatory assessment. To address this gap, this study analyzes publicly available DiGA user reviews from the German Apple App Store (iOS) and the Google Play Store (Android) for all approved and provisionally approved mobile DiGAs as of February 25, 2025.

Using a regulation-informed category scheme and large language model (LLM)–assisted classification, we aim to characterize how users evaluate DiGAs across dimensions relevant to regulatory assessment and to identify where regulatory expectations and lived user experience may converge or diverge.

Methods

Data Identification and Extraction

Data collection involved 2 stages: first, we identified all mobile DTx listed by the BfArM as of February 25th, 2025. Second, we used Selenium-based Python scripts (Version 4.29.0) to extract user reviews from the German Apple App Store (iOS)

XSL•FO
RenderX

and the Google Play Store (Android). Extracted data included review content, ratings, dates, developer responses, and app metadata, all stored in structured JSON format.

Category Development and Regulatory Grounding

The thematic category scheme was developed using a regulatory-driven approach. Categories were derived from legal and regulatory requirements governing DiGAs in Germany, with the aim to ensure regulatory relevance and interpretability of user-reported feedback. Specifically, we systematically reviewed the SGB V, the Digital Health Applications Regulation (DiGAV), and the official BfArM DiGA guideline [5,18]. From these sources, we identified recurring normative evaluation dimensions relevant to DiGA approval, quality assurance, and postmarket monitoring. These include requirements related to content quality, technical stability, usability, effectiveness, reimbursement, and user experience. Based on this review, 10 high-level categories were defined, and each was mapped to one or more regulatory criteria to ensure traceability between app-user feedback and legally relevant quality dimensions (Table 1). The framework was developed deductively, as the study aimed to assess user feedback against the quality dimensions that are legally binding for reimbursable DiGAs. During the human validation, the authors additionally assessed whether statements addressed aspects not covered by the framework.

Table . The 10 feedback categories, the user-reported content they capture, and their regulatory grounding.

Category	Content captured	Regulatory basis
Content	Informational, educational, and therapeutic material; evidence-based scientific soundness, up-to-dateness, readability	section 139e (2) SGB V[a]; section 5 (8) DiGAV[b]
Technology	Technical stability, software reliability, device compatibility, bugs, and system failures	section 139e (2) SGB V; section 5 (2) DiGAV; BfArM[c] guideline
Cost and reimbursement	Perceived cost; reimbursement through statutory health insurance	section 134 (1) SGB V
Login and registration	Initial access; ease of login and registration	section 5 (5) DiGAV
Prescription and approval	Obtaining or redeeming access codes; prescription, insurer approval, and diagnosis requirements	BfArM guideline (Sections 3.2.2, 2.1.2.2)
User experience (UX) and design	User interface, usability, visual, and design aspects (overall interaction beyond initial access)	section 5 (5) DiGAV
Support	Availability, responsiveness, and quality of customer or therapeutic support	BfArM guideline (Section 3.6.2.5)
Tracking and documentation	Documentation of symptoms, nutrition, activities, or behaviors	section 139e (13) SGB V (effective Jan 1, 2026)
Effectiveness	Perceived medical benefit, self-management, health literacy, and therapy adherence	section 139e (13) SGB V; section 8 (2)-(2) DiGAV
Overall impression	Overall subjective rating, perceived utility, and user trust	section 139e (13) No. 2 SGB V

[a]SGB V: German Social Code, Book V.

[b]DiGAV: Digital Health Applications Regulation.

[c]BfArM: German Federal Institute for Drugs and Medical Devices.

A more detailed overview of the legal sources and their mapping to categories is provided in Multimedia Appendix 1.

Model-Based Sentiment and Category Assignment

To systematically assess user sentiment, we developed a custom Python script to automate the classification process using the GPT-4o API provided by OpenAI (steps 1 - 3). GPT-4o was selected since previous research has shown that GPT-based LLMs can achieve strong zero-shot performance in sentiment classification of user-generated texts and online reviews [22]. Each user review was processed individually using a carefully structured prompt designed to guide the model toward consistent and domain-specific classification (Multimedia Appendix 2). The model was instructed to adopt the role of an expert in evaluating user experiences with DiGAs. The prompt was developed iteratively before the full-scale analysis. Initial prompt versions were piloted on a heterogeneous subset of reviews and refined through repeated inspection of ambiguous, inconsistent, or insufficiently specific model outputs. Revisions focused on clarifying category definitions, decision rules for mixed sentiment, and the required structured output format. This iterative prompt-development approach follows emerging methodological guidance emphasizing prompt specification, testing, and refinement for reliable LLM-supported classification tasks [23]. Within this role, it performed the following multistep tasks:

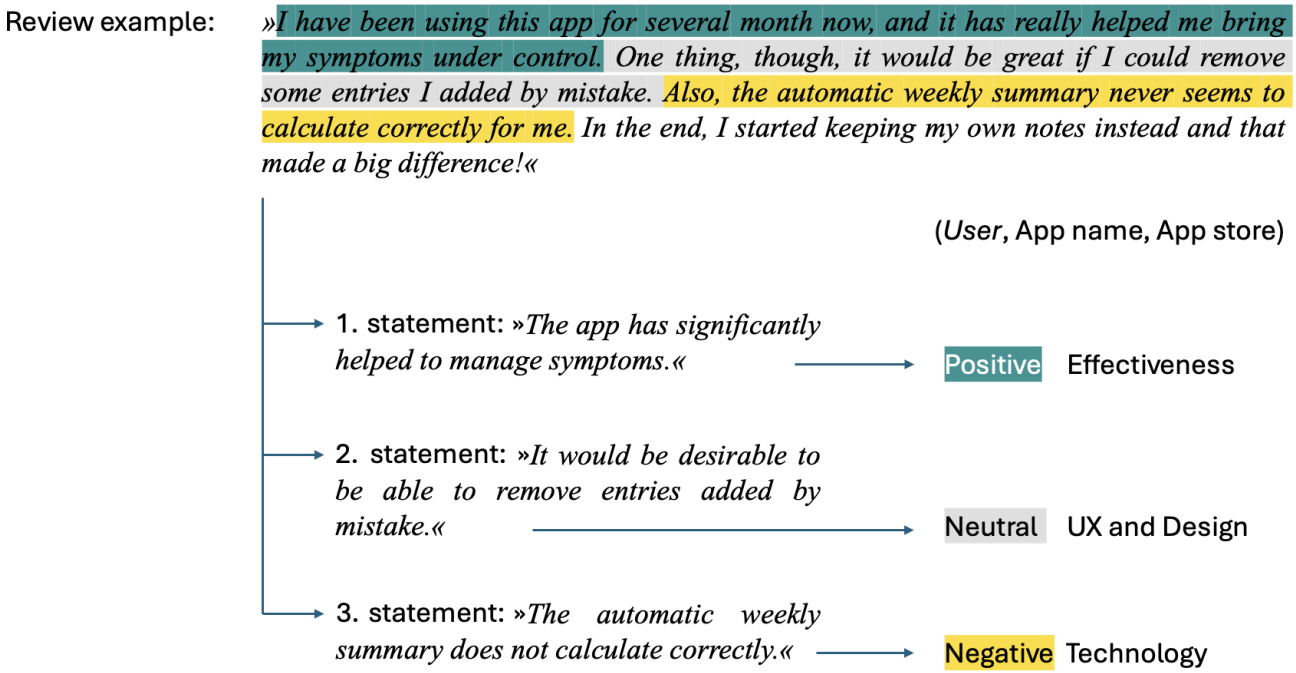
- Extract up to 5 core statements from each user review.
- Assign each statement a sentiment label (positive, neutral, or negative).

XSL•FO
RenderX

- Assign each statement to exactly one of the 10 predefined thematic categories derived from the legal and quality criteria for DiGAs (Multimedia Appendix 1).
- Merge statements that refer to the same theme and sentiment into a single, summarized expression.
- Avoid repeating partial aspects of the same issue across multiple statements.
- Limit each category to one statement per review, unless clearly distinct subtopics are present.
- Exclude overall evaluations if the review indicates that the app could not be used (eg, due to failed registration or missing insurance approval).

The model was provided with a fixed list of 10 valid categories, accompanied by specific guidance on ambiguous cases. For example, access issues after prescription or insurer approval were to be categorized as login and registration, not prescription and approval. User reviews may contain several distinct evaluative aspects. Extracting up to 5 core statements allowed the analysis to capture this multidimensionality while maintaining a consistent upper limit across reviews and preventing unusually long reviews from disproportionately influencing the dataset. Finally, the model was instructed to return its output in a standardized, structured format with numbered statements, each containing 3 elements: the *extracted statement*, *sentiment label*, and *assigned category*. A representative example was included in the prompt to enforce consistent formatting and interpretation. In order to assess whether an unsupervised topic-modeling approach could provide a suitable alternative to the predefined, prompt-based classification [24], we additionally conducted an exploratory BERTopic analysis following Uncovska et al [19]. BERTopic was applied to the complete corpus of user reviews to identify recurring themes without imposing predefined categories. The resulting topic clusters were assessed with respect to thematic coherence, clinical interpretability, and correspondence with the 10 predefined DiGA evaluation categories. In addition, we considered whether the approach could adequately represent multiple distinct issues within a single review. This exploratory comparison informed the selection of the final analytical approach. In the following processing steps, the script parsed the structured model output into a tabular dataset containing the extracted statements and their corresponding sentiment labels and thematic categories (steps 4 - 5). All results were exported in a standardized CSV format for further manual validation (step 6). By assigning the model a clearly defined expert perspective and domain logic, we ensured that the sentiment analysis aligned with both linguistic consistency and health service evaluation standards. A fictitious example of this classification process is shown in Figure 1.

Figure 1. Illustrative example of review classification process. UX: user experience.



Human Validation of Model Output

The accuracy and conceptual validity of the model-generated classifications were evaluated through a structured validation process, as shown in Figure 2. The full set of reviews, including the original review, extracted statement, sentiment label, and categories, was evenly divided among 3 authors (step 7), each of whom independently reviewed one-third of the dataset (step 8). The validation process followed a four-step protocol given below:

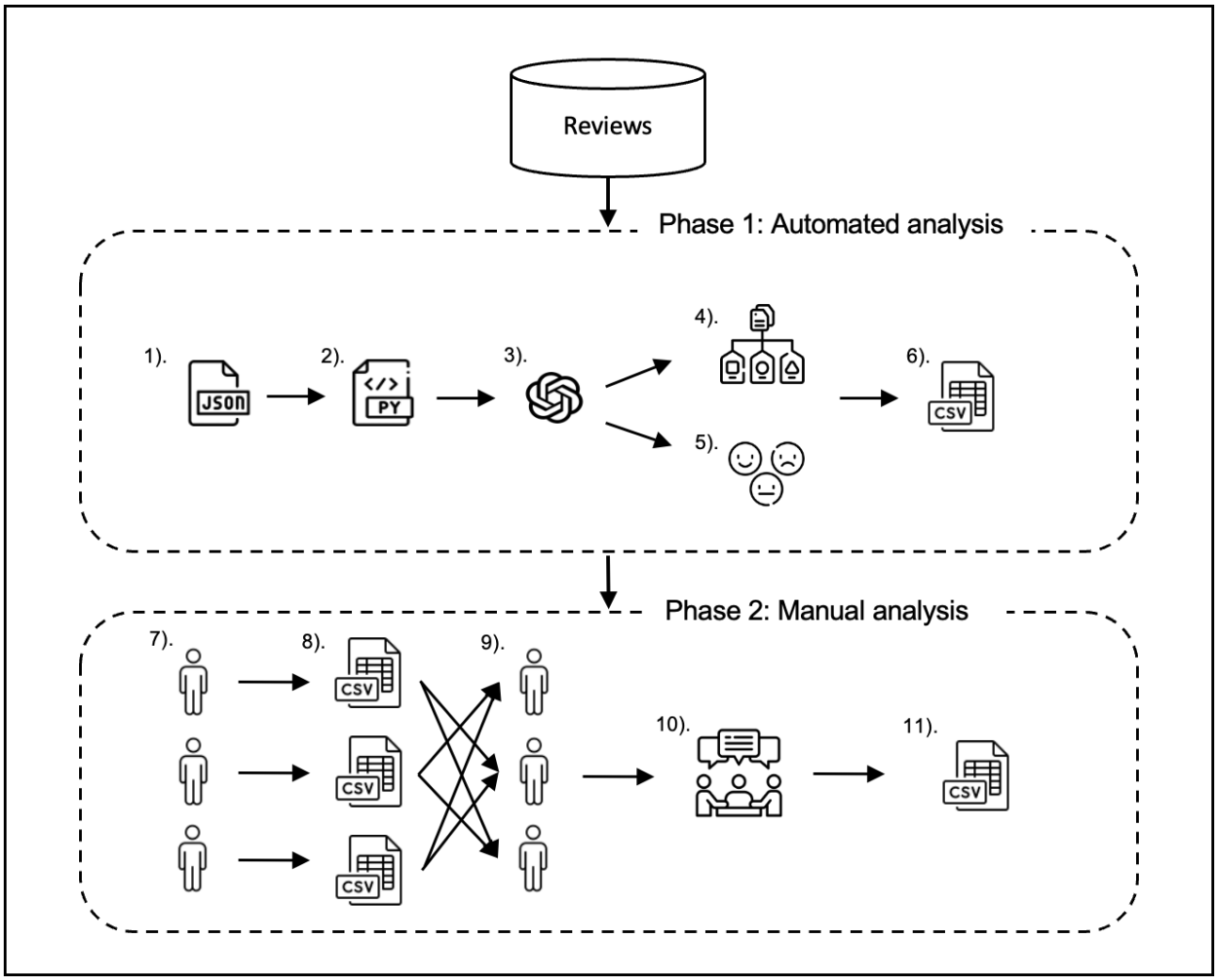
1. Independent review: Each model-generated statement and its assigned sentiment and category were assessed in relation to the original review. Reviewers evaluated whether the extracted statement captured a meaningful aspect of the review and whether the sentiment and category were accurate according to the predefined classification framework.
2. Flagging of inconsistencies: Statements that were incomplete, overly generic, redundant, or incorrectly classified (eg, in sentiment or category) were flagged. Each

XSL•FO
RenderX

flagged item included a brief comment and a suggested correction, if appropriate.

3. Triangulated reassessment: The 2 remaining authors independently evaluated the flagged items, providing their own classifications without being influenced by the previous assessments (step 9). This resulted in 3 independent judgments per flagged entry.
4. Consensus resolution: In cases of disagreement, the authors discussed each item until a final classification was agreed upon (step 10).

Figure 2. Overview of the 2-phase classification and sentiment analysis process, combining automated extraction and classification (phase 1) with manual validation and consensus (phase 2).



Validated statements were finalized and included in the analysis as a consolidated CSV dataset (step 11). Structurally flawed outputs (eg, empty responses or parsing errors) were reprocessed. The manual validation ensured consistency with the classification framework and enabled direct evaluation of the model on real-world data.

Data Analysis

Following the classification and validation process, a descriptive analysis was conducted. The proportions of positive, neutral, and negative statements were calculated for the full dataset and for each of the 10 categories. Data processing and descriptive analyses were performed in Python using the *pandas* library (version 2.2.3), while figures were generated using *matplotlib* (version 1.3.0). In addition, the validated statements within each category were qualitatively reviewed by the authors to identify recurring content patterns among the positive, neutral, and negative statements. The category-specific findings reported in the "Results" section, including descriptions of frequently mentioned issues, are based on this qualitative review. Because a single app accounted for a disproportionately high share of the dataset, a sensitivity analysis was performed in which all analyses were repeated after excluding this app to assess whether its disproportionate weight biased the overall results. In addition, because statements originating from the same app are unlikely to be fully independent, all analyses were additionally repeated using an app-level averaging approach. In this clustering-adjusted analysis, sentiment proportions were calculated separately for each app and then averaged across apps so that every app contributed equally regardless of the number of statements it provided.

Beyond the predefined statement-level classification, an exploratory BERTopic analysis was conducted on all included reviews using a multilingual sentence transformer model (paraphrase-multilingual-mpnet-base-v2) to assess whether unsupervised topic modeling could provide a suitable approach [24]. Topics were inspected qualitatively based on their top

XSL•FO
RenderX

terms and compared with the 10 predefined categories. Based on this assessment, BERTopic was not used for the final quantitative analysis. The full topic modeling output can be found in Multimedia Appendix 3.

Ethical Considerations

This study analyzed publicly accessible user reviews voluntarily submitted to the German Apple App Store and the Google Play Store. The study involved no interaction with review authors, no recruitment of participants, and no access to clinical records, nonpublic personal data, or other restricted information. The analysis was observational and based exclusively on publicly available product evaluations. Following the context-sensitive framework proposed by Eysenbach and Till [25], we considered not only the public accessibility of the reviews but also the nature of the online setting, the noninteractive study design, and potential risks to review authors. App store reviews were treated as publicly accessible, one-directional product evaluations rather than communications within a bounded virtual community. The reviews could be viewed without registration and did not involve an expectation of interpersonal exchange or membership in a restricted group. The analysis of publicly available app store reviews is an established approach in mHealth research and has been used to investigate user experience, usability, perceived benefits, and recurring barriers to app use [13,20-22]. To reduce potential privacy risks, analyses were conducted and reported primarily in aggregate form. No usernames, account identifiers, or other direct identifiers were included in the analytical dataset or reported in the manuscript. Because verbatim quotations may enable reidentification through search engines, no direct quotations from individual reviews are reported in this manuscript. Formal ethics review was not sought because the study involved a noninteractive analysis of publicly accessible app store reviews and did not involve the collection of nonpublic personal data.

Results

Overview of the Dataset

As of February 25, 2025, a total of 69 DiGAs were listed in the official directory maintained by the German BfArM. Of these, 45 DiGAs were available as mobile apps (iOS and Android), and 34 offered a web-based version, with several products available in both formats. Since DiGA-user reviews are only available for mobile apps distributed through public app stores, the analysis was limited to the mobile app versions of DiGAs. Web-based applications were excluded due to the lack of publicly accessible user feedback.

Although 45 DiGAs had a mobile app, 1 DiGA had no publicly available user reviews and could therefore not be analyzed. Thus, 44 DiGAs were included in the analysis. Among them, 29 (65.9%) had been permanently approved, and 15 (34.1%) were listed on a provisional basis at the time of data collection. It is important to note that several DTx share a common mobile app. For example, the *HelloBetter* app includes 6 distinct DiGAs addressing different indications (eg, chronic pain, diabetes, sleep disorders, panic, stress, and vaginismus), all delivered through the same app infrastructure. Similarly, the *Selfapy* app delivers 5 approved DiGAs (eg, for depression, bulimia nervosa, generalized anxiety disorder, and chronic pain) through a single mobile app. Accordingly, the 44 DiGAs available as mobile apps are represented by 35 mobile apps. Consequently, the app-based analysis cannot fully differentiate between user feedback for each individual DiGA when multiple products are bundled within one app. Nevertheless, all reviews associated with such apps were included in the analysis, as they reflect real-world user experiences with the respective DiGA ecosystem.

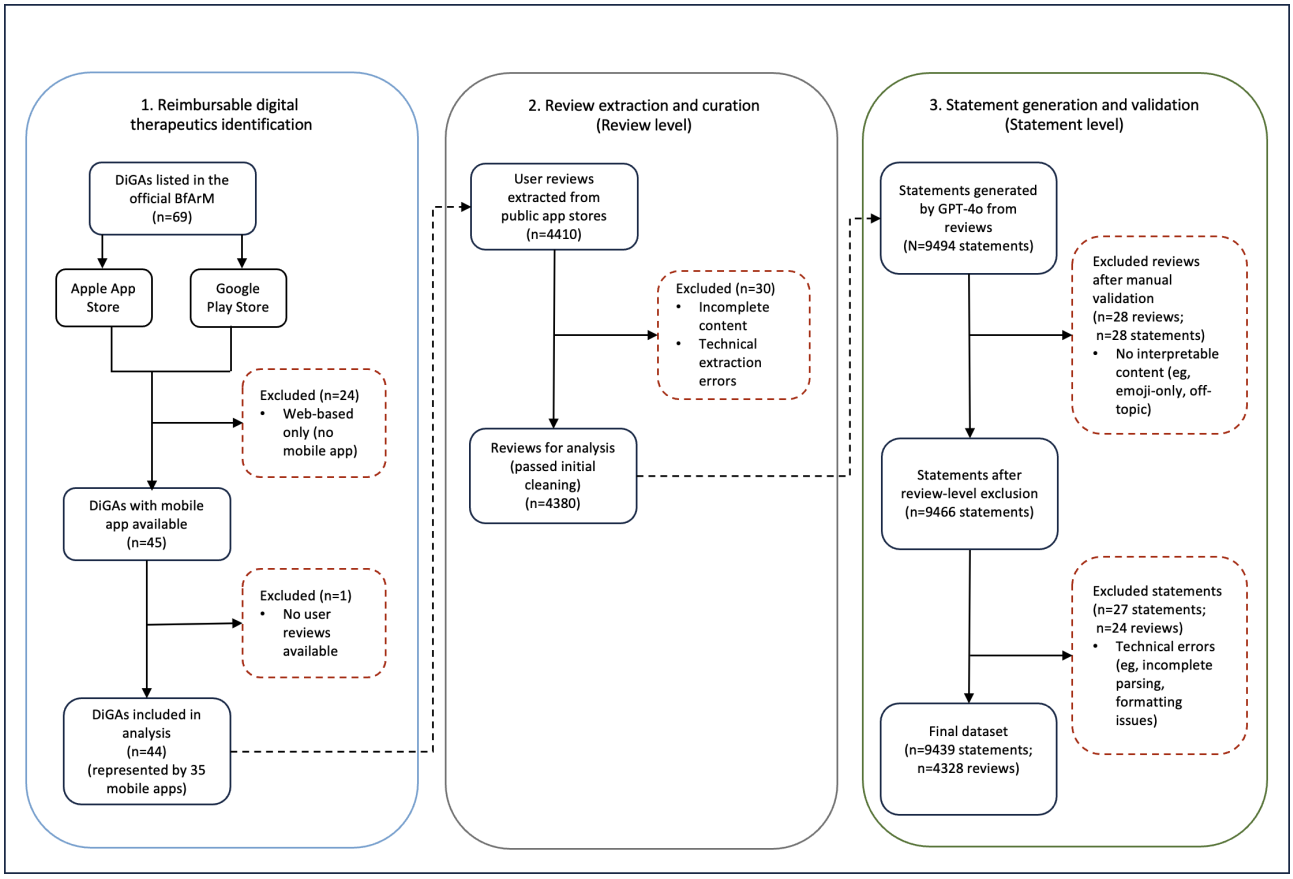
From these apps, 4410 publicly accessible user reviews were extracted via the browser-based versions of the German Apple App Store and the German Google Play Store. For the Apple App Store, only up to 10 written reviews are publicly visible per app when accessed via a browser. During data cleaning and preprocessing, 30 reviews could not be processed further due to incomplete content, parsing errors, or other technical parsing errors. These entries were excluded from the analysis. The remaining dataset of 4380 reviews was fully included in the quantitative and sentiment-based evaluation.

Validation of Model Output

From the 4380 included user reviews, a total of 9494 individual statements were extracted and systematically categorized. Each statement was assigned both a sentiment label (positive, neutral, or negative) and a thematic category. The initial classification was conducted by an LLM, followed by the manual validation. During this manual validation, 28 reviews were excluded due to irrelevance or ambiguous content. These reviews primarily included off-topic comments, incomplete sentences, and noninformative content unrelated to the DTx itself—such as emoji-only responses, promotional slogans, or vague remarks like “I want to be a hero. Greatings, Josef, “just installed,” or “don’t know yet.” In addition, 27 statements (corresponding to 24 reviews) were flagged as technically erroneous and removed. Statements addressing aspects outside the 10 predefined categories were rarely observed during validation. Together, manual-validation exclusions (28 reviews) and the technical exclusions (24 reviews) affected a total of 52 user reviews that were not included in the final dataset. As a result, the final dataset comprises 4328 user reviews containing at least one interpretable statement and a total of 9439 valid and interpretable statements. The complete data processing workflow, including all extraction, classification, and exclusion steps, is summarized in Figure 3. The complete classified dataset is provided in Multimedia Appendix 4. These 9439 statements form the basis for all subsequent quantitative analyses presented in this study. On average, a review consisted of 213.9 characters and 2.2 relevant statements. A detailed overview of the distribution of valid reviews and statements across individual apps can be found in Multimedia Appendix 4.

XSL•FO
RenderX

Figure 3. Workflow of review data identification, extraction, cleaning, and validation. DiGa: digital health application.



Agreement Between Initial Model Output and Human-Validated Classification

To assess the reliability of the automated sentiment and category classification, a systematic manual validation was performed. Agreement between the initial GPT-4o classifications and the final human-validated classifications was high. For sentiment labels, the overall agreement was 99%, with excellent precision and recall for positive (n=6491; F_1 score=1.00) and negative statements (n=2474; F_1-score=0.99) and for neutral statements (n=474; F_1-score=o.92). The sentiment confusion matrix (Figure 4a) shows minimal misclassifications, primarily between neutral and negative labels.

For category classification, agreement between the initial model output and the final human-validated classifications was 95%, with an average F_1-score of 0.95 (Figure 4b). Most thematic agreements were consistent with human reviewers' assessments, particularly for frequently occurring categories such as *overall impression* (n=1431; F_1-score=0.96), *content* (n=2564; F_1-score=0.95), *effectiveness* (n=1609; F_1-score=0.94), and *support* (n=865; F_1-score=0.98). However, lower agreement rates were observed for categories with fewer labeled examples, such as *prescription/approval* (n=234; F_1-score=0.89) and *login/registration* (n=280; F_1-score=0.92).

XSL•FO
RenderX

Figure 4. Confusion matrices for the model-based classification: (a) sentiment and (b) category.



Distribution of Sentiments and Categories

In total, 6491 out of 9439 (68.8) statements were classified as positive, 474 out of 9439 (5.0%) statements as neutral, and 2474 out of 9439 (26.2%) statements as negative. The majority of positive statements were assigned to the category *content* (n=2001, 30.8%), followed by *effectiveness* (n=1440, 22.2%) and *overall impression* (n=1233, 19.0%). Among the 474 neutral statements, most were related to *content* (n=145, 30.6%), *UX/design* (n=83, 17.5%), and *effectiveness* (n=68, 14.3%). Negative classifications (n=2474) occurred most frequently in the categories *Technology* (n=563, 22.8%), *UX/design* (n=340, 13.7%), and *login/registration* (n=263, 10.6%) (Figure 5a). A sensitivity analysis excluding the DTx Zanadio (Sidekick Health Germany GmbH), which accounted for 1433 of 4328 (33.1%) reviews, showed a slightly lower proportion of positive statements (65% vs 68.8% in the full dataset) but did not change the overall pattern of the results, with the same categories dominating positive and negative feedback (Figure 5b).

Figure 5. Sentiment distribution per category across all apps (a) and excluding Zanadio (b).



A clustering-adjusted analysis, in which sentiment proportions were averaged across apps rather than pooled across statements, showed comparable results (64.2% vs 65.0% [3963/6098] positive statements when excluding the most frequently reviewed app) and did not change the ranking of the most positively and most negatively evaluated categories. Larger deviations occurred only in the categories cost and reimbursement and support, in which the median number of statements per app was below 5, making estimates more sensitive to individual reviews. The full comparison is provided in Multimedia Appendix 5.

In addition to the aggregated results, sentiment distributions were also analyzed separately for each individual DTx included

XSL•FO
RenderX

in the dataset. The detailed graphical breakdown of sentiment and thematic categories per app is provided in Multimedia Appendix 6 for reference. This allows a more granular exploration of potential differences in user feedback across specific DTx products, beyond the aggregated trends presented in the "Result" section.

Category-Specific Findings

The following sections provide a detailed overview of user feedback for 3 selected categories: *overall impression*, *effectiveness*, and *login/registration*. These categories were chosen due to their high frequency and relevance across the dataset. Analyses of additional categories, such as *UX/design*, *tracking/documentation*, *technology*, *support*, *prescription/approval*, *cost/reimbursement*, and *content* are provided in Multimedia Appendix 1 for reference.

Overall Impression

A total of 1431 user statements were categorized under "overall impression", providing general user feedback regarding the DTx experience. Positive reviews frequently emphasized the usefulness, clarity, and motivational aspects of the apps. Several users described individual programs as motivating, adaptable to personal needs, and worth recommending. Others highlighted that an app helped them understand and manage specific health concerns and praised its intuitive, clearly structured design, or noted that exploring the relationship between their condition and an everyday factor was insightful and that the app was easy to use. Neutral sentiments reflected initial or uncertain user experiences, with some users indicating only a tentative first impression or that they were not yet convinced. Reviews with negative statements often criticized technical problems or unmet expectations, such as poor performance and a lack of personalization due to a nonadjustable algorithm, or noted that, while an app supported the documentation of symptoms, it did not substitute for therapy and provided no personal benefit.

Login/Registration

In total, 280 user statements addressed the login and registration process. Positive feedback frequently highlighted an easy onboarding experience and simple account handling, with users describing straightforward registration, clear profile setup, and quick login as features that made it easy to get started. Neutral reviews often paired a generally positive impression with specific suggestions for improvement, such as the wish to have their login data or the email address saved to reduce repeated entry. Negative experiences primarily concerned technical login problems or authentication procedures perceived as overly complex. Some users reported losing access to their account after an update and being frustrated by having to log in repeatedly, while others described persistent login failures independent of their email provider. Several users specifically objected to a 2-step login process, which they considered unnecessary for this type of app.

Effectiveness

The category "effectiveness" received 1609 user statements, making it one of the most frequently mentioned aspects of the DTx evaluations. Many users reported positive effects on their health, highlighting improved well-being, symptom relief, or successful integration of the app into daily routines. Users described being satisfied with the range of features and reported that an app had genuinely helped them or that a sleep-focused app had improved how quickly they fell asleep and how well they stayed asleep, while praising its clear structure. Neutral reviews often reflected mixed impressions or external limitations affecting the expected effectiveness, such as finding reminders simultaneously bothersome and necessary or struggling with personal motivation despite valuing the content. Others characterized an app as a helpful first step toward understanding their condition, one that nonetheless required discipline and could not replace professional medical care. Reviews with negative statements often expressed disappointment about insufficient health improvements, with some users noting that helpful tips ultimately did not work for them. A more critical assessment reported no improvement despite full completion of the program, criticized the high cost, reimbursed by health insurance, relative to the benefits, and described the app as poorly designed and unstable.

Discussion

Principal Results

This study presents a large-scale analysis of publicly available app store reviews for reimbursable DiGAs in Germany, based on 9439 classified user statements across 44 listed DTx. Our results show that user feedback was predominantly positive with 6491 of 9439 (68.8%) classified as positive. Positive sentiment was particularly concentrated in the categories effectiveness (1440/1609, 89.5% positive), overall impression (1233/1431, 86.2% positive), and content (2001/2565, 78.0% positive), indicating that users who submitted reviews frequently expressed favorable views of the therapeutic approach and content of these apps. At the same time, 2474 of 9439 (26.2%) statements were negative, with login and registration, technology, and prescription and approval emerging as the most frequently criticized dimensions, pointing to persistent barriers in access, technical reliability, and administrative processes. The high overall classification performance of the model (99% sentiment accuracy, 95% category accuracy) supports the reliability of these patterns.

The predominantly positive overall sentiment, particularly in the categories of effectiveness and content, suggests that many users may perceive DiGAs as therapeutically meaningful and that the content quality of many apps may meet user expectations. This finding should be interpreted in light of known review behavior, as app store reviews tend to be predominantly positive overall, with praise being the most frequently expressed topic [26]. The high proportion of positive statements regarding effectiveness is therefore best understood as an indicator of perceived, self-reported benefit rather than a validated therapeutic effect. Notably, perceived effectiveness from the user perspective has also gained regulatory attention, as the app-accompanying assessment under section 139e (13) SGB V has included patient-reported health status during the DiGA use since January 1, 2026 [3].

In contrast, the high share of negative statements in the categories technology, login and registration, and prescription

XSL•FO
RenderX

and approval points to structural barriers that undermine user access and sustained engagement. Technical malfunctions and app instability are particularly concerning given that DiGAs are regulated medical devices subject to quality and safety requirements under section 139e Abs 2 SGB V. Frequent user criticism of bugs and crashes suggests a gap between regulatory certification and real-world software quality, indicating a need for improved release and quality management by manufacturers [27]. From a requirements engineering perspective, technical stability and reliable access represent basic requirements or "must-be" factors whose absence reliably generates negative feedback, regardless of, for example, an app's therapeutic value [28].

The criticism directed at login and registration processes reflects a broader tension between security requirements and usability. Multifactor authentication, while aligned with national cybersecurity standards, appears to conflict with the low-threshold access that digital regulations expect for a user population with an average age of 55-60 years [6]. Similarly, the prescription and approval category reveals user frustration with the administrative burden of the reimbursement pathway, including lengthy insurer approval processes, eligibility restrictions, and the recurring 90-day prescription cycle. Some users additionally expressed a desire to trial apps before committing to a prescription, which aligns with broader arguments in the literature for more flexible access models in DTx [12]. This concern is in line with broader critiques of the DiGA framework, in which the activation and prescription process and the lack of trial options have been identified as barriers to acceptance among prescribing physicians [11].

Compared to Uncovska et al [19], who applied unsupervised BERTopic modeling to a mixed sample of regulated and nonregulated mHealth apps, our study focuses exclusively on officially listed DiGAs and maps user feedback to a predefined framework grounded in regulatory criteria. Rather than discovering latent topics, our approach systematically evaluates user experience along dimensions that are directly relevant to DiGA approval and postmarket monitoring. This regulatory alignment allows for more targeted interpretation of findings and facilitates direct translation into policy and quality improvement recommendations. However, such reviews cannot replace formal postmarket surveillance or systematically collected patient-reported outcome measures, as they are subject to selection bias and may not represent the broader user population [8,9].

Limitations

Several methodological limitations should be considered when interpreting these findings. App store reviews are voluntary and may be affected by self-selection bias, as users with particularly positive or negative experiences may be more likely to provide feedback, whereas neutral experiences may be underrepresented [26]. The reviews are submitted without a standardized clinical or usage context. Thus, they may reflect isolated incidents rather than sustained use, and cannot be considered representative of all DiGA users. In addition, app store reviews are not a curated or audited source, and prior research has identified potentially incentivized or inauthentic reviews in app stores [29]. No additional bot detection or fraud screening beyond the app stores' own moderation was applied, and neither users' identity nor patient status could be verified. They should therefore be interpreted as complementary insights into perceived user experience and usability rather than as a replacement for formal clinical evaluation or postmarket surveillance. The validation process was not based on a fully independent blinded double-coding design. The 3 reviewers assessed and corrected model-generated statements and labels in relation to the original review rather than independently annotating all reviews from scratch. Although this enabled the validation of a large dataset, it may have introduced anchoring or acceptance effects. The reported agreement measures therefore reflect concordance between the initial model output and the final human-validated dataset, rather than an independent estimate of model accuracy or interrater reliability. The predefined category framework, while based on legal, regulatory, and quality-related criteria for reimbursable DiGAs, may not fully capture user concerns outside these dimensions, such as social support, gamification, emotional well-being, or motivational factors. As the framework was derived from regulatory requirements, it captures the quality dimensions that are legally relevant for DiGAs, but not necessarily features that go beyond these requirements and may generate particular enthusiasm among users. Identifying the latter would require an inductive coding approach, which presents a promising direction for future research. In addition, reviews were available at the level of mobile apps rather than individual DiGAs. Because several apps bundled multiple DiGAs, feedback could not always be assigned reliably to a specific product, indication, or listing status. Moreover, a single app accounted for approximately one-third of all reviews, so the aggregated findings are weighted toward the most frequently reviewed apps. A sensitivity analysis excluding this app confirmed the overall sentiment distribution and category patterns, but the pooled percentages should be interpreted with this imbalance in mind.

Finally, the analysis did not investigate the temporal development of reviews in detail. Future longitudinal analyses could assess whether technology-related complaints were transient, for example following app launches or updates, or reflected persistent barriers [30].

Comparison With Prior Work

Our findings are consistent with prior app review research showing that technical issues, registration barriers, and usability problems are recurring concerns in mHealth apps [14,19]. However, by linking user feedback to regulation-informed categories, our study adds a different perspective: it shows which aspects of the DiGA approval framework are reflected in users' real-world experience and where practical barriers persist despite regulatory certification. In this sense, the study extends prior exploratory work by demonstrating how app store reviews can be used not only to identify general user concerns, but also to generate user-reported real-world evidence relevant to postmarket monitoring and regulatory quality improvement. Similar patterns emerge in other contexts. In an analysis of app store reviews for prescription DTx cleared by the US Food and Drug Administration, access barriers, technical problems, and engagement limitations were among the most frequent themes

XSL•FO
RenderX

[20]. A thematic analysis of 13,549 reviews of mental health apps identified poor usability as the most common reason for app abandonment, along with registration requirements perceived as excessive [21]. Overall sentiment in the US sample was predominantly negative, in contrast to our largely positive findings, which may reflect differences in reimbursement or prescription pathways.

Conclusions

This study provides a large-scale overview of how users experience and evaluate reimbursable DTx in Germany, based on publicly available app store reviews. The findings show that publicly available app store reviews can reveal practical barriers that affect whether reimbursable DTx are accessible, usable, and acceptable in everyday care. Although user feedback was predominantly positive, especially regarding perceived effectiveness, overall impression, and content, recurring criticism of login and registration processes, technical reliability, and prescription and approval procedures indicates that users continue to encounter barriers at key points of access and use.

The practical relevance of these findings lies in the fact that many of the criticized aspects are not peripheral convenience issues, but basic prerequisites for sustained engagement with medical-grade digital health apps. Technical stability, reliable access, understandable onboarding, and low-threshold registration should therefore be treated as essential implementation requirements. App store reviews may help regulators, manufacturers, and researchers identify such barriers early and complement formal evaluation procedures with user-centered real-world feedback. LLM-assisted approaches can support this process by enabling large volumes of unstructured user feedback to be analyzed in a structured and scalable way. As the DiGA framework matures, systematically incorporating user-reported real-world evidence into postmarket surveillance could help ensure that regulatory certification translates into DTx that are not only proven to be effective in clinical trials but also accessible and usable in everyday care.

Acknowledgments

The authors thank the Institute of Medical Informatics and Artificial Intelligence (MIKI), Kiel University, and the University Hospital Schleswig-Holstein, for providing the infrastructure and resources that supported this work.

The authors declare the use of generative AI (GenAI) in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: code generation, code optimization, process automation, creation of algorithms for data analysis, data analysis, proofreading and editing, and translation. GPT-4o (OpenAI) was used for automated sentiment analysis and thematic categorization of app store reviews, code generation and optimization, and language editing of the manuscript. DeepL (July 2025) was used exclusively for the translation of user review quotes from German to English. The scientific conceptualization, literature review, data interpretation, and conclusions are entirely the work of the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Funding

The authors declare no financial support was received for this work.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files. User names extracted from public app store reviews were replaced with pseudonyms prior to publication. A mapping table linking pseudonyms to original user names is retained by the corresponding author and is available on reasonable request. The analysis code is not publicly available, as it was developed specifically for this study and is not maintained for external use, but it is available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: BK
Methodology: BK, HU
Software: BK, HU
Data curation: BK
Formal analysis: BK
Investigation: BK
Validation: BK, HR, HU
Visualization: BK
Writing – original draft: BK
Writing – review & editing: BK, BS, HU
All authors read and approved the final manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Legal basis and results for user feedback categories in DiGA app store reviews.
[PDF File, 333 KB - humanfactors_v13i1e92415_app1.pdf]

Multimedia Appendix 2
LLM-Prompt.
[PDF File, 108 KB - humanfactors_v13i1e92415_app2.pdf]

Multimedia Appendix 3
BERTopic modeling output.
[PDF File, 64 KB - humanfactors_v13i1e92415_app3.pdf]

Multimedia Appendix 4
App store user reviews.
[XLSX File, 1209 KB - humanfactors_v13i1e92415_app4.xlsx]

Multimedia Appendix 5
Clustering-adjusted analysis.
[PDF File, 433 KB - humanfactors_v13i1e92415_app5.pdf]

Multimedia Appendix 6
Category and sentiment analysis per App.
[PDF File, 104 KB - humanfactors_v13i1e92415_app6.pdf]

References

1. Driving the digital transformation of Germany's healthcare system for the good of patients. Federal Ministry of Health. URL: https://www.bundesgesundheitsministerium.de/en/digital-healthcare-act [accessed 2025-08-11]
2. Schmidt L, Pawlitzki M, Renard BY, Meuth SG, Masanneck L. The three-year evolution of Germany's Digital Therapeutics reimbursement program and its path forward. NPJ Digit Med 2024 May 24;7(1):139. [doi: 10.1038/s41746-024-01137-1] [Medline: 38789620]
3. Sozialgesetzbuch (SGB) Fünftes Buch (V) - Gesetzliche Krankenversicherung - (Artikel 1 des Gesetzes v. 20. Dezember 1988, BGBl. I S. 2477)§ 139e Verzeichnis für digitale Gesundheitsanwendungen; Verordnungsermächtigung [Article in German]. Bundesministerium der Justiz und für Verbraucherschutz, Bundesamt für Justiz. URL: https://www.gesetze-im-internet.de/sgb_5/__139e.html [accessed 2026-06-30]
4. DiGA-Verzeichnis [Article in German]. Bundesinstitut für Arzneimittel und Medizinprodukte. URL: https://diga.bfarm.de/de/verzeichnis [accessed 2025-03-15]
5. Verordnung über das Verfahren und die Anforderungen zur Prüfung der Erstattungsfähigkeit digitaler Gesundheitsanwendungen in der gesetzlichen Krankenversicherung (Digitale Gesundheitsanwendungen-Verordnung - DiGAV)§ 8 Begriff der positiven Versorgungseffekte [Article in German]. Bundesministerium der Justiz und für Verbraucherschutz, Bundesamt für Justiz. 2020. URL: https://www.gesetze-im-internet.de/digav/__8.html [accessed 2025-05-10]
6. Bericht nach § 139e Absatz 10 SGB V zur Nutzung, Akzeptanz und Wirkung digitaler Gesundheitsanwendungen – DiGA-Bericht 2024 [Report in German]. : GKV-Spitzenverband; 2025 URL: https://www.gkv-spitzenverband.de/media/dokumente/krankenversicherung_1/telematik/digitales/2024_DiGA-Bericht_final.pdf [accessed 2026-09-20]
7. Mäder M, Timpel P, Schönfelder T, et al. Evidence requirements of permanently listed digital health applications (DiGA) and their implementation in the German DiGA directory: an analysis. BMC Health Serv Res 2023 Apr 17;23(1):369. [doi: 10.1186/s12913-023-09287-w] [Medline: 37069592]
8. Bolinger E, Tyl B. Key considerations for designing clinical studies to evaluate digital health solutions. J Med Internet Res 2024 Jun 17;26:e54518. [doi: 10.2196/54518] [Medline: 38885020]
9. Guo C, Ashrafian H, Ghafur S, Fontana G, Gardner C, Prime M. Challenges for the evaluation of digital health solutions-a call for innovative evidence generation approaches. NPJ Digit Med 2020;3(1):110. [doi: 10.1038/s41746-020-00314-2] [Medline: 32904379]
10. Sunyaev A, Fürstenau D, Davidson E. Reimagining digital health. Bus Inf Syst Eng 2024 Jun;66(3):249-260. [doi: 10.1007/s12599-024-00870-x]

XSL•FO
RenderX

11. Schlieter H, Kähling M, Hickmann E, et al. Digitale Gesundheitsanwendungen (DiGA) im Spannungsfeld von Fortschritt und Kritik : Diskussionsbeitrag der Fachgruppe „Digital Health“ der Gesellschaft für Informatik e. V [Article in German]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2024 Jan;67(1):107-114. [doi: 10.1007/s00103-023-03804-2] [Medline: 38086924]
12. Schwartz DG, Spitzer S, Khalemsky M, et al. Apps don’t work for patients who don’t use them: towards frameworks for digital therapeutics adherence. Health Policy Technol 2024 Jun;13(2):100848. [doi: 10.1016/j.hlpt.2024.100848]
13. Kendziorra J, Seerig KH, Winkler TJ, Gewald H. From awareness to integration: a qualitative interview study on the impact of digital therapeutics on physicians’ practices in Germany. BMC Health Serv Res 2025 Apr 18;25(1):568. [doi: 10.1186/s12913-025-12656-2] [Medline: 40247268]
14. Haggag O, Grundy J, Abdelrazek M, Haggag S. A large scale analysis of mHealth app user reviews. Empir Softw Eng 2022;27(7):196. [doi: 10.1007/s10664-022-10222-6] [Medline: 36246486]
15. Vasa R, Hoon L, Mouzakis K, Noguchi A. A preliminary analysis of mobile app user reviews. Proc 24th Aust Comput-Human Interact Conf 2012 Nov 26:241-244. [doi: 10.1145/2414536.2414577]
16. Kolominsky-Rabas PL, Tauscher M, Gerlach R, Perleth M, Dietzel N. Wie belastbar sind Studien der aktuell dauerhaft aufgenommenen digitalen Gesundheitsanwendungen (DiGA)? Methodische Qualität der Studien zum Nachweis positiver Versorgungseffekte von DiGA [Article in German]. Z Evid Fortbild Qual Gesundhwes 2022 Dec;175:1-16. [doi: 10.1016/j.zefq.2022.09.008] [Medline: 36437182]
17. Xiaozhou L, Zheying Z, Kostas S. Mobile app evolution analysis based on user reviews. In: Fujita H, Herrera-Viedma E, editors. New Trends in Intelligent Software Methodologies, Tools and Techniques: IOS Press; 2018:773-786. [doi: 10.3233/978-1-61499-900-3-773]
18. Das Fast-Track-Verfahren für digitale Gesundheitsanwendungen (DiGA) nach § 139e SGB V [Report in German]. : Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM); 2023 Dec URL: https://www.bfarm.de/SharedDocs/Downloads/DE/Medizinprodukte/diga_leitfaden.pdf?__blob=publicationFile [accessed 2025-07-07]
19. Uncovska M, Freitag B, Meister S, Fehring L. Rating analysis and BERTopic modeling of consumer versus regulated mHealth app reviews in Germany. NPJ Digit Med 2023 Jun 21;6(1):115. [doi: 10.1038/s41746-023-00862-3] [Medline: 37344556]
20. Lakhan SE. Sentiment and thematic analysis of user reviews for FDA-cleared prescription digital therapeutics: a mixed-methods real-world evidence study. Cureus 2025 May;17(5):e84710. [doi: 10.7759/cureus.84710] [Medline: 40416910]
21. Alqahtani F, Orji R. Insights from user reviews to improve mental health apps. Health Informatics J 2020 Sep;26(3):2042-2066. [doi: 10.1177/1460458219896492] [Medline: 31920160]
22. Krugmann JO, Hartmann J. Sentiment analysis in the age of generative AI. Cust Need and Solut 2024 Dec;11(1):3. [doi: 10.1007/s40547-024-00143-4]
23. Homiar A, Thomas J, Ostinelli EG, et al. Development and evaluation of prompts for a large language model to screen titles and abstracts in a living systematic review. BMJ Ment Health 2025 Jul 22;28(1):e301762. [doi: 10.1136/bmjment-2025-301762] [Medline: 40701625]
24. Tejani AS, Ng YS, Xi Y, Fielding JR, Browning TG, Rayan JC. Performance of multiple pretrained BERT models to automate and accelerate data annotation for large datasets. Radiol Artif Intell 2022 Jul;4(4):e220007. [doi: 10.1148/ryai.220007] [Medline: 35923377]
25. Eysenbach G, Till JE. Ethical issues in qualitative research on internet communities. BMJ 2001 Nov 10;323(7321):1103-1105. [doi: 10.1136/bmj.323.7321.1103] [Medline: 11701577]
26. Pagano D, Maalej W. User feedback in the appstore: an empirical study. 21st IEEE Int Requirements Eng Conf 2013:125-134. [doi: 10.1109/RE.2013.6636712]
27. Schramm L, Carbon CC. Critical success factors for creating sustainable digital health applications: a systematic review of the German case. Digit Healt 2024;10:20552076241249604. [doi: 10.1177/20552076241249604] [Medline: 38665883]
28. Hull E, Jackson K, Dick J. Requirements Engineering, 2nd edition: Springer; 2005.
29. Martens D, Maalej W. Towards understanding and detecting fake reviews in app stores. Empir Software Eng 2019 Dec;24(6):3316-3355. [doi: 10.1007/s10664-019-09706-9]
30. ISO/IEC/IEEE International Standard - systems and software engineering -- life cycle processes -- requirements engineering. : International Organization for Standardization (ISO), International Electrotechnical Commission (IEC), and Institute of Electrical and Electronics Engineers (IEEE); 2018 Nov URL: https://www.iso.org/standard/72089.html [accessed 2026-09-16]

Abbreviations

BfArM: Bundesamt für Arzneimittel und Medizinprodukte
DiGA: digital health application
DiGAV: Digital Health Application Regulation
DTx: digital therapeutics
LLM: large language model

XSL•FO
RenderX

mHealth: mobile health
SGB V: German Social Code, Book V
UX: user experience

Edited by S Law; submitted 29.Jan.2026; peer-reviewed by D Schwartz, S Bidmon; revised version received 31.Jul.2026; accepted 17.Aug.2026; published 08.Oct.2026.

Please cite as:
Kinast B, Rohde H, Schreiweis B, Ulrich H
What Users Say About Reimbursable Digital Therapeutics in Germany: Large-Scale App Store Review Analysis Using a Large Language Model
JMIR Hum Factors 2026;13:e92415
URL: https://humanfactors.jmir.org/2026/1/e92415
doi:10.2196/92415

© Benjamin Kinast, Henrik Rohde, Björn Schreiweis, Hannes Ulrich. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 8.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Smartphone-Based Ecological Momentary Assessment to Monitor Opioid Use and Overdose Among People Who Use Opioids: Prospective Observational Feasibility Study

Kamal Gautam[1], PhD; Kiran Paudel[1], MS; Jeffrey A Wickersham[2], PhD; Antoine Khati[1], MD; Anushka Thapa[2], MPH; Sandesh Bhusal[1], MS; Md Safaet Hossain Sujan[1], MPH; Sherry Pagoto[1], PhD; Toan Ha[3], DrPH; Roman Shrestha[1,2], MPH, PhD

[1]Department of Allied Health Sciences, University of Connecticut, 358 Mansfield Road, Unit 1101, Storrs, CT, United States

[2]Section of Infectious Diseases, Department of Medicine, Yale School of Medicine, Yale University, New Haven, CT, United States

[3]Department of Infectious Diseases and Microbiology, School of Public Health, University of Pittsburgh, Pittsburgh, PA, United States

Corresponding Author:
Roman Shrestha, MPH, PhD
Department of Allied Health Sciences, University of Connecticut, 358 Mansfield Road, Unit 1101, Storrs, CT, United States

Abstract

Background: Opioids account for 76% of drug overdose deaths in the United States, with nearly 80,000 opioid overdose deaths annually. The risk of overdose is dynamic and influenced by rapidly changing behaviors and contexts that are not well captured by retrospective or infrequent assessments. Ecological momentary assessment (EMA) allows repeated near real-time reporting of behaviors and experiences in natural settings.

Objective: This study evaluated the feasibility and acceptability of smartphone-based EMA among people who use opioids to monitor opioid use, overdose experiences, and naloxone access.

Methods: Participants were recruited through the New Haven Syringe Services Program in New Haven, Connecticut, and completed fixed-time, twice-daily EMA prompts for 30 days using a smartphone app. EMA measures assessed behaviors and experiences occurring within the previous 12 hours, including opioid craving, plans to use drugs, opioid use, overdose experiences, and naloxone access. Feasibility was assessed through recruitment, retention, and EMA completion rates, as well as the reporting of drug use-related behaviors. Acceptability was evaluated using exit surveys assessing ease of use, burden, and privacy, as well as qualitative interviews.

Results: Of the 13 screened individuals, 10 were enrolled and completed the baseline assessment. A total of 9 (90%) participants completed EMA and responded to 411 prompts, with an overall compliance rate of a mean of 85.0% (SD 8.7%). Compliance was similar across morning (84.2%, SD 13.1%) and evening (85.8%, SD 11.2%) prompts. Participants reported opioid cravings (274 reports), plans to use illicit drugs (252 reports), and opioid use (411 reports) during the study period. Two participants reported overdose events (5 reports) and were carrying naloxone during each overdose event. Acceptability ratings were high: all participants reported that EMA prompts were easy to understand and private, and most did not find them burdensome. Qualitative feedback further highlighted the ease of use, integration into daily routines, and increased self-awareness, although some participants reported emotional discomfort when reflecting on their substance use.

Conclusions: Smartphone-based EMA was feasible and acceptable among people who use opioids recruited through a syringe service program. EMA may support the monitoring of opioid use, overdose experiences, and naloxone access, and it may inform future digital health interventions aimed at reducing harm, including overdose risk, among people who use opioids.

(JMIR Hum Factors 2026;13:e95655) doi:10.2196/95655

KEYWORDS

ecological momentary assessment; opioid overdose; naloxone; harm reduction; syringe service programs; substance use; mobile health

Introduction

Opioid overdose remains a major public health challenge in the United States, with a high overdose mortality rate [1,2]. In 2023, more than 72,000 opioid-involved overdose deaths were reported, largely involving illegally manufactured fentanyl [3], and an estimated 5.6 million individuals aged 12 years or older met the criteria for opioid use disorder [4]. Although evidence-based treatments and harm-reduction strategies, such as expanded access to medications for opioid use disorder and

XSL•FO
RenderX

syringe service programs (SSPs), have expanded [5,6], overdose risk remains high.

People who use opioids often experience fluctuations in drug use, craving, affective states, and contextual exposures, which can influence when and how overdoses occur [7,8]. Prior research has identified factors such as polysubstance use, reduced tolerance, environmental context, and the need for assistance with injections as contributors to overdose risk. Structural and psychosocial vulnerabilities, including homelessness, incarceration, unemployment, psychiatric comorbidity, and suicidal ideation or behavior, have also been linked to increased overdose risk [9,10]. However, most evidence has relied on retrospective self-reports or infrequent clinical assessments [9,11]. These approaches have limited ability to capture the dynamic and context-dependent nature of overdose risk in daily life among people who use opioids, many of whom experience unstable housing [12], unemployment [13], and inconsistent engagement with health care, including early treatment drop-out [14,15]. As a result, clinic-based assessments may not adequately reflect the dynamic conditions and circumstances under which opioid use and overdose events occur.

Ecological momentary assessment (EMA) provides one approach to address this limitation [16,17]. EMA involves collecting repeated self-reports of behaviors and experiences over short time intervals in natural settings as individuals go about their daily lives, thereby reducing recall bias and improving ecological validity [18,19]. Previous research has shown high levels of smartphone access, device ownership, and digital literacy among people who use opioids, including individuals experiencing housing instability, and that EMA has been successfully implemented among these populations [18,20-22]. However, existing EMA studies among this population have primarily focused on momentary predictors of drug use, such as craving, mood, and stress [17,18,23]. Less attention has been given to the use of EMA to monitor overdose-related experiences and naloxone availability and use, and gaps remain regarding the feasibility and acceptability of EMA protocols in populations facing significant social and structural vulnerabilities, such as people who use opioids.

To address these gaps, this pilot study evaluated the feasibility and acceptability of a smartphone-based EMA protocol for people who use opioids for monitoring opioid use, overdose experiences, and naloxone access in daily life. Feasibility was assessed through recruitment, retention, and EMA completion rates, whereas acceptability was assessed through poststudy surveys and qualitative exit interviews. The EMA protocol was designed to capture day-to-day opioid-related experiences and behaviors, including craving, opioid use, overdose experiences, and naloxone access and use through fixed twice-daily EMA prompts. The findings of this study are intended to inform the design and implementation of future EMA-based research and digital interventions targeting opioid use and overdose risk.

Methods

Study Design and Setting

The Smartphone Health Innovation Program was a prospective, observational pilot study that evaluated the feasibility and acceptability of repeated smartphone-based EMA among people who use opioids. The study was conducted in New Haven, Connecticut, at the New Haven Syringe Services Program (NHSSP). The NHSSP is a community harm-reduction program that provides sterile syringe exchange, hygiene supplies, overdose-prevention education, naloxone distribution, and referrals to medical and substance-use treatment services for people who use opioids. Participants were followed for 30 days and completed an in-person baseline visit, a 30-day EMA monitoring period, a structured support check-in approximately 7 days after enrollment, and an end-of-study visit with a qualitative interview.

Recruitment, Eligibility, and Enrollment

Participants were recruited using convenience sampling through flyers posted at the NHSSP and surrounding community locations, word-of-mouth referrals, and referrals from local service organizations in the New Haven area (including treatment and harm-reduction programs, such as the APT Foundation, Inc). Potential participants either approached research staff while visiting the NHSSP for services (eg, syringe exchange visits) or contacted the study team by phone. Eligibility screening was conducted in person at the NHSSP or by telephone. Participants were eligible if they met the following criteria: (1) were aged 18 years or older, (2) reported use of at least one illicit opioid (eg, heroin and fentanyl) within the past 30 days, (3) owned a personal smartphone compatible with the study app (Android or iOS), and (4) resided in the Greater New Haven area.

Eligible participants attended an in-person enrollment visit at the NHSSP. Trained research assistants confirmed eligibility and obtained written informed consent in a private office. Participants then completed baseline questionnaires and onboarding procedures during the same visit. The enrollment and onboarding visit lasted approximately 30 minutes. Baseline questionnaires were administered electronically on a tablet and were primarily self-administered. Research staff read questions aloud and assisted participants when literacy or comprehension barriers were present. Participants were enrolled between November 2024 and February 2025. They completed 30 days of EMA monitoring following enrollment, and end-of-study assessments were conducted between January 2025 and March 2025.

Study Procedures

EMA Protocol Development

EMA prompts were developed based on prior EMA research among people who use substances and were informed by the existing literature on overdose, opioid use, and harm-reduction behaviors [18,23]. The study team and community partners reviewed the prompts to ensure their relevance and appropriateness for the study population and setting.

XSL•FO
RenderX

EMA Monitoring

EMA was conducted using the mEMA app from Ilumivu, a platform designed for smartphone-based EMA studies. Participants received 2 fixed-time-based EMA prompts per day between 9 AM and 10 PM for 30 consecutive days (60 scheduled assessments) on their smartphones. Participants selected their preferred morning and evening times. The prompts were scheduled approximately 12 hours apart to capture nonoverlapping recall periods. This design was selected to support feasibility and engagement by reducing participant burden and allowing integration of EMA into their daily routines [24]. Each EMA remained active for 120 minutes. If unanswered, automated reminders were sent every 30 minutes (a maximum of 4 reminders) for each survey. After 120 minutes, the prompts expired. EMA prompts were brief and required approximately 1 to 2 minutes to complete. Research staff periodically reviewed incoming responses to ensure that prompts were received and completed and to troubleshoot technical problems with participants.

Day-7 Support Contact

Approximately 7 days after enrollment, the research staff conducted structured telephone check-ins. The staff checked on the participants' safety and well-being, asked whether they were receiving notifications, experiencing difficulty charging their phones, or having problems answering EMAs, and reminded them to keep their phones charged and respond to EMAs promptly.

End-of-Study Visit

After 30 days, participants returned to the NHSSP to complete a poststudy questionnaire and a qualitative exit interview describing their experiences using the app to respond to the daily EMA prompts.

Study Measures

EMA Measures

EMA items assessed recent behaviors, risk exposures, and internal states over the prior 12 hours. Participants reported substance use in response to the question, "In the last 12 hours, have you used any drug(s), legal or illegal, that are not prescribed to you?" If the answer was "yes," participants selected the specific drugs used. Additional "yes" or "no" items assessed drug craving ("In the last 12 hours, did you experience any cravings for drug(s)?"), anticipated use ("Do you plan to use any drug(s) in the next 12 hours?"), sexual activity ("In the last 12 hours, have you had vaginal or anal sex with another person?"), overdose exposure ("In the past 12 hours, did you overdose on any drug(s)?" and "In the past 12 hours, did you witness or see another person overdose on drug(s)?"), and naloxone access ("Do you currently have Narcan available to you or someone else, if needed?"). All responses were coded as binary indicators for each EMA item.

Participants also rated affective and physical states using 0-10 numeric rating scales, including happiness ("In the last 12 hours, how happy were you?"), stress ("In the last 12 hours, how stressed were you?"), and physical pain ("In the last 12 hours, how intense was your physical pain?"). Higher scores indicated greater intensity of the respective state.

Baseline Measures

Sociodemographic Characteristics

Participants reported their age, sex, race or ethnicity, education, relationship status, employment status, monthly income, and current housing status.

Substance Use and Overdose History

Participants reported drug use in the past 12 months and the past 30 days, as well as the frequency of opioid use in the past 30 days. The route of administration was categorized as injection, smoking, snorting, oral use, or other. Lifetime overdose was assessed using a structured question that asked whether the participant had ever experienced an opioid overdose. The questionnaire provided symptom examples, including slowed or stopped breathing, blue lips or fingernails, unconsciousness, and inability to talk, to standardize interpretation. Participants also reported past 30-day and lifetime overdoses, as well as whether they had ever witnessed others overdose.

Substance Use Treatment Engagement

Participants were asked whether they were currently enrolled in a treatment program for opioid dependence (methadone, buprenorphine, extended-release naltrexone, or other), whether they had sought treatment in the past 12 months, and whether they had seen a substance use treatment provider in the past 6 months.

Mental Health

Depressive symptoms were measured using the Patient Health Questionnaire-2 [25], a brief, validated screener for depressive symptoms that has been used in drug and alcohol treatment populations [26]. It assesses *how often over the past 2 weeks participants were bothered by little interest or pleasure in doing things* and *feeling down, depressed, or hopeless*. Scores range from 0 to 6, with a cutoff score of ≥3 indicating moderate-to-severe depressive symptoms (Cronbach α=0.83).

Anxiety symptoms were measured using the Generalized Anxiety Disorder 7-item [27], a validated measure of anxiety symptoms that has been used in substance-using populations [28]. The Generalized Anxiety Disorder 7-item assesses how often over the past 2 weeks participants were bothered by 7 symptoms: feeling nervous, anxious, or on edge; not being able to stop or control worrying; worrying too much about different things; trouble relaxing; being so restless that it is hard to sit still; becoming easily annoyed or irritable; and feeling afraid as if something awful might happen. Scores range from 0 to 21, with scores of 10 or more indicating moderate-to-severe anxiety symptoms (Cronbach α=0.94).

Suicidal thoughts and behaviors were assessed using the suicidality module of the World Mental Health Composite International Diagnostic Interview. Participants reported lifetime suicidal ideation, suicide plans, and suicide attempts [29]. Participants also reported whether they had recent contact with a mental health provider (yes or no items). Current pain and

XSL•FO
RenderX

stress levels were assessed using a visual analog scale ranging from 0 to 100, with higher scores indicating greater perceived pain or stress.

Digital Health Literacy

Digital health literacy was assessed using the 3-item Digital Health Care Literacy Scale [30]. Participants rated their agreement with statements indicating that they could independently: (1) use videoconferencing applications, (2) set up a video chat, and (3) solve basic technical problems. Responses were rated on a 5-point Likert agreement scale, ranging from "strongly disagree" to "strongly agree," with a total score ranging from 0 to 12.

Feasibility and Acceptability Outcomes

Feasibility outcomes included recruitment, retention, and EMA compliance. Retention was defined as the completion of the end-of-study visit. EMA compliance was calculated as the proportion of completed prompts out of the total number of expected prompts among participants who initiated EMA monitoring. Morning and evening compliance were calculated separately, and to examine sustained engagement, completion rates were also summarized across 4 weekly intervals (days 1-7, 8-14, 15-21, and 22-30). Consistent with prior EMA literature among people who use substances, compliance rates of approximately 70% to 80% were considered indicative of acceptable feasibility [23,31-33].

Acceptability was assessed using a structured poststudy questionnaire. Participants rated statements such as "The daily EMA questions were easy to understand," "I had privacy and confidentiality while answering the EMA prompts," "The number of daily EMA was burdensome," and their willingness to participate again using multilevel agreement scales. Participants also reported the emotional impact (positive, neutral, or negative), the convenience of EMA timing, the reliability of notifications, the helpfulness of research team reminders, and barriers to responding, including being busy, sleeping, missing notifications, not having a phone, or experiencing technical problems.

Qualitative exit interviews further explored participants' experiences and perceived barriers to daily monitoring. The interview guide was developed by the study team based on the study objectives and prior EMA literature [34]. Participants were asked about their overall experiences and motivations for completing the daily EMA prompts, including perceived benefits and challenges, barriers to responding to prompts, and emotional reactions to questions about drug use (see Multimedia Appendix 1 for the exit interview guide).

Data Analysis

Quantitative data were analyzed using Python version 3.13.4 (Python Software Foundation). Participant characteristics and feasibility outcomes were summarized using descriptive statistics. Continuous variables were summarized using means and SDs, and categorical variables were summarized using counts and percentages. EMA adherence was calculated at the participant level, separately for morning and evening prompts. To assess changes in engagement over time, EMA prompt completion was aggregated into 4 weekly intervals and summarized across participants. EMA behavioral responses (drug use, craving, planned use, overdose events, and naloxone access) were summarized as counts of endorsed events across completed observations. Data management and statistical summaries were conducted using the *pandas* and *NumPy* libraries, and figures were generated using *Matplotlib*.

Qualitative interviews were audio-recorded and transcribed verbatim. Data were analyzed using inductive thematic analysis, following the steps described by Braun and Clarke [35]. Interview transcripts were reviewed multiple times to become familiar with the data. Two coders independently conducted open coding of all transcripts and met regularly to review and refine the coding framework. Any discrepancies were resolved through discussion with the senior coder (KG), and codes were gradually grouped into broader thematic categories. Consistent with recommendations for enhancing rigor in qualitative research, coding decisions and thematic interpretations were reviewed collaboratively throughout the analytic process [36]. Qualitative data were managed and analyzed using Dedoose version 9.0.54 (SocioCultural Research Consultants, LLC).

Ethical Considerations

The study was approved by the institutional review boards (IRB protocol #2000036555) at Yale University and the University of Connecticut. All participants provided written informed consent prior to participation. Participants received US $25 for completing the baseline visit and US $25 for completing the end-of-study visit and interview. EMA completion was incentivized on a per-prompt basis. Participants received US $1 for each completed EMA prompt (maximum of 60 prompts). The maximum total compensation per participant was US $110 over the 30-day study period.

Results

Feasibility

Recruitment and Participation

Thirteen individuals were assessed for eligibility through NHSSP, of whom 10 provided informed consent, were enrolled, and completed the baseline assessment and onboarding for the EMA app. Three eligible individuals did not proceed to enrollment because they did not attend their scheduled baseline interviews and were subsequently unreachable. One enrolled participant withdrew due to a scheduled cataract surgery before completing any EMA prompts. The remaining 9 participants completed the EMA prompts during the 30-day study period. EMA participation remained relatively stable across the study period, with 8 of the 9 participants continuing to respond to EMA prompts through week 4. Seven participants completed the exit survey and qualitative interviews following 30 days of EMA participation.

Participant Characteristics

The mean age of the 10 participants was 43.0 (SD 7.1) years (Table 1). Half identified as male, 4 (40%) as female, and 1 (10%) as transgender. Most participants had a high school education or lower (n=9, 90%), were currently unemployed, and 4 (40%) reported experiencing homelessness. Additionally,

XSL•FO
RenderX

9 (90%) participants reported injection drug use, and 8 (80%) reported daily opioid use in the past 30 days. Furthermore, 7 (70%) participants were receiving medications for opioid use disorder, including methadone maintenance treatment (n=5, 50%) and Vivitrol or other treatment programs (n=2, 20%). A history of opioid overdose was reported by 7 (70%) participants, and most had witnessed an overdose (n=9, 90%) and reported carrying naloxone (n=9, 90%). Digital health literacy was high (mean 10.3, SD 1.94; range 1 - 12).

XSL•FO
RenderX

Table . Demographic characteristics (N=10).

Variable	Values
Age (y), mean (SD)	43.0 (7.1)
Gender, n (%)	
Women	4 (40)
Men	5 (50)
Transgender	1 (10)
Sexual orientation, n (%)	
Gay, lesbian, or bisexual	5 (50)
Straight or heterosexual	5 (50)
Ethnicity, n (%)	
White	5 (50)
Racial or ethnic minority	5 (50)
Current relationship status, n (%)	
Single	4 (40)
With partner	6 (60)
Education, n (%)	
High school or lower	9 (90)
Some colleges or higher	1 (10)
Currently unemployed, n (%)	10 (100)
Currently experiencing homelessness, n (%)	4 (40)
Smartphone type, n (%)	
Apple	2 (20)
Android	8 (80)
Digital health literacy, mean (SD)	10.3 (1.9)
Substance use and overdose history, n (%)	
Currently in MOUD[a]	7 (70)
Injected any drug (past 30 d)	9 (90)
Daily opioid use (past 30 d)	8 (80)
Ever overdosed with opioid	7 (70)
Opioid overdose (past 30 d)	1 (10)
Ever witnessed an overdose	9 (90)
Currently use Naloxone	9 (90)
Physical and psychological symptoms	
Pain, mean (SD)	70.5 (24.5)
Stressed, mean (SD)	75.3 (24.2)
Moderate to severe depressive symptoms, n (%)	4 (40)
Moderate to severe anxiety symptoms, n (%)	3 (30)
Suicidality history (ever), n (%)	
Suicidal thoughts	4 (40)
Suicidal plan	1 (10)
Suicidal attempt	3 (30)
Mental health services use (past 6 mo), n (%)	
Yes	6 (60)

XSL•FO
RenderX

Variable	Values
No, but needed to	3 (30)
No, I did not need to	1 (10)

[a]MOUD: medications for opioid use disorder.

EMA Compliance

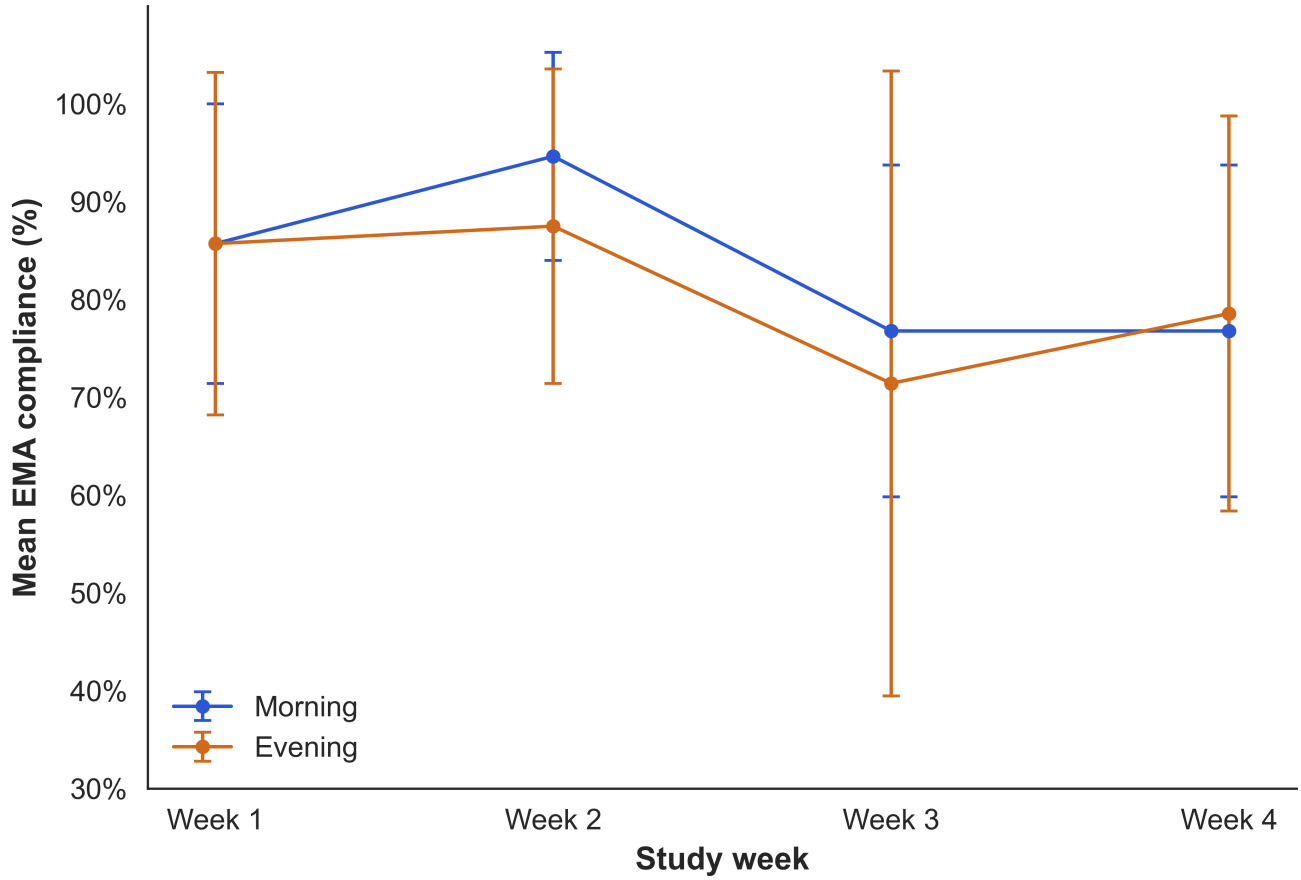
Among the 9 participants who completed at least 1 EMA, 411 out of 540 expected EMA prompts were completed during the study period. Overall, EMA compliance was high, with an average of 85.0% (SD 8.7%). Compliance was similar between morning (mean 84.2%, SD 13.1%) and evening (mean 85.8%, SD 11.2%) prompts (Table 2), with rates ranging from 17% to 97% among the participants (Figure S1 in Multimedia Appendix 1).

Table . Mean Ecological momentary assessment (EMA) compliance by prompt type during the 30-day study period (N=9).

EMA compliance measure	Values, mean (SD)
Overall compliance	85 (8.7)
Morning prompts	84.2 (13.1)
Evening prompts	85.8 (11.2)

Average EMA compliance varied across participants over the 4 weeks of participation, with similar patterns observed for morning and evening prompts (Figure 1). Morning EMA compliance was highest in week 2 (mean 94.6%, SD 10.6%), followed by week 1 (mean 85.7%, SD 14.3%). Compliance declined in week 3 (mean 76.8%, SD 17.0%) and remained stable in week 4 (mean 76.8%, SD 17.0%). Evening EMA compliance followed a comparable trajectory, peaking in week 2 (mean 87.5%, SD 16.1%) and then in week 1 (mean 85.7%, SD 17.5%). Compliance decreased in week 3 (mean 71.4%, SD 31.9%) and increased in week 4 (mean 78.6%, SD 20.2%).

Figure 1. Ecological momentary assessment (EMA) compliance across the 4 weeks of the participation period.



Drug-Related Behavioral Reporting

Weekly reporting patterns for drug-related behaviors across the 30-day study period are presented in Table 3, with participant-level reporting trajectories shown in Figure S2 in Multimedia Appendix 1. Drug craving was reported in 66.7% (274/411) of completed prompts overall, ranging from 58.4% (66/113) to 76.9% (83/108) across study weeks. Plans to use

XSL•FO
RenderX

drugs were reported in 61.3% (252/411) of completed prompts overall, ranging from 59.3% (67/113) to 62.5% (55/88) across weeks. Opioid use was reported in all completed EMA prompts across all study weeks. Overdose experiences were reported by 2 participants, with a total of 5 reports across the study period 1.2% (5/411). Naloxone possession remained consistently high throughout the study period, reported in 87.1% (358/411) of completed prompts overall and ranging from 80.4% (82/102) to 94.7% (107/113) across study weeks.

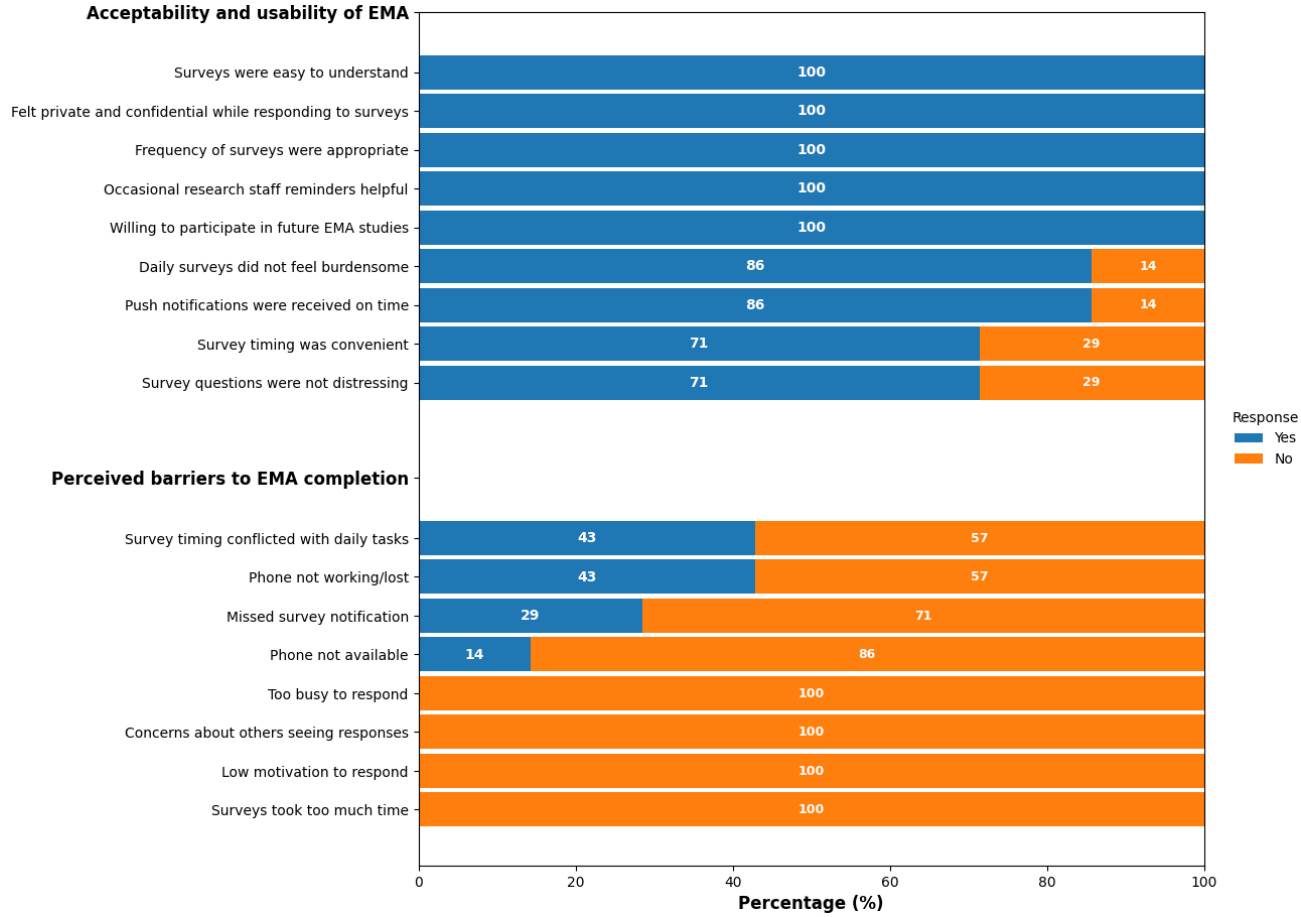
Table . Weekly reporting patterns of drug-related behaviors across ecological momentary assessment (EMA) prompts.

Behavior	Total reports, n (%)	Week 1, n (%)	Week 2, n (%)	Week 3, n (%)	Week 4, n (%)
Drug craving	274 (66.7)	83 (76.9)	66 (64.7)	59 (67.0)	66 (58.4)
Plan to use drugs	252 (61.3)	67 (62.0)	63 (61.8)	55 (62.5)	67 (59.3)
Opioid use	411 (100)	108 (100.0)	102 (100.0)	88 (100.0)	113 (100.0)
Experienced overdose	5 (1.2)	1 (0.9)	0 (0)	2 (2.3)	2 (1.8)
Carry naloxone	358 (87.1)	90 (83.3)	82 (80.4)	79 (89.8)	107 (94.7)

Acceptability

Overall, the acceptability and usability of the EMA protocol were high (Figure 2). All participants reported that the EMA prompts were easy to understand, felt private and confidential, and that the survey frequency was appropriate. All participants also indicated willingness to participate in future EMA studies and found occasional reminders from research staff helpful. Most participants (6/7, 86%) reported receiving push notifications on time and indicated that daily EMA prompts did not feel burdensome. EMA timing was considered convenient by 71% (5/7) of participants, and 71% (5/7) also reported that the survey questions were not distressing.

Figure 2. Participant-reported acceptability, usability, and perceived barriers to ecological momentary assessment (EMA; n=7).



Participants also reported a few barriers to EMA completion. EMA timing conflicts with daily tasks and phone-related issues (eg, phone not working or lost) were each reported by 43% (3/7) of participants. Missed EMA notifications were reported by 29% (2/7) of participants, and lack of phone availability at the time of the EMA was reported by 14% (1/7) of participants. None of the participants reported low motivation, concerns about privacy, or prompt length as barriers to completing the EMA.

XSL•FO
RenderX

Qualitative Findings

Seven participants completed the one-on-one semistructured exit interview. Thematic analysis yielded five themes that reflected the participants' experiences with the EMA study: (1) ease of use and integration into daily routines, (2) self-awareness and reflection related to substance use, (3) perceived accountability through self-monitoring, (4) emotional discomfort related to substance use reflection, and (5) situational and technical barriers to EMA prompt completion.

Ease of Use and Integration Into Daily Routines

Five participants described the EMA prompts as easy to complete and manageable within their daily routines. Participants mentioned that the prompts were brief and delivered at predictable times, which supported routine completion without being perceived as intrusive:

> *It was very convenient, because it was two times a day in the morning and at night, so it wasn't like bugging me or anything.* [P7]
>
> *I already knew what time they were gonna come around, so I kind of prepared to be ready to do it.* [P6]

Self-Awareness and Reflection Related to Substance Use

Four participants reported that repeated reporting increased their awareness of their substance use patterns and daily behaviors. Some participants noted that completing the EMA prompts encouraged them to reflect on their cravings, intentions to use substances, and substance use experiences throughout the day:

> *It showed me that it is true that the first hit makes me crave more… it really showed me that I needed some help.* [P6]
>
> *A lot of the time, I wanted to use drugs, but I didn't plan to, and I didn't actually do it.* [P2]

Perceived Accountability Through Self-Monitoring

Two participants described how regular EMA participation created a sense of accountability, structure, or engagement in their daily routines, even without any intervention or direct interaction with the study staff:

> *It felt good… I looked forward to doing it [EMA prompt] twice a day.* [P1]
>
> *It kept me on my toes… like an assignment or a job, you stay on schedule.* [P6]

Emotional Discomfort Related to Substance Use Reflection

Two participants described experiencing emotional discomfort when reflecting on their substance use through repeated EMA reporting. These experiences included feelings of guilt, sadness, or emotional distress associated with reporting substance use behaviors:

> *It kinda made me feel a little bit guilty when I complete it [EMA prompt] and report drugs use.* [P8]
>
> *It sometimes makes you feel sad and not so good about yourself… it's a hard reality to accept.* [P4]

Situational and Technical Barriers

Two participants described situational or technical barriers that occasionally interfered with EMA prompt completion, including oversleeping, low phone battery, and missed alerts:

> *When I overslept, and I woke up, and I missed the survey, which was challenging.* [P2]
>
> *There were a couple of times that my phone was dead and the notification didn't come through, so I forgot about it.* [P4]

Discussion

Principal Findings

In this study, we explored the feasibility and acceptability of a 30-day smartphone-based EMA protocol among people who use opioids. Recruitment through the SSP was successful and represented an important strength of this study. Overall, 70% of the participants completed the study and follow-up interviews. EMA completion rates were high, with an overall compliance rate of 85%, and similar response rates for morning and evening prompts. Participants regularly reported opioid use, cravings, overdose experiences, and carrying naloxone through EMA, suggesting that twice-daily monitoring of opioid-related behaviors is feasible in community settings.

The EMA compliance rate observed in this study was higher than the pooled compliance rate reported in a meta-analysis of 126 effect sizes from 19,431 participants among EMA studies of people who use substances (75.1%, 95% CI 72.4%-77.7%) [23]. Prior EMA studies among people who use opioids have largely focused on momentary predictors of drug use, such as craving, mood, and stress, rather than on overdose-related experiences or harm-reduction behaviors. Compliance remained generally stable across morning and evening prompts and throughout the 30-day study period, with a modest decline in the third week—a pattern reported in other EMA studies and often attributed to participant fatigue [23-25]. In our study, compliance increased again during the final week, suggesting that participants were able to integrate the EMAs into their daily routines. Predictable EMA prompt timing, brief questionnaires, and regular reminders from the research staff may have contributed to these high response rates. Prior EMA studies often included larger samples and longer study durations, which may partly explain the differences in compliance [17,18,37]. In this context, the high compliance observed in our study may have been influenced by the small sample size, short study duration, and participants' relatively high levels of education and digital literacy [23,38,39].

Importantly, participants reported their drug use behavior and experiences related to overdose and carrying naloxone through daily EMA prompts. Although only a small number of overdose events were reported during the study period, participants reported carrying naloxone on every occasion when an overdose was reported. However, the study did not assess whether naloxone was administered during these overdose events or whether it was used for reversal. These findings suggest that

XSL•FO
RenderX

EMA may be a useful approach for capturing timely information on overdose experiences and harm-reduction responses. Prior EMA-based research among people who use opioids has focused on momentary predictors of drug use, such as craving, mood, pain, and contextual factors, but did not assess overdose experiences or whether participants were carrying naloxone [17,18,40-42]. While preliminary, our findings suggest that EMA may also be useful for capturing overdose-related experiences and harm-reduction behaviors in daily life. Such information may help improve the understanding of the circumstances surrounding overdose events and may inform future real-time prevention and intervention strategies.

Despite disclosing minimal barriers (EMA prompt timing, unavailability of phone, and missed prompt notifications), participants reported that overall participation was acceptable, minimally burdensome, and did not interfere with their daily activities. In similar EMA studies [43,44], privacy and confidentiality were reported as major barriers; however, none of the participants in our study raised these concerns. Participants further reported an overall positive experience, with all indicating that they would participate in a similar study in the future. In real-world settings, it would be helpful to consider the barriers they faced, such as prompt timing, missed push notifications, and the emotional discomfort associated with some questions. Fixed notifications, individualized prompt timing, and participant-friendly question wording may help reduce participant burden and improve engagement when EMA protocols are developed using a user-centered approach. In addition, providing participants with personalized feedback based on EMA data may be more impactful [45], which could motivate them to remain engaged for a longer period. Such EMA data may also support the development of mobile health–based just-in-time adaptive interventions, where timely information on drug craving, planned use, or contextual risk factors could be used to predict periods of elevated overdose risk and deliver timely harm-reduction support, such as reminders to carry and use naloxone [46,47]. However, further research is needed to determine how such approaches can be effectively implemented in real-world settings.

Limitations

Several limitations should be considered when interpreting the findings of this study. First, the sample size was very small, and participants were recruited through a single SSP in New Haven, which limits the stability of the findings and restricts generalizability to the wider opioid-using population. Individuals engaged with harm reduction services may differ from those who are not connected to such programs, including having greater access to harm-reduction resources, higher overdose awareness, and more frequent engagement with health services, potentially limiting the generalizability of the findings to broader populations of people who use opioids. Second, the study sample included individuals with relatively high digital health literacy and smartphone access, which may not reflect the experiences of people who use opioids with lower or more unstable access to digital technologies due to social and structural vulnerabilities, such as housing instability and poverty. Third, EMA items, along with the poststudy acceptability questionnaire, were adapted and were not formally validated, which may limit comparability with other studies. Fourth, all EMA responses and poststudy assessment measures were self-reported and may have been influenced by social desirability or participant fatigue, potentially leading to overreporting of positive experiences and underreporting of barriers. Fifth, although participants reported carrying naloxone, the study did not assess whether naloxone was administered during overdose events. Sixth, the 30-day EMA protocol does not capture long-term adherence or engagement beyond the short monitoring period. Finally, participants received compensation for completing EMA prompts, which may have influenced response rates and engagement observed in this pilot study and may not reflect participation in real-world settings, where similar incentives may not be available.

Conclusions

Our findings suggest that a smartphone-based EMA protocol is feasible and acceptable among people who use opioids. Participants demonstrated high compliance with twice-daily EMA prompts and reported opioid use, overdose experiences, and naloxone possession. These findings highlight the potential of EMA to capture day-to-day patterns of substance use and overdose-related behavior. Future research should examine how EMA data can be leveraged to inform adaptive interventions, such as just-in-time adaptive interventions, to reduce the risk of overdose. Larger and longer-term studies are needed to further evaluate EMA for monitoring opioid use, supporting harm-reduction strategies, and understanding broader aspects of daily functioning and well-being among people who use opioids.

Acknowledgments

Artificial intelligence was not used at any stage in the preparation of this manuscript.

Funding

RS received funding from the National Institute on Drug Abuse (award: K01DA051346).

Data Availability

The datasets used or analyzed during the study are available from the corresponding author upon reasonable request.

XSL•FO
RenderX

Authors' Contributions

Conceptualization: KG (lead), RS (equal)
Data curation: KG (lead)
Formal analysis: KG (lead)
Funding acquisition: RS (lead)
Investigation: AT (lead), AK (supporting)
Methodology: KG (lead), RS (equal)
Project administration: RS (lead), JAW (supporting), AT (supporting), AK (supporting), KG (supporting)
Resources: RS (lead), JAW (supporting)
Supervision: RS (lead), JAW (supporting)
Validation: KG (lead), RS (supporting), JAW (supporting)
Visualization: KG (lead)
Writing – original draft: KG (lead), KP (supporting), AK (supporting)
Writing – review & editing: KG (lead), KP (supporting), JAW (supporting), AK (supporting), AT (supporting), SB (supporting), MSHS (supporting), SP (supporting), TH (supporting), RS (supporting).

Conflicts of Interest

RS is an editorial board member of *JMIR mHealth* and *uHealth* at the time of this publication. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Interview guide and supplementary figures with daily ecological momentary assessment (EMA) completion and reported drug-related behaviors.
[DOCX File, 330 KB - humanfactors_v13i1e95655_app1.docx]

References

1. U.S. overdose deaths in 2021 increased half as much as in 2020 – but are still up 15%. Centers for Disease Control and Prevention, National Center for Health Statistics. URL: https://www.cdc.gov/nchs/pressroom/releases/20220511.html?CDC_AAref_Val=https://www.cdc.gov/nchs/pressroom/nchs_press_releases/2022/202205.htm [accessed 2026-06-17]
2. Spencer MR, Miniño AM, Warner M. Drug overdose deaths in the United States, 2001-2021. NCHS Data Brief 2022 Dec;2022(457):1-8. [Medline: 36598401]
3. Garnett MF, Miniño AM. Drug overdose deaths in the United States, 2003–2023. : National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention; 2024 URL: https://www.cdc.gov/nchs/products/databriefs/db522.htm [accessed 2026-06-17]
4. Key substance use and mental health indicators in the United States: results from the 2023 National Survey on Drug Use and Health. : Substance Abuse and Mental Health Services Administration (SAMSHA); 2024 URL: https://www.drugsandalcohol.ie/41611/1/2023-nsduh-annual-national.pdf [accessed 2026-06-17]
5. Jakubowski A, Fowler S, Fox AD. Three decades of research in substance use disorder treatment for syringe services program participants: a scoping review of the literature. Addict Sci Clin Pract 2023 Jun 10;18(1):40. [doi: 10.1186/s13722-023-00394-x] [Medline: 37301953]
6. Krawczyk N, Allen ST, Schneider KE, et al. Intersecting substance use treatment and harm reduction services: exploring the characteristics and service needs of a community-based sample of people who use drugs. Harm Reduct J 2022 Aug 24;19(1):95. [doi: 10.1186/s12954-022-00676-8] [Medline: 36002850]
7. Pergolizzi Jr JV, Raffa RB, Rosenblatt MH. Opioid withdrawal symptoms, a consequence of chronic opioid use and opioid use disorder: current understanding and approaches to management. J Clin Pharm Ther 2020 Oct;45(5):892-903. [doi: 10.1111/jcpt.13114] [Medline: 31986228]
8. Bichon JA, Bailey AJ, Votaw VR, McHugh RK. Sleep disruption, stress, and craving during inpatient treatment for opioid use disorder. Addict Behav 2025 Nov;170:108427. [doi: 10.1016/j.addbeh.2025.108427] [Medline: 40614370]
9. Dunne N, Casey MB, Ivers JH. What are the risk factors of non-fatal overdose among persons who use opioids? A systematic review and meta-analysis. J Subst Use Addict Treat 2025 Apr;171:209630. [doi: 10.1016/j.josat.2025.209630] [Medline: 39890023]
10. Sujan MSH, Wickersham JA, Khati A, et al. Non-fatal overdose and associated factors among people who use opioids: findings from a cross-sectional study. J Community Health 2025 Oct;50(5):912-921. [doi: 10.1007/s10900-025-01480-8] [Medline: 40381076]
11. Furlan AD, Carnide N, Irvin E, et al. A systematic review of strategies to improve appropriate use of opioids and to reduce opioid use disorder and deaths from prescription opioids. Can J Pain 2018;2(1):218-235. [doi: 10.1080/24740527.2018.1479842] [Medline: 35005381]

XSL•FO
RenderX

12. Chatterjee A, Yu EJ, Tishberg L. Exploring opioid use disorder, its impact, and treatment among individuals experiencing homelessness as part of a family. Drug Alcohol Depend 2018 Jul 1;188:161-168. [doi: 10.1016/j.drugalcdep.2018.04.012] [Medline: 29778009]
13. Matthews TA, Sembajwe G, von Känel R, Li J. Associations of employment status with opioid misuse: evidence from a nationally representative survey in the U.S. J Psychiatr Res 2022 Jul;151:30-33. [doi: 10.1016/j.jpsychires.2022.04.001] [Medline: 35436703]
14. Krawczyk N, Williams AR, Saloner B, Cerdá M. Who stays in medication treatment for opioid use disorder? A national study of outpatient specialty treatment settings. J Subst Abuse Treat 2021 Jul;126:108329. [doi: 10.1016/j.jsat.2021.108329] [Medline: 34116820]
15. Milaney K, Passi J, Zaretsky L, et al. Drug use, homelessness and health: responding to the opioid overdose crisis with housing and harm reduction services. Harm Reduct J 2021 Aug 26;18(1):92. [doi: 10.1186/s12954-021-00539-8] [Medline: 34446034]
16. Burgess-Hull A, Epstein DH. Ambulatory assessment methods to examine momentary state-based predictors of opioid use behaviors. Curr Addict Rep 2021;8(1):122-135. [doi: 10.1007/s40429-020-00351-7] [Medline: 33425652]
17. Lorvick J, Hemberg JL, Browne EN, George MJ, Piontak J, Comfort ML. Ecological momentary assessment study of same-hour polysubstance use among people who use opioids and additional substances. Drug Alcohol Depend 2025 Apr 1;269:112582. [doi: 10.1016/j.drugalcdep.2025.112582] [Medline: 39952168]
18. Alexander K, Sanjuan P, Terplan M. The use of ecological momentary assessment methods with people receiving medication for opioid use disorder: a systematic review. Curr Addict Rep 2023;10(3):366-377. [doi: 10.1007/s40429-023-00492-5]
19. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. Psychol Assess 2009 Dec;21(4):486-497. [doi: 10.1037/a0017074] [Medline: 19947783]
20. Button D, Levander XA, Cook RR, et al. Substance use disorder treatment and technology access among people who use drugs in rural areas of the United States: a cross-sectional survey. J Rural Health 2023 Sep;39(4):772-779. [doi: 10.1111/jrh.12737] [Medline: 36575145]
21. Garett R, Young SD. Potential effects of digital inequality on treatment seeking for opioid use disorder. Int J Ment Health Addict 2023;21(2):890-895. [doi: 10.1007/s11469-021-00629-5] [Medline: 34466133]
22. Venkatasubramanian K, Appleton J, Ranalli TM, Mankodiya K, Solanki D, Carreiro S. Leveraging trauma informed care for digital health intervention development in opioid use disorder. J Med Toxicol 2025 Jan;21(1):60-68. [doi: 10.1007/s13181-024-01040-x] [Medline: 39446308]
23. Jones A, Remmerswaal D, Verveer I, et al. Compliance with ecological momentary assessment protocols in substance users: a meta-analysis. Addiction 2019 Apr;114(4):609-619. [doi: 10.1111/add.14503] [Medline: 30461120]
24. Businelle MS, Hébert ET, Shi D, et al. Investigating best practices for ecological momentary assessment: nationwide factorial experiment. J Med Internet Res 2024 Aug 12;26(3):e50275. [doi: 10.2196/50275] [Medline: 39133915]
25. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003 Nov;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
26. Delgadillo J, Payne S, Gilbody S, et al. How reliable is depression screening in alcohol and drug users? A validation of brief and ultra-brief questionnaires. J Affect Disord 2011 Nov;134(1-3):266-271. [doi: 10.1016/j.jad.2011.06.017] [Medline: 21723619]
27. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
28. Punia K, Levitt E, Taisir R, et al. Psychometric validation of the Generalized Anxiety Disorder Scale (GAD-7) and Patient Health Questionnaire (PHQ-9) in an inpatient substance use disorder treatment program. Psychol Addict Behav 2025 Dec;39(8):723-732. [doi: 10.1037/adb0001098] [Medline: 41196704]
29. Kessler RC, Ustün TB. The World Mental Health (WMH) survey initiative version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). Int J Methods Psychiatr Res 2004;13(2):93-121. [doi: 10.1002/mpr.168] [Medline: 15297906]
30. Nelson LA, Pennings JS, Sommer EC, Popescu F, Barkin SL. A 3-item measure of digital health care literacy: development and validation study. JMIR Form Res 2022 Apr 29;6(4):e36043. [doi: 10.2196/36043] [Medline: 35486413]
31. Stone AA, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 2002;24(3):236-243. [doi: 10.1207/S15324796ABM2403_09] [Medline: 12173681]
32. Turner CM, Coffin P, Santos D, et al. Race/ethnicity, education, and age are associated with engagement in ecological momentary assessment text messaging among substance-using MSM in San Francisco. J Subst Abuse Treat 2017 Apr;75:43-48. [doi: 10.1016/j.jsat.2017.01.007] [Medline: 28237053]
33. Wray TB, Kahler CW, Monti PM. Using ecological momentary assessment (EMA) to study sex events among very high-risk men who have sex with men (MSM). AIDS Behav 2016 Oct;20(10):2231-2242. [doi: 10.1007/s10461-015-1272-y] [Medline: 26746212]
34. Biello K, Salhaney P, Valente PK, et al. Ecological momentary assessment of daily drug use and harm reduction service utilization among people who inject drugs in non-urban areas: a concurrent mixed-method feasibility study. Drug Alcohol Depend 2020 Sep 1;214:108167. [doi: 10.1016/j.drugalcdep.2020.108167] [Medline: 32679521]

XSL•FO
RenderX

35. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
36. Leung L. Validity, reliability, and generalizability in qualitative research. J Family Med Prim Care 2015;4(3):324-327. [doi: 10.4103/2249-4863.161306] [Medline: 26288766]
37. Kowalczyk WJ, Moran LM, Bertz JW, et al. Using ecological momentary assessment to examine the relationship between craving and affect with opioid use in a clinical trial of clonidine as an adjunct medication to buprenorphine treatment. Am J Drug Alcohol Abuse 2018;44(5):502-511. [doi: 10.1080/00952990.2018.1454933] [Medline: 29634425]
38. Yao L, Yang Y, Wang Z, Pan X, Xu L. Compliance with ecological momentary assessment programmes in the elderly: a systematic review and meta-analysis. BMJ Open 2023 Jul 12;13(7):e069523. [doi: 10.1136/bmjopen-2022-069523] [Medline: 37438069]
39. Joo J, Devkota J, Stone BM, et al. Predictors of participant compliance with ecological momentary assessment among individuals with chronic pain who are using cannabis and opioids. Internet Interv 2024 Dec;38:100784. [doi: 10.1016/j.invent.2024.100784] [Medline: 39611051]
40. Heinz MV, Price GD, Singh A, et al. A longitudinal observational study with ecological momentary assessment and deep learning to predict non-prescribed opioid use, treatment retention, and medication nonadherence among persons receiving medication treatment for opioid use disorder. J Subst Use Addict Treat 2025 Jun;173:209685. [doi: 10.1016/j.josat.2025.209685] [Medline: 40127869]
41. Anderson Goodell EM, Nordeck C, Finan PH, Vandrey R, Dunn KE, Thrul J. Feasibility and acceptability of using smartphone-based EMA to assess patterns of prescription opioid and medical cannabis use among individuals with chronic pain. Internet Interv 2021 Dec;26:100460. [doi: 10.1016/j.invent.2021.100460] [Medline: 34646751]
42. Tyler KA, Eisenbraun AR, Habecker P, Khan B. A feasibility and acceptability study using ecological momentary assessment to evaluate drug use patterns among midwestern people who use drugs. J Drug Issues 2025 Jul;55(3):420-433. [doi: 10.1177/00220426241231723] [Medline: 41877898]
43. Roth AM, Rossi J, Goldshear JL, et al. Potential risks of ecological momentary assessment among persons who inject drugs. Subst Use Misuse 2017 Jun 7;52(7):840-847. [doi: 10.1080/10826084.2016.1264969] [Medline: 28426353]
44. Rudolph AE, Bazzi AR, Fish S. Ethical considerations and potential threats to validity for three methods commonly used to collect geographic information in studies among people who use drugs. Addict Behav 2016 Oct;61:84-90. [doi: 10.1016/j.addbeh.2016.05.014] [Medline: 27249807]
45. Bartels SL, van Zelst C, Melo Moura B, et al. Feedback based on experience sampling data: examples of current approaches and considerations for future research. Heliyon 2023 Sep;9(9):e20084. [doi: 10.1016/j.heliyon.2023.e20084] [Medline: 37809510]
46. Perski O, Hébert ET, Naughton F, Hekler EB, Brown J, Businelle MS. Technology-mediated just-in-time adaptive interventions (JITAIs) to reduce harmful substance use: a systematic review. Addiction 2022 May;117(5):1220-1241. [doi: 10.1111/add.15687] [Medline: 34514668]
47. Koike C, Nestor BA, Gelmez PS, Fladger A, Shrier LA, Kossowsky J. Just-in-time adaptive interventions for adolescent and young adult substance use: a systematic review. J Adolesc Health 2025 Dec;77(6):1038-1049. [doi: 10.1016/j.jadohealth.2025.07.027] [Medline: 41065631]

Abbreviations

EMA: ecological momentary assessment
NHSSP: New Haven Syringe Services Program
SSP: Syringe Service Program

Edited by J Sarvestan; submitted 18.Mar.2026; peer-reviewed by J Hodges, M Waselewski, M Mackesy-Amiti, S Proctor; revised version received 04.Jun.2026; accepted 10.Jun.2026; published 29.Jun.2026.

Please cite as:
Gautam K, Paudel K, Wickersham JA, Khati A, Thapa A, Bhusal S, Sujan MSH, Pagoto S, Ha T, Shrestha R
Smartphone-Based Ecological Momentary Assessment to Monitor Opioid Use and Overdose Among People Who Use Opioids: Prospective Observational Feasibility Study
JMIR Hum Factors 2026;13:e95655
URL: https://humanfactors.jmir.org/2026/1/e95655
doi:10.2196/95655

© Kamal Gautam, Kiran Paudel, Jeffrey A Wickersham, Antoine Khati, Anushka Thapa, Sandesh Bhusal, Md Safaet Hossain Sujan, Sherry Pagoto, Toan Ha, Roman Shrestha. Originally published in JMIR Human Factors (https://humanfactors.jmir.org),

XSL•FO
RenderX

29.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Expert Views on Criteria for Evaluation of Human Factors Methods: Qualitative Interview Study

Selvana Awad[1,2], BPharm, MHSM; Rachel Begg[2], BPsych, MPsych; Thomas Loveday[2], BPsych, MPsych, PhD; Andrew Baillie[1,3], BSc, MPsych, PhD; Melissa Baysari[1], BPsych, PhD

[1]The University of Sydney, Camperdown, Sydney, Australia

[2]NSW Health, 1 Reserve Rd, St Leonards, Australia

[3]Sydney Local Health District, Camperdown, Australia

Corresponding Author:
Selvana Awad, BPharm, MHSM
The University of Sydney, Camperdown, Sydney, Australia

Abstract

Background: Human factors (HF), or ergonomics, which explores the interaction between humans and systems, has been used to support design in safety-critical industries such as aviation, transportation, nuclear power, and manufacturing. HF methods have the potential to support the safe design of health IT; however, the evaluation of HF methods to determine their effectiveness and feasibility in this context has been limited.

Objective: The aim of this study was to identify criteria for evaluating HF methods when applied to real-world projects and to use these to propose a framework for method evaluation.

Methods: The study design was qualitative and descriptive and involved semistructured interviews with HF experts working across health and nonhealth industries in academic and/or practitioner roles. HF experts held a relevant degree (eg, ergonomics and HF engineering) and were actively using their HF expertise. Results were thematically analyzed.

Results: A total of 21 participants took part, and interviews lasted, on average, 52 (range 39 - 103) minutes. Participants mentioned that they did not routinely evaluate methods; however, when asked how they would evaluate methods, they outlined a range of criteria to support method evaluation. Overall, 5 criteria and 28 subcriteria were identified. High-level criteria included effectiveness, efficiency, ease of use and acceptability, and impact on the solution.

Conclusions: Results from this study were used to propose a framework for evaluating HF methods used in real-world health IT projects. The framework should provide organizations with valuable information on how to optimize the application and outcomes of HF methods and build HF capability within organizations, particularly where this capability may be lacking.

(JMIR Hum Factors 2026;13:e73324) doi:10.2196/73324

KEYWORDS

human factors; evaluation; methods; quantitative; qualitative; framework

Introduction

Health IT (HIT) enables the processing, storage, and exchange of health information in an electronic environment [1]. In the United States, the Meaningful Use program has led to a significant increase in the adoption of HIT, such as electronic health records [1]. Although clear benefits of HIT have been demonstrated [2,3], research has shown that many HITs suffer from poor usability [4,5]. Usability is a measure of how well a specific user in a specific context can use a product to achieve a defined goal safely, effectively, efficiently, and satisfactorily [5]. Usability can be compromised by poor design and lead to "use errors" [5-7]. The relationship between system design and safety in the context of HIT is well recognized in the literature [5,8-12]. For example, a recent systematic review found that poorly designed electronic health records were associated with usability issues such as poor data entry, lack of workflow support, and inadequate automation [13]. These issues directly contributed to medication errors, such as patient overdoses, and had other negative impacts on medication safety [13]. Human factors (HF) methods have the potential to reduce "use errors" and, in turn, improve patient outcomes [7].

The discipline of HF, or ergonomics, which explores the interaction between humans and systems, has been used to support design in safety-critical industries such as aviation, transportation, nuclear power, and manufacturing [7,14-16]. While some HF methods, particularly those focused on human-computer interaction (eg, usability testing), have been used to support the design and redesign of HIT, the use of safety-focused and systems-based HF methods in this context is limited [8,17]. HF methods commonly used in the HIT context tend to have a linear and micro perspective of potential issues

XSL•FO
RenderX

that may affect system use and user behavior, as they focus on specific problems that may be encountered by the individual user using the system rather than considering the entire sociotechnical system [8,17,18]. HF methods that apply a systems thinking lens to identify problems and risks that may arise from the interactions between components of a complex system are less commonly used in HIT design and evaluation [8,17,19,20].

Potential reasons for the limited use of HF methods include issues with the availability of HF expertise in health care, a research-practice gap, a potential echo chamber effect within health care whereby the same subset of well-known methods is used repeatedly, issues with method usability, and challenges with demonstrating the value added by HF application to justify up-front investment [10,17,20,21]. Further research is required to demonstrate the value of applying HF methods (particularly those with a systems safety focus) to HIT, supported by a robust evaluation approach or framework [22,23].

Evaluation of HF methods to determine their effectiveness and feasibility has been limited. Previous evaluations have mainly been in the context of academic studies evaluating the reliability, validity, and efficacy of some HF methods, for example, when applied by health care participants as compared with experienced HF experts [20]. Despite these studies, empirical data regarding the reliability or validity of many HF methods do not exist [24]. Several challenges associated with conducting these forms of evaluations have been cited, including challenges with recruiting enough experts to enable comparison with a gold standard, time and resource intensiveness, and limited knowledge of appropriate statistical analyses [22,24].

An additional challenge is that there is no consensus or agreement on what constitutes an effective or valuable HF method. How do HF practitioners select and evaluate methods when applying them to real-world projects as part of HF integration processes?

Although a previous paper provides suggestions on potential evaluation criteria for HF methods [18], these are not comprehensive and represent the authors' recommendations rather than findings derived from research methods. The aim of this study was to identify criteria for evaluating HF methods when applied to real-world projects and to use these to propose a framework for method evaluation.

Methods

Overview

This qualitative descriptive study was undertaken as part of a larger study that focused on HF and safety analysis methods for use in the design, redesign, and configuration of HIT. The study design involved semistructured interviews with HF experts working across health and nonhealth industries in academic and/or practitioner roles. A participant was considered an HF expert if they had a relevant degree (eg, ergonomics and HF engineering) and were actively using their HF expertise. Part 1, currently under review, explored what HF methods experts use and how they are selected. Part 2, reported here, explored criteria HF experts use or view as important to evaluate HF methods. Questions used to guide the semistructured interviews were developed by a clinical informatics professional with expertise in design, HF, and safety and quality (SA) and a HF expert (MB) and reviewed by other members of the research team with HF (TL and RB) and implementation science (AB) expertise. After collecting demographic information, the 2 main questions asked relevant to this study were "What makes a 'good' HF method to you?" and "How would you evaluate a HF method?"

Participants were recruited through a combination of purposive, opportunistic, and snowball sampling. Recruitment involved advertising the study through national HF societies and relevant working groups; promotion at an international HF and patient safety conference through networking and word-of-mouth approaches; consultation with HF experts and contacts to recommend potential participants; and a review of common HF textbooks and literature to identify HF authors who could be invited to participate. In addition, HF practitioners known to the research team were directly approached. Suitable participants interested in the study, including those identified by the investigators, were invited to take part in the study via email. Recruitment continued until thematic saturation was reached, that is, no new themes were emerging from the data [25].

The interviews, conducted by SA, occurred via videoconferencing, except for one, which was face to face. Deidentified content from the interview transcripts was independently analyzed by 2 investigators with expertise in HF and qualitative analysis (SA and RB) using a general inductive approach [26]. Each investigator independently coded data for the first 5 interviews and then met to discuss findings and reach consensus on a high-level framework to support the documentation of codes into themes and subthemes. For the remainder of the interviews, each investigator continued to independently assign text to the agreed themes and subthemes using the framework and, at the end of this process, met to discuss any further discrepancies until consensus was reached [27]. Overall, the 2 investigators were generally consistent in their coding and identification of themes and subthemes. Disagreements were minor and resolved via discussion until consensus was reached.

Ethical Considerations

Ethics approval was obtained from the University of Sydney's Human Research Ethics Committee. All participants provided informed consent and agreed to be audio-recorded. All transcripts were de-identified prior to data analysis. All participants provided informed consent and agreed to be audio-recorded.

Results

Participant Demographics

A total of 21 participants took part, and interviews lasted, on average, 52 (range 39 - 103) minutes. Table 1 describes the demographics of the 21 participants.

XSL•FO
RenderX

Table . Characteristics of the participants included in the interviews (N=21).

Core industry and core role		Participants, n	Within industry category, %	Within total participants, %
Health (n=14)				
	Academic	9	64	43
	Practitioner	2	14	10
	Academic and practitioner	3	21	14
	Total	14	100	67
Nonhealth (n=7)				
	Academic	2	29	10
	Practitioner	5	71	24
	Academic and practitioner	0	0	0
	Total	7	100	33

How Should HF Methods Be Evaluated?

Participants explained that they did not routinely evaluate methods; however, when asked how they would evaluate them, they outlined a range of criteria to support method evaluation (Table 2). High-level criteria included effectiveness, efficiency, ease of use and acceptability, and impact on the solution. Other criteria, such as validity, reproducibility, and reliability, were mainly mentioned by academic participants rather than practitioners. Overall, 5 criteria and 28 subcriteria were identified.

XSL•FO
RenderX

Table . Criteria for human factors (HF) method evaluation, as reported by participants.

Criteria and subcriteria	Example quotes
Effectiveness	
1.1 Effectiveness in identifying usability issues	• [N]ice to see whether or not the amount of usability issues related to patient safety. [P2]
1.2 Effectiveness in identifying safety issues	• [N]ice to see whether or not the amount of usability issues related to patient safety. [P2] • ...there's a primary metric that matters most like which method uncovered the most safety issues in advance of implementation or which method uncovered the more severe safety issues [P9]
1.3 Ability to achieve intended impact, change or goal	• What value did it add? Like? Do you feel like added value to the project? Do you feel like it gave you the right answers or the right tools to get to the answers that you wanted? [P16] • Does it make the change that you envision?...And that has to do with your final goal [P6]
1.4 Generates recommendations that are useful and easy to implement	• Does it generate recommendations [that] are easy to implement? And that then do [they] get implemented and lead to risk management? [P7] • [T]he other thing that's really important to us in terms of criteria is going beyond analysis and figuring out a method that can actually produce useful information, and actionable information for redesign for change, for implementation, whatever you want [P11]
1.5 Ability to identify micro, meso and macro (systems) level considerations that have safety implications	• [Y]ou might do an overarching review of the technology in the context of the people using it in the context of the work in the workplace and then on the basis of that, deploy specific things that are the micro, meso, or macro level, to understand a bit more [P7]
1.6 Method's effectiveness in understanding the dynamic nature of the system	• And another important thing is to continue to collect data about the use of the system, what I call the dynamic safety of an application. [P13]
1.7 Ability to cover the required domain areas/constructs (eg, usability, safety, workload, situational awareness, and decision-making)	• And then I think it's the core human factors constructs. It's kind of like, what's the impact on situational awareness, workload? Usability, and decision making, are probably some key constructs that you might want to check [P21]
1.8 Overall usefulness	• [T]he other thing that's really important to us in terms of criteria is going beyond analysis and figuring out a method that can actually produce useful information, and actionable information for redesign for change, for implementation, whatever you want [P11]
Efficiency	
2.1 Demand on time and resources (e.g.eg, workload)	• How fast it was... how much time [and] how much budget it came under. [P16]
2.2 Cost required	• Is it cost effective? [P6]
2.3 Overall efficiency	• It's to look at the efficiency of the methods [P4]
Ease of use and acceptability	
3.1 Ease of use	• Is it usable, based on some objective standard usability? Or is it more usable than the alternative? [P7] • [Y]our method should be easy to use... not too demanding in human resources in training, and not too costly [P1] • [E]ase of use of the method and the ease of understanding the tools from the people who participate [P13]
3.2 Learnability	• How quick it was to learn [P16]

XSL•FO
RenderX

Criteria and subcriteria	Example quotes
3.3 Utility and acceptability of the method	• The types of questions that make sense are acceptability. You know, utility...in [a] very basic sense, we applied this in this way, what did we learn, we learned something new [P7]
3.4 Overall satisfaction of those applying the method	• [E]ase of use of the method and the ease of understanding the tools from the people who participate [P13]
3.5 Participant experience	• [E]ase of use of the method and the ease of understanding the tools from the people who participate [P13] • [W]e asked participants to give us feedback on the method...I think that's an important source of feedback and evaluation of the method, what do they think about it? [P11]
3.6 Likely adoption of the method based on complexity and learnability	• If you develop a very complex method, it may be the best one to identify all the problems that could lead to a risk for the patient. But...nobody's applying it [P1]
3.7 Adoption of the method by non-HF experts	• [T]alked about giving away ergonomics. And in a lot of the things that we've done that has been our goal when we leave, can they actually do things on their own? Even if it's a little bit more simple than the way we would do it? [P11]
3.8 Whether the method met expectations	• And another important thing probably is to clarify, the time needed to apply one method and the expected outcome, like a sort of table of the expectation that one can have through the application of that method, because too often, people look at the ergonomist, either those with the silver bullet in the hand, or those with some annoying requirements to be applied in the design process. [P13]
3.9 Ability of the method to be adapted	• [D]evelop questions that can be added into evaluation about the potential adaptation of these methods [P7]
Impact	
4.1 Impact on patient safety and other outcomes	• [If] you've got a goal of patient safety...the main criterion should be the risk to the patient... the likelihood of patient safety issues...if you want that your method is used, if you want to make it to be used by vendors and hospital later...analyze their human factors or risk, relate the human factors related risks of EHR [electronic health record] [P1] • Is the whole system in total more safe? [P6]
4.2 Impact on end user workflows and workloads	• [Y]ou can also identify workload issues that will slow down the care process even if it doesn't threaten directly patient [P1]
4.3 End user satisfaction (of the health information technology)	• The satisfaction of the end users also could be analyzed [P1]
4.4 Adoption of the system or tool by end users	• Is it used? Is it actually used? By all your users? [P6]
4.5 Impact on redesign (ie, whether changes or improvements were made)	• So when we redesigned the document, how many of those many, many vulnerabilities or problems were identified in the document? How many of those were we able to address? So, you know, having concrete impact on design, I think is a big issue. [P11]
Other	
5.1 Preestablished validity	• I think these reliability and validity studies are really important [P7]
5.2 Reproducibility of results	• [D]id you align to see if you meet that standard? ...and then do all these things that are for measurements, I think it must be reproducible [P9]
5.3 Reliability if applied by different people (interrater reliability)	• I think these reliability and validity studies are really important [P7]

XSL•FO
RenderX

Participants commented on challenges with quantitatively evaluating methods using the criteria identified and suggested a more qualitative, self-assessment–based evaluation; for example:

> *We usually sit down when we've concluded a project as part of the HF team and we will talk about the methods we used and how we feel they performed, whether we had any pitfalls, whether we wished we had something different, or we've done something different.* [P16]

Discussion

This study identified 5 high-level criteria and 28 subcriteria to support the evaluation of HF methods in real-world HIT projects, as reported by HF practitioners across health and nonhealth industries.

Although previous studies have evaluated HF methods by focusing on validity and reliability [23], this study is the first of its kind to identify a broad range of evaluation criteria for HF methods, as recommended by HF practitioners. There is some overlap between our criteria and those developed by Waterson et al [18], which focused on the evaluation of systems-based methods. As such, this study adds to findings generated by Waterson et al [18] by confirming some of the criteria identified in this previous work. This includes outcomes of the method, the method's robustness, the method's usability and support requirements, aspects related to work domains, and aspects related to different levels (ie, individual, team, and organizational) [18].

Our study adds to this by identifying new criteria, particularly those related to operationalized usability, adoption, and other impacts on users (eg, workload); outcome-focused effectiveness, including the link between usability and safety; and more nuanced consideration of system-level considerations (micro, meso, and macro). Furthermore, the findings from our study offer a combined quantitative and qualitative approach to evaluation, with a focus on actionable insights and bridging theory and practice. These points highlight the value of exploring the views of a range of HF experts working across different industries [18].

Our study also elaborates on the method's effectiveness in understanding the dynamic nature of the system, which is aligned with systems safety thinking that defines safety as a dynamic, emergent property of how system components interact with each other [19,28-30]. A dynamic system is a complex and adaptive system that changes behavior due to interactions between system components [31-33]. As such, the safety of a system can change over time and is impacted by many variables, such as human performance, resources, and events at particular points in time. This is consistent with a study that aimed to evaluate methods, such as the functional resonance analysis method, using resilience characteristics as indicators of core safety factors [34].

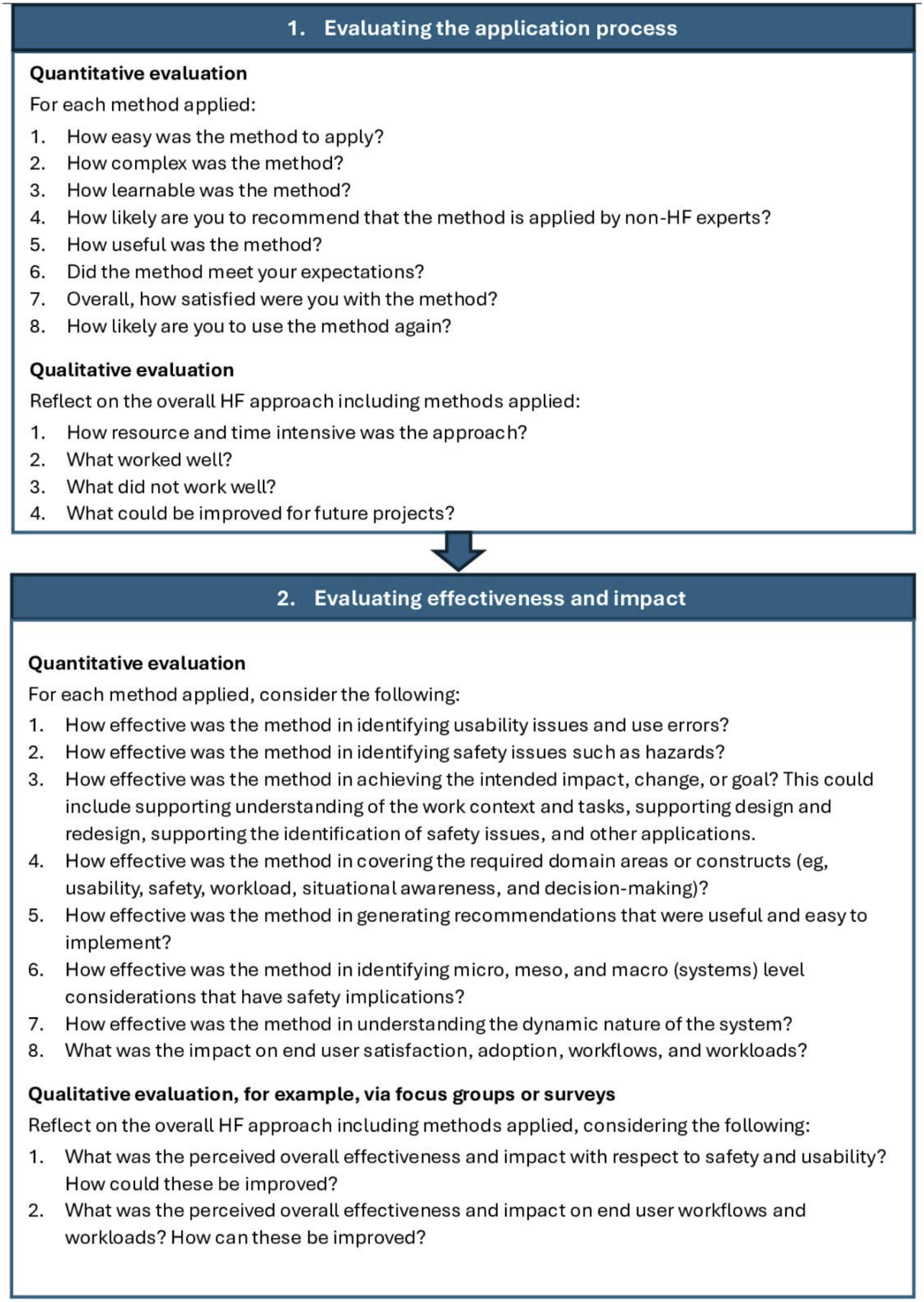
Participants commented on challenges with quantitatively evaluating methods and suggested that a more qualitative approach could be used. This is generally aligned with other literature that suggests that outcome-based quantitative evaluation is not widespread and can be resource intensive [18]. Furthermore, demonstrating the impact of methods can be difficult, as system usability and safety are influenced by a range of factors outside of the method itself [18]. While many factors influence outcomes, health organizations may still benefit from using measures, for example, the number of usability and safety issues identified and user satisfaction scores, to evaluate whether applying methods results in detection of usability and safety issues. Where possible, comparative evaluation to demonstrate that applying methods is better than not applying any method may also be of value. Although such approaches may lack the academic rigor of reliability and validity studies, they may increase an organization's degree of confidence in HF methods and, therefore, willingness to invest in them. Such evaluation may help health staff and managers develop business cases for the application of HF, help organizations reflect on strategies to enhance the use of methods (eg, organizational support), and refine when and how HF methods are applied to deliver the most value within the context of challenging and complex HIT projects. This type of evaluation could occur through self-reported Likert-scale ratings (as done by Waterson et al [18]); reflective qualitative discussions, as suggested by participants in this study; and other relevant metrics.

This study had several limitations. The recruitment process relied on interested experts volunteering to participate, so the sample may have been biased and our results may not represent the views of all HF experts. We included both health and nonhealth participants, as well as practitioners and academics, but did not compare the views of participants from different groups. Although our sample size was modest, we continued interviews until thematic saturation was reached, which is the norm in qualitative research.

We used participant responses and systems safety knowledge to propose a framework for evaluating HF methods in the context of real-world HIT projects (Figure 1). By covering both process- and outcome-based evaluation, this framework loosely aligns with other evaluation frameworks, such as those used in quality improvement and program evaluation [35,36]. While we also recommend further reliability and validity academic studies to validate the robustness of HF methods in the HIT context, our proposed framework offers a subjective yet structured approach for HF method validation that can be applied by organizations to real-world HIT projects where HF methods have been used. The framework, which we recommend be applied flexibly depending on the nature of the project and the resources available, should provide organizations with valuable information on how to optimize the application and outcomes of HF methods and build HF capability within organizations, particularly where this capability may be lacking.

XSL•FO
RenderX

Figure 1. Proposed framework for evaluating human factors (HF) methods.



XSL•FO
RenderX

Funding

SA was supported by an Australian Government Research Training Program (RTP) scholarship.

Conflicts of Interest

None declared.

References

1. Jones SS, Rudin RS, Perry T, Shekelle PG. Health information technology: an updated systematic review with a focus on meaningful use. Ann Intern Med 2014 Jan 7;160(1):48-54. [doi: 10.7326/M13-1531] [Medline: 24573664]
2. Agius S, Cassar V, Bezzina F, Topham L. Leveraging digital technologies to enhance patient safety. Health Technol 2025 Nov;15(6):1053-1063. [doi: 10.1007/s12553-025-01001-6]
3. Insani WN, Zakiyah N, Puspitasari IM, et al. Digital health technology interventions for improving medication safety: systematic review of economic evaluations. J Med Internet Res 2025 Feb 5;27:e65546. [doi: 10.2196/65546] [Medline: 39909404]
4. Adams KT, Pruitt Z, Kazi S, et al. Identifying health information technology usability issues contributing to medication errors across medication process stages. J Patient Saf 2021 Dec 1;17(8):e988-e994. [doi: 10.1097/PTS.0000000000000868] [Medline: 34009868]
5. ISO 9241-210 ergonomics of human-system interaction - part 210: human-centred design for interactive systems. International Organization for Standardization. URL: https://www.iso.org/standard/77520.html [accessed 2025-05-25]
6. ISO 9241-143:2012(e) ergonomics of human-system interaction part 143: forms. International Organization for Standardization. 2012. URL: https://www.iso.org/standard/53590.html [accessed 2025-05-25]
7. Technical basis for user interface design of health IT. National Institute of Standards and Technology (NIST) 15-996. 2015. URL: https://www.nist.gov/publications/technical-basis-user-interface-design-health-it [accessed 2025-05-25]
8. Awad S, Amon K, Baillie A, Loveday T, Baysari MT. Human factors and safety analysis methods used in the design and redesign of electronic medication management systems: a systematic review. Int J Med Inform 2023 Apr;172:105017. [doi: 10.1016/j.ijmedinf.2023.105017] [Medline: 36809716]
9. Brown CL, Mulcaster HL, Triffitt KL, et al. A systematic review of the types and causes of prescribing errors generated from using computerized provider order entry systems in primary and secondary care. J Am Med Inform Assoc 2017 Mar 1;24(2):432-440. [doi: 10.1093/jamia/ocw119] [Medline: 27582471]
10. Carayon P, Du S, Brown R, Cartmill R, Johnson M, Wetterneck TB. EHR-related medication errors in two ICUs. J Healthc Risk Manag 2017 Jan;36(3):6-15. [doi: 10.1002/jhrm.21259] [Medline: 28099789]
11. Ratwani RM, Savage E, Will A, et al. Identifying electronic health record usability and safety challenges in pediatric settings. Health Aff (Millwood) 2018 Nov;37(11):1752-1759. [doi: 10.1377/hlthaff.2018.0699] [Medline: 30395517]
12. Kushniruk A, Nohr C, Borycki E. Human factors for more usable and safer health information technology: where are we now and where do we go from here? Yearb Med Inform 2016 Nov 10;2016(1):120-125. [doi: 10.15265/IY-2016-024] [Medline: 27830239]
13. Cahill M, Cleary BJ, Cullinan S. The influence of electronic health record design on usability and medication safety: systematic review. BMC Health Serv Res 2025 Jan 6;25(1):31. [doi: 10.1186/s12913-024-12060-2] [Medline: 39762805]
14. Guidelines. National Health and Medical Research Council (NHMRC). URL: https://www.nhmrc.gov.au/guidelines [accessed 2025-06-15]
15. Dul J, Bruder R, Buckle P, et al. A strategy for human factors/ergonomics: developing the discipline and profession. Ergonomics 2012;55(4):377-395. [doi: 10.1080/00140139.2012.661087] [Medline: 22332611]
16. Ahlstrom V. Human factors design standard (DOT/FAA/HF-STD-001B). US Department of Transportation Federal Aviation Administration. 2016. URL: https://hf.tc.faa.gov/publications/2016-12-human-factors-design-standard/full_text.pdf [accessed 2025-05-25]
17. King BJ, Read GJM, Hulme A, et al. Evaluating the use of systems thinking methods in healthcare: a RE-AIM analysis of AcciMap and Net-HARMS. Ergonomics 2025 Sep;68(9):1542-1560. [doi: 10.1080/00140139.2024.2423170] [Medline: 39552189]
18. Waterson P, Robertson MM, Cooke NJ, Militello L, Roth E, Stanton NA. Defining the methodological challenges and opportunities for an effective science of sociotechnical systems and safety. Ergonomics 2015;58(4):565-599. [doi: 10.1080/00140139.2015.1015622] [Medline: 25832121]
19. Leveson N. A new accident model for engineering safer systems. Saf Sci 2004 Apr;42(4):237-270. [doi: 10.1016/S0925-7535(03)00047-X]
20. Salmon PM, King B, Hulme A, et al. Toward the translation of systems thinking methods in patient safety practice: assessing the validity of Net-HARMS and AcciMap. Saf Sci 2023 Mar;159:106003. [doi: 10.1016/j.ssci.2022.106003]

XSL•FO
RenderX

21. Carayon P, Salwei ME. Moving toward a sociotechnical systems approach to continuous health information technology design: the path forward for improving electronic health record usability and reducing clinician burnout. J Am Med Inform Assoc 2021 Apr 23;28(5):1026-1028. [doi: 10.1093/jamia/ocab002] [Medline: 33537756]
22. Hulme A, Stanton NA, Walker GH, Waterson P, Salmon PM. Are accident analysis methods fit for purpose? Testing the criterion-referenced concurrent validity of AcciMap, STAMP-CAST and AcciNet. Saf Sci 2021 Dec;144:105454. [doi: 10.1016/j.ssci.2021.105454]
23. Hulme A, Stanton NA, Walker GH, Waterson P, Salmon PM. Testing the reliability and validity of risk assessment methods in human factors and ergonomics. Ergonomics 2022 Mar;65(3):407-428. [doi: 10.1080/00140139.2021.1962969] [Medline: 34328389]
24. Stanton NA. Human Factors Methods: CRC Press.
25. Guest G, Namey E, Chen M. A simple method to assess and report thematic saturation in qualitative research. PLoS One 2020;15(5):e0232076. [doi: 10.1371/journal.pone.0232076] [Medline: 32369511]
26. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
27. Doyle L, McCabe C, Keogh B, Brady A, McCann M. An overview of the qualitative descriptive design within nursing research. J Res Nurs 2020 Aug;25(5):443-455. [doi: 10.1177/1744987119880234] [Medline: 34394658]
28. Cook RI. How complex systems fail. : University of Chicago; 2002 URL: https://how.complexsystems.fail/ [accessed 2025-05-25]
29. Dekker S. Drift into Failure: From Hunting Broken Components to Understanding Complex Systems: Ashgate Publishing Ltd; 2011.
30. Williams K. Safety, systems, complexity, and resilience: what makes organizations safe? In: Haxby E, Hunter D, Jaggar S, editors. An Introduction to Clinical Governance and Patient Safety: Oxford University Press; 2010.
31. Ibrahim Shire M, Jun GT, Robinson S. The application of system dynamics modelling to system safety improvement: present use and future potential. Saf Sci 2018 Jul;106:104-120. [doi: 10.1016/j.ssci.2018.03.010]
32. Qureshi ZH, Campbell A. Systemic safety and accident modelling of complex socio‐technical systems. INCOSE Int Symp 2009 Jul;19(1):21-35. [doi: 10.1002/j.2334-5837.2009.tb00935.x]
33. Zarei E, Khan F, Abbassi R. A dynamic human-factor risk model to analyze safety in sociotechnical systems. Process Saf Environ Prot 2022 Aug;164:479-498. [doi: 10.1016/j.psep.2022.06.040]
34. Resilience assessment based on models of functional resonance. Linköping University. 2008. URL: https://www.resilience-engineering-association.org/download/resources/symposium/symposium-2006/Woltjer_text.pdf [accessed 2025-05-25]
35. Developing and using program logic: a guide. NSW Health. 2023. URL: https://www.health.nsw.gov.au/research/Publications/developing-program-logic.pdf [accessed 2025-05-25]
36. Model for improvement: establishing measures. Institute for healthcare improvement. 2025 Feb 23. URL: https://www.ihi.org/learn/courses/open-school/catalog/qi-102 [accessed 2025-05-25]

Abbreviations

HF: human factors
HIT: health IT

Edited by A Choudhury; submitted 02.Mar.2025; peer-reviewed by J Poots, J Haun; revised version received 11.Dec.2025; accepted 26.Dec.2025; published 03.Feb.2026.

Please cite as:
Awad S, Begg R, Loveday T, Baillie A, Baysari M
Expert Views on Criteria for Evaluation of Human Factors Methods: Qualitative Interview Study
JMIR Hum Factors 2026;13:e73324
URL: https://humanfactors.jmir.org/2026/1/e73324
doi:10.2196/73324

© Selvana Awad, Rachel Begg, Thomas Loveday, Andrew Baillie, Melissa Baysari. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 3.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Interest in and Predictors of Engagement With a Virtual Reality Intervention Among People With Chronic Pain: Cross-Sectional Survey Study

Genevieve R Bryant[1], BA; Samuel A Holzman[2], BA; Hector R Perez[2], MD, MS

[1]SUNY Downstate College of Medicine, Brooklyn, NY, United States

[2]Albert Einstein College of Medicine, Bronx, NY, United States

Corresponding Author:
Hector R Perez, MD, MS
Albert Einstein College of Medicine
3300 Kossuth Ave
Bronx, NY, 10467
United States
Phone: 1 718 920 5756
Email: hector.perez@einsteinmed.edu

Abstract

Background: Although chronic pain (CP) is highly prevalent, current modalities are not sufficient to address the needs of people living with this condition. Pharmacological treatments for CP can have severe side effects and increased likelihood of patients overdosing or developing addiction. Behavioral treatments are often indicated for the treatment of CP, but barriers to treatment are common. Virtual reality (VR)–based interventions have shown promise as an effective and potentially accessible form of treatment for CP. However, previous research on VR interventions for people living with CP has not often included diverse populations, including racial and ethnic minority groups and people with low socioeconomic status.

Objective: This study aimed to gauge the interest of patients with CP in participating in a hypothetical study of at-home VR for CP and to identify predictors of interest. Patients were recruited from a low socioeconomic and racially and ethnically diverse community.

Methods: A total of 48 participants living with CP were recruited from an electronic medical record database, a research participant database, and a pain clinic, and they completed surveys about demographics, pain levels, technology use, and knowledge of VR. Bivariate testing was used to determine which, if any, of the aforementioned variables were associated with interest in a hypothetical study of at-home VR for CP. Stepwise logistic regression models predicting interest were built based on bivariate testing. Finally, we used a thematic analysis framework to analyze an additional open-ended question about reasons for interest in participating in a VR intervention for CP.

Results: Despite low technology use and little knowledge and experience with VR, results showed high interest (42/48, 88%) among patients in participating in a hypothetical study of at-home VR for CP. More frequent email use and using Facebook demonstrated nonsignificant trends toward interest in participating in a VR clinical trial for pain (P=.06 for email use and P=.06 for Facebook use). In stepwise multivariate models controlling for pain score, Facebook use was predictive of being somewhat or very interested in participating in a VR clinical trial for pain (P=.047). Open-ended responses tended to cite the novelty of VR and desperation for pain relief as reasons for participants' interest.

Conclusions: We found high interest in participating in a clinical trial of VR despite low use of technology and low knowledge of VR. Future fully powered studies should seek to confirm the effectiveness of VR treatments for people with CP, especially people from lower socioeconomic, and racially and ethnically diverse backgrounds.

(JMIR Hum Factors 2026;13:e70598) doi:10.2196/70598

KEYWORDS

chronic pain; virtual reality; barriers to treatment; technology-based intervention; accessibility

XSL•FO
RenderX

Introduction

Chronic pain (CP) affects an estimated 50.2 million adults in the United States [1], and it remains one of the most common reasons people seek medical care [2]. Individuals who live in poverty, have less than a high school education, and are on public health insurance are more likely to have high-impact CP [3]. Moreover, CP is linked to numerous adverse physical and mental health outcomes [3].

Despite the burden of CP, there is no gold standard treatment, and current options are insufficient for many people living with CP. Many medications, such as antidepressants and antiepileptic drugs, have severe side effects [4]. Patients are often dissatisfied with conventional pharmacological pain treatment, reporting that CP continues to affect their health, relationships, happiness, and quality of life [5]. Opioid medications can be ineffective and carry significant risk for patients, including the potential for addiction and overdose [6,7]. Nonpharmacologic options are recommended for treating CP, but barriers to accessing them are common, including lack of availability, distance to travel, and high cost of treatment [8].

In numerous studies, patient participants seek new, innovative treatment options for CP. Virtual reality (VR) is a promising new category of therapeutic options for pain management without many of the drawbacks of conventional CP treatments. VR is effective in the treatment of acute pain [9], and emerging evidence suggests it can be effective in the treatment of CP [10]. Additionally, portable VR headsets can be used at home or in a clinic, potentially increasing availability and accessibility of treatment. Interventions involving VR may also become more affordable and more widely available compared with other current treatments as the cost of such technology continues to decrease [11,12].

As VR technology develops, the need for robust study is paramount. Nonetheless, racial and ethnic minority groups and individuals of lower socioeconomic status (SES) may face barriers to entering new VR clinical trials due to preexisting disparities in pain clinical trial enrollment and, potentially, due to low interest in technology interventions [13]. Most VR clinical trials for individuals with CP in the United States to date have included samples primarily composed of non-Hispanic White, high-SES participants [14,15].

To ensure effectiveness in all populations and to ensure equity, further clinical studies of diverse patient populations are necessary. As part of preparations for recruiting participants to a pilot study of VR for CP, we conducted a cross-sectional study of patients with CP in a low SES and racially and ethnically diverse county. The survey was aimed at identifying patients' interest in participating in a hypothetical at-home VR trial and factors associated with interest.

Methods

Overview

We conducted a cross-sectional study using telephone and in-person surveys among patients with CP to evaluate interest in participating in a study of VR for pain management and to examine factors associated with interest.

Ethical Considerations

The study was approved by the Albert Einstein College of Medicine and Montefiore Medical Center Institutional Review Board through the Office of Human Research Affairs (2021-13108**)**. Oral informed consent was obtained over the telephone at the time of the interview. Identifying data was omitted if not essential and are not published in this manuscript. Participants were reimbursed US $10 in electronic gift cards for their time at the conclusion of the survey.

Participants

We recruited patients with CP in the Bronx, New York, both over the phone and in person. Bronx County is the poorest county in New York State and has the lowest internet adoption rates of any borough in New York City [16]. One-third of the Bronx population is Black or African American, 55% is Hispanic or Latino, and 9% is non-Hispanic White [17]. The median household income is US $45,517, with 27.7% of the Bronx population living in poverty [17].

Recruitment and Eligibility

There were 3 methods used for recruitment. First, we used hospital databases to identify potential participants based on *International Classification of Diseases, Tenth Revision (ICD-10)* codes. Second, we used a separate database that consisted of patients with CP that had completed previous research studies at our institution and consented to be recontacted for future research opportunities. Potential participants were identified from these 2 databases and contacted over the phone. Third, patients were recruited in person from a general medicine CP clinic. To be eligible to complete the survey, patients had to be aged at least 18 years, speak English fluently, and be able to provide consent for themselves.

Data Collection

Data were collected between August 2021 and December 2022 and deidentified for analysis. Participants completed the following questionnaires.

Sociodemographic Questions

We asked participants about their age, highest educational level, and race and ethnicity. Educational level was dichotomized into 2 groups: less than high school education and high school education or higher.

Pain Questions

We asked participants about their CP using the Pain, Enjoyment of Life, and General Activity (PEG) scale [18]. The PEG scale is a brief 3-item scale that assesses pain intensity and interference. The 3 questions assess pain on average in the past week, how much pain has interfered with their general activity, and how much pain has interfered with their enjoyment of life. Answer choices for each of the 3 questions were given on a 0- to 10-point scale, where 0 was "no pain" or "it does not interfere" and 10 was "pain as bad as you can imagine" or "it completely interferes." The PEG score is calculated using the mean of the answers to the 3 questions. Additionally, we asked

XSL•FO
RenderX

participants about the level of bodily pain they had experienced in the past week. The answer choices were “none,” “very mild,” “mild,” “moderate,” “severe,” and “very severe.” We also asked participants about all locations where they experienced pain for at least the last 3 months.

Media and Technology Usage Scale

We asked participants about their technology use and attitudes about technology using the Media and Technology Usage Scale [19]. This scale measures (1) the use frequency of a variety of technology and media types and (2) attitudes toward technology. There are 11 usage subscales representing smartphone use, general social media use, internet searching, emailing, media sharing, SMS text messaging, video gaming, online friendships, Facebook (Meta Platforms) friendships, phone calling, and watching television. Answers to questions related to these 11 subscales are scored on a 10-point frequency scale: 1=never, 2=once a month, 3=several times a month, 4=once a week, 5=several times a week, 6=once a day, 7=several times a day, 8=once an hour, 9=several times an hour, 10=all the time. Notably, the smartphone usage scale described above reflects using functions of the smartphone, such as apps, and does not reflect using the smartphone for calling or texting; calling and texting are represented in separate scales. Attitudes toward technology are rated on a 5-point Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly agree). Questions in this domain include (1) positive attitudes toward technology, (2) negative attitudes toward technology, and (3) anxiety or dependence on technology.

VR Questions

In total, we asked 6 questions about VR. First, we asked participants a set of four questions about their experience and perception of VR: (1) whether they had ever heard of VR, (2) whether they had used a VR device, (3) whether they had heard of VR being used to treat pain, and (4) whether they thought VR could work for treating pain. Participants could answer yes or no for the first 3 questions and had the additional option of answering not sure for the fourth question. A fifth question asked if participants would be interested in participating in a hypothetical study of at-home VR for CP and could answer that they were “not interested,” “somewhat interested,” or “very interested.” A sixth question was an open-ended question for participants to provide reasons for their interest or lack of interest.

Data Analysis

All participant data were included in the analysis. Information about gender was extracted from the electronic medical record. We calculated descriptive statistics (mean, median, and SD) for the PEG score, sociodemographic measures, and each of the subscales on the Media and Technology Usage Scale.

Our dependent variable was interest in participating in a hypothetical at-home VR study for CP. For this outcome, we dichotomized responses as “not interested” vs “somewhat interested” or “very interested.” While our dependent variable is subjective, in this open-ended survey, we were testing interest in participating in a future trial; therefore, we felt a subjective outcome was best suited for this study as there are many factors that may ultimately influence a decision to participate in a clinical trial, including availability, location, and incentives. Sociodemographic information, PEG score, pain location (back pain vs no back pain, knee pain vs no knee pain, and hip pain vs no hip pain), and the Media and Technology Usage Scale subscales were independent variables. We ran bivariate tests to determine whether independent variables were associated with patient interest in a hypothetical VR trial. For hypothesis testing in bivariate models, we set the double-sided α to .05. We created multivariate regression models using a stepwise approach. We identified variables a priori, including age, gender, and PEG score, to be included in the initial model regardless of significance. Additional covariates associated with the outcome in bivariate analyses at $P<.20$ were sequentially added to the model. Final model selection was guided by theoretical relevance, model fit, and statistical significance.

To summarize open-ended responses about participants’ interest in a VR trial, we coded responses using a thematic analysis framework. We reviewed codes and developed categories to reflect responses. After each response was categorized, we tabulated categories quantitatively. Where possible, we present illustrative text reflecting categories.

Results

Demographic and Survey Outcomes

Sociodemographic Variables

Of 693 participants contacted, 215 (31%) were successfully reached. Among those successfully reached, 124 (57.7%) declined, and 43 (20%) did not meet the inclusion criteria. The most common reason for exclusion in this study was only speaking Spanish (n=41, 15% of successful contacts). Of 172 potential participants, our final sample consisted of 48 (27.9%) individuals who met the inclusion criteria. The mean age of participants was 58.5 (SD 11.3) years. Of the participants, 32 (67%) were female and 16 (33%) were male. About half (n=23, 47.9%) of the participants had a high school education or less. About one-fifth (n=9, 19%) of participants were non-Hispanic White; 15 (31%) were Black; 18 (38%) were Hispanic; and 6 (13%) were Asian, Native Hawaiian or other Pacific Islander, or other. Please see Table 1 for detailed demographics.

XSL•FO
RenderX

Table 1. Survey participant characteristics (N=48).

Characteristic	Total cohort (n=48)	Interested (n=42)	Not interested (n=6)	*P* value
Age (y), mean (SD; 95% CI)	58.5 (11.3; 55.2-61.9)	58.5 (11.1; 55.0-61.9)	58.7 (13.5; 44.5-72.8)	.97
Gender, n (%; 95% CI)				.99
Woman	32 (67; 52.5-78.9)	28 (67; 51.0-79.9)	4 (67; 29.6-90.9)	
Education, n (%; 95% CI)				.45
High school or less	23 (48; 34.3-62.0)	21 (50; 35.3-64.7)	2 (33; 9.7-70.0)	
Race or ethnicity, n (%; 95% CI)				.99
White, non-Hispanic	9 (19; 10.3-32.0)	8 (19; 9.8-33.6)	1 (17; 3.0-56.4)	
Black, non-Hispanic	15 (31; 19.7-45.4)	13 (31; 19.0-46.3)	2 (33; 9.7-70.0)	
Hispanic	18 (38; 24.9-52.0)	16 (38; 24.7-53.8)	2 (33; 9.7-70.0)	
Other	6 (13; 5.6-25.6)	5 (12; 5.0-26.1)	1 (17; 3.0-56.4)	
Pain severity, n (%; 95% CI)				.89
Moderate or higher	39 (81; 67.4-90.3)	34 (81; 66.7-90.1)	5 (83; 36.5-99.1)	
Pain intensity (PEG[a] score), mean (SD; 95% CI)	6.31 (2.4; 5.61-7.01)	6.17 (2.5; 5.40-6.93)	7.28 (1.9; 5.31-9.25)	.30
Texting use, mean (SD; 95% CI)[a]	5.7 (1.9; 5.2-6.2)	5.8 (1.8; 5.2-6.3)	5.1 (2.9; 2.1-8.1)	.45
Smartphone use, mean (SD; 95% CI)[b]	4.2 (1.9; 3.6-4.7)	4.3 (1.9; 3.7-4.9)	3.3 (1.6; 1.6-5.0)	.21
Email use, mean (SD; 95% CI)[b]	4.4 (2.2; 3.8-5.0)	4.6 (2.2; 4.0-5.3)	2.8 (2.1; 0.6-5.0)	.06
Internet use, mean (SD; 95% CI)[b]	3.6 (3.0-4.2)	3.6 (3.0-4.3)	3.3 (1.3-5.3)	.75
Facebook user[c], n (%; 95% CI)	25 (2; 52.1; 37.9-66.0)	24 (2; 57.1; 41.0-71.9)	1 (2; 16.7; 3.0-56.4)	.06

[a]PEG: Pain, Enjoyment of Life, and General Activity.

[b]Use is rated 1 to 10 on the Media and Technology Usage Scale, with higher values representing more use. The scale is as follows: 1=never, 2=once a month, 3=several times a month, 4=once a week, 5=several times a week, 6=once a day, 7=several times a day, 8=once an hour, 9=several times an hour, 10=all the time.

[c]Participants were asked if they have a Facebook account.

Pain Measures

The mean PEG score was 6.31 (SD 2.4), and 39 (81%) participants reported having pain that was at least moderate. Lower back pain was the most common pain site (n=30, 63%), followed by knee pain (n=17, 35%) and hip pain (n=14, 29%).

Media and Technology Usage Scale

The mean texting usage subscale score was 5.7 (SD 1.9), reflecting text message use between several times a week and once daily. The mean phone use was 5.6 (SD 1.8), again reflecting calling or receiving calls from others between several times a week and once daily. The mean smartphone usage subscale was 4.2 (SD 1.9), which reflects using smartphone functions about once per week. The mean email use was 4.4 (SD 2.2), and the mean internet use was 3.6 (SD 2.0), with both reflecting nondaily use of these technologies. Of the 48 participants, about half (n=25, 52%) reported using Facebook. Participants generally reported mixed attitudes toward technology; mean subscale scores on positive attitude scales with technology were 3.8 (SD 0.8), reflecting a stance approaching agreement; mean subscale scores on negative attitude scales were 3.1 (SD 1.0), reflecting a neutral stance; and mean subscale scores on anxiety or dependence on technology were 2.8 (SD 1.1), reflecting a neutral stance approaching disagreement that participants were anxious or dependent upon technology.

VR Questions

Of the 48 participants, about two-thirds (n=31, 65%) had never heard about VR, 41 (85%) had never tried a VR device, and 2 (4%) had heard about VR being used to treat pain. A small minority of participants (n=7, 15%) thought that VR could work for treating pain, with a majority (n=38, 79%) reporting being unsure. Despite this, about two-thirds (n=31, 65%) of participants were “very interested” in participating in a hypothetical study of at-home VR for CP, and an additional 11 (23%) were “somewhat interested.” As we had defined our primary outcome, interest in VR, as either “very interested” or “somewhat interested,” most participants (n=42, 88%) met the outcome. Only 6 (13%) participants reported not interested in participating in a hypothetical study of at-home VR for CP.

Factors Associated With Interest in Participating in a VR Trial

Relationships between (1) more frequent email use and (2) Facebook use and interest in participating in a VR clinical trial for pain approached statistical significance and were subsequently included in stepwise multivariate models (email subscale score among interested vs not interested participants:

XSL•FO
RenderX

4.6 vs 2.8; *P*=.06 and Facebook use among interested vs not interested participants: 24/42, 57% vs 1/6, 17%; *P*=.06). Age, gender, race, education, pain severity, pain location, smartphone use, and internet use were not predictive of being somewhat or very interested in participating in a VR clinical trial.

We examined predictors of interest across a series of multivariate models. Variables identified a priori included gender, age, and PEG score. These were added to a baseline model along with email use and Facebook use, which were identified through bivariate analyses. The initial model (model 1) was not statistically significant overall (*P*=.14). None of the individual predictors reached significance, though Facebook use approached significance (odds ratio [OR] 9.50, 95% CI 0.68-132.21; *P*=.09). After simplification (model 2: email use, PEG score, and Facebook use), the overall model became statistically significant (*P*=.04), and Facebook use remained a consistent predictor, although it did not reach statistical significance (OR 8.56, 95% CI 0.69-106.38; *P*=.10). In the final reduced model (model 3: PEG score and Facebook use), the model retained statistical significance (*P*=.04), and Facebook use emerged as a statistically significant predictor of VR interest (OR 11.37, 95% CI 1.04-124.81; *P*=.047).

Open-Ended Qualitative Survey Responses: Reasons for Interest in VR

We collected 47 open-ended responses describing reasons for interest or lack of interest in participating in a VR clinical trial. We categorized responses into 7 separate categories. Two categories were related to interest: (1) interest in a new treatment modality and (2) desperation for pain relief. Four categories were related to lack of interest: (1) doubts about VR as a pain treatment, (2) fear of COVID-19, (3) scheduling and traveling concerns, and (4) lack of more information. Three coded responses could not be labeled as interest or lack of interest and instead were marked as "other." The most common category was interest in a new treatment modality (27/47, 58%), followed by desperation for pain relief (9/47, 19%). The most common category representing lack of interest was doubts about VR as a pain treatment (3/47, 6%), followed by scheduling and traveling concerns (2/47, 4%). A sample quote reflecting a participant's interest was as follows:

> *You never know what will work. There's always advancements and new things that they discover outside of pills. I'm open to anything.*

A sample quote reflecting a participant's lack of interest was as follows:

> *My pain can only be helped by a doctor. No way virtual reality works for pain. If your boss is telling you that then they are lying.*

Discussion

Principal Findings

In this telephone survey of 48 persons with CP, we found that, despite a lack of experience with and knowledge of VR, most participants were interested in a study of at-home VR for CP. In this racially and ethnically diverse sample, we found relatively infrequent use of smartphones, email, internet, and social media. Additionally, we found mixed evidence that technology use was predictive of interest in participating in a clinical trial for VR. In open-ended responses, reasons for interest tended to cite new treatment modalities and the high burden of CP. These results indicate that the desire for new and innovative technological interventions exists even among people with CP with relatively low technology use.

Our results did not find patient-level barriers to participating in technology-oriented clinical trials, including a lack of trust. Prior studies have indicated that mistrust of research institutions and investigators is a significant attitudinal barrier to research participation reported by African Americans [20]. Overall, non-Hispanic Black and Hispanic individuals exhibit much higher levels of medical mistrust than their White counterparts [21]. Nonetheless, our qualitative data did not suggest that mistrust was a major barrier for participants. Other patient-level systemic barriers, such as scheduling and transportation, were more prominent in open-ended responses. These barriers may restrict participation in clinical research, and any trial that seeks diverse participation should consider efforts to reduce systemic barriers in recruitment procedures.

By contrast, institutional barriers and researcher-level barriers can be a prominent systemic reason for exclusion in clinical trials. Institutional barriers can include inadequate or untimely reimbursement of participants, insufficient attention to research important to communities of racial and ethnic minority people, and challenges with institutional review boards and community partnerships [22]. Researcher-level barriers can include bias, among many other factors. A notable form of bias in pain research is the underrecognition of pain in minorities by physicians [23]. The belief that Black persons feel less pain than White persons remains prevalent among US physicians, leading to inadequate treatment recommendations and possibly lower referral to pain research [23,24]. Perceptions that individuals from certain racial and ethnic groups are challenging or not ideal study candidates also remain prevalent and lead to less diverse participation [25]. In this study, one important researcher-level barrier participants encountered was that we required English proficiency. Because we were conducting a cross-sectional survey of interest in participation in a VR trial, we elected to only include people who could participate in future VR studies. As of 2025, none of the current US Food and Drug Administration–approved devices for CP are available in Spanish.

The high level of interest in VR in this study implies that institutional and researcher-based systemic barriers could be more prominent for this population than patient-level barriers. Previous literature in other intervention modalities has been mixed on this topic. When given the opportunity to participate in clinical research, some studies show that racial and ethnic minority groups are as likely to participate as White populations, whereas others report the opposite [26-29]. One overarching conclusion of the literature to date is that opportunity to participate or interest in participating alone is often not enough, as actual participation still lags due to a myriad of factors. Ensuring that research opportunities are relevant and significant to populations of interest is absolutely critical, as is lowering

XSL•FO
RenderX

systemic barriers to participation, such as by providing off-hour availability or transportation reimbursement.

Additionally, our results challenge the perception that a population characterized by low technology use has less interest in participating in clinical trials involving technology. We found mixed evidence that technology use was related to interest in VR. Though we found that Facebook use was associated with VR interest in multivariate models, other proxy measures for technology use, including internet use, were not similarly related. We theorize that, perhaps aside from technological familiarity, participants who maintained social ties on the internet may have a particular interest in technological solutions to their CP. For the entire sample, the high interest we observed may also reflect ongoing frustration with the treatment options provided by the health care system among patients with CP. In our open-ended responses, some participants indicated that they were desperate for pain relief and willing to try novel interventions. The extent to which participants feel more comfortable with technology-based research compared to traditional clinical research involving medications or other medical interventions, owing to historic mistrust and inequities of traditional research, is unclear and warrants further research.

Health care systems that adopt VR into routine pain management can increase patient interest and subsequent participation through increased exposure. For example, the Veterans Health Administration (VHA), the largest integrated health care system in the United States, is in the early stages of VR adoption for CP, with patients showing high levels of interest [30]. The VHA has established an Office of Healthcare Innovation and Learning and a specific VA Immersive Program, which has supported literature reviews, conferences, and innovative pilot programs in VR studies [31]. However, because of unique payment incentives, public and private hospital systems lag behind the VHA in technology adoption [32]. Nonetheless, there are new financial options to sustain VR care. Although VR for CP is not routinely covered by Medicare, VR devices can be covered if they are declared reasonable and necessary [33]. Reimbursement through Medicare may lead to increased adoption in private health care systems, such as ours at Montefiore, and affiliated public hospitals that serve diverse populations.

Limitations

The primary limitation of this study is the lack of generalizability of our sample to the patient population. Our sample was small and hospital based, which may have introduced response bias. Some participants may have previously participated in clinical trials. Additionally, we were unable to reach many of our potential participants, and as such, we enrolled a relatively small proportion of patients we attempted to contact. Our telephone response rate is similar to recruitment rates in other studies we have conducted, but there may have been selection bias in the participants that responded to our phone call. However, the directionality of the bias is unclear; patients who answer telephone calls and use telephones more often may be less technologically savvy compared with those who communicate more with friends and relatives via SMS text message or email. By contrast, as observed in our previous studies and those of others withing our health care system and community [34], patients with economic challenges may not have consistent telephone access or consistent telephone numbers and may represent groups that are less technologically savvy. Additionally, we estimated engagement in clinical trial activities by asking about subjective interest, which may overestimate intent to participate because of response bias. We did not assess negative interactions with the health care system and how these may have affected interest in VR clinical trials. Nonetheless, to our knowledge, this is the first study to gauge the level of interest in a VR clinical trial among a predominantly Hispanic and Black patient population with CP.

Conclusions

Patient populations of low SES and racial and ethnic minority groups are not sufficiently included in pain clinical trial recruitment. In this study, we found high interest in participating in a VR study for CP despite low technology use and lack of knowledge of VR, underscoring the importance of not underestimating patient willingness to engage with innovative treatments for CP. Further research is necessary to include diverse populations to ensure equity and effectiveness of VR, especially considering the high burden of CP among these populations. Addressing language accessibility and structural challenges, such as transportation and Medicare or Medicaid reimbursement, will be essential to ensure inclusive clinical trial recruitment and adoption of VR for CP.

Conflicts of Interest

None declared.

References

1. Zelaya CE, Dahlhamer JM, Lucas JW, Connor EM. Chronic pain and high-impact chronic pain among U.S. adults, 2019. NCHS Data Brief 2020 Nov(390):1-8. [Medline: 33151145]
2. Rikard SM, Strahan AE, Schmit KM, Guy Jr GP. Chronic pain among adults - United States, 2019-2021. MMWR Morb Mortal Wkly Rep 2023 Apr 14;72(15):379-385 [FREE Full text] [doi: 10.15585/mmwr.mm7215a1] [Medline: 37053114]
3. Dahlhamer J, Lucas J, Zelaya C, Nahin R, Mackey S, DeBar L, et al. Prevalence of chronic pain and high-impact chronic pain among adults - united states, 2016. MMWR Morb Mortal Wkly Rep 2018 Sep 14;67(36):1001-1006 [FREE Full text] [doi: 10.15585/mmwr.mm6736a2] [Medline: 30212442]
4. Cohen SP, Vase L, Hooten WM. Chronic pain: an update on burden, best practices, and new advances. Lancet 2021 May 29;397(10289):2082-2097. [doi: 10.1016/S0140-6736(21)00393-7] [Medline: 34062143]

XSL•FO
RenderX

5. Purcell N, Zamora K, Gibson C, Tighe J, Chang J, Grasso J, et al. Patient experiences with integrated pain care: a qualitative evaluation of one VA's biopsychosocial approach to chronic pain treatment and opioid safety. Glob Adv Health Med 2019;8:2164956119838845 [FREE Full text] [doi: 10.1177/2164956119838845] [Medline: 31041143]
6. Busse JW, Wang L, Kamaleldin M, Craigie S, Riva JJ, Montoya L, et al. Opioids for chronic noncancer pain: a systematic review and meta-analysis. JAMA 2018 Dec 18;320(23):2448-2460 [FREE Full text] [doi: 10.1001/jama.2018.18472] [Medline: 30561481]
7. Dowell D, Ragan KR, Jones CM, Baldwin GT, Chou R. CDC clinical practice guideline for prescribing opioids for pain - United States, 2022. MMWR Recomm Rep 2022 Nov 04;71(3):1-95 [FREE Full text] [doi: 10.15585/mmwr.rr7103a1] [Medline: 36327391]
8. Becker WC, Dorflinger L, Edmond SN, Islam L, Heapy AA, Fraenkel L. Barriers and facilitators to use of non-pharmacological treatments in chronic pain. BMC Fam Pract 2017 Mar 20;18(1):41 [FREE Full text] [doi: 10.1186/s12875-017-0608-2] [Medline: 28320337]
9. Pourmand A, Davis S, Marchak A, Whiteside T, Sikka N. Virtual reality as a clinical tool for pain management. Curr Pain Headache Rep 2018 Jun 15;22(8):53. [doi: 10.1007/s11916-018-0708-2] [Medline: 29904806]
10. Jones T, Moore T, Choo J. The impact of virtual reality on chronic pain. PLoS One 2016;11(12):e0167523 [FREE Full text] [doi: 10.1371/journal.pone.0167523] [Medline: 27997539]
11. Goudman L, Jansen J, Billot M, Vets N, De Smedt A, Roulaud M, et al. Virtual reality applications in chronic pain management: systematic review and meta-analysis. JMIR Serious Games 2022 May 10;10(2):e34402 [FREE Full text] [doi: 10.2196/34402] [Medline: 35536641]
12. Mallari B, Spaeth EK, Goh H, Boyd BS. Virtual reality as an analgesic for acute and chronic pain in adults: a systematic review and meta-analysis. J Pain Res 2019;12:2053-2085 [FREE Full text] [doi: 10.2147/JPR.S200498] [Medline: 31308733]
13. Clark LT, Watkins L, Piña IL, Elmer M, Akinboboye O, Gorham M, et al. Increasing diversity in clinical trials: overcoming critical barriers. Curr Probl Cardiol 2019 May;44(5):148-172 [FREE Full text] [doi: 10.1016/j.cpcardiol.2018.11.002] [Medline: 30545650]
14. Grassini S. Virtual reality assisted non-pharmacological treatments in chronic pain management: a systematic review and quantitative meta-analysis. Int J Environ Res Public Health 2022 Mar 29;19(7):4071 [FREE Full text] [doi: 10.3390/ijerph19074071] [Medline: 35409751]
15. Ding ME, Traiba H, Perez HR. Virtual reality interventions and chronic pain: scoping review. J Med Internet Res 2025 Feb 18;27:e59922. [doi: 10.2196/59922] [Medline: 39965778]
16. Vogels EA. Digital divide persists even as Americans with lower incomes make gains in tech adoption. Pew Research Center. 2023. URL: https://www.pewresearch.org/short-reads/2021/06/22/digital-divide-persists-even-as-americans-with-lower-incomes-make-gains-in-tech-adoption/ [accessed 2025-05-29]
17. Income and poverty, race and ethnicity. U.S. Census Bureau. URL: https://data.census.gov/profile/Bronx_County,_New_York?g=050XX00US36005#income-and-poverty [accessed 2025-11-07]
18. Krebs EE, Lorenz KA, Bair MJ, Damush TM, Wu J, Sutherland JM, et al. Development and initial validation of the PEG, a three-item scale assessing pain intensity and interference. J Gen Intern Med 2009 Jun;24(6):733-738 [FREE Full text] [doi: 10.1007/s11606-009-0981-1] [Medline: 19418100]
19. Rosen LD, Whaling K, Carrier LM, Cheever NA, Rokkum J. The media and technology usage and attitudes scale: an empirical investigation. Comput Human Behav 2013 Nov 01;29(6):2501-2511 [FREE Full text] [doi: 10.1016/j.chb.2013.06.006] [Medline: 25722534]
20. Tanner A, Kim SH, Friedman DB, Foster C, Bergeron CD. Barriers to medical research participation as perceived by clinical trial investigators: communicating with rural and african american communities. J Health Commun 2015;20(1):88-96. [doi: 10.1080/10810730.2014.908985] [Medline: 25204763]
21. Bazargan M, Cobb S, Assari S. Discrimination and medical mistrust in a racially and ethnically diverse sample of California adults. Ann Fam Med 2021;19(1):4-15 [FREE Full text] [doi: 10.1370/afm.2632] [Medline: 33431385]
22. National Academies of Sciences, Engineering, and Medicine. Improving Representation in Clinical Trials and Research: Building Research Equity for Women and Underrepresented Groups. Washington, DC: National Academies Press; 2022.
23. Hoffman KM, Trawalter S, Axt JR, Oliver MN. Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites. Proc Natl Acad Sci U S A 2016 Apr 19;113(16):4296-4301 [FREE Full text] [doi: 10.1073/pnas.1516047113] [Medline: 27044069]
24. Aronowitz SV, Mcdonald CC, Stevens RC, Richmond TS. Mixed studies review of factors influencing receipt of pain treatment by injured black patients. J Adv Nurs 2020 Jan;76(1):34-46. [doi: 10.1111/jan.14215] [Medline: 31566791]
25. Niranjan SJ, Martin MY, Fouad MN, Vickers SM, Wenzel JA, Cook ED, et al. Bias and stereotyping among research and clinical professionals: perspectives on minority recruitment for oncology clinical trials. Cancer 2020 Jan 01;126(9):1958-1968 [FREE Full text] [doi: 10.1002/cncr.32755] [Medline: 32147815]
26. Kim JY, Botto E, Ford RM. A clinical research interaction scale for racial and ethnic minority participants. JAMA Netw Open 2025 May 01;8(5):e259481 [FREE Full text] [doi: 10.1001/jamanetworkopen.2025.9481] [Medline: 40358951]

XSL•FO
RenderX

27. Igwe E, Woodburn J, Davolos J, Shollenberger C, Miller SM, Hernandez E, et al. Patient perceptions and willingness to participate in clinical trials. Gynecol Oncol 2016 Sep;142(3):520-524. [doi: 10.1016/j.ygyno.2016.06.015] [Medline: 27372403]
28. Pariera KL, Murphy ST, Meng J, McLaughlin ML. Exploring willingness to participate in clinical trials by ethnicity. J Racial Ethn Health Disparities 2017 Aug 7;4(4):763-769. [doi: 10.1007/s40615-016-0280-6] [Medline: 27604379]
29. Chen N, Freeman JQ, Zhao F, Goldberg L, Yarlagadda SR, Terman E, et al. Clinical trial discussion and participation in a breast cancer cohort by race and ethnicity. JAMA Netw Open 2025 Jun 02;8(6):e2515205 [FREE Full text] [doi: 10.1001/jamanetworkopen.2025.15205] [Medline: 40504527]
30. Liu K, Madrigal E, Chung JS, Parekh M, Kalahar CS, Nguyen D, et al. Preliminary study of virtual-reality-guided meditation for veterans with stress and chronic pain. Altern Ther Health Med 2023 Sep;29(6):42-49. [Medline: 34559692]
31. VA Office of Healthcare Innovation and Learning. URL: https://www.innovation.va.gov/hil/home.html [accessed 2025-08-29]
32. Woolhandler S, Toporek A, Gao J, Moran E, Wilper A, Himmelstein DU. Administration's share of personnel in veterans health administration and private sector care. JAMA Netw Open 2024 Jan 02;7(1):e2352104 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.52104] [Medline: 38236601]
33. Tuerk PW, Fortney J, Bosworth HB, Wakefield B, Ruggiero KJ, Acierno R, et al. Toward the development of national telehealth services: the role of Veterans Health Administration and future directions for research. Telemed J E Health 2010;16(1):115-117. [doi: 10.1089/tmj.2009.0144] [Medline: 20043704]
34. Azhar S, Ross AM, Acharya A, Lerner R, Tripathi S, Keller E, et al. “The little I receive is not enough’’: a qualitative study of food insecurity in the South Bronx. Food Cult Soc 2024 Oct 11;28(4):1035-1054. [doi: 10.1080/15528014.2024.2406081]

Abbreviations

CP: chronic pain
OR: odds ratio
PEG: Pain, Enjoyment of Life, and General Activity
SES: socioeconomic status
VHA: Veterans Health Administration
VR: virtual reality

Edited by A Kushniruk; submitted 07.Jan.2025; peer-reviewed by E Maslowski, E Ross; comments to author 07.Apr.2025; revised version received 07.Nov.2025; accepted 25.Nov.2025; published 11.Feb.2026.

Please cite as:
Bryant GR, Holzman SA, Perez HR
Interest in and Predictors of Engagement With a Virtual Reality Intervention Among People With Chronic Pain: Cross-Sectional Survey Study
JMIR Hum Factors 2026;13:e70598
URL: https://humanfactors.jmir.org/2026/1/e70598
doi:10.2196/70598
PMID:

©Genevieve R Bryant, Samuel A Holzman, Hector R Perez. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 11.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Consensus on Core Competencies of Undergraduate Life Sciences Students for Entry-Level Participation in Clinical Trials: RAND-Modified Delphi Study

Ángela Cano[1,2*], PhD; Patricia Cosano-Pérez[2,3*], MSc; Juan José Castón[1,2*], PhD; Diana Corona-Mata[1,2*], PhD; Rosa Ballesteros-Olmedo[4*], MSc; Eva Venegas-Moreno[5*], MD; Jaime Perea-Cortés[5*], MSc; Concepcion Aranda-Valera[6*], PhD; Isabel Machuca[1,2*], PhD; Julian Torre-Cisneros[1,2*], PhD

[1]Department of Medical and Surgical Sciences, Faculty of Medicine, University of Córdoba, Reina Sofía University Hospital, Maimonides Biomedical Research Institute of Cordoba, Cordoba, Spain

[2]Center for Biomedical Research Network on Infectious Diseases, Madrid, Spain

[3]Maimonides Biomedical Research Institute of Cordoba, Cordoba, Spain

[4]Oncology Clinical Trials Unit, Virgen del Rocío University Hospital, Foundation for the Management of Health Research in Seville, Seville, Spain

[5]Endocrinology and Nutrition Management Unit, Institute of Biomedicine of Seville, Virgen del Rocío University Hospital, Spanish National Research Council, University of Seville, Seville, Spain

[6]Rheumatology Clinical Management Unit, Reina Sofía University Hospital, University of Cordoba, Cordoba, Spain

[*]all authors contributed equally

Corresponding Author:
Isabel Machuca, PhD
Department of Medical and Surgical Sciences, Faculty of Medicine, University of Córdoba, Reina Sofía University Hospital, Maimonides Biomedical Research Institute of Cordoba
Avd Menéndez Pidal s/n
Cordoba, 14004
Spain
Phone: 34 957011636
Email: isabelm.machuca.sspa@juntadeandalucia.es

Abstract

Background: The competencies required of life sciences graduates to participate in clinical trial research teams are poorly defined.

Objective: This study aimed to establish a structured consensus on the core competencies that students must acquire during their degree programs.

Methods: A RAND-modified Delphi procedure was conducted using 2 rounds of online questionnaires, a virtual meeting, and a final review. The process began with a questionnaire completed by experts from the University of Córdoba and the Maimonides Institute for Biomedical Research of Córdoba (3 principal investigators, 1 study coordinator, and 1 data manager). The 48 competencies initially proposed were reviewed by a panel of 18 experts representing 6 clinical research groups from universities and biomedical research institutes in Andalusia, Spain. The relevance of each proposed competency was rated on a 9-point Likert scale. Competencies were accepted if they achieved a mean score of 7 points or higher and at least 70% of ratings in the top tertile (scores of 7-9).

Results: The 18 experts achieved 100% consensus on a panel of 48 competencies organized into 3 sections: fundamental concepts (n=23, 47.9% of the competencies), clinical trial development (n=23, 47.9% of the competencies), and certifications (n=2, 4.2% of the competencies).

Conclusions: This consensus defines the core competencies required of undergraduate life sciences students for entry-level participation in clinical trial research. It also provides a framework for designing training strategies in both undergraduate and graduate programs of study.

(JMIR Hum Factors 2026;13:e89449) doi:10.2196/89449

XSL•FO
RenderX

KEYWORDS

clinical research; clinical trials as topic; competency-based education; Delphi technique; undergraduate education; life sciences education

Introduction

Clinical trials [1] are a cornerstone of human clinical research and are often performed to evaluate the efficacy and safety of new pharmaceutical drugs. Many hospitals have dedicated clinical research teams that conduct both registrational clinical trials sponsored by the pharmaceutical industry and "academic" trials led by scientific institutions.

The design and implementation of a clinical trial is a highly complex process involving a diverse team of professionals, typically life sciences graduates [2-4] (Table 1). Protocols must be submitted for evaluation and approval by health authorities such as the Spanish Agency for Medicines and Medical Devices under the Ministry of Health, Social Services, and Equality in Spain [5], as well as by research ethics committees [6,7]. Once approval is granted, trials are highly demanding as they require strict adherence not only to the protocol guidelines but also to rigorous ethical and legal standards established internationally and regulated nationally. Depending on their level of responsibility, each investigator on the team is expected to possess different levels of competency. However, the specific competencies (ie, knowledge, skills, and attitudes) that should be attained during undergraduate life sciences programs are poorly defined. In fact, there is little consensus regarding these general and basic competencies at the undergraduate level despite the emphasis often placed on translational clinical research [8]. Moreover, many life sciences graduates do not perceive clinical research as a potential area for professional development.

In this context, it is argued that competency-based educational frameworks are not static lists of skills but dynamic constructs developed and applied within complex professional environments. Professional expertise is increasingly understood as the integration of cognitive, technical, and nontechnical skills, including context-dependent decision-making [9-11]. From this perspective, competency development can be viewed as a progressive process that begins during undergraduate training and continues through postgraduate education and professional practice. Early exposure to these competencies may facilitate their integration into more complex, context-dependent performance, including the development of higher-order reasoning skills [12]. Importantly, this early exposure does not imply full mastery but rather the acquisition of foundational components that can be further developed in later stages of training. However, many undergraduate curricula in the life sciences do not systematically address these competencies, potentially limiting graduates' integration into multidisciplinary clinical research teams. In this context, competency-based models in medical education support the introduction of foundational competencies at the undergraduate level, which can be progressively developed across subsequent stages of training [8,13]. This approach is particularly relevant in clinical research, where multidisciplinary teams include professionals with diverse backgrounds, such as clinicians, study coordinators, data managers, or methodologists, who require a shared baseline of competencies to collaborate effectively. Furthermore, fostering early engagement in research among physicians and nurses may contribute to the development of clinician investigators and other research-oriented profiles while allowing for progressive specialization in line with individual career choices. Previous work has also shown that undergraduate curricula may provide exposure to foundational knowledge and skills, whereas applied domains requiring clinical judgment, critical appraisal, and decision-making often remain insufficiently developed, particularly in areas such as antimicrobial use and therapeutic decision-making [14].

These findings support the need for structured competency frameworks in clinical research. To this end, this study sought to establish a consensus on the core competencies required for participation in clinical trial research using a RAND-modified Delphi procedure [15,16]. This methodology has been previously validated for defining competencies based on expert opinion [8,17].

XSL•FO
RenderX

Table 1. Definitions.

Term	Definition
General competencies	Basic standards that an investigator must achieve depending on their training and level of responsibility; does not include discipline-specific competencies
Scientific knowledge	An organized body of knowledge obtained through scientific methods, including systematic observation and empirically testable hypotheses
Life sciences graduates	Individuals with undergraduate training in health and biological sciences (eg, medicine, nursing, pharmacy, biology, and biochemistry) who may participate in clinical research teams
Skill	The ability to perform a task correctly and efficiently
Attitude	A learned predisposition to respond consistently to a given object or situation
Clinical trial	A medical research study involving human participants
Academic clinical trial	A clinical trial funded by public or academic institutions to generate scientific knowledge
Commercial clinical trial	A clinical trial funded by industry for commercial and research purposes
Research ethics committee	A body responsible for the ethical review and oversight of research protocols
Protocol	A document describing the design, methodology, and organization of a clinical trial
Informed consent	A document that provides participants with information about a clinical trial, including risks and benefits, and requires signed agreement
GCP[a] certificate	A certification attesting to knowledge of GCP standards for conducting clinical trials
IATA[b] certificate	A certification for the safe handling and shipment of biological samples in accordance with IATA regulations
PI[c]	The individual responsible for conducting a clinical trial at a research site
Study coordinator	A professional responsible for coordinating and managing clinical trial activities
Data manager	A professional responsible for clinical data management and quality control
Clinical trial database	A structured dataset containing data collected from participants during a clinical trial
CRO[d]	A company contracted by the sponsor to manage and monitor clinical trials

[a]GCP: good clinical practice.

[b]IATA: International Air Transport Association.

[c]PI: principal investigator.

[d]CRO: contact research organization.

Methods

Study Design

We conducted a RAND-modified Delphi procedure [15-20] to establish consensus on the general and basic competencies that undergraduate life sciences students must acquire to participate in clinical trial research teams (see the key definitions in Multimedia Appendix 1). The consensus was developed between June 2024 and December 2024 and consisted of four stages: (1) a first round of questionnaires, (2) a virtual meeting of a panel of experts, (3) a second round of questionnaires, and (4) a final review. Depending on participant availability, the rounds of questionnaires were conducted online or via email by means of a survey. The data were analyzed using Microsoft Excel. Examples of the questionnaires are provided in Multimedia Appendix 1. The competencies were defined with a focus on foundational knowledge and awareness appropriate for entry-level participation in clinical trial investigation rather than on full performance of role-specific responsibilities. Additionally, they were structured across three predefined levels: (1) foundational (basic knowledge and understanding of key concepts), (2) intermediate (ability to apply knowledge under supervision in practical settings), and (3) advanced (capacity to perform tasks independently and make informed decisions within the context of clinical research).

Selection of the Panel of Experts

Six research groups from universities and biomedical research institutes in Andalusia, Spain, were selected for the study. Participants were invited through a 2-stage process. In the first stage, principal investigators (PIs) who were experts in the field, clinical investigators, and professors were identified at each institution. Each expert was then asked to invite 2 additional participants with more than 5 years of experience, ideally resulting in a panel of 18 experts comprising 6 clinical investigators with teaching experience, 6 data managers, and 6 clinical trial coordinators. Teleconferences were held with each group to ensure that all panel members had a similar understanding of the study objectives, definitions (Multimedia Appendix 1), and procedures. The panel members were asked to share any insights based on their own experiences. Additionally, all participants were provided with supporting documentation to standardize their understanding of competencies. The material was derived from content already included in the syllabus of the elective undergraduate medical

XSL•FO
RenderX

course Clinical Trials and Good Clinical Practice coordinated by AC and JT-C [21].

First Round of Questionnaires

The questionnaire was designed by a working group of 5 members from the University of Córdoba and Maimonides Institute for Biomedical Research of Córdoba: 3 clinical trial PIs, 1 study coordinator, and 1 data manager. After reviewing the literature, 48 competencies were selected and divided into three blocks: (1) fundamental concepts (n=23, 47.9% of the competencies), (2) clinical trial development (n=23, 47.9% of the competencies), and (3) certifications (n=2, 4.2% of the competencies). A 9-point Likert scale [22] ranging from 1 (not at all important) to 9 (critically important) was used consistently across all rounds.

Consensus was defined via a multistep process. In the first round, competencies were accepted if they achieved a mean score of 7 or higher and at least 70% of ratings in the top tertile (scores of 7-9). Competencies with a mean score of 7 points or higher but with less than 70% of scores in the top tertile were discussed at the expert panel meeting. Agreement was reached through structured discussion, aiming for at least 80% consensus among participants. Competencies with a mean score of less than 7 were not included.

Respondents could provide comments on each competency, including suggestions for rewording or removing items as well as proposals for including other competencies.

Expert Panel Roundtable Discussion

The expert panel received a report detailing the overall scores for each competency assessed in the first round. A virtual roundtable session moderated by AC from the leadership group was then held. The objectives of this session were to (1) present the results of the first round, (2) review competencies for which agreement had not been reached, and (3) discuss whether newly suggested competencies should be included for evaluation by all members of the expert panel in the second round. Discussions continued until 80% consensus was reached among the participants.

Second Round of Questionnaires

In the second round, the expert panel completed a questionnaire that consisted of (1) providing feedback on the decisions made during the roundtable session, (2) assessing the relevance of newly suggested competencies using the Likert scale, and (3) evaluating the revised wording of the modified competencies.

Final Review

The selected competencies were subsequently circulated to the expert panel for final review and confirmation, resulting in unanimous agreement among all members.

Ethical Considerations

According to Spanish Law 14/2007 on Biomedical Research [23], this study did not require approval from an institutional review board or ethics committee as it did not involve patients, clinical data, or identifiable personal information. The research was based on a consensus process involving expert participants and focused on professional opinions rather than personal or sensitive data. Participation was voluntary, and completion of the questionnaires was considered implied consent. No identifiable information was collected or stored. Therefore, in accordance with applicable regulations and institutional policies, formal ethics approval was not required. No financial compensation was provided for participation.

Results

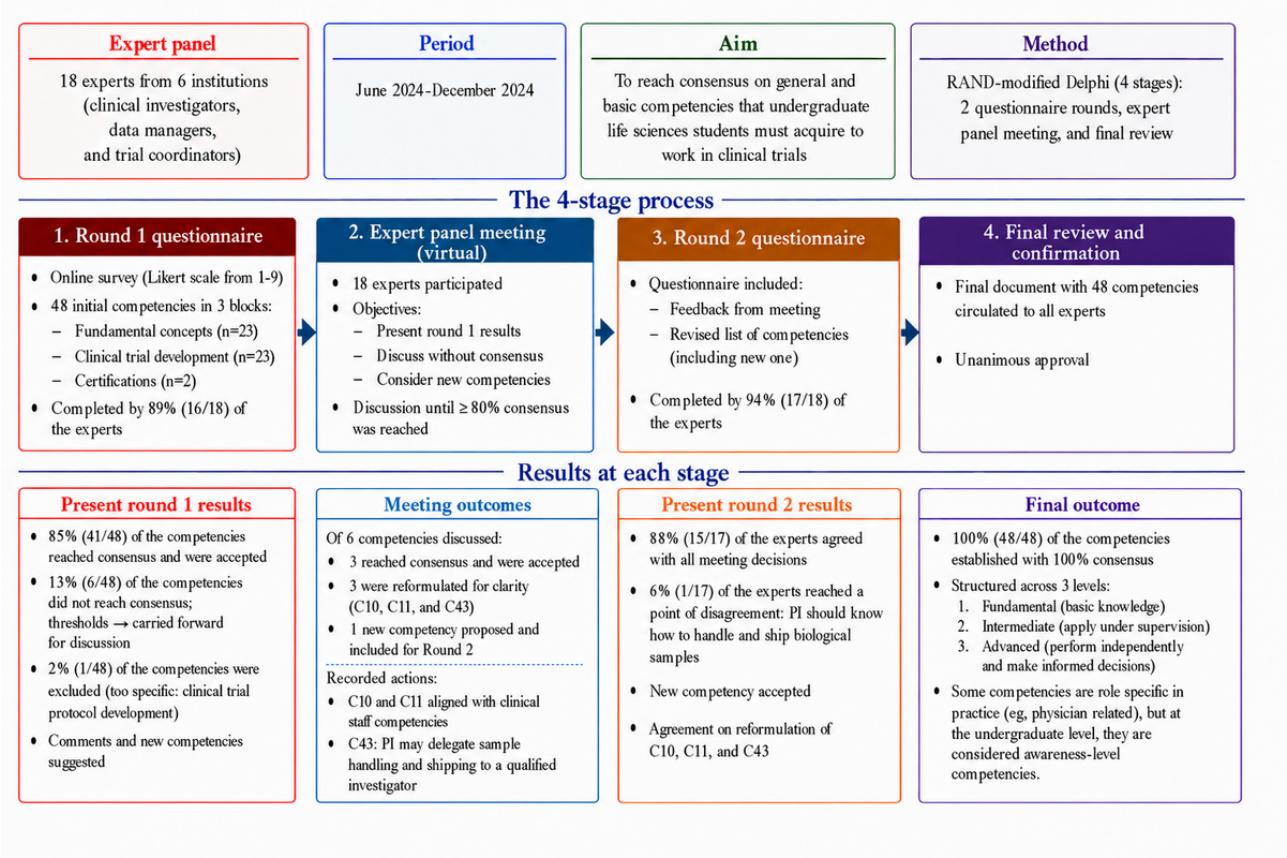
First Round of Questionnaires

The first-round questionnaire was completed by 88.9% (16/18) of the investigators invited to participate. In the first round, consensus was reached on 85.4% (41/48) of the competencies, whereas there was disagreement on a total of 12.5% (6/48) because they did not meet the consensus thresholds in round 1 and were therefore carried forward for discussion during the expert panel meeting. Only 2.1% (1/48) of the competencies, which were related to the development of a clinical trial protocol (section 2, point 1.2 in Multimedia Appendix 1), were excluded because they were considered too specific.

A description of the 18 experts who participated in the consultation is available in Multimedia Appendix 2. Item-level results for all competencies across Delphi rounds are provided in Multimedia Appendix 3. Figure 1 summarizes the results of the 2 questionnaires, the virtual meeting, and the final consensus review.

XSL•FO
RenderX

Figure 1. Overview of the RAND-modified Delphi process and study results. PI: principal investigator.



Expert Panel Roundtable Discussion

All 18 members of the expert panel attended the virtual meeting. Of the 6 competencies discussed, consensus was reached on 3. The remaining 3 competencies (C10, C11, and C43) were reformulated to improve clarity and preserve their original meaning and subsequently included in the second Delphi round. In addition, 1 new competency was proposed during the expert panel discussion and was later accepted after reaching consensus. The reformulation of competencies C10 and C11 consisted of explicitly aligning them with competencies specific to clinical staff. For C43, it was proposed that the PI could delegate the handling and shipping of samples to another qualified investigator, meaning that the PI would not need to perform these tasks personally.

Second Round of Questionnaires

A total of 94.4% (17/18) of the panel members completed the second round of questionnaires. Of these panel members, 88.2% (15/17) stated that they agreed with all the decisions made at the expert panel meeting. The only point of disagreement was that the PI should know how to handle and ship biological samples.

The newly proposed competency was accepted. Agreement was also reached on reformulating the 3 competencies that had been previously modified in the virtual meeting.

Final Review

On the basis of feedback obtained in the second round, a final document outlining 48 competencies was drawn up. The final consensus document (Table 2) was sent to the expert panel and received unanimous approval from all members (18/18, 100%). The extended version of the final competency document is provided in Multimedia Appendix 4.

Some competencies identified in the consensus process are traditionally associated with specific professional roles (eg, physician-related tasks). However, these are interpreted as awareness-level competencies rather than expected performance within the context of undergraduate training. Specifically, we considered that all competencies could be developed across the 3 defined levels (foundational, intermediate, and advanced), although competencies C10, C11, and C12 were considered specific to physicians. The final set of competencies reflects the outcome of a multistage consensus process combining quantitative ratings, expert discussion, and final validation.

The identified competencies can be grouped into higher-order domains reflecting different dimensions of clinical investigation practice, including foundational knowledge, clinical and decision-making skills, research governance, trial management, and transversal competencies such as teamwork and communication. This structure highlights the multidimensional nature of clinical investigation competency by integrating theoretical, practical, and contextual elements into a coherent framework.

XSL•FO
RenderX

Table 2. Consensus on core competencies for life sciences graduates participating in clinical trials.

Section	Competency	Code
Fundamental concepts—clinical research	Principles, scope, types, and phases of clinical research and clinical trials, including legal, ethical, and economic aspects	C1-C9
Clinical method (physicians)	Clinical assessment, prescription of investigational drugs, and evaluation of adverse events	C10-C12
Basic biostatistics	Interpretation of cohorts, variables, and clinical outcomes (mortality, cure, efficacy, and safety), including pharmacokinetics	C13-C19
Teamwork	Identification of team roles, coordinated care, and application of ethical and equality principles	C20-C23
Research design	Formulation of the research question, hypothesis, and study objectives	C24
Trial responsibilities	Understanding sponsor and principal investigator responsibilities under GCP[a]	C25-C26
Ethics and governance	Ethics committee processes and regulatory documentation	C27-C28
Trial management	Clinical trial contract, financial aspects, and protocol structure and implementation	C29-C33
Informed consent	Structure, process, and involved parties	C34-C36
Data and documentation	Source documents and data integrity	C37-C38
Trial procedures	Site initiation, closeout, safety reporting, biological samples, monitoring, and audits	C39-C45
Publication	Compliance with CONSORT[b] reporting standards	C46
Certifications	GCP and IATA[c] certification	C47-C48

[a]GCP: good clinical practice.

[b]CONSORT: Consolidated Standards of Reporting Trials.

[c]IATA: International Air Transport Association.

Discussion

Principal Findings

This study established a consensus-based framework of 48 core competencies for undergraduate life sciences students involved in clinical trial research. Consensus (100% agreement) was achieved among a multidisciplinary panel of experts using a RAND-modified Delphi process. The competencies were organized into three main domains: (1) fundamental concepts, (2) clinical trial development, and (3) certifications. All of them can be interpreted across progressive levels of learning (foundational, intermediate, and advanced). These findings support the view of competencies as integrated and progressively developed capacities rather than isolated skills, thus reinforcing existing competency-based education models.

The competencies also involve higher-order cognitive processes such as critical appraisal and decision-making in complex clinical contexts [12]. In this study, these dimensions are reflected in the competencies identified through the Delphi process, which incorporate both procedural and analytical components.

In addition, clinical research competency involves an epistemic dimension related to how individuals understand, evaluate, and construct knowledge. This dimension is embedded across multiple competencies, including those related to biostatistics (C13-C19) and protocol interpretation (C31-C33), which require the ability to assess the validity and applicability of evidence. Similarly, competencies addressing adverse events (C41 and C42) and clinical decision-making processes (C10-C12) involve judgment under uncertainty. Overall, competency development can be understood as a progression from basic knowledge toward increasingly autonomous and reflective engagement with evidence, consistent with models of epistemic cognition [9].

Comparison With the Literature

It is widely recognized that there is a shortage of professionals trained in human clinical research. Training in this field should begin at the undergraduate level, not only in medicine and nursing but across all life sciences degree programs. Our study addressed this gap by providing a structured, consensus-based competency framework tailored to undergraduate training. Although there is extensive literature describing the competencies required for clinical research professionals and numerous postgraduate programs where advanced competencies can be acquired [24], these frameworks have not been specifically designed for undergraduate education.

Among the most widely recognized competency frameworks [17,24,25] is the one defined by the Joint Task Force for Clinical Trial Competency, which comprises 10 domains and 47 competencies [24]. This framework, validated in 2018, has been widely applied in postgraduate training programs but has not been specifically validated for undergraduate education. More recently, Mendell et al [17] proposed a framework consisting of 59 competencies across 3 levels (fundamental, skilled, and advanced). In our study, we developed a consensus document that defines a set of competencies that all investigators involved in the development of clinical trials should acquire, with the aim of standardizing undergraduate training.

XSL•FO
RenderX

Implications

Although some competencies identified in the consensus process are traditionally associated with specific professional roles (eg, physicians or PIs), these should be interpreted at an introductory level within undergraduate education. Their inclusion reflects the need for early exposure to and understanding of clinical trial processes rather than the expectation of independent performance. Even at an introductory level, participants benefit from a broad understanding of the operational, ethical, and regulatory dimensions of clinical research. From an educational perspective, the competencies identified in this study can be interpreted across different levels of progression (foundational, intermediate, and advanced). This classification is not derived from the Delphi process itself; rather, it is proposed as an interpretative framework to facilitate the integration of these competencies into undergraduate and postgraduate training pathways. More specifically, they aim to promote understanding of the fundamental principles of bioethics and clinical research while providing training in the design, development, and analysis of clinical trial results across all phases.

Importantly, some competencies required reformulation during the consensus process, highlighting the complexity of aligning general research competencies with role-specific clinical responsibilities. This was particularly relevant for competencies initially defined in broader terms, which were subsequently refined to better distinguish between general research competencies and those requiring clinical expertise.

We used a RAND-modified Delphi method [19,20] based on questionnaires. The use of a multistep consensus process combining quantitative thresholds and qualitative discussion is consistent with the established RAND-modified Delphi methodology. The Delphi group included investigators with varying levels of competency. This allowed us to avoid the potential bias that can be introduced when only including investigators with advanced skills due to differences in performance based on their competency level [26]. The competencies are presented in 3 sections. The first section focuses on the acquisition of competencies in general clinical research knowledge. The second section refers to the acquisition of competencies that enable investigators to participate in the design and development of clinical trials. The third section provides the certification required for participation in clinical trials. Of the proposed competencies, only 1 did not obtain the score established for acceptance (section 2, competency 1.2 included in Multimedia Appendix 1). The expert panel concluded that specific and qualified training was required to acquire this competency.

A consistent methodology was used in this study. To ensure that the members of the expert panel fully understood the proposed competencies, all concepts were defined in a consultation document (Table 1). The study objectives, the evaluation system for scoring each competency, and the process leading to the final competency document were explained to all participants. The level of satisfaction (100%) with the final set of competencies was high (100%). This may have been because the members of the expert panel were able to comment on all the proposals via teleconference (face-to-face) or email and because a final round was held to ensure greater consistency across the entire set of competencies.

Any process for acquiring competencies must be evaluated. However, there is no validated system for evaluating these competencies. At advanced levels, competency is often evaluated through scientific output (bibliometric measures), but this approach is not suitable for undergraduate or graduate programs. Experts agree that such evaluations should combine both objective measures and subjective assessment (self-evaluation) of competency levels before and after each acquisition process. Consequently, there is a clear need to develop an evaluation program for the competencies defined in this document.

The competencies outlined in this document could be used by official bodies, such as universities, to establish a standardized training framework [27]. They could also form part of undergraduate programs for students in fields such as medicine or nursing [28] who may be members of clinical trial research teams. However, it should be noted that graduates from other life sciences disciplines also participate in clinical research. The relevance of integrating these competencies into their undergraduate curricula should therefore be carefully evaluated.

Postgraduate programs could also integrate these competencies into their curricula to standardize skill acquisition among research professionals [18]. Additionally, the competencies could serve as a tool for companies to assess the qualifications of professionals involved in clinical trial monitoring [29]. Achieving this, however, would require clearly defining the skills expected of each investigator according to their specific responsibilities within the trial.

The competencies identified in this study extend beyond technical knowledge and reflect key elements of clinical practice in research settings. Competencies such as teamwork, communication, ethical decision-making, and data interpretation are closely linked to core human factors domains, including situational awareness and error prevention. In this context, the proposed framework integrates a human factors perspective and supports not only knowledge acquisition but also the development of nontechnical skills essential for safe and effective participation in clinical research teams.

Limitations

This study has several limitations that should be acknowledged. First, the expert panel comprised 18 participants from a single geographic region in southern Spain (Andalusia), which may limit the generalizability of the findings to other educational or health care contexts. Although the panel included a multidisciplinary group of professionals with diverse roles in clinical research (PIs, study coordinators, and data managers), the consensus reflects a region-specific perspective shaped by local institutional structures, regulatory frameworks, and training pathways. Therefore, caution should be exercised when extrapolating these findings to other national or international settings. Future studies should aim to validate and refine this competency framework in broader, more heterogeneous populations, including experts from different countries, health systems, and academic environments, to enhance its external

XSL•FO
RenderX

validity and global applicability. To minimize potential selection bias inherent in this type of study and improve the relevance of the analysis, the expert panel included clinical investigators with teaching experience as well as a multidisciplinary team of pharmacists, nurses, and biologists engaged in research (Multimedia Appendix 2). In addition, consensus-based approaches such as the Delphi method inherently involve an interpretive component as agreement among experts reflects shared perspectives rather than objective truth. Despite these efforts, the selection process may still have introduced selection bias related to professional background and regional context.

Conclusions

This consensus defines the core competencies that undergraduate life sciences students must acquire for entry-level participation in clinical trial research. It also provides a framework for designing training strategies in both undergraduate and graduate programs of study. The competencies defined in this study should be considered preliminary and will therefore require further validation in broader and more heterogeneous settings before being generalized at the national or international level. Following such validation, this framework may inform the design of undergraduate biomedical curricula aimed at preparing graduates for entry-level participation in clinical research teams.

Acknowledgments

The original draft was written entirely by the authors, and the revisions and final version were reviewed by a native English speaker to ensure grammatical accuracy.

Funding

The final competency document was developed without any financial support.

Data Availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study. Questionnaire responses and consensus data were collected during the Delphi process and are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: AC, PC-P, JJC, DC-M, RB-O, EV-M, JP-C, CA-V, IM, JT-C
Data curation: AC, JT-C
Investigation: AC, PC-P, JJC, DC-M, RB-O, EV-M, JP-C, CA-V, IM, JT-C
Methodology: AC, PC-P, JJC, DC-M, RB-O, EV-M, JP-C, CA-V, IM, JT-C
Validation: AC, PC-P, JJC, DC-M, RB-O, EV-M, JP-C, CA-V, IM, JT-C
Visualization: AC, PC-P, JJC, DC-M, RB-O, EV-M, JP-C, CA-V, IM, JT-C
Writing—original draft: AC, JT-C
Writing—review and editing: AC, PC-P, JJC, DC-M, RB-O, EV-M, JP-C, CA-V, IM, JT-C

Conflicts of Interest

AC receives overarching funding for research from the area of infectious diseases of the Centro de Investigación Biomédica en Red (21/13/00012). JJC received research grants from Pfizer and Menarini and honoraria for training activities from Pfizer, Merck, Shionogi, and Menarini. JT-C has served as scientific advisor for Pfizer, MSD, Shionogi, Advanz, and Menarini; has collaborated in educational activities for Pfizer, MSD, and Shionogi; and has received research grants from MSD and Menarini. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Delphi questionnaire used in round 1.
[DOCX File , 48 KB - humanfactors_v13i1e89449_app1.docx]

Multimedia Appendix 2
Characteristics of the expert panel.
[DOCX File , 16 KB - humanfactors_v13i1e89449_app2.docx]

Multimedia Appendix 3
Item-level results across Delphi rounds.
[DOCX File , 19 KB - humanfactors_v13i1e89449_app3.docx]

XSL•FO
RenderX

Multimedia Appendix 4
Consensus on core competencies for life sciences graduates participating in clinical trials (extended version).
[DOCX File , 19 KB - humanfactors_v13i1e89449_app4.docx]

References

1. Califf RM, Zarin DA, Kramer JM, Sherman RE, Aberle LH, Tasneem A. Characteristics of clinical trials registered in ClinicalTrials.gov, 2007-2010. JAMA 2012 May 02;307(17):1838-1847. [doi: 10.1001/jama.2012.3424] [Medline: 22550198]
2. Kelsey MD, Patrick-Lake B, Abdulai R, Broedl UC, Brown A, Cohn E, et al. Inclusion and diversity in clinical trials: actionable steps to drive lasting change. Contemp Clin Trials 2022 May;116:106740 [FREE Full text] [doi: 10.1016/j.cct.2022.106740] [Medline: 35364292]
3. Kim S. Overview of clinical study designs. Clin Exp Emerg Med 2024 Mar;11(1):33-42 [FREE Full text] [doi: 10.15441/ceem.23.036] [Medline: 37280050]
4. Kandi V, Vadakedath S. Clinical trials and clinical research: a comprehensive review. Cureus 2023 Feb 16;15(2):e35077 [FREE Full text] [doi: 10.7759/cureus.35077] [Medline: 36938261]
5. Agencia Española de Medicamentos y Productos Sanitarios. URL: https://www.aemps.gob.es/ [accessed 2026-04-24]
6. Mehta P, Zimba O, Gasparyan AY, Seiil B, Yessirkepov M. Ethics committees: structure, roles, and issues. J Korean Med Sci 2023 Jun 26;38(25):e198 [FREE Full text] [doi: 10.3346/jkms.2023.38.e198] [Medline: 37365729]
7. Crico C, Sanchini V, Casali PG, Pravettoni G. Evaluating the effectiveness of clinical ethics committees: a systematic review. Med Health Care Philos 2021 Mar;24(1):135-151 [FREE Full text] [doi: 10.1007/s11019-020-09986-9] [Medline: 33219898]
8. Gudiol Munté F, Hernández Quero J, Moreno Guillén S, Pahissa Berga A, de la Torre Cisneros J. Teaching of infectious diseases in medical degree studies [Article in Spanish]. Enferm Infecc Microbiol Clin 2008 Dec;26 Suppl 15:33-39. [doi: 10.1016/s0213-005x(08)76598-4] [Medline: 19195466]
9. Ahmed Rind I, Bhatti J, Chellappan K. From blind trust to critical inquiry: epistemic beliefs and student engagement with ChatGPT in higher education. J Kej 2026;38(1):1-14. [doi: 10.17576/jkukm-2026-38(1)-01]
10. Hofer BK, Pintrich PR. The development of epistemological theories: beliefs about knowledge and knowing and their relation to learning. Rev Educ Res 1997;67(1):88-140. [doi: 10.3102/00346543067001088]
11. Leach DC. Competence is a habit. JAMA 2002 Jan 09;287(2):243-244. [doi: 10.1001/jama.287.2.243] [Medline: 11779269]
12. Rind IA. Conceptualizing the impact of AI on teacher knowledge and expertise: a cognitive load perspective. Educ Sci 2026 Jan 01;16(1):57. [doi: 10.3390/educsci16010057]
13. Torre-Cisneros J, Camacho A, López-Miranda J, Sánchez-Guijo P, Pérez Jiménez F. The teaching process of acquiring clinical skills at the University of Cordoba: a self-assessment by students [Article in Spanish]. Educ Méd 2003;6(2):117-122. [doi: 10.4321/s1575-18132003000200008]
14. Sánchez-Fabra D, Dyar OJ, Del Pozo JL, Amiguet JA, Colmenero JD, Fariñas MD, et al. Perspective of Spanish medical students regarding undergraduate education in infectious diseases, bacterial resistance and antibiotic use [Article in English, Spanish]. Enferm Infecc Microbiol Clin (Engl Ed) 2019 Jan;37(1):25-30 [FREE Full text] [doi: 10.1016/j.eimc.2017.12.003] [Medline: 29429753]
15. Berrevoets MA, Ten Oever J, Oerlemans AJ, Kullberg BJ, Hulscher ME, Schouten JA. Quality indicators for appropriate outpatient parenteral antimicrobial therapy in adults: a systematic review and RAND-modified Delphi procedure. Clin Infect Dis 2020 Mar 03;70(6):1075-1082 [FREE Full text] [doi: 10.1093/cid/ciz362] [Medline: 31056690]
16. Dyar OJ, Beović B, Pulcini C, Tacconelli E, Hulscher M, Cookson B. ESCMID generic competencies in antimicrobial prescribing and stewardship: towards a European consensus. Clin Microbiol Infect 2019 Jan;25(1):13-19 [FREE Full text] [doi: 10.1016/j.cmi.2018.09.022] [Medline: 30414817]
17. Mendell A, Fritter J, Helm S, Capili B, Hildreth L, Johnson K, et al. Team science competencies for clinical research professionals: a multileveled Delphi approach. J Clin Transl Sci 2024 Apr 01;8(1):e112. [doi: 10.1017/cts.2024.509] [Medline: 39655002]
18. Albarqouni L, Hoffmann T, Straus S, Olsen NR, Young T, Ilic D, et al. Core competencies in evidence-based practice for health professionals: consensus statement based on a systematic review and Delphi survey. JAMA Netw Open 2018 Jun 01;1(2):e180281 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.0281] [Medline: 30646073]
19. Dalkey NC. The Delphi method: an experimental study of group opinion. RAND. 1969. URL: https://www.rand.org/pubs/research_memoranda/RM5888.html [accessed 2026-07-31]
20. Fitch K. The Rand/UCLA Appropriateness Method User's Manual. Santa Monica, CA: RAND Corporation; 2001.
21. Teaching guide 2025-26: 145001 [Article in Spanish]. Universidad de Córdoba. 2025. URL: https://www.uco.es/eguiado/guias/2025-26/145001es_2025-26.pdf [accessed 2026-07-29]
22. Likert R. A technique for the measurement of attitudes. Arch Psychol 1932;22(140):1-55 [FREE Full text]
23. Law 14/2007, of 3 July, on Biomedical Research. Instituto de Salud Carlos III. URL: https://www.isciii.es/documents/d/guest/spanishlawonbiomedicalresearchenglish [accessed 2026-07-31]

XSL•FO
RenderX

24. Ianni PA, Samuels EM, Eakin BL, Perorazio TE, Ellingrod VL. Assessments of research competencies for clinical investigators: a systematic review. Eval Health Prof 2021 Sep;44(3):268-278 [FREE Full text] [doi: 10.1177/0163278719896392] [Medline: 31867997]
25. Julé A, Furtado T, Boggs L, van Loggerenberg F, Ewing V, Vahedi M, et al. Developing a globally applicable evidence-informed competency framework to support capacity strengthening in clinical research. BMJ Glob Health 2017 Mar 30;2(2):e000229 [FREE Full text] [doi: 10.1136/bmjgh-2016-000229] [Medline: 28589027]
26. Wang X, Zhao S, Yang H, Miao M, An S, Yao W. Research on core competency elements of clinical investigators. Ther Innov Regul Sci 2025 Jan;59(1):45-53. [doi: 10.1007/s43441-024-00688-5] [Medline: 39612157]
27. Key competences for lifelong learning. European Commission. 2019 Mar. URL: https://education.ec.europa.eu/education-levels/school-education/key-competences-and-basic-skills [accessed 2026-04-24]
28. Dolezel J, Zelenikova R, Finotto S, Mecugni D, Patelarou A, Panczyk M, et al. Core evidence-based practice competencies and learning outcomes for European nurses: consensus statements. Worldviews Evid Based Nurs 2021 Jun;18(3):226-233 [FREE Full text] [doi: 10.1111/wvn.12506] [Medline: 34031973]
29. Nandiwada DR, Kormos W. Interprofessional evidence-based practice competencies: equalizing the playing field. JAMA Netw Open 2018 Jun 01;1(2):e180282 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.0282] [Medline: 30646075]

Abbreviations

PI: principal investigator

Edited by A Kushniruk; submitted 12.Dec.2025; peer-reviewed by S Ismaile, K Wilson-Mitchell, IA Rind; comments to author 31.Mar.2026; revised version received 27.May.2026; accepted 23.Jun.2026; published 02.Nov.2026.

Please cite as:
Cano Á, Cosano-Pérez P, Castón JJ, Corona-Mata D, Ballesteros-Olmedo R, Venegas-Moreno E, Perea-Cortés J, Aranda-Valera C, Machuca I, Torre-Cisneros J
Consensus on Core Competencies of Undergraduate Life Sciences Students for Entry-Level Participation in Clinical Trials: RAND-Modified Delphi Study
JMIR Hum Factors 2026;13:e89449
URL: https://humanfactors.jmir.org/2026/1/e89449
doi:10.2196/89449
PMID:

©Ángela Cano, Patricia Cosano-Pérez, Juan José Castón, Diana Corona-Mata, Rosa Ballesteros-Olmedo, Eva Venegas-Moreno, Jaime Perea-Cortés, Concepcion Aranda-Valera, Isabel Machuca, Julian Torre-Cisneros. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 02.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Change Adaptation Resilience Evaluation Questionnaire for Industry 5.0 Among Workers in High-Tech Manufacturing Companies: Questionnaire Development and Pilot Evaluation Study

Giulia Bassi*, PhD; Angelo Valente*, MSc; Serena Tassoni, MSc; Silvia Salcuni, PhD, Prof Dr

Department of Developmental and Socialization Psychology, University of Padova, Via Venezia 8, Padova, Italy

*these authors contributed equally

Corresponding Author:
Angelo Valente, MSc
Department of Developmental and Socialization Psychology, University of Padova, Via Venezia 8, Padova, Italy

Abstract

Background: Despite increasing recognition of human factors in technologically advanced manufacturing, psychological resilience remains an underexplored dimension. Existing studies often rely on generic instruments that fail to capture the specific adaptive challenges faced by workers in high-tech environments. There is a growing need for context-sensitive tools capable of assessing resilience in line with the human-centric vision of Industry 5.0.

Objective: This study aimed to develop and conduct a preliminary evaluation of the Change Adaptation Resilience Evaluation Questionnaire for Industry 5.0 (CARE-QI 5.0), a multidimensional instrument designed to assess individual and contextual factors contributing to psychological resilience among workers in high-tech manufacturing settings.

Methods: The questionnaire was developed through a 4-phase process: literature review, expert evaluation, focus groups with operators, and pilot testing within 2 manufacturing companies. CARE-QI 5.0 conceptualizes resilience as a higher-order construct comprising 2 second-order dimensions—individual resilience and contextual resilience—further articulated into 18 first-order subscales. Internal consistency and intersubscale correlations were assessed using data from a sample of workers employed in the mechanical and packaging industries in Northeastern Italy.

Results: Findings on the 84 items showed good internal consistency across all subscales and meaningful patterns of intercorrelation. External validity showed that problem-solving self-efficacy, cognitive flexibility, and problem-oriented coping were positively associated with both individual-based resilience and a wide range of contextual resources, including support given by organizations, colleagues, and family, as well as openness to change. In contrast, perseverance in the face of difficulties was linked to cognitive inflexibility, avoidance, and seeking change, suggesting the presence of rigid or less adaptive coping functioning.

Conclusions: CARE-QI 5.0 provides a theoretically grounded and context-sensitive tool for assessing psychological resilience as a multidimensional construct in technologically evolving manufacturing environments. By integrating both individual and contextual protective resources, this instrument captures key adaptive processes relevant to Industry 5.0 scenarios, where human-technology interaction plays a central role. While further psychometric validation is needed, including confirmatory factor analysis and measurement invariance testing, CARE-QI 5.0 shows promise for both theoretical and research use. It can support organizations in monitoring workers' adaptability and well-being, guiding targeted interventions and training strategies that align with workers' resilience profiles and their technological experience.

(JMIR Hum Factors 2026;13:e83754) doi:10.2196/83754

KEYWORDS

resilience; technology; workers; Industry 5.0; manufacturing; instrument; pilot evaluation

Introduction

Background

The manufacturing field constantly faces multiple challenges arising from the elevated production standards imposed by the global market, which increasingly demand high levels of customization and product quality [1]. Within this scenario, the integration of advanced industrial automation and collaborative robotics technologies (eg, cobots) constitutes one of the most significant transformations introduced by Industry 4.0, aimed at enhancing production efficiency through smart and interconnected systems [2]. Despite these technological advancements, manufacturing industries continue to encounter substantial barriers, including difficulties in technological

XSL•FO
RenderX

implementation, as well as insufficient human worker integration into automated systems [3,4]. This limited human-technology integration obstructs the shift toward more adaptive and resilient manufacturing environments, where human-robot interaction (HRI) is expected to play a key role in improving operational flexibility, safety, and decision-making processes [5]. In this context, the human contribution in manufacturing production systems remains a critical factor in shaping production performance and machinery effectiveness [6]. A growing body of literature displays the pivotal mediating role played by Human Capital in technologically advanced firms, facilitating the maximization of both labor output and productivity [7,8]. Concurrently, the effectiveness of human-robot collaboration (HRC) has been strongly linked to increased individual performance, which is largely supported by key human factors [9]. In this context, the development of a new theoretical framework on HRI is considered essential to fostering production systems that are not only more efficient but also more flexible and resilient [10]. Within this evolving paradigm, the role of workers has become increasingly central not only in supporting long-term economic growth through efficient production practices but also in activating resilience strategies that support organizations to recover from potential disruptions [11]. In response to these emerging needs, Industry 5.0 suggests a fundamental shift from a system-centered to a human-centered perspective, keeping the HRI at the core of a collaborative process model [5]. This human-centric vision further emphasizes the importance of human well-being and the preservation of a supportive working environment by promoting the principles of sustainability and resilience [12,13]. Among these principles, resilience stands out as a critical benchmark, defined as the capability to adapt and flexibly cope with vulnerabilities and disruptive changes while maintaining industrial productivity [12]. In this framework, manufacturing workers emerge as the key to such resilience. Through their involvement in HRI processes, they become pivotal agents in supporting both production efficiency and organizational adaptability [14]. To optimize productivity and efficiency within complex industrial environments, several studies converge on the idea that Industry 5.0 can holistically improve workers' well-being and safety by integrating advanced technologies and a human-centered approach [15-17]. This perspective gives rise to the concept of Operator 5.0, a professional figure who goes beyond mere technical skills, by embracing a self-resilience–oriented mindset. Within this framework, *psychological resilience* refers to the individual's ability to emotionally cope with crises and stressors, often supported by advanced technological tools, whereas *cognitive resilience* refers to the operator's ability to sustain mental focus under stressful conditions, thereby reducing the likelihood of error during production processes [18]. Accordingly, Operator 5.0 is not merely a user of technology but also a carrier of psychological resources essential to flourish in hybrid HRC settings, making resilience a key competence in the definition of the Industry 5.0 paradigm.

Recent research examines various dimensions of manufacturing resilience, with particular emphasis on systemic and technological resilience. Nonetheless, despite the increasing recognition of human factors, psychological resilience remains a relatively underexplored area within manufacturing contexts [19]. Existing studies rely on non–sector-specific assessment tools and tend to lack longitudinal or mixed methods designs, which limits their ability to adequately capture the dynamic and evolving nature of resilience in high-tech workplaces [20]. This reveals a crucial gap in the current literature, particularly concerning the potential impact of Industry 5.0 technologies, such as collaborative robots, on workers' psychological resilience. Addressing this gap is pivotal to developing more integrated and context-sensitive assessment frameworks that not only promote human well-being but also enhance workplace adaptability and responsiveness in increasingly manufacturing settings.

Theoretical Background and Current Research Gaps in Resilience at Work

Although various theoretical perspectives on resilience have been proposed, the concept is widely understood as a dynamic and evolving process shaped by individuals' ongoing interactions with their environment [21]. Despite the absence of unanimous consensus upon definition, conceptualization, and measurement frameworks, most assumptions in the literature underline 2 foundational concepts: adversity, typically defined as negative events or circumstances quantitatively associated with adaptive difficulties, and positive outcomes, understood as the capacity to recover or exceed prior levels of psychological functioning following adversity [22]. The American Psychological Association [23] defines resilience as "the ability to adapt in the face of trauma, adversity, tragedy or even significant ongoing stressors." Among the personal resources that support resilient functioning in complex and evolving work environments, 2 constructs have emerged as particularly relevant: change adaptation, the perceived ability to effectively respond to and integrate workplace changes [24], and cognitive flexibility, the capacity to adjust mental strategies and reframe stressful events [25,26]. Both have been linked to improved emotional regulation and adaptive behavior under pressure [26] and can be considered essential components of resilience in technological contexts. Within this framework, protective factors play a fundamental role in promoting resilience by increasing the likelihood of a positive adaptation following exposure to stressors [27]. According to Windle [28], protective factors are pivotal in defining resilience and are typically classified into 2 categories. Internal positive factors refer to intraindividual characteristics that foster resilience and are specific to each individual [29]. These include several traits, such as positive self-perception, self-regulation, cognitive abilities, adaptive coping strategies, and stress inoculation [27,30]. In contrast, external positive factors encompass supportive social and environmental conditions derived from an individual's immediate surroundings [29], such as strong social support (eg, positive relationships from family and friends) and community cohesion [27,30]. These resources affect how individuals perceive and deal with challenges, especially in work-related settings [31].

Increasingly, the literature suggests that resilience cannot be fully understood through the analysis of individual or contextual factors apart, but rather through their dynamic interaction. This is especially relevant in complex and high-pressure environments, such as advanced manufacturing. In such

XSL•FO
RenderX

contexts, resilience emerges as a multilayered construct, where personal cognitive-emotional resources and organizational or social support systems are deeply intertwined. As highlighted in recent research [20], considering both dimensions simultaneously allows for a more accurate understanding of how workers adapt and thrive amidst organizational change, technological disruption, and operational stress.

In line with this perspective, in this study, psychological resilience is conceptualized as a multidimensional and context-sensitive capacity that enables workers to cope with stress, adapt to changes related to technological and organizational demands, and maintain effective functioning in complex and evolving manufacturing environments [20]. More specifically, resilience is framed as a higher-order construct encompassing 2 interrelated domains: individual resilience (eg, self-efficacy, cognitive flexibility, and change adaptation) and contextual resilience (eg, social support and organizational resilience). These dimensions jointly influence workers' capacity to adapt to technological and organizational challenges in advanced manufacturing settings, forming an integrated conceptual model that underpins the development of the Change Adaptation Resilience Evaluation Questionnaire for Industry 5.0 (CARE-QI 5.0) [20]. These dimensions were selected as theoretically grounded components of resilience rather than independent constructs, in line with evidence highlighting the role of both individual resources and contextual protective factors in influencing adaptive functioning under stress [21,27,28]. Given these distinctions, work-related resilience can be broadly considered as the capacity to cope with and overcome adversity in occupational settings, emerging stronger and more competent as a result [32]. While existing research has addressed both individual and environmental dimensions of resilience, there is currently no unified theoretical model specifically tailored to industrial work environments. Despite the growing interest in resilience within high-tech manufacturing settings, the literature lacks psychometrically validated tools designed to assess resilience in such contexts [20]. To date, only Babamiri [33] has developed a tool related to HRI. However, it comprises only 3 items and fails to capture resilience as a multidimensional construct. In light of these limitations, this study introduces a new psychometric instrument: CARE-QI 5.0, specifically developed to assess and conceptualize psychological resilience in advanced manufacturing environments aligned with the principles of Industry 5.0.

Aims

The overarching aim of this study was to develop and conduct an initial evaluation of the CARE-QI 5.0, a new multidimensional self-report instrument designed to assess psychological resilience among industrial workers in high-tech manufacturing environments. In line with the human-centric principles of Industry 5.0, this study pursued the following objectives:

- To conceptualize and operationalize psychological resilience in technologically advanced manufacturing contexts by integrating personal and contextual protective factors derived from existing theoretical and empirical frameworks.
- To develop and refine a context-sensitive set of items through a mixed methods approach that included expert review and focus group feedback, thereby enhancing the semantic clarity and contextual relevance of the instrument.
- To pilot the revised questionnaire in real-world high-tech industrial settings, evaluating its preliminary psychometric properties, including internal consistency, interitem correlations, and subscale coherence.

Methods

Participant Recruitment

Participants were recruited from 2 distinct manufacturing companies located in Northeastern Italy, both operating in technologically advanced production sectors. One company specializes in the design and production of high-precision mechanical components primarily for the pharmaceutical and packaging industries, while the other is a leading provider of flexible packaging solutions, focusing on innovation, sustainability, and advanced printing technologies for food and consumer goods.

For the focus group phase, inclusion criteria were established to ensure representativeness and alignment with the study's aims. Specifically, participants were required to (1) be operators in high-tech companies, (2) ensure a balanced distribution by gender and age, and (3) be fluent in Italian to guarantee full comprehension and effective participation in the qualitative discussions.

For the pilot testing phase, a separate group of workers was recruited. Inclusion criteria for this phase included being operators employed in high-tech manufacturing companies and being fluent in Italian.

It is important to note that the participants involved in the focus groups differed from those who later took part in the pilot testing phase. This methodological separation was implemented to ensure that insights gathered during the qualitative development of the questionnaire did not bias its subsequent evaluation during pilot administration [34].

Ethical Considerations

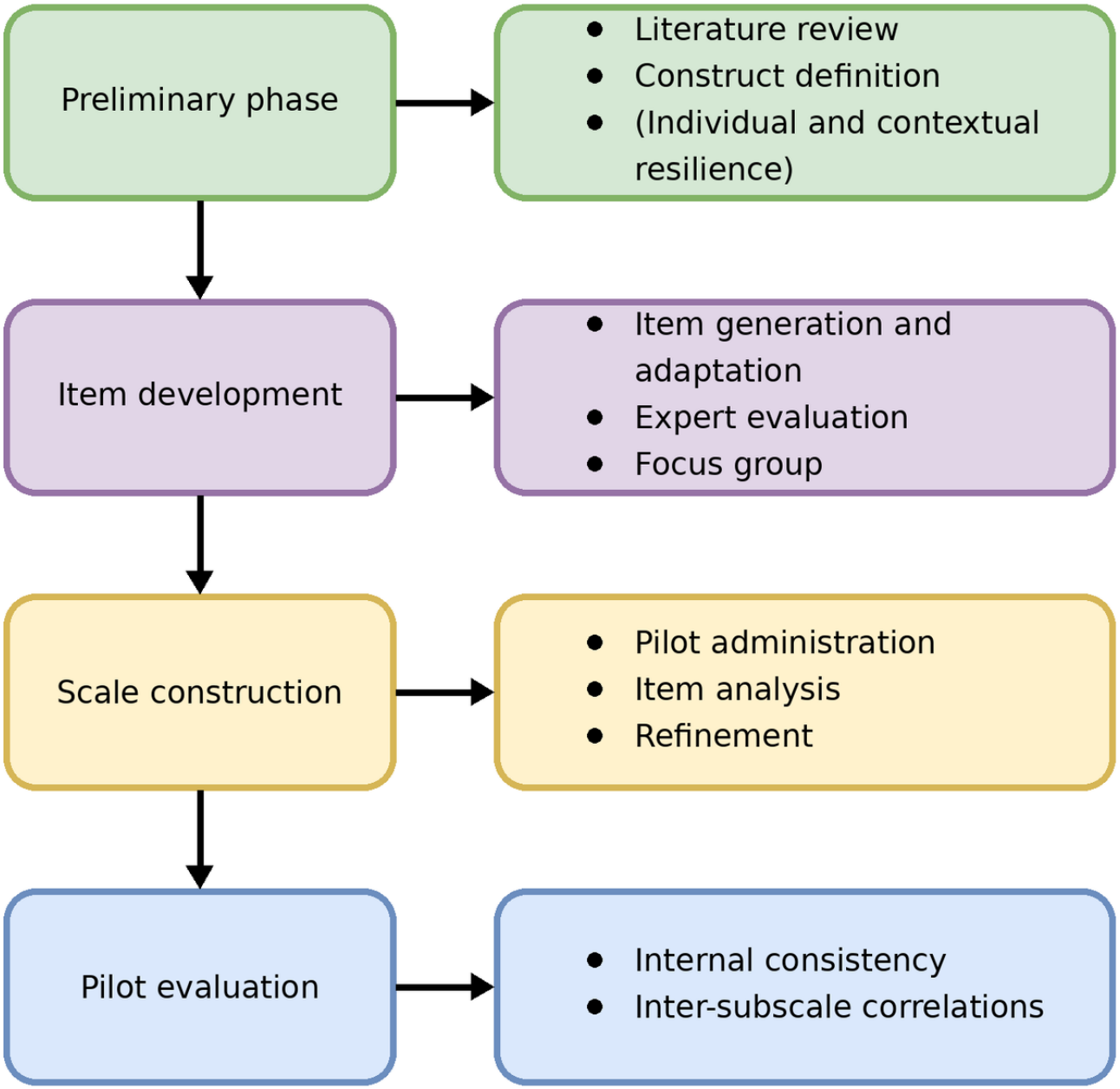
All participants were recruited voluntarily through internal company communications (eg, human resources announcements and mailing lists) as well as via digital and paper flyers specifically created for the purposes of this study. Informed consent was obtained from all participants electronically at the beginning of the digital questionnaire, prior to their involvement. No compensation was provided to participants. Data were collected and stored in compliance with applicable privacy regulations, ensuring the confidentiality and anonymity of participants' data. Data collection procedures complied with the ethical standards outlined in the Declaration of Helsinki [35] and received 2 separate approvals from the Ethics Committee of the University of Padova (protocol 871-c, 2025 for the focus group sessions and 991-b, 2025 for the pilot administration).

XSL•FO
RenderX

Design

An exploratory sequential mixed methods approach was used for the development of the CARE-QI 5.0 questionnaire, following established best-practice guidelines and scale development and validation frameworks [34,36-38]. This comprised a structured, multiphase process encompassing construct definition, item generation, expert evaluation, target population feedback, and pilot psychometric evaluation. The item development process followed an iterative refinement logic, in which they were not strictly retained or excluded based on a single quantitative criterion but were progressively revised, reworded, or replaced across phases, integrating expert judgment and feedback from the target population [34,36-38]. In line with these guidelines, the overall procedure is summarized in Figure 1, comprising 4 key stages.

Figure 1. Development process of the CARE-QI 5.0 following best-practice guidelines for scale development.



First, we conducted a comprehensive literature review across academic databases to identify core constructs and relevant existing instruments. Drawing on these results, we generated and contextually adapted questionnaire items from validated sources. Next, the initial version of the instrument underwent expert review to assess content validity. Third, following item refinement, 2 separate focus groups, lasting 60 minutes each, were conducted with workers from 2 different manufacturing companies to obtain qualitative feedback on item clarity and comprehensibility. Finally, pilot administration of the revised digital questionnaire was conducted as part of a case study, allowing for initial preliminary evaluation aimed at exploring the psychometric properties of the instrument in a real-world industrial context. The digital survey also collected information on participants' sociodemographic, occupational, and technology-related characteristics. Specifically, data were gathered on sex, age, job tenure, job role, work tasks, type of contract, technology readiness level (basic, intermediate, and high), frequency of technology use from 1 ("never") to 5 ("always"), years of experience with such technologies, and

XSL•FO
RenderX

whether they had received any related training. The duration of the compilation was approximately 25 minutes. The 4 major phases are specifically described step by step.

Literature Review for Instrument Development

A deductive strategy for the selection of main constructs was applied, following established methodological guidelines [38]. This process involved a comprehensive review of existing theoretical models and empirical literature through Scopus, Web of Science, PubMed, and Google Scholar. This was followed by a series of iterative brainstorming sessions among the authors to identify the most relevant dimensions of psychological resilience. The selection of main constructs for the questionnaire aimed to provide a comprehensive and representative theoretical structure of worker resilience within advanced manufacturing contexts, in line with Industry 5.0 principles. Items were selected and adapted from validated instruments to ensure theoretical coherence and psychometric robustness, rather than being developed ex novo [34,36,38]. The selection of constructs and items followed a theory-driven approach, conceptualizing them as interrelated components of resilience rather than independent measures. In this framework, resilience was defined as a multidimensional construct influenced by the interaction between individual resources and contextual factors in technologically evolving work environments [20]. Accordingly, each dimension of the questionnaire was derived from established psychometric tools commonly used in psychological research [34-39]. Preference was given to tools that have been validated on Italian populations or adapted to the Italian sociocultural context by reporting good or excellent internal consistency values (Cronbach $\alpha \geq 0.70$) in their validated Italian versions. Items were viewed and selected through a comparative process among the authors of this study. During this phase, items that were most representative of each specific subdimensional construct within the broader model were selected. For each construct, priority was given to items that demonstrated both theoretical clarity and contextual relevance. Subsequently, the authors proposed a context-specific readaptation of the selected items, in line with high-tech manufacturing environments.

Experts' Evaluation: Item Clarity Assessment

The first version of the CARE-QI 5.0 questionnaire was assessed by 3 experts in work and organizational psychology as well as HRI, who independently gave feedback regarding the clarity and comprehensibility of each item, without being informed of the questionnaire's underlying factor structure. The assessors rated each item dichotomously (yes or no), reporting if it was clear or unclear by further providing qualitative comments to suggest any changes. Experts also provided open-ended comments to improve the semantic alignment of the items with the target population.

Two Focus Groups

Two focus group sessions were conducted within 2 manufacturing companies known for their high-tech machinery usage, by implementing the guidelines described by Freeman [39]. This inductive approach, which involves collecting opinions from the target population to corroborate item clarity and their contextual relevance, is recognized as a pivotal step for the questionnaire development and validation process [40,41]. The facilitator (GB) read each item aloud while the text was projected on a monitor, and 2 observers (AV and ST) took notes of the participants' nonverbal expression. The questionnaire was implemented on the *Qualtrics* platform (Qualtrics International Inc). Participants were asked to indicate whether the item was clear, unclear, or ambiguous by responding "yes" or "no" accordingly. This phase was inspired by cognitive interviewing [42], integrating a hetero-directed evaluation. The evaluation primarily focused on participants' judgments of item clarity, but it was also enriched by their spontaneous comments. Although no formal probing scripts or structured think-aloud protocols were used, participants frequently shared their interpretations, suggested alternative wordings, and pointed out ambiguities. These qualitative insights offered valuable information about comprehension processes. All observations were audio recorded and transcribed verbatim to identify potential sources of misunderstanding and to refine the wording of the questionnaire items.

The Pilot Administration

Following the item pool refinement, the CARE-QI 5.0 was administered through a pilot session to the same abovementioned recruited companies. The main purpose of this validation step concerned the examination of the internal validity of the questionnaire, which was filled out using the *Qualtrics* platform. Data were collected over a 2-week period. All items were scored on a 5-point Likert scale, ranging from 1 ("totally disagree") to 5 ("totally agree").

Data Analysis

Statistical analyses were progressively conducted across each phase of the refinement process. As such, each version of the questionnaire integrated adjustments derived from the previous step, with the goal of enhancing item clarity and contextual relevance.

For the experts' evaluation phase, item clarity and comprehensibility were assessed using an agreement index based on dichotomous judgments (clear or unclear) provided by 3 experts. For each item, we calculated the proportion of experts who rated it as clear. Items receiving full agreement (3/3 experts; 100% agreement) were retained without modification, whereas items with lower agreement were revised based on the experts' qualitative feedback. Expert agreement was used to identify items that could be retained without modification, whereas items not reaching full agreement were further refined rather than discarded, in line with the iterative process [34,36,38].

For the focus groups, both quantitative and qualitative analyses were performed. Frequencies and percentages were computed for each item to identify potential comprehension issues, and content analysis was applied to summarize key themes from participants' feedback.

In the pilot phase, all statistical analyses were performed using RStudio (Posit, PBC) [43]. Descriptive statistics were computed for each item and each subscale of the CARE-QI 5.0 questionnaire.

XSL•FO
RenderX

Internal consistency was assessed using Cronbach α and McDonald ω total coefficients, calculated separately for each subscale. For each α estimate, the corresponding 95% CI (lower and upper bounds) was also computed to evaluate the stability and precision of internal consistency estimates.

Pearson correlation coefficients (*r*) were calculated to assess interitem correlations within each subscale to evaluate internal coherence and identify potentially misfitting items. In addition, item-total correlations were examined to verify the contribution of each item to its respective subscale.

Furthermore, intersubscale Pearson correlations (*r*) were computed to explore the relationships among the identified dimensions of the questionnaire. This step aimed to preliminarily assess the internal structure and construct validity of the tool, consistent with the hypothesized hierarchical model of CARE-QI 5.0. Moreover, the subscales were correlated with age, job tenure, technology readiness level (basic, intermediate, and high), frequency of technology use, and years of experience with such technologies. The strength of interfactor relationships was interpreted based on Cohen's [44] guidelines, where *r* values of 0.10, 0.30, and 0.50 are considered small, medium, and large effects, respectively. In addition, intersubscale correlations exceeding $r>0.85$ were interpreted as potential indicators of construct overlap, suggesting that the associated dimensions may be measuring substantially similar or redundant constructs [45].

Results

Results of the 4 Phases

Across the different phases of the study, the CARE-QI 5.0 was developed by integrating and adapting items derived from multiple validated questionnaires, which were progressively refined throughout the process. The results of each development phase are summarized in Tables 1-7. Specifically, Table 1 reports the initial item pool derived from the literature review and existing instruments, Tables 2 and 3 present the results of the expert evaluation and subsequent refinement of the selected items, Tables 4 and 5 describe the characteristics of the pilot sample and the technologies used, Table 6 reports the reliability analyses, and Table 7 presents the final set of questionnaires and subscales included in the CARE-QI 5.0 for the pilot phase. Moreover, item selection and refinement followed an iterative process across the 4 study phases. As illustrated in Figure 2, the initial pool of 82 items was first evaluated by experts, resulting in a subset retained without modification (n=32), and the remaining items were revised based on qualitative feedback. Subsequently, theoretically driven modifications were introduced, including the replacement of the Emotional Style Questionnaire (ESQ; 12 items) with the Coping Orientation to the Problems Experienced (COPE-NVI; 21 items), leading to an expanded pool of 91 items. This version was then evaluated through focus groups, which informed further wording refinements and contextual adaptations. The final version administered in the pilot phase comprised 91 items.

XSL•FO
RenderX

Table . The first set of questionnaires used in the development of Change Adaptation Resilience Evaluation Questionnaire for Industry 5.0.

Construct	Tool	Cronbach α	Subscales selected
Self-efficacy	Occupational Self-Efficacy Scale (OSES) [46]	• Overall α=0.69 - 0.77	Problem solving (OSES-PS); and confidence to persevere in the face of difficulties (OSES-CD)
Emotion regulation strategies	Emotional Style Questionnaire (ESQ) [47] (excluded after expert evaluation)	• Resilience: α=0.70; self-awareness: α=0.62; and sensitivity to context: α=0.71	Resilience (ESQ-RE); self-awareness (ESQ-SA); and sensitivity to context (ESQ-SC)
Locus of control	Locus of Control (LOC-L) [48]	• Total scale: α=0.81	Internal locus of control (LOC-INT); and external locus of control (LOC-EXT)
Change management and adaptation	Acceptance of Change Scale (ACS) [49]	• Predisposition to change: α=0.83; change seeking: α=0.80; and positive reaction to change: α=0.75	Predisposition to change (ACS-PTC); seeking to change (ACS-STC); and positive reaction to change (ACS-PRTC)
Cognitive flexibility	Multidimensional Psychological Flexibility Inventory (MPFI) [50]	• Cognitive flexibility: acceptance: α=0.85; and committed action: α=0.88. • Cognitive inflexibility: fusion: α=0.93; inaction: α=0.94	Cognitive flexibility (acceptance and committed action; MPFI-CF); and cognitive inflexibility (fusion and inaction; MPFI-CINF)
Social support	Multidimensional Scale of Perceived Social Support (MSPSS) [51]; Health and Safety Executive Stress Indicator Tool (SIT) [52]; and Survey of Perceived Organizational Support (SPOS) [53]	• MSPSS: family: α=0.92; and friends: α=0.96. • SIT: supervisor: α=0.85; and colleagues: α=0.86. • SPOS: organization: α=0.81	Friends (SS-Friends); family (SS-Family); colleagues (SS-Coll); supervisors/managers (SS-S/M); and organization (SS-ORG)
Organizational resilience	Self-Evaluation of Resilience (SEOR) [54]	• Skill-based resilience: α=0.90; and individual-based resilience: α=0.81	Skill-based resilience (SEOR-SR); and individual-based resilience (SEOR-IR)

Table . Proportion of expert agreement on item clarity.

Questionnaires	Total items, n	Items with full agreement (3/3 experts), n	Percentage with full agreement (%)
Total	82	32	39
OSES[a]	10	6	60
ESQ[b]	12	1	8
LOC-L[c]	6	3	50
ACS[d]	12	3	25
SS[e]	13	11	85
MPFI[f]	20	5	25
SEOR[g]	9	3	33

[a]OSES: Occupational Self-Efficacy Scale.

[b]ESQ: Emotional Style Questionnaire.

[c]LOC-L: Locus of Control.

[d]ACS: Acceptance of Change Scale.

[e]SS: Social Support.

[f]MPFI: Cognitive Flexibility.

[g]SEOR: Self-Evaluation of Resilience.

XSL•FO
RenderX

Table . The second set of questionnaires used in the development of Change Adaptation Resilience Evaluation Questionnaire for Industry 5.0.

Construct	Tool	Cronbach α	Subscales selected
Self-efficacy	Occupational Self-Efficacy Scale (OSES) [46]	• Overall α=0.69 - 0.77	Problem solving (OSES-PS); and confidence to persevere in the face of difficulties (OSES-CD)
Coping strategies	Coping Orientation to Problems Experienced–Italian Version (COPE-NVI) [55], included after expert evaluation	• Avoidance strategies: α=0.68; positive attitude: α=0.74; problem-oriented coping: α=0.77; and seeking social support: α=0.81	Avoidance strategies (COPE-AV); positive attitude (COPE-PA); problem-oriented coping (COPE-PO); and seeking social support (COPE-SSS)
Locus of control	Locus of Control (LOC-L) [48]	• Total scale: α=0.81	Internal locus of control (LOC-INT); and external locus of control (LOC-EXT)
Change management and adaptation	Acceptance of Change Scale (ACS) [49]	• Predisposition to change: α=0.83; seeking to change: α=0.80; and positive reaction to change: α=0.75	Predisposition to change (ACS-PTC); seeking to change (ACS-STC); and positive reaction to change (ACS-PRTC)
Cognitive flexibility	Multidimensional Psychological Flexibility Inventory (MPFI) [50]	• Cognitive flexibility: acceptance: α=0.85; and committed action: α=0.88. • Cognitive inflexibility: fusion: α=0.93; and inaction: α=0.94	Cognitive flexibility (acceptance and committed action; MPFI-CF); and cognitive inflexibility (fusion and inaction; MPFI-CINF)
Social support	Multidimensional Scale of Perceived Social Support (MSPSS) [51]; Health and Safety Executive Stress Indicator Tool (SIT) [52]; and Survey of Perceived Organizational Support (SPOS) [53]	• MSPSS: family: α=0.92; friends: α=0.96. • SIT: supervisor: α=0.85; colleagues: α=0.86. • SPOS: organization: α=0.81	Friends (SS-Friends); family (SS-Family); colleagues (SS-Coll); supervisors/managers (SS-S/M); and organization (SS-ORG)
Organizational resilience	Self-Evaluation of Resilience (SEOR) [54]	• Skill-based resilience: α=0.90; and individual-based resilience: α=0.81	Skill-based resilience (SEOR-SR); and individual-based resilience (SEOR-IR)

XSL•FO
RenderX

Table . Sociodemographic, occupational-related, and technology-related characteristics (N=44).

Variable	Values
Sex, n (%)	
Male	29 (65.9)
Nationality, n (%)	
Italian	41 (93.2)
Age (y), mean (SD)	45.12 (11.74)
Job tenure (y), mean (SD)	18.42 (8.55)
Job role, n (%)	
Operators	40(90.9)
Technician	4 (9.1)
Task, n (%)	
Assembly	17 (38.6)
Picking	2 (4.5)
Welding	2 (4.5)
Product quality control	2 (4.5)
Type of contract, n (%)	
Open ended	33 (75)
Fixed term	4 (9.1)
Interim	4 (9.1)
Other (ie, employee leasing and work under an independent contractor agreement)	2 (4.5)
Part time	1 (2.3)
Technology readiness level, n (%)	
Base	12 (27.3)
Intermediate	28 (63.6)
High	4 (9.1)
Frequency of technology use for working, mean (SD)	3.93 (1.17)
Years of experience with technology for working, mean (SD)	15.63 (8.91)
Training course for using technologies for working (yes), n (%)	33 (75)

XSL•FO
RenderX

Table . Technologies used for working (in decreasing order).

Technologies	Description	Values, n (%)
Robotic assembly lines	Fully automated robotized assembly lines	16 (36.36)
Measurement instruments	Calipers, micrometers, etc	16 (36.36)
Pillar drills, drilling, and tapping machines	Includes pillar drills, tapping, and other similar tools	12 (27.27)
Semi-manual machining	Electric or manual hybrid workstations	12 (27.27)
Loading and unloading robots	Robots for handling between stations	11 (25.00)
Assembly tables	Fixed or smart tables for manual or semiautomatic assembly	9 (20.45)
Automated handling systems	Conveyors, shuttles, and trolleys	9 (20.45)
Testing machines	For functional or quality testing	9 (20.45)
Lathe	These lathes use a computer to control the machining of the workpiece, ensuring precision and repeatability	8 (18.18)
Screwing, gluing, and welding machines	Automated joining via screws, glue, or welding	8 (18.18)
Digital systems for quality	Classification, defect detection, etc	7 (15.91)
Industrial computers	Computers embedded in manufacturing lines	7 (15.91)
Tracking or traceability software	Batch or product traceability tools	7 (15.91)
Real-time monitoring systems	Real-time performance or quality monitoring	6 (13.64)
Milling machine	They remove material from a workpiece using software to guide the movement of the milling cutter	5 (11.36)
Automatic saws	Programmable saws for cutting profiles and bars	5 (11.36)
Bending machines	Controlled bending equipment (CNC[a])	5 (11.36)
Other	Unspecified or miscellaneous tools	5 (11.36)
Hydraulic and electronic presses	Hydraulic presses for forming or punching	4 (9.09)
Pick-and-place systems	Robotic systems for part handling	4 (9.09)
Cobot (collaborative robots)	Collaborative robots working with humans	3 (6.82)
Gantry, scara, and delta robots	Industrial robots with special architectures	3 (6.82)
Automatic punching machines	Machines that perform punching operations on sheet metal, controlled by software to position and execute punches	2 (4.55)
Waterjet cutting	Plasma-based material cutting	2 (4.55)
Injection molding	Plastic forming by high-pressure injection	2 (4.55)
Programmable logic controllers	Programmable logic controllers for process automation	2 (4.55)
Manufacturing execution system	Manufacturing execution systems	2 (4.55)
Laser scanners and 3D optical measurement systems	Optical scanners for dimension control	2 (4.55)
Laser cutting	Software-guided laser material cutting	1 (2.27)
Automated guided vehicles and autonomous mobile robots	Autonomous material-handling vehicles	1 (2.27)
Computer-aided design and computer-aided manufacturing	Computer-aided design and manufacturing tools	1 (2.27)

[a]CNC: Computer Numerical Control.

XSL•FO
RenderX

Table . Cronbach α and McDonald ω_total for each factor.

Factors	α	95% CI lower	95% CI upper	ω_total	95% CI lower	95% CI upper
OSES-PS[a]	0.79	0.63	0.87	0.71	0.55	0.83
OSES-CD[b]	0.84	0.75	0.90	0.88	0.79	0.93
COPE-AV[c]	0.87	0.70	0.94	0.88	0.80	0.93
COPE-PA[d]	0.63	0.35	0.78	0.69	0.50	0.82
COPE-PO[e]	0.79	0.68	0.87	0.82	0.72	0.89
COPE-SSS[f]	0.78	0.59	0.88	0.74	0.60	0.84
LOCL-INT[g]	*-0.04*	*-0.85*	*0.42*	*0.23*	*0.00*	*0.52*
LOCL-EXT[g]	*0.60*	*0.29*	*0.76*	*0.71*	*0.53*	*0.84*
ACS-PTC[h]	0.83	0.66	0.92	0.83	0.73	0.90
ACS-STC[i]	0.65	0.33	0.81	0.67	0.48	0.81
ACS-PRTC[j]	0.67	0.25	0.84	0.68	0.49	0.82
SS-Friends[k]	0.79	0.48	0.93	0.79	0.60	0.90
SS-Family[l]	0.82	0.63	0.92	0.83	0.68	0.91
SS-S/M[m]	0.89	0.81	0.94	0.90	0.84	0.94
SS-Coll[n]	0.85	0.71	0.92	0.86	0.76	0.92
SS-ORG[o]	0.93	0.88	0.96	0.94	0.89	0.96
MPFI-CF[p]	0.76	0.58	0.86	*0.55*	0.32	0.74
MPFI-CINF[q]	0.92	0.85	0.95	0.92	0.87	0.95
SEOR-SR[r]	0.94	0.89	0.97	0.94	0.90	0.96
SEOR-IR[s]	0.79	0.61	0.88	0.80	0.66	0.89
Total score	0.95	0.91	0.97	0.90	0.82	0.95
Individual resilience	0.93	0.79	0.97	0.84	0.74	0.91
Contextual resilience	0.96	0.94	0.98	0.96	0.93	0.98

[a]OSES-PS: Occupational Self-Efficacy Scale-Problem Solving.

[b]OSES-CD: Occupational Self-Efficacy Scale-Confidence to persevere in the face of difficulties.

[c]COPE-AV: Coping Orientation to Problems Experienced-Avoidance Strategies.

[d]COPE-PA: Coping Orientation to Problems Experienced-Positive Strategies.

[e]COPE-PO: Coping Orientation to Problems Experienced-Problem Oriented.

[f]COPE-SSS: Coping Orientation to Problems Experienced-Seeking Social Support.

[g]Excluded.

[h]ACS-PTC: Acceptance of Change Scale-predisposition to change.

[i]ACS-STC: Acceptance of Change Scale Seeking to Change.

[j]ACS-PRTC: Acceptance of Change Scale Positive Reaction to Change.

[k]SS-Friends: Social Support-Friends.

[l]SS-Family: Social Support-Family.

[m]SS-S/M: Social Support-Supervisors/Managers.

[n]SS-Coll: Social Support-Colleagues.

[o]SS-ORG: Social Support-Organization.

[p]MPFI-CF: Multidimensional Psychological Flexibility Inventory cognitive flexibility.

[q]MPFI-CINF: Multidimensional Psychological Flexibility Inventory cognitive inflexibility.

XSL•FO
RenderX

[r]SEOR-SR: Self-Evaluation of Resilience-Skill-based resilience.

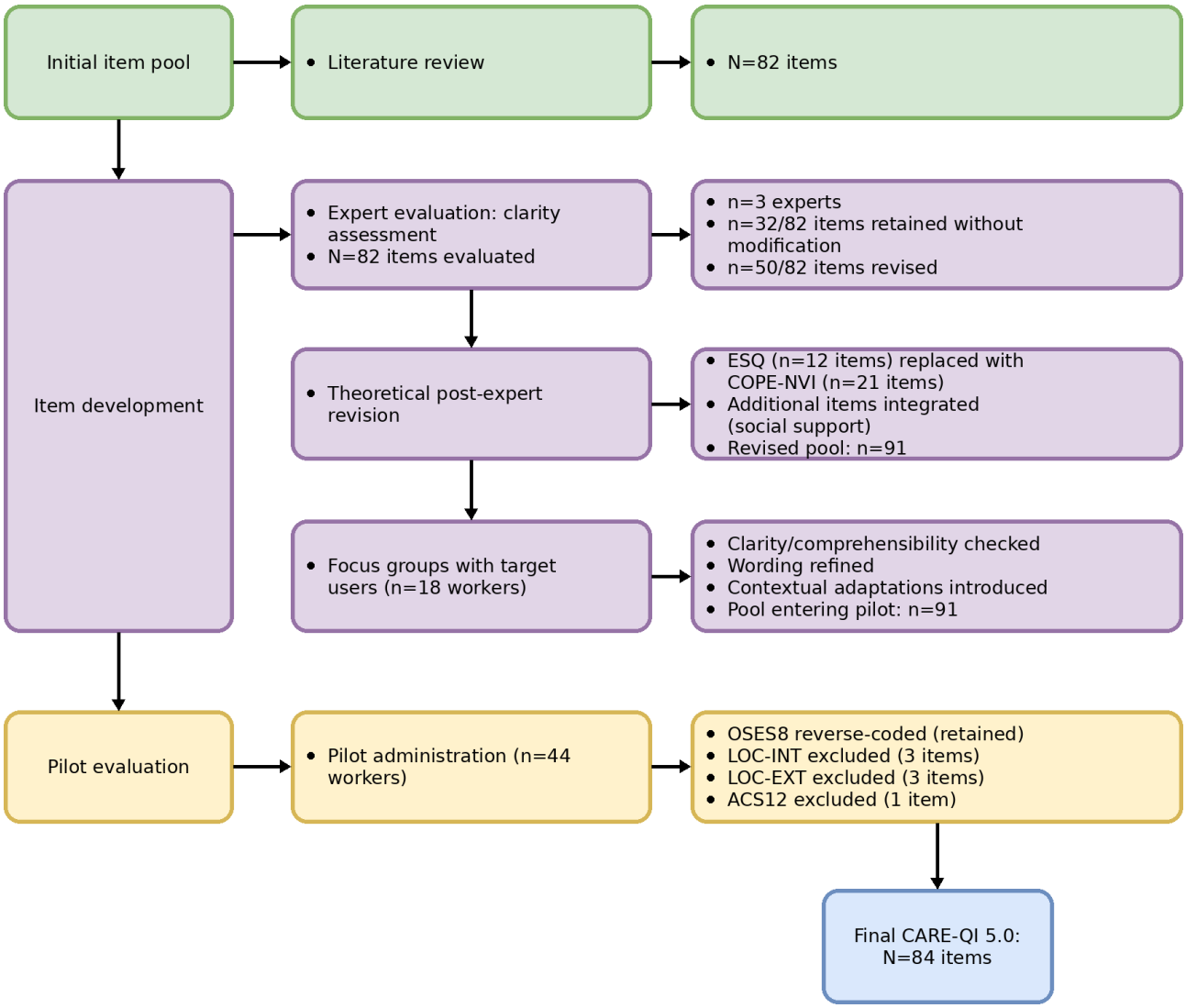
[s]SEOR-IR: Self-Evaluation of Resilience-Individual-based resilience.

Table . The third and final set of questionnaires used in the development of CARE-QI 5.0.

Construct	Tool	Cronbach α	Subscales selected
Self-efficacy	Occupational Self-Efficacy Scale (OSES) [46]	• Overall α=0.69 - 0.77	Problem solving (OSES-PS); and confidence to persevere in the face of difficulties (OSES-CD)
Coping strategies	Coping Orientation to Problems Experienced–Italian Version (COPE-NVI) [55], included after expert evaluation	• Avoidance strategies: α=0.68; positive attitude: α=0.74; problem oriented: α=0.77; and seeking social support: α=0.81	Avoidance strategies (COPE-AV); positive attitude (COPE-PA); problem oriented (COPE-PO); and seeking social support (COPE-SSS)
Change management and adaptation	Acceptance of Change Scale (ACS) [49]	• Predisposition to change: α=0.83; seeking to change: α=0.80; and positive reaction to change: α=0.75	Predisposition to change (ACS-PTC); seeking to change (ACS-STC); and positive reaction to change (ACS-PRTC)
Cognitive flexibility	Multidimensional Psychological Flexibility Inventory (MPFI) [50]	• Cognitive flexibility: acceptance: α=0.85; and committed action: α=0.88. • Cognitive inflexibility: fusion: α=0.93; and inaction: α=0.94	Cognitive flexibility (acceptance and committed action; MPFI-CF); and cognitive inflexibility (fusion and inaction; MPFI-CINF)
Social support	Multidimensional Scale of Perceived Social Support (MSPSS) [51]; Health and Safety Executive Stress Indicator Tool (SIT) [52]; and Survey of Perceived Organizational Support (SPOS) [53]	• MSPSS: family: α=0.92; and friends: α=0.96. • SIT: supervisor: α=0.85; and colleagues: α=0.86. • SPOS: organization: α=0.81	Friends (SS-Friends); family (SS-Family); colleagues (SS-Coll); supervisors/managers (SS-S/M); and organization (SS-ORG)
Organizational resilience	Self-Evaluation of Resilience (SEOR) [54]	• Skill-based resilience: α=0.90; and individual-based resilience: α=0.81	Skill-based resilience (SEOR-SR); and individual-based resilience (SEOR-IR)

XSL•FO
RenderX

Figure 2. Flow diagram of the item selection and refinement process. ACS: Acceptance of Change Scale; CARE-QI 5.0: Change Adaptation Resilience Evaluation Questionnaire for Industry 5.0; COPE-NVI: Coping Orientation to Problems Experienced–Italian Version; ESQ: Emotional Style Questionnaire; LOC-EXT: external locus of control; LOC-INT: internal locus of control; OSES: Occupational Self-Efficacy Scale.



Literature Review

A comprehensive literature review led to the identification of a set of individual and contextual constructs empirically associated with resilience in organizational settings. Personal factors refer to internal dispositions and cognitive-emotional resources that individuals mobilize when facing adversity, such as self-efficacy, locus of control, emotion regulation strategies, and cognitive flexibility [24-30]. Conversely, contextual factors pertain to environmental resources and systemic support mechanisms, such as perceived social support and organizational adaptability, that fall outside the individual's direct control but critically shape their capacity for resilience [29,31].

Accordingly, the CARE-QI 5.0 scale was developed by organizing these constructs into 2 higher-order dimensions under the overarching factor of resilience: (1) personal factors, comprising the subdimensions of self-efficacy, locus of control (internal and external), emotion regulation strategies, change management and adaptation, and cognitive flexibility; and (2) contextual factors, comprising social support and organizational resilience. This theoretical model informed the development of an initial item pool consisting of 82 items, all distributed across the identified subdimensions. The selection of the underlying measurement tools for each construct was based on their conceptual alignment with the target population (ie, industrial workers), empirical validity, psychometric soundness, and adaptability to high-tech occupational contexts. Table 1 summarizes the constructs, the original instruments, Cronbach α values (all >0.70), and the first set of the selected subscales for the development of CARE-QI 5.0.

In particular, for the cognitive flexibility dimension, items were adapted from the Multidimensional Psychological Flexibility Inventory (MPFI) [50], which originally includes 6 first-order factors loading onto two higher-order constructs: flexibility and inflexibility. In this study, only these two broader domains were retained, omitting the subfactors, to provide a more parsimonious and context-sensitive model tailored to high-tech manufacturing environments.

Furthermore, the Social Support (SS) dimension was constructed by integrating items from three validated instruments to represent distinct sources of support relevant to workplace dynamics: (1) 4 items were drawn from the Italian version of

XSL•FO
RenderX

the Multidimensional Scale of Perceived Social Support [51], assessing support from family and friends; (2) 6 items from the Health and Safety Executive Stress Indicator Tool (SIT) [52] measured support from colleagues and supervisors or managers; and (3) 3 items from the Survey of Perceived Organizational Support [53] addressed support from the organization. This integrated approach allowed for a multidimensional and occupationally relevant operationalization of perceived social support.

Experts' Evaluation: Item Clarity Assessment

As summarized in Table 2, expert judgments were used to assess item clarity and comprehensibility. Of the 82 items originally included across 7 questionnaires, 32 items (39%) achieved full expert agreement on clarity. Notably, the SS questionnaire showed the highest agreement, with 85% of its items (11/13) rated as fully clear. In contrast, the ESQ [47] presented significant issues, with only 1 (8%) of 12 items receiving unanimous agreement. Experts provided extensive qualitative feedback regarding the ESQ's lack of contextual relevance and overly complex phrasing, especially in relation to workplace dynamics. As a result, the ESQ was removed from the battery and replaced with the COPE-NVI [55] (21 items), which better aligns with the aim of assessing emotional regulation strategies within occupational contexts and demonstrates solid internal consistency [55]. These revisions, theoretically driven and supported by experts' feedback, were made prior to pilot testing and are reflected in the final pool of 91 items. Table 3 reports the second set of questionnaires used in the development of CARE-QI 5.0.

Focus Groups

Participants of the Focus Groups

A total of 18 operators participated in 2 focus groups (mean age 49.22, SD 9.95 y). The sample was evenly split by sex (50% male and 50% female). Participants had an average job tenure of 29.53 (SD 12.04) years, with a range from 1 to 37 years. Most participants were Italian (n=13, 72.2%), while the remaining 5 (27.8%) were foreign nationals fluent in Italian.

The first focus group consisted of 8 participants (men: n=7, 87.5%), with a mean age of 45.9 (SD 12.12) years. The second group included 10 participants (women: n=8, 80%), with a mean age of 51.9 (SD 6.63) years.

Results Emerged From the Two Focus Groups

The overall clarity assessment indicated that 90.7% of the items were rated as clear by participants, while only 9.3% were identified as unclear. Suggestions for improvement primarily concerned wording adjustments to better reflect the manufacturing context. In particular, participants recommended avoiding the use of anglicisms (eg, *routine* and *feedback*) and suggested replacing generic terms such as "organization" with "company" to increase contextual relevance. Specific concerns emerged regarding the MPFI [50], whose items were perceived as ambiguous in distinguishing between work-related and personal domains. To address this, all MPFI items were reformulated to remove references to the workplace, ensuring they referred solely to individual characteristics across contexts.

Pilot Administration

Participants of the Pilot Study: Sociodemographic and Occupational-Related Characteristics

Table 4 presents the sample's sociodemographic and occupational-related characteristics.

A total of 44 participants completed the revised version of the questionnaire as part of the pilot administration. Age data were available for 43 (97.7%) participants (mean age 45.12, SD 11.74 y). No participants were excluded from the analyses. Most participants were Italian men (n=29, 65.9%) and had a mean job tenure of 18.42 (SD 8.55) years. The majority worked as operators (n=40, 90.9%), *while technicians (*n=4, 9.1%) were less represented. Assembly was the most common task (n=17, 38.6%), followed by welding (n=2, 4.5%), picking (n=2, 4.5%), and product quality control (n=2, 4.5%). Most participants had an open-ended contract (n=33, 75%), with fewer on interim (n=4, 9.1%), fixed-term contract (n=4, 9.1%), part-time contracts (n=1, 2.3%), or other types of contract (n=2, 4.5%), such as employee leasing and work under an independent contractor agreement.

Technology Characteristics

As reported in Table 4, regarding technology readiness, 63.6% reported an intermediate level, 27.3% reported a basic level, and only 9.1% reported a high level. On average, participants reported using technology frequently in their work (mean 3.93, SD 1.17) and had been working with such technologies for an average of 15.63 (SD 8.91) years. The majority had received training (n=33, 75%) for using such technologies within their workplace. Moreover, Table 5 presents the technology used for working. Among the wide range of technologies reported by participants, the most frequently used were robotic assembly lines (n=16, 36.36%) and measurement instruments for calibration and testing (n=16, 36.36%), followed by semimanual machining stations (n=12, 27.27%) and pillar drills or similar drilling or tapping machines (n=12, 27.27%). Other commonly mentioned technologies included loading and unloading robots (n=11, 25.00%), assembly tables (n=9, 20.45%), testing machines (n=9, 20.45%), and automated material handling systems such as conveyors and trolleys (n=9, 20.45%).

Reliability of the CARE-QI 5.0 Subscales

As presented in Table 6, most factors demonstrated acceptable to excellent reliability. Internal consistency was assessed using both Cronbach α and McDonald ω_total. Specifically, alpha coefficients ranged from 0.63 related to the Positive attitude subscale of the COPE-NVI scale (COPE-PA) [55] to 0.94 related to the Skill-based Resilience of the Self-Evaluation of Resilience scale (SEOR-SR) subscale, with CIs generally supporting the stability of internal consistency.

However, the internal (LOC-INT) and external (LOC-EXT) locus of control subscales yielded suboptimal reliability estimates. In particular, the analysis of LOC-INT revealed a negative α coefficient (α=−.04, 95% CI −0.845 to 0.420), suggesting problematic item intercorrelations. LOCL3 ("*A well-prepared person always finds a satisfying job"*) showed a negative correlation with LOCL1 (r=−0.27) and a very weak

XSL•FO
RenderX

correlation with LOCL2 (r=0.05). When LOCL3 was removed, internal consistency improved (α=.48) but still did not reach an acceptable threshold. Moreover, the LOC-EXT showed a low reliability (α=.595, 95% CI 0.287 to 0.759). On the basis of both statistical evidence and conceptual considerations, specifically, the deterministic and potentially externally oriented nature of LOCL3's content, which aligns more with Levenson's [56] depiction of external reinforcement expectancies than with Rotter's [57] original definition of internal locus of control, both the LOCL-INT and LOCL-EXT subscales were excluded from the final version of the resilience scale. A similar issue emerged in the "Confidence to persevere in the face of difficulties" (α=.087, 95% CI −0.452 to 0.309) subscale (OSES-CD) of the Occupational Self-Efficacy Scale (OSES) [46]. Correlational analysis revealed that item OSES8 ("*When unexpected problems happen at work, I don't handle them very well*") was negatively associated with all other items in the factor (eg, OSES8-OSES7, r=−0.15; OSES8-OSES9, r=−0.33), suggesting an inconsistency in directionality. Although the item was conceptually intended to be reverse scored, this transformation had not been applied prior to the initial analysis. After correcting this and reverse-coding OSES8 of the OSES, the internal consistency of the OSES-CD subscale improved substantially (α=.84), supporting its inclusion in the final structure of the scale. Moreover, the MPFI-CF showed a lower ω value compared to α, suggesting that the scale's internal consistency may be partially influenced by its multidimensional structure (acceptance and committed action as distinct but related facets). Despite this, the scale was retained due to its theoretical centrality in resilience models [58]. The global CARE-QI 5.0 score showed excellent reliability (α=.95; ω=0.90). At the domain level, both dimensions demonstrated good to excellent internal consistency. Specifically, both individual and contextual resilience showed excellent reliability (α=.93; ω=0.84; and α=.96; ω=0.96, respectively).

The final pool items were 85. Table 7 presents the third and final set of questionnaires used in the development of CARE-QI 5.0.

To further enhance the transparency of item-level functioning, a supplementary table (Table S4, Multimedia Appendix 1) was included reporting corrected item-total correlations and the change in internal consistency coefficients (Cronbach α and McDonald ω) if each item were deleted. This information provides a detailed overview of the contribution of each item to the internal consistency of its respective subscale.

Description and Contextual Relevance of CARE-QI 5.0 Subscales

Overview

To enhance interpretability within Industry 5.0 contexts, each subscale is briefly defined in relation to workers' adaptive functioning in technologically advanced environments [12]. The full item catalog (Italian and English versions), together with their original sources and type of adaptation, is reported in Table S1 in Multimedia Appendix 2. All items were derived from validated instruments and adapted following a theory-driven approach [34,36-38]. Adaptations varied depending on the source scale and included contextual modifications to the work environment (eg, for ACS and COPE), minor linguistic and semantic refinements (eg, for OSES and SEOR), and temporal reframing from past to present to ensure consistency in item format (eg, for MPFI) while preserving the original construct content.

Self-Efficacy (OSES-PS and OSES-CD)

The self-efficacy subscales reflect workers' perceived capability to manage complex and unpredictable work demands [59]. In Industry 5.0 environments, characterized by increasing HRI and system variability, *problem-solving self-efficacy* supports rapid decision-making and operational continuity, whereas *confidence in persevering* reflects persistence under prolonged uncertainty and disruptions [18,21,46].

Coping Strategies (COPE-AV, COPE-PA, COPE-PO, and COPE-SSS)

The coping strategy subscales represent behavioral and cognitive responses to stressors [60]. In high-tech manufacturing contexts, *problem-oriented coping* is particularly relevant for actively managing technological challenges and workflow disruptions, whereas *avoidance and positive attitude* may reflect regulatory processes that vary in adaptiveness depending on situational demands [61,62]. Seeking social support is critical in collaborative environments where problem-solving relies on interpersonal coordination [31].

Change Management and Adaptation (ACS-PTC, ACS-STC, and ACS-PRTC)

These dimensions capture workers' orientation toward change, a core requirement in Industry 5.0 systems characterized by continuous technological evolution [12]. For example, *openness and proactive engagement with change* have been associated with better adjustment to organizational transformation and innovation [24].

Cognitive Flexibility (MPFI-CF and MPFI-CINF)

Cognitive flexibility is a central resilience mechanism that enables individuals to adapt to changing demands and regulate emotional responses under stress [25,27]. In contrast, *cognitive inflexibility* reflects rigid cognitive patterns that may hinder adaptation, particularly in complex and dynamic technological environments [25,27].

Social Support (SS Subdomains)

The social support subdomains represent a key contextual resource for resilience, influencing both emotional regulation and adaptive functioning [28,32,51-53]. In Industry 5.0 environments, where teamwork and human-technology collaboration are central, support from colleagues, supervisors, and organizations facilitates trust and fluency [63].

Organizational Resilience (SEOR-SR and SEOR-IR)

The organizational resilience subscales refer to the capacity of organizations to adapt and sustain functioning under disruption events [11]. In Industry 5.0, this includes investment in training, skill development, and adaptive work structures that support effective human-technology integration and collective problem-solving [14,54].

XSL•FO
RenderX

Descriptives of the Items and Subscales

Table S2 in Multimedia Appendix 3 reports the descriptive statistics (mean, SD, skewness, and kurtosis) for each item and corresponding subscale of the newly developed CARE-QI 5.0 resilience questionnaire.

Overall, item means ranged from 1.80 to 4.25, suggesting a moderate-to-high endorsement of most items. Several items exhibited moderate negative skewness, indicating a tendency toward higher agreement levels. For example, OSES2 ("When I face difficulties at work, I remain calm because I can rely on my abilities") and item SS7 from the supervisors/managers subscale ("My direct supervisor provides what I need to perform my job effectively") showed pronounced negative skewness (−1.46 and −1.12, respectively), suggesting a potential ceiling effect. Some items also demonstrated leptokurtic distributions, such as OSES2 (kurtosis=5.47), indicating clustered responses around the mean.

At the subscale level, total means ranged from 1.94 (COPE-AV) to 3.94 (COPE-PO), highlighting notable variability in the endorsement of different resilience dimensions. Specifically, higher means were observed for problem-solving self-efficacy (OSES-PS=3.92), problem-oriented coping (COPE-PO=3.94), and cognitive flexibility (MPFI-CF=3.71), whereas lower scores characterized avoidance strategies (COPE-AV=1.94) and confidence to persevere in the face of difficulties (OSES-CD=2.07).

Interitem Correlations Within Subscales and Their Relationships With Total Scale Scores

Interitem correlation analyses within each subscale were conducted to evaluate the internal homogeneity of the items and are reported in Table S3 in Multimedia Appendix 4. Overall, most items showed moderate to strong correlations with one another within their respective subscales, supporting their coherence in measuring the same latent constructs.

However, for the Acceptance of Change Scale (ACS) [49], item ACS12 ("I am aware of the changes and modifications occurring in my workplace") displayed weak correlations with the other items in the Positive Reaction to Change (ACS-PRTC) subscale. Its removal led to an increase in internal consistency (from α=.666 to .673), suggesting limited contribution to the underlying construct. On the basis of both statistical evidence and conceptual clarity, ACS12 was excluded from the final version.

Thus, the final items included in the CARE-QI 5.0 were 84.

Interrelationships Between the CARE-QI 5.0 Subscales: Preliminary Evidence of Structural Coherence

As illustrated in Figure 3, problem-solving self-efficacy (OSES-PS) was positively associated with individual-based resilience (SEOR-IR), as well as with perceived support from the organization, colleagues, and family, cognitive flexibility (MPFI-CF), predisposition to change (ACS-PTC), and problem-oriented coping strategies (COPE-PO).

XSL•FO
RenderX

Figure 3. Correlations between the subscales of the Change Adaptation Resilience Evaluation Questionnaire for Industry 5.0. ACS-PRTC: positive reaction to change; ACS-PTC: predisposition to change; ACS-STC: seeking to change; COPE-AV: coping orientation to problems experienced-avoidance strategies; COPE-PA: coping orientation to problems experienced-positive strategies; COPE-PO: coping orientation to problems experienced-problem oriented; COPE-SSS: coping orientation to problems experienced-seeking social support; MPFI-CF: cognitive flexibility; MPFI-CINF: cognitive inflexibility; OSES-CD: Occupational Self-Efficacy Scale-confidence to persevere in the face of difficulties; OSES-PS: Occupational Self-Efficacy Scale-problem solving; SEOR-IR: Self-Evaluation of Resilience-individual-based resilience; SEOR-SR: Self-Evaluation of Resilience-skill-based resilience; SS-Coll: Social Support-colleagues; SS-Family: Social Support-family; SS-Friends: Social Support-friends; SS-ORG: Social Support-organization; SS-S/M: Social Support-supervisors/managers; TechnoFreq: frequency of technology use; TechnoYears: years of experience with technology.



The dimension of perseverance in facing difficulties (OSES-CD) showed positive correlations with cognitive inflexibility (MPFI-CINF), seeking to change (ACS-STC), positive attitude (COPE-PA), and avoidance strategies (COPE-AV), indicating a less adaptive pattern of coping.

Both positive attitude and avoidance coping strategies were positively associated with cognitive inflexibility. Additionally, positive attitude was also positively correlated with support from the organization and colleagues and with positive reaction to change (ACS-PRTC); conversely, avoidance was positively associated with support from managers or supervisors.

Problem-oriented coping strategies (COPE-PO) were significantly and positively related to both dimensions of organizational resilience (SEOR-SR and SEOR-IR), support from the organization, colleagues, and family, cognitive flexibility, and 2 subdimensions of change management and adaptation (ie, ACS-PRTC and ACS-PTC).

The ACS-PTC subscale was positively correlated with support from colleagues, family, and friends and with cognitive flexibility. The ACS-STC dimension showed positive associations with seeking social support (COPE-SSS) and perceived support from friends, individual-based resilience, and both dimensions of cognitive (in)flexibility. The positive reaction to change dimension (ACS-PRTC) was positively associated with skill-based resilience (SS-Coll), support from colleagues and family, and cognitive flexibility.

Cognitive flexibility itself was positively related to skill-based resilience, as well as to COPE-SSS, the support from organization, colleagues, and friends, reinforcing its central role in adaptive functioning.

XSL•FO
RenderX

Interrelationships Between the CARE-QI 5.0 Subscales and Occupational and Technology Variables

As shown in Figure 3, age was negatively associated with perceived social support from supervisors or managers and with individual-based resilience (SEOR-IR). In contrast, it showed positive correlations with both experience in using technology (TechnoYears) and job tenure. Job tenure was negatively correlated with skill-based resilience (SEOR-SR), while it was positively associated with technological experience. Technology readiness was positively related to both technology use frequency (TechnoFreq) and experience (TechnoYears). Moreover, technology use frequency was positively associated with support from friends. In addition, experience in using technology was negatively correlated with individual-based resilience, as well as with perceived support from supervisors or managers and COPE-SSS.

Discussion

Principal Findings

The overarching aim of this study was to develop and preliminarily evaluate the CARE-QI 5.0, a multidimensional self-report instrument designed to assess resilience in manufacturing environments. The scale is grounded in a global score of resilience, with a dual-factor model, that conceptualizes adaptability as the result of 2 interrelated components: individual resilience (eg, self-efficacy, coping strategies, cognitive flexibility, and change management and adaptation) and contextual resilience (eg, perceived social support and organizational resilience). This framework reflects the human-centered and sustainable vision of Industry 5.0, where the technological transformation of production systems must be accompanied by the promotion of workers' well-being, agency, and psychological adaptability [12].

In particular, this human-centered shift involves increasing forms of HRI and HRC, which require operators to adapt not only to new tools but also to novel relational dynamics with intelligent technologies. Understanding how workers psychologically respond to such transitions is therefore crucial for the sustainable integration of advanced automation.

To ensure theoretical and contextual robustness, the development of the questionnaire followed a 4-phase mixed methods approach: (1) a literature review, (2) expert review, (3) focus groups with manufacturing workers, and (4) a pilot administration aimed at assessing the internal consistency and the correlational structure of the instrument.

Insights From the Preliminary Evaluation of the CARE-QI 5.0

Building on the literature review, a panel of psychologists and HCI experts was involved in a structured content and semantic evaluation of the selected items. This expert review phase was instrumental in validating the proposed dual-factor model of resilience, encompassing both individual and contextual resources, and in defining the preliminary structure of the CARE-QI 5.0. Items that did not meet minimum thresholds for clarity, comprehensibility, or contextual appropriateness were modified or excluded. In particular, although the ESQ was theoretically consistent with core resilience dimensions [47], it was considered too abstract and insufficiently grounded in the lived experiences of industrial workers. Consequently, ESQ-derived items were replaced with items from the COPE-NVI [55], a tool offering a more concrete and behaviorally oriented operationalization of coping strategies and previously validated in occupational settings [55].

Participants generally confirmed the clarity of the items (with over 90% rated as "clear") but also highlighted the presence of language that was either too abstract or unfamiliar, including anglicisms and psychologisms. On the basis of their suggestions, numerous items were reworded to improve operational clarity, and the wording of certain scales, such as cognitive flexibility, was adjusted to ensure their applicability to daily work tasks, interpersonal relations, and problem-solving in the manufacturing context. This process enhanced both the ecological validity and the practical usability of the CARE-QI 5.0 instrument in Industry 5.0 settings.

The pilot testing phase provided evidence of satisfactory internal consistency for the vast majority of subscales, with Cronbach α and McDonald ω coefficients ranging from acceptable to excellent. Notably, at the global level, the CARE-QI 5.0 demonstrated excellent reliability (α=.95 and ω=0.90), suggesting that the overall instrument shows strong internal consistency. At the domain level, both dimensions showed good to excellent reliability: individual resilience (α=.93; ω=0.84) and contextual resilience (α=.96; ω=0.96), further supporting the structural coherence of the instrument. Notably, the cognitive flexibility scale showed a lower ω value compared to alpha, which may reflect the multidimensional structure of the scale, as it comprises acceptance and committed action as theoretically distinct yet related aspects of psychological flexibility. Despite this, the scale was retained, given its theoretical centrality in resilience models: cognitive flexibility, intended as the ability to adapt one's cognitive repertoire in response to changing demands, has been consistently identified as a core component of psychological resilience, particularly in technologically evolving work environments where continuous adaptation to new tools, processes, and roles is required [58]. Only the locus of control dimensions did not meet the reliability threshold and were therefore excluded from the final structure of the questionnaire, which now includes 84 items across the individual and contextual domains. Correlational analyses further supported the structural coherence of the instrument. All intersubscale correlations remained below the critical threshold of r=0.85, indicating an absence of multicollinearity and acceptable discriminant validity. On the basis of Cohen's [44] guidelines, most associations between dimensions fell within the small to moderate range, suggesting related yet distinct constructs, consistent with a hierarchical model of resilience.

Internal Structure and Construct Validity

To assess construct validity, correlations between resilience dimensions and sociodemographic variables were also examined. As expected, age was positively associated with both technological experience and job tenure; yet age was negatively associated with individual-based resilience and perceived

XSL•FO
RenderX

support from supervisors or managers. Similarly, technological experience was negatively related to individual-based resilience and to seeking social support and perceived support from supervisors, suggesting that increased exposure to technology in the workplace may not necessarily be coupled with perceived psychosocial support or internal resilience resources. Conversely, technology readiness was positively associated with both technology use frequency and experience, and the frequency of technology use was positively related to support from friends, indicating that digital engagement may reflect more socially integrated or voluntary patterns of use.

From a psychological perspective, problem-solving self-efficacy emerged as a construct deeply embedded in adaptive functioning. Its positive associations with individual-based resilience, cognitive flexibility, support from key relational and organizational sources, predisposition to change, and problem-oriented coping strategies reinforce its role as a central competence underpinning resilient responses. This pattern reflects a flexible and agentic configuration, in which confidence in one's problem-solving abilities coexists with openness to change and access to supportive contexts. In contrast, the perseverance in facing difficulties (OSESCD) showed a more complex and potentially maladaptive profile. While traditionally viewed as a marker of resilience, its positive associations with cognitive inflexibility, avoidance, positive attitude, and seeking to change, in the absence of links with organizational support or self-efficacy, suggest a form of rigid perseverance, more closely aligned with resignation or effortful persistence than with adaptive coping. This interpretation can be supported by the content of the OSESCD items (eg, "I avoid trying new things when they seem too difficult"), which may reflect passivity or learned helplessness, particularly under conditions of perceived threat or low external support.

A particularly nuanced picture emerged around avoidance and positive attitude strategies, which were both positively associated with cognitive inflexibility but also with cognitive flexibility. This dual alignment underscores their ambivalent nature: while they may serve immediate emotion-regulation functions, their adaptiveness likely depends on the broader cognitive-emotional configuration in which they operate [61]. When enacted within a flexible mindset and embedded in supportive environments, they may contribute to adjustment. However, when embedded in rigid cognitive structures, they risk becoming maladaptive coping habits, promoting disengagement or illusory optimism [62,64].

In line with this dynamic perspective, cognitive inflexibility was largely uncorrelated with external resources (eg, organizational resilience and social support) but was positively associated with rigid or passive profiles, including perseverance in facing difficulties, avoidance, and positive attitude. This reinforces its conceptualization as a marker of limited adaptability, associated with reduced access to both internal and external regulatory mechanisms.

Conversely, cognitive flexibility emerged as a central protective factor, positively linked to both skill- and individual-based resilience, seeking social support and perceived support from colleagues and the organization, and openness to change across all 3 change-related subdimensions. Its wide-ranging associations suggest a pivotal role in integrating coping strategies, relational support, and openness to innovation, especially in complex and evolving work environments, such as those characterizing advanced manufacturing. Problem-oriented coping strategies aligned consistently with cognitive flexibility, resilience dimensions, and social-organizational support, highlighting their embeddedness in broader adaptive networks rather than standing as isolated traits.

Taken together, these findings support a process-oriented view of resilience, which emerges not from single traits or isolated strategies but from the dynamic interplay of flexible cognition, effective coping, change orientation, and social-organizational resources. The value of any strategy appears to depend less on its nominal category and more on how it functions within broader patterns of adaptation and support.

Technology Readiness and Implications for Psychological Resilience

In addition to the psychometric evaluation of CARE-QI 5.0, the pilot phase provided valuable insights into workers' technological exposure, which is a critical contextual factor for interpreting resilience in Industry 5.0 environments [12]. Most participants reported an intermediate (63.6%) or basic (27.3%) level of technology readiness, with only a small subset (9.1%) indicating high readiness. Nevertheless, technology use in daily operations was generally frequent (mean 3.93, SD 1.17), and workers reported an average of over 15 years of experience (mean 15.63, SD 8.91) with advanced manufacturing technologies. The most commonly used tools included robotic assembly lines and precision measurement instruments (both reported by 36.36% of the sample), followed by semimanual machining stations and pillar drills (each 27.27%). A diverse array of technologies was also mentioned, including loading and unloading robots (25.00%), assembly tables, testing machines, and automated material handling systems (each 20.45%). Although less prevalent, the use of collaborative robots (6.82%) and AI-assisted systems, such as MES platforms and real-time monitoring tools, suggests an ongoing, albeit gradual, transition toward more integrated HRC scenarios. Notably, the majority of participants (n=33, 75%) had received specific training for using these technologies, highlighting the importance of organizational investment in upskilling as a facilitator of technological adaptation. This finding reinforces the conceptualization of resilience as an interaction between individual resources and contextual supports, thereby underscoring how technological competence, fostered through structured training, may serve as a protective factor in high-tech industrial settings. Correlational analyses offered further insight into the psychological significance of these technological variables. Technology readiness was positively associated with the frequency of technology use and years of technological experience.

Practical Implications

The CARE-QI 5.0 seems to be a robust and innovative tool for assessing psychological resilience in organizational settings, especially those undergoing significant technological and

XSL•FO
RenderX

structural changes. Developed within the theoretical framework of Industry 5.0, the instrument embraces a human-centered approach that acknowledges the complexity of workers' experiences in increasingly automated and digitized environments.

CARE-QI 5.0 is further characterized by a multidimensional structure, conceptualizing resilience as a higher-order factor composed of 2 second-order domains: individual resilience and contextual resilience. These domains are in turn articulated through a series of first-order subdimensions, including self-efficacy, coping strategies, cognitive flexibility, change management and adaptation, perceived social support, and organizational resilience. This layered architecture enables a comprehensive yet nuanced evaluation of workers' adaptive capacities, considering not only internal dispositions but also the quality of their social and organizational environment. Such a structure is particularly relevant in light of the challenges posed by Industry 5.0 transformations, which involve close interaction between human operators and advanced technologies, such as collaborative robots [20], real-time monitoring systems, and intelligent automation. These changes impact not only task demands and technical skills but also workers' psychological adaptability, stress regulation, and relational dynamics. CARE-QI 5.0 can therefore serve as a valuable instrument for diagnosing resilience profiles and for monitoring how employees adjust to innovation over time. Its use may inform targeted interventions aimed at strengthening both personal and systemic protective resources. Clinical and work and organizational psychologists, HRC experts, and HR departments may use the tool to detect early signals of psychological difficulties, map areas of risk or resilience, and tailor actions to support workers' well-being. Moreover, CARE-QI 5.0 enables organizations to move beyond generic wellness initiatives and adopt evidence-based strategies for promoting sustainable human performance. By identifying which dimensions of resilience are most challenged in a given context, or reduced change readiness, companies can enact specific and measurable improvements in their work environment and organizational culture.

Therefore, CARE-QI 5.0 provides not only a means of assessing workers' adaptability but also a strategic foundation for building more resilient, inclusive, and future-ready workplaces in alignment with the social and technological principles of Industry 5.0.

Limitations and Future Directions

Despite its promising contributions, this study presents several limitations that should be considered, also for future research. First, the psychometric evaluation of the CARE-QI 5.0 remains preliminary. While internal consistency and intersubscale correlations provided initial evidence of structural coherence, no factor-analytic validation was conducted. Given the scale's multidimensional structure, with resilience conceptualized as a general construct composed of 2 higher-order domains (individual and contextual resilience) and multiple first-order subdimensions, future research should use exploratory factor analysis and confirmatory factor analysis, including second-order and bifactor models, to test the hierarchical configuration of the instrument. Second, the sample was small and drawn from a specific regional and sectoral context, namely, manufacturing companies located in Northeastern Italy, primarily within the mechanical and packaging sectors. While these sectors represent relevant use cases in Industry 5.0 transitions, the limited scope may restrict the generalizability of the findings. Further research is needed to assess the applicability and cultural fit of the instrument in other manufacturing domains among a large sample size. Third, an additional limitation concerns the expert review phase. The panel included a small sample size (n=3 experts), which limits the stability and interpretability of item-level agreement estimates. Moreover, the expert evaluation focused exclusively on item clarity and comprehensibility, without formally assessing item essentiality. As such, the current findings should be considered preliminary with respect to content validity. Notably, this limitation was partially mitigated by the inclusion of focus groups with target users, which provided complementary evidence on item comprehensibility and contextual appropriateness. Future research should involve a larger panel of subject matter experts (eg, 5 - 10 or more) and adopt more robust content validity procedures, including the use of indices such as the Content Validity Index, or Aiken V, ideally accompanied by CIs. Future validation efforts should also assess multiple aspects of content validity, including relevance and comprehensiveness, potentially integrating both experts' judgment and feedback representatives of the target population. Finally, although qualitative methods (eg, expert review and focus groups) contributed to the linguistic and contextual refinement of the items, no analysis of measurement invariance was conducted. To ensure that the CARE-QI 5.0 functions equivalently across key demographic groups (eg, gender, age, and professional role), future studies should assess its psychometric stability and fairness using multigroup confirmatory factor analysis.

Conclusions

This study contributes to the growing effort to develop context-sensitive tools for assessing psychological resilience within the evolving landscape of contemporary manufacturing. Through a mixed methods approach with several processes, encompassing theoretical review, expert evaluation, focus groups, and pilot testing, the CARE-QI 5.0 was developed as a multidimensional instrument, grounded in a dual framework of individual and contextual resilience.

Preliminary findings support the internal consistency and conceptual distinctiveness of its subdimensions, revealing the dynamic interplay between personal coping resources, cognitive adaptability, and perceived social-organizational support in shaping workers' capacity to respond to organizational and technological transitions. Correlational analyses suggest that constructs, such as problem-solving self-efficacy, cognitive flexibility, and support from supervisors and colleagues, form core adaptive clusters, while other patterns, such as rigid perseverance and avoidance strategies, highlight the risk of maladaptive coping when flexibility and support are lacking.

By capturing both individual dispositions (eg, self-efficacy and openness to change) and environmental supports (eg, social support and organizational culture), CARE-QI 5.0 offers a

XSL•FO
RenderX

comprehensive and integrative framework for evaluating psychosocial adaptability in high-tech, human-centered contexts. This makes the instrument especially relevant to the principles of Industry 5.0, where technological innovation is increasingly intertwined with sustainability, inclusion, and well-being.

While further psychometric validation and cross-contextual applications are needed, CARE-QI 5.0 holds promise as both a research instrument and a practical tool for high-tech manufacturing environments. It can assist organizations in mapping resilience profiles, guiding the design of targeted interventions, and promoting psychosocial sustainability in technologically advanced workplaces. In doing so, it supports a vision of industrial innovation that places psychological adaptability and human value at the core of progress.

Acknowledgments

We extend our sincere gratitude to the 2 companies and all participating operators for their valuable contribution to this study. Moreover, we thank the 3 independent experts for their suggestions in ameliorating the CARE-QI 5.0: Serena Tassoni, Valeria Orso, and Luca Menghini. No generative artificial intelligence tools were used at any stage of the manuscript preparation or writing process.

Funding

The research is funded within the project CARE-QI 5.0 - Change, Adaptation, Resilience, Evaluation Questionnaire for Industry 5.0: Strumento di valutazione per un'industria resiliente, within the framework of the Ecosystem INEST-Interconnected Nord-Est Innovation (PNRR) project. Notably, the content of the manuscript represents exclusively the authors' point of view, neither the European Union nor the European Commission should be considered accountable for its content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Item-level reliability diagnostic (corrected item-total correlations, Cronbach α, and McDonald ω if item deleted) for the CARE-QI 5.0 subscales.
[DOCX File, 28 KB - humanfactors_v13i1e83754_app1.docx]

Multimedia Appendix 2
CARE-QI 5.0 structure and items selected (N=84 items).
[DOCX File, 39 KB - humanfactors_v13i1e83754_app2.docx]

Multimedia Appendix 3
. Items' descriptive statistics.
[DOCX File, 35 KB - humanfactors_v13i1e83754_app3.docx]

Multimedia Appendix 4
Interitem correlations within subscales with their total scales. *P<.05; **P<.001.
[DOCX File, 39 KB - humanfactors_v13i1e83754_app4.docx]

References

1. Ma C, Jia Z, Zhang J, Duan M, Wang W. Dynamic modeling and analysis of Bernoulli production lines with small-lot orders and product quality problems. IEEE Trans Ind Inf 2024;21(1):258-266. [doi: 10.1109/TII.2024.3452196]
2. Kagermann H. Change through digitization-value creation in the age of industry 4.0. In: Albach H, Meffert H, Pinkwart A, Reichwald R, editors. Management of Permanent Change: Springer Gabler; 2015. [doi: 10.1007/978-3-658-05014-6_2]
3. Rikalovic A, Suzic N, Bajic B, Piuri V. Industry 4.0 implementation challenges and opportunities: a technological perspective. IEEE Systems Journal 2022;16(2):2797-2810. [doi: 10.1109/JSYST.2021.3101673]
4. Segura P, Lobato-Calleros O, Ramírez-Serrano A, Hernández-Martínez EG. Safety assurance in human-robot collaborative systems: a survey in the manufacturing industry. Procedia CIRP 2022;107:740-745. [doi: 10.1016/j.procir.2022.05.055]
5. Dhanda M, Rogers BA, Hall S, Dekoninck E, Dhokia V. Reviewing human-robot collaboration in manufacturing: opportunities and challenges in the context of industry 5.0. Robot Comput Integr Manuf 2025 Jun;93:102937. [doi: 10.1016/j.rcim.2024.102937]
6. Luozzo SD, Starnoni F, Schiraldi MM. On the relationship between human factor and overall equipment effectiveness (OEE): an analysis through the adoption of analytic hierarchy process and ISO 22400. International Journal of Engineering Business Management 2023 Feb;15:18479790231188548. [doi: 10.1177/18479790231188548]

XSL•FO
RenderX

7. Zhao Y, Said R, Ismail NW, Hamzah HZ. Impact of industrial robot on labour productivity: empirical study based on industry panel data. Innovation and Green Development 2024 Jun;3(2):100148. [doi: 10.1016/j.igd.2024.100148]
8. Duan D, Chen S, Feng Z, Li J. Industrial robots and firm productivity. Structural Change and Economic Dynamics 2023 Dec;67:388-406. [doi: 10.1016/j.strueco.2023.08.002]
9. Simone VD, Pasquale VD, Giubileo V, Miranda S. Human-robot collaboration: an analysis of worker's performance. Procedia Comput Sci 2022;200:1540-1549. [doi: 10.1016/j.procs.2022.01.355]
10. Lu Y, Adrados JS, Chand SS, Wang L. Humans are not machines—anthropocentric human–machine symbiosis for ultra-flexible smart manufacturing. Engineering 2021 Jun;7(6):734-737. [doi: 10.1016/j.eng.2020.09.018]
11. Shela V, Ramayah T, Noor Hazlina A. Human capital and organisational resilience in the context of manufacturing: a systematic literature review. JIC 2023 Mar 21;24(2):535-559. [doi: 10.1108/JIC-09-2021-0234]
12. Breque M, De Nul L, Petridis A. Industry 5.0: Towards a Sustainable, Human-Centric and Resilient European Industry: Publications Office of the European Union; 2021. [doi: 10.2777/308407]
13. Gamberini L, Pluchino P. Industry 5.0: a comprehensive insight into the future of work, social sustainability, sustainable development, and career. Australian Journal of Career Development 2024 Apr;33(1):5-14. [doi: 10.1177/10384162241231118]
14. Dhungana D, Haselböck A, Schmidbauer C, Taupe R, Wallner S. Enabling resilient production through adaptive human-machine task sharing. Presented at: 8th Changeable, Agile, Reconfigurable and Virtual Production Conference (CARV 2021) and 10th World Mass Customization & Personalization Conference (MCPC 2021); Nov 1-2, 2021; Aalborg, Denmark p. 198-206. [doi: 10.1007/978-3-030-90700-6_22]
15. Davila-Gonzalez S, Martin S. Human Digital Twin in Industry 5.0: a holistic approach to worker safety and well-being through advanced AI and emotional analytics. Sensors (Basel) 2024 Jan 19;24(2):655. [doi: 10.3390/s24020655] [Medline: 38276347]
16. Trstenjak M, Benešova A, Opetuk T, Cajner H. Human Factors and Ergonomics in Industry 5.0—A systematic literature review. Applied Sciences 2025;15(4):2123. [doi: 10.3390/app15042123]
17. Villani V, Gabbi M, Sabattini L. Promoting operator's wellbeing in industry 5.0: detecting mental and physical fatigue. Presented at: 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC); Oct 9-12, 2022. [doi: 10.1109/SMC53654.2022.9945324]
18. Romero D, Stahre J. Towards the resilient operator 5.0: the future of work in smart resilient manufacturing systems. Procedia CIRP 2021;104:1089-1094. [doi: 10.1016/j.procir.2021.11.183]
19. Gladysz B, Tran TA, Romero D, van Erp T, Abonyi J, Ruppert T. Current development on the Operator 4.0 and transition towards the Operator 5.0: a systematic literature review in light of Industry 5.0. Journal of Manufacturing Systems 2023 Oct;70:160-185. [doi: 10.1016/j.jmsy.2023.07.008]
20. Bassi G, Valente A, Tassoni S, Salcuni S. Psychological resilience in the age of industry 5.0: a systematic review of the manufacturing workforce. Eur Psychol 2026 Jul 7. [doi: 10.1027/1016-9040/a000597]
21. Fletcher D, Sarkar M. Psychological resilience: a review and critique of definitions, concepts, and theory. Eur Psychol 2013;18(1):12-23. [doi: 10.1027/1016-9040/a000124]
22. Vella SL, Pai N. A theoretical review of psychological resilience: defining resilience and resilience research over the decades. Arch Med Health Sci 2019;7(2):233. [doi: 10.4103/amhs.amhs_119_19]
23. The road to resilience. : American Psychological Association; 2014 URL: https://www.apa.org/helpcenter/road-resilience [accessed 2026-07-24]
24. van den Heuvel M, Demerouti E, Bakker AB, Hetland J, Schaufeli WB. How do employees adapt to organizational change? The role of meaning-making and work engagement. Span J Psychol 2020 Dec 21;23:e56. [doi: 10.1017/SJP.2020.55] [Medline: 33345770]
25. Rademacher L, Kraft D, Eckart C, Fiebach CJ. Individual differences in resilience to stress are associated with affective flexibility. Psychol Res 2023 Sep;87(6):1862-1879. [doi: 10.1007/s00426-022-01779-4] [Medline: 36528692]
26. Ouellette S, Longato S, Mayer W, Bouchard F, Boudrias JS. Cognitive flexibility and adaptive performance among workers: testing the moderating effect of the situational strength on this relationship. Can J Adm Sci 2024 Dec;41(4):549-557. [doi: 10.1002/cjas.1744]
27. Abate BB, Sendekie AK, Tadesse AW, et al. Resilience after adversity: an umbrella review of adversity protective factors and resilience-promoting interventions. Front Psychiatry 2024;15:1391312. [doi: 10.3389/fpsyt.2024.1391312] [Medline: 39429523]
28. Windle G. What is resilience? A review and concept analysis. Rev Clin Gerontol 2011 May;21(2):152-169. [doi: 10.1017/S0959259810000420]
29. Turner M, Holdsworth S. Developing resilience: examining the protective factors of early career construction professionals. Construction Management and Economics 2023 Oct 3;41(10):805-819. [doi: 10.1080/01446193.2023.2208238]
30. Sisto A, Vicinanza F, Campanozzi LL, Ricci G, Tartaglini D, Tambone V. Towards a transversal definition of psychological resilience: a literature review. Medicina (Kaunas) 2019 Nov 16;55(11):745. [doi: 10.3390/medicina55110745] [Medline: 31744109]

XSL•FO
RenderX

31. Fiaz S, Muhammad Fahim S. The influence of high-quality workplace relational systems and mindfulness on employee work engagement at the time of crises. Heliyon 2023 Apr;9(4):e15523. [doi: 10.1016/j.heliyon.2023.e15523] [Medline: 37128340]
32. Stephens JP, Heaphy ED, Carmeli A, et al. Relationship quality and virtuousness: emotional carrying capacity as a source of individual and team resilience. J Appl Behav Sci 2013;49(1):13-41. [doi: 10.1177/0021886312471193]
33. Babamiri M, Heidarimoghadam R, Ghasemi F, Tapak L, Mortezapour A. Going beyond general stress scales: developing a new questionnaire to measure stress in human-robot interaction. Int J Soc Robotics 2024 Dec;16(11-12):2243-2259. [doi: 10.1007/s12369-024-01183-5]
34. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Front Public Health 2018;6:149. [doi: 10.3389/fpubh.2018.00149] [Medline: 29942800]
35. WMA Declaration of Helsinki – ethical principles for medical research involving human participants. World Medical Association. 2024. URL: https://www.wma.net/policies-post/wma-declaration-of-helsinki/ [accessed 2026-07-24]
36. DeVellis RF. Scale Development: Theory and Applications, 4th edition: SAGE Publications; 2017.
37. Stefana A, Damiani S, Granziol U, et al. Psychological, psychiatric, and behavioral sciences measurement scales: best practice guidelines for their development and validation. Front Psychol 2024;15:1494261. [doi: 10.3389/fpsyg.2024.1494261] [Medline: 39916786]
38. Hinkin TR. A review of scale development practices in the study of organizations. J Manage 1995 Oct;21(5):967-988. [doi: 10.1177/014920639502100509]
39. Freeman T. “Best practice” in focus group research: making sense of different views. J Adv Nurs 2006 Dec;56(5):491-497. [doi: 10.1111/j.1365-2648.2006.04043.x] [Medline: 17078825]
40. Streiner DL, Norman GR, Cairney J. Health Measurement Scales: A Practical Guide to Their Development and Use, 5th edition: Oxford University Press; 2015. [doi: 10.1093/med/9780199685219.001.0001]
41. Kapuscinski AN, Masters KS. Current status of measures of spirituality: a critical review of scale development. Psycholog Relig Spiritual 2010;2(4):191-205. [doi: 10.1037/a0020498]
42. Willis GB. Cognitive Interviewing: A Tool for Improving Questionnaire Design: SAGE Publications; 2005. [doi: 10.4135/9781412983655]
43. Posit Team. RStudio: Integrated Development Environment for R: Posit; 2025. URL: https://posit.co/ [accessed 2026-07-24]
44. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: Routledge; 1988. [doi: 10.4324/9780203771587]
45. Brown TA. Confirmatory Factor Analysis for Applied Research, 2nd edition: Guilford Publications; 2015.
46. Tani F, Lazzaretti R, Maggino F, Smorti M, Giannini M. An instrument for measuring occupational self-efficacy: an Italian adaptation of the Occupational Self-Efficacy Scale. Giornale Italiano Psicologia 2009;36(3):657-674. [doi: 10.1421/30205]
47. Malandrone F, Catalano A, Carfì F, et al. Psychometric validation of the Italian version of the Emotional Style Questionnaire. PLoS ONE 2022;17(12):e0278715. [doi: 10.1371/journal.pone.0278715] [Medline: 36459526]
48. Argentero P, Vidotto G. LOC-L: Una Scala Di Locus of Control Lavorativo Manuale: Mediatest; 1994. URL: https://www.research.unipd.it/handle/11577/166170?mode=complete [accessed 2026-08-05]
49. Di Fabio A, Gori A. Developing a new instrument for assessing acceptance of change. Front Psychol 2016;7:802. [doi: 10.3389/fpsyg.2016.00802] [Medline: 27303356]
50. Landi G, Pakenham KI, Giovannetti AM, et al. Italian validation of the Italian multidimensional psychological flexibility inventory (MPFI). J Contextual Behav Sci 2021 Jul;21:57-65. [doi: 10.1016/j.jcbs.2021.05.007]
51. De Maria M, Vellone E, Durante A, Biagioli V, Matarese M. Psychometric evaluation of the Multidimensional Scale of Perceived Social Support (MSPSS) in people with chronic diseases. Ann Ist Super Sanita 2018;54(4):308-315. [doi: 10.4415/ANN_18_04_07] [Medline: 30575567]
52. Balducci C, Romeo L, Brondino M, et al. The validity of the short UK Health and Safety Executive Stress Indicator Tool for the assessment of the psychosocial work environment in Italy. European Journal of Psychological Assessment 2017 Jul;33(3):149-157. [doi: 10.1027/1015-5759/a000280]
53. Battistelli A, Mariani MG. Perceived organizational support: validation of the Italian version of the Survey of Perceived Organizational Support (with eight items). Giornale Italiano Psicologia 2011;38(1):189-214. [doi: 10.1421/34845]
54. Pennini A, Caruso R, Conte G, et al. Is the self-evaluation of resilience a valid assessment to measure resilience in healthcare? A confirmatory validation study in Italian healthcare settings. Eval Health Prof 2023 Dec;46(4):396-404. [doi: 10.1177/01632787231170236] [Medline: 37051751]
55. Caricati L, Foà C, Fruggeri L, Tonarelli A. COPE-NVI-25: validazione italiana della versione ridotta della Coping Orientation to the Problems Experienced (COPE-NVI). PSICOLOGIA DELLA SALUTE 2015;2(2):123-140. [doi: 10.3280/PDS2015-002007]
56. Levenson H. Multidimensional locus of control in psychiatric patients. J Consult Clin Psychol 1973 Dec;41(3):397-404. [doi: 10.1037/h0035357] [Medline: 4803272]
57. Rotter JB. Generalized expectancies for internal versus external control of reinforcement. Psychol Monogr 1966;80(1):1-28. [doi: 10.1037/h0092976] [Medline: 5340840]

XSL•FO
RenderX

58. Kashdan TB, Rottenberg J. Psychological flexibility as a fundamental aspect of health. Clin Psychol Rev 2010 Nov;30(7):865-878. [doi: 10.1016/j.cpr.2010.03.001] [Medline: 21151705]
59. Bandura A. Self-Efficacy: The Exercise of Control: Freeman; 1997.
60. Lazarus RS, Folkman S. Stress, Appraisal, and Coping: Springer publishing company; 1984.
61. Bonanno GA, Burton CL. Regulatory Flexibility. Perspect Psychol Sci 2013 Nov;8(6):591-612. [doi: 10.1177/1745691613504116]
62. Aldao A, Nolen-Hoeksema S. The influence of context on the implementation of adaptive emotion regulation strategies. Behav Res Ther 2012 Aug;50(7-8):493-501. [doi: 10.1016/j.brat.2012.04.004] [Medline: 22659159]
63. Bassi G, Orso V, Salcuni S, Gamberini L. Workers' well-being and cognitive load in human-cobot collaboration: systematic review. J Med Internet Res 2025 Aug 27;27:e75658. [doi: 10.2196/75658] [Medline: 40865092]
64. Taylor SE, Brown JD. Illusion and well-being: a social psychological perspective on mental health. Psychol Bull 1988 Mar;103(2):193-210. [doi: 10.1037/0033-2909.103.2.193] [Medline: 3283814]

Abbreviations

ACS: Acceptance of Change Scale
ACS-PRTC: positive reaction to change
ACS-PTC: predisposition to change
CARE-QI 5.0: Change Adaptation Resilience Evaluation Questionnaire for Industry 5.0
COPE: coping orientation to problems experienced
COPE-AV: coping orientation to problems experienced-avoidance strategies
COPE-PA: coping orientation to problems experienced-positive strategies
COPE-PO: coping orientation to problems experienced-problem oriented
COPE-SSS: coping orientation to problems experienced-seeking social support
ESQ: Emotional Style Questionnaire
HRC: human-robot collaboration
HRI: human-robot interaction
LOC-EXT: external locus of control
LOC-INT: internal locus of control
LOC-L: Locus of Control at Work
MPFI: Multidimensional Psychological Flexibility Inventory
MPFI-CF: cognitive flexibility
MPFI-CINF: cognitive inflexibility
MSPSS: multidimensional scale of perceived social support
OSES: Occupational Self-Efficacy Scale
OSES-CD: Occupational Self-Efficacy Scale-Confidence to persevere in the face of difficulties
OSES-PS: Occupational Self-Efficacy Scale-Problem Solving
SEOR-IR: Self Evaluation of Resilience-Individual-based resilience
SEOR-SR: Self Evaluation of Resilience-Skill-based resilience
SS: Social Support

Edited by A Kushniruk; submitted 08.Sep.2025; peer-reviewed by JR de-Moya-Romero, P Larionow; revised version received 27.Apr.2026; accepted 04.May.2026; published 28.Aug.2026.

Please cite as:
Bassi G, Valente A, Tassoni S, Salcuni S
Change Adaptation Resilience Evaluation Questionnaire for Industry 5.0 Among Workers in High-Tech Manufacturing Companies: Questionnaire Development and Pilot Evaluation Study
JMIR Hum Factors 2026;13:e83754
URL: https://humanfactors.jmir.org/2026/1/e83754
doi:10.2196/83754

© Giulia Bassi, Angelo Valente, Serena Tassoni, Silvia Salcuni. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 28.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic

XSL•FO
RenderX

information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Feasibility of the "OA Coach" Mobile App to Support Individuals with Osteoarthritis: Development and Usability Testing

Vicky Duong[1], PhD; Naomi Bloul[1], BSc (Hons); Jocelyn Bowden[1], PhD; Karen Bracken[1], PhD; Kate Bryce[1], MHSc; Leticia Deveza[1,2], PhD; Jillian Eyles[1], PhD; Abdolhay Farivar[1,2], MD; Robin Huang[3], PhD; Sarah Kobayashi[1], PhD; Na Liu[4], PhD; Carin Pratt[5], MAppSc; Charlotte Strong[5], BSc (Hons); Venkatesha Venkatesha[6], PhD; Shirley Yu[1,2], PhD; David J Hunter[1,2], PhD

[1]Sydney Musculoskeletal Research Centre, The Kolling Institute, The University of Sydney, Level 10 Kolling Building, Reserve Road, St Leonards, New South Wales, Australia

[2]Department of Rheumatology, Royal North Shore Hospital, St Leonards, Australia

[3]Vulsen, Sydney, Australia

[4]Business Information Systems, The University of Sydney Business School, The University of Sydney, Sydney, Australia

[5]Department of Physiotherapy, Royal North Shore Hospital, St Leonards, Australia

[6]Research Directorate, Northern Sydney Local Health District, Royal North Shore Hospital, St Leonards, Australia

Corresponding Author:
Vicky Duong, PhD
Sydney Musculoskeletal Research Centre, The Kolling Institute, The University of Sydney, Level 10 Kolling Building, Reserve Road, St Leonards, New South Wales, Australia

Abstract

Background: The OA Coach mobile app was developed to support individuals with knee osteoarthritis in self-managing their condition. The app aims to fill a current gap in the osteoarthritis mobile app field by combining key features such as symptom tracking, objective activity tracking, educational modules, and encouragement notifications underpinned by behavior change theory.

Objective: The aim of this study was to describe the development of the OA Coach mobile app and assess its usability in a 6-week feasibility study.

Methods: The app was designed in consultation with consumers, rheumatologists, physiotherapists, and osteoarthritis researchers. The app prototype contained four screens: (1) a home screen to track goals and activities, (2) a progress page, (3) a learning page with self-directed modules, and (4) an inbox for communication with the study team. For the feasibility study, 30 participants were recruited between March and April 2024 from a database of osteoarthritis trial participants or through the Osteoarthritis Chronic Care Program at Royal North Shore Hospital, Sydney, Australia. Participants were eligible if they were aged 45 years or older, had knee pain ≥4 on an 11-point numerical pain rating scale and knee stiffness lasting <30 minutes in duration, or stiffness >30 minutes and diagnosed with knee osteoarthritis through a health care provider or radiographs. Participants were provided access to the app and asked to interact with it daily for 6 weeks. Outcomes were assessed through online questionnaires or through mobile app data. The primary outcome was usability, assessed using the mHealth App Usability Questionnaire (MAUQ). Secondary outcomes included computer self-efficacy and osteoarthritis knowledge. The quantitative data were summarized descriptively. Qualitative feedback was collected through open-ended survey responses and discussed within the research team to improve the app.

Results: A total of 30 participants completed the study. There was a 1:1 ratio of male to female participants, with an average age of 66.9 (SD 9.1) years and a mean pain level of 6.0 (IQR 5.0-6.8) on an 11-point numerical pain rating scale. Twenty-nine responses from the MAUQ were available for analysis. Most statements scored >5 out of 7 ("somewhat agree"), indicating that the app was easy to use. The mean satisfaction score for the app on the MAUQ was 4.7 (SD 2.0) out of 7. Qualitative feedback from participants indicated the need for clear instructions on how to use and navigate the app, improved structure and integration of the exercise program, and improved tailoring of osteoarthritis education and support.

Conclusions: Overall, the OA Coach app was well accepted by participants. Based on participant feedback, the app will be revised to improve aspects of clarity, ease of use, and personalization. The updated app will be tested against other methods of care delivery in a randomized controlled trial.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12623001223628; https://tinyurl.com/2rzk7ftk

(JMIR Hum Factors 2026;13:e84857) doi:10.2196/84857

XSL•FO
RenderX

KEYWORDS

mobile app; osteoarthritis; digital health; physical activity; self-management

Introduction

Osteoarthritis is the most common form of arthritis and a leading cause of disability worldwide. With the current population growth and aging, the number of Australians living with osteoarthritis is expected to increase, impacting over 3.11 million by the year 2040 [1]. There is no cure for osteoarthritis, and the current management strategies focus on alleviating symptoms. The core treatments for osteoarthritis include education for self-management, physical activity, and exercise [2,3]. These core treatments are appropriate for all individuals with osteoarthritis, regardless of the severity of their symptoms [2].

Many people with osteoarthritis do not receive appropriate care [4]. For example, surgical referral is often prioritized over nondrug treatments, despite conservative, nondrug care being strongly recommended in international guidelines for osteoarthritis management [4-6]. The use of digital health, combined with behavior change techniques (BCTs), may help to improve access to appropriate osteoarthritis care. BCTs (eg, feedback, self-monitoring, and reinforcement) are observable, replicable, and irreducible components of interventions designed to alter or redirect causal processes that regulate behavior [7]. Physical activity interventions that use BCTs are more effective in chronic disease populations [8], and physical activity interventions that include BCTs through digital health have been shown to increase physical activity in people with osteoarthritis in the short term [9]. Digital health is broadly defined as the use of technologies such as wearables, mobile phones or apps, and websites for the treatment, prevention, and maintenance of health [10]. The World Health Organization has identified digital health as a global priority. Greater adoption of digital health initiatives will support accessible, equitable, and universal access to quality health services, as well as strengthen and scale up health promotion, disease prevention, management, and rehabilitation [11]. The COVID-19 pandemic has further accelerated the use of these technologies, and digital health is being increasingly used for the management of knee osteoarthritis [12-16]. Compared to traditional care, the use of digital health care has provided similar improvements in pain and function in osteoarthritis and back pain populations (systematic review of 23 studies, n=4994 participants) [17]. Our research group also found similar results in individuals following knee replacement (systematic review of 17 studies, n=2188 participants) and in a recent randomized controlled trial (n=102 participants) where participants received a combination of digital interventions (exercise mobile app, activity tracker, and online health coaching) [15,18]. In Australia, the majority of households have access to the internet (88%), even in remote and very remote settings (77%). Of those with internet access, 91% have access to a laptop, desktop, or smartphone [19], highlighting the potential to scale up effective digital interventions for osteoarthritis management. Access to evidence-based osteoarthritis care through a digital mobile health app could be a potential solution for low-cost and accessible care.

Mobile apps designed to support the management of osteoarthritis exist; however, no osteoarthritis-specific app has measured or monitored physical activity objectively [20]. As increasing and maintaining physical activity levels is a core aspect of osteoarthritis management, it is important that reliable measures of physical activity are incorporated into these apps. Previous studies have suggested that participants tend to inaccurately report their adherence to exercise and physical activity during home-based programs [21,22]. A recent systematic review also found that existing mobile apps for osteoarthritis management lack quality information, structured exercise programs, and scientific testing [20].

We developed a mobile app called OA Coach to support individuals with osteoarthritis in self-managing their condition. OA Coach aims to fill a current gap in the osteoarthritis mobile health field by combining key features such as symptom tracking, objective activity tracking (step count and sleep), tailored educational modules, and encouragement notifications, all underpinned by behavior change theory and developed with experts in osteoarthritis. The primary aim of this paper was to outline the development of the OA Coach mobile app and examine the usability of the app in participants with knee osteoarthritis in a 6-week feasibility study. Secondary outcomes include changes in knee symptoms, physical activity, sleep, feelings of depression, anxiety, stress, and osteoarthritis knowledge after app use.

Methods

Mobile App Prototype Development

The OA Coach app was co-designed with a multidisciplinary team including osteoarthritis researchers, rheumatologists, physiotherapists, and consumers living with osteoarthritis. The app prototype was developed by Vulsen, a developer with expertise in developing mobile apps for chronic disease management [23,24]. The study team (VD, NB, JE, SK, DJH) tested the mobile app prototype for approximately 2 weeks. A 1-hour co-design workshop was also held with 6 consumers with lived experience of osteoarthritis to gather initial feedback regarding the features and visual aspects of the app prototype. Co-design participants were given access to the mobile app to download onto their personal devices 1 week prior to the workshop. In the workshop, a brief overview of the app was presented, followed by the participants mapping likes or dislikes and potential areas for improvement for the app and a group discussion. Following alpha-testing, issues with the functionality or content of the app were relayed to the developer for resolution prior to the 6-week feasibility study.

Mobile App Features and Functionality

Details on the main features of the app are summarized in Textbox 1. An overview of the Home, Progress, and Learning pages is displayed in Figure 1. The app's log-in page required

XSL•FO
RenderX

participants to enter their date of birth and answer questions about their knee pain (pain severity, stiffness duration). This information was used to make a clinical diagnosis based on the National Institute for Health and Care Excellence guidelines [25], which then determined access to the app. Following this, participants were prompted to complete questions about their weight, height, sleep, mood, and current activity levels in order to trigger corresponding educational modules targeted to the users' needs. For example, if a participant's BMI was above a healthy range or the questionnaire responses indicated poor sleep, participants would be assigned the "Weight Management" or "Sleep" modules, respectively.

Textbox 1. Functionalities associated with each feature of the OA Coach app prototype.

Home screen

- Participants prospectively set personal goals (eg, related to physical activity, exercise, meditation) during the registration process of the app, which are displayed as a checklist on the home screen under Goals.
- Questionnaires for mood and sleep, and logging of pain scores, will also appear under Goals.
- Fitbit data, including the number of steps, hours slept, and number of calories burned, are displayed.

Progress

- Weight is tracked in kilograms. BMI is calculated with the user's height inputted at registration and displayed at the bottom of the graph. If the user's BMI is between 18.5 and 24.9 kg/m^2, it will display "Your weight is within a healthy range." Otherwise, the graph will say "Let's aim to reach a healthy weight range between X kg and X kg."
- Knee pain over the past 24 hours is tracked on a 0 - 10 numerical rating scale, where 0 represents no pain and 10 represents extreme pain. The average pain over the past 7 days is displayed at the bottom of the graph.
- Sleep and step count data are automatically populated with Fitbit data.

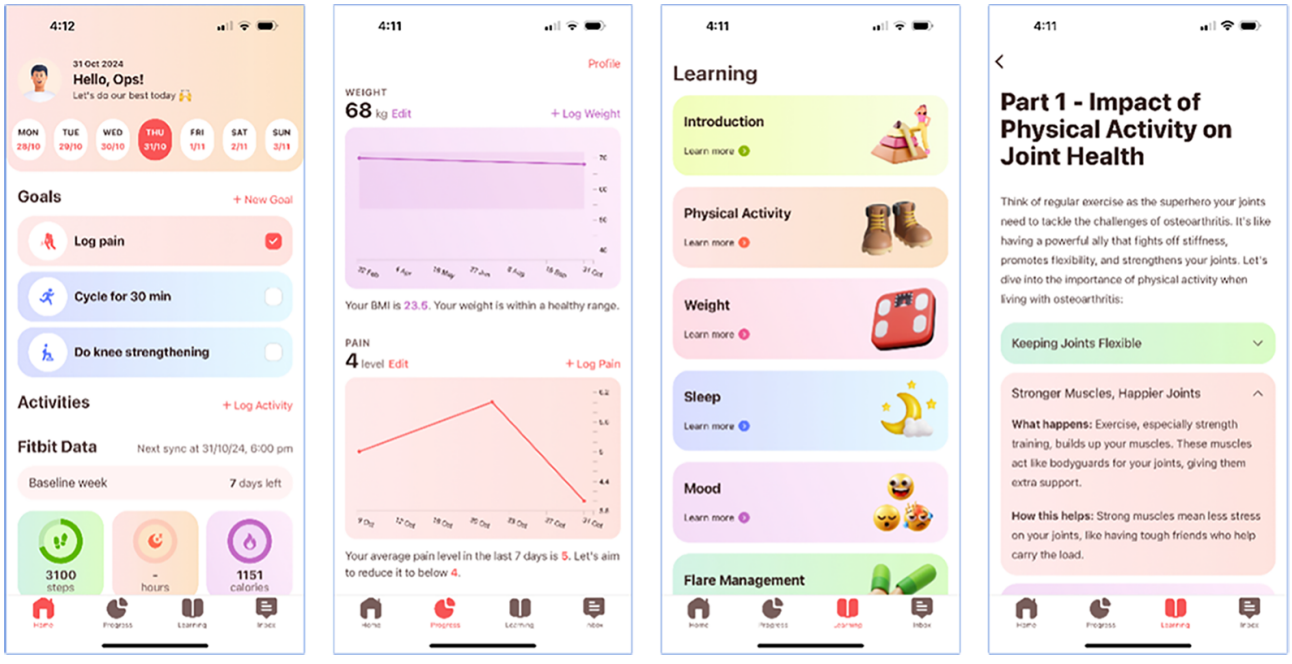
Learning

- Seven learning modules related to the core aspects of osteoarthritis management are displayed (introduction to osteoarthritis, physical activity, weight, sleep, mood, osteoarthritis flare management, and exercise). Each module, apart from physical activity, has 3 units: "Educational resources: Learn the basics, Self-guided help: Digging deeper, and Expert help: Getting more help."
- If participants experience an episode of a pain flare, they will be prompted by a notification to complete the Flare Management module. Similarly, if they have an episode of low mood or sleep, they will be prompted to complete the Mood or Sleep modules, respectively. These modules will appear initially as a pop-up notification and then on participants' Home page in the Goals section.

Messaging

- Initially set up for study staff to send one-way messages to participants related to their progress. The messages are sent from a custom dashboard and appear for participants in the messaging tab, which displays messages in a format similar to text messages on a mobile phone. Participants will also receive a push notification on their phone home screen alerting them to the new message.

Figure 1. Overview of the OA Coach mobile app's Home, Progress, and Learning pages.



XSL•FO
RenderX

Exercise plans for the app were displayed under the "Learning" page. The page included a brief introduction to knee-strengthening exercises, including the benefits of exercise, how to complete the exercise programs (warm up, joint specific exercises, cardiovascular endurance exercises [if possible], warm down), how to assess difficulty and progression, and frequency of program. Separate plans for individuals with good mobility, moderate mobility, and limited mobility were presented, with 3 progressions within each plan. Each exercise plan included instructional photos and descriptions as well as suggested dosage. Participants were free to choose a plan which best aligned with their current mobility status.

Participants were also provided with an activity tracker to wear for the duration of the study. The app requires that participants wear the Fitbit tracker to collect baseline data on sleep and physical activity levels in order to generate appropriate goals. The step and sleep count data are displayed on the "Home" and "Progress" pages of the app. The generated goals are displayed under the "Goals" heading on the "Home" page.

Behavior Change Theory

Behavior change theory was used to help motivate users to increase their physical activity and sleep levels, as well as complete tasks such as logging pain scores or completing surveys. BCTs, such as goal setting, feedback on behaviors, and graded tasks, were integrated into the setup process and push notifications. Participants were asked to set goals during the initial setup process of the app, where they were asked about their current exercise and physical activity levels in a questionnaire. Their responses were used to create weekly goals. For example, if a participant enjoyed playing a sport once a week, a goal would automatically be created to increase the frequency of the sport to align with the World Health Organization physical activity guidelines for older adults [26]. Feedback on behaviors, as well as graded tasks, was provided to participants as push notifications for meeting their step goal at the end of the day. Self-monitoring of pain levels and body weight was also encouraged through notifications and logged in the "Progress" page of the app, along with sleep and step count data. Further details of various OA Coach features mapped to the Behavior Change Taxonomy are outlined in Multimedia Appendix 1.

Feasibility Study Design

The feasibility study involved providing the OA Coach mobile app to study participants for 6 weeks to assess the usability of the developed app. The protocol for the feasibility study was registered on the Australian and New Zealand Clinical Trials Registry before commencement (registry number: 12623001223628). The results of this study are reported in accordance with the CONSORT (Consolidated Standards of Reporting Trials) checklist (Checklist 1).

Ethical Considerations

Ethical approval was granted by the Northern Sydney Local Health District Human Research Ethics Committee (2023/ETH02288). Participant eligibility was assessed using an online survey, and participants who passed the screening survey provided electronic consent through a Research Electronic Data Capture e-consent form.

Participants

Potentially suitable participants were identified by rheumatologists and physiotherapists within the Osteoarthritis Chronic Care Program at Royal North Shore Hospital, Sydney, Australia, and were provided with a link to the eligibility screening survey. Participants from past clinical trials who had consented to be contacted about future research studies were also emailed and invited to participate in the study.

Individuals were eligible for the study if they were aged 45 years or older, had knee pain ≥4 out of 10 on an 11-point Numerical Pain Rating Scale (NPRS; where 0 represents "no pain" and 10 represents "worst imaginable pain"), and experienced knee stiffness lasting no longer than 30 minutes after rising from bed in the morning. Participants were also eligible to participate in the study if they had knee stiffness >30 minutes but had been diagnosed with knee osteoarthritis through radiographs or a health care practitioner. This inclusion criterion was added as an amendment to the study, as a number of participants with a confirmed diagnosis of osteoarthritis were being excluded due to stiffness slightly exceeding 30 minutes. These participants were contacted either from either our database of previous participants or were recruited from the Osteoarthritis Chronic Care Program at Royal North Shore Hospital thus already having a confirmed diagnosis of osteoarthritis. Participants also required ownership of a smartphone with internet access in order to download the OA Coach app. The exclusion criteria for the study were as follows: a diagnosis of inflammatory arthritis, such as rheumatoid arthritis or gout, and any comorbidities or medical conditions that would prevent participation in the physical activity component of the app.

Intervention

All participants received a Fitbit tracker (Fitbit Inspire 2) in the post and instructions on how to download the OA Coach app on their personal smartphones and link the Fitbit tracker. The OA Coach app is freely available on app stores; however, its access remained password-protected during the evaluation period. Participants were instructed to wear the Fitbit tracker for a baseline period of 1 week and were then asked to interact with the app daily for 6 weeks. Participants were instructed to contact the study coordinator through email if they experienced any technical issues in setting up the OA Coach app or Fitbit tracker. The study coordinator aimed to resolve any technical issues through email, prior to contacting participants by phone.

Outcomes

Three sets of data were obtained from the study including baseline and 6-week online questionnaires, data from the mobile app (eg, step count, hours slept, the 21-item Depression, Anxiety and Stress Scale [DASS-21] and Pittsburgh Sleep Quality Index [PSQI] questionnaires), and qualitative feedback from surveys at 6 weeks.

Primary Outcome

The mHealth App Usability Questionnaire (MAUQ) for Standalone mHealth Apps Used by Patients was used to assess

XSL•FO
RenderX

the usability of the OA Coach app for the primary outcome [27]. The questionnaire consists of 18 statements, and participants were asked to rate their agreement with each statement on a 7-point Likert scale, ranging from strongly disagree to strongly agree [27]. To determine the usability of the app, the total score was calculated and averaged across responses, with higher scores representing higher usability.

Secondary Outcomes

The Modified Computer Self-Efficacy Scale (mCSES) and the Osteoarthritis Knowledge Scale (OAKS) were used to assess technology self-efficacy and osteoarthritis knowledge, respectively, at baseline and 6 weeks through surveys. The mCSES asks respondents to rate their confidence using a new technology on a scale from 1 to 10, where 1 represents "not confident at all" and 10 "totally confident" [28]. The OAKS is an 11-item scale assessing important knee osteoarthritis knowledge, with scores ranging from 11 to 55, with higher scores indicating greater knowledge about osteoarthritis [29].

Secondary outcomes collected through the app included changes in knee pain (NRS), BMI, mood, sleep quality and disturbances, physical activity (average step count over the last 7 days), and time asleep (average sleep duration over the last 7 days). The DASS-21 was used to measure the emotional states of depression, anxiety, and stress. The questionnaire consists of 21 statements, with 7 statements pertaining to each subscale. Participants were asked to rate their agreement with each statement from 0 to 3 (where 0 indicates "Never," and 3 indicates "Almost always") [30]. Scores for each subscale were summed and multiplied by 2 for interpretation. The PSQI is a 19-item self-reported questionnaire assessing various aspects of sleep quality (eg, duration, latency, disturbances) over the past month. Scores for each question range from 0 to 3, with higher scores indicating more acute sleep disturbances [31].

Knee pain (NRS) and body weight (kg) were self-reported by participants. Physical activity and time asleep were automatically collected from participants' Fitbit devices, which synced to the mobile app.

Qualitative Feedback

Participants were also asked to provide any feedback regarding the app through open text fields with the option of including their contact details should they wish to discuss the feedback in person with a study team member.

The outcome measures collected in this study were exploratory in nature and were not used for progression, modification, or stopping decisions within the study protocol.

Sample Size

As this study was a feasibility study, no formal sample size calculations were conducted. We aimed to recruit a total of 30 participants for this study, which is an acceptable size for a feasibility study [32].

Statistical Analysis

Participant baseline demographics and clinical characteristics were expressed using means and SDs or medians and IQRs if not normally distributed. The primary outcome (MAUQ) was summarized descriptively. The secondary outcome measures at baseline and 6 weeks were summarized using median and IQR. The change scores (baseline minus 6 weeks) were summarized using median and IQR. Analyses were conducted using STATA (version 18.0). Qualitative data from participant feedback were collated into an Excel spreadsheet, and each comment was mapped to a general category (eg, App download, Fitbit integration, goals, instructions, learning materials, notifications, design, or other) by 1 researcher (NB). These were reviewed by a second researcher (VD) and suggestions on how to improve the app were discussed and documented. The suggestions were then reviewed within the wider study team, including the app developer. Changes that the researchers or developer deemed reasonable and appropriate informed the next iteration of the OA Coach mobile app. Changes that required additional resources and funding were temporarily put on hold.

Results

Participant Demographics

A total of 30 participants were recruited between March 2023 and April 2023. There was an even split of female and male participants (n=15 per sex), and participants had an average age of 66.9 (SD 9.1) years and a median pain level of 6.0 (IQR 5.0-6.8) on an 11-point NRS. The participant baseline demographics are summarized in Table 1. Due to technical issues, there were some participants who did not have complete step, and sleep data at baseline and demographics in Table 1 are based on a complete case analysis.

XSL•FO
RenderX

Table . Participant baseline demographics.

Characteristics	Values (n=30)
Age (y), mean (SD)	66.9 (9.1)
Female, n (%)	15 (50)
NRS[a] pain, median (IQR)	6.0 (5.0-6.8)
Technology self-efficacy (mCSES)[b], median (IQR)	89.0 (74.0-99.0)
Osteoarthritis knowledge (OAKS)[c], median (IQR)	41.0 (38.0-49.0)
BMI (kg/m^2), median (IQR)	26.7(24.4-31.2)
Depression (DASS-21)[d], median (IQR)	6.0[e] (2.0-8.0)
Anxiety (DASS-21), median (IQR)	4.0[e] (0.5-7.0)
Stress (DASS-21), median (IQR)	10.0[e] (4.0-14.0)
Sleep quality and disturbances (PSQI)[f], median (IQR)	6.5 (5.0-10.3)
Physical activity[g], median (IQR)	8073[h] (5636-10642)
Time asleep[i], median (IQR)	7.0[j] (5.6-7.7)

[a]NRS: Numerical Rating Scale.
[b]mCSES: Modified Computer Self-Efficacy Scale.
[c]OAKS: Osteoarthritis Knowledge Scale.
[d]DASS-21: 21-item Depression, Anxiety, and Stress Scale.
[e]Twenty-seven participants.
[f]PSQI: Pittsburgh Sleep Quality Index.
[g]Average step count over the last 7 days.
[h]Twenty-nine participants.
[i]Average sleep duration in hours over the last 7 days.
[j]Twenty-four participants.

Primary Outcome

Overall, participants found the app usable, with the majority of scores greater than 5 out of 7 on the MAUQ. Most participants found the app easy to use (mean 5.8, SD 1.5), easy to learn how to use (mean 6.1, SD 1.2), and were satisfied with the amount of time involved in using the app (mean 6.0, SD 0.9). The lowest scoring statements were regarding improving access to health care services (mean 3.6, SD 1.8) and the app functionalities and capabilities (mean 4.1, SD 1.9), which are in alignment with the qualitative data collected. The results of the MAUQ statements are summarized in Table 2.

XSL•FO
RenderX

Table . Results of the mHealth App Usability Questionnaire (MAUQ).

Statement[a]	Mean (SD)[b]	Min-max
The app was easy to use.	5.8 (1.5)	2-7
It was easy for me to learn to use the app.	6.1 (1.2)	2-7
The navigation was consistent when moving between screens.	5.5 (1.5)	1-7
The interface of the app allowed me to use all the functions (such as entering information, responding to reminders, viewing information) offered by the app.	5.1 (1.9)	2-7
Whenever I made a mistake using the app, I could recover easily and quickly.	5.2 (1.7)	0-7
I like the interface of the app.	5.1 (1.7)	2-7
The information in the app was well organized, so I could easily find the information I needed.	5.2 (1.7)	2-7
The app adequately acknowledged and provided information to let me know the progress of my action.	5.3 (1.6)	2-7
I feel comfortable using this app in social settings.	5.8 (1.6)	1-7
The amount of time involved in using this app has been fitting for me.	6.0 (0.9)	4-7
I would use this app again.	5.2 (1.7)	2-7
Overall, I am satisfied with this app.	4.7 (2.0)	0-7
The app would be useful for my health and well-being.	5.4 (1.8)	0-7
The app improved my access to health care services.	3.6 (1.8)	0-7
The app helped me manage my health effectively.	5.1 (1.2)	2-7
This app has all the functions and capabilities I expected it to have.	4.1 (1.9)	1-7
I could use the app even when the internet connection was poor or not available.	4.3 (1.9)	0-7
This mHealth[c] app provides an acceptable way to receive health care services, such as accessing educational materials, tracking my own activities, and performing self-assessment.	5.3 (1.2)	3-7

[a]n=29; 1 participant had missing data.

[b]Scores range from 1 to 7, where 1 indicates strongly disagree, 2 indicates disagree, 3 indicates somewhat disagree, 4 indicates neither agree nor disagree, 5 indicates somewhat agree, 6 indicates agree, and 7 indicates strongly agree.

[c]mHealth: mobile health.

Secondary Outcomes

The NRS median pain reduced from a median of 5.0 (IQR 4.0 - 6.0) at baseline to 4.0 (IQR 2.0 - 5.0) at 6 weeks, giving a within-patient median reduction of 1 point (IQR 0.0 - 3.0). The modified Computer Self-Efficacy Scale (mCSES) moved in the opposite direction: the median score declined from 89 (IQR 74 - 100) to 81 (IQR 67 - 93), reflecting a median change (reduction) score of 2.5 points (IQR –1.0 to 7.0). The OAKS score did not indicate a substantial change in knowledge (median change of 0.0 points, IQR –2.0 to 3.0).

The magnitude of change in BMI was very small—change from a median 26.7 (IQR 24.4 - 31.2) to 26.6 (IQR 24.5 - 29.1)—corresponding to a median within-patient change of 0.02 (IQR –0.15 to 0.05). There was a small improvement in the psychological distress measure. The median depression scores falling from 6.0 (IQR 2.0 - 8.0) to 4.0 (IQR 2.0 - 8.0), anxiety from 4.0 (IQR 0.0 - 6.0) to 2.0 (IQR 2.0 - 6.0), and stress from 10.0 (IQR 4.0 - 16.0) to 6.0 (IQR 2.0 - 12.0). Sleep quality improved slightly, with the median PSQI score decreasing from 6.5 (IQR 5.0 - 10.5) to 6.0 (IQR 4.0 - 9.0).

The median 7-day step count increased from 8073 (IQR 5392 - 10,726) to 8616 (IQR 6028 - 10,057), a median gain of

XSL•FO
RenderX

446 steps (IQR –1433 to 2,322), while the average nightly sleep duration remained almost equal at median 7.02 (IQR 5.56 - 7.68) hours and 6.90 (IQR 5.40 - 7.65) hours. These changes are minimal and may be partially explained by the cohort that was recruited. Most participants were either previous participants of our clinical trials or had been receiving care from an osteoarthritis management team and therefore were more likely to be knowledgeable about their condition, already be engaged in some form of physical activity and have sufficient sleep hygiene. Secondary outcomes are summarized in Table 3.

Table . Secondary outcomes.

Measure	Baseline, median (IQR)	Six weeks, median (IQR)	Change (baseline to 6 weeks), median (IQR)
NRS[a] pain	5.0 (4.0 to 6.0)	4.0 (2.0 to 5.0)	1 (0.0 to 3.0)
Technology self-efficacy (mCSES)[b]	89.0 (74.0 to 100.0)	81.0 (67.0 to 93.0)	2.5 (–1.0 to 7.0)
Osteoarthritis knowledge (OAKS)[c]	41.0 (38.0 to 49.0)	42 (35 to 48.0)	0.0 (–2.0 to 3.0)
BMI (kg/m^2)	26.7 (24.4 to 31.2)	26.6 (24.5 to 29.1)	–0.02 (–0.15 to 0.05)
Depression (DASS-21)[d]	6.0 (2.0 to 8.0)	4.0 (2.0 to 8.0)	0.0 (0.0 to 6.0)
Anxiety (DASS-21)	4.0 (0.0 to 6.0)	2.0 (2.0 to 6.0)	0.0 (–2.0 to 2.0)
Stress (DASS-21)	10.0 (4.0 to 14.0)	6.0 (2.0 to 12.0)	2.0 (–6.0 to 6.0)
Sleep quality and disturbances (PSQI)[e]	6.5 (5.0 to 10.5)	6.0 (4.0 to 9.0)	1.0 (0.0 to 2.0)
Physical activity[f]	8073 (5392 to 10726)	8616 (6028 to 10057)	446 (–1433 to 2322)
Time asleep[g]	7.02 (5.56 to 7.68)	6.9 (5.4 to 7.65)	0.15 (–0.58 to 0.57)

[a]NRS: Numerical Rating Scale.

[b]mCSES: Modified Computer Self-Efficacy Scale.

[c]OAKS: Osteoarthritis Knowledge Scale.

[d]DASS-21: 21-item Depression, Anxiety, and Stress Scale.

[e]PSQI: Pittsburgh Sleep Quality Index.

[f]Average step count over the last 7 days

[g]Average sleep duration in hours over the last 7 days

Overall, when comparing outcomes at 6 weeks to baseline, there was a small reduction in pain and psychological distress; slight improvements in sleep quality and physical activity; and stability in knowledge, BMI, and sleep duration. These outcomes were expected in a small, uncontrolled feasibility study that was not designed to detect statistically significant changes within the group.

Qualitative Feedback

The qualitative feedback largely focused on usability and suggestions for improving the app, such as better integration of the exercise programs and improved personalization of goals and activity tracking. Additionally, participants highlighted the need for improved onboarding support. Participants also noted confusion over the labeling of certain features. The main areas for improvement and updates made to the next iteration of the OA Coach app are outlined in Table 4.

XSL•FO
RenderX

Table . Summary of qualitative feedback from participants at 6 weeks.

Page	Areas lacking and suggestions for improvement	Updates made to OA Coach app
Goals	• Not possible to set a one-off goal in the app, only recurring goals • Tying goals to specific time was not appropriate • Unable to find DASS-21[a] and PSQI[b] questionnaires	• A new "Activities" section was created to log one-off activities. The goals section will remain for recurring goals. • Clarified that the time next to the goal was for the time for the reminder to be sent. The time has been removed from under the goal so as not to cause confusion but can be edited by selecting the icon next to the goal. • DASS-21 and PSQI questionnaires were renamed to "Sleep" and "Mood" questionnaires to avoid confusion. Clarified instructions on how to complete the questionnaires and goals to fill questionnaires will repeat daily until they have been completed.
Instructions	• Setting up Fitbit goals and duration of exercise activities were confusing	• Clearer instructions and explanatory text were provided. An introductory video was created to explain all features of the app.
Learning	• Some links in the article did not work • Could not find the knee strengthening exercises	• Links in articles were checked and updated. • An introductory video was created to explain how to use the app and features.
Messages	• Feedback notifications were too generic	• Notifications were modified, and changes were made underpinned by behavior change theory. New notifications and messages underwent comprehensive co-design and survey study.

[a]DASS-21: 21-item Depression, Anxiety, and Stress Scale.

[b]PSQI: Pittsburgh Sleep Quality Index.

Discussion

The OA Coach mobile app is designed to provide education to people with osteoarthritis to help them self-manage their condition. The app aims to address the current gaps in mobile apps for the management of osteoarthritis, focuses on all aspects of osteoarthritis management (eg, pain flares, weight management, mood, sleep, physical activity, and exercise), and has been co-designed with end user input. Overall, the usability of the OA Coach mobile app prototype was deemed acceptable by participants when assessed using the MAUQ. The lowest scoring item on the MAUQ was access to health care services, and although the app provides recommendations for how to access services in the learning modules, including the types of support available with relevant links, improvements to the app could be made to make this more explicit, for example, including a directory of health care professionals or service providers who manage different aspects of osteoarthritis and their contact details. The secondary outcomes demonstrated small reductions in pain and psychological distress and slight improvements in sleep quality and physical activity (step count). Interestingly, technology self-efficacy worsened. However, as this was a feasibility study only, the sample size was not powered to reliably detect a true effect. The 6-week duration of the study was also likely too short to detect a change in outcomes.

From the qualitative feedback, study participants recommended improving the clarity and personalization aspects of the OA Coach mobile app. As a result, the app was updated to include more instructions and clarification text within different features of the app. Updates to the app were made, including the addition of an introductory instructional video and an improved interface for the exercise programs and "Inbox." A total of 80 notifications sent through the app also underwent an extensive co-design process and were refined. An additional 12 educational messages were also co-designed for the new "Inbox" feature of the app [33].

A major strength of the study is that the OA Coach mobile app was co-designed and tested with people with lived experience of osteoarthritis. There was also input from researchers and health professionals with various backgrounds and expertise. The developer, Vulsen, is familiar with the requirements for data privacy in Australia, and their mobile apps have been approved for use in collecting patient data within local health districts within Sydney, Australia. Another major strength of the study is that we used objectively measured physical activity and sleep to provide tailored recommendations in the app. We also used validated questionnaires for mood and sleep, allowing the app to tailor learning modules on these topics to participants. The balanced distribution of female to male participants was also a strength of our study, as there is usually a greater proportion of female participants in osteoarthritis clinical trials [34].

Limitations of our study include selection bias—we were only able to include participants in the study who owned a

XSL•FO
RenderX

smartphone and therefore may be more familiar with the use of mobile apps. The usability scores of the app may be biased upward due to the self-selection of participants who were more technology-literate. The participants in our trial were generally older as well and living in metropolitan Sydney, which limits the generalizability of the results. Participants were also recruited from our database of previous clinical trial participants or from the Osteoarthritis Chronic Care Program. As a result, they were more likely to have prior knowledge about their condition and to already be engaging in some form of physical activity or exercise. At baseline, participants were already achieving an adequate level of physical activity as measured by step count and were sleeping the recommended number of hours. This high starting point limits the potential for further improvement, which may partially explain why changes in secondary outcomes were minimal. The study had a short follow-up duration of only 6 weeks, and a longer period of follow-up (eg, 12 wk) would have been ideal to evaluate long-term outcomes and behavior change. For the purposes of this study, we did not analyze the app meta-data, such as time spent interacting with the app and completing learning modules; however, these data may be used to refine future versions of the app. Finally, the educational content featured in our app is mainly limited to knee osteoarthritis; future versions of the app may also target other commonly affected joints, such as the base of the thumb or the hip joint.

The OA Coach mobile app was theory-driven and informed by behavior change theory and best-evidenced osteoarthritis care. The components of the app extend beyond aggregating features and are intended to enhance users' capability, motivation, and self-efficacy. The OA Coach app combines the core aspects of osteoarthritis management, alongside objective monitoring of activity, symptoms, and behaviors. To our knowledge, this is the first mobile app developed for knee osteoarthritis management in the Australian market, which has the ability to measure physical activity objectively using a fitness tracker. Our methodology was consistent with a systematic review published in 2019 by Choi et al [35] that provided a framework for developing mobile apps for osteoarthritis management, particularly apps that have the ability to track pain and symptoms, provide instructional audio/video of exercise regimes, and set goals to maximize adherence. A more recent review by Hensley et al [20] in 2023 also found a lack of quality information, structured exercise programs, and health information protection in the mobile apps evaluated (n=14). OA Coach addresses many of the features missing in previous mobile apps developed for osteoarthritis management. The app will also undergo effectiveness testing in a 3-armed randomized controlled trial before being launched in public app stores [36]. This will also include in-depth interviews with participants to discuss barriers and facilitators to using the app. This will ensure that the final version of the app is effective at managing osteoarthritis-related symptoms. Future updates to the OA Coach mobile app functionality will be informed by consumer feedback as well as data analytics to ensure that the app continues to meet the evolving needs, preferences, and expectations of people living with osteoarthritis.

The results from this feasibility study provide evidence to support the use of the OA Coach mobile app by encouraging patients to self-manage their condition by providing educational resources; addressing comorbidities, such as sleep and mood changes that commonly coexist with osteoarthritis; and encouraging them to be more physically active. The feedback from the feasibility study has been used to refine the current version of the OA Coach app, and it is currently being tested in a randomized controlled trial assessing 3 different models of care. Further development of the app may involve the use of generative artificial intelligence to further tailor notifications to provide personalized recommendations, and the ability to sync with Apple Health or Google Health to track activity levels without the use of an activity tracker.

Acknowledgments

We would like to thank the participants of the study.

No generative artificial intelligence tools were used at any stage in preparation of this manuscript.

Funding

Lenity Australia provided funding for the prototype of the OA Coach mobile app.

Authors' Contributions

VD and DJH conceived the study and all authors contributed to the protocol. RH developed the app with input from VD, NB and DJH. VD and NB conducted the study, with recruitment assistance from LD, AF, CP, CS, SY and DJH. VD and VV conducted the statistical analyses. VD drafted the manuscript, and all authors provided input to the manuscript preparation and approved the final version of the manuscript.

Conflicts of Interest

VD is supported by a Lenity Foundation Fellowship. JE is supported by a National Health and Medical Research Council Investigator Grant (2025504). DJH is the editor of the osteoarthritis section for UpToDate and co-Editor in Chief of Osteoarthritis and Cartilage. DJH provides consulting advice on scientific advisory boards for Haleon, Naia Health, Novartis, Tissuegene, Lilly, Doron, Sanofi, Formation Bio. All other authors declare that there are no conflicts of interest in relation to this work.

XSL•FO
RenderX

Multimedia Appendix 1
Mapping of OA Coach features to Behavior Change Taxonomy.
[DOCX File, 25 KB - humanfactors_v13i1e84857_app1.docx]

Checklist 1
CONSORT Checklist
[PDF File, 220 KB - humanfactors_v13i1e84857_app2.pdf]

References

1. Ackerman IN, Gorelik A, Berkovic D, Buchbinder R. The projected burden of arthritis among adults and children in Australia to the year 2040: a population-level forecasting study. Lancet Rheumatol 2025 Mar;7(3):e187-e196. [doi: 10.1016/S2665-9913(24)00247-9] [Medline: 39647490]
2. Bowden JL, Hunter DJ, Deveza LA, et al. Core and adjunctive interventions for osteoarthritis: efficacy and models for implementation. Nat Rev Rheumatol 2020 Aug;16(8):434-447. [doi: 10.1038/s41584-020-0447-8] [Medline: 32661322]
3. Duong V, Oo WM, Ding C, Culvenor AG, Hunter DJ. Evaluation and treatment of knee pain: a review. JAMA 2023 Oct 24;330(16):1568-1580. [doi: 10.1001/jama.2023.19675] [Medline: 37874571]
4. Basedow M, Esterman A. Assessing appropriateness of osteoarthritis care using quality indicators: a systematic review. J Eval Clin Pract 2015 Oct;21(5):782-789. [doi: 10.1111/jep.12402] [Medline: 26083547]
5. Kolasinski SL, Neogi T, Hochberg MC, et al. 2019 American College of Rheumatology/Arthritis Foundation guideline for the management of osteoarthritis of the hand, hip, and knee. Arthritis Rheumatol 2020 Feb;72(2):220-233. [doi: 10.1002/art.41142] [Medline: 31908163]
6. Bannuru RR, Osani MC, Vaysbrot EE, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthr Cartil 2019 Nov;27(11):1578-1589. [doi: 10.1016/j.joca.2019.06.011]
7. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
8. Greaves CJ, Sheppard KE, Abraham C, et al. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011 Feb 18;11(1):119. [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
9. Berry K, Salter A, Morris R, James S, Bucci S. Assessing therapeutic alliance in the context of mHealth interventions for mental health problems: development of the Mobile Agnew Relationship Measure (mARM) questionnaire. J Med Internet Res 2018 Apr 19;20(4):e90. [doi: 10.2196/jmir.8252] [Medline: 29674307]
10. Ronquillo JG, Zuckerman DM. Software ‐ related recalls of health information technology and other medical devices: implications for FDA regulation of digital health. Milbank Q 2017 Sep;95(3):535-553. [doi: 10.1111/1468-0009.12278] [Medline: 28895231]
11. Global action plan on physical activity 2018–2030: more active people for a healthier world. : World Health Organization; 2018 URL: https://iris.who.int/server/api/core/bitstreams/33339c9c-3a9f-46d4-9f12-ae9ff0dfdc6a/content [accessed 2026-06-27]
12. Shukla H, Nair SR, Thakker D. Role of telerehabilitation in patients following total knee arthroplasty: evidence from a systematic literature review and meta-analysis. J Telemed Telecare 2017 Feb;23(2):339-346. [doi: 10.1177/1357633X16628996] [Medline: 26843466]
13. Bahadori S, Wainwright TW, Ahmed OH. Smartphone apps for total hip replacement and total knee replacement surgery patients: a systematic review. Disabil Rehabil 2020 Apr;42(7):983-988. [doi: 10.1080/09638288.2018.1514661] [Medline: 30362363]
14. Jiang S, Xiang J, Gao X, Guo K, Liu B. The comparison of telerehabilitation and face-to-face rehabilitation after total knee arthroplasty: a systematic review and meta-analysis. J Telemed Telecare 2018 May;24(4):257-262. [doi: 10.1177/1357633X16686748] [Medline: 28027679]
15. Wang X, Hunter DJ, Vesentini G, Pozzobon D, Ferreira ML. Technology-assisted rehabilitation following total knee or hip replacement for people with osteoarthritis: a systematic review and meta-analysis. BMC Musculoskelet Disord 2019 Nov 3;20(1):506. [doi: 10.1186/s12891-019-2900-x] [Medline: 31679511]
16. Safari R, Jackson J, Sheffield D. Digital self-management interventions for people with osteoarthritis: systematic review with meta-analysis. J Med Internet Res 2020 Jul 20;22(7):e15365. [doi: 10.2196/15365] [Medline: 32706657]
17. O'Brien KM, Hodder RK, Wiggers J, et al. Effectiveness of telephone-based interventions for managing osteoarthritis and spinal pain: a systematic review and meta-analysis. PeerJ 2018;6:e5846. [doi: 10.7717/peerj.5846] [Medline: 30397549]
18. Duong V, Robbins SR, Dennis S, Venkatesha V, Ferreira ML, Hunter DJ. Combined digital interventions for pain reduction in patients undergoing knee replacement: a randomized clinical trial. JAMA Netw Open 2023 Sep 5;6(9):e2333172. [doi: 10.1001/jamanetworkopen.2023.33172] [Medline: 37713201]

XSL•FO
RenderX

19. Household use of information technology. Australian Bureau of Statistics. URL: https://www.abs.gov.au/statistics/industry/technology-and-innovation/household-use-information-technology/2016-17 [accessed 2023-01-30]
20. Hensley CP, Witte MM, Cai J, et al. Assessment of mobile health applications for management of knee and/or hip osteoarthritis using the mobile application rating scale. J Clin Rheumatol 2023 Aug 1;29(5):245-253. [doi: 10.1097/RHU.0000000000001896] [Medline: 36256541]
21. Nicolson PJA, Hinman RS, Wrigley TV, Stratford PW, Bennell KL. Self-reported home exercise adherence: a validity and reliability study using concealed accelerometers. J Orthop Sports Phys Ther 2018 Dec;48(12):943-950. [doi: 10.2519/jospt.2018.8275] [Medline: 30053792]
22. Duong V, Dennis S, Ferreira ML, et al. Correlations between objective and self-reported step count adherence following total knee replacement: a longitudinal repeated-measures cohort study. Physiother Res Int 2022 Oct;27(4):e1966. [doi: 10.1002/pri.1966] [Medline: 35830179]
23. Zwi S, Isautier J, Webster AC, et al. A feasibility study of a best practice health literacy app for Australian adults with chronic kidney disease. PEC Innov 2022 Dec;1:100047. [doi: 10.1016/j.pecinn.2022.100047] [Medline: 37213763]
24. Ahn E, Liu N, Parekh T, et al. A mobile app and dashboard for early detection of infectious disease outbreaks: development study. JMIR Public Health Surveill 2021 Mar 9;7(3):e14837. [doi: 10.2196/14837] [Medline: 33687334]
25. Wood G, Neilson J, Cottrell E, Hoole SP, Guideline Committee. Osteoarthritis in people over 16: diagnosis and management-updated summary of NICE guidance. BMJ 2023 Jan 24;380:24. [doi: 10.1136/bmj.p24] [Medline: 36693668]
26. Bull FC, Al-Ansari SS, Biddle S, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020 Dec;54(24):1451-1462. [doi: 10.1136/bjsports-2020-102955] [Medline: 33239350]
27. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth app usability questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500. [doi: 10.2196/11500] [Medline: 30973342]
28. Laver K, George S, Ratcliffe J, Crotty M. Measuring technology self efficacy: reliability and construct validity of a modified computer self efficacy scale in a clinical rehabilitation setting. Disabil Rehabil 2012;34(3):220-227. [doi: 10.3109/09638288.2011.593682] [Medline: 21958357]
29. Darlow B, Krägeloh C, Abbott JH, et al. The osteoarthritis knowledge scale. Musculoskeletal Care 2023 Jun;21(2):516-526. [doi: 10.1002/msc.1727] [Medline: 36573463]
30. Antony MM, Bieling PJ, Cox BJ, Enns MW, Swinson RP. Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. Psychol Assess 1998;10(2):176-181. [doi: 10.1037/1040-3590.10.2.176]
31. Buysse DJ, Reynolds CF III, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
32. Teresi JA, Yu X, Stewart AL, Hays RD. Guidelines for designing and evaluating feasibility pilot studies. Med Care 2022 Jan 1;60(1):95-103. [doi: 10.1097/MLR.0000000000001664] [Medline: 34812790]
33. Bryce KL, Bowden JL, Hunter DJ, et al. Messages and notifications for the “OA Coach” knee osteoarthritis self-management mobile app: codevelopment and evaluation using a participatory research design with focus groups and surveys. J Med Internet Res 2026 May 4;28:e83507. [doi: 10.2196/83507] [Medline: 42081618]
34. Low J, De Oliveira Silva D, Ezzat AM, et al. Blind spots in reporting and representation in knee osteoarthritis research: a systematic review of the sociodemographic and clinical characteristics of 22,443 participants in randomised controlled trials of first-line care. Osteoarthr Cartil 2026 Jul;34(7):998-1006. [doi: 10.1016/j.joca.2026.02.014]
35. Choi W, Zheng H, Franklin P, Tulu B. mHealth technologies for osteoarthritis self-management and treatment: a systematic review. Health Informatics J 2019 Sep;25(3):984-1003. [doi: 10.1177/1460458217735676] [Medline: 29090628]
36. Yuan S, Bowden JL, Eyles JP, et al. Protocol for the comparison of remote versus face-to-face osteoarthritis management programs (COASTAL): a randomized implementation trial. Osteoarthr Cartil Open 2025 Sep;7(3):100647. [doi: 10.1016/j.ocarto.2025.100647] [Medline: 40697621]

Abbreviations

BCT: behavior change technique
CONSORT: Consolidated Standards of Reporting Trials
DASS-21: 21-item Depression, Anxiety, and Stress Scale
MAUQ: mHealth App Usability Questionnaire
mCSES: Modified Computer Self-Efficacy Scale
NRS: Numerical Rating Scale
OAKS: Osteoarthritis Knowledge Scale
PSQI: Pittsburgh Sleep Quality Index

XSL•FO
RenderX

Edited by A Kushniruk; submitted 25.Sep.2025; peer-reviewed by N Shah; revised version received 07.Apr.2026; accepted 04.May.2026; published 15.Jul.2026.

Please cite as:
Duong V, Bloul N, Bowden J, Bracken K, Bryce K, Deveza L, Eyles J, Farivar A, Huang R, Kobayashi S, Liu N, Pratt C, Strong C, Venkatesha V, Yu S, Hunter DJ
Feasibility of the "OA Coach" Mobile App to Support Individuals with Osteoarthritis: Development and Usability Testing
JMIR Hum Factors 2026;13:e84857
URL: https://humanfactors.jmir.org/2026/1/e84857
doi:10.2196/84857

© Vicky Duong, Naomi Bloul, Jocelyn Bowden, Karen Bracken, Kate Bryce, Leticia Deveza, Jillian Eyles, Abdolhay Farivar, Robin Huang, Sarah Kobayashi, Na Liu, Carin Pratt, Charlotte Strong, Venkatesha Venkatesha, Shirley Yu, David J Hunter. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 15.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Telemedicine Service Experience Questionnaire for Chinese Outpatients: Development and Validation Study

Peicheng Wang[1*], PhD; Kun Zhang[2*], PhD; Yanhua Chen[1], PhD; Minan Zhao[1], MS; Peiyao Li[1], MS; Yi Kuang[1], MS; Yanrong Du[1], MS; Weiguo Zhu[3], MD; Xiao Long[3], MD; Leiyu Shi[2], MD; Jiming Zhu[4], PhD

[1]Vanke School of Public Health, Tsinghua University, Beijing, China

[2]Department of Health Policy and Management, Bloomberg School of Public Health, The Johns Hopkins University, Baltimore, MD, United States

[3]International Health Service, Peking Union Medical College of Hospital, Beijing, China

[4]China Medical Board, Dongcheng District, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Jiming Zhu, PhD
China Medical Board, Dongcheng District, Beijing, China

Abstract

Background: Telemedicine has rapidly expanded; however, standardized, telemedicine-specific patient-reported experience measures tailored to outpatient workflows are limited in many settings.

Objective: This study aimed to develop and psychometrically validate the Telemedicine Service Experience Questionnaire (TSEQ) for Chinese outpatients using telemedicine services.

Methods: We conducted a web-based survey among outpatients who completed a telemedicine consultation at Peking Union Medical College Hospital between July 1, 2021, to August 31, 2021, and who had used telemedicine services, using an adapted Chinese Patient Experience Questionnaire that encompasses 15 questions across 4 dimensions, to investigate patients' telemedicine consultation experiences. Item generation was informed by a literature review, workflow mapping, and expert review. We evaluated the factor structure using exploratory factor analysis and confirmatory factor analysis on the full sample with cross-validation. Reliability was assessed using Cronbach α and item-total correlations.

Results: In total, 3338 participants completed the survey (mean age 45.3, SD 17.8 y; n=2182, 65.4% female participants; n=1827, 54.8% with college education or above). The exploratory factor analysis of the final 14-item scale resulted in 4 factors. After scrutinizing the content, these factors were labeled "Service Efficiency," "Post-treatment," "Information Guidance," and "Humanistic Care," and they demonstrated good internal consistency (Cronbach α values of 0.876, 0.840, 0.962, and 0.876, respectively). Moreover, as the average variance extracted values were greater than 0.5 and the composite reliability values were greater than 0.7, the TSEQ scale has high convergent validity. Our findings suggest that the psychometric properties of the 14-item TSEQ are valid and reliable for assessing telemedicine service experience among Chinese outpatients.

Conclusions: The TSEQ demonstrates a stable multidomain structure with satisfactory reliability and validity for evaluating outpatient telemedicine service experience in China. The instrument can support routine quality monitoring and guide targeted workflow improvements. Future studies should validate the TSEQ in multisite and postpandemic samples and examine measurement invariance across key subgroups.

(JMIR Hum Factors 2026;13:e60551) doi:10.2196/60551

KEYWORDS

telemedicine service; outpatients experience questionnaire; validity and reliability; Chinese outpatients; telemedicine; patient satisfaction; questionnaire; China; outpatient; care; information technology; health care system

Introduction

With the rapid development of information technology and the popularization of the internet, telemedicine services are becoming an innovative approach in the medical field [1] and are being used in a variety of settings, including psychiatric counseling [2], home health care services [3], intensive care units [4], and emergency stroke care [5], to provide patients with efficient and convenient medical care. In recent years, the worldwide COVID-19 outbreak has accelerated the spread and adoption of telemedicine services in health care [5]. Faced with the constraints of outbreaks and the need for prevention and control, telemedicine services protect patients, clinicians, and communities from exposure. In addition, they provide a platform where physicians and patients can interact using smartphones or webcam-enabled computers, regardless of time or location

XSL•FO
RenderX

[6]. Particularly in China, a vast country with rural areas and dispersed populations, telemedicine services hold great potential and advantages in addressing the unequal distribution of health care resources and improving the quality of medical services [7]. The Chinese government has explicitly included telemedicine services as a key area in the development of the national health care system. In 2018, the National Health Commission of China launched the Action Plan for Further Improvement of Medical Services (2018 - 2020). The program encourages medical institutions and research institutes to strengthen innovative research and development and vigorously promotes the integration of telemedicine technology with medical resources to facilitate the gradual development of telemedicine [8].

As an emerging health care model, patient satisfaction with telemedicine services is an important indicator of the quality of telemedicine services and an important driving force for the development of telemedicine services. Collecting and analyzing patient feedback can help medical institutions and decision-makers to comprehensively understand patients' demands and expectations, improve and optimize service quality, attract more patients to accept telemedicine services, and promote their wide application in health care. Studies have shown that the use of telemedicine services will only expand if patients and health care professionals are at least as satisfied with telemedicine services as they are with in-person visits [9]. In addition, highly satisfied patients are likely to have higher medication adherence, leading to better health outcomes [10].

Several studies have been conducted to explore the satisfaction with telemedicine services, such as the Telemedicine Satisfaction and Usefulness Questionnaire [11], Telemedicine Satisfaction Survey [12], and Telehealth Usability Questionnaire [13]. Nevertheless, telemedicine-specific, psychometrically validated patient-reported experience instruments tailored to Chinese outpatient workflows remain limited. In China, the scales currently available for measuring telemedicine service satisfaction are mainly based on the extension and improvement of traditional health care service satisfaction scales and lack sufficient consideration of the specific characteristics and unique features of telemedicine services [14]. For example, these scales may not fully capture telemedicine-specific processes (eg, digital information guidance, postconsultation payment and medication delivery, and privacy or security concerns).

To address this gap, we developed the Telemedicine Service Experience Questionnaire (TSEQ), using various items adapted for patient satisfaction surveys among Chinese outpatients. Our primary objective in this paper was to assess the appropriateness of the TSEQ scale as an instrument to measure patient satisfaction across a broad range of outpatient populations in China. Rather than measuring technology usability alone, the TSEQ is designed to capture the service-process experience across the entire outpatient telemedicine journey, including preconsultation guidance, interaction quality, and postconsultation processes (eg, payment and medication delivery), which are particularly salient in platform-based telemedicine services. Accordingly, the objectives of this study were to (1) develop the TSEQ for Chinese outpatient telemedicine services; (2) evaluate its underlying factor structure using exploratory and confirmatory factor analyses; and (3) assess its reliability and validity, including internal consistency and convergent validity.

Methods

Data Collection and Sample

Our research focuses on users of hospital apps at Peking Union Medical College Hospital, a leading medical institution renowned for its comprehensive health care services and advanced medical technology. Data were collected through a web-based questionnaire from July 1, 2021, to August 31, 2021. Eligible participants were adults (aged ≥18 years) who completed a telemedicine consultation and provided informed consent. We excluded responses with substantial missing data, implausibly short completion times, or duplicate submissions where applicable. Convenience sampling was used, which is common in psychometric validation studies. In this study, patients who fully completed the TSEQ questionnaire were included (N=3337). The survey window was selected because telemedicine services at the study hospital had reached stable operational routines and sufficient user volume for psychometric evaluation. As this study focuses on instrument development and validation, the 2021 dataset provides an appropriate foundation to examine the factor structure and reliability and validity of the TSEQ.

Ethical Considerations

A web-based informed consent was obtained before the survey via WeChat. Participants were allowed to withdraw from the survey at any time. All interviews of the survey were kept confidential and anonymous. This study was approved by the Peking Union Medical College Hospital Ethics Committee (S-k1545). Responses were collected anonymously (or deidentified), and no directly identifiable personal information was stored. Data were stored on secure servers with restricted access limited to the research team. Participants did not receive any compensation for participation in this study.

Measures and Scale Development

The TSEQ was adapted from the Chinese Patient Experience Questionnaire (CPEQ; refer to Multimedia Appendix 1 for details) [15]. The CPEQ consists of 2 sections: part A includes 9 questions that collect basic demographic information of the participants, such as age, sex, employment status, and type of insurance and part B comprises 22 questions across 5 dimensions and is designed to comprehensively evaluate patient experience. On the basis of expert discussions on the CPEQ, we developed a 15-item questionnaire for telemedicine service experience, focusing on self-efficacy, information guidance, and humanistic care. Every question was assessed using a 5-point Likert scale (5=strongly agree, 4=agree, 3=unsure, 2=disagree, and 1=strongly disagree). The newly formed TSEQ is presented in Multimedia Appendix 2.

The validation and reliability of the TSEQ are the cornerstones of our research's second objective. We assessed the instrument's internal consistency and test-retest reliability, alongside its validity, to ensure it measures the intended constructs. An

XSL•FO
RenderX

open-ended question was included at the end of the questionnaire to capture valuable, unstructured feedback from the respondents.

Statistical Analysis

Descriptive statistics were used to summarize and describe the demographic characteristics of the participants, such as age, sex, occupation, and education. Categorical variables (eg, sex, occupation, and education) were presented as numbers (percentages), and continuous variables (eg, age) were reported as mean (SD).

Exploratory Factor Analysis

Exploratory factor analysis (EFA) was first performed to explore the number of factors and the factor loading of each item using maximum likelihood estimation, which is the default estimator in Mplus (version 7.0; Muthén & Muthén). Before the factor analysis, Bartlett test of sphericity was usually used to examine whether these variables were independent of each other and suitable for factor analysis. In this study, we computed 1-factor, 2-factor, 3-factor, and 4-factor measurement models using EFA. To identify the model structure, we used a series of indicators such as χ^2, *df*, Tucker-Lewis index (TLI), comparative fit index (CFI), Akaike information criterion, Bayesian information criterion [1,16], standardized root mean square residual (SRMR), and root mean square error of approximation (RMSEA) to evaluate the goodness-of-fit statistics. Acceptable model fit was defined by the following criteria: TLI>0.90, CFI>0.90, SRMR<0.05, and RMSEA<0.08 [17-19].

During EFA, items were considered for removal if they exhibited low factor loadings (eg, <0.40), substantial cross-loadings, low commonalities, weak item-total correlations, or conceptual redundancy. The refined structure was then tested using confirmatory factor analysis (CFA), and model fit indices guided the final scale specification.

Model Validation Using CFA

Following EFA, we conducted the CFA to verify whether the hypothesized structure from EFA was consistent and valid. CFA is a statistical technique used to evaluate the effectiveness of the measurement model and examine the relationship between measured variables and underlying latent constructs or factors. In this study, first-order factors were verified for the dimensions of telemedicine service experience, and a second-order model was established to explain the relationships among multiple first-order factors. We computed and compared the overall fit coefficients, χ^2/df, RMSEA, CFI, TLI, and SRMR of CFA models to remove inappropriate items and confirm the ideal model results. Moreover, the model fit was evaluated using the standardized factor loadings and R^2.

Reliability and Validity

To assess the reliability of CFA-confirmed models, we further computed Cronbach α and split-half reliability coefficients (SRCs) for the total questionnaire, final questionnaire, service efficiency (F1), post-treatment (F2), information guidance (F3), and humanistic care (F4). With regard to the reliability of a research instrument, statistical theory has determined that Cronbach α should be >0.7 [20,21]. Correlation analysis was used to investigate the correlation between the first-order factors and Chinese patients' overall satisfaction and willingness to recommend. Finally, a validity test was conducted to verify the validity of the newly proposed Chinese adaptation of the patient experience measurement instrument by estimating the factor loadings to examine the relationships between observed variables and latent constructs that we wanted to measure. Standard factor loadings >0.70 were considered the criterion for convergent validity [22].

Results

Sociodemographic Characteristics

Among the 3337 participants, the mean age was 45.3 (SD 17.8) years. The majority were female (n=2182, 65.4%), had completed a bachelor's degree or higher education level (n=1827, 54.8%), paid using health insurance (n=2292, 68.7%), and were seeking medical care outside their local area (n=1372, 41.1%; Table 1).

Table 2 presents the performance scores for the 15 individual items of the TSEQ in this study. These items primarily focus on various aspects of hospital services, including waiting times, service convenience, the communication style of physicians, and the respect and privacy protection afforded to patients during their medical visits. The average scores for each item generally range between 4.2 and 4.6, indicating that most respondents are satisfied with the services provided by the hospital and that the scores demonstrate relatively high consistency. These findings support further development of the scale for the EFA and the CFA.

XSL•FO
RenderX

Table . Demographic characteristics of the participants (N=3337).

Variables	Participants
Age (years), mean (SD)	45.3 (17.8)
Sex, n (%)	
Male	1155 (34.6)
Female	2182 (65.4)
Occupation, n (%)	
Government and public institutions	881 (26.4)
State-owned enterprise	536 (16.1)
Nonpublic enterprise	815 (24.4)
Others	1105 (33.1)
Education, n (%)	
Bachelor's degree or higher education level	1827 (54.8)
Others	1510 (45.2)
Registration, n (%)	
Online	1782 (53.4)
Others	1555 (46.6)
Payment, n (%)	
Self-pay	1045 (31.3)
Insurance	2292 (68.7)
Allopatry treatment[a], n (%)	
Yes	1372 (41.1)
No	1965 (58.9)
Virtual physician visits, n (%)	
Yes	1024 (30.7)
No	2313 (69.3)

[a]Allopatry treatment refers to seeking outpatient care outside the patient's usual place of residence or registration (ie, receiving care in a nonlocal setting).

XSL•FO
RenderX

Table . Performance scores for the 17 individual items of the Telemedicine Service Experience Questionnaire.

Question number	Item	Value, mean (SD)
Q1	The waiting time for on-site registration was acceptable.	4.2 (1.0)
Q2	The waiting time for consultation with physicians was acceptable.	4.5 (0.8)
Q3	The consultation length was acceptable.	4.5 (0.7)
Q4	The waiting time for a planned examination was acceptable.	4.5 (0.7)
Q5	The waiting time for payment was acceptable.	4.6 (0.6)
Q6	The waiting time for medicine delivery was acceptable.	4.6 (0.6)
Q7	It was convenient to access self-service information inquiry devices.	4.5 (0.7)
Q8	Medication instruction services provided by the hospital can meet needs.	4.5 (0.7)
Q9	Physicians discussed my condition and care patiently.	4.5 (0.8)
Q10	Physicians explained examination results to me patiently.	4.5 (0.8)
Q11	Physicians discussed treatment with me patiently.	4.5 (0.8)
Q12	I was treated with respect and dignity during this visit.	4.5 (0.8)
Q13	My privacy has been fully protected during this visit.	4.6 (0.6)
Q14	All the medical staff was kind to me during this visit.	4.5 (0.8)
Q15	I got help from medical staff as soon as I had problems.	4.5 (0.8)

EFA Findings

The results of the EFA indicated that the Bartlett test of sphericity yielded a χ^2 value of 785.966 (df=51; P<.001; Kaiser-Meyer-Olkin value=0.950; Table 3), suggesting that the dataset used was suitable for factor analysis. Through this analysis, we successfully established a 4-factor structure, with these factors collectively explaining 78.49% of the total variance (Multimedia Appendix 3). This indicates that the TSEQ tool exhibits good structural validity, accurately reflecting various aspects of patient experiences in remote health care services.

Table . Fit indices for the exploratory factor analysis model.

Model	χ^2 (df)	Tucker-Lewis index	Comparative fit index	Akaike information criterion	Bayesian information criterion	Standardized root mean squared residual	Root mean squared error of approximation (90% CI)
One-factor	9637.302 (90)	0.754	0.789	77359.640	77634.715	0.086	0.178 (0.175-0.181)
Two-factor	2597.384 (76)	0.923	0.944	70347.720	70708.377	0.028	0.100 (0.096-0.103)
Three-factor	1262.718 (63)	0.956	0.974	69039.054	69479.177	0.021	0.076 (0.072-0.079)
Four-factor[a]	*785.966 (51)*	*0.967*	*0.984*	*68586.302*	*69099.780*	*0.016*	*0.066 (0.062-0.070)*

[a]Italicized text indicates the selected final model.

Of the 15 items, 14 loaded meaningfully onto 1 of the 4 factors, with factor loadings ranging from 0.38 to 0.89. However, one item (item 12: “Physicians discussed my condition and care patiently”) had cross-loadings on factors 2 and 4. Despite this, considering the importance of this item, we did not remove it

XSL•FO
RenderX

in the EFA stage. The EFA resulted in a 4-factor solution that explained 79.00% of the variance (Table 4).

Table . Results of the exploratory factor loadings matrix.

Item number	Item	Factor loading			
		Factor 1	Factor 2	Factor 3	Factor 4
Q1	The waiting time for on-site registration was acceptable.	*0.689* [a]	0.038	−0.147	0.018
Q2	The waiting time for consultation with physicians was acceptable.	*0.670*	0.162	−0.053	0.037
Q3	The consultation length was acceptable.	*0.707*	−0.033	0.153	0.004
Q4	The waiting time for a planned examination was acceptable.	*0.744*	−0.019	0.102	0.002
Q5	The waiting time for payment was acceptable.	0.279	−0.024	*0.517*	0.092
Q6	The waiting time for medicine delivery was acceptable.	0.021	0.065	*0.887*	0.006
Q7	It was convenient to make an appointment for a planned examination.	*0.607*	0.021	0.083	0.052
Q8	Medication instruction services provided by the hospital can meet needs.	*0.521*	0.180	0.221	−0.026
Q9	Physicians discussed my condition and care patiently.	0.005	*0.875*	0.001	0.069
Q10	Physicians explained examination results to me patiently.	0.072	*0.928*	0.000	−0.032
Q11	Physicians discussed treatment with me patiently.	−0.011	*0.902*	0.030	0.055
Q12	I was treated with respect and dignity during this visit.	−0.001	*0.444*	0.000	*0.523*
Q13	My privacy has been fully protected during this visit.	0.273	0.186	0.048	*0.382*
Q14	All the medical staff was kind to me during this visit.	−0.001	0.054	0.033	*0.860*
Q15	I got help from medical staff as soon as I had problems.	0.266	−0.025	−0.017	*0.662*

[a]Italicized factor loadings indicate the highest loading for each item and were used to assign the item to the corresponding factor.

CFA Findings

In the CFA, it was observed that the modification index for item 12 was excessively high across the 4 factors (>200; Multimedia Appendix 4). After its removal, the model fit improved in the CFA model. Therefore, the final model excluded item 12 (Table 5).

XSL•FO
RenderX

Given the 4 factors derived from EFA, an interpretation focusing on items with higher factor loadings for each factor, and their comparison to the TSEQ, was performed by the panel (Table 6). Factor 1 was labeled "Service Efficiency," as it consisted of 6 items dealing with the convenience of telemedicine services. The second factor, consisting of 2 items, was called "Post-treatment," as the content involves the time spent on medication dispensing and payment processes. Factor 3 comprised 3 items that captured information-related elements of telemedicine services; therefore, the title of this factor was "Information Guidance." Finally, the 3 items in factor 4, which involve the physician's' attitudes toward patients, were labeled "Humanistic Care."

Table . Overall model fit coefficients for the confirmatory factor analysis model.

Model fit index	Ideal standard	Common standard	Model results	Results after excluding item 12	Second-order model	Conclusion (first-order model)
χ^2/df[a]	<3	<10	23.33	18.70	30.30	Model not ideal
Root mean square error of approximation	<0.08	<0.1	0.082	0.073	0.094	Model ideal
Comparative fit index	>0.9	>0.8	0.959	0.968	0.946	Model ideal
Tucker-Lewis index	>0.9	>0.8	0.948	0.960	0.933	Model ideal
Standardized root mean square residual	<0.08	<0.1	0.035	0.030	0.047	Model ideal

[a]χ^2/df: the ratio of χ^2 to df.

XSL•FO
RenderX

Table . Results of confirmatory factor analysis of the Telemedicine Service Experience Questionnaire scales.

First-order factor and indicators	Standardized factor loading[a]	R^2
Service efficiency (F1)		
Q1[b]	0.599	0.359
Q2	0.763	0.582
Q3	0.795	0.632
Q4	0.798	0.636
Q7	0.729	0.531
Q8	0.812	0.660
Post-treatment (F2)		
Q5	0.836	0.699
Q6	0.867	0.752
Information guidance (F3)		
Q9	0.935	0.873
Q10	0.951	0.904
Q11	0.953	0.909
Humanistic care (F4)		
Q13	0.808	0.652
Q14	0.879	0.772
Q15	0.846	0.716
Second-order model and indicators		
Service efficiency (F1)[c]	0.936	0.876
Post-treatment (F2)	0.826	0.683
Information guidance (F3)	0.820	0.672
Humanistic care (F4)	0.940	0.883

[a]The *P* values for the standardized factor loadings are all <.001.

[b]Q1-Q15 denote questionnaire items, the full wording of which is provided in Table 4.

[c]F1-F4 denote the 4 factors of the second-order model.

Internal Consistency and Validity

As presented in Table 7, the final instrument obtained a Cronbach α coefficient of 0.946 and an SRC value of 0.872 for the total sample. The 4 factors—service efficiency, post-treatment, information guidance, and humanistic care—each obtained Cronbach α coefficients ranging from 0.840 to 0.962, and their SRC values were 0.877 (6 items), 0.840(2 items), 0.965 (3 items), and 0.892 (3 items), respectively. These results demonstrate that the TSEQ has good internal consistency.

The correlation coefficients between overall satisfaction and the 4 factors—service efficiency, post-treatment, information guidance, and humanistic care—were 0.941, 0.781, 0.857, and 0.901, respectively. All the correlations were statistically significant at the .01 level and demonstrate good internal validity. Moreover, as the average variances extracted values were >0.5 and the composite reliability values were >0.7, the TSEQ scale demonstrated high convergent validity (Table 7).

XSL•FO
RenderX

Table . The evidence for the reliability and validity of the Telemedicine Service Experience Questionnaire (TSEQ) scale and 4 factors.

Variable	Correlation[a]				Internal consistency		Convergent validity	
	TSEQ	Service efficiency (F1)	Post-treatment (F2)	Information guidance (F3)	Cronbach α	Split-half reliability coefficients	Average variances extracted	Composite reliability
TSEQ (14-items, mean)	—[b]	—	—	—	0.946	0.872	—	—
Service efficiency (F1)	0.941	—	—	—	0.876	0.876	0.567	0.886
Post-treatment (F2)	0.781	0.805	—	—	0.840	0.840	0.725	0.841
Information guidance (F3)	0.857	0.667	0.520	—	0.962	0.962	0.896	0.962
Humanistic care (F4)	0.901	0.760	0.624	0.788	0.876	0.876	0.714	0.882

[a]All Pearson correlation coefficients were statistically significant (P<.001).

[b]Not applicable.

TSEQ Scoring

The TSEQ measures each item on a 5-point Likert-type scale, ranging from 1 (“strongly disagree”) to 5 (“strongly agree”). Calculating a total score involves summing the response scores for all 14 items of the TSEQ. The total score ranges from 14 to 70, with higher scores indicating higher levels of telemedicine service experience. The Chinese version of the questionnaire is shown in Multimedia Appendix 5.

Discussion

Principal Findings

This study developed and validated the TSEQ, an instrument for measuring telemedicine service experience in Chinese outpatients, based on the CPEQ [15]. In this study, EFA produced a revised 4-factor solution for the TSEQ, leading to modifications in the 5 domains of the CPEQ. Telemedicine services offer convenient medical solutions and effectively tackle the issue of uneven distribution of medical resources, especially in remote and rural areas [23,24]. The global COVID-19 pandemic has recently accelerated the adoption and expansion of telemedicine [25,26]. Importantly, our study does not aim to re-establish the well-known advantage that telemedicine reduces geographic barriers. Instead, it contributes a validated, telemedicine-specific patient experience instrument and highlights which service-process dimensions matter in Chinese outpatient telemedicine workflows. In particular, the emergence of a distinct “post-treatment” domain (eg, payment timing and medication delivery) reflects operational features that are often underrepresented in traditional in-person satisfaction tools and provides actionable targets for quality improvement. Our findings contribute to global efforts to standardize patient-reported evaluation of telemedicine quality. Although the TSEQ was developed for Chinese outpatient workflows, its domain-based structure may inform telemedicine service monitoring in other settings, subject to appropriate cross-cultural adaptation and validation.

In this analysis, the final instrument comprises 4 factors with 14 items, designed to evaluate patient satisfaction with telemedicine services and identify areas for service improvement. The EFA supported the viability of a 4-factor model. Compared to previous studies, the EFA uniquely identified “payment timing (item 5)” and “medicine delivery (item 6)” as a new single factor, while grouping “It was convenient to make an appointment for a planned examination (item 7)” and “Medication instruction services provided by the hospital can meet needs (item 8)” under the first factor, service efficiency. Notably, the item “I was treated with respect and dignity during this visit (item 12)” was originally categorized under “Humanistic Care” in the CPEQ framework but was removed in this study. This change may be attributed to the fact that patients using telemedicine services are more focused on the quality and efficiency of the service and may not experience significant nonverbal communication cues such as eye contact, body language, and tone of voice, which are typically present during face-to-face interactions with health care professionals [24,27].

Additionally, although “privacy and security risks (item 13)” showed lower factor loadings in the EFA, these aspects were still included in the TSEQ due to their critical importance in the telemedicine context [28,29], where patient privacy and security are paramount. Telemedicine also introduces challenges, including technology barriers, digital literacy gaps, privacy or security concerns, and constrained clinician-patient communication. By quantifying patient experience across distinct service domains, the TSEQ can help hospital identify where telemedicine workflows fail to meet expectations and prioritize targeted improvements.

Furthermore, CFA validated the 4-factor structure of the questionnaire. According to the pre-established CFI, and RMSEA thresholds, the second-order model can be considered acceptable. Only the χ^2 test revealed a bad fit for the second-order model, likely due to the large sample size [30]. All standardized factor loadings were statistically significant (all P<.001) and ranged from 0.729 to 0.953. On the basis of

XSL•FO
RenderX

the CPEQ [15], our study has introduced a second factor, defined as "post-treatment," further demonstrating that the efficiency of posttelemedicine services also affects overall patient satisfaction. Analysis of the internal consistency of the Chinese version of the TSEQ showed that the Cronbach α of the entire scale was 0.95, indicating excellent scale reliability. All scales demonstrated good internal consistency, with Cronbach α of >0.80, indicating moderate to good test-retest reliability [30,31]. Moreover, CFA revealed strong correlations among the 4 factors, and further exploration of the relationships among the 4 factors confirmed satisfactory convergent validity based on the average variances extracted and composite reliability values [32,33].

The majority of participants in our study were female (2182/3337, 65.4%) and had completed bachelor's degree or higher education level (n=1827, 54.8%). Nearly half (n=1372, 41.1%) of the participants were seeking medical care outside their local area. The composition of our study population underscores the broad relevance of our findings. This diversity is essential for understanding the multifaceted patient experiences within telehealth services, emphasizing the need for inclusive and accessible health care solutions that cater to varied demographic segments [30,33].

The TSEQ offers actionable value for routine quality monitoring and continuous improvement in outpatient telemedicine services. In practice, health care professionals and platform administrators can integrate the TSEQ as a brief postvisit feedback tool embedded within the telemedicine workflow (eg, immediately after consultation or after completion of postconsultation steps). Domain-level scores can be visualized in a dashboard to identify specific bottlenecks and track trends over time, enabling a closed improvement loop of "measure-diagnose-intervene-reassess." For example, low scores in interaction-related domains may prompt targeted clinician training (eg, communication clarity and shared decision-making in remote settings), whereas low scores in the post-treatment domain can inform workflow redesign for payment, prescription processing, and medication delivery or collection. The TSEQ can also support benchmarking across departments, specialties, or service lines to prioritize resource allocation and evaluate the impact of operational changes. At the policy level, aggregated TSEQ indicators may complement use metrics by capturing patient-centered service quality; helping health systems move from "telemedicine adoption" to "telemedicine performance"; and guiding standards for service transparency, privacy protection, and equitable access for digitally disadvantaged groups.

Limitations

Several limitations should be acknowledged. First, this study relied on self-reported survey responses collected immediately after telemedicine use, which may be subject to social desirability bias, recall bias, and common method variance. Second, participants were recruited using a convenience sampling approach from a single hospital setting, which may limit generalizability to other regions, health care tiers, and telemedicine platforms, especially where service workflows, payment pathways, and medication dispensing or delivery arrangements differ. Third, the sample characteristics may not fully represent telemedicine users with limited digital access or lower digital literacy; therefore, our estimates of service experience may be upwardly biased for populations facing stronger technology barriers. Fourth, data were collected during mid-2021, when telemedicine use and patient expectations were shaped by the COVID-19 pandemic context; experience patterns may shift in postpandemic periods when patients have broader choices between in-person and remote visits. Fifth, the cross-sectional design precludes assessment of temporal stability (eg, test-retest reliability), responsiveness to service improvements, and longitudinal predictive validity. Finally, although we established overall psychometric properties, formal measurement invariance and differential item functioning across key subgroups (eg, age, sex, education, allopatry treatment, and previous telemedicine experience) were not the primary focus of this manuscript and warrant further validation in multisite and postpandemic samples.

Conclusions

In conclusion, this study developed and psychometrically validated the TSEQ for Chinese outpatients. The final instrument demonstrated a stable multidomain structure, indicating that it can be used to capture key dimensions of patients' telemedicine service experience in routine outpatient care. By providing domain-specific scores, the TSEQ enables health care institutions and platform administrators to identify actionable targets for quality improvement, particularly in end-to-end workflow components that are salient in telemedicine settings (eg, communication processes and postconsultation services). Future research should validate the TSEQ in multisite and postpandemic samples, assess test-retest reliability and responsiveness to service changes, and examine measurement invariance or differential item functioning across key subgroups to support broader implementation and benchmarking.

Acknowledgments

The authors thank all study participants and staff members involved in data collection for their contributions.

Funding

The authors declare that no financial support was received for this study.

Data Availability

The data generated and analyzed during this study are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

PW and KZ contributed to methodology, formal analysis, and drafting of the original manuscript. YC, PW, MZ, PL, YK, WZ, and XL contributed to data collection, data analysis, data interpretation, and manuscript review and editing. LS and JZ critically revised the manuscript for important intellectual content. All authors approved the final manuscript and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Original version of the Chinese Patient Experience Questionnaire.
[DOCX File, 19 KB - humanfactors_v13i1e60551_app1.docx]

Multimedia Appendix 2
Newly formed version of the Telemedicine Service Experience Questionnaire.
[DOCX File, 19 KB - humanfactors_v13i1e60551_app2.docx]

Multimedia Appendix 3
Four-factor analysis.
[DOCX File, 18 KB - humanfactors_v13i1e60551_app3.docx]

Multimedia Appendix 4
Modification index.
[DOCX File, 19 KB - humanfactors_v13i1e60551_app4.docx]

Multimedia Appendix 5
Chinese version of the Telemedicine Service Experience Questionnaire.
[DOCX File, 18 KB - humanfactors_v13i1e60551_app5.docx]

Checklist 1
STROBE checklist.
[PDF File, 135 KB - humanfactors_v13i1e60551_app6.pdf]

References

1. Alam MZ, Hoque MR, Hu W, Barua Z. Factors influencing the adoption of mHealth services in a developing country: a patient-centric study. Int J Inf Manag 2020 Feb;50(2):128-143. [doi: 10.1016/j.ijinfomgt.2019.04.016]
2. Marrone S, Mitchell JE, Crosby R, Wonderlich S, Jollie-Trottier T. Predictors of response to cognitive behavioral treatment for bulimia nervosa delivered via telemedicine versus face-to-face. Int J Eat Disord 2009 Apr;42(3):222-227. [doi: 10.1002/eat.20603] [Medline: 18951452]
3. Smith CE, Dauz ER, Clements F, et al. Telehealth services to improve nonadherence: a placebo-controlled study. Telemed J E Health 2006 Jun;12(3):289-296. [doi: 10.1089/tmj.2006.12.289] [Medline: 16796496]
4. Breslow MJ, Rosenfeld BA, Doerfler M, et al. Effect of a multiple-site intensive care unit telemedicine program on clinical and economic outcomes: an alternative paradigm for intensivist staffing. Crit Care Med 2004 Jan;32(1):31-38. [doi: 10.1097/01.CCM.0000104204.61296.41] [Medline: 14707557]
5. Waite K, Silver F, Jaigobin C, et al. Telestroke: a multi-site, emergency-based telemedicine service in Ontario. J Telemed Telecare 2006;12(3):141-145. [doi: 10.1258/135763306776738611] [Medline: 16638235]
6. Garfan S, Alamoodi AH, Zaidan BB, et al. Telehealth utilization during the Covid-19 pandemic: a systematic review. Comput Biol Med 2021 Nov;138:104878. [doi: 10.1016/j.compbiomed.2021.104878] [Medline: 34592585]
7. Cheng N, Zhao Y, Li X, He X, Wang A. Translation and validation study of the Chinese version of the service user technology acceptability questionnaire. Asia Pac J Oncol Nurs 2023 Apr;10(6):100239. [doi: 10.1016/j.apjon.2023.100239] [Medline: 37288350]
8. The action plan for further improvement of medical services (2018-2020) [Article in Chinese]. The State Council Information Office of the People's Republic of China. URL: http://www.scio.gov.cn/xwfb/gwyxwbgsxwfbh/wqfbh_2284/2018n_7138/2018n02y12r/wjxgzc_7305/202207/t20220715_205970.html [accessed 2026-04-30]
9. Nguyen M, Waller M, Pandya A, Portnoy J. A review of patient and provider satisfaction with telemedicine. Curr Allergy Asthma Rep 2020 Sep 22;20(11):72. [doi: 10.1007/s11882-020-00969-7] [Medline: 32959158]

XSL•FO
RenderX

10. Bakar ZA, Fahrni ML, Khan TM. Patient satisfaction and medication adherence assessment amongst patients at the diabetes medication therapy adherence clinic. Diabetes Metab Syndr 2016;10(2 Suppl 1):S139-S143. [doi: 10.1016/j.dsx.2016.03.015] [Medline: 27055354]
11. Bakken S, Grullon-Figueroa L, Izquierdo R, et al. Development, validation, and use of English and Spanish versions of the telemedicine satisfaction and usefulness questionnaire. J Am Med Inform Assoc 2006;13(6):660-667. [doi: 10.1197/jamia.M2146] [Medline: 16929036]
12. Yip MP, Chang AM, Chan J, MacKenzie AE. Development of the Telemedicine Satisfaction Questionnaire to evaluate patient satisfaction with telemedicine: a preliminary study. J Telemed Telecare 2003;9(1):46-50. [doi: 10.1258/135763303321159693] [Medline: 12641893]
13. Parmanto B, Lewis ANJ, Graham KM, Bertolet MH. Development of the Telehealth Usability Questionnaire (TUQ). Int J Telerehabil 2016;8(1):3-10. [doi: 10.5195/ijt.2016.6196] [Medline: 27563386]
14. Ma Q, Sun D, Tan Z, et al. Usage and perceptions of telemedicine among health care professionals in China. Int J Med Inform 2022 Oct;166:104856. [doi: 10.1016/j.ijmedinf.2022.104856] [Medline: 36037593]
15. Hu G, Chen Y, Liu Q, et al. Patient experience of hospital care in China: major findings from the Chinese Patient Experience Questionnaire Survey (2016-2018). BMJ Open 2019 Sep 20;9(9):e031615. [doi: 10.1136/bmjopen-2019-031615] [Medline: 31542764]
16. Bentler PM. Comparative fit indexes in structural models. Psychol Bull 1990 Mar;107(2):238-246. [doi: 10.1037/0033-2909.107.2.238] [Medline: 2320703]
17. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model 1999;6(1):1-55. [doi: 10.1080/10705519909540118]
18. Browne MW, Cudeck R. Alternative ways of assessing model fit. Sociol Methods Res 1992;21:230-258. [doi: 10.1177/0049124192021002005]
19. Byrne BM. Structural Equation Modeling with Mplus: Basic Concepts, Applications, and Programming, 1st edition: Routledge; 2012. [doi: 10.4324/9780203807644]
20. Hair JF, Babin BJ, Anderson RE, Black WC. Multivariate Data Analysis: Cengage India; 2018.
21. Blunch N. Introduction to Structural Equation Modelling Using SPSS and AMOS: SAGE Publications; 2008.
22. Hair JF, Ringle CM, Sarstedt M. PLS-SEM: indeed a silver bullet. J Mark Theory Pract 2011 Apr;19(2):139-152. [doi: 10.2753/MTP1069-6679190202]
23. Schweiberger K, Hoberman A, Iagnemma J, et al. Practice-level variation in telemedicine use in a pediatric primary care network during the COVID-19 pandemic: retrospective analysis and survey study. J Med Internet Res 2020 Dec 18;22(12):e24345. [doi: 10.2196/24345] [Medline: 33290244]
24. Effendi DE, Ardani I, Handayani S, Nugroho AP, Fitrianti Y, Machfutra ED. Patient experience correlates with willingness to retain service utilization in post-pandemic times among telemedicine users in Indonesia. Telemed J E Health 2024 Aug;30(8):2173-2180. [doi: 10.1089/tmj.2024.0008] [Medline: 38752868]
25. Wong DC, Wu Y, Fung H, et al. The provision and perceptions of telemedicine services among traditional Chinese medicine practitioners during COVID-19: a cross-sectional study in Hong Kong, China. Telemed J E Health 2024 Aug;30(8):e2335-e2343. [doi: 10.1089/tmj.2023.0635] [Medline: 38752867]
26. Spaulding EM, Fang M, Commodore-Mensah Y, Himmelfarb CR, Martin SS, Coresh J. Prevalence and disparities in telehealth use among US adults following the COVID-19 pandemic: national cross-sectional survey. J Med Internet Res 2024 May 10;26:e52124. [doi: 10.2196/52124] [Medline: 38728070]
27. Nissinen S, Pesonen S, Toivio P, Sormunen E. Exploring the use, usefulness and ease of use of digital occupational health services: a descriptive correlational study of customer experiences. Digit Health 2024;10:20552076241242668. [doi: 10.1177/20552076241242668] [Medline: 38601187]
28. Hall JL, McGraw D. For telehealth to succeed, privacy and security risks must be identified and addressed. Health Aff (Millwood) 2014 Feb;33(2):216-221. [doi: 10.1377/hlthaff.2013.0997] [Medline: 24493763]
29. Lee A, Fleming JA, Thomson Reuters Accelus. Healthcare information technology. Issue Brief Health Policy Track Serv 2015 Dec 28:1-80. [Medline: 27116770]
30. Vujkovic B, Brkovic V, Pajičić A, et al. Serbian version of the eHealth Literacy Questionnaire (eHLQ): translation, cultural adaptation, and validation study among primary health care users. J Med Internet Res 2024 May 9;26:e57963. [doi: 10.2196/57963] [Medline: 38722675]
31. Chen YC, Cheng C, Osborne RH, Kayser L, Liu CY, Chang LC. Validity testing and cultural adaptation of the eHealth Literacy Questionnaire (eHLQ) among people with chronic diseases in Taiwan: mixed methods study. J Med Internet Res 2022 Jan 19;24(1):e32855. [doi: 10.2196/32855] [Medline: 35044310]
32. Anderberg P, Eivazzadeh S, Berglund JS. A novel instrument for measuring older people's attitudes toward technology (TechPH): development and validation. J Med Internet Res 2019 May 23;21(5):e13951. [doi: 10.2196/13951] [Medline: 31124467]
33. Choi J, Choi S, Song K, et al. Everyday Digital Literacy Questionnaire for older adults: instrument development and validation study. J Med Internet Res 2023 Dec 14;25:e51616. [doi: 10.2196/51616] [Medline: 38095999]

XSL•FO
RenderX

Abbreviations

CFA: confirmatory factor analysis
CFI: comparative fit index
CPEQ: Chinese Patient Experience Questionnaire
EFA: exploratory factor analysis
RMSEA: root mean square error of approximation
SRC: split-half reliability coefficient
SRMR: standardized root mean square residual
TLI: Tucker-Lewis index
TSEQ: Telemedicine Service Experience Questionnaire

Edited by A Kushniruk; submitted 20.May.2024; peer-reviewed by CH Chan, F Amer; revised version received 30.Dec.2025; accepted 09.Feb.2026; published 21.May.2026.

Please cite as:
Wang P, Zhang K, Chen Y, Zhao M, Li P, Kuang Y, Du Y, Zhu W, Long X, Shi L, Zhu J
Telemedicine Service Experience Questionnaire for Chinese Outpatients: Development and Validation Study
JMIR Hum Factors 2026;13:e60551
URL: https://humanfactors.jmir.org/2026/1/e60551
doi:10.2196/60551

© Peicheng Wang, Kun Zhang, Yanhua Chen, Minan Zhao, Peiyao Li, Yi Kuang, Yanrong Du, Weiguo Zhu, Xiao Long, Leiyu Shi, Jiming Zhu. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 21.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Design and Performance of an Email-Based Patient Recruitment Campaign in Primary Care Research: Formative Secondary Analysis

Vanessa T Vaillancourt[1], MSc; Marie-Dominique Poirier[1], MSc; Amélie Fournier[1]; Sabrina T Wong[2,3], PhD, RN; Marie-Eve Poitras[1], PhD, RN

[1]Department of Family Medicine and Emergency Medicine, Université de Sherbrooke, 3001 12 Ave N Immeuble X1, Sherbrooke, QC, Canada

[2]Applied Science and Medicine, School of Nursing, University of British Columbia, Vancouver, BC, Canada

[3]Centre for Health Services and Policy Research, School of Nursing, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Marie-Eve Poitras, PhD, RN
Department of Family Medicine and Emergency Medicine, Université de Sherbrooke, 3001 12 Ave N Immeuble X1, Sherbrooke, QC, Canada

Abstract

Background: Recruiting patients in primary care research remains challenging due to clinical workload, staffing constraints, and the need to limit disruption to routine care. Traditional recruitment methods often place a substantial burden on clinics, prompting research teams to adopt low-burden and scalable approaches such as email-based recruitment. Despite its growing use, limited empirical evidence describes how email recruitment campaigns are designed and how they perform when targeting primary care patients in real-world settings.

Objective: This study aims to descriptively examine engagement metrics from an email recruitment campaign targeting primary care patients.

Methods: We conducted a formative, descriptive secondary analysis of engagement metrics generated during a large-scale email recruitment campaign conducted as part of the Quebec component of the Patient-Reported Indicator Survey-Organisation for Economic Co-operation and Development survey. Between June 2023 and January 2024, a total of 12 primary care clinics invited eligible adult patients (aged ≥45 years) to complete an online survey using a standardized email template distributed via an email marketing platform. Collected engagement metrics included delivery rates, open rates, click-through rates, conversion rates, and device type. Analyses were descriptive and conducted at the clinic level.

Results: For 15,277 patients, invitations were successfully delivered to 14,758 (96.6%). The mean open rate for the initial invitation was 73.4% (10,828/14,757; range 57% - 88%), decreasing with reminders. Most emails were opened on computers (25,868/30,279, 85.4%). Out of a total of 445 emails, 42 (9.4%) were undelivered due to technical issues, and 403 (90.6%) were undelivered due to incorrect email addresses. The overall conversion rate was 9.7% (1430/14,758). Click-through rates varied by content, with the highest engagement observed for the survey link and lower engagement for supplementary video materials. Reminder emails substantially increased survey participation across clinics (759/1138, 66.7%). Participants who completed the questionnaire were predominantly aged 60 to 69 years, identified as female, and had completed postsecondary education.

Conclusions: This formative analysis suggests that email-based recruitment is a feasible and low-burden approach for engaging primary care patients in research. Engagement metrics offer valuable insights at the implementation level to inform the design, adaptation, and monitoring of digital recruitment strategies in real-world primary care settings. These findings provide practical, implementation-oriented insights to inform the design, refinement, and evaluation of email recruitment campaigns in primary care research.

(JMIR Hum Factors 2026;13:e67088) doi:10.2196/67088

KEYWORDS

web-based survey; primary care; email-based recruitment; patient recruitment; engagement metrics

Introduction

The increasing administrative burden and the insufficient workforce place significant pressure on Canadian primary care clinics [1,2]. The schedules of family physicians in primary care are overloaded, and their participation in research initiatives can be hindered by numerous factors, including mismatches between their needs and the research objectives, staff shortages,

XSL•FO
RenderX

administrative burdens, disruptions to clinical routines, and additional documentation requirements [3]. Consequently, their participation in research initiatives must be made as easy and straightforward as possible. Recruiting patients for research participation in clinics is highly challenging [4]. If clinic staff are required to dedicate time to explaining the initiative and responding to patients' questions, this fragments their clinical practice and disrupts the routine workflow. While this added burden is unavoidable, it necessitates the use of low-burden recruitment methods, moving away from traditional on-site or mail-based approaches. Despite many clinics showing interest in participating in research, engaging them remains challenging due to complex procedures, extensive paperwork, staffing shortages, and insufficient office space for recruitment and consent [3,5,6]. Ethical issues concerning the initial contact with patients are also at stake because it "should be made by someone that individuals would expect to have relevant information about them, or in other ways that do not inappropriately intrude on their life or privacy" [7].

Nearly 29 million Canadians use email, with approximately 20% opening marketing emails [8]. As a result, email has become a convenient and efficient method for recruiting research participants across various fields [4,9]. Recruitment emails typically outline the research purpose, potential benefits for participants, and any compensation or incentives that will be offered. Combined with a web-based survey, email recruitment enables primary care researchers to quickly reach a broad audience and track responses, while being less resource intensive than in-clinic or phone recruitment.

Beyond initial recruitment, sustaining participant engagement amidst the myriad of digital stimuli remains challenging. Capturing and maintaining the interest and commitment of potential participants requires deploying specific strategies. Innovative approaches grounded in empirical research are crucial for achieving optimal participant recruitment and engagement outcomes in navigating this complex landscape. Numerous factors have been shown to influence response rates, including shorter survey length, providing incentives, high relevance of the content and survey topic, and attractive visual display [10]. However, there is a lack of information on how patients in primary care engage when invited to participate in research via email. Most existing studies focus on email recruitment of health care professionals [11-13], and those describing email content in sufficient detail or using marketing platforms to monitor engagement metrics are rare.

In recent years, research teams have increasingly relied on email as a recruitment strategy, underscoring the need to better document how such campaigns are designed and how they perform in real-world settings. Accordingly, this study aims to describe the implementation and performance of an email-based recruitment campaign targeting primary care patients using engagement metrics collected from a large-scale survey. The findings are intended to provide practical, implementation-oriented insights to inform the design, adaptation, and monitoring of future email recruitment strategies in primary care.

Methods

Overview

This secondary analysis followed a prespecified descriptive analytic plan focused on characterizing engagement patterns at the clinic level. Primary outcomes included email delivery, open rates, click-through rates, device type, and conversion rates. No hypothesis testing was conducted, as the study objective was exploratory and formative. This paper presents data from reports and indicators gathered during the Patient-Reported Indicator Survey (PaRIS) national survey in Quebec, Canada. The PaRIS is an international survey led by the Organisation for Economic Co-operation and Development (OECD) to develop, standardize, and implement a new generation of indicators that measure the outcomes (patient-reported outcome measures [PROMs]) and experience of health care (patient-reported experience measure [PREMs]) most relevant to primary care patients worldwide. The PaRIS initiative has been described in detail elsewhere [14,15]. The OECD provided guidance on patient recruitment across participating countries, enabling each country to tailor its recruitment methods to its chosen approach, whether by phone, mail, email, or in-person at clinics. Quebec opted to recruit patients through a large-scale email campaign. The sample consisted of patients attending primary care clinics participating in the Canadian PaRIS-OECD national survey in the province of Quebec. The inclusion criteria for patients were as follows: (1) being 45 years or older, (2) having consulted their primary care clinician within the last 6 months, and (3) providing an email address. In accordance with the OECD's guidelines for the PaRIS study, the targeted sample size was established at 75 patients per clinic, although clinics could recruit more if feasible [15].

Our team, comprising a registered nurse and researcher, a research coordinator, a patient partner, and a scientific graphic designer, cocreated an invitation template for patient recruitment. STW and MEP are coleads for PaRIS-OECD in Canada. First, we identified the essential items to include in the email invitation. We then drafted the initial content, refining it before the scientific graphic designer further refined it to enhance its visual clarity and usability. As Canada has 2 official languages, French and English, we created bilingual email recruitment materials.

Building on the Canadian Institutes for Health Research's strategy for patient-oriented research [16] and our previous findings, which demonstrated that recruitment facilitated by a patient partner results in a higher participation rate, we incorporated a patient partner video clip in the email invitation. This 1-minute 50-second French video, featuring English subtitles, showcased a patient partner explaining the project's objectives and emphasizing the importance of participation. The scientific graphic was produced using Adobe Premiere Pro (version 25.4). The email template (Multimedia Appendix 1) was designed using the online marketing platform CyberImpact 2025, which provides delivery and engagement metrics and contains the following 5 sections:

1. Title of the study and catchy wording

XSL•FO
RenderX

2. Greetings and invitation purpose, including an introduction to the patient partner video and a link to it (research team's YouTube channel)
3. Links to French and English PaRIS study questionnaires and contact information to obtain support from the research team
4. Aim of the study, participants' contribution, and what is in it for participants
5. Two links leading to a video about PREMs and PROMs (research team's YouTube channel), previously produced by the research team, and greetings

We also created a reminder template with adjusted wording to include the reminder element and remove links to video clips discussing PREMs and PROMs, thereby streamlining the invitation template. Finally, we created support materials for participating clinics, including a PDF illustrating the steps for importing emails into the marketing platform and how to schedule email campaigns and several examples of email subjects such as "PaRIS Study-Survey on your care experience at [name of the clinic]," "[name of the clinic]: PaRIS research project," and "We need your opinion, please help us!" Each clinic was instructed to program 1 initial invitation and 3 reminders, all scheduled to be sent at 6 AM.

Participating primary care clinics obtained data from the statistics provided by CyberImpact and shared anonymized reports with the research team, as approved by the ethics committee. The following categories of data were used: mailing performance, opening statistics, and statistics on embedded links. Briefly, email open and click rates were measured using hidden tracking pixels. Specifically, the pixel was activated only when the images embedded in the email were downloaded, which helped reduce false openings generated by automated systems that scanned messages. A certain margin of error remained, for example, due to image blocking or user privacy settings. However, CyberImpact considers the reliability of these measures to be very high, estimating data accuracy at approximately 95%. We used the analytic data from YouTube to track metrics (number of views, average view duration, and average percentage of video viewed) for the patient partner introduction video.

The open rate was calculated as the percentage of patients who opened the email out of the total number of patients who received the email. The click-through rate refers to the percentage of people who click a link in an email out of those who open it. The conversion rate is the number of recipients who take the desired action compared with the number of recipients who received the email. The unsubscribe rate refers to the number of patients who unsubscribe out of the number of patients who received the email [17]. Quantitative data were analyzed using descriptive statistics, including measures of central tendency (mean) and variability (SD) using a 95% CI, as well as the minimum and maximum values of the variables' means across clinics, using SPSS (IBM Corp) software. We did not use a reporting guideline. Instead, we emphasized transparency in describing the data sources, metrics examined, unit of analysis, and descriptive analytic approach, consistent with the exploratory objectives of the study and the reporting standards commonly used for formative and implementation-focused research.

Ethical Considerations

This study was in accordance with the ethical standards of the responsible committee on human experimentation and with the World Medical Association Declaration of Helsinki. Ethics approval was obtained from the research ethics board of the Ministry of Health and Social Services of the province of Quebec (CCER-22-23). Patients gave their consent to participate in the PaRIS-OECD national study, and this consent covered secondary analysis. Patients participating in the PaRIS study in Quebec were eligible for a draw to win one of three US $25 grocery gift cards among respondents from their clinic. Demographic characteristics are presented in aggregate, and no identifiable participant information is included in this study.

Results

Between June 2023 and January 2024, a total of 12 primary care clinics participated in the PaRIS-OECD national survey in the province of Quebec. The research coordinator met with a staff member from each participating clinic to assist in programming and sending email invitations and reminders to patients. The clinic staff member then chose their preferred formulation for the email subject, either by creating their own subject line or using 1 of the provided examples. The clinic staff sent a first invitation to patients the day after the meeting. Email distribution was scheduled to accommodate each clinic's operational context, including avoidance of public holidays and alignment with concurrent recruitment activities. All clinics expressed that 3 reminders were excessive and that they did not want to oversolicit their patients. Accordingly, only 2 reminders were sent to comply with this request. All invitations and reminders were sent on weekdays at 6 AM, except for 1 clinic that preferred to send them at 8 AM to coincide with its opening hours.

Twelve primary care clinics sent the initial email invitation to 15,277 patients. Across clinics, 14,758 (96.6%) patients received the email invitation in their inbox, while 14,662 (96%) patients received the first reminder email, and 14,640 (95.8%) patients received the second reminder email. The number of invited patients per practice varied from 47 to 4366 (median 416.5, IQR 143.5-945). Combining the initial invitation and the 2 reminders, a total of 44,771 emails were sent, and 98.4% (n=44,060) were delivered. Patients who had already initiated or completed the questionnaire still received the reminder emails. A total of 629 emails were undelivered due to technical issues (n=119, 18.9%) or an incorrect email address (n=510, 81.1%). A total of 176 patients unsubscribed from the email invitation, corresponding to an unsubscribe rate of 1.2% (176/14,758). Patients provided some reasons explaining their unsubscription; 48 (27.3%) patients unsubscribed because they never subscribed to the emailing list, 34 (19.3%) patients unsubscribed because they no longer wished to receive the communications, and 13 (7.4%) patients unsubscribed because they received too many emails. A total of 7 (4%) patients reported that they did not like the content. Of the 176 patients, 76 (43.2%) patients who unsubscribed did not provide a reason.

XSL•FO
RenderX

A total of 0.3% (39/15,277) of patients contacted the research team by phone or email to obtain more information, validate their eligibility, or seek support with the web-based survey. Among them, 12 (30.7%) asked for a paper copy of the questionnaire.

For the initial invitation, the open rate was 73.4% (10,828/14,758), with a range of 57% to 88% (SD 6.57) across clinics. For the first reminder, the open rate dropped to 68% (9968/14,662; range 52%-86 %; SD 8.51), and for the last reminder, the open rate was 64.8% (9483/14,640), with a range of 47% to 84% (SD 10.09). Of the 30,279 opened emails, most invitation and reminder emails (n=25,868, 85.4%) were opened on a computer, 3803 (12.6%) invitation and reminder emails were opened on a mobile phone, and 643 (2.1%) invitation and reminder emails were opened on a tablet. The information was missing for 19 (0.1%) emails.

The highest click-through rate was 16% (range 7% - 27%), observed with the first reminder and linked to the French questionnaire. The video clip featuring the patient partner obtained the highest click-through rate in the first email invitation (8%; range 1-19) and was viewed 1700 times (direct URL source). The average view duration was 1 minute and 20 seconds, and the average percentage of the video viewed was 73%. The initial invitation and reminders were viewed 863 times in a web browser. Table 1 lists the click-through rates for each link included in the invitations.

Table . Click-through rate for invitation sections.

Link to	Click rate[a] (range)[b] for the first email (n=10,828)	Click rate (range) for the first reminder (n=9968)	Click rate (range) for the second reminder (n=9483)
Patient partner video clip	8.1% (1.2%-18.9%)	5.3% (0.8%-11.7%)	4.0% (0.0%-7.8%)
French Patient-Reported Indicator Survey questionnaires	15.2% (5.3%-25.7%)	16.6% (6.8%-27.5%)	13.9% (3.8%-20.0%)
English Patient-Reported Indicator Survey questionnaires	2.1% (0.0%-14.9%)	0.6% (0.0%-2.9%)	0.4% (0.0%-2.2%)
View invitation in a web browser	4.8% (0.8%-7.1%)	2.8% (0.0%-9.5%)	4.0% (0-8.6%)
Video clip about patient-reported experience measures	1.3% (0.0%-2.7%)	N/A[c]	N/A
Video clip about patient-reported outcomes measures	0.6% (0.0%-2.7%)	N/A	N/A

[a]Number of patients who clicked on the link/number of patients who opened the invitation.

[b]Minimum and maximum click rate across clinics.

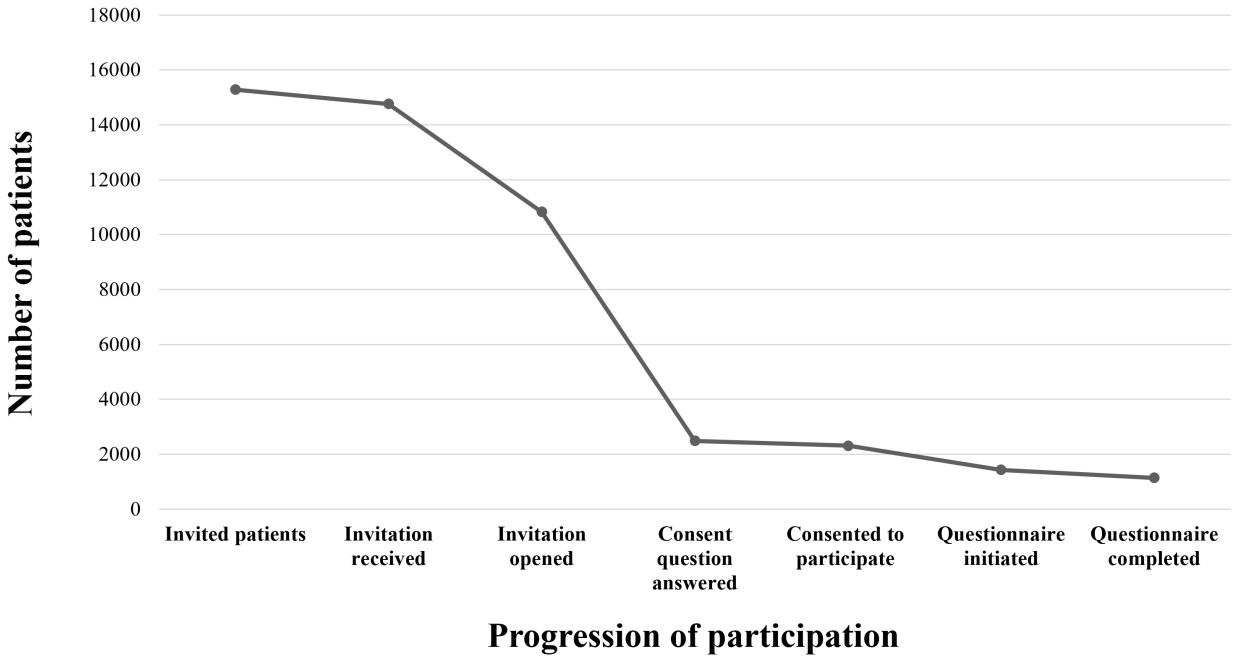
[c]N/A: not applicable. The link to the video clip was not included in the reminders.

In this study, the conversion rate was 9.7% (1430/14,758), defined as the proportion of patients who initiated the questionnaire (n=1430) among those who received the initial email invitation (n=14,758).

Figure 1 illustrates the flow of recruited patients from the initial email invitation to the completion of the web-based survey. Across the initial invitation and the 2 reminders, the emails were opened 30,279 times; however, it was not possible to determine whether these were unique individuals, as the same person could have opened both the initial invitation and 1 or more reminder emails. A total of 2486 patients opened the Research Electronic Data Capture (REDCap; Vanderbilt University) survey and answered the consent question. Among them, 2307 (92.8%) patients provided their consent to participate in the study, and 1430 (62%) initiated the questionnaire. From them, 1138 (79.6%) patients completed all the included items.

XSL•FO
RenderX

Figure 1. Recruitment and retention across study phases.



Across clinics, the first reminder was sent on average on day 10 (range across clinics: 7 - 16 days from initial invitation), and the second reminder was sent on average on day 22 (range across clinics: 3-22 days after the first reminder). Across clinics, questionnaire completion rates increased with the use of reminder emails, with 379 (2.6%) out of 14,758 questionnaires completed after the initial invitation, 437 (3%) out of 14,662 after the first reminder, and 322 (2.2%) out of 14,640 after the second reminder (Table 2). Completion rates varied substantially between clinics, and the inclusion of reminder emails resulted in an approximate 200% (759/379) increase in the cumulative number of completed questionnaires.

Table . Questionnaire completion per clinic after initial invitation and reminders.

Clinic number	Patients who completed the questionnaire after receiving the initial invitation, % (n/N)	Patients who completed the questionnaire after receiving the first reminder, % (n/N)	Patients who completed the questionnaire after receiving the second reminder, % (n/N)
2401	1.7 (16/966)	2.3 (22/955)	4 (34/949)
2402	3.4 (47/1364)	4.6 (63/1361)	2.5 (34/1356)
2403	3.9 (7/178)	2.2 (4/178)	1.1 (2/177)
2405	5 (42/849)	4.8 (41/850)	3.1 (26/850)
2407	3.6 (30/838)	5.0 (42/833)	4.3 (36/832)
2408	5.2 (25/481)	4.6 (22/476)	4.8 (23/476)
2409	3.8 (87/2272)	4.5 (103/2271)	3.4 (78/2264)
2410	3.6 (46/1273)	4 (50/1267)	2.4 (31/1265)
2414	0 (0/83)	7.2 (6/83)	7.2 (6/83)
2418	2.4 (45/1908)	2.8 (53/1903)	1.4 (27/1900)
2419	0.7 (29/4366)	0.6 (24/4348)	0.4 (19/4310)
2420	2.8 (5/179)	5.1 (7/137)	3.4 (6/178)
Total	2.6 (379/14,758)	3 (437/14,662)	2.2 (322/14,640)

We examined the characteristics of patients who completed the sociodemographic section of the questionnaire (Table 3). Respondents were mainly aged between 60 and 69 years, identified as female, had a college or university diploma, and were of White ethnicity.

XSL•FO
RenderX

Table . Sociodemographic characteristics of primary care patients who completed the sociodemographic section of the questionnaire (N=1426).

Characteristics	Patients, n (%)
Age (years)	
<45[a]	118 (8.3)
45 - 49	139 (9.7)
50 - 69	287 (20.1)
60 - 69	485 (34.0)
≥70	397 (27.9)
Gender	
Men	444 (37.9)
Women	723 (61.8)
Nonbinary	3 (0.3)
Education	
Partial high school or less	60 (4.2)
High school diploma	191 (13.4)
Partial college	123 (8.6)
College or university degree	1027 (72.1)
Missing	25 (0.02)
Ethnicity[b]	
Black	16 (1.4)
Métis, Inuit, and First Nations	15 (1.3)
White	1098 (93.8)
Other	27 (3.5)

[a]One clinic made a mistake when extracting patients' email addresses. Patients aged younger than 45 years received the invitation, and some patients responded to the questionnaire, despite being ineligible for the study.

[b]Ethnicity categories were determined by the Organisation for Economic Co-operation and Development and reported as found in the questionnaire.

Discussion

Principal Findings

Overall, this formative secondary analysis provides practical insights into how primary care patients engage with email-based research invitations in real-world clinical contexts. Our findings suggest that email invitations are an efficient and effective method for reaching a large number of primary care patients. High delivery and open rates suggest that emails sent on behalf of trusted primary care clinics are effective at reaching patients. Moreover, conversion rates indicate that a meaningful proportion of recipients proceeded to survey participation. These findings support the feasibility of email recruitment as a low-burden approach for clinics and research teams. These results lead us to the following observations: (1) email is a highly feasible recruitment method in primary care settings; (2) email makes it possible to reach older patients, but equity issues persist; and (3) commitment varies greatly depending on the clinic, which indicates that the organizational and relational context matters.

We found that a large-scale email campaign using a marketing platform effectively reached a significant number of primary care patients. The high delivery rate of 96.6% (14,758/15,277) is consistent with previous studies [9,18]. Only a few email addresses were incorrect, indicating that patients generally provided valid addresses to their clinic. As expected, reminders were effective for primary care patients recruited by their clinic [10]. Unlike paper questionnaires sent by traditional mail, which remain physically present and serve as a constant reminder, digital questionnaires do not provide a lasting prompt [19]. A high proportion of primary care patients (10,828/14,758, 73.4%) opened the invitation, which is high compared to 35% to 51% rates reported in other studies [20,21]. The average open rate in email marketing campaigns related to health care services sent by the industry is 23% [22]. Within the email marketing industry, open rates above 50% are considered outstanding, typically reflecting the presence of a highly loyal or well-targeted niche audience [23]. This may be attributed to the sender's email address being from a reputable authority (ie, a primary care practice), which could lead patients to perceive it as more relevant [24] and reduce the likelihood of them mistaking our invitation for spam or a fraudulent solicitation. Indeed, participants may hesitate to open emails from unknown or suspicious addresses [9,25]. Additionally, the clinic used a catchy email subject to capture patients' attention, which may have contributed to the high open rate [8]. Targeted email campaigns can also enhance recruitment effectiveness, as all invited patients were directly concerned with the study topic

XSL•FO
RenderX

[26]. This confirms the organizational feasibility of large-scale recruitment campaigns while imposing minimal burden on clinics.

Our email invitations achieved high engagement indicators, successfully engaging more than 1000 primary care patients. The conversion rate of our email recruitment strategy was 9.7% (1430/14,758), which is higher than the rates typically reported in similar studies (2.9%-4%) [9,24,25,27-29]. General email marketing campaigns in Canada average a conversion rate of less than 3% [8]. This may also be attributed to the clear division of materials into sections, allowing patients to quickly assess the relevance and interest of the information provided [18]. Including multimedia elements, such as illustrations and video clips, likely contributed to higher performance [18]. A total of 1512 clicks were recorded on the patient partner video link, which suggests that this content seems to be of interest to patients. Moreover, simply knowing a patient partner was involved in the study and made a video may have reassured some patients and added external validity [30]. However, the study design does not allow conclusions regarding its causal impact on participation. These findings suggest that while cocreation may enhance acceptability and trust, its contribution should be understood as contextual and supportive rather than as a measurable driver of engagement. This strategy, along with the contribution of patient partners in research and recruitment, is highly recommended as best practice guidelines for patient-oriented research [16,31].

The age range of patients who responded to the survey was broad. Notably, patients aged 70 years or older also answered the questionnaire, suggesting that email can be an effective method for recruiting older patients. This observation aligns with the increasing internet use among older adults in Canada over recent years, as the proportion of those aged 75 years or older using the internet rose from 62% in 2020 to 72% in 2022 [32]. The online survey received a higher response rate from women, which is consistent with comparable studies [33-37]. Additionally, because patients were invited if they had an encounter with their health care professional within the previous 6 months, and women tend to consult primary care more frequently than men, this may have led to a higher proportion of women being invited to participate in the survey [28,38]. For ethical and privacy reasons, our team did not have access to the identity of admissible invited patients and, consequently, to their sociodemographic characteristics. We also observed a low number of responses among people of non-White ethnic origin. Due to the unavailability of data on the ethnic origins of the invited patients, it is not feasible to calculate or compare specific response rates by ethnicity. However, the literature suggests that ethnicity may be a source of nonresponse bias [39,40]. Heerman et al [28] observed similar results, finding that technology-enabled recruitment approaches are not the most efficient for enrolling non-White patients. Therefore, it is essential to consider strategies to enhance the participation of these underrepresented groups in health surveys, primarily because it is known that ethnic groups other than White have a less positive primary care experience [41].

Engagement metrics varied substantially across participating primary care clinics, including differences in open rates, click-through rates, and conversion rates. These variations suggest that a strictly standardized recruitment protocol may not be optimal in all primary care contexts. Although a common email template and overall recruitment strategy were used, clinics adapted key implementation elements—such as timing of invitations, number of reminders, and communication practices—to better align with their organizational context and patient population. This adaptability appears necessary to maintain acceptability and feasibility, highlighting that digital recruitment strategies in primary care may benefit from a balance between standardization (to ensure efficiency and comparability) and local tailoring (to reflect clinic-specific realities and patient expectations). These findings emphasize the importance of context-sensitive implementation when deploying digital recruitment approaches in real-world primary care settings.

This study should be interpreted as a formative, implementation-oriented analysis rather than a test of behavioral theory or recruitment effectiveness. The use of descriptive engagement metrics offers actionable information for research teams seeking to monitor, refine, and adapt digital recruitment strategies over time. Future studies could build on this work by incorporating comparative designs, experimental testing of message components, or qualitative inquiry to elucidate mechanisms of engagement further.

Limitations

This study had a few limitations that must be underscored. Ideally, participation invitations would have been sent to all patients across clinics simultaneously, and the time between reminders would have been consistent across clinics, allowing for a deeper exploration of factors influencing response rates across clinics. Ideally, to assess the effect of the various strategies embedded in the email, different versions of the message should have been sent to separate groups of patients (for instance, a plain or less visually appealing email or the same message without the patient partner video). However, within the PaRIS study context, it was not feasible to send potentially less-engaging messages to patients, as the primary goal was to maximize participant recruitment. Due to the PaRIS study's design, objectives, and ethical constraints, we could access only the sociodemographic characteristics of patients who responded to the questionnaire, not those who received the email invitation. Sample representativeness is a major concern, given that only 9.7% (1430/14,758) of the invited patients partially or fully completed the questionnaire. Additionally, staff from some clinics discussed the study with some patients, which may have contributed to a higher conversion rate. However, this intervention was not measured, and its effect cannot be assessed. Our study also revealed variation in indicators across clinics, demonstrating that patients do not all respond similarly to email invitations. These differences may be attributed to the type of clinic (private, public, or university-affiliated) and the characteristics of the patients enrolled in the study. However, our study lacks sufficient power to determine these characteristics.

XSL•FO
RenderX

Conclusions

This formative, descriptive secondary analysis documents the functioning of email-based recruitment in real-world primary care research settings. Using routinely collected engagement metrics, we show that emails sent on behalf of trusted primary care clinics are a feasible and low-burden approach for reaching patients and supporting participation in web-based surveys. These findings offer practical, implementation-oriented insights for research teams seeking scalable recruitment approaches. They can inform the design, refinement, and ongoing monitoring of digital recruitment strategies in future primary care studies.

Acknowledgments

The authors would like to acknowledge Catherine Deschenes for her support in extracting data from the emailing platform reports.

Funding

The Academic Research Chair on Optimal Professional Pratices in Primary Care (grant 2020-2025 to MEP) funded this study, and Health Canada funded the Patient-Reported Indicator Survey (PaRIS)-Canada initiative. The PaRIS-Quebec component was funded by the Réseau en recherche en sciences infirmières, the Réseau-1 Québec, and the Centre intégré universitaire en santé et services sociaux du Saguenay-Lac-Saint-Jean. The funders had no role in the conceptualization, design, data collection, analysis, or other aspects of the study.

Data Availability

The datasets used or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

VTV had the original idea for this manuscript and wrote it. She coordinates the Patient-Reported Indicator Survey (PaRIS) study and is responsible for data collection and analysis in the province of Quebec. MDP is a patient partner and shares responsibility for the study in Quebec. She was involved in designing the invitation template and revising the manuscript. MEP is the study's principal investigator and revised and improved the manuscript. AF is a scientific graphic designer and designed the invitation template and the patient partner video. STW is the principal investigator of the national PaRIS survey in Canada and reviewed the manuscript. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Copy of the standardized email template used to invite eligible patients to participate in the study.
[PDF File, 310 KB - humanfactors_v13i1e67088_app1.pdf]

References

1. Rao SK, Kimball AB, Lehrhoff SR, et al. The impact of administrative burden on academic physicians: results of a hospital-wide physician survey. Acad Med 2017 Feb;92(2):237-243. [doi: 10.1097/ACM.0000000000001461] [Medline: 28121687]
2. Canadian Medical Association. Third-party forms (update 2017). The Association. 2017. URL: https://policybase.cma.ca/link/policy13643 [accessed 2025-11-01]
3. Wangler J, Jansky M. Primary care involvement in clinical research - prerequisites, motivators, and barriers: results from a study series. Arch Public Health 2024 Mar 20;82(1):41. [doi: 10.1186/s13690-024-01272-x] [Medline: 38504310]
4. Frampton GK, Shepherd J, Pickett K, Griffiths G, Wyatt JC. Digital tools for the recruitment and retention of participants in randomised controlled trials: a systematic map. Trials 2020 Jun 5;21(1):478. [doi: 10.1186/s13063-020-04358-3] [Medline: 32498690]
5. Messner DA, Moloney R, Warriner AH, Wright NC, Foster PJ, Saag KG. Understanding practice-based research participation: the differing motivations of engaged vs. non-engaged clinicians in pragmatic clinical trials. Contemp Clin Trials Commun 2016 Aug 23;4:136-140. [doi: 10.1016/j.conctc.2016.08.003] [Medline: 29736476]
6. Loskutova NY, Smail C, Ajayi K, Pace WD, Fox CH. Recruiting primary care practices for practice-based research: a case study of a group-randomized study (TRANSLATE CKD) recruitment process. Fam Pract 2018 Jan 16;35(1):111-116. [doi: 10.1093/fampra/cmx064] [Medline: 28985294]
7. CIHR best practices for protecting privacy in health research. Canadian Institutes of Health Research. 2005. URL: https://cihr-irsc.gc.ca/e/documents/et_pbp_nov05_sept2005_e.pdf [accessed 2026-02-18]
8. Bush O. Email marketing statistics in Canada. Made in CA. 2026. URL: https://madeinca.ca/email-marketing-statistics-canada/#:~:text=The%20email%20penetration%20rate%20in,rate%20is%20less%20than%203%25 [accessed 2026-02-18]

XSL•FO
RenderX

9. Ng MY, Olgin JE, Marcus GM, Lyles CR, Pletcher MJ. Email-based recruitment into the Health eHeart Study: cohort analysis of invited eligible patients. J Med Internet Res 2023 Dec 22;25:e51238. [doi: 10.2196/51238] [Medline: 38133910]
10. Dillman DA, Smiyth JD, Christian LM. Internet, Phone, Mail, and Mixed - Mode Surveys: The Tailored Design Method, 4th edition: John Wiley & Sons; 2014. [doi: 10.1002/9781394260645]
11. Murphy CC, Craddock Lee SJ, Geiger AM, et al. A randomized trial of mail and email recruitment strategies for a physician survey on clinical trial accrual. BMC Med Res Methodol 2020 May 19;20(1):123. [doi: 10.1186/s12874-020-01014-x] [Medline: 32429848]
12. Mazurenko O, Sanner L, Apathy NC, et al. Evaluation of electronic recruitment efforts of primary care providers as research subjects during the COVID-19 pandemic. BMC Prim Care 2022 Apr 28;23(1):95. [doi: 10.1186/s12875-022-01705-y] [Medline: 35484491]
13. Harrap B, Taylor T, Russell G, Scott A. A randomised controlled trial of email versus mailed invitation letter in a national longitudinal survey of physicians. PLoS ONE 2023 Aug 22;18(8):e0289628. [doi: 10.1371/journal.pone.0289628] [Medline: 37607168]
14. de Boer D, van den Berg M, Ballester M, et al. Assessing the outcomes and experiences of care from the perspective of people living with chronic conditions, to support countries in developing people-centred policies and practices: study protocol of the International Survey of People Living with Chronic Conditions (PaRIS survey). BMJ Open 2022 Sep 19;12(9):e061424. [doi: 10.1136/bmjopen-2022-061424] [Medline: 36123114]
15. Does healthcare deliver? Results from the Patient-Reported Indicator Surveys (PaRIS). : OECD Publishing; 2025 URL: https://www.oecd.org/en/publications/2025/02/does-healthcare-deliver_978507f1.html [accessed 2026-02-18]
16. Stratégie de recherche axée sur le patient - cadre d'engagement des patients. Instituts de Rechercheen Santé du Canada. URL: https://cihr-irsc.gc.ca/f/48413.html [accessed 2026-02-18]
17. Email marketing terms. Campaign Monitor by Marigold. URL: https://www.campaignmonitor.com/resources/glossary/email-marketing-conversion-rate/ [accessed 2024-04-02]
18. Kirsch K. The ultimate list of email marketing stats that we think you should know. HubSpot. URL: https://blog.hubspot.com/marketing/email-marketing-stats#highligh [accessed 2026-02-18]
19. Becker R. Short- and long-term effects of reminders on panellists' survey participation in a probability-based panel study with a sequential mixed-mode design. Qual Quant 2023 Oct;57(5):4095-4119. [doi: 10.1007/s11135-022-01554-y]
20. Baca-Motes K, Edwards AM, Waalen J, et al. Digital recruitment and enrollment in a remote nationwide trial of screening for undiagnosed atrial fibrillation: lessons from the randomized, controlled mSToPS trial. Contemp Clin Trials Commun 2019 Jun 7;14:100318. [doi: 10.1016/j.conctc.2019.100318] [Medline: 30656241]
21. Nasers C. Performance benchmarks for healthcare marketing emails. 9 Clouds. 2021 Jun 18. URL: https://9clouds.com/blog/performance-benchmarks-for-healthcare-marketing-emails/?utm_source=chatgpt.com [accessed 2026-02-19]
22. Ultimate email marketing benchmarks for 2022: by industry and day. Campaign Monitor by Marigold. URL: https://www.campaignmonitor.com/resources/guides/email-marketing-benchmarks/ [accessed 2024-03-15]
23. Jimenez A. Email marketing benchmarks by industry (+ email open rate data). HubSpot. URL: https://blog.hubspot.com/sales/average-email-open-rate-benchmark [accessed 2025-10-15]
24. Bennett WL, Bramante CT, Rothenberger SD, et al. Patient recruitment into a multicenter clinical cohort linking electronic health records from 5 health systems: cross-sectional analysis. J Med Internet Res 2021 May 27;23(5):e24003. [doi: 10.2196/24003] [Medline: 34042604]
25. Koo M, Skinner H. Challenges of internet recruitment: a case study with disappointing results. J Med Internet Res 2005 Mar 19;7(1):e6. [doi: 10.2196/jmir.7.1.e6] [Medline: 15829478]
26. Morgan AJ, Jorm AF, Mackinnon AJ. Internet-based recruitment to a depression prevention intervention: lessons from the Mood Memos study. J Med Internet Res 2013 Feb 12;15(2):e31. [doi: 10.2196/jmir.2262] [Medline: 23403043]
27. Goldman V, Dushkin A, Wexler DJ, et al. Effective recruitment for practice-based research: lessons from the REAL HEALTH-Diabetes Study. Contemp Clin Trials Commun 2019 May 1;15:100374. [doi: 10.1016/j.conctc.2019.100374] [Medline: 31193095]
28. Heerman WJ, Jackson N, Roumie CL, et al. Recruitment methods for survey research: findings from the Mid-South Clinical Data Research Network. Contemp Clin Trials 2017 Nov;62:50-55. [doi: 10.1016/j.cct.2017.08.006] [Medline: 28823925]
29. Paquin RS, Lewis MA, Harper BA, et al. Outreach to new mothers through direct mail and email: recruitment in the Early Check research study. Clin Transl Sci 2021 May;14(3):880-889. [doi: 10.1111/cts.12950] [Medline: 33382929]
30. Kemper CJ. External validity. In: Zeigler-Hill V, Shackelford TK, editors. Encyclopedia of Personality and Individual Differences: Springer; 2020. [doi: 10.1007/978-3-319-24612-3_1303]
31. Poitras ME, Godbout I, Vaillancourt VT, et al. Step-by-step strategies for an integrated patient-oriented research: lessons learned from a multicentered study. Sci Nurs Health Pract 2020;3(2):1-9. [doi: 10.31770/2561-7516.1068]
32. Canadian internet use survey, 2022. Government of Canada. 2023. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/230720/dq230720b-eng.htm [accessed 2026-02-19]
33. Otufowora A, Liu Y, Young H 2nd, et al. Sex differences in willingness to participate in research based on study risk level among a community sample of African Americans in north central Florida. J Immigr Minor Health 2021 Feb;23(1):19-25. [doi: 10.1007/s10903-020-01015-4] [Medline: 32328873]

XSL•FO
RenderX

34. Smith G. Does gender influence online survey participation? A record-linkage analysis of university faculty online survey response behavior. San José State University. 2008. URL: https://scholarworks.sjsu.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1003&context=elementary_ed_pub [accessed 2026-02-19]
35. Langlais B, Mazza GL, Scherber RM, et al. Impact of imbalanced gender participation in online myeloproliferative neoplasm symptom surveys. J Clin Oncol 2022 Jun 2;40(16_suppl):e19078. [doi: 10.1200/JCO.2022.40.16_suppl.e19078]
36. Fowler FJJ, Cosenza C, Cripps LA, Edgman-Levitan S, Cleary PD. The effect of administration mode on CAHPS survey response rates and results: a comparison of mail and web-based approaches. Health Serv Res 2019 Jun;54(3):714-721. [doi: 10.1111/1475-6773.13109] [Medline: 30656646]
37. Slater M, Kiran T. Measuring the patient experience in primary care: comparing e-mail and waiting room survey delivery in a family health team. Can Fam Physician 2016 Dec;62(12):e740-e748. [Medline: 27965350]
38. Thompson AE, Anisimowicz Y, Miedema B, Hogg W, Wodchis WP, Aubrey-Bassler K. The influence of gender and other patient characteristics on health care-seeking behaviour: a QUALICOPC study. BMC Fam Pract 2016 Mar 31;17:38. [doi: 10.1186/s12875-016-0440-0] [Medline: 27036116]
39. Griffin DH. Measuring survey nonresponse by race and ethnicity. : United States Bureau of the Census; 2002 URL: https://www.census.gov/content/dam/Census/library/working-papers/2002/acs/2002_Griffin_01.pdf [accessed 2026-02-19]
40. Berlin NL, Hamill JB, Qi J, Kim HM, Pusic AL, Wilkins EG. Nonresponse bias in survey research: lessons from a prospective study of breast reconstruction. J Surg Res 2017 Dec 22;224:112-120. [doi: 10.1016/j.jss.2017.11.058] [Medline: 29506826]
41. Lyratzopoulos G, Elliott M, Barbiere JM, et al. Understanding ethnic and other socio-demographic differences in patient experience of primary care: evidence from the English General Practice Patient Survey. BMJ Qual Saf 2012 Jan;21(1):21-29. [doi: 10.1136/bmjqs-2011-000088] [Medline: 21900695]

Abbreviations

OECD: Organisation for Economic Co-operation and Development
PaRIS: Patient-Reported Indicator Survey
REDCap: Research Electronic Data Capture

Edited by A Kushniruk; submitted 01.Oct.2024; peer-reviewed by A Peytremann, GL Mazza, NV Kalckreuth; accepted 09.Feb.2026; published 06.Mar.2026.

Please cite as:
Vaillancourt VT, Poirier MD, Fournier A, Wong ST, Poitras ME
Design and Performance of an Email-Based Patient Recruitment Campaign in Primary Care Research: Formative Secondary Analysis
JMIR Hum Factors 2026;13:e67088
URL: https://humanfactors.jmir.org/2026/1/e67088
doi:10.2196/67088

© Vanessa T Vaillancourt, Marie-Dominique Poirier, Amélie Fournier, Sabrina T Wong, Marie-Eve Poitras. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 6.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Usability Across 3 mHealth Problem-Solving Training Interventions for Diverse Neurodevelopmental and Neurological Populations: Multicase Usability Evaluation

David Ogundairo[1], MEd; Shannon Juengst[2,3], PhD; Avani Modi[4,5], PhD; Shari Wade[4,6], PhD; Matthew Schmidt[1,7], PhD

[1]Department of Workforce Education and Instructional Technology, Mary Frances Early College of Education, University of Georgia, Rivers Crossing 110, Athens, GA, United States

[2]Brain Health and Rehabilitation Research Center, TIRR Memorial Hermann, Houston, TX, United States

[3]Department of Physical Medicine & Rehabilitation, UT Health Houston, Houston, TX, United States

[4]College of Medicine, University of Cincinnati, Cincinnati, OH, United States

[5]Division of Behavioral Medicine and Clinical Psychology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

[6]Division of Physical Medicine and Rehabilitation, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

[7]Department of Clinical and Administrative Pharmacy, College of Pharmacy, University of Georgia, Athens, GA, United States

Corresponding Author:
David Ogundairo, MEd
Department of Workforce Education and Instructional Technology, Mary Frances Early College of Education, University of Georgia, Rivers Crossing 110, Athens, GA, United States

Abstract

Background: mHealth (mobile health) interventions that integrate psychoeducation with structured problem-solving training (PST) hold strong potential for improving self-management of chronic conditions. Evaluating the usability of these interventions requires assessing technological, pedagogical, and sociocultural fit. However, most usability evaluations remain narrowly technocentric, focusing on interface-level metrics while neglecting pedagogical coherence, cultural responsiveness, and patient learning needs.

Objective: This exploratory study aimed to characterize usability challenges and facilitators across 3 psychoeducational mHealth PST interventions and to identify technological, pedagogical, and sociocultural design features that can improve engagement, accessibility, and implementation for diverse users.

Methods: This study used an exploratory, multimethod, collective case study design. Three independent mHealth PST usability studies used structured think-aloud protocols, with sessions conducted remotely via Zoom (Zoom Communications, Inc), except for 1 in-person session for Epilepsy Journey 2.0. A total of 14 participants were enrolled across 3 independent cases via purposive sampling from their respective target populations. Case 1 – Epilepsy Journey 2.0 (n=6; 4 male, 2 female; ages 12 - 18 y; adolescents with epilepsy), case 2 – Survivor's Journey (n=3; 1 male, 2 female; ages 18 - 30 y; adolescent and young adult survivors of brain tumor), and case 3 – electronic problem-solving training (n=5; 2 male, 3 female; ages 30 - 65 y; adults with a history of severe traumatic brain injury). Participants completed a presession technology comfort survey and the Comprehensive Assessment of Usability for Learning Technologies postsession. All sessions were recorded, transcribed, and analyzed thematically. Comprehensive Assessment of Usability for Learning Technologies data were analyzed using descriptive quantitative methods.

Results: Electronic problem-solving training demonstrated the highest usability (mean 87.96, SD 13.73), followed by Survivor's Journey (mean 83.00, SD 7.70), and then Epilepsy Journey 2.0 (mean 79.00, SD 12.75). Findings revealed that usability in health care learning design is shaped by how effectively the technology, learning content, and contextual factors align with patients' needs. Recurring challenges across interventions included unclear navigation, poor mobile responsiveness, instructional ambiguity, insufficient feedback, potential for greater inclusivity, and limited error recovery. Twelve cross-case design principles were derived, emphasizing mobile-first accessibility, cognitive load reduction, context-sensitive feedback, and empathetic, inclusive design.

Conclusions: Usability challenges in mHealth PST interventions arise not only from interface-level issues but also from how effectively the intervention supports users' understanding, decision-making, and real-world application demands. This extends prior mHealth usability research by demonstrating that user difficulties often reflect misalignments between technological features, instructional structure, and the everyday contexts in which individuals engage with PST. The resulting design principles highlight specific, actionable priorities for developers, including mobile-first optimization, clearer task scaffolding, and better feedback

XSL•FO
RenderX

and error recovery. Future work should evaluate these principles in larger samples and clinical settings to determine their impact on engagement, adherence, and downstream health outcomes.

(JMIR Hum Factors 2026;13:e90379) doi:10.2196/90379

KEYWORDS

usability; mobile health; psychoeducation; problem solving; digital health intervention; sociotechnical systems; user experience; mobile phone

Introduction

Background

Digital health interventions have become foundational to modern health care, offering accessible, scalable, and patient-centered models of care [1,2]. Among the most promising are mHealth (mobile health) interventions that integrate psychoeducation with structured problem-solving training (PST). Grounded in social-cognitive and self-regulation frameworks, these interventions equip patients not only with disease-specific knowledge but also with step-by-step strategies for identifying barriers, generating solutions, and applying them to real-world health challenges [3-6]. Across diverse populations, PST-based mHealth interventions have demonstrated consistent benefits, including reductions in depressive and anxiety symptoms, improvements in medication adherence, and increases in caregiver self-efficacy [6-8].

Usability Is Central to the Success of mHealth Interventions

While rigorous, evidence-based design constitutes the methodological bedrock of digital therapeutics, usability, a central dimension of patient experience, is pivotal for uptake and sustained engagement, which in turn mediate clinical effectiveness [9-11]. Usability refers to how effectively, efficiently, and satisfactorily patients can interact with technology to achieve their desired goals [12]. The ISO defines usability as the extent to which a system, product, or service can be used by specified users to achieve a specified goal, encompassing five attributes: effectiveness, efficiency, satisfaction, freedom from risk, and context coverage [13]. Well-designed mHealth interventions that prioritize usability and patients' experiences are more likely to be adopted, maintained, and found beneficial by patients over time [14]. Conversely, poor usability can lead to frustration, disengagement, and abandonment, directly undermining any clinical benefit the intervention might offer [15]. For example, high-school testers in the LIFE4YOUth study reported that poor functionality and difficult navigation made them cease using the app altogether [16], while a systematic review of 60 studies on electronic personal health records and digital mental health tools consistently identifies poor usability as a primary driver of low adoption and high dropout [17,18]. As PST mHealth interventions are complex, serve diverse patient populations, and play a central role in self-management, comprehensive usability evaluations are increasingly regarded as indispensable for designing effective digital tools that support behavior change, chronic disease management, and mental health outcomes [19,20].

General Usability Measures Are Limited for Evaluating mHealth Psychoeducation

Despite growing recognition of usability's importance, most digital health studies rely on general human-computer interaction instruments for usability evaluation [21,22]. The most widely used is the System Usability Scale (SUS) [23], a 10-item questionnaire yielding a composite score from 0 to 100 that captures users' overall perceptions of ease of use and learnability; scores around 50 are generally considered marginal, scores around 70 are considered good, and scores above 85 are considered excellent [24]. Other common instruments include the Computer System Usability Questionnaire [25] and the Software Usability Measurement Inventory [26]. While these tools can efficiently capture surface-level interface mechanics such as ease of use, interface flaws, and navigation flow, they were not designed to evaluate pedagogical alignment, social dynamics, or cultural responsiveness [22,27]. These dimensions are especially important for psychoeducational mHealth interventions, which share structural features with learning technologies (eg, modularized content, formative assessments and feedback, and self-directed practice) and require more than interface usability to support meaningful health behavior change.

Narrowly focused mHealth psychoeducation usability evaluation presents a significant limitation in light of the layered interactions between content, patient needs, and learning objectives in these interventions. Indeed, evidence shows that strong perceived technological usability does not guarantee meaningful impact. In a diabetes self-management study, participants rated the app high on the SUS (mean 80.5, 95% CI 72.3 to 88.7), yet struggled with key tasks, recording the most errors and lowest success rates on activities essential for effective self-management [28]. A pilot randomized trial of an osteoarthritis physical-activity app likewise found acceptable usability ratings (mean 71.3), but found no significant improvement over usual care for most outcomes [29]. Similarly, a usability study of home blood-pressure and pulse-oximeter devices reported satisfactory SUS scores for both a blood pressure monitor (mean 72.4, 95% CI 68.8 to 76.0) and a pulse oximeter (mean 71.52, 95% CI 68.7 to 74.3), despite users making frequent handling errors and some abandoning the devices altogether [30]. These findings illustrate that high usability ratings can coexist with limited learning, behavior change, or health benefits, suggesting a need for more holistic approaches to usability capable of not only evaluating technological aspects, but also aspects such as patient relevance, learning effectiveness, and social acceptability.

XSL•FO
RenderX

Addressing Usability Gaps in Health Care Learning Design Through Multidimensional Evaluation

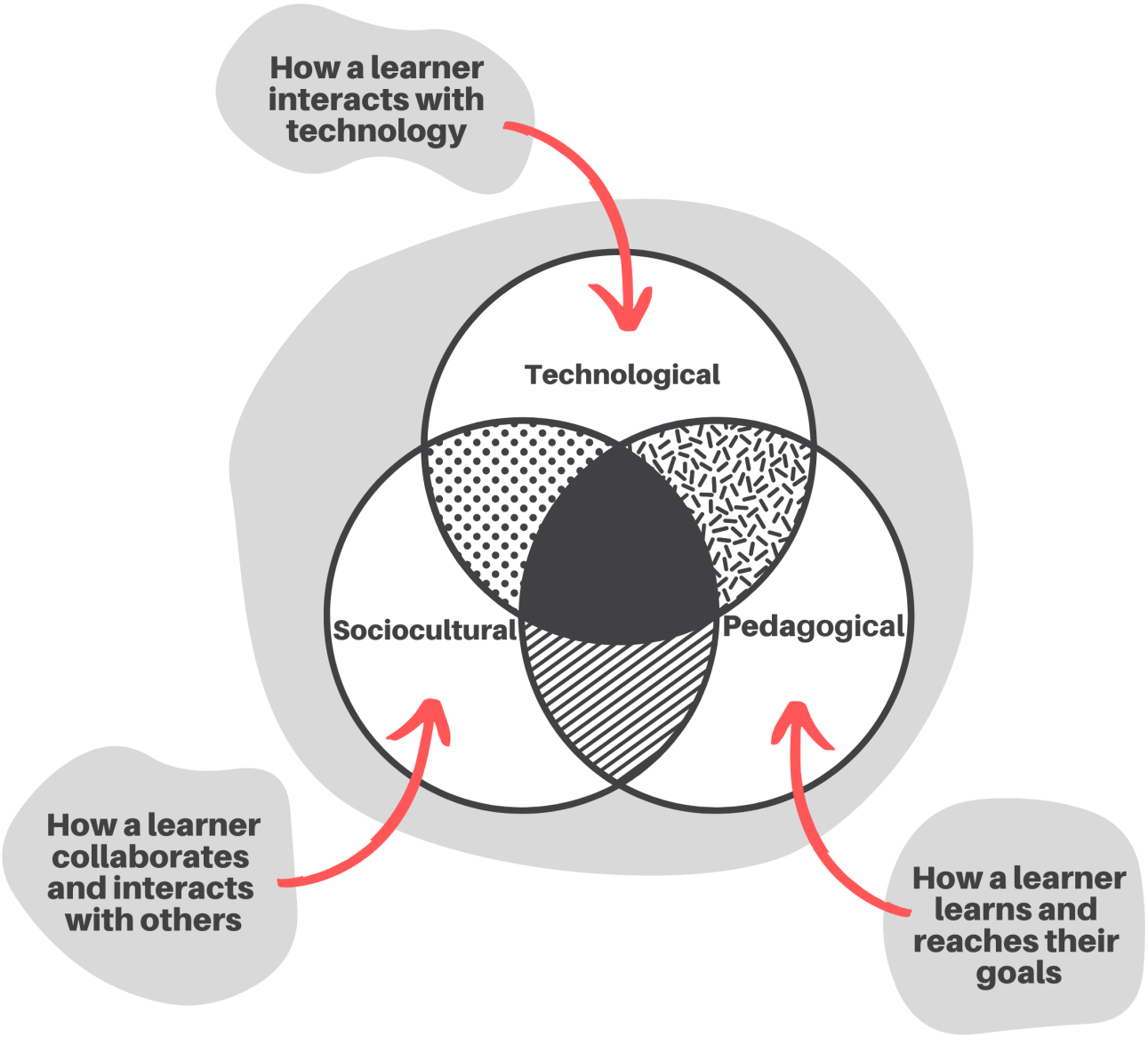
Although more comprehensive evaluation frameworks exist, most usability studies in health care learning remain narrowly focused on interface-level metrics. A growing body of research documents this gap and its consequences. A scoping review of 124 digital health tools for individuals with multiple sclerosis found that only 26% conducted usability testing, and among those, the majority used limited metrics [31]. A review of sensor-based health technologies found that while 83% - 91% assessed usability through satisfaction measures, fewer than 7% evaluated efficiency, and only 9% addressed users' ability to interpret clinical data [32]. A broader analysis of 610 digital health usability studies found that nearly half relied solely on the SUS, with little to no attention to accessibility, instructional utility, or cultural fit [33]. Minimal consideration of contextual or educational dimensions has similarly been noted in health IT evaluations [34] and in health care education specifically, where most mobile app usability studies relied on checklists and surveys, rarely assessing whether tools effectively supported learning or improved patient understanding [35]. The consistent absence of pedagogical and contextual considerations in usability evaluations may explain the high abandonment rates observed for health-related digital tools [36-39], and highlights the necessity of frameworks capable of capturing the complex and multifaceted nature of health care learning environments.

Addressing Usability Gaps in Health Care Learning Design Through Multidimensional Evaluation

The sociotechnical-pedagogical framework addresses this gap by conceptualizing usability as a multimodal, emergent property of 3 interdependent dimensions: technological (eg, interface design and navigation), pedagogical (eg, learning activities, instructional clarity, and task relevance), and sociocultural (eg, cultural values, emotional responsiveness, and social norms), rather than as a property of the interface alone. These dimensions interact to shape users' overall experience of a learning system (Figure 1). Importantly, the sociocultural dimension specifically recognizes that emotional responses, cultural backgrounds, and social contexts profoundly influence how individuals engage with and accept health learning technologies [40]. This integrated view is particularly relevant for PST mHealth interventions designed for populations with neurological or neurodevelopmental conditions, where sociocultural context and learning demands are inseparable from technological use.

XSL•FO
RenderX

Figure 1. Sociotechnical-pedagogical usability framework reprinted with permission from Schmidt et al [40].



To translate the sociotechnical-pedagogical framework into actionable usability assessment, Lu et al [41] developed the Comprehensive Assessment of Usability for Learning Technologies (CAUSLT), a validated instrument specifically designed to assess usability across all three sociotechnical-pedagogical dimensions. The CAUSLT was developed and empirically validated through multiple studies in learning contexts, demonstrating strong psychometric properties: exploratory factor analysis confirmed a coherent 3-factor structure aligned with the sociotechnical-pedagogical model (Kaiser-Meyer-Olkin=0.95; Bartlett test P<.001), with excellent overall internal consistency (Cronbach α=.95) and strong subscale reliability (technological α=.92; pedagogical α=.95; sociocultural α=.82) [41].

Study Rationale and Aims

Despite increasing adoption of psychoeducational mHealth interventions, current usability evaluation practices remain technocentric, focused on interface functionality and user satisfaction, while largely neglecting pedagogical coherence and sociocultural responsiveness [31,33-35]. This limitation is particularly problematic for PST interventions, where learning is not incidental but essential to sustained self-management. The purpose of this collective case study, therefore, is to examine how usability is experienced and manifested across technological, pedagogical, and sociocultural dimensions within 3 psychoeducational PST mHealth interventions. Anchored in the sociotechnical-pedagogical framework [21] and its sociocultural extension [40], and operationalized using the CAUSLT evaluation instrument [41], this study identifies core design principles and recurring usability challenges, with the goal of informing more holistic, patient-centered approaches to digital health design.

Research Questions

The following were the research questions (RQs):

1. How do participants perceive usability across technological, pedagogical, and sociocultural dimensions within mHealth PST interventions?
2. What recurring technological, pedagogical, and sociocultural usability challenges emerge across the three mHealth PST case studies?

XSL•FO
RenderX

3. What design principles can be derived from cross-case patterns in participants' usability experiences?

Methods

Research Design Overview

This study used an exploratory, multimethod, collective case study design [42] in which 3 independently developed psychoeducational mHealth PST interventions were evaluated through structured think-aloud usability sessions. Each of the 3 cases (Epilepsy Journey 2.0, Survivor's Journey, and electronic problem-solving training [ePST]) constituted an independent usability study with its own participant cohort, recruitment context, and data collection procedures. Findings from individual cases were subsequently examined together to identify recurring usability challenges and cross-case design principles. The multimethod approach integrated qualitative think-aloud usability data as the primary analytic source, and quantitative descriptive data from the CAUSLT instrument provided complementary, descriptive usability metrics for triangulation. Consistent with the exploratory and descriptive nature of this study, no inferential statistical tests were conducted across cases, and findings should be understood as preliminary.

Intervention Descriptions

Overview

Three mHealth PST interventions evaluated in this study were independently developed for distinct clinical populations and are described below. While each intervention was developed separately, they share several core foundational features related to the structure and delivery of PST [4]. Each intervention implements the core ABCDE (Assess, Brainstorm, Consider and Choose, Develop, and Evaluate) problem-solving sequence, with ePST adding the F (Flex; ABCDEF [Assess, Brainstorm, Consider and Choose, Develop and Do, Evaluate, and Flex]) step. In addition, all interventions incorporate guided coaching components in certain conditions. Table 1 illustrates design overlaps across interventions.

Table . Common features across interventions.

Feature	Epilepsy Journey 2.0	Survivor's Journey	ePST[a]
Clinical population	Adolescents with epilepsy	Adolescent and young adult survivors of brain tumors	Adults with a history of TBI[b]
PST[c] Framework	ABCDE[d]	ABCDE	ABCDEF[e] (additional Flex step)
Delivery Format	Mobile and web-based modules	Web-based modules	Mobile and web-based modules
Coaching Support	Therapist session	Therapist session	Coach session
Session Structure	10 themed virtual lands, each representing an instructional module	8 learning modules, with supplemental modules	5 learning modules, short lessons with problem-solving prompts
Learning Focus	Executive functioning support	Coping and daily functioning	Independent problem-solving and self-efficacy

[a]ePST: electronic problem-solving training.

[b]TBI: traumatic brain injury.

[c]PST: problem-solving training.

[d]ABCDE: Assess, Brainstorm, Consider and Choose, Develop, and Evaluate.

[e]ABCDEF: Assess, Brainstorm, Consider and Choose, Develop and Do, Evaluate, and Flex.

Case 1: Epilepsy Journey 2.0

Epilepsy Journey 2.0 is a web-based, mHealth PST intervention designed to support adolescents with epilepsy who experience executive functioning difficulties (Figure 2). The intervention includes multimedia instruction, short interactive tasks, and scaffolded guidance to support self-regulation, planning, and problem-solving skills. Instruction includes a 5-step problem-solving sequence using the "ABCDE" mnemonic. Content is organized using a "journey" metaphor, in which users progress through 10 themed virtual lands, each representing an instructional module related to executive functioning [43,44].

XSL•FO
RenderX

Figure 2. Epilepsy Journey 2.0 interface.

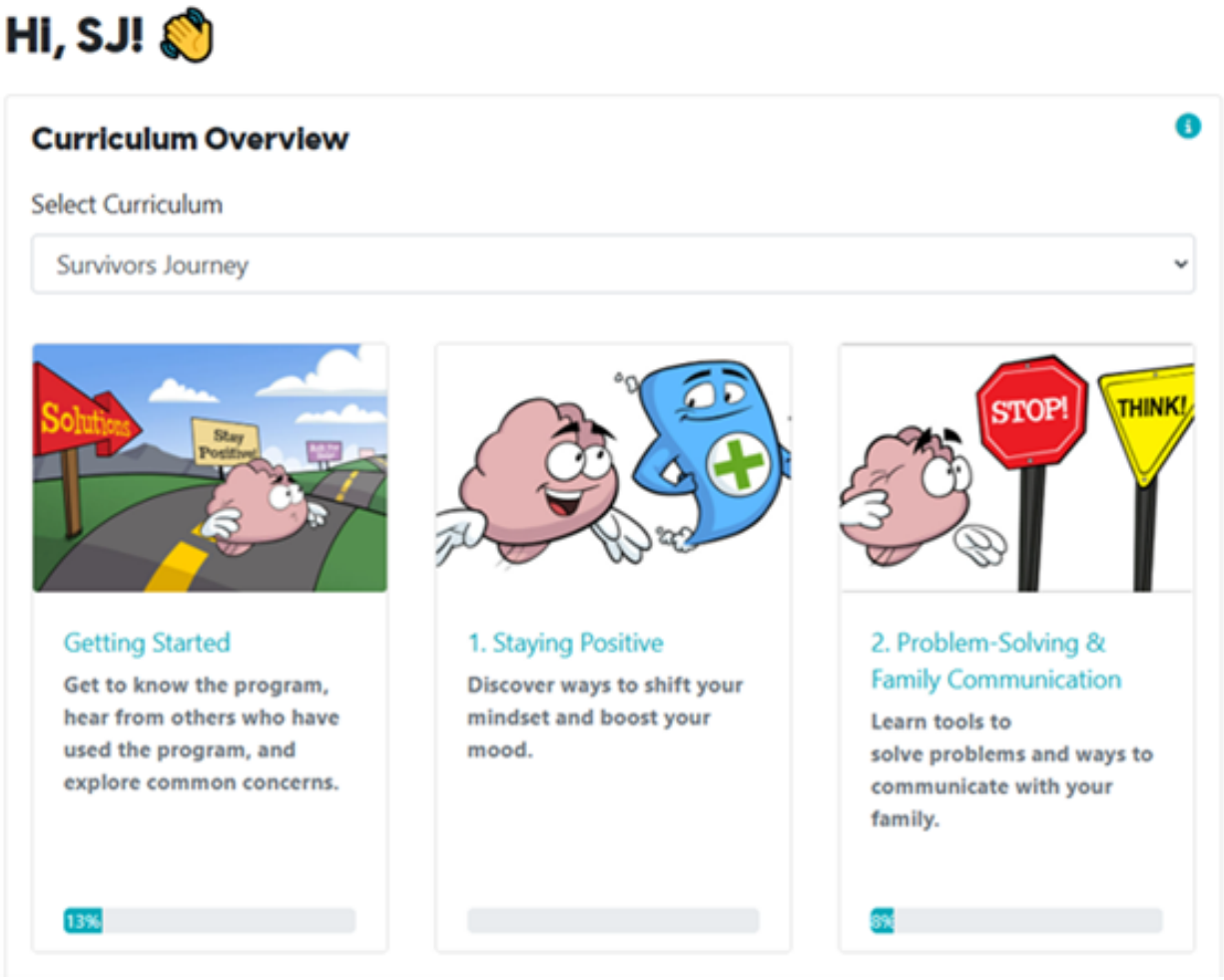


Case 2: Survivor's Journey

Survivor's Journey is a web-based psychosocial intervention designed to provide problem-solving therapy and coping skills training for adolescent and young adult survivors of brain tumors (Figure 3). It delivers PST-based coping and adjustment strategies through interactive multimedia lessons, therapist-supported sessions, and tailored problem-solving activities. The problem-solving sequence also adopts the "ABCDE" mnemonic. Modules focus on managing late effects, emotional regulation, fatigue, and everyday functional challenges among adolescent and young adult survivors of pediatric brain tumors. Instruction is structured through a stepwise problem-solving framework and emphasizes real-world applicability using developmentally appropriate scenarios. The platform has undergone feasibility and preliminary efficacy testing [45].

XSL•FO
RenderX

Figure 3. Survivor's Journey interface.

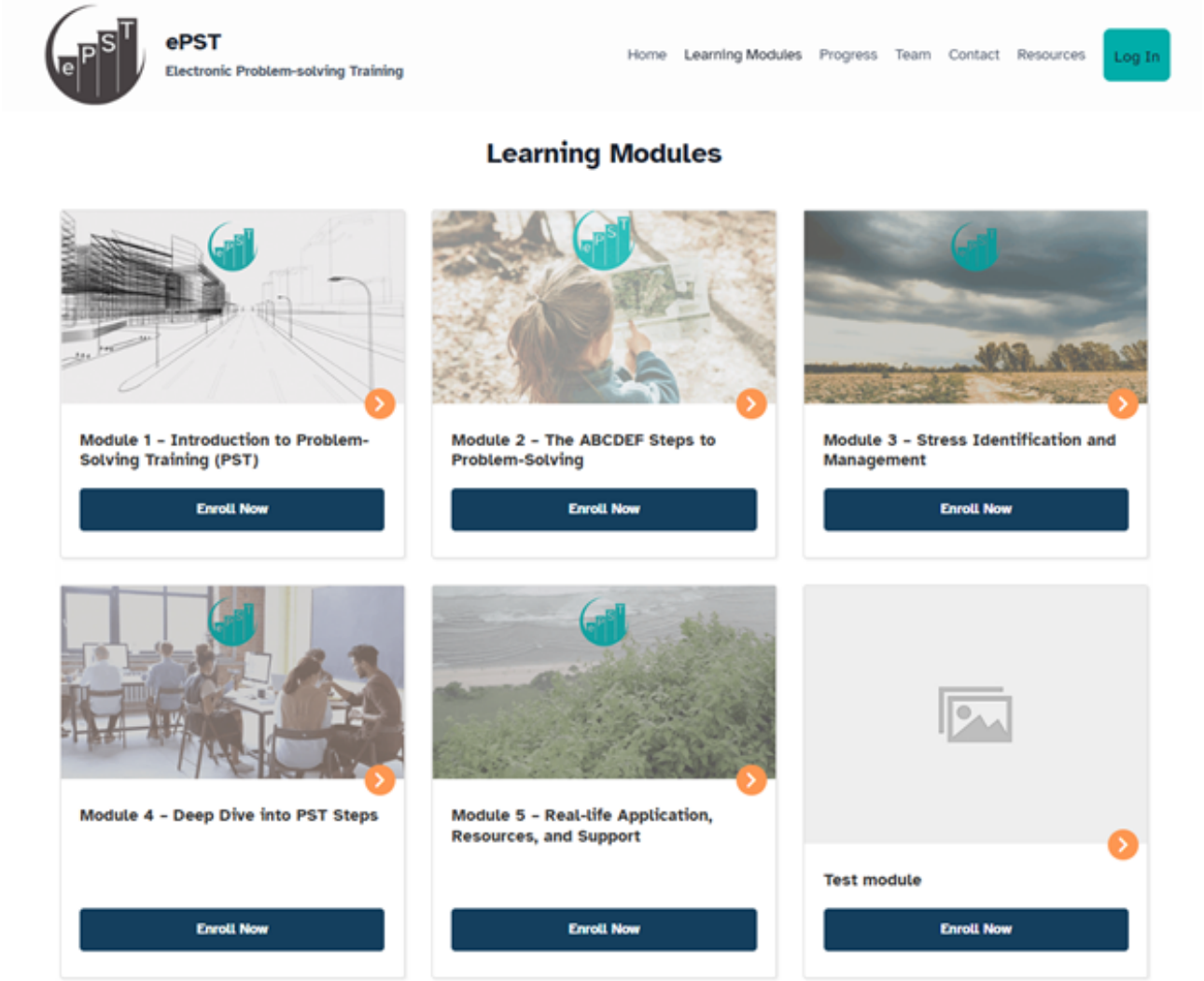


Case 3: ePST

ePST is a self-guided mHealth PST intervention for adults with a history of traumatic brain injury (TBI) with sustained loss of consciousness (characterized as "severe"; Figure 4). ePST builds on a traditional problem-solving framework by adapting it into an interactive, digital format to enhance accessibility and autonomy for users. The program teaches a 6-step problem-solving sequence using the "ABCDEF" mnemonic. Content is delivered through short lessons, interactive decision points, personal examples, and initial guidance from a coach to support users in applying PST strategies independently and at their own pace. Recent development work has described its person-centered adaptation process and community-based participatory design components [46,47].

XSL•FO
RenderX

Figure 4. ePST learning environment. ABCDEF: Assess, Brainstorm, Consider and Choose, Develop and Do, Evaluate, and Flex; ePST: electronic problem-solving training; PST: problem-solving training.



Participant Characteristics

A total of 14 participants were enrolled across the 3 usability studies. Table 2 presents participant characteristics by case, while Figure 5 presents the participant flowchart.

XSL•FO
RenderX

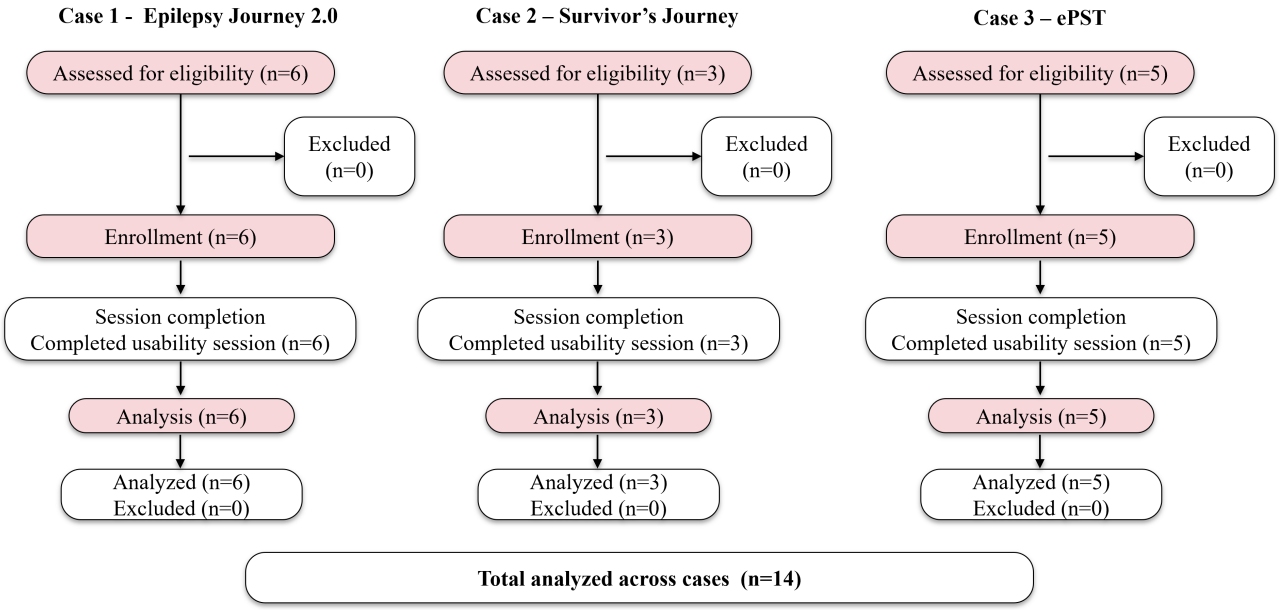
Table . Participants' characteristics and recruitment parameters.

Case	mHealth intervention	Participants, n	Recruitment parameters
1	Epilepsy Journey 2.0	6 (*M*=4, *F*=2)	Adolescents living with epilepsy who were able to navigate a digital platform and provide verbal feedback.
2	Survivor's Journey	3 (*M*=1, *F*=2)	Young adult survivors of brain tumors who were able to participate in remote usability testing and provide feedback on usability, navigation, and support for learning and coping goals.
3	ePST[a]	5 (*M*=2, *F*=3)	Adults who had sustained a severe TBI[b] 11 to 20 years prior and who were able to participate in a virtual session independently

[a]ePST: electronic problem-solving training.

[b]TBI: traumatic brain injury.

Figure 5. Participants flowchart. ePST: electronic problem-solving training.



Inclusion and Exclusion Criteria

Participants were recruited from the intended user populations for each respective intervention. Inclusion and exclusion criteria were defined at the case level. For case 1 (Epilepsy Journey 2.0), eligibility included adolescents with a confirmed clinical diagnosis of epilepsy who were able to access and navigate a web-based digital platform and provide verbal feedback during a usability session. For case 2 (Survivor's Journey), eligibility included young adults aged between 18 and 30 years with a documented history of pediatric brain tumor survivorship who were able to participate independently in a remote usability session and provide verbal feedback on platform navigation, learning content, and problem-solving supports. For case 3 (ePST), eligibility included adults with a documented history of severe TBI occurring 11 to 20 years before participation who were able to engage independently in a virtual session and had access to a personal smartphone or computer with internet connectivity.

Across the 3 cases, participants were excluded if they were unable to complete at least one assigned module during the usability session, could not provide verbal feedback, or lacked the technical access required to participate. No additional clinical exclusion criteria were applied beyond the diagnostic and functional requirements specified for each case in the preceding paragraph.

Sampling Procedures and Participant Recruitment

Purposive sampling was used across all 3 case studies, with participants selected for their direct experiential relevance to the intended user populations of each mHealth PST intervention. This approach was appropriate given the exploratory nature of the study and the goal of obtaining usability feedback from individuals with lived experience relevant to the interventions being evaluated [48].

XSL•FO
RenderX

Case 1 participants (n=6) were recruited through an existing advisory board established as part of the Epilepsy Journey 2.0 development process at Cincinnati Children's Hospital Medical Center (CCHMC). The advisory board comprised adolescents and young adults with epilepsy, their caregivers, health care providers, and educators. Usability participants were drawn from the adolescent member subset of this board. Case 2 participants (n=3) were recruited through clinical and research networks at CCHMC, including the pediatric neuro-oncology program. Eligible young adult survivors of brain tumor were identified by the clinical research team and invited to participate via direct outreach. Case 3 participants (n=5) were recruited through clinical and research settings affiliated with UT (University of Texas) Health Houston and TIRR (The Institute for Rehabilitation and Research) Memorial Hermann. Recruitment materials were shared within these medical center settings, and individuals who expressed interest were screened for eligibility based on TBI history and functional capacity to participate independently.

The sample sizes were determined by the independent protocols, recruitment timelines, and feasibility constraints of each respective usability study. In formative usability research, relatively small samples are commonly used for problem discovery, with foundational studies showing that the first 4 to 5 participants often identify a substantial proportion of major usability problems [49-51]. At the same time, later work has cautioned that 5 users should not be treated as a universal rule, because larger samples improve stability and increase the proportion of problems detected, particularly when systems or user groups are more heterogeneous [52]. Consistent with this broader literature, recommendations for usability studies in health-related digital systems commonly place formative samples in the range of approximately 5 to 10 participants, with larger samples warranted when interface complexity or user diversity is greater. On this basis, the sample sizes for case 1 (n=6) and case 3 (n=5) were considered appropriate for identifying prominent usability issues within their respective intervention contexts, whereas the smaller sample in case 2 (n=3) is interpreted more cautiously and its findings are presented as preliminary.

Measures and Instruments

Quantitative data were collected from two sources: (1) a brief preassessment questionnaire evaluating participants' general technology comfort and familiarity with digital learning platforms, and (2) the CAUSLT scale. For the technology comfort assessment, participants responded to a five-item Likert-type technology comfort survey. It comprised five items which assessed: (1) frequency of eHealth app use; (2) general comfort with digital technology; (3) comfort using digital health or educational tools specifically; (4) confidence navigating new mHealth platforms; and (5) prior experience with interventions similar to the one being tested. Each item was rated on a 5-point scale (1=not at all comfortable/experienced to 5=extremely comfortable/experienced). Responses were summarized using descriptive statistics across all participants. For CAUSLT, the instrument comprises 20 items rated on a 5-point Likert scale (1=strongly disagree to 5=strongly agree), organized into 3 subscales aligned with the sociotechnical-pedagogical framework [41]. Item scores were averaged within each subscale and converted to a 0‐100 standardized scale to facilitate interpretation and cross-subscale comparison.

Ethical Considerations

The 3 studies reported in this paper were reviewed and approved by the appropriate institutional review boards (IRBs). The Epilepsy Journey 2.0 and Survivor's Journey studies were approved by the CCHMC IRB (Federalwide Assurance #00002988). Epilepsy Journey 2.0 recruited an advisory board of adolescents/adults, caregivers, health care providers, and teachers. For the purpose of this study, only adolescents who completed usability testing were included. Due to the advisory board nature, only an agreement form was signed, and participation implied consent, as they were considered consultants, not participants. Survivor's Journey followed the same institutional review requirements and received CCHMC IRB approval under protocol 2024‐0546. The ePST usability study was reviewed by the University of Georgia Human Research Protection Program (IRB #00009943) and was deemed exempt under federal guidelines.

All participants provided written informed consent before data collection. For remote sessions, consent was obtained electronically using Qualtrics; for in-person sessions, consent was documented via approved parental permission, assent, or adult consent forms, as required. Consent documents described study procedures, audio or video recording, risks and benefits, voluntary participation, HIPAA (Health Insurance Portability and Accountability Act) authorization where applicable, and data handling. All data were deidentified before analysis and stored on encrypted institutional servers with access restricted to authorized research personnel. Raw recordings and master linking files will be destroyed at the conclusion of each study. Participants in the ePST study received US $25 per session. Advisory board members in Epilepsy Journey 2.0 were compensated US $50 per advisory board meeting they participated in (up to 3 meetings). Participants in the Survivor's Journey usability study received US $50. No identifying information is presented in this paper or supplementary materials.

Data Collection Procedures

Overview

Usability testing was conducted across the 3 mHealth PST interventions between June 2024 and July 2025. Across all studies, participants completed a structured think-aloud session in which they interacted with selected modules of their assigned intervention, followed by a postsession CAUSLT survey and a presession technology comfort survey. Sessions were conducted remotely via Zoom, except for 1 in-person session for Epilepsy Journey 2.0. A standardized "think-aloud" protocol was used, including facilitator-led introductions, verbal consent, verbal reflection prompts, and observational note-taking. All sessions were audio- and video-recorded and transcribed for analysis.

Case Study 1: Epilepsy Journey 2.0

Conducted in June and August 2024 (n=6; 5 remotely and 1 in person), this study evaluated ease of use, interface design,

XSL•FO
RenderX

content delivery, and interactive features. Participants explored the platform, completed 1 assigned module focused on memory or problem-solving, and provided feedback through verbal protocols.

Case Study 2: Survivor's Journey

Conducted remotely in June and July 2025 (n=3), this study assessed navigation, functionality, and problem-solving support features. Participants explored the platform by completing account sign-up and landing-page tasks, followed by working through a learning module and using the problem-solving tools.

Case Study 3: ePST

Conducted remotely in July 2024 (n=5), this study evaluated the introductory and advanced modules focused on the problem-solving strategies. Each participant completed 2 assigned lessons addressing the full "ABCDEF" problem-solving strategy. Participants accessed the platform via personal smartphones (1 participant used a personal computer) and were prompted to reflect on both ease of use and learning support.

Data Analysis

Quantitative Data Analysis

Quantitative data were analyzed using descriptive statistics appropriate to the exploratory and small-sample nature of each case study. No inferential statistical tests were conducted, and no α level was set. CAUSLT item scores (rated 1 - 5) were converted to a standardized 0 - 100 scale by averaging item responses within each subscale and rescaling proportionally ((mean−1)/4×100). Case-level subscale means and overall usability means were calculated by averaging participant-level standardized scores within each intervention. Technology comfort survey data were analyzed using descriptive statistics (frequencies and percentages) and reported at the case level to contextualize participants' baseline familiarity with digital tools. Regarding missing data, complete CAUSLT and technology comfort survey data were obtained from all 14 enrolled participants. No item-level missing data were identified in any case.

Qualitative Data Analysis

The qualitative component of the study followed established usability testing procedures rather than thematic coding approaches. Session recordings, think-aloud verbalizations, and observer field notes served as the primary qualitative data sources. A semistructured, task-based usability protocol guided each session, during which participants completed a series of predefined tasks while thinking aloud and sharing their screens. Tasks targeted both technological usability (eg, navigation and multimedia interaction) and pedagogical usability (eg, clarity of content and instructional structure).

Transcripts generated from Zoom's automated transcription tool were reviewed for accuracy and used alongside field notes to identify instances where participants encountered navigation difficulties, misunderstandings, or breakdowns in task flow. Two trained members of the research team independently reviewed and discussed each usability session to document observed usability barriers and to verify that identified issues were grounded in direct participant behavior rather than interpretive coding. The team collaboratively reviewed usability issues and assigned severity ratings using the Nielsen 5-point severity framework [53], which categorizes usability problems as: 0=not a usability problem; 1=cosmetic problem only (need not be fixed unless time allows); 2=minor usability problem (low priority to fix); 3=major usability problem (important to fix, high priority); and 4=usability catastrophe (imperative to fix before product release). Disparities in ratings were discussed until consensus was reached. In the case that consensus could not be reached within the research team, the faculty advisor was consulted. This only occurred once, with an issue related to the loading time of 1 video. Representative usability issues and corresponding severity ratings are provided in Multimedia Appendix 1. All team members who participated in usability data analysis received training in applying the Nielsen severity scale as part of the research group's standard usability research training procedures, including review of criteria definitions, worked examples, and calibration exercises before the study. Efficiency-related data, such as time on task and error rates, were extracted from session recordings and incorporated into the analysis. Findings were summarized in structured usability reports for each intervention.

Results

Overview

This section presents findings from the exploratory usability evaluation across the 3 case studies. No cross-case statistical comparisons were made. Quantitative findings are reported descriptively and interpreted within the context of each case. First, participants' technology comfort levels are summarized to contextualize subsequent usability findings. Next, CAUSLT results are reported for overall usability and for the technological, pedagogical, and sociocultural dimensions aligned with the sociotechnical-pedagogical framework. Finally, we highlight recurring usability challenges and corresponding design principles that emerged in relation to these dimensions.

RQ1: How Do Participants Perceive Usability Across Technological, Pedagogical, and Sociocultural Dimensions Within mHealth PST Interventions?

Participants' Technology Comfort Levels

Participants reported varied levels of technology comfort across the 3 interventions. In Epilepsy Journey 2.0, half of the participants (3/6, 50%) identified as experienced technology users; in ePST, 3 of 5 (60%) participants did so; and in Survivor's Journey, two-thirds (2/3, 66.7%) of participants also identified as experienced users. Most participants reported feeling comfortable using technology both in general and for health or educational purposes, including 83.3% (5/6) of participants in Epilepsy Journey 2.0, 100% (3/3) of participants in Survivor's Journey, and 60% (3/5) of participants in ePST. Prior experience with eHealth or digital health tools varied, with the majority in each intervention indicating occasional use. Confidence using technology independently ranged from slightly confident to very confident across all interventions. Detailed

XSL•FO
RenderX

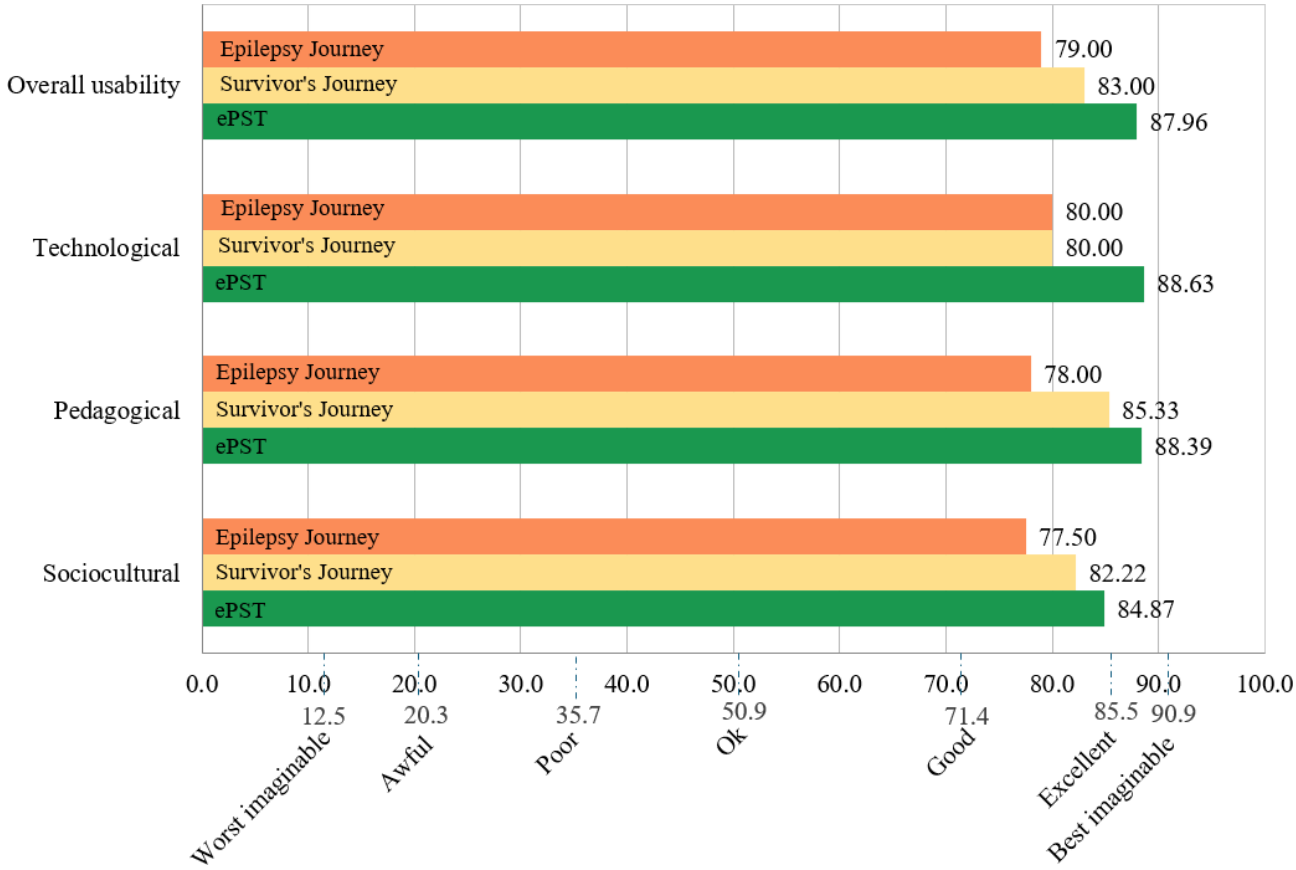
case level technology comfort responses are provided in Multimedia Appendix 2.

CAUSLT Results

Overall Usability

Figure 6 displays average CAUSLT scores across the 3 mHealth PST interventions. Scores are reported on a standardized 0 to 100 scale. All 3 interventions received overall mean scores in the upper range of the scale, indicating generally high participant agreement that the interventions exhibited usability characteristics across the measured dimensions. ePST received the highest overall mean score (87.96, SD 13.73), reflecting consistently high participant agreement across technological, pedagogical, and sociocultural dimensions. Survivor's Journey and Epilepsy Journey 2.0 received an overall mean score of 83.00 (SD 7.70) and 79.00 (SD 12.75), respectively, reflecting similarly strong overall agreement.

Figure 6. Average Comprehensive Assessment of Usability for Learning Technologies scores across the 3 mHealth (mobile health) interventions based on sociotechnical-pedagogical dimensions. ePST: electronic problem-solving training.



Technological Usability

Technological usability was assessed using the 7 technological CAUSLT items (Figure 7). ePST received the highest technological mean scores, reflecting strong participant agreement across items related to ease of navigation and functional clarity. Additionally, Epilepsy Journey 2.0 and Survivor's Journey both received technological mean scores in the upper portion of the scale, with somewhat lower agreement on items related to navigation clarity and user control. These items reflected areas in which participants indicated difficulty understanding interface structure or completing tasks without assistance, suggesting room for technological improvement in both cases.

Pedagogical Usability

Pedagogical usability results are shown in Figure 8. ePST received high mean scores across items reflecting instructional support, content clarity, and learning goal alignment. Epilepsy Journey 2.0 received high mean scores, particularly for items related to engagement and satisfaction, with slightly lower agreement on items concerning instructional clarity and guidance. Survivor's Journey showed consistently strong pedagogical usability, with high mean scores for learning goals, engagement, and pacing, and slightly lower ratings on items related to instructional support and content quality.

XSL•FO
RenderX

Figure 7. Technology usability of 3 interventions. ePST: electronic problem-solving training.

XSL•FO
RenderX

Figure 8. Pedagogical usability of three interventions. ePST: electronic problem-solving training.

XSL•FO
RenderX

Sociocultural Usability

Sociocultural usability was assessed using 3 CAUSLT items focused on inclusivity, lack of prejudice, and opportunities for social interaction (Figure 9). ePST received high mean sociocultural scores, with slightly lower agreement on whether the intervention could be used by a wide range of learners. Survivor's Journey also demonstrated predominantly excellent sociocultural usability, particularly for perceptions of safety and opportunities for social interaction, with somewhat lower mean scores on items related to accommodating diverse users. Epilepsy Journey 2.0 received somewhat lower mean sociocultural scores relative to the other 2 cases, with greater variability across items and more neutral responses regarding social interaction and inclusivity. These scores indicate that participants experienced fewer supports for resolving errors or receiving corrective guidance in this intervention.

XSL•FO
RenderX

Figure 9. Sociocultural usability of 3 interventions. ePST: electronic problem-solving training.

XSL•FO
RenderX

RQ2: What Recurring Technological, Pedagogical, and Sociocultural Usability Challenges Emerge Across the 3 mHealth PST Case Studies?

Overview

The think-aloud sessions and field notes across all 3 case studies yielded a set of preliminary usability challenges that appeared with some consistency across interventions, participants, and sociotechnical-pedagogical dimensions. Representative usability challenges and their associated Nielsen severity ratings are summarized in Multimedia Appendix 1. These challenges are organized below by dimension.

Technological Usability Challenges

Overview

Participants across all 3 interventions identified recurring technological difficulties, including unclear navigation structures, technical performance issues, limited mobile compatibility and readability, confusing progress indicators, and problematic visual design and layout.

Inconsistent and Confusing Navigation Structures

Participants frequently reported uncertainty about how to proceed within modules. In Epilepsy Journey 2.0, a participant stated, “I found it unclear how to start and navigate through the modules and was confused about what button to select at the end of a module.” In ePST, a participant (Pe3) described unexpected behavior when using arrow-based navigation. Similarly, all participants in Survivor’s Journey reported navigation confusion due to misleading labels and screen indicators that did not accurately match where they were in the modules. For example, 1 participant (Ps2) noted that “the system sometimes displays ‘Session Content’ even when I am still on the Start page, and I’m not sure whether clicking Save or Next will complete a task or simply store my progress.”

Poor Mobile Responsiveness and Technical Glitches

Participants noted difficulties when accessing the interventions on smartphones, including slow loading and multimedia glitches. In ePST, a participant (Pe5) remarked, “The video didn’t load properly, and it caused confusion for me.” In Survivor’s Journey, participants encountered system glitches that disrupted smooth task completion. These included the “site tour” repeatedly appearing even after completion and certain features such as the AI-supported Refine button being positioned in ways that made them effectively hidden or difficult to locate.

Lack of Clear Progress Tracking

Participants in Epilepsy Journey 2.0 and Survivor’s Journey reported uncertainty about task completion status, stating that it was not always clear whether activities were finished or how far along they were within a module.

Pedagogical Usability Challenges

Overview

Participants reported several pedagogical issues related to how the interventions supported learning, cognitive engagement, and instructional flow. Key issues included instructional ambiguity, sequencing gaps, information overload, and limited feedback.

Instructional Ambiguity and Sequencing Gaps

Participants across interventions described uncertainty about intended next steps. In Epilepsy Journey 2.0, a participant (Pj5) stated, “I wasn’t sure whether I was supposed to read or click.” ePST participants reported difficulty recalling and following multistep instructions after completing videos. Survivor’s Journey users described similar uncertainty regarding task order and inconsistent message design.

Dense Content Affecting Working Memory

Participants in Epilepsy Journey 2.0 and Survivor’s Journey noted that some pages contained dense text or tightly condensed content that was challenging to process. A participant in Survivor’s Journey (Ps1) explicitly struggled with poor message design regarding text hierarchy and reported confusion caused by redundant and repetitive messages, while Epilepsy Journey 2.0 participants commented on sections being “too wordy.”

Limited Feedback and Progress Cues

Across interventions, several participants expressed uncertainty about whether their actions had been recorded or whether they had completed tasks. Epilepsy Journey 2.0 users noted difficulty determining module completion status, and ePST participants occasionally interpreted badges as indicators of progress. Survivor’s Journey participants also reported uncertainty about overall progress, as they could not distinguish between the progress bar and actual lesson completion.

Sociocultural Usability Challenges

Overview

Participants identified sociocultural concerns related to emotional sensitivity, inclusivity, representational authenticity, and error recovery.

Content Sensitivity and Emotional Impact

In Epilepsy Journey 2.0, a participant (Pj4) expressed concern about a flashing background and questioned, “if it was intended to cause seizures.” In ePST, a participant reported that the orange color scheme was intimidating, and a motivational quote felt misaligned with their expectations as a user with a brain injury. In Survivor’s Journey, 1 participant (Ps1) expressed that the website’s name did not clearly communicate that the intervention was designed for survivors of pediatric brain tumors, which lessened her confidence in its relevance and purpose.

Accessibility and Inclusivity Beyond Technical Features

Participants noted that mobile display challenges affected readability and ease of use for individuals relying exclusively on smartphones. These concerns extended beyond technological accessibility, reflecting broader needs for inclusive visual, cognitive, and layout design.

Relevance and Relatability of Content and Narratives

In ePST, a participant (Pe5) commented that “someone in the lesson does not appear to be brain injured,” raising questions

XSL•FO
RenderX

about the perceived visual authenticity of character portrayals. In contrast, participants in Epilepsy Journey 2.0 noted that characters and testimonials felt relatable.

Lack of Support for Error Recovery and Comprehension Feedback

Participants across cases expressed interest in clearer explanations when answers were incorrect or when navigation errors occurred. For example, a participant in Epilepsy Journey 2.0 (Pj4) stated that understanding "why an answer was wrong would be helpful for learning and retention."

RQ3: What Design Principles Can Be Derived From Cross-Case Patterns in Participants' Usability Experiences?

Cross-case analysis of participants' usability experiences resulted in a set of preliminary design considerations that appeared with directional consistency across the 3 mHealth PST interventions. These considerations were organized into technological, pedagogical, and sociocultural categories to align with the multidimensional structure of the CAUSLT and the broader sociotechnical-pedagogical framework. Table 3 presents the resulting design principles, each linked to observed usability challenges and participant-reported needs.

Table . Usability-driven design principles that emerged across 3 mHealth[a] PST[b] interventions.

Dimension and design principle	Description
Technological	
Intuitive, consistent navigation	Use clear labeling, standardize button behavior, and ensure directional cues are predictable to reduce confusion and improve wayfinding.
Mobile-first design	Optimize layout, font size, multimedia compatibility, and interaction design specifically for smartphone use, particularly for users with no access to a desktop.
Transparent progress tracking	Provide clear, visual indicators of task and module completion to help users track their learning and reduce frustration.
In-platform troubleshooting support	Include tooltips, guided walkthroughs, or pop-ups to help users recover from errors and reduce reliance on external help.
Pedagogical	
Clear instructional guidance	Use explicit instructions, consistent "next step" cues, and scaffolded transitions to guide learners through the intervention without ambiguity.
Reduce cognitive overload	Break up long text blocks, use visual hierarchy, and incorporate multimedia to improve comprehension and decrease fatigue.
Context-sensitive feedback and completion cues	Provide immediate, relevant feedback and clear visual indicators of progress or task completion to reinforce engagement and comprehension.
Use narrative and interactive learning strategies	Incorporate relatable characters, scenarios, or gamified interactions to enhance learner connection and motivation.
Sociocultural	
Safety-conscious and empathetic design	Avoid triggering visuals (eg, flashing), intimidating color schemes, or emotionally insensitive language; test content with end users.
Universal accessibility and flexibility	Design with inclusive features such as adjustable font size, high contrast, and alternative inputs to accommodate diverse users and contexts.
Authentic, relatable content	Include stories, testimonials, and characters that reflect the lived experiences, language, and identities of the target user population.
Provide feedback for error recovery and reflection	Offer explanatory feedback on incorrect answers and missed actions to foster user confidence, learning, and self-efficacy.

[a]mHealth: mobile health.

[b]PST: problem-solving training.

Discussion

Principal Findings

This exploratory study aimed to examine how usability manifests across technological, pedagogical, and sociocultural dimensions within 3 mHealth PST interventions and to identify recurring issues within and across cases that influence patient learning and engagement. Using the sociotechnical-pedagogical framework as an analytic lens, the findings reveal how these dimensions interact to shape user experience and highlight observations regarding challenges within each case. Within and across cases, quantitative CAUSLT findings indicated generally high perceived usability, alongside meaningful variability across dimensions and interventions that highlighted specific areas for improvement. Complementing these findings, qualitative think-aloud data and observational field notes revealed a recurring set of usability challenges spanning navigation

XSL•FO
RenderX

ambiguity, limitations in mobile responsiveness, instructional uncertainty, cognitive overload, insufficient feedback, and condition-specific sociocultural concerns. Despite differences in clinical populations and intervention design, these challenges exhibited notable directional consistency across cases. Synthesis of observations within each case further yielded 12 preliminary design principles organized along sociotechnical-pedagogical dimensions and grounded in observed user behaviors. In the following sections, we discuss key insights associated with each RQ. We conclude by reviewing the study's limitations and outlining implications for future mHealth learning design and evaluation.

RQ1: Participants' Perceptions of Usability Across Technological, Pedagogical, and Sociocultural Dimensions

Across all 3 interventions, participants' perceptions suggested that technological, pedagogical, and sociocultural usability operate as mutually reinforcing dimensions rather than isolated components. Interventions with strong interface clarity, such as ePST and Epilepsy Journey 2.0, appeared to provide a smoother entry point into the learning experience, but participants' comments showed that interface success alone did not determine overall usability. Even when navigation was described as intuitive, uncertainties about task flow, expectations, and progress frequently shaped how users interpreted the system's coherence. This pattern suggests that technological usability is necessary, but not sufficient, in mHealth PST contexts, where learning (not just interaction) drives engagement.

Conversely, the Survivor's Journey case study offered a particularly illustrative, though preliminary, window into the consequences of pedagogical and sociocultural misalignment. Despite receiving relatively strong technological scores, participants in that intervention encountered what Nielsen [53] would classify as "usability catastrophes" in the pedagogical domain, where the pedagogical structure failed to communicate progress accurately, leaving participants unsure whether they had completed learning objectives or inadvertently skipped content. These challenges illustrate that users evaluate usability not solely by whether the system works, but also by whether it supports comprehension, reduces cognitive load, and acknowledges their context of use, in alignment with research suggesting that patients interpret usability through the lens of learning demands, emotional states, and everyday constraints (not just interface performance) [21,54,55].

Taken together, these preliminary observations support the view that usability in learning-oriented mHealth interventions is most productively understood as a relational property, one that materializes at the intersection of how the technology performs, how the instruction guides, and how the content speaks to participants' experiential realities [21,40]. This integrated perspective is consistent with broader arguments in health informatics that narrowly technocentric usability evaluation fails to capture the dimensions of the user experience most consequential for clinical engagement and behavior change [31-34].

RQ2: Recurring Usability Challenges Across Technological, Pedagogical, and Sociocultural Dimensions

Within each intervention case, usability challenges reflected potential misalignments across technological, pedagogical, and sociocultural dimensions. Rather than isolated interface flaws, participants' experiences within each case suggested deeper tensions relative to how users attempted to navigate, learn from, and make sense of the interventions. Within individual cases, when 1 dimension falters, it appeared to trigger or amplify challenges in the others, suggesting an entangled nature of usability in mHealth PST contexts [40]. For example, in Survivor's Journey, participants' difficulties distinguishing between progress storage and task completion (a technological navigation issue) undermined their ability to determine whether learning objectives had been met, resulting in pedagogical uncertainty and disrupted task flow. Similarly, in Epilepsy Journey 2.0, the absence of clear module completion indicators and ambiguous navigation cues left participants uncertain not only about their location within the intervention but also about whether they had fully engaged with core PST components, reflecting cascading pedagogical consequences of technological signaling gaps.

Technological challenges, such as inconsistent navigation pathways, limited mobile responsiveness, and unstable multimedia performance, frequently disrupted participants' ability to orient themselves within the interventions. Even when individual screens appeared well designed, unclear directional cues or difficulty returning to previous sections undermined participants' sense of control and progress. These issues point to a need to consider technological usability not as surface-level interface performance but as a foundational condition that can enable or constrain participants' learning. Indeed, when navigation or responsiveness became unpredictable, participants struggled to maintain focus on problem-solving tasks, indicating that technological clarity is a prerequisite for cognitive engagement. In this case sample, the participants with brain tumors (n=3) indicated a preference for tablets or desktops over smartphones. While this preliminary observation in a limited sample suggests that mobile-only design may introduce accessibility burdens for certain users with specific neurological conditions, validation in larger samples is needed before drawing broader conclusions about device preferences.

Pedagogical challenges emerged when instructional sequences, task expectations, or content density failed to align with participants' cognitive needs. Ambiguity about what to do next, difficulty following multistep processes, and the burden of overly dense or text-heavy material created avoidable cognitive load. These challenges were not simply presentation issues. Instead, they reflect gaps in how systems guide users through the problem-solving process. Even when the interface itself was functional, pedagogical misalignment disrupted participants' ability to interpret, retain, and apply the PST framework. Further, sociocultural challenges shaped participants' experiences by influencing how understood and represented they felt while using a given intervention. Elements that might seem minor from a design perspective, such as color schemes,

XSL•FO
RenderX

character portrayals, or emotionally neutral phrasing, sometimes carried unintended emotional weight for participants with neurological conditions. For example, when examples felt inauthentic or insensitive, participants questioned the intervention's relevance and legitimacy. Participants' judgments about usability were shaped as much by cultural resonance, emotional tone, and instructional clarity as by the mechanics of the interface itself.

Collectively, these recurring challenges demonstrate that usability breakdowns are rarely confined to a single point of failure. Difficulties with navigation, comprehension, or emotional resonance often co-occurred and reinforced one another, illustrating the interdependence of sociotechnical-pedagogical dimensions and underscoring the utility of the sociotechnical-pedagogical framework for identifying these interactions. The systemic perspective offered by the sociotechnical-pedagogical framework underscores why mHealth PST interventions require integrated, context-sensitive design approaches able to anticipate the cognitive and emotional realities of diverse patient populations.

RQ3: Design Principles Derived From Cross-Case Usability Patterns

Building on the recurring usability challenges identified across the 3 interventions, the third RQ focused on how these patterns translate into actionable design principles for mHealth PST environments. Findings from this exploratory study revealed that effective usability is not the outcome of isolated interface improvements, but the result of design decisions that simultaneously support technological clarity, pedagogical coherence, and sociocultural resonance. The preliminary design principles derived illustrate how these dimensions interact to shape patient experience and highlight that usability emerges from the alignment of these layers rather than from any single dimension in isolation.

Across cases, technological design principles emphasize the need for intuitive navigation pathways, mobile-first responsiveness, transparent progress indicators, and embedded troubleshooting support. These principles suggest that ease of use in mHealth PST interventions involves more than functional stability. It requires interfaces that anticipate learners' cognitive flow, device constraints, and need to recover quickly from errors. Participant experiences showed that technical clarity enables deeper engagement with problem-solving content, whereas even small inconsistencies in navigation or responsiveness disrupt orientation and confidence. Thus, technological usability functions as the infrastructural layer that supports, rather than competes with, the learning process.

Pedagogical design principles emphasized instructional clarity, reduced cognitive load, context-sensitive feedback, and the use of narrative or interactive elements to support comprehension. Interventions in which instructions were explicit, sequences were well-paced, and content was visually organized allowed participants to focus on the "ABCDE" or "ABCDEF" steps of PST rather than the mechanics of navigating the system. Conversely, ambiguous instructions or dense content increased cognitive effort and weakened participants' ability to internalize core problem-solving strategies. Such patterns demonstrate the centrality of pedagogical usability to learning-oriented mHealth design, in which the quality of instructional scaffolding can directly shape users' performance and perceptions of system usability.

Finally, sociocultural design principles extended usability beyond technological and pedagogical domains and into the areas of sociocultural authenticity and inclusive representation. Participant perceptions demonstrated that issues such as color choice, narrative tone, and representational accuracy can meaningfully shape trust, comfort, and identification with the content. When these elements aligned with participants' lived experiences, responses suggested that participants perceived the intervention as supportive and credible, whereas when they did not, participants perceived the system as less usable, even if the technology and underlying pedagogy were otherwise sound. This suggests that sociocultural usability mediates how technological and pedagogical decisions are interpreted and underscores the need for co-design practices that meaningfully involve target populations through design and development phases [56].

As a whole, the design principles derived from RQ3 reveal that usability is best understood as a relational construct that materializes when technological, pedagogical, and sociocultural dimensions are coherently integrated. The sociotechnical-pedagogical framework proved particularly effective for identifying where misalignments occurred and how they could be addressed through design. This integrated view offers a path forward for developing digital health PST interventions that are not only functional but also instructionally sound and culturally responsive.

Limitations

Several important limitations should be considered when interpreting these findings. First, and most critical, the sample sizes across all 3 cases were small, and the sample for case 2 (Survivor's Journey; n=3) falls below the commonly cited 5-participant threshold for formative usability testing [50]. The small sample hence limits transferability and precludes inferential cross-case conclusions. Second, demographic data, including race, ethnicity, gender, age, and educational background, were not systematically collected as part of these usability evaluations. This limits the ability to examine how these dimensions interact with condition-specific sociocultural usability experiences. While the sociotechnical-pedagogical framework's sociocultural dimension in this study was operationalized through participants' condition-specific lived experience (as described in the Methods section), it is recognized that demographic factors may intersect with and amplify condition-specific experiences in ways that this study could not detect. Future studies should systematically collect and report demographic data to enable a more complete examination of sociocultural usability. Third, usability testing was predominantly conducted in controlled, virtual environments, potentially affecting participants' natural interactions with the platforms. Real-world usage scenarios might reveal additional usability issues not captured in this structured testing context. Fourth, participant familiarity with technology varied significantly across interventions, which may have influenced

XSL•FO
RenderX

usability perceptions independently of the interventions' inherent design. However, it is notable that the intervention with the highest overall usability also included participants with the lowest average confidence using technology; it may be that those with more technology familiarity also have higher expectations and are more critical than those with less familiarity (rather than those with less familiarity simply struggling more). Finally, this study's primary reliance on observational data without longitudinal follow-up restricts the ability to assess sustained usability perceptions and real-world intervention effectiveness.

Comparison With Prior Work

The preliminary findings of this exploratory study align with and extend prior research that argues for multidimensional approaches to usability in digital health. Consistent with earlier work, technological elements such as navigation clarity, visual readability, and mobile responsiveness remain foundational determinants of user satisfaction and engagement [40,55,56]. However, this study demonstrates that technological usability alone cannot fully account for how individuals evaluate or experience digital health PST interventions. Instead, technological features become meaningful only when aligned with the pedagogical and sociocultural demands of learning and self-management. This perspective aligns with emerging critiques in the mHealth literature that evaluations focusing only on technological factors tend to overlook the bigger picture of how users interpret and make sense of an intervention as a learning environment [57,58].

Pedagogically, this study supports previous arguments by Jahnke et al [21] and others advocating for the incorporation of instructional clarity, cognitive load management, and meaningful feedback into usability design. However, the present findings extend this literature by showing empirically how instructional misalignment often underlies perceived usability failures, emphasizing that interface efficiency must be evaluated alongside how effectively a system teaches, guides, and motivates its patients [21,41]. Socioculturally, the results resonate with scholars [40,59] who emphasize emotional safety, cultural sensitivity, and inclusivity as vital to mHealth engagement. Yet, this study advances prior work by demonstrating that sociocultural usability functions as a mediating layer, shaping how technological and pedagogical design decisions are perceived, trusted, and adopted by diverse users. Empathetic design and authentic representation thus emerge as central to achieving sustained participation and emotional connection [47,56].

More broadly, these findings contribute to a growing body of literature calling for multidimensional usability evaluation in digital health. Reviews of current practice have consistently identified a reliance on single-dimensional instruments, most commonly the SUS, across the field, with little attention to pedagogical or sociocultural dimensions [31-35]. The present study provides a concrete, case-level demonstration of what such evaluations miss: in 2 of the 3 cases, CAUSLT subscale scores revealed variation across sociotechnical-pedagogical dimensions that would not have been visible with a single overall usability score, and think-aloud data revealed condition-specific sociocultural concerns absent from the technological usability narrative. The CAUSLT, as operationalized in this study, proved useful for surfacing these multidimensional patterns, though, as noted in the Methods section, no validated benchmark classifications for the CAUSLT exist yet, and the instrument's psychometric performance in clinical populations beyond those represented here remains to be established [41]. Future research should adopt longitudinal designs, include more diverse populations, and be conducted in real-world clinical settings to examine how technological, pedagogical, and sociocultural factors dynamically interact to shape usability in psychoeducational mHealth interventions. Such work is needed to clarify how multidimensional usability influences engagement, learning processes, and behavioral health outcomes over time.

Implications and Future Directions

The preliminary findings of this exploratory study present several implications for mHealth PST design, evaluation practice, and future research, implications that extend beyond the specific interventions examined here. At the most fundamental level, these findings provide initial evidence that the prevailing practice of evaluating psychoeducational mHealth tools primarily on the basis of interface-level usability metrics is likely to produce an incomplete and potentially misleading picture of whether an intervention is ready for clinical deployment. An mHealth PST tool that scores well on the SUS or similar instruments, but whose instructional scaffolding generates cognitive overload, whose feedback mechanisms fail to support error recovery, or whose characters and scenarios feel inauthentic to the target population, may be perceived as technically usable while failing to deliver meaningful psychoeducational benefit. The CAUSLT, operationalized within the sociotechnical-pedagogical framework, offers a more complete evaluation approach precisely because it surfaces these multidimensional gaps. Broader adoption of multidimensional evaluation tools in the formative development of digital health interventions could improve the alignment between usability, learning, and clinical effectiveness, contributing to stronger engagement, adherence, and ultimately better health outcomes for patients using PST mHealth tools.

For intervention developers and health care teams, the 12 preliminary design principles derived from this study offer a starting framework for building multidimensional usability into the design process from the outset, rather than addressing usability deficits reactively after development. Mobile-first design, explicit instructional scaffolding, context-sensitive feedback, and co-design with condition-specific communities are not complex or expensive design choices; they are design commitments that reflect a genuine prioritization of the user's learning experience. The growing literature on community-based participatory approaches in digital health design [56] suggests that these commitments are most effectively implemented when the target population is involved as a genuine partner in the design process rather than as a late-stage test user.

For researchers, these findings suggest several high-priority directions for future work. First, the 12 design principles identified here require prospective empirical testing to determine

XSL•FO
RenderX

whether implementing them improves usability, learning, engagement, and clinical outcomes in PST mHealth contexts. Second, the CAUSLT requires further psychometric validation within clinical populations with neurological and neurodevelopmental conditions, including the development of population-specific descriptive norms that would enable more meaningful score interpretation than is currently possible. Third, longitudinal usability research is needed to understand how usability experiences at first interaction predict sustained engagement, treatment completion, and behavioral health outcomes over time, a critical but currently understudied link in the mHealth effectiveness chain. Fourth, future studies should use larger, more diverse samples within each condition-specific population and should systematically collect demographic data to enable examination of intersecting sociocultural dimensions of usability experience.

In sum, this exploratory study provides preliminary evidence that usability in psychoeducational mHealth PST interventions is a multidimensional construct shaped by the coherent integration of technological, pedagogical, and sociocultural design decisions. When these dimensions are aligned, the intervention creates conditions for meaningful engagement with PST content; when they diverge, the result is not merely user inconvenience but a potential disruption of the very learning process that the intervention is designed to support. Advancing the field requires evaluation practices and design approaches that take this multidimensionality seriously, and this study offers an initial empirical foundation from which that advancement can proceed.

Acknowledgments

We used ChatGPT (OpenAI) to assist with editing for grammar, creating figure and table captions, and enhancing readability of lengthy and complex sentence formulations. All AI outputs were reviewed, edited, and approved by the named authors.

Funding

The ePST study was supported by the Office of the Assistant Secretary of Defense for Health Affairs through the Congressionally Directed Medical Research Programs TBI and Psychological Health program (HT9425-23-1-0567). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of this paper. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the Departments of Veterans Affairs and Defense or the United States Government. The Epilepsy Journey 2.0 study was supported by funding from the National Institute of Neurological Disorders and Stroke (UG3/UH3 NS129588). The trial sponsor did not have a role in the conceptualization, design, data collection, management, analysis, or dissemination of this study. The Survivor's Journey study was supported by funding from the US Army Medical Research and Development Command (HT9425-24-1-0746). The trial sponsor did not have a role in the conceptualization, design, data collection, management, analysis, or dissemination of this study.

Data Availability

All data are available in this paper.

Authors' Contributions

Conceptualization: MS (lead), DO (supporting), SJ (supporting), AM (supporting), SW (supporting)
Data curation: MS (lead), SJ (supporting), AM (supporting), SW (supporting), DO (supporting)
Formal analysis: DO (lead), MS (supporting)
Funding acquisition: SJ (lead for ePST), AM (lead for Epilepsy Journey 2.0), SW (lead for Survivor's Journey)
Investigation: SJ (lead for ePST), AM (lead for Epilepsy Journey 2.0), SW (lead for Survivor's Journey), MS (supporting)
Methodology: DO (lead for the multicase synthesis), MS (supporting), SJ (supporting), AM (supporting), SW (supporting)
Project administration: DO (lead for manuscript integration), SJ (lead for ePST), AM (lead for Epilepsy Journey 2.0), SW (lead for Survivor's Journey), MS (supporting)
Resources: SJ (lead for ePST), AM (lead for Epilepsy Journey 2.0), SW (lead for Survivor's Journey), MS (supporting)
Supervision: SJ (lead for ePST), AM (lead for Epilepsy Journey 2.0), SW (lead for Survivor's Journey), MS (supporting)
Visualization: DO (lead), MS (supporting)
Writing – original draft: DO (lead)
Writing – review & editing: DO (lead), MS (supporting), SJ (supporting), AM (supporting), SW (supporting)

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Usability issues identified across all 3 cases with Nielsen severity ratings.
[DOCX File, 21 KB - humanfactors_v13i1e90379_app1.docx]

Multimedia Appendix 2
Participant comfort level with teComfort Level with Technology.
[DOC File, 36 KB - humanfactors_v13i1e90379_app2.doc]

References

1. Murray E, Hekler EB, Andersson G, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851. [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
2. eHealth. World Health Organization - Regional Office for the Eastern Mediterranean. 2025. URL: http://www.emro.who.int/health-topics/ehealth/ [accessed 2026-08-18]
3. Lee JA, Choi M, Lee SA, Jiang N. Effective behavioral intervention strategies using mobile health applications for chronic disease management: a systematic review. BMC Med Inform Decis Mak 2018 Feb 20;18(1):12. [doi: 10.1186/s12911-018-0591-0] [Medline: 29458358]
4. Nezu AM, D'Zurilla TJ. Problem-Solving Therapy: A Positive Approach to Clinical Intervention, 3rd edition: Springer Publishing Company; 2006. URL: https://books.google.com/books?id=8L8&dq=Problem-Solving+Therapy:+A+Positive+Approach+to+Clinical+Intervention,+3rd+Edition&source=gbs_navlinks_s [accessed 2026-08-18]
5. Rabelo JL, Cruz BF, Ferreira JDR, Viana BDM, Barbosa IG. Psychoeducation in bipolar disorder: a systematic review. World J Psychiatry 2021 Dec 19;11(12):1407-1424. [doi: 10.5498/wjp.v11.i12.1407] [Medline: 35070785]
6. Palermo TM, Law EF, Bromberg M, Fales J, Eccleston C, Wilson AC. Problem-solving skills training for parents of children with chronic pain: a pilot randomized controlled trial. PAIN 2016 Jun;157(6):1213-1223. [doi: 10.1097/j.pain.0000000000000508] [Medline: 26845525]
7. Economides M, Ranta K, Nazander A, et al. Long-term outcomes of a therapist-supported, smartphone-based intervention for elevated symptoms of depression and anxiety: quasiexperimental, pre-postintervention study. JMIR mHealth uHealth 2019 Aug 26;7(8):e14284. [doi: 10.2196/14284] [Medline: 31452521]
8. Ghanbari E, Yektatalab S, Mehrabi M. Effects of psychoeducational interventions using mobile apps and mobile-based online group discussions on anxiety and self-esteem in women with breast cancer: randomized controlled trial. JMIR mHealth uHealth 2021 May 18;9(5):e19262. [doi: 10.2196/19262] [Medline: 34003138]
9. Hentati A, Forsell E, Ljótsson B, Kaldo V, Lindefors N, Kraepelien M. The effect of user interface on treatment engagement in a self-guided digital problem-solving intervention: a randomized controlled trial. Internet Interventions 2021 Dec;26:100448. [doi: 10.1016/j.invent.2021.100448] [Medline: 34471610]
10. Jakob R, Harperink S, Rudolf AM, et al. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022 May 25;24(5):e35371. [doi: 10.2196/35371] [Medline: 35612886]
11. Milne-Ives M, Homer S, Andrade J, Meinert E. The conceptualisation and measurement of engagement in digital health. Internet Interventions 2024 Jun;36:100735. [doi: 10.1016/j.invent.2024.100735] [Medline: 38558760]
12. Hassan HM, Galal-Edeen GH. From usability to user experience. Presented at: 2017 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS); Nov 24-26, 2017. [doi: 10.1109/ICIIBMS.2017.8279761]
13. ISO 9241. Ergonomics of human-system interaction — part 210: human-centred design for interactive systems. 2019. URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-2:v1:en [accessed 2026-08-18]
14. Cho H, Yen PY, Dowding D, Merrill JA, Schnall R. A multi-level usability evaluation of mobile health applications: a case study. J Biomed Inf 2018 Oct;86:79-89. [doi: 10.1016/j.jbi.2018.08.012] [Medline: 30145317]
15. Goldwater JC. Human factors and usability in mobile health design – factors for sustained patient engagement in diabetes care. Proc Int Symp Hum Factors Ergon Health Care 2014 Jun;3(1):63-70. [doi: 10.1177/2327857914031009]
16. Müssener U, Thomas K, Linderoth C, et al. Development of an intervention targeting multiple health behaviors among high school students: participatory design study using heuristic evaluation and usability testing. JMIR mHealth uHealth 2020 Oct 29;8(10):e17999. [doi: 10.2196/17999] [Medline: 33118942]
17. Niazkhani Z, Toni E, Cheshmekaboodi M, Georgiou A, Pirnejad H. Barriers to patient, provider, and caregiver adoption and use of electronic personal health records in chronic care: a systematic review. BMC Med Inf Decis Mak 2020 Jul 8;20(1):153. [doi: 10.1186/s12911-020-01159-1] [Medline: 32641128]
18. Boucher EM, Raiker JS. Engagement and retention in digital mental health interventions: a narrative review. BMC Digit Health 2024;2(1):52. [doi: 10.1186/s44247-024-00105-9]

XSL•FO
RenderX

19. Holthe T, Halvorsrud L, Karterud D, Hoel KA, Lund A. Usability and acceptability of technology for community-dwelling older adults with mild cognitive impairment and dementia: a systematic literature review. Clin Interv Aging 2018;13:863-886. [doi: 10.2147/CIA.S154717] [Medline: 29765211]
20. Li Z, Lu F, Wu J, et al. Usability and effectiveness of eHealth and mHealth interventions that support self-management and health care transition in adolescents and young adults with chronic disease: systematic review. J Med Internet Res 2024 Nov 26;26:e56556. [doi: 10.2196/56556] [Medline: 39589770]
21. Jahnke I, Schmidt M, Pham M, Singh K. Sociotechnical-pedagogical usability for designing and evaluating learner experience in technology- enhanced environments. In: Learner and User Experience Research: An Introduction for the Field of Learning Design &Amp; Technology 2020:127-145 URL: https://edtechbooks.s3.us-west-2.amazonaws.com/pdfs/36/3246.pdf [accessed 2026-08-18]
22. Lu J, Schmidt M, Lee M, Huang R. Usability research in educational technology: a state-of-the-art systematic review. Educ Tech Res Dev 2022 Dec;70(6):1951-1992. [doi: 10.1007/s11423-022-10152-6]
23. John B. SUS: a “quick and dirty” usability scale. In: Usability Evaluation In Industry: CRC Press; 1996. URL: https://www.taylorfrancis.com/chapters/edit/10.1201/9781498710411-35/sus-quick-dirty-usability-scale-john-brooke [accessed 2026-08-18]
24. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum–Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
25. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Hum-Comput Interact 1995 Jan;7(1):57-78. [doi: 10.1080/10447319509526110]
26. Kirakowski J. The software usability measurement inventory: background and usage. In: Usability Evaluation In Industry: CRC Press; 1996. URL: https://www.taylorfrancis.com/chapters/edit/10.1201/9781498710411-33/software-usability-measurement-inventory-background-usage-kirakowski [accessed 2026-08-18]
27. Zaharias P, Poylymenakou A. Developing a usability evaluation method for e-learning applications: beyond functional usability. Int J Hum-Comput Interact 2009 Jan 14;25(1):75-98. [doi: 10.1080/10447310802546716]
28. Georgsson M, Staggers N. Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. J Am Med Inf Assoc 2016 Jan;23(1):5-11. [doi: 10.1093/jamia/ocv099] [Medline: 26377990]
29. Weber F, Kloek C, Stuhrmann S, Blum Y, Grüneberg C, Veenhof C. Usability and preliminary effectiveness of an app-based physical activity and education program for people with hip or knee osteoarthritis - a pilot randomized controlled trial. Arthritis Res Ther 2024 Apr 10;26(1):83. [doi: 10.1186/s13075-024-03291-z] [Medline: 38600607]
30. Chaniaud N, Métayer N, Megalakaki O, Loup-Escande E. Effect of prior health knowledge on the usability of two home medical devices: usability study. JMIR mHealth uHealth 2020 Sep 21;8(9):e17983. [doi: 10.2196/17983] [Medline: 32955454]
31. Tea F, Groh AMR, Lacey C, Fakolade A. A scoping review assessing the usability of digital health technologies targeting people with multiple sclerosis. NPJ Digit Med 2024 Jun 25;7(1):168. [doi: 10.1038/s41746-024-01162-0] [Medline: 38918483]
32. Tandon A, Cobb B, Centra J, et al. Human factors, human-centered design, and usability of sensor-based digital health technologies: scoping review. J Med Internet Res 2024 Nov 15;26:e57628. [doi: 10.2196/57628] [Medline: 39546781]
33. Maqbool B, Herold S. Potential effectiveness and efficiency issues in usability evaluation within digital health: a systematic literature review. J Syst Software 2024 Feb;208:111881. [doi: 10.1016/j.jss.2023.111881]
34. Yen PY, Bakken S. Review of health information technology usability study methodologies. J Am Med Inf Assoc 2012;19(3):413-422. [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
35. Johnson SG, Potrebny T, Larun L, Ciliska D, Olsen NR. Usability methods and attributes reported in usability studies of mobile apps for health care education: scoping review. JMIR Med Educ 2022 Jun 29;8(2):e38259. [doi: 10.2196/38259] [Medline: 35767323]
36. Kidman PG, Curtis RG, Watson A, Maher CA. When and why adults abandon lifestyle behavior and mental health mobile apps: scoping review. J Med Internet Res 2024 Dec 18;26:e56897. [doi: 10.2196/56897] [Medline: 39693620]
37. Mustafa AS, Ali N, Dhillon JS, Alkawsi G, Baashar Y. User engagement and abandonment of mHealth: a cross-sectional survey. Healthcare (Basel) 2022 Jan 24;10(2):221. [doi: 10.3390/healthcare10020221] [Medline: 35206837]
38. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR mHealth uHealth 2015 Nov 4;3(4):e101. [doi: 10.2196/mhealth.4924] [Medline: 26537656]
39. Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. JMIR mHealth uHealth 2019 Aug 29;7(8):e12983. [doi: 10.2196/12983] [Medline: 31469081]
40. Schmidt M, Earnshaw Y, Jahnke I, Tawfik AA. Entangled eclecticism: a sociotechnical-pedagogical systems theory approach to learning experience design. Educ Technol Res Dev 2024 Jun;72(3):1483-1505. [doi: 10.1007/s11423-024-10353-1]
41. Lu J, Schmidt M, Shin J. Beyond technological usability: exploratory factor analysis of the comprehensive assessment of usability scale for learning technologies (CAUSLT). arXiv. Preprint posted online on Feb 3, 2025. [doi: 10.48550/arXiv.2501.18754]

XSL•FO
RenderX

42. Creswell JW, Poth CN. Qualitative Inquiry and Research Design: Choosing Among Five Approaches, 4th edition: SAGE Publications; 2018. URL: https://collegepublishing.sagepub.com/products/qualitative-inquiry-and-research-design-4-246896 [accessed 2026-08-18]
43. Modi AC, Schmidt M, Smith AW, Turnier L, Glaser N, Wade SL. Development of a web-based executive functioning intervention for adolescents with epilepsy: the Epilepsy Journey. Epilepsy Behav 2017 Jul;72:114-121. [doi: 10.1016/j.yebeh.2017.04.009] [Medline: 28575759]
44. Modi AC, Mara CA, Schmidt M, et al. Epilepsy Journey: a proof of concept trial of a web-based executive functioning intervention for adolescents with epilepsy. Epilepsia 2019 Sep;60(9):1895-1907. [doi: 10.1111/epi.16317] [Medline: 31423591]
45. Wade SL, Narad ME, Moscato EL, et al. A Survivor's Journey: preliminary efficacy of an online problem-solving therapy for survivors of pediatric brain tumor. Pediatr Blood Cancer 2020 Feb;67(2):e28043. [doi: 10.1002/pbc.28043] [Medline: 31724307]
46. Juengst S, Schmidt M. Smart technology, smarter design: person-centered adaptation of problem-solving training to ePST. Presented at: Rehabilitation Psychology Mid-Year Conference; Feb 20-23, 2025 URL: https://rehab.washington.edu/news-and-events/join-us-at-the-rehabilitation-psychology-mid-year-conference/ [accessed 2026-08-29]
47. Weng Y, Ogundairo D, Martin K, Schmidt M. A community-based participatory research to inform the design of electronic problem-solving training intervention for individuals with traumatic brain injury: a case study. Presented at: Mary Frances Early College of Education Research Conference; Apr 29, 2025 URL: https://fyi.coe.uga.edu/2025-04-register-for-the-2025-mary-frances-early-college-of-education-research-conference-on-april-29/ [accessed 2026-08-29]
48. Patton MQ. Qualitative Research & Evaluation Methods: Integrating Theory and Practice: SAGE Publications; 2014. URL: https://study.sagepub.com/patton4e [accessed 2026-08-18]
49. Macefield R. How to specify the participant group size for usability studies: a practitioner's guide. J Usability Stud 2009;5(1):34-45 [FREE Full text]
50. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. Presented at: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems CHI '93; 1993; Amsterdam, The Netherlands p. 206-213. [doi: 10.1145/169059.169166]
51. Virzi RA. Refining the test phase of usability evaluation: how many subjects is enough? Hum Factors 1992 Aug;34(4):457-468. [doi: 10.1177/001872089203400407]
52. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003 Aug;35(3):379-383. [doi: 10.3758/bf03195514] [Medline: 14587545]
53. Nielsen J. Severity ratings for usability problems. Nielsen Norman Group. 1994 Nov 1. URL: https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/ [accessed 2026-08-29]
54. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mHealth apps: meta-ethnographic review of qualitative studies. JMIR mHealth uHealth 2019 Jul 10;7(7):e13817. [doi: 10.2196/13817] [Medline: 31293246]
55. Simola S, Hörhammer I, Xu Y, et al. Patients' experiences of a national patient portal and its usability: cross-sectional survey study. J Med Internet Res 2023;25:e45974. [doi: 10.2196/45974]
56. Schmidt M, Weng Y, Juengst S, Holland A. Designing electronic problem-solving training for individuals with traumatic brain injury: mixed methods, community-based, participatory research case study. J Med Internet Res 2026;28:e83995. [doi: 10.2196/83995]
57. Galavi Z, Montazeri M, Khajouei R. Which criteria are important in usability evaluation of mHealth applications: an umbrella review. BMC Med Inf Decis Mak 2024 Nov 29;24(1):365. [doi: 10.1186/s12911-024-02738-2] [Medline: 39614283]
58. Jacob C, Sezgin E, Sanchez-Vazquez A, Ivory C. Sociotechnical factors affecting patients' adoption of mobile health tools: systematic literature review and narrative synthesis. JMIR mHealth uHealth 2022 May 5;10(5):e36284. [doi: 10.2196/36284] [Medline: 35318189]
59. Schmidt M, Lu J, Luo W, et al. Learning experience design of an mHealth self-management intervention for adolescents with type 1 diabetes. Educ Technol Res Dev 2022;70(6):2171-2209. [doi: 10.1007/s11423-022-10160-6] [Medline: 36278247]

Abbreviations

ABCDE : Assess, Brainstorm, Consider and Choose, Develop, and Evaluate
ABCDEF: Assess, Brainstorm, Consider and Choose, Develop and Do, Evaluate, and Flex
CAUSLT: Comprehensive Assessment of Usability for Learning Technologies
CCHMC: Cincinnati Children's Hospital Medical Center
ePST: electronic problem-solving training
HIPAA: Health Insurance Portability and Accountability Act
IRB: institutional review board
mHealth: mobile health
PST: problem-solving training

XSL•FO
RenderX

RQ: research question
SUS: System Usability Scale
TBI: traumatic brain injury
TIRR: The Institute for Rehabilitation and Research
UT: University of Texas

Edited by A Kushniruk; submitted 27.Dec.2025; peer-reviewed by A Rasool, E Etafo, F Elkourdi, J Megerian; revised version received 25.Jun.2026; accepted 22.Jul.2026; published 09.Sep.2026.

Please cite as:
Ogundairo D, Juengst S, Modi A, Wade S, Schmidt M
Usability Across 3 mHealth Problem-Solving Training Interventions for Diverse Neurodevelopmental and Neurological Populations: Multicase Usability Evaluation
JMIR Hum Factors 2026;13:e90379
URL: https://humanfactors.jmir.org/2026/1/e90379
doi:10.2196/90379

© David Ogundairo, Shannon Juengst, Avani Modi, Shari Wade, Matthew Schmidt. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 9.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Determinants of Adoption of a Mechanical Ventilation Dashboard in Intensive Care: Measurement Instrument for Determinants of Innovations–Based, Single-Center, Cross-Sectional Evaluation

Sarah Frederika Christina Mugge[1], MD; Tobias Daniel van Leijsen[1], RN, MSc; Merijn Constantijn Reuland[1], MD; Annelies Visser[2], PhD; Frederique Paulus[1,3], RN, PhD; Dave Anton Dongelmans[1,4], MD, PhD

[1]Department of Intensive Care, Amsterdam University Medical Centers, Meibergdreef 9, Amsterdam, North Holland, The Netherlands

[2]Department of Surgery, Amsterdam University Medical Centers, Amsterdam, North Holland, The Netherlands

[3]Faculty of Health, Center of Expertise Urban Vitality, Amsterdam University of Applied Sciences, Amsterdam, North Holland, The Netherlands

[4]Amsterdam Public Health Research Institute, Amsterdam University Medical Centers, Amsterdam, North Holland, The Netherlands

Corresponding Author:
Sarah Frederika Christina Mugge, MD
Department of Intensive Care, Amsterdam University Medical Centers, Meibergdreef 9, Amsterdam, North Holland, The Netherlands

Abstract

Background: Lung-protective ventilation (LPV) reduces complications of mechanical ventilation, yet adherence in intensive care units (ICUs) remains inconsistent. Digital dashboards may support LPV by improving situational awareness and supporting protocol adherence. However, adoption of such tools in high-acuity clinical environments depends on a range of cognitive, professional, and contextual determinants. The Measurement Instrument for Determinants of Innovations (MIDI) provides a validated framework to systematically assess these factors.

Objective: This study aims to identify determinants influencing the adoption of a newly piloted mechanical ventilation dashboard in the ICU using the MIDI framework.

Methods: We conducted a single-center, cross-sectional evaluation among ICU health care professionals during a dedicated survey period within a pilot introduction of a mechanical ventilation dashboard at Amsterdam University Medical Center. Participants completed a structured questionnaire consisting of 24 MIDI items adapted to the ICU context rated on a 5-point Likert scale (“completely disagree” to “completely agree”), supplemented by open-ended questions on perceived barriers and facilitators to its use. Determinants were classified as facilitators when ≥80% of respondents selected “agree” or “completely agree” and as barriers when ≥20% selected “disagree” or “completely disagree.” Open-ended responses were analyzed using a general inductive thematic approach.

Results: A total of 71 completed questionnaires were analyzed, including responses from nurses, physicians, intensivists, ventilation specialists, and researchers in mechanical ventilation. Six determinants met the criteria for facilitators: outcome expectations, self-efficacy, procedural clarity, low complexity, correctness, and observability. Two determinants met the criteria for barriers: relevance for client and professional obligation. Analysis of open-ended responses highlighted perceived barriers such as additional workload, the need for an extra device, overlap with existing systems, and limited role-specific relevance. Facilitators included improved situational overview, educational value, easier trend monitoring, and increased efficiency.

Conclusions: This evaluation identified key determinants influencing the adoption of a mechanical ventilation dashboard in the ICU. While the dashboard was generally perceived as useful and easy to understand, adoption was shaped by determinants related to workflow integration, role-specific relevance, and professional responsibility. These findings suggest that successful introduction of digital clinical support tools in intensive care requires attention not only to technical design, but also to how such tools align with users’ roles, daily work processes, and shared clinical responsibilities. Systematic assessment of determinants provides actionable insights into the adoption of digital decision-support tools in high-acuity care settings.

(JMIR Hum Factors 2026;13:e92313) doi:10.2196/92313

KEYWORDS

adoption, barriers and facilitators; dashboard; implementation; intensive care unit; mechanical ventilation; Measurement Instrument for Determinants of Innovations; MIDI questionnaire; lung-protective ventilation; ventilation dashboard

XSL•FO
RenderX

Introduction

Mechanical ventilation is a critical and often lifesaving intervention for patients admitted to intensive care units (ICUs). However, mechanical ventilation can also contribute to lung injury and complications, including acute respiratory distress syndrome and ventilator-induced lung injury, which increase morbidity and mortality [1,2]. To reduce these risks, clinicians use lung-protective ventilation (LPV) strategies, including low tidal volume (4‐8 milliliters/kilogram ideal body weight [V_T/IBW]), limited plateau pressures (<30 cm H_2O), and limited driving pressures (<15 cm H_2O) [2-6].

Despite strong evidence, adherence to LPV in daily ICU practice remains inconsistent [7,8]. Both patient factors, such as body size and disease severity, and organizational factors such as protocols, staffing and unit routines, contribute to variation in LPV delivery [8-10]. In addition, high cognitive workload, competing clinical demands, and limited protocol visibility can hinder consistent application of evidence-based ventilation practices [11,12].

Clinical dashboards may support LPV by integrating relevant data into a single view, improving situational awareness, and reducing cognitive effort during decision-making [13,14]. In the context of LPV, a ventilator management dashboard with tidal volume alerts was associated with improved hospital-wide adherence and facilitated timely correction of suboptimal ventilator settings [15,16]. Beyond adherence, dashboards may also increase efficiency in prerounding data gathering and, when displayed on shared screens, enhance communication accuracy, information exchange, and clinical satisfaction during multidisciplinary discussions [17]. More broadly, prior studies have shown that clinical dashboards can support care delivery by integrating relevant data, reducing cognitive burden, and providing real-time feedback, which may contribute to improved guideline adherence and fewer errors in complex clinical settings [13,14,18].

Adoption of change in the ICU is shaped by perceived role relevance and feasibility within existing workflows, and influenced by a complex and multifaceted set of facilitators and barriers at the personal, organizational, and health system levels, including engagement, available resources, and local organizational context [19-23]. To systematically assess these factors, the Measurement Instrument for Determinants of Innovations (MIDI) has been widely applied in health care settings [24,25]. The MIDI extends beyond technical considerations by capturing determinants such as procedural clarity, compatibility with daily practice, and relevance for professionals and patients. Previous research has demonstrated its utility in identifying barriers and facilitators that influence the adoption of new technologies and interventions in clinical practice [26-30].

The aim of this study was to use the MIDI questionnaire to identify barriers and facilitators influencing the adoption of a newly piloted mechanical ventilation dashboard designed to enhance LPV in the ICU.

Methods

Design, Setting, and Participants

We conducted a single-center, cross-sectional questionnaire study among ICU health care professionals in a Dutch tertiary academic referral center, where a mechanical ventilation dashboard was introduced as part of a pilot implementation. The study had a descriptive and exploratory aim: to identify determinants influencing adoption of the dashboard in the ICU context. The study took place at Amsterdam University Medical Center (Amsterdam UMC).

The study population consisted of ICU health care professionals involved in mechanical ventilation, including intensivists, ICU nurses, physicians, ventilation practitioners (ICU nurses with additional training as ventilation practitioners), and researchers in the field of mechanical ventilation.

Ethical Considerations

The Medical Ethics Review Committee of Amsterdam UMC reviewed the study protocol and confirmed that the Dutch Medical Research Involving Human Subjects Act (WMO) did not apply to this study, and gave approval for this study (reference: 2025.0194). Participation was voluntary and informed consent was implied by completion of the questionnaire. Questionnaire responses were pseudonymized, stored securely, and handled confidentially. No financial or other incentives were provided for participation.

Innovation: VentSight Ventilation Dashboard

The innovation under study was the VentSight ventilation dashboard, a web-based application developed by Hamilton Medical. The dashboard was installed in the ICU of Amsterdam UMC during the first months of 2025 and made available to authorized staff and researchers as part of routine pilot use. The present study evaluated user perceptions collected during a dedicated MIDI survey period conducted within this pilot.

The dashboard was designed to support evaluation of and adherence to LPV by providing continuously updated visualizations of key ventilation parameters. Deviations from predefined target ranges, derived from the hospital's local ventilation protocol, were highlighted to facilitate timely recognition and response. Target ranges varied according to the selected clinical protocol (acute respiratory distress syndrome, obstructive lung disease, normal LPV, or venovenous extracorporeal membrane oxygenation). Protocol selection was performed manually to align with the patient's clinical context. Access to the dashboard for authorized users was provided through a secure environment within the hospital's information technology domain.

For patients receiving invasive or noninvasive ventilation, the dashboard presented 2 levels of information: a unit-wide overview display and individual ventilator detail pages. The overview displayed all active ventilators on a single screen, including positive end-expiratory pressure (PEEP), fraction of inspired oxygen (FiO_2), V_T/IBW, ventilation mode, and basic patient characteristics (sex and height). Parameters were updated every minute, and values outside target ranges were highlighted,

XSL•FO
RenderX

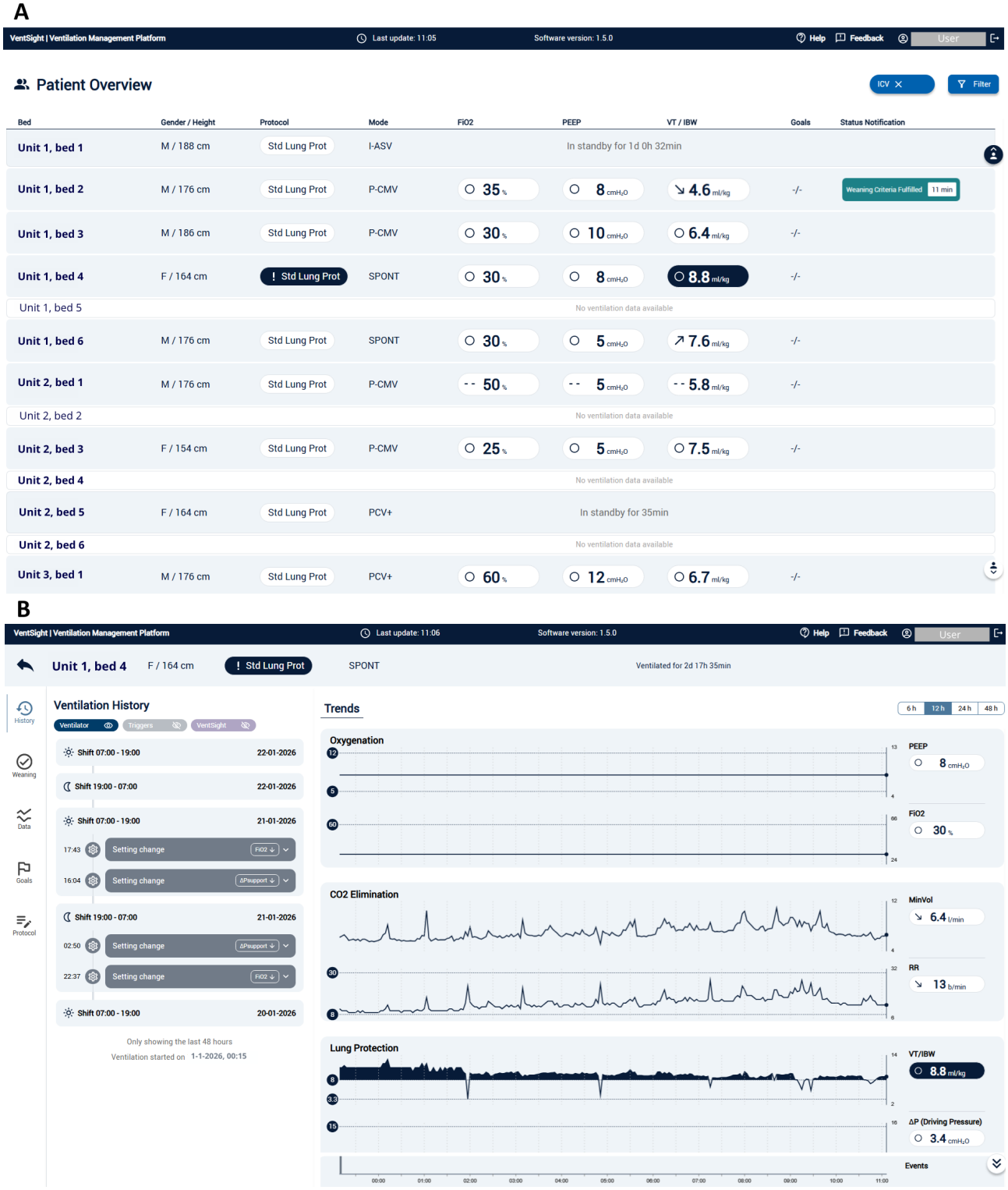
accompanied by trend arrows indicating changes over the preceding hours.

Selecting an individual ventilator opened detailed modules showing graphical trends such as FiO_2 , PEEP, V_T/IBW, respiratory rate and driving pressure, a weaning module aligned with local protocols, continuously refreshed pressure and flow waveforms, and additional ventilator readouts such as cumulative ventilation duration. An example of the dashboard interface is presented in Figure 1. The dashboard and implementation process are further described using the TIDieR (Template for Intervention Description and Replication) guidelines (Checklist 1).

The dashboard's design supported use across professional roles. For nurses, reviewing individual ventilator parameters for assigned patients aligned with their day-to-day bedside responsibilities. For physicians and ventilation practitioners, the unit-wide overview aligned with their roles in monitoring ventilation practices across the whole unit and identifying ventilators requiring additional attention. During the pilot period, the dashboard was not integrated into standard clinical care. Access to and use of the dashboard were limited to study participants during the study procedures and were voluntary. Dashboard use took place under the supervision of the leading researcher (SFCM) to prevent unintended clinical use.

XSL•FO
RenderX

Figure 1. Example of the ventilation dashboard's interface. The dashboard consists of 2 main layers: (A) the intensive care unit (ICU)–wide overview page, displaying all ventilators, and (B) the individual ventilator page, showing trends, waveforms, and protocol-relevant ventilator parameters.



MIDI-Based Determinants Assessment

The study intervention consisted of administering the MIDI questionnaire to identify determinants influencing the adoption of the ventilation dashboard in daily ICU practice [25].

Questionnaire Development

Determinants were assessed using the MIDI framework. The MIDI is intended to be tailored to the specific innovation and implementation context under study. Therefore, in this study, the MIDI was used as intended and tailored to the VentSight dashboard context without redeveloping the questionnaire or changing the underlying MIDI determinant structure. From the original 28 MIDI determinants, 17 were considered relevant to the ICU context and operationalized into 24 questionnaire items rated on a 5-point Likert scale ("completely disagree" to "completely agree"). Items were framed around the use of the

XSL•FO
RenderX

dashboard to support LPV, including timely recognition of deviations and adherence to local ventilation protocols.

Several MIDI determinants, such as personal benefits and outcome expectations, were operationalized using multiple items, which are reported separately in the "Results" section. Additional questions captured professional role, intention to use the dashboard (yes, no, or maybe), and open-ended reflections on perceived barriers and facilitators. A complete overview of items is provided in Multimedia Appendix 1.

The questionnaire was developed by a multidisciplinary team, including an implementation specialist, intensivist, ICU nurse, and physician, and was tailored into 2 role-specific versions. ICU nurses received a version aligned with individual ventilator assessment (individual assessment setup), whereas physicians, ventilation practitioners, and researchers received a version emphasizing the unit-wide overview (ICU overview setup). Both versions contained the same 24 MIDI items; only wording and examples were adapted to reflect role-specific use. The questionnaire was built and administered using Castor EDC.

Pilot Testing

Three ICU ventilation practitioners piloted the questionnaire to evaluate face validity, clarity, and feasibility of completing the questionnaire within the intended time frame (20 min). Feedback indicated that all items were understandable and could be completed within 15 minutes. Minor revisions were made to improve phrasing and logical flow. Pilot testing, therefore, focused on usability and contextual clarity rather than psychometric revalidation of the MIDI instrument itself.

Questionnaire Administration and Recruitment

Information about the study and the dashboard was disseminated 1 week before and during the questionnaire administration period via posters and the digital ICU newsletter. The MIDI questionnaire was administered between July 28, 2025, and August 4, 2025. The target sample size and distribution across professional groups were determined pragmatically, in line with the descriptive and exploratory purpose of this single-center implementation study. As the study aimed to identify determinants of dashboard adoption in the local ICU context rather than to test a predefined hypothesis or estimate an intervention effect, no formal power-based sample size calculation was performed.

Participants were recruited using a stratified convenience sampling approach. Professional groups involved in mechanical ventilation and dashboard use were considered relevant strata, and recruitment was targeted to ensure participation from each of these groups. Within these strata, participants were recruited based on availability and willingness to participate, rather than through random sampling. We aimed for approximately 75 respondents to obtain a broad representation of intended end users. This included approximately 40 ICU nurses, representing about 20% of the approximately 200 ICU nurses working in the ICU, and approximately half of the available physicians and intensivists involved in ICU ventilation care. The higher target number for ICU nurses reflected their larger share of the ICU workforce and their role as key end users of the individual ventilator assessment view. This approach was considered appropriate for the exploratory implementation-focused aim of the study and feasible within the local pilot implementation period.

All participants received a standardized 5- to 10-minute face-to-face onboarding session led by 1 researcher (SM), ensuring consistent exposure to the dashboard's key function. Participants then briefly explored the dashboard using short exploratory questions (Multimedia Appendix 2), followed by completion of the pseudonymized questionnaire via a QR code linking to the secure Castor EDC platform. Therefore, responses reflected initial perceptions after short-term, voluntary exposure, rather than experience with sustained use in routine clinical practice. All sessions were facilitated by the same researcher to ensure consistency of exposure. No modifications were made to the dashboard or study procedures during the questionnaire administration period.

Statistical Analysis

All analyses were descriptive, consistent with the exploratory aim of this single-center implementation study. Questionnaires with less than 70% completion were excluded from analysis to ensure adequate data completeness. For each MIDI item, we calculated the median and IQR, as well as the proportion of respondents selecting "agree" or "strongly agree" (scores 4 or 5) and "disagree" or "strongly disagree" (scores 1 or 2).

In accordance with the MIDI developer recommendations, determinants were classified as facilitators when ≥80% of respondents scored 4 or 5 and as barriers when ≥20% scored 1 or 2 [30]. Determinants were ranked descriptively by these proportions in the combined sample and examined separately by dashboard setup (individual assessment vs ICU overview) to account for differences in use context. Responses to open-ended questions were analyzed using a general inductive approach; responses were read iteratively and grouped into recurring thematic categories to summarize reported determinants. All analyses were performed using R statistical software (version 4.3.3; RStudio).

Results

Participant Characteristics

A total of 78 questionnaires were started. After exclusion of questionnaires with less than 70% item completion, 71 questionnaires were included in the analysis: 36 from the individual assessment setup and 35 from the ICU overview setup. Respondents represented a wide range of ICU professional roles, with ICU nurses forming the majority (39/71, 54.9%), followed by physicians (16/71, 22.5%) and intensivists or intensivist in training (11/71, 15.5%). Five respondents were specialized in mechanical ventilation, either as ventilation practitioners or members of their ICU's ventilation work group. Respondent characteristics are shown in Table 1.

Overall, 51 out of 71 (71.8%) respondents indicated that they intended to use the dashboard, while 20 out of 71 (28.2%) respondents answered "maybe." No respondents indicated that they did not intend to use the dashboard. In the individual assessment setup, 28 out of 36 (77.8%) respondents indicated

XSL•FO
RenderX

an intention to use, compared with 23 out of 35 (65.7%) respondents in the ICU overview setup.

Table . Respondent characteristics (N=71).

Characteristics	Value
Professional role, n (%)	
ICU[a] nurse	39 (54.9)
ICU nurse in training	7 (17.9)
ICU nurse (<5 years of experience)	16 (41)
ICU nurse (>5 years of experience)	16 (41)
Physician	16 (22.5)
Intensivist	10 (14.1)
Intensivist in training	1 (1.4)
Researcher in the field of mechanical ventilation	5 (7.0)
Specialized in mechanical ventilation[b]	5 (7.0)
Comfortable with applying mechanical ventilation on a 9-point Likert-scale, median (IQR)	8 (6-8)

[a]ICU: Intensive care unit.

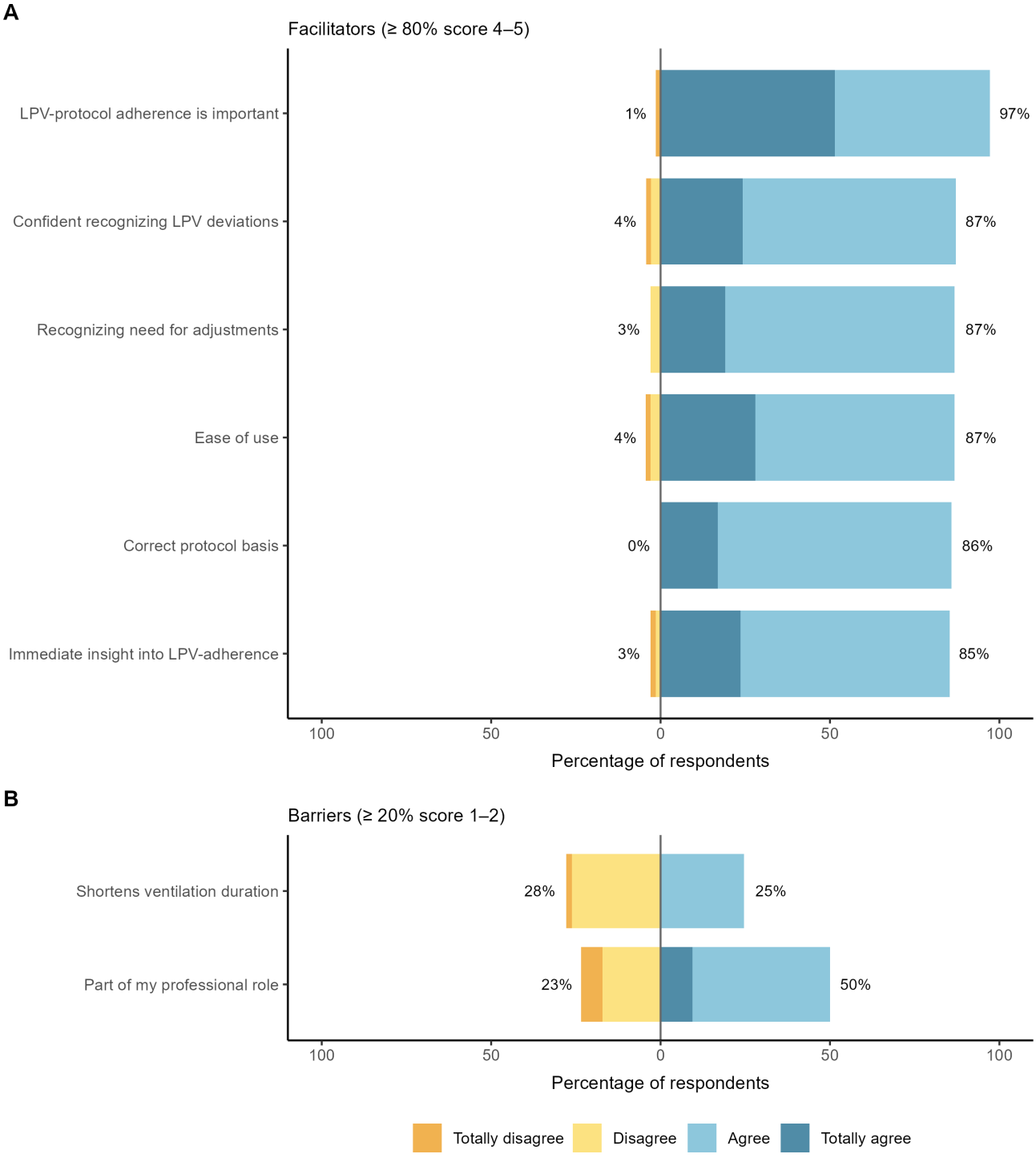
[b]Besides other professional role; for example, a nurse who is also a ventilation practitioner or intensivist who is a member of a ventilation working group.

Facilitators for Implementation

Six determinants met the predefined criteria for classification as facilitator (Figure 2). These included outcome expectations (MIDI item 9a, reflecting the importance of maintaining patients within protocol-defined ventilation targets), self-efficacy (MIDI item 16a, related to recognizing deviations from the ventilation protocol using the dashboard), procedural clarity, low perceived complexity, correctness of the dashboard, and observability. Outcome expectations and self-efficacy received the highest levels of agreement, with 97% (68/70) and 87% (61/70) of respondents, respectively, scoring “agree” or “strongly agree.” Of the 6 determinants classified as facilitators, 2 were related to the end user (outcome expectations and self-efficacy), whereas the remaining 4 facilitators were related to the innovation (procedural clarity, low perceived complexity, correctness, and observability).

XSL•FO
RenderX

Figure 2. Facilitators and barriers based on Measurement Instrument for Determinants of Innovations (MIDI) cutoffs. (A) Facilitators and (B) barriers identified among all participants based on the predefined MIDI thresholds. Facilitators were defined as MIDI items for which ≥80% of respondents scored 4–5, whereas barriers were defined as MIDI items for which ≥20% of respondents scored 1-2. LPV: lung-protective ventilation.

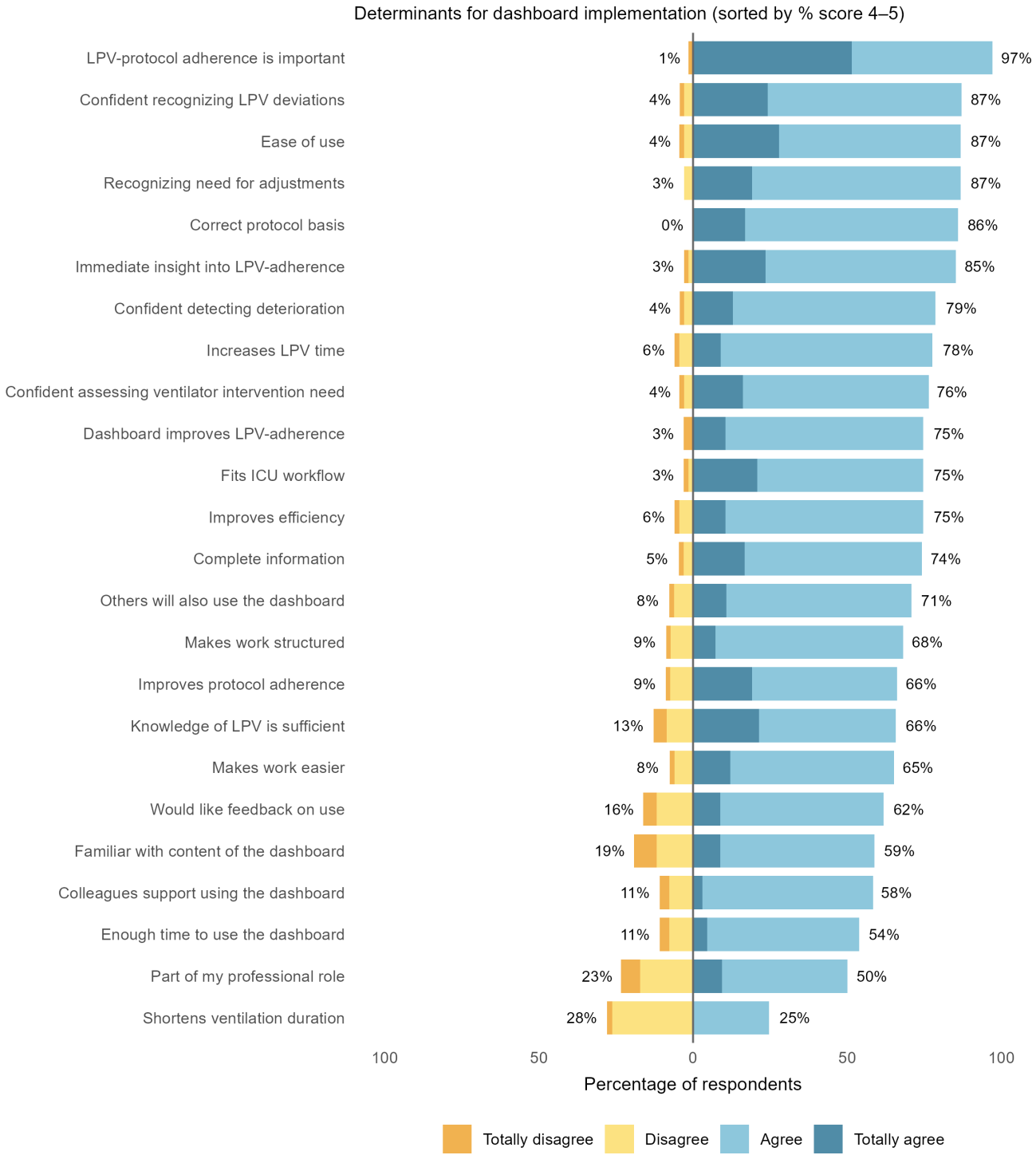


Barriers for Implementation

Two determinants met the predefined criteria for classification as barriers (Figure 2). These were relevance for client (MIDI item 7b, reflecting uncertainty about whether dashboard use would shorten mechanical ventilation); and professional obligation. Disagreement was reported by 28% (17/61) of respondents for relevance for client and by 23% (15/64) for professional obligation. Relevance for client is associated with the innovation, whereas professional obligation is associated with the adopting user. An overview of all determinants is presented in Figure 3.

XSL•FO
RenderX

Figure 3. Likert responses for all Measurement Instrument for Determinants of Innovations (MIDI) items sorted by agreement. Likert-scale responses (1-5) for all MIDI items, ordered by the proportion of respondents scoring 4-5 among all participants. ICU: intensive care unit; LPV: lung-protective ventilation.



Determinants Across Dashboard Setups

Facilitators differed by dashboard use context. Six facilitators were identified in the individual assessment setup, including outcome expectations (MIDI item 9a, importance of maintaining patients within protocol-defined ventilation targets), correctness, observability, self-efficacy related to recognizing protocol deviations and deteriorating ventilator settings, and low perceived complexity. Nine facilitators were identified in the ICU overview setup, including procedural clarity, outcome expectations, multiple self-efficacy items related to protocol monitoring, low perceived complexity, observability, perceived personal benefits related to efficiency, correctness, and compatibility with workflow (Multimedia Appendix 3).

Across both setups, relevance for the client (MIDI item 7b, reflecting uncertainty about whether dashboard use would shorten mechanical ventilation), and professional obligation were consistently identified as barriers. Feedback on performance was an additional barrier in the individual assessment setup, while limited awareness of the innovation's

XSL•FO
RenderX

content emerged as a barrier in the ICU overview setup. Detailed item-level results are provided in Multimedia Appendix 4.

Perceived Barriers and Facilitators

Responses to open-ended questions revealed that perceived barriers included the dashboard being considered extra work and time-consuming, requiring an additional device, presenting information already available in existing systems, and being perceived as not relevant to certain users' roles (particularly nurses).

Facilitators included providing the unit with a broader overview of all patients, supporting education, and enabling easier trend monitoring. Reported benefits were improved efficiency and time savings, enhanced situational overview, and the ability to monitor multiple patients simultaneously. Reported disadvantages were the need for an additional screen and the potential for increased screen-focused work, which could reduce direct patient contact. All open-ended responses are presented in Multimedia Appendix 5.

Discussion

Main Findings

Using the MIDI, we found that adoption of the dashboard aimed at supporting LPV was mainly facilitated by high perceived outcome expectations, self-efficacy, and procedural clarity, while relevance for the client and professional obligation emerged as barriers to implementation. Facilitators and barriers differed slightly by dashboard use context. Most respondents intended to use the dashboard in clinical practice.

Strengths

This study has several strengths. The questionnaire was developed using the validated MIDI framework [24,25], ensuring systematic assessment of a wide range of determinants for implementation. The questionnaire was tailored to 2 user contexts (individual assessment- and ICU overview setup) and piloted for clarity and feasibility. Recruitment during active demonstration of the dashboard ensured that respondents were familiar with the tool before providing feedback. The combination of quantitative Likert-scale items with open-ended questions provided both structured and nuanced insights into perceived facilitators and barriers. The MIDI provided a structured approach to assess determinants of dashboard adoption in the ICU context. Combining structured ratings with open-ended responses allowed us to identify facilitators and barriers and to descriptively compare findings across dashboard use contexts.

Limitations

Some limitations must be considered. First, the study was conducted in a single tertiary ICU in the Netherlands, which may limit generalizability to other settings with different workflows, staffing models, or digital infrastructure. Although the specific determinants identified are context- and innovation-dependent, several reflect broader features of ICU practice and may therefore be relevant beyond this single dashboard. Moreover, this study illustrates how a framework-based approach, such as the MIDI, can be used to systematically identify context-specific barriers and facilitators for digital tool adoption across ICU settings. Second, this was a descriptive and exploratory implementation study with a pragmatically determined sample size. The study was not designed or powered to test predefined hypotheses, estimate intervention effects, or perform formal subgroup comparisons. The findings should therefore be interpreted as context-specific insights into determinants of dashboard adoption, rather than as precise estimates generalizable to all ICU health care professionals. In addition, the professional distribution of respondents was pragmatically determined and was not based on formal stratified random sampling, which may limit interpretation of differences between professional groups. Third, participation was voluntary, which may have introduced selection bias; those more interested in innovation or ventilation practices may have been more likely to respond. Social desirability bias may also have influenced responses, despite pseudonymization, as the dashboard was a local initiative. Fourth, respondents were introduced to the dashboard through a brief standardized onboarding session rather than through prolonged real-world use, and dashboard use was voluntary and limited to study participants. The timing and voluntary context should be considered when interpreting perceived ease of use and intention to use [31]. Initial perceptions shortly after introduction may differ from perceptions after sustained routine use, particularly once users have more opportunity to evaluate workflow integration, peer use, and the dashboard's practical contribution to ventilation care. To partially mitigate limited exposure, participants were given time to explore the dashboard themselves by completing several short explanatory tasks, allowing hands-on interaction that approximated early real-world exposure. Finally, as this was a perception-based questionnaire, findings do not directly measure actual changes in LPV adherence or patient outcomes.

Relation With Prior Work

Our findings align with previous studies showing that adoption of change and implementation of digital tools in ICU settings are shaped by a complex and interconnected set of facilitators and barriers operating across individual, organizational, and health system levels, rather than by technological usability alone [20-23]. Consistent with earlier studies, determinants identified in our study emphasize the importance of staff engagement, professional beliefs, and perceived relevance of the innovation, alongside feasibility within existing workflows and workload constraints [20,23]. Both the MIDI-based assessment and the open-ended responses suggest that barriers to adoption predominantly arise from challenges related to integration into daily practice, competing clinical demands, and unclear professional responsibility, reinforcing the notion that successful uptake of digital tools in critical care depends on alignment with human and organizational factors within the ICU context [13-15,19].

Interpretation of Findings

The high intention-to-use rate suggests initial acceptability of the dashboard among a wide range of ICU professionals, although this should be interpreted in the context of voluntary participation and short-term exposure during a study-specific

XSL•FO
RenderX

onboarding session. This finding can be interpreted in light of technology acceptance models such as the technology acceptance model (TAM) and TAM2, in which perceived usefulness and perceived ease of use are important predictors of usage intention and actual system use [31,32]. Although these constructs were not directly measured by the MIDI, several identified facilitators may reflect conditions that support them. High scores for procedural clarity, low perceived complexity, correctness, and observability suggest that the dashboard was experienced as clear, reliable, and straightforward to interpret, supporting perceived ease of use in a complex clinical environment. High outcome expectations and self-efficacy may similarly reflect perceived usefulness and confidence in using the dashboard for protocol monitoring. Together, these findings may help explain the high initial intention to use the dashboard.

However, our findings indicate that successful implementation is unlikely to be achieved by introducing the technology alone; it requires clear integration strategies into daily workflows. This is particularly relevant in light of TAM2, which emphasizes job relevance as an important determinant of perceived usefulness and user acceptance. In our study, low scores for relevance for client, reflecting uncertainty about patient-related benefit, and professional obligation emerged as barriers [31]. This suggests that the dashboard's relevance for patient care and its place within shared professional responsibilities were not yet fully established across professional groups. The lack of perceived personal benefit, particularly among respondents in the individual assessment setup, may further reduce the likelihood of sustained use. Addressing these barriers should therefore be a key focus of future implementation strategies, including targeted education on the downstream clinical impact of consistent LPV and explicit clarification of how dashboard use contributes to team-based ventilation care.

Because the dashboard was not yet integrated into standard clinical care, respondents' perceptions may differ from perceptions after sustained routine use. In particular, perceived usefulness may increase once users have more opportunity to experience how the dashboard fits into daily workflow, supports shared situational awareness, and contributes to ventilation monitoring over time. Although continuous display of the dashboard at the nursing station was not formally evaluated in this pilot, our findings suggest that such continuous visibility may further reinforce situational awareness and promote routine use, particularly among bedside nurses. In addition, introducing a structured "ventilation review moment" on each ICU, using the dashboard as a central reference tool, could support its use in daily practice, promote protocol adherence, enhance shared situational awareness, align perspectives on ventilation status across professionals, and reinforce shared responsibility for LPV.

Several determinants, including time available, social support, and awareness of the innovation's content, did not meet formal barrier thresholds but showed comparatively low scores. These factors may reflect practical constraints and limited peer reinforcement rather than definitive barriers, yet they represent areas of vulnerability that could hinder adoption if left unaddressed. Embedding dashboard use within a dedicated, team-based ventilation review moment may help mitigate these constraints by normalizing use, strengthening shared responsibility, and increasing familiarity with the dashboard's functionality.

Although the specific configuration and relative importance of determinants identified in this study are shaped by the local ICU context and the characteristics of this dashboard, the findings may point to more general challenges inherent to implementing digital support tools in critical care settings. Determinants such as limited time availability, uncertainty about perceived patient-related benefits, and unclear professional responsibility may reflect underlying features of ICU work, including high workload, task fragmentation, and shared responsibility across professional roles. Rather than implying that these determinants apply uniformly across settings, our findings illustrate how implementation challenges in critical care often arise from the interaction between human, organizational, and contextual factors, extending beyond technical usability alone.

These findings indicate that implementation strategies for nurses should focus on building buy-in and peer support for dashboard use, allocating protected time, and reinforcing the relevance and positive effects of LPV. For physicians, role-specific strategies should emphasize education highlighting the patient benefits of LPV and emphasizing that delivering LPV as a team aligns with professional responsibilities. Practical next steps include tailoring role-specific dashboard views, integrating the dashboard directly into existing systems, and introducing structured moments to jointly assess and evaluate ventilation practices. Finally, clearly communicating the rationale and benefits of LPV, while fostering a culture in which feedback on ventilation practices is normalized and acted upon, may substantially enhance adoption and quality of patient care.

Conclusions

Using the MIDI questionnaire to evaluate adoption of a ventilation dashboard, we found that ICU professionals generally perceived the dashboard as feasible and useful and reported a high intention to use it, while also identifying adoption barriers. Our findings highlight that successful adoption of digital dashboards in critical care depends on human, organizational, and contextual factors at the workplace. Recognizing and addressing these factors is essential for implementing digital tools that aim to improve quality of care in the ICU.

Acknowledgments

The authors would like to thank the 3 intensive care unit (ICU) ventilation practitioners who piloted the questionnaire for their valuable feedback on clarity and feasibility. The authors also gratefully acknowledge the colleagues from the departments of Clinical Engineering Department (Ruud Voorn and Gertjan van der Veldt), ICT (Sanne Terpstra), and the EHR Innovation and

XSL•FO
RenderX

Implementation Center (Jasper van Dijk) at Amsterdam UMC for their essential support in making the practical implementation of the VentSight dashboard possible. Finally, the authors would like to thank the ICU nursing staff, physicians, and ventilation practitioners at Amsterdam UMC for their participation in this study. Generative AI was used solely to check spelling and improve the English writing of the manuscript. ChatGPT (OpenAI) was not used to generate scientific content, perform analyses, or interpret the results.

Funding

This study was partly funded by Hamilton Medical, which covered project costs for piloting and evaluating the VentSight ventilation dashboard. The funder had no role in study design, data collection, analysis, interpretation of data, or writing of the manuscript.

Data Availability

The datasets generated and/or analyzed during the current study are available from the corresponding author upon request.

Authors' Contributions

SFCM, DAD, and FP designed the study. SFCM led the project preparation, coordinated the study project, conducted the pilot, collected the data, performed the analysis, and drafted the manuscript. TDVL contributed extensively to project preparation by piloting the questionnaire and pilot preparation. MCR contributed to the preparation and installation of the piloted dashboard. AV contributed to questionnaire development and study design. DAD and FP supervised the project and contributed to study conception and interpretation of findings. All authors have read and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Items mapped to the Measurement Instrument for Determinants of Innovations (MIDI) determinants and all questionnaire items.
[DOCX File, 29 KB - humanfactors_v13i1e92313_app1.docx]

Multimedia Appendix 2
Questions asked during onboarding to improve interaction with the dashboard.
[DOCX File, 24 KB - humanfactors_v13i1e92313_app2.docx]

Multimedia Appendix 3
Likert scores for the individual assessment and intensive care unit (ICU) overview setup.
[PNG File, 509 KB - humanfactors_v13i1e92313_app3.png]

Multimedia Appendix 4
Results: Likert scores and intention to use.
[DOCX File, 37 KB - humanfactors_v13i1e92313_app4.docx]

Multimedia Appendix 5
Open-ended questions and responses from the questionnaire.
[DOCX File, 40 KB - humanfactors_v13i1e92313_app5.docx]

Checklist 1
TIDieR checklist.
[DOCX File, 26 KB - humanfactors_v13i1e92313_app6.docx]

References

1. Slutsky AS, Ranieri VM. Ventilator-induced lung injury. N Engl J Med 2013 Nov 28;369(22):2126-2136. [doi: 10.1056/NEJMra1208707] [Medline: 24283226]
2. Amato MB, Barbas CS, Medeiros DM, et al. Effect of a protective-ventilation strategy on mortality in the acute respiratory distress syndrome. N Engl J Med 1998 Feb 5;338(6):347-354. [doi: 10.1056/NEJM199802053380602] [Medline: 9449727]
3. Amato MBP, Meade MO, Slutsky AS, et al. Driving pressure and survival in the acute respiratory distress syndrome. N Engl J Med 2015 Feb 19;372(8):747-755. [doi: 10.1056/NEJMsa1410639] [Medline: 25693014]

XSL•FO
RenderX

4. Acute Respiratory Distress Syndrome Network, Brower RG, Matthay MA, et al. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. N Engl J Med 2000 May 4;342(18):1301-1308. [doi: 10.1056/NEJM200005043421801] [Medline: 10793162]
5. Serpa Neto A, Cardoso SO, Manetta JA, et al. Association between use of lung-protective ventilation with lower tidal volumes and clinical outcomes among patients without acute respiratory distress syndrome: a meta-analysis. JAMA 2012 Oct 24;308(16):1651-1659. [doi: 10.1001/jama.2012.13730] [Medline: 23093163]
6. Serpa Neto A, Nagtzaam L, Schultz MJ. Ventilation with lower tidal volumes for critically ill patients without the acute respiratory distress syndrome: a systematic translational review and meta-analysis. Curr Opin Crit Care 2014 Feb;20(1):25-32. [doi: 10.1097/MCC.0000000000000044] [Medline: 24275571]
7. Botta M, Tsonas AM, Pillay J, et al. Ventilation management and clinical outcomes in invasively ventilated patients with COVID-19 (PRoVENT-COVID): a national, multicentre, observational cohort study. Lancet Respir Med 2021 Feb;9(2):139-148. [doi: 10.1016/S2213-2600(20)30459-8] [Medline: 33169671]
8. Midega TD, Bozza FA, Machado FR, et al. Organizational factors associated with adherence to low tidal volume ventilation: a secondary analysis of the CHECKLIST-ICU database. Ann Intensive Care 2020 Jun 1;10(1):68. [doi: 10.1186/s13613-020-00687-3] [Medline: 32488524]
9. Santamaria JD, Tobin AE, Reid DA. Do we practise low tidal-volume ventilation in the intensive care unit? a 14-year audit. Crit Care Resusc 2015 Jun;17(2):108-112. [doi: 10.1016/s1441-2772(23)01050-5] [Medline: 26017128]
10. Umoh NJ, Fan E, Mendez-Tellez PA, et al. Patient and intensive care unit organizational factors associated with low tidal volume ventilation in acute lung injury. Crit Care Med 2008 May;36(5):1463-1468. [doi: 10.1097/CCM.0b013e31816fc3d0] [Medline: 18434907]
11. Abrahamson KA, Fox RL, Doebbeling BN. Facilitators and barriers to clinical practice guideline use among nurses. Am J Nurs 2012 Jul;112(7):26-35. [doi: 10.1097/01.NAJ.0000415957.46932.bf] [Medline: 22705494]
12. Rosa RG, Teixeira C, Sjoding M. Novel approaches to facilitate the implementation of guidelines in the ICU. J Crit Care 2020 Dec;60:1-5. [doi: 10.1016/j.jcrc.2020.07.014] [Medline: 32731099]
13. Ahmed A, Chandra S, Herasevich V, Gajic O, Pickering BW. The effect of two different electronic health record user interfaces on intensive care provider task load, errors of cognition, and performance. Crit Care Med 2011 Jul;39(7):1626-1634. [doi: 10.1097/CCM.0b013e31821858a0] [Medline: 21478739]
14. Khairat SS, Dukkipati A, Lauria HA, Bice T, Travers D, Carson SS. The impact of visualization dashboards on quality of care and clinician satisfaction: integrative literature review. JMIR Hum Factors 2018 May 31;5(2):e22. [doi: 10.2196/humanfactors.9328] [Medline: 29853440]
15. Tandon P, Nguyen KAN, Raut G, Ouyang Y, Cheetirala SN, Ranginwala S, et al. Impact of a real-time ventilator management dashboard with alerts: sustained hospital-wide improvement in lung protective ventilation. CHEST Crit Care 2025;3(4):100167. [doi: 10.1016/j.chstcc.2025.100167]
16. Bourdeaux CP, Thomas MJ, Gould TH, et al. Increasing compliance with low tidal volume ventilation in the ICU with two nudge-based interventions: evaluation through intervention time-series analyses. BMJ Open 2016 May 26;6(5):e010129. [doi: 10.1136/bmjopen-2015-010129] [Medline: 27230998]
17. Lai CH, Li KW, Hu FW, et al. Integration of an intensive care unit visualization dashboard (i-Dashboard) as a platform to facilitate multidisciplinary rounds: cluster-randomized controlled trial. J Med Internet Res 2022 May 13;24(5):e35981. [doi: 10.2196/35981] [Medline: 35560107]
18. Bourdeaux CP, Birnie K, Trickey A, et al. Evaluation of an intervention to reduce tidal volumes in ventilated ICU patients. Br J Anaesth 2015 Aug;115(2):244-251. [doi: 10.1093/bja/aev110] [Medline: 25979150]
19. Johnson MJ, May CR. Promoting professional behaviour change in healthcare: what interventions work, and why? A theory-led overview of systematic reviews. BMJ Open 2015 Sep 30;5(9):e008592. [doi: 10.1136/bmjopen-2015-008592] [Medline: 26423853]
20. Prevedello D, Steckelmacher C, Devroey M, Njimi H, Creteur J, Preiser JC. The burden of implementation: a mixed methods study on barriers to an ICU follow-up program. J Crit Care 2021 Oct;65:170-176. [doi: 10.1016/j.jcrc.2021.06.006] [Medline: 34171692]
21. Kleinpell R, Zimmerman J, Vermoch KL, et al. Promoting family engagement in the ICU: experience from a national collaborative of 63 ICUs. Crit Care Med 2019 Dec;47(12):1692-1698. [doi: 10.1097/CCM.0000000000004009] [Medline: 31567354]
22. Sun Y, Ding M, Liu CC, Dong WS, Xia Y, Wei L. Implementation science research in the intensive care unit: a scoping review. BMJ Open 2025 Nov;15(11):e106654. [doi: 10.1136/bmjopen-2025-106654] [Medline: 41263853]
23. Borges do Nascimento IJ, Abdulazeem H, Vasanthan LT, et al. Barriers and facilitators to utilizing digital health technologies by healthcare professionals. NPJ Digit Med 2023 Sep 18;6(1):161. [doi: 10.1038/s41746-023-00899-4] [Medline: 37723240]
24. Fleuren MAH, Paulussen TGWM, Van Dommelen P, Van Buuren S. Towards a measurement instrument for determinants of innovations. Int J Qual Health Care 2014 Oct;26(5):501-510. [doi: 10.1093/intqhc/mzu060] [Medline: 24951511]
25. Fleuren MAH, Paulussen T, Van Dommelen P, Van Buuren S. Measurement instrument for determinants of innovations (MIDI). : TNO; 2014 URL: https://publications.tno.nl/publication/34640154/3UJxX5/TNO-2014-measurement.pdf [accessed 2026-09-02]

XSL•FO
RenderX

26. van Lummel EVTJ, Meijer Y, Tjan DHT, van Delden JJM. Barriers and facilitators for healthcare professionals to the implementation of multidisciplinary timely undertaken advance care planning conversations at the outpatient clinic (the MUTUAL intervention): a sequential exploratory mixed-methods study. BMC Palliat Care 2023 Mar 15;22(1):24. [doi: 10.1186/s12904-023-01139-y] [Medline: 36922796]
27. van Veenendaal H, Voogdt-Pruis HR, Ubbink DT, et al. Evaluation of a multilevel implementation program for timeout and shared decision making in breast cancer care: a mixed methods study among 11 hospital teams. Patient Educ Couns 2022 Jan;105(1):114-127. [doi: 10.1016/j.pec.2021.05.005] [Medline: 34016497]
28. Siebelink NM, Gaasterland A, Gielissen M, van der Weegen S, Boon B, van der Poel A. Barriers and facilitators influencing implementation of care technology for people with intellectual disabilities: a cross-sectional study among care professionals. J Appl Res Intellect Disabil 2024 Sep;37(5):e13262. [doi: 10.1111/jar.13262] [Medline: 38946655]
29. Dugstad J, Sundling V, Nilsen ER, Eide H. Nursing staff's evaluation of facilitators and barriers during implementation of wireless nurse call systems in residential care facilities. A cross-sectional study. BMC Health Serv Res 2020 Mar 4;20:163. [doi: 10.1186/s12913-020-4998-9] [Medline: 32131815]
30. Verberne LM, Kars MC, Schepers SA, Schouten-van Meeteren AYN, Grootenhuis MA, van Delden JJM. Barriers and facilitators to the implementation of a paediatric palliative care team. BMC Palliat Care 2018 Feb 12;17(1):23. [doi: 10.1186/s12904-018-0274-8] [Medline: 29433576]
31. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manage Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
32. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]

Abbreviations

Amsterdam UMC: Amsterdam University Medical Center
FiO_2 : fraction of inspired oxygen
ICU: intensive care unit
LPV: lung-protective ventilation
MIDI: Measurement Instrument for Determinants of Innovations
PEEP: positive end-expiratory pressure
TAM: technology acceptance model
TIDieR: Template for Intervention Description and Replication
V_T/IBW: tidal volume indexed to ideal body weight
WMO: Dutch Medical Research Involving Human Subjects Act

Edited by A Kushniruk; submitted 28.Jan.2026; peer-reviewed by DS Karbing, M Ding; revised version received 23.Jun.2026; accepted 22.Jul.2026; published 07.Oct.2026.

Please cite as:
Mugge SFC, van Leijsen TD, Reuland MC, Visser A, Paulus F, Dongelmans DA
Determinants of Adoption of a Mechanical Ventilation Dashboard in Intensive Care: Measurement Instrument for Determinants of Innovations–Based, Single-Center, Cross-Sectional Evaluation
JMIR Hum Factors 2026;13:e92313
URL: https://humanfactors.jmir.org/2026/1/e92313
doi:10.2196/92313

© Sarah Frederika Christina Mugge, Tobias Daniel van Leijsen, Merijn Constantijn Reuland, Annelies Visser, Frederique Paulus, Dave Anton Dongelmans. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 7.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

AI-Based Rehabilitation System for Congenital Muscular Torticollis: Formative Usability Evaluation

Seojin Hong[1], MS; Heewon Nam[1], BS; Seongwoo Jeong[1], MS; Hyun Choi[2], PhD; Hyosun Kweon[1], PhD

[1]Department of Clinical Rehabilitation Research, Rehabilitation Research Institute, National Rehabilitation Center, 58 Samgaksan-ro, Gangbuk-gu, Seoul, Republic of Korea

[2]Department of Healthcare and Public Health, Rehabilitation Research Institute, National Rehabilitation Center, Seoul, Republic of Korea

Corresponding Author:
Hyosun Kweon, PhD
Department of Clinical Rehabilitation Research, Rehabilitation Research Institute, National Rehabilitation Center, 58 Samgaksan-ro, Gangbuk-gu, Seoul, Republic of Korea

Abstract

Background: Home-based exercise programs for congenital muscular torticollis (CMT) depend heavily on caregivers, yet adherence is often hindered by cognitive and emotional burdens.

Objective: This study conducted a formative usability evaluation of an AI-based CMT rehabilitation system comprising a clinician web interface and a caregiver mobile app to identify usability-related issues and preliminary design requirements during early development.

Methods: A formative usability evaluation was conducted with 8 clinical experts (rehabilitation physicians, nurses, and physical therapists). Participants engaged in focus group interviews (FGIs) following a demonstration of the system and completed the System Usability Scale (SUS) for the overall system, including a proxy evaluation of the caregiver mobile app. Qualitative data were analyzed using thematic analysis, and SUS scores were analyzed descriptively as a supplementary quantitative measure of perceived usability.

Results: The thematic analysis identified 2 core themes, "Usability Barriers of the CMT Rehabilitation System" and "Improvement Requirements," encompassing 10 subthemes and 21 concepts. Key findings highlighted concerns regarding the reliability and interpretability of AI-generated feedback and the need to simplify data visualization to reduce the potential cognitive burden on caregivers. The mean SUS score for the overall system was 73.44 (SD 18.01). This result was interpreted descriptively and considered supplementary to the qualitative findings.

Conclusions: The formative evaluation identified preliminary usability strengths and areas for improvement from clinicians' perspectives, with particular concerns regarding the reliability and interpretability of AI-generated feedback. The experts emphasized that the AI-based CMT rehabilitation system should function as a therapeutic bridge between clinicians and caregivers. As actual caregivers were not included in this evaluation, the findings represent preliminary formative design inputs based on clinicians' proxy perspectives rather than evidence of usability for caregivers. Future iterations should incorporate the identified improvements and undergo evaluation with actual caregivers in home environments, along with separate technical verification of the accuracy of the AI feedback algorithm.

(JMIR Hum Factors 2026;13:e101427) doi:10.2196/101427

KEYWORDS

congenital muscular torticollis; rehabilitation; formative evaluation; usability; human-centered design; mobile health; mhealth app; caregivers

Introduction

Background

Congenital muscular torticollis (CMT) is a representative musculoskeletal disorder in infancy characterized by an abnormal head and neck posture, in which the head tilts to one side while the chin rotates toward the opposite side [1]. Although CMT is relatively common during early infancy, a lack of timely intervention can lead to long-term functional impairments, such as craniofacial asymmetry, spinal deformities, and developmental delays. Consequently, early diagnosis and prompt intervention for CMT are consistently emphasized as essential elements for ensuring normal growth and functional development [2].

According to previous research, when physical therapy such as stretching and postural correction is initiated within the first month of life, approximately 98% of infants have been reported to recover a normal range of cervical motion within an average

XSL•FO
RenderX

period of 1.5 months. Furthermore, initiating treatment before 3 to 4 months of age typically allows for sufficient improvement through conservative interventions alone. Conversely, after 3 to 4 months, increased resistance as infants develop head control often results in a reduced therapeutic response. Delayed diagnosis and intervention increase the likelihood of requiring invasive treatments, such as botulinum toxin injections or surgery [3]. Therefore, in the treatment of congenital muscular torticollis, it is paramount to promptly initiate conservative rehabilitation interventions including physical therapy, stretching, and postural correction based on an early diagnosis [4,5].

Rehabilitation therapy for CMT relies heavily not only on physical therapy in clinical settings but also on home exercise programs (HEP) performed repeatedly by caregivers in their daily lives [6,7]. However, in actual clinical practice, it is frequently reported that treatment adherence is compromised because caregivers often lack a sufficient understanding of the purpose and methods of the exercises, or struggle to perform them with the correct posture and intensity [8]. In particular, for rehabilitation exercises involving infants, continuity of therapy is significantly influenced not only by the physical burden on caregivers but also by psychological factors, such as anxiety and a lack of confidence in performing the exercises [9,10]. Recent studies have emphasized the necessity of designing rehabilitation programs that extend beyond simple exercise prescriptions by accounting for the emotional experiences and cognitive burdens of caregivers [3,8]. These limitations act as primary factors hindering the effectiveness of torticollis rehabilitation, suggesting an urgent need for systematic tools that support caregivers in performing rehabilitation exercises safely and accurately.

To address these limitations, the CMT rehabilitation system is being developed to support caregivers in performing rehabilitation exercises more accurately and consistently within the home environment. By integrating AI capabilities, this system aims to provide caregivers with visual and cognitive guidance throughout the exercise process, thereby enhancing both the accuracy and consistency of performance. Furthermore, the CMT rehabilitation system possesses the potential to overcome the primary challenges identified in existing practices, such as low treatment adherence, insufficient exercise accuracy, and the limitations of continuous monitoring by clinical professionals. Consequently, the system is expected to facilitate caregiver engagement in the rehabilitation process and, ultimately, contribute to achieving successful clinical outcomes for infants with CMT.

Usability

Usability in medical devices refers to the characteristics of a user interface that support users in achieving effectiveness, efficiency, and user satisfaction when operating the device within its intended use environment [11]. Particularly in software as a medical device (SaMD), usability evaluation is considered a core risk management activity that extends beyond simple improvements in user convenience; it ensures patient safety by guaranteeing that users can operate the device safely and accurately [12]. Software medical devices that fail to adequately incorporate usability principles are highly susceptible to use errors, which can lead to unintended use, improper intervention, and diminished therapeutic efficacy. Consequently, systematic usability evaluation has become an essential procedure in the development process of software medical devices.

Since the CMT rehabilitation system is designed for caregivers to directly perform interventions on infants, various human factors such as the user's level of understanding, cognitive burden, and perception of responsibility directly influence the safety and effectiveness of the product. Due to these characteristics, it is particularly crucial to conduct a formative evaluation during the early development phase to perform an in-depth exploration of the entire usage process and to identify potential use errors.

Focus group interviews (FGIs) are a qualitative research method applicable at any stage of the medical device lifecycle, offering the advantage of comprehensively exploring user experiences, perceptions, design issues, and usage contexts [13]. In particular, through interactions among participants with diverse professional backgrounds, FGIs can provide insights that are difficult to elicit in individual interviews. Therefore, FGIs are a highly appropriate methodology for investigating early-stage usability issues in medical devices with complex usage contexts, such as the CMT rehabilitation system [14].

Aim

The aim of this study was to conduct a formative usability evaluation of a prototype AI-based CMT rehabilitation system, comprising a clinician web interface and a caregiver mobile app, from the perspectives of clinical experts. Through FGIs and a usability assessment, we sought to identify usability-related issues, potential use errors, and preliminary design requirements to guide further system development. As caregivers were not included in this study, the caregiver-facing app was evaluated from clinicians' proxy perspectives, with the findings intended to inform future evaluation with actual caregivers.

Methods

Study Design

This study used a formative usability evaluation that included FGIs and a survey with 8 clinical experts, including rehabilitation medicine physicians, nurses, and physical therapists. The evaluation was designed to identify usability-related issues, potential use errors, and areas for improvement in the prototype CMT rehabilitation system from the perspectives of clinical experts. Rehabilitation medicine physicians and nurses provided perspectives relevant to the clinician-facing web interface, while physical therapists contributed domain expertise in CMT rehabilitation. The caregiver-facing mobile app was evaluated from the clinical experts' proxy perspectives, as actual caregivers were not included in this study. The findings were used to derive preliminary formative design inputs to guide subsequent iterations of the system rather than to establish the usability of the caregiver-facing app for actual caregivers.

XSL•FO
RenderX

Materials

The CMT rehabilitation system used in this study was developed to facilitate the provision of rehabilitation exercises for infants under 12 months of age diagnosed with CMT. The system enables caregivers to perform rehabilitation interventions for their infants using a mobile app. The CMT rehabilitation system consists of a web interface for clinical staff and a mobile app for the infant and caregiver. The intended users of the web interface are health care professionals (physicians and nurses), while the mobile app is intended for caregivers.

Health care professionals access the web interface to diagnose infants who require CMT rehabilitation, prescribe appropriate exercises, and manage or supervise the treatment process. Caregivers use the mobile app to follow guided exercise videos, enabling them to provide direct rehabilitation therapy to the infant. The treatment interface within the mobile app provides AI-based feedback intended to support caregivers during exercise performance. Furthermore, the app allows for the recording of participation and adherence upon completion of each session.

Participants

The formative usability evaluation was designed with reference to the usability engineering process outlined in IEC 62366-1:2015+ AMD1:2020 [15]. For the focus group interviews, we aimed to recruit 5 to 8 clinical experts, including rehabilitation medicine physicians, nurses, and physical therapists involved in the diagnosis, treatment, or outpatient support of patients with CMT, to obtain multidisciplinary perspectives on the clinical workflow [16]. A total of 8 clinical experts participated in the study: 2 rehabilitation medicine physicians, 2 nurses, and 4 physical therapists.

This formative evaluation was limited to clinical experts; caregivers, who are the intended users of the caregiver-facing mobile app, were not recruited. Accordingly, the clinicians evaluated the caregiver-facing app from a proxy clinical perspective, and the usability of the app from the perspective of actual caregivers was not assessed in this study.

The inclusion criteria were as follows: (1) clinical experience in the diagnosis, treatment, or outpatient support of CMT; (2) sufficient proficiency in Korean and English to understand the terminology used in the system; and (3) provision of informed consent to participate in the evaluation. Individuals who did not meet the inclusion criteria or did not provide informed consent were excluded.

Data saturation was not used as a criterion for determining the sample size because the number and composition of participants were predefined for this formative usability evaluation.

Procedure

The FGIs for the formative usability evaluation were conducted in a single session for each group. Participants were organized into 2 groups to capture complementary clinical perspectives based on their professional roles and experience with CMT care: group A consisted of 2 rehabilitation medicine physicians and 2 nurses involved in the clinical management and outpatient support of patients with CMT, and group B consisted of 4 physical therapists with expertise in CMT rehabilitation (Figure 1). Both groups comprised clinical experts, and no caregivers participated in the evaluation.

The formative evaluation procedure began with a briefing for participants regarding the purpose, overview, and methodology of the CMT rehabilitation system. Following the acquisition of informed consent, a preinterview survey was conducted to collect demographic and professional characteristics. The interviews were conducted based on a pre-established scenario and included a live demonstration of the system. Participants were provided with opportunities to interact directly with the product as necessary, while the evaluation facilitator observed and recorded the proceedings. Postinterview surveys were administered to gather additional feedback, and a standardized set of interview questions was applied to both focus groups (Table 1). The question protocol was designed to explore usability-related issues and potential areas for improvement from the participating clinical experts' perspectives, in accordance with the evaluation scenario. Each FGI session lasted approximately 100 minutes (Table 2).

XSL•FO
RenderX

Table . Formative evaluation procedures, involving practitioners and evaluators.

Composition	Details	Time (min)
Orientation	• The facilitator introduces the concept of formative evaluation to participants who are unfamiliar with it. • The facilitator explains the target product, purpose, overview, methodology, and other relevant aspects of formative evaluation to the participant.	30
Guidance on consent	• The facilitator provides the participant with detailed instructions regarding recording and transcription procedures. • The facilitator ensures that the participant fully comprehends the evaluation content and voluntarily provides written consent to participate in formative evaluation.	10
Conduct FGI[a]	• The facilitator conducted user training, demonstrations, and interviews according to a predetermined script. • The participants observed the demonstrations, responded to interview questions, and were allowed to operate the product if necessary. • An observer documented the entire interview process.	40
Perform SUS[b]	• After completing the FGI, the facilitator administered the SUS to assess perceived usability of the overall CMT[c] rehabilitation system, including a proxy evaluation of the caregiver-facing mobile app.	20

[a]FGI: focus group interview.

[b]SUS: System Usability Scale.

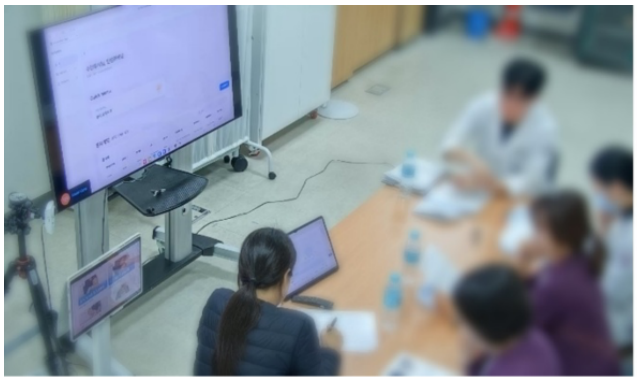
[c]CMT: congenital muscular torticollis.

XSL•FO
RenderX

Table . List of focus group interview questions.

Items	Statement
Opening questions	• General characteristics (age, sex, years of clinical experience, and experience using similar medical devices or providing related treatments) • Are you currently providing treatment for improving torticollis in children?
Transition questions	• Could you describe your clinical experience in treating or rehabilitating torticollis? • What types of rehabilitation treatments have you provided for neonates and infants with torticollis? • In your opinion, what are the most important considerations in the rehabilitation of neonates and infants with torticollis? • What factors should be considered when developing a torticollis rehabilitation system?
Key questions	• The following questions relate to the torticollis rehabilitation system currently under development: • (Instruction manual provided) Do you have any comments on the instruction manual (eg, content structure, length, and clarity)? • (Oral explanation and demonstration provided) Are there any difficulties or areas for improvement when using the clinician web interface of the torticollis rehabilitation system? • Are there any difficulties or areas for improvement when using the child-oriented mobile app of the torticollis rehabilitation system? • Do you have any additional comments regarding the torticollis rehabilitation system?
Closing questions	• Please share any additional considerations that should be taken into account in the development of the torticollis rehabilitation system.

Figure 1. Performance of task scenarios by evaluation participants.





FGI

FGIs are a qualitative research method designed to collect data through in-depth discussions with purposefully selected participants centered on specific topics [16,17]. Information gathered through FGIs is instrumental in identifying potential areas for further exploration or in clarifying themes that may be difficult to capture using other research instruments. The interactive and synergistic nature of focus groups allows participants to refine or elaborate on their thoughts by engaging with the opinions of others. This process often yields richer and more expansive insights than those typically obtained through one-on-one interviews [13].

In this study, FGIs were used to identify usability-related issues, potential use errors, and areas for improvement in the CMT rehabilitation system, thereby deriving preliminary formative design inputs to guide subsequent system development. The question protocol for the FGIs was structured to facilitate meaningful interaction among participants and to elicit comprehensive feedback based on the system's operational scenarios. Following the framework proposed by Krueger and Casey, the interview guide was organized into distinct phases: opening, introductory, transition, key, and concluding questions. The final question set was determined through rigorous deliberation between the evaluation facilitators and the product development team, ensuring that the design promoted robust participant engagement. The key interview questions used for data collection are presented in Table 2.

XSL•FO
RenderX

System Usability Scale

The System Usability Scale (SUS) is an evaluation instrument used to assess the overall usability of a system [18]. The SUS consists of 10 items rated on a 5-point Likert scale (Textbox 1), which are aggregated to produce a composite score ranging from 0 to 100. The scoring procedure involves converting individual item responses to values in the range of 0 to 4 and multiplying the sum of these values by 2.5 to obtain a total score between 0 and 100. Specifically, for odd-numbered items, the score contribution is calculated by subtracting 1 from the participant's response; for even-numbered items, the contribution is determined by subtracting the participant's response from 5 [19].

Textbox 1. System Usability Scale items.

- I think that I would like to use this system frequently
- I found the system unnecessarily complex
- I thought the system was easy to use
- I think that I would need the support of a technical person to be able to use this system
- I found the various functions in this system were well integrated
- I thought there was too much inconsistency in this system
- I would imagine that most people would learn to use this system very quickly
- I found the system very cumbersome to use
- I felt very confident using the system
- I needed to learn a lot of things before I could get going with this system

Ethical Considerations

This study was approved by the Institutional Review Board of the National Rehabilitation Center (NRC-2025-05-036). All participants were provided with a comprehensive explanation regarding the research objectives and methodology prior to the study, and they voluntarily provided written informed consent. Participants were informed that their involvement in the study was entirely voluntary and that they reserved the right to withdraw from the research at any stage without prejudice. To ensure the protection of personal information and the maintenance of confidentiality, all collected data were anonymized and processed in accordance with ethical standards.

Data Analysis

The FGIs were video- and audio-recorded. Two researchers (SH and HN) independently transcribed the recordings verbatim and cross-checked their transcripts against each other. Any discrepancies were reviewed against the original recordings and resolved through discussion, resulting in a finalized transcript for qualitative analysis.

The qualitative data were analyzed using thematic analysis following Braun and Clarke's approach [20]. The analysis involved familiarization with the data, generation of initial codes, identification and organization of recurring patterns into candidate themes, review and refinement of the themes, and definition of the final themes. Two researchers independently reviewed and categorized the finalized transcripts. The coding and categorization results were then cross-checked to identify agreements and discrepancies. Disagreements regarding coding, categorization, or theme interpretation were resolved through discussion and consensus. The final themes reflected recurring usability-related issues and design considerations identified across the FGIs. QualCoder was used to manage and organize the qualitative data throughout the analysis.

The SUS was used as a supplementary quantitative measure of perceived usability. SUS scores were summarized using the mean and SD. Given the small sample size (n=8), the SUS results were interpreted descriptively and considered supplementary to the qualitative findings.

Results

Participant Information

In this study, 8 participants who met the inclusion criteria were recruited. Analysis of the participants' demographic characteristics revealed the following age distribution: 1 in their 20s, 2 in their 30s, 1 in their 40s, and 4 in their 50s. Regarding professional experience, 2 participants had 10 years or less, 2 had 20 years or less, 1 had less than 30 years, and 3 had 30 years or more of clinical experience. Examination of participants' experience with CMT rehabilitation indicated that the 2 rehabilitation medicine physicians had experience with outpatient assistance, while the 2 nurses had experience with treatment observation and/or outpatient assistance. All 4 physical therapists had experience in providing CMT treatment (Table 3).

XSL•FO
RenderX

Table . General characteristics of participants.

ID	Sex	Age (y)	Occupation	Experience in CMT[a] rehabilitation		
				Clinical experience	Clinical experience	Treatment method
P1	Male	29	Rehabilitation physician	Outpatient assistance	4 mo	Outpatient clinic observation and assistance
P2	Female	32	Rehabilitation physician	Outpatient assistance	6 mo	Outpatient clinic observation and assistance
P3	Female	53	Nurse	Treatment observation	30 y	Ward rounds and observation of medical consultation and treatment
P4	Female	53	Nurse	Outpatient assistance	30 y	Ward rounds and observation of medical consultation and treatment
P5	Female	35	Physical therapist	Physical therapy	13 y	Provision of treatment based on assessment of the child's condition
P6	Female	53	Physical therapist	Physical therapy	30 y	Treatment involving stretching of shortened muscles combined with postural education
P7	Female	52	Physical therapist	Physical therapy	28 y	Provision of treatment based on assessment of the child's condition
P8	Male	44	Physical therapist	Physical therapy	19 years	Instrument-assisted physical therapy (stretching and caregiver education)

[a]CMT: Congenital muscular torticollis.

Qualitative Findings: Focus Group Themes

In this study, the data collected through the FGIs were analyzed and categorized according to the thematic analysis framework proposed by Braun and Clarke [20]. The thematic analysis ultimately identified 2 core themes, 10 subthemes, and 21 coded concepts. The inconveniences and areas for improvement regarding the CMT rehabilitation system are summarized in Table 4.

XSL•FO
RenderX

Table . Outcomes of qualitative analysis.

Themes and categories	Concepts
Usability barriers of the CMT[a] rehabilitation system	
Lack of visual readability in the user manual	• Small font size
Insufficient video tutorial structure for understanding the use of mobile apps for children	• Insufficient structure for providing tutorial videos based on treatment prescriptions • Video content within the tutorial video • Insufficient tutorial video structure
Limitations of the provision of treatment information and reliability of mobile apps for children	• AI treatment feedback reliability limitations
Inefficiency of medical staff web pages	• EMR[b] integration not possible • Inefficiency of UI[c] sorting criteria per patient
Limitations of clinical use of web pages for medical professionals	• Inability to verify medical staff's treatment performance data
Improvement requirements of the CMT rehabilitation system	
Strengthening information delivery in the user manual	• Improved visual readability • Provide sufficient information
UI and functional improvements for performing treatment in children's mobile apps.	• Advancing AI treatment feedback • Treatment settings and screen UI configuration • Treatment performance log recording function • Treatment continuous mode
Improving the structure of tutorial videos for children's mobile apps	• Video composition • Playback control • Providing tutorial videos based on treatment prescriptions
Added ability to access treatment performance data on the medical staff web page	• Verification of medical staff treatment performance data
Strengthening the integration and management of clinical information on medical staff web pages.	• EMR integration • Flexibility in treatment prescriptions • Patient clinical information records • Patient-tailored UI structure and visibility improvements

[a]CMT: congenital muscular torticollis.

[b]EMR: electronic medical record.

[c]UI: user interface.

Usability Barriers of the CMT Rehabilitation System

Limited Visual Readability of the User Manual

Participants reported that although the text in the user manual was visually recognizable, the small font size made it difficult to intuitively identify key information. This reduced overall visual readability and interfered with rapid information recognition:

> *I can see it, but the letters are small, so it's not intuitive.* [P3]
>
> *It seems like the font became small because everything was put onto 1 page.* [P4]

Inefficiency of the Clinician Web Interface

Participants indicated that entering a patient registration number in the clinician web interface did not result in electronic medical record (EMR) linkage, making it impossible to verify actual clinical information. This reduced the practical significance of entering patient numbers and highlighted the need for a separate comment section to document medical history or special notes.

Additionally, the dashboard displayed patients based on registration date within a limited time frame, which was perceived as clinically unhelpful. Participants suggested that patient lists should be sortable based on clinically meaningful criteria, such as the most recent prescription date, or should allow clinicians to customize sorting preferences. These limitations were considered to restrict efficient management and use of patient information in clinical contexts:

> *Why do we enter a patient number if it's not linked to the EMR?* [P4]
>
> *I don't think it would be useful if the patient list in the dashboard is sorted only by registration date.* [P2]

Limitations in Clinical Use of the Clinician Web Interface

Participants stated that summary treatment data alone were insufficient to verify inappropriate movements or to provide

XSL•FO
RenderX

specific corrective feedback to caregivers. Without access to treatment videos or objective evidence of incorrect movements, AI feedback was perceived as lacking credibility. Indicators such as "participation" or "How was the treatment?" were considered to have limited clinical significance. Consequently, participants indicated that clinicians would face difficulty identifying areas requiring caregiver education or ensuring the clinical appropriateness of the prescription, potentially reducing clinical applicability:

> *If clinicians could see the exercise videos performed by caregivers or receive more detailed AI feedback, it would be more trustworthy. As it stands, I don't think AI feedback alone is reliable.* [P2]
>
> *If caregivers could provide detailed information about how the child responded, what was difficult, or which posture was challenging, that would be meaningful. But it seems functionally impossible right now.* [P2]

Inadequate Tutorial Video Structure for Supporting the Use of the Mobile App

Participants reported that the tutorial video structure could increase cognitive burden rather than support understanding. Specifically, although only certain exercises were prescribed, caregivers were required to watch all tutorial videos, which was perceived as unnecessary and distracting.

Mandatory viewing upon each login disrupted workflow continuity. The segmented structure of multiple required tutorials was also perceived as interfering with caregivers' ability to understand the rehabilitation process as a cohesive sequence.

Additionally, short video length, fast playback speed, and insufficient guidance regarding movement direction and method were reported as barriers to accurate observation and replication:

> *The tutorial videos are too short, and since only a few exercises are prescribed, I wonder if we really need to watch all of them.* [P2]
>
> *The environment in the video felt distracting, and the child's clothing movement was uncomfortable to watch.* [P3]
>
> *The video speed is too fast, and the voice guidance keeps repeating. It's confusing.* [P5]

Limitations in Treatment Information Provision and Reliability of AI Feedback

Participants questioned whether the AI-generated recognition of therapeutic movements and the accompanying visual feedback adequately reflected the accuracy and clinical relevance of the actual treatment movements. In particular, they perceived that different neck movements were recognized in a similar manner and that the criteria underlying the feedback were not clearly specified. As a result, participants anticipated that caregivers might have difficulty determining the appropriateness of therapeutic movements based solely on AI-generated feedback. These characteristics were perceived as creating a structure in which caregivers could not independently verify whether the movement they were performing aligned with the intended therapeutic action, leading to the perception that participants raised concerns regarding whether the interpretability and reliability of the AI feedback were sufficient to support caregivers' assessment of treatment performance:

> *Since the AI only recognizes the nose and shoulder line, it feels difficult to clearly distinguish whether the movement is left lateral neck flexion or left head rotation. So I'm not sure how meaningful the AI feedback would be for caregivers.* [P2]
>
> *Regarding the green feedback shown on the treatment screen, I don't trust that just because it turns green it necessarily indicates good performance. Also, if the movement is not shown in green, there is no guidance on how to improve, so it's hard to trust it.* [P2]
>
> *Even if the AI feedback appears in green, I question whether that truly means the caregiver is performing the movement correctly.* [P4]

Improvement Requirements of the CMT Rehabilitation System

Strengthening Information Delivery in the User Manual

Participants indicated that key information and safety-related content in the user manual should be conveyed more clearly and intuitively. They suggested increasing font size or using color emphasis to visually distinguish important information, and presenting safety-related content—such as fall risks—using warning colors or graphical illustrations.

For procedures that caregivers may find challenging, such as account registration and entry of child information, participants emphasized the need for explanations accompanied by actual screen examples. In addition, information that may affect user flow—such as rules for rewatching treatment tutorials and the range of prescribable items—should be clearly specified in the manual.

Overall, the user manual was interpreted as requiring enhancement beyond text-centered explanations, with greater emphasis on visual and safety-oriented information grounded in specific interface screens and usage contexts:

> *Important precautions would be more intuitive and visually clear if presented with illustrations. If possible, including pictures or photographs would be helpful.* [P3]
>
> *I think the fall-risk information in the patient user manual is important, so it should be highlighted in pink.* [P5]
>
> *In section 3 of the application user manual, which covers preparation for the child and caregiver, the steps for account registration and entering basic information are explained only in text. However, there are more people than expected who have some level of anxiety about using mobile devices. It would be helpful to include example screenshots showing what those screens look like and which parts need to be checked, including the registration screen and the child information entry screen.* [P8]

XSL•FO
RenderX

> *When reading a manual, people often don't actually read the safety information. So if there were colors that signal 'this must be read' or warning icons such as a fall-risk symbol, it would likely increase awareness and make people look at it again.* [P8]

Addition of Access to Treatment Performance Data on the Clinician Web Interface

Participants indicated that the information provided on the clinician web interface was limited to AI-generated results, making it difficult to fully understand the actual pattern of treatment performance. In particular, beyond the duration classified as "good posture," no detailed information was provided regarding which inappropriate movements had been performed, thereby limiting clinicians' ability to deliver specific corrective feedback. Accordingly, participants expressed the need for detailed treatment performance records or full treatment videos to be transmitted to the clinician web interface, enabling clinicians to directly review patients' actual performance and provide informed feedback. In addition, summary indicators such as participation rates should be accessible on the web interface so that clinicians can reference treatment progress. These findings suggest the necessity of incorporating treatment performance data access functions to support clinicians' clinical judgment and interventions:

> *In addition to AI feedback, it would be helpful to have information about how the remaining time—outside of the 'good posture' period—was spent, so that clinicians can determine whether any movements were performed incorrectly.* [P4]

> *It would be beneficial if the treatment videos performed by caregivers could be transferred to the clinician web interface, allowing clinicians to review them and provide feedback.* [P1]

> *It would also be helpful if the participation records completed by caregivers after each treatment session could be shared with us, so that we can refer to how actively the session was carried out.* [P2]

Strengthening the Integration and Management of Clinical Information on the Clinician Web Interface

Participants indicated that the clinician web interface should be improved to allow integrated management of patient information, prescription settings, and treatment performance summaries within a single screen. They expressed the need for EMR-based medical history and basic patient information to be automatically synchronized through entry of the patient's name or registration number, thereby eliminating redundant data entry. In addition, participants emphasized the need for a comment function that would allow clinicians to record clinical context, such as the direction of torticollis, feeding direction, and relevant medical history. They also requested a function enabling frequently used prescription durations and items to be saved as default or previous settings for reuse. Furthermore, participants suggested that the patient list screen should allow intuitive identification of clinical information—such as torticollis direction and treatment performance summaries (eg, success status or duration)—through color coding or simple visual indicators. Overall, clinicians preferred an integrated management environment that would enable simultaneous review of core clinical information and treatment status on a single screen, rather than navigating across multiple separate pages:

> *It would be helpful if the dashboard patient list clearly indicated which side the child has torticollis. Also, when registering a patient ID, it would be good to have a space to enter information about the child, such as feeding direction.* [P5]

> *Medical staff are very accustomed to looking at patient lists and clinician web interfaces, so even if the screen appears complex, it wouldn't be a problem. We're used to complexity. In fact, it may be better to display as much information as possible. Instead of clicking on a patient's name and navigating to a separate detailed page, it would be more convenient if a panel opened beside the list showing that patient's details, without the current screen disappearing. Since the amount of information isn't excessive, there may be no need to divide it across multiple pages. Even if it includes information such as age, sex, diagnosis, recent condition, and attending physician, although it may look complex, clinicians are accustomed to reviewing such information and may actually prefer it. Having the detailed information displayed alongside the list rather than moving to another page would likely make it more convenient to use.* [P8]

> *It would be ideal if patient information could be directly synchronized with the EMR through the patient's name. If linked properly, there would be no need to manually enter medical history or basic information. Simply entering the patient's name and registration number should automatically retrieve and populate the corresponding information.* [P8]

> *Clinicians usually have specific time durations or prescription items that they frequently use. It would be helpful if they could save their commonly used settings. For example, having a default setting, or recalling the most recent setting and adjusting only the time slightly, would be convenient. It would be beneficial to have functions that allow the previous settings, default settings, or user-saved settings to be easily recalled without having to modify everything manually.* [P8]

UI and Functional Improvements to Support Treatment Performance in the Mobile App

Participants indicated that overall improvements in the user interface and functionality of the caregiver-facing mobile app were necessary to more effectively support treatment performance. In particular, they noted that the current AI recognition method primarily based on the nose and shoulder line was insufficient to adequately distinguish between different therapeutic movements. Accordingly, the need for more sophisticated sensor recognition and enhanced AI-based treatment feedback was emphasized.

XSL•FO
RenderX

Beyond simple visual indicators such as green markings, participants requested reward-based feedback mechanisms such as visual effects that promote a sense of achievement or motivational elements like point accumulation to enhance engagement.

In addition, participants suggested that the treatment settings screen should be improved in terms of font size and information density to allow users to grasp key information at a glance. The need for a practice mode, enabling caregivers to rehearse movements and receive feedback before performing actual treatment sessions, was also raised. Furthermore, participants highlighted the importance of allowing guide videos on the treatment screen to be adjusted or hidden as needed.

Finally, the ability for caregivers to document the treatment process in text form and share this information with clinicians was identified as important for supporting understanding and management of treatment performance:

> *After the AI feedback on the treatment screen is completed, only the time classified as 'good posture' is shown, which may leave caregivers wondering what they are supposed to do next. It would be helpful if more detailed information were provided so they could better understand their performance.* [P3]

> *It might also be useful if caregivers could write something like a treatment log and keep it as a record to share when they meet with the clinician.* [P3]

> *Since certain angles should not be exceeded during exercise therapy, perhaps the AI sensors could be made more precise to capture those limits in greater detail.* [P4]

> *Currently, the AI feedback is limited to red and green lines. It would be good to include motivational features so that when the green indicator appears, there are visual effects like fireworks and mileage accumulation, allowing the child to reach a target point and feel a sense of achievement and enjoyment.* [P8]

> *If caregivers want to start treatment immediately, they should be able to do so, but sometimes they may just want to practice the movement first. It would be good to have a practice mode where they can rehearse the movement and receive feedback, so that after understanding how to perform it properly, they can then begin the actual treatment. Having separate 'Practice' and 'Start' buttons would be helpful.* [P8]

Improvement of the Tutorial Video Structure in the Mobile App

Participants indicated that tutorial videos should be presented in a more intuitive and controllable format to enable caregivers to accurately replicate the therapeutic movements applied to the child. They emphasized the need to improve the filming method (eg, clothing and posture), clarify movement presentation direction (eg, left-right distinction and mirror mode), and include visual aids such as subtitles or arrows highlighting the muscle being stretched.

In addition, participants requested playback control functions, including adjustable playback speed, pause, and repeat options, so that caregivers could sufficiently review and follow specific segments as needed. The ability to automatically select or display guide videos according to the child's affected side (left or right) and prescribed exercises was also identified as an important improvement requirement.

Furthermore, participants suggested that structurally different types of videos—such as therapeutic movement tutorials, device explanations, and diagnostic guidance—should be integrated or reorganized into a simplified framework distinguishing between mandatory and optional tutorials:

> *It would be helpful to include a function that allows the playback speed to be slowed down. In addition, the video could include subtitles or arrows indicating which muscle is being stretched and along which axis. For example, it could show messages such as, 'It has stretched this much,' or 'Stretch only to this point,' to guide users more clearly while they follow along.* [P6]

> *When following the movements on the screen, it would be helpful if the direction of the instructor's explanation and my own direction were clearly specified.* [P7]

> *I don't think the mandatory tutorial needs to be divided into four separate videos. It could simply be combined into one mandatory tutorial that plays continuously. I find myself wondering why it was split into multiple parts. It would be better to have one tutorial for the exercise movements, one for device instructions, and one for diagnosis guidance, and then classify them as either mandatory or optional. Overall, I think the tutorials should be simplified and grouped more clearly.* [P8]

> *The pacing of the video and the narration are both too fast. This makes it difficult for caregivers to understand. There is separate explanatory text beside the video, but instead, it might be better to enlarge the video and include subtitles directly within it. It would also be helpful if arrows or other visual indicators were provided to intuitively guide the current movement status.* [P8]

Quantitative Findings: SUS

The mean SUS score for the overall CMT rehabilitation system, including the clinician web interface and a proxy evaluation of the caregiver mobile app, was 73.44 (SD 18.01)(Table 5). Given the small sample size (n=8) and variability in the scores, the SUS result was interpreted descriptively as a supplementary measure of perceived usability and should be considered alongside the qualitative findings.

XSL•FO
RenderX

Table . System Usability Scale (SUS) results (n=8; unit: point).

Item	Min-max	Mean (SD)	Scaled mean (SD)
Total	—[a]	2.94 (0.37)	73.44 (18.01)
Q1	2.00-5.00	3.63 (1.06)	65.63 (24.80)
Q2	1.00-3.00	2.00 (0.53)	75.00 (12.50)
Q3	2.00-5.00	3.88 (0.83)	71.88 (19.52)
Q4	1.00-4.00	2.00 (1.07)	75.00 (25.00)
Q5	2.00-5.00	3.63 (0.92)	65.63 (21.42)
Q6	1.00-3.00	2.00 (0.53)	75.00 (12.50)
Q7	4.00-5.00	4.38 (0.52)	84.38 (12.10)
Q8	1.00-3.00	1.75 (0.71)	81.25 (16.54)
Q9	3.00-5.00	3.88 (0.64)	71.88 (14.99)
Q10	1.00-4.00	2.25 (0.89)	68.75 (20.73)

[a]Not applicable.

Discussion

Principal Findings

This formative evaluation identified several usability-related issues and preliminary design requirements for the prototype AI-based CMT rehabilitation system from the perspectives of clinical experts. The qualitative findings highlighted 3 main areas: the need to reduce the potential cognitive burden associated with caregiver-delivered rehabilitation, the need to improve the reliability and interpretability of AI-generated feedback, and the need to strengthen coordination between the clinician-facing web interface and caregiver-facing mobile app. The mean SUS score was 73.44 (SD 18.01); however, given the small sample size and variability in scores, this result was interpreted as a supplementary descriptive measure rather than as definitive evidence of overall system usability. Importantly, because caregivers did not participate in this study, findings concerning caregiver experience represent clinicians' proxy perspectives and require confirmation through direct evaluation with actual caregivers.

The qualitative findings identified usability-related concerns that were not adequately represented by the overall SUS score alone. These included the clarity of instructions, delineation of roles between clinicians and caregivers, feedback during exercise performance, and understanding of the system's intended scope of use. Thus, the SUS findings should be regarded as supplementary to the qualitative findings rather than as evidence that the system has achieved an acceptable level of usability. These findings also suggest that usability issues in caregiver-delivered rehabilitation may extend beyond interface operability or functional arrangement to broader human factors considerations associated with the context in which caregivers perform therapeutic exercises with infants.

Human Factors Considerations in Caregiver-Delivered Rehabilitation

The usability issues identified by the clinical experts were less related to basic interface elements, such as button placement, and more closely related to the potential cognitive demands placed on nonprofessional caregivers who are expected to perform therapeutic exercises in home settings. Unlike medical devices operated primarily by trained health care professionals, the caregiver-facing component of the CMT rehabilitation system requires caregivers to perform exercises with infants while simultaneously interpreting instructions and system feedback. From a human factors perspective, this dual role as both system user and provider of home-based exercises may create cognitive and emotional demands that should be considered during system design.

Previous research has indicated that caregivers' disease awareness, level of understanding, and trust in treatment can influence adherence [10]. Consistent with these findings, clinical experts in the present study anticipated that caregivers may experience fear or uncertainty when directly performing therapeutic exercises and may have difficulty simultaneously performing the exercises and interacting with the mobile app. Similarly, Stitt and Operacz [8] reported that caregivers of infants with torticollis may experience psychological stress during home exercise programs because of concerns about causing pain to the child, which may negatively affect adherence. However, the cognitive and emotional burdens identified in the present study were based on clinicians' proxy perspectives and were not directly reported by caregivers.

These findings suggest that the design of CMT rehabilitation software should consider not only basic operability but also the potential cognitive and emotional demands associated with caregiver-delivered rehabilitation. Clear visual materials, comprehensible safety information, and step-by-step guidance may help caregivers understand how exercises should be performed and when corrective action may be necessary. Rather than positioning caregivers as independent clinical decision-makers, the system should support them in performing clinician-prescribed exercises within clearly defined boundaries. From a human factors perspective, these considerations represent preliminary design inputs for reducing potential cognitive burden and use-related errors.

XSL•FO
RenderX

Clear and interpretable feedback may be particularly important when caregivers are expected to determine whether an exercise is being performed appropriately. Subtle aspects of exercise performance, such as movement angle and direction, may be difficult for nonprofessional users to judge without adequate guidance. Clinical experts therefore emphasized the need for visual feedback that communicates exercise performance in an understandable manner. Such feedback may help reduce reliance on subjective judgment and support caregivers in identifying when an exercise requires correction.

In addition, the clinical experts emphasized the importance of clearly defining conditions under which caregivers should discontinue an exercise or seek assistance from a clinician. Home-based rehabilitation requires caregivers to respond appropriately to changes in an infant's condition during exercise. If warning signs, stopping criteria, or safety information are unclear, caregivers may have difficulty determining whether an observed response requires intervention. Accordingly, presenting concrete and observable safety information, including appropriate exercise discontinuation criteria, may represent an important human factors consideration for subsequent system development.

Reliability and Interpretability of AI-Generated Feedback

A particularly important finding was the clinical experts' concern regarding the reliability and interpretability of AI-generated feedback. Participants questioned whether the feedback provided by the prototype could reliably distinguish appropriate from inappropriate exercise performance and whether caregivers would receive sufficient information to understand and correct inappropriate movements. These concerns are particularly relevant in infant rehabilitation, where inaccurate or insufficiently interpretable feedback could contribute to inappropriate exercise performance and potentially increase the risk of harm.

These concerns are consistent with previous research emphasizing the importance of explainability and transparency in medical AI for supporting user understanding and trust [21]. Recent privacy-preserving and interpretable AI approaches for pose-based rehabilitation [22] also provide a useful framework for considering transparent and correction-oriented feedback. Future iterations of the CMT rehabilitation system could therefore explore feedback mechanisms that communicate not only whether a movement is considered appropriate but also why it was classified as such and how an inappropriate movement could be corrected.

Importantly, the present study evaluated clinical experts' perceptions of AI-generated feedback and did not assess the technical accuracy, reliability, or clinical validity of the underlying AI algorithm. Therefore, neither the SUS results nor the qualitative usability findings should be interpreted as evidence of AI performance. Separate technical verification is required to determine whether the algorithm can accurately identify exercise performance and provide reliable feedback. The usability findings presented here should instead be considered preliminary design inputs for improving how AI-generated information is communicated to users.

Integration Within the Clinical Workflow

From the perspective of clinical workflow, participants emphasized the importance of caregiver education before system use and clear delineation of the respective roles of clinicians and caregivers. Clearly defining caregivers' responsibilities and limitations, as well as the circumstances in which clinician monitoring or intervention is required, may help reduce uncertainty and support safer use of the system. In this context, usability should be considered not only in relation to individual interface components but also in relation to how the system fits within the broader rehabilitation workflow.

The clinical experts also requested functions that would allow clinicians to review patient exercise performance, including recorded videos and detailed movement-related data, and suggested potential integration with EMR systems. In CMT management, professional clinical intervention and caregiver education are both important components of rehabilitation [3]. The findings, therefore, suggest that closer coordination between the clinician-facing web interface and caregiver-facing mobile app may facilitate communication and continuity of rehabilitation between clinicians and caregivers. In this sense, the integrated system has the potential to function as a therapeutic bridge between clinical care and caregiver-delivered home rehabilitation.

Methodological Implications

In this study, combining FGIs with the SUS provided complementary qualitative and quantitative information during the formative evaluation. Whereas the SUS provided a descriptive measure of perceived usability, the FGIs enabled more detailed exploration of usability-related concerns involving role perception, perceived responsibility, safety information, AI-generated feedback, and the context of caregiver-delivered rehabilitation. The multidisciplinary composition of the clinical expert groups, including rehabilitation medicine physicians, nurses, and physical therapists, also enabled the system to be considered from different professional perspectives relevant to CMT care.

Given the early developmental stage of the prototype, these findings should be interpreted as preliminary formative design inputs intended to guide subsequent iterations of the system. They should not be interpreted as evidence that usability has been established for actual caregivers or that the system is ready for clinical implementation. Direct evaluation with caregivers will be necessary to determine whether the issues anticipated by clinical experts correspond to caregivers' actual experiences during system use.

Limitations and Future Research

Several limitations should be acknowledged. First, this study included only eight clinical experts, and actual caregivers—the intended users of the caregiver-facing mobile app—did not participate. Therefore, findings concerning caregiver-related usability issues reflect clinicians' proxy perspectives and should not be interpreted as direct evidence of caregivers' experiences or the usability of the caregiver-facing app. Future evaluations should directly involve caregivers to examine their actual

XSL•FO
RenderX

interactions with the system, perceived cognitive and emotional demands, and use-related difficulties.

Second, the small sample size and variability in SUS scores limit the precision and generalizability of the quantitative findings. Accordingly, the SUS results were interpreted descriptively and as supplementary to the qualitative findings rather than as a definitive assessment of overall system usability.

Third, the prototype was evaluated in a controlled setting rather than in actual home environments. Consequently, contextual factors that may influence caregiver interaction with the system during home-based rehabilitation, including environmental distractions, simultaneous caregiving demands, and repeated use over time, were not directly assessed. Future studies should therefore evaluate the system with caregivers under realistic home-use conditions following iterative design refinement.

Finally, this study examined clinical experts' perceptions of AI-generated feedback but did not assess the technical accuracy, reliability, or clinical validity of the underlying AI algorithm. Separate technical verification of the AI feedback algorithm is therefore required before its performance can be established. Following iterative design refinement and technical verification, subsequent usability evaluations involving actual caregivers in representative use environments will be necessary to further assess the safety and usability of the system and to inform progression toward summative evaluation.

Conclusions

This study conducted a formative usability evaluation of an AI-based CMT rehabilitation system from the perspectives of clinical experts. The evaluation identified preliminary usability strengths and areas for improvement, while the mean SUS score of 73.44 (SD 18.01) was interpreted as a supplementary descriptive measure alongside the qualitative findings. The FGIs highlighted several usability-related issues and preliminary design requirements for subsequent system development.

The primary areas for improvement included (1) enhancing the readability of the user manual and visualization of safety information; (2) optimizing the tutorial structure of the caregiver-facing mobile app and improving the reliability and interpretability of AI-generated feedback; and (3) improving clinical workflow efficiency, including potential integration of the clinician web interface with hospital EMR systems. In particular, clinical experts raised concerns regarding whether AI-generated feedback accurately reflected patients' exercise performance and emphasized the need for clinicians to review patient performance videos and detailed movement-related data. These findings highlight the importance of separately verifying the technical accuracy and reliability of the AI feedback algorithm.

The findings suggest that the AI-based CMT rehabilitation system has the potential to support communication and continuity of rehabilitation between clinicians and caregivers. However, because actual caregivers were not included in this study, the findings regarding the caregiver-facing app represent clinicians' proxy perspectives and should be regarded as preliminary formative design inputs rather than evidence of usability for caregivers. Future research should incorporate iterative design refinements based on these findings, evaluate the system with actual caregivers in home environments, and separately verify the technical performance of the AI feedback algorithm.

Acknowledgments

This study was supported by the National Rehabilitation Center Rehabilitation Research & Development Support Program, funded by the Ministry of Health and Welfare (NRCRSP-25TB01, 26TB01).

Generative AI Use

Generative AI tools (eg, ChatGPT or other large language models) were not used in the generation of this manuscript. All content was authored and reviewed by the listed authors.

Funding

The authors declared no financial support was received for this work.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: SH, HN, SJ, HC, HK
Data collection: SH, HN
Formal analysis: SH, HN;
Methodology: SH, HN, SJ
Supervision: HC, HK
Writing – original draft: SH, HN
Writing – review & editing: SH, HN, SJ

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Pradhan P, Norceide DJ, Connolly M, Garayo T, Herman MJ. Congenital muscular torticollis: a current concept review. SurgiColl 2025;3(1). [doi: 10.58616/001c.126718]
2. Sargent B, Kaplan SL, Coulter C, Baker C. Congenital muscular torticollis: bridging the gap between research and clinical practice. Pediatrics 2019 Aug;144(2):e20190582. [doi: 10.1542/peds.2019-0582] [Medline: 31350358]
3. Rodríguez-Huguet M, Rodríguez-Almagro D, Rosety-Rodríguez M, Vinolo-Gil MJ, Ayala-Martínez C, Góngora-Rodríguez J. Effectiveness of the treatment of physiotherapy in the congenital muscular torticollis: a systematic review. Children (Basel) 2023 Dec 20;11(1):8. [doi: 10.3390/children11010008] [Medline: 38275429]
4. Kaplan SL, Coulter C, Sargent B. Physical therapy management of congenital muscular torticollis: a 2018 evidence-based clinical practice guideline from the APTA academy of pediatric physical therapy. Pediatr Phys Ther 2018 Oct;30(4):240-290. [doi: 10.1097/PEP.0000000000000544] [Medline: 30277962]
5. Durguti Z, Murtezani A, Spahiu L, Durguti T, Gara E. Management of infants with congenital muscular torticollis. J Pediatr Neurol 2019 Aug;17(4):138-142. [doi: 10.1055/s-0038-1660829]
6. Kwon DR, Cho SC. Efficacy of intensive inpatient therapy in infants with congenital muscular torticollis involving the entire sternocleidomastoid muscle. Children (Basel) 2023 Jun 20;10(6):1088. [doi: 10.3390/children10061088] [Medline: 37371319]
7. Keklicek H, Uygur F. A randomized controlled study on the efficiency of soft tissue mobilization in babies with congenital muscular torticollis. J Back Musculoskelet Rehabil 2018;31(2):315-321. [doi: 10.3233/BMR-169746] [Medline: 28946532]
8. Stitt A, Operacz R. A qualitative investigation of parent perceptions of home exercises for congenital muscular torticollis. Children (Basel) 2024 Jun 5;11(6):689. [doi: 10.3390/children11060689] [Medline: 38929268]
9. Oledzka MM, Sweeney JK, Evans-Rogers DL, Coulter C, Kaplan SL. Experiences of parents of infants diagnosed with mild or severe grades of congenital muscular torticollis. Pediatr Phys Ther 2020 Oct;32(4):322-329. [doi: 10.1097/PEP.0000000000000738] [Medline: 32991556]
10. Rabino SR, Peretz SR, Kastel-Deutch T, Tirosh E. Factors affecting parental adherence to an intervention program for congenital torticollis. Pediatr Phys Ther 2013;25(3):298-303. [doi: 10.1097/PEP.0b013e318298eb92] [Medline: 23797401]
11. Wiklund ME, Kendler J, Strochlic AY. Usability Testing of Medical Devices, 2nd edition: CRC press; 2015. URL: https://www.taylorfrancis.com/books/9781466595897 [accessed 2026-02-24]
12. Kim J, Kim J, Kim J, Jung M. Development of guideline for heuristic based usability evaluation on SaMD [Article in Korean]. J Biomed Eng Res 2023;44(6):428-442. [doi: 10.9718/JBER.2023.44.6.428]
13. Kitzinger J. Qualitative research. Introducing focus groups. BMJ 1995 Jul 29;311(7000):299-302. [doi: 10.1136/bmj.311.7000.299] [Medline: 7633241]
14. Bae YS, Kim KH, Choi SW, Ko T, Lim JS, Piao M. Satisfaction and usability of an information and communications technology-based system by clinically healthy patients with COVID-19 and medical professionals: cross-sectional survey and focus group interview study. JMIR Form Res 2021 Aug 26;5(8):e26227. [doi: 10.2196/26227] [Medline: 34254946]
15. "IEC 62366-1: 2015+ AMD1: 2020," Medical Devices – Part 1: Application of Usability Engineering to Medical Devices: International Electrotechnical Commission (IEC); 2020, Vol. 1. URL: https://webstore.iec.ch/en/publication/67220 [accessed 2026-11-06]
16. Guest G, Namey E, Taylor J, Eley N, McKenna K. Comparing focus groups and individual interviews: findings from a randomized study. Int J Soc Res Methodol 2017 Nov 2;20(6):693-708. [doi: 10.1080/13645579.2017.1281601]
17. Powell RA, Single HM. Focus groups. Int J Qual Health Care 1996 Oct;8(5):499-504. [doi: 10.1093/intqhc/8.5.499] [Medline: 9117204]
18. Brooke J. SUS - a quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry: CRC Press; 1996. URL: https://www.taylorfrancis.com/chapters/edit/10.1201/9781498710411-35/sus-quick-dirty-usability-scale-john-brooke [accessed 2026-11-06]
19. Lewis JR, Sauro J. Item benchmarks for the System Usability Scale. J Usability Stud 2018;13(3):158-167 [FREE Full text]
20. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
21. Artificial intelligence index report 2025. : AI Index Steering Committee, Institute for Human-Centered AI, Stanford University; 2025 URL: https://hai-production.s3.amazonaws.com/files/hai_ai_index_report_2025.pdf [accessed 2026-11-06]
22. Bačić B, Vasile C, Feng C, Ciucă MG. Towards nation-wide analytical healthcare infrastructures: a privacy-preserving augmented knee rehabilitation case study. In: Mukhopadhyay SC, Senanayake S, Prasad PWC, editors. Innovative Technologies in Intelligent Systems and Industrial Applications: CITISIA 2024 - Volume 1: Springer; 2026, Vol. 1512:378-387. [doi: 10.1007/978-3-032-10898-2_32]

XSL•FO
RenderX

Abbreviations

CMT: congenital muscular torticollis
EMR: electronic medical record
FGI: focus group interview
HEP: home exercise program
SaMD: software as a medical device
SUS: System Usability Scale

Edited by A Bamgboje-Ayodele; submitted 15.May.2026; peer-reviewed by JH Lee, Y Hu; comments to author 29.Sep.2026; revised version received 19.Oct.2026; accepted 29.Oct.2026; published 19.Nov.2026.

Please cite as:
Hong S, Nam H, Jeong S, Choi H, Kweon H
AI-Based Rehabilitation System for Congenital Muscular Torticollis: Formative Usability Evaluation
JMIR Hum Factors 2026;13:e101427
URL: https://humanfactors.jmir.org/2026/1/e101427
doi:10.2196/101427

© Seojin Hong, Heewon Nam, Seongwoo Jeong, Hyun Choi, Hyosun Kweon. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 19.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Geospatially Customizable Culturally Tailored Just-in-Time Adaptive Intervention for Violence-Affected People Living With HIV: Mixed Methods Acceptability, Feasibility, and User Experience Study

Simone J Skeen[1,2], MA, PhD; Stephanie Tokarz[1], MPH; Rayna E Gasik[1], MPH; Ethan A Smith[1], MPH; Katherine P Theall[1,3], MPH, PhD; Gretchen A Clum[1], PhD

[1]Department of Social, Behavioral, and Population Sciences, Celia Scott Weatherhead School of Public Health and Tropical Medicine, Tulane University, 1440 Canal Street, New Orleans, LA, United States

[2]Connolly Alexander Institute for Data Science, Tulane University, New Orleans, LA, United States

[3]Department of Epidemiology, Celia Scott Weatherhead School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, United States

Corresponding Author:
Simone J Skeen, MA, PhD
Department of Social, Behavioral, and Population Sciences, Celia Scott Weatherhead School of Public Health and Tropical Medicine, Tulane University, 1440 Canal Street, New Orleans, LA, United States

Abstract

Background: Posttraumatic stress, along with comorbid mental health challenges and hazardous alcohol use, disproportionately affects people living with HIV. The drivers of these stressors are both intraindividual, rooted in early life adversity and firsthand violence exposures, and contextual, often place-based. Imparting effective coping skills and distinguishing between changeable and unchangeable stressors can improve stress management in the short term, with cascading effects on key HIV continuum of care end points, such as antiretroviral therapy adherence. However, problem- and emotion-based coping skills, delivered via traditional linear in-person group modalities, may falter in the moment. To address this, we adapted the evidence-based Living in the Face of Trauma intervention into an iOS- and Android-native app, featuring daily diary–triggered coping skills recommendations, self-guided Living in the Face of Trauma psychoeducational sessions, and a customizable geofencing function.

Objective: This mixed methods study aimed to examine the acceptability, feasibility, and user experiences of NOLA (New Orleans, Louisiana) Gem, focusing on user interaction costs relative to geographic ecological momentary assessment (GEMA) alone and refining future optimization options.

Methods: People living with HIV (N=32) were recruited across New Orleans and initially randomized 1:1 to treatment (NOLA Gem + GEMA) versus control (GEMA) for 21 days. Feasibility was assessed via enrollment and attrition rates. At the immediate postassessment, participants completed acceptability and usability measures and a brief structured usability interview. Analyses included descriptive statistics, bivariate logit modeling, and synergistic human–large language model deductive coding.

Results: In total, 30 participants (n=22 in the GEMA + NOLA Gem treatment arm) completed the pilot, representing 94% (n=29) of baseline enrollees. Acceptability was very high across the board: 100% (n=30) of users considered NOLA Gem “very” or “somewhat” successful in addressing their daily lives, with 91% (n=28) endorsing increased calm and emotional well-being. In addition, 50% (n=11) of NOLA Gem users were “extremely likely” (Net Promoter Score=10/10) to recommend the app to friends. Eight (27%) GEMA and GEMA + NOLA Gem users reported privacy concerns. Eleven (50%) NOLA Gem users received geofencing alerts; perceptions of this feature’s helpfulness were mixed. No statistically significant sociodemographic or clinical predictors of disparate acceptability or increased privacy concerns were found. No additional frictions were evidenced by GEMA + NOLA Gem versus GEMA users. Qualitatively, NOLA Gem users praised the just-in-time mindfulness, breathing, problem-solving skills delivery, and broader stress control and self-insight benefits. A subset of users pointed out the burdensome length and sometimes inconvenient timing of the daily diaries. Recommendations for next-generation personalization included user-specific dynamic daily diary and geofencing prompt tailoring.

Conclusions: Our small pilot study demonstrated high NOLA Gem acceptability and feasibility, as well as a rich and beneficial user experience among people living with HIV, with clear and actionable opportunities for improvement.

Trial Registration: ClinicalTrials.gov NCT05784714; https://clinicaltrials.gov/study/NCT05784714

International Registered Report Identifier (IRRID): PRR2-10.2196/47151

XSL•FO
RenderX

(JMIR Hum Factors 2026;13:e80705) doi:10.2196/80705

KEYWORDS

mobile health; mHealth; HIV; traumatic stress; posttraumatic growth; coping; geospatial; just-in-time adaptive intervention; JITAI

Introduction

Background

Posttraumatic stress is endured by an estimated 28% of people with HIV, globally [1]. Nearly two-thirds (64%) report symptoms consistent with diagnosable posttraumatic stress disorder (PTSD) [2]. The trauma burden can reach as high as 97.1%, with an average of 4.2 out of 10 recognized adverse childhood experiences (ACEs) on the ACEs Questionnaire, among women with HIV in clinical samples. Lifetime trauma has been linked, in a dose-response manner, to greater likelihoods of self-reported depressive and anxiety symptoms, as well as alcohol and illicit drug use [3]. Indeed, the origins of traumatic stress among many people with HIV in the United States are often attributable to multiple sources, including compounded early life adversities, economic hardship, and, in particular, firsthand violence exposures [4-6]. Brown et al [5], in South Carolina, reported 3 overarching components of traumatic life experiences among people with HIV: firsthand extreme violence and secondhand homicide and suicide exposures, physical and sexual assault, and accidents and natural disasters. Among people with HIV, these experiences can encompass childhood sexual abuse [4], racial trauma, incarceration [5], homelessness, rape, intimate partner and intrafamilial violence, and comorbid life-threatening illness [6].

Brezig et al [7], pointing to the disproportionate prevalence of violent (vs nonviolent) trauma among people with HIV, developed a syndemic theory of trauma alongside the acquisition, forward transmission, and later-life outcomes associated with HIV survivorship. Trauma, per the American Psychological Association, can manifest in heightened fear, vulnerability, rumination, disorientation, and dissociative tendencies [2]. Among some people with HIV, these symptoms are linked to anger, isolation, distrust [6,7], comorbid anxiety, depression, suicidal ideation [8,9], and hazardous drug and alcohol use [2,7].

Through a behavioral medicine lens, posttraumatic stress represents a crucial, robustly demonstrated, within-person determinant of antiretroviral therapy (ART) adherence [2-4]. Among women with HIV, acute trauma is linked to 4 times the odds of lapsed ART [7]. Within the Brown et al [5] sample, people with HIV who experienced past extreme violence–related or death-related trauma were 63% less likely (adjusted odds ratio 0.37, 95% CI 0.15-0.95) to demonstrate adherence than non–trauma-exposed participants; people with HIV exposed to any trauma were 58% less likely (adjusted odds ratio 0.42, 95% CI 0.21-0.86). Among Black gay and bisexual men in Baltimore, Maryland, and Jackson, Mississippi, 24% reported past-year physical, sexual, or psychological violence exposure, and recent and lifetime violence exposure was linked to ART nonadherence (adjusted prevalence risk ratio 1.59, 95% CI 1.25-2.01) [10]. Qualitatively similar effects, entangling lifetime trauma, mental health challenges, and substance use or misuse as drivers of poor ART adherence are consistently demonstrated across independent samples of people with HIV and US geographies [11-13], including New Orleans [14].

Narrowly individualistic perspectives on trauma among people with HIV may overlook important neighborhood-based sources of stress [15,16] and stressors that can change across time and space even within a single day [17,18]. In New Orleans, neighborhood-level alcohol outlet density has been linked to alcohol misuse among men with HIV and an increased likelihood of endorsing depressive symptoms among women with HIV [19]. GPS and daily diary data (from a very small feasibility sample; n=11) demonstrate that people with HIV spend an average of just 19% of their awake time within their own residential Census tracts, exposed to neighborhood crime rates ranging from 1 to 28 per 1000 population and alcohol outlet density rates of 0 to 22 per 1000 as they traversed, on average, 23 Census tracts [20]. Emerging evidence, typically captured via geographic ecological momentary assessment, shows that these dynamic place-based factors can affect mood and rumination among people with HIV in New Orleans [9,21].

Effective coping can interrupt traumatic stress and ART adherence lapses [22-25]. Specifically, *adaptive* coping, grounded in the coping framework of Lazarus and Folkman [26], distinguishes between “changeable” and “unchangeable” stressors. The former are approached via problem-focused strategies, such as structured problem-solving and effective communication skills, along with additional cognitive behavioral therapy–oriented strategies, while the latter are addressed via emotion-focused strategies, including cognitive restructuring, mindfulness meditation, and confronting irrational beliefs such as “all-or-nothing” thinking [23]. The Living in the Face of Trauma (LIFT) intervention by Sikkema et al [25], a Centers for Disease Control and Prevention “best evidence” intervention for improving coping with early life trauma stressors and HIV, as well as reducing trauma symptom burden, exemplifies this approach. Across a series of pilot studies [25] and fully powered randomized controlled trials [27], as well as an adaptation for women with HIV in Cape Town, South Africa [28], LIFT has demonstrated efficacy in reducing intrusive symptom burden, forward transmission risk behaviors, and avoidance and arousal PTSD symptoms, while improving ART adherence.

A 15-session, linear, in-person group intervention, LIFT, by definition, does not respond to momentary ecological stressors that people with HIV may encounter throughout the day. Just-in-time adaptive interventions (JITAIs), delivered via smartphone, and often reliant on ecological momentary assessment (EMA) inputs [29,30], can prompt users with brief psychoeducational and coping skills refreshers whenever necessary [24,31]. Similarly, geofencing interventions deliver intervention content triggered by spatial inputs, wherever necessary [32]. JITAIs have demonstrated preliminary efficacy

XSL•FO
RenderX

in reducing ruminative episodes via cognitive behavioral therapy skills prompts and alcohol misuse among homeless adults in pilot studies [33,34], as well as robust efficacy profiles among adults transitioning from inpatient alcohol use disorder treatment [35], alongside a wide array of mental health endpoints [36,37]. Reported acceptability, feasibility, and engagement, including among sexually minoritized men living with HIV [38], are generally high [34,39-41], with many opportunities for incorporating prospective end user insights [42].

Study Objectives

To our knowledge, no prior studies have examined the acceptability, feasibility, and broader user experience of a geographic ecological momentary assessment (GEMA)–leveraging JITAI that incorporates trauma-informed principles and personalizable geofencing to meet the needs of people with HIV. We report the findings of the pilot of NOLA (New Orleans, Louisiana) Gem, a JITAI adaptation of LIFT, trialed among a sample of people with HIV enrolled across Greater New Orleans. The NOLA Gem app was adapted from the *Living in the Face of Trauma: An Intervention for Coping with HIV and Trauma* manual, in consultation with Sikkema [27,28]. The NOLA Gem study covered Aims I to III. The findings of Aim I are reported in detail in Gasik et al [21]. The results of Aim II, the NOLA Gem design synthesizing the insights of Aims I and II (including a detailed user flow diagram and user story maps), and the pilot trial protocol are described in Kabir et al. Here, we report the results of Aim III.

Methods

The NOLA Gem Aim III study was a randomized controlled trial enrolling adults with HIV across Greater New Orleans, spanning from 2023 to 2024.

GEMA: Control Arm

The NOLA Gem app was designed with a control and a treatment version, which were masked to the user. The assignment was dependent on how the user was created in the admin panel by research staff. Both versions had GPS tracking; latitudinal and longitudinal data points were recorded at 1-minute intervals if the device moved more than 20 m for the duration of the 21-day pilot trial [43]. GEMA alone, combining GPS tracking and daily diary delivery via text message at 9:00 AM (due by 12:00 PM) and 6:30 PM (due by 9:30 PM) intervals, but lacking any interventional component, constituted the control condition.

NOLA Gem + GEMA: Treatment Arm

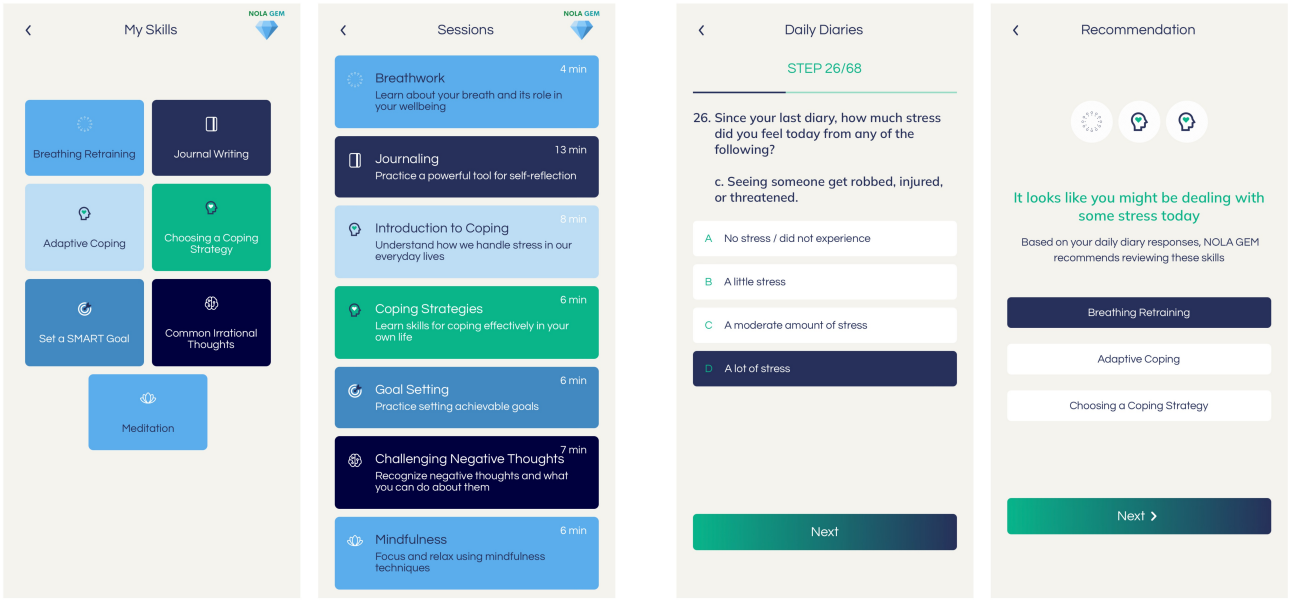
The treatment arm of the NOLA Gem app leveraged the GEMA inputs as tailoring variables, calling on fixed decision rules determined in Aim II, to recommend from among 7 "just-in-time" adaptive coping skills ("Set a SMART Goal," "Common Irrational Thoughts," etc) mapped to predetermined proximal outcomes [39,44]. Each skill was adapted from a LIFT worksheet, culturally tailored to invoke drivers of resilience identified by New Orleanian people with HIV in Aim II. Endorsing more than "a little stress" on daily diary 37f ("Since your last diary, how much stress did you feel today from any of the following? *Had an argument with friend, family member, or romantic partner*") elicited deep-linked recommendations for the "Adaptive Coping" and "Common Irrational Thoughts" in-app brief skills modules, to give just one example. In addition, NOLA Gem users were provided 7 self-directed sessions (estimated 4‐13 minutes in duration), grounded in the traditional, linear delivery of LIFT. These sessions offered psychoeducation on maladaptive or adaptive coping strategies, goal setting, and mindfulness meditation, among other topics. As no generalizable across-case geospatial stressors were evident in Aim I [21], the NOLA Gem treatment arm included a user-determined, essentially personalizable, geofencing feature. Users were asked at baseline to identify 3 locations that they found to be particularly stressful in their lives. Latitude and longitudinal coordinates for these locations were entered into the admin panel with a 150-m geofence; when users entered this geofence, they received a text message asking if they would like to practice their skills, with a link to the skills section of the app. The NOLA Gem JITAI features, built atop and including all aspects of GEMA, constituted the treatment condition for the duration of the 21-day trial period.

Illustrative screenshots of NOLA Gem are shown in Figure 1.

In both arms, the study app was Health Insurance Portability and Accountability Act–compliant and transmitted all data to a Health Insurance Portability and Accountability Act–compliant, Crowdbotics-hosted administrator dashboard.

XSL•FO
RenderX

Figure 1. The New Orleans, Louisiana (NOLA) Gem + GEMA user interface: panel (A) displays the "My Skills" and "Sessions" menus; panel (B) displays a sample daily diary and the resultant skills recommendation prompt. GEMA: geographic ecological momentary assessment.



(A) (B)

Study Participants and Procedures

Participants for Aims I and II were convenience-sampled from the longitudinal New Orleans Alcohol Use in HIV cohort, which, since 2017, has followed 365 people with HIV and hazardous alcohol use histories in outpatient care across Greater New Orleans [45]. For Aim III, this study enrolled participants from this cohort while diversifying recruitment to include a range (n=12) of local HIV outpatient clinics and community organizations serving people with HIV, which were provided with study recruitment flyers.

Participants were eligible if they (1) were 18 years or older; (2) maintained a current ART prescription, verified in person or via a photograph uploaded to a secure Qualtrics link; (3) owned a smartphone; and (4) indicated experiencing at least 1 incident of violence on the screener.

Study enrollment spanned from October 2023 to August 2024. Initially, eligible, consented participants were randomly assigned 1:1 to treatment (NOLA Gem + GEMA) or control (GEMA) arms. Once total enrollment reached 16, the remaining participants were assigned to the treatment arm. This step was taken to address compounding enrollment challenges posed by the ongoing COVID-19 pandemic, specifically the Delta and Omicron variants (during the former wave, Louisiana recorded the highest number of new cases relative to the total population in the United States [46]), and Hurricane Ida, which caused widespread evacuations and a prolonged power outage across Orleans Parish [47]. By favoring treatment-arm enrollment, we aimed to achieve a sufficiently sized pilot sample to gather preliminary mixed methods evidence of NOLA Gem's acceptability, feasibility, and usability, despite these unforeseen obstacles, while maximizing the potential human benefits of research involvement for people with HIV in New Orleans.

The study arm was randomly assigned following eligibility verification, prior to the baseline appointment, to streamline onboarding. At the baseline appointment, participants were read an informed consent form aloud, and verbal consent was recorded in Research Electronic Data Capture (REDCap). The NOLA Gem app was then downloaded from the App Store (iOS) or Google Play (Android) onto participants' phones using the institutional Wi-Fi. The research assistant enabled notifications and location tracking on participants' devices, as well as location services for the NOLA Gem app. Onboarding to both control and treatment conditions was supported by illustrated guides. The baseline assessment was research assistant–administered, and responses were recorded in REDCap.

Measures

Baseline Assessment

Demographics were collected at baseline: name; date of birth (to verify self-reported age); address, including ZIP code; assigned sex; gender (male, female, nonbinary, and a "prefer to self-describe" open-ended response option); race, using US Census categories along with a self-describe option; Latine heritage, using a single-item indicator; and sexual orientation (heterosexual or straight, gay or lesbian, bisexual, or self-describe). Relationship status, educational attainment, employment, and housing details were collected using measures carried over from the NOAH parent study [48]. Participants were administered a 4-item Digital Confidence Score measure (eg, "I use a digital device frequently") adapted from Benson [49].

At baseline, a range of clinical variables were captured. Measures used in these analyses, which exploratorily probed potential drivers of, or disparities in, NOLA Gem acceptability, included the 3-item Alcohol Use Disorder Identification Test–Consumption, administered to detect potentially hazardous alcohol use patterns [50]; the Addiction Severity Index "Alcohol/Drugs" section, which collected past 30-day frequency

XSL•FO
RenderX

of alcohol, cannabis, cocaine or crack, opioid, other, and combined substance use or misuse [48]. The PTSD Checklist for *DSM-5* assessed subthreshold and PTSD symptoms that may benefit from treatment [51]. The Posttraumatic Growth Inventory evaluated positive within-person aftereffects of traumatic stress, such as deepened spirituality and enhanced appreciation of life [52]. The "lost friends due to serostatus," "disclosure risk," "interpersonal discomfort," and "hurtful attitudes toward PWH" items of the HIV Stigma Scale assessed those domains of internalized HIV stigma [53]. The 10-item Kaiser Permanente ACEs Questionnaire collected instances of early-life adversity, neglect, and abuse [54].

Quantitative Postintervention Assessment

At offboarding, both GEMA and GEMA + NOLA Gem users were administered a range of acceptability and usability measures. These end points (preregistered in Skeen et al,[55] crucial for the future trialability and scalability of NOLA Gem, are the focus of this study.

Acceptability was assessed using a series of study-specific items designed to evaluate ease of use, perceived success in enhancing well-being, and privacy concerns. For example, the first item asked, "How often were you able to open and use the NOLA GEM app without confusion, frustration, or asking for help? [Very often, somewhat often, not very often]." Ten items, using a 4-point Likert scale, assessed the NOLA Gem app's multifactorial appeal (eg, likeability, visual appeal), and its perceived helpfulness in attaining proximal outcomes; for example: "I found NOLA Gem to be helpful in changing habits [Agree, somewhat agree, somewhat disagree, disagree]." In addition, to test whether participants were willing to expend their own social capital on behalf of NOLA Gem, an adapted Net Promoter Score (NPS) asked participants about their likelihood of recommending NOLA Gem to a friend (1=not at all likely, 10=extremely likely) [56].

Feasibility was measured via attrition rates postenrollment, daily diary completion rates, and paradata descriptives.

Usability was assessed with the 8-item User Experience Questionnaire (UEQ)–*Short*, an abbreviated version of the UEQ that encompasses both *pragmatic* (determinative of user goals) and *hedonic* (determinative of user pleasure) usability domains. The UEQ uses a 7-point scale with item-specific anchors to deter central tendency biases [57]. For accessibility's sake, we simplified these anchors wherever possible. For example, we replaced "obstructive" with "unhelpful" in item 1: "Rate the app from unhelpful to supportive."

Qualitative Postintervention Assessment

Finally, participants were asked a series of open-ended, atheoretical questions: "What did you not like about the app?" "What were your favorite parts of the app?" "What aspects of the app do you think should be changed?" and "What aspects of the app do you think should be kept?" Responses were digitally recorded and transcribed verbatim into dedicated REDCap fields.

Ethical Considerations

All procedures were reviewed and approved by the Socio-Behavioral Institutional Review Board of the Biomedical and Social Behavioral Institutional Review Boards at Tulane University (reference number 2022‐1861). All participants provided verbal informed consent. All recorded data were private, confidential, and anonymous. Participants were compensated with a US $30 Amazon gift card for completing their baseline appointment, US $20 via ClinCard for their immediate postassessment and 30-day follow-up visit (each), US $2 via ClinCard for each daily diary completed, and a US $10 Walmart gift card for the 90-day follow-up. Starting in March 2024, after switching to treatment-arm only enrollment, participants were also compensated with US $5 via ClinCard per session completed, with a US $15 bonus if all 7 were completed.

Analytic Approach

Overview

Background demographic and quantitative acceptability, feasibility, and usability outcomes are reported descriptively. Although the UEQ "Data Analysis Tool" provides a benchmark usability dataset, these values are not validated for the 8-item UEQ short form [58]. Disparate acceptability was assessed, preliminarily, by recoding dichotomous gender (1=male), race (1=Black), Latine identity (yes=1), Test–Consumption and Addiction Severity Index item-wise (1=any lifetime alcohol or past 30-day cannabis use, cocaine or crack, or opioid use), dichotomizing the PTSD Checklist for *DSM-5* at the recommended >33 cutoff, and median-splitting the Post-Traumatic Growth Inventory and *Digital Confidence Score* sum, for entry into bivariate logit regressions as factor variables [59]. Wherever sufficient variation permitted, acceptability outcomes were dichotomized as well (eg, 1 = "Agree" or "Somewhat agree" to "I found NOLA Gem to be useful"). These transformations were undertaken to account for low variability and a very small sample size.

Particular attention was paid to possible differential drivers of GEMA privacy concerns (recoded such that 1="somewhat" and "very concerned" [60]). Potential privacy concern predictors included sexual orientation (1=sexual minority), and median-split HIV Stigma Scale and *ACEs* scores.

The NPS was computed in the industry-standard manner: scores ≥9 were categorized as *promoters*, scores 7 to 8 as *passives*, and scores ≤6 as *detractors*. The NPS metric was obtained by subtracting the percentage of detractors from the percentage of promoters [56].

While direct control-treatment differentials in acceptability (etc) were not preregistered, we tested for potential GEMA (control) versus NOLA Gem + GEMA (treatment) additive interaction costs exploratorily, via 2-tailed independent-samples *t* test and visual inspection, where warranted [61]. Quantitative analyses were run in Stata/SE 18.0 [62]. Data visualizations were created using Naqvi spider package for Stata [63] and Matplotlib in Python 3.13 [64].

XSL•FO
RenderX

Qualitative responses were analyzed using a rapid, atheoretical, human–large language model (LLM) synergistic approach. First, the lead author (SJS), after conducting immersive close readings of the treatment group responses, developed categories inductively, and formalized them within a codebook. The qualitative data frame was transformed by inserting category-specific columns to (1) dummy code each response for every respective category, as warranted, and (2) extract the most salient segment of text: the "rationale" for the coding application [65]. Next, informed by the work of Dunivin et al [66] and the CHALET framework of Meng et al [67], the "Collaborative Human-LLM Analysis for Empowering Conceptualization in Qualitative Research" technique was used. Category-by-category deductive coding prompts were formulated. The prompts were "few-shot," providing human-synthesized examples, and leveraged contemporary LLMs' chain-of-thought capabilities by requiring a 2-sentence justification for each coding decision [66,67]. To maintain data privacy, the Ollama framework was used to run a local on-device 12 billion–parameter (12B) Gemma 3, a lightweight open-source extension of Google's proprietary Gemini foundation model [68,69]. Matching the initial human approach, the 12B Gemma 3 was tasked with (3) dummy coding expressions of each category and (4) providing coding decision justifications (inclusive of the LLM's own "rationales") in dedicated columns. Initial human-LLM intercoder reliability was computed using Cohen κ. For any category that demonstrated $\kappa < 0.70$, a second human consensus coder (ST) resolved coding application discrepancies. All qualitative analyses were conducted in Python 3.13.

Open Code

Stata .do and Python .ipynb replication scripts, including deductive coding prompts, are hosted in a dedicated GitHub repository [70].

Results

Overview

The overall pilot sample was generally in their mid- to late-50s, with slight over-representation by men (n=17, >55%), a plurality identifying as minoritized sexualities (n=10, <45.45% heterosexual/straight), and roughly 3 quarters Black (n=22, 73.33%). A very large majority (n=30, 93.33%) reported <US $50,000 past-year household income. Digital confidence scores were high, in the "strongly agree"–"agree" range on average.

The pilot sample's demographic details, overall and by study arm, are shown in Table 1.

XSL•FO
RenderX

Table . Baseline sociodemographic attributes and digital confidence: NOLA[a] Gem pilot.

Attribute	Descriptives[b]		
	Overall (N=30)	Treatment: GEMA[c] + NOLA Gem (n=22)	Control: GEMA (n=8)
Age (y), mean (SD; range)	54.2 (11.9; 28 - 74)	51.8 (12.6; 28 - 74)	60.9 (6.06; 50 - 72)
Gender, n (%)[d]			
Male	17 (56.67)	15 (68.18)	2 (25.00)
Female	13 (43.33)	7 (31.82)	6 (75.00)
Nonbinary	0 (0)	0 (0)	0 (0)
Preferred to self-describe[e]	0 (0)	0 (0)	0 (0)
Sexual orientation, n (%)[d]			
Heterosexual or straight	13 (43.33)	10 (45.45)	3 (37.50)
Gay or lesbian	11 (36.67)	9 (40.91)	2 (25.00)
Bisexual	3 (10)	1 (4.55)	2 (25.00)
Preferred to self-describe[e]	3 (10)	2 (9.09)	1 (12.50)
Ethno-racial identification, US Census categories, n (%)[d]			
African-American/Black	22 (73.33)	15 (68.18)	7 (87.50)
Asian	0 (0)	0 (0)	0 (0)
Multiracial	3 (10)	2 (9.09)	1 (12.50)
Native Hawaiian or Pacific Islander	0 (0)	0 (0)	0 (0)
White	4 (13.33)	4 (18.18)	0 (0)
dta[f]	1 (3.33)	1 (4.55)	0 (0)
Hispanic or Latino	2 (6.67)	2 (9.09)	0 (0)
Household income in the past year (US $), n (%)[d]			
<10,000	6 (20)	4 (18.18)	2 (25.00)
10,000-19,999	13 (43.33)	8 (36.36)	5 (62.50)
20,000-34,999	3 (10)	2 (9.09)	1 (12.50)
35,000-49,999	6 (20)	6 (27.27)	0 (0)
50,000-74,999	2 (6.67)	2 (9.09)	0 (0)
75,000-99,999	0 (0)	0 (0)	0 (0)
>100,000	0 (0)	0 (0)	0 (0)
Housing, n (%)[d]			
Stand-alone single-family home	14 (46.67)	10 (45.45)	4 (50)
Apartment, duplex, or multifamily home	13 (43.33)	10 (45.45)	3 (37.50)
Group home or HIV-tailored group facility	1 (3.33)	1 (4.55)	0 (0)
Homeless or shelter	1 (3.33)	0 (0)	1 (12.50)
Nursing home	0 (0)	0 (0)	0 (0)
Preferred to self-describe[e]	1 (3.33)	0 (0)	1 (4.55)
Smartphone OS[g], n (%)[d]			
Android	18 (60)	12 (54.55)	6 (75.00)

XSL•FO
RenderX

Attribute	Descriptives[b]		
	Overall (N=30)	Treatment: GEMA[c] + NOLA Gem (n=22)	Control: GEMA (n=8)
iOS	12 (40)	10 (45.45)	2 (25.00)
Digital confidence scores, mean (SD)[h]			
"I use a digital device frequently"	3.53 (0.73)	3.59 (0.73)	3.38 (0.74)
"Most of my friends use digital devices"	3.63 (0.49)	3.73 (0.46)	3.38 (0.52)
"I can usually get help if I am stuck"	3.37 (0.72)	3.41 (0.80)	3.25 (0.46)
"I feel confident using most digital devices"	3.17 (0.83)	3.27 (0.83)	2.88 (0.83)

[a]NOLA: New Orleans, Louisiana.

[b]Percentages may not total 100 due to rounding.

[c]GEMA: geographic ecological momentary assessment.

[d]Column percentages by attribute are shown.

[e]Details withheld to preserve participant anonymity.

[f]dta: decline to answer.

[g]OS: operating system.

[h]Adapted from Benson [49]; scored on a 4-point scale: 4=strongly agree, 3=agree, 2=neutral, 1=disagree.

Acceptability

Acceptability, similarly, was generally very high across measures in the NOLA Gem + GEMA (treatment) group (n=22). Overwhelmingly, 16 (72.73%) participants reported the easy use of the app, with perceived helpfulness more equivocal, but a very small minority (n=2, 0% - 9.09%) perceiving the app as totally ineffectual. Bivariate logit regressions were null (male, *P*=.40; Black, *P*=.75; Latine, *P*=.26; lifetime alcohol use, *P*=.70; past 30-day drug use, *P*=.29; positive PTSD screen, *P*=.07; >median posttraumatic growth, *P*=.15; >median digital confidence, *P*=.22), suggesting no detectable social patterning to NOLA Gem acceptability among the pilot sample.

Nearly a third (n=7, 31.82%) reported GPS-related privacy concerns, though no statistically significant associations with privacy-related concerns were evidenced (male, *P*=.23; minoritized sexuality, *P*=.70; >median HIV stigma, *P*=.70; >median adverse childhood experiences, *P*=.30; as GEMA was universal, these models used the complete sample, ie, n=30). Consistent with the qualitative findings, nearly a quarter (n=7, 23.33%) found the daily diaries difficult to incorporate into their daily lives. Within the treatment group, 11 activated a custom geofence; though enthusiasm for this feature was mixed, no participants found it "not at all" helpful.

The app usage and satisfaction measures are summarized in Table 2.

Additional measures of multifactorial appeal are plotted in Figure 2.

Measures of NOLA Gem's perceived helpfulness are plotted in Figures 3 and 4.

XSL•FO
RenderX

Table . App usage and satisfaction: geographic ecological momentary assessment (GEMA) + NOLA[a] Gem (treatment group, n=22).

Measure	Response options, n (%)[b]	Mean (SD; range)
How often were you able to open and use the NOLA Gem app without confusion, frustration, or asking for help?	• Very often: 16 (72.73) • Somewhat often: 3 (13.64) • Not very often: 3 (13.64)	2.59 (0.73; 1 - 3)
How successful was the app in addressing your daily life?	• Very successful: 11 (50) • Somewhat successful: 11 (50) • Not very successful: 0 (0)	2.50 (0.51; 1 - 3)
How much has the app increased well-being, sense of calm, or positive feelings?	• A lot: 11 (50) • A little: 9 (40.91) • Not at all: 2 (9.09)	2.41 (0.67; 1 - 3)
Did you feel concerned about your privacy while participating in the GPS tracking part of this study?[c]	• Not concerned: 15 (68.18) • Somewhat concerned: 5 (22.73) • Very concerned: 2 (9.09)	2.67 (0.61; 1 - 3)
Did you receive text messages encouraging or suggesting skills or sessions to try in the NOLA Gem app?	• Yes: 21 (70)	—[d]
If “yes” (n=21): How relevant were these to your daily life or stress?	• Very: 2 (9.52) • Somewhat: 10 (47.62) • A little: 2 (9.52) • Not at all: 1 (4.76) • Did not use: 6 (28.57)	3.48 (1.17; 1 - 4)
Did you receive suggested skills at the end of your daily diaries to try in the NOLA Gem app?	• Yes: 21 (70)	—
If “yes” (n=21): How relevant were these to your daily life or stress?	• Very: 7 (33.33) • Somewhat: 10 (47.62) • A little: 4 (19.05) • Not at all: 0 (0) • Did not use: 0 (0)	3.14 (0.73; 1 - 4)
Geofencing alerts were text messages that appear when you neared areas you identified as stressful at the start of the study. Did you receive any geofencing alert text messages during the study period?	• Yes: 11 (50)	—
If “yes” (n=11): How helpful were these recommendations in the moment?	• Very: 5 (45.45) • Somewhat: 3 (27.27) • A little: 2 (18.18) • Not at all: 0 (0) • Did not use: 1 (9.09)	3.45 (0.93; 1 - 4)
Was the timing of the daily diaries easy for you to incorporate into your daily life?	• No: 7 (23.33)	—
(If “no”) Which timeframe was difficult for you to answer within?	• 9:00 AM-12:00 PM: 4 (57.14) • 6:00 PM-9:00 PM: 0 (0) • Both timeframes: 3 (42.86)	—

[a]NOLA: New Orleans, Louisiana.

[b]Percentages may not add to 100 due to rounding.

[c]Reverse coded.

[d]Not applicable.

XSL•FO
RenderX

Figure 2. NOLA Gem acceptability: domain general. Diamond markers denote mean ratings, and error bars indicate SDs. NOLA: New Orleans, Louisiana.

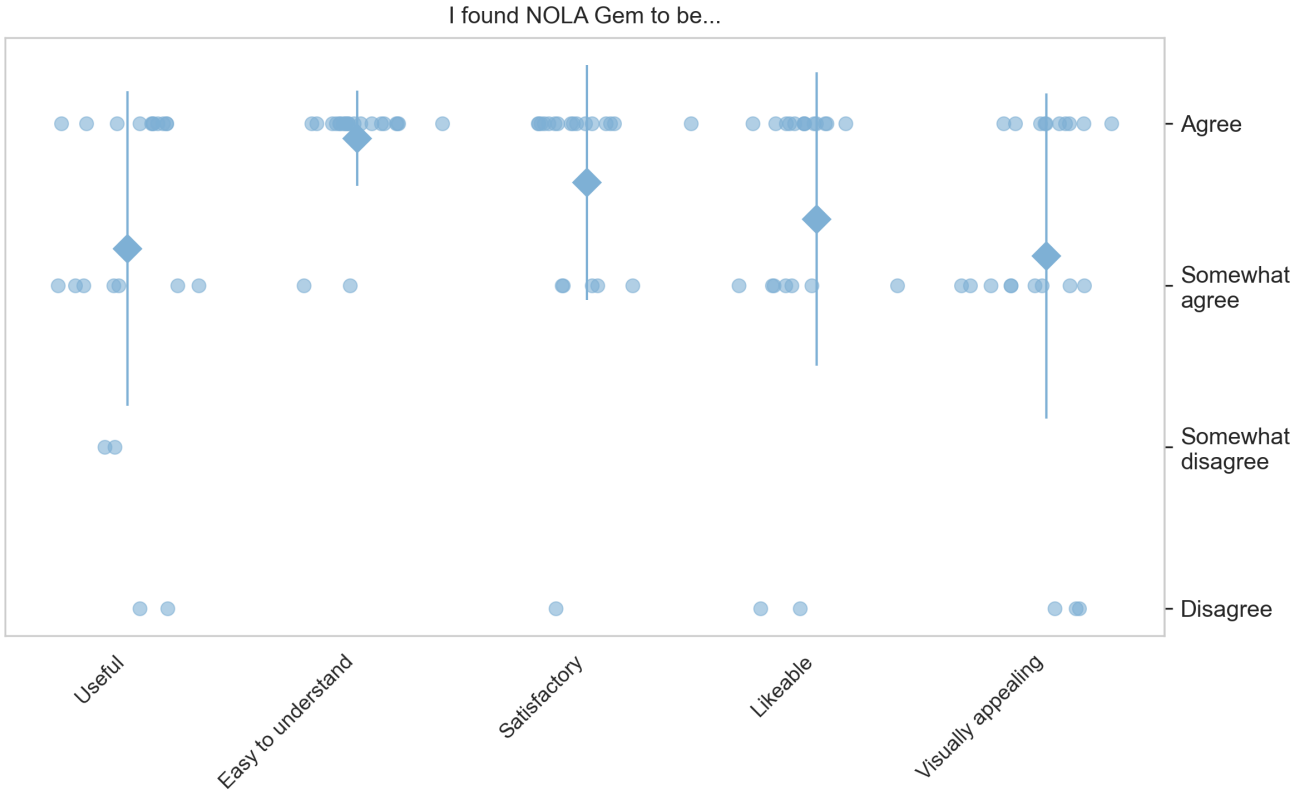
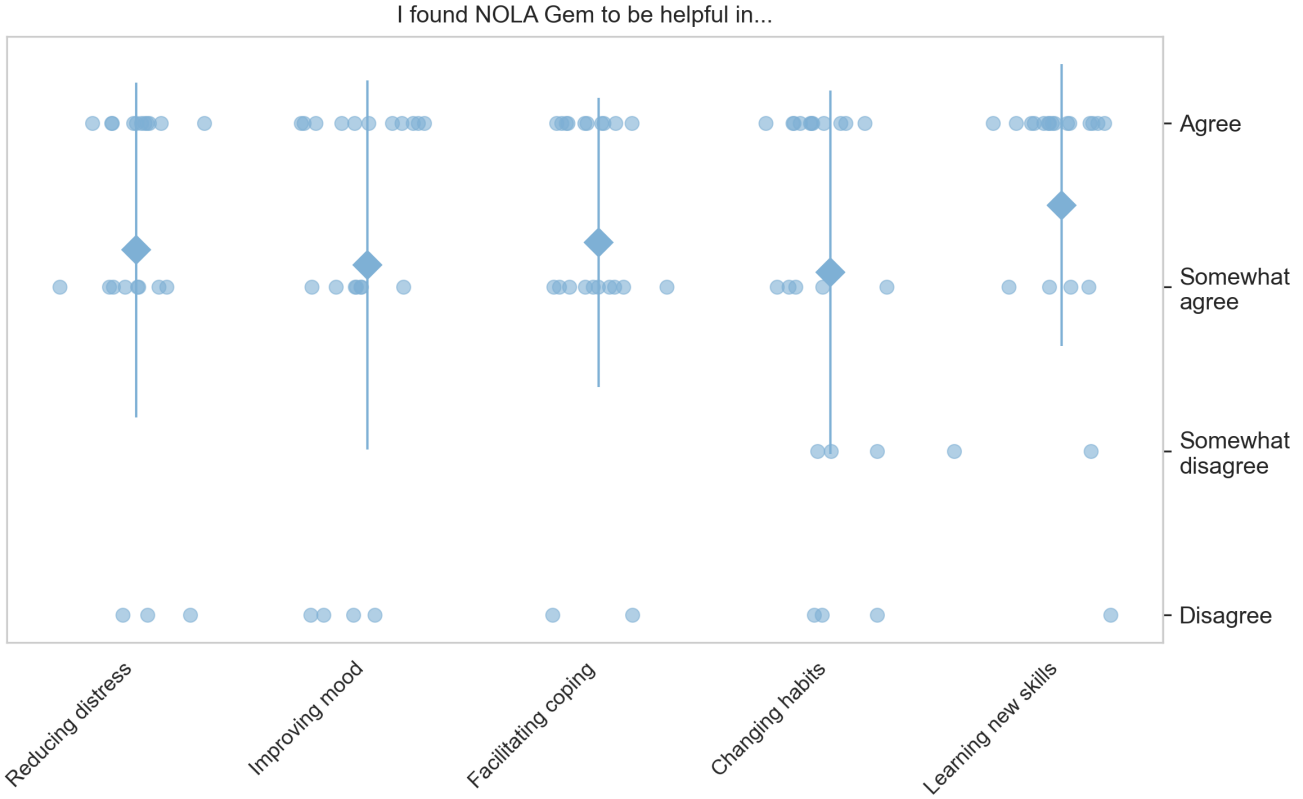
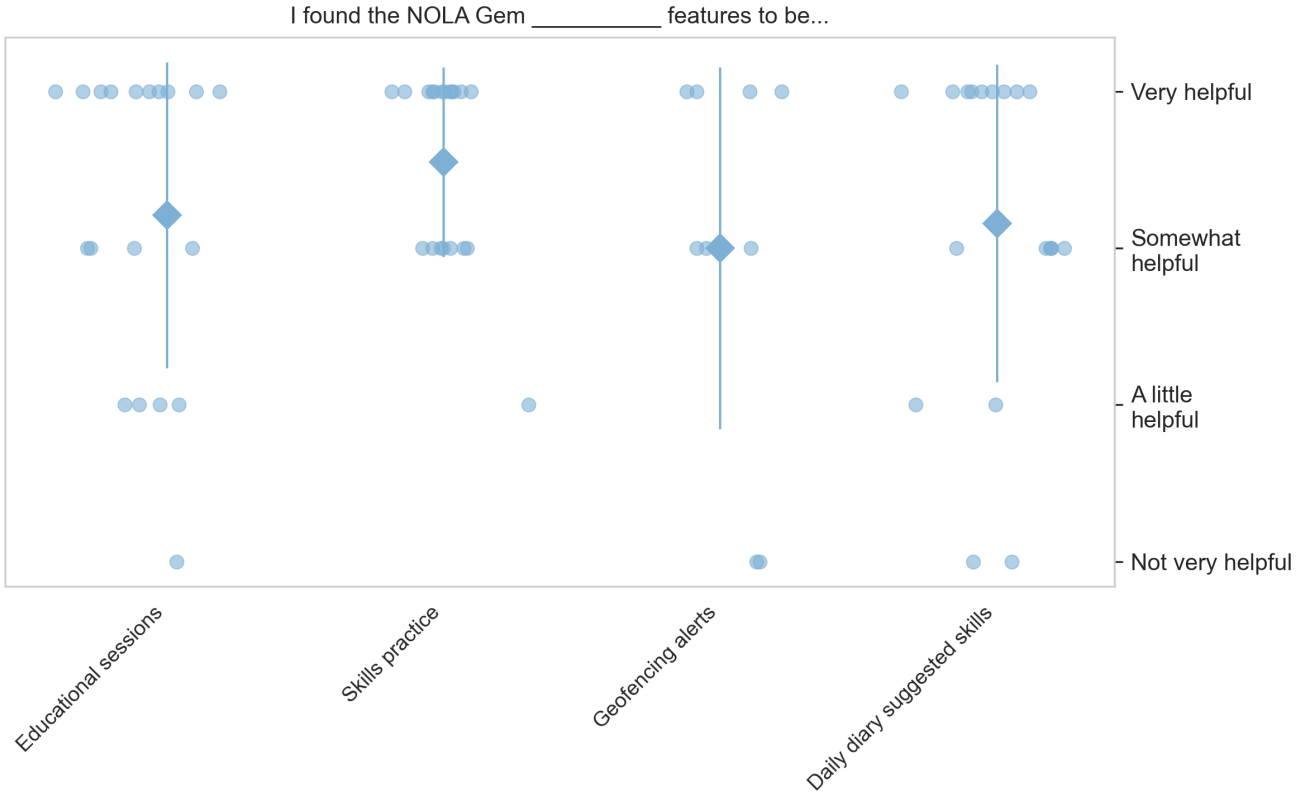


Figure 3. NOLA Gem acceptability: perceived helpfulness. Diamond markers denote mean ratings, and error bars indicate SDs. NOLA: New Orleans, Louisiana.



XSL•FO
RenderX

Figure 4. NOLA Gem acceptability (control): perceived helpfulness. Diamond markers denote mean ratings, and error bars indicate SDs. NOLA: New Orleans, Louisiana.



Because the control-treatment comparison was essentially additive, building features atop the already-demanding GEMA deployment, we tested for between-group differences in acceptability measures. Although, descriptively, GEMA measures tended to be higher (eg, treatment: mean 2.59, SD 0.73 vs control: mean 2.75, SD 0.46 for "able to open the app without confusion"), between-group *t* tests were uniformly null (*P*=.13 - .87).

NPS scores averaged (mean 7.70, SD 3.04). *Promoters* comprised 63.64% (n=14), and *detractors* comprised 22.73% (n=5) of the GEMA + NOLA Gem treatment sample, yielding an NPS metric of 40.91, approaching the generally accepted ≥50 "excellent" score [56].

Feasibility

At baseline, 32 participants were enrolled, with 30 (93.75%; this study's analytic sample) completing the 21-day trial period and completing ≥1 daily diary. Sociodemographic attributes of noncompleters are withheld to maintain participant confidentiality.

NOLA Gem paradata, collected from the treatment arm, showed that in a majority of the 607 completed daily diaries, across all participants (n=22), no skill was recommended: 463 (76.28%). Only 3 (13.64%) of the users completed 0 skills, and in 1 instance, technical errors may have prevented data recording.

These usage distributions are shown in Table 3.

XSL•FO
RenderX

Table . Living in the Face of Trauma (LIFT) skills: recommendations, time spent, completions: geographic ecological momentary assessment (GEMA) + New Orleans, Louisiana (NOLA) Gem (treatment group, n=22).

Skill	Recommendation frequency, n (%)[a]	Minutes spent		Completion frequency, n (%)[b]
		Count of 0-minute users, n_0	Median$_{>0}$ (range$_{>0}$)	
Adaptive coping	14 (9.72)	14	10.0 (1 - 409)	26 (12.68)
Breathing retraining	46 (31.94)	12	6.0 (1 - 13)	46 (22.44)
Choosing a coping strategy	15 (10.41)	17	3.5 (1 - 15)	30 (14.63)
Common irrational thoughts	14 (9.72)	16	3.0 (1 - 31)	27 (13.17)
Journal writing	14 (9.72)	17	36.0 (7 - 137)	16 (7.80)
Meditation	31 (21.53)	12	7.0 (2 - 59)	41 (20)
SMART[c] goal-setting	10 (6.94)	15	8.0 (1 - 74)	19 (9.27)

[a]Calculated as *n* skill-specific recommendations/144 total skills recommendations.

[b]Calculated as *n* skill-specific completions/205 skills completions; completions>recommendations due to user-determined "pull" (nonprompted) skills practice.

[c]SMART: specific, measurable, achievable, realistic, and able to be achieved in time.

User Experience

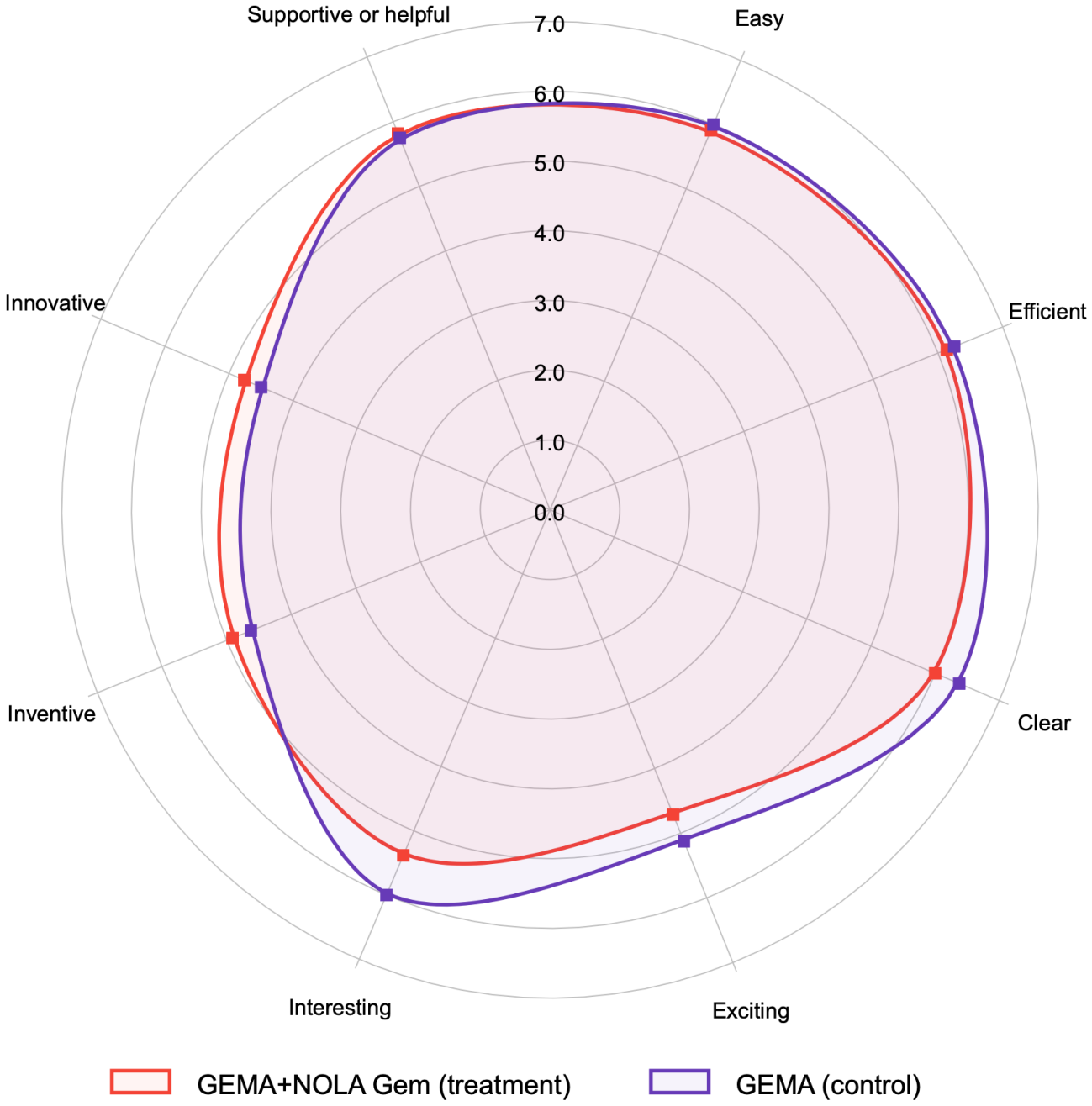
Overview

The short-form UEQ yielded generally good scores for NOLA Gem across the board, particularly on the "Supportive/helpful" (mean 5.82, SD 1.47), "Easy" (mean 5.91, SD 1.93), and "Efficient" (mean 6.14, SD 1.36) criteria. UEQ scores were qualitatively similar across conditions, suggesting negligible additional burden associated with the NOLA Gem components relative to GEMA alone.

The distribution of these scores by study arm is visualized in Figure 5.

XSL•FO
RenderX

Figure 5. Short-form User Experience Questionnaire (UEQ) mean scores spider plot [63]: geographic ecological momentary assessment (GEMA) (control) versus New Orleans, Louisiana (NOLA) Gem + GEMA (treatment).



Qualitative Description

When prompted, many participants were able to specify their favorite NOLA Gem features. These favorites did not, however, exhibit any across-case consistencies: "I like how it was designed, the colors were very calming … It was simple and it was easy to navigate" (participant C); "The meditation, the extra sessions, the definitions. I read everything and it was very reassuring and inventive" (participant K). Some participants intuited the "just-in-time" EMA-triggered mechanisms of action, as captured by a "You seem a little stressed … it will tell you what you can do" in vivo code: "[I liked the app] asking about your health and your condition … When it's checking up on your skills based on how you feel at one point in time in your daily diary of what you're doing, depending on what space you're in at the time" (participant I); "[I liked] the way when it asks you a question you seem a little stressed today it will tell you what you can do to relieve stress. I liked that" (participant R). Participant L described benefiting from a broad range of intervention components:

> *I had a better [understanding] of myself, meditation and breath control, I really enjoyed all parts of the app and most of the time I got satisfactory. Men don't talk and think about their feelings and I really liked how it brought my feelings to the surface. When I do meditation, thanks to your app I learned that I don't have to be professional, I can just do it how it works for me. It also played a role in increasing my self-esteem. I never practiced breathing but thanks to your app I learned how to do it through the stomach.*

XSL•FO
RenderX

These benefits sometimes came with inherent tradeoffs, as pointed out by participant C:

> *My favorite part is the different type of coping mechanisms that are offered. It will be great for those who need them, if they can get past the survey.*

Six persistently evident across-case topics were detected amid the qualitative responses, 4 subordinate to an overarching "Frictions and frustration" category. Cohen κ scores, by code, ranged from 0.49 to 1.00. As such, human-LLM coding applications of "EMA questionnaire length [κ=0.64]," "EMA prompt timing [κ=0.63]," "EMA item attunement [κ=0.68]," and "You seem a little stressed … it will tell you what you can do [in vivo; κ=0.49]" underwent additional human triangulation and review to achieve consensus.

The length and timing of the daily diary surveys were consistently mentioned as frustrations among NOLA Gem participants: "The number of questions. Unable to complete the questions in three hours. Inconvenience part of the three-hour time limit" (participant E); "Survey is too long and redundant. No reason to ask the same questions in a different way" (participant F). Participant E pointed to the daily conflicts potentially raised by study participation (and, by extension, intervention fidelity): "Being that I work on a boat that starts at 6 and doesn't dock until 10, I do not have time to complete the entries"; "The times for diaries were inconvenient due to my work schedule," participant P concurred. The daily diary's *attunement* to the lives of enrolled people with HIV, as well as each question's cultural congruence and ecological validity in tapping routine stressors, was questioned sporadically as well: "the questions were all basically the same, in some cases it did not include emotions that I thought should be included … it asked about seeing trash in the neighborhood and I don't really understand how is that relevant for daily life" (participant J); "I do not drink and I could not answer the questions. Did not feel the questions applied to everyday life experiences for PWH" (participant H). In addition, occasional malfunctions were experienced by certain users: "I don't like that it glitches, last time I used it, it wasn't glitching, this time you could not log in man I didn't like that" (participant G).

Although some participants described the app's "perfection" ("Nothing is going to be perfect, but this app is perfect. All the tools. Keep it authentic, just the way you designed it," participant L), various recommendations for updates were elicited. These included multimedia options: "… incorporating audio to where it's like soothing or calming white noise, ocean, waterfalls, stuff like that" (participant C). Participant J felt that opportunities remain to activate the *adaptive* aspect of JITAI design: "I think that it should be a little more interactive, more interactive and it should adjust to what questions are being asked … doesn't seem to be lifelike; say after it asked about being depressed, it would just move on to the next question rather than going deeper into it." Similarly, participant P saw opportunities for enhanced personalization of the geofencing feature: "Just ask if a person is okay with being reminded that they're in a stressful area before they get alerted about it being stressful if you didn't ask beforehand."

Discussion

Principal Findings

We trialed NOLA Gem, a trauma-informed JITAI leveraging GEMA inputs to provide people with HIV in New Orleans adaptive coping skills refreshers, against a GEMA-exclusive control, over a 21-day pilot period. In general, GEMA + NOLA Gem users (n=22) endorsed high acceptability, particularly in domains of perceived helpfulness, ease of comprehension, and ease of use. Of the 32 participants enrolled, 30 (93.75%) completed at least 1 daily diary, highlighting the app's feasibility. Among GEMA + NOLA Gem (treatment-arm) participants, for every skill, completions exceeded recommendations, indicating a nontrivial rate of "pull," or user-determined, app interactions, and suggesting valuable "Big E" (behavioral) engagement among a subset of dedicated users [39,44]. Most skills saw zero inflation among their "minutes spent" distributions, which were heavily skewed; this pattern might indicate user self-selection into accessible LIFT skills. Conversely, these users might miss out on new and potentially more beneficial coping skillsets. Qualitative user experience concerns converged on the frictions and interaction costs associated with the twice-daily EMA inputs: their sometimes-inconvenient timing and often-burdensome duration. While the relatively low skills recommendation rate *among* daily diary completers suggests those items may not be parsimoniously tapping frequently encountered stressors, the lower recommendation rates also prevent habituation—users "tuning out" a JITAI due to excess "push" notifications of no clear benefit [44]. Nevertheless, many participants declined to offer recommendations for next-generation changes, describing the app as "perfect" and acknowledging the value of the "just-in-time" adaptive coping skills delivery intuitively.

The *protocolization*, or defining the ruleset that determines how intraindividual dynamic inputs trigger just-in-time interventions [44], for each stressor-triggered adaptive coping skills prompt was formally mapped by the study team via feature hypothesis charting in 2022. Work on within-person JITAI optimization has advanced substantially since then, incorporating methods as diverse as reinforcement learning [71], *N*-of-1 model-predictive controller personalization algorithms [72], LLM-enabled free-text user preference updates in real time [73], and, most commonly, microrandomized controlled trials to isolate the most promising, in Nahum-Shani terms, "states of vulnerability/opportunity" and "states of receptivity" for momentary intervention delivery [74]. NOLA Gem, in contrast, was protocolized based on the findings of Aims I and II [21] and HIV biobehavioral domain expertise. These rulesets were static [55], even as they responded daily to EMA inputs. It is evident, based on our qualitative findings, that the sprawling battery of inflexibly scheduled daily diaries [55] was the primary driver of user burden and, as such, the foremost threat to acceptability. Despite some participants' insistence that "nothing" in NOLA Gem should change, a clear opportunity exists to optimize the JITAI mechanisms of action, *re*-protocolizing and pruning surplus daily diary measures in an evidence-based manner.

XSL•FO
RenderX

Though, by most standards, our findings support the advancement to a fully powered randomized controlled trial, in light of the abovementioned emergent methods for JITAI optimization, an additional microrandomized controlled trial cycle using contemporary techniques may further enhance acceptability. Launching intraindividual microrandomized trials, the most robust *stressor → state of receptivity → coping skill → proximal outcome → distal outcome* dynamics can be modeled and protocolized, at a within-person and within-user base (eg, across an outpatient HIV clinic network) level [75]. Indeed, the required in-app feature set already exists, in the form of single-item "SkillCheck" quizzes tap each delivered skill's success in improving a hypothesized JITAI-consistent proximal outcome.

Such improvements would benefit the prospective user base of people with HIV. However, the efficiencies built in by such a "pruning" cycle would benefit future clinical workflow integration—a potential prerequisite to a successful Phase III trial—and, ultimately, scalability and sustainment. The nonadoption, abandonment, scale-up, spread, and sustainability of technological innovation framework recognizes 7 domains. These include the condition, or health issue itself; the technology under consideration; its inherent value proposition, supply- or demand-side; adopters, including staff, patients, caregivers, and providers; organizational climate; political and regulatory considerations; and temporal aspects of implementation and organizational resilience (or lack thereof) [76]. Interventions that incorporate rampant complexity across the sum of these domains are shown, across empirical case studies and hermeneutic literature reviews [77], to rarely, if ever, achieve scale and sustainability. HIV prevention and treatment, despite past policy successes, remain subject to a volatile political climate [78]. People with HIV engaged in outpatient care endure poverty, violence, and medical and psychiatric comorbidities disproportionately [1,7,11,79]. Complexity, across the HIV continuum of care, is all but assured. Even at the pilot stage, it is incumbent upon technologists and HIV digital therapeutics investigators to prioritize simplicity, wherever possible, in intervention design and delivery. The NOLA Gem pilot study, and, in particular, the valuable insights contributed by NOLA Gem pilot users, lay the groundwork for concurrent optimization via simplification to advance next-generation HIV mobile health.

This study is subject to limitations as well. First, the pilot sample was not powered to detect between-group differences, either in GEMA versus GEMA + NOLA Gem acceptability (a proxy for differential interaction costs [61]), or predictors of disparate acceptability across subsets of outpatients with HIV. Null findings on these questions should be understood as highly preliminary. In particular, we are unable to extend the important work of Hubach et al [60], revealing EMA-related confidentiality concerns among a sample of rural men who have sex with men. A confluence of external factors—namely the COVID-19 variants still endemic in Louisiana, and Hurricane Ida [46,47]—delayed the trial launch substantially, hampering our ability to obtain our preregistered enrollment target of 60 participants, and to achieve balanced 1:1 random assignment. Though oversampling treatment-arm participants was ultimately to the benefit of these analyses, focused on the experiences of those NOLA Gem–assigned participants, our capacity to demonstrate equivocal acceptability across GEMA and JITAI paradigms is, similarly, undermined. We are also unable to gauge meaningful engagement metrics via paradata due to the small sample size, beyond a dichotomized "high" versus "low" engager subsetting [80]. Although these findings suggest future directions for NOLA Gem optimization, and support the need for a fully powered randomized controlled trial, attention toward protocolizing user-centered in-app adaptiveness to spur engagement is needed [45,75].

Conclusions

Trauma, particularly trauma rooted in firsthand violence exposures, disproportionately affects the health and wellbeing of people with HIV [1]. Interventions tailored to the needs of people with HIV must not only address environmental stressors across spatial and temporal dimensions [17-20], but they must also embrace principles of trauma-informed care [2,7]. To this end, we developed NOLA Gem, a smartphone-delivered JITAI adaptation of the effective LIFT group intervention [39]. This pilot study demonstrated the acceptability, feasibility, and usability of NOLA Gem, supporting future optimization, and a fully statistically powered evaluation of the app's efficacy among outpatients with HIV.

Acknowledgments

The authors extend their gratitude to Kathleen J. Sikkema, Meghna Daga, and Anudnya Durgule of Crowdbotics; the nursing and research staff at the Louisiana State University Clinical and Translational Research Center; their Tulane-affiliated study staff (Lauren Baker, Claire English, Leah Kimbrough, and Chelsea Solano) and the participants of the New Orleans, Louisiana Gem pilot trial.

Funding

New Orleans, Louisiana Gem is funded by National Institute on Alcohol Abuse and Alcoholism R34AA028961 and P60AA009803. Additional support was provided by National Institute of General Medical Sciences U54GM104940, which funds the Louisiana State University Clinical and Translational Science Center. SJS was supported by a Garvin Shands Saunders Foundation scholarship for the duration of this work.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Tang C, Goldsamt L, Meng J, et al. Global estimate of the prevalence of post-traumatic stress disorder among adults living with HIV: a systematic review and meta-analysis. BMJ Open 2020 Apr 27;10(4):e032435. [doi: 10.1136/bmjopen-2019-032435] [Medline: 32345695]
2. Brown MJ, Adeagbo O. Trauma-informed HIV care interventions: towards a holistic approach. Curr HIV/AIDS Rep 2022 Jun;19(3):177-183. [doi: 10.1007/s11904-022-00603-3] [Medline: 35353271]
3. Cuca YP, Shumway M, Machtinger EL, et al. The association of trauma with the physical, behavioral, and social health of women living with HIV: pathways to guide trauma-informed health care interventions. Womens Health Issues 2019;29(5):376-384. [doi: 10.1016/j.whi.2019.06.001] [Medline: 31303419]
4. Shangani S, van den Berg JJ, Dyer TV, Mayer KH, Operario D. Childhood sexual abuse, alcohol and drug use problems among Black sexual minority men in six US cities: findings from the HPTN 061 study. PLoS One 2022;17(12):e0279238. [doi: 10.1371/journal.pone.0279238] [Medline: 36548261]
5. Brown MJ, Harrison SE, Li X. Gender disparities in traumatic life experiences and antiretroviral therapy adherence among people living with HIV in South Carolina. AIDS Behav 2019 Nov;23(11):2904-2915. [doi: 10.1007/s10461-019-02440-9] [Medline: 30798459]
6. Mujumdar V, Pierson D, Briceño B, et al. Gathering trauma narratives: a qualitative study on the impact of self-identified traumas on people living with HIV (PLWH). N C Med J 2020;81(3):149-156. [doi: 10.18043/ncm.81.3.149] [Medline: 32366621]
7. Brezing C, Ferrara M, Freudenreich O. The syndemic illness of HIV and trauma: implications for a trauma-informed model of care. Psychosomatics 2015;56(2):107-118. [doi: 10.1016/j.psym.2014.10.006] [Medline: 25597836]
8. Hu FH, Liu P, Jia YJ, et al. Prevalence of mental health problems in people living with HIV: a systematic review and meta-analysis. Psychol Health Med 2025 Mar;30(3):397-413. [doi: 10.1080/13548506.2024.2424998] [Medline: 39504439]
9. Lang R, Hogan B, Zhu J, et al. The prevalence of mental health disorders in people with HIV and the effects on the HIV care continuum. AIDS 2023 Feb 1;37(2):259-269. [doi: 10.1097/QAD.0000000000003420] [Medline: 36541638]
10. Wirtz AL, Burns PA, Poteat T, et al. Abuse in the continua: HIV prevention and care outcomes and syndemic conditions associated with intimate partner violence among Black gay and bisexual men in the southern United States. AIDS Behav 2022 Nov;26(11):3761-3774. [doi: 10.1007/s10461-022-03705-6] [Medline: 35661018]
11. Mugavero M, Ostermann J, Whetten K, et al. Barriers to antiretroviral adherence: the importance of depression, abuse, and other traumatic events. AIDS Patient Care STDS 2006 Jun;20(6):418-428. [doi: 10.1089/apc.2006.20.418] [Medline: 16789855]
12. Glynn TR, Mendez NA, Jones DL, et al. Trauma exposure, PTSD, and suboptimal HIV medication adherence among marginalized individuals connected to public HIV care in Miami. J Behav Med 2021 Apr;44(2):147-158. [doi: 10.1007/s10865-020-00191-w] [Medline: 33098541]
13. Brown MJ, Brown AD, Haider MR, Edwards A, Crouch E, Li X. Childhood sexual trauma and antiretroviral therapy adherence: a mixed-methods systematic review. AIDS Behav 2021 Feb;25(2):475-491. [doi: 10.1007/s10461-020-03009-7] [Medline: 32844336]
14. Clum G, Madkour AS, Judson J, et al. Pathways from violence exposure to medication non-adherence: exploration of mental health symptoms and coping in people living with HIV in the U.S. South. AIDS Care 2025 Mar;37(3):372-383. [doi: 10.1080/09540121.2024.2445202] [Medline: 39847686]
15. Rojas D, Melo A, Moise IK, Saavedra J, Szapocznik J. The association between the social determinants of health and HIV control in Miami-Dade County ZIP codes, 2017. J Racial Ethn Health Disparities 2021 Jun;8(3):763-772. [doi: 10.1007/s40615-020-00838-z] [Medline: 32808192]
16. Wright IA, Reid R, Shahid N, et al. Neighborhood characteristics, intersectional discrimination, mental health, and HIV outcomes among Black women living with HIV, southeastern United States, 2019 2020. Am J Public Health 2022 Jun;112(S4):S433-S443. [doi: 10.2105/AJPH.2021.306675] [Medline: 35763751]
17. Duncan DT, Kim B, Al-Aljouni YA, Callander D. Neighborhood-level structural factors, HIV, and communities of color. In: Ojikutu BO, Stone VE, editors. HIV in US Communities of Color, 2nd edition: Springer; 2020:147-168. [doi: 10.1007/978-3-030-48744-7_9]
18. Shacham E, Scroggins SE, Ellis M. Implementing geospatial science and technology to get to zero new HIV infections. Curr HIV/AIDS Rep 2023 Jun;20(3):139-147. [doi: 10.1007/s11904-023-00658-w] [Medline: 37145264]
19. Theall KP, Wallace M, Felker-Kantor E, et al. Neighborhood alcohol environment: differential effects on hazardous drinking and mental health by sex in persons living with HIV (PLWH). AIDS Behav 2019 Dec;23(12):3237-3246. [doi: 10.1007/s10461-019-02632-3] [Medline: 31401740]

XSL•FO
RenderX

20. Theall KP, Felker-Kantor E, Wallace M, Zhang X, Morrison CN, Wiebe DJ. Considering high alcohol and violence neighborhood context using daily diaries and GPS: a pilot study among people living with HIV. Drug Alcohol Depend 2018 Jun 1;187:236-241. [doi: 10.1016/j.drugalcdep.2018.03.005] [Medline: 29684891]
21. Gasik RE, Smith EA, Skeen SJ, et al. Using geographic momentary assessment to explore spatial environment influences on wellbeing in people with HIV. AIDS Behav 2025 Jan;29(1):342-355. [doi: 10.1007/s10461-024-04527-4] [Medline: 39465467]
22. Gill A, Clum G, Molina P, Welsh D, Ferguson T, Theall KP. Life course stressors, latent coping strategies, alcohol use, and adherence among people with HIV. AIDS Behav 2025 Feb;29(2):589-599. [doi: 10.1007/s10461-024-04541-6] [Medline: 39546146]
23. Sikkema KJ, Hansen NB, Kochman A, et al. Outcomes from a group intervention for coping with HIV/AIDS and childhood sexual abuse: reductions in traumatic stress. AIDS Behav 2007 Jan;11(1):49-60. [doi: 10.1007/s10461-006-9149-8] [Medline: 16858634]
24. Sikkema KJ, Ranby KW, Meade CS, Hansen NB, Wilson PA, Kochman A. Reductions in traumatic stress following a coping intervention were mediated by decreases in avoidant coping for people living with HIV/AIDS and childhood sexual abuse. J Consult Clin Psychol 2013 Apr;81(2):274-283. [doi: 10.1037/a0030144] [Medline: 23025248]
25. Sikkema KJ, Hansen NB, Tarakeshwar N, Kochman A, Tate DC, Lee RS. The clinical significance of change in trauma-related symptoms following a pilot group intervention for coping with HIV-AIDS and childhood sexual trauma. AIDS Behav 2004 Sep;8(3):277-291. [doi: 10.1023/B:AIBE.0000044075.12845.75] [Medline: 15475675]
26. Lazarus RS, Folkman S. Stress, Appraisal and Coping: Springer; 1984.
27. Sikkema KJ, Wilson PA, Hansen NB, et al. Effects of a coping intervention on transmission risk behavior among people living with HIV/AIDS and a history of childhood sexual abuse. J Acquir Immune Defic Syndr 2008 Apr 1;47(4):506-513. [doi: 10.1097/QAI.0b013e318160d727] [Medline: 18176319]
28. Sikkema KJ, Mulawa MI, Robertson C, et al. Improving AIDS care after trauma (ImpACT): pilot outcomes of a coping intervention among HIV-infected women with sexual trauma in South Africa. AIDS Behav 2018 Mar;22(3):1039-1052. [doi: 10.1007/s10461-017-2013-1] [Medline: 29270789]
29. Smiley SL, Milburn NG, Nyhan K, Taggart T. A systematic review of recent methodological approaches for using ecological momentary assessment to examine outcomes in U.S. based HIV research. Curr HIV/AIDS Rep 2020 Aug;17(4):333-342. [doi: 10.1007/s11904-020-00507-0] [Medline: 32594365]
30. Balaskas A, Schueller SM, Cox AL, Doherty G. Ecological momentary interventions for mental health: a scoping review. PLoS One 2021;16(3):e0248152. [doi: 10.1371/journal.pone.0248152] [Medline: 33705457]
31. Zainal NH, Liu X, Leong U, Yan X, Chakraborty B. Bridging innovation and equity: advancing public health through just-in-time adaptive interventions. Annu Rev Public Health 2025 Apr;46(1):43-68. [doi: 10.1146/annurev-publhealth-071723-103909] [Medline: 39656954]
32. Tobin K, Heidari O, Volpi C, Sodder S, Duncan D. Use of geofencing interventions in population health research: a scoping review. BMJ Open 2023 Aug 3;13(8):e069374. [doi: 10.1136/bmjopen-2022-069374] [Medline: 37536963]
33. Wang L, Miller L. Assessment and disruption of ruminative episodes to enhance mobile cognitive behavioral therapy just-in-time adaptive interventions in clinical depression: pilot randomized controlled trial. JMIR Form Res 2023 Jan 5;7:e37270. [doi: 10.2196/37270] [Medline: 36602841]
34. Walters ST, Mun EY, Tan Z, et al. Development and preliminary effectiveness of a smartphone-based, just-in-time adaptive intervention for adults with alcohol misuse who are experiencing homelessness. Alcohol Clin Exp Res 2022 Sep;46(9):1732-1741. [doi: 10.1111/acer.14908] [Medline: 35869820]
35. Businelle MS, Perski O, Hébert ET, Kendzor DE. Mobile health interventions for substance use disorders. Annu Rev Clin Psychol 2024 Jul;20(1):49-76. [doi: 10.1146/annurev-clinpsy-080822-042337] [Medline: 38346293]
36. van Genugten CR, Thong MSY, van Ballegooijen W, et al. Beyond the current state of just-in-time adaptive interventions in mental health: a qualitative systematic review. Front Digit Health 2025;7:1460167. [doi: 10.3389/fdgth.2025.1460167] [Medline: 39935463]
37. Perski O, Hébert ET, Naughton F, Hekler EB, Brown J, Businelle MS. Technology-mediated just-in-time adaptive interventions (JITAIs) to reduce harmful substance use: a systematic review. Addiction 2022 May;117(5):1220-1241. [doi: 10.1111/add.15687] [Medline: 34514668]
38. Sizemore KM, Gray S, Wolfer C, et al. A proof of concept pilot examining feasibility and acceptability of the positively healthy just-in-time adaptive, ecological momentary, intervention among a sample of sexual minority men living with HIV. J Happiness Stud 2022 Dec;23(8):4091-4118. [doi: 10.1007/s10902-022-00587-2]
39. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
40. Yang MJ, Sutton SK, Hernandez LM, et al. A just-in-time adaptive intervention (JITAI) for smoking cessation: feasibility and acceptability findings. Addict Behav 2023 Jan;136:107467. [doi: 10.1016/j.addbeh.2022.107467] [Medline: 36037610]

XSL•FO
RenderX

41. Conway FN, Kane H, Bingaman A, et al. User experience of a just-in-time smartphone resonance breathing application for substance use disorder: acceptability, appropriateness, and feasibility. Subst Use Addctn J 2025 Apr;46(2):256-265. [doi: 10.1177/29767342241263675] [Medline: 39087448]
42. Metcalf O, Forbes D, Henry LM, et al. A just-in-time adaptive intervention (Shift) to manage problem anger after trauma: co-design and development study. JMIR Hum Factors 2025 May 22;12:e62960. [doi: 10.2196/62960] [Medline: 40402559]
43. Encore. URL: https://workforce.actsoft.com/encore/ [accessed 2025-06-28]
44. Nahum-Shani I, Naar S. Digital adaptive behavioral interventions to improve HIV prevention and care: innovations in intervention approach and experimental design. Curr HIV/AIDS Rep 2023 Dec;20(6):502-512. [doi: 10.1007/s11904-023-00671-z] [Medline: 37924458]
45. Welsh DA, Ferguson T, Theall KP, et al. The New Orleans alcohol use in HIV study: launching a translational investigation of the interaction of alcohol use with biological and socioenvironmental risk factors for multimorbidity in people living with HIV. Alcohol Clin Exp Res 2019 Apr;43(4):704-709. [doi: 10.1111/acer.13980] [Medline: 30748025]
46. Inside Louisiana's battle to control the Delta variant. Politico. URL: https://www.politico.com/news/2021/09/02/covid-data-gaps-delta-variant-508530 [accessed 2026-05-28]
47. Shultz JM, Trapido EJ, Kossin JP, et al. Hurricane Ida's impact on Louisiana and Mississippi during the COVID-19 Delta surge: complex and compounding threats to population health. Lancet Reg Health Am 2022 Aug;12:100286. [doi: 10.1016/j.lana.2022.100286] [Medline: 35663638]
48. McLellan AT, Luborsky L, Woody GE, O'Brien CP. An improved diagnostic evaluation instrument for substance abuse patients. The Addiction Severity Index. J Nerv Ment Dis 1980 Jan;168(1):26-33. [doi: 10.1097/00005053-198001000-00006] [Medline: 7351540]
49. Benson T. Digital innovation evaluation: user perceptions of innovation readiness, digital confidence, innovation adoption, user experience and behaviour change. BMJ Health Care Inform 2019 Apr;26(1):0. [doi: 10.1136/bmjhci-2019-000018] [Medline: 31039121]
50. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. Arch Intern Med 1998 Sep 14;158(16):1789-1795. [doi: 10.1001/archinte.158.16.1789] [Medline: 9738608]
51. Blevins CA, Weathers FW, Davis MT, Witte TK, Domino JL. The posttraumatic stress disorder checklist for DSM-5 (PCL-5): development and initial psychometric evaluation. J Trauma Stress 2015 Dec;28(6):489-498. [doi: 10.1002/jts.22059] [Medline: 26606250]
52. Tedeschi RG, Calhoun LG. The Posttraumatic Growth Inventory: measuring the positive legacy of trauma. J Trauma Stress 1996 Jul;9(3):455-471. [doi: 10.1007/BF02103658] [Medline: 8827649]
53. Berger BE, Ferrans CE, Lashley FR. Measuring stigma in people with HIV: psychometric assessment of the HIV stigma scale. Res Nurs Health 2001 Dec;24(6):518-529. [doi: 10.1002/nur.10011] [Medline: 11746080]
54. Felitti VJ, Anda RF, Nordenberg D, et al. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study. Am J Prev Med 1998 May;14(4):245-258. [doi: 10.1016/s0749-3797(98)00017-8] [Medline: 9635069]
55. Skeen SJ, Tokarz S, Gasik RE, et al. A trauma-informed, geospatially aware, just-in-time adaptive mhealth intervention to support effective coping skills among people living with HIV in New Orleans: development and protocol for a pilot randomized controlled trial. JMIR Res Protoc 2023 Oct 24;12:e47151. [doi: 10.2196/47151] [Medline: 37874637]
56. What is NPS? The ultimate guide to boosting your Net Promoter Score. Qualtrics. URL: https://www.qualtrics.com/experience-management/customer/net-promoter-score/ [accessed 2026-05-28]
57. Marcus A, editor. Design, user experience, and usability theories, methods, and tools for designing the user experience. Presented at: Third International Conference, DUXU 2014, Held as Part of the HCI International 2014; Jun 22-27, 2014. [doi: 10.1007/978-3-319-07668-3]
58. Schrepp M. User Experience Questionnaire Handbook: UEQ Research Team; 2023. URL: https://www.ueq-online.org/Material/Handbook.pdf [accessed 2025-06-28]
59. Using the PTSD checklist for DSM-5 (PCL-5). : National Center for PTSD; 2013 URL: https://www.ptsd.va.gov/professional/assessment/documents/using-PCL5.pdf [accessed 2025-06-28]
60. Hubach RD, O'Neil A, Stowe M, Giano Z, Curtis B, Fisher CB. Perceived confidentiality risks of mobile technology-based ecologic momentary assessment to assess high-risk behaviors among rural men who have sex with men. Arch Sex Behav 2021 May;50(4):1641-1650. [doi: 10.1007/s10508-019-01612-x] [Medline: 32078710]
61. Interaction cost. Nielsen Norman Group. URL: https://www.nngroup.com/articles/interaction-cost-definition/ [accessed 2025-06-28]
62. What's new in Stata 18. Stata. URL: https://www.stata.com/stata18/ [accessed 2026-05-28]
63. asjadnaqvi/stata-spider. GitHub. URL: https://github.com/asjadnaqvi/stata-spider [accessed 2025-01-14]
64. Hunter JD. Matplotlib: a 2D graphics environment. Comput Sci Eng 2007;9(3):90-95. [doi: 10.1109/MCSE.2007.55]

XSL•FO
RenderX

65. Sharma M, Zhuang D, Bilgic M. Active learning with rationales for text classification. Presented at: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; May 31 to Jun 5, 2015; Denver, Colorado p. 441-451. [doi: 10.3115/v1/N15-1047]
66. Dunivin ZO. Scalable qualitative coding with LLMs: chain-of-thought reasoning matches human performance in some hermeneutic tasks. arXiv. Preprint posted online on Jan 26, 2024. [doi: 10.48550/arXiv.2401.15170]
67. Meng H, Yang Y, Li Y, Lee J, Lee YC. Exploring the potential of human-LLM synergy in advancing qualitative analysis: a case study on mental-illness stigma. arXiv. Preprint posted online on Feb 12, 2024. [doi: 10.48550/arXiv.2405.05758]
68. ollama/ollama. GitHub. URL: https://github.com/ollama/ollama [accessed 2026-05-28]
69. Gemma models overview. Google AI for Developers. URL: https://ai.google.dev/gemma/docs [accessed 2026-05-28]
70. sskeen/nola_gem_acceptability. GitHub. URL: https://github.com/sskeen/nola_gem_acceptability [accessed 2026-06-05]
71. Gönül S, Namlı T, Coşar A, Toroslu İ. A reinforcement learning based algorithm for personalization of digital, just-in-time, adaptive interventions. Artif Intell Med 2021 May;115:102062. [doi: 10.1016/j.artmed.2021.102062] [Medline: 34001322]
72. Park J, Kim M, El Mistiri M, et al. Advancing understanding of just-in-time states for supporting physical activity (Project JustWalk JITAI): protocol for a system ID study of just-in-time adaptive interventions. JMIR Res Protoc 2023 Sep 26;12:e52161. [doi: 10.2196/52161] [Medline: 37751237]
73. Karine K, Marlin BM. Combining LLM decision and RL action selection to improve RL policy for adaptive interventions. arXiv 2025 Jan 13. [doi: 10.48550/arXiv.2501.06980]
74. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: a pragmatic framework. Health Psychol 2015 Dec;34S:1209-1219. [doi: 10.1037/hea0000306] [Medline: 26651462]
75. Liu X, Deliu N, Chakraborty B. Microrandomized trials: developing just-in-time adaptive interventions for better public health. Am J Public Health 2023 Jan;113(1):60-69. [doi: 10.2105/AJPH.2022.307150] [Medline: 36413704]
76. Greenhalgh T, Abimbola S. The NASSS framework—a synthesis of multiple theories of technology implementation. Stud Health Technol Inform 2019 Jul 30;263:193-204. [doi: 10.3233/SHTI190123] [Medline: 31411163]
77. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367. [doi: 10.2196/jmir.8775] [Medline: 29092808]
78. Domestic HIV funding in the White House FY2026 budget request. Kaiser Family Foundation. URL: https://www.kff.org/hivaids/issue-brief/domestic-hiv-funding-in-the-white-house-fy2026-budget-request/ [accessed 2026-05-28]
79. Katz IT, Thomson DR, Ravishankar S, et al. Intersectional forces of urban inequality and the global HIV pandemic: a retrospective analysis. BMJ Glob Health 2025 Apr 9;10(4):e014750. [doi: 10.1136/bmjgh-2023-014750] [Medline: 40204462]
80. Jin S, Kim B, Han K. "I don't know why I should use this app": holistic analysis on user engagement challenges in mobile mental health. Presented at: ACM CHI Conference on Human Factors in Computing Systems (CHI '25); Apr 26 to May 1, 2025. [doi: 10.1145/3706598.3713732]

Abbreviations

ACE: adverse childhood experience
EMA: ecological momentary assessment
GEMA: geographic ecological momentary assessment
JITAI: just-in-time adaptive intervention
LIFT: Living in the Face of Trauma
LLM: large language model
NOLA: New Orleans, Louisiana
NPS: Net Promoter Score
PTSD: posttraumatic stress disorder
REDCap: Research Electronic Data Capture
UEQ: User Experience Questionnaire

Edited by A Kushniruk; submitted 15.Jul.2025; peer-reviewed by R Miranda, SA John; revised version received 27.Mar.2026; accepted 31.Mar.2026; published 26.Jun.2026.

Please cite as:
Skeen SJ, Tokarz S, Gasik RE, Smith EA, Theall KP, Clum GA
A Geospatially Customizable Culturally Tailored Just-in-Time Adaptive Intervention for Violence-Affected People Living With HIV: Mixed Methods Acceptability, Feasibility, and User Experience Study
JMIR Hum Factors 2026;13:e80705
URL: https://humanfactors.jmir.org/2026/1/e80705
doi:10.2196/80705

XSL•FO
RenderX

© Simone J Skeen, Stephanie Tokarz, Rayna E Gasik, Ethan A Smith, Katherine P Theall, Gretchen A Clum. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 26.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Smartphone App Using Reinforcement Learning for Obesity: Single-Arm Feasibility Study

Ken Kurisu[1], MD, PhD; Yoshiharu Yamamoto[2], PhD; Tomohisa Aoyama[3], MD, PhD; Toshimasa Yamauchi[3], MD, PhD; Kazuhiro Yoshiuchi[1], MD, PhD

[1]Department of Stress Sciences and Psychosomatic Medicine, Graduate School of Medicine, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, Japan

[2]Educational Physiology Laboratory, Graduate School of Education, The University of Tokyo, Tokyo, Japan

[3]Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

Corresponding Author:
Kazuhiro Yoshiuchi, MD, PhD
Department of Stress Sciences and Psychosomatic Medicine, Graduate School of Medicine, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, Japan

Abstract

Background: While behavioral interventions remain an evidence-based treatment for obesity, they often require long durations and frequent sessions. To address this, we hypothesized that interventions delivered in daily life via a smartphone app combined with personalized optimization using reinforcement learning may effectively support behavior changes.

Objective: This study aimed to develop and evaluate the feasibility of such an app for individuals with obesity.

Methods: We developed a smartphone app to assist in setting and reviewing daily behaviors related to weight loss. On the screen on which daily behaviors were shown, the order of presentation was optimized using Thompson sampling, a multiarmed bandit algorithm. Twenty individuals with obesity used the app for 4 weeks, and the daily app use rates were quantified. Body weight and mood status were measured daily during the study, and a brief-type self-administered diet history questionnaire and the International Physical Activity Questionnaire were administered at the beginning and end of the study. Changes in these measures were evaluated using the Wilcoxon signed rank test. Furthermore, the longitudinal data collected during this study were analyzed using a linear mixed-effects model to examine factors related to the number of behaviors performed daily.

Results: All 20 recruited individuals with obesity completed the 4-week study schedule. The median app use rate was 98.3% (range 76.9% - 100%). Significant improvements were observed in BMI (median at start 34.9 kg/m^2, range 27.4-52.9; median at end 34.1 kg/m^2, range 26.7 - 51.0; P=.01), as well as daily energy intake and weekend sitting time. The linear mixed-effects model showed a significant association between higher preceding depressive mood levels and fewer behaviors (P<.001).

Conclusions: The feasibility of the smartphone app using reinforcement learning for obesity was sufficient, and the potential effectiveness of the treatment was suggested. Preceding depressive mood may influence daily behaviors related to weight loss.

(JMIR Hum Factors 2026;13:e77323) doi:10.2196/77323

KEYWORDS

cognitive behavioral therapy; ecological momentary intervention; machine learning; multiarmed bandit; obesity; reinforcement learning; smartphone app

Introduction

Obesity leads to diseases such as type 2 diabetes and hypertension, which further increase the risk of serious complications, including cardiovascular diseases and certain types of cancer [1]. Obesity often coexists with psychiatric disorders and has a bidirectional relationship with them [2]. Obesity reduces well-being, resulting in a vicious cycle that leads to further weight gain [3]. The prevalence of obesity is increasing worldwide, including in Japan [4], underscoring the need for effective treatment.

Current obesity treatments include glucagon-like peptide-1 receptor agonists and bariatric surgery, which cause adverse gastrointestinal effects and postoperative psychiatric comorbidities, respectively [5-7]. Behavioral interventions remain important as a low-risk and low-cost treatment [5,8], and successful long-term weight maintenance often relies on behavioral strategies [5]. Although structured cognitive behavioral therapy for obesity has been developed [9,10], the frequent and long duration of sessions burdens both participants and therapists. Furthermore, behavioral science findings indicate that reinforcers should be provided immediately after performing behaviors [11], which is challenging in outpatient settings. To

XSL•FO
RenderX

address these issues, we hypothesized that ecological momentary intervention (EMI) through smartphone apps in daily life would be helpful [12,13].

In the context of EMI for obesity, tailored presentation of behavioral content may play an important role in promoting effective behavior change. In face-to-face cognitive behavioral therapy, behavior changes for diet and physical activity are tailored to individuals' daily lifestyles and preferences [9,10]. From a technical perspective, machine learning and reinforcement learning have been widely deployed for content optimization in modern web applications and smartphone apps [14]. One study used a multiarmed bandit algorithm, an optimization technique using reinforcement learning, to suggest health behaviors [15]. The algorithm explores items with limited prior information to reduce uncertainty and simultaneously exploits items with higher estimated rewards based on accumulated data. Such optimization by multiarmed bandit algorithms is applicable for recommending daily behaviors to individuals with obesity through smartphone apps.

Numerous studies on smartphone apps and web-based interventions for obesity have been conducted, and an umbrella review suggests that they can achieve greater weight loss when combined with usual treatments [16]. Relevant studies have involved sending text messages related to physical activity or dietary intake [17-20]. However, to the best of our knowledge, no study has used an app specifically designed for setting daily behaviors with personalization using machine learning or reinforcement learning. Accordingly, such an app would be novel and could lead to more effective treatments.

Additionally, the completion of behaviors related to weight loss suggested through smartphone apps may be influenced by daily psychosocial factors. Many psychosocial factors reportedly influence the maintenance and progression of obesity [2,3,5]. Studies using ecological momentary assessments, which analyze data collected in daily life, have suggested that preceding mood status and other factors influence daily calorie intake and physical activity in individuals with obesity [21-24]. Examining the relationship between daily behaviors and preceding mood collected through a smartphone app may help improve the app.

In summary, we hypothesized that interventions in daily life delivered via smartphone apps could effectively address current challenges in behavioral interventions and that personalization through multiarmed bandit algorithms could further enhance their effectiveness. Therefore, this study aimed to develop a smartphone app with a multiarmed bandit algorithm for individuals with obesity and evaluate its feasibility. In addition, as psychosocial factors may influence the effectiveness of this intervention, the secondary aim was to examine factors influencing daily behaviors related to weight loss.

Methods

Ethical Considerations

This study was approved by the Institutional Review Board of the University of Tokyo (2022061NI). The research protocol was registered with the University Hospital Medical Information Network Clinical Trials Registry (UMIN000048667). All the participants provided written informed consent. All data were anonymized, and access was limited to authorized research team members only. Participants received up to JP¥ 20,000 (approximately US $150 based on the exchange rate at the time of study start) via QUO Card.

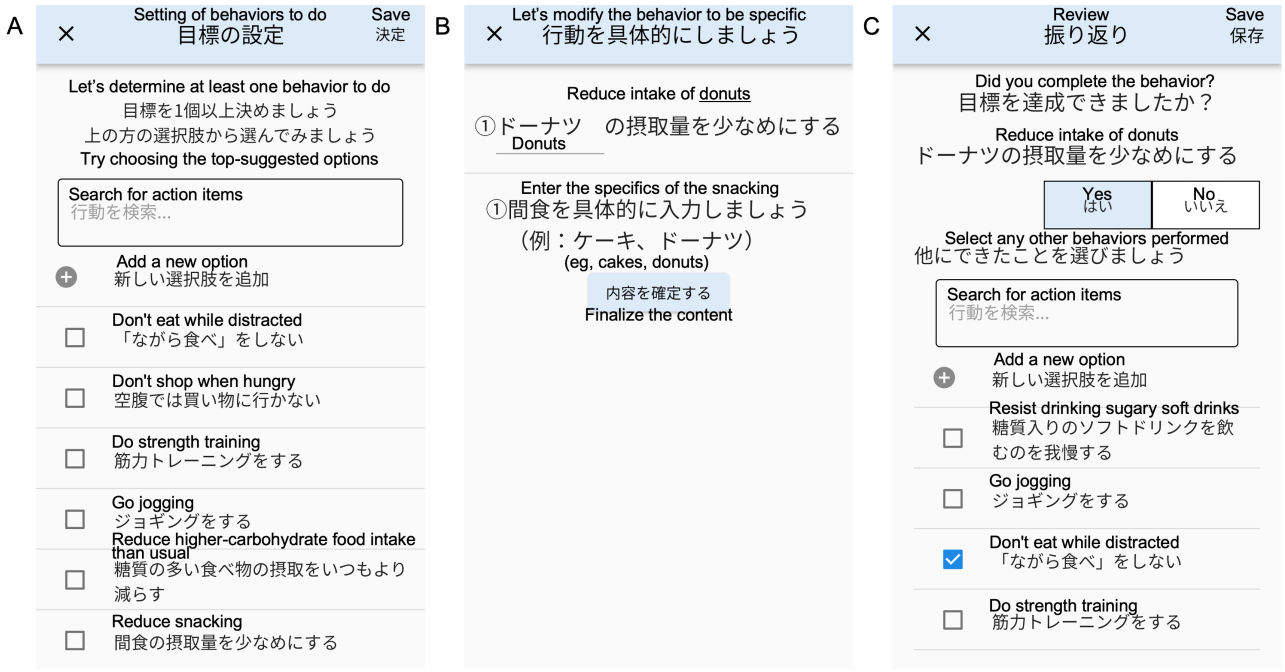
Smartphone App

Figure 1 shows screenshots of the smartphone app we developed for obesity management. Users launch the app upon receiving a notification at the end of the day, typically in the evening at an individually preset time, and select behaviors to perform the following day (Figure 1A). The selected behaviors can be further modified for specificity (Figure 1B). The user receives reminder notifications several times the following day regarding the behaviors they have set. Upon receiving a notification at the end of the following day, users launch the app to record whether they performed the selected behaviors and log any additional behaviors (Figure 1C). Users then select the behaviors to perform the following day. The researchers prepared a behavioral list based on the cognitive behavioral therapy manual [9] and the Japanese guidelines for obesity management [4]. Each participant can also create original items individually. The app displays a complimentary message and a graphical presentation of the number of behaviors performed immediately after recording them.

The presentation order was optimized to display easily achievable, low-effort items earlier on the screen (Figure 1A). This optimization was based on the theory of cognitive behavioral therapy [9,10] and the Fogg behavior model [25], similar to a study on a smartphone app using reinforcement learning to promote health-related activities [15]. A multiarmed bandit algorithm was used, which simultaneously performs exploration (collecting data to determine which items are achievable) and exploitation (presenting achievable items based on the collected data). This has the advantage that it can be applied to even a small number of users while ensuring a diversity of suggestions [14].

XSL•FO
RenderX

Figure 1. Screenshots of the app (translated from Japanese into English): (A) at the end of the day, users select behaviors to perform the following day, with on-screen guidance encouraging them to choose top-listed items suggested by a reinforcement learning algorithm; (B) selected behaviors can be specified further (eg, “Reduce snacking” to “Reduce intake of donuts”); and (C) a notification is sent at the end of the following day to record whether the selected behaviors were performed. Users can also log additional behaviors performed.



We used Thompson sampling, a multiarmed bandit algorithm that does not require prior hyperparameter adjustments [14]. In this study, each behavior was treated as an independent arm and scored by a randomly sampled value of θ from the following beta distribution, where *N* is the number of times the user selects the item and *a* is the number of executions:

$$p\theta a,N=\theta a(1-\theta)N-a\int 01\theta a(1-\theta)N-ad\theta$$

When *N* is small, samples of θ from the beta distribution exhibit a large variance, and the model assigns random scores to behaviors with few selections, which promotes exploration. As *N* increases, the samples converge toward the actual achievable rate of each behavior, which promotes exploitation by assigning higher scores to behaviors with higher completion rates than to those with lower completion rates. In this study, each participant had an independent model trained using only their own behavior data. Each behavior was initialized with a uniform prior over the interval (0, 1).

The app was developed using Flutter (Google) and Firebase (Google). Reinforcement learning was implemented using Python (version 3.9; Python Software Foundation) on the Google Cloud Platform. The app was distributed exclusively to participants’ smartphones (iPhone or Android) through DeployGate (DeployGate, Inc) [26] rather than being publicly released on the Apple App Store or Google Play Store.

Study Participants

This study included individuals diagnosed with “obesity disease” according to criteria established by the Japan Society for the Study of Obesity. This disease is defined as a BMI of ≥25 kg/m^2 with one or more obesity-related disorders such as type 2 diabetes or visceral fat accumulation that likely causes health problems [4,27]. Individuals aged ≥18 years attending the University of Tokyo Hospital were eligible for this study. Exclusion criteria were (1) extremely severe physical or psychiatric diseases, (2) conditions for which the burden of the study would be harmful, and (3) lack of decision-making ability.

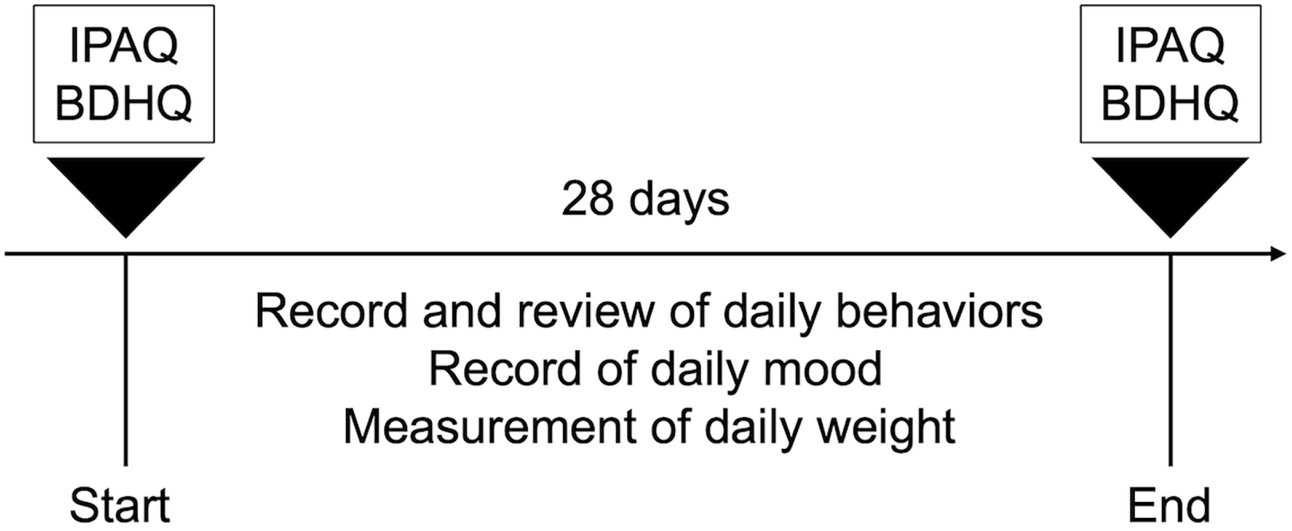
A previous study examining the feasibility of an app using a multiarmed bandit algorithm included approximately 20 participants over 3 weeks [15]. Following this, we set the study period to 4 weeks (28 days) and the sample size to 20. Eligible individuals were recruited continuously until the target sample size was achieved.

Study Procedure

Figure 2 provides the study schedule. At the beginning of the study, the Japanese version of the International Physical Activity Questionnaire (IPAQ) long form, last–7-day version, was used to assess physical activity [28]. The brief-type self-administered diet history questionnaire (BDHQ) was used to assess nutritional intake [29]. The BDHQ consists of questions assessing the consumption frequency of 58 food and beverage items over the preceding month and estimates average daily energy intake and various nutrient intakes.

XSL•FO
RenderX

Figure 2. Study schedule. The study period lasted 28 days, with the International Physical Activity Questionnaire (IPAQ) and brief-type self-administered diet history questionnaire (BDHQ) measured at the start and end of the study. Daily behaviors related to weight loss, mood, and weight were recorded during the study period.



During the study, daily behaviors were recorded using the app. Participants received a notification daily at a random time within a prearranged 1-hour interval and recorded their mood status (Figure 3). Nine items from the Depression and Anxiety Mood Scale were used to calculate scores for anxiety, positive affect, and depression [30]. This was recorded on a 101-point (0 - 100) visual analog scale. Additionally, we provided the participants with a body scale (HN300T2; OMRON) connected via Bluetooth to a smartphone (iPhone or Android) to record weight automatically [31]. Their weight was recorded daily upon waking.

XSL•FO
RenderX

Figure 3. Screenshot of records of daily momentary mood status (translated from Japanese into English).



At the end of the 4-week study, the IPAQ and the BDHQ were administered, as at the start of the study. The top 5 behaviors displayed on the app screen were evaluated on paper using a visual analog scale regarding their subjective likelihood of achievement. In addition, qualitative interviews were conducted to identify potential improvements to the app.

Statistical Analyses

Feasibility was assessed via app use rate, defined as the percentage of days on which the behavior was recorded during the study period. Relevant EMI studies using smartphone apps for obesity or type 2 diabetes have reported recording rates of approximately 80% to 95% [20,32,33], which we set as the benchmark for feasibility.

BMI, physical activity measured using the IPAQ, and calorie and salt intake measured using the BDHQ were compared between the start and end of the study using the Wilcoxon signed rank test, with statistical significance set at .05.

These analyses were performed using R (version 4.3.1; R Foundation for Statistical Computing).

Analysis of Predictors of Daily Behaviors for Weight Loss

In this post hoc analysis, we combined data on daily behaviors with mood status recorded within the preceding 24 hours. Only behavioral records reviewed more than 12 hours and less than 36 hours after setting were included. A linear mixed-effects model was developed using this nested longitudinal dataset, with the daily number of behaviors performed as the outcome. The outcome was defined as the sum of the preselected behaviors that were actually performed and any additional behaviors.

XSL•FO
RenderX

Level 1 variables included (1) preceding mood status (anxiety, positive affect, and depression) measured using the Depression and Anxiety Mood Scale, which was scaled from 0 to 1; (2) the difference between the weight recorded immediately before the behavior review and the weight at the start of the study; (3) the difference between the weight recorded immediately before the behavior review and the weight recorded at one earlier time point; and (4) the number of weeks, ranging from 1 to 4. Level 2 variables included (1) sex; (2) BMI at the start, centered by subtracting 35; and (3) the presence of comorbid psychiatric disorders. The inclusion of these variables and random effects was determined based on the Akaike information criterion.

R (version 4.3.1) and the package *lmerTest* (version 3.1-3) were used for this analysis.

Results

Feasibility Evaluation

Participants were recruited from August 2022 to July 2023, and all 20 individuals with obesity completed the 4-week study. Table 1 provides the descriptive statistics. The median app use rate, defined as the percentage of days on which behaviors were recorded, was 98.3% (range 76.9% - 100%), with a mean of 94.8% (SD 7.1%).

Table . Characteristics of the participants (N=20).

	Values
Age (years), median (range)	43 (21-66)
Sex, n (%)	
Male	9 (45)
Female	11 (55)
BMI at start (kg/m^2), median (range)	34.9 (27.4-52.9)
Height (cm), median (range)	164.4 (151.0-184.0)
Weight at start (kg), median (range)	93.6 (70.5-158.3)
Bariatric surgery (laparoscopic sleeve gastrectomy)	
Individuals who underwent surgery, n (%)	4 (20)
Postoperative period (months), median (range)	17.6 (7.8-45.1)
Glucagon-like peptide-1 receptor agonist use	
Semaglutide, n (%)	7 (35)
Liraglutide, n (%)	1 (5)
Duration of use (months), median (range)	6.3 (0.3-154.9)
Physical comorbidities, n (%)	
Type 2 diabetes	12 (60)
Hypertension	14 (70)
Obstructive sleep apnea	10 (50)
Dyslipidemia	13 (65)
Psychiatric comorbidities, n (%)	
Depression	2 (10)
Bipolar disorder	2 (10)
Adjustment disorder	2 (10)
Sleep disorder	4 (20)
Binge eating disorder	2 (10)

Treatment Effectiveness

Table 2 shows treatment-related results. BMI decreased significantly (median at start 34.9 kg/m^2, range 27.4 - 52.9; median at end 34.1 kg/m^2, range 26.7 - 51.0; *P*=.01). In total, 15% (3/20) of the participants gained weight during the study period, 2 of whom had undergone bariatric surgery. Significant changes were also observed in calorie intake and weekend sitting time.

XSL•FO
RenderX

Table . Changes in treatment-related measurements.

	Before the intervention (baseline), median (range)	After the intervention (4 weeks), median (range)	*P* value
BMI (kg/m^2)	34.9 (27.4 - 52.9)	34.1 (26.7 - 51.0)	.01
BDHQ[a]			
Calorie intake (kcal/day)	1603 (541 - 4191)	1367 (527 - 2572)	.04
Salt intake (g/day)	9.0 (3.7 - 21.1)	8.5 (4.4 - 20.9)	.09
IPAQ[b]			
Physical activity (MET[c] minutes/week)	2716 (99 - 13566)	3413 (537 - 13506)	.70
Weekday sitting (hours/day)	5 (1-14)	4 (1-13)	.06
Weekend sitting (hours/day)	6.5 (2-14)	4.5 (1-12)	.008

[a]BDHQ: brief-type self-administered diet history questionnaire.

[b]IPAQ: International Physical Activity Questionnaire.

[c]MET: metabolic equivalent of task.

Usability Survey

The median number of behaviors performed over the 4 weeks, including individually created behaviors, was 11 (range 6 - 27). The subjective evaluation of the achievability of the items presented at the top of the app screen, which ranged from 0 to 100, had a median score of 85 (range 58 - 100).

In the qualitative interviews, the most frequently raised concern was the following (8/20, 40%):

> *Toward the latter part of the study period, I tended to choose only a few types of items displayed at the top and felt a sense of monotony.*

This implies that the algorithm converged on a small subset of items with higher completion rates toward the end of the study, resulting in these items being repeatedly displayed at the top and predominantly selected by users.

The next most frequently raised concern was the following (6/20, 30%):

> *I was unable to modify some items to be more specific.*

This means that, although each behavior could be further specified, as exemplified in Figure 1B, users had difficulty doing so for several items.

Predictors of Daily Behaviors for Weight Loss

A total of 432 records collected during the study were included in the analysis. The final model developed based on the Akaike information criterion is presented below, where subscript *i* denotes each participant and subscript *j* denotes the time point. Yi,j is the daily number of behaviors performed, Depressioni,j is the preceding depressive mood, and Psychiatryi indicates the presence of psychiatric comorbidities. Fixed and random effects are represented by γ and ζ, respectively.

Level 1:

$$Y_{i,j}=\Pi_{0,i}+\Pi_{1,i}\text{Depression}_{i,j}+\varepsilon_{i,j}$$

Level 2:

$$\Pi_{0,i}=\gamma_{0,0}+\zeta_{0,i}$$

$$\Pi_{1,i}=\gamma_{1,0}+\gamma_{1,1}\text{Psychiatry}_i+\zeta_{1,i}$$

The coefficients of each variable are listed in Table 3. The following can be interpreted from each coefficient: (1) a higher preceding depressive mood was associated with fewer behaviors, and (2) individuals with comorbid psychiatric disorders had smaller coefficients for preceding depressive mood.

Table . Coefficients of the linear mixed-effects model for the daily number of performed behaviors.

	Fixed effects coefficient (95% CI)	*P* value
Intercept ($\gamma_{0,0}$)	4.45 (3.47 to 5.43)	<.001
Depression ($\gamma_{1,0}$)	−1.88 (−2.87 to −0.88)	<.001
Depression × psychiatric comorbidity ($\gamma_{1,1}$)	1.59 (0.34 to 2.85)	.01

Discussion

Principal Results

The smartphone app developed in this study for individuals with obesity was used by 20 recruited participants over 4 weeks, with a median use rate of 98.3%. Participants' BMI decreased significantly during the study period. Furthermore, a linear mixed-effects model based on data obtained through the app showed that a higher preceding depressive mood was associated with a reduced number of completed behaviors related to weight loss.

XSL•FO
RenderX

The median use rate was 98.3%, which exceeded the adherence rates of approximately 80% to 95% reported in relevant studies [20,32,33], indicating sufficient feasibility. Future studies should address the concerns identified in the qualitative interviews. To prevent the selection of behaviors from being monotonous, items that are consistently completed may be retired from the list. In addition, improving the clarity of item explanations may effectively address the challenges in specifying item content.

The median decrease in BMI was 2.3% over 4 weeks. The Japanese clinical guidelines recommend a 3% weight loss for individuals with a BMI of <35 kg/m^2 and a 5% to 10% weight loss for those with a BMI of ≥35 kg/m^2 over 3 to 6 months [4]. Accordingly, the weight loss achieved in this study appeared sufficient. Calorie intake and weekend sitting time improved, and these behavior changes may have contributed to weight loss. Thus, the developed app may be effective for weight loss.

The effect size should be interpreted with caution due to several features of this study design. Daily weight measurements were conducted throughout the study, which may have influenced weight loss [34]. In total, 15% (3/20) of the participants gained weight during the study, 2 of whom had undergone bariatric surgery, a population known to frequently experience weight regain [35]. A previous umbrella review of smartphone app and web application interventions for obesity showed a mean difference of −0.12 kg compared with face-to-face treatment and −4.32 kg compared with no treatment [16]. However, our study was a single-arm trial with a follow-up duration different from those of previous studies, making direct comparisons difficult. Therefore, further studies are warranted to more accurately determine the effect size of the app developed in this study.

Given the high ratings for the subjective likelihood of achievement of the top-suggested items, with a median of 85 on a scale from 0 to 100, the multiarmed bandit algorithm seemingly approached convergence. Qualitative interviews revealed that the participants tended to choose only a few types of items presented at the top and reported a sense of monotony, further suggesting algorithm convergence. Future studies with more accumulated data should evaluate the convergence of the algorithm and may consider applying other optimization approaches such as collaborative filtering [14]. On the basis of the Fogg behavior model [25], this study prioritized behaviors based solely on completion rates; however, future studies should also consider the behaviors' contributions to clinical outcomes such as weight loss, along with the effort and duration required for achievement.

In the post hoc analysis, the linear mixed-effects model showed that preceding depressive mood was related to weight loss behaviors. This aligns with a previous study showing that baseline depression is associated with emotional eating and binge eating [22]. One study showed that delay discounting mediates the association between depressive symptoms and the onset of diabetes [36], which may help explain these findings. Another study suggested that individuals with higher levels of depression exhibited unhealthy behaviors such as physical inactivity [37]. On the basis of these findings, future studies may benefit from providing suggestions on the number of behaviors to select on the app and prioritizing lower-effort behaviors according to daily depression levels. The interaction coefficient of depressive mood and psychiatric comorbidity was positive, indicating that, when adjusted for depressive mood as a covariate, the number of behaviors performed was large for individuals with psychiatric diseases. This may suggest that treating psychiatric disorders positively influences weight loss behaviors; however, further studies are needed to interpret these findings.

Limitations and Future Directions

This study has several limitations. First, the study period was short, and long-term use of the app remains unexplored. Second, the dietary and physical activity questionnaires were administered only at baseline and after the intervention, which limited assessment of temporal changes and intervention feasibility. Third, energy intake estimated using the BDHQ appeared low. Alternative assessment tools such as the Automated Self-Administered 24-hour dietary recall [38] may be preferable in future studies. Fourth, this study used weight alone without distinguishing between fat and muscle mass; this is particularly relevant for participants receiving glucagon-like peptide-1 receptor agonists or who underwent bariatric surgery. Future studies should assess body composition using tools such as bioelectrical impedance analysis. Fifth, in the optimization algorithm, the behaviors were evaluated only with regard to achievability. Their contributions to weight loss may vary and should ideally be incorporated into the algorithm. Sixth, the impact of the multiarmed bandit algorithm on optimization remains unevaluated. Seventh, the effect size for weight loss was inconclusive in this single-arm study. Finally, the analysis of the relationship between preceding depressive mood and daily behaviors used only the subjective input of behavioral execution.

Future research should improve the app based on the findings of this study and assess the treatment efficacy through a randomized controlled trial with a longer follow-up period. Additionally, future studies should assess the contribution of optimization to weight loss using a randomized design that compares optimized and nonoptimized interventions. Furthermore, although this study used an item-fixed measurement based on classic test theory, future research could use computer-adaptive tests [39,40] to measure mood status more efficiently.

Conclusions

The smartphone app for obesity developed in this study demonstrated sufficient feasibility, and its potential effectiveness for weight loss was suggested. Data obtained through this app indicate that a preceding depressive mood may influence daily behaviors related to weight loss.

XSL•FO
RenderX

Funding

This study was supported by a Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research (grants 22K07390 and 25H00388).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. González-Muniesa P, Mártinez-González MA, Hu FB, et al. Obesity. Nat Rev Dis Primers 2017 Jun 15;3:17034. [doi: 10.1038/nrdp.2017.34] [Medline: 28617414]
2. Avila C, Holloway AC, Hahn MK, et al. An overview of links between obesity and mental health. Curr Obes Rep 2015 Sep;4(3):303-310. [doi: 10.1007/s13679-015-0164-9] [Medline: 26627487]
3. Cameron AJ, Magliano DJ, Dunstan DW, et al. A bi-directional relationship between obesity and health-related quality of life: evidence from the longitudinal AusDiab study. Int J Obes (Lond) 2012 Feb;36(2):295-303. [doi: 10.1038/ijo.2011.103] [Medline: 21556045]
4. Guidelines for the management of obesity disease 2022. Japan Society for the Study of Obesity.: Life Science Publishing; 2022. URL: https://www.jasso.or.jp/contents/magazine/journal.html#himan22 [accessed 2025-05-25]
5. Elmaleh-Sachs A, Schwartz JL, Bramante CT, Nicklas JM, Gudzune KA, Jay M. Obesity management in adults: a review. JAMA 2023 Nov 28;330(20):2000-2015. [doi: 10.1001/jama.2023.19897] [Medline: 38015216]
6. Maciejewski ML, Smith VA, Berkowitz TSZ, et al. Association of bariatric surgical procedures with changes in unhealthy alcohol use among US veterans. JAMA Netw Open 2020 Dec 1;3(12):e2028117. [doi: 10.1001/jamanetworkopen.2020.28117] [Medline: 33346846]
7. Konttinen H, Sjöholm K, Jacobson P, Svensson PA, Carlsson LMS, Peltonen M. Prediction of suicide and nonfatal self-harm after bariatric surgery: a risk score based on sociodemographic factors, lifestyle behavior, and mental health: a nonrandomized controlled trial. Ann Surg 2021 Aug 1;274(2):339-345. [doi: 10.1097/SLA.0000000000003742] [Medline: 31850987]
8. Ard JD, Lewis KH, Moore JB. Lifestyle interventions for obesity in the era of GLP-1 receptor agonists. JAMA 2024 Jul 2;332(1):16-18. [doi: 10.1001/jama.2024.7062] [Medline: 38744427]
9. Dalle Grave R, Sartirana M, El Ghoch M. Treatment overview. In: Treating Obesity with Personalized Cognitive Behavioral Therapy: Springer; 2018. [doi: 10.1007/978-3-319-91497-8]
10. Dalle Grave R, Sartirana M, Calugi S. Personalized cognitive-behavioural therapy for obesity (CBT-OB): theory, strategies and procedures. Biopsychosoc Med 2020;14:5. [doi: 10.1186/s13030-020-00177-9] [Medline: 32175002]
11. Reilly MP, Lattal KA. Within-session delay-of-reinforcement gradients. J Exp Anal Behav 2004 Jul;82(1):21-35. [doi: 10.1901/jeab.2004.82-21] [Medline: 15484869]
12. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39. [doi: 10.1348/135910709X466063] [Medline: 19646331]
13. Kim J, Marcusson-Clavertz D, Yoshiuchi K, Smyth JM. Potential benefits of integrating ecological momentary assessment data into mHealth care systems. Biopsychosoc Med 2019;13:19. [doi: 10.1186/s13030-019-0160-5] [Medline: 31413726]
14. Agarwal DK, Chen BC. Statistical Methods for Recommender Systems: Cambridge University Press; 2016.
15. Rabbi M, Pfammatter A, Zhang M, Spring B, Choudhury T. Automated personalized feedback for physical activity and dietary behavior change with mobile phones: a randomized controlled trial on adults. JMIR Mhealth Uhealth 2015 May 14;3(2):e42. [doi: 10.2196/mhealth.4160] [Medline: 25977197]
16. Kupila SKE, Joki A, Suojanen LU, Pietiläinen KH. The effectiveness of eHealth interventions for weight loss and weight loss maintenance in adults with overweight or obesity: a systematic review of systematic reviews. Curr Obes Rep 2023 Sep;12(3):371-394. [doi: 10.1007/s13679-023-00515-2] [Medline: 37354334]
17. Bond DS, Thomas JG, Raynor HA, et al. B-MOBILE--a smartphone-based intervention to reduce sedentary time in overweight/obese individuals: a within-subjects experimental trial. PLOS ONE 2014;9(6):e100821. [doi: 10.1371/journal.pone.0100821] [Medline: 24964010]
18. Finkelstein J, Bedra M, Li X, Wood J, Ouyang P. Mobile app to reduce inactivity in sedentary overweight women. Stud Health Technol Inform 2015;216:89-92. [doi: 10.3233/978-1-61499-564-7-89] [Medline: 26262016]
19. Forman EM, Goldstein SP, Crochiere RJ, et al. Randomized controlled trial of OnTrack, a just-in-time adaptive intervention designed to enhance weight loss. Transl Behav Med 2019 Nov 25;9(6):989-1001. [doi: 10.1093/tbm/ibz137] [Medline: 31602471]
20. Klasnja P, Rosenberg DE, Zhou J, Anau J, Gupta A, Arterburn DE. A quality-improvement optimization pilot of BariFit, a mobile health intervention to promote physical activity after bariatric surgery. Transl Behav Med 2021 Mar 16;11(2):530-539. [doi: 10.1093/tbm/ibaa040] [Medline: 32421187]

XSL•FO
RenderX

21. Bond DS, Thomas JG, Ryder BA, Vithiananthan S, Pohl D, Wing RR. Ecological momentary assessment of the relationship between intention and physical activity behavior in bariatric surgery patients. Int J Behav Med 2013 Mar;20(1):82-87. [doi: 10.1007/s12529-011-9214-1] [Medline: 22203518]
22. Goldschmidt AB, Crosby RD, Engel SG, et al. Affect and eating behavior in obese adults with and without elevated depression symptoms. Int J Eat Disord 2014 Apr;47(3):281-286. [doi: 10.1002/eat.22188] [Medline: 24014067]
23. Goldschmidt AB, Crosby RD, Cao L, et al. Ecological momentary assessment of eating episodes in obese adults. Psychosom Med 2014;76(9):747-752. [doi: 10.1097/PSY.0000000000000108] [Medline: 25373891]
24. Williams-Kerver GA, Steffen KJ, Smith KE, Cao L, Crosby RD, Engel SG. Negative affect and loss of control eating among bariatric surgery patients: an ecological momentary assessment pilot investigation. Obes Surg 2020 Jun;30(6):2382-2387. [doi: 10.1007/s11695-020-04503-6] [Medline: 32125646]
25. Fogg BJ. A behavior model for persuasive design. Presented at: Proceedings of the 4th International Conference on Persuasive Technology; Apr 26-29, 2009. [doi: 10.1145/1541948.1541999]
26. Home page. DeployGate. URL: https://deploygate.com [accessed 2025-05-03]
27. Ogawa W, Hirota Y, Miyazaki S, et al. Definition, criteria, and core concepts of guidelines for the management of obesity disease in Japan. Endocr J 2024 Mar 28;71(3):223-231. [doi: 10.1507/endocrj.EJ23-0593] [Medline: 38123337]
28. Murase N, Katsumura T, Ueda C, Inoue S, Shimomitsu T. Validity and reliability of Japanese version of international physical activity questionnaire. J Health Welf Stat 2002;49(11):1-9 [FREE Full text]
29. Kobayashi S, Honda S, Murakami K, et al. Both comprehensive and brief self-administered diet history questionnaires satisfactorily rank nutrient intakes in Japanese adults. J Epidemiol 2012;22(2):151-159. [doi: 10.2188/jea.je20110075] [Medline: 22343326]
30. Fukui I. The depression and anxiety mood scale (DAMS) – scale development and validation. Jpn J Behav Ther 1997;23:83-93. [doi: 10.24468/jjbt.23.2_83]
31. HN-300T2. OMRON. URL: https://web.archive.org/web/20240727121219/https://www.healthcare.omron.co.jp/product/hbf/hn-300t2.html [accessed 2026-2-13]
32. Thomas JG, Wing RR. Health-e-call, a smartphone-assisted behavioral obesity treatment: pilot study. JMIR Mhealth Uhealth 2013 Apr 17;1(1):e3. [doi: 10.2196/mhealth.2164] [Medline: 25100672]
33. Inada S, Yoshiuchi K, Iizuka Y, et al. Pilot study for the development of a self-care system for type 2 diabetes patients using a personal digital assistant (PDA). Int J Behav Med 2016 Jun;23(3):295-299. [doi: 10.1007/s12529-016-9535-1] [Medline: 26780633]
34. Zheng Y, Klem ML, Sereika SM, Danford CA, Ewing LJ, Burke LE. Self-weighing in weight management: a systematic literature review. Obesity (Silver Spring) 2015 Feb;23(2):256-265. [doi: 10.1002/oby.20946] [Medline: 25521523]
35. Lauti M, Kularatna M, Hill AG, MacCormick AD. Weight regain following sleeve gastrectomy-a systematic review. Obes Surg 2016 Jun;26(6):1326-1334. [doi: 10.1007/s11695-016-2152-x] [Medline: 27048439]
36. Kurisu K, DeAngelis BN, Yoshiuchi K, al'Absi M. Mediating role of delay discounting in the link between depressive symptoms and diabetes onset: findings from a prospective survey of a community sample. Int J Behav Med 2025 Oct;32(5):678-686. [doi: 10.1007/s12529-025-10354-1] [Medline: 40011362]
37. Strine TW, Mokdad AH, Dube SR, et al. The association of depression and anxiety with obesity and unhealthy behaviors among community-dwelling US adults. Gen Hosp Psychiatry 2008;30(2):127-137. [doi: 10.1016/j.genhosppsych.2007.12.008] [Medline: 18291294]
38. Kirkpatrick SI, Subar AF, Douglass D, et al. Performance of the automated self-administered 24-hour recall relative to a measure of true intakes and to an interviewer-administered 24-h recall. Am J Clin Nutr 2014 Jul;100(1):233-240. [doi: 10.3945/ajcn.114.083238] [Medline: 24787491]
39. Kurisu K, Hashimoto M, Ishizawa T, et al. Development of computer adaptive testing for measuring depression in patients with cancer. Sci Rep 2022 May 17;12(1):8247. [doi: 10.1038/s41598-022-12318-x] [Medline: 35581321]
40. Horie T, Kurisu K, Inada S, et al. Development of computer adaptive tests to assess the psychological status of individuals with an eating disorder or type 2 diabetes. Biopsychosoc Med 2025 Feb 14;19(1):2. [doi: 10.1186/s13030-025-00325-z] [Medline: 39953552]

Abbreviations

BDHQ: brief-type self-administered diet history questionnaire
EMI: ecological momentary intervention
IPAQ: International Physical Activity Questionnaire

XSL•FO
RenderX

Edited by A Kushniruk; submitted 11.May.2025; peer-reviewed by NS Akhavan, S Kath; revised version received 06.Jan.2026; accepted 13.Jan.2026; published 26.Feb.2026.

Please cite as:
Kurisu K, Yamamoto Y, Aoyama T, Yamauchi T, Yoshiuchi K
Smartphone App Using Reinforcement Learning for Obesity: Single-Arm Feasibility Study
JMIR Hum Factors 2026;13:e77323
URL: https://humanfactors.jmir.org/2026/1/e77323
doi:10.2196/77323

© Ken Kurisu, Yoshiharu Yamamoto, Tomohisa Aoyama, Toshimasa Yamauchi, Kazuhiro Yoshiuchi. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 26.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Postural Education in School-Aged Populations: Development and Usability Evaluation of a Mobile Biofeedback App (EduBack)

Marco A García-Luna[1], PhD; Miguel García-Jaén[1], PhD; Daniel Ruiz-Fernández[2], PhD; Carmen Manchado[1], PhD; Juan M Cortell-Tormo[1], PhD

[1]Department of General and Specific Didactics, Faculty of Education, University of Alicante, C/ Aeroplano, s/n, San Vicente del Raspeig, Spain
[2]Department of Computer Science and Technology, University of Alicante, San Vicente del Raspeig, Spain

Corresponding Author:
Marco A García-Luna, PhD
Department of General and Specific Didactics, Faculty of Education, University of Alicante, C/ Aeroplano, s/n, San Vicente del Raspeig, Spain

Abstract

Background: Postural education is crucial during childhood and adolescence, yet traditional approaches often lack engaging tools that promote awareness and behavioral change. Wearable technologies and real-time biofeedback systems offer new opportunities to support postural learning through immediate, embodied feedback. However, most existing systems focus on clinical rehabilitation, with few designed specifically for educational use.

Objective: This study aimed to design, develop, and evaluate the usability and technical performance of EduBack, a mobile app that delivers real-time lumbar posture biofeedback through inertial sensors, with a specific focus on educational settings such as schools and physical education environments.

Methods: EduBack was developed using Kotlin (JetBrains) for Android OS (Google; version 8.0 and above) and integrates with 2 inertial measurement units via Bluetooth (2.4 GHz). The app provides visual biofeedback through a dynamic interface showing a virtual spine, corrective messages, and a color-coded alignment bar. The usability evaluation involved 24 undergraduate students (mean age 21.4, SD 1.8 y) who used the app in a controlled session. Participants completed the system usability scale and open-ended qualitative feedback questions. Technical performance data were collected from system logs, latency measurements, and received signal strength indicator values to assess connection stability and sensor-to-app communication.

Results: The average system usability scale score was 83.5 (SD 8.7), indicating excellent usability. Participants reported the interface to be intuitive, the biofeedback visualization clear, and the posture information easy to interpret. Qualitative responses highlighted the app's potential to support postural awareness and motor learning, especially in school-aged populations. From a technical perspective, the system demonstrated robust performance: mean data transfer latency was approximately 120 milliseconds, with less than 1% packet loss across sessions. Received signal strength indicator values consistently remained within the optimal signal range, confirming stable Bluetooth connectivity. All session data were successfully stored and exported without errors. The real-time posture tracking displayed on the app closely matched raw sensor data, ensuring fidelity in feedback.

Conclusions: EduBack is a usable and technically stable mobile app designed to support postural education through wearable sensors and real-time biofeedback. Its user-friendly interface and reliable data transmission make it well-suited for use in schools and educational programs targeting postural health. The app fills a gap in the mobile health field by offering a preventive, educational tool rather than a clinical one. Future research should explore its application in younger populations, integration into physical education curricula, and long-term effects on postural behavior and motor skill acquisition.

(JMIR Hum Factors 2026;13:e79282) doi:10.2196/79282

KEYWORDS

postural education; wearable technology; inertial sensors; mobile health; biofeedback; usability evaluation; motor learning; app development; mobile health innovation

Introduction

Musculoskeletal disorders represent one of the most prevalent health problems across different age groups and populations and are currently recognized as a major global public health concern [1,2]. Among these, low back pain (LBP) is particularly significant, as it is one of the most widespread conditions [3,4], surpassing complaints in other spinal regions such as the cervical or thoracic spine [4,5]. Its global incidence has increased by approximately 55% over the last 3 decades, making it the leading cause of work-related disability today [3-9]. Current estimates suggest that by 2050, nearly 850 million people worldwide will

XSL•FO
RenderX

suffer from LBP, representing a 36% increase from 2020 levels [4].

Unfortunately, children and adolescents are not exempt from this musculoskeletal condition [10,11]. The lifetime prevalence of LBP by ages 10 - 12 years is approximately 20%, increasing significantly during adolescence and reaching rates comparable to adults—up to 70% by late adolescence [12,13]. Several studies have highlighted the upward trend in LBP prevalence among adolescents over the last 2 decades [1,14,15].

LBP is considered a multifactorial syndrome, influenced by a range of biological, biomechanical, psychological, emotional, behavioral, and environmental factors [16-21]. Among the modifiable risk factors, poor postural hygiene, obesity, improper lifting techniques, prolonged sedentary behavior, and low levels of daily physical activity have been identified [5,7,16,20-23]. Although multidisciplinary approaches are often recommended for LBP prevention and treatment [24], there is broad consensus that physical exercise combined with posture-focused educational programs is one of the most effective strategies for reducing both the incidence and severity of LBP [3,25-27].

Numerous researchers and international organizations have advocated for posture education and its integration into school-based programs as a strategy to prevent both present and future LBP [28-39]. This may be due to the strong potential of educational institutions to equip children and adolescents with the knowledge and skills necessary to preserve and improve their health [39,40]. Indeed, one of the most powerful tools to prevent back pain—or at least reduce its frequency and severity—is the implementation of educational interventions that promote long-term acquisition of postural knowledge and healthy habits [39,41].

Technology plays a pivotal role in modern education and has demonstrated clear benefits in enhancing student learning outcomes [42-44]. In this context, biofeedback has emerged as an efficient and effective tool for enhancing postural awareness and promoting corrective behavior in diverse activities and settings [45-48]. Biofeedback is a technique that monitors and quantifies physiological or biomechanical parameters and converts them into real-time visual, auditory, or tactile signals [49,50]. Technological advances have enabled the integration of biofeedback into wearable systems, which offer advantages such as portability, lightweight, low cost, and energy efficiency [51-53].

Despite these advances, wearable tools specifically designed to provide postural biofeedback in educational settings remain limited [45,54]. This scarcity is largely due to the recent emergence of such technologies, as well as the lack of published studies evaluating their implementation and effectiveness in schools. Although some posture-monitoring and correction devices exist, most are designed for clinical or sports use and are poorly adapted for educational contexts. As a result, further research is needed to explore the role of these technologies in teaching and learning healthy postural habits. In particular, such research is envisioned for children and adolescents in both primary and secondary school settings, where early intervention is most relevant for long-term musculoskeletal health.

Therefore, the objective of this study was to develop and preliminarily evaluate EduBack, a mobile app based on inertial sensors and real-time biofeedback. Specifically, this study aimed to assess the usability and technical performance of the system as a first step toward its long-term educational goal of supporting the learning and internalization of correct lumbar posture in school settings. The study followed a descriptive technological development design with an embedded user evaluation, combining the development of the EduBack system with a preliminary assessment of its usability and technical reliability.

Methods

Study Design

This study followed a cross-sectional descriptive technological development design with an embedded user evaluation (mixed methods approach). Detailed methods are presented in the following subsections.

Participants

Although the long-term target population of EduBack is children and adolescents, university students were intentionally selected for this preliminary evaluation to ensure feasibility and technical stability before adapting the system to younger users. Therefore, a convenience sample of 24 undergraduate students (mean age: 21.4, SD 1.8 y; n=10, 42% female participants) from the Faculty of Education was recruited. Inclusion criteria were: (i) no known musculoskeletal or neurological conditions affecting posture and (ii) basic familiarity with smartphone use.

Ethical Considerations

All participants provided written informed consent after receiving detailed information about the study procedures, potential risks, and their right to withdraw at any time without penalty. This study was approved by the Ethics Committee of the University of Alicante (reference number UA-2023-11-16) and conducted in accordance with institutional guidelines and the principles of the Declaration of Helsinki.

All data were anonymized prior to analysis, and no identifying information was collected or stored. Privacy and confidentiality were ensured throughout the study, with data accessible only to the research team. Participants did not receive financial or material compensation for their involvement. The sample consisted of adult undergraduate students, who are not considered a vulnerable population.

Tools and Instruments

System Conceptualization

Based on a review of the literature on wearable tools for postural education [51-53], design principles for educational systems aimed at postural training [36,40,55], usability evaluations of similar health technologies [56,57] and user reviews of commercially available mobile apps for postural correction (Google Play and Apple App Store), 3 core features were defined to build an effective mobile educational system for postural hygiene and training: (i) real-time biofeedback of lumbar posture using inertial sensors, providing immediate visual feedback to raise awareness and enable posture correction;

XSL•FO
RenderX

(ii) local data recording and storage for long-term postural monitoring; and (iii) personalized recommendations based on the user's postural patterns, including suggestions, adaptations, or corrective cues.

Furthermore, for the system to be suitable for long-term educational use, we established the following usability requirements: (i) the interface must be intuitive and user-friendly, requiring no prior technical knowledge from students and (ii) both data collection and postural feedback must be automated and seamlessly integrated into activities to avoid interfering with the learning process.

App Development

During the initial development phase of the EduBack system, a list of key parameters for real-time postural monitoring and biofeedback was formulated. Due to the lack of existing solutions that met our specifications—particularly apps offering intuitive and real-time visualization of lumbar posture in educational settings—we developed a dedicated mobile app to effectively collect and display this information.

Technically, the EduBack system is composed of two primary components: (i) a mobile app designed for local data collection and real-time visual feedback, enabling students to autonomously and effectively monitor their posture and (ii) an inertial measurement unit (IMU) system, consisting of 2 IMU devices placed on the user's lumbar spine. These devices connect via Bluetooth to a smartphone or tablet running the EduBack app and are responsible for capturing data on spinal alignment and movement.

In addition, the EduBack app integrates a computer vision module (MoveNet) to provide a simplified visual representation of the user's skeleton. This representation is based on 17 anatomical reference points (head, shoulders, elbows, wrists, hips, knees, and ankles), detected in real time through the smartphone or tablet camera and displayed using Canvas. The purpose of this visualization is to complement the inertial sensor data with intuitive feedback, enhancing user awareness of posture during educational tasks. Importantly, this skeletal visualization was intentionally designed as a set of abstract points rather than a full video image, in order to respect user privacy. This approach is particularly relevant for the intended target population of children and adolescents, as it allows external recording of postural tasks by another person without displaying identifiable visual features of the participant.

Overview of the EduBack System

EduBack is a mobile app developed using Kotlin for devices running Android OS 8 and above. It features three main screens: Home, List, and Detail. The Home screen basically displays: (i) the sensor status (connection, inclination, etc); (ii) a real-time visual representation of 17 points, based on Canvas and MoveNet, showing the human skeleton detected by the smartphone or tablet camera; and (iii) the integrated biofeedback that includes a spinal visualization with associated messages, a color bar, and a dynamic dot that moves along the bar according to the user's posture. The List screen provides an overview of recorded sessions. In the Detail screen, users can view ratings and key details for each recorded session. Figure 1 displays the 3 main screens of the app, while Figure 2 illustrates some graphical representations of the EduBack app's functionality and its biofeedback, such as when lifting a heavy object from the ground.

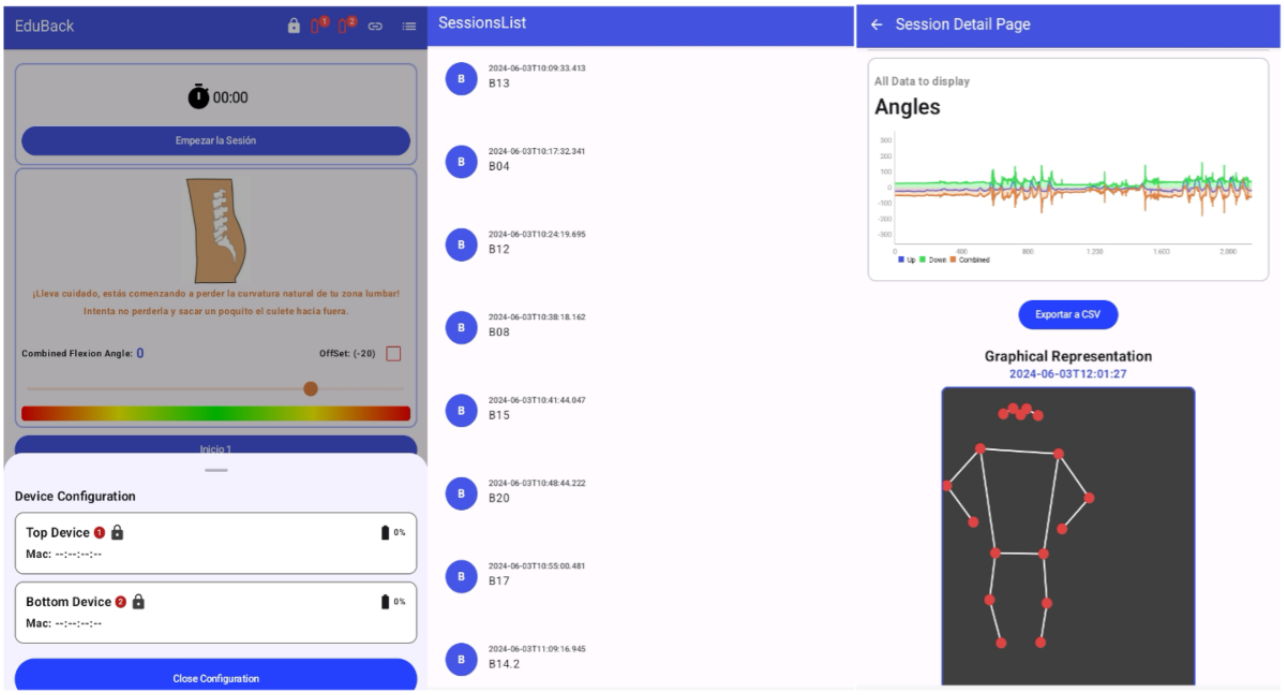
XSL•FO
RenderX

Figure 1. Illustrative examples of the EduBack app functioning and its biofeedback indicators for lumbar posture. Panel A shows the neutral lumbar lordosis, which refers to the natural inward curvature of the lumbar spine in standing posture. Panels B and C represent intermediate deviations from this neutral posture. Panel D illustrates lumbar hyperlordosis, defined as an excessive inward curvature of the lumbar spine. Panel E illustrates lumbar kyphosis, defined as an abnormal outward curvature of the lumbar spine in the lumbar region. Together, these examples demonstrate how the EduBack app provides real-time visual feedback on different postural states.



XSL•FO
RenderX

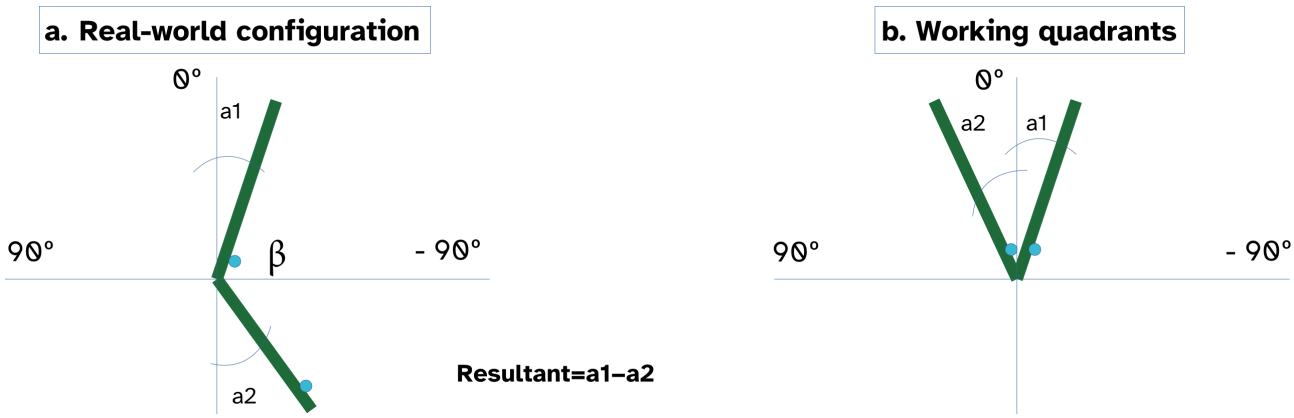
Figure 2. Main screens of the EduBack mobile app. The Home screen (left) displays the sensor status (connection and inclination), a real-time skeleton visualization based on 17 anatomical points detected by the device camera, and integrated biofeedback including a spinal visualization, color bar, and dynamic dot indicating posture. The List screen (center) provides an overview of recorded sessions. The Detail screen (right) shows ratings and key information for each recorded session, allowing users to review their performance and feedback.



The design of the EduBack interface was informed by principles of motor learning, instructional design, and behavior change. Systematic reviews and meta-analyses have consistently shown that color-coded visual cues and dynamic prompts enhance attentional focus, facilitate error correction, and support self-regulated learning in motor tasks [58,59]. Similarly, systematic reviews in instructional design highlight the pedagogical value of visual displays for improving comprehension and engagement [60]. Finally, reviews of digital behavior change interventions confirm that immediate, intuitive feedback is a key driver of adherence and habit formation [61,62]. These findings provide a strong theoretical and empirical foundation for the interface design choices implemented in EduBack.

Figure 3 illustrates how the lumbar angle was estimated from the inclinations of the 2 IMUs placed at L1 and S1. Each sensor provided a flexion angle relative to the vertical axis (A1 and A2). In practice, these angles are calculated in opposite quadrants, with opposite signs. By subtracting them, the resultant angle corresponds to the relative inclination between the 2 sensors, with the sign indicating the orientation direction. The lumbar angle β was then obtained as 180 minus this resultant value. This computational approach has been previously validated in García-Luna et al [63], where IMU-derived pelvic and lumbar tilt angles showed high agreement with a 3D optical motion capture system, considered the gold standard for kinematic assessment.

Figure 3. Representation of the lumbar angle calculation using 2 inertial measurement units. Each sensor (placed at L1 and S1) provides an inclination relative to the vertical axis. By combining these values, the relative lumbar angle (β) is obtained, with the sign indicating orientation. Blue points represent the inertial measurement unit LEDs, showing sensor orientation.



XSL•FO
RenderX

Data Processing Software

The mobile app connects to the inertial sensors using Bluetooth Low Energy, a protocol that supports low-latency communication. Once paired, the device transmits data via notifications, which are received by the smartphone. Internally, the data—originally in byte format—is decoded using bitwise operations and mathematical transformations to convert hexadecimal values into decimal format, allowing for accurate on-screen visualization.

While receiving the data, the app stores it in a local database to enable later retrieval and export in CSV format. This export process transforms the locally stored tabular data into rows and columns, a common practice in this type of system. Storage capacity depends on the available device memory; however, modern smartphones typically do not present limitations in this regard. The app uses a secure local database implemented with Room (Android persistence library) to ensure reliable record keeping.

At the beginning of each session, the user must connect the sensors, triggering real-time data transmission and visualization. During the session, data is temporarily stored in cache memory; once the session ends, the data is transferred to the local database. This process ensures data persistence and prevents information loss if the app is restarted.

Inertial Sensors

The inertial sensors used were Ergotex IMUs (JVTech Solutions), designed specifically for spinal posture monitoring. Each device operates at a frequency of 20 Hz and is equipped with a triaxial gyroscope (±1000 deg/s) and triaxial accelerometer (±2 g), enclosed in a compact unit measuring 21×10 mm and weighing only 8 grams. These sensors record acceleration across 3 axes and transmit data in real time via Bluetooth (2.4 GHz) to a mobile device running the EduBack app.

To ensure optimal performance, each Ergotex IMU integrates the ICM-20602 MEMS MotionTracking sensor (TDK Corp), known for its high accuracy and reliability. This sensor offers a gyroscope sensitivity error within ±1% and gyroscope noise levels of ±4 mdeg/s/√Hz, while accelerometer noise is maintained at 100 µg/√Hz. Additionally, its integration with a 1K-byte first-in, first-out buffer reduces serial bus traffic, enhancing measurement stability and energy efficiency.

The system includes a battery connected to a regulated power supply to ensure a stable voltage output. A microprocessor collects the raw data from the accelerometer and gyroscope through the serial peripheral interface bus and sends the processed information to a radio frequency module for wireless transmission. This system allows for efficient data flow and ensures the precision and reliability of postural data collection.

Importantly, the concurrent validity of these inertial sensors for estimating lumbar and pelvic tilt angles has been previously established against a 3D optical motion capture system, considered the gold standard for kinematic assessment. García-Luna et al [63] reported minimal systematic error (0.08° overall) and high agreement between IMU-derived angles and optical measurements, with correlation coefficients exceeding 0.90 and mean absolute errors around 1°. Bland-Altman analyses confirmed that most data points fell within ±2° limits of agreement, supporting the accuracy of the Ergotex IMUs for sagittal pelvic tilt measurement in both static and dynamic positions. These findings validate the suitability of the sensors for postural monitoring in educational contexts and justify their use in the present study.

Usability Assessment Instrument

To assess the perceived usability of the EduBack app, participants completed the Spanish version of the system usability scale (SUS) [64] immediately after the testing session. The SUS is a widely used, validated 10-item questionnaire that provides a global view of subjective usability [65,66]. Each item is scored on a 5-point Likert scale ranging from “strongly disagree” to “strongly agree.” The total score ranges from 0 to 100, with higher scores indicating better usability. The SUS was selected due to its brevity, ease of application, and frequent use in digital health evaluations, making it an appropriate benchmark for assessing mobile health-related apps [67-69].

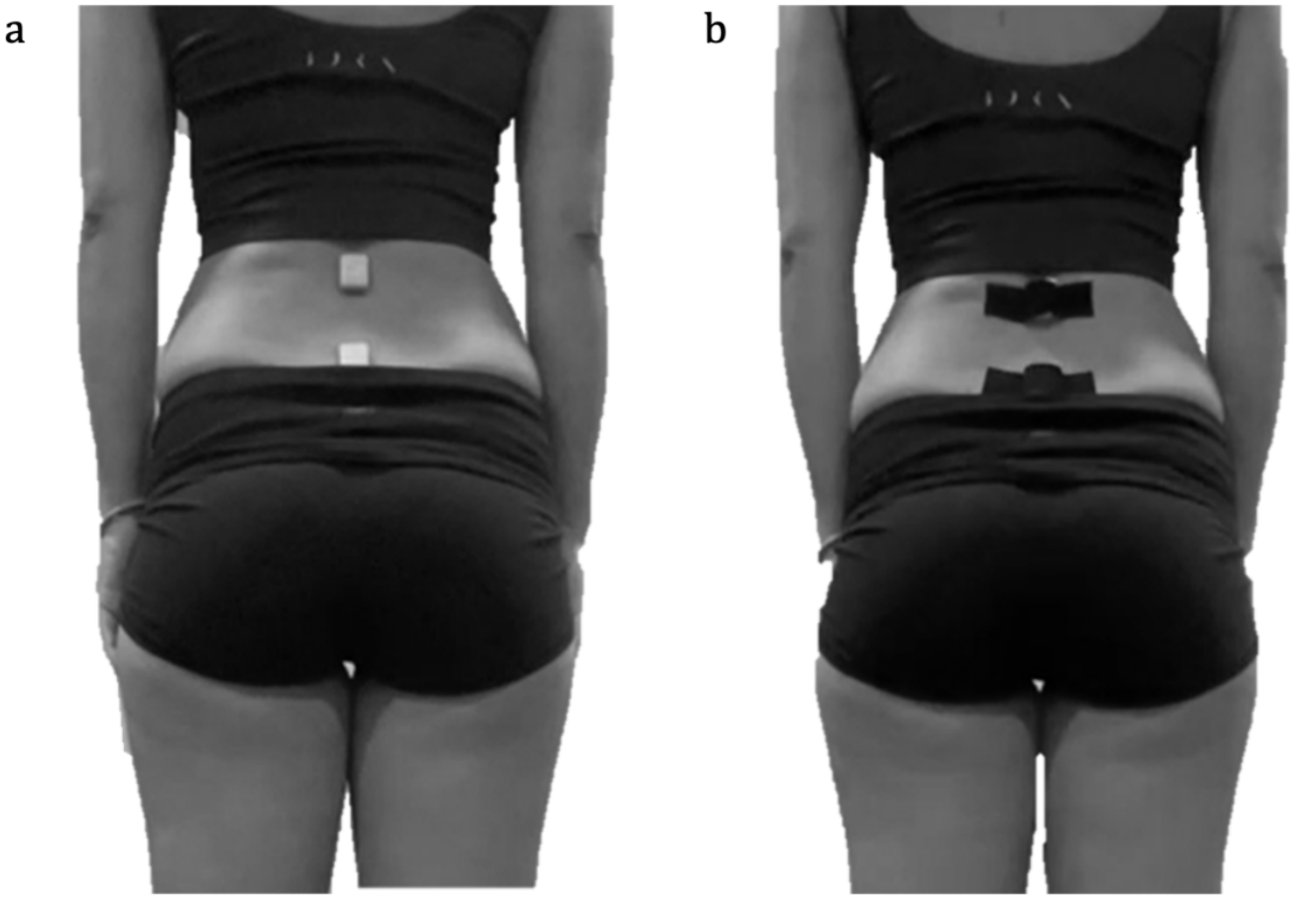
In addition to SUS, qualitative comments on educational value (biofeedback clarity, motor awareness, and school applicability) were collected post-session to complement usability scores.

Procedure

Participants were individually invited to a simulated laboratory setting replicating a postural education session. Sensor placement over the spinous processes of L1 and S1 (Figure 4) was performed directly by a trained researcher to ensure consistency and reliability of the instrumentation. Likewise, all instructions regarding the use of the EduBack app and the inertial sensors were provided orally by the researcher during the session, with no written guide, to guarantee standardized guidance across participants. Each participant then used the EduBack app on a standardized tablet provided in the laboratory during a guided 30-minute session that included a predefined sequence of postural tasks. The protocol included anterior and posterior pelvic tilts in both standing and seated positions, forward and backward trunk inclinations, picking up an object from the floor, and repeated sit-to-stand transitions from chairs of different heights, in addition to maintaining upright and seated postures. All participants performed the same tasks in the same order under standardized instructions. Sessions were conducted individually, with each participant completing the protocol one by one under controlled laboratory conditions. During the session, real-time feedback was displayed on the app. After completing the session, participants were asked to evaluate the app’s usability using a validated questionnaire. In addition, brief qualitative comments were collected regarding the perceived educational value of the biofeedback (clarity, motor awareness, and applicability in school contexts), complementing the quantitative usability scores.

XSL•FO
RenderX

Figure 4. . Placement of the inertial sensors over the spinous processes of L1 and S1. (A) Sensors were initially attached to the skin using hypoallergenic double-sided tape. (B) Sensors were subsequently secured with kinesio tape to ensure stability during the 30-minute postural task protocol.



Data Analysis

Descriptive statistics were used to analyze the usability data. SUS scores were calculated according to standard methods [65], and summarized using means, SDs, and score ranges. Technical performance data were collected through internal logs generated by the EduBack app, including latency (sensor-to-screen transfer time), Bluetooth connection stability, and data packet loss rate. Signal quality was additionally assessed using the Received Signal Strength Indicator, a standard metric for evaluating the strength of wireless signals. Based on predefined thresholds, signal quality was classified as optimal (−50 to −70 dBm), moderate (−70 to −80 dBm), or poor (below −93 dBm), with all sessions falling within the optimal range. These metrics were used to evaluate the robustness of the hardware-software integration and the reliability of real-time postural monitoring. Analyses were conducted using SPSS Statistics (version 27; IBM Corp). Qualitative feedback was collected in the form of brief participant comments at the end of each session and reviewed descriptively to complement the quantitative usability scores.

Results

System Usability

The usability of the EduBack app was assessed using the SUS (Spanish version), a widely validated instrument for evaluating user satisfaction and perceived system usability. The overall mean SUS score obtained was 83.5 (SD 8.7), which corresponds to an "excellent" rating according to established interpretation thresholds [66]. This result places the system well above the average usability benchmark of 68, indicating a high level of user satisfaction and ease of use.

A closer inspection of individual SUS items revealed that the system was particularly well-received in terms of learnability and confidence. For example: (1) 90% (18/20) of participants "strongly agreed" or "agreed" that they felt confident using the system; (2) 85% (17/20) considered that most people would learn to use the app very quickly; and (3) 90% (18/20) agreed that the app was easy to use and that its functionalities were well integrated.

Items related to system complexity and the need for technical support also received favorable evaluations: 90% (18/20) of users reported that they did not require technical assistance to use the app, and 85% (17/20) strongly disagreed with the statement that the system was unnecessarily complex.

Despite these high scores, certain aspects showed slightly more variability. While still favorable, the lowest-rated item was related to the consistency of the interface, where 70% (14/20) users "strongly agreed" or "agreed" that the various system functions were well integrated, and 20% (4/20) selected a neutral response. This suggests a potential area for further refinement in terms of user interface design consistency and alignment across different screens or tasks. No participants rated the system as cumbersome or difficult to use, and no significant usability issues were reported during the testing sessions.

XSL•FO
RenderX

A detailed breakdown of responses to each SUS item is provided in Table 1, which highlights item-level agreement rates and helps pinpoint specific strengths and areas for improvement.

Table . Participant responses to individual SUS[a] items (N=20).

SUS item	Statement (paraphrased)	Agree/strongly agree, n (%)	Neutral, n (%)	Disagree/strongly disagree, n (%)
1	I would use this system frequently	17 (85)	2 (10)	1 (5)
2	The system is unnecessarily complex	1 (5)	2 (10)	17 (85)
3	The system was easy to use	18 (90)	2 (10)	0 (0)
4	I would need help from a technical person	1 (5)	1 (5)	18 (90)
5	The functions were well integrated	14 (70)	4 (20)	2 (10)
6	The system was inconsistent	2 (10)	3 (15)	15 (75)
7	Most people would learn to use it quickly	17 (85)	2 (10)	1 (5)
8	The system is cumbersome to use	0 (0)	1 (5)	19 (95)
9	I felt confident using the system	18 (90)	2 (10)	0 (0)
10	I needed to learn a lot before using the system	1 (5)	2 (10)	17 (85)

[a]SUS: system usability scale.

Technical Performance

The Bluetooth connectivity between the IMU sensors and the EduBack app remained consistently stable throughout all test sessions. System performance logs indicated an average latency of 120 milliseconds for sensor-to-screen data transfer, ensuring a responsive user experience. The rate of data packet loss was minimal, remaining below 1% across all sessions, and no failures in data storage or export were recorded.

Received signal strength indicator values monitored during the sessions consistently remained within the optimal range, confirming stable wireless communication between the sensors and the mobile device as defined by technical thresholds.

Real-time lumbar posture tracking was accurately displayed within the app's biofeedback interface, as confirmed by synchronized screenshots and analysis of the raw data logs. These results support the technical reliability of the EduBack system for continuous, real-time postural monitoring in educational settings.

Qualitative Feedback

Qualitative feedback collected during the usability evaluation provided additional insights beyond the SUS scores. Participants highlighted that the real-time biofeedback display, with its color-coded indicators and dynamic corrective prompts, enhanced their understanding of lumbar movement and posture control. Several students emphasized that the app facilitated motor awareness and contributed to acquiring correct lumbar mobilization patterns, supporting functional learning and the transfer of healthy postural habits to everyday activities. Importantly, participants suggested that EduBack could be particularly beneficial in school settings, especially for children and adolescents, where early postural education is essential for prevention and long-term musculoskeletal health. These comments illustrate the perceived educational value of the system and its potential applicability in real-world contexts.

Discussion

Principal Findings

This study presents the development process (described in the Methods) and reports the preliminary evaluation of EduBack, an innovative mobile app designed to deliver real-time postural biofeedback through wearable IMU sensors. The primary goal of the present study was to evaluate usability and technical reliability, establishing feasibility for future educational apps. While the long-term aim of EduBack is to facilitate postural education and the internalization of correct lumbar posture, these educational outcomes were not directly assessed here and remain directions for future research.

System Usability

From a usability perspective, the results demonstrated that EduBack achieved high acceptance among participants, with SUS scores indicating strong usability. Feedback from undergraduate students was notably positive. Participants appreciated the intuitive interface, particularly the biofeedback visualization that included dynamic color bars, real-time vertebral tracking, and corrective prompts. These findings suggest that the interface design and feedback mechanisms are transferable to educational contexts, although future studies

XSL•FO
RenderX

should confirm this in younger populations. This type of immediate, embodied feedback appears to support postural self-awareness—a key aspect of motor learning—and may foster metacognitive reflection about posture, which is rarely achieved through traditional instruction alone. Unlike clinical systems where usability is often compromised by multisensor burden or complex interfaces, EduBack's simplicity may represent an advantage for adoption in schools. These outcomes align with prior findings on the effectiveness of biofeedback systems in physical rehabilitation and posture correction [45-48], but extend them by highlighting usability in an explicitly educational context.

Technical Performance

Regarding technical performance, the system proved reliable under real-time conditions. The results confirmed that the system performs with low latency and minimal data loss during sensor-to-screen transmission. Participants experienced stable Bluetooth connectivity and consistent visual feedback, which suggests that the system is well-suited for educational environments that demand real-time interaction and reliability. Additionally, session data were securely stored and successfully exported, which is essential for both instructional follow-up and research use. It should be noted that in this context, "reliability" refers to technical robustness rather than psychometric reliability of the instrument. Psychometric reliability (eg, test-retest or interrater consistency) was not assessed in this preliminary study and remains an important direction for future validation. Prior studies have reported frequent technical challenges in wearable systems (eg, unstable wireless connections and battery limitations), which often hinder long-term use [70,71]. By demonstrating stable connectivity and efficient data handling, EduBack addresses some of these recurring barriers, although further testing in uncontrolled environments (eg, crowded classrooms) and formal reliability analysis are needed to confirm its robustness and measurement consistency.

Qualitative Feedback

Qualitative feedback provided complementary insights into the educational value of EduBack. EduBack is distinguished from other wearable-based systems by its explicit focus on postural education rather than clinical intervention. Most prior research in this field has focused on rehabilitation in patients with chronic back pain, scoliosis, or musculoskeletal disorders [45,53,56]. Clinical and occupational apps of wearable posture systems commonly collect user-level information (eg, perceived usefulness, convenience, comfort, and practicality) alongside objective kinematic outcomes. Reviews consistently highlight real-time corrective feedback as promising, but also report heterogeneous and often limited documentation of user preferences and compliance, as well as recurring usability challenges (multisensor burden, attachment or removal, wearability over time, wireless stability, battery or power management) that impact acceptance and scalability. In contrast, the feedback gathered in the present study was explicitly educational: participants emphasized biofeedback clarity, enhanced motor awareness, and applicability in school settings—underscoring shared mechanisms (feedback-guided posture correction) but different aims (therapeutic adherence and symptom management versus prevention and movement literacy in educational contexts). This preliminary evidence of educational value, although not yet linked to measurable learning outcomes, highlights the potential of EduBack to bridge the gap between technological feasibility and pedagogical impact.

Building on this distinction, EduBack is designed for educational training, where the goal is not treatment but awareness, prevention, and acquisition of correct motor habits. This is particularly relevant in school-based educational contexts, where early intervention in postural habits can play a key role in long-term musculoskeletal health. By offering an engaging and interactive biofeedback experience, EduBack could facilitate motor learning of lumbar mobilization and may encourage the correct execution of daily movements from an early age. While the app may also support training in higher education or preprofessional programs as a secondary use case, its primary value lies in promoting postural awareness and self-regulation among children and adolescents in primary and secondary education settings.

Although EduBack was designed primarily for educational settings, several students highlighted its potential for use in clinical teaching environments, suggesting that the system could also help bridge the gap between academic learning and applied health care practice. The visualization of spinal posture, when combined with immediate feedback, may reinforce theoretical content from lectures and provide a meaningful learning experience. This aligns with pedagogical models that emphasize experiential learning and self-regulation in motor skill acquisition [72-74]. Future iterations of EduBack will incorporate gamification strategies (eg, rewards and interactive challenges) to increase engagement among younger users, particularly in primary and secondary school contexts. Such features are expected to enhance motivation, adherence, and the educational impact of postural training. Thus, qualitative insights not only confirm usability but also point toward the system's potential pedagogical relevance, which should be validated with objective educational outcomes in future trials.

Broader Applicability and Future Directions

Moreover, although this study did not include professional educators or clinicians as participants, the students' perspective already points toward broader applicability. Future studies could involve instructors, physical therapists, or school teachers to explore how the system can be integrated into formal curricula or used in clinical decision-making. There is growing interest in using wearable technology in both school-based [75] and outpatient rehabilitation [76] settings, and EduBack may offer a scalable, cost-effective option that meets usability standards in both domains.

In practical terms, EduBack could be integrated into physical education curricula through teacher-guided sessions, where students receive real-time feedback while performing standardized postural exercises. The system may also be incorporated into modules on ergonomics or injury prevention, allowing teachers to demonstrate correct lifting techniques or seated posture with immediate visual reinforcement. Adolescents could use the app independently under supervision, while

XSL•FO
RenderX

younger children would benefit from structured, teacher-mediated activities. These scenarios illustrate how EduBack can complement existing educational practices, providing an interactive tool that supports awareness, prevention, and the acquisition of healthy motor habits within the classroom environment.

EduBack was designed primarily for educational environments, with physical education teachers expected to guide its application. Adolescents may be able to use the system independently, while younger children would require teacher supervision. Clarifying these user profiles is essential to ensure appropriate implementation and will be a key focus of future research.

Future research should also focus on multisite evaluations involving diverse populations and formal educational settings, possibly incorporating mixed method approaches to assess user engagement, learning outcomes, and behavioral adherence. It would also be beneficial to test EduBack in younger populations (eg, primary or secondary school students) and in clinical training contexts, where postural education is part of professional development.

Limitations

This study has several limitations that should be considered when interpreting the findings. First, the sample size was small and limited to university students, which may affect generalizability and highlights the need to test EduBack in younger and more diverse populations. Second, technical performance was evaluated in controlled academic settings; future studies should assess robustness in less stable environments such as classrooms or homes. Third, while the SUS provided a validated measure of usability, future studies could complement it with domain-specific instruments, such as the mobile health App Usability Questionnaire, to capture nuances related to health behavior change and long-term adherence. Finally, the study focused on short-term interactions; longitudinal research is needed to evaluate sustained use, user satisfaction, and potential impacts on posture-related outcomes.

Conclusions

Beyond the usability and technical outcomes, a key contribution of this study is the creation of EduBack itself, a novel educational tool that integrates wearable sensors and real-time biofeedback for postural training. EduBack demonstrates strong usability and technical reliability as a tool for posture-related education. It fills an important gap in the current landscape of mobile health apps by providing real-time, interactive biofeedback specifically tailored for learning rather than therapy. The system offers a promising foundation for future educational interventions in schools, universities, and clinical environments aimed at promoting postural health, awareness, and motor competence.

Acknowledgments

The authors thank all participants for their involvement in the study.

Funding

This work was primarily supported by a predoctoral fellowship awarded to MAGL by the Ministry of Universities of Spain (reference: FPU20/01843). Additional funding was provided by 2 project grants from the Valencian Innovation Agency of Spain (references: INNVA1/2020/81 and INNVA1/2022/34), which contributed to the development of the EduBack app. The funding bodies had no role in the study design, data collection, analysis, interpretation, or manuscript preparation.

Authors' Contributions

Conceptualization: MAGL (lead), JMCT (supporting)
Data curation: MAGL (lead)
Formal analysis: MAGL (lead)
Funding acquisition: MAGL, JMCT
Investigation: MAGL (lead)
Methodology: MAGL (lead), JMCT (supporting)
Project administration: MAGL (lead), JMCT (supporting)
Resources: JMCT (supporting)
Software: MAGL (lead)
Supervision: JMCT (lead), CM (supporting)
Validation: MAGL (lead)
Visualization: MAGL (lead)
Writing – original draft: MAGL (lead)
Writing – review & editing: MAGL (lead), JMCT (supporting), CM (supporting), DRF (supporting), MGJ (supporting)

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. Safiri S, Kolahi AA, Cross M, et al. Prevalence, deaths, and disability-adjusted life years due to musculoskeletal disorders for 195 countries and territories 1990-2017. Arthritis Rheumatol 2021 Apr;73(4):702-714. [doi: 10.1002/art.41571] [Medline: 33150702]
2. Jin Z, Wang D, Zhang H, et al. Incidence trend of five common musculoskeletal disorders from 1990 to 2017 at the global, regional and national level: results from the Global Burden of Disease Study 2017. Ann Rheum Dis 2020 Aug;79(8):1014-1022. [doi: 10.1136/annrheumdis-2020-217050] [Medline: 32414807]
3. Chen S, Chen M, Wu X, et al. Global, regional and national burden of low back pain 1990-2019: a systematic analysis of the Global Burden of Disease Study 2019. J Orthop Translat 2022 Jan;32(1):49-58. [doi: 10.1016/j.jot.2021.07.005] [Medline: 34934626]
4. Ferreira ML, de Luca K, Haile L. Global, regional, and national burden of low back pain, 1990-2020, its attributable risk factors, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. Lancet Rheumatol 2023 Jun;5(6):e316-e329. [doi: 10.1016/S2665-9913(23)00098-X] [Medline: 37273833]
5. Hartvigsen J, Hancock MJ, Kongsted A, et al. What low back pain is and why we need to pay attention. Lancet 2018 Jun 9;391(10137):2356-2367. [doi: 10.1016/S0140-6736(18)30480-X] [Medline: 29573870]
6. Hurwitz EL, Randhawa K, Yu H, Côté P, Haldeman S. The global spine care initiative: a summary of the global burden of low back and neck pain studies. Eur Spine J 2018 Sep;27(Suppl 6):796-801. [doi: 10.1007/s00586-017-5432-9] [Medline: 29480409]
7. Buchbinder R, Underwood M, Hartvigsen J, Maher CG. The Lancet Series call to action to reduce low value care for low back pain: an update. Pain 2020 Sep;161 Suppl 1(1):S57-S64. [doi: 10.1097/j.pain.0000000000001869] [Medline: 33090740]
8. Wang L, Ye H, Li Z, et al. Epidemiological trends of low back pain at the global, regional, and national levels. Eur Spine J 2022 Apr;31(4):953-962. [doi: 10.1007/s00586-022-07133-x] [Medline: 35217914]
9. Lo J, Chan L, Flynn S. A systematic review of the incidence, prevalence, costs, and activity and work limitations of amputation, osteoarthritis, rheumatoid arthritis, back pain, multiple sclerosis, spinal cord injury, stroke, and traumatic brain injury in the United States: a 2019 update. Arch Phys Med Rehabil 2021 Jan;102(1):115-131. [doi: 10.1016/j.apmr.2020.04.001] [Medline: 32339483]
10. Frosch M, Mauritz MD, Bielack S, et al. Etiology, risk factors, and diagnosis of back pain in children and adolescents: evidence- and consensus-based interdisciplinary recommendations. Children (Basel) 2022 Feb 2;9(2):192-205. [doi: 10.3390/children9020192] [Medline: 35204913]
11. Swain MS, Henschke N, Kamper SJ, Gobina I, Ottová-Jordan V, Maher CG. An international survey of pain in adolescents. BMC Public Health 2014 May 13;14(1):447. [doi: 10.1186/1471-2458-14-447] [Medline: 24885027]
12. Kamper SJ, Yamato TP, Williams CM. The prevalence, risk factors, prognosis and treatment for back pain in children and adolescents: an overview of systematic reviews. Best Pract Res Clin Rheumatol 2016 Dec;30(6):1021-1036. [doi: 10.1016/j.berh.2017.04.003] [Medline: 29103547]
13. Calvo-Muñoz I, Gómez-Conesa A, Sánchez-Meca J. Prevalence of low back pain in children and adolescents: a meta-analysis. BMC Pediatr 2013 Jan 26;13:14. [doi: 10.1186/1471-2431-13-14] [Medline: 23351394]
14. Roy R, Galán S, Sánchez-Rodríguez E, et al. Cross-national trends of chronic back pain in adolescents: results from the HBSC study, 2001-2014. J Pain 2022 Jan;23(1):123-130. [doi: 10.1016/j.jpain.2021.07.002] [Medline: 34339858]
15. Wu A, March L, Zheng X, et al. Global low back pain prevalence and years lived with disability from 1990 to 2017: estimates from the Global Burden of Disease Study 2017. Ann Transl Med 2020 Mar;8(6):299. [doi: 10.21037/atm.2020.02.175] [Medline: 32355743]
16. Noll M, Candotti CT, da Rosa BN, Vieira A, Loss JF. Back pain and its risk factors in Brazilian adolescents: a longitudinal study. Br J Pain 2021 Feb;15(1):16-25. [doi: 10.1177/2049463719871751] [Medline: 33633850]
17. Miñana-Signes V, Monfort-Pañego M, Bosh-Bivià AH, Noll M. Prevalence of low back pain among primary school students from the city of Valencia (Spain). Healthcare (Basel) 2021 Mar 3;9(3):270. [doi: 10.3390/healthcare9030270] [Medline: 33802400]
18. Miller MM, Meints SM, Hirsh AT. Catastrophizing, pain, and functional outcomes for children with chronic pain: a meta-analytic review. Pain 2018 Dec;159(12):2442-2460. [doi: 10.1097/j.pain.0000000000001342] [Medline: 30015710]
19. Sá S, Silva AG. Repositioning error, pressure pain threshold, catastrophizing and anxiety in adolescents with chronic idiopathic neck pain. Musculoskelet Sci Pract 2017 Aug;30:18-24. [doi: 10.1016/j.msksp.2017.04.011] [Medline: 28494262]
20. Noll M, Candotti CT, Rosa BND, Loss JF. Back pain prevalence and associated factors in children and adolescents: an epidemiological population study. Rev Saude Publica 2016 Jun 10;50:31. [doi: 10.1590/S1518-8787.2016050006175] [Medline: 27305406]
21. Rosa BND, Noll M, Candotti CT, Loss JF. Risk factors for back pain among southern Brazilian school children: a 6-year prospective cohort study. Int J Environ Res Public Health 2022 Jul 7;19(14):8322. [doi: 10.3390/ijerph19148322] [Medline: 35886179]
22. Mahdavi SB, Riahi R, Vahdatpour B, Kelishadi R. Association between sedentary behavior and low back pain: a systematic review and meta-analysis. Health Promot Perspect 2021;11(4):393-410. [doi: 10.34172/hpp.2021.50] [Medline: 35079583]

XSL•FO
RenderX

23. Bontrup C, Taylor WR, Fliesser M, et al. Low back pain and its relationship with sitting behaviour among sedentary office workers. Appl Ergon 2019 Nov;81:102894. [doi: 10.1016/j.apergo.2019.102894] [Medline: 31422243]
24. Urits I, Burshtein A, Sharma M, et al. Low back pain, a comprehensive review: pathophysiology, diagnosis, and treatment. Curr Pain Headache Rep 2019 Mar 11;23(3):23. [doi: 10.1007/s11916-019-0757-1] [Medline: 30854609]
25. Minghelli B, Nunes C, Oliveira R. Back school postural education program: comparison of two types of interventions in improving ergonomic knowledge about postures and reducing low back pain in adolescents. Int J Environ Res Public Health 2021 Apr 22;18(9):4434. [doi: 10.3390/ijerph18094434] [Medline: 33921952]
26. Huang R, Ning J, Chuter VH, et al. Exercise alone and exercise combined with education both prevent episodes of low back pain and related absenteeism: systematic review and network meta-analysis of randomised controlled trials (RCTs) aimed at preventing back pain. Br J Sports Med 2020 Jul;54(13):766-770. [doi: 10.1136/bjsports-2018-100035] [Medline: 31672696]
27. de Campos TF, Maher CG, Fuller JT, Steffens D, Attwell S, Hancock MJ. Prevention strategies to reduce future impact of low back pain: a systematic review and meta-analysis. Br J Sports Med 2021 May;55(9):468-476. [doi: 10.1136/bjsports-2019-101436] [Medline: 32646887]
28. Bettany-Saltikov J, McSherry R, van Schaik P, et al. PROTOCOL: school-based education programmes for improving knowledge of back health, ergonomics and postural behaviour of school children aged 4-18: a systematic review. Campbell Syst Rev 2019 Jun;15(1-2):e1014. [doi: 10.1002/cl2.1014] [Medline: 37131478]
29. Brink Y, Maart RA, Louw QA. School-based interventions to improve spinal health of children and adolescents: a systematic review. Physiother Theory Pract 2022 Nov;38(13):2378-2401. [doi: 10.1080/09593985.2021.1938305] [Medline: 34157947]
30. Valenciano PJ, Cibinello FU, Neves JCDJ, Fujisawa DS. Effects of postural education in elementary school children: a systematic review. Rev Paul Pediatr 2020;39:e2020005. [doi: 10.1590/1984-0462/2021/39/2020005] [Medline: 33146294]
31. Miñana-Signes V, Monfort-Pañego M, Valiente J. Teaching back health in the school setting: a systematic review of randomized controlled trials. Int J Environ Res Public Health 2021 Jan 22;18(3):979. [doi: 10.3390/ijerph18030979] [Medline: 33499403]
32. Salman M, Bettany-Saltikov J, Kandasamy G, Whittaker V, Hogg J, Racero GA. PROTOCOL: The effect of education programmes for improving knowledge of back health, ergonomics and postural behaviour in university students: a systematic review. Campbell Syst Rev 2022 Mar;18(1):e1213. [doi: 10.1002/cl2.1213] [Medline: 36908660]
33. Fisher D, Louw Q. The effect of classroom-based interventions on sedentary behavior and spinal health in schoolchildren: systematic review. Interact J Med Res 2022 Oct 26;11(2):e39006. [doi: 10.2196/39006] [Medline: 36287590]
34. Fiaschi Ramos AC, Ferreira Silva RM, Bizinotto T, et al. Tools for assessing knowledge of back health in adolescents: a systematic review protocol. Healthcare (Basel) 2022 Aug 22;10(8):1591. [doi: 10.3390/healthcare10081591] [Medline: 36011248]
35. Dugan JE. Teaching the body: a systematic review of posture interventions in primary schools. Educ Rev 2018 Oct 20;70(5):643-661. [doi: 10.1080/00131911.2017.1359821]
36. Galmes-Panades AM, Vidal-Conti J. Effects of Postural Education Program (PEPE study) on daily habits in children. Front Educ 2022;7:935002. [doi: 10.3389/feduc.2022.935002]
37. Galmes-Panades AM, Borràs PA, Vidal-Conti J. Association of postural education and postural hygiene with low back pain in schoolchildren: cross-sectional results from the PEPE study. Health Promot Perspect 2023;13(2):157-165. [doi: 10.34172/hpp.2023.19] [Medline: 37600543]
38. Kovacs FM, Burgos-Alonso N, Martín-Nogueras AM, Seco-Calvo J. The efficacy and effectiveness of education for preventing and treating non-specific low back pain in the hispanic cultural setting: a systematic review. Int J Environ Res Public Health 2022 Jan 12;19(2):825. [doi: 10.3390/ijerph19020825] [Medline: 35055646]
39. Araújo CL, Moreira A, Carvalho GS. Postural education programmes with school children: a scoping review. Sustainability 2023;15(13):10422. [doi: 10.3390/su151310422]
40. Chacón-Borrego F, Ubago-Jimenez JL, La Guardia García JJ, Padial Ruiz R, Cepero González M. Educación e higiene postural en el ámbito de la Educación Física. Papel del maestro en la prevención de lesiones: Revisión sistemática [Article in Spanish]. Retos 2018;34:8-13 [FREE Full text] [doi: 10.47197/retos.v0i34.54319]
41. Miñana V. Effects of a non-randomized educational intervention on knowledge, postural habits and trunk muscle endurance related to back health: a 6-month follow-up study. Eur J Hum Mov 2023;49:104-123. [doi: 10.21134/eurjhm.2022.49.8]
42. Criollo-C S, Guerrero-Arias A, Jaramillo-Alcázar Á, Luján-Mora S. Mobile learning technologies for education: benefits and pending issues. Appl Sci 2021;11(9):4111. [doi: 10.3390/app11094111]
43. Behera DK. Technological interventions in education: an empirical review of their impact on learning outcomes. Alsys Tech J Educ Technol 2023;1(1):62-77. [doi: 10.58578/alsystech.v1i1.1674]
44. Sormunen M, Saaranen T, Heikkilä A, et al. Digital learning interventions in higher education: a scoping review. Comput Inform Nurs 2020 Dec;38(12):613-624. [doi: 10.1097/CIN.0000000000000645] [Medline: 32520782]
45. Battis A, Norrie JP, McMaster H, Beaudette SM. Wearable technology mediated biofeedback to modulate spine motor control: a scoping review. BMC Musculoskelet Disord 2024 Oct 1;25(1):770. [doi: 10.1186/s12891-024-07867-3] [Medline: 39354458]

XSL•FO
RenderX

46. Ribeiro DC, Sole G, Abbott JH, Milosavljevic S. The effectiveness of a lumbopelvic monitor and feedback device to change postural behavior: a feasibility randomized controlled trial. J Orthop Sports Phys Ther 2014 Sep;44(9):702-711. [doi: 10.2519/jospt.2014.5009] [Medline: 25098195]
47. Çelenay Ş, Kaya D, Özüdoğru A. Spinal postural training: comparison of the postural and mobility effects of electrotherapy, exercise, biofeedback trainer in addition to postural education in university students. J Back Musculoskelet Rehabil 2015;28(1):135-144. [doi: 10.3233/BMR-140501] [Medline: 25061037]
48. Kuo YL, Huang KY, Kao CY, Tsai YJ. Sitting posture during prolonged computer typing with and without a wearable biofeedback sensor. Int J Environ Res Public Health 2021 May 19;18(10):5430. [doi: 10.3390/ijerph18105430] [Medline: 34069579]
49. Giggins OM, Persson UM, Caulfield B. Biofeedback in rehabilitation. J Neuroeng Rehabil 2013 Jun 18;10(1):60. [doi: 10.1186/1743-0003-10-60] [Medline: 23777436]
50. Lind CM. A rapid review on the effectiveness and use of wearable biofeedback motion capture systems in ergonomics to mitigate adverse postures and movements of the upper body. Sensors (Basel) 2024 May 23;24(11):3345. [doi: 10.3390/s24113345] [Medline: 38894134]
51. Hodges PW, van den Hoorn W. A vision for the future of wearable sensors in spine care and its challenges: narrative review. J Spine Surg 2022 Mar;8(1):103-116. [doi: 10.21037/jss-21-112] [Medline: 35441093]
52. Papi E, Koh WS, McGregor AH. Wearable technology for spine movement assessment: a systematic review. J Biomech 2017 Nov 7;64:186-197. [doi: 10.1016/j.jbiomech.2017.09.037] [Medline: 29102267]
53. Simpson L, Maharaj MM, Mobbs RJ. The role of wearables in spinal posture analysis: a systematic review. BMC Musculoskelet Disord 2019 Feb 8;20(1):55. [doi: 10.1186/s12891-019-2430-6] [Medline: 30736775]
54. Rodrigues RNG, Tucci AM, de Barros Viana M. The use of biofeedback on students: a systematic review. Trends in Psychol 2022 Jun;30(2):345-366. [doi: 10.1007/s43076-021-00109-8]
55. Menor-Rodríguez MJ, Rodríguez-Blanque R, Montiel-Troya M, Cortés-Martín J, Aguilar-Cordero MJ, Sánchez-García JC. Educational intervention in the postural hygiene of school-age children. Healthcare (Basel) 2022 May 6;10(5):864. [doi: 10.3390/healthcare10050864] [Medline: 35628000]
56. Figueira V, Silva S, Costa I, et al. Wearables for monitoring and postural feedback in the work context: a scoping review. Sensors (Basel) 2024 Feb 19;24(4):1341. [doi: 10.3390/s24041341] [Medline: 38400498]
57. Sabino I, do Carmo Fernandes M, Cepeda C, et al. Application of wearable technology for the ergonomic risk assessment of healthcare professionals: a systematic literature review. Int J Ind Ergon 2024 Mar;100:103570. [doi: 10.1016/j.ergon.2024.103570]
58. Sigrist R, Rauter G, Riener R, Wolf P. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. Psychon Bull Rev 2013 Feb;20(1):21-53. [doi: 10.3758/s13423-012-0333-8] [Medline: 23132605]
59. Moinuddin A, Goel A, Sethi Y. The role of augmented feedback on motor learning: a systematic review. Cureus 2021 Nov;13(11):e19695. [doi: 10.7759/cureus.19695] [Medline: 34976475]
60. Guo D, McTigue EM, Matthews SD, Zimmer W. The impact of visual displays on learning across the disciplines: a systematic review. Educ Psychol Rev 2020 Sep;32(3):627-656. [doi: 10.1007/s10648-020-09523-3]
61. Martín-Martín J, Roldán-Jiménez C, De-Torres I, et al. Behavior change techniques and the effects associated with digital behavior change interventions in sedentary behavior in the clinical population: a systematic review. Front Digit Health 2021;3:620383. [doi: 10.3389/fdgth.2021.620383] [Medline: 34713097]
62. Zhu Y, Long Y, Wang H, Lee KP, Zhang L, Wang SJ. Digital behavior change intervention designs for habit formation: systematic review. J Med Internet Res 2024 May 24;26:e54375. [doi: 10.2196/54375] [Medline: 38787601]
63. García-Luna MA, Jimenez-Olmedo JM, Pueo B, Manchado C, Cortell-Tormo JM. Concurrent validity of the ergotex device for measuring low back posture. Bioengineering (Basel) 2024 Jan 20;11(1):98. [doi: 10.3390/bioengineering11010098] [Medline: 38275578]
64. Sevilla-Gonzalez MDR, Moreno Loaeza L, Lazaro-Carrera LS, et al. Spanish version of the System Usability Scale for the assessment of electronic tools: development and validation. JMIR Hum Factors 2020 Dec 16;7(4):e21161. [doi: 10.2196/21161] [Medline: 33325828]
65. Brooke J. SUS: a “quick and dirty” usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, editors. Usability Evaluation in Industry, 1st edition: Chemical Rubber Company Press; 1996:189-194. [doi: 10.1201/9781498710411]
66. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
67. Lewis JR. The System Usability Scale: past, present, and future. Int J Hum Comput Interact 2018 Jul 3;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
68. Brooke J. SUS: a retrospective. J Usability Stud 2013;8(2):29-40 [FREE Full text]
69. Sauro J, Lewis JR. Quantifying the User Experience: Practical Statistics for User Research, 2nd edition: Morgan Kaufmann; 2016. [doi: 10.1016/B978-0-12-802308-2.00002-3]
70. Kim CY, Lee J, Jeong EY, et al. Wireless technologies for wearable electronics: a review. Adv Elect Mater 2025 Jul;11(10). [doi: 10.1002/aelm.202400884]

XSL•FO
RenderX

71. Qadrie Z, Ashraf H, Maqbool M, et al. Wearable technology in healthcare: opportunities, challenges, and future directions. Smart Wearable Technol 2025;1:1-17. [doi: 10.47852/bonviewSWT52025578]
72. Matuska S, Paralic M, Hudec R. A smart system for sitting posture detection based on force sensors and mobile application. Mob Inf Syst 2020 Nov 18;2020(1):6625797. [doi: 10.1155/2020/6625797]
73. Kim W, Ruiz Garate V, Gandarias JM, Lorenzini M, Ajoudani A. A directional vibrotactile feedback interface for ergonomic postural adjustment. IEEE Trans Haptics 2021;15(1):200-211. [doi: 10.1109/TOH.2021.3112795]
74. Ramachandran V, Schilling F, Wu AR, Floreano D. Smart textiles that teach: fabric‐based haptic device improves the rate of motor learning. Adv Intell Syst 2021 Nov;3(11). [doi: 10.1002/aisy.202100043]
75. Hernández-Mustieles MA, Lima-Carmona YE, Pacheco-Ramírez MA, et al. Wearable biosensor technology in education: a systematic review. Sensors (Basel) 2024 Apr 11;24(8):2437. [doi: 10.3390/s24082437] [Medline: 38676053]
76. Lang CE, Barth J, Holleran CL, Konrad JD, Bland MD. Implementation of wearable sensing technology for movement: pushing forward into the routine physical rehabilitation care field. Sensors (Basel) 2020 Oct 10;20(20):5744. [doi: 10.3390/s20205744] [Medline: 33050368]

Abbreviations

IMU: inertial measurement unit
LBP: low back pain
SUS: system usability scale

Edited by A Kushniruk; submitted 08.Jul.2025; peer-reviewed by JM Aerts, MT Martinez-Romero, O Østerlie, R Sanchis-Sanchis; revised version received 16.Dec.2025; accepted 18.Jan.2026; published 26.Feb.2026.

Please cite as:
García-Luna MA, García-Jaén M, Ruiz-Fernández D, Manchado C, Cortell-Tormo JM
Postural Education in School-Aged Populations: Development and Usability Evaluation of a Mobile Biofeedback App (EduBack)
JMIR Hum Factors 2026;13:e79282
URL: https://humanfactors.jmir.org/2026/1/e79282
doi:10.2196/79282

© Marco A García-Luna, Miguel García-Jaén, Daniel Ruiz-Fernández, Carmen Manchado, Juan M Cortell-Tormo. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 26.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Real-World Evaluation of the Eye+Dot Online Triage Support Tool in Community Optometry Practices: Mixed Methods Evaluation Study

Jen O Lim[1], BA, MB BChir; Sarah Farrell[2], Bsc; Samuel J Newlands[2], MBChB, PGDipOphthBS; Louise Allen[2], MBBS, MA, MD

[1]School of Clinical Medicine, University of Cambridge, Cambridge, United Kingdom

[2]Department of Ophthalmology, Cambridge University Hospitals NHS Foundation Trust, Hills Road, Cambridge, United Kingdom

Corresponding Author:
Louise Allen, MBBS, MA, MD
Department of Ophthalmology, Cambridge University Hospitals NHS Foundation Trust, Hills Road, Cambridge, United Kingdom

Abstract

Background: Many patients attending hospital emergency services with recent-onset eye symptoms could have been managed in the community. This may reflect a lack of specialist experience or triage capacity among primary care providers. Online triage tools collect patient symptomatology and relevant ophthalmic and medical history virtually, compile a report, and suggest an outcome decision which can support the streaming and direction of patients to a suitable service within an appropriate timeframe.

Objective: Our objectives were to implement and analyze the real-world use of a novel online triage tool in a community optometry setting.

Methods: This was a real-world mixed methods evaluation of a web app "eye+dot" (Cambridge Medical Innovation Ltd), which uses a conditional logic multiple-choice questionnaire to produce a symptom report and automated triage recommendation. A preimplementation survey was used to gather views regarding current triage practices and interest in trialing an online triage tool among community optometrists. The app was then implemented in participating practices, with optometrists receiving the symptom report but not the automated triage disposition. The sensitivity and specificity of optometrists and the app at identifying urgent cases (those requiring specialist examination within 24 h) were compared to a gold standard, generated by retrospective review of the symptom reports by hospital specialists. Metadata including patient age, test duration, and usability feedback from patients and optometry staff were analyzed.

Results: Seven optometry practices trialed the app over 5 months. Records of 209 patients with a mean age of 53.5 (SD 17.5; range 13 - 90) years were analyzed. Three patients with potentially sight- or life-threatening symptoms (as defined by Royal College of Ophthalmologists guidelines) received appropriate urgent hospital dispositions by both optometrists and eye+dot. The app's automated disposition had similar sensitivity to optometry triage (75.5% vs 77.3% respectively; P=.99) but significantly better specificity (82.7% vs 54.4%; P<.001) for identification of the 53 cases categorized as requiring either hospital or community specialist examination within 24 hours by gold-standard review. A total of 147 patients gave acceptability feedback, with 121 (82%) of them giving a rating in the top 2 points of a 5-point visual Likert scale. Optometrists in 5 of the 7 practices found the app helpful and were interested in continuing its use.

Conclusions: This real-world study demonstrates the triage tool's accuracy for identifying high-risk symptomatology and its acceptability among patients and optometry services. The significantly lower optometric triage specificity may reflect risk-averse decision-making following direct communication with patients and appointment availability in primary care. In addition, the relatively low mean age of patients suggests possible barriers to uptake among older patients. This and other symptom-based triage tools can provide useful support to eye care providers in the community where access to specialist triage may be limited.

(JMIR Hum Factors 2026;13:e77278) doi:10.2196/77278

KEYWORDS

triage; triage methods; mobile apps; computer-assisted decision making; decision tree; primary health care; emergency service; eye diseases; optometry; mobile health; digital health

XSL•FO
RenderX

Introduction

Background

Patients with new eye or vision symptoms often contact community optometry practices in person or by telephone for advice. Informal triage and advice may be given if there is capacity to do so, but a lack of specific remuneration for triage activities and limited optometrist availability can mean that such patients are redirected to emergency departments (ED) or emergency eye clinics (EECs) with minimal specialist optometrist input. Community optometrist–delivered Minor Eye Condition Services (MECS) and Community Urgent Eyecare Services (CUES) have been commissioned in many areas of the United Kingdom to provide urgent, optometry-led community eye care and reduce pressure on emergency services, and studies indicate that between 37% and 92% of self or general practitioner-referred patients in EECs may be suitable for these services [1-5]. Efficient and early redirection of suitable patients to MECS or CUES could reduce the 870,000 annual ED attendances across the country for eye and vision symptoms and move care into the community, a key goal in the National Health Service (NHS) Long Term Plan [6].

Paper triage support charts are currently used by many optometry practices and EDs. More recently, patient-facing online symptom-based tools with automated triage algorithms have been developed [7-11]. Several have been validated with encouraging results in EDs, but to our knowledge, no digital specialist triage support tool for ophthalmic symptoms has been used to stream and direct patients in the community setting. However, evaluation studies of nonophthalmic digital triage tools in the primary care setting have provided promising results but reveal possible difficulties around triage accuracy, with automated tools often tending to unnecessarily overtriage cases as emergencies, and equity concerns around access and ease of using such online tools for older patients or patients with disabilities [12]. In addition, integration with existing workflows can be challenging, and low patient compliance with suggested triage advice can prevent gains in practice efficiency from being realized [13,14].

Study Aims

The purpose of this study was to evaluate the potential use of an eye and vision symptom-based online triage support tool, examine its real-world implementation, and analyze its accuracy at identifying high-risk ophthalmic symptomatology in the community.

Methods

Study Design

This was a prospective, mixed methods real-world study focusing on how the technology works for patients in everyday community optometry practice. Real-world studies enable testing the technology with real patient scenarios and the inclusion of patients and practitioners who might otherwise not take part in research, making the results more useful for the appropriate context but less strictly controlled than formal validation studies.

Ethical Considerations

Approval of the study's methodology was obtained from the quality and safety committee of the Cambridge University Hospitals NHS Foundation Trust and the Cambridgeshire and Peterborough Integrated Care Service (approval reference: PRN10395). As an anonymized service evaluation that did not involve deviation from the current standard of care, this study was exempt from requiring national ethics committee approval [15]. Patients gave verbal agreement to participate and digitally agreed to the General Data Protection Regulation (GDPR) statement prior to undertaking the online eye+dot questionnaire, and the study was undertaken according to the tenets of the Declaration of Helsinki. Patients could choose to opt out without affecting their care.

The Technology

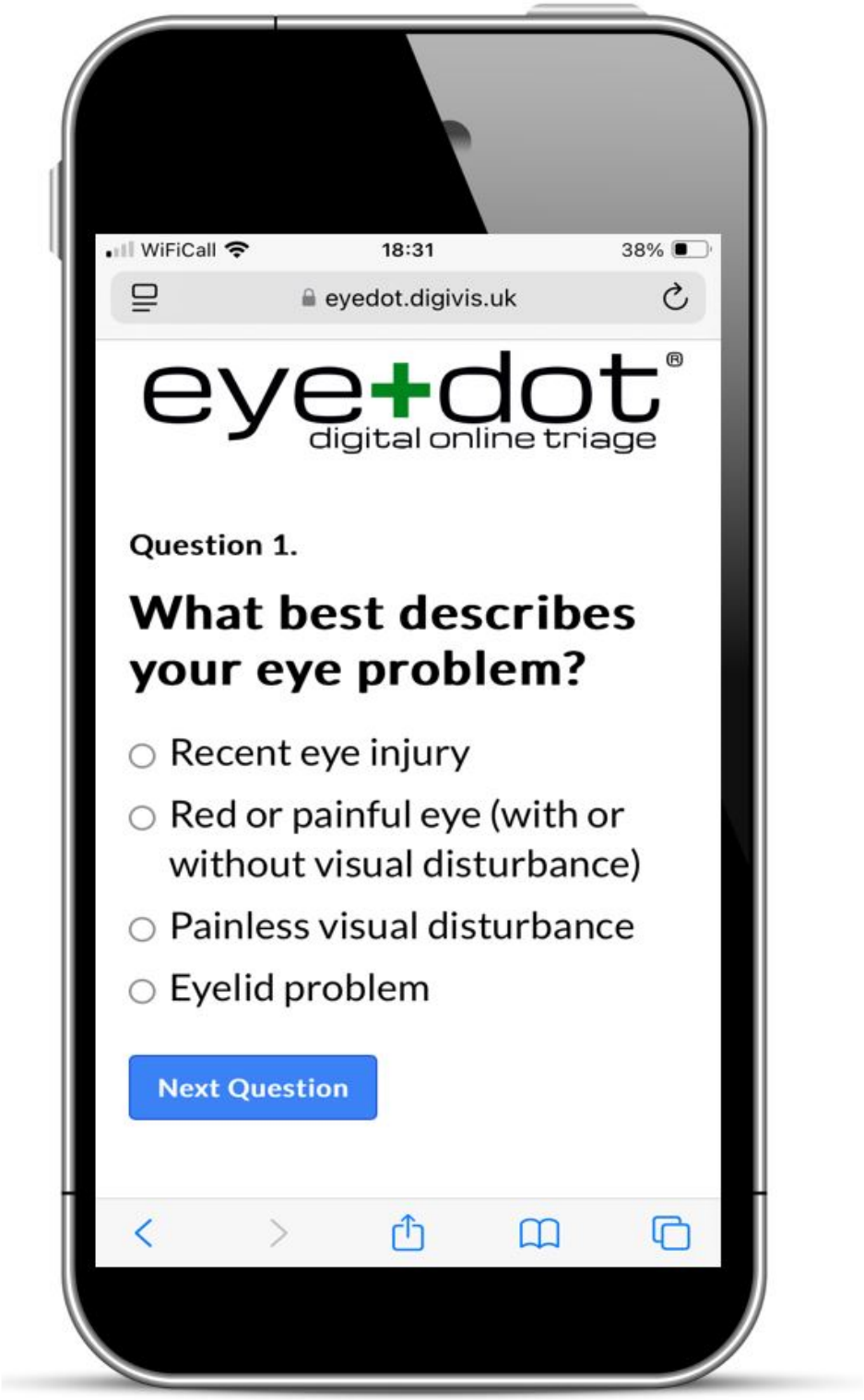
The eye+dot web app (Cambridge Medical Innovation Ltd) is a triage support tool that uses a conditional logic multiple-choice questionnaire of up to 25 questions to collect relevant ophthalmic and clinical history, detect high-risk symptoms, compile a downloadable symptom report, and suggest a triage disposition in terms of appropriate service and urgency. The tool has been developed and iteratively improved over 4 sequential proof-of-concept and validation studies in 3 hospital trusts of the Cambridgeshire and Peterborough Integrated Care Service. The automated triage algorithm used in eye+dot was adapted from the symptom triage guidelines published by the Royal College of Ophthalmologists (RCOphth) [16]. The initial process involved the development of multiple-choice questions and the conditional logic structure. These, and the triage disposition algorithm, were tested and iteratively improved, initially using simulated clinical scenarios, and subsequently evaluation and validation in over 500 real clinical scenarios. Wording, accessibility, and user interface improvements have been made based on patient and clinician feedback. Recent integration with the Electronic Patient Record (Epic) enables clinicians to order and retrieve results directly through the patient record. The app is currently in clinical use in an NHS hospital trust. As the eye+dot app replicates a chart-based symptom scoring chart, it does not meet the criteria for classification as a Medical Device under the UK Medical Devices Regulations based on Medicines and Healthcare Products Regulatory Agency guidance.

The care provider sends the patient a weblink to the online questionnaire by SMS text messaging or email, documenting the unique order number in the patient record. The patient accepts an initial standard disclaimer, GDPR-compliant data usage notice, and inputs their age before starting the questionnaire. The first filter question as seen by the patient is shown in Figure 1. Web browser translation is available but was not used for this study. Symptoms are automatically rated in terms of acuity adapted from previously published triage support tables (Multimedia Appendix 1) [16]. The symptom acuity rating creates an automated disposition, with the suggestion of an appropriate service (ED or EEC, MECS or CUES, and other community optometry services or pharmacist) and urgency (same d, within 24 h, within 48 h, and within a wk or routine). The symptom report and automated disposition are

XSL•FO
RenderX

automatically emailed to the provider's nominated inbox and can also be accessed via the secure online portal (Multimedia Appendix 2). This report can also be used to facilitate effective referrals, which other studies have shown are often affected by incomplete content [17,18]. The report is not routinely made available to the patient unless there is a history of potentially severe chemical, blunt, or high-velocity trauma. In such cases, the patient receives an immediate response with a weblink to the NHS first aid website and advice to attend emergency services urgently.

Figure 1. Initial filter question of the triage tool as it appears on a smartphone.



For the purposes of this study, the ordering provider received the symptom report but was masked to the automated disposition to prevent biasing of their clinical decision.

Preimplementation Survey

Optometrists and managers in the 23 practices registered with the Cambridgeshire Local Optical Committee (a statutory body

XSL•FO
RenderX

representing and supporting optometrists in our region) and accredited in the provision of specialist eyecare with our NHS hospital trust were invited to complete an online survey, designed using Microsoft Forms (Microsoft Corp), regarding current triage practices, and to express interest in participating in a 5-month pilot of a novel online triage tool (Multimedia Appendix 3). These practices serve a population of approximately 920,000. Survey questions were adapted from the validated Normalization Measure Development instrument based on the normalization process theory [19,20] and included questions about patient pathways and volumes, triage tools currently in use, and overall triage outcomes.

Implementation

Practices interested in trialing the triage tool participated in a training webinar (led by LA and SF) to explain the use of the triage tool and the implementation process. Two similar webinars were held at different times of the week to encourage attendance, and a recording was made available online for those practice staff unable to attend. At these webinars, optometrists and administrative staff received instruction in how to order an eye+dot link for patients and complete a study data spreadsheet with the date, name of participating patients, eye+dot reference number, and the outcome of optometry triage. Spreadsheets were pseudonymized prior to being returned to the study team for analysis. Participating practices were asked to use the tool's symptom report in addition to their current triage process for a 5-month period. Patients were directed to services based on their optometry practice's triage decision rather than the eye+dot automated triage disposition.

Criteria for inclusion were that the patient should be aged 13 years or older with normal cognition, a good understanding of English, and recent-onset eye or vision symptoms (within 7 d). Patients may have either walked into the practice or phoned for advice. Patient selection was otherwise highly pragmatic, with practice staff inviting patients to undertake the online questionnaire in addition to the practice's standard triage process. Only patients who agreed to using eye+dot were sent the link to the online tool's questionnaire. Participation was voluntary and patients were informed that the standard of care would not be affected by either agreeing or declining to participate. A further written informed consent process was not necessary for this service evaluation study and was not carried out, although all patients using eye+dot were asked to consent to its GDPR-compliant data use policy before commencing the symptom questionnaire. Patients without digital access were able to undertake the questionnaire over the telephone, with the receptionist documenting the patient's responses on their behalf. The generated symptom report was automatically returned to the nominated practice email inbox to be made available to the triaging optometrist. The unique weblink order number and the optometry triage outcome were recorded by a staff member in a spreadsheet held securely by the practice. To compensate practices for the administration involved, an honorarium of £15 (€17.5 or US $19) was agreed for each completed and deidentified patient record, payable at the end of the study period.

Postimplementation Report Review

Following the 5-month period of implementation, records from optometry practices were deidentified prior to their release to the study team for analysis. Each patient's online symptom report, automated triage disposition, and optometric triage outcome were linked with their unique eye+dot order number. Only patients meeting the eligibility criteria with recorded optometric triage outcome data and a complete online eye+dot report were included for analysis.

Two hospital-based specialists (consultant ophthalmologist LA and hospital optometrist SF), masked to the optometrist and eye+dot triage tool dispositions, reviewed the symptom reports and identified patients with high acuity symptoms who required either urgent hospital care or hospital or community specialist assessment within 24 hours. In cases of disagreement, a third clinician (specialist clinical fellow SN) made the casting decision. This was used as the gold standard triage disposition.

Postimplementation feedback from participating optometrists was communicated by email requesting free-text responses under the following domains: advantages and disadvantages of using eye+dot, barriers and enablers to its uptake, improvements that could be made, and likelihood of adoption in the future.

Statistical Analysis

Preimplementation and postimplementation survey responses were analyzed using descriptive statistics. Differences in sensitivity and specificity of eye+dot automated triage and optometric triage at identifying high-acuity cases as defined above were evaluated using the McNemar test with continuity correction on GraphPad Prism version 10.4.1 (GraphPad Software). In addition, interrater reliability at identifying urgent or nonurgent cases between eye+dot and the other modes of triage, as well as between the 2 main reviewers who determined the gold-standard triage outcome, was evaluated by calculation of Cohen κ. A priori interpretation of κ values was as follows: values ≤0 indicating no agreement, 0.01 to 0.20 (none to slight), 0.21 to 0.40 (fair), 0.41 to 0.60 (moderate), 0.61 to 0.80 (substantial), and 0.81 to 1.00 (as almost perfect agreement). Additionally, adherence to RCOphth guidelines for hospital assessment within 24 hours for presentations suggestive of a sight or life-threatening condition (Multimedia Appendix 4) was assessed. Metadata including duration and completion of the online triage questionnaire, patient age, and feedback on the app using a voluntary visual Likert scale with responses converted to numerical values (Multimedia Appendix 5) were analyzed by descriptive statistics.

Results

Preimplementation Survey

The preimplementation survey (Multimedia Appendix 3) received 53 responses, 50 (94%) of which were from optometrists, and the remaining 3 from practice managers. From the 53 respondents, 36 (68%) reported that they or their practice currently used a paper-based triage tool, and 45 (85%) reported that they were the first point of contact for most patients with recent onset eye symptoms. The most frequent triage outcome prior to implementation of the online triage tool was reported

XSL•FO
RenderX

as a General Ophthalmic Service or private appointment by 27 (51%) respondents, a MECS or CUES appointment by 23 (43%), and advice with or without purchase of eye drops by 2 (4%). One optometrist reported that the most common outcome was advice to attend ED or EEC.

From the 53 respondents, 45 (85%) agreed or strongly agreed that they would be willing to learn how to use an online triage tool, 39 (74%) agreed or strongly agreed that an online tool could improve patient care and make onward referral easier, and 30 (57%) felt that patients would be willing to undertake an online questionnaire for triage purposes.

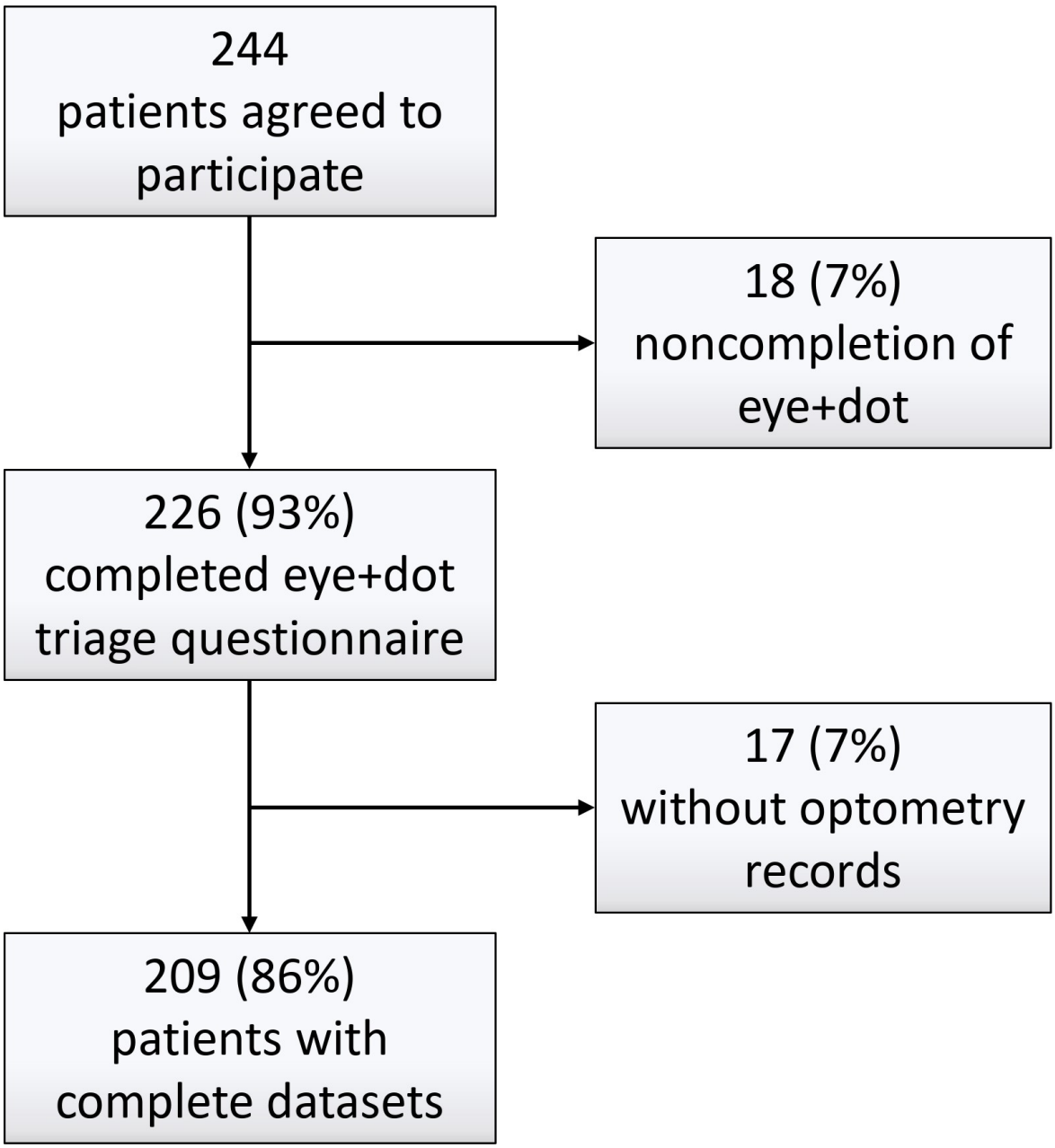
Implementation

The online triage tool was implemented in 7 out of 23 invited optometry practices comprising 4 Specsavers branches and 3 independent practices, all of which also provided MECS services. In total, 14 members of staff attended one of the live webinar sessions. Four other practices expressed interest but were unable to participate due to a lack of administrative or time resources.

During the 5-month implementation period, 244 patients agreed to be sent the online triage link by their practice. Of the 244 patients, 18 (7%) started but failed to complete the online questionnaire, and a further 17 (7%) were excluded due to incomplete optometry triage outcome records. The remaining 209 records were analyzed (Figure 2). The mean age of patients was 53.5 (SD 17.5; range 13 - 90) years. Of the 209 records, 136 (65%) included information on the mode of questionnaire completion. For 19 patients, this was undertaken over the phone with the help of a receptionist rather than online. Based on the responses to the first filter question of the online triage questionnaire, the most frequent primary presenting complaint was red or painful eyes, followed by painless visual disturbance, eyelid problems, and eye injuries (Table 1).

XSL•FO
RenderX

Figure 2. Implementation phase flowchart.



XSL•FO
RenderX

Table . Characteristics of patients with complete online and optometry triage records (N=209).

Feature	Patients, n (%)
Age range (y)	
<20	7 (3.3)
20 - 29	22 (10.5)
30 - 39	16 (7.7)
40 - 49	27 (12.9)
50 - 59	51 (24.4)
60 - 69	48 (23)
70 - 79	27 (12.9)
>79	11 (5.3)
Main presenting complaint	
Red and/or painful eye or eyes	95 (45.5)
Painless visual disturbance	82 (39.2)
Eyelid problems	22 (10.5)
Eye injury	12 (5.7)

Triage Outcomes

Optometric triage outcomes (Multimedia Appendix 6): 106 out of 209 patients (51%) were seen within 24 hours. Of these, 4 (2% of 209) were advised to contact or attend ED or EEC and 31 (15% of 209) attended same-day optometry appointments. 89 (43% of 209) patients were offered less urgent optometry appointments, 13 (6% of 209) given advice or dispensed eyedrops without an appointment and 1 (1% of 209) advised to contact their general practitioner.

Eye+dot triage disposition (Multimedia Appendix 6): 67 out of 209 patients (32%) were allocated a disposition of hospital or community specialist assessment within 24 hours, and 9 (4% of 209) of these were recommended same day ED or EEC appointments. A total of 132 (63% of 209) patients received less urgent triage dispositions to hospital or community optometry services, while pharmacist advice was recommended for a further 10 patients.

Accuracy at Identifying High Acuity Cases

From the 209 patients, 53 (25%) were considered to have potentially urgent eye problems requiring hospital or community specialist eye assessment within 24 hours according to the gold-standard triage dispositions established by expert review. This included 3 patients with presentations meeting the RCOphth definition of an emergency eye condition (Multimedia Appendix 4). There was no significant difference in the proportion of patients with gold-standard categorizations requiring specialist assessment within 24 hours presenting to Specsavers branches and independent optometry practices (25% vs 24% respectively; odds ratio 1.07, 95% CI 0.57 - 2.02; *P*=.84).

All 3 patients with presentations meeting the RCOphth definition of an emergency eye condition [16] were correctly identified as requiring urgent hospital attendance by both optometric and automated triage mechanisms.

Of the 53 patients assigned a gold standard recommendation of hospital or community specialist assessment within 24 hours, 40 (76%) were assigned that outcome by eye+dot triage, with the remainder (n=13, 25%) assigned to be seen within 48 hours. Of the same 53 patients, 41 (77%) were assigned a hospital or optometrist assessment within 24 hours by optometry triage, with a further 4 (8%) to be seen within 48 hours, and 8 (15%) cases considered routine (Table 2). The online triage tool had similar sensitivity as optometry triage (75% vs 77%, respectively; *P*=.99) but significantly better specificity (83% vs 54%, *P*<.001) at identifying cases requiring urgent specialist appointments. Eye+dot's automated triage was found to have moderate-substantial agreement with each of the 2 expert reviewers but lower agreement with the optometrist triage (Table 3).

XSL•FO
RenderX

Table . Disposition outcomes for the 53 patients with gold standard high-acuity symptomatology based on their online questionnaire result with sensitivity and specificity for identifying urgent cases (N=53).

Disposition urgency	Optometry triage (N=53)	Online triage (N=53)	*P* value
Within 24 h, n (%)	41 (77)	40 (75)	
Sensitivity, %	77	75	.99
Specificity, %	54	83	<.001
Within 48 h, n (%)	4 (8)	13 (25)	—[a]
Within a wk/routine/advice, n (%)	8 (15)	0 (0)	—

[a]Not applicable.

Table . Interrater reliability for urgent or nonurgent triage dispositions.

Pair	Cohen κ (95% CI)
Optometrist and eye+dot	0.26 (0.15-0.37)
Reviewer 1 and eye+dot	0.75 (0.64-0.86)
Reviewer 2 and eye+dot	0.61 (0.50-0.72)
Reviewer 1 and Reviewer 2	0.40 (0.28-0.52)

Usability

The mean time taken to complete the online questionnaire was 4 minutes 57 seconds (SD 2 minutes 2 seconds; range 1 minute 30 seconds to 16 minutes 15 seconds). Overall, 147 out of 209 (70%) patients provided voluntary feedback about the eye+dot app via a 5-point visual Likert scale. Of these, 121 (82%) patients rated the app 4 or 5 (mean 4.27, SD 0.76) with 5 being the most positive rating. Of patients whose method of undertaking the questionnaire had been recorded, 71 of the 82 (87%) online respondents rated the app 4 or 5 compared to 9 of the 14 (64%) patients undertaking the questionnaire over the phone (mean 0.71, SD 4.39 and mean 4.21, SD 0.94 for online and telephone groups, respectively; *P*=.70 using the Mann-Whitney *U* test).

Postimplementation Feedback

Qualitative postimplementation feedback responses to a semistructured email questionnaire by participating practices are shown in Textbox 1.

XSL•FO
RenderX

Textbox 1. Semistructured postimplementation feedback from the 7 participating practices.

Were there any advantages of using the eye+dot tool?

- The compiled symptom report improved documentation and supported onward referral
- The report provided useful documentation for records
- Helped reduce the need to interrupt optometrists undertaking scheduled work (3 Specsavers practices mentioned this)

Were there any disadvantages of using the eye+dot tool?

- Older patients were more hesitant to accept online triage
- Optometrists found it quite difficult to interpret the symptom report
- Patients often volunteered different symptomatology when telephoned
- Was not found helpful because there was already capacity to triage and book patient appointments without delay (at 2 independent practices)
- Increased triage time (at 2 independent practices)

What were the barriers and enablers to implementing the innovation?

- Uptake was affected by staffing shortages, changes, and the administrative burden of completing the study documentation
- Uptake was improved by the workflow being accepted as normal practice, with all eligible and able patients being routinely requested to complete the online questionnaire

What improvements could be made to the eye+dot tool?

- Recording eye laterality to enable comparison with previous optometric records
- A free text box for patients after the forced choice questions for clarifications, if needed

Would your practice consider using an online triage tool in the future?

- Yes: 5 (4 Specsavers branches, one independent practice)
- No: 2 (independent practices)

Discussion

Utility and Role of the Eye+Dot Triage Tool

Triage systems should ideally have both high sensitivity and high specificity for urgent cases to ensure timely access to care and efficient use of resources. This real-world study suggests that the eye+dot web app has similar sensitivity as community optometrists at identifying high-risk eye or vision symptomatology requiring specialist assessment within 24 hours and possibly superior specificity. The 3 patients whose symptoms were potentially sight- or life-threatening (based on the RCOphth guidelines; Textbox 1) were accurately identified by both optometric and eye+dot triage. Overall, 12 and 11 patients were underrated by optometry triage and eye.dot triage, respectively, compared to the gold-standard hospital clinician disposition based on review of the eye+dot report. In all 11 cases underrated by eye+dot, the disposition was for specialist review within 48 hours. Of the 12 patients who underrated the optometrists, 4 were given appointments within 48 hours and the remainder routine appointments.

There was substantial correlation between the urgency suggested by eye+dot and expert hospital clinician dispositions (κ=0.61 and 0.75). In contrast, there was only a slight correlation between the hospital clinicians and community optometrists (κ=0.26). Interestingly, most automated triage tools are found to have lower specificity than clinician triage for high-risk presentations, which does not align with our findings for the eye+dot triage tool relative to community optometrist triage [8,12,21]. In this study, the higher acuity triage decisions made by community optometrists may have been influenced by other factors, including additional information from the patient such as their level of anxiety and real-time knowledge of appointment availability. This suggests that human contact and decision-making still have an important role, even when an automatic triage tool is used.

Advantages of the real-world nature of our study design include the inclusion of individuals who might not usually contribute to such studies, understanding how technology can be implemented within service workflows, faster timelines, and cost efficiency. This may be particularly helpful for developers of innovative technology to rapidly understand its feasibility and usability in practice, identify improvements, and optimize its implementation model [22]. The qualitative post-implementation feedback has led to the inclusion of a question about symptom laterality and a free-text box for additional information after the forced-choice questions (Multimedia Appendix 2). A randomized controlled trial would have reduced the risk of bias and ensured optimal data quality, but would not have been feasible in busy, commercial optometry practices.

To the extent that these results represent a real improvement in eye+dot's ability to accurately identify low-urgency cases relative to frontline human triage, they suggest that significant improvements in healthcare system efficiency could be achieved

XSL•FO
RenderX

by using eye+dot or a similar tool to reduce overtriage of patients and consequent pressure on acute hospital services. However, the limitations of the pragmatic design of this study and the possible impact of population or region-specific factors mean that more work is required before definite or generalizable conclusions can be drawn from this observation. Of note, only 4% of patients in this study were advised to attend or contact the ED or EEC by the optometrists in this study, all of whom had received enhanced training as MECS providers. The proportion directed to hospital emergency services is likely to be higher among practices without MECS-trained clinicians.

While it is possible for a member of staff to take a patient through the questionnaire over the phone, only a minority of patients—14% of those for whom the delivery method was documented—chose to use this option. This form of delivery may be less feasible, given the time required. Regular smartphone use is less likely among older individuals, and such patients may prefer to have a 2-way conversation with a clinician if they are anxious about new symptoms [23,24]. These barriers to uptake make it important to ensure other means of triage are available.

The voluntary visual Likert score was not completed by 62 out of 209 or 30% of patients, but 82% (121/147) of those who did gave the eye+dot app a rating in the top 2 of 5 options, suggesting a reasonable level of acceptability for patients. The duration of the test was highly variable, with longer tests reflecting the complexity of the individuals' symptoms and history. The use of web browser translation may have improved the triage experience and quality for individuals whose first language was not English, but this was not assessed during this study.

Limitations

The uptake of digital triage technology in this study was influenced by burdensome study administrative requirements and staff turnover in optometry practices. Those recruiting the most patients had embedded the triage tool into their workflow, routinely inviting suitable patients to complete the questionnaire. Despite only having access to the symptom report (and not the automated triage disposition), all Specsavers practices and one independent optometry practice were interested in continuing to use the tool. Reported benefits included ease of documentation, avoiding interruptions to scheduled appointments, and more efficient MECS booking. The tool may be less useful in practices with more generous optometrist scheduling and greater capacity to triage effectively. Conversely, practices where staff adopt a cautious approach to triage due to time or resource constraints are likely to benefit the most from a triage tool with high specificity for urgent cases [5].

It is possible that patient selection was biased towards the inclusion of younger patients (mean age of 53.5, SD 17.5 y), although we do not have data on the average age of all patients presenting to these optometry practices across the relevant time period. Preferential inclusion of younger patients could potentially bias the range of eye symptoms and previous ophthalmic history elicited.

Retrospective review by expert clinicians is a common method used to establish the reference gold standard triage disposition in studies validating both paper and digital triage tools [9,25-27]. However, minor differences in the precise methodology can introduce varying limitations. In this study, the gold standard disposition may not be ideal for evaluation of optometrist triage as it relied purely on specialist review of symptom reports without an option for them to talk directly with the patient and without the eventual clinical diagnosis being known. On the other hand, the lack of clinical outcome data prevented outcome bias which might arise from knowing that a patient with high acuity red-flag symptoms eventually received a relatively benign diagnosis.

Finally, we were not able to estimate theoretical patient compliance with eye+dot's automated triage advice from this study, as all patients were directed to services in line with their optometrist's triage decision. Other studies have noted that patients advised to seek community health care by remote triage often present to the ED regardless [14]. However, since the eye+dot app is designed to support triage decisions made and communicated by clinical teams rather than issuing advice directly to the patient automatically, this risk may be relatively small.

Conclusion

This real-world evaluation demonstrates that use of the eye+dot symptom-based online triage tool is feasible, acceptable to patients and staff, and can provide safe triage support, potentially improving triage efficiency and decreasing unnecessary hospital attendance if implemented in nonspecialist services.

These benefits may be even more pronounced if implemented in primary care settings without specialist triage resources or expertise. We are also currently evaluating the utility of the triage tool in hospital emergency services, both in supporting nurse-led triage for patients contacting the EEC service by phone and as a kiosk-based service for walk-in patients.

Acknowledgments

The authors thank staff at Specsavers Newmarket, Royston and Fitzroy Street and Lions Yard branches in Cambridge, DH Thomas Opticians in Cambridge, and Saffron Walden and James Oliver Radley Opticians in Buntingford for their contributions to this study.

XSL•FO
RenderX

Funding

This research was funded by the National Institute for Health Research i4i Funding at the Speed of Translation (FAST) award. The views expressed are those of the authors and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care.

Data Availability

The deidentified raw data spreadsheet has been made available via the Open Science Framework [28].

Authors' Contributions

JOL and LA wrote the manuscript with input from the other authors.
JOL carried out statistical analysis.
SF and LA created the preimplementation surveys and undertook training and communication webinars with participating optometry practices.
LA, SF, and SJN reviewed the symptom reports to create gold standard triage dispositions.
LA was responsible for funding acquisition.

Conflicts of Interest

LA is the developer of the eye+dot web app and unsalaried director of Cambridge Medical Innovation Ltd.

Multimedia Appendix 1
Symptom acuity table adapted from the Royal College of Ophthalmologists Emergency Eye Care Guidelines.
[PDF File, 96 KB - humanfactors_v13i1e77278_app1.pdf]

Multimedia Appendix 2
Example of an eye+dot symptom report and triage disposition.
[PDF File, 143 KB - humanfactors_v13i1e77278_app2.pdf]

Multimedia Appendix 3
Online preimplementation survey sent to optometrists registered with the Cambridgeshire Local Optometric Committee.
[PDF File, 363 KB - humanfactors_v13i1e77278_app3.pdf]

Multimedia Appendix 4
Royal College of Ophthalmologists (RCOphth) list of symptoms requiring emergency eyecare [16].
[PNG File, 85 KB - humanfactors_v13i1e77278_app4.png]

Multimedia Appendix 5
Visual Likert scale following completion of eye+dot questionnaire. The pictorial values were converted to numerical ones (least favorable=1, most favorable=5) for analysis.
[PNG File, 760 KB - humanfactors_v13i1e77278_app5.PNG]

Multimedia Appendix 6
Optometric triage outcomes and online triage tool–suggested dispositions. (A) Optometric triage outcomes. (B) Online triage tool suggested dispositions. Redder colors correspond to more urgent triage dispositions. Note that the eye+dot tool directs all urgent (within 24 h) dispositions to a national health service (NHS) hospital or minor eye condition services (MECS) services as urgent appointments are often not available privately or through the general ophthalmic service (GOS). N=209. ED: emergency department; EEC: emergency eye clinic; GOS: general ophthalmic service; MECS: minor eye condition service.
[PNG File, 60 KB - humanfactors_v13i1e77278_app6.png]

References

1. Hau S, Ioannidis A, Masaoutis P, Verma S. Patterns of ophthalmological complaints presenting to a dedicated ophthalmic accident & emergency department: inappropriate use and patients' perspective. Emerg Med J 2008 Nov;25(11):740-744. [doi: 10.1136/emj.2007.057604] [Medline: 18955608]
2. MacIsaac JC, Naroo SA, Rumney NJ. Analysis of UK eye casualty presentations. Clin Exp Optom 2022 May;105(4):428-434. [doi: 10.1080/08164622.2021.1937949] [Medline: 34315329]

XSL•FO
RenderX

3. Rehan SM, Morris DS, Pedlar L, Sheen N, Shirodkar AL. Ophthalmic emergencies presenting to the emergency department at the University Hospital of Wales, Cardiff, UK. Clin Exp Optom 2020 Nov;103(6):895-901. [doi: 10.1111/cxo.13050] [Medline: 32066198]
4. Vernon SA. Analysis of all new cases seen in a busy regional centre ophthalmic casualty department during 24-week period. J R Soc Med 1983 Apr;76(4):279-282. [doi: 10.1177/014107688307600408] [Medline: 6601713]
5. Carmichael J, Abdi S, Balaskas K, Costanza E, Blandford A. Assessment of optometrists' referral accuracy and contributing factors: a review. Ophthalmic Physiol Opt 2023 Sep;43(5):1255-1277. [doi: 10.1111/opo.13183] [Medline: 37395045]
6. The NHS long term plan. : National Health Service; 2019 URL: https://webarchive.nationalarchives.gov.uk/ukgwa/20230418155402/https:/www.longtermplan.nhs.uk/publication/nhs-long-term-plan/ [accessed 2026-03-03]
7. Eijk ESV, Bettink-Remeijer MW, Timman R, Busschbach JJV. From pen-and-paper questionnaire to a computer-assisted instrument for self-triage in the ophthalmic emergency department: process and validation. Comput Biol Med 2015 Nov 1;66:258-262. [doi: 10.1016/j.compbiomed.2015.09.014] [Medline: 26451983]
8. Eijk ESV, Bettink-Remeijer MW, Timman R, Heres MHB, Busschbach JJV. Criterion validity of a computer-assisted instrument of self-triage (ca-ISET) compared to the validity of regular triage in an ophthalmic emergency department. Int J Med Inform 2016 Jan;85(1):61-67. [doi: 10.1016/j.ijmedinf.2015.10.003] [Medline: 26526278]
9. Meer E, Ramakrishnan MS, Whitehead G, Leri D, Rosin R, VanderBeek B. Validation of an automated symptom-based triage tool in ophthalmology. Appl Clin Inform 2023 May;14(3):448-454. [doi: 10.1055/a-2065-4613] [Medline: 36990454]
10. Self-help guides. NHS Inform. URL: https://www.nhsinform.scot/symptoms-and-self-help/self-help-guides [accessed 2025-02-18]
11. Check your symptoms. National Health Service. URL: https://111.wales.nhs.uk/SelfAssessments [accessed 2025-02-18]
12. Chambers D, Cantrell AJ, Johnson M, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. BMJ Open 2019 Aug 1;9(8):e027743. [doi: 10.1136/bmjopen-2018-027743] [Medline: 31375610]
13. Miller S, Gilbert S, Virani V, Wicks P. Patients' utilization and perception of an artificial intelligence-based symptom assessment and advice technology in a British primary care waiting room: exploratory pilot study. JMIR Hum Factors 2020 Jul 10;7(3):e19713. [doi: 10.2196/19713] [Medline: 32540836]
14. Paule A, Abel GA, Parsons J, Atherton H. Patient-facing online triage tools and clinician decision-making: a systematic review. BMJ Open 2025 May 8;15(5):e094068. [doi: 10.1136/bmjopen-2024-094068] [Medline: 40345692]
15. UK Policy Framework for Health and Social Care Research. Health Research Authority. URL: https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/uk-policy-framework-health-social-care-research/uk-policy-framework-health-and-social-care-research/ [accessed 2026-03-10]
16. Commissioning guidance emergency eye care. : The Royal College of Ophthalmologists; 2020 URL: https://www.rcophth.ac.uk/wp-content/uploads/2020/05/Emergency-Eye-Care-Commissioning-Guidance.pdf [accessed 2026-02-18]
17. Khou V, Ly A, Moore L, et al. Review of referrals reveal the impact of referral content on the triage and management of ophthalmology wait lists. BMJ Open 2021 Sep 7;11(9):e047246. [doi: 10.1136/bmjopen-2020-047246] [Medline: 34493511]
18. Davey CJ, Green C, Elliott DB. Assessment of referrals to the hospital eye service by optometrists and GPs in Bradford and Airedale. Ophthalmic Physiol Opt 2011 Jan;31(1):23-28. [doi: 10.1111/j.1475-1313.2010.00797.x] [Medline: 21070302]
19. Finch TL, Girling M, May CR, et al. Improving the normalization of complex interventions: part 2—validation of the NoMAD instrument for assessing implementation work based on normalization process theory (NPT). BMC Med Res Methodol 2018 Nov 15;18(1):135. [doi: 10.1186/s12874-018-0591-x] [Medline: 30442094]
20. Finch TL, Rapley T, Girling M, et al. Improving the normalization of complex interventions: measure development based on normalization process theory (NoMAD): study protocol. Implement Sci 2013 Apr 11;8:43. [doi: 10.1186/1748-5908-8-43] [Medline: 23578304]
21. Gilbert S, Mehl A, Baluch A, et al. How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. BMJ Open 2020 Dec 16;10(12):e040269. [doi: 10.1136/bmjopen-2020-040269] [Medline: 33328258]
22. Real world evidence. National Health Service. URL: https://innovation.nhs.uk/innovation-guides/evidence/real-world-evidence [accessed 2026-02-18]
23. Wilson SA, Byrne P, Rodgers SE, Maden M. A systematic review of smartphone and tablet use by older adults with and without cognitive impairment. Innov Aging 2022;6(2):igac002. [doi: 10.1093/geroni/igac002] [Medline: 35243008]
24. Khanassov V, Ilali M, Ruiz AS, Rojas-Rozo L, Sourial R. Telemedicine in primary care of older adults: a qualitative study. BMC Prim Care 2024 Jul 17;25(1):259. [doi: 10.1186/s12875-024-02518-x] [Medline: 39020277]
25. Chan F, Lai S, Pieterman M, et al. Performance of a new symptom checker in patient triage: Canadian cohort study. PLOS ONE 2021;16(12):e0260696. [doi: 10.1371/journal.pone.0260696] [Medline: 34852016]
26. Fan MMW, Leung LP. Validation of the Hong Kong accident and emergency triage guidelines. Hong Kong Med J 2013 Jun;19(3):198-202. [doi: 10.12809/hkmj133900] [Medline: 23568939]
27. Steiner D, Renetseder F, Kutz A, et al. Performance of the Manchester triage system in adult medical emergency patients: a prospective cohort study. J Emerg Med 2016 Apr;50(4):678-689. [doi: 10.1016/j.jemermed.2015.09.008] [Medline: 26458788]

XSL•FO
RenderX

28. Excel Spreadsheet of Study Data. OSF Registries. URL: https://osf.io/3r4we/overview [accessed 2026-03-10]

Abbreviations

CUES: Community Urgent Eyecare Services
ED: emergency department
EEC: emergency eye clinic
GDPR: General Data Protection Regulation
MECS: Minor Eye Condition Services
NHS: National Health Service
RCOphth: Royal College of Ophthalmologists

Edited by A Kushniruk; submitted 12.May.2025; peer-reviewed by H Anandan, S Rathnayake; revised version received 16.Jan.2026; accepted 09.Feb.2026; published 16.Mar.2026.

Please cite as:
Lim JO, Farrell S, Newlands SJ, Allen L
Real-World Evaluation of the Eye+Dot Online Triage Support Tool in Community Optometry Practices: Mixed Methods Evaluation Study
JMIR Hum Factors 2026;13:e77278
URL: https://humanfactors.jmir.org/2026/1/e77278
doi:10.2196/77278

© Jen O Lim, Sarah Farrell, Samuel J Newlands, Louise Allen. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 16.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Aspects Supporting and Hindering Type 2 Diabetes Self-Management in Web-Based Educational Portals: Usability Testing Study With Updated Framework in Razavi-Khorasan, Iran

Javad Jafari[1,2], MD, PhD; Klas Karlgren[3,4,5], PhD; Hossein Karimi Moonaghi[2,6], MSN, MMedEd, PhD; Mohammad Reza Mazaheri Habibi[7], MSc, PhD; Saeid Eslami[8], PharmD, PhD; Italo Masiello[9], PhD; Stefano Bonacina[3], MSc, PhD

[1]Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

[2]Department of Medical Education, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

[3]Health Informatics Centre - Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Tomtebodavägen 18a, 4th Fl., Stockholm, Sweden

[4]Department of Research, Education, Development and Innovation, Södersjukhuset, Stockholm, Sweden

[5]Department of Health and Functioning, Faculty of Health and Social Sciences, Western Norway University of Applied Sciences, Bergen, Norway

[6]Nursing and Midwifery Care Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

[7]Department of Health Information Technology, Varastegan Institute for Medical Sciences, Mashhad, Iran

[8]Pharmaceutical Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

[9]Department of Computer Science and Engineering, Mälardalen University, Västerås, Sweden

Corresponding Author:
Stefano Bonacina, MSc, PhD
Health Informatics Centre - Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Tomtebodavägen 18a, 4th Fl., Stockholm, Sweden

Abstract

Background: The rising prevalence of diabetes necessitates continuous monitoring and treatment, especially for type 2 diabetes. Patient education plays a crucial role in enabling self-management, and web-based educational portals can support patients effectively. While the literature highlights various issues impacting self-management system design, few studies explore the usability aspects that either facilitate or hinder such systems for this patient group.

Objective: This study aimed to investigate usability issues related to a web-based educational portal for patients with diabetes in Razavi-Khorasan Province, Iran. Additionally, it sought to develop a framework of design principles to address critical usability concerns for diabetes self-management.

Methods: The literature on diabetes self-management was analyzed. The analysis focused on design and usability issues affecting self-management. A think-aloud study was carried out with 10 patients using a web-based educational portal designed for patients with diabetes. The portal comprises 9 sections, with patients performing tasks related to each one. Participants' task completion times were measured, and severity ratings of usability problems were calculated.

Results: Five design principles were proposed that cover (1) understanding and learning about one's condition, (2) motivation and fostering sustained practices, (3) autonomy and confidence, (4) interaction and collaboration, and (5) privacy and security. The study identified 111 usability problems related to the portal's 9 sections. Feedback from participants was considered to refine the principles, and participants' considerations were recorded for any issues that worked against the principles. A total of 16 suggestions for improvement were extrapolated from the data.

Conclusions: This study highlights critical usability and design aspects to consider when developing web-based educational portals to support self-management in type 2 diabetes.

(JMIR Hum Factors 2026;13:e78903) doi:10.2196/78903

KEYWORDS

diabetes mellitus, type 2; self-management; patient portals; user-centered design; patient education as topic; web-based learning

XSL•FO
RenderX

Introduction

Background

Diabetes is a chronic disease rapidly increasing in incidence; in 1990, the number of people with diabetes was 200 million but rose to 830 million in 2022, and in low- and middle-income countries, the prevalence has been rising more rapidly than in high-income countries [1], with type 2 diabetes (T2D) accounting for 90% of all diabetes [2]. As there is no cure for diabetes, solutions have focused on preventing complications by educating patients, more recently through online interventions. Web-based patient education interventions may help improve T2D self-management [3]. The increased availability of the internet and smartphone apps targeting patients with diabetes can offer opportunities and improvement [4]. Websites and mobile apps are readily available, cost-effective, and easy for patients to use anywhere and at any time [5]. Furthermore, studies have also shown that internet-delivered diabetes education is easier for patients to access than traditional education [6].

Today, there is a growing number of websites and other forms of online patient education, especially for patients with diabetes [7], and digital diabetes self-management education has been shown to improve patients' knowledge and clinical outcomes [8]. Many diabetes apps offer a selection of similar functions necessary for diabetes self-management [9]. With the right technical skills, a patient with T2D can take advantage of online patient education and services.

Computer literacy has been defined as "basic skills in using a computer, for example, to save and open a file, use a word processing program, and send and receive e-mail" [10]. Similarly, eHealth literacy is defined as the "ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem" [11]. A study recently conducted in Iran concluded that the level of eHealth literacy regarding health technologies is increasing [12]. In addition, according to statistics presented by the Global Economy, approximately 82% of Iran's 85 million inhabitants were using the internet in 2022 [13], while in 2012, the number was only 53.3% (as identified in the data accessed in 2014 for previous research referenced in a study by Jafari et al [14]). This reveals the rapid increase in the use of the internet in the Iranian population and the need for computer and eHealth literacy for this T2D population to extract the benefits of online patient education and services.

Designing Tools for Self-Management

Self-care focuses on the prevention of a disease and the maintenance of well-being, while self-management focuses on coping with the disease and controlling the clinical parameters [15]. Research shows that the digital health industry has a promising future [16]. Technology can be used to enhance or regulate people's emotions, motivation, and well-being [17]. Proactive communication by health care teams can help patients foster autonomy and encourage them to develop better self-management skills [18]. In addition, technology can support patients with chronic conditions by allowing them to increase their control of the disease [19]. There is a need for a supplementary design in developing supportive health care technology, especially for older people with diabetes [20]. However, it is not obvious how all these issues should be implemented in practice when designing web-based resources for the group of patients with T2D.

When self-managing the disease, patients should take responsibility for interacting with health care providers and have the ability to monitor their own condition [15]. Technology designed for diabetes self-management can be helpful in terms of improving blood sugar control, but it is necessary for designers to know how patients use the technology in their daily routines and over an extended period [19].

Quality of life (QoL) is defined by the World Health Organization as "an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards, and concerns" [21]. QoL is fundamental for human-centered design. In human-centered design, trade-offs are general and unavoidable [21]. Design trade-offs are crucial in any design process, particularly as the digital age continues to evolve in the context of T2D self-management. A "trade-off" is a condition that involves losing one aspect or quality of something in return for gaining another aspect or quality [21]. Trade-offs are associated with matching problem solving and problem framing [22]. Designers need to explore QoL [21] and act in the best interests of all stakeholders, including patients. To achieve this, it is necessary to understand the strengths and weaknesses of the technology to be designed [23]. Technology has resulted in progress, but technology adaptation has not improved QoL for patients [22]. An analysis of design trade-offs in developing self-management systems for patients with heart failure concluded that design trade-off analyses can improve understanding of health self-management technologies [24], and analogous analyses are needed for T2D.

Designing Web-Based Educational Portals

A web-based educational portal is an online tool that allows patients to interact with health care systems and health care providers and follow treatment recommendations for improving their clinical outcomes, while it can also promote preventative measures [25]. In our previous research, a prototype portal was designed and tested for usability in an Iranian context. The development of the web-based educational portal and its preliminary testing has been described in other studies [14,26-28].

Usability is often defined as the "extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [29]. Usability testing is a technique that involves testing products or websites with participants drawn from the target population [30]. Usability testing becomes a crucial part of developing novel health information systems and includes evaluating the degree to which technology fulfills basic usability criteria as the participants perform representative tasks [31]. In addition, usability testing is effective at identifying obstacles and system design issues; therefore, it is highly recommended when developing new health information applications [32]. It is not well known how to design a learning portal to best support

XSL•FO
RenderX

patients with T2D in improving their self-care management in an Iranian context—which aspects are important to the target group and how should design trade-offs be managed?

Study Aim

The aims of this study were to (1) draft a framework of design principles for tools for self-management, (2) explore the usability issues of a web-based educational portal for patients with diabetes, and (3) update the framework of design principles for capturing usability issues important for diabetes self-management.

Manuscript Organization

In the "Methods" and "Results" sections, we present the development of a theoretical framework with several design principles for exploring usability issues related to diabetes self-management. Then, we present how usability testing was conducted on a web-based educational portal and the findings that have led us to propose a modified framework with an added design principle. Finally, these findings and the framework are discussed and related to other findings in the literature in the "Discussion" section.

Methods

Approach to Identifying Design Principles for Diabetes' Self-Management

We conducted an analysis of our previous work on designing a web-based educational portal to support diabetes self-management [14,26-28] and the literature considered for its development [14,26-28]. The portal was developed for the local context of Razavi-Khorasan, considering regional health behaviors and user preferences [28]. To derive the design principles, we systematically mapped the portal's functionalities to the 5 layers of influence as outlined in the Ecological Model of Health Behavior [33]: intrapersonal (personalized education and self-monitoring tools to enhance confidence and control), interpersonal (support for family involvement and social engagement), organizational (integration with health care institutions and professional input), community (adaptation to cultural norms and local resources), and public policy (highlighting the need for culturally adapted digital health solutions in low- and middle-income countries). Finally, through a structured, consensus-based approach involving iterative expert panel discussions among the authors and the synthesis of evidence, we consolidated these insights into a set of design principles for web portals that facilitate effective diabetes self-management.

Study Design and Procedure of the Usability Testing

To describe the usability test, we applied the terminology and the checklist proposed in a study by Martins et al [34]. Following the checklist, we detail the usability assessment moderator, participants, usability evaluation method and usability evaluation technique, tasks, and usability evaluation environment. The checklist filled out for our study is presented in Multimedia Appendix 1. MRMH served as the usability assessment moderator and met the inclusion criteria by having prior experience conducting usability tests. He was external to the portal development team, guaranteeing independence from the product's design and development process. Finally, no observers were included in the study, and therefore, no observer responsibilities or characteristics were collected.

A sample size of 10 participants was considered appropriate for this usability evaluation based on previous studies [35-37]. They indicate that a relatively small number of participants is sufficient to uncover most usability issues, with Nielsen and Landauer's model suggesting that 5 participants can identify approximately 80% of usability problems [36]. Extending the sample to 10 participants further increases problem coverage and reliability. Faulkner [37] reported that 10 users can identify approximately 95% of usability issues across typical tasks while maintaining feasibility in formative testing. The inclusion criteria for participation were as follows: having been diagnosed with T2D more than 1 year prior to this study, not having serious diabetes complications, having more than 9 years of formal education, being physically and cognitively able to participate personally in the study, being able to use a computer and internet accessibility, having no experience in using the portal, and, finally, being willing to participate. Participants (ie, patients with T2D) were recruited through convenience sampling from Mashhad Diabetes Clinic. Participant characteristics gathered for the study comprised age, sex, educational level, digital literacy, and the duration of the disease.

We used a single usability testing method using a concurrent think-aloud technique [38], complemented by screen and interaction logging to capture user events (eg, clicks, navigational paths, and errors), without incorporating additional inquiry methods or techniques. In the think-aloud usability test, participants were asked to verbalize their thoughts while completing predefined tasks, allowing insight into their real-time reasoning and difficulties. All sessions were recorded using the computer's built-in microphone, and participants' verbalizations were transcribed verbatim in Persian. To document screen activity and user interactions, we used Camtasia Studio (version 8.4; TechSmith Inc.), which enabled synchronized recording of both the event flow and participants' spoken commentary. Participants completed 9 tasks corresponding to the 9 sections of the web-based educational portal developed by JJ [39], after which they provided severity ratings. Each task targeted a distinct portal section and was performed using the think-aloud technique. Table 1 lists the 9 tasks included in the study. Participants were asked to verbalize their thoughts, feelings, opinions, and experiences while completing the tasks [40]. In individual sessions, participants were first introduced to how the think-aloud method would be used and then received a brief introduction about the portal before the usability test by MRMH. A scenario (Instructions to participants, Multimedia Appendix 2) was prepared based on the results of our previous research [26,27] and was pilot tested with 2 patients prior to the real usability testing. Participants used a laptop running Windows 7, and their activities during the usability tests were video recorded and audio recorded (as abovementioned). During the sessions, the usability assessment moderator (MRMH) sat beside participants to encourage continuous verbalization and to provide assistance if needed—for example, in cases where

XSL•FO
RenderX

instructions were unclear, medical terminology was difficult to interpret, or a technical glitch occurred.

Table . The 9 tasks and their explanations, including the collected data or material.

Task name	Task explanation
Fill in patient information	The user is asked to type or select from a drop-down list. The data are entered into the system using the keyboard and mouse. They include name and surname, gender, age, education level, occupation, address, weight, height, type of diabetes, age of onset of diabetes, and associated specific disease.
Register laboratory results	The user is asked to type laboratory result values and read medical documentation. FBS[a], 2hpp[b], 5PM[c], HbA1c[d], liver function tests (SGOT[e], SGPT[f], ALP[g]), lipids profile (cholesterol and triglyceride), kidney function tests, urine sugar, BUN[h], creatinine, and blood pressure.
Attending educational courses	The user is asked to attend a course aiming to help them improve their diabetes self-management. Eight courses are available, and they were asked to attend one of these courses. Each course contains a short quiz, which is answered before beginning the course.
Read educational material	The user is asked to read educational material about various aspects of diabetes and its effects on organs and the relationship between stress, physical activity, drugs, and disease progression.
Use the “help-me” steps	The user is asked to follow the “help-me” flow or to follow sequential steps to access specific information.
Check health improvement	The user is asked to access data about their current situation based on their medical condition and the laboratory results recorded in the medical record.
Play educational movies	The user is asked to find and watch one short movie about diabetes for at least 1 minute, regarding diabetes complications, diets tailored for patients with diabetes, or other related subjects. Movies are about 3 to 4 minutes long.
Calculate health indices	The user is asked to use a food calorie counter and a calculator to check the BMI and the glycemic index of different types of food. The task was composed of one calculation based on the user’s interest.
Chat with physicians and nurses	The user is asked to use the chat to communicate with physicians or nurses.

[a]FBS: fasting blood sugar.

[b]2HPP: sugar level 2 hours after breakfast.

[c]5PM: 5-afternoon blood sugar.

[d]HbA_{1C}: hemoglobin $_{A1c}$.

[e]SGOT: serum glutamic oxaloacetic transaminase.

[f]SGPT: Serum glutamic pyruvic transaminase.

[g]ALP: alkaline phosphatase.

[h]BUN: blood urea nitrogen.

As for the usability evaluation environment, the study was conducted under controlled conditions in a room located in the Medical Informatics Department of the Faculty of Medicine of Mashhad University of Medical Sciences, Mashhad, Iran. This setting was selected because it provided a quiet, interruption-free environment that ensured consistent testing conditions across all participants.

Data Analysis

The analysis phase integrated quantitative and qualitative components: quantitatively, we computed time on task (summarized with descriptive statistics); qualitatively, think-aloud verbalizations and moderator notes were subjected to qualitative content analysis [41], with iterative code development. A particular focus was on the observations of the participants’ behaviors and their comments related to self-management and the themes of the framework discussed above. Attention was paid to how the web-based educational portal supported or hindered patients’ self-management activities (eg, due to lack of functionality). The time the participants spent on each task and the issues encountered were collected by Camtasia and analyzed. Descriptive statistics, including means and SDs, were calculated for the time taken by participants to complete the tasks. The usability tests were transcribed, and the resulting data were analyzed through qualitative content analysis [41]. Following the completion of the think-aloud sessions, the severity of identified usability problems was assessed by 3 reviewers: 2 members of the team involved in developing the web-based educational portal (AJ and NF; see Acknowledgments) and the first author (JJ). Severity ratings

XSL•FO
RenderX

were assigned using Nielsen's standardized 0 - to 4-point scale [42], where 0 indicates "not a usability problem," 1 indicates "cosmetic problem only," 2 indicates "minor usability problem," 3 indicates "major usability problem," and 4 represents a "usability catastrophe" that must be resolved prior to release. Ratings were based on 3 factors: frequency (ie, how often the problem occurred), impact (ie, the extent to which the problem affected task completion), and persistence (ie, whether recovery was easy or difficult). The 3 reviewers (coders) worked independently, and coding agreement on severity scores was reached by consensus.

Ethical Considerations

All participants received oral and written information about the study and signed a consent form. They were informed about how the gathered data would be used, that their identities would remain anonymous, and that they had the right to withdraw from the study without having to explain why at any time, if they wished to do so. Participants did not receive any compensation. The Ethics Committee of Mashhad University of Medical Sciences approved the study (IR.MUMS.REC. 1395.108; Multimedia Appendix 3).

Generative Artificial Intelligence Disclosure

Microsoft Copilot has been used to shorten the abstract to the allowed word limit. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Results

Overview

The results are presented in 3 sections. First, the first iteration of the framework ("Design Principles for Diabetes' Self-Management") is outlined, and then, quantitative and qualitative results from the usability tests are explained, and finally, a revised framework is detailed pertaining to design and usability issues in the context of patient self-management of T2D.

Design Principles for Diabetes' Self-Management

Initial Version of the Framework

By the approach explained in the *Methods* section, we proposed 4 design principles for diabetes self-management. The first principle concerns supporting patients' *understanding and learning about the complexities of their condition*. This principle is about creating awareness and supporting patients' learning regarding how to manage their diabetes, including daily measurement routines, diet planning, exercise, medication, illness, alcohol consumption, and stress. When patients do not understand their condition and the importance of managing it, adequate self-management is at risk and is likely to be discontinued. Patients require support on different levels, ranging from performing procedures, such as measurements, to reflecting on how to plan their lives on an overall level to maintain adequate self-management. Patients with diabetes need opportunities to learn specific skills, and environmental and policy contexts influence the psychological factors involved in diabetes management [33]. Opportunities to learn about measuring blood sugar, planning diets, exercising, and addressing challenges specific to diabetes are helpful for supporting self-management and managing diabetes [33].

The second design principle relates to supporting *motivation and fostering sustained practices*. This principle highlights the importance of supporting patients in keeping up their self-management over time and introducing measures to help them avoid skipping self-care practices or giving up entirely. Encouraging, reminding, supporting, and "nudging" can provide the motivation that patients may need to continue their self-management and not become discouraged or distracted. Plans, calendars, schedules, timely reminders, praise, and even rewards could be elements that contribute to keeping patients motivated and supporting them in modifying their everyday practices. Providing updated information, especially information tailored to the individual patient, and giving feedback may maintain patient interest and thereby help them continue to pursue self-management plans over time [33]. Technologies that suggest care activities or adjustments to current treatments based on information on improved care and using these technologies seem to be positive for patients [19]. Mobile apps help patients to self-manage the disease and motivate patients to improve glucose control [4]. Researchers based on numerous studies "recommend designing for continuity in time, space, and the aesthetic dimension and recommend focusing on understandability and learning" [20].

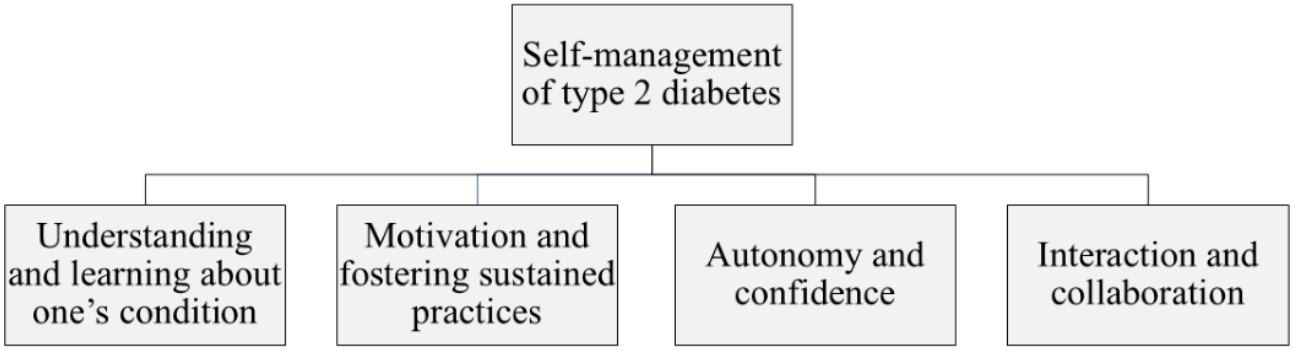
The third design principle concerns the patient's feelings of *autonomy and confidence*. Patients do not only need knowledge and reminders but also need to feel confident and secure when it comes to being able to manage their condition on their own. They need to feel that they can manage their T2D without risking their health and know that they can find and obtain help when needed. It is therefore essential to help patients develop a sense of being in control with clear instructions and, at the same time, help them understand that they have the ability to independently plan their self-management without feeling that their role is just to follow instructions. In previous research, patients with heart conditions have expressed the need to acquire knowledge while also building up their confidence to be effectively involved in decision-making concerning their treatment [43]. Technology, if well designed, can be used to contribute to making patients feel more confident about their disease management, for example, by monitoring patient vitals, giving advice on lifestyle changes, and providing relevant patient education materials [19].

The fourth design principle concerns *interaction and collaboration* with other patients and health care personnel. Supporting contact and communication between patients and health care providers is a key part of supporting patients' self-management [33]. Technology can be a vehicle for sharing information between patients and their caregivers and can provide support. However, technology can also create opportunities for patients to share care activities with informal carers and can contribute to engaging in self-management activities and learning from others with the same chronic condition and enhance collaboration with the care team [19]. Being able to connect with others in a similar situation may be an urgent requirement. Sharing experiences and taking part in

XSL•FO
RenderX

the experiences of others are strong motivators for using social media and are also powerful forces that can be of interest when designing tools for self-management. Figure 1 shows the 4 design principles. These principles draw attention to issues that are relevant for supporting—or hindering—T2D patients' self-management. These principles have served as a background and lens for further developing our framework.

Figure 1. The 4 design principles of diabetes self-management. Results from data collected from the usability testing of the portal.



Description of the Participants' Characteristics

Ten patients from a diabetes outpatient clinic participated in the study. The study participants were 5 males and 5 females with T2D, and the average age was 46.5 (SD 6.4) years. Participants had a mean disease duration of 7.9 (SD 4.5) years from the time of diagnosis. Seven of the 10 participants had received an undergraduate level education (9 - 11 y of education), and 3 had completed postgraduate education (14 - 16 y of education).

Time Spent on Different Tasks and Severity Ratings

The 10 participants completed the usability tests in a mean time of 2068 seconds (34 min and 28 s; Table 2; SD 212.8). The first task, "Fill in patient information," took 513 (SD 131.1) seconds (8 min and 33 s) on average to complete, while it took 442 (SD 33.0) seconds (7 min and 22 s) to complete the second task, "Register laboratory results." The first 2 tasks took longer to complete than the other tasks. Two other tasks on which the participants spent more time were "Calculate health indexes" (task 8) and "Use the 'Help-me' steps" (task 5). Table 2 presents the times that each participant spent completing the tasks and the average time and the SD for each task. In addition, it presents "range" and "range width" for each task completion. As mentioned, participants spent more time on task 1 (mean 513, SD 131.1 s; 25%) and task 2 (mean 442, SD 33 s; 21%) and the least time on task 6 (mean 91, SD 19.7 s; 4%), "Check health improvement," and task 4, "Read educational material" (mean 82, SD 81.5 s; 4%).

XSL•FO
RenderX

Table . Duration of task completion for each user.

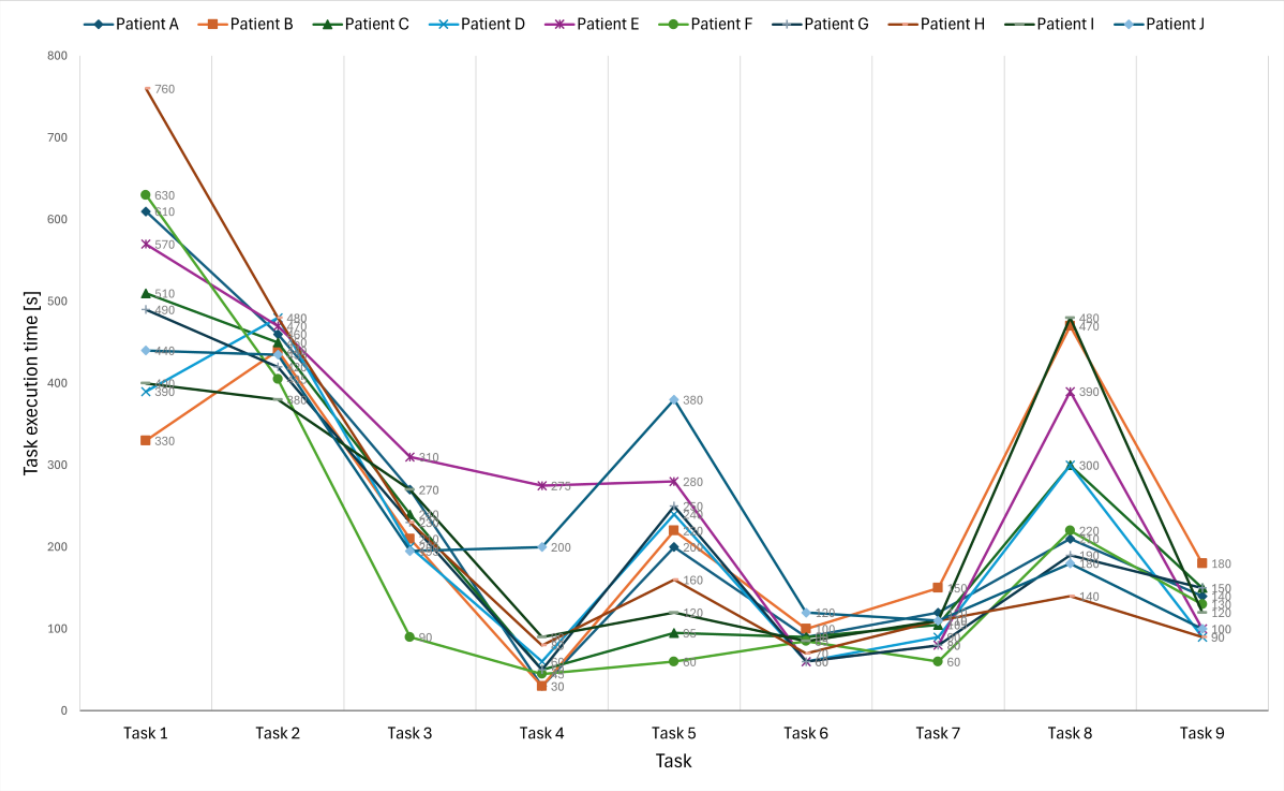
	Task 1—Fill in patient information		Task 2—Register laboratory results		Task 3—Attending educational courses		Task 4—Read educational material		Task 5—Use the "Help-me" steps		Task 6—Check health improvement		Task 7—Play educational movies		Task 8—Calculate health indexes		Task 9—Chat with physicians and nurses		Test duration	
	Value (s)	%	Value (s)	%	Value (s)	%	Value (s)	%	Value (s)	%	Value (s)	%	Value (s)	%	Value (s)	%	Value (s)	%	Value (s)	%
Participant A	610	28.6	460	21.6	270	12.7	30	1.4	200	9.4	90	4.2	120	5.6	210	9.9	140	6.6	2130	100
Participant B	330	15.5	440	20.7	210	9.9	30	1.4	220	10.3	100	4.7	150	7.0	470	22.1	180	8.5	2130	100
Participant C	510	25.6	450	22.6	240	12.1	50	2.5	95	4.8	90	4.5	105	5.3	300	15.1	150	7.5	1990	100
Participant D	390	20.4	480	25.1	200	10.5	60	3.1	240	12.6	60	3.1	90	4.7	300	15.7	90	4.7	1910	100
Participant E	570	22.5	470	18.5	310	12.2	275	10.8	280	11.0	60	2.4	80	3.2	390	15.4	100	3.9	2535	100
Participant F	630	36.5	405	23.5	90	5.2	45	2.6	60	3.5	85	4.9	60	3.5	220	12.8	130	7.5	1725	100
Participant G	490	25.5	420	21.9	230	12.0	50	2.6	250	13.0	60	3.1	80	4.2	190	9.9	150	7.8	1920	100
Participant H	760	35.8	480	22.6	230	10.8	80	3.8	160	7.5	70	3.3	110	5.2	140	6.6	90	4.2	2120	100
Participant I	400	19.5	380	18.5	270	13.1	90	4.4	120	5.8	85	4.1	110	5.4	480	23.4	120	5.8	2055	100
Participant J	440	20.4	435	20.1	195	9.0	200	9.3	380	17.6	120	5.6	110	5.1	180	8.3	100	4.6	2160	100
Time																				
Average	513	25.0	442	21.0	224.5	11.0	91	4.0	200.5	10.0	82	4.0	101.5	5.0	288	14.0	125	6.0	2068	—
SD	131.1	—[a]	33.0	—	59.2	—	81.5	—	95.4	—	19.7	—	25.2	—	122.3	—	30.3	—	208	—
Range	330 - 760	—	380 - 480	—	90 - 310	—	30 - 275	—	60 - 380	—	60 - 120	—	60 - 150	—	140 - 480	—	90 - 180	—	1725 - 2535	—
Range width	430	—	100	—	220	—	245	—	320	—	60	—	90	—	340	—	90	—	810	—

XSL•FO
RenderX

[a]Not applicable.

Figure 2 presents a comparison of the duration of task completion for each user.

Figure 2. Comparison of the duration of task completion for each user.



Five tasks were expected to be particularly informative and relevant to the self-management of diabetes:

- Task 3: “Attending educational courses”; take one pretest and look at the courses.
- Task 4: “Read educational material” about diabetes and body organs.
- Task 5: ”Use the ‘Help-Me’ steps” module and follow the steps.
- Task 7: “Play educational movies” and watch one educational video.
- Task 8: “Calculate health indexes,” namely, BMI and daily calorie intake.

Analysis of tasks 3, 4, 5, 7, and 8 showed that participants reflected during task completion (Table 2 and Figure 2) and provided both feedback and opinions. For example, participants pointed out their interest in finding trustworthy and helpful information, educational materials, and short educational videos about diabetes. In addition, they expressed an interest in finding information that could help them solve diabetes-related complications. Participants pointed out the importance of knowing about healthy foods in all food groups and calculating their daily calorie intakes to create healthy-eating plans relevant to their daily activities. These topics are important for diabetes self-management.

Table 3 presents the number of usability problems participants encountered and the problems that were detected in this study according to severity ratings, which range from 0 to 4, determined by the researchers involved. On the basis of the think-aloud method and the qualitative content analysis, 111 problems were detected that show strengths and limitations derived from the usability testing, and some were repeated by participants. Most usability issues (n=46) were classified as minor with a severity rating of 2 (Table 3), whereas 10 issues were rated as severity 4, indicating catastrophic problems. In addition, the participants encountered 14 issues that were not considered usability problems.

XSL•FO
RenderX

Table . Usability problems were identified by users in terms of tasks and their severity ratings.

Tasks	Severity, n (%)					Total usability problems per task, n (%)
	0	1	2	3	4	
Fill in patient information	8 (21)	10 (26)	10 (26)	6 (16)	4 (11)	38 (34)
Register laboratory results	4 (15)	4 (15)	14 (52)	2 (7)	3 (11)	27 (24)
Attending educational courses	1 (20)	1 (20)	1 (20)	2 (40)	0	5 (4)
Read educational material	0	3 (100)	0	0	0	3 (3)
Use the “Help-me” steps	0	0	3 (75)	1 (25)	0	4 (4)
Check health improvement	0	0	0	0	2 (100)	2 (2)
Play educational movies	0	0	3 (75)	1 (25)	0	4 (4)
Calculate health indexes	0	1 (8)	7 (59)	3 (25)	1 (8)	12 (11)
Chat with physicians and nurses	1 (6)	5 (31)	8 (50)	2 (13)	0	16 (14)
Total usability problems	14 (13)	24 (22)	46 (41)	17 (15)	10 (9)	111 (100)

Participants’ Suggestions for Improving the Portal

In total, 16 suggestions on how to improve the web-based educational portal were noted. In Table 4, the suggestions are sorted in ascending order based on their frequencies.

Table . Suggestions for the improvement of the portal.

User suggestions	Frequencies, n (%)
Need to have a user guide or guideline for the portal	9 (19.1)
Use appropriate messages to interact with the user	8 (17.0)
Embed the list of physicians and nurses in the online chat	5 (10.6)
Use of larger font and readable colors	5 (10.6)
Mark clearly mandatory fields	3 (6.4)
Add more educational content	3 (6.4)
Offer a diet plan	3 (6.4)
Change the titles of the calculator option	2 (4.3)
Do not use medical terms or jargon	2 (4.3)
Classify the content according to patients’ literacy level	1 (2.1)
List specific diseases that a user can select from a list	1 (2.1)
Complete the educational courses	1 (2.1)
Use more pictures and graphs in the educational materials	1 (2.1)
Use simple options to calculate BMI	1 (2.1)
Use more simple options to calculate calories	1 (2.1)
Add an option for booking time to see their physician	1 (2.1)

XSL•FO
RenderX

A Framework of Design Principles

A large number of observations were made regarding the specific complexities and challenges of using the web-based educational portal, which could affect the self-management of patients with T2D. One type of observation concerned difficulties in terms of finding information and following instructions due to poor usability. Another issue related to poor computer literacy adversely affected the possibility of using the portal for self-management due to difficulties with rather basic interactions with buttons, navigation, and finding information in the portal. In some cases, patients needed more time to complete the tasks than expected due to their lack of familiarity with the portal. Various bugs and errors in links, buttons, and forms were also noted, impacting the portal's usability. The first section below presents findings relating to the framework of design principles and is followed by data regarding the time spent on the different tasks (explained in Table 1), severity ratings, and suggestions from participants. In total, 111 codes were derived from analyzing the data that pertained to how the design of the portal supported or hindered self-management behaviors. The results are presented together with the framework of 5 design principles (Figure 1).

Regarding the first design principle about supporting patients in understanding and learning about the complexity of the condition, participants emphasized the importance of finding information and pointed out that interaction within the portal could help them learn more about self-management skills. For example, they expressed that they wanted to learn how to calculate their calorie intake and diets and how to use the BMI calculator. Participants appreciated the educational content as being helpful for dealing with self-management problems but also asked for more information about topics that they would have liked to learn more about, such as diet plans and the role of stress:

> *Here [points to the educational section of the portal], educational content is related to diabetes and body organs that users can [use to] solve their problems.*
>
> *I think it is great; this information helps us know what we have to do because a diabetic patient has a lot of problems.*
>
> *There is nothing written about stress and how it affects the body—I know that stress has a big effect on me.*

While participants appreciated learning about T2D, they pointed out that they had difficulties understanding certain medical terminology, language, and abbreviations on the portal. For 8 of the participants, the term "urine sugar" (glycosuria) was unclear, and for 4 of them, "urea in the blood" (blood urea nitrogen) presented a comprehension problem. All the laboratory items were in English, and it was difficult for participants to enter data about these items. In addition, the use of the abbreviation "BP" for blood pressure was not obvious to users during data entry.

Moreover, participants had difficulties navigating the web-based educational portal to make use of the educational material and in knowing what to do with the information. Six participants had difficulties finding the educational courses and pretest questions on the computer screen. In addition, it was not clear what to do after answering the pretest questions, or, in some cases, the proposed answers were not fully understood. One participant also had an issue with the font size of the text (too small), and 1 patient experienced difficulties accessing the educational course. One participant complained about slow scrolling when studying the educational material. Another participant also complained about the font size during this task. These usability issues risked working against the design principle of supporting understanding and learning. Suggestions from participants included avoiding medical "jargon" and classifying content according to patients' literacy levels.

The second design principle emphasizes supporting motivation and fostering sustained practices. Regarding reminders about how to go about carrying out measurements in their daily lives—such as calculating calories and creating diet plans that are easy to implement—participants expressed that functionalities of the web-based educational portal could help them modify their everyday practices. Features such as reminders, calendars, schedules, rewards, and praise were considered functions that could support participants' motivation by "nudging" them into trying new behaviors and maintaining practices over time. Several participants brought up how educational materials about diabetes could support them in maintaining their interest in the portal. They also expressed how accurate and up-to-date facts about their condition would make them feel that the portal was a place they could return to for resources and advice and something that could assist them in not giving up on self-management. Participants appreciated the idea of finding information about treatments, and knowing that they would find such information on the portal would encourage them to return to consulting the portal. A function that was appreciated was the opportunity to check one's health improvement based on laboratory results, but as the results came from a 3-month data collection period, the participants were not able to see the curve charts on the health progress screen and were therefore not able to test it fully:

> *If that was the case, for example, I would input the food and the site would tell me how much to eat based on my height and weight.*

Several comments and observations highlighted the risks of working against this design principle. As mentioned earlier, participants expressed that they wanted to learn how to calculate their calorie intakes and plan their diets, but they also pointed out that they were hesitant to insert information about their treatment or repeatedly import their weight and height into the BMI calculator, as it was experienced as burdensome. Several interactions, especially relating to inserting data, were in fact experienced as troublesome and referred to as "frustrating." Four participants had difficulties logging in, and 6 participants found it complicated to register, get login codes and passwords, or enter registration information. For all participants, it was unclear what to write in the "patient-specific illness" box in the enrollment portion of the portal, and 4 participants had difficulty entering data into the data entry boxes. Five participants complained about the font size, and 4 of them did not know the date of the onset of their illness, making interaction difficult.

XSL•FO
RenderX

All such experiences were considered annoyances that would lessen their enthusiasm in terms of using the web-based educational portal over time. The participants also experienced difficulties entering laboratory test results, which is a recurring activity that should feel easy to carry out. Eight participants experienced problems entering the value of their "sugar level 2 hours after breakfast (2hpp)." Nine participants had difficulties entering the results of the liver tests such as serum glutamic oxaloacetic transaminase, serum glutamic pyruvic transaminase, and alkaline phosphatase, their five in the afternoon blood sugar (5 PM), and the protein in their urine. Six participants had problems entering the value for serum creatinine, and 3 participants had issues entering their cholesterol and triglyceride levels. A default date was included at the end of the laboratory tests, and it made it difficult for participants to enter the actual laboratory date on the portal.

The third design principle concerned autonomy and confidence. As soon as the participants had understood how to use the web-based educational portal, they explained how it made them feel more secure and willing to continue using it. Participants commented on having mastered carrying out certain steps that they were unable to do before. Being able to use the web-based educational portal without the help of others was reported as satisfying:

> *I know what I have to do in this step. I did this step and can do the next step; others cannot do the tasks the way I'm doing them.*
>
> *This is the one I can do, and I think it is not difficult for me. It is not necessary to ask someone and get help.*

The step-by-step instructions were appreciated and seen as helpful by the participants in learning how to make use of the portal's features. In contrast, participants were disappointed when they were not able to follow the sequential steps of the Help-Me screen to get the necessary information, and they called for an additional step that would more clearly show them how to complete the task. More user guides were suggested as being potentially helpful for building up competence in fully making use of the intentions of the portal.

The fourth design principle highlights the importance of interaction and collaboration with others. An online chat function is provided by the portal, which enables communication between patients and physicians or nurses. Having the possibility to contact health care professionals was viewed as important, and one of the participants commented as follows:

> *I can find the information that I need on this portal. It is great to have contact with nurses and doctors through this portal.*
>
> *How good it is to be able to contact a nurse online! Well, I do not see anything like a doctor here. It is easier to have the doctor's name. It would also be interesting to talk to people who have the same problem as us.*

Participants suggested extensions regarding how they could get in touch with health care professionals, for example, by adding the ability to book a time to meet their physician through the web-based educational portal. Some participants had difficulties finding the names of the physicians or nurses on the lists that were available online, and one suggestion was to embed such a list in the chat function. One concern that was voiced was that the text-based interaction between nurses and physicians should occur on an appropriate, understandable level. Moreover, the distinction between scheduled online face-to-face meetings and offline, or asynchronous, communication with physicians or nurses was unfamiliar to the participants.

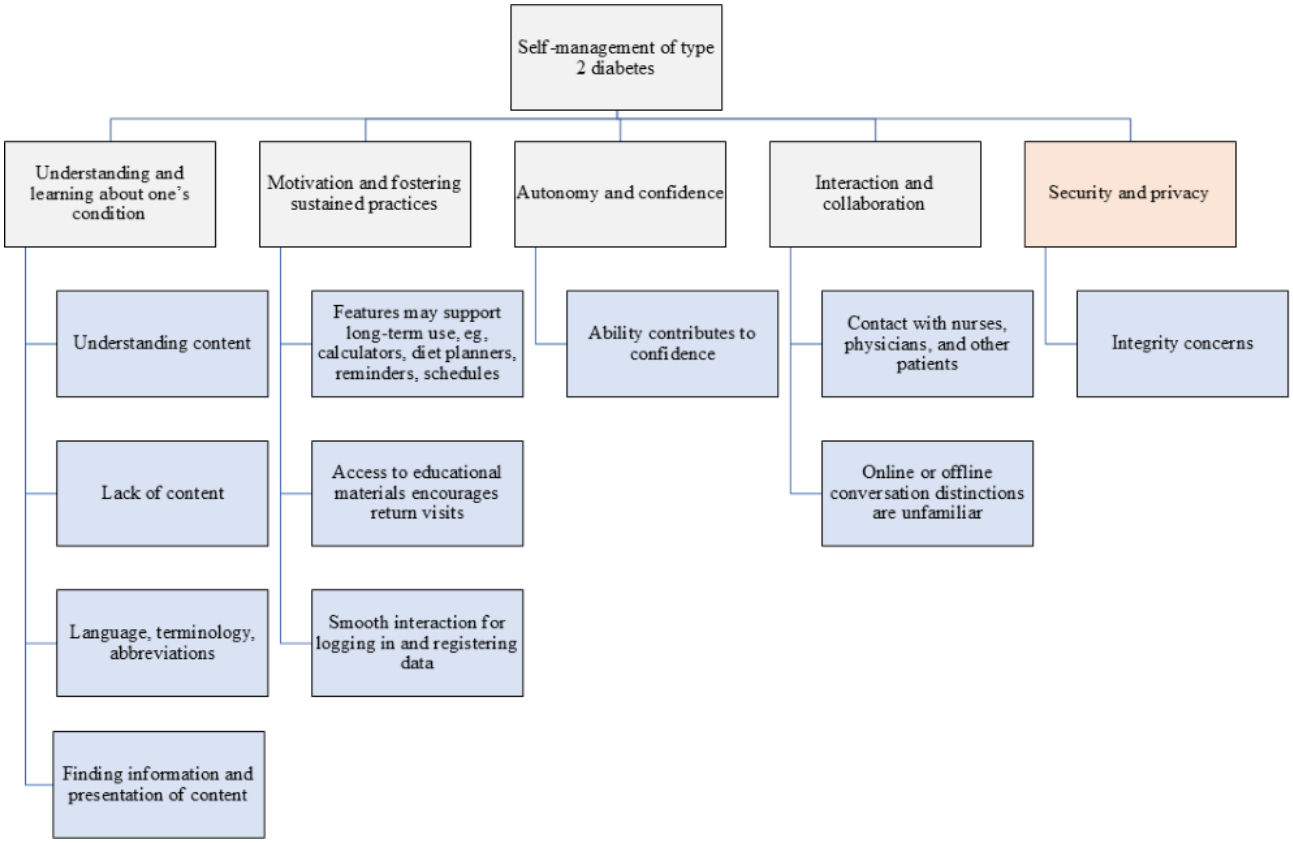
In addition, several observations were made that did not match the framework's 4 principles. A common thread was these observations related to the participants' concerns about their integrity and trusting the system with health data. Participants expressed that they preferred not to share certain information and were uncomfortable being prompted to insert such information. Therefore, the addition of a fifth design principle to the framework relating to security and privacy issues was suggested. Some participants were hesitant about inserting data about their treatments, weight, and height, and there was another aspect to this hesitance. Two participants were unwilling to enter personal information and expressed concerns about security and information privacy and wondered who would be able to see and access the personal data uploaded into the system. The participants referred to the fact that their health information would be in a system that is online as the reason behind their worries. Such concerns, therefore, counter the goal of using the portal. Regardless of whether the users are able to make use of the intended functionalities provided by the system, a lack of confidence in the system risks undermining the overall use of the portal:

> *Is it necessary to enter all this information to register? It would be better if the icons had a star to show which ones are necessary. Because I don't like that, and I don't register myself on some sites that take all my personal information.*

Figure 3 presents the updated framework, which includes 5 principles.

XSL•FO
RenderX

Figure 3. The updated framework.



Discussion

Principal Findings

The principal results are structured as follows: a brief summary of the usability testing is provided first, followed by a detailed discussion of the proposed framework.

The usability testing revealed issues affecting user performance and understanding across the evaluated 9 tasks. Participants successfully completed most tasks, but tasks 1 “Fill in patient information” and 2 “Register laboratory results” required more time and caused confusion about terminology (also because laboratory variables were expressed in English). Think-aloud responses showed that participants reflected on unclear instructions and insufficient system feedback for tasks 3, 4, 5, 7, and 8 (Table 2 and Figure 2). Using Nielsen’s severity rating scale, we identified 46 minor usability problems with a severity rating of 2 and 10 major problems with a severity rating of 4 (Table 3). Minor problems were primarily related to tasks 2, 1, and 9, whereas major problems were related to tasks 1 and 2, related to insufficient support for self-management actions and unclear terminology. Suggestions on how to improve the web-based educational portal were noted (Table 4) and were related to providing a user guide and adopting a suitable terminology for the user. Positive feedback highlighted the usefulness of certain information modules (see section “A Framework of Design Principles”). Overall, these findings guided the refinement of the framework by contributing an additional design principle and providing further detail to existing principles.

This study has proposed a framework of design principles and investigated design and usability aspects of a portal aiming to support the self-management of patients with T2D. The design principles emphasize aspects that are important for the self-management and can guide design work. Design principles have been described as providing a bridge between theory and practice while simultaneously playing several possible roles [44]. They can, for example, be a means of explaining, communicating, and discussing criteria for making selections; suggestions for developing existing practices; a framework for reporting findings; or ideas regarding tool functionalities. In this study, the design principles have fulfilled several of these roles and have provided a way of focusing on aspects of the web-based educational portal that contribute to the overall goals of self-management, as well as to the study and evaluation of the proposed solutions.

These aspects are expected to be crucial to consider when designing tools for T2D self-management, and many have been previously discussed in the literature [4,17-20,33]. Here, the principles have helped us concretize these aspects to empower a specific patient group. The importance of the ideas underlying the principles was in part confirmed by the many observations, but many findings have also added to the understanding of the patient group and its needs. The findings contribute to refining the principles, and negative observations and input from participants have also highlighted the risks and drawbacks of design proposals. Regarding the first design principle, emphasizing the patients’ understanding and learning about their condition, we found overwhelming support for patients’ interest in learning more about their condition and improving their self-management skills. The educational materials that

XSL•FO
RenderX

were provided were appreciated and considered helpful for dealing with self-management problems, but many more topics were called for. Several usability issues also risked countering this design principle, including poorly designed interaction, navigation, presentation, and the use of language and terminology.

The second principle highlights the importance of supporting motivation and fostering sustained practices. Features of the web-based educational portal, such as reminders, calendars, and calculators, which could support, remind, encourage, maintain, and help participants plan their everyday practices, were considered motivating and helpful. However, participants also pointed out that knowing that the portal offers up-to-date educational materials was something that encouraged them to return to it. Poorly designed interaction, especially when registering data, was considered irritating and was identified as a risk when it comes to the sustained use of the portal.

The third principle focuses on patients' autonomy and confidence. Being able to carry out tasks on the portal without help from others was clearly pleasing for participants and contributed to a feeling of confidence. However, there is also a need to educate and guide the patients in how to use the system properly and efficiently. This could be done at the health care centers, where a number of one-to-one sessions with a nurse could be scheduled to direct the patients toward autonomy of use. The usability study shed light on an aspect that we had not previously focused on, namely, the confidence users have in the portal. Confidence in the system and how personal data are handled was a crucial issue for some users, and uncertainty regarding data security was something that risked undermining patients' use of the portal.

The fourth design principle highlights the importance of interaction and collaboration with others and was also confirmed in that participants considered it important and appreciated the possibility to chat with physicians and nurses. In addition, they asked for opportunities to contact other patients and to be able to not only chat with health care personnel but also to book a time to see their physicians. Text-based communication with health care professionals also raised some concerns for those who were unfamiliar with it—what would be an understandable level? In addition, the difference between synchronous and asynchronous communication with physicians or nurses was not clear to all participants.

In addition, as is common to each health information technology application, there are issues associated with security and privacy, which we introduced as a fifth principle because of the usability testing. Patient portals, as clinical information systems, are complex environments that integrate patients' data, information technologies, and human stakeholders. In designing patient portals optimally, it is imperative to protect patient confidentiality and follow health care provider guidelines [45]. Researchers proposed that the privacy principle should be classified into different levels according to importance [46]. To date, in Iran, there is no legal framework in use when it comes to information privacy, but there are legal provisions for cybercrime and protection of personal data [47]. The need to study the European General Data Protection Regulation is also felt in the Iranian legal system. Privacy and individuality are important ethical values and should be respected and protected in the design process [48]. In addition, privacy is gaining more attention in the design of communication technologies [49,50]. Finally, we would like to summarize the elements that support patient engagement. Helpful educational resources, supportive tools for planning and reminders, autonomy in task completion, and easy access to provider communication all encourage and sustain patient portal use for self-management.

Comparison With Prior Work

As designers, there is a need to balance the strengths and weaknesses of a technology [23]. Some of these issues are recognized from previous studies, such as users asking for materials in everyday language without medical jargon while simultaneously requesting accurate and complete information about medical conditions and treatments [28,51]. Such contradictory requirements demand that designers consider the trade-offs. Another such trade-off is, on the one hand, building a web-based educational portal with information about the patient that is as complete as possible, and, on the other hand, burdening patients with having to enter data continuously and also having to worry about the integrity of their personal data, also shown in a study by Derboven et al [24]. Furthermore, the study identified how patients' digital and health literacy and a perception of empowerment were crucial for them to be able to take responsibility for their self-management, which is in line with other studies [14,26-28]. Technology can help patients with chronic diseases take control of their conditions, but research needs to ensure that the technology's design is perceived as safe [23]. In addition, research has previously identified technology as a third agent when generating personalized advice for self-management, which is something that can be automated [24].

This study shows that although design research is time-consuming, usability testing and severity ratings provide invaluable information for design. Going through the scenario and tasks made it possible for participants to indicate which features of the web-based educational portal did not work as intended. It is natural for a usability test to discover problems and glitches—users comment on difficulties rather than when the design works smoothly. A well-known fact in usability design is that users comment more on faults than on features that work well, as the former are unexpected, disturb interactions, and are easier to detect [52]. However, participants also suggested how to improve the design of the portal to better fit their needs. In addition, many of these usability problems can, in most cases, be fixed with a simple redesign, as is reported elsewhere [53]. Inappropriate cultural assumptions and low-level usability issues may, in many cases, lead to nothing more than just annoyance, but seemingly trivial design flaws may ultimately also lead to patient harm [54]. Most of the usability problems that participants encountered in this study concerned entering the data in tasks 1 and 2, which were also found by Mazaheri Habibi et al [55]. Entering personal information and health data on this kind of portal may be a daunting task for patients, and they may also feel insecure about filling in the right data. Developing guidelines to help patients with registration is therefore a necessity [56]. It remains unclear,

XSL•FO
RenderX

especially considering the varying levels of digital and health literacy, whether such guidelines should be provided online, in printed form, or face-to-face. All the participants in this study successfully completed all the tasks, but most of the participants needed help along the way. Similar results were also found in studies of other patient groups [57]. Navigating a health portal is not a straightforward task, and efforts may be needed to produce clearer guidelines in order for portals to be used in practice. A section explaining any changes and improvements regarding the various functionalities can be helpful [15]. One particular difficulty in this case was that while the portal was in Farsi, participants were required to enter their laboratory results in English. Previous studies have shown that patients may find medical terminology on portals difficult [58-60], and they may need more guidance for such interactions. Research has also shown that participants with limited health literacy experience more barriers to using patient portals [56].

Our findings appear to be consistent with prior evaluations of diabetes self-management technologies in similar contexts. Nabovati et al [61] reported high usability for a diabetes app that combined educational content with calculators, reminders, graphical feedback, and physician communication. Our participants similarly endorsed those functions as motivating and practically helpful. Salari et al [62] further showed that personalization guided by behavioral theory and a clinician portal can promote independence and keep users involved. Complementing these results, Ghodousi Moghadam et al [63] demonstrated that a mobile serious game can enhance motivation for diabetes self-management. Together, these studies reinforce our conclusions that educational resources, planning or reminder tools, autonomous task support, and provider communication facilitate sustained portal use, while also reflecting the barriers we found, that is, the need to avoid complex language and to remediate confusing interactions through iterative, user-centered design.

One core aspect of usability is how effectively a design supports users in reaching their desired goal, in this case, self-management. This think-aloud study has investigated aspects that either support or risk being barriers to patients' self-management. Many usability problems and challenges regarding self-management were identified, including negative examples that seem to contradict the design principles. Such observations, however, enable a better understanding of the design principles. The findings enable the principles to be refined to better reach their goals, and the observations from the study, in combination with the input from the participants, have provided several possible ways of addressing these issues.

Limitations

Only literature incorporated in the authors' previous research informed the development of the framework of design principles. Consequently, the range of identified principles may have been influenced by this limited evidence base. Then, this study focused on a group of patients with T2D at a diabetic clinic in one specific city, Mashhad, which may not fully capture the diversity of patients in other settings or geographical locations. This limitation could affect the generalizability of the results. The study also concentrated on specific tasks within the portal's 9 sections, potentially overlooking usability issues or interactions between different sections. To address these gaps, future research should involve more clinical settings and a wider sample of patients. Nevertheless, the research was helpful in refining a framework of design principles pertaining to the design of web-based resources for the self-management of patients with T2D.

Conclusions

A portal designed to enhance patients' understanding, motivation, autonomy, and interaction with health care personnel is likely to be highly appreciated and to effectively support self-management. However, the design must also rigorously address risks that can undermine adoption and sustained use. Poorly designed interactions, such as confusing navigation, unclear workflows, or unintuitive interfaces, can hinder patients' ability to use the portal effectively. Likewise, complex or technical language may reduce comprehension, discourage use, and limit the portal's value for individuals with varying levels of health literacy. In addition, robust safeguards are required to protect patient integrity and data security, as trust in the system is essential for meaningful engagement.

We have proposed a framework of design principles pertaining to the design of web-based resources for the self-management of patients with T2D. Our results imply that usability testing, despite being time-consuming, is necessary when assessing a web-based educational portal, its potential for patients, and its sustained long-term use.

Acknowledgments

The authors received no specific funding for this work. The authors acknowledge the support of the personnel and patients of the Mashhad Diabetes Clinic and Medical Informatics Department at the Faculty of Medicine of Mashhad University of Medical Sciences. The authors would also like to thank Ali Jamali and Niloufar Fouladi, who designed the prototype for the web-based educational portal and participated in the severity analysis.

Funding

The authors declared no financial support was received for this work.

XSL•FO
RenderX

Authors' Contributions

JJ contributed to conceptualization, investigation, methodology, writing the original draft, reviewing and editing, and visualization. KK and IM contributed to conceptualization, supervision, methodology, writing the original draft, reviewing and editing, and visualization. HKM contributed to methodology, writing the original draft, and reviewing and editing. MRMH contributed to methodology, writing the original draft, and usability assessment moderator. SE contributed to methodology, writing the original draft, and reviewing and editing. SB contributed to supervision, methodology, writing the original draft, reviewing and editing, and visualization.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Reporting checklist for our study.
[PDF File, 82 KB - humanfactors_v13i1e78903_app1.pdf]

Multimedia Appendix 2
Instructions to participants.
[PDF File, 90 KB - humanfactors_v13i1e78903_app2.pdf]

Multimedia Appendix 3
Ethical approval.
[PDF File, 236 KB - humanfactors_v13i1e78903_app3.pdf]

References

1. Diabetes. World Health Organization. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/diabetes [accessed 2026-03-11]
2. Ogurtsova K, da Rocha Fernandes JD, Huang Y, et al. IDF Diabetes Atlas: global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes Res Clin Pract 2017 Jun;128:40-50. [doi: 10.1016/j.diabres.2017.03.024] [Medline: 28437734]
3. Hadjiconstantinou M, Byrne J, Bodicoat DH, et al. Do web-based interventions improve well-being in type 2 diabetes? A systematic review and meta-analysis. J Med Internet Res 2016 Oct 21;18(10):e270. [doi: 10.2196/jmir.5991] [Medline: 27769955]
4. Shah VN, Garg SK. Managing diabetes in the digital age. Clin Diabetes Endocrinol 2015 Dec;1(1):16. [doi: 10.1186/s40842-015-0016-2]
5. Stinson J, McGrath P, Hodnett E, et al. Usability testing of an online self-management program for adolescents with juvenile idiopathic arthritis. J Med Internet Res 2010 Jul 29;12(3):e30. [doi: 10.2196/jmir.1349] [Medline: 20675293]
6. Pereira K, Phillips B, Johnson C, Vorderstrasse A. Internet delivered diabetes self-management education: a review. Diabetes Technol Ther 2015 Jan;17(1):55-63. [doi: 10.1089/dia.2014.0155] [Medline: 25238257]
7. Woolley AK, Hadjiconstantinou M, Davies M, Khunti K, Seidu S. Online patient education interventions in type 2 diabetes or cardiovascular disease: a systematic review of systematic reviews. Prim Care Diabetes 2019 Feb;13(1):16-27. [doi: 10.1016/j.pcd.2018.07.011] [Medline: 30131299]
8. Huang Z, Semwal M, Lee SY, et al. Digital health professions education on diabetes management: systematic review by the digital health education collaboration. J Med Internet Res 2019 Feb 21;21(2):e12997. [doi: 10.2196/12997] [Medline: 30789348]
9. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. J Med Internet Res 2014 Apr 9;16(4):e104. [doi: 10.2196/jmir.2968] [Medline: 24718852]
10. Mashroofa MM, Haleem A, Jahufer A, Science Library, South Eastern University of Sri Lanka. Moderating effects of academic position and computer literacy skills on e-learning portal usage: SEM application on theory of planned behaviour. Rupkatha J Interdiscip Stud Humanit 2020;12(4):17. [doi: 10.21659/rupkatha.v12n4.18]
11. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9. [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
12. Riahi A, Mousavi CA. Survey of e-Health literacy among employees of state-owned banks in Tehran during 2020. J Health Literacy 2020;5(3):53-63. [doi: 10.22038/JHL.2020.52891.1133]
13. Iran economic indicators. theGlobalEconomy.com. 2025. URL: https://www.theglobaleconomy.com/Iran/ [accessed 2026-03-11]
14. Jafari J, Hossein KM, Ahmady S, Zary N. Readiness of diabetics to use the internet and mobile services: the case of a large city in a middle-income country. J Community Med Health Educ 2015;05(6). [doi: 10.4172/2161-0711.1000380]

XSL•FO
RenderX

15. Jones MC, MacGillivray S, Kroll T, Zohoor AR, Connaghan J. A thematic analysis of the conceptualisation of self-care, self-management and self-management support in the long-term conditions management literature. J Nurs Healthc Chronic Illn 2011 Sep;3(3):174-185. [doi: 10.1111/j.1752-9824.2011.01096.x]
16. Shah VN, Garg SK. Managing diabetes in the digital age. Clin Diabetes Endocrinol 2015;1:16. [doi: 10.1186/s40842-015-0016-2] [Medline: 28702234]
17. Peters D, Calvo RA, Ryan RM. Designing for motivation, engagement and wellbeing in digital experience. Front Psychol 2018;9:797. [doi: 10.3389/fpsyg.2018.00797] [Medline: 29892246]
18. Robinson SA, Zocchi MS, Netherton D, et al. Secure messaging, diabetes self-management, and the importance of patient autonomy: a mixed methods study. J Gen Intern Med 2020 Oct;35(10):2955-2962. [doi: 10.1007/s11606-020-05834-x] [Medline: 32440998]
19. Nunes F, Verdezoto N, Fitzpatrick G, Kyng M, Grönvall E, Storni C. Self-care technologies in HCI: trends, tensions, and opportunities. ACM Trans Comput Hum Interact 2015 Dec 14;22(6):1-33. [doi: 10.1145/2803173]
20. Ballegaard SA, Hansen TR, Kyng M. Healthcare in everyday life: designing healthcare services for daily life. Presented at: CHI '08: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; Apr 5-10, 2008; Florence, Italy p. 1807-1816. [doi: 10.1145/1357054.1357336]
21. Fischer G. Exploring design trade-offs for quality of life in human-centered design. interactions 2017 Dec 21;25(1):26-33. [doi: 10.1145/3170706]
22. Fischer G, Fogli D, Mørch A, Piccinno A, Valtolina S. Design trade-offs in cultures of participation: empowering end users to improve their quality of life. Behaviour & Information Technology 2020 Jan 2;39(1):1-4. [doi: 10.1080/0144929X.2020.1691346]
23. Fogli D, Piccinno A, Carmien S, Fischer G. Exploring design trade-offs for achieving social inclusion in multi-tiered design problems. Behaviour & Information Technology 2020 Jan 2;39(1):27-46. [doi: 10.1080/0144929X.2019.1634153]
24. Derboven J, Voorend R, Slegers K. Design trade-offs in self-management technology: the HeartMan case. Behaviour & Information Technology 2020 Jan 2;39(1):72-87. [doi: 10.1080/0144929X.2019.1634152]
25. Davis SE, Osborn CY, Kripalani S, Goggins KM, Jackson GP. Health literacy, education levels, and patient portal usage during hospitalizations. AMIA Annu Symp Proc 2015;2015:1871-1880. [Medline: 26958286]
26. Franssen B, Jafari J, Zary N, Masiello I. Designing patient education with distance learning. : Karolinska Institutet; 2015 URL: https://web.archive.org/web/20170812045410id_/http://ki.se/sites/default/files/bastiaan_franssen_designing.pdf [accessed 2026-03-11] [doi: 10.7287/peerj.preprints.1143v1]
27. Jafari J, Karimi Moonaghi H, Zary N, Masiello I. Exploring educational needs and design aspects of internet-enabled patient education for persons with diabetes: a qualitative interview study. BMJ Open 2016 Oct 31;6(10):e013282. [doi: 10.1136/bmjopen-2016-013282] [Medline: 27799245]
28. Jafari J, Karlgren K, Moonaghi HK, Layegh P, Bonacina S, Masiello I. Designing internet-enabled patient education for self-management of T2D diabetes-The case of the Razavi-Khorasan province in Iran. PLoS One 2021;16(4):e0250781. [doi: 10.1371/journal.pone.0250781] [Medline: 33905458]
29. ISO 9241-11:2018(en) - ergonomics of human-system interaction — part 11: usability: definitions and concepts. ISO. 2018. URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en [accessed 2026-03-11]
30. Aiyegbusi OL. Key methodological considerations for usability testing of electronic patient-reported outcome (ePRO) systems. Qual Life Res 2020 Feb;29(2):325-333. [doi: 10.1007/s11136-019-02329-z] [Medline: 31691202]
31. Jaja C, Pares-Avila J, Wolpin S, Berry D. Usability evaluation of the interactive Personal Patient Profile-Prostate decision support system with African American men. J Natl Med Assoc 2010 Apr;102(4):290-297. [doi: 10.1016/s0027-9684(15)30601-5] [Medline: 20437736]
32. Or C, Tao D. Usability study of a computer-based self-management system for older adults with chronic diseases. JMIR Res Protoc 2012 Nov 8;1(2):e13. [doi: 10.2196/resprot.2184] [Medline: 23612015]
33. Sallis JF, Owen N, Fisher EB. Ecological models of health behavior. In: Glanz BKR, Viswanath K, editors. Health Behavior and Health Education: Theory, Research and Practice: Jossey-Bass; 2008:465-485.
34. Martins AI, Santinha G, Almeida AM, et al. Consensus on the terms and procedures for planning and reporting a usability evaluation of health-related digital solutions: Delphi study and a resulting checklist. J Med Internet Res 2023 Jun 6;25:e44326. [doi: 10.2196/44326] [Medline: 37279047]
35. Virzi RA. Refining the test phase of usability evaluation: how many subjects is enough? Hum Factors 1992 Aug;34(4):457-468. [doi: 10.1177/001872089203400407]
36. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. Presented at: CHI '93: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems; Apr 24-29, 1993; Amsterdam, The Netherlands p. 206. [doi: 10.1145/169059.169166]
37. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003 Aug;35(3):379-383. [doi: 10.3758/bf03195514] [Medline: 14587545]
38. McDonald S, Edwards HM, Zhao T. Exploring think-alouds in usability testing: an international survey. IEEE Trans Profess Commun 2012 Mar;55(1):2-19. [doi: 10.1109/TPC.2011.2182569]

XSL•FO
RenderX

39. Jafari J. On designing educational portals for self-management of type 2 diabetes: lessons learnt from an iranian context. : Karolinska Institutet; 2021 URL: https://hdl.handle.net/10616/47752 [accessed 2026-03-11]
40. MacNeela P, Clinton G, Place C, et al. Psychosocial care in mental health nursing: a think aloud study. J Adv Nurs 2010 Jun;66(6):1297-1307. [doi: 10.1111/j.1365-2648.2009.05245.x] [Medline: 20384643]
41. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
42. Nielsen J. Severity ratings for usability problems. NN/Group. 1994. URL: https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/ [accessed 2026-03-11]
43. Siouta E, Hellström Muhli U, Hedberg B, Broström A, Fossum B, Karlgren K. Patients' experiences of communication and involvement in decision-making about atrial fibrillation treatment in consultations with nurses and physicians. Scand J Caring Sci 2016 Sep;30(3):535-546. [doi: 10.1111/scs.12276] [Medline: 26709843]
44. Paavola S, Lakkala M, Muukkonen H, Kosonen K, Karlgren K. The roles and uses of design principles for developing the trialogical approach on learning. Research in Learning Technology 2011 Nov;19(3):233-246. [doi: 10.3402/rlt.v19i3.17112]
45. Mathe JL, Duncavage S, Werner J, Malin BA, Ledeczi A, Sztipanovits J. Towards the security and privacy analysis of patient portals. SIGBED Rev 2007 Apr;4(2):5-9. [doi: 10.1145/1295464.1295467]
46. Lu X, Qu Z, Li Q, Hui P. Privacy information security classification for internet of things based on internet data. International Journal of Distributed Sensor Networks 2015 Aug 1;11(8):932941. [doi: 10.1155/2015/932941]
47. Latifzadeh M, Qabuli Dorafshan SMM, Mohseni S, et al. Analysis of the legal framework for the protection of personal data in the European Union. IJIPM 2021 Dec 1;37(2):439-472. [doi: 10.52547/jipm.37.2.439]
48. van Wel L, Royakkers L. Ethical issues in web data mining. Ethics Inf Technol 2004 Jun;6(2):129-140. [doi: 10.1023/B:ETIN.0000047476.05912.3d]
49. Nordgren A. Privacy by design in personal health monitoring. Health Care Anal 2015 Jun;23(2):148-164. [doi: 10.1007/s10728-013-0262-3] [Medline: 23978898]
50. Zubaydi HD, Varga P, Molnár S. Leveraging blockchain technology for ensuring security and privacy aspects in internet of things: a systematic literature review. Sensors (Basel) 2023 Jan 10;23(2):788. [doi: 10.3390/s23020788] [Medline: 36679582]
51. Kristiansen AM, Svanholm JR, Schjødt I, Mølgaard Jensen K, Silén C, Karlgren K. Patients with heart failure as co-designers of an educational website: implications for medical education. Int J Med Educ 2017;8:47-58. [doi: 10.5116/ijme.5898.309e]
52. Design MA. Design, user experience, and usability. design philosophy, methods, and tools. : Springer Presented at: Second International Conference, DUXU 2013, Held as Part of HCI International 2013; Jul 21-26, 2013 p. 684. [doi: 10.1007/978-3-642-39229-0]
53. Atashi A, Khajouei R, Azizi A, Dadashi A. Erratum: User Interface problems of a nationwide inpatient information system: a heuristic evaluation. Appl Clin Inform 2016 Jan;07(01):e1-e1. [doi: 10.1055/s-0040-1701485]
54. Gibbons MC, Lowry SZ, Patterson ES. Applying human factors principles to mitigate usability issues related to embedded assumptions in health information technology design. JMIR Hum Factors 2014 Dec 18;1(1):e3. [doi: 10.2196/humanfactors.3524] [Medline: 27025349]
55. Mazaheri Habibi MR, Khajouei R, Eslami S, Jangi M, Ghalibaf AK, Zangouei S. Usability testing of bed information management system: a think-aloud method. J Adv Pharm Technol Res 2018;9(4):153. [doi: 10.4103/japtr.JAPTR_320_18]
56. Tieu L, Sarkar U, Schillinger D, et al. Barriers and facilitators to online portal use among patients and caregivers in a safety net health care system: a qualitative study. J Med Internet Res 2015 Dec 3;17(12):e275. [doi: 10.2196/jmir.4847] [Medline: 26681155]
57. Wakefield B, Pham K, Scherubel M. Usability evaluation of a web-based symptom monitoring application for heart failure. West J Nurs Res 2015 Jul;37(7):922-934. [doi: 10.1177/0193945914568813]
58. de Clercq PA, Hasman A, Wolffenbuttel BH. Design of a consumer health record for supporting the patient-centered management of chronic diseases. Stud Health Technol Inform 2001;84(Pt 2):1445-1449. [Medline: 11604965]
59. Pyper C, Amery J, Watson M, Crook C. Patients' experiences when accessing their on-line electronic patient records in primary care. Br J Gen Pract 2004 Jan;54(498):38-43. [Medline: 14965405]
60. Tran DT, Zhang X, Stolyar A, Lober WB. Patient-centered design for a personal health record system. AMIA Annu Symp Proc 2005;2005:1140. [Medline: 16779426]
61. Nabovati E, Rangraz Jeddi F, Tabatabaeizadeh SM, Hamidi R, Sharif R. Design, development, and usability evaluation of a smartphone-based application for nutrition management in patients with type II diabetes. J Diabetes Metab Disord 2023 Jun;22(1):315-323. [doi: 10.1007/s40200-022-01140-x] [Medline: 37255839]
62. Salari R, R Niakan Kalhori S, GhaziSaeedi M, Jeddi M, Nazari M, Fatehi F. Mobile-based and cloud-based system for self-management of people with type 2 diabetes: development and usability evaluation. J Med Internet Res 2021;23(6):e18167. [doi: 10.2196/18167]
63. Ghodousi Moghadam S, Mazloum Khorasani Z, Sharifzadeh N, Tabesh H. A mobile serious game about diabetes self-management: design and evaluation. Heliyon 2024 Sep 30;10(18):e37755. [doi: 10.1016/j.heliyon.2024.e37755] [Medline: 39364243]

XSL•FO
RenderX

Abbreviations

QoL: quality of life
T2D: type 2 diabetes

Edited by P Santana-Mancilla; submitted 11.Jun.2025; peer-reviewed by B Scheckel, R Khajouei; revised version received 26.Jan.2026; accepted 19.Feb.2026; published 01.Apr.2026.

Please cite as:
Jafari J, Karlgren K, Karimi Moonaghi H, Mazaheri Habibi MR, Eslami S, Masiello I, Bonacina S
Aspects Supporting and Hindering Type 2 Diabetes Self-Management in Web-Based Educational Portals: Usability Testing Study With Updated Framework in Razavi-Khorasan, Iran
JMIR Hum Factors 2026;13:e78903
URL: https://humanfactors.jmir.org/2026/1/e78903
doi:10.2196/78903

© Javad Jafari, Klas Karlgren, Hossein Karimi Moonaghi, Mohammad Reza Mazaheri Habibi, Saeid Eslami, Italo Masiello, Stefano Bonacina. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 1.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Impact of Rural Trauma Team Development Education on Prehospital Time, Referral-to-Dispatch Interval, and Neurological and Musculoskeletal Injury Outcomes: Cluster Randomized Controlled Trial

Herman Lule[1,2,3], MMED, MBChB; Micheal Mugerwa[4], MSc; Anne Abio[5,6], PhD, MPH; Benson Oguttu[7,8], MBChB; Andrew Kakeeto[9,10], MBChB; Fiona J Walsh[11], MPH, MIA; Hervé Monka Lekuya[12,13], MMED, MD; Robinson Ssebuufu[14], MMED, MBChB; Patrick Kyamanywa[15], MMED, MBChB; Andreas Deckert[11], Dr Sc; Till Bärnighausen[16,17], MD, ScD; Jussi P Posti[18], MD, PhD; Michael Lowery Wilson[1], MPH, PhD

[1]Injury Epidemiology and Prevention (IEP) Research Group, Turku Brain Injury Centre, Division of Clinical Neurosciences, Turku University Hospital and University of Turku, Vähä Hämeenkatu 1B, Turun Yliopisto, Turku, Finland

[10]Department of Surgery, Kampala International University, Kampala, Uganda

[11]Heidelberg Institute of Global Health, Heidelberg University, Heidelberg, Baden-Wurttemberg, Germany

[12]Department of Neurosurgery, College of Health Sciences, Makerere University, Kampala, Uganda

[13]Department of Human Structure and Repair (GE38), Faculty of Medicine, Ghent University, Flanders, Belgium

[14]Department of Surgery, Mengo Hospital, Kampala, Uganda

[15]Mother Kevin Postgraduate Medical School, Uganda Martyr's University, Nkozi, Uganda

[16]Africa Health Research Institute, Durban, South Africa

[17]Department of Global Health and Population, Havard T H Chan School of Public Health, Havard University, Boston, MA, United States

[18]Neurocenter, Department of Neurosurgery and Turku Brain Injury Center, Turku University Hospital and University of Turku, Turku, Finland

[2]Center for Health Equity in Surgery and Anesthesia, Institute of Global Health Sciences, University of California, San Francisco, CA, United States

[3]Department of Surgery, Kiryandongo Regional Referral Hospital, Kigumba, Uganda

[4]Department of Public Health Sciences, California Baptist University, Riverside, CA, United States

[5]Research Centre for Child Psychiatry, University of Turku, Turku, Finland

[6]INVEST Research Flagship Centre, University of Turku, Turku, Finland

[7]Department of Surgery, Jinja Regional Referral Hospital, Jinja, Uganda

[8]Department of Public Health, Botswana-Havard Health Partnership (BHP), Gaborone, Botswana

[9]Department of Surgery, Hoima Regional Referral Hospital, Hoima, Uganda

Corresponding Author:
Herman Lule, MMED, MBChB
Injury Epidemiology and Prevention (IEP) Research Group, Turku Brain Injury Centre, Division of Clinical Neurosciences, Turku University Hospital and University of Turku, Vähä Hämeenkatu 1B, Turun Yliopisto, Turku, Finland

Abstract

Background: Scarce human resources for health and high injury-related mortality coincide with inequities in accessing quality trauma education programs in low- and middle-income countries. Existing observational studies restrict assessments of trauma training program impacts on providers' knowledge. Evaluation of trauma education programs outside clinical trial settings hinders their effectiveness in influencing clinical practice and policy changes for patient outcomes.

Objective: This study aimed to assess the impact of the Rural Trauma Team Development Course (RTTDC) on clinical processes and patient outcomes of motorcycle-accident–related neurological and/or musculoskeletal injuries in selected Ugandan hospitals.

Methods: This was a pragmatic 2-arm, parallel, multiperiod, cluster randomized controlled trial. The participants were trauma care frontline personnel and patients aged 2 - 80 years at 3 intervention and 3 control Ugandan hospitals (1:1 allocation). Hospitals were randomly allocated to intervention or control groups using permuted block sequences. Sequence codes were generated off-site by an independent statistician using Sealed Envelope (version 1.23.1; Sealed Envelope Ltd). Both patient participants and outcome assessors were blinded to allocation. Hospital allocation codes were concealed until the point of assignment. In the intervention arm, 500 trauma care frontliners received RTTDC, whereas patients received standard care. In the control arm, patients received standard care without RTTDC for staff. The primary outcomes were time from accident to admission and from referral to dispatch. The secondary outcomes were all-cause 90-day mortality and morbidity related to neurological and/or

XSL•FO
RenderX

musculoskeletal injuries. We followed the CONSORT (Consolidated Standards of Reporting Trials) guidelines for reporting cluster randomized trials.

Results: We analyzed 1003 participants (501 intervention and 502 control). The intervention arm had a shorter median (IQR) prehospital time of 1 hour (0.50 - 2) and referral-to-dispatch interval during interfacility transfers of 2 hours (1.25 - 2.75). This contrasted with 2 hours (1.50 - 4) and 4 hours (2.50 - 4.10) in the control arm, respectively *(P*<.001). The 90-day mortality was more than halved in the intervention (5%, 24/457) vs in the control arm (13%, 58/430) *(P*<.001). Fewer participants in the intervention group had unfavorable Glasgow Outcome Scale scores (9%, 42/457) vs (20%, 87/430) *(P*<.001). No difference was found in musculoskeletal injury morbidity outcomes *(P*=.57).

Conclusions: Rural trauma team development training demonstrated potential for improved organizational time efficiency and clinical outcomes for neurological injuries without negatively impacting musculoskeletal injury morbidity outcomes. Evidence from this trial supports that locally contextualized, trainee-led rural trauma team development interventional programs are feasible in low- and middle-income countries. However, despite being a multicenter study conducted across 6 geographically distinct hospitals, the research is limited in generalizability due to its focus on a single health care system within 1 country, retrospective trial registration, exclusion of prehospital deaths, and a relatively small number of clusters, which could introduce measurement bias.

Trial Registration: Pan African Clinical Trial Registry PACTR202308851460352; https://pactr.samrc.ac.za/TrialDisplay.aspx?TrialID=25763

International Registered Report Identifier (IRRID): RR2-10.2196/55297

(JMIR Hum Factors 2026;13:e82591) doi:10.2196/82591

KEYWORDS

trauma; education; training; prehospital; outcomes; neurology; musculoskeletal; injuries; cluster randomized trial

Introduction

Background

Globally, access to essential surgical services and trauma education remains restricted. In Africa, shortages of health care workers and their inequitable distribution further impede access to both emergency and elective surgical care. As a result, a substantial proportion of the population in Africa lacks access to these critical services [1].

Optimizing human resource capacity, particularly for trauma care, is critical for strengthening health care systems, alongside infrastructure development [2]. The provision of trauma education for nonspecialist primary care providers remains a potential remedy to address the shortage of health care human resources in Africa [3], but linking such medical education programs to patient outcomes has been underresearched. While several training programs such as the Advanced Trauma Life Support (ATLS) and Primary Trauma Course (PTC) have been introduced, financial barriers and limited assessment of the impact on patient outcomes have hindered the implementation of these education programs, particularly for frontline trauma care providers in low- and middle-income countries (LMICs), who include lay response personnel such as traffic police [3,4]. According to a scoping review of trauma training courses, the average cost of ATLS per participant was US $820 compared to PTC (US $232) and Rural Trauma Team Development Course (RTTDC; US 50 - US $100) [5].

Locally contextualized trauma education programs are needed in LMICs due to resource constraints limiting ATLS uptake [5]. Other existing courses, such as PTC, face criticism for limited international recognition, a relative bias toward trauma technical skills, which limit lay responders' participation, and their impact on patient outcomes has not been rigorously evaluated through clinical trials, warranting further exploration of need-specific alternatives [4].

Evidence from observational studies in the United States suggests that the RTTDC is a well-recognized form of trauma education. RTTDC can be flexibly implemented through a single-day format or a modular structure over approximately 8 hours, with instructors who are trauma specialists, nurses, or paramedics [6]. RTTDC emphasizes technical skills, including primary and secondary surveys, general trauma care, including for specific populations such as children and older adults, while also highlighting a team approach to injury management in rural environments. This is accomplished by evaluating team performance through simulations, didactic lectures, and case scenarios that focus on local injury burden and trauma nontechnical skills such as leadership, communication, coordination, resource management, and decision-making for interfacility transfers [6-8]. These components are vital for trauma teams in rural areas with limited human and infrastructural resources, enabling participation from nontechnical staff and connecting communities to formal trauma systems. Emphasizing trauma coordination, performance improvement, and situational resource awareness helps minimize delays in interfacility transfers for definitive care of injuries that exceed local capacity. While ATLS and PTC provide comprehensive approaches to technical trauma care, RTTDC offers a cost-effective option for training larger groups, including nontechnical staff, in resource-limited settings [5].

The ability of the RTTDC model to improve each trauma team member's performance makes it applicable in LMICs, where team and organizational efficiency are more desirable to mitigate understaffing. However, researchers have primarily evaluated the course's impact on provider performance rather than patient outcomes, as ethical constraints have precluded the evaluation

XSL•FO
RenderX

of RTTDC interventions through individually randomized prospective clinical trials [8]. Furthermore, the need to mitigate potential contamination between treatment arms necessitates a cluster-randomized study design. Evaluation of surgical interventions and trauma education programs outside of randomized controlled settings contributes to research implementation gaps, limiting their utility in clinical practice [9]. Cumulative research findings from emerging contemporaneous studies are necessary to drive practice and policy changes [9].

While prior systematic reviews have identified improvements in clinician knowledge from trauma education programs in LMICs, the next critical step is to evaluate their impact on organizational efficiency and patient outcomes in the clinical setting [3-5]. Therefore, in this cluster-randomized trial, we examined the impact of a locally contextualized RTTDC on hospital processes and clinical outcomes in an LMIC setting, focusing on time efficiency, mortality, and morbidity of motorcycle accident–related neurological and/or musculoskeletal injuries. These injuries were previously identified as critical for strengthening rural trauma systems in Uganda [10]. Motorcycle-related crashes are the leading cause of fatal trauma in Uganda, and the resultant injuries are predominantly neurological or musculoskeletal in 85% of cases [11]. An injury was deemed motorcycle-related if it involved a motorcycle-motorcycle, motorcycle-pedestrian, motorcycle-vehicle, or motorcycle-static object crash. Motorcycle-related injury mechanism was established during the secondary survey in the evaluation of events surrounding the injury.

This study defined rural trauma team development education as the implementation of RTTDC and the establishment of trauma teams in rural Uganda, which were previously nonexistent. In this study, RTTDC was locally contextualized to fit within the interconnected geopolitical, socioeconomic, and health care system challenges in Uganda, and to conform to the 4 domains of the Kirkpatrick framework for evaluating educational interventions [12]. First, we acknowledged the constraints of limited human resources for health in low-income Uganda, political tensions, and internal conflicts; thus, medical interns and traffic law enforcement professionals formed an integral component of our frontline trauma team participants. These cadres provide the first contact to patients with injury in Uganda due to a weak, inadequately funded prehospital and emergency postcrash care system, largely resulting from a shift of resources to the high infectious disease burden and obstetric emergencies [13].

Second, through engagement with local surgical and medical education experts (university medical faculty deans) and the Ugandan Ministry of Health, we fitted our trauma education intervention within the existing framework of medical curricula. We concurrently timed the recruitment of study participants during medical internship placements and surgical clinical rotations to maximize the course uptake.

Third, we assessed participants' satisfaction regarding the course through self-administered surveys (reaction), determined the effect of RTTDC intervention on participants' knowledge through pre-post training trauma-based multiple-choice questions (knowledge gain and retention), and assessed the change in practice using the pre-post training trauma nontechnical skills scale based on customized trauma case scenarios that suited the local injury burden (behavioral change). The impact of RTTDC on participants' knowledge, skills, and behavior was published elsewhere [14].

Finally, we translated classroom knowledge into clinical practice. We used the RTTDC participants to form well-coordinated, real-world rural trauma teams rooted at subcounties, using trained regional traffic police officers as the focal contact persons at the community level, and trained surgical residents at regional referral hospitals. Continuous reinforcement of RTTDC concepts to address team challenges was attained through weekly clinical audits with the trauma teams via Zoom (Zoom Video Communications, Inc). Consequently, these teams were involved in the development of a pilot Motorcycle Trauma Outcome Registry (MOTOR) using comprehensively pretested, refined, and locally customized quality indicators and measurable outcomes based on feedback from patient-public engagements. The data from the MOTOR registry were used in this study to directly link a trauma education intervention to patient outcomes through a trauma registry, thus completing the fourth level of the Kirkpatrick framework for rigorous evaluation of the effectiveness of RTTDC. The trial reporting followed the CONSORT (Consolidated Standards of Reporting Trials) guidelines for cluster-randomized trials, and the exploratory ancillary analyses due to COVID-19 interruptions adhered to the CONSORT and SPIRIT Extension for Randomized Controlled Trials Revised in Extenuating Circumstances (CONSERVE) guidelines [15,16].

Null Hypothesis

We based the specific objectives of this trial on the null hypothesis that RTTDC training has no effect on the clinical process efficiency and outcomes of neurological and musculoskeletal injuries at both the treatment arm and individual levels.

Specific Objectives

The study aimed to (1) determine the effect of RTTDC training on crash-to-admission and referral decision-to-dispatch intervals during interfacility transfers (primary outcomes); and (2) determine the effect of RTTDC training on all-cause 90-day mortality and morbidity of neurological and musculoskeletal injuries (secondary outcomes).

Methods

Overview

This was a pragmatic 2-arm, parallel, multiperiod, cluster randomized trial conducted from 9/01/2019 to 08/27/2023, with a 1-year suspension from 03/01/2020 to 03/01/2021 due to the extenuating effects of the COVID-19 pandemic. Six (6/17, 35%) Ugandan rural regional referral and teaching hospitals functioning as Level III trauma centers, including Jinja, Hoima, Fort Portal, Mubende, Kiryandongo, and Kampala International University Teaching Hospital, were included. Together, these

XSL•FO
RenderX

hospitals serve a total population of 18 million people in 3 cities and 50 rural districts. Following the national standard of care, all patients requiring referral were transferred to the Level I trauma center, Mulago National Referral Hospital, Uganda's comprehensive resource for multidisciplinary tertiary trauma care and rehabilitation.

Trauma centers are categorized into 5 levels based on their capabilities, with Levels IV and V being the lowest, providing initial stabilization through basic emergency facilities before transferring patients to higher-level centers. These correspond to health centers III and IV in Uganda. Levels III and II function as intermediate centers, offering 24/7 emergency assessment, resuscitation, and initiation of definitive care, including emergency surgeries, under the supervision of specialist surgeons, anesthesiologists, and radiologists, along with ongoing education programs for staff. In Uganda, these are typically represented by district hospitals and regional referral hospitals, respectively. While regional referral hospitals host surgical residency and internship programs to address human resource needs, they operate at Level III due to trauma resource challenges. Level I is a tertiary, superspecialized referral center that provides comprehensive care for severely injured patients, including complex microvascular surgeries, rehabilitation, and conducts trauma research, education, and quality assessment programs.

Randomization and Masking

Hospitals (clusters) were randomized in a 1:1 ratio to the intervention or control arm using permuted-block randomization, assuming equal cluster sizes. Blocked randomization lists with 6 random codes were generated off-site by an independent statistician using Sealed Envelope (version 1.23.1; Sealed Envelope Ltd) [17]. Hospital allocation codes were concealed until the point of assignment; local hospital administrators enrolled staff after allocation, and patient recruitment began thereafter.

In the intervention arm (comprising 3 hospitals), 500 trauma care frontline personnel received the fourth edition of the American College of Surgeons RTTDC training, which was administered in a multiperiod manner, whereas in the control arm (3 hospitals), trauma care frontline personnel did not receive the training [6,18]. The training intervention was delivered by a specialist surgeon and 2 senior surgical residents. Outcome assessors were external medical officers or senior residents who were not involved in training delivery and who remained masked to hospital allocation by using deidentified records; patients were not informed of hospital allocation. The training status of specific providers was not revealed to outcome assessors.

Eligibility Criteria

To be eligible to participate, a trauma center had to be a rural teaching hospital offering 24/7 emergency surgical care with access to blood banks and imaging, such as computed tomography (CT), X-ray, and ultrasound scans, as locally available or accessible in a nearby facility. Additionally, each hospital had to be staffed with surgical consultants, residents, interns, and medical students. Six out of 17 (35%) facilities met these inclusion criteria, accepted participation, and were randomized.

The participants in this trial were either trauma care frontline personnel or patients with trauma. The eligible trauma care frontline personnel constituted rural trauma team networks and were either surgery residents, intern doctors, third- to fifth-year medical or allied health students, or road traffic police officers concerned with the emergency evacuation of injured patients. It was not permissible for trauma care frontline personnel to offer their services in a way that would cross the treatment arms.

Eligible patient participants were those aged 2 - 80 years with motorcycle accident–related neurological and/or musculoskeletal injuries presenting within 24 hours at either the intervention or control hospitals during the study period. The rationale for the restriction of the study to these injuries was threefold. First, to minimize confounding due to diverse injury mechanisms during the impact evaluation of the trauma education intervention. Second, to objectively determine 90-day morbidity using standardized, injury-specific, validated tools such as the Glasgow Outcome Scale for traumatic brain injury (GOS-TBI) and Trauma Outcome Measure Score (TOMS) for musculoskeletal injuries [19,20]. Finally, to attain a sample large enough to reach statistical power in subgroup analyses for the specific injuries.

We excluded medical trainees with no previous exposure to surgical clinical rotation; pregnant women; mentally incapacitated individuals due to psychiatric disorders and without legally authorized representatives to provide informed consent; patients with documented stroke; passengers in a car at the time of collision to control confounding for injury mechanism; and deaths before imaging or arrival at emergency departments.

Prehospital deaths were excluded for 3 reasons. First, the health care system design directs such patients to city mortuaries rather than emergency departments, making accurate assessment of the primary outcome impossible, as our case definition of prehospital time was from the scene to the emergency department. Second, sociocultural norms in rural Uganda, where traditional medicine is prevalent, often discourage postmortems for deaths that are not linked to a public health concern, including postmortem CT imaging, which would hinder analysis of secondary outcomes on all-cause mortality, as caregivers typically view this process as lacking value for money. Finally, the absence of structured "on-scene" emergency care in Uganda at the time of the study meant that recording vital signs before hospital arrival could not be guaranteed, as most patients arrived at emergency departments by public transport; yet, vital signs were essential for assessing injury severity and controlling for potential confounding factors. More detailed eligibility criteria are documented in the trial protocol [21].

Patient and Public Involvement

We assessed the feasibility of trial outcome tools through engagement with trauma care frontline personnel and patient caregivers before study commencement. Further engagements were conducted by establishing rural trauma teams rooted at the subcounty level and obtaining bidirectional feedback on

XSL•FO
RenderX

trauma team performance through weekly audit meetings, as detailed in the study protocols [18,21].

Sample Size Determination and Power Analysis

The published trial protocol details the power calculations and methodology [21]. The total sample size of 1003 patient participants (501 intervention and 502 control) was estimated using an R shiny application (RStudio, Inc), accounting for the clustering design effect, 80% power, 5% type I error, intracluster correlation coefficient (ICC) of 0.02 (0.01 - 0.05), cluster autocorrelation coefficient of 0.8, equal cluster size allowing a coefficient of variation of 0.5, and 18% loss to follow-up [22]. The expected mean difference in primary outcomes was 1.02 hours based on an observational study in the United States [23]. Interim analyses were performed by a blinded statistician on August 28, 2022, in which results showed no patient-reported adverse incidents directly attributable to the intervention, which was the preset termination criteria based on harms; thus, the study continued to completion.

Data Collection

Data on the specified variables and outcome measures were collected from a piloted MOTOR created for this study. The data were entered into a secure Research Electronic Data Capture (REDCap; Vanderbilt University Medical Center), hosted by the University of Turku in Finland. Qualified medical doctors with 2 years of clinical experience or senior surgery residents served as blinded outcome assessors and were trained to use the MOTOR registry datasheet and *International Classification of Diseases, 10th Revision (ICD-10*) coding system. These assessors integrated data from clinical observations, patient interviews, and hospital and police records. Follow-up included accident and emergency departments, surgical wards, outpatient clinics, and home visits, as outlined in the study protocol [21].

Study Variables

We captured asymptomatic and symptomatic traumatic brain injury (TBI) based on the National Institute of Neurological Disorders and Stroke case definition. We obtained head CT images for participants who either met the New Orleans criteria or the Canadian CT Head Rule based on local applicable algorithms used at the study sites, whereas we used radiographs to evaluate limb fractures [24]. Based on our baseline study, we obtained variables earlier identified to influence injury outcomes in our settings, including sex, age, comorbidities, injury mechanism, road user category, helmet use, prehospital factors (transportation, first aid, and time), referral decision intervals, head CT results, definitive surgical treatment, multiplicity of injuries, and injury severity based on Glasgow Coma Scale (GCS) and Kampala Trauma Score II (KTS II) [25]. These variables align with the global surgery agenda to improve national surgical care plans in support of universal health coverage and Sustainable Development Goal 3.

Outcome Measures and Methods of Assessment

Detailed outcome measurements, analysis metrics, and assessment time points are illustrated in the published protocol [21]. The prespecified primary outcomes used measures of central tendency to compare between arms (1) prehospital time (from the accident scene to arrival at the emergency department), and (2) referral dispatch interval (from the time a referral decision was made to the time a patient exited the hospital gate to be transferred to a more resourced facility), measured in hours as indicators of process improvement.

The secondary outcomes compared between arms (1) proportions of all-cause 90-day mortality from the time of injury as a measure of physician-centered clinical outcomes, and (2) morbidity of neurological and musculoskeletal injuries at 90 days from the time of admission using the GOS-TBI (hereafter referred to as Glasgow Outcome Scale [GOS]) and TOMS, respectively, as measures of patient-reported functional outcomes [19,20]. Both GOS and TOMS were measured as final values as reported to outcome assessors, although their baseline equivalents, that is, GCS and Trauma Expectation Factor Score (TEFS), were captured to compare morbidity at baseline [19,26]. GOS was stratified into unfavorable (scores 1 - 3) vs. favorable outcomes (scores 4 - 5), whereas TOMS was dichotomized into unfavorable (TOMS < TEFS) vs favorable (TOMS ≥ TEFS) in accordance with previous studies [20,27].

These outcome measures were selected to capture the complex, multidimensional effects of trauma from diverse perspectives, including physical, social, and psychological domains. Both the GCS and the GOS have consistently demonstrated excellent internal validity, reproducibility, and reliability in multicenter trials; moreover, the tools have been validated for use in LMICs, with versions applicable in both adult and pediatric populations [20]. These tools can be administered physically, telephonically, or by post and are sensitive to severity and changes in injury sequelae over time in a diverse population of patients and examiners [28].

Acknowledging the importance of end-user involvement in research, the study used patient-reported outcomes (10-item Likert TOMS scale reported as a percentage) to assess the level of pain, physical function, disability, satisfaction with treatment, and overall life satisfaction at 90 days, compared to baseline expectations at admission (10-item Likert TEFS scale). Negative domains such as pain, disability, and activity reduction were assigned a minus score, while positive domains relating to satisfaction with life post-injury received a positive score, with total scores ranging from −700 to 300. Since injuries are not mutually exclusive in clinical settings, both GCS and GOS, or TEFS and TOMS, were administered for patients with both neurological and musculoskeletal trauma who were able to verbalize. These tools have demonstrated excellent psychometric properties in previous studies, as illustrated in Multimedia Appendix 1. KTS II captured the number of serious injuries to control for confounding due to polytrauma. The KTS II was used for its validity and suitability in predicting mortality in LMICs compared to other trauma scores, in addition to its ability to integrate physiological and anatomical indicators such as age-related comorbidities and multiplicity of injuries [29].

Clinically Meaningful Differences

For prehospital and referral-dispatch intervals, the minimal clinically meaningful difference was established as 1 hour, reflecting the “golden hour” guideline, which stipulates that

XSL•FO
RenderX

patients with trauma should reach definitive care within this timeframe to avoid increased mortality [30]. For the GOS-TBI, a 1-point difference was identified as the minimum clinically important difference, as it can shift an individual from an unfavorable to a favorable outcome category and correlates with a 1-point decline in the Disability Rating Scale [31]. This is deemed sufficient to demonstrate meaningful efficacy for TBI care and clinical research interventions, as outlined by the US Food and Drug Administration [31]. Conversely, for the TEFS and TOMS, assessed in the context of rehabilitation and improvements in functional performance from the patients' perspectives [32], the minimal clinically important difference was defined as the total TOMS score at 90 days meeting or exceeding the patients' TEFS assessed at admission, consistent with previous studies [33].

Statistical Analysis

Analyses were performed for all randomized clusters and participants in both intervention and control groups. We tested the normality of the distribution and equality of variance using the Shapiro-Wilk and Levene tests, respectively. For primary outcomes, a 2-sample Wilcoxon rank-sum test was used to compare the median (IQR) due to the skewed nature of the data despite log transformation. For secondary outcomes, we used an adjusted chi-square test to compare the difference in proportions of all-cause 90-day mortality and proportions of unfavorable GOS or TOMS for neurological and/or musculoskeletal injuries, respectively. The median (IQR) time to event (injury to death) was compared using a 2-sample Wilcoxon rank-sum test and visualized using Kaplan-Meier survival estimates. To explicitly address cluster dependency and to overcome SE bias due to the small number of clusters, we compared primary and secondary outcomes between treatment arms and reported cluster-level summaries and effect sizes within the framework of restricted (residual) maximum likelihood (REML) mixed-effects logistic regression models, using the Satterthwaite method of adjustment for the degrees of freedom. We used the Hussey and Hughes stratum-by-time extension in mixed-effects models to examine the variation in outcomes in both treatment arms across multiple time periods to address residual misspecification, assuming no study interruptions, and validated these results using cluster robust SEs (variance-covariance matrix of the estimators) in random-effects models. ICCs within periods were computed using the ".estat icc" command for each outcome using the mixed-effects REML regression models with Satterthwaite adjustment as the ratio of between-cluster variance of outcome to total variance of outcome between and within cluster.

Preplanned Subgroup Analyses

We performed preplanned subgroup analyses for sex, road user category, injury mechanisms, severity, and multiplicity of injuries. We used the REML mixed-effects logistic regression models, which permitted adjustment for confounding fixed and random effects to examine factors associated with delayed prehospital and referral-dispatch intervals, all-cause 90-day mortality, and morbidity of neurological and/or musculoskeletal injuries. We used the REML mixed-effects models to accommodate variations in baseline characteristics of random variables, intervention effects, and unbalanced data patterns that could have occurred between treatment groups, resulting from loss to follow-up. The random-effects variable was the treatment arm (intervention vs control) as the unit of analysis, with the odds ratios (ORs) and their corresponding 95% CIs as a direct estimate of the effect size. We examined all covariates for confounding and effect modification using the Cochran-Mantel-Haenszel statistics. Confounders were then added to the REML models individually or in blocks, and percentage change in exposure coefficients, Akaike Information Criterion, Bayesian Information Criterion, and log-likelihood for predictive quality were evaluated. Variables with substantial change in exposure effect were retained, and the models with the smallest Akaike Information Criterion and Bayesian Information Criterion were considered best-fitting. All analyses and data visualization were performed in Stata (version 15.0; StataCorp LLC). We considered $P<.05$ as statistically significant. Due to paper length constraints, ancillary studies detailing results of mixed-effects regression analyses for factors associated with mortality and morbidity of neurological and musculoskeletal injuries, barriers to injury care, and the impact of rural trauma team training on providers' knowledge, which constituted the tertiary outcomes of this trial, are reported separately [14,34,35] (preprint).

Additional Analysis

To account for potential confounding due to factors related to the COVID-19 pandemic's socioeconomic impact on emergency trauma services and livelihood, trial investigators conducted exploratory nonprespecified analyses comparing outcomes between participants recruited before and after the pandemic in accordance with the CONSERVE guidelines [16]. These analyses were performed to reflect a study halt, restricted participant access during lockdowns, and newly mandated participant safety requirements. Post hoc power analyses were conducted to assess whether the sample sizes used in the study were adequate for detecting effect sizes associated with the various study outcomes, given predetermined assumptions.

Handling Missing Data

We compared the baseline characteristics of participants who had missing endpoints due to loss to follow-up between the study arms. We conducted sensitivity analyses to assess the robustness of our findings under different assumptions about missing data. First, we performed a complete case analysis, assuming all missing data were completely missing at random. Second, we considered a worst-case scenario where all participants who lost to follow-up in the intervention group experienced the worst (mortality or unfavorable) trauma outcomes, while all those with missing data in the control group had favorable outcomes or survived. Third, we examined a best-case scenario, assuming all participants with missing data in the intervention group survived with favorable outcomes, while those in the control group had unfavorable outcomes or died. Finally, we performed multilevel multiple imputation to account for clustering and mechanisms of missingness and reported how the interpretation of results varied from the complete case analysis.

XSL•FO
RenderX

Ethical Considerations

The study adhered to ethical guidelines outlined by the Declaration of Helsinki and the 2017 Uganda National Council for Science and Technology guidelines. Ethical approval was obtained from the research and ethics committees of Mbarara University of Science and Technology (ref MUREC 1/7; 05/05/2019) and from the Uganda National Council for Science and Technology (ref SS 5082) before study commencement. Written informed consent was obtained from all participants or their legally authorized representatives before recruitment. The principal investigator facilitated data collection and trauma team coordination with airtime and transport reimbursement of US $7.50 per participant. This reimbursement was approved by the research and ethics committee to enable research-related follow-up of participants outside of their routine clinical appointments. This paper does not contain any personally identifying individual data or images. This study used data from the piloted MOTOR. At the time of protocol submission, the ethics committee did not mandate prior registration for academic, non–industry-funded proof-of-concept projects (policy effective December 31, 2023). Following the evolution of local policy, the investigators pursued registration with the Pan African Clinical Trial Registry (PACTR202308851460352). Due to COVID-19 disruptions, registration was completed on 08/17/2023, after recruitment commenced. The full protocol for this study was published under the International Registered Report Identifier: DERR1-10.2196/55297. The authors confirm that all related trials were included in this registration [21]. The specific impacts of the pandemic on this trial, including suspension, were reported according to CONSERVE guidelines, and subgroup analyses were performed appropriately to control for this confounding [16]. Where applicable, retrospective registration and protocol deviations due to the pandemic are cited in the abstract and paper text; for completeness, a detailed list of all deviations is provided in Multimedia Appendix 2.

Results

Overview

The cluster-level characteristics and resources are provided in Table 1.

XSL•FO
RenderX

Table . Showing cluster-level characteristics and resources (data source: field data).

Treatment arm	Intervention group			Control group		
Cluster ID[a] and site-level characteristics	H1[b]	H2[c]	H3[d]	H4[e]	H5[f]	H6[g]
Distance from national referral hospital (km)	209	84	198	150	294	300
Number of functional CT[h] scanners (n)	[i]0	1	1	1	1	1
Number of functional X-ray machines (n)	1	1	1	1	1	1
Number of functional ultrasound scanners (n)	1	1	1	1	1	1
Number of functional laboratories (n)	1	1	1	1	1	1
Number of functional ambulances (n)	1	2	2	1	2	2
Standard bed capacity (n)	109	500	267	240	300	700
Operating bed capacity (n)	175	600	317	500	333	600
ICU or HDU bed capacity (n)	2	10	20	8	10	12
Number of blood banks (n)	1	1	1	1	1	1
Number of oxygen plants (n)	[j]0	1	1	1	1	1
Staffing capacity (% occupancy of sanctioned vacancies), n (%)	169/349 (48)	389/1269 (31)	360/1195 (30)	348/1200 (29)	338/1200 (28)	380/800 (47)
Catchment population (millions)	0.5	4.5	3.5	2.0	3.5	4.0
Geographical coverage, number of cities (number of districts)	0 (6)	1 (11)	1 (9)	0 (5)	1 (9)	0 (10)
Trained permanent staff involved in the care of patients with trauma, n (%)	16/20 (80)	56/70 (80)	32/40 (80)	–0 (0)	–0 (0)	–0 (0)
[k]Trained staff including surgery residents, interns, and medical trainees, n (%)	344/500 (68.8)	87/500 (17.4)	69/500 (13.8)	–0 (0)	–0 (0)	–0 (0)
Patient cluster size for the 12 periods, n (%)	189 (18.8)	140 (14.0)	172 (17.1)	166 (16.6)	183 (18.2)	153 (15.3)
Average cluster size per period (n)	16	12	14	14	15	13

XSL•FO
RenderX

[a]H: cluster (hospital) identification code (ID).

[b]H1: Kiryandongo Regional Referral Hospital (public).

[c]H2: Jinja Regional Referral Hospital (public).

[d]H3: Hoima Regional Referral Hospital (public).

[e]H4: Mubende Regional Referral Hospital (public).

[f]H5: Fort Portal Regional Referral Hospital (public).

[g]H6: Kampala International University Teaching Hospital (private).

[h]CT: computed tomography.

[i]Outsources CT images from neighboring missionary hospital.

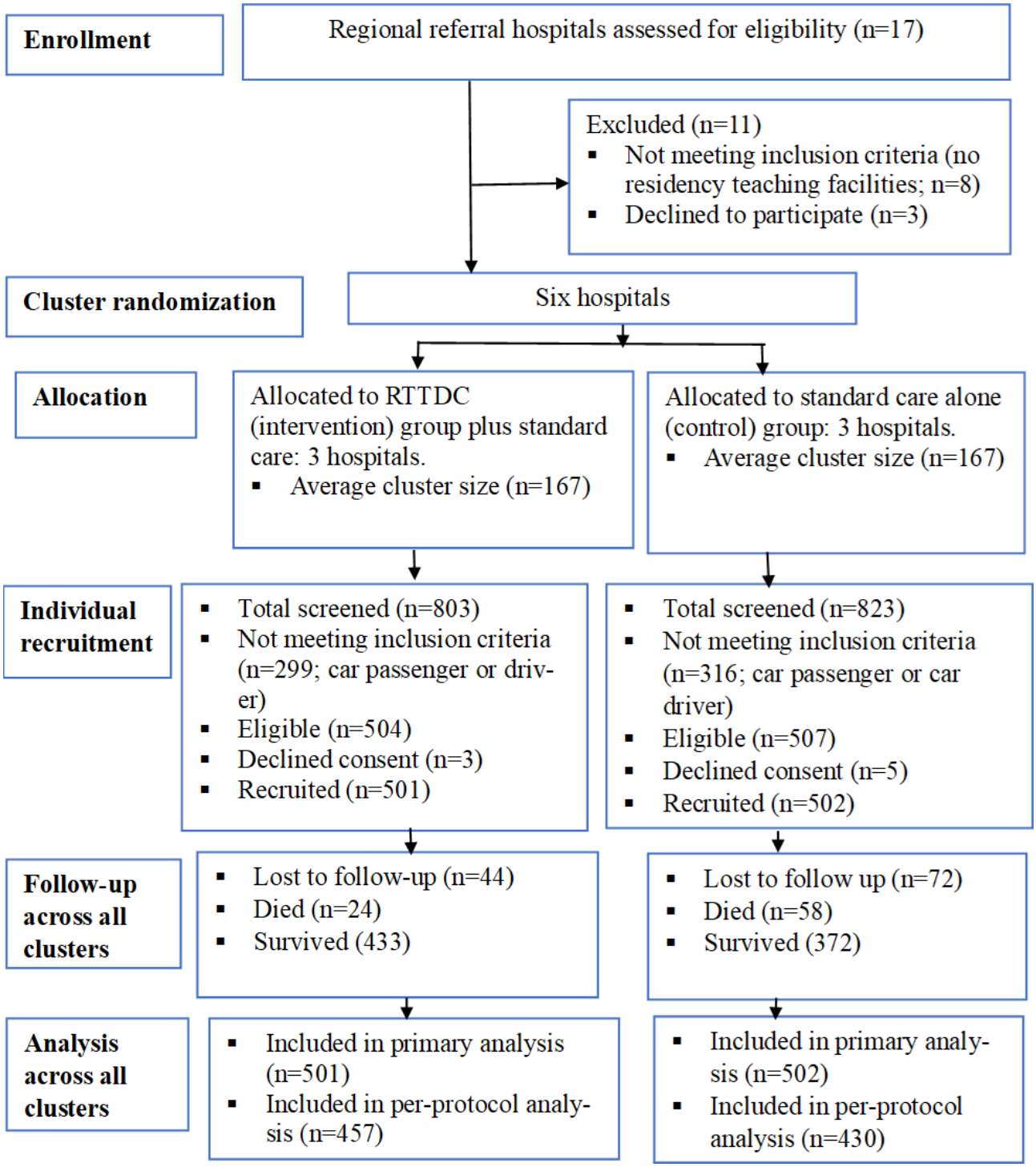
[j]Relies on stored oxygen cylinders and oxygen concentrators.

[k]Proportions commensurate to the average annual number of interns, surgery residents, and medical trainees received at the respective hospitals.

At the end of the study period, there were 1003 participants (501 intervention and 502 control), with an average cluster size of 167 in both groups (14 individuals per cluster per period). The participant CONSORT flow diagram is summarized in Figure 1.

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram. RTTDC: Rural Trauma Team Development Course.



Baseline Sociodemographic and Clinical Characteristics

Cluster- and individual-level baseline sociodemographic and clinical characteristics are summarized in Multimedia Appendix 3. Of 1003 participants, 82% (817) were male and 19% (186) were female (male:female ratio 402:99 in the intervention group vs 415:87 in the control group). The overall median age (IQR) was 28 (22-37) years (median age 28 years, IQR 22-38 years in the intervention group vs median age 28 years, IQR 22-36 years in the control group). There were no significant imbalances in the distribution of baseline characteristics between treatment groups except for 5 out of 37 covariates, namely employment status, proportion sustaining motorcycle-car crash, reported head impact, multiplicity of injuries, and respiratory rate. The overall median injury severity based on KTS II was 8 (IQR 7-9) and did not differ significantly between groups (P=·36).

Most participants (77%, 776/1003) arrived by public means (79%, 398/501 in the intervention group vs 75%, 378/502 in the control group). Helmet use was reported at 25% (254/1003; 23%, 144/501 in the intervention vs 28%, 140/502 in the control group). Irrespective of symptoms, 95% (949/1003) of participants reported head trauma (impact). Symptomatic TBI was present in 70% (699/1003; 69%, 347/501 in the intervention

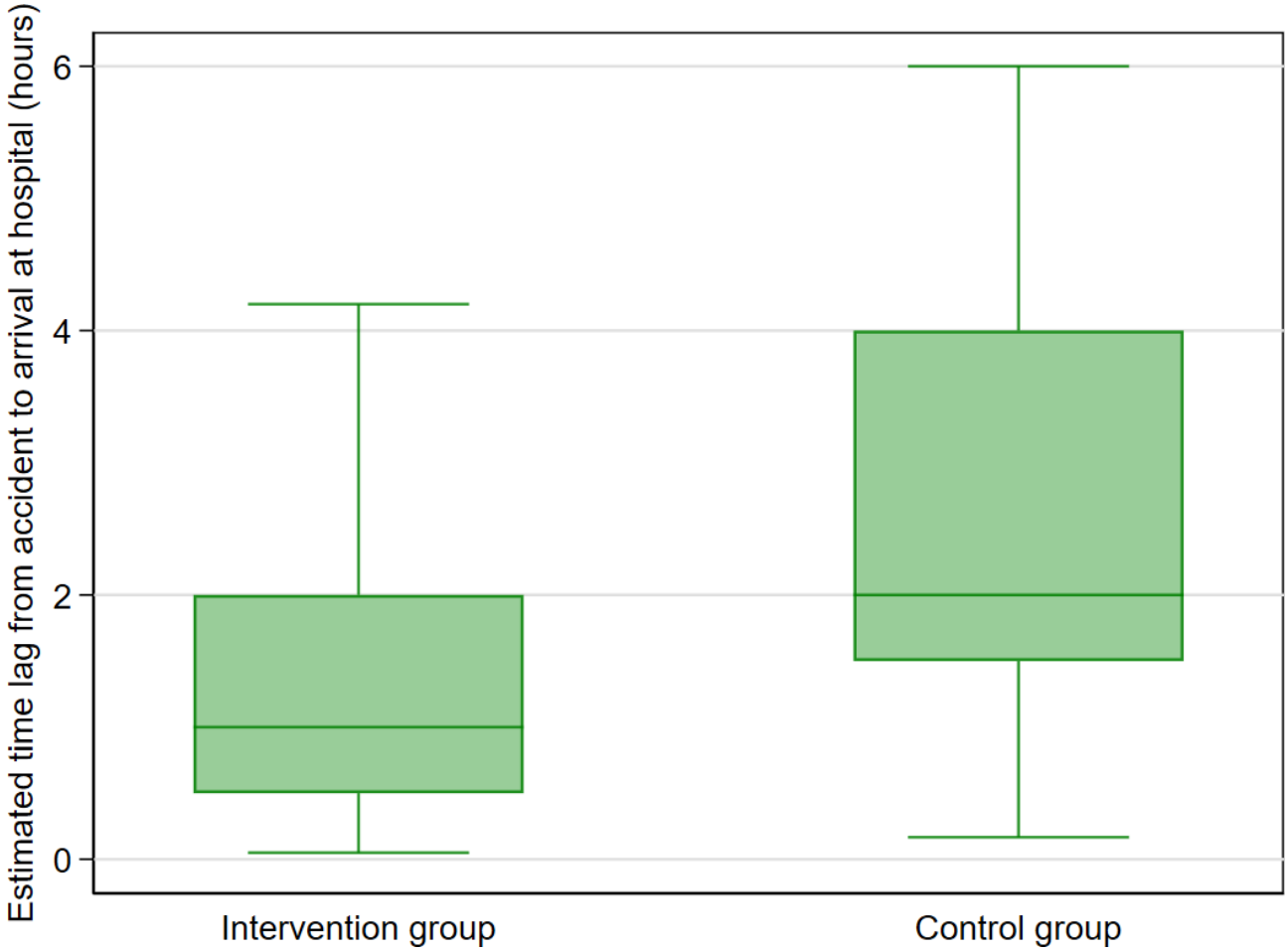
XSL•FO
RenderX

vs 70%, 352/502 in the control group). Over 67% (675/1003) of participants met the head CT criteria (67%, 335/501 in the intervention vs 68%, 340/502 in the control group). Conversely, one-third of the participants did not meet the CT criteria. The median GCS was 14 (IQR 11-15) and did not differ between groups (*P*=.12). The admission systolic blood pressure was ≥90 mm Hg in 82% (822/1003), with oxygen saturation >90% in 80% (802/1003) of participants. Loss of consciousness and posttraumatic headache were the most reported clinical presentations. Acute epidural hematomas were the most observed radiological pathology (17%, 175/1003), followed by subdural hematomas (10%, 95/1003). Emergency craniotomy was the most performed neurosurgical intervention (24%, 236/1003). On the other hand, musculoskeletal injuries were present in 80% (799/1003) (80%, 400/501 in intervention vs 80%, 399/502 in control), of which 27% (268/1003) had limb fractures (27%, 134/501 in intervention vs 27%, 134/502 in control).

Primary Outcome I: Comparison of Prehospital Time

For the 1003 participants, the overall median prehospital time was 2 (IQR 1.00 - 3.00) hours and was shorter in the intervention compared to the control group (1 hour, IQR 0.50 - 2.00 hours vs 2 hours, IQR 1.50 - 4.00 hours; *P*<.001) (Figure 2). Adjusted predictions of prehospital time between arms across study periods are provided in Multimedia Appendix 4. Factors that contributed to delayed prehospital arrival beyond the optimal 1 hour were commute distances longer than 5 km, being a survivor of motorcycle-static object collisions, needing first aid before arrival at emergency departments, and requiring craniotomy, whereas patients' age and head CT findings were the most significant determinants of delayed referral-dispatch execution (Multimedia Appendix 5).

Figure 2. Comparison of median prehospital time lag from accident scenes to arrival at emergency departments.

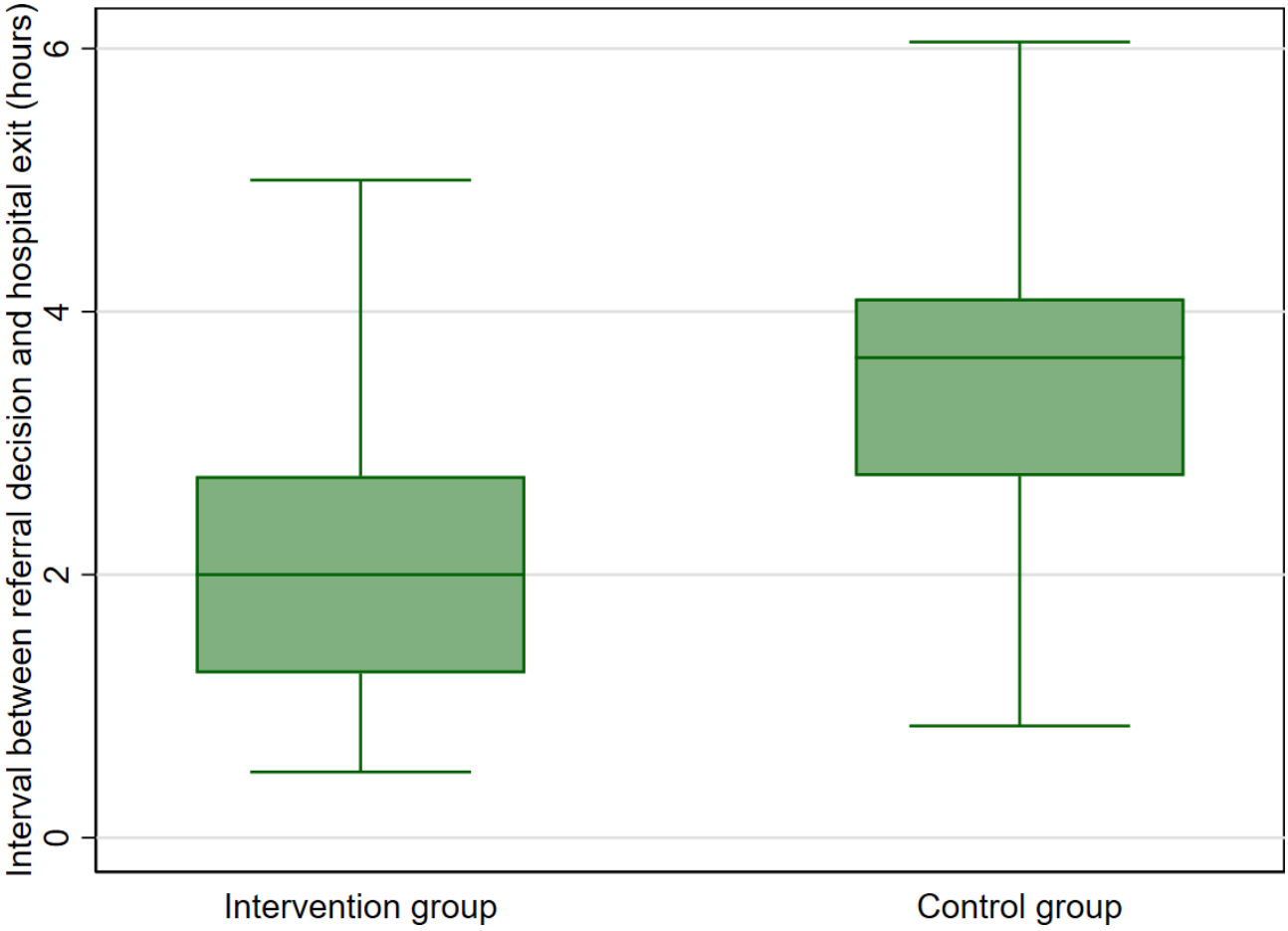


Primary Outcome II: Comparison of Referral Decisions to Dispatch Interval During Interfacility Transfers

For 69% (691/1003) of patients whose care demands exceeded local resources and warranted referral, the overall median referral decision to dispatch interval was 3 hours (IQR 1.75 - 3.85 hours). As provided in Figure 3, the referral decision to dispatch interval was shorter in the intervention (2 hours, IQR 1.25 - 2.75 hours) vs the control group (4 hours, IQR 2.50 - 4.10 hours; *P*<.001]. Adjusted stratum-by-time predictions of referral-dispatch intervals between arms across study periods are provided in Multimedia Appendix 6.

XSL•FO
RenderX

Figure 3. Comparison of referral decision to dispatch interval between groups.

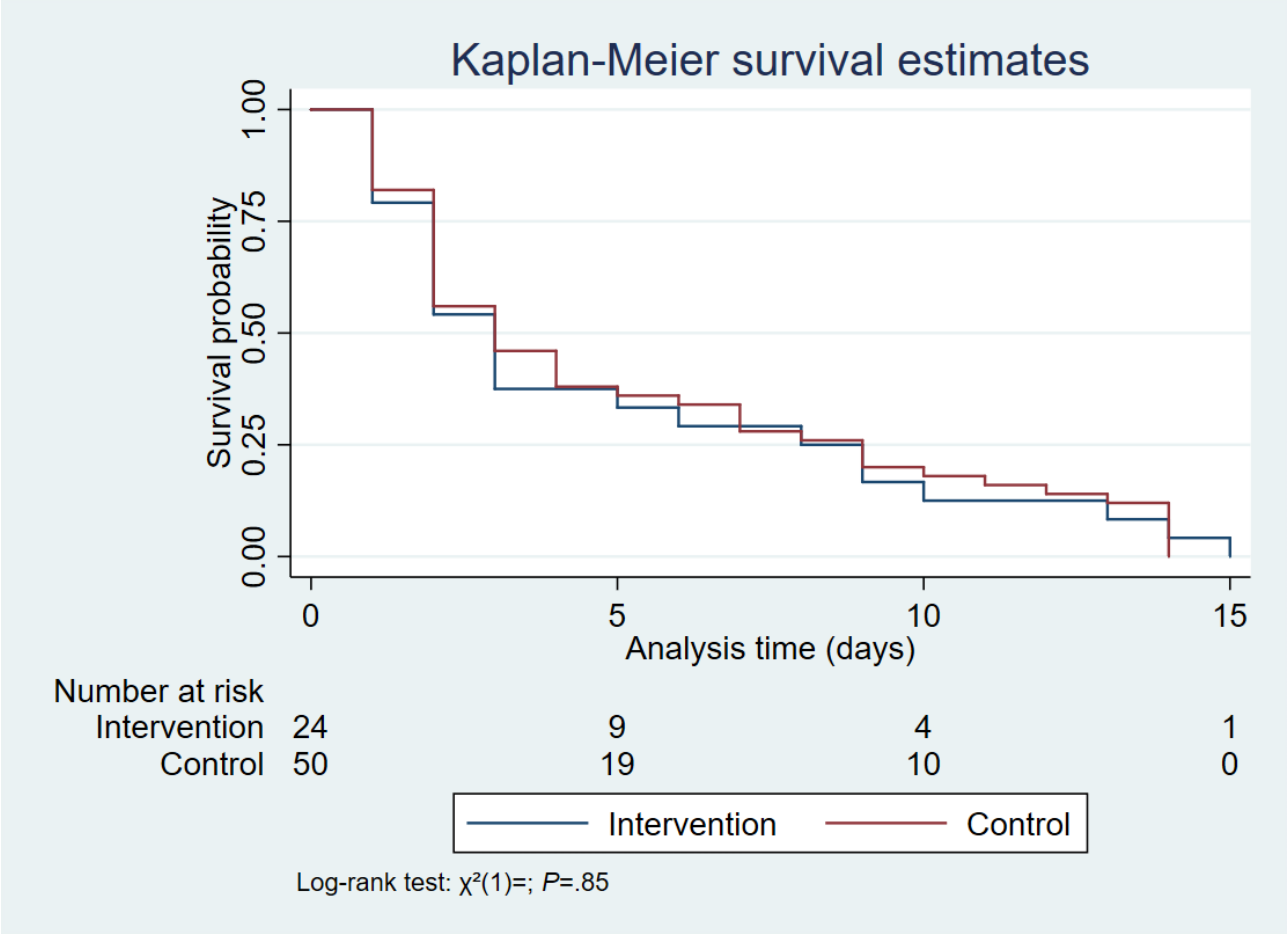


Secondary Outcome I: All-Cause 90-Day Mortality

Of the 88% (887/1003) participants who completed the 90-day follow-up, the overall 90-day mortality was 9% (82/887). Mortality was lower in the intervention at 5% (24/457) compared to the control group at 13% (58/430) (χ^2_1=17·9; *P*<.001). Of the 82 deaths, 8 (9.8%) occurred in the emergency departments of the control group vs 0 (0%) in the intervention group. Figure 4 demonstrates that the overall median survival time among the remaining 74 deaths was 3 (IQR 1-8) days and did not differ significantly between groups (*P*=.48). Stratified Cox regression analysis showed that the only determinant of time to event (death) was receiving a craniotomy (*P*=.01), but the proportions that required this procedure did not differ between groups (*P*=.17) (Multimedia Appendix 7). Although more males (84%, 69/82) died than females (16%, 13/82), preplanned subgroup analyses showed that the median survival time did not differ significantly across study periods by sex (*P*=.71) (Multimedia Appendix 8) and by road user category (*P*=.46) (Multimedia Appendix 9).

XSL•FO
RenderX

Figure 4. Comparison of Kaplan-Meier survival time (days) between groups.



Secondary Outcome II: Comparison of Morbidity of Neurological Injuries

Of the 887 out of 1003 participants who completed the 90-day follow-up, 86% (758/887) had a favorable GOS, with a median of 5 (IQR 4-5). The proportion of participants with favorable outcomes was significantly higher in the intervention group (91%, 415/457) compared to the control group (80%, 343/430) (χ^2_1=21·7; *P*<.001). The median GOS (IQR) was 5 (4-5) in both arms, with the intervention group showing a slightly higher median compared to the control group (*P*=.003). The results of all GOS subcategories among participants are summarized in Multimedia Appendix 10.

Secondary Outcome III: Comparison of Morbidity of Musculoskeletal Injuries

Of the 637 out of 799 patients with musculoskeletal injuries whose TEFS and TOMS were complete at 90 days, the majority (75%, 478/637) had a favorable outcome (TOMS ≥ TEFS). As provided in Figure 5, the proportion of patients whose trauma outcome equaled or exceeded their initial expectations was comparable between the intervention (76%, 262/345) vs the control group (74%, 216/292) (χ^2_1=0.6; *P*=.57).

XSL•FO
RenderX

Figure 5. Comparison of Trauma Expectation Factor Scores (TEFSs) and Trauma Outcome Measure Scores (TOMSs) between groups.

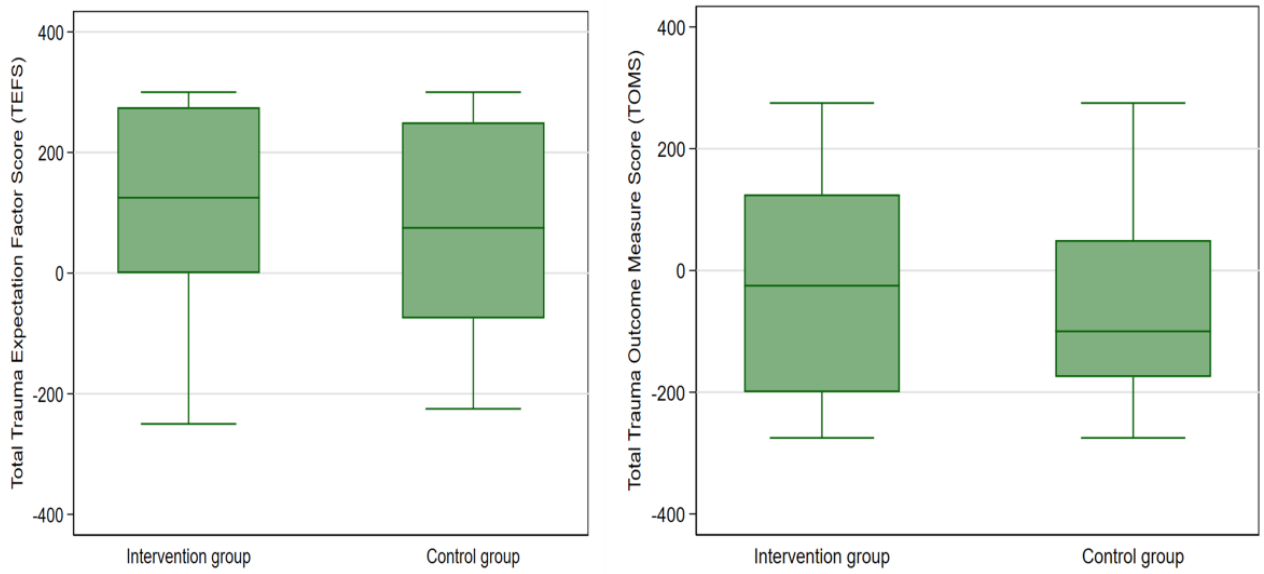


Table . Cluster-level summary and comparison of primary and secondary outcomes between arms (per protocol).

Arm[a] and cluster ID[b]	Cluster size, (n)	Primary outcome 1: prehospital interval in hours, median (IQR)	Primary outcome 2: referral-dispatch interval in hours, median (IQR)	Secondary outcome 1: 90-day mortality, n (%)	Secondary outcome 2: Proportion of unfavorable GOS[c], n (%)	Secondary outcome 3: proportion of unfavorable TOMS[d], n (%)
Both arms	1003	2.0 (1.0 - 3.0)	2.8 (1.8 - 3.9)	82 (9.2)	129 (14.5)	163 (25.6)
Intervention arm	501	1.0 (0.5 - 2.0)	2.0 (1.3 - 2.2)	24 (5.2)	42 (9.1)	86 (24.8)
H1[e]	189	1.0 (0.8 - 2.0)	2.0 (1.5 - 3.0)	15 (8.3)	21 (11.6)	28 (20)
H2[f]	140	1.0 (0.5 - 2.0)	1.5 (0.9 - 2.1)	6 (5.0)	15 (12.6)	27 (27.3)
H3[g]	172	1.2 (0.5 - 2.1)	2.0 (1.2 - 2.8)	3 (1.9)	6 (3.7)	31 (28.7)
Control arm	502	2.0 (1.5 - 4.0)	3.7 (2.8 - 4.1)	58 (13.6)	87 (20.5)	77 (26.6)
H4[h]	166	2.0 (1.0 - 3.0)	3.8 (2.8 - 4.2)	28 (20.0)	41 (29.3)	21 (25.3)
H5[i]	183	2.0 (1.5 - 4.0)	3.6 (2.8 - 4.1)	21 (13.2)	28 (17.6)	22 (19.3)
H6[j]	153	2.0 (1.5 - 4.0)	3.5 (2.5 - 4.0)	9 (7.1)	18 (14.3)	34 (36.6)
Difference (intervention vs. control; 2-sample Wilcoxon rank-sum test *P* value)	—	<.001	<.001	<.001	<.001	.57

[a]Arm: intervention vs control.

[b]Cluster ID: unique identifier for each cluster (hospital).

[c]GOS: Glasgow Outcome Scale.

[d]TOMS: Trauma Outcome Measure Score.

[e]H1: Kiryandongo Regional Referral Hospital.

[f]H2: Jinja Regional Referral Hospital.

[g]H3: Hoima Regional Referral Hospital.

[h]H4: Mubende Regional Referral Hospital.

[i]H5: Fort Portal Regional Referral Hospital.

[j]H6: Kampala International University Teaching Hospital.

The OR of unfavorable to favorable patient-centered trauma outcomes for musculoskeletal injuries did not differ significantly between intervention (OR 0.952, 95% CI 0.804 - 1.128; *P*=.57) vs control (OR 1.073, 95% CI 0.907 - 1.269; *P*=.57). This finding remained consistent after preplanned subgroup analyses for limb fractures in the intervention (OR 1.009, 95% CI

XSL•FO
RenderX

0.716 - 1.368; *P*=.95) vs in the control group (OR 0.990, 95% CI 0.716 - 1.368; *P*=.95) and for isolated tibial fractures in the intervention (OR 0.992, 95% CI 0.740 - 1.332; *P*=.96) vs in the control group (OR 1.008, 95% CI 0.726 - 1.400; *P*=.96) (Multimedia Appendix 11). Adjusted predictions of TEFS and TOMS across study periods are illustrated in Multimedia Appendix 12. The cluster-level summary and comparison of all outcomes using nonparametric Wilcoxon rank-sum tests are provided in Table 2, with the number of clusters in the intervention arm=3; number of clusters in the control arm=3; total number of clusters: 6; number of discrete decay periods=12.

The results of nonparametric tests were congruent with those of mixed-effects regression (REML) comparisons after adjustment for small-sample correction using the Satterthwaite method (Table 3). The corresponding residual estimates are provided in Multimedia Appendix 13.

Table . Per-protocol comparison of primary and secondary outcomes across study periods after adjustment for small-sample correction.

Outcome variable (number of observations)	Average cluster size per period, n (min-max)	Average sample per period, n (min-max)	SE	Satterthwaite adjusted df	Coefficient (intervention arm; 95% CI)	*P* value	Residual ICC[a] at level of study period (95% CI)
Primary outcome 1: prehospital interval in hours (n=1003)	14 (13 - 15)	83 (80 - 88)	0.08	4	–1.13 (–1.29 to –0.97)	<.001	0.02 (0.00-0.07)
Primary outcome 2: referral-dispatch interval in hours (n=691)	10 (7 - 11)	58 (42 - 68)	0.08	4	–1.39 (–1.56 to –1.22)	<.001	0.01 (0.00-0.08)
Secondary outcome 1: percentage of 90-day mortality (n=1003)	14 (13 - 15)	83 (80 - 88)	0.06	4	–0.24 (–0.36 to –0.12)	<.001	0.01 (0.00-0.05)
Secondary outcome 2: percentage of unfavorable Glasgow Outcome Scale (GOS; n=887)	12 (11 - 14)	73 (66 - 84)	0.02	4	–1.11 (–0.16 to –0.07)	<.001	0.01 (0.00-0.07)
Secondary outcome 3: percentage of unfavorable Trauma Outcome Measure Score (TOMS; n=637)	9 (7-11)	53 (42 - 64)	0.04	4	–0.023 (–0.09 to 0.05)	.51	0.02 (0.01-0.10)

[a]ICC: intracluster correlation coefficient.

Adverse Events

There were no crossovers between arms, but there were 8 deaths at the emergency department in the control arm, which can be considered as an adverse injury outcome. At 90-day disposition among 805 retained survivors, a total of 16 (2%) participants were still hospitalized due to pressure ulcer sequelae to neurological injuries. These ulcers required long-term debridement and plastic reconstruction alongside physiotherapy and rehabilitation, which all took longer than 90 days. However, the proportions of those retained (7/433 in the intervention vs 9/372 in the control group; χ^2_1=0.6; *P*=.43) and those discharged home (426/433 in the intervention vs 363/372 in the control group, χ^2_1=0.01; *P*=.94) did not differ significantly.

To examine the extenuating circumstances of COVID-19, we performed exploratory analyses to probe any differences in outcomes, and our results showed no differences in the prehospital interval, referral decision to dispatch interval, and all-cause mortality before and after the COVID-19 pandemic. However, participants who were seen after the onset of the pandemic were twice as likely to report TOMS below their baseline expectations (OR 2·019, 95% CI 1.178 - 3.460; *P*=.01), and this association remained significant even after adjusting for confounding due to age, employment status, marital status, injury mechanisms, injury severity, and multiplicity of injuries (adjusted OR 2.095, 95% CI 1.199 - 3.659; *P*=.01). A detailed account of the pandemic effects and other factors that influenced mortality and morbidity is provided here [34].

Missing Data

Our data quality and handling protocols in REDCap have been previously described in the study protocol [21]. On evaluation of lost to follow-up, the primary outcomes with missing end points (prehospital time or referral dispatch interval) were less than 0.1% (1/1003), whereas 12% (116/1003) were lost to follow-up and had missing secondary outcomes (90-day mortality, GOS, and TOMS). The proportion of loss to follow-up

XSL•FO
RenderX

was lower in the intervention (9%, 44/116; *P*=.001) vs in the control group (14%, 72/116*; P*=.001), but the baseline sociodemographic and clinical characteristics of those lost to follow-up did not differ significantly between the 2 treatment groups (all *P*>.05; Table 4).

XSL•FO
RenderX

Table . Comparison of baseline characteristics among those lost to follow-up. Level of statistical significance: $P<.05$.

Characteristic and category	Intervention group (n=44)	Control group (n=72)	Overall (n=116)	*P* value
Age (years), median (IQR)	28.0 (23.0 - 40.5)	30.0 (23 - 36.5)	28.0 (23.0 - 39.0)	.99
Sex, n (%)				.18
Male	37 (84.1)	54 (75.0)	91 (78.4)	
Female	7 (15.9)	18 (25.0)	25 (21.5)	
Marital status, n (%)				.42
Single	17 (38.6)	28 (38.9)	45 (38.8)	
Married	24 (54.5)	41 (56.9)	65 (56.0)	
Divorced	1 (2.3)	3 (4.2)	4 (3.4)	
Other	2 (4.5)	0 (0.0)	2 (1.7)	
Employment, n (%)				.37
Formal paid employment	1 (2.3)	5 (6.9)	6 (5.2)	
Self-employed	68.2 (41.7)	51 (70.8)	81 (69.8)	
Unemployed	2 (4.5)	4 (5.6)	6 (5.2)	
Student	9 (20.5)	12 (16.7)	21 (18.1)	
Other	2 (4.5)	0 (0.0)	2 (1.7)	
Commute distance, median (IQR)	4 (2.0 - 17.0)	5 (3.0 - 12.5)	5 (2.7 - 14.0)	.53
Road user category, n (%)				.06
Passenger	11 (25.0)	32 (44.4)	43 (37.1)	
Pedestrian	9 (20.5)	16 (22.2)	25 (21.6)	
Motorcyclist	24 (24.5)	24 (33.3)	48 (41.4)	
Mode of arrival, n (%)				.90
Ambulance	4 (9.1)	9 (12.5)	13 (11.2)	
Public means (taxi or cycle)	39 (88.6)	61 (84.7)	100 (86.2)	
Other	1 (1.4)	2 (2.8)	3 (2.6)	
Prehospital care (first aid) received?				.71
Yes	26 (59.0)	40 (55.6)	66 (56.9)	
No	18 (40.9)	32 (44.4)	50 (43.1)	
Who gave first aid?				.83
Health worker	14 (31.8)	24 (33.3)	38 (32.8)	
Lay bystander/police	4 (9.1)	8 (11.1)	12 (10.3)	
Mechanism of injury, n (%)				.36
Motorcycle-motorcycle crash	19 (43.2)	41 (56.9)	60 (51.7)	
Motorcycle-pedestrian	13 (29.5)	20 (27.8)	33 (28.4)	
Motorcycle-car crash	9 (20.5)	9 (12.5)	18 (15.5)	
Motorcycle static object	3 (6.8)	2 (2.8)	5 (4.3)	
Comorbidity (chronic medical illness present), n (%)				.40
Hypertension	3 (100.0)	1 (33.3)	4 (66.7)	
Asthma	0 (0.0)	2 (66.7)	2 (33.3)	
Injury severity (KTS)[a], n (%)				.68
Mild (9-10)	25 (56.8)	41 (56.9)	66 (56.9)	

XSL•FO
RenderX

Characteristic and category	Intervention group (n=44)	Control group (n=72)	Overall (n=116)	*P* value
Moderate (7-8)	14 (31.8)	26 (36.1)	40 (34.5)	
Severe ≤6	5 (11.4)	5 (6.9)	10 (8.6)	
Score for serious injuries, n (%)				.66
>1	4 (9.1)	8 (11.1)	12 (10.3)	
1	32 (72.7)	52 (72.2)	87 (75.00	
None	5 (11.4)	12 (16.7)	17 (14.7)	

[a]KTS: Kampala Trauma Score.

Given the statistically significant differential attrition in loss to follow-up between arms, we simulated the base, worst-, and best-case scenarios. We found that overall, missing data did not impact the interpretation of results of both primary and secondary outcomes under the different scenarios except for mortality (14%, 68/483 in the intervention vs 12%, 58/470 in the control group; *P*=.34) and morbidity based on unfavorable GOS (1-3) for neurological injuries (14%, 68/483 in the intervention vs 19%, 87/470 in the control group; *P*=.06), which did not reach statistical significance in the worst-case scenario (Multimedia Appendix 14). However, the results from multiple imputations were congruent with complete-case analysis, with all primary and secondary outcomes (mortality and GOS) reaching statistical significance (all *P*<.01; Table 5). Note that the estimates for random effects parameters were computed in mixed effects restricted maximum likelihood (REML) regression models using the Satterthwaite method of adjustment of degrees of freedom for small sample correction. Number of study arms=2 (intervention vs control); number of discrete time decay study periods=12; total number of clusters=6; number of clusters per treatment arm=3; sample size may vary due to variation in study outcomes. All analysis variables were included in the model regardless of completeness. The summary presented is a pooled analysis of 20 imputations. Imputed is the minimum across "*m*" of the number of filled-in observations. Imputation method: Multiple imputation by chained equation (MICE). Number of imputations (*m*)=20. Assumption: Missing at random or completely missing at random. . The corresponding imputed variables with loss to follow-up (missing data) are provided in Multimedia Appendix 15, whereas the residual estimates after multiple imputation are provided in Multimedia Appendix 16.

XSL•FO
RenderX

Table . Multiple imputation and comparison of primary and secondary outcomes across study periods after adjustment for small sample correction.

Outcome variable (no of observations)	Average cluster size per period, n (min-max)	Average sample per period, n (min-max)	SE	Satterthwaite adjusted df	Coefficient (intervention group; 95% CI)	*P* value	Residual ICC[a] at level of study period [95% CI]
Primary outcome 1: prehospital interval (hour; n=1003)	14 (13 - 15)	83 (80 - 88)	0.08	4	–1.14 (–1.30 to –0.98)	<.001	0.02 (0.00-0.07)
Primary outcome 2: referral-dispatch interval (hour; n=691)	10 (7 - 11)	58 (42 - 68)	0.08	4	–1.41 (–1.57 to –1.25)	<.001	0.01 (0.00-0.08)
Secondary outcome 1: percentage of 90-day mortality (%; n=1003)	14 (13 - 15)	83 (80 - 88)	0.06	4	–0.24 (–0.35 to 0.11)	<.001	0.01 (0.00-0.06)
Secondary outcome 2: percentage of unfavorable Glasgow Outcome Scale (GOS[b]; %; n=1003)	14 (13-15)	83 (80 - 88)	0.02	4	–1.10 (–0.14 to 0.06)	<.001	0.01 (0.00-0.07)
Secondary outcome 3: percentage of unfavorable Trauma Outcome Measure Score (TOMS[c]; %; n=637)	9 (7-11)	53 (42 - 64)	0.04	4	–0.02 (–0.09 to 0.05)	.56	0.03 (0.01-0.10)

[a]ICC: intracluster correlation coefficient.

Posthoc Power Analyses

The posthoc analyses indicated that, except for secondary outcomes related to the TOMS for musculoskeletal injuries, the sample sizes used in the study were sufficient for detecting effect sizes associated with the various study outcomes at a power level of 80%, based on both predetermined and observed ICC and study assumptions (Multimedia Appendix 17). We did not apply corrections for multiple testing, as the study used marginal rather than disjunctive power. However, with our 5 key study outcomes at a significance level of $P<.05$, the differences in primary outcomes (prehospital time and referral-dispatch interval) would remain statistically significant at $P<.01$ if the Bonferroni method (ie, 0.05/5) were used. Furthermore, in our exploratory analyses of secondary outcomes with distinct constructs (ie, mortality, GOS, and TOMS), the risk of missing a potentially significant finding was deemed a greater concern than the occurrence of false positives.

Discussion

Principal Findings

To our knowledge, this study represents a pioneering cluster-randomized trial assessing the impact of a medical education program (RTTDC) on clinical processes and patient outcomes in an LMIC setting. This trial found that RTTDC training significantly reduced prehospital time, highlighting its potential for improved budget allocation for prehospital logistical systems aimed at enhancing care for neurological and orthopedic injuries in these settings. Prior research, primarily observational or quasi-experimental, has indicated similar reductions in referral decision times and transfer intervals; for instance, a 1.91-hour decrease was reported in Western Virginia hospitals compared to control groups [36]. Further findings from this trial indicated a reduction in the interval from referral decision to hospital dispatch for patients who required interfacility transfer ($P<·001$). These results align with a recent study involving 472 trauma patients in the United States, which documented a similar decrease in emergency department dwell time [7]. However, these studies often faced limitations due to small sample sizes and potential biases. Poor interhospital coordination, dysfunctional referral systems, and transportation were the most important contributing factors to delayed emergency surgical access in rural settings, alongside human, financial, and infrastructural resource challenges, as this trial demonstrated that most trauma patients arrived by public means, which may apply to other LMICs.

The trial's overall 90-day mortality rate was comparable to prior Ugandan studies but higher than rates reported in the US and European cohorts [37,38]. Notably, the intervention group exhibited significantly lower mortality rates than the control

XSL•FO
RenderX

group in both complete-case analysis and best-case scenarios, suggesting that RTTDC implementation and the coordination of rural trauma teams may effectively minimize mortality associated with prolonged referral decision times. In this trial, referral decision to dispatch time of more than 1 hour was associated with a fourfold increase in mortality after controlling for confounding due to injury mechanisms and severity. This finding is particularly relevant given that delays in care have been shown to exacerbate injury-related mortality in both LMICs and high-income countries [37]. Consistent outcomes were observed in previous studies, including a notable 50% reduction in mortality rates associated with RTTDC training in the United States [23]. Despite these potential positive results, variations in mortality outcome rates across studies exist and may stem from different methodological approaches, sample sizes, and settings [7]. In this study, the mortality rate and operative neurosurgical interventions were relatively high, partly due to the restriction of the study to motorcyclists in rural settings where helmet usage and speed limits are poorly regulated, resulting in severe TBI. In the case of 8 deaths in emergency departments of control centers, it was determined that patients were awaiting head CT scans at the request of trainee general surgery residents to decide on transfers for neurosurgical evaluation. However, no local neurosurgeons were available to address potential surgical needs based on the CT results. This highlights the lesson from RTTDC that investigations should not be requested without the requisite human resources to act on the findings.

In our settings, the impact of RTTDC was driven by the actualization of physical trauma teams through pairing junior and senior medical trainees, effective communication and coordination at the community level, facilitated by trained volunteer traffic police, and guided by interns and surgery residents in emergency departments. This collaboration ensured proper selection of destination hospitals, prearrival preparations, and swift interfacility transfers for injuries exceeding local capacity. Key strengths of RTTDC included emphasizing teamwork, closed-loop communication within trauma teams, and optimizing local resources based on facility capabilities. These are elements that other trauma education courses in LMICs should replicate. Recognizing that forgetting is a barrier to skill application, we reinforced knowledge transfer through weekly clinical audits and regular feedback to trauma teams, which helped maintain focus on course objectives. This decision-making support was critical, as previous research indicated that only 31% of district and regional referral hospitals in Uganda had emergency unit protocols [39].

Additionally, prior trauma education interventions, such as the Kampala Advanced Trauma Course, showed limited impact on facility-wide best practices when specialist surgeons and experienced medical officers were the main participants due to inadequate skill transfer to junior faculty [40]. Our experience in audit meetings revealed that local consultant surgeons, although willing to assist, were sometimes unreachable when surgical residents required their expertise in emergency situations. We found the “train the trainer” approach effective for fostering team cohesion by pairing third-year medical students with fifth-year students, interns, and residents during night duty calls. Therefore, a comprehensive understanding of the strengths and weaknesses of the health care system for which a trauma course or a medical education program is to be implemented is critical for its effectiveness. Our study underscores the importance of involving frontline workers, such as trainee surgeons, in trauma training initiatives. To address this gap, 3 local universities in Uganda have initiated postgraduate medical education programs in emergency medicine, although these roles are not yet integrated into the national civil service staffing structure by the government. Long-term local trauma fellowship programs can accelerate this goal and address issues of sustainability.

Knowledge reinforcement through support audit meetings also emphasized the necessity of routine community engagement with traffic police, as there was no structured approach to enhancing community awareness or providing emergency care at accident scenes in Uganda during the study. Ideal Level II and III trauma centers should implement trauma quality assessment and outreach programs for their communities. The meetings revealed practical insights into team dynamics and technological resource challenges, including network connectivity and difficulties in locating ambulances. These issues were addressed by coordinating with volunteer taxi drivers who facilitated rescue evacuations under the guidance of traffic police, general surgery trainees, or intern physicians. This strategy improved the selection of destination facilities and enabled concise handover reports from police, who are first responders in most emergencies, since only 25% of Uganda’s 48 million population had access to ambulances [13].

These lessons highlight the urgent need for investment in surgical education, emergency medical services (EMSs), telemedicine, and remote monitoring systems such as edge computing to address connectivity issues, including the establishment of toll-free emergency numbers that link police to hospitals and ambulances. Additionally, further medical training for lay responders and efforts to increase the number of paramedics are essential. Uganda’s national EMS policy, launched in 2021, aims to enhance emergency care capacity, but fixed staff occupancy in district hospital emergency departments remains below 50% [39,41]. Integrating artificial intelligence–enabled triage, referral decision support, and imaging prioritization into rural trauma workflows may overcome some of the human resource constraints to reduce delays in the trauma care continuum in low-resource settings. However, enhancing patient outcomes through artificial intelligence in these contexts necessitates robust ethical oversight, fair dataset representation from LMICs, and ongoing validation via prospective clinical trials [42].

Contemporaneous evidence indicates that community responders play a significant role in both current and future EMS, even in countries with established prehospital care systems, such as the United Kingdom, due to their timely responses and reduced maintenance costs [43]. The Ugandan EMS policy acknowledges the potential contributions of community responders [41]. However, to actualize these contributions, it necessitates a separate budget to reimburse costs associated with communication and transportation, support training that bridges lay responders to clinicians, and facilitate medical interns and

XSL•FO
RenderX

residents who work overtime to address staffing shortages in emergency departments (resource optimization). Moreover, it is essential to design legal protections for community responders through Good Samaritan laws, establish clear standard operating procedures with decision support algorithms for trainee physicians, develop referral protocols with automated workflow, and implement performance measurement frameworks for improvement. These measures would prevent fatalities during transport, including auditing incidents that occur before ambulance arrival [44]. Consequently, it is crucial for medical training institutions, public-private partnerships, ambulance and insurance companies, and civil society organizations to consider investing in this area.

In addition to mortality, the study explored morbidity outcomes, wherein the intervention group had a lower proportion of unfavorable GOS-TBI scores compared to controls, indicating improved recovery trajectories. However, the analysis of TOMS for musculoskeletal injuries was inconclusive, possibly due to the study being underpowered for this particular outcome, a higher proportion of polytrauma cases, and higher initial trauma outcome expectations among intervention participants. Existing literature corroborates the increased morbidity associated with polytrauma, emphasizing the need for future research into the long-term outcomes of RTTDC interventions in the context of multiple injuries [45]. On the other hand, the nonsignificant results for musculoskeletal injuries may stem from the inherent focus of the RTTDC on life-saving resuscitation measures and the immediate transfer of critical cases such as neurotrauma, which require high care demands. Alternatively, it may be that the evaluation tools (TEFS and TOMS) are less sensitive to assessing outcomes related to musculoskeletal injuries within the trauma education context. To resolve this uncertainty, further investigation is warranted with a larger number of clusters and cluster sizes.

Policy Implications

The implications of our findings are threefold. First, while RTTDC training shows promise for enhancing clinical processes and patient outcomes in LMICs, local and regional evaluations of its impact are essential for effective implementation. Establishing regional continuing medical education certification centers and long-term postgraduate trauma training programs for trauma care providers could facilitate this goal and address concerns about sustainability. Second, the development of context-specific rural trauma transfer guidelines is critical, incorporating local norms and logistical realities. This may necessitate broader participation from various health care professionals, including medical trainees, private ambulance companies, and lay traffic police who provide immediate crash responses, to strengthen rural trauma systems. Finally, a tailored needs assessment for each LMIC is crucial to adapt trauma education programs, ensuring they address specific cultural, technological, and systemic challenges while promoting sustainability. Existing trauma education programs in LMICs lacked robust impact evaluation, cohesive local contextualization, and common outcome metrics for reporting [3]. This trial provided critical clinical processes and patient-reported indices that could constitute a core outcome set for interventions linking medical education to patient outcomes in a trauma context.

Study Strengths and Limitations

This trial offers significant strengths in evaluating a locally contextualized trauma education program and linking it to clinical outcomes in resource-limited rural settings. It effectively demonstrates that training trauma care frontline workers, such as voluntary traffic police officers and medical trainees, can enhance clinical practices and patient outcomes. We previously highlighted the impact of RTTDC on care providers' knowledge and behavioral change [14]. This trial builds on that foundation by showing the translation of acquired classroom knowledge into improved clinical processes. Early engagement with frontline workers ensured that the trial was relevant, and the findings may inform future core outcome sets for similar trials in LMICs, where this study forms an integral component of pioneering efforts to evaluate the real-world impact of an education program (RTTDC) in controlled clinical settings. Studies of this nature are topical for the global surgery community. This study provides an ideal example of how equitable South-West partnerships in global surgery should be implemented. A detailed statement regarding how partnerships and collaborations were actualized is available in Multimedia Appendix 18. There are no previous large-scale cluster randomized controlled trials that have evaluated the effect of trauma education programs on patient workflows and clinical outcomes, but a recent feasibility trial evaluating ATLS course and PTC against standard care in India found justification for full-scale cluster randomized trials [46], thus our publicly available data in the Multimedia Appendix 19 is timely [47].

However, this study has 6 notable limitations. First, a higher loss to follow-up in the control group may introduce potential systematic attrition bias, which must be addressed in future studies by assuming the worst-case situation during sample size computation. Although the overall dropout rate (11.6%) remains within acceptable limits, the conclusions regarding mortality and morbidity must be cautiously interpreted in this context. The differential attrition could suggest that missingness was likely related to the effectiveness of the intervention, with plausibility that participants in the control group without perceived improvement due to unfavorable trauma outcomes or other unmeasured postrandomization experiences were more likely to be lost to follow-up. Furthermore, the ICC was estimated based solely on primary outcomes, which may not accurately reflect variations in cluster sizes for secondary outcomes, as observed with TOMSs for musculoskeletal injuries. This is a common issue in cluster-randomized trial design.

Second, some patients at intervention sites may not have received care from trained providers, limiting the intervention's individual-level impact. Due to significant shortages of health care personnel in Uganda, we set the trained staff capacity in emergency units at 80%, as complete participation in training was not feasible at any given time. This strategy addressed human resource constraints while minimizing ceiling effects associated with training and fostering information sharing among trauma team members in our train-the-trainer model. Given our trauma "team approach" to injury care, which deployed a

XSL•FO
RenderX

5-member in-hospital team for each patient, it was not feasible to provide cluster-level proportions of trained providers and their association with outcomes. However, 80% coverage may result in occasional individual-level exposure misclassification, with a 1:4 likelihood that a patient may encounter 1 untrained vs 4 trained providers during the continuum of care within a typical 5-member trauma team at the same intervention hospital. Consequently, any potential biases in estimates are likely to trend away from the null.

Third, the generalizability of findings is constrained to those who underwent head CT scans in a single country's health care system, excluding emerging prognostic factors for neurological injuries such as blood-based biomarkers.

Fourth, although exploratory analyses indicated no significant changes in prehospital time due to the COVID-19 pandemic, its broader social effects on hospital surroundings, resuscitative capability, injury patterns, and patient-reported morbidity outcomes remained challenging to quantify. Moreover, COVID-19's extenuating effects on the trial, including study halt, retrospective registration, and need for posthoc data stratification by pandemic status, which was not hypothesized beforehand, could introduce bias. However, we are unaware of any other events that could have influenced referral decisions and morbidity outcomes.

Fifth, the absence of preintervention data at the study sites represents a limitation. Ideally, if data had been available before training, a comparison of injury outcomes before and after the training would have been preferable to the current approach of making parallel comparisons. Moreover, the data constraints on prehospital deaths may systemically bias estimates of prehospital time and mortality outcomes, introducing survivor bias. For instance, we could not establish if the intervention reduced prehospital deaths, an issue the Ugandan government can address by mandating all unnatural prehospital deaths to undergo postmortem and forensic analysis to determine the cause of death. Although there were 8 mortalities in the emergency department of control hospitals vs zero in the intervention group, and the principal determinant of time-to-event (death), that is, requiring a craniotomy, did not differ between groups, which would trend any biased estimates away from the null; the authors recommend that the study conclusions should only be applied to patients surviving up to hospital arrival.

Finally, despite sensitivity analyses, we are aware that the relatively smaller number of clusters and cluster sizes, and aggregating intracranial lesions as axial or extra-axial or dichotomizing the GOS and TOMS tools as favorable vs unfavorable outcomes, could have led to loss of information and left unmeasured residual confounding that can potentially contribute to some of the observed associations. Moreover, we were also unable to account for the heterogeneity across the considerably wide range of our participants' age groups, despite trauma not being segregative, an issue future studies should address.

Conclusions

This trial provided reasonable evidence to suggest that rural trauma team development training potentially decreased prehospital time and referral decision intervals, lowering motorcycle accident–related mortality from neurological and musculoskeletal injuries, without negatively impacting patient-reported morbidity. Despite potential measurement biases from the exclusion of prehospital deaths due to data constraints and the relatively smaller number of clusters, these findings highlight the importance of enhancing rural trauma team capabilities and supporting ongoing investment in global trauma education programs. Future research should focus on large, multicountry studies using long-term trauma registries with diverse injury mechanisms and patient-centered quality-of-life outcome measures tailored to local contexts in LMICs.

Acknowledgments

We are grateful to Asiimwe Daniel, Namutosi Esther, Kabarokole Annet, Mwine David, Ssekabembe John, Japheth Comfort, Nalumansi Hilda, and Rwamukaga Nasurudin for their invaluable assistance in the data collection process and for consenting to the publication of their names. Additionally, we thank Mary Nnabagulanyi for her support in the off-site administration and coordination of this trial and appreciate our research participants for their commitment and time dedicated to the study. The authors declare the use of generative artificial intelligence (AI) in the research writing process. Under full human supervision, Microsoft CoPilot and Poe Assistant were used during proofreading, editing, and refinement of grammatical errors. Responsibility for the final paper lies entirely with the authors. Inequalities in access to injury care and trauma training in rural low- and middle-income countries (LMICs) stem from a lack of specialized workforce, trauma resources, and technological advancement. To address this, the principal investigator (PI), a specialist surgeon and perioperative fellow at the Centre for Health Equity in Surgery and Anesthesia at the University of California, San Francisco, was stationed at local trauma centers in Uganda during recruitment. The PI built capacity through the Rural Trauma Team Development Course (RTTDC) training for surgical residents, who then trained interns. This work is part of a broader doctoral project at the University of Turku in Finland, which examines the potential role of allied health and law enforcement trauma teams and registries as underused resources for trauma care. This project underscores the importance of high-fidelity interdisciplinary collaborations in educational and trauma research, as well as the value of community presence in areas designated for global surgery interventions.

XSL•FO
RenderX

Funding

HL was supported by the University of Turku Graduate School, Turku University Hospital (TYKS) Foundation, TYKS Neurocenter, and the University of California, San Francisco (UCSF)–Center for Health Equity in Surgery and Anesthesia (CHESA) through participation in fellowships. JPP was supported by the Academy of Finland (grant nos 17379 and 60063) and the Maire Taponen Foundation. The funders has no role in the study design, implementation, or reporting. All authors had full access to the data and accept responsibility to submit for publication.

Data Availability

The datasets generated or analyzed during this study are available in the MendeleyMendeley repository [47], and are included in this published article. Further details are available in the study protocol and trial registry sections.

Authors' Contributions

HL contributed to conceptualization, data curation, investigation, methodology, project administration, resources, formal analysis, software, visualization, writing (original draft), writing (review and editing), and funding acquisition. BO, AK, HML, RS, and PK contributed to data curation, investigation, validation, and writing (review and editing). MM, AA, and AD contributed to formal analysis, software, visualization, validation, and writing (review and editing). FJW contributed to investigation, validation, and writing (review and editing). TB contributed to investigation, validation, writing (review and editing), and funding acquisition. JPP and MLW contributed to investigation, funding acquisition, supervision, validation, and writing (review and editing).

Conflicts of Interest

The authors declare that no competing interests exist. Early results of this work were shared with stakeholders at the Safety 2024 World Conference on Injury Prevention and Safety Promotion in New Delhi. The views expressed in this report are solely the authors' responsibility and do not represent the official views of their affiliated institutions.

Editorial Notice

The trial used data from the piloted Motorcycle Trauma Outcome Registry (MOTOR) project. At the time of protocol submission, the ethics committee did not mandate prior registration for academic, non–industry-funded proof-of-concept projects (policy effective December 31, 2023). Following changes in local policy, the investigators pursued registration with the Pan African Clinical Trial Registry. Due to COVID-19–related disruptions, registration was completed after recruitment had commenced.

Multimedia Appendix 1
Psychometric properties of data collection tools and interpretation.
[DOCX File, 34 KB - humanfactors_v13i1e82591_app1.docx]

Multimedia Appendix 2
Protocol deviations and implementation fidelity.
[DOCX File, 20 KB - humanfactors_v13i1e82591_app2.docx]

Multimedia Appendix 3
Baseline sociodemographic and clinical characteristics of participants.
[DOCX File, 40 KB - humanfactors_v13i1e82591_app3.docx]

Multimedia Appendix 4
Adjusted predictions of prehospital time between arms across study periods, with CIs.
[DOCX File, 109 KB - humanfactors_v13i1e82591_app4.docx]

Multimedia Appendix 5
Factors associated with delayed prehospital time and referral-dispatch interval of more than 1 hour.
[DOCX File, 35 KB - humanfactors_v13i1e82591_app5.docx]

Multimedia Appendix 6
Adjusted stratum-by-time predictions of referral-dispatch interval between arms across study periods.
[DOCX File, 96 KB - humanfactors_v13i1e82591_app6.docx]

Multimedia Appendix 7
Stratified Cox regression analysis of determinants of time to event (death).

XSL•FO
RenderX

[DOCX File, 21 KB - humanfactors_v13i1e82591_app7.docx]

Multimedia Appendix 8
Subgroup analysis of survival time (days) stratified by sex.
[DOCX File, 74 KB - humanfactors_v13i1e82591_app8.docx]

Multimedia Appendix 9
Subgroup analysis of survival time (days) by road user category.
[DOCX File, 83 KB - humanfactors_v13i1e82591_app9.docx]

Multimedia Appendix 10
Comparison of Glasgow Outcome Score at 90 days between treatment groups.
[DOCX File, 15 KB - humanfactors_v13i1e82591_app10.docx]

Multimedia Appendix 11
Subgroup analysis of Trauma Expectation Factor Score (TEFS) and Trauma Outcome Measure Score (TOMS) for participants with tibial fractures.
[DOCX File, 39 KB - humanfactors_v13i1e82591_app11.docx]

Multimedia Appendix 12
Trauma Expectation Factor Scores (TEFSs) and Trauma Outcome Measure Scores (TOMSs) across study periods.
[DOCX File, 35 KB - humanfactors_v13i1e82591_app12.docx]

Multimedia Appendix 13
Corresponding residual estimates.
[DOCX File, 15 KB - humanfactors_v13i1e82591_app13.docx]

Multimedia Appendix 14
Sensitivity analyses of injury outcomes under different case scenarios.
[DOCX File, 21 KB - humanfactors_v13i1e82591_app14.docx]

Multimedia Appendix 15
Imputed variables with loss to follow-up missing data.
[DOCX File, 15 KB - humanfactors_v13i1e82591_app15.docx]

Multimedia Appendix 16
Residual estimates for random effects parameters after multiple imputations for missing data.
[DOCX File, 14 KB - humanfactors_v13i1e82591_app16.docx]

Multimedia Appendix 17
Post hoc power and sample size analyses for various observed study outcomes.
[DOCX File, 572 KB - humanfactors_v13i1e82591_app17.docx]

Multimedia Appendix 18
Reflexivity statement regarding collaborative partnerships.
[DOCX File, 24 KB - humanfactors_v13i1e82591_app18.docx]

Multimedia Appendix 19
Data analysis code and dataset.
[ZIP File, 10 KB - humanfactors_v13i1e82591_app19.zip]

Checklist 1
CONSORT (Consolidated Standards of Reporting Trials) 2025 checklist.
[PDF File, 221 KB - humanfactors_v13i1e82591_app20.pdf]

XSL•FO
RenderX

Checklist 2
CONSORT (Consolidated Standards of Reporting Trials) 2010 guidelines for cluster randomized trials.
[PDF File, 196 KB - humanfactors_v13i1e82591_app21.pdf]

Checklist 3
Consolidated Standards of Reporting Trials and SPIRIT Extension for Randomized Controlled Trials Revised in Extenuating Circumstances (CONSORT-CONSERVE) checklist version 2021.
[PDF File, 229 KB - humanfactors_v13i1e82591_app22.pdf]

References

1. Alayande B, Chu KM, Jumbam DT, et al. Disparities in access to trauma care in sub-Saharan Africa: a narrative review. Curr Trauma Rep 2022;8(3):66-94. [doi: 10.1007/s40719-022-00229-1] [Medline: 35692507]
2. Ahmat A, Okoroafor SC, Asamani JA, et al. Health workforce policy and plan implementation in the context of universal health coverage in the Africa Region. BMJ Glob Health 2022 May;7(Suppl 1):e008319. [doi: 10.1136/bmjgh-2021-008319]
3. Livergant RJ, Demetrick S, Cravetchi X, et al. Trauma training courses and programs in low‐ and lower middle‐income countries: a scoping review. World J Surg 2021 Dec;45(12):3543-3557. [doi: 10.1007/s00268-021-06283-1] [Medline: 34486080]
4. Kadhum M, Sinclair P, Lavy C. Are Primary Trauma Care (PTC) courses beneficial in low- and middle-income countries - a systematic review. Injury 2020 Feb;51(2):136-141. [doi: 10.1016/j.injury.2019.10.084] [Medline: 31679834]
5. Brown HA, Tidwell C, Prest P. Trauma training in low- and middle-income countries: a scoping review of ATLS alternatives. Afr J Emerg Med 2022 Mar;12(1):53-60. [doi: 10.1016/j.afjem.2021.11.004] [Medline: 35070655]
6. Rural Trauma Team Development Course. American College of Surgeons. URL: https://www.facs.org/quality-programs/trauma/education/rural-trauma-team-development-course/ [accessed 2025-06-21]
7. Bauman ZM, Khan H, Raposo-Hadley A, et al. Rural trauma team development course positively impacts its desired objectives. Am Surg 2024 Jun;90(6):1250-1254. [doi: 10.1177/00031348241227205] [Medline: 38217436]
8. Bauman ZM, Loftus J, Hodson A, et al. Rural trauma team development course instills confidence in critical access hospitals. World J Surg 2020 May;44(5):1478-1484. [doi: 10.1007/s00268-019-05359-3] [Medline: 31894357]
9. Schmidtke KA, Evison F, Grove A, et al. Surgical implementation gap: an interrupted time series analysis with interviews examining the impact of surgical trials on surgical practice in England. BMJ Qual Saf 2023 Jun;32(6):341-356. [doi: 10.1136/bmjqs-2022-015077]
10. Zheng DJ, Sur PJ, Ariokot MG, Juillard C, Ajiko MM, Dicker RA. Epidemiology of injured patients in rural Uganda: a prospective trauma registry's first 1000 days. PLoS ONE 2021;16(1):e0245779. [doi: 10.1371/journal.pone.0245779] [Medline: 33481891]
11. Kalanzi J, Wallis L, Nabukenya M, Okello E, Okong D, Namirembe S. Injury patterns in patients with severe traumatic brain injuries from motor crashes admitted to Mulago hospital accidents & emergency unit. Afr J Emerg Med 2023 Jun;13(2):94-100. [doi: 10.1016/j.afjem.2023.03.003]
12. Anderson LN, Merkebu J. The Kirkpatrick Model: a tool for evaluating educational research. Fam Med 2024 Jun;56(6):403. [doi: 10.22454/FamMed.2024.161519] [Medline: 38996342]
13. Ningwa A, Muni K, Oporia F, et al. The state of emergency medical services and acute health facility care in Uganda: findings from a National Cross-Sectional Survey. BMC Health Serv Res 2020 Dec;20(1):634. [doi: 10.1186/s12913-020-05508-8]
14. Lule H, Mugerwa M, Ssebuufu R, Kyamanywa P, Posti JP, Wilson ML. Prospective multi-centre analysis of rural trauma team development training for medical trainees and traffic law enforcement professionals in Uganda: an interrupted time series study. BMC Med Educ 2025 Jan 29;25(1):148. [doi: 10.1186/s12909-025-06755-1] [Medline: 39881413]
15. Campbell MK, Piaggio G, Elbourne DR, Altman DG, CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. BMJ 2012 Sep 4;345:e5661. [doi: 10.1136/bmj.e5661] [Medline: 22951546]
16. Orkin AM, Gill PJ, Ghersi D, et al. Guidelines for reporting trial protocols and completed trials modified due to the COVID-19 pandemic and other extenuating circumstances. JAMA 2021 Jul 20;326(3):257-265. [doi: 10.1001/jama.2021.9941] [Medline: 34152382]
17. Create a randomisation list. Sealed Envelope Ltd. URL: https://www.sealedenvelope.com/simple-randomiser/v1/lists [accessed 2024-04-11]
18. Lule H, Mugerwa M, SSebuufu R, Kyamanywa P, Posti JP, Wilson ML. Rural trauma team development training amongst medical trainees and traffic law enforcement professionals in a low-income country: a protocol for a prospective multicenter interrupted time series. Int J Surg Protoc 2024 Mar;28(1):12-19. [doi: 10.1097/SP9.0000000000000013] [Medline: 38433864]
19. Suk M, Daigl M, Buckley RE, et al. TEFTOM: a promising general trauma expectation/outcome measure—results of a validation study on Pan-American ankle and distal tibia trauma patients. ISRN Orthop 2013;2013:801784. [doi: 10.1155/2013/801784] [Medline: 24967116]

XSL•FO
RenderX

20. Wilson L, Boase K, Nelson LD, et al. A manual for the Glasgow Outcome Scale-Extended interview. J Neurotrauma 2021 Sep 1;38(17):2435-2446. [doi: 10.1089/neu.2020.7527] [Medline: 33740873]
21. Lule H, Mugerwa M, Ssebuufu R, et al. Effect of rural trauma team development on the outcomes of motorcycle accident–related injuries (Motor Registry Project): protocol for a multicenter cluster randomized controlled trial. JMIR Res Protoc 2024 May 7;13:e55297. [doi: 10.2196/55297] [Medline: 38713507]
22. Ouyang Y, Hemming K, Li F, Taljaard M. Estimating intra-cluster correlation coefficients for planning longitudinal cluster randomized trials: a tutorial. Int J Epidemiol 2023 Oct 5;52(5):1634-1647. [doi: 10.1093/ije/dyad062] [Medline: 37196320]
23. Malekpour M, Neuhaus N, Martin D, et al. Changes in rural trauma prehospital times following the Rural Trauma Team Development Course training. Am J Surg 2017 Feb;213(2):399-404. [doi: 10.1016/j.amjsurg.2016.06.015] [Medline: 27575601]
24. Shaikh F, Munakomi S, Waseem M. Head Trauma (Archived): StatPearls Publishing; 2025. URL: https://www.ncbi.nlm.nih.gov/books/NBK430854/ [accessed 2026-04-09]
25. Ssebakumba MK, Lule H, Olweny F, Mabweijano J, Kiryabwire J. Thirty-day clinical outcome of traumatic brain injury patients with acute extradural and subdural hematoma: a cohort study at Mulago National Referral Hospital, Uganda. Egypt J Neurosurg 2020 Dec;35(1):3. [doi: 10.1186/s41984-020-0073-4]
26. Jain S, Margetis K, Iverson LM. Glasgow Coma Scale: StatPearls Publishing; 2025. URL: https://www.ncbi.nlm.nih.gov/books/NBK513298/ [accessed 2026-04-09]
27. Suk M, Daigl M, Buckley RE, Lorich DG, Helfet DL, Hanson B. Outcomes after orthopedic trauma: are we meeting patient expectations?—a prospective, multicenter cohort study in 203 patients. J Orthop Surg 2017 Jan;25(1):2309499016684089. [doi: 10.1177/2309499016684089]
28. McMillan T, Wilson L, Ponsford J, Levin H, Teasdale G, Bond M. The Glasgow Outcome Scale - 40 years of application and refinement. Nat Rev Neurol 2016 Aug;12(8):477-485. [doi: 10.1038/nrneurol.2016.89] [Medline: 27418377]
29. Damulira J, Muhumuza J, Kabuye U, et al. New Trauma Score versus Kampala Trauma Score II in predicting mortality following road traffic crash: a prospective multi-center cohort study. BMC Emerg Med 2024 Jul;24(1):130. [doi: 10.1186/s12873-024-01048-0]
30. Hsieh SL, Hsiao CH, Chiang WC, et al. Association between the time to definitive care and trauma patient outcomes: every minute in the golden hour matters. Eur J Trauma Emerg Surg 2022 Aug;48(4):2709-2716. [doi: 10.1007/s00068-021-01816-8]
31. Hammond FM, Ketchum JM, Patni VV, Nejadnik B, Bates D, Weintraub AH. Determining the minimally clinically important difference for the disability rating scale in persons with chronic traumatic brain injury. Neurotrauma Rep 2023;4(1):447-457. [doi: 10.1089/neur.2023.0038] [Medline: 37475977]
32. Dvir Z. Difference, significant difference and clinically meaningful difference: the meaning of change in rehabilitation. J Exerc Rehabil 2015 Apr;11(2):67-73. [doi: 10.12965/jer.150199] [Medline: 25960978]
33. Fang C, Platz A, Müller L, et al. Evaluation of an expectation and outcome measurement questionnaire in ankle fracture patients: the Trauma Expectation Factor Trauma Outcomes Measure (TEFTOM) Eurasia study. J Orthop Surg (Hong Kong) 2020;28(1):2309499019890140. [doi: 10.1177/2309499019890140] [Medline: 31916492]
34. Lule H, Mugerwa M, Abio A, et al. Factors associated with all-cause mortality and morbidity of motorcycle crash-related neurological and musculoskeletal injuries in Uganda: the MOTOR cluster randomised trial ancillary study. Inj Prev 2026. [doi: 10.1136/ip-2025-045737]
35. Lule H, Oguttu B, Mugerwa M, Wilson ML, Posti JP. Understanding barriers to effective injury care in low-income contexts: a mixed methods analysis of insights from medical trainees and traffic law enforcement first responders in Uganda. medRxiv. Preprint posted online on Feb 23, 2025. [doi: 10.1101/2025.02.22.25322712]
36. Kappel DA, Rossi DC, Polack EP, Avtgis TA, Martin MM. Does the rural trauma team development course shorten the interval from trauma patient arrival to decision to transfer? J Trauma 2011;70(2):315-319. [doi: 10.1097/TA.0b013e318209589e]
37. Vaca SD, Kuo BJ, Nickenig Vissoci JR, et al. Temporal delays along the neurosurgical care continuum for traumatic brain injury patients at a tertiary care hospital in Kampala, Uganda. Neurosurgery 2019 Jan 1;84(1):95-103. [doi: 10.1093/neuros/nyy004] [Medline: 29490070]
38. Nelson LD, Temkin NR, Barber J, et al. Functional recovery, symptoms, and quality of life 1 to 5 years after traumatic brain injury. JAMA Netw Open 2023 Mar 1;6(3):e233660. [doi: 10.1001/jamanetworkopen.2023.3660] [Medline: 36939699]
39. Wasukira SB, Kambugu CT, Nanyondo S J, et al. Implementation of a cascade training model to enhance emergency care capacity of healthcare workers during the COVID-19 outbreak in Uganda. Afr J Emerg Med 2025 Mar;15(1):565-570. [doi: 10.1016/j.afjem.2025.01.001]
40. Tang Z, Kayondo D, Ullrich SJ, et al. Clinical practices following Train‐The‐Trainer Trauma Course completion in Uganda: a parallel‐convergent mixed‐methods study. World J Surg 2023 Jun;47(6):1399-1408. [doi: 10.1007/s00268-023-06935-4] [Medline: 36872370]
41. National Emergency Medical Services Policy. Ministry of Health - Uganda. URL: https://library.health.go.ug/leadership-and-governance/national-emergency-medical-services-policy [accessed 2025-06-18]
42. Ogut E. Artificial intelligence in clinical medicine: challenges across diagnostic imaging, clinical decision support, surgery, pathology, and drug discovery. Clin Pract 2025 Sep 16;15(9):169. [doi: 10.3390/clinpract15090169] [Medline: 41002784]

XSL•FO
RenderX

43. Siriwardena AN, Patel G, Botan V, et al. Community first responders' role in the current and future rural health and care workforce: a mixed-methods study. Health Soc Care Deliv Res 2024 Jul;12(18):1-101. [doi: 10.3310/JYRT8674] [Medline: 39054745]
44. Idris K, Mubeen Z, Shaikh ZN, Latif A, Hasan S, Khan A. Assessment of ambulance interventions; proposal of a performance measurement framework for healthcare improvement in EMS response to patient collapse. BMC Emerg Med 2025 Apr 12;25(1):56. [doi: 10.1186/s12873-025-01206-y] [Medline: 40221686]
45. Roberts CJ, Barber J, Temkin NR, et al. Clinical outcomes after traumatic brain injury and exposure to extracranial surgery. JAMA Surg 2024 Mar 1;159(3):248-259. [doi: 10.1001/jamasurg.2023.6374] [Medline: 38091011]
46. Gerdin Wärnberg M, Basak D, Berg J, et al. Feasibility of a cluster randomised trial on the effect of trauma life support training: a pilot study in India. BMJ Open 2025 Dec;15(12):e099020. [doi: 10.1136/bmjopen-2025-099020]
47. Lule H, Mugerwa M, Abio A, et al. Motorcycle trauma outcome registry (MOTOR) trial dataset.: Mendeley Data; 2025. URL: https://data.mendeley.com/datasets/bgpmkpcwdt/1 [doi: 10.17632/BGPMKPCWDT.1]

Abbreviations

ATLS: Advanced Trauma Life Support
CONSERVE: CONSORT and SPIRIT Extension for Randomized Controlled Trials Revised in Extenuating Circumstances
CT: Computed Tomography
EMS: Emergency Medical Service
GCS: Glasgow Coma Scale
GOS: Glasgow Outcome Scale
GOS-TBI: Glasgow Outcome Scale for Traumatic Brain Injury
ICC: Intracluster Correlation Coefficient
***ICD-10*:** *International Classification of Diseases, 10th Revision*
KTS II: Kampala Trauma Score II
LMIC: Low- and Middle-Income Country
MOTOR: Motorcycle Trauma Outcome Registry
OR: Odds Ratio
PTC: Primary Trauma Course
REDCap: Research Electronic Data Capture
REML: Restricted Maximum Likelihood
RTTDC: Rural Trauma Team Development Course
TBI: Traumatic Brain Injury
TEFS: Trauma Expectation Factor Score
TOMS: Trauma Outcome Measure Score

Edited by J Sarvestan; submitted 06.Sep.2025; peer-reviewed by E Ogut, N Tuan; revised version received 05.Mar.2026; accepted 06.Mar.2026; published 20.Apr.2026.

Please cite as:
Lule H, Mugerwa M, Abio A, Oguttu B, Kakeeto A, Walsh FJ, Lekuya HM, Ssebuufu R, Kyamanywa P, Deckert A, Bärnighausen T, Posti JP, Wilson ML
Impact of Rural Trauma Team Development Education on Prehospital Time, Referral-to-Dispatch Interval, and Neurological and Musculoskeletal Injury Outcomes: Cluster Randomized Controlled Trial
JMIR Hum Factors 2026;13:e82591
URL: https://humanfactors.jmir.org/2026/1/e82591
doi:10.2196/82591

© Herman Lule, Micheal Mugerwa, Anne Abio, Benson Oguttu, Andrew Kakeeto, Fiona J Walsh, Hervé Monka Lekuya, Robinson Ssebuufu, Patrick Kyamanywa, Andreas Deckert, Till Bärnighausen, Jussi P Posti, Michael Lowery Wilson. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 20.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Quality of Informed Consent and Interface Usability in Primary Care e-Consultation: Cross-Sectional Study

Caitlin Parfitt-Ford[1], BSc; Lisa Ballard[2,3], DHealthPsych; Adriane Chapman[1,3], PhD

[1]School of Electronics and Computer Science, Faculty of Engineering and Physical Sciences, University of Southampton, Southampton, United Kingdom
[2]Primary Care, Population Science, and Medical Education, Faculty of Medicine, University of Southampton, Southampton, England, United Kingdom
[3]Southampton Biomedical Research Centre, National Institute for Health Research, Southampton, England, United Kingdom

Corresponding Author:
Caitlin Parfitt-Ford, BSc
School of Electronics and Computer Science
Faculty of Engineering and Physical Sciences
University of Southampton
University Road
Southampton, SO17 1BJ
United Kingdom
Phone: 44 2380595000
Email: cpf1g20@soton.ac.uk

Related Article:

This is a corrected version. See correction statement: https://humanfactors.jmir.org/2026/1/e103250

Abstract

Background: Patient autonomy through informed consent is a foundational ethical principle for health care practitioners. Online consent processes risk producing "consent in name only," using manipulative or confusing user interfaces to extract consent artificially. This presents a significant danger for safe and ethical remote consultations for primary care providers, which often extract significant amounts of sensitive personal data.

Objective: This study aims to examine the quality of consent obtained through both currently used and novel consent acquisition interfaces for remote e-consultations between a patient and a primary care provider.

Methods: A total of 55 adult participants in the United Kingdom completed an interaction with a mock-up e-consultation system's consent interface for data processing, with 54 completing the full study protocol. The participants were then asked questions regarding what they had provided consent for and the usability of the interface. These responses led to the calculation of an industry-standard System Usability Scale (SUS) score and a novel Quality of Informed Consent Collected Digitally (QuICCDig) score.

Results: Users perceiving interfaces to be more usable (with a greater SUS score) were statistically significantly (n=54; *P*=.004) correlated with an increase in the quality of consent collected from those users (with a higher QuICCDig score). Nonetheless, both existing and novel user interfaces for collecting e-consultation consent were rated poorly, achieving a maximum SUS letter grade of "F." In total, 45% (25/55) of all the participants reported not recalling making a privacy-related decision at all during their consultation, and 87% (48/55) did not recall being offered any alternatives to e-consultation.

Conclusions: The findings demonstrate that current methods for collecting consent in telemedical applications may not be fit for purpose and potentially fail to collect valid informed consent. However, increased usability scores from users do appear to drive improvements in the quality of consent collected. Therefore, decision-makers should place importance on high-quality interface design when building or procuring these systems. We have also provided the QuICCDig score for further use.

(JMIR Hum Factors 2026;13:e78483) doi:10.2196/78483

KEYWORDS

remote consultation; informed consent; electronic consultation; telemedicine; digital triage; privacy; ethics; medical informatics

XSL•FO
RenderX

Introduction

Remote electronic systems for consultations between patients and their primary care providers have been growing rapidly in use since they were first conceptualized—growth that was accelerated by the pressures of the COVID-19 pandemic on the health care system [1]. These systems allow patients to obtain medical advice without having to physically travel to a clinic or speak to a primary care provider over the phone by answering questions and uploading images to an online system for a health care professional to review, thereby expanding access to health care. They allow for rapid symptom triage, which may increase primary care operational efficiency [2]; among some patient populations, they may also lead to increased honesty with health care providers when consulting about symptoms perceived to be embarrassing [3].

Adoption of these systems has varied across international primary care environments. Around the world, asynchronous primary care "e-consultations" conducted via online platforms are known to be used in Norway, Spain, Sweden, the United Kingdom, and the United States [4,5]. In the United Kingdom, access to online consultations from patients to their general practitioner (GP) has been compulsory for GP practices since the April 2021 financial year's contract [6], as part of the National Health Service (NHS) Long Term Plan from 2019 [7]. However, partly owing to unforeseen pressures of the COVID-19 pandemic to implement the technology in a crisis-driven fashion, the success of the rollout has been variable, with some practices seeing substantial efficiency improvements, while others have since reverted to a "strategically traditional" model for meeting the needs of their patient population [8].

Ethical challenges relating to informed consent in telemedicine are nothing new—home monitoring systems have faced them for some time [9]. However, despite the increased uptake of e-consultations [10], issues of safety and consent have not been well evaluated in existing literature [4]. This lack of focus on safety, combined with technocentric policymaking, risks leading clinicians and technologists alike to be overeager to implement new technologies without adequate consideration of the implications for patients' bioethical and legal rights [11]. Regulators and lawmakers internationally have been slow to respond to advances in telemedicine, muddying the waters further [4]—even as disparities in patient understanding of digital health information across age groups and health numeracy have been demonstrated [12].

In the development of websites more broadly, impetus for collecting consent from users for data processing largely comes from regulatory requirements rather than ethical frameworks. Standards for ethical practice for software engineers have been described as "toothless" and criticized for "ethics-washing" without actually changing corporate incentives [13]. Perhaps unsurprisingly, many examples exist of users' data being collected on a "legal minimum" basis or even without proper consent at all [6,7]. Bollinger et al [14] found that, of approximately 30,000 websites studied, 94.7% contained potential European Union General Data Protection Regulation violations related to cookie consent, although that figure is based on automated detection without human verification. In some cases, interfaces are deliberately designed so that users do not remember their choices, even when they regret them upon being reminded of them [9]; however, this has become such standard practice that users simply consider it part and parcel of their internet experience [15].

This resigned acceptance of poor ethical standards in the context of the internet presents a challenge for facilitating ethical telemedical interactions, as the ethical and legal standards for consent in medical contexts are substantially more stringent. Notably, for example, guidance in the United Kingdom consistently highlights the need for clinicians to obtain informed consent from patients for examination and assessment and not just subsequent treatment [16-19]; indeed, government guidance makes it clear that even implied consent for measuring blood pressure by a patient holding out an arm may require the patient "receiving appropriate information" first [20]. The courts have confirmed this; in *R v Hallstrom*, Mr Justice McCullough writes "it goes without saying that, unless clear statutory authority to the contrary exists, no one is […] even to submit himself to a medical examination without his consent" [21]. Failure to collect consent appropriately may be considered a reportable adverse event [22].

There are no currently accepted systems that evaluate clinical informed consent collected through entirely digital systems with no human-in-the-loop involvement. Usability, defined by International Organization for Standardization 9241-11:2018 as the "extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" [23], is measured using well-established metrics such as the Post-Study System Usability Questionnaire or the System Usability Scale (SUS) [24,25]. However, these metrics, when used alone, could, for instance, result in high scores for interfaces that are extremely user-friendly but fail to convey important information, as patients' feelings of being informed in the context of clinical consent may not actually relate to objective measures of knowledge [26]. Therefore, a clear gap exists for a metric to assess the quality of digitally collected informed consent.

This work analyzed the quality of consent that can be collected through remote patient–primary care provider e-consultations. To do this, we (1) developed a scoring system by which consent gathered for electronic patient–primary care provider medical consultations can reliably be evaluated, (2) established an understanding of the validity of current strategies for consent acquisition in remote patient–primary care provider e-consultation via the internet, and (3) evaluated whether changes in user interface design could impact the quality of consent collected.

Methods

Overview

This research developed a novel questionnaire and scoring system, Quality of Informed Consent Collected Digitally (QuICCDig), for evaluating the quality of health care informed

XSL•FO
RenderX

consent collected digitally, based on a synthesis of existing systems from health and technology consent research. A selection of real online consultation systems was then evaluated, using both QuICCDig to evaluate the quality of the consent collected and the industry-standard SUS to evaluate the perceived usability of the interface.

QuICCDig Development

In this study, we synthesized QuICCDig from an analysis of multiple existing criteria for consent evaluation across both health care and digital technology. A narrative literature review was conducted, investigating current methods used for evaluating the quality of consent collection in medical consultations and digital interactions. PubMed, Google Scholar, and HeinOnline were searched for terms including "quality of informed consent," "digital consent evaluation," "medical consent evaluation," "digital consent quality," and "medical consent quality" without restricting to a particular date range, location of publication, or format of publication.

Search results were then assessed for relevance. In particular, results with no full text available to the researchers were excluded as were results whose full text did not contain evaluation methodologies for consent quality or whose methodology was based on parameters such as refusal rate rather than the quality of the consent process itself. Methodologies that were specific to the consent process for a particular subspeciality were excluded, but methodologies used for the evaluation of consent for clinical research were included, as it was hypothesized these might relate to more relevant data-sharing exercises. Where results were already referenced in a broader piece of work, such as a systematic review, the original work was excluded in favor of the broader work to simplify identification of themes.

Interface Evaluation Study Design

To understand how well existing digital consultation systems manage the challenge of acquiring informed consent, we (1) reviewed existing consultation systems approved for use in the United Kingdom to evaluate what consent interfaces they use, (2) created a series of interactive mock-up interfaces (defined as midfidelity interactive prototype user experiences that allow users to complete a specific interaction as relevant for the study) for consent collection in the context of a digital consultation, and (3) provided participants with the interface and subsequently administered the SUS and QuICCDig questionnaires to evaluate the usability of the interface and the quality of the consent collected through it.

All 9 online consultation systems approved for use in the United Kingdom NHS Digital Buying Catalogue [27] as of January 2023 that supported "self-help and signposting" or "symptom checking" were reviewed, of which 7 were noted to be suitable for primary care patient-provider remote consultation use. Only 2 contemporary practices for consent were identified in these 7 systems: either "link and checkbox," in which users are shown a checkbox to confirm their acceptance of privacy information or "summary and checkbox," in which users are shown a checkbox to confirm their acceptance along with a brief explanation of what they are providing consent for and a link to read more. The design of these interfaces reflects different modalities of user interaction as well as different amounts of content presented initially to the user; these differences may reflect attempts to promote user recognition of familiar interaction paths within known user interface paradigms while also allowing appropriate use of users' background knowledge, mapping to the Nielsen design heuristics of "recognition rather than recall" and "match between system and real world" [28].

On the basis of these real consultation systems, 9 mock-up user interfaces were designed, with 2 based on contemporary "link and checkbox" or "summary and checkbox" consent practices and the remainder being novel interfaces not seen to be used in existing approved software (Table 1; for screenshots, refer to Multimedia Appendix 1). Checkbox, drag-and-drop, and swipe interfaces were investigated, as previous research has shown that the quality of consent collected and the perceived usability of the interface may vary across these interface types [29,30].

Participants' demographic data were collected, and they were then presented with a brief to complete a simulated consultation related to a low-acuity mental health presentation. This presentation was chosen because it may result in greater concerns about privacy and confidentiality than a physical health presentation [31,32]. Participants were randomized by computer to interact with 1 of the 9 interfaces, and metadata regarding interactions, such as click heat maps and interaction times, were recorded by the system. The only differences between the interfaces were their consent collection techniques—the other elements were identical. Each participant was only shown the single interface to which they were randomly allocated.

After completing the interaction, participants were presented with a survey evaluating their impressions. First, participants were asked to complete the standard "SUS" questions [24]; they were then presented with the QuICCDig questionnaire described earlier.

Table 1. Mock-up interface reference table according to their features.

	Blanket statement with link to privacy policy	Summary of privacy policy and details page	All policy information contained on a single page
Checkbox	Interface 1	Interface 4	Interface 7
Drag and drop	Interface 2	Interface 5	Interface 8
Swipe	Interface 3	Interface 6	Interface 9

XSL•FO
RenderX

Ethical Considerations

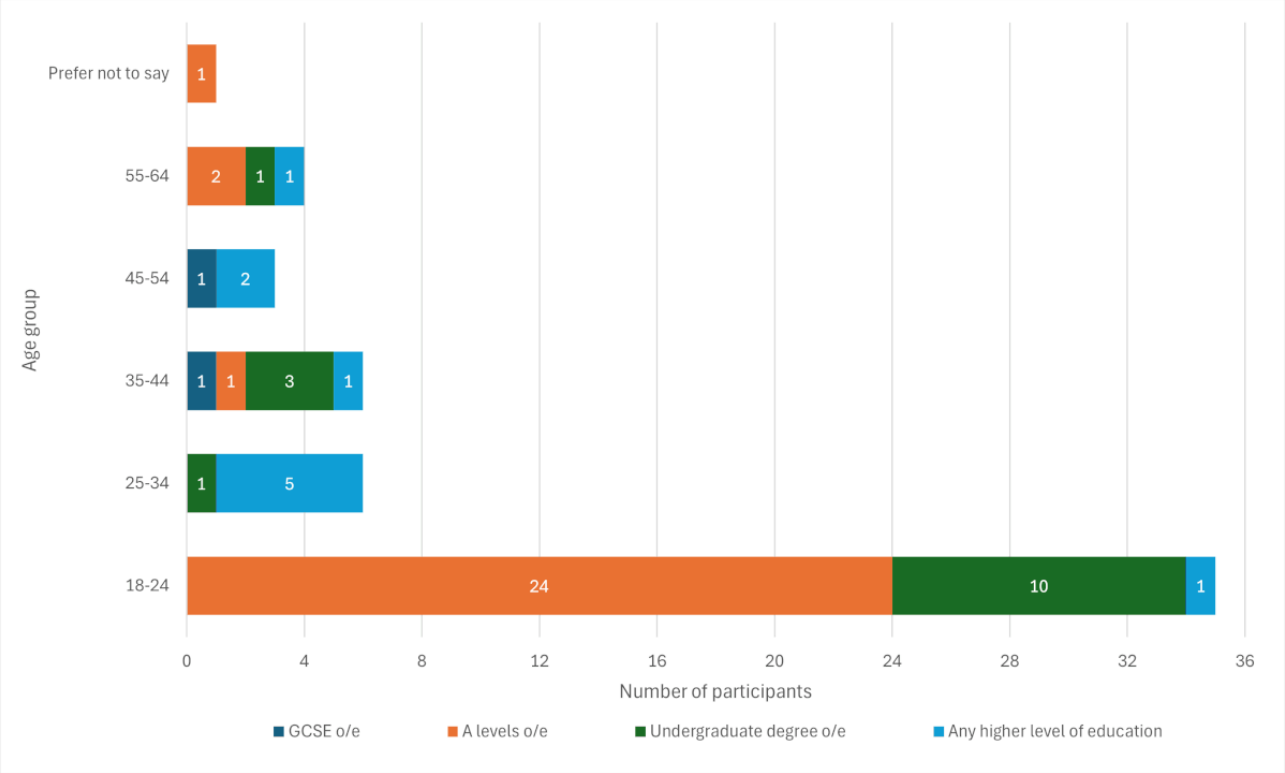
This research was approved by the University of Southampton Faculty of Engineering and Physical Sciences Ethics Committee (FEPS/ERGO/78443). Participants were presented with signposting information for support if the study caused any discomfort or distress and informed of the voluntary, anonymous nature of the study. Informed consent was gained from all participants before their participation, and participants were reminded of their ability to opt-out at any time. Participants were not compensated for their participation. No personal data were used in the analysis.

Recruitment and Statistical Power

Participants aged 18 years or older who had previously interacted with a GP were recruited via snowball sampling from an initial set of posts on social media as well as from advertisements placed on a university campus. A target of 113 participants was set based on a planned linear multiple regression analysis (with an anticipated f^2 value of 0.15, an α value of .05, and a power of 0.8). In actuality, 93 participants were recruited; 66 (71%) completed their interactions enough to answer at least the SUS questions, and 55 (59%) completed the survey well enough to calculate a QuICCDig score. Of these 55 participants, 45% (25/55) had formal computer science-related education, 27% (15/55) had formal medical education, and 76% (42/55) had been prior users of e-consultation technology.

In total, 1 (1%) participant who completed the QuICCDig questionnaire could not have their SUS score calculated due to a technical issue, leaving 54 (58%) of the recruited participants with valid scores for both SUS and QuICCDig. Demographic breakdowns of participants are included in Figure 1.

Figure 1. Demographic breakdowns of participants. GCSE: General Certificate of Secondary Education; o/e: or equivalent.



Results

QuICCDig Synthesis

A total of 11 key papers were identified, and their content was then reviewed for themes. After excluding themes that were either impossible to measure reliably without a human in the loop and were thus impractical to include in an online primary care patient-provider e-consultation system, such as the use of repeat-back testing [33,34], or items that could automatically be collected in a digital system without having to ask a patient (and therefore did not need to be explicitly asked), such as time to complete a consent process or parts of a website that were consulted, 14 key themes were identified as being used in current evaluation criteria. From these themes, corresponding questions were synthesized for inclusion in the QuICCDig metric. The themes included and the questions produced are presented in Table 2, while themes identified but excluded are detailed in Table 3. The scoring metric for these questions is detailed in Multimedia Appendix 2.

XSL•FO
RenderX

Table 2. Quality of Informed Consent Collected Digitally (QuICCDig) themes and questions produced.

Theme	Study	Question
Ability to complete the task	[29,35]	• "Were you able to successfully complete your e-consultation?"
Perceived ease of use	[14,15]	• "How easy did you find it to complete your e-consultation?"

Perceived quality of description and explanation of processes and how they work	[34-37]	• To be asked before the main body of the study, alongside the demographic questions
		  • "What do you think happens during an e-consultation, and where does the information go?"
		• Asked with the rest of the QuICCDig questions after the main body
		  • "Did you get an explanation of what an e-consultation involves?"
		  • "How well do you feel you understand e-consultations now?"
		  • "Did your understanding change from what you thought happened in an e-consultation before? If so, what changed?"

Objective understanding of processes and how they work	[35,38]	• "How long will your consultation data be stored?" • "Which of the below might have access to some data collected during the consultation? [Select multiple: Me, my healthcare provider, other healthcare providers, the company that runs the consultation software, Google, Facebook]."
Why it is being done and the benefits to the patient	[35-38]	• For each group identified in the question about who will have access to consultation data • "Why will NAME have access to your data?"
Risks to the patient inherent to the procedure	[34,36-38]	• "Which of the below are risks involved in the e-consultation process?"
Alternatives	[34,36-38]	• "Do you remember being told about the alternatives to using an e-consultation?" If yes, "What alternatives were there?"
Recollections of consent	[34,35]	• "Do you remember making any privacy-related decisions during your e-consultation?" If yes, "What do you remember deciding?"
Effect on privacy and confidentiality	[37]	• "What impact will this e-consultation have on the confidentiality of your medical records?"
Degree of satisfaction with the decision-making and consent process	[34,37,39]	• "How satisfied with the decision-making process for whether you wanted to continue with an e-consultation were you?"
Whether the patient knew who to ask questions to	[34,37]	• "If you had questions about how your e-consultation would work, or how your data would be used, who would you ask?"
How to revoke consent	[35]	• "If you wanted to change your mind about your e-consultation data being used, what would you do?"
Patient's feeling of involvement with the process	[34,35,39]	• "To what extent did you feel informed and involved with the process of consenting for your e-consultation? [Select one: Not at all, Somewhat, Enough, More than enough, Too much]"
Anything the patient felt was missing	[35]	• "Was there anything else you would have liked to have known that you weren't told about?"

XSL•FO
RenderX

Table 3. Themes identified in the literature but not included in the Quality of Informed Consent Collected Digitally (QuICCDig) scoring system.

Theme	Study	Justification for lack of inclusion
Whether the patient was asked for consent	[34]	In the context of an interaction that is guaranteed to present a prominent consent screen, asking users whether they were prompted for consent would not be expected to yield useful data compared to an in-person medical interaction, where it could be forgotten or made nonobvious.
Timing of information being given	[34,36]	Specific times at which consent information is presented can be automatically collected from an online system but are also less likely to be relevant in the context of an asynchronous patient–primary care provider interaction.
Who gave the information to the patient	[34]	Consent information in online systems, such as primary care patient-provider e-consultation systems, is consistently provided through an automated online interface.
Whether the patient was asked to repeat the explanation	[34]	Assessing repeated explanations for coherence and relevance would be essentially impossible to achieve without the involvement of a clinician, which is not practical for the purposes of a patient-provider e-consultation in primary care.
Questions relating to research consent	[37]	The scope of the QuICCDig score is limited to clinical consent for consultation and assessment.
Questions relating to ignoring consent dialogues and continuing use of the application in any event	[35]	In all the interfaces used and identified, it is not possible for the user to proceed with the consultation without interacting with the consent interface.
Questions relating to specific fine-grained control of (eg, cookies)	[35]	Control of specific aspects of the technical function of the system is outside the scope of the QuICCDig score; rather, the overall consent process is evaluated.
Questions about which parts of the website were consulted	[35]	In this study and in most similar interfaces identified in real-world use in the National Health Service Digital Buying Catalogue, interfaces are presented as single-page applications rather than multipage websites with multiple interaction paths.
Time to complete the consent and consultation process	[29]	This can be captured automatically by online systems.

Interface Evaluation Study

Among participants for whom both a QuICCDig and a SUS score could be calculated, a statistically significant moderate positive correlation was identified (n=54; ρ_{52}=0.388; *P*=.004; 95% CI 0.120-0.604) between a user's calculated SUS score and QuICCDig score (Figure 2). There was no statistically significant difference in the number of study completions between mock-up interfaces (n=93; $F_{8,84}$=1.122; *P*=.36), amount of information presented (n=93; $F_{2,90}$=1.719; *P*=.19), or consent interface type used (n=93; $F_{2,90}$=0.271; *P*=.76). The calculated QuICCDig scores ranged from –0.775 to 0.775 with a mean of 0.371 (SD 0.305); the SUS scores ranged from 13 to 40 with a mean of 32.11 (SD 6.864).

While there were substantial numerical differences in the mean scores between interfaces, as shown in Figure 3, an ANOVA across the different mock-up interfaces for average QuICCDig score (N=55; $F_{8,46}$=0.946; *P*=.49) and average SUS score (n=54, $F_{8,45}$=1.629; *P*=.14) found that these differences were not statistically significant, although the averages did indicate some superiority of the first 2 information levels compared to all information being contained on the same page (Tables 4 and 5). Likewise, no significant difference was found in either SUS or QuICCDig scores based on previous e-consultation use (SUS: n=54; Mann-Whitney *U*=228.0; *P*=.44; QuICCDig: N=55; Mann-Whitney *U*=248.0; *P*=.62). Time elapsed interacting with the interfaces is shown in Table 6 but was not statistically significantly different between experiments (N=55; $F_{8,46}$=0.895; *P*=.53).

Across the board, only 13% (7/55) of the participants recalled being offered alternatives to e-consultation, something presented only in the policy text itself. No correlation was observed between the number of participants who recalled being offered alternatives and the mock-up interface shown. Of the 7 participants who did remember alternatives and stated which ones they remembered, 6 correctly recalled at least 1 alternative option presented by the privacy notice, although only 1 recalled both alternatives (NHS 111 or speaking to the GP by another means), and 1 also produced an alternative which was not included in the notice (NHS 999).

A total of 55% (30/55) of the participants recalled making a privacy-related decision during their e-consultation. Of the 26 participants who responded when asked what they recalled deciding, 12 (46%) referenced the idea of their data being processed; 9 (35%) wrote that they had to agree to privacy information, while 6 (23%) recalled deciding not to read the privacy information. Only 5 (19%) participants gave a specific decision they had to make, for example, "agree [sic] to google analytics," or "if I should use e-consultation at all […] as I was worried about data breaches."

When asked about the impact that the use of e-consultation would have on the confidentiality of their medical records, of the 37 respondents who wrote a response, 3 (8%) gave a response identifying a specific change ("It'll be visible to my medical provider"; "Only the transport risk"; and "The information […] may be accessible to the software company and Google"); 15 (41%) stated they did not believe there to be any change. Some participants (n=8, 22%) also expressed feeling that their data would be generically "more" or "less" secure, while 11 (30%) stated they did not know.

XSL•FO
RenderX

Figure 2. Quality of Informed Consent Collected Digitally (QuICCDig) score against System Usability Scale (SUS) score.

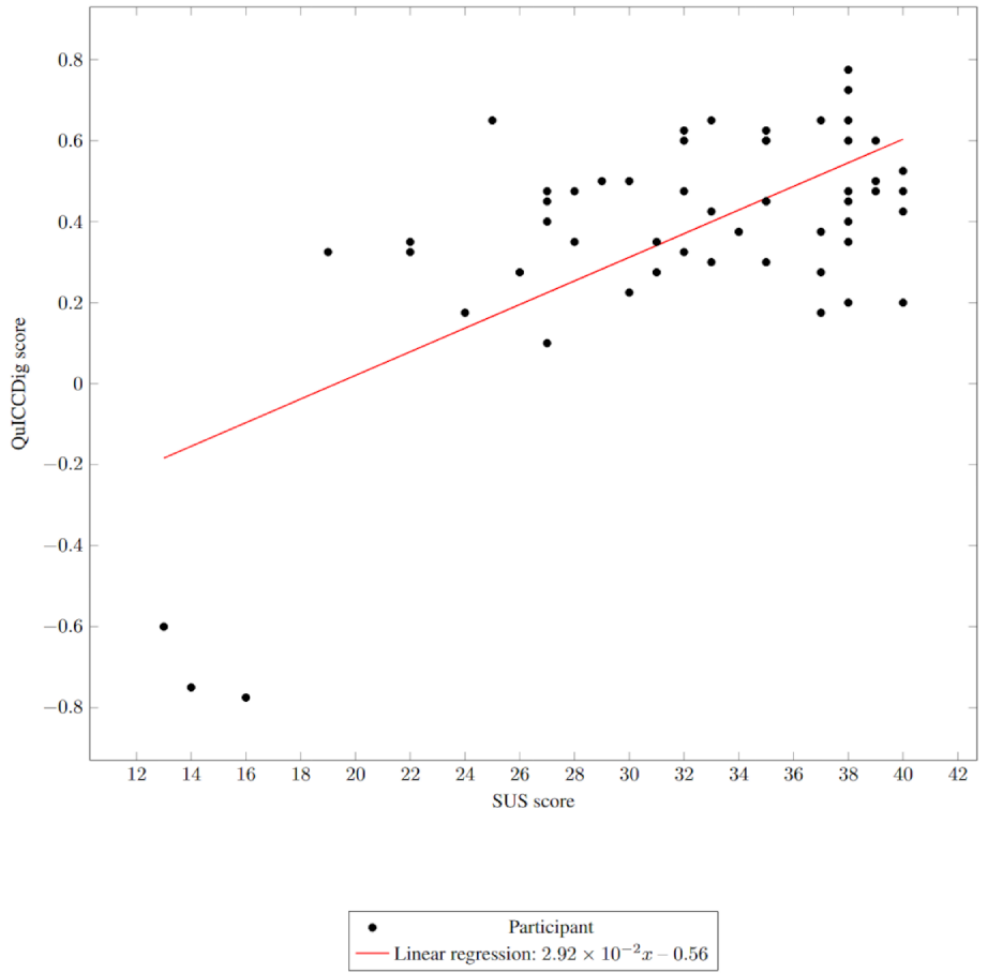
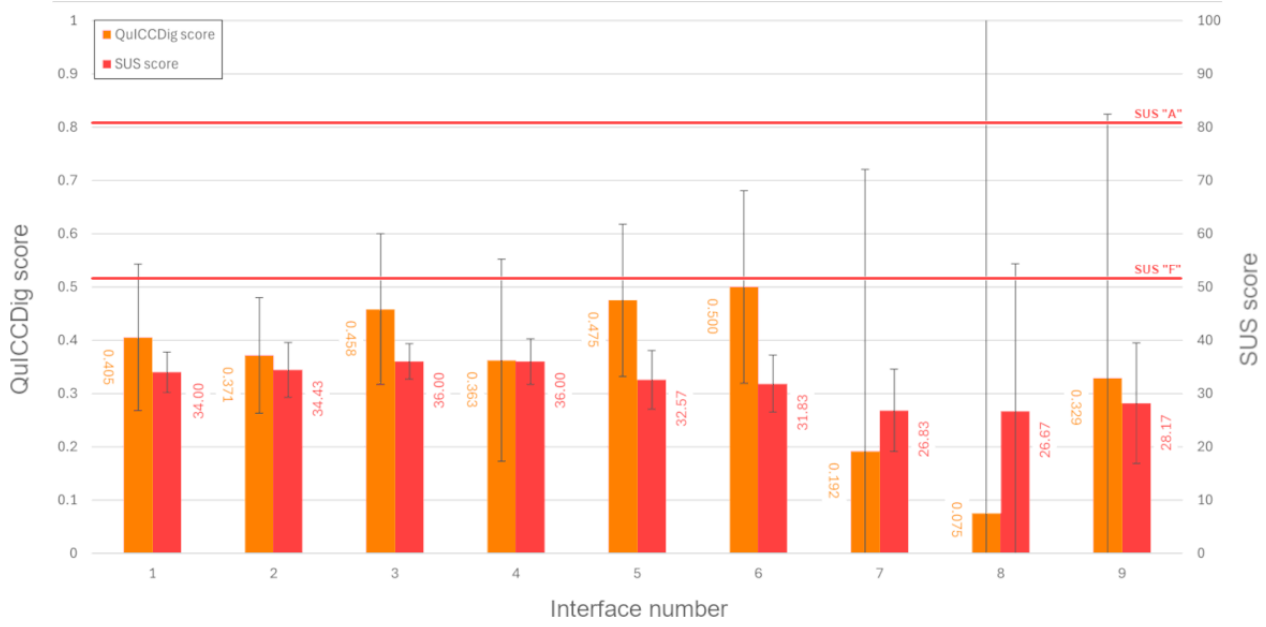


Figure 3. Mean Quality of Informed Consent Collected Digitally (QuICCDig) and System Usability Scale (SUS) scores by interface; 95% CIs for both are displayed in error bars. Grade boundaries for the SUS score are taken from the work of Sauro and Lewis [40].



Notably, 29% (2/7) of the participants from interface 2 (the drag and drop group was presented with a simple privacy policy link) wrote in free text that they would have liked to have what they described as a "simplified" or "short" version of the legal information before providing consent. No participants from any of the other groups made such a comment.

XSL•FO
RenderX

Most of the participants who responded when asked in free text whether their understanding of what an e-consultation was had changed after completing the consultation stated that their understanding did not change (13/23, 57%), while a plurality stated that they thought the mock-up interface was too simple for a full e-consultation (10/23, 43%).

Table 4. Mean System Usability Scale scores per experiment.

	Blanket with privacy policy, mean (SD)	Summary and details page, mean (SD)	All policy information contained on a single page, mean (SD)
Checkbox	Interface 1 34.00 (5.249)	Interface 4 *36.00 (1.732)*[a]	Interface 7 26.84 (7.360)
Drag and drop	Interface 2 34.43 (5.563)	Interface 5 32.57 (5.940)	Interface 8 26.67 (11.150)[b]
Swipe	Interface 3 *36.00 (3.162)*	Interface 6 31.84 (5.115)	Interface 9 28.17 (10.815)

[a]Italics indicate the best result.

[b]The worst result.

Table 5. Mean Quality of Informed Consent Collected Digitally scores per experiment.

	Blanket with privacy policy, mean (SD)	Summary and details page, mean (SD)	All policy information contained on a single page, mean (SD)
Checkbox	Interface 1 0.405 (0.192)	Interface 4 0.363 (0.063)	Interface 7 0.192 (0.504)
Drag and drop	Interface 2 0.371 (0.118)	Interface 5 0.475 (0.155)	Interface 8 0.075 (0.728)[a]
Swipe	Interface 3 0.458 (0.135)	Interface 6 *0.5 (0.172)*[b]	Interface 9 0.329 (0.472)

[a]The worst result.

[b]Italics indicate the best result.

Table 6. Median interaction duration for each mock-up interface.

Interface number	Interaction duration (s), median (IQR)
1	53.3 (35.9-72.4)
2	61.0 (52.7-64.4)
3	61.0 (59.2-62.8)
4	57.9 (32.0-77.2)
5	84.0 (56.1-89.7)
6	69.3 (40.8-79.8)
7	52.0 (39.2-71.1)
8	95.3 (67.7-103.9)
9	86.0 (53.2-97.2)

Discussion

Principal Results

The link found between the quality of consent and perceived interface usability is a significant novel result and strongly supports the importance of high-quality interface design when building remote-access consent collection systems, such as those used for e-consultations. Designers of these systems should consider using user testing with the SUS, QuICCDig, and/or similar scoring systems to evaluate the effectiveness of their designs for gathering informed consent from users. Likewise, organizations responsible for their implementation should consider high-quality user interface design to be a requirement of such systems to effectively gather consent from patients and not merely a “nice-to-have.”

Nonetheless, across all user interfaces tested, both SUS and QuICCDig scores were objectively low. The maximum average SUS score attained by any of the interfaces was 36 for interfaces 3 and 4 (SDs 3.162 and 1.732, respectively), which is between

XSL•FO
RenderX

the fourth and sixth percentile of raw SUS scores and would equate to a letter grade of F [40]. Likewise, the maximum theoretical QuICCDig score is 1, yet the highest attained score by any of the interfaces presented was 0.5. A total of 17% (9/54) of the participants who completed the full study admitted in free text, choosing not to read privacy information thoroughly or at all, even in the context of a study in which participants were informed beforehand of the aim to assess the informedness of consent, which may have had an impact on the SUS scores obtained. An interface required to complete a task that is a prerequisite but ancillary to the user's intended goal in the consultation may therefore always receive low scores. Equally, participants' broader attitudes to privacy and medical consent may have impacted perceived usability or desire to interact with the interface at all. The low number of participants conveying understanding of either positive or negative changes in the confidentiality of their medical records may have arisen for similar reasons. Alternative novel designs may warrant exploration, such as those incorporating voice-over or slideshow interfaces, to examine whether different modalities can deliver improvements.

Fewer than 1 in 7 participants remembered being offered alternatives to completing an e-consultation, even when every design included them in the presented information. This lack of awareness may be particularly concerning for patients less confident in completing their health interactions online, such as older patients, or for patients with what they perceive to be more serious conditions, who may feel less comfortable with online contact with their health care provider [41]. Further research should be conducted targeting these groups specifically. Real-world patients may be aware of existing alternatives anyway, such as out-of-hours services or calling their medical practice by telephone [42]. However, some patients feel pushed into e-consultation use by their primary care provider to avoid alternative methodologies such as face-to-face appointments, which may affect their trust in the service [43]. One study conducted in Norway in 2025 found that 9.5% of the patients who used a primary care e-consultation platform said that sending an e-consultation was not their first choice [42]; clear display of alternative options as part of the process of beginning a consultation may benefit these patients. Similarly, while some patients who are not English speakers may feel more confident completing an e-consultation than a telephone or face-to-face consultation [44], others may prefer to be seen in person [45]; challenges among this patient group with understanding how to contact primary care providers have been specifically highlighted in focus groups conducted by the NHS [46], underscoring the importance of this information being properly conveyed.

Both the drag-and-drop summary and swipe summary interfaces had higher average QuICCDig scores, but lower average SUS scores, than all other interfaces except for the ones containing the entire policy on the consent page. This could potentially indicate a priming effect: perhaps seeing a small amount of information begets interest in more, owing to an increased level of concern, akin to the bulletproof glass effect identified by Brough et al [47]. However, this study's low sample size makes drawing any strong inference from this difficult; future work could investigate this effect in greater detail, perhaps through a more qualitative analysis.

Comparison to Prior Work

Easier-to-use user interfaces appear to be associated with increased quality of consent in this study. This builds on the findings of Lindegren et al [29], demonstrating the effect of consent design patterns on usability and user attention. Habib et al [35] further found that different user interfaces could result in substantially different patterns in recall of consent information, although they did not test the usability of the interfaces they examined.

Research has previously highlighted issues surrounding a lack of reporting on the safety and quality of electronic health care consent collection. Ramos [48], writing in the context of an HIV clinic setting, pointed out that although guidelines clearly assess the accessibility of medicolegal consent information provided in paper form, no standard has existed for evaluating patients' consent following interactions with a digital consent collection user interface. This was once again raised as an important gap in a 2024 systematic review by Leighton et al [4], noting poor reporting on safety— underscoring the fact that in nearly a decade, there has been little progress in this field.

In research medicine, many studies have evaluated the use of technological consent acquisition user interfaces. By virtue of the fact that the end users of these systems are research volunteers, they are likely to have more time and inclination to read consent information than members of the general public accessing health care services, yet even in this group, few definitive findings relating to usability exist [49]. However, usability as measured by SUS scores in research participants appears to be far better than that found in this study [50]. It is possible that participants view providing consent as the primary task in a research consent collection platform, whereas they view it as an ancillary one in a patient–primary care provider consultation interaction, and that this difference is reflected in the reduced SUS scores in this study. This would mirror the findings of Utz et al [51] that some users provide consent for cookies to be placed by websites simply as the easiest path to completing the objective of their visit, obviously with concerning implications for the quality and validity of medical consent.

This study focuses on traditional user interfaces and has not addressed the potential of conversational user interfaces (CUIs) to impact consent. Previous research has shown that users may accept or even prefer information given in this form [52]. CUIs can evoke feelings of social presence [53], and while this has been shown to feel intrusive in the e-commerce context, personal connection to clinicians during consent conversations has been shown to reduce anxiety and improve clinical outcomes [54]. Some research suggests that "chatbot" interfaces that appear overtly as a bot may produce equal perceptions of expertise as those that present themselves as a physician [55]. However, unless a bank of preset responses is used, presenting privacy information this way risks eventual "inevitable" errors [56], potentiating invalid consent based on invalid information. Further research is required in this area to evaluate the safety

XSL•FO
RenderX

and efficacy of modern CUI systems in privacy-related decision-making interfaces.

Limitations

The interpretation of these results should be tempered by the nature and size of the sample and the snowball method of acquisition. While the positive findings are still significant, the negative results may have limited generalizability, owing to the underpowering of the study—the lower-than-targeted sample size may, for instance, have prevented the detection of actual differences in the SUS or QuICCDig scores between the interface types. The sample achieved is also heavily biased toward young, well-educated participants, including those with health care– or computer science–related education. Well-educated participants appear to consistently have higher levels of comprehension and better quality of consent as a consequence [57,58]; education may also have a more direct influence on privacy-protective behavior [59]. Results have varied internationally with respect to the effect that age has on measures of consent quality and recall, although older patients, underrepresented in this study, are known to have higher rates of neurocognitive impairment than younger patients, which presents unique issues for ensuring informed consent is collected effectively wherever possible [60]. Furthermore, some evidence suggests older people may have greater privacy concerns on average than younger people [61].

The variability of health systems' uptake of asynchronous e-consultation systems internationally and the apparent heterogeneity of different populations' digital health literacy [62] may also limit this study's relevance outside of the UK NHS context. The effect of using a simulated clinical interaction in which participants were aware that consent and privacy were being studied, as opposed to using real-world patients, is also unknown and limits the direct applicability of this study to clinical practice. With rigorous ethical safeguards in place, future trials conducted directly in primary care settings with larger, more representative populations should consider using real patient interactions to evaluate these systems and explore in more detail the potential for an intention-behavior gap in these interactions with respect to informed consent.

The QuICCDig scoring system has not been formally externally reviewed, and this is not a formal validation study. There is currently no clear gold-standard metric in the field to validate against, which makes designing a robust validation study challenging. Future research could consider validating against clinician-perceived informedness, although it is important to be mindful of the biases this could introduce, or perhaps simultaneously against validated scores for digital consent, clinical consent, and/or health care research consent. However, this may be difficult to orchestrate in practice.

Conclusions

We find that contemporary user interfaces for collecting informed consent in remote electronic consultations for primary care may not reliably collect fully informed consent from patients. Furthermore, we find the perceived usability of a user interface to be significantly (P=.004) correlated with the quality of consent acquired through that interface. However, even in a study where participants knew consent was being studied, we find that almost half (25/55, 45%) of the participants did not remember making a privacy decision, and more than 87% (48/55) of the participants did not remember being offered alternatives to agreeing to complete their consultation. We recommend that primary care decision-makers consider the quality and usability of a remote consultation and triage application's consent interface when making procurement decisions.

Acknowledgments

The authors thank Dr Reinder Broekstra of the Faculty of Medical Sciences at the University of Groningen in the Netherlands for his input concerning medicolegal aspects of digital consent and Dr Laura Carmichael of the School of Electronics and Computer Science at the University of Southampton for her input in the field of data protection and privacy law.

This study makes use of the System Usability Scale. The System Usability Scale was developed as part of the usability engineering program in integrated office systems development at Digital Equipment Co Ltd, Reading, United Kingdom.

This work is partially supported by the National Institute for Health Research Southampton Biomedical Research Centre.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the 9 user interfaces users were presented with, demonstrating the levels of privacy information presented and the methodologies used for the consent process in each.
[PDF File (Adobe PDF File), 1367 KB - humanfactors_v13i1e78483_app1.pdf]

XSL•FO
RenderX

Multimedia Appendix 2
Quality of Informed Consent Collected Digitally questionnaire and calculation process.
[PDF File (Adobe PDF File), 94 KB - humanfactors_v13i1e78483_app2.pdf]

References

1. Clarke G, Wolters A, Dias A. Access to and delivery of general practice services: a study of patients at practices using digital and online tools. The Health Foundation. 2022. URL: https://www.health.org.uk/reports-and-analysis/briefings/access-to-and-delivery-of-general-practice-services [accessed 2022-10-10]
2. Gottliebsen K, Petersson G. Limited evidence of benefits of patient operated intelligent primary care triage tools: findings of a literature review. BMJ Health Care Inform 2020 May;27(1):e100114 [FREE Full text] [doi: 10.1136/bmjhci-2019-100114] [Medline: 32385041]
3. Wise J. Electronic consultations offer few benefits for GP practices, says study. BMJ 2017 Nov 07:j5141. [doi: 10.1136/bmj.j5141]
4. Leighton C, Cooper A, Porter A, Edwards A, Joseph-Williams N. Effectiveness and safety of asynchronous telemedicine consultations in general practice: a systematic review. BJGP Open 2024 Apr 25;8(1):BJGPO.2023.0177 [FREE Full text] [doi: 10.3399/BJGPO.2023.0177] [Medline: 37783479]
5. Glock H, Jakobsson U, Borgström Bolmsjö B, Milos Nymberg V, Wolff M, Calling S. eVisits to primary care and subsequent health care contacts: a register-based study. BMC Prim Care 2024 Aug 12;25(1):297. [doi: 10.1186/s12875-024-02541-y] [Medline: 39135150]
6. Kanani N, Vautrey R, Kettell R. Supporting general practice in 2021/22. National Health Service England. 2021. URL: https://www.england.nhs.uk/wp-content/uploads/2021/01/C1054-supporting-general-practice-in-21-22.pdf [accessed 2025-09-28]
7. The NHS long term plan. National Health Service England. 2019 Jan 7. URL: https://webarchive.nationalarchives.gov.uk/ukgwa/20230418155402/https:/www.longtermplan.nhs.uk/publication/nhs-long-term-plan/ [accessed 2025-09-28]
8. Greenhalgh T, Clarke A, Byng R, Dakin F, Faulkner S, Hemmings N, et al. After the disruptive innovation: how remote and digital services were embedded, blended and abandoned in UK general practice - longitudinal study. Health Soc Care Deliv Res 2025 Jun;13(31):1-37. [doi: 10.3310/KRWS4334] [Medline: 40515520]
9. Demiris G, Oliver DP, Courtney KL. Ethical considerations for the utilization of tele-health technologies in home and hospice care by the nursing profession. Nurs Adm Q 2006;30(1):56-66. [doi: 10.1097/00006216-200601000-00009] [Medline: 16449885]
10. Solimini R, Busardò FP, Gibelli F, Sirignano A, Ricci G. Ethical and legal challenges of telemedicine in the era of the COVID-19 pandemic. Medicina (Kaunas) 2021 Nov 30;57(12):1314 [FREE Full text] [doi: 10.3390/medicina57121314] [Medline: 34946259]
11. Kaplan B. Ethics, guidelines, standards, and policy: telemedicine, COVID-19, and broadening the ethical scope. Camb Q Healthc Ethics 2022 Jan;31(1):105-118. [doi: 10.1017/S0963180121000852] [Medline: 35049453]
12. Andreadis K, Buderer N, Langford AT. Patients' understanding of health information in online medical records and patient portals: analysis of the 2022 Health Information National Trends Survey. J Med Internet Res 2025 May 30;27:e62696 [FREE Full text] [doi: 10.2196/62696] [Medline: 40446288]
13. Green B. The contestation of tech ethics: a sociotechnical approach to technology ethics in practice. J Soc Comput 2021 Sep;2(3):209-225. [doi: 10.23919/JSC.2021.0018]
14. Bollinger D, Kubicek K, Cotrini C, Basin D. Automating cookie consent and GDPR violation detection. In: Proceedings of the 31st USENIX Security Symposium. 2022 Presented at: USENIX Security '22; August 10-12, 2022; Boston, MA. [doi: 10.3929/ethz-b-000525815]
15. Bhoot AM, Shinde MA, Mishra WP. Towards the identification of dark patterns: an analysis based on end-user reactions. In: Proceedings of the 11th Indian Conference on Human-Computer Interaction. 2020 Presented at: IndiaHCI '20; November 5-8, 2020; Online. [doi: 10.1145/3429290.3429293]
16. Consent and refusal by adults with decision-making capacity: a toolkit for doctors. British Medical Association. 2019. URL: https://www.bma.org.uk/media/txrnpo3s/consent-and-refusal-by-adults-with-decision-making-capacity-guidance-updated-2025.pdf [accessed 2022-10-05]
17. Guidance for providers on meeting the regulations. Care Quality Commission. 2015. URL: https://www.cqc.org.uk/sites/default/files/20150324_guidance_providers_meeting_regulations_01.pdf [accessed 2022-10-05]
18. Arslan R. A review on ethical issues and rules in psychological assessment. J Fam Couns Educ 2018 Jun 01;3(1):17-29. [doi: 10.32568/jfce.310629]
19. COVID-19 Recovery Toolkit Tool 4: virtual consultations. Royal College of Surgeons of England. 2020 Jun. URL: https://www.rcseng.ac.uk/coronavirus/recovery-of-surgical-services/tool-4/ [accessed 2022-10-10]
20. Reference guide to consent for examination or treatment (second edition). United Kingdom Government Department of Health. 2009. URL: https://www.gov.uk/government/publications/reference-guide-to-consent-for-examination-or-treatment-second-edition [accessed 2022-11-09]

XSL•FO
RenderX

21. R v Hallstrom ex p W (No 2), R v Gardner ex p L [1986] QB 1090. In: All England Law Reports vol 2. London: Butterworth & Co; 1986:306-320.
22. Fonseca GM, Belmar-Durán M, Matamala-Santander C. Failure to obtain informed consent should also be considered an adverse event. J Dent Sci 2020 Jun;15(2):232-233 [FREE Full text] [doi: 10.1016/j.jds.2019.11.006] [Medline: 32595908]
23. Ergonomics of human-system interaction part 11: usability: definitions and concepts. International Organization for Standardization. 2018. URL: https://www.iso.org/standard/63500.html [accessed 2026-01-22]
24. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation in Industry. Boca Raton, FL: CRC Press; 1996.
25. Hajesmaeel-Gohari S, Khordastan F, Fatehi F, Samzadeh H, Bahaadinbeigy K. The most used questionnaires for evaluating satisfaction, usability, acceptance, and quality outcomes of mobile health. BMC Med Inform Decis Mak 2022 Jan 27;22(1):22 [FREE Full text] [doi: 10.1186/s12911-022-01764-2] [Medline: 35081953]
26. Sepucha KR, Fagerlin A, Couper MP, Levin CA, Singer E, Zikmund-Fisher BJ. How does feeling informed relate to being informed? The DECISIONS survey. Med Decis Making 2010;30(5 Suppl):77S-84S. [doi: 10.1177/0272989X10379647] [Medline: 20881156]
27. Catalogue solutions. NHS Digital Buying Catalogue. URL: https://buyingcatalogue.digital.nhs.uk/catalogue-solutions?selectedCapabilityIds=43&selectedEpicIds=E00001.E00019 [accessed 2023-01-30]
28. Nielsen J. Enhancing the explanatory power of usability heuristics. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1994 Presented at: CHI '94; April 24-28, 1994; Boston, MA. [doi: 10.1145/191666.191729]
29. Lindegren D, Karegar F, Kane B, Pettersson JS. An evaluation of three designs to engage users when providing their consent on smartphones. Behav Inf Technol 2019 Dec 17;40(4):398-414. [doi: 10.1080/0144929x.2019.1697898]
30. Singh AK, Upadhyaya N, Seth A, Hu X, Sastry N, Mondal M. What cookie consent notices do users prefer: a study in the wild. In: Proceedings of the 2022 European Symposium on Usable Security. 2022 Presented at: EuroUSEC '22; September 29-30, 2022; Karlsruhe, Germany p. 28-39. [doi: 10.1145/3549015.3555675]
31. Rapaport J, Bellringer S, Pinfold V, Huxley P. Carers and confidentiality in mental health care: considering the role of the carer's assessment: a study of service users', carers' and practitioners' views. Health Soc Care Community 2006 Jul;14(4):357-365. [doi: 10.1111/j.1365-2524.2006.00631.x] [Medline: 16787487]
32. Mork MJ, Price S, Best K. Privacy versus care--the shifting balance in mental health. Fam Syst Health 2016 Mar;34(1):56-57. [doi: 10.1037/fsh0000177] [Medline: 26963778]
33. Fink AS, Prochazka AV, Henderson WG, Bartenfeld D, Nyirenda C, Webb A, et al. Enhancement of surgical informed consent by addition of repeat back: a multicenter, randomized controlled clinical trial. Ann Surg 2010 Jul;252(1):27-36. [doi: 10.1097/SLA.0b013e3181e3ec61] [Medline: 20562609]
34. Brezis M, Israel S, Weinstein-Birenshtock A, Pogoda P, Sharon A, Tauber R. Quality of informed consent for invasive procedures. Int J Qual Health Care 2008 Oct;20(5):352-357. [doi: 10.1093/intqhc/mzn025] [Medline: 18625699]
35. Habib H, Li M, Young E, Cranor LF. “Okay, whatever”: an evaluation of cookie consent interfaces. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 Presented at: CHI '22; April 29-May 5, 2022; New Orleans, LA. [doi: 10.1145/3491102.3501985]
36. Spatz ES, Bao H, Herrin J, Desai V, Ramanan S, Lines L, et al. Quality of informed consent documents among US. hospitals: a cross-sectional study. BMJ Open 2020 May 19;10(5):e033299 [FREE Full text] [doi: 10.1136/bmjopen-2019-033299] [Medline: 32434934]
37. Joffe S, Cook EF, Cleary PD, Clark JW, Weeks JC. Quality of informed consent: a new measure of understanding among research subjects. J Natl Cancer Inst 2001 Jan 17;93(2):139-147. [doi: 10.1093/jnci/93.2.139] [Medline: 11208884]
38. Lühnen J, Mühlhauser I, Steckelberg A. The quality of informed consent forms-a systematic review and critical analysis. Dtsch Arztebl Int 2018 Jun 01;115(22):377-383 [FREE Full text] [doi: 10.3238/arztebl.2018.0377] [Medline: 29932049]
39. Gröndahl L. Public knowledge of digital cookies: exploring the design of cookie consent forms. KTH Royal Institute of Technology. 2020. URL: https://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A1470723&dswid=-1949 [accessed 2022-11-12]
40. Sauro J, Lewis JR. Standardized usability questionnaires. In: Sauro J, Lewis JR, editors. Quantifying the User Experience: Practical Statistics for User Research. Burlington, MA: Morgan Kaufmann; 2012:185-240.
41. Mold F, Hendy J, Lai YL, de Lusignan S. Electronic consultation in primary care between providers and patients: systematic review. JMIR Med Inform 2019 Dec 03;7(4):e13042 [FREE Full text] [doi: 10.2196/13042] [Medline: 31793888]
42. Kristiansen E, Atherton H, Austad B, Bergmo TS, Norberg BL, Salisbury C, et al. Patients' use of e-consultations as an alternative to other general practitioner services: cross-sectional survey study. J Med Internet Res 2025 Jan 08;27:e55158 [FREE Full text] [doi: 10.2196/55158] [Medline: 39778196]
43. Khan N, Pitchforth E, Winder R, Abel G, Clark CE, Cockcroft E, et al. What helps patients access web-based services in primary care? Free-text analysis of patient responses to the Di-Facto questionnaire. BMC Prim Care 2024 Jan 10;25(1):20 [FREE Full text] [doi: 10.1186/s12875-023-02257-5] [Medline: 38200431]
44. Ge X, Chappell P, Ledger J, Bakhai M, Clarke GM. The use of online consultation systems and patient experience of primary care: cross-sectional analysis using the general practice patient survey. J Med Internet Res 2024 Jul 26;26:e51272 [FREE Full text] [doi: 10.2196/51272] [Medline: 39058543]

XSL•FO
RenderX

45. Leung K. Use of an electronic consultation system in an inner city general practice: a mixed-methods service evaluation. BMJ Open Qual 2025 Aug 04;14(3):e002741 [FREE Full text] [doi: 10.1136/bmjoq-2024-002741] [Medline: 40759527]
46. NHS South, Central and West Commissioning Support Unit, NHS England. Community languages translation and interpreting services. Support and Transformation for Health and Care. 2024 Dec. URL: https://tinyurl.com/77jfktuv [accessed 2025-09-26]
47. Brough AR, Norton DA, Sciarappa SL, John LK. The bulletproof glass effect: unintended consequences of privacy notices. J Mark Res 2022;59(4):739-754. [doi: 10.1177/00222437211069093]
48. Ramos SR. User-centered design, experience, and usability of an electronic consent user interface to facilitate informed decision-making in an HIV clinic. Comput Inform Nurs 2017 Nov;35(11):556-564 [FREE Full text] [doi: 10.1097/CIN.0000000000000356] [Medline: 28481754]
49. De Sutter E, Zaçe D, Boccia S, Di Pietro ML, Geerts D, Borry P, et al. Implementation of electronic informed consent in biomedical research and stakeholders' perspectives: systematic review. J Med Internet Res 2020 Oct 08;22(10):e19129 [FREE Full text] [doi: 10.2196/19129] [Medline: 33030440]
50. Robins D, Brody R, Jeong IC, Parvanova I, Liu J, Finkelstein J. Towards a highly usable, mobile electronic platform for patient recruitment and consent management. Stud Health Technol Inform 2020 Jun 16;270:1066-1070. [doi: 10.3233/SHTI200325] [Medline: 32570545]
51. Utz C, Degeling M, Fahl S, Schaub F, Holz T. (Un)informed consent: studying GDPR consent notices in the field. In: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. 2019 Presented at: CCS '19; November 11-15, 2019; London, United Kingdom. [doi: 10.1145/3319535.3354212]
52. Brüggemeier B, Lalone P. Perceptions and reactions to conversational privacy initiated by a conversational user interface. Comput Speech Lang 2022 Jan;71:101269. [doi: 10.1016/j.csl.2021.101269]
53. Sohn S. Can conversational user interfaces be harmful? The undesirable effects on privacy concern. In: Proceedings of the ICIS 2019. 2019 Presented at: ICIS '19; December 15-18, 2019; Munich, Germany URL: https://aisel.aisnet.org/icis2019/cyber_security_privacy_ethics_IS/cyber_security_privacy/29/
54. Anandaiah A, Rock L. Twelve tips for teaching the informed consent conversation. Med Teach 2019 Apr;41(4):465-470. [doi: 10.1080/0142159X.2018.1426844] [Medline: 29361861]
55. Jin E, Eastin M. Towards more trusted virtual physicians: the combinative effects of healthcare chatbot design cues and threat perception on health information trust. Behav Inf Technol 2024 May 02;44(4):829-842. [doi: 10.1080/0144929x.2024.2347951]
56. Harkous H, Fawaz K, Shin KG, Aberer K. PriBots: conversational privacy with chatbots. In: Proceedings of the Twelfth Symposium on Usable Privacy and Security. 2016 Presented at: SOUPS '16; June 22-24, 2016; Denver, CO URL: https://infoscience.epfl.ch/server/api/core/bitstreams/a144b30e-a202-47a8-9772-5dbf472d8f7d/content
57. Fons-Martinez J, Murciano-Gamborino C, Ferrer-Albero C, Vergara-Hernandez C, Diez-Domingo J. Digital informed consent/assent in clinical trials among pregnant women, minors, and adults: multicountry cross-sectional evaluation of comprehension and satisfaction. JMIR Hum Factors 2025 Aug 15;12:e65569 [FREE Full text] [doi: 10.2196/65569] [Medline: 40815855]
58. Nnabugwu II, Ugwumba FO, Udeh EI, Anyimba SK, Ozoemena OF. Informed consent for clinical treatment in low-income setting: evaluating the relationship between satisfying consent and extent of recall of consent information. BMC Med Ethics 2017 Dec 02;18(1):69 [FREE Full text] [doi: 10.1186/s12910-017-0227-4] [Medline: 29197378]
59. Boerman SC, Kruikemeier S, Zuiderveen Borgesius FJ. Exploring motivations for online privacy protection behavior: insights from panel data. Commun Res 2018 Oct 05;48(7):953-977. [doi: 10.1177/0093650218800915]
60. Ng IK. Informed consent in clinical practice: old problems, new challenges. J R Coll Physicians Edinb 2024 Jun;54(2):153-158. [doi: 10.1177/14782715241247087] [Medline: 38616290]
61. Prince C, Omrani N, Maalaoui A, Dabic M, Kraus S. Are we living in surveillance societies and is privacy an illusion? An empirical study on privacy literacy and privacy concerns. IEEE Trans Eng Manage 2023 Oct;70(10):3553-3570. [doi: 10.1109/tem.2021.3092702]
62. Zeeb H, Maaß J, Schultz T, Haug U, Pigeot I, Schüz B. Digital Public Health: Interdisciplinary Perspectives. Cham, Switzerland: Springer; 2025.

Abbreviations

CUI: conversational user interface
GP: general practitioner
NHS: National Health Service
QuICCDig: Quality of Informed Consent Collected Digitally
SUS: System Usability Scale

XSL•FO
RenderX

Edited by A Kushniruk; submitted 03.Jun.2025; peer-reviewed by O Akinsola, M Beltrão, R Marshall, M Al Zoubi, D Agbo; comments to author 10.Sep.2025; revised version received 15.Dec.2025; accepted 02.Jan.2026; published 09.Feb.2026.

Please cite as:
Parfitt-Ford C, Ballard L, Chapman A
Quality of Informed Consent and Interface Usability in Primary Care e-Consultation: Cross-Sectional Study
JMIR Hum Factors 2026;13:e78483
URL: https://humanfactors.jmir.org/2026/1/e78483
doi:10.2196/78483
PMID:

©Caitlin Parfitt-Ford, Lisa Ballard, Adriane Chapman. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 09.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Grip Strength Estimation Using Input Data From a Commodity Smartphone: Model Development and Validation Study

Komei Tajima[1], BEng; Kaori Ikematsu[2], PhD; Toshiya Isomoto[2], PhD; Kunihiro Kato[3], PhD; Yuta Sugiura[1], PhD

[1]Department of Information and Computer Science, Faculty of Science and Technology, Keio University, Yokohama, Kanagawa, Yokohama, Japan

[2]LY Corporation, Tokyo, Japan

[3]Japan Women's University, Tokyo, Japan

Corresponding Author:
Yuta Sugiura, PhD
Department of Information and Computer Science
Faculty of Science and Technology
Keio University
3-14-1, Hiyoshi, Kohoku-ku
Yokohama, Kanagawa, Yokohama, 223-8522
Japan
Phone: 81 45 566 1769 ext 48289
Email: sugiura@keio.jp

Abstract

Background: Grip strength is a crucial indicator of muscle deterioration, recovery, sarcopenia, and neurological disorders. However, conventional measurement requires a dedicated dynamometer, which limits accessibility and requires specific movements.

Objective: This study aimed to propose and validate a method for estimating grip strength using standard smartphone operations, thereby eliminating the need for specialized equipment.

Methods: Data were collected from 21 young adults in the main experiment, who performed standard smartphone tasks (tapping, flicking, and dragging) after measuring their grip strength with a dynamometer. A predictive regression model was developed using touch and inertial sensor data. The model was first evaluated using a random split of the entire dataset (random split evaluation). To further assess practical feasibility and generalizability, we conducted leave-one-user-out validation and a few-day calibration validation; the latter simulated a calibration scenario by incorporating 1 to 4 days of user-specific data into the training set.

Results: The regression analysis using a random split of the dataset demonstrated high accuracy, with a mean absolute error of 2.62 (SD 0.18) kg, a mean absolute percentage error (MAPE) of 8.91% (SD 0.57%), and a coefficient of determination of 0.802 (SD 0.036). In the validation of practical scenarios, the leave-one-user-out validation resulted in a MAPE of 15.08% (SD 5.40%). However, the personalized few-day calibration model showed significant improvements as calibration days increased, with the MAPE decreasing to 13.96% (SD 5.57%) after 1 day and reaching 11.64% (SD 5.80%) after 4 days. Furthermore, the National Aeronautics and Space Administration Task Load Index assessment indicated a low overall subjective workload (mean 3.04, SD 2.23 on a scale of 10), confirming the method's suitability for daily use without a significant burden on users.

Conclusions: The proposed method demonstrates that smartphones can serve as a viable, pervasive tool for daily grip strength monitoring, offering a convenient alternative to traditional dynamometers.

(JMIR Hum Factors 2026;13:e90316) doi:10.2196/90316

KEYWORDS

grip strength; mobile health; smartphone; machine learning; human computer interaction

Introduction

Muscle strength is known to decline with aging [1,2], and this decline is used as a diagnostic indicator of sarcopenia and frailty [3,4]. Among various muscle strength measures, grip strength is widely used in medical settings for treatment and rehabilitation [5-7]. Additionally, grip strength has been linked to chronic diseases, with several studies reporting a decline in grip strength among patients with diabetes [8-11]. Furthermore, reduced grip strength has been associated with an increased risk

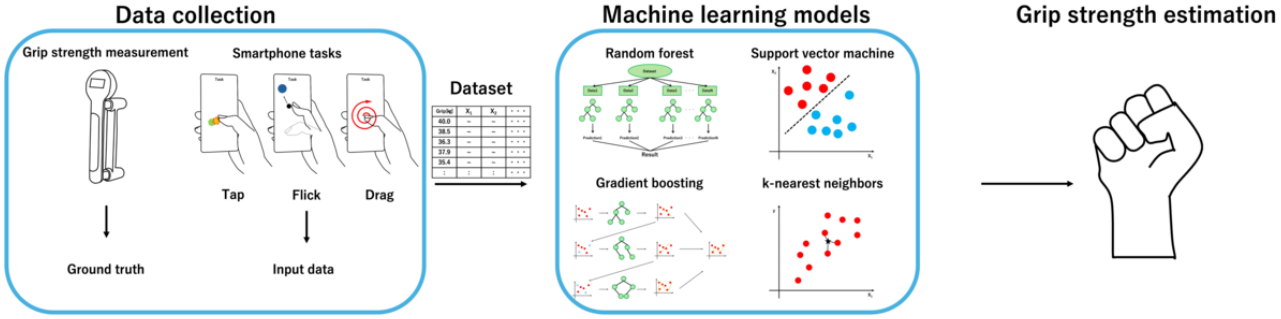
XSL•FO
RenderX

of falls and fractures, highlighting its relevance in assessing overall musculoskeletal health [12]. Therefore, regularly measuring grip strength and monitoring daily variations are essential for health management and disease risk assessment.

However, grip dynamometers are not widely available to the general public, and high-precision devices used in medical settings are expensive, making them difficult to obtain for personal use. Furthermore, electronic grip dynamometers require maintenance and are not easily portable, making measurements outside clinical settings challenging. These limitations hinder the feasibility of continuous and routine monitoring, potentially posing a barrier to effective health management.

We propose a method for estimating grip strength that leverages the widespread availability of commodity smartphones (Figure 1). Specifically, it uses simple and quick touch input tasks to provide practical and accessible grip strength measurements. For model construction, grip strength is first measured using a Jamar-style dynamometer to obtain the ground truth, followed by a smartphone task to collect the corresponding input data. This task involves common smartphone input gestures, such as tapping, flicking, and dragging, during which input coordinates, time stamps, contact area, and sensor data—including acceleration, angular velocity, and orientation—are recorded as feature variables. A regression model is then constructed using the measured grip strength as the target variable and the extracted features as explanatory variables. In total, 21 participants took part in the experiment, and the analysis results showed that the best-performing model had a mean absolute error (MAE) of 2.62 (SD 0.18) kg, a mean absolute percentage error (MAPE) of 8.91% (SD 0.57%), and a coefficient of determination of 0.802 (SD 0.036).

Figure 1. Overview of the proposed grip strength estimation method. Grip strength measurements and smartphone tasks were conducted, and the resulting data were used to construct a machine learning model. Grip strength values were set as the ground truth, while input coordinates, time, and sensor data were used as the input, enabling estimation through regression analysis.



This method enables grip strength estimation using a regular smartphone without the need for specialized equipment, making daily measurements more accessible, enabling use in unsupervised environments, and requiring only a short amount of time (approximately 30 seconds per set of smartphone tasks). A brief overview of the study, including the experimental tasks and key results, is provided in Multimedia Appendix 1.

Methods

Recruitment

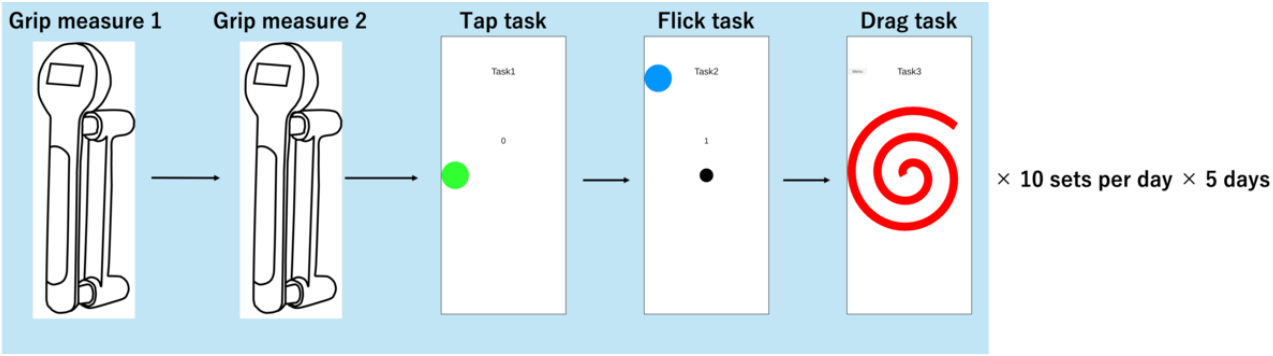
This study included 21 participants (n=16, 76.2% male and n=5, 23.8% female participants), aged 20 to 24 (mean 21.5, SD 1.00) years, with no functional impairments in their hands. All participants were right-handed. We provided participants with a prior explanation of the study and obtained their informed consent before participation.

Procedure

Figure 2 shows an overview of the data collection experiment. Each day, participants performed a set of grip strength measurements and smartphone tasks (tap, flick, and drag tasks) 10 times. There were no breaks between the grip strength measurements and tasks, ensuring continuous execution; this design was deliberately implemented based on previous findings [13] that report a decrease in grip strength immediately after using a handgrip dynamometer, thereby intentionally inducing variability in grip strength over the 10 consecutive task sets. The measurements were conducted over 5 separate days, with at least a 2-day interval between each measurement day. The experimental procedures were standardized across all participants and measurement days.

XSL•FO
RenderX

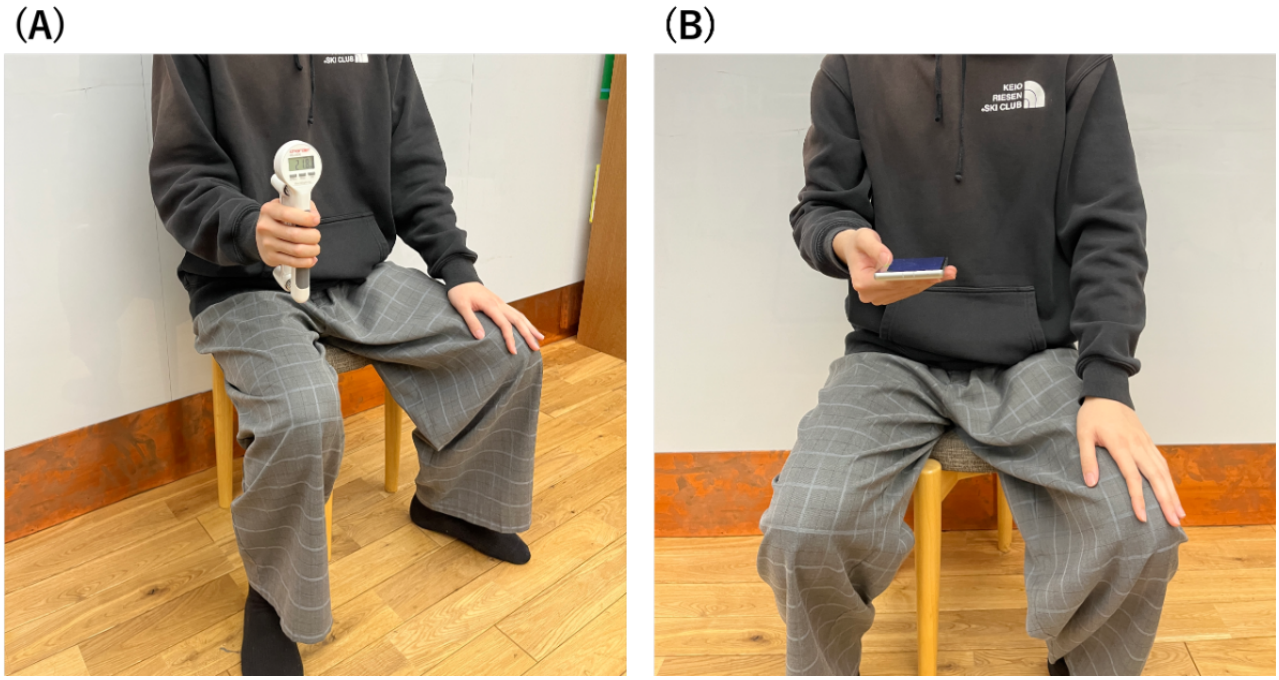
Figure 2. Overview of the data collection experiment. The experiment consisted of grip strength measurements and smartphone tasks (tap, flick, and drag).



Grip Strength Measurement

We used a Jamar-style dynamometer with a standard procedure to measure grip strength. Participants sat on a stable chair without armrests in a relaxed state during the measurement. During the measurement, they kept their upper limb on the measurement side aligned with their body and maintained a posture with the elbow joint flexed at 90° (Figure 3A). The measurement was conducted only on the dominant hand. After the first measurement, a second measurement was taken following a 10-second interval.

Figure 3. Postures used for (A) grip strength measurement and (B) smartphone tasks.

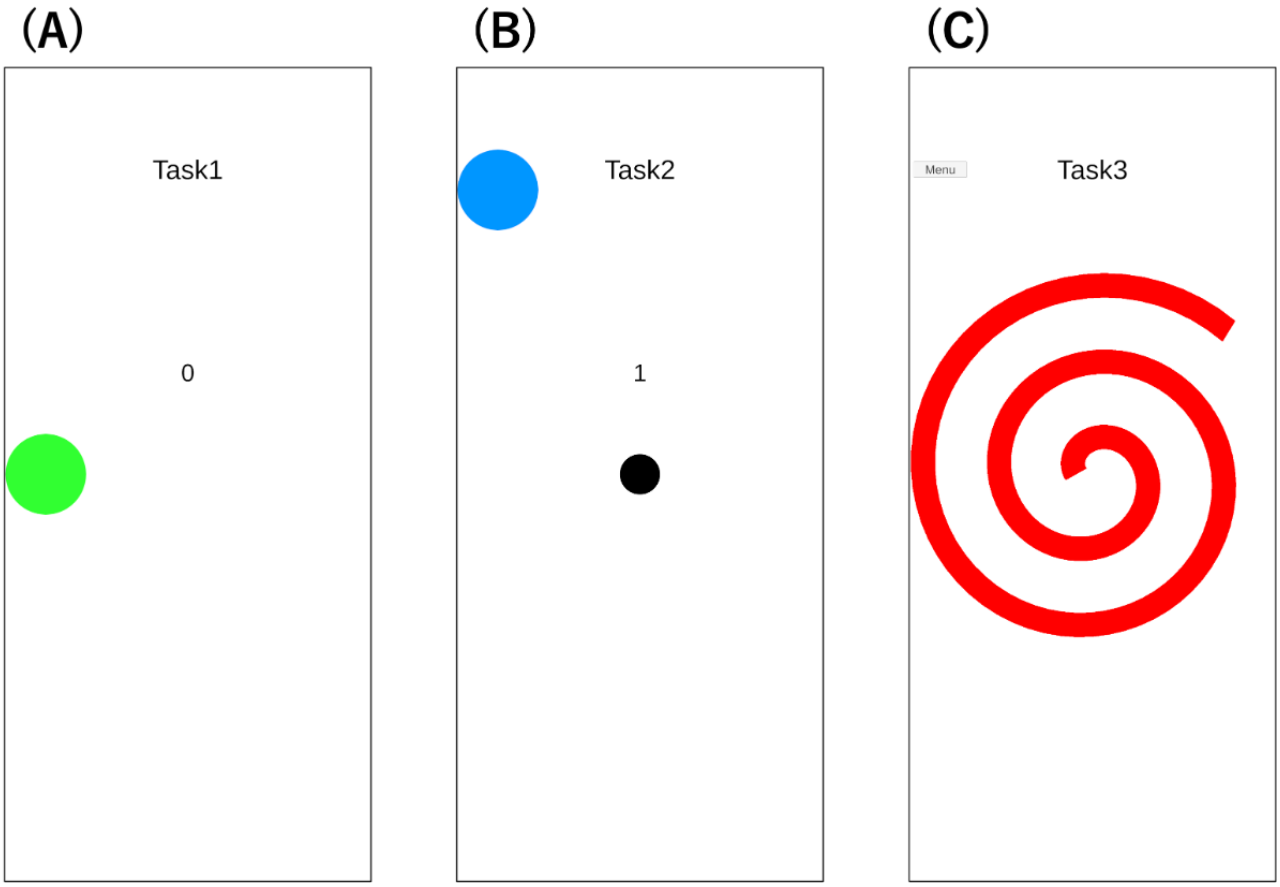


Smartphone Task

A Google Pixel 7a (72.9 mm × 152.0 mm × 9.0 mm) was used for the task, and a custom-developed app built with Unity (version 2022.3.9f1; Unity Software Inc) was used. During the task, data from the device's accelerometer, gyroscope, and orientation sensor were recorded at 60 Hz. Participants sat on a stable chair without a backrest, held the smartphone in their dominant hand, and operated it using their thumb. They were instructed to keep their hand elevated, ensuring that they did not touch the table or their lap (Figure 3B). Additionally, they were asked to perform the following three tasks with careful and deliberate movements: tap task, flick task, and drag task (Figure 4).

XSL•FO
RenderX

Figure 4. Screenshots of the experimental app used for data collection, illustrating (A) the tap task, (B) the flick task, and (C) the drag task.



Tap Task

Participants tapped on a green target (14.6 mm diameter) displayed on the screen. When the target was tapped, a new target appeared at a different location. The task ended after 10 targets were tapped (Figure 4A).

Flick Task

Participants flicked a black circle (7.3 mm diameter) in a specified direction to make it collide with a blue target (14.6 mm diameter). When a successful collision occurred, a new target was generated. The task ended after 10 correct flicks in the designated direction (Figure 4B).

Drag Task

Participants traced a spiral (4.4 mm width) displayed on the screen from the center outward. When using the right hand, a clockwise spiral was used, whereas a counterclockwise spiral was used for the left hand. The task ended after tracing the spiral once (Figure 4C).

Features

We extracted specific behavioral features from each task—tap, flick, and drag—to capture the kinematic characteristics associated with grip strength. For the tap task, the extracted features included the spatial accuracy, defined as the distance between the input coordinates and the center of the target, as well as temporal metrics, such as the time taken to eliminate the target and the duration of the touch interaction. The finger contact area during the tap was also recorded. In the flick task, we calculated the direction of the flick input (represented by x- and y-components), the time taken to eliminate the target, the finger contact area, and the duration of the touch interaction from the start to the end of the flick. For the drag task, features included tracing accuracy, measured as the average distance between the input coordinates and the nearest point on the reference spiral centerline. Additionally, we derived input velocity and acceleration based on the coordinate movement, along with the finger contact area and the total duration of the drag interaction.

Common to all tasks, we used inertial sensor data, including acceleration, angular velocity, and orientation. Linear interpolation was applied to handle any missing values. To capture the hand movement dynamics surrounding the interaction, sensor data were extracted from 3 specific temporal windows: “preinput” (the 10 frames immediately preceding the input event), “during-input” (the average value recorded at the time of input), and “postinput” (the 10 frames immediately following the event) windows. Subsequently, we performed statistical calculations on these extracted data streams to derive the final feature set. These calculations included the differences between consecutive frames to capture the rate of change, as well as the maximum, minimum, range, and mean values to summarize the signal characteristics over the period of interest.

Data Processing

The target variable for our predictive models was the grip strength value, derived by averaging the 2 measurements taken within each set. We then constructed 4 distinct regression

XSL•FO
RenderX

models by pairing this grip strength data with the extracted behavioral features. First, we developed task-specific models, which were trained exclusively on data from each interaction type—namely, the tap, flick, and drag tasks. These models allowed us to evaluate the predictive power of each gesture independently. Second, we constructed a unified model trained on a combined dataset incorporating features from all 3 tasks. By integrating multimodal features across different interaction styles, this unified approach aimed to enhance overall prediction accuracy and establish a comprehensive model capable of estimating grip strength from a variety of daily smartphone gestures.

In this study, we used all the extracted features (touch interaction features, device motion features, and temporal metrics) for model training without applying dimensionality reduction or exclusion techniques. We relied on the inherent feature importance evaluation capabilities of nonlinear models (eg, random forest) to handle potential multicollinearity.

To remove high-frequency noise from the raw sensor data, we applied a low-pass filter with a cutoff frequency of 20 Hz. Before being used as input for the machine learning models, all extracted features were normalized using StandardScaler (scikit-learn; version 1.2.2) to ensure that each feature contributed equally to the model training. For each machine learning model, we performed a grid search to identify the optimal hyperparameters. This systematic search allowed us to select the best-performing configuration for each algorithm, ensuring the robustness of our grip strength estimation.

Model Training and Validation Scenarios

To estimate grip strength, we implemented 4 regression models using the scikit-learn library in Python. The random forest model was configured with 200 estimators, a maximum depth of 20, and the square root of features used for splitting, with the random state fixed at 42 for reproducibility. For support vector regression, we used the radial basis function kernel with a regularization parameter C of 100 and a “scale” kernel coefficient. The gradient boosting (GB) model used 100 estimators, a learning rate of 0.1, and a maximum depth of 5. Finally, the k-nearest neighbors (k-NN) model was set with 9 neighbors, using distance-based weighting and the Manhattan distance metric. Model evaluation was performed using 5-fold cross-validation, comparing performance based on the MAPE and the coefficient of determination.

To evaluate the feasibility of a practical deployment scenario, we assessed the model’s performance when trained on a combination of data from other participants supplemented with varying amounts of the target user’s specific data. Specifically, we conducted evaluations using 2 approaches: leave-one-user-out validation, in which 0 days of the target user’s data were included in the training set, and a personalized calibration approach using 1 to 4 days of the target user’s data (few-day calibration). This experimental design simulates a real-world scenario in which a generic model is calibrated with user-specific data collected during the initial period of use to improve estimation accuracy.

Ethical Considerations

The study protocol was approved by the Research Ethics Committee of Keio University (2025-041). Prior to the experiment, the detailed purpose and procedures of the study were explained to all participants, and written informed consent was obtained. Participants were informed that they could withdraw from the study at any time without penalty. Regarding the exclusion criteria, individuals with preexisting hand impairments or pain were excluded from participation due to safety considerations. As the data collection involved repeated muscle exertion for maximum grip strength measurement, this exclusion was strictly enforced to prevent any risk of injury or symptom exacerbation. All participant data were anonymized and stored on a secure server. Each participant received a cash compensation of ¥2000 (approximately US $13) for their time and participation in the study.

Results

Collected Data

In total, 1050 data points (10 repetitions × 5 days × 21 participants) were collected during the measurements for each tap, flick, and drag task. We used grip strength measured before each task as the ground truth, and we mapped this value onto the data collected in each of the 10 executions of the task. Therefore, each data point consisted of a ground-truth grip strength value and data collected through 1 execution of the task. The average time taken for each task was 6.11 (SD 1.37) seconds for the tap task, 7.45 (SD 1.26) seconds for the flick task, and 3.84 (SD 1.88) seconds for the drag task.

Random Split Cross-Validation Result

Table 1 shows the MAE, Table 2 shows the MAPE, and Table 3 shows the coefficient of determination. In all tasks, the support vector machine (SVM) achieved the best results in terms of both MAPE and coefficient of determination. Additionally, the unified model, which combined features from different tasks, showed a lower MAPE than single-task models. It also achieved the highest coefficient of determination. Figure 5 illustrates a scatter plot of the true values (x-axis) and the estimated values (y-axis) for the test data of the unified model. SVMs, GB, and k-NNs had an average error of approximately 3 kg. For random forest, there were many data points with prediction errors of ≥5 kg for both low and high true values.

XSL•FO
RenderX

Table 1. Mean absolute error of grip strength estimation models across different smartphone tasks.

Task	Random forest (kg), mean absolute error (SD)	Support vector machine (kg), mean absolute error (SD)	Gradient boosting (kg), mean absolute error (SD)	k-nearest neighbors (kg), mean absolute error (SD)
Tap	3.62 (0.24)	3.17 (0.20)	3.27 (0.29)	3.48 (0.29)
Flick	3.99 (0.23)	3.52 (0.17)	3.57 (0.20)	3.60 (0.34)
Drag	3.84 (0.31)	3.46 (0.26)	3.54 (0.31)	3.48 (0.24)
All	3.48 (0.21)	2.62 (0.18)	3.07 (0.28)	2.90 (0.12)

Table 2. Mean absolute percentage error for the proposed estimation method by task type.

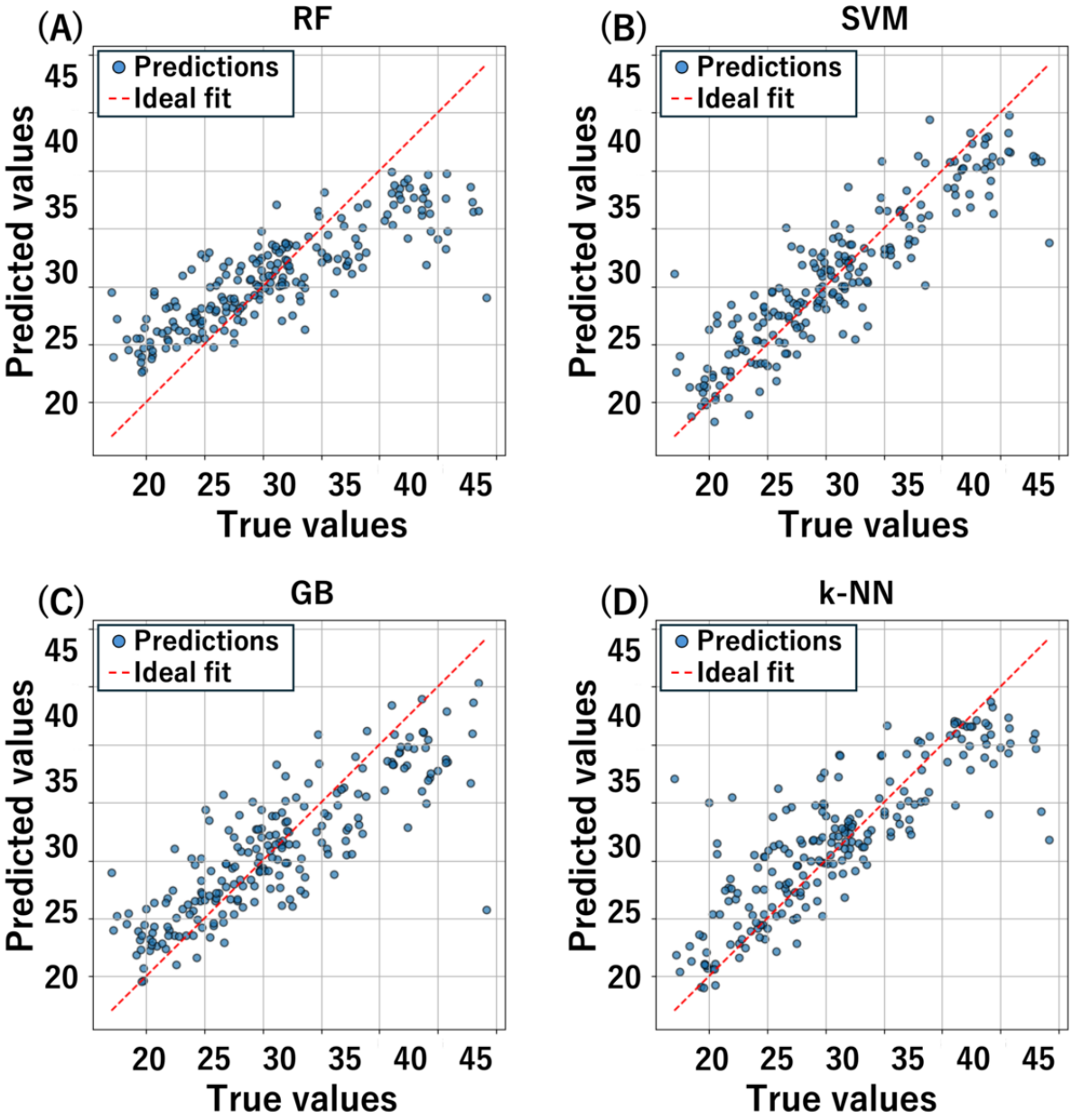
Task	Random forest (%), mean absolute percentage error (SD)	Support vector machine (%), mean absolute percentage error (SD)	Gradient boosting (%), mean absolute percentage error (SD)	k-nearest neighbors (%), mean absolute percentage error (SD)
Tap	12.03 (0.51)	10.74 (0.65)	10.91 (0.63)	11.64 (1.19))
Flick	13.24 (0.55)	11.75 (0.77)	11.94 (0.61)	12.01 (1.25)
Drag	13.02 (0.87)	11.55 (0.71)	11.88 (0.86))	11.69 (0.77)
All	11.87 (0.36)	8.91 (0.57)	10.47 (0.84)	10.15 (1.25)

Table 3. Coefficient of determination (R2) for the performance evaluation of regression models.

Task	Random forest, R^2 (SD)	Support vector machine, R^2 (SD)	Gradient boosting, R^2 (SD)	k-nearest neighbors, R^2 (SD)
Tap	0.610 (0.024)	0.693 (0.039)	0.675 (0.031))	0.650 (0.099)
Flick	0.567 (0.028)	0.644 (0.042)	0.634 (0.030)	0.625 (0.106)
Drag	0.570 (0.033)	0.659 (0.025)	0.648 (0.046)	0.634 (0.012)
All	0.638 (0.026)	0.802 (0.036)	0.731 (0.042)	0.706 (0.095)

XSL•FO
RenderX

Figure 5. Comparison of true and predicted values of grip strength (kg) in the unified model. GB: gradient boosting; k-NN: k-nearest neighbors; RF: random forest; SVM: support vector machine.



Leave-One-User-Out and Few-Day Calibration Validation Results

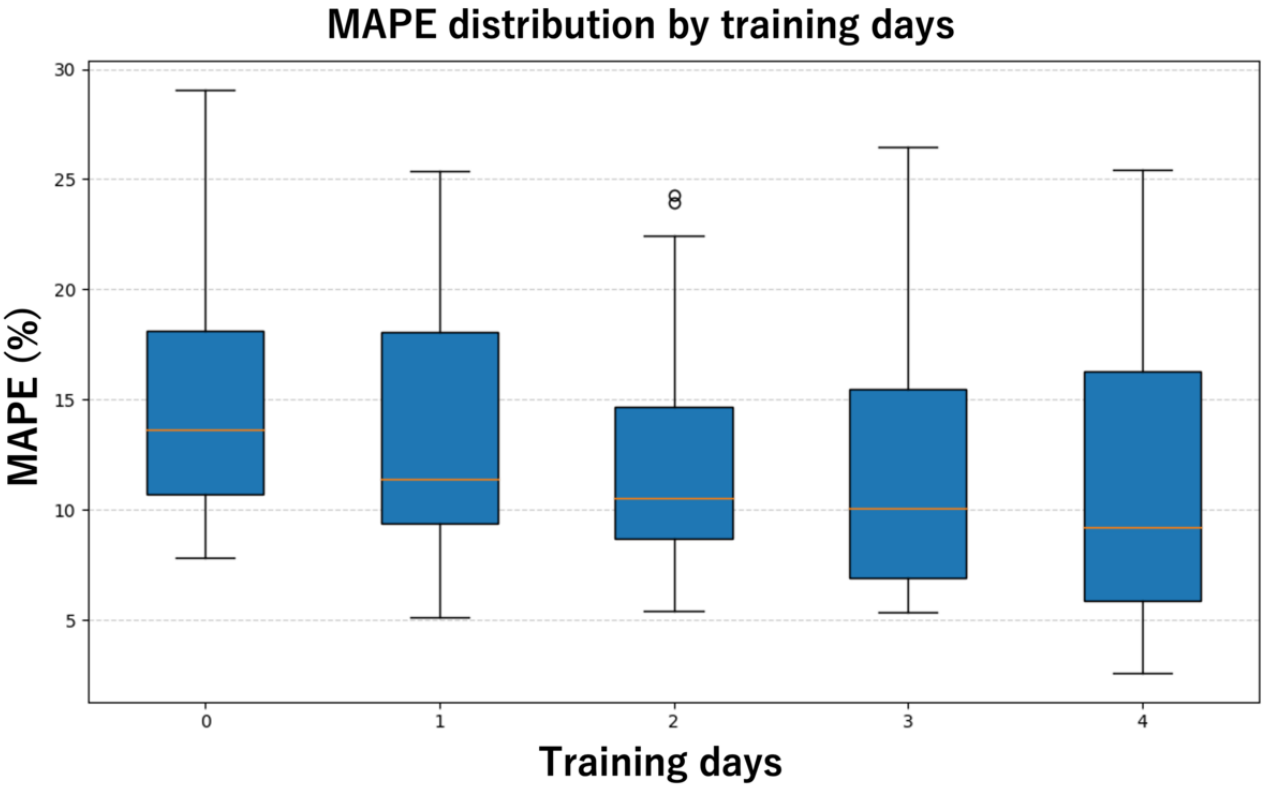
For the leave-one-user-out and few-day calibration validations, the SVM model was used and evaluated based on the MAE and MAPE for each participant. As shown in Table 4 and Figure 6, the estimation accuracy consistently improved as more days of the target user's data were incorporated into the training set. In this evaluation, the case with 0 training days corresponds to leave-one-user-out validation, while the cases with 1 to 4 training days represent the few-day calibration validation.

Table 4. Comparison of mean absolute error and mean absolute percentage error by number of training days.

Variable	Day 0	Day 1	Day 2	Day 3	Day 4
Mean absolute error (kg; SD)	5.03 (2.65)	4.39 (2.28)	3.96 (2.11)	3.80 (2.04)	3.56 (2.23)
Mean absolute percentage error (%; SD)	15.08 (5.40)	13.96 (5.57)	12.60 (5.54)	12.23 (5.68)	11.64 (5.80)

XSL•FO
RenderX

Figure 6. Distribution of estimation errors from the initial state to 4 days of calibration. The box plots illustrate the mean absolute percentage error (MAPE) of the support vector model as the number of user-specific training days increases from 0 to 4. The case with 0 days represents leave-one-user-out validation, while cases with 1 to 4 days represent few-day calibration validation. The horizontal line within each box represents the median error, and the box edges represent the IQR.



Subjective Workload Assessment

To evaluate the subjective workload of the measurement tasks using our proposed method, we administered the National Aeronautics and Space Administration Task Load Index (NASA-TLX) questionnaire to 9 participants. Prior to the assessment, participants were instructed on how to use the app and performed the tap, flick, and drag tasks once. Subsequently, they completed the questionnaire. Following the raw Task Load Index method, participants rated 6 subscales on an 11-point scale ranging from 0 (low) to 10 (high). The results are summarized in Table 5. The overall workload score (mean of all subscales) was 3.04 (SD 2.23), indicating a low level of workload overall. Among the subscales, mental demand (mean 1.89, SD 2.02), effort (mean 1.78, SD 2.10), and frustration (mean 1.78, SD 2.10) were rated particularly low. In contrast, temporal demand showed the highest mean value at 5.11 (SD 3.14).

Table 5. National Aeronautics and Space Administration Task Load Index scores.

Subscales	Descriptions	Mean (SD)
Mental demand	How much mental and perceptual effort did the task require?	1.89 (2.02)
Physical demand	How much physical activity was required?	4.22 (2.86)
Temporal demand	How much time pressure did you experience while performing the task?	5.11 (3.14)
Performance	How successful were you in accomplishing the task?	3.44 (2.50)
Effort	How hard did you have to work?	1.78 (2.10)
Frustration	How irritated, stressed, and annoyed did you feel?	1.78 (2.10)
Overall workload	Average of the 6 subscales	3.04 (2.23)

Discussion

Model Performance and Estimation Characteristics

In the model comparison, SVMs achieved the highest accuracy, followed by GB, k-NNs, and random forest. SVMs and GB effectively captured nonlinear relationships in the data, enabling them to achieve high estimation performance despite the limited dataset. In contrast, k-NNs tend to suffer from decreased accuracy in distance calculations when applied to high-dimensional data, leading to lower overall precision. Additionally, random forest does not focus on reducing errors in the same way as boosting methods, which may result in suboptimal feature selection.

XSL•FO
RenderX

Furthermore, there was a tendency for grip strength to be overestimated when the true value was low and underestimated when the true value was high. This may be attributed to distributional bias in the original data. Most of the data were concentrated between 25 kg and 35 kg, while values >40 kg or <20 kg were relatively scarce. Consequently, the model was likely influenced toward the mean, reducing the estimation accuracy for extreme values.

The observed reduction in estimation errors as the number of calibration days increased suggests that user-specific data significantly enhance the model's ability to account for individual variability. Compared to the leave-one-user-out validation, which yielded the highest error rate, the introduction of even a single day of user data led to a notable improvement. The most substantial improvement occurred between the initial state and the first 2 days of calibration, after which the error rate continued to decrease more gradually. This indicates that while initial personalization is critical to overcoming the limitations of a purely generic model, the system reaches a state of relative stability with approximately 3 to 4 days of data.

The primary dataset involved university students aged 20 to 24 years. To assess the applicability of our framework to broader demographics, we conducted a preliminary case study with a 53-year-old participant (middle-aged adult). We trained a model using the participants' data from 1 day, combined with the entire dataset from the younger group, and tested it on the participants' data from other days. The model achieved an estimation error with an MAE of 1.87 kg (MAPE 11.19%), which is comparable to the accuracy levels obtained from the younger participants. Regarding participants with hand disabilities, we were unable to include them in this study due to ethical considerations. As the data collection tasks involve repeated muscle exertion and impose a certain physical load on the hand, recruiting individuals with existing hand impairments posed a potential risk of exacerbating their conditions.

Practical Applications and Future Directions

In the future, we will implement the proposed model on a smartphone and evaluate its practicality for daily grip strength monitoring. Specifically, we plan to develop a system that inputs data obtained after certain operations into the model and provides feedback on the estimated results. The goal is to track changes in grip strength and use the data for health management and exercise. To achieve this, maintaining accuracy on actual devices and assessing usability will be crucial challenges.

For daily grip strength monitoring, we believe that ease of measurement is one of the most crucial factors. When using the unified model (ie, conducting all tasks), the required time is approximately the same as that for completing one set of smartphone tasks in the data collection experiment. The average completion time for these tasks in our experiment was approximately 30 seconds, which we consider short enough to avoid burdening users in everyday use.

Furthermore, numerous studies have reported on the relationship between smartphone use and grip strength variations, revealing that individuals who spend more time using smartphones daily tend to have lower grip strength [14-16]. Excessive smartphone use can affect hand function and may lead to conditions such as tenosynovitis [17,18]. By using this system, it may be possible to detect grip strength decline during smartphone use, estimate hand fatigue, encourage appropriate use, and reduce the risk of orthopedic disorders.

Subjective Workload and Usability

To evaluate user fatigue and task burden during daily monitoring, we conducted a subjective workload evaluation using the NASA-TLX with 9 participants. The overall workload score was low at 3.04 (SD 2.23) (on a scale of 0 to 10). Specifically, the physical demand score averaged 4.22 (SD 2.86). Given that the task inherently involves physical muscle exertion to estimate grip strength, a moderate level of physical load is expected. However, this value indicates that the task does not impose excessive fatigue compared with maximum voluntary contraction tasks typically performed with a conventional dynamometer.

Notably, the mean temporal demand score was relatively high at 5.11 (SD 3.14), even though participants were explicitly instructed to perform the tasks at their own self-selected pace. This perception may be attributed to the app's feedback design, specifically the display of elapsed time upon task completion. Although participants were instructed to prioritize their own pace, the visualization of a time metric likely created an awareness of *being timed*, inducing a psychological pressure to perform quickly. Consequently, to mitigate this unintended pressure, we plan to redesign the user interface by removing the explicit display of elapsed time in future versions of the app.

Crucially, however, the frustration score remained extremely low at 1.78 (SD 2.10). The clear dissociation between high temporal demand and low frustration implies that this time pressure was not perceived as negative stress or panic but potentially as a gamified element or a source of positive engagement to complete the task efficiently. These results confirm that our proposed method is suitable for continuous daily use without causing significant user burden or stress.

Limitations

To evaluate the generalization of the proposed model across different participants, leave-one-user-out validation was performed. As reported in the Results section, this approach yielded a relatively large estimation error, with an MAE of 5.03 (SD 2.65) kg and an MAPE of 15.08% (SD 5.40%) for the SVM. These results, particularly when contrasted with the improved accuracy observed after just 1 to 4 days of calibration, highlight the significant impact of individual variability in grip strength and smartphone interaction patterns. The high error rate in leave-one-user-out validation suggests that a purely generic model struggles to adequately capture the unique features of each individual's operations. Consequently, our findings indicate that while the model provides a strong foundation, an individual calibration process is currently essential to bridge the gap between a general estimation and the high precision required for practical, personalized health care monitoring.

The accuracy of the proposed model may depend on the smartphone model used. The extracted features, such as device acceleration and angular velocity, could vary due to differences

XSL•FO
RenderX

in device size, weight, and screen dimensions. Consequently, the model may not achieve the same level of accuracy when applied to a different smartphone model. Additionally, the participants in this study used different smartphone models in their daily lives. Some participants struggled with input due to differences in size and shape compared with their usual devices. Their unfamiliarity with the device may have affected task performance. Therefore, it is necessary to explore methods to mitigate the impact of device-related constraints on model performance.

In our experiment, participants were required to hold the smartphone with 1 hand and use their thumb for input. This setup was chosen considering the ease of one-handed operation and the need for consistency in analysis. Meanwhile, factors such as finger length and hand size were not examined in this study. However, because these elements are known to influence touch input accuracy [19], factoring in such variations may improve the accuracy of grip strength estimation across diverse users.

Additionally, using 1 hand with the thumb was the primary interaction style in our experiment. Consequently, the current model is optimized exclusively for this specific posture. We recognize that users often use alternative styles, such as holding the device with 1 hand and operating it with the index finger of the other, which can significantly alter sensor data profiles. To ensure data consistency in a practical deployment, we propose implementing on-screen instructions or a brief tutorial to explicitly guide users to adopt the required one-handed thumb posture during the measurement tasks. Expanding the model to accommodate diverse interaction styles remains an important direction for future work.

In this study, data collection was conducted exclusively with right-handed participants operating the smartphone with their dominant (right) hand. To extend the applicability of our model to left-handed users operating with their dominant hand, we propose a geometric transformation approach. As the anatomical structure of the hands is symmetrical, left-hand interaction data can be mapped to the right-hand feature space by mirroring the horizontal coordinates (x-axis) of touch points and reversing the rotation direction of the spiral drag task (ie, treating counterclockwise input as clockwise). This preprocessing allows the model trained on right-dominant data to be directly applied to left-dominant users.

However, the validity of estimating grip strength using the nondominant hand (regardless of whether it is the left or right hand) remains unverified, as differences in dexterity and motor control may significantly alter sensor patterns compared with the dominant hand.

Consequently, to ensure valid estimation in a real-world deployment, the app requires an input interface where the user explicitly specifies two parameters before use: (1) which hand is currently being used (left or right) and (2) whether it is their dominant or nondominant hand. On the basis of these inputs, the system can determine whether to apply geometric transformations (for left-dominant input) or to notify the user that the current condition (nondominant input) is outside the verified scope of the model.

We acknowledge that this study used a single smartphone model (Pixel 7a). As inertial sensor data (acceleration and angular velocity) are physically influenced by the device's weight and dimensions, verifying reproducibility on different devices is crucial. However, the Pixel 7a (weight 193.5 g; display 15.5 cm) represents a standard form factor in the current smartphone market. Therefore, we anticipate that the fundamental physical interactions between grip strength and device motion would remain consistent, and the influence on estimation accuracy would be limited for devices with similar specifications.

To strictly address the issue of varying sensor characteristics (eg, bias or sensitivity) across different hardware in the future, we propose 3 potential approaches. First, sensor calibration could be introduced as a preprocessing step to identify and correct for device-specific sensor characteristics, thereby standardizing the input data across different models. Second, we suggest focusing on robust feature extraction, which prioritizes features that are independent of raw sensor hardware variations, such as relative motion patterns, rather than relying on absolute sensor values. Finally, if a large-scale dataset across multiple devices becomes available, device-specific features—such as model ID, weight, or screen size—could be explicitly incorporated as input variables to train a more generalized and hardware-agnostic model.

Comparison With Prior Work

Several studies have explored methods for estimating grip strength without using a grip dynamometer. Matsumoto et al [20] proposed a method for estimating grip strength by capturing an image of a hand gripping a soft tennis ball using a monocular camera on a smartphone and analyzing changes in the angle of finger joints. Yamamoto et al [21] proposed a grip strength measurement method using capacitive touch sensors on smartphones, which does not require any electronic components on the measuring device itself. Although these methods offer alternative approaches to measuring grip strength, they still require external equipment beyond a smartphone, and the measurement results may be influenced by the user's posture. In this study, our objective is to estimate the grip strength using only the smartphone, with the aim of establishing an easy and convenient measurement method without specialized equipment.

This study examines the estimation errors of the proposed grip strength model. The Jamar dynamometer manufacturer recommends a measurement error within 5% [22], whereas the error rate in this study slightly exceeds this threshold. We also evaluated the clinical relevance of our estimation accuracy. The minimal clinically important difference for grip strength is reported to be between 5.0 kg and 6.5 kg [23]. Our model achieved an MAE of 2.62 (SD 0.18) kg in the random split and 4.39 (SD 2.28) kg in the single-day case of the few-day calibration, both falling within this minimal clinically important difference range. This indicates that our method is effective for detecting meaningful changes in muscle strength over time, such as recovery or decline. However, a critical limitation arises when applying this method to disease screening based on specific thresholds. For instance, the Asian Working Group for Sarcopenia 2019 defines diagnostic cutoffs of <28 kg for male and <18 kg for female individuals [24]. Given our estimation

XSL•FO
RenderX

error of approximately 2.6 kg to 4.4 kg, there is a risk of misclassification (false positives or false negatives) for individuals whose grip strength lies near these borderline values. Consequently, we position our method not as a replacement for clinical-grade dynamometers for definitive diagnosis but rather as a pervasive tool for daily longitudinal monitoring and initial screening to identify at-risk individuals who should seek professional assessment.

The accuracy of the proposed model is compared with previous grip strength estimation studies. Park et al [25] used national survey data from >20,000 participants, where grip strength was measured using a digital grip strength dynamometer. Their analysis used age, sex, hand preference, height, weight, waist circumference, and 5 physical activity levels (vigorous or moderate work, vigorous or moderate recreation, and travel) as input features, achieving a coefficient of determination of 0.717 with Extreme Gradient Boosting. Hwang et al [26] developed a deep learning model trained on data from 164 participants, incorporating various postures, demographic information, and anthropometric data as input, achieving a coefficient of determination of 0.88. This study, in contrast, focuses on estimating grip strength using everyday smartphone interactions, making the input data fundamentally different from those used in previous studies. However, the limited number of participants and the small training dataset present challenges in terms of reliability and accuracy. Additionally, further investigation into user-independent models is necessary. Future work should aim to expand the dataset and improve the model to enhance estimation accuracy.

Conclusions

In this study, we proposed a model for estimating grip strength using smartphone input data. First, grip strength was measured, followed by the execution of smartphone tasks, during which sensor data from the device were collected. On the basis of the obtained data, multiple machine learning models were constructed and evaluated. The results showed that the model using SVMs achieved the highest accuracy. Furthermore, integrating features obtained from different tasks improved estimation accuracy. In the future, we aim to implement the model on a smartphone to enable real-time grip strength estimation. To achieve this, further accuracy improvements and evaluations under various conditions will be necessary.

Data Availability

The source code for the smartphone app developed in this study and the data analysis scripts are available in the GitHub repository [27].

Funding

This work was partially supported by Japan Science and Technology Agency Advanced Integrated Intelligence Platform Project Acceleration Research (grant JPMJCR25U4) and Japan Society for the Promotion of Science KAKENHI (grant JP25K03168).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Video demonstrating the smartphone tasks (tap, flick, and drag) and providing an overview of the grip strength estimation method.
[MP4 File (MP4 Video), 117611 KB - humanfactors_v13i1e90316_app1.mp4]

References

1. Kallman DA, Plato CC, Tobin JD. The role of muscle loss in the age-related decline of grip strength: cross-sectional and longitudinal perspectives. J Gerontol 1990 May;45(3):M82-M88. [doi: 10.1093/geronj/45.3.m82] [Medline: 2335723]
2. Keller K, Engelhardt M. Strength and muscle mass loss with aging process. Age and strength loss. Muscles Ligaments Tendons J 2013 Oct;3(4):346-350 [FREE Full text] [Medline: 24596700]
3. Cesari M, Leeuwenburgh C, Lauretani F, Onder G, Bandinelli S, Maraldi C, et al. Frailty syndrome and skeletal muscle: results from the Invecchiare in Chianti study. Am J Clin Nutr 2006 May;83(5):1142-1148 [FREE Full text] [doi: 10.1093/ajcn/83.5.1142] [Medline: 16685058]
4. Larsson L, Degens H, Li M, Salviati L, Lee YI, Thompson W, et al. Sarcopenia: aging-related loss of muscle mass and function. Physiol Rev 2019 Jan 01;99(1):427-511 [FREE Full text] [doi: 10.1152/physrev.00061.2017] [Medline: 30427277]
5. Hershkovitz A, Yichayaou B, Ronen A, Maydan G, Kornyukov N, Burstin A, et al. The association between hand grip strength and rehabilitation outcome in post-acute hip fractured patients. Aging Clin Exp Res 2019 Oct;31(10):1509-1516. [doi: 10.1007/s40520-019-01200-y] [Medline: 30993662]
6. Klatt ME, Krakovsky A, O'Callaghan L. Assessment of grip strength across the continuum of care. HAND 2016 Sep 26;11(1_suppl):151S-152S. [doi: 10.1177/1558944716660555KR]
7. Lee SH, Gong HS. Grip strength measurement for outcome assessment in common hand surgeries. Clin Orthop Surg 2022 Mar;14(1):1-12 [FREE Full text] [doi: 10.4055/cios21090] [Medline: 35251535]

XSL•FO
RenderX

8. Boonpor J, Parra-Soto S, Petermann-Rocha F, Ferrari G, Welsh P, Pell JP, et al. Associations between grip strength and incident type 2 diabetes: findings from the UK Biobank prospective cohort study. BMJ Open Diabetes Res Care 2021 Aug;9(1):e001865 [FREE Full text] [doi: 10.1136/bmjdrc-2020-001865] [Medline: 34353878]
9. Cetinus E, Buyukbese MA, Uzel M, Ekerbicer H, Karaoguz A. Hand grip strength in patients with type 2 diabetes mellitus. Diabetes Res Clin Pract 2005 Dec;70(3):278-286. [doi: 10.1016/j.diabres.2005.03.028] [Medline: 15878215]
10. Li JJ, Wittert GA, Vincent A, Atlantis E, Shi Z, Appleton SL, et al. Muscle grip strength predicts incident type 2 diabetes: population-based cohort study. Metabolism 2016 Jun;65(6):883-892. [doi: 10.1016/j.metabol.2016.03.011] [Medline: 27173467]
11. Mainous AG, Tanner RJ, Anton SD, Jo A. Grip strength as a marker of hypertension and diabetes in healthy weight adults. Am J Prev Med 2015 Dec;49(6):850-858 [FREE Full text] [doi: 10.1016/j.amepre.2015.05.025] [Medline: 26232901]
12. Fujita K, Kaburagi H, Nimura A, Miyamoto T, Wakabayashi Y, Seki Y, et al. Lower grip strength and dynamic body balance in women with distal radial fractures. Osteoporos Int 2019 May;30(5):949-956 [FREE Full text] [doi: 10.1007/s00198-018-04816-4] [Medline: 30607458]
13. Severijns D, Zijdewind I, Dalgas U, Lamers I, Lismont C, Feys P. The assessment of motor fatigability in persons with multiple sclerosis: a systematic review. Neurorehabil Neural Repair 2017 May;31(5):413-431. [doi: 10.1177/1545968317690831] [Medline: 28413944]
14. Alshahrani A, Samy Abdrabo M, Aly SM, Alshahrani MS, Alqhtani RS, Asiri F, et al. Effect of smartphone usage on neck muscle endurance, hand grip and pinch strength among healthy college students: a cross-sectional study. Int J Environ Res Public Health 2021 Jun 10;18(12):6290 [FREE Full text] [doi: 10.3390/ijerph18126290] [Medline: 34200762]
15. Bhamra JK, Naqvi WM, Arora SP. Effect of smartphone on hand performance and strength in the healthy population. Cureus 2021 Jun;13(6):e15798 [FREE Full text] [doi: 10.7759/cureus.15798] [Medline: 34306866]
16. Osailan A. The relationship between smartphone usage duration (using smartphone's ability to monitor screen time) with hand-grip and pinch-grip strength among young people: an observational study. BMC Musculoskelet Disord 2021 Feb 15;22(1):186 [FREE Full text] [doi: 10.1186/s12891-021-04054-6] [Medline: 33588812]
17. Dhale T, Verma P, Dhande S, Bandre P, Kumar P. A review on the risk of musculoskeletal disorder in humans originating from excessive use of smartphone. In: Proceedings of the 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things. 2024 Presented at: ICoICI 2024; August 28-30, 2024; Coimbatore, India. [doi: 10.1109/ICoICI62503.2024.10696192]
18. Zirek E, Mustafaoglu R, Yasaci Z, Griffiths MD. A systematic review of musculoskeletal complaints, symptoms, and pathologies related to mobile phone usage. Musculoskelet Sci Pract 2020 Oct;49:102196. [doi: 10.1016/j.msksp.2020.102196] [Medline: 32861360]
19. Larsen JN, Jacobsen TH, Boring S, Bergström J, Pohl H. The influence of hand size on touch accuracy. In: Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Service. 2019 Presented at: MobileHCI '19; October 1-4, 2019; Taipei, Taiwan. [doi: 10.1145/3338286.3340115]
20. Matsumoto N, Fujita K, Sugiura Y. Estimation of grip strength using monocular camera for home-based hand rehabilitation. SICE J Control Meas Syst Integr 2021 Feb 18;14(1):1-11. [doi: 10.1080/18824889.2020.1863612]
21. Yamamoto S, Gu F, Ikematsu K, Kato K, Sugiura Y. Maintenance-free smart hand dynamometer. In: Proceedings of the 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society. 2023 Presented at: EMBC 2023; July 24-27, 2023; Sydney, Australia. [doi: 10.1109/embc40787.2023.10340847]
22. Wozny R, Pratt AL, Pereira C. A study of grip strength among 20–49-year-old British adults and comparison to existing norms. Hand Ther 2015 Aug 24;20(4):115-123. [doi: 10.1177/1758998315599792]
23. Bohannon RW. Minimal clinically important difference for grip strength: a systematic review. J Phys Ther Sci 2019 Jan;31(1):75-78 [FREE Full text] [doi: 10.1589/jpts.31.75] [Medline: 30774209]
24. Chen LK, Woo J, Assantachai P, Auyeung TW, Chou MY, Iijima K, et al. Asian Working Group for Sarcopenia: 2019 consensus update on sarcopenia diagnosis and treatment. J Am Med Dir Assoc 2020 Mar;21(3):300-7.e2. [doi: 10.1016/j.jamda.2019.12.012] [Medline: 32033882]
25. Park M, Kim YH, Lee JS. Machine learning-based model for grip strength prediction in healthy adults: a nationwide dataset-based study. J Clin Med 2025 Feb 25;14(5):1542 [FREE Full text] [doi: 10.3390/jcm14051542] [Medline: 40095470]
26. Hwang J, Lee J, Lee KS. A deep learning-based method for grip strength prediction: comparison of multilayer perceptron and polynomial regression approaches. PLoS One 2021;16(2):e0246870 [FREE Full text] [doi: 10.1371/journal.pone.0246870] [Medline: 33571318]
27. KEIO-LCLAB/GripApp. GitHub. URL: https://github.com/KEIO-LCLAB/GripApp/tree/master [accessed 2026-04-18]

Abbreviations

GB: gradient boosting
k-NN: k-nearest neighbors
MAE: mean absolute error
MAPE: mean absolute percentage error

XSL•FO
RenderX

NASA-TLX: National Aeronautics and Space Administration Task Load Index
SVM: support vector machine

Edited by J Sarvestan; submitted 25.Dec.2025; peer-reviewed by L Herdiman; comments to author 27.Jan.2026; accepted 26.Feb.2026; published 23.Apr.2026.

Please cite as:
Tajima K, Ikematsu K, Isomoto T, Kato K, Sugiura Y
Grip Strength Estimation Using Input Data From a Commodity Smartphone: Model Development and Validation Study
JMIR Hum Factors 2026;13:e90316
URL: https://humanfactors.jmir.org/2026/1/e90316
doi:10.2196/90316

©Komei Tajima, Kaori Ikematsu, Toshiya Isomoto, Kunihiro Kato, Yuta Sugiura. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 23.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Technology Acceptance of 2 mHealth Apps During an Observational Technology Evaluation Study: Qualitative Results of a Mixed Methods Study

Johanna Graeber[1], Dipl Psych; Elke Warmerdam[2,3], PhD; Svenja Aufenberg[4], Dipl.-Biol.; Christopher Bull[5], PhD; Kristen Davies[6], MRes, MBChB; Juliane Döhring[1], Dr Phil; Kirsten Emmert[7,8], PhD; Claire Judd[9], MBBS, MRes; Corina Maetzler[7], MA; Nikolay V Manyakov[10], MSc, PhD; Ralf Reilmann[4,11,12], Dr med; Wan-Fai Ng[6,9,13], MB, BChir, PhD; Victoria Macrae[6]; Walter Maetzler[7], Prof, Dr med; Hanna Kaduszkiewicz[1], Prof, Dr med

[1]Institute of General Practice, University Medical Center Schleswig-Holstein, Kiel University, Kiel, Germany
[2]Innovative Implant Development (Fracture Healing), Division of Surgery, Saarland University, Saarbrücken, Germany
[3]Department of Neurology, University Medical Center Schleswig-Holstein, Kiel University, Kiel, Germany
[4]George Huntington Institute, Münster, Germany
[5]OpenLab, School of Computing, Newcastle University, Newcastle upon Tyne, United Kingdom
[6]Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, United Kingdom
[7]Department of Neurology, University Medical Center Schleswig-Holstein, Campus Kiel, Kiel University, Kiel, Germany
[8]Department of Health Promotion and Prevention, Institute of Public Health and Nursing Research, University of Bremen, Bremen, Germany
[9]NIHR Newcastle Clinical Research Facility, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, United Kingdom
[10]Johnson & Johnson, Beerse, Belgium
[11]Department of Clinical Radiology, University of Münster, Münster, Germany
[12]Department of Neurodegeneration, Hertie Insitute for Clinical Brain Research, University of Tübingen, Tübingen, Germany
[13]Clinical Research Facility, University College Cork, Cork, Ireland

Corresponding Author:
Johanna Graeber, Dipl Psych
Institute of General Practice
University Medical Center Schleswig-Holstein
Kiel University
Michaelisstraße 5, Haus U30
Kiel, 24105
Germany
Phone: 49 431 500 30115
Fax: 49 431 500 30104
Email: jg@allgemeinmedizin.uni-kiel.de

Abstract

Background: Mobile health (mHealth) apps are useful tools for research and disease management. However, implementation of mHealth apps is lacking in many areas. While mHealth apps offer various advantages to researchers and patients, their effectiveness depends on their actual use. Barriers to using mHealth apps are often due to human factors such as usability or technology acceptance. Although prior studies have examined the acceptance of mHealth apps in patient treatment, the key factors driving or hindering the use of mHealth apps in research remain unclear.

Objective: This study explores user perceptions of 2 mHealth apps in the setting of an observational technology evaluation study using the unified theory of acceptance and use of technology. We aim to evaluate the technology acceptance of these specific apps and to investigate challenges in choosing suitable mHealth apps in research. The apps were intended for data collection; no effect on health was expected.

Methods: Patients with chronic diseases as well as healthy participants used a symptom tracking app and a cognitive test app over the course of 4 weeks within the feasibility study of the project “Identifying Digital Endpoints to Assess Fatigue, Sleep and Activities of Daily Living in Neurodegenerative Disorders and Immune-Mediated Inflammatory Diseases.” Thereafter, 61

XSL•FO
RenderX

qualitative interviews were conducted, recorded, and transcribed. A qualitative content analysis using the unified theory of acceptance and use of technology was performed.

Results: An important aspect of motivation for participants was feedback on their health data and performance in the cognitive tests. Effort played a significant role in app use. Patients rated the apps as easy to use and quick. Using the app multiple times per day at fixed times was perceived as disruptive. Participants preferred using their own phone. Social influence as well as facilitating conditions played a lesser role in intention to use the apps. Data security was no concern for most participants. They stressed the importance of good relations with the study team.

Conclusions: In choosing suitable apps, one size will certainly not fit all. For medical research, pretesting of all materials with the potential users is of utmost importance. If the positive effects of the app on users' health are not immediately apparent, other factors may motivate use, for example, feedback, gamification, adjustable functions, applicability on all smartphone operating systems, and good relations to the study team.

(JMIR Hum Factors 2026;13:e70873) doi:10.2196/70873

KEYWORDS

mobile health care; mobile health; mHealth; digital health; technology use; neurodegeneration; app; user experience; mixed method study

Introduction

Technology has become increasingly important in maintaining our health and treating illness [1]. This affects not only care providers but also individuals who take advantage of technologies such as websites, smart devices, and apps. Mobile health (mHealth) apps, for example, monitor users' health, remind themselves of their medication, and connect with other patients and their care providers [2-4]. The increase in use has led to an increase in the availability of these services [2,5-7]. mHealth apps can be useful tools for patients as well as medical staff. They can be particularly effective in the self-management of patients, collection of health data, support of diagnosis, and disease management [2-4,7,8]. In clinical research, mHealth apps deliver convenient access to large amounts of real-time health data from users' everyday lives [2]. This combination of advantages highlights the growing significance of mHealth apps in patient care and medical research [9].

Although a large number of mHealth apps are available and their advantages seem obvious, implementation is still lacking in many areas. Barriers that prevent the implementation of mHealth apps are often due to human issues, such as user acceptance, low familiarity with technology, and resistance to change [10]. A prerequisite to the successful implementation of an mHealth app is its usability. Technology acceptance and the ability to implement the technology in daily life are further requirements [11]. The effectiveness of a new mHealth app depends on its actual use [5]. Therefore, the experiences and opinions of users and patients must be taken into account when evaluating mHealth apps. Understanding what aspects of mHealth apps facilitate or hinder user acceptance is an important step to promote their implementation.

The concept of technology acceptance was originally used in organizational contexts [12], such as the adoption of new technology in the workplace. However, its importance in health research has increased with the growing use of health technology such as mHealth apps [10]. Different models have been used to explain the intention and use of a technology [6], such as the theory of reasoned action or the technology acceptance model [12-14]. The unified theory of acceptance and use of technology (UTAUT) was developed to unify these different models [14] and has been widely used in mHealth research [6,15]. It postulates that the actual use of a technology can be predicted by the behavioral intention to use a technology. The behavioral intention is determined by 4 constructs: performance expectancy, effort expectancy, social influence, and facilitating conditions [13]. Performance expectancy is defined as the degree to which a user expects a technology to help him or her. Several studies on mHealth services find positive effects of performance expectancy on behavioral intention to use a technology [6,16-18]. Effort expectancy describes the assumed ease of use of a technology and has been strongly associated with a greater intention to use in an mHealth context, especially in older populations [6,16]. Venkatesh et al [13] defined social influence as "the degree to which a person perceives that important others believe he or she should use the new system." Social influence also has a positive effect on behavioral intention in mHealth contexts [6,16]. Facilitating conditions describe the perceived organizational and technical infrastructure in support of using a technology, such as instructions or assistance [16].

A number of studies have been conducted on the technology acceptance of mHealth apps using the UTAUT or related models. Many of these studies focus on mHealth apps designed to help patients with self-management of their health such as hypertension or weight management apps [17,19,20]. Studies on diabetes management apps point out the importance of trust, particularly regarding data security and perceived disease threat [21,22]. Similar results were found regarding technology acceptance on COVID-19 contact tracing apps [23]. Technology acceptance of mHealth apps focusing on health information has also been investigated. Intention to use a prenatal care app has been investigated using the UTAUT [24]. Alsyouf et al [25] stressed the importance of data security on intention to use a personal health record system. However, the majority of these studies focus on mHealth apps that provide direct benefits to their users, such as support in health management or providing health information. The notion of perceived usefulness or performance expectancy is a well-described predictor of intention to use mHealth apps [6,7,18]. This influence might

XSL•FO
RenderX

be diminished in the context of research-focused mHealth apps because they do not offer direct benefits for the user.

Here, we report the results of a study on the technology acceptance of 2 mHealth apps developed for research purposes. Patients with Parkinson disease, Huntington disease, rheumatoid arthritis, systemic lupus erythematosus, primary Sjögren syndrome, and inflammatory bowel disease as well as healthy participants used 2 mHealth apps over the course of 4 weeks. This study was part of a larger evaluation study with the goal to examine usability of these mHealth apps as well as other digital technologies. The symptom tracking app included questionnaires regarding the patient's state of health, that is, sleepiness, fatigue, and mood. The cognitive test app included 3 different cognitive tests.

In short, this study serves 2 purposes. First, we assess the technology acceptance of these apps to evaluate their usability for a follow-up study. Second, we explore the specific challenges involved in selecting suitable data-gathering mHealth apps for medical research. We hope these insights will be helpful for researchers in planning future studies using mHealth apps and help prevent data loss due to low technology acceptance.

Methods

Study Overview

Study Design of the IDEA-FAST Feasibility Study

We conducted qualitative semistructured interviews with participants of the feasibility study (FS) of the IDEA-FAST project ("Identifying Digital Endpoints to Assess Fatigue, Sleep and Activities of Daily Living in Neurodegenerative Disorders and Immune-Mediated Inflammatory Diseases"). IDEA-FAST investigates whether digital parameters can be used to assess fatigue and sleep disturbances. The FS was an observational study, in which different technologies and mHealth apps were assessed for their suitability for the larger clinical observation study. The FS included 2 mHealth apps, a symptom tracking app [26] and a cognitive test app. Quantitative and qualitative measures were used to evaluate the suitability of the mHealth apps. For the qualitative evaluation, patients as well as healthy participants were interviewed after using the apps for 4 weeks.

The design and rationale of the FS as well as baseline characteristics have been described previously [27]. In short, digital technologies, including symptom tracking and cognitive test apps, were implemented alongside surveys and patient-reported outcome measures. The FS was conducted from August 2020 until June 2021 at 4 study sites: Erasmus University Medical Centre, Rotterdam (the Netherlands), Newcastle University (the United Kingdom), University Medical Centre Schleswig-Holstein, Kiel, and George Huntington Institute GmbH, Münster (Germany). Participants were introduced to the mHealth apps at a baseline visit with their local site. Afterward, participants used the apps for 4 weeks, each week consisting of a 5-day wearing period and a 2-day rest period. Taken together, the apps were used for 20 days. Additional information and support were offered by clinical staff either via support materials, tutorials, telephone, and/or in-person visits.

mHealth Apps

Overview

Both apps could be installed either on participants' personal phones or on a study phone provided by the research staff. The symptom tracking app was available for Android operating systems, and the cognitive test app for both Android and IOS operating systems.

Symptom Tracking App

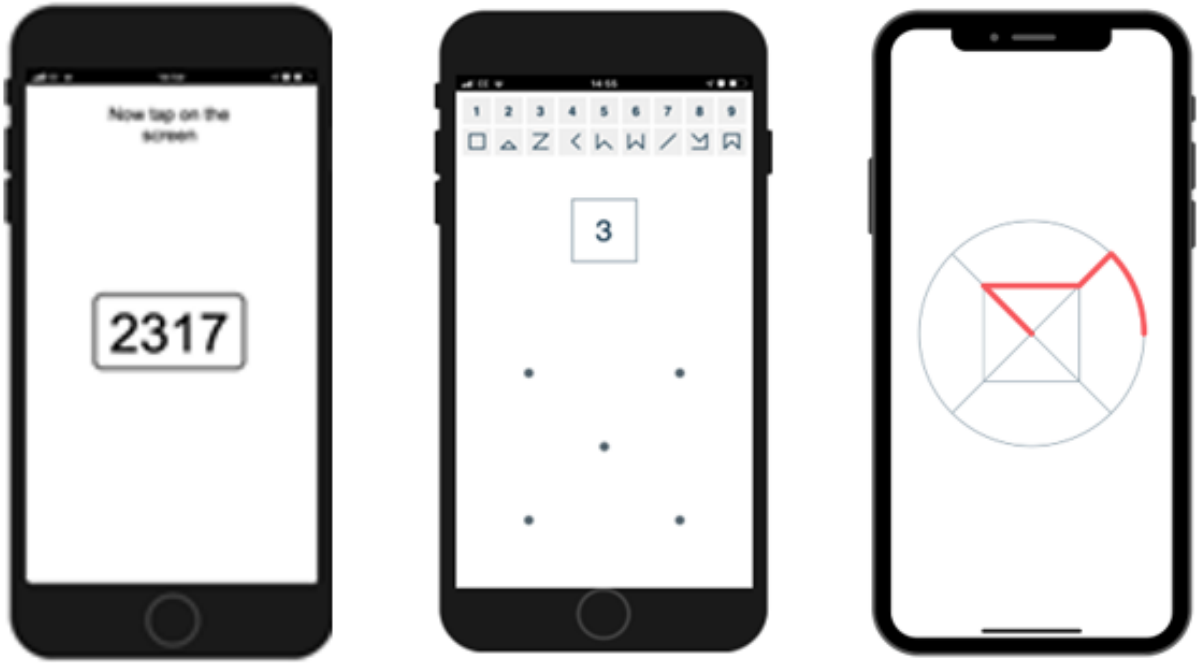
The symptom tracking app has been developed by VTT in accordance with the study design. It was scheduled to be filled out 4 times a day, at 9 AM, 1 PM, 5 PM, and 9 PM. Participants received notifications to answer questions and were given a time frame of several hours for each action within the app. Each action consisted of several questions specially designed for the study regarding current fatigue, sleepiness, mood, and pain. In the morning, users answered additional questions regarding sleep duration, sleep disturbances, and sleep quality. In the evening, physical and mental activity during the day was assessed. In addition, the study team decided to add a voice feature to allow participants to verbally share information about their day. Furthermore, 2 speech tests were included in the app for study purposes: an oral motor skill task, in which participants had to make certain sounds; and a temporal orientation task, where participants stated the time of day. The app is equipped with a reminder function using sounds, which can be turned off.

Cognitive Test App

The cognitive test app has been developed by Cambridge Cognition, also in accordance with the study design. It required participants to complete a cognitive task twice a day. The psychomotor vigilance task (PVT) was administered in the first week, followed by the digit symbol substitution test (DSST) in the second week, and the n-back task (NBX) during weeks 3 and 4 [28]. The tests can be seen in Figure 1. The PVT measures reaction time, alertness, and attention. As participants look at the screen of their smartphone, a number appears, and as soon as they see the number, they have to touch the screen as quickly as possible. The DSST measures cognitive ability and takes 1.5 minutes. Participants see abstract symbols associated with numbers at the screen of their smartphone. When a particular number appears on the screen, they have to draw the corresponding symbol. Participants are asked to draw as many symbols as possible. The NBX measures attention and working memory and takes up to 2 minutes. Participants see a stream of abstract symbols. They must touch the screen whenever a visible symbol matches the one that appeared 2 symbols previously [29].

XSL•FO
RenderX

Figure 1. The 3 tests of the cognitive test app.

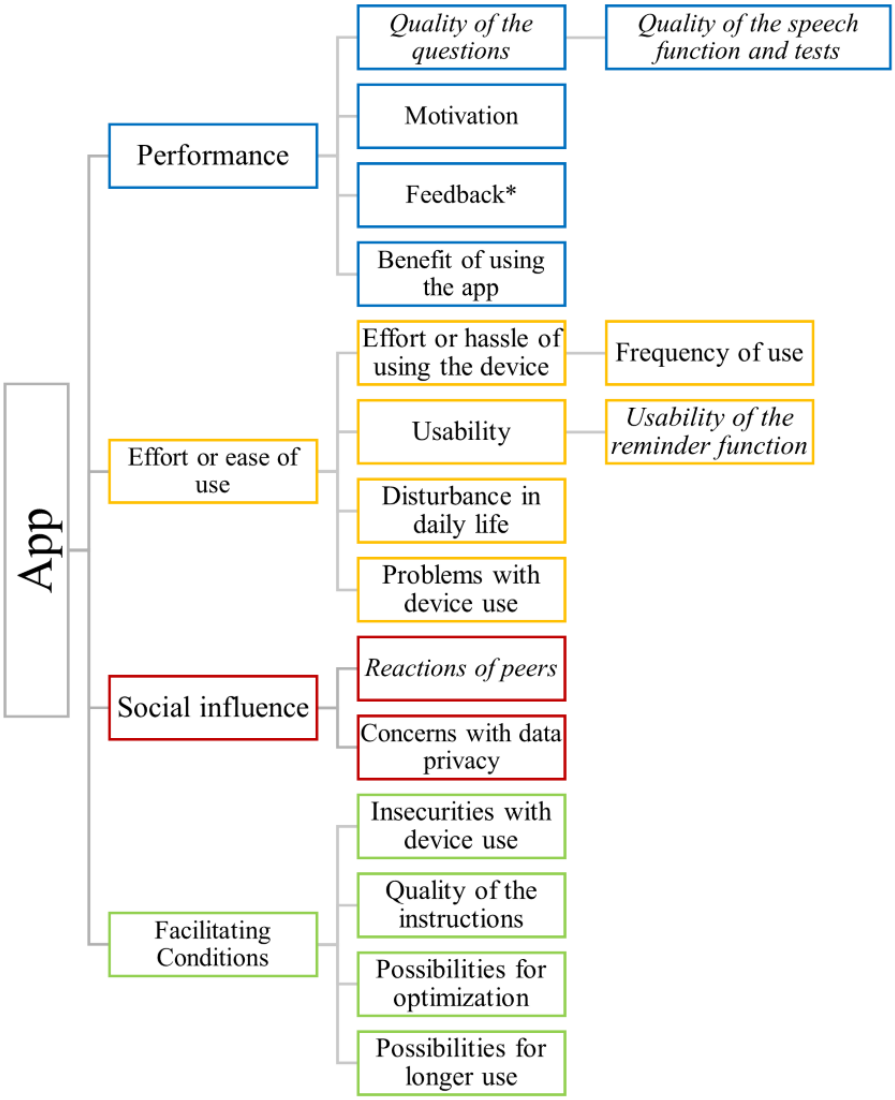


Qualitative Interviews

The general implementation and transcription procedures of the interviews have been described previously [27]. In short, an interview guideline based on the UTAUT model was developed. After an introduction including verbal consent, some basic demographic questions followed. Participants were asked about each app separately. First, participants were asked to describe their general experiences with the app, followed by more detailed questions to cover all aspects of UTAUT. Participants in the FS were contacted during the study to assess their willingness to take part in an interview. Interviews were conducted between October 2020 and June 2021, mostly via videoconference. Written consent to participate in the interviews was obtained beforehand. Interviews were conducted in the first language of the participants and recorded with their consent. They were transcribed verbatim or pseudonymized by a transcription service. Interviews in Dutch were translated into English with an online translation tool and double-checked by a native speaker. Quotations used in this paper were slightly edited for readability without altering their meaning. All interviews were analyzed using content analysis following Mayring [30] with MAXQDA 2022 (version 22.8.0; VERBI GmbH). Overall, 59 complete interviews and 2 summaries of interviews (where recording had not worked) were included in the analysis. Initially, 4 broad top-level codes were defined to classify the interview content according to the 4 constructs of technology acceptance. In the next step, potential subcodes were delineated on the basis of the interview guide. Subsequently, the interviews were analyzed using these preexisting codes, with the additional step of using an inductive method to derive new codes. Applying this method to both apps resulted in comparable coding systems that still account for individual differences. The resulting code system is illustrated in Figure 2.

XSL•FO
RenderX

Figure 2. General code system of the apps. Codes present in italics format are only used for the symptom tracking app, and feedback* is only used for the cognitive test app.



While the UTAUT model was used as a foundation for both the interview guideline and the analysis, it was slightly adapted to fit this study. As participants in a medical study, users did not expect the devices to be personally useful and were unable to judge their usefulness to the study team. In addition, as we interviewed participants after their use of the device, “expectation” seemed to be an inaccurate term in this case. Therefore, we decided to use the terms “performance” and “effort or ease of use” instead of performance expectancy and effort expectancy, respectively.

Participants

Interviews were conducted separately for each participant group until data saturation was reached. To ensure a diverse sample, we conducted at least 1 interview per patient group and research center. When new interviews in a patient group did not lead to new valuable insights, we assumed data saturation. This approach led to different sample sizes. A total of 61 interviews were conducted. As only 1 patient with systemic lupus erythematosus could be interviewed and therefore data saturation is not clear, this interview was excluded from analyses.

Ethical Considerations

Ethics approval was granted by the ethics committee of Kiel University, Germany (Ethikkommission der Medizinischen Fakultät der CAU zu Kiel, approval D491/20) and the Health Research Authority and Health and Care Research Wales (IRAS project ID 282329, REC reference 20/PR/0185). Participants gave written and verbal informed consent before taking part in the study. Interviews were voluntary for participants of the larger FS. To ensure participant privacy, participants were assigned a random code. The code key is stored, encrypted, and password-protected. Interview transcripts were pseudonymized. Voice and video recordings are stored separately in accordance with the General Data Protection Regulation.

Results

Overview

Table 1 provides an overview of the number of interviews per group and site, while Table 2 shows the demographic characteristics of the sample.

XSL•FO
RenderX

Table 1. Number of interviews per group and location.

Location	Healthy participant, n (%)	PD[a], n (%)	HD[b], n (%)	SLE[c], n (%)	PSS[d], n (%)	RA[e], n (%)	IBD[f], n (%)
Kiel	11 (18.03)	12 (19.67)	—[g]	1 (1.64)	—	5 (8.20)	—
Münster	6 (9.84)	—	4(6.56)	—	—	—	—
Newcastle	2 (3.28)	—	—	—	11 (18.03)	1 (1.64)	—
Rotterdam	2 (3.28)	—	—	—	—	—	6 (9.84)

[a]PD: Parkinson disease.
[b]HD: Huntington disease.
[c]SLE: systemic lupus erythematosus.
[d]PSS: primary Sjögren syndrome.
[e]RA: rheumatoid arthritis.
[f]IBD: inflammatory bowel disease.
[g]Not applicable.

Table 2. Demographic data of the interview participants.

Group	Participants, n	Age (years), range	Sex		Living situation	
			Male, n (%)	Female, n (%)	Alone, n (%)	Not alone, n (%)
Healthy participant	21 (34.43)	21-77	12 (46.15)	9 (25.71)	7 (53.85)	14 (29.17)
PD[a]	12 (19.67)	37-80	8 (30.77)	4 (11.43)	1 (7.69)	11 (22.92)
HD[b]	4 (6.56)	29-59	2 (7.69)	2 (5.71)	1 (7.69)	3 (6.25)
PSS[c]	11 (18.03)	52-82	2 (7.69)	9 (25.71)	2 (15.38)	9 (18.75)
SLE[d]	1 (1.64)	53-53	0 (0)	1 (2.86)	0 (0)	1 (2.08)
RA[e]	6 (9.84)	40-68	0 (0)	6 (17.14)	1 (7.69)	5 (10.42)
IBD[f]	6 (9.84)	22-53	2 (7.69)	4 (11.43)	1 (7.69)	5 (10.42)
Total	61 (100)	21-82	26 (100)	35 (100)	13 (100)	48 (100)

[a]PD: Parkinson disease.
[b]HD: Huntington disease.
[c]PSS: primary Sjögren syndrome.
[d]SLE: systemic lupus erythematosus.
[e]RA: rheumatoid arthritis.
[f]IBD: inflammatory bowel disease.

Figure 2 shows the general code system used for both apps. Some of the second- and third-level codes were only used for one of the apps. *Quality of the questions*, *quality of speech function and tests*, as well as *usability of the reminder function* and *reactions of peers* were only needed for the symptom tracking app, while *feedback* was only used for the cognitive test app.

Symptom Tracking App

Overview

Most participants used the symptom tracking app throughout the whole study period, though many participants mentioned missing several appointments. Three participants did not use the app at all. For one participant, it did not work, and some participants only got the notifications a few times for unknown technical reasons. One participant dropped out of the study due to perceived distress caused by the questions.

Performance

While most users rated the quality of the questions specially designed for the observational FS study as “fine,” there were some complaints. A few participants said that they had problems rating their experiences, and others criticized the questions as vague. A few users felt that there were not enough possibilities to answer, especially an option to answer “not at all” was missed. Some users felt that the questions did not fit their symptoms or experiences. In addition, a few participants were unsure about the meaning of some words: “I did not know, what Fatigue is. [...] And after I googled it, I was still unsure whether I had answered it correctly” (translated from German). One participant felt burdened by answering the questions:

> *At some point I said I didn’t want these questions anymore, I felt really bad about it. So for me, such questions don’t work, they also have to be positive. It’s a bit of a self-fulfilling prophecy, in that I always*

XSL•FO
RenderX

> *ask [myself] negative questions. So, I really found that I felt really, really bad afterwards.* [Translated from German]

Many participants found the questions boring or not well-suited to their situation, as they experienced little variation in their symptoms.

Some participants reported not liking the speech function researchers designed for the observational FS study, as they were unsure what to talk about or were embarrassed talking about their day: "I didn't know exactly how precise I was supposed to talk about that" (translated from German). Other participants reported feeling silly while performing the speech tests:

> *The only thing I didn't quite understand was that in the evening you either had to say "papapam" all the time or say "A" for a very long time and I didn't quite understand the point of that, but I'm sure there's a reason for that. [...] I just thought it was a bit silly.* [Translated from Dutch]

Some users pointed out that it was hard to stay motivated to answer the questions: "They were virtually the same questions. You just got fed up with them." On the other hand, some users reported that they enjoyed monitoring their own data:

> *Comparing the beginning of the day to the end of the day [...]. And then looking at how I slept the night before. So, yeah, I enjoyed doing that part.*

A few participants reported that they benefited from using the app, as it raised their awareness of their feelings and symptoms: "I found that very informative to me, to actually stop and think [...]. It was thought-provoking." Another participant said:

> *I actually didn't think the questions were bad at all, because you first had to deal with the question, 'How am I actually doing in this area today?' at that moment. And I also noticed how it changes during the day. When you are tired in the evening, you click on completely different values than in the morning when you have had a good night's sleep.* [Translated from German]

Effort or Ease of Use

Many participants agreed that answering the questions was quickly done, though some still perceived the effort as high. This was mainly due to the number of times it came as a part of study design or because they had to keep 2 phones with them: "If I were going out, having to take two phones out with me [...] was a bit of a nuisance."

While many users were comfortable with the apps' use frequency, a lot of participants found it too frequent. Many users were unable to respond consistently, and some described the timing of the app interactions as incompatible with their daily schedules: "It's a burden for the fixed time that's involved in those things. I've missed [answering the questions] a number of times, because you're in a meeting or whatever" (translated from Dutch), reported one participant.

Using the app was described as easy by most participants. Some users reported problems with opening the app, which could only be done using the notifications on the phone: "That is a bit hidden. You have to scroll down first. I actually didn't find it at first" (translated from German), reported one participant.

A few users reported problems with setting the time with a dial. Most reported not using the reminder function, though some found it quite useful. Some mentioned that remembering to answer the questions was bothersome. Others felt bothered by the reminder.

While most participants did not feel disturbed by using the app, some reported that it disrupted their schedules. "To a certain extent, it almost drove your day," reported one user. It was perceived as more disturbing at work, as there was not always time to use the app: "When I'm at work, in the car or on my bike, I can't just open the app and [answer the questions], so sometimes you get [...] a kind of guilty feeling, those kinds of negative feelings" (translated from Dutch).

There were some technical problems reported with the symptom tracking app. Some participants received the notifications not often enough or not for the whole period—not knowing why. Two users reported not getting any notifications at all. Others reported the questions to be in a mix of English and German: "The questions [were] only in English, it took a few days until we had it [in German]" (translated from German).

Social Influence

While no participants reported concerns regarding reactions by their peers, many said that they avoided answering questions when in company, running errands, or working: "When you are in the car or somewhere in a store, you can't just start taking your phone out" (translated from German). There were no reports about concerns regarding data privacy.

Facilitating Conditions

There were some insecurities mentioned regarding the symptom tracking app. As mentioned in the performance section, some users were unsure what to include during the speech function. Some participants were unsure how long after the scheduled time point they could still respond, which led them to miss some questions. A few suggestions to improve the study design and app use were to personalize the timing of the notifications, increase the time frame, or indicate how long the time frames were. Many participants would have preferred getting the questions less often. Others said that there should be the possibility to go back to correct their inputs. While some users said that the app could be used following the same schedule for a longer time, others said that they would not like that, or that it would only fit for people with more severe symptoms.

Cognitive Test App

Overview

Most participants used the cognitive test app, though some users stopped during the NBX. Others reported that they did not always have time for the tasks. One participant did not use the app.

XSL•FO
RenderX

Performance

Regarding the PVT and DSST, most participants felt positive and motivated. “I enjoyed those little quiz things,” reported one user. A few participants disliked the PVT and described it as boring: “The first one got a bit laborious because you’re doing the same thing all the time.” In contrast, the NBX was disliked by most participants. Many found it too hard, which led to frustration or annoyance. “I just said, I got two right, I thought they’d be thinking I’m the village idiot here. [...] The [NBX] was by far the most frustrating,” said one participant. They reported feeling unmotivated, as they did not see any improvement: “But to actually not feel as if [...] this was really achieving anything. You got to the stage sometimes where you just said, well, I’m just going to press a button.” Only a few participants reported liking this task, seeing it as challenging or fun.

Some participants commented that they would have liked feedback on their performance, which might have improved motivation: “If it would show afterwards whether you had gotten better or worse, or whether that was depending on the time of the day, that would have interested me” (translated from German).

A few participants reported feeling like they benefited from the tasks and could use them for training and getting better. Others reported not feeling any training effect: “I’m not under the impression that I got faster or more experienced [...]. Consistent, probably age-appropriate, I think” (translated from German).

Effort or Ease of Use

Most participants agreed that it was not too much effort, though some said that the tutorial for new tasks was too long. While most participants agreed that the frequency of using the app chosen for the study was appropriate, some reported to have missed a few tests. Other users mentioned that the times were not fitting to their schedules.

Participants generally found the app easy to use. They liked the design as a chat: “It’s a nice way of being greeted: ‘Hello, are you ready for your next exam,’ and concentrate, and off you go. That’s all you have to do,” said one user.

Some participants reported disruptions to their daily routines, as they sometimes had difficulty fitting the app’s scheduled assessments into their workday, especially when tasks required concentration: “You really have to go into the corner, because during the day, when you’re at work, you can’t really do it” (translated from German).

Some technical problems with the app were mentioned. For a few participants, the notifications did not come regularly twice a day, but only once or only at the beginning of their trial period. In other cases, not all tasks were available. A few users had problems solving the tasks because another app notification covered a part of their screen.

Social Influence

No participant reported having concerns regarding data privacy. One user stated:

> *That sort of thing crosses your mind but obviously, the paperwork you’re given beforehand tells you how it’s going to be stored, who will have access, whether it’s being anonymized. [...] If I participate in some sort of medical research then [...] I expect it to be absolutely fine.*

Nobody mentioned any reaction of peers.

Facilitating Conditions

Some participants reported feeling insecure whether they understood the instructions for the third task correctly. While some participants found the instructions sufficient, others felt that the instructions for the last task were vague. Regarding optimization, a few suggestions were made. Some participants suggested to make the NBX easier. Others reported wanting more variety in the tasks: “I would like to have more variety. So, maybe today like this and tomorrow like this, or maybe sometimes more difficult and sometimes easier, so that you are a bit [more] motivated” (translated from German), suggested one participant. While some study participants did not want to use the cognitive test app longer following this study-specific schedule, others reported that they would have liked to do more tasks.

Discussion

Principal Findings

This study considers the technology acceptance of 2 different apps in the context of the IDEA-FAST FS. We interviewed participants who used a symptom and activity app as well as a cognitive test app over 4 weeks. Using the UTAUT [13] model, we identified different themes of technology acceptance in the use of these apps.

For the users, performance played a less crucial role in technology acceptance, whereas the aspect of effort gained importance. Both facilitating conditions as well as social influence seemed to play minor roles in the acceptance of the apps.

Performance

When users expect an mHealth app to be helpful to them, they show more willingness to use it [1,6,16]. This perceived usefulness has been shown to be a good predictor for the intention to use mHealth apps [17,18,21,23]. In this study, the apps did not provide an apparent impact on the participants’ health. Therefore, motivation and adherence to the schedule were issues with both apps. Some participants did not see value in answering the questions of the symptom tracking app and saw it as a burden, while others enjoyed the opportunity to track their behavior and symptoms. The cognitive test app varied in popularity among participants. Some users found the PVT boring, but it was well liked by most, as was the DSST. However, many participants disliked the NBX, found it too hard or frustrating. This led to compliance problems, such as randomly clicking through the test, ignoring or missing scheduled sessions, or ceasing to answer it altogether. A possible reason for this could be that the task was developed for healthy users. The cognitive impairments of participants with Parkinson

XSL•FO
RenderX

disease and Huntington disease, along with the relatively high average age of the sample, contributed to the task being perceived as too difficult. As a result, the task was adjusted during the FS. However, since the adjusted NBX was still perceived as too difficult, it was not included in the subsequent clinical observation study.

An important aspect of motivation for participants seemed to be feedback on their data. While not many participants reported interest in the results from the symptom tracking app, those who enjoyed the app used it as a tool to track their behavior and symptoms, especially fatigue. Possibly, these were patients who were concerned about their fatigue before joining the study. In contrast, many participants voiced a wish for feedback from the cognitive tests. Feedback could help increase motivation if participants were able to see their performance over time. If an mHealth app itself does not seem inherently helpful to the user, feedback might serve as a useful way to engage them. At the same time, feedback could be considered as an intervention, which researchers might wish to avoid. Feedback can be a helpful or interesting resource users can gain from their app use.

Effort or Ease of Use

People tend to use mHealth apps they perceive as easy to understand and to use. This effect has been shown in previous studies [1,6,16,18] and can be observed here as well. While users agreed that the apps did not take too much time, the frequency of use required by the study design was more problematic. Especially having to use the apps at specific time points was seen as disruptive and led to missing data. The 2 apps described in this study had to be used several times per day. Even though the sessions were short, these numerous sessions meant a considerable amount of extra effort, especially since the participants had to adhere to certain time constraints. It was not surprising that almost all participants reported they had not been able to attend all sessions, especially those who were more active or working. Participants suggested adjustable time frames as a solution to account for differences in daily routine or to find individual times they would be able to handle the apps, which should be considered by the clinical team to find a compromise between the patient's needs and study goals. When choosing an app for a scientific study, it is beneficial if it is available for all common smartphone operating systems. Often, participants prefer using their own phone rather than a study phone. Study phones constitute an additional device participants have to familiarize themselves with, charge regularly, carry around, and keep in mind to use. They tend to be forgotten or ignored more easily.

Social Influence and Facilitating Conditions

Both social influence and facilitating conditions did not seem to have significant impact on the acceptance of the apps in this study. An explanation could be the study context. As participants were using the apps without expected impact on their health, it is not surprising that family members and peers did not have specific expectations or motivating sentiments regarding app use. Therefore, opinions of peers seem to have had only a slight impact on their willingness to use the apps. This finding is not entirely inconsistent with the literature, as some studies have found no influence of social influence on the intention to use a technology [20,31]. Other studies have shown that opinions of peers influence the decision to use an mHealth app [21,23]. In the context of study-related apps, social influence may affect the decision to participate in the study rather than the use of the app itself. Another interesting finding was that participants did not voice major concerns regarding data privacy. Data privacy and security have previously been reported as a valid concern for users of mHealth apps [10,16,25,32], though others did not find that privacy concerns predicted intention to use mHealth apps [17,18,21]. As the symptom tracking app gathers quite personal information, concerns regarding data privacy had to be expected. When asked about data privacy, many participants cited trust in the study team and medical data security standards as reasons to feel safe.

Regarding facilitating conditions, salient factors such as costs for the use of mHealth apps or availability [16,23,33] were not an issue in this case. Both apps as well as study smartphones were provided for the participants. It seems that the provided support in the form of instructions and the study team was sufficient for most participants to feel confident about the correct use of the apps. There were some uncertainties concerning the symptom tracking app, like not knowing what to report using the speech function and not knowing how much time there was for answering the questions. Facilitating conditions have been found to be very influential on the intention of using a technology in several studies [1,33], while others did not find a connection [6]. It seems probable that due to the high amount of support and lack of limiting factors, the positive impact of facilitating conditions on the use of the apps was present. This assumption is supported by the fact that many participants stressed the good support during the study and voiced their appreciation for the study team.

Some aspects of these apps were controversial among participants. While many participants disliked the NBX of the cognitive test app, some voiced their appreciation for feeling challenged by the task. Many participants described the frequent documentation of their well-being as tedious, although some appreciated the possibility to track their symptoms. While studying the opinions and experiences of users is a valuable tool to ensure acceptance, users have different desires and needs. In choosing suitable mHealth apps, one size will most certainly not fit all. While some aspects of these apps are necessary for study success, such as the documentation in the symptom tracking app, others might be customizable. An example is the reminder function of the symptom tracking app, where users can decide for themselves if they want to be reminded or not.

Limitations and Strengths

Limitations and strengths of this study align with those previously described [27]. In short, due to the design of the study, participants used multiple devices as well as the mHealth apps at the same time, which might have influenced their opinions. Due to the different number of interviews in each group, disease-specific aspects might not have been entirely captured. In addition, the COVID-19 pandemic might have had an influence on the acceptance of the 2 apps, since participants were partly living in lockdown or had altered their normal

XSL•FO
RenderX

routines. We are also aware that these qualitative interviews primarily reflect app modifications made specifically for this FS, including support provided by the clinical team. Thus, the findings cannot be generalized to the assessment of app use in different contexts. The results represent the experiences of the FS participants with the apps implemented in a specific study with its own design choices. The requirements of apps can be different, depending on their context of use.

This study represents the opinions and voices of a large and diverse sample on the acceptance of 2 mHealth apps with different aims and uses. This enables us to draw conclusions, which might also be useful for other studies planning to use mHealth apps, as well as help in the development of new apps. Conducting these interviews during the duration of the FS helped us to improve and select suitable apps for use in the clinical observation study. As described earlier, the NBX of the cognitive test app was adjusted to fit better to the study population. Its use was discontinued for the clinical observation study. Our experience underlines the importance of involving users and patients in the development and design of mHealth apps in order to achieve a user-friendly and meaningful result [11].

Recommendations

The choice of appropriate mHealth apps for a research project can be a challenging decision to make. The efficacy of any data collection app is contingent upon the extent to which its users actually use it. Consequently, when developing or selecting an mHealth app for a research project, it is essential to consider technology acceptance. The effort expended by the user should be considered as an integral component of the evaluation process. The ideal app is one that is straightforward and intuitive to use [24]. Patients will have divergent preferences regarding the impact of the study on their lives. Consequently, the integration of customizable features, where feasible, may enhance technology acceptance. These may encompass adjustable reminder functions, time points, and time frames for use of the app. The minimization of user effort is also relevant to the decision regarding the use of study phones. Many patients may prefer their personal smartphones as opposed to designated study phones. Therefore, ensuring that the app is available for all common smartphone operating systems might be beneficial [23]. The necessity of carrying 2 phones is thus eliminated, and the potential for patients to forget the phone at home and miss vital data is also minimized. Evidently, the time expended in using the app constitutes a significant proportion of the overall effort involved. Most cognitive testing and symptom tracking apps require patients to use them for a fixed period of time. Should this be required on a daily basis, or several times per day at designated times, it would necessitate a significant degree of effort. To minimize effort, it is vital to keep the required app time as short and infrequent as possible.

Considering the role of performance in technology acceptance, the pivotal question pertains to the extent to which the technology benefits the patient. The use of cognitive testing and symptom tracking apps within the context of research studies may not necessarily result in tangible health benefits for the patient. This is particularly relevant in cases where apps are used over an extended period of time. In such cases, consideration should be given to alternative factors of motivation. These elements may include gamification mechanisms or feedback on data in noninterventional studies. In the context of cognitive testing apps, the motivational factors of enjoyment and challenge can be leveraged. A task that is perceived as too easy may quickly become tedious to the user, while those that are too challenging may result in demotivation over time. The process of identifying an appropriate level of difficulty for the target population can be a difficult yet rewarding endeavor.

The role of social influence in the technology acceptance of mHealth apps in the context of medical studies may be less significant. It is possible that the opinion of significant others may influence patients' decisions regarding participation in the study, thereby reducing the impact of social influence during the actual use of the apps. Data security could be a cause of concern for patients. The provision of information regarding data security measures during the recruitment process is therefore a valuable use of time [25].

Some salient factors that are typically considered facilitating conditions of technology acceptance, such as costs or availability [23], are typically negligible in the context of medical studies. Nevertheless, the pivotal role that technical and organizational support play in facilitating patient compliance has to be taken into account. The provision of adequate instructions and support should be considered a prerequisite [21]. The establishment of positive relations between patients and the study team can be a significant motivating factor, as well as enhancing satisfaction with the study itself.

Conclusions

This study describes the technology acceptance of 2 mHealth apps. The primary objective of this study was to assess the usability of these apps in a follow-up investigation. As part of a larger evaluation study, potential issues were identified. They include demotivation due to the level of difficulty experienced with one of the cognitive tasks and challenges with adhering to the stipulated time schedule for the symptom tracking app. These issues were addressed through a variety of approaches. For instance, the challenging task was set aside for use in the follow-up. To support adherence, time points were limited to those essential for the study's success while allowing as much flexibility as possible. The second objective of this study was to assess the specific challenges researchers encounter when selecting suitable mHealth apps for their own research, and what characteristics an app must possess to be considered a useful tool for health research. The incorporation of patient perspectives in the device selection process has been demonstrated to be beneficial and valuable. The findings of this study provide a valuable contribution to the field of medical research by offering insights that may inform the selection of appropriate mHealth apps. The study emphasizes the importance of effective communication and testing of all materials used in conjunction with patients and intended users prior to the commencement of research studies.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank the participants for their time and contribution to the feasibility study and the interviews. In addition, the authors would like to thank the IDEA-FAST (Identifying Digital Endpoints to Assess Fatigue, Sleep and Activities of Daily Living in Neurodegenerative Disorders and Immune-Mediated Inflammatory Diseases) consortium, consisting of Newcastle University, University Medical Center Schleswig-Holstein, University of Brescia, Erasmus University Medical Center Rotterdam, University of Glasgow, University of Limerick, ECRIN European Clinical Research Infrastructure Network, Queen Mary University of London, Imperial College London, Byteflies NV, DREEM, University of Cambridge, Lixoft SAS, Asociación Parkinson Madrid, Stichting MLC Foundation, TMF eV, Medibiosense Ltd, empirica GmbH, FCiências.ID, Pluribus One Srl, Instituto de Medicina Molecular Joao Lobo Antunes, Teknologian tutkimuskeskus VTT Oy, Cambridge Cognition Ltd, Universidad Autonoma de Madrid, Institut Mines-Télécom, McRoberts BV, George-Huntington-Institut GmbH, Instytut Psychiatrii Neurologii, Medical University Innsbruck, Helse Stanvanger Hf, iXscient Ltd, European Federation of Crohn's and Ulcerative Colitis Association, Leids University Medical Center, University of Manchester, Janssen Pharmaceutica NV, Takeda Development Centre Europe Ltd, AbbVie Inc, AstraZeneca AB, Eli Lilly & Co Ltd, CHDI Foundation, Parkinson's Disease Society of the United Kingdom, Pfizer Ltd, F. Hoffmann-La Roche AG, Sanofi Aventis Recherche et Développement, UCB Biopharma SPRL, Biogen IDEC Ltd, Orion OYJ, and Let it Care.

Funding

The IDEA-FAST (Identifying Digital Endpoints to Assess Fatigue, Sleep and Activities of Daily Living in Neurodegenerative Disorders and Immune-Mediated Inflammatory Diseases) project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (grant 853981). The Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program, European Federation of Pharmaceutical Industries and Associations, and Parkinson's Disease Society of the United Kingdom LBG. This communication reflects the view of the authors, and neither the Innovative Medicines Initiative nor the European Union and European Federation of Pharmaceutical Industries and Associations is liable for any use that may be made of the information contained herein.

Authors' Contributions

WFN, WM, JG, and HK conceived the study and were involved in protocol development with KE. WM, WFN, JG, and KE were involved in gaining ethics approval. JG, EW, KD, VM, SA, RR, and JD were responsible for patient recruitment. Interviews were conducted by JG, EW, HK, and VM. Data analysis and the first draft of the manuscript were completed by JG. A detailed revision was done by JD. All authors reviewed and edited the manuscript and approved the submission version of the manuscript.

Conflicts of Interest

RR is the founding director and owner of the George-Huntington-Institute, a private research institute focused on clinical and preclinical research in Huntington disease, and QuantiMedis, a clinical research organization providing quantitative motor services in clinical trials and research. RR served as an elected member of the Steering Committees of the European Huntington Disease Network and the Huntington Study Group, cochair of the Task Force on Huntington's disease, and member of the Task Force on Technology of the International Parkinson and Movement Disorder Society. He has provided consulting services, advisory board functions, clinical trial services, quantitative motor analyses, and/or lectures for Actelion, Alnylam, Amarin, Annexon, AOP Orphan Pharmaceuticals, AskBio, Cure Huntington Disease Initiative Foundation, Desitin, Hoffmann-La Roche, IONIS, Ipsen, Lundbeck, MEDA Pharma, Medivation, Mitoconix, Neurosearch, Novartis, Omeros, Pfizer, Prana Biotechnology, Prilenia, PTC Therapeutics, Raptor, Siena Biotech, Temmler Pharma, Teva, uniQure, Vaccinex, Voyager, Wave Life Sciences, and Zevra. He has received grant support from the Bundesministerium für Bildung und Forschung, the Cure Huntington Disease Initiative Foundation, the Deutsche Forschungsgemeinschaft, the Deutsches Zentrum für Neurodegeneration und Entzündung, the European Union's 7th Framework Program, Horizon 2020 Innovative Medicine Initiative 2 program, the European Huntington Disease Network, the High-Q-Foundation, the National Institute of Health, and the National Science Foundation. NVM is an employee of Johnson & Johnson and may hold company stocks or stock options. WFN provided consultations for Asahi Kasei Pharma, Amgen, Argenx, Bristol Myers Squibb, EQT, IQVIA, Johnsons & Johnsons Innovations, Novartis, Quotients, Resolves Therapeutics, Sanofi & Veloxis.

References

1. Semiz BB, Semiz T. Examining consumer use of mobile health applications by the extended UTAUT model. bmij 2021;9(1):267-281. [doi: 10.15295/bmij.v9i1.1773]
2. From innovation to implementation: eHealth in the WHO European region. World Health Organization. 2016. URL: https://www.who.int/europe/publications/i/item/9789289051378 [accessed 2025-12-30]
3. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]

XSL•FO
RenderX

4. Moore K, O'Shea E, Kenny L, Barton J, Tedesco S, Sica M, et al. Older adults' experiences with using wearable devices: qualitative systematic review and meta-synthesis. JMIR Mhealth Uhealth 2021;9(6):e23832 [FREE Full text] [doi: 10.2196/23832] [Medline: 34081020]
5. de Korte EM, Wiezer N, Janssen JH, Vink P, Kraaij W. Evaluating an mHealth app for health and well-being at work: mixed-method qualitative study. JMIR Mhealth Uhealth 2018;6(3):e72 [FREE Full text] [doi: 10.2196/mhealth.6335] [Medline: 29592846]
6. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: an extension of the UTAUT model. Int J Med Inform 2017;101:75-84. [doi: 10.1016/j.ijmedinf.2017.02.002] [Medline: 28347450]
7. Cucciniello M, Petracca F, Ciani O, Tarricone R. Development features and study characteristics of mobile health apps in the management of chronic conditions: a systematic review of randomised trials. NPJ Digit Med 2021;4(1):144 [FREE Full text] [doi: 10.1038/s41746-021-00517-1] [Medline: 34611287]
8. Robbins R, Krebs P, Jagannathan R, Jean-Louis G, Duncan DT. Health app use among US mobile phone users: analysis of trends by chronic disease status. JMIR Mhealth Uhealth 2017;5(12):e197 [FREE Full text] [doi: 10.2196/mhealth.7832] [Medline: 29258981]
9. Radbron E, Wilson V, McCance T, Middleton R. The use of data collected from mHealth apps to inform evidence-based quality improvement: an integrative review. Worldviews Evid Based Nurs 2019;16(1):70-77. [doi: 10.1111/wvn.12343] [Medline: 30758133]
10. Abdekhoda M, Ahmadi M, Dehnad A, Hosseini AF. Information technology acceptance in health information management. Methods Inf Med 2014;53(1):14-20. [doi: 10.3414/ME13-01-0079] [Medline: 24248003]
11. Holthe T, Halvorsrud L, Karterud D, Hoel K, Lund A. Usability and acceptability of technology for community-dwelling older adults with mild cognitive impairment and dementia: a systematic literature review. Clin Interv Aging 2018;13:863-886 [FREE Full text] [doi: 10.2147/CIA.S154717] [Medline: 29765211]
12. Davis F. A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results [dissertation]. Cambridge, MA: Massachusetts Institute of Technology; 1985.
13. Venkatesh V, Morris M, Davis G. User acceptance of information technology: toward a unified view. MIS Q 2003;27:478. [doi: 10.2307/30036540]
14. Davis F. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989;13(3):319-340. [doi: 10.2307/249008]
15. Wills M, El-Gayar O, Bennet D. Examining healthcare professionals' acceptance of electronic medical records using UTAUT. Issues Inf Syst 2008;9:396-401. [doi: 10.48009/2_iis_2008_396-401]
16. Petersen F, Jacobs MS, Pather S. Barriers for user acceptance of mobile health applications for diabetic patients: applying the UTAUT model. 2022 Presented at: Responsible Design, Implementation and Use of Information and Communication Technology; April 6-8, 2020; Skukuza, South Africa p. 61-72. [doi: 10.1007/978-3-030-45002-1_6]
17. Breil B, Salewski C, Apolinário-Hagen J. Comparing the acceptance of mobile hypertension apps for disease management among patients versus clinical use among physicians: cross-sectional survey. JMIR Cardio 2022;6(1):e31617 [FREE Full text] [doi: 10.2196/31617] [Medline: 34989683]
18. Mouloudj K, Bouarar AC, Asanza DM, Saadaoui L, Mouloudj S, Njoku AU, et al. Factors influencing the adoption of digital health apps: an extended technology acceptance model (TAM). In: Integrating Digital Health Strategies for Effective Administration. Hershey: IGI Global Scientific Publishing; 2023:116-132.
19. Qu S, Zhou M, Kong N, Campy KS. Factors influencing user acceptance of weight management apps among Chinese obese individuals during the COVID-19 pandemic. Health Policy Technol 2023 Jun;12(2):100758 [FREE Full text] [doi: 10.1016/j.hlpt.2023.100758] [Medline: 37168934]
20. Bäuerle A, Frewer A, Rentrop V, Schüren LC, Niedergethmann M, Lortz J, et al. Determinants of acceptance of weight management applications in overweight and obese individuals: using an extended unified theory of acceptance and use of technology model. Nutrients 2022 May 08;14(9):1968 [FREE Full text] [doi: 10.3390/nu14091968] [Medline: 35565935]
21. Zhang Y, Liu C, Luo S, Xie Y, Liu F, Li X, et al. Factors influencing patients' intentions to use diabetes management apps based on an extended unified theory of acceptance and use of technology model: web-based survey. J Med Internet Res 2019;21(8):e15023 [FREE Full text] [doi: 10.2196/15023] [Medline: 31411146]
22. Schretzlmaier P, Hecker A, Ammenwerth E. Suitability of the unified theory of acceptance and use of technology 2 model for predicting mHealth acceptance using diabetes as an example: qualitative methods triangulation study. JMIR Hum Factors 2022;9(1):e34918 [FREE Full text] [doi: 10.2196/34918] [Medline: 35262493]
23. Chopdar PK. Adoption of Covid-19 contact tracing app by extending UTAUT theory: perceived disease threat as moderator. Health Policy Technol 2022;11(3):100651 [FREE Full text] [doi: 10.1016/j.hlpt.2022.100651] [Medline: 35855013]
24. Asadollahi F, Zagami SE, Roudsari RL. Understanding pregnant women's intention to use mobile health apps and its determinants: applying the UTAUT model in a mixed-methods study. BMC Prim Care 2025;26(1):252. [doi: 10.1186/s12875-025-02950-7] [Medline: 40790558]
25. Alsyouf A, Lutfi A, Alsubahi N, Alhazmi FN, Al-Mugheed K, Anshasi RJ, et al. The use of a technology acceptance model (TAM) to predict patients' usage of a personal health record system: the role of security, privacy, and usability. Int J Environ Res Public Health 2023;20(2):1347 [FREE Full text] [doi: 10.3390/ijerph20021347] [Medline: 36674105]

XSL•FO
RenderX

26. Vildjiounaite E, Kallio J, Kyllönen V, Nieminen M, Määttänen I, Lindholm M, et al. Unobtrusive stress detection on the basis of smartphone usage data. Pers Ubiquit Comput 2018;22:671-688 [FREE Full text] [doi: 10.1007/s00779-017-1108-z]
27. Graeber J, Warmerdam E, Aufenberg S, Bull C, Davies K, Dixon J, et al. Technology acceptance of digital devices for home use: qualitative results of a mixed methods study. Digit Health 2023;9:20552076231181239 [FREE Full text] [doi: 10.1177/20552076231181239] [Medline: 37361435]
28. Cognitive function. Cambridge Cognition. 2023. URL: https://cambridgecognition.com/cognitive-function/ [accessed 2023-07-10]
29. McCue M, Blair C, Fehnert B, King J, Cormack F, Sarkey S, et al. Mobile app to enhance patient activation and patient-provider communication in major depressive disorder management: collaborative, randomized controlled pilot study. JMIR Form Res 2022;6(10):e34923 [FREE Full text] [doi: 10.2196/34923] [Medline: 36301599]
30. Mayring P. Qualitative Inhaltsanalyse: Grundlagen und Techniken. 12th Edition. Weinheim, Basel: Beltz; 2015.
31. Fan W, Liu J, Zhu S, Pardalos PM. Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS). Ann Oper Res 2020;294:567-592 [FREE Full text] [doi: 10.1007/s10479-018-2818-y]
32. Bastoni S, Wrede C, da Silva MC, Sanderman R, Gaggioli A, Braakman-Jansen A, et al. Factors influencing implementation of eHealth technologies to support informal dementia care: umbrella review. JMIR Aging 2021;4(4):e30841 [FREE Full text] [doi: 10.2196/30841] [Medline: 34623314]
33. de Veer AJE, Peeters JM, Brabers AEM, Schellevis FG, Rademakers JJDJM, Francke AL. Determinants of the intention to use e-Health by community dwelling older people. BMC Health Serv Res 2015;15:103 [FREE Full text] [doi: 10.1186/s12913-015-0765-8] [Medline: 25889884]

Abbreviations

DSST: digit symbol substitution test
FS: feasibility study
IDEA-FAST: Identifying Digital Endpoints to Assess Fatigue, Sleep and Activities of Daily Living in Neurodegenerative Disorders and Immune-Mediated Inflammatory Diseases
mHealth: mobile health
NBX: n-back task
PVT: psychomotor vigilance task
UTAUT: unified theory of acceptance and use of technology

Edited by N von Kalckreuth; submitted 04.Jan.2025; peer-reviewed by H Fetrati, K Mouloudj; comments to author 01.Aug.2025; accepted 29.Nov.2025; published 29.Apr.2026.

Please cite as:
Graeber J, Warmerdam E, Aufenberg S, Bull C, Davies K, Döhring J, Emmert K, Judd C, Maetzler C, Manyakov NV, Reilmann R, Ng WF, Macrae V, Maetzler W, Kaduszkiewicz H
Technology Acceptance of 2 mHealth Apps During an Observational Technology Evaluation Study: Qualitative Results of a Mixed Methods Study
JMIR Hum Factors 2026;13:e70873
URL: https://humanfactors.jmir.org/2026/1/e70873
doi:10.2196/70873

©Johanna Graeber, Elke Warmerdam, Svenja Aufenberg, Christopher Bull, Kristen Davies, Juliane Döhring, Kirsten Emmert, Claire Judd, Corina Maetzler, Nikolay V Manyakov, Ralf Reilmann, Wan-Fai Ng, Victoria Macrae, Walter Maetzler, Hanna Kaduszkiewicz. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 29.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Smartphone App to Improve Intuitive Eating and Diet Quality: Design and Usability Study

Mandy Korpusik[1], PhD; Delaram YazdanSepas[1], PhD; Hawley C Almstedt[2], PhD, RDN

[1]Department of Computer Science, Seaver College of Science and Engineering, Loyola Marymount University, Los Angeles, CA, United States

[2]Department of Health and Human Sciences, Seaver College of Science and Engineering, Loyola Marymount University, Los Angeles, CA, United States

Corresponding Author:
Hawley C Almstedt, PhD, RDN
Department of Health and Human Sciences
Seaver College of Science and Engineering
Loyola Marymount University
1 LMU Drive, MS 8888
Los Angeles, CA, 90045
United States
Phone: 1 3103381925
Fax: 1 3105686758
Email: hawley.almstedt@lmu.edu

Abstract

Background: Web-based and mobile phone–based apps have become widely available for dietary self-monitoring; however, their use may increase the risk of disordered eating. College students frequently demonstrate poor nutrient intake despite consumption of sufficient calories. One way to improve diet quality may be via the use of a smartphone app that encourages intuitive eating.

Objective: The purpose of this study was to improve diet quality among college students through the use of a novel smartphone app that promotes intuitive eating rather than calorie counting and weight loss.

Methods: The In2Eat iOS mobile app was developed in SwiftUI and stored user data in a Firebase database. A total of 45 college students completed assessments of intuitive eating, diet quality, and disordered eating before and after 4 weeks of using the In2Eat app. Users evaluated the usability of the app with the System Usability Scale (SUS). Engagement with the app was recorded as the total number of days a meal was logged, the total number of meals logged, and the average number of meals logged per day.

Results: After our 4-week intervention, dietary qualities that protect against chronic disease increased by 28%, fruit consumption increased by 63%, and skin antioxidant levels increased by 6.1% (Hedges g=0.16; mean difference 0.33, 95% bias corrected and accelerated [BCa] CI 0.04-0.61; P=.03). Global intuitive eating did not change during the user study; however, the unconditional permission to eat subscale increased (Hedges g=−0.28; mean difference 0.28, 95% BCa CI 0.07-0.49; P=.01, adjusted P=.07). Overall, disordered eating also did not change with app use, although dietary restraint decreased (Hedges g=−0.23; mean difference 0.30, 95% BCa CI −0.61 to −0.04; P=.04, adjusted P=.22). The average SUS score for the In2Eat app was 67.2 (SD 15.5). The number of days a meal was logged was positively correlated with SUS scores (r=0.28; P=.06), and the total number of meals logged had a monotonic association with app usability (ρ=0.31; P=.04). When divided according to the low (mean 10.2, SD 5.3), medium (mean 26.3, SD 2.8), and high (mean 33.6, SD 3.8) number of days logging meals, participants with higher days of logging reported the app as more usable (H=6.75; P=.03). A regression analysis showed that 8% of the variance in system usability (R^2=0.080; P=.31) was explained by app use; however, none of the individual predictors contributed substantially to the variance.

Conclusions: An intuitive eating smartphone app can improve diet quality without increasing disordered eating. Results suggest that participants who logged more meals tended to rate the app as more usable. Further research is needed with a greater sample size after incorporating features to improve the app's usability.

(JMIR Hum Factors 2026;13:e81439) doi:10.2196/81439

XSL•FO
RenderX

KEYWORDS

disordered eating; nutrition; college age; intuitive eating; mobile app; natural language processing; recommendations; recipes; physical activity

Introduction

Web-based and mobile phone–based apps have become widely available for dietary self-monitoring, with some incorporating recent technological advances, such as speech recognition for logging meals with voice and computer vision for food image understanding to reduce user burden [1-3]. In our prior work, we developed an iOS app called COCO Nutritionist that used natural spoken language understanding and machine learning to automate mapping of self-reported meal logs to the US Department of Agriculture (USDA) food codes and portion sizes, and we demonstrated in a pilot study that there was no significant difference in 3-day reported energy intake between our method and gold-standard 24-hour recalls [4].

Previous research shows that lifestyle habits developed as a college-aged young adult have a lasting influence on physical health, mental health, and general well-being [5,6]. Despite this, there is ample literature demonstrating inadequate nutrient intake among college students [7,8]. Malnutrition exists when there is a lack of calorie intake (ie, undernutrition) and when dietary intake is excessive. Of particular concern is a high intake of calories lacking in nutritional quality, increasing risk for chronic disease. Vitamin and mineral deficiencies can occur even when the diet is abundant in calories; hence, the need to assess diet quality, not just calorie intake [9]. Diet quality assesses the extent to which a diet has sufficient health-promoting foods to avoid vitamin and mineral deficiencies and protect against chronic disease while avoiding overconsumption of dietary factors that increase risk for chronic disease.

One way to improve diet quality may be via the use of a smartphone app that encourages healthy eating. However, research suggests that use of smartphone apps that track calorie intake is related to disordered eating symptoms [10,11]. Repetitive monitoring of calorie intake may lead to obsessive behaviors, anxious feelings, and unhealthy dietary restrictions. Intuitive eating is a conceptual eating behavior that emphasizes the use of internal cues for hunger and satiety to guide food choices [12]. Preoccupation with food, restriction, and rigidity are commonly observed with disordered eating. However, an intuitive approach to eating encourages achieving satisfaction without overeating by using interoceptive awareness while allowing unconditional permission to eat. The 10 principles of intuitive eating include rejecting diet culture, honoring hunger, making peace with food, discovering satisfaction, feeling satiety, challenging diet “rules,” coping with emotions, respecting bodies, joyful movement, and honoring health [12]. Research has found that the use of intuitive eating principles may counteract disordered eating behaviors [13-16].

The purpose of this study was to improve diet quality among college students via a smartphone app designed to encourage intuitive eating and recommend personalized healthy recipes. We hypothesized that consistent use of our smartphone app over 4 weeks would improve diet quality without increasing disordered eating.

Methods

This paper describes the design, development, and evaluation of, to our knowledge, the first mobile app for automated self-report of dietary intake via natural spoken language that promotes intuitive eating principles.

App Development

The In2Eat iOS mobile app was developed by a team of undergraduate computer scientists at Loyola Marymount University. In the preliminary design phase, 2 other apps were found to promote intuitive eating—MyTummy and MEAL. However, neither of these apps supported the kind of self-reported dietary intake and detailed nutrient information that was common among popular diet-tracking apps such as MyFitnessPal, Noom, and Weight Watchers. In2Eat is unique in that it both supports low-burden diet logging and promotes intuitive eating. Diet tracking is required for improving diet quality, which is our goal and is associated with health and disease prevention. However, our intuitive eating approach aims to avoid fostering disordered eating that can occur with apps that emphasize weight loss and calorie counting [10,11].

In2Eat’s natural language processing technology for easily logging meals with voice or text was originally developed for the COCO Nutritionist research prototype [4]. COCO used machine learning, specifically convolutional neural networks (CNNs), to accomplish 2 tasks: parsing a user’s natural language meal description into foods and quantities and mapping the user’s food descriptions to a matching entity in the validated USDA food database using learned embeddings. The CNN automatically learned semantic vector representations, or embeddings, of the user’s unstructured natural language meal description through a binary classification task where the model learned to detect which pairs of USDA foods and meal descriptions were a match. With In2Eat, the 2-step approach of parsing a user’s meal description, followed by database mapping with embeddings, is still the same, but the implementation now uses large language models (LLMs) for parsing meal descriptions and embeddings from LLMs instead of CNNs.

We built our In2Eat app in SwiftUI and stored user data in a Firebase database. A key component of our intuitive eating design is that the app does not display caloric information and instead focuses on diet quality by reporting consumed macro and micronutrients. In addition, there is a daily push notification with an intuitive eating quote of the day, which is drawn from intuitive eating content that we curated. This intuitive eating quote of the day is also displayed on the main dashboard screen upon opening the app. In2Eat supports food logging via written or spoken natural language, along with a hunger satiety check-in at each meal. Several principles of intuitive eating can be fostered by increasing awareness of feelings of hunger and

XSL•FO
RenderX

satiety before or after eating occasions. Thus, experts coach intuitive eating by using a scale of 0 to 10 to represent primal or painful hunger (as 0) and painfully full or stuffed (as 10) [12]. The macronutrients are shown in a pie chart on the dashboard, and the micronutrients are shown as green, yellow, or red progress bars toward reaching the recommended amount based on the user's sex and weight. The recommended targets were based on the Dietary Reference Intakes established by the Institute of Medicine [17]. In2Eat's recipe book recommends healthy recipes, defined as at least 5 g of fiber and no more than 5 g of saturated fat per serving that align with the user's self-reported dietary restrictions, when applicable. Users can save favorite recipes and search for specific recipes as well as filter by cooking and preparation time, cuisine, price, and allergens. Finally, there is a screen for logging the number of minutes of physical activity, without having to label the type of exercise, which aligns with the "joyful movement" principle of intuitive eating.

Recruitment

College-aged adults were recruited via flyers and classroom announcements at a primarily undergraduate institution in an urban setting. Advertising flyers were specifically placed on the walls in the on-campus food pantry, which is available for access 24 hours a day, 7 days per week. Inclusion criteria included being aged 18 years or older and having access to an iOS smart device.

A total of 47 volunteers enrolled in the study; however, 2 (4.3%) were lost to follow-up because they did not use the app and were unresponsive when attempting to schedule follow-up testing. Therefore, 45 (95.7%) participants were included in this analysis. At the time of enrollment, participants completed several online questionnaires and came into the laboratory for measurements of height, weight, and skin antioxidant levels. During the first laboratory visit, participants downloaded the In2Eat app onto their smart device and were given brief instructions on how to use the app. Volunteers were instructed to use the app for the next 4 weeks. After 4 weeks of app use, participants repeated the online questionnaires, as well as height, weight, and skin antioxidant level assessments. Incentives of US $50 grocery store gift cards were provided to participants who completed the baseline and follow-up testing sessions.

Measures

System Usability

The System Usability Scale (SUS) was used at follow-up to assess the app's usability on a scale from 0 to 100. The average SUS score is 68, with 68 through 80.3 considered good and greater than 80.3 considered excellent [18,19]. The instrument consists of 10 questions (eg, "I thought the app was easy to use"), each with 5 options ranging from strongly agree to strongly disagree.

App Use

To investigate the relationship between app engagement and perceived usability, we computed 3 use metrics: LogCount (total number of days a participant logged at least 1 meal), TotalMealsLogged (total number of meals logged throughout the study), and AvgMealsPerDay (mean number of meals logged per day, calculated as TotalMealsLogged divided by LogCount). These metrics capture both breadth and depth of engagement with the app.

Intuitive Eating

Intuitive eating was assessed using the Intuitive Eating Scale version 3 (IES-3) established for reliability and validity [20]. Possible results for the IES-3 range from 1 to 5 with higher scores indicating greater intuitive eating. Scoring the survey provides an overall global score for intuitive eating as well as scores on 4 subscales: unconditional permission to eat, reliance on hunger and satiety cues, eating for physical reasons, and body food choice congruence.

Disordered Eating and Mental Health

The Eating Disorder Examination Questionnaire (EDE-Q) was used to assess disordered eating patterns. This validated and internally consistent questionnaire provides a global disordered eating score and 4 subscales for dietary restraint, eating concern, body shape concern, and weight concern [21,22]. Dietary restraint reflects a tendency to consciously restrict or limit dietary intake. Eating, body shape, and weight concern are subscale scores that reflect an unhealthy preoccupation or worry with these features of the diet or body. The EDE-Q results range from 0 to 6, with higher scores indicating greater disordered eating. Volunteers were also asked to complete the Kessler Psychological Distress Scale (K10) with 10 prompts, scored from 1 to 5 and summed. Results from the K10 range from 10 to 50 with higher scores indicating severe mental disorder [23,24].

Diet Quality

Volunteers submitted the Diet Quality Questionnaire (DQQ), which is validated for use by adults in more than 100 countries [25]. This categorical survey uses food group indicators to calculate how well the diet achieves a global dietary recommendation (possible scores range from 0 to 18) and the degree to which dietary intake on the previous day protects against noncommunicable diseases (NCD-Protect, possible scores range from 0 to 9) and increases risk for noncommunicable diseases (NCD-Risk, possible scores range from 0 to 9). The DQQ also quantifies the proportion of the study group who ate at least 1 serving from all 5 of the food groups recommended in global food-based dietary guidelines (vegetable, fruit, pulse, nuts and seeds, animal-source food, and starchy staple). Diets of high quality also exhibit diversity; thus, the dietary diversity score ranges from 0 to 10 to reflect the average score for the study group, with a higher score indicating the presence of more food groups. Finally, the DQQ quantifies the proportion of the study group who consumed at least one fruit and one vegetable on the preceding day. Data from our investigation were compared against results collected from adults in the Global Diet Quality Project between 2021 and 2023 [26].

As an objective measurement of diet quality, antioxidant levels in the skin were evaluated using a handheld spectrophotometer (Konica Minolta CM-700d). More specifically, fruits and vegetables contain carotenoids that act as antioxidants to reduce

XSL•FO
RenderX

inflammation and cell damage [27]. Carotenoids, and specifically β-carotene, present as a yellow color and typically absorb light in the 450-nm frequency [28]. A spectrophotometer can be used to measure light reflection at specific wavelengths as an indicator of carotenoids, and therefore fruit and vegetable consumption. Using a previously documented noninvasive approach, we recorded the tristimulus b* value at the 450-nm frequency to assess carotenoid levels in the skin [29]. After daily calibration, the thenar eminence of the palm of the hand for each participant was scanned twice with the spectrophotometer. An average of the 4 scans was calculated and used for statistical analysis. As carotenoids are found in fruits and vegetables, measuring skin carotenoid levels is an accurate and less invasive measurement of dietary intake compared with self-reported data or blood tests.

Food Security

Much research has suggested that food security is directly related to diet quality [30]. Thus, we used the USDA Food Security Survey to assess whether our participants had limited or uncertain availability of nutritionally adequate foods [31].

Statistical Analysis

We examined associations between these metrics and SUS scores using correlation and regression analysis. Correlation (Pearson and Spearman) and ordinary least squares regression were conducted using Python (version 3.11; Python Software Foundation) with the Pandas, SciPy, Seaborn, and Statsmodels libraries. To assess the combined predictive value of use behaviors, we included all 3 use metrics as predictors in the regression model. Paired sample, 2-tailed *t* tests were performed using SPSS software (version 29; IBM Corp) to evaluate differences between variables at baseline and after 4 weeks of app use. Bootstrapping was performed with 5000 samples and bias corrected and accelerated (BCa) CI were calculated. Cohen *d* and Hedges correction were calculated as measures of the effect size. The Holm-Bonferroni adjustment was made to correct for multiple comparisons. Assumption testing was performed for outcomes yielding statistically significant results using the Shapiro-Wilk test and Q-Q plots. Where the normality assumption did not hold, a Wilcoxon signed-rank test was conducted as a sensitivity analysis. Missing data were excluded.

Ethical Considerations

Participants were fully informed of the study protocol and provided written informed consent before enrolling. This study was approved by the Loyola Marymount University Institutional Review Board (protocol number 2024FA16-R) according to the Declaration of Helsinki. Participants were assigned an identification number to protect privacy and ensure confidentiality.

Results

Overview

Demographic data for the 45 participants who completed the study are displayed in Table 1. This college-aged group was made up primarily of female participants (n=33, 73.3%) from diverse ethnic and racial groups, with approximately half living on campus (n=22, 48.9%) and using the campus-offered meal plan (n=24, 53.3%). Of the 45 participants in the study, 5 (11.1%) reported low (n=4, 8.8%) or very low (n=1, 2.2%) food security at baseline and follow-up. Food security measures did not correlate with intuitive eating, disordered eating, or diet quality in this population.

XSL•FO
RenderX

Table 1. Demographic characteristics of 45 college-aged volunteers who used the In2Eat app for 4 weeks.

	Value
Age (y), mean (SD)	20.5 (1.2)
Height (cm), mean (SD)	165.2 (8.3)
Weight (kg), mean (SD)	64.6 (13.5)
Sex, n (%)	
Female	33 (73.3)
Male	12 (26.7)
Ethnicity, n (%)	
Hispanic	13 (28.9)
Non-Hispanic	32 (71.1)
Race, n (%)	
Asian	15 (33.3)
Black or African American	10 (22.2)
Mixed (Asian and White)	5 (11.1)
White	10 (22.2)
Declined to state	5 (11.1)
Year in college, n (%)	
First year	7 (15.6)
Sophomore	18 (40)
Junior	9 (20)
Senior	11 (24.4)
Living status, n (%)	
On campus	22 (48.9)
Off campus	23 (51.1)
Meal plan[a] , n (%)	
Yes	24 (53.3)
No	21 (46.7)

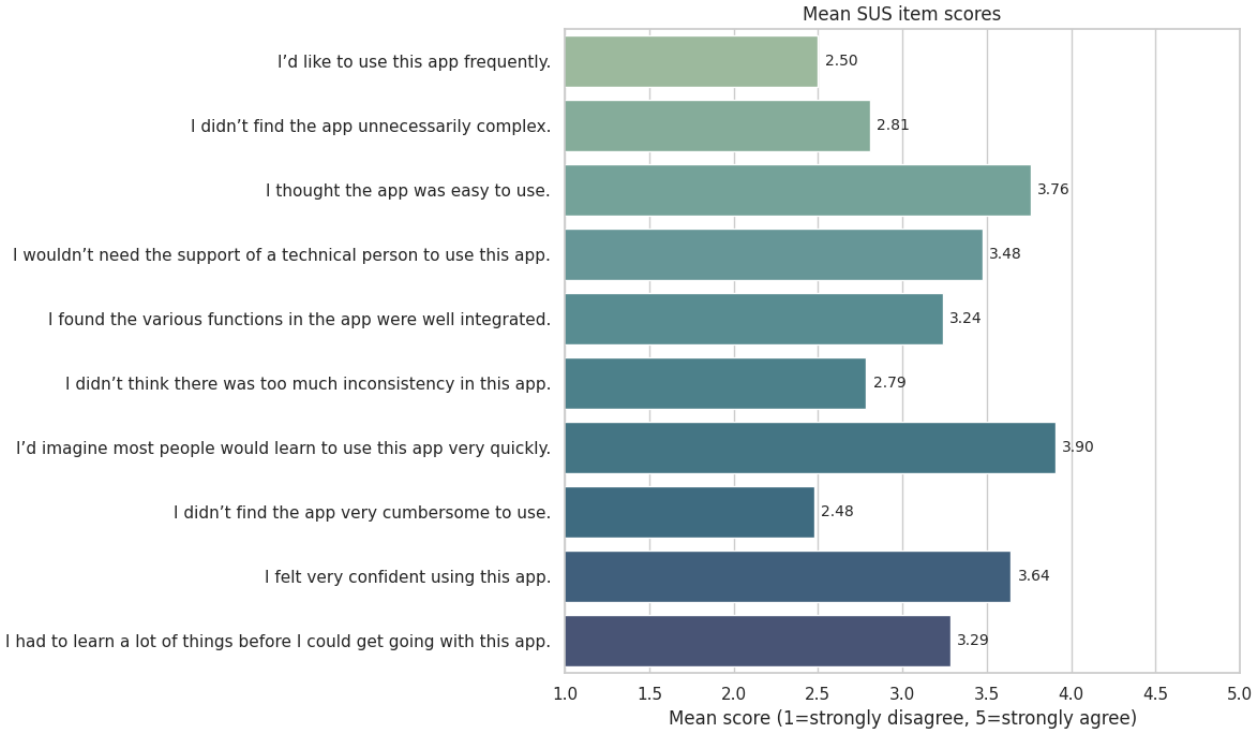
[a]Refers to whether students used the campus-offered meal plan.

System Usability

The average overall system usability score for the In2Eat app at the end of the 4-week pilot study was 67.2 (SD 15.5), which is similar to the commonly cited SUS benchmark score of 68 [18,19]. As shown in Figure 1, the questions with the highest scores were those concerning ease of use, specifically that users thought people would quickly learn to use the app, the app was very easy to use, and they felt confident using the app. By contrast, the questions with the lowest scores were those related to not finding the app cumbersome to use, wanting to use the app frequently, and not finding too much inconsistency or complexity in the app.

XSL•FO
RenderX

Figure 1. System Usability Scale (SUS) scores for each of the 10 survey questions, ranging from the lowest score of 1 (“Strongly disagree”) to the highest score of 5 (“Strongly agree”).



App Use

Figures 2 and 3 visualize the associations between system usability scores and app use using both scatterplots and group-wise boxplots. Pearson and Spearman correlations showed modest and mostly nonsignificant associations between use metrics and SUS scores. LogCount was positively correlated with SUS scores (Pearson r=0.28; P=.06), though this did not reach conventional significance. Notably, TotalMealsLogged had a statistically significant monotonic (but not linear) association with SUS scores (Spearman ρ=0.31; P=.04), suggesting that participants who logged more meals tended to rate the app as more usable.

Table 2 presents descriptive statistics for app use across the full study sample as well as within low, medium, and high use groups (tertiles based on LogCount). As expected, participants in the high use group demonstrated consistently greater engagement across all metrics compared with those in the medium and low use groups. Specifically, high use users logged meals more frequently, entered a greater total number of meals, and maintained a higher average number of meals logged per day. The medium use group showed intermediate values across all engagement metrics. These results confirm a clear separation between groups, capturing both the breadth and intensity of app engagement across the cohort.

Tertile grouping was used to provide an interpretable comparison of usability across distinct levels of engagement (low, medium, and high) and to examine potential nonlinear patterns in the relationship between app use and perceived usability that may not be captured by correlation analyses alone. In addition to group-based comparisons, we report both Pearson and Spearman correlations between use metrics and SUS scores. These analyses showed modest associations, with a statistically significant monotonic relationship observed between TotalMealsLogged and SUS scores (Spearman ρ=0.31; P=.04), indicating that greater meal logging was associated with higher perceived usability.

We further explored differences in usability scores across low, medium, and high use groups (tertiles based on LogCount). A Kruskal-Wallis test revealed a significant difference in SUS scores across LogCount groups (H=6.75; P=.03), indicating that participants with higher day-level logging frequency perceived the app as more usable. No substantial group differences were observed for TotalMealsLogged or AvgMealsPerDay.

Finally, we performed an ordinary least squares regression, including all 3 use metrics as predictors of SUS scores. The overall model explained only 8% of the variance (R^2=0.080; P=.31), and none of the individual predictors were statistically significant. While LogCount had the largest coefficient (β=.55), its contribution was not statistically robust (P=.31). These results suggest a trend toward more consistent engagement being associated with higher usability ratings, but the effect sizes are modest and should be interpreted cautiously given the limited sample size.

XSL•FO
RenderX

Figure 2. Relationship between app use metrics and perceived system usability. Each panel shows a scatterplot with a linear regression line, comparing (A) the total number of days users logged meals (LogCount), (B) the total number of meals logged, and (C) the average number of meals per day with the System Usability Scale (SUS) scores. Across all 3 use metrics, greater engagement with the app tended to be associated with higher perceived usability.

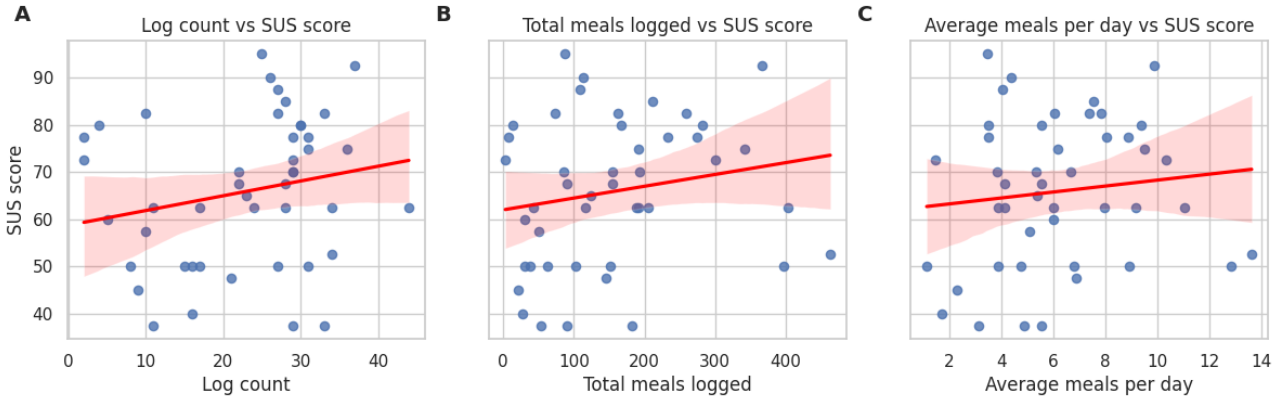


Figure 3. Distribution of System Usability Scale (SUS) scores grouped by use level (low, medium, and high) for each app engagement metric. Participants with higher app use more frequently reported higher usability scores, particularly those with higher LogCount values.

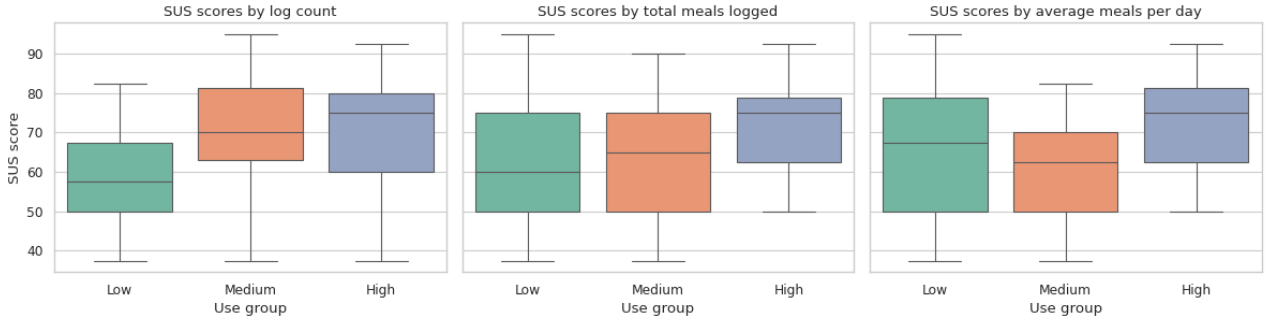


Table 2. App use metrics among the 45 participants overall and by group.

Use metrics	Overall, mean (SD)	Low, mean (SD)	Medium, mean (SD)	High, mean (SD)
LogCount	23.11 (10.34)	10.2 (5.28)	26.28 (2.76)	33.62 (3.8)
TotalMealsLogged	157.09 (115.43)	57.8 (52.95)	144.17 (64.41)	289.54 (96.13)
AvgMealsPerDay	6.19 (2.85)	5.02 (2.66)	5.45 (2.2)	8.57 (2.59)

Intuitive Eating

The global score for intuitive eating did not change after 4 weeks of using the In2Eat app (Table 3); however, the unconditional permission to eat subscale of the IES-3 did increase significantly (P=.01), though with a small effect size (Hedges g=−0.28) and not after Holm-Bonferroni adjustment to control for type I error across related outcomes with multiple comparisons (adjusted P=.07). Normality assumptions were met.

XSL•FO
RenderX

Table 3. Intuitive eating, disordered eating, and psychological distress at baseline and after 4 weeks (n=42).

Outcome	Baseline, mean (SD)	Mean difference[a] (95% BCa[b] CI)	*P* value	Adjusted *P* value[c]	Hedges *g*
Intuitive eating					
Global score	3.78 (0.53)	0.02 (−0.08 to 0.12)	.68	.68	0.04
Unconditional permission to eat	3.60 (1.03)	0.28 (0.07 to 0.49)	.01[d]	.07	0.28
Eating for physical reasons	3.83 (0.75)	−0.06 (−0.23 to 0.13)	.55	>.99	−0.08
Reliance on hunger satiety cues	3.79 (0.74)	−0.15 (−0.34 to 0.02)	.12	.48	−0.19
Body food choice congruence	3.89 (0.60)	0.05 (−0.08 to 0.17)	.49	>.99	−0.08
Disordered eating					
Global score	1.24 (1.02)	−0.10 (−0.31 to 0.09)	.32	>.99	−0.10
Dietary restraint	0.96 (1.38)	−0.30 (−0.61 to −0.04)	.04[d]	.22	−0.23
Eating concern	0.66 (0.65)	−0.05 (−0.21 to 0.12)	.55	>.99	−0.07
Body shape concern	1.77 (1.41)	−0.05 (−0.33 to 0.21)	.74	>.99	−0.03
Weight concern	1.56 (1.34)	−0.03 (−0.33 to 0.25)	.84	.84	−0.02
Psychological distress	17.0 (5.23)	−0.69 (–2.45 to 1.02)	.41	—[e]	−0.13

[a]Mean difference was calculated as week 4−baseline.

[b]BCa: bias corrected and accelerated.

[c]Holm-Bonferroni adjusted *P* value.

[d]Statistical significance at $P<.05$.

[e]Not applicable.

Disordered Eating and Mental Health

Internal consistency at baseline was excellent for the EDE-Q scale ($\alpha=0.95$) and the K10 ($\alpha=0.89$). Baseline and follow-up assessments of disordered eating via the EDE-Q revealed no major changes during app use, except for dietary restraint which decreased significantly ($P=.04$), though with a small effect size (Hedges $g=-0.23$) and not after Holm-Bonferroni adjustment to control for familywise error (adjusted $P=.22$). For dietary restraint, the normality assumption did not hold due to an S curve in the Q-Q plot and the Shapiro-Wilk test result ($P<.001$); however, the Wilcoxon signed-rank sensitivity analysis ($P=.03$) confirmed the finding, indicating that the result was robust to nonnormality.

Diet Quality

Diet quality among app users also appeared to improve during the study (Table 4). Dietary features that are known to protect against noncommunicable diseases (NCDs) increased by 28%, while there was a negligible decrease (4%) in dietary features that increase the risk for NCDs. Of particular note is the 63% increase in fruit consumption during the app intervention. Furthermore, paired *t* tests (N=47) revealed a significant increase in skin antioxidant levels ($P=.03$) with a small effect size (Hedges $g=0.16$).

We additionally examined whether outcomes differed between students living on campus and those living off campus. No significant differences were observed between groups across all primary outcomes, including diet quality (global dietary recommendation score, NCD-Protect, and NCD-Risk), intuitive eating (IES-3), disordered eating (EDE-Q), and skin antioxidant levels (all $P>.05$). Similarly, SUS scores did not differ between groups ($P=.36$). These findings suggest that residential status did not meaningfully influence study outcomes.

XSL•FO
RenderX

Table 4. Diet quality assessed before and after 4 weeks of using the In2Eat app.

	Baseline	4 weeks	Change (%)	United States average[a]
Global dietary recommendations	9.13	9.98	9.3	9.5
NCD[b]-protect[c]	2.63	3.37	28.1	3.9
NCD-risk[d]	2.50	2.39	−4.3	3.5
All-5[e]	0.21	0.28	34.6	0.44
Dietary diversity[f]	5.19	5.43	4.7	6.4
At least 1 fruit[g]	0.44	0.72	63.0	0.77
At least 1 vegetable[h]	0.71	0.70	−2.0	0.85
Skin antioxidant levels	15.38	15.71	6.1	—[i]

[a]Average scores for adults in the United States [26].

[b]NCD: noncommunicable disease.

[c]Degree to which diet is protective against noncommunicable diseases.

[d]Degree to which diet increases risk for noncommunicable diseases.

[e]Proportion of participants who ate all 5 food groups on the previous day.

[f]Number of food groups consumed on the previous day.

[g]Proportion of participants eating at least one fruit on the previous day.

[h]Proportion of participants eating at least one vegetable on the previous day.

[i]Not applicable.

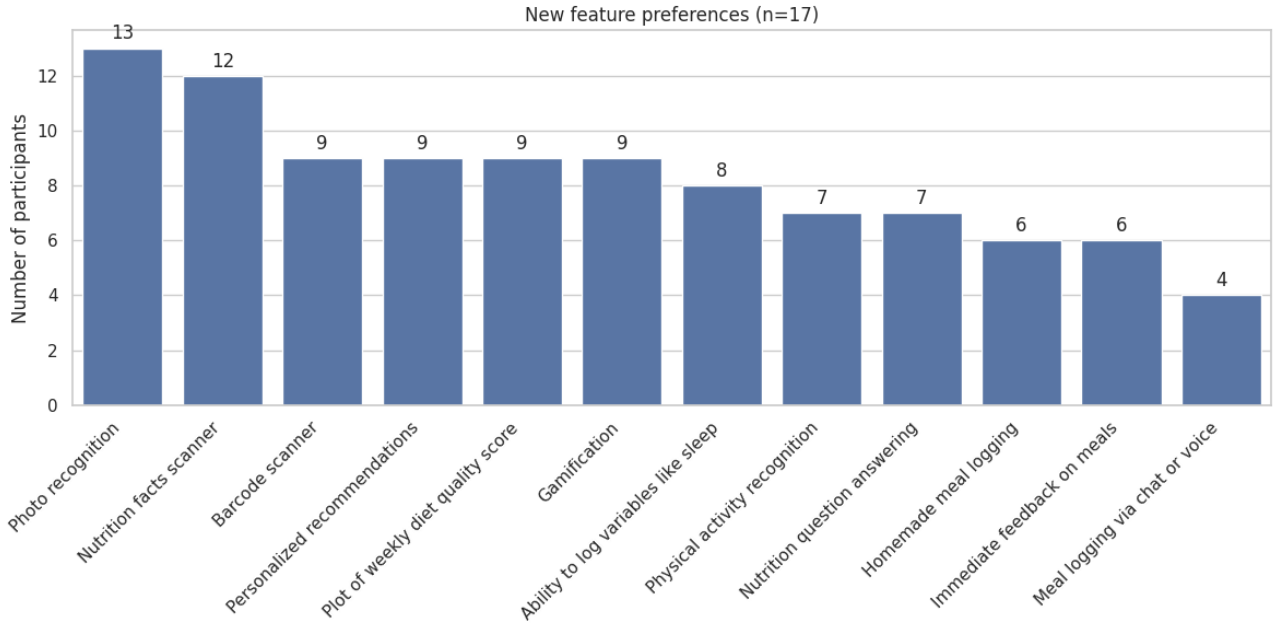
Qualitative Feedback

We collected qualitative feedback from 17 study participants via a Qualtrics survey to inform future strategies for improving the app and to investigate why participants did not frequently use the app despite the monetary incentive. The survey asked 2 free-response questions, one about why they may not have used the app often, and the second about how we could improve the In2Eat app. The final question asked them to select which of 12 hypothetical new features would make them more likely to use the app more often. As shown in Figure 4, by far the most popular new feature was computer vision—13 (76.5%) out of 17 respondents selected the ability to log photos of their meals, and 12 (70.6%) selected the ability to take a photo of a nutrition facts label.

Patterns emerged from the free-response questions as well. Four (23.5%) respondents each mentioned forgetting to log meals and that reminders would be helpful, that they would like incentives or rewards for logging meals, and that the data entry was too time-consuming and/or annoying. Three (17.6%) respondents each mentioned wanting computer vision; worries about their own logging precision, especially portion size, as well as the accuracy of the app itself; and the app missing foods, including non-American cuisines. Other feedback included speeding up meal logging with “quick add” of common foods or meals; making the interface more interactive, visual, and informative; adding a social tab to see friends; helping them to track and achieve their nutrition goals; and adding a barcode scanner. This qualitative data analysis strongly supported our current work on incorporating computer vision as well as gamification features (ie, incentives, goals, and friends).

XSL•FO
RenderX

Figure 4. Frequency of interest in hypothetical new features from 17 app users.



Discussion

Principal Findings

The overarching goal of this study was to investigate whether the use of a mobile phone app that promotes intuitive eating principles (through hunger satiety checks at each meal and push notifications with daily intuitive eating quotes of the day) would improve diet quality, without increasing disordered eating symptoms as often occurs when using other diet-tracking apps [10,11]. Our results from a 4-week user study with 45 college students demonstrated that diet quality improved, as measured by a significant increase in skin antioxidant levels [27], a 63% increase in fruit consumption, and a 28% increase in dietary features known to protect against NCDs. Compared with data reported for American adults, a smaller proportion of our college-aged participants consumed all 5 food groups, 1 fruit, 1 vegetable, and showed a lower dietary diversity score. On a positive note, the global dietary recommendation score for our population was above the national average for US adults, while the NCD-Protect score became closer to the national average after 4 weeks of app use. Notably, the NCD-Risk score among our participants was lower than national norms, indicating a diet with fewer contributing factors to disease. The unconditional permission to eat subscale of the intuitive eating IES-3 increased significantly. Finally, unlike existing diet-tracking apps, ours did not increase disordered eating; rather, for every disordered eating metric, the mean slightly decreased after the intervention, and there was a significant decrease in dietary restraint. However, due to the small sample size, the effects were small, and Holm-Bonferroni adjusted *P* values were not significant.

Strengths

This work has several strengths. First, our study is the first intervention that we know of to provide a validated alternative to the expensive and time-consuming current approach of working one-on-one with a registered dietician, without the harm caused by most apps that emphasize weight loss and calorie counting. Second, our results demonstrated significant improvement in diet quality after a 4-week intervention using our app. Improved diet quality is associated with health promotion and disease prevention. Finally, our research used validated measures of intuitive eating, disordered eating, and psychological distress, as well as robustness checks, including bootstrapped CIs and nonparametric sensitivity analyses to strengthen confidence in the findings.

Comparison With Prior Work

Prior work demonstrated the accuracy of logging meals via natural language within a mobile app [4] and showed that the use of intuitive eating principles may counteract disordered eating behaviors [13-16]. Our work builds upon this literature by investigating the use of a new mobile app that not only allows users to log meals with natural language but also promotes intuitive eating principles and asks users to log their hunger satiety.

Limitations

Limitations of the study include the small sample size (n=45), which resulted in small effect sizes, and the lack of diversity because we only recruited college students, who have been shown in prior work to have inadequate nutrient intake [7,8]. Another limitation is the wide range of app use, resulting in study participants who did not use the app often despite the monetary incentive. Other limitations include the lack of a control group, the inclusion of participants with eating disorders and medically supervised diets, and the absence of socioeconomic data.

Future work on In2Eat will focus primarily on improving the app's usability and engagement, as well as incorporating more intuitive eating features. On the basis of feedback from our study participants, we are currently adding computer vision and nutrition facts label scanning so users can more easily log their meals with photos of meals or nutrition facts labels. We are also adding tooltips with explanations for each nutrient and more

XSL•FO
RenderX

useful feedback and visualizations, including a detailed analysis after logging a meal and a weekly diet quality score with feedback for improvement. Furthermore, we are expanding from iOS to the Android platform.

Conclusions

To our knowledge, our In2Eat iOS app is the first mobile app to combine diet tracking with intuitive eating, and ours is the first study to demonstrate that the use of a diet-tracking app with a focus on intuitive eating instead of calorie counting improves diet quality without increasing disordered eating.

Acknowledgments

This study would not have been possible without the support of the Seaver College of Science and Engineering at Loyola Marymount University through an interdisciplinary seed grant.

The authors thank undergraduate computer science students Catalina Barnes, Lauren Campbell, Funmilola Idowu, and Ngoc Kim Ngan Tran, who designed and developed the In2Eat iOS app, and undergraduate health and human sciences students Alexis Arle, Gabrielle Besio, Samantha Chuakay Lim, and Cameron Sumaquial, who conducted the user study and analyzed the data. The authors also thank undergraduate computer science student Natasha Cordova-Diba for her preliminary work on the In2Eat website and Riley Kehoe for his assistance with the development of the recipe recommender system.

Funding

This study was supported by an interdisciplinary seed grant from the Seaver College of Science and Engineering at Loyola Marymount University.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to institutional review board requirements regarding the privacy of sensitive personal data collected during this study, including mental health and food security.

Authors' Contributions

Conceptualization: HCA (lead), MK (equal), DY (equal)
Data curation: HCA
Formal analysis: HCA, DY (supporting)
Funding acquisition: MK (lead), HCA (equal), DY (equal)
Investigation: HCA (lead), MK (supporting), DY (supporting)
Methodology: HCA
Project administration: MK (lead), HCA (equal), DY (equal)
Resources: MK (lead), HCA (equal), DY (equal)
Supervision: MK (lead), HCA (equal), DY (equal)
Validation: HCA
Visualization: DY
Writing—original draft: HCA (lead), MK (equal), DY (equal)
Writing—review and editing: MK (lead), HCA (equal), DY (equal)

Conflicts of Interest

MK is the owner of MealMate Inc. All other authors declare no other conflicts of interest.

References

1. Ngo J, Engelen A, Molag M, Roesle J, García-Segovia P, Serra-Majem L. A review of the use of information and communication technologies for dietary assessment. Br J Nutr 2009 Jul;101 Suppl 2:S102-S112 [FREE Full text] [doi: 10.1017/S0007114509990638] [Medline: 19594959]
2. Griffiths C, Harnack L, Pereira MA. Assessment of the accuracy of nutrient calculations of five popular nutrition tracking applications. Public Health Nutr 2018 Jun;21(8):1495-1502 [FREE Full text] [doi: 10.1017/S1368980018000393] [Medline: 29534771]
3. Winzer E, Luger M, Schindler K. Using digital photography in a clinical setting: a valid, accurate, and applicable method to assess food intake. Eur J Clin Nutr 2018 Jun;72(6):879-887. [doi: 10.1038/s41430-018-0126-x] [Medline: 29563639]
4. Taylor S, Korpusik M, Das S, Gilhooly C, Simpson R, Glass J, et al. Use of natural spoken language with automated mapping of self-reported food intake to food composition data for low-burden real-time dietary assessment: method

XSL•FO
RenderX

comparison study. J Med Internet Res 2021 Dec 06;23(12):e26988 [FREE Full text] [doi: 10.2196/26988] [Medline: 34874885]
5. Nelson MC, Story M, Larson NI, Neumark-Sztainer D, Lytle LA. Emerging adulthood and college-aged youth: an overlooked age for weight-related behavior change. Obesity (Silver Spring) 2008 Oct;16(10):2205-2211 [FREE Full text] [doi: 10.1038/oby.2008.365] [Medline: 18719665]
6. Stok FM, Renner B, Clarys P, Lien N, Lakerveld J, Deliens T. Understanding eating behavior during the transition from adolescence to young adulthood: a literature review and perspective on future research directions. Nutrients 2018 May 24;10(6):667 [FREE Full text] [doi: 10.3390/nu10060667] [Medline: 29794986]
7. Bernardo GL, Jomori MM, Fernandes AC, da Costa Proença RP. Food intake of university students [Article in Portuguese]. Rev Nutr 2017;30(6):847-865. [doi: 10.1590/1678-98652017000600016]
8. Shorrock FH, Almstedt H. Nutrient intake and diet quality according to meal plan use in college students. J Food Nutr Sci 2022;4(2):108-121. [doi: 10.1057/jfns000040]
9. Burchi F, Fanzo J, Frison E. The role of food and nutrition system approaches in tackling hidden hunger. Int J Environ Res Public Health 2011 Feb;8(2):358-373 [FREE Full text] [doi: 10.3390/ijerph8020358] [Medline: 21556191]
10. Messer M, McClure Z, Norton B, Smart M, Linardon J. Using an app to count calories: motives, perceptions, and connections to thinness- and muscularity-oriented disordered eating. Eat Behav 2021 Dec;43:101568. [doi: 10.1016/j.eatbeh.2021.101568] [Medline: 34543856]
11. Levinson CA, Brosof LC, Ma J, Fewell L, Lenze EJ. Fear of food prospectively predicts drive for thinness in an eating disorder sample recently discharged from intensive treatment. Eat Behav 2017 Dec;27:45-51 [FREE Full text] [doi: 10.1016/j.eatbeh.2017.11.004] [Medline: 29145095]
12. Tribole E, Resch E. Intuitive Eating: A Revolutionary Anti-Diet Approach, Fourth Edition. New York, NY: St. Martin's Publishing Group; 2020.
13. Anderson LM, Reilly EE, Schaumberg K, Dmochowski S, Anderson DA. Contributions of mindful eating, intuitive eating, and restraint to BMI, disordered eating, and meal consumption in college students. Eat Weight Disord 2016 Mar;21(1):83-90. [doi: 10.1007/s40519-015-0210-3] [Medline: 26243300]
14. Katcher JA, Suminski RR, Pacanowski CR. Impact of an intuitive eating intervention on disordered eating risk factors in female-identifying undergraduates: a randomized waitlist-controlled trial. Int J Environ Res Public Health 2022 Sep 23;19(19):12049 [FREE Full text] [doi: 10.3390/ijerph191912049] [Medline: 36231362]
15. Linardon J, Fairburn CG, Fitzsimmons-Craft EE, Wilfley DE, Brennan L. The empirical status of the third-wave behaviour therapies for the treatment of eating disorders: a systematic review. Clin Psychol Rev 2017 Dec;58:125-140. [doi: 10.1016/j.cpr.2017.10.005] [Medline: 29089145]
16. Thomson J, Almstedt HC. Intuitive eating and the female athlete triad in collegiate runners. Nutrients 2025 Jul 17;17(14):2337 [FREE Full text] [doi: 10.3390/nu17142337] [Medline: 40732962]
17. Institute of Medicine. In: Meyers LD, Hellwig JP, Otten JJ, editors. Dietary Reference Intakes: The Essential Guide to Nutrient Requirements. Washington, DC: National Academies Press; 2006.
18. Sauro J. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. Denver, CO: Measuring Usability LLC; 2011.
19. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
20. Tylka TL, Maïano C, Fuller-Tyszkiewicz M, Linardon J, Burnette CB, Todd J, et al. The Intuitive Eating Scale-3: development and psychometric evaluation. Appetite 2024 Aug 01;199:107407. [doi: 10.1016/j.appet.2024.107407] [Medline: 38729580]
21. Fairburn CG, Cooper Z, Shafran R. Cognitive behaviour therapy for eating disorders: a "transdiagnostic" theory and treatment. Behav Res Ther 2003 May;41(5):509-528. [doi: 10.1016/s0005-7967(02)00088-8] [Medline: 12711261]
22. Mond JM, Hay PJ, Rodgers B, Owen C, Beumont PJ. Validity of the Eating Disorder Examination Questionnaire (EDE-Q) in screening for eating disorders in community samples. Behav Res Ther 2004 May;42(5):551-567. [doi: 10.1016/s0005-7967(03)00161-x]
23. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SL, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002 Sep 26;32(6):959-976. [doi: 10.1017/S0033291702006074]
24. Andrews G, Slade T. Interpreting scores on the Kessler Psychological Distress Scale (K10). Aust N Z J Public Health 2001 Dec;25(6):494-497 [FREE Full text] [doi: 10.1111/j.1467-842x.2001.tb00310.x] [Medline: 11824981]
25. Herforth AW, Wiesmann D, Martínez-Steele E, Andrade G, Monteiro CA. Introducing a suite of low-burden diet quality indicators that reflect healthy diet patterns at population level. Curr Dev Nutr 2020 Nov 25;4(12):nzaa168 [FREE Full text] [doi: 10.1093/cdn/nzaa168] [Medline: 33344879]
26. Global Diet Quality Project: enabling diet quality monitoring globally with tools and data. Harvard Dataverse. URL: https://dataverse.harvard.edu/dataverse/dqqdata [accessed 2026-06-17]
27. Krinsky NI, Johnson EJ. Carotenoid actions and their relation to health and disease. Mol Aspects Med 2005 Dec;26(6):459-516. [doi: 10.1016/j.mam.2005.10.001] [Medline: 16309738]

XSL•FO
RenderX

28. Stahl W, Heinrich U, Jungmann H, von Laar J, Schietzel M, Sies H, et al. Increased dermal carotenoid levels assessed by noninvasive reflection spectrophotometry correlate with serum levels in women ingesting Betatene. J Nutr 1998 May;128(5):903-907. [doi: 10.1093/jn/128.5.903] [Medline: 9567001]
29. Alaluf S, Heinrich U, Stahl W, Tronnier H, Wiseman S. Dietary carotenoids contribute to normal human skin color and UV photosensitivity. J Nutr 2002 Mar;132(3):399-403. [doi: 10.1093/jn/132.3.399] [Medline: 11880562]
30. Hanson KL, Connor LM. Food insecurity and dietary quality in US adults and children: a systematic review. Am J Clin Nutr 2014 Aug;100(2):684-692 [FREE Full text] [doi: 10.3945/ajcn.114.084525] [Medline: 24944059]
31. U.S. Adult Food Security Survey Module: three-stage design, with screeners. Economic Research Service, USDA. 2012 Sep. URL: https://www.ers.usda.gov/media/6424/us-adult-food-security-survey-module.pdf [accessed 2026-06-13]

Abbreviations

BCa: bias corrected and accelerated
CNN: convolutional neural network
DQQ: Diet Quality Questionnaire
EDE-Q: Eating Disorder Examination Questionnaire
IES-3: Intuitive Eating Scale version 3
K10: Kessler Psychological Distress Scale
LLM: large language model
NCD: noncommunicable disease
NCD-Protect: the degree to which dietary intake on the previous day protects against noncommunicable diseases
NCD-Risk: the degree to which dietary intake on the previous day increases risk for noncommunicable diseases
SUS: System Usability Scale
USDA: US Department of Agriculture

Edited by K Sellen; submitted 30.Jul.2025; peer-reviewed by G Downer, PhD, FAND, RD, LD, CNS, M Cooper; comments to author 09.Feb.2026; accepted 17.Jun.2026; published 17.Jul.2026.

Please cite as:
Korpusik M, YazdanSepas D, Almstedt HC
A Smartphone App to Improve Intuitive Eating and Diet Quality: Design and Usability Study
JMIR Hum Factors 2026;13:e81439
URL: https://humanfactors.jmir.org/2026/1/e81439
doi:10.2196/81439
PMID:

©Mandy Korpusik, Delaram YazdanSepas, Hawley C Almstedt. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 17.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Examining Telemedicine Adoption After the COVID-19 Pandemic and Evaluating the Impact of a Virtual Hospital on Clinician Engagement: Longitudinal Survey Study

Galia Barkai[1,2], MBA, MD; Shany Edelman-Yaron[1], MSc; Hadar Amir[1], MBA; Havi Inbar-Lankry[3], MA; Rachel Sarafraz[1], MPH

[1]Sheba BEYOND Virtual Hospital, Sheba Medical Center, Ramat Gan, Israel

[2]Gray Faculty of Medical & Health Sciences, Tel Aviv University, Tel-Aviv, Israel

[3]Sheba Medical Center, Ramat Gan, Tel Aviv, Israel

Corresponding Author:
Galia Barkai, MBA, MD
Sheba BEYOND Virtual Hospital
Sheba Medical Center
2 Sheba Road
Ramat Gan, 5265601
Israel
Phone: 972 526668304
Email: galiabar69@gmail.com

Abstract

Background: After the emergence of telemedicine as a major health care delivery mode during the COVID-19 pandemic, Sheba Medical Center (SMC) rapidly transitioned its outpatient clinics to telemedicine. A postimplementation satisfaction study revealed that while nearly 90% of patients reported a positive experience, only 38% of clinicians expressed high satisfaction. SMC soon established Sheba BEYOND, a dedicated virtual hospital. This initiative was accompanied by extensive efforts to improve the telemedicine experience. In 2025, a follow-up survey addressing clinician satisfaction and acceptance was conducted.

Objective: This study aimed to explore whether clinician satisfaction improved and which barriers to clinician engagement persisted.

Methods: An anonymous survey was distributed via a link to the staff of SMC's outpatient clinics. The survey inquired about user experience with the telemedicine digital platform, integration in workflow, barriers to use, perceived benefit to patients, and overall satisfaction. Clinicians who never or rarely perform telemedicine visits were classified as nonusers; all others were classified as users. The questionnaire included 21 questions, 2 additional questions for nonusers, and 5 additional questions for users. Overall satisfaction was ranked on a 10-point Likert scale. Descriptive statistics were calculated using SPSS (version 29.02; IBM Corp) and Microsoft Excel. A chi-square test of independence was used for categorical variables. We used a 2-tailed *t* test, with a probability of <.05 considered statistically significant.

Results: The survey was sent to 2735 clinicians, and 274 (10%) responded. Telemedicine users accounted for 52.6% (144/274). Among all respondents, 54.4% (149/274) were physicians, 30.3% (83/274) were paramedical staff, and 15.3% (42/274) were nurses. Clinical specialties across all respondents were internal medicine (36/274, 13%), psychiatry (34/274, 12%), pediatrics (36/274, 13%), oncology (31/274, 11%), and other specialties (137/274, 50%). Compared to the 2020 survey, a 41.9% increase was recorded in the number of users reporting a high level of satisfaction, from 37.7% (61/162) to 53.5% (77/144; *P*=.006). This increase was largely driven by a significant improvement in the oncology clinic, where users reporting a high level of satisfaction rose from 23.3% (10/43) to 63.6% (14/22; *P*=.01). In total, 64.4% (176/274) of respondents believed that telemedicine was advantageous for patient care and 66% (95/144) of users reported that telemedicine reduced their workload. Among the 103 video-visit users who answered the technical feedback questions, 62.1% (64/103) experienced problems in connecting to the patients' network and 32% (33/103) experienced connectivity problems in SMC. In fact, 80.6% (83/103) reported that they had to switch to the telephone at least once. Most of the 98 nonusers (80/98, 81.6%) stated that they did not perform telemedicine because it was mostly clinically inappropriate.

XSL•FO
RenderX

Conclusions: This study demonstrates the positive impact of institutional change management strategy in promoting clinician engagement with telemedicine. However, persistent technical barriers and varying engagement highlight opportunities for refinement.

(JMIR Hum Factors 2026;13:e93911) doi:10.2196/93911

KEYWORDS

telemedicine; telehealth; health personnel attitudes; clinician engagement; clinician satisfaction; video consultations; implementation barriers; virtual hospital; change management

Introduction

The COVID-19 pandemic catalyzed an unprecedented transformation in health care delivery worldwide, with telemedicine emerging as a vital solution to maintain continuity of care during lockdowns and social distancing mandates [1]. At Sheba Medical Center (SMC), outpatient services rapidly transitioned to telemedicine in early 2020. A postimplementation survey revealed a definite contrast in satisfaction levels: while nearly 90% of patients reported a positive experience, only 38% of clinicians expressed high satisfaction, citing technical challenges, administrative burden, and difficulty integrating telemedicine into clinical workflows [2]. This discrepancy echoed broader global trends observed at the time: a high level of patient satisfaction [3-5] with usability challenges on the clinicians' side [6]. Although these studies examined short-term satisfaction during the pandemic, there is limited evidence on whether sustained institutional change can lead to durable cultural shifts in clinician engagement with telemedicine.

Soon after the onset of the pandemic, the SMC established Sheba BEYOND, a dedicated virtual hospital that launched new telemedicine-based, hybrid care services, including home hospitalization, remote rehabilitation, and chronic patient management. In addition to ensuring continuity of care during the pandemic, Sheba BEYOND was established as part of a broader organizational strategy aimed at improving accessibility to specialist care, reducing geographic barriers, enhancing timeliness and continuity of ambulatory services, and supporting long-term digital transformation through hybrid models of care delivery. This initiative was accompanied by extensive efforts to facilitate clinician engagement and improve the telemedicine experience through the establishment of a video telemedicine helpdesk for technical support, frequent technical upgrades, interoperability improvements, administrative streamlining, and leadership training and incentives for telemedicine adoption.

Between 2020 and 2024, the number of yearly telemedicine visits at SMC more than doubled, increasing from approximately 60,000 in 2020 to >114,000 in 2024. Yearly patient experience surveys have continued to show high levels of patient satisfaction, with minor fluctuations over the years. In 2025, we conducted a follow-up survey addressing clinicians to assess whether the organizational efforts to facilitate telemedicine adoption translated into increased clinician satisfaction and acceptance over time.

Unlike prior studies focused primarily on short-term telemedicine implementation during the acute pandemic phase, this study examines long-term clinician engagement several years later within the context of an institutional virtual hospital framework and sustained organizational change efforts and explores whether the significant increase in telemedicine visits reflects a true cultural shift or if barriers to clinician engagement persist.

Methods

Overview

To assess the experience of telemedicine clinicians, an anonymous survey link was distributed through both text message and email to SMC's outpatient clinic staff. The survey questions assessed users' experiences interacting with the telemedicine digital platform, integration of telemedicine in clinician workflow, barriers to use, perceived benefit to patients, and overall satisfaction and thoughts on potential improvements to the service. Clinicians who reported that they never or rarely perform telemedicine visits were classified as nonusers; all others were classified as users. The questionnaire included 21 questions for all respondents, 2 additional questions for nonusers, and 5 additional questions for users. Three reminders, sent through both text message and email, spaced 1 week apart, were sent to the clinicians who had yet to respond to the survey.

Eligibility

The study did not define formal patient-level inclusion or exclusion criteria, as the survey focused on clinician experience across existing outpatient telemedicine services routinely performed at the institution. Decisions regarding telemedicine appropriateness were made according to specialty-specific clinical judgment and institutional practice.

Surveys

Overall satisfaction was ranked on a 10-point Likert scale [7]. Satisfaction ratings of 8 to 10 were considered as a high level of satisfaction, ratings of 4 to 7 as an intermediate level of satisfaction, and ratings of 1 to 3 as a low level of satisfaction. Responses to operational questions such as ease of use and adherence to appointment schedule were rated on a 5-point Likert scale, with scores of 4 to 5 considered high, 3 considered intermediate, and 1 to 2 considered low. For patient benefit, ratings of 4 to 5 were considered as a high level of support, 3 was considered as an intermediate level of support, and 1 to 2 were considered as a low level of support. Several survey items were adapted from the previously published institutional telemedicine survey conducted in 2020, while additional questions were developed by the investigators to assess workflow integration, operational barriers, and modality-specific experiences [2]. Comparison to the 2020 survey was made only

XSL•FO
RenderX

when questions were similar in the 2025 survey. The questionnaires (translated from Hebrew) are presented in Multimedia Appendix 1.

Statistical Analysis

Descriptive statistics were calculated with Microsoft Excel and SPSS software (version 29.02; IBM Corp). A chi-square test of independence was used for categorical variables. We used a 2-tailed *t* test, with a probability of <.05 considered statistically significant.

Ethical Considerations

The study received approval from SMC's institutional review board for retrospective analysis of prospectively collected data (2516-SMC-D). Participation in the survey was voluntary and anonymous. Completion of the questionnaire was considered implied informed consent. No identifiable personal information was collected.

Results

The survey was distributed to 2735 clinicians and 274 responded, representing a 10% response rate. Of the 274 respondents, 61% (n=167) accessed the survey link via text message. Telemedicine users accounted for 52.5% (144/274) of the respondents. Clinical characteristics and mode of communication are presented in Table 1.

Table 1. Clinician characteristics and mode of communication (n=274).

	Clinicians, n (%)
Clinician characteristics (n=274)	
Clinical specialty	
Internal medicine	36 (13)
Psychiatry	34 (12)
Pediatrics	36 (13)
Oncology	31 (11)
Other specialties	137 (50)
Clinician type	
Physicians	149 (54.4)
Paramedical staff	83 (30.3)
Nurses	42 (15.3)
Mode of communication (n=143[a])	
Mostly video visits	69 (48.3)
Both telephone and video	56 (39.2)
Primarily telephone	14 (9.8)
Alternative modes	4 (2.8)

[a]Notably, of the 143 clinicians who responded to the question regarding the mode of communication, none of the 19 psychiatrists indicated telephone visits as a significant feature of their telemedicine practice; 36.8% (7/19) of the psychiatrists engaged predominantly in video visits, while 63.2% (12/19) exclusively conducted video-based telemedicine visits.

User Satisfaction: Comparison to 2020

There was a 41.9% increase in users reporting a high level of satisfaction, from 37.7% (61/162) in 2020 to 53.5% (77/144) in 2025 (*P*=.006). Additionally, the proportion of respondents reporting a low level of satisfaction decreased from 11.1% in 2020 to 4.2% in 2025 (*P*=.02; Figure 1).

The observed overall increase in satisfaction was largely driven by a significant improvement among clinicians in the oncology clinic, where users reporting high satisfaction increased from 23.3% (10/43) in 2020 to 63.6% (14/22) in 2025 (*P*=.01). A high level of satisfaction in other clinics, including internal medicine, psychiatry, pediatrics, and other specialties, remained largely unchanged during the same period (Figure 2).

Changes in overall satisfaction levels across different sectors were predominantly nonsignificant. However, the high level of satisfaction among paramedical staff increased from 40.3% (29/72) to 59.3% (32/54; *P*=.03), while a low level of satisfaction among physicians decreased from 21.8% (19/87) to 9.5% (7/74; *P*=.03).

When queried about the perceived benefits of telemedicine for patients overall, 64.4% (176/274) of respondents expressed the belief that telemedicine was advantageous for patient care. No significant variation was observed across medical disciplines. Among telemedicine users, the proportion who considered telemedicine beneficial for their patients remained high and unchanged over time, at 87% (141/162) in 2020 and 86.6% (123/142) in 2025. In contrast, among nonusers, this proportion decreased from 68% (34/50) in 2020 to 41.9% (49/117) in 2025 (*P*<.001).

XSL•FO
RenderX

Figure 1. Comparison of overall user satisfaction in 2020 (n=162) and 2025 (n=144).

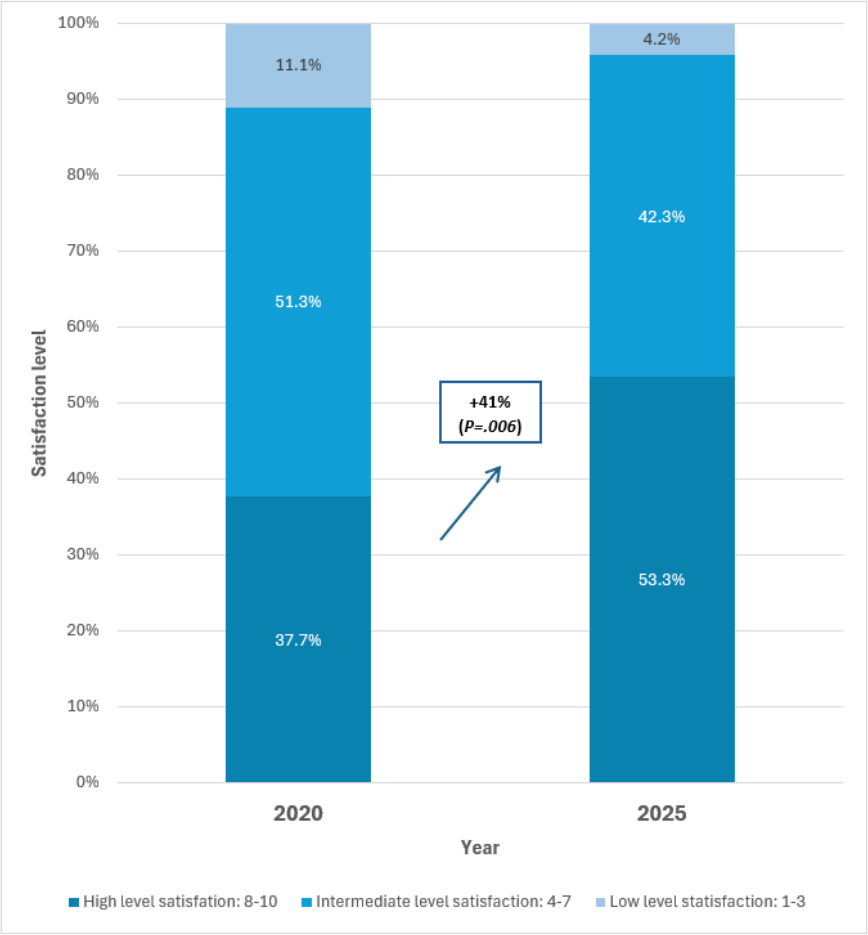
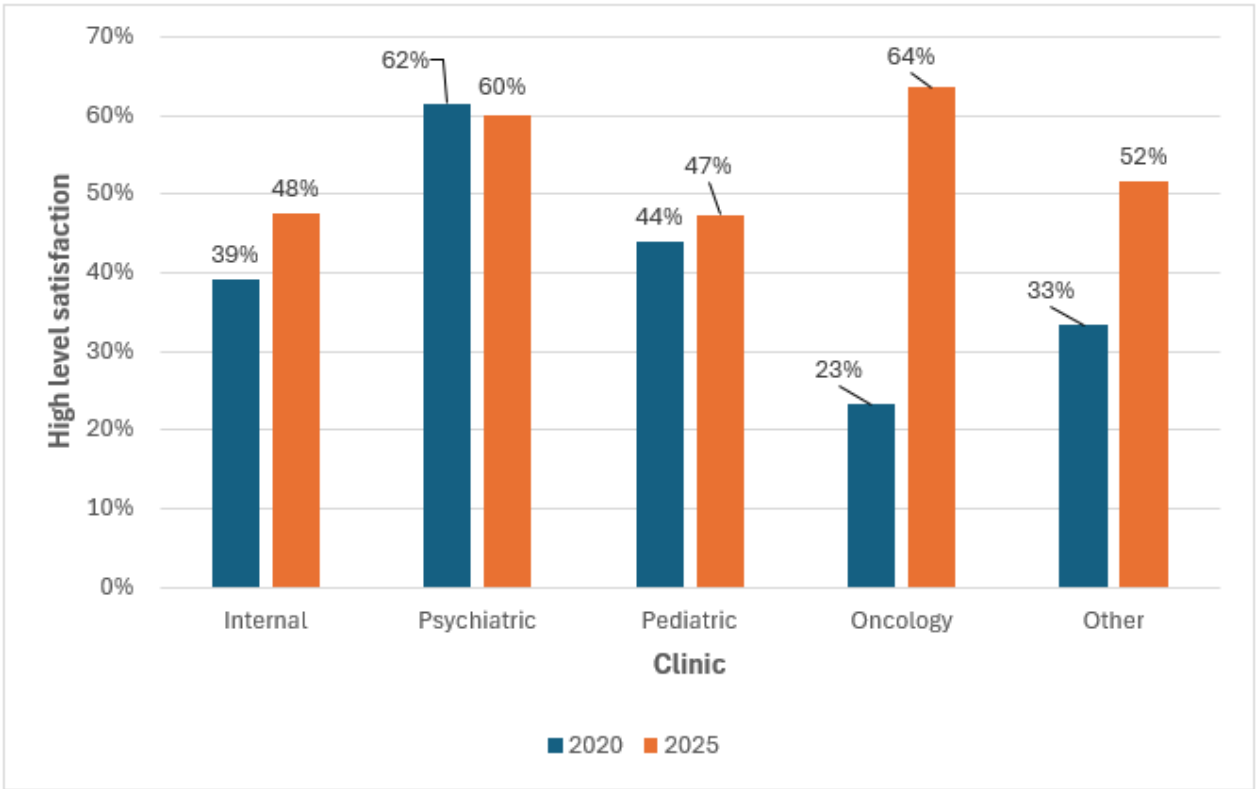


Figure 2. Comparison of numbers of users reporting a high level of satisfaction in different clinics in 2020 and 2025.



XSL•FO
RenderX

Efficiency and Integration Into Workflow

The majority of users (95/144, 66%) reported that telemedicine reduced their workload and only 9% (13/144) felt their workload increased when performing telemedicine visits, with no significant differences observed across clinics. Nurses experienced the greatest benefit, with 90.5% (38/42) reporting reduced workload, compared to 63.8% (95/149) of physicians (P>.05) and 61.4% (51/83) of paramedical staff (P=.04).

Half of the users (73/144, 50.7%) felt that telemedicine visits were shorter and only 3.5% (5/143) of clinicians felt they were longer. Most internal medicine (16/21, 76.2%) and oncology (15/22, 68.2%) clinicians felt telemedicine visits were shorter, and most psychiatric (17/20, 85%) and pediatric (11/19, 57.9%) clinicians felt it took the same time as in-person visits.

Overall, 63.2% (91/144) of telemedicine users reported that patient appointments typically began on time. Punctuality rates were significantly higher in psychiatry (17/20, 85%) and pediatrics (15/19, 78.9%) compared to other specialties (37/62, 59.7%; P=.01). Clinicians who reported having delays (53/144, 36.8%) identified several factors contributing to the delays in telemedicine visits, including clinician-related delays (23/53, 43.4%), patient unavailability at the scheduled time (5/53, 9.4%), technical issues (12/53, 22.6%), and other miscellaneous reasons (13/53, 24.5%).

Video Visits

Of the 143 users performing video visits, 51.7% (n=74) reported actively using SMC's telemedicine platform (Datos Health), while 13.3% (n=19) indicated they had previously used it but discontinued and 35% (n=50) stated they had never used the platform. The most commonly used alternative video platforms were Zoom (Zoom Communications, Inc; 36/104, 34.6%) and WhatsApp (Meta Platforms, Inc; 18/104, 17.3%), with multiple responses permitted. Among the 103 video-visit users who answered the technical feedback questions, the Datos platform was considered very easy to use by 53.4% (55/103) of users, intermediately easy by 23.3% (24/103), and hard to use by 23.3% (24/103).

The proportion of clinicians performing video visits reporting technical or connectivity problems was high. Most clinicians (64/103, 62.1%) experienced problems in connecting to the patients' network, 32% (33/103) experienced connectivity problems in SMC and only 7.8% (8/103) reported never having connectivity problems, with no significant differences among specialties. In fact, 80.6% (83/103) reported that they had to switch to telephone conversation at least once, and 23.3% (24/103) stated that it occurred at least half of the time. Only 19.4% (20/103) of users never had to switch to telephone.

A third (83/274, 30.3%) reported that they had contacted SMC's helpdesk regarding technical difficulties with the video platform; 86.9% (238/274) stated that they were satisfied with the service they received from the helpdesk staff.

Nonusers

In total, 20% (26/130) of nonuser respondents reported they had previously performed telemedicine and stopped. The proportion of clinicians who ceased to perform telemedicine visits was higher in psychiatry compared with other clinics although the absolute numbers were small (6/14, 42.9%). When asked for reasons for not performing telemedicine, of the 98 respondents who answered this question, 81.6% (n=80) stated that it was mostly clinically inappropriate, 23.4% (n=23) stated that they had experienced difficulties, 15.3% (n=15) thought it would interfere with their workflow, 9.2% (n=9) reported that they did not know how to perform telemedicine visits, and 8.2% (n=8) believed that patients would find it less convenient (multiple responses were allowed).

Discussion

Principal Findings

This study aimed to examine clinician engagement and satisfaction with telemedicine following institutional efforts to enhance uptake through technical improvements, administrative streamlining, and clinician support initiatives at SMC's virtual hospital, Sheba BEYOND. The findings demonstrate that while significant progress has been made in improving clinician satisfaction, persistent obstacles continue to affect overall adoption.

The results demonstrate a significant improvement in clinician satisfaction with telemedicine services, with users reporting a high level of satisfaction increasing from 37.7% (61/162) in 2020 to 53.5% (77/144) in 2025 (P=.006), alongside a notable decrease in low levels of satisfaction.

The doubling of telemedicine visits between 2020 and 2024, along with a significant increase in clinician satisfaction levels, highlight a fundamental cultural shift in clinician attitudes toward telemedicine, reflecting the success of Sheba BEYOND's robust infrastructure and multifaceted change management strategies. Similar interventions in other health care systems have successfully increased clinicians' confidence in telemedicine platforms by addressing usability and logistical challenges and providing appropriate support [8]. Oncology clinicians showed the most substantial improvement in satisfaction (from 10/43, 23.3% to 14/22, 63.6%; P=.01), likely due to improvements to interoperability in the oncology clinic. It is possible that the increase is also due to exceptional receptiveness to telemedicine by oncology patients [9]. Satisfaction in psychiatry remained high; prior research has shown that psychiatrists believe most patients to be good candidates for telemedicine and that it is useful as a means to encourage patient engagement [10]. Satisfaction in pediatrics did not see meaningful change. Pediatric physicians may be suspicious of their ability to provide care and build rapport with patients when using telemedicine, a perception reported in other studies on pediatric telemedicine [11,12]. The greater improvement observed by paramedical staff may reflect the longitudinal nature of their clinical activities, which often include patient education and rehabilitation support that may integrate more naturally into telemedicine workflows. Physicians, in contrast, may remain more affected by limitations related to physical examination requirements and diagnostic complexity.

XSL•FO
RenderX

The sustained increase in telemedicine visits and improved satisfaction mirror findings from previous studies highlighting the importance of institutional frameworks in fostering a cultural shift toward telemedicine. For example, the study by Ramaswamy et al [4] reported similar trends in higher satisfaction and increased adoption when supported by extensive training, workflow compatibility, and leadership-driven initiatives. Sheba BEYOND's multifaceted strategies align closely with global approaches to managing change in telemedicine adoption after the COVID-19 pandemic [8].

The results highlight a significant disparity in perceptions between telemedicine users and nonusers, with 87% (141/162) of users believing telemedicine benefits their patients compared to only 41.9% (49/117) of nonusers. Interestingly, confidence in telemedicine's utility among nonusers decreased significantly, from 68% (34/50) in 2020 to 41.9% (49/117) in 2025 (P<.001). This decrease may not reflect a true change in perspective but rather be related to the fact that over time, clinicians who believe in the utility of telemedicine have gradually embraced it, leaving the "nonuser" group with those who continue to be skeptical about this mode of health care delivery. Nonusers repeatedly cited clinical inappropriateness (80/98, 81.6%), difficulties with workflow integration (15/98, 15.3%), and technical problems (23/98, 23.5%) as primary reasons for opting out of telemedicine services. The literature supports these concerns, noting that telemedicine may be less applicable to high-complexity or tactile specialties such as surgery or certain subspecialty consults [13].

These findings align with barriers highlighted in the study by McMaster et al [14], which identified difficulties in integrating telemedicine into traditional workflows and its appropriateness for certain specialties as key obstacles to clinician adoption. Two-thirds (95/144, 66%) of telemedicine users reported reduced workloads, and 50.7% (73/144) perceived telemedicine visits as shorter than in-person consultations. Internal medicine (16/21, 76.2%) and oncology (15/22, 68.2%) clinicians were particularly likely to report shorter visits, whereas psychiatry (17/20, 85%) and pediatrics (11/19, 57.9%) found visit durations comparable to in-person care. Such efficiency gains are consistent with studies showing that telemedicine facilitates streamlined care delivery [15].

However, adherence to scheduled appointment times remained a challenge, with only 63.2% (91/144) of telemedicine users reporting that visits typically began on time. Punctuality rates varied across specialties, with psychiatry (17/20, 85%) and pediatrics (15/19, 78.9%) performing better than others. Clinician-related delays (23/53, 43.4%) and technical problems (12/53, 22.6%) were identified as the main contributors to scheduling inefficiencies, emphasizing the need for ongoing optimization of workflows and technology.

A significant proportion of telemedicine users performing video visits reported connectivity issues, with 62.1% (64/103) experiencing problems with patients' networks and 32% (33/103) encountering connectivity problems within SMC's network. This constrained user experience led to 80.6% (83/103) of clinicians resorting to telephone consultations at least once, undermining the platform's reliability. While SMC's telemedicine platform was deemed very easy to use by 53.4% (55/103) of users, alternative platforms such as Zoom and WhatsApp enjoyed widespread use, albeit with varying consistency. SMC encourages the use of its institutional telemedicine platform but does not restrict clinicians from using alternative video communication platforms. The study design did not allow assessment of the reasons underlying clinicians' preference for telephone versus video consultations. We speculate that some clinicians may have favored telephone visits due to their simplicity and reduced vulnerability to connectivity issues, while others may have preferred video visits because visual interaction was important for clinical assessment and patient communication.

The literature emphasizes the importance of stable and user-friendly telemedicine platforms in ensuring clinician satisfaction and adoption. A systematic review by Pogorzelska and Chlabicz [3] identified platform usability and seamless integration into clinical workflows as critical factors influencing telemedicine success during the pandemic. For institutions with large-scale telemedicine implementation, customized solutions that address specialty-specific needs, such as pediatric consultations or psychiatric evaluations, can significantly enhance adoption.

Institutional efforts to support telemedicine adoption must go beyond technology upgrades and training. Telemedicine adoption may be encouraged or hindered by specialty-specific concerns. To ensure wider engagement, there is a need to address clinician skepticism regarding the clinical appropriateness of telemedicine and to tailor solutions to their specialty needs and circumstances. SMC's institutional efforts to expand telemedicine adoption, specifically in the nonuser group, continue to focus on specialty-specific workflow adaptation, training initiatives, and hybrid care model development, while recognizing that telemedicine may remain less suitable for certain clinical scenarios and specialties.

Additionally, addressing technical barriers is paramount. Streamlined software platforms, increased access to stable internet connections, and ongoing support are essential to improving reliability and user satisfaction [8]. Patient education and technological assistance can also mitigate connectivity issues, helping reduce clinician workload related to troubleshooting technical problems during visits.

This study has several limitations. Its 10% response rate may limit representativeness, as clinicians more actively engaged with telemedicine may have been more likely to respond. The study was conducted at a single tertiary medical center, which may limit generalizability. In addition, the study design did not allow correlation between clinician satisfaction and objective telemedicine use metrics at the individual level. Furthermore, patient experience was outside the scope of this clinician-focused survey. Future studies should aim to address these gaps with broader sampling and data regarding clinician uptake.

Conclusions

This study demonstrates the positive impact of institutional change management strategy in promoting clinician engagement with telemedicine services. The doubling of telemedicine visits

XSL•FO
RenderX

and significant improvements in clinician satisfaction reflect an adaptive cultural shift, driven by organizational commitment and infrastructure development. However, persistent technical barriers and varying specialty-specific engagement highlight opportunities for refinement. Addressing these challenges through targeted interventions is crucial to achieve sustained engagement and maximize telemedicine's potential in outpatient care.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Funding

The authors declared no financial support was received for this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questionnaire (translated to English).
[DOCX File , 31 KB - humanfactors_v13i1e93911_app1.docx]

References

1. Ohannessian R, Duong TA, Odone A. Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: a call to action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
2. Barkai G, Gadot M, Amir H, Menashe M, Shvimer-Rothschild L, Zimlichman E. Patient and clinician experience with a rapidly implemented large-scale video consultation program during COVID-19. Int J Qual Health Care 2021 Feb 20;33(1):mzaa165 [FREE Full text] [doi: 10.1093/intqhc/mzaa165] [Medline: 33313891]
3. Pogorzelska K, Chlabicz S. Patient satisfaction with telemedicine during the COVID-19 pandemic-a systematic review. Int J Environ Res Public Health 2022 May 17;19(10):6113 [FREE Full text] [doi: 10.3390/ijerph19106113] [Medline: 35627650]
4. Ramaswamy A, Yu M, Drangsholt S, Ng E, Culligan PJ, Schlegel PN, et al. Patient satisfaction with telemedicine during the COVID-19 pandemic: retrospective cohort study. J Med Internet Res 2020 Sep 09;22(9):e20786 [FREE Full text] [doi: 10.2196/20786] [Medline: 32810841]
5. Andrews E, Berghofer K, Long J, Prescott A, Caboral-Stevens M. Satisfaction with the use of telehealth during COVID-19: an integrative review. Int J Nurs Stud Adv 2020 Nov;2:100008 [FREE Full text] [doi: 10.1016/j.ijnsa.2020.100008] [Medline: 33083791]
6. Nguyen M, Waller M, Pandya A, Portnoy J. A review of patient and provider satisfaction with telemedicine. Curr Allergy Asthma Rep 2020 Sep 22;20(11):72 [FREE Full text] [doi: 10.1007/s11882-020-00969-7] [Medline: 32959158]
7. Likert R. A technique for the measurement of attitudes. Arch Psychol 1932;22(140):1-55 [FREE Full text]
8. Pong C, Tseng RM, Tham YC, Lum E. Current implementation of digital health in chronic disease management: scoping review. J Med Internet Res 2024 Dec 12;26:e53576 [FREE Full text] [doi: 10.2196/53576] [Medline: 39666972]
9. West HJ, Bange E, Chino F. Telemedicine as patient-centred oncology care: will we embrace or resist disruption? Nat Rev Clin Oncol 2023 Oct;20(10):659-660. [doi: 10.1038/s41571-023-00796-5] [Medline: 37365437]
10. Uscher-Pines L, Parks AM, Sousa J, Raja P, Mehrotra A, Huskamp HA, et al. Appropriateness of telemedicine versus in-person care: a qualitative exploration of psychiatrists' decision making. Psychiatr Serv 2022 Aug 01;73(8):849-855 [FREE Full text] [doi: 10.1176/appi.ps.202100519] [Medline: 35080419]
11. Pakanati V, Raol N, Ching Siong T, Govil N. Patient and physician satisfaction with telemedicine in pediatric otolaryngology. Int J Pediatr Otorhinolaryngol 2022 May;156:111097. [doi: 10.1016/j.ijporl.2022.111097] [Medline: 35259634]
12. Finkelstein JB, Tremblay ES, Van Cain M, Farber-Chen A, Schumann C, Brown C, et al. Pediatric clinicians' use of telemedicine: qualitative interview study. JMIR Hum Factors 2021 Dec 02;8(4):e29941. [doi: 10.2196/29941] [Medline: 34860669]
13. Martinez RN, Hogan TP, Lones K, Balbale S, Scholten J, Bidelspach D, et al. Evaluation and treatment of mild traumatic brain injury through the implementation of clinical video telehealth: provider perspectives from the Veterans Health Administration. PM R 2017 Mar;9(3):231-240. [doi: 10.1016/j.pmrj.2016.07.002] [Medline: 27423365]
14. McMaster T, Mori K, Lee S, Manasa S, Stelmach W, To H. Innovations and implementation of telemedicine in surgical clinics beyond COVID-19: a narrative review. Telemed J E Health 2023 Jan;29(1):50-59. [doi: 10.1089/tmj.2021.0409] [Medline: 35736794]

XSL•FO
RenderX

15. Mullen-Fortino M, Rising KL, Duckworth J, Gwynn V, Sites FD, Hollander JE. Presurgical assessment using telemedicine technology: impact on efficiency, effectiveness, and patient experience of care. Telemed J E Health 2019 Feb;25(2):137-142. [doi: 10.1089/tmj.2017.0133] [Medline: 30048210]

Abbreviations

SMC: Sheba Medical Center

Edited by A Kushniruk; submitted 21.Feb.2026; peer-reviewed by B Cross, C Doarn; comments to author 04.May.2026; revised version received 20.May.2026; accepted 23.Jun.2026; published 09.Sep.2026.

Please cite as:
Barkai G, Edelman-Yaron S, Amir H, Inbar-Lankry H, Sarafraz R
Examining Telemedicine Adoption After the COVID-19 Pandemic and Evaluating the Impact of a Virtual Hospital on Clinician Engagement: Longitudinal Survey Study
JMIR Hum Factors 2026;13:e93911
URL: https://humanfactors.jmir.org/2026/1/e93911
doi:10.2196/93911
PMID:

©Galia Barkai, Shany Edelman-Yaron, Hadar Amir, Havi Inbar-Lankry, Rachel Sarafraz. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 09.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Usability and Workflow Integration of a Machine Learning–Derived Neonatal Risk Predictor in Kenyan Neonatal Units: Multisite User-Centered Pilot Evaluation

Ronald Danny Nyatuka[1], PhD; Paul Macharia[1], PhD; Esther Kakhata[1], PhD; Faith Siva[1], M.Sc; Betsy Muriithi[1], PhD; Md Shafiqur Rahman Jabin[2,3], PhD

[1]School of Computing and Engineering Sciences, Strathmore University, Nairobi, Nairobi County, Kenya

[2]Department of Medicine and Optometry, Linnaeus University, Kalmar, Kalmar, Sweden

[3]Faculty of Health and Social Care, University of Bradford, Bradford, England, United Kingdom

Corresponding Author:
Md Shafiqur Rahman Jabin, PhD
Department of Medicine and Optometry
Linnaeus University
Universitetsplatsen 1
Kalmar, Kalmar, 39231
Sweden
Phone: 44 7915673612
Email: mdshafiqur.rahmanjabin@lnu.se

Abstract

Background: Neonatal mortality remains a leading contributor to under-5 deaths globally, particularly in low- and middle-income countries (LMICs). While machine learning (ML)–based risk prediction models show promise for identifying high-risk neonates, published evidence describing the real-world usability and practical implementation of ML-derived neonatal risk prediction tools within routine clinical workflows in LMIC neonatal units remains limited.

Objective: This study aimed to evaluate the usability, user experience, and perceived clinical utility of a paper-based neonatal risk predictor tool derived from an ML model and implemented across 3 Kenyan health facilities.

Methods: A postimplementation, cross-sectional usability evaluation was conducted following a 4-month implementation period from August through November 2025. The study was embedded within a longitudinal mixed methods project. Frontline neonatal health care workers (n=10) completed standardized usability instruments, including adapted global usability items (System Usability Scale [SUS]), selected Questionnaire for User Interaction Satisfaction (QUIS) domains, and the Post-Study System Usability Questionnaire (PSSUQ), alongside a project-specific Post-Study Neonatal Utility Questionnaire (PSNUQ). Descriptive statistics (medians, IQRs, and category percentages) were computed. A total of 3 purposively selected neonatal unit leaders participated in semistructured key informant interviews (KIIs), which were analyzed using thematic analysis. Quantitative and qualitative findings were triangulated to contextualize perceptions of usability.

Results: Among participating frontline health care workers, overall perceptions of usability were generally favorable. Around 75% (6/8) of respondents reported being willing to use the tool frequently, and 55% (5/9) indicated confidence in using it independently. Around half (5/10) disagreed that the tool was complex, while 22% (2/9) agreed, indicating moderate polarization in perceived complexity. Median PSSUQ composite scores were below 3 across subscales, reflecting positive usability ratings. Overall, 8 of 10 (80%) respondents agreed that the tool supports early identification of high-risk neonates and improves care prioritization within the first 48 hours. However, workflow integration was workload-sensitive: 40% (4/10) reported an increased documentation burden during periods of high patient volume. KIIs identified staffing shortages, parallel documentation systems, and the importance of administrative endorsement as key structural influences on adoption.

Conclusions: The neonatal risk predictor tool demonstrated acceptable usability, learnability, and perceived clinical relevance across 3 diverse Kenyan facilities. However, variability in perceived complexity and workload sensitivity highlights the importance of structured onboarding, workflow-aligned integration, and context-aware implementation planning. These findings underscore that translating ML-derived predictor models into clinical practice requires not only technical validity but also strong usability and system-level readiness within routine neonatal care settings.

XSL•FO
RenderX

International Registered Report Identifier (IRRID): RR2-10.2196/81996

(JMIR Hum Factors 2026;13:e94828) doi:10.2196/94828

KEYWORDS

health technology adoption; frontline health care workers; decision-support systems; early-life morbidity; cognitive load theory; documentation burden; health care quality improvement; mixed-methods evaluation; health systems strengthening; low-resource settings

Introduction

Neonatal mortality remains a major global health challenge, accounting for almost half of all under-5 deaths worldwide, with an estimated 2.3 million newborn deaths in 2022 [1]. Despite global progress in reducing child mortality, declines in neonatal deaths have been slower and progress uneven across regions, making early identification and management of high-risk newborns a priority for achieving Sustainable Development Goal (SDG) Target 3.2 [2,3].

Structured, timely risk identification at birth and during the first 48 hours is critical because many life-threatening conditions (such as prematurity-related complications, intrapartum injury, and early-onset infection) present within this window [3]. In low- and middle-income country (LMIC) settings, reliance on subjective clinical judgment and inconsistent documentation often reduces the reliability of early triage, contributing to delays in escalation of care [3]. These realities create a pressing need for practicable risk-stratification tools that fit routine neonatal workflows and improve early detection of clinical deterioration.

Machine learning (ML) and other data-driven approaches show promise in neonatal risk prediction, with systematic reviews demonstrating that ML models can predict neonatal mortality and other adverse perinatal outcomes using routinely collected predictors [4-6]. However, most published ML models remain limited by modest external validation, variable predictor availability across settings, and sparse evidence on real-world deployment [4,5]. This gap between technical model development and safe operational deployment has also been documented in broader health information technology (HIT) literature, where system functionality issues, configuration errors, and workflow misalignment have resulted in unintended care disruptions [7-10]. Consequently, translating ML predictors into clinical practice requires 2 parallel lines of evidence: (1) demonstration of acceptable technical performance, and (2) evidence that the predictor tool is usable, acceptable, and practically feasible for frontline health care workers within the constraints of their working environments [6,11-13].

Usability and user experience (UX) are central determinants of health-technology adoption: internationally recognized standards define usability in terms of effectiveness, efficiency, and satisfaction in a specific context of use [9,14], and empirical work shows that poor usability undermines both uptake and the intended effect of clinical decision-support tools [14,15]. Evidence from national incident-report analyses indicates that poorly integrated digital systems often cause workflow disruptions, documentation inconsistencies, and patient safety risks when usability and contextual fit are not adequately addressed [8-10,16]. In health care, and especially in high-stress, time-sensitive clinical areas such as neonatal units, even small frictions in interface clarity, learnability, or workflow fit can materially reduce consistent use and thereby erode potential clinical benefit [9,14,15].

User-centered design (HCD) approaches that iteratively involve end-users in design, testing, and refinement increase the probability of adoption and sustained use of digital health tools in LMICs [17,18]. Prior work examining digital health testbeds and system configuration failures underscores the importance of early-stage contextual testing and iterative refinement to prevent downstream operational disruptions [9,10]. HCD emphasizes aligning an intervention's information architecture, terminology, and workflow demands with the cognitive load and priorities of frontline staff, an especially important requirement when staffing shortages and high nurse-to-patient ratios constrain the time available for documentation [17]. For neonatal risk predictors, HCD therefore implies not only designing an accurate predictor tool but also ensuring the tool's presentation, learning curve, and integration into routine tasks are acceptable to users [18-20].

Despite the theoretical consensus, there remains remarkably little real-world evidence on the usability and perceived utility of neonatal risk predictor tools in LMIC neonatal units [11]. Existing studies of ML-based neonatal prediction largely focus on model building and internal validation, with few studies examining frontline user perceptions, interface quality, and the practical trade-offs staff make when integrating new documentation tasks into busy clinical workflows [3,5]. This gap is consequential: without evidence that frontline users can and will use a predictor tool reliably, implementation at scale risks failure even if the underlying model performs well on retrospective data [3,5].

This study presents the completed user-centered evaluation of a neonatal risk predictor tool derived from an ML model [4] and implemented as a paper-based risk tool across 3 Kenyan health facilities, following the methodology outlined in the previously published study protocol [21]. Our objectives were to (1) assess perceived usability and user experience using validated instruments and bespoke UX items, (2) describe perceived usefulness and likely impact on routine neonatal workflows, and (3) identify design and training priorities to strengthen adoption in resource-constrained neonatal units. By focusing on frontline user perspectives and practical workflow constraints, the paper aims to provide the implementation-oriented evidence needed to move from ML model development toward safe, usable, and context-appropriate deployment.

XSL•FO
RenderX

Methods

Overview

The study adhered to good practices for pragmatic usability assessment in clinical settings. The study is reported in accordance with the Good Reporting of A Mixed Methods Study (GRAMMS) recommendations [22], and the completed GRAMMS checklist is provided as Multimedia Appendix 1.

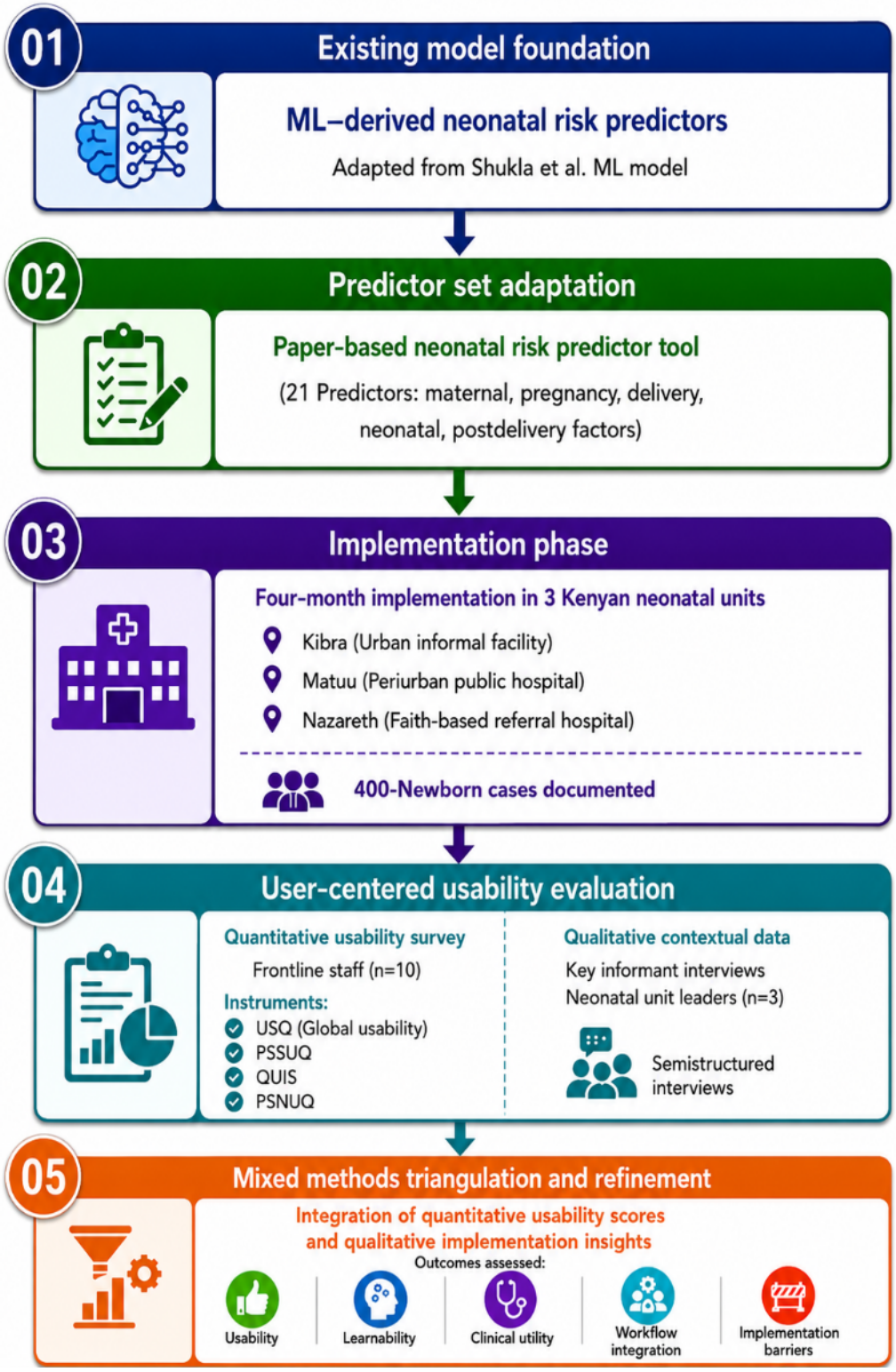
Study Design

This study used a postimplementation, cross-sectional usability, and UX evaluation embedded within a longitudinal mixed-methods project. The use of mixed methods triangulation aligns with established approaches for evaluating complex health care interventions, combining quantitative usability metrics with qualitative contextual insight to enhance interpretive depth [23-25]. The evaluation combined standardized usability instruments with a project-specific Post-Study Neonatal Utility Questionnaire (PSNUQ) and semistructured key informant interviews (KIIs) to triangulate frontline users' perceptions with contextual implementation data. The quantitative usability component reported here was administered after a 4-month implementation period of the paper-based neonatal risk predictor tool. The study followed good practice for pragmatic usability assessment in clinical settings [14].

This study represents the completed user-centered usability evaluation described in the previously published study protocol [21]. This paper reports the implementation and postimplementation usability findings from that protocol-defined study. Preintervention key informant interviews were conducted in August 2025, followed by implementation of the paper-based neonatal risk predictor and newborn data collection from August through November 2025. The postimplementation usability survey was administered in December 2025. The overall study design and mixed-methods usability evaluation framework are illustrated in Figure 1.

XSL•FO
RenderX

Figure 1. Study design and mixed methods usability evaluation framework. ML: machine learning; PSNUQ: Post-Study Neonatal Utility Questionnaire; PSSUQ: Post-Study System Usability Questionnaire; QUIS: Questionnaire for User Interaction Satisfaction; USQ: Usability Scale Questionnaire.



The ML–derived neonatal risk predictor tool was implemented as a paper-based documentation instrument across 3 Kenyan neonatal units over a 4-month period. Following implementation, usability and user experience were evaluated using a mixed methods approach that combined standardized usability questionnaires (System Usability Scale [SUS], Post-Study System Usability Questionnaire [PSSUQ], Questionnaire for User Interaction Satisfaction [QUIS], and PSNUQ) completed by frontline neonatal staff (n=10) with qualitative KIIs with neonatal unit leaders (n=3). Quantitative and qualitative findings were triangulated to assess usability, learnability, clinical utility, workflow integration, and implementation context.

Study Setting

The study was conducted in 3 Kenyan health facilities representing distinct service contexts: an urban informal

XSL•FO
RenderX

settlement facility in Kibra (Nairobi County), a periurban public hospital in Matuu (Machakos County), and a faith-based referral hospital in Nazareth (Kiambu County). Facilities were located in Nairobi County (Kibra), Machakos County (Matuu), and Kiambu County (Nazareth). Contextual details and facility characteristics are described in the project report and site summaries.

Participants and Sampling

The published study protocol specified a purposive sample of 10 health care workers across the 3 participating facilities, with a planned allocation of 4 participants from Kibera, 4 from Matuu, and 2 from Nazareth [21]. In the completed study, however, recruitment was based on the availability of eligible frontline neonatal health care workers during the implementation period. All eligible and available staff involved in the implementation were invited to participate, and 10 frontline staff completed the postimplementation survey.

The protocol's prespecified allocation of 4 participants from Kibra, 4 from Matuu, and 2 from Nazareth was intended to ensure representation across the 3 participating facilities. The allocation was a planned recruitment target rather than a statistically derived proportional allocation. In the completed study, the final numbers of participants from each facility corresponded to these planned numbers, although recruitment was based on the availability of eligible frontline neonatal health care workers during the implementation period.

A total of 2 participant groups contributed to the usability evaluation:

1. Key informants (qualitative KIIs): Purposive sampling was used to select senior neonatal unit leaders (one per facility) for semistructured interviews exploring implementation context, governance structures, workflow integration, and perceived barriers and enablers to adoption. A total of 3 KIIs were conducted across the participating facilities. These interviews have also been reported in a separate published analysis examining governance and health system determinants of scalability and institutionalization of the neonatal risk stratification tool [26]. In this study, the same interview data were used to contextualize the user-centered evaluation, with analysis focused specifically on perceptions of usability, workflow integration, and implementation-related factors relevant to the frontline use of the paper-based predictor tool. The semistructured interview guide is provided in Multimedia Appendix 2.
2. Frontline neonatal health care workers (quantitative postsurvey): The 3 participating facilities had a combined workforce of 43 neonatal health care workers: 11 at Kibra, 15 at Matuu, and 17 at Nazareth. All eligible neonatal nurses and midwives involved in the 4-month implementation of the tool were invited to complete the post-implementation survey. Due to staff availability during the study period, 10 frontline health care workers participated in the survey. Participants were distributed across the 3 facilities as follows: 4 from Matuu, 4 from Kibra, and 2 from Nazareth.

All participants provided information and consent for the use of anonymized aggregate responses in research outputs.

Development of the ML-Derived Neonatal Risk Predictor

The paper-based neonatal risk predictor evaluated in this study was derived from a previously developed and validated ML model for neonatal mortality prediction reported by Shukla et al [4]. The original model was developed using routinely collected maternal, obstetric, delivery, neonatal, and early postnatal variables and demonstrated good predictive performance in LMIC settings (area under the receiver operating characteristic curve >0.80). Rather than evaluating the predictive performance of the ML algorithm itself, this study focused on the feasibility, usability, user experience, and workflow integration of a paper-based implementation of the predictor set within routine neonatal care across 3 Kenyan health facilities.

The complete list of predictor variables and the paper-based neonatal risk predictor form used during implementation are provided in Multimedia Appendix 2.

Intervention and Exposure

The object of evaluation was a paper-based neonatal risk predictor tool (the "predictors set") that had been integrated into routine documentation for newborns at delivery (Day 1) and at a postdelivery check (Day 2). The tool captured 21 candidate predictors, organized into 5 domains: maternal characteristics, pregnancy-related factors, delivery characteristics, neonatal clinical indicators, and early postdelivery observations. The complete list of predictors and the paper-based neonatal risk predictor form are provided in Multimedia Appendix 2 to enhance transparency and reproducibility. During the 4-month implementation period, the tool was used prospectively within routine neonatal care, resulting in 400 newborn cases recorded across the 3 participating sites: 100 at Kibra, 100 at Nazareth, and 200 at Matuu. These site-specific totals represent the cases captured during the implementation period and were not determined through a statistically derived per-site sampling allocation.

Instruments

Standard Instruments

We combined well-established usability measures to capture complementary aspects of usability and user experience:

1. PSSUQ: The PSSUQ (James R Lewis) is a validated instrument designed to measure users' perceived system usability after task completion; the most commonly used short form has 16 items across dimensions such as system usefulness, information quality, and interface quality, typically scored on a 7-point Likert scale (lower scores = better perceived usability). The PSSUQ has been psychometrically evaluated and widely used in health technology research [15,27]. PSSUQ scores were summarized using the standard System Usefulness, Information Quality, and Interface Quality subscales, together with the overall score.
2. QUIS: The QUIS is a user satisfaction instrument developed by Chin et al [28] that measures overall reaction and

XSL•FO
RenderX

multiple interface dimensions (eg, screen factors, terminology, and learnability). We used QUIS items to assess interaction satisfaction and perceptions of content and layout [28]. Selected QUIS domains assessed overall user satisfaction, ease of use, interface characteristics, and learning.

3. SUS: While our project labels one instrument "SUS" in the field documents, the SUS is the canonical short usability scale commonly used to obtain a global usability score (10 items and 5-point Likert) [29]. In our implementation, we used a short version of SUS, adapted from standard global usability measures (SUS and PSSUQ constructs), to capture overall perceived ease of use and willingness to reuse.

Project-Specific Instrument Content (PSNUQ and SUS Items)

The PSNUQ includes domains on perceived usefulness, learnability, task support, content quality, interface quality, and perceived productivity and impact. The SUS items in the survey covered willingness to use, perceived complexity, need for assistance, and overall ease of use. The QUIS subscales assessed overall reaction, layout and text clarity, terminology, and interaction satisfaction.

Questionnaire Administration and Data Collection Procedures

Postsurvey questionnaires were administered face-to-face by the research team to frontline neonatal staff at the end of the implementation period. Respondents completed the combined instrument (SUS+QUIS+PSNUQ) in a private setting; the survey took approximately 20-30 minutes. KIIs with neonatal unit leads were conducted using a semistructured guide, with audio recording with permission; interview transcripts were used for thematic analysis and triangulation.

The complete questionnaires, including all item wording, response anchors, scoring procedures, and the PSSUQ subscale allocation, are provided in Multimedia Appendix 2. The SUS, PSSUQ, and selected QUIS domains are validated usability instruments, whereas the PSNUQ was developed for this study to evaluate perceived clinical relevance, feasibility, and implementation considerations of the neonatal risk predictor tool.

Scoring and Quantitative Analysis

Survey items used Likert-style ordinal response scales (typically 5- or 7-point, depending on the instrument and item). For standardized instruments (PSSUQ, QUIS, and SUS-derived items), we followed canonical scoring procedures where applicable (eg, SUS scoring and PSSUQ subscales) and computed descriptive statistics (median, IQR, and category percentages) for each item and subscale. Given the small sample size (n=10), analyses were descriptive, focusing on response distributions, identifying polarized items, and cross-tabulations by facility or cadre where numbers allowed. We also computed simple case-level completeness and item-level missingness for the survey dataset. These methods align with pragmatic usability evaluation approaches in clinical contexts [14,15].

Qualitative Analysis and Triangulation

KIIs were audio-recorded with participant consent and transcribed verbatim. The transcripts were analyzed using thematic analysis through an iterative coding process, with coding conducted manually using a structured coding framework informed by the interview guide and patterns emerging from the data [30,31]. EK and FS conducted the initial coding, while RDN and MSRJ independently reviewed the coding and emerging themes. Coding discrepancies were discussed among the coders and resolved through consensus. The analysis focused on identifying themes relevant to usability, workflow integration, and implementation context, with the qualitative findings used primarily to contextualize and triangulate the quantitative usability findings rather than as a standalone qualitative outcome analysis. In particular, qualitative insights helped explain divergent or polarized survey responses, including concerns related to staffing shortages, documentation duplication, workload pressure, infrastructure limitations, and resistance to workflow changes. The researchers acknowledged that interpretation of the qualitative data was informed by their experience in digital health and neonatal care, and efforts were made to minimize interpretive bias through regular discussion of emerging themes. Given the exploratory nature of this pilot study and the inclusion of one key informant from each participating facility, thematic saturation was not formally assessed. Representative deidentified quotations supporting the themes are provided in Multimedia Appendix 3. Findings from the quantitative postsurvey (SUS, QUIS, and PSNUQ components) and the qualitative KIIs were then triangulated to generate integrated interpretations. This triangulation enabled identification of both tool-specific usability issues (eg, clarity, learnability, and layout) and system-level constraints affecting adoption (eg, staffing capacity, governance processes, and infrastructure readiness). The triangulated findings informed practical recommendations for design refinement and implementation planning.

The reporting checklist, complete postimplementation survey instrument (SUS, selected QUIS domains, PSSUQ items, and PSNUQ), semistructured interview guide, and the paper-based neonatal risk predictor form are provided as Multimedia Appendices 1-3 to support transparency and reproducibility.

Ethical Considerations

Ethical approval (SU-ISERC2949/25) was obtained from the Strathmore University Institutional Scientific Ethics Review Committee (SU-ISERC), Nairobi, Kenya, on 28 July 2025, prior to commencement of the study. All participants provided informed consent before participation after receiving information about the study aims, voluntary participation, confidentiality protections, and data management procedures. Interview transcripts were anonymized before analysis to protect participant confidentiality and institutional privacy. The study adhered to established ethical principles for qualitative health systems research throughout data collection, analysis, and reporting.

XSL•FO
RenderX

Results

Participant Characteristics

A total of 10 frontline neonatal health care workers completed the postintervention usability evaluation across the 3 study sites. The majority of respondents were mid-career nurses with diploma-level qualifications, reflecting the cadre primarily responsible for routine neonatal documentation and triage. The characteristics and study-site distribution of the participating frontline neonatal health care workers are summarized in Table 1.

Table 1. Characteristics of frontline neonatal health care worker participants (n=10).

Characteristics	Values, n (%)[a]
Study site (n=10)	
Kibra	4 (40)
Matuu	4 (40)
Nazareth	2 (20)
Professional cadre (n=10)	
Nurses	8 (80)
Midwives	2 (20)
Age (years; n=9)	
18-30	2 (22.2)
31-40	6 (66.7)
51-60	1 (11.1)
Educational qualification (n=9)	
Certificate	7 (77.8)
Bachelor's degree	2 (22.2)

[a]Percentages are calculated using the number of respondents for each characteristic. One participant had missing responses for age and educational qualification; therefore, the denominator for these variables is 9 rather than 10.

Overall Usability Perception: Global Usability (SUS Items)

Among participating frontline health care workers, overall perceptions of usability were generally favorable. Most respondents reported positive overall perceptions of the neonatal risk predictor tool's usability. Specifically, 75% (6/8) reported a willingness to use the tool frequently in routine practice, and 55% (5/9) indicated confidence in using it independently. Regarding perceived complexity, around half (5/10) disagreed that the tool was complex, while 22% (2/9) agreed. Together, these findings suggest the tool's overall acceptability, with moderate variability in users' perceptions of complexity.

PSSUQ Composite Scores

PSSUQ subscales were computed using canonical scoring procedures (lower scores indicate better usability). Median scores across all subscales were below 3, indicating favorable perceptions of usability.

Information quality and interface quality showed slightly greater dispersion, reflecting variability in respondents' perceptions of clarity and layout. Median PSSUQ subscale scores are presented in Table 2. All median scores were below 3, indicating generally favorable perceptions of system usability.

Table 2. PSSUQ subscale scores.

Subscale	Values, median (IQR)[a]
System Usefulness	2.1 (1.5-2.8)
Information Quality	2.3 (1.7-3.0)
Interface Quality	2.4 (1.8-3.2)
Overall PSSUQ[b] score	2.3 (1.6-3.0)

[a]PSSUQ responses were available for 10 participants (n=10). Scores are presented as median (IQR). Lower scores indicate greater perceived usability.
[b]PSSUQ: Post-Study System Usability Questionnaire.

XSL•FO
RenderX

User Experience Domains (QUIS and PSNUQ)

Content and Layout Clarity

A majority of respondents (8/10, 80%) agreed that the content of the tool was clearly structured, and the same proportion (8/10, 80%) indicated that the terminology was understandable. Additionally, 7 respondents (7/10, 70%) agreed that the information was logically grouped. However, 3 respondents (3/10, 30%) expressed concerns regarding potential redundancy with existing documentation forms.

Learnability and Confidence

A total of 8 respondents (8/10, 80%) agreed that the neonatal risk predictor tool was easy to learn after initial exposure, whereas 2 respondents (2/10, 20%) reported some initial difficulty during early use. Regarding confidence in using the tool independently, 6 respondents (6/10, 60%) indicated that they felt confident completing the tool without assistance following the implementation period, while 4 respondents (4/10, 40%) reported requiring occasional clarification or support during early stages of use.

In addition, 7 respondents (7/10, 70%) reported that their comfort level with the tool increased over the 4-month implementation period, suggesting progressive familiarization with the structured documentation format. These findings indicate that while initial onboarding may require guidance for a subset of users, perceived learnability improved with continued exposure.

Perceived Clinical Utility

PSNUQ responses demonstrated high perceived clinical utility of the neonatal risk predictor tool. A total of 8 respondents (8/10, 80%) agreed that the tool supports early identification of high-risk neonates, while 7 respondents (7/10, 70%) agreed that it enhances structured clinical thinking during neonatal assessment. Additionally, 8 respondents (8/10, 80%) reported that the tool may improve care prioritization within the first 48 hours of life.

Workflow Integration

Although the tool was perceived as clinically helpful, integration into the routine workflow was described as dependent on staffing capacity. Overall, 6 respondents (6/10, 60%) indicated that the tool was manageable under normal staffing conditions, whereas 4 respondents (4/10, 40%) reported that high patient volumes increased the perceived documentation burden. Among those reporting higher-workload environments, 3 respondents (3/10, 30%) specifically noted that competing clinical priorities made consistent completion more challenging.

Polarization and Divergent Perceptions

Item-level distribution analysis demonstrated moderate polarization on complexity-related items. Item-level responses demonstrated moderate variation in perceptions of tool complexity. A total of 5 respondents (5/10, 50%) disagreed or strongly disagreed that the tool was complex, 3 (3/10, 30%) selected a neutral response, and 2 (2/10, 20%) agreed or strongly agreed.

Triangulated Interpretation

Qualitative insights from KIIs provided contextual interpretation of the usability findings. All 3 informants (n=3, 100%) identified staffing shortages as a structural factor influencing perceived documentation burden. A total of 2 informants (n=3, 67%) reported that parallel documentation requirements contributed to concerns about duplication, particularly in facilities where existing paper records must be maintained alongside new tools. Additionally, 2 informants (n=3, 67%) emphasized that visible administrative endorsement influenced staff motivation to adopt and consistently use the tool.

Triangulation of quantitative survey findings with qualitative themes indicates that perceived usability was shaped not only by interface clarity and learnability but also by broader structural workflow conditions within the neonatal units.

In Summary, the neonatal risk predictor tool demonstrated favorable usability performance across multiple domains. Survey responses indicated generally positive global usability ratings, acceptable learnability following initial exposure, and strong perceived clinical relevance for the early identification and prioritization of high-risk neonates. At the same time, item-level distribution revealed moderate variability in perceptions of complexity, with a subset of respondents reporting increased burden under high-workload conditions. These patterns collectively reflect broad usability across participants, alongside measurable differences in experience related to workflow intensity and staffing context.

Discussion

Principal Findings and Interpretation

This pilot, multisite usability evaluation suggests that the neonatal risk predictor tool was generally perceived as usable, learnable, and clinically relevant by participating frontline neonatal health care workers across the 3 Kenyan facilities. Most respondents reported willingness to use the tool regularly and increasing confidence with continued exposure; median composite usability scores were in the favorable range. At the same time, item-level responses revealed measurable polarization on perceived complexity and workload sensitivity, with a minority of users reporting initial difficulty and persistent concerns about documentation burden under high patient loads. These findings suggest that while the tool is broadly acceptable to users, its successful integration into routine neonatal workflows will depend on workload-aware implementation strategies and structured onboarding.

Our results align with broader evidence emphasizing the importance of usability and contextual fit for clinical decision-support tools in LMICs. The neonatal period remains the most vulnerable window for child survival globally, and interventions that improve early identification of high-risk neonates must be operationally feasible at the point of care to affect outcomes [1,3]. The finding that frontline health care workers value clinically relevant predictors (eg, Appearance, Pulse, Grimace, Activity, and Respiration [APGAR], gestational age, and vital signs) is consistent with prior ML prediction reviews that identify these variables as strong, widely used

XSL•FO
RenderX

predictors of neonatal outcomes [5]. However, systematic reviews and expert commentaries have repeatedly highlighted that technical model performance alone is insufficient for impact: external validation, predictor availability, and deployment logistics are essential for translation [5]. Our triangulated usability and KII data confirm this; technical validity must be matched by usability and health-system readiness for effective adoption [3,5].

Our findings are also consistent with previous usability evaluations of digital health interventions in maternal and neonatal care. For example, Balderas-Díaz et al [32] demonstrated that user-centered design, iterative proof-of-concept evaluation, and usability testing are essential for promoting user acceptance and successful implementation of adaptive mHealth systems in perinatal care. Consistent with their findings, our study highlights that, beyond technical performance, successful adoption depends on usability, workflow integration, and alignment with users' clinical needs.

Usability Implications and Standards

The observed variability in perceived complexity and information-quality dispersion highlights the need to follow established usability standards and HCD principles when refining and scaling the tool. ISO 9241-11 defines usability as effectiveness, efficiency, and satisfaction in a specified context of use; our findings show acceptable satisfaction and improved efficiency with use, but situational constraints reduce effectiveness in high-workload settings [14]. Similarly, reviews of usability instruments in mHealth underline the importance of combining standardized psychometric measures (eg, PSSUQ, QUIS, and SUS) with context-specific items that capture workflow and documentation interactions [15]. Our approach, combining standardized scales with a project-specific PSNUQ and KIIs, therefore, aligns with recommended practice for pragmatic usability evaluation in health settings [14,15].

Interpretation of Variability and Polarization

The observed polarization in perceived complexity is not merely noise—it is a signal that the intervention interacts with heterogeneous user contexts, capacities, and expectations. Some frontline staff found the tool intuitive and readily incorporated it into routine care, while others experienced friction that increased perceived burden. This pattern commonly appears in early-stage usability evaluations and often reflects differences in prior exposure to structured documentation, individual triage strategies, and local workflow realities (eg, staffing and shift patterns). Interpreting polarization, therefore, requires moving beyond averages to examine subgroups (by facility, cadre, or prior digital familiarity) and to identify threshold conditions: small increases in caseload or complexity may flip a user from "comfortable" to "overloaded," producing bimodal response distributions. Empirically, such subgroup heterogeneity is a cue for targeted refinements (eg, simplified pathways for high-workload contexts) rather than wholesale redesign [33].

From a theoretical perspective, polarization can also reflect differences in cognitive load and schema availability: users with established mental models for structured documentation experience lower intrinsic and extraneous cognitive load and therefore report lower perceived complexity; those lacking such schemas are more affected by extraneous load and report higher complexity (cognitive load theory). Targeted training that promotes rapid schema formation (eg, worked examples and supervised practice) is therefore likely to reduce polarization by elevating baseline competence across the team [34].

Finally, polarization carries implementation implications: rather than treating the tool as binary (adopt vs reject), implementation plans should include adaptive pathways, for example, a "light" checklist for high-pressure shifts and a richer version for less busy times, to maintain fidelity while accounting for context. This approach preserves core clinical value while lowering the activation energy for use under strain (implementation outcomes and feasibility guidance) [35].

Translation Gap Between ML and Clinical Workflow

The literature on ML for neonatal prediction is clear: many models demonstrate high retrospective performance but fail to translate into routine care due to gaps in predictor availability, data quality, and workflow fit (external validation, calibration, and deployment are often missing). In other words, algorithmic accuracy is necessary but not sufficient for clinical impact; the "last mile" is practical integration into health care workers' daily tasks [9,10]. This study therefore complements previous model-development research by evaluating the implementation of an ML-derived predictor set under routine clinical conditions rather than reassessing algorithmic performance. Consistent with broader analyses of health information technology incidents, system-level fragilities often emerge not from algorithmic logic itself but from workflow integration failures, configuration issues, and documentation burdens.

Our findings reinforce this translation gap: frontline staff valued clinically relevant predictors (APGAR, gestational age [GA], and vitals), but even clinically sensible predictors must be routinely and rapidly available in usable form for an ML-based tool to function in practice. This reinforces calls in the literature for deployment-focused research that explicitly couples model development with operational capture and usability testing [5,36].

Operational requirements for ML deployment, therefore, extend beyond model metrics to include measures of data completeness, real-time availability, and process fit. Our mixed results, reasonable data capture for many predictors but workload-sensitive usability, illustrate why ML systems often falter post deployment: data pipelines may look good in retrospective datasets but break when frontline realities impose delays, missing fields, or inconsistent recording habits. Addressing this requires co-designed data workflows, automated data extraction when possible, and continuous monitoring of data quality during deployment [5,37].

Although the neonatal risk predictor evaluated in this study was derived from a validated ML model, participants were not asked to evaluate or compare the underlying algorithm. Instead, the usability evaluation focused on frontline users' experiences with the paper-based implementation of the predictor set within routine clinical workflows. Consequently, the findings should be interpreted as evidence regarding implementation and

XSL•FO
RenderX

usability rather than perceptions of algorithmic transparency or trust in specific ML approaches.

Cognitive and Workflow Load Considerations

Clinical tasks in neonatal units are time-pressured and cognitively intensive. Cognitive load theory highlights that working memory is limited and that extraneous load (poorly designed interfaces, unnecessary duplication) reduces the capacity available for core clinical reasoning. Our results, which showed that some users reported perceived complexity and workload-related friction, map precisely onto this theory. In practice, design choices that reduce extraneous cognitive load (prefilled defaults, binary checkboxes rather than free-text, and minimized decision branches) and scaffold intrinsic load (clear prompts and quick-reference thresholds) are most likely to improve consistent uptake. Empirical work on electronic health records (EHRs) and clinical decision support system **(**CDSS) shows that poor interface design increases cognitive load and error risk; conversely, simplified, context-aware interfaces reduce time-on-task and decision fatigue [34,38].

From a workflow standpoint, the tool must respect "attention budgets." During high-volume shifts, health care workers have very limited time to document; even an additional 30 seconds per baby can be prohibitive. Therefore, interventions should be evaluated in terms of marginal time cost under realistic caseloads (time-motion studies), and design should aim to minimize per-case time through automation, task reallocation, or condensed item sets for critical screening. This attention to microeconomics of time turns usability into an outcome (time cost and adherence), not just a perception metric [38,39].

Our findings can also be interpreted through the Systems Engineering Initiative for Patient Safety (SEIPS) framework, which conceptualizes health care quality and patient safety as emerging from interactions among people, tasks, tools, and technologies, the physical environment, and organizational conditions [40,41]. From a SEIPS perspective, the documentation burden reported during periods of high workload reflects a work-system issue rather than a limitation of the neonatal risk predictor itself. Similarly, the qualitative findings regarding staffing shortages, parallel documentation systems, and organizational support indicate that successful implementation depends not only on the tool's usability but also on optimizing the broader clinical work system and integrating workflows. In low-resource settings, these considerations extend to the availability and reliability of the infrastructure supporting digital health implementation; evidence from Kenyan health facilities has highlighted the importance of addressing electricity and connectivity constraints and considering context-appropriate approaches to maintain continuity of digital health services [42,43].

Implementation Science Perspective

Implementation science provides conceptual tools to interpret our findings and plan next steps. The Implementation Outcomes Framework (Proctor et al [35]) identifies acceptability, feasibility, appropriateness, and adoption as proximal outcomes that must precede penetration and sustainability. Our data show reasonable acceptability and perceived appropriateness, but feasibility and adoption are conditional on staffing and documentation realities, precisely the intermediate implementation outcomes that predict later penetration and sustainability. Framing results in this taxonomy clarifies operational priorities (eg, improving feasibility via workflow changes) and defines measurable implementation end points for subsequent trials.

Normalization process theory is also useful: it articulates how new practices become embedded through coherence (sense-making), cognitive participation (engagement), collective action (operational work), and reflexive monitoring [44]. Our KIIs and survey data indicate that coherence exists (staff see the value), but cognitive participation and collective action are hindered by workload and governance barriers, meaning that embedding will require active facilitation, local champions, and adaptive workflows to transform individual acceptance into routinized practice. These theoretical lenses suggest concrete implementation strategies (champion identification, cocreation, iterative feedback loops, and fidelity monitoring) that align with our empirical patterns [45].

Contribution to LMIC Digital Health Literature

This study contributes, both practically and conceptually, to the emerging LMIC digital health literature in several ways. First, it presents real-world usability evidence from neonatal units, a high-impact but under-studied clinical domain in LMICs, and ties usability to operational capture metrics, bridging a common gap between model-focused and deployment-focused work [39]. Second, by triangulating standardized usability instruments with context-specific qualitative insights, the work demonstrates a pragmatic mixed methods evaluation blueprint that other LMIC implementers can replicate. Third, it underscores that pilot usability samples (n≈5-15) are methodologically defensible for early user-centered design validation while pointing toward necessary next steps (larger trials and psychometric validation), thereby setting realistic methodological expectations for the field. Finally, it emphasizes implementation equity: without attention to staffing, governance, and data burden, digital tools risk widening care quality gaps across facilities [46]. Digital inequities and differential adoption across populations have been documented in other contexts, where infrastructure, governance, and sociocultural factors shape digital uptake [47,48]. These contributions respond directly to calls in recent mHealth and implementation reviews for context-sensitive, deployment-oriented evidence.

Implications for Practice

Based on our findings, implementers and facility managers should consider the following practical steps before scaling:

1. Structured onboarding and mentorship: Provide brief, practical training sessions with supervised use during the initial weeks and ongoing mentorship for new or rotating staff to reduce early-use errors and perceived complexity [49]. Evidence from health care quality improvement and digital training interventions demonstrates that structured onboarding and supervised adoption phases improve implementation fidelity and long-term usability outcomes [23,50,51].

XSL•FO
RenderX

2. Workflow-sensitive adaptation: Integrate the tool into existing delivery registers or consolidate parallel documentation forms to reduce duplication; where duplication is unavoidable for regulatory reasons, design the tool to minimize additional data entry (eg, checkboxes and prefilled fields) [52].
3. Task allocation and scheduling: Assign clear responsibilities (eg, primary recorder during delivery and secondary verifier) to ensure consistent completion during peak periods [53].
4. Hybrid implementation planning: For facilities with unreliable IT and infrastructure, maintain a hybrid approach (paper with planned digitization) and plan for offline digital options with synchronization capability [52,54].
5. Monitoring and iteration: Embed short feedback cycles (weekly and monthly) to capture user problems and quickly iterate on interface or process changes [54,55].
6. User engagement and co-design: Actively involve frontline health care workers in iterative refinement of the tool through feedback sessions, participatory design workshops, or pilot discussions. Engaging end users in the adaptation process can enhance local ownership, improve contextual fit, and increase the likelihood of sustained adoption within routine clinical workflows [17,20].

These pragmatic steps align with implementation science principles and can increase the likelihood that technically valid predictor models translate into practice improvements.

Strengths and Limitations of the Study

Several features increase the study's utility to researchers and implementers: (1) real-world exposure, that is, the usability assessment followed a 4-month implementation period rather than a short laboratory test, increasing ecological validity); (2) multisite design, that is, 3 diverse facility contexts (urban informal, periurban public, and faith-based referral) enhance transferability of implementation insights across LMIC delivery settings; (3) triangulated methods, that is, combining standardized usability instruments (SUS, QUIS and PSSUQ-derived items) with KIIs provided depth and helped explain quantitative polarization; and (4) operational realism, that is, the usability evaluation was explicitly tied to routine documentation and staffing realities, yielding actionable recommendations for onboarding and workflow adaptation.

We must acknowledge several limitations that constrain the generalizability and strength of conclusions. First, the postsurvey sample size was small (n=10), limiting statistical power and precluding robust inferential testing; this study should therefore be understood as a pilot usability evaluation rather than a definitive validation. Accordingly, the findings should be interpreted as preliminary evidence intended to inform future larger-scale usability and implementation studies rather than to support broad generalization. Second, although multi-site, qualitative key informant data were limited to 3 senior health care workers (n=3), which constrains claims about saturation for broad thematic generalization; however, those informants were purposively selected leaders with facility-level oversight, increasing the relevance of governance and workflow insights.

Third, the usability measures relied on self-report; objective measures of time on task, error rates, or direct observation could provide complementary evidence of efficiency and effectiveness. Fourth, the PSNUQ instrument is study-adapted; while combining it with validated scales strengthens construct coverage, formal psychometric validation (eg, reliability and factor structure) was not possible with this sample. Finally, the study used a paper-based implementation of the predictor tool; findings may differ with a fully digital interface and therefore should not be assumed to directly generalize to other delivery modalities. These limitations should guide the interpretation and design of future studies.

Future Research Recommendations

To build on these pilot findings, we recommend:

1. Larger multisite usability and effectiveness trials: Expand sample sizes across more facilities and include objective time-motion measures, task completion rates, and downstream clinical process outcomes (eg, time to antibiotic administration, referral timeliness) [56].
2. Psychometric validation of PSNUQ: With a larger sample, assess internal consistency (Cronbach α) and factor structure to formally validate the adapted instrument [37].
3. Compare delivery modalities: Evaluate differences between paper, hybrid, and fully digital interfaces for the same predictor tool to identify modality-specific strengths and weaknesses [57].
4. Implementation trials linked to process outcomes: Conduct stepped-wedge or cluster-randomized evaluations to test whether increased usability and workflow integration translate into measurable improvements in neonatal care processes and, ultimately, clinical outcomes [58].
5. Cost-effectiveness and sustainability analysis: Assess the resource and financing implications of scaling, including staff time, training costs, and information communication technology (ICT) investments. These analyses will help policymakers weigh trade-offs between manual and digital approaches [59].

Conclusions

This pilot usability evaluation suggests that participating frontline neonatal health care workers generally perceived the neonatal risk predictor tool as usable, learnable, and clinically relevant, while also identifying workflow-sensitive barriers, particularly staffing pressures and documentation duplication, that may influence implementation. These findings support proceeding to larger, methodologically rigorous evaluations while simultaneously investing in pragmatic onboarding, workflow integration, and infrastructure planning to maximize the potential for real-world impact. Given the global priority of reducing neonatal mortality, pragmatic usability evidence such as ours is a necessary step toward responsibly translating ML into safer, more timely neonatal care.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank all health care professionals and facility leaders who participated in this study for their time, insights, and contributions. We are particularly grateful to the participating health facilities in Kenya for their collaboration and support during data collection and implementation activities. We also acknowledge the contributions of research assistants and local coordinators who supported data collection, transcription, and field logistics. Their efforts were essential to the successful completion of this study. The authors used OpenAI to assist with language editing, manuscript refinement, and figure refinement for publication. All AI-assisted outputs were critically reviewed, verified, and edited by the authors, who take full responsibility for the accuracy, integrity, and final content of the manuscript.

Data Availability

The datasets generated and analyzed during the current study are not publicly available due to ethical and confidentiality restrictions. The data include deidentified qualitative interview transcripts collected from healthcare professionals within specific health facilities in Kenya. Given the small sample size and context-specific nature of the facilities, public sharing could potentially compromise participant confidentiality. Deidentified data may be made available from the corresponding author upon reasonable request, subject to institutional and ethical approval and applicable data protection requirements.

Funding

A Frontiers Seed Funding (Grant number FS-2425-23-127) from the Royal Academy of Engineering has been received to support this study.

Authors' Contributions

Conceptualization: RDN, PM, EK, BM, MSRJ
Methodology: RDN, PM, MSRJ
Formal analysis: RDN, PM, EK, FS, MSRJ
Investigation: RDN, FS, BM
Writing – review & editing: RDN, PM, MSRJ
Manuscript structure and refinement: RDN
Literature review: EK, BM
Study tool design: FS
Data curation: FS
Study coordination: FS
Writing – original draft: MSRJ
Mentoring: MSRJ
Supervision: MSRJ
Overall study coordination: MSRJ

Conflicts of Interest

None declared.

Multimedia Appendix 1
GRAMMS Checklist.
[PDF File (Adobe PDF File), 89 KB - humanfactors_v13i1e94828_app1.pdf]

Multimedia Appendix 2
Data collection tools.
[PDF File (Adobe PDF File), 7547 KB - humanfactors_v13i1e94828_app2.pdf]

Multimedia Appendix 3
Qualitative verbatim quotes.
[PDF File (Adobe PDF File), 95 KB - humanfactors_v13i1e94828_app3.pdf]

References

1. Newborn mortality. World Health Organization. 2024 Mar 14. URL: https://www.who.int/news-room/fact-sheets/detail/newborn-mortality? [accessed 2026-02-21]
2. Levels and trends in child mortality. UNICEF. 2024. URL: https://data.unicef.org/resources/levels-and-trends-in-child-mortality-2024/ [accessed 2026-08-31]

XSL•FO
RenderX

3. Lawn JE, Bhutta ZA, Ezeaka C, Saugstad O. Ending preventable neonatal deaths: multicountry evidence to inform accelerated progress to the sustainable development goal by 2030. Neonatology 2023;120(4):491-499. [doi: 10.1159/000530496] [Medline: 37231868]
4. Shukla VV, Eggleston B, Ambalavanan N, McClure EM, Mwenechanya M, Chomba E, et al. Predictive modeling for perinatal mortality in resource-limited settings. JAMA Netw Open 2020;3(11):e2026750 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.26750] [Medline: 33206194]
5. Mangold C, Zoretic S, Thallapureddy K, Moreira A, Chorath K, Moreira A. Machine learning models for predicting neonatal mortality: a systematic review. Neonatology 2021;118(4):394-405 [FREE Full text] [doi: 10.1159/000516891] [Medline: 34261070]
6. Preti LM, Ardito V, Compagni A, Petracca F, Cappellaro G. Implementation of machine learning applications in health care organizations: systematic review of empirical studies. J Med Internet Res 2024;26:e55897 [FREE Full text] [doi: 10.2196/55897] [Medline: 39586084]
7. Jabin MSR. A qualitative, multi-framework methodology for analysing health information technology-related patient safety incidents. Front Digit Health 2026;8:1786358 [FREE Full text] [doi: 10.3389/fdgth.2026.1786358] [Medline: 42294047]
8. Jabin M. Why health information technology safety problems remain invisible. Front Digit Health 2026;8:1785141 [FREE Full text] [doi: 10.3389/fdgth.2026.1785141] [Medline: 42221162]
9. Nyatuka D, Jabin MSR, Dionne-Morris L. Contextual barriers to health information systems optimization in underserved communities in Kenya: qualitative study informed by frugal innovation and information and communication technologies for development. JMIR Form Res 2026;10:e78950 [FREE Full text] [doi: 10.2196/78950] [Medline: 41813110]
10. Nyatuka RD, Jabin MSR, Dionne-Morris L. Frugal-oriented information and communication technology for development framework toward low-cost digital maternal health in low- and middle-income countries: quantitative descriptive study. JMIR Form Res 2026;10:e77330 [FREE Full text] [doi: 10.2196/77330] [Medline: 42044317]
11. de Hond AAH, Leeuwenberg AM, Hooft L, Kant IMJ, Nijman SWJ, van Os HJA, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. NPJ Digit Med 2022;5(1):2 [FREE Full text] [doi: 10.1038/s41746-021-00549-7] [Medline: 35013569]
12. Adlung L, Cohen Y, Mor U, Elinav E. Machine learning in clinical decision making. Med 2021;2(6):642-665 [FREE Full text] [doi: 10.1016/j.medj.2021.04.006] [Medline: 35590138]
13. Choi Y, Yu W, Nagarajan MB, Teng P, Goldin JG, Raman SS, et al. Translating AI to clinical practice: overcoming data shift with explainability. Radiographics 2023;43(5):e220105 [FREE Full text] [doi: 10.1148/rg.220105] [Medline: 37104124]
14. Ergonomics of human-system interaction (ISO 9241-11:2018); Part 11: Usability: Definitions and concepts. ISO. 2018. URL: https://www.iso.org/standard/63500.html? [accessed 2026-02-21]
15. Hajesmaeel-Gohari S, Khordastan F, Fatehi F, Samzadeh H, Bahaadinbeigy K. The most used questionnaires for evaluating satisfaction, usability, acceptance, and quality outcomes of mobile health. BMC Med Inform Decis Mak 2022;22(1):22 [FREE Full text] [doi: 10.1186/s12911-022-01764-2] [Medline: 35081953]
16. Jabin M. Why digital health fails silently: a sociotechnical theory of health information technology-related risk. Front Digit Health 2026;8:1785086 [FREE Full text] [doi: 10.3389/fdgth.2026.1785086] [Medline: 42245003]
17. Kang B, Poddar M, Luitel A, Rimal RN, Melaku B, Black DP. Narrative review of human-centered design in public health interventions in low- and middle-income countries: recommendations for practice, research, and reporting. Glob Health Sci Pract 2025;13(1):e2400164 [FREE Full text] [doi: 10.9745/GHSP-D-24-00164] [Medline: 40813242]
18. Tzimourta K. Human-centered design and development in digital health: approaches, challenges, and emerging trends. Cureus 2025;17(6):e85897. [doi: 10.7759/cureus.85897] [Medline: 40548156]
19. Göttgens I, Oertelt-Prigione S. The application of human-centered design approaches in health research and innovation: a narrative review of current practices. JMIR Mhealth Uhealth 2021;9(12):e28102 [FREE Full text] [doi: 10.2196/28102] [Medline: 34874893]
20. Triplett NS, Munson S, Mbwayo A, Mutavi T, Weiner BJ, Collins P, et al. Applying human-centered design to maximize acceptability, feasibility, and usability of mobile technology supervision in Kenya: a mixed methods pilot study protocol. Implement Sci Commun 2021;2(1):2 [FREE Full text] [doi: 10.1186/s43058-020-00102-9] [Medline: 33413688]
21. Nyatuka RD, Macharia P, Esther K, Siva F, Muriithi B, Rahman Jabin MS. Kenyan neonatal mortality risk predictor: protocol for a user-centered design evaluation. JMIR Res Protoc 2026;15:e81996 [FREE Full text] [doi: 10.2196/81996] [Medline: 41894529]
22. NA. 610-1990 - IEEE Standard Computer Dictionary: a compilation of IEEE standard computer glossaries. IEEE Xplore. Piscataway, NJ: IEEE; 1991. URL: https://ieeexplore.ieee.org/document/182763 [accessed 2026-08-21]
23. Jabin MSR, Schultz T, Mandel C, Bessen T, Hibbert P, Wiles L, et al. A mixed-methods systematic review of the effectiveness and experiences of quality improvement interventions in radiology. J Patient Saf 2022;18(1):e97-e107. [doi: 10.1097/PTS.0000000000000709] [Medline: 32433438]
24. Jabin M, Schultz T, Hibbert P, Mandel C, Runciman W. Effectiveness of quality improvement interventions for patient safety in radiology: a systematic review protocol. JBI Database System Rev Implement Rep 2016;14(9):65-78. [doi: 10.11124/JBISRIR-2016-003078] [Medline: 27755318]

XSL•FO
RenderX

25. Jabin MSR, Samuriwo R, Chilaka M, Yaroson EV. Effectiveness and experiences of quality improvement interventions in older adult care: protocol for a mixed methods systematic review. JMIR Res Protoc 2024;13:e56346 [FREE Full text] [doi: 10.2196/56346] [Medline: 38635311]
26. Nyatuka RD, Siva F, Jabin MSR. Governance pathways for scaling neonatal risk stratification tools in LMIC health systems: a multi-site qualitative study in Kenya. Front Digit Health 2026;8:1883524 [FREE Full text] [doi: 10.3389/fdgth.2026.1883524] [Medline: 42577112]
27. Sauro J. 10 Things to know about the post study System Usability Questionnaire. MeasuringU. 2019. URL: https://measuringu.com/pssuq/ [accessed 2026-02-22]
28. Chin JP, Diehl VA, Norman KL. Development of an instrument measuring user satisfaction of the human-computer interface. 1988 Presented at: CHI '88: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; May 15-19, 1988; Washington DC p. 213-218. [doi: 10.1145/57167.57203]
29. Brooke J. SUS: a quick and dirty usability scale. In: Usability Evaluation in Industry. United Kingdom: Taylor & Francis; 1995:189.
30. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014;9:26152 [FREE Full text] [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]
31. Clarke V, Braun V. Teaching thematic analysis: overcoming challenges and developing strategies for effective learning. Psychologist 2013;26(2). [doi: 10.4135/9781036232078.n9]
32. Balderas-Díaz S, Rodríguez-Fórtiz MJ, Garrido JL, Bellido-González M, Guerrero-Contreras G. A psycho-educational intervention programme for parents with SGA foetuses supported by an adaptive mHealth system: design, proof of concept and usability assessment. BMC Med Inform Decis Mak 2022;22(Suppl 4):291 [FREE Full text] [doi: 10.1186/s12911-022-02036-9] [Medline: 36357878]
33. Nielsen J. Why you only need to test with 5 users. NN/G. 2018. URL: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ [accessed 2026-02-22]
34. Baxter KA, Sachdeva N, Baker S. The application of cognitive load theory to the design of health and behavior change programs: principles and recommendations. Health Educ Behav 2025;52(4):469-477 [FREE Full text] [doi: 10.1177/10901981251327185] [Medline: 40145544]
35. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
36. Seneviratne MG, Shah NH, Chu L. Bridging the implementation gap of machine learning in healthcare. BMJ Innov 2019;6(2):45-47. [doi: 10.1136/bmjinnov-2019-000359]
37. Husain L, Kitchens K, Raja S, Memon M. Bridging implementation gaps in digital health: a translational research imperative for equitable healthcare innovation. Clin Transl Sci 2025;18(10):e70375. [doi: 10.1111/cts.70375] [Medline: 41044865]
38. Asgari E, Kaur J, Nuredini G, Balloch J, Taylor AM, Sebire N, et al. Impact of electronic health record use on cognitive load and burnout among clinicians: narrative review. JMIR Med Inform 2024;12:e55499 [FREE Full text] [doi: 10.2196/55499] [Medline: 38607672]
39. McCool J, Dobson R, Whittaker R, Paton C. Mobile health (mHealth) in low- and middle-income countries. Annu Rev Public Health 2022;43:525-539 [FREE Full text] [doi: 10.1146/annurev-publhealth-052620-093850] [Medline: 34648368]
40. Carayon P, Wooldridge A, Hoonakker P, Hundt AS, Kelly MM. SEIPS 3.0: human-centered design of the patient journey for patient safety. Appl Ergon 2020;84:103033 [FREE Full text] [doi: 10.1016/j.apergo.2019.103033] [Medline: 31987516]
41. Holden RJ, Carayon P, Gurses AP, Hoonakker P, Hundt AS, Ozok AA, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 2013;56(11):1669-1686 [FREE Full text] [doi: 10.1080/00140139.2013.838643] [Medline: 24088063]
42. Jabin MSR, Nyatuka RD. Assessing the feasibility of solar-powered digital health infrastructure in rural and peri-urban Kenya: a mixed-methods study. Front Digit Health 2026;8.
43. Jabin MSR, Nyatuka RD. Bridging the data-power paradox: a conceptual-empirical analysis of solar-powered digital health systems in low-resource settings. Front Digit Health 2026;8:1848840. [doi: 10.3389/fdgth.2026.1848840] [Medline: 42294048]
44. May CR, Mair F, Finch T, MacFarlane A, Dowrick C, Treweek S, et al. Development of a theory of implementation and integration: normalization process theory. Implement Sci 2009;4:29 [FREE Full text] [doi: 10.1186/1748-5908-4-29] [Medline: 19460163]
45. Murray E, Treweek S, Pope C, MacFarlane A, Ballini L, Dowrick C, et al. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. BMC Med 2010;8:63 [FREE Full text] [doi: 10.1186/1741-7015-8-63] [Medline: 20961442]
46. Poulsen A, Hickie IB, Alam M, Crouse JJ, Ekambareshwar M, Loblay V, et al. Overcoming barriers to mHealth co-design in low- and middle-income countries: a research toolkit. Inf Technol Dev 2024;30(3):542-561. [doi: 10.1080/02681102.2023.2295480]

XSL•FO
RenderX

47. Wepa D, Thomas S, Jabin MSR. The experience and impact of digital technologies on indigenous populations in New Zealand during the COVID-19 pandemic and cyclone gabrielle: the kaupapa māori methodology. JMIR Aging 2025;8:e73974 [FREE Full text] [doi: 10.2196/73974] [Medline: 41150869]
48. Wepa D. Digital disparities: tech solutions for indigenous communities. Open Access Gov 2025:26-27. [doi: 10.56367/oag-046-11905]
49. Muñoz PM, Cardinal-Fernández P, Morales Rodríguez Á, Ruiz-Zaldibar C, de la Cuerda López A. The impact of an onboarding plan for newly hired nurses and nursing assistants: results of a quasi-experimental study. Nurs Rep 2025;15(11):398 [FREE Full text] [doi: 10.3390/nursrep15110398] [Medline: 41295822]
50. Rahman Jabin MS, Ashfaq A, Bi N, Nilsson E. Impact of training and education programs for health care professionals on video and text-based meetings in ensuring health care quality: protocol for a scoping review. JMIR Res Protoc 2025;14:e69963 [FREE Full text] [doi: 10.2196/69963] [Medline: 40633092]
51. Jabin MSR, Bi N, Thomas S, Ashfaq A, Nilsson E. The impact of training and education programs for healthcare professionals on video- and text-based consultations in ensuring healthcare quality: a scoping review. Front Digit Health 2026;8:1861018 [FREE Full text] [doi: 10.3389/fdgth.2026.1861018] [Medline: 42528790]
52. Bowens FM, Frye PA, Jones WA. Health information technology: integration of clinical workflow into meaningful use of electronic health records. Perspect Health Inf Manag 2010;7(Fall):1d [FREE Full text] [Medline: 21063545]
53. Patterson ES, Gibbons MC, Latkany P, Ramaiah M, Brick D, Lowry SZ. Integrating electronic health records into clinical workflow: an application of human factors modeling methods to two specialty care areas. 2015 Presented at: Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care; June 17-19, 2015; Garden Grove, CA p. 42-49. [doi: 10.1177/2327857915041018]
54. Loo RTJ, Nasta F, Macchi M, Baudot A, Burstein F, Bove R, et al. Recommendations for successful development and implementation of digital health technology tools. J Med Internet Res 2025;27:e56747 [FREE Full text] [doi: 10.2196/56747] [Medline: 40499040]
55. Goldhaber N, Jacobs MB, Laurent LC, Knight R, Zhu W, Pham D, et al. Integrating clinical research into electronic health record workflows to support a learning health system. JAMIA Open 2024;7(2):ooae023 [FREE Full text] [doi: 10.1093/jamiaopen/ooae023] [Medline: 38751411]
56. Rosa C, Marsch LA, Winstanley EL, Brunner M, Campbell AN. Using digital technologies in clinical trials: current and future applications. Contemp Clin Trials 2021;100:106219 [FREE Full text] [doi: 10.1016/j.cct.2020.106219] [Medline: 33212293]
57. Mumtaz H, Riaz MH, Wajid H, Saqib M, Zeeshan MH, Khan SE, et al. Current challenges and potential solutions to the use of digital health technologies in evidence generation: a narrative review. Front Digit Health 2023;5:1203945 [FREE Full text] [doi: 10.3389/fdgth.2023.1203945] [Medline: 37840685]
58. Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. BMC Med Res Methodol 2006;6:54.
59. Gentili A, Failla G, Melnyk A, Puleo V, Tanna GLD, Ricciardi W, et al. The cost-effectiveness of digital health interventions: a systematic review of the literature. Front Public Health 2022;10:787135 [FREE Full text] [doi: 10.3389/fpubh.2022.787135] [Medline: 36033812]

Abbreviations

APGAR: Appearance, Pulse, Grimace, Activity, and Respiration
CDSS: clinical decision support system
EHR: electronic health record
GRAMMS: Good Reporting of A Mixed Methods Study
HCD: user-centered design
HIT: health information technology
ICT: information communication technology
KII: key informant interview
LMIC: low- and middle-income countries
ML: machine learning
PSNUQ: Post-Study Neonatal Utility Questionnaire
PSSUQ: Post-Study System Usability Questionnaire
QUIS: Questionnaire for User Interaction Satisfaction
SDG: Sustainable Development Goal
SEIPS: Systems Engineering Initiative for Patient Safety
SU-ISERC: Strathmore University Institutional Scientific Ethics Review Committee
SUS: System Usability Scale
UX: user experience

XSL•FO
RenderX

Edited by A Bamgboje-Ayodele; submitted 07.Mar.2026; peer-reviewed by CJ Engstrom, G Guerrero-Contreras; comments to author 25.Jun.2026; revised version received 15.Aug.2026; accepted 17.Aug.2026; published 09.Sep.2026.

Please cite as:
Nyatuka RD, Macharia P, Kakhata E, Siva F, Muriithi B, Jabin MSR
Usability and Workflow Integration of a Machine Learning–Derived Neonatal Risk Predictor in Kenyan Neonatal Units: Multisite User-Centered Pilot Evaluation
JMIR Hum Factors 2026;13:e94828
URL: https://humanfactors.jmir.org/2026/1/e94828
doi:10.2196/94828
PMID:

©Ronald Danny Nyatuka, Paul Macharia, Esther Kakhata, Faith Siva, Betsy Muriithi, Md Shafiqur Rahman Jabin. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 09.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Patient Perspectives on a Symptom Assessment App in Emergency Care: Interview Study Among Patients in the AkuSym Trial

Claudia Schorr[1], MSc; Malte L Schmieding[2], Dr med, MA; Doreen Kuschick[1], Dr rer medic, MSc; Felix Balzer[2], Dr med, Dr rer nat; Myrto Bolanaki[3], Dr med; Christoph Heintze[1], Dr med, MPH; Marvin Kopka[2,4], MSc, MPH, Dr rer medic, PhD; Martin Möckel[3], Dr med; Anna Slagman[3], Dr rer medic, MSc; Alica Thissen[2], Dr phil, Dipl-Psych; Cornelia Wäscher[1], MPH; Konrad Schmidt[1,5], Dr med; Hendrik Napierala[1], MSc, Dr med

[1]Institute of General Practice and Family Medicine, Charité - Universitätsmedizin Berlin, Charitéplatz 1, Berlin, Germany

[2]Institute of Medical Informatics, Charité - Universitätsmedizin Berlin, Berlin, Germany

[3]Emergency and Acute Medicine, Charité - Universitätsmedizin Berlin, Berlin, Germany

[4]Department of Psychology and Ergonomics (IPA), Division of Ergonomics, Technische Universität Berlin, Berlin, Germany

[5]Institute of General Practice, Faculty of Health Sciences Brandenburg, Brandenburg Medical School Theodor Fontane, Neuruppin, Brandenburg, Germany

Corresponding Author:
Claudia Schorr, MSc
Institute of General Practice and Family Medicine, Charité - Universitätsmedizin Berlin, Charitéplatz 1, Berlin, Germany

Abstract

Background: Symptom assessment apps (SAAs) allow laypersons to input their medical history and symptoms, generating a report of probable diagnoses and recommending next steps for medical care. Numerous and varied research findings exist regarding their diagnostic and triage accuracy, usability, acceptance, and user characteristics, including use in various settings. However, the perspective of low-acuity emergency patients on the benefits and challenges of using SAAs has not yet been explored.

Objective: This study aimed to explore patients' experiences and attitudes toward the use of SAA before consultation in the emergency department (ED) and to identify perceived benefits and challenges of integrating SAAs into emergency care from the patients' perspective.

Methods: We conducted semistructured interviews with self-referred, walk-in patients at 2 EDs and an ED-associated outpatient on-call practice in Berlin, who were allocated to the intervention group of the AkuSym trial. Participants had self-assessed their health complaints using the Ada Health SAA before their first encounter with the treating physician. Interviews were audio-recorded, transcribed, and analyzed using qualitative content analysis with deductively and inductively developed categories.

Results: Eighteen patients were interviewed between July 2022 and November 2022. Most participants stated that they would use an SAA again, including outside the emergency care setting. Key factors influencing reuse were precise symptom queries and clear, layperson-friendly medical content. Patients' trust in medical professionals was higher than in SAA assessments, and personal doctor-patient contact remained essential. The SAA was perceived primarily as a source of preliminary information and a supplement to doctor-patient contact. It helped participants reflect on and articulate their symptoms. Participants also saw potential for SAAs to support decisions about whether and where to seek medical care, to provide self-help recommendations for minor complaints, and to enable more informed participation in diagnostic discussions. Concerns included inaccurate or incomplete symptom input, difficulties understanding and interpreting SAA results, and receiving alarming or unclear information that could increase anxiety or complicate the subsequent consultation. Identified challenges included enabling information transfer through digital interfaces and ensuring data protection. Participants suggested solutions such as voluntary access via waiting-room QR codes, linking SAA reports to electronic patient records, and guiding patients to appropriate health care services.

Conclusions: Meaningful integration of SAA use into emergency care requires a clear understanding of how SAA assessments should be used during the doctor-patient consultation, and ensuring accessibility for patients with limited digital health literacy. Such tools have the potential to improve health literacy and empower patients, potentially leading to more efficient health care use. Linking SAA reports to a public website that provides evidence-based, layperson-friendly information and contact details for appropriate health care providers could help patients understand their results and identify suitable care options.

Trial Registration: German Clinical Trials Register DRKS00028598; https://drks.de/search/en/trial/DRKS00028598

(JMIR Hum Factors 2026;13:e101285) doi:10.2196/101285

XSL•FO
RenderX

KEYWORDS

symptom checker; emergency care; patient-physician interaction; health literacy; patient empowerment; clinical decision support; AI in health care; qualitative content analysis

Introduction

Symptom assessment apps (SAAs) are designed for individuals seeking information on their health concerns. Users input their symptoms, and the SAA generates a report outlining probable diagnoses and/or suggesting next steps for medical care. Research on SAAs has covered their diagnostic accuracy and effectiveness in guiding patients to appropriate care levels, including use in primary care settings and emergency departments (EDs) [1-16]. Diagnostic accuracy varies considerably between SAAs [2,3,10,12,15], as does their capability to perform triage [10,15,17]. SAAs show deficiencies in both diagnostic and triage accuracy [8]. Their diagnostic accuracy has been found to be generally low [2,3,15] or acceptable when confined to patients with urgent symptoms [5]. The triage capabilities of SAAs are found to be more accurate than their diagnostic capabilities [3] and comparable to telephone triage capabilities [5,7,13]. However, SAAs tend toward risk aversion [2,8,9,12].

Studies have also examined the ease of use of SAAs [5,18-20] and identified the characteristics distinguishing users from nonusers [13,21]. The usability and/or acceptability of SAAs were positively perceived by laypersons [5,18-20,22,23]. User characteristics or predictors identified for SAA use were higher socioeconomic status [21,24], younger age [13,21], female sex [13,21,24], hypochondria, and self-efficacy [25].

Current literature addresses the impact of SAA use among healthy individuals, patients, and primary care physicians, as well as the impact on treatment consultations in primary care [19,20,23,26-30]. Most participants in a US cross-sectional survey perceived the tested SAA as a useful tool for receiving information on their health complaints and for getting closer to plausible diagnoses [24]. In an Italian study with patients and physicians of general practice settings, SAA use was positively viewed by patients, whereas general practitioners voiced criticism concerning "the usefulness of patients' self-management and reduction of unnecessary visits" [20]. Wetzel et al [28] determined neither positive nor negative associations between SAAs and primary care in a cross-sectional survey with a mixed method design among the German population.

In a mixed methods pilot observational study, Scheder-Bieschin et al [22] investigated the potential benefits of implementing an AI-driven symptom-assessment tool in EDs. The pilot confirmed strong feasibility, alongside high usability among both patients and health care providers, with strong endorsement for facilitating communication [22]. Additionally, the tool demonstrated clear potential to improve patient-provider interactions and contribute to ED efficiency [22]. However, this study was funded by the app developer, Ada Health. Independent, comprehensive studies examining patients' views on the application of SAAs in emergency settings are scarce. The current study aims to fill this gap by exploring patients' experiences and attitudes toward the use of SAAs before their consultation in the ED. Furthermore, we sought to uncover the possible benefits and drawbacks of adopting SAAs as a standard practice in emergency care from the patients' perspective.

In this study, we focus on the perspective of self-referred walk-in patients with acute symptoms triaged as low priority (Manchester Triage System levels 3 to 5), who were recruited from the randomized controlled trial (RCT) [31] of the same research project (AkuSym). Patients who were pre-examined and came to the ED or the outpatient on-call practice with a referral or with known symptoms (eg, from a chronic condition) were not suitable, as they already had a doctor-patient contact and/or an established diagnosis.

Our research was guided by the following 2 questions: How do patients perceive the use of an SAA before their consultation in the ED or in on-call practice? What are the perceived benefits and challenges of integrating SAAs into the emergency care process, according to patients?

Methods

Study Design

This study was conducted under the umbrella of the "Symptom Checker Apps in Acute Care" (AkuSym) project, which received funding from the German Federal Ministry of Health. AkuSym's methodology encompassed a cross-sectional survey, a multicenter RCT [31], focus groups involving physicians, and patient interviews.

The RCT component was executed across 3 locations in Berlin: 2 EDs at Charité-Universitätsmedizin Berlin (Campus Berlin Mitte and Campus Virchow Klinikum) and a statutory health insurance physicians' on-call practice located at the Jewish Hospital in Berlin. After triage, patients were screened for eligibility by the recruiting study personnel (study nurse, study research assistant, or study physician). Eligible and consenting participants were then randomly assigned by the blinded study team to either the intervention group or control group in a 1:1 ratio. Participants in the intervention group used an SAA on a tablet before their consultation with a physician, while those in the control group proceeded with their visit without using the SAA. Study personnel provided instructions on the use of the SAA (for the intervention group) and the tablet to complete the surveys (for both groups) and were available for assistance throughout the study. For the intervention, the SAA, Ada [32], was chosen from numerous options, based on existing literature regarding its accuracy for diagnosis and triage, its usability, the safety of its recommendations, the range of conditions and main complaints encompassed [5,10,18,33], and its availability in both German and English. None of the researchers were involved in the development of the SAA. Both patients and health care providers were given access to the assessment reports generated by the Ada app before the encounter. However, providers could decide at their discretion whether to view the

XSL•FO
RenderX

report at all and, if so, how much of its content to incorporate into the consultation. Patients were not given specific instructions on whether to address the report during the consultation.

To gain insights into the intervention group's experiences and perceptions regarding the use of SAAs, and their potential benefits and challenges in emergency care, we conducted interviews with participants. These semistructured interviews were conducted primarily over the phone unless participants preferred in-person meetings, to facilitate timely discussions post-RCT participation, thereby ensuring participants' fresh recollection of their experiences with the SAA.

Participant Selection

Participants in the intervention group were contacted in case they indicated their interest in taking part in an interview when consenting to the RCT. The eligibility criteria included: (1) participation in the intervention group, (2) willingness to participate in an interview, and (3) proficiency in the German language. To ensure semantic equivalence and avoid data loss, distortion, or bias introduced by translation, all interviews were conducted in German. As shown in the demographics, approximately one-third of the sample reported a nationality other than German, indicating that diverse international perspectives were represented while maintaining methodological consistency.

The first 8 interviews conducted relied on convenience sampling to ensure a steady initial influx of participants. For the subsequent interviews, purposive sampling was used to actively balance the sample regarding age and sex. Since recruitment for the RCT was largely complete at that stage, this approach ensured that the total sample of interviewees mirrored the age and sex structure of the broader intervention cohort. The researchers had no prior acquaintance with the participants.

Data Collection

Semistructured interviews were conducted using a guide developed within the project team and discussed at an interdisciplinary researchers' workshop (Multimedia Appendix 1). The interview guide was developed deductively based on the study objectives and the patient journey, resulting in five thematic blocks corresponding to the chronological stages of the study on site: (1) description of the situation prior to study participation, including the reason for visiting the emergency room/on-call practice and participants' basic attitudes toward and expectations of using SAAs; (2) recalling the concrete use of the SAA during the study; (3) personal reaction to and handling of the SAA assessment report while waiting for the doctor; (4) subjective impression of the doctor-patient contact; (5) personal conclusions after treatment regarding the use of SAAs, adequate user groups, application scenarios, and potential and risks concerning the implementation of the SAA as a standard procedure in the acute care setting.

To facilitate the interviewee's speech and to focus on the research questions, each thematic block began with an introductory question, followed by several guiding questions. Each interview was audio-recorded. Before each interview, the participants gave their written consent. For telephone interviews, participants received the study information and consent form by mail, together with a prepaid return envelope, and returned the signed consent form to the Institute of General Practice and Family Medicine at Charité–Universitätsmedizin Berlin prior to the interview. For in-person interviews, participants received the documents in advance by email or mail, according to their preference. They either signed the consent form on site before the interview or brought the previously signed form to the interview. After each interview, the participants answered a short questionnaire on sociodemographic data (Table 1) and received a small expense allowance in the form of an online shopping voucher valued at €30 (€1=US $1.04 as of July 1, 2022). Data saturation was reached after 16 interviews; 2 additional interviews were conducted to confirm saturation, resulting in a total sample of 18 participants.

XSL•FO
RenderX

Table . Participant demographic data (N=18).

Characteristic	Value
Sex, n	
Male	8
Female	10
Age (y), mean (SD; min-max)	35.6 (12.94; 19-64)
Highest school-leaving qualifications, n	
Lower secondary education	3
Secondary education	5
General qualification for university entrance/vocational baccalaureate diploma	10
Highest educational qualifications, n	
No professional training (until now)	3
Vocational school/apprenticeship [34]	5
Technical school/master school/specialist academy	1
University/university of applied sciences	7
Other	2
Country of birth, n	
Germany	13
Other European Union country	3
Non–European Union country	2
Study center, n	
Charité Universitätsklinikum Campus Berlin Mitte	6
Charité Universitätsklinikum Campus Virchow	8
On-call practice of the KV-Berlin[a]	4

[a]KV-Berlin: Kassenärztliche Vereinigung Berlin (Association of Statutory Health Insurance Physicians Berlin).

Data Analysis

We used qualitative content analysis, following the approach outlined by Kuckartz [35] for our data analysis. Initial codes were deductively generated based on the interview guide. As the coding proceeded, additional codes were inductively derived from the content itself. CS, DK, and HN were involved in the coding process, comprising 4 “coding rounds,” including a test coding session. The transcripts were coded independently by the 3 researchers (CS, DK, and HN) and then cross-checked. Discrepancies were discussed within the team, and several adjustments to the categories were made before finalizing the code book. Codes with the same content were grouped together. Each code was paraphrased and interpreted.

The software MAXQDA (version 2022; VERBI GmbH) facilitated the data analysis process, with coding and analysis of the transcripts performed in German. For the purposes of this paper, quotes from the interviews were translated into English by CS, with support from the AI translation software DeepL Pro (DeepL SE).

Ethical Considerations

Ethical approval was granted by the ethics committee of Charité–Universitätsmedizin Berlin (EA2/284/21). Data safety was adhered to in line with the General Data Protection Regulation [36] and local law. All participants provided written informed consent prior to their interviews. The study was prospectively registered in the German Clinical Trials Registry (DRKS00028598). We followed the COREQ (Consolidated Criteria for Reporting Qualitative Research) [37] and the SRQR (Standards for Reporting Qualitative Research) guidelines [38] to ensure comprehensive and transparent reporting of our qualitative research findings.

Results

In total, 18 interviews were conducted in the period from July 2022 to November 2022. Of these, 17 interviews were held on the telephone, and 1 interview was conducted face-to-face. The length of the interviews ranged from 12 to 56 minutes. The shortest interview, lasting 12 minutes, involved a participant who strongly rejected the SAA and provided brief responses.

Demographics

The 10 female and 8 male participating patients had an average age of 35.6 (SD 12.94) years. Most interviewees attained the highest level of school-leaving qualifications (n=10) and the highest level of educational qualifications (n=8). All 18

XSL•FO
RenderX

interviewees had their place of residence in an urban area. Nearly one-third of the interviewees were born abroad. For details see Table 1.

Overview of Categories

Our analysis led to the creation of a system comprising 5 principal categories, as detailed in Table 2. Each of these main categories was further divided into 2 to 7 subcategories. The system of categories, including the definitions of the 5 main categories and their respective subcategories, can be found in the appendix (Multimedia Appendix 2). A table of selected interview quotes is also included (Multimedia Appendix 3).

Table . Overview of categories and subcategories.

Category	Subcategories
Health-related information gathering	• Previous experience with SAAs[a] • Experience with alternative/present information paths/types • Basic attitude toward digital information gathering, particularly regarding SAAs
Doctor-patient relationship	• Evaluation of the doctor-patient relationship during the study • Impact of SAA use on doctor-patient interaction during the study • Basic attitude of patients toward medical doctors • Differences/comparison between SAA and standard medical care • Patients' trust in SAA, medical professionals' expertise, and their own health literacy • Tension/dilemma between informed patients and medical expertise • Patient dissatisfaction with missing/unclear diagnosis
Experiences with using the SAA "Ada"	• Congruence or incongruence of the "SAA diagnosis" with the medical diagnosis/with own assumption • Patients' expectations toward the SAA before the first use • Influence of SAA assessment report and the entire SAA use on own understanding of disease and emotional state • App: positive issues/advantages • App: negative issues/disadvantages • Suggestions for improvement and requests regarding the use of the Ada app • Personal conclusions after SAA use
User groups/application scenarios for SAAs	• Appropriate users + "necessary competencies" • Inappropriate users + "impeding/aggravating properties" • Application scenarios/criteria
Future/potential integration of SAAs in the care setting	• Advantages/potentials of (future) SAA use • Disadvantages/risks of (future) SAA use

[a]SAA: symptom assessment app.

Patient Perspectives on Using an SAA Prior to Consultation in the ED or on-Call Practice

We present detailed findings from the 4 subcategories that yielded data directly relevant to our research questions.

Impact of SAA Use on Doctor-Patient Interaction During the Study

Of the 18 interviewees, 3 (16.7%) reported that the use of the SAA had an influence on the doctor-patient interaction; specifically, they felt better prepared and more reassured when their suspected diagnosis was confirmed by the SAA, or they appreciated the opportunity to compare the results with the physician's assessment:

> *I definitely think that, yeah, you already have it in the back of your mind. And I think that you also somehow go into a conversation like that differently, because you already have some information about what it could be, and then, of course, you compare that with the information the doctor gives you.* [ID13, 59]

Two participants were uncertain about its impact, while the majority, 13 out of 18 (72.2%), stated that the use of the SAA had no influence on the doctor-patient interaction. Moreover, 16 participants reported that the use of the SAA was not addressed during the consultation, neither by the physician nor by the patient. Participants cited different reasons for this, including not feeling the need to address it because they considered the physician's assessment plausible or identical to the SAA's or to their own assessment, not bringing it up because the doctor did not, not thinking of it, forgetting to mention it, feeling that the doctor was under time pressure, being too tired to discuss it, or feeling that the SAA use did not play any role in the encounter:

> *I have to admit, the moment I was sitting in that room waiting for the doctor—or when the doctor walked*

XSL•FO
RenderX

> *in—I wasn't really thinking about the app anymore. It had faded into the background.* [ID04, 51]

Two participants mentioned their involvement in the AkuSym study to their doctors; however, the use of the SAA itself was not discussed. One participant had prior experience with a different SAA. Of the 18 interviewees, 17 (94.4%) stated that they would consider using an SAA again in the future, including remote use.

Patients' Trust in Medical Professionals' Expertise, the SAA Assessment, and Their Own Health Literacy

The possibility of an SAA substituting for direct doctor-patient interaction was not a focus of the study, a detail made clear to participants from the start. However, many interviewees still linked SAA use to traditional medical care, viewing it as an alternative form of consultation. They ultimately viewed SAAs as a supplement to, or a second opinion for, the traditional doctor-patient dialogue, rather than a replacement. This preference was deeply connected to a greater trust in doctors over SAAs and was attributed to the value of physical examination and the belief that an app cannot match the quality of medical care and information provided by a human doctor:

> *Well, I would say this: the trust in a doctor who can see with their own eyes or maybe even hear how I am doing is definitely higher than the trust in an app.* [ID01, 82]

Interviewees skeptical about the trustworthiness of SAAs also voiced concerns about the credibility of advice given by SAAs, like being urged to seek emergency care. Some interviewees doubted their capacity to accurately understand the SAA assessment in terms of its content and medical jargon. They believed it was nearly impossible for nonexperts to determine the relevance or accuracy of any diagnosis proposed in an SAA assessment report to their specific situations. According to them, diagnosing is strictly a doctor's responsibility:

> *Well, I would say that the language is more scientific, medical, so to speak. And as a layperson, it would be difficult for me to say, okay, it is this or that or that? I would really say that a doctor definitely has to make the decision.* [ID10, 19]

Apart from comprehending the SAA assessment, some interviewees doubted their ability to correctly input data into the SAA's symptom query. As a result, certain participants expressed concern that doctors might be swayed by inaccurate outcomes in the SAA report caused by improper data entry. Concurrently, they anticipated that their doctor would identify any inaccuracies in data entry. They wished for doctors to maintain trust in their professional judgment when diagnosing (ID04, 69).

Nonetheless, most interviewees valued the use of an SAA for offering preliminary findings or an initial evaluation (ID13, 11). Moreover, a few expressed a preference for SAAs over search engines when seeking evidence-based information on health issues (ID10, 87).

Advantages and Potential of (Future) SAA Use

Care Recommendations and Guidance

When discussing the potential benefits of integrating SAAs into the standard practice of emergency medical care, many interviewees noted that, if proven reliable, these tools could help determine the necessity and type of medical assistance required, such as whether to consult a general practitioner, visit the ED, or an outpatient on-call practice.

Furthermore, the study participants expressed a desire for SAAs to provide self-help tips or action recommendations for minor ailments. They believe that a more informed approach to seeking medical care could help reduce unnecessary medical consultations, including visits to the general practitioner or the ED:

> *I think I would also use this to assess in advance whether it is serious enough that I need to see a doctor. Or is it so bad that I need to go to the emergency room? Wouldn't the general practitioner be sufficient for now? Or could I perhaps help myself first? I think this could also help prevent unnecessary doctor visits.* [ID13, 67]

Additionally, the importance of possessing preliminary information to actively participate in the diagnostic process was highlighted. According to the participants, this involves discussing potential diagnoses collaboratively with the physician, on an equal basis, while still honoring the physician's professional expertise (ID13, 81).

The precision and detail of the particular SAA's symptom query were appreciated. A few mentioned that it allowed them to share more details about their symptoms than is typically possible in a doctor-patient conversation, which can be limited by time constraints (ID14, 62). An advantage pointed out by many participants was that the SAA use enabled them to reflect on and articulate their symptoms more thoroughly:

> *It definitely gave me a sense of more security. I believe that having an overview in my mind of what points have actually occurred, what the symptoms are, was also good for me to gain clarity. Since when have I had which symptoms, and so on.* [ID14, 56]

Some interviewees believed that being able to describe their symptoms more accurately, or having fewer questions after using an SAA, could lead to more time-efficient medical consultations.

Moreover, some participants suggested that SAAs, due to their AI algorithms, might excel at generating differential diagnoses compared with doctors. They speculated that comparing the SAA report with a physician's evaluation could also offer practitioners valuable insights, ensuring that all potential diagnoses are considered or providing fresh perspectives (ID01, 90).

Building on these perceived benefits, participants also reflected on how SAAs could be practically integrated into the emergency care process. Their suggestions covered 4 key areas: accessible and voluntary use, integration into clinical encounters, care navigation, and digital infrastructure and data sharing.

XSL•FO
RenderX

Accessible and Voluntary Use

It was suggested that SAA use in the waiting room should be optional, as not every patient may be able to use the SAA due to a lack of technical skills or poor health conditions. Given the widespread use of smartphones, it was proposed to offer a QR code to patients in the waiting room for voluntary SAA use to support the medical assessment process (ID10, 85).

Integration Into Clinical Encounters

When using the SAA remotely, the question was raised as to whether patients needed to present their SAA reports during treatment discussions or whether health care providers can access this information in advance. One participant suggested making the SAA report available to the ED/doctor via a connection to the electronic patient record. However, some participants questioned the utility of providing doctors with the SAA report because it had not been discussed during their consultation (ID09, 68).

In addition, it was suggested that it could be beneficial for medical treatment if the same SAA used for recording symptoms remotely is also used in the waiting room, particularly when tracking symptoms over extended periods. To support this, a diary feature was proposed, as recalling and describing intermittent or recurrent symptoms during doctor-patient interactions can be challenging.

Care Navigation

If the SAA recommends seeking medical aid, some participants wished to be directed to the appropriate level of health care (general practitioner, on-call practice, and ED) and to receive the respective phone numbers or addresses (ID09, 60) and (ID10, 81).

Digital Infrastructure and Data Sharing

Participants raised questions about how to implement digital interfaces for transferring SAA-generated information to health care professionals while ensuring data protection and security.

In the context of digitalization efforts in Germany, there was a suggestion to develop a central online portal that consolidates various health care services. This would offer a unified access point for all health care apps and services (ID10, 115).

Disadvantages and Risks of (Future) SAA Use

When exploring the downsides and risks associated with using SAAs, patients interviewed frequently highlighted the absence of a human element and the importance of having direct, personal conversations with a doctor about their concerns:

> *Well, I'd rather go to the doctor because I prefer face-to-face rather than entering something into a machine and having it spit something out for me.* [ID02, 13]

> *I think it's a clear disadvantage that it lacks a bit of humanity. I mean, when it comes to illnesses, it's of course very important that there is a certain form of humaneness involved. And an app, well, of course, it hardly has that.* [ID05, 78]

Patients implicitly worried that the app could become a substitute for the doctor, although the SAA was explicitly meant as a supplement in the trial.

Some interviewees found the symptom query of the SAA to be imprecise and incomplete (ID17, 12), voicing concerns that this might lead to inaccuracies or omissions in the information conveyed to doctors. A few participants also worried that their ability to use the app accurately could be affected by poor physical or psychological condition while in the waiting room, potentially leading to incorrect symptom reporting (ID14, 74). Both concerns raised fears that doctors might make incorrect diagnoses, especially if they relied on the SAA's assessment while being distracted or overworked (ID14, 78).

Without having experienced it themselves, participants suggested that alarming diagnoses indicated by the SAA could heighten anxiety and worry, particularly in patients prone to anxiety. It was assumed by some interviewees that this could complicate the treatment consultation, as doctors might need to first address the patient's anxiety instead of proceeding with medical history taking and providing information (ID18, 75). Additionally, the doctor-patient interaction might be prolonged unnecessarily if patients seek further explanations, for example, regarding why certain differential diagnoses suggested by the SAA were dismissed.

The main findings of this study are summarized in Table 3.

XSL•FO
RenderX

Table . Main findings from patient interviews on symptom assessment app (SAA) use prior to consultation in the emergency department (ED) or on-call practice.

Subcategory	Main findings
Impact of SAA use on doctor-patient interaction during the study	• Influence on interaction: only 3 out of 18 interviewees felt the SAA influenced the interaction (eg, feeling reassured or being able to compare results). 13 stated it had no influence, and 2 were unsure whether it had an impact. • Lack of discussion: 16 out of 18 participants did not discuss the SAA with their doctor, due to the perceived plausibility of SAA assessment or the physician's assessment, time pressure, tiredness, forgetting to mention it, or finding it irrelevant to the encounter. • Future Intentions: 17 out of 18 would consider using an SAA again in the future (including remotely).
Patients' trust in medical professionals' expertise, the SAA assessment, and their own health literacy	• Role of SAAs vs doctors: SAAs are viewed strictly as a second opinion or preliminary tool, never a substitute for human doctors. According to participants, physical exams and human empathy generate substantially higher trust in traditional medical care. • Challenges for laypersons: complex medical jargon makes it difficult to evaluate SAA accuracy; concerns exist about inputting data correctly and potentially misleading the doctor. • Perceived value: valued for providing initial findings, self-help information, and being preferred over general search engines for obtaining evidence-based health information.

Advantages and potential of (future) SAA use	• Care guidance and efficiency: helps determine the appropriate level of medical assistance needed (general practitioner, ED, or self-care) to reduce unnecessary visits.
	• Preparation and articulation: enables patients to reflect on, structure, and detail their symptoms more thoroughly, fostering active and equal participation in diagnostic discussions.
	• Clinical advantages: AI algorithms may assist in generating comprehensive differential diagnoses and fresh perspectives that doctors might overlook.
	• Patients' suggestions for integrating SAAs into emergency care:
	  • Accessible and voluntary use: keep SAA use completely optional in waiting rooms via QR codes to accommodate varying tech skills and health conditions.
	  • Clinical integration: connect SAA reports to electronic patient records, use the same SAA consistently across settings (remotely and in the waiting room), and implement a diary feature for tracking intermittent symptoms over time.
	  • Care navigation: direct users to appropriate care levels and provide immediate contact numbers or addresses when medical aid is advised.
	  • Digital infrastructure: establish secure data-sharing interfaces and a unified central online portal for health services.

Disadvantages and risks of (future) SAA use	• Lack of human element: absence of personal, face-to-face empathy, and human touch during vulnerable moments of illness. • Inaccuracies and misdiagnosis: flawed queries or improper data entry (eg, due to poor patient condition) could mislead doctors if relied upon blindly. • Increased anxiety: alarming automated diagnoses could heighten patient anxiety. • Prolonged consultations: consultation time could increase if doctors need to address patients' concerns about alarming app-generated diagnoses or explain why certain app-suggested differential diagnoses were not considered clinically relevant.

Discussion

Principal Findings

This study examined the patient perspective on SAA use in the emergency care setting. Experiences and assessments made by our study participants concerning this topic were mixed and partially conflicting; for example, some interviewees found the symptom query of the app very precise, while others found it somewhat inaccurate or nonexhaustive. However, even those participants who criticized or mentioned insecurities regarding

XSL•FO
RenderX

the process of symptom input or the comprehension of their SAA report intend to reuse the SAA again. Whereas some interviewees believed that the use of an SAA could enhance the efficiency of doctor-patient interactions, some of them also expressed concerns that SAA usage might complicate the treatment conversation. Such a variety of findings concerning the patient perspective on SAA use is congruent with the subjective perspectives of SAA users investigated by Müller et al [30]. The authors revealed a "broad spectrum of experiences but also inconsistent aspects" among their interviewees who had previous experience with SAAs, though apart from the emergency care process [30].

The majority of our interviewees reported that the findings presented in the SAA report had no impact on their conversation with the doctor. This could be due to the fact that the report was not discussed during the consultation, either by the physicians or the patients. A comparison with the results from the physician focus groups, which have been conducted as part of the AkuSym project but are not yet published, may offer further insights into these dynamics. The impact of SAA use on the doctor-patient interaction may also have been limited by the fact that not all physicians reviewed the SAA report prior to seeing their patients, as shown in the RCT conducted within the AkuSym project [31]. Nevertheless, our interviewees frequently stated that using the SAA helped them feel better prepared for the doctor-patient contact, either because it prompted them to reflect on their symptoms or on the diagnoses presented in the report, or because they found it helpful to have preliminary information. The aforementioned RCT did not show any statistically significant positive or negative effects of SAA use prior to the consultation, neither on the doctor-patient relationship nor on patient satisfaction [31]. However, although no measurable effects were found, the influence of the SAA was more frequently perceived as positive rather than negative by both physicians and patients [31].

All but 1 participant in our study concluded to use an SAA again for obtaining preliminary information about health complaints. They expressed a high interest in information on possible causes of their symptoms to decide whether and what kind of professional medical help should be sought or to know whether and what kind of self-help measures are applicable in case of acute but mild or minor complaints. These findings are similar to the results from a feasibility study in general practices in Italy: half of the participating patients concluded to reuse the SAA from home, valuing it as a potential tool for gaining information on initial health complaints, reducing unnecessary medical consultations, and supporting their physicians [20,23].

Our study participants' interest in information about health complaints relates to their striving for health literacy and patient empowerment. Cerezo et al [39] describe patient empowerment as "an enabling process involving a shift in the balance of power, or as an outcome of this process." The statements of our study participants support the assumptions of Altendorf et al [40], who argue that following the evidence-based health information provided by SAAs can strengthen patients' health literacy, thereby facilitating their empowerment. Health care providers and patient-centered care can further enhance the empowerment process, for example, by providing understandable information [41]. A precise and exhaustive symptom query or the provision of comprehensible content and medical terminology for laypersons were key criteria for our study participants to use an SAA again for online health searching.

The findings on health literacy and patient empowerment discussed above conform to the results of a scoping review exploring the impact of SAA use on health care professionals in primary care [26] and to the aforementioned Italian feasibility study in primary care [20,23]. Radionova et al [26] identified 2 potentials of SAAs for layperson empowerment: (1) "providing laypersons with knowledge" on health-related issues as assumable superior alternative to search engines and (2) "layperson activating potential" as laypersons might become more active by assuming self-management and responsibility. Wiedermann et al [23] concluded that SAAs have the potential to empower patients through "offering a sense of agency and participation".

Our participants' aim for patient empowerment and health literacy did not challenge the authority of medical professionals. On the contrary, having experienced the use of an SAA, the interviewed patients clearly stressed out that they continue to have most trust in the expertise of a medical doctor. Similarly, Müller et al [30] figured out that SAA users prioritized their own judgment and the physicians' assessment over the app findings.

Although it was emphasized beforehand that the study's purpose was not to determine whether the SAA could replace doctor-patient contact, participants repeatedly underlined that the use of the SAA could not replace this contact. This attitude was strongly linked to the higher level of trust in doctors, as well as to the importance placed on "the human touch" that comes with personal doctor-patient contact. Consequently, participants perceived the SAA not merely as a supportive tool, but as a potential third actor within the clinical encounter, aligning with the assumptions of Sauerbrei et al [42]. To ensure that the implementation of AI tools positively influences the patient-doctor relationship, the authors recommend actively shaping their role through assistive use and adapting medical education accordingly [42]. This perspective also resonates with work on eHealth ethics in primary care describing a triangulated relationship between patient, eHealth, and primary care professionals, and calling for renewed reflection on the role of human interaction and how shared decision-making is shaped within this triad [43].

Our interviewees not only shared their opinions on the potential integration of SAAs into the emergency care process but also provided suggestions for their practical implementation. One of these suggestions was offering voluntary SAA use via a QR code displayed in the waiting room, while another proposed making the SAA report available to health care providers through the electronic patient record. Such options could be helpful in further exploring the potential of SAAs to contribute to a more efficient medical assessment process for patients assigned a low triage level in an emergency care setting. To implement these options, however, it would be necessary to address challenges in information transfer related to digital

XSL•FO
RenderX

interfaces, as well as data protection and security regulations. This was also pointed out by our interviewees.

If the SAA is to be integrated as a standard procedure in emergency care, attending physicians should be instructed to actively address the SAA results during conversations with patients. This is important to avoid patients questioning the purpose of using the SAA prior to their conversation with the treating physician, a concern expressed by some of our participants. How extensively physicians discuss the SAA results with patients could remain flexible.

Further challenges to the use of SAAs in emergency care relate to patients' digital health literacy. Bundsgaard Andersen et al [44] identified a substantial digital divide among patients receiving emergency care, with 43% (203/468) demonstrating low digital literacy. Patients with lower levels of digital literacy were older, frailer, had more hospital admissions during the preceding year, and were less likely to have attained a high level of education [44]. This high proportion suggests that limited digital literacy may constitute a relevant barrier to effective SAA use and potentially contribute to disparities in patients' access to and engagement with these tools.

In a scoping review of eHealth literacy in emergency care, Paller et al [45] reported heterogeneous findings, highlighting demographic and language factors as areas requiring further investigation. The authors recommend that health care providers consider patients' prior internet use and provide appropriate guidance during the ED encounter [45]. They also advocate educating both patients and health care professionals on the effective use of online health information [45].

These findings suggest that the integration of SAAs into emergency care should account for differences in patients' digital health literacy and ensure that the information provided through such tools is accessible and understandable to diverse patient groups.

In the SAA used in this study (Ada), it is possible to save individual symptom queries, but it does not include the diary feature suggested by 1 participant for tracking symptoms over extended periods. However, there exists a wide range of SAAs, including those that provide recommendations for actions, as well as more extensive functions, such as symptom diaries [26].

If seeking medical aid is recommended as part of the SAA assessment, some interviewees requested to be provided with the contact details of the appropriate health care providers. In Germany, SmED (Structured Initial Medical Assessment in Germany [*Strukturierte medizinische Ersteinschätzung in Deutschland*]; in4medicine AG)-Patient is a chatbot for the structured initial medical assessment of patients [46]. This is a digital tool that can support those seeking adequate medical help for treating their symptoms. After guiding users through a symptom query, it provides an initial assessment recommendation and contacts for health care providers in the surrounding area, including EDs, on-call practices, and general practitioners. In contrast to SAAs, it does not offer any treatment recommendations or self-help tips. To date, no empirical studies have been conducted specifically on SmED-Patient. However, "SMASS" (Swiss Medical Assessment System; in4medicine, Inc), which serves as the basis for SmED-Patient, has been evaluated as safe for use in clinical routine [47].

One participant's suggestion goes beyond the integration of the SAA into emergency care: the idea of developing an online portal centralizing all health services and apps. An evidence-based health information web portal [48,49] was implemented by the Dutch College of General Practitioners in 2012 and, 2 years later, resulted in a 12% less health care usage [50]. Although not app-based, the project exemplarily shows that the provision of evidence-based online health information using understandable language can improve resource use in the health care system, potentially through higher health literacy and patient empowerment.

Based on our study findings, we suggest linking the SAA report to a public website that provides further evidence-based, layperson-friendly information, as well as contact details for the appropriate health care providers.

Strengths and Limitations

Qualitative interviews facilitate a thorough examination of personal views on a particular subject. Consequently, the findings from this study enrich the other components of the AkuSym project and expand current knowledge of SAAs. Researcher bias was mitigated through a collaborative coding process involving 3 independent coders.

One limitation of the study concerns the relatively short duration of the interviews. This may be partly due to the interview guideline, which did not focus on biographical aspects or long-term experiences, topics that might have elicited more extensive responses. Additionally, the use of the SAA lasted no more than 10 minutes. Except for 1 participant, none of the interviewees had prior experience using an SAA, which further limited the extent they could report based on past use.

However, this aspect can also be seen as a strength of the study: the perspectives gathered primarily reflect the impressions of first-time users who did not actively seek out the SAA themselves. While the sample is not representative in a statistical sense, it may still offer valuable insights into how individuals in the general population might perceive SAAs when encountering them in a broader, system-integrated context.

All 3 study sites were in a central city district of Berlin. The findings of this study, therefore, might not be directly applicable to more rural regions, where outpatient on-call practices are organized differently, or to health care systems with more integrated solutions, such as that in the Netherlands [48,49]. However, the accessibility of a digital tool can be considered a strength in rural regions, where access to traditional health care is often more difficult than in urban settings. Specifically, using an SAA may prompt someone to drive the distance to care when they might not have done so otherwise.

The study participants sought medical assistance in the ED or the on-call practice on their own, without referral from another doctor and without showing symptoms being part of a known chronic disease. They thereby met the recruitment criteria of the RCT. This probably explains the mean age of the interviewees representing the group of middle adulthood.

XSL•FO
RenderX

Besides that, people using digital and online assessment services, including SAAs, are more likely to be "younger and more highly educated" [2]. Also in this study, the interviewees showed high levels of school leaving and educational qualifications. A notable proportion of our study participants (n=7) held a university degree. This characteristic was also observed in the underlying RCT population from which the interviewees were recruited, suggesting that the relatively high level of educational attainment in our qualitative sample may partly reflect the educational characteristics of the RCT population. Given that higher educational attainment is closely linked to increased health literacy [51], this demographic characteristic may have contributed to a more positive perception of the SAA. Consequently, the target group of the interviews for this study was met, but the results obtained from the interviewees are not transferable to other groups.

Conclusions

This qualitative study explored patient perspectives on SAA use in emergency care, particularly the perceived benefits and challenges of integrating SAAs into the emergency care process. Participants valued the SAA for providing preliminary information about their health complaints and helping them reflect on and articulate their symptoms. They also saw potential for SAAs to support decisions about whether and what type of medical care might be appropriate, provide self-help recommendations for minor complaints, and enable more active participation in discussions with physicians. However, these benefits depend on reliable, comprehensive, and layperson-friendly information.

Participants also identified potential drawbacks and risks associated with SAA use in emergency care, including difficulties with symptom input and with the interpretation of SAA assessments, as well as the potential for alarming or unclear information to increase anxiety or complicate the subsequent consultation. The meaningful integration of SAAs into emergency care requires a clear understanding of how SAA assessments should be used during the doctor-patient consultation. Challenges include ensuring secure information transfer and data protection, as well as accessibility for patients with limited digital health literacy. Participants suggested practical measures to address these challenges, including voluntary access to SAAs via a QR code in the waiting room, linking SAA reports to the electronic patient record, and guiding patients to appropriate health care services.

The findings of this study suggest that SAA use may support health literacy, patient empowerment, and more informed decisions about seeking care, potentially contributing to more efficient use of health care. Linking SAA reports to a public website that provides evidence-based, layperson-friendly information and contact details for appropriate health care providers could help patients understand their results and identify suitable care options. Further research should investigate how such approaches can be implemented and evaluated in emergency care, including remote use, while keeping personal doctor-patient interaction at the center of care.

Acknowledgments

The authors thank all physicians who participated in the AkuSym project.

Generative AI Use

The authors attest that there was no use of generative AI technology in the generation of text, figures, or other informational content of this manuscript.

Funding

This study was supported by the German Ministry of Health (grant 2521TEL500).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: FB (lead), MLS (colead)
Supervision: CH, KS
Funding acquisition: FB, AS, KS, MLS, MM, MK, HN
Investigation: AS, MB, MM, CS (supporting contributions from DK and CW)
Formal analysis: CS, HN, DK
Writing – original draft: CS, HN
Writing – review & editing: MLS, MK, AT, AS, MB, MM, CH, KS
All authors reviewed the final manuscript.

Conflicts of Interest

CS, DK, AT, KS, CW and CH have reported no conflicts of interest.

XSL•FO
RenderX

HN reports investments in exchange traded funds (ETFs) that invest in technology companies that might profit from the results of this project.
MB reports grants from Thermo Fisher BRAHMS, and Roche Diagnostics outside the submitted work.
MK invests in ETFs that hold shares of various companies in the technology and health care sectors, and holds stocks in organizations that do not gain financially from the publication of this manuscript but might do so in the future;
MM has received grants from Health Care Research Projects and Biomarker Research, as well as personal fees from consulting outside the submitted work.
FB has received personal fees from Elsevier Publishing and personal fees from Medtronic, outside the submitted work.
MLS invests in ETFs that hold shares of various companies in the technology and health care sectors, which do not gain financially from the publication of this manuscript but might do so in the future. In 2014 and 2015, he was an employee at Ada Health GmbH (formerly known as medx GmbH), the developer of the investigated health app.
AS reports grants from Deutsche Forschungsgemeinschaft, grants from Bundesministerium für Bildung und Forschung, grants and personal fees from Zentralinstitut für die Kassenärztliche Versorgung, grants from Thermo Fisher Scientific, grants from Roche Diagnostics, grants from Bundesministerium für Gesundheit, grants from Innovationsfonds des Gemeinsamen Bundesausschuss, outside the submitted work

Multimedia Appendix 1
Interview guide for the AkuSym interview study.
[PDF File, 97 KB - humanfactors_v13i1e101285_app1.pdf]

Multimedia Appendix 2
Codebook for the AkuSym interview study.
[PDF File, 200 KB - humanfactors_v13i1e101285_app2.pdf]

Multimedia Appendix 3
Selected quotes from the AkuSym interview study.
[PDF File, 101 KB - humanfactors_v13i1e101285_app3.pdf]

References

1. Cotte F, Mueller T, Gilbert S, et al. Safety of triage self-assessment using a symptom assessment app for walk-in patients in the emergency care setting: observational prospective cross-sectional study. JMIR mHealth uHealth 2022 Mar 28;10(3):e32340. [doi: 10.2196/32340] [Medline: 35343909]
2. Chambers D, Cantrell AJ, Johnson M, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. BMJ Open 2019 Aug 1;9(8):e027743. [doi: 10.1136/bmjopen-2018-027743] [Medline: 31375610]
3. Wallace W, Chan C, Chidambaram S, et al. The diagnostic and triage accuracy of digital and online symptom checker tools: a systematic review. NPJ Digit Med 2022 Aug 17;5(1):118. [doi: 10.1038/s41746-022-00667-w] [Medline: 35977992]
4. Hennemann S, Kuhn S, Witthöft M, Jungmann SM. Diagnostic performance of an app-based symptom checker in mental disorders: comparative study in psychotherapy outpatients. JMIR Ment Health 2022 Jan 31;9(1):e32832. [doi: 10.2196/32832] [Medline: 35099395]
5. Fraser HSF, Cohan G, Koehler C, et al. Evaluation of diagnostic and triage accuracy and usability of a symptom checker in an emergency department: observational study. JMIR mHealth uHealth 2022 Sep 19;10(9):e38364. [doi: 10.2196/38364] [Medline: 36121688]
6. Gräf M, Knitza J, Leipe J, et al. Comparison of physician and artificial intelligence-based symptom checker diagnostic accuracy. Rheumatol Int 2022 Dec;42(12):2167-2176. [doi: 10.1007/s00296-022-05202-4] [Medline: 36087130]
7. Chan F, Lai S, Pieterman M, et al. Performance of a new symptom checker in patient triage: Canadian cohort study. PLoS One 2021;16(12):e0260696. [doi: 10.1371/journal.pone.0260696] [Medline: 34852016]
8. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015 Jul 8;351:h3480. [doi: 10.1136/bmj.h3480] [Medline: 26157077]
9. Schmieding ML, Mörgeli R, Schmieding MAL, Feufel MA, Balzer F. Benchmarking triage capability of symptom checkers against that of medical laypersons: survey study. J Med Internet Res 2021 Mar 10;23(3):e24475. [doi: 10.2196/24475] [Medline: 33688845]
10. Schmieding ML, Kopka M, Schmidt K, Schulz-Niethammer S, Balzer F, Feufel MA. Triage accuracy of symptom checker apps: 5-year follow-up evaluation. J Med Internet Res 2022 May 10;24(5):e31810. [doi: 10.2196/31810] [Medline: 35536633]
11. Kopka M, Wang SM, Kunz S, Schmid C, Feufel MA. Technology-supported self-triage decision making. Npj Health Syst 2025 Jan 25;2(1):3. [doi: 10.1038/s44401-024-00008-x] [Medline: 42527456]
12. Hill MG, Sim M, Mills B. The quality of diagnosis and triage advice provided by free online symptom checkers and apps in Australia. Med J Aust 2020 Jun;212(11):514-519. [doi: 10.5694/mja2.50600] [Medline: 32391611]

XSL•FO
RenderX

13. Morse KE, Ostberg NP, Jones VG, Chan AS. Use characteristics and triage acuity of a digital symptom checker in a large integrated health system: population-based descriptive study. J Med Internet Res 2020 Nov 30;22(11):e20549. [doi: 10.2196/20549] [Medline: 33170799]
14. Gilbert S, Mehl A, Baluch A, et al. How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. BMJ Open 2020 Dec 16;10(12):e040269. [doi: 10.1136/bmjopen-2020-040269] [Medline: 33328258]
15. Riboli-Sasco E, El-Osta A, Alaa A, et al. Triage and diagnostic accuracy of online symptom checkers: systematic review. J Med Internet Res 2023 Jun 2;25:e43803. [doi: 10.2196/43803] [Medline: 37266983]
16. Kopka M, Napierala H, Privoznik M, Sapunova D, Zhang S, Feufel MA. The RepVig framework for designing use-case specific representative vignettes and evaluating triage accuracy of laypeople and symptom assessment applications. Sci Rep 2024 Dec 23;14(1):30614. [doi: 10.1038/s41598-024-83844-z] [Medline: 39715767]
17. Kopka M, von Kalckreuth N, Feufel MA. Accuracy of online symptom assessment applications, large language models, and laypeople for self-triage decisions. NPJ Digit Med 2025 Mar 25;8(1):178. [doi: 10.1038/s41746-025-01566-6] [Medline: 40133390]
18. Knitza J, Muehlensiepen F, Ignatyev Y, et al. Patient's perception of digital symptom assessment technologies in rheumatology: results from a multicentre study. Front Public Health 2022;10:844669. [doi: 10.3389/fpubh.2022.844669] [Medline: 35273944]
19. Miller S, Gilbert S, Virani V, Wicks P. Patients' utilization and perception of an artificial intelligence-based symptom assessment and advice technology in a British primary care waiting room: exploratory pilot study. JMIR Hum Factors 2020 Jul 10;7(3):e19713. [doi: 10.2196/19713] [Medline: 32540836]
20. Mahlknecht A, Engl A, Piccoliori G, Wiedermann CJ. Supporting primary care through symptom checking artificial intelligence: a study of patient and physician attitudes in Italian general practice. BMC Prim Care 2023 Sep 4;24(1):174. [doi: 10.1186/s12875-023-02143-0] [Medline: 37661285]
21. Kopka M, Scatturin L, Napierala H, et al. Characteristics of users and nonusers of symptom checkers in Germany: cross-sectional survey study. J Med Internet Res 2023 Jun 20;25(1):e46231. [doi: 10.2196/46231] [Medline: 37338970]
22. Scheder-Bieschin J, Blümke B, de Buijzer E, et al. Improving emergency department patient-physician conversation through an artificial intelligence symptom-taking tool: mixed methods pilot observational study. JMIR Form Res 2022 Feb 7;6(2):e28199. [doi: 10.2196/28199] [Medline: 35129452]
23. Wiedermann CJ, Mahlknecht A, Piccoliori G, Engl A. Redesigning primary care: the emergence of artificial-intelligence-driven symptom diagnostic tools. J Pers Med 2023 Sep 15;13(9):1379. [doi: 10.3390/jpm13091379] [Medline: 37763147]
24. Meyer AND, Giardina TD, Spitzmueller C, Shahid U, Scott TMT, Singh H. Patient perspectives on the usefulness of an artificial intelligence-assisted symptom checker: cross-sectional survey study. J Med Internet Res 2020 Jan 30;22(1):e14679. [doi: 10.2196/14679] [Medline: 32012052]
25. Wetzel AJ, Klemmt M, Müller R, Rieger MA, Joos S, Koch R. Only the anxious ones? Identifying characteristics of symptom checker app users: a cross-sectional survey. BMC Med Inform Decis Mak 2024 Jan 23;24(1):21. [doi: 10.1186/s12911-024-02430-5] [Medline: 38262993]
26. Radionova N, Ög E, Wetzel AJ, Rieger MA, Preiser C. Impacts of symptom checkers for laypersons' self-diagnosis on physicians in primary care: scoping review. J Med Internet Res 2023 May 29;25(1):e39219. [doi: 10.2196/39219] [Medline: 37247214]
27. Haase CB, Ajjawi R, Bearman M, Brodersen JB, Risor T, Hoeyer K. Data as symptom: doctors' responses to patient-provided data in general practice. Soc Stud Sci 2023 Aug;53(4):522-544. [doi: 10.1177/03063127231164345] [Medline: 37096688]
28. Wetzel AJ, Koch R, Koch N, et al. "Better see a doctor?" Status quo of symptom checker apps in Germany: a cross-sectional survey with a mixed-methods design (CHECK.APP). Digit Health 2024;10:20552076241231555. [doi: 10.1177/20552076241231555] [Medline: 38434790]
29. Gottliebsen K, Petersson G. Limited evidence of benefits of patient operated intelligent primary care triage tools: findings of a literature review. BMJ Health Care Inform 2020 May;27(1):e100114. [doi: 10.1136/bmjhci-2019-100114] [Medline: 32385041]
30. Müller R, Klemmt M, Koch R, et al. "That's just future medicine" - a qualitative study on users' experiences of symptom checker apps. BMC Med Ethics 2024 Feb 16;25(1):17. [doi: 10.1186/s12910-024-01011-5] [Medline: 38365749]
31. Schmieding ML, Kopka M, Bolanaki M, et al. Impact of a symptom checker app on patient-physician interaction among self-referred walk-in patients in the emergency department: multicenter, parallel-group, randomized, controlled trial. J Med Internet Res 2025 Apr 2;27:e64028. [doi: 10.2196/64028] [Medline: 40173434]
32. Ada [Article in German]. URL: https://ada.com/de/ [accessed 2025-06-01]
33. Ceney A, Tolond S, Glowinski A, Marks B, Swift S, Palser T. Accuracy of online symptom checkers and the potential impact on service utilisation. PLoS One 2021;16(7):e0254088. [doi: 10.1371/journal.pone.0254088] [Medline: 34265845]
34. Apprenticeship [Artilce in German]. Linguee. URL: https://www.linguee.de/englisch-deutsch/uebersetzung/apprenticeship.html

XSL•FO
RenderX

35. Kuckartz U. Qualitative Inhaltsanalyse Methoden, Praxis, Computerunterstützung [Book in German], 4th edition: Beltz Juventa; 2018.
36. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). : European Union; 2016 URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj [accessed 2026-10-21]
37. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
38. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251. [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
39. Cerezo PG, Juvé-Udina ME, Delgado-Hito P. Concepts and measures of patient empowerment: a comprehensive review. Rev Esc Enferm USP 2016;50(4):667-674. [doi: 10.1590/S0080-623420160000500018] [Medline: 27680054]
40. Altendorf M, Luckmann M, Lägel R, Amelung VE. Kompetente Patient:innen und bessere Entscheidungen? Einsatz von Symptom-Checkern zur Steigerung der Versorgungseffizienz [Article in German]. Monitor Versorgungsforschung 2023:66-70. [doi: 10.24945/MVF.02.23.1866-0533.2487]
41. Hickmann E, Richter P, Schlieter H. All together now - patient engagement, patient empowerment, and associated terms in personal healthcare. BMC Health Serv Res 2022 Sep 2;22(1):1116. [doi: 10.1186/s12913-022-08501-5] [Medline: 36056354]
42. Sauerbrei A, Kerasidou A, Lucivero F, Hallowell N. The impact of artificial intelligence on the person-centred, doctor-patient relationship: some problems and solutions. BMC Med Inform Decis Mak 2023 Apr 20;23(1):73. [doi: 10.1186/s12911-023-02162-y] [Medline: 37081503]
43. Boers SN, Jongsma KR, Lucivero F, et al. SERIES: eHealth in primary care. Part 2: exploring the ethical implications of its application in primary care practice. Eur J Gen Pract 2020 Dec;26(1):26-32. [doi: 10.1080/13814788.2019.1678958] [Medline: 31663394]
44. Bundsgaard Andersen L, Juul Larsen J, Marsaa K, Rosenmai G, Seemann H, Andersen Schmidt T. Quantitative evaluation of patients' digital capability evaluated in an emergency department setting: a cross-sectional study. Emerg Med J 2025 Nov 21;42(12):785-790. [doi: 10.1136/emermed-2024-213999] [Medline: 40461065]
45. Paller J, Haugen AS, Großschädl F. eHealth literacy in emergency care: a scoping review. BMC Health Serv Res 2026 May 7;26(1):984. [doi: 10.1186/s12913-026-14637-5] [Medline: 42121136]
46. SmED - Strukturierte medizinische Ersteinschätzung in Deutschland: Übersicht [Article in German]. Zentralinstitut für die kassenärztliche Versorgung (Zi). URL: https://www.zi.de/themen/medizin/smed/uebersicht#c3328 [accessed 2025-06-17]
47. Meer A, Rahm P, Schwendinger M, et al. A symptom-checker for adult patients visiting an interdisciplinary emergency care center and the safety of patient self-triage: real-life prospective evaluation. J Med Internet Res 2024 Jun 27;26:e58157. [doi: 10.2196/58157] [Medline: 38809606]
48. Thuisarts [Article in Dutch]. URL: https://www.thuisarts.nl/ [accessed 2026-10-13]
49. Drenthen T, Beijaert RPH, Jansen PWM, Korevaar JC, Smeele IJM. What do you think of Thuisarts.nl? Experiences after 3 years of www.Thuisarts.nl [Article in Dutch]. Ned Tijdschr Geneeskd 2014;158:A8282. [Medline: 25515391]
50. Spoelman WA, Bonten TN, de Waal MWM, et al. Effect of an evidence-based website on healthcare usage: an interrupted time-series study. BMJ Open 2016 Nov 9;6(11):e013166. [doi: 10.1136/bmjopen-2016-013166] [Medline: 28186945]
51. Santillan-Garcia A, Gea-Caballero V, Frutos-Bernal E, et al. Health literacy and self-care in patients with heart failure: a cross-sectional study. Front Public Health 2025;13:1657807. [doi: 10.3389/fpubh.2025.1657807] [Medline: 41648760]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
ED: emergency department
RCT: randomized controlled trial
SAA: symptom assessment app
SMASS: Swiss Medical Assessment System
SmED: Structured Initial Medical Assessment in Germany
SRQR: Standards for Reporting Qualitative Research

XSL•FO
RenderX

Edited by A Kushniruk; submitted 13.May.2026; peer-reviewed by L Alpay, M Hill; revised version received 16.Sep.2026; accepted 02.Oct.2026; published 16.Nov.2026.

Please cite as:
Schorr C, Schmieding ML, Kuschick D, Balzer F, Bolanaki M, Heintze C, Kopka M, Möckel M, Slagman A, Thissen A, Wäscher C, Schmidt K, Napierala H
Patient Perspectives on a Symptom Assessment App in Emergency Care: Interview Study Among Patients in the AkuSym Trial
JMIR Hum Factors 2026;13:e101285
URL: https://humanfactors.jmir.org/2026/1/e101285
doi:10.2196/101285

© Claudia Schorr, Malte L Schmieding, Doreen Kuschick, Felix Balzer, Myrto Bolanaki, Christoph Heintze, Marvin Kopka, Martin Möckel, Anna Slagman, Alica Thissen, Cornelia Wäscher, Konrad Schmidt, Hendrik Napierala. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 16.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Workforce Perceptions of a Conceptual Hospital-at-Home Digital Twin: Qualitative Study

Faiza Yahya[1], MPharm; Matthew Cooper[1,2], BSc, MSc, PhD; Lauren Lawson[1], BSc, MSc; Hamde Nazar[1,3], MPharm, PhD

[1]National Institute for Health and Care Research (NIHR) Newcastle Patient Safety Research Collaboration, Newcastle University, Newcastle upon Tyne, United Kingdom

[2]School of Psychology, Newcastle University, Newcastle upon Tyne, United Kingdom

[3]School of Pharmacy and Biomolecular Sciences, Royal College of Surgeons in Ireland (RCSI), Dublin, Ireland

Corresponding Author:
Faiza Yahya, MPharm
National Institute for Health and Care Research (NIHR) Newcastle Patient Safety Research Collaboration, Newcastle University, Newcastle upon Tyne, United Kingdom

Abstract

Background: Hospital-at-home (HaH) services are transformative for delivering acute care in a patient's home, with support from a multidisciplinary team and remote monitoring technologies, and there is a growing demand for innovation and digital solutions. Digital twin (DT) technology offers the possibility of leveraging real-time data to personalize care, predict risk, and improve patient management. However, perceptions of the workforce regarding the application of HaH DTs remain underexplored, and addressing this is crucial for successful adoption and sustainability.

Objective: This qualitative study aimed to explore health and care professionals' perceptions of the concept of a HaH DT model of care and its perceived impact.

Methods: A total of 26 semistructured interviews were conducted between June and December 2025. Participants were based in the United Kingdom and were recruited through purposive sampling via known contacts, professional networking, and social media. A short vignette on DTs was presented to participants at the start of the interview. An interview topic guide centered around acceptability and the domains of the Non-adoption, Abandonment, Scale-up, Spread, and Sustainability (NASSS) framework was used to explore complexity in implementation and potential management. Data analysis involved a hybrid approach drawing on reflexive thematic analysis.

Results: Seven overarching themes were identified, highlighting a mixture of perceived benefits, concerns, and recommendations. Participants noted value in personalized, holistic care enabled by real-time data-driven interventions and appropriate patient-clinician contact. However, concerns included trust in DTs, with participants highlighting that implementation requires adapting workforce roles, strong education and training, system-level infrastructure, and a supporting evidence base.

Conclusions: This work provides the foundations for stakeholders and developers to better understand the likely adoption of a HaH DT and the likely impact on individuals, organizations, and systems. Findings from this study highlight key recommendations, and further research and implementation efforts should prioritize stakeholder involvement to develop clear actionable plans.

(JMIR Hum Factors 2026;13:e96645) doi:10.2196/96645

KEYWORDS

hospital-at-home; digital twin; artificial intelligence; machine learning; risk prediction; digital health; staff; workforce; implementation; adoption

Introduction

Hospital-at-home (HaH) services (sometimes referred to as virtual wards) are innovative models aimed at delivering hospital-level care within a patient's home, using a multidisciplinary team approach and remote monitoring technology [1]. Evidence has shown that patients prefer to be cared for at home, and HaH can reduce the need for residential care after patients are discharged from the hospital [2]. There is a significant need to ensure patients are cared for safely and effectively, in the right place, and at the right time. However, caring for patients at home who are acutely unwell comes with its own clinical and operational challenges, particularly for aging populations with increasingly complex health needs [3]. Despite the potential benefits of HaH models, their implementation has been variable, especially the adoption of technology and the consistency in how patients are managed [1,4-9]. Unlike hospital wards, these innovative workflows mean that clinicians are often managing patients remotely with

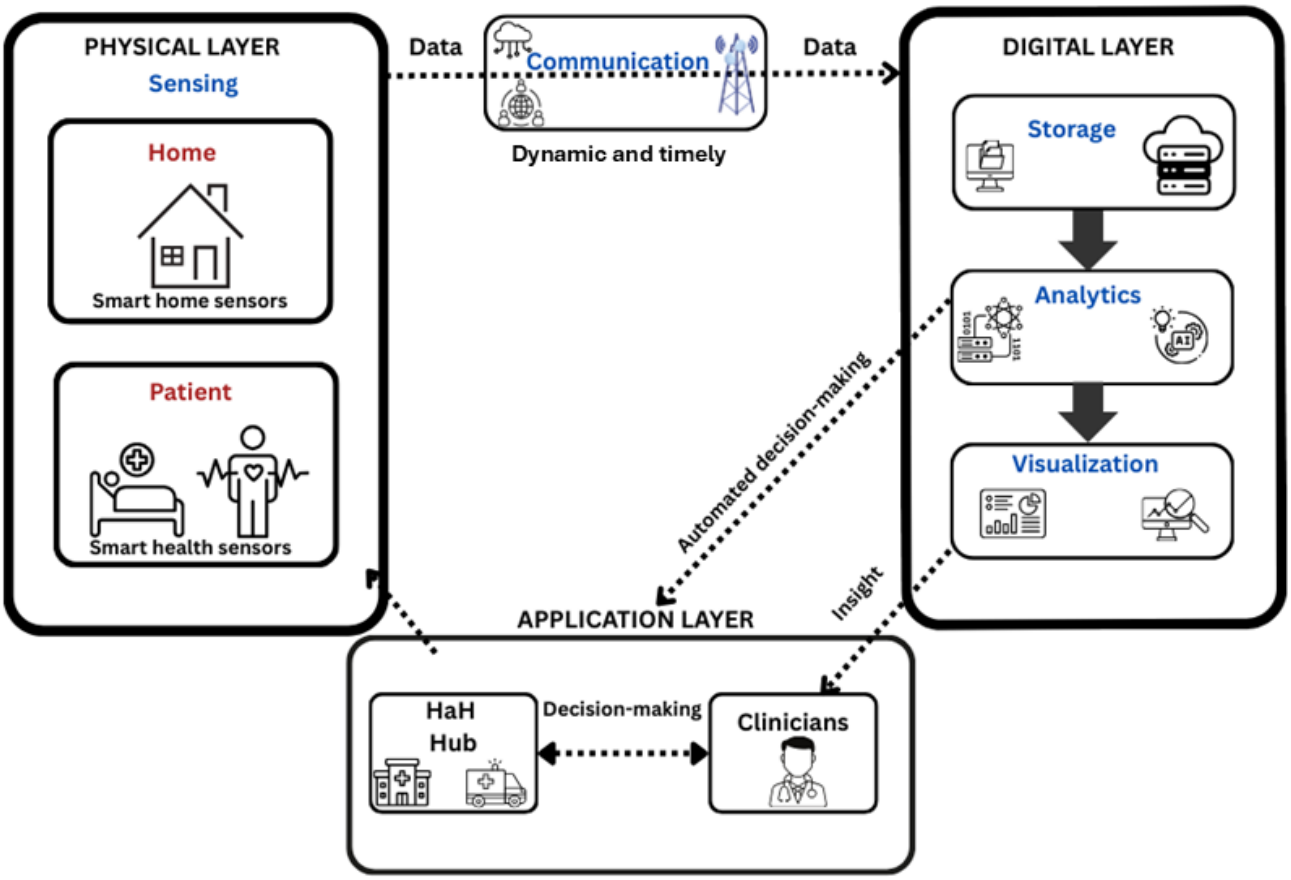
XSL•FO
RenderX

limited direct observation, relying on decisions that are sometimes made on patient-reported intermittent physiological measurements. This can create safety blind spots, making it difficult to identify early deterioration, prioritize caseload, and maintain oversight. Increasing service demands and workforce pressures further challenge clinicians' abilities to maintain remote situational awareness, requiring approaches that can support and streamline their cognitive workload and case management for delivering timely interventions. Compared with typical clinical settings, there is a greater need for proactive and effective risk assessment, status prediction, and care management rather than reactive care [10].

In the era of advancing digital solutions, there has been a strong emphasis on embracing technology to support streamlined, more efficient health care delivery. In England, the National Health Service (NHS) 10-year health plan focuses on 3 major shifts: hospital to community, sickness to prevention (more proactive care), and analog to digital [11]. Current HaH models use remote patient monitoring (RPM) technologies, and some have virtual dashboards to support decision-making. However, their adoption has been variable. These models primarily present current or historical data, failing to reflect a patient's live status, which can fluctuate rapidly in acute care. In contrast, a digital twin (DT) is a dynamic virtual representation of a person and/or their environment, and it continuously integrates real-time data to forecast future possibilities and inform patient management decisions [10,12,13]. Rather than just displaying static physiological observations, a DT can simulate how a patient's condition may evolve over time under different scenarios, enabling the prediction of risk, assessment of intervention options, and prioritization of workflows [10,12,13]. Consequently, DTs have the capability to personalize care for an individual patient and/or setting and offer bidirectional data exchange, predictive simulation, and autonomous decision support functions to support patient management. These capabilities extend above and beyond conventional RPM and dashboards as they are supported by data-driven care planning, leveraging AI and machine learning algorithms. The DT concept involves 5 core architectural components and is presented in Figure 1 [10,13]. The risk prediction capability supports clinicians and service providers in introducing early interventions, leading to improved safety. However, beyond this, DTs can optimize care pathways by integrating patient physiological parameters, home environmental parameters, and service-level operational information. Current evidence points to an expanding body of research [14-16], the development of risk prediction models [17,18], and an increasing application of command centers [10,19] to manage caseloads. Despite this, there is little evidence on the use of DT technology in health care in real-world practice, with many DT applications in health care being in early stages of development and nonexistent in HaH models of care [20]. Building on previous research, we know that the adoption of technology or innovation comes with its own challenges, especially when it comes to human factors and multilevel factors (eg, acceptability, cognitive workload, and human-automation interaction) [21]. As these factors influence sustainability, a fundamental understanding of workforce perspectives is essential, taking into consideration the inherent sociotechnical complexity and the interactions among people, services, and technology. Guided by the Non-adoption, Abandonment, Scale-up, Spread, and Sustainability (NASSS) framework of health and care technologies [21], the aim of this study was to explore the perceptions of health and care professionals and the perceived impact of the concept of a DT-enabled HaH model of care in order to anticipate and plan for the potential adoption or nonadoption of such technology in the future.

XSL•FO
RenderX

Figure 1. Simplified conceptual model for the hospital-at-home (HaH) digital twin architecture [10].



Methods

Study Approach

This qualitative study used semistructured interviews and reflexive thematic analysis. The study has been reported in line with the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist (Checklist 1) [22].

Study Design

Theoretical Framework

This study sought to gain an understanding of health and care professionals' perspectives on a HaH DT concept and its acceptability. The semistructured interviews were guided by a combination of the theoretical framework of acceptability (TFA) [23] and the NASSS framework [21]. The TFA [23] is useful to explore the individual-level factors of those delivering or receiving a health care intervention, and the NASSS framework [21] is useful to explore organizational factors and the wider system across the following 7 domains: condition, technology, value proposition, adopters, organization, wider system, and adaptation over time (Multimedia Appendix 1).

The NASSS framework was selected as it provides a sociotechnical perspective on the complexity of the implementation of health and care technologies and has been widely applied to digital health interventions, including telemedicine and virtual care [21,24].

Underpinned by theories of individual, organizational, and system-level change [21,24,25], the NASSS framework provides a structured approach to understanding implementation complexities and potential pragmatic solutions in complex (adaptive) systems such as the HaH DT concept [24,25]. This approach enabled us to gain a deeper understanding of the perspectives of the workforce at the micro, meso, and macro levels, using a robust methodology and reflexive thematic analysis.

Participant Selection

Participants were recruited by purposive sampling via known contacts, professional networks, forums, conferences, and snowball sampling. Social media were used to support recruitment via professional and social platforms (eg, LinkedIn). Recruitment via these methods aimed to gain a variety of perspectives from a diverse group of health and care professionals based in the United Kingdom and from different backgrounds in any setting (hospital, primary care/community, or HaH).

Consistent with reflexive thematic analysis, decisions regarding sample size were guided by the adequacy and richness of the dataset. Data collection and analysis occurred concurrently, with familiarization and preliminary coding undertaken alongside ongoing recruitment. This was determined by the lead researcher (FY) and regular discussion among the co-researchers after every third interview. Recruitment ceased when additional interviews were no longer contributing sufficiently novel insights relevant to the study aims, and this was determined by the lead researcher's interpretative judgment aligning with reflexive thematic analysis [26].

XSL•FO
RenderX

Data Collection

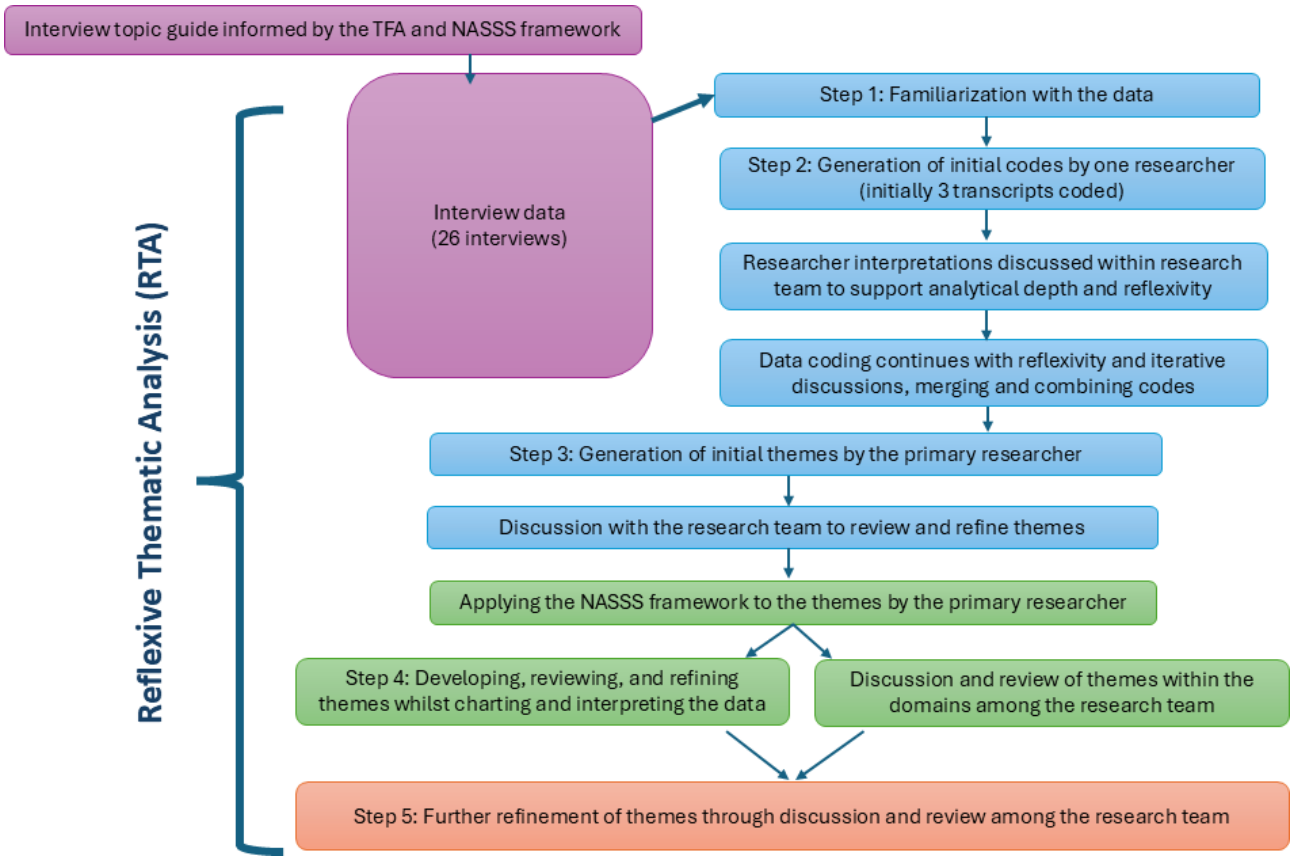
Interviews took place online via Microsoft Teams. An interview guide, which was informed by the TFA [23] and NASSS framework [21], enabled addressing the key themes in the research aims and objectives. Moreover, a short vignette on DTs was presented to participants at the start of the interview (Multimedia Appendix 2). This included a short 4-minute video prepared by the lead author, which explained the concept of HaH, the potential of digital technology, and the advancements of DTs in health care, leaving participants with an understanding of how a DT can operate, its goals, and its constituent parts. Further information can be found in Multimedia Appendix 2. Demographic data were recorded to measure the diversity and inclusion of the sample. The data included professional background, role, geographical location, and level of expertise/training. Interview data were recorded and transcribed verbatim via Microsoft Teams. The interview transcripts were not returned to the participants for comment or checking.

Data Analysis

Overview

The anonymized data were managed in NVivo software (Lumivero), and the development of codes, categories, and themes was an iterative process. Data analysis drew on reflexive thematic analysis following the 6 steps outlined by the Braun and Clarke approach as a general guide [27,28]. The steps are outlined below and illustrated in Figure 2. The process was iterative, moving back and forth between the data and phases to facilitate rigor in data interrogation and engagement [28]. The process was adapted to incorporate the NASSS framework as an interpretative lens to contextualize the findings and build an in-depth understanding for real-world application.

Figure 2. Stepwise methods used for qualitative data analysis. NASSS: Non-adoption, Abandonment, Scale-up, Spread, and Sustainability; TFA: theoretical framework of acceptability.



Step 1

The first step involved familiarization with the data in the transcripts through a systematic review of each transcript. The primary researcher (FY) read and reread each transcript and made reflective notes after reviewing each transcript, becoming immersed and engaged with the data.

Step 2

Initial codes were generated by the primary researcher (FY), with consideration of the research aims and objectives. An inductive coding approach was used, with both semantic and latent coding where applicable. This was done reflexively, with interpretations discussed within the research team to support analytical depth and reflexive engagement with the data. These discussions were intended to enrich interpretation rather than achieve coding agreement or consensus, and they enabled merging and refining of codes before moving to the next step. Examples are provided in Multimedia Appendix 3.

Step 3

Initial themes were generated by the primary researcher (FY), which involved critically engaging with the raw data to organize the initial codes into themes. This was then reviewed multiple times to ensure they captured meaningful patterns before applying the 7 domains of the NASSS framework to the

XSL•FO
RenderX

generated themes. Examples are provided in Multimedia Appendix 4. The iterative analytical process using the NASSS framework aimed to allow greater interpretation by exploring use in the wider context, organizational readiness, and long-term viability [21]. This was done to enable key recommendations to be drawn for further work or implementation in practice.

Step 4

The development, review, and refinement of themes involved discussion with the research team (FY, MC, HN, and LL) as a form of sense-checking to clarify interpretations and challenge assumptions. The coded data extracts were reviewed to refine, combine, or discard any themes as appropriate, ensuring they were meaningful and representative of the data.

Steps 5 and 6

The refinement, definition, and naming of themes involved further iterative, detailed analysis of each theme and drafting of the analytical narrative to define the scope of each theme within the NASSS domains and contextualize the data in relation to the coded data extracts. This process involved flexibility and iterative discussions with the research team. Multimedia Appendices 5 and 6 illustrate parts of this process and provide detailed descriptions.

Research Team and Reflexivity

It is acknowledged that researcher reflexivity inevitably shapes the research and engagement with data [27]. The primary researcher (FY), a female PhD candidate, has a health care professional background (clinical pharmacist) with experience in various health care sectors but has very minimal experience in providing care at a patient's home and has no clinical experience in HaH care. This experience can be viewed as an asset [27], as the researcher presents a degree of subjectivity that allows engagement with participants, supporting the attainment of richness and meaning from the data. The researcher's values as a health care professional and previous knowledge of what matters to patients and the workforce were also considered. Functional reflexivity and previous interviews with HaH clinicians also supported and shaped the development of this research to further enhance understanding. Reflexivity was considered throughout the study, and a reflexive journal was kept by the primary researcher (FY) after each interview to note any influences of their own existing beliefs or experiences in order to minimize any bias in data interpretation. To acknowledge this reflexivity while enhancing transparency and analytical rigor, iterative discussions with experienced qualitative researchers (MC, LL, and HN) were undertaken. Ongoing discussions took place during inductive coding and analysis to support answering the research questions in relation to the NASSS framework.

Ethical Considerations

This study obtained ethical approval from The Faculty of Medical Sciences Research Ethics Committee at Newcastle University (reference number: 61011/2023).

Informed consent was collected from participants before the interview and was rechecked at the interview. Privacy and confidentiality were maintained by anonymizing any participant-identifiable data. To protect anonymity, generic group codes were assigned to professional roles and applied to transcripts (eg, manager, clinical lead, staff nurse, and pharmacist). Any participant details were kept secure on a university password-protected domain in line with university regulations. Participants were remunerated for their participation in the form of a gift voucher (£27.50 [US $36.30]) funded by the National Institute for Health and Care Research (NIHR) Newcastle Patient Safety Research Collaboration in line with the national NIHR guidance [29].

Results

Overview of the Participant Sample

Between June and December 2025, 44 individuals expressed an interest in the study. Twenty-eight consented to participate, but a final total of 26 health and care professionals participated in the semistructured interviews. Participant characteristics are presented in Table 1. Among the participants, 4 had leadership responsibilities within their team, organization, or integrated care board (ICB); 2 were clinical directors; and 1 was a geriatrician.

XSL•FO
RenderX

Table . Participant characteristics.

Characteristic	Value (N=26), n (%)
Profession	
Nurse	8 (31)
Pharmacist	6 (23)
Care coordinator	2 (8)
Physiotherapist	2 (8)
Consultant	2 (8)
Medical registrar	1 (4)
Occupational therapist	1 (4)
Paramedic	1 (4)
Pharmacy technician	1 (4)
Social prescriber	1 (4)
General practitioner	1 (4)
Sector	
Secondary care	8 (31)
Primary care	5 (19)
Community	6 (23)
Both/integrated care	5 (19)
Integrated care board	2 (8)
Region (based on integrated care board locality) [30]	
North East and Yorkshire	6 (23)
North West	6 (23)
Midlands	11 (42)
London	1 (4)
South East	1 (4)
South West	1 (4)

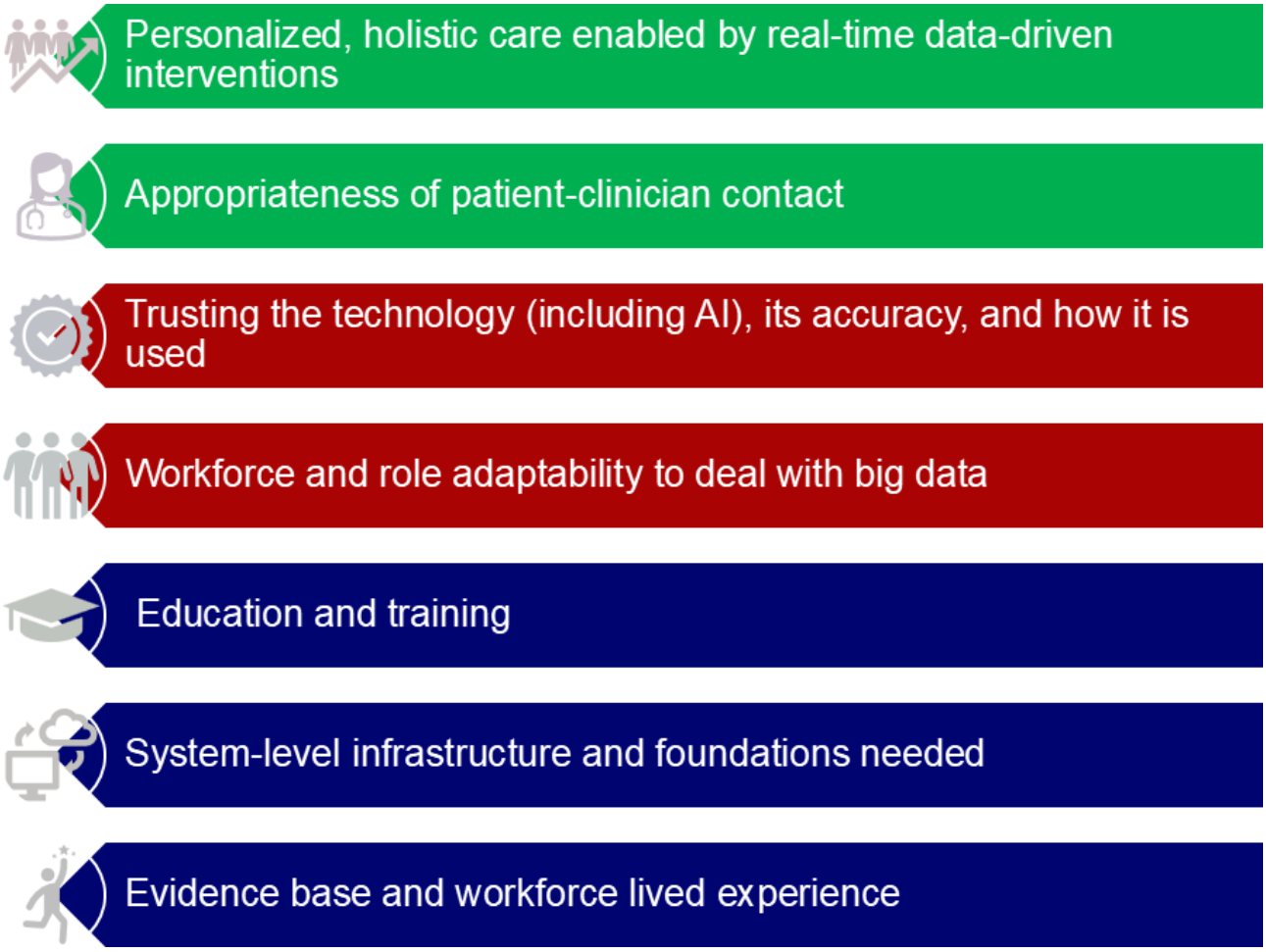
Interviews lasted between 45 and 65 minutes. The average number of years of experience in health care overall was 16.7 years (median 15 years). Among the 26 participants, 16 (62%) had experience in HaH (average experience of 2.5 years; range: 0.5 - 10 years), and of these, the majority (13/16, 81%) had 5 years or less of experience in HaH/virtual wards.

Key Themes From Thematic Analysis

Seven overarching themes were generated from in-depth analysis of the interview data. These are presented in Figure 3 and described in further detail below. Themes 1 and 2 relate to the positive benefits of DTs as perceived by the workforce (this covers factors such as remote situational awareness and individualized predictive algorithms), and themes 2 and 3 relate to current concerns that are likely to change with time (this encompasses factors such as human-automation interaction and sociotechnical factors). It is important to note that there were mixed views (both positive and negative in theme 2). Finally, themes 4 - 7 relate to what the workforce believes needs to happen to support the adoptability of HaH DTs (these may relate to implementation, sociotechnical complexity, and supporting information processing).

XSL•FO
RenderX

Figure 3. Key themes.



Theme 1: Personalized, Holistic Care Enabled by Real-Time Data-Driven Interventions

Participants recognized that DTs enable personalized, individualized care, enhancing a holistic picture. DTs were seen to have substantial transformational value to patients, allowing personalization of care within a patient's context and improving access to care for those who may not normally engage with hospitals when needed. This real-time and connected data facilitated "getting patients the right care at the right time" (participant #10), and the data could potentially "connect the dots" between multiple conditions or health care services in a disjointed health care system. This perceived potential of DTs goes beyond conventional monitoring, giving the opportunity to detect emerging functional or social care needs.

The real-time data capture supports clinicians in tracking micro-level trends (for the patient) and predicting trajectories, and it also provides strategic insights at the meso-level (for workforce and organizational planning) and macro-level (for better population health management).

> *Being able to use a tool like this and figure out on our current caseload which of the people who we could safely defer for a visit or safely not need someone to go out to today in order to take more people in. Now that you know, so not just the in the moment prediction of escalation, but looking at a patient or looking at a person over time on an NHS at home pathway and trying to predict and work out how long they might need to stay, when might they be discharged? When not when? So there's something about that forecasting not just an acute deterioration, but forecasting, Yeah, length of stay and the pattern of patterns beyond … just a deterioration.* [Participant #26]

Theme 2: Appropriateness of Patient-Clinician Contact

This theme grouped both positive and negative feelings from participants. Although there was a perception of reducing physical patient-clinician contact, which can sometimes impact relationships and increase anxiety, it was appreciated that a DT can increase the appropriateness and value of subsequent in-person contact by providing teams with valuable information to send the right clinician to the right patient at the right time. Thus, increasing the value of patient-clinician contact.

> *That's two visits that have had to happen, whereas we could have avoided that by sending the right person the first time to go and review them.* [Participant #9]

Participants reported the need to ensure "the right patients and the right problems" (participant #25) are considered when implementing DTs. Considering this, participants felt that patients with multimorbidity, cognitive and learning disabilities, or severe mental health problems would have complexities

XSL•FO
RenderX

deemed to be unpredictable and high risk for a DT. Participants felt that DTs would be better accepted if patients were given support or an informed choice, drawing comparisons to previous experiences with digital transformation in practice.

> *...just don't tell people, just don't send out a message saying you can't access your prescriptions by phoning this … anymore. Explain why, explain how you can help them.* [Participant #23]

Theme 3: Trusting the Technology (Including AI), Its Accuracy, and How It Is Used

A recurrent theme from participants was human-automation trust and trust in the accuracy of the technology, emphasizing that it takes time to build trust with new technologies like DTs. To clinicians, DTs were seen as a revolutionary decision aid to be used as a tool to support decision-making, but they emphasized that DTs should not replace it. Participants reported that "this is a tool to help with your decision-making, but it is the whole clinical picture that [you need]" (participant #9) and repeatedly mentioned the importance of "clinician instinct and gut-like pattern recognition" (participant #18). They felt technology could not always replicate this, emphasizing that there remains space for skills with clinical judgment and patient choice.

There were differing opinions regarding workforce acceptability, highlighting that this can vary depending on confidence, trust, experience level, digital literacy, and readiness. These initial hesitancies revolved around confidence in trusting the technology and clinical accountability.

> *…they need to be reliant that they're keeping their patients safe because that you do often get a fear that well, if anything were to happen. If that device didn't pick something up, I as a nurse or a registered practitioner would then be liable accountable for that patient deteriorating because I've signed off to use a device and it's failed.* [Participant #21]

Participants felt that with time, embedded training, and increasing familiarity with DTs and AI, the current younger generations and upcoming generations will be more accepting and trusting of DTs. Similarly, participants perceived that this would be the case with patient adopters.

Theme 4: Workforce and Role Remodeling to Deal With Big Data

Participants expressed concerns about the amount of data coming into DTs, worrying that this could "just open the floodgates" (participant #14), with consequences for capacity with increasing data volumes.

> *...you're kind of opening yourself up to getting tons and tons of data, that potentially isn't all relevant, but also that we don't have the staff to deal with either.* [Participant #25]

This requires different ways of working, with staff having to adopt different roles and be adaptable in order to enable data to be reviewed and information to be processed and actioned where necessary. To do this effectively and efficiently, a dedicated role and 24-hour resource were viewed as crucial. Some participants saw this as an opportunity for "a better use of resources and more and more timely" (participant #12), while others had concerns that this is a change from their usual work and clinical training (requiring behavioral adaptation).

> *There's going to be a lot more sitting down and looking at a computer screen rather than going and seeing patients, and I think that's for a lot of people. That's not why they came into this profession.* [Participant #25]

Theme 5: Education and Training

An underlying theme throughout discussions with participants was the need for robust and continual training with DT technology.

> *...ongoing support that we need when we're introducing new things and change… And a proper… ongoing plan and that that is visible to staff and that they know that they're not going to be left in the lurch.* [Participant #26]

A lack of training and understanding was a key factor influencing some of the hesitancies that participants experienced, negatively impacting adoption. Participants explained that education and training need to clearly present the benefits of DTs and be provided to patients, staff, and organizations. This should factor in dedicated resources for training and protocol updates, as well as a dedicated contact for dealing with technical issues. The presence of trained "DT champions" was suggested to support adoptability and reduce training burden on organizations and teams. The data highlighted that the different ways of working with DTs require future training to be embedded at the undergraduate and preregistration levels to ensure that incoming clinicians are accustomed to using such technology, ultimately enhancing its sustainability.

Theme 6: System-Level Infrastructure and Foundations Needed

Current systems still require improvements to ensure digital readiness and capability to leverage the potential of DTs. Participants felt strongly about the lack of interoperability and organizational ergonomics that exist in current health care systems and felt that there is an overwhelming need to ensure that the foundations are rooted first.

> *You see, the foundations are just so different everywhere. And they're so unsteady in areas that you wouldn't think that you'd have to reinforce in order for all of this to work. It's like you've got a drip, a tap dripping at your sink. And before you know it, you're out in the streets digging up the pipes you know because the source of the problem is actually so far away, but you can't progress until you fix that first.* [Participant #15]

This includes interoperability of systems and data sharing to ensure seamless flow of real-time data and shared visibility centered around the patient. Moreover, reliable Wi-Fi access and connectivity are paramount for DTs to be adoptable. This concern focused on avoiding a digital and social divide, which

XSL•FO
RenderX

was previously experienced with technology by some participants.

> *…they were just incredibly rural, so they had no mobile or Internet connection. They couldn't use their services. So straight away it disadvantaged those groups of people.* [Participant #2]

This commitment to system-level infrastructure is required at organizational and wider system levels. In addition to the digital infrastructure required, participants highlighted the need for human resources, flexible workforce structures, and adaptability to support digital readiness and its continued capability.

> *...what is the infrastructure that's going to be in place to what infrastructure do you need to support that? And I don't mean, not only just data infrastructure, but what's the human infrastructure for it, because I think if you're adding in all these additional layers, you're going to need additional IT support additional people. The key bit there was you're sending them home with the enough support and the right care, and then that very quickly.* [Participant #2]

The value proposition was perceived to be high for both the system and individual patients. From a system perspective, participants mentioned that while the initial investment is significant, there is a substantial return on the investment in the long term—giving rise to better population health management, improved workflows, reduced hospital admissions, and better patient outcomes, ultimately supporting the statement, "you've got to spend money to save money" (participant #3). However, having a clear "contingency plan as well for when it's not working" (participant #20) was seen as important, which would require integrated care systems to work together.

Theme 7: Evidence Base and Workforce Lived Experience

DTs are still an unfamiliar concept. To aid successful adoption and sustainability, participants suggested that DTs would require a promising evidence base. This could include positive case studies of use or lived experiences of the workforce, which can build familiarity and trust over time.

> *I think anything, anything that's new is always taken up with skepticism, skepticism. It's like not easily implemented, but once it's implemented and people see the effect and the positive impact. Of course they accept it.* [Participant #13]

Participants generally expressed a positive outlook, agreeing that DTs align with the 10-year plan and long-term health care strategies. They emphasized the need to embrace change, recognizing that workforce culture will change with time, knowledge, and adaptability.

Participants viewed the alignment of DTs with national health policy as a positive facilitator for implementation. Moreover, participants felt that as familiarity with DT technology improves and its benefits for holistic, personalized care become clear, the mindset and adoption will likely change over time.

> *I think it I think it's happening whether we like it or not to be honest and I think we need to just adapt to this happening because it would be madness not to do it really. It's part of the NHS strategy.* [Participant #25]

There was a consensus that evidence (or proof of concept) is needed for clinicians to build trust with DTs. Some participants suggested that initially a phased approach is needed to help build the evidence and proof of concept in the early stages.

This would help identify any unforeseen challenges, uncover technical considerations, and increase familiarity, with participants suggesting "you can double run it, can't you, to start with" (participant #12). Participants questioned whether an acute setting was the best starting point, and they suggested starting with long-term conditions or "starting with relatively low risk standardized care pathways" (participant #25). Disease groups with "clear sort of clinical scenarios" (participant #25) were preferred for proof of concept—such as respiratory/chronic obstructive pulmonary disease, heart failure, influenza, and pneumonia—to identify key learnings that support wider implementation and scale-up over time. Participants provided suggestions for different settings for the initial implementation (Table 2).

Table . Quotes from participants regarding different settings to begin the implementation plan.

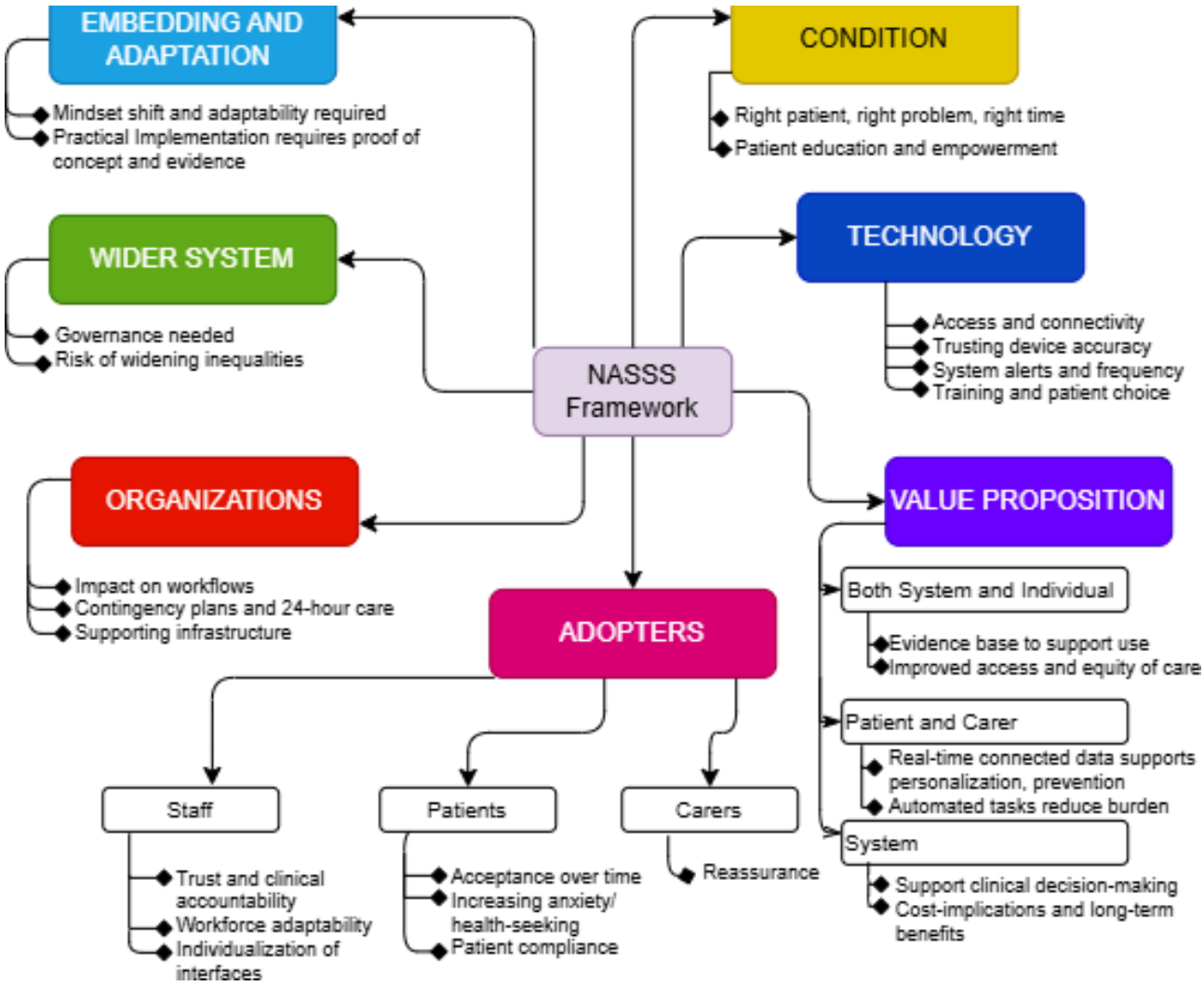
Setting	Quote from participant
Predischarge care (step-down hospital at home)	"…I think, you know, a day or two days before they leave hospital just to get them accustomed to using the technology." (participant #21)
Care homes	"If you were to go to a trial, trial it in like a care home setting, because then obviously you have got people that are able to deal with technology and able to follow instructions to try and see how that's going to pan out." (participant #4)
Long-term condition monitoring (admission avoidance)	"I think you'd probably do it in long term because you would also then learn the acute stuff because you'd start to learn about when people have deteriorating and you'd get that information about what you know, what are the parameters for that person that are changing, how do we pick them up earlier... You'll have those kind of baseline…findings as well." (participant #26)

Integrating the NASSS Framework in Thematic Analysis

We provide an overview of the 7 domains of the NASSS framework for implementation (Figure 4) and any perceived complexity. These are also described in more detail in Multimedia Appendix 6.

XSL•FO
RenderX

Figure 4. Analytical themes integrated into the Non-adoption, Abandonment, Scale-up, Spread, and Sustainability (NASSS) framework for implementation.



Using the NASSS Framework for Implementation and Sustainability

From the synthesis of the data, an in-depth understanding was obtained to identify areas of uncertainty and interdependencies. The 7 NASSS domains can be classified as simple (straightforward, predictable, and few components), complicated (multiple interacting components or issues), or complex (dynamic, unpredictable, and not easily disaggregated into constituent components) [25]. This supports the development of potential actionable solutions that stakeholders may wish to consider to mitigate/reduce complexity during the implementation and sustainability of HaH DTs (Table 3).

XSL•FO
RenderX

Table . Use of the NASSS[a] framework and participant-informed implementation and sustainability considerations for HaH[b] DTs[c].

NASSS domain	Complexity level	Rationale	Proposed action plans	Actors
Condition	Complex	Conditions in acute HaH may present as unpredictable and high risk, with complex multimorbidities. Complexity is further enhanced by continuous monitoring of patient data or by the presence of cognitive impairments.	• Selection criteria for suitability • Appropriate training and education • Escalation pathways • Tailoring monitoring to individual patient need (eg, timed at 15 min intervals) and using patient baseline data	Practitioners, clinicians, service leads, and clinical leads
Technology	Complex	Currently, the technology is still unfamiliar and not yet fully developed. Existing technologies, such as wearables, may not have a clear linkage between data and current health care systems. The predictive capabilities remain conceptual and innovative, presenting uncertainty. Over time, supportive infrastructure is likely to adapt with improved understanding, evidence, and familiarity in practice.	• Training and support • Governance • Interoperability • Data sharing and accurate patient baseline information • Enhanced connectivity for all	Technology developers, system developers, commissioners, integrated care systems, and digital health boards
Value proposition	Complicated	The value to the patient and carer is likely to be significant, reducing burden and ensuring a sense of safety and security. Based on staff perspectives, while this high value to patients can be classified as simple, there is no evidence of desirability, safety, efficacy, or cost. The value to the system is currently uncertain, depending on many factors, including initial costs, cost-effectiveness, and impact on outcomes, safety, and efficacy. The value to clinicians is yet unknown, as no trial has been conducted in practice.	• Proof-of-concept study • Stakeholder engagement • Patient and public involvement and engagement • Qualitative study with patients and carers • Case studies and evaluations around models of care, clinical effectiveness, patient outcomes, and time and economic implications, including environmental impact	Service leads, organizations, and commissioners
Adopters	Simple (workforce-perceived; patients/carers), complicated (staff)	As perceived by workforce participants, the DT would be more acceptable to patient and carer adopters who experience a lower burden. There are more factors to consider with staff adopters, as it would involve learning new skills and a change to routine tasks and potentially roles.	• Education and training for patients, carers, and the workforce • Embed training at the undergraduate level and placements for staff • Stakeholder engagement • Patient and public involvement and engagement • DT champions • Review and plan workflow, roles, and capacity	Service users, patients, carers, family members, and patient and public representatives (including patient and public involvement and engagement)

XSL•FO
RenderX

NASSS domain	Complexity level	Rationale	Proposed action plans	Actors
Organizations	Complex	In current health care organizations, systems are complex, with a varying appetite for innovation. This variance depends on many factors, including willingness for innovation, formal links across multiple organizations, shared vision, and budget implications. There is a need for well-led resources to establish successful adoption and readiness.	• Education and training • Linking between organizations • Shared systems and interoperability • Shared vision and pathways to avoid duplication of care • Accessible funding and shared resources for the workforce and infrastructure • 24-hour response pathways	Health system, organization board members, acute or community trusts, and integrated care systems
Wider system	Complex	Financial and regulatory requirements are not yet understood.	• Governance including information governance, data security, cyber security, and ethical considerations • Connectivity for the population	National representative bodies, HaH society, NHS England, and Department of Health and Social Care
Embedding and adaptation over time	Complicated	A positive outlook is present, with agreement that DTs align with national policy changes. Increased use of digital technology and adaptability will be seen while familiarity and an emerging evidence base develop. There will need to be robust resilience to deal with unforeseen circumstances and adaptations over time, as well as behavioral change.	• Improved evidence base including proof of concept and evaluations • Data management • Internet connectivity • Digital interoperability • Raise shared awareness across an integrated care system • Aligning with national and local strategies • Collective action and reflective action	Clinicians, practitioners, integrated care systems, service users, and policymakers

[a]NASSS: Non-adoption, Abandonment, Scale-up, Spread, and Sustainability.

[b]HaH: hospital-at-home.

[c]DT: digital twin.

Discussion

This study aimed to explore health and care professionals' perceptions of a DT-enabled HaH model of care and its perceived impact. Although this technology has not yet been demonstrated in practice, data from participants were used to explore a conceptual model, providing an in-depth understanding of acceptability, benefits, and challenges across multiple levels: individual (patients and staff adopters), meso (organizational), and macro (systemic). The use of the NASSS framework, which is based on a synthesis of multiple theories of technology implementation [21], enabled a rich understanding of multiple sociotechnical factors influencing the potential adoption of DT technology and its integration within existing HaH models of care. The findings showed that there are many uncertainties and multilevel factors that can potentially impact adoption. Some include the current conceptual nature of HaH DTs and the unknowns around interoperability and integration. Furthermore, the potential change to staff routines and the need to learn new skills may involve new structures within organizations to embed DTs, leading to initial apprehension among staff. However, in general, professionals' views of the concept were positive, and they reported that embedding DTs in practice over time was manageable.

This study highlights key themes to guide recommendations for stakeholder consideration, supporting the implementation of HaH DTs and long-term sustainability. First, the results show that improvements are required from a system-wide perspective to enable successful adoption, scale-up, and spread. These may include digital infrastructure, connectivity, and interoperability with a shared vision centered around the patient. These can enable true person-centered care and holistic insights, and facilitate the predictive capabilities of DTs, if accurate data are inputted. This could facilitate raising awareness across multiple organizations, which can support long-term sustainability. Additionally, supportive infrastructure includes human resources and workforce capacity, which may require remodeling and new ways of working to deal with large amounts of data. There is potential for hybrid roles, working across boundaries, and a level of flexibility. While there is a scarcity of studies around

XSL•FO
RenderX

the adoption of DT technology in health care, particularly in a patient's home, this study advances new knowledge that contributes to and builds upon what is already known about the adoption of existing technologies such as telehealth and remote monitoring. However, this study was conducted in a systematic manner using a robust implementation framework and differs conceptually by considering the potential leverage that DTs offer, such as real-time simulation and predictive analytics. Findings from studies examining telehealth adoption have noted that system-wide improvements and service enhancement or redesign are needed to achieve sustainability [31]. A study by Thomas et al [32], which explored telehealth adoption in Australia using the NASSS framework, similarly found that implementation strategies are needed at multiple levels to support this sociotechnical complexity. The authors noted the importance of raising awareness, as well as clinicians' desires for flexible hybrid models, dedicated spaces, and structured staff training. The findings broadly correlate with the work of Sugarhood et al [33], who described that innovative technology cannot be seen as "plug and play" but as a complex innovation requiring strengthened linkage between different parts of the system, with input and coordination between people and organizations. This may require extensive changes to systems, structures, and ways of working, focusing on optimizing human-system integration. The authors describe the technology itself as a "hard core" (the unchangeable features of the innovation itself) that requires a "soft periphery" (systems and services essential to full implementation) [33]. Comparably, this "soft periphery" requires "organizational innovativeness" and receptiveness among organizations and leaders, which can vary [33]. Similar to our study, other research has described that technological innovations face challenges such as competing priorities and organizational financial pressures [33].

Second, adapting working models and using advanced technology require robust and ongoing training and support for reassurance, sense-making, and enhancing human-system interaction. Technical support and in-house DT champions were highlighted by participants as potential key facilitators to improve adoptability and provide supportive reassurance for any initial reluctance in the workforce. This training needs to be incorporated into clinicians' undergraduate and heuristic training to ensure that DTs are used to complement and enhance care and do not replace clinical judgment or lead to overreliance. This aligns with the findings of Sugarhood et al [33], who conducted a key informant interview study exploring telecare technologies as a system innovation. Their study identified that technological innovations require people to think and behave differently, recognizing the importance of staff training and workforce development in adoption. However, while initial successes were noted with designated digital "champions," these trained staff were found to often move on to new roles, impacting sustainability. Cartledge et al [31] highlighted that continuous support and improvement were key themes impacting adoption, particularly focusing on staff adopters and organizational domains.

Third, this study showed that adoption and implementation require reassuring evidence, suggesting that initial proof-of-concept studies can help identify any initial technical challenges before wider implementation. Participants felt that this would also help clarify which patients would be suitable for DTs, shaping any selection criteria. The recommendations for initial implementation suggest that stakeholder engagement is important throughout, with the involvement of individuals from various levels to cover all domains of the NASSS framework. Some of these may include patients, members of the public, clinicians, allied health practitioners, service leads, commissioners, technology developers and suppliers, digital leads, organization board members, and national policymakers. This is consistent with broader literature on technological adoption, which suggests that trialability (opportunity to experiment before committing to adopt) and observability (visible benefits) are both facilitators for adoption [33]. In the literature, examples for telecare include having a "show flat" or demonstration area attached to a day center for older people [33], and this approach could be considered for DTs in the community [10]. From a different angle, other literature has mentioned that new technology should not just focus on proof-of-concept but rather use co-production from the early stages of design and development to identify how technologies might better fit in with users' material surroundings, networks, and lifestyles [33]. This is critical as the needs of end users may differ from the needs and capacities of providers [33].

The findings from this study contribute to the growing international literature on DTs, specifically addressing the scarce research on HaH. National drivers advocate for increased use of technology, care closer to home, and strategies for prevention [11]. However, remote monitoring and the use of technology in HaH remain variable and uncertain [7]. Moreover, it has been argued that increased use of technology can support increased use of HaH [7]. Maximizing the potential of technology requires recognizing that technology can automate time-consuming tasks, such as data entry, and safely manage larger patient groups with the use of remote monitoring [7,34]. In their systematic review, Shi et al [35] recognized that current pressures on urgent and emergency services are encouraging innovative care models that use technology, representing a paradigm shift in acute health care delivery [35]. However, available evidence on high-intensity technologies, such as wearables for remote monitoring, remains scarce [8,35].

The findings of this study are consistent with international literature by Serrano et al [36], who similarly reported that most practitioners support the use of remote monitoring technology and perceived that this will become increasingly relevant in the future and advantageous for clinical practice. Serrano et al [36] described common benefits, including continuous monitoring of patients to provide prompt care, efficient communication, and visualization of health trends within a patient's own environment, allowing personalized care. Furthermore, they reported that the way data are presented via wearables or devices can influence patient understanding and self-care decisions [36]. Some challenges, such as increased workload, data inaccuracy, and higher patient anxiety, were also noted [36]. However, other studies, such as one exploring workforce perceptions on the use of a pregnancy remote monitoring intervention by Lanssens et al [37], found no actual increase in workload when using the technology. Additionally, the systematic review by Serrano et

XSL•FO
RenderX

al [36] identified practitioners' concerns regarding privacy, which was not a key focus found in our data. While these studies are not entirely comparable to studies of HaH DTs, they explored key components of DTs, such as sensing technology (remote monitoring) for data collection, communication in some aspects (telecare), and visualization using dashboards, which are also used in clinical practice. Nonetheless, there remains a lack of literature specifically addressing HaH DTs, particularly regarding predictive analytics and simulation capabilities.

The application of this multilevel integrated framework highlights some of the contextual challenges within complex health care systems, showing that findings do not always fit separately within individual domains. For example, the theme around trust impacts staff and patient adopters as well as the technology domain and value proposition, and it is recognized that trust strengthens with a growing evidence base and time. This also applies to training, which overlaps across multiple domains and has financial impacts that affect the value proposition, organization, wider system, and embedding over time. These domains should not be taken in isolation, but they are useful as a guide to uncover some of the complexities that need to be considered before implementation, ultimately informing action planning and targeted efforts. This builds on previous literature, which highlights that health care systems can be complex, such as the systematic review by Greenhalgh et al [38] that explored the diffusion of innovations in health service organizations. This reinforces that multiple components can impact the diffusion, dissemination, implementation, and sustainability of innovations [38]. This needs to be recognized as an ongoing process [39]. Moreover, different clinical specialities may have varying priorities in a system, and these need to be further understood. This can be embedded into the design and development process, as has been demonstrated by Waddell et al [40], who integrated the NASSS framework into a well-known design process, the double diamond, for their human-centered intervention. This recognizes that stakeholder involvement and an interdisciplinary approach are important throughout the process. In practice, organizations and teams applying DTs in HaH services may wish to use the NASSS-Complexity Assessment Toolkit (NASSS-CAT) [41] to understand their own context-specific complexity across each of the 7 NASSS domains. This also supports stakeholder engagement involving questioning to understand suitability, feasibility, the capacity of organizations, and the readiness of adopters, including patients. Furthermore, the toolkit can be used as an ongoing tool at different points in time.

The use of a diverse sample with a range of participants supported the aims of this study. Participants had varying experiences, both within and outside of HaH and across multiple settings. The included participants had experience as digital leads and information officers, with some having responsibilities for commissioning decisions and others having no prior experience of high-level technology use and having only clinical experience. However, this study may have attracted interest from keen participants who may be viewed as potential early adopters of technology or those who have negative experiences of technology, introducing potential sampling bias. Furthermore, the inclusion of only participants based in the United Kingdom may limit the generalizability of the findings to international contexts. Ultimately, participants' perceptions were based on theorization of a DT concept rather than first-hand experience of this technology, meaning that further work tailored to specific contexts would be required.

In conclusion, the findings of this qualitative study show that the adoption of innovations, such as HaH DTs, is shaped by multiple factors across several levels. While there is a perception of great value to patient and carer adopters, based on perceived reassurance and reduction in the burden with DTs compared to other technologies, there is a need for patient and public involvement or co-production to strengthen the findings. The use of a robust framework in this study allowed us to uncover some of the challenges with physical, cognitive, and organizational ergonomics that may be present with a HaH-DT concept and assess how stakeholders can build on this moving forward. The findings of this study highlight that system-level improvements and proof-of-concept evaluations support adoption, and digital champions support ongoing sustainability, awareness, and training. This work provides an early foundation for understanding likely adoption, and further research and implementation efforts should prioritize stakeholder involvement to develop clear, actionable plans.

Acknowledgments

The authors declare that no generative AI tool was used in any portion of manuscript generation.

Funding

This study/research was funded by the National Institute for Health and Care Research (NIHR) Newcastle Patient Safety Research Collaboration (PSRC). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Data Availability

The datasets generated and/or analyzed during the study are not publicly available to protect the privacy of the participants but are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: FY, MC, HN

XSL•FO
RenderX

Data curation: FY
Formal analysis: FY, MC, LL, HN
Methodology: FY, MC, HN
Writing – original draft: FY
Writing – review & editing: FY, MC, LL, HN

Conflicts of Interest

None declared.

Multimedia Appendix 1
Comparison between the theoretical framework of acceptability (TFA) and the Non-adoption, Abandonment, Scale-up, Spread, and Sustainability (NASSS) framework.
[DOCX File, 15 KB - humanfactors_v13i1e96645_app1.docx]

Multimedia Appendix 2
Interview vignette (slides) and topic guide.
[DOCX File, 50822 KB - humanfactors_v13i1e96645_app2.docx]

Multimedia Appendix 3
Coding extracts: inductive and deductive (steps 2 and 3).
[DOCX File, 446 KB - humanfactors_v13i1e96645_app3.docx]

Multimedia Appendix 4
Table mapping themes to the Non-adoption, Abandonment, Scale-up, Spread, and Sustainability (NASSS) framework.
[XLSX File, 20 KB - humanfactors_v13i1e96645_app4.xlsx]

Multimedia Appendix 5
Theme refinement according to the Non-adoption, Abandonment, Scale-up, Spread, and Sustainability (NASSS) framework.
[XLSX File, 16 KB - humanfactors_v13i1e96645_app5.xlsx]

Multimedia Appendix 6
Themes with descriptions and complexity.
[DOCX File, 35 KB - humanfactors_v13i1e96645_app6.docx]

Checklist 1
COREQ checklist.
[PDF File, 496 KB - humanfactors_v13i1e96645_app7.pdf]

References

1. Virtual wards operational framework. NHS England. 2024. URL: https://www.england.nhs.uk/long-read/virtual-wards-operational-framework/ [accessed 2025-04-02]
2. Shepperd S, Butler C, Cradduck-Bamford A, et al. Is comprehensive geriatric assessment admission avoidance hospital at home an alternative to hospital admission for older persons?: a randomized trial. Ann Intern Med 2021 Jul;174(7):889-898. [doi: 10.7326/M20-5688] [Medline: 33872045]
3. Wang X, Stewart C, Lee G. Patients' and caregivers' perceptions of the quality of hospital-at-home service: a scoping review. J Clin Nurs 2024 Mar;33(3):817-838. [doi: 10.1111/jocn.16906] [Medline: 37817557]
4. Hospital at home for frailty: current situation and future potential. British Geriatric Society. 2025. URL: www.bgs.org.uk/HospitalAtHomeFrailty [accessed 2026-07-21]
5. Virtual inequality? British Red Cross. 2024. URL: https://assets.redcross.org.uk/82b1e254-5524-0172-0612-9ce813c7824c/669a9d2d-5315-4439-b0ff-ad2b53332d24/Virtual%20inequality%20v05.pdf [accessed 2026-07-21]
6. Malhis RA, Bond SE, Sadeq AA, et al. Virtual healthcare compared to hospital care for acute and post-acute illness in adults: a systematic review and meta-analysis of randomized controlled trials. Br J Clin Pharmacol 2026 Feb;92(2):374-395. [doi: 10.1002/bcp.70348] [Medline: 41316908]
7. Cathers S, Lally C. Virtual wards and hospital at home. UK Parliament. 2025. URL: https://post.parliament.uk/research-briefings/post-pn-0744/ [accessed 2025-05-19]

XSL•FO
RenderX

8. Cheng SA, Tan SI, Goh SLE, Ko SQ. The value of remote vital signs monitoring in detecting clinical deterioration in patients in hospital at home programs or postacute medical patients in the community: systematic review. J Med Internet Res 2025 May 26;27:e64753. [doi: 10.2196/64753] [Medline: 40418800]
9. 2025/26 priorities and operational planning guidance. NHS England. 2025. URL: https://www.england.nhs.uk/long-read/2025-26-priorities-and-operational-planning-guidance/ [accessed 2025-04-03]
10. Yahya F, Cooper M, Saif W, Kassem M, Nazar H. Development of a hospital-at-home digital twin for patients with frailty: scoping review. J Med Internet Res 2025 Dec 10;27:e81510. [doi: 10.2196/81510] [Medline: 41370790]
11. 10 year health plan for England: fit for the future. GOV.UK. 2025. URL: https://www.gov.uk/government/publications/10-year-health-plan-for-england-fit-for-the-future [accessed 2026-08-05]
12. Dhanaraj RK, Murugesan S, Balusamy B, Balas VE. Digital Twin Technologies for Healthcare 4.0: Institution of Engineering & Technology; 2022.
13. Yahya F, Cooper M, Kassem M, Nazar H. Hospital at home digital twin for the management of patients with frailty: a scoping review protocol. BMJ Open 2025 Jun 17;15(6):e093418. [doi: 10.1136/bmjopen-2024-093418] [Medline: 40527560]
14. Al-Ali AR, Gupta R, Zaman Batool T, Landolsi T, Aloul F, Al Nabulsi A. Digital twin conceptual model within the context of Internet of Things. Future Internet 2020 Oct 1;12(10):163. [doi: 10.3390/fi12100163]
15. Farias FD, Dagostini CM, Bicca YDA, Falavigna VF, Falavigna A. Remote patient monitoring: a systematic review. Telemed J E Health 2020 May;26(5):576-583. [doi: 10.1089/tmj.2019.0066] [Medline: 31314689]
16. Patel R, Thornton-Swan TD, Armitage LC, et al. Remote vital sign monitoring in admission avoidance hospital at home: a systematic review. J Am Med Dir Assoc 2024 Aug;25(8):105080. [doi: 10.1016/j.jamda.2024.105080] [Medline: 38908399]
17. Bradley J, Schelbert EB, Bonnett LJ, et al. Predicting hospitalisation for heart failure and death in patients with, or at risk of, heart failure before first hospitalisation: a retrospective model development and external validation study. Lancet Digit Health 2022 Jun;4(6):e445-e454. [doi: 10.1016/S2589-7500(22)00045-0] [Medline: 35562273]
18. Lewis C. Predictors of adverse outcomes and factors associated with a risk-stratification based community virtual ward model for older persons with complex health and social care needs [Thesis]. : Royal College of Surgeons in Ireland; 2020 URL: https://repository.rcsi.com/articles/thesis/Predictors_of_adverse_outcomes_and_factors_associated_with_a_risk-stratification_based_community_virtual_ward_model_for_older_persons_with_complex_health_and_social_care_needs/13234541?file=25490546 [accessed 2026-11-11] [doi: 10.25419/rcsi.13234541]
19. Mebrahtu TF, McInerney CD, Benn J, et al. Effect of a hospital command centre on patient safety: an interrupted time series study. BMJ Health Care Inform 2023 Jan;30(1):e100653. [doi: 10.1136/bmjhci-2022-100653] [Medline: 36697032]
20. Ringeval M, Etindele Sosso FA, Cousineau M, Paré G. Advancing health care with digital twins: meta-review of applications and implementation challenges. J Med Internet Res 2025 Feb 19;27(1):e69544. [doi: 10.2196/69544] [Medline: 39969978]
21. Greenhalgh T, Abimbola S. The NASSS framework - a synthesis of multiple theories of technology implementation. Stud Health Technol Inform 2019 Jul 30;263:193-204. [doi: 10.3233/SHTI190123] [Medline: 31411163]
22. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
23. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88. [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
24. Shin HD, Hamovitch E, Gatov E, et al. The NASSS (Non-Adoption, Abandonment, Scale-Up, Spread and Sustainability) framework use over time: a scoping review. PLOS Digit Health 2025 Mar;4(3):e0000418. [doi: 10.1371/journal.pdig.0000418] [Medline: 40096260]
25. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367. [doi: 10.2196/jmir.8775] [Medline: 29092808]
26. Braun V, Clarke V. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. Qualitative Research in Sport, Exercise and Health 2021 Mar 4;13(2):201-216. [doi: 10.1080/2159676X.2019.1704846]
27. Braun V, Clarke V. Thematic Analysis: A Practical Guide: SAGE; 2021.
28. Doing reflexive TA. Thematic Analysis. URL: https://www.thematicanalysis.net/doing-reflexive-ta/ [accessed 2026-02-13]
29. Payment guidance for researchers and professionals involving people in research. National Institute for Health and Care Research. 2024. URL: https://www.nihr.ac.uk/about-us/who-we-are/policies-and-guidelines/payment-guidance-researchers-and-professionals [accessed 2026-03-22]
30. Find your local integrated care board (ICB). NHS. URL: https://www.nhs.uk/nhs-services/find-your-local-integrated-care-board/ [accessed 2026-02-13]
31. Cartledge S, Rawstorn JC, Tran M, Ryan P, Howden EJ, Jackson A. Telehealth is here to stay but not without challenges: a consultation of cardiac rehabilitation clinicians during COVID-19 in Victoria, Australia. Eur J Cardiovasc Nurs 2022 Aug 29;21(6):548-558. [doi: 10.1093/eurjcn/zvab118] [Medline: 34935940]

XSL•FO
RenderX

32. Thomas EE, Taylor ML, Ward EC, et al. Beyond forced telehealth adoption: a framework to sustain telehealth among allied health services. J Telemed Telecare 2024 Apr;30(3):559-569. [doi: 10.1177/1357633X221074499] [Medline: 35130099]
33. Sugarhood P, Wherton J, Procter R, Hinder S, Greenhalgh T. Technology as system innovation: a key informant interview study of the application of the diffusion of innovation model to telecare. Disabil Rehabil Assist Technol 2014 Jan;9(1):79-87. [doi: 10.3109/17483107.2013.823573] [Medline: 23919407]
34. Donnelly T. AI, virtual wards and digital scribes: embracing digital technology in health. The Health Foundation. 2023. URL: https://www.health.org.uk/features-and-opinion/features/ai-virtual-wards-and-digital-scribes-embracing-digital-technology-in [accessed 2026-02-26]
35. Shi C, Dumville J, Rubinstein F, et al. Inpatient-level care at home delivered by virtual wards and hospital at home: a systematic review and meta-analysis of complex interventions and their components. BMC Med 2024 Apr 2;22(1):145. [doi: 10.1186/s12916-024-03312-3] [Medline: 38561754]
36. Serrano LP, Maita KC, Avila FR, et al. Benefits and challenges of remote patient monitoring as perceived by health care practitioners: a systematic review. Perm J 2023 Dec 15;27(4):100-111. [doi: 10.7812/TPP/23.022] [Medline: 37735970]
37. Lanssens D, Vandenberk T, Lodewijckx J, et al. Midwives', obstetricians', and recently delivered mothers' perceptions of remote monitoring for prenatal care: retrospective survey. J Med Internet Res 2019 Apr 15;21(4):e10887. [doi: 10.2196/10887] [Medline: 30985286]
38. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629. [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
39. Robert G, Greenhalgh T, MacFarlane F, Peacock R. Adopting and assimilating new non-pharmaceutical technologies into health care: a systematic review. J Health Serv Res Policy 2010 Oct;15(4):243-250. [doi: 10.1258/jhsrp.2010.009137] [Medline: 20592046]
40. Waddell A, Seguin JP, Wu L, et al. Leveraging implementation science in human-centred design for digital health. Presented at: 2024 CHI Conference on Human Factors in Computing Systems; May 11-16, 2024. [doi: 10.1145/3613904.3642161]
41. eNASSS-CAT. Nuffield Department of Primary Care Health Sciences, University of Oxford. URL: https://www.phc.ox.ac.uk/research/groups-and-centres/interdisciplinary-research-in-health-sciences/enasss-cat [accessed 2026-03-02]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
DT: digital twin
HaH: hospital-at-home
ICB: integrated care board
NASSS: Non-adoption, Abandonment, Scale-up, Spread, and Sustainability
NHS: National Health Service
NIHR: National Institute for Health and Care Research
RPM: remote patient monitoring
TFA: theoretical framework of acceptability

Edited by S Law; submitted 30.Mar.2026; peer-reviewed by E Poon, GI Han, M Galani, MF Islam; revised version received 17.Sep.2026; accepted 17.Sep.2026; published 05.Nov.2026.

Please cite as:
Yahya F, Cooper M, Lawson L, Nazar H
Workforce Perceptions of a Conceptual Hospital-at-Home Digital Twin: Qualitative Study
JMIR Hum Factors 2026;13:e96645
URL: https://humanfactors.jmir.org/2026/1/e96645
doi:10.2196/96645

© Faiza Yahya, Matthew Cooper, Lauren Lawson, Hamde Nazar. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 5.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Feasibility and Acceptability of the SISKA System for Continuous Student Emotional and Mental Health Guidance: Mixed Methods Study

Tantut Susanto[1], PHN, PhD; Heni Dwi Windarwati[2], MHN, PhD; Niken Asih Laras Ati[3], MN; Rizkiyani Istifada[1], MN, PHN; Liasari Armaijn[4], MD, MSc; Andi Eka Pranata[5], RN, MSc; Hanik Endang Nihayati[6], PhD; Meidiana Dwidiyanti[7], PhD; Thika Marliana[8], MHN, PhD; Ermanto Fahamsyah[9], PhD; Slamin[10], PhD; Yunita Arihandayani[11], MD, MPH

[1]Department of Community, Family & Geriatric Nursing, Faculty of Nursing, Universitas Jember, Jalan Kalimantan 37, Jember, East Java, Indonesia

[10]Faculty of Computer Sciences, Universitas Jember, Jember, East Java, Indonesia

[11]Directorate of Health Services for Vulnerable Groups, Ministry of Health of Indonesia, Jakarta, Indonesia

[2]Department of Mental Health Nursing, Faculty of Health Sciences, University of Brawijaya, Puncak Dieng Eksklusif, Kunci, Kalisongo, Malang, Jawa Timur, Indonesia

[3]Department of Mental Health Nursing, Faculty of Nursing, Universitas Jember, Jember, East Java, Indonesia

[4]Faculty of Medicine, Khairun University, Ternate, North Maluku, Indonesia

[5]Postgraduate Study Program, Universitas Jember, Jember, East Java, Indonesia

[6]Faculty of Nursing, Airlangga University, Surabaya, East Java, Indonesia

[7]Department of Nursing, Faculty of Medicine, Diponegoro University, Semarang, Central Java, Indonesia

[8]Faculty of Health Sciences, Universitas Respati Indonesia, Depok, Jakarta, Indonesia

[9]Faculty of Law, Universitas Jember, Jember, East Java, Indonesia

Corresponding Author:
Heni Dwi Windarwati, MHN, PhD
Department of Mental Health Nursing, Faculty of Health Sciences, University of Brawijaya, Puncak Dieng Eksklusif, Kunci, Kalisongo, Malang, Jawa Timur, Indonesia

Abstract

Background: University students face substantial underserved mental health needs. Digital “front doors” that combine brief screening, decision support, and timely counselor follow-up may help close access and workflow gaps on campus. SISKA (Sistem Informasi Layanan Konseling Akademik*;* “academic counseling information system”) was developed as an Android app with a counselor-facing web dashboard to support routine self-assessment, risk-tiering, just-in-time guidance, referral activation, and continuous follow-up within real university workflows.

Objective: The study aims to evaluate the feasibility, acceptability, and usability of SISKA in Indonesian public universities and to identify strengths and friction points to inform implementation at scale.

Methods: This exploratory sequential mixed-methods study recruited trained lecturer-counselors (n=20) who provided qualitative input on practicality and integration and 3 student cohorts to assess (1) usability (n=35 students who completed all 10 System Usability Scale [SUS] questions), (2) use analytics (n=1057 active accounts from December 2024 to October 2025), and (3) changes in mental health before and after the intervention (n=119 students with ≥1 counseling session and paired daily mental health assessment scores). This was a single-arm feasibility study without a comparator group. Analyses included descriptive summaries and paired tests; qualitative data were analyzed thematically to explain quantitative patterns.

Results: Monthly uptake showed a launch spike, a mid-year trough, and a late-year uptick. Usability metrics showed a short learning curve and manageable cognitive load (SUS mean score 83.3, SD 10.5) with strong item agreement on ease, quick learning, and integration. In the pre- and postintervention cohort, daily mental health assessment scores declined from a mean 33.9 (SD 8.07) to a mean 32.8 (SD 8.94; Δ=−1.14; P=.03), indicating a statistically significant but small preliminary reduction in distress following counseling. Qualitatively, students described feelings of psychological safety, that the psychoeducation was helpful, and that the scheduling was oriented to the services; friction points included interface inconsistencies, multistep confirmation, and intermittent connectivity or notification issues.

Conclusions: SISKA appears feasible and acceptable for campus deployment, enabling earlier support pathways and showing proximal improvement among counseled students. Priorities for scaling up the app include user interface standardization, streamlined confirmation flows, and reliable, context-sensitive notifications, alongside embedding measurement-based care loops

XSL•FO
RenderX

to convert routine self-reports into timely feedback and action. These findings support the readiness of the app for structured implementation in university settings.

(JMIR Hum Factors 2026;13:e87647) doi:10.2196/87647

KEYWORDS

students; mental health; mobile apps; counseling; feasibility studies

Introduction

University students carry a persistent burden of common mental disorders, with approximately 1 in 3 screening positive within a year, reflecting the onset of disorders in adolescence that persist into the first college year. Evidence from the World Mental Health Surveys International College Student (WMH-ICS) program across 19 colleges in 8 countries indicates that over their lifetime, 35.3% of students, and the past year, 31.4%, screen positive for *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* mood, anxiety, or substance disorders [1]. Recent WMH-ICS findings also reveal that over half of first-year students screened positive for at least one mental disorder, underscoring how widespread and demographically diffuse these problems have become [2]. Against this backdrop, stigma, low mental-health literacy, and self-reliance norms further suppress help-seeking [3], allowing needs to go unmet and symptoms to escalate in the absence of early detection and timely care. Accordingly, campuses should adopt a whole-university proactive model that integrates light-touch screening and stepped care and uses a clear digital "front door" to normalize help-seeking, reduce barriers, and move students from informal supports to timely formal care; this should be accompanied by rigorous cost-effectiveness evaluations and implementation readiness assessments [4,5].

Research indicates that common mental health problems such as depression and anxiety harm students' academic performance by lowering engagement; they predict a lower grade point average and increase the likelihood of dropping out [6,7]. Evidence from routine psychological services shows that students are less likely to reliably recover and are at greater risk of poorer outcomes, suggesting that care pathways and intervention content should be adapted to address student-specific needs [8,9]. In parallel, university services should embed routine outcome monitoring with session-by-session feedback to deliver data-informed care and timely course corrections that help close workflow and feedback gaps [10]. On Indonesian campuses, needs assessments reveal strong demand for counseling and for early electronic counseling initiatives that provide appointments and access to psychologists as well as to peer counselors who can help students clarify difficulties and develop solution-focused responses [11]. Recently, a Kotlin-based mental health prototype app demonstrated technical promise, with about 85% validation using a bidirectional long short-term memory model, but the app also showed integration gaps, including an unintegrated machine-learning module, a nonfunctional map API, and an empty history view because it did not save notes and analyses [12]. An Indonesian electronic group counseling platform has also been pilot-tested with counselors and hundreds of students, offering scheduling, assessments, virtual sessions, and auto-generated reports, which suggests a feasible digital front door for campus care [13]. These findings support a shift toward student-tailored pathways with integrated routine outcome monitoring and a mature digital front door, while providing initial confirmation of feasibility, acceptability, and usability in real campus workflows and benefits for students' psychological outcomes.

In response to the campus-specific gaps in access, fragmented information flow, and limited decision support described above, SISKA (Sistem Informasi Layanan Konseling Akademik; "academic counseling information system") was developed as a student-facing Android app paired with a counselor-facing web dashboard that operate as a single, integrated workflow. The deployment of SISKA at a university in East Java is a strategic response to the need for improving the quality of student counseling and advisory services through a digital, preventive, and integrative approach. SISKA is designed to support students' academic functioning and well-being, promote a mentally healthy and caring campus environment, and strengthen data-driven and technology-based student service quality. The app supports brief daily or weekly self-assessments, risk stratification into actionable tiers, just-in-time guidance, self-help and peer-support options, referral activation and scheduling, crisis pathways, configurable notifications, and psychoeducation; the dashboard aggregates risk flags, visualizes routine-outcome trends, and provides concise decision support so counselors can prioritize outreach, schedule sessions, and document care with a lower administrative load. These design choices are consistent with eHealth principles that emphasize continuous measurement and feedback in real-world settings where engagement typically declines without intentional experience design; the choices also align with evidence that an online routine outcome monitoring and feedback system is feasible and supports clinicians' session-by-session decision-making [14,15]. They also align with the behavioral intervention technology model, which maps clinical aims to intervention strategies and technical elements to guide the design and evaluation of eHealth and mobile health interventions [16].

Responding to the growing campus mental health burden and access barriers, this study introduces and evaluates SISKA as a scalable, Android-based platform for self-assessment, counseling guidance, and continuous follow-up in higher education settings. We assessed usability with the Indonesian-adapted System Usability Scale (SUS) and triangulated SUS findings with student and counselor feedback, drawing on measurement-based care to convert screening data into concise decision support with rapid feedback. Therefore, our objectives are to determine whether SISKA is feasible, practical, and acceptable; to determine whether it shows early

XSL•FO
RenderX

efficacy signals by meeting SUS usability benchmarks for both quantitative scores and qualitative feedback; to determine whether there are improvements in mental health outcomes from before to after counseling facilitated through SISKA; and to use qualitative insights to steer iterative refinement and increase implementation readiness.

Methods

Design

This study used a single-arm pre- and postintervention feasibility design without a comparator group. The study aimed to evaluate the feasibility, acceptability, usability, and initial efficacy of the SISKA Android app and web-based dashboard for student mental and emotional health services in a real-world university setting rather than to determine intervention effectiveness. Therefore, the study was not structured to control for external influences. These factors should be examined in future controlled trials. An exploratory sequential mixed-methods approach was used with an initial quantitative phase (examining system usability, self-assessment scores, and use analytics), which informed a subsequent qualitative phase (comprising open-ended surveys of students and counselors) to explain quantitative patterns and contextual determinants of use.

Setting and Participants

This study was conducted in public universities in Indonesia and involved 2 participant groups: students as primary SISKA app users and counselors as dashboard users. Students accessed the Android app for periodic self-assessment, immediate feedback, psychoeducation, or in-app counseling requests. Counselors, who were lecturers trained in counseling, used the web-based dashboard for privacy-aware monitoring, case management, scheduling, documentation, and continuous follow-up. A total of 20 lecturer-counselors participated.

Counselors were included if they were lecturers formally employed in academic and mental health counseling, had completed standardized SISKA counselor training prior to deployment, had role-appropriate dashboard access, agreed to data-protection procedures, and provided informed consent. Counselors were excluded if they failed to complete training before data collection, took a leave of absence or had workload barriers preventing dashboard use and documentation, or refused or withdrew consent to participate.

A total of 26 counselors actively conducted counseling sessions during the study period, and 20 counselors who met the inclusion criteria voluntarily completed the survey. While additional counselors completed SISKA training, only those who conducted counseling sessions during the study period were eligible to participate. As students were free to select their preferred counselor, some were not selected for sufficient sessions and did not have enough experience using SISKA to provide relevant feedback.

Student eligibility for this study was as follows: age ≥18 years, enrollment in a public university, able to provide informed consent, regular access to an Android device with basic digital literacy, and sufficient fluency in Bahasa Indonesia to complete app items and surveys. The exclusion criteria were as follows: acute psychiatric crisis or imminent risk requiring specialized care and severe cognitive, sensory, or language limitations.

Because this study was designed as a feasibility evaluation of the SISKA digital counseling system, the analytical cohorts were recruited sequentially (Multimedia Appendix 1). First, the usability cohort included students who completed the SUS after ≥1 week of system exposure (n=35) to obtain an early assessment of the usability of the SISKA app. In feasibility and early-stage usability studies, relatively small samples are commonly used to identify usability issues and generate preliminary usability estimates [17,18]. Second, a use analytics cohort included all active SISKA users during the evaluation period (n=1057) to characterize system uptake and engagement under real-world implementation conditions. Students in this group without a recorded counseling session were excluded from the outcome analyses (n=938). Although many users received an automated referral to a counselor following the initial mental health screening (978/1057, 92.5%), referral did not necessarily indicate service use. Therefore, the counseling outcome analysis included only students who attended at least one SISKA counseling session and had paired daily mental health assessment records (n=119). A total of 755 students had only a single recorded daily mental health assessment, with a mean score of 30.76 (SD 8.57; the assessment score ranged from 10 to 50, with higher scores indicating greater psychological distress and lower scores indicating better mental health), and they were therefore not included in the paired pre- and postintervention analysis. In parallel, a counselor-focused practicality cohort was specified including all lecturer-counselors with valid dashboard access who completed the qualitative survey (n=20).

SISKA App and System Implementation

The SISKA app and web-based system was designed with 2 primary interfaces: student and counselor. The student interface (a mobile app) included several key features, such as a validated daily self-assessment based on a standardized, culturally adapted scale for the Indonesian context, real-time feedback to support counseling, initial self-help resources (articles and posters), and an option to request a counseling session. The counselor interface (a web- and app-based dashboard) included a monitoring dashboard for aggregated and individual student self-assessment results (with appropriate privacy safeguards), a case management module, scheduling tools, and a system for documenting guidance sessions and continuous follow-up.

The SISKA implementation procedure included 3 phases. First, all participating students received brief orientation on downloading, installing, and using the self-assessment and initial resource features. All participating counselors received comprehensive training on the counselor dashboard, data interpretation, guidance logging, and follow-up procedures. Second, the students completed a baseline self-assessment within the SISKA app at the start of the study period (from December 1, 2024, to October 15, 2025). Third, the system was implemented for 8 weeks. During this period, students were encouraged to complete biweekly reassessments and use the available self-help resources.

XSL•FO
RenderX

Students who were identified as needing support based on their self-assessment results or who actively requested assistance through the app were offered at least one counseling session with a trained lecturer-counselor. Counseling sessions were conducted using the SISKA dashboard and followed a structured workflow that included case monitoring, scheduling, documentation, and follow-up. While the system provided standardized guidance for counselors, the timing and number of counseling sessions could vary depending on student needs and counselor availability, reflecting real-world campus service conditions. Counselors used the system to identify students who needed support, initiate counseling guidance, and log all ongoing follow-up activities.

Data Collection and Measurements

This study used the SUS to assess the perceived usability of the SISKA app. The SUS is a 10-item, technology-agnostic instrument widely used to quantify usability [19]. The Indonesian adaptation has demonstrated good internal consistency (Cronbach α=0.841) [20]. To obtain a more detailed assessment of usability, the SUS was treated as comprising 2 interpretive subscales: usability (items 1, 2, 3, 5, 6, 7, 8, and 9) and learnability (items 4 and 10), following the 2-factor interpretation of the SUS described by Lewis and Sauro [21]. The usability and learnability components contain different numbers of items and therefore have different score ranges. Consequently, both component scores were rescaled to a 0 to 100 metric by dividing each adjusted score by its maximum possible score and multiplying by 100. Respondents rated each statement on a 5-point Likert scale (1=strongly disagree to 5=strongly agree). Item responses were scored using the standard SUS procedure, where adjusted item contributions are summed and multiplied by 2.5 to produce a total score ranging from 0 to 100, with higher scores indicating better usability. A SUS score ≥68 indicates acceptable or good usability [19,20]. The total SUS score was also interpreted using the adjective-rating scale proposed by Bangor et al [22], which provides a more meaningful and robust basis for interpreting SUS scores through descriptive usability labels. In addition to the overall SUS score, descriptive summaries of the usability and learnability subscales were examined to provide additional insights for improving SISKA's interface design and workflow.

This study used an in-app daily mental health assessment as the continuous pre- to postintervention outcome measure. This assessment is a brief, routine self-report completed inside SISKA that aggregates core indicators of current mental and emotional status using Likert-type items summed to a total score, with higher scores indicating greater distress. The instrument was adapted from the Kessler Psychological Distress Scale (K10), a widely used screening tool for nonspecific psychological distress [23]. The K10 has demonstrated strong psychometric properties, including high internal consistency and good screening accuracy for depressive and anxiety disorders, with receiver operating characteristic values ranging from 0.78 to 0.86 [24]. Evidence from Indonesian samples also supports its reliability, with Cronbach α=0.859 [25]. The daily mental health assessment was used as a pragmatic in-app monitoring measure to capture short-term changes in psychological distress within the SISKA counseling workflow. For each student in the pre- and postintervention cohort, the preintervention score was the nearest valid daily mental health assessment recorded before the index counseling session, and the postintervention score was the nearest valid daily mental health assessment after that session. When multiple daily mental health assessments were available within a window, the observation closest in time to the index session was selected.

To promote response quality and reduce common-method bias, SUS items were presented in the validated order and with mixed polarity, accompanied by brief instructions that emphasized that there were no right or wrong answers and that responses should reflect the respondent's personal experience with the SISKA app. We administered a qualitative survey to complement the capture of contextual determinants of the SISKA app (Multimedia Appendix 2). The qualitative survey included open-ended questions designed to capture participants' experiences with the SISKA system. To encourage balanced responses and reduce potential compliance or social-desirability bias, the participants were asked to reflect on both the system's strengths and weaknesses. This research aimed to enhance the interpretation of SUS scores by connecting quantitative usability patterns with explanatory narratives about real-world use.

Both quantitative (feasibility; Table 1) and qualitative (acceptability, usability, and context; Table 2) data were collected.

Table . Quantitative data collection.

Metric category	Data source	Measurement	Indicator
Usability	Student survey	SUS[a]	Mean SUS score ≥68 (acceptable)
Practically	Counselor logs	Number of follow-ups initiated through the system, time saved/spent on administrative tasks (self-reported or system tracked)	≥X student cases managed per counselor; Y% reduction in paperwork time
Initial efficacy	Student self-assessment	Change in self-assessment score from daily self-assessment report	Statistically significant positive change (reduction in distress/improvement in well-being)

[a]SUS: System Usability Scale.

XSL•FO
RenderX

Table . Qualitative data collection.

Data type	Participants	Method	Focus areas
User experiences	Students	Qualitative survey	Perceived ease of use, interface design, relevance of feedback/resources, privacy concerns
System acceptability	Counselors	Qualitative survey	Utility for case identification, integration into existing workflow, perceived effectiveness for follow-up, suggestions for improvement

Quantitative and Qualitative Data Analysis

Quantitative data analysis was performed on descriptive and inferential statistics. Descriptive statistics, including frequency, percentage, mean, and SD, were used to describe demographic characteristics, system use rates, and SUS scores. Inferential statistics were determined using paired-sample *t* tests to analyze pre- and postintervention daily mental health assessment scores, and Cohen *d* with 95 % CIs was used to obtain a standardized estimate of the practical magnitude of change. Analyses were conducted using complete case data. For the usability analysis, only participants who completed all 10 items of the SUS were included. One SUS response was incomplete (1/36, 2.8%) and was excluded from the usability analysis, resulting in a final usability sample of 35 participants. Similarly, the counseling outcome analysis included only students who had daily mental health assessment records before and after counseling sessions (n=119). Entries with missing follow-up daily mental health assessment records were excluded from the paired analysis to ensure valid pre- and postintervention comparisons.

Qualitative survey transcripts were analyzed using a thematic analysis approach [26]. To ensure the rigor and trustworthiness of the qualitative analysis, coder triangulation was used, with 2 researchers independently reviewing and coding the interview responses. Discrepancies in coding were discussed among the research team to reach consensus. The analysis followed an iterative process in which codes were repeatedly reviewed and refined into themes. The analysis involved iterative coding, development of themes and subthemes, and organization of findings into categories to facilitate interpretation of participant experiences. Data collection and analysis continued until thematic saturation was reached. These procedures were implemented to enhance the credibility and transparency of the qualitative findings. Data coding was conducted to identify recurrent themes, patterns, and insights related to usability, practicality, acceptability, and contextual barriers to or facilitators of the SISKA system. This analysis also examined barriers to and enablers of continuous counseling guidance and follow-up.

Credibility was enhanced through coder triangulation and team discussions to ensure that interpretations accurately reflected participants' perspectives. Dependability was supported by documenting coding decisions and using a consistent analytical process across researchers. Confirmability was strengthened through collaborative review of codes, themes, and supporting quotations to ensure that findings were grounded in participant responses. Transferability was facilitated by providing detailed descriptions of the participants and implementation context. The research team comprised researchers from diverse disciplinary backgrounds, including nursing, mental health, and computer science. Throughout the analytic process, reflexive discussions were conducted to consider how the researchers' professional expertise, disciplinary perspectives, and prior assumptions might influence data interpretation. These discussions helped promote critical reflection and enhance the transparency of the analytic process.

Descriptive statistics were used to map the prevalence of emotional distress (eg, depression and anxiety). The qualitative phase involved a thematic analysis of textual data, including students' specific concerns recorded during the self-assessment and the subsequent personalized recommendations and guidance provided by academic counselors during the follow-up process. Data integration focused on how the qualitative insights informed the interpretation of the quantitative findings and shaped the follow-up process, thus assessing the system's efficacy in facilitating a stepped-care approach.

Integration of Data

Quantitative data gathered through surveys, including user satisfaction scores, frequency of system use, and initial ratings of mental health symptom severity, provided measurable insights into feasibility (eg, use rates), acceptability (eg, average satisfaction scores), and usability (eg, SUS scores). These numerical results established the magnitude and distribution of key outcomes. Qualitative data collected through interviews or focus groups with students and counselors regarding their experiences, challenges, and perceptions regarding the SISKA system provided rich, in-depth contextual insight. This narrative information explained why the system achieved certain satisfaction levels, how it was used in practice, and what specific features were considered helpful or hindered adoption, thereby illuminating the "how" and "why" behind the quantitative findings.

In the analysis, integration often occurred through triangulation or by explaining quantitative results with qualitative data. For instance, low satisfaction scores (quantitative) for a particular system feature could be explored and explained by direct quotes and themes from interviews (qualitative) detailing specific technical glitches or usability issues. Conversely, high reported use frequency (quantitative) might be validated and enriched by qualitative accounts of students finding the system genuinely convenient and beneficial for immediate stress management. This convergence of findings from different data types provided a more robust and complete answer to the research questions about feasibility, acceptability, and usability than either method

XSL•FO
RenderX

could achieve alone, offering actionable insights for system improvement.

Ethical Considerations

Several measures were implemented to ensure privacy and data security. Student responses were stored in a secure database within the SISKA system and were accessible only to authorized counselors and the student mental health service team. All data were handled in accordance with institutional data protection policies.

This study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. Ethical approval was obtained centrally from the Research Ethics Committee of Universitas Respati Indonesia (715/SK.KEPK/UNRI/X/2025) for the multicenter IN-SIGHT research program, in which SISKA was implemented as one of the digital applications across participating sites in East Java, Central Java, and Jakarta. The approval covered the implementation of the SISKA system and the analysis of anonymized student data for research purposes. All participants were informed about the study objectives, procedures, potential risks, and benefits prior to participation. The study adhered to the Indonesian Ministry of Health's national ethical guidelines for research involving human participants [27] .

Results

User Adoption

The observed pre- and postintervention changes are reported descriptively as preliminary findings within a feasibility framework. These changes should not be interpreted as evidence of causal effectiveness. At baseline, 1057 SISKA users were assessed between December 2024 and October 2025. Most students were from social and humanities programs (n=564, 53.4%). Initial daily mental health assessments indicated that 79 students (7.5%) reported no mental health problems, 554 (52.4%) were at risk of mental health problems, and 424 (40.1%) were at risk of mental health disorders. Accordingly, 978 users (92.5%) were referred to a counselor (Table 3). The characteristics of the 35 SUS respondents were compared with those of the 1022 SUS nonrespondents to examine whether the usability estimate reflected the broader SISKA user base. Overall, the groups were broadly similar in faculty distribution and referral status. However, among the 35 SUS respondents, there was a higher proportion classified as at risk of mental health disorders at baseline than nonrespondents (57.1% vs 39.5%), suggesting that the usability estimate may overrepresent students with higher baseline distress.

Table . Participant characteristics for initial daily mental health assessment of Sistem Informasi Layanan Konseling Akademik (SISKA)'s users.

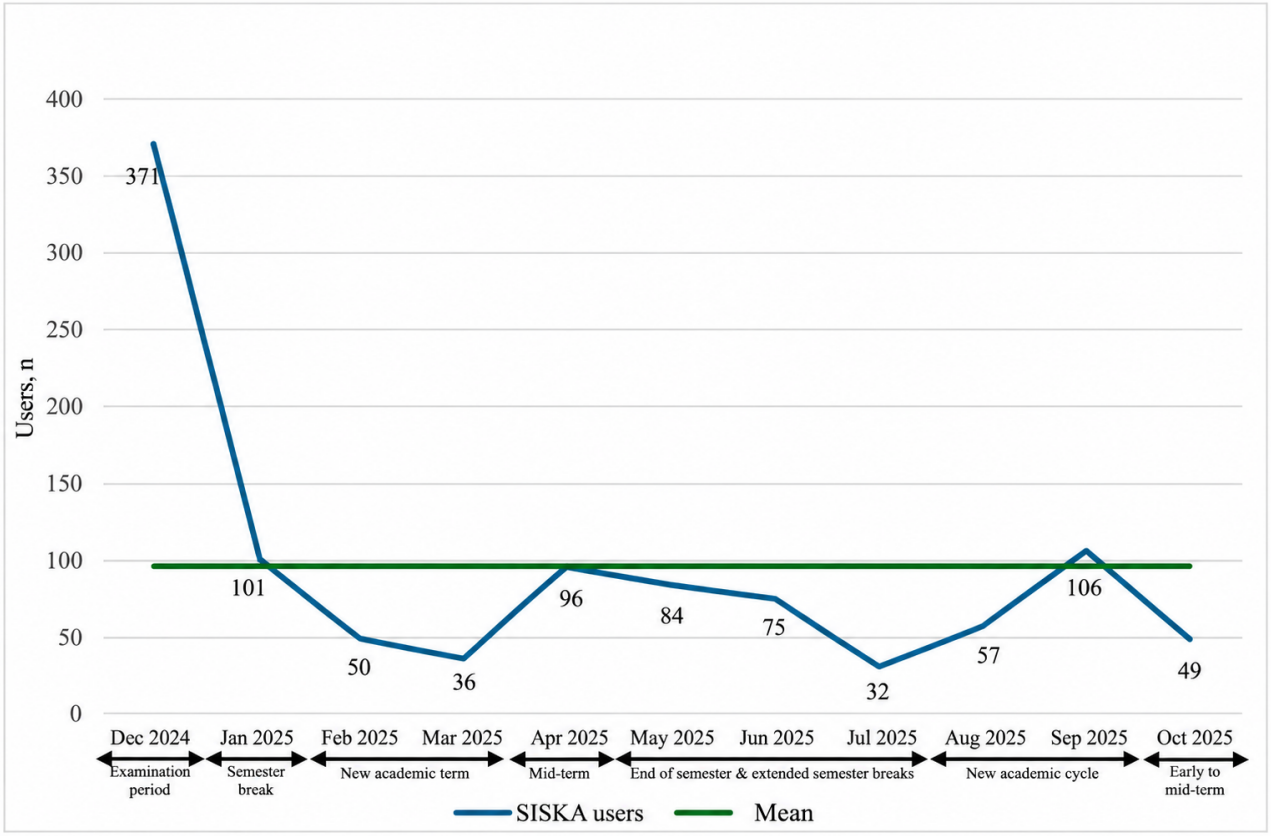
Characteristic	SISKA users (N=1057), n (%)	SUS[a] nonrespondents (n=1022), n (%)	SUS respondents (n=35), n (%)	Pre- and postintervention counseling subsample, n (%)	
				Precounseling (n=119)	Postcounseling (n=119)
Faculty, n (%)					
Social and humanities	564 (53.4)	546 (53.4)	18 (51.4)	68 (57.1)	68 (57.1)
Science, technology, engineering, and mathematics	317 (29.9)	312 (30.5)	12 (34.3)	27 (22.7)	27 (22.7)
Health sciences	176 (16.7)	164 (16.1)	5 (14.3)	24 (20.2)	24 (20.2)
Daily mental health assessments					
Did not experience mental health problems	79 (7.5)	77 (7.5)	2 (5.7)	4 (3.4)	8 (6.7)
At risk of mental health problems	554 (52.4)	541 (52.9)	13 (37.2)	52 (43.7)	61 (51.3)
At risk of mental health disorders	424 (40.1)	404 (39.5)	20 (57.1)	63 (52.9)	50 (42)
Referral to a counselor					
Yes	978 (92.5)	945 (92.5)	33 (94.3)	115 (96.6)	111 (93.3)
No	79 (7.5)	77 (7.5)	2 (5.7)	4 (3.4)	8 (6.7)

[a]SUS: System Usability Scale.

The final total was 1057 users. Monthly uptake of users spiked at launch in December 2024 (n=371, 35.1%), declined through March 2025 (n=36, 3.4%), returned to the mean in April 2025 (n=96, 9.1%), reached a mid-year trough in July 2025 (n=32, 3.0%), and then rebounded in August 2025 (n=57, 5.4%), with a secondary peak in September 2025 (n=106, 10.0%) before easing in October 2025 (n=49, 4.6%) (Figure 1). This pattern represents a typical postlaunch decline, mid-year trough, and late-year uptick.

XSL•FO
RenderX

Figure 1. Uptake of Sistem Informasi Layanan Konseling Akademik (SISKA) users from 2024 to 2025, with a final total of 1057 users.



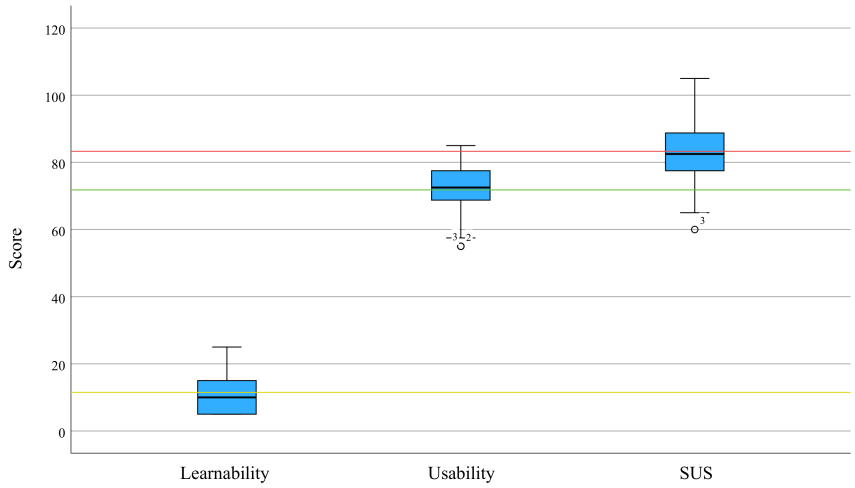
The initial spike in December 2024 coincided with the examination period, and the subsequent decline in January 2025 aligned with the semester break, followed by relatively lower engagement in February and March 2025, which may have reflected the transition into a new academic term, including orientation activities. There was a moderate increase in April 2025 during another examination period (midterm examinations), and a gradual decline from May to July 2025 overlapped with the end-of-semester phase and extended semester breaks in June and July 2025. The rise in August 2025 and the peak in September 2025 may be associated with the start of a new academic cycle and examination period.

In Figure 2, the SUS distribution indicates a tight central tendency with moderate spread: the boxplot shows a median score of approximately 83 (95% bootstrap CI 80.0-87.5) with a mean of 83.29 (SD 10.5), indicating that most observations clustered in the high usability range. The histogram approximated a near-normal distribution, with the majority of scores concentrated between 75 and 90, and the mean and median positions were closely aligned, suggesting relatively symmetrical data with only minor skewness. The usability component showed a mean score of 71.8 (SD 8.4), equivalent to 89.8 on a 1 to 100 scale, while the learnability component had a mean score of 11.5 (SD 5.8), equivalent to 57.5 on a 1 to 100 scale. The histograms for both components also had approximately normal shapes. In Figure 3, item-level patterns show fast pick-up and routine use among the SUS respondents: “learn quickly” had 40% “strongly agree” responses, and 31.4% “agree” responses (71.4% for both; item mean 4.0), while “ease” had 57.1% “strongly agree” responses and 25.7% “agree” responses (82.8% for both; mean 4.4). These results align with the near-normal usability distribution and overlapping mean and median lines, indicating a short learning curve and rapid routinization. Overall, the SUS mean score of 83.3 (SD 10.5) on the 0 to 100 scale exceeds the commonly accepted usability benchmark of 68, indicating excellent perceived usability and qualitatively characterizing the system as easy to learn, well integrated, and confidence-building, with minimal cumbersomeness and low support needs. In addition, based on the Bangor adjective rating scale, the SISKA SUS score of 83.3 corresponds approximately to the “excellent” usability range [22].

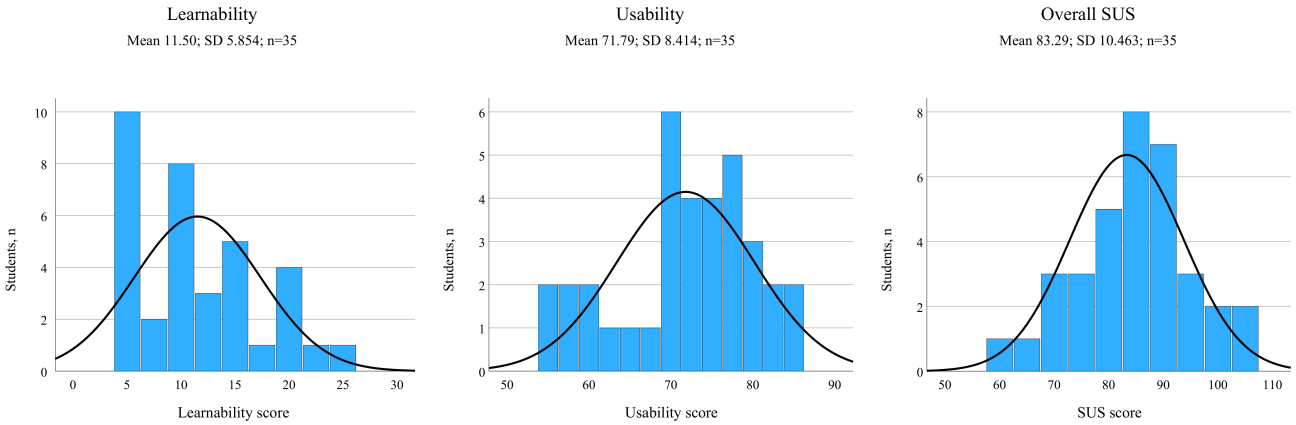
XSL•FO
RenderX

Figure 2. Analysis of system usability testing. A: B: C: SUS System Usability Scale.

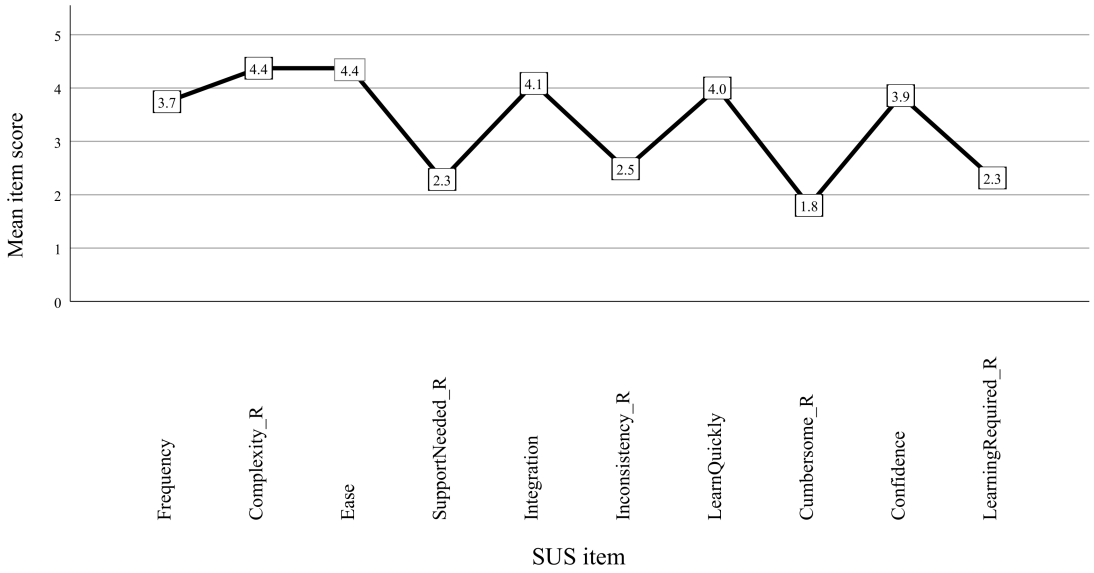
A



B

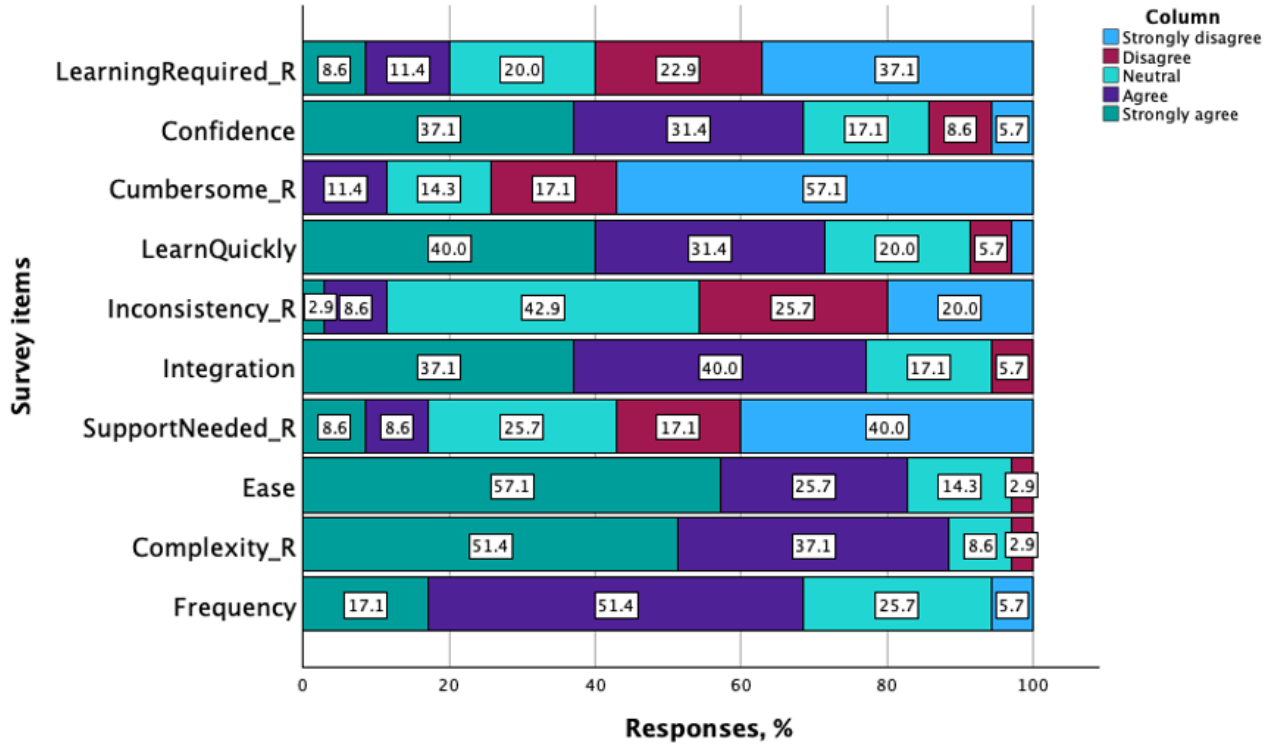


C



XSL•FO
RenderX

Figure 3. Stacked bar graph of System Usability Scale (SUS) results for Sistem Informasi Layanan Konseling Akademik (SISKA).



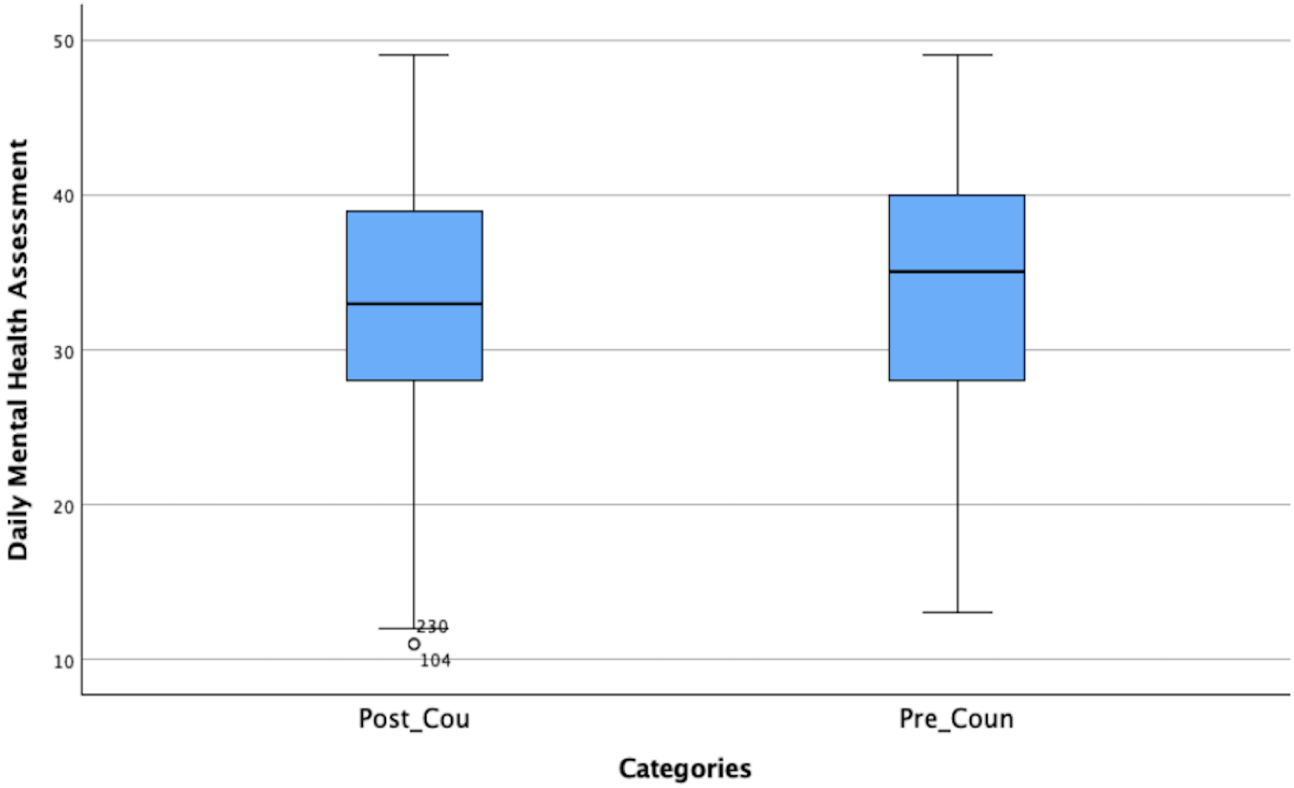
The SISKA app had an integrated and easy-to-learn system, but there were gaps in user confirmation of counseling schedules after registration (Multimedia Appendix 3). The comprehensive and simple counselor menu facilitated access to self-assessments, counseling scheduling, and mental health educational content. Initially, users may have felt confused about accessing SISKA, but after further attempts, it became easier to use. This indicates that users adapted quickly to accessing the SISKA app.

Help-Seeking Enablement and Early Support

The pre- and postintervention subsample comprised 119 students who received at least one SISKA counseling session and had valid paired self-assessments around that index session, drawn from the social and humanities fields (n=68, 57.1%); science, technology, engineering, and mathematics (n=27, 22.7%); and health sciences (n=24, 20.2%). Among the 119 students, 112 (94.12%) received 1 counseling session, 4 (3.36%) received 2 sessions, 2 (1.68%) received 3 sessions, and 1 (0.84%) received 4 sessions. From pre- to postcounseling, the proportion of students classified as at risk of mental health disorders decreased from 63 (52.9%) to 50 (42.0%), while the proportion reporting no mental health problems increased from 4 (3.4%) to 8 (6.7%) (Table 3). Referral remained high but tapered modestly as needs abated, decreasing from 96.6% (n=115) to 93.3% (n=111), with the complementary no-referral proportion increasing from 4 (3.4%) to 8 (6.7%). On the continuous metric, daily mental health scores decreased from a preintervention mean of 33.9 (SD 8.07) to a postintervention mean of 32.8 (SD 8.94) with a statistically significant paired-sample test result (Δ=−1.14; P=.03; Figure 4). The standardized effect size was small (Cohen d=0.18, 95% CI −0.005 to –0.358), indicating that the statistically significant change represented a small practical magnitude. Because higher scores indicate greater distress, the decrease in score suggests a small preliminary improvement in students' daily mental health assessments after counseling.

XSL•FO
RenderX

Figure 4. Analysis of pre- and postintervention daily mental health assessments of Sistem Informasi Layanan Konseling Akademik (SISKA) users.



The SISKA app enabled students to seek help early and provided support that felt like a psychologically safe space. Students felt psychologically comfortable in expressing their concerns and emotions. The safe space allowed them to share their stories, thereby reducing anxiety. SISKA also helped identify early symptoms and students' mental state. These indications strengthened students' readiness to continue treatment with professional services on campus. The features in SISKA also helped determine students' steps in choosing a counselor and provided guidance on selecting the next treatment steps. SISKA served as an initial gateway to seek help by improving early detection and decision support. SISKA not only featured system integration, but also provided psychological empowerment for students from their initial contact.

Strengths of SISKA

Quantitatively, SISKA's strengths were reflected in the SUS total clustering high on the 0 to 100 scale (mean 83.3) and in item-level agreement highlighting quick learning and ease: strong scores of "agree" or "strongly agree" for "learn quickly" (71.4%) and "ease" (82.8%), as well as for "integration" (77.1%) and "confidence" (68.5%). These patterns, together with the near-normal SUS distribution and overlapping mean and median lines, point to an intuitive, coherent interface that supported rapid uptake and routine use (Figures 2 and 3).

Interview results indicated that SISKA's strengths lay in its ease of access and the value of its features, which directly demonstrated the counseling process. The app was easily accessible to counselors and students, simplifying counseling services. Furthermore, SISKA was helpful for periodic self-assessments and monitoring user progress, along with providing mental health education. Several features that supported the app's strengths included health articles, a counseling schedule management menu, and a counselor portfolio.

Weaknesses of SISKA

Figure 3 shows residual friction in a subset of user flows. Reverse-keyed items still showed modest disagreement totals, including "support needed" (8.6%+8.6%=17.2%), "learning required" (8.6%+11.4%=20.00%), "cumbersome" (11.4%+14.3%=25.7%), as well as sizeable "neutral" segments on "inconsistency" (42.9%; mean 2.5, SD 1.0) and "strongly agree" on "complexity" (51.4%; mean 4.4, SD 0.8 after reverse scoring). These patterns, together with the SUS distribution (Figure 2), indicate pockets of uneven labeling and feedback and occasional multistep burdens; prioritizing user interface standardization, clearer labeling, and simplified confirmation flows should reduce perceived inconsistency and residual effort. Several participants reported that the user interface and experience were inconsistent and that the system could sometimes only be accessed via a PC. Furthermore, if the connection was weak, the app page became inaccessible, loading was slow, and notifications were not delivered in real time. Some users also complained that the counseling confirmation process was unfriendly, as there was no contact person to follow up on schedules. Users desired integrated and real-time counseling confirmation. These qualitative results explain the neutral perception regarding system inconsistency and complexity.

App analytics showed 502 monthly active users during the most recent monitoring period, although user retention remained low, with a 7-day retention value of 2 users. Additional system analytics demonstrated increasing device acquisitions and first-time app openings, indicating continued user reach and onboarding activity. However, the decline in long-term

XSL•FO
RenderX

active-user retention suggests that there were challenges in sustaining engagement over time. These findings provide preliminary operational evidence regarding the feasibility of SISKA implementation while also highlighting the need for future evaluation of counselor workload distribution, follow-up intensity, and long-term service sustainability.

Implementation Readiness

Readiness was supported by early statistically significant improvement in continuous daily mental health scores (P=.03) within the counseled pre- and postintervention cohort and by the usability profile indicating fast learning and routine use. To strengthen deployment at scale, implementation should prioritize the quantitative friction points identified in item-level diagnostics, including streamlining confirmations, reinforcing consistency cues, and mitigating complexity in longer flows while preserving the high-performing aspects (ease, learnability, and integration) that underpinned adoption and sustained engagement.

Ultimately, users provided suggestions regarding the readiness of SISKA's implementation, taking into account design aspects, system design aesthetics, service and channel expansion, and system socialization. Users desired a more intuitive and effective interface to avoid confusion. Furthermore, the application needed to add new features, such as relaxation music, a chat feature with counselors, and accessibility via Apple iOS. Users also suggested improvements to the aesthetic appearance of the app without increasing the burden of access.

Discussion

Principal Findings

This research indicates that SISKA is learnable, acceptable, and able to open earlier support pathways for students. Quantitatively, the compact SUS profile with strong agreement on ease, quick learning, and integration signals a short learning curve and manageable cognitive load, supporting early adoption and repeat use. In the counselling-exposed pre- and postintervention cohort, a small yet statistically significant decline in daily mental health scores reflects proximal improvement when structured self-monitoring and brief in-app feedback were paired with timely counselor follow-up.

Qualitatively, students described psychological safety, easier expression of concerns, and useful psychoeducation, alongside scheduling features that oriented them toward professional services. At the same time, users identified friction points, including interface inconsistencies, multistep confirmation flows, and intermittent connectivity and notification delays that could disrupt momentum. Prioritizing user interface standardization, simplifying confirmations, and strengthening reliable, context-sensitive notifications while embedding ongoing monitoring and feedback loops would position SISKA at a feasibility stage, ready for structured implementation in campus settings.

User Adoption

The compact SUS profile with strong endorsement of ease, quick learning, and cross-screen integration indicates a short learning curve and manageable cognitive load, conditions known to support early retention in digital health tools. Framed by the person-based approach, which emphasizes designing around users' lived contexts and barriers, these signals suggest that SISKA's onboarding and screen flows were sufficiently light to convert initial curiosity into routine use [28]. Layering just-in-time guidance at first contact further reduces early uncertainty and helps users navigate key actions without excessive support, a design principle repeatedly linked to sustained engagement in mobile health [29]. More broadly, conceptual work on engagement indicates that low-friction access, coherent workflows, and timely feedback jointly shape both the experiential (usability and learnability) and behavioral (continued use) facets of engagement; the SISKA patterns aligned with this integrated view [30]. In student settings specifically, systematic reviews highlight that user experience is pivotal for adoption and persistence; thus, a mid-to-upper SUS performance with high item agreement on ease and learnability is consistent with conditions that enable early uptake on campus [31]. Overall, the SUS mean of 83.3, coupled with the item-level profile, is best interpreted as indicating early learnability with low support needs, providing a pragmatic foundation for subsequent engagement and pathway activation [28]. When mapped to the Bangor adjective rating scale, this score further indicates that SISKA was perceived as having excellent usability, supporting its feasibility-stage readiness [22].

Participants explained that they had initial confusion when using the SISKA app, but this improved after guidance. This finding is consistent with research on app-based digital mental health interventions, which emphasizes the importance of providing just-in-time guidance during the onboarding phase to enhance learnability. Past work also explains that there is a decrease in anxiety and depression among students when a workflow is broken down into simple steps [32]. Research also shows that a chatbot app designed in a simple and integrated way can demonstrate rapid acceptance by students [33]. A similar mechanism in chatbot interventions showed feasibility, acceptability, and symptom improvement when users received guidance on how to use the app [34]. Features designed for easy adoption can indicate a lower learning burden and a reduced risk of anxiety.

Help-Seeking Enablement and Early Support

Considered together, the movement of students out of the most severe category and the small but statistically significant pre- and postintervention positive change in daily mental health scores are consistent with proximal improvement when digital self-monitoring is coupled to timely human follow-up. Although the pre- and postintervention daily mental health assessment change was statistically significant, the magnitude of change was small. This is supported by the standardized effect size, which indicated a small effect. Therefore, this finding should be interpreted as a preliminary feasibility-stage signal rather than definitive evidence of clinical effectiveness. In addition, the DMHA was designed as a monitoring tool embedded within the SISKA platform and did not have an established minimal clinically important difference threshold. The present analysis included students who attended at least one counseling session.

XSL•FO
RenderX

As many participants received only an initial counseling contact, the observed change in daily mental health assessment score likely reflects early or proximal improvement rather than sustained clinical change. Additional counseling sessions and longer follow-up periods may be required to produce larger or clinically meaningful improvements in student mental health. Therefore, the observed difference should be considered an exploratory signal of potential improvement rather than definitive evidence of clinically meaningful change. In campus settings, structured self-assessment, in-app decision support, and rapid escalation to counselors have repeatedly been linked to earlier pathway activation and short-term symptom gains when compared with reactive, crisis-driven care models [31,35]. Framing SISKA's workflow within measurement-based care helps explain this pattern: repeated symptom data and brief feedback loops enable quicker adjustments to support and increase user agency, which in turn can translate into modest near-term relief [36,37]. While these gains came from feasibility-stage research and are proximal, they indicate that integrating self-report streams with actionable prompts and fast counselor escalation can bridge the critical early step from identification to first contact on campus [35,37].

The SISKA app facilitated expression of concerns, raised awareness of symptoms, provided decision support for students, and sped up counselor follow-ups. In line with previous studies, self-assessment with the app encouraged students to seek help to improve short-term symptoms [38]. Other studies have confirmed that systems integrated within the campus can encourage students to access counseling services [39]. Based on the experiences of students who have accessed SISKA services, speedy follow-up with counselors helped anxious students find immediate help. Support in the form of prompts and just-in-time content was associated with reduced anxiety and depression, as well as improved well-being [40]. The availability of rapid escalation within the app to immediate assistance was a key component in bridging the initial phase to treatment [41].

Additionally, the app helped students determine the next steps in selecting a counselor with whom they felt comfortable after the assessment. In a pilot study, student engagement increased when the app provided clear and actionable guidance [32]. The principle of measurement-based care supports informed decision-making through continuous symptom data and periodic feedback within the app [42]. An app with integrated campus links made it easier for students to access initial mental health support.

Strengths of SISKA

The usability signal, reflected in the compact SUS profile and strong agreement on ease, quick learning, integration, and confidence, indicates a low-friction and coherent interface that keeps cognitive demands manageable. In digital mental health contexts, such simplicity and consistency have repeatedly been linked to better user experience and greater persistence among university students. Reviews in the JMIR family of journals show that interventions with good adoption on campuses tend to be those that minimize entry costs, provide clear navigation, and deliver relevant content that helps users decide on next steps for support [31,43] The person-based approach further explains why these features matter, since design that is grounded in users' lived contexts and barriers is more acceptable and more likely to translate first-time use into routine engagement [28]. Attention to these fundamentals also guards against the "law of attrition," where burdensome or inconsistent experiences erode use over time; maintaining lightweight performance, consistent workflows, and clear in-app guidance helps counter early drop-off and supports continued self-assessment, progress monitoring, and timely access to counseling [14,31,43].

The SISKA app has key strengths, including low friction that makes it easy for users to access; features that support self-assessment, monitoring, and scheduling; and relevant psychoeducational content. These strengths are consistently in line with the dimensions of usability, usefulness, and satisfaction. The success of mental health apps relies on their lightweight design and precise workflow [44,45]. Maintaining an app with low-friction access, simple login procedures, consistent navigation, and lightweight performance enhances initial engagement and increases the likelihood of continued use. Another strength lies in the SISKA app's features that make it easy for users to access mental health services. The self-assessment feature enhanced students' awareness of mental health symptoms, allowing them to receive feedback that accelerated their orientation toward professional help [46]. SISKA users' favorite feature was the psychoeducation content and mental health articles. Quality and easily accessible psychoeducation content can help users improve their literacy and encourage help-seeking [45]. The present study did not directly analyze the relationship between individual content use and help-seeking outcomes, although the SISKA system included psychoeducational content. Therefore, the relationship between these components and help-seeking should be interpreted as suggestive rather than empirically demonstrated.

Weaknesses of SISKA

The diagnostic pattern of residual burden on reverse-keyed items, together with large neutral plateaus in inconsistency and complexity, signals microfrictions that increased cognitive effort and, over time, could erode engagement if left unaddressed. Prior work shows that small mismatches in labels, icons, and feedback, as well as multistep burdens within a use flow, cumulatively raise the cost of use and accelerate attrition; targeted user interface standardization, clearer labels, and compression of critical paths are therefore priority remedies [14,47]. In mobile health specifically, timing and reliability of prompts matter for maintaining momentum. Just-in-time adaptive intervention principles recommend tailoring prompts to moments of opportunity and minimizing disruption, which aligns with evidence that notifications can produce strong near-term opening of the app but require optimization to avoid habituation and long-term drop-off [29,48]. For SISKA, this implies a dual emphasis on interface coherence and prompt strategy. Standardizing terminology and status cues across screens, simplifying confirmation flows, and ensuring dependable, context-sensitive notifications can reduce perceived inconsistency and effort, support smoother hand-offs to counselors, and protect against the gradual decline in use documented in digital health interventions [14,29,47,48].

XSL•FO
RenderX

The SISKA app was still under development during the study period and had performance and connectivity issues. The recommendations for user interface standardization, confirmation flow simplification, and notification optimization were derived from usability observations and qualitative feedback. Some of the failings identified by participants were an unfriendly design with inconsistent user interface and experience. The limitations of this app make students vulnerable, and confirmations are only one-way [49]. In addition to performance and design weaknesses, connectivity issues caused notifications to be not be delivered in real time, leading to users dropping off when accessing the app [49]. The quality of an app determines whether it continues to be used.

Implementation Readiness

Taken together, a statistically significant pre- and postintervention improvement in the daily mental health score and a usability profile characterized by fast learning and routine use support feasibility-stage readiness for a structured rollout in the campus setting [35]. To translate these proximal gains into sustainable practice, deployment should systematically resolve quantified friction points, including clarifying confirmation paths, reinforcing cross-screen consistency cues, and reducing multistep complexity, so that adoption and maintenance remain favorable across the reach, efficacy, adoption, implementation and maintenance (RE-AIM) dimensions [35,50]. Embedding measurement-based care loops that convert repeated self-reports into timely feedback and rule-based actions can operationalize day-to-day decision support for students and counselors [37]. For scale-up across diverse faculties and services, aligning the next phase with implementation frameworks will strengthen generalizability and fit: RE-AIM helps balance reach, effectiveness, adoption, implementation quality, and maintenance in real workflows [50], while assessing implementation outcomes such as acceptability, appropriateness, feasibility, and fidelity guides site-level tailoring [51], and Consolidated Framework for Implementation Research (CFIR)–informed diagnostics surface local barriers and facilitators before wider rollout [52]. Together, these steps provide a pragmatic pathway from feasibility to scalable service integration on campus, with continuous feedback driven by measurement-based care ensuring that improvements are retained and amplified over time.

Several user suggestions addressed identified weaknesses. Users expected the SISKA app design to be simple, with extensive socialization within the system, additional features, and improved aesthetics. In line with the CFIR framework [52], the health service implementation framework was based on organizational readiness, stakeholder engagement, user-friendliness, and social norms. In the implementation of mental health services on campus, a tiered and multiplatform socialization strategy increases reach and reduces barriers to access [31]. The implications of this research inform the design of service expansion by considering capacity, integration, procedural standards, limited trials, and rebuilding design to maintain simplicity, appeal, and user sustainability in accessing the SISKA app. The recommendations should be interpreted as preliminary design considerations for future system development studies, as this study did not conduct feature-level analyses linking system use to help-seeking behavior.

The SISKA system also conceptually aligns with several established frameworks in digital mental health and implementation science. First, the use of real-time self-assessment and automated feedback mechanisms reflects core principles of just-in-time adaptive interventions, which emphasize delivering support at moments of need using dynamic data signals [29]. Second, the repeated daily mental health assessments embedded within SISKA are consistent with principles of measurement-based care, where routine symptom monitoring informs clinical decision-making and treatment adjustment [53]. Finally, the feasibility evaluation reflects several early implementation dimensions described in the RE-AIM framework, particularly reach, adoption, and implementation in real-world service settings [50,54]. However, the present study did not formally test these frameworks. Instead, they were used as conceptual lenses to interpret the system design and early feasibility signals; future studies should evaluate these mechanisms using more rigorous implementation and effectiveness designs.

Limitations

Limitations include a nonrandomized, single-arm pre- and postintervention design without controls, so changes may have reflected regression to the mean, maturation, or concurrent events rather than use of SISKA. The precision of the usability estimate was limited by the relatively small SUS subsample (n=35), although small samples are commonly used in feasibility and early-stage usability evaluations to generate preliminary usability insights. Reliance on self-report measures (daily mental health assessments and SUS scores) may have introduced recall and social-desirability biases. Generalizability was restricted to Android users at Indonesian public universities with trained lecturer-counselors. In addition, use patterns may have been influenced by academic-calendar seasonality and unmeasured support services, and voluntary participation in qualitative feedback may have led to overrepresentation of more engaged or satisfied users.

Conclusion

SISKA demonstrated promising feasibility, acceptability, and usability for campus-based mental health support. A small but statistically significant improvement in daily mental health scores was observed among counseled students. These findings should be interpreted as preliminary. Future implementation should prioritize standardizing user interface elements, simplifying confirmation user flows, and strengthening reliable and context-sensitive notifications while embedding measurement-based care loops to convert routine self-reports into timely feedback and actions. The generalizability of these findings may be limited by the specific institutional context, as SISKA was implemented within Indonesian universities and relied on a lecturer-counselor model for student mental health support. Next steps include pragmatic comparative trials, longer follow-up to assess the durability and equity of outcomes, evaluation of counselor workload and service capacity, and cost-effectiveness analyzes to inform sustainable scale-up.

XSL•FO
RenderX

Acknowledgments

The authors would like to thank the Ministry of Research, Technology, and Higher Education, Directorate General of Resources for Research, Technology, and Higher Education, for facilitating the 2025 Riset Konsorsium Unggulan Berdampak (RIK UB). We also gratefully acknowledge the Ministry of Health of the Republic of Indonesia as our research partner. We are grateful for the support from the Institute of Learning Development and Quality Assurance and the Research and Community Service Institution, Universitas Jember, in facilitating this study. The authors would like to acknowledge Bambang Toni Prayogi for his contribution as the developer of the SISKA system. Finally, we extend our appreciation to the 5 RIK UB teams (Universitas Brawijaya, Universitas Airlangga, Universitas Diponegoro, Universitas Jember, and Universitas Respati Indonesia) and to all consortium research teams under the PRISMA Research Center for their valuable collaboration and contributions.

Generative AI Use

The authors declare the use of generative AI (GenAI) in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: proofreading, editing, and translation. The GenAI tool used was ChatGPT (GPT-5.5; OpenAI). Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Funding

This work was supported by the Ministry of Research, Technology, and Higher Education, Directorate General of Resources for Research, Technology, and Higher Education under the Riset Konsorsium Unggulan Berdampak (RIK UB; 009/C3/DT.05.00/RIKUB/2025).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the confidential nature of student mental health data but are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: TS, HDW, NALA, RI, HEN
Data curation: NALA, RI, LA, AEP
Formal analysis: NALA, RI, AEP
Funding acquisition: HDW, TS, HEN, MD
Investigation: HDW, LA, MD, S, YA
Methodology: TS, HDW, MD, TM
Project administration: RI, AEP, TM
Software: HDW, LA, MD, EF, S
Supervision: TS, HDW, HEN, EF, YA
Validation: TS, LA, HEN, TM, S, YA
Visualization: NALA, AEP, TM, EF
Writing – original draft: TS, NALA, RI, LA, AEP, HEN, MD
Writing – review & editing: TS, HDW, NALA, RI, TM, EF, S, YA
All authors have read and approved the final manuscript and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Diagram of participation flow.
[DOCX File, 26 KB - humanfactors_v13i1e87647_app1.docx]

Multimedia Appendix 2
Items in qualitative survey.
[DOCX File, 14 KB - humanfactors_v13i1e87647_app2.docx]

Multimedia Appendix 3
Thematic analysis of student and counselor experiences.
[DOCX File, 17 KB - humanfactors_v13i1e87647_app3.docx]

XSL•FO
RenderX

References

1. Auerbach RP, Mortier P, Bruffaerts R, et al. WHO World Mental Health Surveys International College Student project: prevalence and distribution of mental disorders. J Abnorm Psychol 2018 Oct;127(7):623-638. [doi: 10.1037/abn0000362] [Medline: 30211576]
2. Mason A, Rapsey C, Sampson N, et al. Prevalence, age-of-onset, and course of mental disorders among 72,288 first-year university students from 18 countries in the World Mental Health International College Student (WMH-ICS) initiative. J Psychiatr Res 2025 Mar;183:225-236. [doi: 10.1016/j.jpsychires.2025.02.016] [Medline: 40010072]
3. Lui JC, Sagar-Ouriaghli I, Brown JSL. Barriers and facilitators to help-seeking for common mental disorders among university students: a systematic review. J Am Coll Health 2024 Nov;72(8):2605-2613. [doi: 10.1080/07448481.2022.2119859] [Medline: 36084266]
4. Wiljer D, Shi J, Lo B, et al. Effects of a mobile and web app (Thought Spot) on mental health help-seeking among college and university students: randomized controlled trial. J Med Internet Res 2020 Oct 30;22(10):e20790. [doi: 10.2196/20790] [Medline: 33124984]
5. Ozamiz-Etxebarria N, Idoiaga Mondragon N, Tsybuliak N. Strengthening mental health among university students. Front Psychol 2025;16:1689173. [doi: 10.3389/fpsyg.2025.1689173] [Medline: 41280192]
6. Awadalla S, Davies EB, Glazebrook C. A longitudinal cohort study to explore the relationship between depression, anxiety and academic performance among Emirati university students. BMC Psychiatry 2020 Sep 11;20(1):448. [doi: 10.1186/s12888-020-02854-z] [Medline: 32917172]
7. Sinval J, Oliveira P, Novais F, Almeida CM, Telles-Correia D. Exploring the impact of depression, anxiety, stress, academic engagement, and dropout intention on medical students' academic performance: a prospective study. J Affect Disord 2025 Jan 1;368:665-673. [doi: 10.1016/j.jad.2024.09.116] [Medline: 39303881]
8. Barnett P, Saunders R, Buckman JEJ, et al. Are students less likely to respond to routinely delivered psychological treatment? A retrospective cohort analysis. Compr Psychiatry 2022 Nov;119:152348. [doi: 10.1016/j.comppsych.2022.152348] [Medline: 36191389]
9. Barnett P, Arundell LL, Saunders R, Matthews H, Pilling S. The efficacy of psychological interventions for the prevention and treatment of mental health disorders in university students: a systematic review and meta-analysis. J Affect Disord 2021 Feb 1;280(Pt A):381-406. [doi: 10.1016/j.jad.2020.10.060] [Medline: 33227669]
10. Barkham M, Broglia E, The SCORE Consortium. Routine outcome monitoring (ROM) and feedback in university student counselling and mental health services: considerations for practitioners and service leads. Couns and Psychother Res 2024 Jun;24(2):459-471. [doi: 10.1002/capr.12694]
11. Aryani F, Mewarni S, Rais M, Rafli M. Optimalisasi layanan kesehatan mental di kalangan mahasiswa melalui pengembangan model konseling online. Seminar Nasional Hasil Penelitian 2024;1(11):1510-1518 [FREE Full text]
12. Priharsanto AS, Ajhari AA. Pengembangan aplikasi kesehatan mental berbasis kotlin: studi kasus aplikasi hugme. Jurnal Kajian Teknik Elektro 2024;9(2):102-108. [doi: 10.52447/jkte.v9i2.7690]
13. Eddy Wibowo M, Mahfud A, Khiyarusoleh U. Development of electronic group counseling (EGC) software to optimize counseling services in Indonesia. Konselor 2023;12(3):141-149. [doi: 10.24036/0202312321-0-86]
14. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11. [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
15. Wiebe DE, Remers S, Nippak P, Meyer J. Evaluation of an online system for routine outcome monitoring: cross-sectional survey study. JMIR Ment Health 2021 Dec 1;8(12):e29243. [doi: 10.2196/29243] [Medline: 34855615]
16. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 5;16(6):e146. [doi: 10.2196/jmir.3077] [Medline: 24905070]
17. Lewis JR. Binomial confidence intervals for small sample usability studies. In: Proceedings of the 1st International Conference on Applied Ergonomics: USA Publishing; 1996:732-737.
18. Sauro J, Lewis JR. In: Elliot S, editor. Quantifying the User Experience: Practical Statistics for User Research: Elsevier Inc; 2012.
19. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
20. Hasibuan DP, Santoso HB, Yunita A, Rahmah A. An Indonesian adaptation of the E-Learning Usability Scale. J Phys: Conf Ser 2020 Jun 1;1566(1):012051. [doi: 10.1088/1742-6596/1566/1/012051]
21. Lewis JR, Sauro J. The factor structure of the System Usability Scale. In: International Conference on Human Centered Design: Springer; 2009. URL: https://link.springer.com/chapter/10.1007/978-3-642-02806-9_12 [doi: 10.1007/978-3-642-02806-9_12]
22. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123. [doi: 10.5555/2835587.2835589]
23. Kessler RC, Andrews G, Colpe LJ, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. Psychol Med 2002 Aug;32(6):959-976. [doi: 10.1017/s0033291702006074] [Medline: 12214795]

XSL•FO
RenderX

24. Tran TD, Kaligis F, Wiguna T, et al. Screening for depressive and anxiety disorders among adolescents in Indonesia: formal validation of the Centre for Epidemiologic Studies Depression Scale - Revised and the Kessler Psychological Distress Scale. J Affect Disord 2019 Mar 1;246:189-194. [doi: 10.1016/j.jad.2018.12.042] [Medline: 30583144]
25. Savitri RIW, Oktari S, Mafaza Y, et al. The effect of coping strategies toward psychological distress in orphanage adolescents. Jurnal Psibernetika 2024;17(2):104. [doi: 10.30813/psibernetika.v17i2.6168]
26. Braun V, Clarke V. Conceptual and design thinking for thematic analysis. Qual Psychol 2022;9(1):3-26. [doi: 10.1037/qup0000196]
27. Komite Etik Penelitian dan Pengembangan Kesehatan. Pedoman dan Standar Etik Penelitian dan Pengembangan Kesehatan Nasional/ Komite Etik Penelitian dan Pengembangan Kesehatan [Health Research and Development Ethics Committee. National Guidelines and Ethical Standards for Health Research and Development. Jakarta, Indonesia: Health Research and Development Ethics Committee]. 2021 URL: https://kepkn.kemkes.go.id/assets/img/keppkn/file/BUKU_PEDOMAN_DAN_STANDAR_ETIK_PENELITIAN_DAN_PENGEMBANGAN_KESEHATAN_NASIONAL(KEPKN).pdf [accessed 2026-10-19]
28. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30. [doi: 10.2196/jmir.4055] [Medline: 25639757]
29. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
30. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
31. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869. [doi: 10.2196/12869] [Medline: 31333198]
32. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 6;4(2):e19. [doi: 10.2196/mental.7785] [Medline: 28588005]
33. Mustafa SH, Mohammed EA, Salih AM, et al. A qualitative exploration of acceptance of a conversational chatbot as a tool for mental health support among university students in Sudan. Int J Med Sci and Nurs Res ;4(1):16-23. [doi: 10.55349/ijmsnr.2024411623]
34. Klos MC, Escoredo M, Joerin A, Lemos VN, Rauws M, Bunge EL. Artificial intelligence-based chatbot for anxiety and depression in university students: pilot randomized controlled trial. JMIR Form Res 2021 Aug 12;5(8):e20678. [doi: 10.2196/20678] [Medline: 34092548]
35. Taylor ME, Liu M, Abelson S, Eisenberg D, Lipson SK, Schueller SM. The reach, effectiveness, adoption, implementation, and maintenance of digital mental health interventions for college students: a systematic review. Curr Psychiatry Rep 2024 Dec;26(12):683-693. [doi: 10.1007/s11920-024-01545-w] [Medline: 39392547]
36. Fortney JC, Unützer J, Wrenn G, et al. A tipping point for measurement-based care. Psychiatr Serv 2017 Feb 1;68(2):179-188. [doi: 10.1176/appi.ps.201500439] [Medline: 27582237]
37. Lewis CC, Boyd M, Puspitasari A, et al. Implementing measurement-based care in behavioral health: a review. JAMA Psychiatry 2019 Mar 1;76(3):324-335. [doi: 10.1001/jamapsychiatry.2018.3329] [Medline: 30566197]
38. Martinengo L, Stona AC, Tudor Car L, Lee J, Griva K, Car J. Education on depression in mental health apps: systematic assessment of characteristics and adherence to evidence-based guidelines. J Med Internet Res 2022 Mar 9;24(3):e28942. [doi: 10.2196/28942] [Medline: 35262489]
39. Lattie E, Cohen KA, Winquist N, Mohr DC. Examining an app-based mental health self-care program, IntelliCare for college students: single-arm pilot study. JMIR Ment Health 2020 Oct 10;7(10):e21075. [doi: 10.2196/21075] [Medline: 33037874]
40. Edge D, Newbold A, Ehring T, Rosenkranz T, Frost M, Watkins ER. Reducing worry and rumination in young adults via a mobile phone app: study protocol of the ECoWeB (Emotional Competence for Well-Being in Young Adults) randomised controlled trial focused on repetitive negative thinking. BMC Psychiatry 2021 Oct 21;21(1):519. [doi: 10.1186/s12888-021-03536-0] [Medline: 34674669]
41. Meyerhoff J, Popowski SA, Lakhtakia T, et al. Automated digital safety planning interventions for young adults: qualitative study using online co-design methods. JMIR Form Res 2025 Feb 26;9:e69602. [doi: 10.2196/69602] [Medline: 40009840]
42. Moggia D, Lutz W, Brakemeier EL, Bickman L. Treatment personalization and precision mental health care: where are we and where do we want to go? Adm Policy Ment Health 2024 Sep;51(5):611-616. [doi: 10.1007/s10488-024-01407-w] [Medline: 39172281]
43. Ferrari M, Allan S, Arnold C, et al. Digital interventions for psychological well-being in university students: systematic review and meta-analysis. J Med Internet Res 2022 Sep 28;24(9):e39686. [doi: 10.2196/39686] [Medline: 36169988]

XSL•FO
RenderX

44. Kaveladze BT, Wasil AR, Bunyi JB, Ramirez V, Schueller SM. User experience, engagement, and popularity in mental health apps: secondary analysis of app analytics and expert app reviews. JMIR Hum Factors 2022 Jan 31;9(1):e30766. [doi: 10.2196/30766] [Medline: 35099398]
45. Madrid-Cagigal A, Kealy C, Potts C, et al. Digital mental health interventions for university students with mental health difficulties: a systematic review and meta-analysis. Early Interv Psychiatry 2025 Mar;19(3):e70017. [doi: 10.1111/eip.70017] [Medline: 40033658]
46. Hawley S, Yu J, Bogetic N, et al. Digitization of measurement-based care pathways in mental health through REDCap and electronic health record integration: development and usability study. J Med Internet Res 2021 May 20;23(5):e25656. [doi: 10.2196/25656] [Medline: 34014169]
47. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen J. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152. [doi: 10.2196/jmir.2104] [Medline: 23151820]
48. Bell L, Garnett C, Bao Y, et al. How notifications affect engagement with a behavior change app: results from a micro-randomized trial. JMIR Mhealth Uhealth 2023 Jun 9;11:e38342. [doi: 10.2196/38342] [Medline: 37294612]
49. Giebel GD, Abels C, Plescher F, et al. Problems and barriers related to the use of mHealth apps from the perspective of patients: focus group and interview study. J Med Internet Res 2024 Apr 23;26:e49982. [doi: 10.2196/49982] [Medline: 38652508]
50. Glasgow RE, Harden SM, Gaglio B, et al. RE-AIM planning and evaluation framework: adapting to new science and practice with a 20-year review. Front Public Health 2019;7:64. [doi: 10.3389/fpubh.2019.00064] [Medline: 30984733]
51. Proctor E, Silmere H, Raghavan R, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76. [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
52. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated Consolidated Framework for Implementation Research based on user feedback. Implement Sci 2022 Oct 29;17(1):75. [doi: 10.1186/s13012-022-01245-0] [Medline: 36309746]
53. Scott K, Lewis CC. Using measurement-based care to enhance any treatment. Cogn Behav Pract 2015 Feb;22(1):49-59. [doi: 10.1016/j.cbpra.2014.01.010] [Medline: 27330267]
54. Glasgow RE, Vogt TM, Boles SM, Glasgow E. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
K10: 10-item Kessler Psychological Distress Scale
RE-AIM: reach, efficacy, adoption, implementation and maintenance
SISKA: Sistem Informasi Layanan Konseling Akademik
SUS: System Usability Scale
WMH-ICS: World Mental Health Surveys International College Student

Edited by A Kushniruk; submitted 12.Nov.2025; peer-reviewed by M Pulier, Y Chu; revised version received 12.May.2026; accepted 23.Jun.2026; published 06.Nov.2026.

Please cite as:
Susanto T, Windarwati HD, Ati NAL, Istifada R, Armaijn L, Pranata AE, Nihayati HE, Dwidiyanti M, Marliana T, Fahamsyah E, Slamin , Arihandayani Y
Feasibility and Acceptability of the SISKA System for Continuous Student Emotional and Mental Health Guidance: Mixed Methods Study
JMIR Hum Factors 2026;13:e87647
URL: https://humanfactors.jmir.org/2026/1/e87647
doi:10.2196/87647

© Tantut Susanto, Heni Dwi Windarwati, Niken Asih Laras Ati, Rizkiyani Istifada, Liasari Armaijn, Andi Eka Pranata, Hanik Endang Nihayati, Meidiana Dwidiyanti, Thika Marliana, Ermanto Fahamsyah, Slamin, Yunita Arihandayani. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 6.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly

XSL•FO
RenderX

cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Insights on Designing the MINT (Micro-Interventions for Navigating Tension) Chatbot System for Stress Management: Collaborative Autoethnography

Marco Bolpagni[1,2], BSc, MSc; Sofia Asteria Piffer[2], BSc; Marco Pozza[1,2], BSc, MSc; Munesu Maminimini[1,2], BSc, MSc; Silvia Gabrielli[2], MSc, PhD

[1]Department of General Psychology, University of Padua, Via Venezia, 8, Padua, Veneto, Italy

[2]Centre for Digital Health and Wellbeing, Digital Health Research, Fondazione Bruno Kessler, Trento, Italy

Corresponding Author:
Marco Bolpagni, BSc, MSc
Department of General Psychology, University of Padua, Via Venezia, 8, Padua, Veneto, Italy

Abstract

Background: Digital microinterventions are increasingly used to support stress self-management in everyday life. While previous research has examined their feasibility, adherence, and preliminary effectiveness, much less is known about how these interventions are experienced over extended periods of real-world use.

Objective: This study used collaborative autoethnography (CAE) to examine the everyday experience of living with MINT (Micro-Interventions for Navigating Tension), a Telegram-based chatbot for stress management. Our goal was not to evaluate clinical outcomes but to understand how scheduled microinterventions fit into everyday life, when they were ignored, and when people relied on other ways of coping instead, and to identify implications for the design of future digital mental health tools.

Methods: We conducted a 4-week CAE involving 5 digital mental health researchers, who used the MINT chatbot as both participants and researchers. The experience was documented through multiple qualitative and quantitative sources, including reflective diaries, ecological momentary assessments (EMAs), workshop discussions, and wearable data, and analyzed collaboratively using reflexive thematic analysis (RTA).

Results: Six interconnected themes described how MINT was experienced in everyday life: (T1) the importance of timing and emotional receptivity, (T2) the need for reciprocity and relational continuity, (T3) the tendency to favor existing coping strategies over suggested exercises, (T4) the meaning of nonengagement during periods of high stress, (T5) the collapse of self-initiated use after reminders were removed, and (T6) how everyday experiences with consumer wearables and general-purpose large language models reshaped expectations of digital mental health systems.

Conclusions: Our findings show that CAE complements conventional evaluations of digital mental health interventions by revealing aspects of everyday use that are difficult to capture through feasibility studies, large-scale trials, or looking at usage data. Beyond the specific case of MINT, they suggest that future stress-management systems should better account for users' receptivity, provide greater continuity across interactions, build on existing coping strategies, and help users move from externally prompted engagement to autonomous self-management.

(JMIR Hum Factors 2026;13:e89876) doi:10.2196/89876

KEYWORDS

digital health; stress management; digital microinterventions; mental health chatbots; ecological momentary assessment; just-in-time adaptive interventions; collaborative autoethnography; user engagement; wearable sensors; generative AI

Introduction

Stress is a pervasive challenge that affects psychological well-being, cognitive functioning, and overall quality of life [1]. Despite its prevalence, individuals often lack accessible, personalized tools to help them recognize and regulate stress. Digital mental health interventions (DMHIs), including mindfulness apps [2], chatbots [3], ecological momentary interventions (EMIs) [4], and wearable devices [5], have attempted to fill this gap by delivering support outside clinical environments. One approach that has attracted growing attention is the use of digital microinterventions. Microinterventions are short, structured therapeutic activities, such as cognitive reappraisal [6], guided breathing [7], positive imagery [8], or mindfulness exercises, that can be completed in under a few minutes [9]. Their brevity and adaptability make them well suited for real-world stress self-management, where users often have only small windows of time available. Research has already explored several ways of delivering microinterventions in

XSL•FO
RenderX

everyday settings. For example, Microsoft researchers [10] have developed a just-in-time adaptive intervention (JITAI) system for workplace stress management, showing that context-triggered support, derived from indicators, such as email load, meeting frequency, time of day, facial expression cues, and heart rate, can be feasibly delivered and meaningfully integrated into daily activities. Other work [11] has focused on wearable-based approaches, where smartwatches periodically prompt users to report how they feel and then offer brief, breathing-oriented microinterventions to support emotional regulation during moments of elevated arousal. Further analysis [12] has evaluated a range of cognitive behavioral therapy (CBT)–based techniques to identify particularly promising exercises and examine how different stress levels may align with specific intervention strategies. Finally, a study [13] conducted a microrandomized trial with parents using the commercial mindfulness app Headspace to assess the feasibility and acceptability of delivering mindfulness-based microinterventions for stress regulation in daily life.

However, most field evaluations focus on feasibility, adherence, or preliminary outcomes and not on the lived experience of interacting with microinterventions across an extended period. As a result, while many of the practical challenges surrounding engagement with digital microinterventions have already been identified, we still know comparatively little about how these challenges play out in everyday life: how such interventions fit (or fail to fit) into daily routines, how users make sense of the activities, how emotional context shapes engagement, or how motivation fluctuates across weeks. This limitation reflects a broader concern in digital mental health research. Large trials and log-based analyses, while valuable for quantifying use, often miss how engagement is shaped by everyday contexts and circumstances, making it difficult to explain why people sustain or abandon DMHIs. Prior work [14] on "trial bias" further suggests that engagement observed in controlled studies may not reflect real-world use, underscoring the need for methodologies that foreground real-life contexts and subjective experiences. To better understand these everyday dynamics, researchers have increasingly turned to methods that foreground lived experience. This shift has prompted researchers both to involve users as collaborators, helping to interpret how digital mental health tools shape and are shaped by identity, emotion, and daily routines [15,16], and to turn the analytic lens inward, using autoethnographic methods to generate richly contextual insights from their own experiences [17]. This growing interest is reflected in recent work. Authors of the study [18], for example, used autoethnography to explore how low-tech event-logging can be used for smartphone disengagement. Another study [19] tracked 16 weeks of sleeping with a companion robot, highlighting how emotional and relational dynamics emerge through extended use. More recently, the authors of the study [20] used a duoethnographic design to examine how wearable-triggered prompts and large language model (LLM)–based interventions unfold in daily life.

This paper presents a collaborative autoethnography (CAE) [21] of MINT (Micro-Interventions for Navigating Tension), a Telegram chatbot delivering brief stress-management exercises through scheduled prompts and user-initiated access. Our goal was not to evaluate clinical efficacy or stress-contingent delivery. Instead, we used MINT as a formative case to examine a more basic design problem: how scheduled microinterventions are taken up, resisted, or displaced in everyday life. Over 4 weeks, 5 researchers used the system while documenting their own experiences as both users and designers. Through this case, we explore the conditions under which scheduled microinterventions fail to become part of a person's own coping repertoire and derive implications for the design of future stress-management systems.

Methods

Overview of the MINT Prototype

MINT is a Telegram chatbot designed to support stress self-management through brief, evidence-based microinterventions. The system aims to provide immediate stress relief while helping users develop skills for everyday resilience by teaching exercises that foster resilience. Users can log mood, stress, sleep, physical activity, and coping strategies, as well as record reflections in a diary-style format (Figure 1). Exercises (microinterventions) are available both through scheduled ("push") delivery and user-initiated ("pull") access, allowing users to actively select both the exercise and its delivery style. MINT's interventions are grounded in CBT techniques and include cognitive reappraisal [6], controlled breathing [7], positive mood induction [8], smiling [22], and guided visualization of one's best possible self [23]. Each exercise lasts under 5 minutes and can be tailored to match the user's preferred learning style according to the visual, aural, read/write, kinesthetic (VARK) framework [24], adapting its delivery as a visual diagram, audio guide, written instructions, or kinesthetic prompt (see an example in Multimedia Appendix 1). VARK variants were intentionally introduced to support learning and skill acquisition and to eventually help users internalize stress-management techniques. The flows of exercises were designed by the research team and delivered through scripted conversations in the chatbot interface.

XSL•FO
RenderX

Figure 1. Screenshots of the MINT (Micro-Interventions for Navigating Tension) Telegram chatbot illustrating core interaction types: (A) delivery of a scheduled microintervention, (B) manual stress and mood recording, (C) menu for user-initiated ("pull") selection of any exercise, and (D) ecological momentary assessment (EMA). All information related to real people has been censored. All remaining conversations are fictional; therefore, consent is not required.



A B C D

At the start, MINT collects an initial user profile, including personality (Big Five Inventory-10 [BFI-10; 25]), coping style (COPE-28 [Coping Orientation to Problems Experienced-28; 26]), cognitive style (Rational Experiential Inventory-13 [REIm-13; 27]), and learning style preferences (VARK v8.01 [28]); while most measures are used for research purposes, only the VARK preference guides personalization. Optional wearable integration is also supported and enables passive monitoring of physiological signals, such as electrodermal activity (EDA), heart rate variability (HRV), and skin temperature, to gather exploratory insights into the feasibility of wearing a smartwatch for multiple weeks and user perceptions of such data. Architecturally, MINT was developed as a rule-based conversational agent using Node.js and the grammY framework for Telegram API integration. Interventions and ecological momentary assessment (EMA) delivery followed scripted, finite-state decision trees, with user responses, logged metrics, and interaction timestamps stored in a MongoDB database. Scheduled push notifications were triggered via automated background cron jobs on a dedicated self-hosted server.

Study Design

As anticipated in the *Introduction*, we chose CAE to explore our lived, day-to-day experiences of using the MINT chatbot. The CAE [21] is a qualitative methodology in which researchers document and interpret their own experiences in a systematic and collective way, making visible aspects of everyday experience that are difficult to capture otherwise.

Participants and Researcher-Participant Positionality

All 5 authors served as researchers-participants (RPs; Table 1), engaging with MINT over a 4-week period in June 2025 while reflecting on its role in our daily lives. Each of us had prior experience studying or designing digital mental health technologies, and 1 (RP2) had led the development of the MINT prototype. Recognizing that this shared expertise inevitably shaped our expectations and interpretations, we paid explicit attention to positionality throughout the study, regularly discussing in debriefing sessions how our disciplinary backgrounds, design knowledge, and personal histories influenced our experiences and sensemaking of MINT.

XSL•FO
RenderX

Table . Characteristics of the 5 researchers-participants (RPs), including their roles, backgrounds, involvement in MINT (Micro-Interventions for Navigating Tension) development, PSS-10[a] scores, and visual, aural, read/write, kinesthetic (VARK) preferences.

ID	Sex/age (y)	Role (background)	Expertise	Involvement in MINT	PSS-10 (T0→T1)[b]	VARK (1st→2nd)
RP1	Female/55	Senior researcher (PhD in psychology, MSc in psychology)	HCI[c] digital health (psychology interventions)	Supervisor: aware of the overall concept	17→14	K[d]→R
RP2	Male/33	PhD researcher (MSc in DS[e], BSc in psychology)	Digital health (AI and psychology interventions)	Lead developer: built the bot and refined exercises/EMA[f] logic	17→17	K→V[g]
RP3	Male/30	PhD researcher (MSc in CS[h], BSc in CS)	Digital health (AI)	Regular research group member: user only	23→22	K→R[i]
RP4	Female/25	Technologist (MA in UX[j] psychology, BSc in design)	UX psychology	Designer: found and drafted many of the exercises	18→11	K→V
RP5	Male/25	PhD researcher (MSc in CS, BEng in electronics)	Digital health (engagement)	Regular research group member: user only	23→23	K→V

[a]PSS-10: Perceived Stress Scale.

[b]T0, measurement at the beginning, and T1, measurement at the end of the autoethnography.

[c]HCI: human computer interaction.

[d]K: kinesthetic.

[e]DS: data science.

[f]EMA: Ecological Momentary Assessment.

[g]V: visual.

[h]CS: computer science.

[i]R: read/write.

[j]UX: user experience.

Procedures

We interacted with MINT through the Telegram bot for 28 consecutive days following a structured weekly progression (Table 2), with each week focusing on different modes of engagement and intervention delivery to introduce frictions, sustain learning, and elicit responses under different conditions.

Table . Overview of the 4 experimental phases (weeks 1 - 4) and their respective intervention delivery strategies.

Week	Phase name	Description	Key focus/purpose
#1	Preferred modality delivery	One daily afternoon microintervention delivered in the participant's dominant VARK[a] learning preference.	Establish a baseline of "comfortable," intuitive, and preferred engagement.
#2	Nonpreferred modality delivery	Daily interventions continued but were intentionally mismatched with the participant's VARK preferences.	Create mild friction; explore resistance, adaptability, and emotional reactions to nonpreferred formats.
#3	Participant-selected activities	No specific exercises suggested; participants received only 1 daily prompt inviting them to choose any intervention themselves.	Shift responsibility to the user; emphasize autonomy and the cognitive/emotional effort of self-selection.
#4	Free-use period	All scheduled prompts removed; participants could interact (or not interact) with MINT[b] entirely on their own terms.	Capture fully self-directed patterns of engagement and disengagement.

[a]VARK: visual, aural, read/write, kinesthetic.

[b]MINT: Micro-Interventions for Navigating Tension.

XSL•FO
RenderX

Each weekday, we received an EMA in the morning at 9:15 AM, another EMA in the afternoon at 7:15 PM, and a microintervention at 6:15 PM (Table 2). To reduce burden, we received no interventions on weekends, and on these days we only completed the morning and afternoon EMAs. At the end of the 4 weeks, we hosted 2 in-person postintervention reflective workshops [29] (each approximately 90 min long) to revisit our experiences and develop a shared interpretation of the 4-week journey. The sessions were audio-recorded and transcribed to support subsequent analysis.

Ethical Considerations

This study was conducted in full accordance with the Declaration of Helsinki and with current Italian regulations governing research involving human participants. In this case, in line with Italian Law No. 3/2018 [30] (Art. 2, paragraphs 7 and 10), formal ethics committee approval was not required because the investigation focused exclusively on the usability and lived experience of a prototype digital intervention, did not involve patients or vulnerable individuals, and did not entail any clinical treatment. Nevertheless, all RPs provided voluntary, written informed consent before enrollment, were free to withdraw from the study at any time without consequence, and were fully informed about data handling procedures.

Data Collection

To capture our experiences from different perspectives, we collected multiple forms of data throughout the 4-week study period. Each of us kept a personal diary, noting episodes of stress, how the chatbot fit into our daily routines, and our emotional reactions to each interaction (eg, comfort, resistance, irritation, or relief). These diaries served as our primary narrative record that captured not only what happened but also how it felt to navigate ordinary routines with the system in the background. Alongside these individual accounts, we maintained a shared reflexive log where we posted brief observations, patterns we noticed, or questions we wanted to revisit later. These shared notes often prompted new reflections in our personal diaries, creating a fluid back-and-forth between private and communal sense-making. The interactions with MINT also generated an additional source of data. Our EMA responses and free-text messages, collected through the chatbot, provided brief snapshots of mood and stress that enriched the corpus of personal reflections. Finally, 3 of us used wearable devices for continuous physiological monitoring: RP2 wore an EmbracePlus (Empatica Inc) [31], RP4 wore a Galaxy Watch 4 (Samsung Electronics) [32], and RP3 wore a Watch SE2 (Apple Inc) [33]. These devices generated continuous physiological data throughout the study, complementing our subjective accounts while also revealing what it was like to live with continuous stress monitoring.

Data Analysis

All qualitative data were analyzed through reflexive thematic analysis (RTA) [34], following the 6-phase analytic process described by Braun and Clarke [35]. Consistent with the logic of CAE, the 2 postintervention workshops served as the primary site of analysis, providing a space for collective sense-making through which researchers reflected on, interpreted, and negotiated the meaning of their experiences. These discussions served as the main point of theme generation, capturing shared and divergent experiences with the MINT chatbot. The resulting themes were then revisited across the broader dataset, including personal diaries, the shared reflexive log, stress entries, EMAs, and summaries of wearable data, to refine and contextualize the emerging interpretations. All materials were stored in a shared Google Drive folder, cleaned for clarity, and anonymized using nonidentifying placeholders (eg, [NAME] to replace actual names of people cited in the notes/transcripts) and participant codes (RP1-RP5). Throughout the analysis, we documented emerging themes and analytic decisions in a shared codebook. The final themes reported in this paper reflect what emerged through this collaborative process of integration, shared discussions, and individual reflections across all data sources.

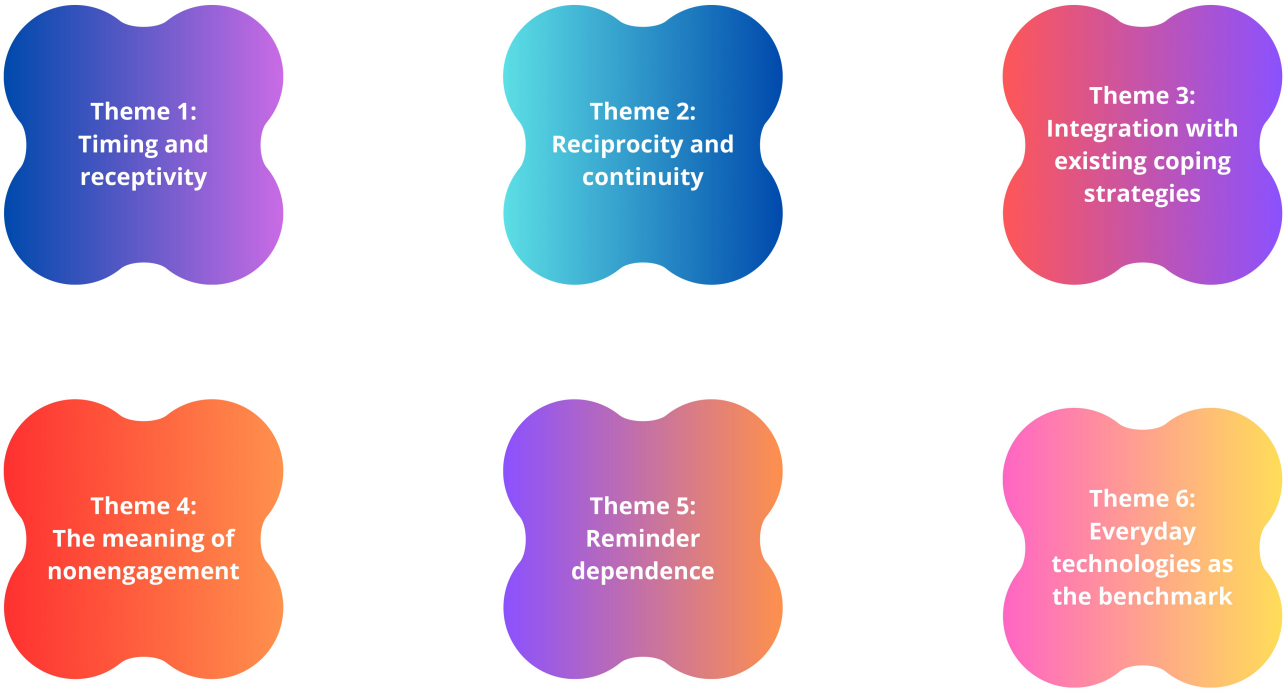
Results

Themes Identified Through Interactions With MINT

Over the 4 weeks, our collective engagement with MINT evolved in ways that were often surprising and sometimes frustrating. Across the entire corpus analyzed, 1 pattern became clear: our responses to the chatbot were not simply reactions to individual exercises. They depended just as much on when the prompts arrived, what we were already doing to cope, how much effort the interaction required, and whether the system seemed able to give anything back. We identified 6 interrelated themes (Figure 2) that describe core design challenges that emerged through interaction with MINT (more details on themes, subthemes, and coding instances are reported in Multimedia Appendix 2).

XSL•FO
RenderX

Figure 2. Overview of the 6 themes identified through collaborative reflexive thematic analysis (RTA).



Theme 1: Timing and Receptivity

From the first few days, the timing of MINT's notifications made a noticeable difference in how we received its exercises. There were moments when an intervention arrived just as we were settling into routines we had intentionally carved out for ourselves, like cooking dinner or finally sitting down to relax. RP2 captured this most vividly: "Very often [notifications from MINT] arrived when I was doing something else… For example, one day I had just turned on the Xbox, I was finally relaxing, and the reminder came. At that moment the bot was stealing the only space of peace I had managed to carve out." That feeling of intrusion was shared by all of us: even brief prompts could turn a low-stress moment into a jolt of irritation, not relief. But it was not only the schedule that mattered; our emotional state at the moment of delivery was equally decisive. Exercises framed with positivity, like visualizing the best possible self or the smiling exercise, were polarizing. When we were relatively calm, these exercises could feel energizing or pleasant. But on days when tension was high, the same prompts sometimes backfired. RP4 put it perfectly: "The best possible self is a perfect example of a double-edged tool… when I was already feeling down it made me feel worse, because I was imagining a version of myself that was completely unreachable at that moment." Smiling exercises could feel hollow or even irritating under stress, while other exercises, like controlled breathing, felt neutral and safe, something we could do without pretending that everything was fine. They demanded little emotional effort and could be accessed regardless of mood. These experiences showed us how timing and emotional alignment worked together as gatekeepers. The same exercise could either be welcomed or rejected depending on when it arrived and how we felt. For instance, even an exercise we normally enjoy could trigger annoyance if it appeared at a moment when our attention or emotional bandwidth was low. As the weeks went on, we noticed this pattern more clearly. We started anticipating how we might feel before even opening MINT, sometimes ignoring notifications in advance, preemptively judging the exercise's fit with our current state. Over time, we realized that acceptability was not just about content, but about context, about how each exercise intersected with the unpredictable rhythms of daily life.

Theme 2: Reciprocity and Continuity

Another recurring experience concerned the balance between effort and return. Using MINT required regular attention: daily EMAs, diary entries, exercise completion, and occasionally syncing wearables. Yet, despite investing this energy, there was little sense that the bot responded to us or incorporated our input in any meaningful way. RP5's reflection expressed this plainly: "You give the chatbot a lot of data but you get nothing back… no real insight." This sentiment resonated across the group. At times, the bot felt like an "empty box" (RP2) or "just another sequence of homework," (RP5), with no acknowledgment of what we had already done. Over time, we realized how much relational cues mattered. Small conversational touches, such as brief comments, acknowledgment of previous entries, and even light humor, could have made exercises feel less transactional. RP2 captured this perfectly: "The role of small talk is fundamental… Those circumstantial chats are what unblock everything in in-person counselling sessions. If you arrive at the therapy room and the counselor immediately applies a cognitive-behavioural technique, you feel you are only receiving directives." We noticed that without any sense that the bot "knew" us or remembered our past engagement, its exercises sometimes felt cold, prescriptive, and burdensome. Some of us also wanted visible feedback as evidence that our engagement mattered. RP1 wrote in her notes, "I would have loved a simple dashboard, even just a weekly message saying 'Last week you felt better on the days you walked more … shall we use that?'… something that shows the bot is actually doing something with everything I'm giving it." Without this, even well-designed

XSL•FO
RenderX

exercises sometimes felt like chores. Over time, this shaped how and when we engaged, as the perceived effort of participation without visible reciprocity began to weigh more heavily than the potential benefits of the exercises themselves.

Theme 3: Integration With Existing Coping Strategies

The moments when we truly felt the tension dissolve rarely lined up with finishing a MINT exercise. More often, relief arrived through small, familiar things we reached for on our own, like going for a run (RP1), walking the dog in the evening light (RP3), turning the car into a private concert hall and singing at the top of our lungs while stuck in traffic (RP2), curling up with a novel for 10 quiet pages (RP4), sketching, or simply loading a favorite video game. These strategies also shared another characteristic: they worked quickly and intuitively. RP1 noted, "I already know what works for me … running gives me the endorphins, loud rhythmic music settles my breathing, a few pages of fiction melt the noise in my head… it just does." Even when MINT exercises were tailored to our VARK preferences, they often felt slower or more effortful compared with strategies we had already internalized. We noticed ourselves comparing the 2: the structured exercises vs our own spontaneous coping tactics. The comparison was stark: while the exercises were often interesting or novel, they did not always provide the immediate, visceral relief that our own strategies did. Over time, we started to integrate MINT exercises selectively, favoring moments when our usual strategies were unavailable or when we wanted to experiment with a different approach. This realization prompted deeper reflection on how we individually manage stress and how digital interventions fit into those practices. It became clear that while MINT offered structured, evidence-based exercises, the most meaningful support often came from us and the routines we had developed over years of trial and error. Over the weeks, our engagement with the bot shifted from following instructions to observing, reflecting, and occasionally experimenting, blending our spontaneous coping strategies with its prompts when it seemed useful.

Theme 4: The Meaning of Nonengagement

One pattern became impossible to ignore over the 4 weeks. Our engagement, or lack of it, was closely tied to moments of stress. The periods when we did not interact with MINT often turned out to be the most informative. On low-stress days, we were conscientious, completing EMAs, logging diary entries, and experimenting with exercises. Yet, when stress peaked, our interactions dropped sharply. RP2 described this clearly: "My engagement pattern made me think of missing not at random [MNAR] … the classic example to understand MNAR is the alcoholic who doesn't fill in the questionnaire when he drinks. Here, when I was really overwhelmed, I simply didn't record anything." This was true for all of us: periods of silence were rarely accidental; they were protective withdrawals from additional effort or exposure to stress. For instance, one of us (RP3) missed every EMA on the day of a major work crisis, while another (RP2) did not engage with interventions and left diary entries blank for 3 consecutive days, later recalling them as "the worst [part] of the entire month." In these moments, engaging with MINT felt like adding friction rather than offering relief. As we revisited our diaries, EMAs, and workshop discussions, the blank spaces themselves became meaningful signals. They came to represent moments when the support we most needed was, paradoxically, the hardest to access. We began to treat missing entries as informative, signaling not a failure of adherence but the contours of our emotional landscape. Interestingly, these gaps also shaped how we thought about the system's design. We realized that MINT had no way to detect absence or adapt when we withdrew. In contrast to times when engagement was high, peak stress moments were invisible to the system, and its push notifications often went unanswered, further emphasizing the mismatch between system design and lived reality. By the end of the study, we collectively understood that these silences were not just "missing data" in a technical sense; they were data in themselves, communicating both our states of mind and our coping capacity. Recognizing this made us reflect differently on the metrics of success for digital interventions: completion rates alone were insufficient. The pattern of nonengagement became a story in itself, narrating where MINT could not reach us and where the system might have needed to adapt.

Theme 5: Reminder Dependence

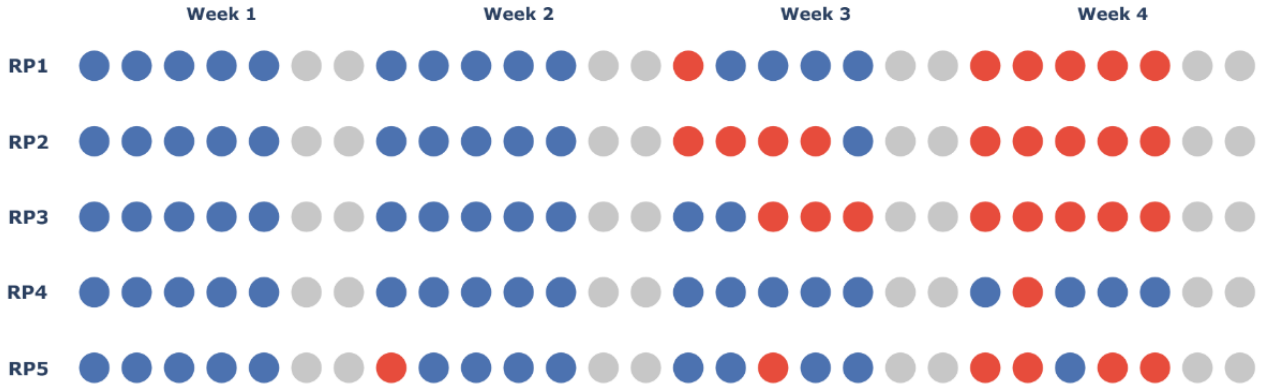
Week 4 was revealing for us in a way we hadn't anticipated. We made a provocative and deliberate choice: at the start of the week we switched off every scheduled notification overnight. No gentle tapering, no transitional phase, just silence. In real-world deployments, this would be unusual and potentially harmful, but for us, it was intentional. We wanted to force the question: once the external scaffolding disappears, what remains? The answer came within hours and could not have been clearer.

Within 2 days, 4 of the 5 participants had essentially abandoned the bot (Figure 3). As RP2 summed it up in the second workshop, "Without reminders the intervention crashed… we all crashed, except you RP4." RP4 was the quiet exception. She continued to use controlled breathing and deliberate walking at genuine moments of tension. The techniques had become hers; the bot had successfully faded away. For the rest of us, when the prompts vanished, so did any impulse to seek the bot out. The radical week 4 switch acted like a stress test that exposed a truth we would otherwise have discovered only after months of gradual disengagement: the first 3 weeks had trained dependence, not independence. We had inadvertently built a system that functioned like a daily pill, something you take because it arrives, not because you have learned to live differently. Curiosity and our duty as researchers had kept us compliant, but MINT had never become a companion in genuine behavior change. This realization hit hard. Microintervention systems are often conceived as short-term analgesics: deliver evidence-based techniques, reduce symptoms, repeat. But week 4 showed that symptom relief without skill transfer is little more than digital pharmacology: effective only as long as the external dose keeps coming. True behavior change companions, by contrast, must do the opposite of a pill: they must work themselves out of a job by helping the person internalize new patterns until the tool itself becomes optional. RP2 articulated how we should have used those crucial early days instead: "The week should be a protected onboarding and alliance-building

XSL•FO
RenderX

phase. The bot should be deliberately present, even chatty, but with one clear message: 'For the next week I'll be around a lot, not to give you homework, but so we get to know each other and so you learn exactly how to use me when you actually need me.' It teaches transparently: how stress works in the body, why breathing is an instant reset, why movement is medicine. It gives three or four one-word commands ('BREATHE,' 'WALK,' 'PAUSE,' 'VENT'). Every evening it asks one gentle question: 'One thing that helped today?' or 'Anything you want me to remember?' By the end of that week the user trusts the bot, feels seen, and knows precisely how to call for help. Only then do you start fading the prompts." The abrupt week 4 silence was therefore not a flaw; it was the perfect diagnostic. By removing the external "dose" all at once, it revealed that we had built a pharmacological tool rather than a behavior change companion. It proved that genuine autonomy is cultivated from day 1 through transparency, psychoeducation, felt understanding, and explicit pathways for self-initiation. In other words, our prototype taught us how to design another version of it that, like the best therapist or coach, eventually makes itself proudly obsolete.

Figure 3. Participant timeline showing daily engagement patterns (completed microinterventions) with the MINT (Micro-Interventions for Navigating Tension) chatbot. Red indicates that the RP did not engage with MINT, whereas blue indicates that the RP engaged with MINT. RP: researcher-participant.



Theme 6: Everyday Technologies as the Benchmark

Finally, 3 of us (RP2, RP3, and RP4) wore wearable devices to capture physiological signals during the study. In practice, the data provided by those devices proved far less useful than we had hoped. Signals were noisy and frequently mismatched with how we actually felt. As RP2 recalled a frightening moment on the road, "A car almost hit me on a bend… I expected a massive peak [in EDA]… instead I saw nothing at all." RP4 regularly experienced the reverse: "I felt calm and it said I was super stressed…" Practical difficulties added to the disappointment. Two of us (RP1 and RP4) developed skin irritation within the first week, even with the medical-grade device, and daily charging became another small but persistent hassle. By the second workshop, we agreed that, for the moment, wearable-derived stress metrics remain too inconsistent and burdensome to serve as a reliable foundation for everyday intervention. They generated interesting patterns for post hoc reflection but could not be trusted in real time. Wearables, however, were not the only benchmarks that shaped how we experienced MINT. An even more unexpected comparison emerged from our own everyday habits. We ourselves, while testing MINT, began turning to general-purpose LLMs for the emotional support that the prototype was not providing. As RP4 admitted with a half-smile, "I write everything I'm feeling on ChatGPT… it helps just to put the words down… it's my free psychologist … well, cheap psychology anyway. [laugh]" The joke landed because it was true for several of us (RP2, RP3, and RP4). During the same 4 weeks we were dutifully completing MINT's EMAs and exercises, we were also quietly opening ChatGPT (OpenAI) to vent after a hard day, draft angry messages more calmly, externalize racing thoughts, or simply to feel heard when the structured bot stayed silent. These parallel conversations were never part of the protocol; they happened spontaneously, often late at night or in moments of peak frustration. Some exchanges were brief check-ins, others long, meandering threads that continued over days. We noticed the same patterns: the model remembered what we had said before, asked gentle follow-up questions, and responded with a warmth that, however simulated, felt caring. A few of us also observed that, at times, chatting with an LLM could have downsides, like enabling prolonged rumination rather than gently interrupting it. Instead of recognizing the loop and helping the person close or rework the situation, the model simply kept the conversation going. We had deliberately designed MINT to be short, focused, and emotionally neutral in order to respect time and clinical rigor. In practice, we found ourselves seeking out something more verbose, curious, and unmistakably human-like. Current openly available LLMs, tools never intended for therapeutic use, quietly showed us the level of continuity, attentiveness, and conversational space that users now expect from any digital companion. The unreliable wearables and our own unsolicited use of LLMs reminded us that users (even when the users are the designers) do not evaluate a new intervention in isolation. We compare it, consciously or not, with everything else already in our hands.

Discussion

Principal Results and Comparison With Prior Work

Contribution of CAE

Digital mental health research has already identified many of the challenges discussed in this paper, including issues related

XSL•FO
RenderX

to timing, engagement, personalization, attrition, and the limitations of current sensing technologies. We do not present these themes as entirely new findings. Their value lies instead in the level of detail with which CAE allowed us to examine them from the inside: not only as design problems but also as experiences that unfolded, accumulated, and changed meaning across ordinary days. As anticipated in the *Introduction*, traditional trials, log analyses, and feasibility studies can show when engagement drops, which prompts are completed, and which features are preferred. CAE allowed us to ask a different question: what does it feel like when those patterns happen, and how do users make sense of them while they are living with the system? These forms of evidence are complementary rather than competing. In this discussion, we therefore use the literature as a point of orientation but focus on what the CAE perspective added to each theme: the mechanisms, tensions, and lived interpretations that are often flattened in aggregate engagement metrics.

Theme 1

The first example is about timing. JITAI research has already shown that receptivity depends strongly on when support is delivered [36]. Poorly timed prompts can reduce engagement, while better-timed ones can make digital support feel more relevant. What our CAE made visible, however, was the deeper meaning of "poor timing." A mistimed prompt was not simply inconvenient. At times, it interrupted a moment participants had actively created to recover from stress. This matters because the system was not only competing with work, distraction, or busyness; it was also sometimes competing with users' own successful attempts at restoration.

This also aligns with the occupational psychology literature on the importance of psychological detachment and psychological restoration, which underscores the importance of preserving personal downtime for the recovery of depleted cognitive and emotional resources [37,38]. From this perspective, the design issue is not just interruption management; it is the risk that an intervention aimed at supporting regulation may disturb an ongoing act of self-regulation.

The CAE also helped us distinguish need from receptivity. In everyday use, higher stress did not automatically translate into greater openness to support. Some exercises required a degree of emotional availability that was absent precisely when stress was high. Others, especially lower-effort regulation strategies such as breathing, felt easier to access across a wider range of states. This points to a narrow and unstable window: enough distress for support to matter, but enough attentional and emotional bandwidth to engage with it. Current wearable approaches often equate physiological arousal with need [39-41], but our experience suggests an inverted-U pattern: too little arousal yields little or no benefit, whereas excessive arousal triggers protective withdrawal; this underscores why linear "more arousal equals more need" algorithms risk delivering interventions precisely when users are least able or willing to receive them.

Theme 2

Findings on the perceived lack of reciprocity and relational cues instead extend previous work showing that relational agents support sustained engagement [42,43]. Our experience suggests that the issue goes beyond conversational style. Throughout the 4 weeks, participants rarely questioned the quality of the exercises themselves. Instead, frustration accumulated because the effort invested in the interaction seemed to disappear into the system. Logging moods, completing EMAs, and recording reflections gradually came to feel like one-way transactions, with little indication that MINT was doing anything with that information. Seen in this light, reciprocity is not simply about making a chatbot warmer or more socially engaging. It is about making users feel that their contributions have consequences. Even small signals (eg, a reference to a previous interaction, a brief acknowledgment, or a simple synthesis of recurring patterns) could have changed how participants interpreted the same interaction. The absence of these cues shifted the perceived balance between effort and value, until the exercises themselves began to feel increasingly transactional. This resonates with previous work showing that users disengage when conversational agents fail to establish a sense of relationship or continuity [44,45]. Reading back the workshops, what stood out was that we rarely asked for a more empathic chatbot. We asked for signs that the chatbot was listening. Those are related ideas, but they are not the same. The issue was less the absence of warmth than the absence of evidence that previous interactions had shaped subsequent ones.

Theme 3

Theme 3, on the other hand, resonates with longitudinal studies showing that people rarely adopt the full repertoire of techniques offered by mindfulness or CBT apps, instead selecting only those that fit their everyday lives [46,47]. Our experience suggests that this selectivity is not simply a matter of personal preference. Throughout the study, participants seemed to evaluate each proposed exercise against an already existing repertoire of coping strategies that had been built through years of experience. The decision was rarely "Is this exercise good?" It was closer to "Is this more useful than what I would naturally do right now?" Familiar activities such as walking, running, reading, or listening to music were not necessarily perceived as more evidence-based or therapeutically sophisticated. They simply required little deliberation, carried no learning cost, and already belonged to participants' habits. Structured microinterventions, therefore, entered an environment where effective coping strategies were already in place, rather than an empty space waiting to be filled. This perspective is consistent with the person-based approach to intervention development [48], which emphasizes understanding how interventions fit within people's existing lives rather than expecting people to adapt to the intervention. Our findings reinforce this principle by suggesting that future systems should explicitly discover and build upon users' own coping repertoires from the outset. This does not reduce the value of evidence-based exercises. Instead, it positions them as additions to an existing repertoire, introduced when they genuinely expand what users can already do rather than asking them to replace or take the place of strategies that already work.

XSL•FO
RenderX

Theme 4

Nonengagement during periods of elevated stress is a phenomenon that has already been described as a form of MNAR data in digital mental health research literature [49], and similar patterns have been reported in studies of depression and bipolar disorder, where drop out often coincides with symptom worsening [50,51]. Even in this case, our experience did not challenge this interpretation. It added something more specific: the opportunity to understand what those missing moments actually represented from the participant's perspective. Looking back together at diaries, workshop discussions, and EMA records, we found that periods of nonengagement rarely appeared as simple lapses in adherence. They reflected moments in which interacting with the system itself had become one demand too many. This perspective echoes recent calls to treat nonuse as meaningful data rather than simply as adherence failure [49,52,53]. Our findings suggest that the next step is not only to detect these patterns computationally but also to understand the circumstances that produce them. From a design perspective, prolonged silence should perhaps be interpreted less as disengagement from the intervention itself and more as a potential indicator that the demands of engaging with the intervention have temporarily exceeded the user's available capacity.

Theme 5

Similarly, the sharp decline in self-initiated use after scheduled reminders were removed reflected a challenge that is increasingly being documented across many DMHIs. Longitudinal evaluations of commercial apps consistently report high initial engagement followed by substantial attrition over time [54,55]. Our experience suggests that this transition deserves attention not only as an engagement problem but also as a question of learning (in the sense of internalizing intervention strategies). What remains once the external structure disappears? Considering this theme alongside theme 3 offers 1 possible interpretation. Throughout the study, participants continued to rely on coping strategies that had become part of their everyday routines through years of personal experience. By contrast, most of the chatbot exercises remained closely associated with the chatbot itself. They were completed when prompted, but they rarely became something participants reached for independently. RP4 was the only participant who continued using exercises after the reminders had disappeared. In her case, the strategies had become detached from the chatbot that had introduced them and had entered her everyday coping repertoire. This distinction may be more important than continued use of the chatbot itself. A successful intervention is not necessarily one that people keep opening indefinitely but one whose underlying strategy gradually becomes part of their own repertoire. When that transfer does not occur, engagement risks remaining dependent on the external structure that originally sustained it.

Theme 6

Finally, theme 6 broadens the lens beyond the chatbot itself. As we reflected on our experience, it became increasingly clear that we were not evaluating MINT against an abstract idea of a digital intervention. We were comparing it, often implicitly, with technologies that had already become part of everyday life and had quietly reshaped our expectations. General-purpose LLMs were the clearest example of this shift. We did not expect MINT to become an open-ended conversational agent, nor did we see that as its purpose. Instead, interacting daily with LLMs had changed what a conversational system was expected to do. When read alongside theme 2, the issue was not simply conversational warmth. It was continuity. We expected signs that previous interactions still mattered: that the chatbot remembered, connected, and built on what had already been shared. As these behaviors become commonplace in everyday AI systems, their absence becomes more noticeable even in highly structured digital health interventions. Recent work is beginning to document similar shifts in users' expectations of specialized mental health tools [56-58].

Just as LLMs have changed our expectations of conversation, the increasing presence of wearable devices in our lives has also shaped our expectations of physiological sensing. Simply wearing a device that continuously monitors physiological signals creates the expectation that the system will recognize meaningful changes in one's state. Across our reflections, the critical issue was not occasional inaccuracies themselves but what happens after repeated mismatches between sensed data and lived experience. Those moments gradually weaken confidence that the system can recognize when support was actually needed. Previous work has extensively discussed the technical challenges of using consumer wearables for stress detection in everyday settings [59-61]. Our experience suggests an additional consequence that is less often examined: repeated mismatches may gradually erode users' willingness to rely on physiological sensing as the basis for intervention delivery. From the user's perspective, sensing is not evaluated signal by signal but through the accumulation of moments in which physiological measures either correspond or fail to correspond with lived experience.

Design Implications and Directions for Future Research

Designing DMHIs often begins with deciding what support to provide. Our experience suggests that just as much attention should be given to what happens after that support is introduced into everyday life. Many of the design implications that emerged from this study concerned not the content of the interventions themselves but the conditions that shaped how they were experienced and, over time, whether they became part of everyday routines. To translate these observations into practical guidance for both the further development of MINT and the design of future digital stress-management interventions, Table 3 synthesizes the key design challenges emerging from our findings together with their corresponding design implications.

XSL•FO
RenderX

Table . Design implications for digital microintervention systems for stress management derived from the 6 collaborative autoethnography (CAE) themes.

Design focus	Key empirical insight	Design implication
Receptivity-aware intervention delivery	Theme 1	Develop receptivity-aware triggering that goes beyond physiological arousal, estimating not only when support is needed but also when users are likely to have sufficient emotional and attentional capacity to engage with it.
Relational continuity and reciprocity	Theme 2	Treat continuity across interactions (eg, remembering previous inputs and acknowledging earlier exchanges) as a core design feature rather than a conversational enhancement.
Integration with existing coping strategies	Theme 3	Design microinterventions to extend users' existing coping repertoires rather than compete with or replace them.
Accounting for nonengagement	Theme 4	Redefine nonengagement patterns as meaningful signals (MNAR[a]) that should inform adaptation and evaluation, rather than interpreting them only as adherence failure.
Transition to autonomous use	Theme 5	Design explicit pathways through which users gradually move from prompted engagement to self-initiated coping, allowing reminders to fade as strategies become part of everyday life.
Appropriate use of wearable sensing	Theme 6	Use wearable sensing primarily to enrich contextual understanding rather than as the sole trigger for intervention delivery in everyday settings.
Balancing everyday AI expectations with evidence-based interventions	Theme 6	Recognize that users increasingly expect continuity across interactions and explore architectures that can provide it while remaining clinically appropriate.

[a]MNAR: missing not at random.

Several questions naturally follow from these observations. One concerns receptivity: future work should explore how multimodal sensing and lightweight contextual information can better estimate when support is both needed and welcome. Another concerns learning over time. Combining microrandomized trials with longer qualitative follow-up could help clarify how people move from externally prompted engagement to genuinely self-initiated use. Finally, the widespread adoption of everyday generative AI raises new design questions. Hybrid scripted-generative architectures [62,63] offer one promising direction for exploring how conversational continuity can be achieved while preserving the structure and safety of evidence-based interventions.

Limitations and Role of CAE in Digital Health Research

Before concluding, it is important to acknowledge several limitations of this study. First, our RPs were all digital mental health researchers, and this inevitably shaped both the usage of the prototype and the depth of interpretations. Compared with the broader population, we were likely more attentive to design decisions, more accustomed to reflecting on our own engagement, and more familiar with stress-management strategies. The study also lacked, by definition, cultural, socioeconomic, and clinical diversity, and RP2's dual role as a prototype developer may have influenced parts of the collective sense-making despite our explicit attention to positionality throughout the study. The research lasted 4 weeks, longer than most feasibility pilots but still short relative to the development of longer-term habits. Importantly, because our analysis of nonengagement (theme 4) relies on retrospective self-reports and collaborative reflections, the study design cannot conclusively determine the causes of nonuse or distinguish whether missed interactions were driven by emotional overload or by competing daily demands, participant burden, and reduced motivation over time. These observations should therefore be understood as subjective participant interpretations rather than empirically established causal claims. In addition, MINT relied on scheduled rather than stress-contingent intervention delivery. This was an intentional design choice, as our goal was to understand how scheduled microinterventions are experienced in everyday life rather than to evaluate JITAI algorithms. Likewise, the abrupt removal of reminders during the final week was designed to probe self-initiated engagement and whether participants had integrated the strategies into their everyday coping repertoire, not to simulate a realistic deployment strategy. As a result, the observed decline in engagement cannot be attributed solely to reminder dependence but should instead be interpreted as highlighting the need to better understand how users transition from externally prompted support to independent use, including within more context-aware intervention designs. At the same time, this study also illustrates what CAE can contribute during the early stages of designing digital health interventions. Several of the insights discussed in this paper did

XSL•FO
RenderX

not emerge from single interactions but from returning to the same experiences over time and discovering that they acquired a different meaning once considered alongside later events. What initially felt like isolated moments of frustration, comfort, or disengagement gradually became interpretable as recurring patterns. This depth of interpretation depended not only on prolonged engagement with the system but also on the collective sense-making that sits at the heart of CAE. Revisiting experiences together allowed us to question initial interpretations, surface alternative readings, and gradually build richer explanations than any one perspective could have produced alone. Perhaps the most valuable aspect of this process was not that it generated more data but that it gave us repeated opportunities to elaborate on what those data meant. For us, this reinforces the role of CAE as a complementary approach within the broader evaluation of digital health interventions, particularly during the early stages of design, where understanding how design decisions are experienced can meaningfully inform what is tested next.

Conclusions

This study used CAE to examine the everyday experience of using a stress-management chatbot over 4 weeks. Using MINT as part of our daily routines allowed us to look beyond individual microinterventions and consider how they were experienced over time: how expectations developed, how support was interpreted in different situations, and how the strategies introduced by the chatbot related to our existing ways of coping. The study also illustrates what CAE can contribute to the evaluation of digital health interventions. Returning to the same experiences over time and making sense of them collectively often enriched our initial interpretations. Experiences that initially seemed unrelated acquired a different meaning when revisited alongside later events, revealing patterns that were difficult to recognize in the moment. Instead of introducing entirely new design challenges, this process helped us develop richer interpretations of familiar ones, moving beyond description toward explanations and hypotheses that can inform subsequent stages of intervention design and evaluation. We see this as complementing more established evaluation approaches, particularly during the formative stages of design. More broadly, this experience reminded us that design features rarely have a single, fixed meaning. The same feature can be reassuring, frustrating, motivating, or simply irrelevant depending on when it is encountered, what has happened before, and how it fits within the broader context of everyday life.

Generative AI Use

ChatGPT (OpenAI) was used in the preparation of this manuscript to assist with English-language editing and phrasing. All scientific content, study design, data collection, analysis, interpretation, and final decisions remain the sole responsibility of the human authors.

Funding

The authors declared no financial support was received for this work.

Data Availability

The qualitative data supporting the findings of this study (personal reflective diaries, shared reflexive log entries, workshop transcripts, and free-text responses from ecological momentary assessments) contain sensitive personal reflections. Therefore, the full raw data cannot be made publicly available. A curated subset of the dataset, with identifying details (eg, names, places, institutions, and other contextual identifiers) removed or masked where feasible, is available from the corresponding author upon reasonable request and subject to a data-sharing agreement.

Authors' Contributions

Conceptualization: MB, SAP, SG
Data curation: MB
Formal analysis: MB, SAP, MP, MM, SG
Investigation: MB, SAP, MP, MM, SG
Methodology: MB, SAP, SG
Project administration: MB
Software: MB
Supervision: SG
Visualization: MB
Writing – original draft: MB, SAP, SG
Writing – review & editing: MP, MM

Conflicts of Interest

None declared.

Multimedia Appendix 1
MINT (Micro-Interventions for Navigating Tension) study implementation files.

XSL•FO
RenderX

[DOCX File, 7 KB - humanfactors_v13i1e89876_app1.docx]

Multimedia Appendix 2
Reflexive thematic analysis outputs.
[DOCX File, 11 KB - humanfactors_v13i1e89876_app2.docx]

References

1. Lovallo WR. Stress and Health: Biological and Psychological Interactions, 3rd edition: SAGE Publications; 2016. [doi: 10.4135/9781071801390]
2. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR mHealth uHealth 2015 Aug 19;3(3):e82. [doi: 10.2196/mhealth.4328] [Medline: 26290327]
3. Milne-Ives M, de Cock C, Lim E, et al. The effectiveness of artificial intelligence conversational agents in health care: systematic review. J Med Internet Res 2020 Oct 22;22(10):e20346. [doi: 10.2196/20346] [Medline: 33090118]
4. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39. [doi: 10.1348/135910709X466063] [Medline: 19646331]
5. Hickey BA, Chalmers T, Newton P, et al. Smart devices and wearable technologies to detect and monitor mental health conditions and stress: a systematic review. Sensors (Basel) 2021 May 16;21(10):3461. [doi: 10.3390/s21103461] [Medline: 34065620]
6. Pogrebtsova E, Craig J, Chris A, O'Shea D, González-Morales MG. Exploring daily affective changes in university students with a mindful positive reappraisal intervention: a daily diary randomized controlled trial. Stress Health 2018 Feb;34(1):46-58. [doi: 10.1002/smi.2759] [Medline: 28512859]
7. Feldman G, Greeson J, Senville J. Differential effects of mindful breathing, progressive muscle relaxation, and loving-kindness meditation on decentering and negative reactions to repetitive thoughts. Behav Res Ther 2010 Oct;48(10):1002-1011. [doi: 10.1016/j.brat.2010.06.006] [Medline: 20633873]
8. Siedlecka E, Denson TF. Experimental methods for inducing basic emotions: a qualitative review. Emotion Rev 2019 Jan;11(1):87-97. [doi: 10.1177/1754073917749016]
9. Johnson JA, Zawadzki MJ, Materia FT, White AC, Smyth JM. Efficacy and acceptability of digital stress management micro-interventions. Procedia Comput Sci 2022;206:45-55. [doi: 10.1016/j.procs.2022.09.084]
10. Suh J, Howe E, Lewis R, et al. Toward tailoring just-in-time adaptive intervention systems for workplace stress reduction: exploratory analysis of intervention implementation. JMIR Ment Health 2024 Sep 12;11:e48974. [doi: 10.2196/48974] [Medline: 39264703]
11. Greiner K, Müller E, Mädche A. Designing a smartwatch-based micro-intervention to support students emotion regulation. Proc Mensch Comput 2024 2024:432-436. [doi: 10.1145/3670653.3677520]
12. Oba T, Takano K, Sugawara D, Kimura K. Just-in-time delivery of cognitive behavioral therapy-based exercises: single-case experimental design with random multiple baselines. JMIR Form Res 2025 Jul 24;9:e69556. [doi: 10.2196/69556] [Medline: 40705402]
13. Militello L, Sobolev M, Okeke F, Adler DA, Nahum-Shani I. Digital prompts to increase engagement with the Headspace app and for stress regulation among parents: feasibility study. JMIR Form Res 2022 Mar 21;6(3):e30606. [doi: 10.2196/30606] [Medline: 35311675]
14. Baumel A, Edan S, Kane JM. Is there a trial bias impacting user engagement with unguided e-mental health interventions? A systematic comparison of published reports and real-world usage of the same programs. Transl Behav Med 2019 Nov 25;9(6):1020-1033. [doi: 10.1093/tbm/ibz147] [Medline: 31689344]
15. Kilfoy A, Hsu TCC, Stockton-Powdrell C, Whelan P, Chu CH, Jibb L. An umbrella review on how digital health intervention co-design is conducted and described. NPJ Digit Med 2024 Dec 23;7(1):374. [doi: 10.1038/s41746-024-01385-1] [Medline: 39715947]
16. Wells I, Thelwell E, Giacco D. Guidance on how to involve people with lived experience in research on digital mental health interventions. Eur Psychiatr 2024 Apr;67(S1):S59-S60. [doi: 10.1192/j.eurpsy.2024.171]
17. Kaltenhauser A, Stefanidi E, Schöning J. Playing with perspectives and unveiling the autoethnographic kaleidoscope in HCI – a literature review of autoethnographies. Proc 2024 CHI Conf Hum Factors Comput Syst 2024 May 11:1-20. [doi: 10.1145/3613904.3642355]
18. Senabre Hidalgo E, Greshake Tzovaras B. "One button in my pocket instead of the smartphone": a methodological assemblage connecting self-research and autoethnography in a digital disengagement study. Methodol Innov 2023 Jul;16(2):149-162. [doi: 10.1177/20597991231161093]
19. Biswas B, Dooley E, Coulthard E, Roudaut A. Autoethnography of living with a sleep robot. Multimodal Technol Interact 2024 Jun 18;8(6):53. [doi: 10.3390/mti8060053]
20. Neupane S, Dongre P, Gracanin D, Kumar S. Wearable meets LLM for stress management: a duoethnographic study integrating wearable-triggered stressors and LLM chatbots for personalized interventions. Ext Abstr CHI Conf Hum Factors Comput Syst 2025:1-8. [doi: 10.1145/3706599.3720197]

XSL•FO
RenderX

21. Chang H, Ngunjiri FW, Hernandez KAC. Collaborative Autoethnography, 1st edition: Left Coast Press; 2013. [doi: 10.4324/9781315432137]
22. Moore G, Galway L, Donnelly M. Remember to smile: design of a mobile affective technology to help promote individual happiness through smiling. Proc EAI Int Conf Pervasive Comput Technol Healthc 2017 May 23:348-354. [doi: 10.1145/3154862.3154936]
23. Renner F, Schwarz P, Peters ML, Huibers MJH. Effects of a best-possible-self mental imagery exercise on mood and dysfunctional attitudes. Psychiatry Res 2014 Jan 30;215(1):105-110. [doi: 10.1016/j.psychres.2013.10.033] [Medline: 24252218]
24. Fleming ND, Mills C. Not another inventory, rather a catalyst for reflection. To Improve Acad 1992;11(1):137-155. [doi: 10.3998/tia.17063888.0011.014]
25. Rammstedt B, Kemper CJ, Klein MC, Beierlein C, Kovaleva A. A short scale for assessing the Big Five dimensions of personality: 10 item Big Five Inventory (BFI-10). Methods Data Anal 2013;7(2):233-249. [doi: 10.12758/MDA.2013.013]
26. Carver CS. You want to measure coping but your protocol's too long: consider the brief COPE. Int J Behav Med 1997;4(1):92-100. [doi: 10.1207/s15327558ijbm0401_6] [Medline: 16250744]
27. McGuiness CE, Zajac I, Wilson C, Turnbull D. The REIm-13: a brief measure of thinking style. Eur J Psychol Assess 2019;35(4):451-458. [doi: 10.1027/1015-5759/a000421]
28. The VARK questionnaire (version 8.01). : VARK-Learn Limited; 2019 URL: https://www.coalservices.com.au/wp-content/uploads/2023/05/VARK-Questionnaire-1.pdf [accessed 2026-10-15]
29. Trettin B, Agerskov H, Lehmkuhl L, Rothmann MJ, Jensen CM, Nielsen C. Workshops as a research method in health science: epistemological and methodological perspectives. J Adv Nurs 2026 Apr;82(4):3797-3811. [doi: 10.1111/jan.17050] [Medline: 40371990]
30. LEGGE 11 gennaio 2018, n. 3: delega al governo in materia di sperimentazione clinica di medicinali nonche' disposizioni per il riordino delle professioni sanitarie e per la dirigenza sanitaria del Ministero della salute [Article in Italian]. Gazzetta Ufficiale della Repubblica Italiana. 2018 Jan 31. URL: https://www.gazzettaufficiale.it/eli/id/2018/1/31/18G00019/sg [accessed 2026-10-14]
31. EmbracePlus. Empatica. URL: https://www.empatica.com/en-eu/embraceplus/ [accessed 2025-12-02]
32. Galaxy Watch4 Bluetooth (40mm) [Article in Italian]. Samsung. 2021. URL: https://www.samsung.com/it/watches/galaxy-watch/galaxy-watch4-black-bluetooth-sm-r860nzkaitv/buy/ [accessed 2023-11-07]
33. Apple Watch SE (2a generazione) - Specifiche tecniche [Article in Italian]. Apple. URL: https://support.apple.com/it-it/111853 [accessed 2025-12-02]
34. Campbell K, Orr E, Durepos P, et al. Reflexive thematic analysis for applied qualitative health research. Qual Rep 2021 Jun 20;26(6):2011-2028. [doi: 10.46743/2160-3715/2021.5010]
35. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
36. Borghouts J, Eikey E, Mark G, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387. [doi: 10.2196/24387] [Medline: 33759801]
37. Sonnentag S, Fritz C. The Recovery Experience Questionnaire: development and validation of a measure for assessing recuperation and unwinding from work. J Occup Health Psychol 2007 Jul;12(3):204-221. [doi: 10.1037/1076-8998.12.3.204] [Medline: 17638488]
38. Chen Y, Pan L, Lu F, Sun D, Liao C, Na M. Psychological detachment in Chinese higher education: a multitheoretical model of academic stress, cultural pressure, and coping resources. Front Psychol 2025;16:1647184. [doi: 10.3389/fpsyg.2025.1647184] [Medline: 41164721]
39. Leonard NR, Casarjian B, Fletcher RR, et al. Theoretically-based emotion regulation strategies using a mobile app and wearable sensor among homeless adolescent mothers: acceptability and feasibility study. JMIR Pediatr Parent 2018;1(1):e1. [doi: 10.2196/pediatrics.9037] [Medline: 30637376]
40. Schwerdtfeger AR, Tatschl JM, Rominger C. Effectiveness of 2 just-in-time adaptive interventions for reducing stress and stabilizing cardiac autonomic function: microrandomized trials. J Med Internet Res 2025 Aug 7;27:e69582. [doi: 10.2196/69582] [Medline: 40773285]
41. Sadeghi M, McDonald AD, Sasangohar F. Posttraumatic stress disorder hyperarousal event detection using smartwatch physiological and activity data. PLoS One 2022;17(5):e0267749. [doi: 10.1371/journal.pone.0267749] [Medline: 35584096]
42. Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. ACM Trans Comput-Hum Interact 2005 Jun;12(2):293-327. [doi: 10.1145/1067860.1067867]
43. Ly KH, Ly AM, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. Internet Interv 2017;10:39-46. [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]
44. Xu Z, Lee YC, Stasiak K, Warren J, Lottridge D. The digital therapeutic alliance with mental health chatbots: diary study and thematic analysis. JMIR Ment Health 2025 Oct 10;12:e76642. [doi: 10.2196/76642] [Medline: 41072011]
45. Croes EAJ, Antheunis ML. Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. J Soc Pers Relat 2021;38(1):279-300. [doi: 10.1177/0265407520959463]

XSL•FO
RenderX

46. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
47. Stawarz K, Preist C, Tallon D, Wiles N, Coyle D. User experience of cognitive behavioral therapy apps for depression: an analysis of app functionality and user reviews. J Med Internet Res 2018 Jun 6;20(6):e10120. [doi: 10.2196/10120] [Medline: 29875087]
48. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30. [doi: 10.2196/jmir.4055] [Medline: 25639757]
49. Goldberg SB, Bolt DM, Davidson RJ. Data missing not at random in mobile health research: assessment of the problem and a case for sensitivity analyses. J Med Internet Res 2021 Jun 15;23(6):e26749. [doi: 10.2196/26749] [Medline: 34128810]
50. Karyotaki E, Kleiboer A, Smit F, et al. Predictors of treatment dropout in self-guided web-based interventions for depression: an "individual patient data" meta-analysis. Psychol Med 2015 Oct;45(13):2717-2726. [doi: 10.1017/S0033291715000665] [Medline: 25881626]
51. Halabi R, Mulsant BH, Alda M, et al. Not missing at random: missing data are associated with clinical status and trajectories in an electronic monitoring longitudinal study of bipolar disorder. J Psychiatr Res 2024 Jun;174:326-331. [doi: 10.1016/j.jpsychires.2024.04.036] [Medline: 38692162]
52. Markowski KL, Smith JA, Gauthier GR, Harcey SR. Patterns of missing data with ecological momentary assessment among people who use drugs: feasibility study using pilot study data. JMIR Form Res 2021 Sep 24;5(9):e31421. [doi: 10.2196/31421] [Medline: 34464327]
53. Cho YW, Chow SM, Li J, et al. Within- and between-individual compliance in mobile health: joint modeling approach to nonrandom missingness in an intensive longitudinal observational study. JMIR mHealth uHealth 2025 Oct 30;13:e65350. [doi: 10.2196/65350] [Medline: 41167252]
54. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567. [doi: 10.2196/14567] [Medline: 31573916]
55. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
56. Siddals S, Torous J, Coxon A. "It happened to be the perfect thing": experiences of generative AI chatbots for mental health. Npj Ment Health Res 2024 Oct 27;3(1):48. [doi: 10.1038/s44184-024-00097-4] [Medline: 39465310]
57. Kang B, Hong M. Development and evaluation of a mental health chatbot using ChatGPT 4.0: mixed methods user experience study with Korean users. JMIR Med Inform 2025 Jan 3;13:e63538. [doi: 10.2196/63538] [Medline: 39752663]
58. Rządeczka M, Sterna A, Stolińska J, Kaczyńska P, Moskalewicz M. The efficacy of conversational AI in rectifying the theory-of-mind and autonomy biases: comparative analysis. JMIR Ment Health 2025 Feb 7;12:e64396. [doi: 10.2196/64396] [Medline: 39919295]
59. Shen S, Qi W, Zeng J, et al. Passive sensing for mental health monitoring using machine learning with wearables and smartphones: scoping review. J Med Internet Res 2025 Aug 14;27:e77066. [doi: 10.2196/77066] [Medline: 40811794]
60. Pinge A, Gad V, Jaisighani D, Ghosh S, Sen S. Detection and monitoring of stress using wearables: a systematic review. Front Comput Sci 2024 Dec 18;6:1478851. [doi: 10.3389/fcomp.2024.1478851]
61. Bolpagni M, Pardini S, Dianti M, Gabrielli S. Personalized stress detection using biosignals from wearables: a scoping review. Sensors (Basel) 2024 May 18;24(10):3221. [doi: 10.3390/s24103221] [Medline: 38794074]
62. Sanna L, Bolpagni M, Fietta V, et al. LLM-enriched finite-state chatbots for mental health support: a case study on self-help+. In: Bellazzi R, Juarez Herrero JM, Sacchi L, Zupan B, editors. Artificial Intelligence in Medicine: 23rd International Conference, AIME 2025, Pavia, Italy, June 23–26, 2025, Proceedings, Part I: Springer; 2025:386-396. [doi: 10.1007/978-3-031-95838-0_38]
63. Lenh PPC, Huan TP, Phuong NDT, Luan NT, Anh LT. A hybrid dialogue system for student mental health assessment and support. In: Nghia PT, Thai VD, Thuy NT, Huynh VN, Van Huan N, editors. Advances in Information and Communication Technology Proceedings of the 3rd International Conference ICTA 2024: Springer; 2025:816-826. [doi: 10.1007/978-3-031-80943-9_86]

Abbreviations

BFI-10: Big Five Inventory-10
CAE: collaborative autoethnography
CBT: cognitive behavioral therapy
COPE-28: Coping Orientation to Problems Experienced-28
DMHI: digital mental health intervention
EDA: electrodermal activity
EMA: Ecological Momentary Assessment

XSL•FO
RenderX

EMI: ecological momentary intervention
HRV: heart rate variability
JITAI: just-in-time adaptive intervention
LLM: large language model
MINT: Micro-Interventions for Navigating Tension
MNAR: missing not at random
REIm-13: Rational Experiential Inventory-13
RP: researcher-participant
RTA: reflexive thematic analysis
VARK: visual, aural, read/write, kinesthetic

Edited by A Bamgboje-Ayodele; submitted 18.Dec.2025; peer-reviewed by S Neupane; revised version received 03.Sep.2026; accepted 15.Sep.2026; published 27.Oct.2026.

Please cite as:
Bolpagni M, Piffer SA, Pozza M, Maminimini M, Gabrielli S
Insights on Designing the MINT (Micro-Interventions for Navigating Tension) Chatbot System for Stress Management: Collaborative Autoethnography
JMIR Hum Factors 2026;13:e89876
URL: https://humanfactors.jmir.org/2026/1/e89876
doi:10.2196/89876

© Marco Bolpagni, Sofia Asteria Piffer, Marco Pozza, Munesu Maminimini, Silvia Gabrielli. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 27.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Patient and Designer Collaboration in Cocreating Technical Innovations in a Hospital-Based Makerspace: Qualitative Interview Study

Jiske van der Giessen[1], MSc; Onno Helder[2], PhD; Thijs van Houwelingen[3], PhD

[1]Nursing Sciences, Program in Clinical Health Sciences, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

[2]Department of Create4Care, Erasmus MC University Medical Centre, Rotterdam, The Netherlands

[3]Research Group Technology Use in Health and Social Care, Research Centre for Healthy and Sustainable Living, University of Applied Sciences Utrecht, Heidelberglaan 7, Utrecht, The Netherlands

Corresponding Author:
Thijs van Houwelingen, PhD
Research Group Technology Use in Health and Social Care, Research Centre for Healthy and Sustainable Living, University of Applied Sciences Utrecht, Heidelberglaan 7, Utrecht, The Netherlands

Abstract

Background: Hospital-based makerspaces have emerged as collaborative environments for technical innovation, where designers cocreate with health care professionals and patients to resolve specific problems experienced in practice. As end users, patients can offer unique insights that could drive the development of patient-centered health care services or research. However, cocreation with patients within a makerspace environment is still rare—representing a missed opportunity to use their insights to develop innovations that meet their needs. Patient-designer interactions are underaddressed in the literature, and a deeper understanding could enhance their effectiveness.

Objective: The study aimed to explore the experiences of designers and patients (or their carers) cocreating in an academic hospital–based makerspace, using a bottom-up approach.

Methods: A generic qualitative study was conducted using semistructured individual interviews based on the Sunnybrook team–based competencies and design thinking methodology. A heterogeneous sample of 12 participants was recruited from a makerspace in an academic hospital in the Netherlands, comprising 6 with a designer's perspective and 6 with a patient's perspective. Most participants were involved in collaborative pairs, representing both perspectives within the same innovation project. Data were analyzed thematically.

Results: Four key themes emerged from the data: (1) dealing with the patient's situations, (2) integrating different perspectives, (3) feeling valuable and useful, and (4) dynamic interplay and engagement. Collaboration with patients in this hospital-based makerspace was perceived as valuable for innovation development, benefiting from diverse perspectives. However, the degree of patient involvement varied, influenced by factors such as the patient's health status, reachability, and interest in innovation development.

Conclusions: This study highlights the complexity of developing technical innovations that involve hospitalized patients. It offers valuable insights into how both patients and designers generally viewed their collaboration in the makerspace positively. Practical recommendations for hospital-based makerspaces include respecting patient autonomy, involving patients early and consistently throughout all development phases, maintaining communication beyond the formal collaboration period, and actively seeking feedback.

(JMIR Hum Factors 2026;13:e85926) doi:10.2196/85926

KEYWORDS

patient involvement; cocreation; patient; innovation; qualitive study; nursing development; delivery of health care; communication

Introduction

Over the past decade, makerspaces in hospitals have contributed to a new movement in health care known as Health 4.0 [1-3]. This movement uses modern technology to improve the quality, accessibility, and cost-efficiency of health care [4,5]. Such improvement is needed to address the increasing health care demand and to enhance staff empowerment and the work environment in the face of nursing workforce shortages [6-8]. However, technical innovations designed for health care may fail to meet the needs of nurses and patients [9]. These end users themselves may need to develop innovations that align with their daily practice, and makerspaces can assist in this process.

XSL•FO
RenderX

A makerspace is a collaborative physical department equipped with prototyping facilities, including 3D printers and laser cutters [9]. Professionals such as industrial product designers, user experience designers, software engineers, and researchers are employed to facilitate innovation development. Hospital-based makerspaces aim to find suitable solutions to problems experienced in health care practice, such as poor ergonomics and time-consuming tasks. These problems are identified using a bottom-up approach, meaning that nurses and patients address their problems and cocreate with the makerspace staff to remedy the problem or develop innovations to solve them [10,11]. By recognizing patients or their carers as end users, involving them in a cocreation process offers valuable insights to address issues and develop innovations that could improve patient-centered health care [12-15].

For conceptual clarity, this study adopts the definition of cocreation proposed by Pearce et al [16], who identified 4 interrelated collaborative processes—coideation, co-design, coimplementation, and coevaluation—as the core components of cocreation in health interventions. These 4 processes align closely with the stages of design thinking: inspiration (problem identification), ideation (prototyping and testing), and implementation (diffusion of innovations). The design thinking approach is commonly used in makerspaces and provides a structured lens for examining stakeholder involvement [17].

In health care, cocreation approaches are widely promoted as a means of strengthening patient involvement, improving the relevance of interventions, and bridging the gap between knowledge generation and practice. A systematic review by Grindell et al [18] found that cocreation fosters shared understanding, responsibilities, and decision-making among stakeholders, and enables diverse types of knowledge, including patient experience, to shape innovation. Moreover, a qualitative study by Voorheis et al [19] shows that patient and public involvement does not merely inform isolated design choices but influences the broader direction and viability of an innovation, including the implementation and long-term relevance of innovations. Their study also highlights that participating in cocreation can be empowering for patients themselves. At the same time, the authors emphasize that without clear roles, sustained engagement, and supportive structures, much of this potential value remains underused. A scoping review found that patient involvement is often limited to the first 2 phases of innovation development [20,21]. Excluding patients and their carers from all phases represents a missed opportunity and could be due to a mismatch in perception of the usefulness of solutions or a lack of public awareness and guidance on how to effectively collaborate with patients in technological innovation development [22,23].

While concepts such as public and patient involvement and community-based participatory research offer valuable insights into patient involvement, they are structured around research-oriented participation models and projects focusing on the public, where larger groups of stakeholders are included. Therefore, they differ from the makerspace context, where innovation emerges through bottom-up, practice-driven problem-solving; patients are involved due to an intrinsic motivation to solve their experienced problems and are part of a smaller team within the hospital.

Although the importance of involving patients throughout innovation processes is repeatedly emphasized in the literature, empirical research examining how patients and designers actually experience cocreation within hospital-based makerspaces remains scarce. Most existing studies focus on how makerspaces function organizationally and their role, rather than on the human interactions that shape cocreation [2,3,20,24].

Understanding the collaborative experiences of collaborations between designers and patients (or their carers) is essential. Gaining insight into the barriers and facilitators of such collaboration can improve future cocreation efforts within hospital makerspaces. Therefore, this study aims to explore the experiences of designers and patients (or their carers) cocreating in an academic hospital-based makerspace, using a bottom-up approach.

Methods

Study Design

A generic qualitative study was performed using semistructured interviews and thematic analysis. This approach enabled an in-depth understanding of how designers and (carers of) patients perceived the cocreation and what could be learned from their experiences, rather than testing or producing a theory or explicating the essential structure of a single phenomenon [25]. The study was conducted between February 2024 and June 2024 and adhered to the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist (Checklist 1) [26].

Setting

The study was conducted at a makerspace with a professionally equipped prototype studio, housed within an academic hospital, including a large children's hospital in the Netherlands. The makerspace is financed through the hospital's internal funding. The staff comprises industrial product designers, user experience designers, electrotechnicians, a quality adviser, a manager, and a coordinator who is also an academic researcher. The staff collaborates with individuals who identify a problem, such as patients, to develop technical innovations that address the experienced problem. Cocreation is embedded throughout all phases of the design thinking process. Depending on the situation, patients either visit the makerspace studio or designers meet them at the hospital ward. This makerspace aims to deliver functional prototypes that comply with relevant regulatory requirements. It operates in accordance with the Medical Device Regulation and produces medical devices for in-house use under the hospital's quality management system, which includes preparing required documentation such as an innovation justification statement, a statement of requirements, and risk analyses. The makerspace has successfully produced over 30 innovative products for use in daily practice. The makerspace staff proactively disseminates their knowledge and products within the hospital and beyond.

Patients or carers typically approach the makerspace with a specific problem and cocreate with the makerspace staff to develop a solution. Examples of innovations include a silicone

XSL•FO
RenderX

cover designed to fit the rigid connection piece of a kidney drain, reducing discomfort during daily activities and sleep (Figure 1A); an easily accessible emergency bag to support rapid response when a child's tracheostomy tube becomes obstructed (Figure 1B); a pillbox enabling hospitalized patients to manage their own medication while allowing nurses to maintain oversight of medication administration (Figure 1C); a waterproof shower cape for patients with a Port-a-Cath for intravenous chemotherapy; and a bedside table that can be positioned over the bed to facilitate play, homework, or family meals.

Figure 1. Examples of technical innovations developed in cocreation with patients within the makerspace: (A) Silicone cover kidney drain, (B) Emergency bag tracheostomy tube, and (C) Pillbox for hospitalized patients.



Population and Recruitment

To gain insight into the perspectives of both designers and end users, a heterogeneous group of designers and patients was selected using purposive sampling [27]. The first perspective, referred to as the "designer" perspective, included makerspace staff such as industrial product designers and user experience designers. The second, the "patient" perspective, included individuals involved in cocreating innovations addressing problems experienced by patients. This group included patients and representatives such as informal caregivers (eg, parents in the case of young children) and professionals who supported patients during the cocreation process and were able to describe their experiences with patients from their perspective. Eligible participants were individuals who had participated in a collaborative project with the hospital's makerspace to develop an innovation addressing issues experienced by patients and who represented either the designer or patient perspective. Designers who had no direct interaction with patients during collaborative projects were excluded. Additionally, patients under the age of 18 years or those who were hospitalized and in a delicate condition at the time of the study were excluded, as they were considered unable to participate adequately in an interview or participation could impose an undue burden.

Eligible designers and patients (or their representatives) were identified through the makerspace's network of previous collaborations. For recruitment, individuals who had cocreated innovations at the makerspace, preferably pairs who had collaborated on projects, were approached and informed about the study by the makerspace's coordinator, who is also senior researcher OH. Eligible participants had to actively make contact with the study's first researcher, JVDG, via email. Once contact was initiated, they were sent the participant information letter and informed consent form (ICF). JVDG then scheduled an interview with those who had confirmed their willingness to participate.

Data Collection

Researcher JVDG conducted individual face-to-face and online interviews to collect the data. With a background in nursing and no affiliation with the hospital, JVDG brought relevant experience while maintaining impartiality. A semistructured interview guide was developed, based on the Sunnybrook team–based competencies [28] and the design thinking [17,29] approach to innovation development (Multimedia Appendix 1). The guide was chronologically structured around the phases of innovation: inspiration, ideation, and implementation. A single interview guide was developed, reflecting the equal role of patients within interprofessional teams as emphasized in the Sunnybrook model [28]. The content and language of the interview guide were reviewed by an external expert in patient-driven innovation and senior researchers OH and TVH. It was pilot-tested in 1 interview and subsequently refined to ensure relevance and clarity. After the first 4 interviews, the guide was further adjusted based on emerging insights.

At the beginning of each interview, participants completed a brief demographic form, which included categorical questions about age, innovation experience, and professional perspective based on their employment. All interviews were audio-recorded. A constant comparison method was applied during data collection and analysis, leading to iterative adjustments to both the interview guide and the codebook (for further details, see the "Data Analysis" section). Data collection stopped once data saturation was confirmed.

Data Analysis

Researcher JVDG transcribed the interviews verbatim. Data were analyzed using a thematic approach proposed by Boeije and supported by NVivo (version 14; Lumivero) qualitative research software [30]. This approach consists of 2 phases: segmenting and reassembling. In the segmenting phase, JVDG reread the transcripts and inductively assigned codes to relevant text fragments. In vivo coding was applied to the first 7 interviews. During axial coding, these initial codes were grouped into broader categories, resulting in a preliminary codebook with broader codes, which was then used to code the remaining interviews. This phase was also carried out by JVDG. Before the reassembling phase, JVDG created a mind map to become familiar with the coded data. In the reassembling phase, the axial codes were interpreted and discussed with senior

XSL•FO
RenderX

researchers OH and TVH, leading to the development of conceptual themes and categories. An independent researcher with expertise in qualitative research reviewed the conceptual themes, whereupon JVDG established the final themes. Consensus on the final themes was reached among researchers OH, TVH, and JVDG. Representative quotes and descriptions of themes and categories were derived from the original Dutch data and translated into English.

Study Rigor

The study's credibility was enhanced through member checking and researcher triangulation. Participants received a summary of the key points from their interviews and were invited to provide feedback. The feedback received was incorporated into the data to ensure alignment with participants' perceptions. Researcher triangulation was achieved through weekly meetings between researchers OH, TVH, and JVDG to discuss and monitor the study's progress. Additionally, 2 external researchers—1 with expertise in patient-driven innovation and 1 in qualitative research—were consulted during various phases of the study. To enhance confirmability, a digital logbook containing observational, methodological, and analytical memos was maintained throughout the research process [31].

Ethical Considerations

The study adhered to the principles outlined in the Declaration of Helsinki [32], as well as the European [33] and Dutch codes of conduct [34]. Participants received verbal information about the research and were sent an information letter and ICF via email by the researcher JVDG. All participants signed the ICF prior to the start of the interview. Data confidentiality was ensured through pseudonymization. The study was reviewed and approved by the medical ethics review committee of the Erasmus Medical Center for a nonmedical scientific research application (MEC-2024 - 0019).

Results

Overview

In total, 12 individuals participated: 6 with the designer's perspective and 6 with the patient's perspective (Table 1). The patient perspective was represented by 2 patients, 2 parents, 1 staff member of the Children's Advisory Council who helped patients during the cocreation with the makerspace, and 1 nurse who cocreated with patients and the makerspace to develop an innovation to enhance patients' self-sufficiency in medication administration. All invited eligible individuals agreed to participate. Among these 12 participants, 4 designer-patient pairs were included. From the 8 cocreated innovations discussed in the interviews, 4 innovations were developed for children, 3 innovations for adults, and 1 innovation was primarily developed for a child's caregiver. Most of the participants were equally distributed between the age ranges of 18 to 35 and 36 to 50 years. All designers had more than 3 years of previous experience in innovation projects; patients mostly had none or up to 3 years of such experience.

Table . Demographics of participants (N=12).

Characteristics and categories	Participants, n (%)
Age range (years)	
18-35	5 (42)
36 - 50	5 (42)
51 - 65	1 (8)
>65	1 (8)
Previous experience with innovation (y)	
None	2 (17)
<0.5	0 (0)
0.5 - 1	1 (8)
1 - 3	2 (17)
>3	7 (58)
Perspective	
Industrial product designer	5 (42)
User experience designer	1 (8)
Parent	2 (17)
Patient	2 (17)
Staff member of Children's Advisory Council	1 (8)
Nurse	1 (8)

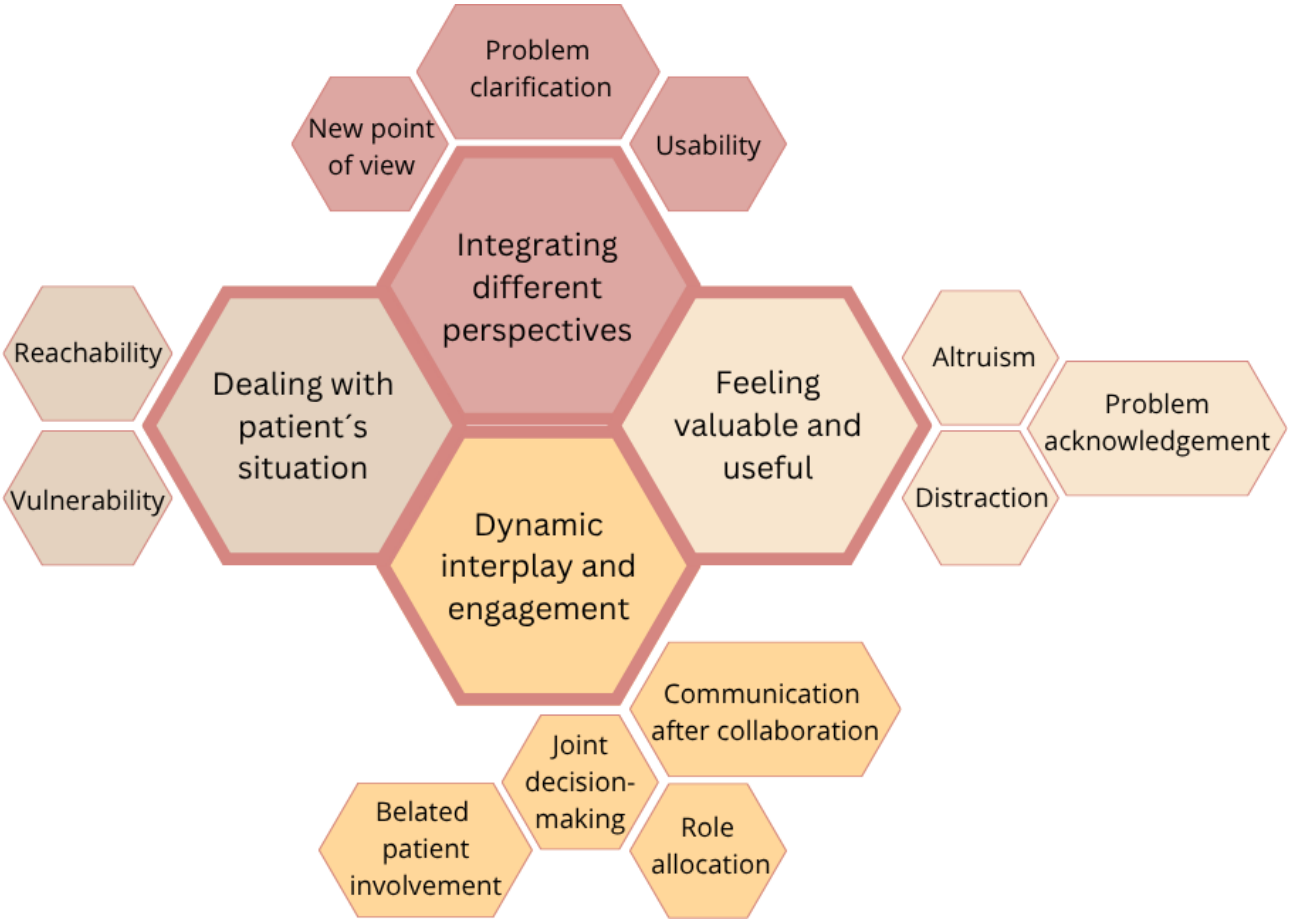
After 8 interviews, no new categories emerged, indicating data saturation. To confirm and assess the identified themes, 4 additional interviews were undertaken. The interviews lasted an average of 51 (SD 19) minutes, with durations ranging from 25 to 90 minutes, depending largely on the participant's level of involvement in the cocreation. Three interviews were

XSL•FO
RenderX

conducted online for convenience, while 9 interviews took place at the hospital. One interview included the patient's partner. Member-check feedback was obtained from 10 participants, all of whom agreed with the summary made by JVDG. Two minor corrections were made based on their feedback. Despite an email reminder, responses from 2 participants were not received.

The following themes emerged: dealing with the patient's situation, integrating different perspectives, feeling valuable and useful, and dynamic interplay and engagement (Figure 2).

Figure 2. Four themes (represented by the larger hexagons) and 12 associated subthemes regarding the experiences of designers and (carers of) patients cocreating in a hospital's makerspace.



Theme: Dealing With the Patient's Situation

The circumstances of patients differ significantly from the everyday experiences of healthy individuals. This theme explores the challenges and implications that these differences create for patients and designers during their collaboration.

Vulnerability Due to the Patient's Health

Participants noted that a patient's compromised health often limits their ability to engage fully in the innovation process, although they may have wanted otherwise. Illness brings competing concerns such as fatigue. Designers—particularly younger ones—are often aware of this vulnerability. For many, cocreation with patients is a new experience, and they may hesitate to impose an undue burden on patients. Some designers struggle with knowing what is appropriate to ask or expect from patients under these circumstances:

> *I was sometimes a bit ashamed to ask things, because I felt like the person would then literally have to think back to such a moment [traumatic experience] you know, just to answer my question.* [P3, user experience designer]

Reachability

Participants suggested that maintaining low-key contact and close involvement increases the patient's reachability. However, they also noted that a patient's reachability is often limited by their health circumstances. First, rapid changes in the patient's condition can lead to unexpected transfers to another ward or home. Designers found it challenging to reach a patient after discharge:

> *It was by chance that [child's name] was admitted, so I could just walk back and forth, but in the future, I would like to be more physically available. If someone says, "come and have a look," I would like to be able to do that, and now that's not possible.* [P12, parent]

Second, designers were hesitant to visit patients on their wards because they do not interact with patients on a daily basis, and extra effort was needed. This was perceived as a barrier. Once on the ward, some designers felt uncomfortable or rushed due to the ward's often hectic environment:

XSL•FO
RenderX

> *But that is just how the intensive care unit is. It is hectic. You don't know at what time someone will come to the bedside [of the patient]. So that was for me a challenge.* [P3, user experience designer]

Additionally, the makerspace is not yet well-known among patients. This lack of knowledge has made it difficult for a patient to access the makerspace. In many collaborations, intermediaries often facilitate contact between designers and patients:

> *Yeah well, I was lucky my nurse knew [industrial product designer 1], otherwise I would never have come up with it.* [P4, patient]

Patients emphasized that, for both parties, it simply comes down to asking. They value their autonomy in decision-making and felt encouraged to express whether or not they wished to participate in a cocreation.

One less experienced designer wondered why she had not collected the patient's contact information during their hospital stay, which would have allowed for continued collaboration after discharge.

Theme: Integrating Different Perspectives

This theme explores the value of incorporating multiple perspectives and the challenges that may arise as a result.

New Point of View

Incorporating multiple perspectives provides a range of viewpoints during the innovation project. Both designers and patients noted that differing perspectives can generate fresh insights into the problem or proposed solution. Both parties felt as though the other party had been given carte blanche:

> *Sometimes, you just have to give people carte blanche because you're already too involved. I had my own bag [tracheostomy tube emergency bag], so I was too much into that concept. You know; this is how we do it. So, I would just show that. But that is not helpful at all, because sometimes you have to start from scratch.* [P7, parent]

> *No, as blank as possible, without any prior knowledge that you just shout what is possible, what you think is possible. That's just beautiful [...]. And then someone else who knows more about it can say, well, this might be possible, and this might not be. I wish they did that more often, [...]. No idea is crazy enough, so to speak.* [P1, industrial product designer]

Problem Clarification and Usability

Designers emphasized the importance of the patient's perspective in clearly defining the problem. Participants also agreed that integrating multiple perspectives had a positive impact on the usability of the innovation in practice. Some noted that when an innovation is not adopted in practice, it is often because the needs of all stakeholders were not adequately identified:

> *And it turned out they [patients] did not like it at all. And then I looked back, and I thought, yes, I can understand that, because we didn't involve them at all in that initial new design. We did ask those parents; what do you prefer? What is important to you? What should come out of it for you? We actually completely forgot to include the patient.* [P6, staff member of Children's Advisory Council]

However, 1 designer pointed out that the inclusion of multiple perspectives, such as a project involving a patient, parent, and nurse, could also cause interference. The differing needs made it harder to clearly define the core problem, and the designer felt that the patient's needs were overshadowed:

> *And that is what I mean by those different perspectives: I get a lot of information but then I wonder, how is it from the child's point of view? It was not clear to me which pain points the child was really experiencing.* [P3, user experience designer]

Theme: Feeling Valuable and Useful

This theme captures how cocreation gave both designers and patients a sense of contributing to something valuable and useful.

Altruism

All participants expressed a strong motivation to contribute to the improvement of health care, even if they did not benefit directly from the outcomes. They desired to help solve a problem and do something for others:

> *I always like to participate in these kinds of experiments. To make people's lives more comfortable.* [P9, patient]

> *Well, someone who is sick, who has a problem and addresses this to ultimately help others sort of, with that approach, running such a project, I think is really great.* [P1, industrial product designer]

Distraction

For patients, participation also served as a distraction from their medical situation. It provided them with a sense of activity and usefulness, allowing them to contribute even in the midst of an unpleasant situation, "And then … yeah, I was bored to death anyway, so I thought yes let's do it" (P4, patient).

Problem Acknowledgment

Both designers and patients felt that the innovation process validated the underlying problem. They appreciated that it was taken seriously and actively addressed. This acknowledgment gave participants a sense that they did not have to simply accept problems or settle for inadequate solutions:

> *Once I had a parent standing in front of me, almost in tears, and she said something like: I think it is really wonderful that there are people who think about those small, seemingly insignificant things, but it makes life easier for such people.* [P10, industrial product designer]

XSL•FO
RenderX

Theme: Dynamic Interplay and Engagement

General Outline

This theme explores the depth of cocreation throughout the innovation process, with a focus on the interplay between designers and patients. The interviews revealed instances where the perceived level of cocreation did not align with the patient's actual involvement.

Participants highlighted that cocreation is highly contextual, varying depending on factors such as the stage of the innovation process and the patient's role. The level of patient involvement often depended on their personal interest, enthusiasm, or health condition:

> *Genuinely working together, yeah, it's possible, but it's difficult. And I think it really depends on the patient or depends on the setting and context.* [P2, industrial product designer]

Role Allocation

Some participants wondered if the cocreation could be considered truly collaborative, as patients were not fully involved throughout the entire design process. Designers often took a leading role, with patients only providing feedback on developed concepts:

> *It's constantly; you go and test something, you receive feedback, you go and redesign it, and you go and test it again.* [P11, industrial product designer]

One designer suggested that genuine cocreation would involve patients actively contributing ideas and helping to design prototypes—essentially taking on a co-designer role. However, the designer also questioned whether this expectation was realistic for patients. Many patients expressed that their level of involvement—whether throughout the entire process or within a single phase—felt appropriate. Contributing to just one part of the cocreation required less time and energy, yet still gave them a sense of satisfaction, particularly knowing that their input could benefit others:

> *If I had been in my full power, I might have done more exploring or for example thought what can we do with this, you know? [...] But now I had so much on my mind, I left it as it was. But I enjoyed it, though.* [P4, patient]

Belated Patient Involvement

Designers expressed concerns about burdening patients, which led them to consider involving patients only in a single phase of the design process. They also noted that excluding patients early on benefited their individual creative process, enabling them to freely explore multiple design options. Often, designers developed a physical prototype before involving patients, as they believed that patients provided more useful feedback on tangible products than on abstract ideas on paper:

> *Because if you start describing it [idea of concept of the innovation] then they might also have questions something like "yes but how about that?" So, seeing a model or something may lead also to other ideas.* [P11, industrial product designer]

Designers highlighted the risk of making unintentional assumptions during the process, which can impact the final design and limit the ability to make changes later. Two designers expressed that they would have preferred to involve the patient earlier in the process. Consequently, participants confirmed that valuable information related to the problem or solution could have been overlooked by excluding patients' perspectives in the early phases:

> *And then you have to take them along from the very beginning, not asking them to test it when you already have developed the [innovation], because that [innovation] might work, but the question is whether that is the best fitting innovation for those parents, because maybe there was something much better.* [P6, staff member of Children's Advisory Council]

Joint Decision-Making

Participants indicated that decisions were largely made collaboratively, which they perceived as pleasant and fostered a sense of equality. This approach facilitated open communication and contributed to a good atmosphere in the makerspace. Patients felt free to express their opinions about the design. While joint decision-making sometimes led to shared disappointment when a design fell short, all participants ultimately reflected positively on the cocreation and were satisfied with the final innovation.

> *It was always a very open conversation, and yeah, we, she was like the designer, and we were the ones presenting the problem, so to speak, but we could really meet each other halfway.* [P5, patient's perspective represented by nurse]

Communication After Collaboration

Participants highlighted the importance of letting patients know—as a sign of appreciation—how their feedback was incorporated. However, many patients said that they missed such communication and wondered how the innovation was further developed.

> *After that, I didn't hear how it ended, and I find that a shame, do you understand? So, then you think, well, he said maybe we'll apply it throughout the whole hospital during that treatment. It would be nice for those who in this case participated to know what happens next to it.* [P9, patient]

One patient discovered that the innovation had not been further developed, and they felt that the innovation was "just" a project that had been shelved. Several participants indeed noted that innovations are sometimes not further developed due to a lack of interest and funding.

Discussion

Principal Results

This study focused on the cocreation between designers and patients (or their carers), exploring their experiences throughout the innovation process. Four key themes were identified. The first theme, dealing with the patient's situation, highlighted

XSL•FO
RenderX

challenges related to patients' vulnerability and limited availability due to their condition. The second theme, integrating different perspectives, emphasized the value of incorporating diverse viewpoints, which helped clarify the problem. However, it also revealed that differing perspectives could complicate problem definition. The third theme, feeling useful and valuable, described how participation fostered a sense of meaningful contribution among patients, offering both a positive distraction and a sense of validation regarding their experienced problem.

The fourth theme, dynamic interplay and engagement, focused on the nature of the interactions within the cocreation. During the interviews and analysis, questions arose about how to distinguish genuine cocreation from surface-level involvement. Despite this, participants generally viewed each other as equal partners. Key aspects within this theme included role allocation, the extent of patient involvement, decision-making, and communication. While existing literature provides limited insights into patient-designer interaction within the cocreation process in hospital-based makerspaces, this study provides a better understanding from both perspectives.

Strengths and Limitations

A strength of this study was the high quality of the data collected using an interview guide based on the Sunnybrook team–based competencies in combination with the design thinking model. Additionally, the use of the constant comparison technique allowed for iterative adjustments to the interview guide, enabling interviews to more effectively address relevant topics. This approach facilitated deeper exploration and assessment of conceptual themes, ultimately leading to data saturation. The high response rate in the member check procedure confirmed that the data accurately reflected participants' experiences, thereby increasing the study's credibility. Another strength was the inclusion of designer-patient cocreation pairs across diverse innovation projects, each with varying levels of patient involvement. This diversity provided detailed insights into how different perspectives interpreted the same collaborative experience.

A limitation of this study is that it was conducted at a single hospital and with a relatively small population of eligible participants. However, for each cocreation between designers and patients within the makerspace, all eligible individuals were contacted. Both perspectives were ultimately represented by 6 participants. Literature suggests that data saturation in heterogeneous samples is typically reached between 9 and 16 interviews, depending on the sample characteristics [35]. We observed a consistent overlap in overarching categories after 8 interviews. Therefore, we consider our sample size to be appropriate for this study.

The patient perspective was represented by 2 patients and 4 individuals who were not patients themselves: 2 parents, a nurse, and a member of the children's advisory board. Because half of the cocreated innovations focused on children, some findings may be slightly oriented toward pediatric contexts. While these participants drew on their own experiences and their interpretations of patients' experiences during cocreation, it remains uncertain how their inclusion may have shaped the patient perspective described in this study. Certain overarching themes, such as dealing with the patient's situation, may initially appear misaligned with a strictly patient-focused view. However, emotional and physical vulnerability were also central to the parents' experiences and therefore remained relevant to our analysis.

Importantly, the nonpatient participants contributed substantial experiential knowledge based on their involvement in cocreation processes with patients and the makerspace. Given that our findings are consistent with existing literature from comparable settings, we consider these complementary perspectives as meaningful additions to the patient perspective rather than a distortion of it.

Another limitation is the potential for recall bias, as some projects took place several years ago. While fading memories may have influenced participants' perceptions, many reported that the interview process helped revive their recollections.

Comparison With Prior Work

Our study highlights how incorporating the patient's perspective can enhance the understanding of user needs, leading to better problem clarification and greater usability of innovations. These findings support previous research on patient involvement in service innovation, as well as literature emphasizing broader stakeholder engagement in health care [19,36-38]. However, our study also sheds light on the complexity of developing technical innovations that involve various end users, including hospitalized patients. Identifying and addressing the diverse needs of stakeholders—such as patients, nurses, infection prevention specialists, and manufacturers—while also complying with existing regulations poses a significant challenge.

Within the makerspace of this study, the average time between the start of development and the implementation of the innovation is 19 months, ranging from 6 to 48 months. This extended timeline often exceeds the duration of hospital stays for patients and is further complicated by the high turnover of patients, making it difficult for designers to reach and involve patients throughout the development process, as observed in our study. Compromised reachability due to these short admissions could explain the belated involvement of patients.

The vulnerability experienced by designers and patients is likewise seen in the literature, where patients' physical and mental conditions influence their preferred level of involvement [39,40]. In a qualitative study by Rowland et al [38] exploring the patient's perspective, vulnerability was identified as an essential element of the patient experience, contributing to a unique form of embodied knowledge. While participants in our study acknowledged the value of the patient perspective, this same vulnerability often led to hesitation and hindered involvement. This finding offers new insights into how patient vulnerability can impact cocreation and innovation development within a makerspace context. A mixed methods study on motives for involvement in service innovation found that patient engagement fostered feelings of accomplishment and was often driven by a desire to give back, as well as by social connection and enjoyment [41]. These motivations closely align with the theme of feeling valuable identified in our study.

XSL•FO
RenderX

A systematic review on service innovation highlighted the gap between the intention to involve patients and the practical challenges of doing so—particularly in joint decision-making processes [42]. The review emphasized that care providers require proper attitudes, skills, and a deeper understanding of patients to facilitate meaningful involvement. Our finding about the lack of communication with patients after the collaboration has not been explicitly reported in previous literature. However, a co-design framework for health care innovation includes a final step of revisiting co-design participants to share the outcomes of the project [23]. This step appears to have been missed by the patients in our study.

Delayed patient involvement, as observed in our study, limited the optimal use of patient contributions. Previous literature similarly indicates that stakeholder involvement often occurs primarily during the front-end prototyping phase, rather than during the earlier inspiration phase of design thinking or consistently across all phases of development [20,43,44]. A qualitative study identified 17 strategies for stakeholder involvement, 1 of which involved real-time modification of prototypes in collaboration with stakeholders [43]. However, this strategy was found to be underused, which aligns with our findings regarding limited patient participation across all development phases.

Implications for Practice

This study underscores the importance of involving patients and their carers in the development of health care innovations. Their unique perspectives are highly valuable, particularly for innovations related to patient care. The findings offer insights into how patients and designers collaborate in innovation development and provide practical guidance for hospital-based makerspaces seeking to strengthen cocreation with patients (Textbox 1). Makerspaces should actively evaluate their cocreations and ensure that cocreation experiences are systematically assessed through patient feedback.

While this study explored the makerspace's previous patient cocreations, it raises the question of how to increase the frequency of patient cocreation. Patients in our study were previously unaware of the existence of the makerspace. Therefore, makerspaces should research strategies to enhance their visibility and accessibility to patients. It is crucial for patients who experience troubling issues to know that a makerspace is available to help them find solutions. Additionally, makerspaces should strive to involve the patients' perspectives in a broader range of projects, beyond those directly related to patient-experienced issues. By doing so, innovations will be better aligned with patients' needs and experiences, ultimately leading to more effective and user-centered health care solutions.

Textbox 1. Practical tips for hospital-based makerspaces regarding cocreation with patients.

The tips are as follows:

- Let patients speak for themselves. Respect patient autonomy in decision-making. Do not hesitate to initiate contact out of concern it may burden the patient. Participation can evoke positive emotions.
- Involve patients early. Engage patients from the problem formulation phase, as they can offer valuable clarification. When their condition allows it, involve them throughout all development phases.
- Maintain communication after collaboration. Keep patients informed about the further development of the innovation. Continued communication reinforces their sense of involvement and appreciation.
- Gather ongoing feedback. Regularly collect feedback from patients on their experiences and suggestions. This helps to refine the cocreation process and ensures that future collaborations are more inclusive and effective.

Conclusions

This study explores the cocreation between designers and patients (or their carers) from both perspectives and offers valuable insights into the cocreation process of patient-initiated technical innovations within a hospital-based makerspace. While cocreation is widely perceived as valuable in innovation development, its success is influenced by the patient's individual situation and their level of involvement. To strengthen future cocreation, makerspaces may benefit from applying practical strategies to effectively involve patients.

Acknowledgments

The authors thank Ko Hagoort for editorial advice.

Funding

The authors declared no financial support was received for this work.

Data Availability

The datasets generated or analyzed during this study are not publicly available because the participants involved in this research did not provide written consent authorizing the public dissemination of their data.

XSL•FO
RenderX

Authors' Contributions

Conceptualization: JVDG, OH, TVH
Formal analysis: JVDG
Investigation: JVDG
Methodology: JVDG, OH, TVH
Project administration: JVDG
Recruitment: OH
Supervision: OH, TVH
Validation: OH, TVH
Visualization: JVDG
Writing—original draft: JVDG
Writing—review and editing: OH, TVH

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File, 1566 KB - humanfactors_v13i1e85926_app1.docx]

Checklist 1
COREQ checklist.
[PDF File, 470 KB - humanfactors_v13i1e85926_app2.pdf]

References

1. Laurisz N, Ćwiklicki M, Żabiński M, Canestrino R, Magliocca P. Co-creation in health 4.0 as a new solution for a new era. Healthcare (Basel) 2023 Jan 27;11(3):363. [doi: 10.3390/healthcare11030363] [Medline: 36766938]
2. Sedini C, Rasmussen AN, Real M, Cipriani L. Co-creating social and sustainable innovation in Makerspaces and Fab Labs lessons learnt from the SISCODE European project. Presented at: Proceedings of the Fab 16 Research Papers Stream (Fab16); Aug 9-12, 2021; Montréal, Canada p. 141-153. [doi: 10.5281/zenodo.5169840]
3. Molloy SJ. Innovation labs in healthcare: a review of design labs as a model for healthcare innovation. : Creative Commons Attribution; 2018 URL: https://openresearch.ocadu.ca/id/eprint/2364/13/MOLLOY_SEAN_2018_MDES_SFI_MRP.pdf [accessed 2023-10-05]
4. Rezaei Aghdam A, Watson J, Cliff C, Miah SJ. Improving the theoretical understanding toward patient-driven health care innovation through online value cocreation: systematic review. J Med Internet Res 2020 Apr 24;22(4):e16324. [doi: 10.2196/16324] [Medline: 32329736]
5. Roddy L, Polfuss M. Employing design thinking methods in nursing to improve patient outcomes. Nurs Forum 2020 Nov;55(4):553-558. [doi: 10.1111/nuf.12461] [Medline: 32497281]
6. Akselrod S, Barron GC, Dain K, et al. Getting health back on the highest political agenda-the UN High-level Meetings on health in 2023. Lancet Glob Health 2023 Jun;11(6):e819-e820. [doi: 10.1016/S2214-109X(23)00166-3] [Medline: 37001539]
7. Securing the future of Europe's ageing population by 2050. Lancet Reg Health Eur 2023 Dec;35:100807. [doi: 10.1016/j.lanepe.2023.100807] [Medline: 38115962]
8. Tamata AT, Mohammadnezhad M. A systematic review study on the factors affecting shortage of nursing workforce in the hospitals. Nurs Open 2023 Mar;10(3):1247-1257. [doi: 10.1002/nop2.1434] [Medline: 36303066]
9. Rigtering C, Spaans LJ, de Jong JPJ. How to bridge the nurse innovation-diffusion gap? An in-depth case study of Create4Care. Front Public Health 2023;11:1209965. [doi: 10.3389/fpubh.2023.1209965] [Medline: 37601181]
10. Patey AM, Soong C. Top-down and bottom-up approaches to low-value care. BMJ Qual Saf 2023 Feb;32(2):65-68. [doi: 10.1136/bmjqs-2022-014977] [Medline: 36517225]
11. van Steenis S, Helder O, Kort HSM, van Houwelingen T. Impact of bottom-up cocreation of nursing technological innovations: explorative interview study among hospital nurses and managers. JMIR Hum Factors 2025 Mar 31;12:e60543. [doi: 10.2196/60543] [Medline: 40163045]
12. Thorsen R, Bortot F, Caracciolo A. From patient to maker - a case study of co-designing an assistive device using 3D printing. Assist Technol 2021 Nov 2;33(6):306-312. [doi: 10.1080/10400435.2019.1634660] [Medline: 31237803]
13. Schiavone F. User Innovation in Healthcare: How Patients and Caregivers React Creatively to Illness, 1st edition: Springer Charm; 2020. [doi: 10.1007/978-3-030-44256-9]
14. Nickel GC, Wang S, Kwong JCC, Kvedar JC. The case for inclusive co-creation in digital health innovation. NPJ Digit Med 2024 Sep 16;7(1):251. [doi: 10.1038/s41746-024-01256-9] [Medline: 39284925]

XSL•FO
RenderX

15. Suna J, Tanti D, Merlo G, et al. Evaluating co-design approaches with parents in paediatric healthcare: a systematic review. BMC Health Serv Res 2025 Oct 28;25(1):1416. [doi: 10.1186/s12913-025-13495-x] [Medline: 41152904]
16. Pearce T, Maple M, Shakeshaft A, Wayland S, McKay K. What is the co-creation of new knowledge? A content analysis and proposed definition for health interventions. Int J Environ Res Public Health 2020 Mar 26;17(7):2229. [doi: 10.3390/ijerph17072229] [Medline: 32224998]
17. Brown T. Design thinking. Harv Bus Rev 2008 Jun;86(6):84-92. [Medline: 18605031]
18. Grindell C, Coates E, Croot L, O'Cathain A. The use of co-production, co-design and co-creation to mobilise knowledge in the management of health conditions: a systematic review. BMC Health Serv Res 2022 Jul 7;22(1):877. [doi: 10.1186/s12913-022-08079-y] [Medline: 35799251]
19. Voorheis P, Petch J, Pham Q, Kuluski K. Maximizing the value of patient and public involvement in the digital health co-design process: a qualitative descriptive study with design leaders and patient-public partners. PLoS Digit Health 2023 Oct;2(10):e0000213. [doi: 10.1371/journal.pdig.0000213] [Medline: 37878566]
20. Cluley V, Ziemann A, Feeley C, Olander EK, Shamah S, Stavropoulou C. Mapping the role of patient and public involvement during the different stages of healthcare innovation: a scoping review. Health Expect 2022 Jun;25(3):840-855. [doi: 10.1111/hex.13437] [Medline: 35174585]
21. Merchán-Baeza JA, Borralleras Andreu C, Minobes-Molina E, Grau Carrión S, Romero-Mas M, Ramon-Aribau A. Co-created technological solutions for caregivers in health care: systematic review. J Med Internet Res 2023 May 1;25:e41260. [doi: 10.2196/41260] [Medline: 37126384]
22. Gagnon MP, Tantchou Dipankui M, Poder TG, Payne-Gagnon J, Mbemba G, Beretta V. Patient and public involvement in health technology assessment: update of a systematic review of international experiences. Int J Technol Assess Health Care 2021 Feb 5;37(1):e36. [doi: 10.1017/S0266462321000064] [Medline: 33541449]
23. Bird M, McGillion M, Chambers EM, et al. A generative co-design framework for healthcare innovation: development and application of an end-user engagement framework. Res Involv Engagem 2021 Mar 1;7(1):12. [doi: 10.1186/s40900-021-00252-7] [Medline: 33648588]
24. Zakoth D, Mauroner O, Emes J. The role of makerspaces in innovation processes: an exploratory study. R&D Manag 2024 Mar;54(2):398-428. [doi: 10.1111/radm.12594]
25. Kahlke RM. Generic qualitative approaches: pitfalls and benefits of methodological mixology. Int J Qual Methods 2014 Feb;13(1):37-52. [doi: 10.1177/160940691401300119]
26. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
27. Holloway I, Galvin K. Qualitative Research in Nursing and Healthcare, 4th edition: Wiley Blackwell; 2016.
28. McLaney E, Morassaei S, Hughes L, Davies R, Campbell M, Di Prospero L. A framework for interprofessional team collaboration in a hospital setting: advancing team competencies and behaviours. Healthc Manage Forum 2022 Mar;35(2):112-117. [doi: 10.1177/08404704211063584] [Medline: 35057649]
29. Hendricks S, Conrad N, Douglas TS, Mutsvangwa T. A modified stakeholder participation assessment framework for design thinking in health innovation. Healthc (Amst) 2018 Sep;6(3):191-196. [doi: 10.1016/j.hjdsi.2018.06.003] [Medline: 29960878]
30. Boeije HR. Analysis in Qualitative Research, 2nd edition: SAGE Publications; 2010.
31. The research audit trail: methodological guidance for application in practice. Electron J Bus Res Methods 2020 Dec 1;18(2). [doi: 10.34190/JBRM.18.2.008]
32. World Medical Association. World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human participants. JAMA 2025 Jan 7;333(1):71-74. [doi: 10.1001/jama.2024.21972] [Medline: 39425955]
33. The European code of conduct for research integrity – revised edition 2023. : All European Academies (ALLEA); 2023 URL: https://allea.org/wp-content/uploads/2023/06/European-Code-of-Conduct-Revised-Edition-2023.pdf [accessed 2026-05-11]
34. Nederlandse gedragscode wetenschappelijke integriteit [Article in Dutch]. DANS Data Station Physical and Technical Sciences. URL: https://doi.org/10.17026/DANS-2CJ-NVWU [accessed 2026-05-11]
35. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022 Jan;292:114523. [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]
36. Shah SGS, Robinson I. Benefits of and barriers to involving users in medical device technology development and evaluation. Int J Technol Assess Health Care 2007;23(1):131-137. [doi: 10.1017/S0266462307051677] [Medline: 17234027]
37. Carlini J, Muir R, McLaren-Kennedy A, Grealish L. Transforming health-care service through consumer co-creation: directions for service design. J Serv Mark 2024 Feb 29;38(3):326-343. [doi: 10.1108/JSM-12-2022-0373]
38. Rowland P, McMillan S, McGillicuddy P, Richards J. What is "the patient perspective" in patient engagement programs? Implicit logics and parallels to feminist theories. Health (London) 2017 Jan;21(1):76-92. [doi: 10.1177/1363459316644494] [Medline: 27130651]

XSL•FO
RenderX

39. Sehlbach C, Bosveld MH, Romme S, Nijhuis MA, Govaerts MJB, Smeenk FWJM. Challenges in engaging patients in feedback conversations for health care professionals' workplace learning. Med Educ 2024 Aug;58(8):970-979. [doi: 10.1111/medu.15313] [Medline: 38415960]
40. Witell L, Engström J, Elg M, Snyder H. The influence of disease and context on patient participation in healthcare service development. Presented at: HELIX Conference on Innovation and Practice in Work, Organization, and Regional Development; Jun 12-14, 2013 URL: https://www.diva-portal.org/smash/get/diva2:717875/FULLTEXT01.pdf [accessed 2026-05-08]
41. Engström J, Elg M. Innovating service while fighting cancer?: user involvement, motivation, and patient well-being. Presented at: HELIX Conference on Innovation and Practice in Work, Organization, and Regional Development; Jun 12-14, 2013 URL: https://www.diva-portal.org/smash/get/diva2:717875/FULLTEXT01.pdf [accessed 2026-05-08]
42. Snyder H, Engström J. The antecedents, forms and consequences of patient involvement: a narrative review of the literature. Int J Nurs Stud 2016 Jan;53:351-378. [doi: 10.1016/j.ijnurstu.2015.09.008] [Medline: 26602069]
43. Rodriguez-Calero IB, Coulentianos MJ, Daly SR, Burridge J, Sienko KH. Prototyping strategies for stakeholder engagement during front-end design: design practitioners' approaches in the medical device industry. Des Stud 2020 Nov;71:100977. [doi: 10.1016/j.destud.2020.100977]
44. Oliveira M, Zancul E, Fleury AL. Design thinking as an approach for innovation in healthcare: systematic review and research avenues. BMJ Innov 2021;7(2):491-498. [doi: 10.1136/bmjinnov-2020-000428]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
ICF: informed consent form

Edited by C Jacob; submitted 15.Oct.2025; peer-reviewed by C Kunze, H Thongdaeng; revised version received 04.Mar.2026; accepted 02.Apr.2026; published 20.May.2026.

Please cite as:
van der Giessen J, Helder O, van Houwelingen T
Patient and Designer Collaboration in Cocreating Technical Innovations in a Hospital-Based Makerspace: Qualitative Interview Study
JMIR Hum Factors 2026;13:e85926
URL: https://humanfactors.jmir.org/2026/1/e85926
doi:10.2196/85926

© Jiske van der Giessen, Onno Helder, Thijs van Houwelingen. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 20.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Perspectives and Needs Regarding Remote Monitoring Technologies Among South Asian Individuals Living With Long-Term Conditions in the United Kingdom: Semistructured Interview and Focus Group Study

Syed Mustafa Ali[1,2*], PhD; Yumna Masood[1*], PhD; Karen Staniland[1], PhD; William G Dixon[1,3], PhD; Sabine N van der Veer[1,2], PhD; Caroline Sanders[2,4], PhD

[1]Division of Informatics, Imaging and Data Science, Manchester Academic Health Science Centre, The University of Manchester, Portsmouth St, Manchester, United Kingdom

[2]National Institute for Health and Care Research (NIHR) Applied Research Collaboration – Greater Manchester (ARC-GM), Manchester, United Kingdom

[3]Northern Care Alliance NHS Foundation Trust, Salford, United Kingdom

[4]Centre for Primary Care and Health Services Research, Division of Population Health, Health Services Research and Primary Care, The University of Manchester, Manchester, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Syed Mustafa Ali, PhD
Division of Informatics, Imaging and Data Science, Manchester Academic Health Science Centre, The University of Manchester, Portsmouth St, Manchester, United Kingdom

Abstract

Background: South Asian individuals face a higher burden of long-term conditions while also experiencing more inequities in health care access and outcomes. Despite the potential of remote monitoring technologies to improve management of long-term conditions, South Asian individuals are less likely to engage with digital health interventions and are underrepresented in health research, partly due to language barriers.

Objective: This study explored the perspectives and needs regarding remote monitoring technologies of South Asian individuals living with a long-term condition in the United Kingdom who did not have English as their first language. We used Pakistanis as an example subgroup of South Asian individuals and rheumatoid arthritis and early inflammatory arthritis as example long-term conditions.

Methods: We conducted semistructured interviews and a focus group discussion with Pakistani adults diagnosed with rheumatoid or early inflammatory arthritis who did not have English as their first language. Audio-recordings were transcribed verbatim, deidentified, and analyzed thematically.

Results: Seventeen adults participated in this study (n=9, 53% in an individual interview and n=8, 47% in the focus group); none of them had previous experience of remote monitoring technologies. We identified three themes: (1) the perceived value and challenges of using remote monitoring technologies for disease self-management, (2) differences in perceived needs and capacity for using remote monitoring technologies between first- and later-generation immigrants related to social determinants, and (3) the role of community and family support in using remote monitoring technologies. Participants perceived remote monitoring technologies as useful, particularly where they were dissatisfied with current health care services. Language and the role of family and community members in supporting technology use were considered important factors, but needs in these areas varied between first-generation (migrated to the United Kingdom) and second- or third-generation immigrants (born in the United Kingdom to parents or grandparents who migrated to the United Kingdom). For first-generation immigrants, these factors intersected with other social and digital determinants, such as gender and literacy, resulting in additional requirements.

Conclusions: Addressing language and literacy barriers, alongside leveraging family and community support, will contribute to equitable remote monitoring technologies to facilitate self-managing long-term conditions among South Asian ethnic minority groups. Future efforts should focus on developing tailored, culturally responsive approaches, particularly for first-generation immigrants, to ensure remote monitoring technologies decrease rather than exacerbate existing ethnic health inequities.

(JMIR Hum Factors 2026;13:e82333) doi:10.2196/82333

XSL•FO
RenderX

KEYWORDS

patient-generated health data; remote monitoring; digital inclusion; South Asian individuals; ethnic health inequalities; language barriers

Introduction

South Asian individuals represent a substantial ethnic minority in many high-income countries, including the United Kingdom [1] and the United States [2]. In these settings, South Asian individuals experience a higher prevalence of long-term conditions [3-5] and related mortality [6,7] compared with their White counterparts. They are also more likely to live with multiple long-term conditions [8], but their needs often remain unmet [9,10]. These differences may partly be attributed to culturally influenced health and treatment beliefs and to language barriers that hinder understanding and engagement with treatment [11-14]. Delayed or limited access to health services further compounds poor experiences and health outcomes [13], as well as quality of life and work productivity [15].

Culture and language shape people's perceptions of and communication about health [16,17], as well as their interactions with health care services and subsequent outcomes [18]. For South Asian individuals, language proficiency is crucial to accessing health information and services. Those with limited formal education may need interpretation support [19], which is not always available due to a shortage of interpreters and logistic complexities [20]. Specific subgroups among South Asian individuals, such as older adults and immigrant women, may face additional challenges, including misalignment between their culturally influenced health beliefs and behaviors and how health services are offered [21]. Compounding these issues is the lack of cultural confidence and competence among health care professionals [22,23], which in turn can result in poor patient and health care provider communication and a lack of consideration of sociocultural factors in care planning. For instance, managing chronic pain effectively requires sensitivity to a patient's cultural context, yet such considerations are often overlooked, leading to suboptimal treatment decisions [24,25].

Remote monitoring technologies, such as symptom-tracking apps or connected home blood pressure monitors, have the potential to improve patient and health care provider communication when managing long-term conditions [26-28], for example, by providing visual summaries of changes in symptoms [29]. The UK National Health Service (NHS) Race and Health Observatory further recognized the potential of digital health technologies to address inequities in health service access and outcomes for ethnic minority populations [30]. However, the successful implementation of such technologies relies on patients' digital literacy skills [31] and meaningful, digitally enabled engagement between patients and health care providers [32]. These requirements may present particular challenges for South Asian communities, where language barriers and culturally inappropriate health practices may hinder uptake and sustained use. Applying digital equity frameworks [33,34] and engaging with South Asian communities during the early stages of technology development may help address these issues and contribute to digital health solutions that are equitable, inclusive, and culturally appropriate.

However, despite this growing attention to digital health technologies, South Asian individuals have lower use rates of digital health tools within the NHS [35] and remain underrepresented in related research [36,37]. This is likely due to intersecting factors at individual and system levels, resulting in language barriers, health beliefs, behaviors [38], and challenges in accessing digital health services and research [39]. These factors leave South Asian individuals' perspectives and needs regarding remote monitoring technologies largely unknown.

Therefore, this study aimed to explore the perspectives and needs regarding remote monitoring technologies among South Asian individuals in the United Kingdom living with a long-term condition, specifically those who do not speak English as their first language. We used Pakistanis as the largest example subgroup of South Asian individuals living in the United Kingdom and rheumatoid arthritis and early inflammatory arthritis as example long-term conditions.

Methods

We used a narrative study design as we wanted to focus on people's narrative accounts of their unique experiences and perceptions regarding various aspects of human interactions and culture [40]. We used the Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist to report this study [41] (Checklist 1).

Ethical Considerations

The study received ethics approval from the UK Health Research Authority South Central - Berkshire B Research Ethics Committee (22/SC/0103). All participants provided written informed consent before taking part, ensuring they understood the purpose of the study, data confidentiality measures, and their right to withdraw from the study without impact on their care and treatment. All participants received compensation for their time. The collected data have been anonymized to safeguard participants' information. The study was conducted in accordance with the Declaration of Helsinki and its later amendments.

Study Context

We used rheumatoid arthritis as the example long-term condition for our study because it is more prevalent among South Asian individuals [42]. They also face significant delays in being diagnosed with rheumatoid arthritis [43] and are less likely to receive treatment [9]. Similar to conditions such as asthma, heart failure, and inflammatory bowel disease, management of rheumatoid arthritis relies on patient reports of symptoms and flares. Many patients struggle to describe these accurately during their brief and infrequent outpatient consultations [44], which may lead to incomplete and inaccurate information, poor

XSL•FO
RenderX

treatment decisions, and ultimately worse outcomes [45,46]. This may be harder for people who do not have English as their first language. Remote monitoring technologies may help to address this by enabling all patients to track their symptoms at home using a smartphone app and share this symptom data with their health care team for discussion during consultations. As an example technology for this study, we used the Remote Monitoring of Rheumatoid Arthritis (REMORA) symptom-tracking smartphone app, which served as a prompt during interviews and focus groups. The app is currently being evaluated as part of a complex intervention in a randomized controlled trial [47].

Study Participants and Recruitment

We recruited participants using a purposive sampling method. People were eligible to take part if they were aged 18 years or above, did not have English as their first language, self-identified as Pakistani, had a diagnosis of rheumatoid arthritis or early inflammatory arthritis, and were able to give informed consent. We focused on Pakistanis as a South Asian subgroup because they form the largest ethnic minority in the United Kingdom [1], and we had 2 research team members from the same background: SMA, a male digital and public health researcher, and YM, a female qualitative health service researcher. Both had Urdu as their first language and had experience working with the Pakistani community in the United Kingdom. SMA and YM were mindful of their possible influence on the data collection, analysis, and interpretation, and they used reflexivity to bring insights, which were relevant for addressing the study objectives.

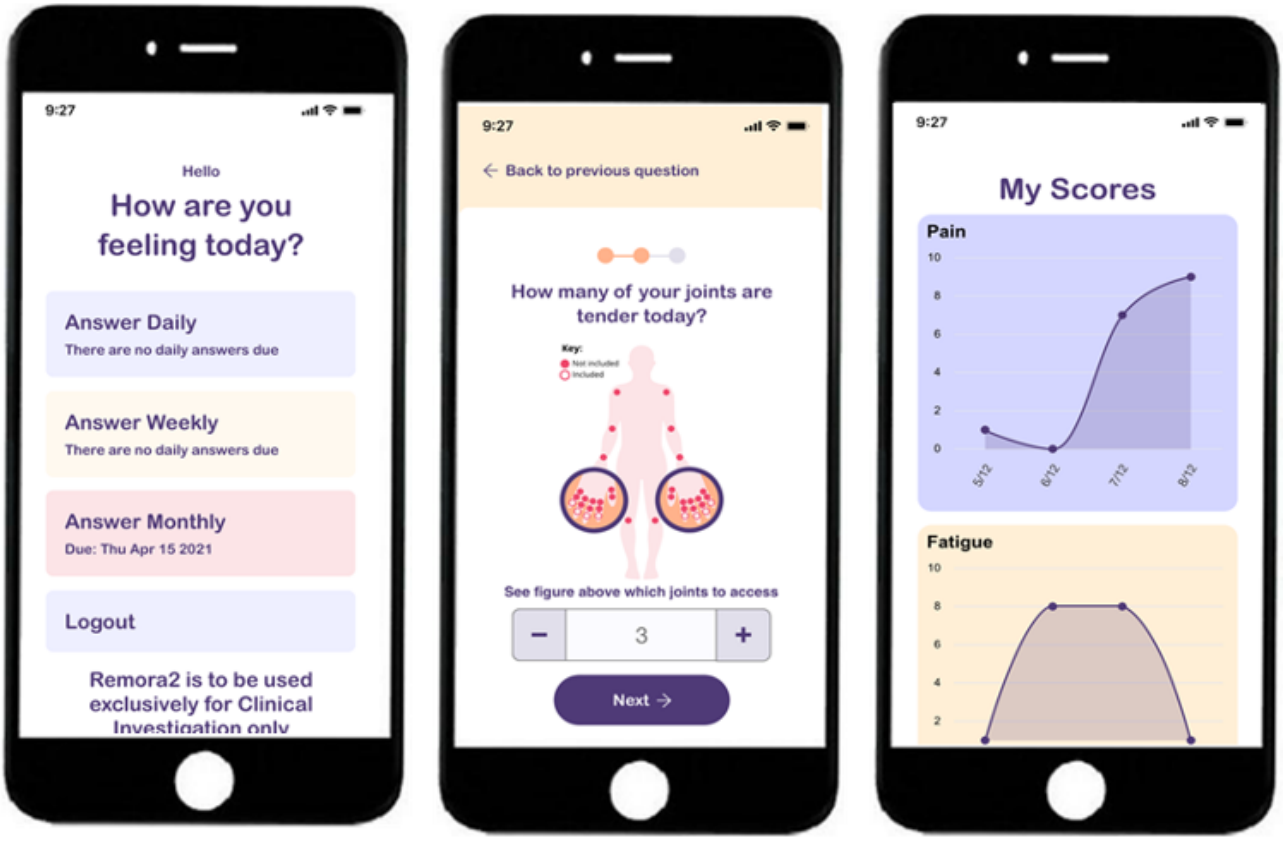
We identified and recruited potential participants via a rheumatology outpatient clinic and a community center for South Asian women in Greater Manchester. Local research nurses and a community center manager approached people to introduce the study, share or read the participant information sheet, and obtain consent to contact from those who were interested. One researcher (YM) contacted potential participants to determine whether they had any questions and were willing to participate, after which they were asked to provide written informed consent. Recruitment continued until data saturation was achieved, with no new ideas emerging from the interviews.

Data Collection and Analysis

We used a brief questionnaire to collect data on participants' age, gender, language proficiency, and their diagnosis. People were invited to participate in either a focus group or an individual interview, whichever they preferred. We conducted interviews and a focus group discussion with the help of topic guides (Multimedia Appendices 1 and 2), which were developed based on published literature [34,37,39,48] and input from patient representatives (KS and others). Topics included experiences of living with rheumatoid arthritis or early inflammatory arthritis, experiences of health care services, experiences of using digital technologies, views on symptom monitoring, and views on using digital health care technologies, including potential benefits or barriers. Previous studies have highlighted the importance of the wider context in which people manage their health care problems, including any social support and existing health care arrangements, when researching perspectives regarding implementation of new health care technologies [42]. Consequently, interviews and the focus group began with a discussion of this wider context before focusing on views about remote monitoring technologies. The topic guides provided a structure for the interviews and focus group discussion while leaving flexibility for participants to introduce topics that researchers had not considered. In addition to the topic guides, we used screenshots from the REMORA app as visual prompts to give participants an idea about what a remote monitoring technology could look like (Figure 1).

XSL•FO
RenderX

Figure 1. Screenshots of the Remote Monitoring of Rheumatoid Arthritis (REMORA) app, an example remote monitoring technology (Copyright University of Manchester).



Individual interviews were conducted in Urdu or in English, whichever language participants preferred. The focus group discussion was in English, as participants were bilingual and were comfortable sharing their thoughts in English. All data collection happened in a community setting familiar to participants, which facilitated a more comfortable and culturally sensitive interaction. All interviews and the focus group were audio-recorded and transcribed verbatim in the language in which they were conducted. The interviews lasted approximately 45 minutes, while the focus group discussion lasted approximately 90 minutes.

SMA and YM independently reviewed transcripts line by line and assigned codes using NVivo software (version 12; Lumivero) to manage the data. For interviews in Urdu, SMA and YM coded the Urdu transcripts and selected and translated relevant quotes into English through discussion. Selected and translated quotes were then discussed with the wider research team (CS, KS, SMA, SNvdV, and YM) to guide iterative coding, thematic analysis, quote selection, and interpretation. We also shared a summary of the findings with our patient representatives to sense-check our interpretations. Finally, SMA and YM double-checked translations of selected quotes used to illustrate themes in the manuscript.

Results

Overview of Participants

Table 1 shows the characteristics of the 17 participants, of whom 10 (59%) participated in an individual interview (9 in Urdu and 1 in English). Most were women (n=16, 94%), aged 60 years and above (n=11, 65%), migrated to the United Kingdom during their life (hereafter referred to as first-generation immigrants; n=15, 88%), bilingual (proficient in speaking and understanding English but with limited proficiency in written English language; n=16, 94%), and living with multiple long-term conditions (hypertension and diabetes were most common; n=11, 65%).

XSL•FO
RenderX

Table . Characteristics of participants (n=17).

Variables	Value, n (%)
Sex	
Female	16 (94)
Male	1 (6)
Age group (y)	
45 - 59	6 (35)
≥60	11 (65)
First-generation immigrant[a]	
Yes	15 (88)
No	2 (12)
Working	
Yes	7 (41)
No	10 (59)
Recruitment route	
Community center	14 (82)
Outpatient clinic	3 (18)

[a]Defined as having migrated to the United Kingdom during their life.

Perspectives and needs related to remote monitoring technologies were explored within a wider context of managing rheumatoid arthritis or early inflammatory arthritis and experiences of health care services. None of the participants had prior experience using any remote monitoring technology. We identified three main themes that reflected participants' views regarding remote monitoring technologies and were developed through an iterative process of analysis: (1) perceived value and challenges of using remote monitoring technologies for disease self-management, (2) differences in perceived needs and capacity for using remote monitoring technologies between first- and later-generation immigrants related to social determinants, and (3) the role of community and family support for using remote monitoring technologies. We described the themes in more detail in the following sections, illustrated with participants' quotes and characteristics (individual interviews only).

Perceived Value and Challenges of Using Remote Monitoring Technologies for Disease Self-Management

Participants perceived both challenges and value in using remote monitoring technologies for disease self-management. Particularly, their self-management practices or lay remedies and their experiences with health care services provided a relevant context for discussing the potential value of integrating remote monitoring technologies in their care.

The use of culturally influenced lay remedies was reflected strongly in people's health beliefs and behaviors regarding managing their condition. All participants mentioned the use of one or more lay remedies, ranging from consuming turmeric, garlic, vinegar, honey, ginger, and herbs to applying massaging oil and turmeric paste on painful sites. Participants also mentioned incorporating lifestyle modifications, such as exercise, yoga, and changes in diet (eg, consuming more vegetables) to manage their condition.

> *I use herbal products too...sometimes I ask someone to bring them from Pakistan or sometimes I get them from Birmingham. Consuming them makes me feel better.* [Female, aged 70 - 80 years, retired, first-generation immigrant]

> *My condition gets worse when I have flare ups...for that time...someone told me to apply black seed oil on my heels...I also add vicks to it...it feels better.* [Male, aged 50 - 60 years, working, first-generation immigrant]

Many participants reported unsatisfactory experiences with health care services, seemingly stemming from a long and difficult diagnostic process, a lack of treatment options being offered, delayed appointments with health care professionals, and limited consultation time. One of the participants said:

> *Firstly, you don't get an appointment...if you get it...then they are not interested in knowing about what has been happening with you.* [Female, aged 60 - 70 years, retired, first-generation immigrant]

They also shared their expectations about support between clinical consultations:

> *I think so between consultations there should be a person who can advise how to manage your pain and your arthritis. This can help a lot, because not everyone is alike.* [Female, aged 70-80 years, retired, first-generation immigrant]

Participants also mentioned how memory affected their day-to-day living and how this might have affected the quality of their consultations with clinicians. One participant described it as follows:

XSL•FO
RenderX

> *I go to my appointment myself...I leave home, and when I reach for my appointment, I forget what I am here for.* [Female, aged 50-60 years, working, first-generation immigrant]

Following these previous experiences, participants acknowledged the potential value of remote monitoring technologies in aiding patient and health care provider interaction, helping patients between consultations, scheduling appointments, and helping them remember and communicate important information about flares, their underlying causes, and their frequency during clinical consultations. One of the participants described how the use of remote monitoring technology could help in memorizing key information about their conditions:

> *[E]ven if I open that app and start using it, I may forget what I wanted to do...but then it would also refresh my mind. If you have already put some information in it then it won't go anywhere.* [Focus group participant]

Remote monitoring technologies could potentially benefit patients by capturing changes in their pain and guiding them in managing their conditions better, as described by one of the participants:

> *It [remote patient monitoring] can help you monitor your condition, if it is getting better or worse, and what are the underlying causes...and what are the alternatives...if your condition is getting better, then how can you improve it further and if it is getting worse then you can identify if a consultation or help is needed.* [Female, aged 60-70 years, retired, first-generation immigrant]

Recognizing the potential benefits of remote monitoring technologies, all participants expressed positive views about sharing their data with health care professionals. However, one of the participants raised concerns about data sharing:

> *Well...it depends where it [data] is going, where it's going to be used, or how far will it go...I don't know.* [Female, aged 70-80 years, retired, first-generation immigrant]

In addition, some older women expressed a lack of motivation for learning to use new digital tools. One woman aged >70 years, who opposed using remote monitoring technologies, shared her frustration regarding trying different tools or approaches for managing rheumatoid arthritis:

> *I have accepted now that this the way I am going to live my life...initially I was advised to maintain a diary, which I used to do. Now I have threw everything away...that I will not do anything now...I feel like I am fed up of all this.* [Female, aged 70-80 years, retired, first-generation immigrant]

When asked how barriers to accessing and using remote monitoring technologies among Pakistani people could be addressed, participants suggested some technology features such as in-app reminders to submit symptom scores, questions or instructions presented in translated and/or audiovisual formats, information explaining how to interpret symptom summary reports, and biometrics-enabled login instead of using passwords. Participants shared challenges or difficulties related to their experience of using technologies:

> *The password is the worst...when you have to make it and they say it's not right, do this, there's no capital. I hate passwords. I have to get my sons to do it.* [Focus group discussion]

Differences in Perceived Needs and Capacity for Using Remote Monitoring Technologies Between First- and Later-Generation Immigrants Related to Social Determinants

Participants suggested differences in needs and capacity for using remote monitoring technologies between first- and second- or third-generation (ie, those born in the United Kingdom to parents or grandparents who migrated to the United Kingdom during their lifetime) Pakistani immigrants and related this generational aspect to other social and digital determinants of health, such as gender. For example, one of the participants referred to gender norms and said:

> *Women are usually at home and are not exposed to technology, and in our culture, it's very common that women are least engaged with new technologies.* [Male, aged 50-60 years, working, first-generation immigrant]

Limited proficiency in English was mentioned by participants as a key potential barrier to using remote monitoring technologies, mainly for first-generation immigrants. Among these immigrants, participants suggested young immigrant women (ie, who moved to the United Kingdom at a young age) were a subgroup at greater risk of developing inequities because of their language barriers. For example, some young immigrant women can only communicate effectively in their local language (such as Punjabi or Pothwari) rather than in Urdu, the national language of Pakistan. One of the participants described it as:

> *The ladies group that I attend regularly [organized by local charities]...most of the group members don't know how to speak English...they sometimes find it hard to communicate in Urdu...these are the ladies who came to the UK for marriage when they were very young, and they typically lack good education and digital skills.* [Female, aged 60-70 years, retired, first-generation immigrant]

One of the participants suggested how language requirements could be addressed:

> *[Referring to the REMORA app] see the content of the app in English, which is a barrier. This should be translated into Urdu...may be in audio format. Certain words such as tiredness and others in the app can be best described through images or visuals.* [Male, aged 50-60 years, working, first-generation immigrant]

There were also different language requirements between first- and second-generation immigrants, suggesting that not every individual with a Pakistani background would require Urdu to access digital health tools and resources (ie, understanding written materials or content). For example, second-generation

XSL•FO
RenderX

immigrants can understand spoken Urdu but may not be able to read Urdu. Similarly, first-generation immigrants are proficient in spoken English but may struggle to read English. One of the participants described it as follows:

> *Having the app in English and Urdu languages will be helpful...our kids understand Urdu...but if you speak to them in Urdu, they feel shy in answering in Urdu, but they do understand it...they just not speak Urdu.* [Female, aged 60-70 years, retired, first-generation immigrant]

Participants considered second- or third-generation immigrants to be more educated and digitally skilled:

> *This second generation is very educated and intelligent. They are well settled and capable to do anything. They are capable to use digital health technologies too.* [Male, aged 50-60 years, working, first-generation immigrant]

Among second- or third-generation immigrants, participants also perceived an emerging need for digital health tools because Pakistani people are being diagnosed with rheumatoid arthritis or early inflammatory arthritis at younger ages. Acknowledging the hesitation among older adults, one of the participants highlighted the need for older adults or first-generation immigrants to use digital health tools for managing their health conditions:

> *...if you can't look after yourself who's going to look after the rest of the family? So, take the time out for yourself. I encourage other ladies to have a free computer lesson. It's a big help to them. But they are reluctant to go, they are scared to go.* [Focus group participant]

According to participants, most first- and second- or third-generation immigrants either lived together in the same household or remained in close contact to support each other in health and non–health-related matters. Related to this, participants identified older women living alone as an example of a subgroup at risk of inequities because of their living and social circumstances. One of the older women said:

> *There are women in their older age, who live alone, their husbands are dead, their kids are settled elsewhere, so normally they are left with their own challenging circumstances...I personally think that women of my [older] age are either not interesting in using digital health technologies or they don't have enough resources to support the use of these technologies.* [Female, aged 60-70 years, retired, first-generation immigrant]

Role of Community and Family Support for Using Remote Monitoring Technologies

Participants acknowledged family and communal values of helping each other, particularly older people, individuals living in challenging circumstances, and those with less education. For example, one of the participants said about older adults:

> *Older people normally would not be able to use these tools [remote monitoring technologies] on their smartphones. Unless there are community groups...just like the one I attend...where someone sits with them and explains and shows them.* [Male, aged 50-60 years, working, first-generation immigrant]

Similarly, participants described how some women got help from local community groups, where they received support for day-to-day living and managing health issues. In addition to community support, children and grandchildren could provide family support for their parents and grandparents to handle their day-to-day affairs and manage health issues. For example, they could find and share relevant health-related information or show them how to search for relevant information online themselves (eg, YouTube videos, which participants mentioned as a common source of information, often shared by their family or friends). One of the participants said:

> *I am not very educated...so kids can help us in recording important information in the app so that doctor could access that information. Otherwise, I would not be able to remember and communicate this information during consultation.* [Female, aged 50-60 years, working, first-generation immigrant]

Sometimes, help available from children and grandchildren was a reason for older adults to lose interest in learning new skills and tools that could otherwise enable them to manage their own conditions better without receiving help from their children. One of the participants said:

> *I am not interested in learning how to use the [remote patient monitoring] app, when all such needs are fulfilled at home by family members.* [Female, aged 60-70 years, retired, first-generation immigrant]

Finally, participants proposed different activities and resources that might be helpful for patients to improve their health and digital literacy, particularly regarding the use of remote monitoring technologies. For example, disease-specific awareness sessions are regularly conducted by local physicians in community settings. Participants proposed organizing similar sessions to promote the potential value of remote monitoring technologies and train people how to use them effectively. In addition to in-person sessions, participants also proposed audiovisual content to help enable them to use such technologies. Group exercise sessions were also suggested, as one of the participants described the difficulty they faced:

> *Say, a physiotherapist gave you exercise, then they say go and just do it. There is no follow-up, there is nothing. You are just left to your own devices. If there is something that people every so often kept on top of called you in, there was some groups or sessions, and it's like how you are doing, and has it worked or not worked?* [Female, aged 70-80 years, retired, first-generation immigrant]

Discussion

Principal Findings

This qualitative study explored perspectives and needs regarding the use of remote monitoring technologies for self-managing

XSL•FO
RenderX

long-term conditions among South Asian individuals living in the United Kingdom who did not speak English as their first language. Study participants were dissatisfied with health care services and perceived remote monitoring technologies as a potential solution to these issues. For example, they believed these technologies could help address memory problems, enrich patient and health care provider communication, and help them feel supported between consultations. However, some subgroups (such as older women living alone in challenging circumstances or young immigrant women with language barriers and poor educational backgrounds) were considered at greater risk of developing further inequities if remote monitoring technologies are introduced in care pathways. Moreover, first-generation immigrants were expected to face more inequities than later-generation immigrants because of differences in intersecting social and digital determinants. Participants highlighted how family and communal values within the South Asian community could be leveraged to guide the development of support for people using remote monitoring technologies, facilitated by audiovisual content and in-person awareness sessions.

Relation to Other Studies

Despite recognizing the perceived benefits of remote monitoring technologies, common authentication methods (eg, passwords), along with a lack of motivation to learn how to use these technologies, may create accessibility barriers [49], as reported by participants in our study. In addition to addressing these accessibility barriers, the provision of timely and personalized care may further promote sustained engagement with remote monitoring technologies [28].

Our study showed intergenerational differences in perspectives and needs regarding the use of remote monitoring technologies to support South Asian individuals in managing their condition. In keeping with Aldosari et al [39], our study highlighted that literacy levels and language requirements varied between first- and second-generation immigrants. This supports previous research advocating language proficiency as an important determinant for equitable access to and use of digital health technologies for South Asian individuals [39] and for ethnic minority groups more generally [50,51]. While technological capabilities to translate English content are improving, people with language barriers may still prefer face-to-face communication [38] and therefore remain reliant on available interpretation support via professional interpreters or informal carers [52]. In turn, this influences their trust in digital health research and services [37,53], making it crucial to explore and address language requirements to ensure equitable and inclusive digital health research and technologies.

We also found that the role of family and community members is important in supporting the use and adoption of remote monitoring technologies. Their role is increasingly being recognized in facilitating South Asian individuals' access to digital health tools and services [38,54] and building their skills and trust in using digital health technologies [55,56]. Particularly in the context of multigenerational households, living arrangements may improve digital literacy and health care access [57] and help navigate usability issues related to digital health technologies [58]. This dependence might not necessarily be perceived as negative, as in Asian cultures more broadly, interdependence tends to be valued (rooted in filial piety, a concept where social norms shape intergenerational relationships [59]). However, we also found that the availability of support from family carers (eg, children or grandchildren) could have a simultaneous negative impact on people's motivation to obtain digital skills. This aligns with findings from another study in South Asian individuals living with cardiometabolic diseases, suggesting that family members and South Asian community members acted as both positive and negative drivers for adopting healthier versions of cultural food [38]. Overall, this suggests that social context could be leveraged to enhance the adoption of remote monitoring technologies, but future research should further explore how family and community support moderate interdependent relationships and ultimately people's engagement with digitally enabled health care services.

Limitations

One of our eligibility criteria was that participants did not have English as their first language. Our aim was to recruit a sample at risk of experiencing language barriers when engaging with digital health technologies and research. Overall, in the United Kingdom, around 20% of people who do not have English as their main language report limited or no English language proficiency [60]. However, most participants in our study were able to speak English very well and did not consider language an important barrier for themselves. Although they highlighted the English language as a potential barrier for some other first-generation immigrant women, our sample did not allow in-depth exploration of perspectives from those with more limited English language proficiency.

Most study participants were recruited via a community-based organization that organized regular health-related activities (eg, awareness sessions and exercise sessions). This may partly explain why community support was identified as a potential enabler for using remote monitoring technologies. Patients recruited through other routes or less health-focused community-based organizations might have had different perspectives on the role of community-based resources in addressing ethnic health inequities.

Implications for Practice and Research

Researchers and technology developers should consider proficiency in reading and writing English or people's native language as an important determinant of health when designing equitable remote monitoring technologies for end users from South Asian ethnic minority groups. This should include exploring the language proficiency of end users' family members (eg, children or grandchildren) to examine if and how information should be offered to ensure it is accessible to multiple generations, enabling both the end user and their "supporters" to engage with digital health tools and resources. Requirements to involve family carers are also reflected in NHS England guidance on access to health services, including the implementation guidance on accessible information standards [61] and the framework for action on digital inclusion [62]. However, the literature provides limited specific examples of how language needs are addressed in practice. Therefore, future

XSL•FO
RenderX

research should examine how aspects of language requirements for digital health technologies can be effectively implemented, particularly within the context of multigenerational households. These explorations should form part of understanding people's wider social context and of proactively involving family carers in co-designing remote monitoring technologies, as has been done previously in the context of dementia [54] and diabetes [63]. While digital skills-building training, as well as educational and cognitive-behavioral programs, have been developed for carers to enhance their pain-related knowledge and self-efficacy, robust evidence of their effectiveness remains limited. Therefore, future research should focus on strengthening this evidence base [64,65]. Overall, involvement of carers will support the development of more equitable remote monitoring technologies that contribute to better health outcomes among South Asian ethnic minority groups.

Conclusions

Remote monitoring technologies have the potential to enhance care for South Asian individuals by addressing language barriers, cultural needs, and issues of trust. Language barriers often intersect with other social and digital determinants, such as age, educational attainment, and digital health literacy. This creates further obstacles, especially for first-generation South Asian immigrants, who may require additional support. In this context, the role of family carers and the wider community in supporting access to and use of remote monitoring technologies is particularly important. Researchers and technology developers must therefore consider the social and cultural environments in which people from South Asian minority backgrounds live to ensure that remote monitoring technologies are inclusive, accessible, and capable of addressing existing health inequities for those living with long-term conditions.

Acknowledgments

The authors would like to thank members of the public involvement group who guided the research. The authors would also like to thank Ms Humera Haqqani from Let's Talk, who supported participant recruitment and helped organize interviews in a community venue.

Funding

The Remote Monitoring of Rheumatoid Arthritis (REMORA) research program is funded by the National Institute for Health and Care Research (NIHR) with a contribution from Versus Arthritis, under NIHR's Programme Grants for Applied Research scheme (grant NIHR202030). The views expressed are those of the authors and not necessarily those of the NIHR, Versus Arthritis, or the Department of Health and Social Care.

Data Availability

Anonymized data will be made available by the corresponding author upon reasonable request.

Authors' Contributions

SNvdV, WGD, CS, and KS conceived the idea of this study. YM drafted topic guides and used them for collecting data. YM and SMA coded and analyzed textual data. SMA, YM, SNvdV, CS, and KS interpreted the data. SMA and YM contributed to the first draft and all authors reviewed and approved the final draft of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Topic guide for interviews.
[PDF File, 197 KB - humanfactors_v13i1e82333_app1.pdf]

Multimedia Appendix 2
Topic guide for workshop.
[PDF File, 165 KB - humanfactors_v13i1e82333_app2.pdf]

Checklist 1
COREQ checklist.
[PDF File, 178 KB - humanfactors_v13i1e82333_app3.pdf]

References

1. Ethnic group, England and Wales: census 2021. Office for National Statistics. 2022. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/bulletins/ethnicgroupenglandandwales/census2021 [accessed 2026-05-11]

XSL•FO
RenderX

2. Health of Asian population. National Center for Health Statistics. URL: https://www.cdc.gov/nchs/fastats/asian-health.htm [accessed 2026-05-11]
3. Yom S, Lor M. Advancing health disparities research: the need to include Asian American subgroup populations. J Racial Ethn Health Disparities 2022 Dec;9(6):2248-2282. [doi: 10.1007/s40615-021-01164-8] [Medline: 34791615]
4. Raleigh V. The health of people from ethnic minority groups in England. The King's Fund. 2023. URL: https://www.kingsfund.org.uk/publications/health-people-ethnic-minority-groups-england [accessed 2026-05-11]
5. Shah AD, Vittinghoff E, Kandula NR, Srivastava S, Kanaya AM. Correlates of prediabetes and type II diabetes in US South Asians: findings from the Mediators of Atherosclerosis in South Asians Living in America (MASALA) study. Ann Epidemiol 2015 Feb;25(2):77-83. [doi: 10.1016/j.annepidem.2014.10.013] [Medline: 25459085]
6. Stafford M, Knight H, Hughes J, et al. Associations between multiple long-term conditions and mortality in diverse ethnic groups. PLoS One 2022;17(4):e0266418. [doi: 10.1371/journal.pone.0266418] [Medline: 35363804]
7. Mathew Joseph N, Strunk FA, Dhar B, Strong LL. Cardiovascular disease in South Asian Americans: a serious health concern. J Nurse Pract 2023 Nov;19(10):104789. [doi: 10.1016/j.nurpra.2023.104789]
8. Eto F, Samuel M, Henkin R, et al. Ethnic differences in early onset multimorbidity and associations with health service use, long-term prescribing, years of life lost, and mortality: a cross-sectional study using clustering in the UK Clinical Practice Research Datalink. PLoS Med 2023 Oct;20(10):e1004300. [doi: 10.1371/journal.pmed.1004300] [Medline: 37889900]
9. Adas MA, Norton S, Balachandran S, et al. Worse outcomes linked to ethnicity for early inflammatory arthritis in England and Wales: a national cohort study. Rheumatology (Oxford) 2022 Dec 23;62(1):169-180. [doi: 10.1093/rheumatology/keac266] [Medline: 35536178]
10. Vakil K, Desse TA, Manias E, et al. Patient-centered care experiences of first-generation, South Asian migrants with chronic diseases living in high-income, western countries: systematic review. Patient Prefer Adherence 2023;17:281-298. [doi: 10.2147/PPA.S391340] [Medline: 36756536]
11. Bosworth A, Dubey S, Adebajo A, et al. Patient empowerment: apni jung (our fight) against rheumatoid arthritis for South Asian population. Mediterr J Rheumatol 2021 Jun;32(2):93-95. [doi: 10.31138/mjr.32.2.93] [Medline: 34447903]
12. Kumar K, Raza K, Nightingale P, et al. Determinants of adherence to disease modifying anti-rheumatic drugs in White British and South Asian patients with rheumatoid arthritis: a cross sectional study. BMC Musculoskelet Disord 2015 Dec 29;16:396. [doi: 10.1186/s12891-015-0831-8] [Medline: 26714853]
13. Kumar K, Raizada SR, Mallen CD, Stack RJ. UK-South Asian patients' experiences of and satisfaction toward receiving information about biologics in rheumatoid arthritis. Patient Prefer Adherence 2018;12:489-497. [doi: 10.2147/PPA.S153741] [Medline: 29670337]
14. Jeraj S, Butt J. Musculoskeletal conditions and Black, Asian and minority ethnic people: addressing health inequalities. Race Equality Foundation. 2020. URL: https://raceequalityfoundation.org.uk/wp-content/uploads/2022/10/MSK-Report-Addressing-Health-Inequalities-2.pdf [accessed 2026-05-11]
15. Jain A, Aggarwal A, Adams J, et al. Work productivity loss among rheumatoid arthritis patients in India: a qualitative study. Rheumatol Adv Pract 2019;3(2):rkz046. [doi: 10.1093/rap/rkz046] [Medline: 32016165]
16. Nisar M, Khan A, Kolbe-Alexander TL. "Cost, culture and circumstances": barriers and enablers of health behaviours in South Asian immigrants of Australia. Health Soc Care Community 2022 Sep;30(5):e3138-e3149. [doi: 10.1111/hsc.13759] [Medline: 35181973]
17. Ali SM, Lee RR, McBeth J, et al. Exploring the cross-cultural acceptability of digital tools for pain self-reporting: qualitative study. JMIR Hum Factors 2023 Feb 8;10:e42177. [doi: 10.2196/42177] [Medline: 36753324]
18. Georgopoulou S, Prothero L, D'Cruz DP. Physician-patient communication in rheumatology: a systematic review. Rheumatol Int 2018 May;38(5):763-775. [doi: 10.1007/s00296-018-4016-2] [Medline: 29582095]
19. Whitaker KL, Krystallidou D, Williams ED, et al. Understanding uptake and experience of interpreting services in primary care in a South Asian population in the UK. JAMA Netw Open 2022 Nov 1;5(11):e2244092. [Medline: 36445711]
20. Rhodes P, Nocon A, Wright J. Access to diabetes services: the experiences of Bangladeshi people in Bradford, UK. Ethn Health 2003 Aug;8(3):171-188. [doi: 10.1080/1355785032000136407] [Medline: 14577994]
21. Chowdhury D, Stolee P, Sims-Gould J, Tong C. "Think positive and don't die alone" - foreign-born, South Asian older adults' perceptions on healthy aging. Int J Qual Stud Health Well-being 2023 Dec;18(1):2253576. [doi: 10.1080/17482631.2023.2253576] [Medline: 37691478]
22. Kumar K, Stack RJ, Adebajo A, Adams J. Health-care professionals' perceptions of interacting with patients of South Asian origin attending early inflammatory arthritis clinics. Rheumatol Adv Pract 2019;3(2):rkz042. [doi: 10.1093/rap/rkz042] [Medline: 31768498]
23. Deshmukh A, Roberts L, Adebajo A, et al. Development and testing of a bespoke cultural intervention to support healthcare professionals with patients from a diverse background. Rheumatology (Oxford) 2024 May 2;63(5):1352-1358. [doi: 10.1093/rheumatology/kead383] [Medline: 37505459]
24. Bull E, Young D, Etchebarne A, Malpus Z. Understanding ethnic minority service user experiences of being invited to and attending group pain programmes: a qualitative service evaluation. Br J Pain 2023 Feb;17(1):58-70. [doi: 10.1177/20494637221129196] [Medline: 36815070]

XSL•FO
RenderX

25. Rogger R, Bello C, Romero CS, Urman RD, Luedi MM, Filipovic MG. Cultural framing and the impact on acute pain and pain services. Curr Pain Headache Rep 2023 Sep;27(9):429-436. [doi: 10.1007/s11916-023-01125-2] [Medline: 37405553]
26. Austin L, Sharp CA, van der Veer SN, et al. Providing “the bigger picture”: benefits and feasibility of integrating remote monitoring from smartphones into the electronic health record. Rheumatology (Oxford) 2020 Feb 1;59(2):367-378. [doi: 10.1093/rheumatology/kez207] [Medline: 31335942]
27. Gandrup J, Ali SM, McBeth J, van der Veer SN, Dixon WG. Remote symptom monitoring integrated into electronic health records: a systematic review. J Am Med Inform Assoc 2020 Nov 1;27(11):1752-1763. [doi: 10.1093/jamia/ocaa177] [Medline: 32968785]
28. Thomas EE, Taylor ML, Banbury A, et al. Factors influencing the effectiveness of remote patient monitoring interventions: a realist review. BMJ Open 2021 Aug 25;11(8):e051844. [doi: 10.1136/bmjopen-2021-051844] [Medline: 34433611]
29. Al-Attar M, Crawford A, Gambin A, Ali SM, Dixon WG, van der Veer SN. Assessing the health equity impacts of smartphone-based remote monitoring in people with rheumatoid arthritis: a stakeholder engagement workshop. Rheumatology (Sunnyvale) 2023 Apr 24;62(Supplement_2). [doi: 10.1093/rheumatology/kead104.089]
30. Wadhawan M, Vu ZN, Linden B, Barry C, Abiola A. Digital apps and reducing ethnic health inequalities: driving progress through digital, design, and organisational transformation. : NHS Race & Health Observatory; 2023 URL: https://nhsrho.org/research/digital-apps-and-reducing-ethnic-health-inequalities/ [accessed 2026-05-11]
31. Hamann P, Knitza J, Kuhn S, Knevel R. Recommendation to implementation of remote patient monitoring in rheumatology: lessons learned and barriers to take. RMD Open 2023 Dec 6;9(4):e003363. [doi: 10.1136/rmdopen-2023-003363] [Medline: 38056918]
32. Watson K, Sheldon H, Pallari E, et al. Remote monitoring of patients with rheumatoid arthritis in a low disease activity state: a mixed methods evaluation across six hospitals in London, UK. Rheumatology (Oxford) 2024 Oct 1;63(10):2721-2733. [doi: 10.1093/rheumatology/keae112] [Medline: 38402509]
33. Crawford A, Serhal E. Digital health equity and COVID-19: the innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020 Jun 2;22(6):e19361. [doi: 10.2196/19361] [Medline: 32452816]
34. Richardson S, Lawrence K, Schoenthaler AM, Mann D. A framework for digital health equity. NPJ Digit Med 2020;5(1):119. [doi: 10.1038/s41746-022-00663-0] [Medline: 35982146]
35. Zhang J, Gallifant J, Pierce RL, et al. Quantifying digital health inequality across a national healthcare system. BMJ Health Care Inform 2023 Nov 24;30(1):e100809. [doi: 10.1136/bmjhci-2023-100809] [Medline: 38007224]
36. Quay TA, Frimer L, Janssen PA, Lamers Y. Barriers and facilitators to recruitment of South Asians to health research: a scoping review. BMJ Open 2017 Jun 1;7(5):e014889. [doi: 10.1136/bmjopen-2016-014889] [Medline: 28576896]
37. Ali SM, Gasteiger N, Masood Y, et al. Addressing ethnic health inequities by improving the inclusiveness of digital health research for South Asians. Exchanges Interdiscip Res J 2024;11(3):284-303. [doi: 10.31273/eirj.v11i3.1566]
38. Ramasawmy M, Sunkersing D, Persson DR, et al. “If it all goes digital, we’ll have to learn”: facilitators and barriers to uptake of digital health in British South Asians with cardiometabolic disease. BMC Digit Health 2024;2:32. [doi: 10.1186/s44247-024-00083-y]
39. Aldosari N, Ahmed S, McDermott J, Stanmore E. The use of digital health by South Asian communities: scoping review. J Med Internet Res 2023 Jun 12;25:e40425. [doi: 10.2196/40425] [Medline: 37307045]
40. Steffen V. Life stories and shared experience. Soc Sci Med 1997 Jul;45(1):99-111. [doi: 10.1016/s0277-9536(96)00319-x] [Medline: 9203275]
41. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
42. Siebert S, Lyall DM, Mackay DF, et al. Characteristics of rheumatoid arthritis and its association with major comorbid conditions: cross-sectional study of 502 649 UK Biobank participants. RMD Open 2016;2(1):e000267. [doi: 10.1136/rmdopen-2016-000267] [Medline: 27403335]
43. Stack RJ, Nightingale P, Jinks C, et al. Delays between the onset of symptoms and first rheumatology consultation in patients with rheumatoid arthritis in the UK: an observational study. BMJ Open 2019 Mar 4;9(3):e024361. [doi: 10.1136/bmjopen-2018-024361] [Medline: 30837252]
44. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
45. Yun H, Chen L, Xie F, et al. Do patients with moderate or high disease activity escalate rheumatoid arthritis therapy according to treat‐to‐target principles? Results from the Rheumatology Informatics System for Effectiveness Registry of the American College of Rheumatology. Arthritis Care Res (Hoboken) 2020 Feb;72(2):166-175. [doi: 10.1002/acr.24083] [Medline: 31566905]
46. Price E, Galloway J, Alveyn E, Coalwood C, Buck J, Farrell G. National early inflammatory arthritis audit: state of the nation summary report 2024. : Healthcare Quality Improvement Partnership; 2024 URL: https://www.hqip.org.uk/wp-content/uploads/2024/10/Ref.-428-NEIAA-SoN-Report-2024.pdf [accessed 2026-05-12]

XSL•FO
RenderX

47. van der Veer SN, Griffiths-Jones D, Parkes M, et al. Remote monitoring of rheumatoid arthritis (REMORA): study protocol for a stepped wedge cluster randomized trial and process evaluation of an integrated symptom tracking intervention. Trials 2024 Oct 15;25(1):683. [doi: 10.1186/s13063-024-08497-9] [Medline: 39407290]
48. Sanders C, Rogers A, Bowen R, et al. Exploring barriers to participation and adoption of telehealth and telecare within the Whole System Demonstrator trial: a qualitative study. BMC Health Serv Res 2012 Jul 26;12:220. [doi: 10.1186/1472-6963-12-220] [Medline: 22834978]
49. Walker RC, Tong A, Howard K, Palmer SC. Patient expectations and experiences of remote monitoring for chronic diseases: systematic review and thematic synthesis of qualitative studies. Int J Med Inform 2019 Apr;124:78-85. [doi: 10.1016/j.ijmedinf.2019.01.013] [Medline: 30784430]
50. Mishra V, Blasi DE, Dexter JP. Bridging ethics and evidence: language as a critical determinant of health equity. Am J Bioeth 2024 Nov;24(11):66-69. [doi: 10.1080/15265161.2024.2402179] [Medline: 39401726]
51. Blum K. Digital determinants of health important with growing technology usage. Association of Health Care Journalists. 2024. URL: https://healthjournalism.org/blog/2024/11/digital-determinants-of-health-important-with-growing-technology-usage/#:~:text=More recently%2C some experts have,Algorithmic bias [accessed 2026-05-12]
52. Seuren LM, Shaw S. How informal carers support video consulting in physiotherapy, heart failure, and cancer: qualitative study using linguistic ethnography. J Med Internet Res 2024 May 31;26:e51695. [doi: 10.2196/51695] [Medline: 38819900]
53. Daggett A, Abdollahi S, Hashemzadeh M. The effect of language concordance on health care relationship trust score. Cureus 2023 May;15(5):e39530. [doi: 10.7759/cureus.39530] [Medline: 37366455]
54. Farnood A, Johnston B, Evans C. The role of digital health interventions in supporting family caregivers for people with dementia to improve quality of life. Curr Opin Support Palliat Care 2024 Dec 1;18(4):224-229. [doi: 10.1097/SPC.0000000000000734] [Medline: 39373310]
55. Ali SM, Saiyed MM, McAvoy A, Mackin R, Jay C, van der Veer SN. Determinants of mistrust in digital health research and approaches to address them among Muslim ethnic minorities living in the United Kingdom: a qualitative study. Int J Equity Health 2025 Aug 15;24(1):225. [doi: 10.1186/s12939-025-02583-3] [Medline: 40817223]
56. Teke J, Olawade DB, Leena N, Weerasinghe K, Mc Lernon S, Moorley C. Digital health disparities: a review of barriers and solutions for racially diverse groups. Int J Med Inform 2026 Feb;206:106173. [doi: 10.1016/j.ijmedinf.2025.106173] [Medline: 41205583]
57. Zhang Z, Yang L, Cao H. Effect of inter-generational living arrangement and digital exclusion on unmet healthcare needs among older adults: findings from two national cohort studies. J Clin Nurs 2026 May;35(5):2159-2180. [doi: 10.1111/jocn.17638] [Medline: 39789828]
58. Fakolade A, Cardwell KL, Chow E, Ross-White A, Pilutti LA. Exploring usability characteristics in computer-based digital health technologies for family caregivers of people with chronic progressive conditions: a scoping review. Digit Health 2025;11:20552076251365141. [doi: 10.1177/20552076251365141] [Medline: 40808709]
59. Mai HV, Le HV. The interdependence of happiness and filial piety within the family: a study in Vietnam. Health Psychol Rep 2023;12(2):124-132. [doi: 10.5114/hpr/172091] [Medline: 38628278]
60. Language, England and Wales: census 2021. Office for National Statistics. 2022. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/language/bulletins/languageenglandandwales/census2021 [accessed 2026-05-12]
61. Accessible Information Standard – implementation guidance. NHS England. 2025. URL: https://www.england.nhs.uk/long-read/accessible-information-standard-implementation-guidance/ [accessed 2026-05-12]
62. Inclusive digital healthcare: a framework for NHS action on digital inclusion. NHS England. 2023. URL: https://www.england.nhs.uk/long-read/inclusive-digital-healthcare-a-framework-for-nhs-action-on-digital-inclusion/ [accessed 2026-05-12]
63. Prinjha S, Ricci-Cabello I, Newhouse N, Farmer A. British South Asian patients' perspectives on the relevance and acceptability of mobile health text messaging to support medication adherence for type 2 diabetes: qualitative study. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e15789. [doi: 10.2196/15789] [Medline: 32310150]
64. Latter S, Hopkinson JB, Richardson A, Hughes JA, Lowson E, Edwards D. How can we help family carers manage pain medicines for patients with advanced cancer? A systematic review of intervention studies. BMJ Support Palliat Care 2016 Sep;6(3):263-275. [doi: 10.1136/bmjspcare-2015-000958] [Medline: 27150294]
65. Chi NC, Barani E, Fu YK, et al. Interventions to support family caregivers in pain management: a systematic review. J Pain Symptom Manage 2020 Sep;60(3):630-656. [doi: 10.1016/j.jpainsymman.2020.04.014] [Medline: 32339651]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
NHS: UK National Health Service
REMORA: Remote Monitoring of Rheumatoid Arthritis

XSL•FO
RenderX

Edited by A Kushniruk, E Borycki; submitted 13.Aug.2025; peer-reviewed by A Gupta, M Ramasawmy, S Mittal, T Manavi; revised version received 25.Mar.2026; accepted 13.Apr.2026; published 26.May.2026.

Please cite as:
Ali SM, Masood Y, Staniland K, Dixon WG, van der Veer SN, Sanders C
Perspectives and Needs Regarding Remote Monitoring Technologies Among South Asian Individuals Living With Long-Term Conditions in the United Kingdom: Semistructured Interview and Focus Group Study
JMIR Hum Factors 2026;13:e82333
URL: https://humanfactors.jmir.org/2026/1/e82333
doi:10.2196/82333

© Syed Mustafa Ali, Yumna Masood, Karen Staniland, William G Dixon, Sabine N van der Veer, Caroline Sanders. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 26.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Digital Assistive Technology Acceptance and Use by Caregivers of Older Adults With Cognitive Impairment: Qualitative Interview Study

Xingyue Su[1], MSW; Honglin Chen[1], PhD; Pasi A Karjalainen[2], PhD

[1]Department of Social Science, University of Eastern Finland, Yliopistonranta 8, Kuopio, Finland

[2]Department of Technical Physics, University of Eastern Finland, Kuopio, Finland

Corresponding Author:
Honglin Chen, PhD
Department of Social Science, University of Eastern Finland, Yliopistonranta 8, Kuopio, Finland

Abstract

Background: Population aging is associated with a rising number of people living with cognitive impairment (CI), straining traditional caregiving systems. More than 55 million individuals live with CI worldwide, while care delivery faces severe capacity constraints. China's rapid aging provides a context for examining the acceptance and use of digital assistive technologies (DATs) in CI care. Although DATs offer the potential to address this care gap, systematic evidence on acceptance and use remains limited.

Objective: This study examined caregivers' acceptance and use of DATs in CI care, applying the Technology Acceptance Model (TAM) framework to identify key facilitators and barriers shaping acceptance and effective technology integration in caregiving environments.

Methods: This qualitative study used semistructured interviews and thematic analysis to examine the use of DATs and identify facilitators and barriers shaping caregivers' acceptance of DATs. Purposive sampling recruited 15 primary caregivers of people with CI across varying decline stages. The study focused on caregivers of people with CI aged 60 years and older. Caregivers were selected as participants because they serve as the primary decision-makers and technology facilitators in CI care.

Results: In CI care, DATs served 5 primary functional domains. Caregivers identified perceived usefulness and ease of use as primary acceptance facilitators. Factors shaping acceptance emerged across 3 dimensions: technological design factors, user psychological factors, and external environmental factors.

Conclusions: The findings suggest that traditional TAM constructs may benefit from contextual adaptation. The acceptance of DATs is not merely an individual assessment of usefulness or ease of use but also a relational and socially mediated process shaped by emotional dependence, family attitudes, technology anxiety, and support systems. Technology may therefore need to evolve dynamically across different stages of cognitive decline, while human caregivers remain irreplaceable in providing emotional support, personalized judgment, and crisis management. Effective DAT integration requires coordinated efforts among relevant stakeholders.

(JMIR Hum Factors 2026;13:e80614) doi:10.2196/80614

KEYWORDS

digital assistive technology; cognitive impairment; Technology Acceptance Model; caregiver; older adults

Introduction

Background

China faces severe challenges from population aging, with the population aged 65 years and older exceeding 220 million, accounting for 15.6% of the total population by 2025 [1]. This demographic transformation has generated substantial demands for innovative care solutions, particularly as cases of cognitive impairment (CI) are growing alarmingly from 55 million globally in 2021 to an estimated 139 million by 2050, with China bearing the largest share worldwide [2]. China currently has approximately 9.83 million patients with Alzheimer disease, one of the most prevalent forms of CI, which imposes substantial medical costs and caregiving burdens [3].

Digital Assistive Technologies in Care for People With CI

In response, digital assistive technologies (DATs) have been increasingly promoted to support the care for people with CI. DATs refer to technology-based tools and systems that are designed to support individuals in maintaining functional ability, independence, and quality of life, particularly in the context of aging and disability [4]. In CI care, these technologies include cognitive training applications, wearable monitoring devices, smart home systems, and other digital solutions that support

XSL•FO
RenderX

daily living, safety, and health monitoring [5,6]. While some DATs require active user interaction, others operate passively through ambient sensing and background monitoring, highlighting the need to align technological solutions with users' cognitive capacities and care environments.

The Chinese government has incorporated CI care into national strategic planning and recent policies promoting smart health and older care models, including wearable devices, service robots, smart antiwandering terminals, medication and care reminder systems, and products for CI assessment and training [7]. Despite growing needs and technological development, a persistent gap remains between the potential of DATs and their effective use in everyday CI care. The usability and effectiveness of DATs are shaped by both technological design and practical deployment [8]. Usability refers to how well a technology enables users to achieve their goals, while effectiveness concerns the accuracy and completeness of that achievement [9]. However, these objective characteristics may not fully explain why some technologies are accepted in practice, while others are not accepted. A device may perform well by objective standards yet still be perceived as difficult to use or of limited value in a specific care context. Focusing solely on objective outcomes, therefore, risks overlooking the subjective factors that determine whether a technology is accepted. Accordingly, this study focuses on caregivers' perceived usefulness and perceived ease of use, rather than on objective measures of system performance.

Caregiving in People With CI

CI encompasses a spectrum of conditions affecting memory, attention, language, and executive function, ranging from mild to severe CI, and is associated with progressive loss of independence and increased reliance on caregivers [2]. Caregivers play a central role in mediating the use of DATs for older adults with CI, particularly when individuals experience diminished capacity to evaluate or communicate their preferences. As many older adults with moderate to severe cognitive decline may have difficulty articulating their needs, caregivers serve as key decision-makers, facilitators, and evaluators of technology use in real-world settings. Accordingly, this study examined both formal (trained professionals providing paid care) and informal (nonprofessional individuals, such as family members, relatives, or friends) caregivers to reflect the varied contexts in which DATs are introduced, integrated, and evaluated in CI care.

Theoretical Frameworks and Related Work

The Technology Acceptance Model (TAM) has been widely used to explain technology acceptance in health care settings [10,11]. TAM proposes that perceived usefulness and perceived ease of use are the primary determinants of users' behavioral intention to use a technology [12]. It has also been applied to studies of technology uptake in aged care, including mobile health services, wearables, and smart home systems [13].

However, TAM does not fully capture the complexity of CI care settings, where acceptance may also be shaped by personalized training, safety concerns, cultural relevance, disease progression, and digital inequality [14]. A further consideration is that technology use does not necessarily reflect acceptance. Caregivers may continue using a device out of obligation or lack of alternatives, even when they do not perceive it as useful or easy to use. Given that DATs in CI care have not yet been widely embedded in routine practice, understanding the factors shaping acceptance is especially important. This study adopts TAM as its primary theoretical framework, with the interview guide structured around its core constructs. The subsequent thematic analysis was conducted inductively, remaining open to themes beyond TAM's original scope. As technology continues to evolve, the factors shaping acceptance may also shift. The findings are therefore discussed through the lens of TAM and its extensions to explore whether the specific context of care for people with CI raises considerations not fully addressed by existing models.

Factors Shaping Technology Acceptance in CI Care

Existing research has identified multiple factors influencing technology acceptance among caregivers and older adults with CI. A systematic review mapping barriers to TAM domains found that most relate to behavioral intentions to use technology, including cost, privacy concerns, stigma, and fear of misuse [15]. Physical and cognitive factors also pose significant challenges, including cognitive decline, reduced physical functioning, and visual, auditory, and language impairments [13]. While prior research has examined caregivers' acceptance of DATs in the care of people with CI, factors beyond perceived usefulness and perceived ease of use have not been fully examined. This study therefore aims to explore other factors that may shape caregivers' acceptance of DATs.

Research Questions

To address these gaps, this study adopts a qualitative approach to explore caregivers' experiences with DATs in the care of people with CI. Specifically, it examines how caregivers use DATs in everyday care scenarios and identifies the factors influencing their acceptance and sustained use. The study is guided by 2 research questions (RQs):

- RQ1: How are DATs used in CI care settings?
- RQ2: What factors shape caregivers' acceptance of DATs?

Methods

Study Design and Participants

This qualitative study used semistructured interviews to explore caregivers' experiences with DATs in CI care among older adults aged 60 years and older. Caregivers were selected as the primary participants because they serve as the key decision-makers in technology acceptance and use in this context. As cognitive capacity varies across stages and types of CI, caregivers play a central role in evaluating, implementing, and sustaining the use of DATs in real-world care settings.

Purposive sampling was used with the following inclusion criteria: (1) age ≥18 years, (2) >1 year of experience caring for older people with CI, (3) used at least one DAT, (4) informal or formal caregiver, and (5) communication ability to clearly express care experiences. It should be noted that inclusion criterion (3) restricted participation to caregivers who had

XSL•FO
RenderX

already adopted at least one DAT; the experiences and views of caregivers who had not used any DAT were therefore not captured in this study.

Fifteen caregivers were recruited through online communities, offline activities, and institutional recommendations, including 11 informal and 4 formal caregivers. The mean age of the participants was 39.67 (SD 15.0; range 21-78) years (9 women and 6 men), and the mean duration of care was 4.3 (SD 4.0; range 1-16) years.

Data saturation was assessed after each interview by evaluating thematic redundancy, adequacy of data to address the RQs, and the overall coherence of the emerging thematic structure. Initial data saturation indicators emerged by the 10th interview, when core themes around technology acceptance factors began stabilizing. No new themes emerged after the 13th interview, and data collection concluded after the 15th interview. Participant characteristics, caregiving contexts, and types of DATs used are summarized in Table 1.

XSL•FO
RenderX

Table . Technology use among informal and formal caregivers in cognitive impairment (CI) care[a].

ID[b]	Gender	Age (y)	Relationship with older adult	Condition of older adult[c]	Duration of care	Technology used	Usage scenario
ICG1	Female	44	Mother-daughter	Severe CI	5 years	Social robot and smart home	Home-based
ICG2	Male	50	Father-son	Mild CI	3 years	GPS shoes and smartwatch	Home-based
ICG3	Male	25	Grandfather-grandson	Mild CI	2 years	Smartwatch	Home-based
ICG4	Female	50	Father-daughter	Severe CI	7 years	Social robot and GPS wristband	Home-based
ICG5	Female	23	Grandfather-granddaughter	Mild CI	1 year	Smart voice assistant	Home-based
ICG6	Male	44	Father-son	Severe CI	3 years	Smartwatch and online mini-program	Home-based
ICG7	Female	46	Father-daughter	Moderate CI	9 years	Social robot and tablet games	Home-based
ICG8	Male	24	Grandmother-grandson	Severe CI with aphasia	2 years	"Xiao Ai" smart assistant and smart mattress	Home-based
ICG9	Female	21	Grandfather-granddaughter	Moderate CI	3 years	Walking aid, smartwatch, and smart home	Home-based
ICG10	Male	78	Spouse	Moderate CI	16 years	Walking aid and smart home	Home-based
ICG11	Female	39	Mother-in-law and daughter-in-law	Severe CI	6 years	Walking aid and emergency alarm system	Home-based
FCG1	Female	44	/[d]	/	1.5 years	"Xiao Ai" smart assistant and smart mattress	Home-based
FCG2	Female	44	/	/	1 year	Smart mattress and senior landline phone	Institutional
FCG3	Male	40	/	/	4 years	Virtual reality (VR) video system	Institutional
FCG4	Female	23	/	/	1 year	Smart wristband and smart mattress	Institutional

[a]To facilitate the identification of participants' roles in the analysis, each respondent was assigned a unique identifier. Informal caregivers are labeled as ICG (informal caregiver), and formal caregivers as FCG (formal caregiver). No standardized cognitive assessment was conducted. Detailed descriptions of the digital assistive technology listed in this table are provided in Multimedia Appendix 1.

[b]The number following each label indicates the individual participant's ID.

[c]CI levels reported in this study reflect caregivers' descriptions informed by prior clinical diagnoses.

[d]Institutional caregivers who provide care to multiple older adults.

Data Collection and Analysis

Semistructured interviews lasting 40 to 60 minutes were conducted online and offline. The interview guide covered participants' background information, DAT use experience, perceived ease of use and usefulness, and factors influencing technology acceptance. TAM informed the interview guide by sensitizing the study to perceived usefulness and perceived ease of use, but it did not determine the full scope of questioning. The semistructured format allowed participants to introduce concerns, experiences, and interpretations beyond the initial TAM-informed prompts. All interviews were recorded and transcribed verbatim, forming approximately 100,000 Chinese words of textual data.

The subsequent thematic analysis was conducted inductively rather than deductively. Codes and themes were generated from participants' accounts following a systematic 6-step approach:

XSL•FO
RenderX

(1) familiarization with data through repeated reading of transcripts and immersion in the dataset; (2) initial code generation to systematically identify and extract key concepts and patterns across the entire dataset; (3) searching for themes by collating codes into potential overarching themes and gathering relevant coded data extracts; (4) reviewing themes through iterative refinement to ensure internal coherence and clear distinctions between themes; (5) naming and defining themes, with later contextualization in relation to the TAM constructs where appropriate; and (6) producing the final analysis report with compelling evidence and analytical narrative [16,17].

The transcribed interviews were read repeatedly by author (XS). Guided by the study aims, relevant data segments were identified and extracted from individual responses, and meaning units were developed with attention to their content and context. These meaning units were interpreted at a semantic level rather than at a latent level. Codes were produced using keywords that captured the essence of the meaning units. The coded meaning units were compared and grouped into categories. After reviewing and refining the categories, they were organized into themes named using short sentences. These themes were finally contextualized and classified according to TAM through a process involving all authors, with authors (HC and PAK) providing final verification and validation. Details of the coding table for data analysis are provided in Multimedia Appendix 2. Research quality was ensured through expert consultation, supervisor discussions, double-checking procedures, diverse participant selection, and regular review of the analysis.

Ethical Considerations

This nonmedical qualitative study was conducted in accordance with the Finnish National Board on Research Integrity (TENK) guidelines [18]. After the discussion with the Ethical Committee of Research at the University of Eastern Finland, no formal review board number was issued for this study. According to Finnish research ethics principles, interview-based human sciences research should follow the guidelines of TENK, including voluntary participation, informed consent, confidentiality, and protection of personal data. As this study involved qualitative interviews with caregivers rather than medical interventions or direct experimentation with older adults with CI, it would normally be assessed as nonmedical human sciences research under Finnish ethical review principles. Particular attention was nevertheless paid to the sensitive caregiving context, participants' emotional burden, and anonymization of all interview data. Participation was voluntary, informed consent was obtained prior to data collection, and confidentiality and anonymity were strictly maintained throughout the study. To acknowledge participants' time investment and valuable contributions, appropriate compensation was provided as an expression of gratitude for their participation. Strict confidentiality protocols were maintained throughout the study, with all personal information and sensitive data anonymized during data processing. Interview recordings were securely stored and deleted on completion of analysis, and all participants are referred to anonymously in research outputs to ensure privacy protection.

Results

Overview

This section presents the key findings in responding to the 2 RQs. The first part addresses how DATs are used in CI care settings (RQ1), highlighting 5 core functional domains. The second part examines factors shaping caregivers' acceptance of these technologies (RQ2), outlining both barriers and facilitators. These themes emerged from the thematic analysis of caregiver interviews.

It is important to note that the functional capabilities of a technology do not automatically translate into perceived benefits for users. A device may be designed to perform a specific function, yet caregivers may not experience it as useful or manageable in their particular care context. This gap between objective functionality and subjective perception is central to why this study focuses on caregivers' perceived usefulness and ease of use, rather than on what technologies are technically designed to do.

DAT Applications: From Daily Living to Cognitive Intervention

The following findings address RQ1 by examining how DATs are used in CI care settings. The analysis of caregiver interviews revealed that DATs serve 5 functional domains in CI care, with applications spanning safety support, emotional and social engagement, daily living assistance, health monitoring, and cognitive function support.

Safety Support Functions

Safety emerged as the primary concern for caregivers, with DATs showing significant value in this domain. Caregivers reported that smart home devices had an impact on safety and caregiver burden reduction and perceived that these technologies enabled individuals with CI to maintain greater independence.

> *My mother's condition is serious, and these technologies help a lot. The robot alerts me if she doesn't move for a while. I can also monitor her remotely on my phone, which makes me feel much more at ease.* [ICG1]

Caregivers described that safety technologies helped reduce the incidence of falls, wandering, and other safety incidents and reported that real-time monitoring enhanced their ability to respond quickly to emergencies.

> *My dad fell in the yard. The watch immediately sent an alarm to my phone, and I rushed over. Fortunately, it wasn't serious. At least I could know it in time…* [ICG2]

> *My father once got lost near the community, but the wristband gave an alarm. We checked the location and found him quickly. The function is very crucial.* [ICG4]

> *We installed smart lights that turn on when they sense movement, and the door locks notify us if Grandpa tries to go out.* [ICG9]

XSL•FO
RenderX

Emotional and Social Support Functions

Social robots are perceived as valuable tools for addressing emotional isolation and social disconnection among individuals with CI. These anthropomorphic devices provided consistent companionship through conversational interactions, entertainment content, and emotional regulation support. Caregivers noted significant reductions in patient agitation, anxiety, and behavioral disturbances when social robots were integrated into care routines.

> *At first, I was confused about the robot, but it has helped our family a lot. My mother now chats with it daily, greeting it every morning and even confiding in it. This relieved much of my stress and gave me more time for other tasks.* [ICG1]

The personalization capabilities of social robots were described as especially beneficial, with devices able to simulate familiar voices or faces, play preferred music, and engage in individualized conversations. Caregivers demonstrated perceived usefulness of DATs in reducing caregiver emotional burden while providing older adults with meaningful social experiences that resembled interactions with family members or friends.

> *Before using the robot, my father would wander looking for my deceased mother daily. The robot's programmed video tasks helped distract him, and now he no longer roams or leaves home. His smiles have even increased from 'chatting' with the robot.* [ICG4]

Caregivers also perceived social robots as supporting emotional stability and social interaction among care recipients while providing themselves with valuable respite and emotional support tools. Caregivers also reported that the continuous availability of these companions helped maintain social connections.

> *This robot is very important to my father. When he was depressed, it suggested playing his favorite music. After listening, he began chatting about the past and gradually recovered.* [ICG7]

Daily Living Assistance Functions

Caregivers reported that smart voice assistants reduced the demands of daily care coordination, particularly in managing medication schedules and appointment reminders. The voice-based interface was seen as especially well-suited to care recipients with declining cognitive abilities by participants, as it required minimal operational effort.

> *We use an intelligent voice assistant to set daily medication reminders. It's very convenient—I don't have to watch over Grandpa all the time. The assistant reminds him when to take which medicine, which is much better than constant supervision.* [ICG5]

> *Sometimes I forget what the medicine is for... the voice assistant tells us its functions, side effects. This saves us the trouble of reading instructions or asking doctors.*
>
> *[ICG5]*

The medication management capabilities of voice assistants extended beyond simple reminders to include drug information provision, dosage verification, and safety monitoring. According to caregivers, when patients or caregivers had questions about medication purposes, administration methods, or potential side effects, voice assistants could provide immediate, accurate information, reducing the risk of medication errors.

Health Monitoring Functions

Caregivers highlighted the reassurance provided by wearable devices and smart sensors through continuous monitoring of vital signs, including heart rate, blood pressure, and body temperature. The ability to receive real-time data and alerts was described as enhancing caregivers' confidence in managing health risks at a distance.

> *I bought a watch for my grandpa to monitor his heart rate. It connects to my phone, so I can know if his heart rate or blood pressure is abnormal.* [ICG3]

Caregivers also reported health monitoring technologies as tools that supported routine observation and early notification in institutional care settings. These devices were presented as assisting caregivers by generating alerts when unusual conditions were detected, which helped staff become aware of potential problems in a timelier way.

> *The smart mattress at the nursing home where I work monitors vital signs, such as when residents get up or remain immobile for long periods. It automatically alerts the responsible nurse in case of abnormalities.* [FCG3]

Cognitive Function Support

Cognitive training applications delivered through various digital platforms, including WeChat mini-programs (lightweight applications embedded within WeChat, a widely used instant messaging and social media platform in China), tablet-based cognitive games, and mobile-based apps, were perceived as showing promise in maintaining cognitive engagement among care recipients.

WeChat mini-programs allow users to access services, including cognitive games without installing separate software, offering convenient cognitive training, including memory exercises, attention training, and logic tasks.

> *My father plays a memory card matching game on WeChat daily. Though he sometimes forgets details, he enjoys it greatly. The group also has small games like arithmetic and riddles, which keep him engaged and prevent excessive daydreaming or sleep.* [ICG6]

Tablet-based cognitive games provided more comprehensive cognitive training through touchscreen interfaces and intuitive game mechanics. These applications were reported to address multiple cognitive domains, including memory, attention, problem-solving, and reasoning skills through engaging, gamified activities such as puzzles, spot-the-difference games, and simple mathematical problems by caregivers.

> *I downloaded a tablet game with puzzles and spot-the-difference for my dad. He plays it daily, which helps improve his reaction speed and concentration.* [ICG7]

XSL•FO
RenderX

Caregivers described the gamified format as helpful for sustaining care recipients' engagement and motivation. In their accounts, game elements such as challenges and progression systems made cognitive exercises feel more enjoyable and supported more regular participation.

> *Grandpa played a tablet game where he had to identify potential dangers like kitchen knives or falling objects. This helped enhance his risk judgment.* [ICG3]

According to caregivers, language-focused games appeared to support verbal communication and delay language decline, although these observations reflect caregiver perceptions rather than clinical assessments.

> *These games include language exercises such as word puzzles that delay language decline and improve vocabulary.* [ICG7]

Factors Influencing Technology Acceptance in CI Care

The following findings address RQ2 by examining the factors that shape caregivers' acceptance of DATs in CI care. These factors are organized into 3 categories: technology design, user psychology, and external environment, each of which encompasses both facilitators and barriers to acceptance.

Technology Design Factors

Design Usability and Operational Challenges

Participants identified operational complexity and user interface design as significant barriers to technology acceptance among individuals with CI and their caregivers. They reported that complex operational procedures and nonintuitive interfaces created difficulties for older adults with declining cognitive abilities, which could reduce willingness to use these technologies.

> *Smart home devices have many functions, but the operation is too complicated. My parents often forget which button to press and sometimes give up. Some functions expected to be automatic aren't used, so I must operate them manually, which takes more time.* [ICG1]

> *My dad struggles with the tablet. There are too many options, and he gets confused easily. He often presses the wrong buttons and ends up on unfamiliar pages with ads. He can only use it with my help.* [ICG2]

Participants further described a gap between their initial expectations and their actual experience of the technologies. Caregivers reported that they had expected these tools to reduce care burden and support care recipient autonomy, but technical limitations and inconsistent performance often limited these perceived benefits.

> *The smart assistant doesn't understand grandma's dialect well; she only speaks Teochew, and her Mandarin isn't standard.* [ICG8]

> *The positioning shoes sometimes fail to give timely alarms, which worries me.* [ICG2]

> *My grandfather often can't hear the voice assistant's medicine reminders clearly and sometimes just ignores repeated instructions.* [ICG3]

> *I thought the smartwatch would reduce my caregiving burden by reminding medicine and monitoring health, but unstable networks and inaccurate data made me spend more time managing it.* [ICG6]

> *I hoped the social robot could ease my dad's emotional issues so I wouldn't have to be with him constantly. But it often misunderstands him, and my daughter or I must stay nearby to help.* [ICG4]

> *The sensor light helps at night when my spouse gets up, but the camera isn't sensitive enough and sometimes misses parts of the room, which worries me.* [ICG10]

Taken together, participants' accounts suggested that perceived usefulness in the care of people with CI was shaped not only by a technology's intended functions but also by the extent to which it performed reliably and aligned with caregivers' expectations in everyday care practice.

Personalization and User-Specific Needs

Participants identified insufficient personalization and limited inclusive design as significant constraints on the perceived usefulness of DATs in the CI care. They reported that many smart devices followed a "one-size-fits-all" design and did not adequately respond to individual care needs, which could reduce willingness for sustained use.

> *Our robot can only customize one spouse image for my dad, which isn't enough. He used to have many friends. If we could add images of old friends or comrades, he'd be happier. We'd pay for better customization, but current options are very limited.* [ICG4]

> *I bought my dad a big tablet, but it has too many functions he can't use. He can't remember complex steps. Many functions are unnecessary for him. The design doesn't fit his needs.* [ICG7]

Although some devices provided simple customization options, these typically failed to fully meet the actual needs of care recipients and caregivers. Participants also expressed a need for more flexible customization across different stages of CI and noted that the absence of stage-appropriate functions limited the usefulness of existing devices in everyday care.

User Psychological Factors

Technology Anxiety and Psychological Adaptation

In the interviews, technology anxiety was presented as an important barrier to technology acceptance among both individuals with CI and their caregivers. Caregivers reported that unfamiliar technologies often gave rise to confusion and apprehension, which could reduce willingness to use these devices.

Caregivers' accounts suggested that technology anxiety became stronger when devices required abilities that people with CI could no longer easily manage. They described complex

XSL•FO
RenderX

technology as increasing psychological burden and sometimes leading care recipients to avoid or resist use. Repeated operational difficulties were also reported as weakening users' confidence and heightening uncertainty about their own ability to use the device. Caregivers further noted that they experienced similar pressure when device setup was complicated or when they had to repeatedly solve technical problems and explain unfamiliar functions during care.

> *Every time my grandfather tries to use the smartwatch, he gets nervous. He fears pressing wrong buttons or accidental payments. Even with repeated teaching, when alone, he dares not operate the device and prefers to give up.* [ICG9]

Caregiver Technology Burden and Support

Although DATs were reported to reduce caregiver burden and improve care efficiency, caregivers reported that learning new technologies alongside daily care work increased their workload, particularly when they also needed to manage maintenance and provide repeated support during use. Limited technical support was described as a further challenge. Several participants noted that device malfunctions often had to be handled independently, and some formal caregivers reported that they had received little training and mainly relied on trial and error.

> *My father can't use mini-programs independently. Whenever issues arise, I have to help him reset or troubleshoot. I've taught him repeatedly, but he can't remember, so I'm always the one managing it.* [ICG6]

> *My dad enjoys interacting with the robot, but it often fails to recognize his speech or responds incorrectly. I must stay with him to correct errors. When no one helps, he becomes anxious and frustrated, sometimes telling the robot to shut up. I worry he'll lose willingness to try.* [ICG4]

> *When equipment malfunctions, we resort to online guides or customer service, but problems often remain unresolved. Instead of reducing workload, these devices increase our burden.* [FCG2]

> *The smart mattresses at our nursing home need training, but it was brief and insufficient. When malfunctions occur, we rely on customer support or technicians, which is time-consuming. Sometimes the alerts misidentify patients, and fixing these issues is complex.* [FCG4]

External Environmental Factors

Family Attitudes and Cultural Values

Caregivers' accounts indicated that family member attitudes and support were important to the acceptance of DATs. They reported that encouragement or resistance from family members, as well as the presence or absence of emotional and practical support, could shape care recipients' acceptance of these technologies in everyday care.

Across interviews, some caregivers described family involvement as facilitating technology use. Support with selecting, purchasing, installing, and explaining devices was presented as helping reduce hesitation and making technologies easier to introduce into daily care.

> *I was hesitant at first because these high-tech products are expensive and unfamiliar. But my daughter was very supportive, helping with installation and teaching my dad step-by-step. Now, although my dad learns slowly, he can basically use the devices, and I feel more secure knowing my daughter is there to help.* [ICG7]

> *My daughter suggested installing smart home systems and helped me choose and set up lights and alarms. I was initially afraid of improper operation, but with her help, I gradually found the devices easy to use. Sometimes, they really save us a lot of trouble.* [ICG10]

Interview accounts also showed that family attitudes could constrain the introduction of DATs. Some caregivers reported that objections from relatives, concerns about cost, or discomfort with the idea of technology-mediated care limited their acceptance of such devices. In these accounts, hesitation was related not only to practical issues but also to concerns about whether technology-based care might be seen as inadequate, impersonal, or inconsistent with family responsibility.

> *I wanted to install a monitoring system to know if grandma is safe when I go out, but my parents opposed it. Grandma said, "What's the point? Treating me like a show?" She often gets angry and resents the nanny. Since grandma disliked it, I didn't argue and didn't buy the equipment.* [ICG8]

> *I wanted to buy my mother-in-law a talking robot, but my husband said it's too expensive and unnecessary...* [ICG11]

> *My mother feels that using robots makes her seem irresponsible or disrespectful to the elderly, who need human companionship more. Although these technologies might reduce burden, she feels uneasy.* [ICG9]

> *People here think elderly care should be done personally. When I mentioned smart devices to peers, many thought I was being lazy or irresponsible. I felt conflicted because the intention was to reduce burden, but it was seen negatively.* [ICG4]

Taken together, these accounts suggest that acceptance of DATs was shaped not only by practical considerations but also by family norms and cultural expectations surrounding care. For some caregivers, technology was viewed as potentially helpful, while for others, it raised concerns about responsibility, appropriateness, and the place of human care in family life.

Affordability and Sustained Use

Economic conditions were an important consideration in caregivers' acceptance and use of DATs. Several caregivers raised concerns about the cost of devices, particularly in families with limited financial resources where purchase, maintenance, and continued use could place a substantial burden on the household.

XSL•FO
RenderX

> *Some technologies like health monitors are excellent, but our family can't afford the high cost, especially for long-term use and maintenance.* [ICG2]
>
> *If prices were reasonable, with public resources or subsidies, I would consider buying more useful devices. Often the technology is good, but we can't afford it.* [ICG6]
>
> *I wanted to buy my mother-in-law a talking robot, but my husband thought it was too expensive and unnecessary. We rely on my income alone, so it's not affordable.* [ICG11]

Across interviews, caregivers indicated that they would be more willing to try or continue using DATs if costs were lower or if financial support, such as subsidies, were available. Some participants acknowledged the potential value of these technologies but described cost as a practical barrier to uptake and continued use. Ongoing expenses, including service fees and maintenance costs, were also mentioned as limiting sustained use in everyday care.

Discussion

Principal Findings

This study explored how DATs are used in care for people with CI and what factors shape caregivers' acceptance. The discussion first relates 4 identified themes to TAM-related models, including family attitudes, social support, technology anxiety, and emotional dependence. It then considers stage-specific care needs, the essential role of human caregivers, and coordinated support for DAT integration.

Technology Acceptance Beyond Established TAM Constructs

Overview

The findings of this study are broadly consistent with TAM's core constructs. Caregivers identified perceived usefulness and perceived ease of use still as central factors shaping their acceptance of DATs. However, the findings differentiate this study from conventional technology acceptance studies by showing how social, emotional, and care-contextual dimensions shape acceptance alongside perceived usefulness and perceived ease of use. Beyond TAM's core constructs, participants described 4 themes shaping their acceptance: family attitudes, social support, technology anxiety, and emotional dependence. Examined through the lens of TAM, TAM3, and Unified Theory of Acceptance and Use of Technology (UTAUT), the identified factors showed different degrees of correspondence with established constructs, although some factors point to dimensions of technology acceptance that existing models have not fully addressed in the care context for people with CI. These observations are offered tentatively, as they are grounded in a small qualitative sample and warrant further empirical investigation.

Family Attitudes

Family attitudes toward technology shaped DAT acceptance in care for people with CI by influencing both attitudes toward technology and the practical conditions for continued use. Sustained technology use is often less a matter of older adults' individual acceptance than of family members' decisions and ongoing involvement, particularly in China, where older adults often rely on family support. Family caregivers typically take responsibility for selecting, setting up, and maintaining technologies, as well as providing day-to-day guidance during use. Prior research on technology socialization highlights that adult children and other family members frequently act as key facilitators by embedding technologies into everyday family practices [19]. Evidence from China further shows that intergenerational support enhances perceived ease of use, perceived usefulness, and trust in digital health services [20]. These findings correspond in part to UTAUT's social influence construct because relatives' expectations shaped attitudes toward DATs [21]. It also relates to facilitating conditions because family members often provide practical support for use.

Social Support

Social support, particularly from social workers and community programs, enhances technology acceptance by providing training and psychological support during initial acceptance phases. The existence of social support enables some caregivers to receive timely assistance when they encounter difficulties, enhancing their confidence and persistence in the use of technology [20]. This corresponds to UTAUT's facilitating conditions construct, which highlights the importance of support structures for technology use [21] because community programs, social workers, and service providers shaped whether caregivers had the training, troubleshooting support, and reassurance needed to continue using DATs. In this study, such support was not limited to formal organizational or technical arrangements, but also included community-based and interpersonal forms of assistance.

Technology Anxiety

Technology anxiety, identified as a barrier to acceptance among both people with CI and caregivers, corresponds closely to the computer anxiety construct in TAM3, where anxiety is understood as shaping perceived ease of use [22]. In this study, this was reflected in caregivers' accounts that unfamiliar or cognitively demanding devices often created apprehension and reduced willingness to engage. Among people with CI, such anxiety was especially strong when repeated operational difficulties occurred, while among caregivers, it was linked to complex setup requirements and responsibility for troubleshooting.

Emotional Dependence

Emotional dependence was the theme least readily aligned with existing TAM or UTAUT constructs. Prior research on human-robot interaction has identified emotional connection as a factor shaping older adults' engagement with assistive technologies, alongside more functional considerations such as perceived usefulness and ease of use [23]. From a relational care and sociotechnical perspective, this finding suggests that DATs enter existing care relationships rather than being added simply as neutral tools. Their acceptance depended partly on whether caregivers and care recipients perceived them as fitting into routines of companionship, reassurance, and emotional

XSL•FO
RenderX

presence. Some caregivers remained hesitant about technology use when they felt that devices could not provide warmth or emotional presence. Others reported that when older adults developed emotional attachment to a device, they were sometimes more willing to engage with it. This theme therefore extends beyond traditional TAM constructs in a narrow functional sense. Devices became more acceptable when they gained emotional meaning within care routines but were resisted when perceived as unable to provide human warmth or relational presence.

Technology Needs Across Stages of Cognitive Decline

Caregivers Described Technology Needs as Evolving Alongside Cognitive Decline

Caregivers described care priorities and technology use as changing across stages of cognitive decline, ranging from cognitive support and memory assistance in mild stages to safety monitoring and passive sensing in later stages. Their accounts suggested 3 broad patterns. In mild stages, caregivers tended to prioritize independence and described using memory aids and cognitive training tools to support everyday functioning. In moderate stages, greater emphasis was placed on safety monitoring and personalized interaction, with social robots and monitoring systems described as more relevant in responding to behavioral change. In severe stages, caregivers referred more often to comprehensive monitoring and nonverbal forms of interaction in the context of high care dependency. These patterns appeared to vary across individuals and care settings. Taken together, the findings suggest that technologies for the care of people with CI may need to be flexible enough to respond to changing care needs across stages of decline.

Human Caregivers and Technology in Collaborative Care

Participants' accounts suggested that human caregivers remained essential in 3 aspects of care: emotional support, personalized judgment, and response to complex situations. Technology was not perceived as fully capable of meeting these needs, particularly when care involved unexpected circumstances. These accounts suggest that participants viewed DATs mainly as supportive tools, while human caregivers continued to be seen as central to care.

> *My mother feels that using robots makes her seem irresponsible or disrespectful to the elderly, who need human companionship more. Although these technologies might reduce burden, she feels uneasy.* [ICG9]

> *The smart mattress means I don't have to stay by the patient's side all the time, but when something more complex happens, you still need a person there to handle it.* [FCG2]

Implications for Coordinated Technology Integration in Care for People With CI

The findings suggest that effective integration of DATs in the care of people with CI may depend on coordinated support across multiple stakeholder groups [24]. Previous research has highlighted the role of clinicians in evaluating and responsibly introducing DATs in this context [25]. In this study, participants' accounts pointed to 3 areas of practical relevance. For technology developers, the findings highlight the importance of age-friendly interfaces and functions that can respond to changing care needs across stages of cognitive decline. For social workers, the findings suggest a role in supporting technology use through guidance, training, and emotional support during implementation. For policymakers, the findings point to the need to improve affordability and reduce structural barriers to access, including through financial support and appropriate regulation. Taken together, these implications suggest that DAT integration requires coordinated support across technology design, caregiver training, implementation support, and policy measures addressing affordability and access.

Limitations

This study acknowledges several limitations. First, the sample was relatively small, comprising 15 caregivers primarily from family care settings in China, which limits representativeness across diverse geographical and cultural contexts. More detailed sociodemographic characteristics, such as education level, socioeconomic background, and digital literacy, were not systematically examined, and these factors may influence technology acceptance. As inclusion criteria (3) required prior DAT use, the sample reflects the experiences of adopters only, and the views of caregivers who had not used any DAT are not captured. The diversity of the sample in terms of caregiver type, age, and technologies used also means that findings cannot be attributed to any specific caregiver group or device category.

The qualitative methodology, while providing rich insights into caregiver experiences, inherently carries subjective interpretation risks and lacks quantitative validation. Additionally, the study focuses on currently available technologies rather than emerging innovations such as AI-powered robots or VR/AR applications, which remain limited in practical implementation due to cost and accessibility barriers [24].

A further limitation concerns the source of the data, as this study captures caregiver perspectives rather than direct experiences of care recipients, who may face communication challenges that complicate assessment of their actual technology preferences. In addition, CI stages reported in Table 1 were based on caregivers' descriptions informed by prior clinical diagnoses, and no standardized cognitive assessments were conducted by this research. Future research should include larger, more diverse samples; use mixed method approaches, longitudinal designs to assess long-term technology adaptation, and develop innovative methods to capture the authentic experiences of care recipients themselves.

Conclusions

This study examined how DATs are used in care for people with CI and what factors shape caregivers' acceptance of these technologies. Caregivers described DATs as being used across 5 functional domains: safety support, emotional and social engagement, daily living assistance, health monitoring, and cognitive function support. Their accounts also showed that acceptance was shaped not only by perceived usefulness and perceived ease of use but also by family attitudes, social support,

XSL•FO
RenderX

technology anxiety, and emotional dependence. Technologies were described as more acceptable when they were reliable, easy to manage, and responsive to stage-specific needs. They were less acceptable when they increased caregiver burden, failed to match everyday care demands, or conflicted with family expectations and care values. DATs may be used in care, but acceptance still depends on perceived usefulness, ease of use, relational care, family attitudes, and support conditions. Overall, the findings suggest that DAT integration depends on relational, cultural, and social contexts rather than technical functionality alone. Across interviews, DATs were generally viewed as supportive tools within care, while human caregivers remained essential for emotional support, individualized judgment, and response to complex situations.

These findings have 3 practical implications. First, DAT design should prioritize simplicity, personalization, and flexibility across different stages of CI. Second, implementation should be accomplished by accessible training and sustained support for caregivers. Third, wider uptake is likely to depend on enabling family, community, and policy conditions. In summary, improving DAT integration in care for people with CI requires attention not only to what technologies are designed to achieve but also to how caregivers and recipients experience, assess, and sustain their use in everyday care. Effective DAT integration must be grounded in caregivers' perceptions and in the broader social and cultural contexts of technologies. Future research should further examine these dynamics in larger and more diverse samples.

Funding

This work was partially supported by NordForsk through the project "AI for Mindful Care and Aging: Navigating vulnerability in the use and implementation of automated home care" (Project 232018).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of digital assistive technologies (DATs).
[DOCX File, 33 KB - humanfactors_v13i1e80614_app1.docx]

Multimedia Appendix 2
Coding table.
[DOCX File, 34 KB - humanfactors_v13i1e80614_app2.docx]

References

1. National Bureau of Statistics of China. China Statistical Yearbook 2025: China Statistics Press; 2025. URL: https://www.stats.gov.cn/sj/ndsj/2025/indexch.htm [accessed 2026-01-23]
2. Li J, Li J. Trends, inequalities, and cross-location similarities in global dementia burden and attributable risk factors across 204 countries and territories: a systematic analysis for the Global Burden of Disease Study 2021. Int J Surg 2025;111(8):5298-5310. [doi: 10.1097/JS9.0000000000002628]
3. Ren R, Qi J, Lin S, et al. The China Alzheimer Report 2022. Gen Psychiatr 2022;35(1):e100751. [doi: 10.1136/gpsych-2022-100751] [Medline: 35372787]
4. Assistive technology. World Health Organization. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/assistive-technology [accessed 2025-05-25]
5. Thordardottir B, Malmgren Fänge A, Lethin C, Rodriguez Gatta D, Chiatti C. Acceptance and use of innovative assistive technologies among people with cognitive impairment and their caregivers: a systematic review. Biomed Res Int 2019;2019:9196729. [doi: 10.1155/2019/9196729] [Medline: 30956989]
6. Zang GY, Rao K, Wu J, He Y, Tang Y, Shi L. Digital therapeutics for cognitive impairment: exploring innovations, challenges, and future prospects. J Med Internet Res 2025 Jun 12;27:e73689. [doi: 10.2196/73689] [Medline: 40504630]
7. General Office of the State Council. Opinions of the General Office of the State Council on developing the silver economy and promoting the welfare of the elderly. Central People's Government Portal. 2024 Jan 15. URL: https://www.gov.cn/zhengce/zhengceku/202401/content_6926088.htm [accessed 2025-05-23]
8. Meiland F, Innes A, Mountain G, et al. Technologies to support community-dwelling persons with dementia: a position paper on issues regarding development, usability, effectiveness and cost-effectiveness, deployment, and ethics. JMIR Rehabil Assist Technol 2017 Jan 16;4(1):e1. [doi: 10.2196/rehab.6376] [Medline: 28582262]
9. Ergonomics of human-system interaction — part 210: human-centred design for interactive systems. International Organization for Standardization. 2019. URL: https://www.iso.org/standard/77520.html [accessed 2026-01-23]
10. Ammenwerth E. Technology Acceptance Models in health informatics: TAM and UTAUT. Stud Health Technol Inform 2019 Jul 30;263:64-71. [doi: 10.3233/SHTI190111] [Medline: 31411153]

XSL•FO
RenderX

11. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A systematic review of the Technology Acceptance Model in health informatics. Appl Clin Inform 2018 Jul;9(3):604-634. [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
12. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep 1;13(3):319-340. [doi: 10.2307/249008]
13. Zhou C, Ai Y, Wang S, et al. Barriers and facilitators to participation in electronic health interventions in older adults with cognitive impairment: an umbrella review. BMC Geriatr 2024 Dec 26;24(1):1037. [doi: 10.1186/s12877-024-05645-3] [Medline: 39725926]
14. Boyle LD, Husebo BS, Vislapuu M. Promotors and barriers to the implementation and adoption of assistive technology and telecare for people with dementia and their caregivers: a systematic review of the literature. BMC Health Serv Res 2022 Dec 23;22(1):1573. [doi: 10.1186/s12913-022-08968-2] [Medline: 36550456]
15. Malden S, McGill K, Guthrie B, et al. Patient and carer-related facilitators and barriers to the adoption of assistive technologies for the care of older adults: systematic review. JMIR Aging 2025 Nov 27;8:e73917. [doi: 10.2196/73917] [Medline: 41313153]
16. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014;9:26152. [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]
17. Braun V, Clarke V. Thematic analysis: a practical guide. bpsqmip 2022;1(33):46-50. [doi: 10.53841/bpsqmip.2022.1.33.46]
18. The Ethical Principles of Research with Human Participants and Ethical Review in the Human Sciences in Finland, 2nd edition: Finnish National Board on Research Integrity (TENK). URL: https://tenk.fi/sites/default/files/2021-01/Ethical_review_in_human_sciences_2020.pdf [accessed 2025-05-23]
19. Tang Z, Shah M, Jamal A. Exploring the process of technology socialization (TS) in the family: ICT adoption for middle-aged parents with the influence of adult children. Inf Syst Front 2024 Oct;26(5):1873-1892. [doi: 10.1007/s10796-023-10428-z]
20. Jin H, Qu Y. Association between intergenerational support, technology perception and trust, and intention to seek medical care on the internet among Chinese older adults: cross-sectional questionnaire study. J Med Internet Res 2025 Jan 6;27:e65065. [doi: 10.2196/65065] [Medline: 39761564]
21. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view1. MIS Q 2003 Sep 1;27(3):425-478. [doi: 10.2307/30036540]
22. Venkatesh V, Bala H. Technology Acceptance Model 3 and a research agenda on interventions. Decision Sciences 2008 May;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
23. Sirizi D, Sabet M, Yahyaeian AA, Bacsu JDR, Smith ML, Rahemi Z. Evaluating human-robot interactions to support healthy aging-in-place. J Appl Gerontol 2025 Jul;44(7):1109-1124. [doi: 10.1177/07334648241301485]
24. Watanabe K, Miwa H, Wakui T. Technology integration in care service systems: the required actions of technology developers. IEEE Eng Manag Rev 2023;51(1):129-142. [doi: 10.1109/EMR.2023.3237516]
25. Ienca M, Fabrice J, Elger B, et al. Intelligent assistive technology for Alzheimer's disease and other dementias: a systematic review. J Alzheimers Dis 2017;56(4):1301-1340. [doi: 10.3233/JAD-161037] [Medline: 28222516]

Abbreviations

CI: cognitive impairment
DAT: digital assistive technology
RQ: research question
TAM: Technology Acceptance Model
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by C LeRouge; submitted 14.Jul.2025; peer-reviewed by G Spalla, H Heppner, S Obeidat; revised version received 24.May.2026; accepted 25.May.2026; published 19.Jun.2026.

Please cite as:
Su X, Chen H, Karjalainen PA
Digital Assistive Technology Acceptance and Use by Caregivers of Older Adults With Cognitive Impairment: Qualitative Interview Study
JMIR Hum Factors 2026;13:e80614
URL: https://humanfactors.jmir.org/2026/1/e80614
doi:10.2196/80614

© Xingyue Su, Honglin Chen, Pasi A Karjalainen. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 19.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium,

XSL•FO
RenderX

provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring Patients' Experiences in a Blended Transdiagnostic Group Treatment: Qualitative Study

Noelia Jiménez-Orenga[1], PhD; Alba Fadrique-Jiménez[2,3], MSc; Azucena García-Palacios[1,4], PhD; Jorge Osma[2,3], PhD; Amanda Díaz-García[2], PhD; Juana Bretón-López[1,4], PhD

[1]Department of Basic, Clinical Psychology and Psychobiology, Universitat Jaume I, Castellón, Spain

[2]Department of Psychology and Sociology, Faculty of Social and Human Sciences, Universidad de Zaragoza, C/ Ciudad Escolar, 4, Teruel, Spain

[3]Health Research Institute of Aragon (IIS Aragón), Zaragoza, Spain

[4]CIBER de Physiopathology of Obesity and Nutrition (CIBEROBN), Instituto de Salud Carlos III, Madrid, Spain

Corresponding Author:
Amanda Díaz-García, PhD
Department of Psychology and Sociology, Faculty of Social and Human Sciences, Universidad de Zaragoza, C/ Ciudad Escolar, 4, Teruel, Spain

Abstract

Background: Emotional disorders (EDs) are the most prevalent mental disorders worldwide. Health services face significant difficulties in attending to the high demand and applying evidence-based psychological treatments. Combining the transdiagnostic approach with the group and blended formats could help the accessibility of treatment for ED. It is important to assess the feasibility and acceptability of new interventions from a qualitative perspective.

Objective: This study aimed to explore the experiences and opinions of patients with ED who have received transdiagnostic cognitive behavioral therapy treatment in a group and blended format.

Methods: Two subsamples of participants were included: those who completed the intervention (n=18) and those who did not (n=4). The completers' subsample participated in focus groups, which were transcribed verbatim and subsequently analyzed using a consensus qualitative research methodology. The noncompleters subsample completed a brief online questionnaire. Key themes identified by 2 independent researchers were described and used as complementary to focus group findings. Results were reported following the COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines.

Results: After analyzing the focus group interviews using the consensus qualitative research method, 8 domains (or topic areas) were identified, namely experience with the online platform, configuration of the blended intervention, therapeutic content, experience with the group sessions via videoconference, role of the therapists, overall assessment of the blended treatment, elements that help maintain adherence, and suggestions for improvement. Overall, participants reported benefits from this type of intervention, expressed satisfaction with the treatment, and highlighted the perceived improvement and usefulness of what they learned. They also noted the good usability of the platform and the possibility of establishing a good therapeutic alliance in this format. Participants also pointed out unfavorable aspects and offered suggestions on possible areas for improvement. For noncompleters, various reasons for dropping out were identified: lack of treatment efficacy, intervention-related stress, difficulty in applying learned strategies, low involvement, lack of motivation, time constraints, or need for different types of help. Strategies to enhance adherence to treatment were also identified, such as changes in the format and configuration of the sessions, allowing more time to discuss individual problems or reducing the number of homework assignments.

Conclusions: These qualitative results support the feasibility, acceptability, and clinical utility of a blended group transdiagnostic intervention for the treatment of ED. This novel format could be a scalable and well-valued option within mental health services, although it can still be further optimized based on the results of this study.

Trial Registration: ClinicalTrials.gov NCT04008576; https://clinicaltrials.gov/study/NCT04008576

(JMIR Hum Factors 2026;13:e86016) doi:10.2196/86016

KEYWORDS

qualitative research; opinion; acceptability; completers; dropouts; transdiagnostic; CBT; group; blended; emotional disorders; anxiety; depression

XSL•FO
RenderX

Introduction

Depressive, anxiety, and other related disorders, commonly known as emotional disorders (EDs) [1], are the most prevalent mental disorders worldwide [2]. In Spain, the cases of ED reach almost 17% of the general population, being the most frequently diagnosed psychological disorders in health care services [3] and generating a high economic burden [4].

In this regard, the National Health System (NHS) faces significant challenges in meeting the high demand for mental health care, with a limited number of psychology professionals [5]. While international recommendations suggest a caseload of approximately 86 patients per year per professional, psychologists in the Spanish NHS attend approximately 330 patients annually [6]. This contributes to delayed care, long waiting lists, and extended time between sessions, thereby hindering the implementation of evidence-based treatments (EBTs).

To address these limitations, alternative therapeutic approaches with a more dimensional perspective have been proposed. The transdiagnostic approach stands out among them. This approach focuses on intervening in common factors involved in the origin and maintenance of various psychological disorders [7], such as the various EDs [8], which could facilitate the implementation of EBT and the management of comorbidities [9].

In parallel, digital mental health interventions, particularly internet-based therapy, have emerged to improve treatment accessibility. Online cognitive behavioral therapy (CBT) has been shown to be effective for treating anxiety and depression [10], even when using a transdiagnostic approach to address ED [11]. However, some patients require more therapist contact and customized treatment [12,13]. Blended treatments, which combine elements of face-to-face and online therapy, have emerged as a promising alternative, leveraging the strengths of both approaches while reducing some of their barriers [14]. Additionally, the blended format has been effective in reducing symptoms of depression and anxiety [15-17].

Another alternative that can help to improve psychological care, especially in the NHS, is to carry out interventions in group format. This type of intervention has been shown to be as effective as individual therapy in treating ED [18,19] and also has advantages for patients, such as peer support and reducing stigma [20,21]. Furthermore, it is a more efficient option than individual therapy as it allows treating many individuals at once and reducing the burden of services [21,22].

Consequently, there is growing interest in combining transdiagnostic, blended, and group approaches into a single treatment protocol [23]. This integration could help more people with ED access EBT and allow a greater number of individuals with similar issues to be treated simultaneously. However, beyond effectiveness, it is crucial to evaluate the feasibility and acceptability of these innovative interventions from both quantitative and qualitative perspectives [24]. Qualitative research is a valuable method for collecting and analyzing nonnumerical data, enabling an in-depth exploration of individual experiences and providing detailed information about concepts and opinions [25].

There are some qualitative studies on blended interventions from the patients' [26,27] and the therapists' points of view [28,29]. However, literature in this area remains scarce, and, to our knowledge, no studies have reported the opinion of patients who have received a transdiagnostic CBT intervention for ED applied in a blended and group format.

This qualitative study is part of a broader feasibility trial (trial registration: NCT04008576) [23] and its main objective was to explore in depth the experiences and opinions of patients who received a transdiagnostic CBT intervention in blended and group formats for ED, in order to understand their degree of acceptance, satisfaction, and possible areas for improvement from their perspective. Quantitative results on feasibility, acceptability, and preliminary efficacy are reported elsewhere [30].

Methods

Participants

Participants were selected through convenience sampling from a community sample of patients who had received a blended transdiagnostic group intervention for EDs at the Emotional Disorders Clinic of Universitat Jaume I [30]. Patients were required to be 18 years or older; no upper age limit was imposed. The initial sample included both individuals who had received psychological treatment prior to the feasibility trial and participants who were receiving psychological therapy for the first time.

Patients began the same intervention but differed in their level of treatment adherence; 23 out of 34 participants (68%) completed the full intervention, and 11 out of 34 (32%) discontinued treatment [30]. Therefore, this study included 2 subsamples based on treatment completion status. This was done to explore the experiences of those who finished the intervention and those who dropped out.

The first subsample consisted of 18 patients (14 women and 4 men) who completed the full intervention protocol, with a mean age of 33.94 (SD 11.03; range 20 - 54) years. Sociodemographic and clinical characteristics of participants at the beginning of the treatment can be found in Table 1.

XSL•FO
RenderX

Table . Sociodemographic and clinical characteristics of subsample 1.

Code	Sex	Age	Marital status	Educational level	Main diagnosis	Comorbid diagnoses
P[a]1#FG[b]1	Female	23	In a relationship	HE[c]	MDD[d]	3 (PD[e], A[f], GAD[g])
P2#FG1	Female	54	Married	HE	GAD	0
P3#FG1	Female	27	Single	HE	GAD	3 (MDD, PD, A)
P4#FG1	Male	35	In a relationship	HE	MDD	2 (GAD, SAD[h])
P5#FG1	Female	38	Married	ME[i]	GAD	3 (MDD, SAD, A)
P6#FG1	Female	52	Married	HE	GAD	2 (OCD[j], MDD)
P7#FG2	Female	23	In a relationship	HE	PDD[k]	1 (GAD)
P8#FG2	Female	23	In a relationship	HE	GAD	1 (MDD)
P9#FG2	Female	27	In a relationship	HE	NS anxiety disorder[l]	0
P10#FG2	Female	20	In a relationship	HE	GAD	3 (OCD, MDD, A)
P11#FG3	Male	40	Single	HE	A	2 (PD, GAD)
P12#FG3	Female	27	Single	HE	MDD	2 (SAD, GAD)
P13#FG3	Female	51	Married	HE	MDD	1 (GAD)
P14#FG4	Female	40	Married	EE[m]	A	2 (PD, SAD)
P15#FG4	Male	34	In a relationship	HE	GAD	1 (MDD)
P16#FG4	Female	35	In a relationship	HE	MDD	3 (GAD, SAD, A)
P17#FG4	Female	20	Single	HE	GAD	0
P18#FG4	Male	42	Single	HE	PD	3 (PDD, A, GAD)

[a]P: participant.
[b]FG: focus group.
[c]HE: higher education.
[d]MDD: major depressive disorder.
[e]PD: panic disorder.
[f]A: agoraphobia.
[g]GAD: generalized anxiety disorder.
[h]SAD: social anxiety disorder.
[i]ME: middle education.
[j]OCD: obsessive-compulsive disorder.
[k]PDD: persistent depressive disorder (dysthymia).
[l]NS anxiety disorder: nonspecified anxiety disorder.
[m]EE: elementary education.

The second subsample consisted of 4 patients (all women) who initiated but discontinued the intervention before completing it (dropout). Their mean age was 28.5 (SD 12.39; range 21 - 47) years. Sociodemographic and clinical characteristics of these participants at the beginning of the intervention can be found in Table 2.

XSL•FO
RenderX

Table . Sociodemographic and clinical characteristics of subsample 2.

Code	Sex	Age	Marital status	Educational level	Main diagnosis	Comorbid diagnoses
D[a]1	Female	24	In a relationship	HE[b]	SAD[c]	1 (A[d])
D2	Female	22	Single	HE	OCD[e]	1 (MDD[f])
D3	Female	21	In a relationship	HE	SAD	1 (MDD)
D4	Female	47	In a relationship	HE	GAD[g]	3 (MDD, SAD, A)

[a]D: dropout.

[b]HE: higher education.

[c]SAD: social anxiety disorder.

[d]A: agoraphobia.

[e]OCD: obsessive-compulsive disorder.

[f]MDD: major depressive disorder.

[g]GAD: generalized anxiety disorder.

Treatment

The treatment in which all participants previously took part was a transdiagnostic CBT intervention designed to address ED, composed of 16 modules and lasting 24 weeks. It was delivered in a group and blended format, so that patients attended a total of 8 group sessions by videoconference (2 hours in duration), and, in the period between sessions (3 weeks), they worked on the therapeutic content through online modules on a web platform. In this platform, the recommendation was to advance at a rate of approximately one module per week, allowing participants to consolidate therapeutic content between sessions. During the intervention, 2 messages of support per week were sent to participants (an email and an SMS text messaging). These messages served as reminders and motivational prompts to encourage continued participation. Further details regarding the intervention structure and content can be found in the study protocol [23] and in the paper on quantitative results [30].

Procedures

The qualitative data collection procedures differed according to participants' treatment completion status. On the one hand, those who completed the treatment (n=23) were invited to participate in focus groups (in-depth group opinion interviews). This was done with the goal that interview participants had sufficient knowledge about the intervention and could provide more complete opinions on all the aspects explored. Invitations were made during the final treatment sessions and by email. Finally, a total of 18 participants accepted and took part in the interviews (subsample 1), while the remaining 5 were unable to attend.

A total of 4 focus groups were conducted, one per treatment group, each lasting approximately one hour. Meetings were held via videoconference between 2 and 5 weeks after the end of treatment. All interviews were recorded and subsequently transcribed verbatim for later analysis. No field notes were taken.

Each focus group was conducted by 2 psychologists from the research team. In total, 5 psychologists (4 women and one man) conducted the interviews; 3 of them PhD students (NJ-O, Macarena Paredes-Mealla, and Rosa Lorente-Català) and 2 were PhDs in psychology with previous experience in qualitative research (AD-G and Alberto González-Robles). All interviewers were also experienced clinicians and researchers in transdiagnostic and internet-based interventions. These psychologists were the therapists for the therapy groups in the feasibility study, so to ensure an unbiased procedure, they only interviewed those participants who had not previously been their patients (eg, Macarena Paredes-Mealla and Alberto González-Robles moderated an interview with participants whose therapy group had been led by AD-G and NJ-O). During the focus groups, there were no observers; only the 2 interviewers and the participants were present. Participants were informed that the aim of the study was to explore their opinions and experiences in order to improve the future implementation of the intervention.

Focus groups continued until all available participants from the previous study had been interviewed, corresponding to one group interview per therapy group. The research team determined that data saturation had been reached. According to Krueger and Casey [31], a commonly accepted rule of thumb is to plan 3 or 4 focus groups for a specific audience and, once they have been conducted, determine whether data saturation has been reached, that is, the point at which ideas begin to be repeated and no new information emerges. Furthermore, in the consensus qualitative research (CQR) methodology, it is recommended to include at least 8 - 15 participants to achieve data saturation [32,33].

On the other hand, participants who did not complete treatment (n=11) were invited via email to complete a brief open-ended online questionnaire aimed at exploring their reasons for dropping out. A total of 4 participants completed this questionnaire (subsample 2), while the remaining participants did not respond to the questionnaire nor to the email.

Interview Protocol

Following recommendations [31,33], a semistructured interview with 12 open-ended questions was developed to explore participants' perceptions and experiences in depth through focus group discussions. To that end, 2 doctoral students reviewed

XSL•FO
RenderX

the literature of qualitative studies on psychological interventions, including those using blended treatment formats. Afterward, each doctoral student independently designed a list of relevant interview questions, which were then shared and discussed collaboratively until consensus was reached. Finally, 2 senior researchers (AD-G; JB-L) and full professors reviewed the proposal of questions, and the final version was defined. An effort was made to reach a balance between obtaining the greatest amount of information possible and asking the fewest number of questions. The interview script can be found in Multimedia Appendix 1. Two questions that constituted the dropout questionnaire can be seen below in Table 3.

Table . Responses to the dropout questionnaire.

Participant	Responses	
	Question 1: Please write a comment indicating the reasons why you have discontinued the treatment.	Question 2: Please write a comment indicating any strategy that, in your opinion, could help to improve the treatment or enhance adherence to it.
D[a]1	"It hasn't helped me and it overwhelmed me more than it helped me."	"That it be individual or that it doesn't have so many mandatory tasks."
D2	"I wasn't fully committed, and since I thought this study was helpful in a more general way, I thought in-person therapy would be more effective for me."	"I think it would help if the sessions were in person, although that would mean the study could only be conducted with people from Castellón. Also, sometimes I didn't know how or at exactly what moments to apply the treatment, and I didn't know where to start."
D3	"I've gotten worse and I'm not able to do things on my own. My excessive lack of motivation and energy prevented me from studying and applying the tools I was given. That's why I decided I needed more specialized help, because I'm not able to follow through on my own."	"It could be improved by dedicating part of the group session to sharing problems that are currently worrying each person and talking about them a bit with everyone."
D4	"Lack of time to follow treatment guidelines."	"Well, I'm not sure, but maybe, personally, weekly in-person sessions would help me get more organized. With so many things going on in my daily life, I still had to access the platform, and by the time I realized it, 3 weeks had already passed."

[a]D: dropout.

Data Analysis

The CQR methodology [32,33], which is a structured and rigorous method for analyzing qualitative information, was used to analyze the focus group data. This methodology combines elements from constructionism and from postpositivism, and seeks to understand people's subjective reality while simultaneously reporting that reality as objectively as possible [33].

Two doctoral students trained in qualitative analysis acted as judges (NJ-O and AF-J), and 2 full professors with prior experience in qualitative research acted as auditors (AD-G and JB-L). All researchers involved in the analysis were women. One of the judges also served as an interviewer in the study.

First, the 2 judges independently identified the domains and then pooled them until a consensus was reached. The 2 auditors reviewed the proposed domains and again a consensus was reached between all of them. Subsequently, the judges independently identified and placed the core ideas from each interview into the corresponding domains and then shared them until a consensus was reached. Core ideas were also shared interview by interview with the auditors, and the necessary modifications were made until consensus was reached again. Finally, cross-analysis was carried out and the categories were established, following the same consensus procedure explained above. The information has been identified from the data (data-driven coding), although the domains partially correspond to the interview questions.

Following the suggestions of Hill et al [33], the frequency of each category was quantified to indicate its representativeness. Categories were labeled as "General" if they applied to all or all but one case (17 - 18 cases), as "Typical" if they applied to more than half of the cases (10 - 16 cases), as "Variant" if they applied to half or fewer of the cases (3 - 9 cases), and as "Rare" if they applied to only 1 or 2 cases.

The information available regarding the reasons for the dropping out of the participants who discontinued the intervention was very limited, so it was not considered appropriate to conduct a formal qualitative analysis based on an established methodology. Instead, this information was treated descriptively as complementary to the insights gathered in the focus groups. Nonetheless, the key themes identified independently by the 2 judges are briefly reported.

The COREQ (Consolidated Criteria for Reporting Qualitative Research; Checklist 1) [34] guideline was used to report the results.

XSL•FO
RenderX

Ethical Considerations

The study received approval from the Ethics Committee of Universitat Jaume I (CD/40/2019) and all participants provided informed consent prior to participation. Participant data were handled in accordance with applicable data protection regulations. Privacy and confidentiality were ensured by anonymizing all data and storing them securely on password-protected institutional servers accessible only to the research team. No personally identifiable information is reported in the manuscript. Participants did not receive financial compensation for their participation in the study.

Results

Focus Group Results

A total of 8 domains and 47 categories were identified after analysis using the CQR method, 10 of which were categorized as "Typical," 31 as "Variant," and 6 as "Rare." The results obtained from the focus group interviews are shown in Table 4, which includes several illustrative core ideas according to the frequency of the category.

XSL•FO
RenderX

Table . Results from focus group interviews.

Domains and categories (frequency)	Illustrative core ideas
Experience with the online platform	
Overall positive perception; typical (10)	• P[a]4#FG[b]1 says he has had a good experience with the platform • P13#FG3 indicates that she likes the platform • P14#FG4 notes that the experience with the platform has been very positive for her and that it has been very good
Usability; variant (9)	• P9#FG2 highlights that the platform is very easy to use • P12#FG3 indicates that she found the platform to be quite simple, intuitive, and easy to use • P15#FG4 considers that it has been a simple and easy-to-use platform
Availability of resources and content; variant (7)	• P5#FG1 appreciates having a tool with all the contents, multimedia, tasks, and monitoring logs • P10#FG2 says that one of the advantages is being able to review all the resources once the treatment is over
Positive assessment of the tools offered; variant (6)	• P12#FG3 appreciates having PDF summaries at the end of the modules because it is like a “photograph” that helps her to review • P16#FG4 adds that being able to download the PDFs and have them at hand to implement the strategies has been a key point and that she was very pleased to be able to see the progress graphs
Utility; rare (2)	• P9#FG2 reports that the platform has been very useful to complement what was seen in the sessions
Difficulties; variant (7)	• P5#FG1 agrees with the disconnection of the platform during use and indicates annoyance at not being able to access the specific part of the module where she had reached • P8#FG2 mentioned that the “calendar” tool did not indicate the days she had accessed the platform
Configuration of the blended intervention	
Positive aspects (pace of progress, frequency and duration of sessions); typical (14)	• P1#FG1 says she feels it is a good pace to advance at one module per week and feels it works • P10#FG2 notes that if the duration of the sessions had been longer, they would have been tedious, and if they had been shorter, there would not have been enough time to deal with all the topics • P13#FG3 is satisfied with the timing, she thought it was good that the contents were explained in the sessions and then they had a 3-week period to work on them, she thinks it was adequate • P16#FG4 commented that she has found the length of the sessions and the spacing between them to be very good
Unfavorable aspects (demand, frequency and duration of sessions); typical (10)	• P3#FG1 said that the duration of the sessions was a bit long and tiring • P4#FG1 expresses that he felt that the waiting time until the next session was too long and that he would have liked them to be more frequent • P10#FG2 indicates that some modules were longer and required a little more time
Ambivalent opinion on the duration of the intervention; rare (2)	• P18#FG4 points out that it is a long therapy and it is possible to fall behind and lose the motivation to continue
Individual variables that modulate the adequacy of the configuration; variant (7)	• P8#FG2 explains that during the first modules she was a bit down in the dumps and found it hard to feel motivated to do them, although this did not pose a problem for her • P13#FG3 indicates that her problem has been her lack of time to keep up with the platform and complete the tasks due to her busy life
Therapeutic contents	

XSL•FO
RenderX

Domains and categories (frequency)	Illustrative core ideas
Suitability and richness of the contents; typical (14)	• P5#FG1 notes that there is a lot of information and many strategies that are useful • P6#FG1 says that he really liked the content and finds it very useful, and agrees that everyone should have access to this syllabus • P8#FG2 considers that the contents are very well explained and that it is a very complete program • P9#FG2 considers that the richness lies in the fact that different strategies are provided and you can choose the one that best suits you
Notable usefulness of the different therapeutic components; typical (11)	• P8#FG2 points out that the strategy that has helped her the most is cognitive flexibility • P12#FG3 says that some modules have helped her more than others, especially those that are “action-oriented,” such as mindfulness and those related to positive affect • P18#FG4 highlights the usefulness of mindfulness in his case and explains that he liked the exercises very much and noticed a difference when practicing them regularly
Difficulty in choosing strategies; rare (2)	• P4#FG1 agrees that he lacked clarity on when to use each strategy
Influence of the level of prior knowledge about the contents; variant (4)	• P4#FG1 mentions feeling overwhelmed by the amount of information due to his lack of prior knowledge
Experience with group sessions via videoconference	

XSL•FO
RenderX

Domains and categories (frequency)	Illustrative core ideas
Group bonding; typical (11)	• P6#FG1 mentions looking forward to the session time to talk to everyone • P15#FG4 considers that the exchange has helped them a lot to create synergies and that the group has also served as social support • P17#FG4 points out feeling comforted because her peers said things that she had been through or she could say things that they could also identify with, and highlights that she would recommend group therapy because you feel more understood and a bond is generated
Positive aspects of the group format; typical (11)	• P6#FG1 positively values being able to share with other people and that it helps to not feel lonely and strange • P7#FG2 mentions being able to share your problems with others as an advantage • P10#FG2 adds that an advantage of group therapy is learning a lot from the other participants, seeing how others deal with and manage their problems
Initial reluctance toward the group format; variant (7)	• P7#FG2 acknowledges that when group therapy was proposed to her, she did not like the idea because she had a previous experience in which she had not felt well • P15#FG4 notes that the first day he was very nervous and self-conscious, because he didn't know what to say and if he could open up so easily
Limitations of the group format; variant (3)	• P3#FG1 believes that the group format may limit the ability to share certain topics that could be shared privately
Relevant variables for the proper functioning of the group; variant (4)	• P2#FG1 believes that they were a very heterogeneous group in terms of problems and life stages, and that it would have been better to find people with greater homogeneity
Positive aspects of videoconferencing; variant (8)	• P3#FG1 appreciates the fact that the sessions were online since she would not have been able to attend in person due to the geographical distance • P9#FG2 is grateful that the sessions were held via videoconference because wherever you are, you can connect to the session
Limitations of videoconferencing; variant (7)	• P1#FG1 believes there may be connection problems and that this will determine the course of the session • P7#FG2 agrees on the difficulty in expressing what she really felt and thought if there was someone at home
Preference for individual format; variant (5)	• P2#FG1 expresses preference for individual treatment • P13#FG3 shares that she prefers individual therapy that, since everyone has their own story, she likes more individual therapy because of the way she wants to work on her problem
Preference for in-person format; variant (8)	• P14#FG4 mentions that it was difficult for her to speak via videoconference and that she would have preferred in-person format • P17#FG4 points out that she prefers face-to-face versus online sessions, but believes that this aspect has been managed quite well in the sessions and was not bad
Satisfaction with the sessions; variant (4)	• P9#FG2 mentions that she would always prefer the group format via videoconference and would recommend it that way because it worked very well for her, and also explains that online group therapy allows her to feel less exposed and invaded as well as to have freedom to intervene and to be able to be in a familiar environment during the sessions
Positive assessment of the dynamics followed in the sessions; rare (2)	• P15#FG4 notes that reviewing previous modules at the beginning of the sessions helped refresh and remember what had been seen previously and made it easier to solve doubts and stay up to date

XSL•FO
RenderX

Domains and categories (frequency)	Illustrative core ideas
Role of therapists	
Positive evaluation of the therapists (abilities, aptitudes, and availability); typical (12)	• P5#FG1 highlights that the therapists are wonderful, they have listened to them, explained to them, and emphasized when they had doubts • P7#FG2 agrees that the therapists are very understanding, and highlights that they always gave each one time to talk and encouraged them to share • P12#FG3 highlights that the therapists were always available when you asked them and responded fairly quickly and in detail • P18#FG4 describes the therapists as excellent people, emphasizes their warmth and their excellent organization and preparation, and is very grateful for the availability of them to help and answer questions
Importance of therapist support; variant (9)	• P5#FG1 considers the role of therapists to be fundamental; she believes that they provide the professional vision, help you to talk about certain topics, and make recommendations • P9#FG2 considers that without the therapists they could not have done it and highlights that their support was vital throughout the process • P11#FG3 believes that the support from the therapists has been good and that it always goes well; he explains that the therapists have been supportive for some things that they did not know or did not know how to interpret well
Establishment of a good therapeutic alliance; variant (9)	• P6#FG1 reported having felt a bond with the therapists • P10#FG2 agrees that the therapeutic alliance can be formed in the same way (as in face-to-face individual therapy) and considers that the commitment of the therapists to the patients and also of the patients to them and to the therapy has been noted • P17#FG4 mentions that they have had a very good alliance with the therapists
Greater difficulty in creating the bond; variant (3)	• P4#FG1 considers that in group it is more difficult to create a bond with the therapists, and that a different type of bond is created compared to individual therapy
Overall assessment of the blended treatment	

XSL•FO
RenderX

Domains and categories (frequency)	Illustrative core ideas
Perceived improvements and usefulness of learning; typical (15)	• P5#FG1 indicates that she has learned many things and has found new methods that work for her • P8#FG2 considers that she has changed a lot since the intervention and that it has helped her a lot, and adds that, although her life circumstances have not changed, she feels that her coping has improved tremendously • P9#FG2 indicates that at the beginning she did not expect to get so many results and explains that thanks to the intervention her anxiety has decreased and she has been able to learn to manage crises, understand them, and know what to do when they occur • P11#FG3 agrees that treatment has been useful to him, especially finding ways to deal with the problems that arise
Satisfaction with the treatment; variant (7)	• P1#FG1 considers that the intervention has satisfactorily fulfilled everything expected in 70 - 80% • P10#FG2 is very grateful and happy for the intervention received, she values her evolution very positively and emphasizes that most of her initial objectives have been met
Recommended treatment; variant (5)	• P7#FG2 notes that she would recommend this treatment to other people with emotional disorders • P12#FG3 mentions that she would recommend the intervention and would invite anyone with an emotional disorder to do it because she believes it is really useful
Advantages; typical (11)	• P11#FG3 considers that the combination of online content and sessions with the therapists has been very helpful and worth doing this way • P8#FG2 highlights the temporal flexibility that this format allows and adds the spatial flexibility as an advantage, in the sense of being able to do the intervention from any location, which allows it to continue despite inconveniences or unforeseen circumstances such as COVID • P18#FG4 adds that the main advantage he enjoyed was accessing the treatment from another country, explaining that otherwise he could not have done it and that it does not exist in his country, so for him it is a great advantage
Disadvantages; variant (9)	• P5#FG1 indicates that the treatment takes too long because you have to know all the strategies to know which one you can use in your case • P6#FG1 indicates that the treatment was more general and does not focus on what she believes her problem is • P10#FG2 believes that this intervention format, in comparison to traditional therapy, does not allow to delve into the why of the problems or to understand where they come from at the individual level
Elements that help to maintain adherence	

XSL•FO
RenderX

Domains and categories (frequency)	Illustrative core ideas
Therapist support; variant (8)	• P13#FG3 adds that the fact that 2 psychologists explain the topics helps to continue, she considers it as the main thing. A total of 2 psychologists explain the topics help to continue; she considers it as the main thing • P18#FG4 points out that what helped him not to drop out despite being behind was that the therapists continued to invite him to attend the sessions, which gave him confidence
Therapeutic resources and materials; variant (6)	• P13#FG3 mentions that the videos have helped her a lot to get motivated and continue with the treatment because she saw herself reflected in it • P17#FG4 indicates that in moments of overwhelm or low mood, having the decisional balance table visible helped her a lot to move forward
Having graphs that represent progress; variant (5)	• P1#FG1 highlights the progress graphs on the platform as motivating elements • P16#FG4 mentions that seeing her week-by-week progress on the platform's graphs helped motivate her to continue with the treatment
Having scheduled sessions; variant (5)	• P9#FG2 highlights that the schedule of having to meet every 3 weeks for the sessions is what helped her stay consistent with the tasks, keep going, and complete the treatment • P12#FG3 highlights the fact that there are established sessions as a motivating element that helps to keep going
Commitment to the group; variant (3)	• P6#FG1 indicates that her groupmates are one of the elements that helped her the most to continue with treatment
Perceived impact of the intervention; variant (3)	• P18#FG4 mentions that what motivated him the most was seeing that the tools were working, that watching the videos and applying the strategies really had an effect on his mood
Readiness to change; rare (2)	• P5#FG1 highlights the desire to change as a person and to find new ways of doing things as a motivating element
Having reminder messages; rare (1)	• P12#FG3 adds that the constant reminder messages, although subtle, have been very important for her in keeping going
Proposals for improvement	

XSL•FO
RenderX

Domains and categories (frequency)	Illustrative core ideas
Modifications to the platform; variant (6)	• P1#FG1 suggests that the platform could be more interactive and incorporate activities that allow decision-making to be put into practice • P16#FG4 notes that she has missed a forum or a similar tool to be able to use between sessions and try to make the group closer
Adjustments in the configuration (pace of progress, duration and periodicity of sessions); variant (5)	• P3#FG1 suggests that sessions of one or one and a half hours and weekly frequency would be more bearable • P11#FG3 notes that it would have been better if the time between sessions had been adapted according to the content to be worked on in the corresponding modules
Suggested changes regarding the sessions; variant (3)	• P1#FG1 proposes doing some of the group sessions in person
Greater individualization of the intervention; variant (4)	• P8#FG2 suggests having, from time to time, an individual space with the therapist for about 10 minutes to delve deeper into individual difficulties
Suggestions for improving adherence; variant (7)	• P3#FG1 suggests that individualized weekly follow-up, including by email, could help improve treatment adherence • P18#FG4 proposes leaving less space between sessions and doing them every 2 weeks to improve adherence for people who need more follow-up or who are a little disorganized

[a]P: participant.

[b]FG: focus group.

Experience With the Online Platform

This domain includes all experiences and opinions related to the online platform. It does not include what is clearly identified as proposals for improvement of the platform (included in the domain "proposals for improvement"). The categories found are described and exemplified below.

Overall Positive Perception (Typical)

Participants expressed an overall positive experience with the use of the platform. We can see an example in the following excerpt:

> *Let's see, for me the experience (with the platform) has also been very positive. And as my colleagues say, you had it there and you could go in and do it little by little. And the truth is that yes, it was very good. I liked it.* [Participant (P) 14#focus group (FG) 4]

Usability (Variant)

Participants highlighted the simplicity of the platform and its ease of use. An example of this is the following extract:

> *As for the use of the platform, I found it to be a fairly simple and intuitive platform, uh well, easy to use, also… all the different elements were quite clear, simple.* [P14#FG4]

Availability of Resources and Content (Variant)

Some participants highlighted having the contents accessible through the platform as positive.

> *I imagine in individualized therapy it won't work the same way, that it will be totally different, they won't give you… an application where you can read all the information, watch videos, do exercises, monitoring logs… so in that sense I think that… I liked it, it works well.* [P5#FG1]

Positive Assessment of the Tools Offered (Variant)

Some participants appreciated the different tools found on the platform (eg, videos, downloadable PDFs, review tool for revisiting completed modules, and progress graphs).

> *Eh, I'd add that for me a point that I really liked about the application is, eh, the section called "Weekly evolution" or something like that, the fact of physically seeing, that is, graphically seeing the evolution you've been having week by week. […] And besides, as P15#FG4 had said at the beginning, being able to download the PDFs later to have them at hand to be able to put them into practice in case you forget something or whatever, I also think that has been a key point.* [P16#FG4]

Utility (Rare)

Some participants referred to the usefulness of the platform from a clinical point of view. The following fragment can serve as an example:

> *And… I think it (the platform) has helped us a lot to complement everything that was discussed in each session.* [P9#FG2]

Difficulties (Variant)

Participants mentioned perceived technical difficulties or aspects that they would have liked to be another way. We can see an example below:

XSL•FO
RenderX

> *Eh, it's true that when you went a while without using it, it disconnected immediately, you had to re-enter the program and what really bothered me the most was that maybe I didn't have time to see the entire module and I only viewed maybe a couple of pages, and when I connected another day and re-entered, it wasn't saved where I had left off, you had to start the module again and look for where you had left off. For me, this is the inconvenience that has caused me the most discomfort, but otherwise, in general, very well.* [P5#FG1]

Configuration of the Blended Intervention

This domain includes all aspects related to the structure of the blended intervention and how its different components were combined or organized.

Positive Aspects (Typical)

Most participants identified positive aspects of the configuration regarding the pace of progress on the platform, the frequency of the sessions, and the duration of the sessions.

> *As for me, eh, I think the same as you. I thought they were good, the sessions were also quite good. If they had lasted longer, they would have been a bit tedious, and if they had been shorter, there would have been a lack of topics to address.* [P10#FG2]

Unfavorable Aspects (Typical)

Participants also identified unfavorable aspects of the configuration in relation to the demand of the pace of progress as well as the periodicity and duration of sessions.

> *It seemed like too much time to me. I would have liked a little more, closer together, but well, I understand that that's what we had, but for me the group sessions were what helped me the most, and I would have liked them to be more frequent, and the duration of 2 hours seemed fine to me.* [P4#FG1]

Ambivalent Opinion on the Duration of the Intervention (Rare)

Some participants expressed their opinion about the duration of the treatment. While one person pointed to the duration as a positive aspect, another person pointed to it as a more unfavorable aspect:

> *The fact that it has lasted so long, for so many months, I think it has helped us a lot, yes, it has not been short, you see your progress month after month.*
>
> *[P9#FG2]*
>
> *That can happen, you can fall behind because it's a long course, a long therapy, so you lose a little bit, you can also lose the motivation to keep going if you're too far behind.*
>
> *[P18#FG4]*

Individual Variables that Modulate the Adequacy of the Configuration (Variable)

Some participants also mentioned personal variables that may influence the appropriateness of the configuration in each case. The following excerpt is an example of this:

> *My problem is that lately I've had a very busy life, but just as I haven't been able to do my homework, I had to go to the physiotherapist and I haven't gone, so I'm not justifying, but well, for failing something, it has been my time.* [P12#FG3]

Therapeutic Contents

This section includes all the experiences and opinions regarding the treatment modules or therapeutic strategies, which were taught both in the group sessions and on the online platform. The categories found are described and exemplified below.

Suitability and Richness of the Contents (Typical)

Most participants emphasized the adequacy of the contents as well as the richness of the program. We can see an example in the following extract:

> *Personally, I thought the program was super complete, I really liked the fact that different tools and methodologies are provided because that way everyone chooses, tests, and at the end looks or explores what works best for him/her, and it is very well explained.* [P8#FG2]

Notable Usefulness of the Different Therapeutic Components (Typical)

Participants highlighted those activities, strategies, or therapeutic components that they had found most useful in their case.

> *Yes, in my case the same, mindfulness. I think, like everyone else, I really liked the exercises, I did the exercises as they came and yes, I noticed a difference. After about 3 or 4 days, I noticed a difference, and I think that if you do it constantly, it really helps you to relax. What other one? I also liked the cognitive flexibility one a lot, to see a fact from different perspectives, because one normally goes with the belief… that one already has programmed, right? The beliefs, distressing or anxious beliefs usually. So, trying to see it from different perspectives, that's very useful. Very useful. And the identification of maladaptive behaviors as well, that normally one acts automatically, right? Again, according to the anxious perspective that you have, and you don't realize that those behaviors are generating again the anxiety and the anxiousness, reinforcing the anxiety, the anxiousness. So those 3. I would highlight 3 very useful tools.* [P18#FG4]

Difficulty in Choosing Strategies (Rare)

Some participants mentioned difficulty in knowing when or in which situations to use each of the strategies. An example of this is the following intervention:

XSL•FO
RenderX

> *I haven't been able to grasp, understand everything and so on, and especially what P5#FG1 was saying, that I don't know when to use each thing, and that's what really… I fell short on, I missed that part I think, of knowing when to use each thing and so on.* [P4#FG1]

Influence of the Level of Prior Knowledge About the Contents (Variant)

Some participants indicated how their prior knowledge of the treatment contents influenced their evaluation. While some emphasized the importance of having some background, for others, having more prior knowledge made the contents seem somewhat repetitive:

> *What happened to me was that… it overwhelmed me a little. I think the information was very good and all, but the thing is that since I didn't have much (prior) knowledge, I suppose it overwhelmed me a little* [P4#FG1]

> *Eh, for me it was a bit the opposite, I informed myself a lot beforehand, so there were many things that felt very obvious or very repetitive to me. Well, I guess it depends on the person and their knowledge about the topic. And overall, well, yes, I obviously learned a lot of new things, and I also enjoyed that part… but yeah, in general, some of the information felt a bit boring because… I already knew it beforehand.* [P3#FG1]

Experience With Group Sessions via Videoconference

All experiences and opinions related to group sessions by videoconference are included here. Aspects related to the frequency and/or duration of the sessions, already included in the configuration domain, are not included in this section.

Group Bonding (Typical)

Participants reported a connection or bond with the group, which they identified as a source of support or understanding and acknowledged looking forward to the sessions to talk to everyone and share. The following excerpt is an example of this category:

> *It comforted me because sometimes my groupmates would say things that I had also gone through, or I could say things that they could identify with. So, I think it has been quite good and it has changed my perspective. And what's more, I would recommend it because you feel more understood and maybe there's a bond of mmm I don't know, there's a bond there.* [P17#FG4]

Positive Aspects of the Group Format (Typical)

Participants highlighted advantages or positive aspects of the group format, as exemplified in the following fragment:

> *I think you also learn a lot. I mean, I think there are things I wouldn't have learned if the others had not been there, because… everyone has their own experiences and their own way of dealing with things, and seeing how other people deal with the problem, even if it's not included in the modules, or whatever, also helps a lot, because you can try it that way or not, and the truth is that in an individual therapy you don't get that, that is, having other examples from real life, seeing a person who is doing it and living it, it helps a lot.* [P10#FG2]

Initial Reluctance Toward the Group Format (Variant)

Some participants admitted to being uncertain about the group format and expressed initial feelings of fear, strangeness, or nervousness about having to speak in front of other people.

> *At the beginning, before starting the group therapy, I was looking for individual therapy, but they recommended me 'there is this group that is going to be formed, I think you would be interested in joining, try it, you don't lose anything', obviously. And it's true that the first day I was very nervous, I was very self-conscious, like 'I don't know what to say, I don't know if I can open up so easily'. But in the end, you mention your problem, you see that another person comments on his/her problem, there are points in common, what I can say can help a person or what another person says can help me, and I believe that throughout the sessions it has also been seen that we were talking more and more, it was less difficult for us to start talking.* [P15#FG4]

Limitations of the Group Format (Variant)

A few participants also highlighted the possible limitations of the group format, as we can see in the following excerpt:

> *I haven't had experience of other types of therapy either, but I think that there are certain things that you might not be able to share in a group and that maybe in private you would be able to. I don't know. That's what I believe.* [P3#FG1]

Relevant Variables for the Proper Functioning of the Group (Variant)

Some participants mentioned some important factors that may influence the group therapy experience, such as how the particular group gets along, how homogeneous the group is, or how participative people are:

> *Well, I think that ehh I agree with everything my groupmates are saying, but I think that we were a very heterogeneous group. Ehh each one had a different problem, was in a different stage of life,… so… […] Maybe it would have been better to find groups of people among whom… people among whom there would have been a greater homogeneity, maybe.* [P2#FG1]

Positive Aspects of Videoconferencing (Variant)

Participants highlighted advantages or positive aspects of the videoconference, as exemplified in the following extract:

> *I really appreciated the fact that they are by videoconference, because each one has his/her own life rhythms and schedules, and no matter where you are in the world, you can connect 2 hours for the session.* [P9#FG2]

XSL•FO
RenderX

Limitations of Videoconferencing (Variant)

Participants also highlighted possible limitations of conducting the sessions via videoconference, as we can see in the following intervention:

> *The same as P8#FG2, that many times I have wanted to express what I felt, but like for example, I didn't know if my roommates were there, I couldn't say everything I really thought for fear of being heard.* [P7#FG2]

Preference for Individual Format (Variant)

Some participants expressed a preference for individual therapy, as can be seen in this case:

> *Well, I'm going to give my opinion first because I've gone through individual and group, and I think it is, I can be more objective. I prefer individual. I also like to have met mm the groupmates, but it's true that as everyone has their own story… I like it more individual. I do… but that doesn't mean group was bad for me, nooo, but I liked individual more, maybe because of the way I want to work on my problem.* [P13#FG3]

Preference for In-Person Format (Variant)

Some participants expressed a preference for the in situ format, as exemplified in the following excerpt:

> *Eh. I, like P15#FG4, was also looking for something individual, and they offered me this and I said I'm going to try it. And the truth is that it's good, but I would prefer face-to-face. […] In videoconference I have felt that it's more difficult for me to talk. The truth is that I have found it difficult to speak. Maybe it's personal. I don't know, it's because I'm like that, but the truth is that face-to-face would have been better.* [P14#FG4]

Satisfaction With the Sessions (Variant)

Some participants highlighted their satisfaction with the sessions received, as we can see in this fragment:

> *I would add the fact that I would always prefer to do it online and in group rather than face-to-face and individual because in the individual sessions I've done I feel very exposed, very invaded, like now is the moment and you have to talk, say everything, you have this schedule, everybody, you know? Doing it in a group online, things come up, right now I don't feel like talking, I listen to the other, you're in your space, in your home, you're not in a psychologist's office, it's different, I… I think I would always recommend doing it this way. It worked really well for me.* [P9#FG2]

Positive Assessment of the Dynamics Followed in the Sessions (Rare)

Some participants positively valued the dynamics followed in the development of the group sessions via videoconference, for example:

> *And also one thing we did in the sessions, which I imagine we'll discuss later, was that at the beginning of the session, for example, the beginning of session 2, we reviewed a little bit what we had seen in the previous modules, so that also helped you refresh and remember what you had seen, in case there was something you had missed or you had doubts about, that also helped you to keep up to date. […] It's also true that the dynamics within the sessions made everything quite, quite free-flowing and the feeling of tiredness wasn't so common.* [P15#FG4]

Role of Therapists

This domain encompasses all the experiences, assessments, and opinions regarding the role of the therapists in the intervention, including therapeutic support received, therapeutic alliance, etc. The identified categories are outlined and illustrated below.

Positive Evaluation of the Therapists (Typical)

Participants positively rated the therapists in terms of their abilities, their aptitudes, and the availability shown throughout the intervention.

> *Yes, apart from that, I totally agree with my groupmates, for example, they always gave each one of us our own time to talk, if for example at a certain moment we didn't have time to talk, they always told us, 'oh, how about you?', and I don't know, they are very understanding, really, we are very grateful that they have been there.* [P7#FG2]

Importance of Therapist Support (Variant)

Participants emphasized the importance of having therapeutic support, that is, the presence of the therapists. We can see an example in the following excerpt:

> *I think it's fundamental, I mean, the role of therapists in this case. They are the ones who open you up to talk about certain topics, who help you, who… advise you, who give you their point of view, who is the… the wise one, so to speak… So it's a fundamental part of it.* [P5#FG1]

Establishment of a Good Therapeutic Alliance (Variant)

Participants felt that a bond with therapists could be established in this format and reported having established a good therapeutic alliance.

> ***Interviewer:*** *Did you feel that this alliance was able to develop in the same way that perhaps you would have developed it face-to-face, though, in this online version?*
>
> ***P9#FG2:*** *I think so.*
>
> ***P10#FG2:*** *I also, I mean, even though they don't know absolutely everything about me and everything that has happened to me, I think I felt that commitment, didn't I? I mean, that commitment from them to us and also from us to them and to the treatment. So, yes.*

XSL•FO
RenderX

Greater Difficulty in Creating the Bond (Variant)

Some other participants felt that it was more difficult to create a therapeutic bond in group therapy than in individual therapy or that this bond is different. An example of this is the following intervention:

> *I also agree, it's another type of bond, no, I don't think it's the bond that… […] It's another type of bond and regarding the ease of creating it or not, I think it's more difficult in a group because there are many people, it's more risky eh with, with one person alone it's much easier, isn't it?, more risky, but well, if it works, it's also very good.* [P4#FG1]

Overall Assessment of the Blended Treatment

This section includes all experiences, opinions, and assessments regarding the intervention as a whole, including its impact or effects, its advantages and disadvantages, etc.

Perceived Improvements and Usefulness of Learning (Typical)

Most participants highlighted their improvement after the intervention as well as the usefulness of the learning achieved during the treatment in their lives, as exemplified below:

> *Ehh, in my case, I also agree with my groupmates, especially, let's see, I had enough knowledge, I think, ehh… to practice and all that, but I hadn't taken the time to do it, I mean, the theory was 'ok, very good,' but I had never taken the time, and it seemed to me that the treatment has been like an opening key to something that was already there, you know? Mm… especially the fact of continuing to practice, of saying 'jeez, it's fundamental…,' like going to the gym, right? to keep practicing and so on. It has helped me a lot to become aware that… that I need it, and… and I have also changed a lot from before having started it to now, it has helped me a lot. […] I would like to add that in my case the situation hasn't changed, I mean, there are some groupmates whose situation has changed and so on, but in my case it has not, that's why I'm happier in life, because the situation hasn't changed, but I've improved and I'm handling things much better.* [P8#FG2]

Satisfaction With the Treatment (Variant)

Some participants also emphasized their satisfaction and positive feelings toward the treatment received, as can be seen in the following case:

> *I feel that… you can't cure a problem, none, in a few online sessions and that you have to work and for that what they give you are tools to achieve it individually and then, in that, I think it has helped me a lot, I think they have given me many tools to know how to manage what happens to me, more than cure me, and also… ehh well, the importance of emotions and what happens to me in general I have also learned a lot and, and well, I think that 70-80% has satisfactorily fulfilled everything what, what I expected.* [P1#FG1]

Recommended Treatment (Variant)

Some participants went a step further and indicated that they would recommend the treatment to others with emotional problems.

> *It is true that it requires more work and more time, but I would recommend it. The truth is that it is quite useful, and I would invite anyone who has any emotional disorder to do this treatment because the truth is that it is really useful.* [P12#FG3]

Advantages (Typical)

Most participants mentioned advantages of this type of blended treatment, which combines group sessions delivered via videoconference with self-applied work through an online platform. The following excerpt is an example of an outstanding advantage:

> *Well, the advantages. Mmm. I think the convenience, as my groupmates mentioned, it's very easy to attend the sessions, very easy to access the platform, the course. […] Um, another thing. The main advantage that I enjoyed was having taken this course from another country. I wouldn't have been able to take it. I wouldn't have. I wouldn't have. It doesn't exist in Mexico. So, well, for me it's a great advantage.*' [P18#FG4]

Disadvantages (Variant)

Participants also mentioned disadvantages of this type of blended intervention, among them:

> *I think if I had been alone with one of them (therapists), I think it would have been, I would have focused more on, on my problems or I don't know… This was more, more general, a help, […] it has been very helpful, but […] it doesn't focus on what I think my problem is.* [P6#FG1]

Elements That Help to Maintain Adherence

This domain includes participants' opinions and experiences regarding the aspects of the intervention that helped them to stay motivated, continue progressing with the treatment, and promoting adherence to it.

Therapist Support (Variant)

Participants identified therapeutic support as a key factor that helped them maintain motivation and continue with treatment.

> ***Interviewer:*** *Did you notice any other aspects? Because it's a procedure that involves some work at home, then the sessions, eh, there's a lot of work there. Is there anything that's helped you stay motivated and keep going? What do you notice the most?*
>
> ***Interviewer:*** *From any of the aspects of the treatment.*
>
> ***P13#FG3:*** *Well, obviously everything helps. Having 2 psychologists explain the topic to be addressed, of course that helps. […] I think it's the main thing.*

XSL•FO
RenderX

Therapeutic Resources and Materials (Variant)

Participants identified therapeutic resources and materials, such as the modules, the videos, and the decisional balance table, as important factors for adherence.

> *Well, I was going to say that I, for example, haven't had the thought of quitting, not at all, but it's true that there was a time when I was really swamped with exams and many other things […] So it's true that in that moment of overwhelm, […] the table we made at the beginning of the pros and cons, I think it was, of starting treatment, of changing, and all that, helped me a lot. Well, for me, the table was, I mean, I made it to my liking, so I have it posted and that helps me a lot when I'm feeling more down or whatever, having seen it and saying, look, here's the reminder of this. Well, that's been very good for me.* [P17#FG4]

Having Graphs That Represent Progress (Variant)

Participants highlighted that being able to see their evolution reflected in the progress graphs of the platform was a motivation element for them. The following is an illustrative example of this category:

> *Seeing the progress graphically week by week is something that, at least for me, also helped motivate me to continue with the treatment.* [P16#FG4]

Having Scheduled Sessions (Variant)

Participants highlighted having established sessions as an element that helped them to continue with the treatment and do the homework tasks.

> *I think it has been the constancy, seeing each other, seeing each other every 3 weeks, I think it has been that, wait, today I don't feel like doing it, okay, tomorrow neither, but in 3 weeks these people are going to ask me what… 'what have you worked on?' you know? So, it's like a duty, like going to class, for me that has helped a lot. It was like I have to do my homework for college or my thesis or whatever, I have to do my psychology homework, I mean the psychologist, where I'm going, the treatment, because I started it and I have to finish it, and in 3 weeks I have to see them, so it's been like a constant and that's how it's been every month.* [P9#FG2]

Commitment to the Group (Variant)

Some participants also identified groupmates and commitment to the group as an element that helped them maintain motivation and not drop out of treatment, as we can see in the following case:

> ***Interviewer:*** *What elements of this treatment helped you to stay motivated, to stay connected, to keep moving forward in this treatment?*
>
> ***P6#FG1:*** *In my case, more than the modules and all that, the therapists and my groupmates the most.*

Perceived Impact of the Intervention (Variant)

Some participants highlighted that the perceived impact of the treatment, that is, noticing the effect it had on them or the progress they were making, helped them to continue with the intervention.

> *In my case, what motivated me the most, even though as I mentioned before, I was behind in the course, was to see that the tools worked. Knowing that the videos and applying the tools that were mentioned in them really showed an effect on my mood, at least in the short term, I mean, because we know that this problem is long, it's not a, you have to practice these tools in the long-term to notice a change, but a change was noticeable in the short-term, at least in the short-term. And that was quite motivating for me.* [P18#FG4]

Readiness to Change (Rare)

Some participants mentioned the readiness to change and commitment to oneself as a motivating element.

> *Also the motivation of wanting to change ehh yourself as a person, yourself… and knowing how to control yourself, knowing… well, to control! Okay, it's accepting your emotions and ways of facing the things that happen to you, well, for me to have the motivation to find new, new ways of doing things, that were not the ones I used to do previously.* [P5#FG1]

Having Reminder Messages (Rare)

One participant also noted how important the weekly motivational reminder messages they received were to her for continuing.

> *'Well, let's see, in my opinion the fact that there are established sessions and also the constant, constant online reminders, I mean, by email, which I think have been very important, although subtle, but for me they have been very important.* [P12#FG3]

Proposals for Improvement

This domain refers to any suggestion, proposal, or idea to improve the intervention in any way from the patients' point of view.

Modifications to the Platform (Variant)

Participants suggested some modifications to the platform, aimed at improving the current functionality or incorporating new tools. We can see an example in the following extract:

> *P1#FG1: Yeah, well, my, my idea is, it's also related to the platform, but I've also missed like the interaction, I mean, it was like just listening to the video, it would be interesting to do activities in umm like games… from the video and it explained to you Maria's situation, for example, Maria has anxiety and this. Like a kind of 'What would you do in this situation if you were Maria?' Because that also puts you a little bit in the role of what you're going through, although, although you don't necessarily*

XSL•FO
RenderX

have anxiety, but… something like more interactive with the platform that also helps you to understand a little bit the type of decisions, or if it's a good decision, that kind of thing, it would be interesting ehh that the platform itself motivates a little bit more and not just listening to a video.'

Adjustments in the Configuration (Variant)

Participants proposed some adjustments regarding the pace of progress on the platform and the duration and periodicity of sessions, as exemplified below:

> ***Interviewer:*** *Do you think that the duration of the group sessions that you had, 2 hours of online sessions every 3 weeks, is adequate? That is, 2 hours every 3 weeks, did you see it as adequate or how did you feel that temporality?*
>
> ***P11#FG3:*** *I think it's fine. To be there for 2 hours, ehh it's really like attending a therapy session or even a class.*
>
> ***Interviewer:*** *And what about it being every 3 weeks? Did you find that spaced out enough, too spaced out?*
>
> ***P11#FG3:*** *Mm maybe… it depends, it depends on the subject that we were doing… more time or less time would have been better.*

Suggested Changes Regarding the Sessions (Variant)

Some participants made suggestions about the sessions, proposing some additional sessions or that some of the sessions be held in person.

> *Even though it's online, I like it, but later, when we were doing about 2 or 3 sessions, I would also have liked it to be in-person format because well I actually miss in-person attendance, because I got used to it at the university and then when we went back to in-person, the difference was also noticeable, so, well, doing at least one or 2 in person I think it would also help a lot.* [P1#FG1]

Greater Individualization of the Intervention (Variant)

Some participants emphasized the importance of personalizing or individualizing the intervention further, as we can see in this case:

> ***P8#FG2:*** *Maybe what was missing… I don't know, a proposal also, like maybe 10 minutes, once in a while, not every week, obviously it cannot be done, to go deeper into the problem or more individual difficulties. You know? […] It's true that there were sessions when we were asked about the problems, 'how did you deal with them?' and each of us talked a little bit. But, yes, I was referring maybe to one more difficulty that you have, and you can't share with the group or whatever, no… maybe not frequently, but it would be good.*
>
> ***Interviewer:*** *It would be like a space for, an individual space with the therapist, right? To share the things that in the group are not…*
>
> ***P8#FG2:*** *Exactly.*

Suggestions for Improving Adherence (Variant)

Participants made proposals to improve adherence to treatment and reduce dropouts. This category includes proposals specifically linked to enhancing adherence, although some of them mention, for example, changes in the configuration or suggestions for greater individualization.

> ***Interviewer:*** *Great, okay. Emm can you think of any strategies that you can or we can recommend, we can include improving adherence, so that people can get to finish this treatment, prevent them from dropping out of therapy, that you want to suggest to us?*
>
> ***P3#FG1:*** *I don't know if it could be done, or if it would be useful, but maybe ehh as a kind of weekly follow-up by the, by the therapists, even if it's by email, asking you how the week went or… I don't know.*

Perspective of Dropouts

Participants who discontinued the intervention identified various reasons for dropping out. Participants mentioned treatment-related variables (lack of results, treatment-related stress, or difficulties applying the strategies), person-related variables (low involvement, lack of motivation and energy, or time constraints), and the need for other types of help (in-person or more specialized professional help). As strategies to improve adherence to the treatment, participants suggested adjustments in the format and configuration of the sessions (individual and in-person or weekly) and other aspects (less homework tasks and more time to discuss individual problems). Table 3 shows the literal responses to the dropout questionnaire.

Discussion

Principal Findings

The goal of this study was to explore in depth the experiences and opinions of patients who received a transdiagnostic CBT intervention for ED, delivered in a blended and group format, in order to understand their degree of acceptance and improve the future implementation of these types of interventions.

Overall, the results of this study indicate a general satisfaction with the transdiagnostic intervention, consistent with prior evidence on the feasibility and utility of blended psychological treatments [26,35]. Positive perceptions were observed across multiple dimensions, including the usability of the web platform, the perceived richness and relevance of the therapeutic contents, the added value of group format, and the perceived support provided by the therapists. Furthermore, participants generally reported perceived improvements following the intervention and emphasized the advantages of this blended format. Consequently, this intervention appears to be both feasible and well-accepted by patients for the treatment of ED. The results of this study also provide relevant information on specific measures that could be adopted in the future to further optimize treatment and minimize dropout rates.

Generally, participants reported an overall positive user experience with the platform, particularly regarding its intuitive design and the availability of multimedia materials and contents.

XSL•FO
RenderX

They also appreciated the different tools offered by the platform. These results are consistent with the previous feasibility and acceptability study in which the patients included in this study participated [30]. In the aforementioned study, adequate adherence to the platform and high-excellent usability according to the System Usability Scale [36,37] scores were observed. Previous studies on other treatment protocols administered through the same web platform also reported high usability scores [38,39]. In addition, these findings align with research suggesting that ease of use, clarity, usability, attractive appearance, and perceived utility are critical factors in the acceptance and engagement with digital mental health interventions [40,41]. Participants also identified some technical difficulties with the platform, such as progress within each module not being saved and disconnection of the platform after a period of inactivity that required them to log in again. However, this feature was intentionally implemented as a security measure to protect users' privacy if the platform was accidentally left open. A future challenge will be to maintain robust security standards that enable patients to use digital devices confidently and safely [42] without compromising usability and acceptability. Technical problems have also been reported in other studies of blended interventions [12,26,27], suggesting that the difficulties identified in this study are common in this field.

The mixed evaluations of pace of progress on the platform reflect a common challenge in online and group interventions: balancing standardization with individualization. While participants were generally satisfied and considered the weekly pace adequate, others wished for a more flexible rhythm, particularly for modules involving complex skills such as emotional avoidance or acceptance. Prior literature suggests that personalizing the configuration of blended treatments may enhance adherence and treatment benefits [43]. As the systematic review by Erbe [14] indicates, there is no single optimal ratio between face-to-face and online components in blended interventions; thus, patient-specific customization may be key. The length of the sessions was considered appropriate by most participants in focus groups, although a few felt they were slightly long. There were also differing opinions on session frequency; some found the timing adequate to work on the content of the modules, while others would have preferred more frequent sessions or less time between them. These findings are in line with the heterogeneity of session lengths reported in the literature. For example, Osma et al [44] implemented 2-hour unified protocol group sessions, while Schuster et al [45] noted that therapists were concerned that 90-minute weekly sessions did not allow enough time for discussion. This variability of configurations in the literature highlights the need to customize the specific combination of in-person and online elements, as well as the timing. This personalization of the intervention has been emphasized by Fernández-Álvarez et al [12] and Patel et al [13], especially within transdiagnostic frameworks where symptom profiles and levels of functional impairment can vary significantly. Although not asked directly, a couple of focus group participants talked about the duration of the treatment. While one participant felt that the duration had allowed them to track progress month by month, another participant believed that it could cause a loss of motivation. Interestingly, participants mentioned personal variables that had an impact on the optimal fit of the planned configuration. These insights suggest that individual factors, such as workload, time management challenges, or fluctuating symptoms, may impact the optimal engagement and highlight the value of adapting to the patient's life circumstances.

Participants highlighted the appropriateness and richness of the content. They valued the material positively, indicating that it is a very comprehensive program and suggesting that everyone should have access to the modules. They stated that the content was well explained and relevant to their issues. They also pointed out which content had been most useful for their specific cases. Mindfulness was the strategy mentioned by more participants, followed by activities belonging to the positive affect component—such as value-based behavioral activation or the identification of psychological strengths—and emotional avoidance and emotion-driven behavior modules. It may be thought that patients could be reluctant to adopt strategies like mindfulness, as avoidance is a common trait among people with ED [46]. Therefore, it is very important to note that participants responded positively to mindfulness and felt it was very helpful, as it was the strategy most frequently mentioned. They also highlighted the exposure component, cognitive flexibility, and relapse prevention as other useful strategies, providing a comprehensive overview of the entire program. Overall, these results align with previous quantitative findings. When participants were asked how useful each module was on a scale from 0 to 10, the average scores were 7 or higher for all modules [30]. Another important point from participants' feedback was the variability in prior knowledge and its impact during treatment. Some participants felt unsure about which strategies to use in specific situations, while others sometimes found the content repetitive due to their previous familiarity with similar material. Previous research emphasizes the need to tailor the complexity of therapeutic content to participants' prior knowledge and learning styles [47,48]. These findings suggest the value of preassessment tools to gauge prior familiarity with CBT concepts, which could help guide each patient's pacing or personalized use of modular content.

Regarding the group format, participants noted some limitations and expressed initial reluctance toward this approach. However, most participants recognized positive aspects, such as learning from others and feeling a bond with the group. This aligns with literature highlighting the therapeutic value of shared experiences, group cohesiveness, and peer learning among individuals with similar issues [20,49,50]. Numerous studies on the acceptability of transdiagnostic interventions report good acceptance and high treatment satisfaction, and this is held in a group setting [51,52]. In terms of videoconferencing, participants identified limitations like potential connection issues and challenges in finding a private space. However, they also highlighted several advantages, such as the convenience of joining sessions from anywhere and the fact that emotions surfaced in the same way through videoconferencing as in face-to-face therapy. These benefits and additional advantages of this form of psychotherapy have been noted in previous literature [49,53]. This research indicates that, in general, patients have a positive attitude toward videoconferencing as a

XSL•FO
RenderX

therapy option [54-56]. In line with this, some participants in this study reported satisfaction with the sessions and positively valued the dynamics of group interactions. However, despite these benefits, some participants expressed a personal preference for individual or in-person therapy. This preference was one of the recruitment challenges faced in the treatment feasibility study [30]. Previous literature has also reported that some individuals prefer in-person groups [49], and many patients with ED often favor individual face-to-face therapy [57]. Therefore, it is crucial to explore ways to make more acceptable those interventions that enhance the cost-benefit balance.

Most participants highlighted the qualities of the therapists and their availability, such as their explanations, their empathy and closeness, their expressiveness, or their response to doubts that arose, among other issues. Additionally, in general, participants emphasized the importance of the therapist's role, indicating that it was fundamental and that without therapists they could not have done it. Likewise, they reported having established a good alliance with the therapists in the same way as in classical psychotherapy. This is consistent with previous studies in which the importance of support from therapists was highlighted [12,13]. The establishment of a strong therapeutic alliance, even in an online group format, was reported by several participants, consistent with previous research demonstrating the feasibility of alliance building in digital contexts [58]. Furthermore, research shows that patients and therapists experience a good therapeutic alliance with videoconferencing therapy, comparable to that obtained with face-to-face therapy [55,59-61]. Studies on blended interventions also show that it is possible to establish a good therapeutic alliance between therapists and patients in this format [26,35].

Participants identified several general advantages of the blended intervention they received, notably temporal flexibility and, particularly, spatial flexibility, which allows treatment to be accessed from anywhere without the need to travel. Additionally, the combination of online modules and sessions was highlighted as a positive aspect, with both elements considered useful. These advantages have been noted in other studies of blended interventions. For instance, in the study by Etzelmueller et al [26], patients also emphasized the spatial flexibility and convenience of working from home, and they found the combination of online modules and synchronous video-based sessions to be beneficial. The main disadvantage identified by participants in focus groups was the more general nature of the treatment, which limits in-depth exploration of individual circumstances or specifics. This is related to the transdiagnostic nature of the intervention and likely its group format. Despite this, most participants reported perceived improvements and the usefulness of the skills learned in their daily lives, noting that they have discovered methods that work for them, that they can better cope with difficult situations or their own symptoms, or that they feel better overall. This is consistent with the preliminary efficacy found in our previous study [30], where significant reductions in anxious and depressive symptoms were obtained at posttreatment and follow-up compared to baseline, and where significant changes were found in other clinical measures, such as negative affect, neuroticism, emotion regulation, or quality of life. In other blended interventions mentioned above [26], patients also reported positive treatment effects. Focus group participants also expressed satisfaction with the intervention, and some even said they would recommend the treatment to others with ED. This is consistent with the quantitative satisfaction reported in our previous study, where patients showed high satisfaction with the treatment [30]. It is also in line with previous research, which generally shows positive opinions [26,27] and high satisfaction with blended interventions [16].

Regarding the domain related to elements that helped them maintain adherence, the most frequent category was therapeutic support, which once again highlights the importance of the therapist's presence. The existence of scheduled sessions and the commitment to groupmates were mentioned, which also appears to reinforce the importance of support and guidance from therapists as well as one of the advantages of the group format [20,21]. Other important points were the therapeutic materials themselves and the effects of the treatment on the patients, both in terms of their personal experience of improvement and the possibility to see their progress reflected in the platform's graphs. In addition, although less frequently mentioned, some participants described internal factors, such as readiness to change and self-commitment, pointing to the role of intrinsic motivation. Taken together, these findings suggest that successful adherence in blended formats likely results from a combination of external (therapist, structure, characteristics, and content of treatment) and internal resources (motivation and self-management), supporting recent models of digital intervention engagement [62]. Future adaptations might benefit from integrating dynamic feedback systems and tailored support strategies to further optimize adherence.

With regard to this, during focus groups, participants suggested incorporating some changes to the platform, such as a discussion forum or a greater degree of interactivity. They also proposed adapting the configuration to the specific modules to be worked on or making the sessions more frequent, or even some of them in person. In addition, they proposed a greater individualization of the intervention, suggesting creating brief individual spaces with the therapist or combining some individual sessions with the current intervention. They also made proposals to improve adherence to treatment, some of which were also related to more individualized follow-up or leaving less space between sessions. Thus, the importance for patients of the presence of the therapists throughout the intervention and the customization of treatment is once again evident.

Finally, participants who dropped out of treatment provided reasons for dropping out and suggested some strategies to improve treatment adherence. Although they did not complete the program, we considered it important to include their experiences to enrich understanding of the intervention's acceptability and identify potential barriers to engagement. Similar to what was found in previous studies [12], participants mentioned that the treatment was not working for them, that they had difficulty applying the strategies learned, or that they identified the need for other types of assistance, such as in-person or more specialized help. The relevance of addressing these aspects has been previously evidenced because noncompletion of tasks generates dissonance, may lead to

XSL•FO
RenderX

increased discouragement, and is related to treatment success [27]. Dropout participants proposed changes regarding the format and configuration of the sessions with the aim of improving adherence to treatment, suggesting that they could be individual, in-person, or held weekly. They also believed that having more time to discuss individual problems in sessions or having fewer homework assignments could help to increase adherence. It has been shown that patients who engage more with homework assignments in online CBT interventions achieve a greater reduction in anxiety and depression symptoms [63], so it would be important to assess patients' engagement in this regard to be able to implement preventive strategies. Several of the identified barriers and proposed suggestions for improvement coincide with aspects mentioned by participants in the focus groups, which reinforces the value of the findings obtained. Taken together, the importance of focusing on the personalization of treatment emerges.

Although this format may not be suitable for all individuals, the findings from the feasibility study [30], together with the qualitative results from this study, support the viability and acceptability of this blended intervention, which also appears to have clinical utility. Additionally, it may help to overcome some of the disadvantages of fully face-to-face and fully self-administered formats and may be a good resource to facilitate access to EBT for a broader population. This is particularly relevant in the context of public health systems such as the Spanish NHS, where there are long waiting lists and sessions occur infrequently. Therefore, a key recommendation for future research and clinical practice would be to explore which patient profiles and under what circumstances this type of blended intervention is most effective [43], as well as to identify cases where alternative formats may be more beneficial.

Limitations

This study has several limitations. One of them is that retrospective responses pose a methodological limitation due to possible recall bias. However, this is a common limitation in qualitative research and not specific to this study. Nonetheless, the focus groups were conducted shortly after the end of the intervention, which likely reduced the extent of this bias. An additional limitation is that 2 of the participants in the focus groups hardly spoke during the interview, resulting in very few ideas from them, which made it difficult to reach the "General" category. Another limitation is that very few participants who dropped out of treatment completed the dropout questionnaire, so formal qualitative analyses could not be conducted. Finally, it would be advisable for future studies to be conducted with people treated in other settings, such as the Spanish NHS, in order to increase the generalizability of the results.

Despite the aforementioned limitations, this study contributes to completing the established conclusions regarding the feasibility, acceptability, and preliminary efficacy of the intervention studied [30]. Furthermore, to our knowledge, it is the first qualitative study to explore the opinions of patients (completers and noncompleters) who have received a transdiagnostic intervention applied in group and blended formats, thus contributing to the development of this novel format.

Conclusions

This study reinforces the feasibility and acceptability of a blended transdiagnostic group treatment for ED. These qualitative findings provide valuable insight into patient experiences, perceived benefits, and areas for improvement while also offering guidance for future research and implementation. This study supports this innovative format as a potentially scalable, acceptable, and effective option within mental health services.

Acknowledgments

Thanks to Rosa Lorente-Català, Alberto González-Robles, and Macarena Paredes-Mealla for their collaboration in conducting the focus groups.

Generative artificial intelligence (AI) tools were not used for data analysis, interpretation, or the development of the study findings. However, AI-based tools were used solely for language editing and stylistic improvements of the manuscript text. The authors take full responsibility for the accuracy, originality, and integrity of the work.

Funding

This study is funded by the Research Promotion Plan of the Universitat Jaume I (Ref. UJI-B2022-54; Code: 22I600) and supported by a predoctoral grant awarded to NJ-O from the Conselleria d'Educació, Cultura, Universitats i Ocupació of the Generalitat Valenciana and the European Social Fund (ACIF/2021/220). JO has received funding by the Government of Aragon (Department of Science, University and Knowledge Society; research team: S31_23R). The funders have no role in design, data collection and analysis or preparation of the manuscript.

Data Availability

The data analyzed are the transcripts of the interviews and are not publicly available as they may contain sensitive information. Access to the transcripts will be available upon reasonable request to the authors.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview script.
[DOCX File, 21 KB - humanfactors_v13i1e86016_app1.docx]

Checklist 1
COREQ checklist.
[DOCX File, 27 KB - humanfactors_v13i1e86016_app2.docx]

References

1. Bullis JR, Boettcher H, Sauer‐Zavala S, Farchione TJ, Barlow DH. What is an emotional disorder? A transdiagnostic mechanistic definition with implications for assessment, treatment, and prevention. Clin Psychol Sci Pract 2019 Jun;26(2):e12278 [FREE Full text] [doi: 10.1111/cpsp.12278]
2. World Health Organization. World Mental Health Today: Latest Data: World Health Organization; 2025:1-64 URL: https://iris.who.int/server/api/core/bitstreams/31714489-1345-4439-8b37-6cbdc52e15ca/content [accessed 2026-06-28]
3. Los trastornos de ansiedad tienen una prevalencia del 12,6% en la población. Infocop. 2024. URL: https://www.infocop.es/los-trastornos-de-ansiedad-tienen-una-prevalencia-del-126-en-la-poblacion-general/?cn-reloaded=1 [accessed 2025-06-21]
4. Ruiz-Rodríguez P, Cano-Vindel A, Muñoz Navarro R, et al. Impacto económico y carga de los trastornos mentales comunes en España: una revisión sistemática y crítica. Ansiedad y Estrés 2017 Jul;23(2-3):118-123. [doi: 10.1016/j.anyes.2017.10.003]
5. Fernández-García X. Situación de la psicología clínica en el Sistema Nacional de Salud (SNS) y perspectivas de crecimiento. Ansiedad y Estrés 2021;27(1):31-40. [doi: 10.5093/anyes2021a5]
6. Cuéllar Flores I, Fernández Garzón L, Félix-Alcántara MP, et al. Indicadores asistenciales y estándares de calidad asistencial para la psicología clínica en los centros de salud mental del Sistema Madrileño de Salud evaluados por sus profesionales. Apunt Psic 2022 Oct 11;40(2):71-86. [doi: 10.55414/ap.v40i2.1414]
7. Sauer-Zavala S, Gutner CA, Farchione TJ, Boettcher HT, Bullis JR, Barlow DH. Current definitions of "Transdiagnostic" in treatment development: a search for consensus. Behav Ther 2017 Jan;48(1):128-138. [doi: 10.1016/j.beth.2016.09.004] [Medline: 28077216]
8. Barlow DH, Farchione TJ, Sauer-Zavala S, et al. Unified Protocol for Transdiagnostic Treatment of Emotional Disorders: Therapist Guide: Oxford University Press; 2017. [doi: 10.1093/med-psych/9780190685973.001.0001]
9. Allen LB, White KS, Barlow DH, Shear MK, Gorman JM, Woods SW. Cognitive-behavior therapy (CBT) for panic disorder: relationship of anxiety and depression comorbidity with treatment outcome. J Psychopathol Behav Assess 2010 Jun;32(2):185-192. [doi: 10.1007/s10862-009-9151-3] [Medline: 20421906]
10. Hedman-Lagerlöf E, Carlbring P, Svärdman F, Riper H, Cuijpers P, Andersson G. Therapist-supported internet-based cognitive behaviour therapy yields similar effects as face-to-face therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. World Psychiatry 2023 Jun;22(2):305-314. [doi: 10.1002/wps.21088] [Medline: 37159350]
11. Kolaas K, Berman AH, Hedman-Lagerlöf E, Lindsäter E, Hybelius J, Axelsson E. Internet-delivered transdiagnostic psychological treatments for individuals with depression, anxiety or both: a systematic review with meta-analysis of randomised controlled trials. BMJ Open 2024 Apr 3;14(4):e075796. [doi: 10.1136/bmjopen-2023-075796] [Medline: 38569713]
12. Fernández-Álvarez J, Díaz-García A, González-Robles A, Baños R, García-Palacios A, Botella C. Dropping out of a transdiagnostic online intervention: a qualitative analysis of client's experiences. Internet Interv 2017 Dec;10:29-38. [doi: 10.1016/j.invent.2017.09.001] [Medline: 30135750]
13. Patel S, Akhtar A, Malins S, et al. The acceptability and usability of digital health interventions for adults with depression, anxiety, and somatoform disorders: qualitative systematic review and meta-synthesis. J Med Internet Res 2020 Jul 6;22(7):e16228. [doi: 10.2196/16228] [Medline: 32628116]
14. Erbe D, Eichert HC, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017 Sep 15;19(9):e306. [doi: 10.2196/jmir.6588] [Medline: 28916506]
15. Leterme AC, Behal H, Demarty AL, et al. A blended cognitive behavioral intervention for patients with adjustment disorder with anxiety: a randomized controlled trial. Internet Interv 2020 Sep;21:100329. [doi: 10.1016/j.invent.2020.100329] [Medline: 32523873]
16. Ferrao Nunes-Zlotkowski K, Shepherd HL, Beatty L, Butow P, Shaw JM. Blended psychological therapy for the treatment of psychological disorders in adult patients: systematic review and meta-analysis. Interact J Med Res 2024 Oct 29;13:e49660. [doi: 10.2196/49660] [Medline: 39470720]

XSL•FO
RenderX

17. Romijn G, Batelaan N, Koning J, et al. Acceptability, effectiveness and cost-effectiveness of blended cognitive-behavioural therapy (bCBT) versus face-to-face CBT (ftfCBT) for anxiety disorders in specialised mental health care: a 15-week randomised controlled trial with 1-year follow-up. PLoS One 2021;16(11):e0259493. [doi: 10.1371/journal.pone.0259493] [Medline: 34767575]
18. Burlingame GM, Strauss B, Joyce AS. Change mechanisms and effectiveness of small group treatment. In: Lambert MJ, editor. Bergin and Garfield's Handbook of Psychotherapy and Behavior Change, 6th edition: John Wiley & Sons; 2013:640-689 URL: https://www.researchgate.net/profile/Bernhard-Strauss/publication/303105298_Change_mechanisms_and_effectiveness_of_small_group_treatments/links/5774fe3308aeb9427e257e36/Change-mechanisms-and-effectiveness-of-small-group-treatments.pdf [accessed 2026-06-28]
19. Rosendahl J, Alldredge CT, Burlingame GM, Strauss B. Recent developments in group psychotherapy research. Am J Psychother 2021 Jun 1;74(2):52-59. [doi: 10.1176/appi.psychotherapy.20200031] [Medline: 33745284]
20. Psychotherapy: understanding group therapy. American Psychological Association. 2019. URL: https://www.apa.org/topics/psychotherapy/group-therapy [accessed 2025-06-21]
21. Pappas S. Group therapy is as effective as individual therapy, and more efficient. Here's how to do it successfully. Monit Psychol 2023;54(2) [FREE Full text]
22. Peris-Baquero Ó, Moreno JD, Osma J. Long-term cost-effectiveness of group unified protocol in the Spanish public mental health system. Curr Psychol 2023 Sep;42(26):22462-22477. [doi: 10.1007/s12144-022-03365-8]
23. Díaz-García A, González-Robles A, García-Palacios A, et al. Blended transdiagnostic group CBT for emotional disorders: a feasibility trial protocol. Internet Interv 2021 Mar;23:100363. [doi: 10.1016/j.invent.2021.100363] [Medline: 33520670]
24. Creswell JW, Plano Clark VL. Designing and Conducting Mixed Methods Research, 3rd edition: SAGE Publications; 2018. [doi: 10.1111/j.1753-6405.2007.00096.x]
25. Fernández-Felipe I, Díaz-García A, Marco JH, García-Palacios A, Guillén Botella V. "Family Connections", a DBT-based program for relatives of people with borderline personality disorder during the COVID-19 pandemic: a focus group study. Int J Environ Res Public Health 2022;19(1):79. [doi: 10.3390/ijerph19010079]
26. Etzelmueller A, Radkovsky A, Hannig W, Berking M, Ebert DD. Patient's experience with blended video- and internet based cognitive behavioural therapy service in routine care. Internet Interv 2018 Jun;12:165-175. [doi: 10.1016/j.invent.2018.01.003] [Medline: 30135780]
27. Schuster R, Sigl S, Berger T, Laireiter AR. Patients' experiences of web- and mobile-assisted group therapy for depression and implications of the group setting: qualitative follow-up study. JMIR Ment Health 2018 Jul 11;5(3):e49. [doi: 10.2196/mental.9613] [Medline: 29997106]
28. Schuster R, Topooco N, Keller A, Radvogin E, Laireiter AR. Advantages and disadvantages of online and blended therapy: replication and extension of findings on psychotherapists' appraisals. Internet Interv 2020 Sep;21:100326. [doi: 10.1016/j.invent.2020.100326] [Medline: 32477885]
29. Titzler I, Saruhanjan K, Berking M, Riper H, Ebert DD. Barriers and facilitators for the implementation of blended psychotherapy for depression: a qualitative pilot study of therapists' perspective. Internet Interv 2018 Jun;12:150-164. [doi: 10.1016/j.invent.2018.01.002] [Medline: 30135779]
30. Jiménez-Orenga N, Díaz-García A, Lorente-Català R, et al. Feasibility, acceptability, and preliminary efficacy of a blended transdiagnostic group CBT for the treatment of emotional disorders. Internet Interv 2025 Jun;40:100823. [doi: 10.1016/j.invent.2025.100823] [Medline: 40235849]
31. Krueger RA, Casey MA. Focus Groups: A Practical Guide for Applied Research, 5th edition: SAGE; 2015. URL: https://tinyurl.com/3p826j7p [accessed 2026-06-28]
32. Hill CE, Thompson BJ, Williams EN. A guide to conducting consensual qualitative research. Couns Psychol 1997 Oct;25(4):517-572. [doi: 10.1177/0011000097254001]
33. Hill CE, Knox S, Thompson BJ, Williams EN, Hess SA, Ladany N. Consensual qualitative research: an update. J Couns Psychol 2005 Apr;52(2):196-205. [doi: 10.1037/0022-0167.52.2.196]
34. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
35. Vernmark K, Hesser H, Topooco N, et al. Working alliance as a predictor of change in depression during blended cognitive behaviour therapy. Cogn Behav Ther 2019 Jul;48(4):285-299. [doi: 10.1080/16506073.2018.1533577] [Medline: 30372653]
36. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum-Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
37. Castilla D, Jaen I, Suso-Ribera C, et al. Psychometric properties of the Spanish full and short forms of the System Usability Scale (SUS): detecting the effect of negatively worded items. Int J Hum-Comput Interact 2024 Aug 2;40(15):4145-4151. [doi: 10.1080/10447318.2023.2209840]
38. Diaz-Sanahuja L, Suso-Ribera C, Lucas I, et al. A self-applied psychological treatment for gambling-related problems via the internet: a pilot, feasibility study. J Gambl Stud 2024 Sep;40(3):1623-1651. [doi: 10.1007/s10899-024-10318-2] [Medline: 38795233]

XSL•FO
RenderX

39. Rachyla I, Quero S, Pérez-Ara M, Molés M, Mira A. Web-based, self-help intervention for adjustment disorders: acceptance and usability. Annu Rev Cybertherapy Telemed 2017:207-210 [FREE Full text]
40. Borghouts J, Eikey E, Mark G, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387. [doi: 10.2196/24387] [Medline: 33759801]
41. Smith KA, Blease C, Faurholt-Jepsen M, et al. Digital mental health: challenges and next steps. BMJ Ment Health 2023 Feb;26(1):e300670. [doi: 10.1136/bmjment-2023-300670] [Medline: 37197797]
42. Bennett K, Bennett AJ, Griffiths KM. Security considerations for e-mental health interventions. J Med Internet Res 2010 Dec 19;12(5):e61. [doi: 10.2196/jmir.1468] [Medline: 21169173]
43. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen J. Mixing online and face-to-face therapy: how to benefit from blended care in mental health care. JMIR Ment Health 2016 Feb 9;3(1):e9. [doi: 10.2196/mental.4534] [Medline: 26860537]
44. Osma J, Peris-Baquero O, Suso-Ribera C, Farchione TJ, Barlow DH. Effectiveness of the Unified Protocol for transdiagnostic treatment of emotional disorders in group format in Spain: results from a randomized controlled trial with 6-months follow-up. Psychother Res 2022 Mar;32(3):329-342. [doi: 10.1080/10503307.2021.1939190] [Medline: 34132170]
45. Schuster R, Kalthoff I, Walther A, et al. Effects, adherence, and therapists' perceptions of web- and mobile-supported group therapy for depression: mixed-methods study. J Med Internet Res 2019 Apr 28;21(5):e11860. [doi: 10.2196/11860] [Medline: 31066700]
46. Spinhoven P, Drost J, de Rooij M, van Hemert AM, Penninx BW. A longitudinal study of experiential avoidance in emotional disorders. Behav Ther 2014 Nov;45(6):840-850. [doi: 10.1016/j.beth.2014.07.001] [Medline: 25311292]
47. Reid GM, Grills AE, Mian ND, Reid AA, Merson RA, Langer DA. Using research-informed pedagogical practices to maximize learning in youth cognitive behavioral therapy. Evid Based Pract Child Adolesc Ment Health 2017;2(2):82-95. [doi: 10.1080/23794925.2017.1290511] [Medline: 28983515]
48. Ricci L, Villegente J, Loyal D, Ayav C, Kivits J, Rat AC. Tailored patient therapeutic educational interventions: a patient-centred communication model. Health Expect 2022 Feb;25(1):276-289. [doi: 10.1111/hex.13377] [Medline: 34816546]
49. Banbury A, Nancarrow S, Dart J, Gray L, Parkinson L. Telehealth interventions delivering home-based support group videoconferencing: systematic review. J Med Internet Res 2018 Feb 2;20(2):e25. [doi: 10.2196/jmir.8090] [Medline: 29396387]
50. Yalom ID, Leszcz M. The Theory and Practice of Group Psychotherapy, 6th edition: Basic Books; 2020. URL: https://tinyurl.com/yk6yx2ku [accessed 2026-06-28]
51. Huynh AT, Gaboury I, Provencher MD, Norton PJ, Roberge P. Patient acceptability of group transdiagnostic cognitive behavior therapy for the treatment of anxiety disorders in community-based care: a qualitative study. Clinical Psychologist 2022 May 4;26(2):119-128. [doi: 10.1080/13284207.2022.2041363]
52. Norton PJ. Transdiagnostic group CBT for anxiety disorder: efficacy, acceptability, and beyond. Rev Psicopatología Psicol Clínica 2012;17(3):205-217 [FREE Full text] [doi: 10.5944/rppc.vol.17.num.3.2012.11840]
53. Leuchtenberg S, Gromer D, Käthner I. Videoconferencing versus face‐to‐face psychotherapy: insights from patients and psychotherapists about comparability of therapeutic alliance, empathy and treatment characteristics. Couns Psychother Res 2023 Jun;23(2):389-403 [FREE Full text] [doi: 10.1002/capr.12538]
54. Bleyel C, Hoffmann M, Wensing M, Hartmann M, Friederich HC, Haun MW. Patients' perspective on mental health specialist video consultations in primary care: qualitative preimplementation study of anticipated benefits and barriers. J Med Internet Res 2020 Apr 20;22(4):e17330. [doi: 10.2196/17330] [Medline: 32310139]
55. Steel K, Cox D, Garry H. Therapeutic videoconferencing interventions for the treatment of long-term conditions. J Telemed Telecare 2011;17(3):109-117. [doi: 10.1258/jtt.2010.100318] [Medline: 21339304]
56. Trondsen MV, Bolle SR, Stensland G, Tjora A. Video-confidence: a qualitative exploration of videoconferencing for psychiatric emergencies. BMC Health Serv Res 2014 Dec;14(1):544. [doi: 10.1186/s12913-014-0544-y]
57. Osma J, Suso-Ribera C, Peris-Baquero Ó, et al. What format of treatment do patients with emotional disorders prefer and why? Implications for public mental health settings and policies. PLoS One 2019;14(6):e0218117. [doi: 10.1371/journal.pone.0218117] [Medline: 31181100]
58. Berger T. The therapeutic alliance in internet interventions: a narrative review and suggestions for future research. Psychother Res 2017 Sep;27(5):511-524. [doi: 10.1080/10503307.2015.1119908] [Medline: 26732852]
59. Backhaus A, Agha Z, Maglione ML, et al. Videoconferencing psychotherapy: a systematic review. Psychol Serv 2012 May;9(2):111-131. [doi: 10.1037/a0027924] [Medline: 22662727]
60. Norwood C, Moghaddam NG, Malins S, Sabin‐Farrell R. Working alliance and outcome effectiveness in videoconferencing psychotherapy: a systematic review and noninferiority meta‐analysis. Clin Psychology and Psychoth 2018 Nov;25(6):797-808 [FREE Full text] [doi: 10.1002/cpp.2315]
61. Simpson SG, Reid CL. Therapeutic alliance in videoconferencing psychotherapy: a review. Aust J Rural Health 2014 Dec;22(6):280-299. [doi: 10.1111/ajr.12149] [Medline: 25495622]

XSL•FO
RenderX

62. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
63. Sapkota RP, Peynenburg V, Dear BF, Titov N, Hadjistavropoulos HD. Engagement with homework in an internet-delivered therapy predicts reduced anxiety and depression symptoms: a latent growth curve analysis. J Consult Clin Psychol 2023 Feb;91(2):112-117. [doi: 10.1037/ccp0000775] [Medline: 36395032]

Abbreviations

CBT: cognitive behavioral therapy
COREQ: Consolidated criteria for reporting qualitative research
CQR: consensual qualitative research
EBT: evidence-based treatment
ED: emotional disorder
NHS: National Health System

Edited by A Kushniruk; submitted 16.Oct.2025; peer-reviewed by M Pritchard; revised version received 15.Apr.2026; accepted 04.May.2026; published 06.Jul.2026.

Please cite as:
Jiménez-Orenga N, Fadrique-Jiménez A, García-Palacios A, Osma J, Díaz-García A, Bretón-López J
Exploring Patients' Experiences in a Blended Transdiagnostic Group Treatment: Qualitative Study
JMIR Hum Factors 2026;13:e86016
URL: https://humanfactors.jmir.org/2026/1/e86016
doi:10.2196/86016

© Noelia Jiménez-Orenga, Alba Fadrique-Jiménez, Azucena García-Palacios, Jorge Osma, Amanda Díaz-García, Juana Bretón-López. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 6.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

User Perspectives on Conversational Agents for Preventive Alcohol Self-Help: Qualitative Exploratory Study

Marissa De Vries[1], MSc; Melissa Begeman[1], MSc; Marloes Derksen[2], PhD; Anna Goudriaan[1,3], Prof Dr; Matthijs Blankers[1,4,5], Prof Dr

[1]Department of Research, Arkin Mental Health Care, Amsterdam Institute for Addiction Research, Klaprozenweg 111, Amsterdam, The Netherlands

[2]Department of Medical Informatics, eHealth Living & Learning Lab Amsterdam, University of Amsterdam, Amsterdam University Medical Centers, Amsterdam, The Netherlands

[3]Department of Psychiatry, Amsterdam University Medical Centers, Amsterdam, The Netherlands

[4]Department of Drugs, Institute of Mental Health and Addiction, Trimbos Institute, Utrecht, The Netherlands

[5]Tranzo Scientific Center for Care and Wellbeing, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, The Netherlands

Corresponding Author:
Marissa De Vries, MSc
Department of Research, Arkin Mental Health Care, Amsterdam Institute for Addiction Research, Klaprozenweg 111, Amsterdam, The Netherlands

Abstract

Background: Alcohol use remains a major public health concern, and although preventive alcohol self-help interventions aim to support individuals in need of indicated prevention, they continue to face challenges related to low engagement and high attrition. Chatbots, also known as conversational agents (CAs), powered by artificial intelligence, may enhance engagement by offering personalized guidance and 24/7 availability. Yet, user perspectives on CAs in preventive alcohol self-help care remain largely unexplored.

Objective: This paper aimed to gain insight into the diverse factors shaping the adoption of CAs in preventive alcohol self-help by qualitatively exploring users' expectations and concerns, complemented by descriptive questionnaire data on perceived usefulness, intention to use, comfort, and trust.

Methods: This qualitative, exploratory study with a small purposive sample (N=12) included adults recruited from a Dutch digital alcohol self-help intervention and a client advisory board. Data were collected during an online focus group (n=8, 66.7%), 4 semistructured interviews (n=4, 33.3%), and a brief questionnaire administered to all participants in March 2025 and April 2025. The qualitative data were analyzed inductively using thematic analysis.

Results: Qualitative findings showed that participants appreciated the 24/7 availability, practical guidance, and nonjudgmental stance of CAs but doubted their capacity for genuine empathy and authenticity. Participants emphasized that the conversational tone was important: CAs should be supportive without being overly agreeable and should be confrontational once trust is established. Concerns about undue influence arose occasionally, while privacy was viewed by 1 (8.3%) participant as a major issue. Descriptive questionnaire results showed that 11 (92%) participants had prior experience with CAs, whereas only 6 (50%) had experience with digital health interventions. Most participants considered CAs potentially useful for reducing alcohol use, yet few expressed strong intentions to use one themselves. Comfort and trust ratings were largely neutral, reflecting a cautious but generally open attitude toward CA use.

Conclusions: CAs appear to hold promise as complementary tools for preventive alcohol self-help, although they are not perceived as replacements for human contact. Future design efforts should emphasize personalized interactions; an appropriate conversational tone, supportive yet not overly agreeable; and transparent communication about data use to enhance engagement and acceptance. The findings are descriptive and are not intended to be broadly generalizable.

(JMIR Hum Factors 2026;13:e86752) doi:10.2196/86752

KEYWORDS

conversational agents; preventive health care; self-help; alcohol use; qualitative study

Introduction

Although global alcohol consumption has declined since 2010, alcohol use remains a significant public health concern [1]. Alcohol use is associated with increased risks of physical illness, psychological distress, and adverse social outcomes [2]. Despite these consequences, globally only around 1 in 6 individuals with alcohol use disorder receives any form of treatment [3].

XSL•FO
RenderX

This treatment gap is driven by a range of factors, including limited availability of services, inadequate screening and referral procedures, and stigma surrounding alcohol-related problems [4].

In response to the limited availability of services, digital interventions have emerged as promising tools to increase access to care. These interventions, delivered via websites or mobile apps, can be used independently or alongside professional treatment. They are typically low-threshold, flexible in timing and location, and have shown small to moderate effects in reducing alcohol use and improving mental health outcomes [5,6]. There is growing recognition that digital interventions should not only target individuals with alcohol use disorders but also those engaging in risky drinking behavior. Preventive digital programs can reach this broader group early, potentially reducing harm before dependence develops [7,8].

Preventive digital interventions can vary in intensity, and a key distinction concerns the degree of guidance. Guided interventions—those that include contact with a therapist or coach—tend to show higher adherence than fully automated self-help formats [5,9,10]. This suggests that some form of personalized feedback or interaction is important for maintaining engagement. Conversational agents (CAs), also known as chatbots, offer a potential solution to deliver such guidance in a scalable way. A CA is an artificial intelligence (AI) system designed to engage users in natural, humanlike dialogue. Preliminary evidence suggests that CAs can enhance user satisfaction and engagement compared to unguided digital interventions and may even support reductions in depressive symptoms [11-13].

Although evidence for the effectiveness of CAs in digital mental health care is increasing, it is also important to understand users' perspectives on these technologies. A recent qualitative study with users of a CA for anxiety reported that participants valued its accessibility but expressed concerns about the agent's ability to demonstrate empathy and maintain privacy. Many participants also emphasized the need for human supervision [11]. A review highlighted that factors such as personalization, reliable responses, and ease of use are important for user engagement with these technologies [14]. Additionally, although some users remain skeptical about the role of CAs in providing health advice, they appear somewhat more open to using them as a supplement to mental health support [15]. Reviews further emphasize that personalization, reliability, and transparent communication are central to sustained engagement [14].

Although existing research provides important insights into the general use of CAs for mental health, much less is known about their potential role in alcohol-related prevention and self-help. Alcohol consumption involves distinct behavioral, social, and motivational dimensions that may influence how users perceive and interact with automated support. Understanding these perceptions is important for assessing whether CAs can effectively complement existing preventive self-help approaches and for identifying factors that may facilitate or hinder their adoption. This study therefore examines the experiences, expectations, and concerns of participants in a Dutch preventive alcohol self-help program and of members of a client advisory board, focusing on their perspectives on CAs in digital alcohol prevention and the reasons why they would or would not use such tools. Given the limited existing evidence, the study adopted an exploratory qualitative design without predefined hypotheses or preregistration, with the goal of generating inductive insights to inform the future development and evaluation of CAs in digital alcohol prevention interventions.

Methods

Study Design

This study used an exploratory qualitative design based on an online focus group, 4 semistructured interviews, and a brief questionnaire to gather contextual information. The focus group primarily generated interactive discussions around shared experiences and general expectations, while the individual interviews provided more nuanced and personal reflections, particularly regarding doubts about empathy, trust, and personal willingness to use CAs. The questionnaire was administered prior to the focus group and interviews to obtain participants' independent views on CAs, uninfluenced by group dynamics, and to serve as a basis for more in-depth exploration during the discussions. The questionnaire data were used descriptively and for contextual triangulation of the qualitative findings, rather than for inferential analysis. Given the small sample size, the quantitative results (eg, mean scores) should be interpreted with caution and are intended to provide indicative insights only.

Recruitment

The qualitative and quantitative components of the study included 12 participants recruited between March 2025 and April 2025 from a Dutch online self-help program for alcohol use and the client advisory board of a large mental health care institution (Arkin, the Netherlands). The self-help program consists of 6 evidence-based modules based on both motivational interviewing and cognitive behavioral therapy. Other intervention features include a support forum, a personal diary, and encouragement for users to log their daily substance use and craving in accordance with cognitive behavioral therapy principles.

Inclusion criteria were participation in Jellinek's digital self-help alcohol intervention, familiarity with the program, proficiency in Dutch, and willingness to participate in an online group discussion or interview. Due to the anonymous nature of the self-help program, participants' contact details were not initially available. However, individuals who expressed interest in participating in the study by responding to a forum message or clicking a banner within the self-help program were asked to provide their phone number or email address, enabling the researchers to provide more information and formally invite them to the focus group. A total of 24 participants provided contact details, all of whom were contacted multiple times to schedule participation. Of these, 2 (16.7%) participants did not speak Dutch, 6 (50%) were unable or unwilling to join an online group discussion or interview, and 7 (58.3%) did not respond to phone calls or emails. Ultimately, 8 (66.7%) participants from the self-help community joined the focus group. In addition to the focus group, 4 semistructured interviews were conducted. Three interviews were held with members of Arkin's client

XSL•FO
RenderX

advisory board, and 1 was conducted with another self-help user who did not want to join a group discussion but was willing to participate in an interview. Because of the limited number of participants (n=12, 100%), formal data saturation could not be established; however, recurring patterns and overlapping themes emerged across the focus group and interviews. This suggests a degree of thematic consistency, even if full saturation was not attained [16].

All participants completed a questionnaire prior to the focus group and interviews and received a €45 (€1=US $1.14 as of June 23, 2026) gift voucher as compensation.

Data Collection Procedures

Data were collected between March 2025 and April 2025 through 1 online focus group session (n=8, 66.7%) and 4 online semistructured interviews (n=4, 33.3%). The focus group followed a semistructured discussion format exploring attitudes toward the integration of a CA. The focus group lasted approximately 3 hours, although it was planned for 2.5 hours. The same thematic areas were addressed in the individual interviews, allowing deeper reflection and more personal nuance. The interviews lasted approximately 1.5 hours. The focus group was moderated by an independent researcher, a female psychotherapist (MSc) with prior experience in moderating focus groups. She had no relationship with any participants. She was independent because she did not have assumptions or personal interests related to the study, which was important to minimize the risk of leading questions or undue influence. The primary researcher (first author) was present during the focus group in a nonparticipatory role. She introduced herself as the researcher of this project, explained why an independent person moderated the focus group, and explained the rationale for conducting this research. The researcher was also present to address potential procedural questions from participants and to become familiar with the data. This allowed the researcher to gain a contextual understanding of the discussions, including tone and group dynamics, which may not be fully captured in written transcripts. The researcher also made field notes during the focus group and interviews.

Prior to the focus group and interviews, participants were asked to complete a short, researcher-developed questionnaire about their prior experiences with CAs and their perspectives on how these could be used in the self-help program. All participants were asked for what purposes they would use a CA after being given several examples of possible functions. They were invited to write down as many desired functions as they wished. The questionnaire also included several demographic items (Table 1). The questionnaire was pretested by researchers involved in the study prior to data collection to ensure that it functioned properly and that the wording was comprehensible for individuals with varying levels of language proficiency. Several questionnaire items were also discussed during the focus group and interviews to gain deeper insight into participants' responses and to stimulate group-level discussion (Checklist 1).

Table . Characteristics of participants (N=12).

Characteristics	Participants
Gender, n (%)	
Men	8 (66.7)
Women	4 (33.3)
Age (years), mean (SD)	61.1 (8.3)
Education level, n (%)	
Lower	3 (25.0)
Higher	9 (75.0)
Country of birth, n (%)	
The Netherlands	11 (91.7)
Belgium	1 (8.3)
Previous addiction treatment, n (%)	
Yes	6 (50.0)
No	6 (50.0)
Previous experience with digital interventions[a], n (%)	
Yes	6 (50.0)
No	6 (50.0)

[a]Excludes the Jellinek digital alcohol self-help intervention.

Data Analysis

The qualitative data were analyzed using inductive thematic analysis [17,18]. This method was selected for its suitability for generating in-depth, exploratory insights into participants' experiences, perceptions, and attitudes regarding complex and potentially ambivalent topics such as AI in mental health care. In addition, the inductive approach enables researchers to maintain an open perspective toward the data. Alternative

XSL•FO
RenderX

qualitative approaches, including grounded theory, content analysis, and framework analysis, were considered in consultation with a qualitative data supervision group before thematic analysis was selected as the most appropriate method, given the exploratory aim of the study.

The focus group and interviews were audio-recorded and transcribed verbatim. Transcripts were anonymized prior to analysis and were not returned to participants. The focus group and interviews were conducted in Dutch, and the quotes were translated into English and checked by a native speaker for accuracy.

In the first phase, the primary researcher (first author) familiarized herself with the data by repeatedly reading the transcripts of the focus group and interviews. An initial phase of open, inductive coding was then conducted, in which segments of the data relevant to the research question were identified. Given the nature of focus group and interview data, nonrelevant conversational content was not coded. Codes were generated at a semantic level, staying close to participants' own words and meanings without imposing predefined categories.

After completion of the initial coding, codes were systematically compared and grouped based on similarities and patterns, resulting in the development of candidate themes. These themes were iteratively reviewed and refined in relation to both the coded extracts and the full dataset. The coding process and theme development were discussed with a second independent researcher (M Begeman). Discrepancies in interpretation were resolved through discussion until consensus was reached. A documentation process was maintained by documenting coding decisions, code definitions, and changes to the coding structure, supported by analytic memos and versioning within MAXQDA 2022 (VERBI Software GmbH). Participants did not provide feedback on the findings.

The questionnaire data were analyzed descriptively using means and SDs to provide contextual information about participants' familiarity with and attitudes toward CAs.

Ethical Considerations

All participants provided informed consent and received €45 as compensation for participating. The Medical Research Ethics Committees United determined that the study was not subject to the Dutch Medical Research Involving Human Subjects Act (Dutch: *wet medisch-wetenschappelijk onderzoek met mensen*) and granted approval (2024.1147). To minimize potential bias, the focus group was moderated by an independent facilitator who was not involved in the research project. The primary researcher has a background in clinical psychology and addiction research, with an interest in digital mental health. She was not involved in the development of the self-help program under study but is involved in planned future adaptations and improvements to the program. To minimize potential bias, the researcher adopted a nondirective role during the focus group, avoiding leading questions and allowing participants to freely express their perspectives.

Results

Participant Demographics

Participant characteristics are presented in Table 1. The sample consisted of 8 (66.6%) men and 4 (33.3%) women, with a mean age of 61.1 (SD 8.3) years. Previous research among users of the same self-help program has shown that the average age in this population is relatively high, at around 51 years [19]. Most participants had completed higher education, and half reported prior experience with digital interventions (excluding the Jellinek digital alcohol self-help intervention).

Qualitative Data

The thematic analysis of the focus group data identified 2 overarching themes related to the use of CAs: motivating factors that encouraged participants to consider CAs as supportive tools and hindering factors that limited their willingness to engage with such systems. Within each theme, several subthemes captured both functional and experiential dimensions of how participants envisioned interacting with CAs in the context of a self-help program to reduce or quit their alcohol use.

Many of the themes raised by self-help participants during the focus group were echoed in the interviews with members of the advisory board, indicating a high degree of consistency across the 2 participant groups.

Motivating Factors for CA Use

Immediate Support in Moments of Need

A central theme was the perceived value of CAs as a source of immediate support, particularly during moments of craving or relapse. Participants described situations in which traditional support options, such as friends, family, or health professionals, would be unavailable, noting that cravings often occur at night or during other moments of isolation. The promise of a tool that is available "anytime, anywhere" was repeatedly emphasized as a unique strength of CAs.

One participant explained the following:

> *When you wake up at 2:30 a.m. craving a drink, it would be great to have something right then.*

For several participants, this possibility filled a gap that existing services or peer support could not reliably address. Others stressed that even if the CA could not "solve" the craving, the act of engaging in dialogue might provide enough distraction or motivation to resist the urge.

Information Source

Participants also emphasized the value of CAs as a reliable and comprehensive source of information. Many described using tools such as ChatGPT (OpenAI) in their daily lives as a quick reference point, highlighting their utility for clarifying questions about health, lifestyle, or substance use.

One participant shared the following:

> *I use ChatGPT like Google now.*

Several saw this practical, informational function as more appealing than emotional support, which they preferred to seek

XSL•FO
RenderX

from human contacts. The CA's broad knowledge base was described as particularly useful for offering "unexpected angles" or "new ideas" that might not occur to the user.

One participant noted the following:

> *Sometimes it gives you ideas that make you think, wow. A little spooky, but interesting.*

Such input was considered helpful for self-reflection, even when participants did not act directly on the advice.

Positive Design and Interaction Characteristics

Beyond specific functions, participants valued certain interactional characteristics of CAs. Among these were reliability (ie, being always available and never tired) and a nonjudgmental tone. Several participants contrasted this with human helpers, who might express disappointment, irritation, or fatigue in response to relapse disclosures. By contrast, the CA was viewed as a safer and more consistent space for sharing sensitive information.

One participant stated the following:

> *It doesn't get angry with you or sigh when you've relapsed again.*

In terms of design, participants expressed divergent preferences.

Some preferred a minimalist, text-only interface, emphasizing simplicity and focus:

> *Give me just text. No human-like figure.*

Others felt that adding visual or customizable elements could make interactions more engaging, even playful:

> *I'd find it funny if you could create the character yourself. With a weird hat and glasses, so it feels like someone you know.*

This design flexibility was highlighted as a motivating factor because it could adapt to different user needs and preferences.

Hindering Factors for CA Use

Reluctance and Unfamiliarity

For several participants, reluctance to engage with CAs stemmed less from negative experiences and more from a lack of familiarity with the technology. The novelty of AI-based systems created hesitation, described by 1 participant as "cold feet." This sense of uncertainty was tied to limited exposure and the perception that AI was still a "black box" or unknown territory. Others indicated that although they were curious, they preferred to observe developments from a distance before adopting such tools themselves. This hesitation therefore reflected not outright rejection but rather a cautious, wait-and-see attitude.

Privacy and Security Concerns

Privacy did not emerge as a dominant barrier in this study. Most participants distinguished between highly sensitive information, such as financial or medical records, which they would never entrust to a CA, and less sensitive disclosures about drinking behavior, which they perceived as relatively harmless.

One participant explained the following:

> *Who cares whether I drink or not? If that information leaks, that doesn't bother me.*

However, this general downplaying of privacy did not mean that all concerns were absent.

One participant raised questions about where their data might end up, mentioning companies or countries with geopolitical associations:

> *Which cloud does it end up in? To Elon Musk? The Chinese? The Russians?*

These statements suggest that although privacy was not top of mind for everyone, uncertainties about data storage and control could still undermine trust if left unaddressed.

Lack of Human Empathy and Authenticity

One of the most frequently cited barriers was the perceived inability of CAs to replicate essential human qualities, particularly empathy. Several participants emphasized that emotional support during moments of vulnerability required a genuine human connection.

One participant stated the following:

> *When it comes to moral support, my experience is that it only works with people.*

The absence of warmth or emotional resonance limited the value participants attributed to CAs in deeply personal situations, such as feelings of shame. Moreover, some participants found the communication style of existing chatbots unconvincing. Responses were described as "too smooth" or overly scripted, lacking the imperfections and nuances of human dialogue.

Rather than offering comfort, this polished style sometimes reduced credibility:

> *I don't want a super smooth answer. I want something I can actually use.*

These reflections highlight a tension around tone: responses that are overly smooth or pleasing may come across as inauthentic, making the CA feel less trustworthy or connected to the user's real struggles.

Off-Putting Design and Potential Influence

Design elements were another source of hesitation. Although some participants appreciated the possibility of customizing avatars or visual interfaces, others described these features as deceptive.

The idea of interacting with a humanlike avatar provoked discomfort, with 1 participant remarking the following:

> *I'd feel betrayed if a fake person was sitting across from me.*

Concerns about undue influence underscored the importance of transparency regarding the capabilities and limitations of CAs. Some worried that users with limited digital literacy or critical thinking skills might overrely on the system or accept its advice uncritically. The risk of undue influence underscored the importance of transparency about the limits of what CAs can and cannot do.

XSL•FO
RenderX

Finally, trust was repeatedly described as something that could not be assumed but had to be earned. Participants suggested that credibility would need to be built gradually, through consistent and reliable interactions over time. Without this foundation, early negative experiences could quickly erode the willingness to use the CA.

Survey: Familiarity With and Opinions on CAs

Participants' knowledge of and experiences with AI and CAs are presented in Table 2. The desired functions of a CA are presented in Figure 1.

The most frequently desired functions were helping users find information, providing advice, and offering motivation when struggling or after a relapse.

On a 5-point scale, participants reported moderate knowledge of AI (mean 3.0, SD 1.2) and perceived the usefulness of CAs for reducing alcohol use as fairly high (mean 3.8, SD 1.1). Comfort and trust levels were largely neutral (mean 3.0, SD 1.2 and mean 2.8, SD 0.9, respectively), while willingness to use a CA was mixed (mean 3.3, SD 1.3).

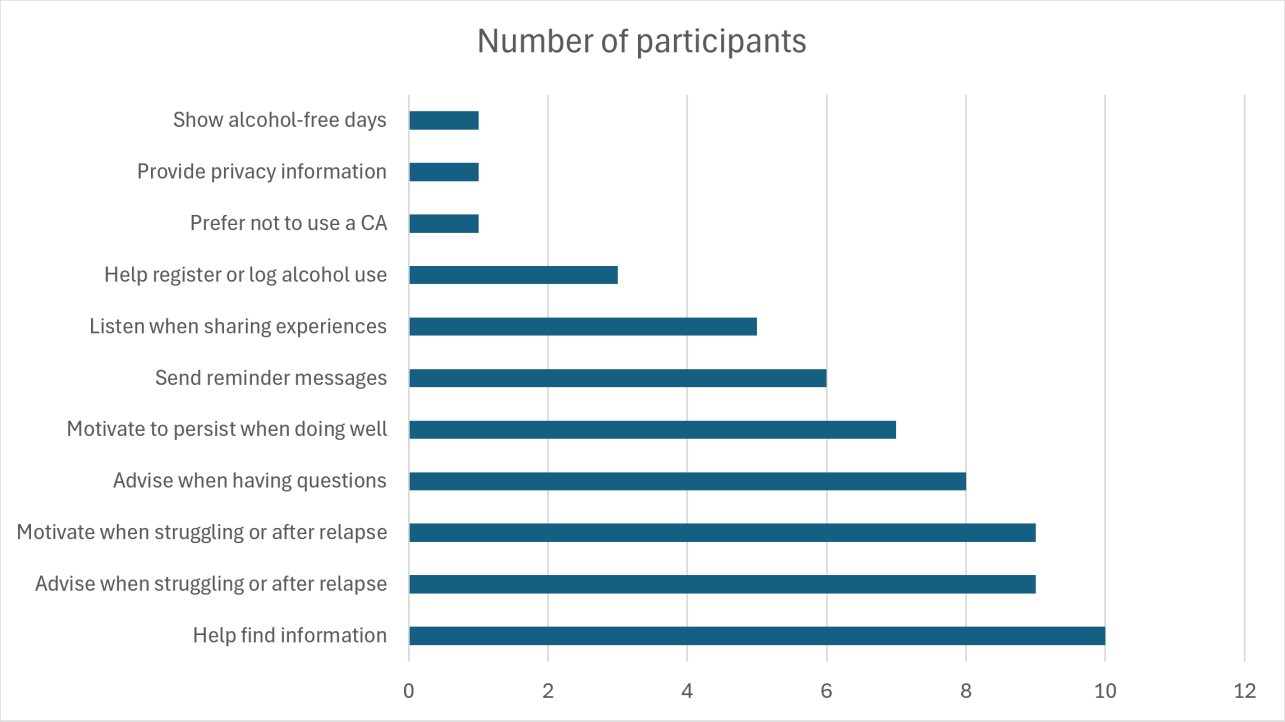
Table . Participants' knowledge of and experiences with conversational agents (CAs; N=12).

Items	Response scale (5-point)	Participants, mean (SD)
Knowledge of AI[a]	"Very low" to "very high"	3.0 (1.2)
Usefulness of AI for reducing alcohol use	"Not useful" to "very useful"	3.8 (1.1)
Willingness to use a CA	"Definitely not" to "definitely yes"	3.3 (1.3)
Comfort discussing alcohol use with a CA	"Very uncomfortable" to "very comfortable"	3.0 (1.2)
Trust in a CA ability to give good advice	"Very low" to "very high"	2.8 (0.9)
Satisfaction with chatbots[b]	"Very dissatisfied" to "very satisfied"	3.6 (0.8)

[a]AI: artificial intelligence.

[b]Only participants with prior chatbot experience answered this question (n=11).

Figure 1. Desired functions of a conversational agent (CA) for alcohol self-help among participants (N=12).



Discussion

Principal Findings

This qualitative study examined how users of a Dutch preventive alcohol self-help program and members of a client advisory board perceive the potential role of CAs in digital alcohol prevention. Across both qualitative and survey data, findings indicate a general perception of CAs as potentially useful, albeit accompanied by notable reservations regarding trust, authenticity, and willingness to adopt them.

Participants broadly valued the accessibility and functionality of CAs, particularly their around-the-clock support and practical guidance during moments of craving or relapse. They emphasized the appeal of an always-available, nonjudgmental companion that could offer both information and motivation without the social barriers often associated with human help.

XSL•FO
RenderX

These qualitative perceptions were supported by survey results showing relatively positive evaluations of usefulness for reducing alcohol use (mean 3.8, SD 1.1), suggesting that participants recognize the potential functional value of CAs in preventive contexts.

Despite the perceived usefulness, both datasets consistently indicated ambivalence regarding personal uptake. Although survey responses indicated only moderate willingness to use a CA themselves (mean 3.3, SD 1.3), participants in the focus group and interviews expressed ambivalence about empathy and authenticity. This tension suggests that perceived usefulness does not directly translate into intended use.

Concerns about empathy and authenticity were central in both methods. In interviews and focus groups, participants questioned whether CAs could provide meaningful emotional understanding or whether responses might feel too "scripted" or "pleasing," particularly in vulnerable situations. These concerns were also reflected in the survey data, with neutral trust ratings (mean 2.8, SD 0.9) and moderate comfort levels when discussing alcohol-related issues with a CA (mean 3.0, SD 1.2), indicating cautious attitudes toward relational aspects of CAs despite recognition of their functional benefits. Privacy concerns, in contrast, were mentioned less frequently and were generally not perceived as a major barrier.

Several factors that may facilitate engagement include trust, consistency, and flexibility in design. Participants viewed trust as something that must be earned gradually through reliable, transparent, and contextually appropriate interactions. The combination of practical functionality and emotional credibility appeared central to the willingness to use a CA in preventive care. Taken together, these findings provide a more nuanced picture of user perspectives on CAs in alcohol self-help: interest exists, but readiness to adopt CAs is conditional.

The study adds to existing knowledge by situating these insights specifically in a preventive context, where users are not seeking treatment but exploring support to maintain control over their drinking. This distinction helps explain the observed ambivalence, as CAs are perceived as promising complements rather than substitutes for human contact or established self-help methods.

Comparison With Prior Work

The finding that accessibility, neutrality, and broad knowledge are seen as major advantages aligns with prior digital mental health research [20,21]. Our results extend this work by showing that these qualities may be particularly relevant in preventive contexts, where individuals often hesitate to seek formal treatment because of stigma or low perceived need. In such cases, the neutrality of a CA may not be a limitation but rather a functional strength, lowering the threshold for disclosure and support.

Personalization emerged as a strong desire in our sample, ranging from tailored interaction styles to flexible interface design. This resonates with broader evidence that adaptive systems can foster engagement and adherence [22]. At the same time, personalization brings to the forefront a central tension in digital health: the more interventions are tailored to individuals, the more data collection is required. Prior research has shown that increased data demands can create privacy concerns and reduce trust [15]. Interestingly, most of our participants downplayed privacy as an obstacle in hypothetical discussions. This aligns with previous research on the self-help intervention, which showed limited concern for ethical aspects of applying machine learning to adapt the digital intervention [23].

This discrepancy points to a possible gap between imagined use and real-world adoption. Our survey data support this interpretation. Although most respondents described AI as useful for reducing alcohol use (5="very useful" and 3="quite useful"), only a minority expressed strong intentions to use a CA themselves (3="definitely yes"), with trust and comfort ratings largely neutral. These findings suggest that enthusiasm for CA functionalities in principle does not automatically translate into personal uptake.

Our findings also add nuance to debates about the emotional capacity of CAs. Previous studies have questioned whether they can provide empathy and called for human supervision [11]. Some participants in our study echoed these doubts, but others highlighted neutrality as preferable to human interaction, which can evoke feelings of disappointment or judgment. This suggests that the role of CAs may not be to replicate human warmth but to provide a complementary form of support in moments where neutrality is an advantage.

Finally, our results speak to the broader challenge of adherence in digital self-help programs. Dropout and inconsistent use remain persistent problems in unguided interventions [10]. The motivating factors identified by our participants—24×7 availability, nonjudgmental stance, and personalized feedback—mirror elements that prior research has associated with higher adherence in digital interventions, particularly studies showing that guidance and personalization increase completion rates [5,9,10,22].

Broader Implications and Ethical Considerations

The findings highlight both practical and ethical implications for the design and implementation of CAs in preventive alcohol self-help. From a design perspective, participants expressed diverse preferences regarding visual appearance and interaction style. Some favored a simple, text-based interface, while others appreciated the idea of a more humanlike avatar or personalized appearance. However, across these differences, participants consistently emphasized that the *tone of* communication mattered most. The CA should respond in an empathetic, respectful, and nonjudgmental way, without becoming overly pleasing or superficial. Participants valued honesty and even confrontation when appropriate, but only once trust had been established. This suggests that CAs in this context should be capable of adjusting their interactional style over time, matching their tone and feedback to the user's stage of engagement and level of trust.

Beyond design, and building on participants' emphasis on 24×7 availability and low-threshold access, CAs may serve as complements to existing self-help or prevention programs by offering immediate support during moments when human help is unavailable. Integrating such systems into alcohol prevention

XSL•FO
RenderX

and broader eHealth strategies could potentially increase accessibility, although this remains to be tested in practice. At the same time, the very features that make CAs attractive, such as their constant availability and responsiveness, may also introduce new ethical challenges. Participants' concerns about undue influence and the lack of genuine empathy suggest that CAs are not merely neutral tools but may function as moral technologies that shape decision-making and redistribute responsibility within care relationships. For example, when users rely on a CA for advice during moments of vulnerability, questions arise regarding who holds responsibility for the outcomes of such interactions—whether this lies with the user, the developers, or the system itself.

Limitations

Several limitations should be acknowledged. The sample was small, relatively homogeneous, and skewed toward older, higher-educated participants, which limits generalizability. However, previous research among users of the same self-help program has shown that the average age in this population is approximately 51 years [19], suggesting that the sample is broadly consistent with the program's user base. Another limitation concerns recruitment. Because participants were informed that the study involved discussing AI-based tools, it is possible that the sample included individuals who were more interested in or open to technological innovations in health care than the average self-help user. This may have resulted in a group with relatively well-developed or positive opinions about CAs, potentially limiting the range of perspectives represented.

Finally, although recurring patterns and overlapping themes were observed across the focus group and interviews, formal data saturation could not be confirmed because of the limited number of participants. Future studies with a larger and more diverse sample could build on these findings to further explore differences in user expectations, trust development, and CA design preferences.

Future Directions

Future research should replicate this study in larger and more diverse samples to verify whether these findings hold across different populations and drinking profiles and to achieve stronger data saturation. Observational and experimental studies involving actual interaction with CAs would allow researchers to examine how trust, comfort, and engagement evolve over time. Directly observing user-CA exchanges, for example, within existing preventive self-help platforms, could provide valuable behavioral data and reveal when and how users disengage.

It would also be valuable to explore how specific design features, such as text-only vs avatar-based interfaces and conversational tone, affect engagement. Finally, longitudinal research could clarify whether CAs can meaningfully support adherence and sustained behavior change over time, beyond initial curiosity or short-term use.

Conclusions

This study shows that users see potential for CAs in preventive alcohol self-help, mainly because of their constant availability, practical guidance, and neutral stance. However, CAs are not perceived as replacements for human contact. Moreover, enthusiasm in principle does not automatically translate into willingness to use such tools, as doubts about empathy, authenticity, and trust remain. CAs should therefore be viewed as complementary support within broader self-help systems. Future design efforts should emphasize personalized design; an appropriate conversational tone, supportive yet not overly agreeable; and transparent communication about data use to enhance engagement and acceptance. Their future value will depend on design, user involvement, and real-world evaluation to strengthen trust and engagement.

Acknowledgments

The authors thank all participants for their valuable contributions to the focus group and interviews. The authors are especially grateful to the chair of the focus group for her guidance and support during the discussions. The authors declare the use of generative artificial intelligence (GenAI) tools during the research and manuscript preparation process. According to the GenAI Disclosure Taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: proofreading, editing, and translation. The GenAI tools used were ChatGPT (version 4.5; OpenAI) and DeepL (DeepL SE). The authors take full responsibility for the final content of the manuscript. GenAI tools are not listed as authors and bear no responsibility for the final outcomes. DeepL was used to translate text into scientific English. As English is not the authors' native language, portions of the manuscript were initially drafted in Dutch and subsequently translated using DeepL. The authors reviewed all translated content for accuracy and fidelity to the original meaning.

In addition, selected portions of the Introduction and Discussion sections were proofread and edited using suggestions generated by ChatGPT-4.5 to improve readability.

Funding

This study was funded by ZonMw (The Netherlands Organisation for Health Research and Development; grant 10690032320014).

XSL•FO
RenderX

Data Availability

The datasets generated and analyzed during this study are not publicly available because of privacy and ethical restrictions. Data may be available from the corresponding author upon reasonable request and with permission from the Medical Research Ethics Committees United.

Authors' Contributions

Conceptualization: MDV, M Begeman, AG, M Blankers
Data curation: MDV
Formal analysis: MDV, M Begeman
Investigation: MDV
Methodology: MDV, MD, AG, M Blankers
Project administration: MDV
Resources: MDV
Software: MDV, M Begeman
Supervision: MD, AG, M Blankers
Visualization: MDV
Writing—original paper: MDV
Writing—review and editing: MDV, M Begeman, MD, M Blankers

Conflicts of Interest

None declared.

Checklist 1
COREQ checklist.
[PDF File, 433 KB - humanfactors_v13i1e86752_app1.pdf]

References

1. Global status report on alcohol and health and treatment of substance use disorders. : World Health Organization; 2024 URL: https://www.who.int/publications/i/item/9789240096745 [accessed 2026-06-22]
2. Rehm J, Gmel GE Sr, Gmel G, et al. The relationship between different dimensions of alcohol use and the burden of disease-an update. Addiction 2017 Jun;112(6):968-1001. [doi: 10.1111/add.13757] [Medline: 28220587]
3. Mekonen T, Chan GC, Connor J, Hall W, Hides L, Leung J. Treatment rates for alcohol use disorders: a systematic review and meta-analysis. Addiction 2021 Oct;116(10):2617-2634. [doi: 10.1111/add.15357] [Medline: 33245581]
4. Koob GF. Alcohol use disorder treatment: problems and solutions. Annu Rev Pharmacol Toxicol 2024 Jan 23;64:255-275. [doi: 10.1146/annurev-pharmtox-031323-115847] [Medline: 38261428]
5. Riper H, Hoogendoorn A, Cuijpers P, et al. Effectiveness and treatment moderators of internet interventions for adult problem drinking: an individual patient data meta-analysis of 19 randomised controlled trials. PLoS Med 2018 Dec;15(12):e1002714. [doi: 10.1371/journal.pmed.1002714] [Medline: 30562347]
6. Philippe TJ, Sikder N, Jackson A, et al. Digital health interventions for delivery of mental health care: systematic and comprehensive meta-review. JMIR Ment Health 2022 May 12;9(5):e35159. [doi: 10.2196/35159] [Medline: 35551058]
7. Kaner EF, Beyer FR, Garnett C, et al. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. Cochrane Database Syst Rev 2017 Sep 25;9(9):CD011479. [doi: 10.1002/14651858.CD011479.pub2] [Medline: 28944453]
8. Sohi I, Shield KD, Rehm J, Monteiro M. Digital interventions for reducing alcohol use in general populations: an updated systematic review and meta‐analysis. Alcohol Clin Exp Res (Hoboken) 2023 Oct;47(10):1813-1832. [doi: 10.1111/acer.15175]
9. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions — a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
10. Shams F, Tai AM, Kim J, et al. Adherence to e-health interventions for substance use and the factors influencing it: systematic review, meta-analysis, and meta-regression. Digit Health 2023;9:20552076231203876. [doi: 10.1177/20552076231203876] [Medline: 37780062]
11. Lee S, Yoon J, Cho Y, Chun J. A systematic review of chatbot-assisted interventions for substance use. Front Psychiatry 2024;15:1456689. [doi: 10.3389/fpsyt.2024.1456689] [Medline: 39319358]
12. Chen TH, Chu G, Pan RH, Ma WF. Effectiveness of mental health chatbots in depression and anxiety for adolescents and young adults: a meta-analysis of randomized controlled trials. Expert Rev Med Devices 2025 Mar;22(3):233-241. [doi: 10.1080/17434440.2025.2466742] [Medline: 39935147]

XSL•FO
RenderX

13. Heinz MV, Mackin DM, Trudeau BM, et al. Randomized trial of a generative AI chatbot for mental health treatment. NEJM AI 2025 Mar 27;2(4). [doi: 10.1056/AIoa2400802]
14. Khosravi M, Azar G. Factors influencing patient engagement in mental health chatbots: a thematic analysis of findings from a systematic review of reviews. Digit Health 2024;10:20552076241247983. [doi: 10.1177/20552076241247983] [Medline: 38655378]
15. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019;5:2055207619871808. [doi: 10.1177/2055207619871808] [Medline: 31467682]
16. Guest G, Bunce A, Johnson L. How many interviews are enough?: An experiment with data saturation and variability. Field Methods 2006;18(1):59-82. [doi: 10.1177/1525822X05279903]
17. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019;11(4):589-597. [doi: 10.1080/2159676X.2019.1628806]
18. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria: striving to meet the trustworthiness criteria. Int J Qual Methods 2017;16(1). [doi: 10.1177/1609406917733847]
19. Derksen M, van Beek M, Blankers M, et al. Effectiveness of machine learning-based adjustments to an eHealth intervention targeting mild alcohol use. Eur Addict Res 2025;31(1):47-59. [doi: 10.1159/000543252] [Medline: 39709958]
20. Li H, Zhang R, Lee YC, Kraut RE, Mohr DC. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. NPJ Digit Med 2023 Dec 19;6(1):236. [doi: 10.1038/s41746-023-00979-5] [Medline: 38114588]
21. Skjuve M, Følstad A, Fostervold KI, Brandtzaeg PB. My chatbot companion - a study of human-chatbot relationships. Int J Hum Comput Stud 2021 May;149:102601. [doi: 10.1016/j.ijhcs.2021.102601]
22. Hornstein S, Zantvoort K, Lueken U, Funk B, Hilbert K. Personalization strategies in digital mental health interventions: a systematic review and conceptual framework for depressive symptoms. Front Digit Health 2023;5:1170002. [doi: 10.3389/fdgth.2023.1170002] [Medline: 37283721]
23. Derksen ME, van Beek M, de Bruijn T, Stuit F, Blankers M, Goudriaan AE. Ethical aspects and user preferences in applying machine learning to adjust eHealth addressing substance use: a mixed-methods study. Int J Med Inform 2025 Jul;199:105897. [doi: 10.1016/j.ijmedinf.2025.105897] [Medline: 40157245]

Abbreviations

AI: artificial intelligence
CA: conversational agent

Edited by A Kushniruk, S Law; submitted 31.Oct.2025; peer-reviewed by P Martinez-Angulo, S Mohanadas; revised version received 01.Jun.2026; accepted 01.Jun.2026; published 14.Jul.2026.

Please cite as:
De Vries M, Begeman M, Derksen M, Goudriaan A, Blankers M
User Perspectives on Conversational Agents for Preventive Alcohol Self-Help: Qualitative Exploratory Study
JMIR Hum Factors 2026;13:e86752
URL: https://humanfactors.jmir.org/2026/1/e86752
doi:10.2196/86752

© Marissa De Vries, Melissa Begeman, Marloes Derksen, Anna Goudriaan, Matthijs Blankers. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 14.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Video-Algorithmic Patient Monitoring in Mental Health Inpatient Settings: Qualitative Study of Patient or Consumer, Clinician, and Vendor Perspectives

Piers Gooding[1], BA, PhD; Hamilton Kennedy[2], BSW, MA; Simon D'Alfonso[3], BA, BSc, PhD; Timothy Kariotis[4], BSocSci, MPH; Catherine Daniel[5], BPsychNurs, PGDipN(MtlHlth), MN, PhD; Bridget Hamilton[2], BSN, PhD

[1]La Trobe Law School, La Trobe University, Bundoora, Melbourne, Australia

[2]Centre for Mental Health Nursing, The University of Melbourne, Melbourne, VIC, Australia

[3]School of Computing & Information Systems, The University of Melbourne, Melbourne, Australia

[4]The University of Melbourne, Melbourne, Australia

[5]School of Nursing, Midwifery and Paramedicine, Australian Catholic University, Melbourne, Victoria, Australia

Corresponding Author:
Piers Gooding, BA, PhD
La Trobe Law School, La Trobe University, Bundoora, Melbourne, Australia

Abstract

Background: Video-algorithmic patient monitoring (VAPM) combines remote, noncontact sensors and algorithmic analysis and is increasingly trialed in acute psychiatric and other care settings. While promoted for improving safety and reducing risk, it raises ethical concerns regarding safety, privacy and surveillance. Little is known about how those encountering VAPM in mental health care contexts anticipate its use and potential impacts, including where it has not yet been implemented.

Objective: This study aimed to explore the views of patients or mental health consumers, specialized mental health nurses and nurse academics, hospital managers, and technology vendors regarding the appropriateness and anticipated implications of VAPM in mental health inpatient care.

Methods: This qualitative study identified key stakeholders in Australia via networking techniques for participation in a deliberative workshop. A deliberative workshop was held, and the workshop discussion was audio-recorded, transcribed, and thematically analyzed, consistent with methods in health technology research, which enable exploration of different viewpoints, including convergences and divergences across stakeholder groups.

Results: In total, 16 stakeholders participated, exploring themes concerning (1) contestation over the rationale for VAPM in mental health settings, (2) VAPM reshaping care and relationships, (3) perceived harms of VAPM, (4) perceived observational support for safety and reduced disruption, (5) serious privacy implications of VAPM, (6) the need for appropriate governance, and (7) the potential for VAPM to transform, not augment, service delivery. General views differed across groups. Patients or service users expressed concerns about privacy, coercion, and the potential to intensify stigma. Mental health nurses were cautious but interested in possible benefits for safety and suicide prevention. Hospital managers and technology vendors largely emphasized safety gains.

Conclusions: The findings suggest that the anticipated risks of VAPM are primarily experienced subjectively, as infringements on privacy, dignity, and trust, while purported benefits remain largely untested and unquantified. From a utilitarian perspective, direct comparison is therefore difficult—the risks are set out in the anticipated experiences of those with lived experience, and the benefits remain hypothetical. From this view, robust, independent evidence of real-world outcomes is required. Yet, for some participants, the very premise of such calculation was rejected, with privacy, dignity, and trust regarded as nonnegotiable, rather than items for trade-off. If VAPM is to be pursued at all, it should proceed only with extreme caution, with transparent evidence of outcomes, and with meaningful participation from those whose lives and care are most directly impacted.

(JMIR Hum Factors 2026;13:e85518) doi:10.2196/85518

KEYWORDS

vision-based patient monitoring and management; passive remote monitoring technology; video-algorithmic patient monitoring; monitoring and surveillance in mental health settings; health and artificial intelligence

XSL•FO
RenderX

Introduction

Technology used to monitor and surveil is proliferating in many areas of contemporary life. However, there are not always opportunities for people to consider the implications, personally and socially, for monitoring and surveillance in specific settings.

This paper refers to "video-algorithmic patient monitoring" (VAPM) to describe a form of monitoring that incorporates remote sensors, video technology, and algorithmic analysis. This technology is being trialed in varied health and social care settings, including mental health inpatient wards. VAPM involves video cameras or other sensors, artificial intelligence (AI) software, and an interface for nurses and other mental health practitioners to make remote, noncontact clinical observations and data collection about patients in hospital environments [1]. Other descriptors include "vision-based patient monitoring and management" and "passive remote monitoring technology" [2,3].

We conceptualize this monitoring arrangement as a "sociotechnical system" insofar as it is a "system involving the interaction of hard systems and human beings, in ways that either cannot be separated or are thought to be inappropriate to separate" [4,5]. This framing offers a way to analyze the dynamic interaction of people and technology, rather than narrowly focusing on the technology itself.

Traditional patient monitoring in hospitals involves in-person checks by health care professionals or simple alerts provided by clinical technology that tracks certain vital signs (eg, a pulse oximeter). VAPM is an iteration of these monitoring techniques. Proponents advocate for their use to improve upon in-person monitoring, including traditional forms of vital sign monitoring, and to reduce disruption for patients. Others have raised concerns that these technologies pose a threat to service user safety, privacy and dignity [6,7].

In mental health inpatient settings, VAPM involves patients' bedrooms being fitted with sensors [1], typically designed to focus on a person's pulse, breathing rate, and movement. Signals are detected remotely and continuously, based on physiological phenomena such as the chest rising and falling or the use of photoplethysmography (a simple optical technique that uses light to measure blood volume changes in peripheral circulation). Unlike vital sign monitoring, which is used extensively in health care settings, VAPM entails remote monitoring that does not rely on sensors attached to a person's body (on-body sensors). Attempts are underway to use such data as alerts to reduce self-harm, assault, and suicide in mental health wards [8].

At the time of data collection (December 2021), there was limited research on the impacts and use of VAPM [1,2]. Although several studies emerged since this time [9-13], empirical studies into VAPM remain limited. A systematic review of surveillance-based technology initiatives in inpatient and acute mental health settings [14] identified 9 studies that covered VAPM (the systematic review authors used the term "Vision-Based Patient Monitoring and Management" [14]), all from the United Kingdom; all were rated as "low quality," except one, which was rated "high quality" [3]. In total, 8 of the 9 studies also reported conflicts of interest, with the high-quality study being the only paper by Dewa et al [3], which did not report any conflicts of interest [14]. As we will discuss, the post-2021 studies—which are UK-centered—are broadly consistent with participants' anticipation that VAPM might reduce disturbance from in-person observations and improve efficiency (particularly overnight), with potential flow-on safety and cost implications [8-11,14]. However, participants also anticipated that these benefits would remain contestable and difficult to justify without clear evidence and strong governance, and that concern is reinforced by the still-limited and often conflicted UK evidence base [14].

The stated aims of VAPM include reducing disturbance to patients by making observations less intrusive, which may have several flow-on benefits to people's health and recovery [1], improving efficiency with night-time safety and reducing reliance on intensive monitoring, which may lead to reduced costs [10], improved quality and safety in care [10,11], reduction in incidents of assaults, and reductions in the associated involuntary tranquilization of patients and service users [13].

Ethical and legal concerns have arisen in the United Kingdom, with several complaints about the technology being reportedly made by rights organizations, nurses, and patient groups on the basis that VAPM raises privacy concerns, is objectionable, and is potentially harmful to patients [6,15]. In at least 1 National Health Service (NHS) Trust, complaints reportedly led to a cessation of VAPM [16]. These concerns suggest that even if further studies confirmed benefits to patients' sleep or other areas, such as reduced self-harm or suicide, such findings would have to be considered alongside other social, ethical, and legal implications of the technology. The legal implications of VAPM, for example, are not well understood [17].

This paper reports the findings of a qualitative exploratory study, via a workshop with key stakeholders which asks about the anticipated implications of VAPM in mental health inpatient settings. The study occurred in Australia, where VAPM is not widespread but has been applied in aged care settings [18] and seems likely to be trialed in mental health care settings [19]. The study aims to consider not just the practical and clinical implications of VAPM, but also the social, ethical, and legal issues, situating VAPM within broader debates about surveillance technologies in health care and their effects on privacy, consent, trust, and so on. This paper also draws on recent scholarship on new technologies of automation, such as AI ethics and algorithmic accountability, to analyze how VAPM may impact governance, decision-making, and trust in inpatient mental health care. We use these perspectives to assess the implications of data-driven monitoring for matters such as privacy, agency, and relationships of care.

Methods

Overview

We adopted a deliberative workshop format to elicit multistakeholder perspectives on the use of VAPM in mental health inpatient settings. Methods such as citizens' juries, co-design workshops, and deliberative panels are increasingly

XSL•FO
RenderX

used in health technology and medicolegal research to qualitatively explore ethical, legal, and social implications of new technologies [20,21]. They enable structured engagement with complex dynamics and trade-offs concerning issues of privacy and safety and are supported by recent empirical examples demonstrating their value in AI health care deliberation and digital tool development [22,23].

An online workshop was held in December 2021 with 16 participants and the 6 study authors, as well as an additional research assistant and an independent facilitator. The authors worked with an independent facilitator with expertise in group process design to develop an online workshop that was interactive and would provide participants with the opportunity to engage with evidence, deliberate, and respond both in verbal and written formats.

Recruitment

Using networking techniques, each coauthor identified 4 individuals who were invited to the workshop in a process of purposive sampling, informed by criteria concerning participants' appropriateness, capacity, and willingness to participate [24]. We aimed to recruit participants from 5 broad groups: consumers, nurses and nurse academics, health care managers, technology vendors, and ethico-legal and psychology academics. Invitations were sent directly to participants via email. As academics, practitioners, and advocates who research extensively in the field, we drew on our networks, as well as "cold-calling" potential participants to meet recruitment targets.

Participants

The 16 participants included 3 mental health consumers, 7 specialized mental health nurses (employed across direct care and academic and policy roles), 2 managers of acute mental health care services, 2 patient monitoring technology vendors, 1 legal academic, and 1 public health academic who specializes in inpatient hospital settings. The research team included a consumer academic, a legal academic, a computing academic who specializes in digital mental health, 2 mental health nurse academic researchers, and a health informatician.

Given that patients are most directly impacted by VAPM, the authors prioritized the recruitment of mental health "consumers," a term used in Australia to refer to someone who has used or continues to use mental health care services. ("Patient" is sometimes used throughout this paper, acknowledging that involuntary patients in mental health inpatient settings do not choose to be there in the sense conveyed by "consumer.") We sought a diversity of consumer perspectives on the topic of VAPM [25], but only 3 were able to attend the workshop, all of whom had served in representative and advocacy roles that required them to understand and articulate the views of diverse consumers. This latter experience enabled them to engage with both the governance and policy dimensions of VAPM and to consider VAPM based on their personal experience of inpatient mental health settings. A total of 7 consumers were invited to participate, all of whom expressed interest, but only 3 of whom were able to make the scheduled workshop.

We also prioritized the recruitment of nurses, given the implications of VAPM for care work. Psychiatrists and health service managers were invited, as they would be expected to make resourcing and care design decisions and interact with VAPM in mental health inpatient settings. Two vendors of VAPM who worked in a private technology company were invited, given technology vendors are involved in the creation, sale, and implementation of these technology types. Several scholars of health law and ethics were invited, but did not reply or could not attend, except for 1 legal academic.

The participants have been coded according to their stakeholder group, which is presented in Table 1.

XSL•FO
RenderX

Table . Participant codes and stakeholder groups.

Participant or researcher code	Participant group
P1	Technology vendor
P2	Consumer
P3	Health service manager
P4	Nurse
P5	Health informatics academic
P6	Legal academic
P7	Consumer academic
P8	Nurse
P9	Nurse academic
P10	Technology vendor
P11	Consumer
P12	Nurse academic
P13	Nurse
P14	Nurse academic
P15	Health service manager (with a clinical background)
P16	Nurse academic

Ethical Considerations

This study received ethics approval from the University of Melbourne Ethics Committee (reference: 2021-22674-23512-4). All participants provided written and verbal informed consent before participating in the workshop. Participants were informed about the purpose of the study, the nature of their participation, and the intended use of their contributions, including publication of anonymized findings. Participation was voluntary, and participants could decline to answer questions or withdraw from the study prior to publication. To protect participant privacy and confidentiality, all contributions were anonymized, and participants are identified only by participant codes and broad stakeholder groups in the paper. No identifying information is reported. Participants were offered an Aus $50 (US $34.50) honorarium to participate if they were not able to otherwise attend in a professional capacity.

Reflexivity and Researcher Positionality

All authors approached VAPM without settled views and with an exploratory orientation. One author (the legal academic) had previously written about the ethico-legal issues raised by VAPM [17], and the consumer academic's interest in the topic started with experience serving on a health service advisory group that had been asked to share views on VAPM. Reflexivity was addressed through ongoing team discussion, with attention throughout to separating normative concerns about VAPM from faithfully reporting participant accounts. Throughout the analysis, we aimed to let the data guide theme development, while acknowledging that researchers' standpoints shape interpretation and therefore warrant explicit disclosure rather than being glossed over.

Data Collection

The 3-hour workshop was moderated by an independent facilitator, with breakout discussions in small groups facilitated by the 6 authors. The workshop first entailed a plenary session in which members of the research team summarized existing empirical research on the use of the monitoring systems in acute psychiatric settings. Participants were provided information before the workshop, which included a summary of the literature on VAPM. Participants were separated into small groups and asked to share their views. Each group comprised a diverse range of participants (rather than being separated into cohort groups).

The small group discussions were an adapted form of focus group, a method offering rich descriptions by relevant actors of a phenomenon and the context in which it operates [26]. Breakout discussions were facilitated with the aim of offering a neutral, semistructured approach. Facilitators sought to guide the discussion to remain focused on the workshop prompts while refraining from evaluative commentary on the merits of particular viewpoints.

The audio recording of all workshop group discussions was transcribed for analysis. In total, 5.5 hours of audio-recorded data were generated.

Analysis

Analysis proceeded using thematic content analysis to capture broader group-level themes across the data [27], and a descriptive phenomenological method to frame and analyze the subjective, experiential accounts from workshop participants (particularly if they shared personal narratives, which occurred with some consumers and mental health nurses). Descriptive phenomenology takes an interpretivist stance, well-suited to practice-oriented research that seeks to record and understand

XSL•FO
RenderX

participants' lived experience of a phenomenon while suspending researcher assumptions [28].

Four of the coauthors (SD, BH, TK, PG) used a blended approach of inductive and abductive coding [29] to code the workshop transcripts. This was done in person (not using coding software), using printed transcripts, and codes were written on sticky notes before being grouped into themes on a physical whiteboard. Each transcript was read by at least 2 researchers to ensure coding consistency. The research team met to compare interpretations, discuss points of divergence, and refine the developing thematic structure.

Differences in coding were resolved through discussion and consensus. Consistent with a reflexive thematic analysis approach, we did not calculate a formal statistical measure of intercoder reliability. Rather than seeking statistical reproducibility, our aim was analytic rigor through collaborative scrutiny, reflexive dialogue, and iterative theme development.

Quotes were selected to exemplify themes. Participants did not review findings before publication.

We did not aim to achieve thematic saturation in the sense of exhaustively capturing all possible views on VAPM. Rather, the purpose of the workshop was exploratory, designed to elicit and compare the reasoning, concerns, and anticipated benefits articulated by key stakeholder groups and to identify points of convergence and contestation relevant to decisions about VAPM, including whether it ought to be used, and if so, how it should be governed. In qualitative approaches aimed at interpretive thematic analysis and exploratory issue-mapping, "saturation" can be a poor fit where the aim is to illuminate key tensions and meaning-making rather than to enumerate all views [30].

Results

The analysis led to the development of 7 themes, some of which included subthemes. In Table S1 in Multimedia Appendix 1, we provide illustrative quotes for the cross-cutting theme and each major theme and subtheme to convey the findings overall, before summarizing each.

Contestation Over the Rationale for VAPM in Mental Health Settings

A recurring thread across the findings was contestation over the rationale for using VAPM in acute psychiatric settings and its unique implications in the mental health care context. A fundamental tension appeared in the discussion between the utilitarian ambition for enhanced safety and efficiency (often driven by administrators and technology vendors) and what might be described as a relational ethics of care approach, which prioritizes the ethical imperative to protect human dignity, privacy, and the relational essence of care (strongly advocated by consumers and many nurse and academic participants).

A utilitarian account of the technology considers its potential as a versatile tool that provides an "additional layer" of monitoring and "visibility" (Participant 10) in scenarios where continuous human observation is not feasible. From this view, the technology has the capacity, principally, to enhance patient safety, reduce physical disruptions, and provide valuable operational support to clinical staff. Those sympathetic to this view highlighted its potential utility in proactively identifying disturbed sleep patterns, detecting medical emergencies, and enhancing safety measures against potential harm from other consumers or staff—and thus viewed it as potentially amenable to ethical and high-quality care. From this view, VAPM can support and extend the reach of clinical care, particularly when direct physical presence is limited.

However, most participants focused on the potential negative effects of VAPM. Some—particularly the consumers—expressed profound concerns that it represents a dehumanizing, intrusive, and potentially harmful form of surveillance that undermines the core principles of high-quality, contemporary mental health care. Key aspects of this account include invasion of privacy, exacerbation of distress, erosion of trust and therapeutic relationship, potential for misuse and lack of agency, and complacency and deskilling of staff. Consumer participants with this view expressed a profound rejection of the technology in patient bedrooms.

These 2 broad accounts seemed to be underpinned by differing accounts of the ways in which the mental health care context was seen as unique compared to other areas of medicine. Most participants, particularly consumers and nurses, queried whether a tool primarily designed for physiological monitoring would align with the distinct needs and expectations of mental health care, where therapeutic engagement and trust are paramount. For example, one consumer said:

> *There's no electronic monitoring that will tell you that my hope was just crushed [...] The only way you can know that we're thinking and feeling that is to talk with us and listen to us.* [Participant 2]

A nurse (Participant 4) felt the proposal aligns with an "overly restrictive non-therapeutic program" and moves away from "strengths-based consumer-led models of care."

In contrast, a hospital manager (Participant 15) raised a comparison with general intensive care units, noting that in physical hospital care settings, staff "don't blink an eye" at monitoring physiological data. He added, "[a]s a rule, we've never had that facility in our ICU settings" within mental health, implicitly suggesting that the disparity in technological monitoring between the 2 creates a disparity in the safety and quality of care. Here, mental health care was viewed as different because it did not have parity of care with general health, which VAPM could help remedy.

All participants acknowledged the potential for coercion in public psychiatric hospital units, with several participants noting that this technology will have unique implications, given that patients or consumers are already significantly disempowered and monitored in (often locked) mental health care contexts.

VAPM Will Reshape Care and Relationships

Overview

An overarching theme among participants was the impact of VAPM on care and the people who give and receive it. Most participants expressed a concern with the potentially

XSL•FO
RenderX

dehumanizing impact of monitoring and surveilling people's bodies for "signs" and movements, instead of asking people about their distress or checking on them face-to-face. The extent to which VAPM was indeed viewed in this way varied in three ways, between (1) those who felt certain it was dehumanizing, (2) those who felt it potentially dehumanized but that more detail was required alongside consideration of potential benefits, and (3) those who felt it was a potentially exciting opportunity to transform care, and indeed to "rehumanize" care by freeing up care professionals' time for more valuable face-to-face care.

VAPM May Dehumanize Care Depending on How It Is Used

Participants raised competing concerns about whether VAPM would support or undermine care.

In total, 3 participants expressly described VAPM as "dehumanising" (Participant 2, Participant 4, and Participant 13). One participant said it felt "grotesque" and "the complete opposite direction that we should be heading for good care" (Participant 2). Such accounts conveyed not only fear of misuse but also anger at the idea that replacing observation with surveillance contradicts the relational ethos of mental health practice. Participants also worried that remote viewing could reduce therapeutic contact and intensify privacy intrusion; during in-person monitoring, a patient could at least know when they are being observed, unlike remote continuous monitoring.

Concerns extended beyond consumers to staff. One nurse warned that VAPM may be "dehumanising for the person you're serving, but also the workforce" (Participant 4), moving practice away from strengths-based and consumer-led approaches. For these participants, the possibility that biometric monitoring could detect crises did not outweigh the loss of direct contact, trust, and the complex facets of human care.

VAPM Could Humanize Care by Creating More Time for It

Some participants suggested that VAPM could actually create conditions for more humanized care. By automating routine and time-consuming observation tasks, the technology was seen as potentially freeing staff to focus on therapeutic interaction. A nurse academic explained that much of the current workload "is really not about actually caring for the person," but if monitoring was managed by technology, staff could "use our time and resources to actually support the person in a more therapeutic way" (Participant 14). This was echoed in specific clinical contexts, such as eating disorder units, where constant physical checks can be disruptive; VAPM was imagined as a way to reduce unnecessary physical intrusion while still safeguarding health.

Others envisioned VAPM as a "backup" safety mechanism that might enable staff to spend more time present on the ward. A hospital manager (Participant 3) linked this possibility to wider efforts to design units that keep staff "out there with the consumers" rather than tied to nursing stations. In this framing, the technology could reduce the burden of repetitive checks while fostering closer interaction, more group activity, and ultimately safer environments by reducing suicides (discussed in the VAPM as Observational Support for Safety and Reduced Disruption section). These perspectives framed VAPM not as a substitute for relational care, but as a tool that might support and expand it, if used judiciously.

Consent to Being Monitored

Some participants proposed that VAPM could be made more acceptable—and thus more "humanizing"—through a clear consent process, giving patients the choice to opt in or out depending on their needs. A nursing academic (Participant 12) suggested that for some, the technology might be preferable to the disruption of in-person checks during the night, allowing people to choose "what's going to be better for me." In this framing, VAPM could be a tool of choice, reducing intrusive physical monitoring and supporting therapeutic rest where it aligned with a consumer's comfort and sense of safety.

Yet, these proposals for consent were met with significant skepticism and distrust by some. Consumers worried that promises of choice would not be honored in practice: one participant doubted whether the system would truly be switched off if they declined and feared it could be misused by staff, undermining any sense of agency (Participant 2). Others warned that systemic pressures, such as staff shortages, could lead to VAPM becoming a default, regardless of individual wishes (Participant 11 and Participant 4).

Contestation also arose over whether consent requirements should differ from current observation practices, with a vendor questioning why AI monitoring should need explicit permission when in-person night checks do not (Participant 10). Nurses responded that, unlike a glance through the door, VAPM automatically records physiological data, which in other contexts would require active consent for a patient with capacity (Participant 12). For one nurse academic, the idea that VAPM should be installed so there was "no way of avoiding it," and that this could be presented as a "selling point" was described as "very disturbing" (Participant 9). Taken together, these perspectives highlight potential issues with relying on consent as a bridge to acceptability in inpatient settings, where structural constraints and fears of surveillance may undermine aspirations for genuine choice. The distinct privacy implications of VAPM (including 'privacy of place' and 'privacy of data') are addressed in the section "VAPM Has Serious Privacy Implications."

Perceived Harms of VAPM

There was broad agreement that VAPM could cause harm, though participants differed on the extent of this harm and how it compared to potential benefits. Consumer participants, along with several nurses, voiced the strongest reservations. For some, the concern was visceral: Participant 5, a public health academic who specializes in researching inpatient settings, described her "gut" as:

> *really uncomfortable ... if I was in that position, I wouldn't like it. I wouldn't like the experience of having a video camera in a place where I was sleeping.*

Nurses also expressed apprehension, with Participant 4 warning of "human rights and ethical considerations and safety," a theme echoed by others.

XSL•FO
RenderX

Several participants argued that VAPM could in fact undermine, rather than enhance, safety. One consumer explained:

> *I'd be trying to hit it with the chair, put water on it to anything I could to damage it because I'd be I'd feel really unsafe from it. The more distressed I was, the more I'd feel like that.* [Participant 2]

She added that surveillance might not stop self-harm, but simply shift it elsewhere: "if there's monitoring in the bedroom, then I'd self-injure on the toilet, you know." A nurse (Participant 13) cautioned that the technology could encourage "complacency" among staff, reducing engagement on the assumption that safety was assured. One of the vendors (Participant 1) acknowledged the same risk, noting that staff might fall into a "false sense of security" and miss critical issues.

Participants also highlighted the possibility of secondary and longer-term harms. Participant 7, a consumer, warned, "If I went into a ward that had these in every single bedroom, I'd probably flip lid." Both consumers and nurses suggested that VAPM could "exacerbate paranoia" (Participant 7), "play into ... someone's mental state" (Participant 1, technology vendor), or worsen anxiety. The technology was also described by two participants as potentially retraumatizing (Participant 12 and Participant 13), particularly for people with experiences of domestic violence in which surveillance had been used as a tool of abuse. Others raised concerns that awareness of such monitoring might deter people from seeking help in services where VAPM is in place. Some participants thus framed privacy intrusion as a harm in itself and a source of distress and mistrust. Detailed privacy implications (privacy of place and privacy of data) are addressed in theme "VAPM has serious privacy implications."

VAPM as Observational Support for Safety and Reduced Disruption

Participants who saw potential benefits framed VAPM as a way to support observation in two related ways: (1) risk detection or prevention and (2) less disruptive routine checks, especially overnight. Comments included that the monitoring offers an "exciting opportunity if it's used well" (Participant 14), provides an "opportunity to augment our care" and "prevent deaths" (Participant 15), and helps provide "continuous monitoring ... [and] constant data feeding back" (Participant 13).

Preventing Rare but Severe Adverse Events

Some participants argued that VAPM could add an extra layer of visibility to help detect and prevent rare but serious events, including suicide, sexual assault, or medical deterioration. Participant 15 (hospital manager) stated that although rare, deaths in hospitals "do occur frequently enough" and said his "primary concern is always the safety of the consumers or patients":

> *some of our patients unfortunately pass away in the middle of the night ... And of course, the other issue is the suicide risk, in the night. So the extent that this technology may help us manage prevent deaths, I think that's an important consideration.*

Participant 13 (nurse) highlighted the importance of mitigating interpersonal risks, particularly for vulnerable consumers, stressing the need for "sexual safety" and "being able to ensure that there's not a person entering another person's bedroom. I feel like that would offer significant reassurance to particularly our female consumers."

Physical health monitoring was seen as potentially beneficial when predicting, preventing, or responding to someone having a physical medical episode (eg, a stroke, fall, or asthma attack) or attempting suicide. Participant 15 (hospital manager) explicitly mentioned preventing deaths from issues like respiratory depression and managing suicide risk at night as key concerns occupying his mind, stating that "[a]s a clinician, this is what makes me deeply uncomfortable and I think this is an opportunity to augment our care."

Reducing Disruption From Routine Observation

Others focused on using VAPM to reduce sleep disruption and intrusive checks while still supporting physical monitoring. Participant 16 (nurse academic) noted that consumers frequently comment about the "very untherapeutic environment that nighttime [routine observations of every consumer] have on people and that disruption to sleep" and invasion of privacy, suggesting that the technology could offer "some positive aspects to that." Participant 13 (nurse) said that "lack of overnight disruption would be great, too." Participant 12 (nurse) agreed and described current night checks as "terrible," recalling that staff "always feel really bad" about waking patients up and shining a torch on them.

VAPM Has Serious Privacy Implications

Most participants raised concerns about intrusions on privacy. Participants distinguished between privacy of place—bedrooms and other intimate spaces—and privacy of data—the capture, storage, access, and potential secondary uses of video and physiological information.

Privacy of Place: VAPM Reshaping Privacy in Intimate Spaces

It is accepted that patients within inpatient mental health settings are subject to significant observation, even to the extent that observation is the primary objective of some hospital admissions. However, participants viewed VAPM as enabling observation of a different nature and scale, with heightened risks of intrusion and misinterpretation.

A participant warned that VAPM could intrude on "a personal moment" (Participant 8, nurse) and "a private intimate moment" (Participant 9, consumer), such as masturbation or other personal activities. Without control over when monitoring occurs, observation becomes constant and intrusive, and misinterpretation of what is seen could negatively affect the care someone receives.

Participant 15 (hospital manager) suggested that privacy and confidentiality risks could be mitigated by limiting the visual detail available to staff, for example, by displaying only a "pixelated body form" (sometimes referred to as a "point cloud") rather than a live video feed or detailed image of a person

XSL•FO
RenderX

"sleeping or going to the toilet or whatever and doing something private."

Privacy of Data: Handling, Retention, and Secondary Use

Concerns about data handling, storage, accessibility, and the eventual use of the generated consumer data were widely discussed among participants, demonstrating apprehension closely related to privacy and surveillance. Consumers expressed fears about how information could be retained and later misused, with one saying "[W]e hear lots of examples of from consumers of incidences from 10, 15 years ago being brought up now ... that affects their treatment now" (Participant 8) and querying whether that might occur with data collected by VAPM. Others questioned whether routine physiological data collected through VAPM would become part of a person's permanent medical record, with implications for how future admissions or treatments might be judged.

Nurses and consumers raised uncertainties about where and how the data would be used and stored, including whether it could be subpoenaed or whether it might be used in staff performance management, likening the situation to "police walking around with our cameras on" (Participant 9). Beyond immediate clinical settings, participants worried about the commodification of data, even in deidentified form. As one consumer academic put it, such datasets could be sold or repurposed to create behavioral profiles, with the risk of reinforcing surveillance and targeting of certain groups (Participant 11). Together, whether these perceptions were correct or not, these sentiments reflect concerns about how data generated by VAPM may pose risks not only for individual privacy but also for equity, trust, and the long-term integrity of mental health care.

VAPM Requires Appropriate Governance

Procedures and Guidelines to Support Use of VAPM

Most participants emphasized the importance of carefully developed guidance and rules on use, should a VAPM trial proceed. A nursing academic described the need for "strict guidelines necessary on how it supplements care rather than replacing care," which "would need to be really clear [as to] the expectations on staff on how it was to be used appropriately" (Participant 16). Concerns were also raised about responsibility in the event of a critical incident. A consumer (Participant 11), who had a legal background, noted that good "regulation and governance" must be in place, as these new technologies "stretch the existing structures that we have" and may even create liability under legal obligations that require a certain standard of care (evaluating the accuracy of this claim is outside the scope of this paper). A hospital manager described the need for a "fairly robust procedural implementation process to make sure that people's concerns were addressed" (Participant 15).

Legal and clinical participants worried that because the technology is proprietary, its inner workings would remain opaque, preventing proper scrutiny or challenge. Participant 6 (lawyer) noted that descriptions of the technology seem to be "always really vague." She questioned the nature of the AI and algorithm behind VAPM, stating that the inner workings of privately developed technologies more generally, which are sold to health services, are:

> *not explained because it's proprietary. Someone owns it and so on, which then makes it even harder to kind of scrutinise, let alone to kind of challenge if something goes wrong.*

Liability was thus not only understood in terms of potential harm in individual cases, but also as matters of organizational governance and responsibility: managing sensitive data appropriately, ensuring adherence to staff protocols, establishing private sector accountability mechanisms, and safeguarding the human rights of vulnerable consumers. Participants also questioned who would authorize the technology, how it would fit within established models of care, and whether it would support or displace existing clinical practice. No clear answers emerged, but several participants emphasized that such technologies should ultimately support, rather than replace, the work of health care professionals.

Safeguards in the Adoption of VAPM

Participants who supported trialing VAPM generally agreed on the need for safeguards. Proposed safeguards included clear guidelines for use, technological modifications to protect privacy, meaningful consent processes, and, in Australian jurisdictions with human rights legislation, ensuring that VAPM is "assessed [as to] it's human rights compliance" (Participant 11). Several agreed that additional "pilot" studies would be needed to test claims being made about the technology. One nurse raised the point that she would want to learn from "a mental health unit that's using it, [in which] people have responded, [and in which] their experiences are positive" (Participant 4).

Is VAPM a Technological Solution to the Wrong Problem?

One consumer described VAPM as:

> *a solution to problems that shouldn't have existed to begin with. That's kind of, I think where I'm sitting, we're providing solutions to problems that just shouldn't be problems.* [Participant 7]

She elaborated by stating:

> *Oh, you won't be getting a light shone in your eyes every 15 minutes. That's that sounds amazing—but can't we just stop doing it anyway?*

This encapsulated a critique of the plan to rely on technology to fix issues that stem from what they view as poor clinical practice or systemic failures. Another consumer, who was trained as a nurse, noted that nurses do not need to shine a torch in someone's eyes; she was taught to "shine the torch on the roof of the ceiling and you get an ambient light that that gives you enough to actually read obs, but not disturb the patient" (Participant 8). This concern relates to the earlier point, raised by consumer participants, that VAPM bypasses the need for genuine, therapeutic human interaction, which is the foundational element of mental health care.

XSL•FO
RenderX

Confusion Over the Specifics of VAPM

Most participants identified several aspects of VAPM that they felt were unclear, and several questions they would want answered before forming a view of the technology and its likely impact on the service system in which it would be used.

Because the workshop was held in a national context where VAPM had not been used, participants were not assessing a single, fully specified VAPM system. Limited technical detail about system configuration, performance (eg, false alarms), data flows, access controls, data retention, and workflow integration influenced how participants responded. Given the speculative nature of the discussion, participants had to reason from general principles, such as privacy and dignity, and draw analogies with other care contexts, such as in-person monitoring. At the same time, this uncertainty was itself salient to participants and contributed to the strong emphasis on transparency, independent evaluation, and governance safeguards as prerequisites for any trial or adoption.

Participant 6 worried that if the training datasets for the AI were biased (intentionally or unintentionally), the technology might be "more accurate for some people" than others (eg, men more than women), information which staff would not know because the workings are proprietary. Participant 3, the hospital manager whose service was considering this technology, stated that a key goal of their planned pilot was to determine if the technology was actually "going to be able to deliver what it can from an alerting perspective"—a matter which was unknown. Participant 9 (nurse academic) suggested that evidence would be needed to determine whether VAPM kept nurses "in the nurses station more and [on] the floor less," and questioned whether monitoring would be restricted to nighttime use or applied more broadly. Participant 7 (consumer) questioned if it would be used only for a "couple of beds" or for the "whole ward."

For most participants, therefore, views on the appropriateness of the technology in acute psychiatric care hinged on unresolved questions about the empirical basis for the technology, its scope and functionality, and questions about how it would be governed, such as what procedures were in place to protect patient privacy and prevent misuse.

VAPM Has the Potential to Transform Not Just Augment Service Delivery

The potential service transformation created by the use of this sociotechnical system was another theme in the workshop, which the researchers subdivided into 3 themes: function creep, cost minimization and corporatization, and confusion about unresolved evidentiary and governance matters.

Risk of VAPM Function Creep

Some participants worried that the monitoring scope would quickly expand from passive overnight checks in some rooms to encompass day shifts, hallways, and potentially all consumer and patient rooms, eroding privacy and human interaction. Participant 8 (consumer) described a "slippery slope" in which allowing VAPM in some rooms today would mean "eventually they'll just be in every room," to the point "we stop having any interaction, any sort of relational space between clinical staff and the consumers." A hospital manager (Participant 15) described the reality of function creep in health systems generally: "[t]he way systems work, the way organisations work, these things just happen without people thinking about it properly and putting the correct controls in place."

VAPM as a Form of Cost Minimization and Privatization

A recurring fear by some participants was that VAPM would be used as a cost-cutting measure rather than a tool to improve care, with some participants concerned that it would justify a push for nurses to take on larger numbers of patients or consumers. Some participants worried that, in a risk-averse system with finance as a major driver, the technology might provide what a public health researcher described as "just... an excuse to reduce staff numbers to actually do things that will undermine the system improving, not actually support improvements" (Participant 5). A nursing academic (Participant 14) who supported its use in certain circumstances, similarly cautioned that it could be "used as an excuse to reduce the nursing resource and also that human interaction between nurses and consumers." A broader concern was raised by Participant 11 about the role of the private technology sector in the development of health care technologies used in public health settings.

Discussion

Principal Findings

This study used a workshop with a range of key stakeholders to explore the potential implications of VAPM in acute psychiatric settings. The findings provide key insights into the impact VAPM might have, primarily for patients or consumers and nurses.

VAPM raised issues that sit within broader debates about surveillance technologies in health care, which are often justified in terms of safety and efficiency but raise concerns about consent, privacy, data governance, and the impact of monitoring on trust and therapeutic relationships [31]. Our findings reflect these tensions in a mental health inpatient context, highlighting how anticipated benefits are entangled with concerns about intrusion, misuse, and the conditions required for legitimate and acceptable use.

Participants who saw potential benefits emphasized patient safety and care efficiency, but these were framed tentatively and with caution, as potential benefits. Some tension emerged between groups: hospital managers and technologists broadly supported its testing and development; nurses and consumers varied somewhat but generally viewed VAPM negatively, with some qualifications. Most consumer participants were of the view that VAPM should be rejected, following its characterization as a form of intrusive surveillance that dehumanized patients and those who were supposed to care for them (for a similar viewpoint in the literature, see Powell et al [7]). Some participants, particularly nurses, saw potential use in specific circumstances and only if there was compelling empirical evidence to support its use, and ideally with patient consent. Most consumer and nursing participants articulated

XSL•FO
RenderX

the view that the technology risked undermining nursing practice by replacing nursing functions and limiting the ability of nurses to develop therapeutic relationships. Across groups, the lack of robust pilot studies was a recurring concern, leaving many participants hesitant to endorse the technology, while its strongest critics rejected the premise of its use altogether.

These findings could be used to argue that these technologies should not be adopted due to the risks they pose to the privacy and dignity of patients. Alternatively, they could also be used to inform the design and implementation of these technologies to ensure that the sociotechnical system of VAPMs contributes not only to clinical and management outcomes but also to dignity, experience, and privacy outcomes for patients.

Regarding therapeutic interaction, participants were clear that VAPM should not replace human care. Nurses acknowledged the value of reducing intrusive overnight checks but emphasized that fewer face-to-face interactions might also limit opportunities for rapport and engagement. Although some were cautiously open to exploring alternatives that would minimize sleep disruption, nurses and consumers wanted clarity about how the technology would be used, how privacy would be safeguarded, and how such matters would be communicated to patients. Participants alluded to the cost minimization objective seeming to butt up against claims that VAPM would be an adjunct to—rather than replacement of—face-to-face care, echoing the observation of Griffiths et al [14] that "it is unclear how this adjunctive role is envisioned alongside the stated aim of cost reduction." These concerns were often expressed by participants who explicitly or tacitly seemed to acknowledge that consumers and frontline staff are rarely the ultimate decision-makers in mental health hospitals.

In relation to privacy, VAPM was described as potentially less invasive if used in ways that reduce human intrusion and rely on abstracted physiological data rather than live video streams. From this view, VAPM could enhance care without reproducing the direct visual surveillance of closed-circuit television. However, participants cautioned that whether or not video feeds were actively viewed might matter little to patients, whose perceptions or experience of constant monitoring would still be experienced as intrusive.

Consent was a further complication, and our findings highlight challenges to individual autonomy, particularly within the coercive context of involuntary inpatient care. Some participants' skepticism about opt-out arrangements and the fear that monitoring would continue regardless of refusal suggests that consent is, at best, a partial safeguard for VAPM in these settings. Consent processes may improve acceptability for some (particularly where monitoring is time-limited and genuinely optional), but they cannot be relied on as the primary ethical justification for deployment.

While some argued for an opt-in process, others noted that in mental health inpatient settings, where involuntary psychiatric intervention is common [32], it is unclear whether genuine consent for VAPM will be possible in some cases or legally required. Legal status does not map neatly onto decision-making capacity, which may fluctuate, but the conditions of admission, power imbalance, and role of risk-aversion in services could make refusal difficult in practice. Furthermore, opting out may not necessarily be respected or even feasible in a resource-constrained system. Gooding and Clifford [17] have even queried (tentatively) whether VAPM in mental health inpatient settings would conform with current law in some Australian jurisdictions, particularly if video is used. In the United Kingdom, the nongovernment organization Rethink Mental Illness argues that in addition to being required to meet medical-device regulatory requirements, VAPM should be conceptualized as "a form of restrictive practice that breaches people's rights, with potential short and longer-term effects on people" [33]. These considerations also suggest that VAPM may (inadvertently) facilitate algorithmic paternalism [34], where algorithmic risk-management overrides a patient's agency or autonomy under the guise of safety. An individual's lack of agency over data captured from their most private moments, such as sleep or personal activities, may represent a fundamental breach of dignity that cannot be solved by a simple opt-in or opt-out checkbox. These concerns suggest that consent is, at best, a partial safeguard for VAPM in these settings. Consent processes may improve acceptability for some (particularly where monitoring is time-limited and genuinely optional), but they cannot be relied on as the primary ethical justification for deployment.

Uncertainty on matters of regulation and governance among the participants suggested that considerable planning and oversight would be needed before monitoring systems were trialed or adopted. Participants seemed to stress (and we agree) that any plans for implementation must go beyond minimum legal requirements to include consideration of social or ethical acceptance and perceived benefit among those most affected. On a related note, VAPM illustrates the insufficiency of perceived usefulness and perceived ease of use, the 2 pillars of the original technology acceptance model, which is used to explain whether, how, and why users adopt or reject a new technology [35]. Technology acceptance, particularly in health care, is not just about individual perceptions of usefulness and ease of use, but also about trust, perceived risk, and the social context in which the technology is introduced.

Again, such perspectives on implementation diverged along stakeholder lines. Consumers and nurses, whose daily experiences would be most directly affected, tended to be skeptical or resistant. They saw risks to therapeutic relationships and to their sense of agency and privacy. Technology vendors, by contrast, sought to promote their products by identifying problems to solve, and managers focused on organizational priorities rather than the details of nursing practice. Such contrasts highlight the importance of recognizing not only official intentions for new technologies but also the way they are interpreted and taken up.

On this latter point, there is some evidence from research on electronic health records that perceptions strongly shape adoption, where early concerns by stakeholders with electronic health records continued to influence professional practice long after implementation [36]. Management and technology providers may communicate their intention for a technology, but it is the users' interpretation of the technology that plays a key role in how it is used and whether it is used as planned or

XSL•FO
RenderX

not. Here, "users" are ostensibly nurses and other health practitioners at the ward level and could potentially extend to patients or consumers, though "users" seem an unsatisfactory description of this group. (On this latter point, Boetto [37] raises concerns in social care contexts that "end users" of AI are often envisaged as policymakers, managers, and practitioners, but not the service users who will ultimately experience its impacts and are best placed to identify risks and benefits.) Regardless, early perceptions such as those identified in this study can help us understand how end users and those affected will use and engage with the technology if implemented.

The theory of contextual integrity [38] may offer a useful lens for these concerns. Privacy, on this view, is defined by the appropriate flow of information within a specific context. What is appropriate is defined by context-specific information norms. Information norms that guide appropriate information flow can be mapped using the 5 parameters of actors (sender, receiver, and subject), information type, and transmission principle (which are constraints on information flow, such as consent). VAPM challenges established norms across several of these parameters by enabling more extensive data collection (eg, continuous rather than occasional checks), involving more actors (beyond the nurse at the door), and removing established principles of reciprocity and warning (patients can no longer see or hear when they are being observed). These shifts raise questions about whether such information flows are justified by reference to contextual values and whether less intrusive alternatives might achieve the same ends. Nissenbaum [39] suggests that, even where new technologies breach existing information norms, consideration should be given to whether the new information flows introduced by the technology serve the interests of different stakeholders, and whether they align with the ethical and political values of the specific context. Such an evaluation may result in breaches of existing information flows being viewed as justified. In mental health care, this question of alignment is especially difficult because ethical and political values are contested, particularly in inpatient settings involving involuntary psychiatric intervention, where competing views persist about whether risk and safety management should take precedence over the choices and rights of service users.

Applying contextual integrity also points toward possible design modifications. For instance, proactive notifications could alert patients when monitoring is active, or monitoring could be restricted to times when patients are asleep, thereby reducing unnecessary intrusion. Patients might also be given access to the data collected, for example, via a display screen. Such features potentially mitigate some of the privacy concerns raised in this study and afford patients a degree of control.

The erosion of digital trust [40] identified by participants is linked to the violation of contextual integrity. Trust in a mental health setting is built on the expectation that information remains within the boundaries of the clinician-patient relationship. VAPM challenges these established norms by involving new, often invisible actors, such as third-party technology vendors and data analysts. This in turn raises concerns about function creep and the commodification of patient data, that data collected for safety monitoring might eventually be used for purposes such as staff performance management or sold to create behavioral profiles. Establishing digital trust in VAPM, therefore, requires more than technical accuracy; it demands algorithmic accountability and transparent data stewardship that ensures information flows remain aligned with the ethical values of the psychiatric context.

Finally, there remains the question of how VAPM affects mental health itself. Paranoia and persecutory beliefs are common among people admitted to mental health inpatient settings [41], with evidence suggesting more than 70% of patients at first-episode psychosis experiencing some form of persecutory delusion [42]. Early evidence suggests that surveillance technologies may heighten such fears for some patients who have reported concerns that devices could "emit damaging rays" or "control them" [9], notwithstanding the potential benefits noted earlier. Future research is needed to assess whether purported benefits outweigh the potential negative psychological impacts of VAPM in these contexts.

Implications for Policy and Practice

The findings of this study raise serious issues for future research, policy, and practice, which we summarize in Textbox 1.

Textbox 1. Recommended steps for hospital managers, policymakers, and so on.

- Start with an appraisal decision, not implementation: use structured deliberative processes (like this workshop) to decide whether video-algorithmic patient monitoring is appropriate in particular inpatient mental health settings.
- Define the "nonnegotiables" up front: identify minimum requirements relating to dignity, privacy, trust, and therapeutic relationships that must be met before video-algorithmic patient monitoring can be considered.
- Specify the proposed system and use-case before any decision: clarify what is being deployed (video vs abstracted data, alerting functions, where or when used, and who can access) so stakeholders are not evaluating an undefined intervention.
- Require independent evidence as a precondition: treat claims about safety, efficiency, and cost as provisional unless supported by independent, publicly available evaluation.
- Assess consent feasibility in involuntary contexts: determine whether meaningful consent or opt-out is possible in practice; if not, treat this as a major barrier requiring a higher threshold of justification.
- If proceeding, set governance first: establish binding rules on purpose limitation, oversight and accountability, transparency, data handling or retention, auditability, and safeguards against misuse and function creep, with meaningful consumer participation.

We elaborate on 3 points here.

First, given the contested and still limited evidence base, the findings support the need for independent, coproduced research

XSL•FO
RenderX

that examines both the effects of VAPM and the conditions under which it might be regarded as legitimate and acceptable. Participants consistently questioned whether claimed benefits would be realized in practice, or whether they would be offset by risks to privacy, dignity, and therapeutic relationships. Independent, high-quality research that centers lived experience perspectives alongside frontline staff perspectives is therefore critical for informing any decision to proceed. Where VAPM is already in use, such evaluations should examine uptake across contexts, impacts on professional practice and ward culture, and the experiences of patients or consumers and staff. If a trial is contemplated, it should be treated as a research activity with transparent protocols and publicly reported outcomes, rather than a step toward routine adoption.

Second, participants made clear that decisions cannot be made about an ill-defined intervention. To reduce confusion and enable informed appraisal, any proposal to use VAPM should be accompanied by a clear description of the system and use-case: whether monitoring relies on video or abstracted data; when and where it is active (eg, selected beds vs whole wards and night-time vs broader use); measurable outcomes (eg, reducing false alerts, patient safety metrics, trust levels, incident reduction, and staff time spent with patients); who can access what information; and how data are stored, retained, and governed.

Processes for avoiding concerns about conflicts of interest must be clear [43]. On this latter point, one of the post-2021 studies reporting reduced self-harm following implementation was subsequently retracted [43]. While the retraction relates to an omitted conflict-of-interest declaration rather than an adjudicated finding of data error, it nonetheless highlights fragility in the VAPM evidence base and reinforces the importance of transparent disclosure and independent evaluation. As such, a transparent evaluation plan should be put in place to determine the accuracy of the technology and its impact on staff behavior, with results made public before any wider roll-out. By providing this clarity to all key groups, services can directly address the questions raised by staff, managers, and consumers in this study.

Finally, if VAPM is to be pursued, we would suggest that governance is established before implementation and should be guided by clear and collaboratively developed standards. Participants' concerns point to the need for standards on purpose limitation, oversight and accountability, consent and choice (including how these operate in involuntary settings), auditability, and safeguards against function creep and secondary uses (including performance management). Policymakers and hospital administrators may wish to draw on participatory governance and co-design models used in digital mental health, including standing lived-experience advisory groups, coproduced implementation planning, and deliberative panels used to set acceptable data practices and safeguards [44-46]. Examples include Australia's National Digital Mental Health Standards, which includes a "Consumer Partnership Principle" requiring service-user partnership across the planning, design, delivery, measurement, review, and evaluation of services [44], which can be adopted through practices like sustained lived-experience advisory groups that shape design and implementation over time [47].

Strengths and Limitations

The strength of this study included the relatively diverse sample. This mix of participants—consumers, mental health nurses, health managers, and technology vendors—increased the diversity of views and minimized the risk of a single perspective dominating. Another strength was the coproduction process whereby the workshop was coplanned, co-designed, and coproduced with traditional academics and a consumer academic. This study appears to be the only empirical study on VAPM in mental health settings with consumer academic leadership and which provided for consumers to be equally important contributors to understanding the social, ethical, and legal implications of the technology. Finally, it could be considered a strength that this is an independent academic study with no vested interests, unlike most studies in the field (for an overview, see Griffiths et al [14]).

One limitation, regarding diverse consumer perspectives, is that we recruited only 3 consumers. Recruitment was constrained primarily by scheduling. Although additional consumers were invited to participate, they were not available at the workshop time. The 3 consumer participants had extensive experience in consumer advocacy roles, which may be associated with higher confidence in formal deliberation, greater familiarity with service systems and rights-based discourse, and an ability to articulate concerns in more conceptual terms than may be the case for others. Accordingly, the findings should not be treated as representative of the broader service user population, and the consumer accounts should be read as informed perspectives rather than a cross-section of typical inpatient views. However, our aim was exploratory: to elicit and compare stakeholder reasoning about the acceptability and implications of VAPM. On that basis, the consumer contributions were substantively valuable, foregrounding how privacy, coercion, dignity, and trust may be experienced and reasoned about in relation to VAPM. We do not claim statistical generalizability; instead, the findings offer analytic insights and issues for governance and practice that should be tested in future research with larger and more diverse consumer samples.

Purposive sampling was used, which supported a somewhat diverse sample but left the potential to amplify the authors' biases, given the sample was effectively determined by author networks. Author positions may have influenced the framing of questions and interpretation. To mitigate this, we used reflexive practices across the study (including explicit efforts to set aside normative claims and a commitment to grounding themes in participants' accounts and illustrative quotations). Nonetheless, positionality cannot be eliminated in qualitative research; hence, we have sought to make it transparent so readers can assess how the team's perspectives may have shaped the analysis. The potential to amplify bias is somewhat offset by the consistency of the study aims (an exploratory study of viewpoints among people most likely to be impacted by the introduction of a technology), criteria (participants belonging to the main groups impacted), and epistemological basis of the research (valuing the qualitative input of those most affected in generating knowledge).

XSL•FO
RenderX

Response bias is a limitation. In this mixed-stakeholder deliberative workshop, participants may have tempered comments due to social desirability, professional norms around safety, or the presence of managers (in the case of nurses) or vendors. We mitigated this through independent moderation, neutral semistructured facilitation, and options for anonymous reporting, but some more critical or personal views may nonetheless be understated.

A further limitation is the time lag between data collection (December 2021) and publication. In the intervening period, the VAPM landscape—particularly in the United Kingdom—has evolved, with additional service evaluations and reviews published and wider use reported in some NHS mental health settings. There has also been increased public scrutiny and organized opposition, alongside policy attention, including an NHS-commissioned evidence review and work toward national guidance on vision-based monitoring. These developments mean that some issues raised in our workshop (eg, potential efficiency gains overnight) can now be compared against a larger, albeit still limited, evidence base, while other concerns (privacy, consent, dignity, and trust) remain live and contested in implementation debates.

As such, the study was somewhat hampered by the lack of empirical research at the time of the workshops, which had evaluated VAPM on its own terms, meaning that claims about its appropriateness could not be contrasted with claims about efficacy. Nor could we invite people who had encountered VAPM in the mental health care context, whether as patients, nurses, or hospital managers, given that it has not been trialed in Australian mental health inpatient settings to our knowledge, making the discussion largely hypothetical without the benefit of a specific VAPM definition or configuration for these inpatient settings. On the other hand, what this study suggests is that before technical development or empirical testing even takes place, the perspectives—if not leadership—of consumers and those working in mental health settings on the proposal should be sought.

Nevertheless, the workshop provided a valuable forum for persons likely to be impacted by VAPM to voice their views. There is value in deliberative input in the early stages of technology adoption or experimentation, particularly sociotechnical systems like VAPM. Our approach potentially offers a template for health services and researchers seeking to test views on proposed technologies, including for the purposes of risk mitigation, social acceptance, and safety.

Conclusions and Future Research

This study highlights the considerable uncertainty surrounding the anticipated use of VAPM in mental health inpatient settings. Currently, anticipated risks are primarily articulated as subjective, as potential interference with privacy, dignity, and trust, while the claimed benefits lack robust and independent evaluation. Ethical and practical concerns voiced by consumers and frontline staff suggest that wide adoption would be premature, and some accounts that tend toward an ethics of care approach rejected the entire premise of VAPM. Any future trials must clearly define the technology's purpose and mechanisms and be independently evaluated against not only safety and efficiency claims but also impacts on dignity, privacy, and therapeutic relationships. If VAPM is trialed, the findings of this study could inform both the technology configuration and model of care in which it is implemented. Crucially, research must be coproduced with those most directly affected—both staff and consumers, but particularly the latter, we would argue—to ensure that the technology serves the people it is intended to benefit.

Acknowledgments

The authors wish to thank the participants who took their time to be a part of this research. The authors would also like to thank Rusty Benson for his assistance as the independent facilitator who hosted the workshop.

Funding

This work was funded through a grant from the Centre for AI and Digital Ethics, University of Melbourne. PG acknowledges support for his contribution from the Australian Research Council (DE200100483).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Illustrative quotations from the deliberative workshop, organized by theme and subtheme.
[DOCX File, 19 KB - humanfactors_v13i1e85518_app1.docx]

References

1. Barrera A, Gee C, Wood A, Gibson O, Bayley D, Geddes J. Introducing artificial intelligence in acute psychiatric inpatient care: qualitative study of its use to conduct nursing observations. Evid Based Ment Health 2020 Feb;23(1):34-38. [doi: 10.1136/ebmental-2019-300136] [Medline: 32046991]
2. Borges PVK, Nourani-Vatani N. Vision-based detection of unusual patient activity. Stud Health Technol Inform 2011;168(16-23):16-23. [Medline: 21893907]

XSL•FO
RenderX

3. Dewa LH, Broyd J, Hira R, et al. A service evaluation of passive remote monitoring technology for patients in a high-secure forensic psychiatric hospital: a qualitative study. BMC Psychiatry 2023 Dec 14;23(1):946. [doi: 10.1186/s12888-023-05437-w] [Medline: 38098066]
4. Socio-technical system. Oxford Reference. URL: https://www.oxfordreference.com/view/10.1093/oi/authority.20110803100515814 [accessed 2026-06-22]
5. Baxter G, Sommerville I. Socio-technical systems: from design methods to systems engineering. Interact Comput 2011 Jan;23(1):4-17. [doi: 10.1016/j.intcom.2010.07.003]
6. Batty D. NHS trusts criticised over system that films mental health patients in their bedrooms. The Guardian. 2021. URL: https://www.theguardian.com/society/2021/dec/13/nhs-trusts-urged-to-ditch-oxevision-system-covert-surveillance-mental-health-patients [accessed 2022-03-21]
7. Powell R, Porter H, Head N, Thomas E, Mariette S. Towards a Mad technoscience: revealing, repurposing, and reimagining carceral mental health technologies. Justice Power Resist ;9(1):27-44. [doi: 10.1332/26352338Y2025D000000045]
8. Lloyd-Jukes H, Gibson OJ, Wrench T, Odunlade A, Tarassenko L. Vision-based patient monitoring and management in mental health settings. J Clin Eng 2021;46(1):36-43. [doi: 10.1097/JCE.0000000000000447]
9. Clark H, Edwards A, Davies R, et al. Non-contact physical health monitoring in mental health seclusion. J Psych Intensive Care 2022 Apr 1;18(1):31-37. [doi: 10.20299/jpi.2021.009]
10. Buckley C, Malcolm R, Hanlon J. Economic impact of a vision-based patient monitoring system across five NHS mental health trusts. PLoS Digit Health 2024 Sep;3(9):e0000559. [doi: 10.1371/journal.pdig.0000559] [Medline: 39259712]
11. Malcolm R, Shore J, Stainthorpe A, Ndebele F, Wright K. Economic evaluation of a vision-based patient monitoring and management system in an acute adult and an older adult mental health hospital in England. J Med Econ 2022;25(1):1207-1217. [doi: 10.1080/13696998.2022.2147753] [Medline: 36377376]
12. Ndebele F, Wright K, Gandhi V, Bayley D. Non-contact health monitoring to support care in a psychiatric intensive care unit. J Psych Intensive Care 2022 Sep 1;18(2):95-100. [doi: 10.20299/jpi.2022.008]
13. Ndebele F, Wright K, Gandhi V, Bayley D. Non-contact monitoring to support care in acute inpatient mental health. J Ment Health 2024 Jun;33(3):320-325. [doi: 10.1080/09638237.2023.2245882] [Medline: 37589442]
14. Griffiths JL, Saunders KRK, Foye U, et al. The use and impact of surveillance-based technology initiatives in inpatient and acute mental health settings: a systematic review. BMC Med 2024 Nov 29;22(1):564. [doi: 10.1186/s12916-024-03673-9] [Medline: 39614242]
15. Stop Oxevision. URL: https://stopoxevision.wordpress.com/ [accessed 2026-06-26]
16. Batty D. NHS trust to stop filming mental health patients in their bedrooms. The Guardian. 2022. URL: https://www.theguardian.com/society/2022/mar/14/nhs-trust-to-stop-filming-mental-health-patients-in-their-bedrooms [accessed 2026-06-20]
17. Gooding PM, Clifford DM. Semi-automated care: video-algorithmic patient monitoring and surveillance in care settings. J Bioeth Inq 2021 Dec;18(4):541-546. [doi: 10.1007/s11673-021-10139-7] [Medline: 34787791]
18. Hendry J. 'Botched' aged care AI camera trial generates 12,000 false alerts. InnovationAus.com. 2022 Sep. URL: https://www.innovationaus.com/botched-aged-care-ai-camera-trial-generates-12000-false-alerts/ [accessed 2026-06-22]
19. LIO. Alpha global partners with LIO to redefine mental health care in Australia. PR Newswire. 2025. URL: https://www.prnewswire.com/news-releases/alpha-global-partners-with-lio-to-redefine-mental-health-care-in-australia-302529071.html [accessed 2025-10-09]
20. van der Veer SN, Riste L, Cheraghi-Sohi S, et al. Trading off accuracy and explainability in AI decision-making: findings from 2 citizens' juries. J Am Med Inform Assoc 2021 Sep 18;28(10):2128-2138. [doi: 10.1093/jamia/ocab127] [Medline: 34333646]
21. Carter SM, Aquino YSJ, Carolan L, et al. How should artificial intelligence be used in Australian health care? Recommendations from a citizens' jury. Med J Aust 2024 May 6;220(8):409-416. [doi: 10.5694/mja2.52283] [Medline: 38629188]
22. Oortwijn W, Husereau D, Abelson J, et al. Designing and implementing deliberative processes for health technology assessment: a good practices report of a Joint HTAi/ISPOR Task Force. Value Health 2022 Jun;25(6):869-886. [doi: 10.1016/j.jval.2022.03.018] [Medline: 35667778]
23. Street J, Duszynski K, Krawczyk S, Braunack-Mayer A. The use of citizens' juries in health policy decision-making: a systematic review. Soc Sci Med 2014 May;109:1-9. [doi: 10.1016/j.socscimed.2014.03.005] [Medline: 24657639]
24. Oliver P. Purposive sampling. In: Jupp V, editor. The SAGE Dictionary of Social Research Methods: SAGE Publications:244-245. [doi: 10.4135/9780857020116]
25. Daya I, Hamilton B, Roper C. Authentic engagement: a conceptual model for welcoming diverse and challenging consumer and survivor views in mental health research, policy, and practice. Int J Ment Health Nurs 2020 Apr;29(2):299-311. [doi: 10.1111/inm.12653] [Medline: 31538723]
26. Sofaer S. Qualitative methods: what are they and why use them? Health Serv Res 1999 Dec;34(5 Pt 2):1101-1118. [Medline: 10591275]
27. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

XSL•FO
RenderX

28. Sinfield G, Goldspink S, Wilson C. Waiting in the wings: the enactment of a descriptive phenomenology study. Int J Qual Methods 2023 Oct;22:16094069231207012. [doi: 10.1177/16094069231207012]
29. Graebner ME, Martin JA, Roundy PT. Qualitative data: cooking without a recipe. Strateg Organ 2012 Aug;10(3):276-284. [doi: 10.1177/1476127012452821]
30. Braun V, Clarke V. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. Qual Res Sport Exerc Health 2021 Mar 4;13(2):201-216. [doi: 10.1080/2159676X.2019.1704846]
31. Althobaiti K. Surveillance in next-generation personalized healthcare: science and ethics of data analytics in healthcare. New Bioeth 2021 Dec;27(4):295-319. [doi: 10.1080/20502877.2021.1993055] [Medline: 34720071]
32. Seclusion and restraint. Australian Institute of Health and Welfare. 2021. URL: https://www.aihw.gov.au/mental-health/topic-areas/restrictive-practices [accessed 2026-06-20]
33. Our position on Oxevision, the new monitoring system in mental health units. Rethink Mental Illness. 2023. URL: https://www.rethink.org/news-and-stories/media-centre/2023/11/our-position-on-oxevision-the-new-monitoring-system-in-mental-health-units/ [accessed 2025-10-09]
34. McCradden MD, Kirsch RE. Patient wisdom should be incorporated into health AI to avoid algorithmic paternalism. Nat Med 2023 Apr;29(4):765-766. [doi: 10.1038/s41591-023-02224-8] [Medline: 36823303]
35. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep 1;13(3):319-340. [doi: 10.2307/249008]
36. Barrett AK. Technological appropriations as workarounds: integrating electronic health records and adaptive structuration theory research. Inf Technol People 2018;31(2):368-387. [doi: 10.1108/ITP-01-2016-0023]
37. Boetto H. Artificial intelligence in social work: an EPIC model for practice. Aust Soc Work 2025:1-14. [doi: 10.1080/0312407X.2025.2488345]
38. Nissenbaum H. Privacy in Context: Technology, Policy, and the Integrity of Social Life: Stanford University Press; 2009.
39. Nissenbaum H. Contextual integrity up and down the data food chain. Theor Inq Law 2019 Mar 16;20(1):221-256. [doi: 10.1515/til-2019-0008]
40. Saveljeva J, Volkova T. A survey on digital trust: towards a validated definition. Digital 2025;5(2):14. [doi: 10.3390/digital5020014]
41. Appelbaum PS, Robbins PC, Roth LH. Dimensional approach to delusions: comparison across types and diagnoses. Am J Psychiatry 1999 Dec;156(12):1938-1943. [doi: 10.1176/ajp.156.12.1938] [Medline: 10588408]
42. Freeman D, Garety P. Advances in understanding and treating persecutory delusions: a review. Soc Psychiatry Psychiatr Epidemiol 2014 Aug;49(8):1179-1189. [doi: 10.1007/s00127-014-0928-7] [Medline: 25005465]
43. Kekic M, Rose A, Baker C, Bayley D. RETRACTED: Reduced self-harm on acute mental health wards following the implementation of a vision-based patient monitoring system: Evidence from five NHS trusts. J Psychiatr Ment Health Nurs 2024 Aug;31(4):e1-e9. [doi: 10.1111/jpm.13036] [Medline: 38441289]
44. National Safety and Quality Digital Mental Health Standards. Australian Commission on Safety and Quality in Health Care. 2022. URL: https://www.safetyandquality.gov.au/national-standards/national-safety-and-quality-digital-mental-health-standards [accessed 2026-06-26]
45. Muchamore I, Karanikolas P, Gooding P. How lived experience expertise shapes research and development in digital mental health: a review of literature and insights. : Wellcome; 2024 URL: https://apo.org.au/node/328006 [accessed 2024-11-04]
46. Gilbert K, Muchamore I, Katterl S, et al. Digital futures in mind: why lived experience collaboration must guide digital mental health technologies. Aust J Social Issues 2025 Mar;60(1):196-215. [doi: 10.1002/ajs4.355]
47. Patrickson B, Musker M, Thorpe D, van Kasteren Y, Bidargaddi N, The Consumer and Carer Advisory Group (CCAG). In-depth co-design of mental health monitoring technologies by people with lived experience. Future Internet 2023;15(6):191. [doi: 10.3390/fi15060191]

Abbreviations

AI: artificial intelligence
NHS: National Health Service
VAPM: video-algorithmic patient monitoring

XSL•FO
RenderX

Edited by A Kushniruk; submitted 08.Oct.2025; peer-reviewed by C Asuzu, N Jelani; revised version received 13.Mar.2026; accepted 19.Mar.2026; published 20.Jul.2026.

Please cite as:
Gooding P, Kennedy H, D'Alfonso S, Kariotis T, Daniel C, Hamilton B
Video-Algorithmic Patient Monitoring in Mental Health Inpatient Settings: Qualitative Study of Patient or Consumer, Clinician, and Vendor Perspectives
JMIR Hum Factors 2026;13:e85518
URL: https://humanfactors.jmir.org/2026/1/e85518
doi:10.2196/85518

© Piers Gooding, Hamilton Kennedy, Simon D'Alfonso, Timothy Kariotis, Catherine Daniel, Bridget Hamilton. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 20.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Feasibility of a Mobile-Based Home Monitoring System for Patients With Heart Failure: Mixed Methods Pilot Study

Finna E Indriany[1], MD, PhD; Kemal N Siregar[1], Prof Dr, PhD; Bambang Budi Siswanto[2], Prof Dr, SpJP; Budhi Setianto Purwowiyoto[2], Prof Dr, SpJP; Rarsari Soerarso[2], PhD, SpJP; Indrajani Sutedja[3], PhD; Benny Setiawan[4]; Hendy R Wijaya[5], PhD; Raymond Ching Chiew Wong[6], Prof Dr Med, PhD; Salvatore Di Somma[7], Prof Dr Med, PhD; Peter MacDonald[8], Prof Dr Med, PhD

[1]Faculty of Public Health, Universitas Indonesia, Jalan Lingkar Kampus UI, Depok, Indonesia

[2]Cardiology and Vascular Medicine Department, Faculty of Medicine, Universitas Indonesia, Jakarta, Indonesia

[3]Information Systems Department, School of Information Systems, Bina Nusantara University, Jakarta, Indonesia

[4]Pharmacy Study Program, Faculty of Health Sciences, Pelita Harapan University, Tangerang, Indonesia

[5]Research and Technology Transfer Office, Bina Nusantara University, Jakarta, Indonesia

[6]Department of Cardiology, National University Heart Center, Singapore, Singapore

[7]Great Health Science, AISC Italian Association of Patients with Heart Failure, Rome, Italy

[8]St Vincent's Clinical School, Faculty of Medicine, University of New South Wales, Sydney, Australia

Corresponding Author:
Finna E Indriany, MD, PhD
Faculty of Public Health, Universitas Indonesia, Jalan Lingkar Kampus UI, Depok, Indonesia

Abstract

Background: The risk of rehospitalization in patients with heart failure (HF) has initiated various efforts to prevent and simultaneously improve quality of life. Self-monitoring at home is one option, and technology is increasingly being used for this purpose.

Objective: This pilot study aimed to evaluate the feasibility and preliminary effects of a digital home monitoring intervention on patient-reported outcomes and 30-day readmissions among patients with HF in Indonesia.

Methods: A mixed methods pilot study was conducted, combining qualitative system development and quantitative evaluation. Patients were assigned to an intervention group (digital monitoring) or control group (standard care). Readmission rates were compared using chi-square tests and odds ratios. Changes in Kansas City Cardiomyopathy Questionnaire scores were analyzed using linear mixed-effects models.

Results: A total of 60 patients were included (n=30, 50% in the intervention group; n=30, 50% in the control group). Readmission occurred in 20% (6/30) of patients in the intervention group and 43.3% (13/30) of patients in the control group (odds ratio 0.33, 95% CI 0.10 - 1.09; *P*=.10). Linear mixed-effects analysis showed greater improvement in Kansas City Cardiomyopathy Questionnaire overall summary score in the intervention group (*P*=.02). Improvements were observed in the physical limitation, symptom frequency, symptom burden, quality of life, and social limitation domains. During follow-up, 3.3% (1/30) of the patients in the intervention group died of non–HF-related causes, and 10% (3/30) of the patients in the control group died due to HF.

Conclusions: This pilot study suggests that digital home monitoring is feasible and associated with improvements in patient-reported outcomes, with a potential signal toward reduced readmission. Larger studies are needed to confirm effectiveness.

(JMIR Hum Factors 2026;13:e84309) doi:10.2196/84309

KEYWORDS

heart failure; digital health; remote monitoring; Kansas City Cardiomyopathy Questionnaire; pilot study

Introduction

Heart failure (HF) is associated with high morbidity, mortality, and frequent hospital readmissions [1]. It affects over 20 million people worldwide, with a 5-year mortality rate of 60% to 70% [2]. It is a major public health problem, particularly among older adults, with a prevalence of 1% to 2% in low-income countries [3]. Reducing readmissions remains a key priority in HF management [4,5]. Digital health interventions, including home monitoring systems, have been increasingly used to support postdischarge management [6,7]. These systems enable early detection of symptom deterioration and may improve patient self-management. However, most evidence comes from high-income settings, with limited data from resource-limited

XSL•FO
RenderX

environments such as Indonesia [8,9]. This study aimed to evaluate the feasibility of a digital home monitoring intervention and explore preliminary effects on patient-reported outcomes and readmissions in a real-world Indonesian setting.

In many health care settings, postdischarge monitoring of patients with HF still relies on periodic outpatient visits and manual patient reporting [10]. This approach limits the ability of health care providers to detect early clinical deterioration. Consequently, innovative strategies are needed to support continuous monitoring and improve patient engagement in self-management [11,12].

Digital health technologies, particularly mobile health apps, have emerged as promising tools to support chronic disease management. Mobile-based monitoring systems enable patients to record symptoms, physiological parameters, and medication adherence while allowing health care providers to review patient-reported data remotely. Several studies have suggested that remote monitoring may improve patient engagement, facilitate early detection of clinical deterioration, and potentially reduce hospital readmissions [13]. However, the implementation of digital monitoring systems in real-world clinical settings remains challenging, particularly in terms of patient adherence, usability, and integration into existing health care workflows.

In Indonesia, the use of mobile-based monitoring apps for HF management is still limited. Understanding feasibility and patient adherence to such systems is essential before larger-scale implementation can be considered [14]. Therefore, this study developed a mobile-based monitoring app called FineHeart to support home monitoring and communication between patients with HF and health care providers.

This study aimed to evaluate the feasibility, patient adherence, and preliminary clinical outcomes associated with the use of the FineHeart mobile app among patients with HF following hospital discharge.

Methods

Study Design

This was a mixed methods pilot study consisting of qualitative needs assessment and system development combined with quantitative evaluation of feasibility and preliminary outcomes. This study was conducted from January 1, 2024, to November 30, 2024, at the National Cardiovascular Center (NCVC) Harapan Kita, Jakarta, Indonesia.

Participants and Setting

A qualitative needs assessment was conducted through interviews with hospital staff and document studies at the NCVC Harapan Kita. For the quantitative study, patients diagnosed with HF were recruited from a clinical setting. Eligibility criteria were applied consistently across groups. Participants were assigned to intervention or control groups based on predefined operational criteria (nonrandomized allocation). Baseline differences between groups were assessed and are acknowledged as potential confounders.

Intervention

The digital home monitoring system, named FineHeart, included daily symptom reporting, automated alerts for clinical deterioration, and remote clinician monitoring. The detailed system architecture and workflow are outlined in Multimedia Appendices 1 and 2, respectively. The development of the prototype was also based on the results of previous research using a machine learning method that developed a prediction model for HF readmissions at the same site implemented in a prototype mobile app [15].

Data Collection

The user acceptance test (UAT) was administered to both health workers and patients. The quantitative study involved posthospitalization patients with an I50 primary diagnosis code (*International Classification of Diseases, 10th Revision*). In the intervention group, participants filled out the Kansas City Cardiomyopathy Questionnaire (KCCQ) and were instructed on the use of the app prototype. A prototype link was given to patients and their families or caregivers, equipped with user-friendly instruments (digital sphygmomanometers, digital body scales, and digital oximeters), and were trained on how to use them. Patients filled out the app prototype daily for 30 days according to the parameters requested and could be monitored online by hospital nurses. If signs required a clinical response, indicated by the presence of green, yellow, and red notifications on the app, patients were followed up on by nurses or cardiologists. Readmission data were obtained from hospital records and verified during follow-up from patient reports. Readmission was defined as all-cause hospitalization within 30 days after index discharge. We used the 23 official and licensed items from the KCCQ from Outcomes Instruments, LLC, which has been translated into Indonesian and tested for reliability and validity, focusing on 4 domains that are closely related to public health: quality of life, self-efficacy, social limitation, and overall summary score [16,17]. Scores were obtained at baseline and 30 days. Missing outcome data were primarily due to incomplete reporting, loss to follow-up, or death during the study period. Adherence was defined as the proportion of days during which patients completed the required daily monitoring entries. After 30 days, participants were asked to re–fill out the KCCQ. In the control group, participants filled out the KCCQ before discharge and 30 days after discharge. No monitoring was carried out at home; only the follow-up schedule at the hospital was followed as usual.

Data Analysis

Qualitative data were analyzed using thematic analysis [18]. Coding was performed iteratively, and themes were derived through consensus among the researchers. For statistical analysis, baseline characteristics were summarized descriptively. Readmission rates were compared using chi-square tests, and odds ratios (ORs) with 95% CIs were calculated. Changes in KCCQ scores were analyzed using linear mixed-effects models that included all available data without imputation. Due to the pilot nature and limited sample size, multivariable adjustment was not performed to avoid overfitting. A P value below .05 was considered statistically significant.

XSL•FO
RenderX

Ethical Considerations

Ethics approval was granted by the Research and Community Service Ethics Committee, Faculty of Public Health, Universitas Indonesia (Ket-718/UN2.F10.D11/PPM.00.02/2024), and by the institutional review board of the NCVC Harapan Kita (DP.04.03/KEP003/EC001/2024). All participants provided written informed consent.

Results

Qualitative Results

A total of 3 cardiologists and 7 nurses participated in the qualitative portion of this study. This qualitative analysis identified 7 major themes reflecting the needs of health care professionals regarding the development of a home-based monitoring system for patients with HF. They can be grouped into 3 overarching domains: digital transformation needs, service delivery optimization, and organizational readiness. The findings indicated that existing care was reactive rather than proactive, fragmented rather than integrated, and manual rather than digital. These limitations contributed to delayed detection of patient deterioration and inefficiencies in care delivery.

Participant Allocation

Patients who met the eligibility criteria were consecutively recruited during the study period (nonrandomized). A total of 60 patients with HF were enrolled in the study, of whom 30 (50%) were assigned to the intervention group using the FineHeart mobile app, whereas 30 (50%) were assigned to the control group receiving standard care.

During the 30-day follow-up period, 13.3% (4/30) of the patients in the intervention group and 20% (6/30) of the patients in the control group did not complete the quality of life assessment. Therefore, data from 86.7% (26/30) of the patients in the intervention group and 80% (24/30) of the patients in the control group were included in the final KCCQ analysis.

Baseline Characteristics

The baseline demographic and clinical characteristics of the participants in the quantitative portion of this study are shown in Table 1. Baseline differences were observed between groups in several variables, including educational level and clinical characteristics. Overall, the intervention and control groups showed comparable baseline characteristics in terms of age, gender distribution, and HF diagnosis. However, minor variations were observed in several clinical variables, which should be considered when interpreting the study findings.

XSL•FO
RenderX

Table . Baseline characteristics.

Variable	Intervention group (n=30)	Control group (n=30)	*P* value
Age (y), mean (SD)	69.0 (4.3)	68.2 (4.1)	.56
Sex (male), n (%)	21 (70)	18 (60)	.42
Marital status (married), n (%)	27 (90)	28 (93.3)	.63
Payer (government health insurance), n (%)	29 (96.7)	30 (100)	.32
Living with partner (yes), n (%)	30 (100)	30 (100)	>.99
Educational level, n (%)			.003
Elementary school	3 (10)	9 (30)	
Junior high school	7 (23.3)	0 (0)	
Senior high school	14 (46.7)	20 (66.7)	
University	6 (20)	1 (3.3)	
Occupation, n (%)			.07
Unemployed	16 (53.3)	24 (80)	
Self-employed	5 (16.7)	1 (3.3)	
Government or private sector employee	9 (30)	5 (16.7)	
Religion, n (%)			.36
Muslim	29 (96.7)	27 (90)	
Catholic	1 (3.3)	1 (3.3)	
Buddhist	0 (0)	2 (6.7)	
Outpatient compliance (regularly), n (%)	23 (76.7)	12 (40)	.004
HF[a] family history (yes), n (%)	2 (6.7)	4 (13.3)	.39
Smoking (yes), n (%)	7 (23.3)	9 (30)	.56
Alcohol consumption (yes), n (%)	1 (3.3)	2 (6.7)	.55
Substance abuse (yes), n (%)	1 (3.3)	0 (0)	.31
Anemia (yes), n (%)	2 (6.7)	5 (16.7)	.23
Aortic valve disorder (yes), n (%)	3 (10)	0 (0)	.08
Asthma (yes), n (%)	0 (0)	0 (0)	>.99
Cancer (yes), n (%)	0 (0)	0 (0)	>.99
Arrhythmia (yes), n (%)	4 (13.3)	4 (13.3)	>.99
Stroke (yes), n (%)	3 (10)	1 (3.3)	.30
Coronary artery disease (yes), n (%)	12 (40)	7 (23.3)	.17
Dementia (yes), n (%)	1 (3.3)	0 (0)	.31
Diabetes mellitus (yes), n (%)	11 (36.7)	17 (56.7)	.12
Dyslipidemia (yes), n (%)	7 (23.3)	8 (26.7)	.77
Hypertension (yes), n (%)	11 (36.7)	12 (40)	.79
Liver disease (yes), n (%)	0 (0)	0 (0)	>.99
Lung disease (yes), n (%)	6 (20)	2 (6.7)	.13
Protein caloric malnutrition (yes), n (%)	0 (0)	0 (0)	>.99
Psychiatric disorder (yes), n (%)	0 (0)	0 (0)	>.99
Renal disease (yes), n (%)	11 (36.7)	12 (40)	.79

XSL•FO
RenderX

Variable	Intervention group (n=30)	Control group (n=30)	*P* value
Rheumatic disease (yes), n (%)	4 (13.3)	5 (16.7)	.72
Thyroid disease (yes), n (%)	1 (3.3)	1 (3.3)	>.99
Cardiac devices (yes), n (%)	1 (3.3)	0 (0)	.31
Cardiac surgery (yes), n (%)	1 (3.3)	0 (0)	.31
Coronary angioplasty (yes), n (%)	4 (13.3)	3 (10)	.69
Mechanical ventilator (yes), n (%)	0 (0)	0 (0)	>.99
ACE[b] inhibitors (yes), n (%)	10 (33.3)	14 (46.7)	.29
ARBs[c] (yes), n (%)	5 (16.7)	8 (26.7)	.35
ARNIs[d] (yes), n (%)	11 (36.7)	4 (13.3)	.04
SGLT2[e] inhibitors (yes), n (%)	1 (3.3)	0 (0)	.31
Diuretics (yes), n (%)	23 (76.7)	29 (96.7)	.02
β-blockers (yes), n (%)	27 (90)	27 (90)	>.99
Spironolactone (yes), n (%)	20 (66.7)	24 (80)	.24
NYHA[f] classification, n (%)			.49
III	15 (50)	18 (60)	
IV	15 (50)	12 (40)	
LVEF[g], n (%)			.04
HFrEF[h]	25 (83.3)	17 (56.7)	
HFmEF[i]	1 (3.3)	7 (23.3)	
HFpEF[j]	4 (13.3)	6 (20)	

[a]HF: heart failure.
[b]ACE: angiotensin-converting enzyme inhibitor.
[c]ARB: angiotensin receptor blocker.
[d]ARNI: angiotensin receptor–neprilysin inhibitor.
[e]SGLT2: sodium-glucose cotransporter 2.
[f]NYHA: New York Heart Association.
[g]LVEF: left ventricular ejection fraction.
[h]HFrEF: HF with reduced ejection fraction.
[i]HFmEF: HF with midrange ejection fraction.
[j]HFpEF: HF with preserved ejection fraction.

Patient Adherence to the FineHeart App

Among patients assigned to the intervention group, adherence to the FineHeart app varied. Most participants were able to use the app regularly during the monitoring period, although several patients (4/30, 13.3%) demonstrated lower adherence due to difficulties in operating the app or limited caregiver support.

Readmission Outcome

At 30 days, readmission occurred in 20% (6/30) of the patients in the intervention group, which was lower than the 43.3% (13/30) in the control group. This difference was not statistically significant (*P*=.10), but the OR suggested lower odds of readmission in the intervention group (OR 0.33, 95% CI 0.10 - 1.09).

KCCQ Outcomes

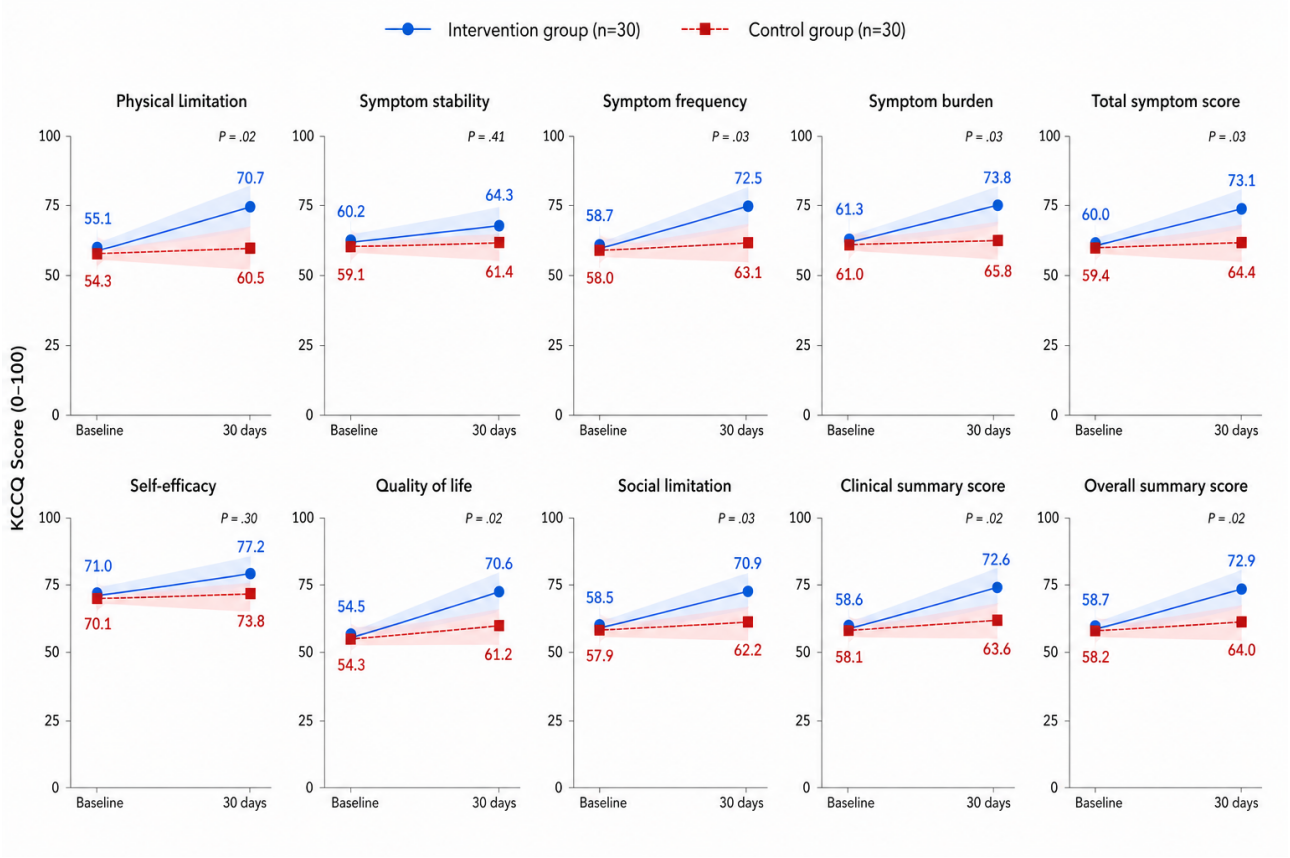
Linear mixed-effects analysis demonstrated a significant time-by-group interaction for overall summary score (*P*=.02), indicating greater improvement in the intervention group (Table 2). Improvements were observed in physical limitation, symptom frequency, symptom burden, quality of life, and social limitation. These changes exceeded clinically meaningful thresholds (>5 points). No significant differences were observed for symptom stability and self-efficacy (Figure 1).

XSL•FO
RenderX

Table . Changes in Kansas City Cardiomyopathy Questionnaire scores (linear mixed-effects model).

Domain	Change in the intervention group	Change in the control group	Interaction *P* value
Physical limitation	+15.6	+6.2	.02
Symptom stability	+4.1	+2.3	.41
Symptom frequency	+13.8	+5.1	.03
Symptom burden	+12.5	+4.8	.03
Total symptom score	+13.1	+5.0	.03
Self-efficacy	+6.2	+3.7	.30
Quality of life	+16.1	+6.9	.02
Social limitation	+12.4	+4.3	.03
Clinical summary	+14.0	+5.5	.02
Overall summary	+14.2	+5.8	.02

Figure 1. Improvement in Kansas City Cardiomyopathy Questionnaire scores from baseline to 30 days: intervention vs control group. Values shown are estimated marginal means from linear mixed-effects models with fixed effects for time, group, and time × group interaction and a random intercept for participant. Shaded areas represent 95% CIs. Higher scores indicate better health status. *P* values represent the time × group interaction.

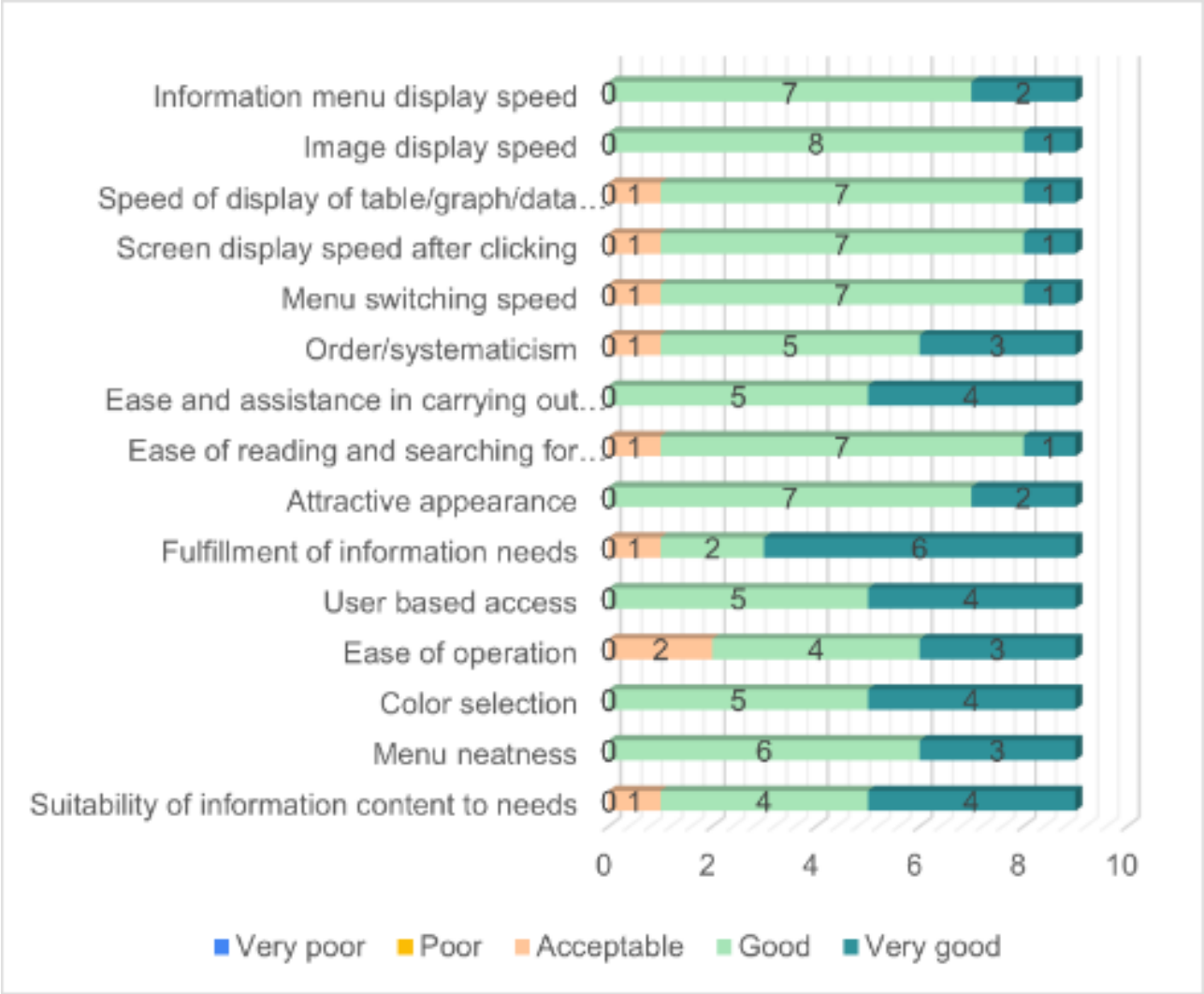


User Acceptance Test

The UAT results among 9 health workers showed that the app prototype was complete and could be used by health care workers and patients at home. The results of the UAT among 25 patients with HF showed that patient acceptance of the FineHeart app prototype was predominantly either good or very good (Figures 2 and 3).

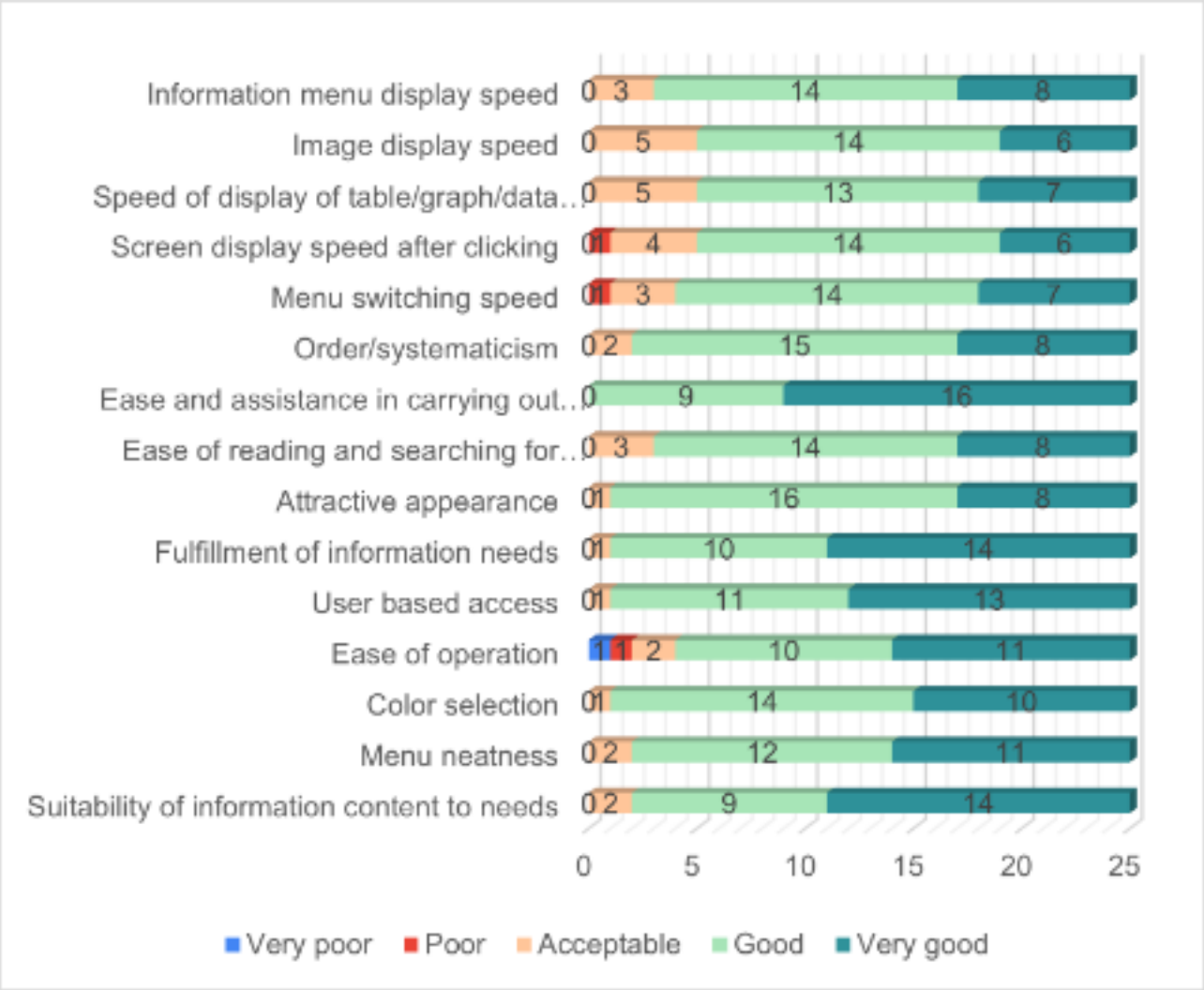
XSL•FO
RenderX

Figure 2. User acceptance test (UAT) results in health workers. Most of the 9 health workers who took part in the UAT provided good and very good ratings in all parameters.



XSL•FO
RenderX

Figure 3. User acceptance test (UAT) results in patients with heart failure. Most of the patients who took part in the UAT (90% on average) provided good and very good ratings in all parameters.



Missing Data and Mortality

During the 30-day follow-up, some KCCQ data were missing due to incomplete reporting or loss to follow-up. In addition, 3.3% (1/30) of the patients in the intervention group died of non–HF-related causes, and 10% (3/30) of the patients in the control group died due to HF.

Discussion

Principal Findings

This pilot study evaluated the feasibility of a digital home monitoring intervention for patients with HF in a resource-limited setting [19]. The findings suggested that the intervention was feasible to be implemented and associated with improvements in patient-reported outcomes, as reflected by significant changes in several KCCQ domains. In addition, a lower proportion of 30-day readmissions was observed in the intervention group than in the control group, although this difference did not reach statistical significance.

Importantly, the improvement in KCCQ overall and domain-specific scores indicated a positive effect on health-related quality of life [20]. These improvements were most pronounced in the physical limitation, symptom frequency, symptom burden, quality of life, and social limitation domains, suggesting that the intervention may support both symptom management and functional status. The observed improvements in KCCQ scores were clinically meaningful as changes exceeding 5 points are generally considered relevant for patients with HF [17]. The magnitude of change observed in this study suggests that patients in the intervention group experienced perceptible benefits in their daily functioning and symptom burden.

In contrast, while the readmission rate was lower in the intervention group (6/30, 20% vs 13/30, 43.3%), the difference was not statistically significant. This might be due to the limited sample size and the pilot nature of this study, which was not powered to detect differences in clinical outcomes.

Taken together, these findings should be interpreted as preliminary signals rather than definitive evidence of effectiveness. The results were hypothesis generating and provided a rationale for larger, adequately powered studies.

An important observation in this study was the difference in mortality between groups [21]. During the follow-up period, 10% (3/30) of the patients in the control group died due to HF,

XSL•FO
RenderX

whereas 3.3% (1/30) of the patients in the intervention group died from a non–HF-related cause. This imbalance might reflect differences in baseline clinical stability or disease severity. The higher rate of HF-related mortality in the control group might also partially explain the lower KCCQ scores and higher readmission rates observed in that group. Patients with more advanced diseases or greater instability are more likely to experience worse symptoms, reduced quality of life, and increased health care use.

The findings of this study were broadly consistent with those of previous research on digital and telemonitoring interventions in HF, which have demonstrated variable effects on clinical outcomes but more consistent improvements in patient-reported measures. Several studies have suggested that digital monitoring may enhance early detection of symptom deterioration, improve adherence, and support self-management [13,22]. However, the impact on readmission has been heterogeneous across studies, likely reflecting differences in intervention design, patient populations, health care systems, and implementation strategies [23].

Study Limitations and Future Research

This study had several limitations that should be considered when interpreting the findings.

First, the study involved a relatively small sample size and was conducted at a single center. As this was a pilot study, the sample size was intended to explore feasibility and preliminary outcomes rather than provide definitive evidence of effectiveness.

Second, this study used a quasi-experimental design without randomization. Although efforts were made to ensure comparability between groups, baseline differences may still have influenced the observed results.

Third, the duration of follow-up was limited to 30 days following hospital discharge. While this period allowed for initial evaluation of patient adherence and early outcomes, longer follow-up periods are necessary to better assess the long-term impact of digital monitoring on hospital readmission and quality of life.

Fourth, not all eligible patients were familiar with smartphone technology, and some declined participation due to concerns about using digital devices. This might limit the generalizability of the findings to populations with lower digital literacy.

Future studies should include larger multicenter cohorts and randomized study designs to better evaluate the clinical effectiveness of mobile-based monitoring systems. In addition, further research should explore strategies to improve patient engagement, the usability of digital platforms, and the integration of remote monitoring systems into routine clinical workflows.

Conclusions

This pilot study evaluated the feasibility and preliminary outcomes of the FineHeart mobile-based monitoring app for patients with HF following hospital discharge. The findings suggested that the app is feasible for supporting patient self-monitoring and engagement in postdischarge care. Patients who used the app showed favorable trends in several quality of life domains and demonstrated varying levels of adherence to digital monitoring. Although this study was not designed to establish definitive clinical effectiveness, the results provide preliminary signals that mobile-based monitoring may support patient self-management and potentially contribute to improved clinical outcomes. Larger studies are needed to confirm effectiveness.

Acknowledgments

The authors want to thank all the participants and the heart failure team of the National Cardiovascular Center Harapan Kita that were involved in this study for the facilities, data, and assistance provided. The authors declare the use of generative AI (GenAI) in the research and writing process. According to the Generative AI Delegation Taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: translation. The GenAI tool used was GPT-5.5 (OpenAI). Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Funding

This study was supported by Universitas Indonesia (grant NKB-354/UN2.RST/HKP.05.00/2024).

Data Availability

The datasets generated or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: FEI (lead), KNS (equal), BBS (equal), BSP (equal), RCCW (supporting), SDS (supporting), PM (supporting)
Data curation: FEI (lead), IS (equal), HRW (equal), BS (supporting)
Formal analysis: FEI (lead), IS (equal), HRW (supporting), BS (supporting)
Funding acquisition: KNS
Investigation: FEI (lead), RS (equal)
Methodology: FEI (lead), KNS (equal)

XSL•FO
RenderX

Project administration: FEI (lead), BBS (equal), RS (supporting)
Resources: FEI (lead), BS (supporting)
Supervision: KNS (lead), BBS (equal), BSP (equal), RS (equal)
Validation: KNS (lead), IS (equal)
Visualization: FEI (lead), IS (equal), HRW (supporting)
Writing—original draft: FEI (lead), KNS (supporting), RCCW (supporting), SDS (supporting), PM (supporting)
Writing—review & editing: FEI (lead), KNS (equal), RCCW (equal), SDS (equal), PM (equal)

Conflicts of Interest

None declared.

Multimedia Appendix 1
FineHeart app system architecture.
[PDF File, 69 KB - humanfactors_v13i1e84309_app1.pdf]

Multimedia Appendix 2
FineHeart app workflow diagram.
[PDF File, 83 KB - humanfactors_v13i1e84309_app2.pdf]

References

1. Rosano GM, Vitale C, Adamo M, Metra M. Roadmap for the management of heart failure patients during the vulnerable phase after heart failure hospitalizations: how to implement excellence in clinical practice. J Cardiovasc Med (Hagerstown) 2022;23(3):149-156. [doi: 10.2459/JCM.0000000000001221] [Medline: 34937849]
2. Hlaváček D, Haluzík M, Mahrík J, Popivnyak G, Kasperová BJ, Ivák P. Metabolic consequences of advanced chronic heart failure and its modification by implantation of a durable left ventricular assist device. Rev Cardiovasc Med 2024;25(11):388. [doi: 10.31083/j.rcm2511388] [Medline: 39618874]
3. Sopek Merkaš I, Slišković AM, Lakušić N. Current concept in the diagnosis, treatment and rehabilitation of patients with congestive heart failure. World J Cardiol 2021 Jul 26;13(7):183-203. [doi: 10.4330/wjc.v13.i7.183] [Medline: 34367503]
4. Hafkamp FJ, Tio RA, Otterspoor LC, et al. Optimal effectiveness of heart failure management - an umbrella review of meta-analyses examining the effectiveness of interventions to reduce (re)hospitalizations in heart failure. Heart Fail Rev 2022 Sep;27(5):1683-1748. [doi: 10.1007/s10741-021-10212-8] [Medline: 35239106]
5. Masip J, Frank Peacok W, Arrigo M, et al. Acute heart failure in the 2021 ESC Heart Failure Guidelines: a scientific statement from the Association for Acute CardioVascular Care (ACVC) of the European Society of Cardiology. Eur Heart J Acute Cardiovasc Care 2022 Feb 8;11(2):173-185. [doi: 10.1093/ehjacc/zuab122] [Medline: 35040931]
6. Davoudi M, Najafi Ghezeljeh T, Vakilian Aghouee F. Effect of a smartphone-based app on the quality of life of patients with heart failure: randomized controlled trial. JMIR Nurs 2020 Nov 2;3(1):e20747. [doi: 10.2196/20747] [Medline: 34406971]
7. Senanayake S, Kularatna S, Lee AS, et al. Health services costs of clinical heart failure with reduced ejection fraction in Singapore. Value Health Reg Issues 2025 Jan;45:101037. [doi: 10.1016/j.vhri.2024.101037] [Medline: 39226724]
8. Zakiyah N, Sinuraya RK, Kusuma AS, Suwantika AA, Lestari K. Cost-effectiveness analysis of sacubitril/valsartan compared to enalapril for heart failure patients in Indonesia. Clinicoecon Outcomes Res 2021;13:863-872. [doi: 10.2147/CEOR.S322740] [Medline: 34675566]
9. Hany A, Vatmasari RA. Correlation between nurse-patient interaction and readiness to care for post-treated heart failure patients in the intensive care room Malang, Indonesia. J Public Health Res 2021 Apr 14;10(2):2229. [doi: 10.4081/jphr.2021.2229] [Medline: 33855414]
10. Caggianelli G, Alivernini F, Chirico A, et al. The relationship between caregiver contribution to self-care and patient quality of life in heart failure: a longitudinal mediation analysis. PLoS One 2024;19(3):e0300101. [doi: 10.1371/journal.pone.0300101] [Medline: 38470867]
11. Lee KC, Breznen B, Ukhova A, Martin SS, Koehler F. Virtual healthcare solutions in heart failure: a literature review. Front Cardiovasc Med 2023;10:1231000. [doi: 10.3389/fcvm.2023.1231000] [Medline: 37745104]
12. Bohanec M, Tartarisco G, Marino F, et al. HeartMan DSS: a decision support system for self-management of congestive heart failure. Expert Syst Appl 2021;186. [doi: 10.1016/j.eswa.2021.115688]
13. Tan SY, Sumner J, Wang Y, Wenjun Yip A. A systematic review of the impacts of remote patient monitoring (RPM) interventions on safety, adherence, quality-of-life and cost-related outcomes. NPJ Digit Med 2024 Jul 18;7(1):192. [doi: 10.1038/s41746-024-01182-w] [Medline: 39025937]
14. Portz JD, Vehovec A, Dolansky MA, Levin JB, Bull S, Boxer R. The development and acceptability of a mobile application for tracking symptoms of heart failure among older adults. Telemed e-Health 2018;24(2):161-165. [doi: 10.1089/tmj.2017.0036]

XSL•FO
RenderX

15. Indriany FE, Siregar KN, Purwowiyoto BS, Siswanto BB, Sutedja I, Wijaya HR. Predicting the risk of severity and readmission in patients with heart failure in Indonesia: a machine learning approach. Healthc Inform Res 2024 Jul;30(3):253-265. [doi: 10.4258/hir.2024.30.3.253] [Medline: 39160784]
16. Spertus JA, Jones PG, Sandhu AT, Arnold SV. Interpreting the Kansas City Cardiomyopathy Questionnaire in clinical trials and clinical care: JACC state-of-the-art review. J Am Coll Cardiol 2020 Nov 17;76(20):2379-2390. [doi: 10.1016/j.jacc.2020.09.542] [Medline: 33183512]
17. Nassif M, Fine JT, Dolan C, et al. Validation of the Kansas City Cardiomyopathy Questionnaire in symptomatic obstructive hypertrophic cardiomyopathy. JACC Heart Fail 2022 Aug;10(8):531-539. [doi: 10.1016/j.jchf.2022.03.002] [Medline: 35902155]
18. Braun V, Clarke V. Supporting best practice in reflexive thematic analysis reporting in Palliative Medicine: a review of published research and introduction to the Reflexive Thematic Analysis Reporting Guidelines (RTARG). Palliat Med 2024 Jun;38(6):608-616. [doi: 10.1177/02692163241234800] [Medline: 38469804]
19. Bocchino M, Agazio E, Damiano C, et al. Telecardiology activities in hospital and university cardiology facilities in Italy: survey study. JMIR Cardio 2025 Dec 5;9:e73747. [doi: 10.2196/73747] [Medline: 41348919]
20. Sammour Y, Spertus JA, Austin BA, et al. Outpatient management of heart failure during the COVID-19 pandemic after adoption of a telehealth model. JACC Heart Fail 2021 Dec;9(12):916-924. [doi: 10.1016/j.jchf.2021.07.003] [Medline: 34857175]
21. Emmons-Bell S, Johnson C, Roth G. Prevalence, incidence and survival of heart failure: a systematic review. Heart 2022 Aug;108(17):1351-1360. [doi: 10.1136/heartjnl-2021-320131] [Medline: 35042750]
22. Choi EY, Park JS, Min D, Ahn S, Ahn JA. Heart Failure-Smart Life: a randomized controlled trial of a mobile app for self-management in patients with heart failure. BMC Cardiovasc Disord 2023 Jan 9;23(1):7. [doi: 10.1186/s12872-023-03039-8] [Medline: 36624373]
23. Johnson AE, Bilderback AL, Boninger M, et al. Remote monitoring for heart failure: assessing the risks of readmission and mortality. Am Heart J Plus 2021;10:100045. [doi: 10.1016/j.ahjo.2021.100045] [Medline: 38550399]

Abbreviations

HF: heart failure
KCCQ: Kansas City Cardiomyopathy Questionnaire
NCVC: National Cardiovascular Center
OR: odds ratio
UAT: user acceptance test

Edited by A Kushniruk; submitted 28.Sep.2025; peer-reviewed by AM Ibrahim, KH Gjeilo; revised version received 30.May.2026; accepted 23.Jun.2026; published 29.Jul.2026.

Please cite as:
Indriany FE, Siregar KN, Siswanto BB, Purwowiyoto BS, Soerarso R, Sutedja I, Setiawan B, Wijaya HR, Wong RCC, Di Somma S, MacDonald P
Feasibility of a Mobile-Based Home Monitoring System for Patients With Heart Failure: Mixed Methods Pilot Study
JMIR Hum Factors 2026;13:e84309
URL: https://humanfactors.jmir.org/2026/1/e84309
doi:10.2196/84309

© Finna E Indriany, Kemal N Siregar, Bambang Budi Siswanto, Budhi Setianto Purwowiyoto, Rarsari Soerarso, Indrajani Sutedja, Benny Setiawan, Hendy R Wijaya, Raymond Ching Chiew Wong, Salvatore Di Somma, Peter MacDonald. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 29.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Using Theory-Based Frameworks to Identify Barriers, Facilitators, Expectations, and Willingness to Pay for Online Medical Consultation Services in China: Qualitative Study

Shujie Dong[1,2*], PhD; Qiushi Cai[1,3*], MRes; Jingyi Ye[4], MPH; Yanya Chen[5], PhD; Rongsheng Zhao[1], PhD; Wai-kit Ming[2], MPH, MMSc, MD, PhD

[1]Department of Pharmacy, Peking University Third Hospital, Beijing, China

[2]Department of Infectious Diseases and Public Health, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, To Yuen Building, 31 To Yuen Street, Hong Kong, China (Hong Kong)

[3]Department of Metabolism, Digestion and Regeneration, School of Medicine, Imperial College London, London, England, United Kingdom

[4]School of Public Health, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong, China (Hong Kong)

[5]School of Nursing, Jinan University, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Wai-kit Ming, MPH, MMSc, MD, PhD
Department of Infectious Diseases and Public Health, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, To Yuen Building, 31 To Yuen Street, Hong Kong, China (Hong Kong)

Abstract

Background: Online medical consultation (OMC) services have gained considerable attention as an integral component of telemedicine. Recently, AI has been increasingly integrated into OMC platforms, facilitating more efficient consultations and clinical decision-making. AI-driven OMC services can provide preliminary triage, medication guidance, and diagnostics for multiple medical conditions. Despite the availability and potential benefits of AI-driven OMC services, public acceptance and willingness to pay (WTP) for these services remain low.

Objective: This study aimed to identify perceived barriers, facilitators, expectations, and factors shaping public acceptance of and stated WTP for AI-driven OMC services.

Methods: We conducted semistructured qualitative interviews with patients, caregivers, and health care professionals to explore barriers, facilitators, expectations, and factors shaping public acceptance of and stated WTP for AI-driven OMC services. The study was informed by the theories of perceived risk and perceived benefit, which guided the development of the interview guide. All interviews were audio-recorded and transcribed verbatim. Data were analyzed using NVivo (version 15; Lumivero) with deductive thematic analysis guided by these theories. Coding was conducted independently and cross-checked by 2 researchers to ensure credibility and consistency.

Results: Thematic analysis of 20 in-depth interviews identified 2 main themes and 11 subthemes. Perceived risks and perceived benefits emerged as 2 key perspectives influencing participants' acceptance and WTP. Psychological, governance, social, functional, health, and financial risks reduced acceptance, whereas convenience, diversity, reliability, efficiency, and educational benefits promoted it. Participants' self-reported WTP ranged from RMB 0 to RMB 200 (US $0-$27.28; RMB 1=US $0.1364 as of January 15, 2025), with participants who had prior experience with OMC generally reporting higher values than those without prior OMC experience.

Conclusions: This study identified facilitators and barriers influencing public acceptance of and WTP for AI-driven OMC services using theoretical constructs. Our findings offer valuable insights into the development and refinement of AI-driven OMC services, enabling more targeted pricing strategies and tailored services that address public preferences and concerns, as well as supporting the development of standardized regulatory governance for digital medical consultation platforms.

(JMIR Hum Factors 2026;13:e96448) doi:10.2196/96448

KEYWORDS

online medical consultation; public perception; willingness to pay; qualitative research; facilitators; barriers; in-depth interview; AI

XSL•FO
RenderX

Introduction

Telemedicine is the delivery of health care services in which distance is a critical factor and health care professionals use information and communication technologies to exchange valid information for the diagnosis, treatment, and prevention of diseases and injuries, as well as for research, evaluation, and education [1]. Online medical consultation (OMC) is an important element of telemedicine and has become increasingly popular over the past few years, especially since the COVID-19 pandemic [2]. Recently, the use of AI in health care has grown rapidly, particularly in the field of telemedicine. Integrating AI into telemedicine has transformed a wide array of medical practices, from diagnostics to medical consultation [3]. The integration of AI into OMC services enables real-time data analysis and decision support, thereby facilitating remote consultations with greater efficiency and accuracy [4]. AI-driven OMC services have been demonstrated to be an effective means of addressing consultations for various medical conditions, such as surgery [5], cancer [6], emergency care [7], psychiatry [8], and cardiovascular diseases [9].

Despite the availability and potential benefits of AI-driven OMC services, their uptake among the public remains low [10]. This gap in adoption highlights the need to understand the multifaceted factors influencing their use [11,12]. Previous studies have predominantly focused on facilitators and barriers associated with OMC services without AI integration. For example, multiple factors have been identified as facilitators of OMC services, including cost savings, reduced waiting and travel times, ease of use, familiarity with the system, trust in technology, family involvement, and patient age [11]. Several factors have also been identified as barriers to OMC services, including slow internet speed, poor audio or video quality, limited internet access, weak signal coverage, difficulty expressing emotions, lack of body language, and user resistance [11]. However, evidence regarding public perceptions of acceptance and willingness to pay (WTP) for AI-driven OMC services remains limited. Although the integration of AI into telemedicine can reshape OMC services, promoting adoption requires addressing both the perceived benefits and risks of these services from users' perspectives.

The theory of perceived risk suggests that consumers perceive risk due to uncertainty and potentially undesirable consequences associated with paying for or using services [13,14]. It has been widely applied in previous studies to explore public perceptions of health care services, such as mobile health (mHealth) technologies [15], e-consultation services [16], online pharmacy platforms [17], and AI-enabled health care apps [18]. Meanwhile, the theory of perceived benefit emphasizes consumers' beliefs about the extent to which they will gain positive outcomes from using a specific service [19]. Perceived benefits tend to enhance the intention to use and are associated with health care consumers' greater willingness to accept the vulnerability (ie, service uncertainty and potential negative consequences) resulting from such uncertainty [13,20]. Understanding perceived risks and benefits is crucial for designing health care services that meet consumer expectations, especially in the context of the rapid development of digital health services and AI technologies. Therefore, the theories of perceived risk and perceived benefit can serve as useful frameworks to evaluate the willingness to accept and pay for AI-driven OMC services.

This study used a qualitative approach informed by the theories of perceived risk and perceived benefit to identify perceived barriers, facilitators, expectations, and factors shaping public acceptance of and stated WTP for AI-driven OMC services. By doing so, it seeks to shed light on the development and refinement of AI-driven OMC services, including targeted pricing strategies and tailored service designs to address patient needs.

Methods

Ethical Considerations

This study was approved by the Ethics Committee of Peking University Third Hospital (approval number M20241044). All participants provided informed consent prior to participation. All information collected during the study was treated as strictly confidential. Interview data were anonymized before analysis and stored in NVivo (version 15; Lumivero) on secure, password-protected institutional servers, with access restricted to authorized members of the research team.

Study Design

In this study, AI-driven OMC services refer to OMC services in which AI technologies are used to support medication-related information provision, preliminary guidance, triage, or decision support. These services may include fully automated AI responses as well as AI-assisted models involving review by health care professionals. As this remains an emerging service model, the study explored participants' perceptions of this broader service category rather than evaluating a single standardized platform. This study adopted a descriptive qualitative research design to comprehensively explore the factors influencing patients' acceptance of and WTP for AI-driven OMC services. In-depth semistructured interviews were conducted to collect detailed information from participants.

Setting and Participants

Purposive sampling was used to recruit participants with diverse characteristics, including sex, age, geographic location, identity characteristics, prior OMC experience, and educational level. The research team approached potential participants from different cities in China through clinical, professional, and personal networks. Some participants also referred other eligible individuals to the research team. Recruitment continued until the sample included patients, caregivers, and health care professionals with varied backgrounds and experiences related to OMC services.

Eligible participants were individuals aged 18 years or older who were able to provide informed consent and participate in a Chinese-language interview. Patients, caregivers, and health care professionals were included to capture diverse perspectives on the factors influencing acceptance of and WTP for AI-driven OMC services. Participants were excluded if they were unable

XSL•FO
RenderX

to provide informed consent, unable to complete the interview, or unwilling to be audio-recorded.

In-depth semistructured interviews were conducted in Chinese either face-to-face or remotely via telephone in a quiet and private setting between December 2024 and January 2025. The sample size was determined based on data saturation, with interviews discontinued when no new information emerged [21].

Data Collection

Before the interviews, participants were informed of the study's purpose and provided written informed consent. Each interview lasted for 30-45 minutes and was audio-recorded with AI-assisted real-time transcription using a speech-to-text device. All transcripts were checked against the original audio recordings and verified against the recordings within 24 hours to ensure transcription accuracy. All interviews were led by QC, while SD served as the note-taker. Both researchers had backgrounds in pharmacy and qualitative research and were familiarized with the interview guide before data collection. A semistructured interview outline designed based on relevant literature and participant characteristics [22] was pilot tested with 3 volunteers and refined before the formal interviews. The outline consisted of 5 primary questions and was detailed in Textbox 1. Participants were not shown a standardized platform, workflow, use case, or pricing scenario before the interviews. Instead, the interviews used open-ended questions to explore participants' perceptions, concerns, expectations, and WTP regarding AI-driven OMC services as a broad and emerging service model. Interviews began with open-ended prompts (eg, "what" and "how") to elicit detailed responses, followed by probing questions for deeper exploration of emerging themes [23,24]. Plain and accessible language was used to encourage open communication. SD documented contextual information to complement the audio-recorded data, including facial expressions and body language in face-to-face interviews and tone, pauses, and emotional reactions in telephone interviews. Participants could withdraw at any time or refuse to answer any question.

Textbox 1. Semistructured interview outline questions for participants.

This semistructured interview outline was developed based on the research objectives and core research questions. Relevant interview questions are listed below:

1. Could you briefly share your personal information, including your age, place of residence, family situation, health condition, and internet proficiency?
2. Have you ever used any online medical consultation (OMC) service, particularly one driven by AI technologies? If yes, when would you choose to use such a service? If not, why?
3. How do you feel about AI-driven OMC services? What are their strengths, weaknesses, and key aspects you care about?
4. Would you be willing to use or pay for such services, and what factors might encourage or discourage your willingness to use or pay for them?
5. What are your expectations for AI-driven OMC services, and what areas do you think could be improved or optimized in the future?

Data Coding and Analysis

The data analysis process was conducted concurrently with data collection, and further sampling decisions and interview questions were guided by the findings of the ongoing data analysis [24]. Interviews were conducted until ongoing coding indicated that no new codes or subthemes relevant to the study objectives emerged [21].

Transcripts were sequentially numbered and imported into NVivo (version 15) for coding. To minimize potential bias and enhance the credibility and consistency of the analysis, 2 researchers independently coded and analyzed the data, followed by cross-checking and discussion to resolve discrepancies. An initial codebook was developed through comparison of independently generated codes and was refined iteratively during the analysis.

The analysis followed Clarke and Braun's 6-phase thematic analysis approach [25,26]. First, 2 researchers (SD and QC) independently read the transcripts to become familiar with the data. Second, each researcher independently generated initial codes across the dataset, focusing on participants' perceptions of AI-driven OMC services without applying a fixed coding template. Third, the researchers independently organized related codes into preliminary themes. Initial coding remained open to concepts emerging from the data, while the theories of perceived risk and perceived benefit were used as a sensitizing framework to support theme organization and interpretation. Although the final themes were organized under the broad dimensions of perceived risk and perceived benefit, the subthemes were not predefined by these theories. Instead, they were developed and named by the research team based on recurring patterns in participants' responses. The 2 sets of codes and preliminary themes were then compared. Differences in coding interpretations, theme boundaries, and theme names were discussed until consensus was reached. Where uncertainty remained, coding and thematic decisions were discussed with the wider research team. Analytic notes were used to document coding decisions, emerging interpretations, and points requiring further discussion. Fourth, the agreed themes were reviewed and refined to ensure internal consistency and clear boundaries. Fifth, the themes and subthemes were refined and named based on the coded data within the broader risk-benefit framework. Finally, the results of the analysis were compiled and presented in a structured report to support the interpretation of WTP and service acceptance.

The final set of codes, themes, and subthemes was jointly reviewed and confirmed by all team members. Representative quotations were selected from the Chinese transcripts to illustrate each theme, translated into English by bilingual

XSL•FO
RenderX

researchers, and checked against the original text to preserve participants' intended meanings.

Results

Participant Characteristics

A total of 20 participants were included in this study, comprising 5 male and 15 female participants, aged 22-68 years. Of the 20 interviews, 8 were conducted face-to-face and 12 were conducted remotely via telephone. The sample comprised 15 patients, 3 caregivers, and 2 health care professionals. In terms of geographic distribution, 11 participants lived in urban areas and 9 in rural areas. Regarding educational background, 5 participants held a master's degree or higher, 9 held a bachelor's degree, and 6 had completed high school education or below. Detailed participant characteristics are shown in Table 1.

Table . Characteristics of study participants. An OMC[a] experience refers to participants' prior experience with OMC services and does not necessarily indicate prior experience with AI-driven consultation services.

Participant number	Sex	Age (years)	Residence	Identity characteristics	Prior online medical consultation experience	Educational Level
1	Male	41	Urban	Patient	Yes	Master's degree or above
2	Female	36	Urban	Patient	Yes	Master's degree or above
3	Male	41	Urban	Caregiver	Yes	Master's degree or above
4	Female	35	Urban	Patient	Yes	Bachelor's degree
5	Female	22	Rural	Patient	No	Bachelor's degree
6	Female	41	Urban	Caregiver	Yes	Bachelor's degree
7	Female	39	Urban	Patient	Yes	Master's degree or above
8	Female	34	Urban	Health professional (pharmacist)	No	Bachelor's degree
9	Female	35	Rural	Patient	No	Bachelor's degree
10	Female	50	Urban	Patient	No	High school or below
11	Male	24	Rural	Patient	No	Bachelor's degree
12	Male	30	Rural area	Patient	Yes	Master's degree or above
13	Female	22	Rural	Patient	No	Bachelor's degree
14	Female	22	Urban	Patient	Yes	Bachelor's degree
15	Female	62	Rural	Patient	No	High school or below
16	Male	54	Rural	Health professional (doctor)	Yes	Secondary education, vocational education, or below
17	Female	68	Rural	Patient	No	High school or below
18	Female	25	Urban	Caregiver	Yes	Bachelor's degree
19	Female	22	Urban	Patient	No	Bachelor's degree
20	Female	53	Rural	Patient	No	High school or below

[a]OMC: online medical consultation.

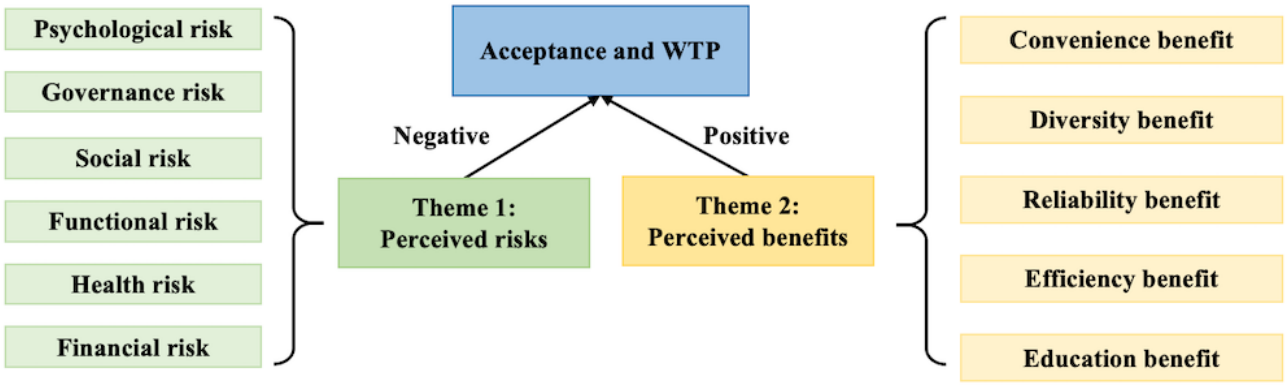
Thematic Analysis Outcomes

Generated Themes: an Overview

Thematic analysis of 20 in-depth interviews identified 2 main themes and 11 subthemes, revealing the multifaceted factors influencing public acceptance and WTP for AI-driven OMC services (Figure 1). Representative participant quotations for themes, subthemes, and WTP are shown in Table 2. These findings reflect the underlying psychological mechanisms and decision-making processes in participants' evaluations of perceived risks and benefits.

XSL•FO
RenderX

Figure 1. Themes and subthemes identified from thematic analysis of public acceptance and willingness to pay (WTP) of AI-driven online medical consultation (OMC) services, based on perceived risk and perceived benefit theories. WTP: willingness to pay.



XSL•FO
RenderX

Table . The themes, subthemes, WTP[a], and participant quotations.

Themes, subthemes, and WTP	Quotations
Theme 1: perceived risks	
Subtheme 1: psychological risk	• "I think the process of AI-driven service was complicated for me, so I didn't use it." (Participant 17; patient without prior OMC[b] experience) • "I'm not really familiar with it yet, and I haven't had the chance to try it myself, so I don't really get how it works. But I guess once I get the hang of it, it might feel a lot easier. I hope the government could promote it more, maybe even have some kind of volunteers or people who can go around teaching how to use it. That would be great, right?" (Participant 20; patient without prior OMC experience) • "I don't really know what AI is. I'm getting on a bit, and I don't fancy trying out new things like that, so I probably won't be using it in the future." (Participant 15; patient without prior OMC experience)
Subtheme 2: governance risk	• "Yeah, I mean, although I haven't used these platforms myself, but I am concerned about the privacy." (Participant 5; patient without prior OMC experience) • "In the future, when AI becomes widely used, how can we ensure that it is not manipulated by certain individuals to create false information? In terms of medical treatment, certain safety assurances need to be provided, particularly in regulatory oversight and control." (Participant 16; service-participant doctor) • "Emm, it's not an official platform backed by the government right now. So, sometimes I'm not feeling confident about it." (Participant 3; service-experienced caregiver) • "If we're consulting at a physical place, when there's a problem, we can quickly identify the person and seek compensation. But for online consultation based on AI, we can't seek accountability." (Participant 11; patient without prior OMC experience)
Subtheme 3: social risk	• "With AI consultations, we are talking to a robot, not a real person. There's no face-to-face interaction, which makes it feel less natural. Unlike a doctor..., the AI-driven platform relies only on what you type in, so it might miss important details." (Participant 12; service-experienced patient) • "I feel like, since the AI cannot do a face-to-face consultation, the responses can sometimes be more conservative. This may be because it lacks direct observation and relies solely on the information I provide." (Participant 1; service-experienced patient)
Subtheme 4: functional risk	• "Also, since the consulting service provided by AI is just a computer program answering questions, it doesn't have the authority to prescribe medicine or order tests." (Participant 4; service-experienced patient) • "Health care professionals will ask about lots of different things when they're diagnosing, like your medical history..., AI might not be quite as good at handling these individual details." (Participant 12; service-experienced patient) • "AI lacks extensive clinical experience, and unlike doctors..., AI tends to rigidly follow standard guidelines." (Participant 18; service-experienced caregiver) • "Online consultations aren't linked to offline medical records, making it hard to review and consolidate information." (Participant 14; service-experienced patient) • "AI still needs training. It might not be well-developed enough to respond like health care professionals." (Participant 7; service-experienced patient) • "Sometimes the system crashed or the internet connection is unstable. It just keeps loading, or the conversation gets wiped out completely. It's really frustrating when that happens." (Participant 7; service-experienced patient)

XSL•FO
RenderX

Themes, subthemes, and WTP	Quotations
Subtheme 5: financial risk	• "What worries me is, if the AI gets it wrong or misses something important, I'll have to go see a doctor anyway. So I'd be paying twice, first for the AI, then for an actual consultation." (Participant 3; service-experienced caregiver) • "Even if the AI gives some medication advice, I still need a doctor for a prescription. So in the end, I'm paying for both, and that's just extra hassle and cost." (Participant 4; service-experienced patient)
Subtheme 6: health risk	• "The service should be available the whole day because you don't know when you might need to ask for advice. In extremely urgent situations, delayed responses can lead to severe outcomes." (Participant 10; patient without prior OMC experience) • "um, sometimes there might even be misunderstandings, you know, there's a risk of things going off track." (Participant 7; patient without prior OMC experience) • "Sometimes, AI doesn't get it right, ...the information and references aren't always accurate. If you don't double-check, you might end up trusting something that's not actually true." (Participant 18; service-experienced caregiver)
Theme 2: Perceived Benefits	

XSL•FO
RenderX

Themes, subthemes, and WTP	Quotations
Subtheme 7: convenience benefit	• "If I can get an appointment, I'd definitely go offline, but often there are no slots available. So, I use online consultations a lot. It's really accessible." (Participant 4; service-experienced patient) • "AI can handle many users at the same time. People can ask it questions anytime, and it gives us a quick response. It is more flexible" (Participant 11; patient without prior OMC experience) • "In rural or remote areas, people may not conveniently go to big cities or top hospitals for consultations. But through this way, they can get medical guidance effectively, and even without having to leave their homes." (Participant 16; doctor without prior OMC experience) • "...But for some small issues that I'm unsure about or I have a general understanding of my condition but encounter further questions or uncertainties later, I think these can easily be addressed through online consultations conducted by AI." (Participant 1; service-experienced patient)
Subtheme 8: diversity benefit	• "Personally, I'm not very fond of oral communication. For text-based consultations, it's easier for me to explain things. I prefer offline interactions with doctors to avoid any potential awkwardness." (Participant 12; service-experienced patient) • "A nice thing about online consultation is I can choose different consultation ways according to my needs. And there's no limit on how many messages you can send." (Participant 2; service-experienced patient) • "For example, if I feel that I have a mental illness and choose to take medication, I will go to the Internet for consultation, because I do not want to let others know." (Participant 9; patient without prior OMC experience)
Subtheme 9: reliability benefit	• "Theoretically speaking, if it's powerful enough, I think it might integrate information from different sources and give more comprehensive advice. In an ideal scenario, it might even be more professional because its memory and storage capacity could be stronger compared to humans. Also, it can retrieve information faster in some ways, so yeah, that's what I think." (Participant 18; service-experienced caregiver) • "If AI can supplement the answers from health care professionals it would be more comprehensive and reliable." (Participant 1; service-experienced patient) • "I'd trust AI more if a doctor reviewed its recommendations before they were given to patients, ensuring accuracy and safety." (Participant 19; patient without prior OMC experience)
Subtheme 10: efficiency benefit	• "AI consultations save time because I can quickly get basic medical information without waiting for a doctor to explain everything. Plus, it's free, so there's no cost at all." (Participant 6; service-experienced caregiver) • "Sometimes, I don't have time to visit a doctor in person, so using AI is much more convenient. It's free or at least much cheaper than online consultations with real doctors." (Participant 2; service-experienced patient) • "I can describe my symptoms to AI, and it helps me figure out which department I should visit. That way, I don't waste time wondering where to go when I actually need to see a doctor." (Participant 14; service-experienced patient)
Subtheme 11: education benefit	

XSL•FO
RenderX

Themes, subthemes, and WTP	Quotations
	• "If patients first use AI to get some initial information about their medication or health concerns, they can communicate more clearly and confidently with pharmacists. This often helps us give them more accurate and effective treatment." (P8; pharmacist without prior OMC experience) • "Health care professionals might not always be familiar with every medication, especially if it's outside their specialty area. In such cases, AI can really help by providing more detailed information." (P16; service-experienced doctor) • "I can first check with the AI, get some background knowledge, and this helps me discuss things better with the doctor later. Also, by using AI, I get to learn more about medication use, drug interactions, and overall health education." (P14; service-experienced patient)
WTP	• "To be honest, I just believe AI consultations should be free. It's not like there's an actual doctor involved. It's just pulling stuff from the internet, so why should we pay for that?" (Participant 11; patient without prior OMC experience) • "If it's just an AI spitting out answers, I wouldn't pay for it. But if there's a proper doctor checking things and making sure the advice is right, then yeah, that's different. That's worth paying for. I'd be willing to pay up to US $30 for that." (Participant 12; service-experienced patient) • "Maybe they could do it in levels—basic AI stuff for free, but if you want a real person to review it, then you pay a bit. At least that way, people can decide what they need." (Participant 18; service-experienced caregiver)

[a]WTP: willingness to pay.

[b]OMC: online medical consultation.

Theme 1: Perceived Risks

Subtheme 1 (Psychological Risk): Unfamiliarity and Skepticism Toward the Services

Many participants without prior OMC experience were skeptical about the reliability of online and AI-driven services and tended to prefer in-person interactions with doctors or pharmacists. Some expressed unfamiliarity with AI-driven medical consultation services, leading to apprehension and reluctance to use the platform. One participant had never heard of AI being used for medical consultations and showed little interest in adopting such technology. Additionally, several participants without prior OMC experience found the AI system difficult to navigate and often required assistance from others to access the service. Their hesitation was further reinforced by a lack of prior experience with similar digital health care platforms. Without adequate information, they found it difficult to recognize the potential benefits of AI-driven consultations. Training provided by health care professionals through video tutorials or user manuals could help them feel more confident in using the service.

Subtheme 2 (Governance Risk): Data Privacy, Security, and Institutional Accountability

Quite a few participants voiced worries over inadequate governance of AI-driven OMC platforms, covering 2 core dimensions, namely data privacy and protection and institutional accountability. Many feared that their confidential medical and personal data might be disclosed to unauthorized third parties, which would erode their trust in digital consultation services. More importantly, participants raised prominent governance loopholes regarding regulatory supervision and liability clarification. Some noted that current AI consultation platforms are not operated by nationally certified authorities, making it unclear where to file complaints or seek redress in cases of inaccurate advice or service failures. The lack of defined responsibility and regulatory oversight left them hesitant to rely on AI-driven consultations. Many participants indicated that they would only consider using and paying for such services if robust data protection measures and transparent mechanisms for complaint handling and accountability were established. Some noted that current AI consultation platforms are not operated by nationally certified authorities, making it unclear where to file complaints or seek redress in cases of inaccurate advice or service failures. The lack of defined responsibility and regulatory oversight left them hesitant to rely on AI-driven consultations. Many participants indicated that they would only consider using and paying for such services if robust data protection measures and transparent mechanisms for complaint handling and accountability were established.

Subtheme 3 (Social Risk): Limited Interpersonal Interaction and Communication Quality

Participants pointed out that one of the shortcomings of AI-driven OMC services was the lack of face-to-face interaction. Without body language or eye contact, they felt uneasy and found it difficult to convey complex information. Some also criticized the overly cautious nature of AI responses, which often suggested possibilities such as "might have" or "could be" instead of giving clear conclusions, leaving users uncertain. The quality of advice also depended on the completeness of the

XSL•FO
RenderX

information provided. When users lacked medical knowledge, struggled to describe symptoms, or had no test results, responses became vague and less useful.

Subtheme 4 (Functional Risk): Limited Clinical, Technical, and Usability Functionality of Platforms

Service-experienced participants indicated that current AI-driven consultation platforms still have functional limitations. Although these systems can analyze user inputs, provide basic drug information, and suggest possible treatment options, they cannot prescribe medications, request laboratory tests, or perform full medical evaluations. Patients must still consult health care professionals for accurate diagnoses, prescriptions, and examinations to ensure safe and effective care. Many participants also noted that AI performs poorly in handling highly individualized medical problems because it lacks clinical experience. It cannot independently access biomarkers or imaging data and depends entirely on user-provided information, which restricts its capacity to offer personalized advice. In addition to these clinical limitations, participants also raised usability-related concerns, including poor network connections, system errors, software glitches, and difficulty navigating the platform. These issues reduced their confidence in the stability and reliability of AI-driven OMC services. As these technologies are still at an early stage, some participants expressed concerns about technical stability, incomplete data integration, and limited accuracy due to insufficient real-world validation.

Subtheme 5 (Financial Risk): Uncertain Pricing and Potential Additional Health Care Costs

Participants expressed concerns that if AI-generated recommendations were inaccurate or incomplete, they might spend more on unnecessary treatments or follow-up consultations with health care professionals. Additionally, since AI lacks the authority to prescribe medications, users still need to consult doctors for prescriptions and laboratory tests, incurring further costs. These financial uncertainties made participants hesitant to invest in AI-driven consultation services, highlighting the need for clear pricing structures and transparent value propositions to justify any associated fees.

Subtheme 6 (Health Risk): Potential Health Harm From Delayed or Inaccurate Advice

Several participants mentioned that some AI-driven consultation services were embedded in company-developed platforms operating within fixed service hours. Although AI was always technically available, responses were sometimes constrained by platform schedules, leading to concerns about the ability to address urgent medical problems during the night. Some participants felt that delayed professional feedback, due to restricted hours or limited consultant availability, might increase health risks. Others expressed worries about the accuracy of AI-generated recommendations. Service-experienced users noted that communication during online consultations was less smooth than in face-to-face settings, which sometimes led to misunderstandings between patients and health care professionals. A few participants further reported that AI occasionally produced incorrect information or cited nonexistent references, creating a false sense of reliability.

Theme 2: Perceived Benefits

Subtheme 1 (Convenience Benefit): Accessible and Flexible Services

Participants highlighted accessibility as a major factor promoting the use of AI-driven OMC services. Those living in rural or underserved areas described these platforms as a convenient alternative for obtaining medical information and guidance, helping them overcome geographical barriers. Participants from larger cities also recognized the convenience of AI-driven consultations compared with traditional health care visits. Many appreciated the flexibility to seek medical advice at times that suited their schedules, often from home or other comfortable locations. Several participants mentioned that AI-driven consultations were useful for addressing minor health issues or follow-up questions after outpatient visits.

Subtheme 2 (Diversity Benefit): Multiple Consultation Formats and Language Options

By offering diverse communication methods and language options, AI-driven online consultation services foster a more inclusive health care experience. Participants indicated that they could choose between voice, text, or video formats according to their preferences. These communication options were also considered advantageous for discussing sensitive health topics, such as mental health, as they offered greater privacy and reduced the discomfort associated with face-to-face communication. In particular, some participants who were more introverted preferred text-based consultations because they found them less stressful than speaking directly with health care providers. Additionally, multilingual support further enhanced inclusivity, enabling users from different linguistic or dialect-speaking backgrounds to seek medical advice without language barriers.

Subtheme 3 (Reliability Benefit): Comprehensive Medical Information via Data Integration

Most participants regarded AI-driven medical consultation services as relatively reliable due to their ability to process large volumes of data and integrate multiple information sources. Unlike human consultations, these systems can rapidly analyze clinical guidelines, drug information, and statistical data to provide evidence-based recommendations. Several participants emphasized that AI's capacity to efficiently retrieve and cross-reference information could enable more comprehensive advice. Furthermore, several participants suggested that integrating AI-generated insights with human consultations could provide more comprehensive and trustworthy guidance, as it supplements health care professionals' responses with additional information. Others believed that implementing a system in which AI-generated recommendations undergo human review before being delivered to patients would further enhance reliability, accuracy, and patient safety.

Subtheme 4 (Efficiency Benefit): Reduced Health Care Costs and Time

Most participants highlighted the cost-effectiveness and time-saving advantages of AI-driven medical consultation services. Unlike traditional online consultations, which can be

XSL•FO
RenderX

expensive, AI services are either free or substantially more affordable, making them a cost-efficient option. They also help save time by providing quick access to basic medical information, reducing the need for lengthy explanations from health care professionals. Additionally, some participants noted that AI can assist in identifying potential health issues and recommending appropriate specialist departments, thereby streamlining the process before in-person visits.

Subtheme 5 (Education Benefit): Improved Health and Medication Literacy

Most participants emphasized that AI-driven medical consultation services play an important role in improving patient education and medical literacy. Participants found that AI's ability to clearly explain medication indications, dosage instructions, mechanisms of action, potential drug interactions, and relevant medical concepts helped them better understand their treatment plans. This enhanced their readiness for face-to-face consultations and facilitated communication with health care providers. AI-driven consultations were perceived as effective tools for empowering patients, promoting informed decision-making, and increasing confidence in managing their medications and health conditions.

WTP

The majority of participants assumed that AI-driven services should be free. Some believed that AI-driven online consultation services could be priced according to their level of human involvement. For example, fully automated AI-driven services that rely solely on algorithmic responses were expected to remain free or low-cost, as they primarily aggregate publicly available medical information. In contrast, AI-assisted services involving health care professionals could be charged, as expert validation and personalized advice make a paid model more justifiable. A tiered pricing model could balance accessibility and quality, offering free basic AI consultations while charging for expert-reviewed recommendations. Participants who preferred traditional offline consultations reported lower WTP, whereas those satisfied with prior OMC experiences showed higher WTP for such services. In summary, participants reported WTP values ranging from RMB 0 to RMB 200 (US $0-$27.28; RMB 1=US $0.1364 as of January 15, 2025), with some indicating a higher WTP for services reviewed or validated by senior professionals or tertiary hospitals.

Discussion

Principal Findings

This qualitative study explored perceived barriers, facilitators, expectations, and factors shaping public acceptance of and stated WTP for AI-driven OMC services. Through in-depth interviews with 20 participants from diverse backgrounds, perceived risks and perceived benefits were identified as 2 key dimensions influencing participants' acceptance and WTP for these services. Regarding perceived risks, psychological, governance, social, functional, health, and financial risks were found to negatively affect acceptance, serving as potential barriers that may discourage engagement with AI-driven OMC services. For perceived benefits, convenience, diversity, reliability, efficiency, and educational benefits emerged as key motivating factors encouraging acceptance. Participants reported WTP values ranging from RMB 0 to RMB 200 (US $0-$27.28; RMB 1=US $0.1364 as of January 15, 2025), with participants' views appearing to vary by their prior OMC experience and understanding of the service model. These findings provide valuable insights for the development and optimization of AI-driven OMC services, supporting targeted pricing strategies and personalized service design that address public preferences and concerns.

The theory of perceived risk provides a valuable framework for interpreting participants' hesitancy toward AI-driven OMC services [27]. As novel health care technologies often face initial resistance due to unfamiliarity [28], psychological risk was reflected in participants' lack of familiarity with AI-driven platforms, leading to skepticism about reliability and a preference for face-to-face interactions. The preference for in-person consultations reflects not only technological hesitancy but also established expectations regarding health care delivery models [29]. Governance risks emerged as major concerns, with participants expressing fears of data leakage and inadequate regulatory oversight. These findings align with previous studies that highlighted concerns about ethical dilemmas, unclear regulations, and data privacy in health care–related AI applications [30,31]. Lack of transparency remains a prevalent issue across most existing AI technologies [32]. Social risk was evident in participants' discomfort with non–face-to-face communication, which they felt hindered the expression of complex health information. Deficiencies in empathetic responses and human-centered care make AI interactions less acceptable to some users [33]. Functional risks were associated with both clinical and technical limitations of the platforms, including the inability to prescribe medications, conduct tests, or deliver highly personalized advice, which reduced the perceived value of the service. Beyond clinical functionality, usability appears to be an important factor shaping trust and acceptance, as technical stability, ease of navigation, and platform reliability influence whether users feel confident relying on AI-driven OMC services. AI-driven consultations were considered more reliable for managing straightforward tasks and therefore more acceptable [34], although technical failures may compromise reliability and erode trust among health care professionals [32,35]. Health risks centered on urgent care availability and potential misinformation, while financial risks reflected uncertainty about whether AI-driven services would ultimately reduce or increase health care costs. These multifaceted risks collectively present substantial barriers to the adoption of and WTP for AI-driven OMC services, particularly among participants without prior OMC experience.

Perceived benefits emerged as key motivating factors for WTP for AI-driven OMC services. Importantly, these benefits refer to participants' perceived benefits rather than demonstrated performance of the AI-driven OMC model, as this study did not evaluate a standardized platform or clinically validate the accuracy, safety, or reliability of medication advice. OMC may extend health care to marginalized populations through remote consultations, monitoring, and technology-assisted diagnosis [36], and AI is expected to provide more real-time predictions

XSL•FO
RenderX

of patient outcomes in the future [37]. AI has expanded access to integrated health care by enhancing provider communication, facilitating interactions, and serving as a knowledge hub [38,39]. Such flexibility aligns with increasing consumer demand for on-demand health care [40,41]. Service diversity, including multiple communication modes and language support, created a more inclusive health care experience appealing to diverse user preferences and communication styles. Reliability was attributed to AI's ability to process and integrate large datasets [42], although hybrid models with human oversight were considered more trustworthy. While user-friendly AI platforms are essential for implementation, integrating them into existing health care systems remains challenging and often requires substantial changes to established operational workflows [43]. Potential cost-effectiveness was perceived as another benefit, with participants expecting time or cost savings compared with traditional consultations. Moreover, participants perceived educational value in these services, particularly for improving medication knowledge, preparing for subsequent consultations, and supporting communication with health care professionals. However, whether these perceived benefits translate into improved clinical outcomes requires further evaluation.

Regarding WTP, our findings reveal that participants' WTP was shaped by how they understood the service model and the level of human involvement. Many participants believed that fully automated AI-driven OMC services should be free or low-cost, whereas services involving review by health care professionals were considered more acceptable as paid services. These findings align with the broader digital health literature showing that WTP is highly context-dependent and varies by service modality, user characteristics, and the perceived value of the service. A systematic review of eHealth WTP studies identified substantial heterogeneity across studies: one-time WTP ranging from US $0.88 to US $191.84 and monthly WTP ranging from US $5.25 to US $45.64, with WTP differing by population characteristics and delivery modality [44]. Another study of primary care web portals similarly suggested that patients may be willing to pay modest fees for online services that offer direct practical value, such as physician messaging, medical record access, and medication refills [45]. More recent evidence from health apps in Hong Kong also showed that WTP was associated with prior app experience and perceived benefit, while unwillingness to pay was commonly linked to expectations that such services should be free, distrust, and limited awareness of benefits [46]. In line with previous studies, our participants appeared to attach value less to AI involvement itself than to clinically meaningful service attributes, particularly professional oversight, trustworthiness, convenience, and affordability. Since WTP was elicited through open-ended qualitative questions and participants discussed different service models, including AI-only and professionally reviewed services, this range should be interpreted as exploratory evidence of perceived value rather than as a pricing estimate. As highlighted by a previous study on cost-aware frameworks, health care AI development expenses—including data acquisition, computational resources, and developer labor—are frequently underestimated in conventional cost assessments [47]. Rather than supporting a specific payment model, our findings suggest that future payment models for AI-driven OMC services need to balance affordability, perceived value, professional oversight, and equitable access. The strategic design of AI payment models is therefore crucial to enhance WTP while optimizing financial efficiency and ensuring equitable access across populations [48]. In the future, insurance companies or health networks may leverage reimbursement policies and out-of-pocket costs to influence or restrict patients' provider choices as part of cost-effectiveness–based payment strategies [49]. Prior OMC experience may have shaped how some participants described service value and payment conditions; however, this study did not formally compare WTP across experience groups. The experience-related observations, which refer to prior OMC experience with or without the involvement of AI, should be interpreted as exploratory and hypothesis-generating rather than as evidence of subgroup differences in WTP. Some participants with prior OMC experience provided more concrete descriptions of payment conditions, whereas participants without such experience more often expressed uncertainty or expected AI-only services to be free. Variations in preferred service models may also reflect personal characteristics, prior digital health experiences, perceived needs, and expectations [50]. These findings suggest that trial periods or freemium models may be worth further exploration to build initial trust, although future quantitative or experimental studies using standardized service descriptions and payment scenarios are needed before drawing conclusions about pricing models.

Implications

By identifying participants' perceived barriers, facilitators, expectations, and payment considerations, our findings provide user-centered implications for the design, implementation, and future evaluation of AI-driven OMC services. This study used a qualitative approach, informed by the theories of perceived risk and perceived benefit, to identify perceived barriers, facilitators, expectations, and factors shaping public acceptance of and stated WTP for AI-driven OMC services. These findings suggest that service design may need to consider not only technical performance but also usability, trust, privacy protection, professional oversight, and affordability from the user perspective. For policymakers, establishing clear regulatory frameworks for AI-driven OMC services could help address privacy concerns and build public trust. Regulations should balance innovation with consumer protection. For telemedicine developers, our results highlight the importance of intuitive, user-friendly interfaces to overcome psychological barriers associated with unfamiliarity. Integrating AI-generated recommendations with human oversight could mitigate concerns about accuracy, personalization, and accountability. From a payment perspective, tiered service models that distinguish between basic AI-only support and professionally reviewed services may be worth further exploration, although future quantitative studies are needed to evaluate WTP and pricing models more precisely.

Limitations

Several limitations should be considered when interpreting our findings. First, the sample size of 20 participants, while appropriate for qualitative research, limits the transferability of the results to broader populations and settings. Although

XSL•FO
RenderX

patients, caregivers, and health care professionals were included to capture different stakeholder perspectives, the small numbers of caregivers and health care professionals limited the ability to compare views across groups or determine whether saturation was achieved within each stakeholder group. Therefore, subgroup-related findings should be interpreted as exploratory rather than representative of each stakeholder group.

Second, the study was conducted during a period when AI-driven OMC services were in the early stages of adoption in China, suggesting that perceptions may evolve as these services become more mainstream. Third, the focus on stated rather than revealed WTP may not fully capture actual payment behaviors in real-world settings. In addition, WTP was explored through open-ended qualitative questions rather than a formal contingent valuation approach, and participants were not asked to evaluate one standardized service model or payment scenario. Therefore, the reported WTP values should be interpreted as exploratory observations of perceived value rather than pricing estimates. Participants' descriptions of AI as efficient, comprehensive, or reliable reflected perceived or expected attributes rather than empirically tested properties of a specific AI-driven OMC platform. Future studies should evaluate the clinical accuracy, safety, and reliability of AI-generated medical advice under standardized service scenarios. Fourth, because the interviews were conducted in Chinese and selected quotations were translated into English for reporting, some linguistic nuances may have been affected during translation, although translated quotations were checked against the original transcripts to preserve participants' intended meanings. Although some findings may be relevant to countries with similar health care contexts, such as uneven medical resource distribution, growing telemedicine use, and early-stage AI integration, transferability should be considered cautiously. Health care financing, regulation, pharmacist roles, and public trust in AI may differ across settings. Future research should address these limitations through larger quantitative studies, cross-cultural comparisons, and experimental designs measuring actual payment behaviors rather than stated intentions.

Conclusion

Through semistructured in-depth interviews with 20 participants from diverse backgrounds, perceived risks (psychological, governance, social, functional, health, and financial risks) and perceived benefits (convenience, diversity, reliability, efficiency, and educational benefits) were identified as 2 key dimensions associated with participants' acceptance and WTP for AI-driven OMC services. These findings provide exploratory insights into participants' concerns, perceived value, and expectations, which may inform future service design and further research on payment models for AI-driven OMC services.

Acknowledgments

We would like to thank all the researchers who participated in this study and the respondents who took part in the survey.

Disclosure of Delegation to Generative AI (GenAI)

The authors declare the use of GenAI in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GenAI tools under full human supervision: Proofreading and editing

The GenAI tool used was: ChatGPT-5.6 Sol.

Responsibility for the final manuscript lies entirely with the authors.

GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Declaration submitted by: Qiushi Cai

Funding

This research was funded by the National Natural Science Foundation of China [grant number: 72304009].

Data Availability

Due to the sensitive nature of qualitative interview data and privacy considerations, the datasets generated and analyzed during the current study are not publicly available. Deidentified data may be provided by the corresponding author upon reasonable request and with approval from the Ethics Committee of Peking University Third Hospital (Approval No. M20241044).

Authors' Contributions

Conceptualization: SD, QC, JY, YC
Data curation: SD, QC, JY, YC
Formal analysis: SD, QC
Investigation: SD, QC
Methodology: SD, QC, RZ, WKM
Project administration: JY, YC
Resources: RZ, WKM
Supervision: RZ, WKM

XSL•FO
RenderX

Writing – original draft: SD, QC
Writing – review & editing: SD, QC, JY, YC, RZ, WM
RZ and WKM are co-corresponding authors.

Conflicts of Interest

None declared.

Checklist 1
COREQ checklist.
[PDF File, 551 KB - humanfactors_v13i1e96448_app1.pdf]

References

1. World Health Organization. Consolidated telemedicine implementation guide. In: Geneva: World Health Organization: World Health Organization; 2022.
2. Dorsey ER, Topol EJ. Telemedicine 2020 and the next decade. The Lancet 2020 Mar;395(10227):859. [doi: 10.1016/S0140-6736(20)30424-4]
3. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. In: Artificial Intelligence in Healthcare: Elsevier; 2020:25-60. [doi: 10.1016/B978-0-12-818438-7.00002-2]
4. Cunha Reis T. Artificial intelligence and natural language processing for improved telemedicine: before, during and after remote consultation. Aten Primaria 2025 Aug;57(8):103228. [doi: 10.1016/j.aprim.2025.103228] [Medline: 39955812]
5. Shiraishi M, Tanigawa K, Tomioka Y, et al. Blepharoptosis consultation with artificial intelligence: aesthetic surgery advice and counseling from chat generative pre-trained transformer (ChatGPT). Aesth Plast Surg 2024 Jun;48(11):2057-2063. [doi: 10.1007/s00266-024-04002-4]
6. Siglen E, Vetti HH, Lunde ABF, et al. Ask Rosa - the making of a digital genetic conversation tool, a chatbot, about hereditary breast and ovarian cancer. Patient Educ Couns 2022 Jun;105(6):1488-1494. [doi: 10.1016/j.pec.2021.09.027] [Medline: 34649750]
7. Yau JYS, Saadat S, Hsu E, et al. Accuracy of prospective assessments of 4 large language model chatbot responses to patient questions about emergency care: experimental comparative study. J Med Internet Res 2024 Nov 4;26:e60291. [doi: 10.2196/60291] [Medline: 39496149]
8. Franco D'Souza R, Amanullah S, Mathew M, Surapaneni KM. Appraising the performance of ChatGPT in psychiatry using 100 clinical case vignettes. Asian J Psychiatr 2023 Nov;89:103770. [doi: 10.1016/j.ajp.2023.103770] [Medline: 37812998]
9. Thouil A, Rhazari M, Alaoui H, Serraj K, Kouismi H. Bronchial fissuring: a severe complication of arterial aneurysm. Cureus . [doi: 10.7759/cureus.43029]
10. Gao S, He L, Chen Y, Li D, Lai K. Public perception of artificial intelligence in medical care: content analysis of social media. J Med Internet Res 2020 Jul 13;22(7):e16649. [doi: 10.2196/16649] [Medline: 32673231]
11. Almathami HKY, Win KT, Vlahu-Gjorgievska E. Barriers and facilitators that influence telemedicine-based, real-time, online consultation at patients' homes: systematic literature review. J Med Internet Res 2020 Feb 20;22(2):e16407. [doi: 10.2196/16407] [Medline: 32130131]
12. Azar KMJ, Koliwad S, Poon T, et al. The Electronic CardioMetabolic Program (eCMP) for patients with cardiometabolic risk: a randomized controlled trial. J Med Internet Res 2016 May 27;18(5):e134. [doi: 10.2196/jmir.5143] [Medline: 27234480]
13. Yang M, Jiang J, Cameron AF, Liu X. How do you cope? Online medical consultation service uncertainty, coping strategies, and subsequent payment. Electron Commer Res Appl 2023 Sep;61:101294. [doi: 10.1016/j.elerap.2023.101294]
14. Demirgüneş BK. Relative importance of perceived value, satisfaction and perceived risk on willingness to pay more. Int Rev Manag Mark 2015 Mar(4):211-220 [FREE Full text]
15. Schnall R, Higgins T, Brown W, Carballo-Dieguez A, Bakken S. Trust, perceived risk, perceived ease of use and perceived usefulness as factors related to mHealth technology use. Stud Health Technol Inform 2015;216:467-471. [doi: 10.3233/978-1-61499-564-7-467] [Medline: 26262094]
16. Qi M, Cui J, Li X, Han Y. Perceived factors influencing the public intention to use e-consultation: analysis of web-based survey data. J Med Internet Res 2021 Jan 20;23(1):e21834. [doi: 10.2196/21834] [Medline: 33470934]
17. Limbu YB, Huhmann BA. What influences consumers' online medication purchase intentions and behavior? A scoping review. Front Pharmacol 2024;15:1356059. [doi: 10.3389/fphar.2024.1356059] [Medline: 38414739]
18. Gupta S, Kamboj S, Bag S. Role of risks in the development of responsible artificial intelligence in the digital healthcare domain. Inf Syst Front 2023 Dec;25(6):2257-2274. [doi: 10.1007/s10796-021-10174-0]
19. Kim DJ, Ferrin DL, Rao HR. A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. Decis Support Syst 2008 Jan;44(2):544-564. [doi: 10.1016/j.dss.2007.07.001]

XSL•FO
RenderX

20. Gong Z, Han Z, Li X, Yu C, Reinhardt JD. Factors influencing the adoption of online health consultation services: the role of subjective norm, trust, perceived benefit, and offline habit. Front Public Health 2019;7:286. [doi: 10.3389/fpubh.2019.00286] [Medline: 31637229]
21. Moser A, Korstjens I. Series: practical guidance to qualitative research. Part 3: sampling, data collection and analysis. Eur J Gen Pract 2018 Dec;24(1):9-18. [doi: 10.1080/13814788.2017.1375091] [Medline: 29199486]
22. Pyo J, Lee W, Choi EY, Jang SG, Ock M. Qualitative research in healthcare: necessity and characteristics. J Prev Med Public Health 2023 Jan;56(1):12-20. [doi: 10.3961/jpmph.22.451] [Medline: 36746418]
23. Barroga E, Matanguihan GJ. A practical guide to writing quantitative and qualitative research questions and hypotheses in scholarly articles. J Korean Med Sci 2022 Apr 25;37(16):e121. [doi: 10.3346/jkms.2022.37.e121] [Medline: 35470596]
24. Dicicco-Bloom B, Crabtree BF. The qualitative research interview. Med Educ 2006 Apr;40(4):314-321. [doi: 10.1111/j.1365-2929.2006.02418.x] [Medline: 16573666]
25. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
26. Kiger ME, Varpio L. Thematic analysis of qualitative data: AMEE Guide No. 131. Med Teach 2020 Aug;42(8):846-854. [doi: 10.1080/0142159X.2020.1755030] [Medline: 32356468]
27. Sarkar S, Chauhan S, Khare A. A meta-analysis of antecedents and consequences of trust in mobile commerce. Int J Inf Manage 2020 Feb;50:286-301. [doi: 10.1016/j.ijinfomgt.2019.08.008]
28. Safi S, Thiessen T, Schmailzl KJG. Acceptance and resistance of new digital technologies in medicine: qualitative study. JMIR Res Protoc 2018 Dec 4;7(12):e11072. [doi: 10.2196/11072] [Medline: 30514693]
29. Alpert JS. Face-to-face versus digital encounters in the clinic. Am J Med 2024 May;137(5):379-380. [doi: 10.1016/j.amjmed.2023.06.015] [Medline: 37423431]
30. Farhud DD, Zokaei S. Ethical issues of artificial intelligence in medicine and healthcare. ijph 2021(11). [doi: 10.18502/ijph.v50i11.7600]
31. Naik N, Hameed BMZ, Shetty DK, et al. Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility? Front Surg 2022;9:862322. [doi: 10.3389/fsurg.2022.862322] [Medline: 35360424]
32. Ahmed MI, Spooner B, Isherwood J, Lane M, Orrock E, Dennison A. A systematic review of the barriers to the implementation of artificial intelligence in healthcare. Cureus 2023 Oct;15(10):e46454. [doi: 10.7759/cureus.46454] [Medline: 37927664]
33. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019;5:2055207619871808. [doi: 10.1177/2055207619871808] [Medline: 31467682]
34. Lambert SI, Madi M, Sopka S, et al. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. NPJ Digit Med 2023 Jun 10;6(1):111. [doi: 10.1038/s41746-023-00852-5] [Medline: 37301946]
35. Bedaf S, Marti P, Amirabdollahian F, de Witte L. A multi-perspective evaluation of a service robot for seniors: the voice of different stakeholders. Disabil Rehabil Assist Technol 2018 Aug;13(6):592-599. [doi: 10.1080/17483107.2017.1358300] [Medline: 28758532]
36. Anawade PA, Sharma D, Gahane S. A comprehensive review on exploring the impact of telemedicine on healthcare accessibility. Cureus 2024 Mar;16(3):e55996. [doi: 10.7759/cureus.55996] [Medline: 38618307]
37. Yelne S, Chaudhary M, Dod K, Sayyad A, Sharma R. Harnessing the power of AI: a comprehensive review of its impact and challenges in nursing science and healthcare. Cureus . [doi: 10.7759/cureus.49252]
38. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: capabilities, features, barriers, and applications. Sens Int 2021;2:100117. [doi: 10.1016/j.sintl.2021.100117] [Medline: 34806053]
39. Sharma S, Rawal R, Shah D. Addressing the challenges of AI-based telemedicine: best practices and lessons learned. J Educ Health Promot ;12(1):338. [doi: 10.4103/jehp.jehp_402_23]
40. Alowais SA, Alghamdi SS, Alsuhebany N, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ 2023 Sep 22;23(1):689. [doi: 10.1186/s12909-023-04698-z] [Medline: 37740191]
41. Ezeamii VC, Okobi OE, Wambai-Sani H, et al. Revolutionizing healthcare: how telemedicine is improving patient outcomes and expanding access to care. Cureus 2024 Jul;16(7):e63881. [doi: 10.7759/cureus.63881] [Medline: 39099901]
42. Collins C, Dennehy D, Conboy K, Mikalef P. Artificial intelligence in information systems research: a systematic literature review and research agenda. Int J Inf Manage 2021 Oct;60:102383. [doi: 10.1016/j.ijinfomgt.2021.102383]
43. Maleki Varnosfaderani S, Forouzanfar M. The role of AI in hospitals and clinics: transforming healthcare in the 21st century. Bioengineering (Basel) 2024 Mar 29;11(4):337. [doi: 10.3390/bioengineering11040337] [Medline: 38671759]
44. Erion G, Janizek JD, Hudelson C, et al. A cost-aware framework for the development of AI models for healthcare applications. Nat Biomed Eng 2022 Dec;6(12):1384-1398. [doi: 10.1038/s41551-022-00872-8] [Medline: 35393566]
45. Xie Z, Chen J, Or CK. Consumers' willingness to pay for eHealth and its influencing factors: systematic review and meta-analysis. J Med Internet Res 2022 Sep 14;24(9):e25959. [doi: 10.2196/25959] [Medline: 36103227]
46. Adler KG. Web portals in primary care: an evaluation of patient readiness and willingness to pay for online services. J Med Internet Res 2006 Oct 26;8(4):e26. [doi: 10.2196/jmir.8.4.e26] [Medline: 17213045]
47. Liu H, Xie Z, Or C. Willingness to pay for health apps, its sociodemographic correlates, and reasons for being unwilling to pay. Digit Health 2024;10:20552076241248925. [doi: 10.1177/20552076241248925] [Medline: 38698831]

XSL•FO
RenderX

48. Parikh RB, Helmchen LA. Paying for artificial intelligence in medicine. NPJ Digit Med 2022 May 20;5(1):63. [doi: 10.1038/s41746-022-00609-6] [Medline: 35595986]
49. Robertson C, Woods A, Bergstrand K, Findley J, Balser C, Slepian MJ. Diverse patients' attitudes towards artificial intelligence (AI) in diagnosis. PLoS Digit Health 2023 May;2(5):e0000237. [doi: 10.1371/journal.pdig.0000237] [Medline: 37205713]
50. Murry LT, Witry MJ, Urmie JM. A qualitative exploration of patient preferences for Medicare Part D consultation services offered in a community pharmacy setting. J Am Pharm Assoc (2003) 2023;63(1):97-107. [doi: 10.1016/j.japh.2022.08.020] [Medline: 36151025]

Abbreviations

mHealth: mobile health
OMC: online medical consultation
WTP: willingness to pay

Edited by A Bamgboje-Ayodele; submitted 28.Mar.2026; peer-reviewed by A Hartanto, C Or; revised version received 06.Jul.2026; accepted 19.Aug.2026; published 06.Oct.2026.

Please cite as:
Dong S, Cai Q, Ye J, Chen Y, Zhao R, Ming WK
Using Theory-Based Frameworks to Identify Barriers, Facilitators, Expectations, and Willingness to Pay for Online Medical Consultation Services in China: Qualitative Study
JMIR Hum Factors 2026;13:e96448
URL: https://humanfactors.jmir.org/2026/1/e96448
doi:10.2196/96448

© Shujie Dong, Qiushi Cai, Jingyi Ye, Yanya Chen, Rongsheng Zhao, Wai-kit Ming. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 6.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluation of Wellby, a Cocreated Mobile App and Wearable to Support Stress Management and Overall Well-Being: Mixed Methods Acceptability and Usability Study

Justin Laiti[1], BS; Sina Javadpour[2], MS; Jenna Mullen[3], BS; Claire Conway[2], PhD; Elaine Byrne[1], PhD; Pádraic Dunne[1], PhD

[1]Centre for Positive Health Sciences, Royal College of Surgeons in Ireland, 123 St Stephen's Green, Dublin, Ireland

[2]Tissue Engineering Research Group, Royal College of Surgeons in Ireland, Dublin, Ireland

[3]School of Medicine, Royal College of Surgeons in Ireland, Dublin, Ireland

Corresponding Author:
Justin Laiti, BS
Centre for Positive Health Sciences, Royal College of Surgeons in Ireland, 123 St Stephen's Green, Dublin, Ireland

Abstract

Background: Digital well-being support tools can offer adolescents tailored interventions embedded in their digital environments. However, there is a lack of high-quality, evidence-based digital interventions specifically designed for young people's well-being needs. Wellby is a mobile app and wearable device cocreated with Irish secondary school students to support stress management and overall well-being.

Objective: This study aimed to evaluate the acceptability and usability of the Wellby system using the technology acceptance model, focusing on ease of use, perceived usefulness, and behavioral intention among Irish secondary school students.

Methods: A mixed methods acceptability and usability study was conducted with Irish secondary school students (n=43) across 3 schools: 2 mainstream secondary schools and 1 Youthreach center for early school leavers. Students accessed the Wellby mobile app for 8 weeks and received the custom wearable device for the final 4 weeks (April-May 2024). Wellby included mood tracking, to-do lists, educational resources, text-based health coaching, and heart rate variability biofeedback with guided breathing exercises. Acceptability and usability were evaluated using the user version of the Mobile App Rating Scale (uMARS), structured focus groups, and app engagement analytics. Secondary measures included baseline well-being questionnaires (Perceived Stress Scale, Pittsburgh Sleep Quality Index, and EPOCH Measure of Adolescent Well-Being).

Results: Among survey respondents (n=29), uMARS subscale scores were high: information quality (mean 4.2/5.0), functionality (4.1/5.0), and aesthetics (4.1/5.0). The perceived impact subscale showed the highest scores for well-being awareness (3.8/5.0) and help-seeking intention (3.9/5.0). Of survey respondents who participated in prior co-design sessions, 88% (21/24) and 83% (20/24) felt the wearable and app, respectively, aligned with their co-design feedback. App engagement more than doubled following the introduction of the wearable and revealed that the wearable tab was most accessed (643/1688, 38% of interactions), followed by the home (430/1688, 25%), resources (338/1688, 20%), and coaching (277/1688, 16%) tabs. Focus group feedback highlighted the mood tracker as the most valued feature and identified technical improvements, including better battery life, enhanced Bluetooth connectivity, and more personalization options.

Conclusions: This study demonstrates that cocreated digital well-being tools can achieve high acceptability and usability with Irish secondary school students. Applying the technology acceptance model effectively captured student experience through mixed methods feedback. Students particularly valued self-tracking and personalization features over coaching or educational features. These findings highlight the importance of aligning digital technologies with adolescents' needs and preferences, such as features that encourage increased autonomy and identity formation. Future iterations of Wellby should address the technical limitations while continuing the involvement of students to adapt to their developmental needs and maintain high acceptability.

Trial Registration: ClinicalTrials.gov NCT06294210; https://clinicaltrials.gov/study/NCT06294210

(JMIR Hum Factors 2026;13:e79381) doi:10.2196/79381

KEYWORDS

adolescents; stress; sleep; wearable electronic devices; mobile apps; cocreation; acceptability; technology acceptance model; TAM

XSL•FO
RenderX

Introduction

Background

Digital tools present novel pathways to enhance the health and well-being of young people, particularly given the ubiquity of digital technology in many adolescents' lives [1]. Mobile apps offer scalable and accessible routes for adolescents to avail themselves of educational tools and interventions that address their well-being needs [2]. These can facilitate healthy lifestyle habits during the formative period of adolescence, which can encourage positive health later in life and reduce the burden of preventable noncommunicable diseases [3,4]. Additionally, connected technologies can augment the well-being support capabilities of mobile apps, such as wearable devices, which are increasingly used for tracking activities, health behaviors, and continuous physiological data [5]. While mobile apps include affordances that show promise in supporting adolescent well-being, there is limited research into adolescent-specific digital tools [6,7]. Given the potential for these tools to provide crucial support for young people, there is a need to further investigate and develop digital tools that are specifically designed to consider the health and well-being needs of this age group [2,8]. Because social, environmental, and demographic factors significantly impact adolescent health outcomes [9,10], understanding local contexts and cultures is critical for designing tailored interventions for specific cohorts of adolescents [7,11]. Several studies have demonstrated the importance of adopting adolescent-centered designs to promote intervention acceptability, usability, and effectiveness [12-15].

Co-Design and Participatory Approaches

Participatory design practices are well documented throughout history [16], and there are several modern frameworks for engaging participants in the cocreation of new interventions with young people [17,18]. Radical participatory design is one framework for any population that stresses the need to shift power from researchers or designers to the participants throughout each phase of the intervention creation process [16]. While many design processes involve users only after a prototype has been developed, co-design and cocreation offer a more inclusive approach by positioning participants as equal partners throughout the entire process [19]. The term co-design tends to focus solely on the technology development process, while cocreation includes broader involvement of participants such as during planning and implementation phases [19]. This collaborative approach is particularly important when designing for adolescents, as their unique contexts and preferences require specialized consideration.

When developing digital tools for this age group, it is also essential to acknowledge the risks of these technologies so that potential privacy and safety concerns are considered in the design. Digital tools present several potential risks, including concerns about data privacy, inappropriate content exposure, overuse or dependence, algorithmic bias, and insufficient safeguarding mechanisms [20,21]. Specifically, an increased reliance on mobile apps, particularly social media, has been associated with increased anxiety and depression among young people [22,23]. Often, these apps implement addictive features, aiming to maximize people's time on the platform, which can be particularly harmful for adolescents [24]. Participatory engagement and acceptability testing can help to mitigate these risks by centering the experiences, input, and impact on the people who will use these technologies.

Current Gaps in Adolescent Digital Health

Despite the potential of digital tools to improve adolescent health [1,6,7], there remains a gap in evidence-based digital interventions focused on supporting adolescent well-being. A review of over 1000 mobile apps available for wellness and stress management found that only 2% were supported by scientific research papers, while 4% of the apps focused on either young adults or youth [25]. Similarly, reviews of mobile apps targeting adolescents have found a need for more rigorous evaluation and the incorporation of co-design processes to promote relevant and engaging interventions for young people [6,26]. Among platforms co-designed with young people, many focus on chronic disease management or clinical mental health conditions rather than stress or overall well-being [17].

In response to the lack of adolescent-specific and evidence-based interventions, researchers have increasingly called for more rigorous evaluation of these tools to better understand their effectiveness, safety, and long-term impact on adolescent well-being [7,27,28]. Fields such as positive technology have emerged to focus on the minimization of risk while maximizing features that promote healthy and meaningful engagement with technology [29]. Additionally, participatory design processes can help to mitigate risks by fostering trust between adolescents, caregivers, and researchers or developers, while also ensuring alignment with the adolescents' goals and preferences [30].

Related Acceptability and Usability Studies

The assessment of acceptability and usability can help inform the iterative development of technology [31] and enhance the impact on participants' well-being [32]. This can be measured in several ways, including quantitative assessments using validated instruments, engagement patterns captured through analytics, and qualitative feedback on people's experience with technology. Inal et al [31] conducted a systematic review of usability evaluations for mobile mental health technologies, analyzing 42 studies across various populations. In this review, they found that data were most often collected through questionnaires, field studies, and interviews, with user reception and people's perception of the intervention being the main outcomes of focus. Notably, only 3 of the 42 included studies incorporated participatory design in the development of the technologies that were tested. Another systematic review by Creaser et al [33] synthesized results from 33 studies on the acceptability of wearable activity trackers for promoting physical activity in youth. This found that wearables increased motivation in young people to be more physically active, despite technical difficulties and novelty effects of wearables use.

A similar study investigated the app engagement, usability, and satisfaction of adolescents, aged 13 - 20 years, interacting with the BeMe app, focused on supporting adolescent mental health [34]. The BeMe app included content based on cognitive behavioral therapy, interactive activities, text-based coaching,

XSL•FO
RenderX

and links to clinical services. The majority of participants in this study engaged with the content, mood ratings, and skill training activities on the app, and greater participant engagement was associated with increased app satisfaction. Another study assessed the usability and acceptability of Fitbits for Australian adolescents aged 13 - 14 years during a 6-week period [35]. The adolescents in this study found the Fitbit easy to use and useful for monitoring physical activity, with barriers including comfort, design issues, and lack of activity-specific feedback. Additionally, Martin et al [36] outlined the co-design of a mobile app to promote a healthy lifestyle in European adolescents aged 13 - 16 years. Based on the co-design process, a series of recommendations were established for the mobile app including personalization features, accessible language and design, and comprehensive app tutorials. Finally, Swahn et al [37] summarized feedback from female adolescent participants using a wearable in an urban setting in Uganda. Despite general positive feedback from participants, they also shared concerns about comfort, surveillance, and data privacy, and strong reactions from community members. Although these studies focus on adolescent feedback on digital well-being tools, most evaluate either mobile apps or wearable devices in isolation.

Furthermore, few studies employ theoretical frameworks to systematically assess the acceptability and usability of these tools. The technology acceptance model (TAM) was chosen for this study as it offers a simple yet effective and adaptable framework for acceptability that has been applied in diverse contexts, such as education [38]. Other common frameworks include the unified theory of acceptance and use of technology (UTAUT) [39] and the theoretical framework of acceptability (TFA) [40]. The UTAUT is based on several frameworks, including the TAM, and outlines 4 key factors and modifiers that influence intention and usage, while the TFA was established for health-related interventions and includes prospective, concurrent, and retrospective components of acceptability. The TAM offers several advantages for this study. It presents a robust measure with previous application in similar adolescent and educational contexts [38,41], enabling comparison to related work and promoting reproducibility and adaptability of this method for future studies.

Aim and Hypothesis

This paper outlines student evaluations of the acceptability and usability of Wellby, a cocreated mobile app and wearable, to help to iteratively develop this intervention for student well-being. The acceptability and usability testing detailed in this paper helps to maintain alignment with students' needs and promote beneficial technology usage. The mixed methods approach of this study is transferable to other cocreated interventions, and students offer feedback that can be used to enhance digital tools more generally to support adolescent-specific well-being needs.

The aim of this study is to assess the acceptability and usability of the Wellby mobile app and wearable as part of a cocreation process with Irish secondary school students. The primary outcomes are ease of use, perceived usefulness, and behavioral intention of Wellby, which are measured using the TAM framework both quantitatively and qualitatively [42]. Secondary outcomes are students' stress, sleep, and overall well-being, measured by baseline questionnaires and demographic information to further contextualize the results. Wellby has been created through an iterative cocreation process with students (aged 15 - 19 years) at 4 Irish secondary schools to address students' well-being needs through a mobile app and wearable that offer educational resources, heart rate variability (HRV) tracking, and text-based coaching. Based on participatory design principles, we hypothesize that the cocreation process will lead to a system that aligns with students' well-being goals and demonstrates high acceptability and usability among students. This study was conducted across 3 of the 4 Irish secondary schools that were involved in the co-design process. Acceptability and usability feedback was gathered to adapt the next version of Wellby to provide relevant stress and well-being management tools for students.

Methods

Overview

This mixed methods acceptability and usability study evaluated Wellby over 8 weeks with Irish secondary school students (n=43) based on the TAM [42].

Ethical Considerations

Ethical approval for this study was obtained through the Research Ethics Committee at the Royal College of Surgeons in Ireland (ID 202403004). Informed consent was obtained from students' parents and/or legal guardians, with written assent provided by adolescents. This study was registered with ClinicalTrials.gov (NCT06294210). Participant data were pseudonymized, and qualitative feedback was assessed for identifiable information to preserve participant privacy. No compensations or incentives were provided to participants besides the potential benefits of the intervention itself.

Participants

Students from 3 schools participated in this study. Two of the schools were mainstream schools, while one school was a Youthreach center. Youthreach is an alternative coeducational program with locations around Ireland, providing specialized education for early mainstream school leavers [43]. The participants at the mainstream schools (schools A and B) were in their 5th year, aged 16 - 17 years. School A is a coeducational school in County Wexford, and school B is an all-female school in County Kildare. The survey participants at Youthreach (school C), in County Dublin, were in their first or second year of the program, aged 15 - 19 years. Eighty-eight percent of student participants were also involved in previous co-design sessions (Table 1). In usability studies, a baseline of 5 - 10 participants is recommended by Macefield [44]. This target number was reached for each school in this study to ensure representative feedback since these schools previously showed slight differences in well-being needs [45].

XSL•FO
RenderX

Table . Student participant demographics by school.

Characteristics	School A (n=16), n	School B (n=11), n	School C (n=16), n	Total (n=43), n
Age (years)				
16	5	2	5	12
17	11	7	7	25
18	0	2	1	3
19	0	0	3	3
Race and ethnicity				
White Irish	12	5	13	30
White/other European	4	2	1	7
Asian	0	2	1	3
Black or Black Irish	0	1	1	2
White African	0	1	0	1
Gender				
Female	8	10	9	27
Male	8	0	5	13
Nonbinary	0	0	1	1
Prefer not to say	0	1	1	2
Previous research engagement				
Prior co-design participants	13	10	15	38
New participants	3	1	1	5
Wearable user				
Yes	3	1	2	6
No	13	10	14	37

Wellby Mobile App and Wearable Description

The co-designed Wellby mobile app was published on the iOS Apple Store and the Google Play Store for students to access. The app was developed by JL, with educational well-being and HRV content curated by JM, PD, and JL. The mobile app includes four tabs (Figure 1): (1) a home tab with student-requested tools, including an optional daily quote, mood check-in and tracker, a clickable calendar with daily to-do lists, and settings where customizable app colors can be chosen; (2) an educational resource tab with evidence-based articles, infographics, and videos on topics requested by students, including stress management, sleep, digital well-being, and time management; (3) a wearable tab that facilitates connection to the Wellby wearable device and offers options to start new recordings, see past recordings, learn about HRV, and engage in breath work; and (4) a health coaching tab that offers optional text-based messaging with certified health coaches around student-requested well-being focus areas. Two certified health coaches were recruited through the Centre for Positive Health Sciences at the Royal College of Surgeons in Ireland (RCSI) and were vetted by RCSI and the participating schools. In future implementations, coaches could be school counselors, health and well-being staff, or community health workers.

The Wellby wearable is a wrist-worn device that includes a photoplethysmography sensor for tracking heart activity, a rechargeable battery, a Bluetooth connection module, and a light indicator of battery and connection status (Figure 2). This device was designed and assembled by engineering researchers based at RCSI (JL and SJ). Developing a custom device in this study allowed for increased adaptability and customizability of the intervention based on student input throughout the iterative cocreation process. It also enhanced data control, analysis, and security, enabling intentional and transparent data collection and processing. Students indicated in the co-design that the appearance of the device was important to them. Therefore, they were able to request the color of the nylon case, strap design, and strap material for the wearable in this study.

XSL•FO
RenderX

Figure 1. Wellby mobile app tabs showing the home tab (far left), resource tab topics (second to the left), wearable tab (second to the right), and coaching tab (far right).

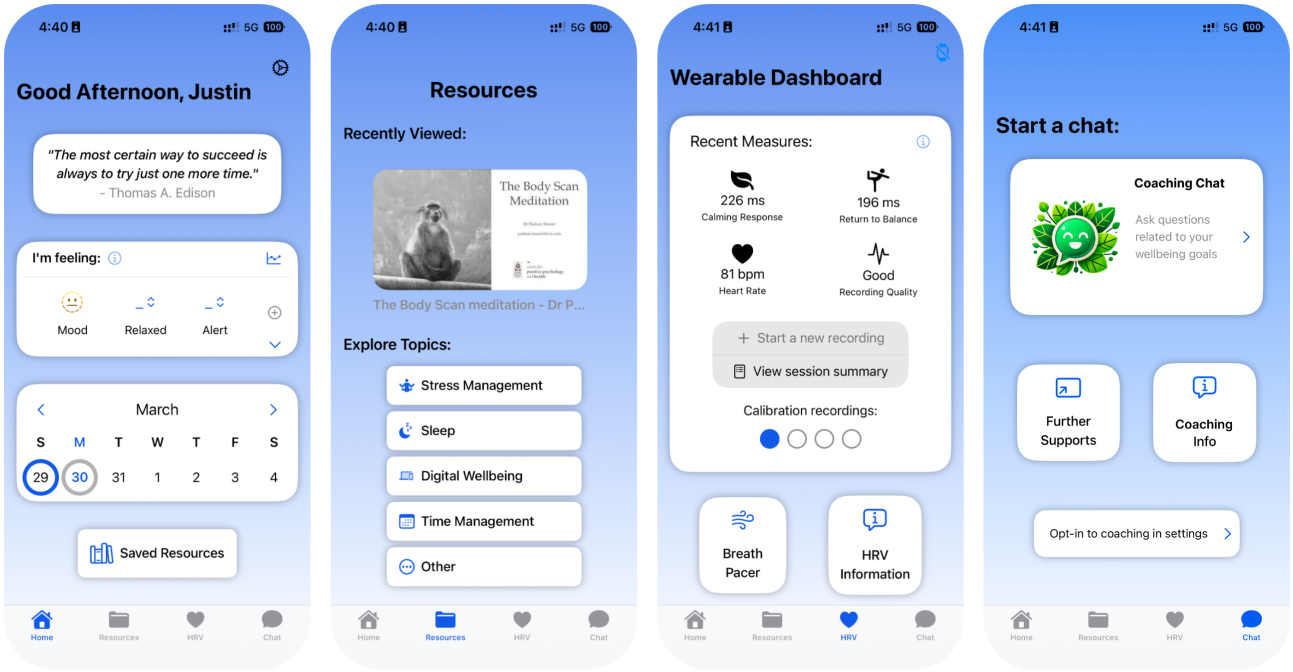
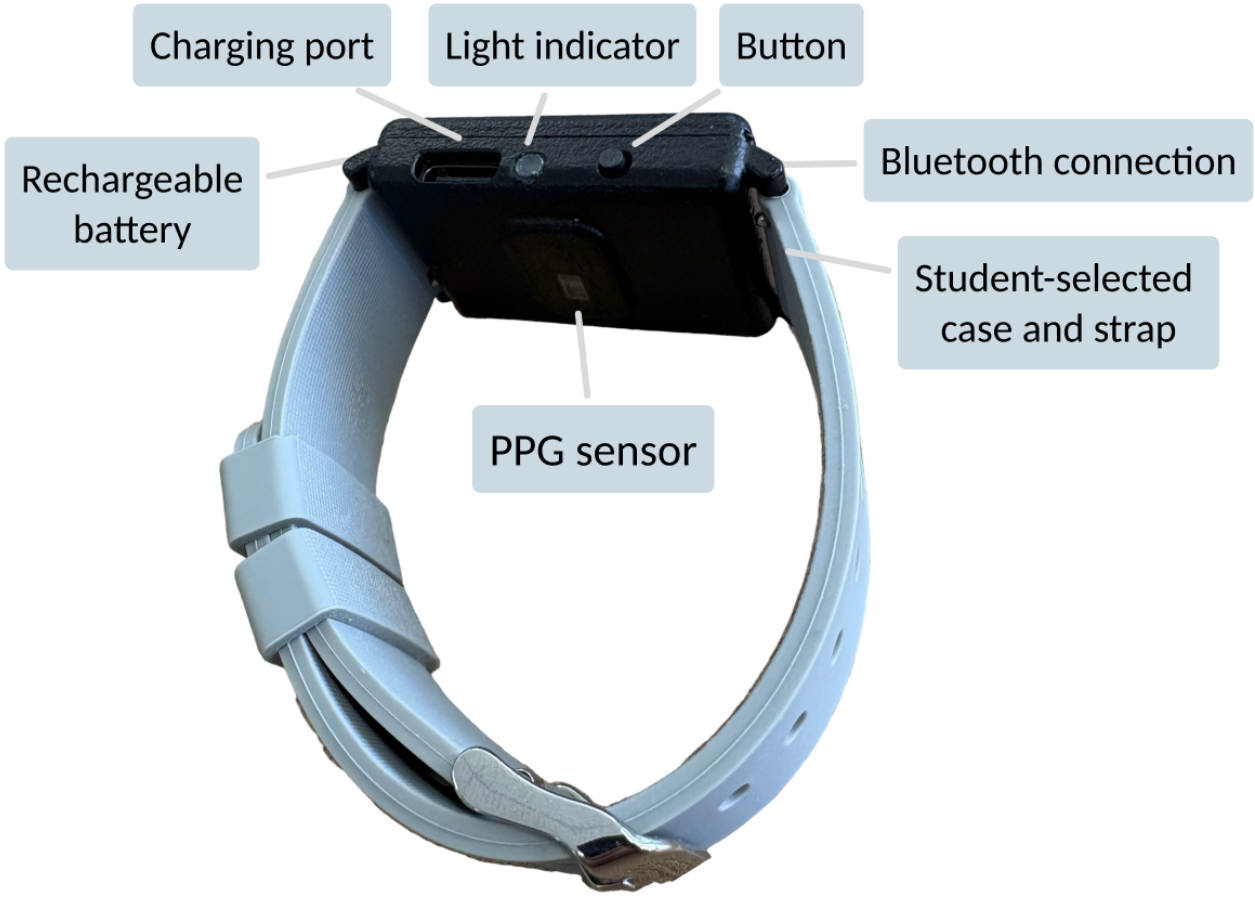


Figure 2. Wellby device feature schematic, including one option for the customizable strap and case colors. PPG: photoplethysmography.



Study Design

In this study, students tested Wellby for 8 weeks (April-May 2024). The primary outcome measure of acceptability and usability was assessed using the user version of the Mobile App Rating Scale (uMARS) [46], survey questions, qualitative feedback from follow-up focus group discussions, and app engagement measured throughout the study. As secondary outcome measures, well-being questionnaires were collected at

XSL•FO
RenderX

baseline. Perceived stress, sleep quality, and overall well-being were measured using the 10-item Perceived Stress Scale (PSS) [47], the Pittsburgh Sleep Quality Index (PSQI) [48], and the EPOCH (engagement, perseverance, optimism, connectedness, and happiness) Measure of Adolescent Well-Being [49], respectively. Well-being questionnaires were only collected at baseline because of students' limited availability and increased pressure during the final weeks of their school year. We prioritized collecting the primary outcomes for this acceptability and usability study, which was designed to inform a follow-up longitudinal efficacy study with repeated well-being measures. Additional students were recruited at schools A and B who completed the well-being questionnaires without testing Wellby. This group was recruited to compare the well-being of students with and without access to Wellby for potential future efficacy studies on the impact of Wellby on student well-being. No additional students were recruited at school C because all students were involved in testing Wellby.

Students first had access to the Wellby mobile app alone for 4 weeks before receiving the connected wrist-worn HRV monitor for the final 4 weeks of the study. This allowed for the investigation of the acceptability and usability of the mobile app alone and in combination with the wearable device. Four weeks represents a common duration of acceptability studies of wearables for adolescents [33]. The phased approach aligns with the iterative development of cocreated tools [36] and enabled comparison between the intervention components. Students were encouraged to use Wellby as much as they found useful throughout the study period. The uMARS and hour-long focus group sessions were completed at the end of the 8-week study.

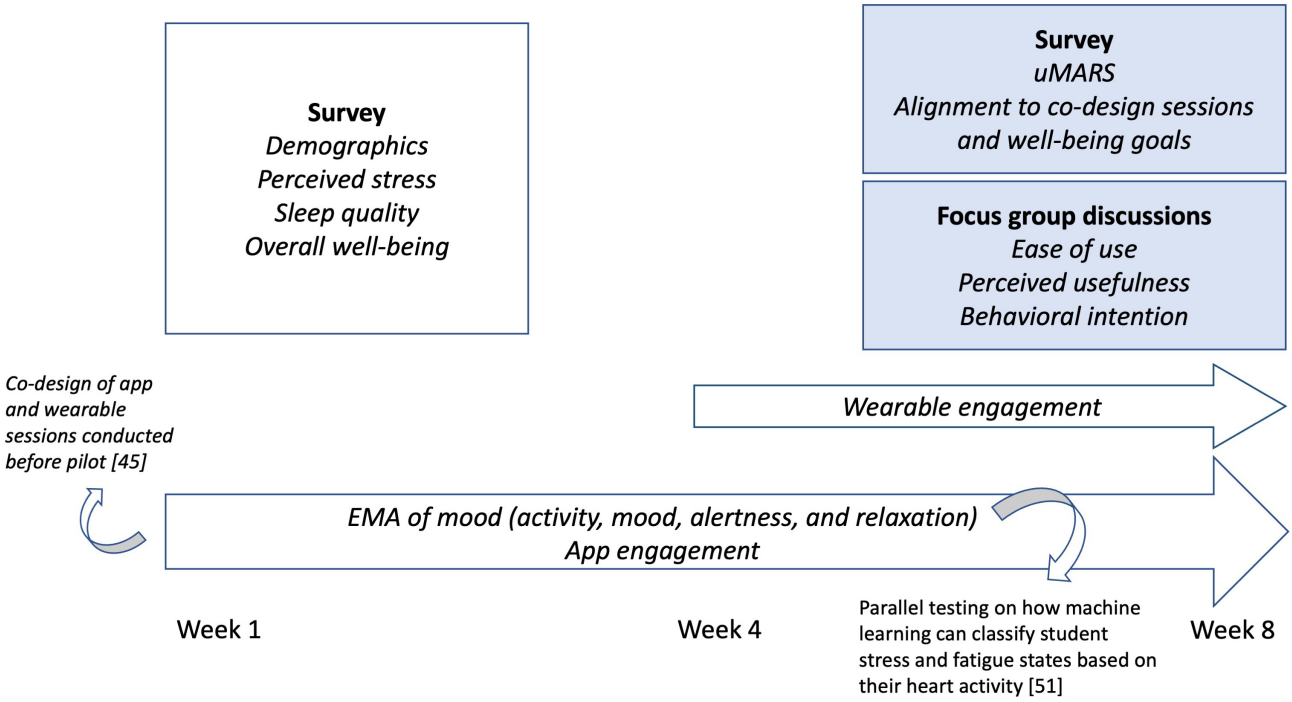
Positive design provided the framework to develop the Wellby system [29], and Radical Participatory Design [16] provided the methodological guidelines to promote continuous collaboration with students throughout the cocreation process. As outlined in a previous work [45], students were involved in a needs assessment and co-design sessions to shape the focus and design of Wellby.

Through the study protocol and cocreation approach, we aimed to reduce the potential risks that are presented by digital tools, such as privacy concerns, overuse, and fixation on negative data [20,50]. Data were encrypted in Google Firebase, and access was reserved to approved data controllers (JL and PD). We established an incident reporting protocol for in-app coaches to notify researchers of messages indicating potential violence or harm to students. During participant onboarding, we demonstrated the capabilities of the technology in person and provided multiple avenues for the students to raise questions or concerns during the study, such as through their teacher or an in-app feedback form. To avoid fixation on negative data, evidence-based text and diagrams were provided in the biofeedback tab about the meaning of HRV metrics, and signal quality metrics were included for each recording to enhance transparency about the reliability of the HRV data.

Additionally, students were encouraged to complete ecological momentary assessments (EMAs) throughout the study and wearable recordings during the final 4 weeks. The EMAs included questions about the student's current activity, mood (represented as an emoji), and alertness and relaxation level on a scale from 1 to 5. The focus on alertness and relaxation was connected to sleep and stress, previously identified as being the well-being priorities of this cohort [45]. Check-in reminders to engage with the app and complete the EMAs were sent at 3 random times during the school week, provided students had enabled notifications. A separate and parallel study was conducted based on the EMAs and wearable recordings from the testing of Wellby. The objective of this parallel study was to understand how machine learning could potentially be used to classify student stress and fatigue states based on their HRV metrics [51]. Figure 3 illustrates the data collection methods and timepoints throughout this study.

XSL•FO
RenderX

Figure 3. Wellby data collection methods outline of the 8-week study, including pre- and postintervention surveys, continuous data collection, focus group discussions, and connected research papers [45,51]. EMA: ecological momentary assessment; uMARS: user version of the Mobile App Rating Scale.



Questionnaire Measures

The postintervention survey included questions about the study's alignment with students' well-being goals and the uMARS questionnaire, providing quantitative data associated with the primary outcome of this study. The uMARS includes 5 subscales and has previously demonstrated an overall internal consistency (Cronbach α) of 0.90 [46]. Similarly, from the responses in our study, the uMARS resulted in good internal consistency (α=0.89). The first 4 subscales of the uMARS focused on engagement, functionality, aesthetics, and information quality, grouped into the "ease of use" of Wellby. The final subscale of the uMARS focused on the perceived impact on awareness, knowledge, attitudes, intention to change, help seeking, and behavior change based on the use of Wellby. Each subscale included questions with responses ranked on a scale from 1 to 5, with 1 being the lowest and 5 being the highest score.

The baseline well-being questionnaires, including the PSS, PSQI, and EPOCH, have been validated for adolescent populations where they demonstrated internal consistencies (Cronbach α) of 0.62, 0.72, and 0.88, respectively [52-54]. Our study demonstrated comparable internal consistency for the PSS (α=0.78), PSQI (α=0.74), and EPOCH (α=0.89). The overall scores for these questionnaires range from 0 to 40 for the PSS, 0‐21 for the PSQI, and 0‐25 for the EPOCH questionnaire, where higher scores indicate greater stress, lower sleep quality, and greater well-being, respectively.

Data Analysis

To analyze the uMARS results, the first 4 subscales were grouped to evaluate the ease of use of Wellby, while the fifth subscale was used to evaluate the behavioral intention and perceived usefulness in supporting student well-being. Survey questions about the alignment of Wellby with the outcomes of previous co-design sessions and students' well-being goals were summarized using descriptive statistics.

Qualitative feedback from the focus group discussions was analyzed by two researchers (JL and EB) using reflexive thematic analysis based on Braun and Clarke's approach [55]. The identified themes were then grouped into ease of use, perceived usefulness, and behavioral intention based on the TAM [42]. Pen profiles were constructed to illustrate these groupings and provide example quotations associated with each subtheme along with an indication of the number of mentions [56]. Focus group participants included students who were able to attend the in-person, postintervention meeting. These took place during the final weeks of the school year. Demographic information of focus group participants was not available for analysis as this was only collected in the onboarding survey.

App engagement was recorded on Google Firebase, which saved data to indicate the app tabs and features that the students interacted with throughout the study. The engagement data and well-being metrics were summarized across each school using descriptive statistics.

Results

Student Well-Being Assessments

The well-being questionnaires administered as part of the onboarding survey are summarized based on demographic categories in Table 1. Notably, the results demonstrate a higher perceived stress for students at school B, the all-female school, and for female participants overall. Similarly, perceived sleep quality was lowest at school B and for female students. The EPOCH assessment of overall adolescent well-being indicated the highest well-being in students at school A and in male students (Table 2).

XSL•FO
RenderX

Table . Student well-being questionnaire results by demographic for the Perceived Stress Scale (PSS), the Pittsburgh Sleep Quality Index (PSQI), and the EPOCH[a] Measure of Adolescent Well-Being.

Category	PSS, mean (SD)	PSQI, mean (SD)	EPOCH, mean (SD)
School			
School A (n=16)	18.5 (4.0)	6.8 (2.3)	16.7 (3.1)
School B (n=11)	24.2 (5.6)	9.5 (4.0)	15.3 (3.2)
School C (n=16)	19.8 (7.8)	8.1 (3.6)	15.1 (4.0)
Age (years)			
16 (n=12)	16.5 (7.4)	7.5 (4.3)	15.9 (4.9)
17 (n=25)	21.2 (4.9)	7.7 (3.0)	15.7 (2.9)
18 (n=3)	25.3 (7.6)	10.0 (3.6)	17.8 (1.0)
19 (n=3)	25.0 (5.6)	9.7 (1.5)	13.7 (2.3)
Gender			
Female (n=27)	23.1 (5.5)	8.7 (3.3)	15.0 (2.7)
Male (n=13)	15.7 (5.6)	5.9 (2.9)	18.0 (2.9)
Nonbinary (n=1)	16 (N/A[b])	13 (N/A)	20 (N/A)
Prefer not to say (n=2)	18.0 (1.4)	8.5 (0.7)	10.1 (7.3)
Race and ethnicity			
Asian (n=3)	26.7 (8.4)	13.3 (3.8)	10.8 (5.2)
Black or Black Irish (n=2)	25.0 (12.7)	12.0 (1.4)	18.4 (2.3)
White/other European (n=7)	17.7 (3.5)	6.6 (2.9)	16.3 (4.1)
White African (n=1)	19 (N/A)	8 (N/A)	15.3 (N/A)
White Irish (n=30)	20.2 (6.1)	7.5 (3.0)	16.0 (3.0)
Total (N=43)	20.4 (6.3)	8.0 (3.4)	15.8 (3.5)

[a]EPOCH: engagement, perseverance, optimism, connectedness, and happiness.

[b]N/A: not available.

Because of the notable gender differences in the well-being questionnaires and the potential confounding effect of the all-female school, results were analyzed by gender within each school. To increase sample size and generalizability, data from the comparison group (who did not test Wellby) were combined with the intervention group for this gender-based analysis.

Only data from female and male students are presented in Table 3 because these groups had sample sizes that enabled the use of the Mann-Whitney *U* tests for schools A and C to compare results from female and male students. School B was excluded from statistical testing due to the absence of male participants. For school A, Mann-Whitney *U* tests revealed no significant gender differences for PSS scores (U=79.5, P=.90, r=0.031), PSQI scores (U=77.5, P=.82, r=0.051), or EPOCH total scores (U=77.5, P=.82, r=0.051). At school C, a significant gender difference was found for PSS scores, with females reporting higher stress than males (U=0, P=.007, r=0.770). While effect sizes suggested greater sleep quality in males, no significant differences were found for PSQI scores (U=6, P=.08, r=0.514) or EPOCH scores (U=9, P=.19, r=0.385).

Students' well-being goals set at the beginning of the study included elements of lifestyle medicine (daily activity, eating well, relationships, sleep, and stress management) and digital well-being, identified by students as relevant well-being needs in the previous co-design sessions. Sleep and stress management were the most chosen categories for the total student selections (Table S1 in Multimedia Appendix 1). Daily activity and eating well were also frequently chosen, particularly at schools A and C.

XSL•FO
RenderX

Table . Student well-being questionnaire results based on gender at each school.

School[a]	PSS[b], mean (SD)		PSQI[c], mean (SD)		EPOCH[d], mean (SD)	
	Female	Male	Female	Male	Female	Male
School A (n=15/11)	19.3 (4.4)	19.2 (3.4)	6.8 (2.0)	7.5 (3.0)	17.3 (3.9)	17.4 (2.7)
School B (n=16/0)	23.8 (6.1)	N/A[e]	9.5 (4.1)	N/A	15.2 (2.9)	N/A
School C (n=9/5)	24.8 (4.4)	15.3 (3.0)	9.3 (2.5)	5.5 (3.0)	14.3 (1.9)	16.7 (3.1)
Total (n=40/15)	22.4 (5.4)	18.1 (3.6)	8.3 (3.1)	6.9 (3.1)	15.9 (3.4)	17.2 (2.7)

[a]Participant numbers presented as n=female/male.

[b]PSS: Perceived Stress Scale.

[c]PSQI: Pittsburgh Sleep Quality Index.

[d]EPOCH: EPOCH Measure of Adolescent Well-Being, which assesses 5 positive psychological characteristics (engagement, perseverance, optimism, connectedness, and happiness).

[e]N/A: not applicable, all-female school.

Survey Data

Survey Overview

At the end of the 8-week study, students (n=29) completed the postintervention survey, aligning with the final weeks of their school year, which may have contributed to the 67% (29/43) response rate.

Ease of Use

On the first 4 uMARS subscales, the mean of the total scores was 4.1 (SD 0.7). School B had the highest average score for each subscale, 4.0 (SD 0.9) for engagement and 4.4 for functionality (SD 0.3), aesthetics (SD 0.8), and information quality (SD 0.7), compared to the other schools. Based on the student responses grouped across all schools, information quality had the highest overall score of 4.2 (SD 0.6), while engagement had the lowest overall score of 3.8 (SD 0.7). The functionality, aesthetic, and information quality subscales all reported average scores greater than 4.0 for the grouped responses (Table 4).

Table . The uMARS[a] "ease of use" subscale scores by school site[b].

uMARS subscale	School A (n=6)	School B (n=7)	School C (n=16)	Total (n=29)
Engagement	4.0 (0.7)	4.0 (0.9)	3.7 (0.7)	3.8 (0.7)
Functionality	4.3 (0.4)	4.4 (0.3)	3.9 (0.6)	4.1 (0.6)
Aesthetic	3.8 (0.9)	4.4 (0.8)	4.2 (0.8)	4.1 (0.8)
Information quality	4.2 (0.4)	4.4 (0.7)	4.2 (0.7)	4.2 (0.6)

[a]uMARS: user version of the Mobile App Rating Scale.

[b]Data are presented as mean (SD).

Perceived Usefulness and Behavioral Intention

On the final uMARS subscale, the perceived usefulness of Wellby was indicated by the highest scores of 3.9 (0.8) and 3.8 (1.0) for increased help seeking and well-being awareness, respectively (Table 5). Regarding behavioral intention, uMARS results indicated moderate levels of intention to change (mean 3.7, SD 1.0) and anticipated behavior change (mean 3.6, SD 1.2) because of Wellby.

Of the student respondents who indicated prior participation in the co-design sessions for Wellby (n=24), 88% (21/24) and 83% (20/24) felt that the wearable and the app, respectively, aligned "extremely well" or "somewhat well" with their feedback from co-design sessions. Remaining responses indicated students felt "neutral" about the alignment of Wellby with the co-design session outcomes. Responses from all students, including those not involved in the co-design sessions, indicated that a majority (22/29, 76%) felt Wellby aligned "extremely well" or "somewhat well" with their well-being needs (Table S2 in Multimedia Appendix 1).

XSL•FO
RenderX

Table . The uMARS[a] perceived impact scores at each school site[b].

uMARS perceived impact questions	School A (n=6)	School B (n=7)	School C (n=16)	Total (n=29)
Awareness: Wellby has increased my awareness of the importance of addressing aspects of my well-being	3.5 (1.0)	3.9 (1.3)	3.9 (0.8)	3.8 (1.0)
Knowledge: Wellby has increased my understanding of my well-being	3.2 (1.0)	3.6 (1.1)	3.8 (0.9)	3.6 (1.0)
Attitudes: Wellby has changed my attitudes toward improving my well-being	3.2 (1.6)	3.6 (1.1)	3.9 (0.9)	3.7 (1.1)
Intention to change: Wellby has increased my motivation to address certain aspects of my well-being	3.7 (0.8)	3.6 (1.5)	3.9 (0.9)	3.7 (1.0)
Help seeking: Wellby would encourage me to seek further well-being support (if I needed it)	4.2 (0.8)	3.6 (0.8)	3.9 (0.8)	3.9 (0.8)
Behavior change: Use of Wellby will increase aspects of my well-being	3.0 (0.9)	3.9 (1.3)	3.6 (1.2)	3.6 (1.2)

[a]uMARS: user version of the Mobile App Rating Scale.

[b]Data are presented as mean (SD).

Focus Group Feedback

Focus Group Overview

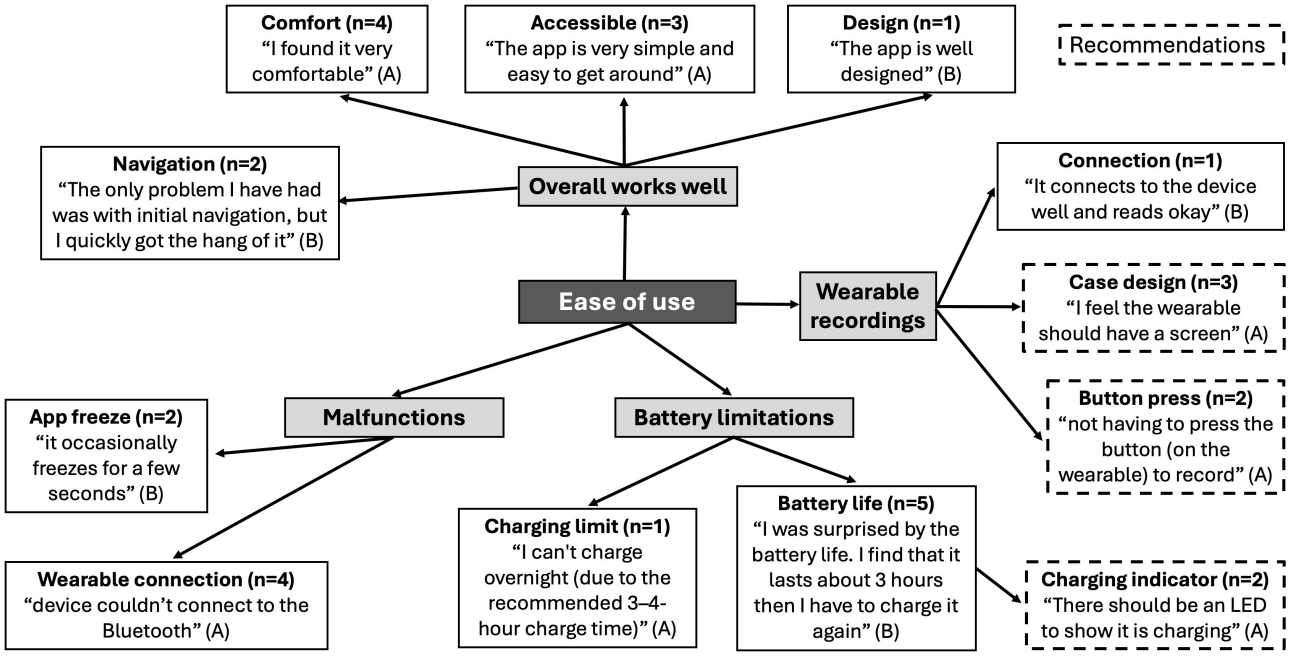
Structured focus groups were conducted in person at the end of the intervention with students at school A (n=6) and school B (n=5). Due to time constraints at school C, less structured feedback from students was gathered, which is summarized under the “ease of use” section. Themes from the discussion were grouped based on TAM categories, and subthemes are detailed using the number of mentions for each theme, a student quotation, and a school indicator for each quote.

Ease of Use

There was an overall positive response to the ease of use of Wellby, with many students noting that the wearable was comfortable and that the app was “easy to get around” (school A) (Figure 4). Some students noted that it took time to initially understand the app navigation. While one student noted that the app “connects to the device well” (school B), others (n=4) noted challenges with Bluetooth connection, particularly on the Android version of the mobile app. Other limitations included occasional app freezing and battery life. Several students noted that the battery life was not long enough (n=5), and one student specified that they would like to charge it overnight but were unable to due to the recommended 3 - 4 hour charging time. Students also recommended ways to improve the ease of use of Wellby by including a light indicator for different charging stages on the wearable and making it possible to start a heart activity recording without having to press the button on the wearable. Additionally, some students (n=3) noted that they would like something on the face of the wearable, such as a screen or clock. Feedback from school C indicated that students didn’t see the wearable as socially acceptable with their school’s fashion norms. They recommended changes including making the device rounder, having a shiny or metallic material, and including more color options.

XSL•FO
RenderX

Figure 4. Student focus group feedback on the "ease of use" of Wellby grouped into major themes. The subthemes include the frequency (n), an example quote, and the school associated with each student quote (represented by A, B, or C). Subthemes that were recommendations from students are indicated with dashed borders.



Perceived Usefulness

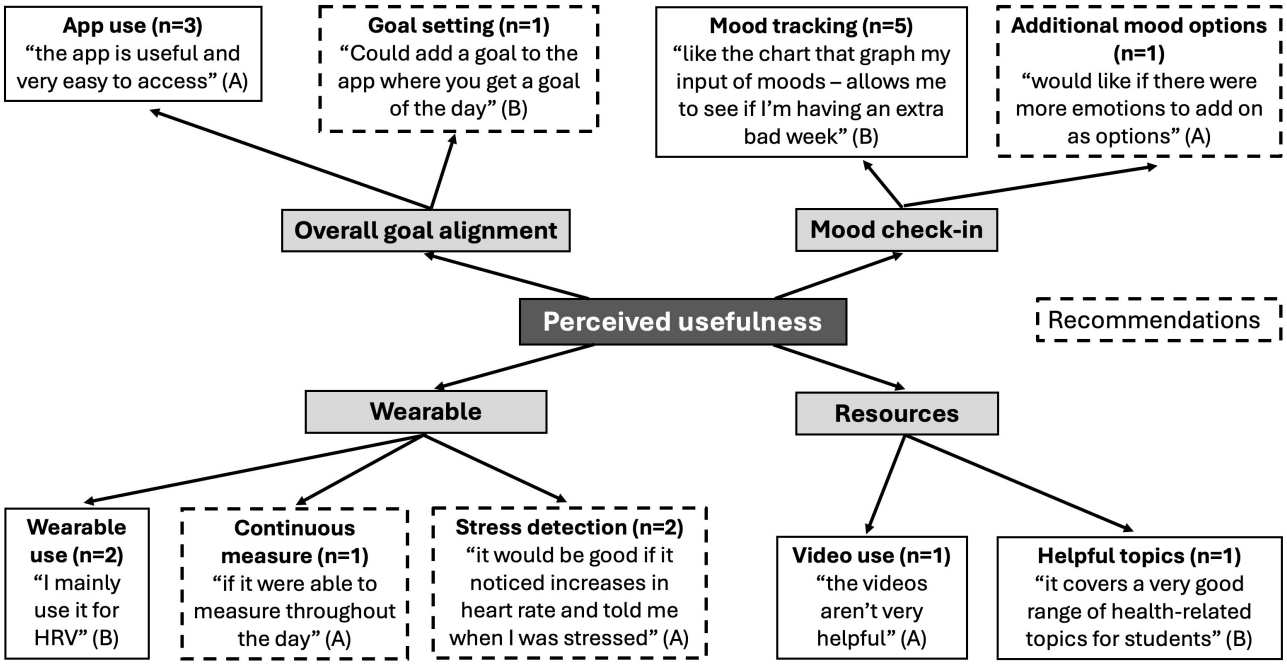
Students indicated an overall alignment of Wellby with their well-being goals (Figure 5). One student recommended adding a feature to the app to specify and track student goals. The most mentioned app feature was the mood tracker, one student sharing that it "allows [them] to see if [they're] having an extra bad week" (school B) and another that they "like the check-in feature" and "spend ages looking through the emojis to choose [their] mood" (school B). To further enhance the mood check-in, a student recommended adding more options to this feature so that they can input more details about their current moods.

Students had mixed opinions about the usefulness of app resources, with one student noting that they "cover a very good range of health-related topics for students" (school B), while another noted that the video resources "aren't very helpful" (school A).

Students also noted feedback about the wearable, one describing that they mainly use Wellby to track their HRV. They also noted that they would like for the wearable to be able to monitor their heart activity continuously throughout the day and notify them of potential increases in stress levels.

XSL•FO
RenderX

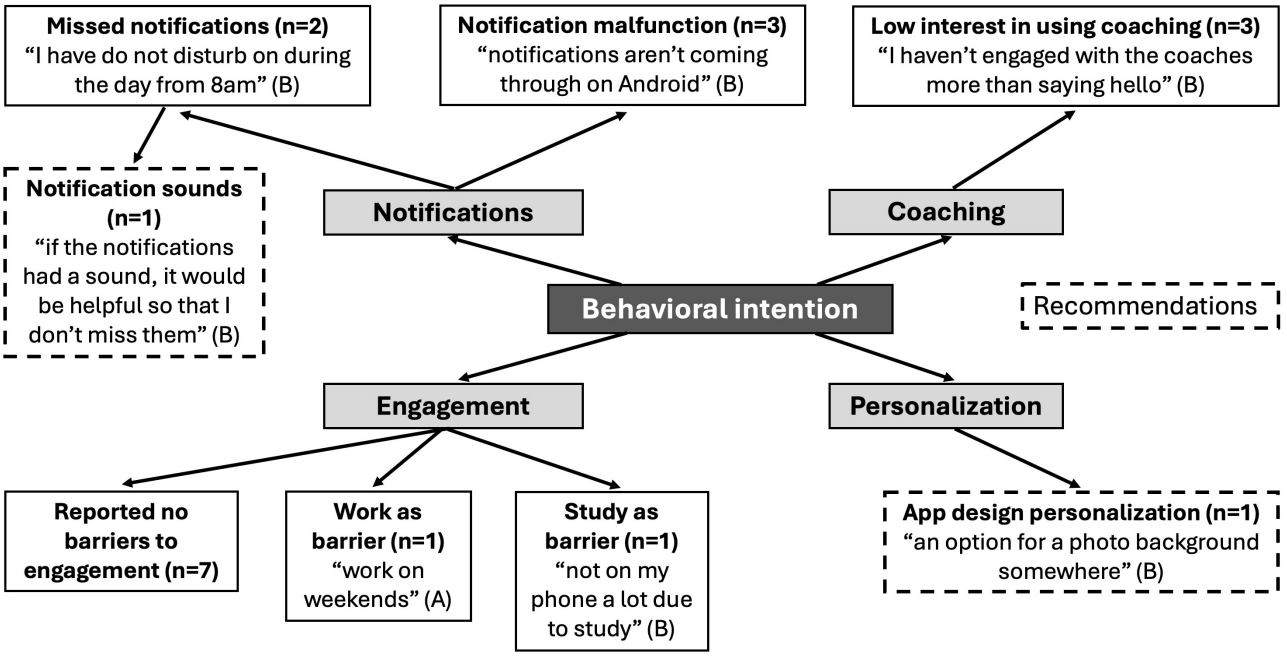
Figure 5. Student focus group feedback on the "perceived usefulness" of Wellby grouped into major themes. The subthemes include the frequency (n), an example quote, and the school associated with each student quote (represented by A, B, or C). Subthemes that were recommendations from students are indicated with dashed borders.



Behavioral Intention

Student feedback from the focus group also included barriers and facilitators of students' engagement with Wellby. Several students (n=3) noted that they had little interest in messaging the health coach provided through the app (Figure 6). Other students missed notifications due to focus features on their phones or malfunctions on Android. One student recommended adding sounds to notifications so that they remember to engage with the app. When asked if there were external factors preventing them from using Wellby, one student indicated that they had work on the weekends, another noted their focus on studying, while most (n=7) indicated no external barriers to engagement. Students also commented on the benefits of personalizing Wellby, such as the app background color or wearable strap material. One student recommended enhanced personalization through choosing custom background images for the app.

Figure 6. Student focus group feedback on the "behavioral intention" of Wellby grouped into major themes. The subthemes include the frequency (n), an example quote, and the school associated with each student quote (represented by A, B, or C). Subthemes that were recommendations from students are indicated with dashed borders.



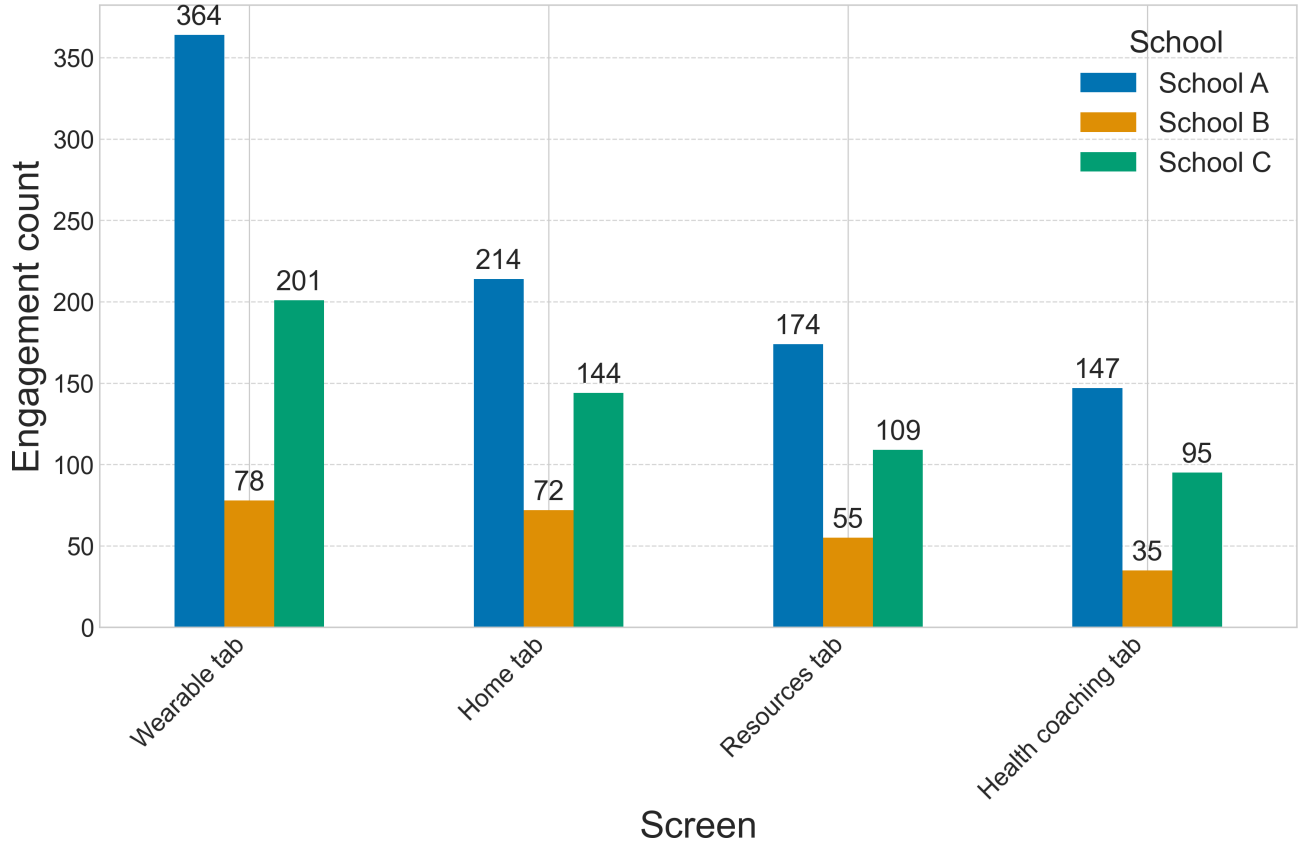
XSL•FO
RenderX

Wellby Engagement Metrics

App engagement data throughout the 8 weeks was summarized based on the total engagement with the 4 app tabs, separated by school, and the overall engagement with each app feature. The wearable tab had the highest engagement, followed by the home tab, resource tab, and coaching tab, for each school. Engagement was highest at schools A and C, which had 5 more students involved in the testing than school B (Figure 7). The app features that allowed students to customize the background color on the app had the highest engagement, with 417 total clicks throughout the study period. The following 4 features with the most engagement included completing mood check-ins, viewing specific resource topics, recording HRV, and viewing HRV summary (Table S3 in Multimedia Appendix 1).

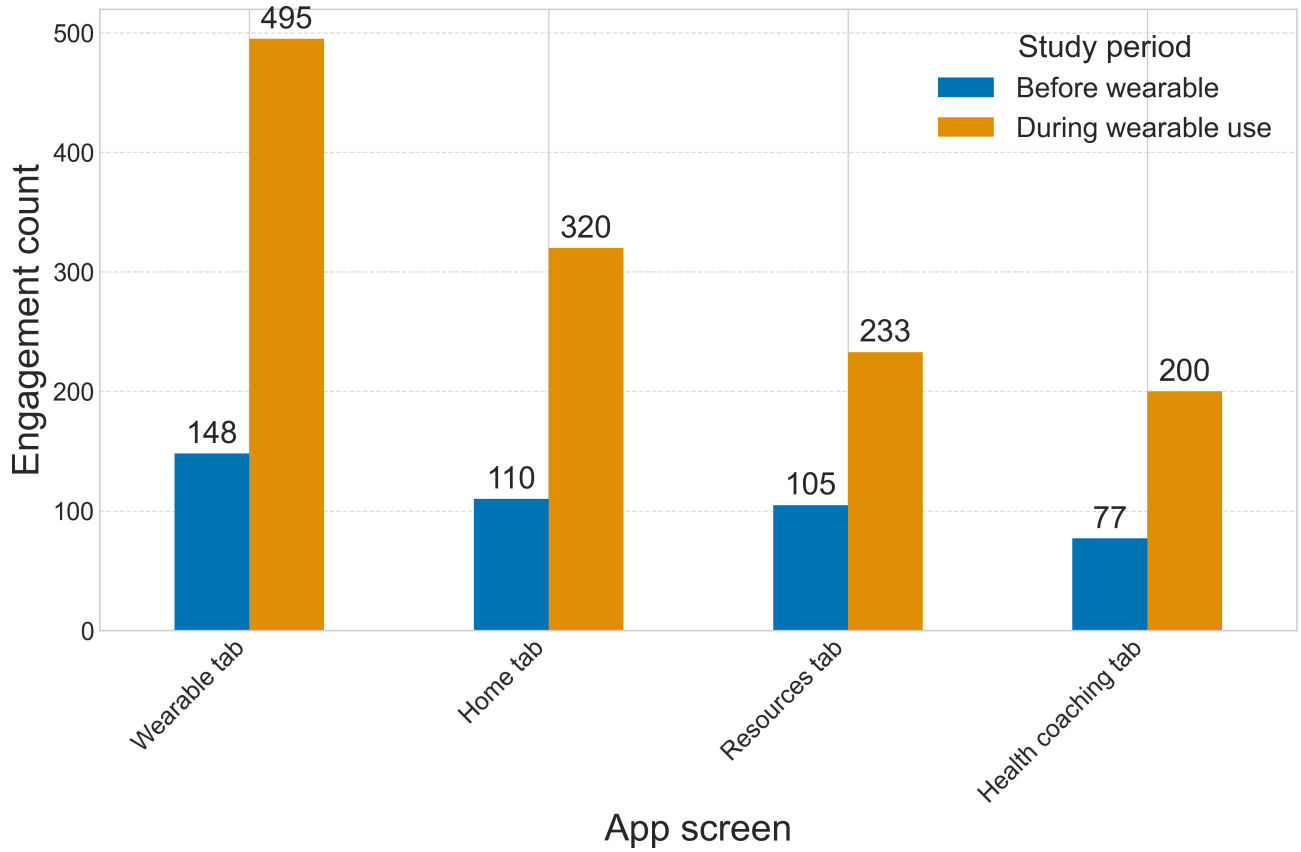
Since students had access to the mobile app alone for the first 4 weeks of the study before being given wearables, the engagement was analyzed during both time periods. Student engagement with each app tab increased substantially, ranging from 123 to 347 additional interactions per tab, during the final 4 weeks of the study when they had access to the wearable device (Figure 8). The wearable tab included access to HRV data summaries, a wearable connection to begin a new recording, information on interpreting HRV, and a customizable breath pacer.

Figure 7. Student engagement counts with the Wellby app represented across each school and grouped by app tab.



XSL•FO
RenderX

Figure 8. Student engagement counts with the Wellby app tabs during the first 4 weeks of the study (before wearable access) compared to the final 4 weeks (when students had access to the wearable device).



Discussion

Principal Findings

This mixed methods acceptability and usability study demonstrated the successful alignment of the co-designed Wellby intervention with the needs and preferences of Irish secondary school students. Additionally, secondary baseline well-being measures captured student well-being needs.

The identification of stress and sleep as the most selected well-being goals aligned with the baseline questionnaire results that indicated high stress and low sleep quality in student participants. The mean overall PSS score (mean 20.4, SD 6.3) was substantially higher than scores reported from a similar cohort in Greece (mean 12.6, SD 8.1) [57], while the mean overall PSQI score (mean 8.0, SD 3.4) was above the 6.0 cutoff for poor sleep in young adults [58]. Additionally, the results of the well-being questionnaires varied across demographics and schools, with the highest perceived stress and lowest sleep quality in female students and at the all-female school. This is consistent with stress-related gender differences identified in similar studies of adolescents using digital health tools [59]. Our findings highlight the context-specific well-being needs that exist both between and within different school environments, echoing the results of the needs assessment and co-design phase of Wellby [45]. The gender differences observed from the well-being questionnaires at school C may indicate that the needs of Youthreach students when they join the program differ between genders. Future research could further examine the well-being and acceptability differences among demographics, including potential gender-specific differences, and how personalized features can enhance the adaptability of digital tools to meet a range of student needs.

The high uMARS scores and strong alignment with students' goals likely reflect Wellby's close adherence to student input during co-design. The overall average score for the first 4 subscales in the uMARS (mean 4.1, SD 0.7) is comparable to the uMARS evaluations of the highest rated available mental health apps (mean 4.1 - 4.3) assessed in a review by Ko and Woo [60], compared to those with the lowest ratings (mean 2.9 - 3.3) out of the 41 eligible apps reviewed on the same 4 subscales. The results also revealed that the alignment of Wellby with students' goals was higher for those in the co-design sessions (21/24, 88%) compared to students overall (22/29, 76%). Additionally, students shared recommendations for adjusting the wearable appearance to match their style and for additional features, such as goal tracking. These requested improvements to Wellby may explain why the alignment responses were not higher. Overall, these findings add to existing evidence on the effectiveness of participatory design processes in producing tools that are appropriate and acceptable for adolescents [17].

Barriers to engagement that emerged in the focus groups included technical challenges and environmental constraints such as schoolwork and nonacademic commitments. While technological issues are common barriers in mobile health research, there is a need to more thoroughly examine the system-level influences on students' engagement with digital well-being tools [61]. Researchers might investigate these by

XSL•FO
RenderX

working with students' guardians, teachers, and school staff members to gain a wider perspective on the implications and considerations for implementing digital well-being supports in an educational context. Together they could explore how to integrate these tools into schools' existing support pathways for students such as through school counselors, physical education, or health teachers. The financing for these supports could be allocated through governmental budgets for school well-being resources, which schools can then allocate to the most appropriate resources for their students. For example, the national health service in Ireland provides free digital cognitive behavioral therapy through various referral pathways, and Harty et al [62] demonstrated how this enables digital well-being support tools to be distributed at scale.

The recurring themes related to challenges with the study technology also have important implications for participant engagement. Technological barriers noted by students, such as the short battery life of the wearable and Android compatibility errors, likely reduced sustained engagement. This was underscored by the engagement subscale having the lowest "ease of use" score on the uMARS. Several feedback items from students were also recommendations for features that could increase the interactivity and student interest in Wellby, such as goal trackers, updates to the mood tracking and wearable feedback, and increased options for personalization.

Students additionally expressed their enjoyment in customizing the wearable case and strap, potentially increasing feelings of ownership and trust. While this study aimed to foster collaboration with students, there are several ethical considerations when working with adolescents [30], and technical issues can reduce trust in these types of digital interventions [63]. Notably, previous research demonstrated that trust in technology is a crucial driver of acceptability, particularly for adolescents [64]. Therefore, potential technical issues should be identified early, for example, through short-term tests with students before longer interventions. Future research could further investigate how to promote trust among participants through selected features and collaborative approaches to these interventions.

Student engagement patterns with Wellby indicated that app color personalization, mood check-ins, viewing educational resources, and HRV recordings were the most common features used during this study. This agrees with qualitative feedback indicating that students were most interested in personalized features, the mood check-in, and the wearable component of Wellby. Despite conflicting opinions on the usefulness of the in-app resources in focus groups, the information quality achieved the highest score (mean 4.2, SD 0.6) out of the uMARS subscales. Along with students' frequent engagement with the educational resources, this high uMARS score suggests that several students found the evidence-based well-being content to be relevant and beneficial to them.

After students received wearable devices at the study midpoint, engagement increased across all Wellby app tabs. Although technical challenges with the wearable were reported in focus group discussions, students emphasized their interest in the wearable and in features for future iterations. The engagement increases following wearable implementation, alongside qualitative feedback and moderate uMARS scores for intention to change and behavior change, suggests that the devices were a main driver of student engagement and demonstrate the potential of this tool to encourage behavior change in adolescents. Rather than providing an additional mobile app alone, wearables offer a tangible representation and reminder of students' commitment to enhancing their well-being. Wearable devices may provide valuable tools to catalyze student behavior change and the establishment of habits that can support them as they grow older, particularly by providing a device that can bridge the digital and physical sides of many adolescents' experiences.

Student engagement patterns and feedback also indicated students' overall interest in self-tracking by recording their heart activity and mood over time. This suggests that adolescents may prefer self-directed wellness approaches over guided interventions like coaching or prescriptive content, potentially connecting to their developmental needs for greater autonomy and identity formation. As self-tracking tools become more widely accessible and used, there is a need to provide targeted and mindful technologies that offer tools for adolescents' inclination toward self-exploration while avoiding the potential for excessive self-monitoring or overreliance on digital tools [65].

Students provided recommendations for the iterative development of Wellby, including increased stress-tracking features, mood-tracking options, and customizable features. This aligns with a main recommendation from a similar co-design study of European adolescents that emphasized the importance of a personalized interface [36]. Social acceptability is another important consideration for this cohort, which emerged in this study, and must be considered for adolescent-centered technologies, particularly wearable technologies that are highly visible. Understanding the psychosocial impacts of technologies can be aided using scales such as the Wearable Acceptability Scale, which offer ways to measure the social acceptability of wearables [66]. Mindful feature design and evaluation helped students to gain an increased sense of ownership and autonomy when using the Wellby system, aligning with the overall goal of cocreating Wellby with students.

Limitations

One limitation of this study is the short duration. A longer study with follow-up assessments could identify any novelty effects associated with the introduction of a wearable device and capture sustained engagement and actions resulting from students' behavioral intention beyond 8 weeks. Additionally, the custom wearable device experienced technical difficulties, including battery limitations, Bluetooth connectivity issues, and Android-specific bugs, which may have influenced students' perceptions of acceptability. These highlight the challenge of developing custom technologies for adolescents and managing students' expectations. The sample size, while meeting recommended guidelines for usability studies, was relatively small with 43 participants across 3 schools, limiting the generalizability of findings to broader adolescent populations

XSL•FO
RenderX

and the analysis of subgroup differences. The postintervention survey response rate of 67% (29/43) and the small focus group sizes should also be considered when interpreting findings, as this may introduce response bias toward students who were more engaged with the study. Future studies should consider timing intervention endpoints when students are more likely to engage by collaborating with school staff to avoid busy academic periods.

The substantial overlap of students involved in both the co-design sessions and intervention testing (38/43, 88% of participants) presents another potential limitation and methodological consideration. This overlap may have introduced familiarity bias due to students' prior interest and knowledge in the intervention, which could have inflated the acceptability scores. However, this also enabled assessment of participatory fidelity by gathering feedback on whether the intervention reflected students' input from the co-design session and revealed higher alignment of Wellby with the well-being goals of this cohort compared to students not involved in co-design sessions. When designing future interventions that include participatory design methods, researchers should consider the impacts of prior participatory engagement on the intervention outcomes. Involving the same participants throughout may be appropriate for cocreation studies focused on iterative design and fidelity [67]. However, for studies focused on the generalizability of intervention efficacy and usability, researchers could consider separating participants into cohorts based on prior engagement to control for familiarity bias and enable comparison of acceptability and usability outcomes between these groups.

Future studies should aim to capture a diverse adolescent population by recruiting from different geographic and socioeconomic areas. For example, in Ireland, this could include schools identified by the Delivering Equality of Opportunity in Schools program in both rural and urban areas. Additionally, researchers should collect comprehensive demographic data, such as socioeconomic indicators, to contextualize the outcomes and promote generalizability and reproducibility of their findings. Finally, as this study was conducted in the Irish secondary school context, the findings may not be directly transferable to other cultural or educational environments without consideration of local contexts and adolescent well-being needs.

Conclusions

This acceptability and usability study provides evidence that co-designed digital well-being tools can achieve high levels of user acceptance among Irish secondary school students when developed through participatory processes. The strong alignment between Wellby and students' well-being goals, as shown by both quantitative uMARS scores and qualitative feedback, supports the value of involving adolescents as equal partners throughout the design and evaluation process. Students particularly valued self-tracking features including mood tracking and HRV monitoring features, while they engaged less with educational resources and text-based coaching. This was underscored by the considerable increases in app engagement following the introduction of the wearable. This reflects the importance of creating adolescent-centered tools that consider developmental needs such as increased autonomy and identity formation in this age group. Despite some technical limitations, the TAM framework effectively captured students' experiences and highlighted the importance of ease of use, perceived usefulness, and behavioral intention in capturing the acceptance and usability of digital health interventions.

As adolescents increasingly avail themselves of technologies to aid in self-expression, identity formation, and habit building, these technologies become intertwined with their well-being journeys. Consequently, the design choices and features offered by these technologies influence the ability of adolescents to pursue their unique well-being goals and navigate challenges throughout adolescence. Participatory engagement of relevant stakeholders, well-being–informed design, and iterative testing can help to ensure that new technologies provide adolescents with appropriate support. In exploring the acceptability and usability of Wellby, this study highlighted the value that young people have for personalizable and aesthetic technology, features for self-exploration, and the ability to develop a sense of ownership and autonomy through these tools.

The presented findings provide insights for developers and researchers working to create adolescent-centered digital well-being tools. In addition to actionable student insights for Wellby and similar technologies, this mixed methods study contributes to the growing evidence base supporting participatory design methodologies in creating relevant technologies that can help adolescents to thrive.

Acknowledgments

The authors would like to acknowledge the students for their valuable time and contributions, which have been instrumental in shaping this research project. Additionally, the teachers (Anne Smyth, Elaine Brady, Brendan Boner, Marie O'Grady, Christine Hughes, Colm Feighery, Liam McCotter, Luke Coyne Curry, and Eimear O'Rourke) at each participating school contributed a substantial amount of work to coordinate these research activities, providing their insights and frequently communicating with the students.

Funding

This research was funded as part of a PhD scholarship partnership between the U.S. Fulbright Program and the Royal College of Surgeons in Ireland (RCSI; to JL) and the RCSI Strategic Academic Recruitment Start-Up Lectureship fund (to CC) to support PhD scholarship (to SJ).

XSL•FO
RenderX

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the vulnerable nature of the participant cohort. Summary data presented in this paper are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: JL (lead), PD (equal), EB (equal)
Data curation: JL
Formal analysis: JL (lead), EB (supporting)
Funding acquisition: PD (lead), JL (equal), CC (supporting)
Investigation: JL
Methodology: JL (lead), PD (equal), EB (equal)
Project administration: JL (lead), PD (equal), EB (equal)
Resources: JL (lead), SJ (supporting), JM (supporting)
Software: JL (lead), SJ (supporting)
Supervision: PD (lead), EB (equal)
Validation: JL (lead), EB (supporting)
Visualization: JL
Writing – original draft: JL
Writing – review & editing: JL (lead), PD (supporting), EB (supporting), SJ (supporting), JM (supporting), CC (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional data tables outlining student well-being goals, survey responses, and app engagement counts.
[DOCX File, 22 KB - humanfactors_v13i1e79381_app1.docx]

References

1. Giovanelli A, Ozer EM, Dahl RE. Leveraging technology to improve health in adolescence: a developmental science perspective. J Adolesc Health 2020 Aug;67(2S):S7-S13. [doi: 10.1016/j.jadohealth.2020.02.020] [Medline: 32718517]
2. Freeman JL, Neff G. The challenge of repurposed technologies for youth: understanding the unique affordances of digital self-tracking for adolescents. New Media Soc 2023 Nov;25(11):3047-3064. [doi: 10.1177/14614448211040266]
3. Sawyer SM, Afifi RA, Bearinger LH, et al. Adolescence: a foundation for future health. Lancet 2012 Apr 28;379(9826):1630-1640. [doi: 10.1016/S0140-6736(12)60072-5] [Medline: 22538178]
4. Patton GC, Sawyer SM, Santelli JS, et al. Our future: a Lancet commission on adolescent health and wellbeing. Lancet 2016 Jun 11;387(10036):2423-2478. [doi: 10.1016/S0140-6736(16)00579-1] [Medline: 27174304]
5. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016;13(2):e1001953. [doi: 10.1371/journal.pmed.1001953]
6. Punukollu M, Marques M. Use of mobile apps and technologies in child and adolescent mental health: a systematic review. Evid Based Ment Health 2019 Nov;22(4):161-166. [doi: 10.1136/ebmental-2019-300093] [Medline: 31358537]
7. Wong CA, Madanay F, Ozer EM, et al. Digital health technology to enhance adolescent and young adult clinical preventive services: affordances and challenges. J Adolesc Health 2020 Aug;67(2S):S24-S33. [doi: 10.1016/j.jadohealth.2019.10.018] [Medline: 32718511]
8. Ferretti A, Hubbs S, Vayena E. Global youth perspectives on digital health promotion: a scoping review. BMC Digit Health 2023;1(1):25. [doi: 10.1186/s44247-023-00025-0]
9. Macintyre AK, Torrens C, Campbell P, et al. Socioeconomic inequalities and the equity impact of population-level interventions for adolescent health: an overview of systematic reviews. Public Health 2020 Mar;180:154-162. [doi: 10.1016/j.puhe.2019.11.008] [Medline: 31923881]
10. Viner RM, Ozer EM, Denny S, et al. Adolescence and the social determinants of health. Lancet 2012 Apr 28;379(9826):1641-1652. [doi: 10.1016/S0140-6736(12)60149-4] [Medline: 22538179]
11. Maharjan SM, Poudyal A, van Heerden A, et al. Passive sensing on mobile devices to improve mental health services with adolescent and young mothers in low-resource settings: the role of families in feasibility and acceptability. BMC Med Inform Decis Mak 2021 Apr 7;21(1):117. [doi: 10.1186/s12911-021-01473-2] [Medline: 33827552]
12. Clement A, Ravet M, Stanger C, Gabrielli J. Feasibility, usability, and acceptability of MobileCoach-Teen: a smartphone app-based preventative intervention for risky adolescent drinking behavior. J Subst Use Addict Treat 2024 Apr;159:209275. [doi: 10.1016/j.josat.2023.209275] [Medline: 38110119]

XSL•FO
RenderX

13. Coulter RWS, Dougherty M, Clark MB, Wilson A, Miller E, Ragavan MI. Online human-centered design methods are acceptable, appropriate, and feasible for generating adolescent relationship abuse intervention ideas with sexual and gender diverse youth. J Adolesc Health 2024 Oct;75(4):656-664. [doi: 10.1016/j.jadohealth.2024.06.001] [Medline: 39066750]
14. Reen GK, Muirhead L, Langdon DW. Usability of health information websites designed for adolescents: systematic review, neurodevelopmental model, and design brief. J Med Internet Res 2019 Apr 23;21(4):e11584. [doi: 10.2196/11584] [Medline: 31012856]
15. Müssener U, Thomas K, Linderoth C, et al. Development of an intervention targeting multiple health behaviors among high school students: participatory design study using heuristic evaluation and usability testing. JMIR Mhealth Uhealth 2020 Oct 29;8(10):e17999. [doi: 10.2196/17999] [Medline: 33118942]
16. Udoewa V. An introduction to radical participatory design: decolonising participatory design processes. Des Sci 2022;8. [doi: 10.1017/dsj.2022.24]
17. Malloy J, Partridge SR, Kemper JA, Braakhuis A, Roy R. Co-design of digital health interventions with young people: a scoping review. Digit Health 2023;9:20552076231219117. [doi: 10.1177/20552076231219117] [Medline: 38107986]
18. Youth-centered digital health interventions: a framework for planning, developing and implementing solutions with and for young people. : World Health Organization; 2020 Jul 19 URL: https://www.who.int/publications/i/item/9789240011717 [accessed 2026-03-19]
19. Vargas C, Whelan J, Brimblecombe J, Allender S. Co-creation, co-design, co-production for public health – a perspective on definition and distinctions. Public Health Res Pract 2022 Jun 15;32(2):3222211. [doi: 10.17061/phrp3222211] [Medline: 35702744]
20. Hollis C, Falconer CJ, Martin JL, et al. Annual research review: digital health interventions for children and young people with mental health problems – a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
21. Orben A, Dienlin T, Przybylski AK. Social media's enduring effect on adolescent life satisfaction. Proc Natl Acad Sci USA 2019 May 21;116(21):10226-10228. [doi: 10.1073/pnas.1902058116]
22. Primack BA, Shensa A, Escobar-Viera CG, et al. Use of multiple social media platforms and symptoms of depression and anxiety: a nationally-representative study among U.S. young adults. Comput Human Behav 2017 Apr;69:1-9. [doi: 10.1016/j.chb.2016.11.013]
23. Vannucci A, Flannery KM, Ohannessian CM. Social media use and anxiety in emerging adults. J Affect Disord 2017 Jan 1;207:163-166. [doi: 10.1016/j.jad.2016.08.040] [Medline: 27723539]
24. De D, El Jamal M, Aydemir E, Khera A. Social media algorithms and teen addiction: neurophysiological impact and ethical considerations. Cureus 2025 Jan;17(1):e77145. [doi: 10.7759/cureus.77145] [Medline: 39925596]
25. Lau N, O'Daffer A, Colt S, et al. Android and iPhone mobile apps for psychosocial wellness and stress management: systematic search in app stores and literature review. JMIR Mhealth Uhealth 2020 May 22;8(5):e17798. [doi: 10.2196/17798] [Medline: 32357125]
26. Raeside R, Halim N, Partridge S, Redfern J. A review of mobile apps for prevention of chronic disease in adolescents across a range of behaviours. Heart Lung Circ 2020;29(Suppl 2):S357. [doi: 10.1016/j.hlc.2020.09.720]
27. Eisenstadt M, Liverpool S, Infanti E, Ciuvat RM, Carlsson C. Mobile apps that promote emotion regulation, positive mental health, and well-being in the general population: systematic review and meta-analysis. JMIR Ment Health 2021 Nov 8;8(11):e31170. [doi: 10.2196/31170] [Medline: 34747713]
28. Wani C, McCann L, Lennon M, Radu C. Digital mental health interventions for adolescents in low- and middle-income countries: scoping review. J Med Internet Res 2024 Oct 29;26:e51376. [doi: 10.2196/51376] [Medline: 39471371]
29. Riva G, Baños RM, Botella C, Wiederhold BK, Gaggioli A. Positive technology: using interactive technologies to promote positive functioning. Cyberpsychol Behav Soc Netw 2012 Feb;15(2):69-77. [doi: 10.1089/cyber.2011.0139] [Medline: 22149077]
30. Badillo-Urquiola K, Shea Z, Agha Z, Lediaeva I, Wisniewski P. Conducting risky research with teens: co-designing for the ethical treatment and protection of adolescents. Proc ACM Hum Comput Interact 2021;4(CSCW3):1-46. [doi: 10.1145/3432930]
31. Inal Y, Wake JD, Guribye F, Nordgreen T. Usability evaluations of mobile mental health technologies: systematic review. J Med Internet Res 2020 Jan 6;22(1):e15337. [doi: 10.2196/15337] [Medline: 31904579]
32. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
33. Creaser AV, Clemes SA, Costa S, et al. The acceptability, feasibility, and effectiveness of wearable activity trackers for increasing physical activity in children and adolescents: a systematic review. Int J Environ Res Public Health 2021 Jun 8;18(12):6211. [doi: 10.3390/ijerph18126211] [Medline: 34201248]
34. Prochaska JJ, Wang Y, Bowdring MA, Chieng A, Chaudhary NP, Ramo DE. Acceptability and utility of a smartphone app to support adolescent mental health (BeMe): program evaluation study. JMIR Mhealth Uhealth 2023 Aug 28;11:e47183. [doi: 10.2196/47183] [Medline: 37639293]

XSL•FO
RenderX

35. Ridgers ND, Timperio A, Brown H, et al. Wearable activity tracker use among Australian adolescents: usability and acceptability study. JMIR Mhealth Uhealth 2018 Apr 11;6(4):e86. [doi: 10.2196/mhealth.9199] [Medline: 29643054]
36. Martin A, Caon M, Adorni F, et al. A mobile phone intervention to improve obesity-related health behaviors of adolescents across Europe: iterative co-design and feasibility study. JMIR Mhealth Uhealth 2020;8(3):e14118. [doi: 10.2196/14118]
37. Swahn MH, Gittner KB, Lyons MJ, et al. Advancing mHealth research in low-resource settings: young women's insights and implementation challenges with wearable smartwatch devices in Uganda. Sensors (Basel) 2024;24(17):5591. [doi: 10.3390/s24175591]
38. Granić A, Marangunić N. Technology acceptance model in educational context: a systematic literature review. Brit J Educational Tech 2019 Sep;50(5):2572-2593. [doi: 10.1111/bjet.12864]
39. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003 Sep 1;27(3):425-478. [doi: 10.2307/30036540]
40. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Dec;17(1):88. [doi: 10.1186/s12913-017-2031-8]
41. Drehlich M, Naraine M, Rowe K, et al. Using the technology acceptance model to explore adolescents' perspectives on combining technologies for physical activity promotion within an intervention: usability study. J Med Internet Res 2020;22(3):e15552. [doi: 10.2196/15552]
42. Venkatesh V, Bala H. Technology acceptance model 3 and a research agenda on interventions. Decis Sci 2008 May;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
43. Smyth E, Banks J, O'Sullivan J, McCoy S, Redmond P, McGuinness S. Evaluation of the National Youthreach Programme: Research Series 82: Economic and Social Research Institute; 2019.
44. Macefield R. How to specify the participant group size for usability studies: a practitioner's guide. J Usability Stud 2009;5(1):34-45. [doi: 10.5555/2835425.2835429]
45. Laiti J, Donnelly J, Byrne E, Dunne PJ. Co-creating Wellby—a mobile app and wearable for student well-being management guided by a needs assessment and co-design. Front Digit Health 2025;7. [doi: 10.3389/fdgth.2025.1560541]
46. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016;4(2):e72. [doi: 10.2196/mhealth.5849]
47. Cohen S, Kessler RC, Gordon LU, editors. Measuring Stress: A Guide for Health and Social Scientists: Oxford Academic; 1995. [doi: 10.1093/oso/9780195086416.001.0001]
48. Buysse DJ, Reynolds CF III, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4]
49. Kern ML, Benson L, Steinberg EA, Steinberg L. The EPOCH measure of adolescent well-being. Psychol Assess 2016 May;28(5):586-597. [doi: 10.1037/pas0000201] [Medline: 26302102]
50. Gui M, Büchi M. From use to overuse: digital inequality in the age of communication abundance. Soc Sci Comput Rev 2021 Feb;39(1):3-19. [doi: 10.1177/0894439319851163]
51. Laiti J, Liu Y, Dunne PJ, Byrne E, Zhu T. Real-world classification of student stress and fatigue using wearable PPG recordings. IEEE Trans Affective Comput 2025;17(1):587-601. [doi: 10.1109/TAFFC.2025.3628467]
52. Chen W, Liang Y, Cheng J, Xie E, Zhu Q, Liu J. Psychometric properties of the Perceived Stress Scale Short Form (PSS-10) in Chinese adolescents: validity, measurement invariance across gender and left-behind status. Child Youth Serv Rev 2024 Jan;156:107252. [doi: 10.1016/j.childyouth.2023.107252]
53. de la Vega R, Tomé-Pires C, Solé E, et al. The Pittsburgh Sleep Quality Index: validity and factor structure in young people. Psychol Assess 2015 Dec;27(4):e22-e27. [doi: 10.1037/pas0000128] [Medline: 26653055]
54. Buerger S, Holzer J, Yanagida T, Schober B, Spiel C. Measuring adolescents' well-being in schools: the adaptation and translation of the EPOCH measure of adolescent well-being—a validation study. School Ment Health 2023 Jun;15(2):611-626. [doi: 10.1007/s12310-023-09574-1]
55. Byrne D. A worked example of Braun and Clarke's approach to reflexive thematic analysis. Qual Quant 2022 Jun;56(3):1391-1412. [doi: 10.1007/s11135-021-01182-y]
56. Mackintosh KA, Knowles ZR, Ridgers ND, Fairclough SJ. Using formative research to develop CHANGE!: a curriculum-based physical activity promoting intervention. BMC Public Health 2011 Oct 27;11(1):831. [doi: 10.1186/1471-2458-11-831] [Medline: 22032540]
57. Kourmousi N, Kounenou K, Pezirkianidis C, Kalamatianos A, Chrousos GP. Validation of Perceived Stress Scale-10 among Greek middle adolescents: associations between stressful life events and perceived stress. Soc Sci (Basel) 2025;14(6):344. [doi: 10.3390/socsci14060344]
58. Dietch JR, Taylor DJ, Sethi K, Kelly K, Bramoweth AD, Roane BM. Psychometric evaluation of the PSQI in U.S. college students. J Clin Sleep Med 2016 Aug 15;12(8):1121-1129. [doi: 10.5664/jcsm.6050]
59. Wright M, Reitegger F, Cela H, Papst A, Gasteiger-Klicpera B. Interventions with digital tools for mental health promotion among 11-18 year olds: a systematic review and meta-analysis. J Youth Adolesc 2023 Apr;52(4):754-779. [doi: 10.1007/s10964-023-01735-4] [Medline: 36754917]
60. Ko S, Woo H. Users' needs for mental health apps: quality evaluation using the user version of the mobile application rating scale. JMIR Mhealth Uhealth 2025 Jul 4;13:e64622. [doi: 10.2196/64622] [Medline: 40614100]

XSL•FO
RenderX

61. Giebel GD, Speckemeier C, Abels C, et al. Problems and barriers related to the use of digital health applications: scoping review. J Med Internet Res 2023 May 12;25:e43808. [doi: 10.2196/43808] [Medline: 37171838]
62. Harty S, Enrique A, Akkol-Solakoglu S, et al. Implementing digital mental health interventions at scale: one-year evaluation of a national digital CBT service in Ireland. Int J Ment Health Syst 2023 Oct 10;17(1):29. [doi: 10.1186/s13033-023-00592-9] [Medline: 37817270]
63. Ejdys J. Building technology trust in ICT application at a university. Int J Emerg Mark 2018 Nov 29;13(5):980-997. [doi: 10.1108/IJoEM-07-2017-0234]
64. Sawrikar V, Mote K. Technology acceptance and trust: overlooked considerations in young people's use of digital mental health interventions. Health Policy Technol 2022 Dec;11(4):100686. [doi: 10.1016/j.hlpt.2022.100686]
65. Feng S, Mäntymäki M, Dhir A, Salmela H. How self-tracking and the quantified self promote health and well-being: systematic review. J Med Internet Res 2021 Sep 21;23(9):e25171. [doi: 10.2196/25171] [Medline: 34546176]
66. Kelly N, Gilbert S. The WEAR scale: developing a measure of the social acceptability of a wearable device. : Association for Computing Machinery Presented at: CHI'16: CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, CA, USA. [doi: 10.1145/2851581.2892331]
67. Trickett E, Rasmus SM, Allen J. Intervention fidelity in participatory research: a framework. Educ Action Res 2020 Jan 1;28(1):128-141. [doi: 10.1080/09650792.2019.1689833]

Abbreviations

EMA: ecological momentary assessment
EPOCH: engagement, perseverance, optimism, connectedness, and happiness
HRV: heart rate variability
PSQI: Pittsburgh Sleep Quality Index
PSS: Perceived Stress Scale
RCSI: Royal College of Surgeons in Ireland
TAM: technology acceptance model
TFA: theoretical framework of acceptability
uMARS: user version of the Mobile App Rating Scale
UTAUT: unified theory of acceptance and use of technology

Edited by A Kushniruk; submitted 20.Jun.2025; peer-reviewed by A Adepoju, A Knapp; revised version received 07.Jan.2026; accepted 18.Jan.2026; published 09.Apr.2026.

Please cite as:
Laiti J, Javadpour S, Mullen J, Conway C, Byrne E, Dunne P
Evaluation of Wellby, a Cocreated Mobile App and Wearable to Support Stress Management and Overall Well-Being: Mixed Methods Acceptability and Usability Study
JMIR Hum Factors 2026;13:e79381
URL: https://humanfactors.jmir.org/2026/1/e79381
doi:10.2196/79381

© Justin Laiti, Sina Javadpour, Jenna Mullen, Claire Conway, Elaine Byrne, Pádraic Dunne. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 9.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Acuity, Diagnosis, and Community Factors in Virtual and In-Person Treatment: Qualitative Study

Brianna Cerrito[1], MS; Alexa Connors[2], MSW; Amanda Fialk[1,2], PhD; Jamie Xiao[1], BA; Josephine Usow[1], MPH; Frank D Buono[3], PhD

[1]Research & Innovation Department, The Dorm, New York, NY, United States

[2]Clinical Department, The Dorm, New York, NY, United States

[3]Department of Psychiatry, Yale School of Medicine, 300 George Street, New Haven, CT, United States

Corresponding Author:
Frank D Buono, PhD
Department of Psychiatry, Yale School of Medicine, 300 George Street, New Haven, CT, United States

Abstract

Background: The COVID-19 pandemic accelerated the adoption of virtual psychotherapy, making videoconferencing a vital tool for maintaining continuity of care. While virtual sessions offer flexibility and accessibility, concern has also been raised about the impact of virtual care on therapeutic relationships, a sense of community, and the ability to support clients with higher levels of clinical need.

Objective: This study sought to examine how mediums of care (ie, virtual vs in-person) impact the therapeutic experience for clients and providers in critical domains such as community, clinical presentation, and clinicians' use of clinical skills. The following themes were evaluated: (1) How therapeutic relationships and community dynamics differ across virtual and in-person mediums of mental health care? (2) How acuity and diagnosis influence the perception of effectiveness and safety of these mediums?

Methods: This study is a grounded theory design. Participants were intensive outpatient mental health clients and clinicians. Participants were interviewed in focus groups of either 6 - 8 clinicians or clients, assessing for important themes (ie, attendance, engagement, and therapeutic alliance) across the 2 modes of care (ie, virtual vs in-person). A total of 6 structured focus groups were conducted. Thematic analysis was conducted to analyze the transcripts until theoretical saturation was achieved.

Results: Results highlighted the nuanced differences in client and clinician perceptions of relational dynamics and effectiveness of care in virtual versus in-person mental health care. This study revealed, the perspectives of the client and clinician, that in-person treatment fosters more engagement and interpersonal connection (ie, therapeutic alliance and peer-to-peer) than virtual settings. Additionally, in-person care better supports the ability to exercise important clinical skills, especially for highly acuity clients with complex diagnoses.

Conclusions: The results of this study outline how the medium of care (ie, virtual vs in-person) impacts therapeutic experience from the perspective of both client and clinician. It is understood that critical domains such as community, interpersonal relationships, and clinicians' use of clinical skills are affected by the medium of care, and key factors such as clinical presentation (ie, acuity and diagnosis) must be accounted for when choosing a particular mode of care for a client.

(JMIR Hum Factors 2026;13:e82863) doi:10.2196/82863

KEYWORDS

face-to-face; online group psychotherapy; online treatment; qualitative methods; therapeutic relationship; video-delivered psychotherapy; acuity; diagnosis; community

Introduction

Virtual psychotherapy appointments saw an increase during the COVID-19 pandemic, helping to maintain continuity of care without violating social guidelines [1]. Years after the pandemic, virtual psychotherapy is still relevant for many populations, including those in rural areas, of marginalized communities, and with demanding careers, offering flexibility, convenience, and access to care for clients and providers [2,3]. In-home therapy sessions have increased, and some clients experience greater openness and relaxation in therapy, allowing for deeper conversations about their mental health struggles [4]. However, virtual sessions also present additional challenges to therapy, such as disrupting the ability to read body language and pick up on nonverbal cues. For individuals who struggle with activities of daily living (ADLs), a therapist may not be able to thoroughly assess whether a client has showered or brushed their teeth, due to the lack of physical presence [5]. Other

XSL•FO
RenderX

challenges to virtual therapy include concerns about establishing strong therapeutic relationships via online mediums, as well as the ability to care for clients with higher clinical acuity, as these clients may need a more involved approach [6-8].

In therapy, human connection (ie, therapeutic alliance and community-building) is crucial to positive mental health outcomes (ie, depression, anxiety, stress, and suicidal ideation and attempts) [9,10]. Therapeutic alliance is a foundation of psychotherapy, representing the trust and effort maintained by both parties to increase mental well-being [11]. The link between therapeutic alliance and positive outcomes is critical, where the therapeutic relationship acts as a mediator of change in mental health outcomes [12]. Research has also consistently demonstrated the importance of having strong relationships with others, reporting that a sense of community and higher levels of social support are positively correlated with reduced mental health symptoms and improved well-being [10,13]. Social relationships both within and outside treatment spaces are a critical part of the therapeutic process, and the changed nature of relationships over virtual mediums raises new questions into the quality of online relationships, and whether these online relationships may have an impact on mental health outcomes.

Research conducted during and since the COVID-19 pandemic highlights differing perspectives on the nature of virtual relationships in therapeutic settings. Some research has shown that clients are able to create a strong connection in virtual mental health care settings, with reports of no change in the quality of therapeutic relationship [1,14,15]. However, systematic reviews and meta-analyses following the COVID-19 pandemic suggest that teletherapy may hinder therapeutic alliance, as it has been shown to foster a weaker therapeutic alliance when compared to face-to-face settings [16,17]. Other research reports therapists' perceptions of virtual mediums as less effective due to the barrier of a screen and reduced authenticity when providing care [6,18]. The varied responses call for further research on differences between the perceived quality of in-person and virtual therapeutic relationships.

Virtual mediums may also impact a therapist's ability to appropriately treat high acuity clients. In virtual settings, therapists have reported a decreased ability to practice clinical skills and respond in crisis or emergency, which is especially important for high acuity clients, as they need more robust care [8,19]. Research has also shown that clients with higher baseline symptom severity demonstrate greater mental health outcomes as well as shorter length of stay in face-to-face versus remote treatment settings, suggesting that there may be a potential benefit to in-person care for high acuity clients [7,19].

While existing literature emphasizes the importance of in-person connection for therapeutic quality, there is a gap in existing literature on how clinicians and clients perceive key therapeutic factors, such as client acuity, diagnosis, and engagement, within virtual versus in-person treatment contexts. Notably, much of the existing teletherapy literature focuses on symptom outcomes or therapeutic alliance broadly, with limited qualitative examination of how clinicians and clients jointly experience acuity, safety, and clinical skill use across care modalities, particularly in higher acuity settings. This study explored how mediums of care (ie, virtual vs in-person) impact the therapeutic experience for young adult clients and providers in an intensive outpatient mental health setting across critical domains such as community, clinical presentation, and clinicians' use of clinical skills. Given there are benefits and drawbacks across virtual and in-person mediums of care, this paper explored 2 main themes from the perspective of clinicians and their clients. The research questions evaluated the ways in which therapeutic relationships and community dynamics differ across virtual and in-person mediums of mental health care and how acuity and diagnosis influence the perception of effectiveness and safety of these mediums.

Methods

Study Design

This study uses a grounded theory design, a qualitative research method aimed at developing a theory based on the data collected [20]. Initially, grounded theory methods were used to arrange and code the data, allowing the development of tentative categories. Reflexive thematic analysis, rooted in constructivist principles, then facilitated the identification and analysis of patterns across the data [20]. Through iterative coding and constant comparison, grounded theory methods helped in refining these patterns into themes, ultimately leading to an explanatory framework that elucidates the relationships among identified themes. The present researchers sought to develop a theory directly based on the data collected to understand the participants' experiences. The research questions were as follows: How do therapeutic relationships and community dynamics differ across virtual and in-person care? How does acuity of mental health diagnoses influence these perceptions?

Participants

Participants included clinical staff (ie, those with social work degrees, professional counseling degrees, nursing degrees, Master of Arts degrees, mental health counseling degrees, alcohol and substance use degrees, and dietetic degrees) and clients seeking intensive outpatient mental health treatment in New York, NY, and Washington, DC. The study included 32 individuals (14 clients and 18 clinical staff members) broken into 6 homogenous groups, 3 of which were client groups and 3 of which were staff groups.

Groups did not exceed 4 to 6 individuals, and this was intentional to provide ample opportunity for engagement of each respondent in the groups. During the focus groups, the client respondents were in 1 of 2 locations: home or a mental health treatment center. However, all clinicians were on site at the treatment program for focus group interviews.

Client demographics, including age, race or ethnicity, and gender identity, and clinical credentials for staff were consistent with those previously reported in Cerrito et al [5]. From the point of view of the analyses conducted, the distribution of the sociodemographic factors remains important, as the sample was predominantly White (n=12, N=14, 86%), and majority were cisgender male and female (n=11, N=14, 79%), with a smaller

XSL•FO
RenderX

portion identifying as transgender nonbinary (n=3, N=14, 21%), which may limit generalizability.

Ethical Considerations

This study is approved by the institutional review board at Yale School of Medicine (IRB #2000032626). Clinical staff and clients who had sought intensive outpatient mental health treatment since the pandemic provided insight about their experience through qualitative focus groups. All participants provided informed consent prior to participation. Data were deidentified and stored securely to protect participant privacy and confidentiality. No financial compensation was provided for participation.

Data Collection

This research followed the methodology outlined in Cerrito et al [5]. Recruitment occurred via an email invitation outlining the study's aims, the voluntary nature of participation, and the opportunity to contribute to advancing understanding of mediums of mental health care. Data collection occurred between March 20 and April 3, 2023, with participation invitations sent via email to staff and clients. Inclusion criteria were as follows: (1) at least 18 years of age, (2) are a client previously admitted to The Dorm or are a provider employed by The Dorm, and (3) able to read and understand the consent form. Exclusion criteria included the following: (1) are unfit to complete the survey due to medical or psychological constraints and (2) require a legally authorized representative.

In qualitative research, the social context of the participants' lives, time, place, and biography are important [21]. Client participant perspectives were interpreted within a shared social and clinical context, including engagement in intensive outpatient mental health care during the pre- and postpandemic period, while clinical staff participants had experience delivering both in-person and virtual treatment modalities. This was also the subject of attention both during the interview and the analytical process.

The invitation to participate in this study was sent out in emails that included detailed descriptions of the subject of the study, its course, and the purpose of the data being solely for scientific purposes. All the respondents gave informed verbal consent to participate in the study at the start of the focus groups. To minimize interviewer preconceptions, focus groups were conducted by a qualitative researcher trained in neutral, open-ended interviewing techniques that use nonleading prompts. These prompts were consistent with Braun and Clarke's reflexive thematic analysis approach, which emphasizes reflexivity, openness to participant meaning, and minimizing assumptions [22].

The first part of the interview guide was related to attendance and engagement in the 2 modes of care, and the second portion began to explore detailed topics, referring to a list of auxiliary questions, which had a semistructured form. To obtain the best quality of data possible, the researcher asked detailed questions about experiences and reflections related to engagement, friendships, rapport-building, trust, honesty, communication, and the benefits of virtual versus in-person. The questions were open and nonleading, and follow-up items were prewritten to ensure comprehensive dialogue, for example, "Tell me about attendance in virtual care settings."

During the interview, all respondents were consistently encouraged to share their experiences as openly and honestly as possible. The interviewer used nonjudgmental language and flexible pacing to ensure that all respondents had a chance to contribute. Much effort was made to create optimal comfort and safety in the space for this purpose.

Data Analysis

Throughout the data analysis process, the interpretative perspectives of clients and clinical staff were prioritized, consistent with grounded theory methodological assumptions. The analysis began with grounded theory techniques like open coding and constant comparison, which helped in the initial organization of the data [23]. Reflexive thematic analysis was then used to further analyze and report patterns, guided by Braun and Clarke's 6-phase framework [21]. This structured approach included familiarization with data, generating initial codes, and identifying themes [21,22]. Constant comparison was used to refine these themes into conceptual categories, contributing to the development of theoretical insights into the impacts of different care mediums [21]. For instance, themes such as client engagement, friendship, and rapport were identified as central to the sense of community. Through constant comparison, these themes were linked to broader constructs like therapeutic alliance and community dynamics, leading to theoretical insights into the differential effects of virtual and in-person care on these constructs.

The authors remained open to the depth of the transcripts and were responsive to emerging insights throughout the analytic process. The stages of this research study were adapted from Braun and Clarke's 6-phase thematic analysis framework, which include familiarization with data, generating initial codes, searching for themes, reviewing themes, and defining and naming themes, and the final phase 6, writing the report [22].

Thematic analysis in this study went in the following stages.

- Stage 1: All 3 transcripts were reviewed by each of the 3 authors (BC, AC, and AF) to uncover central codes that repeated throughout the 6 transcripts across clients and clinical staff.
- Stage 2: Codes were selected and would be used as the basis for coding in the next stage.
- Stage 3: Line-by-line coding was conducted, using the previously identified codes from stage 2 to guide reading and identifying analytically important text in the dialogue.
- Stage 4: Theoretical coding occurred, where all codes were considered collectively to generate conceptual categories (ie, themes).

Results

Overview

The in-depth qualitative focus group interviews were recorded on Zoom (Zoom Video Communications) and lasted up to 1 hour. As a result of the 6 focus group interviews (3 client groups and 3 staff groups of 4 to 8 individuals per group),

XSL•FO
RenderX

Zoom-recorded audio material of 4 hours 40 minutes was obtained, which was then subjected to transcription. Groups were labeled (ie, client groups 1, 2, and 3; clinician groups 1, 2, and 3) solely for organizational purposes and do not reflect hierarchy or analytic distinctions. The 6 interviews, 3 from clients and 3 from clinical staff, were rich in depth and content. Following iterative engagement with the data, it was identified that saturation was reached once no new themes emerged, and there was no need to conduct repeat interviews. Iterative engagement consisted of repeated transcript review, reflexive memoing, and ongoing refinement of themes as analysis progressed, in line with Braun and Clarke's reflexive thematic analysis [22]. Saturation was achieved when the data no longer yielded new insights or properties relevant to the emerging theoretical categories, and all core themes were fully developed and defined conceptually [22].

The identified codes that emerged were considered of greatest analytical importance due to their salience across interviews and relevance to the research questions. Analytic priority was given to codes that captured meaningful clinical patterns, rather than just those that appeared most frequently. The identified codes were as follows: client engagement, friendship, rapport, acuity, ADLs, client diagnosis, and use of clinical skills. These codes created 3 themes: community (client engagement, friendship, and rapport), clinical presentation (acuity, ADLs, and client diagnosis), and use of clinical skills.

Theme 1: Community

Overview

The theme of community reflects the depth of connection fostered in in-person versus virtual mediums of mental health care. Across client and clinician groups, participants noted that while virtual formats provided continuity and flexibility, they also introduced barriers to interpersonal relating and opportunities for spontaneous interaction, all of which were described as essential to a sense of therapeutic community. For example, one clinician said:

> *We've gotten to our ideal ... in person, no mask. We can see it all, and we have 100% of the information available to us in that moment ... it's ... just—was the last barrier down in terms of like physical body language and stuff.* [Staff group 1]

In this theme, community is explored thoroughly to fully saturate the concept through 3 subthemes: client engagement, friendship, and rapport. This section will explore how virtual care can serve as a valuable supplement to treatment, and where it may lack the relational qualities that contribute to the community, connection, and accountability of which in-person care naturally provides.

Client Engagement

Client engagement captured participants' observations about attention, presence, and participation, particularly in group settings. Many described virtual sessions as easier to get distracted from or tune out of, especially when cameras were off or notifications were popping up on the client's screen. If notifications are not manually turned off, it is understood that notifications may be distracting or cause the desire to "impulsively" check notifications instead of focusing on the session. Clinicians added that it is harder to address this lack of engagement when on Zoom (Textbox 1).

Textbox 1. Lack of engagement.

"It's a lot easier to be distracted. Just by being in a different space, you have different things around you that can go off. Your phone can go off. If you're on your computer, unless you make a conscious decision to turn off your notifications, you're getting notifications. Versus if you're in a group, it's like you just put your phone away and if it buzzes you feel awkward because it's buzzing and someone is talking ... And it's more of a conscious decision to not be distracted on a computer screen than it is to be in group" [Client group 3].

"I found that in virtual sessions, while they're still useful, there's inevitably going to be implicit distractions. Even if you're not explicitly trying to surf the web, you see your notifications everywhere. There's a top clock in the corner of your screen. And impulsively sometimes I feel like I just click my email, and I go back but I think that takes away from the interpersonal experience and it allows me to talk about things at a superficial level where maybe the therapist wouldn't notice" [Client group 1].

"I think that on Zoom or virtually, it's harder to address engagement or lack of engagement, so that's ... where it becomes a little different because ... it's very easy for someone to shut off their video even though we ask them not to. But it's hard for us to sort of pull them back out, whereas in person, it's a lot easier to ask them a question and look them in the eye ... and sort of feel ... the group pressure or groupthink process to get them involved. But ... I think engagement is a little bit less on—virtually 'cause it is easy to just sort of check out, look at other things" [Staff group 2].

Conversely, in-person groups were described as more immersive, often due to innate social pressure in being physically seen, subtle nonverbal cues, or eye contact that naturally invite participation, learning, and connection.

> *When you're in breakout rooms on Zoom, one, you're probably in a room alone because you're on Zoom, so you yearn for that social connection and you do that by talking about other things whereas in-person, you have time after group to talk about other things. I'm ... more focused on the subject matter of therapy with in-person groups.* [Client group 1]

> *I think it is largely more effective to be in person like everyone else has said too, but to have the flexibility and option can be nice and maybe encourage more participation if someone doesn't feel like, "Oh, I have to be in person no matter what even if X, Y, Z comes up."* [Staff group 1]

Friendship

Friendship highlighted the organic ways that peer relationships tend to form before and after in-person sessions, like "water-cooler" talk. Participants emphasized that between-group sessions is where true connection and trust were built, a dynamic

XSL•FO
RenderX

not easily replicated over Zoom, where interaction typically ends the moment the session does. For example:

> *Definitely, proximity is huge in making friends and making connections with others. I've always found the people that I'm closest with are the people that I find myself in really close proximity with and being really close involves having a shared space like [peer] said. That's a very helpful factor in making strong connections with others.* [Client group 3]

> *I mean, we have the clubhouse here and ... even if it's not anything super clinical, like the ability after group to like all go down to the clubhouse, play ping-pong, play pool, go get some lunch. That definitely wasn't happening virtually, and so I think it was definitely probably a lot harder for the clients to kinda connect with one another over Zoom, especially if they didn't already know each other.* [Staff group 1]

Rapport

Rapport focused on the alliance between clients and clinicians. While some participants described ways in which virtual care could support connection, such as learning clues about someone based on the way they decorate their home environment, many felt that unspoken exchanges and small talk were more organically cultivated in person. Additionally, a client shared that connection and community in person greatly reduced insecurities, leading to better engagement in treatment (Textbox 2).

Textbox 2. Nonverbal communication and connection.

"... let's say we're doin' like an online DBT or somethin' like that and we're going through stuff. It was more of giving someone support with what they said instead of something like—it seemed very clinical. It didn't seem organic" [Staff group 1].

"If I've heard somebody say something that I resonate with on a Zoom call, for example, I'll make a point to ... reflect on that ... and really giving them a verbal hint that I'm understanding what they're saying and nodding my head. Versus in-person, it's easier to see body language and better response than virtually whereas you would verbally need to say it instead of in-person cues" [Client group 3].

"... being able to relate to people on a personal level and getting details about them that I would not otherwise know profoundly reduced my insecurities. When I came here, I was really insecure about the fact that I was a student on a leave of absence and I talked to people and they're like, yeah, I'm also on a leave. And I'm sure maybe that would have came up but there's just details and stuff you learn about people that you just wouldn't know otherwise. And I think that's really important in building a foundation so when you get to groups [where] you're comfortable sharing, you feel better about your treatment" [Client group 1].

Some emphasized that the absence of shared physical space could subtly disrupt the therapeutic experience, due to rapport being lost. The concern is that this makes it harder to build trust and emotional connection. Some examples include:

> *I think that there's also the ability to pick up on ... body language and cues ... we'll have clients who are anxious. Like their leg is bouncing or they're fidgeting, and there's things that you can pick up that you can't pick up on virtually. And I also think having a space separate from their home and where they are allows for a different level of openness, and I think it can be hard to kind of really jump into the rapport-building in the same way virtually ...* [Staff group 1]

> *The disconnect comes from a place of personal experience with picking up on body language. And really connecting with others has a lot, for me, to do with experiencing others in their fullest sense and gaining a full experience of what others are saying versus on Zoom when we're virtual and there's not a full opportunity to gain the full experience with interacting with others. And granted, I'm definitely an extrovert, so I gain energy from being with others.* [Client group 3]

Theme 2: Clinical Presentation

Overview

Researchers also looked at clinical presentation, particularly as it relates to how a client shows up diagnostically (ie, how their symptoms are manifesting within the context of individual and group sessions and in virtual and in-person spaces). Clinical presentation was explored through 3 codes: client acuity, ADLs, and client diagnosis.

Acuity

When exploring client diagnostic acuity, openness, and engagement in individual and group therapy sessions across both mediums of care were discussed. Some staff felt that for clients in early recovery or in more of an acute psychiatric state, virtual sessions made it easier for clients to avoid addressing therapeutic issues head-on and even potentially engage in harmful behaviors behind the screen. Similarly, it was sometimes more difficult for clinicians to pick up on these behaviors virtually, decreasing the effectiveness of the therapeutic process. For example:

> *If they're in ... an environment that typically reinforces substance use or unhealthy behaviors ... I have heard stories about clients who were drinking ... alcohol out of a water bottle during sessions, and those sort of behaviors are obviously more difficult to do in an in-person setting. I think the high-acuity substance use also like takes a big hit when it's virtual as opposed to in person.* [Staff group 1]

> *Yeah, I think similarly with eating disorders as well, any with high acuity, very similar to substance use ...* [Staff group 1]

While some clinicians believed that in-person sessions were more effective for clients who needed higher levels of care and more oversight due to the severity of their symptoms, other clinicians also noted that virtual sessions could be a good option for those who are more stable (ie, in a maintenance stage of

XSL•FO
RenderX

treatment, able to cope more effectively, and need less clinical oversight). For example:

> *... if a client was then going back to college and ... wanting to be part of [mental health treatment] at like [outpatient level] it makes it a lot more accessible for them to be at [mental health treatment] if it's virtual if their university is far away, but if a client were to come straight from [residential treatment program] or a hospital setting, whatever it is ... we would probably like ... to work with them more in person to get that important work done and then—until they maybe are a little bit more stabilized. Then we can kind of transfer it over to virtual.* [Staff group 1]

> *I think that in-person therapy is absolutely imperative, especially with groups. But I can see the functionality and use of Zoom therapy as a useful supplement or tool for busy young adults.* [Client group 1]

Activities of Daily Living

ADLs explored a client's ability to take care of their mental, physical, and emotional needs. Some clinicians felt that it was more difficult to pick up on how a client was taking care of their hygiene or physical health during virtual sessions, whereas in person, there may be more obvious signs if a person did not take good care of themselves (ie, whether a person showered, shaved, and took their medication).

> *Even things like your ADLs. You can just walk past a client and smell whether or not this person took a shower or whatever. You could just see in a snapshot if they had a long night or whatever the situation is just walking around the place, seeing how they either get somethin' to eat, how they socialize with other people.* [Staff group 1]

> *I think some of the thought disorders. I think without having certain clients come in person, there are certain hygiene factors that you ... you don't know ... like when I have a client ... I can see other than just facial hair or their hair if there is—see or smell if there's different hygiene. It tells me like more about if they're taking their medication, how are they in their living, are they likely leaving their apartment at all or not, things like that.* [Staff group 1]

Clients and staff felt that virtual sessions enabled clients to engage in avoidant behaviors, such as not needing to engage in ADLs like getting out of bed, getting dressed, showering, or taking medication. This avoidant behavior sometimes even made it difficult for clients to transition into things like jobs or other settings that were not virtual.

> *You don't have to get dressed, put shoes on, go outside, take the train, all that stuff, think about transit time.* [Client group 1]

> *... being at an intensive like IOP level of care, kind of like what [therapist]'s—, like the getting up and getting to a place is part of the treatment. It's—a lot of people who haven't been leaving their house or haven't—whether it's anxiety based, depression based, executive functioning ... I think even just getting up and getting dressed and the basic kind of behavioral activation things that that help with mood and anxiety are taken away in a virtual setting.* [Staff group 1]

Client Diagnosis

When looking at diagnosis, researchers explored whether there were certain mental health diagnoses that were easier or harder to treat in a virtual session. Specifically, clinicians felt that virtual care was less effective for disorders like depression, agoraphobia, or social anxiety, where avoidance is prevalent, as it allowed clients to continue to stay in their homes and hide behind a screen, instead of addressing the issue more directly.

> *I feel like it's harder virtually to treat anyone. But I'm thinking of specifically people who like to avoid ... things, so maybe that's because of anxiety or depression. But a lot of clients here wanna avoid, and I've had lots of clients who were like, "Can we meet virtual?" because they wanna avoid taking transportation, they wanna void coming to [mental health treatment] ... but part of doing those things is like facing your anxiety head on, so by keeping things virtual for some people, that keeps them in their avoidance zone.* [Staff group 2]

> *Social anxiety. Anxiety in general. For some people, part of the therapeutic process is leaving their place to go to another office, and that obviously takes out a big part of it. I had a client who would say like, "Oh, I can't make it in today. I'm gonna be virtual," when they were hybrid, and they definitely used it as avoidance. I think that with clients who one of their goals is to be in like a more public space and interacting, virtual therapy like takes out a huge chunk of that. And I think that ... anxious and socially anxious conditions ... like virtual really is incompatible with it.* [Staff group 1]

Clinicians also felt that mental health conditions such as attention-deficit/hyperactivity disorder or thought disorders were often times more difficult to treat due to the amount of stimuli in the background and the potential for distractions.

> *What comes to mind is folks with executive functioning disorders, attention deficit disorder, and folks with psychotic disorders are distractible from their own internal stimuli and don't seem to do well with virtual.* [Staff group 2]

> *I feel like Attention Deficit Hyperactivity Disorder (ADHD), too, is another one because ... instead of you being in like a curated environment, you can be wherever. You can like walk around and someone might be in the other room, or ... there's just so ... much room for more distractions.* [Staff group 1]

Some clinicians also believe that eating disorders and substance use disorders are more difficult to treat virtually due to the lack of oversight in a virtual space. Specifically, due to the secretive nature and maladaptive behaviors of both disorders, it is harder

XSL•FO
RenderX

to assess whether clients are using substances, purging, or even eating their meals through a screen (Textbox 3).

Textbox 3. Visibility and accountability.

"I've also had a couple clients with like body image issues who didn't like to see their face on screen, and they'd hide behind like ... a ... photograph of a dog and just like would not engage and ... it was kind of disruptive in the group. I understood where they were coming from, but it was hard to find a workaround for that" [Staff group 1].

"I think it's impossible to do a truly beneficial EDO meal group virtually as opposed to in person just because of the amount of ED behaviors that would even be an option for the clients. Like I can imagine that like just having that sort of opportunity to engage in these behaviors would be either tempting or distressing to those in the group.I think that ... a successful virtual [eating disorders] meal group would be next to impossible" [Staff group 1].

"I think being in person and like seeing someone walk and sit down and being with them face to face is really helpful when you're dealing with folks who use substances. So there's just ... an added ... ability to, I think, go unnoticed and for folks to evade detection or toxicologies or any sort ... anything that will like get them found out. I think that's probably similar with eating disorders. We've seen people just completely refuse to like show how much they've eaten, or they just deny that they like went to the restroom right after eating" [Staff group 3].

There were some disorders, however, where clinicians and clients both thought that virtual sessions or groups could be appropriate and helpful depending on things like client age and the content of the therapy group, sometimes making it easier for clients to use coping skills in a comfortable space after an intense group.

> *My adolescent client ... we can do really, really good work, but due to her anxiety, like she is able to do makeup and like work on her hair and stuff like in session, which ... it also feels like very age appropriate, and I know she gets like super fidgety with her anxiety.* [Staff group 3]

> *... in terms of groups that are entirely virtual, the participation does depend group to group but there have been times where especially at trauma groups it's helpful to be virtual. I think there's less of a sense of vulnerability and I realize ... the goal is to be comfortable with vulnerability. But especially early on in treatment being in trauma groups ... it was helpful because I could sit on my bed and I could do my little things that help me calm down and I didn't feel like everyone was seeing me even though I was on camera, so in that way it was a lot more helpful.* [Client group 2]

Theme 3: Use of Clinical Skills

Finally, researchers looked at the use of their clinical skills and how the medium of care (ie, virtual vs in-person) impacts this, if at all, within the context of individual and group sessions. Use of clinical skills was both the theme and code for this category, with particular topics coming up around the importance of engagement, rapport building, and body language being a cornerstone for using clinical skills.

For example, engaging clients is a skillset that clinicians highlighted as difficult when comparing virtual and in-person therapeutic spaces. Clinicians reported that client engagement skills are negatively impacted in virtual spaces:

> *It was a deep, deep challenge to have the clients reflect until you ... found that sweet spot ... But if it was left up to, "Okay, you guys, what did you wanna talk about?" dead silent. And then working through just the different personalities ... so I didn't know where my leverage was with them or what I could use to engage them.* [Client group 3]

Fostering therapeutic alliance and rapport is another clinical skill that appeared in the transcripts when discussing the use of virtual versus in-person therapeutic spaces. Clinicians found that rapport building was easier in person. For example:

> *A big way I develop rapport is through humor, and I found that it's harder to be humorous or to exude my type of humor on Zoom 'cause just you get less ... bodily feedback and body language confirmation, so it can make me proceed a little bit more shyly on Zoom ... it's definitely about ... making sure to take that time and not like speed through something just because you have their attention and you're face to face with them.* [Staff group 2]

Clients and clinicians also noted that picking up on body language cues, an important domain of identifying signs and symptoms of mental health concerns, is more challenging in a virtual space. Being able to only see one's head prevents the clinician from accessing important nonverbal cues (ie, tapping fingers and legs shaking; Textbox 4).

XSL•FO
RenderX

Textbox 4. Challenges in a virtual space.

"I think it's just as difficult for [clinicians] to understand and realize my mannerisms ... I have a very difficult time, so I think anybody else would have a difficult time as well" [Client group 3].

"... all you can see is the square of their head. You can't see what they're doing with their legs, with their hands, who's on each side of them. It's not a real representation of therapy" [Staff group 2].

"I was trained as a psychoanalyst and group psychotherapist so seeing and lookin' at people's body language or a facial expression ... You have to really do a extra tune-in to someone that's virtual to see because again if it's just chest and above, you're not seein' other things. Is their leg movin' rapidly? Are they tappin' their fingers? Just you can't see as much as if they were in person, so it does take a shift in the way you read the room or read all group members, so there is that limitation there" [Staff group 2].

Finally, eye movement desensitization and reprocessing (EMDR) was mentioned multiple times as a modality that clinicians felt strongly about being conducted in person. The below quote summarizes the consensus:

> *I personally would prefer Eye Movement Desensitization and Reprocessing (EMDR) in person ... because it is so intense and because so much stuff can come up for a client, it's important to have a safety zone of being in an office with a clinician. And you really, really need to be able to read your clients during EMDR.* [Staff group 2]

Discussion

Principal Findings

While virtual care is acknowledged by the study participants as a valuable tool for maintaining treatment continuity, particularly in the face of logistical barriers (ie, distance, scheduling, school, or work), participants repeatedly emphasized the relational and clinical limitations innate to virtual platforms. This study uncovers, from the perspective of client and clinician focus groups, that in-person treatment is more supportive and fosters engagement in treatment, interpersonal connection (ie, therapeutic alliance and peer-to-peer), and the ability to exercise clinical skills than virtual settings.

Clinicians reported that their ability to use essential clinical skills (ie, engaging clients, fostering therapeutic alliance, and reading body language) is hindered in virtual settings. Certain modalities, such as EMDR, were considered more effective in person due to the emotional intensity they evoke and the importance of clinicians closely monitoring clients for safety in this type of therapeutic modality. While existing literature supports the feasibility of EMDR in virtual settings, there is a need for more robust methodological research to examine efficacy [22]. This study supports in-person treatment as the ideal medium for complex modalities, such as EMDR, and high acuity cases of all diagnoses, given the emotional and relational nuance that is missing when treatment is conducted online.

In examining how therapeutic relationships and community dynamics differ across virtual and in-person care, it was found that loss of connection occurs when therapy switches to a virtual format. Clients described distractibility and a reduction in therapeutic and peer-to-peer accountability when in virtual groups. Body language, nonverbal communications, and water-cooler conversations between sessions were some of the factors that were reported as being essential for therapeutic alliance and peer connection; yet, these were also reported to mainly occur in an in-person setting. Participants' reports made clear that there are qualities of in-person care that help build rapport and foster therapeutic alliance, naturally, in a way that only virtual settings can forcibly achieve. Existing literature supports the finding that virtual mental health treatment limits access to nonverbal cues, body language, and the type of connection that is central to therapeutic and peer alliance [5,17,24].

Acuity of mental health diagnoses also strongly influenced participant perceptions of the medium of care. When discussing clinical presentation, clinicians and clients admitted to the absence of and difficulty in noticing subtle cues (eg, smell and fidgeting) that are essential for picking up on signs and symptoms of mental health concerns. For this reason, in-person care was reported as a necessity for clients with higher acuity mental health concerns and complex diagnoses, such as active eating, substance use, or thought disorders.

Implications

This study extends previous work that examines the benefits and drawbacks of virtual mental health care [5]. The ability to connect, be present, and experience therapeutic alliance as well as support, particularly for young adults, is understood to be most effective when in person. While each setting offers distinct advantages, participants emphasized that clients with higher clinical acuity and complex diagnoses experience more effective support, safety, and peer-to-peer as well as therapist-to-client connection when in person. Certain diagnoses may need more support and the added element of in-person monitoring for safety, due to the complexity and/or acuity of one's mental health condition. These findings underscore the importance of accounting for therapeutic and diagnostic factors when deciding the medium of care, as this can offer an effective framework for optimizing client engagement and experience.

Limitations

The following limitations are consistent with Cerrito et al [5]. In summary, these include the following: (1) the analyzed group is not representative of the broader populations of mental health providers and young adults seeking mental health treatment, and future research should examine these hypotheses in a larger, more diverse population. (2) Considering factors that qualify a client for virtual versus in-person care is warranted in future research, as we understand that these results represent a specific group of individuals at a particular point in history (ie, post–COVID-19 pandemic), from large urban settings (ie, New York, NY, and Washington, DC), and social contexts (ie, predominantly high socioeconomic status). (3) Due to the

XSL•FO
RenderX

necessary exclusion criteria, the sample may not fully represent all vulnerable populations. Individuals were excluded if they were unable to complete the survey due to medical or psychological limitations or if participation required a legally authorized representative.

Conclusions

The results of this study outline how the medium of care (ie, virtual vs in- person) impacts therapeutic experience from the perspective of both client and clinician. It is understood that critical domains such as community, interpersonal relationships, and clinicians' use of clinical skills are affected by the medium of care, and key factors such as clinical presentation (ie, acuity and diagnosis) must be accounted for when choosing a particular mode of care for a client.

Funding

The authors declared no financial support was received for this work.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the confidential nature of this information.

Authors' Contributions

Conceptualization: BC, AF, FDB
Data curation: FDB
Formal analysis: FDB
Investigation: BC, JX, AF, FDB, AC
Methodology: AF, FDB, BC
Project administration: FDB, BC
Resources: AF, BC
Supervision: FDB
Validation: FDB
Visualization: FDB, BC
Writing—original draft: BC, AC, AF, JX, JU, FDB
Writing—review and editing: BC, JX, AC, AF, FDB, JU

Conflicts of Interest

FDB is a consultant for Addinex Technologies, NY. All other authors have no conflicts of interest to declare.

References

1. Giordano C, Ambrosiano I, Graffeo MT, Di Caro A, Gullo S. The transition to online psychotherapy during the pandemic: a qualitative study on patients' perspectives. Res Psychother 2022 Nov 4;25(3):638. [doi: 10.4081/ripppo.2022.638] [Medline: 36373388]
2. Gangamma R, Walia B, Luke M, Lucena C. Continuation of teletherapy after the COVID-19 pandemic: survey study of licensed mental health professionals. JMIR Form Res 2022 Jun 1;6(6):e32419. [doi: 10.2196/32419] [Medline: 35584317]
3. Zulman DM, Verghese A. Virtual care, telemedicine visits, and real connection in the era of COVID-19: unforeseen opportunity in the face of adversity. JAMA 2021 Feb 2;325(5):437-438. [doi: 10.1001/jama.2020.27304] [Medline: 33528520]
4. Adams L, Adamo N, Hollocks MJ, et al. Autistic young people's experiences of remote psychological interventions during COVID-19. Autism 2023 Aug;27(6):1616-1627. [doi: 10.1177/13623613221142730] [Medline: 36645009]
5. Cerrito B, Connors A, Fialk A, Xiao J, Buono FD. Understanding therapeutic experience: a qualitative evaluation of virtual and in-person treatment. Cureus 2024 Dec;16(12):e75392. [doi: 10.7759/cureus.75392] [Medline: 39781144]
6. Békés V, Aafjes-van Doorn K, Roberts KE, Stukenberg K, Prout T, Hoffman L. Adjusting to a new reality: consensual qualitative research on therapists' experiences with teletherapy. J Clin Psychol 2023 May;79(5):1293-1313. [doi: 10.1002/jclp.23477] [Medline: 36704974]
7. Salazar de Pablo G, Pascual-Sánchez A, Panchal U, Clark B, Krebs G. Efficacy of remotely-delivered cognitive behavioural therapy for obsessive-compulsive disorder: an updated meta-analysis of randomised controlled trials. J Affect Disord 2023 Feb 1;322:289-299. [doi: 10.1016/j.jad.2022.11.007] [Medline: 36395988]
8. Solimini R, Busardò FP, Gibelli F, Sirignano A, Ricci G. Ethical and legal challenges of telemedicine in the era of the COVID-19 pandemic. Medicina (Kaunas) 2021 Nov 30;57(12):1314. [doi: 10.3390/medicina57121314] [Medline: 34946259]
9. Huggett C, Gooding P, Haddock G, Quigley J, Pratt D. The relationship between the therapeutic alliance in psychotherapy and suicidal experiences: a systematic review. Clin Psychol Psychother 2022 Jul;29(4):1203-1235. [doi: 10.1002/cpp.2726]

XSL•FO
RenderX

10. Park EY, Oliver TR, Peppard PE, Malecki KC. Sense of community and mental health: a cross-sectional analysis from a household survey in Wisconsin. Fam Med Community Health 2023 Jun;11(2):e001971. [doi: 10.1136/fmch-2022-001971] [Medline: 37399294]
11. Stubbe DE. The therapeutic alliance: the fundamental element of psychotherapy. Focus (Am Psychiatr Publ) 2018 Oct;16(4):402-403. [doi: 10.1176/appi.focus.20180022] [Medline: 31975934]
12. Baier AL, Kline AC, Feeny NC. Therapeutic alliance as a mediator of change: a systematic review and evaluation of research. Clin Psychol Rev 2020 Dec;82:101921. [doi: 10.1016/j.cpr.2020.101921] [Medline: 33069096]
13. Bjørlykhaug KI, Karlsson B, Hesook SK, Kleppe LC. Social support and recovery from mental health problems: a scoping review. Nordic Soc Work Res 2022 Dec 5;12(5):666-697. [doi: 10.1080/2156857X.2020.1868553]
14. Eichenberg C, Aranyi G, Rach P, Winter L. Therapeutic alliance in psychotherapy across online and face-to-face settings: a quantitative analysis. Internet Interv 2022 Sep;29:100556. [doi: 10.1016/j.invent.2022.100556] [Medline: 35942217]
15. Stewart C, Konstantellou A, Kassamali F, et al. Is this the "new normal"? A mixed method investigation of young person, parent and clinician experience of online eating disorder treatment during the COVID-19 pandemic. J Eat Disord 2021 Jun 30;9(1):78. [doi: 10.1186/s40337-021-00429-1] [Medline: 34193291]
16. Aafjes-Van Doorn K, Békés V, Luo X, Hopwood CJ. Therapists' perception of the working alliance, real relationship and therapeutic presence in in-person therapy versus tele-therapy. Psychother Res 2024 Jun;34(5):574-588. [doi: 10.1080/10503307.2023.2193299] [Medline: 37011405]
17. Mercadal Rotger J, Cabré V. Therapeutic alliance in online and face-to-face psychological treatment: comparative study. JMIR Ment Health 2022 May 2;9(5):e36775. [doi: 10.2196/36775] [Medline: 35499910]
18. Lin T, Stone SJ, Heckman TG, Anderson T. Zoom-in to zone-out: therapists report less therapeutic skill in telepsychology versus face-to-face therapy during the COVID-19 pandemic. Psychotherapy (Chic) 2021 Dec;58(4):449-459. [doi: 10.1037/pst0000398] [Medline: 34881922]
19. Bulkes NZ, Davis K, Kay B, Riemann BC. Comparing efficacy of telehealth to in-person mental health care in intensive-treatment-seeking adults. J Psychiatr Res 2022 Jan;145:347-352. [doi: 10.1016/j.jpsychires.2021.11.003] [Medline: 34799124]
20. Chun Tie Y, Birks M, Francis K. Grounded theory research: a design framework for novice researchers. SAGE Open Med 2019;7:2050312118822927. [doi: 10.1177/2050312118822927] [Medline: 30637106]
21. Ahmed SK, Mohammed RA, Nashwan AJ, et al. Using thematic analysis in qualitative research. J Med Surg Pub Health 2025 Aug;6:100198. [doi: 10.1016/j.glmedi.2025.100198]
22. Guest G, Namey E, Chen M. A simple method to assess and report thematic saturation in qualitative research. PLoS One 2020;15(5):e0232076. [doi: 10.1371/journal.pone.0232076] [Medline: 32369511]
23. Glaser BG, Holton J. Staying open: the use of theoretical codes in grounded theory. In: Grounded Theory Review: Institute for Research and Theory Methodologies; 2023, Vol. 22:4-16.
24. von Below C, Bergsten J, Midbris T, Philips B, Werbart A. It turned into something else: patients' long-term experiences of transitions to or from telepsychotherapy during the COVID-19 pandemic. Front Psychol 2023;14. [doi: 10.3389/fpsyg.2023.1142233]

Abbreviations

ADL: activities of daily living
EMDR: eye movement desensitization and reprocessing

Edited by A Kushniruk; submitted 25.Aug.2025; peer-reviewed by A Waheed, M Hirschi; revised version received 10.Mar.2026; accepted 10.Mar.2026; published 20.Apr.2026.

Please cite as:
Cerrito B, Connors A, Fialk A, Xiao J, Usow J, Buono FD
Acuity, Diagnosis, and Community Factors in Virtual and In-Person Treatment: Qualitative Study
JMIR Hum Factors 2026;13:e82863
URL: https://humanfactors.jmir.org/2026/1/e82863
doi:10.2196/82863

© Brianna Cerrito, Alexa Connors, Amanda Fialk, Jamie Xiao, Josephine Usow, Frank D Buono. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 20.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The

XSL•FO
RenderX

complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Role of Explanations in AI-Generated Alerts: Qualitative Study of Clinical Views on Explainable AI in Predictive Tools

Jessica Rahman[1], PhD; Alana Delaforce[2,3], PhD; DanaKai Bradford[2], PhD; Jane Li[1], PhD; Farah Magrabi[4], PhD; David Cook[2,5], PhD; Aida Brankovic[2], PhD

[1]Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation (CSIRO), 160 Hawkesbury Road, Sydney, New South Wales, Australia

[2]Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Brisbane, Queensland, Australia

[3]School of Nursing and Midwifery, Griffith University, Brisbane, Australia

[4]Australian Institute of Health Innovation, Macquarie University, Sydney, Australia

[5]Intensive Care Unit, Princess Alexandra Hospital, Brisbane, Australia

Corresponding Author:
Jessica Rahman, PhD
Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organisation (CSIRO), 160 Hawkesbury Road, Sydney, New South Wales, Australia

Abstract

Background: Artificial intelligence (AI)–driven clinical decision support (CDS) tools offer promising solutions for health care delivery by optimizing resource allocation, detecting deterioration, and enabling early interventions. However, adoption remains limited due to insufficient validation and a lack of transparency and trust. Explainable AI (XAI) seeks to improve user understanding of AI outputs; however, how clinicians interpret and integrate these explanations into their decision-making remains underexplored. Furthermore, discrepancies in explanations, known as the "disagreement problem," can undermine trust and, at worst, lead to poor clinical decisions.

Objective: This study examines clinicians' perspectives on the role and value of explainability in AI-driven CDS tools within Australian critical care settings and the impact of discrepancies in AI-generated explanations on clinical decision-making.

Methods: Qualitative data were collected using semistructured interviews with 14 clinical experts, incorporating scenario-based exercises, and were analyzed using inductive thematic analysis.

Results: Clinicians valued explainability, particularly in complex or unfamiliar situations, when explanations were clear, plausible, and actionable. Trust and perceived usefulness extended beyond explanation quality, encompassing factors such as system accuracy, alignment with clinicians' reasoning, workflow integration, and perceived reliability. Discrepancies in explanations generated by different XAI methods were not a major concern, provided that the AI-generated predictive alerts were accurate.

Conclusions: This study provides design recommendations for developing trustworthy, user-centric CDS tools that incorporate XAI. Findings highlight that explainability is critical for establishing initial trust in AI-driven tools by supporting perceived usefulness, but its importance diminishes over time and with user expertise and familiarity, as learned usefulness takes precedence. Recommendations highlight the importance of aligning the design and implementation of AI tools with clinicians' needs to enhance trust, mitigate risks, and promote successful adoption for improved patient outcomes.

(JMIR Hum Factors 2026;13:e81460) doi:10.2196/81460

KEYWORDS

artificial intelligence; explainable artificial intelligence; explainable AI; disagreement in XAI; clinical decision support; qualitative research; human factors

Introduction

Artificial intelligence (AI)–driven clinical decision support (CDS) systems are emerging as powerful tools in health care, leveraging large-scale medical data to identify patterns and generate insights that help clinicians deliver more informed and effective care [1-3]. Beyond supporting clinical decision-making, these systems have the potential to optimize resource allocation and streamline workflows by processing large volumes of electronic health data and deploying scalable, cost-effective algorithms that enhance efficiency while minimizing additional resource burdens [4-8]. Furthermore, CDS tools can detect subtle patterns of deterioration and offer predictive insights to help clinicians intervene earlier [9,10]. However, despite the promise of AI in health care, these tools have not been widely adopted for several reasons. One

XSL•FO
RenderX

significant barrier is the lack of real-world testing and validation [11-13]. While early evidence suggests that AI-driven tools hold promise beyond controlled academic settings, their effectiveness in dynamic, real-world clinical environments remains underexplored, and further research is needed to understand the factors influencing their adoption, including patient heterogeneity, system integration, and workflow compatibility [13,14].

Recent research on deployed decision support tools also highlights the lack of human factors consideration as a key driver behind their limited adoption in clinical settings [1,15-19]. This includes a lack of attention to how clinicians interact with technology, usability issues, and how well the tool fits into existing workflows [1,16-19]. Another major concern is the lack of explainability and transparency in many AI-driven CDS tools [20,21]. Clinicians are understandably hesitant to rely on tools they do not fully understand, especially in high-stakes environments such as hospitals [2,22]. When an AI-driven CDS system provides a recommendation or prediction, it is crucial for health care providers to understand the reasoning behind it. Without clear and understandable explanations, there is a risk that clinicians will not trust the tools, regardless of their accuracy [23].

Without adequate transparency, there is also a risk that clinicians may overrely on these tools, potentially leading to harmful outcomes [24-27]. This trust gap between health care providers and AI-driven CDS tools is a significant obstacle to widespread adoption [2,28]. Effective implementation of AI-driven CDS systems requires careful trust calibration, ensuring that clinicians develop an appropriate level of trust aligned with the system's real-world performance, to support safe and effective use in practice [27,29].

To support trust calibration of AI-driven CDS tools, explainable AI (XAI) methods have been designed to provide transparency by not only presenting the predictions but also revealing the underlying rationales that drive those predictions [27,30-33]. By allowing medical professionals to see the logic behind the AI's decision, XAI aims to bridge the gap between the complex algorithms used in machine learning and the human need for clarity and accountability in clinical practice. For tasks such as classification and regression, post hoc explanation methods are increasingly being used to shed light on the behavior of complex models after training [15,34,35]. Rather than altering the model, these methods aim to reveal how it makes decisions by producing clinically informative approximations or visual representations. However, recent findings in responsible AI have highlighted significant challenges in creating robust, user-centric, and context-sensitive explanations that can be trusted and easily comprehended [36,37].

While much of the existing literature has focused on the technical reliability or consistency of XAI methods, less attention has been paid to how clinicians interpret, evaluate, and use these explanations in practice. Understanding these human-centered aspects of explainability is critical for ensuring that AI-driven CDS tools support, rather than disrupt, clinical workflows and reasoning. One key challenge is that post hoc explanations generated by models can be overly sensitive to small input changes or offer convincing justifications for incorrect predictions [26,38]. Research has shown that when these explanations are presented alongside model outputs, they can unintentionally foster increased trust in the model, even if that trust is not justified, and lead to potentially harmful decisions [33,39-41]. Another challenge is the "disagreement problem," which refers to discrepancies among state-of-the-art methods in the explanations they provide, as revealed by recent research [37,42-46].

The disagreement issue arises when different XAI techniques produce conflicting explanations, even when applied to the same input and underlying model prediction. While several studies have shown that AI-generated explanations can increase clinicians' likelihood of accepting recommendations from CDS tools [1,47-50], they often rely on the implicit assumption that a correct model prediction is accompanied by a valid explanation. Previous studies have examined the impact of incorrect explanations on users when the AI's prediction is correct [51]. However, they often overlook the fact that different XAI methods can produce conflicting explanations for the same prediction. Moreover, recent research has shown that such explanations can inflate users' trust in the model even when that trust is not warranted [26,33,39,40]. Despite these concerns, little is known about how clinicians understand and respond to such discrepancies. This gap not only limits our understanding of explanation reliability in high-stakes settings but also complicates the task of selecting appropriate XAI methods for use in CDS tools.

Building on these insights, this study explores clinicians' perspectives on the role and value of explainability in AI-driven CDS tools, focusing on how explanations are interpreted, trusted, and incorporated into clinical decision-making through qualitative interviews with clinical experts. It also explores how discrepancies between different explanation methods affect clinical judgments through a scenario-based approach. Additionally, we seek to identify key considerations for effectively integrating explainability into AI-driven CDS tools. Our study is focused on Australian health care settings, where there is a growing need for exploration of AI-driven CDS tools [52]; yet, limited empirical evidence exists regarding their acceptance and impact in real-world clinical practice. This study seeks to contribute to this area by exploring 3 key research questions (RQs) listed below:

- RQ1: How does explainability influence clinicians' use, trust, and perceived usefulness of AI-driven CDS tools?
- RQ2: How do clinicians interpret and respond when different explainability methods provide differing explanation factors, despite the AI prediction being correct?
- RQ3: What key design considerations can support the trustworthy and user-centric integration of AI-driven CDS tools with explainability as a key component?

Methods

Study Design

Semistructured interviews were conducted with clinical experts to explore their views on XAI and the discrepancies observed

XSL•FO
RenderX

between different XAI methods. During the interviews, participants were given a scenario-based exercise using an AI-based predictive CDS tool to assess how discrepancies in explanations influenced decision-making.

Participant Recruitment

The study participants comprised clinical experts who were potential users of CDS tools designed for use in the intensive care unit (ICU) environment. Recruitment began by reaching out to current and former clinical collaborators to identify suitable participants. A purposive, snowball sampling method was used, allowing the identification of additional stakeholders as the research progressed [53]. A recruitment flyer was prepared and emailed to all potential participants. After participants signed the consent form, they scheduled an interview time using an online scheduler. Recruitment and data collection were conducted concurrently with data analysis, continuing until data saturation was reached [54].

Data Collection

Semistructured interviews were conducted from February to July 2024. The interview was divided into 3 main segments, each containing multiple open-ended questions. The questions for the first 2 segments were adapted from the study by Schwartz et al [55]. The first segment focused on understanding the participants' clinical experience and their general views on CDS and XAI. The second segment explored participants' thoughts on factors influencing their trust in CDS tools and how XAI contributes to it. In the final segment, participants were presented with 2 scenarios based on an AI-based predictive CDS and explainability modules developed in a previous study using electronic medical record data [36]. Detailed descriptions of these scenarios are provided in the "Scenarios" subheading. A pilot interview was conducted with 1 participant, leading to further refinement of the questions and interview duration. The interview guide is available in Multimedia Appendix 1. Interviews were conducted online via Microsoft Teams and recorded using the platform's built-in feature. Four team members (JR, AD, JL, and DB) with doctoral qualifications and expertise in qualitative research conducted the interviews, with at least 2 present at each session. Each session lasted between 45 and 60 minutes. Notes were taken during the interviews. All audio recordings were transcribed by a professional transcription service and deidentified upon return.

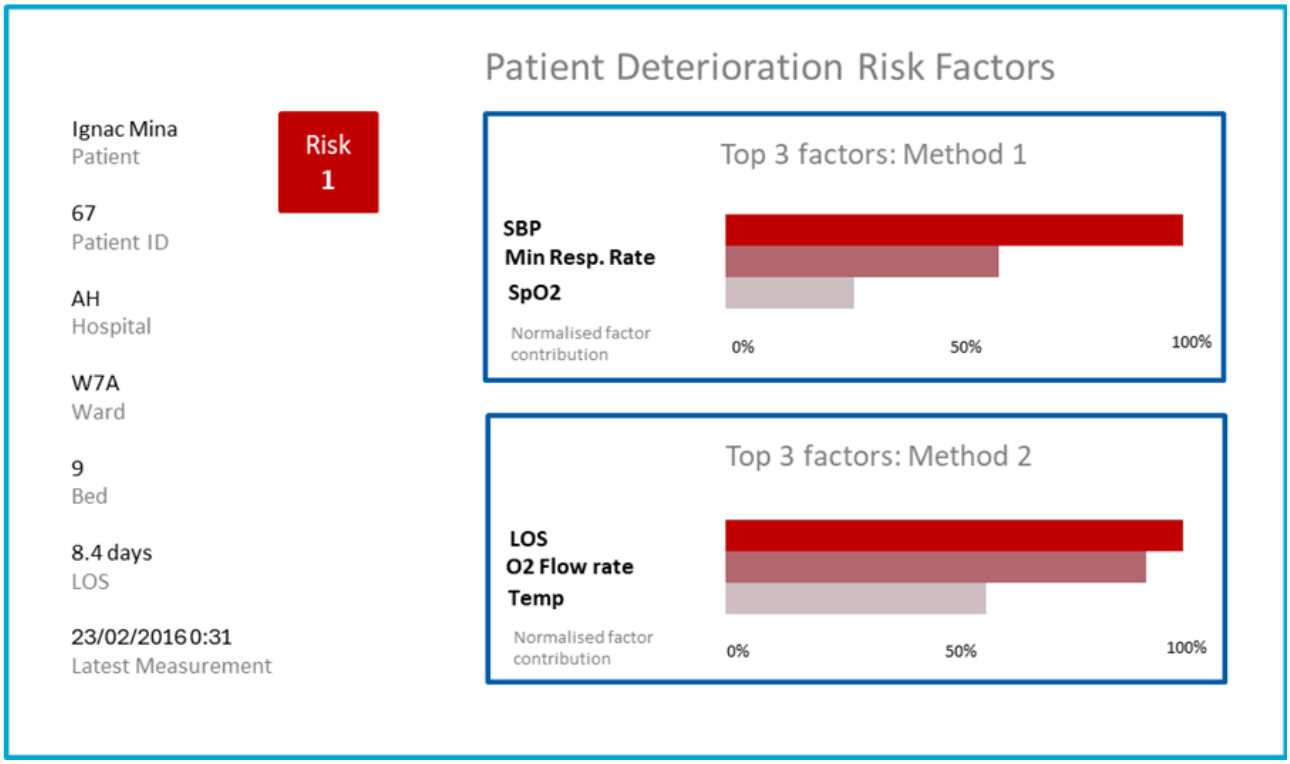
Scenarios

The CDS tool selected as the use case for this study was an algorithm that predicts the activation of an early warning score system, providing up to 8 hours of warning of potential patient deterioration. The algorithm was developed and validated using real-world data from a tertiary hospital in Queensland, Australia, and further methodological and technical details are available in prior publications [36,56]. The scenarios included encounter records, extracts with primary and additional diagnosis Systematized Nomenclature of Medicine codes, extracts indicating deteriorating alerts (red and yellow flags), and vital signs data. Explanations were generated by several post hoc explainers using real examples from the model's training and testing datasets. The selected use cases were constructed to demonstrate instances in which the model's predictions were accurate, while the corresponding explanations varied in their degree of correctness. This approach aimed to assess how domain experts not only see the explanations but also interpret and respond to discrepancies between model accuracy and explanation fidelity and to ensure that correct model predictions would not be disregarded solely due to imperfections in the accompanying explanations. Information was presented as a prototype of the tool planned for pilot implementation (Figures 1 and 2). Further information on patient diagnoses and procedures was presented to the participants when the scenarios were shown, to provide contextual information about the considered cases and to realistically simulate the clinical scenarios. These details are provided in Multimedia Appendix 2.

XSL•FO
RenderX

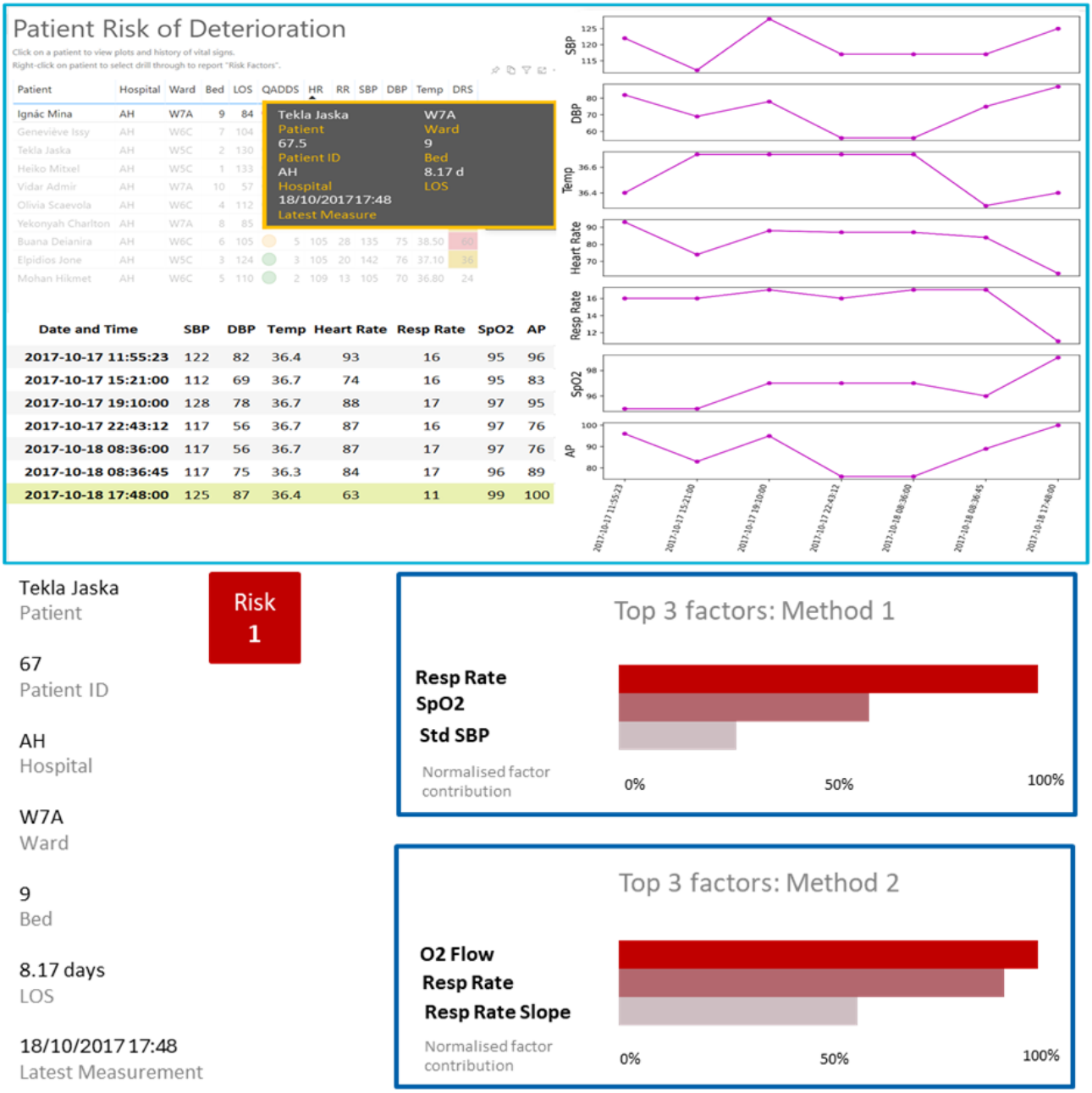
Figure 1. The first part of the clinical decision support (CDS) tool "(A)" showed a list of patients at risk of deterioration and recorded values of vital signs for selected patients in the last 24 hours, alongside the basic demographics. The second part of the tool "(B)" presents additional available information corresponding to the patient regarding the predicted deterioration risk. In addition to basic demographics about the patient, 3 main factors rationalizing the model prediction were presented, obtained using 2 different methods. The factors were ordered based on their normalized contribution in descending order. Both methods correctly predicted patient deterioration. The main contributor obtained with Method 1 matched the vital sign that triggered the alarm, unlike any of the contributors obtained with Method 2. AH: anonymized hospital identification; AP: arterial pressure; DBP: diastolic blood pressure; DRS: deterioration risk score; HR: heart rate; ID: identification; LOS: length of stay; Min: minimum; QADDS: Queensland Adult Deterioration Detection System; Resp: respiratory; RR: respiratory rate; SBP: systolic blood pressure; SPO2: oxygen saturation; Temp: temperature.





XSL•FO
RenderX

Figure 2. (A) and (B) illustrate results for a different patient. In this case, both predictive methods accurately identified patient deterioration, and both explainable artificial intelligence (XAI) techniques recognized the same contributing factor as indicated by the vital sign that triggered the alarm, but with differing ranks of importance. AH: anonymized hospital identification; AP: arterial pressure; DBP: diastolic blood pressure; DRS: deterioration risk score; HR: heart rate; ID: identification; LOS: length of stay; Min: minimum; QADDS: Queensland Adult Deterioration Detection System; Resp: respiratory; RR: respiratory rate; SBP: systolic blood pressure; SPO2: oxygen saturation; Temp: temperature.



Data Analysis

The analysis aimed to uncover domains related to what influences trust in XAI and how discrepancies affect it. Due to the lack of suitable existing frameworks that adequately addressed the issues discussed by participants for deductive analysis, an inductive open coding approach using grounded theory [57] was used, allowing themes to emerge organically from the data without any preconceived categories or theoretical framework. The data analysis was conducted using Microsoft Excel. The COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist is provided in Checklist 1.

Initial coding of each interview transcript was carried out by a qualitative researcher (JR). The coding was then sent to a second qualitative researcher (AD) for agreement checking. Consensus discussions were held, and all disagreements were resolved between the 2 researchers. Using the inductive approach, an initial codebook was developed after coding the first transcript. This codebook guided the coding of subsequent transcripts and was iteratively reviewed and updated following each consensus discussion. After completing the initial coding, the codes were grouped into broader domains. For each domain, the 2 qualitative researchers conducted a consistency check to refine the constructs and ensure all data were coded consistently. At this stage, 1 researcher (JR) also inductively summarized each text artifact into a statement that captured its meaning, referred

XSL•FO
RenderX

to as subconstructs. A consensus discussion was then conducted with the second researcher (AD), using the same method as the initial coding discussions, to further refine and condense the statements where needed. This iterative process of consistency checking at each stage enabled the study team to triangulate the data, incorporating multiple perspectives to strengthen confidence in the study's findings. After refinement, the preliminary findings were emailed to all interview participants, following their agreement to participate in a "member checking" process. Participants were asked to provide feedback on whether the preliminary findings captured the most important issues and on the clarity of the phrasing used for constructs and statements. Responses were received from 57% (8/14) of participants, comprising affirmative and/or constructive feedback. This feedback informed further refinement and changes to the final set of domains and constructs.

Ethical Considerations

Ethics approval for this study was obtained from the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Health and Medical Research Ethics Committee (reference no 2023_074_LR). A privacy review of the study was also conducted by a legal counsel at CSIRO. A recruitment flyer, which included a link to the online participant information sheet and digital consent form, was prepared and emailed to all potential participants. The digital consent form was hosted in the CSIRO's secure REDCap (Research Electronic Data Capture; Vanderbilt University) application. The information sheet provided details about the study's purpose, privacy considerations, potential risks and benefits, and how the data would be collected and used. Upon completing their participation, participants were provided with an AUD $50 (approximately US $35) gift card.

Results

Participant Demographics

Clinical experts (N=14) took part in the interviews (Table 1). Experience levels were categorized as follows: "Junior"= <10 years of experience, "Mid"= 10 - 20 years, and "Senior"= >20 years. Overall, there was equal participation from both doctors (7/14) and nurses (7/14) working in the ICU environment, with 64% (9/14) in the adult, 29% in the pediatric (4/14), and 7% (1/14) in the neonatal ICU. Of the participants, 50% (7/14) were identified as senior clinical experts (4 females), whereas 29% (4/14, 2 females) and 21% (3/14, 2 females) were identified as mid and junior level, respectively. Participants were based in 5 states and territories of Australia, comprising the Australian Capital Territory, New South Wales, Victoria, Queensland, and Western Australia.

Table . Demographics of the interview participants.

Category and subcategory	Participants, n (%)
Role	
Doctor	7 (50)
Nurse	7 (50)
Experience	
Junior	3 (21.4)
Mid	4 (28.6)
Senior	7 (50)
Sex	
Male	6 (42.9)
Female	8 (57.1)
ICU[a] environment	
Adult	9 (64.3)
Pediatric	4 (28.6)

[a]ICU: intensive care unit.

Participants were also asked about their current use of CDS systems, which could be AI-powered or not. The majority (n=11) of participants indicated they had experience with CDS tools; however, they were not AI-driven. Crude rule-based scoring systems were still prevalent, and while they provided a framework for assessing risk, they lacked the maturity needed for real-time individualized care adjustments. Three participants indicated that they did not have any experience with computer-based CDS tools because their current systems were paper-based and relied on manual charting of patient observations. All participants recognized the potential of AI to address clinical challenges, and many felt that the systems currently in use were relatively simple and would benefit from further development and integration into everyday clinical practice.

Findings

Overview

An inductive thematic analysis using open coding identified 6 overarching domains. Of these, 3 focused on AI-driven CDS with XAI as a key component and are therefore reported here: (1) explainability and actionability, (2) trust, and (3) design

XSL•FO
RenderX

recommendations for AI-driven CDS with XAI. The constructs and subconstructs associated with these 3 domains are presented in Table 2. The number of participants refers to the number of participants who raised a construct at least once across interviews. For example, a participant may have mentioned an issue with high frequency in their interview, but it was only counted once to avoid potentially skewed data. Exemplar quotes for each of the subconstructs are provided. The remaining 3 domains identified broader use, concerns, and limitations surrounding AI-powered CDS tools, as well as implementation considerations for successful adoption. These additional findings are presented in Multimedia Appendix 3.

XSL•FO
RenderX

Table . Constructs and subconstructs of the 3 key domains identified in the qualitative analysis.

Domain, construct, and subconstruct	Exemplar quote
Explainability and actionability	
Need and usefulness of explainability (n=7)	
Complexity of both the patient condition and decisions to be made (n=6).	"…there may be less need for complex explanation when it's saying put patient X in bed X, because they're likely to move by this date as opposed to if I'm trying to diagnose a pneumonia which might be confused for a COPD type thing…" (P11, Nurse, Mid, Pediatric)
Clinicians' experience level (n=2).	"…but I think there's very much a role with the more junior staff who are learning…they're the ones who need the explanations more so than the senior people" (P10, Nurse, Senior, Adult)
AI[a] explanations, cognitive processes, and decision-making (n=13)	
Offset cognitive overload and prompt safety checks (n=7).	"…when there are times of cognitive overload, something that can sort through more information and facilitate a clear direction, that would be much better." (P9, Nurse, Junior, Pediatric)
Provide deterministic explanations that fit the clinician's mental model (n=7).	"…So the method two makes a lot more sense because…it fits with my ground truth..." (P14, Doctor, Senior, Neonatal)
Provide plausible explanations that can either reinforce or challenge a clinician's existing perspective (n=4).	"If it's a plausible explanation, then obviously I would have to rethink to determine whether that is the right thing." (P7, Doctor, Senior, Adult)
Avoid ambiguous explanatory factors as they may cause confusion (n=3).	"…the point you provide me with some information that comes out of context, am I better off or not? It is sufficient to make me less certain…" (P1, Doctor, Senior, Adult)
Application of AI CDS[b] explanations (n=13)	
Prompt patient review and be considered within the broader context of patient-related factors (n=12).	"…there's so many variables outside of obs that allow us to make decisions." (P5, Doctor, Junior, Adult)
In the setting of multiple explanations, can prompt the same clinical decision (n=3).	"The fact they don't agree is not [an issue], I can think of many examples where there are evidence that allowed us to reach the same conclusion..." (P1, Doctor, Senior, Adult)
Trust	
Influencing factors (n=12)	
Clinician age and experience (n=5).	"…the younger the clinicians are, the less they're frightened of it, maybe the more accepting they are of it…" (P10, Nurse, Senior, Adult)
Accurate predictions, which are trusted more than accurate explanations (n=6).	"…if it triggers that there's something going wrong, I would probably always take that seriously and so even if the explanation's not that great, it's up to us to…rule out what's going on with the patient…" (P13, Nurse, Senior, Adult)
Conflicts in decision-making stemming from information provided by AI-driven CDS (n=1).	"…these sorts of tools…become kind of bargaining chips between clinical team, because…data that is in black and white…will always be trusted more than someone's gut feeling or someone's intuition etc." (P5, Doctor, Junior, Adult)
Lived clinical experience and clinicians' understanding of the patient's situation (n=3).	"…the benefit for these systems is for less experienced people that haven't got…lived experience clinically…or my senior colleagues because we can't be at the bedside 24/7." (P14, Doctor, Senior, Neonatal)
Opinion leaders (n=2).	"…we want the clinicians who have some expertise in managing deteriorating patients to endorse and support the algorithm." (P7, Doctor, Senior, Adult)
Evidence-based and clinical utility (n=9).	"…if you demonstrated that in the type of hospital I am in, with the type of electronic systems and everything being similar, that the system always was better than all of the clinicians…. I would have to trust it more quickly...." (P2, Doctor, Mid, Adult)
Clinician internal calibration (n=11)	
Repeated exposure, consistent performance, reflection on clinical outcomes, and accuracy of predictions (n=11).	"…having that explainable element and…familiarity with using and trying previous AI or decision support tools successfully and that being helpful in my practice - would influence my views..." (P4, Doctor, Mid, Adult)
Impact of agreement level (n=12)	

XSL•FO
RenderX

Domain, construct, and subconstruct	Exemplar quote
Alignment with a clinician's mental model fosters trust (n=12).	"…what is it about various elements of the patient's profile that have made that tool perform in such a different way to what I was thinking..." (P4, Doctor, Mid, Adult)
Design recommendation for AI-driven CDS with XAI[c]	
Explanations are actionable and useful (n=10)	
Systems should present actionable data such as physiological trends, rather than standalone routine metrics, in a timely manner (n=10).	"…that's fairly well understood that the longer somebody stays the more at risk they are for hospital acquired complications.... It doesn't necessarily value add in that, but the physiological variables definitely." (P10, Nurse, Senior, Adult)
Systems should demonstrate in-depth predictive ability with detailed analysis of potential reasons for patient deterioration (n=3).	"…rather than just triggering for a temperature or a heart rate…, really putting it up front to say that it would be a sepsis trigger." (P9, Nurse, Junior, Pediatric)
Patient complexity and contextual factors (n=9)	
Provision of information beyond physiological data, such as the patient's history and comorbidities (n=9).	"…get the machine learning to pick up that…historical background of pre-admission and intra-admission factors..." (P9, Nurse, Junior, Pediatric)
Other recommendations (n=13)	
Simple, optional, and concise explanations (n=7).	"…the prediction that you had popping up as an overlay above the observations so that people could read them together would be helpful. Along with an explanation as to why, well even do you want the explanation?" (P12, Nurse, Senior, Adult)
The level of certainty and the potential for bias in prediction (n=6).	"…you need shades of confidence…maybe with confidence intervals, not just a score but some idea of how confident the system is." (P2, Doctor, Mid, Adult)
At-a-glance insights using clear graphics and minimal text (n=5).	"…if it took a lot of time and you had to make a decision about something…, then that would be kind of wasteful or distracting you from a patient." (P4, Doctor, Mid, Adult)
Color coding to indicate level of deterioration and transparent explanation (n=3).	"…whatever pops up on screen there's a little bit of explanatory detail and a bit of color coding will be that traffic lights or something else, basically saying this is what the AI model is predicting or directing…and this is why..." (P11, Nurse, Mid, Pediatric)
Information confined to specific intended use cases (n=2).	"So I trust this to give me a list of people that I should see, as a high priority and I can plan my workflow.... I do not trust it to tell me what dose of drug to use and what drugs to use on people." (P1, Doctor, Senior, Adult)
User-friendly, intuitive, and requiring minimal effort to use (n=1).	"The ones that actually get used clinically are the ones that require the least amount of effort..." (P5, Doctor, Junior, Adult)

[a]AI: artificial intelligence.

[b]CDS: clinical decision support.

[c]XAI: explainable artificial intelligence.

Explainability and Actionability

The domain explored when and why explainability is needed in CDS tools. It identified how effective explanations can enhance decision-making by helping clinicians manage cognitive load and focus on key details. It also explored how explanations influence decision-making, particularly when varying levels of disagreement arise between different explanation methods.

Subconstructs explored the context-dependent usefulness of explanations, where the need for explainability varies with clinician experience, task complexity, and decision type. Junior clinicians would benefit more from explanations for understanding AI outputs, especially in unfamiliar or complex situations, while reliance on detailed explanations would decrease in routine tasks or for experienced clinicians who intuitively assess clinical indicators. Junior clinicians may benefit from detailed explanations because they are still developing their clinical reasoning and may rely on AI outputs as a learning tool. In contrast, senior clinicians, drawing on extensive experience, may use AI mainly as a confirmatory aid, focusing on explanations primarily when outputs diverge from their expectations or offer novel insights. As AI tools may fail to capture the multifactorial nature of patient conditions and physiology, their role should be to provide direction on which physiological parameters to investigate further, with final decision-making remaining a human responsibility.

Trust

The trust domain examined factors influencing trust in AI-driven CDS, including clinicians' age, experience, intuition alignment, and the tool's predictive accuracy. It further highlighted how trust builds gradually through reliable performance and repeated use, enabling integration with clinical decisions. Specifically,

XSL•FO
RenderX

confidence would grow as users understand the evidence behind the algorithm and how it was developed. Validation of the XAI tools by multiple studies and investigators, particularly if in similar settings, would similarly foster trust. One of the most important determinants of trust was alignment with clinical instinct. A system that consistently produced reliable outcomes containing few discrepancies with expert clinical judgment would increase trust, while contradictions would undermine trust, particularly when there is a disagreement with the tool's predictions.

Design Recommendations for AI-Driven CDS With XAI

This domain outlines key design considerations for AI-driven CDS tools from the participants, focusing on when and how explanations are presented. It highlights delivering actionable, contextualized information through clear, concise explanations that focus on physiological trends and patient history. While centered on XAI-supported CDS, the findings also underscore the importance of communicating prediction certainty and potential biases to enhance decision-making without overwhelming users.

This means that insights must be actionable, focusing on physiological trends rather than routine metrics, and offering timely and relevant data. A detailed analysis of potential deterioration causes would be preferred over just alerts, with the potential to suggest actions such as specific tests to order. The ability to integrate additional variables could better predict outcomes, such as mental state deterioration, past clinical history, environmental factors, laboratory values, historical data, and, in pediatric settings, parental concerns.

From a user interface perspective, AI tools with concise and customizable presentation options would allow for variations in experience and complexity. Clear, fact-based explanations using bullet points or likelihood percentages were optimal for busy critical care settings, with additional guidance available if required. The provision of weighted graphs, CIs, graded risk indicators, and visual cues such as traffic-light systems allowed clinicians to assess and mitigate risk.

Discussion

Principal Findings

This study explored how explainability influences clinicians' trust, interpretation, and use of AI-driven CDS tools in the context of Australian critical care settings, with particular attention to how clinicians respond when different explainability methods produce different explanations for the same AI prediction. While trust and explainability have been explored in prior research [1,27,29,30,32,47-50], this study contributes novel insights by addressing the underexamined issue of discrepancies between explanation methods, a phenomenon increasingly recognized in technical XAI research [37,42-46] but not studied from a human-centered or clinical perspective. Our findings highlighted a broad range of factors that influence clinicians' use of and trust in AI-driven CDS tools, including how explanations are perceived, applied, and deemed useful, all of which directly informed key design and implementation considerations for XAI-driven CDS systems. The study also offers a novel contribution to the understanding of how explainability influences the determinants for establishing initial trust and overtime trust in AI-driven tools. Specifically, we delineate the positioning of explainability in supporting both perceived usefulness (for initial trust) and learned usefulness (for overtime trust) [58].

The study investigated the perceptions of Australian clinical experts on using explanations as part of their clinical decision-making and how explanations impact trust in XAI-driven CDS. The findings directly address RQ1 by answering how explainability influences clinicians' use, trust, and perceived usefulness of AI-driven CDS tools. The results showed that explainability plays a dynamic and context-dependent role in shaping clinicians' interactions with these systems. Consistent with the literature [29,59], the clinical experts in our study agreed that explanations are a valuable feature that enhances trust and the usefulness of AI-driven CDS tools. However, the necessity and value of explanations vary depending on the context and the clinician's experience. The importance of context-based explanations was similarly highlighted in the findings of Barda et al [60]. For more fundamental decision-making (such as bed allocation, ordering routine blood tests, and administering pain relief), explanations are often unnecessary and can be viewed as an optional resource. Additionally, experienced clinicians may rely more on their own clinical intuition and use explanations less frequently, while less experienced clinicians may prefer more detailed explanations to guide their decision-making. A similar finding was reported by Panigutti et al [49], where the explanation interface was found to be valuable in helping novices avoid mistakes and supporting collaborative decision-making tasks.

Drawing upon the integrative trust model proposed by Cabiddue et al [58], which identified perceived usefulness as a crucial determinant of initial trust in AI-driven tools, our findings reveal a stronger need for explainability during this nascent stage. Explainability is shown to be particularly instrumental in supporting perceived usefulness, especially when users are developing familiarity with the tool. Furthermore, the value of explainability appears to be moderated by user expertise. Novice users, who lack extensive clinical experience, may find explanations more useful in establishing initial trust compared to their expert counterparts. Conversely, the model highlighted learned usefulness as the key determinant for establishing overtime trust. A key finding of our study is that the importance of explainability diminishes overtime as users transition from perceived usefulness to learned usefulness. For expert users, in particular, the significance of explainability decreases notably once they have established confidence in the tool's predictions and rely more heavily on their clinical judgment. Once this level of trust is solidified, the contribution of explainability to learned usefulness becomes less pronounced.

The study further investigated whether partial or complete discrepancies between different XAI methods affect the trust and perceived usefulness of the tool. Participants viewed AI recommendations as an adjunct in decision-making and were not concerned by discrepancies in explanations, prioritizing their own patient assessments. They stressed the need for accurate AI predictions to maintain trust and avoid alarm fatigue.

XSL•FO
RenderX

Explanations were found most helpful when they aligned with clinicians' mental models and the patient's holistic condition. However, participants emphasized that the initial prognostic or diagnostic model predictions must be accurate, as too many false positives would erode trust in the system and lead to alarm fatigue. Alignment with clinicians' mental models is crucial for adopting AI-driven tools [1], as explanations that match these models foster trust, while large discrepancies can cause skepticism. Our study found that discrepancies in explanations enabled reflection through skepticism [59], a fundamental concept in human-centered XAI that encourages critical thinking over passive acceptance when evaluating explanations and building trust in the system [61]. In addition, clinicians often face complex decisions where different approaches yield the same outcome; explanations aligning with their thinking, despite discrepancies in key factors, did not significantly affect trust. However, repeated poor performance in predictions and explanations can lead to loss of trust, consistent with prior findings linking high AI performance to greater acceptance [62,63]. While previous studies have explored XAI's usefulness [64-67] and compared AI's prediction and clinical judgments [19,67,68], few have focused on how explanation discrepancies impact clinician decision-making. Addressing RQ2, our findings show that clinicians may tolerate some level of discrepancies between explanation methods as long as the AI prediction is accurate and the explanations remain clinically meaningful. Rather than dismissing the system, they respond evaluatively, using discrepancies to scrutinize the output, while maintaining trust when the system's outputs are accurate, actionable, and aligned with their clinical reasoning.

Finally, the study identified several design recommendations for successful use of XAI systems in clinical settings. Explanations should focus on actionable data, emphasizing physiological trends (eg, respiratory rate and blood pressure) over routine metrics such as length of stay. Instead of a single metric, the trend over a specific period of time should be shown to assist in making specific clinical decisions. Providing actionable recommendations, such as potential diagnostic triggers, was also suggested by some participants, a finding that was also echoed in related studies [60,69]. Alongside the explanation when a prediction is generated, AI-driven systems should incorporate contextual information on the patient's clinical history and clinical complexity. This is consistent with the literature that explored factors influencing medical experts' use of AI-driven CDS systems [1,70,71]. Multiple studies identified that a lack of a holistic perspective of the patient in the CDS recommendations provided hindered the system's adoption in the health care settings [50,70]. In adult settings, information on comorbidities and previous admissions provides valuable insights, while in pediatric settings, incorporating concerns from both parents and clinicians offers a more holistic view of the patient's condition. Some additional contexts that could be included are the underlying causes of physiological changes, including medication effects, movements, and so on. Providing contextual information along with explanations aligns with the mental model of collaborative practice, enabling colleagues to deliberate and discuss the broader context of a patient's condition to make informed clinical decisions [19].

Explanations need to be simple, concise, and presented in an optional format to avoid overwhelming clinicians. Some ways of visualization include using bullet points, likelihood percentages, and summaries to communicate essential information effectively. The level of detail should be customizable, enabling clinicians of different expertise to adjust the information shown according to their preferences. Ensuring that explanations are digestible would allow clinicians to focus on critical insights without sifting through lengthy text, with additional background and secondary details accessible on demand [72]. Aligned with our findings to prioritize accurate predictions, CDS systems should clearly convey prediction certainty and explanations, along with disclaimers about potential biases. This can be achieved by including CIs or graded risk indicators, alongside visual aids such as weighting graphs on both the predictions and corresponding explanations. Displaying AI confidence scores has been found to enhance trust calibration, leading people to trust the AI more when it shows higher confidence [27,73]. Predictions and their accompanying explanations should be presented within an optimal timeframe to ensure they are actionable, as alerts that come too early or too late can heighten clinician anxiety. The appropriate timing will vary depending on the hospital setting, patient cohort, and available resources, making it crucial to determine this in consultation with expert clinicians. Together, these design considerations demonstrate how explainability, when thoughtfully implemented, underpins trustworthy, context-aware, and user-centered integration of AI-driven CDS tools in clinical settings.

Our study strengthens the human-centered XAI literature by providing evidence from the Australian health care context, addressing a gap in prior research that has predominantly been conducted internationally [1,29,50,59,60,64,65,69,71,74]. We find that key insights from the state-of-the-art literature—such as the value of actionable, context-based explanations, and the variation in reliance on AI by clinician experience—are also relevant in Australian settings. Additionally, the analysis reveals new insights into clinicians' views and tolerance for discrepancies between explanations. Finally, our results highlight local nuances, including workflow considerations and concerns about overreliance in high-pressure environments, offering practical guidance for the design and deployment of AI-driven CDS tools with XAI in Australia.

Limitations and Future Work

One limitation of our study is the composition of the participant cohort. ICU settings are among the most critical in health care, where timely and precise decision-making is vital. This focus motivated our approach and guided our selection of clinical experts from this environment. Nonetheless, many of our findings, particularly those related to clinician trust, explanation usefulness, and balancing guidance with overreliance, may also apply to other clinical settings such as emergency departments or surgical units. However, it is important to note that differences in workflow, patient acuity, and team structures could influence how AI-driven CDS tools are used. Though our findings may apply to other domains with certain domain adaptation, further research is needed to assess the generalizability of our results beyond the ICU. Future studies

XSL•FO
RenderX

could examine perspectives from these additional settings and incorporate participants from other countries outside Australia to provide a more comprehensive understanding. Due to the study's qualitative nature, the findings should be interpreted with caution when considering generalizability. It is important to acknowledge that the scenarios used in this study were based on a desktop prototype, which may have constrained participants' ability to fully consider the scenarios within the context of real-world decision-making environments. Future research could explore the use of more interactive and immersive prototypes, such as those integrated into clinical workflows or simulated environments, to better capture the nuances of decision-making in practice. These future studies could include large-scale usability studies to assess the impact of these actionable insights and further support our qualitative conclusions. Further research should explore scenarios in which the CDS provides plausible yet conflicting predictions relative to clinicians' expected reasoning. Examining such cases would offer deeper insights into how explanations influence trust calibration and decision-making under conditions of disagreement. Future research could also focus on validating the design recommendations for AI-driven CDS tools, ensuring they are adaptable to various clinical scenarios and effectively address trust, safety, and usability concerns, thereby supporting widespread and ethical adoption in health care. This study contributed to the development of a checklist for XAI tools [75]. Evaluating explanations in development and trial phases using the checklist could help enhance trust, reduce risks, foster AI adoption, and determine the true clinical potential of applied XAI.

Conclusions

This study identified recommendations for developing trustworthy, user-centric XAI-supported CDS tools. Three key domains were identified, emphasizing the importance of trust and explainability in gaining clinician acceptance of AI tools. It also highlighted that discrepancies between different explanation techniques are not inherently problematic, provided the explanations are consistent with clinicians' mental model. While developing and deploying algorithms is relatively low-cost, the greater challenge lies in integrating them into clinical workflows, requiring acceptance by health care professionals. By examining user needs, our findings support the development of clinician-informed, actionable insights throughout the AI tool lifecycle. This approach can enhance trust, mitigate risks, and promote sustainable health care by aligning AI systems with clinical workflows and improving long-term patient outcomes through informed and effective use.

Acknowledgments

The authors would like to thank all clinical experts who participated in this study and gave their valuable opinions. This study was funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Responsible Innovation Future Science Platform.

Funding

This project was funded by the CSIRO Responsible Innovation Future Science Platform and the Australian e-Health Research Centre.

Disclaimer

The authors used Microsoft Co-pilot to assist with language editing and improving clarity. All content was reviewed and verified by the authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Semistructured interview guide.
[DOCX File, 57 KB - humanfactors_v13i1e81460_app1.docx]

Multimedia Appendix 2
Detailed scenarios presented to participants in the semistructured interview.
[PDF File, 717 KB - humanfactors_v13i1e81460_app2.pdf]

Multimedia Appendix 3
Additional qualitative findings that did not directly address the research question.
[DOCX File, 256 KB - humanfactors_v13i1e81460_app3.docx]

Checklist 1
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.

XSL•FO
RenderX

[DOCX File, 55 KB - humanfactors_v13i1e81460_app4.docx]

References

1. Cai CJ, Winter S, Steiner D, Wilcox L, Terry M. “Hello AI”: uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. Proc ACM Hum-Comput Interact 2019 Nov 7;3(CSCW):1-24. [doi: 10.1145/3359206]
2. Liu CF, Chen ZC, Kuo SC, Lin TC. Does AI explainability affect physicians’ intention to use AI? Int J Med Inform 2022 Dec;168:104884. [doi: 10.1016/j.ijmedinf.2022.104884]
3. Zhang S, Yu J, Xu X. Rethinking human-AI collaboration in complex medical decision making: a case study in sepsis diagnosis. 2024 Presented at: Proceedings of the CHI Conference on Human Factors in Computing Systems; May 11-16, 2014. [doi: 10.1145/3613904.3642343]
4. Ahmadi A. Artificial intelligence revolution: a comprehensive review of its transformative impact on hospital data management in the future. Int J BioLife Sci 2024;3(3):115-133. [doi: 10.22034/ijbls.2024.199024]
5. Elhaddad M, Hamam S. AI-driven clinical decision support systems: an ongoing pursuit of potential. Cureus 2024 Apr;16(4):e57728. [doi: 10.7759/cureus.57728] [Medline: 38711724]
6. Sharma V, Jamwal A, Agrawal R, Pratap S. A review on digital transformation in healthcare waste management: applications, research trends and implications. Waste Manag Res 2025 Jun;43(6):828-849. [doi: 10.1177/0734242X241285420] [Medline: 39352741]
7. Soori M, Jough FKG, Dastres R, Arezoo B. AI-based decision support systems in Industry 4.0, a review. JET 2026;4:206-225. [doi: 10.1016/j.ject.2024.08.005]
8. Malviya R, Sundram S, Dhanaraj RK, Kadry S, editors. Digital Transformation in Healthcare 50: Volume 1: IoT, AI and Digital Twin: De Gruyter; 2024.
9. Giordano C, Brennan M, Mohamed B, Rashidi P, Modave F, Tighe P. Accessing artificial intelligence for clinical decision-making. Front Digit Health 2021;3:645232. [doi: 10.3389/fdgth.2021.645232] [Medline: 34713115]
10. Khalifa M, Albadawy M. Artificial intelligence for clinical prediction: exploring key domains and essential functions. Comput Methods Programs Biomed Update 2024;5:100148. [doi: 10.1016/j.cmpbup.2024.100148]
11. Pinsky MR, Bedoya A, Bihorac A, et al. Use of artificial intelligence in critical care: opportunities and obstacles. Crit Care 2024;28(1):113. [doi: 10.1186/s13054-024-04860-z]
12. Kamel Rahimi A, Pienaar O, Ghadimi M, et al. Implementing AI in hospitals to achieve a learning health system: systematic review of current enablers and barriers. J Med Internet Res 2024 Aug 2;26:e49655. [doi: 10.2196/49655] [Medline: 39094106]
13. Yin J, Ngiam KY, Teo HH. Role of artificial intelligence applications in real-life clinical practice: systematic review. J Med Internet Res 2021 Apr 22;23(4):e25759. [doi: 10.2196/25759] [Medline: 33885365]
14. Singh RP, Hom GL, Abramoff MD, Campbell JP, Chiang MF, AAO Task Force on Artificial Intelligence. Current challenges and barriers to real-world artificial intelligence adoption for the healthcare system, provider, and the patient. Transl Vis Sci Technol 2020 Aug;9(2):45. [doi: 10.1167/tvst.9.2.45] [Medline: 32879755]
15. Antoniadi AM, Du Y, Guendouz Y, et al. Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: a systematic review. Appl Sci (Basel) 2021;11(11):5088. [doi: 10.3390/app11115088]
16. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons for physicians not adopting clinical decision support systems: critical analysis. JMIR Med Inform 2018 Apr 18;6(2):e24. [doi: 10.2196/medinform.8912] [Medline: 29669706]
17. Kohli A, Jha S. Why CAD failed in mammography. J Am Coll Radiol 2018 Mar;15(3 Pt B):535-537. [doi: 10.1016/j.jacr.2017.12.029] [Medline: 29398499]
18. Taneva S, Sara W, Julian G, Peter R, Emily N, Joseph C. The meaning of design in healthcare: industry, academia, visual design, clinician, patient and hf consultant perspectives. 2014 Presented at: CHI’14 Extended Abstracts on Human Factors in Computing Systems; Apr 26 to May 1, 2014. [doi: 10.1145/2559206.2579407]
19. Yang Q, Steinfeld A, Zimmerman J. Unremarkable AI: fitting intelligent decision support into critical, clinical decision-making processes. 2019 Presented at: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, United Kingdom p. 1-11. [doi: 10.1145/3290605.3300468]
20. Čartolovni A, Tomičić A, Lazić Mosler E. Ethical, legal, and social considerations of AI-based medical decision-support tools: a scoping review. Int J Med Inform 2022 May;161:104738. [doi: 10.1016/j.ijmedinf.2022.104738] [Medline: 35299098]
21. Kasperbauer TJ. Conflicting roles for humans in learning health systems and AI-enabled healthcare. J Eval Clin Pract 2021 Jun;27(3):537-542. [doi: 10.1111/jep.13510] [Medline: 33164284]
22. Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable AI: a review of machine learning interpretability methods. Entropy (Basel) 2021;23(1):18. [doi: 10.3390/e23010018]
23. Carter SM, Rogers W, Win KT, Frazer H, Richards B, Houssami N. The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. Breast 2020 Feb;49:25-32. [doi: 10.1016/j.breast.2019.10.001] [Medline: 31677530]
24. Gaube S, Suresh H, Raue M, et al. Do as AI say: susceptibility in deployment of clinical decision-aids. NPJ Digit Med 2021 Feb 19;4(1):31. [doi: 10.1038/s41746-021-00385-9] [Medline: 33608629]

XSL•FO
RenderX

25. Kaur H, Adar E, Gilbert E, Lampe C. Sensible AI: re-imagining interpretability and explainability using sensemaking theory. 2022 Presented at: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency; Jun 21-24, 2022; Seoul, South Korea p. 702-714. [doi: 10.1145/3531146.3533135]
26. Sivaraman V, Bukowski LA, Levin J, Kahn JM, Perer A. Ignore, trust, or negotiate: understanding clinician acceptance of AI-based treatment recommendations in health care. : ACM; 2023 Presented at: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems CHI '23; Apr 23-28, 2023; Hamburg, Germany p. 1-18. [doi: 10.1145/3544548.3581075]
27. Zhang Y, Liao QV, Bellamy RKE. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. : ACM; 2020 Presented at: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency FAT* '20; Jan 27-30, 2020; Barcelona, Spain p. 295-305. [doi: 10.1145/3351095.3372852]
28. Stephanidis C, Salvendy G, Antona M, et al. Seven HCI grand challenges. Int J Hum Comput Interact 2019 Aug 27;35(14):1229-1269. [doi: 10.1080/10447318.2019.1619259]
29. Schmidt P, Biessmann F, Teubner T. Transparency and trust in artificial intelligence systems. J Decis Syst 2020 Oct 1;29(4):260-278. [doi: 10.1080/12460125.2020.1819094]
30. Goebel R, Chander A, Holzinger K, et al. Explainable AI: the new 42? 2018 Presented at: International Cross-Domain Conference for Machine Learning and Knowledge Extraction; Aug 27-30, 2018 URL: https://inria.hal.science/hal-01934928v1/file/cdMake18-XAI-Lecue-et-al.pdf [accessed 2026-04-25]
31. Kostopoulos G, Davrazos G, Kotsiantis S. Explainable artificial intelligence-based decision support systems: a recent review. Electronics (Basel) 2024;13(14):2842. [doi: 10.3390/electronics13142842]
32. Wang D, Yang Q, Abdul A, Lim BY. Designing theory-driven user-centric explainable AI. Presented at: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, United Kingdom p. 1-15. [doi: 10.1145/3290605.3300831]
33. Wang X, Yin M. Are Explanations Helpful? A Comparative Study of the Effects of Explanations in AI-Assisted Decision-Making. 2021 Apr 14 Presented at: IUI '21; Apr 14-17, 2021; College Station TX USA p. 318-328 URL: https://dl.acm.org/doi/proceedings/10.1145/3397481 [doi: 10.1145/3397481.3450650]
34. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 2020 Jun;58:82-115. [doi: 10.1016/j.inffus.2019.12.012]
35. Chari S, Acharya P, Gruen DM, et al. Informing clinical assessment by contextualizing post-hoc explanations of risk prediction models in type-2 diabetes. Artif Intell Med 2023 Mar;137:102498. [doi: 10.1016/j.artmed.2023.102498] [Medline: 36868690]
36. Brankovic A, Cook D, Rahman J, Khanna S, Huang W. Benchmarking the most popular XAI used for explaining clinical predictive models: untrustworthy but could be useful. Health Informatics J 2024;30(4):14604582241304730. [doi: 10.1177/14604582241304730] [Medline: 39716973]
37. Krishna S, Han T, Gu A, Wu S, Jabbari S, Lakkaraju H. The disagreement problem in explainable machine learning: a practitioner's perspective. arXiv. Preprint posted online on Feb 3, 2024. [doi: 10.48550/arXiv.2202.01602]
38. Slack D, Hilgard A, Singh S, Lakkaraju H. Reliable post hoc explanations: modeling uncertainty in explainability. 2021 Presented at: Proceedings of the 35th International Conference on Neural Information Processing Systems; Dec 6-14, 2021; Virtual p. 9391-9404. [doi: 10.5555/3540261.3540980]
39. Bussone A, Stumpf S, O'Sullivan D. The role of explanations on trust and reliance in clinical decision support systems. 2015 Presented at: 2015 International Conference on Healthcare Informatics (ICHI); Oct 21-23, 2015; Dallas, TX, USA p. 160-169. [doi: 10.1109/ICHI.2015.26]
40. Chromik M, Eiband M, Buchner F, Krüger A, Butz A. I think i get your point, AI! The illusion of explanatory depth in explainable AI. 2021 Apr 14 Presented at: Proceedings of the 26th International Conference on Intelligent User Interfaces; Apr 14-17, 2021; College Station, TX p. 307-317. [doi: 10.1145/3397481.3450644]
41. Cabitza F, Fregosi C, Campagner A, Natali C. Explanations considered harmful: the impact of misleading explanations on accuracy in hybrid human-AI decision making. In: World Conference on Explainable Artificial Intelligence: Springer; 2024:255-269. [doi: 10.1007/978-3-031-63803-9_14]
42. Brughmans D, Melis L, Martens D. Disagreement amongst counterfactual explanations: how transparency can be misleading. TOP 2024 Oct;32(3):429-462. [doi: 10.1007/s11750-024-00670-2]
43. Koenen N, Wright MN. Toward understanding the disagreement problem in neural network feature attribution. In: World Conference on Explainable Artificial Intelligence: Springer; 2024:247-269. [doi: 10.1007/978-3-031-63800-8_13]
44. Renard X, Laugel T, Detyniecki M. Understanding prediction discrepancies in classification. Mach Learn 2024 Oct;113(10):7997-8026. [doi: 10.1007/s10994-024-06557-4]
45. Roy S, Laberge G, Roy B, Khomh F, Nikanjam A, Mondal S. Why don't XAI techniques agree? characterizing the disagreements between post-hoc explanations of defect predictions. 2022 Presented at: 2022 IEEE International Conference on Software Maintenance and Evolution (ICSME); Oct 3-7, 2022; Limassol, Cyprus p. 444-448. [doi: 10.1109/ICSME55016.2022.00056]

XSL•FO
RenderX

46. Verma S, Boonsanong V, Hoang M, Hines K, Dickerson J, Shah C. Counterfactual explanations and algorithmic recourses for machine learning: a review. ACM Comput Surv 2024 Dec 31;56(12):1-42. [doi: 10.1145/3677119]
47. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 2020 Nov 30;20(1):310. [doi: 10.1186/s12911-020-01332-6] [Medline: 33256715]
48. Liao QV, Sundar SS. Designing for responsible trust in AI systems: a communication perspective. 2022 Jun 21 Presented at: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency; Jun 21-24, 2022; Seoul, South Korea p. 1257-1268 URL: https://dl.acm.org/doi/proceedings/10.1145/3531146 [accessed 2026-04-25] [doi: 10.1145/3531146.3533182]
49. Panigutti C, Beretta A, Giannotti F, Pedreschi D. Understanding the impact of explanations on advice-taking: a user study for AI-based clinical decision support systems. 2022 Apr 29 Presented at: CHI Conference on Human Factors in Computing Systems; Apr 30 to May 6, 2022; New Orleans, LA p. 1-9. [doi: 10.1145/3491102.3502104]
50. Yang Q, Hao Y, Quan K. Harnessing biomedical literature to calibrate clinicians' trust in AI decision support systems. 2023 Presented at: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems CHI '23; Apr 23-28, 2023. [doi: 10.1145/3544548.3581393]
51. Morrison K, Spitzer P, Turri V, Feng M, Kühl N, Perer A. The impact of imperfect XAI on human-AI decision-making. Proc ACM Hum-Comput Interact 2024 Apr 17;8(CSCW1):1-39. [doi: 10.1145/3641022]
52. Coiera E, Magrabi F. A national policy roadmap for artificial intelligence in healthcare. Australian Alliance for Artificial Intelligencein Healthcare. 2023. URL: https://figshare.com/articles/conference_contribution/A_National_Policy_Roadmap_for_Artificial_Intelligence_in_Healthcare/28646342?file=53167781 [accessed 2026-05-01]
53. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015 Sep;42(5):533-544. [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
54. Moser A, Korstjens I. Series: practical guidance to qualitative research. Part 3: sampling, data collection and analysis. Eur J Gen Pract 2018 Dec;24(1):9-18. [doi: 10.1080/13814788.2017.1375091] [Medline: 29199486]
55. Schwartz JM, George M, Rossetti SC, et al. Factors influencing clinician trust in predictive clinical decision support systems for in-hospital deterioration: qualitative descriptive study. JMIR Hum Factors 2022 May 12;9(2):e33960. [doi: 10.2196/33960] [Medline: 35550304]
56. Brankovic A, Huang W, Cook D, Khanna S, Bialkowski K. Elucidating discrepancy in explanations of predictive models developed using EMR. Stud Health Technol Inform 2024 Jan 25;310:865-869. [doi: 10.3233/SHTI231088] [Medline: 38269932]
57. Glaser B, Strauss A. Discovery of Grounded Theory: Strategies for Qualitative Research: Routledge; 2017. [doi: 10.4324/9780203793206]
58. Cabiddu F, Moi L, Patriotta G, Allen DG. Why do users trust algorithms? A review and conceptualization of initial trust and trust over time. Eur Manag J 2022 Oct;40(5):685-706. [doi: 10.1016/j.emj.2022.06.001]
59. Solano-Kamaiko IR, Mishra D, Dell N, Vashistha A. Explorable explainable AI: improving AI understanding for community health workers in India. 2024 May 11 Presented at: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems CHI '24; May 11-16, 2024; Honolulu, HI p. 1-21. [doi: 10.1145/3613904.3642733]
60. Barda AJ, Horvat CM, Hochheiser H. A qualitative research framework for the design of user-centered displays of explanations for machine learning model predictions in healthcare. BMC Med Inform Decis Mak 2020 Oct 8;20(1):257. [doi: 10.1186/s12911-020-01276-x] [Medline: 33032582]
61. Ehsan U, Riedl MO. Human-centered explainable AI: towards a reflective sociotechnical approach. Presented at: HCI International 2020-Late Breaking Papers: Multimodality and Intelligence: 22nd HCI International Conference, HCII 2020; Jul 19-24, 2020; Copenhagen, Denmark p. 449-466. [doi: 10.1007/978-3-030-60117-1_33]
62. Lai V, Tan C. On human predictions with explanations and predictions of machine learning models: a case study on deception detection. 2019 Jan 29 Presented at: Proceedings of the Conference on Fairness, Accountability, and Transparency; Jan 29-31, 2019; Atlanta, GA p. 29-38 URL: https://dl.acm.org/doi/proceedings/10.1145/3287560 [accessed 2026-04-25] [doi: 10.1145/3287560.3287590]
63. Yin M, Wortman Vaughan J, Wallach H. Understanding the effect of accuracy on trust in machine learning models. 2019 May 2 Presented at: CHI '19; May 4-9, 2019; Glasgow Scotland Uk p. 1-12 URL: https://dl.acm.org/doi/proceedings/10.1145/3290605 [accessed 2026-04-25] [doi: 10.1145/3290605.3300509]
64. Humer C, Hinterreiter A, Leichtmann B, Mara M, Streit M. Reassuring, misleading, debunking: comparing effects of XAI methods on human decisions. ACM Trans Interact Intell Syst 2024 Sep 30;14(3):1-36. [doi: 10.1145/3665647]
65. Panigutti C, Beretta A, Fadda D, et al. Co-design of human-centered, explainable AI for clinical decision support. ACM Trans Interact Intell Syst 2023 Dec 31;13(4):1-35. [doi: 10.1145/3587271]
66. Wolf CT. Explainability scenarios: towards scenario-based XAI design. 2019 Presented at: Proceedings of the 24th International Conference on Intelligent User Interfaces; Mar 17-20, 2019; Marina del Ray, CA p. 252-257. [doi: 10.1145/3301275.3302317]

XSL•FO
RenderX

67. Wysocki O, Davies JK, Vigo M, et al. Assessing the communication gap between AI models and healthcare professionals: explainability, utility and trust in AI-driven clinical decision-making. Artif Intell 2023 Mar;316:103839. [doi: 10.1016/j.artint.2022.103839] [Medline: 41550460]
68. Kayser MG, Menzat B, Emde C, et al. Fool me once? Contrasting textual and visual explanations in a clinical decision-support setting. In: Al-Onaizan Y, Bansal M, Chen YN, editors. Presented at: Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing; Nov 12-14, 2024; Miami, FL p. 18891-18919 URL: https://aclanthology.org/2024.emnlp-main [accessed 2026-04-25] [doi: 10.18653/v1/2024.emnlp-main.1051]
69. Bhattacharya A. Towards directive explanations: crafting explainable AI systems for actionable human-AI interactions. 2024 May 11 Presented at: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems; May 11-16, 2024; Honolulu, HI p. 1-6 URL: https://dl.acm.org/doi/proceedings/10.1145/3613905 [accessed 2026-04-25] [doi: 10.1145/3613905.3638177]
70. Kilsdonk E, Peute LW, Jaspers MWM. Factors influencing implementation success of guideline-based clinical decision support systems: a systematic review and gaps analysis. Int J Med Inform 2017 Feb;98:56-64. [doi: 10.1016/j.ijmedinf.2016.12.001] [Medline: 28034413]
71. Yildirim N, Richardson H, Wetscherek MT, et al. Multimodal healthcare AI: identifying and designing clinically relevant vision-language applications for radiology. 2024 Presented at: Proceedings of the CHI Conference on Human Factors in Computing Systems; May 11-16, 2024; Honolulu, HI p. 1-22 URL: https://dl.acm.org/doi/proceedings/10.1145/3613904 [accessed 2026-04-25] [doi: 10.1145/3613904.3642013]
72. Payne TH, Hines LE, Chan RC, et al. Recommendations to improve the usability of drug-drug interaction clinical decision support alerts. J Am Med Inform Assoc 2015 Nov;22(6):1243-1250. [doi: 10.1093/jamia/ocv011] [Medline: 25829460]
73. Rechkemmer A, Yin M. When confidence meets accuracy: exploring the effects of multiple performance indicators on trust in machine learning models. 2022 Presented at: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; Apr 30 to May 6, 2022; New Orleans, LA p. 1-14 URL: https://dl.acm.org/doi/proceedings/10.1145/3491102 [accessed 2026-04-25] [doi: 10.1145/3491102.3501967]
74. Jacobs M, He J, F. Pradier M, et al. Designing AI for trust and collaboration in time-constrained medical decisions: a sociotechnical lens. 2021 May 6 Presented at: CHI '21; May 8-13, 2021; Yokohama Japan p. 1-14 URL: https://dl.acm.org/doi/proceedings/10.1145/3411764 [doi: 10.1145/3411764.3445385]
75. Brankovic A, Cook D, Rahman J, et al. Clinician-informed XAI evaluation checklist with metrics (CLIX-M) for AI-powered clinical decision support systems. NPJ Digit Med 2025 Jun 14;8(1):364. [doi: 10.1038/s41746-025-01764-2] [Medline: 40517152]

Abbreviations

AI: artificial intelligence
CDS: clinical decision support
COREQ: Consolidated Criteria for Reporting Qualitative Research
CSIRO: Commonwealth Scientific and Industrial Research Organisation
ICU: intensive care unit
REDCap: Research Electronic Data Capture
RQ: research question
XAI: explainable artificial intelligence

Edited by A Choudhury; submitted 30.Jul.2025; peer-reviewed by D Corsar, D Fotopoulos, M Elbattah; revised version received 02.Feb.2026; accepted 19.Feb.2026; published 01.May.2026.

Please cite as:
Rahman J, Delaforce A, Bradford D, Li J, Magrabi F, Cook D, Brankovic A
The Role of Explanations in AI-Generated Alerts: Qualitative Study of Clinical Views on Explainable AI in Predictive Tools
JMIR Hum Factors 2026;13:e81460
URL: https://humanfactors.jmir.org/2026/1/e81460
doi:10.2196/81460

© Jessica Rahman, Alana Delaforce, DanaKai Bradford, Jane Li, Farah Magrabi, David Cook, Aida Brankovic. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 1.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly

XSL•FO
RenderX

cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Challenges and Strategies of Managing Mental Health Medication in Real-World Practices: Interview Study Among University Students

Jiachen Li, MS; Justin Steinberg, MS; Elizabeth Mynatt, PhD; Varun Mishra, PhD

Khoury College of Computer Science, Northeastern University, 360 Huntington Ave, Boston, MA, United States

Corresponding Author:
Jiachen Li, MS
Khoury College of Computer Science, Northeastern University, 360 Huntington Ave, Boston, MA, United States

Abstract

Background: Mental health has become a growing concern among university students. According to the Healthy Minds Study National Report 2022 - 2023, 46% of students have been diagnosed with a mental health condition by a health professional, reflecting a nearly 50% increase from 2013 to 2021. While researchers have developed various technologies to help this issue, one significant aspect of the mental health management journey is often understudied—the role of mental health medication. Understanding how university students manage their medication for mental health symptoms in real-world practice has not been fully explored.

Objective: In this study, we aimed to examine university students' experiences in taking and managing their mental health medications. We want to understand the unique challenges in the mental health medication management process and their coping strategies, particularly examining the role of various technologies in this process.

Methods: We conducted semistructured interviews with 10 participants who are self-identified as currently taking medication for mental health diagnosis and are students at a university. Two researchers (JL and JS) open-coded the interview transcripts and notes and conducted a thematic analysis to identify 3 primary themes, which we detail in the Results section.

Results: We discovered that due to struggles with self-acceptance and the interdependent relationship between medication, symptoms, schedules, and life changes, the medication management process for students was a highly dynamic journey involving frequent dosage changes. Thus, students adopted flexible strategies of using minimal technology to manage their medication in different situations while maintaining a high degree of autonomy. Based on our findings, we propose design implications for future technologies to seamlessly integrate into their daily lives and assist students in managing their mental health medications.

Conclusions: In the paper, we conducted semistructured interviews with university students currently undergoing medication for mental health issues. The interviews aimed to uncover the challenges they encountered in their university lives, along with the coping strategies and technologies used to navigate these challenges. We found that students often rely on creative ways to use fixed alarms on their phones; however, they encounter numerous difficulties due to factors such as the flexibility of daily routines. Based on these findings, we proposed 6 design implications for how advanced technology could better assist university students in managing their mental health medications.

(JMIR Hum Factors 2026;13:e67579) doi:10.2196/67579

KEYWORDS

mental health; qualitative study; personal health informatics; mHealth; medication management

Introduction

Background

Researchers have recently observed that mental health challenges pose a significant concern, particularly among university students. According to the Healthy Minds Study National Report 2022 - 2023, 46% of students have been diagnosed with a mental health condition by a health professional, encompassing depression (30%), anxiety (36%), and other symptoms [1]. The reported data represent the highest rate in the 15-year history of the survey, reflecting a nearly 50% increase from 2013 to 2021 [2]. Recognizing the severity of these issues, many researchers have developed various technologies to assist, including chatbots [3-5], digital therapeutic applications [6,7], and wearables for passive monitoring [8-10]. Although these technologies have demonstrated positive effects, one significant aspect of the mental health management journey is often understudied—the role of mental health medication. According to the report, 40% of students with positive depression or anxiety screens had taken at least one type of medication in the preceding 12 months.

XSL•FO
RenderX

Since these medications often have a significant effect in mitigating and managing mental health symptoms, it is crucial to examine people's experiences in taking and managing these medications.

Despite the large population of university students using medications for mental health conditions, their strategies in managing those medications are relatively understudied. Most prior works on medication management strategies focused on older adults [11,12] and individuals with disabilities [13], often categorizing younger populations as "healthier." With the increase in mental health diagnoses, however, more university students are regularly taking medications to manage their mental health. Further, given their busy schedules and dynamic routines, ensuring adherence among this group is also difficult. For researchers aiming to develop technologies to address this issue, it is important to first understand the real-world practices of individuals in adopting various technologies for mental health medication management, the challenges they encounter during the process, and the hesitations with being more consistent with the medications. Recognizing the significance of this issue, our research delved into the challenges university students face in managing their mental health medication and the coping strategies they use to tackle these issues.

In this paper, we make three important contributions. (1) We pinpoint the unique challenges in the medication management process for university students with mental health conditions. (2) We delve into the coping strategies used by university students in addressing these challenges, with a particular focus on examining the role of various technologies in the process. (3) Based on our findings, we outline and discuss design implications for future systems aimed at assisting university students in managing their mental health medications.

Related Work

Mental Health Issues Among University Students

The prevalence of mental health issues among university students has been escalating, posing a growing concern. According to a recent study, nearly half of all lifetime mental health conditions start by the mid-teens and three-quarters by the mid-20s [14]. The discussion of mental health within university communities underscores the challenges that many students face during their collegiate journeys. The combination of heavy academic workloads, social pressures, and the inherent uncertainties associated with this exploratory phase of life contributes to an environment where mental health issues can manifest [15-17].

In recent years, there has been a discernible uptick in mental health concerns among students, with a notable rise in conditions such as anxiety, depression, eating disorders, and suicidal ideation [1,16]. These issues not only cast shadows over the overall university experience but also impede students' active participation in academic and extracurricular activities, leading to a considerable decline in their overall quality of life [2,18]. Due to the gravity of the situation, there has been a stronger push to spread mental health awareness. Between 2013 and 2023, there has been a substantial 7.8% reduction in the perceived stigma surrounding mental health issues, fostering an environment where individuals are more inclined to seek assistance [1,2]. This reduction in stigma has correlated with a 12.8% increase in the usage of psychotropic medications among university students during the same period [2].

While the surge in medication use reflects a growing acknowledgment of mental health challenges, the issue of mental medication adherence among university students remains underexplored. The efficacy of mental health therapy and psychiatric medication is contingent on consistent adherence to prescribed regimens. Factors such as academic pressures, social dynamics, and lifestyle transitions unique to the university experience may significantly influence medication adherence [19,20], thereby necessitating a nuanced understanding of this critical aspect of mental health treatment.

Mental Health Treatment: The Role of Medications

The treatment of mental illness frequently involves a combination of medication and mental therapy. Medications are typically prescribed when psychotherapies on their own are ineffective [18,21]. Medications play a crucial role in alleviating symptoms by influencing brain chemicals affecting mood and behavior [22]. Commonly prescribed medications include antidepressants, antipsychotics, and mood stabilizers [23]. Antidepressants are used to treat a wide variety of illnesses, such as major depressive disorder, general anxiety disorder, and obsessive-compulsive disorder [24]. When taking antidepressants, the dosage of the medication will help determine the effectiveness of the treatment. However, the higher the dose, the more likely a patient will experience side effects [25].

While psychiatric medications are integral components of mental health treatment, their effectiveness is also contingent on factors such as accurate diagnosis, tailored prescription regimens, and, notably, adherence to prescribed medications [26-28]. Adherence is the execution of a prescribed treatment plan with little to no error. This includes taking the correct medication at the appropriate dose, at the proper time, and with a proper frequency [29]. Adherence to psychiatric medication regimens is a critical determinant of treatment success, influencing the ability to manage symptoms and enhance overall functioning [30,31]. There are many factors that can influence a patient's capacity to adhere to medication, such as age, educational level, attitudes and beliefs toward medication, routines and schedule, stigma, dosage changes, and medication side effects [20,28,32,33]. We contribute to the existing literature by understanding how these factors affect adherence within the university student demographic, specifically for mental health medications.

Technology for Medication Management

In order to combat the issue of medication nonadherence, many studies have tested different types of adherence-related technologies. There is no universally recognized standard in medication adherence technology, making it difficult for patients to identify the most effective solution for their needs. These technologies encompass a wide range of tools such as smart pill bottles, sensors, self-reporting technology, mobile apps, and chatbots [34-37]. Smart pill bottles, for instance, can connect to the internet and record the time when a patient opens the

XSL•FO
RenderX

bottle, providing a timeline of adherence [38]. For mobile phones, there are many different types of modalities that can be used, such as phone calls, SMS text messaging, online surveys, as well as audio and video content [36,37,39,40]. However, depending on the features used, there can be challenges with user engagement, battery life, storage capacity, and ease of use [40].

Notably, the combination of mobile interventions with smart pillboxes has emerged as an effective method in supporting antidepressant treatment [41]. Mobile apps, in this context, serve not only to monitor adherence but also to provide medication education, track side effects and symptoms, and act as a communication channel with health care providers. This serves to address multiple pain points that are related to adherence, such as a lack of knowledge, forgetfulness, and low motivation [41]. This multilayered approach showcases the potential of technology to address medication adherence challenges while offering additional benefits in terms of education, monitoring, and communication within the health care ecosystem.

The exploration of using mobile adherence solutions also extends to artificial intelligence (AI). AI smartphone apps have also been investigated for their effectiveness in delivering reminders and encouragement for medication adherence for adults with noncommunicable diseases [26]. Conversational agents, also known as chatbots, are able to simulate human interaction in text, audio, or visual formats [42]. Conversational AI systems have been shown to help older adults remember to take their medications by tracking their daily activities and symptoms [26]. AI systems are able to collect and interpret data from a wide variety of different data structures, such as from electronic health records, clinical notes, and real-time physiological data in conjunction with sensors [42]. Furthermore, because of an AI system's ability to detect meaningful patterns in data, they also have the ability to provide personalized responses to better support user needs as well as foster patient/provider relationships [26,42]. In addition, the integration of these new technology-based approaches comes with several advantages, including easy implementation, cost-effectiveness, and extensive outreach [27,31]. Our research focuses on discerning effective versus ineffective technologies for medication adherence among university students and understanding how technology-specific factors influence adoption.

Methods

Interview Study Design

Recruitment materials were distributed through public university email lists and social media channels, initially targeting two universities and later expanding through snowball sampling. Participants were required to provide a ".edu" email to verify their student status. The study received approval from the Department of Human Research at Northeastern University (#23-04-19), and no direct contact with universities beyond publicly accessible channels was made. The recruitment materials clearly stated that participation was voluntary and would not affect participants' academic standing. After expressing initial interest, we emailed the consent form to the participants, outlining the study's purpose and content to prepare them for potentially sensitive interview topics before the interview. Students could contact the research team for additional information or consultation before providing consent, given the potentially sensitive nature of the study. If the participants wanted to proceed, we scheduled the virtual semistructured interviews through Zoom, which typically lasted an hour. After eligibility screening, 12 students were contacted for scheduling, and 10 completed both the survey and the interview. Two members of the research team were present for all interviews; one researcher led the interview while the other documented the discussion. We obtained audio recordings and transcripts only after the participants' initial consent. Each participant received a US $20 Amazon gift card via email after the interview. Given the sensitive subject of the interview, we provided participants information about the resources at the university's Health and Counseling center, and also provided them with the university's 24-hour helpline number to connect with a mental health clinician who is there to listen, guide, and help.

The interviews primarily revolved around each participant's unique journey managing their mental health medication. We delved into their medication schedules, ranging from basic details like frequency to deeper explorations of personal challenges. Our focus also extended to the technologies used during this process, exploring both successful and challenging experiences. Acknowledging the individuality of each journey, we let the participant drive the interview and describe their experience in the later part of the interview, with some guiding prompts like, "Describe your most recent experience taking medication, such as today or yesterday." Researchers dynamically adjusted the interview questions based on each participant's unique challenges, aiming to capture a comprehensive understanding of their personal experiences in managing medications within real-world scenarios.

In addition to the interviews, we requested the participants to complete an online survey. This survey collected demographic information, basic medication information, and 1 scale measuring self-efficacy in medication use (SEAMS) [43]. The SEAMS is a 13-item scale assessing patients' confidence in taking medications as prescribed, scored on a 3-point Likert scale ranging from 13 to 39, with higher scores indicating greater self-efficacy. The survey was deidentified using participant IDs to protect anonymity.

Participants

In this study, we conducted 10 semistructured interviews with participants who are self-identified as currently taking medication for mental health conditions and are students at a university. We used the participants' university email addresses to verify their eligibility for the study.

Data Analysis

We carefully reviewed the interview transcripts and notes, adopting an open coding approach [44]. The data were coded independently by two researchers (JL and JS), with each focusing on specific segments. Subsequently, both researchers exchanged their individual codebooks, engaging in discussions

XSL•FO
RenderX

to harmonize and merge the codes by consensus. Following this collaborative process, we conducted a thematic analysis, adhering to the principles of grounded theory [45,46]. We reached data saturation with 10 participants and stopped further recruitment [47,48]. This analysis resulted in the identification of 3 primary themes, which we detail in the Results section.

Ethical Considerations

The study obtained the institutional review board approval from the Department of Human Research at Northeastern University (#23-04-19). Informed consent from participants was obtained before each interview. Given the sensitive nature of the data collected, personal identification was immediately removed postinterview. Participants were given the autonomy to self-report their diagnosis voluntarily and retained the right to withhold this information if they chose. Realizing the interview topic's sensitivity and the potential risk of disclosing private information, we ensured that recruitment of new participants ceased as soon as we reached data saturation [47,48]. Each participant received a US $20 Amazon gift card via email after the interview.

Results

Participants' Demographics

Two of our participants identified as male, while 8 identified as female (Table 1). In terms of racial identification, 5 participants identified as White, 1 as Hispanic, 1 as Black or African American, 1 as Asian, and 2 as Other. Among the participants, 3 were enrolled in undergraduate school, and the remaining 7 were attending graduate school. On average, participants reported taking 1.6 (SD 0.7) types of medications. All participants took medications at least once a day, with one individual taking additional medications weekly and another multiple times a day (Table 2). Regarding medication storage, 8 out of 10 participants only used the original pill bottle or blister; one used a 7-day drugstore-style pillbox, and another used both methods. In terms of technology used in the medication management process, 4 relied solely on memory, 4 used general digital calendar/reminder apps, and 2 used other technologies such as third-party medication and self-report apps. The average score for medication self-efficacy was 25.3 (scale value ranging from 13 as nonefficacious to 39 as efficacious), respectively. These scores suggest that there is no obvious ceiling effect in our sample, indicating that participants do face some challenges in their medication management.

XSL•FO
RenderX

Table . Demographic information of the participants (N=10).

Characteristics	Participants, n
Gender	
Female	8
Male	2
Age (years)	
18 - 24	4
25 - 34	6
Race/ethnicity	
White	5
Hispanic	1
Black or African American	1
Asian	1
Other	2
Education level	
Some college	3
Professional degree	4
Doctorate	3
Marital status	
Never married	8
Married	1
Other	1
Income (in US$)	
Less than 50,000	3
50000 - 59999	3
More than 59,999	3
Blank	1

XSL•FO
RenderX

Table . Baseline information related to mental health and medication (N=10).

Characteristics	Values
Diagnosis length, n	
1 - 3 years	2
3 - 5 years	4
More than 5 years	4
Medication length, n	
1 - 3 years	6
3 - 5 years	0
More than 5 years	1
Not answered	3
Total medication counts, mean (SD)	1.6 (0.7)
Frequency	
Daily (once a day)	10
Weekly (once a week)	1
Multiple times a day	1
Pill containers, n	
Original pill bottle	8
Original blister	1
7-day drug store like pillbox	2
Obtain medication, n	
On-campus health care/clinic	1
Off-campus health care provider	5
Local pharmacy/chemist	2
Online pharmacy services	1
Other	1
Medication management strategies, n	
Establish a stable routine/schedule	8
Digital reminders/alarms	6
Physical reminders/visual cues	5
Self-report on digital devices	2
Families/friends remind me	2
No, I depend on my memory	1
Technology used, n	
Digital calendar/reminder apps	6
Third-party medication reminding apps	1
Digital self-report apps	1
No, I depend on my memory	4
SEAMS[a] self-efficacy (“13=nonefficacious” to “39=efficacious”), mean (SD)	25.3 (4.74)

[a]SEAMS: Self-Efficacy for Appropriate Medication Use Scale.

XSL•FO
RenderX

Theme 1: A Dynamic Voyage: Managing Mental Health Medication Is Never a Static Process

A standout takeaway from all the interviews was the dynamic nature of managing mental health medication. In this section, we aim to illustrate the causes of this dynamic nature and its effects on medication management strategies.

Theme 1.1: Struggling With Self-Acceptance: Expectations of Reducing or Ceasing Medication

We noticed significant concern and hesitation regarding taking medication for mental health concerns, setting apart the expectations for mental health medication from those for treating chronic diseases related to nonmental conditions. Of the 10 participants, 9 did not commence medication immediately after receiving a diagnosis. For instance, Participant 5 had been diagnosed with a mental health condition for 16 years but only began taking medication approximately 10 years ago. This delay in medication initiation underscores that medication is not typically the first treatment choice for mental health challenges.

> *I just wasn't seeking medical help like medical or medicine intervention for a while.* [Participant 4]

Starting medication was often considered "a tough decision" by our participants, generating feelings of anxiety. This anxiety stemmed from various reasons. Initially, many associated taking medication with "having serious issues" and struggled to reconcile this notion within themselves.

> *For example, once you start going on medication, it's like, you have problems. It's more like admitting to yourself that you need help.* [Participant 2]

Moreover, uncertainty about the effects of medication caused anxiety. Participant 2 described embarking on taking medications as "taking a risk" since it is not clear why some people have different symptoms even on the same medications. Concerns also extended to potential long-term effects, particularly as mental health medication "has some influence on our brain" [Participant 8]. Another intriguing factor was the involvement of parents in this decision-making process, especially considering the changing relationships between university students and their parents during this pivotal time. Participants mentioned instances where their parents decided if they could take medication during their teenage years, or they felt the need for parental confirmation before starting medication in university while gradually assuming greater autonomy later on. We delve deeper into this transitional period in the latter part of this paper.

Due to concerns about taking mental health medication, although many patients ended up using it for a long time, they did not expect nor desire a long-term reliance on medications. Participant 4 and many others expressed that they did not anticipate medication being a lifelong necessity; instead, they hoped to gradually reduce and eventually stop taking it. Participant 8 expressed strong feelings on this:

> *I want to decrease overall medication usage. And I really hope that one day I will not rely on the medication.* [Participant 8]

This unique desire leads to more instability in managing mental health medication, unlike managing many chronic diseases where patients are comfortable maintaining a stable dose for extended periods [49]. Our participants consistently reevaluate their symptoms, actively seeking opportunities to adjust the dosage. This frequent reassessment significantly destabilizes the management of medication.

Theme 1.2: Dosage Changes and Medication Adjustments Are Common

In addition to the participant's willingness to reduce dosage, several other factors in real life may contribute to changes in dosage or even the type of medication. A common cause for adjusting dosage is the fluctuation in symptom severity. This occurrence is surprisingly frequent among our participants, given that mental health symptoms can be significantly and rapidly influenced by life changes, a common experience for university students. Participant 10 increased dosage after entering university to manage the heightened stress of schoolwork, while Participant 8 found a better balance and relied less on medication upon studying abroad for graduate school.

Insufficient effectiveness and significant side effects are substantial factors leading participants to alter their dosage or medication. We observed a noteworthy number of participants expressing the challenge of finding the right medication that works for them, particularly at the onset of their medication journey. Participant 4 described this process in detail:

> *You feel elaborate when you have to figure out what works for you. If you visit with a professional like a psychiatrist, and they prescribe something to the best of their knowledge. You have to stay on it for like 2 months to see the full effects, which means if it's something [that] doesn't work for you, you have to commit to it for 2 months, and then go back to the drawing board. Sometimes you have to wean off of it. It's like, hmm, start over.* [Participant 4]

Participant 1 also described people underestimating the risks when starting a mental health medication:

> *I know people expect sometimes it has an effect right away. I think it's not commonly known that within the first couple of weeks of taking medication are actually, if you're at risk for suicide, those are particularly volatile weeks for you in starting a medication, because it's a frequent occurrence that starting a medication can actually give people a sort of boost of motivation, almost that they need to follow through with any sort of suicidal ideation or self-harm.* [Participant 1]

This revealed a gap between theory and real-world experiences of taking medication, where there is more back and forth, which would complicate medication management strategies. To summarize, the previous descriptions explain why there are frequent changes in medication for university students with mental health issues, emphasizing the importance of recognizing and addressing these issues.

XSL•FO
RenderX

Theme 1.3: Medication Interlaces With Mental Health Symptoms

Aside from treating symptoms, medication for mental health establishes a complex and intimate relationship with mental health status. While many of our participants recognized the importance of taking medication on time, they also found this additional task to be stressful, which in turn deteriorated the symptoms.

> *I also constantly feel the pressure of stress which reminds me to take the medication that has been prescribed for that stress.* [Participant 1]

Participant 4 also found this process distressing "because you're altering your mental state in an attempt to fix your mental state." Mental health symptoms also impact participants' capacity to take medication that alleviates those symptoms. For instance, Participant 5 mentioned that during particularly depressive episodes, getting out of bed and accessing medication stored in the kitchen became challenging. As a result, the distinctive connection between mental health symptoms and managing medication designed to address those symptoms further complicates the process.

Theme 1.4: Transition Period for University Students

Besides the instability caused by medication and mental symptoms, the unique phase of being university students introduced further fluctuations in the medication management process. Many participants indicated significant changes in their medication routines upon entering school. This involved practical challenges such as differences in accessing medical services due to relocation and changes in insurance (typically transitioning from family to individual coverage). Moreover, this life stage shift represented a change in responsibility, prompting university students to manage their medication independently. As Participant 1 noted:

> *My mom would remind me a lot when I was at home. Almost daily or just even ask me if I had taken it, so I definitely lack that (at school).*

Participants also expressed that the decision to take medication was no longer solely their parents' responsibility, granting them more autonomy in decision-making. Predictably, this life transition resulted in unstable medication management strategies as university students adapted to a new lifestyle and worked on establishing new habits.

In this section, we discovered that the process of taking mental health medication was marked by instability and continual fluctuations, influenced by various factors related to the self, medication, and external events. This inherent instability prompted university students to develop flexible strategies for managing their medication, a topic we will delve into in detail in the next section.

Theme 2: Strategies for Managing Mental Health Medication Among University Students: Flexible and Creative Ways of Using Simple Technology

This section reflects the diversity of needs and corresponding methods university students use to better incorporate medication management into their dynamic lifestyles. Before exploring different strategies, it is crucial to delve into the motivation and consequences behind university students taking medication. This context is essential for understanding why they opt for specific strategies.

Theme 2.1: Motivation and Consequences: Symptoms

A large reason why our participants have developed or adopted strategies in the first place is due to the consequence of side effects or decline in mental health and/or performance that comes with missing dosages. Many of the consequences are experienced firsthand, while the motivation of others is driven by worry or anxiety. According to Participant 7,

> *If not taking my meds means that I feel horrible, or I have a depressive episode, or can't get the work done, or I don't show up in the way I want to enough times, I am motivated to take my meds on time. If I can't feel the difference, I'll spontaneously stop taking them, because I don't think they're making a difference. I think it's rarely me being motivated to take my meds. It's me being motivated to not feel like s**t.*

Further, due to the nature of many mental health medications, missing doses do not have an immediate effect on the user, which can cause the user to become more relaxed with their routine. Nevertheless, all of our participants recognized the importance of "being more consistent" and generally experienced heightened stress when missing a dose for an extended period, which could lead to a severe outcome. Participant 2 once took their own initiative to quit medication after forgetting to take medications for a few days and "was just kinda like, f**k it, I'm fine." However, this did not yield positive results, and Participant 2 had to resume their medication after a while when they felt the symptoms returning. This nicely characterizes why our participants find it so important to use different tools and strategies to aid them in their medication adherence.

Theme 2.2: Fixed Reminders Are Common But Not Enough

Many participants developed strategies involving alarms and reminders as parts of their daily routines. Of the 10 participants, 6 used the preinstalled iPhone clock app due to its easy access, and only a minority (n=2) sought out specialized apps such as Apple Health. Despite the use of various apps, all reminders were set at fixed times, typically aligned with their regular medication schedule. However, participants expressed that this approach is often insufficient for various reasons.

First, our participants sometimes do not immediately take their medication when the alarm goes off. For example,

> *If I'm just chilling at that point I'll just take my pill. If I am wrapping something up, like I'm playing a video game or watching a show and I have 5 more minutes left, I just gotta finish this thing.* [Participant 2]

Depending on their current activities, some individuals may choose to snooze the alarm if they are in the middle of a task. Moreover, if there is a longer delay before taking the medication, participants dynamically adjust the reminder time. For example,

XSL•FO
RenderX

if Participant 2 is not at home at 9 PM (the usual time of the recurring alarm), they turn off the 9 PM alarm and manually set a new alarm for 11:30 PM to be reminded later. They then repeat the snooze method until they have taken their medication. Some participants even use a backup reminder strategy at the beginning, setting two recurring alarms at different times in case they miss the first one.

Second, their medication routines might be closely tied to other schedules that do not adhere to fixed times. Participant 5's perspective highlights the significance of a system that acknowledges the variability in waking times of university students and emphasizes the need for reminders to adapt accordingly.

> *I wake up at not exactly the same time every day, and my medication routine is tied to my morning routine rather than a specific time. So setting an alarm that I need to take it at 9 am or 10 am every day isn't as useful for me.* [Participant 5]

In addition to the sleep schedule, Participant 3 also mentioned that their medications "don't do very well with skipping breakfast." This implies that having a meal is another crucial activity that participants need to consider when managing their medication.

The earlier discussions on schedules illustrate how routine activities related to medications occur at flexible times. Moreover, the situation could worsen if our participants have an irregular schedule on some days that disrupts their normal plan. For instance,

> *If I wake up sort of later in the day, and there's a lot that I need to do, I'm more likely to skip that normal morning routine that I stick with, which tends to just sort of push my medication to the back of my mind.* [Participant 1]

Academic tasks such as meetings are sometimes events that often get in the way. In those cases, taking medication will not be prioritized, and "sometimes I just end up forgetting because I don't have the allowance or the reminders anymore" [Participant 3]. In more complicated situations where Participant 5's medications were linked with meals, they missed a dose because the 7 AM meeting was earlier than their usual wake-up time, disabling them to take breakfast, which resulted in forgetting medications.

Due to the limitations of fixed alarms, some participants did not consider reminders and alarms to be the prioritized strategy, but rather viewed them as a double assurance. For Participant 9, they had a daily reminder but still usually took medications before the alarm.

To summarize, this perspective calls for a shift from a rigid, time-based reminder to a more adaptive and context-aware system. Participants expressed desire for technology that understands and integrates into their established daily routines. This ensures that medication reminders are not only timely but also align with their natural routines.

Theme 2.3: Physical Cues: A Simple But Effective Method

While acknowledging the limitations of digital reminders, several participants in our study expressed a preference for using physical cues to manage their medication. For instance, Participant 7 highlighted their reliance on a physical pill organizer right next to their bed. Similarly, Participant 10 strategically places their medications on their desk near their computer and backpacks. By keeping the pillbox visible and within their line of sight, they create visual cues that serve as reminders for medication intake. The preference for physical instruments is rooted in a sense of reliability and simplicity. This insight sheds light on the significance of environmental cues and the impact of visibility on adherence behavior.

Theme 2.4: Gaps Between Reminder and Actual Intake: The Paradox of Logging

While reminders are helpful in medication management, there is still a gap between setting up the reminders and actually taking the medications on time. Participant 3 shared a sentiment about how easy it is to become distracted and cast doubt upon oneself regarding whether they have actually taken their medication. When the alarm goes off, they have to walk over to the place where they keep their medication. If they are thinking about something else at that specific moment, they can forget whether they have taken it. Participant 2 also shared a strategy of turning the phone alarm off after, but not before taking medications to avoid uncertainty caused by distractions. In such cases, proper tracking and logs for medication become crucial. While a minority of our participants were logging their medication intakes, the majority of them (9 out of 10) were not doing so, even if they recognized the possibility and importance. This did not exactly align with our expectations of university students, who are generally more tech-savvy.

When discussing the requirements for a future system to assist with medication management, the most commonly mentioned keywords were minimal and simple. They do not want to "add an additional task" to their established routines, with self-reporting typically being considered one of those tasks. Participant 4 complained:

> *I'm like, not great at logging symptoms like, not great, basically just any longitudinal logging, like, I just don't love doing it.*

The need for logging was not just for medication but also for symptoms related to medication, such as moods and side effects. Similar to Theme 1.3, participants express a desire to monitor their mood and medication dosage to deduce a potential relationship, given the intricate connections between symptoms and medications. However, these efforts often face challenges, primarily due to the complexity and additional effort required for ongoing self-reporting. Here, we observed a paradox between the necessity of having this information and the reluctance to put in extra effort to acquire it.

XSL•FO
RenderX

Theme 2.5: Preserving Autonomy in Medication Management: Managed by Symptoms

In addition to finding ways to ensure they take medication on time, we also realized that our participants might occasionally decide to skip medication for various reasons. Among those reasons, an interesting factor was mental health symptoms. Because the primary motivation for taking medication is to treat the symptoms, some participants also manage medication based on symptoms. Participant 4 mentioned they never forgot to take medication for more than 3 days because, in that case, they would experience symptoms that reminded them to take their medications.

> *Is there a reason why today's harder than other days? Is there a reason why I'm doing more work to take care of myself mentally than other days? And then usually it'll be like, did I take my meds? No, I didn't, oh, shoot!*

In some cases, our participants would even take medication immediately once they felt the symptoms, usually for short-term effect meds. For instance, one of Participant 5's medications has a short-term effect. They expressed themselves as not having a good routine and taking medication immediately if they felt the symptoms, but would skip the next dose if it was too close.

All of these reflect participants' urgent needs to preserve autonomy in managing medication. That is part of the reason why, even if they acknowledge the limitations of oversimplified technology such as default alarms, they still choose to use it and adapt it in their own ways so that they can have full control over it.

Theme 3: Overcoming Obstacles in Seeking External Support

Besides coping with personal concerns, motivation, and the inconvenience of managing medication independently, our participants encountered numerous obstacles while seeking external support from various stakeholders. One of the most common barriers involved the difficulty in obtaining medication. In our study, all mental health medications were prescribed by doctors, necessitating patients to find a doctor, obtain a diagnosis, and then acquire a prescription. Many participants found this process overwhelming, primarily due to the multiple rounds of interactions with different professionals, particularly as most medications required issuance by specialists, such as psychiatrists. "It just ended up being a little bit too many phone calls and a little bit too much like I didn't have enough understanding on how," expressed one participant, who eventually gave up seeking medication due to the complex procedures on the first try. Participant 4 highlighted the frustration of starting the process repeatedly because of frequent relocations for school and changes in general practitioners. Additionally, Participants 3 and 8, who were international students, mentioned the extended difficulty in transferring prescriptions due to the US medical system's inability to validate previous diagnoses and prescriptions. Despite having local student insurance, Participant 8 continued to consult their doctors remotely in their home country and shipped the medication overseas. Some participants initially sought help from school medical centers but often faced long wait times.

Getting a doctor or prescription marked only the beginning, particularly for participants grappling with mental health issues. They placed immense importance on not just obtaining medical assistance but on finding a trustworthy and empathetic doctor. Participant 2, for instance, experimented with two therapists before settling on the current one, emphasizing the difficulty in locating a psychiatrist they genuinely trusted and perceived as caring. Participant 4 also remarked: "It's really hard to find a psychiatrist that you trust, that you think is caring." The recurrent theme of "caring" resonated strongly throughout the interviews. Participant 4, voicing frustration, explained,

> *It's not that they couldn't understand what I was saying. They were just like, that's normal, so it doesn't matter. (OR) I think that has a lot to do with being a woman and being young, and the experiences of people in those demographics within the medical system (said by doctor).*

Participants also voiced concerns about doctors overlooking the side effects of medications. One participant lamented, "Doctors are prescribing. Practitioners are not listening to me when I say that my side effects from this medication are harmful to me, and I do not want to take it anymore." Even Participant 5, who had a longstanding trust in their doctor, refrained from discussing concerns about the medication's long-term effects because, during the initial attempt, the doctor "kind of brushed it off a little bit and said I don't have dementia, and it's fine, and that maybe it's just depression." Similarly, Participant 10 shared discomfort with having the same doctor as their mother and found satisfaction in switching to a psychiatric professional who demonstrated genuine care and was willing to adjust the dosage to alleviate symptoms. Given the subjective and self-evaluated nature of mental health diagnoses and treatment, the trust placed in doctors became even more critical. Many participants sought to establish long-term relationships with trusted doctors and hesitated to switch, underscoring the pivotal role of trust in the management of mental health medications. Understanding this dynamic is essential in navigating the complexities of mental health care.

To summarize, these unique challenges related to various stakeholders hindered students' access to necessary medication, further disrupting their regular medication management routines. Considering these external factors tied to the broader system is crucial when designing technologies to aid in regulating mental health medication.

Discussion

Summary

In this paper, we conducted formative interviews with 10 university students and identified 3 themes and 9 subthemes reflecting their strategies and challenges in managing medications for mental health conditions (Table 3). In this section, we situate our findings within prior research and propose design implications for future technologies that support university students in managing their mental health medications.

XSL•FO
RenderX

Table . Mappings of themes and design implications.

Themes and subthemes	Design implications
1. A dynamic voyage: managing mental health medication is never a static process	
1.1. Struggling with self-acceptance: expectations of reducing or ceasing medication	Implications 1 and 5
1.2. Dosage changes and medication adjustments are common	Implications 1 and 3
1.3. Medication interlaces with mental health symptoms	Implication 1
1.4. Transition period for university students	Implication 1
2. Flexible and creative ways of using simple technology	
2.1. Motivation and consequences: symptoms	Implications 1 and 3
2.2. Fixed reminders are common but not enough	Implications 2 and 3
2.3. Physical cues: a simple but effective method	Implication 2
2.4. Gaps between reminder and actual intake: the paradox of logging	Implications 2 and 4
2.5. Preserving autonomy in medication management: managed by symptoms	Implications 1, 3, and 4
3. Overcoming obstacles in seeking external support	Implications 5 and 6

Identify State Changes in University Students Managing Mental Health Medication: Responsibility, Mental Status, and Self-Acceptance

During our study, we realized that mental health medication management for university students is a very dynamic process. Therefore, it is important to discuss the transitions between different states during the management process. In previous work, Caldeira et al [32] identified 4 states within the medication management framework: wellness, new task, erratic, and disruption. These states were categorized along two dimensions: medication regularity (routine vs nonroutine) and time scale (long-term vs short-term). The authors stated that changes in routines and medications can disrupt the most stable "wellness" state. This statement aligns with our findings. Based on the interviews, we further classified the specific changes in routines and medications that university students might encounter when managing mental health medications. For changes in medication, this includes but is not limited to starting medication, adjusting dosage, and changing the medication. For changes in routine, relocation, and shifts in personal status (examples specific to university students include entering a new school, graduating, and moving out from parents), these are special events that university students are facing. Although the overall categories of factors influencing state changes align with previous work, we do find and want to emphasize some differences. One important observation is that for psychiatric medications, the dynamic transitions are much more frequent, making the "new task state" more common than the "wellness state." Additionally, we found that the "wellness state" is not always the most desirable state for university students managing their mental health medication. According to our interviews, many participants, despite recognizing the chronic nature of their mental health symptoms, have a strong desire to lower their dosage or stop taking medication altogether. As researchers, it is crucial to acknowledge that the desire to reduce dosage can be an additional cause of state changes, beyond the introduction of new medication as mentioned in previous work.

In our study, we also want to highlight 3 additional factors that may contribute to changes in states, beyond routine and medication. The first factor is the change in responsibility. For our participants, who are university students, we found that many are gradually taking on more responsibility for managing their medications independently. This shift in mindset not only causes routine differences, such as the absence of parents' reminders (Participant 1), but also grants them more autonomy in decision-making. This change in responsibility leads to a series of changes in how they manage their medications. Therefore, including changes in responsibility within the framework is crucial for thoroughly discussing state transitions.

Mental status is another factor that can influence medication management states. Participant 5 mentioned that during particularly depressive episodes, even getting out of bed and accessing medication stored in the kitchen became challenging. For our participants, changes in mental status are common due to their diagnoses and the effects of medications. It is important to recognize that for the population we are targeting, fluctuations in mental status can significantly affect how they manage their medications and should be considered as a common factor to cause state changes.

Another factor is self-acceptance. While the stigma associated with mental health conditions may not be as prevalent among university students as in the past [50], there remains a hesitancy regarding medication, with many participants not viewing it as the best option. We observed variability in these concerns among our participants. Some mentioned feeling more at peace after several years of medication use. Given that intentional nonadherence remains a significant issue [51]—especially in our context—it is crucial to include the self-acceptance of medications in our spectrum of considerations to capture possible state changes.

XSL•FO
RenderX

Mutual Relationships Between Mental Status/Symptoms and Medication

In the earlier discussions, a key point emphasized by our participants is the complex process to "alter your mental state in an attempt to fix your mental state." This intricate relationship makes it challenging to ascertain whether changes in symptoms are attributed to medication or other factors. In many past studies discussing technology developed for medication management, the relationship between medications and mental symptoms was often underexplored, with the focus primarily on the management of medications alone [37]. For example, in a study by Stawarz et al [37] reviewing the functionality of 229 medication reminder apps, none of the 3 design requirements they summarized addressed symptom tracking. This is understandable since most chronic diseases are relatively stable, and traditionally, decisions regarding medication dosage and changes are based solely on clinical advice, not self-report. Consequently, the primary responsibility of a patient in managing their medication is simply to "adhere to that medication" [52]. In studies focusing on psychiatric medications, researchers are beginning to recognize this unique aspect of managing mental health medications related to symptoms, though they still encounter numerous barriers. For example, Papoutsaki et al [33] explored how individuals in an online health community track their symptoms and psychiatric medication dosages, particularly when attempting to withdraw from a medication. Members of the forums emphasized the use of daily diaries to track symptoms, and some even created charts to monitor the exponential decay of their previous dosage. However, this approach still heavily relies on self-regulation [53]. In our study, none of our participants used similar tools or social platforms to manage their medication and symptoms.

To further elaborate on the relationship between medication and mental symptoms, it is important to recognize their interdependence. Tracking changes in symptoms can help determine the effectiveness of the medication and the appropriate dosage. Conversely, fluctuating mental states can also affect a participant's ability to manage their medication. This intrinsic relationship complicates the medication management process. Some of our participants even mentioned managing their medication based on their symptoms rather than adhering to more routinized clinical advice. Although these actions are less standardized compared to those observed in online communities by Papoutsaki et al [33], the underlying rationale behind these behaviors is the same. The real-world practices we discovered suggest that a standardized medication regimen may not be adequate for participants who are actively engaged in a sensemaking process to understand the impact of their medications on their symptoms. In this context, technology should be designed to actively assist university students in understanding the relationship between their prescribed medications and their mental symptoms to prevent intentional nonadherence.

Design Implications

Based on our findings in interviews and discussions with previous works, we would like to present several detailed design implications. We hope that those implications can help researchers to design future technology that helps university students better manage their medications in 6 aspects.

Implication 1: Tracking the Relationship Between Mental Status/Symptoms and Medication

First, the system should have the capability to directly track and display potential relationships between medication and mental health symptoms. Other than an app only tracking medication dosage or mental status and mood, the system should proactively establish a relationship between those two factors. Second, it is also essential to record significant events that may influence mental status during the medication transition process, both those self-identified and automatically tracked. These events include but are not limited to starting medication, adjusting dosage, changing the medication, relocation, and shifts in personal status (examples specific to university students include entering a new school, graduating, and moving out from parents) according to insights from theme 1. With these functions in place, the system should be able to aid students in evaluating the impact of medication, thereby fostering trust and belief in the treatment.

Implication 2: Dynamically Adjust Reminders Based on Location and Schedule

We observed that participants typically linked their medication intake with routines and specific locations. In this context, they often encounter dissatisfaction with fixed reminders, particularly when they have flexible routines. This barrier prompts them to establish their own methods, such as creating backup reminders (theme 2). This finding aligns with previous research; for example, Stawarz et al [37] mentioned that medication reminder apps need to support the creation of backup reminders and routines. We aim to further elaborate on these broad design implications drawn from previous works. To design a future system that acknowledges and accommodates participants' existing approaches to medication management, we proposed the following actionable design suggestions based on our findings.

Broad Delivery Time Window Based on Sleep, Eating Schedule, Last Medication Intake Time, and Availability

The reminder's delivery time window should be set widely, especially in situations where medication does not necessitate a fixed intake time (which is the case in 100% of our interviews). Sleep and eating routines are major factors influencing medication timing; reminders should dynamically adjust to these daily patterns. The timing should also consider the previous intake, since many mental health medications have defined effectiveness periods and dosage limits. If one dose is taken early, subsequent reminders should be delayed accordingly. Finally, reminders must align with moments of actual availability. Sending notifications during classes, meetings, or other commitments often leads to missed doses. For university students, academic schedules and obligations should therefore inform the optimal reminder timing.

Deliver at the Right Place: Storage and Activity

Participants frequently used location cues as medication reminders (theme 2.3). They reported that generic phone pop-ups were often ignored, highlighting the importance of delivering

XSL•FO
RenderX

reminders in contextually relevant places. Effective locations typically include where the medication is stored and where related activities occur—sometimes the same place, such as keeping medication by the bed for nighttime doses. Delivering reminders near these locations can improve visibility and accessibility; for instance, an audio cue near a pillbox can prompt action as participants pass by.

Implication 3: Recognize the Potential for Immediate Medication Intake in Response to Symptoms

In our interviews, we also noted that, at times, medication management was loosely based on symptoms (theme 2.1). In severe cases, missing a dose and not promptly taking it could lead to a potential disaster (eg, Participant 7 unable to drive and needing to take the subway home). While the system should provide participants with considerable autonomy to decide in such situations, it is important to acknowledge that participants might take action when faced with these situations, and follow-up management should be adjusted accordingly. Although more discussion is needed, especially considering the limited accuracy of sensing those symptoms and the complicated decision-making process to ensure safety, immediate intervention might be necessary if medication nonadherence occurs in a dangerous situation (eg, while driving).

Implication 4: Striking a Balance Between Autonomy and Minimal Effort

We recognized a paradox in the medication management process: students desire control over their own routines but are hesitant to invest substantial effort in additional tasks, such as self-reporting (themes 2.4 and 2.5). While many of these challenges may arise from limitations in the technology they use [54], which could be better designed following the previous design implications, additional design strategies could enhance the overall experience.

Automatic tracking and background reminders can significantly alleviate the efforts students invest in the medication management process. However, false detections might lead to increased efforts for correction [55]. In such instances, these tracking and reminders should be easily modifiable with simple interactions akin to the one-click snooze technique for a phone alarm. This ensures that participants can maintain minimal effort while achieving more accurate and flexible management results.

Implication 5: Safeguarding Privacy in a Creative Manner

Many individuals express concerns about third-party apps accessing sensitive information or the potential for unauthorized access. Here are two strategies to alleviate these concerns. (1) Create pseudonyms for medication, especially for reminders displayed on the interface that other people might be able to see. This creative approach was suggested by one of the participants. By integrating a “personal encryption rule” into the process, the app can significantly desensitize the information related to medication management. (2) Deidentification, especially regarding academic information. In discussions about collaboration between academic institutions and medical services, despite understanding the regulations, many individuals remain concerned that disclosure of their symptoms, especially medications, might negatively impact their academic well-being (theme 3). One way to address this concern is to completely deidentify the data (name, email, phone number, and academic program). While complete anonymity may be challenging [56,57], implementing some form of redirection in the information transition process can help students build trust in the technology and increase adoption.

Implication 6: Enhance Communication With Other Stakeholders to Build Trust

Many participants conveyed that their doctors did not show sufficient care for their symptoms and side effects (theme 3). In a paper by Jo et al [58], the authors discuss how to support clinicians in planning the discontinuation of psychiatric drugs, including the types of information needed. While they considered factors such as trust issues, patient medical history, and insurance issues, they did not address how patient feedback might influence decisions about the regimen. In our study, we found that the relationship between symptoms and medications is crucial for managing mental health medications among university students. This relationship is typically best understood by participants through their daily experiences. Considering the lack of trust and communication expressed by some of our participants, we propose that participants should have the ability to use data previously collected on the mutual relationship between symptoms and medications. This would empower them to effectively discuss and influence medical decisions with their doctors. However, this should be under full control of the participants, taking into consideration privacy issues.

Limitations

This study is subject to certain limitations. First, the sample size is relatively small, restricting our ability to generalize the findings to a wider population. This limitation stems from the reluctance to disclose sensitive information. In our research, we specifically focus on a particular group of university students, requiring them to provide a “.edu” email address. Participants may be concerned about linking their academic details to mental health challenges, thereby impeding their willingness to take part. To address these concerns, aside from implementing deidentification measures, we only requested participants to disclose information related to diagnoses and medications if they were comfortable doing so. However, this introduces the potential for selection bias, as those who willingly participated in the interviews might already be predisposed to sharing such information. Despite acknowledging these limitations, our analysis of the interview data revealed numerous common patterns and reached data saturation, affirming the validity of our findings. Additionally, we provide the survey results encompassing demographic information, medication schedule, medication adherence, and self-efficacy. This information can assist future researchers in better contextualizing our results. Given the diversity in daily routine management, we are cautious about generalizing our findings to other population groups. Future research could target a broader population using less sensitive measures, such as surveys, to further validate our findings.

In our study, we refrained from collecting specific medication names for each participant due to privacy considerations. While

XSL•FO
RenderX

many participants willingly shared information about symptoms, medication working mechanisms (eg, immediate effects or delays), or even the exact names of medications, this practice impacted our capacity to systematically analyze varied strategies corresponding to different types of medications. Acting as an exploratory endeavor, we aspire for future studies to gather more comprehensive information, such as electronic health records, to further delve into these issues.

Conclusions

In the paper, we conducted semistructured interviews with university students currently undergoing medication for mental health issues. The interviews aimed to uncover the challenges they encountered in their university lives, along with the coping strategies and technologies used to navigate these challenges. We found that deciding to take medications is often a challenging choice for many students. The expectations for dosage reduction, the inherent uncertainty of medication effects, the intricate relationship between medications and symptoms, and the frequent life changes associated with university all contribute to a highly dynamic journey in medication management. As a result, students have developed various strategies to cope with these challenges, often relying on creative ways to use fixed alarms on their phones. However, they encounter numerous difficulties due to factors such as the flexibility of daily routines. Despite university students' strong desire to maintain autonomy in the medication management process, they still struggle to devote additional effort to self-reporting, creating additional challenges in building trust with their doctors. Based on these findings, we proposed 6 design implications for how advanced technology could better assist university students in managing their mental health medications.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Conflicts of Interest

None declared.

References

1. Healthy Minds Study among colleges and universities, year (2022-2023) [Data set].: Healthy Minds Network; 2023. URL: https://healthymindsnetwork.org/research/data-for-researchers [accessed 2026-02-12]
2. Lipson SK, Zhou S, Abelson S, et al. Trends in college student mental health and help-seeking by race/ethnicity: findings from the National Healthy Minds Study, 2013-2021. J Affect Disord 2022 Jun 1;306:138-147. [doi: 10.1016/j.jad.2022.03.038] [Medline: 35307411]
3. De Nieva JO, Joaquin JA, Tan CB, Marc Te RK, Ong E. Investigating students' use of a mental health chatbot to alleviate academic stress. In: CHIuXiD '20: 6th International ACM In-Cooperation HCI and UX Conference: Association for Computing Machinery; 2021. [doi: 10.1145/3431656.3431657]
4. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 6;4(2):e19. [doi: 10.2196/mental.7785] [Medline: 28588005]
5. Morris RR, Kouddous K, Kshirsagar R, Schueller SM. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. J Med Internet Res 2018 Jun 26;20(6):e10148. [doi: 10.2196/10148] [Medline: 29945856]
6. Bauer AM, Iles-Shih M, Ghomi RH, et al. Acceptability of mHealth augmentation of collaborative care: a mixed methods pilot study. Gen Hosp Psychiatry 2018;51:22-29. [doi: 10.1016/j.genhosppsych.2017.11.010] [Medline: 29272712]
7. Richards P, Simpson S. Beyond the therapeutic hour: an exploratory pilot study of using technology to enhance alliance and engagement within face-to-face psychotherapy. Br J Guid Counc 2015 Jan;43(1):57-93. [doi: 10.1080/03069885.2014.936824]
8. Mishra V. From sensing to intervention for mental and behavioral health. In: UbiComp/ISWC '19 Adjunct: Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers: Association for Computing Machinery; 2019:388-392. [doi: 10.1145/3341162.3349304]
9. Mishra V, Sen S, Chen G, et al. Evaluating the reproducibility of physiological stress detection models. Proc ACM Interact Mob Wearable Ubiquitous Technol 2020 Dec 17;4(4):1-29. [doi: 10.1145/3432220]
10. Xu X, Chikersal P, Doryab A, et al. Leveraging routine behavior and contextually-filtered features for depression detection among college students. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Sep 9;3(3):1-33. [doi: 10.1145/3351274]
11. Conn VS, Hafdahl AR, Cooper PS, Ruppar TM, Mehr DR, Russell CL. Interventions to improve medication adherence among older adults: meta-analysis of adherence outcomes among randomized controlled trials. Gerontologist 2009 Aug;49(4):447-462. [doi: 10.1093/geront/gnp037] [Medline: 19460887]

XSL•FO
RenderX

12. Yap AF, Thirumoorthy T, Kwan YH. Systematic review of the barriers affecting medication adherence in older adults. Geriatr Gerontol Int 2016 Oct;16(10):1093-1101. [doi: 10.1111/ggi.12616] [Medline: 26482548]
13. Schwartz JK, Unni E. Inclusion of people with disabilities in research to improve medication adherence: a systematic review. Patient Prefer Adherence 2021;15:1671-1677. [doi: 10.2147/PPA.S314135] [Medline: 34345167]
14. Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S, Ustün TB. Age of onset of mental disorders: a review of recent literature. Curr Opin Psychiatry 2007 Jul;20(4):359-364. [doi: 10.1097/YCO.0b013e32816ebc8c] [Medline: 17551351]
15. Kitzrow MA. The mental health needs of today's college students: challenges and recommendations. NASPA Journal 2003 Dec;41(1):167-181. [doi: 10.2202/1949-6605.1310]
16. Pedrelli P, Nyer M, Yeung A, Zulauf C, Wilens T. College students: mental health problems and treatment considerations. Acad Psychiatry 2015 Oct;39(5):503-511. [doi: 10.1007/s40596-014-0205-9] [Medline: 25142250]
17. Zivin K, Eisenberg D, Gollust SE, Golberstein E. Persistence of mental health problems and needs in a college student population. J Affect Disord 2009 Oct;117(3):180-185. [doi: 10.1016/j.jad.2009.01.001] [Medline: 19178949]
18. Doran C. Prescribing Mental Health Medication: The Practitioner's Guide, 3rd edition: Routledge; 2021.
19. Shahin W, Kennedy GA, Stupans I. The association between social support and medication adherence in patients with hypertension: a systematic review. Pharm Pract (Granada) 2021;19(2):2300. [doi: 10.18549/PharmPract.2021.2.2300] [Medline: 34221197]
20. Wagoner ST, Schaefer MR, Rawlinson AR, Shapiro SK, Kavookjian J, Gray WN. Barriers to treatment adherence among college students with attention-deficit/hyperactivity disorder. J Dev Behav Pediatr 2020 Jan;41(1):9-15. [doi: 10.1097/DBP.0000000000000723] [Medline: 31449195]
21. Chou T, Golik AM, DeSerisy M, Comer JS. Disorders of negative affect. In: Friedman HS, editor. Encyclopedia of Mental Health, 2nd edition: Elsevier; 2016:69-74. [doi: 10.1016/B978-0-12-397045-9.00005-7]
22. Bushnell GA, Compton SN, Dusetzina SB, et al. Treating pediatric anxiety: initial use of SSRIs and other antianxiety prescription medications. J Clin Psychiatry 2018;79(1):16m11415. [doi: 10.4088/JCP.16m11415] [Medline: 29099547]
23. Merikangas KR, He JP, Rapoport J, Vitiello B, Olfson M. Medication use in US youth with mental disorders. JAMA Pediatr 2013 Feb;167(2):141-148. [doi: 10.1001/jamapediatrics.2013.431] [Medline: 23403911]
24. Leppien E, Bystrak T, Doughty B. Chapter 2: Antidepressant medications. In: Ray SD, editor. A Worldwide Yearly Survey of New Data in Adverse Drug Reactions: Elsevier; 2020, Vol. 42. [doi: 10.1016/bs.seda.2020.08.007]
25. Walkup JT, Albano AM, Piacentini J, et al. Cognitive behavioral therapy, sertraline, or a combination in childhood anxiety. N Engl J Med 2008 Dec 25;359(26):2753-2766. [doi: 10.1056/NEJMoa0804633] [Medline: 18974308]
26. Babel A, Taneja R, Mondello Malvestiti F, Monaco A, Donde S. Artificial intelligence solutions to increase medication adherence in patients with non-communicable diseases.. Front Digit Health 2021;3:669869. [doi: 10.3389/fdgth.2021.669869] [Medline: 34713142]
27. Arshed M, Mahmud AB, Minhat HS, Ying LP, Umer MF. Effectiveness of mHealth interventions in medication adherence among patients with cardiovascular diseases: a systematic review. Diseases 2023 Mar 1;11(1):41. [doi: 10.3390/diseases11010041] [Medline: 36975590]
28. Marrero RJ, Fumero A, de Miguel A, Peñate W. Psychological factors involved in psychopharmacological medication adherence in mental health patients: a systematic review. Patient Educ Couns 2020 Oct;103(10):2116-2131. [doi: 10.1016/j.pec.2020.04.030] [Medline: 32402489]
29. Jimmy B, Jose J. Patient medication adherence: measures in daily practice. Oman Med J 2011 May;26(3):155-159. [doi: 10.5001/omj.2011.38] [Medline: 22043406]
30. Mu Y, Zhao L, Shen L. Medication adherence and pharmaceutical design strategies for pediatric patients: an overview. Drug Discov Today 2023 Nov;28(11):103766. [doi: 10.1016/j.drudis.2023.103766] [Medline: 37708932]
31. Rootes-Murdy K, Glazer KL, Van Wert MJ, Mondimore FM, Zandi PP. Mobile technology for medication adherence in people with mood disorders: a systematic review. J Affect Disord 2018 Feb;227(613):613-617. [doi: 10.1016/j.jad.2017.11.022] [Medline: 29172054]
32. Caldeira C, Bhowmick P, Komarlingam P, Siek KA. A state-based medication routine framework. In: CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems: Association for Computing Machinery; 2022. [doi: 10.1145/3491102.3517519]
33. Papoutsaki A, So S, Kenderova G, Shapiro B, Epstein DA. Understanding delivery of collectively built protocols in an online health community for discontinuation of psychiatric drugs. Proc ACM Hum-Comput Interact 2021 Oct 13;5(CSCW2):1-29. [doi: 10.1145/3479564]
34. Griffin AC, Xing Z, Mikles SP, et al. Information needs and perceptions of chatbots for hypertension medication self-management: a mixed methods study. JAMIA Open 2021 Apr;4(2):ooab021. [doi: 10.1093/jamiaopen/ooab021] [Medline: 33898936]
35. Pal P, Sambhakar S, Dave V, et al. A review on emerging smart technological innovations in healthcare sector for increasing patient's medication adherence. Global Health Journal 2021 Dec;5(4):183-189. [doi: 10.1016/j.glohj.2021.11.006]
36. Park LG, Howie-Esquivel J, Dracup K. Electronic measurement of medication adherence. West J Nurs Res 2015 Jan;37(1):28-49. [doi: 10.1177/0193945914524492] [Medline: 24577868]

XSL•FO
RenderX

37. Stawarz K, Cox AL, Blandford A. Don't forget your pill! Designing effective medication reminder apps that support users' daily routines. In: CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Association for Computing Machinery; 2014:2269-2278. [doi: 10.1145/2556288.2557079]
38. Rice DR, Kaplan TB, Hotan GC, et al. Electronic pill bottles to monitor and promote medication adherence for people with multiple sclerosis: a randomized, virtual clinical trial. J Neurol Sci 2021 Sep 15;428:117612. [doi: 10.1016/j.jns.2021.117612] [Medline: 34392138]
39. Appiah B, Kretchy IA, Yoshikawa A, Asamoah-Akuoko L, France CR. Perceptions of a mobile phone-based approach to promote medication adherence: a cross-sectional application of the technology acceptance model. Explor Res Clin Soc Pharm 2021 Mar;1:100005. [doi: 10.1016/j.rcsop.2021.100005] [Medline: 35479503]
40. Mason M, Cho Y, Rayo J, Gong Y, Harris M, Jiang Y. Technologies for medication adherence monitoring and technology assessment criteria: narrative review. JMIR Mhealth Uhealth 2022 Mar 10;10(3):e35157. [doi: 10.2196/35157] [Medline: 35266873]
41. Corden ME, Koucky EM, Brenner C, et al. MedLink: a mobile intervention to improve medication adherence and processes of care for treatment of depression in general medicine. Digit Health 2016;2:2055207616663069. [doi: 10.1177/2055207616663069] [Medline: 29942564]
42. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence-based chatbots for promoting health behavioral changes: systematic review. J Med Internet Res 2023 Feb 24;25:e40789. [doi: 10.2196/40789] [Medline: 36826990]
43. Risser J, Jacobson TA, Kripalani S. Development and psychometric evaluation of the Self-Efficacy for Appropriate Medication Use Scale (SEAMS) in low-literacy patients with chronic disease. J Nurs Meas 2007;15(3):203-219. [doi: 10.1891/106137407783095757] [Medline: 18232619]
44. Corbin JM, Strauss A. Grounded theory research: procedures, canons, and evaluative criteria. Qual Sociol 1990;13(1):3-21. [doi: 10.1007/BF00988593]
45. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
46. Glaser BG, Strauss AL, Strutzel E. The discovery of grounded theory; strategies for qualitative research. Nurs Res 1968 Jul;17(4):364. [doi: 10.1097/00006199-196807000-00014]
47. Ando H, Cousins R, Young CA. Achieving saturation in thematic analysis: development and refinement of a codebook. Comprehensive Psychology 2014 Jan;3:03. [doi: 10.2466/03.CP.3.4]
48. Guest G, Bunce A, Johnson L. How many interviews are enough?: An experiment with data saturation and variability. Field methods 2006 Feb;18(1):59-82. [doi: 10.1177/1525822X05279903]
49. Coleman CI, Limone B, Sobieraj DM, et al. Dosing frequency and medication adherence in chronic disease. J Manag Care Pharm 2012 Sep;18(7):527-539. [doi: 10.18553/jmcp.2012.18.7.527] [Medline: 22971206]
50. Gabunia L, Verulava T. Exploring stigma associated with mental health services among college students. Health Policy, Economics and Sociology 2024 Jan;8(1). [doi: 10.52340/healthecosoc.2024.08.01.02]
51. Lehane E, McCarthy G. An examination of the intentional and unintentional aspects of medication non-adherence in patients diagnosed with hypertension. J Clin Nurs 2007 Apr;16(4):698-706. [doi: 10.1111/j.1365-2702.2005.01538.x] [Medline: 17402951]
52. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 4;353(5):487-497. [doi: 10.1056/NEJMra050100] [Medline: 16079372]
53. Cadel L, Cimino SR, Rolf von den Baumen T, James KA, McCarthy L, Guilcher SJT. Medication management frameworks in the context of self-management: a scoping review. Patient Prefer Adherence 2021;15:1311-1329. [doi: 10.2147/PPA.S308223] [Medline: 34163148]
54. Kelley C, Lee B, Wilcox L. Self-tracking for mental wellness: understanding expert perspectives and student experiences. In: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems: Association for Computing Machinery; 2017:629-641. [doi: 10.1145/3025453.3025750]
55. Lam WY, Fresco P. Medication adherence measures: an overview.. Biomed Res Int 2015;2015(1):217047. [doi: 10.1155/2015/217047] [Medline: 26539470]
56. El Emam K, Moher E. Privacy and anonymity challenges when collecting data for public health purposes. J Law Med Ethics 2013 Mar;41 Suppl 1(Suppl 1):37-41. [doi: 10.1111/jlme.12036] [Medline: 23590738]
57. Walker N. All or nothing: the false promise of anonymity. bioRxiv. Preprint posted online on Nov 3, 2016. [doi: 10.1101/084921]
58. Jo E, Ryu M, Kenderova G, et al. Designing flexible longitudinal regimens: supporting clinician planning for discontinuation of psychiatric drugs. In: CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems: Association for Computing Machinery; 2022. [doi: 10.1145/3491102.3502206]

Abbreviations

AI: artificial intelligence
SEAMS: Self-Efficacy for Appropriate Medication Use Scale

XSL•FO
RenderX

Edited by M Waite; submitted 21.Oct.2024; peer-reviewed by H Foord-Warren, V Zhu; revised version received 10.Dec.2025; accepted 27.Jan.2026; published 19.Mar.2026.

Please cite as:
Li J, Steinberg J, Mynatt E, Mishra V
Challenges and Strategies of Managing Mental Health Medication in Real-World Practices: Interview Study Among University Students
JMIR Hum Factors 2026;13:e67579
URL: https://humanfactors.jmir.org/2026/1/e67579
doi:10.2196/67579

© Jiachen Li, Justin Steinberg, Elizabeth Mynatt, Varun Mishra. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 19.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Health Care Professionals' and Patients' Perceptions and Experiences of Who Uses Video Consultations, and Why: Qualitative Study

Irene Muli[1], PhD; Åsa Cajander[2], PhD; Nadia Davoody[3], PhD; Lovisa Jäderlund Hagstedt[3,4], MD; Helena Hvitfeldt[1], PhD; Maria Hägglund[1,5], PhD; Marina Taloyan[6,7,8], PhD

[1]Participatory eHealth and Health Data Research Group, Department of Women's and Children's Health, Uppsala University, Dag Hammarskjölds väg 14B, Uppsala, Sweden

[2]Department of Information Technology, Uppsala University, Uppsala, Sweden

[3]Department of Learning Informatics, Health Informatics Centre, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

[4]Norrtälje Hospital, Vårdbolaget Tiohundra, Stockholm, Sweden

[5]Uppsala University Hospital, Uppsala, Sweden

[6]Academic Primary Healthcare Centre, Region Stockholm, Stockholm, Sweden

[7]Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden

[8]Tbilisi State Medical University, Tbilisi, Georgia

Corresponding Author:
Irene Muli, PhD
Participatory eHealth and Health Data Research Group, Department of Women's and Children's Health, Uppsala University, Dag Hammarskjölds väg 14B, Uppsala, Sweden

Abstract

Background: Video consultations (VCs) are effective and beneficial, yet their use is being discontinued, and there is a preference for face-to-face consultations.

Objective: This study investigates how patients and health care professionals (HCPs) perceive patients' introduction to VCs, who use them, and what drives their use in Swedish primary care.

Methods: Six focus group interviews with 27 HCPs and 13 individual interviews with patients in primary care were conducted between August 2022 and May 2023. The interviews examined VC implementation and were analyzed using rapid assessment procedures.

Results: A total of five themes were identified: (1) challenging start with unprepared users and immature technology; (2) users and nonusers are perceived to have different characteristics, needs, and circumstances; (3) patient-related drivers: based on patients' preferences and opportunities; (4) HCP-related drivers: clinical suitability, assessment of patient needs, and preferences; and (5) societal and organizational drivers: the pandemic, demographics, and infrastructure. Patients and HCPs described the introduction of VCs as rushed and confusing, with limited guidance and support (theme 1). HCPs struggled to assist patients due to a lack of training and limited access to the patient-facing interface (theme 1). VC users were typically perceived as younger, digitally literate, and motivated by convenience or urgency, while older adults and those with language or cognitive barriers were often assumed to be nonusers (theme 2). VC use was shaped by patient preferences, accessibility, and clinical urgency (theme 3), as well as by HCPs' professional judgment and convenience (theme 4). Assumptions held by HCPs about patients' digital skills and preferences influenced whether VCs were offered, while patients' own assumptions about complexity or suitability affected whether they accepted them. Broader factors, such as digital infrastructure, platform usability, reimbursement policies, and the COVID-19 pandemic, also significantly influenced use (theme 5).

Conclusions: The rushed implementation potentially deterred some patients and HCPs from use. Misguided preconceptions and biases negatively influenced VC use and risked reinforcing existing disparities and contributing to digital exclusion. In addition, HCPs' and patients' preferences, which were related to their needs, waiting times, and different circumstances, and potentially misguided judgments of appropriateness, influenced VC use. Lastly, infrastructure, reimbursement, sociodemographics, and organizational type also drive VC use. To support more sustainable and equitable use of VC in primary care, developers should optimize VC applications' usability, implementers should deploy multiple strategies, health care providers should consider the potential of VC in care delivery, and policymakers should increase digital readiness. Further research should evaluate the effectiveness of different strategies for introducing patients to VCs, explore younger patients' and nonusers' perspectives, characteristics of HCP users, and differences between professional roles, as well as between consultation types.

XSL•FO
RenderX

(JMIR Hum Factors 2026;13:e68658) doi:10.2196/68658

KEYWORDS

video consultations; primary care; introduction; perceptions; drivers; Sweden

Introduction

Background

The rapid increase in video consultations (VCs) is an example of the accelerating digitalization of health care [1,2]. However, despite this progress, some studies indicate that VC use, which peaked during the COVID-19 pandemic, has since declined [3-5].

Teleconsultations, such as VCs, are generally as effective as face-to-face consultations regarding clinical outcomes in primary care and mental health [6] and could, in some cases, be a superior alternative to telephone consultations (TCs) [7]. Additionally, both patients and health care professionals (HCPs) have reported high satisfaction with VCs [6,8-10]. Despite these positive findings, some patients and HCPs still prefer face-to-face consultations [9-11].

The way that the use of VCs is implemented might explain the discontinuation of VC use and preference for face-to-face consultations [12]. However, research on implementation is limited to only a few regions or countries and secondary care settings [13-15], with studies mainly investigating HCPs' experiences and organizational challenges [13]. While existing studies often focus on postimplementation satisfaction, they tend to overlook the implementation process, particularly from the patients' perspectives [16,17]. When patient introduction to VCs is described, it is often vague and lacks sufficient detail [18,19]. No studies have been found that conduct an in-depth investigation of patients' introduction to VCs.

In the research project ePrIm (implementation of eHealth in primary care), the overall aim was to explore how the implementation of eHealth in Swedish primary care affects the work environment of HCPs and patient satisfaction with care. The implementation of VCs was 1 of 3 case studies conducted in the project, the other 2 focusing on patients' online record access [20] and digital self-management of asthma. In this study, we focus on results from the VC case study, specifically on patients' introduction to VCs and perceptions of who may or may not want to use VCs.

Perceptions of who the users and nonusers of VCs are might influence VC use [21]. Although there is considerable research on the perceptions of VCs, little is known about the perceptions of who the users and nonusers of VCs are. In the context of this study, "perceptions" comprise the attitudes, beliefs, and subjective judgments held by HCPs and patients about VCs. This includes how they view the characteristics, needs, and potential use of VCs for different patient groups, as well as their opinions regarding the suitability of VCs for various medical scenarios. Further, few studies have explored the drivers of VC use and nonuse in-depth [21-23]; drivers in this study refer to factors influencing use either positively or negatively.

This study addresses notable gaps in the current research: patients' introduction to VCs, perceptions of users and nonusers, and the drivers of VC use, in a Swedish context. By collecting data from patients and HCPs, this study offers a comprehensive perspective on the factors influencing VC adoption and use. This dual approach is valuable because it allows for a holistic understanding of the potentially conflicting motivations, experiences, and barriers perceived by the 2 groups. By examining both sides, this study can identify discrepancies and areas of potential improvement, providing a richer and more nuanced understanding of VC use. Given the reported benefits of VCs for patients, HCPs, and society, it is vital to expand our understanding of why there is a risk of discontinuing their use in various contexts and why there remains a preference for face-to-face consultations.

Aim

This study aimed to investigate the dual perspectives of patients and HCPs on patients' introduction to VCs, perceptions of VCs users, and drivers of use in Swedish primary care. More specifically, this study investigated: (1) HCPs' and patients' experiences of patients' introduction to VC, (2) HCPs' and patients' perceptions of who uses or does not use VC, and (3) HCPs' and patients' experiences and perceptions of drivers of use or nonuse of VC.

Methods

Study Design

This study had a qualitative approach, using focus groups and individual interviews to capture the experiences and perceptions of HCPs and patients on the introduction and use of VCs.

Study Setting

Digitalization of health care services is widespread in Sweden. According to a national assessment, digital consultations (ie, consultations using applications that replace in-person consultations, eg, VCs) with physicians have increased from 10% in 2019 to 24% in 2022 [24,25]. Nearly all Swedes have internet access at home, and over 90% use it daily [24].

This study was conducted in Region Stockholm, 1 of 21 regions responsible for providing tax-funded health care in Sweden [26]. Tax-funded health care can be either publicly or privately managed, and a third of the care in Region Stockholm is privately managed [27]. This study involved both privately and publicly managed health care centers. The care fees are the same within the region but will vary depending on the patient's citizenship [28]. VCs cost the same as face-to-face consultations for the patient, although the care provider's reimbursement is lower for VCs [29].

The region's over 200 primary care centers (PCCs) and family doctor offices are the gateway to health care from which referrals

XSL•FO
RenderX

to specialized care are made when necessary [28] PCCs typically comprise a mix of HCPs.

Initially, direct-to-patient VCs were provided through online platforms and applications developed and managed by private health care providers [30]. Progressively, especially during the COVID-19 pandemic, platforms and applications have emerged in the public sector. *Alltid öppet* [always open] is such a platform, which was developed by the public health care provider Region Stockholm in 2018. *Alltid öppet* supports both VCs and consultations through chats between patients and HCPs and currently (2024) has 2 million users. PCCs can choose which digital services they offer through the platform, and, in some cases, this includes online appointment booking by the patients themselves. Parents can use the application on behalf of children younger than 13 years of age. Use requires authorization with electronic identification (eID). *Alltid öppet* was the VC service used by both HCPs and patients in this study.

Recruitment of Study Participants

All participants were recruited from 10 PCCs in Region Stockholm; 5 rural PCCs in the most northern part of Stockholm (Norrtälje), and 5 more centrally located PCCs with a variation in sociodemographic contexts (Jakobsberg, Hässelby, Liljeholmen, Gustavsberg, and Huddinge PCC).

All HCPs from the participating health care centers with experience in conducting VCs through *Alltid öppet were* invited to participate. All eligible HCPs were sent invitations via email with information about this study and were followed up with help from the PCC manager, who helped schedule the focus groups at times convenient for the volunteering HCPs. The participating HCPs also received information about this study before the interview. The focus groups consisted of HCPs of different professions: general practitioners (GPs), nurses, counselors, psychologists, and physiotherapists. The HCPs were also diverse with regard to their experience of using VCs with patients and as patients themselves. A total of 6 focus group interviews were conducted with 3 to 6 participants in each; in total, 27 HCPs participated. All but 1 focus group consisted of HCPs from the same PCC; the remaining group consisted of participants from 2 smaller PCCs.

Patients were recruited to the current study through a previously conducted survey about VC use through *Alltid öppet* [11]. Survey participants were eligible to participate if they were 16 years or older and had a VC through the *Alltid öppet* application from March to May 2022. Invitations to participate in the survey study were sent via text message. After completing the survey, respondents were invited to participate in the interview study. Those interested in participating received a link to a short questionnaire, separate from the survey, which also provided further information about the interview study. The questionnaire collected information regarding: contact details, demographics, and frequency of VC usage. A diverse group of respondents (Table 1) was then contacted via telephone and invited to participate. All participants received full study information; consent to participate and record the interviews was obtained before the interview. In total, 13 patients were interviewed.

XSL•FO
RenderX

Table . Characteristics of the interviewed patients, Region Stockholm, 2022.

Patient characteristics	Values
Sex	
Female	7
Male	5
Other or do not want to answer	1
Age (years)	
25 - 34	2
35-44	3
45-54	1
55-64	4
65-74	2
75-84	1
Highest education level	
Primary school	1
High school	3
Adult education [29]	1
University or university college <3 years	2
University or university college >2 years	6
Total number of VCs[a] via Alltid öppet[b]	
1-5	7
6-10	2
More than 10	1

[a]VC: video consultations.

[b]Alltid öppet: ”always open”; an application for video consultations.

Data Collection

Data were collected using focus group interviews with HCPs and individual semistructured interviews with patients. The focus group interviews were conducted between December 2022 and May 2023, and the patient interviews between August and December 2022. The dates were determined by the project plan and the participants’ availability. The choice to use focus group interviews with HCPs and individual semistructured interviews with patients was motivated by the desire to capture the collective dynamics and shared experiences among HCPs in a group setting, while allowing for a more in-depth exploration of personal experiences and perceptions in the one-on-one patient interviews. Both the focus groups and the individual interviews were audio-recorded and transcribed verbatim.

All focus groups, except one, which was performed in person, were conducted through videoconferencing at their workplace. A semistructured interview guide was used in the focus group interviews. The guide was based on the overall project aim to investigate the implementation of VCs through *Alltid öppet*, its impact on the work environment and work processes, as well as the HCPs’ perceptions of patients’ introduction to and use of VCs. In this study, our focus is on the latter. Due to our limited access to HCPs’ time, we chose not to pilot the focus group interview guide but instead critically assessed it after each focus group to ensure all important aspects were covered. The interviews were 45 - 60 minutes long and, because they were conducted during the HCPs’ lunch break, lunch was provided as a token of appreciation.

The interviews with patients were conducted via telephone or videoconferencing, based on participant preference. The semistructured interview guide used in the interviews was pilot-tested, leading to minor changes. The interview guide was based on the overall project aim to investigate how and why patients use VCs and included questions about patients’ introduction to VCs and factors influencing VC use. The interviews were, on average, 30 minutes long and were conducted until data saturation was reached, at which point no new information emerged. After the interview, participants received a gift card as a token of appreciation.

The focus group interviews with HCPs were conducted by MH, the moderator, with at least 1 facilitator. IM, LJH, ND, and MT facilitated at least 1 focus group interview each. The facilitators assisted the discussions by asking follow-up questions or providing the moderator with questions to ask, and by observing nonverbal communication. IM conducted the individual patient interviews. The moderator and the facilitators were all female. MH is a researcher or lecturer with a PhD in medical

XSL•FO
RenderX

informatics, IM and LJH were both PhD students in digital health. LJH is also a practicing medical doctor who works at one of the PCCs included in this study but did not facilitate the focus group interview with HCPs from her PCC. ND also has a PhD in medical informatics, and MT has a PhD in medical science. MH, who moderated the focus group interviews, and IM, who conducted the individual interviews, had extensive knowledge and experience in qualitative research at the time of the interviews. MT, who, together with IM, conducted the analysis, also had extensive knowledge and experience in qualitative research. To minimize potential bias and maintain a neutral interview environment, the researchers did not disclose any personal goals or motivations related to this study to participants.

While data saturation was reached for patient interviews, it was not fully achieved for HCPs due to practical constraints. Due to HCPs' busy schedules, many were unable to participate, and for those who could, interviews were conducted only during their lunch breaks.

Analysis

Both the focus groups and individual interviews were analyzed using rapid assessment procedures (RAP) [31]. RAP was informed by the time-sensitive nature of the broader project and the need for structured yet efficient analysis. The broader project's different parts were conducted simultaneously, and as is often the case with research, the resources were limited. To conduct all the studies and meet the strict deadline, we had to choose methods that offered efficiency without compromising rigor. RAP allows for a systematic and flexible approach to analyzing qualitative data, facilitating the rapid identification and synthesis of key themes and patterns without compromising the depth and reliability of the findings. RAP has been successfully used in health care research, particularly in implementation and evaluation studies where timely insights are crucial [32,33]. RAP was thus selected because it is an established qualitative approach that enables efficient yet rigorous data analysis; in other words, it does not compromise depth when properly executed. To support consistency and reliability in this study, we used structured templates, independently piloted the coding approach, and held regular analytic discussions to align interpretations across researchers.

This study follows an exploratory qualitative design, allowing themes to emerge organically from the data rather than fitting them into a predetermined theoretical model.

Following the RAP methodology, the first step involved creating neutral domain-based templates based on this study's aim. For step 2, the templates were independently piloted with 1 interview by IM and MT. In step 3, the analysis and usability of the template were discussed to ensure intercoder reliability, and the domains were clarified and further piloted. Of step 4, when consistency in analysis was reached, the remaining interviews were divided between IM and MT, who summarized vital points. The final domains in the RAP template were (1) patients' introduction to VC, (2) perceptions of patient users and nonusers, and (3) drivers of patients' use and nonuse. In step 5, emerging key points were discussed and categorized, after which categories were grouped into themes. We recognize that perceptions of patient users and nonusers and drivers of patient use and nonuse overlap to some degree; however, we believe that there is a distinction that is worth exploring further. IM used NVivo (version 1.4; Lumivero) and MT used pen and paper.

This study was reported in accordance with the COREQ (Consolidated Criteria for Reporting Qualitative Research) [34]. This study used the generative artificial intelligence technology Grammarly (Superhuman Platform) to enhance the grammatical accuracy and vocabulary richness.

Ethical Considerations

This study received ethical approval from the Swedish Ethical Review Authority (reference number 2021 - 05096). All study participants provided written or oral informed consent before participating in this study. In accordance with the Swedish Ethical Review Authorities' requirements, participants were informed that their participation was voluntary and that they could choose to revoke their participation at any time. Transcribed interviews and focus groups were deidentified to protect the participants' privacy, and confidentiality was preserved when reporting results. In accordance with the ethical approval, patients were compensated with a gift certificate (200 SEK, equivalent to US $20), whereas HCPs received lunch during the focus group interviews.

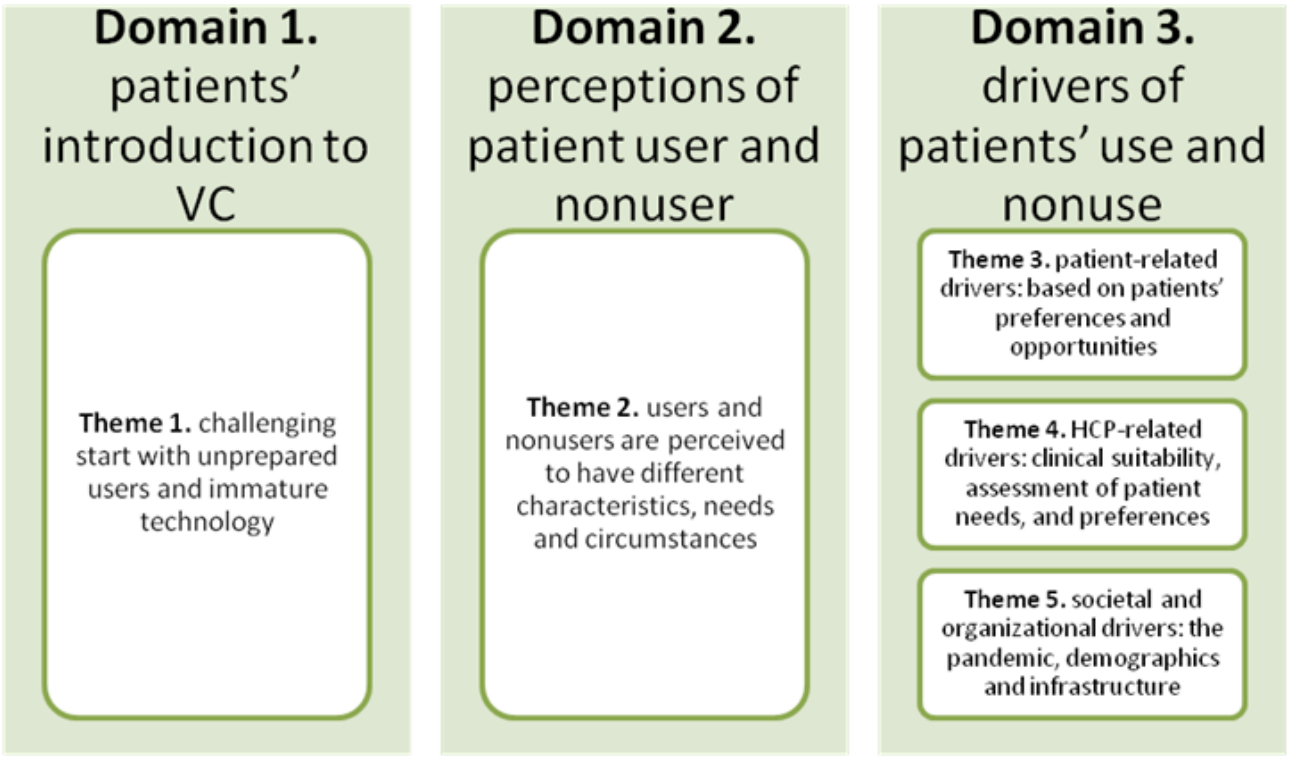
Results

Overview

A total of five themes were identified under the three predefined domains: (1) patients' introduction to VC, (2) perceptions of patients as users and nonusers, and (3) drivers of patients' use and nonuse. The first and second domains had 1 theme each, while the third had 3 themes (Figure 1).

XSL•FO
RenderX

Figure 1. The identified themes under the 3 predefined domains, rapid assessment procedures, patients, and health care professionals, Region Stockholm, 2022 - 2023. HCP: health care professional; VC: video consultation.



Domain 1. Patients' Introduction to VC

Theme 1. Challenging Start With Unprepared Users and Immature Technology

In this study, the patients' introduction refers to all activities aiming to inform patients about the service and help them start using it.

Patients' introduction to VCs encountered many challenges. On the one hand, some patients had acquired the necessary eID and accessed the application due to the COVID-19 pandemic, using it to schedule COVID-19 tests and vaccinations, which facilitated the introduction of VCs. On the other hand, the need to promptly implement VCs due to the pandemic left little room for a structured introduction of VCs.

According to HCPs, patients were provided with brochures and instruction videos online but needed guidance to find them. With the invitation to a VC, patients were also provided with instructions for downloading the application and could receive additional assistance with the download, if required, from HCPs, either over the phone or physically at the PCC. Unforeseen issues occurred during the installation of the VC application, which HCPs had to assist patients with. Significant guidance was provided over the phone in the beginning. Assisting patients in using the services was challenging for HCPs because they lacked manuals and were themselves unfamiliar with the service. They also lacked access to the patients' user interface, which made helping patients navigate the VC tool difficult. Communicating with patients with language barriers and instructing older adults with physical impairments, such as poor eyesight, had also been challenging. According to HCPs, informal caregivers or relatives were important in assisting with the use and scheduling of VCs.

HCPs also expressed that there was some resistance from patients who did not want to acquire the necessary eID; the patients did not trust it due to reports of related fraud in the media. This was experienced as challenging by HCPs because they were not able to help those without an eID, and explaining this to patients was time-consuming. Getting started was particularly challenging for older patients, according to HCPs, presumably due to low digital literacy. Older adults were stressed and worried about technical difficulties. According to HCPs, patients were inexperienced, confused, and often became disconnected due to a lack of knowledge, leading to many planned VCs being converted to TCs. HCPs tried to find what worked best from case to case. Some of this confusion was believed to be related to the application's usability. The service itself was also described as unstable.

> *There was a lot of calling back and forth in the beginning since patients couldn't connect or didn't know how to.* [FG 1]
>
> *The application or the video feature had many bugs initially, it could…I mean, half of the consultations worked, and half didn't.* [FG 2]

These reported issues improved over time as both patients and HCPs became more experienced, and the application itself was updated and became more stable.

Patients' experiences aligned with those of the HCPs. They initially experienced technical difficulties and uncertainties at both ends. Although the introduction was confusing for some patients, they were familiar with the technical difficulties, which was something they had expected. According to patients, they

XSL•FO
RenderX

received information from HCPs, found information themselves, or coincidentally found out about the service.

> *[…] then they had long waiting times, so they said; "Would you consider having a VC?" Yes, that works…I wasn't aware it was possible to book a VC.* [P1]

Domain 2. Perceptions of Patient User and Nonuser

Theme 2. Users and Nonusers are Perceived to have Different Characteristics, Needs, and Circumstances

In this study, "perceptions" refers to the attitudes, beliefs, and subjective judgments held by HCPs and patients about patient users of VCs.

According to HCPs, patient users of VCs were usually young. Users were perceived to have a positive attitude toward VCs and varied needs. They saw the benefit of VCs, wanted contact urgently, had minor complaints, or were on sick leave. Patient users were almost always follow-up patients who did not need to come to the PCC. People with psychosocial needs were also common users. Further, patient users of VCs were perceived to have special circumstances. They usually had longer geographical distances to the PCC, worked full-time, or experienced other obstacles or barriers to physically attending a meeting at the PCC.

> *Those with social phobia thought VCs were great.* [FG 3]

According to HCPs, users were perceived to have certain capabilities; they were those who scheduled a VC themselves and those who often had less serious health care needs. HCPs expressed that by doing so, they used resources that should be allocated to patients with greater needs, something they perceived as unfair.

In contrast, according to HCPs, nonusers were usually older, although not all older patients were nonusers. Nonusers usually preferred or required face-to-face visits.

Patients also perceived VCs to be great for people struggling with mental health issues, for example, social anxiety, which prevented them from seeking care in person at the PCC. Moreover, patients believed that older people lacking digital literacy were nonusers. Having at least some digital literacy or experience of internet use was seen as a prerequisite to using VCs in order to be comfortable with them. VCs were also considered to exclude people lacking technical and language skills.

> *I think a large proportion of patients are excluded. Partly because they can't handle the technology, and then there is the language barrier.* [P2]

Domain 3. Drivers of Patients' Use and Nonuse

Theme 3. Patient-Related Drivers: Based on Patients' Preferences and Opportunities

Drivers in this study refer to factors influencing use either positively or negatively.

According to HCPs, the use of VCs was significantly based on patients' preferences. The majority of patients requested face-to-face meetings or TCs. These requests were assumed to be due to familiarity with face-to-face consultations and the need for contact. On the other hand, some VCs were scheduled at patients' requests, and patients sometimes requested a VC when HCPs would have preferred a face-to-face meeting. Moreover, some HCPs always checked the patient's ability to have a VC even when they thought VCs were appropriate.

> *For those scheduled for a VC, it's my impression that it's often the patient who requested it for some reason.* [FG 4]

Impatience or need for immediate contact was another factor influencing VC use, according to HCPs. Due to long waiting times for face-to-face consultations, a VC could be chosen in order to see an HCP sooner. According to HCPs, quick contact or the ability to have a consultation instead of rescheduling due to mild but contagious illnesses were influential factors for preferring VCs. Having sick children could be a factor influencing their caregivers' use of VC. Being on sick leave or geographical distance to the PCC were not only common characteristics of users, but also acted as drivers of VC use. Patients who had moved may want to remain in contact with a health care provider with whom they had an established relationship, ensuring continuity of care. Difficulties patients experienced in booking a VC were identified by HCPs as a factor negatively influencing use.

According to HCPs, immigrants, older persons, and teenagers are, in some instances, unable to acquire an eID. Moreover, some people could be constrained by law, for example, due to conservatorship, where they are unable to make their own decisions. Other patients simply did not wish to have eIDs, according to HCPs; as previously mentioned, they did not trust this due to reports of related fraud in the media.

Some patients expressed that VCs felt impersonal compared to a consultation at the PCC, saw no need to have VCs, and preferred face-to-face and TCs for inquiries. Others preferred face-to-face as an initial contact, while VCs were more appropriate for follow-up visits, while others thought that VCs could be used for the initial assessment. The ability to make the necessary assessments would also influence use, according to the patients, as would the type of care needed and their ability to reach the PCC. Patients thought VCs worked well when avoiding infections and when they were at home with children. Additionally, the opportunity to meet with a physician and the ability to have quick contact were factors that influenced use.

> *I think it's a great compliment, especially if you're like me. I'm quite sensitive to infections, so instead of scheduling a meeting when you have a cold or are at home with a child with a cold, you could schedule a VC.* [P3]

Theme 4. HCP-Related Drivers: Clinical Suitability, Assessment of Patient Needs, and Preferences

When deciding on whether to offer a VC or not, HCPs focused on the patient's complaint, what type of care this would require, their assessment of the patient's needs, and what combinations

XSL•FO
RenderX

of these factors made a VC suitable. Making these assessments had been learned over time, as there were no clinical guidelines they could follow.

While there are some routines surrounding infections, in cases where in-person assessments and investigations were required and where VCs would be inappropriate, most decisions of whether or not to use VCs were still based on the individual HCP's assessment. According to some HCPs, VCs were potentially inappropriate for mental health patients needing cognitive behavior therapy. VCs were also perceived as inappropriate in cases where communication could be challenging, for example, patients with cognitive impairment, or severe autism, and for complex patient cases, for example, patients with co-morbidities or very old patients. There was also an initial assumption that VCs would not work well for older patients, and while some HCPs expressed that this perception had changed over time, for others, this assumption was still a driver of VC use.

> *People with multiple illnesses, maybe older individuals. That's a patient group to which I, maybe wrongly, don't offer VCs.* [FG 5]

When physical assessments were not necessary, use was based on patients' needs. Patients' safety and integrity during a VC were considered important factors. If there were concerns for patients' safety or integrity during a VC, HCPs might choose to schedule a follow-up face-to-face consultation at the PCC instead of a VC. Moreover, third parties involved in the consultation, for example, translators, could further complicate the use of VCs, but the technology could also facilitate participation of, for example, sign language interpreters. HCPs also considered VCs functional for follow-ups and saw a time advantage in having shorter VCs, potentially releasing time to serve more seriously ill patients at the PCC.

Furthermore, VC use was based on HCPs' preferences and convenience. HCPs preferred a face-to-face first encounter with patients. Some HCPs only offered their patients a face-to-face visit, and others offered VCs, sometimes without an option to choose a face-to-face visit. Physicians doing their residency would also opt for face-to-face appointments for educational reasons. Some HCPs would schedule VCs when they worked from home or elsewhere, for example, due to mild infectious diseases, thereby avoiding sick leave and rescheduling of appointments. Further, some HCPs opted for TCs due to initial technical issues. Technical limitations could also affect their assessment, for example, making VCs suboptimal for skin conditions due to the low image resolution.

According to patients, continuity of care, that is, being able to have a VC with a familiar HCP, influenced their use. Patients relied on the HCPs' judgment, and HCPs' recommendations influenced their use of VCs. Further, 1 patient expressed a willingness to consider VCs if it benefited HCPs and their work environment.

Theme 5. Societal and Organizational Drivers: the Pandemic, Demographics, and Infrastructure

According to HCPs, the COVID-19 pandemic greatly influenced and accelerated VC use for both patients and HCPs. Social distancing was a catalyst for HCPs scheduling VCs to reduce the number of patients in the waiting room.

Sociodemographic factors were other influential factors according to HCPs. They believed that having younger and older patients unable to use VCs, for example, due to a lack of necessary software, might have influenced overall VC usage. Moreover, internet connectivity, or lack thereof, for example, in rural areas, was an issue that greatly influenced VC use. Changes to the software that enable VCs, and various updates, influenced the ability to conduct VCs. Seasonal differences had also been observed, with patients requesting VCs more during the winter and summer.

> *It feels like everyone wants digital [consultations] during summer or winter, but in spring and autumn, which are working seasons, they come in [to the PCC] more.* [FG]

HCPs expressed that the lower threshold for having a VC compared to a physical appointment at the PCC had led to an increase in overall appointments. Another influencing factor was the reimbursement system, which included fees and compensation for VCs but not for TCs. HCPs felt guilty about making patients pay for issues that could be resolved through a phone call, yet felt pressured to do so to increase PCC resources. Moreover, existing work processes also influenced the use of VCs. Nurses were sometimes unable to have VCs during emergency hours because they needed to be able to respond to potential emergencies at the PCC. Furthermore, VCs were viewed not as a replacement but rather as a complement to physical care.

Patients confirmed the influence of the pandemic on their use of VCs. For some patients, being able to use a public sector application for VCs was also important.

Discussion

Principal Findings

This is one of the few studies investigating patients' introduction to VCs, perceptions of VCs' users, and drivers of use in primary care. While the COVID-19 pandemic forced a rapid introduction of VCs into Swedish primary care, positively influencing VC use, the rushed, unstructured process might also have negatively affected VC use. Perceptions and preconceptions about users and nonusers were also influential. Among other drivers of VC use were both HCPs' and patients' preferences and judgments of appropriateness, as well as several broader organizational and societal factors.

Pandemic, HCPs, and Preconceptions Influence VC Use

One of the key findings of this study was that the COVID-19 pandemic and patients' introduction to VCs both positively and negatively influenced VC use. The pandemic's influence aligns with earlier research [17,23,35]. The pandemic positively influenced VC use by facilitating its introduction and accelerating its use. The pandemic, however, also rushed implementation, contributing to a challenging start. The technology was immature, and patients and HCPs were

XSL•FO
RenderX

unprepared, which also highlighted broader issues of societal and organizational readiness for eHealth adoption. The extraordinary circumstances may have increased openness to digital solutions and partially offset the effects of a suboptimal introduction. However, the hurried implementation may have led to poor user experiences, undermining trust and satisfaction with the service. These early frustrations may also have shaped long-term attitudes toward VCs, deterring some patients and HCPs from continued use. Under normal circumstances, a more structured rollout, one that included efforts to understand and address patients' concerns, might have improved acceptance and usability. However, given the urgency of the pandemic, such user-centered planning was likely unfeasible.

Another key finding of this study was how important HCPs were in the introduction of VCs. Most patients were introduced to the service by HCPs [36]. However, the lack of preparedness among HCPs contributed to many of the challenges they experienced. This issue has also been expressed by other HCPs, where initial usability and technical issues were attributed to the lack of training of some staff [18,37].

Additionally, HCPs' preconceptions influenced VC use. Several HCPs admitted being less likely to suggest VCs to older patients, based on assumptions about digital literacy and willingness to engage with technology. Although some later recognized that age alone was not a reliable indicator of VC use, these assumptions likely shaped access to digital services in practice. These findings raise significant concerns about equity and inclusion in digital health care. When HCPs make assumptions about patients' abilities or preferences, particularly based on age, disability, language skills, or other perceived vulnerabilities, they risk reinforcing existing disparities and contributing to digital exclusion. Previous research indicates that most people with disabilities have an overall positive experience of VCs [38]. Additionally, although they encountered difficulties, patients who faced language barriers also reported high satisfaction with VC services [39].

Notably, the role of HCPs extended beyond simply offering or withholding VCs. Patients often reported relying heavily on HCPs' recommendations when deciding whether to accept a VC, underscoring the strong influence of provider endorsement. This dynamic aligns with prior research demonstrating the significant impact of HCPs on patients' willingness to adopt digital services [40].

In contrast, there is also reason to believe that bias regarding older adults and technology use exists among patients themselves. Several participants perceived older individuals as less likely to use VCs, reflecting common stereotypes around age and digital competence. This perception aligns with the concept of effort expectancy, a key factor influencing both the intention to use and actual adoption of technology [40], which may be negatively affected by assumptions that older adults lack digital literacy or confidence. If older patients expect VCs to be difficult or complex, these beliefs can undermine their self-efficacy and willingness to engage, thus reducing actual uptake.

Moreover, the initial challenges in introducing VCs in this case, particularly for older users, may have reinforced these negative perceptions, contributing to a cycle of continued nonuse or disrupted engagement. This suggests that both individual and systemic factors may interact to maintain digital exclusion among older populations. Without proactive efforts to support digital confidence and ensure accessible, user-friendly design, such exclusion risks becoming self-perpetuating.

In addition to individual perceptions, broader concerns about fairness also emerged as a factor potentially limiting the adoption of VCs. Some participants expressed the view that VCs might unfairly disadvantage certain patient groups, particularly those with more complex care needs, a concern echoed in earlier studies [17,41,42]. If VCs are perceived as excluding vulnerable or high-need populations, this could undermine their perceived value and legitimacy as a mode of care delivery more generally.

To summarize, the pandemic, HCPs, and preconceptions about and among patients influenced VC use both positively and negatively.

Preferences Influence Use and Are Related to Needs, Waiting Times, and Different Circumstances

Another key finding of this study was that VC use seemed to be greatly influenced by HCPs' and patients' preferences, which in turn were influenced by a range of different factors.

In our study, patients' preference for face-to-face consultations was related to their need for physical contact, which has been previously highlighted [17]. This could be related to the quality of VCs, where VCs do not fully emulate face-to-face interactions. Familiarity with face-to-face interaction was also thought to be related to its preference. To some extent, therefore, preference could be a matter of exposure and experience. The dynamic of an in-person meeting can also differ from a virtual one, negatively affecting the VC experience and preference. HCPs have, for example, reported less prepared patients who might try to multitask during the VC [30,37].

Challenges with waiting times within Swedish health care seem to be another significant driver of preference and VC use, according to both HCPs and patients. While VCs are offered and used for quicker contact, the waiting time could also drive inappropriate use of VCs. If VCs are chosen for their shorter waiting times, rather than their clinical appropriateness, this might negatively influence experiences and the intention to continue using VCs [11]. Moreover, as previously highlighted, VCs could reduce waiting times [37,43], whereas shorter VCs for simpler cases could allow time for severe cases. Since VCs could be offered as an alternative to rescheduling, a visit could reduce treatment delays.

Other factors that appear to influence preferences include geographical distance, seasonal factors or weather conditions, and health in general. Geographical distance could be related to accessibility for people in rural areas and those with busy schedules or mobility issues [23]. Weather conditions could also be related to mobility issues and comfort; for example, winter storms can hinder travel. It could also be related to conditions during these seasons, such as the increase in flu infections in the winter and allergies in the summer, where examinations and in-person consultations might be unnecessary.

XSL•FO
RenderX

Canadian HCPs have reported that virtual visits enable them to provide care during bad weather conditions [37]. This highlights the potential for VCs to address some of the seasonally related care delivery challenges, which could be considered during resource planning.

The ability of HCPs to choose when to have VCs offers control over their work and flexibility, which has previously been reported [35], and might positively influence their work environment, health, and preferences. However, technical issues could negatively influence their preferences and their perceived patient preferences.

Additionally, VCs have also previously been viewed as complementary due to a lack of research and quality, and usability issues have been reported as causing the preference for telephone or face-to-face consultations over VCs [9,23,37]. Further, the reports of patients requesting VCs could be evidence of an overassumption that most patients want or prefer a face-to-face consultation.

In a nutshell, personal preferences for VC or a face-to-face consultation, which heavily influence use, stem from different needs, waiting times, and different circumstances.

Potentially Misguided Judgments of Appropriateness Influence Use

Beyond individual preferences, another key finding was that both patients' and HCPs' judgments about when VCs are appropriate influence their use of VCs. However, these judgments did not always align with other research, revealing a tension between personal preference and actual clinical appropriateness.

For instance, patients in this study found VCs useful for initial contact, whereas many HCPs preferred face-to-face meetings for first-time encounters, viewing VCs as more appropriate for follow-ups. This perspective aligns with findings from a Norwegian study, where GPs expressed reluctance to use VCs with new patients [44]. Nevertheless, a systematic review of telehealth use during the COVID-19 pandemic showed that both initial and follow-up visits were commonly conducted via VC, suggesting that in practice, preferences and needs often adapt to context [45]. However, according to HCPs with experience from the pandemic, for some complex issues and with some patients, virtual care had been too time-consuming and had required more effort [37,46]. These concerns, while valid, also raise questions about how much of the perceived burden stems from a lack of support, training, or established workflows for virtual care.

According to both patients and HCPs, the ability to make clinically sound assessments was an important driver of VC use. A study on the transferability of in-person GP consultations to telehealth for patients with diabetes and cardiovascular disease found no clinical tasks that were not transferable [47]. The growing use of home monitoring tools further supports this potential. Yet, limited awareness among HCPs about what can be effectively managed virtually may unnecessarily restrict use.

Health conditions were also seen as factors influencing the use of VCs. HCPs have previously raised usability issues when assessing mental health patients as a reason for preferring face-to-face consultations [9]; however, there seems to be some nuances regarding mental health conditions. According to our findings, VCs are considered suitable for psychosocial issues in general, but some have not deemed VCs suitable for cognitive behavior therapy. Another group of conditions specifically mentioned by HCPs was skin conditions. As 1 HCP expressed, VCs are inappropriate for skin conditions due to poor imaging quality. Canadian HCPs have also reported challenges assessing skin conditions [33]. Appropriateness could depend on the quality of the devices used by patients and HCPs for skin conditions.

To sum it up, both patients' and HCPs' judgments of appropriateness influence VC use; however, the line between personal preference and actual clinical appropriateness is slightly blurred.

Infrastructure, Reimbursement, Sociodemographic, and Organizational Type Impact Use

Beyond preference and judgments of appropriateness, several organizational and societal drivers of VC use were uncovered.

This study highlighted a lack of infrastructure, potentially excluding some patients from using VCs. Restricted access to the necessary eID appears to exclude some patient groups, such as teenagers, a group that also faces barriers in other countries [18]. Language and disability-related barriers have also been attributed to the technology's shortcomings [18]. Internet connectivity issues, which have previously been described [18,37], can be more challenging in rural areas and contribute to some difficulties during the introduction, influencing substantial use. In addition, reliable systems and proper integration of the systems have been considered essential by HCPs [10].

As previously reported, reimbursement is another factor that influences VC use, both positively and negatively [23]. The reported guilt HCPs feel, in some cases, for charging for VCs highlights a potential issue with the current reimbursement system. There are also reports of work processes influencing use, where nurses might not schedule VCs at certain times. The influence of the work process on VC use has been previously reported [23], and the lack of redevelopment of workflows and routines has been described as being among the everyday challenges of VC implementation [13].

Further, the sociodemographics of the community served by the PCC are another influential factor, according to HCPs. For example, if the infrastructure, that is, eID, is not accessible to most patients, this might influence overall use and impact the HCPs' perceived usefulness of VCs and intention to use. This could also be further exacerbated by other clinicians' nonuse [19].

Whether or not the VCs were provided by a private or public organization also influenced patients' willingness to use the service. The choice between private and public health care has been extensively discussed in the Swedish media and has also been previously reported to influence the intention to use VCs [21]. This could be related to trust and credibility, as eHealth

XSL•FO
RenderX

users have expressed this as a necessity, where public actors were considered more credible than private actors [48].

To summarize, suboptimal infrastructure, organizational issues with reimbursement and work processes, as well as sociodemographics and organizational type, impact overall VC use.

Strengths and Limitations

A strength of this study is its inclusion of both HCPs and patients, offering a comprehensive view of VC use. The diversity among participants, including HCPs from various professions and centers, representing different socioeconomic areas, and patients with different characteristics and VC experiences, adds to this study's robustness.

However, the youngest patient user group was not included in this study, which is a limitation. As digital natives, younger populations may have unique perspectives on VCs, including different levels of digital literacy, preferences for communication modalities, and expectations of telehealth services. Their absence limits our ability to generalize findings to all patient demographics.

The relatively small sample size may also limit the generalizability of the findings, especially for HCPs, where data saturation was not fully achieved. However, including a diverse sample of HCPs across multiple roles and settings ensured sufficient information power, which was a strength that also reduced the risk of profession-related bias. Moreover, thematic consistency across discussions suggests key insights were captured despite the limited number of focus groups.

Additionally, there is a potential for selection bias, as patients and HCPs who chose to participate might have been more positively inclined toward VCs, potentially skewing the data toward more favorable perceptions. The focus group interview guide was not pilot-tested before data collection, which may have affected the clarity and consistency of the questions; however, we critically assessed the interview guide after each focus group to ensure all important aspects were covered and to improve the clarity of the questions. Furthermore, different tools were used in the data analysis process, both pen and paper and NVivo, which could have introduced inconsistency in the coding and interpretation of the data. Yet, the coding and analysis method used was discussed in detail between the analysts to ensure that the results would not be affected by the different tools used.

As most interviews were conducted remotely, there is a potential loss of details such as body language. However, studies comparing online and face-to-face interviews have identified strengths and weaknesses with both approaches [49], and given that we were interviewing patients and HCPs with experience of VCs, we deemed the method appropriate in this case.

A potential limitation of this study is recall bias, as interviews were conducted 2 years after the implementation of VCs. Participants' recollections may have been influenced by more recent experiences, making it difficult to capture initial reactions or challenges with complete accuracy. This may have led to an emphasis on enduring issues rather than transient concerns experienced during the early adoption phase. However, by including both HCPs and patients with varying degrees of VC experience, we aimed to mitigate this effect and capture a balanced perspective on long-term usability and adoption.

Implications and Future Work

Our findings have several key implications for stakeholders involved in the development and implementation of VCs in health care.

For developers of VC applications, this study highlights the importance of optimal usability to promote successful introduction and long-term use. Developers should prioritize extensive usability testing with both HCPs and patients before launch, ensuring the interface is intuitive and accessible to different user groups, such as older adults and people with disabilities. Continuous improvement based on user feedback is essential.

For implementers, a successful introduction of VCs requires multiple strategies. Patients' concerns need to be understood and addressed. Their digital literacy and self-efficacy also need improvement. Providing multilingual guidance, video tutorials, and accessible materials for nonnative speakers, individuals with visual impairment, and older adults is vital. As key actors in the implementation of VC, HCPs need to be better prepared, adequately trained, and supported. Educational and informational materials developed by physicians have been perceived as useful for clinics lacking prior experience and could be a potential solution to issues HCPs face [18]. Practical information regarding the guidance of people speaking other languages, those with visual impairment, and older people might also be beneficial. Additionally, training programs should help HCPs address biases and assumptions that could hinder adoption, particularly among vulnerable populations. Further, for filling the knowledge gap regarding the appropriateness of VCs and by being in discussion with both patients and HCPs, formulating clear guidelines may be key to fostering patient trust and expanding equitable use. Implementers should also assess and adjust workflows within health care organizations to minimize disruptions to service use.

Health care providers could consider the potential for VCs as a solution to some of the seasonal-related care delivery issues when resource planning.

For policymakers, this study emphasizes the importance of digital readiness, including reliable internet access and the necessary tools such as eIDs. Policymakers should invest in upgrading digital infrastructure to provide universal access, particularly in underserved areas. Additionally, supportive reimbursement systems for VCs should be established to ensure appropriate use, along with long-term investments in health care systems to reduce waiting time and improve service delivery. Enhancing health care providers' ability to implement new digital interventions is also crucial for building resilience in the face of future crises.

Further research is needed to evaluate the effectiveness of different strategies for introducing patients to VCs, with a focus on addressing inequalities. Studies should also explore younger patients' perspectives, as their digital habits and preferences

XSL•FO
RenderX

may influence adoption. Additionally, more research is required to understand how the characteristics of HCP users and nonusers affect VC usage. Future research could also explore differences in drivers of use across professional roles and between consultations with new vs known patients. Further, to better understand nonuse, future research should involve nonusers to gain their perspectives.

Conclusions

The pandemic facilitated the introduction of VCs and accelerated use. However, the rushed implementation of VC contributed to a challenging start, potentially deterring some patients and HCPs from use. HCPs played a vital role in the introductions and influenced VC use both positively and negatively. They were unprepared, which contributed to many challenges, and their misguided preconceptions risk reinforcing existing disparities and contributing to digital exclusion. However, patients also relied heavily on HCPs' recommendations when deciding to accept a VC. Moreover, patients' biases against themselves, reinforced by initial introduction challenges and broader concerns about fairness, may also influence VC use. In addition, HCPs' and patients' preferences, which were related to their needs, waiting times, and different circumstances, influence use. Further, judgments of appropriateness that also influenced use may be misguided. Lastly, infrastructure, reimbursement, sociodemographics, and organizational type were also uncovered as drivers of use.

Based on these findings, developers of VC applications should optimize usability, and implementers should include multiple strategies, including training programs, guidance, and assessing and adjusting workflows. Health care providers should consider the potential of VCs as a solution to some care-delivery issues when planning resources. Additionally, policymakers should invest in digital readiness, including upgrading digital infrastructure, creating supportive reimbursement systems, and long-term investments in health care systems. Further research should evaluate the effectiveness of different strategies for introducing patients to VCs, explore younger patients' perspectives, characteristics of HCP users and nonusers, differences between professional roles, as well as between consultation types. Nonusers' perspectives should also be explored.

Acknowledgments

We thank the participants for their contributions and the Alltid öppet management in Region Stockholm and Tiohundra administrators for facilitating data collection.

Funding

AFA Försäkring funded this work through the project "Recommendations for implementation of eHealth in primary care - ePrIm" (190210). The study was also supported by Uppsala MedTech Science and Innovation [https://medtech.uu.se/https://medtech.uu.se/], a joint strategic initiative between Uppsala University and Uppsala University Hospital.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to the qualitative nature of the materials and the terms of the ethical approval, but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Singla A, Meenu D. Telemedicine: transforming healthcare delivery in a digital age. SSJMRA 2024;1(1):71-79. [doi: 10.36676/ssjmra.v1.i1.09]
2. Ekman B, Thulesius H, Wilkens J, Lindgren A, Cronberg O, Arvidsson E. Utilization of digital primary care in Sweden: descriptive analysis of claims data on demographics, socioeconomics, and diagnoses. Int J Med Inform 2019 Jul;127:134-140. [doi: 10.1016/j.ijmedinf.2019.04.016]
3. Kirk UB, Vestergaard CH, Bech BH, Christensen MB, Kallestrup P, Huibers L. Video consultation in general practice during COVID-19: a register-based study in Denmark. BJGP Open 2024 Jul;8(2):BJGPO.2023.0208. [doi: 10.3399/BJGPO.2023.0208] [Medline: 38191187]
4. Hagstedt LJ, Hägglund M, Petersson G, Davoody N, Hvitfeldt H. Impact of the COVID-19 pandemic on use of video consultations in a Swedish primary care setting. Presented at: 18th Scandinavian Conference on Health Informatics; Aug 22-24, 2022; Tromsø, Norway p. 212-213. [doi: 10.3384/ecp187037]
5. Etz RS, Solid CA, Gonzalez MM, Britton E, Stange KC, Reves SR. Telemedicine in primary care: lessons learned about implementing health care innovations during the COVID-19 pandemic. Ann Fam Med 2023;21(4):297-304. [doi: 10.1370/afm.2979] [Medline: 37487734]
6. de Albornoz SC, Sia KL, Harris A. The effectiveness of teleconsultations in primary care: systematic review. Fam Pract 2022 Jan 19;39(1):168-182. [doi: 10.1093/fampra/cmab077] [Medline: 34278421]

XSL•FO
RenderX

7. Rush KL, Howlett L, Munro A, Burton L. Videoconference compared to telephone in healthcare delivery: a systematic review. Int J Med Inform 2018 Oct;118:44-53. [doi: 10.1016/j.ijmedinf.2018.07.007]
8. Gabrielsson-Järhult F, Kjellström S, Josefsson KA. Telemedicine consultations with physicians in Swedish primary care: a mixed methods study of users' experiences and care patterns. Scand J Prim Health Care 2021 Jun;39(2):204-213. [doi: 10.1080/02813432.2021.1913904] [Medline: 33974502]
9. Thiyagarajan A, Grant C, Griffiths F, Atherton H. Exploring patients' and clinicians' experiences of video consultations in primary care: a systematic scoping review. BJGP Open 2020;4(1):bjgpopen20X101020. [doi: 10.3399/bjgpopen20X101020] [Medline: 32184212]
10. Donaghy E, Atherton H, Hammersley V, et al. Acceptability, benefits, and challenges of video consulting: a qualitative study in primary care. Br J Gen Pract 2019 Sep;69(686):e586-e594. [doi: 10.3399/bjgp19X704141] [Medline: 31160368]
11. Muli I, Hvitfeldt H, Cajander Å, et al. To connect or not connect: long-term adoption of video consultations, and reasons for discontinuing use. J Telemed Telecare 2025 May;31(4):572-584. [doi: 10.1177/1357633X231203267] [Medline: 37787186]
12. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implementation Sci 2009 Dec;4(1):50. [doi: 10.1186/1748-5908-4-50]
13. James HM, Papoutsi C, Wherton J, Greenhalgh T, Shaw SE. Spread, scale-up, and sustainability of video consulting in health care: systematic review and synthesis guided by the NASSS framework. J Med Internet Res 2021 Jan 26;23(1):e23775. [doi: 10.2196/23775] [Medline: 33434141]
14. Greenhalgh T, Shaw S, Wherton J, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. J Med Internet Res 2018 Apr 17;20(4):e150. [doi: 10.2196/jmir.9897] [Medline: 29625956]
15. Assing Hvidt E, Atherton H, Keuper J, et al. Low adoption of video consultations in post–COVID-19 general practice in Northern Europe: barriers to use and potential action points. J Med Internet Res 2023;25:e47173. [doi: 10.2196/47173]
16. Barkai G, Gadot M, Amir H, Menashe M, Shvimer-Rothschild L, Zimlichman E. Patient and clinician experience with a rapidly implemented large-scale video consultation program during COVID-19. Int J Qual Health Care 2021 Feb 20;33(1):mzaa165. [doi: 10.1093/intqhc/mzaa165] [Medline: 33313891]
17. Hvidt EA, Christensen NP, Grønning A, Jepsen C, Lüchau EC. What are patients' first-time experiences with video consulting? A qualitative interview study in Danish general practice in times of COVID-19. BMJ Open 2022 Apr;12(4):e054415. [doi: 10.1136/bmjopen-2021-054415] [Medline: 35772823]
18. Srinivasan M, Asch S, Vilendrer S, et al. Qualitative assessment of rapid system transformation to primary care video visits at an academic medical center. Ann Intern Med 2020 Oct 6;173(7):527-535. [doi: 10.7326/M20-1814] [Medline: 32628536]
19. Glassberg B, Weiss K, Hughes T, et al. The transition to telehealth: a pilot model in a New York City student-run free clinic during the COVID-19 pandemic. J Stud Run Clin 2021;7(1). [doi: 10.59586/jsrc.v7i1.247]
20. Muli I, Cajander Å, Hvitfeldt H, et al. To read or not to read - a cross-sectional study of Swedish primary care patients' adoption of patient accessible electronic health records. Digit Health 2024;10:20552076241287636. [doi: 10.1177/20552076241287636] [Medline: 39381824]
21. Landgren S, Cajander Å. Non-use of digital health consultations among Swedish elderly living in the countryside. Front Public Health 2021;9:588583. [doi: 10.3389/fpubh.2021.588583] [Medline: 34568247]
22. Greenhalgh T, Ladds E, Hughes G, et al. Why do GPs rarely do video consultations? Qualitative study in UK general practice. Br J Gen Pract 2022 May;72(718):e351-e360. [doi: 10.3399/BJGP.2021.0658] [Medline: 35256385]
23. De Guzman KR, Snoswell CL, Giles CM, Smith AC, Haydon HM. GP perceptions of telehealth services in Australia: a qualitative study. BJGP Open 2022 Mar;6(1):BJGPO. [doi: 10.3399/BJGPO.2021.0182]
24. Internet Stiftelsen. Svenskarna och internet 2019. : Stockholm; 2019 URL: https://svenskarnaochinternet.se/app/uploads/2019/10/svenskarna-och-internet-2019-a4.pdf [accessed 2026-02-20]
25. Svenskarna och internet. 2022. URL: https://svenskarnaochinternet.se/rapporter/svenskarna-och-internet-2022/ [accessed 2026-02-20]
26. Region Stockholm. URL: https://www.regionstockholm.se/en/sprak/english/ [accessed 2026-02-20]
27. Ekonomi- och verksamhetsstatistik för regioner [Article in Swedish]. Sveriges Kommuner och Regioner. URL: https://skr.se/halsaochsjukvard/ekonomiochavgiftersjukvard/ekonomiochverksamhetsstatistiksjukvard.7872.html [accessed 2026-02-26]
28. Healthcare in Stockholm county. 1177. URL: https://www.1177.se/en/Stockholm/other-languages/other-languages/soka-vard/hitta-ratt-vard-nara-dig---andra-sprak---stockholms-lan/ [accessed 2026-02-20]
29. Verksamhetsadministration | vårdgivarguiden [Article in Swedish]. Region Stockholm. URL: https://vardgivarguiden.se/administration/verksamhetsadministration/ [accessed 2026-02-20]
30. Hägglund M, Ekwall AK, Davoody N, Farrokhnia N. Escape to the future - a qualitative study of physicians' views on the work environment, education, and support in a digital context. BMC Med Inform Decis Mak 2023 Oct 19;23(1):231. [doi: 10.1186/s12911-023-02337-7] [Medline: 37858147]
31. Renfro CP, Rome Z, Gatwood J, Hohmeier KC. Use of Rapid Assessment Procedures when analyzing qualitative data in pharmacy research. Res Soc Adm Pharm 2022 Jan;18(1):2249-2253. [doi: 10.1016/j.sapharm.2021.05.013]

XSL•FO
RenderX

32. Beebe J. Rapid Assessment Process: An Introduction: Rowman Altamira; 2001.
33. Vindrola-Padros C, Johnson GA. Rapid techniques in qualitative research: a critical review of the literature. Qual Health Res 2020 Aug;30(10):1596-1604. [doi: 10.1177/1049732320921835] [Medline: 32667277]
34. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
35. Wanderås MR, Abildsnes E, Thygesen E, Martinez SG. Hammering nails with a screwdriver: how GPs perceive video consultations. BJGP Open 2024 Dec;8(4):BJGPO.2024.0010. [doi: 10.3399/BJGPO.2024.0010] [Medline: 38866403]
36. Muli I, Taloyan M, Hagstedt LJ, Hägglund M. Patients' introduction to online video consultations in primary healthcare. Stud Health Technol Inform 2023 May 18;302:1011-1012. [doi: 10.3233/SHTI230329] [Medline: 37203556]
37. Hedden L, Spencer S, Mathews M, et al. "There's nothing like a good crisis for innovation": a qualitative study of family physicians' experiences with virtual care during the COVID-19 pandemic. BMC Health Serv Res 2023 Apr 4;23(1):338. [doi: 10.1186/s12913-023-09256-3] [Medline: 37016330]
38. Lawford BJ, Hinman RS, Morello R, Oliver K, Spittle A, Bennell KL. Perceptions about the efficacy and acceptability of telephone and video-delivered allied health care for adults with disabilities during the COVID-19 pandemic: a cross-sectional national survey. Arch Phys Med Rehabil 2022 Jul;103(7):1368-1378. [doi: 10.1016/j.apmr.2021.12.017] [Medline: 35041838]
39. Payvandi L, Parsons C, Bourgeois FC, Hron JD. Inpatient telehealth experience of patients with limited english proficiency: cross-sectional survey and semistructured interview study. JMIR Form Res 2022 Apr 19;6(4):e34354. [doi: 10.2196/34354] [Medline: 35438641]
40. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view1. MIS Q 2003 Sep 1;27(3):425-478. [doi: 10.2307/30036540]
41. Fernemark H, Skagerström J, Seing I, Ericsson C, Nilsen P. Digital consultations in Swedish primary health care: a qualitative study of physicians' job control, demand and support. BMC Fam Pract 2020 Nov 24;21(1):241. [doi: 10.1186/s12875-020-01321-8] [Medline: 33234111]
42. Murphy M, Scott LJ, Salisbury C, et al. Implementation of remote consulting in UK primary care following the COVID-19 pandemic: a mixed-methods longitudinal study. Br J Gen Pract 2021;71(704):e166-e177. [doi: 10.3399/BJGP.2020.0948] [Medline: 33558332]
43. Mueller M, Knop M, Niehaves B, Adarkwah CC. Investigating the acceptance of video consultation by patients in rural primary care: empirical comparison of preusers and actual users. JMIR Med Inform 2020 Oct 22;8(10):e20813. [doi: 10.2196/20813] [Medline: 32969339]
44. Johnsen TM, Norberg BL, Kristiansen E, et al. Suitability of video consultations during the COVID-19 pandemic lockdown: cross-sectional survey among Norwegian general practitioners. J Med Internet Res 2021 Feb 8;23(2):e26433. [doi: 10.2196/26433] [Medline: 33465037]
45. Ward K, Vagholkar S, Sakur F, Khatri NN, Lau AYS. Visit types in primary care with telehealth use during the COVID-19 pandemic: systematic review. JMIR Med Inform 2022 Nov 28;10(11):e40469. [doi: 10.2196/40469] [Medline: 36265039]
46. Halas G, Baldwin A, LaBine L, MacKay K, Singer A, Katz A. A phenomenological inquiry of the shift to virtual care delivery: insights from front-line primary care providers. Healthcare (Basel) 2024;12(8):861. [doi: 10.3390/healthcare12080861]
47. Lane J, David K, Ramarao J, et al. Translating primary care to telehealth: analysis of in-person consultations on diabetes and cardiovascular disease. BJGP Open 2023 Mar;7(1):BJGPO.2022.0123. [doi: 10.3399/BJGPO.2022.0123] [Medline: 36450404]
48. Neher M, Nygårdh A, Broström A, Lundgren J, Johansson P. Perspectives of policy makers and service users concerning the implementation of eHealth in Sweden: interview study. J Med Internet Res 2022 Jan 28;24(1):e28870. [doi: 10.2196/28870] [Medline: 35089139]
49. Lobe B, Morgan DL, Hoffman K. A systematic comparison of in-person and video-based online interviewing. Int J Qual Methods 2022 Apr;21:16094069221127068. [doi: 10.1177/16094069221127068]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
eID: electronic identification
ePrIm: implementation of eHealth in primary care
GP: general practitioner
HCP: health care professional
PCC: primary care center
RAP: rapid assessment procedure
TC: telephone consultation
VC: video consultation

XSL•FO
RenderX

Edited by A Kushniruk; submitted 11.Nov.2024; peer-reviewed by B Austad, C Eze, J Muschol, T Awofala, U Kirk; revised version received 17.Jan.2026; accepted 09.Feb.2026; published 05.Mar.2026.

Please cite as:
Muli I, Cajander Å, Davoody N, Hagstedt LJ, Hvitfeldt H, Hägglund M, Taloyan M
Health Care Professionals' and Patients' Perceptions and Experiences of Who Uses Video Consultations, and Why: Qualitative Study
JMIR Hum Factors 2026;13:e68658
URL: https://humanfactors.jmir.org/2026/1/e68658
doi:10.2196/68658

© Irene Muli, Åsa Cajander, Nadia Davoody, Lovisa Jäderlund Hagstedt, Helena Hvitfeldt, Maria Hägglund, Marina Taloyan. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 5.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Enhancing Data-Driven Decision-Making in HIV Care With Viral Load and Early Infant Diagnosis Data Dashboards in Côte d'Ivoire: Qualitative Study

Yao He[1], MPH, PhD; Yves-Rolland Kouabenan[2], MPH; Paul Henri Assoa[2], MSc; Nancy Puttkammer[1], PhD, MPH; Stephen Gloyd[3], MD, PhD; Noah Hoffman[4], MD, PhD; Bradley H Wagenaar[3,5], MPH, PhD; Casey Iiams-Hauser[1], MIA; Pascal Komena[2]; N'zi Pierre Fourier Kamelan[2]; Adama Sanogo Pongathie[6], MD; Jan Flowers[1], MSc; Nadine Abiola[2], PharmD; Natacha Kohemun[7], PharmD; Jean-Bernard Koffi Amani[7], MSc; Christiane Adje-Toure[7], MSc, MIH, PhD; Lucy A Perrone[3,4,8], MSPH, PhD

[1]Digital Initiatives Group, International Training and Education Center for Health, Department of Global Health, Schools of Public Health and Medicine, University of Washington, Seattle, WA, United States

[2]International Training and Education Center for Health- Côte d'Ivoire (I-TECH CIV), Abidjan, Côte d'Ivoire

[3]Department of Global Health, University of Washington, Seattle, WA, United States

[4]Department of Laboratory Medicine and Pathology, University of Washington, Seattle, WA, United States

[5]Department of Epidemiology, Schools of Public Health and Medicine, University of Washington, Seattle, WA, United States

[6]Direction de l'Informatique et de la Santé Digitale Sanitaire, Ministère de la Santé, de l'Hygiène Publique et de la Couverture Maladie Universelle, Abidjan, Côte d'Ivoire

[7]Division of Global HIV and TB, United States Centers for Disease Control and Prevention, Abidjan, Côte d'Ivoire

[8]Department of Pathology and Laboratory Medicine, University of British Columbia, 2211 Wesbrook Mall, Vancouver, BC, Canada

Corresponding Author:
Lucy A Perrone, MSPH, PhD
Department of Global Health, University of Washington, Seattle, WA, United States

Abstract

Background: Data dashboards are popular tools for supporting routine monitoring and decision-making in public health. Two dashboards were developed in Côte d'Ivoire to visualize laboratory data on HIV viral load (VL) and early infant diagnosis (EID) testing.

Objective: This study assessed the attitudes and experiences regarding data-driven decision-making and the VL and EID dashboards among existing and potential dashboard users in Côte d'Ivoire.

Methods: We conducted a qualitative study including 2 focus group discussions (FGDs) and 12 in-depth interviews (IDIs). The conceptual framework for the use of health data in decision-making guided the FGDs, and the Consolidated Framework for Implementation Research informed the IDIs. We used deductive and inductive approaches to analyze the interview data.

Results: The 26 participants were from 17 organizations; 11 (42.3%) were female. The participants reported a supportive data culture that valued data-driven decision-making and external pressure from the United States President's Emergency Plan for AIDS Relief (PEPFAR) that motivated data use. The dashboards were considered useful for monitoring performances and making decisions for service delivery and laboratory operations. Existing users used the dashboards regularly. Potential users expressed interest in the speed and ability to track progress. The participants considered the dashboards simple and straightforward compared to other analytical tools but suggested updating the dashboards more frequently and visualizing more data.

Conclusion: The study highlighted the importance of supportive data culture and the potential of dashboards to promote data use. However, challenges such as limited access to the internet and equipment for potential users need to be addressed.

(JMIR Hum Factors 2026;13:e76550) doi:10.2196/76550

KEYWORDS

health data dashboards; data-driven decision making; data dashboard; data visualization; user experience; HIV/AIDS

XSL•FO
RenderX

Introduction

Progress around HIV/AIDS prevention and treatment has stagnated around the world [1], and the HIV response needs to be reinvigorated to achieve the 95-95-95 goals set by the Joint United Nations Programme on HIV/AIDS (UNAIDS) [2]. HIV viral load (VL) suppression and early infant diagnosis (EID) are important health and service delivery outcomes for monitoring timely and effective treatment, preventing transmissions, and assessing prevention of mother-to-child transmission. Using data and evidence strengthens the decision-making by country leadership and contributes to successful HIV responses [3]. To help achieve the targets for VL suppression and EID in Côte d'Ivoire based on high-quality data, the national VL and EID data dashboards provide information that supports the monitoring, evaluation, and decision-making in HIV-related services and programs.

Understanding the context, culture, and behavior around data demand and use is important to ensure the effectiveness of the tools that aim to support data-driven decision-making, especially in low- and middle-income countries (LMICs). According to a 2023 literature review, the main barriers to data-driven decision-making in LMICs include insufficient data use competencies, low data quality, inadequate data availability, lack of systems design and user-centered design, disconnect between data producers and users, unsupportive leaders, absence of organizational supports, misalignment in data needs between donors and local stakeholders, weak data use culture, and low individual motivation [4]. Evidence is limited to whether and how these barriers influence data-driven decision-making in HIV responses in Côte d'Ivoire, and such evidence will provide useful context for understanding the successes and issues with the VL and EID dashboards.

Dashboards are tools designed to support data-driven decision-making by visualizing quantitative data and supplying audit and feedback to decision makers and service providers [5,6]. Applications of dashboards are ubiquitous in public health, with purposes ranging from disease surveillance [7-9] and monitoring of quality and efficiency [10-14] to clinical decision support [15-18] and stock management, among others [19,20].

Despite the proliferation of dashboards in public health, studies that assess data needs and user experience of dashboard users are limited. After reviewing 1191 papers, the authors of a systematic review on public health dashboards identified only 18 user studies, while most studies describe only the design and technical development process of a specific dashboard [21]. Only a few of the 18 user studies considered data demands and self-efficacy of dashboard users [21], highlighting a notable gap in research that other reviews already identified previously [22,23].

This study aimed to fill the research gaps described above by answering the following questions: What were the attitudes and experiences of existing and potential dashboard users on data-driven decision-making in their organizations? What were the participants' impressions of the VL and EID dashboards?

Methods

Intervention Description

To respond to the need for a tool that consolidates and presents data in an accessible format, the publicly accessible, online dashboard for VL was developed in 2017 through the collaboration among the International Training and Education Center for Health (I-TECH), the Côte d'Ivoire Ministry of Health (MOH), the United States President's Emergency Plan for AIDS Relief (PEPFAR), the United States Centers for Disease Control and Prevention (CDC), The United States Agency for International Development, the Clinton Health Access Initiative, and the Kenyan MOH [24,25]. I-TECH obtained the software source code in the HTML for the Kenyan VL dashboard and adapted it for the Ivorian context to visualize monthly aggregate VL data from the open-source enterprise-level laboratory information system (OpenELIS; Digital Initiatives Group at the University of Washington [26]), the electronic laboratory information system in Côte d'Ivoire [27,28]. The visualization tool was Highcharts, and the data sources were monthly OpenELIS datasets at the laboratory level. The use case was to provide analyzed, critical information for monitoring program performance and progress toward achieving the goals for viral load testing and suppression and EID. The target users were the CDC and the PEPFAR implementing partners, policy makers at different levels of the MOH in Côte d'Ivoire, and clinical and laboratory service providers.

Data visualizations produced by the VL dashboard included bar charts and pie charts for the number of tests, turnaround time, and viral suppression with disaggregation by gender, age group, and reasons for seeking care. All charts except for the national average turnaround time could be disaggregated by month, region, district, clinical laboratory, or PEPFAR implementing partners. Users could access the disaggregated visualizations through the navigation tabs at the top of the dashboard webpage. A dedicated tab showed charts about viral suppression by gender, age, reason for the test request, and antiretroviral regimen. All data visualizations, source datasets, and a dashboard user guide were globally accessible and available for download.

In 2020 the EID dashboard was created based on the VL dashboard to display monthly aggregate EID data from OpenELIS [29]. The types of visualizations and functions are similar to those described above for the VL dashboard. More details about the development process, information visualized, and types of visualizations of the VL and EID dashboards were described elsewhere [24].

Study Design and Data Collection

We conducted a qualitative study including 2 focus group discussions (FGDs) and 12 semistructured in-depth interviews (IDIs). Specifically, the FGDs were intended to gather general perspectives that the participants converged or diverged on about data-driven decision-making and the VL and EID dashboards. The IDIs were intended to probe for greater depth of response with themes that arose in the FGDs. Data collection took place from September 2021 to April 2022 through in-person or online meetings.

XSL•FO
RenderX

The study population consisted of existing and potential users of the VL and EID dashboards. Existing users were staff members of national and regional clinical laboratories, the CDC office in Côte d'Ivoire, and PEPFAR implementing partners. Potential users were from MOH divisions and clinical laboratories that reportedly had not used the dashboards, but the I-TECH project team and existing users thought these entities would find the dashboards useful.

To enroll existing users, we purposefully sampled individuals who were already using the dashboards based on previous meetings to participate in the first FGD and subsequent IDIs. We then used snowball sampling to ask the FGD participants to introduce us to other existing users from additional organizations to participate in the IDIs.

To enroll potential users, we first purposefully sampled a staff member of the Programme National de Lutte contre le Sida (PNLS; National AIDS Control Program) who had shared with the study team that they had not used the dashboards. The study team also identified individuals who had not used the dashboards and were heads or deputy heads of the regional clinical laboratories in Côte d'Ivoire. We purposefully chose PNLS because, as the national technical group created by the MOH to guide the national HIV/AIDS response, PNLS knew about the dashboards but reported that they did not routinely use them. Regional laboratories might also be interested in the dashboards because they were referral sites that processed all VL and EID samples referred from other laboratories in the region. We invited identified potential users to participate in the second FGD and subsequent IDIs. Recruitment stopped once we confirmed that all targeted types of organizations were represented and that thematic saturation had been reached.

The conceptual framework for the use of health data in decision-making (referred to as the "data use framework" hereafter) informed the interview guides for the 2 FGDs [4,30]. The data use framework builds on the Performance of Routine Information System Management (PRISM) framework created by the MEASURE Evaluation group [4,30,31]. The framework highlights key activities that most directly influence data demand and use and subsequently the building blocks in the health system and ultimately health outcomes. We chose to focus on the activities that would influence data demand and use to see if and how they were reflected in the participants' responses.

The relevant constructs of the Consolidated Framework for Implementation Research (CFIR) informed the interview guides for the IDIs (Table 1) [32,33]. The CFIR is a meta-theoretical framework designed to explain the factors that affect translation of interventions into practice and their sustainment in practice across different contexts [32,33].

The interview guides were initially developed in English and later translated into French. The FGDs and IDIs were recorded, and comprehensive notes were taken. On average, the FGDs lasted 59 minutes, and the IDIs lasted 32 minutes. The recordings were transcribed verbatim in French and translated into English for analysis.

A combination of deductive and inductive approaches was used to analyze the data. Codes were created using the activities that influence data demand and use from the data use framework, the CFIR constructs, as well as those derived from the data. All transcripts were coded by one primary coder (YH). YH first chose one FGD transcript and one IDI transcript to code with the initial codebook that consisted of the deductive codes. After adding inductive codes to the codebook, YH revised the coding in the first two transcripts and continued coding the rest of the transcripts. Any subsequent changes in the codebook or coding approach were applied to transcripts that had been coded previously. We summarized the coded information into themes that corresponded to the research questions. The two frameworks served as the basis for the themes, with additional themes derived from the data that did not fit the constructs of the two frameworks. Two additional team members (YRK and PHA) reviewed the coded data and thematic summaries to validate the analysis and ensure consistency in interpretation. The coding and thematic analysis were conducted in ATLAS.ti (8 Windows; ATLAS.ti Scientific Software Development GmbH).

XSL•FO
RenderX

Table . The *a priori* components of interest from the conceptual framework for the use of health data in decision-making and the Consolidated Framework for Implementation Research (CFIR).

Data collection method	Conceptual framework for the use of health data in decision-making[a]	CFIR constructs[b]
• FGD1[c]: existing dashboard users • FGD2: potential dashboard users	• Data use context • Data demand and use infrastructure	None.

IDIs[d] with individual existing or potential dashboard users	• Data use context
	• Data demand and use infrastructure
	• Data users and data producers
	• Data quality
		• Innovation
		  • Relative advantage
		  • Adaptability
		  • Design
		• Inner setting
		  • Access to knowledge and information
		• Outer setting
		  • External pressure – market pressure
		  • Policies and laws
		  • Financing
		• Individuals
		  • Capability

[a]Also referred to as the data use framework.

[b]Both the 2019 and 2022 versions of the CFIR were referenced.

[c]FGD: focus group discussion.

[d]IDI: in-depth interview.

Ethical Considerations

This study was determined to be nonhuman participants research by the University of Washington Institutional Review Board and approved by Côte d'Ivoire Comité National d'Ethique des Sciences de la Vie et de la Santé (CNESVS, Ivorian Institutional Review Board; reference number 006 - 21/MSHP/CNESVS-km). All participants were unpaid volunteers and provided written informed consent. To ensure data confidentiality, the data collected were stored in a project-specific, encrypted instance of Google Drive managed by the secure server environment of the University of Washington that is compliant with the US Health Insurance Portability and Accountability Act (HIPAA) guidelines.

Results

Overview

We conducted one FGD and 7 IDIs with existing users and another FGD and 5 IDIs with potential users. There were 26 participants across the FGDs and IDIs; 11 (42.3%) were female; 17 organizations were represented (Table 2). The participants from the CDC and PNLS were program officers; the participants from the PEPFAR implementing partners were clinical program leads or strategic information team members; and the participants from national or regional referral laboratories were heads or deputy heads of the laboratories.

XSL•FO
RenderX

Table . Characteristics of the study participants who were existing or potential users of the HIV viral load and early infant diagnosis data dashboards in Côte d'Ivoire in 2021 - 2022.

Characteristics	Overall	FGD1[a] with existing users	FGD2 with potential users	IDIs[b] with existing users	IDIs with potential users
Participants from each type of organization, n (%)	26 (100)	9 (34.6)	5 (19.2)	7 (26.9)	5 (19.2)
US CDC[c]	3 (11.5)	2 (22.2)	1 (20.0)	0 (0.0)	0 (0.0)
PEPFAR[d] IP[e]	9 (34.6)	7 (77.8)	0 (0.0)	2 (28.6)	0 (0.0)
PNLS[f]	1 (3.8)	0 (0.0)	0 (0.0)	0 (0.0)	1 (20.0)
National referral laboratory	6 (23.1)	0 (0.0)	1 (20.0)	4 (57.1)	1 (20.0)
Regional referral laboratory	7 (26.9)	0 (0)	3 (40.0)	1 (14.3)	3 (40.0)
Organizations represented (mins)	17	6	5	6	5
Participants by sex, n (%)					
Female	11 (42.3)	3 (33.3)	2 (40.0)	2 (28.6)	4 (80.0)
Male	15 (57.7)	6 (66.7)	3 (60.0)	5 (71.4)	1 (20.0)
Mean recording length in minutes	36	70	48	36	26

[a]FGD: focus group discussion.

[b]IDI: in-depth interview.

[c]US CDC: United States Centers for Disease Control and Prevention.

[d]PEPFAR: United States President's Emergency Plan for AIDS Relief.

[e]IP: implementing partner.

[f]PNLS: Programme National de Lutte contre le Sida (National AIDS Control Program).

Data-Driven Decision-Making

Data Use Context

According to all the participants in the FGDs and IDIs, their organizations used data in decision-making, and there was no difference between the organizations of the existing users and those of the potential users. Not only would the leaders use data in decision-making, but different units and levels within each organization also used data in their work, even if some units or levels were not seen as decision makers. Data were used routinely to monitor technical and operational performance as well as ad hoc under special circumstances to assess the situation, identify problems or needs, and make decisions to solve problems or make improvements. The most common areas where the participants used data to monitor were laboratory operational processes (eg, nonconformity of samples, sample flows, and turnaround time) and lost-to-follow-up clients for VL testing.

Data Demand, Data Use Infrastructure, Data Users, and Data Producers

The FGD with existing users revealed convergence among PEPFAR implementing partners in demands for data, data use infrastructure, and internal processes that engaged data users and producers (Table 3). PEPFAR implementing partners, not directly providing services, used both laboratory and clinical data to monitor laboratory operational processes and progress of HIV program implementation. Laboratory data were from OpenELIS, and clinical data were exported from the electronic health record for HIV clients into Microsoft Excel spreadsheets. These organizations had data management units that compiled and validated the data collected from the clinical laboratories in the implementing districts. The programmatic units then received the processed data and conducted data analyses to monitor or evaluate progress achievements and shortfalls. The analyses informed leaders in deciding how their organizations could support the clinical laboratories to address areas of improvement.

While 3 referral laboratories (11.5% of the 26 participants) used data to inform laboratory operations and clinical decision-making, 4 other laboratories (15.4%) used data only for laboratory operations (Table 3). Three participants (11.5%) from 2 national referral laboratories and one regional referral laboratory shared that they not only used operational data from OpenELIS to monitor nonconformities and improve sample flows and result validation, but they also actively contributed to clinical decision-making by examining data on testing results and whether a client was on time for a test. The 10 participants (38.5%) from the rest of the referral laboratories were most interested in laboratory operational data and viewed themselves as suppliers of data to the clinicians for making decisions. As an existing user noted, "we give [the clinician] his result so it's

XSL•FO
RenderX

a chain which we play a part in," describing laboratory staff as data providers.

Table . Data demanded, data use infrastructure, and uses of the data demanded at the 2 types of organizations where most participants of this study worked in Côte d'Ivoire in 2021 - 2022.

Themes	CDC[a], PEPFAR[b] implementing partners	Referral laboratories
What data were demanded	• Laboratory data on VL[c], EID[d] • Clinical data on active client case • Pharmacy data on filled prescriptions	• Data on laboratory operational process
What infrastructure supported data use	• OpenELIS[e] • Clinical records (electronic health records for HIV clients; Excel spreadsheets)	• OpenELIS
What issues were understood or addressed using data	• Identify districts or laboratories with performance issues • Track lost-to-follow-up clients for VL testing or treatment • Identify clients with unsuppressed VL	• Bottlenecks in sample flow • Nonconforming samples • Turnaround time • Workload
What decisions or strategies were made using data	• Follow-up with specific districts, laboratories, or clients	• Improve workflow • Follow-up with satellite laboratories about nonconforming samples

[a]CDC: US Centers for Disease Control and Prevention.

[b]PEPFAR: US President's Emergency Plan for AIDS Relief.

[c]VL: viral load.

[d]EID: early infant diagnosis.

[e]OpenELIS: open-source enterprise level laboratory information system.

External Pressure

The participants from PEPFAR implementing partners and the Retro-CI laboratory, established and funded by the US CDC, shared that how they used data in decision-making was heavily influenced by PEPFAR and the MOH.

The participants from the laboratories in the public health system shared that their data use behavior already existed to support internal performance monitoring and address MOH data requirements. They would share their existing data with donors or partners in the areas that these external entities supported, eg, HIV testing and donated equipment. An existing user stated that they reported data to PEPFAR every month because they were supported by PEPFAR, "but it's really not (PEPFAR) who influence our activity," indicating that the laboratory maintained operational independence despite donor reporting requirements.

Data Quality

When asked how data quality issues affect how often they used data for decision-making, the participants shared that data timeliness had a more direct impact on the frequency of data use. All existing users who participated in the first FGD shared that they had internal processes to ensure the quality of the data in the dashboards. Four participants from the second FGD with potential users shared that, since they had technical validations, the data quality should be fine. However, one participant disagreed, stating that, despite the processes and units that were intended to ensure data quality, there were still missing data or poorly documented data on paper forms that decreased data quality.

VL and EID Dashboards

Frequency of Using the Dashboards

Three existing users shared that they used the dashboards monthly; 2 said quarterly; and one said every 6 months because they were most interested in stock shortages, but they said that the dashboards were not up to date on stock. One existing user said they only used the dashboards on an ad hoc basis because the dashboards were not updated as frequently as they preferred.

Reason for Using the Dashboards

The existing users considered the dashboards useful tools for identifying and raising awareness of problems and important references for communicating with other laboratories. An existing user from a national referral laboratory felt excited and relieved that the dashboards were available because they provided direct evidence of laboratory performances, and there was no need to guess how the performances had been. An existing user at a PEPFAR implementing partner agency was interested in the performances of other similar organizations, which the dashboards showed, so that they could cross-compare and "boost our team to work more."

After examining the dashboards briefly during the FGD and IDIs, potential users shared the reasons that would motivate them to use the dashboards were serving as a model to peers ("people outside the laboratory can copy us and seek to get in touch with us to see how we have achieved this" [IDI with potential user 5]); accessing information quickly for decision-making ("very quickly we know where [tasks] get stuck [in the workflow] and what quick decision to take." [FGD potential user 5]); and monitoring progress toward targets ("good

XSL•FO
RenderX

to know that we have feedback [from the dashboards] about our implementation of the national guidelines." [IDI with potential user 1]).

Facilitators

All existing users were confident that they themselves and the colleagues who used the dashboards routinely at work were capable of using the dashboards. The FGD participants within each group converged that the design of the dashboards was nice, the colors were beautiful and distinct, and the layout was easily readable. Two existing users (of the 16; 12.5%) shared that the need for the dashboards was common across the PEPFAR implementing partners and the US CDC, so the motivation for using the dashboards already existed.

The participants shared that the relative advantages of the dashboards were simplicity and straightforwardness compared to other data analysis and visualization tools such as Microsoft Excel or Microsoft Power BI. Compared to static data spreadsheets, the visualizations on the dashboards showed a lot more information in an interactive way. The participants explained that, since the configurations and structures of the dashboard interface were predefined, accessing specific information was faster and easier for people who monitor performances routinely. One potential user at a regional referral laboratory thought that, compared to Excel, which could also visualize data, the dashboards were faster and more straightforward, and it was also a direct way to convey information to others who wanted to learn about the VL and EID activities of the laboratory or the region. One potential user and one existing user highlighted the benefit of an interactive tool that could quickly display information that they needed, for example, meaningful disaggregation by population group, whereas Excel spreadsheets were static. As one potential user noted, ("with a simple click, […]we do not need to go and redo the calculations again [in Excel] to be able to access the information").

Although one existing user perceived Excel and Power BI as more versatile in analyzing data exported from OpenELIS, they acknowledged that the dashboards were "accessible" (IDI with existing user 4). Sometimes if she "had not finished analyzing [data exported from OpenELIS], [she] will take the information from the dashboards," indicating that the dashboards provided faster access to analyzed information that was useful for decision-making.

In terms of market pressure, the existing users converged during the first FGD that they were aware of other organizations, especially PEPFAR implementing partners, that were using the dashboards. In the subsequent IDIs, an existing user shared that

> *generally all the partners in Côte d'Ivoire use the dashboards, I exchange with colleagues in other organizations, and everyone refers to the dashboards* [IDI with existing user 5]

Most potential users (9 out of 10; 90%) did not know of any organization that used the dashboards. The only participant who had prior knowledge came from PNLS, which oversees all HIV-related activities in the country.

Potential Barriers

Three existing users (of the 16; 18.8%) shared that, although they were able to use the dashboards without difficulty, they were concerned that new users might need internet connections and computers or tablets to access these online dashboards. Both groups of existing and potential users would like to have more information and training or coaching on using the dashboards. One existing user thought that training or coaching would be useful for new users to orient themselves around the layout and functions of the dashboards, noting that "[new users] may be impressed, but the use will be difficult, they need explanations and coaching." Indeed, all but 2 potential users (80%) had not heard of the dashboards through any training or awareness session, making it the primary reason why they had not used them.

Suggestions for Improving the Dashboards

The most mentioned suggestion was updating the data displayed in the dashboards more frequently. Four existing users (25%) and 3 potential users (30%) suggested updating at least every other week, with more near–real-time updates being preferred. One potential user cautioned the others during the FGD about having an ideal world with near real-time data updates, since there was a natural trade-off between the speed of posting data and the necessary biological and technical validations in the laboratory operational process. Therefore, weekly updates might be the most ideal scenario that the dashboards could realistically achieve (FGD potential user 5).

The participants also suggested various information to add to the dashboards so they can meet the needs of the organizations better. Three existing users (18.8%) and 2 potential users (20%) would like to have more indicators about the laboratory operational process and service quality, eg, the number of noncompliant and rejected samples and testing validity. Information about when and why major disruptions in laboratory services happened was highly requested by all but 2 existing users to provide context around suboptimal performance. The potential user from PNLS would like the dashboards to extend to all laboratory testing related to people living with HIV rather than just VL and EID.

Discussion

Principal Findings

This study highlights the integral role of data-driven decision-making in HIV program implementation and laboratory operations in Côte d'Ivoire. The findings reveal that data are widely used across organizational levels to monitor performance, address challenges, and inform decisions, but application varies between referral laboratories and PEPFAR implementing partners. While some organizations focus on operational monitoring, others integrate data into clinical decision-making, reflecting differences in roles and priorities. While existing users valued the dashboards for tracking performance and benchmarking, potential users faced barriers related to access, awareness, and training. Concerns about the need for additional indicators suggest opportunities for improvement. Strengthening capacity-building and data integration could enhance the impact

XSL•FO
RenderX

of the dashboards on decision-making. Articles that describe dashboards displaying HIV-related data are not uncommon [34-40]. This study not only describes the features of the VL and EID dashboards in Côte d'Ivoire but also explores contextual factors, dashboard complexity, hardware availability, and internet connectivity that influenced the dashboard user experience.

The participants not only recognized the importance of data but also reported concrete behaviors of using data in decision-making. The recognition and behaviors were shared by both leaders and other staff members in an organization. This data culture, especially among leadership, seemed to be more supportive than in some LMICs reported in other studies [41,42]. The finding in our study could reflect a sample of participants who highly valued data, since more than half of the participants were existing dashboard users and had monitoring and evaluation as one of their key responsibilities.

We observed a virtuous cycle starting from a supportive data use culture at PEPFAR partner organizations and clinical laboratories in Côte d'Ivoire, which made people willing to use the VL and EID dashboards in their work. MOH supervisors can use the dashboards to provide regular feedback and comparative results to health facilities and laboratories, showcasing to both the data producers and users how the submitted data were used and enhancing the value of data collection and use [43-45]. The perceived value of data ultimately fed back into a supportive data use culture.

The participants at clinical laboratories shared that they submitted data to PEPFAR implementing partners monthly in addition to the routine data collection and use process established by the MOH. This indicates parallel data systems due to donors' data demands, which may have been burdensome and led service providers to view themselves only as data producers [4,46]. Having access to useful tools such as the VL and EID dashboards that visualize the data that the laboratories submitted might motivate them to assume the role of data users [47].

All participants reported that their organizations had processes to validate data or ensure data quality, which gave them confidence in the quality of the data they collected, reported, and used. Key data validation or quality assurance activities include using the data validation report in OpenELIS, the electronic laboratory information system, and conducting routine data quality assessments. Our prior research supports this impression by showing that OpenELIS, the electronic laboratory information system feeding data into the dashboards, improved data timeliness and completeness significantly [48]. However, since poor data quality remains a barrier to data-driven decision-making in LMICs [41], more targeted research is necessary to understand the correlation between data quality, users' confidence in data quality, and actual data use behaviors.

The findings about how the study participants viewed the dashboards conformed with what other studies reported, but the perceived facilitators and barriers for using the dashboards were an addition to the evidence. Ease of use was one of the most assessed dashboard characteristics [21], and the study participants found the VL and EID dashboards simple and straightforward. Simple innovations have a higher likelihood of effectiveness due to their ability to enhance user satisfaction and shorten the learning curve [49-51]. For dashboards that intend to simplify the data synthesis and analysis process for users, low complexity is an important determinant of effectiveness.

The market pressure from peer organizations may have motivated existing users to use the dashboards. This phenomenon was especially present among the organizations that worked directly with PEPFAR. It would be interesting to explore among the potential users who work indirectly with PEPFAR if the same phenomenon would apply when some potential users start using the dashboards while others do not. Late adopters often face significant pressure to implement innovations, especially when competitors or colleagues have already adopted them; when others in the field are using an innovation, individuals and organizations may feel compelled to follow suit [52]. In order for market pressure to exist, potential users first must know about the innovation. Marketing the dashboards through training and awareness activities as tools for building an organization's credibility could be necessary before market pressure can build.

The existing users, all based in Abidjan or other urban areas, were concerned that potential users in other regions of the country might not have the necessary equipment or Internet connection to access the online dashboards and that they might need training to learn to use the dashboards. It is important to ensure the precondition of necessary equipment is met for achieving the outcome of increased user pool in the future [53]. Providing training or a support hotline might also be useful strategies to engage potential users [54].

Limitations

The FGDs and IDIs were carried out by I-TECH personnel who were involved in the development and implementation process of the dashboards, potentially leading to social desirability bias in reporting barriers or issues. However, since we explained to the participants that the study was formative rather than evaluative, the social desirability bias might not have manifested to a great extent. Other limitations include potential difficulties in recalling or misunderstanding the questions. To partially address these limitations, we provided explanations of the interview questions and encouragement to offer honest critiques and suggestions for improvement.

In addition, the study sample was purposively selected and not intended to be statistically representative of all existing or potential users of the dashboards. Since this was a qualitative study, we did not collect objective observations or produce quantitative estimates of the phenomena discussed by the informants. Thus, we were unable to assess how common particular experiences or opinions may have been. Future mixed methods research could help corroborate the qualitative insights reported by triangulating them with quantitative data on dashboard use patterns and implementation outcomes.

Finally, since a single study team member coded all transcripts, there is a possibility of subjective bias or limited interpretive diversity in the coding process. This risk was mitigated by

XSL•FO
RenderX

independent review and validation of the thematic analysis by 2 additional team members.

Conclusion

This study demonstrated that decision makers and program staff in Côte d'Ivoire's HIV response all valued data-driven decision-making. They recognized the utility of the HIV VL and EID dashboards in monitoring performance and guiding decisions. While PEPFAR partners valued laboratory, clinical, and pharmacy data and used them in client- or program-level monitoring, laboratories demanded data on laboratory operational processes to improve workflow and quality control. These insights contribute to the broader understanding of effective data use and dashboard design in Côte d'Ivoire.

Acknowledgments

We thank the Ministry of Health, Public Hygiene, and Universal Health Coverage (MSHPCMU) in Côte d'Ivoire and all implementing laboratories for the continuous collaboration. We thank Marisa Van Osdale at I-TECH for program management; Ahoua Koné at the University of Washington for her guidance and insight on the Ivorian context; and the project interns and data consultants at I-TECH CIV.

The US Centers for Disease Control and Prevention reviewed the paper but had no influential role in study design, data collection, analysis, and interpretation.

Funding

This work was supported by PEPFAR through the US Centers for Disease Control and Prevention under the terms of cooperative agreement #NU2GGH001968-05-00 awarded to the University of Washington. The findings and conclusions in this paper are those of the authors and do not necessarily represent the official position of the funding agencies.

Data Availability

The data that support the findings of this study are not openly available due to protection of privacy and confidentiality and are available from the corresponding author, LP, upon reasonable request.

Conflicts of Interest

None declared.

References

1. In danger: UNAIDS global AIDS update 2022. Joint United Nations programme on HIV/AIDS (UNAIDS). 2022. URL: https://www.unaids.org/sites/default/files/media_asset/2022-global-aids-update_en.pdf [accessed 2023-07-19]
2. Understanding fast-track: accelerating action to end the AIDS epidemic by 2030. Joint United Nations programme on HIV/AIDS (UNAIDS). 2015. URL: https://www.unaids.org/sites/default/files/media_asset/201506_JC2743_Understanding_FastTrack_en.pdf [accessed 2023-07-19]
3. The path that ends AIDS: UNAIDS global AIDS update 2023. Joint United Nations programme on HIV/AIDS (UNAIDS). 2023. URL: https://www.unaids.org/sites/default/files/media_asset/2023-unaids-global-aids-update_en.pdf [accessed 2023-07-19]
4. Barriers to use of health data in low- and middle-income countries: a review of the literature. MEASURE evaluation. 2018. URL: https://www.measureevaluation.org/resources/publications/wp-18-211.html [accessed 2023-07-19]
5. Dasgupta N, Kapadia F. The future of the public health data dashboard. Am J Public Health 2022 Jun;112(6):886-888. [doi: 10.2105/AJPH.2022.306871] [Medline: 35613427]
6. Powell BJ, Waltz TJ, Chinman MJ, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb 12;10(1):21. [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
7. Grange ES, Neil EJ, Stoffel M, et al. Responding to COVID-19: the UW medicine information technology services experience. Appl Clin Inform 2020 Mar;11(2):265-275. [doi: 10.1055/s-0040-1709715] [Medline: 32268390]
8. Lourenço C, Tatem AJ, Atkinson PM, et al. Strengthening surveillance systems for malaria elimination: a global landscaping of system performance, 2015-2017. Malar J 2019 Sep 18;18(1):315. [doi: 10.1186/s12936-019-2960-2] [Medline: 31533740]
9. Cheng CKY, Ip DKM, Cowling BJ, Ho LM, Leung GM, Lau EHY. Digital dashboard design using multiple data streams for disease surveillance with influenza surveillance as an example. J Med Internet Res 2011 Oct 14;13(4):e85. [doi: 10.2196/jmir.1658] [Medline: 22001082]
10. Cassim N, Tepper ME, Coetzee LM, Glencross DK. Timely delivery of laboratory efficiency information, part I: developing an interactive turnaround time dashboard at a high-volume laboratory. Afr J Lab Med 2020;9(2):947. [doi: 10.4102/ajlm.v9i2.947] [Medline: 32391244]
11. Woo JS, Suslow P, Thorsen R, et al. Development and implementation of real-time web-based dashboards in a multisite transfusion service. J Pathol Inform 2019;10(1):3. [doi: 10.4103/jpi.jpi_36_18] [Medline: 30915257]

XSL•FO
RenderX

12. Austin J, Barras M, Sullivan C. Interventions designed to improve the safety and quality of therapeutic anticoagulation in an inpatient electronic medical record. Int J Med Inform 2020 Mar;135:104066. [doi: 10.1016/j.ijmedinf.2019.104066] [Medline: 31923817]
13. Waller RG, Wright MC, Segall N, et al. Novel displays of patient information in critical care settings: a systematic review. J Am Med Inform Assoc 2019 May 1;26(5):479-489. [doi: 10.1093/jamia/ocy193] [Medline: 30865769]
14. Crofts J, Moyo J, Ndebele W, Mhlanga S, Draycott T, Sibanda T. Adaptation and implementation of local maternity dashboards in a Zimbabwean hospital to drive clinical improvement. Bull World Health Organ 2014 Feb 1;92(2):146-152. [doi: 10.2471/BLT.13.124347] [Medline: 24623908]
15. Van Dort BA, Zheng WY, Sundar V, Baysari MT. Optimizing clinical decision support alerts in electronic medical records: a systematic review of reported strategies adopted by hospitals. J Am Med Inform Assoc 2021 Jan 15;28(1):177-183. [doi: 10.1093/jamia/ocaa279] [Medline: 33186438]
16. Dowding D, Randell R, Gardner P, et al. Dashboards for improving patient care: review of the literature. Int J Med Inform 2015 Feb;84(2):87-100. [doi: 10.1016/j.ijmedinf.2014.10.001] [Medline: 25453274]
17. Khairat SS, Dukkipati A, Lauria HA, Bice T, Travers D, Carson SS. The impact of visualization dashboards on quality of care and clinician satisfaction: integrative literature review. JMIR Hum Factors 2018 May 31;5(2):e22. [doi: 10.2196/humanfactors.9328] [Medline: 29853440]
18. Micallef C, Chaudhry NT, Holmes AH, Hopkins S, Benn J, Franklin BD. Secondary use of data from hospital electronic prescribing and pharmacy systems to support the quality and safety of antimicrobial use: a systematic review. J Antimicrob Chemother 2017 Jul 1;72(7):1880-1885. [doi: 10.1093/jac/dkx082] [Medline: 28369528]
19. Agarwal S, Glenton C, Henschke N, et al. Tracking health commodity inventory and notifying stock levels via mobile devices: a mixed methods systematic review. Cochrane Database Syst Rev 2020 Oct 28;10(10):CD012907. [doi: 10.1002/14651858.CD012907.pub2] [Medline: 33539585]
20. Mabirizi D, Phulu B, Churfo W, et al. Implementing an integrated pharmaceutical management information system for antiretrovirals and other medicines: lessons from Namibia. Glob Health Sci Pract 2018 Dec 27;6(4):723-735. [doi: 10.9745/GHSP-D-18-00157] [Medline: 30591578]
21. Schulze A, Brand F, Geppert J, Böl GF. Digital dashboards visualizing public health data: a systematic review. Front Public Health 2023;11:999958. [doi: 10.3389/fpubh.2023.999958] [Medline: 37213621]
22. Budd J, Miller BS, Manning EM, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug;26(8):1183-1192. [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]
23. Murphy DR, Savoy A, Satterly T, Sittig DF, Singh H. Dashboards for visual display of patient safety data: a systematic review. BMJ Health Care Inform 2021 Oct;28(1):e100437. [doi: 10.1136/bmjhci-2021-100437] [Medline: 34615664]
24. Kirk M, Assoa PH, Iiams-Hauser C, et al. Adaptation of an electronic dashboard to monitor HIV viral load testing in Côte d'Ivoire. Afr J Lab Med 2021;10(1):1284. [doi: 10.4102/ajlm.v10i1.1284] [Medline: 34192117]
25. Viral load dashboard. Chargevirale.openelisci.org. URL: https://chargevirale.openelisci.org/vl_dashboard/ [accessed 2023-07-19]
26. OpenELIS global – open-source laboratory information system. OpenELIS Global. URL: https://openelis-global.org/ [accessed 2026-02-10]
27. Kenya viral load dashboard. Viral Load Dashboard (NASCOP). URL: https://viralload.nascop.org/ [accessed 2023-07-19]
28. He Y, Iiams-Hauser C, Henri Assoa P, et al. Development and national scale implementation of an open-source electronic laboratory information system (OpenELIS) in Côte d'Ivoire: sustainability lessons from the first 13 years. Int J Med Inform 2023 Feb;170:104977. [doi: 10.1016/j.ijmedinf.2022.104977] [Medline: 36608629]
29. HIV early infant diagnosis (EID) dashboard. Chargevirale.openelisci.org. URL: https://chargevirale.openelisci.org/eid_dashboard/fr/ [accessed 2023-07-19]
30. Nutley T, Reynolds HW. Improving the use of health data for health system strengthening. Glob Health Action 2013 Feb 13;6(1):20001. [doi: 10.3402/gha.v6i0.20001] [Medline: 23406921]
31. Aqil A, Lippeveld T, Hozumi D. PRISM framework: a paradigm shift for designing, strengthening and evaluating routine health information systems. Health Policy Plan 2009 May;24(3):217-228. [doi: 10.1093/heapol/czp010] [Medline: 19304786]
32. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated consolidated framework for implementation research based on user feedback. Implement Sci 2022 Oct 29;17(1):75. [doi: 10.1186/s13012-022-01245-0] [Medline: 36309746]
33. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implementation Sci 2009 Dec;4(1). [doi: 10.1186/1748-5908-4-50]
34. Braunstein SL, Coeytaux K, Sabharwal CJ, et al. New York City HIV care continuum dashboards: using surveillance data to improve HIV care among people living with HIV in New York City. JMIR Public Health Surveill 2019 Jun 19;5(2):e13086. [doi: 10.2196/13086] [Medline: 31219053]
35. Jwanle P, Ibiloye O, Obaje M, et al. Accelerating HIV epidemic control in Benue state, Nigeria, 2019-2021: the APIN program experience. Ther Adv Infect Dis 2023;10:20499361231153549. [doi: 10.1177/20499361231153549] [Medline: 36814516]

XSL•FO
RenderX

36. Lahuerta M, Syowai M, Vakil S, et al. Monitoring the transition to new antiretroviral treatment regimens through an enhanced data system in Kenya. PLoS ONE 2020;15(4):e0232104. [doi: 10.1371/journal.pone.0232104] [Medline: 32324800]
37. Metz M, Smith R, Mitchell R, et al. Data architecture to support real-time data analytics for the population-based HIV impact assessments. JAIDS Journal of Acquired Immune Deficiency Syndromes 2021;87(1):S28-S35. [doi: 10.1097/QAI.0000000000002703]
38. Michael S, Gompels M, Sabin C, Curtis H, May MT. Benchmarked performance charts using principal components analysis to improve the effectiveness of feedback for audit data in HIV care. BMC Health Serv Res 2017 Jul 24;17(1):506. [doi: 10.1186/s12913-017-2426-6] [Medline: 28738800]
39. Radin AK, Abutu AA, Okwero MA, et al. Confronting challenges in monitoring and evaluation: innovation in the context of the global plan towards the elimination of new HIV infections among children by 2015 and keeping their mothers alive. J Acquir Immune Defic Syndr 2017 May 1;75 Suppl 1(Suppl 1):S66-S75. [doi: 10.1097/QAI.0000000000001313] [Medline: 28398999]
40. Vrazo AC, Sullivan D, Ryan Phelps B. Eliminating mother-to-child transmission of HIV by 2030: 5 strategies to ensure continued progress. Glob Health Sci Pract 2018 Jun 27;6(2):249-256. [doi: 10.9745/GHSP-D-17-00097] [Medline: 29959270]
41. Akhlaq A, McKinstry B, Muhammad KB, Sheikh A. Barriers and facilitators to health information exchange in low- and middle-income country settings: a systematic review. Health Policy Plan 2016 Nov;31(9):1310-1325. [doi: 10.1093/heapol/czw056] [Medline: 27185528]
42. Mate KS, Bennett B, Mphatswe W, Barker P, Rollins N. Challenges for routine health system data management in a large public programme to prevent mother-to-child HIV transmission in South Africa. PLoS ONE 2009;4(5):e5483. [doi: 10.1371/journal.pone.0005483] [Medline: 19434234]
43. Rendell N, Lokuge K, Rosewell A, Field E. Factors that influence data use to improve health service delivery in low- and middle-income countries. Glob Health Sci Pract 2020 Sep 30;8(3):566-581. [doi: 10.9745/GHSP-D-19-00388] [Medline: 33008864]
44. Harrison T, Nutley T. A review of constraints to using data for decision making: recommendations to inform the design of interventions. MEASURE Evaluation. 2010. URL: https://www.measureevaluation.org/resources/publications/tr-10-77/at_download/document [accessed 2010-07-19]
45. Qazi MS, Ali M. Health management information system utilization in Pakistan: challenges, pitfalls and the way forward. Biosci Trends 2011 Dec;5(6):245-254. [doi: 10.5582/bst.2011.v5.6.245] [Medline: 22281538]
46. Ikamari L, Adewuyi A, Akinlo A. Decision maker perceptions in Kenya and Nigeria: an assessment of data use constraints. MEASURE Evaluation. 2007. URL: https://www.measureevaluation.org/resources/publications/tr-07-44/at_download/document [accessed 2023-07-19]
47. Kim D, Sarker M, Vyas P. Role of spatial tools in public health policymaking of Bangladesh: opportunities and challenges. J Health Popul Nutr 2016 Feb 27;35(1):8. [doi: 10.1186/s41043-016-0045-1] [Medline: 26922788]
48. He Y, Kouabenan YR, Assoa PH, et al. Laboratory data timeliness and completeness improves following implementation of an electronic laboratory information system in Côte d'Ivoire: quasi-experimental study on 21 clinical laboratories from 2014 to 2020. JMIR Public Health Surveill 2024 Mar 20;10:e50407. [doi: 10.2196/50407] [Medline: 38506899]
49. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629. [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
50. Gustafson DH, Sainfort F, Eichler M, Adams L, Bisognano M, Steudel H. Developing and testing a model to predict outcomes of organizational change. Health Serv Res 2003 Apr;38(2):751-776. [doi: 10.1111/1475-6773.00143] [Medline: 12785571]
51. Klein KJ, Conn AB, Sorra JS. Implementing computerized technology: an organizational analysis. J Appl Psychol 2001 Oct;86(5):811-824. [doi: 10.1037/0021-9010.86.5.811] [Medline: 11596799]
52. Walston SL, Kimberly JR, Burns LR. Institutional and economic influences on the adoption and extensiveness of managerial innovation in hospitals: the case of reengineering. Med Care Res Rev 2001 Jun;58(2):194-228. [doi: 10.1177/107755870105800203] [Medline: 11398646]
53. Lewis CC, Klasnja P, Powell BJ, et al. From classification to causality: advancing understanding of mechanisms of change in implementation science. Front Public Health 2018;6:136. [doi: 10.3389/fpubh.2018.00136] [Medline: 29868544]
54. Hamoy GL, Amoranto AJP, Evangelista-Sanchez AMA, et al. Real-time regular routine reporting for health (R4Health): lessons from the implementation of a large scale mobile health system for routine health services in the Philippines. Acta Med Philipp 2016;50(4). [doi: 10.47895/amp.v50i4.775]

Abbreviations

CDC: Centers for Disease Control and Prevention
CFIR : Consolidated Framework for Implementation Research
CNESVS: Comité National d'Ethique des Sciences de la Vie et de la Santé
EID: early infant diagnosis

XSL•FO
RenderX

FGD: focus group discussion
HIPAA: Health Insurance Portability and Accountability Act
I-TECH: International Training and Education Center for Health
IDI: in-depth interview
LMICs : low- and middle-income countries
MOH: Ministry of Health
OpenELIS: open-source enterprise-level laboratory information system
PEPFAR: United States President's Emergency Plan for AIDS Relief
PNLS: Programme National de Lutte contre le Sida (National AIDS Control Program)
PRISM: Performance of Routine Information System Management
UNAIDS: Joint United Nations Programme on HIV/AIDS
VL: viral load

Edited by Y Senathirajah; submitted 25.Apr.2025; peer-reviewed by B Abimbola, M Padilla; revised version received 11.Dec.2025; accepted 08.Jan.2026; published 31.Mar.2026.

Please cite as:
He Y, Kouabenan YR, Assoa PH, Puttkammer N, Gloyd S, Hoffman N, Wagenaar BH, Iiams-Hauser C, Komena P, Kamelan NPF, Pongathie AS, Flowers J, Abiola N, Kohemun N, Amani JBK, Adje-Toure C, Perrone LA
Enhancing Data-Driven Decision-Making in HIV Care With Viral Load and Early Infant Diagnosis Data Dashboards in Côte d'Ivoire: Qualitative Study
JMIR Hum Factors 2026;13:e76550
URL: https://humanfactors.jmir.org/2026/1/e76550
doi:10.2196/76550

© Yao He, Yves-Rolland Kouabenan, Paul Henri Assoa, Nancy Puttkammer, Stephen Gloyd, Noah Hoffman, Bradley H Wagenaar, Casey Iiams-Hauser, Pascal Komena, N'zi Pierre Fourier Kamelan, Adama Sanogo Pongathie, Jan Flowers, Nadine Abiola, Natacha Kohemun, Jean-Bernard Koffi Amani, Christiane Adje-Toure, Lucy A Perrone. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 31.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Patients' and Health Care Professionals' Experiences of a Digital Self-Management System for Asthma: Qualitative Study

Lovisa Jäderlund Hagstedt[1,2], MD; Helena Hvitfeldt[2,3], PhD; Sara Riggare[3,4], PhD; Göran Petersson[5], Prof Dr Med; Nadia Davoody[1], PhD; Maria Hägglund[3,6], PhD

[1]Department of Learning Informatics Management and Ethics, Health Informatics Center (HIC), Karolinska Institutet, Stockholm, Sweden

[2]R&D, Tiohundra AB, Norrtälje, Sweden

[3]Department of Women's and Children's Health, Participatory eHealth and Health Data, Uppsala University, Dag Hammarskjölds väg 14B, Uppsala, Sweden

[4]Centre for Disability Studies, Uppsala University, Uppsala, Sweden

[5]Department of Medicine and Optometry, Linnaeus University, Kalmar, Sweden

[6]Uppsala MedTech Science & Innovation, Uppsala University Hospital, Uppsala, Sweden

Corresponding Author:
Maria Hägglund, PhD
Department of Women's and Children's Health, Participatory eHealth and Health Data, Uppsala University, Dag Hammarskjölds väg 14B, Uppsala, Sweden

Abstract

Background: Living with asthma—especially in its severe forms—can significantly impact daily life, including social activities, work, travel, and household responsibilities. Collaboration between patients and health care professionals (HCPs) is frequently lacking, particularly regarding treatment goals. Self-management has been shown to mitigate the negative effects of asthma. Technical solutions might support self-management for patients with chronic diseases and their collaboration with HCPs.

Objective: This explorative study aims to understand how patients and HCPs experience the use of a digital self-management system for asthma monitoring.

Methods: This qualitative study was conducted at 5 primary care centers in Sweden and involved 20 participants: 14 patients who had utilized the digital self-management system Asthmatuner for at least 6 months and 6 specialist asthma nurses. Individual semistructured interviews were analyzed using qualitative content analysis to explore patterns and relationships within the data.

Results: We identified 1 main theme, that is, "data-supported empowerment," and 3 subthemes, that is, (1) empowerment by awareness, knowledge, and learning; (2) contact health care-patient; and (3) managing the monitoring. The theme of data-supported empowerment emerged as a synthesis of these findings, reflecting how the self-management system enabled patients to take a more active role in managing their medications and health. While most patients did not monitor their data continuously, they engaged with it when they felt it was necessary. Some patients expressed expectations of personalized follow-up from HCPs based on their monitoring data; however, these expectations were not always fulfilled. We also revealed a need to adapt and clarify the overlapping responsibilities of patients and HCPs.

Conclusions: The digital self-management system for asthma was well received by both patients and HCPs, as it promoted empowerment. Clear communication about changes in workflow and responsibilities is essential to ensure the successful implementation of digital systems and improved health care delivery.

(JMIR Hum Factors 2026;13:e79866) doi:10.2196/79866

KEYWORDS

asthma; digital health; empowerment; primary care; self-management; user experience

Introduction

Health care faces multiple challenges, including an aging population, a growing number of individuals living with chronic diseases and multimorbidity, increasing demands from the public, limited resources, and expanding medical possibilities [1]. These challenges highlight the need for innovative approaches to health care delivery, particularly in primary care, where most chronic disease management takes place [2-4].

Person-centered care and patient empowerment are necessary to meet the needs of persons with chronic conditions [5-8]. The World Health Organization defines empowerment as "a process through which people gain greater control over decisions and actions affecting their health …" [9]. To achieve patient empowerment, the patient needs to understand his or her role,

XSL•FO
RenderX

possess the necessary knowledge and skills, and be in a facilitating environment.

The concept of self-management is widely used in chronic care contexts and often refers to patients' ability to manage symptoms, treatment, and lifestyle changes. Individuals with chronic conditions are often expected to take significant responsibility for their daily care and disease self-management [10]. However, this has also been critiqued for placing responsibility on individuals without sufficiently acknowledging structural, relational, and contextual factors that shape health behaviors [11,12]. While there is no universally accepted definition of self-management, this study adopts the definition: "… the individual's ability to manage the symptoms, treatment, physical and psychosocial consequences, and lifestyle changes inherent in living with a chronic condition" [13]. Self-monitoring is a key component of self-management, as it enhances awareness and provides actionable information [10,14]. Effective self-management can improve health outcomes, reduce health care utilization, and lower overall health care costs [15-18]. In this study, we focus on the use of digital tools that support patients in their everyday self-management of asthma, while recognizing that such tools are embedded in broader care practices and relationships.

Digital health technologies have shown promise in supporting self-management of chronic conditions [19], with positive effects reported for both patients and health care professionals (HCPs) [20,21]. Asthma, a chronic respiratory disease, affects approximately 10% of the Swedish population, with higher prevalence among adult women [22,23]. Living with asthma—especially in its severe forms—can significantly impact daily life, including social activities, work, travel, and household responsibilities. Patients often report psychological distress and dissatisfaction with the continuity of care in primary settings. Moreover, collaboration between patients and HCPs is frequently lacking, particularly regarding treatment goals. These findings underscore the importance of structured patient education, which may be enhanced by digital health solutions [24].

In Sweden, asthma and chronic obstructive pulmonary disease nurses are registered nurses with several years of clinical experience and specialized training in respiratory care, including university-level courses in asthma, allergy, and chronic obstructive pulmonary disease. Their responsibilities encompass patient assessment, spirometry testing, education on disease management and inhalation techniques, smoking cessation support, and coordination of individualized care plans. These nurses play a central role in long-term follow-up and in supporting patients' self-management [25].

Digital technology has the potential to improve both the health and care of patients with asthma [26-29], but there is limited evidence concerning patients' experiences and their views of digital intervention in the self-management of asthma [28,30]. Customizable treatment plans and self-management through the self-monitoring of lung function and symptoms appear to be more effective than other forms of asthma self-management [31]. The digital tool AsthmaTuner (MediTuner AB) is a Conformité Européenne–marked self-management system that offers personalized treatment recommendations based on daily lung function and symptom monitoring [32].

The user experience of digital self-management interventions will have an impact on patients' and HCPs' adoption and continued use of such tools and is therefore important to understand. According to ISO 9241 - 210, user experience is "a person's perceptions and responses that result from the use or anticipated use of a product, system or service" [33] and includes emotions, beliefs, behaviors, and responses before, during, and after use, shaped by system functionality, prior experiences, and context.

This explorative study aims to understand how patients and HCPs experience the use of a digital self-management system (AsthmaTuner) for asthma monitoring.

Methods

Study Design

A qualitative approach was used to understand the experiences of patients and HCPs. The data were collected through semistructured interviews with patients and specialist nurses who had used the self-management system AsthmaTuner for at least 6 months. The COREQ (Consolidated Criteria for Reporting Qualitative Research) has guided the reporting of this study [34] (Checklist 1).

Study Setting

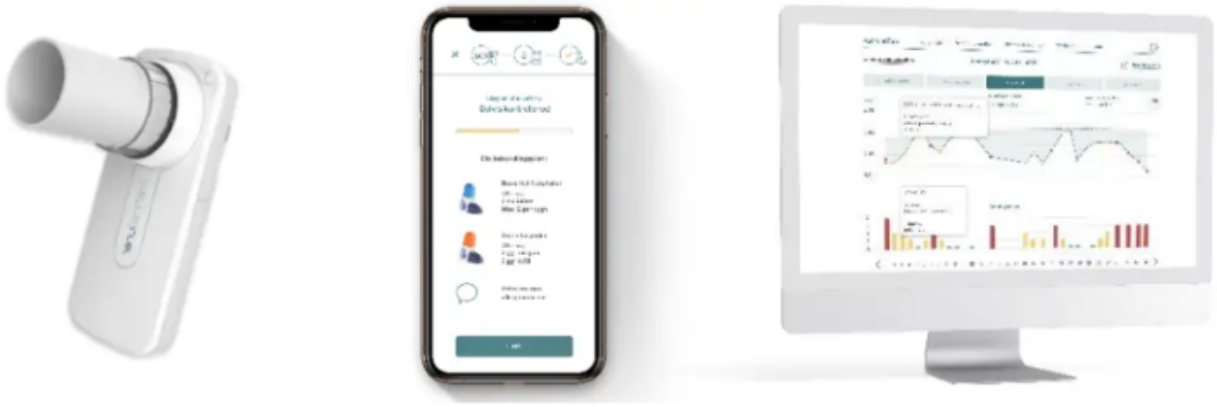
The study took place at 5 primary care centers in Region Stockholm, Sweden. These centers are part of the primary care service in Tiohundra AB in Norrtälje. Norrtälje is the largest municipality area-wise in Region Stockholm; it is sparsely populated and has an aging population, and its inhabitants have lower socioeconomic status than other parts of Region Stockholm. This demographic and socioeconomic profile more closely resembles that of many municipalities across Sweden outside the major urban centers, making Norrtälje a relevant case for understanding primary care in rural and semirural contexts nationwide.

The Self-Management System—Asthmatuner

Asthmatuner is a Conformité Européenne–marked self-management system for monitoring asthma. It consists of a portable wireless spirometer that can be used to measure lung function and a mobile app through which patients answer questions about their current asthma symptoms in conjunction with measuring their lung function. The app provides patients with a treatment plan based on their current symptom estimates and lung function. Figure 1 demonstrates an overview of the AsthmaTuner components. The treatment plan is presented in 3 levels—controlled, partly controlled, or uncontrolled—and suggests how the patient should take their medication. The app presents a view of trends in their symptoms and their lung function [32].

XSL•FO
RenderX

Figure 1. Overview of the components of the AsthmaTuner self-management system (wireless spirometer, patient app, and health care professional [HCP] interface).



In addition, the self-management system includes an interface for the HCPs where patient data can be viewed and treatment plans prescribed. Being able to view patient data allows the HCPs to identify patients with poor asthma control in need of follow-up.

Study Participants and Recruitment

Patients were recruited from the intervention arm of an ongoing randomized controlled trial in which 200 patients were given access to the self-management system, while the 200 patients in the control group were assigned to usual care and received a written or verbally communicated treatment plan. Written informed consent was obtained from all participants before inclusion.

Letters with information about the qualitative study were sent from the primary care centers to all patients who had used the self-management system for at least 6 months and were 18 years or older. Patients who responded to the invitation were subsequently contacted to schedule an interview. Invitation letters were sent out in 4 rounds as new patients became eligible. In total, 14 patients, 11 female/3 male, aged 44 to 76 years, mean age 64.5 (SD 10.1 y), participated in the study.

To ensure variation in the sampling, we applied a broad inclusion strategy to capture a wide range of experiences relevant to our study objectives. We used participant recruitment across different primary care centers from various geographic locations to reach a diverse audience, enhancing the representativeness of our sample. Despite a small sample, it corresponded roughly to the larger group regarding age and gender.

All specialist asthma nurses working at the participating primary care centers were approached and agreed to be interviewed (N=6). No further details (age or gender) about the nurses will be reported to ensure their privacy.

Data Collection

Semistructured interview guides were designed by the research team and used for the interviews (Multimedia Appendices 1 and 2). Semistructured interviews were chosen to allow for both consistency across interviews and flexibility to explore participants' individual experiences in depth. This method is well suited for qualitative research aiming to understand complex and personal phenomena, such as the use of a self-management system in health care settings [35]. The guides were based on previously constructed interview guides and modified using the team's experiences of qualitative interviews. They were not pilot-tested due to time constraints. The interviews were conducted in Swedish either by phone or online, recorded, and transcribed verbatim. This approach allowed for greater flexibility in scheduling and reduced the need for travel or physical presence, which was particularly important given the participants' varying health conditions, time constraints, and geographical locations. The interviews lasted 20 to 45 minutes, and no repeat interviews were performed. Only the interviewer and the participant were present during the interview. None of the interviewing researchers had any prior relationship with the patients, and they were not involved in patient care. The first author worked as a primary care physician at one of the participating care centers and, therefore, did not participate in data collection to avoid biased responses.

The first 10 patient interviews and all the nurses' interviews were performed during 2021 to 2022. To reach data saturation from the patient interviews, 4 additional interviews were conducted in 2023. Since these interviews did not present any new findings, we concluded that saturation had been reached.

The participant checking of transcripts and analysis was not done to reduce the burden of participating in the study.

Data Analysis

The interviews were analyzed using an inductive content analysis approach guided by Graneheim and Lundman [36].

LJH, MH, ND, and SR performed the initial coding. They read all the transcribed interviews and shared the work of identifying relevant meaning units and the initial coding. To identify codes and subcategories, the coding process was first performed on the interviews with patients and nurses separately. ND, MH, and SR performed the initial coding of the patient interviews, and LJH and MH took the main responsibility for coding the nurses' interviews. When subcategories had been identified, they were compared, and similarities and differences between the nurses and patients emerged. At this stage, the decision was made to continue the categorization and identification of subthemes jointly for both groups. This not only reduced the repetition of the findings due to similarities between groups but

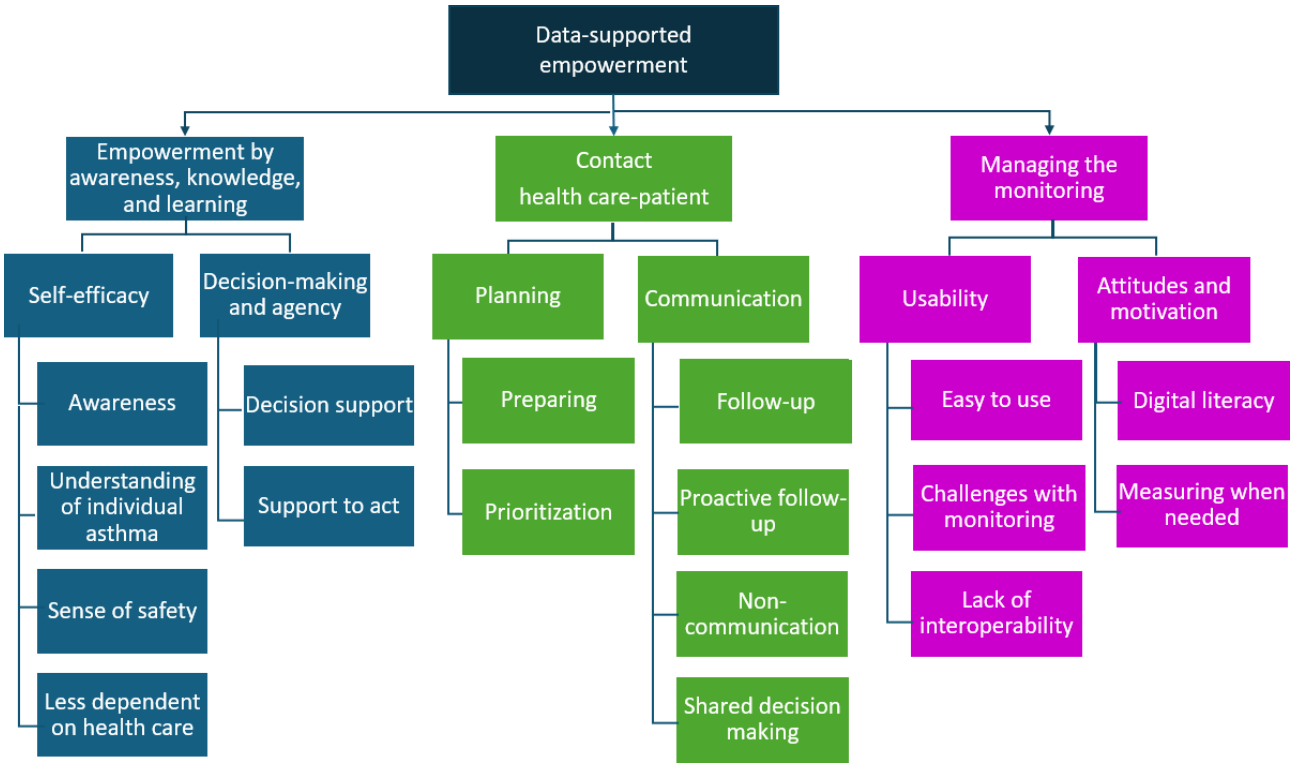
XSL•FO
RenderX

also enabled us to highlight the subcategories where nurses and patients expressed different experiences or opinions. To enhance transparency and support the trustworthiness of the analysis, we include an example of the coding process in line with the COREQ checklist [34] in Table 1. The complete results, with all themes and categories, are presented in the results section (Figure 2). The coding was structured according to the approach described by Graneheim and Lundman [36], where meaning units were condensed, coded, and grouped into subcategories and categories through an iterative interpretative process.

Table . Example of the analytical process, showing how meaning units were condensed, coded, and grouped into subcategories and categories, following the qualitative content analysis approach described by Graneheim and Lundman [36].

Subthemes of data-supported empowerment and their categories	Subcategory	Meaning unit
Empowerment by awareness, knowledge, and learning		
Decision-making and agency	Support to act	"I notice tendencies, like now it's getting a little worse. And then I can start premedication before the values get too affected, so my asthma doesn't have to become so bad." [Patient 2, female, 58 years old]
Self-efficacy	Understanding of individual asthma	"I get an understanding of my asthma and when it gets worse. By measuring continuously, I know how I'm doing, and I can control my activities and my medication." [Patient 14, female, 52 years old]
Contact health care-patient		
Communication	Proactive follow-up	"Sometimes they have called and said, 'I keep ending up in the red zone. What's wrong?' They have contacted us then, just as they should when the device tells them to." [Specialist nurse]
Communication	Noncommunication	"… I would have liked to have more feedback. But that doesn't really have anything to do with the app." [Patient 11, male, 71 years old]

Figure 2. Main theme, subthemes, categories, and subcategories.



Consensus regarding the categories and themes was reached through an iterative process in a dialogue between all four coders and HH, who also read all the interviews. The meaning units were translated into English after the analysis was completed.

XSL•FO
RenderX

Ethical Considerations

The study received ethical approval by the Swedish Ethical Review Authority (2019 - 01898). All study participants gave written informed consent. All transcribed interviews were pseudonymized before analysis, to ensure the participants privacy was preserved. Study participants were not reimbursed for their participation, in accordance with the ethical approval.

Results

Overview

The analysis resulted in the identification of 1 overarching theme: data-supported empowerment, which built on three subthemes: (1) empowerment by awareness, knowledge, and learning; (2) contact health care-patient; and (3) managing the monitoring. Each subtheme encompassed several categories relevant to both patients and HCPs, with some subcategories applying to only one of these groups (Figure 2).

Rather than aiming to identify a single main theme from the outset, the analysis explored patterns and relationships within the data. The theme of data-supported empowerment emerged as a synthesis of these findings, reflecting how the self-management system enabled patients to take a more active role in managing their medications and health. While most patients did not monitor their data continuously, they engaged with the data when they felt it was necessary. Some patients expressed expectations of personalized follow-up from HCPs based on their monitoring data; however, these expectations were not always fulfilled.

Empowerment by Awareness, Knowledge, and Learning

Overview

The self-management system facilitated awareness, understanding, and knowledge of the patient's condition and supported patients in taking a more active part in their care and treatment. Self-efficacy and decision-making and agency thus became the main categories of the subtheme empowerment by awareness, knowledge, and learning.

Self-Efficacy

The category self-efficacy refers to participants' perceived ability to manage their asthma effectively, including interpreting symptoms, adjusting medication, and making informed decisions about daily activities. This interpretation is grounded in Bandura's concept of self-efficacy, which emphasizes individuals' belief in their capacity to influence outcomes through their actions [37]. Most patients experienced that monitoring gave them a better awareness of how asthma affects their well-being, and HCPs described how patients who used the self-management system became more interested in their condition and treatment.

For some patients, monitoring not only gave them a better awareness of asthma in general but also a higher level of understanding of their individual asthma.

> *It gives me an understanding of my asthma and when it gets worse. By measuring continuously, I know how I'm doing, and I can control my activities and my medication.* [Patient #14, female, 52 years old]

Patients who experienced this increase in awareness, which was also acknowledged by HCPs, used the monitoring to follow their lung function and adjust their treatment, and the monitoring also helped them to identify situations and factors that triggered exacerbations, such as different allergens or infections.

Several patients stated that the monitoring gave them a sense of safety because it helped them to determine whether their symptoms were caused by their asthma or something else, but also because they were encouraged to adjust their treatment if needed.

> *I get good advice through it [the system]. It makes me feel safe. And my asthma has been great since I started using it.* [Patient #4, female, 68 years old]

Some patients felt less dependent on health care as a result of using the self-management system since they could manage their asthma based on the treatment recommendations provided.

Decision-Making and Agency

The self-management system provided patients with information about their asthma control and suggested dosage adjustments of medications based on their current symptoms and lung function. Both patients and HCPs found that the self-management system served as an important decision support tool for the patients and gave them support to act on the suggestions.

> *Yes, it feels like when I get bad measurements, I should increase [the dose of my medication]. Or it says I should take a bit more Bricanyl. I think it's brilliant.* [Patient #6, male, 61 years old]

> *I notice tendencies, like now it's getting a little worse. And then I can start premedication before the measurements get too affected, so my asthma doesn't have to become so bad.* [Patient #2, female, 58 years old]

Contact Health Care-Patient

Interactions between patients and HCPs were affected by the self-management system, and 2 main categories were identified: planning and communication.

Planning

The self-management system was used by both patients and HCPs to prepare for planned follow-ups. Patients used it to recall variations in asthma control since the last visit, enabling them to give more informed answers.

HCPs used the data for prioritization when deciding which patients were most in need of active follow-up.

> *This way, I get better control of how the patients with asthma are doing. And I can call the right patients; I can prioritize the patients who need to come here.* [Specialist nurse #1]

XSL•FO
RenderX

Communication

Patient monitoring enabled digital follow-up. Patients and HCPs could discuss the data via phone or video consultations. The system was used for proactive follow-up, where nurses contacted patients with deteriorating asthma control and offered guidance. HCPs appreciated it when patients contacted them based on system alerts.

> *Sometimes they've called and said, "I keep ending up in the red zone. What's wrong?" They've contacted us then, just as they should when the device tells them to.* [Specialist nurse #3]

However, there were also cases of noncommunication, where some patients experienced a lack of feedback from the HCPs, and some had problems contacting their specialist nurse.

During regular follow-ups, the data provided by the system were used for shared decision-making, where patients and HCPs examined and reflected on the collected data together.

> *During a follow-up visit, you look at the registered data together with the patient, discuss how they have been, and try to find out what it was that made them deteriorate.* [Specialist nurse #5]

Managing the Monitoring

Usability and attitudes and motivation were the 2 main categories related to the management of monitoring.

Usability

Patients and HCPs agreed that the self-management system was easy to use. The HCPs noted that most patients did not have any problems using the system and were pleased to discover that older patients also managed to use the technology successfully.

> *I am surprised at how many patients, even older people, can handle the digital technology, electronic identification, and an app like this* [Specialist nurse #1]

Even though all patients thought the self-management system was easy to use, they also experienced some challenges with the monitoring. The most important challenge was remembering to use the device. Patients found that the option to use a reminder in the app was helpful, but after using it for some time, some patients got so used to the reminder that they no longer reacted to it.

HCPs did not report any challenges with monitoring, but instead a lack of interoperability, specifically concerning the need for the double documentation of the patient's medications and treatment plan in the system's care portal and the patient's electronic health record.

Attitudes and Motivation

Digital literacy affects the patient's attitude and motivation when deciding whether to try the self-management system. HCPs described how patients with high digital literacy were eager to try the self-management system, while those with low digital literacy were more reluctant and needed more support to make it work. For some patients, wanting to challenge themselves and learn how to use a new technology was a motivator.

> *I thought, "this is going to be interesting. I must try this!" And then I also thought, as I often do, if you find that something is difficult, then you should definitely learn it.* [Patient 8, female, 74 years old]

Patients were encouraged to use the self-management system at least once a week, although after using it for a while, most patients started measuring when needed, usually in connection with periods of exacerbation of their asthma. Some patients, however, planned to use it more regularly before a health care appointment to provide additional data on their current asthma control for the HCPs, indicating some variation in patients' tracking strategies.

> *I might not use it regularly, but when I start to feel that something is off, I immediately start using it, and then I can see, Yes, my asthma's not good now. I'll have to take more medication.* [Patient #2, Female, 58 years old]

Discussion

Principal Findings

We found that the use of a digital self-management system for asthma was appreciated by patients and HCPs, as it supported empowerment and proved beneficial for both patients and HCPs. The system helped patients gain better knowledge about their chronic condition and provided them with tools and skills to manage the disease. The data generated by the system enabled HCPs to prioritize patients with greater needs for follow-up, thereby supporting a more efficient allocation of health care resources. By empowering patients who are willing and able to take greater responsibility for their own care, the system also allowed HCPs to dedicate more time to those requiring direct personal contact. However, self-monitoring is not without challenges. It demands additional effort and management from both patients and HCPs [38,39], highlighting the need to adapt existing health care delivery models to accommodate these new responsibilities.

The main theme, data-supported empowerment, was facilitated by patient-generated health data. This was particularly evident in 2 of the 3 subthemes: empowerment by awareness, knowledge, and learning, and contact health care-patient. The system helped patients learn about factors affecting their asthma symptoms, making them better equipped to manage their condition and proactively adjust their medications to prevent exacerbations. These findings are supported by previous research that has shown that digital technology can support patients in better understanding their symptoms and condition [30,40-43].

Consistent with earlier studies, the patient-generated health data provided by the system functioned as decision support for HCPs, helped patients become more informed, supported self-care, and contributed to shared decision-making [44,45]. These experiences align with Bandura's definition of self-efficacy as the belief in one's ability to organize and execute actions required to manage prospective situations [37].

XSL•FO
RenderX

Including both patients and HCPs in the study enabled the identification of differences in their experiences of using digital self-management tools. The subtheme Empowerment by awareness, knowledge, and learning was emphasized more by patients than by HCPs. However, HCPs appreciated the patients' improved empowerment and self-efficacy and believed that this would positively affect the patients' asthma control. Physicians have expressed similar views on patient-generated health data and have advocated for digital solutions to improve care [46].

Although patients can independently use the system, it may also create expectations that health care services cannot always meet. Some patients expected feedback from HCPs based on their reported data and were disappointed when this did not occur. However, some HCPs did proactively contact patients when they noticed a deterioration in asthma control. Personalized feedback is a key factor in achieving optimal outcomes and quality of care [41,47,48], but both patients and HCPs must have clear expectations and understand their respective responsibilities [49]. To ensure equity in care, it is also essential to establish routines for HCPs so that monitoring is streamlined and not dependent on the individual HCPs' engagement.

In contrast to the findings from self-tracking in Parkinson disease, where patients found tracking burdensome and were unsure what to monitor [50], our study showed that patients were guided by the system and adapted their monitoring to periods when they experienced a greater need. Thus, the patients were not bound to monitor at certain times, but when they were motivated. This ad hoc pattern, rather than a continuous approach to self-monitoring, adopts a symptom-triggered, pragmatic strategy shaped by perceived relevance, burden, and confidence in their self-awareness. While this adaptive use may support sustainable engagement, it can also limit the early detection of gradual changes in asthma control. This needs to be considered in both the design of the self-management system, including in the development of algorithms to give treatment advice, and in the guidance to patients. The latter may include instructing patients to repeat measurements when they feel well on a regular basis, even if not measuring continuously, to ensure that the system does not interpret poor results as a new normal if patients only measure when experiencing worse symptoms. In illnesses where symptoms are not easily acknowledged, for example, diabetes or hypertension, the incentives to track may be experienced differently.

Certain findings in our study are specific to asthma, such as the importance of monitoring lung function and adjusting medication in response to symptoms. The system's ability to support these tasks was central to participants' positive experiences. At the same time, several themes, such as the need for tailored support, usability, and integration into daily routines, are likely relevant to other chronic conditions. By distinguishing asthma-specific needs from general principles of digital self-management, the study contributes both to condition-specific knowledge and broader digital health research.

This study shows that digital self-management systems, such as the one described in our study, may contribute to a novel way of delivering health care for individuals with chronic conditions, but it also highlights difficulties that must be addressed. Our findings support the potential of digital tools to facilitate self-management but also highlight the importance of considering the broader context in which such management occurs. As May et al [51] argue, self-management can impose a *burden of treatment*, particularly when health care systems shift responsibilities to patients without adequate support. This underscores the need to view self-management not only as an individual behavior but as a socially situated practice shaped by interactions with HCPs, access to resources, and systemic structures [52].

Well-functioning self-management systems have the potential to optimize care, save time for HCPs, and deliver more cost-effective health care. Further research is needed to determine whether digital self-management systems improve patient outcomes. The ongoing randomized controlled trial will provide additional insights into this. Future studies should also focus on establishing the factors necessary to implement and integrate self-management for patients, and its use by HCPs in primary care in order to maximize its gain.

The findings of this study suggest several implications for the design of digital self-management systems. For instance, patients valued features that provided immediate feedback on their asthma status, which enhanced their sense of control. However, some struggled with interpreting data, indicating a need for clearer visualizations and integrated guidance. Nurses emphasized the importance of seamless integration with existing systems to avoid the duplication of work. These insights point to the need for user-centered design that accommodates varying levels of digital literacy and supports both autonomous use and professional collaboration.

Strengths and Limitations

In line with qualitative research traditions, particularly those using inductive content analysis, we formulated a broad aim rather than predefined research questions. This approach allows for openness to participants' perspectives and supports the exploratory nature of the study, where themes and insights emerge from the data rather than being constrained by specific hypotheses [36]. The use of semistructured interviews also facilitated an in-depth exploration of participants' experiences. Multiple researchers independently coded the data, enhancing credibility and reliability. Discrepancies were resolved through thorough discussions within the research team.

The data for this study were collected in 2021 to 2023. While digital self-management technologies continue to evolve rapidly, the key challenges identified by patients and HCPs, such as interoperability problems, varying levels of digital literacy, and their impact on attitude and motivation in using digital tools, remain highly relevant today. Nevertheless, there is a need for ongoing research to explore how digital health technologies may influence patients' and HCPs' experiences, engagement, and interactions with these tools.

A limitation of the study is the predominance of older patients (average age 65 years) and female patients. Age is a known barrier to adopting digital technology. However, participants

XSL•FO
RenderX

were generally satisfied with the system's usability, and younger users were unlikely to face additional challenges.

The interview guide was not pilot tested; however, it was carefully designed (by LJH, HH, MH), leveraging collective expertise and wide experience in the field. While it was neither formally validated nor pilot-tested, we ensured its robustness through several internal reviews and revisions. To facilitate participation among both patients and HCPs, interviews were conducted either online (via Zoom) or by telephone, based on the respondents' preference. While this approach enabled flexibility and accessibility, it may have limited the ability to observe nonverbal cues and establish rapport in the same way as face-to-face interviews.

We consider the number of interviews with patients to have been sufficient to achieve saturation. That only 6 specialist asthma nurses participated is a limitation; however, this reflects the total number of eligible nurses at the participating units.

The relatively low number of patients who volunteered to be interviewed potentially creates a selection bias where those who did not respond to the invitation might have had a more negative perception of the self-management system. There is also a selection bias concerning which patients initially chose to try the self-management system; however, this would likely be replicated in a real implementation, making our results relevant to future users of the digital self-management system. Yet, further studies involving a broader and more diverse group of patients, including those who chose not to engage with the self-management system, are needed to better understand barriers to adoption and to capture a wider range of experiences.

While the qualitative nature of the study does not guarantee the generalizability of the results, we argue that they are transferable to similar patient groups and contexts. Still, we cannot exclude that some patient groups cannot take advantage of the self-management system. To counteract that, HCPs must be observant and identify those and provide physical contact when needed.

Conclusions

This study provides novel insights into the real-world use of a digital self-management system for asthma, drawing on the perspectives of both patients and specialist nurses after extended use. A key contribution of this study is the identification of how digital self-management systems can both empower users and simultaneously shift responsibilities in ways that require careful negotiation between patients and HCPs. These findings underscore the importance of clear communication about expectations, roles, and responsibilities when implementing digital tools in chronic care.

Moreover, the study emphasizes the need for systems to be adaptable to individual needs and seamlessly integrated with existing digital infrastructures to support sustainable use. Patients and HCPs must know how to interact with and what to expect from these systems for them to promote a better way of delivering health care for people with chronic conditions. Clear communication about changes in workflow and responsibilities is essential to ensure successful implementation and improved health care delivery.

By focusing on asthma, a condition with specific self-monitoring and treatment demands, this study adds to the understanding of how digital tools can be tailored to support condition-specific needs while also offering transferable lessons for the broader field of chronic disease management.

Acknowledgments

We want to thank Björn Nordlund and colleagues at Medituner who were responsible for the randomized controlled quantitative study.

We also like to thank the primary care of Tiohundra in Norrtälje.

Funding

This work was supported by AFA Försäkring (190210).

Data Availability

The datasets generated and analyzed during the current study are not publicly available since the qualitative data collected cannot be shared without risking the study participants' privacy. For inquiries regarding the dataset, contact the corresponding author.

Authors' Contributions

MH, HH, and LJH were responsible for designing the study. MH, ND, and SR collected the data. LJH, MH, ND, and SR performed the initial analysis of the data. LJH, MH, ND, SR, GP, and HH were responsible for the final interpretation of the data. All authors contributed to writing, reading, and approving the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Interview guide health care professionals.
[DOCX File, 16 KB - humanfactors_v13i1e79866_app1.docx]

Multimedia Appendix 2
Interview guide patients.
[DOCX File, 17 KB - humanfactors_v13i1e79866_app2.docx]

Checklist 1
COREQ checklist.
[PDF File, 574 KB - humanfactors_v13i1e79866_app3.pdf]

References

1. Maun A, Björkelund C, Arvidsson E. Primary care utilisation, adherence to guideline-based pharmacotherapy and continuity of care in primary care patients with chronic diseases and multimorbidity—a cross-sectional study. BMC Prim Care 2023 Nov 13;24(1):237. [doi: 10.1186/s12875-023-02191-6] [Medline: 37957554]
2. Innovative care for chronic conditions: building blocks for actions. WHO. URL: https://iris.who.int/server/api/core/bitstreams/e3ce2c65-c0d8-496c-ae93-4709403fe79a/content [accessed 2025-10-22]
3. Wallace E, Salisbury C, Guthrie B, Lewis C, Fahey T, Smith SM. Managing patients with multimorbidity in primary care. BMJ 2015 Jan 20;350:h176. [doi: 10.1136/bmj.h176] [Medline: 25646760]
4. Glasgow RE, Huebschmann AG, Krist AH, Degruy FV. An adaptive, contextual, technology-aided support (ACTS) system for chronic illness self-management. Milbank Q 2019 Sep;97(3):669-691. [doi: 10.1111/1468-0009.12412] [Medline: 31424137]
5. Small N, Bower P, Chew-Graham CA, Whalley D, Protheroe J. Patient empowerment in long-term conditions: development and preliminary testing of a new measure. BMC Health Serv Res 2013 Jul 8;13(1):263. [doi: 10.1186/1472-6963-13-263] [Medline: 23835131]
6. Hudon C, Fortin M, Haggerty JL, Lambert M, Poitras ME. Measuring patients' perceptions of patient-centered care: a systematic review of tools for family medicine. Ann Fam Med 2011;9(2):155-164. [doi: 10.1370/afm.1226] [Medline: 21403143]
7. Håkansson Eklund J, Holmström IK, Kumlin T, et al. "Same same or different?" A review of reviews of person-centered and patient-centered care. Patient Educ Couns 2019 Jan;102(1):3-11. [doi: 10.1016/j.pec.2018.08.029] [Medline: 30201221]
8. Ahmed A, van den Muijsenbergh M, Vrijhoef HJM. Person-centred care in primary care: what works for whom, how and in what circumstances? Health Soc Care Community 2022 Nov;30(6):e3328-e3341. [doi: 10.1111/hsc.13913] [Medline: 35862510]
9. WHO guidelines on hand hygiene in health care. WHO. URL: https://iris.who.int/server/api/core/bitstreams/b7cdc469-d662-4958-adfd-949a750e5ad9/content [accessed 2025-10-22]
10. Wilde MH, Garvin S. A concept analysis of self-monitoring. J Adv Nurs 2007 Feb;57(3):339-350. [doi: 10.1111/j.1365-2648.2006.04089.x] [Medline: 17233653]
11. Greenhalgh T. Patient and public involvement in chronic illness: beyond the expert patient. BMJ 2009 Feb 17;338:b49. [doi: 10.1136/bmj.b49] [Medline: 19223339]
12. Bury M. Chronic illness, self-management and the rhetoric of empowerment. In: Scambler G, Scambler S, editors. New Directions in the Sociology of Chronic and Disabling Conditions: Assaults on the Lifeworld: Springer; 2010:161-179. [doi: 10.1057/9780230297432_9]
13. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]
14. Huygens MWJ, Swinkels ICS, de Jong JD, et al. Self-monitoring of health data by patients with a chronic disease: does disease controllability matter? BMC Fam Pract 2017 Mar 20;18(1):40. [doi: 10.1186/s12875-017-0615-3] [Medline: 28320330]
15. Lorig KR, Sobel DS, Ritter PL, Laurent D, Hobbs M. Effect of a self-management program on patients with chronic disease. Eff Clin Pract 2001;4(6):256-262. [Medline: 11769298]
16. Ory MG, Ahn S, Jiang L, et al. Successes of a national study of the chronic disease self-management program: meeting the triple aim of health care reform. Med Care 2013 Nov;51(11):992-998. [doi: 10.1097/MLR.0b013e3182a95dd1] [Medline: 24113813]
17. Hibbard JH, Greene J, Overton V. Patients with lower activation associated with higher costs; delivery systems should know their patients' "scores". Health Aff (Millwood) 2013 Feb;32(2):216-222. [doi: 10.1377/hlthaff.2012.1064] [Medline: 23381513]
18. Hibbard JH, Mahoney ER, Stock R, Tusler M. Do increases in patient activation result in improved self-management behaviors? Health Serv Res 2007 Aug;42(4):1443-1463. [doi: 10.1111/j.1475-6773.2006.00669.x] [Medline: 17610432]

XSL•FO
RenderX

19. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS ONE 2016;11(5):e0156164. [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
20. Morton K, Dennison L, May C, et al. Using digital interventions for self-management of chronic physical health conditions: a meta-ethnography review of published studies. Patient Educ Couns 2017 Apr;100(4):616-635. [doi: 10.1016/j.pec.2016.10.019] [Medline: 28029572]
21. Bonoto BC, de Araújo VE, Godói IP, et al. Efficacy of mobile apps to support the care of patients with diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2017 Mar 1;5(3):e4. [doi: 10.2196/mhealth.6309] [Medline: 28249834]
22. Borna E, Nwaru BI, Bjerg A, et al. Changes in the prevalence of asthma and respiratory symptoms in western Sweden between 2008 and 2016. Allergy 2019 Sep;74(9):1703-1715. [doi: 10.1111/all.13840] [Medline: 31021427]
23. Astmaepidemiologi – en överblick. Läkartidningen. 2025. URL: https://lakartidningen.se/vetenskap/astmaepidemiologi-en-overblick/ [accessed 2025-11-04]
24. Papapostolou G, Tunsäter A, Binnmyr J, Telg G, Roslind K. Patient perspectives on living with severe asthma in Denmark and Sweden. Eur Clin Respir J 2021 Jan 1;8(1):1856024. [doi: 10.1080/20018525.2020.1856024]
25. Kompetensbeskrivning för sjuksköterskor inom astma, allergi och kol. Svensk sjuksköterskeförening. 2024. URL: https://swenurse.se/publikationer/kompetensbeskrivning-for-sjukskoterskor-inom-astma-allergi-och-kol [accessed 2025-11-09]
26. Mosnaim G, Safioti G, Brown R, et al. Digital health technology in asthma: a comprehensive scoping review. J Allergy Clin Immunol Pract 2021 Jun;9(6):2377-2398. [doi: 10.1016/j.jaip.2021.02.028] [Medline: 33652136]
27. McLean G, Murray E, Band R, et al. Interactive digital interventions to promote self-management in adults with asthma: systematic review and meta-analysis. BMC Pulm Med 2016 May 23;16(1):83. [doi: 10.1186/s12890-016-0248-7] [Medline: 27215329]
28. Morrison D, Wyke S, Agur K, et al. Digital asthma self-management interventions: a systematic review. J Med Internet Res 2014 Feb 18;16(2):e51. [doi: 10.2196/jmir.2814] [Medline: 24550161]
29. Frey U, Brodbeck T, Majumdar A, et al. Risk of severe asthma episodes predicted from fluctuation analysis of airway function. Nature 2005 Dec 1;438(7068):667-670. [doi: 10.1038/nature04176] [Medline: 16319891]
30. Merchant R, Inamdar R, Henderson K, et al. Digital health intervention for asthma: patient-reported value and usability. JMIR Mhealth Uhealth 2018 Jun 4;6(6):e133. [doi: 10.2196/mhealth.7362] [Medline: 29866644]
31. Gibson PG, Powell H, Coughlan J, et al. Self-management education and regular practitioner review for adults with asthma. Cochrane Database Syst Rev 2003(1):CD001117. [doi: 10.1002/14651858.CD001117] [Medline: 12535399]
32. Ljungberg H, Carleborg A, Gerber H, et al. Clinical effect on uncontrolled asthma using a novel digital automated self-management solution: a physician-blinded randomised controlled crossover trial. Eur Respir J 2019 Nov;54(5):1900983. [doi: 10.1183/13993003.00983-2019] [Medline: 31481605]
33. ISO 9241-210:2019 ergonomics of human-system interaction — part 210: human-centred design for interactive systems. : International Organization for Standardization (ISO); 2019 URL: https://www.iso.org/standard/77520.html [accessed 2025-11-04]
34. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
35. Kallio H, Pietilä AM, Johnson M, Kangasniemi M. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. J Adv Nurs 2016 Dec;72(12):2954-2965. [doi: 10.1111/jan.13031] [Medline: 27221824]
36. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
37. Bandura A. Self-Efficacy: The Exercise of Control: W. H. Freeman and Company; 1997. URL: https://www.scribd.com/document/394892404/Albert-Bandura-Self-Efficacy-The-Exercise-of-Control-W-H-Freeman-Co-1997-pdf [accessed 2026-02-28]
38. Simpson AJ, Honkoop PJ, Kennington E, et al. Perspectives of patients and healthcare professionals on mHealth for asthma self-management. Eur Respir J 2017 May;49(5):1601966. [doi: 10.1183/13993003.01966-2016] [Medline: 28461291]
39. Öberg U, Orre CJ, Isaksson U, Schimmer R, Larsson H, Hörnsten Å. Swedish primary healthcare nurses' perceptions of using digital eHealth services in support of patient self-management. Scand J Caring Sci 2018 Jun;32(2):961-970. [doi: 10.1111/scs.12534] [Medline: 28960451]
40. Li J, Varnfield M, Jayasena R, Celler B. Home telemonitoring for chronic disease management: perceptions of users and factors influencing adoption. Health Informatics J 2021;27(1):1460458221997893. [doi: 10.1177/1460458221997893] [Medline: 33685279]
41. Korpershoek YJG, Vervoort S, Trappenburg JCA, Schuurmans MJ. Perceptions of patients with chronic obstructive pulmonary disease and their health care providers towards using mHealth for self-management of exacerbations: a qualitative study. BMC Health Serv Res 2018 Oct 4;18(1):757. [doi: 10.1186/s12913-018-3545-4] [Medline: 30286761]

XSL•FO
RenderX

42. Jaana M, Sherrard H, Paré G. A prospective evaluation of telemonitoring use by seniors with chronic heart failure: adoption, self-care, and empowerment. Health Informatics J 2019 Dec;25(4):1800-1814. [doi: 10.1177/1460458218799458] [Medline: 30247080]
43. Thomas NA, Drewry A, Racine Passmore S, Assad N, Hoppe KK. Patient perceptions, opinions and satisfaction of telehealth with remote blood pressure monitoring postpartum. BMC Pregnancy Childbirth 2021 Feb 19;21(1):153. [doi: 10.1186/s12884-021-03632-9] [Medline: 33607957]
44. Patoz MC, Hidalgo-Mazzei D, Blanc O, et al. Patient and physician perspectives of a smartphone application for depression: a qualitative study. BMC Psychiatry 2021 Jan 29;21(1):65. [doi: 10.1186/s12888-021-03064-x] [Medline: 33514333]
45. Knitza J, Simon D, Lambrecht A, et al. Mobile health usage, preferences, barriers, and eHealth literacy in rheumatology: patient survey study. JMIR Mhealth Uhealth 2020 Aug 12;8(8):e19661. [doi: 10.2196/19661] [Medline: 32678796]
46. Vallo Hult H, Hansson A, Svensson L, Gellerstedt M. Flipped healthcare for better or worse. Health Informatics J 2019 Sep;25(3):587-597. [doi: 10.1177/1460458219833099] [Medline: 30887867]
47. Jacob C, Sezgin E, Sanchez-Vazquez A, Ivory C. Sociotechnical factors affecting patients' adoption of mobile health tools: systematic literature review and narrative synthesis. JMIR Mhealth Uhealth 2022 May 5;10(5):e36284. [doi: 10.2196/36284] [Medline: 35318189]
48. Blonigen D, Harris-Olenak B, Kuhn E, Humphreys K, Timko C, Dulin P. From "Step Away" to "Stand Down": tailoring a smartphone app for self-management of hazardous drinking for veterans. JMIR Mhealth Uhealth 2020 Feb 13;8(2):e16062. [doi: 10.2196/16062] [Medline: 32053118]
49. Wannheden C, Roczniewska M, Hasson H, Karlgren K, von Thiele Schwarz U. Better self-care through co-care? A latent profile analysis of primary care patients' experiences of e-health-supported chronic care management. Front Public Health 2022;10:960383. [doi: 10.3389/fpubh.2022.960383] [Medline: 36211687]
50. Riggare S, Scott Duncan T, Hvitfeldt H, Hägglund M. "You have to know why you're doing this": a mixed methods study of the benefits and burdens of self-tracking in Parkinson's disease. BMC Med Inform Decis Mak 2019 Aug 30;19(1):175. [doi: 10.1186/s12911-019-0896-7] [Medline: 31470832]
51. May CR, Eton DT, Boehmer K, et al. Rethinking the patient: using Burden of Treatment Theory to understand the changing dynamics of illness. BMC Health Serv Res 2014 Jun 26;14(1):281. [doi: 10.1186/1472-6963-14-281] [Medline: 24969758]
52. Franklin M, Willis K, Lewis S, Smith L. Chronic condition self-management is a social practice. J Sociol (Melb) 2023 Mar;59(1):215-231. [doi: 10.1177/14407833211038059]

Abbreviations

COREQ: Criteria for Reporting Qualitative Research
HCP: health care professional

Edited by M Waite; submitted 30.Jun.2025; peer-reviewed by J Eriksen, J Viitanen; revised version received 23.Nov.2025; accepted 19.Feb.2026; published 20.Mar.2026.

Please cite as:
Jäderlund Hagstedt L, Hvitfeldt H, Riggare S, Petersson G, Davoody N, Hägglund M
Patients' and Health Care Professionals' Experiences of a Digital Self-Management System for Asthma: Qualitative Study
JMIR Hum Factors 2026;13:e79866
URL: https://humanfactors.jmir.org/2026/1/e79866
doi:10.2196/79866

© Lovisa Jäderlund Hagstedt, Helena Hvitfeldt, Sara Riggare, Göran Petersson, Nadia Davoody, Maria Hägglund. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 20.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Using the Theoretical Domains Framework to Identify Barriers to and Enablers of Patient Telemedicine Services Use in China: Qualitative Study

Ke Liu[1,2*], MPA; Yuting Yang[3*], MS; Zixuan Song[1,2], MS; Huixian Li[1,2], MS; Yanli Lyu[1,2], MS; Ke Zhang[2,4], MD; Xinxia Wu[1,2], MM; Zheng Hou[2,5], MD; Yipei Wang[3], MBA, PhD

[1]Department of Medical Affairs, Peking University Third Hospital, Beijing, China

[2]Office of Internet Hospital, Peking University Third Hospital, Beijing, China

[3]Institute of Hospital Management, Peking University Third Hospital, North Huayuan Road 49, Beijing, China

[4]Department of Otolaryngology, Peking University Third Hospital, Beijing, China

[5]Department of Gynecology and Obstetrics, Peking University Third Hospital, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Yipei Wang, MBA, PhD
Institute of Hospital Management, Peking University Third Hospital, North Huayuan Road 49, Beijing, China

Abstract

Background: Telemedicine has rapidly expanded worldwide due to its convenience and accessibility. In China, an increasing number of hospitals have begun offering telemedicine services; however, patient utilization remains relatively low. Limited research has examined patients' behaviors during the process of adopting telemedicine services.

Objective: This study aimed to identify barriers to and enablers for patients using telemedicine services and to formulate implementation strategies.

Methods: We conducted semistructured qualitative interviews based on the Theoretical Domains Framework (TDF) to identify barriers and enablers to telemedicine utilization. Twenty-one patients who had used Peking University Third Hospital's telemedicine services were included in the interviews. Data were analyzed using NVivo 12.0 with deductive thematic analysis guided by the TDF. Moreover, a group of experts was assembled to devise potential intervention strategies.

Results: A total of 28 themes were identified, including 14 barriers and 14 enablers across 5 of the 14 TDF domains. The most frequently reported barriers were operational challenges, prolonged waiting periods from asynchronous communication, and doubts about therapeutic efficacy, whereas the most frequently mentioned enablers were the convenience of telemedicine, time conservation, and support from hospitals. On the basis of these factors, we devised 6 intervention strategies.

Conclusions: This study demonstrated that patients' utilization of telemedicine services was affected by several barriers and enablers, including system architecture and design, patient interactions using telemedicine, and external assistance. To enhance the utilization, these factors must be meticulously considered. This study also suggests strategies to enhance the utilization of telemedicine.

(JMIR Hum Factors 2026;13:e78457) doi:10.2196/78457

KEYWORDS

telemedicine services; Theoretical Domains Framework; barriers; enablers; patients; qualitative study

Introduction

With the rapid advancement of information technology, telemedicine services have expanded globally [1,2]. In China, the provision of telemedicine services by medical institutions has increased significantly since the COVID-19 pandemic [3]. Telemedicine services have expanded patients' health care options [4]. Telemedicine is thought to enhance the efficiency of health care services [5,6], decrease medical costs [7,8], improve patient satisfaction [9], and facilitate the equitable distribution of medical resources worldwide [10,11]. Despite the growing number of institutions providing telemedicine, patient utilization remains low. As of 2023, more than 3000 institutions had provided telemedicine services, delivering nearly 100 million virtual visits. Despite the seemingly large service volume, telemedicine accounts for only 1% of all visits [12], a proportion further validated by patient interviews [13]. Conversely, almost 20% of medical consultations in the United States were performed via telemedicine in 2020 [14]. The establishment of telemedicine services without corresponding

XSL•FO
RenderX

utilization has resulted in a significant waste of financial and human resources.

To enhance telemedicine utilization, it is crucial to identify and understand the factors influencing patient engagement. Previous studies have identified demographic factors, such as age [15], co-residence with children [16], and education [17], but these factors are difficult to change. Consequently, researchers turned to behavioral frameworks, such as the unified theory of acceptance and use of technology (UTAUT) [18], the technology acceptance model (TAM) [19], and the theory of planned behavior [20]. These frameworks have helped identify factors, such as perceived usefulness [21], ease of use [22], and social influence [23], as critical precursors to user intention. However, these models primarily focus on reflective processes, namely, how cognitive evaluations lead to the intention to use. Therefore, these classical models offer less granularity regarding the contextual barriers, habitual behaviors, and skill gaps that prevent intention from translating into actual usage in real-world health care environments.

The Theoretical Domains Framework (TDF) addresses this gap not by discarding prior theories but by synthesizing them into a more granular, intervention-oriented structure rooted in the capability, opportunity, and motivation system [24]. Thus, adopting the TDF allows for a true integration of classical constructs: for instance, the "perceived ease of use" in TAM is no longer just a perception but is further defined in operational context as "capability" (ie, the knowledge and skills domains); "social influence" in the UTAUT and the theory of planned behavior is recontextualized within "opportunity" (ie, the social influences domain); and "perceived usefulness" aligns well with "motivation" (ie, the beliefs about capabilities and beliefs about consequences domains) [25].

Therefore, the TDF effectively extends the theoretical boundary beyond the cognitive focus of TAM and UTAUT. By incorporating less-explored domains, such as environmental context and resources, the TDF captures the contextual drivers of behaviors that traditional models often overlook [26,27]. Its comprehensiveness enables our research to move beyond a simple prediction of adoption intention. Instead, it provides a diagnostic structure to identify relevant barriers and enablers affecting telemedicine usage [28,29]. Adopting the TDF thus situates our work within the existing literature, bridging the well-established cognitive constructs of TAM and UTAUT with the implementation-focused depth required for designing interventions aimed at improving patient utilization of telemedicine services in China.

Building on this theoretical foundation, this study targets patients who have used telemedicine services in China, aiming to examine their behaviors and psychological aspects during the telemedicine process. It used semistructured qualitative interviews based on the TDF to ascertain barriers and enablers to telemedicine adoption. Our research may offer actionable evidence to support the refinement of telemedicine platforms and the development of patient-centered service models in future practice.

Methods

Study Design

We conducted the survey in Beijing, a leading Chinese metropolis in the development of telemedicine. This study focused on tertiary public hospitals, which deliver the majority of telemedicine services in China and are highly trusted and sought after by patients [30,31]. Peking University Third Hospital, a tertiary public hospital, serves as the governing institution of the Beijing Telemedicine Quality Control Center [32]. It ranks among the top 10 hospitals in Beijing in terms of telemedicine user volume, with patients coming from dozens of provinces nationwide over recent years. The hospital provides both synchronous and asynchronous telemedicine consultations, covering the 2 predominant service models in China. Its extensive online services span the entire care process, including prediagnosis, diagnosis, and postdiagnosis.

Participants Recruitment

We used convenience sampling to select patients engaged in real-time video and asynchronous graphic consultations at Peking University Third Hospital. We accounted for patients' varied geographic locations, disease classifications, and socioeconomic statuses to ensure that the sample accurately represented the telemedicine-using patient community. The inclusion criteria were as follows: (1) at least one prior use of telemedicine at Peking University Third Hospital within the past 6 months, (2) willingness to participate in the interview, and (3) adequate communication skills to engage in the interview. Previous literature suggests that thematic saturation in qualitative interview studies is generally achieved before reaching 21 interviews [33]. Accordingly, the intended sample size was 20 to 22 participants in the preliminary phase. When the data collection process yielded no new information about the enablers and barriers influencing access to telemedicine services, data saturation was reached [34].

Interview Guide Development

We applied the TDF throughout the interview design, data collection, and analysis. The TDF guided the development of the interview guide, which included 1 to 3 questions for each of the 14 domains (eg, goals and intentions) [35]. For example, one of the questions was, "What do you know about telemedicine services?" Investigators with expertise in implementation science provided guidance to develop the interview guide. Moreover, we consulted with experts in hospital management, health policy management, and physicians to refine the outline. We conducted 2 preexperiments after the outline was initially finalized. The interview outline was refined slightly after the pretests, and the samples from the pretests were not included in the study due to slight changes. Table 1 presents the TDF domains and corresponding interview questions.

XSL•FO
RenderX

Table . Interview questions aligned with the Theoretical Domains Framework (TDF) domains.

TDF domains	Definition [26]	Definition in this study	Interview questions
Goals	Mental representations of outcomes or end states that an individual wants to achieve	Goals that patients hope to resolve using telemedicine	• What problems can telemedicine help you solve?
Intention	A conscious decision to perform a behavior or resolve to act in a certain way	Patients' willingness or plan to use telemedicine	• How did you know about telemedicine, and why did you try it? • How do you feel about it compared to the offline service?
Knowledge	An awareness of the existence of something	The patient's understanding of telemedicine	• What do you know about telemedicine services?
Skills	An ability or proficiency acquired through practice	The patient's practical ability to operate telemedicine independently	• Are you familiar with using telemedicine services? • Can you independently use telemedicine for online consultations?
Beliefs about capabilities	Acceptance of the truth, reality, or validity about an ability, talent, or facility that a person can put to constructive use	Patients' confidence in their ability to use telemedicine successfully	• To what extent do you feel capable of using telemedicine? • When you use telemedicine, is there someone to help you? If so, who and how?
Optimism	The confidence that things will happen for the best or that desired goals will be attained	Patients' positive expectations about the benefits of telemedicine	• Do you think telemedicine can benefit patients? • Is it beneficial to you? Please provide some examples.
Memory, attention, and decision processes	The ability to retain information, focus selectively on aspects of the environment, and choose between 2 or more alternatives	Cognitive aspects influencing telemedicine patients' memory, concentration, and decision-making	• Are there aspects that complicate the utilization of telemedicine services? Have you faced any particular challenges?
Emotions	A complex reaction pattern, involving experiential, behavioral, and physiological elements, by which the individual attempts to deal with a personally significant matter or event	Patients' emotional experiences when engaging with telemedicine	• How do you feel when you use telemedicine services? • How do you feel about telemedicine services compared to offline service?
Beliefs about consequences	Acceptance of the truth, reality, or validity of outcomes of a behavior in a given situation	Patients' perceptions of the benefits and risks of telemedicine	• What are the enablers of telemedicine services? • What are the barriers to telemedicine services?
Reinforcement	Increasing the probability of a response by arranging a dependent relationship, or contingency, between the response and a given stimulus	Incentives or feedback that motivate patients to use telemedicine	• What incentives can motivate you to use telemedicine services?
Social or professional role and identity	A coherent set of behaviors and displayed personal qualities of an individual in a social or work setting	Patients' perception of their role in health management and its alignment with telemedicine	• Have people around you (family members, friends, etc) used telemedicine services? • What attitudes do they hold about telemedicine services? • Does their perspective affect your decision regarding the visit?
Social influences	Those interpersonal processes that can cause individuals to change their thoughts, feelings, or behaviors	Impact of family, peers, and social norms on telemedicine use	• Do social environment and policies affect your use of telemedicine services? How?

XSL•FO
RenderX

TDF domains	Definition [26]	Definition in this study	Interview questions
Environmental context and resources	Any circumstance of a person's situation or environment that discourages or encourages the development of skills and abilities, independence, social competence, and adaptive behavior	External conditions and resources that facilitate or hinder telemedicine use	• Do you believe that the social environment and policies will influence your utilization of telemedicine services? What is its impact? • Does the availability of resources and support affect your use of telemedicine? How?
Behavioral regulation	Anything aimed at managing or changing objectively observed or measured actions	Patients' strategies to ensure effective use of telemedicine	• What changes are necessary for telemedicine services to succeed? • Do you have any other suggestions to enhance telemedicine usage?

Interview and Data Collection

The formal interview was performed via telephone and approved in January 2024. We secured the participants' informed consent and subsequently audio recorded the interviews using the software's exclusive features. Each interview lasted approximately 10 to 20 minutes. The interviews were conducted in Mandarin Chinese by the authors KL and YY, both of whom had professional training in qualitative interviewing and extensive experience in qualitative research (n=2, one male and one female, health management practitioners). Before each interview, the interviewer provided a brief self-introduction and overview of the interview content to establish rapport with the participant. Only 2 of the interviewers were present during the interview to create a conducive environment for communication. The recordings were transcribed into textual materials within 48 hours after the interviews, and the researcher recorded and sorted the data to ensure consistency. After each interview, participants were sent a prearranged text message, and their data were collected.

Data Coding and Analysis

Interview transcripts were transcribed, organized, and imported into the software NVivo (version 12.0; QSR International Pty Ltd) for coding. We conducted a deductive thematic analysis guided by the TDF and followed 3 steps [36,37]. First, the 14 TDF domains served as an a priori coding framework. Second, the transcripts were reviewed in detail, with relevant content identified and coded into themes. Third, these themes were categorized as either barriers or enablers within the corresponding TDF domains. When a theme was relevant to more than 1 TDF domain, it was assigned to the domain judged to be more closely aligned. Two authors (KL and YY) independently coded the transcripts. Coding discrepancies were resolved through discussion with a third author (YW) until consensus was reached. The coded κ for each level was greater than 0.90. To prepare the results, we tabulated the themes within each TDF domain and tabulated more prominent themes into barriers and enablers separately. A detailed description of the coding process, including example excerpts and coding framework, was provided in Multimedia Appendix 1.

Strategy Development

Following encoding and analysis, we assembled a multidisciplinary panel of hospital experts, including specialists in information technology, clinical practice, and hospital administration leadership. This collaborative endeavor integrated technical, clinical, and operational viewpoints via monthly interdepartmental meetings and methodically structured brainstorming sessions employing implementation science frameworks. Using the TDF model, the panel systematically formulated various intervention options aimed at addressing the identified barriers and enablers.

Ethical Considerations

This study was approved by the Medical Science Research Ethics Committee of Peking University Third Hospital (IRB00006761-M2024029). Participants provided informed consent. We stored the survey data, which were strictly managed to remain confidential. We offered a compensation of 100 yuan (US $13.8) as a token of appreciation to all participants for their involvement in the study.

This study followed established reporting standards for qualitative research and was reported in accordance with the COREQ checklist (Checklist 1).

Results

Participants and Institutions

All interviews were conducted from January to March 2024, covering a duration of 3 months. Twenty-five telephone interviews were conducted, of which 21 were considered legitimate for analysis. Of the 21 interviewees, 13 (62%) were women, and 19 (91%) held a bachelor's degree or higher. The majority (18/21, 85%) of the participants had full-time jobs. The mean interview duration was 13.7 (SD 5.7) minutes. All participants had used telemedicine services at least once, with a mean of 3.46 (SD 2.8) visits as specified in Table 2.

XSL•FO
RenderX

Table . Interview patient characteristics (N=21).

Characteristics	Participants, n (%)
Sex	
Male	8 (38)
Female	13 (62)
Age (y)	
<30	4 (19)
30 - 35	5 (24)
36 - 40	7 (33)
>40	5 (24)
Education	
Senior high school or below	2 (9)
College	10 (48)
Postgraduate degree or above	9 (43)
Employment status	
Full-time employment	18 (85)
Unemployed	1 (5)
Retired	1 (5)
Student	1 (5)
Home location	
Beijing	13 (62)
Outside of Beijing	8 (38)
Having a chronic disease	
Yes	3 (15)
No	18 (85)
Undergone surgery	
Yes	3 (15)
No	18 (85)
Medical insurance type	
Employee insurance	14 (67)
Publicly funded free medical care	6 (28)
Urban resident insurance	1 (5)
Number of uses	
1	8 (38)
2 - 5	8 (38)
6 - 10	5 (24)

Barriers to and Enablers of Patient Use of Telemedicine Services

A total of 190 principal assertions were documented, covering 5 domains and 28 themes, including 14 barriers and 14 enablers. Tables 3 and 4 display the themes within each domain and their frequencies.

XSL•FO
RenderX

Table . Barriers to use of telemedicine services (n=71).

TDF[a] domains and themes	Frequency, n (%)
Environmental context and resources (n=33)	
The interaction of the app is poor	12 (17)
The functions of the telemedicine services are limited	8 (11)
The process of telemedicine services and treatment is not smooth enough	5 (7)
The operation is not intelligent enough and is rather cumbersome	4 (6)
The propagandizing is insufficient	3 (4)
The environment is too noisy	1 (1)
Memory, attention, and decisions process (n=23)	
Patients have to wait passively for a long time by asynchronous communication	10 (14)
The operation is too complicated for the patients	10 (14)
It is hard to find the functions that suit oneself	3 (4)
Beliefs about consequences (n=8)	
Doubts and distrust about the effectiveness of telemedicine services	7 (10)
The risk of privacy leakage and security issues	1 (1)
Skills (n=6)	
Mobile phone operation skill is poor	4 (6)
It is difficult to collect information online	2 (3)
Beliefs about capabilities (n=1)	
Patients have no confidence in telemedicine services	1 (1)

[a]TDF: Theoretical Domains Framework.

XSL•FO
RenderX

Table . Enablers to use of telemedicine services (n=119).

TDF[a] domains and themes	Frequency, n (%)
Beliefs about consequences (n=67)	
The convenience of telemedicine services	36 (30)
Telemedicine services can save time	16 (13)
Telemedicine services can enhance appointment accessibility	10 (8)
Telemedicine services can reduce transportation costs	5 (4)
Environmental context and resources (n=37)	
The system is easy to operate	10 (8)
Hospitals or app can provide guidance	8 (7)
Stable and fast network	8 (7)
Provide prompt consultation services	4 (3)
There is a quiet and suitable environment	4 (3)
Has the function of prescribing medicine	3 (3)
Beliefs about capabilities (n=8)	
The patient has self-efficacy in digital health	5 (4)
The patient has successful experience in online diagnosis and treatment	3 (3)
Skills (n=7)	
The patient's mobile phone operation skills are relatively good	5 (4)
The patient has the ability to collect information	2 (2)

[a]TDF: Theoretical Domains Framework.

The barriers to theoretical domains were referenced a total of 71 times. Specifically, "environmental context and resources" was referenced 33 times; "memory, attention, and decision processes" was referenced 23 times; "beliefs about consequences" was referenced 8 times; "skills" was referenced 6 times; and "beliefs about capabilities" was referenced once. In contrast, the enablers of theoretical domains were referenced a total of 119 times. "Beliefs about consequences" was referenced most frequently, occurring 71 times"; "environmental context and resources" was referenced 37 times; "beliefs about capabilities" was referenced 8 times; and "skills" was referenced 7 times.

Barriers

Environmental Context and Resources

"Environmental context and resources" was the primary barrier. It primarily emphasized user-environment interaction and the functional categories of telemedicine services. On the one hand, user-environment interaction explicitly examined how the design attributes of app influence patient engagement. It encompassed interface usability and system intelligence. On the other hand, the limited variety of functions created barriers. If patients' medical needs were not fully met, they did not use telemedicine services.

> *The software design seems flawed, insufficient prompts and guidance mechanisms make navigation unnecessarily challenging.* [LQ, male, 2-time user]

> *The functions of online consultation are too limited. It would be great if they could be more diverse to meet more demands.* [KJ, female, 8-time user]

Memory, Attention, and Decision Process

The patients' memory, attention, and decision-making processes might cause barriers, including asynchronous communication and the complexity of the process. First, barriers arose from asynchronous communication between patients and physicians. Following patients' inquiries, they frequently endured prolonged delays before receiving a response from the physician, which significantly impacted patients' readiness to use telemedicine. Second, barriers emerged from the complexity of the process. The complexity of the process was referenced 8 times in the interview, exclusively by first-time users. Patients' attention was limited, but too many steps and jumps caused patients to resist the use of telemedicine services.

> *Sometimes we want to contact the physician proactively and leave a message there, but the physician can't see it. We can only wait for the physician to contact us. This is too passive.* [LJT, female, 2-time user]

> *I searched endlessly for the payment page but couldn't find it. The design isn't user-friendly.* [YY, female, 1-time user]

Beliefs About Consequences

"Beliefs about consequences" manifested as barriers from two primary concerns: skepticism about treatment effectiveness and fears of privacy infringements. Primarily, owing to inadequate

XSL•FO
RenderX

communication during telemedicine consultations, patients doubted the efficacy of telemedicine services. In addition, telemedicine necessitated the accumulation of extensive patient data, which led to concerns regarding the potential breach of personal privacy.

> *I believe that online medical consultations are not as reliable as in-person visits. In a physical setting, physicians can conduct thorough examinations, whereas online consultations rely solely on verbal descriptions.* [WWH, male, 1-time user]
>
> *I am genuinely concerned about the potential for my personal information or medical data to be leaked. Given the prevalence of telecom fraud, I prefer to go to hospital.* [YYX, male, 3-time user]

Skills

"Skills" emerged as a critical barrier domain, primarily manifested through deficits in digital health literacy and device operation proficiency. During the interviews, numerous patients indicated that telemedicine presented specific challenges for the older adults.

> *At our age, we're barely comfortable with basic smartphone functions like WeChat. Navigating these sophisticated medical platforms feels overwhelmingly complex.* [ST, female, 3-time user]

Enablers

Beliefs About Consequences

"Beliefs about consequences" was repeatedly highlighted as an enabler, reflecting patients' recognition of the benefits and potential outcomes of telemedicine. First, nearly all polled patients indicated that convenience facilitated their utilization of telemedicine services, with several patients reiterating this point consistently. Second, time conservation emerged as a key enabler, as telemedicine significantly reduced time expenditure, particularly for younger patients. Third, improved appointment accessibility was identified as a crucial enabler. Patients were able to schedule follow-up tests and obtain medications via telemedicine services without concerns regarding registration.

> *After we use it, we think it is really convenient, so we don't have to go to the hospital anymore, and it is not easy for us to go there.* [YY, female, 1-time user]
>
> *I think using this [telemedicine services] can communicate with the physician, the effect is similar, and it is much easier than offline, no need to ask for leave or go to the hospital.* [SST, female, 1-time user]
>
> *I've found departments are readily available on the telemedicine services platform. That's really helpful.* [TXM, female, 7-time user]

Environmental Context and Resources

"Environmental context and resources" was identified as an enabler influencing patients' use of telemedicine services. First, app usability emerged as an enabler, as patients typically perceived that simple operation enhanced their engagement with the telemedicine services. Second, support from the app or hospital proved essential, helping patients easily locate access to telemedicine services and substantially improving utilization. Finally, the comprehensiveness of platform functionalities was recognized as a crucial enabler, as an extensive range of telemedicine services facilitated patients' adoption and use.

> *Telemedicine services are straightforward. I like it.* [WZ, male, 5-time user]
>
> *The hotline service is particularly helpful. Whenever I encounter difficulties, their staff guides me through the process step-by-step.* [LJT, female, 2-time user]
>
> *We hope they'll keep expanding features so we don't have to make endless trips to the hospital anymore.* [PY, male, 2-time user]

Beliefs About Capabilities

"Beliefs about capabilities" evolved as enablers, reflected in two aspects: self-efficacy and prior successful experience. On the one hand, self-efficacy in digital health was largely associated with usage frequency, as expressed by confident users. On the other hand, patients' previous successful experiences influenced their utilization behaviors, as they drew upon earlier eHealth encounters.

> *I'm generally good at figuring out tech stuff. People around me always ask me for advice when they need online medical help.* [YY, female, 1-time user]
>
> *Having used other hospitals' internet services, I assume this APP should work fairly similarly.* [KJ, female, 8-time user]

Skills

"Skills" emerged as a crucial enabler in the use of telemedicine services. In particular, expertise in using mobile phones and information retrieval was frequently mentioned. Individuals with advanced mobile device skills were more inclined to use telemedicine services. In addition, proficiency in information-seeking behaviors was recognized as an enabler.

> *I'm quite comfortable using smartphones daily, so navigating these platforms seems easy.* [LLZ, male, 1-time user]
>
> *I independently researched all relevant information about telemedicine services. Honestly, I'm pretty good at finding what I need.* [YY, female, 1-time user]

Interventions to Improve Patient Utilization of Telemedicine Services

On the basis of these findings, we designed intervention measures addressing barriers and enablers through expert group discussions (Table 5). We assessed the importance of every domain by determining the frequency of citations throughout all transcripts. The domains "beliefs about consequences" and "environmental context and resources" were the most frequently cited, with 75 and 70 mentions, respectively, accounting for 40% and 37% of the total citations.

XSL•FO
RenderX

Table . Evidence-based interventions to improve telemedicine services utilization (n=190).

TDF[a] domains	Frequency, n (%)	Intervention	Implementation details
Beliefs about consequences	75 (39)	Enhance support for first-time users	• Pilot postdischarge coaching in orthopedics • Train family physicians to assist virtual visits
Environmental context and resources	70 (37)	Simplify app functions	• Suspend development of complex features • Optimize core functionalities • Streamline the patient's usage process
Memory, attention, and decision process	23 (12)	Enhance APP guidance and 24/7 support hotline	• Add real-time operation prompts • Enable automatic step progression • Develop an AI[b]-powered patient assistant • Provide step-by-step guidance • Establish rapid complaint resolution • Implement AI-assisted feedback tracking
Skills	13 (7)	Multichannel publicity	• Distribute instructional videos via WeChat or social media • Install interactive kiosks in clinics • Create a physician's promotional webpage
Beliefs about capabilities	9 (5)	Age-friendly adaptation	• Launch large-text interface • Create elderly-specific tutorials • Integrate voice command features

[a]TDF: Theoretical Domains Framework.

[b]AI: artificial intelligence.

In the "beliefs about consequences" domain, convenience emerged as a key factor encouraging patients to use telemedicine. Enhancing the experiences of first-time users was considered essential. It was advisable to offer supplementary services for patients using telemedicine for the first time, thereby enhancing their experience of its convenience.

Furthermore, in the "environmental context and resources" domain, the complexity of using the patient application represented the most prominent barrier factor. It was suggested to simplify the patient interface and limit the proliferation of complex features, hoping to facilitate patients' focus on critical functions to meet the needs of the majority.

Discussion

Principal Findings

This study conducts an in-depth exploration of the relatively underresearched area of telemedicine among Chinese patients. It is the first instance of using the TDF model for conducting interview-based surveys with patients regarding telemedicine, thereby enhancing the scientific rigor of the research. This study found that telemedicine services adoption is jointly shaped by 14 barriers and 14 enablers, including 5 of the 14 TDF domains. This study has found more behavioral and psychological influencing factors in patients compared to other studies. Barriers arose from operational complexity, concerns about privacy and efficacy, asynchronous communication, and so on. Enablers include high convenience and accessibility, simplified system functionalities, and external support. On the basis of identified barriers and enablers, we developed targeted intervention measures and prioritized them based on their relative significance as expressed by the interviewees. In conclusion, this study advances a comprehensive understanding of the determinants influencing this adoption and establishes a theoretical foundation for developing targeted interventions in telemedicine services.

Positive Feedback Mechanisms

Our exploratory analysis of usage frequency indicates that barriers and enablers change alongside patients' cumulative experiences with telemedicine services. First-time or infrequent users were more inclined to highlight challenges associated with accessing the platform and operating system, indicating the significant learning curve of initial usage. In contrast, frequent users recognized constraints in the scope and profundity of

XSL•FO
RenderX

accessible system functionalities while concurrently emphasizing convenience and efficiency as primary enablers. This transition suggests a possible positive feedback loop: as patients become accustomed to the system, operational obstacles lessen, and the perceived advantages of telemedicine become more prominent, hence encouraging further utilization.

This corresponds with previous research [13,38]. Similarly, in the TAM, it is believed that "perceived usefulness" influences behavioral intention [39]. Our research delineates a self-perpetuating feedback loop: utilization of telemedicine services leads to increased satisfaction and reinforced intention to use. This phenomenon highlights the importance of the first usage experience in telemedicine service adoption. Similarly, the anchoring effect in initial interactions profoundly influences long-term commitment. Facilitating smooth navigation and practical convenience during initial interactions might create positive behavioral anchors, promoting regular use of online telemedicine services. Previous research indicates that Michigan Medicine (Ann Arbor) has established a system that incorporates template instructions and anticipatory advice, which can enhance the patient utilization process [40]. Consequently, we advocate for the deployment of multimodal guiding systems, including artificial intelligence–driven triage and a human-assisted hotline.

Duality of Barriers and Enablers

Our study reveals that the utilization of telemedicine services is directionally modulated by coexisting enablers and barriers within the same theoretical domains. The 2 domains, "environmental context and resource" and "the beliefs about consequences," involve many barriers and enablers. Within the identical structural domain, both barriers and enablers arose, and the correlation between the two was notably strong.

The initial point is the discrepancy between functional complexity and operational simplicity. Within the "environmental context and resource" domain, patients simultaneously demand operational simplicity and comprehensive functionalities. To address patients' requirements of operational simplicity, we must reduce the challenges to assure accessibility. This aligns with the findings of our and other research from the viewpoint of physicians [25,41,42]. The complexity of system operation significantly affects both physicians and patients. However, excessive feature expansion paradoxically increases interface complexity, hindering core functionalities. This phenomenon aligns with intrinsic cognitive load theory [43]. In the development of telemedicine services, it is imperative to address this equilibrium. This is a critical aspect that we meticulously consider while devising potential intervention strategies.

The second aspect is the trade-off between convenience and privacy. The "beliefs about consequences" domain encompasses opposing factors: perceived convenience and privacy concerns. In the study, the convenience of telemedicine services is the most frequently mentioned enabler by patients. However, this depends on patients supplying comprehensive personal information, medical history, and privacy. This also results in an elevated danger of personal privacy breaches. Although the issue of privacy breaches is only mentioned once in this study, it is frequently cited as a barrier in many studies on the use of telemedicine services [44]. Patients tend to resist using telemedicine services due to the risk of privacy breaches [45]. This also results in a simultaneous presence of dependence on convenience and sensitivity to contradictions, leading to conflicting and hesitant feelings in patients during the usage process [46].

The Mismatch Between the Demand and Usage Capacity

In skills, digital literacy is a significant influencing element. Numerous interviewers highlighted the concern of digital literacy for older adults, despite the interviewers in the study being youthful. Older patients, despite increasing demand resulting from mobility restrictions and multimorbidity, face barriers due to limited digital literacy and adaptability [47,48]. In contrast, younger groups with technical skills demonstrate reduced demand. This generational mismatch hinders the use of telemedicine services by patients.

In establishing intervention strategies for telemedicine services, it is essential to address the requirements and features of both generations, rather than focusing on only one group. The alteration of text size was frequently referenced during the interview. Consequently, we devised an app modification strategy for the older people, offering extensive assistance while preserving the core functionality. We have also developed the "family affection account" feature, which enables children to assist the older adults in completing telemedicine-related processes.

Legal and Security Concerns

In research primarily focusing on Western populations, privacy concerns are often cited as a major barrier to adoption [41,49-51]. Yet, this barrier was not prominent in our study, with only one participant explicitly raising privacy-related concerns. We speculate on the possible reasons from both macro levels and micro levels.

At the macro level, the social discourse and privacy awareness in China and Western countries differ markedly. In Europe and the United States, privacy protection is widely discussed [52], whereas in China, discussions regarding medical data privacy are limited, and most patients have low awareness, leading to infrequent expression of privacy concerns [52,53].

At the micro level, Chinese patients typically exhibit a high degree of trust in health care institutions, largely due to their public and government-regulated nature [54], which may further reduce concerns about potential data misuse [55].

Limitations of the Study

This study has 3 main limitations. First, the sample consisted only of individuals who had already used telemedicine and were recruited from a single tertiary public hospital in Beijing, which may limit the generalizability of the findings. Future studies should consider including more diverse populations. Second, using a deductive approach to map participants' statements to TDF domains may result in inconsistencies across coders. To mitigate this inherent limitation of deductive analysis, we used double independent coding, with any discrepancies resolved through adjudication by a third coder. Given the high

XSL•FO
RenderX

consistency of participant narratives regarding barriers and enablers, we believe any residual impact is minimal. Third, as this study represents the first paper in a planned series, its primary goal was to identify barriers to and enablers of telemedicine adoption, thereby laying the foundation for intervention development. Thus, this study did not assess the effectiveness of proposed interventions. In future work, we will incorporate considerations of feasibility and other factors to prioritize specific interventions and rigorously evaluate their effectiveness.

Conclusions

This study offered vital insights by identifying barriers to and enablers of patients using telemedicine services based on the TDF model. These findings not only genuinely represent patients' experiences but also reflect the behavioral and psychological factors in using telemedicine services. This study revealed barriers, including operational difficulties, extended waiting times, and doubts concerning therapeutic effectiveness. Simultaneously, the enablers included the convenience of telemedicine, time conservation, and support from hospitals. Considering these criteria, we developed intervention strategies to enhance patients' access to telemedicine. This research offers a solid foundation for developing more targeted intervention strategies and holds substantial practical implications for improving the telemedicine experience for patients.

Acknowledgments

The authors would like to appreciate the generous time that the interviewed participants gave during data collection. The authors sincerely thank the reviewers for their constructive comments and valuable suggestions, which greatly improved the quality of this manuscript.

Funding

This work was supported by the National Natural Science Foundation (72204005) and Capital's Funds for Health Improvement and Research under grant number 2024-2-40915.

Data Availability

The datasets used in this study are available from the corresponding author upon reasonable request.

Authors' Contributions

KL and YY are co-first authors of this manuscript and contributed equally to this research. KL, YY, and YW planned and designed the study. KL and YY conducted the interview and collected and coded the data. ZS, HL, and YL assisted in organizing the interview records. KZ, XW, and ZH contributed the administrative and technical support of the study. KL, YY, and YW wrote the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The coding process and results (excerpt).
[DOCX File, 24 KB - humanfactors_v13i1e78457_app1.docx]

Checklist 1
COREQ checklist.
[DOCX File, 20 KB - humanfactors_v13i1e78457_app2.docx]

References

1. Klaassen B, van Beijnum BJF, Hermens HJ. Usability in telemedicine systems-a literature survey. Int J Med Inform 2016 Sep;93:57-69. [doi: 10.1016/j.ijmedinf.2016.06.004] [Medline: 27435948]
2. Zhang X, Lai KH, Guo X. Promoting China's mHealth market: a policy perspective. Health Policy Technol 2017 Dec;6(4):383-388. [doi: 10.1016/j.hlpt.2017.11.002]
3. Hong Z, Li N, Li D, et al. Telemedicine during the COVID-19 pandemic: experiences from Western China. J Med Internet Res 2020 May 8;22(5):e19577. [doi: 10.2196/19577] [Medline: 32349962]
4. Wang Y, Zhang P, Xing Y, et al. Telemedicine integrated care versus in-person care mode for patients with short stature: comprehensive comparison of a retrospective cohort study. J Med Internet Res 2024 Nov 19;26:e57814. [doi: 10.2196/57814] [Medline: 39560983]
5. Carrillo de Albornoz S, Sia KL, Harris A. The effectiveness of teleconsultations in primary care: systematic review. Fam Pract 2022 Jan 19;39(1):168-182. [doi: 10.1093/fampra/cmab077] [Medline: 34278421]

XSL•FO
RenderX

6. Peters GM, Kooij L, Lenferink A, van Harten WH, Doggen CJM. The effect of telehealth on hospital services use: systematic review and meta-analysis. J Med Internet Res 2021 Sep 1;23(9):e25195. [doi: 10.2196/25195] [Medline: 34468324]
7. Elbert NJ, van Os-Medendorp H, van Renselaar W, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014 Apr 16;16(4):e110. [doi: 10.2196/jmir.2790] [Medline: 24739471]
8. Snoswell CL, Taylor ML, Comans TA, Smith AC, Gray LC, Caffery LJ. Determining if telehealth can reduce health system costs: scoping review. J Med Internet Res 2020 Oct 19;22(10):e17298. [doi: 10.2196/17298] [Medline: 33074157]
9. Tan H, Zhang X, Yang Y. Satisfaction or gratitude? Exploring the disparate effects of physicians' knowledge sharing on patients' service evaluation in online medical consultations. Inf Syst J 2023 Sep;33(5):1186-1211. [doi: 10.1111/isj.12440]
10. Wang Z, Wang Y, Qiu Y, Jin C. Strategic resource allocation in dual-channel healthcare systems: the role of telemedicine in physician assignment. IEEE Trans Autom Sci Eng 2025;22:13569-13587. [doi: 10.1109/TASE.2025.3549822]
11. Caulfield BM, Donnelly SC. What is connected health and why will it change your practice? QJM 2013 Aug;106(8):703-707. [doi: 10.1093/qjmed/hct114] [Medline: 23676416]
12. Statistical bulletin on the development of China's health and wellness industry in 2023 [Report in Chinese]. : National Health Commission of the People's Republic of China; 2024 URL: https://www.yi123.cc/doc/7e50qoexjc [accessed 2026-01-09]
13. Wang S, Wang X, Zhou Y, Xu J. Utilization of, satisfaction toward, and challenges for Internet-based healthcare services provided by primary health institutions: evidence from China. Front Public Health 2023;10:1100634. [doi: 10.3389/fpubh.2022.1100634] [Medline: 36743153]
14. Weiner JP, Bandeian S, Hatef E, Lans D, Liu A, Lemke KW. In-person and telehealth ambulatory contacts and costs in a large US insured cohort before and during the COVID-19 pandemic. JAMA Netw Open 2021 Mar 1;4(3):e212618. [doi: 10.1001/jamanetworkopen.2021.2618] [Medline: 33755167]
15. Zhou M, Zhao L, Kong N, Campy KS, Qu S, Wang S. Factors influencing behavior intentions to telehealth by Chinese elderly: an extended TAM model. Int J Med Inform 2019 Jun;126:118-127. [doi: 10.1016/j.ijmedinf.2019.04.001] [Medline: 31029253]
16. de Veer AJE, Peeters JM, Brabers AEM, Schellevis FG, Rademakers JJDJM, Francke AL. Determinants of the intention to use e-Health by community dwelling older people. BMC Health Serv Res 2015 Mar 15;15:103. [doi: 10.1186/s12913-015-0765-8] [Medline: 25889884]
17. AlBar AM, Hoque MR. Patient acceptance of e-Health services in Saudi Arabia: an integrative perspective. Telemed J E Health 2019 Sep;25(9):847-852. [doi: 10.1089/tmj.2018.0107] [Medline: 30452333]
18. Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: systematic review. J Med Internet Res 2019 May 21;21(5):e13117. [doi: 10.2196/13117] [Medline: 31115340]
19. Lee AT, Ramasamy RK, Subbarao A. Understanding psychosocial barriers to healthcare technology adoption: a review of TAM technology acceptance model and unified theory of acceptance and use of technology and UTAUT frameworks. Healthcare (Basel) 2025 Jan 27;13(3):250. [doi: 10.3390/healthcare13030250] [Medline: 39942440]
20. Li D, Hu Y, Pfaff H, et al. Determinants of patients' intention to use the online inquiry services provided by internet hospitals: empirical evidence from China. J Med Internet Res 2020 Oct 29;22(10):e22716. [doi: 10.2196/22716] [Medline: 33006941]
21. Harris MA, Freeman KA, Duke DC. Seeing is believing: using Skype to improve diabetes outcomes in youth. Diabetes Care 2015 Aug;38(8):1427-1434. [doi: 10.2337/dc14-2469] [Medline: 26033508]
22. Green SM, Lockhart E, Marhefka SL. Advantages and disadvantages for receiving internet-based HIV/AIDS interventions at home or at community-based organizations. AIDS Care 2015;27(10):1304-1308. [doi: 10.1080/09540121.2015.1051503] [Medline: 26357907]
23. Li P, Luo Y, Yu X, et al. Patients' perceptions of barriers and facilitators to the adoption of e-Hospitals: cross-sectional study in Western China. J Med Internet Res 2020 Jun 11;22(6):e17221. [doi: 10.2196/17221] [Medline: 32525483]
24. Xu P, Jin Y, Guo P, et al. Barriers and enablers of pelvic floor rehabilitation behaviours in pregnant women with stress urinary incontinence: a qualitative analysis using the Theoretical Domains Framework. BMC Pregnancy Childbirth 2023 Apr 28;23(1):300. [doi: 10.1186/s12884-023-05633-2] [Medline: 37118702]
25. Wu X, Yang Y, Li Y, et al. Using theory-based frameworks to identify barriers and enablers of physicians' telemedicine adoption and develop intervention strategies in China: multicenter qualitative study. J Med Internet Res 2025 Sep 8;27:e73412. [doi: 10.2196/73412] [Medline: 40920450]
26. Cane J, O'Connor D, Michie S. Validation of the Theoretical Domains Framework for use in behaviour change and implementation research. Implement Sci 2012 Apr 24;7:37. [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
27. Shachak A, Kuziemsky C, Petersen C. Beyond TAM and UTAUT: future directions for HIT implementation research. J Biomed Inform 2019 Dec;100:103315. [doi: 10.1016/j.jbi.2019.103315] [Medline: 31629923]
28. Atkins L, Francis J, Islam R, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implement Sci 2017 Jun 21;12(1):77. [doi: 10.1186/s13012-017-0605-9] [Medline: 28637486]

XSL•FO
RenderX

29. Yamada J, Kouri A, Simard SN, Segovia SA, Gupta S. Barriers and enablers to using a patient-facing electronic questionnaire: a qualitative Theoretical Domains Framework analysis. J Med Internet Res 2020 Oct 8;22(10):e19474. [doi: 10.2196/19474] [Medline: 33030437]
30. Han Y, Lie RK, Guo R. The internet hospital as a telehealth model in China: systematic search and content analysis. J Med Internet Res 2020 Jul 29;22(7):e17995. [doi: 10.2196/17995] [Medline: 32723721]
31. Xu D, Zhan J, Cheng T, Fu H, Yip W. Understanding online dual practice of public hospital doctors in China: a mixed-methods study. Health Policy Plan 2022 Apr 12;37(4):440-451. [doi: 10.1093/heapol/czac017] [Medline: 35266518]
32. Peking University Third Hospital has been approved as the chair unit of the Beijing internet-based medical service quality control and improvement center [Web page in Chinese]. Peking University Third Hospital. 2025. URL: https://www.puh3.net.cn/info/1971/21191.htm [accessed 2026-01-09]
33. Francis JJ, Johnston M, Robertson C, et al. What is an adequate sample size? Operationalising data saturation for theory-based interview studies. Psychol Health 2010 Dec;25(10):1229-1245. [doi: 10.1080/08870440903194015] [Medline: 20204937]
34. Talkhan H, Stewart D, McIntosh T, et al. Exploring determinants of antimicrobial prescribing behaviour using the Theoretical Domains Framework. Res Social Adm Pharm 2024 Apr;20(4):401-410. [doi: 10.1016/j.sapharm.2023.12.009] [Medline: 38320948]
35. Bele S, Cassidy C, Curran J, Johnson DW, Bailey JAM. Using the Theoretical Domains Framework to identify barriers and enablers to implementing a virtual tertiary-regional telemedicine rounding and consultation for kids (TRaC-K) model: qualitative study. J Med Internet Res 2021 Dec 22;23(12):e28610. [doi: 10.2196/28610] [Medline: 34941561]
36. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13(117):24047204. [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
37. Hearn JH, Bryson K, Barsauskaite L, Bullo S. A COM-B and Theoretical Domains Framework mapping of the barriers and facilitators to effective communication and help-seeking among people with, or seeking a diagnosis of, endometriosis. J Health Commun 2024 Mar 3;29(3):174-186. [doi: 10.1080/10810730.2023.2299003] [Medline: 38262475]
38. Wang L, Wu T, Guo X, Zhang X, Li Y, Wang W. Exploring mHealth monitoring service acceptance from a service characteristics perspective. Electron Commer Res Appl 2018;30:159-168. [doi: 10.1016/j.elerap.2018.06.003]
39. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172. [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
40. Eyrich NW, Andino JJ, Fessell DP. Bridging the digital divide to avoid leaving the most vulnerable behind. JAMA Surg 2021 Aug 1;156(8):703-704. [doi: 10.1001/jamasurg.2021.1143] [Medline: 33769441]
41. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. J Telemed Telecare 2018 Jan;24(1):4-12. [doi: 10.1177/1357633X16674087] [Medline: 29320966]
42. Yang Y, Zhu X, Song R, Zhang X, Guo F. Not just for the money? An examination of the motives behind physicians' sharing of paid health information. J Inf Sci 2023 Feb;49(1):145-163. [doi: 10.1177/0165551521991029]
43. Haji FA, Rojas D, Childs R, de Ribaupierre S, Dubrowski A. Measuring cognitive load: performance, mental effort and simulation task complexity. Med Educ 2015 Aug;49(8):815-827. [doi: 10.1111/medu.12773] [Medline: 26152493]
44. Ansarian M, Baharlouei Z. Applications and challenges of telemedicine: privacy-preservation as a case study. Arch Iran Med 2023 Nov 1;26(11):654-661. [doi: 10.34172/aim.2023.96] [Medline: 38310426]
45. Churruca K, Foo D, Turner A, et al. Users' perspectives of direct-to-consumer telemedicine services: survey study. JMIR Form Res 2025 Feb 3;9:e68619. [doi: 10.2196/68619] [Medline: 39899749]
46. Zhang X, Guo X, Guo F, Lai KH. Nonlinearities in personalization-privacy paradox in mHealth adoption: the mediating role of perceived usefulness and attitude. Technol Health Care 2014;22(4):515-529. [doi: 10.3233/THC-140811] [Medline: 24763205]
47. Choi N. Relationship between health service use and health information technology use among older adults: analysis of the US National Health Interview Survey. J Med Internet Res 2011 Apr 20;13(2):e33. [doi: 10.2196/jmir.1753] [Medline: 21752784]
48. Jing R, Lai X, Li L, Wang H. Internet use and healthcare utilization among older adults in China: a nationally representative cross-sectional study. BMC Geriatr 2025 Mar 5;25(1):150. [doi: 10.1186/s12877-025-05798-9] [Medline: 40045191]
49. Mishkin AD, Zabinski JS, Holt G, Appelbaum PS. Ensuring privacy in telemedicine: ethical and clinical challenges. J Telemed Telecare 2023 Apr;29(3):217-221. [doi: 10.1177/1357633X221134952] [Medline: 36349356]
50. Pool J, Akhlaghpour S, Fatehi F, Gray LC. Data privacy concerns and use of telehealth in the aged care context: an integrative review and research agenda. Int J Med Inform 2022 Apr;160:104707. [doi: 10.1016/j.ijmedinf.2022.104707] [Medline: 35131698]
51. Solimini R, Busardò FP, Gibelli F, Sirignano A, Ricci G. Ethical and legal challenges of telemedicine in the era of the COVID-19 pandemic. Medicina (Kaunas) 2021 Nov 30;57(12):1314. [doi: 10.3390/medicina57121314] [Medline: 34946259]
52. Sun L. Health data governance in China: emphasizing 'sharing' and 'protection' based on the right to health. Med Law Int 2023 Mar;23(1):26-43. [doi: 10.1177/09685332221140416]

XSL•FO
RenderX

53. Zhang H, Zhang H, Zhang Z, Wang Y. Patient privacy and autonomy: a comparative analysis of cases of ethical dilemmas in China and the United States. BMC Med Ethics 2021 Feb 2;22(1):8. [doi: 10.1186/s12910-021-00579-6] [Medline: 33531011]
54. Huang YN, Yin WQ, Liu Q, Wen QZ, Huang DM, Chen ZM, et al. Analysis on the relationship between residents' medical trust and social trust [in Chinese]. Chin Health Serv Manage 2018;35(10):777-779.
55. Mi J, Du X, Aili Y, Li M, Liu Y. Unveiling patient trust in physicians across China: insights from a nationwide cross-sectional study. Risk Manag Healthc Policy 2024;17:1039-1052. [doi: 10.2147/RMHP.S456585] [Medline: 38680478]

Abbreviations

mHealth: mobile health
TAM: technology acceptance model
TDF: Theoretical Domains Framework
UTAUT: unified theory of acceptance and use of technology

Edited by A Kushniruk; submitted 03.Jun.2025; peer-reviewed by J Claggett, X Zhang; revised version received 03.Dec.2025; accepted 28.Dec.2025; published 21.Jan.2026.

Please cite as:
Liu K, Yang Y, Song Z, Li H, Lyu Y, Zhang K, Wu X, Hou Z, Wang Y
Using the Theoretical Domains Framework to Identify Barriers to and Enablers of Patient Telemedicine Services Use in China: Qualitative Study
JMIR Hum Factors 2026;13:e78457
URL: https://humanfactors.jmir.org/2026/1/e78457
doi:10.2196/78457

© Ke Liu, Yuting Yang, Zixuan Song, Huixian Li, Yanli Lyu, Ke Zhang, Xinxia Wu, Zheng Hou, Yipei Wang. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 21.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Hospital-at-Home for South Asian Communities in British Columbia, Canada: Qualitative Interview Study

Emma Wong[1], MPH; Mahabhir Kandola[1], MSc; Kamal Arora[1], PhD; Harroop Sharda[2], PhD; Roman Deol[1], BSc, MBA; Mary Jung[3], PhD; Robert Paquin[4], BSN, MSc; Maria Montenegro[1], MSc; Megan MacPherson[1,5], PhD

[1]Department of Virtual Health, Fraser Health, 13450 102 Ave, Surrey, BC, Canada

[2]British Columbia Institute of Technology, Burnaby, BC, Canada

[3]School of Health and Exercise Sciences, University of British Columbia, Kelowna, BC, Canada

[4]King's College London, London, United Kingdom

[5]Department of Occupational Science and Occupational Therapy, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Megan MacPherson, PhD
Department of Virtual Health, Fraser Health, 13450 102 Ave, Surrey, BC, Canada

Abstract

Background: South Asian communities in Canada face significant disparities in access to health care and experience higher rates of chronic conditions such as cardiovascular disease, diabetes, and hypertension. Hospital-at-Home services have the potential to improve access and outcomes, yet little is known about how these services are perceived and experienced by South Asian patients and caregivers. Understanding both barriers and facilitators is critical for culturally responsive implementation.

Objective: This study aimed to explore the experiences of South Asian community members with in-person hospital care and their perceptions, attitudes, and expectations regarding virtual Hospital-at-Home services, with the goal of identifying culturally tailored strategies to improve access, quality, and satisfaction.

Methods: A qualitative study using semistructured interviews was conducted with 20 South Asian community members in the Fraser Health region in British Columbia, Canada. Interviews explored experiences with in-person hospital care, perceptions of a virtual hospital service (also known as Hospital-at-Home), and recommendations for enhancing awareness and accessibility. Interviews were audio-recorded, transcribed, and analyzed thematically to identify key patterns in perceptions, experiences, and needs.

Results: Participants described multiple systemic barriers to in-person hospital care, including long wait times, overcrowding, transportation challenges, and difficulty navigating the health system. Cultural and religious needs, such as gender-concordant care and culturally appropriate food, were frequently unmet, while language-concordant care and family involvement were critical to positive experiences. Discrimination and assumptions based on ethnicity or age further shaped perceptions of care. Virtual hospital services were valued for convenience, comfort, reduced exposure to hospital-acquired infections, and support for family involvement. However, participants raised concerns about clinical quality, the absence of physical examinations, digital literacy, privacy, and home-based responsibilities. Acceptance varied by age, immigration status, and familiarity with technology. Participants emphasized the importance of culturally tailored outreach, leveraging community leaders, ethnic media, and peer testimonials to increase awareness and trust.

Conclusions: South Asian patients and caregivers recognize both challenges in traditional hospital care and potential benefits of Hospital-at-Home services. Implementation strategies that address systemic barriers, integrate cultural and linguistic considerations, and engage trusted community networks are essential to improving equity, access, and satisfaction. Findings highlight the need for culturally responsive, patient-centered approaches in the design and delivery of virtual health services for racialized populations.

(JMIR Hum Factors 2026;13:e79675) doi:10.2196/79675

KEYWORDS

telemedicine; South Asian people; qualitative research; health equity; Hospital-at-Home

XSL•FO
RenderX

Introduction

Background

The South Asian population, which includes individuals with ethnic roots originating from Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, and Sri Lanka, constitutes the largest visible minority or racialized group in Canada [1]. South Asians in Canada face a higher prevalence of cardiovascular disease, diabetes, and hypertension compared to other racialized groups [2,3]. While research suggests that genetic predisposition is a partial contributor [4], the observed health disparities likely also result from social and environmental determinants of health, such as access to higher education, employment opportunities, food security, experiences of discrimination, and access to culturally appropriate health care services [5-7]. It is important to recognize that South Asian populations in Canada are highly diverse, encompassing a range of languages, religions, socioeconomic backgrounds, and migration histories, all of which influence health experiences and outcomes. Previous research has shown that South Asians encounter significant barriers in accessing health care resources for chronic disease prevention and management [8]. Given the inequities and disparities faced by South Asian community members within Canada, innovative care strategies are needed to enhance access to culturally relevant care and improve health outcomes.

Virtual Health Care

Virtual health care, which uses remote technologies to deliver health services, has been identified as a key mechanism for improving access to care [9]. Specifically, virtual hospitals-at-home (commonly referred to as simply Hospital-at-Home) provide an alternative to traditional "brick and mortar" hospital admissions, allowing patients to receive the care they need from the comfort of their own home [10]. This care model has been increasingly used worldwide, accelerated by hospital capacity strains during the COVID-19 pandemic [11,12]. While admitted to a Hospital-at-Home, patients are monitored remotely using digital devices that transmit data to their multidisciplinary care team [10] and have daily virtual consultations via telephone or video, supplemented with home visits when necessary [10].

Research has shown that Hospital-at-Home models are both feasible and effective across various conditions [13]. Studies report that patients opt into virtual hospitals at high rates, and that this approach delivers high-quality care with fewer complications (eg, shorter hospital stays, lower incidence of delirium, and reduced rate of falls [14]). Patients and caregivers frequently express higher satisfaction with home-based care, citing increased comfort, privacy, and family engagement as key benefits [13].

Beyond individual benefits, Hospital-at-Home can alleviate system-wide pressures by reducing wait times, freeing up hospital bed capacity, and expanding access to care [15]. Despite these advantages, challenges remain. Some studies have noted higher readmission risk for patients with complex medical conditions, as well as increased caregiver burden in home settings [13]. Additionally, concerns about communication barriers and the need for structured care coordination highlight areas where further improvements are needed to ensure equitable and effective implementation of Hospital-at-Home [13].

Given the potential of virtual hospitals to enhance access to care, particularly for underserved populations, it is critical to understand and address the barriers to their adoption and use. This is especially important for racialized communities, such as South Asians in Canada, who already experience health disparities and face unique challenges in accessing virtual health services [16].

Virtual Health Use in South Asian Communities

While virtual care has the potential to alleviate barriers experienced by South Asian communities [17] in Canada, and research suggests positive attitudes toward integrating virtual services [16], significant barriers still exist for accessing these services. Studies have shown that age, gender, income, education, and language are associated with the ability to use health-related technologies among Panjabi-speaking (also known as "Punjabi") South Asian individuals in Canada [18] and that language preferences, education, age, and sex predict the use of virtual care among South Asians in Alberta [19]. In both studies, older adults, women, those with a preference for non-English communication, and those with lower educational levels were less likely to engage with virtual care [18,19]. A recent community-based participatory action research project in Surrey, British Columbia, Canada further supported these findings [16]. Through focus groups and photo voice activities, participants identified concerns about their ability to navigate new technologies as they age, gendered responsibilities limiting women's access to technology, and communication difficulties due to language barriers, literacy levels, and digital literacy [16].

Although virtual care is designed to support disease prevention and self-management, these services are often developed without equity in mind, are not culturally responsive, and are generally underused by South Asian populations in Canada [16]. The divide in access and engagement with virtual care services could further exacerbate health disparities faced by South Asian communities [8]. Given the relative novelty of virtual hospitals, little is known about their acceptance among South Asian communities. Future virtual care models must address known barriers to ensure equitable access.

Setting

Health care in British Columbia, Canada, is provided by a provincial health authority, 5 regional health authorities, and a First Nations health authority [20]. The Fraser Health Authority is the largest regional health authority, serving over 2 million people across 20 diverse and rapidly growing communities [21]. According to the 2021 Census, nearly 1 in 5 Fraser Health residents identifies as South Asian [22]. Fraser Health emphasizes understanding how religion, language, and cultural practices influence health to develop customized programs and services.

To better support South Asian communities, Fraser Health established the South Asian Health Institute (SAHI) in 2013, focusing on research, innovation, and evidence-based care [5]. Fraser Health's Virtual Health team supports the development

XSL•FO
RenderX

and implementation of new virtual care pathways across the region. One such pathway is the virtual hospital service. As Fraser Health serves 75% of South Asian residents in the province, it is essential to conduct patient-oriented research, or research done in direct partnership with patients, to understand what matters to South Asian communities when developing the virtual hospital [5].

Research Objectives

This patient-oriented research study aims to engage South Asian community members through 3 primary objectives:

1. Capture and understand the emotional experiences, as well as the perceived barriers and facilitators, of members within South Asian communities in the context of both in-person and Hospital-at-Home systems.
2. Understand the lived experiences of South Asian community members in navigating the health care system and identify gaps in accessing care.
3. Identify opportunities for how Hospital-at-Home services can be implemented to improve the access and flow of care among South Asian communities.

Methods

Overview

This qualitative study used semistructured interviews to explore the experiences, perceptions, and attitudes of South Asian community members in British Columbia, Canada, toward in-person hospital care and the emerging Virtual Hospital-at-Home model. The study was guided by a patient-oriented research approach, prioritizing the perspectives of patients and caregivers to inform culturally relevant virtual care strategies.

This study was guided by a steering group including members from Progressive Intercultural Community Services (PICS) Society [23], Fraser Health's Virtual Health department [24], Fraser Health's SAHI [25], and Fraser Health's Equity, Diversity, and Inclusion team [26], and academic partners from the University of British Columbia and King's College London.

Interviews were guided by a pragmatic epistemology, which seeks to generate practical findings that can be readily applied in real-world settings while also tolerating multiple truths [27]. Although all South Asian community members may be exposed to the same care environments, everyone has unique experiences.

Recruitment

Participants were recruited through convenience sampling using outreach by the PICS Society and the SAHI, and supplemented with community outreach efforts including social media, places of worship, and public libraries. The inclusion criteria were that participants must (1) identify as South Asian, (2) reside within the Fraser Health region, and (3) have been a patient in the hospital in the past year or have been a caregiver to someone in the hospital in the past year. Recruitment continued until thematic saturation was reached, when no new concepts emerged from subsequent interviews.

Data Collection

South Asian community members were interviewed to understand their lived experiences navigating the virtual health care system. A topic guide, adapted from previous Virtual Health patient partner workshops, was revised with feedback from community members and edited by Fraser Health's plain language team for clarity. The guide covered 2 primary domains: (1) experiences with in-person hospital care, including barriers, facilitators, and family involvement; and (2) perceptions of, and attitudes toward, the Virtual Hospital-at-Home model, including benefits, concerns, and strategies for community engagement. Interviews began with a screening question to classify participants as patients, caregivers, or both. A full interview topic guide is provided in Multimedia Appendix 1.

Thirty- to 60-minute, one-on-one interviews were conducted via Microsoft Teams, with the PICS Society providing technical support when needed to ensure participant access to the platform. To ensure accessibility, interpretation services were available in Panjabi, Hindi, Urdu, and Pashto. All interviews were audio-recorded with participant consent, transcribed verbatim, and translated where required by members of the research team. Interviews were conducted by EW and MK, who both had master's degrees and worked as a "Research Assistant" and the "Research and Knowledge Translation Lead" for the Virtual Health department at the time of interviews. All interviewees had no prior relationships with those conducting the interviews.

Data Analysis

Thematic analysis guided by the approach of Braun and Clarke [28] was conducted using NVivo (QSR International). First, transcripts were read repeatedly for familiarization. An initial set of open codes was inductively generated to capture salient features of the data. Codes were iteratively refined and grouped into higher-order categories. Themes were then developed through constant comparison across transcripts, paying attention to both convergent and divergent perspectives. Theme prevalence was assessed approximately by noting how many participants raised each theme, though exact tallies were not feasible given the conversational variability of interviews. Findings were also interpreted with attention to cultural safety and equity considerations relevant to South Asian communities in the Fraser Health region.

Coding was conducted by MK and cross-checked by MM; discrepancies were resolved by consensus. For multilingual interviews, translations were reviewed by bilingual team members to ensure conceptual equivalence. Final themes were reviewed and refined in consultation with the steering committee to enhance cultural validity. The steering committee did not influence the coding of raw data. Thematic saturation was achieved by the 17th interview; 3 additional interviews confirmed that no new themes emerged.

Ethical Considerations

This study received approval from the Fraser Health Research Ethics Board (H23-03144). Interested participants were sent a consent form, the topic guide, and a video explaining the Hospital-at-Home to review beforehand. Informed consent for

XSL•FO
RenderX

participation and interview recording was obtained verbally prior to the interview. To protect privacy and confidentiality, all transcripts were anonymized, and identifiers were removed prior to analysis. Data were stored securely on password-protected institutional servers. Participants were compensated with a US $14.86 gift card at the conclusion of their interview.

Results

Participant Characteristics

Twenty participants were interviewed (n=9 male, n=11 female). Participants were aged 19 - 79 years, most commonly between 60 - 69 years (6/20, 30%). Younger adults aged 19 - 29 (n=5), 30 - 39 (n=4), and 40 - 49 (n=4) years were also represented, along with 1 participant in their 70s. Interviews were conducted in English (n=15), Panjabi (n=3), Pashto (n=1), and Hindi (n=1).

Most participants were originally from India (14/20, 70%), with others from Afghanistan (3/20, 15%), and smaller representations from Iraq (n=1), Pakistan (n=1), and India and Fiji (n=1). Half had lived in the community for more than 10 years (n=10); the remainder were more recent arrivals (n=4 lived in Canada <1 year, n=5 lived here 1 - 3 years). Regarding immigration status, 9 were Canadian citizens and 11 were permanent residents.

Self-rated health was generally positive—6 participants (30%) rated their health as "very good," 6 (30%) as "good," and 8 (40%) as "average." Six participants (30%) reported having a chronic condition. Participants included caregivers only (7/20, 35%), patients only (6/20, 30%), and both caregivers and patients (5/20, 25%). Two participants also contributed professional perspectives, working in newcomer services and health care.

All participants had access to a smartphone (20/20, 100%), while 11 (55%) had access to a tablet and 11 (55%) had access to a laptop. Ten participants (50%) had all 3 devices, while 8 (40%) had only a smartphone. Device use was frequent—14 (70%) used devices daily or multiple times per day, 4 (20%) used devices regularly but not daily, and 2 (10%) did not regularly use devices.

Experiences of In-Person Hospital Care

Overview

Participants described a mix of challenges and positive experiences with in-person hospital care. Six overarching themes emerged: systemic barriers to accessing care, communication and language barriers, cultural and religious needs not met, family and caregiver roles, discrimination and bias, and positive experiences and gratitude. Experiences varied by age, gender, immigration status, and length of time in Canada. Older and more recently arrived participants reported more barriers than younger or Canadian-born participants.

Systemic Barriers to Accessing Care

Participants frequently reported challenges accessing timely and appropriate hospital services, including long wait times, overcrowding, transportation difficulties, and navigation challenges. These barriers contributed to frustration, fatigue, and inequitable access.

Long Wait Times

Wait times were among the most common challenges mentioned. Many participants described hours-long waits in emergency departments, sometimes extending overnight or across multiple days. As 1 participant shared, "we have waited with a broken wrist 28 hours" (P10, female, aged 60 - 69 years, >10 years in community). Another recalled, "you have to wait maybe sometimes 10 hours or more than that and you have a critical illness" (P20, female, aged 60 - 69 years, >10 years in community). Such prolonged waits were perceived as not only inconvenient but also unsafe when health concerns were urgent.

Overcrowding and Understaffing

Participants described hospitals as overcrowded and short-staffed, with patients left in hallways and nurses visibly stretched thin. One participant reported, "patients are all outside in the hallways. They don't even have proper rooms" (P13, female, aged 60 - 69 years, <1 year in community). Another observed, "sometimes you can see like the nurses are really tired, and then you see them overwhelmed with the amount of work" (P11, female, aged 19 - 29 years, >10 years in community). These accounts reflect how structural resource constraints were felt by both patients and providers.

Transportation Difficulties

Transportation presented another layer of inequity, particularly for patients without private vehicles. Some relied on long bus rides with multiple transfers, which became especially challenging for older adults and those with mobility issues. One participant explained,

> *The nearest hospital is like a 45 minute transit ride, but the bus doesn't go directly there. So I have to walk. For me as an able-bodied male, mid-20s, it's fine. But...for my grandma who doesn't drive and can't speak much English, it's very hard.* [P1, male, aged 20 - 29 years, >10 years in community]

Another recalled, "I rely on HandyDART, and if that's not available then my son would have to take time off work to drive me" (P17, female, aged 30 - 39 years, >10 years in community). These logistical challenges shaped how and when participants sought hospital care.

Navigation Challenges for Newcomers

For newcomers to Canada, navigating the health system added further barriers. Several participants spoke about confusion around referrals, insurance, and entitlements. One reflected,

> *I took around a year's time to understand what was MSP and how to access it. What is covered under it and what do I need other insurance for?* [P8, female, aged 30 - 39 years, 1 - 3 years in community]

Another highlighted misunderstanding about referral pathways:

> *You have to first go to a family doctor, and then if the problem is serious then you go to hospital. Many of the people don't have that understanding.* [P6, male, aged 40 - 49 years, >10 years in community]

XSL•FO
RenderX

Such experiences suggest that lack of system literacy can delay or complicate access to hospital care.

Language and Communication

Communication challenges were central to participants' accounts, particularly for older adults and newcomers with limited English proficiency, with many describing difficulties conveying symptoms, understanding instructions, or navigating interactions without the help of family. While some praised interpreters or language-concordant providers, gaps in communication were widely felt and often shaped the overall hospital experience.

Limited English Proficiency

Participants often highlighted the difficulty of explaining their health concerns in English. For example, one explained,

> *If I talk about my grandfather, they don't know how to speak in English...if they are taking them alone into the rooms, they can't express how they're feeling or they can't tell about their problems.* [P5, male, aged 19 - 29 years, 1 - 3 years in community]

Another noted, "when the other person is not from my community, the only hurdle I feel is the language" (P9, female, aged 30 - 39 years, 4 - 6 years in community). These challenges not only caused stress but also risked miscommunication in clinical encounters and increased reliance on family members acting as interpreters. One participant described, "she had a family member of hers that spoke on her behalf and communicated with the doctors" (P4, female, aged 60 - 69 years, >10 years in community). Similarly, another shared, "mom didn't speak English, so until I got there and started helping her she didn't have any good idea as to what was happening" (P14, male, aged 60 - 69 years, >10 years in community). While this reliance on family interpreters ensured patients could understand their care, it also created dependence on family members, sometimes leading to delays or added burdens. Participants also noticed changes over time, including greater availability of translators and culturally diverse staff. As one shared, "now they have translators everywhere, even in the parking. Greeters in the hospital who speak their language" (P17, female, aged 30 - 39 years, >10 years in community). Another noted, "everywhere there are translators in hospitals" (P18, male, aged 70 - 79 years, >10 years in community).

Language-Concordant Care

Participants expressed strong appreciation for health care staff who spoke their language, emphasizing that this reduced stress and improved trust. As one noted, "we prefer to go to [specific hospital] just because the nurses and all the other staff member do speak Panjabi. So it's easier for my grandparents" (P11, female, aged 19 - 29 years, >10 years in community). Another explained, "if the person is from South Asian communities and they are feeling shy or hesitating to say something in front of a White person, a South Asian doctor or nurse can help" (P5, male, aged 19 - 29 years, 1 - 3 years in community). Such accounts highlight how language-concordant care fosters both clarity and cultural comfort.

Cultural and Religious Needs Not Met

Participants frequently described hospital environments that did not align with their cultural or religious values. These included concerns about gender concordance with providers, lack of culturally appropriate food, and perceived insensitivity to religious practices.

Gender-Concordant Care

Gender of providers was an important factor for some participants, particularly for women from Muslim and Sikh backgrounds. One explained, "for me as a Muslim lady, I don't want anybody to touch me if he's a man and get naked in front of him while there is another lady available" (P20, female, aged 60 - 69 years, >10 years in community). Another recalled, "she would have preferred a female nurse, but they had a male" (P1, male, aged 19 - 29 years, >10 years in community). These accounts underscore how gender concordance shaped comfort and dignity in care.

Dietary and Religious Respect

Food was another area where participants felt their needs were not adequately met. Some reported being served meals that conflicted with their dietary or religious practices. For instance, 1 participant explained, "she was a very strict vegetarian, but they fed her beef" (P1, male, aged 19 - 29 years, >10 years in community). Others described hospital food as unappealing or culturally inappropriate:

> *It's a challenge because the hospital food is a challenge for them. They don't like it, obviously, because it's not their cultural food.* [P17, female, aged 30 - 39 years, >10 years in community]

Cultural Insensitivity

In some cases, participants felt their cultural or religious practices were disrespected. One recalled, "she had her religious books and statues by her bed and they moved to a corner by garbage, which is offensive to our religious practices" (P1, male, aged 19 - 29 years, >10 years in community). These incidents, though not universal, left lasting negative impressions and diminished participants' sense of respect for care. Participants recounted instances where staff failed to accommodate dietary and religious needs, sometimes in ways that were profoundly disrespectful.

Family and Caregiver Roles

Family members were described as essential to the hospital experience, serving as translators, advocates, and sources of emotional support. Yet participants also reported that hospital policies and circumstances sometimes limited family involvement, creating both barriers and additional burdens.

Restrictions on Family Presence

Some participants described frustration with restrictions on family members accompanying patients, even when needed for support or translation. One participant recalled, "they were not allowing my family members to come and sit with me" (P2, male, aged 19 - 29 years, <1 year in community). Another shared, "sometimes the person needs emotional support or they need someone from the family...but I can't go with her because

XSL•FO
RenderX

they don't have enough space" (P5, male, aged 19 - 29 years, 1 - 3 years in community). Such restrictions were perceived as leaving patients vulnerable and isolated.

Emotional and Practical Burden on Families

While family involvement was crucial, it often created strain. One participant described, "her son would take her, but he would have to take time off work in order to take her" (P4, female, aged 60 - 69 years, >10 years in community). Another reflected on balancing childcare with caregiving: "sometimes I have to be without food for long times...I cannot also go to the cafeteria with kids" (P9, female, aged 30 - 39 years, 4 - 6 years in community). These stories illustrate how the health system's reliance on families carries significant costs.

Preference for Family Caregiving

While family involvement can be burdensome, many participants preferred it over professional care for trust and cultural reasons. As 1 participant explained, "South Asian people...are more comfortable with their family member, rather than a nurse, doing certain tasks" (P17, female, aged 30 - 39 years, >10 years in community). This highlights how family involvement was not only necessary but often desired.

Discrimination and Bias

Although many participants spoke positively of their care, some described experiences of stereotyping, dismissal, or unequal treatment.

Stereotyping and Assumptions

A few participants reported feeling judged on the basis of cultural stereotypes. One explained, "because they're Indian, they obviously eat unhealthy and because of that they have high blood pressure and heart issues...without doing further research" (P1, male, aged 19 - 29 years, >10 years in community). Such assumptions left participants feeling unseen and unfairly blamed.

Unequal Quality of Care

Others felt that seniors and immigrants received less attentive or less culturally appropriate care. As one participant described, "I've seen that personally with my grandma, that they don't take seniors that seriously" (P11, female, aged 19 - 29 years, >10 years in community). Another reflected, "it's not very culturally appropriate...my grandparents' experience is not the same" (P1, male, aged 19 - 29 years, >10 years in community). These perceptions shaped trust in the system and influenced decisions about where and when to seek care.

Positive Experiences and Gratitude

Despite systemic and cultural challenges, participants consistently recognized skilled, compassionate care. One said,

> *The staff was very good. They were peacefully talking. The doctor was very nice. It was very good.* [P18, male, aged 70 - 79 years,>10 years in community]

Another reflected, "I have the best experience and I was very happy" (P3, male, aged 40 - 49 years, 1 - 3 years in community). Additionally, trust in providers' skill was often expressed as gratitude, even when other barriers were present. For instance, 1 parent stated, "I'm very grateful to all the doctors and nurses who provided care to my child" (P12, female, aged 40 - 49 years, <1 year in community). Such reflections reveal the duality of participants' experiences: systemic and cultural barriers coexisted with respect for the expertise and dedication of individual providers.

Perceptions and Expectations of Hospital-at-Home

Overview

Participants shared diverse perspectives on virtual hospital services, highlighting benefits, limitations, and conditions under which such care is acceptable. Five major themes emerged—convenience, comfort, and family support; cultural and generational differences; quality of care and clinical limitations; emotional safety, privacy, and trust; and barriers to access and technology. Participants' experiences and expectations reflected both practical considerations, such as travel and technology, and cultural and interpersonal factors influencing acceptance. Views often diverged across age, gender, and immigration status, with younger or Canadian-born participants more enthusiastic about virtual care and elders or newcomers expressing hesitation.

Convenience, Comfort, and Family Support

Overview

Participants consistently highlighted the practical and emotional advantages of receiving care at home. Virtual hospital services were valued for reducing stress, increasing comfort, and enabling family involvement, which was seen as supportive both physically and emotionally. Many participants described how being in a familiar environment could enhance recovery, while also reducing exposure to illness compared with hospital settings.

Avoiding Travel and Wait Times

Many participants appreciated the time-saving and stress-reducing aspects of virtual care. One noted,

> *I can be at home and go about my daily things and just wait for a phone call. I don't have to actually sit at all at an actual office anyway.* [P6, male, aged 40 - 49 years, >10 years in community]

Others highlighted avoidance of transportation challenges and hospital stress, stating it would allow them "to get help like at home...I don't have to stuck in the traffic and get stressed out" (P10, female, aged 60 - 69 years, >10 years in community) and "virtual hospitals would be more convenient. He won't have to worry about transportation waiting. Time to come and go" (P7, male, aged 60 - 69 years, 1 - 3 years in community).

Healing at Home

Being at home with family and in a familiar environment was viewed as beneficial for recovery. Participants described emotional and physical advantages: "you are around your family, that motivates you to be more good. I think that's the biggest thing that virtual can provide" (P2, male, aged 19 - 29 years, <1 year in community), and "if this virtual hospital help was available, then definitely he would have felt more better at home rather than at the hospital...we can read books for him, we can sit beside him" (P17, female, aged 30 - 39 years, >10 years in

XSL•FO
RenderX

community). Others emphasized comfort and food preferences, noting,

> *I feel more comfortable in my space...Benefits I can be with my family. That's the best benefit I can eat the food that I want.* [P16, female, aged 30 - 39 years, <1 year in community]

Reduced Exposure to Illness

Participants recognized the safety advantage of avoiding hospital-acquired infections: "It can help many people...rushing to the hospitals because of. danger of getting communicable diseases" (P12, female, aged 40 - 49 years, <1 year in community), and "with the virtual, then there wouldn't be so many people waiting for beds. hospitals have a lot of diseases going around" (P15, female, aged 19 - 29 years, >10 years in community).

Smooth Transitions from Hospital to Home

Several participants highlighted the importance of smooth transitions from hospital to home, emphasizing clear communication, rapid response in emergencies, and reliable monitoring. The availability of equipment and the teach-back style training were viewed as critical.

> *...how the South Asian will be educated...equipment will be there, they'll teach you how to use it. It's important that you use it properly.* [P4, female, aged 60 - 69 years, >10 years in community]

> *If something gets worse, who do I call first? That should be very clear.* [P15, female, aged 19 - 29 years, >10 years in community]

Home-Based Challenges

Participants noted that home environments can pose challenges during recovery, such as stairs, inaccessible bathrooms, or absence of hospital equipment:

> *My bedroom is upstairs and the bathroom is downstairs. After surgery, how can I manage the stairs?* [P7, male, aged 60 - 69 years, 1 - 3 years in community]

> *We don't have special hospital beds at home. Sometimes you need that equipment to recover properly.* [P12, female, aged 40 - 49 years, <1 year in community]

Routine household responsibilities, such as childcare, also limited the ability to rest: "I know if I go home, I have to cook for the other two children and I have to take care of them" (P20, female, aged 60 - 69 years, >10 years in community). Participants worried that home-based care could shift responsibilities onto family members, especially women and elders:

> *Yes, absolutely, definitely. When I am at the hospital, I'm just quite confident that I've been taken care of 100%. But if I am at home, it will put more pressure on me* [P10, female, aged 60 - 69 years, >10 years in community]

Cultural Norms and Generational Differences

Overview

Cultural norms, family expectations, and generational comfort with technology strongly influenced participants' perceptions of virtual hospital care. While younger participants reported ease with online platforms, older adults often struggled with technology and valued in-person monitoring. Cultural beliefs regarding family roles and appropriate care also shaped attitudes toward remote health care.

Generational Divide in Comfort with Technology

Participants highlighted that younger generations were more comfortable with technology, while older adults often struggled. One reflected,

> *For me, growing up in this generation, I'm not shy for technology, but for example, my grandparents, it's very a new world to them. So for them there is more of that disconnect.* [P1, male, aged 19 - 29 years, >10 years in community]

Similarly, a participant noted, "if I imagine my grandmother using this service, she might be overwhelmed with the tablet thing...she would just be like, I just wanna meet people in real" (P16, female, aged 30 - 39 years, <1 year in community).

Cultural Norms and Expectations of Care

Some participants described South Asian cultural preferences for in-hospital monitoring, emphasizing family support and safety:

> *I believe for the South Asian people this will not work very well...I would say at the hospital, I feel the care is around me 24 hours...So I think at the hospital we can get the maximum care.* [P20, female, aged 60 - 69 years, >10 years in community]

Others stressed the importance of culturally sensitive approaches, including language support: "I think having that community-centered culturally sensitive approach to healthcare is really important because this generic advice doesn't help the community" (P1, male, aged 19 - 29 years, >10 years in community).

Absence of Support Persons at Home and Immigrant Social Networks

Some participants identified the absence of support persons at home as a barrier to virtual care. As immigrants, many noted that they no longer have the robust support systems they relied on in their native countries. While they can manage their daily lives, participants expressed doubts about their ability to maintain the same level of self-care when unwell, particularly those living alone or with family members who work full time.

> *It's very hard because you don't have much people around you...You don't have a community...You don't have neighbors as we use to have uh back in in Asia or South East Asia, there's a lot of people around you...But here, because we are immigrant, it's very limited people that we know and it's not the bond even it's not that because of the type of living here as*

XSL•FO
RenderX

we all work or just us to take care of the family...So our relationship is very limited. [P20, female, aged 60 - 69 years, >10 years in community]

Quality of Care and Clinical Limitations

Overview

While participants acknowledged the convenience of virtual care, many expressed concerns about clinical quality. The inability to conduct physical examinations and the perception that remote care could be "lesser" than in-person attention were recurring issues. Participants suggested virtual care might be suitable for routine or minor issues, but not for serious medical concerns.

Value of Physical Examination

Many participants expressed concern that virtual care could not substitute for hands-on assessment:

In my view...the Virtual Hospital will not be very effective if the patient is not present in front of the doctor...virtual care will be only communication. [P3, male, aged 40 - 49 years, 1 - 3 years in community]

My knee problem...I wasn't sure how I could show my knees over the phone? [P13, female, aged 60 - 69 years, <1 year in community]

Perceptions of "Lesser Care"

Some participants worried that virtual hospital services could be perceived as inferior, noting, "I just feel like if it's online like people would think like, oh, you're not taking me seriously" (P11, female, aged 19 - 29 years, >10 years in community) and "they wouldn't fully be supportive...they would think ohh I would receive better care at a hospital rather than me taking my medications on my own" (P15, female, aged 19 - 29 years, >10 years in community). Participants suggested virtual care is appropriate for routine or minor concerns, while serious conditions require in-person attention:

If it's a small thing, then Virtual Hospital is better. But if it's like a big issue, then she would prefer in-person care. [P4, female, aged 60 - 69 years, >10 years in community]

Emotional Safety, Privacy, and Trust

Overview

Emotional well-being, privacy, and trust emerged as key considerations. Participants emphasized the importance of empathy and interpersonal connection, while also noting concerns about privacy at home and the potential for fraud or security issues.

Need for Interpersonal Connection

Participants emphasized empathy and human touch, even via remote communication:

Having that interpersonal touch as much as you can...sometimes having that personal touch makes them feel better. But through a phone, you don't really get that personal touch, so still making sure whoever is on the line is caring and kind. [P1, male, aged 19 - 29 years, >10 years in community]

Concerns About Privacy at Home

Some expressed difficulties discussing sensitive issues at home:

If I want to tell you something, but I don't want to tell other people at home it could be a concern. So it's better to go to Doctor [in person]. [P10, female, aged 60 - 69 years, >10 years in community]

Trust in Fraser Health's Professionalism

Participants generally assumed privacy and security were managed by a reputable provider: "I would also assume because Fraser Health is a reputed professional organization, they would have already taken all of these requisite steps to make sure their privacy is covered" (P6, male, aged 40 - 49 years, >10 years in community), and "I feel like Fraser Health has really good system for privacy" (P11, female, aged 19 - 29 years, >10 years in community).

Apprehension About Fraud or Scams

Some participants noted the risk of unknown numbers or digital communication: "Only thing is sometimes when you get a call from an unknown number, it seems like it might be a fraud" (P7, male, aged 60 - 69 years, 1 - 3 years in community).

Barriers to Access and Technology

Despite interest in virtual care, participants noted practical barriers related to technology and digital literacy. These challenges were particularly salient for older adults and recent immigrants.

Digital Literacy Challenges

Older adults may lack experience with technology: "Maybe having like info sessions at temples. helping teach the elders about how it works" (P1, male, aged 19 - 29 years, >10 years in community), and "if I imagine my grandmother using this service, she might be overwhelmed with the tablet thing" (P16, female, aged 30 - 39 years, <1 year in community).

Access to Equipment

Not all households have necessary devices: "Not just like assuming like they have a laptop on them...my mom's mom has never touched a computer in her life" (P1, male, aged 19 - 29 years, >10 years in community).

Suggestions for Building Awareness of Virtual Hospitals at Home

Participants consistently highlighted the importance of proactive outreach and awareness-building, noting that many newcomers and elders were unaware of existing programs. Outreach strategies were rooted in community-based trust networks, emphasizing cultural familiarity, word of mouth, and visible engagement in South Asian gathering spaces.

Leveraging Trusted Spaces and Leaders

Participants recommended outreach through temples, gurdwaras, and mosques, where health authority representatives could host training or information sessions. They stressed that engagement

XSL•FO
RenderX

would be more effective if facilitated by respected community leaders.

> *...meeting with the community leaders and they can like bring it to the community. I know at one of my temples...they have community members teaching them...maybe having someone from Fraser Health going in and doing a training session.* [P6, male, aged 40 - 49 years, >10 years in community]

Testimonials and Peer Influence

Most participants expressed hesitation about being the first to try the new program, preferring to hear positive testimonials from others before committing. They also mentioned that even a few negative reviews could deter their interest in the program. This reflects the strong sense of mutual trust and reliance on community members, which participants valued just as much as official information provided by the health authority. Hearing positive testimonials from peers was seen as essential to building trust and encouraging adoption.

> *...having like some kind of testimonial or proven proof that it is already worked before...that would just remove the stress or the doubt. Once it becomes like accepted as a normal thing, that would make it easy.* [P7, male, aged 60 - 69 years, 1 - 3 years in community]

Community Events and Visible Outreach

To engage South Asian communities, participants recommended venues such as libraries, community centers, places of worship, and networking events where community members regularly gather. Additionally, cultural and religious events, such as the annual Vaisakhi Day Parade held in Surrey, were identified as valuable outreach opportunities. Participants suggested setting up information booths equipped with printed materials, visual media, and representatives from the health authority to disseminate information about virtual care in the languages spoken by the community. They noted that organizations providing employment or settlement support (such as PICS) already use these channels effectively, demonstrating their potential for outreach. Furthermore, participants recommended collaborating with local South Asian–based nonprofit organizations.

> *...we have like festivals of the year, you know Vaisakhi, Culture Fest...maybe having like a Fraser Health pop up so community members can engage with it and educate it that way. Trying to get in the community where you can would be the best.* [P6, male, aged 40 - 49 years, >10 years in community]

Ethnic Media and Social Platforms

To reach a broader demographic, participants suggested using local South Asian newspapers, radio, or TV networks for individuals to share their lived experiences. This would particularly target the working population who often listen to the radio while driving. Additionally, some participants recommended using social media, a popular platform for news consumption, to further raise awareness.

> *I think doing more awareness about this virtual...talking with the people, providing more awareness to the people that they should know that such kind of program is existed.* [P7, male, aged 60 - 69 years, 1 - 3 years in community]

Discussion

Principal Findings

This qualitative study explored the experiences of South Asian community members in the Fraser Health region with in-person hospital services, and their perceptions and expectations of Hospital-at-Home models. Participants described significant systemic and cultural barriers in traditional hospital settings, alongside gratitude for providers' skill and compassion. When discussing Hospital-at-Home, participants articulated both enthusiasm for its convenience and comfort and concerns about quality of care, cultural alignment, and feasibility in their home environments. Taken together, these findings provide important insights for health system planners seeking to advance culturally safe and equitable virtual hospital services.

Summary of Key Findings

Participants' hospital experiences reflected 6 key themes: (1) systemic barriers, including long wait times, overcrowding, transportation, and navigation challenges; (2) communication and language barriers; (3) unmet cultural and religious needs; (4) essential but burdensome family and caregiver roles; (5) experiences of discrimination and bias; and (6) positive encounters characterized by gratitude and respect. These issues varied across age, gender, and immigration status, with older and more recently arrived participants reporting more barriers.

Expectations of Hospital-at-Home revealed 5 major themes: (1) convenience, comfort, and family support; (2) cultural and generational differences; (3) quality of care and clinical limitations; (4) emotional safety, privacy, and trust; and (5) barriers to access and technology. Across both sets of findings, participants consistently highlighted the central role of family, the importance of cultural sensitivity, and the need for accessible information delivered through trusted community channels.

Participants' mixed experiences with in-person hospitals provide essential context for understanding perceptions of Hospital-at-Home. Structural challenges, such as prolonged emergency department waits and overcrowded wards, reduced trust in the timeliness and equity of hospital care. For newcomers, health system literacy gaps compounded these difficulties, as participants struggled to understand coverage, referral pathways, and navigation. These frustrations may increase openness to Hospital-at-Home as an alternative, particularly if it can reduce wait times and enhance comfort. However, systemic inequities that disadvantage newcomers and elders risk being reproduced in virtual formats unless addressed proactively.

Comparison to Existing Literature

Previous research has extensively documented barriers to virtual health care access for South Asians. Studies by Zibrik et al [18] and Makowsky et al [19] highlight disparities in virtual care

XSL•FO
RenderX

access, while Dahal et al [29] and Hyman et al [16] emphasize the impact of socioeconomic factors on virtual health engagement. Notably, few studies have examined Hospital-at-Home programs within other racial or ethnic minority communities, limiting opportunities for direct comparison. This scarcity of literature underscores the novelty of exploring Hospital-at-Home acceptability specifically among South Asian populations in Canada and highlights the unique structural, cultural, and familial considerations identified in this study.

Findings from this study align with existing literature in confirming that language barriers, trust in health care providers, and digital literacy shape perceptions toward adopting virtual care. Moreover, consistent with Bhalla et al [30], community-driven strategies, such as the role of family members in health decision-making and the influence of social networks such as WhatsApp, play a crucial role in health care engagement for the South Asian community. These findings were echoed by participants of this study who noted their attitudes toward Hospital-at-Home care could be highly influenced by peer testimonials.

A key distinction of this study is its focus on Hospital-at-Home models rather than primary care or outpatient virtual services, which have been the emphasis of much prior research. While previous research has explored South Asians' engagement with digital tools for chronic disease management and primary care, this study surfaces concerns that are specific to the delivery of hospital-level care in the home. In particular, it contributes new insights on the role of the home environment, including the physical suitability of the space, the availability and involvement of family caregivers, the complexities of intergenerational living arrangements, and the burden of household responsibilities during recovery from acute episodes.

These burdens often fall disproportionately on women, particularly mothers, reflecting traditional caregiving roles common in South Asian cultures that persist postimmigration. Female caregivers frequently manage both direct patient care and household duties, reinforcing the gendered division of labor described in previous literature [16]. Recognizing this, Hospital-at-Home programs should incorporate gender-sensitive supports, such as education, respite resources, and guidance for safe patient care, to mitigate additional burdens on women.

Another critical finding relates to the health care system education to improve virtual care access. Dahal et al [29] emphasize the role of community organizations in supporting newcomers through education on health care navigation. Participants in this study similarly reported limited awareness of services such as urgent and primary care centers and the 8-1-1 line, resulting in unnecessary reliance on emergency departments. This highlights that health system literacy is a foundational barrier: without addressing these basic knowledge gaps, new service models such as Hospital-at-Home risk reproducing existing inequities rather than improving access.

Chowdhury et al [31] identified concerns about virtual care creating distance between patients and providers, a theme also present in this study. Participants feared that reduced in-person interactions could diminish care quality and trust. This included concerns about developing weaker relationships with providers due to the impersonal nature of virtual interactions, which participants felt could compromise the depth of care and empathy received. This concern reinforces the importance of maintaining strong communication strategies within Hospital-at-Home models to ensure services remain patient-centered and do not further alienate those who experience barriers to care.

While age, gender, and immigration status emerged as individual factors shaping experiences with Hospital-at-Home, our findings underscore the importance of viewing these dimensions through an intersectional lens. For example, older newcomers with limited English proficiency described significant challenges with both language and digital literacy, creating compounded barriers to adoption that were distinct from those faced by younger, Canadian-born participants who reported greater comfort with system navigation and technology use. Similarly, caregiving responsibilities were disproportionately reported by women in multigenerational households, highlighting how gendered expectations intersect with cultural norms and family structures to intensify the caregiving burden. Recent immigrants, regardless of age, also described limited awareness of alternatives to the emergency department, illustrating how immigration status interacts with health system literacy to shape health care–seeking behaviors. These examples illustrate that barriers to Hospital-at-Home adoption are not experienced uniformly across South Asian communities but rather reflect the cumulative and intersecting effects of multiple social identities. Recognizing these layered experiences is critical for designing flexible, equity-oriented models that can adapt to the diverse needs within South Asian populations.

Implications for Hospital-at-Home Implementation

Despite substantial evidence demonstrating the effectiveness of Hospital-at-Home programs in improving patient outcomes and reducing health care costs, significant knowledge gaps remain regarding how to scale these services across diverse populations [32]. This study addresses this gap by exploring the adoption barriers and enablers of Hospital-at-Home services among South Asian communities in the Fraser Health region, offering insights for culturally tailored implementation strategies.

Increasing awareness of Hospital-at-Home services in ways that resonate with South Asian communities is essential to enhance willingness to participate. Findings from this study indicate that South Asian communities prefer learning about Hospital-at-Home through trusted channels such as community organizations, cultural events, and peer testimonials. Additionally, benefits such as increased privacy at home, proximity to family members, and greater control over aligning care with cultural and religious practices were identified as factors that enhance the appeal of Hospital-at-Home services for South Asian patients.

From a systems perspective, raising awareness about how Hospital-at-Home fits within the broader health care system is equally important. Educating patients and caregivers about the structure of health care services and how Hospital-at-Home complements traditional hospital care can build trust, increase

XSL•FO
RenderX

understanding of its role in acute care delivery, and foster confidence in available escalation pathways. This system-level approach can help patients see Hospital-at-Home as an integral part of a coordinated healthcare continuum rather than as a stand-alone service.

Addressing individual barriers is also critical for sustainable and equitable adoption. Limited digital literacy skills among South Asian populations can be mitigated through multilingual education and training strategies that accommodate diverse learning styles. Designing user-friendly digital interfaces for Hospital-at-Home services is also critical to ensure accessibility. Furthermore, the study highlights the importance of explicitly considering family members and caregivers in Hospital-at-Home implementation. This includes developing caregiver-focused training materials, addressing family dynamics in care planning, and regularly evaluating caregiver experiences.

This study further emphasizes that preferences and attitudes toward Hospital-at-Home services are not uniform across South Asian communities, nor do they universally apply to all patient needs. For instance, while some participants appreciated the comfort and cultural familiarity of home-based care, others voiced significant concerns around privacy, data security, and the adequacy of clinical care delivered virtually. These diverging views underscore the need for flexible, patient-centered models that allow for individual choice.

Housing environments emerged as an important contextual factor influencing the feasibility of Hospital-at-Home care. Participants described challenges such as stairs, limited space for medical equipment, and crowded or multigenerational households, all of which could constrain the safety and acceptability of receiving acute care at home. These findings point to the need to consider how physical living environments intersect with the delivery of virtual hospital models. While a detailed policy analysis is beyond the scope of this study, it is important to acknowledge that housing affordability and adequacy are pressing issues in Canada, particularly in urban centers such as Surrey. Future research should explore how these broader structural determinants, including housing conditions, shape equitable access to home-based care models and whether supportive housing or community infrastructure can mitigate such barriers.

To ensure equitable and effective implementation, the design of Hospital-at-Home programs must incorporate sufficient resources to continuously engage diverse patient, provider, and community voices. This ongoing feedback process is critical for adapting services to reflect the cultural nuances, varied needs, and evolving preferences of South Asian populations, enabling the development of culturally informed strategies that foster sustained adoption and trust in Hospital-at-Home services. This study's qualitative findings informed a subsequent Experience-Based Co-Design initiative, which actively engaged South Asian community members in the cocreation of culturally responsive solutions to the challenges identified. Specific outcomes of this process included the development of culturally adapted training materials, multilingual communication tools, and revised care workflows to strengthen trust and clarity in care delivery. By linking this foundational research with participatory design methods, the broader project not only addressed knowledge gaps but also demonstrated a model for translational research, showing how community-engaged approaches can directly inform the design of culturally safe virtual care innovations.

Policy and System-Level Considerations

The findings from this study highlight the need for equity-focused policies and system-level interventions to enhance virtual care accessibility for South Asian communities in Fraser Health and beyond. A central implication is that health system literacy must be addressed as a prerequisite to virtual adoption; simply introducing Hospital-at-Home services without parallel education on how and when to use them may not improve access. South Asian patients may face barriers due to limited English proficiency and a lack of access to culturally competent care [33]. Policies should mandate the integration of multilingual virtual health services, including real-time interpretation and culturally responsive health communication strategies. This aligns with Fraser Health's ongoing efforts to improve accessibility through initiatives such as the SAHI.

Another critical policy recommendation is the standardization of digital literacy support within virtual care services. The findings of this study identified the problematic impact of technological barriers, particularly among older adults and individuals unfamiliar with digital tools. Policies should ensure that virtual care platforms are user-friendly and that comprehensive digital literacy training programs are offered as part of virtual hospital enrollment. This could include step-by-step tutorials, one-on-one technical assistance, and hybrid learning options that combine online and in-person training [34]. In addition, although only a minority of participants mentioned it, data privacy and security remain essential considerations in the design of virtual hospital services. Policies should ensure that multilingual and culturally tailored digital platforms also meet high standards of technological ethics, including secure data management and patient confidentiality, to build and maintain trust among diverse communities.

System-level reforms should focus on streamlining care pathways for virtual hospital patients. Participants expressed concerns about delays in care escalation and emergency response times. To address this, policies should define clear transition protocols between virtual and in-person hospital care, ensuring that patients requiring higher levels of medical attention can be seamlessly transferred. Emergency response times can be optimized through virtual triage systems to ensure virtual hospital patients receive timely interventions when needed [35] .

An additional consideration is the inclusion of family caregivers in virtual health planning. Individuals in South Asian households rely on family members and loved ones for health care navigation and support. Virtual care policies should recognize and integrate the role of caregivers, facilitating access to training, resources, and decision-making tools to support patient care effectively.

XSL•FO
RenderX

Finally, data-driven policy improvements are essential for ensuring that virtual care models address the specific needs of South Asian communities. These findings of this study highlight the importance of ongoing patient engagement and feedback mechanisms. There is a clear need to implement structured monitoring and evaluation frameworks that assess adoption, patient experiences, health outcomes, and engagement levels in virtual care programs. These insights should inform policy refinements to ensure equitable and effective virtual health services.

Strengths and Limitations

This study provides novel insights into Hospital-at-Home acceptability among South Asian communities in Canada, a population underrepresented in Hospital-at-Home research. Semistructured interviews in multiple languages enabled diverse perspectives across gender, age, and immigration status. Importantly, the study included voices of marginalized groups—such as elders, recent immigrants, and family caregivers—who are often excluded from virtual health research.

While this study provides valuable insights into the experiences of South Asian communities with virtual health care, several limitations should be noted. Recruitment challenges limited geographic diversity, with most participants residing in Surrey. As access to health services and digital infrastructure varies across municipalities, future studies should aim for broader regional representation to capture a wider range of experiences.

Another limitation is the reliance on referrals from PICS, which may have introduced selection bias. Participants recruited through PICS are likely to have stronger community connections and greater awareness of health resources, leading to a sample more engaged with the health care system than the broader South Asian population. This may result in an underrepresentation of those facing greater systemic and socioeconomic barriers. Nonetheless, the inclusion of participants with limited English proficiency, lower digital literacy, and newcomer status partially mitigates this concern by ensuring that marginalized perspectives were represented.

The reliance on online interviews is another limitation. While virtual interviews increased accessibility for some, they may have excluded participants uncomfortable with digital technologies or lacking reliable internet access, potentially missing the perspectives of those most affected by digital exclusion. Future research should incorporate alternative data collection methods, such as in-person or telephone interviews, to ensure more inclusive participation.

Despite these limitations, this study offers valuable insights into the structural and cultural barriers influencing virtual health adoption among South Asian communities. Future research should use diverse recruitment strategies and mixed methods approaches to enhance sample representativeness and ensure that equity-driven virtual care models are more inclusive of underserved populations.

Conclusion

This study highlights the complex and context-specific factors that influence South Asian communities' engagement with Hospital-at-Home and virtual care services. Barriers such as limited digital literacy, language challenges, caregiver burden, and lack of awareness about health care alternatives must be addressed through culturally and linguistically responsive strategies. At the same time, key facilitators such as trust in health care providers, the comfort of home-based care, and alignment with cultural and familial values offer important starting points for building more inclusive models of care. Ensuring the equitable implementation of Hospital-at-Home services will require sustained investment in community-specific education, caregiver inclusion, and ongoing feedback mechanisms that reflect the diversity within South Asian populations.

Acknowledgments

The authors sincerely thank the South Asian community members who participated in this study and shared their experiences, as well as the Steering Committee for their guidance and support throughout the project.

Funding

This work was supported by Health Research BC and the Health Research Foundation of Innovative Medicines Canada Team Grant in Virtual Care. The funders had no role in the study design, data collection, analysis, interpretation, writing of the manuscript, or the decision to submit the paper for publication.

Data Availability

The deidentified interview transcripts and other data generated or analyzed during this study are available from the corresponding author on reasonable request, subject to ethical approval and participant confidentiality agreements. Due to the sensitive nature of the data and the small community context, data cannot be publicly shared.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full interview guide.

XSL•FO
RenderX

[DOCX File, 17 KB - humanfactors_v13i1e79675_app1.docx]

Checklist 1
COREQ checklist.
[PDF File, 1171 KB - humanfactors_v13i1e79675_app2.pdf]

References

1. Census profile, 2016 census. Statistics Canada. URL: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E [accessed 2025-04-16]
2. Bainey KR, Jugdutt BI. Increased burden of coronary artery disease in South-Asians living in North America. Need for an aggressive management algorithm. Atherosclerosis 2009 May;204(1):1-10. [doi: 10.1016/j.atherosclerosis.2008.09.023] [Medline: 18980768]
3. Rana A, de Souza RJ, Kandasamy S, Lear SA, Anand SS. Cardiovascular risk among South Asians living in Canada: a systematic review and meta-analysis. CMAJ Open 2014 Jul;2(3):E183-E191. [doi: 10.9778/cmajo.20130064] [Medline: 25295238]
4. Palaniappan L, Garg A, Enas E, et al. South Asian cardiovascular disease & cancer risk: genetics & pathophysiology. J Community Health 2018 Dec;43(6):1100-1114. [doi: 10.1007/s10900-018-0527-8] [Medline: 29948525]
5. South Asian health report. Fraser Health. 2015. URL: https://www.fraserhealth.ca/-/media/Project/FraserHealth/FraserHealth/Health-Topics/South-Asian-Health/201606_South_Asian_Health_Report.pdf [accessed 2025-10-14]
6. Islam F, Khanlou N, Tamim H. South Asian populations in Canada: migration and mental health. BMC Psychiatry 2014 May 26;14:154. [doi: 10.1186/1471-244X-14-154] [Medline: 24884792]
7. Godley J. Everyday discrimination in Canada. Can J Sociol 2018;43(2):111-142. [doi: 10.29173/cjs29346]
8. van Draanen J, Shafique A, Farissi A, et al. How to offer culturally relevant type 2 diabetes screening: lessons learned from the South asian diabetes prevention program. Can J Diabetes 2014 Oct;38(5):329-333. [doi: 10.1016/j.jcjd.2013.11.008] [Medline: 24797496]
9. Jennett PA, Affleck Hall L, Hailey D, et al. The socio-economic impact of telehealth: a systematic review. J Telemed Telecare 2003;9(6):311-320. [doi: 10.1258/135763303771005207] [Medline: 14680514]
10. Melman A, Maher CG, Machado GC. Virtual hospitals: why we need them, how they work and what might come next. J Physiother 2021 Jul;67(3):156-157. [doi: 10.1016/j.jphys.2021.06.018] [Medline: 34148815]
11. Ryan PP, Hawkins KL, Altman S, et al. A novel virtual hospital at home model during the coronavirus disease 2019 (COVID-19) pandemic. Infect Control Hosp Epidemiol 2021 Sep;42(9):1140-1142. [doi: 10.1017/ice.2020.435] [Medline: 32829735]
12. Shepperd S, Doll H, Angus RM, et al. Admission avoidance hospital at home. Cochrane Database Syst Rev 2008 Oct 8(4):CD007491. [doi: 10.1002/14651858.CD007491] [Medline: 18843751]
13. Chua CMS, Ko SQ, Lai YF, Lim YW, Shorey S. Perceptions of hospital-at-home among stakeholders: a meta-synthesis. J Gen Intern Med 2022 Feb;37(3):637-650. [doi: 10.1007/s11606-021-07065-0] [Medline: 34363185]
14. Patel HY, West DJ. Hospital at Home: an evolving model for comprehensive healthcare. Glob J Qual Saf Healthc 2021 Nov;4(4):141-146. [doi: 10.36401/JQSH-21-4] [Medline: 37261225]
15. McIntyre D, Chow CK. Waiting time as an indicator for health services under strain: a narrative review. Inquiry 2020;57:46958020910305. [doi: 10.1177/0046958020910305] [Medline: 32349581]
16. Hyman A, Stacy E, Mohsin H, et al. Barriers and facilitators to accessing digital health tools faced by South Asian Canadians in Surrey, British Columbia: community-based participatory action exploration using photovoice. J Med Internet Res 2022 Jan 13;24(1):e25863. [doi: 10.2196/25863] [Medline: 35023842]
17. Jong M, Mendez I, Jong R. Enhancing access to care in northern rural communities via telehealth. Int J Circumpolar Health 2019;78(2):1554174. [doi: 10.1080/22423982.2018.1554174] [Medline: 31066652]
18. Zibrik L, Khan S, Bangar N, Stacy E, Novak Lauscher H, Ho K. Patient and community centered eHealth: exploring eHealth barriers and facilitators for chronic disease self-management within British Columbia's immigrant Chinese and Punjabi seniors. Health Policy Technol 2015 Dec;4(4):348-356. [doi: 10.1016/j.hlpt.2015.08.002]
19. Makowsky MJ, Jones CA, Davachi S. Prevalence and predictors of health-related internet and digital device use in a sample of South Asian Adults in Edmonton, Alberta, Canada: results from a 2014 community-based survey. JMIR Public Health Surveill 2021 Jan 8;7(1):e20671. [doi: 10.2196/20671] [Medline: 33416506]
20. Health authorities. Government of British Columbia. URL: https://www2.gov.bc.ca/gov/content/health/about-bc-s-health-care-system/partners/health-authorities [accessed 2025-04-07]
21. About Fraser Health. Fraser Health. URL: https://www.fraserhealth.ca/about-us/about-fraser-health [accessed 2025-04-07]
22. Census profile, 2021 census of population. Statisics Canada. 2021. URL: https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/index.cfm?Lang=E [accessed 2025-04-07]
23. PICS. URL: https://pics.bc.ca/ [accessed 2025-04-24]
24. Virtual health. Fraser Health. URL: https://www.fraserhealth.ca/patients-and-visitors/virtual-health [accessed 2025-04-07]

XSL•FO
RenderX

25. South Asian Health Institute (SAHI). Fraser Health. URL: https://www.fraserhealth.ca/health-topics-a-to-z/south-asian-health/south-asian-health-institute [accessed 2025-04-24]
26. Equity, Diversity and Inclusion (EDI). Fraser Health. URL: https://www.fraserhealth.ca/health-topics-a-to-z/EDI [accessed 2025-04-24]
27. Dewey J. Context and Thought, Volume 12: University of California Publications in Philosophy; 1931.
28. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
29. Dahal R, Naidu J, Bajgain BB, et al. Patient-identified solutions to primary care access barriers in Canada: the viewpoints of Nepalese immigrant community members. J Prim Care Community Health 2022;13:21501319221141797. [doi: 10.1177/21501319221141797] [Medline: 36476102]
30. Bhalla M, Boutros H, Meyer SB. Aunties, WhatsApp, and “haldi da doodh”: South Asian communities’ perspectives on improving COVID-19 public health communication in Ontario, Canada. Can J Public Health 2022 Dec;113(Suppl 1):46-53. [doi: 10.17269/s41997-022-00712-x] [Medline: 36449223]
31. Chowdhury D, Tong C, Lopez K, Neiterman E, Stolee P. “When in Rome…”: structural determinants impacting healthcare access, health outcomes, and well-being of South Asian older adults in Ontario using a multilingual qualitative approach. Front Public Health 2024;12:1405851. [doi: 10.3389/fpubh.2024.1405851] [Medline: 39741940]
32. Lai YF, Ko SQ. Time to shift the research agenda for Hospital at Home from effectiveness to implementation. Cochrane Database Syst Rev 2024 Mar 5;3(3):ED000165. [doi: 10.1002/14651858.ED000165] [Medline: 38441225]
33. Budhwani S, Fujioka J, Thomas-Jacques T, et al. Challenges and strategies for promoting health equity in virtual care: findings and policy directions from a scoping review of reviews. J Am Med Inform Assoc 2022 Apr 13;29(5):990-999. [doi: 10.1093/jamia/ocac022] [Medline: 35187571]
34. Donetto S, Pierri P, Tsianakas V, Robert G. Experience-based co-design and healthcare improvement: realizing participatory design in the public sector. Des J 2015 Jun;18(2):227-248. [doi: 10.2752/175630615X14212498964312]
35. Gellert GA, Rasławska-Socha J, Marcjasz N, et al. How virtual triage can improve patient experience and satisfaction: a narrative review and look forward. Telemed Rep 2023;4(1):292-306. [doi: 10.1089/tmr.2023.0037] [Medline: 37817871]

Abbreviations

PICS: Progressive Intercultural Community Services
SAHI: South Asian Health Institute

Edited by A Kushniruk; submitted 25.Jun.2025; peer-reviewed by G Wang, M Nisar; revised version received 23.Sep.2025; accepted 16.Oct.2025; published 05.Jan.2026.

Please cite as:
Wong E, Kandola M, Arora K, Sharda H, Deol R, Jung M, Paquin R, Montenegro M, MacPherson M
Hospital-at-Home for South Asian Communities in British Columbia, Canada: Qualitative Interview Study
JMIR Hum Factors 2026;13:e79675
URL: https://humanfactors.jmir.org/2026/1/e79675
doi:10.2196/79675

© Emma Wong, Mahabhir Kandola, Kamal Arora, Harroop Sharda, Roman Deol, Mary Jung, Robert Paquin, Maria Montenegro, Megan MacPherson. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 5.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Bridging Gaps in Women's Heart Health: User-Centered Needs Assessment Informed by Patient and Clinician Interviews

Christine Jacob[1], PhD; Sangeetha-Rose Puthanveettil[2], MSc; Patrick Vavken[4,5,3*], MD, PD, MSc, MBA; Emel Kaplan[6*], MD, PhD; Christine S Zuern[6*], MD, Prof Dr

[1]Institute for Information Systems, University of Applied Sciences Northwestern Switzerland (FHNW), Windisch, Aargau, Switzerland

[2]University of Applied Sciences Northwestern Switzerland FHNW, Windisch, Switzerland

[3]Vavken Health Labs, Zurich, Switzerland

[4]ETH, Zurich, Switzerland

[5]University of St. Gallen, St. Gallen, Switzerland

[6]Cardiovascular Research Institute Basel (CRIB) and Department of Cardiology, University Hospital Basel, University of Basel, Basel, Switzerland

[*]these authors contributed equally

Corresponding Author:
Christine Jacob, PhD
Institute for Information Systems
University of Applied Sciences Northwestern Switzerland (FHNW)
Bahnhofstrasse 6
Windisch, Aargau, 5210
Switzerland
Phone: 41 56 202 74 64
Email: christine.k.jacob@gmail.com

Abstract

Background: Women with cardiovascular disease (CVD) remain underserved due to gaps in recognition, diagnosis, and care tailored to sex-specific risks. Digital health tools have the potential to address these inequities, but many fail to reflect the distinct needs of women. In a prior review, we assessed 20 CVD apps and 22 wearables and found that only 25% (5/20) of apps and 40% (9/22) of wearables included any sex-specific content, such as hormone cycle tracking and life-stage considerations related to pregnancy or menopause. These findings confirm that current digital tools largely mirror the gender gaps seen in traditional care.

Objective: This study aimed to define the user requirements for a CVD app designed specifically for women. We sought to explore the unmet needs and challenges faced by female patients and their clinicians that current tools fail to address, and also to identify and prioritize features that would be most valuable and feasible to implement.

Methods: We conducted a qualitative study using semistructured interviews to explore the needs, preferences, and expectations of women living with CVD and their treating clinicians. Guided by the human-centered design framework, this work focused on the "Define" phase. A total of 20 participants in Switzerland were interviewed, including 11 women with CVD, 7 cardiologists, and 2 experts in regulatory and reimbursement. Participants were recruited through purposive sampling, and interviews were conducted online between April and July 2025. Thematic analysis was used to synthesize the data, highlighting design priorities and contextual factors relevant for developing a patient-centered and system-aware digital health tool.

Results: The interviews with women living with CVD and cardiologists confirmed the consistent gaps between existing care pathways and the specific needs of female patients. Both groups highlighted the lack of early symptom recognition, insufficient sex-specific guidance, and limited tools tailored to women's lived experience. While patients prioritized personalized education, emotional support, and features that address hormonal and life-stage–specific risks, clinicians emphasized clinical use, workload integration, and actionable summaries. Success was defined experientially by patients (eg, empowerment and reduced anxiety), and operationally by clinicians (eg, earlier detection and improved adherence). Willingness to pay was moderate among both groups, with patients favoring simplicity and clinicians emphasizing workflow integration and proven clinical use.

Conclusions: These findings highlight the importance of designing an artificial intelligence–enabled CVD app for women that meaningfully integrates patient empowerment with clinical workflows. A dual-value approach is essential, offering personalized tools that address emotional and lifestyle needs for patients, while supporting clinicians with concise, actionable insights. Early

XSL•FO
RenderX

reflections on regulatory and reimbursement considerations suggest that a modular, evidence-based rollout strategy would be key for long-term adoption and scale.

(JMIR Hum Factors 2026;13:e82916) doi:10.2196/82916

KEYWORDS

eHealth; mobile health; mHealth; digital health; user research; technology adoption; technology implementation; qualitative study; usability; cardiovascular disease; women's health; human centered design

Introduction

Background

Cardiovascular disease (CVD) remains the world's leading cause of death, claiming nearly 18 million lives each year and impacting more than half a billion people globally [1-3]. Despite decades of progress in cardiology, these numbers underscore a persistent challenge: our current approaches to prevention, diagnosis, and treatment are not reaching everyone equally [3]. One critical blind spot is the under-recognition of sex-based physiological differences in cardiovascular health [4]. Women, in particular, are often underdiagnosed, undertreated, and underserved, partly due to atypical symptom presentation and a longstanding male-centric model of research and care [4]. This has serious consequences, including delays in diagnosis and poorer outcomes for women across the CVD continuum [5].

Digital health technologies offer a promising avenue to change this narrative. Mobile health (mHealth) tools, remote patient monitoring, wearable sensors, and artificial intelligence (AI)–driven decision support systems are reshaping how individuals manage their cardiovascular risk, bringing prevention, detection, and self-care into everyday life [1]. According to the World Health Organization, mHealth refers broadly to health care and public health services supported by mobile devices such as smartphones, wearables, and wireless sensors [6]. These tools can improve access to timely care, enable more tailored interventions, and potentially reduce long-term costs to the health system [7].

Evidence is steadily accumulating in favor of mHealth for cardiovascular care. For example, a systematic review by Coorey et al [8] found that mobile apps can support better blood pressure (BP) control, encourage healthy dietary habits, and reduce hospital readmissions for patients with CVD. Smartphone-based photoplethysmography, a technology that measures changes in blood volume using infrared light, has shown promise in detecting atrial fibrillation (AF) and assessing heart rate (HR) variability, offering an accessible, noninvasive, and scalable solution for early risk detection [9].

Wearable devices are also gaining ground as tools for continuous cardiovascular monitoring. Positioned on the wrist, chest, or hip, these devices can monitor HR, blood oxygen levels, sleep patterns, and physical activity using either photoplethysmography or ECG technology [10,11]. For instance, a study by Guo et al [12] involving more than 187,000 users identified more than 260,000 potential AF episodes, with confirmatory testing validating the diagnosis in most cases. A broader review of smartwatch-based interventions echoed these findings, highlighting improvements in lifestyle behaviors, medication adherence, AF detection, and reductions in unplanned hospitalizations [13].

Current State of CVD Apps and Wearables

Building on the well-documented unmet needs in CVD for women, we evaluated how effectively existing digital health tools, specifically mobile apps and wearable devices, address sex-specific factors in CVD. To do this, we conducted a structured review of 20 patient-facing CVD apps and 22 commercially available wearables. Each tool was assessed using the foundational and contextual dimensions of the sociotechnical framework for evaluating patient-facing eHealth interventions [14], which emphasizes both technical functionality and integration into real-world health contexts.

The results of this assessment, published in a separate study [15], revealed a significant gap: only 25% (5/20) of the reviewed apps and 40% (9/22) of the wearables incorporated sex-specific content. This included considerations such as the impact of hormonal changes, menopause, or pregnancy on cardiovascular health, factors known to influence symptom presentation, disease progression, and treatment needs in women. These findings reinforce the conclusion that digital health tools are not exempt from the systemic gaps that characterize traditional CVD care pathways [15]. Rather, they mirror the underrepresentation of women's needs in CVD. Addressing this gap is critical if we aim to develop digital interventions that are both equitable and clinically effective for women living with CVD.

Objectives

To address this gap, this study explored the user needs and requirements for a digital health app designed specifically for women with CVD. The primary objectives were (1) to understand the specific challenges and unmet needs that female patients and health care professionals encounter in cardiovascular care that are not adequately addressed by current digital tools, and (2) to identify and prioritize the features and functionalities that a new app should incorporate to better support sex-specific cardiovascular health management. Furthermore, the study did not only focus on the preferences and expectations of the primary users, such as patients and clinicians, but also considered broader system-level enablers and constraints, including clinical integration, regulatory requirements, and reimbursement considerations, to guide a practical and scalable development strategy.

XSL•FO
RenderX

Methods

Overview

We adopted a qualitative research approach to explore the nuanced needs and expectations of both patients and clinicians. This methodology was chosen for its strength in capturing complex, context-dependent experiences and sociocultural factors that are often overlooked by quantitative methods [16].

Qualitative methods are increasingly used in health services and technology research, as they allow researchers to uncover the "why" behind user behaviors and preferences [17]. In our case, this approach helped surface the specific ways in which digital tools can support women across the CVD journey, and how clinicians view their potential integration into care pathways. These insights provide a strong foundation for user-centered feature development and iterative design.

Scope and Conceptual Framework

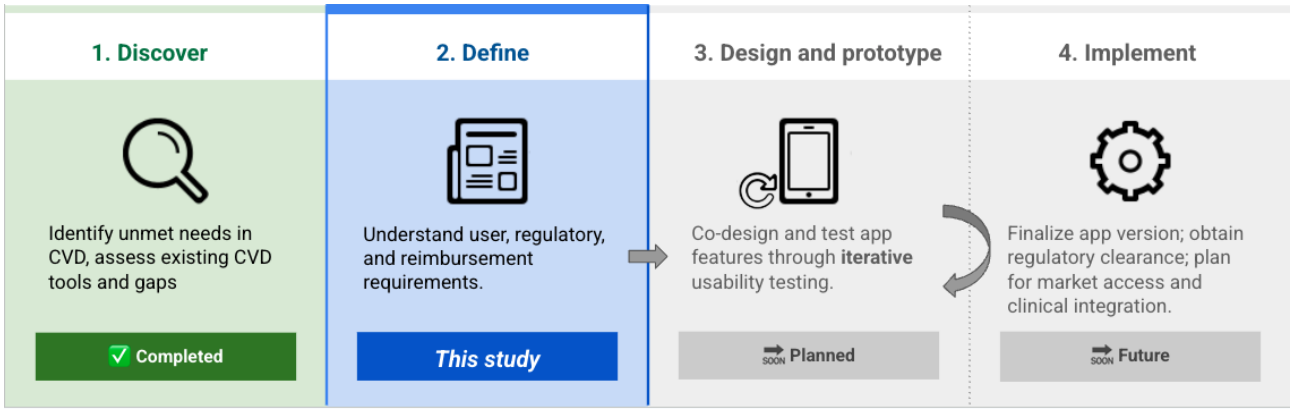
This study was guided by the human-centered design (HCD) framework, which places the needs, experiences, and preferences of end users, here, women living with CVD and the clinicians who care for them, at the core of innovation [18-20]. HCD is a structured, iterative methodology that unfolds across 4 key phases: discover, define, design and prototype, and implement. Each phase plays a critical role in ensuring that health technologies are developed not only for users, but with them [21,22].

The discover phase, which focuses on identifying the problem space and understanding unmet needs, was addressed in our previously published study that assessed the extent to which existing CVD apps and wearables account for sex-specific considerations [15]. That foundational work revealed significant gaps in digital tools for women with CVD, particularly around life-stage–specific guidance and personalized risk tracking [15].

Building on those findings, this study focused on the define phase. This stage aimed to deepen understanding of user needs through qualitative inquiry, synthesize priorities across user groups, and translate insights into clear design criteria. Specifically, we investigated the expectations, preferences, and contextual considerations of both female patients and clinicians to guide the conceptualization of a CVD app tailored to women's unique health trajectories.

The subsequent design and prototype and implementation phases, where iterative development, user testing, and real-world deployment take place, are beyond the scope of this study but will build directly on the requirements defined in this study. Figure 1 illustrates the HCD process guiding this study, with this study focusing specifically on the define phase, following earlier gap analysis in the discover phase.

Figure 1. Human-centered design approach for developing an app for women with cardiovascular disease. CVD: cardiovascular disease.



Sampling Strategy and Participant Recruitment

We used purposive sampling, a commonly used strategy in qualitative research aimed at capturing rich, experience-based insights from relevant stakeholders [23]. Patient participants were recruited in collaboration with the Women's Heart Health Program at the University Hospital Basel, a specialized cardiology outpatient program focusing on women's cardiovascular health. Inclusion criteria required participants to be women aged 18 years or older, have a confirmed diagnosis of CVD, access to email and the internet, and be comfortable using a teleconferencing tool.

Eligible patients were approached directly by the clinical staff in the Women's Heart Health Program, who explained the study and obtained written informed consent. Once consent was given, the signed forms and contact details were forwarded to the core research team at the University of Applied Sciences and Arts Northwestern Switzerland, who subsequently managed the coordination and communication with participants.

Cardiologists were recruited independently by the research team through targeted online searches and professional outreach via email. The participating cardiologists were recruited from 5 different hospitals and clinics, reflecting 3 care settings across Switzerland, including university hospitals, cantonal hospitals, and private practices, offering a diverse clinical perspective. And unlike clinical or patient experience, which can vary widely, regulatory and reimbursement requirements are governed by defined legal frameworks. Therefore, expert input was sought from two highly specialized professionals with national and EU (European Union)-level expertise, who provided focused guidance on certification and financing pathways for digital health tools in Switzerland and Europe.

All participants received study information and consent materials in both English and German and were given the option to choose

XSL•FO
RenderX

their preferred language. Interview duration ranged from 28 to 63 minutes, depending on the depth of participant responses. Recruitment took place between April and June 2025 and continued until thematic saturation was reached, that is, when no new themes emerged from additional interviews [23,24]. This was assessed continuously during data collection and analysis. For both patient and clinician interviews, saturation was monitored separately. We noted that by the tenth patient and sixth clinician interview, no substantially new themes were identified, and the final interviews largely confirmed and elaborated on existing patterns. The English version of the participant information sheet and consent form is included in Multimedia Appendix 1.

A total of 20 participants based in Switzerland were recruited: 11 women living with CVD, 7 cardiologists, 1 regulatory expert, and 1 reimbursement expert. The sample characteristics of the participating patients and clinicians are summarized in Multimedia Appendix 2.

Data Collection and Synthesis

Data were collected through in-depth, semistructured interviews conducted online between April and July 2025. In total, 2 tailored interview guides were developed: 1 focused on the patient experience and day-to-day disease management, and 1 directed at clinicians, emphasizing workflow, system integration, and implementation considerations. Both guides are provided in Multimedia Appendix 3.

A structured thematic analysis was conducted following the framework by Braun and Clarke [25,26], including familiarization with the data, initial coding, theme development, review, refinement, and final synthesis (see Multimedia Appendix 4 for detailed steps). We used a hybrid coding approach, combining deductive codes informed by prior work and study objectives, with inductive codes that emerged from the data, as described in Multimedia Appendix 4. Codes were grouped into broader thematic categories and, where relevant, mapped to stages of the patient journey (eg, early diagnosis support, treatment, and long-term management). Coding was supported using NVivo (Lumivero) qualitative data analysis software. Patient and clinician interviews were conducted and primarily analyzed by SP, while CJ and PV led the interviews with the regulatory and reimbursement experts. CJ further refined the thematic analysis, synthesis, and reporting. To mitigate potential researcher bias, coding was triangulated across the 3 coauthors (SRP, CJ, and PV), and discrepancies were resolved through iterative discussion. Any coding discrepancies between SP and CJ were addressed through discussions with PV until a consensus was reached. This process spanned from May to August 2025.

Ethical Considerations

The Ethics Committee of Northwest and Central Switzerland determined that ethics approval was not needed for this study, according to the Federal Act on Research involving Human Beings, article 2, paragraph 1 (reference number Req-2025-00491). All participants were briefed about the research background and signed a consent form agreeing to participate. Participant data were anonymized, access to identifiable source files was restricted to the first and second authors, and no participants received financial compensation, with the exception of the regulatory and market access experts who received consulting fees for their professional advisory contributions.

Results

Understanding the User Journey and Unmet Needs

The patient interviews revealed a consistent pattern of late recognition, emotional burden, and lack of gender-sensitive care in the management of CVD. Many women described experiencing a delayed or unclear diagnosis, often attributed to atypical symptom presentation and a lack of accessible, sex-specific information. Many had to self-educate or navigate their condition alone, especially at the outset, before a formal diagnosis was made. Emotional distress, stress, and fear, especially post diagnosis, were prominent themes. Participants spoke about the emotional toll that followed their diagnosis, including feelings of shock, sadness, fear, and uncertainty about the future. For some, the psychological burden was compounded by life-altering restrictions, such as being advised against future pregnancies or having to give up previously enjoyed activities like vigorous exercise.

There was a widespread perception among the participants that hormonal and life-stage influences (eg, pregnancy and menopause) are underacknowledged in treatment pathways. Only 18% (2/11) of patients received any sex-specific guidance, mostly about the potential dangers of getting pregnant with heart disease. Participating patients stated that they generally struggle with lifestyle adjustments, medication adherence, and managing comorbidities. When discussing the complexity of symptom management, participants frequently described difficulties in understanding medical instructions and uncertainty about how to respond to different symptoms or side effects. This was further complicated by external factors such as stress, an acknowledged trigger that patients knew could worsen their condition, yet felt largely powerless to control.

Digital tools are underused; only 18% (2/11) of patients had experience using CVD apps or wearables, for example, for ECG measurement, while the others expressed confusion, lack of awareness, or anxiety about using them. All participating patients agreed that gender-specific tracking would be beneficial. Table 1 highlights the main challenges and unmet needs faced by women with CVD along the care continuum, their current use of digital health tools, and how many participants referenced each theme in the interviews.

XSL•FO
RenderX

Table 1. Understanding the patient journey and unmet needs (n=11).

Themes and subthemes	Prevalence, n (%)
Early symptoms and diagnosis	
Unfamiliar with the disease at diagnosis	7 (64)
Atypical or no symptoms prior to diagnosis	6 (55)
Self-initiated education (eg, googling and support groups)	5 (45)
Emotional distress after diagnosis (shock, sadness, and fear)	6 (55)
Stress and psychological burden of the disease	5 (45)
CVD[a] perceived as a burden or life-changing event	6 (55)
Life stage and gender-specific care	
Not informed or addressed by HCPs[b]	9 (82)
Was informed about the impact of life stages (eg, pregnancy and menopause)	2 (18)
Perceived treatment as male-centered	4 (36)
Belief that women's CVD symptoms are often dismissed	3 (27)
Ongoing disease management challenges	
Difficulty with lifestyle and diet changes	6 (55)
Struggles with medication adherence or side effects	5 (55)
Difficulty understanding symptoms and instructions	4 (36)
Difficulty managing comorbidities	3 (27)
Anxiety before appointments because of uncertainties (eg, on potential progression)	2 (18)
Dealing with constant emotional overwhelm and fatigue	3 (27)
Use of existing digital tools	
Has not used any apps for CVD	9 (82)
Not aware of any helpful CVD apps	7 (64)
Prefers face-to-face care over digital tools	4 (36)
Feels overwhelmed or anxious by the idea of apps	3 (27)
Has used a step counter, hydration reminder, or menstrual tracker	3 (27)
Has used a CVD-related app or wearable (eg, for ECG[c] measurement)	2 (18)

[a]CVD: cardiovascular disease.

[b]HCP: health care professional.

[c]ECG: electrocardiogram.

Clinicians reported several diagnostic challenges stemming from atypical symptom presentation in women, compounded by limited awareness and insufficient referral pathways. They emphasized the mismatch between traditional risk models and real-world female presentations, which are often underrecognized or misattributed. Time constraints and health care system overload further exacerbate diagnostic delays. Despite acknowledging the lack of tools to capture hormonal or life-stage influences, 5 of 7 (71%) participating clinicians reported that they continue to follow standard monitoring protocols without sex-specific adjustments.

Only 1 of 7 (17%) interviewed cardiologists recommended CVD apps, while 6 of 7 (86%) cited lack of familiarity or trust in their use, with usability for older patients as the main concern. Integration of digital tools into clinical workflows remains controversial; some see potential for structured summaries and AI-assisted risk insights, while others worry about time burden or data security. There was strong agreement on the need for decision support tools that surface earlier warnings during high-risk windows such as menopause, pregnancy, or postevent follow-ups, with 5 of 7 (71%) participants citing this as an existing gap. Table 2 presents the key challenges and unmet needs encountered by clinicians throughout the clinical workflow, alongside their current engagement with digital health tools, and how many participants referenced each theme in the interviews.

XSL•FO
RenderX

Table 2. Understanding the clinician workflow and unmet needs (n=7).

Themes and subthemes	Prevalence, n (%)
Diagnostic challenges	
Atypical symptoms in women make the diagnosis more challenging	7 (100)
Lack of awareness and, accordingly, referrals	5 (71)
Limited time and capacity in care	3 (43)
Sex-specific gaps	
Symptoms misattributed or dismissed	7 (100)
Treatment paths not sex-sensitive and mostly male-focused	5 (71)
Monitoring practices	
Standardized tests, no sex differences	5 (71)
Limited tools for life-stage tracking in women (eg, pregnancy and menopause)	3 (43)
Workflow integration needs	
Desire for structured reports	6 (86)
Need decision support during diagnosis and follow-ups	5 (71)
EHR[a] integration	3 (43)
Use of existing digital tools	
Does not recommend CVD[b] apps to their patients	6 (86)
Low awareness of validated apps	6 (86)
Too many apps with low usability for older adults	5 (71)

[a]EHR: electronic health record.

[b]CVD: cardiovascular disease.

Desired Features and Functionality

Based on the patient interviews, several key themes emerged regarding desired features and functionalities for a CVD app tailored to women. Patients emphasized the importance of integrated support across the full care continuum, from early symptom awareness and diagnosis to daily management and communication with health care providers.

Core priorities included the ability to track vital signs, log symptoms, and receive tailored educational content specific to women's cardiovascular health. Medication reminders, stress management features, and dietary guidance were commonly requested. Also, 9 of 11 (82%) patients expressed interest in syncing the app with wearables to streamline data collection and support longitudinal tracking. Communication features, such as automated report generation and the ability to share real-time data with health care providers, were considered highly valuable, with 8 of 11 (73%) patients supporting real-time feedback or alerts when symptoms are concerning. Importantly, all participants favored sex-specific, personalized recommendations, but some voiced concerns about privacy, complexity, and the risk of being overwhelmed by notifications. Table 3 summarizes the key features and functionalities patients desire in a digital health solution, along with the main barriers and concerns, and indicates how many participants raised each point during the interviews.

Based on the interviews with cardiologists, several priorities emerged. The overarching emphasis was on improving diagnostic precision, treatment adherence, and patient-provider communication, while minimizing time burden and clinical noise. Clinicians expressed strong interest in receiving structured patient-generated data (eg, BP, symptom trends, medication adherence, and hormonal cycle data), especially when aggregated into concise, longitudinal reports. There was broad consensus (7/7, 100%) on the value of sex-specific insights (eg, menopause and pregnancy risks), particularly if tailored to life stages and actionable. However, real-time monitoring or alerts received more mixed responses. Most clinicians (5/7, 71%) preferred periodic summaries over continuous alerts, citing concerns around workload, liability, and alert fatigue.

Communication through the app was not favored by most clinicians, who instead preferred to retain current channels such as email, phone, or in-person visits. While 6 out of 7 (86%) found the ability to customize patient goals important, they highlighted that this must not increase cognitive or administrative burden. Concerns were raised about data quality, patient over-reliance on technology, and integration with existing clinical infrastructure (eg, electronic health records [EHRs]). Despite some openness to features such as AI-driven alerts or remote monitoring, most clinicians emphasized the need for evidence of clinical use and a clear focus on reducing, not increasing, workload. Table 4 summarizes the key features and functionalities clinicians desire in a digital health solution, along with the main barriers and concerns, and indicates how many participants raised each point during the interviews.

XSL•FO
RenderX

Table 3. Desired features and functionality: the patient's perspective (n=11).

Themes and subthemes	Prevalence, n (%)
Diagnosis support and education	
Educational content tailored to women's heart health	10 (91)
Summaries and videos explaining medications or symptoms	3 (27)
Symptom awareness and early management	
Symptom logging	3 (27)
Hormonal cycle insights	2 (18)
Personalized alerts based on symptoms or wearable input	5 (45)
Treatment management and adherence support	
Medication reminders	7 (64)
Diet and lifestyle guidance	5 (45)
Personalized health recommendations based on lifestyle and symptoms	5 (45)
Appointment reminders	4 (36)
Monitoring and tracking	
Vital signs tracking (BP[a], HR[b], ECG[c], etc)	9 (82)
Integration with wearables	9 (82)
Sleep and stress tracking	3 (27)
Cycle tracking	2 (18)
Communication and feedback loops	
Real-time feedback or alerts when symptoms are concerning	8 (73)
Automated report generation	6 (55)
Ability to share data with HCPs[d]	6 (55)
In-app messaging with the care team	4 (36)
Motivation and well-being	
Exercise and mindfulness tips	2 (18)
Community and patient support forums	2 (18)
Barriers and concerns	
App complexity	4 (36)
Privacy concerns	3 (27)
Over-reliance or tech-induced anxiety	3 (27)
Over-generalization of content	2 (18)
No concerns	4 (36)

[a]BP: blood pressure.
[b]HR: heart rate.
[c]ECG: electrocardiogram.
[d]HCP: health care professional.

XSL•FO
RenderX

Table 4. Desired features and functionality: the clinician's perspective (n=7).

Themes and subthemes	Prevalence, n (%)
Treatment planning and customization	
Use of sex-specific guidance (pregnancy and menopause)	6 (86)
Important to customize goals and treatment plans	6 (86)
Monitoring preferences	
Prefer active patient monitoring	3 (43)
Prefer alerts-only model	3 (43)
Would not use the app for monitoring (education use only)	1 (14)
Communication features	
Prefer automated patient reports	5 (71)
In-app direct communication not preferred (prefer phone, email, and in-person follow-up)	5 (71)
Alert preferences	
Prefer periodic summaries only	5 (71)
Prefer both real-time alerts plus summaries	1 (14)
Prefer alerts only for high-risk events	1 (14)
Patient-generated data priorities	
Blood pressure readings	7 (100)
Heart rate and rhythm	5 (71)
Symptom tracking	4 (57)
Medication adherence reports	4 (57)
Hormonal cycle fluctuations	4 (57)
Stress, sleep, and mental health indicators	3 (43)
Exercise and activity levels	3 (43)
Weight and BMI changes	2 (29)
Lifestyle factors (eg, diet and smoking)	2 (29)
CVD[a] risk score integration	2 (29)
Diagnostic data (eg, lab values and cholesterol)	2 (29)
Barriers and concerns	
Excess workload and nonactionable data (noise)	4 (57)
Data accuracy and false alarms	3 (43)
Legal liability if alerts go unaddressed	3 (43)
Poor EHR[b] integration	3 (43)
Privacy and data security concerns	3 (43)
Lack of clinical use	2 (29)
Patients' over-reliance on the app	2 (29)
Motivation and sustained engagement by patients	1 (14)

[a]CVD: cardiovascular disease.

[b]EHR: electronic health record.

Contrasting the two perspectives revealed strong alignment between patients and clinicians on the value of tracking vital signs, providing sex-specific guidance, and offering educational content tailored to women's heart health. However, notable divergences emerged around communication, monitoring preferences, and the diagnostic support role of the app. Patients favored real-time feedback, interactive features, and personalized support, while clinicians expressed concerns about workload, data reliability, and legal liability, preferring structured summaries and limited app-mediated communication. These differences highlight the need for a dual-pathway design that

XSL•FO
RenderX

balances patient empowerment with clinical workflow and safety.

Table 5 compares the perspectives of participating women with CVD versus cardiologists regarding desired app features across the patient journey. It highlights areas of alignment and divergence across different phases of the patient journey, and identifies where design focus is most needed to address gaps and support user-centered implementation.

Table 5. User requirements comparison across the cardiovascular disease patient journey (patients vs clinicians).

Feature cluster	Patient perspective	Clinician perspective	Alignment
Symptom onset and early risk tracking	Many patients want tools to log symptoms, track hormonal cycles, and receive alerts for serious changes.	Clinicians track symptoms but prefer periodic summaries over real-time alerts due to workload and liability concerns.	Partial
Diagnosis support	Patients often seek symptom explanations and hope for diagnostic guidance from the app.	Clinicians are wary of misdiagnosis; some see the app as educational but not diagnostic.	Divergent
Treatment adherence and medication	High interest in medication reminders, side effect information, and adherence support.	Clinicians value adherence tracking and side effect reporting if actionable, summarized, and integrated into their workflow.	Strong
Education and empowerment	Nearly all patients want tailored, female-focused educational content (eg, videos, articles, and lifestyle advice).	Clinicians agree that education is a key benefit of the app, especially if it supports better patient engagement.	Strong
Sex-specific and life stage guidance	All patients emphasized the need for advice linked to hormonal changes, pregnancy, and menopause.	Most clinicians strongly support sex-specific features to address overlooked risks and life-stage changes.	Strong
Vital signs and lifestyle monitoring	Most want to track vitals (BP[a] and HR[b]) and sync with wearables for exercise, diet, and sleep insights.	Clinicians prioritize BP and HR and accept wearable data cautiously, questioning accuracy and clinical validity.	Moderate
Personalization and motivation	Strong interest in personalized advice and goal-setting features to support self-management.	Clinicians support customizable goals but are concerned about overburdening users or implying unsupported precision.	Partial
Communication with care teams	Many patients want in-app messaging, shared data, and reminders to feel more connected and supported.	Clinicians prefer structured reports and oppose app-based messaging due to time limits and workload.	Divergent
Monitoring preferences	Most prefer real-time feedback if needed, while some fear over-monitoring or anxiety from alerts.	The majority prefer summary reports; only a few support real-time data access or alerts, citing resource constraints.	Divergent
Trust, privacy, and usability	Simplicity and ease of use matter most; some express concerns about privacy and overalerting.	Clinicians worry about liability, data overload, and integration challenges; they prefer actionable, low-burden tools.	Moderate

[a]BP: blood pressure.

[b]HR: heart rate.

Success Metrics and Willingness to Pay

Patients defined success with the app in terms of greater control, improved understanding, and emotional reassurance. Key success indicators include feeling more in control of their heart health (7/11, 64%), increased awareness of sex-specific symptoms and conditions (6/11, 55%), and reduced anxiety or stress (5/11, 45%), along with better adherence to medication and more personalized guidance. For many, success is also tied to the app's ability to improve communication with health care providers and deliver trustworthy, up-to-date content.

Regarding willingness to pay, the average score was 7.1 on a scale of 1 to 10, with individual responses ranging from 3 to 10. Preferences for pricing models were mixed, with a slight preference for one-time payments (6/11, 55%), reflecting a desire for financial simplicity and predictability. Monthly models were preferred by 4 of 11 (36%) respondents for their flexibility, though some requested a free trial period as a prerequisite.

The primary factors influencing willingness to pay included the app's demonstrated usefulness, degree of personalization, sex- and age-specific features, trustworthiness, and perceived value for improving cardiovascular health. A few respondents flagged affordability and usability as potential barriers, highlighting the importance of designing an accessible and demonstrably beneficial tool. Table 6 provides an overview of patient-reported success metrics and expectations around payment for digital health solutions, including the number of participants who mentioned each aspect during the interviews.

XSL•FO
RenderX

Table 6. Patient-reported success metrics and payment expectations (n=11).

Themes and subthemes	Prevalence, n (%)
Perceived success indicators	
Feeling more in control of heart health	7 (64)
Increased awareness of sex-specific symptoms and conditions	6 (55)
More personalized recommendations	5 (45)
Reduced stress or anxiety	5 (45)
Increased motivation to manage health	4 (36)
Access to reliable and up-to-date health information	3 (27)
Improved medication adherence	2 (18)
Improved communication with doctors (eg, being taken seriously)	2 (18)
Fewer emergency visits	1 (9)
Preferred payment model	
One-time payment	6 (55)
Monthly subscription	4 (36)
Mixed (sees pros and cons in both)	1 (9)
Payment decision drivers	
Proven usefulness and effectiveness	6 (55)
Personalized recommendations or tracking	5 (45)
Availability of sex- or age-specific features	4 (36)
Free trial option	3 (27)
Affordability or current financial situation	3 (27)
Recommendation by a doctor or a trusted source	2 (18)
Simplicity and ease of use	2 (18)
Positive user reviews	1 (9)
Direct communication or data sharing with HCP[a]	1 (9)

[a]HCP: health care professional.

Clinicians defined the success of a CVD app through its impact on patient engagement, adherence, and health outcomes, rather than direct clinical decision-making support. The most frequently cited indicators of success included improved patient engagement and participation in care management (6/7, 86%), greater adherence to treatment (5/7, 71%), and better patient education and symptom recognition (5/7, 71%), particularly regarding sex-specific cardiovascular risks. Several clinicians emphasized the need for such outcomes to be validated through research before adoption in routine practice.

Adoption incentives included features that support remote monitoring, patient education, and AI-supported risk assessments, provided they are accurate and well-integrated. Seamless integration into clinical workflows and EHRs, a simple interface, and support for research and data sharing were also valued.

Regarding willingness to pay, clinicians expressed moderate interest, with a mean score of 5.3 (SD 1.3) on a scale of 1 to 10. Preferences leaned toward monthly or yearly pricing models, with a desire for a free trial period. Willingness to pay was highly conditional on the app's proven benefit to patient outcomes, usability, compliance with data protection, and ability to support research or institutional adoption. Table 7 provides an overview of clinician-reported success metrics and expectations around payment for digital health solutions, including the number of participants who mentioned each aspect during the interviews.

XSL•FO
RenderX

Table 7. Clinician-reported success metrics and payment expectations (n=7).

Themes and subthemes	Prevalence, n (%)
Perceived success indicators	
Improved patient engagement and participation in care management	6 (86)
Improved adherence to treatment	5 (71)
Better symptom recognition and earlier referral	5 (71)
Better patient education and self-management	5 (71)
Reduction in preventable cardiovascular events or hospitalizations	3 (43)
Must be tested in a research context	3 (43)
Easier or faster appointments due to prefilled questionnaires	2 (29)
Long-term lifestyle improvements	1 (14)
Increased patient satisfaction and trust	1 (14)
Must-have features for adoption	
Remote monitoring or real-time patient tracking	4 (57)
Educational content (videos, disease explanations, etc.)	4 (57)
AI[a]-driven risk assessments	3 (43)
Ability to use data for research	3 (43)
Patient engagement tools (behavior change and motivation)	3 (43)
Simple, intuitive user interface	3 (43)
Integration with EHR[b] systems	3 (43)
Preferred payment model	
Monthly payment	3 (43)
Yearly payment (with trial preferred)	2 (29)
One-time payment	1 (14)
No preference	1 (14)
Payment decision drivers	
Proven improvement in health outcomes and adherence	4 (57)
Ability to support research (questionnaires and data export)	3 (43)
Data security and privacy guarantees	2 (29)
Institutional or guideline endorsement	2 (29)
Direct benefit to patients (motivation and education)	2 (29)
Ease of use (for both patients and providers)	2 (29)
Broad adoption or market share	1 (14)
Gender-specific features are fully integrated	1 (14)

[a]AI: artificial intelligence.

[b]EHR: electronic health record.

Regulatory and Reimbursement Considerations

To complement the user requirements gathered from patients and clinicians, we conducted additional interviews with one regulatory expert (SN [nonauthor]) and one reimbursement expert (MF [nonauthor]), both with in-depth knowledge of the Swiss and broader European health care landscapes. This supplementary perspective is critical, as the app will be developed primarily for use in Switzerland, with a medium-term vision for deployment in other European contexts.

The regulatory expert assessed that an AI-enabled CVD app (AI-driven alerts and AI-supported risk assessments), designed to support women across the spectrum of symptom recognition, diagnosis, treatment adherence, and self-management, is likely to be considered a medical device under Swiss law. Specifically, if the app includes features such as AI-driven risk alerts, symptom tracking, and clinical report generation intended to support medical decision-making or influence health outcomes, it would fall under the Medical Device Ordinance as governed by Swissmedic. Depending on the final functionality, the app

XSL•FO
RenderX

would likely be classified as a Class IIa medical device, requiring compliance with regulatory obligations related to safety, performance, and quality assurance.

Given that the app would likely use algorithms to generate health risk scores and may eventually integrate with electronic health records or wearable data for clinical review, transparency of its AI functionalities (eg, AI-driven alerts and AI-supported risk assessments) and proper documentation of its performance are essential. Although Switzerland is not part of the EU, it has harmonized its medical device regulatory system with the EU Medical Device Regulation (MDR). Therefore, developers must ensure conformity with Annex I of the MDR, particularly concerning software validation, cybersecurity, and human oversight of AI outputs. Early consultation with Swissmedic, including use of their pre-submission guidance services, was strongly recommended by the expert to confirm classification and identify the most appropriate pathway to conformity assessment.

From a data protection standpoint, the expert added that the app must comply with the Swiss Federal Act on Data Protection, which aligns closely with the EU's General Data Protection Regulation. This includes obtaining explicit consent for processing health data, ensuring transparency in AI-based recommendations, and conducting a Data Protection Impact Assessment if sensitive data are used for profiling or personalized recommendations. Data must be stored securely in accordance with Swiss and European data security standards.

In terms of market access within Switzerland, the reimbursement expert explained that there is currently no established reimbursement pathway for digital health tools of this nature under the standard TARMED system (soon to be replaced by the new tariff system called TARDOC). However, several viable strategies exist. These include partnerships with supplemental insurance providers, integration into occupational health programs, or collaborative pilots with cantonal public health authorities focused on prevention and chronic disease management. Additionally, public-private partnerships involving academic hospitals could support real-world evidence generation, which will be critical for clinical validation and acceptance.

As the app evolves, modular feature deployment (eg, launching initially with educational content and lifestyle tracking before integrating AI-based alerts or EHR connectivity) may offer a lower-risk route to initial uptake. Positioning the app as an adjunct to care rather than a diagnostic tool can help mitigate medico-legal risks and facilitate clinician adoption. For future scaling into EU markets, preparation for CE (Conformité Européenne) marking and consideration of fast-track pathways like Germany's Digitale Gesundheitsanwendungen (Digital Health Applications) process would be key strategic steps. A summary of regulatory and reimbursement considerations is provided in Multimedia Appendix 5, focusing primarily on Switzerland, with additional insights relevant to potential midterm expansion into the EU market.

Discussion

Primary Findings

This study offers detailed insights into how digital health tools can be designed to better serve women with CVD by directly addressing long-standing gaps in diagnosis, treatment, and self-management. What distinguishes this work is its grounding in sex-specific lived experiences and clinician workflows; most existing tools and studies either generalize across populations or fail to capture the unique challenges women face across the cardiovascular care continuum.

Our findings confirmed the diagnostic blind spots long documented in the literature but go a step further by detailing how these gaps are experienced by women, especially the emotional toll of feeling dismissed or misdiagnosed due to atypical symptom presentation, and how clinicians themselves acknowledge the limitations of current care pathways and tools. We also identified a mismatch between what patients expect from digital tools (eg, personalized, real-time, and educational support) and what clinicians view as feasible or desirable (eg, structured summaries and low-disruption integration), echoing broader implementation challenges in digital health but offering concrete, user-validated features to bridge this divide.

Furthermore, this work goes beyond user needs to surface actionable priorities for design, regulation, and reimbursement. By mapping unmet needs to specific app features and clarifying stakeholder requirements, we contribute an applied framework for the development of a digital tool that is not only patient-centered but also clinically and systemically grounded.

Addressing Unmet Needs Across the Care Continuum

Our interviews with patients and clinicians confirmed the consistent disconnect between current cardiovascular care pathways and the lived experiences and needs of women with CVD. Both groups emphasized a clear gap in early recognition and diagnosis of CVD in women, often driven by atypical symptom presentation, a finding well-supported in the literature [4]. Women frequently report vague or nonspecific symptoms (eg, fatigue, anxiety, and back pain) that are either misattributed or dismissed, leading to delayed diagnosis and poorer outcomes. This aligns with clinical studies showing that atypical symptoms in women are associated with underdiagnosis and delayed care, a pattern historically described as the Yentl syndrome and reaffirmed in recent studies [27].

Clinicians similarly confirmed that CVD does not present in women as it does in men, often requiring multiple diagnostic steps and greater clinical vigilance. Yet, most participating clinicians admit to using standardized, male-centered diagnostic tools, with little to no adaptation for sex- or life-stage-specific risks (eg, menopause and pregnancy-related complications). This is consistent with research papers that highlight the lack of integration of sex-specific risk factors into clinical practice and guidelines [28,29].

Our findings also showed that there is a shared perception that awareness and education are insufficient, both for patients and providers. Patients often feel overwhelmed and underinformed, particularly about how hormonal changes might affect their

XSL•FO
RenderX

heart health. Clinicians acknowledge this gap, noting that many are not trained to assess or communicate about sex-specific CVD risks, confirming concerns raised in the literature about persistent gender bias in cardiovascular medicine [30].

On the digital health front, while most patients have not used cardiovascular apps, citing lack of awareness, complexity, or emotional overwhelm, many express openness to tools that are personalized, educational, and easy to use. Importantly, all patients agreed that features like hormone-aware symptom tracking and cycle or menopause-specific guidance would be valuable. Clinicians were generally supportive of a digital tool, provided it reduces workload and offers actionable insights. However, many remain skeptical due to workflow integration concerns, unclear clinical use, and insufficient validation, echoing findings from other studies that cite similar adoption barriers [14,31-33].

This shows a strong concordance between user-reported unmet needs in our research and evidence in the scientific literature. The development of a CVD app specifically designed for women presents an opportunity to bridge the identified diagnostic support and management gaps by offering sex-specific decision support and longitudinal symptom tracking across life stages, in a way that integrates meaningfully into both patient journeys and clinician workflows.

Balancing Patient-Centered Design With Clinical Workflow Integration

The synthesis of clinician and patient requirements revealed meaningful overlap in priorities for a CVD app tailored to women, particularly around core features such as BP, HR, cycle and symptom tracking, medication adherence tools, and educational content specific to women's cardiovascular risks. Both groups also expressed interest in incorporating wearable data, lifestyle tracking, and AI-driven alerts, but with key differences in expectations around data flow and engagement. To ensure alignment with human factors (HF) principles, app features and interface components should be conceptualized to minimize cognitive load, support workflow compatibility, and accommodate differentiated engagement preferences between patients and clinicians.

Patients consistently emphasized the importance of holistic, personalized support, especially around lifestyle guidance (eg, diet and stress management), menstrual and hormonal tracking, and emotional well-being. Our recent research, which reviewed 20 commercially available CVD apps, showed that only 25% (5/20) offered sex-specific content, reinforcing that sex-specific app features in CVD are rare and needed [15]. Patients also valued real-time feedback and communication, but with flexibility to self-regulate interaction frequency, a finding supported by previous research [34].

Clinicians, by contrast, largely favored periodic summaries over real-time alerts, citing concerns about information overload, workflow disruption, and medico-legal liability, concerns echoed across several reviews on eHealth implementation challenges [14,32,33,35], underscoring the importance of minimizing cognitive workload and supporting efficient information processing in clinical settings. There was strong support for automated, concise reporting tools to enable faster clinical decision-making, particularly if data is aggregated and actionable (eg, medication nonadherence flags, high BP trends, and cycle fluctuation). Importantly, clinicians stressed the need for sex-specific clinical insights (eg, pregnancy-safe medication guidance), aligning with calls in the literature to address gender gaps in cardiovascular risk assessment and care pathways [36].

Despite agreement on many core features, divergence between patients' and clinicians' requirements remains. Clinicians were skeptical of communication via the app, preferring existing channels (email and phone), whereas many patients desired low-barrier communication tools or asynchronous Q&A features. Additionally, clinician trust in wearable data remains limited due to concerns about accuracy, despite evidence that wearables can support prevention strategies and risk monitoring when used appropriately [10,37].

These findings underscore the need to balance patient-centered design with clinical workflow integration, ensuring that features valued by patients (eg, education, personalization, symptom explanation, and risk prevention) do not impose undue burden on clinicians. The app should prioritize structured reporting, passive data aggregation, and modular engagement options, and be framed as an adjunct, not a replacement, for in-person care. Differentiated user models, with patient-facing tools focused on behavior change, motivation, and real-time support, and clinician-facing tools optimized for decision efficiency and minimal disruption, can support better task-technology fit for both groups. Interface design should adhere to HF principles such as customizability by role, transparency of AI outputs, clear visual hierarchies, and user autonomy in alert settings, all of which were strongly echoed in user preferences. Usability must remain a central design principle, interpreted not only as ease of use but also through HF constructs such as interaction efficiency, information clarity, and minimization of user burden. Successful uptake will hinge on transparent validation, privacy safeguards, and clinical evidence.

Articulating Meaningful Success Narratives

The analysis of success metrics and willingness to pay highlighted distinct yet interrelated priorities between patients and clinicians that carry significant implications for the app's value proposition and sustainable adoption. While both groups identify improved patient adherence, better education, and enhanced engagement as key indicators of effectiveness, they differ in how success is defined and operationalized.

Patients tend to frame success in terms of personal empowerment and perceived well-being, such as increased awareness of sex-specific symptoms and conditions, feeling in control of their heart health, reduced anxiety, and improved ability to interpret symptoms and make informed decisions. Most patients cited greater motivation and personalized insights as important outcomes, and several emphasized a desire for the app to help them feel taken seriously in clinical encounters. These findings align with research demonstrating that digital interventions can significantly boost patient empowerment, knowledge, and self-management in chronic illness [13,38]. This experiential framing of success suggests that patients value both functional benefits (eg, symptom recognition and adherence

XSL•FO
RenderX

reminders) and emotional reassurance, and are likely to judge effectiveness through lived experience rather than clinical endpoints alone.

Clinicians, in contrast, emphasized evidence-based outcomes, such as reduced hospitalizations, earlier detection of complications, and measurable improvements in disease management. While some acknowledged softer benefits like improved communication or decreased workload due to better-informed patients, there was a consistent call for formal evaluation of impact, preferably through research trials, factors frequently noted in clinician acceptance studies [32,33]. This divergence highlights the importance of building a dual feedback system, one capturing patient-reported outcomes and engagement metrics, and another tracking clinical indicators that can be aggregated and validated over time.

In terms of willingness to pay, patients showed relatively high interest. Key influencing factors included the app's ability to improve cardiovascular health, offer sex-specific and personalized recommendations, and provide reliable, easily digestible information, echoing user preferences reported in patients' mHealth adoption literature [34]. Several patients noted the appeal of free trial options and transparent pricing, especially in light of economic constraints.

By contrast, clinicians expressed moderate willingness to pay and viewed app adoption as contingent on demonstrated clinical use, alignment with existing systems, and endorsement by institutions or guidelines. Most clinicians would expect the app to be provided at the health system or institutional level, particularly if it supports research or integrates with existing systems, factors shown as crucial for sustained clinician adoption in previous research [32,33,35].

Payment feasibility was primarily seen through the lens of institutional procurement rather than individual expenditure. This indicates that a tiered strategy could be considered, potentially offering direct-to-patient subscriptions for patients and licensing options or integration pathways for clinical and research settings. Such a dual-strategy may combine offering core low-risk features and premium add-ons directly to motivated patients, with an institutional offering that includes more advanced research data modules such as an AI-predictive model, EHR integration, and structured reporting and smart alerts. Free trials, outcome-based pricing, or inclusion in reimbursement schemes could improve uptake across both segments, as emphasized in similar research [39,40].

Overall, the findings suggest that to drive adoption and value across stakeholder groups, the app must articulate distinct success narratives, one rooted in empowerment and lived experience for patients, and one grounded in clinical efficiency and measurable outcomes for providers.

Regulatory and Reimbursement Implications

Incorporating regulatory and reimbursement expertise early in the development process helps ensure that the solution is not only clinically and user-relevant but also aligned with the legal, safety, and financial frameworks required for real-world adoption. Failure to consider these aspects at an early stage often results in substantial implementation barriers and limits the scalability and sustainability of digital health innovations [14,32-35,41].

From a regulatory standpoint, the expert confirmed that if the app includes AI-driven functionalities, such as personalized risk alerts, clinical report generation, or features that support diagnosis or influence medical decision-making, it would likely qualify as a Class IIa medical device under Switzerland's Medical Device Ordinance, harmonized with the EU MDR. This classification implies that the app must meet formal requirements related to software performance, cybersecurity, human oversight, and clinical evaluation, as outlined in MDR Annex I. Given the inclusion of health data processing and algorithmic personalization, the app must also comply with the Swiss Federal Act on Data Protection, which closely mirrors the general data protection regulation, including requirements for explicit user consent and transparency of algorithmic outputs. These recommendations align with recent regulatory reviews emphasizing the increased scrutiny of AI-based medical software and the importance of early consultation with regulatory bodies to validate classification and compliance pathways [42].

On the reimbursement side, the expert highlighted that Switzerland currently lacks a dedicated reimbursement pathway for digital health applications within the standard outpatient tariff system (TARMED), which is soon to be replaced by the new TARDOC system. In the midterm, TARDOC will be the key framework to monitor. According to the factsheet on digital health applications provided by the Federal Office of Public Health, digital tools, including AI applications, are expected to be classified as part of the infrastructure or personnel services [43]. This means they could potentially be reimbursed under specific TARDOC positions, such as those related to telemedicine. However, several alternative access models were identified, including partnerships with private health insurers, occupational health initiatives, and cantonal prevention programs. These strategies reflect recent calls in the literature for shifts in European digital health financing, which recognize the need for novel evidence-generation mechanisms to support the adoption and scaling of digital tools [44,45]. In Switzerland, reimbursement will likely depend on demonstrated value in real-world settings, particularly in enhancing adherence, promoting prevention, and reducing unnecessary health care use.

Design Implications and Next Steps

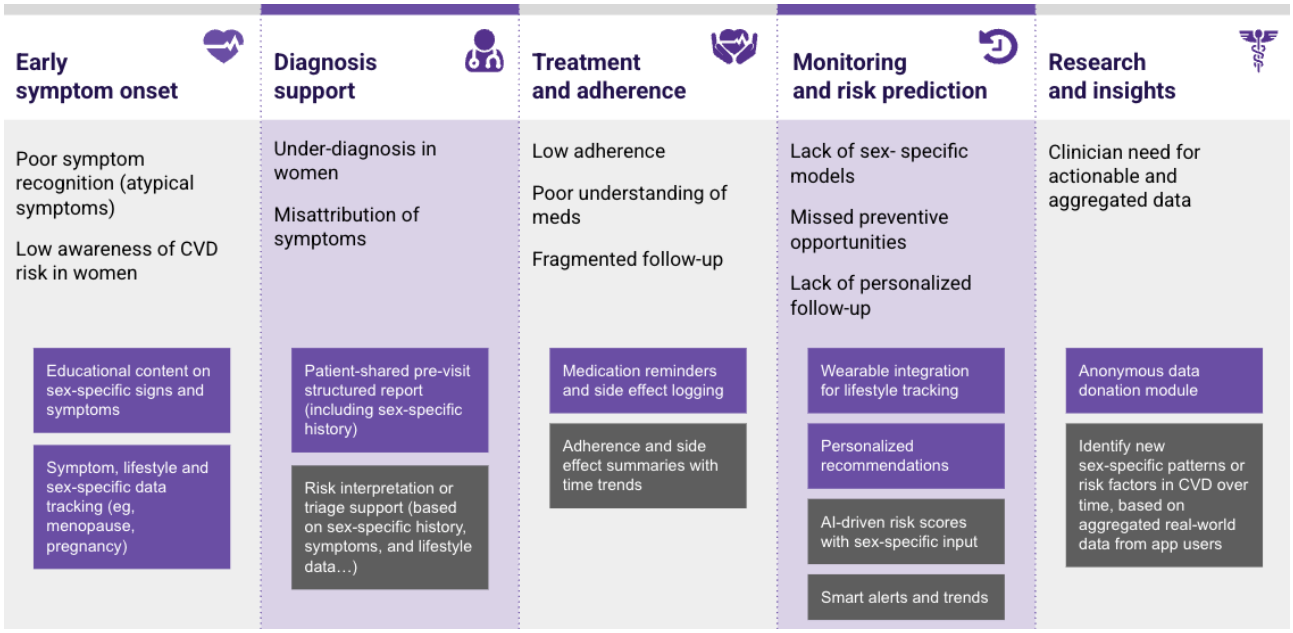
To translate the diverse and specific user requirements identified in this study into actionable design features, we mapped patient and clinician needs along the cardiovascular care continuum. Thematic synthesis of interview data revealed critical gaps and opportunities at multiple points in the patient journey, from early symptom recognition to diagnosis support, treatment adherence, ongoing self-management, and broader contributions to cardiovascular research. These findings aligned with persistent shortcomings in the literature, particularly around under-recognition of sex-specific symptoms [46], insufficient tailoring of risk assessment tools, and limited integration of digital health tools into routine care for women with CVD [15].

Figure 2 presents a high-level design framework illustrating how these needs will be addressed through core app features,

XSL•FO
RenderX

organized by phase of care. Patient-facing features are shown in dark violet boxes and focus on supporting the lived experience of women navigating cardiovascular conditions, such as personalized education, symptom logging, and wearable-integrated tracking. Clinician-facing features, in the dark gray boxes in the figure, prioritize workflow efficiency, diagnostic support, and data integration. It's also worth noting the interconnectedness of some of these elements. For example, the AI-driven risk alerts and clinical report generation depend on continuous, structured input from patient-generated data, including wearable metrics, symptom tracking, and medication adherence. Similarly, the research and insights feature, designed for clinician relevance, builds on aggregated app data to identify emerging sex-specific patterns and close gaps in clinical evidence.

Figure 2. Design implications across the cardiovascular patient journey in women. AI: artificial intelligence; CVD: cardiovascular disease.



This feature map provides a consolidated foundation for the next phase of development. It will serve as the blueprint for the initial prototype, which will undergo usability testing in an iterative design process, as highlighted in Figure 1 under the methodology section. Through repeated feedback loops with both patients and health care professionals, the prototype will be refined to ensure alignment with real-world needs and system-level constraints, ultimately enhancing adoption, clinical relevance, and long-term impact.

Limitations and Future Research

This study was conducted exclusively in Switzerland, which may limit the generalizability of the findings to other health systems or cultural contexts. Although all patients were recruited via a single university hospital partner, the clinic serves a broad public population and includes individuals with varying levels of health and digital literacy. Nonetheless, this limitation will be addressed in the next phase of the project through iterative prototyping with a more demographically and geographically diverse patient sample. Given that the primary goal is to develop a tool for use in Switzerland, with longer-term plans to expand into other German-speaking countries such as Germany and Austria, we believe the study was appropriately scoped for this stage of development. Broader geographic and cultural representation will be integrated in the next phase of the development, particularly during prototype testing and validation.

The sample size was relatively small, in line with typical qualitative research. While this limits the breadth of perspectives captured, we achieved thematic saturation, suggesting that the findings are sufficiently robust to inform the current design phase. Nonetheless, some stakeholder groups were not represented in this study. In particular, general practitioners and nurses were not included due to recruitment challenges and resource constraints. Their perspectives, especially those of general practitioners, who often serve as the first point of contact during early symptom onset, are crucial for understanding barriers to timely recognition and referral. Future phases of the development will include these stakeholders to ensure that insights from across the care continuum are reflected in the final design.

Conclusions

This study confirmed a persistent disconnect between existing cardiovascular care pathways and the specific needs of women living with CVD. Patients shared experiences of feeling underinformed, underheard, and underserved, particularly during early diagnosis support, due to atypical symptoms and a lack of sex-specific guidance. Clinicians acknowledged these challenges but admitted that current tools and workflows are still largely designed around male-centric risk models. Both groups agreed that digital health has the potential to bridge this gap but also highlighted different expectations; patients seek personalized education and empowerment, while clinicians prioritize clinical use, low burden, and workflow integration.

The participants' input showed strong overlap in desired core features, such as sex-specific symptom and cycle tracking, medication adherence tools, and tailored educational content.

XSL•FO
RenderX

However, divergences appeared around data communication and feedback. Patients favored more interactive, real-time support, while clinicians preferred summarized, actionable insights to avoid alert fatigue and liability concerns. These findings highlight the importance of modular, adaptable design, an app that offers flexibility in use, differentiated interfaces for patients and providers, and clearly defined roles for AI and automation.

These findings also inform key HF considerations for interface design. For clinicians, minimizing cognitive workload was critical; participants favored tools that provide actionable, low-effort insights rather than real-time data streams. Features such as passive data aggregation, structured dashboards, and summary reports were seen as more aligned with clinical workflows and information processing constraints. For patients, usability was tied to personalization, transparency, and flexible communication options. To ensure task-technology fit, the design must allow differentiated engagement based on user role, with modular, customizable features that minimize burden while maximizing relevance.

Success metrics also differed by perspective. Patients emphasized psychological and motivational benefits such as feeling in control, recognized, and informed, whereas clinicians focused on measurable outcomes such as adherence, hospital avoidance, and engagement. However, both agreed that success requires trust, usability, and relevance to daily life. These insights informed a dual-value strategy for app development, aligning patient empowerment with clinical efficiency.

Finally, our expert interviews highlighted that integrating regulatory and reimbursement perspectives early is key to a successful implementation. With features such as AI-supported insights and clinical data sharing, the app will likely qualify as a medical device under Swiss and European laws. Planning for staged implementation, starting with nonregulated educational features and expanding as validation and infrastructure grow, emerges as a practical strategy. However, long-term success and widespread adoption would largely depend on demonstrating real-world value and forming partnerships with insurers and public health programs.

Acknowledgments

The authors thank the Women's Heart Health Program at the University Hospital Basel for their instrumental role in patient recruitment. Artificial intelligence was not used to create, modify, or interpret the scientific content of the article. It was solely used as a proofreading tool to correct spelling and grammar, enhance clarity, and streamline phrasing. The authors take full responsibility for the integrity and originality of the work.

Funding

This study was sponsored by Innosuisse (Swiss Innovation Agency, grant 75940.1 INNO-ICT). The role of Innosuisse is to promote science-based innovation in the interests of industry and society in Switzerland.

Conflicts of Interest

PV is the founder of Vavken Health Labs, the technology provider developing the application being studied. CJ is an editorial board member of *JMIR Human Factors*, a JMIR Publications journal, at the time of this publication.

Multimedia Appendix 1
Participant information sheet and consent form.
[PDF File (Adobe PDF File), 333 KB - humanfactors_v13i1e82916_app1.pdf]

Multimedia Appendix 2
Sample characteristics of participating patients and clinicians.
[PDF File (Adobe PDF File), 80 KB - humanfactors_v13i1e82916_app2.pdf]

Multimedia Appendix 3
Interview guides.
[PDF File (Adobe PDF File), 123 KB - humanfactors_v13i1e82916_app3.pdf]

Multimedia Appendix 4
Thematic Analysis Process.
[PDF File (Adobe PDF File), 115 KB - humanfactors_v13i1e82916_app4.pdf]

Multimedia Appendix 5
Summary of regulatory and reimbursement considerations.
[PDF File (Adobe PDF File), 95 KB - humanfactors_v13i1e82916_app5.pdf]

XSL•FO
RenderX

References

1. Tromp J, Jindal D, Redfern J, Bhatt A, Séverin T, Banerjee A, et al. World Heart Federation roadmap for digital health in cardiology. Glob Heart 2022;17(1):61 [FREE Full text] [doi: 10.5334/gh.1141] [Medline: 36051317]
2. Leading causes of death. Centers for Disease Control and Prevention (CDC). URL: https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm [accessed 2025-01-10]
3. Cardiovascular diseases. World Health Organization. URL: https://www.who.int/health-topics/cardiovascular-diseases#tab=tab_1 [accessed 2025-01-17]
4. Bairey Merz CN, Andersen H, Sprague E, Burns A, Keida M, Walsh MN, et al. Knowledge, attitudes, and beliefs regarding cardiovascular disease in women: the women's heart alliance. J Am Coll Cardiol 2017 Jun 22;70(2):123-132 [FREE Full text] [doi: 10.1016/j.jacc.2017.05.024] [Medline: 28648386]
5. Al Hamid A, Beckett R, Wilson M, Jalal Z, Cheema E, Al-Jumeily Obe D, et al. Gender bias in diagnosis, prevention, and treatment of cardiovascular diseases: a systematic review. Cureus 2024 Feb 15;16(2):e54264 [FREE Full text] [doi: 10.7759/cureus.54264] [Medline: 38500942]
6. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. World Health Organization Institutional Repository for Information Sharing. 2011. URL: https://apps.who.int/iris/handle/10665/44607 [accessed 2022-11-27]
7. Singh S, Kolekar MH, Tanwar S, Wierzchoń ST, Bhatnagar RK, editors. Emerging Technologies for Computing, Communication and Smart Cities. Singapore: Springer Nature Singapore; 2022.
8. Coorey GM, Neubeck L, Mulley J, Redfern J. Effectiveness, acceptability and usefulness of mobile applications for cardiovascular disease self-management: Systematic review with meta-synthesis of quantitative and qualitative data. Eur J Prev Cardiol 2018;25(5):505-521. [doi: 10.1177/2047487317750913] [Medline: 29313363]
9. Castaneda D, Esparza A, Ghamari M, Soltanpur C, Nazeran H. A review on wearable photoplethysmography sensors and their potential future applications in health care. Int J Biosens Bioelectron 2018;4(4):195-202 [FREE Full text] [doi: 10.15406/ijbsbe.2018.04.00125] [Medline: 30906922]
10. Scheid JL, Reed JL, West SL. Commentary: is wearable fitness technology a medically approved device? Yes and no. Int J Environ Res Public Health 2023;20(13) [FREE Full text] [doi: 10.3390/ijerph20136230] [Medline: 37444078]
11. Cowie MR, Lam CSP. Remote monitoring and digital health tools in CVD management. Nat Rev Cardiol 2021:457-458. [doi: 10.1038/s41569-021-00548-x] [Medline: 33824486]
12. Guo Y, Wang H, Zhang H, Liu T, Liang Z, Xia Y, et al. Mobile photoplethysmographic technology to detect atrial fibrillation. J Am Coll Cardiol 2019;74(19):2365-2375 [FREE Full text] [doi: 10.1016/j.jacc.2019.08.019] [Medline: 31487545]
13. Triantafyllidis A, Kondylakis H, Katehakis D, Kouroubali A, Alexiadis A, Segkouli S, et al. Smartwatch interventions in healthcare: a systematic review of the literature. Int J Med Inform 2024;190:105560. [doi: 10.1016/j.ijmedinf.2024.105560] [Medline: 39033723]
14. Jacob C, Lindeque J, Müller R, Klein A, Metcalfe T, Connolly SL, et al. A sociotechnical framework to assess patient-facing eHealth tools: results of a modified Delphi process. NPJ Digit Med 2023;6(1):232 [FREE Full text] [doi: 10.1038/s41746-023-00982-w] [Medline: 38102323]
15. Chauhan GK, Vavken P, Jacob C. Mobile apps and wearable devices for cardiovascular health: narrative review. JMIR Mhealth Uhealth 2025;13:e65782-e65782. [doi: 10.2196/65782]
16. Pope C, Mays N. Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. BMJ 1995;311(6996):42-45 [FREE Full text] [doi: 10.1136/bmj.311.6996.42] [Medline: 7613329]
17. Mays N, Pope C. Qualitative research in health care. Assessing quality in qualitative research. BMJ 2000;320(7226):50-52 [FREE Full text] [doi: 10.1136/bmj.320.7226.50] [Medline: 10617534]
18. Norman DA, Draper SW. User Centered System Design: New Perspectives on Human-Computer Interaction. Hillsdale, NJ: Lawrence Erlbaum Associates; 1986.
19. The Field Guide to Human-Centered Design. IDEO. URL: https://www.ideo.com/journal/design-kit-the-human-centered-design-toolkit [accessed 2022-06-10]
20. Or CK, Holden RJ, Valdez RS. Human factors engineering and user-centered design for mobile health technology: enhancing effectiveness, efficiency, and satisfaction. In: Human-Automation Interaction: Mobile Computing. Cham, Switzerland: Springer; 2023:97-118.
21. Göttgens I, Oertelt-Prigione S. The application of human-centered design approaches in health research and innovation: a narrative review of current practices. JMIR Mhealth Uhealth 2021;9(12):e28102 [FREE Full text] [doi: 10.2196/28102] [Medline: 34874893]
22. Jacob C, Bourke S, Heuss S. From testers to cocreators—the value of and approaches to successful patient engagement in the development of ehealth solutions: qualitative expert interview study. JMIR Hum Factors 2022;9(4):e41481. [doi: 10.2196/41481]
23. Nabid MR, Sedghi R, Behbahani M, Arvan B, Heravi MM, Oskooie HA. Application of poly 1,8-diaminonaphthalene/multiwalled carbon nanotubes-COOH hybrid material as an efficient sorbent for trace determination of cadmium and lead ions in water samples. J Mol Recognit 2014;27(7):421-428. [doi: 10.1002/jmr.2361] [Medline: 24895274]

XSL•FO
RenderX

24. Creswell J, Poth C. Qualitative Inquiry and Research Design: Choosing Among Five Approaches. Thousand Oaks, CA: SAGE Publications, Inc; 2017.
25. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014;9:26152 [FREE Full text] [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]
26. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
27. Laborante R, Borovac JA, Galli M, Rodolico D, Ciliberti G, Restivo A, et al. Gender-differences in antithrombotic therapy across the spectrum of ischemic heart disease: time to tackle the Yentl syndrome? Front Cardiovasc Med 2022;9 [FREE Full text] [doi: 10.3389/fcvm.2022.1009475] [Medline: 36386309]
28. de Marvao A, Alexander D, Bucciarelli-Ducci C, Price S. Heart disease in women: a narrative review. Anaesthesia 2021;76 Suppl 4:118-130 [FREE Full text] [doi: 10.1111/anae.15376] [Medline: 33682102]
29. Keteepe-Arachi T, Sharma S. Cardiovascular disease in women: understanding symptoms and risk factors. Eur Cardiol 2017;12(1):10-13 [FREE Full text] [doi: 10.15420/ecr.2016:32:1] [Medline: 30416543]
30. Panahiazar M, Bishara AM, Chern Y, Alizadehsani R, Islam SMS, Hadley D, et al. Gender-based time discrepancy in diagnosis of coronary artery disease based on data analytics of electronic medical records. Front Cardiovasc Med 2022;9 [FREE Full text] [doi: 10.3389/fcvm.2022.969325] [Medline: 36505372]
31. Park LG, Kariuki JK, Hendriks JM, Gallagher R, Burke LE. A review and promise of digital health technologies to promote global cardiovascular health: a call to action for nursing. J Cardiovasc Nurs 2025. [doi: 10.1097/JCN.0000000000001205] [Medline: 40241412]
32. Jacob C, Sanchez-Vazquez A, Ivory C. Understanding clinicians' adoption of mobile health tools: a qualitative review of the most used frameworks. JMIR Mhealth Uhealth 2020;8(7):e18072 [FREE Full text] [doi: 10.2196/18072] [Medline: 32442132]
33. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR Mhealth Uhealth 2020;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]
34. Jacob C, Sezgin E, Sanchez-Vazquez A, Ivory C. Sociotechnical factors affecting patients' adoption of mobile health tools: systematic literature review and narrative synthesis. JMIR Mhealth Uhealth 2022;10(5):e36284 [FREE Full text] [doi: 10.2196/36284] [Medline: 35318189]
35. Jacob C, Brasier N, Laurenzi E, Heuss S, Mougiakakou S, Cöltekin A, et al. AI for IMPACTS framework for evaluating the long-term real-world impacts of AI-powered clinician tools: systematic review and narrative synthesis. J Med Internet Res 2025;27:e67485 [FREE Full text] [doi: 10.2196/67485] [Medline: 39909417]
36. Piani F, Baffoni L, Strocchi E, Borghi C. Gender-specific medicine in the european society of cardiology guidelines from 2018 to 2023: where are we going? J Clin Med 2024;13(14) [FREE Full text] [doi: 10.3390/jcm13144026] [Medline: 39064064]
37. Gill SK, Barsky A, Guan X, Bunting KV, Karwath A, Tica O, et al. Consumer wearable devices for evaluation of heart rate control using digoxin versus beta-blockers: the RATE-AF randomized trial. Nat Med 2024;30(7):2030-2036. [doi: 10.1038/s41591-024-03094-4] [Medline: 39009776]
38. Fomo M, Borga LG, Abel T, Santangelo PS, Riggare S, Klucken J, et al. Empowering capabilities of people with chronic conditions supported by digital health technologies: scoping review. J Med Internet Res 2025;27:e68458 [FREE Full text] [doi: 10.2196/68458] [Medline: 40577044]
39. Avşar TS, Elvidge J, Hawksworth C, Kenny J, Németh B, Callenbach M, et al. Linking reimbursement to patient benefits for advanced therapy medicinal products and other high-cost innovations: policy recommendations for outcomes-based agreements in Europe. Value Health 2024;27(11):1497-1506 [FREE Full text] [doi: 10.1016/j.jval.2024.07.007] [Medline: 39094693]
40. van Kessel R, Srivastava D, Kyriopoulos I, Monti G, Novillo-Ortiz D, Milman R, et al. Digital health reimbursement strategies of 8 European countries and Israel: scoping review and policy mapping. JMIR Mhealth Uhealth 2023;11:e49003 [FREE Full text] [doi: 10.2196/49003] [Medline: 37773610]
41. Jacob C, Lindeque J, Klein A, Ivory C, Heuss S, Peter MK. Assessing the quality and impact of eHealth tools: systematic literature review and narrative synthesis. JMIR Hum Factors 2023;10:e45143 [FREE Full text] [doi: 10.2196/45143] [Medline: 36843321]
42. Fraser AG, Biasin E, Bijnens B, Bruining N, Caiani EG, Cobbaert K, et al. Artificial intelligence in medical device software and high-risk medical devices - a review of definitions, expert recommendations and regulatory initiatives. Expert Rev Med Devices 2023;20(6):467-491 [FREE Full text] [doi: 10.1080/17434440.2023.2184685] [Medline: 37157833]
43. Factsheet on the reimbursement of digital health applications under the mandatory health insurance (OKP) [Report in German]. Federal Office of Public Health (FOPH). Bern, Switzerland: Bundesamt für Gesundheit (BAG); 2024. URL: https://backend.bag.admin.ch/fileservice/sdweb-docs-prod-bagadminch-files/files/2025/03/18/00d21ad2-1f9d-4c04-803b-622ea70a127f.pdf [accessed 2025-08-06]
44. Arcà E, Heldt D, Smith M. Comparison of health technology assessments for digital therapeutics in Germany, the United Kingdom and France. Digit Health 2025;11 [FREE Full text] [doi: 10.1177/20552076241308704] [Medline: 39845519]

XSL•FO
RenderX

45. Gehder S, Goeldner M. Unpacking performance factors of innovation systems and studying Germany's attempt to foster the role of the patient through a market access pathway for digital health applications (DiGAs): exploratory mixed methods study. J Med Internet Res 2025;27:e66356. [doi: 10.2196/66356]
46. Wang LY, Chiang GSH, Wee CF, Chan SWK, Lau JXX, Taeihagh A. Preventing ischemic heart disease in women: a systematic review of global directives and policies. npj Womens Health 2024;2(1). [doi: 10.1038/s44294-024-00040-0]

Abbreviations

AF: atrial fibrillation
AI: artificial intelligence
BP: blood pressure
CE: Conformité Européenne
CVD: cardiovascular disease
EHR: electronic health record
EU: European Union
HCD: human-centered design
HF: human factors
HR: heart rate
MDR: Medical Device Regulation
mHealth: mobile health

Edited by J Sarvestan; submitted 25.Aug.2025; peer-reviewed by SCL Au, V P?rez Jover; comments to author 24.Nov.2025; revised version received 12.Dec.2025; accepted 15.Dec.2025; published 13.Jan.2026.

Please cite as:
Jacob C, Puthanveettil SR, Vavken P, Kaplan E, Zuern CS
Bridging Gaps in Women's Heart Health: User-Centered Needs Assessment Informed by Patient and Clinician Interviews
JMIR Hum Factors 2026;13:e82916
URL: https://humanfactors.jmir.org/2026/1/e82916
doi:10.2196/82916
PMID:

©Christine Jacob, Sangeetha-Rose Puthanveettil, Patrick Vavken, Emel Kaplan, Christine S Zuern. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 13.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring Video Consultations Across the Public and Private Sectors in Norway: Semistructured Interview Study

Mari Skoge[1,2], ClinPsy; Sofie Ragnhild Aminoff[1], ClinPsy, PhD; Henrik Myhre Ihler[1], MD, PhD; Kari Jorunn Kværner[1,3], MD, PhD; Linn Nathalie Støme[1], MS, PhD; Kristin Lie Romm[1,2], MD, PhD

[1]Early Intervention in Psychosis Advisory Unit for South East Norway, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway

[2]Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway

[3]Department of Strategy and Entrepreneurship, BI Norwegian Business School, Oslo, Norway

Corresponding Author:
Mari Skoge, ClinPsy
Early Intervention in Psychosis Advisory Unit for South East Norway
Division of Mental Health and Addiction
Oslo University Hospital
Sognsvannsveien 21, Bygg 12, 2. etg
Oslo, 0372
Norway
Phone: 47 93624119
Email: maskoge@uio.no

Abstract

Background: Delivering therapy through video consultations can increase the reach and impact of mental health care services. However, adoption varies, and there is a lack of professional consensus about the usefulness of video consultations in therapy settings.

Objective: This study aimed to explore mental health professionals' experiences with and attitudes toward video consultations across different clinical environments in the private and public health care sectors in Norway to inform the design of future digitalized services.

Methods: In this qualitative study, we recruited leaders and clinicians from public hospitals and private clinics. We conducted semistructured interviews that mapped individual experiences and attitudes concerning video consultations, as well as contextual aspects concerning the participants' professional environments. We used reflexive thematic analysis with an inductive, essentialist, and experiential orientation to analyze the data.

Results: A total of 24 mental health professionals (16 from public hospitals and 8 from private clinics) participated. Variations in their attitudes did not follow patterns reflecting the type of service or sector they worked in. Rather, attitudes seemed related to higher-level assumptions rooted in professional culture, societal values, and previous experiences. We generated six themes capturing and structuring the professional perspectives: (1) meta-perspectives on the digitalization of therapeutic rooms, (2) the "how" of service integration, (3) challenging therapist culture, (4) negotiating the limits of the digital therapy room, (5) creating clinical value from the digital format, and (6) adapting techniques and technology in digital therapy sessions.

Conclusions: To strengthen the adoption and impact of video consultations, we should direct attention toward higher-level societal and cultural aspects that shape attitudes and practices. We suggest incorporating digitalized therapy in education, facilitating personal experiences with video consultations, increasing the sharing of knowledge between clinical environments, and sparking innovation of both service models and technology.

(JMIR Hum Factors 2026;13:e80812) doi:10.2196/80812

KEYWORDS

clinician; hybrid psychotherapy; private sector; public sector; telemental health; telepsychiatry; therapist; video consultations; videoconferencing

XSL•FO
RenderX

Introduction

Background

Implementation of video consultations (VC) as an alternative for mental health care delivery addresses several practical barriers to patients' access to treatment [1-4]. The clinical outcomes of therapy delivered through VC are comparable to those of traditional in-person therapy for a wide range of mental health conditions, including depression, suicidal ideation, generalized anxiety disorder, and posttraumatic stress disorder [5-9]. Thus, studies investigating adaptations of psychotherapy practices to the VC format have emerged [10-14]. The literature on VC in services for severe mental illness and other less common mental health conditions is scarcer and inconclusive, but remains optimistic and encourages further research [15-19].

Hybrid therapy—the combination of VC and in-person therapy sessions—allows patients and clinicians to experience the benefits of both formats [20-24]. Previous research on the clinical use of VC has largely focused on comparing the service delivery alternatives in isolation. However, the flexible integration of both VC and in-person treatment modalities is increasingly treated as a promising approach for future mental health care services, potentially providing added value compared to using a single format [25-27]. Studies supporting hybrid models for delivering different types of clinical interventions have appeared in the literature [28-30], but there is a need for more research to develop ideal strategies for integrating VC as a component in routine hybrid services.

The adoption of VC varies significantly across mental health care [31,32], and the attitudes of mental health professionals contribute to this variation [33-36]. Some studies have shown that mental health professionals have largely positive attitudes toward VC, perceiving the format as acceptable and feasible for several clinical purposes [2,33,34,37-41]. Other studies have found greater diversity in perceptions of VC [42,43], and some have identified fundamental drawbacks reported by mental health professionals, such as feeling less present and experiencing a reduction in the quality of the therapeutic relationship [44-46]. Across studies, many report a lack of VC competency or an unmet need for support [14,35,36,47,48]. Furthermore, more experience with VC is associated with attitudes that are more positive [49-51]. Overall, mental health professionals show a preference for the traditional format [24,33,39], which is more evident than that of patients [37,48].

Organizational context is another factor that has been suggested as significant in influencing mental health professionals' attitudes toward and intention to use VC [35]. The public and private health sectors represent distinct organizational conditions. In Norway, the number of private actors offering on-demand mental health care services has increased during the last decade, typically including digital treatment alternatives and opportunities for quicker access to services [52]. In parallel, the proportion of the population with private health insurance (as a supplement to the basic public coverage) has grown significantly [53], which influences the demand in the market for private health services. This development might indicate a more innovative climate, higher responsiveness to user demands, and quick adaptations to digitalization trends among private practitioners. However, we know little about how the private mental health sector approaches VC [54-56]. Health care systems are generally complex and characterized by organizational silos that challenge knowledge sharing across clinical environments [57-59]. Such gaps are highly noticeable within the public health care sector but are even more evident between the public and private sectors [60,61]. Another important reason for the lack of knowledge about private mental health actors is that the private sector in Norway typically has fewer incentives to spend resources on research activities, while the public sector finances and conducts most of the studies on mental health care services.

Study Objectives

In summary, the topic of VC in mental health care lacks professional consensus. In this study, we aim to explore the experiences, attitudes, and contextual aspects related to mental health professionals' use of VC. The study will include professionals working in various services across public and private mental health care to capture a broad range of perspectives. Our ambition is to develop knowledge about factors that contribute to differences in VC adoption and attitudes to inform the design of accessible future health care services that reach more people.

Methods

Study Design

This study is a qualitative investigation of mental health professionals' experiences with and attitudes toward VC, as well as the professional contexts of the participants. The study uses a combination of convenience and snowball sampling strategies [62], digital semistructured interviews, and reflexive thematic analysis [63].

Study Setting

The study was conducted in Norway. The Norwegian national public health system provides universal health coverage. The private health expenditure of the population is generally low (14% in 2021) [64]. Mental health care is provided through primary and specialized services, the latter requiring the presence of moderate or severe psychiatric conditions and a referral from the primary services. Specialized psychiatric services comprise both district psychiatric centers (local outpatient clinics) and hospitals. The services are semidecentralized and organized based on catchment areas, aiming to provide equal access to care across all social strata and geography [65].

The 2 public sector sites of our study are St Olav's Hospital (Trondheim University Hospital) and Oslo University Hospital. The catchment area of St Olav's Hospital consists of both urban and rural areas in Central Norway and is characterized by greater geographical distances between inhabitants and services than the catchment area of Oslo University Hospital, which serves urban areas only. The 4 private sector sites of the study are all located in major cities of Norway. These private clinics provide services to people who are self-referred. Their clients can pay for the services themselves or through private health insurance.

XSL•FO
RenderX

Three of the private clinics in our study allow clients to book consultations directly through their websites, where clients can access information about available therapists and time slots. The fourth clinic requires clients to call, email, or submit a contact form. Two of the private clinics specialize in certain therapy orientations and patient populations, while the others offer more general mental health services. To protect the anonymity of the participants, we have chosen not to disclose the names of these clinics, as they are small, and the mental health professionals could be identified.

Recruitment

The sample of participants in this study was obtained through convenience sampling, which included a snowball strategy [62]. We contacted the leaders of psychiatric clinics at St Olav's Hospital, Oslo University Hospital, and several private clinics and invited them to take part as recruitment sites in our study (see Figures S1-S3 in Multimedia Appendix 1 for more details). The 2 public hospitals and 4 of the private clinics agreed to participate and approved the recruitment and data collection procedures. Two private clinics did not respond to our invitations. Both management and study participants at the clinical sites facilitated further recruitment of participants by informing colleagues about the project and sharing the project group's contact information.

The inclusion criteria were as follows: (1) working as a clinician or working as a leader of other clinicians and (2) employed at one of the public hospitals or affiliated with one of the private clinics. All heads of the units had approved their participation and accepted that the participants spent an hour of their day on the interview. We did not aim to achieve a representative sample to generate representative findings; rather, we aimed to develop empirical thematic insights that were transferable to both research and clinical contexts [66]. We sought to access a broad range of viewpoints and contexts that would allow an exploration of whether expected characteristics (eg, sector affiliation) and other nonpredefined aspects influence approaches to VC. Importantly, we also included participants working in services targeting severe mental illnesses. A lesson learned from our previous work is that researchers and services should consider patients living with severe diagnoses as candidates for digitalized services, although they are often excluded from the target groups of such interventions [67].

Data Collection

Data collection was carried out between May 2024 and December 2024. We conducted digital semistructured interviews addressing individual experiences and attitudes toward VC, as well as the participants' workplace and professional environment. The interviews were based on an interview guide developed by the authors for the purpose of this study (Multimedia Appendix 2). We used the secure web-based VC platform Join, delivered by the state-owned company Norsk helsenett. The duration of the interviews ranged from 45 to 60 minutes. We recorded each interview with an app developed by Services for Sensitive Data (TSD), owned by the University of Oslo. The interviews were automatically transferred to the secure TSD platform and transcribed by the University of Oslo's OpenAI Whisper technology.

Data Analysis

We analyzed the interview transcripts from a critical realist position, applying a method largely based on Braun and Clarke's [63] reflexive thematic analysis. We approached the data with an inductive, explorative, experiential, and contextualist approach [68]. The analysis was data-driven. It placed a direct focus on the participants' utterances rather than on language and latent meaning. A contextualist orientation was a natural choice to capture the features of professional cultures and clinical environments. Choosing the reflexive school of thematic analysis, we aimed to transparently communicate the impact of researcher participation, as well as analytic decisions and processes. We selected this method with the ambition of developing meaningful themes that can carry transferable insights across the mental health field [66]. Moreover, the analysis was conducted from a pragmatic perspective that considered current and future clinical practices. We followed the 6 phases of thematic analysis outlined by Braun and Clarke [69], with some minor adaptations (Multimedia Appendix 3).

User Representation

Two user representatives with lived experience, recruited through the Bipolar Association Norway, contributed to this study. The representatives provided feedback on the study in the early design phase and during the writing process. The purpose of this involvement was to ensure that the research aims, the methods applied to achieve these, and the angle used to present and discuss the findings are relevant, interesting, and valuable to people living with mental disorders, including severe conditions such as bipolar and psychotic disorders.

Ethical Considerations

The project obtained approval from the data protection officer at Oslo University Hospital (24/04658). All participants provided written consent to take part in the study. The consent forms included information about the purpose of data collection and how the data would be stored and processed. The forms emphasized voluntary participation, including the option to leave the study at any time without risking any consequences. Signed consents and interview transcripts were stored in the secure TSD database. To protect the privacy of participants, identifiable features of participants were removed by MS within the secure TSD environment before the data were exported for analysis.

The economic compensation offered only to the participants working in private clinics entails ethical consideration. Importantly, private practitioners' financial models conflict with involvement in research, which contributes to the maintenance of a knowledge gap concerning private health care services. Moreover, the project group holds the view that the compensation was not high enough to create any form of pressure for the private practitioners to participate in the study, only enough to compensate for lost income during the interview. Therefore, we evaluate the economic compensation to the participants working in private clinics as both ethically acceptable and reasonable.

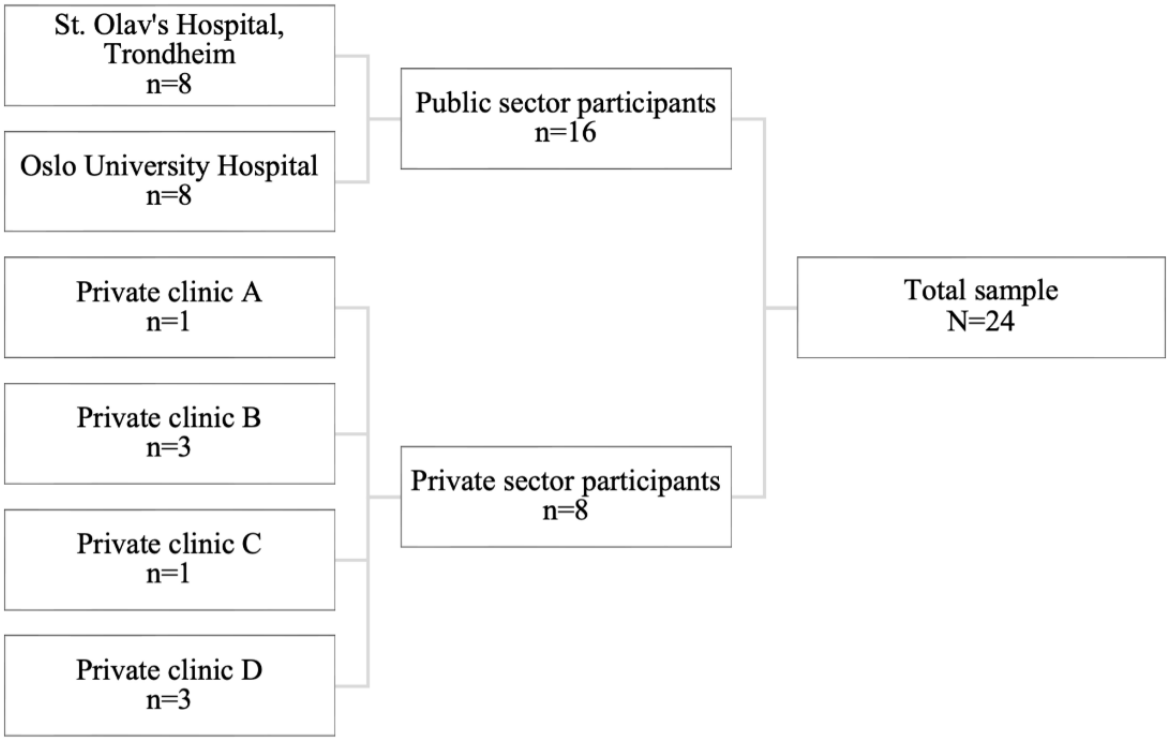
XSL•FO
RenderX

Results

Overview

A total of 24 mental health professionals participated in our study (Figure 1 and Table 1), representing different demographic characteristics, geographical catchment areas, professional backgrounds, types of services, and target patient populations, as well as affiliation to either the public or private health sector. From the earliest phases of the analysis, we observed great variation in the participants' attitudes toward VC and in the amount of personal experience with VC as a therapy format. Figure 2 is a result of early familiarization with the data material (Multimedia Appendix 3). The figure provides a simplified overview of the 3 subgroups in regard to levels of VC experience and attitudes. The figure is based on subjective and exploratory screening of the main tendencies in the participant accounts. The most extreme values on each of the axes represent the most extreme participant views identified in this study sample, and the position of each data point is relative to the others. Taken together, the distribution of data points in the figure suggests that there is a tendency that VC experiences and VC attitudes are related and that the relationship between the two concepts is stronger at more extreme levels. Moreover, the figure indicates differences in the amount of experience with VC between the 3 subgroups. Nevertheless, a continuum of attitudes is represented in each group. Early screening of attitudes in relation to participants' characteristics indicated that variations in attitudes did not systematically follow affiliation with the public or private sector or the type of service the mental health professionals worked with.

Figure 1. Chart of recruitment sites and study participants.

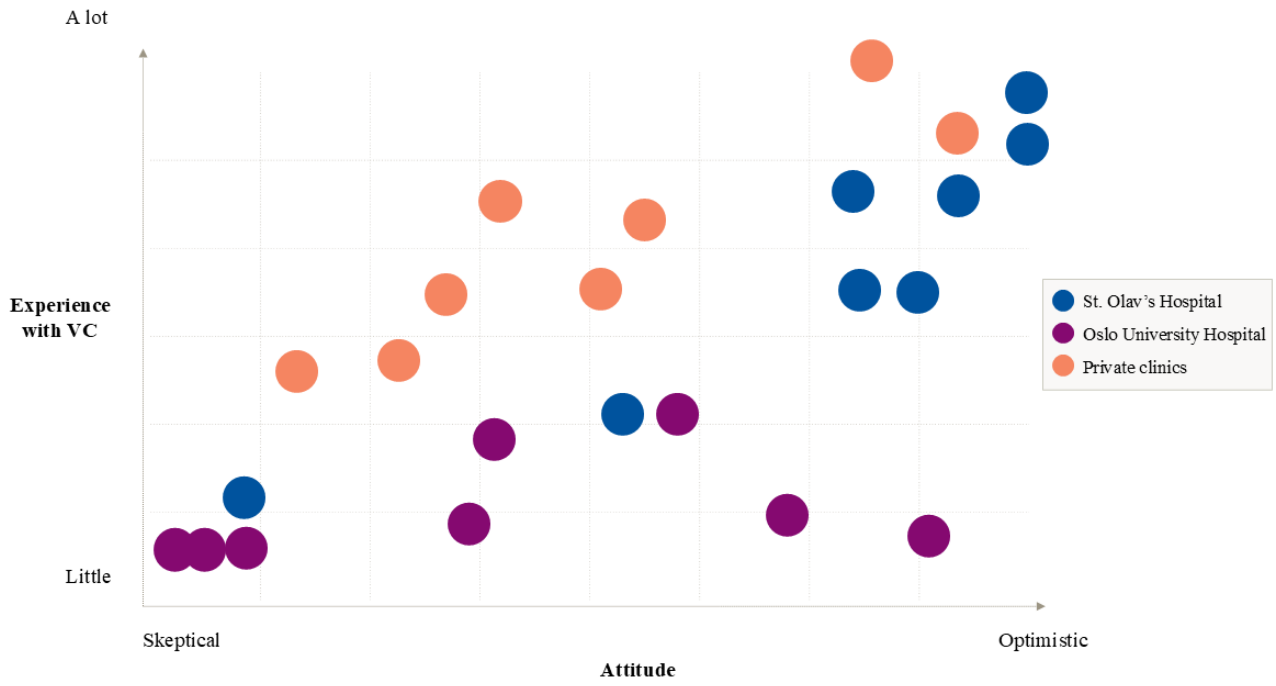


XSL•FO
RenderX

Table 1. Participant characteristics (N=24).

Characteristics	Values
Participant age (years), mean (SD; range)	45.5 (9.9; 31-65)
Sex assigned at birth, n (%)	
Female	17 (71)
Male	7 (29)
Role, n (%)	
Clinician	16 (67)
Leader (few clinical tasks in everyday workflow)	3 (12)
Combined (significant amount of clinical tasks in everyday workflow)	5 (21)
Professional background, n (%)	
Specialist in clinical psychology	9 (38)
Licensed clinical psychologist	5 (21)
Psychiatrist	4 (17)
Resident doctors in psychiatry	3 (12)
Psychiatric nurse	2 (8)
Nurses	1 (4)
Type of mental health service, n (%)	
Open booking/no referral	8 (33)
Psychotic disorders	5 (21)
General psychiatric services	4 (17)
Short-term general psychiatric services	3 (12)
Bipolar disorder	3 (12)
Currently not working in a clinical service	1 (4)

Figure 2. Early qualitative mapping of the experiences and attitudes of participants. VC: video consultations.



XSL•FO
RenderX

Through the following phases of the reflexive thematic analysis, we constructed a set of 6 themes, with each theme belonging to a distinct level of abstraction (Textbox 1).

Textbox 1. Themes from the reflexive thematic analysis listed according to level.

Meta level

- Theme 1: meta-perspectives on the digitalization of therapeutic rooms

Service level

- Theme 2: the "how" of service integration

Therapist culture level

- Theme 3: challenging therapist culture

Clinical level

- Theme 4: negotiating the limits of the digital therapy room
- Theme 5: creating clinical value from the video format

Practice level

- Theme 6: adapting techniques and technology in digital therapy

Theme 1: Meta-Perspectives on the Digitalization of Therapeutic Rooms

The first theme connects the practice of VC to the overarching digitalization trends in society and health care services. The participants demonstrated opposing underlying values that naturally influenced their standpoints at lower levels (ie, the 5 other themes). Interestingly, these meta-perspectives did not seem to follow other key characteristics of the participants, such as age, gender, private or public workplace, clinic, or severity of the condition of their patient population.

Some viewed the increased attention paid to VC in mental health care as a symptom of the fast-paced digitalization of our communities, which reduces arenas for healthy ways of experiencing and relating to others. Using this rationale, VC practices contribute to an unwanted development toward more individualism and isolation. Thus, there is a need to safeguard some of these analogue spaces, especially traditional therapy rooms, which might be particularly important.

> *Personally, and on behalf of others, I find it very exhausting to interact so much digitally. It has an overhead cost, whose extent I think we don't fully understand yet.* [B4, resident doctor in psychiatry, public hospital]

> *Digital (encounters) enhance the extremely rational, intellectualizing way we relate to ourselves and others, which I believe results in more psychopathology.* [D1, specialized psychologist, private clinic]

The opposite viewpoint places digitalization in relation to new dynamics in the distribution of power and agency between patients and services. Since offering digital treatment can increase access to care and help balance the relational asymmetry of health care services, this viewpoint treats the expansion of therapy alternatives as a moral obligation.

> *They experience a higher degree of autonomy and control when they meet me at a screen's distance.* [A1, psychiatrist, public hospital]

> *We deprive some patients of the opportunity to show up—patients who would otherwise struggle because they must enter our arena. There is already an imbalance—an asymmetrical relationship. Therefore, I think we owe it to them.* [A2, specialized psychologist, public hospital]

Theme 2: The "How" of Service Integration

The second theme captures the diversity in VC integration strategies and emphasizes the relevance of exploring creative possibilities of VC use further. The data material contained descriptions of various service models involving VC in different ways, serving different purposes. See Textbox 2 for an overview of the service models identified in the study. The list demonstrates that using VC does not mean one thing. Rather, VC can be adapted to the specific needs of different clinical contexts. In practice, the participants often applied a combination of elements from the listed service models. We identified the most evident differences between the sectors in these dimensions. Some models were more available or exclusively implemented in the private sector, such as the digital hub of specialists. The hub allows personalization on several levels.

XSL•FO
RenderX

Textbox 2. Service models that integrate videoconferencing identified in the data material.

Only digital

- All consultations in a course of treatment are digital. One private therapist worked exclusively with video consultations (VC) from home. One public clinician worked with a fully digital intervention for relatives.

Digital first session

- All new patients attend their first session through VC. Later, they decide on the format. Used in a public service for bipolar disorder that served many patients referred directly from emergency wards.

Hybrid strategy (relational booster rationale)

- Hybrid courses of treatment where traditional consultations comprise a minority of meetings and are used to strengthen the therapeutic relationship. Patients typically live far away.

Hybrid strategy (content rationale)

- Hybrid courses of treatment where the format depends on the session's content. For instance, parts of psychiatric assessments where only verbal reports are needed can indicate a VC session.

Hybrid strategy (practical rationale)

- Hybrid courses where there is a preplanned strategy of using both formats to reduce travel or absence from daily activities. Target patients may have a network and activities they attend regularly.

Hybrid strategy (symptom-based rationale)

- Hybrid courses where anxiety-related symptoms are barriers to receiving services, and therapy starts off as VC, while the proportion of physical meetings gradually increases based on symptom reduction.

Hybrid strategy (staff flexibility rationale)

- Clinicians work from home on days when only digital appointments are scheduled to increase variety, flexibility, and satisfaction with work.

Ad hoc hybrid courses (short-notice changes)

- Traditional sessions may change into digital ones at short notice due to patients experiencing unanticipated events or problems with executive functions. VC sessions secure continuity.

Open booking (only private clinics)

- A client-centered service that allows anyone to book a VC session at a suitable time with the therapist they prefer, typically directly through a website. Some clinics also offer different durations of sessions.

Digital hub (only private clinics)

- Clinics provide information about therapists' competencies and preferred therapy orientations, offering a menu of therapeutic alternatives from which clients can choose the best match for their needs.

> *Here, you can choose someone who matches where you are in the therapy process, what your goals are, and how you like to work. I think this leads to faster improvement compared to going to someone at random.* [D1, specialized psychologist, private clinic]

Furthermore, working independently with more flexible opening hours in a private clinic allows more staff flexibility for practitioners:

> *It is nice to have some days at home, deciding for myself how I want to structure the day. I might have a consultation at nine, the next consultation at one. Then one at six, and the last one at eight. Depending on what the client needs. And meanwhile, I can just live my life.* [E2, psychologist, private clinic]

Hybrid, preplanned strategies for delivering therapy for practical and symptom-related reasons were applied across the sample. However, a few therapists also talked about using traditional sessions to boost the relational quality of mainly digital courses of treatment, where the effect of the physical conversation would be maintained for a while:

> *... then we can keep leveraging from that (physical) booster session in the other (digital) conversations.* [E3, specialized psychologist, private clinic]

An insight from this dimension was that the clinicians' presentation of how a VC service could work out is important, as it can shape patients' expectations. One of the participants worked with a completely digital support service for the relatives of patients with severe mental illness. The clinician described how the relatives accepted the digital service they were offered and that there were no requests for in-person meetings:

> *The thing about the (digital service to relatives) conversations is that they are video-only. Right? Even*

XSL•FO
RenderX

> *from the first session. So, you go into it with the expectation that we will only meet on video.* [B5, psychiatric nurse, public hospital]

Theme 3: Challenging Therapist Culture

The third theme is based on the participants' descriptions of VC as a deviation from tradition and something that challenges assumptions established in professional culture. Some participants described video practice as a threat to therapists' integrity, associating it with something less therapeutic than traditional therapy. Others said that VC could trigger a feeling of helplessness and less confidence in their role as therapists, a threat to their professional identity.

> *It's kind of taboo in a way ... I think some view it as somewhat poorer. That you don't really connect with the patient properly, when you don't meet and look this person properly in the eyes.* [B7, nurse, public hospital]

> *There is sometimes a need to initiate acute interventions, and then you don't necessarily get in the right position to act ... you feel more helpless on video.* [B2, resident doctor in psychiatry, public hospital]

Nevertheless, several participants challenged their preconceptions through experience and developed new, accepting approaches to VC. They reflected on the norms and myths of psychotherapy and identified culture-based assumptions about digital therapies:

> *If people were used to sitting at home and talking, and that was the standard, I am pretty sure going to an office would have felt strange.* [A6, specialized psychologist, public hospital]

> *It has become imperative that the patient comes to the outpatient clinic and stays there for 45 minutes. I do not have any evidence to state that it is very effective or useful for either the patient or us. ... It is deeply ingrained in healthcare culture. Things like video consultations are not part of education programs. ... This influences what one thinks as a clinician.* [A7, specialized psychologist, public hospital]

As observed early in the familiarization process of the dataset, there seemed to be a trend that the most experienced VC therapists held more positive views, and vice versa. However, we also observed that private practitioners with more experience than employees from public hospitals conveyed diverse opinions. Interestingly, in private clinics, patients decide on the therapy format themselves when they book a consultation. Thus, some private practitioners may practice an even amount of VC therapy, even if they subscribe to a therapist culture that favors the physical therapy space.

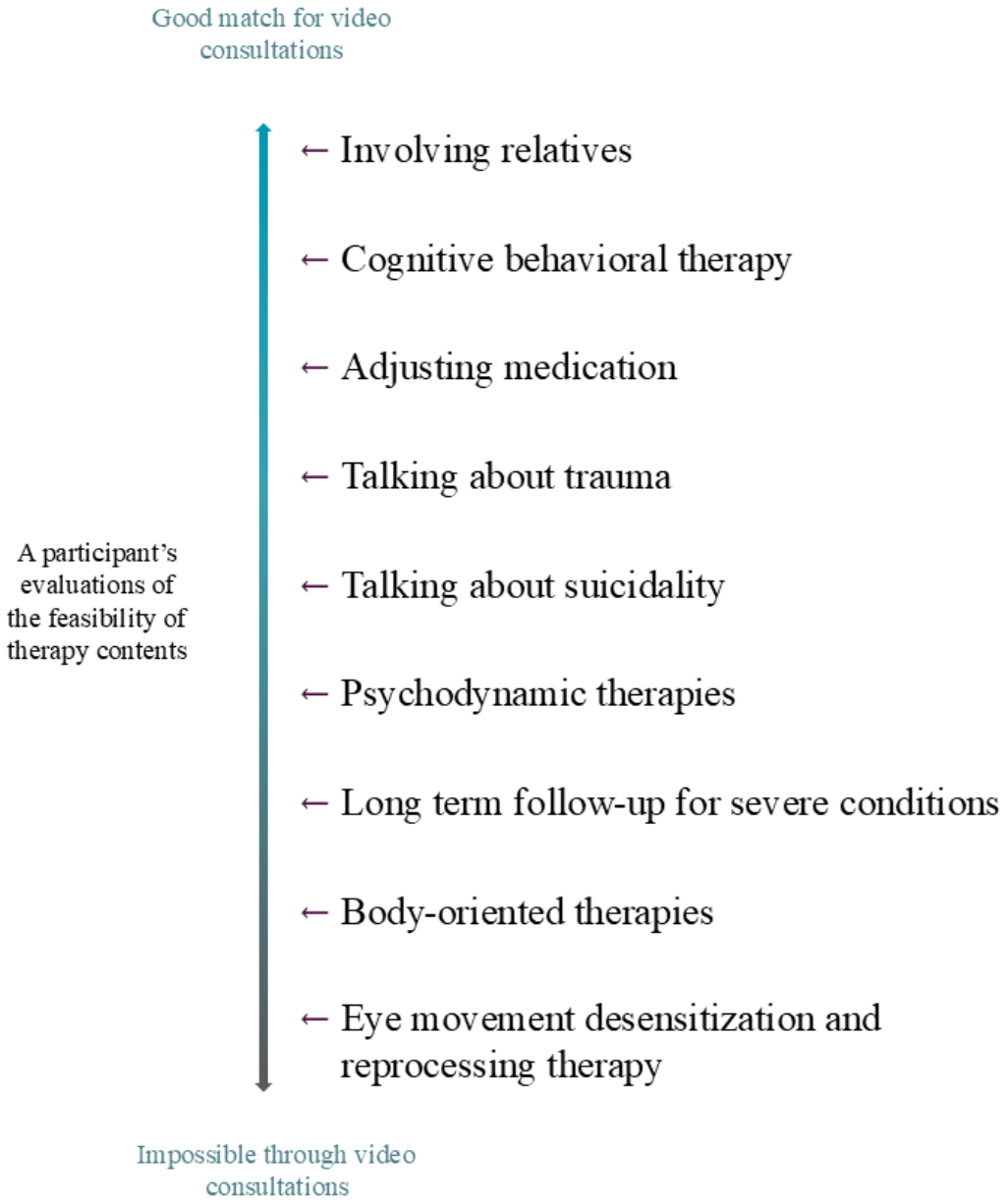
Theme 4: Negotiating the Limits of the Digital Therapy Room

Theme 4 demonstrates the contrasts in professional opinions and rationales regarding the clinical purposes for which VC is suitable. All participants acknowledged that the digital therapy room was less appropriate when patients lacked social contact and activities in their daily lives or needed exposure to work with avoidance. Apart from this, their reflections on the feasibility of different therapy contents varied.

Figure 3 illustrates the continuum of the perceived feasibility of different therapy contents when delivered through VC, based on a participant who worked within the cognitive therapy orientation. The case presented in the figure demonstrates how participants typically had clear logic and presented a continuum of feasibility, defining the limits of VC. From the viewpoint in this case, eye movement desensitization and reprocessing therapy (EMDR) is the least feasible intervention using VC.

XSL•FO
RenderX

Figure 3. Illustration of a typical continuum of feasibility concerning different therapy contents when delivered through videoconferencing, based on 1 participant.



However, others expressed contrasting opinions. These perspectives on feasibility are demonstrated in the following extracts:

> *There are certain things that are not easy to do on video—that cannot be done on video. For example, trauma therapy with EMDR. It is completely impossible.* [A3, psychiatrist, public hospital]

> *I might think that EMDR works better digitally than other forms of therapy, because there are very few words.* [D1, specialized psychologist working with EMDR in traditional sessions, private clinic]

Similarly, while most participants talked about patients with psychotic disorders as the patient population least probable of experiencing value from VC, one of the clinicians working with this patient group remarked,

> *I think the patient can often surprise the clinician in that area, and they can often be better (than what the clinicians expect).* [B7, nurse, public hospital]

Finally, some mental health professionals demonstrated the importance of understanding the difference between the feasibility of carrying out interventions and the actual therapeutic value that is realized. In the context of services aimed at patients struggling with severe diagnoses and complex life challenges, a VC may not achieve the same outcomes as in-person contact. A participant stated that services and clinicians are at risk of misleading themselves in the following way:

> *"Now that we have digital tools, we actually get in touch with the patient. The patient attends the appointments. We are now in a position where we can talk to this person." However, the patient just sits there in their municipal apartment, isolated,*

XSL•FO
RenderX

> *without needing to go out. "But at least we get in touch with him!" I imagine that we might end up just believing that we have done something good.* [B4, resident doctor in psychiatry, public hospital]

Theme 5: Creating Clinical Value From the Video Format

The fifth theme revolves around the process of realizing therapeutic value from VC. The data material contained examples demonstrating that the video format, in some cases, comes with unique additional clinical benefits:

> *I have many clients whom I have never met in person. However, it was particularly interesting that with those I had met in person, I felt that more happened when we could talk on video, and I experienced more therapeutic breakthroughs with them after we started with video consultations. I found that a bit interesting.* [E2, psychologist, private clinic]

Several clinicians explained that the practical benefits of the video format mediated some of the extra therapeutic value. When patients do not need to spend much time travelling or spend energy overcoming practical obstacles and stress related to a physical meeting, more resources may be available for the purpose of the therapy session. Moreover, remote sessions often come with proximity to family members, which can result in more effective therapeutic processes:

> *They might have to sit on a bus for two hours each way, and then the entire consultation would end up talking about how terribly frustrating it had been. When you have a video consultation, you can really concentrate on what the underlying problem is.* [A3, psychiatrist, public hospital]

> *I had the opportunity to do things more quickly. I found out that it might be a good idea to talk to (the patient's) relative just to calm the situation and asked if that relative was present at home. The patient said yes, and then I got 15 minutes with the relative right away. And then that situation was solved. It wasn't an issue anymore.* [A6, specialized psychologist, public hospital]

Another application of VC is as an aid to regulate patients' level of activation. Facilitating a safe, less confronting therapy environment reduces the level of activation, providing patients with more opportunities to take part in and experience the effects of emotionally challenging therapeutic interventions:

> *If there is a really difficult conversation with a client where we are going to talk about our alliance, I might take the client out of the therapy room to help downregulate feelings, not having to maintain eye contact. We can use video to do the same. Sometimes, when we want feelings to be more intense, I prefer to be in the office.* [E1, psychologist, private clinic]

Theme 6: Adapting Techniques and Technology in Digital Therapy Sessions

The sixth theme is placed at the concrete level of therapeutic practice, covering pragmatic adaptations of therapy interventions and technical aspects of the technologies used for VC. One clinician always wore headphones during VC to optimize the dynamics of the conversation, rather than using the speakers:

> *I get the sound directly in the ears instead of it coming from speakers. That makes a difference.* [E3, specialized psychologist, private clinic]

Another clinician talked about how basic VC therapy skills provide great opportunities to carry out therapeutic interventions that traditionally depend on the physical environment of the therapy room. This clinician emphasized the importance of careful verbal guidance and attentive assessment of the patient's needs during VC:

> *I don't feel (that the video format) is an obstacle because I can guide them and see if there's something they need, find a slightly better chair to sit on, or if they have something they can hold in their hand, or if they're working on regulating themselves a bit. Grab a pillow or something like that. They can do that at home as well.* [D3, specialized psychologist using body-oriented methods, private clinic]

Several participants shared the opinion that there is potential for improvement for future VC practices. Some said that simply improving the information about VC directed at patients could strengthen implementation. Others talked about potential improvements in the technology; for example, how AI might enhance the liveliness of VC. Finally, multiple participants talked about the lack of safe and clinically validated software that can mimic whiteboards and paper-based tools that are typically used in traditional assessment and therapy sessions. Such technology already exists and is integrated into VC systems used in other professional contexts, but is not available for clinical use in mental health care.

> *If we get an AI as a video chat function, then we could have more visual dynamics. And I think these dynamics will elevate the experience.* [B2, resident doctor in psychiatry, public hospital]

> *If there were a function where you could incorporate questionnaires or some kind of interactive board in the video call (...) That could perhaps also address some challenges related to forms and documentation.* [A8, specialized psychologist, public hospital]

Discussion

Principal Results

This study explores experiences, attitudes, and current debates related to the role of VC in mental health care and challenges preconceptions identified in our own research environment and beyond. To our knowledge, this is the first qualitative study to explore the experiences and attitudes of mental health professionals across public and private settings on the topic of VC. Moreover, the study included perspectives from a broad selection of services, ranging from those targeting bipolar and psychotic disorders to on-demand services targeting clients with milder conditions and not requiring referrals. Our study identifies differences in VC implementation, experiences, and professional attitudes that support the existing literature

XSL•FO
RenderX

[31,32,42,43]. Contrary to the authors' expectations, key characteristics such as public and private sector affiliation and symptom severity of the target patient populations did not seem to shape the participants' views to a significant degree. Building on this, our study points out important areas to direct future efforts to increase the adoption and impact of VC.

Primarily, our findings demonstrate that higher-level aspects, such as societal values and cultural perceptions, are important in shaping professional attitudes. Psychotherapy and the therapist role are defined through the cultural value systems in society and in professional environments [70]. Thus, fundamental changes in mental health care services, such as the digitalization of the delivery of therapy, can be experienced as a cultural deviation from traditional models of psychotherapy [71,72]. In accordance with this notion, several participants talked about how traditional ideas about psychotherapy weaken the credibility of VC, although many forms of VC-delivered psychotherapy are supported by robust evidence [5-9]. This tendency indicates a lack of cultural familiarity with digitalized treatment and a lack of knowledge about the evidence-based applications of VC.

While the status of VC in clinical environments remains unclear, polarized approaches have been developed. In our study, the participants with the most critical attitudes typically conceptualized VC as part of purely digital, fast-paced, cost-reduction–focused services and, thus, a symptom of an unwanted development in health care. Research on the resistance to digitalized mental health services has identified similar viewpoints of clinicians who experience that technology conflicts with fundamental humanistic values in mental health care, resulting in resistance and stigma related to the digitalization of therapy [72]. This approach facilitates an all-or-nothing view of digitalized services instead of a nuanced perception of VC as a flexible tool that can be integrated in various ways in hybrid services, a model that is becoming increasingly applied in the literature [25-27].

Interestingly, even among the participants who held the most positive attitudes, we found conflicting views on whether VC is feasible for patients with severe conditions. These clinicians described the patient group as vulnerable and conveyed that it was their professional responsibility to protect these patients accordingly. However, adjustments to the service alternatives based on the vulnerable status of patient groups may not always result in better outcomes for the patients [73]. Furthermore, the exclusion of such patient populations may be based on stigmatic perceptions of the patients, which have been found to be relatively frequent among clinicians working in mental health care services [74,75]. Differentiating the treatment offered based on patients' diagnoses may preserve such preconceptions. Although the existing evidence pertaining to VC in services targeting these populations is still scarce and calls for more research [15-19], we encourage an inclusive and personalized approach and advocate refraining from excluding patients based on diagnosis alone.

Importantly, the education of mental health professionals plays a fundamental role in defining the scope of psychotherapy and conceptualizing VC. Currently, VC is mostly absent from the training of mental health personnel, which may promote more trust in traditional cultural perceptions of therapy and the limits of VC. Moreover, mental health professionals report low self-efficacy and a lack of training and support in clinics [14,35,36,47,48]. At the same time, the research field continues to produce new knowledge about VC, such as through studies focused on adapting basic therapeutic skills to the video format and studies on how we can maintain a feeling of connectedness in video sessions [10,14]. In addition, the body of literature covers adaptations of interventions within therapy orientations that are often closely associated with traditional settings and that emphasize sensory experiences and the atmosphere of the therapy room, such as psychodynamic transference [11], emotion-focused therapy exercises [12], and EMDR [13]. Our findings encourage the inclusion of knowledge about VC in the basic training of professionals to ensure that VC is incorporated from an early point.

We also found that the amount of personal experience with VC plays an important role in shaping attitudes, as it directly challenges established cultural assumptions. Several participants expressing optimistic attitudes stated that they had not been able to perceive VC as feasible and valuable before they had practiced it themselves and adjusted their preconceptions. These observations partly align with previous research showing that more experience with VC is associated with positive attitudes and less self-doubt as a therapist [49-51]. These findings are supported by research suggesting that one of the main drivers of adoption is clinicians' own personal experiences, proving that digital tools can result in clinical benefits [76]. Together with the literature, our findings emphasize the importance of personal experience to perceiving the value of VC. Thus, we suggest that services become aware of this tendency, demonstrate the beneficial effects for both service users and providers, and develop incentives for clinicians to integrate VC in their practice.

Finally, our study presents a range of creative and innovative therapy practices and service designs, suggesting that there is potential synergistic value in sharing local insights between professional environments. Several of the service models identified have been implemented only to a very limited degree across the clinical sites of the study, such as the digital hub or digital first sessions (see Theme 2 in the Results section). Some of these service designs have been mentioned in the literature but have not been tested in public settings in Norway [27]. Using insights and service models from both the private sector and innovative public sites may be valuable in future decision-making regarding service designs in public health care. Interestingly, few innovative activities revolve around the technologies used for VC, and clinicians report that the platforms they use lack the basic tools used in traditional therapy [24,41,46]. This indicates that there is room for improvement and further innovation in terms of clinical VC software, which has barely changed since the introduction of VC in mental health care.

Limitations

Although the method chosen for this study does not aim to produce generalizable findings, and although our findings can

XSL•FO
RenderX

still be transferable, we wish to explicitly acknowledge that the composition of our sample may have moved the analytical focus and the results in a more optimistic and progressive direction. For instance, several participants recruited from one of the public hospitals were involved in research and development, while two were involved in digitalization projects. We are also aware that the participants who decided to participate might have had more positive attitudes toward the topic of the research project than those who did not want to participate. This, combined with the authors' optimistic attitudes and roles in other digitalization projects closely related to the topic of this study, possibly influenced the angle we chose for this paper, which encourages digitalization in mental health care. Another limitation relates to the practical transferability of findings produced in this study in an international context, as all participants were recruited from Norwegian health care sites. Furthermore, the analysis is not exhaustive in discussing aspects that have an impact on the participants' viewpoints. For instance, regarding the private actors' contextual characteristics, the paper discusses the impact of incentives for practicing VC and whether VC use is voluntary or not to a very limited degree. An important research-related issue resulting from such differences in public and private sector contexts revolved around the recruitment for this study. Participants from private clinics received economic compensation of about US $110 because of slow recruitment due to the nature of the private practice financial model. Lost activity with paying clients results in reduced income, which is a barrier related to participation in research projects. The other participants did not receive similar economic compensation since they were all employed at university hospitals, where participation in research and development projects is integral. Future research should explore the effects of different organizational conditions in detail, which may be of great importance in developing attractive workplaces for mental health professionals and sustainable hybrid services.

Conclusions

This qualitative exploration emphasizes the multidimensional nature of professional perspectives on VC. Our findings underline how overarching societal and cultural ideas influence professional attitudes, logic, and, thus, clinical practice. Our ambition is that the paper challenges preconceptions about digitalized therapy and that it points to useful strategies for targeting problems concerning the limited adoption of VC. Educational institutions and services must promote a nuanced, evidence-based concept of VC and demonstrate benefits for both service users and providers. In addition, incentives for mental health professionals to practice VC should be developed and presented. There is unrealized potential in applying existing therapy adaptations and hybrid service models established in separate parts of health care systems, particularly for public clinics. In addition, there is room for innovation with regard to service designs that integrate VC and the technology platforms used to create unique value in hybrid services and to increase the attractiveness of the VC format in mental health care.

Acknowledgments

We wish to thank all the study participants and the two user representatives for their contributions to this study. This study is funded by the Research Council of Norway through grant 309264 (Forhelse Research Centre for Digital Health Services, a centre for research-based innovation).

Data AVailability

To protect the anonymity of the participants, the qualitative dataset generated during this study will not be available to readers. More information concerning the study settings is available from the corresponding author on reasonable request.

Authors' Contributions

MS wrote the manuscript. The other authors reviewed, modified, and suggested changes during the writing process. KLR, SRA, LNS, KJK, and MS designed and planned the study. KLR, SRA, HMI, and MS recruited participants. MS was the main interviewer. KLR participated as a supervisor and cointerviewer in the first 3 interviews. MS led the analysis with guidance from the other authors. KLR is the main supervisor of MS as a PhD candidate. SRA and LNS are cosupervisors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary figures.
[DOCX File , 80 KB - humanfactors_v13i1e80812_app1.docx]

Multimedia Appendix 2
Interview guide.
[DOCX File , 17 KB - humanfactors_v13i1e80812_app2.docx]

Multimedia Appendix 3

XSL•FO
RenderX

Detailed description of the analysis.
[DOCX File , 18 KB - humanfactors_v13i1e80812_app3.docx]

References

1. Picillo B, Yu-Lefler H, Bui C, Wendt M, Sripipatana A. Telehealth-facilitated mental health care access and continuity for patients served at the Health Resources and Services Administration-funded health centers. Telemed J E Health 2025;31(7):838-847. [doi: 10.1089/tmj.2025.0011] [Medline: 40014364]
2. Marini C, Steinberg MH, Miron CD, Irwin M, Schantz BL, Ginsberg DL, et al. Examination of patient and provider satisfaction, benefits, and challenges with psychiatric outpatient and hospital-based telehealth treatment during the COVID-19 pandemic. Psychiatr Q 2025;96(4):869-885. [doi: 10.1007/s11126-025-10150-w] [Medline: 40268852]
3. Cummings JR, Kalk T, Trello S, Walker ER, Graetz I. Therapists' perspectives on access to telemental health among medicaid-enrolled youth. Am J Manag Care 2023;29(11):e339-e347 [FREE Full text] [doi: 10.37765/ajmc.2023.89430] [Medline: 37948654]
4. Galvin E, Desselle S, Gavin B, Quigley E, Flear M, Kilbride K, et al. Patient and provider perspectives of the implementation of remote consultations for community-dwelling people with mental health conditions: a systematic mixed studies review. J Psychiatr Res 2022;156:668-678 [FREE Full text] [doi: 10.1016/j.jpsychires.2022.10.051] [Medline: 36399859]
5. Hagi K, Kurokawa S, Takamiya A, Fujikawa M, Kinoshita S, Iizuka M, et al. Telepsychiatry versus face-to-face treatment: systematic review and meta-analysis of randomised controlled trials. Br J Psychiatry 2023;223(3):407-414 [FREE Full text] [doi: 10.1192/bjp.2023.86] [Medline: 37655816]
6. Oh J, Ho J, Lee S, Park J. Effects of digital psychotherapy on suicide: a systematic review and meta-analysis. Healthcare (Basel) 2024;12(14) [FREE Full text] [doi: 10.3390/healthcare12141435] [Medline: 39057578]
7. Shaker AA, Austin SF, Storebø OJ, Schaug JP, Ayad A, Sørensen JA, et al. Psychiatric treatment conducted via telemedicine versus in-person modality in posttraumatic stress disorder, mood disorders, and anxiety disorders: systematic review and meta-analysis. JMIR Ment Health 2023;10:e44790 [FREE Full text] [doi: 10.2196/44790] [Medline: 37277113]
8. Giovanetti AK, Punt SEW, Nelson E, Ilardi SS. Teletherapy versus in-person psychotherapy for depression: a meta-analysis of randomized controlled trials. Telemed J E Health 2022;28(8):1077-1089. [doi: 10.1089/tmj.2021.0294] [Medline: 35007437]
9. Krzyzaniak N, Greenwood H, Scott AM, Peiris R, Cardona M, Clark J, et al. The effectiveness of telehealth versus face-to face interventions for anxiety disorders: a systematic review and meta-analysis. J Telemed Telecare 2024;30(2):250-261. [doi: 10.1177/1357633X211053738] [Medline: 34860613]
10. Kumar MS, Krishnamurthy S, Dhruve N, Somashekar B, Gowda MR. Telepsychiatry netiquette: connect, communicate and consult. Indian J Psychol Med 2020;42(5 Suppl):22S-26S [FREE Full text] [doi: 10.1177/0253717620958170] [Medline: 33354059]
11. Sayers J. Online Psychotherapy: transference and countertransference issues. Brit J Psychotherapy 2021;37(2):223-233. [doi: 10.1111/bjp.12624]
12. Banack KD. Adapting emotion-focused therapy for teletherapy. Pers Centered Exp Psychother 2021;20(4):303-311. [doi: 10.1080/14779757.2021.1993968]
13. Kaptan SK, Kaya ZM, Akan A. Addressing mental health need after COVID-19: a systematic review of remote EMDR therapy studies as an emerging option. Front Psychiatry 2023;14:1336569 [FREE Full text] [doi: 10.3389/fpsyt.2023.1336569] [Medline: 38250261]
14. Werbart A, Dominicus E, Sheahan L, Philips B. When the screen deepens the sense of connectedness: what therapists have found effective in successful cases of teletherapy. Couns and Psychother Res 2024;25(1). [doi: 10.1002/capr.12820]
15. Farrell A, George N, Amado S, Wozniak J. A systematic review of the literature on telepsychiatry for bipolar disorder. Brain Behav 2022;12(10):e2743 [FREE Full text] [doi: 10.1002/brb3.2743] [Medline: 36102239]
16. Greenwood H, Krzyzaniak N, Peiris R, Clark J, Scott AM, Cardona M, et al. Telehealth versus face-to-face psychotherapy for less common mental health conditions: systematic review and meta-analysis of randomized controlled trials. JMIR Ment Health 2022;9(3):e31780 [FREE Full text] [doi: 10.2196/31780] [Medline: 35275081]
17. Randall L, Raisin C, Waters F, Williams C, Shymko G, Davis D. Implementing telepsychiatry in an early psychosis service during COVID-19: experiences of young people and clinicians and changes in service utilization. Early Interv Psychiatry 2023;17(5):470-477 [FREE Full text] [doi: 10.1111/eip.13342] [Medline: 35943177]
18. Lal S, Abdel-Baki A, Lee H. Telepsychiatry services during COVID-19: a cross-sectional survey on the experiences and perspectives of young adults with first-episode psychosis. Early Interv Psychiatry 2023;17(4):368-377 [FREE Full text] [doi: 10.1111/eip.13332] [Medline: 35731034]
19. Best MW, Romanowska S, Zhou Y, Wang L, Leibovitz T, Onno KA, et al. Efficacy of remotely delivered evidence-based psychosocial treatments for schizophrenia-spectrum disorders: a series of systematic reviews and meta-analyses. Schizophr Bull 2023;49(4):973-986 [FREE Full text] [doi: 10.1093/schbul/sbac209] [Medline: 37066769]
20. Steidtmann D, McBride S, Mishkind M, Shore J. From couch to click: opportunities and challenges of hybrid work models in mental health care. Curr Psychiatry Rep 2025;27(6):385-392. [doi: 10.1007/s11920-025-01607-7] [Medline: 40186706]

XSL•FO
RenderX

21. Yellowlees P, Richard Chan S, Burke Parish M. The hybrid doctor-patient relationship in the age of technology - telepsychiatry consultations and the use of virtual space. Int Rev Psychiatry 2015;27(6):476-489. [doi: 10.3109/09540261.2015.1082987] [Medline: 26493089]
22. Aafjes-van Doorn K, Békés V, Stukenberg KW, Roberts KE. Therapists' views and recommendations on the ins and outs of practicing teletherapy. Couns Psychol Q 2023;37(3):398-418. [doi: 10.1080/09515070.2023.2227133]
23. Honey A, Hines M, Barton R, Berry B, Gilroy J, Glover H, et al. Preferences for telehealth: a qualitative study with people accessing a new mental health service. Digit Health 2023;9:20552076231211083 [FREE Full text] [doi: 10.1177/20552076231211083] [Medline: 37928331]
24. Leuchtenberg S, Gromer D, Käthner I. Videoconferencing versus face-to-face psychotherapy: insights from patients and psychotherapists about comparability of therapeutic alliance, empathy and treatment characteristics. Couns Psychother Res 2022;23(2):389-403. [doi: 10.1002/capr.12538]
25. Chen K, Huang JJ, Torous J. Hybrid care in mental health: a framework for understanding care, research, and future opportunities. NPP Digit Psychiatry Neurosci 2024;2(1):16. [doi: 10.1038/s44277-024-00016-7] [Medline: 41360926]
26. Nwokedi CN, Olowe KJ, Alli OI, Iguma DR. The role of telehealth in psychological counseling: a comprehensive review. World J Adv Res Rev 2025;25(1):1639-1649. [doi: 10.30574/wjarr.2025.25.1.0228]
27. Mucić D, Hilty D, Shore J. Telepsychiatry: from "telepathy" to the mainstream. In: Mucić D, Hilty DM, editors. Digital Mental Health: The Future is Now. Cham, Switzerland: Springer Nature Switzerland; 2024:215-254.
28. Hjelle KM, Eide TO, Thorsen AL, Kvale G, Hagen K, Snorrason I, et al. The Bergen 4-day treatment for panic disorder: adapting to COVID-19 restrictions with a hybrid approach of face-to-face and videoconference modalities. BMC Psychiatry 2023;23(1):570 [FREE Full text] [doi: 10.1186/s12888-023-05062-7] [Medline: 37550696]
29. Kassing F, Seeley JR, Rizvi SL, Compton SN, Sinclair J, Oshin LA, et al. Assessing the implementation of suicide-focused treatments delivered in hybrid telemental health format in a real-world setting. Psychother Res 2025;35(7):1198-1212. [doi: 10.1080/10503307.2024.2415104] [Medline: 39439041]
30. Peifer JS, Bradley E, Taasoobshirazi G. Pilot testing a brief partner-inclusive hybrid intervention for perinatal mood and anxiety disorders. Front Psychiatry 2022;13:735582 [FREE Full text] [doi: 10.3389/fpsyt.2022.735582] [Medline: 35633795]
31. Appleton R, Barnett P, Vera San Juan N, Tuudah E, Lyons N, Parker J, et al. Implementation strategies for telemental health: a systematic review. BMC Health Serv Res 2023;23(1):78 [FREE Full text] [doi: 10.1186/s12913-022-08993-1] [Medline: 36694164]
32. Rains LS, Dalton-Locke C, Landau S, Needle JJ, Johnson S. Variations in the uptake of telemental health technologies in community and crisis mental health services during the early pandemic: a survey of mental health professionals in the UK. BMC Psychiatry 2022;22(1):776 [FREE Full text] [doi: 10.1186/s12888-022-04385-1] [Medline: 36494646]
33. Cowan KE, McKean AJ, Gentry MT, Hilty DM. Barriers to use of telepsychiatry: cinicians as gatekeepers. Mayo Clin Proc 2019;94(12):2510-2523. [doi: 10.1016/j.mayocp.2019.04.018] [Medline: 31806104]
34. Connolly SL, Miller CJ, Lindsay JA, Bauer MS. A systematic review of providers' attitudes toward telemental health via videoconferencing. Clin Psychol (New York) 2020;27(2) [FREE Full text] [doi: 10.1111/cpsp.12311] [Medline: 35966216]
35. Negev M, Magal T, Kaphzan H. Attitudes of psychiatrists toward telepsychiatry: a policy Delphi study. Digit Health 2023;9:20552076231177132 [FREE Full text] [doi: 10.1177/20552076231177132] [Medline: 37312951]
36. Haydon HM, Fowler JA, Taylor ML, Smith AC, Caffery LJ. Psychological factors that contribute to the use of video consultations in health care: systematic review. J Med Internet Res 2024;26:e54636 [FREE Full text] [doi: 10.2196/54636] [Medline: 39661977]
37. Moeller AM, Christensen LF, Hansen JP, Andersen PT. Patients' acceptance of video consultations in the mental health services: a systematic review and synthesis of qualitative research. Digit Health 2022;8:20552076221075148 [FREE Full text] [doi: 10.1177/20552076221075148] [Medline: 35154803]
38. Almathami HKY, Win KT, Vlahu-Gjorgievska E. Barriers and facilitators that influence telemedicine-based, real-time, online consultation at patients' homes: systematic literature review. J Med Internet Res 2020;22(2):e16407 [FREE Full text] [doi: 10.2196/16407] [Medline: 32130131]
39. Samad L, Teague B, Elzubeir K, Moreira K, Agarwal N, Bagge S, et al. Mental health care using video during COVID-19: service user and clinician experiences, including future preferences. MHRJ 2023;28(1):93-111. [doi: 10.1108/MHRJ-06-2022-0036]
40. Guinart D, Marcy P, Hauser M, Dwyer M, Kane JM. Mental health care providers' attitudes toward telepsychiatry: a systemwide, multisite survey during the COVID-19 pandemic. Psychiatr Serv 2021;72(6):704-707. [doi: 10.1176/appi.ps.202000441] [Medline: 33593104]
41. Buckman JEJ, Saunders R, Leibowitz J, Minton R. The barriers, benefits and training needs of clinicians delivering psychological therapy via video. Behav Cogn Psychother 2021;49(6):696-720. [doi: 10.1017/S1352465821000187] [Medline: 33966666]
42. Roddy JK, Gabriel L, Sheehy R, Charura D, Dunn E, Hall J, et al. Mental health practitioners' perceptions of online working: a literature review. Br J Guid Couns 2024;53(2):190-205. [doi: 10.1080/03069885.2024.2373179]

XSL•FO
RenderX

43. Békés V, Aafjes-van Doorn K, Roberts KE, Stukenberg K, Prout T, Hoffman L. Adjusting to a new reality: Consensual qualitative research on therapists' experiences with teletherapy. J Clin Psychol 2023;79(5):1293-1313. [doi: 10.1002/jclp.23477] [Medline: 36704974]
44. Aafjes-Van Doorn K, Békés V, Luo X, Hopwood CJ. Therapists' perception of the working alliance, real relationship and therapeutic presence in in-person therapy versus tele-therapy. Psychother Res 2024;34(5):574-588. [doi: 10.1080/10503307.2023.2193299] [Medline: 37011405]
45. Gullslett MK, Kristiansen E, Nilsen ER. Therapists' experience of video consultation in specialized mental health services during the COVID-19 pandemic: qualitative interview study. JMIR Hum Factors 2021;8(3):e23150 [FREE Full text] [doi: 10.2196/23150] [Medline: 34096505]
46. Åhs JW, Ranheim A, Eriksson H, Mazaheri M. Encountering suffering in digital care: a qualitative study of providers' experiences in telemental health care. BMC Health Serv Res 2023;23(1):418 [FREE Full text] [doi: 10.1186/s12913-023-09367-x] [Medline: 37127655]
47. Lipschitz JM, Connolly SL, Van Boxtel R, Potter JR, Nixon N, Bidargaddi N. Provider perspectives on telemental health implementation: lessons learned during the COVID-19 pandemic and paths forward. Psychol Serv 2023;20(Suppl 2):11-19 [FREE Full text] [doi: 10.1037/ser0000625] [Medline: 35201809]
48. Sharma G, Devan K. The effectiveness of telepsychiatry: thematic review. BJPsych Bull 2023;47(2):82-89 [FREE Full text] [doi: 10.1192/bjb.2021.115] [Medline: 34915955]
49. Burbach FR, Stiles KM, Brown K, Allen VME, Mc Kenny R. 'More collaborative and open': as systemic psychotherapists become more familiar with videoconferencing, are the advantages of the medium becoming more apparent? A focus group study with UK systemic practitioners. Hum Syst 2022;2(2):82-101. [doi: 10.1177/26344041221085227]
50. Aafjes-van Doorn K, Békés V, Prout TA. Grappling with our therapeutic relationship and professional self-doubt during COVID-19: will we use video therapy again? Counsel Psychol Q 2020;34(3-4):473-484. [doi: 10.1080/09515070.2020.1773404]
51. Kaphzan H, Sarfati Noiman M, Negev M. The attitudes and perceptions of israeli psychiatrists toward telepsychiatry and their behavioral intention to use telepsychiatry. Front Psychiatry 2022;13:829965 [FREE Full text] [doi: 10.3389/fpsyt.2022.829965] [Medline: 35386519]
52. Psykiske helsetjenester og tilgang til psykologtjenester i Norge En analyse av tilgang til og fordeling av helsetjenester for psykiske lidelser og psykologenes rolle i tjenestetilbudet. Oslo Economics. 2022. URL: https://osloeconomics.no/wp-content/uploads/2022/06/Psykiske-helsetjenester-og-tilgang-til-psykologtjenester-i-Norge-26.09.22.pdf [accessed 2025-05-01]
53. Helseforsikring. Finans Norge. 2024. URL: https://www.finansnorge.no/tema/statistikk-og-analyse/forsikring/helseforsikring/ [accessed 2025-05-01]
54. Achtyes ED, Glenn T, Monteith S, Geddes JR, Whybrow PC, Martini J, et al. Telepsychiatry in an era of digital mental health startups. Curr Psychiatry Rep 2023;25(6):263-272 [FREE Full text] [doi: 10.1007/s11920-023-01425-9] [Medline: 37166622]
55. Larsson A, Elf O, Gross C, Elf J. Welfare services in an era of digital disruption: how digitalization reshapes the health care market. In: Digital Transformation and Public Services. Milton Park, Oxfordshire, UK: Routledge; 2019:33-57.
56. Frank HE, Cruden G, Crane ME. Private practice, private insurance, and private pay mental health services: an understudied area in implementation science. Adm Policy Ment Health 2024;51(1):1-6 [FREE Full text] [doi: 10.1007/s10488-023-01306-6] [Medline: 37880471]
57. Mohler MJ. Collaboration across clinical silos. Front Health Serv Manage 2013;29(4):36-44. [Medline: 23858987]
58. Pedersen ERG, Sudzina F, Rosati F. A multi-dimensional study of organisational boundaries and silos in the healthcare sector. Health Serv Manage Res 2024;37(4):200-208. [doi: 10.1177/09514848231218617] [Medline: 38011078]
59. Snell T. Breaking down silos: the value of interdisciplinary production of knowledge for (health) innovation. In: Handbook of Ethnography in Healthcare Research. Milton Park, Oxfordshire, UK: Routledge; 2020:388-400.
60. Ii SS, Fitzgerald L, Morys-Carter MM, Davie NL, Barker R. Knowledge translation in tri-sectoral collaborations: an exploration of perceptions of academia, industry and healthcare collaborations in innovation adoption. Health Policy 2018;122(2):175-183. [doi: 10.1016/j.healthpol.2017.11.010] [Medline: 29191387]
61. Lander B. Sectoral collaboration in biomedical research and development. Scientometrics 2012;94(1):343-357. [doi: 10.1007/s11192-012-0776-8]
62. Robinson OC. Sampling in interview-based qualitative research: a theoretical and practical guide. Qual Res Psychol 2013;11(1):25-41. [doi: 10.1080/14780887.2013.801543]
63. Braun V, Clarke V. One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qual Res Psychol 2020;18(3):328-352. [doi: 10.1080/14780887.2020.1769238]
64. Saunes I, Durvy B, Litvinova Y. Norway: health system summary. European Observatory on Health Systems and Policies, WHO Regional Office for Europe. 2024. URL: https://eurohealthobservatory.who.int/publications/i/norway-health-system-summary-2024 [accessed 2026-01-08]
65. The basic structure of the Norwegian health service. Ministry of Health and Care Services. 2024. URL: https://www.regjeringen.no/en/topics/health-and-care/innsikt/the-basic-structure-of-the-norwegian-health-service/id227440/ [accessed 2025-05-01]

XSL•FO
RenderX

66. Slevin E, Sines D. Enhancing the truthfulness, consistency and transferability of a qualitative study: utilising a manifold of approaches. Nurse Researcher 2000;7(2):79-98. [doi: 10.7748/nr2000.01.7.2.79.c6113]
67. Skoge M, Aminoff SR, Barrett EA, Bryhni GE, Kling K, Kværner KJ, et al. A mobile app designed to promote shared decision-making in the treatment of psychotic disorders: feasibility and acceptability study. JMIR Hum Factors 2025;12:e68813 [FREE Full text] [doi: 10.2196/68813] [Medline: 40644621]
68. Braun V, Clarke V. Conceptual and design thinking for thematic analysis. Qual Psychol 2022;9(1):3-26. [doi: 10.1037/qup0000196]
69. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2008;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
70. Tseng WS. Culture and psychotherapy: review and practical guidelines. Transcult Psychiatry 1999;36(2):131-179. [doi: 10.1177/136346159903600201]
71. Docherty JP, Colbert BM. The evolution of psychotherapy: from Freud to prescription digital therapeutics. Front Psychiatry 2024;15:1477543 [FREE Full text] [doi: 10.3389/fpsyt.2024.1477543] [Medline: 39507283]
72. Berardi C, Antonini M, Jordan Z, Wechtler H, Paolucci F, Hinwood M. Barriers and facilitators to the implementation of digital technologies in mental health systems: a qualitative systematic review to inform a policy framework. BMC Health Serv Res 2024;24(1):243 [FREE Full text] [doi: 10.1186/s12913-023-10536-1] [Medline: 38408938]
73. Atkinson JM. Protecting or empowering the vulnerable? mental illness, communication and the research process. Res Ethics 2007;3(4):134-138. [doi: 10.1177/174701610700300410]
74. Sullivan G, Mittal D, Reaves CM, Haynes TF, Han X, Mukherjee S, et al. Influence of schizophrenia diagnosis on providers' practice decisions. J Clin Psychiatry 2015;76(8):1068-74; quiz 1074. [doi: 10.4088/JCP.14m09465] [Medline: 26335084]
75. Wang K, Link BG, Corrigan PW, Davidson L, Flanagan E. Perceived provider stigma as a predictor of mental health service users' internalized stigma and disempowerment. Psychiatry Res 2018;259:526-531 [FREE Full text] [doi: 10.1016/j.psychres.2017.11.036] [Medline: 29156425]
76. Feijt MA, de Kort YA, Bongers IM, IJsselsteijn WA. Perceived drivers and barriers to the adoption of emental health by psychologists: the construction of the levels of adoption of eMental health model. J Med Internet Res 2018;20(4):e153 [FREE Full text] [doi: 10.2196/jmir.9485] [Medline: 29691215]

Abbreviations

EMDR: eye movement desensitization and reprocessing therapy
TSD: Services for Sensitive Data
VC: video consultations

Edited by A Kushniruk; submitted 17.Jul.2025; peer-reviewed by M Pulier, J Muschol; comments to author 04.Dec.2025; revised version received 11.Dec.2025; accepted 02.Jan.2026; published 26.Jan.2026.

Please cite as:
Skoge M, Aminoff SR, Ihler HM, Kværner KJ, Støme LN, Romm KL
Exploring Video Consultations Across the Public and Private Sectors in Norway: Semistructured Interview Study
JMIR Hum Factors 2026;13:e80812
URL: https://humanfactors.jmir.org/2026/1/e80812
doi:10.2196/80812
PMID:

©Mari Skoge, Sofie Ragnhild Aminoff, Henrik Myhre Ihler, Kari Jorunn Kværner, Linn Nathalie Støme, Kristin Lie Romm. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 26.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Perspectives of Frontline Clinicians and End-Line Users on Smartphone-Based Photography for Assessing Traumatic Dental Injuries: Focus Group Interview Study and Thematic Analysis

Emily C Schultz[1,2], RDH, MS; Boyen Huang[1], DDS, MHA, PhD; Margaret Shenouda[1], BA; Mohamed Estai[3], PhD; Sarbin Ranjitkar[4], BDS, BScDent (Hons), PhD; Jeffrey P Louie[5], MD; Patimaporn Pungchanchaikul[6], DDS, MSc, PhD

[1]Department of Primary Dental Care, University of Minnesota School of Dentistry, Minneapolis, MN, United States

[2]Department of Dental Education, College of Allied Health and Nursing, Minnesota State University, Mankato, MN, United States

[3]School of Human Sciences, The University of Western Australia, Perth, Australia

[4]School of Dentistry, College of Health, Adelaide University, Adelaide, Australia

[5]Department of Pediatrics, University of Minnesota Medical School, Minneapolis, MN, United States

[6]Faculty of Dentistry, Khon Kaen University, Khon Kaen, Thailand

Corresponding Author:
Boyen Huang, DDS, MHA, PhD
Department of Primary Dental Care
University of Minnesota School of Dentistry
515 Delaware Street SE
Minneapolis, MN, 55455
United States
Phone: 1 612 624 7848
Email: huan2321@umn.edu

Abstract

Background: Mobile health (mHealth) is increasingly used in teledentistry for telediagnosis and other services; yet, the perceptions of frontline clinicians and end-line users regarding these technologies remain unexplored.

Objective: This study examined the acceptability, feasibility, and usability of an mHealth model for telediagnosis from the perspectives of frontline clinicians and end-line users.

Methods: A qualitative study using focus group interviews was conducted with 15 participants, including frontline clinicians and end-line users. Frontline clinicians captured dental images via a smartphone app, while end-line users assessed them through an mHealth platform. Interview transcriptions were thematically analyzed using consensus coding.

Results: Thematic analysis identified 9 key themes: feasibility and perceived ease of use, perceived usefulness, compatibility, self-image and social influences, self-efficacy, voluntariness and behavior intention, anxiety, facilitating conditions, and attitudes toward a behavior. Participants considered smartphone-based photography acceptable and feasible for remote dental assessment. Facilitators and barriers to implementing the mHealth model were highlighted, and recommendations for improvements were proposed.

Conclusions: Cyclical education and professional development are essential to enhancing user confidence and technology usability. Addressing patient and clinician resistance through targeted education, improved communication, and operational upgrades such as camera grids, system integration, and simplified login can support adoption. This study highlights mHealth's potential in emergency dental assessment and screening, particularly for underserved populations, and underscores opportunities for interprofessional collaboration. Future research should explore broader clinical applications across oral health conditions.

(JMIR Hum Factors 2026;13:e82668) doi:10.2196/82668

KEYWORDS

acceptability; dental trauma; feasibility; image quality; mHealth; teledentistry; telediagnosis; usability; user perception

XSL•FO
RenderX

Introduction

Traumatic dental injuries (TDIs) are prevalent [1] and demand urgent attention when accompanied by pain and/or bleeding [2]. Most patients with TDI presenting in emergency departments need consultation and evaluation from dental professionals [2]. However, the limited on-call availability of dental professionals in emergency settings often leads to unnecessary patient transfers and increased health care costs [3], an issue that worsened as dental emergency services declined during the COVID-19 pandemic [4]. To address these challenges, technologies for virtual dental examinations have been developed and implemented [5].

Mobile health (mHealth) facilitators have incorporated mobile phones and wireless technological devices into health care practices, supporting the promotion and maintenance of health, enhancing preventive care, improving clinical decision-making and operational efficiency, and enabling remote communication and interaction. This approach has gained widespread adoption among health care professionals and caregivers [6]. Key benefits of mHealth include increased accessibility to health care services, the provision of anonymous consultation, and decreased reliance on travel and physical contact. These advantages were particularly valuable during the COVID-19 pandemic, when social distancing measures were necessary [7].

The concept of mHealth as a modality of teledentistry, or more specifically, m-oral health [8], includes teleconsultation for treatment planning and review [9], telediagnosis with virtual examination [9-14], telemonitoring using patient-generated health data [15-18], telesupport with interactive multimedia [19], and teleintervention to improve therapy adherence and complication management [20]. Prior mHealth studies have demonstrated adequate diagnostic performance for the remote assessment of TDIs [10], dental caries [11], and oral cancer [12] using smartphone-acquired photographs. In a survey conducted in Saudi Arabia, more than 60% of dentists have used mHealth technologies to capture and/or transmit clinical photographs, and the majority of them were confident in the diagnostic accuracy using this approach [21].

Our recent work described a workflow for telediagnosis of TDI, where trained clinicians and students used a smartphone camera and app to capture dental photographs, uploading them to secure cloud storage. Remote dentists and dental therapists then reviewed the images, recorded dental findings, and provided clinical recommendations through the same pathway in reverse. While the quantitative performance metrics of this mHealth practice were promising, areas for improvement, such as image quality and professional development on dental photography and remote assessment, were also indicated [10].

Within the Technology Acceptance Model (TAM), perceived ease of use and perceived usefulness are the determinants for an individual's intention to use a new technology [22]. Based on different roles and concerns, Wallis et al [23] categorized mHealth users as patients, frontline clinicians (point-of-care clinicians), and end-line users (academic experts). Building on a similar classification, the users of our telediagnosis model comprised human participants (patients and nonpatient individuals who provided consent for dental photography), frontline clinicians (those who captured dental photos), and end-line users (those who remotely assessed dental photos) [10]. When searching literature with keywords of feasibility, acceptability, usability, or user perception, we found that only a few studies have reported frontline clinicians' and end-line users' perspectives on specific mHealth models, practices, or technologies for dental care or oral health promotion [13,15-18,20]. Among these, only 1 study focused on telediagnosis [13], but their diagnostic method was not photography-related. Of further note, an earlier mHealth study reported the perspectives of parents and caregivers on a photographic telediagnosis model for the detection of dental caries [14], but they did not investigate the perspectives of frontline clinicians and end-line users.

Despite evidence that smartphone-acquired photographs can achieve acceptable diagnostic performance for several oral conditions, few studies have examined the perspectives of the 2 professional user groups essential to photographic telediagnosis, frontline clinicians who capture images and end-line users who interpret them. Prior work has focused on patients, caregivers, or nonphotography telemethods, leaving clinician-centered implementation questions unanswered. This gap matters because diagnostic accuracy alone does not guarantee real-world adoption. Human factors such as privacy concerns, workflow fit, training needs, and system integration also directly influence acceptability and sustained use. By exploring these clinician perspectives, this study addresses practical barriers and facilitators that must be resolved to translate promising diagnostic performance into scalable, safe, and equitable teledentistry services. This study therefore aimed to explore frontline clinicians' and end-line users' perspectives on the feasibility, usability, and acceptability of smartphone-based photographic telediagnosis for TDI.

Methods

Study Design and Setting

This qualitative study was part of a larger project and was reviewed and approved by the University of Minnesota Institutional Review Board (Study ID: STUDY00014736). The study setting consisted of multiple sites located in Minnesota (United States) and Khon Kaen (Thailand), including the University of Minnesota clinical site, Khon Kaen University clinical sites, and Minnesota State Fair research facilities. The quantitative component of the larger project has been reported recently [10]. Following that, a qualitative study design grounded in the TAM [22] and interpretivism [24] was used to evaluate the perspectives of frontline clinicians (dental photographers) and end-line users (remote reviewers) on their use of the mHealth technology to assess TDI. Qualitative study data were collected in November and December 2023 and analyzed from January to August 2024. The conduction and presentation of this study adhered to the COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines [25].

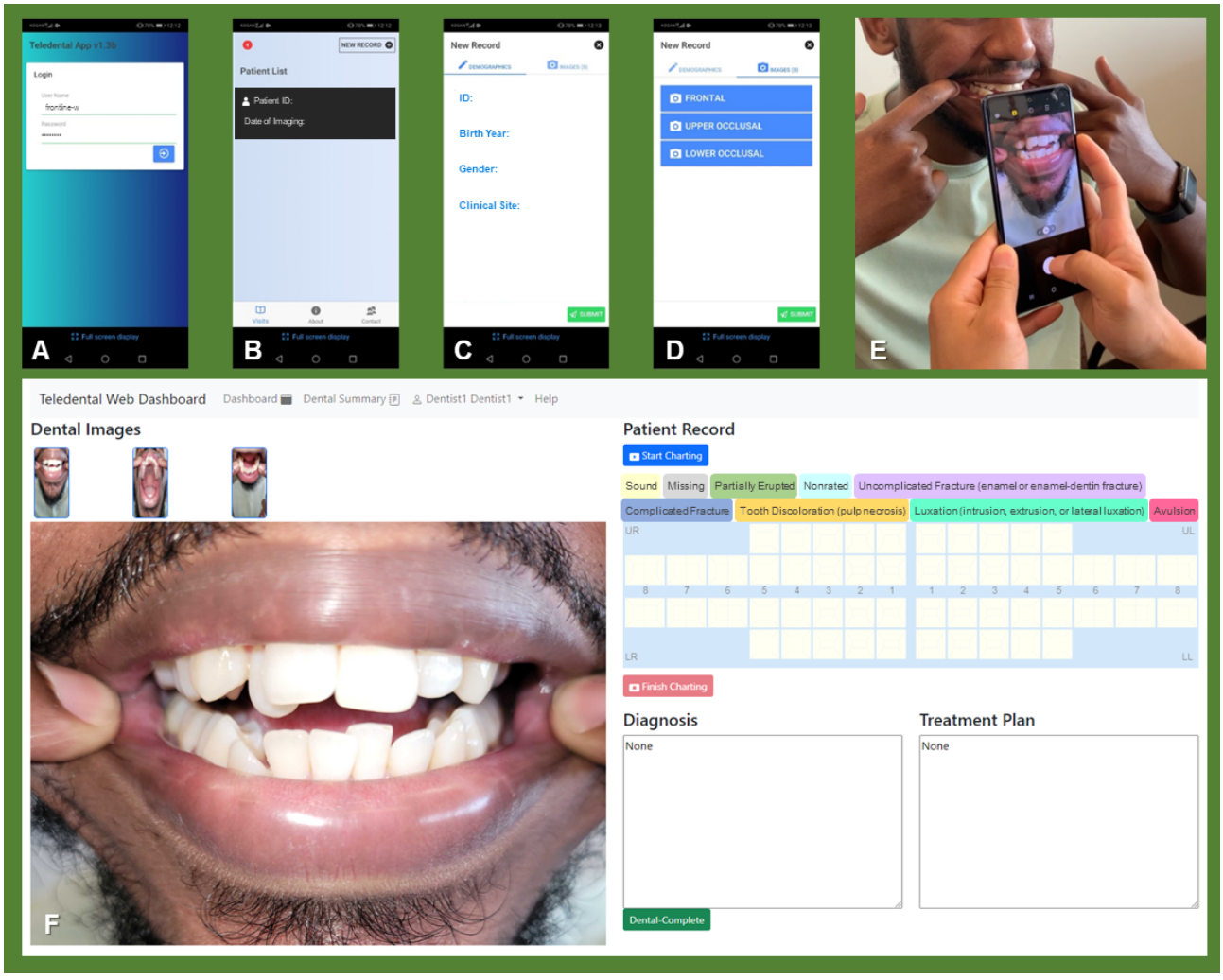
This study built upon our quantitative work of TDI remote assessments using an image acquisition app "Teledental" (Commonwealth Scientific and Industrial Research Organisation

XSL•FO
RenderX

[CSIRO]) and a web-based data management platform "Remote-I" (CSIRO; Figure 1). From August 2022 to July 2023, the photographs of teeth were captured by 11 trained frontline clinicians using the Teledental app, and then remotely reviewed by 5 licensed dental professionals (end-line users) using the Remote-I system [10].

In Figure 1, panels A-C show the "Teledental" app interface used by frontline clinicians to log in with individual passwords and enter patient data, including patient ID, birth year, sex, and clinical site. Panel D displays the app's imaging screen, which prompts users to capture 3 standardized intraoral views: frontal, upper occlusal, and lower occlusal. Panel E depicts 2 frontline clinicians participating in one-on-one calibration after completing a training session designed to ensure consistent image quality and adherence to capture protocols. Panel F shows the "Remote-I" platform interface used by end-line users to access the patient record created during calibration; when the cursor hovers over a thumbnail, the corresponding dental image is enlarged for review. End-line users examined all 3 images and charted dental conditions using a color-coded odontogram that included diagnostic categories such as uncomplicated fractures, luxations, and partially erupted teeth. The clinicians provided consent for the use of their images and data in the publication.

Figure 1. Workflow and interfaces of the mHealth app and platform used in this study for photographic telediagnosis of traumatic dental injuries.



Basic Security Measures

All researchers, frontline clinicians, and end-line users adhered to institutional review board–approved procedures governing data access and privacy. To limit image persistence on devices, the Teledental app was configured to prevent local storage of captured images; images were transmitted directly to a US-owned cloud service selected jointly by the research team and the vendor (CSIRO). Remote image review required authenticated access to the Remote-I platform; the platform's settings disallowed image downloads to end-line users' workstations. Study smartphones were purchased new from the vendor (Samsung), configured with restricted user permissions, and connected only to password-protected institutional Wi-Fi (eg, eduroam or hospital internal networks); installation of additional apps was prohibited; devices were centrally stored and managed by a key researcher in Minnesota and a key researcher in Khon Kaen. Individual user accounts and passwords were required for both Teledental and Remote-I and the passwords were changed periodically. Two-factor authentication was not used by study participants during data collection. All study personnel completed training in privacy and data-handling best practices, including HIPAA (Health

XSL•FO
RenderX

Insurance Portability and Accountability Act) principles [26], and compliance with those requirements was monitored.

Participants

All eligible frontline clinicians and end-line users who contributed to the preceding quantitative study were invited (consecutive sampling) and the full available cohort consented to participate, with the only exclusion being a frontline clinician who is a coauthor to avoid conflict of interest. The sample included a purposeful mix of roles and experience levels (dental students, general dentists, dental therapists, and dental specialists) across 2 countries. No participants were selected or excluded by demographic or professional criteria. To acknowledge potential recruitment influences, we report that University of Minnesota employees and student workers received modest incentives (US $40 and US $15, respectively) while participants based in Thailand did not. All participants were notified of the requirement to attend a 60-minute focus group interview through the Zoom app (Zoom Video Communications). Each focus group was composed of 1 to 6 participants according to their schedule availability. The participants based in the United States and Thailand were separated into different focus groups due to different time zones. The interviews were conducted in English by 2 US-based researchers together. All participants were capable of communicating in English during the Zoom interviews.

Data Collection and Instrument

Two moderators, ECS (a female dental hygiene faculty member with a Master of Science degree) and MS (a female Honors college student with a Bachelor of Arts degree) conducted each focus group interview, while BH (a male senior dentistry faculty member and the principal investigator [PI] with a Doctor of Philosophy and Doctor of Dental Surgery degrees) observed all sessions. Prior to the first interview, the moderators underwent training to adopt a neutral interviewing stance, ensuring that participants' perspectives were accurately captured. Since participants were recruited from frontline and end-line users of a previous study led by the observer, the moderators had a limited prior acquaintance with 3 participants. To maintain transparency, the moderators and observer introduced themselves and outlined the research objectives and content on the informed consent form and at the beginning of each interview. Before the interview with participants based in Thailand, the observer and moderators interacted with them in English for approximately 15 minutes via Zoom to confirm their proficiency and ensure effective communication. Only the participants and researchers were present during the interviews. Field notes were taken by the moderators during the sessions, and all interviews were audio-recorded and securely stored on a cloud server. After each focus group, participants were emailed the full study instrument and given 2 weeks to provide additional comments. To safeguard data quality and reduce risks of misinterpretation, a researcher (PP), a native Thai speaker fluent in English with graduate training in Britain, attended the Thai focus group to provide Zoom technical support and reviewed the transcripts from that session for accuracy. Although data saturation was defined as the point at which no new information emerged across 3 consecutive focus groups [27], we chose to interview all 15 participants in an effort to generate more evidence and decrease bias due to our small but very diverse sample size. Repeat interviews were not conducted.

The study instrument used for the interviews consisted of guided interview questions, exploring the perspectives of the frontline clinicians and end-line users. The interview questions were adapted from published resources [22,28-32] and were rephrased to reflect the context of this qualitative study and the mHealth technologies used. Subsequently, a multimember researcher panel with diverse expertise, including qualitative study design, medical informatics, clinical dentistry, dental public health, and special needs dentistry, reviewed and finalized the questions. The interview questions included in this study are presented in Table 1. The purpose of using this guided interview instrument for the interview was to gain a well-rounded perspective from the participants on their experience with the mHealth model and maintain consistency among all focus group interviews.

XSL•FO
RenderX

Table 1. Guided interview questions adapted from the Technology Acceptance Model.

Interview questions	References
• Tell us about the feasibility of using the mHealth[a] app or platform. What factors played a role?	[22,28]
• How do you feel this mHealth app or platform can contribute to working with dental trauma cases? (In terms of efficiency, usefulness, etc)	[22,28]
• How does this mHealth app or platform fit with the way dental trauma cases are examined and diagnosed?	[28,29]
• How do you feel the use of the mHealth app or platform may change the way you and other people in your profession, specialty or field are viewed?	[28,29]
• Did your previous experiences influence your use of the mHealth app or platform? Anything positive? Anything negative?	[28,30]
• How willing are you to use the mHealth app or platform to examine and diagnose dental trauma cases in the future?	[28,29]
• How likely are you to use an mHealth app or platform like the one we used in the future for dental trauma cases?	[28,31]
• How comfortable or uncomfortable did you feel about using the technology for dental trauma cases? • Only to frontline clinicians: How might you have observed either potential distress or comfort from the patients or fairgoers enrolled in this study?	[28,30]
• What attitudes do you believe your superiors have around your use of the mHealth app or platform to examine and diagnose dental trauma cases?	[28,31]
• What resources, knowledge, or assistance do you feel should be available when using the mHealth app or platform to examine and diagnose dental trauma cases?	[28,32]
• What do you feel went well with the mHealth app or platform? What improvements do you feel could be made with the mHealth app or platform?	[28,31]

[a]mHealth: mobile health.

Data Analysis

A consensus coding process was used to summarize and synthesize the qualitative data from the interview questions and determine the results [33,34]. Coders calibrated via Zoom before conducting any focus group interviews. This early calibration ensured alignment in coding practices and interpretation prior to engaging with the data, to maintain consistency across all transcripts. Memoing was used throughout the coding process to document coder reflections, emerging themes, and the rationale behind coding decisions. These memos supported transparency and helped track the coding development. The research team used a negotiated agreement strategy for intercoder reliability. Prior to data collection, coders collaboratively established a process for resolving discrepancies. When disagreements arose during coding, coders revisited the transcript, discussed their individual interpretations, and reached a consensus. The PI, who observed all focus group interviews, was also consulted to help resolve disagreements and ensure consistency. We used a hybrid coding approach that integrated both deductive and inductive strategies to analyze the focus group data. Initially, we developed a set of a priori codes grounded in our research questions and informed by the TAM [22], which provided a theoretical framework for examining participants' responses. As coding progressed, we remained open to emergent patterns and concepts, allowing for the creation of new inductive codes that captured unanticipated insights arising directly from the data. This flexible approach enabled us to systematically explore expected themes while also incorporating novel findings that enriched our understanding of participants' experiences [34,35]. After all interview sessions had taken place, a Zoom meeting among the research team was held to analyze and discuss the qualitative data, use the transcripts to identify common themes, and extract appropriate evidence, such as descriptive terms and quotes, from the participants' responses. The transcripts and findings were not returned to participants for review.

Ethical Considerations

This study was reviewed and approved by the University of Minnesota Institutional Review Board (Study ID: STUDY00014736). Written informed consent was obtained from all participants prior to data collection. All researchers and study personnel adhered to institutional review board–approved procedures governing data access and privacy. Compensation was provided to University of Minnesota employees and student workers, but not Thailand-based participants due to University out of country payment restrictions.

Results

Overview

A total of 7 interview sessions were conducted. Among the 15 study participants, 10 were frontline clinicians who took the dental photographs and 5 were end-line users who interpreted

XSL•FO
RenderX

the dental photographs from a distance. The 10 frontline clinicians included 5 general dentists, 1 pediatric dental specialist, 1 dental therapist, and 3 dental students. The 5 end-line users consisted of 1 pediatric dental specialist, 2 general dentists, and 2 dental therapists. The 6 dentists who acted as frontline clinicians and took dental photographs were all based in Khon Kaen (Thailand), while the other 4 frontline clinicians and the 5 end-line users were based in Minnesota (United States). Table 2 lists the focus groups, roles, professions, sex, and countries of the participants included in this qualitative study.

Data saturation was achieved within the 7 interviews as no new codes or themes emerged after the first 4 interviews. Thematic analysis of frontline clinicians' and end-line users' structured interviews revealed 9 main themes: feasibility and perceived ease of use, perceived usefulness, compatibility, self-image and social influences, self-efficacy, voluntariness and behavior intention, anxiety, facilitating conditions, and attitudes toward a behavior (Table 3).

Table 2. Focus groups, roles, professions, sex, countries, and years in practice of the participants (9 in the United States and 6 in Thailand); years in practice ranged from 0 to 20 years. All consented to publication.

Focus group and role	Profession	Sex	Country	Years in practice
1				
End-line user	Dental therapist	Male	United States	10
End-line user	General dentist	Male	United States	15
2				
End-line user	Dental therapist	Female	United States	2
3				
Frontline clinician	Dental student	Male	United States	0
4				
Frontline clinician	Dental therapist	Female	United States	12
Frontline clinician	Dental student	Male	United States	0
5				
End-line user	General dentist	Male	United States	3
End-line user	Dental specialist	Male	United States	20
6				
Frontline clinician	Dental student	Female	United States	0
7				
Frontline clinician	General dentist	Female	Thailand	5
Frontline clinician	General dentist	Female	Thailand	6
Frontline clinician	General dentist	Female	Thailand	12
Frontline clinician	General dentist	Male	Thailand	5
Frontline clinician	General dentist	Male	Thailand	8
Frontline clinician	Dental specialist	Male	Thailand	15

XSL•FO
RenderX

Table 3. Subthemes (code words), definitions, and representative transcript excerpts that contributed to theme construction.

Main theme	Definition	Subtheme	Example from transcripts
• Feasibility and perceived ease of use	• Statements that describe the feasibility and ease of use of the mHealth[a]	• Consistency • Login • Template	• "I think it was fairly easy for me just because I was comfortable using intraoral cameras before. I think it's something that would be kind of intuitive to a new user. It wasn't that difficult." • "I think the only negative thing, which we kind of figured out, was if the app was left unutilized for a certain amount of time, then it would time out and you'd have to sign back in." • "I want a guideline with the application while using. Kind of like a template to make sure pictures are being taken at correct angles."
• Perceived usefulness	• Comments relating to the usefulness of mHealth	• Access • Timely • Useful	• "I think it's beneficial because, in terms of increasing access, people might not always have access to dental services right away following trauma." • "mHealth could help get the treatment done faster if you could get diagnosis back quicker and in turn get patients treatment faster." • "Specifically speaking about usability, it was really user-friendly in terms of clicking on patient, reviewing what is in odontogram and chart."
• Compatibility	• Statements referring to the familiarity and integration of mHealth technologies	• Technology familiarity	• "I think with our generation in this day and age it is easier to use apps and the app wasn't that complicated at all, it was kind of like straightforward."
• Self-image and social influences	• Comments that describe the social perception of mHealth use	• Acceptance • Resistance	• "I feel like this would be viewed in a positive manner, in that we are taking extra steps to expand access to care through teledentistry and through this mHealth method of like this is easily accessible for patients." • "Some people might say well that's not truly seeing the patient or diagnosing the issue if you're not seeing them in-person since it's kind of a new area of dentistry, but I don't think I'd have any issue with performing teledentistry."
• Self-efficacy	• Statements that describe the user's beliefs in themselves to use mHealth	• Confidence	• "I was very comfortable with how to utilize the technology and then since I've used my phone a lot and taken a lot of images before, it was pretty streamline for me to be able to use."
• Voluntariness and behavior intention	• Comments referring to the user's intention to adopt or integrate mHealth into practice	• Reluctance • Intention	• "For me, I don't feel like it adds a lot to either the patient experience or patient care at the moment." • "I think I would definitely be willing to use it, like I was saying, like for screening patients, I think it's a very valuable asset to see how acute or how soon they would need to get in."

XSL•FO
RenderX

Main theme	Definition	Subtheme	Example from transcripts
• Anxiety	• Statements that describe the user's comfort level related to the use of mHealth	• Comfortable	• "I think it was very easy for me. I think it was something that I felt very comfortable being able to use, taking the images, making sure the images came out well. I felt very comfortable doing that. I know it may be more challenging for people who are not as familiar with being able to use that technology."
• Facilitating conditions	• Statements that describe user's view on what conditions are needed to implement mHealth	• Resources • More practice	• "I think the research team did a good job training us through this, the app, and showing us how to use it, and they even did a one-on-one calibration when we got the phones from them." • "I think if we did a few practice sessions it would help us a lot to be able to emulate what type of images we wanted to take and how we wanted them to look like. So providing some model images of what the appropriate images would look like, I think that would be beneficial."
• Attitudes toward a behavior	• Statements that describe a user's feelings about the use of mHealth	• Positive attitudes • Improvements	• "I was excited when I was using it. I thought it's something that I can definitely make a difference." • "Technology that would streamline the photo taking process, instead of needing to click. Something that would eliminate the need to save or change to the next photo manually. Also not having to login every time, save login for quicker future login."

[a]mHealth: mobile health.

Feasibility and Perceived Ease of Use

As a whole, the frontline clinicians based in Thailand and the United States found the mHealth app feasible and straightforward to use, stating the app itself was self-explanatory, simple, and easy to use. The frontline clinicians at the Minnesota State Fair (United States) used the technology more often and more consistently, while the frontline clinicians at clinical sites in Minnesota (United States) and Khon Kaen (Thailand) used the technology more sporadically. With this being said, frontline clinicians who used it more consistently found it easier to use than those who used it sporadically as they were more familiar with the technology.

Almost all frontline clinicians felt that a more streamlined login process is necessary to improve mHealth app use. They criticized the need to log in each use with a username and password, as it could be time-consuming. Frontline clinicians in the United States suggested some options for streamlining the login process, including facial recognition for login, the ability to stay logged in and/or the ability to save username and password on the login page. Both frontline clinicians in Thailand and the United States mentioned lighting and picture scale as potential barriers. Both parties suggested some type of template or guide to help with image flow, image angles, and image scale and believe this would make images more consistent. A frontline clinician also mentioned that at times he would have trouble focusing the image,

> *For some reason, when I was taking photos, I felt like I couldn't focus on what I wanted to focus on. Sometimes, I would take multiple pictures of the same thing and try to get a really good picture, but I couldn't get a clear image of things.*

End-line users had similar feelings to the frontline clinicians as far as the feasibility of the mHealth data management platform. Generally speaking, end-line users mentioned the platform itself was feasible, simple to use, user-friendly and straightforward. Regarding the mHealth platform, a dentist end-line user stated, "I think it's something that would be kind of intuitive to a new user."

A dental specialist and a general dentist mentioned that if a charting error was made, there was not a simple way to correct the mistake. They found the easiest way to correct it was to exit out of the platform and start over. They mentioned the importance of fixing this, because they believe teledentistry should be fairly efficient due to the volume one may be reviewing. A dental therapist end-line user stated that the most difficult thing was that the platform could be slow at times. Despite the hardships, overall end-line users found the learning curve to be manageable.

Perceived Usefulness

As a whole, frontline clinicians perceived the mHealth app to be useful. Several frontline clinicians believed that the mHealth app was beneficial, particularly for increasing access to dental

XSL•FO
RenderX

care and reducing the burden on emergency services. Frontline clinicians based in the United States commented, “People might not have access to dental services right away following trauma.” They also stated, “I believe this app is very useful and beneficial, especially like when kids get trauma and they don’t have a dental home, it gives them quick access to care.” Frontline clinicians in Thailand also thought that the app would be a game changer for teledentistry. Specifically, they thought the mHealth app was great for those in remote areas where they do not have a dental specialist.

Frontline clinicians in both countries did suggest they think the app would be more useful if it could include more information, such as patient age, radiographs, and better image quality. They also felt some limitations with its current scope of only TDI and suggested widening it to a broader and more general dentistry scope.

End-line users deemed the mHealth data management platform useful for teledentistry, especially for initial assessments. They mentioned its potential for timely diagnosis and usefulness when there is not a dental provider in the nearby area.

Similar to the frontline clinicians, end-line users also felt having more patient information, such as pain, chief complaint, chart notes, etc, would be beneficial and useful. They also mentioned that suboptimal image quality of some dental photos may hinder the diagnosis of minor trauma and recommended that this be addressed for future use.

Compatibility

Both frontline clinicians and end-line users felt that the mHealth app and platform were compatible with the way existing dental cases and TDI cases are treated. Meaning both parties felt the technology was not too out of the realm and could be useful for handling trauma cases. A frontline clinician and an end-line user eluded to the fact that “as a society we use technology every day, and some of the younger generations rely on technology so we are already familiar with it.” They stated that most people are using smartphones every day, so there likely would not be a learning curve for widespread TDI-based teledentistry. Future improvements for the mHealth app and platform, as recommended by the interviewees, included fine-tuning of the odontogram and patient chart displayed on the platform, and integration of the mHealth model with the practice management software available in the market.

Self-Image and Social Influences

The frontline clinicians expressed that using the mHealth app did not significantly alter their self-image or how they were perceived by others. They felt the app was generally viewed positively as a helpful tool in the dental care process. A frontline clinician in Thailand stated, “there have been no bad attitudes from superiors when using the application.”

Although frontline clinicians generally had a positive perception of using an mHealth app, some concerns were raised about patients questioning the provider’s competence if they relied heavily on the app for diagnosis. One participant mentioned, “If the parent sees a provider using the application on their child they may question their intelligence as to why they are not confident in diagnosing on their own.” They also alluded that older generations may potentially have similar concerns.

Four out of 5 end-line users felt that using a teledentistry platform, such as mHealth, would have a positive impact on their professional images. They mentioned it would showcase their adaptability to using new technologies. The other end-line user, a general dentist, thought that professionals in the field would not view him or his work any differently whether he used the technology or not.

The only concern mentioned was initially, there may be some pushback, as health professionals have seen this when implementing other new technologies until the entire team is familiar and comfortable with it.

Self-Efficacy

Frontline clinicians felt confident using the mHealth app after the initial training, citing their familiarity with smartphone technology as a contributing factor. One frontline clinician stated, “Once we had the training, I was very comfortable with how to utilize the technology and then since I’ve used my phone a lot and taken a lot of images before, it was pretty streamline for me to be able to use.” Another made a similar comment,

> *I felt pretty comfortable. I felt familiar with using a lot of apps and utilizing that app wasn’t very challenging, and then taking the images were pretty easy.*

All end-line users also felt confident in their ability to use the mHealth platform effectively after receiving the initial training. End-line users attributed their confidence to the training session and prior experiences with other electronic charting and dental record systems. There were no barriers or negative experiences in regards to self-efficacy reported by end-line users.

Voluntariness and Behavioral Intention

Frontline clinicians indicated a willingness to use the app voluntarily, recognizing its potential benefits in improving access to care and streamlining the diagnostic process. One participant said, “I think it’s beneficial because, in terms of increasing access, people might not always have access to dental services right away following trauma.”

They also suggested the app could be more widely adopted if it included additional features and addressed current limitations, such as the need for frequent logins and better image quality mentioned earlier.

There was a strong willingness among end-line users to adopt the mHealth platform for future use, especially for screening and remote consultations. Most of them saw it as a great triage tool to get their patients to the right place in a timely manner. End-line users saw it as a valuable tool for patient care and were open to integrating it into their daily practices, despite some initial hesitations. Per one end-line user, “I would definitely be willing to use it, I think it’s a very valuable asset.” Another end-line user stated, “I’d say I’m very likely to use it in the future.”

One end-line user, the dental specialist, was concerned with how the app would integrate into other dental practice software

XSL•FO
RenderX

in an attempt to have more patient information available for the reviewer. Another end-line user, a general dentist, thought it would be beneficial to be able to communicate with the patient through the app, whether that be through voice, FaceTime (Apple Inc), or text messaging.

Anxiety

The frontline clinicians overall seemed comfortable with using and capturing photos on the smartphones, as they use them regularly. Initial anxiety among frontline clinicians was noted due to the unfamiliarity with using the specific mobile app for dental assessments. Their anxiety lessened over time as they became more familiar with the process. Those frontline clinicians witnessed a wide range of anxiety among patients and state fairgoers who were getting their dental photos taken. The most anxiety from patients was reported in a Minnesota community clinical setting. Frontline clinicians attributed this mostly to privacy issues, stating that some of the patients are potentially illegal immigrants, and became hesitant once they found out their picture would be acquired. However, anxiety decreased among some patients once they fully understood the study purpose and applications as well as the photos taken were limited to their mouths. Minnesota state fairgoers were less anxious, as they had agreed to come to the research facilities and learn more about the study, and willingly came in to learn about the study, versus being at the dentist for an oral health issue. The frontline clinicians in Thailand did not notice much anxiety from their patients. They credited this to previous use of intraoral cameras during dental examinations and the fact that participants had previously gone through and provided an informed consent.

When it came to using the mHealth platform, end-line users were overall comfortable and felt the dental charting they did was straightforward. They did report anxiety when it came to system charting glitches, such as forgetting which teeth they had already marked as sound, and the accuracy of their diagnoses when the quality of some images was less than ideal.

Facilitating Conditions

Frontline clinicians appreciated the training sessions and written materials provided, which helped them navigate the app effectively. One frontline clinician in the United States stated,

> *I think the research team did a good job training us through this, the app, and showing us how to use it, and they even did a one-on-one calibration when we got the phones from them.*

The same interviewee thought that a couple of practice sessions and being provided with model images would have helped them have a better understanding and emulate the photo expectations. The frontline clinicians based in Thailand thought a video tutorial would have helped them better understand what angles they should take photos from. They found the most difficult part was needing to log in to the app and connect to the Wi-Fi every time they went to use the app.

When it came to resources, knowledge, or assistance, end-line users emphasized the importance of adequate training and support to maximize the platform's effectiveness. They felt the training they received from the research team was adequate, and liked the guidelines so everyone was working off the same standards. However, an end-line user suggested that going through a couple of cases together would have been even more of help.

As stated earlier in this article, end-line users felt that including more patient information, or integrating with patients' dental practice software to have access to patient dental history would make the diagnosis process easier.

Attitudes Toward a Behavior

Overall, the frontline clinicians had a positive attitude toward using the mHealth app, as it was straightforward to use and they recognized its potential to enhance dental care delivery. They thought that if this dental model were integrated more widely, it could make a significant difference, especially in areas with a shortage of dental health professionals. The mHealth model was seen as a step forward in leveraging technology to improve health care access and efficiency.

Frontline clinicians thought the experience could be even more positive by suggesting things, such as streamlining the technology by automating the photo-taking process to save and switch to the next photo without manual intervention, and user interface enhancements, such as making the app more child-friendly with colors or cartoons.

Overall, the end-line users' attitudes toward using the mHealth platform were positive. End-line users acknowledged its potential to streamline workflows and improve patient outcomes. They appreciated the convenience and accessibility it offered. Specifically, they noted the value of having standardized views for each patient and appreciated the ability to zoom in on the photos and the ability to chart multiple conditions on the tooth.

Despite the positive attitudes toward the platform and use, end-line users recognized some areas of improvement and future development. One suggestion in an attempt to standardize the diagnosis was to integrate the trauma guidelines into the platform for quick reference. Another end-line user suggested making the reviewing platform into an app that could be easily accessed with a smartphone to improve convenience during emergency situations.

Discussion

Principal Findings

This study revealed participant perspectives toward the mHealth app and platform. Participants highlighted improved and faster diagnosis, underscoring perceived usefulness, and generally found the system easy to use despite login and photo clarity challenges. The app and platform seemed compatible with current dental practices. Most users responded positively, though some worried patients might perceive overreliance on technology. Training increased confidence and reduced anxiety, leading to stronger self-efficacy. Some were ready to keep using the app, while others wanted improvements first, showing mixed intentions. Overall, users expressed positive attitudes toward its role in TDI assessment.

The 9 themes illustrate how TAM constructs operate in practice. Perceived usefulness related to faster diagnosis and improved

XSL•FO
RenderX

access, while ease of use was reflected in straightforward operation. Compatibility was evident through integration with existing practices, and self-efficacy was reinforced by training and calibration. Social influences appeared in concerns about patient perceptions and peer resistance, while voluntariness and behavioral intention were expressed in willingness or reluctance to continue use. Anxiety, facilitating conditions, and attitudes toward use further shaped adoption, underscoring the interplay between technical usability and human factors.

When implementing teledentistry in clinical settings, its impacts on oral health outcomes [35], patient satisfaction [36], and cost-effectiveness [35] are commonly taken into account, and these effects have been regularly reported in literature [16,17,20,35,36]. Despite our belief in patient-centered care [37,38], health care workers' roles are central to mHealth adoption. In this study, smartphone-based photography was regarded as feasible, acceptable, and usable for remote TDI assessment. The results suggested potential applications in the initial assessment of emergency cases (such as trauma) and in triage or screening contexts (such as dental caries), warranting empirical testing in settings serving underserved populations and within interprofessional emergency workflows [2]. A prior caregiver study using the same mHealth app for dental screening and caries assessment in children also supported its feasibility [14], but direct comparisons are limited and require confirmatory research. Several other studies have also highlighted the relevance of including frontline and end-line users' perspectives in the determination of feasibility, acceptability, and usability of mHealth technologies for dental care and oral health promotion [13,15-18,20]. Those studies reported benefits including diagnostic accuracy [13], communication [17], patient-provider relationship [17], efficiency [18], and medication adherence [20]. Compellingly, some users of a non–photography-related mHealth app recommended adding a photo-capture-and-forward function to it [15], which happens to be a key characteristic of our mHealth model. This could also imply the viability of smartphone-based photography in the mHealth modality of teledentistry.

Training and calibration were key facilitators for the use of this mHealth model, increasing comfort and confidence among both frontline clinicians and end-line users. Literature suggested that adequate training decreases uncertainty, strengthens self-efficacy, and enhances mHealth usability [10,15,39,40]. A recent scoping review has found approximately 20 teledentistry educational programs [41]. Building on this evidence, curricula should incorporate topics such as virtual communication and assessment, technology troubleshooting, billing, and ethics, to help learners develop teledentistry competencies. These could be complemented by simulated training and clinical practice [42]. Because familiarity enhances usability, periodic and continuous training is recommended to refresh skills and sustain competence.

Patient and clinician perceptions may pose barriers. Some older patients might view smartphone use during diagnosis as a sign of incompetence, while younger or tech-savvy patients were more accepting [43]. Education to improve digital literacy and highlight benefits can support adoption. Resistance from colleagues may also hinder uptake. Inclusive development, professional development, enhanced communication, and reduced operational burdens can facilitate acceptance [44].

This study is exploratory and focused on user perceptions rather than on product design or feature validation; nevertheless, participants also proposed technical suggestions. Consistent image quality is essential, and camera grids or templates could help. Integration of electronic patient records would strengthen utility, while governance and ethical frameworks addressing data linkage, artificial intelligence (AI), and security are needed [45,46]. Biometric or voice login could reduce screen contact while preserving security. Finally, a risk-averse "ethical brake" approach—safeguards that trigger additional verification before data sharing or high-stakes decisions—has been proposed to inform the design of login, data sharing, and reviewer feedback loops, helping minimize potential harms as mHealth and AI features are introduced [46].

To ensure trustworthiness, investigator and data source triangulation were used to test the validity [47]. The interview questions were created and reviewed by a multimember researcher panel. Two moderators and an observer were jointly involved in data collection and analysis. These approaches broadened the focus of this research and reduced researcher biases. Additionally, the data were collected from the participants of different roles (frontline clinicians and end-line users), qualifications (dental specialists, general dentists, dental therapists, and dental students), sexes (females and males), and locations (Thailand and the United States). This participant pool contributed to diverse perspectives and validation of data.

Using a focus group interview method presents certain limitations. Group composition was based on participants' availability rather than self-selection, which may have influenced dynamics. As noted in prior literature, dominant voices can shape group discourse [48]. To mitigate this, moderators were trained to foster inclusive participation, and follow-up emails were sent to all participants to invite additional reflections. Cultural differences between Thailand and the United States also warrant consideration, even though interviews were conducted separately by country. Prior research on communication styles suggested Thais might be more inclined to accommodate audience expectations, whereas Americans might prioritize balancing personal goals with audience interests [49]. These differing cultural norms could have influenced how frontline clinicians and end-line users expressed their views. Furthermore, because interviews with Thai clinicians were conducted in English rather than their primary language, some nuance may have been lost in cross-cultural expression; however, presession proficiency checks, member checking, and transcript review by a bilingual researcher helped mitigate this risk.

While no participants were selected or excluded by demographic or professional criteria, the small sample size remains a limitation. The operation of this mHealth model required 11 frontline clinicians to take dental photos and 5 end-line users to diagnose TDI; all except 1 frontline clinician (a coauthor) participated in this qualitative study. Limited generalizability due to small samples in telemedicine and telehealth research has been noted in an integrated review [50]. Nevertheless,

XSL•FO
RenderX

thematic saturation was reached early, with no new codes emerging after the fourth interview, and data collection concluded after 7 sessions. This suggests that the dataset adequately captured the range of user perspectives for this workflow, consistent with Guest et al [51] findings that as few as 6 interviews may be sufficient for thematic analysis and interpretation. Future work should include larger and more culturally diverse clinician samples and explore similar mHealth models to enhance transferability.

The incentive differences between countries stemmed from institutional funding restrictions, which prohibit disbursing school-supported funds to individuals not legally eligible to work in the United States. Although they did not affect participation rates, such disparities may have introduced subtle response biases or influenced perceived power dynamics. To promote equitable participation in future international research, investigators should consider globally inclusive funding mechanisms or secure domestic support within each participating country.

Interpersonal dynamics [52] and prior researcher-participant familiarity [53] may also influence participant responses. To mitigate potential courtesy bias and power dynamics arising from interviewer familiarity, reflexivity and bracketing strategies were used throughout the study. Moderators received training to foster inclusive dialogue, actively invite dissenting views, and minimize hierarchical influence during interviews. A structured interview guide (Table 1) was used to ensure consistency across sessions and reduce the influence of moderator bias or familiarity on discussion flow. Although the PI observed all sessions, he did not moderate or participate; his webcam and microphone remained off throughout. Participants were not supervised, assessed, or taught by the PI during or after the study period. Their feedback, whether positive or critical, was valued equally and anonymized during reporting to encourage candor. Follow-up emails invited dissenting views, further reducing social pressure and courtesy bias. Prior researcher-participant relationships have also been identified as a potential asset in qualitative research, helping to build rapport and facilitate open dialogue [53].

Future research should incorporate role-specific interview elements, mixed methods approaches, and patient perspectives, alongside cost-effectiveness analyses. With evidence supporting remote assessment of TDI [10], dental caries [11,14], and oral cancer [12], diagnostic capacities could expand to other oral conditions such as tooth wear [54], developmental defects [55], malocclusion [56], and temporomandibular disorders [57].

Scaling mHealth tools in emergency dental settings requires technical and training enhancements as well as policy and infrastructure support. Regulatory frameworks for teledentistry differ across jurisdictions and may lack explicit guidance for photographic diagnosis, raising concerns around licensure, liability, and documentation standards. Reimbursement for asynchronous services—such as image-based consultations—remains inconsistent, potentially limiting uptake in emergency departments and underserved communities. Moreover, integration with existing electronic health record systems and secure data transmission protocols is vital to ensure interoperability, privacy compliance, and operational efficiency. Addressing these regulatory, reimbursement, and infrastructure barriers is essential to move beyond pilot models and establish sustainable, scalable mHealth solutions.

Equally important is the role of educational and professional development in supporting adoption and innovation. Integrating photography-based mHealth workflow into emergency care requires structured training and clear protocols for photo capture and clinical assessment. Competency development in teledentistry should be incorporated into continuing education and dental curricula to prepare both current and future providers. At the system level, health care organizations can leverage the insights from this study to guide infrastructure planning, resource allocation, and policy for effective integration of mHealth technologies.

Conclusion

Within the study limitations, frontline clinicians and end-line users regarded smartphone-based photography as a feasible, acceptable, and usable mHealth tool for remote dental assessment. Strengthening self-efficacy through ongoing education and professional development will be essential, while patient and clinician resistance may be mitigated by strategies such as inclusive development, communication, reduction of operational burdens, and targeted training. Technical refinements, including standardized photo templates, integration with patient management systems, and streamlined login, could further enhance usability. Future research should pilot this model in larger and more diverse clinical populations, evaluate cost-effectiveness, and explore its integration into broader oral health workflows beyond TDI.

Acknowledgments

The authors acknowledge Minnesota Masonic Charities, International Association for Dental Research, the University of Minnesota Center for Global Health and Social Responsibility, and the University of Minnesota School of Dentistry for funding support. We also thank the research participants and university faculty and staff who made this possible through their involvement. Part of this work has been presented in MS's Honors thesis [58]. A publicly available preprint of this work was posted online following the manuscript's submission to the journal for publication [59].

Funding

This study was supported with a Department of Pediatrics Driven to Discover Grant by Minnesota Masonic Charities, a Regional Development Program Grant by the International Association for Dental Research, a Global Health Seed Grant by the University

XSL•FO
RenderX

of Minnesota Center for Global Health and Social Responsibility, and a Faculty Start-up Funding for BH by the University of Minnesota School of Dentistry.

Data Availability

The datasets generated or analyzed during this study are not publicly available because participants of this study did not give written consent for their recordings to be publicly shared, but the transcribed data are available from the corresponding author (BH) on reasonable request.

Conflicts of Interest

None declared.

References

1. Huang B, Marcenes W, Croucher R, Hector M. Activities related to the occurrence of traumatic dental injuries in 15- to 18-year-olds. Dent Traumatol 2009;25(1):64-68. [doi: 10.1111/j.1600-9657.2008.00685.x] [Medline: 19208012]
2. Hall E, Hickey P, Nguyen-Tran T, Louie J. Dental Trauma in a pediatric emergency department referral center. Pediatr Emerg Care 2016;32(12):823-826. [doi: 10.1097/PEC.0000000000000961] [Medline: 27898626]
3. Schiappa M, Palla B, Callahan NF. What are the incidence and costs associated with interfacility emergency department transfer patients treated by oral and maxillofacial surgeons? Oral Surg Oral Med Oral Pathol Oral Radiol 2024;138(5):608-612. [doi: 10.1016/j.oooo.2024.06.008] [Medline: 39217075]
4. Eggmann F, Haschemi AA, Doukoudis D, Filippi A, Verna C, Walter C, et al. Impact of the COVID-19 pandemic on urgent dental care delivery in a Swiss university center for dental medicine. Clin Oral Investig 2021;25(10):5711-5721 [FREE Full text] [doi: 10.1007/s00784-021-03872-1] [Medline: 33710460]
5. Namakian M, Subar P, Glassman P, Quade R, Harrington M. In-person versus "virtual" dental examination: congruence between decision-making modalities. J Calif Dent Assoc 2012;40(7):587-595. [Medline: 22916380]
6. Davies A, Mueller J. Developing Medical Apps and mHealth Interventions. Cham, Switzerlad: Springer; 2020.
7. Harris B, Ajisola M, Alam RM, Watkins JA, Arvanitis TN, Bakibinga P, et al. Mobile consulting as an option for delivering healthcare services in low-resource settings in low- and middle-income countries: a mixed-methods study. Digit Health 2021;7:20552076211033425 [FREE Full text] [doi: 10.1177/20552076211033425] [Medline: 34777849]
8. Mariño RJ, Uribe SE, Chen R, Schwendicke F, Giraudeau N, Scheerman JFM. Terminology of e-oral health: consensus report of the IADR's e-Oral Health Network Terminology Task Force. BMC Oral Health 2024;24(1):280 [FREE Full text] [doi: 10.1186/s12903-024-03929-z] [Medline: 38419003]
9. Xiao J, Meyerowitz C, Ragusa P, Funkhouser K, Lischka TR, Mendez Chagoya LA, National Dental Practice-Based Research Network Collaborative Group. Assessment of an innovative mobile dentistry ehygiene model amid the covid-19 pandemic in the national dental practice-based research network: protocol for design, implementation, and usability testing. JMIR Res Protoc 2021;10(10):e32345 [FREE Full text] [doi: 10.2196/32345] [Medline: 34597259]
10. Huang B, Estai M, Pungchanchaikul P, Quick K, Ranjitkar S, Fashingbauer E, et al. Mobile health assessment of traumatic dental injuries using smartphone-acquired photographs: a multicenter diagnostic accuracy study. Telemed J E Health 2024;30(10):2592-2600 [FREE Full text] [doi: 10.1089/tmj.2024.0012] [Medline: 38946689]
11. Estai M, Kanagasingam Y, Huang B, Checker H, Steele L, Kruger E, et al. The efficacy of remote screening for dental caries by mid-level dental providers using a mobile teledentistry model. Community Dent Oral Epidemiol 2016;44(5):435-441. [doi: 10.1111/cdoe.12232] [Medline: 27111291]
12. Birur PN, Sunny SP, Jena S, Kandasarma U, Raghavan S, Ramaswamy B, et al. Mobile health application for remote oral cancer surveillance. J Am Dent Assoc 2015;146(12):886-894. [doi: 10.1016/j.adaj.2015.05.020] [Medline: 26610833]
13. Abuabara A, de Castro J, Locks M, Pezzin A, Mattos N, de Araújo CM, et al. Evaluation of Endo 10 mobile application as diagnostic tool in endodontics. J Clin Exp Dent 2023;15(8):e612-e620 [FREE Full text] [doi: 10.4317/jced.60342] [Medline: 37674608]
14. Azimi S, Fernando C, Estai M, Patel J, Silva D, Tennant M. Experience of primary caregivers in utilising an mHealth application for remote dental screening in preschool children. Aust Health Rev 2023;47(5):545-552. [doi: 10.1071/AH23110] [Medline: 37580061]
15. Ibarra-Noriega A, Yansane A, Mullins J, Simmons K, Skourtes N, Holmes D, et al. Evaluating and improving the usability of a mHealth platform to assess postoperative dental pain. JAMIA Open 2024;7(1):ooae018 [FREE Full text] [doi: 10.1093/jamiaopen/ooae018] [Medline: 38476372]
16. Van Cleave JH, Fu MR, Bennett AV, Concert C, Riccobene A, Tran A, et al. The usefulness of the Electronic Patient Visit Assessment (ePVA) as a clinical support tool for real-time interventions in head and neck cancer. Mhealth 2021;7:7 [FREE Full text] [doi: 10.21037/mhealth-19-250] [Medline: 33634190]

XSL•FO
RenderX

17. Tokede B, Yansane A, Ibarra-Noriega A, Mullins J, Simmons K, Skourtes N, et al. Evaluating the impact of an mHealth platform for managing acute postoperative dental pain: randomized controlled trial. JMIR Mhealth Uhealth 2023;11:e49677 [FREE Full text] [doi: 10.2196/49677] [Medline: 37933185]
18. Kalenderian E, Tungare S, Mehta U, Hamid S, Mungia R, Yansane A, et al. Patient and dentist perspectives on collecting patient reported outcomes after painful dental procedures in the National Dental PBRN. BMC Oral Health 2024;24(1):201 [FREE Full text] [doi: 10.1186/s12903-024-03931-5] [Medline: 38326805]
19. Del Carmen MDC, Cagigas-Muñiz D, García-Robles R, Oprescu AM. Reducing dental anxiety in children using a mobile health app: Usability and user experience study. JMIR Form Res 2023;7:e30443 [FREE Full text] [doi: 10.2196/30443] [Medline: 37889521]
20. Pereira-Salgado A, Westwood JA, Russell L, Ugalde A, Ortlepp B, Seymour JF, et al. Mobile health intervention to increase oral cancer therapy adherence in patients with chronic myeloid leukemia (The REMIND System): clinical feasibility and acceptability assessment. JMIR Mhealth Uhealth 2017;5(12):e184 [FREE Full text] [doi: 10.2196/mhealth.8349] [Medline: 29212628]
21. Almazrooa SA, Mansour GA, Alhamed SA, Ali SA, Akeel SK, Alhindi NA, et al. The application of teledentistry for Saudi patients' care: a national survey study. J Dent Sci 2021;16(1):280-286 [FREE Full text] [doi: 10.1016/j.jds.2020.04.014] [Medline: 33384810]
22. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989;13(3):319-339.
23. Wallis L, Blessing P, Dalwai M, Shin SD. Integrating mHealth at point of care in low- and middle-income settings: the system perspective. Glob Health Action 2017;10(sup3):1327686 [FREE Full text] [doi: 10.1080/16549716.2017.1327686] [Medline: 28838302]
24. Scauso M. Interpretivism: Definitions, Trends, and Emerging Paths. Oxford, UK: Oxford Research Encyclopedia of International Studies; 2024.
25. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
26. Health Insurance Portability and Accountability Act of 1996. Office of the Assistant Secretary for Planning and Evaluation. 1996. URL: https://aspe.hhs.gov/reports/health-insurance-portability-accountability-act-1996 [accessed 2025-11-21]
27. Guest G, Namey E, Chen M. A simple method to assess and report thematic saturation in qualitative research. PLoS One 2020;15(5):e0232076 [FREE Full text] [doi: 10.1371/journal.pone.0232076] [Medline: 32369511]
28. Klingberg A, Sawe HR, Hammar U, Wallis LA, Hasselberg M. m-Health for burn injury consultations in a low-resource setting: an acceptability study among health care providers. Telemed J E Health 2020;26(4):395-405 [FREE Full text] [doi: 10.1089/tmj.2019.0048] [Medline: 31161967]
29. Moore GC, Benbasat I. Development of an instrument to measure the perceptions of adopting an information technology innovation. Inf Syst Res 1991;2(3):192-222. [doi: 10.1287/isre.2.3.192]
30. Compeau D, Higgins C. Computer self-efficacy: development of a measure and initial test. MIS Q 1995;19(2):189-211.
31. Fishbein M, Ajzen I. Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Boston, USA: Addison-Wesley; 1975.
32. Thompson RL, Higgins CA, Howell JM. Personal computing: toward a conceptual model of utilization. MIS Q 1991;15(1):125-143.
33. Naeem M, Ozuem W, Howell K, Ranfagni S. A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. Int J Qual Methods 2023;22(11). [doi: 10.1177/16094069231205789]
34. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
35. Scheerman JFM, Qari AH, Varenne B, Bijwaard H, Swinckels L, Giraudeau N, et al. A systematic umbrella review of the effects of teledentistry on costs and oral-health outcomes. Int J Environ Res Public Health 2024;21(4):407 [FREE Full text] [doi: 10.3390/ijerph21040407] [Medline: 38673320]
36. Emami E, Harnagea H, Shrivastava R, Ahmadi M, Giraudeau N. Patient satisfaction with e-oral health care in rural and remote settings: a systematic review. Syst Rev 2022;11(1):234 [FREE Full text] [doi: 10.1186/s13643-022-02103-2] [Medline: 36309732]
37. Rathert C, Wyrwich MD, Boren SA. Patient-centered care and outcomes: a systematic review of the literature. Med Care Res Rev 2013;70(4):351-379. [doi: 10.1177/1077558712465774] [Medline: 23169897]
38. Howell SE, Fukuoka B. Teledentistry for patient-centered screening and assessment. Dent Clin North Am 2022;66(2):195-208. [doi: 10.1016/j.cden.2022.01.002] [Medline: 35365273]
39. National Research Council. In: Druckman D, Bjork RA, editors. Learning, Remembering, Believing: Enhancing Human Performance. USA: National Academies Press; 1994.
40. Birur N, Patrick S, Bajaj S, Raghavan S, Suresh A, Sunny S, et al. A novel mobile health approach to early diagnosis of oral cancer. J Contemp Dent Pract 2018;19(9):1122-1128 [FREE Full text] [Medline: 30287715]
41. Bakhshaei A, Ramachandran S, Brondani M. Teledentistry within oral health care providers' training: a scoping review. Eur J Dent Educ 2024;28(2):631-644. [doi: 10.1111/eje.12990] [Medline: 38279780]

XSL•FO
RenderX

42. McLeod CD, Adatorwovor R, Brame JL, White BA, Weintraub JA. Teledentistry: dental hygiene students' knowledge, attitudes, and curriculum recommendations. J Dent Hyg 2021;95(5):64-72. [Medline: 34654717]
43. Giles-Smith L, Spencer A, Shaw C, Porter C, Lobchuk M. A study of the impact of an educational intervention on nurse attitudes and behaviours toward mobile device use in hospital settings. J Can Health Libr Assoc 2017;38(1):12-29.
44. Safi S, Thiessen T, Schmailzl KJ. Acceptance and resistance of new digital technologies in medicine: Qualitative study. JMIR Res Protoc 2018;7(12):e11072 [FREE Full text] [doi: 10.2196/11072] [Medline: 30514693]
45. Cavanagh R, Gerson SM, Gleason A, Mackey R, Ciulla R. Competencies needed for behavioral health professionals to integrate digital health technologies into clinical care: a rapid review. J Technol Behav Sci 2022:1-14 [FREE Full text] [doi: 10.1007/s41347-022-00242-w] [Medline: 35194561]
46. Thurzo A, Thurzo V. Embedding fear in medical AI: A risk-averse framework for safety and ethics. AI 2025;6(5):101. [doi: 10.3390/ai6050101]
47. Carter N, Bryant-Lukosius D, DiCenso A, Blythe J, Neville AJ. The use of triangulation in qualitative research. Oncol Nurs Forum 2014;41(5):545-457. [doi: 10.1188/14.ONF.545-547] [Medline: 25158659]
48. Leung FH, Savithiri R. Spotlight on focus groups. Can Fam Physician 2009;55(2):218-219 [FREE Full text] [Medline: 19221085]
49. Knutson TJ, Komolsevin R, Chatiketu P, Smith VR. A cross-cultural comparison of Thai and US American rhetorical sensitivity: implications for intercultural communication effectiveness. Int J Intercult Relat 2003;27(1):63-78.
50. Groom LL, McCarthy MM, Stimpfel AW, Brody AA. Telemedicine and telehealth in nursing homes: an integrative review. J Am Med Dir Assoc 2021;22(9):1784-1801.e7 [FREE Full text] [doi: 10.1016/j.jamda.2021.02.037] [Medline: 33819450]
51. Guest G, Bunce A, Johnson L. How many interviews are enough? Field Methods 2006;18(1):59-82. [doi: 10.1177/1525822x05279903]
52. de Souza J, Gillett K, Salifu Y, Walshe C. Changes in participant interactions. Using focus group analysis methodology to explore the impact on participant interactions of face-to-face versus online video data collection methods. Int J Qual Methods 2024;23(2). [doi: 10.1177/16094069241241151]
53. Roiha A, Iikkanen P. The salience of a prior relationship between researcher and participants: reflecting on acquaintance interviews. Res Method Appl Linguistics 2022;1(1):100003. [doi: 10.1016/j.rmal.2021.100003]
54. Lee J, Fung S, Yong R, Ranjitkar S, Kaidonis J, Evans AR, et al. Tooth wear development in the Australian Aboriginal dentition from Yuendumu: a longitudinal study. PLoS One 2021;16(7):e0254151 [FREE Full text] [doi: 10.1371/journal.pone.0254151] [Medline: 34242291]
55. Pitiphat W, Luangchaichaweng S, Pungchanchaikul P, Angwaravong O, Chansamak N. Factors associated with molar incisor hypomineralization in Thai children. Eur J Oral Sci 2014;122(4):265-270. [doi: 10.1111/eos.12136] [Medline: 24924351]
56. Tanny L, Huang B, Shaweesh A, Currie G. Characterisation of anterior open bite in primary school-aged children: a preliminary study with artificial neural network analysis. Int J Paediatr Dent 2021;31(5):576-582. [doi: 10.1111/ipd.12759] [Medline: 33222319]
57. Huang B, Takahashi K, Sakata T, Kiso H, Sugai M, Fujimura K, et al. Increased risk of temporomandibular joint closed lock: a case-control study of ANKH polymorphisms. PLoS One 2011;6(10):e25503 [FREE Full text] [doi: 10.1371/journal.pone.0025503] [Medline: 22003394]
58. Shenouda M. Dental photographer perception of participant attitude with the use of teledentistry [Thesis]. Minneapolis, MN, USA: University of Minnesota; 2024.
59. Schultz EC, Huang B, Shenouda M, Estai M, Ranjitkar S, Louie JP, et al. In-Review: Perspectives of Front-Line Clinicians and Remote Reviewers on Smartphone-Based Photography for Assessing Traumatic Dental Injuries: A Qualitative Study. United States: University of Minnesota URL: https://conservancy.umn.edu/items/51c73533-eb65-4257-a6c2-33895147b00b [accessed 2026-01-06]

Abbreviations

AI: artificial intelligence
COREQ: Consolidated Criteria for Reporting Qualitative Research
CSIRO: Commonwealth Scientific and Industrial Research Organisation
HIPAA: Health Insurance Portability and Accountability Act
mHealth: mobile health
PI: principal investigator
TAM: Technology Acceptance Model
TDI: traumatic dental injury

XSL•FO
RenderX

Edited by A Kushniruk; submitted 19.Aug.2025; peer-reviewed by R Musa, A Thurzo; comments to author 12.Sep.2025; revised version received 15.Dec.2025; accepted 02.Jan.2026; published 26.Jan.2026.

Please cite as:
Schultz EC, Huang B, Shenouda M, Estai M, Ranjitkar S, Louie JP, Pungchanchaikul P
Perspectives of Frontline Clinicians and End-Line Users on Smartphone-Based Photography for Assessing Traumatic Dental Injuries: Focus Group Interview Study and Thematic Analysis
JMIR Hum Factors 2026;13:e82668
URL: https://humanfactors.jmir.org/2026/1/e82668
doi:10.2196/82668
PMID:

©Emily C Schultz, Boyen Huang, Margaret Shenouda, Mohamed Estai, Sarbin Ranjitkar, Jeffrey P Louie, Patimaporn Pungchanchaikul. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 26.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Evaluation of the Implementation of a Mobile Health App to Support Dutch Primary Care for Diabetes: Qualitative Study

Liselot N van den Berg[1,2], MSc; Lisenka te Lindert[1,2], MSc; Jiska J Aardoom[1,2], PhD; Anke Versluis[1,2], PhD; Sofie H Willems[1,2], MD, PhD; Niels H Chavannes[1,2], MD, PhD; Marise J Kasteleyn[1,2], PhD

[1]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands

[2]National eHealth Living Lab, Leiden, The Netherlands

Corresponding Author:
Liselot N van den Berg, MSc
Department of Public Health and Primary Care
Leiden University Medical Center
Hippocratespad 21
Leiden, 2300 RC
The Netherlands
Phone: 31 71 526 57 61
Email: L.N.van_den_Berg@lumc.nl

Abstract

Background: Over 1 million Dutch people have diabetes, of whom 90% have type 2 diabetes. Studies show that lifestyle plays an important role in the course of type 2 diabetes. MiGuide (MiGuide Ltd) is an online platform that helps people adopt and sustain lifestyle changes. The platform is integrated into existing diabetes care within primary care. Previous research has shown that implementing new (eHealth) interventions is challenging and may reduce effectiveness. Mapping out the barriers and success factors in the implementation process is essential so that eHealth interventions such as MiGuide can be used effectively in regular health care.

Objective: This study aimed to evaluate the implementation of MiGuide within Dutch primary care.

Methods: A qualitative study design was used, supplemented by quantitative data from patients. Five general practices participated. Three focus groups (FGs; at baseline, after 6 months, and after 12 months) were conducted with 3 general practitioners, 3 FGs with 8 specialized practice nurses (divided into 2 separate groups with 4 participants per group), 2 FGs (at 6 months and after 12 months) with 5 patients, and 2 FGs (at baseline and after 12 months) with 4 stakeholders from the management of the care group. The implementation process was discussed with health care professionals and management, and usage and user-friendliness were discussed with patients. The framework method was used to analyze the data. The following quantitative data were collected: patient characteristics, user data, and questionnaires at baseline and 6 months, assessing quality of life, usability, and diabetes self-care. The quantitative data were examined using exploratory analyses.

Results: Four themes were found in the qualitative data: “innovation,” “capability, motivation, and opportunity,” “processes,” and “setting.” Different factors within these themes played an essential role throughout the implementation process, such as facilities, technical difficulties, motivation, COVID-19, and the work processes. Areas for improvement were also identified. The supplemented quantitative data showed that usability scored below average at 6 months (mean 53.8; SD 9.3; n=8). Participants had a mean score of 0.84 (SD 0.13) on the EuroQoL-5 dimension and 81.9 (SD 13.4) on the EuroQoL visual analogue scale at baseline. Moreover, the average number of days someone exercised was 4.2 (SD 1.7), and the number of days someone ate a generally healthy diet was 5.1 (SD 1.3). Insufficient data on quality of life and diabetes self-care were collected at 6 months and therefore not presented in this study.

Conclusions: Implementation is a complex process with multiple barriers and facilitators. It is essential to explore the use of context-specific strategies that are aligned with the implementation process phase. Further research is needed to evaluate the next version of the MiGuide platform, which is being implemented in another setting with lifestyle coaches.

(JMIR Hum Factors 2026;13:e54431) doi:10.2196/54431

XSL•FO
RenderX

KEYWORDS

diabetes care; implementation; eHealth; lifestyle; primary care; evaluation; mobile health; mHealth; diabetes; qualitative study; Dutch; type 2 diabetes; MiGuide; framework method; quality of life; barriers; facilitators; context-specific; quantitative data; patients

Introduction

In 2018, almost 1 million people were registered with type 2 diabetes mellitus (T2DM) in Dutch primary care [1]. This number has more than doubled since 2004 [2]. Approximately 90% of these people receive treatment in primary care [1].

Following Dutch treatment guidelines for standard diabetes care, patients receive an annual consultation with their general practitioner (GP) and a quarterly or biannual consultation with a specialized practice nurse (SPN). The consultations with the SPN include monitoring of weight, systolic blood pressure, hemoglobin A_{1c} levels, foot condition, and the occurrence of hypo- or hyperglycemia. Individuals with T2DM are also advised to lose weight and adopt a healthy lifestyle, such as sufficient physical activity and a healthy diet. Previous studies have shown that lifestyle plays an essential role in the disease progression of people with T2DM [3-8]. For example, increased physical activity, a healthier diet, and weight loss can lead to remission of T2DM [3,5]. According to the guideline, education on a healthier lifestyle should be provided alongside self-management strategies [9].

Self-management, which is the ability to manage symptoms, treatment, physical consequences, psychosocial consequences, and lifestyle changes, is inherent in living with diabetes [10]. The usage of eHealth self-management interventions for T2DM, such as smartphone apps, has demonstrated potential for adopting a healthier lifestyle. This especially applies to apps targeting diet and physical activity, which can improve health outcomes such as hemoglobin A_{1c}. Although self-management apps have potential, it is recommended to offer eHealth as a blended care solution [11].

The success of self-management interventions regarding lifestyle and diabetes depends on the fidelity of the implementation process [12]. Implementation fidelity pertains to the extent to which an intervention is used as intended. Key elements include adherence, intervention complexity, facilitation strategies, quality of delivery, and participant responsiveness [13]. By evaluating whether an intervention has been implemented with fidelity, researchers and practitioners can better understand how and why an intervention works and the extent to which outcomes can be improved. This is crucial because there is often a gap between theory and practice in eHealth implementation. Often, eHealth solutions do not make it to practical use, and when they do, the adoption process is slower than anticipated due to various complex, multilevel challenges [14-17]. For instance, the patient, health care, and organizational levels should be considered [17].

For this study, we followed the implementation of MiGuide (MiGuide Ltd) in primary care. MiGuide is implemented as an integrated part of diabetes care and consists of an app and an online portal. The app allows individuals to track their food intake, physical activity, and specific goals. Lifestyle challenges, such as completing exercise goals each day and consultation preparation, are also available. Measurements, such as blood pressure and weight, are uploaded to the portal, which is accessible to GPs and SPNs. The content of MiGuide is based on the Dutch professional guideline for standard diabetes primary care. This study aims to examine the implementation of MiGuide in Dutch primary care. Our objectives are to identify the barriers and facilitators to the implementation of MiGuide in Dutch primary care and to evaluate the usage of MiGuide.

Methods

Setting

This study was conducted among Dutch general practices affiliated with the care group named the GP team (in Dutch: “Het Huisartsenteam” [HHT]) located in the province of Noord-Brabant. A care group is a partnership of several general practices to support logistics and the quality of care [12,18]. The care group HHT cooperated with MiGuide Ltd in an implementation project for the platform MiGuide. Five practices, for which it was technically feasible to link with the MiGuide portal, implemented MiGuide as part of their usual diabetes care. A management team with implementation experts guided the implementation process. It consisted of developers from MiGuide Ltd, eHealth implementation advisors from an external party, and the care group’s board members. MiGuide Ltd provided all instruction materials, provided the experts for training the GPs and SPNs, and was on standby for problem-solving and support. The care group had an intermediary role and informed the general practices.

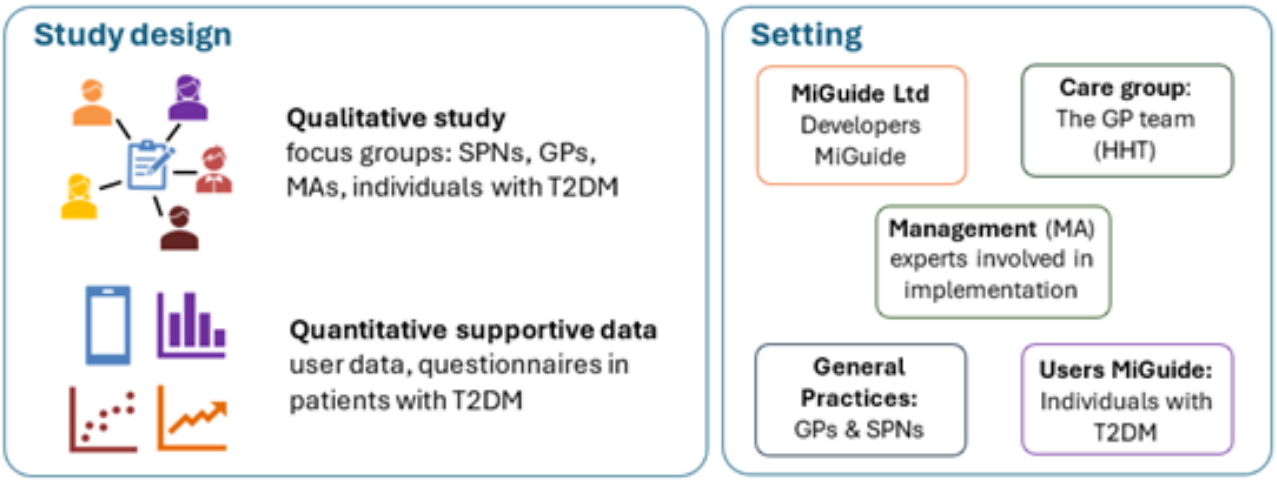
Study Design

This was a descriptive qualitative study, supplemented by quantitative data from individuals with T2DM using MiGuide. The study was conducted in Dutch primary care between October 2020 and August 2022.

The “implementation fidelity model” [13] was used to examine whether the intervention was used as intended. To evaluate the implementation process, the following aspects were examined: barriers and facilitators, the impact on care pathways, including the transition of care from practice to home, and the role and experiences of GPs and SPNs. We assessed the implementation process at three levels: (1) the practice level from the perspective of GPs and SPNs, (2) the management level from the perspective of experts from the care group and MiGuide Ltd involved in the implementation process (ie MAs), and (3) the user level from the perspective of patients. Quantitative data were used to assess patients’ objective usage of MiGuide, their perceived system usability, and the impact on quality of life and self-care. Figure 1 gives an overview of the design, setting, and outcomes.

XSL•FO
RenderX

Figure 1. Overview of the study design, setting, timeline, and outcomes. EQ-5D: EuroQoL-5 dimension; EQ-VAS: EuroQoL visual analogue scale; GP: general practitioner; HHT: the GP team (in Dutch: Het Huisartsenteam); MA: management; QoL: quality of life; SDSCA: Summary of Diabetes Self-Care Activities; SPN: specialized practice nurse; SUS: System Usability Scale; T2DM: type 2 diabetes mellitus.



Timeline and outcomes

Qualitative

Focus group	Timepoint 1 Nov and Dec 2020	Timepoint 2 Apr and May 2021	Timepoint 3 Oct and Nov 2021
SPN group 1 (n)	4	3, and 1 interview	3
SPN group 2 (n)	4	3	2
GP (n)	3	2, and 1 interview	3
MA (n)	4		5
Patients (n)		4	3, and 1 interview

Quantitative

Data	Measurements	T0 (baseline)	T1 (6 months after start)
Baseline characteristics	Diabetes duration, medication, gender, length, weight, year of birth	x	
Usage MiGuide patient level	Number of logins; number and type of selected challenges; challenge completed yes/no; physical activity; type of physical activity; duration; number of consultation preparations; number of registered food diary entries		x
Questionnaires	QoL (EQ-5D and EQ-VAS)	x	x
	SUS		x
	SDSCA	x	x

Study Population and Procedure

Patients were eligible to participate in the quantitative study if they (1) were diagnosed with T2DM, (2) were able to read and communicate in Dutch, (3) were treated in a general practice within the care group, (4) had a smartphone (Android version 6.0 or higher [Google], or iOS version 11.0 or higher [Apple Inc]) with access to data connectivity, and (5) were not being treated in a hospital. During the initial phase of MiGuide's release, one of the exclusion criteria was the participant's unwillingness to share their data with their health care provider. This criterion was dropped when the second release of the app was launched in July 2021. The inclusion criteria for GPs and SPNs in the qualitative study were (1) being affiliated with the care group and (2) working with MiGuide. Patients also needed to use MiGuide to participate in the focus groups (FGs).

XSL•FO
RenderX

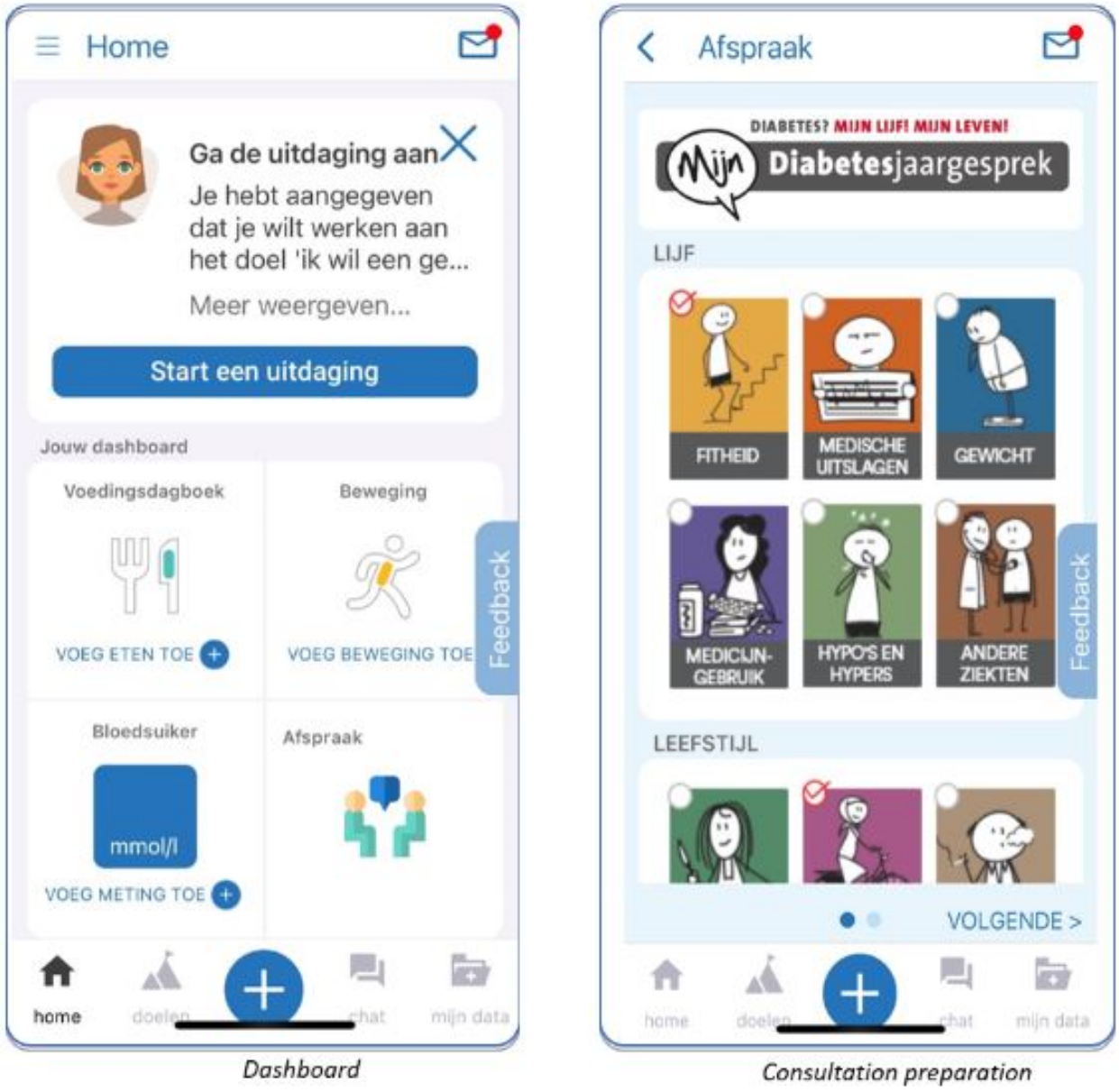
The general practices sent an invitation letter to eligible patients describing MiGuide and the study. Patients could download the app from the App Store or the Google Play Store. At onboarding, individuals indicated whether they agreed to share (1) their user data with their general practice and (2) their email address for research participation. MiGuide could still be used when an individual did not agree. Next, MiGuide Ltd asked for consent via email. Before the second release, users could only use MiGuide when they agreed to share their data for research purposes. As this was not a voluntary choice nor a valid consent, all participants who started using MiGuide before the second release were actively approached to ask for their participation in this study. At the end of the onboarding, patients could visit their GP to verify their identity and receive a unique QR code, but this was not mandatory to use MiGuide. The QR code facilitated the data connection between MiGuide, the health care professional (HCP), and the patient's medical data. No reimbursement was given for research participation.

For the qualitative part of the study, the care group invited SPNs, GPs, and MAs via email. Patients already using MiGuide were invited to participate via newsletters sent by MiGuide Ltd. All FG members were asked to consent to FG participation via email by the study team. After participation, they received a book about lifestyle and a reimbursement from MiGuide Ltd.

MiGuide Platform

MiGuide Ltd developed a platform consisting of an app and an online portal to guide individuals with T2DM in changing their behavior to adopt a healthier lifestyle (refer to Figure 2). The content is based on the content of consultations with GPs and SPNs (as described in Dutch treatment guidelines for diabetes in primary care).

Figure 2. Components of MiGuide.



XSL•FO
RenderX

The components of MiGuide are organized into the following categories: food diary, physical activity diary, goal setting, and data. The food diary has been improved since the second release by connecting it to an existing app from the Netherlands Nutrition Centre. The physical activity diary offers the possibility to connect with Google Fit. Goal setting offers different goals to enhance control of the following diabetes determinants: blood sugar, blood pressure, sufficient physical activity, prevention or minimization of complications, cholesterol levels, healthy eating and drinking, and healthy weight. The data category comprises the possibility to record blood pressure, blood glucose, weight, and consultation preparation. This data can be shared with the information system of the user's general practice if the user agrees to do so. The consultation preparation component allows users to send notes to their general practice about themes they want to discuss during a consultation with their GP or SPN. The themes they can choose from consist of physique, lifestyle, well-being, and self-care.

Data Collection

Qualitative Study

FGs and interviews were conducted using video conferencing platforms (Microsoft Teams or Jitsi). The FGs and interviews were audio-recorded.

FGs with GPs, SPNs, and MAs took place (1) during the initial adoption phase, (2) at 3 months after the implementation, and (3) at 12 months after the implementation (end of the study period). The FGs with individuals with T2DM took place (1) 3 months after initiating the use of MiGuide and (2) at the end of the study period.

To reach data saturation, we aimed for 5 GPs, 10 SPNs (divided into 2 groups), 4 MAs, and 4 patients. In cases where it was not possible to gather all members of an FG simultaneously, separate interviews were conducted to collect additional data.

Quantitative Study

Figure 1 provides the usage data of MiGuide, along with the questionnaire data collected at the user level. The usage data were gathered through an automated process. Users received in-app notifications at baseline and at 6 months to complete the questionnaires, with demographic characteristics collected only at baseline. Subsequently, MiGuide Ltd sent manual email reminders when participants had not completed the questionnaires.

Analysis

Qualitative Study

A thematic analysis based on the framework method procedure [19] was performed. At stage 1, the audio of all interviews was automatically converted to transcripts and pseudonymized. Two researchers (LvdB and LtL) became familiar with the interviews (stage 2) by checking and supplementing the transcription and making notes and summaries of the interviews. At stage 3, the researchers LvdB and LtL started coding the first transcripts using a mainly deductive approach. A concept codebook was created before coding, based on the implementation fidelity model and the above-mentioned aspects to evaluate the implementation process. The first 2 FGs with SPNs and the first of the other predefined FGs were coded independently by LvdB and LtL. At stage 4, the analytical framework (codebook) was further developed by comparing the independently coded FGs and agreeing on the definition of codes and categories by LvdB and LtL. If applicable, new codes were added to the codebook. The codes and categories were then applied to the remaining transcripts at stage 5. These transcripts were divided by LvdB and LtL and coded by one researcher. At stage 6, the data were charted into the framework matrix. Data from each transcript were summarized and provided with illustrative quotes. Each predefined FG was treated as a different unit of analysis, thereby assessing the implementation process at different levels: patient, GP, SPN, and MA. Finally, at stage 7, characteristics and differences between the data were identified. This process was supervised by senior researchers MJK and SHW.

Quantitative Study

Normally distributed numerical data were described as mean (SD). Nonnormally distributed numerical data were described as median with IQR. Categorical data were described as frequencies with percentages. Descriptive statistics were performed for the collected quantitative data using IBM SPSS Statistics (version 25) [20].

Ethical Considerations

The study did not fall within the scope of the Dutch Medical Research Involving Human Subjects Act, according to the Medical Research and Ethics Committee of VU Amsterdam Medical Center (2020.0708). Subsequently, a declaration of no objection was obtained from the Medical Research and Ethics Committee. Participants provided informed consent and were able to opt out from the study (refer to the "Procedure" section). Both quantitative and qualitative data were collected pseudonymously.

Results

Demographic Characteristics

In the qualitative study, 3 patients were male (3/5, 60%), all 9 SPNs were female (9/9, 100%), all 3 GPs (3/3, 100%) were male, and 5 MAs were male (5/6, 83%).

In the quantitative study, 58 participants were enrolled. Half were male (29/58, 50%), the mean age was 60.1 (SD 12.1), the median BMI was 30.3 (IQR 25.9-35.8), and 19% (11/58) of the participants were paired with the MiGuide portal. The demographic characteristics and number of paired patients per participating practice are provided in Table 1.

XSL•FO
RenderX

Table 1. Baseline data of participating practices.

Characteristic	Practice 1 (n=4)	Practice 2 (n=42)	Practice 3 (n=7)	Practice 4 (n=5)	Total (N=58)
Paired patients, n (%)	0 (0)	8 (19)	2 (28.6)	1 (20)	11 (19)
Age (years), mean (SD)	60.6 (7.0)	61.6 (11.3)	50.5 (11.7)	60.7 (19.3)	60.1 (12.1)
Male, n (%)	2 (50)	23 (56.1)	1 (14.3)	2 (50)	29 (50)
BMI[a], median (IQR)	27.9 (24.8-32.9)	30.3 (26.8-36.4)	32.3 (28.7-35.8)	25.7 (22.6-40.3)	30.3 (25.9-35.8)

[a]BMI data were complete for all participants except one individual in practice 2.

Quality of Life and Self-Care Baseline

At baseline, the mean EuroQoL-5-dimension score was 0.84 (SD 0.13), and the mean EuroQoL visual analogue scale score was 81.9 (SD 13.4). The mean number of days participants exercised was 4.2 (SD 1.7), and the mean number of days they followed a generally healthy diet was 5.1 (SD 1.3; refer to Table 2). Insufficient data were collected at 6 months (by 3, 4, and 6 participants, respectively); therefore, these data are not presented.

Table 2. Quality of life and self-care scores at baseline.

Questionnaire	Baseline value
EQ-5D (n=21), mean (SD)	0.84 (0.13)
EQ-VAS[a] (n=16), mean (SD)	81.9 (13.4)
SDSCA[b] (n=13), mean (SD)	
No of days of physical activity or exercise	4.2 (1.7)
No of days of glucose monitoring	1.4 (2.0)
No of days of foot care	1.5 (1.7)
No of days of a healthy diet (general)	5.1 (1.3)
No of days of a healthy diet (specific)	5.3 (0.9)
Nonsmoking, n/N (%)	9/11 (81.8)

[a]EQ-VAS: EuroQoL visual analogue scale.

[b]SDSCA: Summary of Diabetes Self-Care Activities.

Innovation

Patients were enthusiastic about the MiGuide concept but mentioned that the app was not intuitive for navigation. Later, during FG2, they reported that the app's design improved and looked clean and clear. They often used the MiGuide app daily, ranging from 5 to 60 minutes.

The 2 most frequently used functions were the food diary and physical activity tracking. Nevertheless, patients thought the food diary was too generic. They were also critical of the physical activity component: the point system was difficult to interpret. Some used pedometers and wondered why their results were not directly visible in the app. MiGuide Ltd immediately processed the food diary feedback after FG1, resulting in more enthusiastic patients during FG2. They were also positive regarding plans to change the physical activity functionality at a later stage. In the quantitative study, MiGuide usability at 6 months had a mean score of 53.8 (SD 9.3; n=8; range 20-90), which was below average.

Patients indicated that they rarely used consultation preparation; some used challenges, goals, and daily tips. This was also seen in the quantitative data (refer to Figures S1-S3 in Multimedia Appendix 1). During the FGs, patients stated that consultation preparation was not used because they thought it was not very useful for the SPN or themselves. For example, measurements were already available for SPNs, or patients wanted to discuss their T2DM during a consultation. They also mentioned that they already used the care group's app (ie, the HHT [Pharmeon BV] app) to access blood results and visits.

All patients, except one, said they expected to keep using MiGuide long-term. They expected a positive course of their T2DM while using MiGuide because it could motivate them to increase physical activity and eat healthier. Greater insight allowed them to make more impactful lifestyle changes. During FG2, they stated that they became more aware of their T2DM and that the app provided insight into their lifestyle.

> *For me, the app works very well. I mean, if I decide that day to misbehave and drink some craft beers, delicious, on Saturday night. Then I will put it in the app and think, "yes, this was not the best for me, but it was delicious." I think if you improve the app further, you will become aware of yourself. I think that is the most important part of the app. This is the reason I fill it in loyally.* [PT3 during FG1; April 22, 2021]

XSL•FO
RenderX

> *I have to admit, I never thought I would when they told me to lose 30 kilos first… At the time, I did not think it was a dot on the horizon; it was a dot behind the dot. And afterwards, with some attention, you can lose 20 kilos quite easily.* [PT3 during FG2; September 23, 2021]

SPNs and GPs mentioned that the MiGuide platform has potential, may be insightful, and looks nice. Initially, navigation on the platform was suboptimal. SPNs mentioned during all FGs that they were not working much with MiGuide and were not actively promoting it in their practices (refer to "Processes" and "Setting" sections). The GPs stated that mainly the SPNs worked with the app. They both experienced a low number of users (2-5 patients using the app per practice); while downloads increased over the year, SPNs indicated that patients stopped using the app, mainly due to the large number of notifications. One GP also said that patients stated that they were already using the HHT app or another app that helps with lifestyle changes and T2DM. The SPNs and GPs noticed that consultation preparation was not being used. Moreover, patients decided which features they wanted to use in the app without the guidance of the SPNs.

> *One of my SPNs has also downloaded the app herself and also tries to fill in things herself, but it does not seem to be very user-friendly. For example, with filling in food: if you have eaten salmon, for example, you have to click a lot to fill in the right amount of what you have eaten. The moment you have entered certain goals and you do not do it, you receive messages at the most impossible times that you have not achieved it or have not done it yet. This then causes people to say, "I do not need to get messages like that at midnight." Then they are done with it very quickly.* [GP1 during FG2; April 15, 2021]

At all 3 time points, the SPNs and GPs reported technical difficulties, mainly in linking the app with their information system via the portal and the QR code. The question arose among SPNs during FG1 if the connection with the information system was necessary and would add value. One of the GPs stated that they were happy that MiGuide was not directly linked to the information system and that they could find the data in a separate section instead of directly in the PT's record.

> *The problem with the fact that you cannot send via a safe environment is, of course, a pity. The link has not yet been made with [name of the information system]. Yeah, otherwise, it did not work out this week, so I have a hotline with [name of MiGuide employee], and they are resetting it, and yet, it is not running yet.* [SPN4 during FG1; November 10, 2020]

During FG1, the SPNs were asked about their expectations of MiGuide on PTs' lifestyle and self-management. They believed the app would enable better monitoring of individuals with T2DM and facilitate exchanging information. They found it useful for patients with an affinity for eHealth, and the app could provide patients with valuable insights into their lifestyle. However, there was some concern about whether the app would meet the needs of patients and whether they would view it as a self-management tool or would instead rely heavily on the SPNs for support. One of the GPs also stated that it is a great app for patients, with many interesting features, and that it would be better to focus on the self-management part of the app.

> *If PTs enjoy using such an app, I would like to encourage them to do so. Because I think it is a great app for PTs in itself, with a lot of interesting things in it, such as lifestyle advice and things like that. […] I think those things that the PT can do themselves that they are good. That is where I would focus. So, then it is a self-management app.* [GP2 during FG3; November 9, 2021]

During FG1, someone from MiGuide Ltd explained that MiGuide aimed to improve the patients' self-management. The patients could get more insight into their disease, improve their lifestyle using MiGuide, and hopefully have another type of conversation with their HCP due to the use of the consultation preparation function. A care group (HHT) representative emphasized that MiGuide aligned perfectly with their 5-year plan to prioritize the creation of a continuum between the patient and practice and the establishment of a connection between HCPs and patients, thereby providing blended care.

> *Which, of course, we see in recent years, that piece of lifestyle medicine, that it is increasingly well substantiated and that it really will become the leading way to apply it. At the same time, it is also complex to implement and introduce. So, it is not just the chronic disease itself, but also the rise of lifestyle medicine where we will find substance.* [MG1 during FG1; December 2, 2020]

During FG1, the MAs discussed that all patients should be invited to use MiGuide instead of inviting based on assumptions (eg, not asking because of the patient's age). However, accepting that the app is not for everyone is essential. HCPs could become frustrated if they aim to reach 100% of patients. The first 10%-20% could be easily reached; thus, HCPs should focus on this group. Knowing the benefits of using the platform could make it more appealing to those not initially enthusiastic about it. It may not work to convince individuals already resistant to new technology. Instead, a more promising approach may be to target individuals who already do a lot online (eg, planning e-consultations).

> *So yes, ask everyone, but also accept the no. There is already enough extra energy to get this off the ground, but do not try to pull the person up to their knees in the mud. They will follow. Focus on the group that wants it now and is also aware that it is not a solution for everyone, so the "digital unless" is also one that often passes by; there are always patients who are not language-skilled enough or who do not master this part. Accept that too, because if you think you have to achieve 100 percent, you will only frustrate yourself.* [CZ1 during FG1; December 2, 2020]

During FG2, the MAs discussed the requirements to initiate changes in the process at the patient level. A transition is needed for patients to have autonomy over their health. Concerns

XSL•FO
RenderX

remained about whether the process was clearly defined for the SPN and the patient and whether they could check the progress of their goals. Someone from HHT wondered whether the MiGuide platform could create an extra burden for SPNs.

The MAs envisioned a transition toward increased self-management for the patients, less workload in the practices, and eventually a decrease in the number of consultations.

Capability, Motivation, and Opportunity

The SPN or GP informed the patients about MiGuide and invited them to use it. During FG1, patients mentioned that they had not yet paired MiGuide with the QR code. One patient was wondering whether the connection would be acceptable to the GPs, as they had a high workload. However, the enthusiasm from their HCP encouraged the patients to use the app.

One patient reported difficulties logging into the app; 9 out of 10 times, they needed to start the app twice before logging in. This problem emerged a couple of weeks (2 or 3 weeks) before FG2 and resulted in less motivation to use it. Due to the availability and use of the HHT app, some patients explained that they did not need MiGuide (refer to the "Innovation" section).

The patients expressed that it is a matter of self-discipline, structure, and persistence to use or keep using the app.

> *I think the fact that you have diabetes should be enough. I think the motivation to fill it in should come from within yourself. If it does not come from yourself, then I do not think you will get it done.* [PT3 during FG1; April 22, 2021]

The enthusiasm of both patients and MiGuide Ltd stimulated the SPNs to work with MiGuide. The importance of enthusiasm from others was repeated throughout all FGs. The SPNs stated during the last FG that it would motivate and activate their enthusiasm if they were reminded by personal visits instead of only reminder emails from MiGuide Ltd or the care group.

> *In our role, we are really like, "Oh, that is new, and we want that, and we are going to work on that." Well, yes, SPN7 and I, if I speak for the two of us, we are also really always at the forefront of rolling out new things, implementing and informing patients about that, and collaborating…* [SPN9 during FG1; December 10, 2020]

During FG1 and FG2, SPNs expressed that they had little experience with MiGuide and needed more knowledge before working with the platform. They agreed that extra training was necessary and that more experience would help them use MiGuide in the long run. One GP stated that it should become completely "dummy-proof." MiGuide Ltd appreciated this feedback and organized extra demo sessions and masterclasses for HCPs to learn about the platform. The demo sessions helped motivate and activate the SPNs, as they became more confident in informing patients. Only some GPs followed the masterclass and demo session, but they were satisfied with the information they received.

During the initial implementation of MiGuide, SPNs mentioned that GPs did their best to be involved and to stimulate the use of MiGuide. However, over time, GP involvement decreased, and the SPNs felt less supported. They also mentioned that they did not experience much facilitation and support from the care group (refer to the "Processes" section). On the other hand, MiGuide Ltd and its helpdesk were accessible and helpful, which was valued.

Moreover, the FGs helped motivate SPNs by exchanging thoughts about MiGuide. The SPNs and GPs stated that the motivation and enthusiasm of HCPs were essential facilitators. Many HCPs were skeptical, leading to friction when patients wanted to use the app. It would help to see why the innovation is valuable (eg, a decrease in the workload). The added value could also be used in other practices to motivate them to implement the platform.

> *You know, if I tell a patient that they can come twice a year, I do not want to have to answer 80 more messages in those six months, because then I would rather have that they just come every three months, instead of twice a year, if I need to have so much other contact with the patient.* [SPN5 during FG3; October 6, 2021]

A participant from HHT explained that the 5 practices were informed by MiGuide Ltd, and documentation was prepared for these practices (eg, a manual and webinar). The initial release date was delayed due to technical issues (refer to the "Innovation" section). In the starting phase, the care group frequently contacted the first 2 practices and MiGuide Ltd about the implementation, which was acknowledged as a facilitator. During FG2, a participant from MiGuide Ltd stated that contact with the practices decreased while communication seemed essential. Nonetheless, practices were able throughout the implementation phase to contact MiGuide Ltd.

During FG2, the MAs also noticed resistance from the HCPs to use MiGuide as a standard method. SPNs and GPs felt that the benefit was too minimal and not immediately noticeable (it cost a lot of energy or time, and they did not immediately gain something in return). The MAs experienced the same issues while changing telephone appointments to digital (online) appointments.

> *That is true, and I am certainly not going to argue that online appointments are an advantage for the assistants. But make no mistake that they spent quite a while sabotaging you to prevent people from making online appointments. And why? Because that effect was not immediately apparent. They thought only when they experienced the effect, "We have to start enthusing about it." But before that, we were almost literally on top of it, just to keep saying, "No, do it now, do it now. The benefit will come to use later." That it will come later, that is just tricky.* [HHT2 during FG2; October 14, 2021]

Funding was an essential facilitator for implementing and integrating MiGuide into the practices. However, this could be a problem in the long term because there is no structural funding yet. The MAs also noticed that the FGs resulted in increased motivation among the SPNs to use MiGuide.

XSL•FO
RenderX

Processes

The SPNs expressed in FG1 and FG3 that the implementation of MiGuide was imposed upon them by HHT (ie, in a top-down manner). They indicated that the implementation process could have been improved if they had been allowed to provide input. The GPs also acknowledged these issues concerning the implementation within their practices. The top-down approach, along with other factors outlined under "setting," resulted in MiGuide having lower priority for both SPNs and GPs.

> *I think HHT wants to be upfront and do new things and try new things, and that is also a lot of fun. And sometimes you think, "Yes, it is already completely thought through in all layers," because the bosses behind their desks say, "Okay, we are going to do this, and MiGuide and is fun; we give it a chance." And eventually, we are the ones who need to do it. Yes, and then the question is whether it is thought through, like, do the SPNs have time? How should they do it? Do they have enough training, support, etc.? Yes, it sometimes is; yes, maybe not high enough on the list, I think.* [SPN7 during FG3; October 13, 2021]

As mentioned by GPs and MAs under the "Capability, Motivation, and Opportunity" subheading, it would motivate HCPs to know that the implementation of MiGuide would help reduce their workload in the longer term. SPNs were aware that there would initially be an increased workload due to the changes in their work processes (eg, becoming accustomed to the app and preparation time). During FG3, SPNs anticipated that the workload would further increase due to the potential for patients to ask questions via the app and the need to personalize the program for all users. However, they reported no differences in MiGuide use on daily care, workload, and job satisfaction since the previous FG. On the other hand, GPs expected during FG1 that the function of consultation preparation could positively influence SPNs' consultations because of the possibility of low-key communication.

> *With the daily process, not a lot, I would say. With diabetes, I have an average of 100; thus, then we still have over 2000 patients. So yes, I think it is a bit, as I stated earlier. I hope that patients are motivated to do more with their lifestyle and that they will receive feedback that it will work. And I hope, on the other hand, that the health care professional gets something out of it because patients will get to the consultation better prepared. That you have already answered their questions for an important part. Yes, and I hope that the app's usage over a longer period can replace the number of consultations. That these can be dropped.* [GP2 during FG1; November 17, 2020]

> *I had one patient who did prepare, but only 20 minutes before their appointment. They stated, "Oh yes, I forgot, but I need to do this for you." Yes, then I think, you do not need to do it for me; you should do it for yourself. And what they select, I cannot work with that, because this patient has a lot of mental issues, and they want to discuss this during the consultation. Yes, I cannot help them with that.* [SPN5 during FG3; October 6, 2021]

During FG1, MAs recognized the need for management to implement an innovation. They play a crucial role in providing necessary support, such as organizing webinars, creating guides, offering workarounds, being available for questions, motivating and stimulating individuals involved in the implementation process, and showing the progress of the implementation. One important aspect highlighted by the participants was that digital projects often differ from initial expectations. However, this difference is not viewed as a problem since continuous development and integration of feedback usually lead to improvements.

One aspect that needs to be considered is that it is difficult during implementation to immediately have a fully operational innovation. Therefore, managing expectations and ensuring clear communication about what aspects work and what aspects do not work is crucial. Failing to address these factors from the beginning could result in disappointment. Providing workarounds (ie, guidelines on what to do when encountering issues with the platform) and offering more intensive implementation coaching could facilitate a smoother transition. The first experience with the platform and the timing of the implementation were also pivotal, and greater emphasis could have been placed on the testing phase.

> *[…] but there are also a lot of them who take that step with a bit more difficulty than we do, and when they run into things like that, you quickly have an excuse to say, "Hey, wait a minute. That should not be done yet." That does then cause a delay, and it also ensures that you do not yet have the results, as you just asked for the study OZ1. If it is all there, and what do we think of it in practice?* [CZ1 during FG1; December 2, 2020]

A representative from HHT experienced difficulty with promoting MiGuide in the practices since it deviated from the regular work processes. Adopting MiGuide required HCPs to shift from a traditional medicine approach to a lifestyle medicine approach. The question remained as to how MiGuide could be integrated into daily routines. As mentioned under the "Capability, Motivation, and Opportunity" subheading, gaining insight into reports could be essential in demonstrating the results of SPNs' efforts. Communicating and gathering feedback could contribute to improving the implementation.

Setting

During all FGs, the SPNs and GPs expressed that there was too much going on in their practices, which negatively affected the implementation of MiGuide. The MiGuide platform was implemented during the COVID-19 pandemic, and this influenced the implementation process (eg, other work processes were needed during the pandemic). Furthermore, other factors in the setting greatly influenced the implementation process of MiGuide: reorganization, absenteeism, turnover, and being too occupied with usual diabetes care (ie, many disrupted patients due to COVID-19).

XSL•FO
RenderX

> *I have not paired any new people, and I also notice that I have experienced a very hectic summer here in the practice, so I was actually happy to be able to do my daily tasks, and because of that, I have made MiGuide a lower priority. Maybe not the idea, but I was glad that my days were done at some point or that I had done what I had to do, and yes, I found the summer quite tough, let's put it that way.* [SPN5 during FG3; October 6, 2021]

During FG1, SPNs stated that it would be easier to implement MiGuide when sitting across patients to show the app. After a year, GPs noticed that patients were coming back to the practice after the COVID-19 pandemic for regular consultations. These consultations could be the perfect opportunity to discuss MiGuide with the patients. They noticed that the strength of the MiGuide platform is the connection with the practice and that MiGuide Ltd could focus more on promoting this aspect (refer to "Innovation" subheading).

The MAs also recognized that the COVID-19 pandemic influenced the implementation process of MiGuide. For example, the explanation about MiGuide was held digitally instead of physically in the practices. This resulted in uncertainties about how the MiGuide platform, especially the app, worked. They also acknowledged other barriers in the setting of the practices: pregnancy leaves, absenteeism, turnover, and multiple programs (eg, Acceleration Program Information Exchange Patient and Professional 5 [in Dutch: Versnellingsprogramma Informatie-uitwisseling Patiënt en Professional 5] and the transition from one information system to another) being implemented and used in the practices.

During FG1, the ecosystem of the care group was mentioned as a great facilitator. The fit between MiGuide and the goals of HHT helped with the commitment to the project. The MAs also noticed that the process could have changed if people had met physically in the practices. It also helps when practices have embraced lifestyle medicine (refer to the "Processes" subheading).

> *At the beginning, of course, there was also the part that everyone was going to work more from home and more digitally. If that had not been the case, we might have been able to take MiGuide physically into the practices in the beginning, as was initially planned. So, I do not know if that would have changed the process, but that is perhaps still a point of attention.* [HHT1 during FG1; December 2, 2020]

Suggestions for Improvement

During the FGs, improvement possibilities regarding the MiGuide app were discussed. In this section, only feedback regarding not yet implemented features or processes is discussed.

FGs With Patients

Patients suggested several improvements for the MiGuide app. These included (1) the implementation of speech-to-text functionality, which would not only streamline the app's usability for the patients themselves but would also be helpful for the older population and individuals with visual impairments; (2) enhanced tracking of physical activity, with a preference for incorporating a pedometer or a calorie-based system for the physical activity feature; (3) the addition of medication tracking, including a medication management feature within the app; (4) a streamlined login procedure to simplify the login process and enhance user experience; and (5) the introduction of engaging physical activity challenges, reflecting a desire for more challenging elements regarding the physical activity feature.

FGs With SPNs and GPs

Suggestions regarding the MiGuide platform included (1) reduction of questionnaires to minimize the number of questionnaires for patients; (2) customizable notifications, allowing patients to personalize the frequency of notifications; (3) improved personalization, such as linking automatic questions to specific outcomes—for instance, when an outcome shows an increase, the platform can generate questions regarding potential factors such as medication changes or stress levels, which can then be discussed during consultations between the SPN and patient to identify strategies for improving outcomes; (4) a notification for consultation preparation integrated into the information system to alert HCPs when a patient has prepared their consultation; and (5) automated synchronization of blood pressure measurements to enhance user-friendliness.

Suggestions regarding the implementation of MiGuide included (1) preimplementation testing, in which a willing person could test the app and give feedback on their experience, allowing SPNs to familiarize themselves with the app, gain valuable insights, and ensure a smooth integration into their practices. This test phase also facilitates a foolproof system; (2) streamlining the manual by creating a concise version to enhance clarity and comprehension of MiGuide; (3) facilitating knowledge exchange among SPNs and GPs, including exchanging tips to improve the implementation process, and evaluating the current implementation process with colleagues; (4) involvement of GPs and the care group (HHT) to contribute to smoother adoption and usage of the MiGuide platform, with 2 or 3 responsible individuals designated to oversee implementation to ensure effective coordination and support; (5) involvement of other HCPs, such as physiotherapists or lifestyle coaches, to foster greater engagement and usage of the platform; (6) improving visibility via commercials, such as health insurance campaigns or newsletters from HHT, to encourage patients to use the app; (7) emphasis on report insights and best practices, providing practical examples to offer valuable guidance and facilitate evidence-based decision-making during implementation, while highlighting that workload would decrease after the implementation of MiGuide because this is experienced as the biggest motivator for HCPs to implement and use MiGuide; and (8) integration with the HHT app rather than functioning as a stand-alone tool, ensuring seamless integration into the existing HHT app and prioritizing ease of use for both patients and HCPs.

FGs With MAs

Suggestions regarding the MiGuide platform included (1) expansion of the MiGuide platform to include other chronic diseases beyond its current scope, such as CVRM and obesity. Suggestions regarding the implementation of MiGuide included (1) preimplementation planning and testing, in which it is

XSL•FO
RenderX

recommended to incorporate a phase for testing and developing a plan or workaround before the actual implementation; this tailored approach ensures that all aspects, including timing, motivation, technical processes, and user-friendliness, are thoroughly addressed to optimize the overall implementation process; (2) improved communication with HCPs and enhanced expectation management, as clear communication and proactive management of expectations will contribute to a smoother adoption and usage of the MiGuide platform, with interventions, comprehensive instructions, gathering feedback, and fostering a collaborative approach helping to improve communication; (3) securing structural funding to ensure the sustainability of the implementation, with efforts directed toward securing funding for ongoing support and maintenance of the MiGuide platform; (4) focus on engaging participating patients, acknowledging that achieving 100% engagement may not be feasible and that the focus should be on the group that is motivated and enthusiastic; (5) emphasis on report insights and best practices, as reports and practical examples offer valuable guidance and facilitate evidence-based decision-making during the implementation process; (6) integration with lifestyle programs or interventions, considering that the MiGuide platform is not yet a part of diabetes care and exploring opportunities to integrate the platform with existing lifestyle programs or interventions to enhance its utility, with integrating a lifestyle coach could further enhance the app's effectiveness; (7) involvement of other HCPs, such as physiotherapists or lifestyle coaches, who have more frequent interactions with individuals with T2DM and can foster greater engagement and usage of the platform; and (8) acceptance of trial and error, recognizing the iterative nature of the implementation process and being aware of the trial-and-error approach that allows for continuous learning, adjustment, and refinement.

Discussion

Principal Findings

The study examined the implementation of the MiGuide platform within 5 Dutch practices of a care group. During the implementation, it became clear that the adoption phase took longer than expected. The group of users was also lower than anticipated; however, this group was motivated and enthusiastic. When implementing an innovation, such as MiGuide, it is necessary to focus on engaging a small group of enthusiastic participants, continuously learning and reporting insights from this small group via an iterative process, and concentrating beforehand on expectation management. Furthermore, developing clear guidelines for the implementation and a workaround (ie, guidelines on what to do when encountering issues with the platform) and thorough testing of the innovation could contribute to a successful implementation. As discussed with the HCPs and MAs, the innovation should fit within their daily work, and there should be sufficient involvement from all parties within the practice to stimulate the adoption.

The COVID-19 pandemic had a big impact on the implementation of MiGuide. Other studies also examined the impact of COVID-19 on implementation and found that the pandemic resulted in unexpected situations and that the setting in which the innovation was delivered was essential for the success of the implementation [21-23]. This was also the case for MiGuide: an unforeseen transition from a physical implementation process toward a digital implementation process and more disrupted patients due to the pandemic, which increased the workload of HCPs. Therefore, MiGuide had a lower priority within the list of daily tasks of the HCPs.

Compared to the framework for implementation fidelity [13], adherence was relatively low, and therefore, implementation fidelity was relatively low in this study. Most patients used the app every day for 5-60 minutes, which aligned with the recommended "dose" and "duration"; however, the HCPs were still in the start-up phase (eg, sending out invitations) and were not actively promoting MiGuide 1 year after the start of the implementation. The potential moderators in the framework [13] should also be considered. Some participants stated that the platform was easy to use, while others had more difficulties, especially in terms of navigation. The strategies to facilitate the implementation and quality of delivery were relatively high due to the manual, webinar, and accessible support desk of MiGuide Ltd. This could be further improved by creating a workaround and discussing with the practices what their needs are to facilitate fitting strategies. The last moderator, participant responsiveness (eg, enthusiasm), was high among most patients, while it declined over time for HCPs. All these factors may have affected the implementation process of the MiGuide platform.

A scoping review conducted on the implementation of eHealth interventions for hypertension or diabetes [24] revealed similar barriers hindering the adoption. Among these barriers were technical issues, complexities associated with the integration of nonroutine processes, time and attention requirements for usage, centralized decision-making structures, high turnover, low HCP engagement, limited prioritization of the intervention compared to existing initiatives, shortage of funding, and ambiguities surrounding staff responsibilities during implementation. This review also emphasized that the likelihood of successful implementation increased when interventions align with and become an integral part of existing organizational goals and workflows [24]. Other studies also found similar barriers and facilitators and stated that the challenges around implementation are multilevel and complex [14,16,25]. Furthermore, when innovations are adopted, implementation science is not always considered, even though implementation science may ensure that innovations are making sustainable positive differences [26]. To successfully implement an eHealth innovation, such as MiGuide, it is essential to explore the use of context-specific strategies that are aligned with the implementation process phase [15]. MiGuide Ltd considered the importance of the fit of their innovation with current workflows and therefore changed the implementation setting from GPs and SPNs toward lifestyle coaches [27]. The upscaling of MiGuide is currently being supported through two programs: (1) CooL-MiGuide (a lifestyle intervention program reimbursed under the basic health insurance package in the Netherlands) and (2) MiGuide-Direct (a program focusing on the combined app of glucagon-like peptide-1 medication and lifestyle

XSL•FO
RenderX

interventions). This could help to scale up the implementation of MiGuide to reach a larger population.

Strengths and Limitations

This study had several strengths and limitations. An essential strength was the observational setting, which minimized the interference with the dynamics of daily practice and facilitated the inclusion of perspectives of HCPs and patients. A second strength was that MiGuide Ltd gathered the input after each FG, whereby the most frequently discussed suggestions were discussed and, when possible, addressed.

A limitation was that the quantitative data were insufficient for profound analyses due to the low sample size. This was partly due to the required adjustments in the consent procedure and partly due to a representative reflection of the real situation in which data are not easily obtained because of high dropout. Consequently, we were not able to conduct an extensive quantitative analysis as originally planned. It is therefore necessary to carefully interpret the study data due to the limited generalizability of the findings. In addition, it is essential to be aware of potential selection and social desirability bias, which may have occurred during data collection due to the pseudonymous character of this study. Furthermore, as previously discussed, the COVID-19 pandemic had a great impact on the implementation and the study.

Implications for Future Research and Practice

In this study, a minimal viable product was examined. In a future study, an improved version could be evaluated via a hybrid type II trial [28] with a focus on both implementations, usage, and outcomes. The focus could also be on the integration of the MiGuide platform with a lifestyle program or intervention. It would also be beneficial to include a more diverse range of HCPs, such as lifestyle coaches, and patients from different demographic backgrounds to enhance the generalizability of the findings. This study and the new setting could then be compared regarding barriers and facilitators for the implementation process. Moreover, potential risks—such as an over-reliance on technology or decreased face-to-face interaction between patients and HCPs, and possibilities for mitigating these risks—could be examined. The usage of artificial intelligence could also be a theme to further explore. MiGuide Ltd already incorporates artificial intelligence for two primary purposes: (1) the development of personalized content and (2) the analysis of user data and lifestyle patterns. Further possibilities with artificial intelligence could be explored.

Furthermore, it is essential to acknowledge that an eHealth innovation is a dynamic product that needs to be regularly evaluated and adjusted where necessary. The maintenance of eHealth also needs proper funding, as both are associated with ongoing costs. As stated by MAs and found in previous studies [15,29], not having structural funding is a barrier during the implementation process. Therefore, costs should be considered during further implementation of MiGuide. In a future study, the cost-effectiveness and economic impact of MiGuide, including how potential benefits may be distributed across stakeholders, such as patients and HCPs, could be assessed. A comprehensive cost-effectiveness analysis could also inform evidence-based policy decisions and support the development of sustainable reimbursement models for the broader implementation of digital health interventions such as MiGuide.

In addition, a designated implementation team could improve the success rate of the implementation. For example, a previous study found that implementation efforts with designated teams led to an 80% success rate versus 14% during implementation efforts that did not have such teams [30]. Although MiGuide Ltd gathered a team of experts in implementation, this could be further enhanced throughout the process and adjusted to the implementation phase.

Conclusions

This study evaluated the implementation of MiGuide, an eHealth innovation that promotes a healthier lifestyle among individuals with T2DM, within primary care. The implementation is a complex process, and barriers, also found in previous studies, such as technical issues and time and attention requirements for usage, were established by the participants. Essential factors for a successful implementation are engaging a small group of enthusiastic participants, continuously learning and reporting insights, and concentrating beforehand on expectation management and testing of the app. In addition, it is important to explore the use of context-specific strategies that are aligned with the implementation process phase. Further research is needed to evaluate the new setting with lifestyle coaches in which an improved platform is integrated.

Acknowledgments

We would like to thank Arjen Huizinga, Chief Executive Officer of MiGuide Ltd, who contributed to the data collection and the reviewing and editing of the manuscript.

Funding

This study was funded via a ZorgTech voucher from Innovation Quarter and by the National eHealth Living Lab. Innovation Quarter was not involved in the study design, data collection, interpretation, or manuscript writing.

Authors' Contributions

LNvdB, LtL, JJA, AV, SHW, NHC, and MJK contributed to the conceptualization and reviewed and edited the manuscript. LNvdB and LtL prepared the original draft and conducted the qualitative analysis. MJK conducted the quantitative analysis. LtL provided the visualization. LNvdB and MJK contributed to the methodology, validation, investigation, resources, and data curation.

XSL•FO
RenderX

JJA, AV, SHW, NHC, and MJK supervised the project and the funding acquisition. MJK was responsible for the project administration.

Conflicts of Interest

None declared.

Multimedia Appendix 1
User data MiGuide.
[DOCX File , 27 KB - humanfactors_v13i1e54431_app1.docx]

References

1. Geurten RJ, Elissen AMJ, Bilo HJG, Struijs JN, van Tilburg C, Ruwaard D. Identifying and delineating the type 2 diabetes population in the Netherlands using an all-payer claims database: characteristics, healthcare utilisation and expenditures. BMJ Open 2021;11(12):e049487 [FREE Full text] [doi: 10.1136/bmjopen-2021-049487] [Medline: 34876422]
2. Overbeek JA, Nijpels G, Swart KMA, Blom MT, Elders PJM, Herings RMC. Type 2 diabetes, the epidemic: trends in prevalence and incidence, 2004-2020. Diabetes Metab Syndr Obes 2024;17:1503-1509 [FREE Full text] [doi: 10.2147/DMSO.S445288] [Medline: 38562280]
3. Lean M, Leslie WS, Barnes AC, Brosnahan N, Thom G, McCombie L, et al. Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial. Lancet 2018;391(10120):541-551 [FREE Full text] [doi: 10.1016/S0140-6736(17)33102-1] [Medline: 29221645]
4. Look AHEAD Research Group, Wadden TA, West DS, Delahanty L, Jakicic J, Rejeski J, et al. The look AHEAD study: a description of the lifestyle intervention and the evidence supporting it. Obesity (Silver Spring) 2006;14(5):737-752 [FREE Full text] [doi: 10.1038/oby.2006.84] [Medline: 16855180]
5. Ried-Larsen M, Johansen MY, MacDonald CS, Hansen KB, Christensen R, Wedell-Neergaard AS, et al. Type 2 diabetes remission 1 year after an intensive lifestyle intervention: a secondary analysis of a randomized clinical trial. Diabetes Obes Metab 2019;21(10):2257-2266 [FREE Full text] [doi: 10.1111/dom.13802] [Medline: 31168922]
6. Johansen MY, Karstoft K, MacDonald CS, Hansen KB, Ellingsgaard H, Hartmann B, et al. Effects of an intensive lifestyle intervention on the underlying mechanisms of improved glycaemic control in individuals with type 2 diabetes: a secondary analysis of a randomised clinical trial. Diabetologia 2020;63(11):2410-2422. [doi: 10.1007/s00125-020-05249-7] [Medline: 32816096]
7. Magkos F, Hjorth MF, Astrup A. Diet and exercise in the prevention and treatment of type 2 diabetes mellitus. Nat Rev Endocrinol 2020;16(10):545-555. [doi: 10.1038/s41574-020-0381-5] [Medline: 32690918]
8. Pot GK, Battjes-Fries MC, Patijn ON, Pijl H, Witkamp RF, de Visser M, et al. Nutrition and lifestyle intervention in type 2 diabetes: pilot study in the Netherlands showing improved glucose control and reduction in glucose lowering medication. BMJ Nutr Prev Health 2019;2(1):43-50 [FREE Full text] [doi: 10.1136/bmjnph-2018-000012] [Medline: 33235957]
9. Barents ESE, Bilo HJG, Bouma M, Dankers M, de Rooij A, Hart HE, et al. Diabetes mellitus type 2. NHG-Richtlijnen. 2021. URL: https://richtlijnen.nhg.org/standaarden/diabetes-mellitus-type-2 [accessed 2022-04-09]
10. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]
11. van Beugen S, Ferwerda M, Hoeve D, Rovers MM, Spillekom-van Koulil S, van Middendorp H, et al. Internet-based cognitive behavioral therapy for patients with chronic somatic conditions: a meta-analytic review. J Med Internet Res 2014;16(3):e88 [FREE Full text] [doi: 10.2196/jmir.2777] [Medline: 24675372]
12. van Bruggen S, Kasteleyn MJ, Rauh SP, Meijer JS, Busch KJG, Numans ME, et al. Experiences with tailoring of primary diabetes care in well-organised general practices: a mixed-methods study. BMC Health Serv Res 2021;21(1):1218 [FREE Full text] [doi: 10.1186/s12913-021-07198-2] [Medline: 34753463]
13. Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S. A conceptual framework for implementation fidelity. Implement Sci 2007;2:40 [FREE Full text] [doi: 10.1186/1748-5908-2-40] [Medline: 18053122]
14. van Gemert-Pijnen JEWC, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
15. Versluis A, van Luenen S, Meijer E, Honkoop PJ, Pinnock H, Mohr DC, et al. SERIES: eHealth in primary care. Part 4: addressing the challenges of implementation. Eur J Gen Pract 2020;26(1):140-145 [FREE Full text] [doi: 10.1080/13814788.2020.1826431] [Medline: 33025820]
16. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]

XSL•FO
RenderX

17. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
18. Benchmark Transparante ketenzorg — Rapportage 2019 Zorggroepen: diabetes mellitus, VRM, COPD en Astma.: InEen; 2020. URL: https://ineen.nl/wp-content/uploads/2020/07/200715-Benchmark-Transparante-ketenzorg-2019-web.pdf [accessed 2022-04-09]
19. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
20. SPSS statistics for windows, version 25.0. IBM. 2017. URL: https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-25 [accessed 2020-03-24]
21. Chernogorova Y, Bliznakov Z, Bliznakova K. Management challenges in implementing scientific projects during COVID-19 pandemic. PJMS 2021;23(1):136-150. [doi: 10.17512/pjms.2021.23.1.09]
22. Chambers DA. Considering the intersection between implementation science and COVID-19. Implement Res Pract 2020;1:0020764020925994 [FREE Full text] [doi: 10.1177/0020764020925994] [Medline: 36915854]
23. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. Adoption of digital technologies in health care during the COVID-19 pandemic: systematic review of early scientific literature. J Med Internet Res 2020;22(11):e22280 [FREE Full text] [doi: 10.2196/22280] [Medline: 33079693]
24. Khalid A, Dong Q, Chuluunbaatar E, Haldane V, Durrani H, Wei X. Implementation science perspectives on implementing telemedicine interventions for hypertension or diabetes management: scoping review. J Med Internet Res 2023;25:e42134 [FREE Full text] [doi: 10.2196/42134] [Medline: 36917174]
25. Tossaint-Schoenmakers R, Versluis A, Chavannes N, Talboom-Kamp E, Kasteleyn M. The challenge of integrating ehealth into health care: systematic literature review of the Donabedian model of structure, process, and outcome. J Med Internet Res 2021;23(5):e27180 [FREE Full text] [doi: 10.2196/27180] [Medline: 33970123]
26. Moir T. Why is implementation science important for intervention design and evaluation within educational settings? Front. Educ 2018;3 [FREE Full text] [doi: 10.3389/feduc.2018.00061]
27. Cool-MiGuide. 2023. URL: https://www.loketgezondleven.nl/interventies-zoeken#/InterventionDetails/2300152 [accessed 2023-10-30]
28. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]
29. Ahmed B, Dannhauser T, Philip N. A systematic review of reviews to identify key research opportunities within the field of eHealth implementation. J Telemed Telecare 2019;25(5):276-285. [doi: 10.1177/1357633X18768601] [Medline: 29703127]
30. Fixsen DL, Blasé KA, Timbers GD, Wolf MM. In search of program implementation: 792 replications of the teaching-family model. Behav Anal Today 2007;8(1):96-110. [doi: 10.1037/h0100104]

Abbreviations

FG: focus group
GP: general practitioner
HCP: health care professional
HHT: Het Huisartsenteam
MA: management
SPN: specialized practice nurse
T2DM: type 2 diabetes mellitus

Edited by A Kushniruk; submitted 09.Nov.2023; peer-reviewed by E Spanakis, L Aljerf; comments to author 12.Mar.2024; revised version received 07.Aug.2024; accepted 12.May.2025; published 09.Feb.2026.

Please cite as:
van den Berg LN, te Lindert L, Aardoom JJ, Versluis A, Willems SH, Chavannes NH, Kasteleyn MJ
Evaluation of the Implementation of a Mobile Health App to Support Dutch Primary Care for Diabetes: Qualitative Study
JMIR Hum Factors 2026;13:e54431
URL: https://humanfactors.jmir.org/2026/1/e54431
doi:10.2196/54431
PMID:

XSL•FO
RenderX

©Liselot N van den Berg, Lisenka te Lindert, Jiska J Aardoom, Anke Versluis, Sofie H Willems, Niels H Chavannes, Marise J Kasteleyn. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 09.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Facilitators and Barriers to Over-the-Counter Hearing Aid Use in People With Dementia: Semistructured Interview Study

Dana P Urbanski[1,2], AuD, PhD; Alexander M Hungs[3]; Peggy B Nelson[4,5], PhD; Joseph E Gaugler[6,7], PhD

[1]Department of Speech, Language and Hearing Sciences, College of Arts and Sciences, Indiana University Bloomington, Bloomington, IN, United States

[2]Regenstrief Institute, Indiana University Center for Aging Research, Indianapolis, IN, United States

[3]University of Minnesota School of Public Health, Center for Healthy Aging and Innovation, Minneapolis, MN, United States

[4]Department of Speech-Language-Hearing Sciences, College of Liberal Arts, University of Minnesota, Minneapolis, MN, United States

[5]University of Minnesota Center for Applied and Translational Sensory Science, Minneapolis, MN, United States

[6]Division of Health Policy and Management, School of Public Health, University of Minnesota, Minneapolis, MN, United States

[7]Center for Healthy Aging and Innovation, School of Public Health, University of Minnesota, Minneapolis, MN, United States

Corresponding Author:
Dana P Urbanski, AuD, PhD
Regenstrief Institute
Indiana University Center for Aging Research
1101 W 10th St
Indianapolis, IN, 46202
United States
Phone: 1 812 855 7709
Email: urbanski@iu.edu

Abstract

Background: Over-the-counter (OTC) hearing aids were introduced to improve the affordability and accessibility of hearing health care for adults with perceived mild-to-moderate hearing loss. While these devices have demonstrated effectiveness in cognitively healthy older adults—particularly in the domains of audibility, self-reported hearing ability, and speech recognition in quiet—their use and outcomes in people with dementia remain underexplored. This issue warrants further attention, as people with dementia often experience co-occurring hearing loss and may rely on OTC hearing aids to overcome cost and access barriers to prescription amplification. However, given the cognitive and functional challenges of dementia, it is unclear whether and how OTC hearing aids can support the hearing care needs of these individuals.

Objective: To explore interest-holder perspectives on the feasibility and acceptability of OTC hearing aids for community-dwelling older adults with dementia, identifying key facilitators and barriers that influence their use in this population.

Methods: Semistructured interviews were conducted with 45 participants across three interest-holder groups (15 per group): (1) community-dwelling older adults with dementia and hearing loss, (2) family caregivers of community-dwelling older adults with dementia and hearing loss, and (3) geriatric direct care professionals. Interviews were conducted and recorded via secure Zoom (Zoom Communications) videoconferencing, then transcribed and analyzed using thematic analysis.

Results: Participants endorsed several facilitators and barriers to OTC hearing aid use in people with dementia. Facilitators included increased accessibility, perceived affordability and value, and enhanced autonomy and control. Barriers included mistrust of OTC hearing aids, difficulty assessing candidacy due to unreliable self- and proxy reports of hearing status, caregiver uncertainty regarding device programming and adjustment, challenges evaluating device effectiveness, and concerns about caregiver burden and burnout from long-term device management.

Conclusions: OTC hearing aids offer meaningful advantages for people with dementia and their family caregivers. However, significant barriers must be addressed to ensure their feasibility and acceptability for this population. Future research should further examine and quantify these barriers to inform the development of tailored devices, services, and delivery models that promote successful OTC hearing aid use in people with dementia and their family caregivers.

(JMIR Hum Factors 2026;13:e83857) doi:10.2196/83857

XSL•FO
RenderX

KEYWORDS

over-the-counter hearing aids; dementia; hearing loss; older adults; caregiver; caregiving; assistive technology

Introduction

Background

Newly introduced over-the-counter (OTC) hearing aids have shown effectiveness among cognitively healthy older adults with mild-to-moderate hearing loss, particularly in the domains of audibility, self-reported hearing ability, and speech recognition in quiet [1-5]. However, there is limited evidence for their use in people with dementia. In this manuscript, we explore the perspectives of dementia-care interest-holders [6] on the benefits, challenges, and opportunities of OTC hearing aid use in community-dwelling people with dementia.

OTC Hearing Aids: Reducing Cost and Improving Accessibility

Among American older adults, age-related hearing loss affects about two-thirds of people older than 70 years and more than 80% of those aged 80 years and older [7,8]. For most of these individuals, hearing aids are the most effective therapy for age-related hearing loss [9], offering well-documented health benefits [10-13]. However, hearing aid adoption is low among older Americans with hearing loss—typically ranging from 30% to 40% [8,14,15]. To address this problem, in October 2022, the United States Food and Drug Administration (FDA) introduced regulations for OTC hearing aids, creating a new category of hearing aids available directly to consumers without a prescription. Previously, all hearing aids required a professional evaluation and fitting by a licensed hearing care professional. Today, OTC hearing aids—now widely available in stores and web-based retailers—allow adults with self-perceived mild-to-moderate hearing loss to purchase and fit hearing aids of their choosing. This change is intended to make hearing aids more affordable and accessible by removing the added expense and travel associated with seeing a licensed professional. Proponents of the new regulations anticipate that more affordable, widely available OTC hearing aids will substantially lower cost and access barriers to amplification, thereby facilitating greater hearing aid uptake among approximately 15 million Americans living with untreated hearing loss [8].

The reasons for low hearing aid adoption are multifactorial and include barriers other than cost and access, such as stigma, low readiness to address hearing loss, and poor hearing aid self-efficacy, among other factors [16-19]. Nevertheless, cost and access are formidable barriers to prescription hearing aids for many older adults [20,21]. First, conventional prescription hearing aids are expensive and not covered by Medicare. On average, a pair of prescription devices costs US $3600 [22]. This price may be cost-prohibitive for many older Americans with low or fixed incomes. Indeed, a recent analysis found that an out-of-pocket cost of US $2500 for hearing aids is a catastrophic, unaffordable expense for upwards of three-quarters of American adults with hearing loss [21]. Moreover, prescription hearing aids are geographically inaccessible to many Americans. Recent data indicate that audiologists are unevenly and inequitably distributed across the United States, clustering in metropolitan areas with low rates of self-reported hearing difficulty to the disservice of rural areas with greater self-reported hearing difficulty [23]. In short, those in rural or underserved communities may lack access to prescription hearing aids, irrespective of cost or insurance coverage. Together, the poor affordability and inequitable availability of prescription hearing health care necessitate an alternative approach. For many older Americans, OTC hearing aids offer this much-needed alternative.

OTC Hearing Aids: Considering the Needs of People With Dementia

While OTC hearing aids address cost and access, they also carry broader implications for older adults with hearing loss. OTC hearing aids introduce a new “do-it-yourself” service delivery model in which older adults assume responsibility for selecting, fitting, using, managing, and troubleshooting hearing aids. To date, a small but growing body of research shows that well-selected, generally healthy older adults with intact cognition can successfully self-program, use, and manage OTC hearing aids [1-5]. However, OTC hearing aids are not tested or designed for people with dementia, who may lack the cognitive ability to understand and use these devices. Although traditional prescription hearing care can also present challenges for people living with dementia, the lack of clinician involvement in OTC hearing aids introduces potential challenges that remain unexamined. This gap in the evidence for OTC hearing aids warrants further consideration. Older people with co-occurring dementia and hearing loss are a critical subpopulation of people with untreated hearing loss who—like the general older adult population—will rely on OTC hearing aids to overcome existing cost and access barriers to prescription hearing aids. If people with dementia are to benefit from OTC hearing aids, it is essential to explore their needs, perspectives, and capabilities for using these devices.

OTC Hearing Aids for People With Dementia: Potential Benefits, Challenges, and Opportunities

Like age-related hearing loss, dementia is prevalent in older adults. Approximately 10% of American older adults aged 65 years and older live with dementia, increasing with age to 20% of those aged 80 years and older and 40% for those older than 90 years [24-26]. Many individuals with dementia also experience co-occurring hearing loss. Among those older than 70 years with dementia, 79% have clinically significant hearing loss, rising to 94% of those aged 85 years and older [27]. For those with both conditions, untreated hearing loss can compound the effects of dementia on communication, exacerbating dementia-related behavioral and emotional changes such as depression, agitation, apathy, and social withdrawal [28-30]. Importantly, hearing aids may relieve the compounding effects of hearing loss on dementia. In people with dementia, hearing aid use is associated with reduced depressive, neuropsychiatric, and other problematic behavioral symptoms [28,31-33]. These health benefits broadly mirror the social and emotional health

XSL•FO
RenderX

benefits observed for hearing aid use in the general older adult population. However, fewer than one-third of people with co-occurring dementia and hearing loss use hearing aids [27,34]. In short—like the general older adult population—people with dementia benefit from hearing aids but demonstrate low adoption rates while facing high cost and access barriers to prescription hearing aids. Although people with dementia may face additional barriers to hearing aid adoption, cost and access are important barriers for people with and without dementia. It is thus essential to explore whether and how OTC hearing aids can provide inclusive access to affordable hearing health care for people with dementia.

It remains unclear whether OTC hearing aids are feasible and acceptable for community-dwelling people with dementia and their family caregivers. Without robust evidence on this topic, one can draw parallels between OTC hearing aids and the broader literature on everyday and at-home assistive technology use in older people with dementia. Here, studies consistently find that older adults with cognitive impairment, including dementia, both demonstrate and self-report greater difficulty understanding and managing a variety of everyday technologies compared with their peers without cognitive impairment [35-37]. In quantitative studies, these findings are statistically significant and show large effect sizes [35,37,38]. In qualitative studies, older adults with cognitive impairment report an array of overlapping, compounding challenges for at-home technology use, including memory deficits, difficulty attending to multiple processes, inability to follow sequential instructions, confusion interpreting technology-generated messages, low self-efficacy for technology use, and sensitivity to technology-related stress, among others [36,39,40]. Some people with cognitive impairment even forget what a given technology is intended to do or how it is meant to function [36]. Importantly, studies have found that manufacturer instruction manuals are typically unhelpful for older adults with cognitive impairment, including dementia, who may express optimism in their ability to follow written instructions but often fail to implement them correctly [36].

Considering these challenges, people with dementia will likely need family or other caregiver support to use OTC hearing aids. This is not unexpected, as people with dementia typically need help performing some or all instrumental activities of daily living (IADL), such as housekeeping, preparing food, taking medication, using transportation, shopping, managing finances, and using the telephone [41]. Similarly, people with dementia usually require some level of caregiver assistance for using traditional prescription hearing aids [42,43]. Yet OTC hearing aids have been developed for independent use, without explicit consideration of caregiver roles. In the absence of ongoing professional hearing services and counseling, it is unclear if OTC hearing aids are designed and delivered in ways that facilitate effective family caregiver involvement. It is further unknown if family caregivers are open to and comfortable assuming responsibilities for OTC hearing aid programming, use, care, and maintenance. Research is needed that explores the needs, perspectives, and capabilities of community-dwelling people with dementia and their family caregivers for using OTC hearing aids. This knowledge can inform the development of tailored devices, services, and supports that promote successful OTC hearing aid use in people with dementia and their family caregivers, thus ensuring that people with dementia can benefit from lower-cost, more accessible OTC hearing aids.

Study Objectives

This study explores perspectives on the feasibility and acceptability of OTC hearing aids as a pathway to hearing health care in community-dwelling older people with dementia. We conducted individual semistructured interviews with participants from 3 dementia-care interest-holder groups, including community-dwelling people with dementia and hearing difficulty, family caregivers of people with dementia and hearing difficulty, and geriatric direct care professionals with expertise in the care needs of older people with dementia and hearing difficulty. Using thematic analysis of the interview data, we aimed to identify (1) facilitators or advantages of OTC hearing aid use for community-dwelling older adults with dementia, and (2) barriers or disadvantages of OTC hearing aid use in these individuals. Study reporting follows the Consolidated Criteria for Reporting Qualitative Research (COREQ) guidelines.

Methods

Participant Groups and Eligibility Criteria

Participants were recruited to represent key dementia-care interest-holders directly affected by, or with knowledge pertaining to, OTC hearing aid use in community-dwelling people with dementia. Interest-holders represented three groups (15 each): (1) community-dwelling older adults with self- or caregiver-reported early to midstage dementia and self-perceived mild-to-moderate hearing loss, (2) family caregivers of community-dwelling older people with caregiver-reported early to midstage dementia and caregiver-reported mild-to-moderate hearing difficulty, and (3) geriatric care professionals who provide direct health care services to community-dwelling older adults with co-occurring dementia and hearing loss. Participants were recruited through a research registry of geriatric care professionals and family caregivers of people living with dementia, as well as via web-based advertising and word-of-mouth referrals.

To be eligible for inclusion, community-dwelling older adults with dementia met the following criteria: (1) aged 65 years or older, (2) non–nursing home resident, (3) self- or caregiver-reported diagnosis of early to midstage Alzheimer disease or related dementia, (4) education-adjusted Telephone Interview for Cognitive Status–modified (TICS-m) score of ≤27 out of 50 points consistent with the presence of dementia [44,45], (5) self-perceived mild-to-moderate hearing loss, and (6) ability to communicate in English and follow study procedures. We relied on self-perceived mild-to-moderate hearing loss rather than clinical audiometric test results to align with the FDA's printed candidacy guidelines for OTC hearing aids. Participants were not required to provide audiometric test results and could have any or no prior or current hearing aid experience.

XSL•FO
RenderX

Family caregivers were recruited separately from older persons with dementia (ie, caregivers and individuals with dementia were not enrolled as care dyads) and met the following inclusion criteria: (1) aged 18 years or older, (2) self-identified provider of informal or unpaid health care assistance to a family member who is a community-dwelling older person aged 65 years or older with co-occurring early to midstage Alzheimer disease or related dementia and mild-to-moderate hearing difficulty, (3) self-reported negative history of cognitive impairment or dementia, and (4) able to communicate in English and follow study procedures. Family caregivers could have any or no experience using hearing aids and assisting others in using hearing aids. To ensure study feasibility and accessibility, persons with dementia and family caregivers were not required to supply medical records confirming the dementia or hearing loss status of the person with dementia.

Geriatric direct care professionals were required to be at least 18 years old, able to communicate in English, and providers of direct health care services for community-dwelling older adults, including those with co-occurring dementia and hearing loss. They could be from any profession or geriatric care specialty outside of audiology, otolaryngology, and speech-language pathology. No specific hands-on hearing aid experience or knowledge was required.

Ethical Considerations

This study was approved by the University of Minnesota Institutional Review Board (STUDY00016265). Prior to enrollment, all participants gave informed consent or assent in accordance with institutional review board requirements. Individuals with dementia gave their verbal assent for participation, while written informed consent was obtained from a legally authorized representative. Family caregivers and geriatric direct care professionals provided written informed consent. Participant data were deidentified and stored on secure, password-protected servers to protect privacy and confidentiality. All participants were compensated for their time and contributions.

Descriptive Measures

We collected sociodemographic information from all participants. Additionally, several screening and descriptive measures were administered to characterize participants with dementia, including (1) TICS-m [45], (2) Hearing Handicap Inventory for the Elderly (HHIE) [46], and (3) Lawton-Brody IADL Scale (Lawton and Brody [47]).

We used the TICS-m as a quantitative measure of cognitive function [45]. The TICS-m is a widely used telephone-administered screening test of cognitive functioning designed for use when in-person assessment is impractical or infeasible. The test consists of 11 items (50 points total) covering several cognitive domains, including orientation to time and place, receptive and expressive language ability, verbal memory, calculation, and verbal abstraction. Scores have shown good sensitivity for detecting dementia in older adults when using a cutoff of ≤27 points for education-adjusted TICS-m scores [44]. In this study, TICS-m scores were adjusted for years of education using the corrections in Knopman et al [44] and required to meet the ≤ 27 cutoff, providing additional evidence for participants' cognitive function beyond their self- or caregiver-reported dementia status.

As mentioned previously, we relied on self-perceived hearing difficulty, consistent with FDA candidacy guidelines for OTC hearing aids. To provide additional descriptive information regarding the hearing abilities of participants with dementia, we administered the HHIE [46]. The HHIE is a validated self-assessment of hearing impairment consisting of 25 questions that measure situational and emotional effects of hearing loss in older adults. Each item presents a potential everyday impact of hearing loss, which the respondent rates as "yes," "sometimes," or "no." Each "yes" response is worth 4 points, "sometimes" is 2 points, and "no" is 0 points, giving a maximum possible score of 100 points for the highest level of perceived hearing difficulty. The results give a useful measure of subjective hearing difficulty; however, it should be noted that the HHIE is not specifically validated for use in people with dementia, who face challenges when completing health-related questionnaires. To improve feasibility for our participants, we introduced necessary flexibility into our administration of the HHIE. Whenever possible, we encouraged the older person with dementia to be actively involved in the completion of the HHIE—whether independently or together with a family caregiver. However, when necessary, family caregivers completed the HHIE for the individual with dementia based on their own perception of the person with dementia's hearing difficulty. HHIE scores were only used descriptively and were not used to determine eligibility or classify participants, given the lack of validation in dementia populations.

We used the Lawton-Brody IADL Scale to quantify independent living skills and the level of assistance participants with dementia require for completing IADL tasks that involve self-management skills that may be relevant to OTC hearing aid use [47]. The Lawton-Brody scale includes eight questions covering the following functional domains: using the telephone, shopping, food preparation, housekeeping, laundry, transportation, medication management, and finances. Each item is rated dichotomously as able (one point) or less able (0 points). Points are summed to give a total score, and dependence is defined as a score <8 points for women and <5 points for men due to differences in traditional gender roles [48]. The questionnaire may be administered to either persons with dementia or family caregivers. In this study, we requested that a family caregiver complete the Lawton-Brody whenever possible, as people with dementia may be hesitant to report functional deficiencies in the home. If a family caregiver was unavailable, the person with dementia completed the questionnaire independently.

Interviews

Participants completed an individual interview conducted via secure Zoom (Zoom Communications) videoconferencing. Interviews followed a semistructured interview guide consisting of open-ended questions designed to gather interest-holder-perceived facilitators and barriers to OTC hearing aids for community-dwelling older people with co-occurring early to midstage dementia and hearing loss (Multimedia

XSL•FO
RenderX

Appendix 1). Consistent with semistructured qualitative interviewing techniques, each interview followed a set of predefined questions; however, the interviewer followed the participant's lead in conversation and asked tailored follow-up questions as appropriate. Participants were allowed and encouraged to explore any issue of interest to them at the intersection of hearing loss, communication, hearing aid use, dementia, and dementia caregiving.

Each interview began with a brief introduction to OTC hearing aids and their main differences from prescription hearing aids, focusing on factual distinctions between regulatory categories to avoid biasing participant responses. Subsequent interview questions followed the main steps of the OTC hearing aid consumer pathway, eliciting participant views on determining OTC hearing aid candidacy, purchasing OTC hearing aids, configuring OTC hearing aids, assessing OTC hearing aid benefits, and using, maintaining, and troubleshooting OTC hearing aids. The interview guide was adapted for each participant group but maintained the same overall structure and sequencing of topics to ensure the similarity of interview content across participants. Interview guides for each participant group are included in the Multimedia Appendix 1.

Several accommodations were needed for individuals with dementia. For most participants with dementia, a family member assisted with Zoom and other study logistics. In these cases, family caregivers were instructed to allow the person with dementia to answer interview questions independently and to assist only as necessary to help the participant understand study directions and remain on task. Additionally, interview questions were modified as needed to ensure understandability for participants with dementia, who present with individual differences in cognitive and language ability not well predicted by their self- or caregiver-reported dementia status and quantitative measures of cognitive functioning. However, even with question modifications, participants with dementia often could not answer all interview questions. In these cases, participants with dementia were included in the study if they demonstrated an understanding of the definition of OTC hearing aids through their responses and engaged in the interview by providing independent and relevant answers to a subset of the interview questions.

The interviews were conducted by the first author, a female audiologist (AuD) who, at the time of the study, was a doctoral candidate and clinical researcher with formal training and experience in clinical audiology, qualitative interviewing, and dementia-related research. No prior relationship was established between the interviewer and any participant before study commencement. Participants were informed that the interviewer was an audiologist and researcher conducting a study on perspectives regarding hearing aid use and dementia. To minimize potential bias, participants were not informed that the study focused specifically on OTC hearing aids until the interview. The interviewer has a professional interest in improving hearing care for people living with dementia and remained attentive to ensuring that participants' perspectives guided the interviews.

All interviews were audio-recorded, transcribed verbatim, and anonymized. Interviews typically lasted between 30 and 60 minutes each. No repeat interviews were conducted. Field notes were not taken during or after interviews. The University of Minnesota Institutional Review Board approved the interview guide and all study procedures.

Data Analysis

Interview transcripts were reviewed and analyzed in NVivo 12 (QSR International Pty Ltd, 2022). We used reflexive thematic analysis as our methodological orientation, following Braun and Clarke's [49] six steps for thematic analysis: (1) familiarization, (2) generation of initial codes, (3) search for themes, (4) review, (5) name and define themes, and (6) write-up. First, 2 qualitative coders independently read the deidentified transcripts to familiarize themselves with participant responses (step 1). Then, the coders convened to discuss emerging patterns, using an inductive approach, and generate an initial coding framework (step 2). The coding framework considered all participants together—rather than partitioning them into their interest-holder groups—to facilitate subsequent comparison and contrast of participant perspectives. Next, the 2 coders individually coded one randomly selected transcript from each participant group and convened to compare interpretations and refine/clarify the coding framework. The first author coded all remaining transcripts, with the second coder double-coding 20% of transcripts to ensure consistency of code application. During this process, regular meetings were held to review codes, and disagreements were resolved through discussion. After coding was complete, the coders collated codes into themes and subthemes (step 3), which were subsequently reviewed, named, and defined by the research team (steps 4 and 5). The resulting themes were grouped under 2 main headings, which frame the results narrative (step 6): (1) facilitators or advantages of OTC hearing aid use for community-dwelling older adults with dementia, and (2) barriers to or disadvantages of OTC hearing aid use in these individuals. Formal data saturation thresholds were not prespecified; however, no substantially new themes emerged during later analysis, suggesting that saturation had been reached. Transcripts and themes were not returned to participants for comment, correction, or feedback.

Results

Participant Characteristics

The study included a total of 45 participants split equally between the 3 interest-holder groups (15 each). No eligible participants declined to participate, and no enrolled participants withdrew from the study. Most family caregivers and direct care professionals were female, consistent with published demographics of both groups [50-52]. Family caregivers had a mean age of 62.7 (range 32-79, SD 13.1) years, and just over half had experience assisting their care recipient with hearing aids (n=8). Direct care professionals had a mean age of 46.7 (range 26-74, SD 14) years and all but one (n=14) reported assisting their patients with routine hearing aid use, such as insertion and removal, changing batteries/charging hearing aids, and routine cleaning/care. In contrast with family caregivers

XSL•FO
RenderX

and direct care professionals, people with dementia were predominantly male. This outcome was somewhat unexpected, as Alzheimer or related dementia disproportionately affects women; however, it is also well documented that men are more likely than women to have hearing loss [53]. Participants with dementia had a mean age of 82.1 (range 68-95, SD 7.7) years, and most were current hearing aid users (n=11). Detailed participant characteristics are provided in Tables 1-3.

Table 1. Participant characteristics for persons with dementia (N=15).

Characteristic	Values
Sex, n (%)	
Female	4 (27)
Male	11 (73)
Age (years), mean (SD)	82.1 (7.7)
Education, n (%)	
Technical	2 (13)
High School	3 (20)
Some college	2 (13)
Undergraduate	3 (20)
Graduate	5 (33)
Marital status, n (%)	
Married	9 (60)
Widowed	5 (33)
Divorced	1 (7)
Lives alone, n (%)	
Yes	4 (27)
No	11 (73)
Hearing aid user, n (%)	
Yes	11 (73)
No	4 (27)
TICS-m[a], mean (SD)	17.2 (6.2)
HHIE[b], mean (SD)	36 (21.4)
Lawton-Brody IADL[c], mean (SD)	2.27 (2.2)

[a]Education-adjusted Telephone Interview for Cognitive Status–modified (TICS-m) calculated using corrections from Knopman et al [44].

[b]Hearing Handicap Inventory for the Elderly (HHIE); completed independently by the person with dementia (n=3), by a family caregiver (n=5), or jointly (n=7).

[c]Lawton-Brody Instrumental Activities of Daily Living (IADL) Scale; completed by a family caregiver (n=14) or the person with dementia (n=1).

XSL•FO
RenderX

Table 2. Participant characteristics for family caregivers (N=15).

Characteristic	Values
Sex, n (%)	
Female	14 (93)
Male	1 (7)
Age (years), mean (SD)	62.7 (13.1)
Education, n (%)	
Some college	1 (7)
Undergraduate	5 (33)
Graduate	9 (60)
Relationship to care recipient, n (%)	
Spouse	8 (53)
Child	6 (40)
Grandchild	1 (7)
Lives with care recipient, n (%)	
Yes	11 (73)
No	4 (27)
Care recipient uses hearing aids, n (%)	
Yes	8 (53)
No	7 (47)

Table 3. Participant characteristics for direct care professionals (N=15).

Characteristic	Values
Sex, n (%)	
Female	14 (93)
Male	1 (7)
Age (years), mean (SD)	46.7 (14.0)
Education, n (%)	
Undergraduate	9 (60)
Graduate	6 (40)
Occupation, n (%)	
Registered nurse	8 (53)
Physical therapist	2 (13)
Certified nursing assistant	1 (7)
Clinical mental health counselor	1 (7)
Nurse practitioner	1 (7)
Occupational therapy assistant	1 (7)
Social worker	1 (7)
Years of experience, n (%)	
1-5	3 (20)
6-10	4 (27)
11-15	3 (20)
16-20	2 (13)
21+	3 (20)

XSL•FO
RenderX

Interview Themes

In the following sections, we present key themes and subthemes generated from the thematic analysis process described above. Table 4 outlines these themes and provides exemplar quotes for each.

Table 4. Themes and descriptions.

Theme	Description
Facilitators	
Accessibility	Participants emphasized the ease of obtaining OTC[a] hearing aids and the benefit of eliminating the need for recurrent office visits for hearing aid fittings and adjustments.
Affordability	Participants appreciated the lower cost of OTC hearing aids, with some highlighting their potential to offer better overall value for individuals with dementia.
Autonomy and control	Participants valued the freedom, flexibility, and autonomy to try OTC hearing aids independently, noting that this approach could also reduce stigma associated with traditional hearing care.
Barriers	
Mistrust of OTC hearing aids	Participants expressed skepticism about the quality, customization, and ethics of OTC hearing aids, particularly regarding their ability to meet the needs of people with dementia.
Assessing OTC hearing aid candidacy in people with dementia	Participants were concerned about the ability of people with dementia to accurately self-classify their hearing loss and questioned the reliability of family caregivers' assessments without professional guidance.
Caregiver-facilitated programming and adjustment	Participants doubted whether family caregivers could effectively program and adjust OTC hearing aids without professional support.
Assessing the effectiveness of OTC hearing aids	Evaluating the benefit of OTC hearing aids was seen as challenging, particularly when distinguishing the effects of hearing loss versus cognitive decline on communication.
Ongoing OTC hearing aid use and caregiver burden/burnout	Family caregivers and direct care professionals expressed concerns about the additional responsibilities OTC hearing aids might place on family caregivers, particularly given the challenges of caregiver burden and burnout.

[a]OTC: over-the-counter.

Facilitators of OTC Hearing Aid Use in People With Dementia

Overview

Across the 3 interest-holder groups, participants identified important facilitators/advantages of OTC hearing aid use for community-dwelling older people living with co-occurring dementia and hearing loss. Under this heading, we identified 3 main themes described below.

Facilitator Theme #1: Accessibility

Consistent with the FDA's rationale for OTC hearing aids, participants in all 3 groups expressed optimism that OTC hearing aids can increase access to hearing aids for older people with dementia. Improved accessibility, participants noted, benefits many older adults with hearing loss and not solely those with dementia. However, participants also explained that better access to amplification has particularly high salience in the dementia caregiving context. Family caregivers and direct care professionals highlighted the advantage of eliminating the need for recurrent office visits to fit and adjust prescription hearing aids. Both family caregivers and care professionals emphasized that dementia caregiving is "consuming" [caregiver, wife, age (years) in the range of 60s], involving round-the-clock care, including providing transportation to and from frequent health care appointments for dementia and its complications. From this perspective, a subtheme emerged that family caregivers may view recurrent hearing aid office appointments as "burdensome" [care professional, female, age (years) in the range of 50s], a "struggle" [care professional, female, age (years) in the range of 40s], and "too hard." [caregiver, daughter, age (years) in the range of 60s] As a result, many family caregivers are "apt to not show up for the appointments or not do it at all." [care professional, female, age (years) in the range of 50s] Conversely, participants across groups viewed accessing OTC hearing aids as a fast, easy, and convenient alternative. As one person with dementia remarked, OTC hearing aids are "probably the easiest way to start." [person with dementia, male, age (years) in the range of 90s]

Participants explained that the time and travel involved in prescription hearing aid fittings are not only inconvenient but can also have financial, social, and emotional ramifications. Family caregivers and care professionals explained that people with dementia spend large amounts of time seeing health care providers. This "time investment" [care professional, female, age (years) in the range of 40s] is costly for individuals with dementia and their family caregivers. For caregivers, the cost is often measured in terms of the financial and employment consequences of taking time off work to transport care recipients with dementia to health care appointments.

As one family caregiver remarked:

> *If I have to schedule with the audiologist … it requires me to take time off work and then travel there, so it's*

XSL•FO
RenderX

> *costly in a number of ways, you know?* [caregiver, daughter, age (years) in the range of 60s]

For people with dementia, the cost is measured indirectly in lost time for social engagement and other activities. As one family caregiver commented:

> *Going to the audiologist is not the most exciting thing for her [care recipient]. I mean, she's kind of missing out on other activities and social opportunities and things if she's having to go to a lot of appointments.* [caregiver, granddaughter, age (years) in the range of 30s]

Many family caregivers and care professionals also raised concerns about the emotional impact of health care appointments on people with dementia. They shared that health care visits, including audiology appointments, can be difficult experiences for people with dementia, who may struggle to recognize, navigate, and understand large medical buildings and unfamiliar health care providers. One care professional commented on this challenge for prescription hearing care and shared how OTC hearing aids might help:

> *As they get further into the dementia, going to a new doctor or a place that's different—like having their hearing checked—being in that office situation can be scary for them. So if it's [hearing aid fitting] something they could do at home with a loved one, that would definitely be helpful for them.* [care professional, female, age (years) in the range of 40s]

Similarly, one care professional observed:

> *Sometimes the medical buildings are hard to navigate… and that can be overwhelming for people with dementia. So, I think it'd be easier being able to go to the pharmacy up the street or something they're familiar with to buy hearing aids.* [care professional, female, age (years) in the range of 20s]

Put simply, one person with dementia anticipated that if they decided to purchase OTC hearing aids, "it would be easy to get there." [person with dementia, female, age (years) in the range of 80s]

Facilitator Theme #2: Affordability

Participants in all 3 groups identified affordability as a key advantage or facilitator of OTC hearing aid use in people with dementia. Indeed, participants felt that more affordable hearing aids benefit all older adults, including those with dementia. However, perceptions of cost took on greater nuance in the context of dementia caregiving. Rather than emphasizing cost alone, many family caregivers spoke about value (ie, expected benefit/worth relative to price). The value subtheme was exemplified by a subset of family caregivers who shared that although they could afford prescription hearing aids, they questioned their benefit for people with dementia. Several family caregivers shared that the progression of dementia seemed to make their care recipient content to avoid conversation. They wondered if their relative was interested in hearing better and participating in conversations—and, in turn, whether expensive hearing aids were worth the cost. For example, one family caregiver reflected:

> *The interesting thing about Alzheimer's is like … he's happy in his own little world and so he likes being with people—but no, he doesn't necessarily want to talk … or be part of the conversation. That's not necessarily important to him right now. So, it makes it like … what's it gonna cost me? What's gonna be the benefit?* [caregiver, wife, age (years) in the range of 60s]

Family caregivers also raised concerns about whether people with dementia can take full advantage of the more expensive advanced features and professional fitting services that come with prescription hearing aids. Many noted that Bluetooth technologies, smartphone apps, and other hearing aid accessories are "technically more challenging than people with dementia can work with" [caregiver, daughter, age (years) in the range of 70s] and would, therefore, go unused. Several family caregivers also questioned whether people with dementia can provide sufficiently meaningful, reliable input to justify the added cost of professional hearing aid adjustments. As one family caregiver said,

> *It might be better that they go to an audiologist and really get the exact fit for what they need, but my husband is at the point right now where he can't tell you what the exact right fit is, so that's not applicable to him.* [caregiver, wife, age (years) in the range of 60s]

Finally, family caregivers expressed concern about the expense of prescription hearing aids relative to their high risk of being lost or damaged by people with dementia. One family caregiver explained:

> *A lot of dementia patients lose things. So honestly, if I think he's [care recipient] gonna lose his hearing aids again, I'm gonna go with the cheapest ones around because they're only gonna last until the next time he loses 'em.* [caregiver, wife, age (years) in the range of 60s]

Importantly, many family caregivers felt that lower-cost OTC hearing aids could mitigate their concerns about the value of hearing aids for people with dementia. For example, one family caregiver remarked:

> *I would hate to get those professional hearing aids and either they're lost or damaged, or she can't utilize them in the way they need to be. Yeah, I could swallow it a little bit easier if it's a thousand-dollar OTC hearing aids.* [caregiver, daughter, age (years) in the range of 50s]

This and other family caregiver responses suggest OTC hearing aids may help shift caregivers' value calculations in favor of purchasing and trying amplification.

Facilitator Theme #3: Autonomy and Control

Another facilitator theme emerged around control—or the freedom, flexibility, and autonomy to efficiently try OTC hearing aids and make decisions about them. Across all 3 groups, several participants saw OTC hearing aids as an opportunity for people with dementia and their family caregivers to "scope out hearing aids for themselves" [care professional,

XSL•FO
RenderX

female, age (years) in the range of 50s] and more rapidly evaluate their benefits. As one family caregiver put it:

> *I think that, in general, to have a fast way … to just get a hearing aid to your person and see if it works is a wonderful, wonderful thing.* [caregiver, wife, age (years) in the range of 60s]

Relatedly, many family caregivers and care professionals highlighted the advantage of quickly gauging the likelihood of hearing aid tolerance and acceptance in people with dementia. One care professional explained:

> *With OTC hearing aids, they [family caregivers] could probably get a general idea of if they're [care recipients] gonna be compliant with hearing aids right away.* [care professional, female, age (years) in the range of 30s]

Additionally, a prominent subtheme emerged around stigma. Direct care professionals stressed the deep-seated stigma and "embarrassment" [care professional, female, age (years) in the range of 30s] that accompany a diagnosis of age-related hearing loss. Receiving this diagnosis, participants explained, means a person is "labeled as old now" [care professional, female, age (years) in the range of 40s] and—especially for community-dwelling persons with dementia—may invoke fear of losing independence. As a result, those with dementia may avoid seeing a hearing care professional altogether. Several care professionals felt OTC hearing aids could help circumvent stigma by giving older adults with dementia and their caregivers the autonomy and control to try hearing aids "on the sly without actually having it in their medical records." [care professional, female, age (years) in the range of 50s] One care professional explained:

> *They [community-dwelling people with dementia] don't want hearing loss to be another thing that makes them feel incapable of taking care of themselves … They're at home. They don't want to move, so then they don't want to bring attention to their hearing.* [care professional, female, age (years) in the range of 20s]

Many care professionals and family caregivers felt that some people with dementia would be more open to OTC hearing aids, which allow them to try amplification without "admitting to a loss." [care professional, female, age (years) in the range of 50s] or "involving other people." [person with dementia, male, age (years) in the range of 90s] As one family caregiver reflected:

> *It might be easier for him [care recipient] to accept not going in to see a doctor … There might be an advantage that I could at least introduce him to hearing aids this way.* [caregiver, wife, age (years) in the range of 60s]

Barriers to OTC Hearing Aid Use in People With Dementia

Overview

Although participants described meaningful potential facilitators/advantages of OTC hearing aids for people with dementia, they also identified several unaddressed barriers or disadvantages. Below, we describe 5 main barrier themes.

Barrier Theme #1: Mistrust of OTC Hearing Aids

Across all 3 groups, many participants expressed a general mistrust of OTC hearing aids. This theme emerged from a variety of statements revealing apprehension, hesitance, and skepticism about the quality and effectiveness of OTC hearing aids and their associated support services. Specifically, some participants raised concerns that OTC hearing aids may lack sufficient customization of their frequency responses to match individual hearing losses. These participants viewed OTC hearing aids as simple "amplifiers" [caregiver, wife, age (years) in the range of 70s; care professional, female, age (years) in the range of 60s; person with dementia, male, age (years) in the range of 70s], which they believed to be too "generic" [caregiver, daughter, age (years) in the range of 60s; caregiver, wife, age (years) in the range of 60s] for people with hearing loss—especially those with dementia whose hearing loss is further complicated by the effects of cognitive decline. A participant with dementia commented:

> *I am suspicious of [OTC hearing aids] … because it's not talking about my hearing in each ear. It is trying to be an amplifier. I got a friend who doesn't wanna pay for [prescription hearing aids]. He says, 'I'll a buy a set of those [OTC hearing aids].' And I said, 'Good luck. You're gonna waste your money.'* [person with dementia, male, age (years) in the range of 70s]

A subtheme emerged in which participants questioned the competence, training, and expertise of OTC hearing aid vendors and customer support staff. Across groups, participants overwhelmingly endorsed the need for readily available, knowledgeable, and trustworthy technical support to facilitate OTC hearing aid use in people with dementia. At the same time, participants expressed uncertainty that OTC hearing aids would include "any technical assistance that you can count on." [caregiver, daughter, age (years) in the range of 70s] Specifically, participants raised concerns that the people who sell and service OTC hearing aids may have insufficient training and experience in hearing loss and hearing aids—and, therefore, might not provide helpful, actionable recommendations. This sentiment was especially prominent among people with dementia. As one participant with dementia remarked:

> *Well, not being familiar with buying over the counter, the clerk would not know how to explain properly how to do it versus a doctor at the clinic.* [person with dementia, male, age (years) in the range of 80s]

Along the same lines, another participant with dementia said:

> *I don't know if it [OTC hearing aids] would be easier. 'Cause sometimes they don't know exactly what you need or to tell you how to use it.* [person with dementia, male, age (years) in the range of 70s]

Participants with dementia, in particular, expressed doubt about the ethics and morality of the people who sell and service OTC hearing aids. They described feeling skeptical—"iffy" [person with dementia, male, age (years) in the range of 70s]—about

XSL•FO
RenderX

the veracity of the "smiley ads" [person with dementia, male, age (years) in the range of 70s] for OTC hearing aids and questioned their underlying motives. Many participants with dementia felt that OTC hearing aids only emphasized sales and profits at the expense of older adults living on fixed incomes. As one participant with dementia explained, "you could be fooled by somebody that's just trying to make some money." [person with dementia, female, age (years) in the range of 80s] Another participant with dementia asked:

> *Who's on the other side of the counter? … I wouldn't even consider that. Not at all. To me that would be like a crapshoot.* [person with dementia, male, age (years) in the range of 80s]

Barrier Theme #2: Assessing OTC Hearing Aid Candidacy in People With Dementia

An important barrier theme emerged around assessing OTC hearing aid candidacy in people with dementia. According to FDA regulations, OTC hearing aids are intended for adults with perceived mild-to-moderate hearing loss. However, many care professionals and family caregivers questioned the accuracy and reliability of perceived hearing loss in people with dementia. Both participant groups overwhelmingly agreed that people with dementia cannot accurately self-classify their degree of hearing loss. Specifically, a subtheme emerged surrounding the memory demands of self-assessing one's hearing loss. As several family caregivers and care professionals noted, age-related hearing loss gradually "creeps up" [care professional, female, age (years) in the range of 50s] on older adults, thus requiring them to recall and reflect on their hearing experiences over both long and short periods of time. Many people with dementia, participants explained, do not have the memory and cognitive capacity to remember and compare their listening experiences. As a result, family caregivers and care professionals were doubtful that people with dementia could accurately describe their hearing loss progression and categorize its severity relative to others' hearing ability. One care professional explained:

> *I think people with dementia could identify profound hearing loss—that they can't hear. But anything less, you add the complication of dementia. What are they not remembering? To say mild or moderate versus getting moderate to severe, they wouldn't know.* [care professional, female, age (years) in the range of 40s]

During semistructured interviews, many participants with dementia struggled to answer questions about determining their own candidacy for OTC hearing aids. However, those who could respond shared a similar opinion. For example, when asked if he was confident in self-classifying his degree of hearing loss, one participant with dementia responded:

> *Probably not because it comes on so gradually. And after you've been living with the condition, you sort of just accept that's the way your hearing is … It gets integrated into your life.* [person with dementia, male, age (years) in the range of 80s]

Family caregivers and care professionals agreed that in many cases, family members would be responsible for determining OTC hearing aid candidacy for people with dementia. However, family caregivers lacked confidence that they could accurately classify their care recipient's degree of hearing loss without a professional hearing examination. Here, a notable subtheme emerged regarding the challenge of separating the effects of cognitive decline/dementia and hearing loss on communication in people with dementia. Many family caregivers shared that they were unsure if they were observing symptoms of cognitive decline or hearing loss. As one family caregiver reflected:

> *I don't know if she's [care recipient] not holding that conversation because she doesn't have the capacity and the wherewithal to hold that conversation, or if she's really not hearing the other person speak.* [caregiver, wife, age (years) in the range of 70s]

Given the challenges described above, participants in all 3 groups felt it was important that people with dementia have a professional hearing examination regardless of their or their family's interest in OTC hearing aids. A care professional commented:

> *Even though it's over the counter, I still feel like there would be a need for… some type of evaluation and referral from a specialist. Not just that the caregiver says, 'Well, I've been talking recently, and they've not been responding back.'* [care professional, female age (years) in the range of 30s]

Similarly, one participant with dementia explained:

> *People try to guess their hearing loss. But you have to use technology to figure it out. You have to sit in that box. And they do the little noises back and forth, and then when you're done each ear is tested for its own hearing.* [person with dementia, male, age (years) in the range of 70s]

Barrier Theme #3: Caregiver-Facilitated Programming and Adjustment

Participants in all 3 groups overwhelmingly expressed uncertainty and trepidation about the process of programming and adjusting OTC hearing aids for people with dementia. First and foremost, there was strong agreement that by and large, people with dementia cannot independently self-program and customize OTC hearing aids—regardless of the specific self-programming or fitting process used. Indeed, many family caregivers and care professionals pointed out that people with dementia—even in the early stages of the disease—often need help using everyday technologies such as smartphones, televisions, computers, and tablets. Similarly, participants anticipated that family caregivers would play a significant role in programming and adjusting OTC hearing aids for people with dementia.

However, participants questioned whether family caregivers possess sufficient knowledge, skills, and experiences to program OTC hearing aids for care recipients with dementia. Some participants highlighted that family caregivers are not "educated in hearing" [person with dementia, female, age (years) in the range of 80s], while others stressed that many dementia caregivers are themselves older adults who may face barriers to learning and using new technologies. As one family caregiver reflected:

XSL•FO
RenderX

> *When you think about it, a lot of the people that need hearing aids are old, right? And most people my age do not understand technology very well and they wouldn't be able to program hearing aids for their spouse.* [caregiver, wife, age (years) in the range of 70s]

Most participants expressed a preference for and confidence in professional hearing aid programming and adjustments. Across groups, participants emphasized the complexity of human hearing and hearing aids, along with the education and training of licensed hearing care professionals— "there's a lot to hearing," [caregiver, daughter, age (years) in the range of 60s] one family caregiver commented. Participants especially valued professional expertise in programming and adjusting hearing aids based on an individual's audiometric thresholds, behavioral responses, and preferences. Specifically, several family caregivers and participants with dementia viewed hearing care professionals as uniquely attuned to "the correct questions to ask" [person with dementia, female, age (years) in the range of 80s], a person with hearing loss—including those with dementia—when tailoring and optimizing an individual's hearing aid settings. As one family caregiver explained:

> *I'd prefer [care recipient's hearing aids] be programmed by a professional who asks the questions and compares, 'How is it now? How is it not now?' Whether that same quality of observation could happen with me doing it—I don't know.* [caregiver, husband, age (years) in the range of 70s]

Similarly, one participant with dementia shared:

> *I'd rather go to someone that [programmed hearing aids] all the time…. Because they had probably done a lot of work with people with hearing aids, so they would know a lot of questions to ask about how I was hearing.* [person with dementia, female, age (years) in the range of 80s]

In particular, family caregivers were unsure if they would ask the right questions of their care recipient when programming OTC hearing aids. This concern was magnified by a subtheme regarding the challenge of obtaining reliable self-reports of sensory symptoms from people with dementia. Both family caregivers and direct care professionals emphasized that people with dementia lose the "self-reflective ability" [caregiver, wife, age (years) in the range of 70s] and "correct words" [care professional, female, age (years) in the range of 40s] to clearly identify and describe bodily and sensory sensations. Consequently, participants expressed concern that family caregivers would not be able to elicit and properly interpret feedback from their care recipient to facilitate effective OTC hearing aid programming. As a result, caregivers anticipated that they "[would] not trust" [caregiver, wife, age (years) in the range of 60s] and "would not feel confident" [caregiver, daughter, age (years) in the range of 60s] in their "amateur" [caregiver, husband, age (years) in the range of 70s] programming results. Several participants with dementia echoed this sentiment, one of whom explained:

> *She's [family caregiver] good with a lot of things, but I don't know that she's an expert in that [OTC hearing aid programming]. She is not educated in hearing, and I believe that you have to be educated with the body in order to work with improving hearing.* [person with dementia, female, age (years) in the range of 80s]

One family caregiver summarized her concerns about using an OTC hearing aid for her care recipient:

> *I think [programming is] the major disadvantage, not knowing if you've programmed it for the best quality of hearing, especially for somebody else* [caregiver, daughter, age (years) in the range of 60s]

Barrier Theme #4: Assessing the Effectiveness of OTC Hearing Aids

Following OTC hearing aid programming, participants raised concerns about their ability to assess the effectiveness of OTC hearing aids in people with dementia. Across groups, most participants felt they could determine if OTC hearing aids were helpful to some degree—but they questioned their ability to judge whether the devices had been "optimized" [caregiver, wife, age (years) in the range of 60s] or had achieved a person with dementia's full hearing potential. Here, participants sought assurance that OTC hearing aids had delivered "the very best benefit" [caregiver, wife, age (years) in the range of 70s] for speech understanding and communication. However, many participants felt such assurance could only be obtained by seeing a hearing care professional for a prescription hearing aid fitting. Participants in all 3 groups expressed trust and confidence in a hearing care professional's ability to ensure hearing aids are "correct… and fit beautifully enough for [people with dementia] to hear." [care professional, female, age (years) in the range of 30s] As one participant with dementia explained when discussing hearing aid programming:

> *Since I'm a perfectionist, I wanna talk to somebody who is in the [hearing aid] business.* [person with dementia, female, age (years) in the range of 80s]

Without the assistance of a professional, participants in all 3 groups were unsure how they would assess and quantify the effectiveness of OTC hearing aids. Specifically, family caregivers and direct care professionals noted two prominent subthemes described earlier: (1) the challenge of separating the effects of cognitive decline/dementia and hearing loss on communication in people with dementia, and (2) the challenge of obtaining a reliable self-report of sensory symptoms from people with dementia. Together, these 2 challenges made family caregivers and direct care professionals apprehensive about evaluating the amount and extent of OTC hearing aid benefit in people with dementia—"it's tricky because you're also observing cognition," [caregiver, wife, age (years) in the range of 70s] one family caregiver noted. Another family caregiver reflected:

> *When you have cognitive decline, all parts of the brain are dying literally and not functioning as well. It takes [care recipient] longer to process things that are said. And I might have less confidence in [care recipient's] ability to describe it [hearing aid performance].* [caregiver, wife, age (years) in the range of 70s]

XSL•FO
RenderX

Most participants emphasized the importance of programming OTC hearing aids for optimal speech understanding and communication; however, several family caregivers and direct care professionals raised concerns about the possibility of sensory agitation from poorly fit OTC hearing aids. Responses in this subtheme emphasized the sensitivity of people with dementia to changes in their environment and sensory inputs. Namely, participants wondered if suboptimal OTC hearing aids could produce uncomfortable sound quality or acoustic feedback that might make a person with dementia startled, agitated, frustrated, or withdrawn. Participants raised specific concerns about "loud sounds," [caregiver, daughter, age (years) in the range of 40s] "echoing," [care professional, female, age (years) in the range of 70s] "ringing," [care professional, female, age (years) in the range of 20s] and "buzzing" [care professional, female, age (years) in the range of 50s] as likely sources of sensory agitation from hearing aids that might be overlooked or poorly managed in OTC hearing aids. Notably, participants lacked confidence in both people with dementia and their family caregivers to accurately identify, assess, and resolve these issues. One care professional explained:

> *If they [people with dementia] can't tell you that the [hearing aid] volume is too loud, it could be echoing and stuff. You never would know that because they have dementia, and they just won't talk at all because the sound in their ear is not correct… like a buzz or a ringing.* [care professional, female, age (years) in the range of 70s]

Adding further challenge, caregivers and care professionals felt they might have limited time to identify and resolve OTC hearing aid issues before their care recipient might become frustrated or uncooperative. Here, a subtheme emerged for the tendency of people with dementia to "give up early" [care professional, female, age (years) in the range of 40s] on new technologies. Several participants stressed that if the effectiveness and benefit of OTC hearing aids were not quickly apparent, a person with dementia might abandon the devices altogether. One family caregiver explained:

> *My husband [care recipient] would expect me to be the expert on it [OTC hearing aids], and then if it didn't work, he'd get cranky at me, and he might give up.* [caregiver, wife, age (years) in the range of 60s]

A participant with dementia shared his own expectations for immediate OTC hearing aid benefit:

> *If it's [OTC hearing aids] wrong, I wouldn't have any idea what to do. It's so simple…. You put it [hearing aid] in your ear and press the button. If it doesn't work… that isn't right.* [person with dementia, male, age (years) in the range of 80s]

Barrier Theme #5: Ongoing OTC Hearing Aid Use and Caregiver Burden/Burnout

Family caregivers and direct care professionals agreed that the challenges of OTC hearing aid use in people with dementia extend beyond the initial purchase and programming. Indeed, both groups stressed that OTC hearing aid use is an ongoing commitment involving daily responsibilities for hearing aid use, care, maintenance, and troubleshooting. Across groups, participants had differing views on the amount and type of hearing aid tasks a person with dementia could complete independently, reflecting individual differences between care recipients. Nevertheless, participants concurred that family caregivers would need to assume ultimate responsibility for overseeing and managing continued OTC hearing aid use for a person with dementia—especially as their cognitive and functional abilities further declined. This responsibility emerged as a significant concern. Nearly all family caregivers emphasized their multiple competing caregiving responsibilities, with many describing themselves as overwhelmed, overloaded, and burned out. For this reason, many family caregivers questioned their patience and capacity for self-learning and implementing OTC hearing aid tasks without the assistance of a hearing care professional. "The caregiver needs to be supported and educated to be successful… that's pretty important in dementia care," one family caregiver explained [caregiver, daughter, age (years) in the range of 60s]. Another family caregiver shared her initial reaction to learning about OTC hearing aids on television:

> *What I thought about [OTC hearing aids] is, I would find it challenging to figure out. I have to say as a caregiver and being kind of burned out that I'm not as confident doing things as I used to be. I mean, I'm aging, too, but I think it has more to do with being overwhelmed and burned out. It's harder for me to learn things right now that I would normally be able to manage.* [caregiver, daughter, age (years) in the range of 60s]

Another family caregiver echoed this perspective:

> *I can figure out technology without any problem…but I get very frustrated very easily when I have to do anything on my own. It's not that I can't do it, it's just that you have so much on your plate that if somebody puts one more thing on your plate, it puts you on overload.* [caregiver, wife, age (years) in the range of 70s]

Family caregivers were unsure if the challenges of long-term OTC hearing aid use would be worth their effort. Here, family caregivers returned to the subtheme of value discussed earlier; however, they augmented their previous calculation of price versus expected hearing aid benefit with a new calculation of their caregiving work (cost) versus observed hearing aid benefit. In this context, family caregivers anticipated that their work would be substantial while OTC hearing aid effectiveness would be uncertain or unclear (returning to an earlier barrier theme—see Barrier #4). Furthermore, family caregivers questioned whether their care recipient would be cooperative, compliant, or equally invested in using OTC hearing aids. As one family caregiver described:

> *It's [OTC hearing aid use] one more thing to do. It's questionable benefit…. And unclear how much cooperation you're gonna get from your partner. And you learn where to fight your battles.* [caregiver, husband, age (years) in the range of 70s]

Another family caregiver elaborated:

XSL•FO
RenderX

> *It's always gonna be how much work is it versus the benefit. That's always gonna be, as a caregiver, the scale you're in.* [caregiver, wife, age (years) in the range of 60s]

Finally, when considering whether she would use OTC hearing aids with her spouse, who has dementia, one family caregiver shared her competing priorities as a dementia caregiver: "I need to make my husband's [care recipient] life easier, but I also need to make our lives easier." [caregiver, wife, age (years) in the range of 60s].

Discussion

Principal Findings

This qualitative study explored the feasibility and acceptability of now widely available OTC hearing aids for community-dwelling older people with dementia and their family caregivers. We approached this question from the perspectives of key dementia-care interest-holders via semistructured interviews with individuals from three interest-holder groups: (1) community-dwelling older adults with co-occurring dementia and hearing difficulty, (2) family caregivers of community-dwelling older adults with co-occurring dementia and hearing difficulty, and (3) geriatric direct care professionals. Using thematic analysis, we identified and described specific facilitators and barriers to OTC hearing aid use in community-dwelling older people with dementia. Below, we discuss our findings and their clinical implications.

Facilitators of OTC Hearing Aid Use in People With Dementia

Facilitators or advantages of OTC hearing aid use included accessibility, affordability, and autonomy/control. While these benefits extend to older adults in general, our participants highlighted their specific importance within the dementia caregiving context. In terms of accessibility, participants endorsed the benefit of OTC hearing aids in eliminating the recurrent in-person appointments typically required for prescription hearing aid fitting and adjustments. Participants felt that removing the need for in-person appointments could make hearing health care more feasible and palatable for dementia caregivers and their care recipients. Additionally, participants appreciated the affordability of OTC hearing aids. However, family caregivers tended to emphasize value considerations rather than cost. Here, family caregivers viewed OTC hearing aids as a potentially better value than prescription hearing aids when considering their care recipients' simple listening environments, low technology proficiency, and high proclivity for losing/damaging hearing aids and other technologies. Finally, participants also identified the benefits of OTC hearing aids in the realm of autonomy and control. Across interest-holder groups, participants appreciated the autonomy and flexibility OTC hearing aids offer for deciding when, where, and how to try hearing aids for people with dementia. Participants viewed this enhanced control as particularly important within the context of stigma, noting that the combined stigmas of dementia and hearing loss may prevent individuals with dementia from seeking or agreeing to formal hearing testing or prescription hearing aids. Participants in all 3 groups agreed that OTC hearing aids could offer a more discreet alternative by allowing people with dementia to circumvent formal hearing testing and documentation of prescription hearing aid use.

These findings generally align with the larger hearing aid and dementia caregiving literature. Regarding accessibility, the benefits of OTC hearing aids are evident. Studies show family dementia caregivers bear heavy responsibility for transporting their care recipient to health care appointments, which can contribute to caregiving-related emotional, psychological, financial, and physical burden [54,55]. Furthermore, health care appointments may—as our participants discussed—cause emotional and physical distress for people with dementia, particularly when health care providers are not specifically trained in dementia [56]. Given these considerations, our participants' desire to avoid recurrent in-person hearing aid appointments is understandable and may constitute a significant advantage for OTC hearing aids in the dementia caregiving context.

Our participants' preference for lower-cost, more affordable hearing aids is unsurprising and consistent with the underlying rationale for OTC hearing aids. However, our family caregiver participants made an important distinction between value and cost. To date, most studies on barriers to hearing aid adoption emphasize cost and affordability. Research, however, has shown that value perceptions can also drive hearing aid purchase decisions [57,58]. For older adults broadly, perceived hearing aid value may include emotional, social, quality, and price considerations [58]. Our results suggest that these factors may operate differently in the dementia caregiving context, where family caregivers consider their care recipient's functional limitations when forming expectations of hearing aid benefits relative to price. By lowering the cost of amplification, OTC hearing aids may play a unique role in improving perceived hearing aid value for people with dementia and their family caregivers.

Our participants' emphasis on control and patient autonomy is consistent with the priorities of early policy and critique documents that led to the formal introduction of OTC hearing aids, though somewhat less consistent with the OTC hearing aid regulations themselves [59]. Autonomy is seldom discussed in the context of OTC hearing aids, possibly because it is viewed as an aspect of accessibility. Indeed, accessibility does contribute to the ease of obtaining OTC hearing aids. However, our family caregivers and care professionals viewed control as having significance beyond more easily obtaining hearing aids. Specifically, our participants emphasized the importance of autonomy for overcoming stigma-related barriers to hearing aid adoption in people with dementia. Among older adults, dementia and hearing loss each carry stigma—but together, their impact may be even more harmful. For older adults living with both dementia and hearing loss, the 2 stigmas may combine to produce unique fears around loss of independence, aging, and social isolation [60-65]. Although the adoption of OTC hearing aids still requires some acknowledgment of hearing loss, these devices may help overcome stigma by giving older people with dementia and their caregivers the autonomy to circumvent

XSL•FO
RenderX

formal medical documentation of hearing loss and hearing aid use.

Barriers to OTC Hearing Aid Use in People With Dementia

Although participants endorsed several facilitators of OTC hearing aid use in community-dwelling older people with dementia, they also described meaningful barriers that tempered their enthusiasm. From our participants' responses, we identified the following barriers or disadvantages of OTC hearing aid use: mistrust of OTC hearing aids; assessing OTC hearing aid candidacy in people with dementia; caregiver-facilitated programming/adjustment; assessing the effectiveness of OTC hearing aids; and ongoing OTC hearing aid use and caregiver burden/burnout.

Our findings broadly support and add nuance to the emerging literature on consumer attitudes toward and barriers to OTC hearing aid use. In our study, mistrust emerged as a substantial barrier to OTC hearing aid use for people with dementia and their family caregivers. A recent study of consumer attitudes toward OTC hearing aids also describes trust barriers to OTC hearing aid use. In a survey of 1377 adults without prior hearing aid experience, Singh and Dhar [66] found overall low interest in OTC hearing aids, which they largely attributed to a lack of knowledge and trust in OTC hearing aids. In our study, some participants' mistrust stemmed from misinformation or lack of knowledge, such as the belief that OTC hearing aids are simple, one-size-fits-all amplifiers. However, participants also conveyed deeper mistrust regarding the competence, motivations, and ethics of those who sell and service OTC hearing aids. Here, it is plausible that people with dementia and their family caregivers may have heightened mistrust of OTC hearing aids compared with the general population. Research indicates that individuals with dementia are particularly vulnerable to financial fraud and exploitation [67,68], potentially leading those with earlier-stage disease and higher self-awareness to approach new products and sales claims with increased caution. Additionally, family dementia caregivers have been shown to harbor mistrust for a variety of dementia care services and interventions, stemming from previous negative experiences and a strong desire to protect their care recipients [69,70].

As discussed previously, both family caregivers and care professionals raised serious concerns about the accuracy and reliability of self-perceived hearing loss in people with dementia. This concern is supported by numerous studies in the general older adult population demonstrating that even without known cognitive impairment, self-reported hearing ability is generally unreliable compared with measured audiometric thresholds [71-73]. However, as our participants highlighted, this problem may be even more pronounced in people with dementia. In a recent study, Kim et al [74] found lower concordance between self-reported and audiometric hearing status in older adults with cognitive impairment than in those without. Notably, the authors also found poor sensitivity of proxy-reported hearing ability in older adults with cognitive impairment, suggesting that family or friends' hearing loss judgments are an unreliable substitute for self-report in this population [74]. Our qualitative results echo these findings, indicating that family caregivers not only lack confidence in their care recipient's self-reported hearing status—but also in their own ability to assess and categorize their care recipients' hearing and communication abilities. Interestingly, our family caregiver participants attributed this lack of confidence, in large part, to the challenge of distinguishing the effects of hearing loss and dementia on communication—an observation echoed in other published studies [31,75]. This insight could help explain the finding of Kim et al [74] of poor sensitivity for proxy-reported hearing loss in their sample of older people with dementia.

Our participants also expressed concerns about the process of programming and adjusting OTC hearing aids. Participants in all 3 groups agreed that people with dementia cannot independently program and adjust OTC hearing aids. This sentiment is well supported by a host of studies documenting the myriad challenges and barriers faced by people with dementia when using both everyday and assistive technologies [35-37]. However, our family caregiver participants also lacked confidence in their own ability to assist their care recipients with OTC hearing aid programming.

While there are few studies of OTC hearing aid use in community-dwelling dementia caregiving dyads, Convery et al [76] found that in a sample of older adults with a range of cognitive abilities, lay partner participation in the OTC hearing aid self-fitting process did not significantly influence or improve outcomes. The authors suggest that fitting outcomes might have improved with knowledgeable support from trained personnel [76]. Here, it is important to consider—as some of our participants mentioned—that many partners and caregivers of older people with co-occurring dementia and hearing loss are themselves older adults who may encounter challenges in using technology. Studies show that older adults face a variety of barriers to technology use, including lack of knowledge, age-related vision loss, and fine motor difficulties, as well as negative attitudes and anxiety toward technology [77-79]. Our family caregiver participants, who were primarily older adults, exemplified these barriers, expressing great apprehension about their technology proficiency and hearing aid knowledge. Notably, in this context, our caregivers' concerns about technology use were compounded by the previously mentioned barrier surrounding self- and proxy-report of hearing ability in people with dementia. In short, our family caregivers lacked confidence in their care recipients' ability to self-report changes in hearing to guide caregivers' programming decisions, and caregivers were also unsure if they could intuit whether they made the correct programming decisions.

Across interest-holder groups, participants were unsure if they could accurately assess the effectiveness of OTC hearing aids in people with dementia. Specifically, our participants raised concerns that self- or caregiver-programmed settings might be suboptimal in ways that could go undetected—potentially leading to sensory agitation, missed benefits, and hearing aid abandonment in people with dementia. To our knowledge, no research has examined the effectiveness of self- or caregiver-fit OTC hearing aids in people with dementia. In the general older adult population, self-programmed OTC hearing aids have consistently shown comparable outcomes to prescription hearing aids [1-5]. However, outcomes of OTC hearing aids in dementia

XSL•FO
RenderX

caregiving dyads are yet unknown. This uncertainty may have contributed to our participants' strong preference for prescription hearing aids and professional guidance. However, this finding is broadly consistent with the results of recent consumer surveys, which show that most adults with self-reported hearing difficulty prefer having professional assistance with their health care [66,80]. Research is urgently needed to determine how OTC hearing aid outcomes compare to those of prescription hearing aids in community-dwelling people with dementia and their family caregivers.

Last, our family caregiver and care professional participants stressed the need for ongoing caregiver involvement to support long-term OTC hearing aid use in people with dementia. Here, our family caregiver participants raised serious concerns about adding OTC hearing aid responsibilities to their already significant caregiving load. Many described feeling overwhelmed and burned out, echoing a large and well-established literature on caregiver burden, burnout, and stress among informal dementia caregivers [81-84]. Furthermore, several of our caregiver participants questioned whether OTC hearing aids would be worth their caregiving effort, especially if the OTC hearing benefit was unapparent and their care recipient was apathetic or uncooperative about using hearing aids.

Considerations of caregiver burden/burnout augment the value theme described earlier to include caregiver perceptions of their work/effort for OTC hearing aid use versus the expected or observed OTC hearing aid benefit. Caregivers may use this calculation of work/effort versus benefit to decide whether to (1) initiate OTC hearing aid use for their care recipient, or (2) continue using OTC hearing aids with their care recipient. Research shows that although people with dementia prefer to be included in important decisions, caregivers often make important medical and treatment decisions for them [85]. In this case, caregivers may ultimately decide whether to invest time and energy into using OTC hearing aids with and for their care recipient. It is important to note that this situation is not unique to OTC hearing aids; it also applies to prescription hearing aids. Indeed, research shows that caregiver assistance is an important determinant of successful prescription hearing aid use in people with dementia [42]. However, the lack of professional support and services for OTC hearing aids could add additional stress on caregivers, potentially increasing the caregiver burden involved in hearing aid use. Research is needed that compares caregiver outcomes for prescription versus OTC hearing aids.

Clinical Implications and Future Directions

Overview

Findings from this study highlight several ways in which OTC hearing aid devices and service delivery models could be better aligned with the needs of people with dementia and their caregivers. Unless and until major cost and accessibility barriers to prescription hearing aids are removed, people with dementia may turn to OTC hearing aids to afford and access hearing health care. Thus, addressing the barriers identified in this study is essential.

Implications for Device Design

Participants emphasized perceived challenges related to device complexity, difficult-to-navigate controls, and unclear or insufficient user feedback during use. These findings point to actionable opportunities for OTC hearing aid manufacturers to improve usability for people with dementia and their caregivers. Design considerations informed by participant perspectives include the following:

- Simplified, intuitive programming and troubleshooting workflows that reduce the cognitive load required to set up and adjust devices.
- Dedicated caregiver-assisted programming pathways that allow caregivers to complete or oversee key steps in the setup and adjustment process.
- Caregiver-oriented tools or interfaces that help validate and verify whether amplification settings are appropriate and functioning as intended.
- Clear, interpretable system feedback that communicates both correct and incorrect steps in formats accessible to users with dementia and to caregivers providing assistance.
- Physical design features that minimize handling complexity, including clearly marked controls and components that are easy for both people with dementia and caregivers to understand and manipulate.

Implications for Service Delivery and Caregiver Support

Participants consistently emphasized that people with dementia are unlikely to use OTC hearing aids successfully without meaningful caregiver involvement. Perceived challenges included difficulties with initial setup, troubleshooting, and day-to-day management of the devices. These findings point to several opportunities for strengthening service delivery models and caregiver supports:

- Brief caregiver-focused training modules delivered in person, by phone, or through app-based tutorials to assist with device setup, programming, and troubleshooting.
- Simplified written, visual, or video-based instructions tailored to dementia caregiving dyads that reduce cognitive load and clarify step-by-step processes.
- Optional caregiver-tailored remote onboarding or coaching sessions to compensate for the absence of professional fitting services, especially during initial use.
- Caregiver-centered guidance and expectation-setting, including strategies for monitoring benefit, recognizing device malfunction, and supporting consistent use.

Together, these supports reflect opportunities for behavioral and educational interventions aimed at reducing caregiver burden and improving the likelihood of successful device use.

Future Research Directions

Further research is needed to evaluate the feasibility and acceptability of OTC hearing aids for people living with dementia and their caregivers. Priority areas include the following:

- Hands-on trials in both laboratory and home settings to observe device use, errors, and points of breakdown.

XSL•FO
RenderX

- Quantifying the magnitude and consequences of identified barriers, including which barriers are most consequential for real-world outcomes.
- Examining caregiver capabilities and roles, including how caregiver burden, health literacy, and technology experience shape outcomes.
- Developing and pilot-testing tailored interventions, device refinements, and service delivery supports informed by the themes identified in this study.

In addition, research is needed to better understand the effectiveness of OTC hearing aids for people living with dementia compared with traditional prescription hearing aids. Clinical trials should evaluate outcomes and benefit domains to clarify when, how, and for whom OTC hearing aids may provide appropriate amplification to support communicative function.

Limitations

A key strength of this study is its inclusion of multiple different perspectives, providing a more robust exploration of the question than any one interest-holder group alone. However, several important limitations should be noted.

One consideration is that most of our participants with dementia were current hearing aid users. This outcome resulted from our inclusion criterion requiring that participants with dementia self-report mild-to-moderate hearing difficulty. During recruitment, we found that people with dementia who already used hearing aids were more likely to self-identify as having hearing difficulty than those who did not use hearing aids. Additionally, our sample of participants with dementia was predominantly male. This imbalance may reflect the dynamics of our caregiver-mediated recruitment pathways, in which caregivers and community partners more frequently referred male care recipients. Gender differences in the prevalence of age-related hearing loss may also have contributed to this pattern [8].

We also relied on self- or caregiver-reported hearing loss and dementia status rather than confirmed medical diagnoses. This approach was chosen to ensure the feasibility and accessibility of this study for dementia caregivers and people with dementia, for whom additional testing or records requests would have been cumbersome and likely infeasible. In terms of hearing loss, our use of self-reported hearing ability aligns with OTC hearing aid candidacy, which does not require a formal hearing evaluation. Although audiometric results would have been informative for cross-checking self- and caregiver-reported hearing ability, all participants reported mild-to-moderate hearing difficulty consistent with OTC hearing aid candidacy guidelines. Regarding dementia status, our reliance on self- or caregiver reports introduces some uncertainty regarding participants' cognitive abilities and underlying conditions. However, the combination of self- or caregiver report with the TICS-m cognitive test increases confidence that our sample includes people with early to midstage dementia.

Finally, our study required that participants speak English, limiting the potential racial and ethnic diversity of our sample. Family caregivers and people with dementia who do not speak English may face distinct challenges in using OTC hearing aids, which warrants further investigation.

Given these limitations, caution is warranted when generalizing the findings to women, new hearing aid users, and individuals who do not speak English. Even so, the themes identified in this study point to several opportunities to improve the user experience for many potential OTC hearing aid users with dementia. Future research should more comprehensively characterize the needs of users from different backgrounds and with different levels of hearing aid experience.

Conclusions

This study identified and described interest-holder-perceived facilitators and barriers to OTC hearing aid use in community-dwelling people with dementia and their family caregivers. Key facilitators or advantages of OTC hearing aid use included improved accessibility and affordability, as well as enhanced autonomy and control. However, significant barriers underscore the complexity of OTC hearing aid use in this population. Interest-holder–perceived barriers included mistrust of OTC hearing aids; assessing OTC hearing aid candidacy in people with dementia; caregiver-facilitated programming/adjustment; assessing the effectiveness of OTC hearing aids; and ongoing OTC hearing aid use and caregiver burden/burnout. Future research should further explore and quantify these and other barriers toward the development of tailored devices, services, and supports that promote successful OTC hearing aid use in people with dementia and their family caregivers. Such work is essential for ensuring that people with dementia can equitably benefit from the introduction of lower-cost, more accessible OTC hearing aids.

Acknowledgments

We are deeply grateful to the participants who contributed their time, insights, and energy to this study. The authors did not use generative artificial intelligence to generate any part of this manuscript.

Funding

This work was supported by the University of Minnesota Graduate School's Interdisciplinary Doctoral Fellowship and Doctoral Dissertation Fellowship awarded to the corresponding author.

XSL•FO
RenderX

Data Availability

The qualitative data generated and analyzed during this study are not publicly available in order to protect participant confidentiality, as the data cannot be sufficiently deidentified. Researchers interested in further information about the study procedures or analytic approach may contact the corresponding author.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Semistructured interview guides.
[DOCX File , 18 KB - humanfactors_v13i1e83857_app1.docx]

References

1. Humes LE, Rogers SE, Quigley TM, Main AK, Kinney DL, Herring C. The effects of service-delivery model and purchase price on hearing-aid outcomes in older adults: a randomized double-blind placebo-controlled clinical trial. Am J Audiol 2017;26(1):53-79 [FREE Full text] [doi: 10.1044/2017_AJA-16-0111] [Medline: 28252160]
2. Nelson PB, Perry TT, Gregan M, VanTasell D. Self-adjusted amplification parameters produce large between-subject variability and preserve speech intelligibility. Trends Hear 2018;22:2331216518798264 [FREE Full text] [doi: 10.1177/2331216518798264] [Medline: 30191767]
3. Urbanski D, Hernandez H, Oleson J, Wu Y. Toward a new evidence-based fitting paradigm for over-the-counter hearing aids. Am J Audiol 2021;30(1):43-66 [FREE Full text] [doi: 10.1044/2020_AJA-20-00085] [Medline: 33264578]
4. De Sousa KC, Manchaiah V, Moore DR, Graham MA, Swanepoel DW. Effectiveness of an over-the-counter self-fitting hearing aid compared with an audiologist-fitted hearing aid: a randomized clinical trial. JAMA Otolaryngol Head Neck Surg 2023;149(6):522-530 [FREE Full text] [doi: 10.1001/jamaoto.2023.0376] [Medline: 37052929]
5. Venkitakrishnan S, Urbanski D, Wu Y. Efficacy and effectiveness of evidence-based non-self-fitting presets compared to prescription hearing aid fittings and a personal sound amplification product. Am J Audiol 2023;33(1):1-24. [doi: 10.1044/2023_AJA-23-00121] [Medline: 37956699]
6. Akl EA, Khabsa J, Petkovic J, Magwood O, Lytvyn L, Motilall A, et al. "Interest-holders": a new term to replace "stakeholders" in the context of health research and policy. Cochrane Evid Synth Methods 2024;2(11):e70007. [doi: 10.1002/cesm.70007] [Medline: 40475280]
7. Sharma RK, Lalwani AK, Golub JS. Prevalence and severity of hearing loss in the older old population. JAMA Otolaryngol Head Neck Surg 2020;146(8):762-763 [FREE Full text] [doi: 10.1001/jamaoto.2020.0900] [Medline: 32584401]
8. Reed NS, Garcia-Morales EE, Myers C, Huang AR, Ehrlich JR, Killeen OJ, et al. Prevalence of hearing loss and hearing aid use among US medicare beneficiaries aged 71 years and older. JAMA Netw Open 2023;6(7):e2326320 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.26320] [Medline: 37505496]
9. Thai A, Khan SI, Choi J, Ma Y, Megwalu UC. Associations of hearing loss severity and hearing aid use with hospitalization among older US adults. JAMA Otolaryngol Head Neck Surg 2022;148(11):1005-1012 [FREE Full text] [doi: 10.1001/jamaoto.2022.2399] [Medline: 36048464]
10. Bridges JA, Bentler RA. Relating hearing aid use to well-being among older adults. Hear J 1998;51(7):39. [doi: 10.1097/00025572-199807000-00002]
11. Dillard LK, Pinto A, Mueller KD, Schubert CR, Paulsen AJ, Merten N, et al. Associations of hearing loss and hearing aid use with cognition, health-related quality of life, and depressive symptoms. J Aging Health 2023;35(7-8):455-465. [doi: 10.1177/08982643221138162] [Medline: 36412130]
12. Lin FR, Pike JR, Albert MS, Arnold M, Burgard S, Chisolm T, ACHIEVE Collaborative Research Group. Hearing intervention versus health education control to reduce cognitive decline in older adults with hearing loss in the USA (ACHIEVE): a multicentre, randomised controlled trial. Lancet 2023;402(10404):786-797. [doi: 10.1016/S0140-6736(23)01406-X] [Medline: 37478886]
13. Mulrow CD, Tuley MR, Aguilar C. Sustained benefits of hearing aids. J Speech Hear Res 1992;35(6):1402-1405. [doi: 10.1044/jshr.3506.1402] [Medline: 1494282]
14. Jorgensen L, Novak M. Factors influencing hearing aid adoption. Semin Hear 2020;41(1):6-20 [FREE Full text] [doi: 10.1055/s-0040-1701242] [Medline: 32047345]
15. Picou E. Hearing aid benefit and satisfaction results from the MarkeTrak 2022 survey. Semin Hear 2022;survey:301-316. [doi: 10.1055/s-0042-1758375]
16. Ferguson MA, Woolley A, Munro KJ. The impact of self-efficacy, expectations, and readiness on hearing aid outcomes. Int J Audiol 2016 Jul;55 Suppl 3:S34-S41 [FREE Full text] [doi: 10.1080/14992027.2016.1177214] [Medline: 27266542]
17. Jorbonyan A, Momtaz YA, Foroghan M, Mehrkian S. Factors affecting the adoption and use of hearing aids among older adults: a systematic review. Curr Psychiatry Res Rev 2021;17(3):172-187. [doi: 10.2174/2666082217666211109113920]

XSL•FO
RenderX

18. Kochkin S. MarkeTrak VII: Obstacles to adult non-user adoption of hearing aids. Hear J 2007;60(4):24-51. [doi: 10.1097/01.hj.0000285745.08599.7f]
19. Ng JH, Loke AY. Determinants of hearing-aid adoption and use among the elderly: a systematic review. Int J Audiol 2015;54(5):291-300. [doi: 10.3109/14992027.2014.966922] [Medline: 25640403]
20. Coco L, Titlow KS, Marrone N. Geographic distribution of the hearing aid dispensing workforce: a teleaudiology planning assessment for Arizona. Am J Audiol 2018;27(3S):462-473 [FREE Full text] [doi: 10.1044/2018_AJA-IMIA3-18-0012] [Medline: 30452750]
21. Jilla AM, Johnson CE, Huntington-Klein N. Hearing aid affordability in the United States. Disabil Rehabil Assist Technol 2023;18(3):246-252. [doi: 10.1080/17483107.2020.1822449] [Medline: 33112178]
22. Nassiri A, Ricketts T, Carlson M. Current estimate of hearing aid utilization in the United States. Otol Neurotol Open 2021;1(1):e001 [FREE Full text] [doi: 10.1097/ONO.0000000000000001] [Medline: 38549728]
23. Planey AM. Audiologist availability and supply in the United States: a multi-scale spatial and political economic analysis. Soc Sci Med 2019;222:216-224. [doi: 10.1016/j.socscimed.2019.01.015] [Medline: 30660682]
24. Langa KM, Larson EB, Crimmins EM, Faul JD, Levine DA, Kabeto MU, et al. A comparison of the prevalence of dementia in the United States in 2000 and 2012. JAMA Intern Med 2017;177(1):51-58 [FREE Full text] [doi: 10.1001/jamainternmed.2016.6807] [Medline: 27893041]
25. Matthews KA, Xu W, Gaglioti AH, Holt JB, Croft JB, Mack D, et al. Racial and ethnic estimates of Alzheimer's disease and related dementias in the United States (2015-2060) in adults aged ≥65 years. Alzheimers Dement 2019;15(1):17-24 [FREE Full text] [doi: 10.1016/j.jalz.2018.06.3063] [Medline: 30243772]
26. Plassman B, Langa K, Fisher G, Heeringa S, Weir D, Ofstedal M, et al. Prevalence of dementia in the United States: the aging, demographics, and memory study. Neuroepidemiology 2007;29(1-2):125-132 [FREE Full text] [doi: 10.1159/000109998] [Medline: 17975326]
27. Nieman CL, Garcia Morales EE, Huang AR, Reed NS, Yasar S, Oh ES. Prevalence of hearing loss and hearing aid use among persons living with dementia in the US. JAMA Netw Open 2024;7(10):e2440400 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.40400] [Medline: 39432304]
28. Atef RZ, Michalowsky B, Raedke A, Platen M, Mohr W, Mühlichen F, et al. Impact of hearing aids on progression of cognitive decline, depression, and quality of life among people with cognitive impairment and dementia. J Alzheimers Dis 2023;92(2):629-638. [doi: 10.3233/JAD-220938] [Medline: 36776058]
29. Dawes P, Littlejohn J, Bott A, Brennan S, Burrow S, Hopper T, et al. Hearing assessment and rehabilitation for people living with dementia. Ear Hear 2022;43(4):1089-1102 [FREE Full text] [doi: 10.1097/AUD.0000000000001174] [Medline: 34966160]
30. Mamo S, Oh E, Lin F. Enhancing communication in adults with dementia and age-related hearing loss. Semin Hear 2017;38(2):177-183 [FREE Full text] [doi: 10.1055/s-0037-1601573] [Medline: 28522892]
31. Dawes P, Wolski L, Himmelsbach I, Regan J, Leroi I. Interventions for hearing and vision impairment to improve outcomes for people with dementia: a scoping review. Int Psychogeriatr 2019;31(2):203-221 [FREE Full text] [doi: 10.1017/S1041610218000728] [Medline: 30244688]
32. Mamo SK, Nirmalasari O, Nieman CL, McNabney MK, Simpson A, Oh ES, et al. Hearing care intervention for persons with dementia: a pilot study. Am J Geriatr Psychiatry 2017;25(1):91-101 [FREE Full text] [doi: 10.1016/j.jagp.2016.08.019] [Medline: 27890543]
33. Palmer CV, Adams SW, Bourgeois M, Durrant J, Rossi M. Reduction in caregiver-identified problem behaviors in patients with Alzheimer disease post-hearing-aid fitting. J Speech Lang Hear Res 1999;42(2):312-328. [doi: 10.1044/jslhr.4202.312] [Medline: 10229449]
34. Nirmalasari O, Mamo SK, Nieman CL, Simpson A, Zimmerman J, Nowrangi MA, et al. Age-related hearing loss in older adults with cognitive impairment. Int Psychogeriatr 2017;29(1):115-121 [FREE Full text] [doi: 10.1017/S1041610216001459] [Medline: 27655111]
35. Malinowsky C, Almkvist O, Kottorp A, Nygård L. Ability to manage everyday technology: a comparison of persons with dementia or mild cognitive impairment and older adults without cognitive impairment. Disabil Rehabil Assist Technol 2010;5(6):462-469. [doi: 10.3109/17483107.2010.496098] [Medline: 20545577]
36. Nygård L, Starkhammar S. The use of everyday technology by people with dementia living alone: mapping out the difficulties. Aging Ment Health 2007;11(2):144-155. [doi: 10.1080/13607860600844168] [Medline: 17453547]
37. Rosenberg L, Kottorp A, Winblad B, Nygård L. Perceived difficulty in everyday technology use among older adults with or without cognitive deficits. Scand J Occup Ther 2009;16(4):216-226. [doi: 10.3109/11038120802684299]
38. Schmidt L, Wahl H. Predictors of performance in everyday technology tasks in older adults with and without mild cognitive impairment. Gerontologist 2019;59(1):90-100. [doi: 10.1093/geront/gny062] [Medline: 29878116]
39. Albers EA, Mikal J, Millenbah A, Finlay J, Jutkowitz E, Mitchell L, et al. The use of technology among persons with memory concerns and their caregivers in the United States during the COVID-19 pandemic: qualitative study. JMIR Aging 2022;5(1):e31552 [FREE Full text] [doi: 10.2196/31552] [Medline: 35134748]

XSL•FO
RenderX

40. Dequanter S, Fobelets M, Steenhout I, Gagnon M, Bourbonnais A, Rahimi S, et al. Determinants of technology adoption and continued use among cognitively impaired older adults: a qualitative study. BMC Geriatr 2022;22(1):376 [FREE Full text] [doi: 10.1186/s12877-022-03048-w] [Medline: 35484488]
41. Altieri M, Garramone F, Santangelo G. Functional autonomy in dementia of the Alzheimer's type, mild cognitive impairment, and healthy aging: a meta-analysis. Neurol Sci 2021;42(5):1773-1783. [doi: 10.1007/s10072-021-05142-0] [Medline: 33738665]
42. Hooper E, Brown LJ, Cross H, Dawes P, Leroi I, Armitage CJ. Systematic review of factors associated with hearing aid use in people living in the community with dementia and age-related hearing loss. J Am Med Dir Assoc 2022;23(10):1669-1675.e16 [FREE Full text] [doi: 10.1016/j.jamda.2022.07.011] [Medline: 35988590]
43. Hooper E, Brown LJ, Cross H, Dawes P, Leroi I, Armitage CJ. Enablers and barriers to hearing aid use in people living with dementia. J Appl Gerontol 2024;43(7):978-989 [FREE Full text] [doi: 10.1177/07334648231225346] [Medline: 38235997]
44. Knopman DS, Roberts RO, Geda YE, Pankratz VS, Christianson TJH, Petersen RC, et al. Validation of the telephone interview for cognitive status-modified in subjects with normal cognition, mild cognitive impairment, or dementia. Neuroepidemiology 2010;34(1):34-42 [FREE Full text] [doi: 10.1159/000255464] [Medline: 19893327]
45. Welsh K, Breitner J, Magruder-Habib K. Detection of dementia in the elderly using telephone screening of cognitive status. Cogn Behav Neurol 1993;6(2):103-110 [FREE Full text]
46. Ventry IM, Weinstein BE. The hearing handicap inventory for the elderly: a new tool. Ear Hear 1982;3(3):128-134. [doi: 10.1097/00003446-198205000-00006] [Medline: 7095321]
47. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 1969;9(3):179-186. [Medline: 5349366]
48. Chen A, An E, Yan E, He D, Saripella A, Butris N, et al. Incidence of preoperative instrumental activities of daily living (IADL) dependence and adverse outcomes in older surgical patients: a systematic review and meta-analysis. J Clin Anesth 2023;89:111151. [doi: 10.1016/j.jclinane.2023.111151]
49. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
50. The Alzheimer's Association. 2023 Alzheimer's disease facts and figures. Alzheimer's Dement 2023;19(4):1598-1695. [doi: 10.1002/alz.13016]
51. Boxer R, Norman M, Abadir P, Beizer JL, Dierich M, Lau S, et al. When women rise, we all rise: American geriatrics society position statement on achieving gender equity in geriatrics. J Am Geriatr Soc 2019;67(12):2447-2454. [doi: 10.1111/jgs.16195]
52. Dunkin JJ, Anderson-Hanley C. Dementia caregiver burden. Neurology 1998;51(1_suppl_1):-1. [doi: 10.1212/wnl.51.1_suppl_1.s53]
53. Lin FR. Hearing loss prevalence in the United States. Arch Intern Med 2011;171(20):1851. [doi: 10.1001/archinternmed.2011.506]
54. Liddle J, Tan A, Liang P, Bennett S, Allen S, Lie DC, et al. “The biggest problem we’ve ever had to face”: how families manage driving cessation with people with dementia. Int Psychogeriatr 2016;28(1):109-122. [doi: 10.1017/s1041610215001441]
55. Taylor BD, Tripodes S. The effects of driving cessation on the elderly with dementia and their caregivers. Accid Anal Prev 2001;33(4):519-528. [doi: 10.1016/s0001-4575(00)00065-8]
56. Challen R, Low L, McEntee M. Dementia patient care in the diagnostic medical imaging department. Radiography 2018;24:S33-S42. [doi: 10.1016/j.radi.2018.05.012]
57. Amlani A, Taylor B, Weinberg T. Increasing hearing aid adoption rates through value-based advertising and price unbundling. Hear Rev 2011;18(13):10-17 [FREE Full text]
58. Amlani A. Influence of perceived value on hearing aid adoption and re-adoption intent. Hear Rev 2013;20(3):8-12 (forthcoming) [FREE Full text]
59. Menon KN, Hoon-Starr M, Shilton K, Hoover EC. Over-the-counter hearing aids challenge the core values of traditional audiology. J Speech Lang Hear Res 2024;67(2):657-667. [doi: 10.1044/2023_jslhr-23-00306]
60. da SJ, de AC, Lüders D. The self-stigma of hearing loss in adults and older adults: a systematic review. Ear Hear 2023;44(6). [doi: 10.1097/aud.0000000000001398]
61. Liu C, Mavrommatis MM, Govindan A, Cosetti MK. The stigma of hearing loss: a scoping review of the literature across age and gender. Otolaryngol Head Neck Surg 2025;172(6):1874-1881. [doi: 10.1002/ohn.1246]
62. Milne A. The ‘D’ word: reflections on the relationship between stigma, discrimination and dementia. J Ment Health 2010;19(3):227-233. [doi: 10.3109/09638231003728166]
63. Rapaport P, Burton A, Leverton M, Herat-Gunaratne R, Beresford-Dent J, Lord K, et al. “I just keep thinking that I don’t want to rely on people.” a qualitative study of how people living with dementia achieve and maintain independence at home: stakeholder perspectives. BMC Geriatr 2020;20(1). [doi: 10.1186/s12877-019-1406-6]
64. Swaffer K. Dementia: stigma, language, and dementia-friendly. Dementia 2014;13(6):709-716. [doi: 10.1177/1471301214548143]

XSL•FO
RenderX

65. Wallhagen MI. The stigma of hearing loss. Gerontologist 2010;50(1):66-75 [FREE Full text] [doi: 10.1093/geront/gnp107] [Medline: 19592638]
66. Singh J, Dhar S. Assessment of consumer attitudes following recent changes in the US hearing health care market. JAMA Otolaryngol Head Neck Surg 2023;149(3):247-252 [FREE Full text] [doi: 10.1001/jamaoto.2022.4344] [Medline: 36656571]
67. Ebner N, Pehlivanoglu D, Shoenfelt A. Financial fraud and deception in aging. Adv Geriatr Med Res. 2023 2023;5(3):e230007. [doi: 10.31219/osf.io/ay4jw]
68. Fenge L, Lee S. Understanding the risks of financial scams as part of elder abuse prevention. Br J Soc Work 2018;48(4):906-923. [doi: 10.1093/bjsw/bcy037]
69. Macleod A, Tatangelo G, McCabe M, You E. "There isn't an easy way of finding the help that's available." barriers and facilitators of service use among dementia family caregivers: a qualitative study. Int Psychogeriatr 2017;29(5):765-776. [doi: 10.1017/S1041610216002532] [Medline: 28351450]
70. Teel CS, Carson P. Family experiences in the journey through dementia diagnosis and care. J Fam Nurs 2003;9(1):38-58. [doi: 10.1177/1074840702239490]
71. Kamil RJ, Genther DJ, Lin FR. Factors associated with the accuracy of subjective assessments of hearing impairment. Ear Hear 2015;36(1):164-167 [FREE Full text] [doi: 10.1097/AUD.0000000000000075] [Medline: 25158982]
72. Kiely KM, Gopinath B, Mitchell P, Browning CJ, Anstey KJ. Evaluating a dichotomized measure of self-reported hearing loss against gold standard audiometry: prevalence estimates and age bias in a pooled national data set. J Aging Health 2012;24(3):439-458. [doi: 10.1177/0898264311425088] [Medline: 22205434]
73. Kim SY, Kim H, Kim M, Park B, Kim J, Choi HG. Discrepancy between self-assessed hearing status and measured audiometric evaluation. PLoS One 2017;12(8):e0182718 [FREE Full text] [doi: 10.1371/journal.pone.0182718] [Medline: 28792529]
74. Kim AS, Betz JF, Albert M, Deal JA, Faucette SP, Oh ES, et al. Accuracy of self- and proxy-rated hearing among older adults with and without cognitive impairment. J Am Geriatr Soc 2022;70(2):490-500 [FREE Full text] [doi: 10.1111/jgs.17558] [Medline: 34813080]
75. Slaughter SE, Hopper T, Ickert C, Erin DF. Identification of hearing loss among residents with dementia: perceptions of health care aides. Geriatr Nurs 2014;35(6):434-440. [doi: 10.1016/j.gerinurse.2014.07.001] [Medline: 25212262]
76. Convery E, Keidser G, Hickson L, Meyer C. Factors associated with successful setup of a self-fitting hearing aid and the need for personalized support. Ear Hear 2019;40(4):794-804. [doi: 10.1097/AUD.0000000000000663] [Medline: 30285978]
77. Chen K, Lou V. Measuring senior technology acceptance: development of a brief, 14-Item scale. Innov Aging 2020;4(3):igaa016 [FREE Full text] [doi: 10.1093/geroni/igaa016] [Medline: 32617418]
78. Chen K, Chan AHS. Gerontechnology acceptance by elderly Hong Kong Chinese: a senior technology acceptance model (STAM). Ergonomics 2014;57(5):635-652. [doi: 10.1080/00140139.2014.895855]
79. Gitlow L. Technology use by older adults and barriers to using technology. Phys Occup Ther Geriatr 2014;32(3):271-280. [doi: 10.3109/02703181.2014.946640]
80. Edwards B. Emerging technologies, market segments, and marketrak 10 insights in hearing health technology. Semin Hear 2020;41(1):37-54 [FREE Full text] [doi: 10.1055/s-0040-1701244] [Medline: 32047347]
81. Connors MH, Seeher K, Teixeira-Pinto A, Woodward M, Ames D, Brodaty H. Dementia and caregiver burden: a three-year longitudinal study. Int J Geriatr Psychiatry 2020;35(2):250-258. [doi: 10.1002/gps.5244] [Medline: 31821606]
82. Etters L, Goodall D, Harrison BE. Caregiver burden among dementia patient caregivers: a review of the literature. J Am Acad Nurse Pract 2008;20(8):423-428. [doi: 10.1111/j.1745-7599.2008.00342.x] [Medline: 18786017]
83. Chiao C, Wu H, Hsiao C. Caregiver burden for informal caregivers of patients with dementia: a systematic review. Int Nurs Rev 2015;62(3):340-350. [doi: 10.1111/inr.12194] [Medline: 26058542]
84. Liu Z, Heffernan C, Tan J. Caregiver burden: a concept analysis. Int J Nurs Sci 2020;7(4):438-445. [doi: 10.1016/j.ijnss.2020.07.012] [Medline: 33195757]
85. Miller LM, Whitlatch CJ, Lyons KS. Shared decision-making in dementia: a review of patient and family carer involvement. Dementia (London) 2016;15(5):1141-1157. [doi: 10.1177/1471301214555542] [Medline: 25370075]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
FDA: Food and Drug Administration
HHIE: Hearing Handicap Inventory for the Elderly
IADL: instrumental activities of daily living
OTC: over-the-counter
TICS-m: Telephone Interview for Cognitive Status–modified

XSL•FO
RenderX

Edited by J Sarvestan; submitted 09.Sep.2025; peer-reviewed by L Coco; comments to author 08.Oct.2025; revised version received 19.Dec.2025; accepted 19.Dec.2025; published 01.Apr.2026.

Please cite as:
Urbanski DP, Hungs AM, Nelson PB, Gaugler JE
Facilitators and Barriers to Over-the-Counter Hearing Aid Use in People With Dementia: Semistructured Interview Study
JMIR Hum Factors 2026;13:e83857
URL: https://humanfactors.jmir.org/2026/1/e83857
doi:10.2196/83857
PMID:

©Dana P Urbanski, Alexander M Hungs, Peggy B Nelson, Joseph E Gaugler. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 01.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

e-Mental Health Intervention Preferences Among Informal Caregivers of Adults Living with Chronic Kidney Disease: Semistructured Interview Study

Chelsea Coumoundouros[1,2], PhD; Abla Sami[1], MSc; Paul Farrand[2], PhD; Robbert Sanderman[3], PhD; Louise von Essen[1], PhD; Joanne Woodford[1], PhD

[1]CIRCLE – Complex Intervention Research in Health and Care, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden
[2]Clinical Education, Development and Research (CEDAR), Psychology, University of Exeter, Exeter, United Kingdom
[3]Department of Health Psychology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

Corresponding Author:
Joanne Woodford, PhD
CIRCLE – Complex Intervention Research in Health and Care
Department of Women's and Children's Health
Uppsala University
Dag Hammarskjölds väg 14B
Uppsala, 751 05
Sweden
Phone: 1 729 999 211
Email: joanne.woodford@uu.se

Abstract

Background: Informal caregivers of people living with chronic kidney disease commonly experience mental health difficulties. However, there is currently limited access to scalable psychological interventions such as those using e-mental health platforms. Furthermore, there is an absence of such interventions tailored for this group of caregivers, which may limit acceptability.

Objective: To inform future e-mental health intervention development, this study aimed to explore preferences for e-mental health interventions among informal caregivers of people living with chronic kidney disease.

Methods: A total of 13 informal caregivers of people living with chronic kidney disease participated in semistructured interviews. Eligible informal caregivers lived in the United Kingdom and were recruited via nonprofit organizations and social media advertisements. Semistructured interviews followed a topic guide exploring caregivers' e-mental health intervention preferences. Data were analyzed using manifest content analysis.

Results: A total of 4 categories and 13 subcategories were generated. "Design preferences" illustrates key intervention design features that were important to caregivers, such as involving caregivers and other interest-holders in the design process. "Content preferences" describes the information caregivers valued in an intervention, such as information about chronic kidney disease and resources to support caregivers. "Support preferences" reflects the characteristics of support caregivers preferred if support was available as part of the intervention, such as peer support and providing support that accommodates individual preferences. Lastly, "Dissemination preferences" describes methods caregivers recommended to promote intervention uptake, such as engaging communication strategies and trusted and accessible sources of information.

Conclusions: Caregivers identified a number of e-mental health intervention design and delivery preferences that should be considered when developing e-mental health interventions for caregivers of people living with chronic kidney disease. To ensure such interventions are acceptable, feasible, and relevant, co-design approaches should be adopted during the intervention development process.

(JMIR Hum Factors 2026;13:e80962) doi:10.2196/80962

KEYWORDS

chronic kidney disease; e-mental health; informal caregiver; intervention development

XSL•FO
RenderX

Introduction

Background

Chronic kidney disease (CKD) impacts over 800 million people worldwide, is a leading cause of death globally, and significantly impacts quality of life [1-3]. Informal caregivers (caregivers) play a vital role in supporting people living with CKD, providing medical, practical, and emotional support [4-7]. Research suggests caregivers experience worse mental health (eg, symptoms of depression and stress) compared to noncaregivers [8], with caregivers reporting negative impacts, such as high levels of burden, poor quality of life, and mental health difficulties [4,6,9-11]. For example, studies have shown that between 30% and 60% of caregivers of people living with CKD experience symptoms of depression [7,10,11], which is higher compared to the general adult population [12,13]. Consistent with these findings, qualitative research shows that caregivers of people living with end-stage kidney disease experience mental health difficulties due to the caregiving role and could benefit from psychological and emotional support [14]. Ensuring caregivers receive acceptable, feasible, evidence-based mental health support is important not only to ensure caregivers maintain their own well-being, but also because their mental health can also impact the mental health and well-being of the person living with CKD [15,16]. However, caregivers experience multiple barriers to accessing mental health support, such as mental health stigma and negative views of mental health support, limited time due to caregiving responsibilities, and prioritization of the needs of the care recipient [17-19]. All of these barriers may contribute to the low uptake of mental health support among caregivers [20-22].

e-Mental Health Interventions to Support Caregiver Mental Health

e-Mental health (e-MH) interventions offer one way to provide accessible mental health support to caregivers, given these interventions can be accessed flexibly, without the need for travel or being available for regular appointments [23,24]. The flexibility offered by e-MH interventions can overcome some access barriers experienced by caregivers, such as a lack of time or reliable availability to attend appointments with a mental health professional [17,18]. Additionally, a priority setting initiative highlighted the use of technology as a key research and clinical priority to enhance support for caregivers [25]. e-MH interventions, specifically internet-based cognitive behavioral therapy (CBT) interventions, have been shown to be as effective as comparable in-person interventions among the general population [26], with evidence showing e-MH interventions are also effective for caregivers [27,28].

Although e-MH interventions have been developed for other groups of caregivers (especially caregivers of people living with cancer or dementia) [29], to our knowledge, there are no e-MH interventions tailored to caregivers of people living with CKD, despite interest in this type of intervention [7]. Tailoring e-MH interventions to the intended population has been shown to be important to improve intervention acceptability [30,31]. Additionally, existing mental health interventions for caregivers of people living with CKD identified in the literature are in-person and group-based [32], which may not align with intervention preferences identified by caregivers of people living with CKD [7]. Therefore, the development of an e-MH intervention tailored to the preferences of caregivers of people living with CKD is important to provide accessible mental health support for this group, while ensuring the e-MH intervention is acceptable, feasible, and relevant in practice.

Approach to e-MH Intervention Development

To develop an e-MH intervention tailored to the preferences of caregivers of people living with CKD, the Medical Research Council framework for the development and evaluation of complex interventions will be followed [33]. This framework emphasizes the importance of considering context and involving interest-holders during the intervention development process as a strategy to enhance the acceptability and future implementation potential of interventions [33]. Our previous research has explored self-help intervention preferences among caregivers of people living with CKD [7] and how interest-holders who may be involved in implementing interventions for this group of caregivers (eg, health and social care professionals, mental health care professionals, and staff at nonprofit organizations) viewed e-MH health interventions for caregivers of people living with CKD [34]. We have also explored caregivers' current experiences accessing and receiving support from both informal, such as friends and family, and formal, such as health and social care professionals, sources [18]. This study further explores caregivers' e-MH intervention preferences to provide insights that can be used to inform the development of an e-MH intervention that is acceptable, feasible, and relevant to caregivers of people living with CKD.

Aim

This study aimed to explore e-MH intervention preferences among informal caregivers of people living with CKD.

Methods

Study Design

A qualitative study using semistructured interviews and manifest content analysis [35] was conducted. Results are reported following the SRQR (Standards for Reporting Qualitative Research) checklist [36] (Multimedia Appendix 1). The study sample and data collection procedures are the same as our previous qualitative study exploring caregivers' experiences of accessing and receiving support [18].

Ethical Considerations

Ethical approval was obtained from the University of Exeter Psychology Research Ethics Committee (reference 513911) and from the Swedish Ethical Review Authority (dnr 2022-03068-01). Informed consent was obtained from participants prior to data collection. A professional transcription company transcribed audio files, and the resulting transcripts were deidentified by the research team prior to data analysis.

Sampling

Convenience sampling was used to recruit caregivers using advertisements shared by CKD and/or caregiver nonprofit

XSL•FO
RenderX

organizations (eg, via newsletters, email listserv, organization websites, and social media posts), paid and unpaid social media posts on study-specific social media pages, and contacting caregivers who participated in a related study [7] and expressed interest in taking part in an interview. Study advertisements included images of people with a range of genders, ages, and ethnicities to encourage participation of caregivers with diverse backgrounds. Caregivers interested in taking part in an interview contacted the research team via telephone or email. Caregivers were eligible to participate if they were at least 18 years of age, caring for an adult living with CKD, and residing in the United Kingdom, with eligibility confirmed by a research team member via telephone or email prior to the interview.

Data Collection

Between May 2022 and June 2023, a total of 13 semistructured interviews with a mean length of 63 (range 43-87) minutes were conducted by CC via Microsoft Teams (n=12) or telephone (n=1). The interview guide was partly informed by related qualitative research [17,37,38] and explored topics including support needs, e-MH intervention preferences, and potential barriers and facilitators to intervention use (Multimedia Appendix 2). Before the interview, caregivers were provided with a brief, written description of an e-MH intervention (Multimedia Appendix 2) based on existing low-intensity CBT interventions that have been developed for caregivers experiencing anxiety and/or depression [31,39]. The e-MH intervention was described as a CBT-based online program (eg, app or website) designed to support caregivers experiencing anxiety and/or depression. The intervention content was described as including CBT-based activities and materials (eg, goal setting and behavioral activation), in addition to including information relevant to caregivers of people living with CKD (eg, diet and relevant support services). The intervention was also described as including support from a trained professional, with the option for support to be provided in-person at a convenient location (eg, community center, GP office, or kidney care unit) or via telephone, video-call, email, or SMS text message. The types of support the trained professional would provide were described as including understanding the caregiver's mental health needs, supporting the caregiver to use the intervention and work through the CBT-based activities, and directing caregivers to additional sources of support (eg, programs within the community, hospital, etc). Interviews were conducted until a rich understanding of intervention preferences was achieved, and no new categories were generated [40]. This was discussed iteratively during data collection by JW and CC.

Data Analysis

Manifest content analysis with inductive coding was conducted by 2 researchers independently (CC and AS) [35]. All transcripts were read and then coded independently by CC and AS for data familiarization. Coding was undertaken by first creating condensed meaning units, followed by developing codes to describe the condensed meaning units. CC and AS generated categories and subcategories based on their initial coding, resulting in the development of 2 sets of categories and subcategories. CC and AS then held a workshop where they reviewed both sets of categories and subcategories, discussed similarities and differences, and came to a shared understanding of the categories and subcategories that best reflected the data. A single, revised version of categories and subcategories was developed by CC and shared for feedback with AS. This version of categories and subcategories was then shared for feedback with JW and PF, resulting in the development of the final set of categories and subcategories. Trustworthiness was established by having 2 independent researchers code the data prior to discussion, keeping memos throughout analysis, peer examination of categories and subcategories by 2 additional research team members, and inclusion of quotes to support the categories generated during the analysis process [41].

Researcher Characteristics

Interviews were conducted by CC, who also led data analysis. CC has a public health background and, at the time of the study, was a PhD candidate with experience conducting qualitative research related to caregiver needs and intervention preferences. AS has a community and public health background and has completed training courses in qualitative research. JW has a PhD in Psychology, was the doctoral supervisor of CC, and has extensive experience conducting, teaching, and supervising qualitative research. JW's research focuses on informal care and mental health. PF is an expert in CBT with extensive qualitative research experience, including research related to kidney disease. Additionally, PF has lived with CKD for over 25 years, during which he has experienced all forms of kidney replacement therapy and is currently on hemodialysis. LvE is a female professor in caring sciences, a psychologist, and specializes in mental health in people affected by somatic disease and their informal caregivers. LvE has extensive experience in teaching, conducting, and supervising qualitative research. RS is a health psychologist and an expert in research on coping with chronic somatic illnesses, interventions (both face-to-face and eHealth) in this area, and has worked extensively with both quantitative and qualitative methods. There are no preexisting relationships between the research team and participants.

Results

Sample Characteristics

Caregivers were all female, aged 52 (SD 13) years on average, lived in England (n=12) or Wales (n=1), and had a White (n=11) or South Asian (n=2) background. Caregivers often worked full- or part-time (n=7) while supporting their spouse (n=8), and had been providing care for an average of 7.2 (SD 7.6) years. About half (n=7) of care recipients were receiving kidney replacement therapy (dialysis or transplant). Caregivers provided support with a mean of 10 (SD 6) care-related activities, including cooking, communicating with the medical care team, emotional support, managing symptoms, and running errands. Sample characteristics are summarized in Table 1, with select individual characteristics with caregiver pseudonyms available in Multimedia Appendix 3 to facilitate the interpretation of supporting quotations.

XSL•FO
RenderX

Table 1. Summary of caregiver characteristics (n=13).

Characteristic	Value
Age in years, mean (SD; range)	52 (13; 31-67)
Gender (women), n (%)	13 (100)
Country of residence, n (%)	
England	12 (92)
Wales	1 (8)
Ethnic background, n (%)	
White	11 (85)
South Asian	2 (15)
Educational level, n (%)	
ISCED[a] 6 or higher	7 (54)
ISCED 5	2 (15)
ISCED 3	2 (15)
Other	2 (15)
Employment status, n (%)	
Full-time	4 (31)
Part-time	3 (23)
Homemaker	2 (15)
Not working or retired	4 (31)
Caregiver's relationship to the care recipient, n (%)	
Spouse/partner	8 (62)
Sibling	3 (23)
Parent	2 (15)
Child	1 (8)
Receipt of kidney replacement therapy, n (%)	
Dialysis	3 (23)
Transplant	4 (31)
No[b]	6 (46)
Living with care recipient (yes), n (%)	8 (62)
Time in caregiving role (years), mean (SD; range)	7.2 (7.6; 0.25-0.30)
How well the caregiver is coping, n (%)	
Very well	2 (15)
Well	2 (15)
Neither well nor not well	6 (46)
Not well	2 (15)
Very unwell	1 (8)

[a]ISCED: International Standard Classification of Education.

[b]Individuals not receiving kidney replacement therapy were either not at a stage of disease that required dialysis (n=5) or were receiving conservative kidney management (n=1).

Qualitative Results

Content analysis generated 4 categories and 13 subcategories reflecting caregivers' e-MH intervention preferences (Textbox 1).

XSL•FO
RenderX

Textbox 1. Overview of categories and subcategories.

Design preferences

- Considering safety
- Engaging and user-friendly design features
- Involving caregivers and other interest-holders in the design process
- Supporting flexible usage
- Tailoring to cultural and chronic kidney disease (CKD) caregiving contexts

Content preferences

- Providing information on CKD and how to support the care recipient
- Supporting emotional well-being

Support preferences

- Accommodating individual support preferences
- Knowing support is available
- Ensuring support providers have the right knowledge and skills
- Providing peer support

Dissemination preferences

- Engaging communication strategies
- Leveraging trusted and accessible sources

Design Preferences

Overview

This category describes e-MH intervention design features and characteristics that caregivers felt could help ensure an e-MH intervention is considered relevant and could enhance their engagement and trust in the intervention. This category includes the following five subcategories: (1) considering safety, (2) engaging and user-friendly design features, (3) involving caregivers and interest-holders in the design process, (4) supporting flexible usage, and (5) tailoring to cultural and CKD caregiving contexts.

Considering Safety

Including design features that would ensure caregivers feel safe using an e-MH intervention was expressed by caregivers in relation to 2 dimensions of safety, that is, psychological safety and data security. Psychological safety related to e-MH interventions that include online discussion forums, where anyone can respond to questions or comments posted by caregivers. Caregivers emphasized the importance of ensuring questions were answered respectfully and the potential need for monitoring in case incorrect information was shared or responses were disrespectful, judgmental, or negative.

> *I've seen things [forums] before where people would genuinely put something there, and you think they're crying for help, and somebody says, "What are you saying that for? Somebody's said that before." And I think that just straight away puts people right off, doesn't it? And people leave because of those sorts of things.... And if I put something on there and somebody said, "Oh, that's just stupid. What are you asking that question for?" That would make me think, "well, I'm just not going to bother anymore."* [Emily]

Data security related to caregivers' concerns regarding the security of information that would be shared within the intervention, and ensuring caregivers were informed about how their data were being handled.

> *You always want to know what's happening with the information that you give, who uses it, and how.... I think people are open-minded about that [sharing information] when they know.* [Olivia]

Engaging and User-Friendly Design Features

Ensuring e-MH interventions were engaging, not relying solely on written material, and incorporated opportunities for interaction was valued by caregivers. Opportunities for interaction that could be included within an e-MH intervention suggested by caregivers included creative activities, such as art-based activities and opportunities to interact with people, including other caregivers (eg, an online forum where caregivers could ask questions) and health care professionals (eg, a live session or talk with health care professionals, including time for caregivers to ask questions).

> *But I think you can't beat human contact. And if we've learned anything through the pandemic, it's that people need people.* [Olivia]

Ensuring the intervention was user-friendly was viewed as another important element to support caregiver engagement. Key design considerations suggested by caregivers to ensure the design of a user-friendly intervention included offering the intervention as an app to allow caregivers to access the

XSL•FO
RenderX

intervention quickly on their mobile phone and presenting information in an "easy to digest" (Freya) way using easy-to-understand language. Designing an intervention to meet the needs of caregivers with different levels of digital literacy was also emphasized as important to ensure the intervention is "straightforward" (Emily) to use and does not "presume that everybody's quite techy" (Emily).

Involving Caregivers and Other Interest-Holders in the Design Process

Involvement of people with lived experience, such as caregivers, people living with CKD, and other interest-holders with relevant expertise (eg, kidney care providers), during intervention development was recommended by caregivers as a key strategy to enhance their trust and positive views of an e-MH intervention.

> *I think when you're setting it [the e-MH intervention] up, when you're developing it, get people like me and people you've interviewed to test drive it for you. Because end users will tell you how it's working, [and if] it's meeting their need.* [Zainab]

Sustained involvement of people with lived experience and relevant expertise, such as having a project advisory group including public contributors and other interest-holders during initial testing and later intervention implementation, was also suggested as a strategy to ensure their perspectives continue to inform refinement of the intervention design. Such involvement was also perceived as important to demonstrate that the research and intervention development team is responsive to caregivers' preferences.

Supporting Flexible Usage

The online format was perceived as making e-MH interventions flexible and thus feasible for caregivers to access and use. Caregivers expressed the importance of the ability to choose when and where to access the intervention and how long to engage with it at one time. This flexibility was perceived as supporting caregivers' ability to balance their caregiving responsibilities and needs (eg, the need to limit contact with others due to COVID-19, with intervention use).

> *Also, if it's online, it means that we can do it whilst we are cooking, it means we can do it after work, or whatever.* [Priya]

Additionally, caregivers wanted to be able to access information in the intervention based on their own needs rather than needing to work through all the content in a set order. The intervention's online format was viewed as meeting this preference for flexible usage.

Tailoring to Cultural and CKD Caregiving Contexts

An e-MH intervention specifically for caregivers was valued as a way to provide a tailored support resource that would align with caregivers' needs and context. Tailoring the intervention to caregivers of people living with CKD was seen as important to enhance the perceived relevance of the intervention, which could encourage caregiver engagement. Caregivers expressed preferences for e-MH interventions tailored to recognize the caregiving experience. For example, addressing the "practical side and the emotional side" (Zainab) of caregiving, and including information and language relevant to caring for someone living with CKD.

> *I would want, as you mentioned, some intervention where it is tailored towards people that are caring for people with kidney disease and renal related issues. So, things like [including] useful terminology, using the language that kidney patients use, and carers use. I think that would be really helpful because it would set the tone, it would feel like the right environment.* [Priya]

Two caregivers had South Asian backgrounds, and they also mentioned the importance of cultural tailoring and cultural sensitivity when developing the e-MH intervention. For example, being aware of various religious holidays when scheduling live events or support sessions.

Content Preferences

Overview

This category describes the information and resources caregivers felt should be included within an e-MH intervention. This category has two subcategories: (1) providing information on CKD and how to support the care recipient, and (2) supporting emotional well-being.

Providing Information on CKD and How to Support the Care Recipient

Caregivers were interested in receiving information about CKD, including the stages, treatments (eg, dialysis, transplants, and side effects), diet, and information about living with someone who is living with a transplant and is immunosuppressed. Information was desired not only to understand the care recipient's current CKD stage but also to prepare for the progression of CKD to other stages and to understand future care and treatment needs.

> *I think for me, it [intervention content] would be what the future would be like, about dialysis, transplant, and what life would be like. Just so you can prepare yourself emotionally—this is what my life is going to be like.* [Freya]

Caregivers also wanted information that would help them to support the care recipient. This included information on how to provide them with emotional support so they "feel better in themselves as well" (Sarah), motivate them to do things to support their health and well-being (eg, attend medical appointments), and communicate effectively with the care recipient's care team to ensure they receive optimal care. Preferences for the intervention to include information about CKD were also connected to caregivers' desire for an intervention that would not only support them as caregivers, but also help them understand what the care recipient was experiencing and how to help monitor their health.

Supporting Emotional Well-Being

Caregivers expressed a need for resources offering practical tips and strategies they could use to practice self-care and manage the different emotions they had to cope with while

XSL•FO
RenderX

caring, such as coping with uncertainty and anxiety or feeling low:

> *I think ways of distracting yourself and ways of being kind to yourself really, suggestions of something you can do that's mindfulness, that takes your mind away from the whole caring situation.* [Sarah]

In addition to resources supporting their emotional well-being, caregivers were interested in strategies to help them balance different roles and responsibilities while identifying when to reach out for support. An e-MH intervention with practical, concrete activities caregivers could use in their daily routines, rather than just providing information, was viewed as something that would increase the perceived value of the intervention.

Support Preferences

Overview

This category describes characteristics of both the type of support that should be available within an e-MH intervention and the characteristics of individuals who would support the intervention. This category has the following four subcategories: (1) accommodating individual support preferences, (2) knowing support is available, (3) ensuring support providers have the right knowledge and skills, and (4) providing peer support.

Accommodating Individual Support Preferences

Support that could be customized to accommodate caregivers' individual preferences was preferred to ensure support would be accessible and acceptable to caregivers. Caregivers expressed mixed preferences regarding the support format, with in-person, written, telephone, and video-call support being mentioned. Flexibility to select the support format and change it based on needs and preferences was expressed, with Kate explaining that "if somebody doesn't want to physically talk, there is that option of just sending something [written]." Support should also be responsive to the type of support the caregiver wanted, for example, whether a caregiver wants emotional support or informational support (ie, to be directed to relevant resources and information). In-person support was viewed as an important way to connect with the support provider and facilitate relationship building and trust. However, in-person support sessions were not always viewed as feasible due to a lack of time and the potential burden of having to travel to the session. In addition to format preferences, the ability to select a support provider based on the provider's characteristics, such as gender or cultural background, was viewed as a way to further meet caregivers' preferences and help caregivers feel more comfortable seeking support:

> *I think having a mix of genders, sexualities [among support providers]—I think it's important because men might feel comfortable talking to men, women might feel comfortable [talking to women], or it might be the opposite. I think having access to both would really help.* [Priya]

Knowing Support Is Available

Caregivers expressed a preference for support to be available as part of an e-MH intervention, given it would be "nice to know that there is support there [in the intervention]" (Sarah). Caregivers felt that support providers should take active steps to support them by reaching out periodically rather than relying on the caregiver to initiate contact with the support provider. Support is needed to address caregivers' emotional support needs by building a relationship with the caregiver and providing encouragement:

> *And I think some, just some positive reinforcement along the way that you're doing the best that you can in this moment.* [Olivia]

Additionally, practical support to assist caregivers in using the intervention, such as technical support, understanding and applying intervention content in practice, and checking on caregivers' progress to motivate intervention usage, was important:

> *When I come across something I don't understand or I don't feel confident about, that's when I need support.* [Holly]

Ensuring Support Providers Have the Right Knowledge and Skills

Caregivers felt support providers would need to have a range of skills to meet caregivers' support needs, including cultural sensitivity, emotional support skills (eg, empathy and active listening), knowledge of CKD, knowledge of the intervention (eg, navigating intervention content), and an understanding of the caregiving experience:

> *Maybe somebody who's had some training on counselling or somebody who's empathetic. It [the support provider] needs to be somebody who can try and understand where you're at and where you are coming from.* [Rebecca]

It was acknowledged that a single support provider may not have the full range of knowledge and skills to meet all support needs. It was suggested that support providers could vary based on the type of support needed. For example, a support provider could be available for CKD medical information needs, and potentially a different provider for emotional support needs.

Providing Peer Support

Caregivers wanted an e-MH intervention to include opportunities to connect with other caregivers with shared characteristics, such as a similar age range or lived-experience (eg, caring for a young family and someone living with CKD). Shared lived experience was viewed as important to foster understanding and provide validation throughout the caregiving journey. Peer support was seen as a way to receive advice from other caregivers, including caregivers with more experience who may be able to provide insights on caring for people at different stages of CKD:

> *I think what I would really love is to meet somebody to say 'I was in that situation'.... It's all very well reading about it [CKD], but you still, you've got all those questions to ask. In terms of how much medicine is going to be in the house or how many hospital visits will there be and how will it impact on the whole family and things like that. So yes, first-hand experience, I think would be very helpful.* [Sarah]

XSL•FO
RenderX

However, depending on the format of peer support, guidelines and monitoring were perceived to be needed to ensure peer support provided is nonjudgmental and confidential.

Dissemination Preferences

Overview

This category describes key elements that should be considered when sharing information about the e-MH intervention with caregivers. This category includes 2 subcategories: (1) engaging communication strategies and (2) leveraging trusted and accessible sources.

Engaging Communication Strategies

Caregivers expressed that communication strategies used to promote the e-MH intervention would need to create interest and excitement in the intervention to motivate uptake. They suggested implementing communication strategies using short, easy-to-understand language that clearly expresses the purpose of the intervention so caregivers can quickly understand who should use the intervention and how it will meet their needs:

> *You can get bogged down, especially on our journey, with lots of words that you either don't understand or they're not in your day-to-day vocabulary. So, you would spend more time trying to Google what it meant. So, if something is bright, if it's clear, if it's short, if it's concise, those are the things that engage me.* [Kate]

Caregivers suggested that communication strategies should leverage factors they identified as motivating their uptake of an e-MH intervention, including an interest and desire to learn, the opportunity to try a new intervention, and evidence (eg, testimonials that other caregivers benefited from the intervention).

> *And knowing that you are part of a...pilot scheme, it's quite exciting, it's something new, it's something exciting, you know. And it's something that you want to be actively participating in.* [Claire]

Leveraging Trusted and Accessible Sources

Caregivers expressed that an e-MH intervention should be endorsed by a trusted source to enhance their confidence that the intervention and the information provided within it would be trustworthy. A variety of sources were considered trustworthy, including the kidney care unit, nonprofit CKD and caregiver organizations, and other caregivers:

> *I think using the kidney charities would be a really good way [to promote the intervention]. If they promote it, then you'd feel this is endorsed by someone I trust to have good information. I think if you just saw a random advert you'd think, 'is this actually legit?'* [Freya]

In addition to considering the trustworthiness of the intervention source, the accessibility and reach of the intervention source were other important considerations regarding who should disseminate information about the e-MH intervention. Not all potential intervention sources were considered accessible to all caregivers; for example, some caregivers had limited contact with the kidney care unit due to living far from the unit or not always attending appointments with the person they were caring for. Additionally, some caregivers felt that health care professionals can be overburdened and focused on the person living with CKD, and therefore, may not be the best source of information about interventions for caregivers.

Discussion

Principal Results

Overview

This study explored the preferences of caregivers of people living with CKD that require consideration to inform the development of an acceptable and relevant e-MH intervention for this group of caregivers. Study findings illustrated preferences for an e-MH intervention tailored to the context of caregivers in relation to intervention design, intervention content, support options, and dissemination approaches. Findings align with the Medical Research Council framework for the development and evaluation of complex health care interventions, which highlights the importance of considering context (eg, the caregivers' background and their caregiving situation) as a core component of the intervention development process [33].

e-MH Intervention Preferences

Many preferences identified in this study align with intervention preferences identified in other caregiving literature, such as the need for practical, disease-specific content, the provision of empathetic and knowledgeable support from a trained provider, and flexible options for how to use and engage with the intervention [7,42-44]. Although caregivers commonly report wanting information about the care recipient's illness [7,42], this study identified an interest in not only information related to the care recipient's current stage of disease, but also advanced stages as a way to prepare for the future. In contrast, a study focused on caregivers of people living with cancer found a preference for information specific to the care recipient's current stage of disease [42]. This difference may relate to the progressive nature of the CKD trajectory and the increase in care needs as the disease progresses to more advanced stages [45]. Additionally, other work suggests that seeking and obtaining information is important to support self-management of CKD by people living with CKD and their caregivers and that unmet information needs impact caregivers' mental health [46,47]. This highlights the importance of addressing caregiver information and education needs within an e-MH intervention.

Although the preferences identified in this study align with other research, it is important to recognize that this study only captured the preferences of women caregivers. Many e-MH intervention preferences reported by men align with the findings of this study, with men valuing easy-to-use e-MH interventions with flexible engagement options (eg, content available in multiple formats or the option to receive support from a trained provider) [48]. However, some e-MH intervention preferences may vary by gender; for example, one study found that men preferred e-MH interventions with a video-game format more than women [49]. Potential gender differences in e-MH

XSL•FO
RenderX

intervention preferences highlight the importance of including caregivers with diverse gender identities in the intervention development process to enhance intervention acceptability.

Contrasting Caregiver and Other Interest-Holder e-MH Intervention Preferences

To enhance the future implementation potential of an e-MH intervention for caregivers, the perspective of other interest-holders who would be involved in intervention delivery and dissemination is vital [33]. Our related research exploring potential implementers' (ie, professionals working in kidney care, mental health care, and nonprofit organizations) perspectives of an e-MH intervention for caregivers of people living with CKD [34] identified a number of similarities to the caregiver perspectives explored in this study. Potential implementers also viewed e-MH interventions as a flexible means of providing mental health support that could fit within caregivers' caring responsibilities, and that tailoring of intervention content and availability of support within the intervention were important design features [34]. However, they raised concerns regarding e-MH intervention access and digital exclusion [34]. Although caregivers in this study acknowledged designing an intervention to suit different digital literacy levels was important, they did not express similar concerns regarding digital exclusion. Since interest-holder views of e-MH intervention design and caregivers' ability to use the intervention are important implementation determinants [50], concerns regarding digital exclusion should be addressed as the intervention development process continues. Strategies to address these concerns and support intervention access for caregivers with lower digital literacy should include training to show caregivers how to use the intervention [41,42] and providing a nondigital version [34]. Continued engagement of caregivers and interest-holders in the co-design of the e-MH intervention will be needed to ensure the intervention is user-friendly and meets the preferences of relevant user groups [51].

Trust in e-MH Interventions

Trust was identified by caregivers as an important design element of e-MH interventions in relation to the credibility of the intervention and its content, trust in the organization disseminating or endorsing the intervention, and data security. Trust in e-MH interventions has only been minimally explored in the caregiving literature [29,34,52,53]; trust is an important element of eHealth intervention design [54]. For example, persuasive system design, an approach to designing digital interventions and technologies to change end users' attitudes or behaviors, highlights technology credibility as an important intervention feature that can be built through strategies such as highlighting the expertise of the intervention developers and the intervention's evidence base [55]. Another study suggests that aligning intervention features with user preferences is another way to build trust in e-MH interventions [54]. Applying this to the user preferences identified in this study, designing an e-MH intervention with multiple support format options that caregivers can choose from, such as written or in-person support, and having a gender and culturally diverse group of support providers such that caregivers can choose a support provider they feel most comfortable working with, could help foster trust of the e-MH intervention [54]. Additional strategies to enhance trust based on our findings and related research include leveraging trusted sources, such as health care providers and established nonprofit organizations, to disseminate information about the intervention, and involving caregivers and other interest-holders with relevant expertise in intervention development [34]. Incorporating strategies to build trust and credibility in the development of e-MH interventions will be important to support intervention adoption [54,56].

Implementation Considerations

An e-MH intervention for caregivers could fit well within a collaborative care model [57,58], whereby a central contact could link caregivers to health and social care professionals who can address their individual support needs. These could include facilitating a connection with the kidney health care team for support with medical information needs or a connection to a mental health professional for support with mental health needs. This approach to the provision of support aligns with the preference expressed in this study for support to be provided by providers with the appropriate competencies based on the caregivers' specific support needs. Collaborative care models have been shown to benefit caregivers, for example, improving caregivers' ability to cope and providing reassurance through the provision of accessible and acceptable support [59,60]. A collaborative care model to support caregivers of people living with CKD could be supported by the Assistant Wellbeing Practitioner (Kidney) role that has been proposed to be integrated within kidney units to provide mental health support to people living with CKD and their caregivers while working with multidisciplinary kidney care teams [61]. Competencies for the Assistant Wellbeing Practitioner (Kidney), including having knowledge related to mental health and CKD, and using psychologically informed approaches to support caregivers of people living with CKD, such as supporting caregiver use of digital interventions [61], align with the support preferences expressed by caregivers in this study. However, this role has not yet been implemented throughout the health care system.

Alternatively, nonprofit, community organizations have also been endorsed by caregivers [7] as an alternative implementation context for an e-MH intervention for caregivers of people living with CKD. Interest-holders working within community organizations also viewed an e-MH intervention for caregivers as aligning with the mission of their organization [34]. Implementation of e-MH interventions for caregivers within community organizations could align with the dissemination preferences expressed in this study, which highlighted that health care professionals can be viewed as less accessible to caregivers, given that not all interact with the care recipient's care team. Another factor that may facilitate implementation of an e-MH intervention for caregivers within community organizations is the preference for intervention support to be provided by peers (ie, people with lived experience of caring for someone living with CKD) and the high value caregivers place on the knowledge gained from lived experience [18,62]. Shared lived experience can support relationship building, which is an important common factor in mental health interventions [63]. There is also evidence that e-MH interventions supported

XSL•FO
RenderX

by peers are as effective as e-MH interventions supported by mental health care professionals [64]. Leveraging peers to provide support as part of an e-MH intervention for caregivers may also enhance the feasibility of implementing an e-MH intervention, given existing capacity challenges within the mental health care system [65] and the lack of psychosocial care professionals within the kidney care sector [66]. Therefore, community organizations should also be considered as a possible implementation setting for an e-MH intervention for caregivers.

Limitations

The perspectives of some potentially important subgroups of caregivers may not have been captured in this study, including the perspective of men caring for someone living with CKD, and caregivers from a wider range of cultural backgrounds. Caregiver mental health status was not assessed in this study; therefore, the preferences identified may not reflect the preferences of caregivers experiencing common mental health difficulties, such as anxiety or depression, who are the intended users of the e-MH intervention. However, the perspectives of caregivers who were coping well and not well were included in this work. Future studies may want to focus on recruiting caregivers reporting current or past experiences of common mental health difficulties to further inform e-MH intervention development. Caregivers were recruited using advertisements on various online platforms. Therefore, the views of caregivers with lower digital literacy may not have been captured in this study. As e-MH intervention development continues, strategies to engage caregivers with low digital literacy (eg, recruitment via newspaper or printed advertisements and in-person data collection opportunities) should be explored. Finally, caregivers' e-MH intervention preferences were obtained based on a text-based description of e-MH interventions, which could be interpreted differently based on caregivers' previous experiences and interactions with e-MH or other digital health interventions. Future work developing the e-MH intervention for caregivers of people living with CKD should continue to engage caregivers, leveraging recruitment strategies to enhance participant diversity [67], in the co-design process as tangible intervention prototypes are developed to ensure the intervention continues to align with caregivers' needs and preferences.

Conclusions

This study identified a number of preferences regarding e-MH interventions among caregivers of people living with CKD that can be used to inform future intervention development. Ensuring interventions are tailored toward caregivers' context, include content that meets caregivers' information needs and preferences, have flexible access and support options to accommodate caregivers' caring responsibilities and support needs, and are disseminated using engaging messaging via trusted sources were key design considerations. Future work should use the results of this study to codevelop e-MH intervention prototypes to continue the intervention development process, ensuring continued collaboration and feedback from caregivers and other interest-holders with relevant expertise.

Acknowledgments

We thank everyone who shared information about the study and all the caregivers who shared their experiences with us during interviews.

Funding

This study received funding from the European Union's Horizon 2020 research and innovation program under the Marie-Sklodowska Curie grant agreement number 814072, and was supported by U-CARE, a strategic research environment funded by the Swedish Research Council (dnr 2009-1093). The sponsors were not involved in the design, collection, analysis, and interpretation of data, or writing of the manuscript.

Data Availability

The datasets generated and/or analyzed during this study are not publicly available to maintain participants' privacy and confidentiality, but may be provided upon reasonable request from JW.

Authors' Contributions

Conceptualization: CC, JW
Methodology: CC, PF, JW
Formal analysis: CC, AS, PF, JW
Investigation: CC
Data curation: CC
Writing—original draft: CC, AS
Writing—review and editing: CC, AS, PF, RS, LvE, JW
Visualization: CC
Project administration: CC
Funding acquisition: RS, LvE
Supervision: PF, RS, LvE, JW

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
SRQR checklist.
[PDF File (Adobe PDF File), 154 KB - humanfactors_v13i1e80962_app1.pdf]

Multimedia Appendix 2
Caregiver interview guide.
[DOCX File , 142 KB - humanfactors_v13i1e80962_app2.docx]

Multimedia Appendix 3
Individual caregiver characteristics.
[DOCX File , 17 KB - humanfactors_v13i1e80962_app3.docx]

References

1. Francis A, Harhay MN, Ong ACM, Tummalapalli SL, Ortiz A, Fogo AB, American Society of Nephrology, European Renal Association, International Society of Nephrology. Chronic kidney disease and the global public health agenda: an international consensus. Nat Rev Nephrol 2024;20(7):473-485 [FREE Full text] [doi: 10.1038/s41581-024-00820-6] [Medline: 38570631]
2. Jager KJ, Kovesdy C, Langham R, Rosenberg M, Jha V, Zoccali C. A single number for advocacy and communication-worldwide more than 850 million individuals have kidney diseases. Kidney Int 2019;96(5):1048-1050. [doi: 10.1016/j.kint.2019.07.012] [Medline: 31582227]
3. Fletcher BR, Damery S, Aiyegbusi OL, Anderson N, Calvert M, Cockwell P, et al. Symptom burden and health-related quality of life in chronic kidney disease: a global systematic review and meta-analysis. PLoS Med 2022;19(4):e1003954 [FREE Full text] [doi: 10.1371/journal.pmed.1003954] [Medline: 35385471]
4. Hoang VL, Green T, Bonner A. Informal caregivers' experiences of caring for people receiving dialysis: a mixed-methods systematic review. J Ren Care 2018;44(2):82-95. [doi: 10.1111/jorc.12235] [Medline: 29357407]
5. DePasquale N, Cabacungan A, Ephraim PL, Lewis-Boyér L, Powe NR, Boulware LE. Family members' experiences with dialysis and kidney transplantation. Kidney Med 2019;1(4):171-179 [FREE Full text] [doi: 10.1016/j.xkme.2019.06.001] [Medline: 32734197]
6. Gilbertson EL, Krishnasamy R, Foote C, Kennard AL, Jardine MJ, Gray NA. Burden of care and quality of life among caregivers for adults receiving maintenance dialysis: a systematic review. Am J Kidney Dis 2019;73(3):332-343. [doi: 10.1053/j.ajkd.2018.09.006] [Medline: 30454885]
7. Coumoundouros C, Farrand P, Hamilton A, von Essen L, Sanderman R, Woodford J. Cognitive behavioural therapy self-help intervention preferences among informal caregivers of adults with chronic kidney disease: an online cross-sectional survey. BMC Nephrol 2023;24(1):4 [FREE Full text] [doi: 10.1186/s12882-022-03052-7] [Medline: 36600229]
8. Pinquart M, Sörensen S. Differences between caregivers and noncaregivers in psychological health and physical health: a meta-analysis. Psychol Aging 2003;18(2):250-267. [doi: 10.1037/0882-7974.18.2.250] [Medline: 12825775]
9. Shah KK, Murtagh FEM, McGeechan K, Crail SM, Burns A, Morton RL. Quality of life among caregivers of people with end-stage kidney disease managed with dialysis or comprehensive conservative care. BMC Nephrol 2020;21(1):160 [FREE Full text] [doi: 10.1186/s12882-020-01830-9] [Medline: 32366220]
10. Hawamdeh S, Almari AM, Almutairi AS, Dator WLT. Determinants and prevalence of depression in patients with chronic renal disease, and their caregivers. Int J Nephrol Renovasc Dis 2017;10:183-189 [FREE Full text] [doi: 10.2147/IJNRD.S139652] [Medline: 28740416]
11. Adejumo OA, Iyawe IO, Akinbodewa AA, Abolarin OS, Alli EO. Burden, psychological well-being and quality of life of caregivers of end stage renal disease patients. Ghana Med J 2019;53(3):190-196 [FREE Full text] [doi: 10.4314/gmj.v53i3.2] [Medline: 31741490]
12. Arias de la Torre J, Vilagut G, Ronaldson A, Dregan A, Ricci-Cabello I, Hatch SL, et al. Prevalence and age patterns of depression in the United Kingdom. A population-based study. J Affect Disord 2021;279:164-172. [doi: 10.1016/j.jad.2020.09.129] [Medline: 33059219]
13. Steel Z, Marnane C, Iranpour C, Chey T, Jackson JW, Patel V, et al. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. Int J Epidemiol 2014;43(2):476-493 [FREE Full text] [doi: 10.1093/ije/dyu038] [Medline: 24648481]
14. Matthews M, McKeaveney C, Noble H, Reid J. Exploring the unmet needs and experiences of informal caregivers of patients with end-stage kidney disease (ESKD) receiving haemodialysis - a qualitative study. PLoS One 2024;19(5):e0302525 [FREE Full text] [doi: 10.1371/journal.pone.0302525] [Medline: 38722989]

XSL•FO
RenderX

15. Fu L, Wu Y, Zhu A, Wang Z, Qi H. Depression of caregivers is significantly associated with depression and hospitalization of hemodialysis patients. Hemodial Int 2022;26(1):108-113. [doi: 10.1111/hdi.12967] [Medline: 34227223]
16. Gerogianni G, Polikandrioti M, Babatsikou F, Zyga S, Alikari V, Vasilopoulos G, et al. Anxiety-depression of dialysis patients and their caregivers. Medicina (Kaunas) 2019;55(5):168 [FREE Full text] [doi: 10.3390/medicina55050168] [Medline: 31137563]
17. Mosher C, Given B, Ostroff J. Barriers to mental health service use among distressed family caregivers of lung cancer patients. Eur J Cancer Care (Engl) 2015;24(1):50-59 [FREE Full text] [doi: 10.1111/ecc.12203] [Medline: 24761985]
18. Coumoundouros C, Farrand P, Sanderman R, von Essen L, Woodford J. "Systems seem to get in the way": a qualitative study exploring experiences of accessing and receiving support among informal caregivers of people living with chronic kidney disease. BMC Nephrol 2024;25(1):7 [FREE Full text] [doi: 10.1186/s12882-023-03444-3] [Medline: 38172754]
19. Messina A, Amati R, Albanese E, Fiordelli M. Help-seeking in informal family caregivers of people with dementia: a qualitative study with iSupport as a case in point. Int J Environ Res Public Health 2022;19(12):7504 [FREE Full text] [doi: 10.3390/ijerph19127504] [Medline: 35742751]
20. Wolff JL, Dy SM, Frick KD, Kasper JD. End-of-life care: findings from a national survey of informal caregivers. Arch Intern Med 2007;167(1):40-46. [doi: 10.1001/archinte.167.1.40] [Medline: 17210876]
21. Mosher CE, Champion VL, Hanna N, Jalal SI, Fakiris AJ, Birdas TJ, et al. Support service use and interest in support services among distressed family caregivers of lung cancer patients. Psychooncology 2013;22(7):1549-1556 [FREE Full text] [doi: 10.1002/pon.3168] [Medline: 22941782]
22. Jegermalm M. Informal care and support for carers in Sweden: patterns of service receipt among informal caregivers and care recipients. Eur J Soc Work 2004;7(1):7-24. [doi: 10.1080/13691914504200021746]
23. Lal S, Adair CE. E-mental health: a rapid review of the literature. Psychiatr Serv 2014;65(1):24-32. [doi: 10.1176/appi.ps.201300009] [Medline: 24081188]
24. Sin J, Henderson C, Spain D, Cornelius V, Chen T, Gillard S. eHealth interventions for family carers of people with long term illness: a promising approach? Clin Psychol Rev 2018;60:109-125 [FREE Full text] [doi: 10.1016/j.cpr.2018.01.008] [Medline: 29429856]
25. Kent EE, Rowland JH, Northouse L, Litzelman K, Chou WS, Shelburne N, et al. Caring for caregivers and patients: research and clinical priorities for informal cancer caregiving. Cancer 2016;122(13):1987-1995 [FREE Full text] [doi: 10.1002/cncr.29939] [Medline: 26991807]
26. Hedman-Lagerlöf E, Carlbring P, Svärdman F, Riper H, Cuijpers P, Andersson G. Therapist-supported Internet-based cognitive behaviour therapy yields similar effects as face-to-face therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. World Psychiatry 2023;22(2):305-314 [FREE Full text] [doi: 10.1002/wps.21088] [Medline: 37159350]
27. Zhao Y, Feng H, Hu M, Hu H, Li H, Ning H, et al. Web-based interventions to improve mental health in home caregivers of people with dementia: meta-analysis. J Med Internet Res 2019;21(5):e13415 [FREE Full text] [doi: 10.2196/13415] [Medline: 31066680]
28. Yu Y, Xiao L, Ullah S, Meyer C, Wang J, Pot AM, et al. The effectiveness of internet-based psychoeducation programs for caregivers of people living with dementia: a systematic review and meta-analysis. Aging Ment Health 2023;27(10):1895-1911 [FREE Full text] [doi: 10.1080/13607863.2023.2190082] [Medline: 36951611]
29. Coumoundouros C, Mårtensson E, Ferraris G, Zuidberg JM, von Essen L, Sanderman R, et al. Implementation of e-mental health interventions for informal caregivers of adults with chronic diseases: mixed methods systematic review with a qualitative comparative analysis and thematic synthesis. JMIR Ment Health 2022;9(11):e41891 [FREE Full text] [doi: 10.2196/41891] [Medline: 36314782]
30. Farrand P, Mullan E, Rayson K, Engelbrecht A, Mead K, Greenberg N. Adapting CBT to treat depression in armed forces veterans: qualitative study. Behav Cogn Psychother 2019;47(5):530-540. [doi: 10.1017/S1352465819000171] [Medline: 30894234]
31. Woodford J, Farrand P, Hagström J, Hedenmalm L, von Essen L. Internet-administered cognitive behavioral therapy for common mental health difficulties in parents of children treated for cancer: intervention development and description study. JMIR Form Res 2021;5(7):e22709 [FREE Full text] [doi: 10.2196/22709] [Medline: 34142662]
32. Hovadick AC, Jardim VR, Paúl C, Pagano A, Reis I, Torres H. Interventions to improve the well-being of family caregivers of patients on hemodialysis and peritoneal dialysis: a systematic review. PeerJ 2021;9:e11713 [FREE Full text] [doi: 10.7717/peerj.11713] [Medline: 34322322]
33. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
34. Coumoundouros C, El Arab RA, Farrand P, Hamilton A, Sanderman R, von Essen L, et al. Potential implementers' perspectives on the development and implementation of an e-mental health intervention for caregivers of adults with chronic kidney disease: qualitative interview study. JMIR Hum Factors 2023;10:e51461 [FREE Full text] [doi: 10.2196/51461] [Medline: 37792676]

XSL•FO
RenderX

35. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
36. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014;89(9):1245-1251. [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
37. Woodford J, Månberg J, Cajander Å, Enebrink P, Harila-Saari A, Hagström J, et al. Help-seeking behaviour and attitudes towards internet-administered psychological support among adolescent and young adults previously treated for cancer during childhood: protocol for a survey and embedded qualitative interview study in Sweden. BMJ Open 2021;11(4):e041350 [FREE Full text] [doi: 10.1136/bmjopen-2020-041350] [Medline: 34155002]
38. Law E, Levesque JV, Lambert S, Girgis A. The "sphere of care": a qualitative study of colorectal cancer patient and caregiver experiences of support within the cancer treatment setting. PLoS One 2018;13(12):e0209436 [FREE Full text] [doi: 10.1371/journal.pone.0209436] [Medline: 30586391]
39. Farrand P, Woodford J, Coumoundouros C, Svedin F. Supported cognitive-behavioural therapy self-help versus treatment-as-usual for depressed informal caregivers of stroke survivors (CEDArS): feasibility randomised controlled trial. Cogn Behav Therap 2020;13:e23. [doi: 10.1017/s1754470x20000239]
40. Braun V, Clarke V. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. Qual Res Sport Exerc Health 2019;13(2):201-216. [doi: 10.1080/2159676x.2019.1704846]
41. Williams EN, Morrow SL. Achieving trustworthiness in qualitative research: a pan-paradigmatic perspective. Psychother Res 2009;19(4-5):576-582. [doi: 10.1080/10503300802702113] [Medline: 19579089]
42. Köhle N, Drossaert CH, Oosterik S, Schreurs KM, Hagedoorn M, van Uden-Kraan CF, et al. Needs and preferences of partners of cancer patients regarding a web-based psychological intervention: a qualitative study. JMIR Cancer 2015;1(2):e13 [FREE Full text] [doi: 10.2196/cancer.4631] [Medline: 28410157]
43. Biliunaite I, Dumarkaite A, Kazlauskas E, Sanderman R, Andersson G. ICBT program for improving informal caregiver well-being: a qualitative study. Internet Interv 2021;23:100361 [FREE Full text] [doi: 10.1016/j.invent.2021.100361] [Medline: 33489781]
44. Yu Y, Xiao L, Ullah S, Meyer C, Wang J, Pot AM, et al. The experiences of informal caregivers of people with dementia in web-based psychoeducation programs: systematic review and metasynthesis. JMIR Aging 2023;6:e47152 [FREE Full text] [doi: 10.2196/47152] [Medline: 37247218]
45. Burckhardt P, Nagin DS, Padman R. Multi-trajectory models of chronic kidney disease progression. AMIA Annu Symp Proc 2016;2016:1737-1746 [FREE Full text] [Medline: 28269932]
46. Donald M, Beanlands H, Straus S, Ronksley P, Tam-Tham H, Finlay J, et al. Identifying needs for self-management interventions for adults with CKD and their caregivers: a qualitative study. Am J Kidney Dis 2019;74(4):474-482. [doi: 10.1053/j.ajkd.2019.02.006] [Medline: 30952486]
47. Matthews M, Reid J, McKeaveney C, Noble H. Knowledge requirements and unmet needs of informal caregivers of patients with end-stage kidney disease (ESKD) receiving haemodialysis: a narrative review. Healthcare (Basel) 2021;10(1):57 [FREE Full text] [doi: 10.3390/healthcare10010057] [Medline: 35052221]
48. Opozda MJ, Oxlad M, Turnbull D, Gupta H, Smith JA, Ziesing S, et al. Facilitators of, barriers to, and preferences for e-mental health interventions for depression and anxiety in men: metasynthesis and recommendations. J Affect Disord 2024;346:75-87 [FREE Full text] [doi: 10.1016/j.jad.2023.11.015] [Medline: 37949238]
49. Smail-Crevier R, Powers G, Noel C, Wang J. Health-related internet usage and design feature preference for e-mental health programs among men and women. J Med Internet Res 2019;21(3):e11224 [FREE Full text] [doi: 10.2196/11224] [Medline: 30882361]
50. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated Consolidated Framework for Implementation Research based on user feedback. Implement Sci 2022;17(1):75 [FREE Full text] [doi: 10.1186/s13012-022-01245-0] [Medline: 36309746]
51. Donald M, Beanlands H, Straus SE, Smekal M, Gil S, Elliott MJ, et al. A web-based self-management support prototype for adults with chronic kidney disease (My Kidneys My Health): co-design and usability testing. JMIR Form Res 2021;5(2):e22220 [FREE Full text] [doi: 10.2196/22220] [Medline: 33560245]
52. Christie HL, Schichel MCP, Tange HJ, Veenstra MY, Verhey FRJ, de Vugt ME. Perspectives from municipality officials on the adoption, dissemination, and implementation of electronic health interventions to support caregivers of people with dementia: inductive thematic analysis. JMIR Aging 2020;3(1):e17255 [FREE Full text] [doi: 10.2196/17255] [Medline: 32401217]
53. Christie HL, Boots LMM, Peetoom K, Tange HJ, Verhey FRJ, de Vugt ME. Developing a plan for the sustainable implementation of an electronic health intervention (Partner in Balance) to support caregivers of people with dementia: case study. JMIR Aging 2020;3(1):e18624 [FREE Full text] [doi: 10.2196/18624] [Medline: 32584261]
54. Westheimer JL, Smith RP, Iacobelli P, Oh H, Tezino L, Khan R, et al. J Affect Disord 2024;367:318-323. [doi: 10.1016/j.jad.2024.08.206] [Medline: 39226937]
55. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. CAIS 2009;24. [doi: 10.17705/1CAIS.02428]

XSL•FO
RenderX

56. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021;21(1):1556 [FREE Full text] [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]
57. Thota AB, Sipe TA, Byard GJ, Zometa CS, Hahn RA, McKnight-Eily LR, Community Preventive Services Task Force. Collaborative care to improve the management of depressive disorders: a community guide systematic review and meta-analysis. Am J Prev Med 2012;42(5):525-538. [doi: 10.1016/j.amepre.2012.01.019] [Medline: 22516495]
58. Reist C, Petiwala I, Latimer J, Raffaelli SB, Chiang M, Eisenberg D, et al. Collaborative mental health care: a narrative review. Medicine (Baltimore) 2022;101(52):e32554 [FREE Full text] [doi: 10.1097/MD.0000000000032554] [Medline: 36595989]
59. Martel S, Dowedoff P, Marcinow M, Arora R, Kerr J, Shulman R, et al. Patient and family caregiver experience with a collaborative care pilot project to support seniors managing multi-morbidity in a primary care setting. Can J Community Ment Health 2021;40(1):53-66. [doi: 10.7870/cjcmh-2021-003]
60. Mavandadi S, Wray LO, DiFilippo S, Streim J, Oslin D. Evaluation of a telephone-delivered, community-based collaborative care management program for caregivers of older adults with dementia. Am J Geriatr Psychiatry 2017;25(9):1019-1028. [doi: 10.1016/j.jagp.2017.03.015] [Medline: 28433550]
61. Farrand P, Hamilton A, Strickland S. Development of a competency framework for the assistant wellbeing practitioner (Renal) role. J Ren Care 2024;50(1):63-75. [doi: 10.1111/jorc.12449] [Medline: 36461879]
62. Elliott MJ, Donald M, Farragher J, Verdin N, Love S, Manns K, et al. Priorities for peer support delivery among adults living with chronic kidney disease: a patient-oriented consensus workshop. CMAJ Open 2023;11(4):E736-E744 [FREE Full text] [doi: 10.9778/cmajo.20220171] [Medline: 37582622]
63. Wampold BE. How important are the common factors in psychotherapy? An update. World Psychiatry 2015;14(3):270-277 [FREE Full text] [doi: 10.1002/wps.20238] [Medline: 26407772]
64. Lidbetter N, Grigoroglou C, Seccombe N, Taxiarchi VP, Court K, Karachaliou D. Clinician- vs non-clinician-supported eTherapy for anxiety and depression. MHDT 2024;2(1):32-42. [doi: 10.1108/mhdt-04-2024-0011]
65. Al-Haboubi Y, Oladimeji K. Awareness, loneliness, and demand for mental health services in the NHS. BMJ 2022;377:o1178. [doi: 10.1136/bmj.o1178] [Medline: 35537744]
66. Seekles ML, Ormandy P, Coyne E. Mapping the UK renal psychosocial workforce: the first comprehensive workforce survey. BMC Nephrol 2019;20(1):100 [FREE Full text] [doi: 10.1186/s12882-019-1287-0] [Medline: 30898110]
67. Dawson S, Ostrovska B, Treweek S, Gillies K, Witham M, Devane D, et al. Sometimes different, often the same: guidance on how recruitment and retention in trials need to be tailored to enhance participation of ethnic minority groups: a qualitative interview study in England. J Clin Epidemiol 2025;188:112005 [FREE Full text] [doi: 10.1016/j.jclinepi.2025.112005] [Medline: 41038543]

Abbreviations

CBT: cognitive behavioral therapy
CKD: chronic kidney disease
e-MH: e-mental health
SRQR: Standards for Reporting Qualitative Research

Edited by A Kushniruk; submitted 20.Jul.2025; peer-reviewed by C Carswell, A AL-Asadi; comments to author 02.Jan.2026; revised version received 22.Jan.2026; accepted 17.Feb.2026; published 07.Apr.2026.

Please cite as:
Coumoundouros C, Sami A, Farrand P, Sanderman R, von Essen L, Woodford J
e-Mental Health Intervention Preferences Among Informal Caregivers of Adults Living with Chronic Kidney Disease: Semistructured Interview Study
JMIR Hum Factors 2026;13:e80962
URL: https://humanfactors.jmir.org/2026/1/e80962
doi:10.2196/80962
PMID:

©Chelsea Coumoundouros, Abla Sami, Paul Farrand, Robbert Sanderman, Louise von Essen, Joanne Woodford. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 07.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly

XSL•FO
RenderX

cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

A Chronic Pain Self-Management Mobile App (Dolodoc): Cross-Sectional Acceptability Study

Julie Guebey[1], MA; Laetitia Gosetto[1], PhD; Benno Rehberg[2,3], Prof Dr Med; Christian Lovis[1,3], Prof Dr Med; Frederic Ehrler[1], PhD; Aude Molinard-Chenu[1,4], PhD

[1]Division of Medical Information Science, University Hospital of Geneva, Geneva, Switzerland

[2]Anesthesiology, University Hospital of Geneva, Geneva, Switzerland

[3]Medecine Faculty, University of Geneva, Geneva, Switzerland

[4]Division of Consultative Psychiatry and Crisis Intervention, University Hospital of Geneva, Geneva, Switzerland

Corresponding Author:
Julie Guebey, MA
Division of Medical Information Science
University Hospital of Geneva
Boulevard de la Tour, 8
Geneva, 1205 Genève
Switzerland
Phone: 41 22 372 62 88
Email: julie.guebey@hug.ch

Abstract

Background: Approximately 19% of adults in Europe are affected by chronic pain, which reduces the quality of life. Pain-management mobile health (mHealth) apps offer a promising solution for self-management, but user engagement and adherence can limit their clinical impact. User experience design and research play a crucial role in optimizing usability and long-term adoption of digital health interventions.

Objective: This study aims to evaluate the user experience of Dolodoc, a mobile app for chronic pain self-management, using a mixed methods approach that assesses acceptability through a content quality survey and examines use by analyzing overall use patterns.

Methods: A cross-sectional acceptability study of the main content of Dolodoc was conducted among patients with chronic pain recruited from the Geneva University Hospitals pain center and through snowball sampling. Participants rated 84 evidence-based self-management strategies by using a 5-point Likert scale based on 5 acceptability criteria: understandability, motivation, feasibility, relevance, and alignment with the related quality-of-life dimension. To reduce participant fatigue and avoid random responses, the 84 strategies were randomly divided across survey versions. Use was assessed through metrics collected over 6 months with Piwik PRO Analytics to observe real-world use behaviors among Dolodoc users.

Results: In the acceptability study, 33 participants rated the self-management strategies positively across all dimensions. On a scale from –2 to 2, the strategies were well understood (mean 1.47, SD 0.76), motivational (mean 1.12, SD 0.96), feasible (mean 1.01, SD 1.05), relevant (mean 0.99, SD 1.09), and aligned with the dimensions (mean 1.33, SD 0.89). The use study demonstrated that 60% (486/802) of the patients used Dolodoc only once, indicating that long-term adherence remains a challenge. Within Dolodoc, pain tracking, useful links, and medication logging were the most actively used features.

Conclusions: This study highlights the gap between acceptability and long-term adherence to mHealth solutions. Improving personalization and accessibility could increase user engagement and long-term adherence. Future iterations of the app should incorporate tailored interventions and real-time feedback mechanisms. In addition, leveraging a digital navigation follow-up could facilitate user adoption and sustained engagement.

(JMIR Hum Factors 2026;13:e77163) doi:10.2196/77163

KEYWORDS

self-management; chronic pain; user experience; digital health; mobile health; mHealth

XSL•FO
RenderX

Introduction

Chronic pain is a major health care problem that affects about 19% of adults in Europe [1]. This pernicious symptom affects patients worldwide and across many dimensions of their quality of life [2], underscoring the need to develop new approaches to care. Chronic pain often forces individuals to explore multiple approaches to alleviate pain [3]. Managing chronic pain involves an integrative medicine approach based on pharmaceutical or nonpharmaceutical treatments, psychological support, and self-management strategies [4]. Optimizing self-management is paramount in the promotion of self-efficacy [5], which stands out as a powerful lever [6] to improve clinical outcomes related to chronic pain, such as disability, affective distress, and pain severity [7].

In this context, digital mobile health (mHealth) solutions represent a promising sector [8]. There is a growing number of smartphone apps that leverage these innovations to improve the care trajectories of patients with pain [9,10]. In fact, mobile apps for chronic pain management have demonstrated benefits, including improvements in quality of life and pain severity [11,12]. However, we postulate that their effectiveness depends on how well they address the needs and perceptions of the target users. Motivational factors are key to patient engagement in self-management [13] and multimodal treatment [14]. However, long-term engagement and adherence have emerged as key limiting factors for digital interventions in medicine [15,16]. One way to increase adherence and engagement of target users is user experience (UX) design, which aims to clarify the understanding of the human-product interaction, identifying key components and dimensions to improve design practices [17]. It serves as an essential framework at every stage of the design process, aiming to enhance the overall quality of the final product based on the needs of the target user [18]. UX design in health innovations aligns with the contemporary concept of person-centered medicine, which prioritizes patient engagement and individualized care [19,20]. Commonly used methods in UX studies include use metrics analysis, user interviews, usability testing, surveys, and behavioral analytics. These methods can be applied at various stages of the innovation process, ensuring adaptability to different research and design needs. Mixed methods studies facilitate understanding of UX, enabling better adoption and increased effectiveness of digital mHealth tools [21]. Globally, UX studies aim to understand 2 main characteristics of an mHealth solution: acceptability and usability [22].

Dolodoc is a mobile app designed to facilitate self-management of chronic pain. It features pain self-management strategies focusing on 7 quality-of-life dimensions, the ability to add information about the user's pain, and the ability to share a summary of pain-related activities and outcomes with health care professionals [23]. A recent study shows significant gaps between the functionality being developed and end-user expectations, highlighting the need for close collaboration among designers, developers, and users throughout the co-design process [24].

In this study, we describe users' perceptions of acceptability and study their perceptions of the information contained in Dolodoc using 5 selected acceptability criteria based on previous literature: understandability [25], motivation [26], feasibility [26], relevance [27], and alignment with the related quality-of-life dimension. Acceptability is a multifaceted construct reflecting the extent to which individuals delivering or receiving a health intervention consider it appropriate based on their anticipated or experienced cognitive and emotional responses [28]. Usability can be inferred from use studies that examine how systems are used in natural contexts. The aim is to observe and analyze how users interact with a product, service, or technology in their everyday or professional environments, paying particular attention to behaviors, routines, frequency of use, and changes in practice over time. It is mostly used in studies to understand the adoption and effective use of technologies, beyond mere usability performance [29]. In this study, we observe the use habits of real-world users. Overall, this work aims to describe key UX outcomes of an mHealth solution for chronic pain using a mixed methods approach.

Methods

Study Design

This mixed methods cross-sectional study assesses UX outcomes of a chronic pain self-management mobile app, including acceptability through content quality assessments by users and use through global use metrics.

Ethical Considerations

This study was submitted to the Geneva Cantonal Commission for Research Ethics and was deemed to fall outside the scope of the Swiss Federal Act on Research involving Human Beings (Human Research Act); therefore, formal ethics committee approval was not required. All participants were informed about the purpose of the study, and informed written consent was obtained prior to participation. Participation was voluntary, and participants could withdraw at any time without consequence. All data were collected and pseudonymized and handled in accordance with applicable data protection regulations. No personally identifiable information was collected, and confidentiality of participants was strictly maintained. No financial compensation was provided to participants for their participation in this study.

Mobile App Description

Dolodoc is a mobile app developed to assist patients in self-managing chronic pain. It provides self-management strategies and encourages users to maintain a digital journal to dynamically monitor their pain and related quality-of-life indicators. The development and implementation of this digital tool have been described elsewhere [30], and they follow UX design principles for optimization. The core content of the app consists of a library of 84 evidence-based self-management strategies that users can test in real life, tracking their effects within the app on various pain severity indicators, including quality-of-life measures. More details about the app, including illustrative examples, are available in Multimedia Appendix 1 and in previous publications [23,30]. The strategies were

XSL•FO
RenderX

developed through consultations with a focus group of patients with chronic pain and professionals specializing in chronic pain care. Participants first discussed the impact of pain on their daily life in various contexts. From these discussions, 7 quality-of-life dimensions emerged: mood, social support, sleep, work, intimacy, relaxation, and daily activities. Drawing on these dimensions and scientific evidence, a team of psychologists developed motivational pain-management strategies: each quality-of-life dimension corresponds to 4 to 20 strategies, totaling 84 self-management strategies. The self-management strategies were reviewed at the time by a pain specialist and a communication expert.

Participants

Acceptability Study

Participants in the content evaluation survey were eligible for inclusion if they were aged 18 years or older, had chronic pain, were fluent in French, and were able to provide informed consent. Participants were primarily recruited from the pain center at the University Hospital of Geneva, and additional patients meeting the inclusion criteria were recruited through snowball sampling.

There was no requirement to download or use the app to participate, as the study focuses on its content rather than the app itself. It allowed the study to include feedback from experienced users, novices, and nonusers alike. To ensure that each participant fully understood the framework within which the strategies were provided, a standardized contextual presentation of the mobile app was given before participants shared their opinions.

Use Study

As detailed in the following section, the use study included data from anonymous Dolodoc users. These data reflect the behavior of all Dolodoc users throughout the study period, not just the participants who took part in the evaluation interviews.

Setting, Recruitment, and Data Collection

Acceptability Study: Content Quality Assessment by Users

Recruitment took place from June to September 2024. Some participants were recruited passively through flyers in the waiting room, but most were recruited actively by members of the pain center team. All 84 strategies were randomly divided into 5 groups to reduce the total duration of the survey. On the basis of the time of their recruitment, participants were allocated to a random sample of 5 to 22 strategies. This methodological choice was made to reduce participants' burden. In this study, we evaluated the content of the app, not the app itself. Participants were asked to complete an online questionnaire covering demographics, pain characteristics (duration and impact on the 7 quality-of-life dimensions included in Dolodoc), and an evaluation of each sampled strategy. Each strategy was assessed by 5 to 10 participants against 5 acceptability criteria: understandability, motivational impact, feasibility, relevance, and alignment with the relevant quality-of-life dimension (refer to Multimedia Appendix 2 for an example). Responses were collected on a 5-point Likert scale ranging from "strongly disagree" to "strongly agree," allowing for a quantitative understanding of participants' perceptions of each strategy.

Use Study: Real-World Use Metrics

Use data were obtained from the Piwik PRO Analytics Suite (Piwik PRO GmbH), an online usage analytics platform that collects Dolodoc use metrics via anonymous tracking using cookies and session data. Data were collected over 6 months, from January to June 2024.

A cookie (named _pk_id) is collected to determine if a user is recurring and to calculate the number of visits and the duration of each session. Another cookie (named _pk_ses) is used to indicate whether a session is active. The information contained in these cookies is sent to Piwik with each event during the use of Dolodoc (eg, a page view).

The unique identifier from the _pk_id cookies is used as a proxy to identify unique visitors. At the same time, session data are extracted from _pk_ses, including operating system, session number, page title, time spent on the page, and session duration. Although most of our analyses included both iOS and Android users, the number of sessions per visitor could not be calculated for iOS users, likely because on iOS, _pk_id is deleted after 30 minutes of inactivity due to Apple's stricter cookie management. As a result, Piwik always identifies iOS users as unique visitors, preventing us from distinguishing between recurring and unique visitors.

Statistical Methods

Descriptive and inferential statistics were applied depending on the nature of the data. Microsoft Excel, Jupyter Notebook [31], and Orange Data Mining (University of Ljubljana) [32] were used for data analysis and visualization.

Results

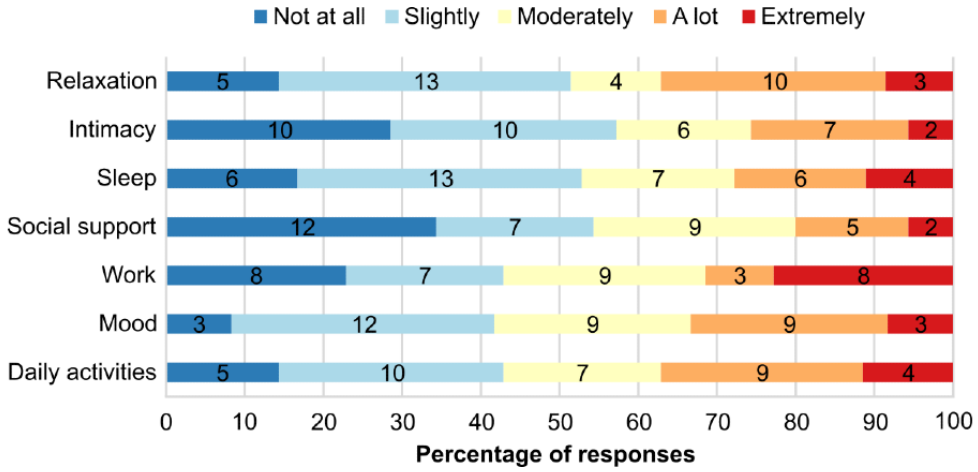
Acceptability Study

Demographics and Pain-Related Characteristics

A total of 33 participants (n=20, 60.6% female and n=13, 39.4% male individuals) were enrolled in the study. A total of 3 (9.1%) participants evaluated 2 different subsamples of strategies; therefore, the total number of assessments per question in the survey is up to 36. For these 3 (9.1%) recurrent participants, we minimized intraindividual variability using only their most recent responses for demographic variables. The participants' ages ranged from 23 to 79 years, with a mean age of 38.1 (SD 15.63) years. Most participants reported chronic pain for more than 1 year (n=24, 72.7%), while 9 (27.3%) participants reported chronic pain for less than 1 year. Each quality-of-life dimension was affected to varying degrees, but overall, quality-of-life dimensions were not homogeneously affected by chronic pain in our sample (χ^2_{24}=21.4; P<.001; Figure 1). These results support the idea of an individual profile in the impact of pain on quality of life.

XSL•FO
RenderX

Figure 1. Impact of pain on the 7 quality-of-life dimensions targeted in the self-management strategies incorporated in the Dolodoc mobile app, as reported by the 33 participants. The number of assessments in each subfraction (35-36 in total) is indicated within the bars; 5 missing values were excluded.



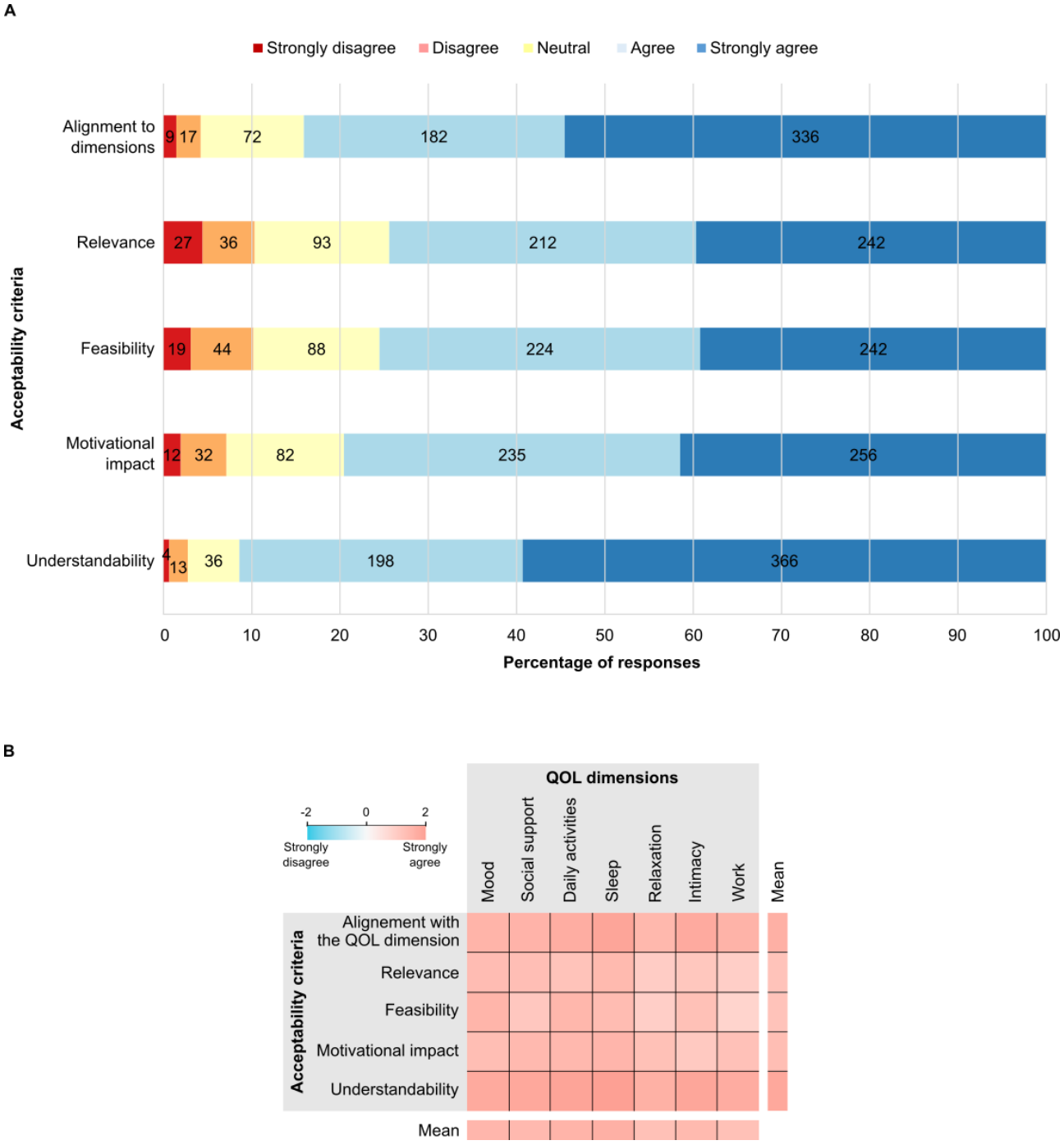
Content Evaluation Survey

Overall, the content of Dolodoc was well perceived by participants: on a Likert scale from –2 to 2 (from strongly disagree to strongly agree), the average rating across all acceptability criteria was 1.18. Moreover, of the 84 strategies, 62 (73.8%) received strong agreement (mean >1), whereas 22 (26.2%) were agreed upon (mean >0), demonstrating that acceptability criteria were globally fulfilled according to our user sample. More specifically, understandability was the highest-rated acceptability criterion (mean 1.47) while relevance (mean 0.99) and feasibility (mean 1.01) received the lowest ratings. The strategies were also perceived as motivational (mean 1.12) and aligned with the specific dimensions (mean 1.33; Figure 2A). Each acceptability criterion received between 610 and 617 ratings. In this study, missing values accounted for 0.26% (8/3085) of the data because the online questionnaire did not require respondents to answer all questions before submitting the survey.

To precisely assess perceptions of the strategies in the context of well-being, we calculated a mean rating for each of the 7 previously described quality-of-life dimensions, based on each acceptability criterion (Figure 2B). The mean scores range between 1 (agree) and 2 (strongly agree), with values from 1.08 for the “relaxation” dimension and 1.40 for the “sleep” dimension. The detailed, numbered results are provided in Multimedia Appendix 3. These results suggest a positive consensus among participants regarding Dolodoc self-management strategies within all quality-of-life dimensions.

XSL•FO
RenderX

Figure 2. Pain-management strategies in Dolodoc are well perceived by users across 5 acceptability criteria and 7 quality-of-life (QOL) dimensions. (A) Bar plot showing the distribution of ratings across 5 acceptability criteria. The number of assessments in each subfraction is indicated within the bars (B) Heat map displaying the average rating score for acceptability criteria and QOL dimensions (n=3077 ratings in total).

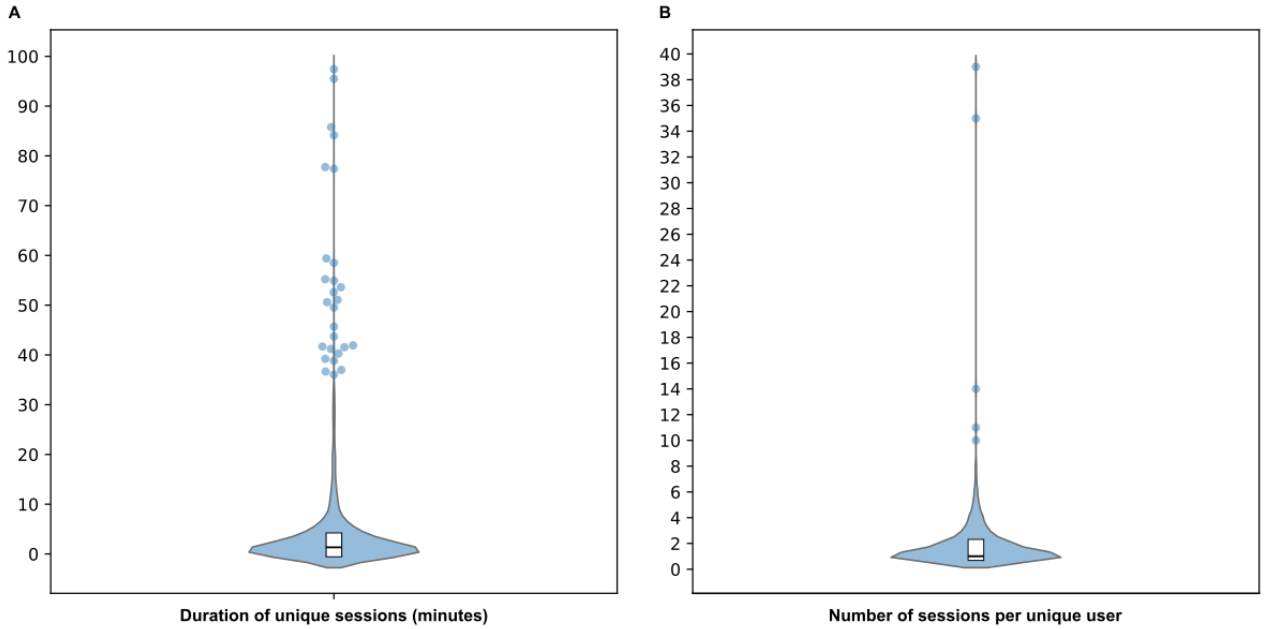


Use Study

Over a 6-month period, from January to June 2024, a total of 3589 unique sessions were recorded, with 59% (2116/3589) on iOS and 41% (1473/3589) on Android. The mean duration of each session was 3 minutes and 21 seconds (201 s; Figure 3A and Multimedia Appendix 4 for percentile distribution). On average, 10.39 page views occurred per session. We observed 0.07% (25/37,255) missing values in the page titles data because these events were not counted as page views. Unique sessions containing only missing values were excluded from analysis (n=7) The bounce rate (ie, the percentage of sessions where only 1 page view occurred) was 23.34% (836/3582).

XSL•FO
RenderX

Figure 3. Analysis of user interaction with the Dolodoc mobile app based on the number of sessions and duration. Violin plots (box plots represent the IQR and the median) show (A) the distribution of the number of sessions per unique user and (B) the duration of each unique session. (A) 1473 sessions from Android users; (B) 3589 sessions from Android and iOS users. Outliers (above 95th percentile) are displayed as individual points. Underlying data are presented in Multimedia Appendix 4.



Among Android users, 1473 sessions were observed among 802 unique visitors. A total of 60.6% (486/802) of these visitors visited the app only once, and the mean number of sessions per visitor was 1.84, although a few outliers displayed more than 10 sessions during this period (Figure 3B and Multimedia Appendix 4 for percentile distribution). As mentioned in the methods section, iOS data were not available for this analysis.

Table 1 shows the top 10 most visited pages for all users. When navigating the app, the most visited page was the "first launch page," showing that a substantial proportion of users used Dolodoc only once. Next, visitors most often viewed the home page and the page where they could monitor quality-of-life dimensions through a 4-point visual scale. Interestingly, the page where the strategies are compiled was the first action users took in the mobile app, because the 3 previous pages are automatically displayed upon connection. These results are consistent with the architecture and purpose of Dolodoc, guiding users toward self-observation and exploration of the included strategies for self-managing chronic pain.

Table 1. Top 10 most visited pages in Dolodoc over 6 months.

Page	Fraction of total views (n=37,230), n (%)	Functionality	Type of user interaction
First launch	9830 (26.4)	Welcome	Passive
Home	5737 (15.41)	Welcome and monitoring	Passive
Quality-of-life dimension evaluation	4333 (11.64)	Monitoring	Passive
Strategies	4312 (11.58)	Management	Active
Report export	2860 (7.68)	Monitoring	Active
Details of activities	2509 (6.74)	Management	Active
User profile	1479 (3.97)	Personalization	Active
Graphical view of collected data	843 (2.26)	Monitoring	Active
History of pain	789 (2.12)	Monitoring	Active
Dashboard	661 (1.78)	Monitoring	Active

Interestingly, the time spent on each page of the app does not follow the same architectural logic as the most visited pages (Table 2). Users spent the most time on less accessible pages (ie, pages that required more clicks to be accessed) of the app, such as a page where they could describe and add characteristics to their pain named "add your pain," a page containing useful links, and a page where they could add their current medication named "add medication." These results showed what users actively sought as they navigated the app, providing valuable feedback and insight into their expectations and needs.

XSL•FO
RenderX

Table 2. Top 10 most engaging pages in the Dolodoc mobile app based on average visit duration over 6 months.

Page	Time spent (s), mean (SD)	App content functionality	Type of user interaction
Add your pain	78.97 (107.23)	Personalization	Active
Useful links	59.70 (209.62)	Education	Active
Add medication	48.15 (62.67)	Personalization	Active
Association	39.65 (147.96)	Social support	Active
Home	39.08 (108.24)	Welcome and monitoring	Passive
Graphical view of collected data	28.89 (102.56)	Monitoring	Active
Create an activity	28.64 (69.30)	Management	Active
Report	23.71 (96.94)	Monitoring	Active
Print report	21.31 (59.46)	Monitoring	Active
Personal information	20.32 (27.91)	Monitoring	Active

Discussion

Key Results

This study provides a detailed evaluation of specific dimensions of UX, focusing on content acceptability and real-world use patterns, rather than on interaction design or interface usability. The population that evaluated the app's content was unevenly affected by pain when assessing its impact on 7 quality-of-life dimensions, highlighting the variability in the experience of chronic pain and the need for personalized approaches to pain management. Self-management strategies included in Dolodoc were perceived as engaging by users, based on 5 acceptability criteria: understandability, motivational aspect, feasibility, relevance, and alignment with quality-of-life dimensions. This result was not straightforward, because although the app was developed using a user-centered design, end users were not involved in developing this content [23]. Although participants in the acceptability study perceived the content as acceptable and engaging, real-world use data showed that 60% (486/802) of users discontinued the app after a single visit. This suggests a potential discrepancy between perceived acceptability in an experimental context and actual acceptability in real-life conditions. However, users were actively engaged with less accessible features that required deeper navigation and/or active data entry, which may explain the longer visit durations. Personal pain and medication descriptions underscore the importance of being active with the app, and the need for personalization to enhance engagement and motivation [27]. Pages offering educational and social support were also among the most visited in the app, underscoring their relevance. Overall, this mixed methods pilot study provides preliminary insights into UX parameters of a mobile app for the self-management of chronic pain, highlighting directions for future research rather than serving as a comprehensive review of the topic.

Limitations

The main limitation of the acceptability study is the relatively small number of patients (N=33), which limits generalizability; therefore, the results must be interpreted as preliminary. Additionally, patients were recruited primarily by health care professionals, which may lead to selection bias, with the most motivated patients participating. Further exploration of the difference between active and passive approaches would have brought valuable information to the study. However, our sample was diverse for demographics and pain-related characteristics.

A major limitation of this study is the lack of qualitative feedback from users who discontinued the app early. While the use analytics indicate that approximately 60% (486/802) of users accessed the app only once, the reasons underlying this early dropout remain unknown. Without direct user feedback, it is not possible to determine whether the discontinuation was due to usability issues, a lack of perceived relevance, insufficient onboarding, or unmet expectations. This limits our ability to fully interpret acceptability in real-world conditions, particularly for users who did not engage beyond an initial session. Taken together, these findings indicate that good content acceptability alone is not sufficient to ensure sustained use or even initial adoption, highlighting the multifaceted nature of UX in mHealth interventions.

Another limitation of the acceptability study is the assessment of tailored acceptability criteria, instead of a standard measurement instrument, such as the user version of the Mobile Application Rating Scale (uMARS) [33]. However, our study focuses on a French-speaking population, and at the time of the study's conception, the uMARS had not been translated into French. As mentioned in the methods section, the main limitation of the use study was excluding iOS users from the number of sessions per visitor calculation. Furthermore, interpreting bounce rate and session data requires caution, as these metrics do not directly reflect user satisfaction or use.

Interpretation

This study highlights the challenges of adopting a pain self-management app and identifies an important gap between acceptability and long-term engagement. This rare combination of real-world and experimental UX observations in the context of chronic pain provides useful, patient-centered insights that could help improve Dolodoc and inspire similar mHealth innovations. Indeed, postproduction UX studies in the context of mHealth solutions for chronic pain are sparse and usually based on small user samples, lacking real-world use data [24,34-36]. Using both ecological and survey-based data helps

XSL•FO
RenderX

mitigate the "trial bias," which has been previously described as an important confounder in user engagement studies of unguided eHealth interventions [34].

Users were satisfied with the content of this mHealth solution, suggesting that the user-centered design that guided its development [23] led to good acceptability of Dolodoc, similar to other mHealth interventions for chronic pain. In comparison, other therapeutic modalities, such as cognitive behavioral therapy [37], mindfulness-based interventions [38], physical activity [39], and multimodal group therapy [40], are also generally well accepted by patients experiencing chronic pain. However, real-world use habits in our study demonstrate the need to enhance long-term engagement further. This study does not identify specific barriers to adoption and sustained use; however, poor long-term engagement has already been described for other mHealth interventions in chronic disorders [16] and chronic pain [41], highlighting important barriers to use, such as the lack of reminders [26,42].

The diversity of pain profiles among our sample of users in the acceptability study, and the fact that global users actively spent most of their time on pages that were dedicated to personalization features, reveal the importance of personalizing pain-management mHealth solutions, in accordance with previous reports [25,26]. The personalization of mHealth apps positively influences users' trust, thereby increasing the likelihood of their adoption [43].

Generalizability

Despite the limitations mentioned, the study provides a strong initial assessment of the acceptability and use of Dolodoc. Although these UX results are specific to this mHealth solution, they align with reports on similar interventions regarding use [44,45] and acceptability [46].

Perspectives

On the basis of the results of this study, it seems relevant to add features to improve engagement over time, such as a reward system and additional features [26]. Additionally, it would be useful to revise the app navigation logic to facilitate access to the most frequently used pages. Future improvements could also help us collect UX data longitudinally by letting users rate content within the app using like or dislike buttons.

Future research should prioritize qualitative investigations to understand the reasons for early discontinuation. Semistructured interviews or focus groups involving users who abandoned the app after initial use could provide valuable insights into barriers related to usability, perceived value, emotional experience, or contextual factors. Such qualitative feedback would complement use analytics and content acceptability measures, enabling a more comprehensive understanding of UX and informing more targeted design improvements.

However, optimizing digital therapeutic alliance [47] remains a key challenge for the future of mHealth. In this regard, digital navigation follow-ups have emerged as a potential solution, grounded in human interactions, to maximize patient engagement with mHealth solutions [48] and increase their clinical impact [49].

Acknowledgments

The authors want to thank Myriam Perrier for her help in navigating the use study results.

Funding

The development of Dolodoc was funded by a grant from the Private Foundation of the Geneva University Hospitals (grant QSP2-03).

Authors' Contributions

Data curation: JG, AM-C
Formal analysis: JG, AM-C
Investigation: JG
Supervision: AM-C, FE, LG, CL
Resources: BR
Visualization: JG, AM-C
Writing – original draft: JG, AM-C, LG
Writing – review & editing: JG, AM-C, FE, LG, CL, BR
Project administration: AM-C

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example of screenshots of the Dolodoc mobile app from a previous study [30].
[DOCX File , 119 KB - humanfactors_v13i1e77163_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Example of a strategy and the evaluation items.
[DOCX File , 14 KB - humanfactors_v13i1e77163_app2.docx]

Multimedia Appendix 3
Mean ratings of adoption criteria (columns) by quality-of-life dimensions (rows).
[DOCX File , 23 KB - humanfactors_v13i1e77163_app3.docx]

Multimedia Appendix 4
Percentile distribution of user interaction.
[XLSX File (Microsoft Excel File), 216 KB - humanfactors_v13i1e77163_app4.xlsx]

References

1. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain 2006 May;10(4):287-333. [doi: 10.1016/j.ejpain.2005.06.009] [Medline: 16095934]
2. Dueñas M, Ojeda B, Salazar A, Mico JA, Failde I. A review of chronic pain impact on patients, their social environment and the health care system. J Pain Res 2016 Jun 28;9:457-467 [FREE Full text] [doi: 10.2147/JPR.S105892] [Medline: 27418853]
3. Takai Y, Yamamoto-Mitani N, Abe Y, Suzuki M. Literature review of pain management for people with chronic pain. Jpn J Nurs Sci 2015 Jul;12(3):167-183. [doi: 10.1111/jjns.12065] [Medline: 25407249]
4. Mills S, Torrance N, Smith BH. Identification and management of chronic pain in primary care: a review. Curr Psychiatry Rep 2016 Feb;18(2):22 [FREE Full text] [doi: 10.1007/s11920-015-0659-9] [Medline: 26820898]
5. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Adv Behav Res Ther 1978;1(4):139-161. [doi: 10.1016/0146-6402(78)90002-4]
6. van Hooff ML, Vriezekolk JE, Kroeze RJ, O'Dowd JK, van Limbeek J, Spruit M. Targeting self-efficacy more important than dysfunctional behavioral cognitions in patients with longstanding chronic low back pain; a longitudinal study. BMC Musculoskelet Disord 2021 Sep 25;22(1):824 [FREE Full text] [doi: 10.1186/s12891-021-04637-3] [Medline: 34563141]
7. Jackson T, Wang Y, Wang Y, Fan H. Self-efficacy and chronic pain outcomes: a meta-analytic review. J Pain 2014 Aug;15(8):800-814 [FREE Full text] [doi: 10.1016/j.jpain.2014.05.002] [Medline: 24878675]
8. Rejula V, Anitha J, Belfin RV, Peter JD. Chronic pain treatment and digital health era-an opinion. Front Public Health 2021 Dec 10;9:779328 [FREE Full text] [doi: 10.3389/fpubh.2021.779328] [Medline: 34957031]
9. Lalloo C, Jibb LA, Rivera J, Agarwal A, Stinson JN. "There's a pain app for that": review of patient-targeted smartphone applications for pain management. Clin J Pain 2015 Jun;31(6):557-563. [doi: 10.1097/AJP.0000000000000171] [Medline: 25370138]
10. Rosser BA, Eccleston C. Smartphone applications for pain management. J Telemed Telecare 2011;17(6):308-312. [doi: 10.1258/jtt.2011.101102] [Medline: 21844177]
11. Thurnheer SE, Gravestock I, Pichierri G, Steurer J, Burgstaller JM. Benefits of mobile apps in pain management: systematic review. JMIR Mhealth Uhealth 2018 Oct 22;6(10):e11231 [FREE Full text] [doi: 10.2196/11231] [Medline: 30348633]
12. Moman RN, Dvorkin J, Pollard EM, Wanderman R, Murad MH, Warner DO, et al. A systematic review and meta-analysis of unguided electronic and mobile health technologies for chronic pain-is it time to start prescribing electronic health applications? Pain Med 2019 Nov 01;20(11):2238-2255. [doi: 10.1093/pm/pnz164] [Medline: 31386151]
13. Jung MJ, Jeong Y. Motivation and self-management behavior of the individuals with chronic low back pain. Orthop Nurs 2016;35(5):330-337. [doi: 10.1097/NOR.0000000000000233] [Medline: 26918776]
14. Stewart JA, Egloff N, von Känel R, Grolimund J, Studer M, Grosse Holtforth M. Motivation for psychological treatment predicts favorable outcomes in multimodal interdisciplinary treatment for chronic somatoform pain. Psychother Psychosom 2017;86(1):60-61 [FREE Full text] [doi: 10.1159/000449471] [Medline: 27884009]
15. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
16. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119 [FREE Full text] [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
17. Berni A, Borgianni Y. Making order in user experience research to support its application in design and beyond. Appl Sci 2021 Jul 29;11(15):6981. [doi: 10.3390/app11156981]
18. M N, Nagaraj A, Gattu H, Shetty PK. Research study on importance of usability testing/ user experience (UX) testing. Int J Comput Sci Mob Comput 2014 Oct;3(10):78-85 [FREE Full text]
19. Sullivan M. The new subjective medicine: taking the patient's point of view on health care and health. Soc Sci Med 2003 Apr;56(7):1595-1604. [doi: 10.1016/s0277-9536(02)00159-4] [Medline: 12614708]

XSL•FO
RenderX

20. NEJM Catalyst. What is patient-centered care? NEJM Catal 2017 [FREE Full text]
21. Shania M, Handayani PW, Asih S. Designing high-fidelity mobile health for depression in Indonesian adolescents using design science research: mixed method approaches. JMIR Form Res 2023 Jul 03;7:e48913 [FREE Full text] [doi: 10.2196/48913] [Medline: 37399059]
22. Lemon C, Huckvale K, Carswell K, Torous J. A narrative review of methods for applying user experience in the design and assessment of mental health smartphone interventions. Int J Technol Assess Health Care 2020;36(1):64-70 [FREE Full text] [doi: 10.1017/S0266462319003507] [Medline: 31973787]
23. Molinard-Chenu A, Ehrler F, Rehberg D, Lovis C. Dolodoc: a virtual pain management coach [Article in French]. Rev Med Suisse 2024 Sep 04;20(885):1568-1573. [doi: 10.53738/REVMED.2024.20.885.1568] [Medline: 39238460]
24. Tong C, Kernoghan A, Lemmon K, Fernandes P, Elliott J, Sacco V, et al. Lessons and reflections from an extended co-design process developing an mHealth app with and for older adults: multiphase, mixed methods study. JMIR Aging 2022 Oct 28;5(4):e39189 [FREE Full text] [doi: 10.2196/39189] [Medline: 36306166]
25. Ledel Solem IK, Varsi C, Eide H, Kristjansdottir OB, Mirkovic J, Børøsund E, et al. Patients' needs and requirements for eHealth pain management interventions: qualitative study. J Med Internet Res 2019 Apr 01;21(4):e13205 [FREE Full text] [doi: 10.2196/13205] [Medline: 30877780]
26. Bostrøm K, Varsi C, Eide H, Børøsund E, Kristjansdottir Ó, Schreurs KM, et al. Engaging with EPIO, a digital pain self-management program: a qualitative study. BMC Health Serv Res 2022 Apr 29;22(1):577 [FREE Full text] [doi: 10.1186/s12913-022-07963-x] [Medline: 35488295]
27. Adams P, Murnane EL, Elfenbein M, Wethington E, Gay G. Supporting the self-management of chronic pain conditions with tailored momentary self-assessments. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17; May 6-11, 2017; Denver, CO. [doi: 10.1145/3025453.3025832]
28. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
29. Ledo D, Houben S, Vermeulen J, Marquardt N, Oehlberg L, Greenberg S. Evaluation strategies for HCI toolkit research. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18; April 21-26, 2018; Montreal, QC URL: https://dl.acm.org/doi/10.1145/3173574.3173610 [doi: 10.1145/3173574.3173610]
30. Ehrler F, Guebey J, Rochat J, Gosetto L, Rehberg B, Lovis C, et al. Dolodoc, an app to leverage self-management of chronic pain: design, development, and implementation report. JMIR Med Inform 2025 Aug 08;13:e71597 [FREE Full text] [doi: 10.2196/71597] [Medline: 40779762]
31. Kluyver T, Ragan-Kelley B, Pérez F, Granger B, Bussonnier M, Frederic J, et al. Jupyter Notebooks—a publishing format for reproducible computational workflows. In: Loizides F, Schmidt B, editors. Positioning and Power in Academic Publishing: Players, Agents and Agendas. Amsterdam, The Netherlands: IOS Press; 2016:87-90.
32. Demsar J, Curk T, Erjave A, Crt Gorup, Hocevar T, Milutinovic M, et al. Orange: data mining toolbox in Python. J Mach Learn Res 2013;14:2349-2353 [FREE Full text]
33. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
34. Liu Y, Encelle B, Sehaba K. A user-centered approach to design a mobile application for chronic pain management. Model Meas Control C 2020 Dec;81(1-4):24-29. [doi: 10.18280/mmc_c.811-405]
35. Reynoldson C, Stones C, Allsop M, Gardner P, Bennett MI, Closs SJ, et al. Assessing the quality and usability of smartphone apps for pain self-management. Pain Med 2014 Jun;15(6):898-909. [doi: 10.1111/pme.12327] [Medline: 24422990]
36. Shah UE, Chiew TK. A systematic literature review of the design approach and usability evaluation of the pain management mobile applications. Symmetry 2019 Mar 19;11(3):400. [doi: 10.3390/sym11030400]
37. Bernardy K, Klose P, Welsch P, Häuser W. Efficacy, acceptability and safety of cognitive behavioural therapies in fibromyalgia syndrome - a systematic review and meta-analysis of randomized controlled trials. Eur J Pain 2018 Feb;22(2):242-260. [doi: 10.1002/ejp.1121] [Medline: 28984402]
38. Ruskin DA, Gagnon MM, Kohut SA, Stinson JN, Walker KS. A mindfulness program adapted for adolescents with chronic pain: feasibility, acceptability, and initial outcomes. Clin J Pain 2017 Nov;33(11):1019-1029. [doi: 10.1097/AJP.0000000000000490] [Medline: 28328699]
39. van Doormaal MC, Meerhoff GA, Vliet Vlieland TP, Peter WF. A clinical practice guideline for physical therapy in patients with hip or knee osteoarthritis. Musculoskelet Care 2020 Jul 09;18(4):575-595. [doi: 10.1002/msc.1492]
40. Ochandorena-Acha M, Dalmau-Roig A, Dürsteler C, Vilchez-Oya F, Ferrer A, Martin-Villalba I, et al. Acceptability of multimodal and multidisciplinary group-based program for chronic low back pain: a qualitative study. Physiother Theory Pract 2025 May;41(5):981-997. [doi: 10.1080/09593985.2024.2377343] [Medline: 38994708]
41. Forbes G, Newton S, Cantalapiedra Calvete C, Birch J, Dodds J, Steed L, et al. MEMPHIS: a smartphone app using psychological approaches for women with chronic pelvic pain presenting to gynaecology clinics: a randomised feasibility trial. BMJ Open 2020 Mar 12;10(3):e030164 [FREE Full text] [doi: 10.1136/bmjopen-2019-030164] [Medline: 32165549]

XSL•FO
RenderX

42. Kristjánsdóttir Ó, Fors EA, Eide E, Finset A, van Dulmen S, Wigers SH, et al. Written online situational feedback via mobile phone to support self-management of chronic widespread pain: a usability study of a web-based intervention. BMC Musculoskelet Disord 2011 Feb 25;12:51 [FREE Full text] [doi: 10.1186/1471-2474-12-51] [Medline: 21352516]
43. Guo X, Zhang X, Sun Y. The privacy–personalization paradox in mHealth services acceptance of different age groups. Electron Commer Res Appl 2016;16:55-65. [doi: 10.1016/j.elerap.2015.11.001]
44. Bostrøm K, Børøsund E, Eide H, Varsi C, Kristjansdottir Ó, Schreurs KM, et al. Short-term findings from testing EPIO, a digital self-management program for people living with chronic pain: randomized controlled trial. J Med Internet Res 2023 Aug 25;25:e47284 [FREE Full text] [doi: 10.2196/47284] [Medline: 37624622]
45. Linardon J. Rates of attrition and engagement in randomized controlled trials of mindfulness apps: systematic review and meta-analysis. Behav Res Ther 2023 Nov;170:104421 [FREE Full text] [doi: 10.1016/j.brat.2023.104421] [Medline: 37862854]
46. Shi JL, Sit RW. Impact of 25 years of mobile health tools for pain management in patients with chronic musculoskeletal pain: systematic review. J Med Internet Res 2024 Aug 16;26:e59358 [FREE Full text] [doi: 10.2196/59358] [Medline: 39150748]
47. Henson P, Wisniewski H, Hollis C, Keshavan M, Torous J. Digital mental health apps and the therapeutic alliance: initial review. BJPsych Open 2019 Jan;5(1):e15 [FREE Full text] [doi: 10.1192/bjo.2018.86] [Medline: 30762511]
48. Wisniewski H, Torous J. Digital navigators to implement smartphone and digital tools in care. Acta Psychiatr Scand 2020 Apr;141(4):350-355 [FREE Full text] [doi: 10.1111/acps.13149] [Medline: 31930477]
49. Camacho E, Chang SM, Currey D, Torous J. The impact of guided versus supportive coaching on mental health app engagement and clinical outcomes. Health Informatics J 2023;29(4):14604582231215872 [FREE Full text] [doi: 10.1177/14604582231215872] [Medline: 38112116]

Abbreviations

mHealth: mobile health
uMARS: user version of the Mobile Application Rating Scale
UX: user experience

Edited by A Parush; submitted 08.May.2025; peer-reviewed by J Katz, R Marshall, M Betrancourt; comments to author 14.Aug.2025; revised version received 02.Feb.2026; accepted 09.Feb.2026; published 14.May.2026.

Please cite as:
Guebey J, Gosetto L, Rehberg B, Lovis C, Ehrler F, Molinard-Chenu A
A Chronic Pain Self-Management Mobile App (Dolodoc): Cross-Sectional Acceptability Study
JMIR Hum Factors 2026;13:e77163
URL: https://humanfactors.jmir.org/2026/1/e77163
doi:10.2196/77163
PMID:

©Julie Guebey, Laetitia Gosetto, Benno Rehberg, Christian Lovis, Frederic Ehrler, Aude Molinard-Chenu. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 14.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Applying the Information-Motivation-Behavioral Skills Model to a Virtual Reality–Assisted Cognitive Health Education Intervention to Promote Screening Uptake: Qualitative Study

Tingting Yin[1*], PhD; Yulei Song[1*], PhD; Bingqing Yang[1], MS; Jingjing Yan[1], MS; Yan Wang[1], MS; Haiyan Yin[1], PhD; Xueqing Zhang[1], MS; Guihua Xu[1], PhD; Yamei Bai[1], PhD

School of Nursing, Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China

[*]these authors contributed equally

Corresponding Author:
Tingting Yin, PhD
School of Nursing
Nanjing University of Chinese Medicine
138 Xianlin Avenue
Nanjing, Jiangsu,
China
Phone: 86 18851835056
Email: yori1998@163.com

Abstract

Background: Uptake of mild cognitive impairment (MCI) and dementia screening among community-dwelling older adults remains suboptimal globally and is shaped by limited dementia literacy, fear, stigma, and difficulties navigating services. Virtual reality (VR)–assisted education may enhance engagement, but older adults' experiences of such education and the contextual conditions shaping screening intentions require further understanding.

Objective: This study aims to explore community-dwelling older adults' understanding, concerns, and attitudes regarding MCI and screening after exposure to a community-based VR-assisted cognitive health education intervention and to examine their perceived information, motivation, and behavioral skills related to screening using the information-motivation-behavioral skills (IMB) model.

Methods: Twenty-two community-dwelling older adults were recruited from participating communities between August 23 and September 17, 2025, using purposive sampling with attention to variation in age and educational attainment, supplemented by snowball sampling. All participants completed a 40-minute VR-assisted cognitive health education session comprising a 10-minute immersive 360° video delivered through a Meta Quest 3 head-mounted display followed by 30 minutes of nurse-led face-to-face education. Four researchers coded the original Mandarin transcripts; themes and subthemes were derived inductively from the data and subsequently mapped to the IMB model.

Results: Sixteen data-derived subthemes were subsequently mapped to the 3 domains of the IMB model. Information-related accounts encompassed misunderstanding and lack of knowledge about MCI, the relationship between MCI and dementia, knowledge acquisition and clarification, timing and psychological readiness, information overload, attention to peripheral details, and uncertain awareness of cognitive health status. Motivation-related accounts included fear and stigma associated with screening, perceived reduction in stigma and greater screening acceptability, reconciliation with aging, low self-prioritization, and a pessimistic outlook on treatment. Behavioral-skills accounts included low confidence in self-management, knowledge to action, community-based referral and navigation, and difficulty changing established habits. Across domains, participants' accounts suggested that greater understanding and screening acceptability did not necessarily translate into action when practical navigation and referral support were lacking.

Conclusions: Participants perceived VR-assisted cognitive health education as engaging and helpful for making MCI-related information more concrete, but their accounts also indicated that screening intentions may not translate into action without clear, trusted navigation and referral support. These findings describe participants' experiences and perceived barriers and facilitators rather than estimating intervention effectiveness.

(JMIR Hum Factors 2026;13:e91657) doi:10.2196/91657

XSL•FO
RenderX

KEYWORDS

mild cognitive impairment; dementia screening; virtual reality; IMB model; community nursing; health service navigation; qualitative research; information-motivation-behavioral skills model

Introduction

Dementia is a leading cause of disability and dependency in older adults and poses growing challenges for health systems worldwide as populations age [1,2]. In China, rapid demographic aging, large absolute numbers of older people, and uneven access to specialist services make cognitive impairment a major public health and nursing priority [3]. Mild cognitive impairment (MCI) represents an intermediate clinical state between normal aging and dementia, characterized by objective cognitive decline with relatively preserved functional independence, and it is widely regarded as an important window for early identification, risk reduction, and timely support [4,5]. Although disease-modifying treatment development has faced substantial setbacks and methodological challenges [6-8], prevention and risk reduction remain central components of dementia care [4,9]. Timely recognition may enable individuals and families to plan ahead, access services, and consider appropriate interventions and support [10]. However, the potential benefits of early detection depend on whether older adults are willing and able to engage with screening and follow-up services [10,11].

Despite increasing attention to cognitive health, dementia screening hesitancy remains common [11], and dementia may remain undetected in primary care [12]. Prior studies have identified multiple barriers, including limited dementia literacy [13,14], uncertainty about dementia risk reduction and assessment [13], fear of receiving a diagnosis and the psychological consequences of a high-risk result [15], and shame or anticipated stigma associated with dementia [15,16]. In addition, many older adults interpret memory lapses as a normal part of aging and therefore not warranting assessment [17]. These barriers involve not only information but also motivation, identity, family considerations, and social norms; system-level constraints and communication problems can further delay recognition and diagnosis [18].

Community-based primary care has a pivotal role in addressing these barriers. In China, community health centers are a core component of the primary health care system, with responsibilities for prevention and chronic disease management [19,20]. Primary care is also increasingly recognized as an important setting for early detection of cognitive impairment in older adults [21]. Strengthening coordination between primary care institutions and hospitals is a recognized priority in China [20], providing a structural basis for referral and navigation support when further cognitive assessment is needed.

Immersive virtual reality (VR) has been applied to cognitive assessment [22], dementia-related nursing interventions [23], and rehabilitation research [24], while systematic reviews have examined its health applications among community-dwelling older adults [25]. In dementia education specifically, a randomized trial in Japan evaluated a multielement VR program combining simulation and VR films with short films, lectures, and discussion; the intervention improved intention to help people with dementia, although the primary attitude outcome did not differ significantly between groups [26]. More recent work in older adults with MCI has also reported improvements in dementia knowledge, critical health literacy, and self-efficacy following a VR-based dementia prevention program [27]. These findings support the potential of VR as an educational format, but they do not explain how older adults experience VR-assisted cognitive health education in community settings or how such experiences intersect with screening decisions. The literature on immersive VR in community-dwelling older adults has also highlighted the need for further study of usability and acceptability [25]. In the Chinese context, qualitative evidence remains limited regarding how older adults perceive VR-assisted cognitive health education, how fear or stigma shape screening readiness, and what practical service-navigation barriers may persist after education. Clarifying these perceived processes and contextual conditions may help inform screening-promotion interventions that can be integrated with routine community care.

Behavior change theory provides a useful lens for understanding how educational interventions translate into health actions. The information-motivation-behavioral skills (IMB) model [28] proposes that people are more likely to adopt a health behavior when they (1) possess relevant and usable information, (2) are sufficiently motivated at personal and social levels, and (3) have the behavioral skills and self-efficacy to act, including the practical ability to navigate services. This model has been applied to a range of preventive and self-management behaviors [29-31] and is particularly relevant to screening contexts, where knowledge and willingness may be present but follow-through fails because the “next step” is unclear or difficult. Applying IMB to cognitive screening in community settings can clarify which components of an intervention are most salient for older adults and what additional nursing or system supports are needed to reduce the gap between intention and action.

Against this background, this study explored a community-based VR-assisted cognitive health education intervention. The aims of this qualitative study were to (1) describe community-dwelling older adults’ understanding, concerns, and attitudes regarding MCI and screening after exposure to VR-assisted education; (2) examine participants’ perceived information, motivation, and behavioral skills relevant to screening, guided by the IMB model; and (3) identify implications for community nursing practice and service processes that may support screening uptake.

Methods

Study Design

This study used a descriptive qualitative design to explore community-dwelling older adults’ experiences and perceptions of VR-assisted cognitive health education in relation to MCI screening. The qualitative component was initiated during the recruitment and health education phase of a broader

XSL•FO
RenderX

community-based cluster-randomized study after the research team observed that some older adults were hesitant to participate in cognitive screening. It focused on participants' accounts, meanings, and contextual factors and was not designed to test intervention effectiveness or estimate causal effects on screening uptake [32,33]. Reporting was guided by the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist [34] (Multimedia Appendix 1).

Broader Cluster-Randomized Study and VR-Assisted Cognitive Health Education

The broader community-based cluster-randomized study comprised 3 study arms: VR-based adaptive-difficulty cognitive training, VR-based fixed-difficulty cognitive training, and a control condition without additional cognitive training. The community was the unit of randomization, and the 9 participating communities were allocated to the 3 study arms using a random-number-table method. A total of 178 older adults who met the parent-trial eligibility criteria, including confirmation of MCI, were formally enrolled in the 3 randomized trial groups. The shared VR-assisted cognitive health education component was provided across the study conditions, whereas the groups differed in the subsequent cognitive training component. The present qualitative sample was recruited earlier, during the common recruitment and health education phase, and participation in the qualitative interviews did not require a confirmed MCI diagnosis or depend on subsequent randomized-group allocation.

The shared VR-assisted cognitive health education session lasted approximately 40 minutes and comprised 2 components. First, participants viewed a 10-minute Chinese-language immersive 360° video using a Meta Quest 3 head-mounted display. Viewing was individual and seated. The experience was observational rather than interactive: participants could explore the 360° visual environment through head movements but did not manipulate virtual objects. The video presented first-person scenarios intended to depict cognitive change, including a familiar living room transforming into a floating island and the intermittent appearance of relatives' faces in a "memory maze." A local person living with dementia and a family member also described experiences of early screening, result notification, diagnosis, and treatment. Chinese subtitles were provided. Immediately after the VR session, participants received a 30-minute face-to-face educational session delivered by a nurse, who provided additional information about MCI and dementia and explained cognitive screening and the screening process.

Semistructured interviews were conducted with 22 older adults who had completed the shared VR-assisted cognitive health education session. The qualitative study specifically examined participants' experiences of this common educational component and was not designed to compare experiences or outcomes across the randomized cognitive training groups. The interviews explored participants' prior experiences of cognitive health education, reactions to the VR format, understanding and concerns regarding MCI and dementia, and views about screening and follow-through. The interview guide (Table 1) was informed by the IMB model and covered information-related understanding and misconceptions, motivational factors such as fear and stigma, and behavioral skills such as confidence and service navigation.

XSL•FO
RenderX

Table 1. Semistructured interview guide questions informed by the information-motivation-behavioral skills (IMB) model.

Number	Interview question	Core IMB constructs	Relevance to screening uptake or intervention experience
1	Before viewing the VR[a] session, what did "MCI"[b] and "dementia" mean to you? How did you interpret memory changes as "normal aging" versus "a problem"?	MCI or dementia literacy; symptom appraisal	Probing baseline misconceptions that shape perceived need for screening
2	Before today, what did you know about cognitive screening (what it involves, where it happens, and what results mean)?	Screening process knowledge; perceived consequences	Identifying knowledge gaps that create uncertainty and avoidance of screening
3	How do you usually experience community health talks or printed materials about cognitive health (eg, terminology, pace, and amount of information)?	Information accessibility; cognitive load	Examining barriers in conventional education that limit comprehension and engagement
4	After viewing the VR session, what information became clearer for you, and what (if anything) remained confusing?	Information acquisition; message clarity	Exploring whether participants perceived the VR session as clarifying information for decision-making
5	How acceptable was the VR format for you (eg, comfort, attention, and distraction by novelty or details)?	Information processing conditions; intervention acceptability	Clarifying delivery-related conditions that may strengthen or dilute key messages
6	When you think about cognitive screening, what feelings or concerns arise (eg, fear, being labeled, stigma, and "loss of face")?	Stigma; anticipated social judgment; affective motivation	Exploring how stigma-related emotions may influence willingness to initiate screening
7	How do you view the value of screening and early action (eg, "knowing is useful" vs "nothing can be done," treatment pessimism, fatalism)?	Outcome expectations; response efficacy; fatalism	Exploring how participants described the perceived value and manageability of screening and early action
8	Do responsibilities and self- prioritization influence your decision (eg, "not worth spending time or money on myself," caregiving duties, and reconciliation with aging)?	Self-worth; aging narratives; competing priorities	Uncovering identity- and values-based barriers that persist despite improved knowledge
9	If you decided to be screened, what would you do first (where to go, which department, how to register), and what would be the hardest step?	Service navigation skills; self-efficacy	Mapping the intention-action gap into concrete capability and system-friction points
10	What support would make it easier to complete screening (eg, community referral, appointment help, reminders, and accompaniment)?	Social support; action planning; enabling resources	Identifying implementation supports participants considered helpful for translating readiness into screening action

[a]VR: virtual reality.

[b]MCI: mild cognitive impairment.

Participants and Recruitment

Participants were recruited from 9 participating communities in Nanjing, Jiangsu Province, China, between August 23 and September 17, 2025, using purposive sampling supplemented by snowball sampling. Community staff posted standardized recruitment notices in community WeChat (Tencent Holdings Limited) group chats (WeChat is a widely used messaging and social media platform in China), including brief information about the study and the research team's contact details. Interested older adults contacted the researchers directly and voluntarily. Among eligible individuals who expressed interest, the research team purposively selected participants with attention to variation in age and educational attainment to capture a range of perspectives. Snowball sampling was used as a supplementary strategy, whereby some participants learned about the study through older adults who had already participated. Eligibility criteria for the qualitative study differed from those of the parent trial because the qualitative study focused on experiences of cognitive health education and screening rather than cognitive-training outcomes among people with confirmed MCI. Eligibility criteria for the qualitative study were an age of 60-85 years, completion of the shared VR-assisted cognitive health education session, the ability to communicate sufficiently to participate in an in-depth interview, and willingness to discuss perceptions and experiences concerning MCI, dementia, and cognitive screening. A confirmed MCI diagnosis was not required. Individuals who had previously undergone MCI or other cognitive impairment screening at another institution and medical professionals were excluded. Participants provided written informed consent before the interview and received a small gift worth approximately CNY 10 (US $1.50) after participation.

XSL•FO
RenderX

Data Collection and Analysis

Before each interview, the researcher collected participant characteristics including age, retirement status, marital status, educational level, family history of dementia, and prior MCI screening. The semistructured interview guide was developed in Mandarin Chinese and pilot-tested with 3 older adults, who were included in the final sample of 22. No substantive changes were made to the interview guide after pilot testing; therefore, the 3 pilot interviews were retained in the final analytic sample. During the pilot stage, interview notes were reviewed with the 3 pilot participants to confirm that the recorded content accurately reflected their intended meaning. Interviews were conducted in Mandarin by TY, a researcher trained in qualitative methods who was involved in volunteer recruitment and intervention delivery in the broader study. Before the qualitative interviews, TY had prior study-related contact with 7 of the 22 participants; she had no prior relationship with the remaining 15. All interviews were conducted face-to-face in private, quiet indoor rooms provided by the participating communities, with only the interviewer and participant present. Interviews lasted 39-152 minutes. The variation reflected differences in the depth and richness of participants' responses; some participants provided particularly detailed and reflective accounts, leading to additional probing and clarification and consequently longer interviews. With participants' consent, interviews were audio-recorded and transcribed verbatim in Mandarin within 24 hours after each interview. Field notes were recorded during and immediately after the interviews to document contextual observations, relevant nonverbal responses, and issues requiring clarification; these notes were considered alongside the transcripts during coding and interpretation. Thematic saturation was initially judged to have been reached after the 19th interview. Three additional interviews were conducted to confirm saturation, and no new themes emerged in interviews 20-22. Full verbatim transcripts from the 22 formal interviews were not returned to participants for checking.

Coding, theme development, and interpretation were conducted using the original Mandarin transcripts. All 4 researchers (TY, BY, JY, and YW) coded the full set of 22 transcripts using Microsoft Word and NVivo (version 15; Lumivero). Codes, subthemes, and themes were developed iteratively from the interview data. Coding decisions and emerging interpretations were compared regularly among the 4 researchers; disagreements regarding coding, code definitions, theme boundaries, or interpretation were discussed until consensus was reached. Related codes were grouped into candidate subthemes, which were reviewed against the coded extracts and original transcripts and refined until the research team agreed on the final thematic structure. After the subthemes had been developed from the data, they were mapped to the IMB model domains according to their predominant conceptual meaning. Thus, the IMB model served as a theoretical framework for organizing and interpreting the final thematic structure rather than as a rigid set of predetermined themes. English translations were prepared for the participant quotations selected for reporting after the qualitative analysis had been completed.

Patient and Public Involvement

Patient and public involvement occurred at different stages of the broader research program and was separate from the 22 semistructured interviews reported in this study. First, formative qualitative work was conducted before this interview study to characterize community perceptions and stigma related to MCI and dementia; findings informed development of the VR-assisted cognitive health education content. During intervention development, community-dwelling older adults and community staff viewed the VR educational materials and provided feedback on the content and presentation. Their comments were considered by the research team when refining the materials before implementation. During recruitment for the present qualitative study, community health coordinators supported logistical organization and dissemination of recruitment information.

Ethical Considerations

This study was approved by the Ethics Committee of the First Affiliated Hospital of Nanjing Medical University (number 2025-SR-015) and conducted in accordance with the Declaration of Helsinki [35]. Participants provided written informed consent before the semistructured interviews and received the stated compensation after participation.

Results

Characteristics of Participants

Twenty-two community-dwelling older adults from participating communities participated in the interviews. Their mean age was 73.27 (SD 4.95) years, and 15 (68.18%) were women. Most participants had completed secondary education or higher (Table 2). The relationship between the qualitative interview sample and the formally enrolled parent cluster-randomized trial sample is summarized in Table S1 in Multimedia Appendix 2. The two samples were recruited at different stages and under different eligibility criteria and were not statistically compared.

XSL•FO
RenderX

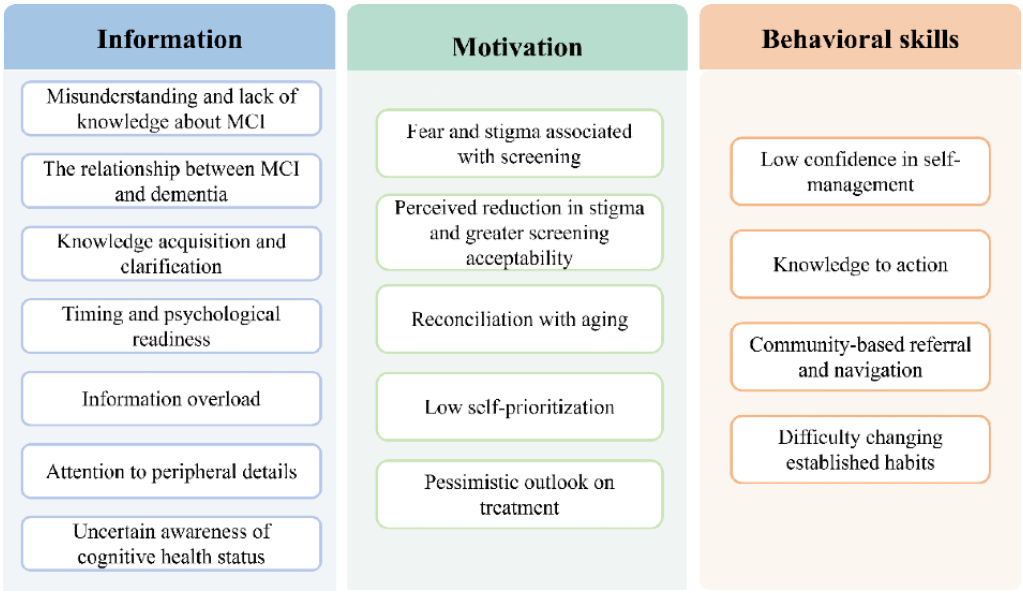
Table 2. Characteristics of participants in the semistructured interviews (N=22).

Demographic information	Value
Age (years), mean (SD)	73.27 (4.95)
Sex, n (%)	
Male	7 (31.82)
Female	15 (68.18)
Marital status, n (%)	
Single	1 (4.54)
Married	20 (90.92)
Widowed	1 (4.54)
Education level, n (%)	
Primary or less	1 (4.54)
Middle school	8 (36.36)
High school	9 (40.92)
College and above	4 (18.18)
Retirement status, n (%)	
No	1 (4.54)
Yes	21 (95.45)
Family history of dementia, n (%)	
No	19 (86.36)
Yes	3 (13.64)

Themes, Subthemes, and Codes

Participants described varied prior understanding of MCI and dementia, concerns about screening, and experiences of the VR-assisted education session. Several participants perceived the VR content as making concepts easier to understand or less threatening, whereas others described information overload, distraction by delivery details, or persistent uncertainty. Across interviews, participants also emphasized practical barriers to acting on screening intentions, particularly uncertainty about where to go and how to obtain referral support. Sixteen subthemes were developed from the interview data and subsequently mapped to the 3 IMB domains (Figure 1).

Figure 1. Subthemes mapped to the information-motivation-behavioral skills (IMB) model across 22 semistructured interviews. MCI: mild cognitive impairment.



XSL•FO
RenderX

Information

Misunderstanding and Lack of Knowledge About MCI

Participants commonly described limited prior understanding of MCI, often equating memory problems with "dementia" or "mental problems." Their accounts indicated that these misconceptions shaped how they interpreted the value and possible consequences of screening:

> *I always thought poor memory was senile dementia and there's no cure for it, so I didn't want to get screened.* [N14, female, 68 years old]

Participants' accounts also reflected uncertainty about the nature of MCI and what, if anything, could be done after early identification. Some participants linked screening to additional worry or family burden:

> *In our village, they all say it's the "confused disease," and getting screened won't help anyway; it just makes the children worried.* [N21, female, 76 years old]

The Relationship Between MCI and Dementia

Participants frequently described uncertainty about the distinction and relationship between MCI and dementia, and some interpreted MCI as inevitably progressing to dementia:

> *The VR said it's not dementia, but it's the early stage of dementia. So wouldn't that just be a matter of time? Hearing this made me even more scared.* [N12, male, 68 years old]

> *Now I understand. It's like the relationship between high blood pressure and cerebral infarction. If it's managed early, it might not progress to that stage.* [N15, male, 71 years old]

Knowledge Acquisition and Clarification

Several participants perceived the visual and immersive presentation as easier to understand than professional terminology or dense written materials and described the session as helping them clarify previously confusing concepts about MCI and dementia:

> *Before, just listening to the terms, I couldn't understand. The animation in the VR clearly demonstrated "aging of brain circuits" and "complete disconnection." Now I really understand.* [N5, male, 67 years old]

Some participants also valued the individually delivered format and described it as allowing them to engage with the information at a more comfortable pace, without the social discomfort they associated with asking questions during face-to-face health education:

> *When the doctor was explaining, I was too embarrassed to ask. With this VR, I can watch it slowly and even go back to review it. Finally, I understood that "forgetfulness" and "pathological forgetfulness" are different.* [N22, male, 71 years old]

Timing and Psychological Readiness

Participants described timing and psychological readiness as important to how cognitive-health messages were received. When education occurred during periods of anxiety or poor physical health, some participants interpreted screening suggestions defensively or negatively:

> *During that time, I wasn't feeling well physically either. Anyone who suggested I have my brain checked, I thought it was cursing me.* [N4, female, 79 years old]

Participants also described greater receptiveness at certain life transitions, such as retirement, suggesting that readiness to engage with prevention messages varied by personal context:

> *Before retirement, I never thought about these things. When I just retired, I was nervous. At this time, when someone told me about prevention, I could listen to them.* [N20, female, 65 years old]

Information Overload

Participants described conventional health-education materials and expert talks as difficult to follow when information was dense, highly technical, or delivered too quickly. These experiences were especially salient in accounts of frustration or disengagement:

> *With this VR, it feels like there's a patient teacher guiding me step by step. Today, I'll understand how the heart beats, and tomorrow I'll learn about the blood vessels. This way of learning is not tiring, and I have a clear understanding. As a result, everything is firmly imprinted in my mind.* [N8, female, 75 years old]

Participants' accounts suggested a preference for information delivered in smaller, layered segments rather than large amounts of technical content presented at once:

> *They gave me a booklet with densely packed text and many English abbreviations. After reading it for a while, I got a headache and threw it aside directly.* [N1, female, 77 years old]

> *The experts spoke too fast, with one professional term after another. I didn't even have time to take notes and after listening, I became even more confused.* [N10, male, 79 years old]

Attention to Peripheral Details

Some participants reported that the novelty of VR or peripheral visual and operational details initially drew attention away from the core educational message. Questions about image accuracy or the realism of examples sometimes became a focus of discussion:

> *Why is the color of the brain model in the VR program pink? I heard it should be gray. Could this be unscientific?* [N7, female, 69 years old]

For these participants, attention to non–core details appeared to compete with attention to the intended message about MCI and screening:

XSL•FO
RenderX

> *The example given is that buying vegetables and making a mistake with the money. I think this example is not good. Our generation's mental arithmetic is not as good as that of the young people.* [N6, female, 79 years old]

Uncertain Awareness of Cognitive Health Status

Participants often described uncertainty in appraising their own cognitive health, moving between interpretations of memory change as "normal aging" and as a possible problem. Some accounts combined worry with avoidance, including a preference not to know whether a problem was present:

> *I just occasionally forget things. Everyone does this. There's no need to specifically check; finding out would only add to my worries.* [N16, female, 75 years old]

> *Sometimes I feel my reaction is slower, but I'm afraid it's really a problem. It's better not to know.* [N13, female, 71 years old]

Motivation

Fear and Stigma Associated With Screening

Participants described screening as carrying negative social meanings, including being labeled, diagnosed, or viewed as "abnormal." These anticipated consequences contributed to fear and reluctance in some accounts:

> *I don't want to be screened. What if something goes wrong during the test, will others think I'm "abnormal"? Even my grandson might not let me do it.* [N2, female, 71 years old]

Some participants described avoiding cognitive screening because they feared discovering a problem or being judged by family, peers, or the wider community. Their accounts positioned stigma as a social and emotional barrier to screening:

> *I don't go for this screening test even when my unit organizes it. I'm afraid it will affect my reputation and be said that "his mind is not good."* [N3, male, 82 years old]

Perceived Reduction in Stigma and Greater Screening Acceptability

Some participants reported that the VR session made cognitive decline feel less shameful and screening more acceptable, particularly when the content framed cognitive change as a health issue that could be discussed and managed.

A man who had previously refused screening said after the session:

> *Previously, I thought being diagnosed meant being "mentally retarded." Now I understand that it's just a risk condition like high blood pressure, and early detection can also be controlled. I don't fear being laughed at by others anymore.* [N18, male, 81 years old]

Participants who identified with the community-based cases described less embarrassment and greater openness to considering screening. These accounts suggest a perceived reduction in stigma rather than a measured change in screening behavior:

> *The cases in the VR were all normal. They were like neighbors to me, making me feel that having such a condition is not embarrassing. I should get checked if there's a need.* [N11, female, 69 years old]

Reconciliation With Aging

Participants sometimes normalized cognitive decline as an inevitable part of aging. For some, this interpretation functioned as an adaptive way of maintaining calm and self-identity, but it also reduced the perceived need to seek assessment for memory changes:

> *As one gets older, the machine becomes old and the parts don't work properly. This is normal. Why bother?* [N3, male, 82 years old]

Participants' accounts indicated that acceptance of aging could coexist with reduced motivation to pursue screening, particularly when memory changes were interpreted as universal and unavoidable:

> *Having a poor memory also has its advantages. Not being able to remember the troubles, life becomes simpler.* [N9, female, 73 years old]

Low Self-Prioritization

Some participants placed their own health needs behind family responsibilities and described screening as not worth the time or expense. These accounts reflected low self-prioritization rather than lack of awareness alone.

A rural woman whispered:

> *It's not easy for the children to earn money. I won't bother to have my health checked. It's a waste of money.* [N1, female, 77 years old]

Another man who had been caring for his ill spouse for a long time said:

> *My job is to take good care of my wife. I don't care about my own situation.* [N10, male, 79 years old]

These accounts suggest that prioritizing family responsibilities over personal health could reduce motivation to seek screening.

Pessimistic Outlook on Treatment

Several participants expressed pessimistic expectations about the value of early identification or treatment for MCI and dementia. For some, the belief that progression was inevitable reduced the perceived benefit of screening:

> *I've seen what dementia is like in its advanced stage. Currently, there is simply no cure. Once diagnosed, there's only waiting.* [N20, female, 65 years old]

> *Even with early intervention, it's only a delay. Eventually, it will still progress. This awareness is actually a burden.* [N14, female, 68 years old]

XSL•FO
RenderX

Behavioral Skills

Low Confidence in Self-Management

Some participants doubted that self-management or lifestyle changes could meaningfully protect cognitive health. These accounts reflected low self-efficacy and limited confidence in their ability to act on prevention advice:

> *Memory is determined by nature. What you eat and what you practice are all psychological effects.* [N16, female, 75 years old]

> *I can't even manage my diabetes well. How can I manage my brain? Let it be as it is.* [N19, female, 77 years old]

Knowledge to Action

Participants described a gap between understanding screening recommendations and knowing how to act on them. Even when they considered screening worthwhile, some reported low confidence in making appointments, choosing a department, or initiating the process.

A retired teacher who was aware of the benefits of early intervention confessed:

> *I know I should get checked, but I'm completely confused about how to make an appointment, where to go, and what tests to take. It's so confusing that I just put it off and forget about it eventually.* [N9, female, 73 years old]

Participants' accounts repeatedly returned to the question of "what to do next," highlighting the need for practical instructions, prompts, and support to initiate screening:

> *After the lecture, I thought it made sense, but when I got home and was busy with daily chores, I forgot all about it. There was no reminder, no one to guide me through the first step, and it was very difficult for me to start.* [N21, female, 76 years old]

Community-Based Referral and Navigation

Participants described familiar community health services as a preferred first point of contact for cognitive screening and referral. They emphasized the value of clear, convenient, and trusted guidance about where to go and how to access the appropriate service:

> *If the community hospital can conduct simple initial screenings, or if the doctor can refer me to the right department, I will definitely be willing to go. I myself would go to the hospital for registration, but I couldn't even figure out whether to go to the neurology department or the geriatrics department.* [N17, female, 69 years old]

Participants' accounts portrayed community staff and local networks as sources of practical guidance, trust, and reassurance. These relationships appeared important when older adults felt uncertain about navigating hospital departments or registration processes:

> *Our building's group leader organized everyone to go to the community health center for assessment. With companions and familiar doctors, I felt at ease and I participated.* [N12, male, 68 years old]

Difficulty Changing Established Habits

Participants also described difficulty changing long-established lifestyle and social routines, even when they understood health recommendations. These accounts reflected challenges with self-regulation and translating advice into sustained behavior:

> *The doctor told me to go out more and play cards with others, but I'm used to staying at home alone. I don't know who to go to when I go out. Thinking about it, it's more comfortable to stay at home.* [N3, male, 82 years old]

> *I know eating lightly is good for the brain, but I've been eating salty food all my life and the dishes have lost their flavor. I just can't eat them. My willpower is not strong enough.* [N5, male, 67 years old]

Participants' accounts suggested that established habits could make both preventive lifestyle change and participation in screening activities more difficult. They described these routines as obstacles to acting on health advice and attending screening:

> *I usually enjoy going fishing and playing cards. These are all static activities. I just don't like going out for activities, let alone participating in screening events in the community or at the hospital. I don't like it at all and I can't change that.* [N5, male, 67 years old]

Discussion

Principal Findings

This qualitative study explored how community-dwelling older adults experienced VR-assisted cognitive health education and how their accounts related to the IMB model. Participants described the immersive educational content as making some information about MCI and screening more concrete and understandable, and some reported feeling less fearful or embarrassed about cognitive screening after the session. However, participants also described persistent uncertainty about how to proceed with screening, including where to go, how to register, and whom to approach for assistance. Taken together, these accounts suggest potential information-related, motivation-related, and behavioral skill–related processes through which VR-assisted cognitive health education may support screening readiness, while also highlighting practical conditions that may prevent willingness from translating into action. These findings represent participants' experiences and perceived processes and should not be interpreted as evidence that VR caused an increase in screening uptake.

Information: Dementia Literacy and Cognitive Load

Participants described substantial gaps in basic knowledge about MCI, commonly conflating MCI with dementia or mental illness, and equating screening with being diagnosed and socially labeled. These accounts are consistent with studies documenting limited dementia knowledge in diverse community populations [13,14,36] and qualitative work showing that biomedical terminology, shame, and negative social meanings can shape how people understand dementia [16]. Help-seeking research

XSL•FO
RenderX

further suggests that early cognitive changes are often discounted as normal aging, left uncertain, or misattributed, which can delay recognition of the need for assessment [17]. In our study, misconceptions were not only about disease concepts but also about the purpose and consequences of screening, with some participants anticipating social harm (loss of reputation) or family disruption (worrying adult children), which reduced perceived benefit.

Compared with their prior experiences of conventional health education, participants often described the VR session as easier to understand because it provided concrete visual explanations and could be followed at an individual pace. This perceived benefit is consistent with emerging intervention evidence showing that VR-based dementia education can support dementia-related knowledge, health literacy, or behavioral intentions in some settings [26,27]. In this study, however, participants' descriptions of the immersive presentation should be interpreted as perceived clarity rather than evidence that knowledge objectively improved.

At the same time, participants described limitations related to novelty, peripheral details, and dense information. These concerns are consistent with systematic review evidence emphasizing the need to consider usability and acceptability when applying immersive VR with older adults [25]. Participants' accounts in this study also suggested a preference for information that is paced and layered rather than densely presented. A practical implication is therefore to keep VR education brief and focused on a small set of core messages: what MCI is and is not, why screening may be considered, what screening involves, and what the next steps are in local services.

Motivation: Stigma, Avoidance, Fatalism, and Reconciling With Aging

In the motivation domain, participants' accounts highlighted fear of being labeled, anticipated discrimination, and concerns about face and reputation. This pattern is consistent with prior research identifying stigma, fear, and social consequences of a cognitive label as barriers to screening, diagnostic assessment, or dementia-service use [11,15,37]. Some participants described the community-based cases in the VR session as relatable and less stigmatizing, suggesting that familiar narratives may make screening feel more acceptable. Because these are retrospective qualitative accounts, however, the data do not establish the magnitude or durability of any change in stigma.

Several participants also expressed low perceived self-efficacy and fatalism, including skepticism that lifestyle changes or early identification could help. Pessimism about treatment and low self-prioritization were also evident in some accounts. These themes are consistent with prior research showing that reluctance toward cognitive screening or follow-up can be shaped by perceived limited benefits, psychological burden, stigma, and concerns about the consequences of diagnosis for patients and families [11,13,37]. Some participants described MCI as feeling more manageable after the session, particularly when it was framed as a risk state for which early management may still be meaningful; this should be interpreted as a perceived shift in understanding rather than a measured intervention effect.

A contribution of this study is the detailed description of reconciliation with aging as an adaptive psychological stance. Many participants normalized cognitive decline as part of aging and described acceptance as a way to maintain calm and identity. Similar narratives appear in help-seeking research, where memory changes are often discounted as normal aging or their significance is left uncertain [17]. In our data, this stance was double-edged: it protected well-being but reduced intrinsic motivation to pursue screening. Participants' accounts suggest that the educational content did not necessarily require directly challenging acceptance of aging. For some participants, framing screening in relation to self-care, maintaining independence, and reducing potential family burden appeared to provide an alternative way of considering screening. For nursing practice, motivational work may therefore benefit from culturally sensitive and values-based communication rather than emphasizing disease threat alone.

Women accounted for 15 of the 22 interview participants. Although several female participants discussed health education in relation to family roles and caregiving, the study was not designed to compare women and men, and the overrepresentation of women may reflect self-selection into health education or interview participation. These accounts therefore should not be interpreted as evidence of a sex or gender difference in interest in the intervention.

Behavioral Skills: The Intention-Action Gap and the Centrality of Community Referral

A practice-relevant pattern in the behavioral-skills domain was the gap participants described between considering screening worthwhile and knowing how to complete it. Participants frequently reported uncertainty about where to go, how to register, which department to choose, what screening would involve, and how to take the first step. In their accounts, this intention-action gap was closely tied to complex service navigation and low confidence in navigating hospital systems.

Consistent with research on primary care and community-based models of care, participants emphasized the value of a trusted and convenient referral route. Prior studies have highlighted the role of primary care in cognitive detection [21], the importance of coordination with specialist services in dementia diagnosis [38], and the value of integrated teams and navigation support for older adults [39,40]. In this sample, participants' accounts suggest that education alone may be insufficient when there is no immediate, low-barrier next step; this is an implementation hypothesis that requires prospective evaluation.

For practice, the findings suggest a potential role for community nurses and primary care teams in linking education to action. Future VR-assisted education could be kept brief and structured, followed by facilitated discussion to check understanding and address concerns; multielement VR dementia education has previously combined immersive content with lectures and discussion [26]. It could also end with concrete service information: where screening is available, how to book, what to expect, and who can assist. Nurse-led navigation strategies such as appointment assistance, reminders, or accompaniment should be evaluated rather than assumed to improve uptake.

XSL•FO
RenderX

For education, nursing curricula and continuing professional development should strengthen skills in stigma-sensitive communication, motivational interviewing with older adults, and age-friendly digital health facilitation. Training should also cover how to recognize and respond to reconciliation-with-aging narratives in a way that respects values while supporting preventive action. For policy and service planning, investing in technology alone is unlikely to be sufficient. This emphasis is consistent with evidence supporting stronger coordination between primary care institutions and hospitals in China [20], navigation support for older adults using community-based primary care [39], and interdisciplinary team support in dementia primary care [40]. Future implementation strategies could evaluate standardized community-to-hospital referral processes, clear responsibility for follow-up, and protected staff time for navigation support. Program evaluation could consider not only knowledge-related outcomes but also screening completion and appropriate follow-up care.

Strengths, Limitations, and Future Research

A strength of this study is its theory-informed qualitative design conducted within a community intervention context, which allowed exploration of experiences and contextual factors that quantitative outcomes may not capture. Developing themes from the interview data and subsequently mapping the subthemes to the IMB model provided a structured way to distinguish information, motivation, and behavioral-skill issues while retaining patterns grounded in participants' accounts.

Several limitations should be considered. The sample was drawn from communities in one province and included many participants with secondary education or higher, which may limit transferability to rural or lower-literacy populations who could face different barriers to VR use and screening access. Participants initially volunteered in response to community recruitment information and were then purposively selected with attention to age and educational variation; this process may still have favored older adults who were more willing to engage with cognitive health education or research. Women were overrepresented (15/22), and interview participation may therefore have been subject to self-selection; apparent gender-related patterns should be interpreted cautiously. Future research should prospectively evaluate integrated approaches that combine VR-assisted education with structured nurse-led clarification, referral, reminders, and follow-up and should assess equity, feasibility, and screening completion in underserved settings.

Conclusion

This qualitative study suggests that VR-assisted cognitive health education may be experienced as engaging and understandable by some community-dwelling older adults, while fear, stigma, uncertain self-appraisal, and practical service-navigation barriers can persist. Participants' accounts particularly emphasized the value of trusted community-based guidance when considering what to do after receiving screening information. These findings support further evaluation of models that pair age-friendly digital education with clear referral and navigation support, but they do not establish that VR itself increases screening uptake.

Acknowledgments

We sincerely thank the older adults who generously dedicated their time to participate in this study. We also express our appreciation to the staff at the home-based older adult care service centers for their assistance in facilitating data collection.

Funding

This work is supported by the National Key Research and Development Program of China (2023YFC3603600), the National Natural Science Foundation of China (72174095), the Open Research Project of Nursing Discipline under the Fourth Phase of Jiangsu Province's Advantageous Discipline Program (YSHL202503), and Jiangsu Province Postgraduate Research and Practical Innovation Project (26CXJH3684).

Data Availability

The datasets used and analyzed during this study are available from the corresponding author on reasonable request.

Generative AI Use

During preparation and revision of this manuscript, the authors used ChatGPT and Gemini to assist with English-language editing, including grammar, syntax, clarity, and readability. These tools were not used to conduct the primary qualitative coding or independently interpret the original interview data. The authors reviewed and edited all AI-assisted output and take full responsibility for the content of the publication.

Authors' Contributions

Conceptualization: TY, YB
Data curation: TY, YS
Formal analysis: TY, BY, JY, YW
Funding acquisition: YB, YS, TY

XSL•FO
RenderX

Investigation: YS, XZ
Methodology: TY, HY
Project administration: YS, BY, GX
Resources: TY, HY
Software: TY, YS
Supervision: BY, GX
Validation: BY, JY
Visualization: YW, XZ
Writing – original draft: TY, BY
Writing – review & editing: TY, BY, GX

Conflicts of Interest

None declared.

Multimedia Appendix 1
COREQ checklist.
[DOCX File , 16 KB - humanfactors_v13i1e91657_app1.docx]

Multimedia Appendix 2
Relationship between the qualitative interview sample and the parent cluster-randomized trial sample.
[DOCX File , 18 KB - humanfactors_v13i1e91657_app2.docx]

References

1. Alzheimer's Association. 2025 Alzheimer's disease facts and figures. Alzheimers Dement 2025 Apr 29;21(4):e70235. [doi: 10.1002/alz.70235]
2. Dementia. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/dementia [accessed 2026-01-10]
3. Jia L, Du Y, Chu L, Zhang Z, Li F, Lyu D, COAST Group. Prevalence, risk factors, and management of dementia and mild cognitive impairment in adults aged 60 years or older in China: a cross-sectional study. Lancet Public Health 2020 Dec;5(12):e661-e671 [FREE Full text] [doi: 10.1016/S2468-2667(20)30185-7] [Medline: 33271079]
4. Livingston G, Huntley J, Liu KY, Costafreda SG, Selbæk G, Alladi S, et al. Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. Lancet 2024 Aug 10;404(10452):572-628. [doi: 10.1016/S0140-6736(24)01296-0] [Medline: 39096926]
5. Petersen RC. Early diagnosis of Alzheimer's disease: is MCI too late? Curr Alzheimer Res 2009 Aug;6(4):324-330 [FREE Full text] [doi: 10.2174/156720509788929237] [Medline: 19689230]
6. Cacabelos R. What have we learnt from past failures in Alzheimer's disease drug discovery? Expert Opin Drug Discov 2022 Apr;17(4):309-323. [doi: 10.1080/17460441.2022.2033724] [Medline: 35129021]
7. Ritchie CW, Terrera GM, Quinn TJ. Dementia trials and dementia tribulations: methodological and analytical challenges in dementia research. Alzheimers Res Ther 2015;7(1):31 [FREE Full text] [doi: 10.1186/s13195-015-0113-6] [Medline: 25788988]
8. Jobke B, McBride T, Nevin L, Peiperl L, Ross A, Stone C, as the PLOS Medicine Editors. Setbacks in Alzheimer research demand new strategies, not surrender. PLoS Med 2018 Feb;15(2):e1002518 [FREE Full text] [doi: 10.1371/journal.pmed.1002518] [Medline: 29486005]
9. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 2020 Aug 08;396(10248):413-446 [FREE Full text] [doi: 10.1016/S0140-6736(20)30367-6] [Medline: 32738937]
10. Robinson L, Tang E, Taylor J. Dementia: timely diagnosis and early intervention. BMJ 2015 Jun 16;350:h3029 [FREE Full text] [doi: 10.1136/bmj.h3029] [Medline: 26079686]
11. Jin J, Li BJ, Theng YL. Why people hesitate about dementia screening: a systematic review and meta-analysis. Soc Sci Med 2025;382:118360. [doi: 10.1016/j.socscimed.2025.118360] [Medline: 40651239]
12. Ford E, Rooney P, Oliver S, Hoile R, Hurley P, Banerjee S, et al. Identifying undetected dementia in UK primary care patients: a retrospective case-control study comparing machine-learning and standard epidemiological approaches. BMC Med Inform Decis Mak 2019 Dec 02;19(1):248 [FREE Full text] [doi: 10.1186/s12911-019-0991-9] [Medline: 31791325]
13. Vrijsen J, Maeckelberghe ELM, Broekstra R, de Vries JJ, Abu-Hanna A, De Deyn PP, et al. Knowledge, health beliefs and attitudes towards dementia and dementia risk reduction among descendants of people with dementia: a qualitative study using focus group discussions. BMC Public Health 2021 Jul 07;21(1):1344 [FREE Full text] [doi: 10.1186/s12889-021-11415-2] [Medline: 34233658]

XSL•FO
RenderX

14. Zhang H, Loi SM, Zhou S, Zhao M, Lv X, Wang J, et al. Dementia literacy among community-dwelling older adults in urban China: a cross-sectional study. Front Public Health 2017;5:124 [FREE Full text] [doi: 10.3389/fpubh.2017.00124] [Medline: 28638820]
15. Bajwa R, Hanjari M, Al-Oraibi A, Akyea R, Brar M, Robinson L, et al. Qualitative study exploring knowledge and attitudes towards dementia risk prediction, barriers to dementia services and service improvement recommendations with diverse populations in England. BMJ Open 2025 May 30;15(5):e092370 [FREE Full text] [doi: 10.1136/bmjopen-2024-092370] [Medline: 40447431]
16. Kłosińska U, Leszko M. Do I have symptoms of dementia: a discursive study of awareness and shame among people with advanced dementia. Gerontologist 2024 Aug 01;64(8):gnae067 [FREE Full text] [doi: 10.1093/geront/gnae067] [Medline: 38835189]
17. Perry-Young L, Owen G, Kelly S, Owens C. How people come to recognise a problem and seek medical help for a person showing early signs of dementia: a systematic review and meta-ethnography. Dementia (London) 2018 Jan;17(1):34-60 [FREE Full text] [doi: 10.1177/1471301215626889] [Medline: 26764265]
18. Bradford A, Kunik ME, Schulz P, Williams SP, Singh H. Missed and delayed diagnosis of dementia in primary care: prevalence and contributing factors. Alzheimer Dis Assoc Disord 2009;23(4):306-314 [FREE Full text] [doi: 10.1097/WAD.0b013e3181a6bebc] [Medline: 19568149]
19. Li X, Lu J, Hu S, Cheng KK, De Maeseneer J, Meng Q, et al. The primary health-care system in China. Lancet 2017 Dec 09;390(10112):2584-2594. [doi: 10.1016/S0140-6736(17)33109-4] [Medline: 29231837]
20. Li X, Krumholz HM, Yip W, Cheng KK, De Maeseneer J, Meng Q, et al. Quality of primary health care in China: challenges and recommendations. Lancet 2020 Jun 06;395(10239):1802-1812 [FREE Full text] [doi: 10.1016/S0140-6736(20)30122-7] [Medline: 32505251]
21. Fowler NR, Partrick KA, Taylor J, Hornbecker M, Kelleher K, Boustani M, et al. Implementing early detection of cognitive impairment in primary care to improve care for older adults. J Intern Med 2025 Jul;298(1):31-45. [doi: 10.1111/joim.20098] [Medline: 40410933]
22. Tsai CF, Chen CC, Wu EH, Chung CR, Huang CY, Tsai PY, et al. A machine-learning-based assessment method for early-stage neurocognitive impairment by an immersive virtual supermarket. IEEE Trans Neural Syst Rehabil Eng 2021;29:2124-2132. [doi: 10.1109/TNSRE.2021.3118918] [Medline: 34623270]
23. Wen J, Yan H, Wang S, Xu J, Zhou Z. The effectiveness of nursing interventions for elderly dementia patients based on virtual reality technology: a systematic review and meta-analysis. Ageing Res Rev 2024 Jan;93:102135. [doi: 10.1016/j.arr.2023.102135] [Medline: 37995900]
24. Bryant L, Stubbs P, Bailey B, Nguyen V, Bluff A, Hemsley B. Interacting with virtual characters, objects and environments: investigating immersive virtual reality in rehabilitation. Disabil Rehabil Assist Technol 2025 Jan;20(1):107-117 [FREE Full text] [doi: 10.1080/17483107.2024.2353284] [Medline: 38781087]
25. Dermody G, Whitehead L, Wilson G, Glass C. The role of virtual reality in improving health outcomes for community-dwelling older adults: systematic review. J Med Internet Res 2020 Jun 01;22(6):e17331 [FREE Full text] [doi: 10.2196/17331] [Medline: 32478662]
26. Suzuki H, Igarashi A, Matsumoto H, Kugai H, Takaoka M, Sakka M, et al. A dementia-friendly educational program using virtual reality for the general public in Japan: a randomized controlled trial for DRIVE. Gerontologist 2024 Nov 01;64(11):gnae113 [FREE Full text] [doi: 10.1093/geront/gnae113] [Medline: 39140623]
27. Chang C, Huang K, Kuo L, Cheng Y, Huang S, Chuang T, et al. Pilot randomized controlled study on the effectiveness of a virtual reality-based dementia prevention program using self-regulated learning strategies among older adults with mild cognitive impairment. Healthcare (Basel) 2025 May 07;13(9):1082 [FREE Full text] [doi: 10.3390/healthcare13091082] [Medline: 40361860]
28. Fisher JD, Fisher WA. Changing AIDS-risk behavior. Psychol Bull 1992 May;111(3):455-474. [doi: 10.1037/0033-2909.111.3.455] [Medline: 1594721]
29. Choi H, Dion H, Huang M, Sathyan L, Herfel E, Makhulo B, et al. Applying the information-motivation-behavioral skills model to a video-assisted HPV intervention to promote self-screening uptake: a qualitative study in Western Kenya. BMJ Glob Health 2025 May 06;10(5):e017616 [FREE Full text] [doi: 10.1136/bmjgh-2024-017616] [Medline: 40335076]
30. Lee MH, Park Y. The effectiveness of the information-motivation-behavioral skills model-based intervention on preventive behaviors against respiratory infection among community-dwelling older adults. Patient Educ Couns 2021 Aug;104(8):2028-2036 [FREE Full text] [doi: 10.1016/j.pec.2021.01.016] [Medline: 33531157]
31. Kim SR, Kim S, Kim HY, Cho K. Predictive model of self-management in patients with stroke based on the information-motivation-behavioral skills model. J Cardiovasc Nurs 2023;38(2):158-167. [doi: 10.1097/JCN.0000000000000883] [Medline: 35030109]
32. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g] [Medline: 10940958]
33. Doyle L, McCabe C, Keogh B, Brady A, McCann M. An overview of the qualitative descriptive design within nursing research. J Res Nurs 2020 Aug;25(5):443-455 [FREE Full text] [doi: 10.1177/1744987119880234] [Medline: 34394658]

XSL•FO
RenderX

34. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
35. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human participants. JAMA 2025 Jan 07;333(1):71-74. [doi: 10.1001/jama.2024.21972] [Medline: 39425955]
36. Ta Park V, Nam B, Huang D, Yun SW, Phan N, Lee EJ, et al. Knowledge, attitudes, and behaviors related to dementia prevention and caregiving among Korean Americans (the KIMCHI Project): pre- and posttest evaluation study. JMIR Aging 2025 Aug 15;8:e72147 [FREE Full text] [doi: 10.2196/72147] [Medline: 40815825]
37. Henderson JT, Martin A, Patnode CD, Henrikson NB. A synthesis of qualitative studies on patient and caregiver experiences with cognitive impairment screening and diagnosis. Aging Ment Health 2023;27(8):1506-1517. [doi: 10.1080/13607863.2022.2126431] [Medline: 36193565]
38. Abe M, Tsunawaki S, Dejonckheere M, Cigolle CT, Phillips K, Rubinstein EB, et al. Practices and perspectives of primary care physicians in Japan and the United States about diagnosing dementia: a qualitative study. BMC Geriatr 2021 Oct 11;21(1):540 [FREE Full text] [doi: 10.1186/s12877-021-02457-7] [Medline: 34635089]
39. Lafortune C, Huson K, Santi S, Stolee P. Community-based primary health care for older adults: a qualitative study of the perceptions of clients, caregivers and health care providers. BMC Geriatr 2015 Apr 30;15:57 [FREE Full text] [doi: 10.1186/s12877-015-0052-x] [Medline: 25925552]
40. Angomas E, Blinka MD, Kelly E, Oniha O, McGuire M, Colburn JL, et al. "It's Usually Not Dementia That's the Focus": barriers and facilitators of dementia care in primary care. J Gen Intern Med 2025 Jul;40(10):2265-2275. [doi: 10.1007/s11606-025-09493-8] [Medline: 40229605]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
IMB: information-motivation-behavioral
MCI: mild cognitive impairment
VR: virtual reality

Edited by A Kushniruk; submitted 18.Jan.2026; peer-reviewed by D Anyanwu, A Maafs, N Kumar; comments to author 06.Aug.2026; revised version received 14.Aug.2026; accepted 02.Oct.2026; published 21.Oct.2026.

Please cite as:
Yin T, Song Y, Yang B, Yan J, Wang Y, Yin H, Zhang X, Xu G, Bai Y
Applying the Information-Motivation-Behavioral Skills Model to a Virtual Reality–Assisted Cognitive Health Education Intervention to Promote Screening Uptake: Qualitative Study
JMIR Hum Factors 2026;13:e91657
URL: https://humanfactors.jmir.org/2026/1/e91657
doi:10.2196/91657
PMID:

©Tingting Yin, Yulei Song, Bingqing Yang, Jingjing Yan, Yan Wang, Haiyan Yin, Xueqing Zhang, Guihua Xu, Yamei Bai. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 21.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Stakeholder Perspectives on Information About Digital Health Applications in German Outpatient Mental Health Care: Qualitative Interview Study

Klemens Höfer[1], MA; Pascal Raszke[1], MA; Jürgen Wasem[1], Prof Dr; Felix Plescher[1], MA; Anna Bußmann[2], Dr rer medic; Sarah Schlierenkamp[2], MA; Michael Minor[2], MA; Luisa Friedrich[2], BA; Carsten Volland[2], MA; Sophia Zander[3], MSc; Anja Wadeck[3], MD; Udo Schneider[3], PD Dr rer pol; Josepha Katzmann[4], Dr. rer. medic.; Enno Maaß[4], Dr. rer. nat.; Carina Abels[1], Dr rer medic

[1]Institute for Healthcare Management and Research, University of Duisburg-Essen, Essen, Germany

[2]Essener Forschungsinstitut für Medizinmanagement, Essen, Germany

[3]Techniker Krankenkasse, Hamburg, Germany

[4]DeutschePsychotherapeutenVereinigung, Berlin, Germany

Corresponding Author:
Klemens Höfer, MA
Institute for Healthcare Management and Research
University of Duisburg-Essen
Thea-Leymann-Straße 9
Essen, 45127
Germany
Phone: 49 201 18 34037
Email: klemens.hoefer@medman.uni-due.de

Abstract

Background: Digital health applications (digitale gesundheitsanwendungen; DiGA) were integrated into the German statutory health care system in 2020 and can be prescribed by physicians and psychotherapists. Nevertheless, their use within routine care, particularly in the treatment of mental disorders, remains limited. Previous research suggests that a lack of information and knowledge among health care providers and patients may contribute to these difficulties.

Objective: This study aimed to explore the perspectives of stakeholders on the information-related aspects of DiGA implementation in outpatient mental health care in Germany. This included investigating the perceived information status, information needs, suitable information providers, and potential approaches to improve information provision.

Methods: We conducted semistructured interviews with experts representing stakeholder groups relevant to DiGA in outpatient mental health care in Germany. Participants were recruited using purposive stakeholder sampling. Potential participants were identified through existing professional contacts and web-based searches and were subsequently approached by email. The web-based interviews were conducted via Microsoft Teams between September and November 2023. The transcripts were analyzed using qualitative content analysis based on the method described by Kuckartz, combining deductive and inductive category development.

Results: A total of 13 interviews with 14 participants were conducted, comprising 5 health care provider representatives, 4 statutory health insurance representatives, 4 DiGA developer representatives, and 1 patient representative. The findings were grouped into 3 main themes: perceived information status, improvement of the information status, and information suppliers. Representatives of health care providers described a general awareness of DiGA but said they lacked more detailed knowledge required to assess and compare individual applications. They also perceived information gaps among patients. Participants distinguished between general information on aspects such as regulatory requirements and organizational processes and more detailed product-specific information. Different potential information suppliers were associated with different opportunities and problems, particularly regarding expertise, objectivity, available resources, and the ability to provide product-specific information.

Conclusions: The findings indicate that information-related challenges in the implementation of DiGA extend beyond general awareness. Health care providers may require accessible information with different levels of detail, while the source of information

XSL•FO
RenderX

also appears relevant. Addressing these different information needs may support the integration of DiGA into outpatient mental health care.

International Registered Report Identifier (IRRID): RR2-10.2196/56205

(JMIR Hum Factors 2026;13:e82514) doi:10.2196/82514

KEYWORDS

DiGA; digital health application; digitale gesundheitsanwendung; health care research; information; mental disorders; mental health care; mHealth; mobile health; qualitative research

Introduction

Mental disorders represent a significant challenge to the health care system in Germany. Approximately 28% of adults experienced a mental disorder within a 12-month period between 2009 and 2012 [1]. Furthermore, absenteeism from work due to mental disorders increased by 146% between 2000 and 2023, highlighting the significant societal burden of mental disorders [2]. Timely access to outpatient psychotherapy remains challenging; German claims data for patients with depression indicate an average time of approximately 13 weeks between initial contact with a psychotherapist and the start of psychotherapy [3]. These challenges highlight the need for additional approaches to support existing structures. Digital mental health interventions are one low-threshold approach for expanding the range of options available in outpatient care.

Germany introduced a specific regulatory and reimbursement framework for mobile health (mHealth) apps with the Digital Healthcare Act in 2019, and the first digital health applications (Digitale Gesundheitsanwendungen; DiGA) became available in October 2020 [4]. DiGA are certified and quality-assured mHealth apps that can be prescribed by physicians or psychotherapists and are reimbursed by statutory health insurance (SHI). Alternatively, insured individuals have the option of applying directly to their SHI fund, should they be able to provide up-to-date proof of a relevant diagnosis. To qualify for reimbursement, DiGA must undergo the DiGA-Fast-Track Assessment conducted by the Federal Institute for Drugs and Medical Devices (Bundesinstitut für Arzneimittel und Medizinprodukte; BfArM). As part of this procedure, the BfArM evaluates whether the app meets requirements concerning suitability for use, data protection, quality, interoperability, and whether a positive health care effect has been or can be demonstrated during a provisional listing period [4]. DiGA can be listed permanently or provisionally; in the latter case, the app is already reimbursable while additional evidence of its positive effects is generated during an evaluation period of up to 12 months, which can be extended by up to another 12 months [4]. As of October 2026, 63 DiGA were listed in the directory, of which 30 were designed for the treatment of mental disorders [5].

Despite their formal integration into the health care system, low use rates show that DiGA have not yet become an established part of routine care [6]. Previous research suggests that limited information and knowledge may contribute to difficulties in their implementation. In a cross-sectional survey study in 2022, only 22% of general practitioners reported feeling confident about advising patients on DiGA [7]. More recent qualitative research among general practitioners identified information needs regarding evidence, usability, and cost-related aspects of DiGA [8]. Similar information needs were reported among psychotherapists, who additionally emphasized aspects such as data privacy and guidance on the prescription process [9]. Awareness remains limited among the general population in Germany, with a recent representative study showing that over 90% of respondents were unfamiliar with DiGA [10]. This suggests that adequate information may be relevant both for those recommending DiGA and for those potentially using them. Thus, the current level of information and awareness may represent a barrier to the broader implementation of DiGA, while appropriately tailored information could support their integration into routine care.

The implementation of DiGA involves several stakeholders with different roles in providing, receiving, and communicating information. These include health care providers, SHI funds, DiGA developers, and patients, whose perspectives may differ according to their respective roles in the care process. Against this background, the present study focuses on the information dimension of DiGA implementation in outpatient mental health care in Germany. The analysis provides an in-depth examination of information-related topics across stakeholder perspectives. Specifically, the study explores the perceived information status regarding DiGA, approaches to improving the information status through general and product-specific information, and the roles of different information suppliers in outpatient mental health care.

Methods

Overview

The manuscript reports an analysis of information-related aspects within an interview study. The other publication, based on the same interview material, examined the pathways to accessing DiGA [11]. Reporting was guided by the 32-item COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist [12]. The completed checklist is provided in Multimedia Appendix 1.

Study Design and Reporting

This exploratory qualitative study used semistructured expert interviews to investigate stakeholder perspectives on the implementation of DiGA in outpatient mental health care in Germany. The interviews form part of the broader mixed methods project, “Realizing the Potential of Digital Health Applications in Outpatient Mental Health Care” (DiGAPsy), which analyzed the implementation of DiGA in outpatient

XSL•FO
RenderX

mental health care from various angles. The main objective was to identify barriers and develop health policy recommendations to improve the integration of DiGA into mental health care [13]. The mixed methods project included a scoping review [14], an analysis of SHI claims, qualitative analysis of focus groups and expert interviews [11], patient and health care provider surveys, and preference evaluation among both groups [15].

Theoretical Framework and Interview Guide Development

The interviews aimed to explore the barriers to implementing DiGA and potential approaches to addressing these issues. The interview guides covered several areas relevant to the broader DiGAPsy project, including general experiences with DiGA, implementation barriers, information needs and provision, adherence, organizational structures, and potential models for integrating DiGA into outpatient mental health care.

Four semistructured interview guides were developed for representatives of the following stakeholder groups: health care providers, patients, DiGA developers, and SHI funds. The interview guides were based on the research questions and objectives of the DiGAPsy project and on the findings of a previously conducted scoping review [14]. KH initially developed the guides, which were subsequently revised by CA and discussed within the project consortium. Before data collection began, each stakeholder-specific guide underwent an internal pretest within the project consortium to assess comprehensibility, thematic structure, and feasibility. Based on this pretest, minor adjustments were made to the wording and structure of individual questions. No pilot interviews with representatives of the target stakeholder groups were conducted. The guides followed a common thematic structure, but individual questions and their wording were slightly adapted to the role of the respective stakeholder group. Participants were informed about the topics of the interviews but did not receive the interview questions in advance. The complete interview guides are provided in Multimedia Appendix 2.

Sampling and Recruitment

A purposive stakeholder sampling strategy was defined before recruitment. Four stakeholder groups were considered relevant because of their roles in the DiGA care process and in providing, receiving, and communicating information about DiGA. Health care provider representatives were included because they can prescribe DiGA, integrate them into routine care, receive information from other stakeholders, and interact with patients. SHI representatives were included because SHI funds provide an alternative point of access to DiGA, pay for the applications, and have established channels for communicating information to their insured. DiGA developer representatives were included because developers provide the applications and communicate product-specific information about their own products. A patient representative was included because patients receive information from these different sources and may also contribute to the information exchange by communicating with other stakeholder groups. The selection of these stakeholder groups and potential institutions was discussed within the DiGAPsy project consortium before recruitment started.

Potential institutions were identified through existing professional contacts of the research team and additional web-based searches for organizations belonging to the predefined stakeholder groups. Institutions were selected to represent different perspectives within these groups. For SHI funds, different types of SHI funds were considered. Among DiGA developers, both larger developers offering several DiGA and smaller developers offering a single DiGA were approached. In addition, developers of both provisionally and permanently listed DiGA were considered. Health care provider representatives were recruited from medical and psychotherapist chambers, Associations of Statutory Health Insurance Physicians, and professional associations representing health care providers. Patient organizations were approached to represent the patient perspective.

Participants were eligible if they represented one of the predefined stakeholder groups and their professional role involved DiGA, digitalization, or the treatment of people with mental disorders. No additional formal exclusion criteria were defined. The participating institutions selected or referred individuals with relevant professional expertise. Potential participants were subsequently contacted directly by email.

In total, 15 institutions were contacted, of which 2 declined participation. Because participants were recruited as institutional experts on the basis of their professional roles and expertise rather than individual characteristics, sociodemographic characteristics such as age and gender were not considered for participant selection or systematically collected. Individual previous experience with DiGA was likewise not systematically assessed. One patient representative was included in the expert interview sample. Within the broader DiGAPsy project, patient perspectives were examined in more detail in 2 separate components: a qualitative focus group study involving 31 patients and a cross-sectional survey involving 1413 patients.

Sample Size and Thematic Coverage

The target number of 12 expert interviews was predefined during the planning of the DiGAPsy project [13]. Recruitment was concluded after this predefined target had been reached. One additional interview was conducted because a stakeholder who had already been approached agreed to participate at a later stage. Overall, 13 interviews involving 14 participants were conducted; 1 interview included 2 SHI representatives. The final sample comprised 5 health care provider representatives, 4 SHI representatives, 4 DiGA developer representatives, and 1 patient representative. A detailed overview of the interviews and stakeholder groups is provided in Multimedia Appendix 3.

No formal assessment of theoretical or thematic saturation was conducted during data collection or analysis. The sample size was determined by the predefined study plan rather than by a saturation criterion. Accordingly, saturation was not used as a criterion for ending data collection.

Data Collection and Research Team

The interviews were carried out online using Microsoft Teams between September and November 2023 and lasted between 52 and 82 minutes. All interviews were conducted in German. No persons other than the participants and members of the research

XSL•FO
RenderX

team were present. No repeat interviews were conducted. Audiovisual recordings of the interviews were made after participants had provided written informed consent.

All 3 researchers who conducted the interviews had an academic background in economics, specializing in health care management. KH and FP held master's degrees, while CA held a PhD. At the time of the study, KH and FP were male researchers at the Institute for Health Care Management and Research at the University of Duisburg-Essen, while CA was a female researcher and head of the research group. All 3 researchers had previous experience conducting qualitative interviews in other research projects. During each interview, 1 researcher primarily moderated the discussion, while another researcher followed the conversation and contributed follow-up or clarification questions where appropriate. Brief notes on potential follow-up questions and specific participant statements were made during the interviews and used to support the interview. These notes were not analyzed as a separate data source.

At the start of each interview, the researchers introduced themselves and their professional roles and explained the purpose of the research project. Two of the interviewers (CA and FP) had previously met 2 participants in the context of an interview or workshop conducted as part of a previous project ("Continuous Quality Assurance of Digital Health Applications," funding code: 01VSF20007). For the remaining participants, contact prior to the interviews was limited to recruitment and organizational arrangements. None of the 3 interviewers (KH, CA, and FP) had a predefined position in favor of or against the use of DiGA. No conflicts of interest related to the conduct of the interviews were reported.

Transcription and Translation

The audiovisual recordings were fully transcribed in accordance with the established transcription rules of Claussen et al [16] by student research assistants. Identifying information relating to individuals, organizations, or institutions was pseudonymized during transcription. Each completed transcript was subsequently reviewed by a second member of the research team (CA, KH, or FP). The transcripts were checked for completeness and plausibility, and unclear or discrepant passages were compared with the original recordings and corrected if necessary. The audiovisual recordings were deleted after transcription and quality assurance had been completed.

The full coding and qualitative analysis were conducted in German. For this manuscript, selected quotations were translated into English. These were then checked against the original German statements to ensure that their substantive meaning was retained.

Data Analysis

The pseudonymized interview transcripts were analyzed using MAXQDA 2022 (VERBI GmbH), based on the approach to qualitative content analysis described by Kuckartz [17]. The analysis combined deductive and inductive category development. The initial deductive coding framework was developed by KH and CA based on the main themes of the interview guides. No separate trial coding was carried out. PR conducted the initial coding of the interview material. During this first coding phase, relevant passages were assigned to the predefined deductive categories, while additional inductive codes were developed to cover themes and subthemes that were not sufficiently captured by the initial coding framework. In a second round, CA reviewed the initial coding and performed quality assurance. This review focused particularly on the assignment of individual text passages to the main categories and on the consistency of the deductive and inductive coding. Passages that were difficult to categorize, as well as ambiguities regarding the structure and definitions of individual codes, were subsequently discussed between KH and CA. Remaining differences were resolved through discussion until consensus was reached. The final information-related coding tree is provided in Multimedia Appendix 4.

The interviews and the coding framework covered several dimensions of DiGA implementation. For the present article, the analysis and reporting were restricted to the codes and coded material relating to information. Other thematic dimensions of the interview material were outside the scope of the present analysis and are addressed separately [11]. Transcripts were not returned to participants for correction, and no participant checking of the coding or study findings was performed.

Ethical Considerations

The study was conducted as part of the DiGAPsy project and was reviewed and approved by the Ethics Committee of the Medical Faculty of the University of Duisburg-Essen (reference number 23-11209-BO). Prior to participating in the study, all interviewees received written information about the project and provided written informed consent using a consent form. This covered participation in the study, the recording and transcription of the interview, the pseudonymized analysis of the data, and the scientific publication of anonymized quotations. Identifying information was stored separately from the interview transcripts, and access to the study data was restricted to the research team. Participants were offered a €100 (US $111.89 as of October 7, 2026) allowance for taking part in the interview.

Results

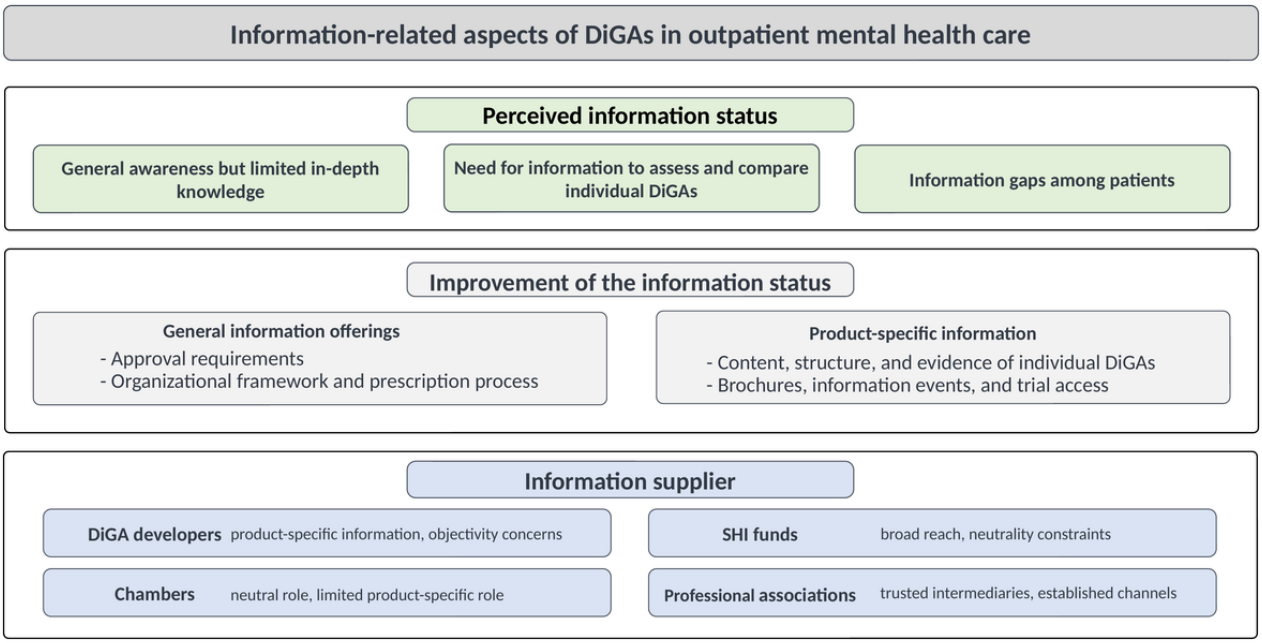
Overview

A total of 14 participants were recruited for the interviews, resulting in 13 interviews, with 1 double interview. An overview is provided in Multimedia Appendix 3.

The results reflect the experts' perspectives on DiGA for outpatient mental health care. Opinions and statements regarding information on DiGA were grouped into the 3 main categories "perceived information status," "improvement of the information status," and "information supplier." These categories and the subcategories into which the results are grouped are presented in Figure 1 and are described in the following sections.

XSL•FO
RenderX

Figure 1. Information-related aspects of digitale gesundheitsanwendungen (DiGA) in outpatient mental health care. SHI: statutory health insurance.



Perceived Information Status

Representatives of health care providers indicated that they were familiar with the existence of DiGA but described their knowledge as mostly superficial. According to them, existing awareness did not translate into sufficient knowledge for health care providers to feel confident using them.

> *I think that is the main hurdle, that we are just not informed enough.* [Health care provider representative 3]

Similarly, DiGA developer representatives perceived that, although DiGA were generally known among health care providers, only a limited proportion had incorporated them into their routine practice.

> *The proportion of doctors who have really internalized DiGA for themselves and use it, regardless of psychology, but psychology is of course the biggest indication, is still vanishingly small.* [DiGA developer representative 4]

Information gaps were also perceived on the patient side. DiGA developer representatives described awareness and knowledge of DiGA among patients as limited, and the patient representative agreed with this assessment by also highlighting information gaps.

Beyond general awareness, representatives of health care providers described uncertainties regarding the evidence supporting individual DiGA. They were particularly concerned about the extent to which findings from clinical studies could be transferred to real-world settings and whether there was sufficient evidence on how DiGA perform in routine care.

> *Yes, so we need more, even more knowledge about how the apps work in the real world. The studies with a small group of subjects at the university who are then selected or who report under certain conditions is certainly something different than if you suggest it to patients who may already have an indication for psychotherapy and then use it until the waiting time is bridged. We have virtually no data on this yet.* [Health care provider representative 4]

Another challenge was the large number of available DiGA and the effort required to gain an understanding of each. Representatives of health care providers described difficulties in obtaining or maintaining detailed knowledge of available products, particularly when several DiGA addressed the same indication area.

> *And especially if you now imagine that 50% of DiGA go into the psych field, then I can't expect psychotherapists to be informed about all DiGA.* [Health care provider representative 2]

Participants talked about a lack of detailed, application-specific knowledge. It was indicated that the challenges concerned not mainly the availability of information but more the ability to navigate the range of applications and assess and compare the information relevant to their use.

Improvement of the Information Status

The perception of low levels of information led to considerations of how these could be enhanced. Several options for achieving this goal have been described, which can be split into 2 fields. First, general offerings to increase awareness of DiGA in general. Second, information offerings at a product-specific level with the aim of increasing awareness of one DiGA in particular.

General Information Offerings

Participants described improving general information on DiGA as a broader task involving different stakeholder groups. DiGA developer representatives emphasized that increasing knowledge would require efforts beyond individual organizations. They viewed information on aspects such as evidence, data privacy,

XSL•FO
RenderX

and regulatory requirements as a possibility for increasing confidence in the quality and benefits of DiGA.

> *And to provide good information about DiGA, and also to address the issues of privacy and evidence, for example, at some point, to get involved in these processes and to communicate them so that not every health care provider feels that they have to look behind the scenes themselves.* [DiGA developer representative 3]

Representatives of health care providers particularly emphasized the need for clearer information on the differences in the evidence requirements for the 2 types of approval status. Some expressed concerns regarding the evidence base of provisionally listed DiGA and considered it important to communicate more clearly that the strength of evidence may differ between applications.

> *Then it would certainly be important to provide information about this, and also to say that there are also some DiGA that have really good evidence.* [Health care provider representative 4]

In addition to evidence and regulatory information, health care provider representatives expressed a need for practical guidance and an organizational framework for using DiGA in routine care. This included information on the prescription process, which was considered particularly relevant for psychotherapists, as prescribing is not traditionally part of their routine practice.

> *However, psychotherapists are not typically involved in issuing prescriptions. This is a long-standing practice that they are not familiar with.* [Health care provider representative 3]

Continuing medical education was discussed as one possible way of providing such general and organizational information. Participants considered educational formats suitable for explaining what DiGA are, how they are structured and used, and how the prescription process works. At the same time, health care provider representatives pointed to limitations of this approach when it comes to individual applications, as general educational formats cannot provide detailed information on every specific DiGA.

> *So, there are training courses on digitization that can be widely distributed, you can say we're doing a training course on how to use DiGA in psychotherapy, how to use them, but you can't focus an individual DiGA.* [Health care provider representative 2]

The usability and structure of existing information sources were also discussed. In particular, participants identified limitations of the DiGA directory. Its structure and information hierarchy were described as insufficiently user-friendly, making it difficult to identify relevant information quickly in routine care. SHI representatives emphasized that this was especially problematic when health care providers had little time to assess suitable alternatives.

> *If you imagine you have a full waiting room, then he can't go to the DiGA directory and filter for mental disorders and take a look - that's not possible.* [SHI representative 2]

Participants therefore emphasized the importance of presenting information in a more structured and standardized way. Such a framework could facilitate access to relevant characteristics of individual applications and improve comparability between DiGA, thereby supporting health care providers in assessing available options.

> *But actually, of course, I would prefer to have a relatively highly structured information framework based on certain criteria and thus be able to establish comparability between DiGA.* [DiGA developer representative 3]

Product-Specific Information

In addition to providing general information, health care provider representatives emphasized the need for more detailed, product-specific information to support the assessment of individual applications. DiGA developer representatives described product-specific information provision as a continuous process rather than a one-time distribution of information materials. They indicated that information may need to be provided repeatedly to maintain awareness among health care providers. Developer representatives also described regular information events as an established approach to providing interested health care providers with more comprehensive product-specific information. These events could be integrated into continuing medical education and were described as an opportunity to explain the content, structure, and functioning of individual applications in greater detail. The need for such product-specific information was also linked to difficulties in assessing an application before recommending it to a patient.

> *Well, if I were to recommend a book to a patient, I would always look at it first. But I can't really take a closer look at a DiGA.* [Health care provider representative 1]

Health care provider representatives therefore described a need for product-specific information that is brief and easy to grasp while still providing sufficient detail on the content, structure, and available evidence. The existing materials provided by developers were not always considered to meet these requirements, especially when their focus was primarily on promoting the application rather than providing sufficient information for its evaluation.

> *It must also be said that the DiGA developers do not provide very comprehensive information. So, when I get a flyer in the office, the product is advertised on three pages, but there is very little specific information. That's not enough for me.* [Health care provider representative 1]

The patient representative additionally emphasized that product-specific information should communicate what patients can realistically expect from using a DiGA, including the active involvement that may be required during use.

> *This is not a miracle pill, but you now have to work on yourself, it is a new experience and the patients who use it should be informed about this beforehand.* [Patient representative 1]

XSL•FO
RenderX

Trial access for health care providers was discussed from different perspectives. Overall, participants viewed such access positively, and it was described as already being used in practice. Health care provider and DiGA developer representatives considered trial access a way of increasing transparency and reducing uncertainty by allowing health care providers to familiarize themselves directly with the content, structure, and functioning of an application. Developer representatives also described providing trial access as part of their contribution to supporting the uptake of DiGA and as an opportunity to present their own products to health care providers.

> *And I think that's the best advertising you can do, to actually give a physician the opportunity to see the program and experience it from the patient's point of view.* [DiGA developer representative 3]

At the same time, DiGA developer representatives also expressed reservations. They questioned the extent to which the experience of a health care provider testing a DiGA could reflect that of patients for whom the application was developed.

> *That's the famous bait that has to taste good to the fish and not to the fisherman, that is - and then the doctor comes along and says, hmm, and then he tries to put himself in the patient's shoes somehow, it just doesn't work.* [DiGA developer representative 4]

As an alternative to relying on health care providers to test individual applications in detail, developer representatives also emphasized the importance of designing DiGA to be simple and intuitive to use without requiring substantial additional time from health care providers.

Trial access for patients was also discussed from different perspectives. SHI representatives were generally open to this approach and considered it a simple way of increasing transparency and allowing patients to gain an impression of an application before regular use.

> *Perhaps also for the insured, so that they can simply see what the product actually is, i.e., get to grips with it.* [SHI representative 1]

DiGA developer representatives were more critical. They emphasized that indications and contraindications would need to be clarified before access could be provided. They also raised concerns about reimbursement, arguing that a DiGA may already provide benefits during an initial trial period and that these benefits would not be adequately reimbursed if access were provided before the start of regular reimbursed use.

Information Suppliers

Participants also discussed which stakeholders could provide information on DiGA.

DiGA Developers

DiGA developers were discussed as an important but also controversial source of information. For developer representatives, providing information was of particular importance as it also formed part of the marketing of their products. They described focusing their efforts primarily on health care providers, as this was considered a more efficient use of their limited resources than directly addressing patients.

> *For us, it's also a little bit of a resource issue whereby we have to look at how we can use the manpower or resources that we have wisely. We are working more with doctors right now.* [DiGA developer representative 2]

It was also noted by developer representatives that a significant proportion of DiGA developers are small companies. This means that they have fewer resources for information provision and sales activities than larger companies, such as pharmaceutical manufacturers. Nevertheless, health care provider representatives expressed criticism of the primary reliance on developers for product-specific information, as they questioned the objectivity of information that also serves marketing purposes.

SHI Funds, Chambers, and Professional Associations

DiGA developer representatives also expressed the view that other stakeholders, particularly SHI funds and chambers, should contribute more to information provision. SHI funds were considered to have established communication channels through which information on DiGA could reach insured individuals. SHI representatives themselves emphasized, however, that their information activities were subject to requirements of objectivity, limiting their ability to provide product-specific recommendations.

At the same time, DiGA developer representatives questioned the neutrality of SHI funds in this context, referring to competing in-house or selective contractual products that may be similar to DiGA. SHI representatives also described high DiGA prices and the availability of similar products as reasons for being reluctant to proactively promote DiGA to insured individuals.

> *For example, we could mobilize our call centers. And if we get a call from somebody who needs mental health counselling somewhere, we could say, look here, there's a DiGA in this area, check it out. But we don't do that and no SHI fund will do that because of the prices, because of the high prices. Maybe also because there are similar products, which might also play a role.* [SHI representatives 3 and 4]

Medical and psychotherapeutic chambers were also discussed as potential information providers. Health care provider representatives were generally open to such involvement. However, representatives of the chambers described similar limitations regarding product-specific information and emphasized their neutral role. They considered their involvement particularly appropriate when concerns regarding individual DiGA arose but did not see DiGA-specific training as part of their role.

> *If we have a reason, for example if there is imminent danger, or if we see that there is a DiGA and patients are suffering, then we can intervene. But we can't do training on a single DiGA.* [Health care provider representative 5]

Professional associations were discussed more positively as potential information providers. DiGA developer and health care provider representatives described them as having established communication channels as well as the expertise and capacity to assess relevant information and communicate

XSL•FO
RenderX

it to health care providers. Health care provider representatives also referred to the role professional associations had played in communicating previous changes in the health care system.

> *I believe that the professional associations in particular must play an important multiplier role here.* [DiGA developer representative 3]

Overall, participants attributed different opportunities and limitations to the potential information providers. Developers were able to provide detailed product-specific information but faced resource constraints and concerns regarding objectivity, whereas SHI funds and chambers described limitations related to their neutral role. Professional associations were viewed as potential intermediaries with established communication channels and relevant professional expertise.

Discussion

Principal Findings

This article examines the role of information in the implementation process of DiGA in outpatient mental health care in Germany. Health care provider representatives described a general awareness of DiGA, while more detailed knowledge, for example, required to differentiate between individual applications and assess their evidence, was often perceived as limited. Information needs therefore appear to extend beyond raising general awareness and include both system-level information on regulatory requirements and organizational processes and more detailed information that enables the assessment and comparison of individual DiGA. In addition, the results show that the source of information is relevant, as different potential information providers were associated with different opportunities and problems. Together, these findings suggest that the information challenge surrounding DiGA concerns not only awareness but also the type of information provided and those responsible for providing it.

The distinction between general awareness and more detailed knowledge found in our interviews is also reflected in previous research. In a survey of German general practitioners, Wangler and Jansky [7] found that 87% of respondents were aware that DiGA could be prescribed, whereas only 22% felt confident in competently advising patients about DiGA and only 25% reported having a general overview of the available applications. More recent qualitative studies point in a similar direction. Sahan et al [8] identified persistent knowledge gaps among general practitioners despite experience with DiGA prescriptions, while Panitz et al [9] reported moderate knowledge among psychological psychotherapists and a shared need for further information, including information on evidence, data privacy, and prescription procedures. The distinction between general awareness and more detailed knowledge can be further interpreted using Rogers' Diffusion of Innovations Theory, which describes innovation adoption as a process comprising 5 stages [18]. The initial knowledge phase is primarily concerned with becoming aware of an innovation and developing a basic understanding of how it functions. During the subsequent persuasion phase, individuals seek more comprehensive information and form a positive or negative attitude toward the innovation. In the decision phase, a decision is made to adopt or reject the innovation, followed by the implementation and confirmation phases. Particularly during the first 3 stages, limited knowledge and experience can create uncertainty, while the type and depth of information required change throughout the adoption process [18]. When evaluated in relation to the aforementioned framework, the results of the present study indicate that health care providers have information requirements that extend beyond the initial stage and align with the persuasion and decision phases. Consequently, measures designed solely to raise awareness may be inadequate. Health care providers require more detailed information to compare DiGA, assess their evidence and suitability, and make informed judgments about their use in routine care.

Information about the evidence supporting individual DiGA emerged as an important topic. Health care provider representatives particularly expressed uncertainty regarding provisionally listed DiGA and the differences in evidence requirements according to listing status. Similar concerns have been reported by Schroeder et al [19], whose qualitative study among German health care providers identified uncertainty regarding the evidence requirements for DiGA and particularly provisionally listed applications. Understanding these regulatory differences is important, but our findings also indicate a need for information that supports the assessment and comparison of DiGA. This is consistent with the work of Mäder et al [20], who identified limitations in the transparency and completeness of evidence-related information in the DiGA directory, emphasizing the importance of information that supports health care providers in assessing applications. More structured and comparable information could therefore facilitate the evaluation of DiGA in routine care. Continuing medical education could complement this by providing more general knowledge about DiGA and their organizational and regulatory framework. Similarly, Wangler and Jansky [7] found that 52% of surveyed general practitioners considered systematic further training to be a promising approach. However, our findings also indicate limitations of general educational formats for conveying detailed information on individual applications and suggest that developer-provided information may be perceived critically because information provision and marketing overlap.

Trial access was another approach discussed in the interviews to provide more detailed, product-specific information. By allowing health care providers to familiarize themselves with the content and functioning of individual DiGA, trial access may increase transparency and support the evaluation of an application before recommending or prescribing it. This is consistent with previous findings, in which the absence of testing options was identified as a barrier among psychotherapists [9]. The German Society for Internal Medicine has similarly called for long-term access to test accounts for physicians [21]. In relation to Rogers' model, such opportunities may be particularly relevant during the decision phase, when individuals seek additional information to evaluate an innovation before deciding for or against its adoption [18]. However, the interviews also highlighted practical limitations of this approach. A detailed assessment of individual DiGA is time-consuming, particularly if health care providers need to familiarize themselves with

XSL•FO
RenderX

several applications within the same indication. Similarly, Kendziorra et al [22] found that the time required to acquire and evaluate information regarding DiGA was a significant barrier for health care providers, hindering their integration into routine practice. Furthermore, developer representatives in our study questioned the extent to which health care providers' own experience with a DiGA can reflect the experience of patients for whom the application was developed. Trial access may therefore complement product-specific information by providing direct insight into an application, but it cannot substitute for information on evidence, suitability, and patient experience.

The limitations of individual product assessment also raise the question of which organizations could support implementation by providing and evaluating information. Our interviews showed that different information providers were associated with different opportunities and limitations. DiGA developers can provide detailed product-specific information, but health care provider representatives also questioned the objectivity of information that simultaneously serves marketing purposes. SHI funds and chambers, in contrast, have established communication channels but described limitations in providing product-specific information due to their neutral role. Against this background, professional associations were considered relevant potential intermediaries, as they combine professional expertise with established channels for communicating information to health care providers. This is consistent with Sahan et al [8], whose interviewed general practitioners also identified professional associations among the preferred sources of DiGA-related information and more broadly emphasized the importance of independent information sources.

This role can also be interpreted in light of Rogers' Diffusion of Innovations Theory [18]. Rogers emphasizes that, while media and other impersonal communication channels are particularly relevant for creating initial awareness of an innovation, interpersonal channels become increasingly important when individuals form an attitude toward an innovation and decide whether to adopt it. In this context, evaluations and experiences communicated by peers and opinion leaders can influence adoption decisions. Professional associations could therefore act as intermediaries by assessing and structuring available information and communicating it through established channels. More recently, digital health navigators represent another possible form of such an intermediary role. Freisberg et al [23] described digital health navigators as potential links between patients, health care professionals, and digital interventions and emphasized the need for specific digital, technical, clinical, and communication competencies to fulfill this role. Similarly, Schwarz et al [24] proposed digital navigators as specially trained medical assistants who support health care providers in selecting digital interventions and assist patients during their use.

Limitations

Several limitations should be considered when interpreting the findings. First, the sample comprised 14 participants representing 4 different stakeholder groups. Given this relatively small and heterogeneous sample, the range of perspectives captured within each individual stakeholder group is limited, and the findings should be understood as an exploratory account of stakeholder perspectives. In addition, sociodemographic characteristics and individual previous experience with DiGA were not systematically collected, as participants were primarily recruited on the basis of their institutional roles and professional expertise. Consequently, potential differences in perspectives according to these characteristics or previous DiGA experience could not be examined. However, including health care provider representatives, SHI representatives, DiGA developer representatives, and a patient representative meant that perspectives from different relevant positions within the DiGA care process could be examined within one analysis. Representation across stakeholder groups was unequal, with only one patient representative included. Consequently, this perspective is particularly limited, and some statements concerning patients reflect the perceptions of other stakeholders. Although this does not compensate for the limited patient representation in the present analysis, patient perspectives were examined more extensively in other components of the wider DiGAPsy project, including a focus group study with 31 patients and a cross-sectional survey involving 1413 patients.

Second, the target sample size was predefined during project planning, and no formal assessment of theoretical or thematic saturation was conducted. The study therefore does not claim that all potentially relevant perspectives on information provision were captured. Although the purposive sampling strategy allowed the inclusion of predefined stakeholder perspectives, the recruitment of participants through selected institutions may have introduced selection effects. The use of semistructured interview guides also provided a common thematic framework across stakeholder groups, although the interview guides may have influenced which aspects were discussed in greater depth. Finally, the present article deliberately focuses on information-related aspects of a broader interview study. Other implementation barriers and organizational dimensions were therefore outside the scope of this focused analysis rather than being interpreted as less relevant.

Third, all interviews and the qualitative analysis were conducted in German, whereas selected quotations were translated into English for publication. Although translation can lead to subtle changes in meaning, the translated quotations were checked against the original German statements to preserve their substantive content. The findings also need to be interpreted within the specific German DiGA framework, which differs from regulatory models for digital health interventions in other countries and may limit transferability. Nevertheless, the study provides a detailed account of information-related implementation issues within an established reimbursement system and may therefore also inform discussions about information provision for digital health interventions beyond the German context.

Conclusion

This study provides insights into stakeholder perspectives on information-related aspects of the implementation of DiGA in outpatient mental health care in Germany. The findings suggest that health care providers require more detailed and accessible

XSL•FO
RenderX

information to assess and compare the evidence, characteristics, and suitability of individual applications.

The findings highlight practical opportunities for enhancing the existing provision of information. In particular, the DiGA directory could be developed further to make it more user-centered and facilitate access to structured, comparable information on individual DiGA. The addition of standardized explanatory or demonstration videos could provide patients and health care providers with a quick overview of an application's content, structure, and functionality. These measures could complement existing information formats and support a user-friendly assessment of individual DiGA. However, information provision represents only one component of the broader implementation process.

Acknowledgments

The authors acknowledge support from the Open Access Publication Fund of the University of Duisburg-Essen and thank the fund for its support.

The authors declare the use of generative AI (GenAI) in the research and writing process. According to the GAIDeT (Generative AI Delegation Taxonomy; 2025), the following tasks were delegated to GenAI tools under full human supervision: literature search and systematization, proofreading and editing, translation, reformatting, quality assessment, and publication support. The GenAI tools used were ChatGPT-5.4 (OpenAI) and DeepL.

Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Funding

This study was funded by the Innovation Fund of the German Federal Joint Committee (G-BA; funding code 01VSF22029). The funder was not involved in the study design, data collection, data analysis, interpretation of the results, manuscript preparation, or the decision to submit the manuscript for publication.

Data Availability

The qualitative interview data generated and analyzed during this study are not publicly available because the transcripts contain contextual information that could potentially allow identification of individual participants or represented organizations. Due to these confidentiality constraints, the interview transcripts are not available for external access, including upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
COREQ checklist.
[PDF File (Adobe PDF File), 248 KB - humanfactors_v13i1e82514_app1.pdf]

Multimedia Appendix 2
Complete interview guides.
[DOCX File , 20 KB - humanfactors_v13i1e82514_app2.docx]

Multimedia Appendix 3
Overview of the conducted interviews and stakeholder groups.
[DOCX File , 15 KB - humanfactors_v13i1e82514_app3.docx]

Multimedia Appendix 4
List of codes.
[DOCX File , 24 KB - humanfactors_v13i1e82514_app4.docx]

References

1. Jacobi F, Höfler M, Siegert J, Mack S, Gerschler A, Scholl L, et al. Twelve-month prevalence, comorbidity and correlates of mental disorders in Germany: the Mental Health Module of the German Health Interview and Examination Survey for Adults (DEGS1-MH). Int J Methods Psychiatr Res 2014;23(3):304-319 [FREE Full text] [doi: 10.1002/mpr.1439] [Medline: 24729411]

XSL•FO
RenderX

2. Gesundheitsreport 2024: Arbeitsunfähigkeiten. Techniker Krankenkasse. URL: https://www.tk.de/resource/blob/2168508/fd1b72979121501d601808bc2a45fda7/gesundheitsreport-2024-data.pdf [accessed 2026-09-07]
3. Schlierenkamp S, Werner S, Schlesiger P, Friedrich L, Abels C, Höfer K, et al. Treatment pathways in the outpatient care of patients with depression: an analysis of statutory health insurance claims data. Psychiatr Prax 2025;52(8):442-450 [FREE Full text] [doi: 10.1055/a-2712-9615] [Medline: 41224096]
4. DiGA-Leitfaden: Stand 10.12.2025, Version 3.6. Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM). URL: https://www.bfarm.de/SharedDocs/Downloads/DE/Medizinprodukte/DiGA_leitfaden [accessed 2026-09-07]
5. DiGA-Verzeichnis. DiGA-Verzeichnis, Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM). URL: https://DiGA.bfarm.de/de [accessed 2026-09-07]
6. DiGA-Bericht des GKV-Spitzenverbandes: Berichtszeitraum 1. September 2020 bis 31. Dezember 2025. GKV-Spitzenverband (National Association of Statutory Health Insurance Funds). 2026. URL: https://www.gkv-spitzenverband.de/media/dokumente/krankenversicherung_1/telematik/digitales/2025_DiGA_Bericht_GKV_SV.pdf [accessed 2026-09-07]
7. Wangler J, Jansky M. What potential and added value do DiGA offer for primary care? Results of a survey of general practitioners in Germany. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2022;65(12):1334-1343 [FREE Full text] [doi: 10.1007/s00103-022-03608-w] [Medline: 36269336]
8. Sahan F, Panitz K, Wagenaar C, Kairies-Schwarz N, Vomhof M, Zinn M, et al. Genie in the bottle? A qualitative study of general practitioners' perspectives and information needs concerning digital mental health applications in Germany. BMC Prim Care 2025;26(1):393. [doi: 10.1186/s12875-025-03115-2] [Medline: 41366646]
9. Panitz K, Wagenaar C, Sahan F, Kairies-Schwarz N, Guthardt L, Bau JT, et al. Perspectives on digital therapeutic prescribing: a qualitative study among German psychological psychotherapists. Front Digit Health 2026;8:1656614 [FREE Full text] [doi: 10.3389/fdgth.2026.1656614] [Medline: 41737814]
10. Christine Eicher S, Cohrdes C. [Awareness, use and acceptance of digital health applications (DiGA) for mental health in the general adult population in Germany]. Gesundheitswesen 2026;88(5):330-336. [doi: 10.1055/a-2711-1021] [Medline: 41285397]
11. Giebel G, Höfer K, Raszke P, Wasem J, Bußmann A, Schlierenkamp S, et al. Access to digital health applications: prescription by physicians vs patient requests made directly to health insurance companies: interviews with experts. Gesundheitswesen 2026;88(7):455-461 [FREE Full text] [doi: 10.1055/a-2800-1191] [Medline: 42235549]
12. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
13. Höfer K, Plescher F, Schlierenkamp S, Solar S, Neusser S, Schneider U, et al. mHealth apps in German outpatient mental health care: protocol for a mixed methods approach. JMIR Res Protoc 2024;13:e56205 [FREE Full text] [doi: 10.2196/56205] [Medline: 39657173]
14. Plescher F, Höfer K, Wasem J, Bußmann A, Solar S, Minor M, et al. Concepts for the integration and implementation of mHealth apps for patients with mental disorders: scoping review. J Med Internet Res 2025;27:e66340 [FREE Full text] [doi: 10.2196/66340] [Medline: 40921068]
15. Volland C, Plescher F, Raszke P, Friedrich L, Speckemeier C, Schlierenkamp S, et al. Prescriber's preferences for digital health applications in mental health care: cross-sectional best-worst scaling study of general practitioners and psychotherapists in Germany. J Med Internet Res 2026;28:e99203 [FREE Full text] [doi: 10.2196/99203] [Medline: 42418252]
16. Claussen J, Jankowski D, Dawid F. Aufnehmen, Abtippen, Analysieren: Wegweiser zur Durchführung von Interview und Transkription. Hannover, Germany: BoD–Books on Demand; 2020.
17. Kuckartz U. Qualitative Inhaltsanalyse: von Kracauers Anfängen zu heutigen Herausforderungen [Qualitative content analysis: from Kracauer's beginnings to today's challenges]. Forum Qual Sozialforsch 2019;20(3). [doi: 10.17169/fqs-20.3.3370]
18. García?Avilés J. Diffusion of innovation. In: The International Encyclopedia of Media Psychology. Hoboken, New Jersey: Wiley; 2020:1-8.
19. Schroeder T, Haug M, Georgiou A, Seaman K, Gewald H. Evidence of how physicians and their patients adopt mHealth apps in Germany: exploratory qualitative study. JMIR Mhealth Uhealth 2024;12:e48345 [FREE Full text] [doi: 10.2196/48345] [Medline: 38231550]
20. Mäder M, Timpel P, Schönfelder T, Militzer-Horstmann C, Scheibe S, Heinrich R, et al. Evidence requirements of permanently listed digital health applications (DiGA) and their implementation in the German DiGA directory: an analysis. BMC Health Serv Res 2023;23(1):369 [FREE Full text] [doi: 10.1186/s12913-023-09287-w] [Medline: 37069592]
21. DGIM beklagt fehlende DiGA-Testmöglichkeiten für Ärzte. Deutsches Ärzteblatt. URL: https://www.aerzteblatt.de/news/dgim-beklagt-fehlende-DiGA-testmoeglichkeiten-fuer-aerzte-2eddce6a-2700-4ced-b3bf-e0a2a2bcb798 [accessed 2026-09-29]
22. Kendziorra J, Seerig KH, Winkler TJ, Gewald H. From awareness to integration: a qualitative interview study on the impact of digital therapeutics on physicians' practices in Germany. BMC Health Serv Res 2025;25(1):568 [FREE Full text] [doi: 10.1186/s12913-025-12656-2] [Medline: 40247268]
23. Freisberg L, Meier-Diedrich E, Heinze M, Schönbeck J, Schubert D, Speck J, et al. Expected competencies and personal attributes of digital health navigators to support digital mental health care: focus group and interview study with patients

XSL•FO
RenderX

and health care professionals. JMIR Ment Health 2026;13:e83073 [FREE Full text] [doi: 10.2196/83073] [Medline: 42024869]
24. Schwarz J, Chen K, Dashti H, Heinze M, Schönbeck J, Schubert D, et al. Piloting digital navigators to promote acceptance and engagement with digital mental health apps in German outpatient care: protocol for a multicenter, single-group, observational, mixed methods interventional study (DigiNavi). JMIR Res Protoc 2025;14:e67655 [FREE Full text] [doi: 10.2196/67655] [Medline: 40996088]

Abbreviations

BfArM: Bundesinstitut für Arzneimittel und Medizinprodukte
COREQ: Consolidated Criteria for Reporting Qualitative Research
DiGA: digitale gesundheitsanwendung
DiGAPsy: Realizing the Potential of Digital Health Applications in Outpatient Mental Health Care
mHealth: mobile health
SHI: statutory health insurance

Edited by S Law; submitted 17.Aug.2025; peer-reviewed by J Schwarz, M Scheibe, D Prinz; comments to author 25.Jun.2026; revised version received 07.Sep.2026; accepted 25.Sep.2026; published 04.Nov.2026.

Please cite as:
Höfer K, Raszke P, Wasem J, Plescher F, Bußmann A, Schlierenkamp S, Minor M, Friedrich L, Volland C, Zander S, Wadeck A, Schneider U, Katzmann J, Maaß E, Abels C
Stakeholder Perspectives on Information About Digital Health Applications in German Outpatient Mental Health Care: Qualitative Interview Study
JMIR Hum Factors 2026;13:e82514
URL: https://humanfactors.jmir.org/2026/1/e82514
doi:10.2196/82514
PMID:

©Klemens Höfer, Pascal Raszke, Jürgen Wasem, Felix Plescher, Anna Bußmann, Sarah Schlierenkamp, Michael Minor, Luisa Friedrich, Carsten Volland, Sophia Zander, Anja Wadeck, Udo Schneider, Josepha Katzmann, Enno Maaß, Carina Abels. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 04.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Reducing Information Uncertainty in Hospital Wayfinding for People With Specific Needs: Findings From a Qualitative, Multimethod Study

Ladislava Zbiejczuk Suchá[1], PhD; Kateřina Jochecová[1], PhD; Radek Pavlíček[2], MA; Petr Červenka[2], BA; Karel Sobol[2], ING; Henrieta Greplová[1], BA; Gabriela Godišková[1], MA; Alica Kúdelová[1], BA; Pavla Pospíšilová[1], MA; Ondřej Kvarda[1], PhD; Natálie Káčová[1], MA; Michaela Holubec Birtusová[3], MA

[1]Department of Information and Library Studies, Faculty of Arts, Masaryk University, Brno, South Moravia, Czech Republic

[2]Support Centre for Students with Special Needs, Masaryk University, Brno, South Moravia, Czech Republic

[3]CZECHDESIGN Center CZ, Prague, Czech Republic

Corresponding Author:
Ladislava Zbiejczuk Suchá, PhD
Department of Information and Library Studies
Faculty of Arts
Masaryk University
Arna Nováka 1
Brno, South Moravia, 60200
Czech Republic
Phone: 420 777035620
Email: sucha@phil.muni.cz

Abstract

Background: Hospital wayfinding is a well-documented source of patient stress, yet people with sensory, cognitive, or mobility-related needs face additional information barriers that generic wayfinding research rarely resolves. Information uncertainty, rather than a simple lack of signage, has recently been proposed as a unifying construct for locating and diagnosing these barriers across the patient journey.

Objective: This study empirically applies a Stage-Touchpoint-Uncertainty coding framework across 2 structurally contrasting hospitals to identify the elements that constitute information uncertainty during hospital navigation, how patients with specific needs perceive and prioritize these elements, and how the resulting uncertainty peaks can inform the prioritization of wayfinding interventions.

Methods: We conducted a qualitative, multimethod study with 4 rounds of data collection between May 2024 and July 2025 at 2 university hospitals that differ in scale and spatial organization (hospital A: more than 1000 beds, decentralized multipavilion campus; hospital B: 890-bed comprehensive center with a historic core). Methods comprised expert accessibility walkthroughs, website accessibility and usability testing with users with severe visual impairment, user walkthroughs with mobile eye-tracking and think-aloud protocols, and focus groups homogeneous by impairment type (hearing, developmental/learning, physical/mobility, and older adults) to evaluate candidate signage designs. In total, 38 participants took part (36 research participants and 2 accessibility experts who conducted the heuristic walkthroughs). The mean age of user-walkthrough and focus-group participants was 38.0 (range 21-75) years; age was not recorded for experts or website-testing participants.

Results: Coding organized observations into a Stage-Touchpoint-Uncertainty matrix spanning 5 consolidated journey stages, from digital prearrival search through physical navigation to exit. Four recurring uncertainty mechanisms emerged: absence, where a needed cue is simply missing; equivocality, where cues are present but fail to converge on one meaning; perceptual degradation, where a cue is legible in principle but difficult to see or read; and cognitive-translation demand, where a legible cue still requires additional mental work to act on. Absence characterized the earliest stages, involving digital search and entry, whereas equivocality became dominant once participants were navigating the building physically and remained the most frequently coded pattern through exit. Priorities also diverged by impairment group and data source: participants with visual impairment prioritized the digital retrievability of information, whereas focus-group participants with developmental/learning impairments prioritized redundant multichannel coding, and those with physical/mobility impairments prioritized functional-eligibility information.

XSL•FO
RenderX

Conclusions: Information uncertainty in hospital wayfinding is best understood along 2 axes—the stage of the journey at which it occurs and the mechanism that produces it. Together, these axes support wayfinding interventions that target specific stages and are tailored to specific types of breakdown, with potential transferability to comparable hospital settings beyond the 2 studied here. Breakdowns corroborated across multiple sources and impairment groups should be prioritized first.

(JMIR Hum Factors 2026;13:e106882) doi:10.2196/106882

KEYWORDS

wayfinding; hospital accessibility; information uncertainty; patient journey mapping; usability testing; universal design; assistive technology; disability

Introduction

Background

Wayfinding is a key part of the patient experience. It is 2-fold in nature, representing both a design element and a medical problem. Well-designed wayfinding systems in hospitals offer numerous benefits, including reduced stress and anxiety for patients and visitors [1,2]. Effective wayfinding can save costs by decreasing missed appointments, minimizing staff time spent giving directions, and improving overall efficiency [1,3], as well as reducing complaints [4]. Implementing clear signage with symbols and pictograms can enhance patient outcomes and staff satisfaction [3]. Digital wayfinding systems can further improve the user experience by providing interactive guidance and reducing task complexity [5,6]. A holistic approach to wayfinding design, incorporating multiple communication channels and considering the broader health care journey, can empower health seekers and improve their overall experience [7,8]. Ultimately, well-designed wayfinding systems contribute to a more comfortable and user-friendly hospital environment [9]. Several studies point out that supporting wayfinding also involves paying attention to digital tools, such as information on hospital websites, digital navigation systems, and other specialized applications [10,11]. However, digital touchpoints introduce distinct accessibility challenges of their own. Usability testing of health-related digital systems with assistive technology users has repeatedly identified structural and navigational barriers that persist even on well-resourced digital touchpoints, suggesting that digital wayfinding tools cannot be assumed to be accessible by default [12].

Wayfinding in Hospitals in the Context of Specific Patient Needs

In recent years, several studies have addressed the issue of wayfinding in hospitals [13]. Researchers continue to examine how various groups navigate and orient themselves in physical environments, with a strong emphasis on both their strategies and effectiveness. However, the diverse needs of users, including those with special needs, must be taken into account in both research and design [6]. To address the challenges that arise from differences in spatial perception, studies have focused on diverse populations, such as older adults [14,15], children [16], and individuals with specific needs, including those with visual impairments [17], mobility impairments [18], autism [16], dementia [19,20], and other disabilities, to identify solutions that improve their wayfinding experiences. Methodologically, understanding how these populations navigate has increasingly relied on eye-tracking and think-aloud protocols, which allow researchers to capture attentional and behavioral evidence of navigation difficulties in situ rather than relying solely on retrospective self-reports. However, these approaches remain underused in naturalistic, real-world settings compared with controlled laboratory settings [21].

Information Uncertainty in Wayfinding

Previous studies conceptualized wayfinding as spatial problem-solving involving information processing, decision-making, and plan execution [22]. Uncertainty in wayfinding can arise from multiple factors, such as environmental complexity, the legibility and effectiveness of navigational aids, and individual differences in spatial cognition. Information uncertainty may result from gaps in information that patients experience in hospital settings [23]. Uncertainty significantly influences wayfinding by affecting both cognitive processes and navigational behaviors [24,25]. A thorough understanding of the sources of information uncertainty and how they influence both cognitive and behavioral aspects of navigation is essential for designing effective wayfinding systems and built environments.

Patient Journeys as a Method for Mapping Patients' Experience

Patient journey mapping is an emerging method for understanding and improving patient experiences in health care systems [26]. It provides valuable insights into wayfinding challenges in complex hospital environments [7]. This approach can inform clinic placement, signage design, and resource allocation to enhance patient experience and operational efficiency [27] and optimize signage placement [28]. Studies have shown that patients benefit from simple building layouts, nature-themed visual displays in diagnostic settings, and a variety of social spaces during treatment, including when navigating health care facilities [29]. Building on a replicable Stage-Touchpoint-Uncertainty framework [30], this study applies the framework empirically across 2 hospitals and 4 data-collection methods, focusing specifically on patients with diverse impairment-related needs. The framework operationalizes information uncertainty in hospital wayfinding through a 7-stage patient journey model spanning digital and physical touchpoints. It also codes observable, time-stamped indicators of uncertainty rather than relying solely on retrospective self-report.

Patient journey mapping has also been proposed as a boundary object that helps multidisciplinary teams reconcile competing priorities in service design by externalizing tacit knowledge about the patient experience for stakeholders outside the clinical

XSL•FO
RenderX

field [31,32]. Digital applications of journey mapping in hospital settings—for instance, mobile tools guiding patients through the perioperative period—illustrate both the potential of structuring the patient journey around discrete stages and the persistent challenge of tailoring these structures to individual variations in needs [33].

Research Questions

While previous studies have emphasized the importance of understanding patient experience as a driver of health care service improvement, they have also highlighted the methodological challenge of capturing this experience holistically across the complex trajectory of hospital care [23]. Building on this line of inquiry, our study seeks to further unpack the patient journey by focusing on the constituent elements (touchpoints) that shape the hospital wayfinding experience. In doing so, we aim to address 3 research questions: (1) What elements constitute information uncertainty in the patient's experience during navigation and interaction with hospital services? (2) How are these elements perceived and prioritized by patients with specific needs? and (3) How can the uncertainty peaks identified across the patient journey inform the design and prioritization of wayfinding interventions for people with specific needs?

Taken together, these strands of literature suggest that information uncertainty in hospital wayfinding is shaped jointly by physical and digital touchpoints, the user's specific perceptual and cognitive profile, and the methodological lens used to observe it. However, few studies have combined field-based behavioral methods, website accessibility testing, and evaluative design methods within a single, cross-impairment framework.

Methods

Study Design and Hospital Settings

This study applies the patient journey framework for mapping information uncertainty in hospital wayfinding [30] to 2 university hospitals that differ in scale and spatial organization. We adopted a qualitative, multimethod design with methodological triangulation across 4 rounds of data collection between May 2024 and July 2025. The design combines expert heuristic walkthroughs, accessibility and usability testing of the hospital website, user walkthroughs with mobile eye-tracking accompanied by think-aloud protocols and posttask interviews, and focus groups evaluating candidate signage designs. Each method captures a different layer of the wayfinding experience—professional assessment, digital interaction, embodied use, and deliberative evaluation—consistent with our conceptualization of wayfinding as a sociotechnical, cross-channel information behavior [34,35].

Data were collected at 2 large hospitals selected for their contrasting wayfinding contexts. Hospital A has more than 1000 beds and is organized as a sprawling, multipavilion campus on a hillside site. Approximately 25 departments and 7 clinics are distributed across numerous freestanding historic pavilion buildings dating from the hospital's original construction between 1926 and 1934. These buildings are connected by outdoor pathways and later infill buildings rather than a single unified structure, resulting in a decentralized, low-legibility spatial layout typical of large teaching hospitals that have grown incrementally. Hospital B is an 890-bed university hospital comprising 39 specialized departments and designated as a comprehensive cardiovascular, cerebrovascular, and oncology center. It is organized around a historic core in the old town district, with the first building completed in 1868 and several pavilions added later for cardiac surgery, intensive care, and dialysis. This spatial organization reflects historical growth on a constrained inner-city site and the partial relocation of departments to a newer campus.

Ethical Considerations

The study protocol was reviewed and approved by the Ethics Committee of Masaryk University (approval number EKV-2022-121) in accordance with relevant laws and institutional guidelines. All participants provided written informed consent before each session and received a financial reward of approximately €20 (US $22.45).

Analytical Framework

The analysis applies the Stage-Touchpoint-Uncertainty coding scheme [30]. The journey stages, spanning digital and physical contexts, were consolidated into 5 groups:

- Prearrival (digital orientation): hospital website, department page, and route-planning tools
- Entry and initial orientation: vestibule/lobby, reception/volunteer desk, and first decision node
- En route navigation: corridors, junctions, vertical circulation, and confirmation cues
- Destination interface: department front of house (doors, counters, and check-in)
- Exit and postvisit: Return path, payment and administration, discharge, and follow-up information

For each stage, the analysis enumerates 2 classes of wayfinding touchpoints: digital (web pages, map widgets, PDF maps, contact information, and accessibility notes) and physical (exterior totems, building identifiers, building numbers, directories, directional signs, confirmation signs, landmarks, floor graphics, elevator call panels, and door labels). Social and organizational touchpoints (reception desks and ad hoc staff guidance) are outside the scope of this study, as the analytical focus is on the artifactual layer of wayfinding. Help-seeking interactions with other people and the use of personal devices were not allowed during walkthroughs.

The unit of analysis is a stage-bounded interaction with a wayfinding touchpoint—an observable episode in which a participant engages with a digital or physical cue at a specific journey stage (eg, entry and initial orientation: "reads lobby directory for 18 s and selects elevator A"). Each unit is coded along 5 indicator families: task state, operationalizing the microprocess of wayfinding [22,23]; uncertainty markers, directly observable in eye-tracking and audio data [24,25]; touchpoint qualities, translating environmental communication principles into codable variables [3,28]; accessibility factors, surfacing equity-relevant issues across digital and physical channels [11,17]; and context modifiers, reducing attribution

XSL•FO
RenderX

error when failures are driven by the environment rather than the cue itself [4].

Coded units are aggregated into a Stage-Touchpoint-Uncertainty matrix that overlays task states and quality flags by stage, enabling the identification of uncertainty peaks, recurring breakdown patterns, and impairment-specific issues across participants and sites. As the number of coded events differs across stages and hospitals for reasons related to route length and structure rather than uncertainty severity (see the "Data Sources and Procedures" section), an uncertainty peak is defined here not as a raw event-count maximum but as a stage-mechanism combination corroborated by multiple independent data sources or affecting a majority of participants within that stage. The 4-mechanism typology (absence, equivocality, perceptual, and cognitive-translation) used throughout this paper was applied as an interpretive classification by the coding team to each stage-touchpoint pattern, rather than logged as a discrete, systematically quantified code for every individual event.

Data Sources and Procedures

Overview

The study comprised 4 rounds of data collection, summarized in Table 2. Each round is described below.

Expert Walkthroughs

In May 2024, 2 accessibility and inclusion experts (E1, experts with mobility impairment, and E2, experts with visual impairment) conducted heuristic walkthroughs of hospital A to establish a baseline of design weaknesses before user testing. The protocol included 3 approach routes to the hospital (from public transport and the parking lot), 3 routes between buildings within the hospital complex, and 1 indoor route. Each route comprised several destination departments arranged in sequence. The experts documented their observations through structured field notes, photographs, and short postwalk debriefs with the research team. Findings from this round informed the design of subsequent user walkthroughs and provided a heuristic comparison layer for triangulation.

Conducting expert-led field walkthroughs of the physical hospital environment follows established practice in accessibility assessment of the built environment, commonly termed an access audit. Access audits typically involve a structured assessment of a building along the journey of a user with disability, from arrival at the site to every area they might reasonably need to access. This approach is reflected in our protocol, which combined 2 approach routes (from public transport and the parking lot), 3 interbuilding routes, and 1 indoor route. This journey-based approach is consistent with the broader conceptualization of accessibility as a relational property rather than a fixed attribute of the building alone. Iwarsson and Ståhl [36] proposed an instrumental definition of accessibility that integrates personal and environmental components, arguing that accessibility must be analyzed through the interaction between the 2 rather than through the environment in isolation. This framing justified our selection of 2 experts with complementary profiles (ie, with mobility/visual impairment) rather than a single generalist auditor, as each expert's assessment reflects a different facet of the person-environment relationship that a single evaluator could not fully represent. Consistent with standard access-audit methodology, an access audit evaluates how well an environment and its service delivery meet the needs of existing and potential users by assessing features against predetermined criteria, thereby identifying physical barriers that may compromise access to services. This approach underpinned our experts' structured field notes and photographic documentation, which subsequently served as the heuristic baseline against which findings from later user walkthroughs were contextualized.

Website Accessibility and Usability Testing

In September 2024, the hospital A website was tested for digital accessibility and usability with 4 users with severe visual impairment (2 women and 2 men) of varying ages. The testers used different assistive technologies—JAWS (Freedom Scientific), NVDA (NV Access), VoiceOver (Apple Inc), and ZoomText (Ai Squared/Freedom Scientific)—with moderate to advanced proficiency. Selecting users with high assistive technology proficiency was a deliberate sampling strategy to minimize confounding arising from limited technology skills, allowing observed breakdowns to be attributed primarily to the website rather than to user competence.

Testing was conducted manually with experienced screen reader users rather than through automated conformance checking, following a methodological distinction established in a foundational comparative study of web accessibility evaluation methods. Mankoff et al [37] compared a laboratory study involving users with blindness with an automated tool, expert reviews conducted by designers with and without a screen reader, and remote testing. Focusing on techniques for evaluating accessibility for users with blindness, they found that multiple developers using a screen reader were the most consistently successful at identifying different classes of accessibility problems, although they identified only around half of the known problems [37]. This finding supports our decision to prioritize testing with real assistive technology users (JAWS, NVDA, VoiceOver, and ZoomText) over an automated audit, while also tempering expectations that a single testing round would be exhaustive. We address this limitation through triangulation with expert walkthroughs and user walkthroughs at the physical site.

Testers completed 10 task scenarios reflecting common prearrival information needs, such as locating a specific clinic, identifying the correct entrance, and finding parking and accessibility information. The sessions were unmoderated, and screen and audio recordings were captured throughout. Participants were asked to think aloud during testing and complete a structured report afterward.

User Walkthroughs With Mobile Eye-Tracking

Between June 2025 and July 2025, participants completed accompanied walkthroughs of hospital A and hospital B. Routes at hospital A were traversed within a single large, multistory pavilion (building 3), accessed via 2 elevators and 2 stairwells across 6 floors (basement to fourth floor). All participants

XSL•FO
RenderX

completed the same fixed sequence of 5 destination tasks: (1) the electroencephalogram unit within the neurology outpatient clinic; (2) the ENT (ear-nose-throat; otorhinolaryngology) outpatient clinic; (3) the specialized examination rooms of the ophthalmology department; (4) the urology department office; and (5) exiting the building via a route not previously used during the session. An additional brief directional probe ("Which direction is the phoniatrics unit?") was administered to assess signage-reading accuracy independently of overall wayfinding performance. Routes at hospital B spanned multiple interconnected buildings and, for most participants, followed a fixed sequence of 5-6 destination tasks: the gynecology department, the patient admissions office, the neurology outpatient clinic, an exterior building-identification task, the computed tomography scanning facility, and a return to the main reception. The exit tasks were tailored to the physical characteristics, spatial organization, and available patient through-routes of each site rather than scripted identically. Hospital A's walkthrough took place within a single multistory pavilion, making an exit via a previously unused route a meaningful test of independent orientation. By contrast, hospital B's decentralized, multibuilding layout meant that a return to the shared main reception better reflected the through-routes available to patients navigating that site.

The decision to conduct eye-tracking in situ using mobile rather than stationary equipment reflects a well-established methodological concern in spatial cognition research regarding the ecological validity of eye-tracking data. Mobile eye-tracking enables participants to move unrestrictedly through real environments, allowing researchers to examine the relationship between visual behavior and the characteristics of the physical environment [38] in ways that static, screen-based stimuli cannot. Concurrent think-aloud protocols were added specifically to address a known interpretive limitation of gaze data collected in complex, multicue environments, such as hospital corridors and junctions, where several plausible fixation targets (signs, doors, and landmarks) may be visible simultaneously, and gaze position alone cannot disambiguate participant intent.

Participants met the researchers at the main entrance of each hospital, where informed consent, financial compensation, and eye-tracking calibration were completed. Eye movements were recorded using SMI Eye Tracking Glasses 2, while concurrent audio captured think-aloud comments. Participants then navigated to a sequence of target destinations, each of which was communicated orally and provided on paper. One researcher accompanied each participant to prevent interference with hospital operations but did not provide directions or evaluative feedback during the task. Participants with hearing impairment who used sign language were additionally accompanied by an interpreter who was briefed in advance not to assist with navigation. Eye-tracking video served primarily to time-stamp and disambiguate moments of uncertainty identified in the concurrent think-aloud audio and posttask interviews rather than to provide independent quantitative gaze metrics (eg, fixation counts or dwell time). No systematic quantitative eye-tracking analysis is reported.

After the final destination was reached, the researcher conducted a semistructured posttask interview covering both general impressions of the wayfinding system and specific aspects of the participant's experience, including decisions made at junctions, sources of confusion, observed breakdowns, and instances of help-seeking. Each walkthrough, including the interview, lasted approximately 45 minutes.

Focus Groups

The team conducted 4 rounds of focus groups to evaluate 3 candidate navigation design proposals developed by external design studios. Group sizes ranged from 2 to 8 participants, and the groups were homogeneous by impairment type—hearing (n=8), developmental/learning (n=5), mobility (n=2), and older adults (n=4)—enabling each cohort to focus on needs specific to its modality. The focus groups evaluated candidate signage designs developed independently of the walkthrough findings. They were not used to validate or member-check specific breakdowns identified during the user walkthroughs but served as an independent evaluative data source.

Each session followed a 4-part protocol covering (1) pictogram comprehensibility, (2) the interior signage system, (3) the exterior signage system, and (4) the main complex map. Participants completed a short, structured scoring sheet indicating studio preference for each category and each pictogram and listing 5 descriptive adjectives for each design, followed by open deliberation. Focus groups involving participants with hearing impairment who used sign language were additionally supported by an interpreter. Sessions were audio-recorded, and the scoring sheets were retained as supplementary structured data.

Structuring focus groups to be homogeneous by impairment type follows long-standing guidance on focus-group composition in qualitative health research. Kitzinger's [39] foundational methodological paper on focus groups provides specific guidance on group composition, and homogeneous grouping is recommended because it allows group dynamics to surface shared, modality-specific concerns that might otherwise be diluted or reframed in a mixed group. This choice is further supported by participatory disability research, which argues that people with disabilities should evaluate design solutions directly rather than be represented by proxies. Bennett and Rosner [40] examined how empathy-based design activities, in which designers attempt to understand people with disabilities on their behalf, can serve as a powerful but potentially problematic substitute for direct engagement with the people the design is intended to serve. This argument is consistent with the disability rights principle "nothing about us without us" [41]. Our 4-group, impairment-specific approach operationalizes this principle by giving each cohort direct evaluative authority over the candidate signage systems rather than filtering its perspectives through a single mixed panel.

Participants

A total of 38 participants were included in the analysis across the 4 data sources (Table 1): 36 research participants who were involved in website testing, user walkthroughs, and focus groups, and 2 accessibility experts who conducted the heuristic

XSL•FO
RenderX

walkthroughs. These individuals were counted together because Table 1 organizes participation by data source rather than by role. Of the 15 participants originally recruited for the user walkthroughs, 13 were retained for coding: 1 session was excluded because of failed eye-tracking calibration, attributed to the participant's eyeglasses, and 1 recording was lost due to a technical fault. Table 1 presents the characteristics of the 13 user-walkthrough participants retained for coding. All walkthrough-based analyses reported in the "Results" and "Discussion" sections, including the 451 coded stage-touchpoint events, are based on these same 13 participants. Participants were recruited through the university's participant database, personal contacts, and interest groups representing people with specific needs. The sole exclusion criterion was prior familiarity with the layout of hospital A or hospital B. Familiarity was assessed alongside self-reported orientation skills and experience with unfamiliar environments using an online intake questionnaire.

To capture the heterogeneity of patient experiences and address concerns related to equity, inclusion, and access within health care systems, recruitment deliberately prioritized diversity across 3 axes: impairment type (visual, hearing, mobility, cognitive, neurodevelopmental, and learning), age, and assistive technology use. The resulting sample included participants with hearing impairment, visual impairment, dyslexia, dyscalculia, attention-deficit/hyperactivity disorder, autism spectrum disorder, mobility limitations, lung fibrosis, and cognitive impairment, distributed across the 4 data sources, as shown in Table 1.

Table 1. Participant characteristics by data collection method.[a]

Characteristic	Expert walkthroughs (n=2)	Website testing (n=4)	User walkthroughs (n=13)	Focus groups (n=19)	Total (N=38)
Gender, n (%)					
Male	2 (100)	2 (50)	7 (54)	8 (42)	19 (50)
Female	0 (0)	2 (50)	6 (46)	11 (58)	19 (50)
Age, mean (range)[b]	N/A[c]	N/A	32.7 (21-61)	41.6 (21-75)	38.0 (21-75)
Hospital, n (%)					
Hospital A	2 (100)	4 (100)	6 (46)	N/A	12 (32)
Hospital B	0 (0)	0 (0)	7 (54)	N/A	7 (18)
Participant cohort/impairment category, n (%)					
Visual	1 (50)	4 (100)	0 (0)	0 (0)	5 (13)
Hearing	0 (0)	0 (0)	5 (38)	8 (42)	13 (34)
Physical (mobility or chronic condition)	1 (50)	0 (0)	0 (0)	2 (11)	3 (8)
Cognitive/neurodevelopmental (cognitive, learning, attention-deficit/hyperactivity disorder, or autism spectrum disorder)	0 (0)	0 (0)	9 (69)	5 (26)	14 (37)
Older adults[d]	0 (0)	0 (0)	0 (0)	4 (21)	4 (11)

[a]Percentages are of column N. Counts exceed column N in User walkthroughs because 1 participant (P6) had co-occurring impairments spanning 2 categories (hearing and cognitive/neurodevelopmental) and is counted in both.

[b]The reported overall mean age (38.0 years) is calculated across the user-walkthrough (mean 32.7) and focus-group (mean 41.6) participants only; age was not recorded for expert-walkthrough or website-testing participants.

[c]N/A: not applicable.

[d]"Older adults" reflects a dedicated focus group recruited by age rather than by a specific impairment.

Data Analysis

All audio and screen-capture materials were uploaded to Condens.io (Condens Insights GmbH) and transcribed. Eye-tracking videos were reviewed alongside think-aloud audio to identify uncertainty events with temporal precision. Field notes from expert walkthroughs and structured scoring sheets from focus groups were imported as supplementary data sources.

Coding proceeded in 2 rounds, applying the framework described above. The first round used descriptive coding to identify individual barriers and elements of information uncertainty at the touchpoint level. The second round applied process coding to situate the identified elements within stages of the patient journey and capture sequences of action, decision, and recovery as they unfolded in context [42]. Coded units were entered into the Stage-Touchpoint-Uncertainty matrix. Aggregation by stage and impairment group supported the identification of uncertainty peaks, recurring breakdown patterns, and impairment-specific issues. Reports from focus groups and expert walkthroughs were used to triangulate and enrich the data from website testing and user walkthroughs. Two members (LZS and HG) of the research team conducted the 2 coding rounds, dividing the transcripts and cross-checking each other's coding. One coder coded a given portion of the transcripts while the other reviewed and checked the coding; these roles were reversed for the remaining portion. Thus, every

XSL•FO
RenderX

transcript was coded by 1 team member and independently checked by the other. A third team member (KJ) adjudicated cases in which the coder and checker did not reach agreement. This coder-checker approach, appropriate for a small, closely coordinated coding team, followed established qualitative coding guidance [42] rather than a formal interrater reliability procedure.

Mapping of Data Sources to Journey Stages and Triangulation Logic

The 4 data sources contribute unevenly to the analysis across journey stages (Table 2). The prearrival stage is investigated through a single source—website accessibility testing—which yields behavioral observations and assistive technology evidence on digital touchpoints. The remaining stages, from arrival through exit, are investigated primarily through user walkthroughs with mobile eye-tracking and think-aloud protocols, complemented by posttask interviews. Expert walkthroughs, conducted earlier in the project, serve as a heuristic baseline against which user-derived evidence is contextualized rather than as a primary source of findings. Focus groups operate across stages, providing an evaluative layer focused on participants' preferences and prioritization of design alternatives.

Methodological triangulation operates along 2 axes. First, *within-stage triangulation*, where for the arrival-to-exit stages, 3 independent sources—expert heuristic evaluation, user walkthroughs with attentional and behavioral evidence, and posttask interviews providing reflective evidence—converge on the same touchpoints. This enables cross-validation of findings and identification of breakdowns that are observable in behavior, articulated by participants, *and* recognized by experts as design weaknesses. Second, *cross-source elaboration*, where focus groups add an evaluative dimension that the in situ methods cannot capture, allowing participants to articulate priorities and preferences in a deliberative setting within groups organized by impairment type. This layered approach follows the principle that information uncertainty in wayfinding is a multidimensional phenomenon [24,25] and that no single method can adequately capture both its behavioral and attitudinal manifestations (Figure 1).

The decision to combine 4 methodologically distinct data sources rather than rely on any single source reflects a deliberate application of methodological triangulation, understood as the use of multiple methods or data sources to develop a comprehensive understanding of a phenomenon and test the validity of findings through their convergence [43,44]. Denzin [43] and Patton [44] distinguish 4 types of triangulation—method, investigator, theory, and data-source triangulation. Our design draws on the first and last of these simultaneously: 4 different methods (expert walkthroughs, accessibility testing, eye-tracking walkthroughs, and focus groups) are applied to the same underlying phenomenon (information uncertainty in hospital wayfinding), while drawing on 4 structurally different data sources (expert judgment, assistive technology–mediated interaction, embodied in situ behavior, and deliberative group evaluation). This dual structure allows the same touchpoints to be examined from professional, behavioral, and attitudinal vantage points rather than from a single perspective.

The rationale for combining methods in this particular way, rather than using an arbitrary combination, follows directly from the known limitations of each method in isolation, as discussed above. Expert walkthroughs are efficient but reflect professional judgment rather than lived user experience; screen reader–based website testing captures functional barriers at digital touchpoints but provides no information about the physical environment; mobile eye-tracking captures embodied, real-time behavior but is resource-intensive and necessarily limited in the number of participants and routes it can cover; and focus groups capture considered, comparative preferences but are evaluative and hypothetical rather than based on observations in situ. Every research method has inherent strengths and weaknesses. Surveys can reach large populations but may miss deeper meanings, whereas interviews offer rich detail but can be shaped by how questions are framed. Combining complementary approaches allows the weaknesses of one method to be offset by the strengths of another. Structuring the 4 rounds around this principle of complementary offsetting, rather than redundancy, enables the Stage-Touchpoint-Uncertainty matrix to aggregate types of evidence that no single method could have produced independently. This approach is consistent with the broader methodological pluralism that characterizes contemporary hospital wayfinding research. Related work applying a design-research framework to the same domain demonstrates this pluralism, drawing on methods ranging from surveys and experiments to ethnographic and physiological approaches and combining techniques such as usability testing, eye-tracking, virtual reality prototyping, and participatory concept testing across diverse patient groups [45].

Finally, the temporal sequencing of the 4 rounds (expert walkthroughs → website testing → user walkthroughs → focus groups) reflects a further, often under-specified rationale for combining methods: triangulation involves not only combining methods but also using earlier rounds to inform and focus later ones, thereby developing a progressively more comprehensive understanding of the phenomenon under study. In our design, expert walkthroughs generated a baseline set of anticipated touchpoint issues; website testing and user walkthroughs then examined whether these anticipated issues occurred for specific user groups under real-world conditions; and focus groups, conducted last, allowed participants to evaluate candidate design responses to the issues identified in the preceding 3 rounds. This progressive logic justifies treating the 4 methods not as parallel, independent data collection exercises but as components of a single, cumulative research design.

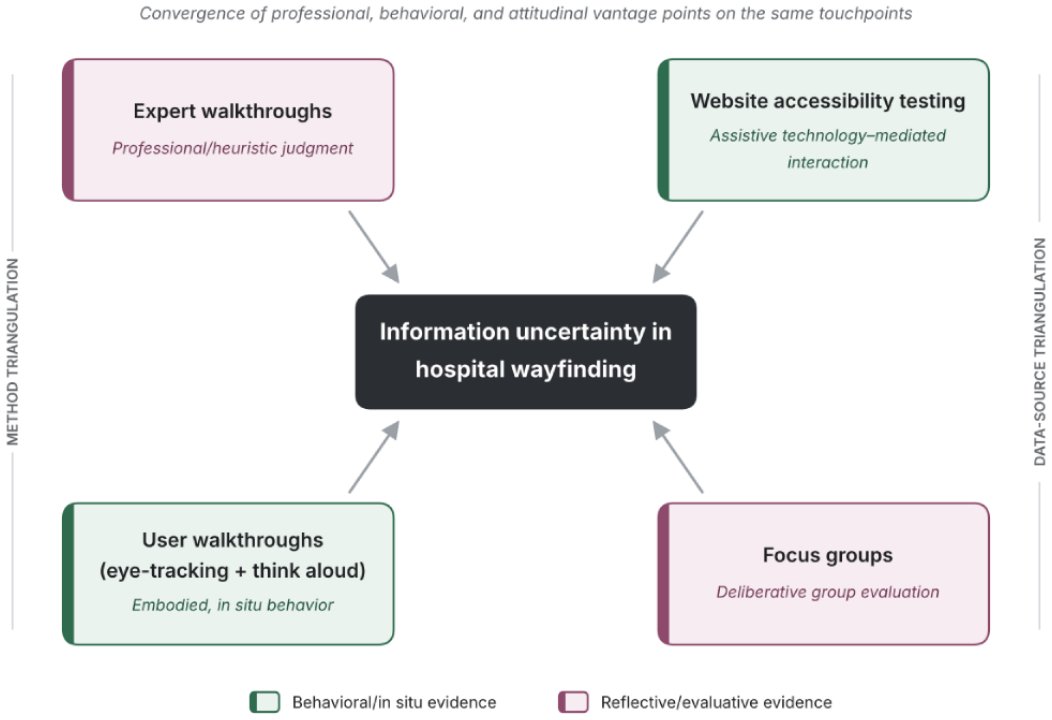
We acknowledge an asymmetry in coverage: the prearrival stage is supported by a single data source and by testers from a single impairment category (visual impairment). Focus groups partially compensate for this asymmetry through their cross-stage contributions and inclusion of participants with hearing, mobility, developmental, and learning differences. A comprehensive, multiimpairment investigation of the digital prearrival stage remains a direction for future research and is identified as a limitation in the "Discussion" section.

XSL•FO
RenderX

Table 2. Mapping data sources to journey stages, evidence type, framework dimensions, and principal limitations.

Data source	Journey stages covered	Evidence type	Framework dimensions populated	Principal limitations
Expert walkthroughs (n=2)	Arrival → exit (hospital A)	Heuristic field notes and photographs	Touchpoint qualities and context modifiers	Single site; expert perspective only; and no behavioral data
Website accessibility testing (n=4)	Prearrival	Screen capture, think-aloud audio, and task completion notes	Touchpoint qualities and accessibility factors	Single impairment category (visual) and single site
User walkthroughs with mobile eye-tracking (n=8 at A and n=7 at B)	Arrival → exit (hospitals A and B)	Eye-tracking video, think-aloud audio, and posttask interviews	Task state, uncertainty markers, touchpoint qualities, and context modifiers	Limited prearrival coverage, observer presence, and 1-shot exposure to site
Focus groups (4 groups; n=19)	Cross-stage (evaluative)	Audio recordings and structured scoring sheets	Touchpoint qualities and accessibility factors	Hypothetical evaluation of designs rather than in situ use and single site

Figure 1. Triangulation logic: behavioral and reflective triangulation, and method and data source triangulation.



Results

Prearrival: Digital Orientation

The prearrival stage was investigated through accessibility testing of the hospital website with 4 testers with severe visual impairments using JAWS, NVDA, VoiceOver, and ZoomText across 10 task scenarios. Observations were coded into stage-touchpoint events (n=40), each characterized by the touchpoint engaged and the uncertainty trigger(s) it produced. The resulting touchpoint structure is summarized in Table 3.

Touchpoint-quality families (visibility, legibility, consistency, and accessibility) were not quantified for this stage. As every observation was mediated by a specific assistive technology, the same underlying breakdown could plausibly be attributed to different families depending on which technology exposed it. For instance, a missing cross-link may manifest as a discoverability problem for a screen reader user and as a magnification-specific problem for a ZoomText user. Quantifying these categories would imply a level of precision that the coding cannot support at this stage; therefore, the recurring qualitative patterns are reported instead.

Several patterns recurred across multiple touchpoints. First, *information fragmentation*, that is, the data required to complete a single task were systematically split across pages (eg, clinic, accessibility, contact, and vaccination pages) without cross-links. This prompted backtracking by most testers and outright task failure or partial completion in several scenarios, notably those involving parking information for the maternity and emergency buildings, the disability-assistance service, and the route to the pediatric ward. Second, *avoidance strategies* were triggered whenever a primary touchpoint failed: testers resorted to the

XSL•FO
RenderX

footer site map, external search engines, or Google Maps satellite imagery. This was particularly evident when seeking parking information, which was absent from the website for at least two of the buildings tested, indicating that the website did not reliably fulfill its informational function for this user group. Third, *full-text search proved unreliable* for semantically equivalent queries, returning no results or an unrelated result for at least two query pairs (eg, "COVID testing" vs "COVID test"), despite search being testers' first recourse once navigation failed. Fourth, information about a *disability-assistance service* for visitors was absent from every location where testers expected to find it—the home page, the "For patients" and "About the hospital" sections, and full-text search. Only a general "Patient Rights" page mentioned people with disabilities, without identifying a bookable service. Fifth, the building-accessibility page used the term *accessibility* ambiguously, leaving it unclear whether it referred to step-free access or merely physical reachability, and described entrances using cardinal directions that were not actionable on-site.

Several scenarios resulted in unresolved information needs that carried forward into the next stage. Patients who arrive without confirmed information about building entrances, parking, or internal routes enter the physical environment without the cognitive anchors that prearrival orientation is intended to provide [46]. This continuity is examined further in the following sections.

Table 3. Dominant touchpoints and uncertainty triggers in the prearrival stage.

Touchpoint	Uncertainty triggers	Impairment-specific notes
Centralized "Building accessibility" page	Ambiguous use of the term *accessibility*; entrance descriptions given via cardinal directions ("entry from the north") not actionable on-site; parking information completely absent for several buildings; no cross-link from department pages	All testers
Individual clinic and department pages	Missing route descriptions and department-specific access detail; no cross-link to the accessibility page	All testers
Full-text search	Semantically equivalent queries return no or unrelated results; not consistently available from every page	All testers
Main dropdown navigation	Hover-activated menu disappears outside viewport at magnification; subitems inaccessible without static fallback	ZoomText user only
Breadcrumb navigation	Incomplete paths (no intermediate categories); low-contrast gray text; does not branch, preventing stepwise return	ZoomText user only
Home page department list	Alphabetic list lacks clear heading structure, hindering discovery without prior familiarity	1 of 4 testers
Information about parking (cross-page)	Completely absent for several buildings; forces avoidance strategies (eg, Google Maps satellite view)	All testers
Information about disability assistance	Absent or buried in unrelated sections (the "Patient Rights" page)	All testers

Entry and Initial Orientation

Observations were coded into stage-touchpoint events (n=59), distributed unevenly between sites (11 at hospital A and 48 at hospital B). This difference reflects route structure rather than uncertainty severity: the hospital A route unfolds within a single multistory pavilion and therefore generates a single entry event per session, whereas the hospital B route crosses several freestanding buildings, each triggering a separate entry event (see Table 4).

Two patterns recurred across multiple participants. First, *map information gaps at the point of entry*: the general orientation maps available in hospital lobbies frequently lacked the specific department, floor level, or directional cue needed to plan a route. As one participant put it after repeatedly scanning the lobby map without locating the target department, "It's annoying that I don't know where this is. That means I have to go all the way back to the main entrance again" [P15, hospital B]. Another participant, unable to locate any relevant information on the map, decided to abandon it altogether: "I won't find anything here. I'll just have to go inside somewhere" [P10, hospital B]. Second, building-level identification was inconsistently supported at hospital B, where participants sometimes could not confirm which building they had entered: "This might have been marked here at some point, but..." [P13, hospital B]. The participant trailed off before consulting the map again, illustrating a breakdown that carried forward into the en route stage described below.

XSL•FO
RenderX

Table 4. Dominant touchpoints and uncertainty triggers in the entry and initial orientation stage.

Touchpoint	Uncertainty triggers	Site/impairment-specific notes
General orientation map (lobby or entrance)	Lacks floor-level or directional detail; department not locatable on the map at all	Both sites
Paper/portable map	Consulted repeatedly at the entry point before committing to a direction	Both sites
Signage/building identifiers	Absent or insufficiently prominent building-level identification at entry	Both sites

En Route Navigation

Observations were coded into stage-touchpoint events (n=241; 67 at hospital A and 174 at hospital B). This is the most heavily represented stage in the dataset, reflecting the greater number and length of interbuilding segments at hospital B rather than differences in underlying uncertainty (see Table 5).

Several patterns recurred across sites and participants. First, *floor-numbering inconsistency* was the most consistently identified breakdown across the walkthrough data, affecting 8 of the 13 coded participants: “Honestly, it’s not clear to me whether I’m supposed to go up or down the stairs” [P13, hospital B]. A similar issue arose at hospital A: “I’m pissed that I have to take the elevator — there was no warning anywhere on the main sign about this” [P8, hospital A]. Second, *department-name overlap* led several participants to question whether differently worded signs referred to the same destination: “It says ‘patient reception’ here, but I don’t know — maybe it’s something else?” [P12, hospital B]. Third, at hospital B, *adjacent conflicting doorsignage* caused momentary confusion: “I don’t know if this is really the way through. They’re sending pretty mixed signals” [P13, hospital B]. Fourth, at hospital A, *elevator or door availability conflicted with posted signage*: “The elevator isn’t running, so now I don’t know — if I can only get there by elevator, and the elevator isn’t running, I don’t know how I’m supposed to get there” [P6, hospital A]. Another participant, finding that a door would not open from the side approached, exclaimed: “Damn it! So I have to go back up?” [P8, hospital A].

Table 5. Dominant touchpoints and uncertainty triggers in the en route navigation stage.

Touchpoint	Uncertainty triggers	Site/impairment-specific notes
Floor/level signage	Inconsistent floor-numbering conventions; color coding not aligned with numbering	Both sites
Department/room signage	Overlapping department names referring ambiguously to the same or different destinations	Both sites
Directional signage/general maps	Missing directional detail; information overload on large general maps	Hospital B
Door signage	Conflicting adjacent signs (“Do not open” next to “Exit”)	Both sites
Elevator panel/door	Elevator marked operational on signage but out of service on arrival; doors not opening from the approached side	Hospital A

Destination Interface

Observations were coded into stage-touchpoint events (n=84; 48 at hospital A and 36 at hospital B; Table 6).

Two patterns stood out. First, *terminology inconsistency at the point of arrival* was most pronounced at hospital A: the same clinical service was labeled inconsistently across signs approaching the same destination. One participant, confronted with a term she had not encountered elsewhere, explained: “I had no idea that outside regular hours this is already called the emergency service — that’s probably general knowledge I’m just missing” [P1, hospital A]. Another compared what she had seen inside the elevator with the signage at the corresponding floor exit: “In the elevator I saw something written a bit differently — there was nothing there about the acute ambulance...I’ll check again once the elevator arrives” [P1, hospital A]. At a signage board listing several overlapping department names simultaneously, a participant summarized the experience directly: “ENT ambulance — back entrance by the snack stand — that’s the exit — speech therapy...use the stairs only for floors 1, 3, 4...it’s total [chaos]” [P4, hospital A]. Second, at hospital B, breakdowns at this stage were more often attributable to physical legibility than to naming: “That’s a tiny sign, and faded on top of it” [P14, hospital B].

XSL•FO
RenderX

Table 6. Dominant touchpoints and uncertainty triggers in the destination interface stage.

Touchpoint	Uncertainty triggers	Site/impairment-specific notes
Departmental/terminology signage	Same service labeled inconsistently across signs; in-patient-ward and ambulance labels adjacent and easily confused	Both sites
Elevator-car versus floor-exit signage	Information seen inside the elevator car differs from, or is absent at, the floor-exit signage	Hospital A
Printed schedule/opening-hours notices	Outdated or contradictory opening-hours information posted near the destination	Hospital A
Door/room signage	Small print, low-contrast, or faded signage at the final approach to the destination	Both sites

Exit and Postvisit

Following reclassification of the hospital B "return to reception" task to this stage, it now comprises 27 units across 10 of the 13 coded participants (12 units from 5 participants at hospital A and 15 units from 5 participants at hospital B). The 2 sites' exit tasks are not fully equivalent: hospital A asked participants to leave the building via a previously unused route, whereas hospital B asked participants to return to the main reception, a task that does not necessarily involve exiting the building. This difference should be interpreted as a difference in task demand rather than underlying site quality (Table 7).

Two site-specific patterns emerged. At hospital A, exit-related information was often available only inside the elevator: "I would have had to remember that. When we rode the elevator, I saw it there...and here's the building exit" [P1, hospital A]. At hospital B, breakdowns during the return to reception centered on wing/section disambiguation. Encountering a sign that contradicted the expected route, one participant reasoned aloud: "I don't think we can get through this way. I think we have to go around the outside" [P16, hospital B]. Another, mistaking a previously visited section for the target section, asked: "This is A1, that's fine — but we need B1. So where is it?" [P15, hospital B]. A third opted to retrace a familiar route: "I think I'd just go back the same way I came" [P11, hospital B].

Table 7. Dominant touchpoints and uncertainty triggers in the exit and postvisit stages.

Touchpoint	Uncertainty triggers	Site-specific notes
Elevator-car signage	Exit/bridge-route information available only inside the elevator, absent from floor-level signage at the exit itself	Hospital A
Unlabeled doors near restricted-looking areas	Doors near apparent staff-only areas, with no confirming sign	Hospital A
Printed schedule notices	Outdated schedule reused to infer a department's current location during exit	Hospital A
Evacuation signage/doors	Evacuation-route signage repurposed as a wayfinding aid; a usually-locked evacuation door found open	Hospital A
Building/wing signage	A previously visited section wrongly assumed to be the current target section	Hospital B
Door/exit signage	An explicit "not an exit" sign contradicting the participant's expected route	Hospital B

Cross-Source Synthesis: Stages, Uncertainty Types, and Convergence

Aggregating the 451 coded stage-touchpoint events across the 5 journey stages (prearrival, n=40; entry, n=59; en route, n=241; destination interface, n=84; and exit, n=27), together with the evaluative focus-group data, yielded 4 recurring types of uncertainty. These types were defined by what fails between a touchpoint and the patient's ability to act on it, rather than by where the failure occurs. *Absence-based* uncertainty denotes a needed cue that is missing at the point of need. *Equivocality-based* uncertainty denotes cues that are present—sometimes redundantly—but do not converge on a single, stable meaning. *Perceptual* uncertainty denotes an unambiguous cue that fails to reach the patient because the signal itself is degraded. *Cognitive-translation* uncertainty denotes a correct, legible, and unambiguous cue that nonetheless requires an additional internal operation before it can guide action.

Mapping every touchpoint-level finding onto this typology (Table 8) revealed a distribution not apparent in the stage-by-stage tables. Absence-based findings were concentrated in the digital prearrival and physical entry stages, whereas equivocality-based findings were the most frequently coded type in the en route, destination interface, and exit stages and across the dataset as a whole. As event counts reflect route length and structure as much as uncertainty severity (see the "Methods" section), this pattern is reported as a coding frequency rather than as evidence of relative severity. Two equivocality patterns were explicitly cross-stage. First, inconsistencies in information about vertical circulation recurred during the en route, destination-interface, and exit stages at hospital A, with information displayed inside elevator cars differing from or missing from floor-level signage. Second, inconsistent naming of services and locations occurred across multiple journey stages: entry (building identification), en route (overlapping department names), destination interface (inconsistent service labels), and exit (confusion over wings or

XSL•FO
RenderX

sections). Perceptual uncertainty (small print, low-contrast, faded signage, and information-dense maps) was the least frequent type. Cognitive-translation uncertainty appeared at prearrival (cardinal-direction entrance descriptions) and in the focus-group data (nonviewer-oriented floor plans and color-only staircase coding).

The focus-group (evaluative) data triangulated the in situ findings rather than duplicating them. All focus groups—involving participants with hearing, developmental/learning, and physical/mobility impairments, as well as older adults—contributed coded data to this analysis. The strongest cross-source convergence concerned parking and vehicle information, which was corroborated across all impairment groups and the prearrival website-testing data. This was the most extensively triangulated finding in the dataset, appearing as absence at the digital prearrival stage and as equivocality at the physical, symbolic level on-site (eg, a parking-payment pictogram was independently misread as a bus by both an older-adult participant and a developmental/learning participant). Entrance/exit-pictogram ambiguity and the trade-off involving active/dynamic wheelchair symbols were also corroborated across all 4 groups, while water closet–pictogram ambiguity, floor-directory information density, the salience of the "you are here" marker, and campus-map color semantics were each corroborated across focus groups. Priorities also diverged in kind: participants with developmental/learning impairments emphasized redundant multichannel coding and completeness over minimalism; participants with physical/mobility impairments emphasized functional-eligibility information; older-adult participants relied heavily on everyday conventions from unrelated domains to decode pictograms (eg, interpreting a parking icon as a bus or a red cross as a traffic-light "stop" signal), were more sensitive to time pressure when viewing information-dense panels, and most often identified asking hospital staff as their preferred fallback strategy; and participants with hearing impairments emphasized the visual salience and contrast of symbols and the use of plain written language. Participants with visual impairment, whose website accessibility testing was reported under prearrival above, separately prioritized the digital retrievability of information.

Table 8. Uncertainty types mapped across journey stages, with representative examples.

Uncertainty type	Prearrival	Entry	En route	Destination	Exit
Absence	Parking and disability-assistance information was absent from the hospital website and not linked from relevant department pages.	General and entrance maps and door signs repeatedly lacked department names or directional cues, forcing participants to return to the main information point for guidance.	Directional signs and supplementary maps were absent from stairwells and interbuilding corridors, leaving participants without cues to confirm their floor or direction.	The destination building or department lacked clear arrival markers, leaving participants unsure whether they had reached their goal; corridors leading to the destination department lacked directional signage, forcing participants to guess which way to go.	N/A[a]
Equivocality	Ambiguous use of the term "accessibility," unresolved interpretations of parking and vehicle pictograms across 2 focus groups, and inconsistent terminology across website pages.	Building-identity uncertainty	Inconsistent floor numbering, conflicting door signs, discrepancies between indicated and actual elevator status, and overlapping department names.	Inconsistent service labels; discrepancies between elevator-car and floor signage; ambiguity in vehicle, water closet, and elevator pictograms; multiple overlapping department names on a single board; and outdated or contradictory opening-hours notices.	"Not an exit" signs conflicting with the expected route; confusion between A1 and B1; and evacuation signage repurposed for other directions.
Perceptual	Low-contrast breadcrumb text hindered stepwise navigation back through the website for a ZoomText user.	General map in the entrance hall with content positioned above comfortable eye level and information density that hindered rapid orientation.	Information overload on large maps, with equally sized bilingual text competing with the primary language and reducing legibility.	Door signage that was small, faded, or low in contrast.	N/A
Cognitive-translation	Cardinal-direction entrances	The entrance-door sign and the preceding instruction were each individually correct and legible, but required patients to mentally integrate the floor number and elevator/level instruction to construct a coherent route.	Floor plan not designed from the viewer's perspective; staircases distinguished by color alone, without text labels linking them to destination departments, making the coding difficult to interpret intuitively (focus-group finding).	"You are here" marker useful only when the map is correctly oriented; a mismatch between map orientation and the viewer's facing direction requires mental rotation to determine the correct route (focus-group finding).	N/A

[a]N/A: not applicable.

XSL•FO
RenderX

Discussion

Principal Findings

This study characterized information uncertainty in hospital wayfinding along 2 axes: *where* in the patient journey a breakdown occurs and *what kind* of breakdown it is. The journey-stage structure and the 4-type mechanism typology are complementary rather than competing: the stage locates a problem in a form that is administratively actionable (ie, identifies which team or physical or digital location must act), while the type specifies the appropriate remedy (adding missing content, standardizing terminology or adding a disambiguating channel, strengthening the signal, or reducing the cognitive translation demanded of the patient).

The stage structure was not merely an organizing convenience; it was a precondition for the study's 2 most consequential findings. First, the uncertainty mechanism differed across the journey, from absence-based deficits during digital prearrival search and physical entry to equivocality-based conflicts during on-site navigation, arrival, and exit. This finding describes variation across the care pathway that no single-touchpoint analysis could reveal. Second, tracking findings by stage revealed that an apparently scattered set of complaints could be explained by a small number of recurring cross-stage mechanisms. What might otherwise have been logged as 3 separate elevator complaints was, in fact, a single equivocality mechanism—information asymmetry between the elevator car and floor-level signage—recurring across 3 consecutive stages. Similarly, terminological drift was inherently a cross-stage phenomenon whose 4 stage-specific manifestations became apparent only because the analysis preserved the sequence of the journey rather than pooling all touchpoint events. The typology thus explains *why* particular touchpoints recur as problem areas, while the stage structure explains *where* each recurrence occurs along the pathway and who is responsible for addressing it. Across focus groups, the ability to locate the emergency department was considered categorically more urgent than the ability to locate other destinations. Participants explicitly linked psychological stress to situations in which even simple directional cues failed to provide useful guidance, suggesting an affective dimension of information uncertainty that the Stage-Touchpoint-Uncertainty framework, as applied here, does not explicitly model.

Regarding how patients with specific needs perceive and prioritize these mechanisms, evidence across data sources revealed clearly divergent priorities: digital retrievability of information for participants with visual impairment (website accessibility testing) and, among focus-group participants, redundant multichannel coding for the developmental/learning group and functional-eligibility information and standards-related concerns for the physical/mobility group. These differences indicate that a single wayfinding intervention is unlikely to serve all groups equally.

Comparison With Prior Work

The 2-axis structure operationalizes the distinction between locating a breakdown and diagnosing its mechanism more explicitly than the single-dimension Stage-Touchpoint-Uncertainty matrix from which it derives. The mechanism typology draws on Daft and Lengel's [47] foundational distinction between *uncertainty* (ie, the absence of information) and *equivocality* (ie, the presence of information that nonetheless admits multiple or conflicting interpretations). Their central claim that the more serious organizational problem is often a lack of clarity rather than a lack of data [47] maps directly onto our finding that equivocality, rather than absence, was the most frequently identified mechanism in the physical navigation stages. This finding reflects a qualitative coding-frequency pattern, not relative severity. To our knowledge, this distinction has not previously been applied at this level of granularity to patient wayfinding. Framing the recurring elevator and terminology findings as equivocality rather than as isolated site defects is consistent with treating wayfinding as a temporally and cross-channel distributed information behavior [34,35] and with the distributed-cognition premise that meaning is assembled across touchpoints rather than delivered complete by any one of them [48].

The predominance of equivocality over absence in the on-site data reinforces standardization arguments in the signage literature, including the reduction of environmental complexity through uniform color, typography, and symbol conventions across a signage system [49] and the specific emphasis on standardization in hospital wayfinding [50]. However, our findings extend these arguments by showing that inconsistency propagates across functionally sequential stages, rather than occurring only within a single sign type. The cognitive-translation type is supported by evidence that mental rotation of a map measurably increases cognitive load and reduces performance on orientation tasks [51]. This evidence provides a plausible cognitive basis for the difficulties associated with the nonoriented floor plans and cardinal-direction entrance descriptions observed in this study, rather than treating them as mere inconveniences. Finally, the divergent priorities across impairment groups are consistent with the universal-design literature's recognition that no single design serves all users equally, given persistent disagreement over how accessibility itself should be defined across groups [52]. They are also consistent with recent research in human-computer interaction, which argues that conflicting access needs are a substantive design consideration that requires careful deliberation rather than a technical problem that can simply be engineered away [53].

Practical Implications

The 2-axis structure translates directly into a division of labor among the teams responsible for hospital wayfinding, rather than a generic call for "better signage." Absence-type breakdowns are, by definition, content-management problems: the missing parking and disability-assistance information identified at prearrival requires a web-content solution—adding a persistent, cross-linked entry accessible from every relevant department page, not only from the centralized accessibility page—and no physical-plant intervention. Equivocality-type breakdowns require a different kind of fix: disambiguation and consistency across an existing set of cues, rather than the addition of new content. The ambiguous use of accessibility on the website should be resolved by separating

XSL•FO
RenderX

step-free/wheelchair access from general reachability into 2 distinct, consistently labeled categories. The recurring asymmetry between elevator-car and floor-exit signage at hospital A, which surfaced independently at en route, destination, and exit, points to a single fix: synchronizing elevator-car and floor-level signage content. This would resolve 3 stage-level findings at once, illustrating why locating a breakdown by mechanism rather than by isolated touchpoint is practically, not just analytically, useful. Terminological drift across entry, en route, and destination similarly calls for a single controlled vocabulary for service and department names, maintained centrally and applied uniformly across digital and physical signage, consistent with standardization arguments in the signage literature [49,50]. Perceptual-type breakdowns (small print, low-contrast, and faded signage at final approach) require conventional print-production fixes, such as minimum contrast ratios and type sizes at the door or room level. Cognitive-translation breakdowns, by contrast, call for a different kind of redesign: replacing cardinal-direction entrance descriptions with landmark-anchored instructions that are actionable on arrival and orienting floor plans to the viewer's direction of travel rather than to an abstract north. These changes would reduce the internal translation step that our data suggest is otherwise demanded of the patient [46].

As impairment groups differed in the mechanisms they prioritized—digital retrievability, redundant multichannel coding, and functional-eligibility information—no single fix among those described above will serve every user group equally well. This reinforces the need to validate these corrections directly with the affected groups rather than assume they are sufficient once implemented [40,41,52,53]. We therefore recommend that hospital wayfinding teams use the stage-mechanism matrix presented here as a triage tool: the stage identifies which team is responsible for the fix (web content, facilities/signage, or elevator maintenance), while the mechanism type identifies the kind of fix required. This approach allows interventions corroborated across multiple sources and impairment groups—most clearly, those addressing parking and vehicle information—to be prioritized over findings supported by a single source.

Strengths and Limitations

The principal strength of this study is methodological triangulation across 4 data sources and 2 structurally contrasting hospitals, allowing behavioral, digital-accessibility, and evaluative evidence to be considered together rather than in isolation. Several limitations qualify the findings. First, impairment status is reported unevenly: the focus-group findings are organized by impairment group, whereas the entry-through-exit walkthrough findings are not disaggregated by impairment category in this analysis, despite participants with hearing and cognitive/neurodevelopmental impairments forming the 2 largest walkthrough subgroups. Claims linking specific uncertainty types to specific impairment groups therefore remain provisional beyond the focus-group data. We deliberately did not attempt a quantitative or descriptive comparison between the walkthrough subgroups with hearing impairment and cognitive/neurodevelopmental impairments. In addition to their small and unequal sizes (n=4 and n=8, respectively), events for both subgroups were pooled across the 2 hospital sites, which, as noted below, differ substantially in scale, layout, and signage design. Any apparent difference between the 2 impairment groups could therefore not be cleanly separated from site-level differences in the volume and type of touchpoints encountered. A robust group comparison would require a larger sample stratified by both impairment group and site, which is beyond the scope of this study. Second, the elevator-equivocality patterns were documented at only 1 hospital site, so whether this pattern reflects a site-specific flaw or a general vertical-circulation problem cannot be determined. Third, the focus-group (symbol-level) and walkthrough (touchpoint-level) data operate at different levels of granularity and have not been fully cross-validated. Fourth, the 2 hospitals' exit tasks differed in demand, which should be considered when comparing findings. Fifth, the study covers 2 hospitals in 1 national context and may not generalize to facilities with different architectural or signage traditions. In addition, because the 2 hospitals differ in scale, layout, and available through-routes, as also reflected in their nonequivalent exit tasks, cross-site comparisons are limited to the typology of underlying uncertainty mechanisms (absence, equivocality, perceptual, and cognitive-translation) rather than direct comparisons of event frequencies between hospitals. Sixth, a researcher and, for participants using sign language, an interpreter accompanied each walkthrough participant. Although researchers did not provide directions or evaluative feedback, their presence may have influenced navigation behavior (a Hawthorne-type effect) relative to fully unaccompanied navigation. Seventh, eye-tracking data were used qualitatively to time stamp and disambiguate think-aloud utterances rather than analyzed as independent quantitative gaze metrics; a dedicated quantitative eye-tracking analysis was beyond the scope of this study. Eighth, focus-group comprehension of pictograms was tested within a fixed sequence of designs that itself supplied contextual cues (eg, adjacent icons and prior slides). Comprehension observed in these sessions likely overestimates comprehension of the same symbols encountered in isolation in the field.

Conclusions

Information uncertainty in hospital wayfinding is best understood along 2 axes: the journey stage at which a breakdown occurs and the underlying mechanism—missing information, conflicting information, degraded signal, or unmet cognitive-translation demand. The journey-stage structure remains essential because it allowed a scattered set of complaints to be consolidated into a few recurring cross-stage mechanisms and revealed that the type of uncertainty changes as patients move from digital prearrival search to physical navigation. The typology, in turn, replaces a generic call for "better signage" with a differentiated remedy for each mechanism. Together, the 2 axes support an intervention strategy that is sufficiently stage-specific to assign responsibility and potentially transferable to comparable settings beyond the 2 hospitals studied. This strategy prioritizes breakdowns corroborated across multiple sources and impairment groups—most clearly, those involving parking and vehicle information—over findings supported by a single source.

XSL•FO
RenderX

Acknowledgments

The authors thank the participants for their time and insights, the external design studios for making candidate signage materials available for focus-group evaluation, and both hospitals and their staff for their cooperation in this research.

Funding

This research was conducted as part of the project Design and Optimization of Orientation Labeling in Hospitals (WayDesign), supported by the Technology Agency of the Czech Republic (grant TQ01000352).

Data Availability

The datasets generated and analyzed during this study (interview transcripts, eye-tracking recordings, and coded matrices) are not publicly available due to the potential identifiability of participants with rare impairment profiles and the terms of the informed consent obtained. Deidentified excerpts of the coded matrices may be made available by the corresponding author (LZS) upon reasonable request, subject to approval by the Ethics Committee of Masaryk University.

Generative AI Use

Generative AI was used to support the automated transcription of audio recordings (Condens.io) and the copyediting and visualization of sections of this manuscript (Claude; Anthropic) under direct author supervision. All AI-assisted output was reviewed, verified against source data, and edited by the authors, who take full responsibility for the accuracy and integrity of the content. Generative AI was not used to generate original data, participant quotations, or citations.

Authors' Contributions

Conceptualization: LZS, KJ
Data collection – expert walkthroughs: KS, PČ
Data collection – focus groups: KJ, AK, GG
Data collection – user walkthroughs: KJ, NK, PP, MHB
Data collection – website testing: RP
Formal analysis/coding: LZS, KJ, HG
Funding acquisition: MHB
Methodology: LZS, KJ
Writing – original draft: LZS
Writing – review and editing: LZS, KJ, RP, PČ, KS, HG, GG, AK, PP, OK, NK, MHB

Conflicts of Interest

None declared.

References

1. Morag I, Heylighen A, Pintelon L. Evaluating the inclusivity of hospital wayfinding systems for people with diverse needs and abilities. J Health Serv Res Policy 2016 Oct 23;21(4):243-248. [doi: 10.1177/1355819616642257] [Medline: 27095135]
2. Petermans A, Mattiello L. Wayfinding in a hospital environment: proposing methodologies to impact positively well-being of patients, visitors, and staff. The Evolving Scholar 2022:1-7 [FREE Full text] [doi: 10.24404/622FA20A06178A1CB8717A6B]
3. Potter J. Best practices for wayfinding in a hospital setting. Scholar's Bank/University of Oregon. 2017. URL: https://hdl.handle.net/1794/22565 [accessed 2025-07-16]
4. Kaya S, Ileri Y, Yuceler A. Importance of hospital way-finding system on patient satisfaction. In: Bilgin MH, Danis H, Demir E, Can U, editors. Business Challenges in the Changing Economic Landscape (Volume 2). Cham, Switzerland: Springer International Publishing; 2016:33-40.
5. Harper C, Avera A, Crosser A, Duke T, Jefferies S. Designing wayfinding systems for hospitals. Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care 2018 Jun 29;7(1):122. [doi: 10.1177/2327857918071024]
6. Morag I, Pintelon L. Digital wayfinding systems in hospitals: a qualitative evaluation based on managerial perceptions and considerations before and after implementation. Appl Ergon 2021 Jan;90:103260. [doi: 10.1016/j.apergo.2020.103260] [Medline: 32950756]

XSL•FO
RenderX

7. Short EJ, Reay S, Gilderdale P. Wayfinding for health seeking: exploring how hospital wayfinding can employ communication design to improve the outpatient experience. The Design Journal 2017 Sep 06;20(sup1):S2551-S2568. [doi: 10.1080/14606925.2017.1352767]
8. Giraldi L, Maini M, Meloni D. Way-finding and communication design as strategic systems to improve the well-being of children in paediatric hospitals. In: Bagnara S, Tartaglia R, Albolino S, Alexander T, Fujita Y, editors. Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018). Cham, Switzerland: Springer International Publishing; 2019:799-810.
9. MacKenzie S, Krusberg J. Can I get there from here? Wayfinding systems for healthcare facilities. Leadersh Health Serv 1996;5(5):42-46. [Medline: 10161460]
10. Geng S, Yan S, Chau H, Zhang W, Zhang C. Wayfinding strategies for non-emergency services in Australian hospitals. Encyclopedia 2024 Dec 27;5(1):2. [doi: 10.3390/encyclopedia5010002]
11. Lo C, Liao C, Chen I, Hsieh T. Taiwanese people's wayfinding personas and tool preferences. In: Marcus A, Wang W, editors. Design, User Experience, and Usability: Understanding Users and Contexts. Cham, Switzerland: Springer International Publishing; 2017:249-264.
12. Reuschel W, McDonnall M, Burton D. The accessibility and usability of online job applications for screen reader users. Journal of Visual Impairment and Blindness 2023 Dec 07;117(6):479-490. [doi: 10.1177/0145482x231216757] [Medline: 38529353]
13. Deng L, Romainoor NH. A bibliometric analysis of published literature on healthcare facilities' wayfinding research from 1974 to 2020. Heliyon 2022 Sep;8(9):e10723 [FREE Full text] [doi: 10.1016/j.heliyon.2022.e10723] [Medline: 36177223]
14. Deng L, Zhang B, Shi G, Zhang C. Elderly users' perceptions of signage systems from tertiary hospitals in Guangzhou. Heliyon 2024 Feb 15;10(3):e25003 [FREE Full text] [doi: 10.1016/j.heliyon.2024.e25003] [Medline: 38317991]
15. Pouyan AE, Ghanbaran AH, Hosseinzadeh A, Shakibamanesh A. The elderly wayfinding performance in an informative healthcare design indoors. Journal of Building Engineering 2024 Jun;87:108843. [doi: 10.1016/j.jobe.2024.108843]
16. Irish JEN. An exploratory study testing environmental wayfinding aids as an intervention for children with autism. HERD 2022 Oct 12;15(4):114-130. [doi: 10.1177/19375867221111467] [Medline: 35822221]
17. Jeamwatthanachai W, Wald M, Wills G. Indoor navigation by blind people: behaviors and challenges in unfamiliar spaces and buildings. British Journal of Visual Impairment 2019 May 08;37(2):140-153. [doi: 10.1177/0264619619833723]
18. Morag I, Sonmez V, Van Puyvelde A, Pintelon L. Improving wayfinding in hospitals for people with diverse needs and abilities: an exploratory approach based on multi-criteria decision making. Appl Ergon 2024 Jan;114:104149. [doi: 10.1016/j.apergo.2023.104149] [Medline: 37847986]
19. Passini R, Rainville C, Marchand N, Joanette Y. Wayfinding in dementia of the Alzheimer type: planning abilities. J Clin Exp Neuropsychol 1995 Dec;17(6):820-832. [doi: 10.1080/01688639508402431] [Medline: 8847388]
20. Ali N, Luther SL, Volicer L, Algase D, Beattie E, Brown LM, et al. Risk assessment of wandering behavior in mild dementia. Int J Geriatr Psychiatry 2016 Apr 30;31(4):367-374. [doi: 10.1002/gps.4336] [Medline: 26223779]
21. Asan O, Yang Y. Using eye trackers for usability evaluation of health information technology: a systematic literature review. JMIR Hum Factors 2015 Apr 14;2(1):e5 [FREE Full text] [doi: 10.2196/humanfactors.4062] [Medline: 27026079]
22. Passini R. Wayfinding: a conceptual framework. Urban Ecology 1981 Apr;5(1):17-31. [doi: 10.1016/0304-4009(81)90018-8]
23. Gualandi R, Masella C, Viglione D, Tartaglini D. Exploring the hospital patient journey: what does the patient experience? PLoS One 2019 Dec 5;14(12):e0224899 [FREE Full text] [doi: 10.1371/journal.pone.0224899] [Medline: 31805061]
24. Chien Y, Tsai P, Lin Y, Wu S, Chen K, Han Y, et al. Uncertainty and Mental Workload Among Wayfinding Strategies. Cham, Switzerland: Springer Cham; 2016.
25. Rosenbloom L. Real-time continuous uncertainty annotation (RCUA) for spatial navigation studies. arXiv Preprint published online on July 18, 2023 [FREE Full text] [doi: 10.48550/arXiv.2207.14651]
26. Bulto LN, Davies E, Kelly J, Hendriks JM. Patient journey mapping: emerging methods for understanding and improving patient experiences of health systems and services. Eur J Cardiovasc Nurs 2024 May 28;23(4):429-433 [FREE Full text] [doi: 10.1093/eurjcn/zvae012] [Medline: 38306596]
27. Cheng AC, White J. Using wayfinding data to understand patient travel within a medical center. AMIA Annu Symp Proc 2018;2018:1216-1223 [FREE Full text] [Medline: 30815163]
28. Guo W, He Y. Optimized wayfinding signage positioning in hospital built environment through medical data and flows simulations. Buildings 2022 Sep 11;12(9):1426. [doi: 10.3390/buildings12091426]
29. Zijlstra E, Hagedoorn M, van DSC, Mobach M, Jylhä T. The patient journey in a hospital environment. 2020 Presented at: European Facility Management International Conference; June 3-4, 2020; Online URL: https://www.researchgate.net/publication/367464418_The_patient_journey_in_a_hospital_environment
30. Suchá LZ. Mapping information uncertainty in hospital wayfinding through patient journey maps. IR 2026 May 25;31(ISIC):752-761. [doi: 10.47989/ir31isic65290]
31. McCarthy S, O'Raghallaigh P, Adam F. Integrated patient journey mapping: a sensemaking approach for health information systems research. In: Research Handbook on Health Information Systems Edward Elgar Publishing. Cheltenham, UK: Edward Elgar Publishing; 2025:224-238.

XSL•FO
RenderX

32. Joseph AL, Monkman H, Kushniruk A, Quintana Y. Exploring patient journey mapping and the learning health system: scoping review. JMIR Hum Factors 2023 Feb 27;10:e43966 [FREE Full text] [doi: 10.2196/43966] [Medline: 36848189]
33. Willems SJ, Coppieters MW, Pronk Y, Diks MJF, van der Heijden KWAP, Rooker S, et al. A clinical journey mobile health app for perioperative patients: cross-sectional study. JMIR Hum Factors 2021 Feb 08;8(1):e20694 [FREE Full text] [doi: 10.2196/20694] [Medline: 33555262]
34. Fidel R. An Ecological Approach to Information Behavior. Cambridge, MA: The MIT Press; 2012.
35. Short EJ, Reay SD, Douglas RA. Designing wayfinding systems in healthcare: from exploratory prototyping to scalable solutions. Design for Health 2019 Apr 24;3(1):180-193. [doi: 10.1080/24735132.2019.1575659]
36. Iwarsson S, Ståhl A. Accessibility, usability and universal design--positioning and definition of concepts describing person-environment relationships. Disabil Rehabil 2003 Jan 21;25(2):57-66. [doi: 10.1080/dre.25.2.57.66] [Medline: 12554380]
37. Mankoff J, Hayes G, Kasnitz D. Disability studies as a source of critical inquiry for the field of assistive technology. New York, NY: ACM; 2010 Presented at: ACM SIGACCESS Conference on Computers and Accessibility; October 25-27, 2010; Orlando, FL. [doi: 10.1145/1878803.1878807]
38. Schneider A, Vollenwyder B, Krueger E, Mühlethaler C, Miller DB, Thurau J, et al. Mobile eye tracking applied as a tool for customer experience research in a crowded train station. J Eye Mov Res 2023 Jan 16;16(1):e1 [FREE Full text] [doi: 10.16910/jemr.16.1.1] [Medline: 37927371]
39. Kitzinger J. Qualitative research. Introducing focus groups. BMJ 1995 Jul 29;311(7000):299-302 [FREE Full text] [doi: 10.1136/bmj.311.7000.299] [Medline: 7633241]
40. Bennett C, Rosner D. The promise of empathy: design, disability, and knowing the other. New York, NY: ACM; 2019 Presented at: Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland. [doi: 10.1145/3290605.3300528]
41. Charlton J. Nothing About Us Without Us: Disability Oppression and Empowerment. Oakland, CA: University of California Press; 2004.
42. Saldaña J. The Coding Manual for Qualitative Researchers, 3rd Edition. Los Angeles, CA: SAGE; 2016.
43. Denzin N. The Research Act: A Theoretical Introduction to Sociological Methods. London, UK: Routledge; 2017.
44. Patton MQ. Enhancing the quality and credibility of qualitative analysis. Health Serv Res 1999 Dec;34(5 Pt 2):1189-1208 [FREE Full text] [Medline: 10591279]
45. Zbiejczuk SL, Holubec BM, Jochecová K, Kvarda O, Godišková G. Exploring and designing wayfinding behavior in hospitals: the double 'double diamond' approach. Illinois Library. 2026. URL: https://www.ideals.illinois.edu/items/139998 [accessed 2026-07-13]
46. Hölscher C, Meilinger T, Vrachliotis G, Brösamle M, Knauff M. Up the down staircase: wayfinding strategies in multi-level buildings. Journal of Environmental Psychology 2006 Dec;26(4):284-299. [doi: 10.1016/j.jenvp.2006.09.002]
47. Daft RL, Lengel RH. Organizational information requirements, media richness and structural design. Management Science 1986 May;32(5):554-571. [doi: 10.1287/mnsc.32.5.554]
48. Hutchins E. Cognition in the Wild. Cambridge, MA: The MIT Press; 1995.
49. Zhou T, Xia P, Zhu Q, Du J. Cognition-driven navigation assistive system for emergency indoor wayfinding (CogDNA): proof of concept and evidence. Safety Science 2023 Jun;162:106100. [doi: 10.1016/j.ssci.2023.106100]
50. Kalyan BA, Rani M, Shouri S. A study on patients' perception of signage system in a tertiary care teaching hospital. HVT 2024 Jul 22;8(Issue 3):356. [doi: 10.24969/hvt.2024.499]
51. Zhao M, Liu J, Liu Y, Kang P. Effects of mental rotation on map representation in orienteers-behavioral and fNIRS evidence. PeerJ 2023;11:e16299 [FREE Full text] [doi: 10.7717/peerj.16299] [Medline: 37868057]
52. Persson H, Åhman H, Yngling AA, Gulliksen J. Universal design, inclusive design, accessible design, design for all: different concepts—one goal? On the concept of accessibility—historical, methodological and philosophical aspects. Univ Access Inf Soc 2014 May 7;14(4):505-526. [doi: 10.1007/s10209-014-0358-z]
53. Tang X, Piper A. Designing for collective access: in search of a solution to accessible communication in a mixed-ability non-profit. arXiv Preprint published online on May 11, 2026 [FREE Full text] [doi: 10.48550/arXiv.2605.10085]

Abbreviations

ENT: ear-nose-throat

XSL•FO
RenderX

Edited by J Viitanen; submitted 13.Jul.2026; peer-reviewed by L Herdiman, MA Mohd Bakri; comments to author 11.Sep.2026; revised version received 28.Sep.2026; accepted 16.Oct.2026; published 09.Nov.2026.

Please cite as:
Zbiejczuk Suchá L, Jochecová K, Pavlíček R, Červenka P, Sobol K, Greplová H, Godišková G, Kúdelová A, Pospíšilová P, Kvarda O, Káčová N, Holubec Birtusová M
Reducing Information Uncertainty in Hospital Wayfinding for People With Specific Needs: Findings From a Qualitative, Multimethod Study
JMIR Hum Factors 2026;13:e106882
URL: https://humanfactors.jmir.org/2026/1/e106882
doi:10.2196/106882
PMID:43003400

©Ladislava Zbiejczuk Suchá, Kateřina Jochecová, Radek Pavlíček, Petr Červenka, Karel Sobol, Henrieta Greplová, Gabriela Godišková, Alica Kúdelová, Pavla Pospíšilová, Ondřej Kvarda, Natálie Káčová, Michaela Holubec Birtusová. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 09.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Dynamic Augmented Reality Cues for Telementoring in Laparoscopic Surgery: Usability Study

Jhasketan Padhan[1], MCA; Hawa Hamza[1], BSc; Dehlela Shabir[1,2], MSc; Julien Abinahed[1], PhD; Panagiotis Tsiamyrtzis[3,4], PhD; Zhigang Deng[5], PhD; Nikhil V Navkar[2], PhD

[1]Itqan Clinical Simulation and Innovation Center, Hamad Medical Corporation, PO Box 3050, Doha, Qatar

[2]Artificial Intelligence Research and Innovation Hub, Hamad Medical Corporation, PO Box 3050, Doha, Qatar

[3]Department of Mechanical Engineering, Politecnico di Milano, Milan, Italy

[4]Department of Statistics, Athens University of Economics and Business, Athens, Greece

[5]Department of Computer Science, University of Houston, Houston, TX, United States

Corresponding Author:
Nikhil V Navkar, PhD
Artificial Intelligence Research and Innovation Hub, Hamad Medical Corporation, PO Box 3050, Doha, Qatar

Abstract

Background: Surgical telementoring enables a remote expert surgeon (mentor) to guide an operating surgeon (mentee) during surgery and facilitates the transfer of surgical skills. However, commonly used audio and static visual cues are inadequate to demonstrate complex tool-tissue interactions. To overcome this limitation, dynamic augmented reality (AR)–based visual cues are overlayed on the operative field to demonstrate precise instrument movements.

Objective: The objective of this work was to evaluate dynamic AR cues and identify the most suitable cue that effectively demonstrates the required motions of surgical instruments during laparoscopy.

Methods: A user study was conducted in a simulated environment among mentor-mentee pairs using 3 dynamic AR cues (hand gestures, a 3D pointer, and a virtual tool). The task assessed how closely the mentee was able to follow the mentor. The outcomes measured were (1) dynamic time warping distance, representing the closeness of the paths followed; (2) angular error in tooltip orientation; and (3) the NASA Task Load Index, assessing cognitive workload during telementoring.

Results: Telementoring using the virtual tool resulted in a reduced dynamic time warping distance compared with hand gestures (P<.01 for 12/13, 92.3% of the trials) and the 3D pointer (P<.05 for 11/13, 84.6% of the trials). Lower orientation error was also noted while using the virtual tool as compared with hand gestures (P<.05 for 5 of 6 poses). There were no substantial differences in the NASA Task Load Index scores.

Conclusions: Use of a virtual tool (as a dynamic AR cue) enabled the mentee to follow the mentor's instructions with fewer errors than both hand gestures and a 3D pointer without increasing cognitive workload. Further research is needed to assess the clinical effectiveness of the virtual tool during live surgery.

(JMIR Hum Factors 2026;13:e86566) doi:10.2196/86566

KEYWORDS

laparoscopy; minimally invasive surgical procedure; telementoring; augmented reality; dynamic visual cues

Introduction

A growing need for expert surgeons' guidance during laparoscopic surgery necessitates the adoption of telementoring technologies [1]. During telementoring, an expert surgeon (mentor) can guide an operating surgeon (mentee) in real time even if they are continents apart. Operating room footage is sent to the remote mentor, who then provides instructions to the mentee using audio cues. Further enhancing the mode of instruction, the telementoring system may also integrate augmented reality (AR) technologies, whereby computer-simulated virtual objects are superimposed on the user's field of view [2]. This includes the use of static AR-based visual cues such as annotations or *telestrations* (ie, sketches or illustrations drawn over a video feed by the mentor in real time) overlaid on the mentee's operative view displayed on a screen. This is useful during in-person guidance as well, as the mentor can point at anatomical structures instead of describing them verbally [3]. However, static AR cues remain stationary within the operating field, and the mentor may not be able to communicate the required tool positions, orientations, and movements. To overcome this limitation, the use of dynamic AR cues, including but not limited to the mentor's hand gestures, a 3D pointer, and a virtual tool, has been proposed [4]. The remotely located expert surgeon can communicate with the mentee by controlling these cues in real time, thereby

XSL•FO
RenderX

demonstrating complex tool-tissue interactions required during laparoscopic surgery. This can also be used to enhance in-person guidance, thereby eliminating the need for the mentor surgeon to take control of the instruments to demonstrate specific tool movements.

Advancements in dynamic AR cues for telementoring during minimally invasive surgery (laparoscopic and robot assisted) have been tested extensively (Table 1). They include procedures such as cholecystectomy, prostatectomy, and suturing in various clinical settings such as virtual simulator, synthetic phantom, animals (cadavers), and humans (live). The dynamic AR cues frequently used for guiding mentees during minimally invasive surgeries can be categorized as hand gestures, pointers, or surgical tools [5]. While some of the studies have provided a description of the technologies used, others have compared dynamic AR cues with different modes of mentoring. They include assessment of hand gestures with audio cues [3,6,7] or in-person guidance [8], 3D pointers with audio cues [9,10], and virtual tools with in-person guidance [11-15]. A comparison of dynamic AR cues against static 2D telestrations has also been conducted [16]. As shown in Table 2, previous articles have already reported favorable results while using dynamic AR cues when compared with standard practices such as audio cues, in-person guidance, and static 2D telestrations. Both hand gestures and 3D pointers have shown improved performance compared with audio cues. Virtual tools have resulted in comparable performance to and lower operating time than that with in-person guidance. In these studies, all 3 dynamic AR cues were preferred over static 2D telestrations. While some of the related studies have introduced these telementoring technologies in cadaveric and live surgical settings, they have not objectively compared the 3 dynamic AR cues, namely, hand gestures, 3D pointers, and virtual tools, against each other. To the best of our knowledge, there are no existing studies that have performed a comparative analysis of the different dynamic AR cues. To enhance telementoring and improve the learning experience for the mentee, there is a need to standardize and optimize the use of dynamic AR cues. This requires comparing different types of dynamic AR cues among themselves to identify the most effective modality for the mentor to render the information to the mentee during surgical telementoring.

XSL•FO
RenderX

Table . Related works on dynamic augmented reality (AR) cues during minimally invasive surgical telementoring.

Telementoring system	Dynamic AR cue			Comparison	Clinical setting	Outcome
	Hand gestures	3D pointer	Virtual tool			
iSurgeon [3,6,7,17]	✓			Audio cue	Cholecystectomy on animals (cadavers)	Better GOALS[a] and OSATS[b] scores, fewer complications (P<.001), and reduced workload (NASA-TLX[c]; P<.02)
HoloPointer [9,10]		✓		Audio cue	Cholecystectomy on virtual simulator and humans (live)	Improved movement economy (P=.047), error rate (P=.047), and performance (P=.03) and favorable ratings by participants
Long et al [18]	✓		✓	—[d]	Prostatectomy on humans (live)	Successful execution of telementoring system
Shabir et al [11-13,19,20]			✓	In-person guidance	Suturing on synthetic phantom	Comparable performance among groups receiving telementoring and in-person guidance
Lowry et al [14]			✓	No cue; in-person guidance	Suturing and basic tasks on synthetic phantom	Comparable performance among groups receiving telementoring and in-person guidance
Proximie [8,21]	✓			In-person guidance	Prostatectomy and aquablation on humans (live)	Overall favorable ratings by participants
Feng et al [22,23]		✓		Audio cue	Cholecystectomy on synthetic phantom	Better movement economy (P=.01) and performance (P<.001)
Jarc et al [16,24]	✓	✓	✓	Static visual cue (2D telestration)	Suturing and basic tasks on animals (live)	3D tools were preferred over static visual cues by the participants
VIPAR[e] platform [25]	✓	✓		—	Ventriculostomy on humans (live)	Successful execution of VIPAR platform
ART[f] [15]			✓	In-person guidance	Suturing on synthetic phantom	Shorter operating time (P=.01) and fewer failed attempts while using ART

[a]GOALS: Global Operative Assessment of Laparoscopic Skills.

[b]OSATS: Objective Structured Assessment of Technical Skills.

[c]NASA-TLX: NASA Task Load Index.

[d]Not applicable.

[e]VIPAR: virtual interactive presence and AR.

[f]ART: AR telementoring.

XSL•FO
RenderX

Table . Advantages of dynamic augmented reality (AR) cues over conventional mentoring methods.

Dynamic AR cue	Comparison		
	Audio cue	In-person guidance	Static 2D telestration
Hand gestures	Better performance score, fewer complications, and reduced workload [3,6,7,17]	Favorable ratings by participants [8,21]	Preferred over static visual cues by the participants [16,24]
3D pointer	Improved movement economy, performance [9,10,22,23], and error rate and favorable ratings by participants [9,10]	—[a]	Preferred over static visual cues by the participants [16,24]
Virtual tool	—	Similar performance [11-14,19,20], shorter operating time, and fewer failed attempts [15]	Preferred over static visual cues by the participants [16,24]

[a]Not available.

In this study, we aimed to quantitatively assess the dynamic AR cues during telementoring and compare the mentee's performance while receiving instructions in each mode, namely, hand gestures, a 3D pointer, or a virtual tool. In the following sections, we present the experimental setting used for this study, the results obtained during the trial, and a discussion of the major implications of using dynamic AR cues for telementoring during laparoscopic surgery.

Methods

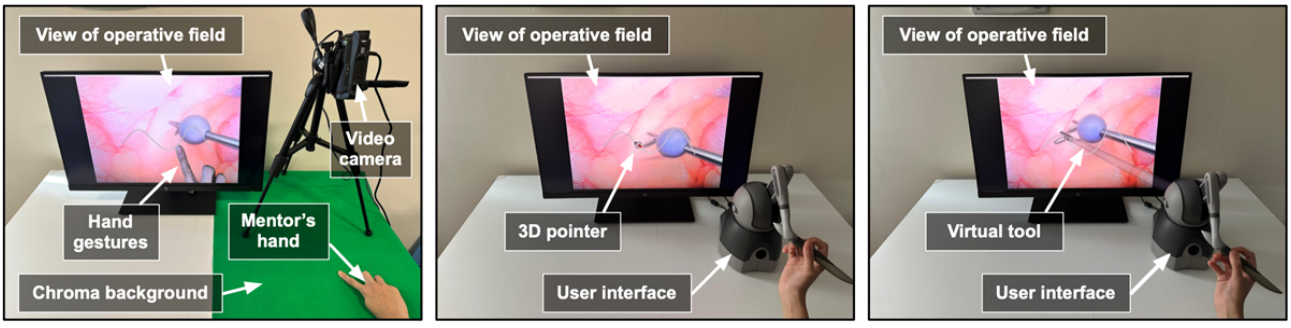
Participant Enrollment

A user study was conducted among mentor-mentee pairs as participants to evaluate the dynamic AR cues for telementoring. Participants were recruited from the Department of Surgery at Hamad General Hospital in Qatar. They were researchers working on the development of surgical technologies. None of them had prior experience performing real surgeries; however, they had experience using surgical simulators and were familiar with the hand-eye coordination required for maneuvering laparoscopic tools. Hand-eye coordination is crucial for laparoscopic tasks, especially when the operative field is displayed to the user on a 2D screen [26]. A total of 10 participants as mentees were included in the user study. One participant, a senior researcher from the Department of Surgery who was familiar with the telementoring system, was recruited as the mentor to demonstrate the required laparoscopic instrument motions to the mentees. The mentor was well versed in using user interfaces at the mentor's site to demonstrate the required motion of a laparoscopic tool using different dynamic AR cues.

Setup at the Mentor and Mentee Sites

The mentor used 3 different dynamic AR cues, namely, *hand gestures*, a *3D pointer*, and a *virtual tool*, to guide the mentees. The setup at the mentor's site for generating each of the dynamic AR cues is shown in Figure 1. To capture hand gestures (Figure 1A), the mentor moved the hand over a chroma background, and a camera was used to acquire the movements. The background was then removed from the acquired video, and the hand gestures were overlaid onto the live view of the operative field. In the case of the 3D pointer (Figure 1B), the mentor controlled the position and orientation (pose) of a stylus (Geomagic Touch; 3D Systems). The relative motion of the stylus was mapped to the 3D pointer rendered onto the live view of the operative field. Finally, the same stylus was also used to control the movements of the virtual tool (Figure 1C) based on our previous work [27].

Figure 1. The setup at the mentor's site to generate the 3 types of dynamic augmented reality cues: (A) hand gestures, (B) a 3D pointer, and (C) a virtual tool.

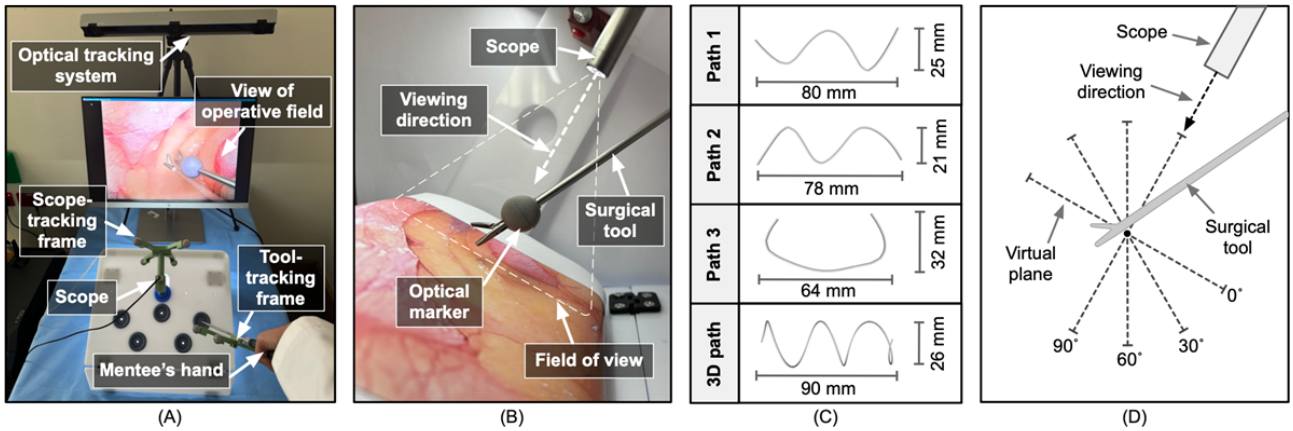


At the mentee's site, the study was performed on a low-fidelity synthetic phantom setup using a laparoscopic box trainer (Figure 2). The setup is shown in Figure 2A. A one-to-one comparison of the 3 dynamic AR cues requires a controlled environment that remains constant throughout the trials to eliminate confounding variables related to anatomical variability, tissue deformation, and procedural differences. For this reason, we designed this study using a phantom setup, which provides a stable and reproducible environment. This approach allowed for the isolation and objective assessment of the effect of each

XSL•FO
RenderX

dynamic AR cue on tooltip movement and spatial guidance independent of the surgical procedure itself. Second, it avoided patient-related surgical risk and ethical concerns while still providing meaningful insights on the use of dynamic AR cues. The type of task that was selected was surgical tool maneuvering. It was chosen as it is an essential skill for laparoscopy requiring hand-eye coordination and precise instrument control [28]. Improvements in instrument control and spatial guidance are broadly transferable to more complex surgical tasks.

Figure 2. (A) Setup at mentee's site comprising an optical tracking system, a screen rendering the view of the operative field, a box trainer, a surgical scope, and a laparoscopic instrument. (B) Setup inside the box trainer. (C) Three virtual 2D paths and one 3D path used in scenario A. (D) The orientation at which the 2D paths were rendered in front of the scope on a virtual plane.

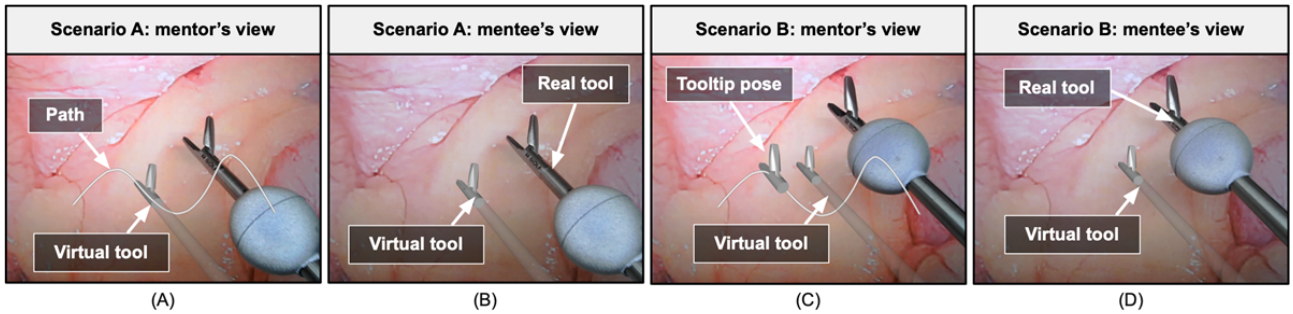


The dynamic AR visual cues were overlaid on the mentees' view of the operative field in real time. The dynamic AR cues assisted the mentees in comprehending the required pose of the laparoscopic instrument and maneuvering it accordingly. To record the movements of the laparoscopic instrument with respect to the scope, tracking frames were used. An optical tracking system (V120-Trio; NaturalPoint) was used to track the motion of the laparoscopic instrument and the scope. In addition, the distal end of the laparoscopic instrument was fitted with an optical marker (Figure 2B). This assisted in capturing accurate positions of the laparoscopic instrument tip during the study. As the setup required placing an optical marker on the laparoscopic instrument, the user study was conducted in a simulated environment instead of on humans.

Telementoring Scenarios

Two scenarios, scenario A and scenario B, were simulated to assess the dynamic AR cues. Scenario A evaluated dynamic AR cues to facilitate continuous tool movements. The movements of the laparoscopic instrument were characterized by 4 different paths (Figure 2C). They consisted of 2D paths, each positioned at 0°, 30°, 60°, and 90° orientations (Figure 2D), and one 3D path. The paths were rendered onto the mentor's screen (Figure 3A), whereas the mentees were able to see the dynamic AR cue generated by the mentor (Figure 3B). The mentor was asked to generate dynamic AR cues that would guide the mentees to follow the paths using the laparoscopic instrument. The mentees observed the dynamic AR cues in real time and performed the movements of the laparoscopic instrument to follow the mentor.

Figure 3. (A) Mentor's view in scenario A. (B) Mentee's view in scenario A. (C) Mentor's view in scenario B. (D) Mentee's view in scenario B.



Scenario B evaluated dynamic AR cues to facilitate precise orientation of the laparoscopic instrument's jaw. Six different feasible poses for orienting the jaw of the laparoscopic instrument were rendered to the mentor (Figure 3C). The mentor observed these poses and generated dynamic AR cues. The dynamic AR cues were rendered to the mentees (Figure 3D), and the mentees maneuvered the laparoscopic instrument to orient the jaw. As it was impossible to indicate this using 3D pointers, only hand gestures and virtual tools were used for guidance in scenario B.

In both scenarios, verbal instructions from the mentor were not provided. No verbal instructions were given to assess the dynamic AR cues independently as feedback in the form of audio cues can assist the mentees and alter the performance during the task. After completion of both scenarios, the mentees were required to complete the NASA Task Load Index (NASA-TLX) questionnaire [29] for each dynamic AR cue.

XSL•FO
RenderX

Study Design and Outcomes Measured

We chose to study within-subject factors by having the same individuals participate in all 3 dynamic AR cues. In this way, the participants acted as their own controls, and we eliminated any variation in task performance arising from individual differences. Additionally, to eliminate bias arising from the learning effect, we randomized the order in which the AR cues were allocated to the participants, as well as the order of the different tasks performed using each AR cue. Before the start of the study, each mentee went through a preparatory step to familiarize themselves with the telementoring setup and dynamic AR cues. Each mentee performed both scenarios with all 3 dynamic AR cues. The sequence in which the dynamic AR cues were used by the mentor-mentee pair was randomized. Furthermore, in scenario A, the sequence in which the 2D and 3D paths were rendered to the mentor and the orientations in the case of the 2D path were also randomized. The primary outcomes measured as part of this study included objective assessments in the form of (1) the average dynamic time warping (DTW) distance between the paths followed by the mentees and the required paths shown by the mentor during scenario A and (2) angular errors (measured in degrees) in orienting the laparoscopic instrument jaw during scenario B. DTW is a technique for measuring similarity between time-series data using temporal alignment [30]. Unlike the Euclidean distance, which relies on point-to-point comparison, DTW allows for one-to-many and many-to-one comparison, which is useful for paths that vary in time and speed. DTW has been applied for surgical performance and instrument motion demonstrations [19,26,27,31]. A lower value of DTW indicates a greater similarity and better alignment between the mentees' path and the reference path. The secondary outcome consisted of subjective assessment scores rated by the mentees using the NASA-TLX. These measured mental, physical, and temporal demands; effort; performance; and frustration on a scale from 1 to 10, with a lower score indicating a more positive rating.

Statistical Analysis

In scenario A for paths 1 to 3, an ANOVA was performed to determine whether there were significant mean values over the 3 dynamic AR cues and the 4 orientations of the virtual plane. The dynamic AR cue model variable was highly statistically significant ($P<.01$) in all cases, whereas the orientation of the virtual plane was significant only in path 2. It was then decided to study all pairwise comparisons of the 3 dynamic AR cues for a path and a specified orientation, and as the same mentee participated in all 3 AR cues, we used the 2-tailed paired *t* test. For path 4, ANOVA established significant differences among the 3 AR cues. Therefore, a paired *t* test was used to examine statistically significant mean differences over all 3 pairs. Finally, in scenario B, a paired *t* test was used to determine the statistical significance between the 2 dynamic AR cues (hand gestures and virtual tool). In all the above models, the standard model assumptions were violated, and so the response variable was transformed using the natural logarithmic transformation to alleviate the problem. All relevant descriptive statistics and *P* values are reported in the Results section.

Ethical Considerations

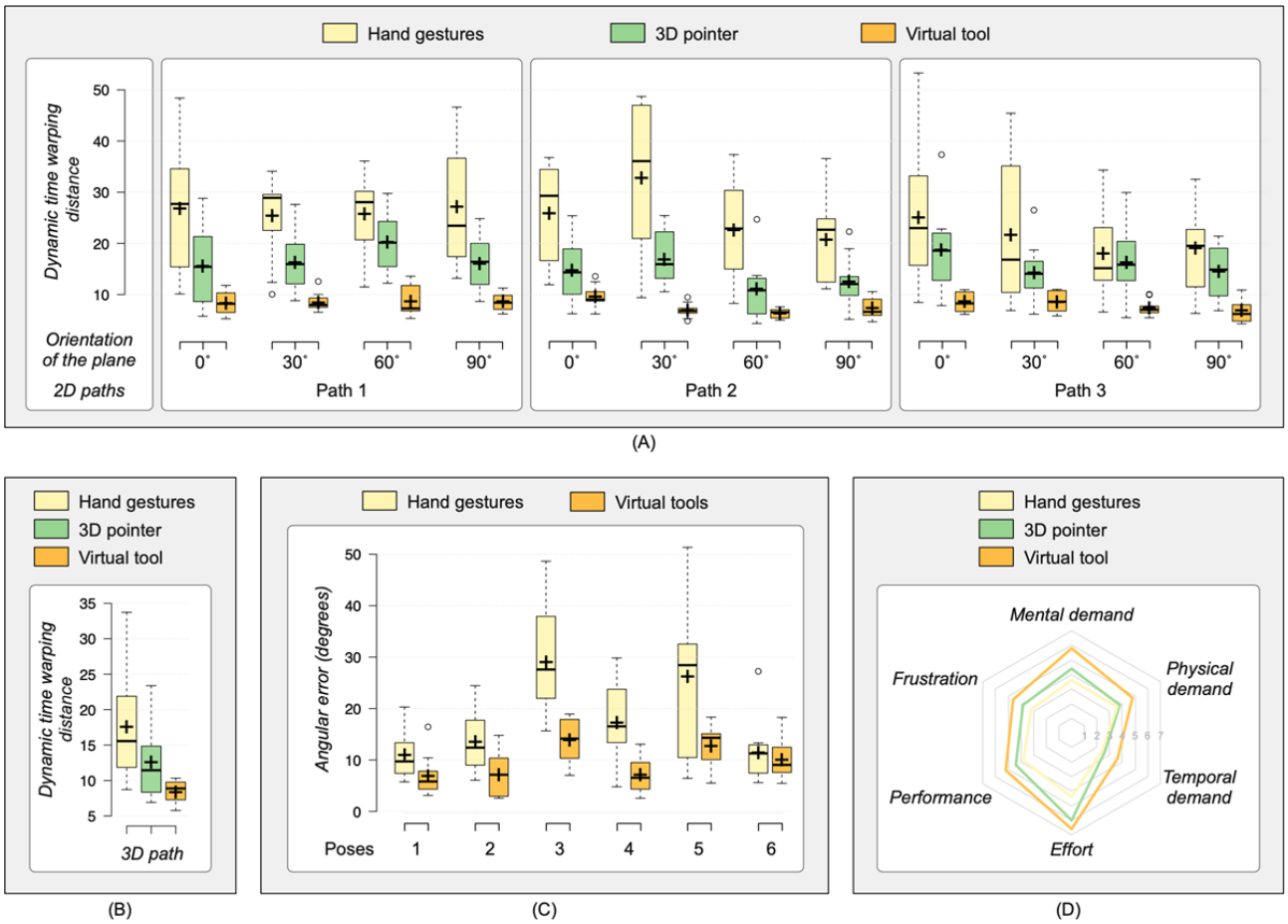
Institutional review board approval (MRC-03-23-786) was obtained from the Medical Research Center at Hamad Medical Corporation, Qatar. A research information sheet was used to obtain consent from the participants taking part in the study. All data collected was anonymized and stored securely. The participants enrolled in the study on a voluntary basis and were not provided any financial or other forms of compensation.

Results

All participants successfully completed the tasks in both scenarios. Figure 4 presents the observed results when the mentees received guidance from the mentor in the form of dynamic AR cues. During scenario A, the paths followed by the mentees had improved DTW distances when they were guided by a virtual tool as opposed to hand gestures and a 3D pointer. As shown in Table 3, the reduction was statistically significant when the virtual tool was used by the mentor compared with hand gestures ($P<.01$ in 12/13, 92.3% of the trials) and the 3D pointer ($P<.05$ in 11/13, 84.6% of the trials). This improvement in performance was consistently observed across the three 2D paths in each plane orientation (Figure 4A), as well as the 3D path (Figure 4B).

XSL•FO
RenderX

Figure 4. Average dynamic time warping distance from the intended path for each dynamic augmented reality cue while the mentees followed along: (A) three 2D paths at different orientations and (B) the 3D path during scenario A, (C) angular error orienting the jaw of the laparoscopic instrument for the 6 poses in scenario B, and (D) NASA Task Load Index scores reported by the mentees.



XSL•FO
RenderX

Table . Dynamic time warping distance during scenario A and angular error (in degrees) during scenario B.

Scenario, path or pose, and virtual plane orientation	Hand gestures, mean (SD)	3D pointer, mean (SD)	Virtual tool, mean (SD)	*P* value		
				Hand gestures vs 3D pointer	3D pointer vs virtual tool	Virtual tool vs hand gestures
Scenario A—dynamic time warping distance						
Path 1						
0°	26.9 (11.9)	15.6 (7.3)	8.3 (2.4)	.048	.03	.001
30°	25.5 (8.2)	16.2 (5.6)	8.4 (1.8)	.03	.08	<.001
60°	25.8 (7.9)	20.2 (5.6)	8.6 (2.9)	.22	.001	<.001
90°	27.1 (12.0)	16.0 (5.3)	8.6 (1.6)	.03	.004	<.001
Path 2						
0°	25.9 (10.0)	14.8 (5.9)	9.6 (2.1)	.03	<.001	<.001
30°	32.8 (15.0)	16.9 (5.1)	6.9 (1.4)	.001	.06	.001
60°	22.6 (10.0)	11.1 (5.8)	6.4 (0.9)	.15	.03	<.003
90°	20.7 (8.0)	12.6 (5.0)	7.4 (2.1)	.03	.03	<.001
Path 3						
0°	25.1 (13.6)	18.8 (8.1)	8.7 (1.9)	.65	.004	.001
30°	21.7 (13.0)	14.3 (5.7)	8.6 (2.2)	.03	.001	<.001
60°	18.0 (8.6)	16.3 (7.0)	7.4 (1.5)	.04	.02	.001
90°	19.2 (8.8)	14.5 (5.4)	6.9 (2.3)	.17	.002	.03
3D path	17.6 (7.3)	12.6 (5.1)	8.4 (1.6)	.08	.03	<.001
Scenario B—angular error (degrees)						
Pose 1	11.0 (4.4)	—[a]	6.9 (4.0)	—	—	.03
Pose 2	13.5 (6.1)	—	7.2 (4.2)	—	—	.01
Pose 3	29.0 (10.4)	—	13.9 (4.3)	—	—	<.001
Pose 4	17.3 (8.1)	—	7.1 (3.4)	—	—	<.001
Pose 5	26.2 (15.0)	—	12.8 (4.5)	—	—	.03
Pose 6	11.5 (6.2)	—	10.1 (3.7)	—	—	.66

[a]Not applicable.

In scenario B, a substantial decrease in angular error was noted when the mentees were guided by the virtual tool as compared with hand gestures (P<.05 for 5 out of 6 poses). Pose 6, located toward the edge of the screen, did not have a statistically significant difference in angular error (Figure 4C).

The scores from the NASA-TLX (Figure 4D) suggest that the use of the 3D pointer and the virtual tool was relatively demanding for the mentees compared with hand gestures. However, the overall differences in NASA-TLX scores were not substantial (Table 4).

XSL•FO
RenderX

Table . Scores from the NASA Task Load Index questionnaire.

Item	Question	Scale	Hand gestures, mean (SD)	3D pointer, mean (SD)	Virtual tool, mean (SD)
Mental demand	"How mentally demanding were the tasks in both the scenarios?"	1 - 10 (low-high)	3.6 (1.7)	4.4 (1.9)	5.8 (2.4)
Physical demand	"How physically demanding were the tasks in both the scenarios?"	1 - 10 (low-high)	3.6 (2.3)	3.8 (2.0)	4.8 (2.8)
Temporal demand	"How hurried or rushed was the pace of the tasks in both the scenarios?"	1 - 10 (low-high)	2.4 (1.5)	2.6 (1.3)	3.6 (1.5)
Performance	"How successful were you in accomplishing what you were asked to do?"	1 - 10 (perfect-failure)	5.6 (2.3)	4 (1.2)	3.4 (2.1)
Effort	"How hard did you have to work to accomplish your level of performance?"	1 - 10 (low-high)	3.8 (2.2)	4.4 (1.5)	5.2 (2.4)
Frustration	"How insecure, discouraged, irritated, stressed, and annoyed were you?"	1 - 10 (low-high)	3.2 (1.8)	3.8 (1.3)	4.6 (1.8)

Discussion

This study presents an assessment of 3 different dynamic AR cues, namely, hand gestures, a 3D pointer, and a virtual tool, used in telementoring during laparoscopic surgery. Integrating dynamic AR cues is necessary as static AR cues (telestrations) alone cannot fully reproduce the instructions conveyed during in-person guidance. The aim of this user study was to identify the most suitable dynamic AR cue for surgical telementoring. Overall, our results showed that the mentees were able to replicate the motions indicated by the mentor and follow the required path, demonstrating the feasibility of using dynamic AR cues. An important result of this study is that there was an improvement in performance when the mentor used the virtual tool to guide the mentees (as opposed to hand gestures or the 3D pointer). This was depicted in Figures 4A and B, where the nonoverlapping box plots across different plane orientations indicate a significant decrease in the DTW distance during scenario A. This indicates that the mentees could follow the required paths accurately in not only the 2D space but also the 3D space. As a result, it can be inferred that virtual tools would be useful during laparoscopic surgery. Such surgeries are performed in the insufflated body cavity with multiple tissue structures that vary in distance from the laparoscope's distal end. Therefore, having cues that allow for depth perception has the potential to improve accuracy while reducing procedure time and incidence of errors [32]. Similarly, in scenario B, lower angular errors were noted for the poses guided by the virtual tool than for those guided by hand gestures. This is a valuable observation as the ability to communicate and reproduce the precise jaw orientation of a laparoscopic tool is essential for avoiding damage to vital anatomical structures during live surgery [4]. Examples of such surgical steps include placing the initial snip on tissue structure, cutting a vessel, or placing staples for incision closure. Virtual tools would be more efficient compared with hand gestures while guiding the mentee to easily communicate the optimal position and orientation of the sharp scissors used for cystic duct or artery dissection during cholecystectomy. The direction of the curved side of the sharp scissors with respect to anatomical landmarks can be conveyed effortlessly [4]. Similarly, the appropriate orientation of a harmonic scalpel (consisting of an active blade and inactive jaw) can also be conveyed while separating the gallbladder from the liver. Conveying such information is critical as incorrect handling can result in damage to organs and bleeding, compromising patient safety [33].

The proposed work is not intended to replace the use of static 2D cues such as annotations or telestrations during laparoscopic surgical telementoring. The optimal cue for a particular surgical telementoring session can vary with type and frequency of instructions that a mentor (expert surgeon) must convey to a mentee (operating surgeon). This is dependent upon the complexity of the surgical procedure as well as the difference in expertise level between the mentor and mentee. When the mentor and mentee are of comparable expertise, telementoring using audio or static AR cues would be sufficient. Dynamic AR cues can assist in gaining microskills particular to an unfamiliar surgical technique when the mentee already possesses sufficient macroskills such as anatomical expertise [19]. Additionally, a prior review of dynamic AR cues used in minimally invasive surgeries has noted extensive use of surgical tools for telementoring during suturing among medical students and the use of pointers during cholecystectomy procedures among surgical trainees [5]. Further studies with surgeons will help assess the impact of dynamic AR cues in various surgical scenarios in which significant differences in expertise exist

XSL•FO
RenderX

between the mentor (expert surgeon) and mentee (operating surgeon).

Previous studies have shown that mentees report lower workload while receiving instructions using dynamic AR cues than using audio cues alone [3,7]. However, when assessing the workload perceived by the mentees for each of the dynamic AR cues in this study, there was no statistically significant difference between the NASA-TLX scores for hand gestures, the 3D pointer, or the virtual tool. Some of the mentees noted higher mental and physical demand while being guided by the virtual tool than by the other cues. This was attributed to the extra care taken by the mentee to align exactly with the mentor's virtual tool. Although virtual tools might have resulted in a perceived higher demand, the results indicate that the mentees were able to understand the required position and orientation of the tool (which was not conveyed by the other cues). It is important to bear in mind that the use of the NASA-TLX as an outcome measure would receive a relatively lower rating according to the Medical Education Research Study Quality Instrument, indicating a weaker methodological quality, as it measures subjective user perceptions [34]. A reasonable approach would be to conduct multiple trials and observe the difference in perceived workload for each dynamic AR cue over the course of the trials.

Laparoscopic surgery is considered the standard approach for many procedures (such as cholecystectomy and colectomy) due to its advantage for the patient in terms of lower risk of complications and shorter recovery period [6]. Nevertheless, adoption of laparoscopic surgery continues to be limited worldwide due to its steep learning curve and the complex instrument-handling skills required. Given its clear advantage over the open approach, laparoscopic surgery might be initiated without in-person supervision in some countries [35,36]. However, nonmentored laparoscopic surgeries are associated with higher rates of conversion to open surgery and complications and longer hospital stays than mentored ones [36,37]. This is due to the challenges faced by inexperienced surgeons working with an operative view that is drastically different from that in open surgery, resulting in dissection of incorrect layers or improper placement of the initial incision [36]. Telementoring overcomes this challenge by connecting experienced surgeons with the operating surgeon. However, the amount of information that can be transferred between the experienced and operating surgeons is dependent upon the quality of the cues rendered. Dynamic AR cues enhance the information by depicting the exact tool-tissue interaction required during the surgery.

In this study, the dynamic AR cues used by the mentor were displayed to the mentees on a 2D screen. However, a typical surgical field consists of anatomical structures located at varying depth levels. With limited depth perception, it is challenging to accurately convey instructions, especially when the target anatomy is located deep within the operative field. For future work, it would be useful to incorporate multiple cameras to capture and render a 3D model of the mentor's hand. In this way, the mentee will be able to perceive depth while being guided by hand gestures (which is lacking in the current setup). Furthermore, low-cost immersive head-mounted display (HMD) devices (such as the mixed reality–based Meta Quest 3 headset or Apple Vision Pro) could be used by the remote mentor to perceive a stereoscopic view of the surgical field obtained from a 3D laparoscopic camera used at the mentee's side. This would also allow the expert surgeon to convey the dynamic AR cues in a 3D space and better demonstrate the required motions to the operating surgeon. In addition, instead of using a traditional bulky workstation, such as a laptop or desktop connected to external input devices to manipulate dynamic AR cues, these HMD devices incorporate built-in sensors that can directly track hand gestures. The captured hand movements can then be translated into dynamic AR cues in real time, eliminating the need for intermediary input devices. Furthermore, these HMD devices are portable and just need an internet connection to link to the operating room.

Several limitations and future directions should be considered when interpreting the findings of this study. First, this study was limited to 10 participants as mentees due to budget constraints. The primary aim of this study was exploratory in nature, and despite the small sample size, the findings provide preliminary evidence that may help inform and guide future large-scale studies. Second, while previously published articles have already established favorable results while using dynamic AR cues when compared with audio cues alone [6,9,23], it would be nevertheless useful to incorporate audio cues to simulate a real surgical training setting. The audio cues are useful when both mentor and mentee possess comparable surgical macroskills, such as anatomical knowledge, instrument maneuvering, and identification of critical structures. However, when a mentor remotely guides a less experienced mentee through a newly acquired surgical technique, the mentee may not yet have mastered technique-specific microskills, including visual tactility, economy of movement, and tissue handling. In such cases, the overlaid dynamic AR cues are expected to expedite skill acquisition by providing clear, unambiguous visual guidance independent of verbal instruction (audio cue). Third, implementing dynamic AR cues in a live surgical setting where objective clinical outcomes (such as operating time, error rate, or incidence of intraoperative complications) are measured will be required to fully evaluate the effectiveness of using the AR cues. While this study focused on laparoscopic surgery, it would be interesting to explore the application of AR cues for robot-assisted procedures where the surgical instruments have additional degrees of freedom.

In conclusion, this work has indicated that the use of virtual tools as dynamic AR cues results in improved performance compared with hand gestures and 3D pointers. However, there was no significant difference in the perceived workload among the 3 AR cues. Moving forward, further exploration on a multi-institutional scale is recommended. Studies integrating various dynamic AR cues during real-world surgery would also be essential to fully assess the AR cues' clinical value during surgical telementoring.

XSL•FO
RenderX

Acknowledgments

All opinions, findings, conclusions, or recommendations expressed in this work are those of the authors and do not necessarily reflect the views of their sponsors.

Funding

The research reported in this publication was supported by the Qatar Research, Development, and Innovation Council Academic Research Grant award ARG01-0430-230047.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

JP and DS prototyped the system. HH assisted in manuscript writing. JA provided input from the surgical perspective. PT performed the statistical analyses. ZD and NVN were responsible for the conception and design of the study.

Conflicts of Interest

None declared.

References

1. Fernandes RD, Ghasroddashti A, Sorefan-Mangou F, et al. Educational effectiveness of telementoring as a continuing professional development intervention for surgeons in practice: a systematic review. Ann Surg Open 2023 Sep 28;4(4):e341. [doi: 10.1097/AS9.0000000000000341] [Medline: 38144497]
2. Lin Z, Xu Z, Liu H, et al. Robotic manipulator‐assisted omnidirectional augmented reality for endoluminal intervention telepresence. Adv Intell Syst 2024 Jan;6(1):2300373. [doi: 10.1002/aisy.202300373]
3. Wild C, Lang F, Gerhäuser AS, et al. Telestration with augmented reality for visual presentation of intraoperative target structures in minimally invasive surgery: a randomized controlled study. Surg Endosc 2022 Oct;36(10):7453-7461. [doi: 10.1007/s00464-022-09158-1] [Medline: 35266048]
4. Lambert S, Voros S, Canlorbe G, Troccaz J, Avellino I. Understanding takeovers and telestration in laparoscopic surgery to inform telementoring system design. In: CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems: Association for Computing Machinery. [doi: 10.1145/3613904.3641978]
5. Hamza H, Aboumarzouk OM, Al-Ansari A, Navkar NV. Dynamic augmented reality cues for telementoring in minimally invasive surgeries: scoping review. J Med Internet Res 2025 Feb 3;27:e63939. [doi: 10.2196/63939] [Medline: 39899360]
6. Cizmic A, Müller F, Wise PA, et al. Telestration with augmented reality improves the performance of the first ten ex vivo porcine laparoscopic cholecystectomies: a randomized controlled study. Surg Endosc 2023 Oct;37(10):7839-7848. [doi: 10.1007/s00464-023-10360-y] [Medline: 37612445]
7. Felinska EA, Fuchs TE, Kogkas A, et al. Telestration with augmented reality improves surgical performance through gaze guidance. Surg Endosc 2023 May;37(5):3557-3566. [doi: 10.1007/s00464-022-09859-7] [Medline: 36609924]
8. El-Asmar JM, Labban M, El-Hajj A. Integration of aquablation through telemetry: an alternative to onsite proctoring? World J Urol 2021 Sep;39(9):3473-3479. [doi: 10.1007/s00345-021-03603-x] [Medline: 33547926]
9. Huettl F, Heinrich F, Boedecker C, et al. Real-time augmented reality annotation for surgical education during laparoscopic surgery: results from a single-center randomized controlled trial and future aspects. J Am Coll Surg 2023 Aug 1;237(2):292-300. [doi: 10.1097/XCS.0000000000000712] [Medline: 37042553]
10. Heinrich F, Huettl F, Schmidt G, et al. HoloPointer: a virtual augmented reality pointer for laparoscopic surgery training. Int J Comput Assist Radiol Surg 2021 Jan;16(1):161-168. [doi: 10.1007/s11548-020-02272-2] [Medline: 33095424]
11. Shabir D, Balakrishnan S, Padhan J, et al. Evaluating a remote tele-mentoring gameplay setup for teaching laparoscopic suturing skills. Presented at: 2023 IEEE 11th International Conference on Serious Games and Applications for Health (SeGAH); Aug 28-30, 2023. [doi: 10.1109/SeGAH57547.2023.10253770]
12. Shabir D, Kharbech S, Padhan J, et al. Telementoring system assessment integrated with laparoscopic surgical simulators. Presented at: 2023 IEEE 5th Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS); Jun 2-4, 2023. [doi: 10.1109/ECBIOS57802.2023.10218557]
13. Shabir D, Balakrishnan S, Padhan J, et al. Tele-mentoring using augmented reality: a feasibility study to assess teaching of laparoscopic suturing skills. Presented at: 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS); Jun 22-24, 2023. [doi: 10.1109/CBMS58004.2023.00193]
14. Lowry B, Johnson GGRJ, Vergis A. Merged virtual reality teaching of the fundamentals of laparoscopic surgery: a randomized controlled trial. Surg Endosc 2022 Sep;36(9):6368-6376. [doi: 10.1007/s00464-021-08939-4] [Medline: 34981231]

XSL•FO
RenderX

15. Vera AM, Russo M, Mohsin A, Tsuda S. Augmented reality telementoring (ART) platform: a randomized controlled trial to assess the efficacy of a new surgical education technology. Surg Endosc 2014 Dec;28(12):3467-3472. [doi: 10.1007/s00464-014-3625-4] [Medline: 24962856]
16. Jarc AM, Shah SH, Adebar T, et al. Beyond 2D telestration: an evaluation of novel proctoring tools for robot-assisted minimally invasive surgery. J Robot Surg 2016 Jun;10(2):103-109. [doi: 10.1007/s11701-016-0564-1] [Medline: 26914650]
17. Kowalewski KF, Hendrie JD, Schmidt MW, et al. Development and validation of a sensor- and expert model-based training system for laparoscopic surgery: the iSurgeon. Surg Endosc 2017 May;31(5):2155-2165. [doi: 10.1007/s00464-016-5213-2] [Medline: 27604368]
18. Long Y, Li C, Dou Q. Robotic surgery remote mentoring via AR with 3D scene streaming and hand interaction. Comput Methods Biomech Biomed Eng Imaging Vis 2023 Jul 4;11(4):1027-1032. [doi: 10.1080/21681163.2022.2145498]
19. Shabir D, Abdurahiman N, Padhan J, et al. Preliminary design and evaluation of a remote tele-mentoring system for minimally invasive surgery. Surg Endosc 2022 May;36(5):3663-3674. [doi: 10.1007/s00464-022-09164-3] [Medline: 35246742]
20. Shabir D, Abdurahiman N, Padhan J, et al. Towards development of a tele-mentoring framework for minimally invasive surgeries. Int J Med Robot 2021 Oct;17(5):e2305. [doi: 10.1002/rcs.2305] [Medline: 34256415]
21. Youssef SC, Sabbubeh B, Haram K, et al. Augmented reality robot-assisted radical prostatectomy with PROXIMIE: preliminary clinical experience. Urol Video J 2022 Oct;16:100187. [doi: 10.1016/j.urolvj.2022.100187]
22. Feng Y, Li K, Semsar A, et al. Communication cost of single-user gesturing tool in laparoscopic surgical training. Presented at: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019. [doi: 10.1145/3290605.3300841]
23. Feng Y, McGowan H, Semsar A, et al. A virtual pointer to support the adoption of professional vision in laparoscopic training. Int J Comput Assist Radiol Surg 2018 Sep;13(9):1463-1472. [doi: 10.1007/s11548-018-1792-9] [Medline: 29796835]
24. Jarc AM, Stanley AA, Clifford T, Gill IS, Hung AJ. Proctors exploit three-dimensional ghost tools during clinical-like training scenarios: a preliminary study. World J Urol 2017 Jun;35(6):957-965. [doi: 10.1007/s00345-016-1944-x] [Medline: 27671899]
25. Davis MC, Can DD, Pindrik J, Rocque BG, Johnston JM. Virtual interactive presence in global surgical education: international collaboration through augmented reality. World Neurosurg 2016 Feb;86:103-111. [doi: 10.1016/j.wneu.2015.08.053] [Medline: 26342783]
26. Aghazadeh F, Zheng B, Tavakoli M, Rouhani H. Spatiotemporal characteristics of eye-hand coordination among different skill levels in laparoscopic surgery. Surg Endosc 2026 Feb;40(2):1199-1209. [doi: 10.1007/s00464-025-12368-y] [Medline: 41225171]
27. Shabir D, Anbatawi M, Padhan J, et al. Evaluation of user-interfaces for controlling movements of virtual minimally invasive surgical instruments. Int J Med Robot 2022 Oct;18(5):e2414. [doi: 10.1002/rcs.2414] [Medline: 35486635]
28. Breedveld P, Stassen HG, Meijer DW, Jakimowicz JJ. Manipulation in laparoscopic surgery: overview of impeding effects and supporting aids. J Laparoendosc Adv Surg Tech A 1999 Dec;9(6):469-480. [doi: 10.1089/lap.1999.9.469] [Medline: 10632507]
29. Hart SG, Staveland LE. Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock PA, Meshkati N, editors. Advances in Psychology: North-Holland; 1988:139-183. [doi: 10.1016/S0166-4115(08)62386-9]
30. Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans Acoust Speech Signal Process 1978 Feb;26(1):43-49. [doi: 10.1109/TASSP.1978.1163055]
31. Despinoy F, Zemiti N, Forestier G, Sánchez A, Jannin P, Poignet P. Evaluation of contactless human-machine interface for robotic surgical training. Int J Comput Assist Radiol Surg 2018 Jan;13(1):13-24. [doi: 10.1007/s11548-017-1666-6] [Medline: 28914409]
32. Alaraimi B, El Bakbak W, Sarker S, et al. A randomized prospective study comparing acquisition of laparoscopic skills in three-dimensional (3D) vs. two-dimensional (2D) laparoscopy. World J Surg 2014 Nov;38(11):2746-2752. [doi: 10.1007/s00268-014-2674-0] [Medline: 25002241]
33. Ebina K, Abe T, Yan L, et al. A surgical instrument motion measurement system for skill evaluation in practical laparoscopic surgery training. PLoS ONE 2024 Jun 25;19(6):e0305693. [doi: 10.1371/journal.pone.0305693] [Medline: 38917181]
34. Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association between funding and quality of published medical education research. JAMA 2007 Sep 5;298(9):1002-1009. [doi: 10.1001/jama.298.9.1002] [Medline: 17785645]
35. Kim JH, Lee IK, Kang WK, Oh ST, Kim JG, Lee YS. Initial experience of a surgical fellow in laparoscopic colorectal cancer surgery under training protocol and supervision: comparison of short-term results for 70 early cases (under supervision) and 73 late cases (without supervision). Surg Endosc 2013 Aug;27(8):2900-2906. [doi: 10.1007/s00464-013-2851-5] [Medline: 23436093]
36. Bakula B. Is non-mentored initiation of laparoscopic colorectal surgery safe? Single surgeon initial experience with the first 40 cases. Front Surg 2023 Sep 5;10:1196037. [doi: 10.3389/fsurg.2023.1196037] [Medline: 37744727]

XSL•FO
RenderX

37. Natarajan L, Gosavi AT, Wataganara T, et al. Mentoring a surgical team towards procedural competence in the early learning curve for selective fetoscopic laser photocoagulation. Singapore Med J 2022 May;63(5):274-282. [doi: 10.11622/smedj.2020137] [Medline: 36043312]

Abbreviations

AR: augmented reality
DTW: dynamic time warping
HMD: head-mounted display
NASA-TLX: NASA Task Load Index

Edited by A Kushniruk; submitted 27.Oct.2025; peer-reviewed by B Niroomand, KW Rio; revised version received 13.Jan.2026; accepted 09.Feb.2026; published 04.Mar.2026.

Please cite as:
Padhan J, Hamza H, Shabir D, Abinahed J, Tsiamyrtzis P, Deng Z, Navkar NV
Dynamic Augmented Reality Cues for Telementoring in Laparoscopic Surgery: Usability Study
JMIR Hum Factors 2026;13:e86566
URL: https://humanfactors.jmir.org/2026/1/e86566
doi:10.2196/86566

© Jhasketan Padhan, Hawa Hamza, Dehlela Shabir, Julien Abinahed, Panagiotis Tsiamyrtzis, Zhigang Deng, Nikhil V Navkar. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 4.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

An Automated Curriculum to Support Behavioral Health Counseling Among Pediatric Residents: Usability Study

Liam Fleck[1], BA; Rachel Herbst[1,2], PhD; Andrea Meisman[1], MA; Thomas Talbot[3,4], MD; Michael Glisson[1], BAS; Max Remington[4], BFA; Micah Goldson[1], MS; Francis Real[1,2], MEd, MD

[1]Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

[2]University of Cincinnati, 3333 Burnet Ave, Cincinnati, OH, United States

[3]USC Institute for Creative Technologies, Los Angeles, CA, United States

[4]BreakAway, Ltd, Hunt Valley, MD, United States

Corresponding Author:
Francis Real, MEd, MD
Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

Abstract

Background: Behavioral health concerns are common in pediatric practice, with pediatricians reporting a lack of skills related to providing effective behavioral management strategies to parents. A prior human-facilitated, screen-based virtual reality training curriculum proved effective in enhancing behavioral health communication skills among pediatric residents. However, barriers to spread and scale of the curriculum included the need for human facilitation.

Objective: This study explored the usability of an automated virtual reality–based behavioral health anticipatory guidance curriculum for pediatric residents.

Methods: Through a partnership with BreakAway Ltd, an automated prototype was developed that required verbalization by users to support progress through simulated scenarios and included the receipt of personalized feedback to enable deliberate practice of communication skills. Usability of the prototype was assessed using mixed methods that included in-person completion of the prototype, a semistructured interview, and completion of survey instruments including the System Usability Scale and the Measurement, Effects, Conditions: Spatial Presence Questionnaire.

Results: Nine individuals completed usability testing. Qualitatively, users indicated that the system was easy to use, realistic, gamified, and likely most helpful for novice learners. Quantitatively, the ease of system usability was rated highly with some limitations related to spatial presence noted.

Conclusions: Usability testing of an automated curriculum to support behavioral health counseling skills in pediatric residents was completed, providing data to support adaptations in preparation for implementation.

(JMIR Hum Factors 2026;13:e77438) doi:10.2196/77438

KEYWORDS

virtual reality; artificial intelligence; AI; medical education; health communication; behavioral health; motivational interviewing

Introduction

Background

Although pediatricians frequently discuss behavioral health concerns with families [1], they report a lack of training on how to effectively address these concerns using evidence-based approaches and collaborative communication [2,3]. This represents a critical training gap, as the implementation of effective behavioral management strategies by parents to address typical childhood behaviors (eg, tantrums) can mitigate the development of future behavioral health disorders [4]. With approximately 20% of children currently having a diagnosis of a behavioral health disorder [5], identification of training strategies to support future pediatricians' skills in delivering behavioral health anticipatory guidance (BHAG) is needed [6].

Recent curricula aiming to support BHAG skills in pediatric residents have primarily relied upon experts to deliver content via didactic or case-based discussions, limiting the opportunity for spread and scale of effective interventions [7,8]. Simulation-based medical education in the form of patient actors has been incorporated into training pediatric residents on BHAG; however, limitations related to accessibility, realism, and psychological safety have been reported [9]. Advances in technology offer an opportunity for broad dissemination of theory-based education. Screen-based virtual reality (VR) is one technology that can be leveraged to support BHAG skill development in pediatric residents to optimize pediatric health outcomes.

XSL•FO
RenderX

Prior Work

We previously developed a novel screen-based VR curriculum to enhance BHAG competencies among pediatric residents through deliberate practice of skills [10-12]. A human facilitator drove both the VR responses of graphical characters in scenarios and provided personalized performance feedback following each simulated scenario. The VR curriculum proved highly acceptable and efficacious in enhancing BHAG and health care communication skills [10,12-14]. During usability testing of the human-facilitated curriculum, residents reported the curriculum as realistic, engaging, practical, appropriately scaffolded, and psychologically safe. They indicated the curriculum was easy to use and reported high levels of immersion and spatial presence within the virtual environment [10]. The spread and scale of this education was limited due to the need for highly trained human facilitators. Artificial intelligence (AI) offers an opportunity to explore how automation may support maintenance and scale of such VR-based educational interventions.

AI allows users to interact with computer systems that use data sources to interpret user inputs and deliver curated responses [15]. The use of AI in medical education and simulation has been limited; prior research indicates limitations related to realism and accuracy [16,17]. Thus, we sought to assess the feasibility and acceptability of an automated curriculum to support BHAG and communication skills in pediatric residents.

Methods

Ethical Considerations

This study was reviewed and determined to be exempt by the Cincinnati Children's Hospital Medical Center Institutional Review Board (2023-0314). A waiver of documentation of informed consent was granted; therefore, prior to data collection, participants reviewed a study information sheet, and participation in study procedures indicated their consent. All study team members adhered to institutional procedures governing data access and privacy. Compensation was provided to all participants.

Curriculum Development

To develop the automated intervention, we partnered with BreakAway Ltd, a game developer that created a conversational interaction system titled Standard Patient Studio (SPS) [18]. SPS uses a cloud-based AI platform in which progression through a simulated scenario is driven by a user's verbalizations. A multidisciplinary team that included 2 pediatricians, 1 pediatric psychologist, 3 software developers with expertise in AI, and 2 research assistants adapted the human-facilitated VR curriculum to the SPS platform. Procedures included scenario storyboarding with branching logic using the prior human-facilitated VR scenarios as a guide, with iterative scenario development testing by study team members. The final curricular prototype yielded 2 scenarios that replicated the human-facilitated VR curriculum's learning objectives and used identical verbal statements by the graphical parent. Potential answer options for the learner were displayed on the screen during each scenario to support scaffolded learning and to inform a learner's spontaneous verbalizations, thereby ensuring progression through a scenario. The answer option most closely aligned with the user's verbalizations was selected by the natural language understanding system. This system categorizes user input via a medical taxonomy reference that assigns an appropriate patient action. In prior trials, this system has demonstrated a 92% appropriate response rate [19]. On the basis of the selected answer option, the user received immediate feedback (ie, excellent, acceptable, or poor response) via color-coded emojis that were previously embedded in the SPS platform. Upon scenario completion, users received feedback indicating why an answer selection was preferred or not preferred. The feedback mirrored the human-facilitated feedback by identifying the specific BHAG and motivational interviewing skills demonstrated during a scenario and where there were missed opportunities to demonstrate such skills. A low performance score required users to repeat the scenario, which mirrored the deliberate practice approach used for the human-facilitated VR training.

Usability Testing

Usability testing, a critical component of intervention development, which seeks to systematically evaluate the extent to which an intervention achieves its intended purpose by engaging experts to identify weaknesses, occurred using mixed methods [20,21]. Health care team members, including senior faculty, psychologists, simulation educators, and pediatric residents, were purposefully sampled to pilot the curriculum. Usability testing occurred in person, where users were asked to "think aloud," verbalizing their perspectives as they navigated the automated curriculum [22]. Following completion, participants underwent a semistructured interview to explore their overall impressions (Textbox 1). Interviews were recorded, transcribed, and analyzed by 2 raters (AM and FR) using the rigorous and accelerated data reduction (RADAR) technique, a rapid qualitative analysis approach [23]. RADAR uses a 5-step approach to data reduction through the iterative refinement of data tables and is well suited for studies with narrow research questions. The raters ensured consensus on reduction decisions at each step. Trustworthiness of the qualitative data was also supported through the use of an interview guide that generated descriptive responses (credibility), multiple analyzers and peer debriefing (dependability), and recording of detailed procedures (confirmability) [24]. Participants also completed the System Usability Scale (SUS) [25] to examine the platform's ease of use and a subset of items from the Measurement, Effects, Conditions: Spatial Presence Questionnaire (MEC-SPQ) [26] to assess for immersion in the virtual environment.

XSL•FO
RenderX

Textbox 1. Interview guide.

Usability interview guide items

- What are your overall impressions of the automated curriculum?
- How easy or difficult was the curriculum to use? Tell me more.
- How was your interaction with the avatar characters?
  - What would you change?
- What were your impressions of the feedback you received after the simulation?
  - Did it seem to accurately reflect your performance?
  - How easy or difficult was it to interpret?
- What one change would you make to the automated curriculum?
- Would you recommend the automated curriculum for learners? Why or why not?

Results

A total of 9 individuals, including 2 pediatric psychologists, 3 senior pediatric faculty, 2 pediatric residents, and 2 simulation educators, underwent testing during a 1-week period in August 2024. Participants had a mean age of 36 (SD 7.9) years and were mostly women (n=8, 89%), White (n=8, 89%), and non-Hispanic (n=9, 100%).

Qualitative

Of the 9 participants, 8 (89%) provided qualitative data. Participants described the scenarios as relevant and realistic. Overall, they reported that the system was easy to use and correctly coded their verbalizations. The study team observed, and participants reported, that there was a learning curve associated with using the online microphone. Participants indicated appreciation for the nuanced answer options and level of difficulty, although some participants desired decreased scaffolding and increased free agency. Some participants described a game-based experience, attributed to the immediate feedback in the form of color-differentiated emojis. Some viewed the game mechanics integration positively, while others felt it detracted from realism. All participants (n=8) indicated that they would recommend the curriculum for learners. Due to the inclusion of game-based elements and the select option format, some participants believed the curriculum was best suited for novice learners (eg, medical students and residents) rather than practicing pediatricians. All participants indicated that the feedback accurately reflected their performance. When asked what one change participants would make to the automated curriculum, 3 (38%) requested more free agency and less scaffolding during the simulations, 2 (25%) wanted shorter answer options, 2 (25%) wanted more variability in terms of graphical character responses, and 1 (13%) wanted more specific feedback regarding incorrect answer selections.

Table 1 includes exemplar quotes informing these data interpretations.

XSL•FO
RenderX

Table . Principal themes and supporting quotes.

Theme	Supporting quotes
Easy to use	• "I thought it was very easy. I had no problems using the system and getting through the scenario." [ID 26] • "It's very self-explanatory, and you just have to like push a button. It's not super high, not high-tech in a way that I'm like I have no clue what to do with this." [ID 21]
Realistic scenarios	• "I think it's a practical way, some of the things that the parents said are very realistic and real life that we hear all the time, so I thought that was realistic." [ID 24] • "I felt the situation was very realistic, and it felt like the, it just, it felt accurate. That's my biggest reaction. These things come up all the time, and kids do have tantrums in the office all the time, so this is very relevant." [ID 26]
Accurate character responses	• "I feel like she responded appropriately to what I said." [ID 22] • "I think I would say it feels representative in some ways of what like parents might say." [ID 27]
Game-based design	• "I felt more like it was like a game to find the best answer than an actual interaction with the person. It felt more like a trivia game." [ID 26] • "It [the feedback] felt like a score, like a game and a score." [ID 27]
Potential for more realistic characters	• "I would like the room to look more realistic, and I'd like Phyllis [the graphical parent] to look more realistic, and I'd like Alex [the graphical child] to actually be with Phyllis and not a pop-out video." [ID 30] • "I mean, if it looked more realistic, obviously, that would be nice. And her voice sounds a little bit monotone. Like if it could sound a little bit more natural, then I think I would feel more bought in that it feels like a more realistic patient interaction." [ID 22]
Accurate and clear feedback	• "I like this [feedback] page, especially because it breaks it up, and it gives specific feedback. I guess sometimes I didn't pick the most egregious answers, so I didn't always get like the bad feedback. But I do like that you could see kind of what the other choices are, and you could kind of see what the scale is." [ID 29] • "I like that it's relatively clear [the feedback]. Obviously, you know, there's the green, yellow, red. Green is good. Yellow is kind of that middle of the road. Red is bad." [ID 23]
Most helpful for novice learners	• "I can see how it could have a lot of value for resident training and definitely helping residents get more experience." [ID 27] • "I think I would, especially for like interns or early learners because, I mean, it's low stakes with the avatar. And I also like the microphone aspect of it, because I think speaking the words instead of just where you've been clicking goes a long way." [ID 29]

Quantitative

The overall SUS mean score among participants was 78 (SD 9.8; range 65.0-92.5), indicating good to excellent usability of the system [27]. On the MEC-SPQ, 44% (4/9) of participants agreed or strongly agreed that it felt like they actually participated in the action of the simulation. Only 11% (1/9) agreed or strongly agreed that they felt as though they were physically present in the environment. (Table 2)

XSL•FO
RenderX

Table . Quantitative outcome metrics assessing system usability (System Usability Scale [25]) and spatial presence (adapted Measurement, Effects, Conditions: Spatial Presence Questionnaire [MEC-SPQ] [26]), which both use a 5-point Likert scale from strongly disagree (score=1) to strongly agree (score=5).

Items	Score, mean (SD)
System Usability Scale	
I think that I would like to use this system frequently	3.6 (0.9)
I found the system unnecessarily complex	1.9 (1.1)
I thought the system was easy to use	4.1 (0.6)
I think that I would need the support of a technical person to be able to use this system	1.4 (0.5)
I found the various functions in this system were well integrated	3.8 (0.8)
I thought there was too much inconsistency in this system	1.9 (0.9)
I would imagine that most people would learn to use this system very quickly	4.3 (0.5)
I found the system very cumbersome to use	1.6 (0.5)
I felt very confident using the system	4.0 (0.7)
I needed to learn a lot of things before I could get going with this system	1.7 (0.5)
Total	78.3 (9.8)
MEC-SPQ (subset)	
The virtual reality experience captured my senses (ie, it held my attention)	3.3 (1.0)
I dedicated myself completely to the virtual reality experience. (ie, I was not distracted)	3.4 (1.0)
I was able to make a good estimate of the size of the presented space	3.2 (0.8)
I felt as though I was physically present in the environment of the simulation	2.6 (1.0)
It seemed as though I actually took part in the action of the simulation	3.1 (0.9)
The objects in the simulation gave me the feeling that I could do things with them	2.4 (1.0)
Even now, I still have a concrete mental image of the spatial environment	3.6 (1.0)

Discussion

Principal Results

Adaptation of a human-facilitated VR curriculum on BHAG was successfully automated through the use of the SPS platform developed by BreakAway Ltd. Qualitative and quantitative data indicated that users could easily navigate the platform and interact with its components. Compared with the human-facilitated VR curriculum [10], users reported less spatial presence in the virtual environment. However, users still described the automated training as realistic and relevant. Users also endorsed a game-based aspect to the training experience, which was not reported with the prior human-facilitated VR curriculum [10]. Overall, users indicated that the current automated training might be most appropriate for novice learners. Our decision to scaffold the learning experience by displaying answer options supported reliable progression through the scenarios and safeguarded against misclassification of verbalizations, which can occur when using generative AI models [17,28-30]. However, more advanced learners may benefit from removing scaffolded answer options at a limited number of interaction points.

Limitations

This study had several limitations. First, our small, purposeful sample, ideal for usability testing [31-33] and intervention development, limited generalizability. However, recent literature by Guest et al [31,32] indicates that 6 to 7 interviews are likely sufficient to support theme generation for homogenous samples such as ours. Second, the rapid qualitative analysis method may have limited exploration of interview data. However, the RADAR technique is well suited for research questions that are narrow in scope [23].

Comparison with Prior Work

Behavioral health medical education lacks effective, innovative training curricula necessary to develop practice competencies [3]. By removing the need for a human facilitator or patient actors, AI-based communication training provides an opportunity for spread and scale [16,34]. Additional research is needed on how to best incorporate AI into education interventions to optimize learning outcomes that support

XSL•FO
RenderX

behavior change. We specifically sought to align the automated curriculum with deliberate practice principles, as this served as the theoretical foundation for the prior human-facilitated VR curriculum [11]. Deliberate practice is a personal, goal-oriented approach to training that focuses on attempting a behavior (eg, providing BHAG), receiving immediate feedback, and then repeating it until demonstrating skill mastery. Consistent with this theoretical framework, the automated curriculum provided a platform for rehearsal of specific verbiage, feedback based on each interaction point, repetition of skills over the two scenarios, and the opportunity to repeat scenarios. Aligning AI-based educational interventions with evidence-based adult learning theories will allow us to establish a literature base on how AI might most effectively support learning. This is particularly important for the topic of behavioral health, as recent studies describing curricular interventions often do not specifically indicate a theoretical foundation for the work [8,35,36]. As AI becomes integrated routinely into training, we should consider how we might best facilitate trust for learners interacting with AI systems as well as provide transparency regarding AI algorithms and source data [37].

Conclusions

Our study demonstrated the feasibility of developing an automated curriculum to support BHAG skills in pediatric residents through collaboration with an experienced industry partner. Automation of training curricula adds value by enhancing the capacity for scalability by reducing dependence on human facilitation. As such, automation also decreases the costs for participation, supporting equity among learners in accessing novel, evidence-based education. Next steps include updating the curriculum based on usability testing and implementing it among pediatric residents to assess its acceptability and effectiveness among that population and to explore how its outcomes compare to other training modalities, such as human-facilitated VR interventions. An evaluation of implementation strategies will also be critical to understand how such novel education can be integrated into resident curricula. Moreover, given the varying perceptions related to the game-based elements incorporated into the intervention, further exploration of how such game-based features impact attitudes and learning outcomes is warranted. We also plan to add an orientation regarding learning goals prior to intervention completion to align with best practices for promoting a constructive digital feedback environment [38]. In the future, we aim to refine the automated curriculum, including adjusting the level of free agency and gamified elements, to support learning across the continuum of health care providers.

Funding

This work was supported through the Cincinnati Children's Hospital Medical Innovation Fund award. Funders played no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Conflicts of Interest

TT and MR are employed or consult with BreakAway, Ltd. To minimize bias, these authors did not participate in the collection, management, or analysis of data. All other authors declare no other conflicts of interest.

References

1. Committee on Psychosocial Aspects of Child and Family Health and Task Force on Mental Health. Policy statement--the future of pediatrics: mental health competencies for pediatric primary care. Pediatrics 2009 Jul;124(1):410-421. [doi: 10.1542/peds.2009-1061] [Medline: 19564328]
2. McMillan JA, Land MJ, Leslie LK. Pediatric residency education and the behavioral and mental health crisis: a call to action. Pediatrics 2017 Jan;139(1):e20162141. [doi: 10.1542/peds.2016-2141] [Medline: 28011943]
3. Green CM, Foy JM, Earls MF, Committee on Psychosocial Aspects of Child and Family Health, Mental Health Leadership Work Group. Achieving the pediatric mental health competencies. Pediatrics 2019 Nov;144(5):e20192758. [doi: 10.1542/peds.2019-2758] [Medline: 31636144]
4. Fostering Healthy Mental, Emotional, and Behavioral Development in Children and Youth: A National Agenda: National Academies Press; 2019. URL: https://books.google.co.in/books/about/Fostering_Healthy_Mental_Emotional_and_B.html?id=PhXHDwAAQBAJ&source=kp_book_description&redir_esc=y [accessed 2026-04-02]
5. Sappenfield O, Alberto C, Minnaert J, Donney J, Lebrun-Harris L, Ghandour R. Adolescent mental and behavioral health, 2023. In: National Survey of Children's Health Data Briefs: Health Resources and Services Administration; 2018. URL: https://www.ncbi.nlm.nih.gov/books/NBK608531/ [accessed 2026-04-02]
6. McMillan JA, Land MJ, Tucker AE, Leslie LK. Preparing future pediatricians to meet the behavioral and mental health needs of children. Pediatrics 2020 Jan;145(1):e20183796. [doi: 10.1542/peds.2018-3796] [Medline: 31818865]
7. Manning A, Weingard M, Fabricius J, French A, Sendak M, Davis N. Be ExPeRT (behavioral health expansion in pediatric residency training): a case-based seminar. MedEdPORTAL 2023;19:11326. [doi: 10.15766/mep_2374-8265.11326] [Medline: 37534018]
8. Agazzi H, Dickinson S, Plant RM. Helping our toddlers, developing our children's skills: innovative behavioral management training for pediatric residents. Adv Pediatr 2022 Aug;69(1):13-21. [doi: 10.1016/j.yapd.2022.03.007] [Medline: 35985705]

XSL•FO
RenderX

9. Jones MR, Dadiz R, Baldwin CD, Alpert-Gillis L, Jee SH. Integrated behavioral health education using simulated patients for pediatric residents engaged in a primary care community of practice. Fam Syst Health 2022 Dec;40(4):472-483. [doi: 10.1037/fsh0000738] [Medline: 36508618]
10. Herbst R, Rybak T, Meisman A, et al. A virtual reality resident training curriculum on behavioral health anticipatory guidance: development and usability study. JMIR Pediatr Parent 2021 Jun 29;4(2):e29518. [doi: 10.2196/29518] [Medline: 34081601]
11. Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. Psychol Rev 1993;100(3):363-406. [doi: 10.1037/0033-295X.100.3.363]
12. Real FJ, Whitehead M, Ollberding NJ, et al. A virtual reality curriculum to enhance residents' behavioral health anticipatory guidance skills: a pilot trial. Acad Pediatr 2023;23(1):185-192. [doi: 10.1016/j.acap.2022.07.011] [Medline: 35870801]
13. Rollnick S, Miller WR. What is motivational interviewing? Behav Cogn Psychother 1995 Oct;23(4):325-334. [doi: 10.1017/S135246580001643X]
14. Erickson SJ, Gerstle M, Feldstein SW. Brief interventions and motivational interviewing with children, adolescents, and their parents in pediatric health care settings: a review. Arch Pediatr Adolesc Med 2005 Dec;159(12):1173-1180. [doi: 10.1001/archpedi.159.12.1173] [Medline: 16330743]
15. OpenAI. URL: https://openai.com [accessed 2026-04-02]
16. Merritt C, Glisson M, Dewan M, Klein M, Zackoff M. Implementation and evaluation of an artificial intelligence driven simulation to improve resident communication with primary care providers. Acad Pediatr 2022 Apr;22(3):503-505. [doi: 10.1016/j.acap.2021.12.013] [Medline: 34923145]
17. Rodgers DL, Needler M, Robinson A, et al. Artificial intelligence and the simulationists. Simul Healthc 2023 Dec 1;18(6):395-399. [doi: 10.1097/SIH.0000000000000747] [Medline: 37747487]
18. USC Standard Patient Studio. BreakAway Games. URL: https://www.breakawaygames.com/case-studies/usc-standard-patient-studio [accessed 2026-04-02]
19. Talbot TB, Kalisch N, Christoffersen K, Lucas G, Forbell E. Natural language understanding performance & use considerations in virtual medical encounters. Stud Health Technol Inform 2016;220:407-413. [Medline: 27046614]
20. Jake-Schoffman DE, Silfee VJ, Waring ME, et al. Methods for evaluating the content, usability, and efficacy of commercial mobile health apps. JMIR Mhealth Uhealth 2017 Dec 18;5(12):e190. [doi: 10.2196/mhealth.8758] [Medline: 29254914]
21. Real FJ, Meisman A, Rosen BL. Usability matters for virtual reality simulations teaching communication. Med Educ 2020 Nov;54(11):1067-1068. [doi: 10.1111/medu.14314] [Medline: 32914443]
22. Johnson WR, Artino ARJ, Durning SJ. Using the think aloud protocol in health professions education: an interview method for exploring thought processes: AMEE guide no. 151. Med Teach 2023 Sep;45(9):937-948. [doi: 10.1080/0142159X.2022.2155123] [Medline: 36534743]
23. Watkins DC. Rapid and rigorous qualitative data analysis: the "RADaR" technique for applied research. Int J Qual Methods 2017;16:1-9. [doi: 10.1177/1609406917712131]
24. Hanson JL, Balmer DF, Giardino AP. Qualitative research methods for medical educators. Acad Pediatr 2011;11(5):375-386. [doi: 10.1016/j.acap.2011.05.001] [Medline: 21783450]
25. Brooke J. SUS: a 'quick and dirty' usability scale. In: Usability Evaluation In Industry: CRC Press; 1996. URL: https://www.taylorfrancis.com/chapters/edit/10.1201/9781498710411-35/sus-quick-dirty-usability-scale-john-brooke [accessed 2026-04-02]
26. Vorderer P, Wirth W, Gouveia FR, et al. MEC Spatial Presence Questionnaire (MEC-SPQ): short documentation and instructions for application. : Project Presence; 2004.
27. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123. [doi: 10.5555/2835587.2835589]
28. Stokel-Walker C, Van Noorden R. What ChatGPT and generative AI mean for science. Nature 2023 Feb;614(7947):214-216. [doi: 10.1038/d41586-023-00340-6] [Medline: 36747115]
29. Shoja MM, Van de Ridder JM, Rajput V. The emerging role of generative artificial intelligence in medical education, research, and practice. Cureus 2023 Jun;15(6):e40883. [doi: 10.7759/cureus.40883] [Medline: 37492829]
30. Misra SM, Suresh S. Artificial intelligence and objective structured clinical examinations: using ChatGPT to revolutionize clinical skills assessment in medical education. J Med Educ Curric Dev 2024;11:23821205241263475. [doi: 10.1177/23821205241263475] [Medline: 39070287]
31. Guest G, Bunce A, Johnson L. How many interviews are enough?: an experiment with data saturation and variability. Field Methods 2006;18(1):59-82. [doi: 10.1177/1525822X05279903]
32. Guest G, Namey E, Chen M. A simple method to assess and report thematic saturation in qualitative research. PLoS One 2020;15(5):e0232076. [doi: 10.1371/journal.pone.0232076] [Medline: 32369511]
33. Sandars J, Lafferty N. Twelve Tips on usability testing to develop effective e-learning in medical education. Med Teach 2010;32(12):956-960. [doi: 10.3109/0142159X.2010.507709] [Medline: 21090948]
34. Liaw SY, Tan JZ, Lim S, et al. Artificial intelligence in virtual reality simulation for interprofessional communication training: mixed method study. Nurse Educ Today 2023 Mar;122:105718. [doi: 10.1016/j.nedt.2023.105718] [Medline: 36669304]

XSL•FO
RenderX

35. Kiger ME, Fowler L, Eviston M, et al. A case-based, longitudinal curriculum in pediatric behavioral and mental health. MedEdPORTAL 2024;20:11400. [doi: 10.15766/mep_2374-8265.11400] [Medline: 38686119]
36. Meyers N, Maletz B, Berger-Jenkins E, et al. Mental health in the medical home: a longitudinal curriculum for pediatric residents on behavioral and mental health care. MedEdPORTAL 2022;18:11270. [doi: 10.15766/mep_2374-8265.11270] [Medline: 35990196]
37. Zhang X, Wei X, Ou CX, Caron E, Zhu H, Xiong H. From human-AI confrontation to human-AI symbiosis in Society 5.0: transformation challenges and mechanisms. IT Prof 2022;24(3):43-51. [doi: 10.1109/MITP.2022.3175512]
38. Zhang X, Wang X, Tian F, Xu D, Fan L. Anticipating the antecedents of feedback-seeking behavior in digital environments: a socio-technical system perspective. Internet Res 2023 Mar 28;33(1):388-409. [doi: 10.1108/INTR-09-2021-0672]

Abbreviations

AI: artificial intelligence
BHAG: behavioral health anticipatory guidance
MEC-SPQ: Measurement, Effects, Conditions: Spatial Presence Questionnaire
RADAR: rigorous and accelerated data reduction
SPS: Standard Patient Studio
SUS: System Usability Scale
VR: virtual reality

Edited by A Kushniruk; submitted 16.May.2025; peer-reviewed by M Mansoor, X Zhang; revised version received 09.Jan.2026; accepted 28.Feb.2026; published 22.Apr.2026.

Please cite as:
Fleck L, Herbst R, Meisman A, Talbot T, Glisson M, Remington M, Goldson M, Real F
An Automated Curriculum to Support Behavioral Health Counseling Among Pediatric Residents: Usability Study
JMIR Hum Factors 2026;13:e77438
URL: https://humanfactors.jmir.org/2026/1/e77438
doi:10.2196/77438

© Liam Fleck, Rachel Herbst, Andrea Meisman, Thomas Talbot, Michael Glisson, Max Remington, Micah Goldson, Francis Real. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 22.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluating the Viability of Virtual Reality for Children's Food Choice Research: Comparative Mixed Methods Study

Deepti Aggarwal[1], PhD; Thuong Hoang[1], PhD; Sze-Yen Tan[2], PhD; Mohammadreza Mohebbi[3], PhD; Catherine G Russell[2], PhD

[1]School of Information Technology, Faculty of Science Engineering and Built Environment, Deakin University, Geelong, Australia

[2]School of Exercise and Nutrition Sciences, Faculty of Health, Deakin University, Geelong, Australia

[3]Faculty of Health, Biostatistics Unit, Deakin University, Geelong, Australia

Corresponding Author:
Deepti Aggarwal, PhD
School of Information Technology, Faculty of Science Engineering and Built Environment, Deakin University, Geelong, Australia

Abstract

Background: Virtual reality (VR) systems offer promising potential as a controlled platform to investigate human behaviors specifically related to food choices. Yet, little is known about its viability for conducting food choice studies with children, thus limiting the development of public health research.

Objective: This study aimed to investigate the viability of VR technology for understanding children's food choices, focusing specifically on perceptual differences between VR and real-life (RL) settings. We examined how children perceive and interact with food portion sizes and container sizes in VR as compared with an equivalent RL scenario.

Methods: A within-subject, mixed methods study was conducted with 437 children aged 5 - 12 years at a science museum. Participants engaged in a standardized food selection task for a simulated breakfast scenario, choosing portions of cereal and milk in 2 conditions: a head-mounted VR environment and a corresponding RL physical setup. Children's food selection behaviors were quantitatively compared across 3 independent variables: condition (VR vs RL), food healthiness (healthy vs unhealthy options), and container size (small, medium, and large). Qualitative and quantitative data were collected via postsession questionnaires assessing presence, embodiment, and simulator sickness, alongside detailed interaction logs from the VR environment. Data analysis used statistical comparisons and thematic analysis.

Results: The findings revealed both behavioral consistency and significant perceptual differences between the VR and RL conditions. A behavioral similarity was identified, as children served significantly larger portions of unhealthy food compared with healthy food in both conditions. Crucially, a difference was observed in size perception: children struggled to accurately match the size of bowls and glasses between the VR and RL conditions. Furthermore, while children reported low feelings of presence and embodiment within the VR scenario, they demonstrated a high degree of control and engagement in the virtual task.

Conclusions: Our findings suggest that current state-of-the-art VR technology presents limitations in its viability for conducting food choice studies with children, particularly concerning accurate size and volume perception. Based on the findings, we provide 4 practical recommendations to guide the future development of immersive food environments, thereby supporting more reliable and ecologically valid food choice research with young populations.

(JMIR Hum Factors 2026;13:e88476) doi:10.2196/88476

KEYWORDS

virtual reality; portion size; food choices; children; embodiment; immersion

Introduction

Background

Within the field of human-computer interaction (HCI), there is a growing interest in designing for and around food practices [1-3], underlined by research in dietary-monitoring systems [4,5], technologies for supporting rich conversations during mealtimes [6,7], and awareness about our food ecosystems such as food waste and carbon emission [8]. Recent works further blend food research with interactive experiences, such as combining real-life (RL) eating with virtual reality (VR) gaming experiences [9], encouraging healthy eating practices through tangible systems [10,11], and creating novel food experiences using emerging technologies such as 3D printers [12]. Recently, Covaci and colleagues [13] presented multiple design illustrations, through several co-design workshops, speculating the design of VR-based food experiences for the everyday context. As such, existing research collectively suggests that it

XSL•FO
RenderX

is feasible to enhance eating experiences with interactive technologies for different contexts. In this work, we take a step further and explore the viability of using VR technology as a tool to conduct food choice studies. Unlike existing works that focus on the adult population, we investigate the use of VR technology for younger children (aged 5 - 12 years).

On the other hand, existing literature in the nutrition science domain suggests the viability of using VR systems to conduct food choice studies, due to its capability to simulate realistic environments and food-related cues [14-18]. For instance, studies with adults using VR have highlighted key insights into their personal eating practices and the ways they feed their children [14,15,19]. There is, however, a notable gap in understanding the firsthand experience of children about their food choices. This is a crucial area to explore, as a rich understanding of how children make food choices can guide the development of effective programs that promote positive eating practices from an early age [20,21]. Given the rising rates of childhood obesity in many countries [22,23] and the serious health risks it poses later in life [20,22], this research addresses a timely and important topic.

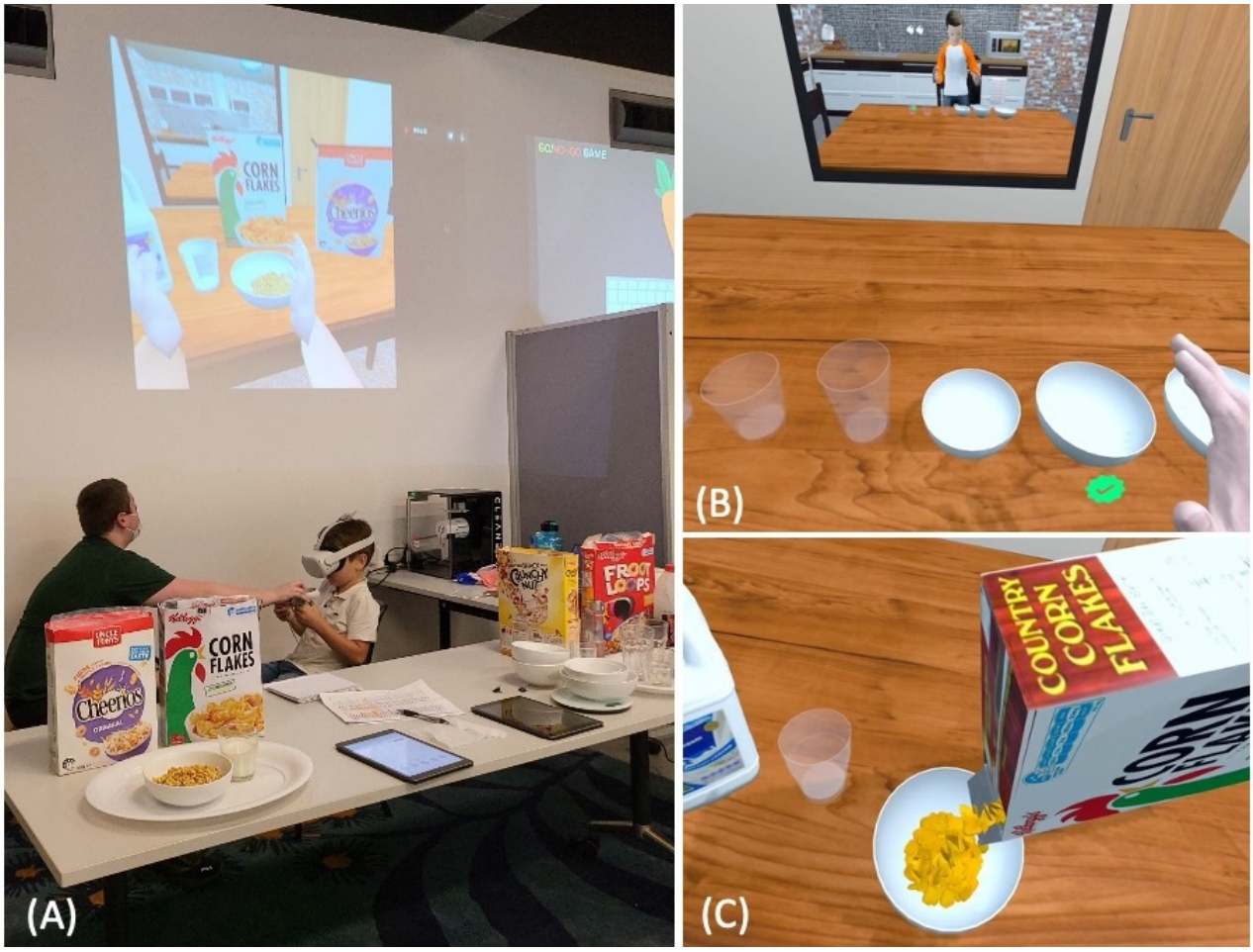
Engaging children in research can, however, be challenging, as the scientific tasks may seem repetitive or boring to them [24]. Researchers have emphasized the need to employ developmentally appropriate procedures and measures for robust data collection [25] as well as to modernize study procedures to increase participation of children in research [26,27]. VR systems offer an engaging environment for children to learn and explore, and their popularity among children and teenagers is increasing [28,29]. A recent survey suggests that 1 in 5 families with children younger than 17 years has a VR headset [30] and that around 70% of 2 - to 15-year-old children in the United States and United Kingdom reported being fairly or extremely interested in VR [29]. Previous works have investigated the potential of VR to help children deal with painful clinical procedures [31,32], as well as to teach them a wide variety of topics in school settings such as learning to cook [33], Swedish language [34], and science simulations [35]. This research further extends the applications of VR with children by investigating how children interact with simulated food environments in VR and how their interactions differ from RL settings.

This paper investigates the viability of using VR technology to understand children's food choices. We investigated how children perceive food portion sizes and container sizes in VR and how these perceptions differ from RL scenarios. We specifically focused on only 2 tasks—food portioning (ie, how much can the individual serve themselves) and container selection (ie, which container the individual should select)—as these are the basic tasks required to conduct food choice studies. We conducted a controlled trial with 437 children aged 5-12 years at a science museum (as shown in Figure 1). Not all participants completed the same set of tasks; hence, the sample size for each research question (RQ) varied. Employing a within-subjects study design and mixed methods data collection approach, we present a comparative account of participants' interactions in a virtual food environment (ie, a virtual kitchen) with an RL museum setting.

This work makes the following contributions: First, our study investigates the firsthand experience of children in making food decisions related to food portion sizes and container sizes. Our findings showed that children served larger portions of unhealthy food than healthy food in both the VR and RL conditions. Second, we discovered that children struggled to accurately match the size of bowls and glasses between the VR and RL conditions. Both findings raise concerns on the viability of using VR technology for food choice studies. Moreover, our study also revealed that children found the VR experience so engaging and fun that it could potentially compromise the scientific rigor needed for data collection. In response, we present 4 recommendations to guide future development of playful VR experiences for children that can still maintain high scientific standards and data quality. Finally, this work extends HCI practices into nutrition science by exploring a novel application of VR technology as a research tool for conducting food choice studies. By introducing methods for testing concepts such as food portion sizes, container size mapping, and health factors related to food items, we have opened new avenues for HCI researchers and nutrition scientists to create robust and impactful VR systems in the future.

XSL•FO
RenderX

Figure 1. A snapshot of the virtual reality (VR) study booth at the science museum. (A) Children were tasked to make food choice decisions for breakfast items in both VR and real-life scenarios. (B) Participant's viewpoint in VR during bowl size selection task. (C) Participant's viewpoint while pouring cereal into a virtual bowl.



Prior Work

We discuss the related literature around 3 topics: VR as a tool for food research, VR as an engagement tool for children, and children's embodiment in VR.

VR as a Tool for Food Research

Researchers in health and nutrition are just beginning to explore the use of VR technology to study eating habits and food choices [14,16]. This field is still new [36], but initial studies have already shown that VR can create realistic food environments that produce convergent results with RL settings [14,36-38]. In food research, participants often perform the same tasks in VR and real-world settings, allowing researchers to directly compare their choices. These studies use a range of virtual environments, such as virtual buffets [14,15], cafeterias or food courts [39,40], and supermarkets [17,37,38,41], to offer participants a wide range of food options. On the other hand, VR systems also allowed researchers to easily change factors such as food color and texture, food venues (bar or restaurant), and social settings, as per the study requirements. It also provided a way to collect objective data related to food selection such as the order in which items are chosen, portion sizes of different food items, and the total weight of the selected food.

As such, prior research suggests the viability of using VR to conduct food studies [14-18]. Researchers have argued that the realism offered by VR technology addresses the existing issues with traditional measures to record personal food choices [25,42,43]. For instance, assessing food choices through real food is challenging due to the associated high cost of buying real food while also managing food safety (such as food handling) and food preferences (eg, allergies, vegan or vegetarian, etc). Limiting food choices to overcome the cost issues is not a viable solution as it can impact the generalizability of the study findings [43]. On the other hand, conducting studies with fake foods is another potential solution, which allows researchers to include a wide variety of food items for greater experimental control. However, fake foods lack in simulating realistic environments that significantly influence the food choices, hence limiting generalizability and ecological validity of the study findings [42].

To understand the food practices of children, several studies have employed proxy measures and investigated parents' feeding practices [14-16]. Persky and colleagues [15] compared parents' pasta selection for their children in VR versus RL conditions and reported a high correlation. Parents in the study also felt that the VR environment was realistic and that their virtual food choices accurately reflected their RL habits, both in terms of portion size and healthiness. Similarly, another study [44,45] used a VR buffet to raise awareness about childhood obesity among mothers whose children were genetically vulnerable to

XSL•FO
RenderX

being obese in childhood and later. The study was conducted with overweight and obese mothers, where they were required to select food they believed their child would eat through pointing-based interaction in VR. The researchers observed the mothers' decisions and measured the portion sizes they chose and accordingly educated them on appropriate food portion sizes for their children.

There is a limited understanding of the firsthand experiences of children's food selection in a virtual food environment, which this research aims to address. Our work builds upon the recent HCI research [13] that has explored the potential applications of VR for food-related environments. The authors described 18 themes to develop "food metaverse" that can help people connect with food and adopt sustainable practices. To design similar food metaverse experiences for children, it is important to understand their perceptions of food portion sizes and container size mapping, as these are fundamental tasks in any food choice study. Furthermore, by comparing how children perceive and interact with food in both VR and RL, we can determine whether VR is a valid tool for studying this age group. Our goal is to address these gaps by providing a detailed comparison of children's interactions in a virtual food environment versus an RL setting, focusing on portion sizes and container size estimation.

VR as an Engagement Tool for Children

VR has mainly been used as an educational tool to engage children in topics that are otherwise complex to teach in real-world settings. For instance, Gorman and colleagues [33] used VR with 360-degree images and videos to teach middle schoolers about food hygiene and safe cooking practices. The VR setup successfully overcame challenges related to the shortage of cooking facilities and teaching staff and created an engaging classroom where children were highly motivated to learn. In another study, researchers [46] designed a virtual kitchen where 7‐ to 9-year-old children could prepare their breakfast and get relevant feedback on their food choices. The study revealed that VR offered a promising environment, as children enjoyed learning about the nutritional value of various breakfast items. Another work [45] used a VR supermarket to teach 6- to 13-year-olds about the environmental impact of their food choices. The study found that all participants could recall the information presented in text and images format, but only those children older than 10 years were able to fully understand the messages. Interestingly, those who understood the information reported feeling sorry about the negative environmental impact of their food habits and expressed a desire to change their habits for the better. We extend the applications of VR for children beyond education and investigate their interactions with food and containers in a food choice study.

Children's Interactions in VR

A significant amount of works have explored how adults and older adults embody themselves in VR environments for different contexts, such as for physiotherapy training [47], combating social isolation in aged care facilities [48], and for gaming purposes [9]. However, there is a notable gap in understanding how children interact and embody themselves in VR (as also emphasized by a recent review paper) [49]. Among the few studies, Keenaghan and colleagues [50] compared how adults and 5-year-old children perceive different-sized avatars and objects in VR. Participants were asked to estimate sizes of their body and objects in a virtual "tea party" environment. The virtual avatar was manipulated relative to the participant's actual height. The study revealed that the reported degrees of embodiment across adults and children were comparable. Participants' perception of the virtual objects was influenced by their virtual body size, that is, they felt that the virtual objects were bigger when they embodied a smaller body and vice versa in a larger body. The notable finding was that unlike adults, children also felt that their (actual) body size increased in the large body condition; however, they did not report getting smaller in the small body size condition. The authors hypothesized that since children are more familiar with their growing body, they are more likely to notice a change in body size than adults. Also, children mainly reported feeling bigger and not smaller because shrinking of bodies is not a natural phenomenon. Previous research also suggests that children are more sensitive to their own-body representation as they must keep track of a constantly changing and growing body along with other sensory developments [51].

Weijs and colleagues [52] similarly compared a full-body embodiment of adults and 8- to 12-year-old children. The authors investigated participants' responses to synchrony of the avatar's movements (ie, avatar moving with the same or different speed as that of the participants) and appearance of the avatar as human and skeleton. Their study highlighted that both children and adults showed highest ownership for the virtual avatar for the synchronous condition, and that the ownership reduced with increased asynchrony. Also, higher ownership and agency were reported for human avatars than the skeleton avatars by children and adults. The notable difference was that children demonstrated a stronger sense of agency than adults, both with increasing temporal mismatches and with skeleton avatars. The difference in agency (ie, the sense of being in control) suggests the developmental differences in children and adults. Prior research [53] revealed that children aged up to 8 years move from relying solely on visual information to integrating multiple senses to develop a better sense of their own bodies.

In a recent study [54], researchers compared children's perception of human, animal, and anthropomorphized creatures in VR and revealed that the animal avatars received the most positive descriptions from children aged 5-9 years. Since animal creatures are farther away from human-likeness, realistic behaviors were not expected from them. On the other hand, human and anthropomorphized creatures were expected to have highly realistic behaviors. The virtual character's movements, blinking of eyes, and facial features such as head size or hair color were found strange and unreal by children. Similarly, another study [55] investigated the perception of 6- to 8-year-old children related to presence and realism in VR. Through a qualitative study of the VR experience where children swam in a pool with dinosaurs, the study highlighted how VR challenged children's understanding of reality when dinosaurs were treated as real. Children also tested the reality of VR by doing certain activities such as holding breaths and comparing the number of

XSL•FO
RenderX

fish the dinosaurs consumed in VR as opposed to their actual diet. While some children verbally described the VR experience as fake, they still reported physiological and emotional responses to VR exposure. The study concluded that although children knew that dinosaurs were extinct, VR was still able to create alternate perceptions of what is real for them. We take inspiration from these works and further extend the literature on VR by investigating children's perceptions of food and containers in a food choice study. Next, we describe our study design.

Methods

Ethical Considerations

The project was approved by the Deakin University ethics committee (ethics number 2019 - 393), including approval for the use of commercial VR headsets with children younger than the minimum recommended age of 13 years. Consent from parents and assent from children were obtained prior to the study. Once consent was obtained, parents provided the demographic details of the participants at the entry point through a questionnaire. The study was conducted exclusively with child participants. Upon recruitment at the study's entry point, each participant was assigned a unique alphanumeric ID code. This ID was utilized across all data collection platforms—including VR system logs and Qualtrics survey responses—to ensure seamless data integration. All subsequent statistical analyses were conducted using these IDs to maintain participant confidentiality.

Study Design

Our study was a part of a larger children's eating and weight project, which was conducted at a science museum in a metropolitan city in Australia. The museum is popular for featuring interactive installations and exhibitions related to science and technology and aims to expose children to innovation and its societal impact. The project was set up with multiple activities within a dedicated room within the museum, with a single entry and exit point. Our study was one of the activities within the room. Interested families approached the room at the museum, where banners were placed to invite participants. The volunteers introduced the families to the project and the associated activities. All participants, including children and parents, were recruited at the entrance. Parents were informed that VR activity may cause motion sickness, so children with motion sickness history should not participate. However, no parent mentioned any motion sickness history for their children.

To mitigate the risk for this age group, we ensured that the participants used the VR headset in a seated position. Additionally, to help prevent simulation sickness in children, we limited the duration of the VR experience to less than 5 minutes. We used Oculus Quest 2, as the study occurred before the rebranding of Oculus Quest 2 to Meta Quest 2. The study was conducted over 5 consecutive weekends (Saturdays and Sundays) for a total of 10 days over a school holiday period. In total, 437 children aged 5 - 12 (mean 8.2, SD 2.0) years participated in the study. Out of the 437 participants, 175 were female and 190 were male. Gender information for the remaining 72 participants was unavailable, which was attributed to either unstable internet connection at the museum or incomplete demographic information provided by the parents. Since the study was conducted at the museum, we had no control over the gender distribution and allowed all eligible participants to participate in the study. Participants did not receive any monetary compensation for their participation. We also noted that children younger than 16 years had free access to the museum; however, the accompanying parents were required to buy a ticket.

Research Questions

Building upon existing works that highlight the potential of VR technology for conducting food choice research [15,18,46], we designed a VR experience to investigate children's interactions in the virtual food environment. The main RQ we explored in our study was *what factors affect the use of VR as a medium for food choice research with children?* We compared the participants' interactions within a simulated food environment in VR with that of RL setting at the museum.

The main RQ is broken down into different aspects.

1. RQ1: How does VR influence the measurement of key variables in food choice studies (food portion size, container size, and additional characteristics of healthy and unhealthy)?
2. RQ2: How does the children's experience (presence, embodiment, simulator sickness, and repeatability) of being in VR affect its usage as a medium for food choice research?
3. RQ3: How do children's demographic characteristics (gender and age) affect the VR usage as a medium for food choice research?

Study Parameters

The study was designed around the everyday activity of breakfast, as this is when children are most likely to serve themselves food. Participants were asked to imagine that it was breakfast time, and they needed to serve themselves cereal and milk. The 2 main conditions for the study were VR and RL. The setting for the VR condition was a virtual kitchen environment, and the setting for the RL condition was the museum. The independent variables were (1) food healthiness with 2 levels: healthy and unhealthy, and (2) container size with 3 levels: small, medium, and large. Table 1 provides a summary of the study design. These 2 independent variables were applied to 2 food options: cereals and milk. Figure 2 provides a snapshot of the study station.

XSL•FO
RenderX

Table . Summary of our 2×2×3 study design. Food options included breakfast cereal and milk, and container options included bowl and milk.

Independent variable	Levels
Condition	Real life and virtual reality
Food healthiness	Healthy and unhealthy
Container size	Small, medium, and large

Figure 2. A snapshot of the study station for the real-life condition: 3 sizes of bowls and glasses (small, medium, and large), a large plate to prevent milk spillage, a tablet to fill out the questionnaire, and a scale to weigh the results afterward.



Our choice of cereal and milk was guided by their health star rating and sugar content information available on the package: health star rating is a standard method to define the healthiness of a product in our country [56,57]. A large amount of sugar is not desirable in young children, as sugar intake has been associated with potentially adverse consequences, such as dental caries, obesity, and coronary heart diseases, in the long run [58]. Two healthy cereals in the study included Uncle Toby's Cheerios (Health Star Rating of 4 [59] and sugar content as 14.6 g/100 g) and Kellogg's corn flakes (Health Star Rating of 3.5 and sugar content as 10 g/100 g). Two unhealthy cereals included Kellogg's Froot Loops (Health Star Rating of 2 and sugar content as 38.8 g/100 g) and Kellogg's crunchy nuts (Health Star Rating of 2 and sugar content as 35 g/100 g).

For cereals, we also maintained the consistency in shape to avoid any other potential influence on the findings [60]. For example, using dinosaur-shaped cereal can influence children's choice as opposed to using flat-shaped cereals (ibid.). Hence, we selected breakfast cereals of similar shapes for both healthy and unhealthy options. For instance, Uncle Toby's Cheerios (healthy option) and Kellogg's Froot Loops (unhealthy option) both are ring-shaped. Similarly, Kellogg's corn flakes (healthy option) and Kellogg's crunchy nuts (unhealthy option) both are flat.

XSL•FO
RenderX

On the other hand, options for milk included plain milk as a healthy option (with sugar content as 4 g/100 g) and chocolate milk as unhealthy (with sugar content as 10.6 g/100 g). For cereal bowls and milk glasses as containers, we purchased crockery sets (bowls and glasses) with 3 sizes: small, medium, and large from the local department store. Participants were instructed not to consume any food during the RL trial to prevent potential risks related to food allergies and food hygiene.

In the study, we did not ask nor expect participants to distinguish between healthy and unhealthy foods. Instead, the research team randomly assigned children with different food options. The healthiness variable was used mainly to investigate food portion size in RQ1. Consequently, no visual labels were added on the boxes to categorize them as healthy or unhealthy. In previous research, factors such as taste, health, and social settings have all been shown to affect how much food people serve themselves [3,61]. While we did not investigate taste or social influence in our study, we specifically investigated how the health factor impacts food portion size. Some studies suggest that foods with high sugar content are often perceived as tastier than those with less sugar [62], thus making the health-taste-portion size connection relevant for our study.

In the RL condition, products were replenished as needed to maintain consistency across participants. To preserve the structural integrity of the cereal, items were never handled directly; instead, they were poured into and out of the containers, preventing physical breakdown. In contrast, milk was replenished on a daily basis due to its shorter shelf life.

Due to the challenges of letting children handle real food in a public museum, our study adopted a different approach for the RL and VR conditions (more details under “Methodological Challenges” section). In the VR condition, children served the food themselves. In the RL condition, however, a researcher served the food based on the child’s verbal instructions. We made this choice to maintain a level of agency for the child in both conditions. This difference, particularly in how the stop pouring” action was executed (verbally in RL vs by hand movements in VR), does create a confounding variable. However, we expect that it will not impact our results, as RQ1 aims to investigate children’s decision-making process about food portion sizes as a cognitive task and not as a motor skill task of serving food. Therefore, we expect the final portion sizes to be comparable in both conditions.

VR Experience

We designed a virtual environment of a typical home kitchen with a range of food products, crockery, and cutlery along with other typical everyday objects (eg, vase and microwave) to create a realistic environment (Figure 1 provides a summary of the study). Participants appeared as an avatar seated at a kitchen table, with virtual cereal boxes and milk cans within their reach. We used the same cereal and milk brands for both the VR and RL conditions. A virtual bowl and glass were placed in front of the participants on the kitchen table. Participants were required to press the controller buttons to grab an object (bowl, glass, cereal box, and milk can) from the table and tilt the controller to pour cereals and milk into the containers (Figure 1C). We placed a virtual mirror in front of the participant, on the other side of the kitchen table, to induce ownership of the self-avatars [63] (Figure 1B). Using Unity meter units, virtual bowls and glasses were matched with the dimensions of store-bought physical containers for all sizes: small, medium, and large (default scale of 1 unit equals 1 meter).

To calculate the cereal weight in VR, we applied a particle system that provided an estimate weight of each cereal type. We measured the average weight of cereal for each of the 4 types, performed a sample test by measuring the real weight of full and half-full bowls for all sizes, and finally calibrated the measurement values in VR. To calculate the volume of milk, we applied the rate of pouring based on the tilt angle of the milk container to increase the visual content within the virtual glass. We measured the capacity of the physical glasses for full and half-filled and adjusted the visualization of the virtual glasses to match accordingly. The weight of the RL containers (bowls and glasses) was measured prior to the study.

Study Procedure

Before starting the study, we conducted a pilot study with 10 children (4 boys and 6 girls; mean age 8.1, SD 1.9 years) recruited through the researchers’ network. The pilot study was conducted 4 weeks before the final study in a VR laboratory at the university campus to check whether the study tasks were age appropriate and whether the children were able to finish the entire study within 10 minutes. Given that the study took place in a crowded science museum, it was essential for each child to complete the tasks quickly to manage the flow of participants within the large room with multiple study stations. The pilot study confirmed the risk of food spillage as some of the younger pilot participants could not comfortably handle the milk can. Thus, we adopted the procedure for a researcher pouring the milk in the RL condition instead of the children. On the other hand, the data collection questionnaire was revised to minimize the time spent at the VR station. Specifically, we simplified the Likert scales by reducing the number of response categories and reduced the number of questions to ensure that they were developmentally appropriate and easily interpretable for children aged 5 - 12 years (more about questionnaires is discussed under “Data Collection” section).

We employed a within-subjects study design, where all participants were asked to complete the main task of making the breakfast cereal in 2 conditions: VR and RL. For the first condition, participants were given the following instructions:

> *Imagine that you are in your kitchen. It’s a normal school day, and you are preparing breakfast for yourself before you head to school. You will not get to eat again for another 3 hours or so, so prepare yourself a sufficient breakfast. Serve yourself a breakfast cereal and pour yourself a glass of milk to drink.*

Participants were asked to pour milk and cereal into separate containers, so that we could accurately measure the portion sizes of each item individually. For the second condition, participants received the same instruction to serve the portion size that they would like to eat or drink, with the assumption that it would not have changed from the first condition. The first condition and

XSL•FO
RenderX

the second condition were alternated between VR first–RL second and RL first–VR second.

For the RL condition, participants were asked to sit on the provided chair, accompanied by a researcher. Each participant was assigned to either the healthy or unhealthy food option. The order was alternated to ensure a balanced number of participants in both groups. A bowl and a glass were placed directly in front of them. Participants did not handle the food directly in RL; rather, the researcher poured the food, following their instructions. The researcher first provided the task instructions to the participants and then picked up the selected cereal and started pouring it into the bowl. Participants were asked to say "Stop" when the desired amount was reached. A similar process was repeated for the drink. The researcher slowly poured the food and constantly checked with the children whether the amount was appropriate. After that, the bowl and glass with the food were weighed on a table scale and the data were recorded in Qualtrics. The actual weight of the corresponding container was later subtracted to retain the weight of the poured food content. Milk was measured by weight rather than volume for consistency. All food contents were poured back into their respective containers for reuse. Figure 2 shows the arrangement of the study station for the RL condition.

For the VR condition, participants were first asked to sit comfortably in the provided chair. A researcher assisted the participants in wearing the VR headset (Oculus Quest 2) and 2 handheld controllers. The interpupillary distance of the headset was adjusted for each participant for the best visual clarity, but the adjustment values were not recorded as data. The researcher ensured that the participant remained seated throughout the session for safety purposes, with a protocol in place to stop the session immediately if a child experienced any VR-related discomfort. However, no adverse effects were observed, and all participants successfully finished the session. The researcher explained to the participants how to operate the handheld controllers and interact with the virtual environment. A practice task was completed before starting the study tasks to avoid novelty. In the virtual kitchen, participants were represented as a virtual avatar seated against a table, with gender and skin tone matched according to the researcher's observations. The researcher confirmed whether the participant agreed to the chosen gender and skin tone of the virtual avatar before starting the study procedures.

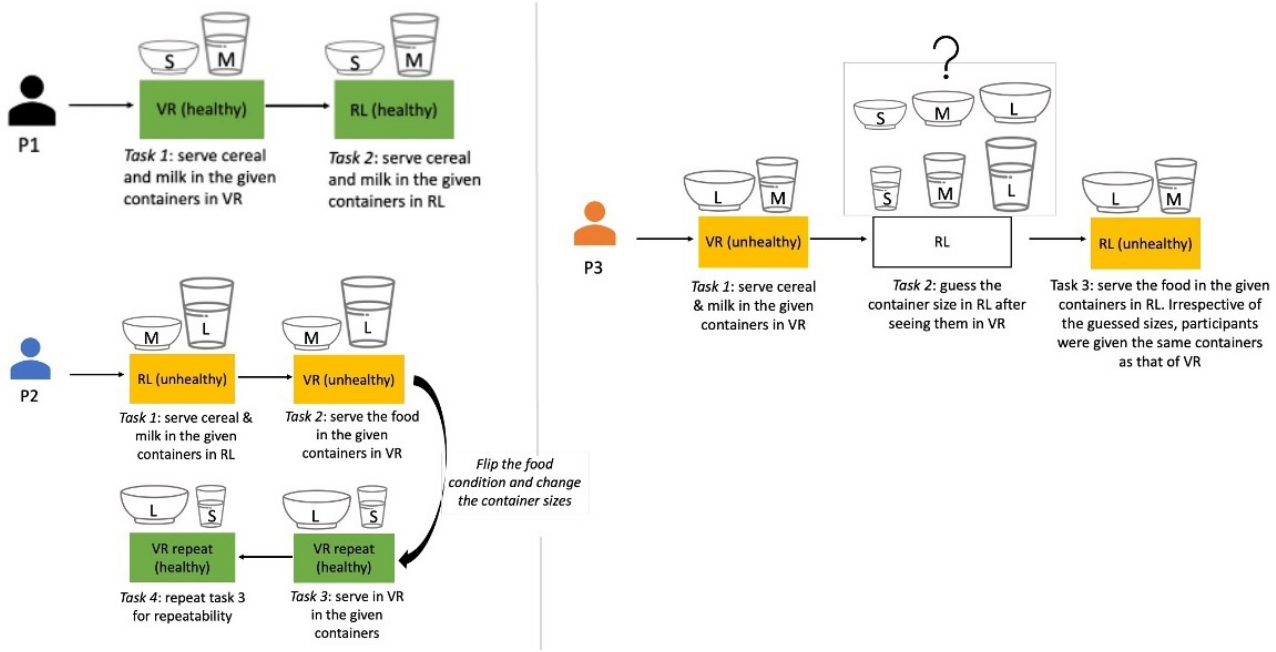
In the virtual kitchen, participants were asked to wave at themselves (virtual avatar) in the mirror and explore how their movements matched the avatar before beginning with the study tasks. The participants then proceeded to grab the desired cereal box and pour it into the virtual bowl placed in front of them to the desired amount or to match with the previous RL condition. The participants then completed the task of pouring milk (plain or chocolate) into the virtual glass (refer P1 case in Figure 3). Figure 1B and C show the point of view of a participant for the bowl size selection task and cereal-serving task.

A subset of participants was asked to perform additional tasks related to RQ2 (repeatability) and RQ1 (container size estimation). For instance (refer to P2 case in Figure 3), within the same VR session, participants were first instructed to serve the food in RL and VR. Then they were instructed to serve again by imagining that they were serving breakfast for a subsequent day—with the goal of matching the exact portion of the food items as performed in steps 1 and 2. This task contributed data collection for RQ2 (repeatability). To avoid learning effects, they were presented with a different set of cereal and milk, and container sizes.

To answer RQ1 about container size estimation (refer P3 case in Figure 3), participants in the first condition were presented with a specific-sized bowl and glass. In the second condition (VR or RL), they were presented with 3 different sizes of bowls and glasses and were asked to identify the sizes they had seen in the first condition. "VR first" refers to the condition when participants were initially shown the bowl or glass in VR condition and then were asked to guess the container size in RL condition and vice versa for "RL first." The sizes selected by participants were recorded. Regardless of their answer, the correct-sized bowl and glass were presented to them to help them complete the tasks for the second condition. In total, participants spent less than 5 minutes in VR and less than 10 minutes in total.

XSL•FO
RenderX

Figure 3. Sample order of tasks for 3 participants completing VR and RL conditions. P1 completed only the main task related to RQ1 portion size for healthy condition. P2 first performed the portion size task and then completed additional tasks to serve the same amount of food in VR for tasks 3 and 4 for repeatability testing (RQ2). P3 was asked to guess the glass and bowl sizes (RQ1) before performing the portion size task (RQ1). Sizes of bowls and glasses were randomized and counterbalanced. RL: real-life; L: large; M: medium; RQ: research question; S: small; VR: virtual reality.



Data collection procedures varied across RQs to accommodate the high volume of museum visitors. While the P1 case was conducted during peak hours, the P2 and P3 cases (Figure 3) were restricted to off-peak periods (such as late afternoons) as they required additional time and physical setup. This scheduling was essential to manage the long queues for the popular VR exhibit and ensure a steady participant flow. Consequently, these variations in study design resulted in distinct participant pools for each RQ, as detailed in Table 2.

XSL•FO
RenderX

Table . Summary of data points for each research question (RQ).

Research questions and subtasks	Participants, n	Final data points after cleaning
RQ 1		
Food portion size	317	• 317 data points for cereal (both VR[a] and RL[b] conditions) • 305 data points for milk (both VR and RL conditions)
Food healthiness	284	• 284 data points for cereal in RL condition • 264 data points for cereal in VR condition • 276 data points for milk in both VR and RL conditions
Container size	195	• 195 data points for bowl size guesses (72 RL first and 123 VR first) • 193 data points for glass size guesses (70 RL first and 123 VR first)
RQ 2		
Presence, embodiment, and simulator sickness	411	• 411 total responses: 240 responses from age group of 5-8 years, and 171 responses from age group of 9 - 12 years.
Repeatability	140	• 134 data points for cereal and 140 data points for milk—all data points were only in the VR conditions
RQ 3		
Demographics	420	• 180 male and 169 female participants; gender data not recorded for 71 participants • 245 participants in age group of 5 - 8 years; 175 participants in age group of 9 - 12 years • 195 data points for glass size estimation; 193 data points for bowl size estimation

[a]VR: virtual reality.

[b]RL: real-life.

Methodological Challenges

Our study procedures were guided by the challenges of conducting a controlled study in a real-world setting with children. Challenges of testing prototypes in real-world settings and the need to adapt study design have been highlighted for field studies [64], which also guided our study design. First, we aimed at reducing the VR exposure time for children to manage potential simulation sickness. Hence, we developed short VR activities for children and simpler and shorter questionnaires for data collection.

Second, to ensure a smooth and safe experience in the busy science museum, we made a few adjustments to the study tasks for RL and VR conditions that were guided by the pilot testing. Unlike in VR, participants in the RL setting were not allowed to handle the food directly. This approach not only safeguarded our electronic equipment from food spillage but also aligns with how parents typically serve food to young children to avoid spillage. By having a researcher serve the food, we were also able to mitigate risks related to food allergies and hygiene.

Finally, due to the different affordances of virtual and real worlds, we used a mirror in VR as it is a common strategy [63] to induce ownership of a virtual body. Participants can connect to the virtual avatar by seeing their mirror reflection and associating their bodily movements with the avatar. However, in RL, it was not natural to see yourself in a mirror in a public setting (ie, amidst the crowd) when participants could directly see their body movements without any interruption. Hence, a mirror was not included in RL.

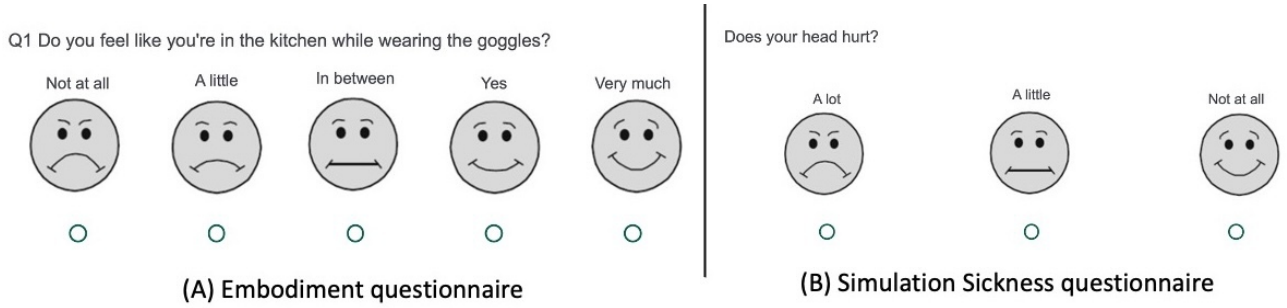
Data Collection

A researcher carried a tablet with Qualtrics questionnaires for data collection. Demographics information was noted based on the researcher's observations of the participant's gender, skin tone, and hair color, which were used to match the VR avatar. Participants' demographic details (age and gender) were collected from their parents during recruitment. After completing the study tasks, participants sat down with a researcher to complete a questionnaire with 17 questions structured in the following way: (1) 1 question on previous experience with VR; (2) 8 combined presence and embodiment questions, including 3 questions (Q1-Q3) about presence adopted from the iGroup Presence Questionnaire [65] and 5 questions (Q4-Q8) about embodiment adopted from Gonzalez-Franco [66]; (3) 7 questions on discomfort felt in completing the activity through an adopted

XSL•FO
RenderX

sickness questionnaire [67,68]; and finally, (4) 1 question on subjective experience with VR. We combined the presence and embodiment questionnaires and reduced the number of questions to simplify data collection from our participants and to keep the study length manageable for children. The researcher read each question to the participants while showing the tablet with the visual Likert scales, as shown in Figure 4, and asked the participants to touch the desired option. The researcher also asked the participant to verbally explain their reasoning wherever relevant, which were noted down as a response to the last question. Researchers' observations were also noted through a separate Qualtrics form.

Figure 4. Snapshots of visual scales for (A) embodiment questionnaire and (B) Simulator Sickness questionnaire.



Presence, embodiment, and sickness questions followed a visual scale as shown in Figure 4A. For the presence and embodiment questionnaire, the Likert scale was coded as not at all=1, a little=2, in between=3, yes=4, and very much=5, while the Likert scale for the Simulator Sickness questionnaire was coded as not at all=1, a little=2, and a lot=3. The Likert scale for presence and embodiment was reduced from a 7-point scale to a 5-point scale, while the sickness questionnaire was reduced from a 4-point scale to a 3-point scale. These changes were guided by our pilot testing with children, where children naturally used the descriptors—not at all, a little, and a lot—to respond to presence-, embodiment-, and sickness-related questions. These changes are aligned with a previous study [69], which emphasizes the importance of using simple and visual scales when designing questionnaires for children.

Study variables related to the virtual kitchen were recorded in the VR headset as a csv file. The csv files were transferred from the headset to a safe storage at the end of the day. The csv included data related to study condition (VR or RL), food type (healthy or unhealthy), given cereal type (Froot Loops and crunchy nuts as unhealthy cereals; Cheerios and cornflakes as healthy cereals), given milk type (chocolate milk as unhealthy option and plain milk as healthy option), and size of the given bowl or glass (small, medium, and large). These data also included the cereal weight (in g) poured in the virtual bowl, milk weight (in g) poured in the virtual glass, guess made by the participant about the container size (as small, medium, and large), and time taken to pour cereal or milk. For the RL condition, the researcher manually maintained an Excel sheet to record participants' selection such as cereal weight poured in the bowl, milk weight poured in the glass, and guess made by the participant about the container size.

Data Analysis

To answer our RQs, we conducted statistical tests using STATA (version 17; StataCorp LLC) on the data collected through the VR headset. We conducted chi-square tests, correlations, 2-tailed *t* tests, and kappa analysis to compare the level of agreement between study parameters. Kappa statistics [70] is more suitable for assessing agreement when multiple categorical factors are involved. Agreement cannot be measured through a chi-square test or a regression test, as they can only suggest correlation between the outcome and independent variables. All significant results are reported in the "Results" section. Descriptive analysis was conducted to analyze presence, embodiment, and simulator sickness collected through Qualtrics. Finally, we conducted deductive thematic analysis [71] on the subjective data collected through the researcher's notes in Qualtrics, and participants' responses on the last question in the poststudy questionnaire. Thematic analysis was conducted to mainly complement the findings gathered from the quantitative analysis, as per the mixed methods approach [72]. Initially, the first author (DA) created a separate Excel workbook with participants' quotes and researchers' notes, which had a total of 63 entries. She then coded the 63 entries deductively around the RQs, with code names as kitchen characteristics, avatar characteristics, presence, sickness, VR experience, VR issues, study tasks, and other issues. After finalizing the quantitative analysis, the first (DA) and second authors (TH) worked together to select the best representing quotes for the generated quantitative insights.

Table 2 presents a summary of the data points that guided all 4 RQs. Data were cleaned by excluding incomplete records and identifying outliers based on physical plausibility. Data points were discarded if the poured volume exceeded the maximum capacity of the assigned container (small, medium, or large). These logical bounds ensured that the final dataset included only realistic serving behaviors. To establish a baseline for outlier detection, the maximum capacity for each of the 3 container sizes (small, medium, and large) was determined by measuring the physical weight of the bowls and glasses when filled to capacity with cereal and milk. Consequently, the total number of valid data points (column 3 in Table 2) differs from the initial number of recruited participants. Because each child performed tasks in both VR and RL environments, they contributed multiple data points to the final dataset. In total, 437 participants contributed data to this study, with each participant contributing to 1 or multiple research data points. Hence, Table 2 distinguishes between the participant count and the valid data points used for analysis in each condition.

To address RQ1 related to food portion size, a paired 2-sample *t* test was conducted to find relations between the amount of

XSL•FO
RenderX

food (milk or cereal) poured by the participants in VR and RL conditions. A total of 317 data points were used for cereal and 305 data points were used for milk.

To address RQ1 regarding food healthiness, we conducted a 2-sample *t* test assuming equal variances to compare the weight of unhealthy and healthy cereal and milk in both conditions. Healthy and unhealthy portion sizes were compared separately for RL and VR conditions. We had 284 data points for cereal in the RL condition, 264 data points for cereal in the VR condition, and 276 data points for milk in both VR and RL conditions.

In response to RQ1 regarding the container size, we analyzed how close participants' guesses were to the correct container size. We first performed a weighted kappa analysis on the study conditions (VR or RL) and participants' guesses to the container (bowl and glass) size. For weighted kappa analysis, we used weighted metrics for calculating agreement in a way that a perfect agreement was assigned a 100% weight, while the next closest option received a 75% score, and the third option received zero weight. We also conducted a chi-square test to supplement the results of weighted kappa analysis. There were 195 data points for bowl size guesses, and 193 data points for glass size guesses.

To analyze RQ2 regarding presence, embodiment, and sickness, we excluded incomplete responses, which led to a total of 411 responses. We analyzed these responses for 2 age groups: those aged 5 - 8 years (240 responses) and 9 - 12 years (171 responses). We split our analysis into 2 groups, as previous research suggests that children undergo different motor skill development at different ages [53,73,74], which may influence their interactions within VR.

The Likert responses were coded as numeric values from 1 to 5 for "Not at all" to "Very much," and from 1 to 3 for "Not at all" to "A lot." We calculated the mean of the responses for all questions under the same category. For instance, the average of Q1, Q2, and Q3 scores formulated the score for "Sense of Presence." Finally, an embodiment score was calculated by taking an average of the responses for 3 categories: body ownership, self-location, and control.

To investigate the repeatability of tasks in VR (RQ2), we had 134 data points for cereal and 140 data points for milk. A paired 2-sample *t* test was performed to compare the amount of cereal and milk poured by participants in VR.

Finally, for RQ3 regarding the effect of demographics, we again conducted kappa analysis to find the influence of age and gender on the amount of food (milk or cereal) served in VR and RL conditions, as well as in guessing the container sizes. The analysis was done separately for boys and girls, younger than and older than 8 years, that is, aged 5 - 8 and 9 - 12 years. There were 306 data points for cereal and 295 data points for milk.

Results

We structure our findings around the 3 RQs. A total of 437 children participated in our VR study. Out of 437 participants, 328 children had no experience with VR and 109 children had used VR at least once.

RQ1: Food Portion Size (Being Served in RL Versus Serving Themselves in VR)

As per the paired 2-sample *t* test, the mean value of milk poured in VR is 242.5 (SD 8623.6) g and in RL is 215.9 (SD 9073.9) g. For both conditions, the SDs for the mean value of milk poured are high. The high SDs are due to the high variability of milk content poured by the participants. The mean value of milk poured in VR was significantly higher than in RL (t_{304}=3.87; P<.001).

We also conducted the paired 2-sample *t* test to understand the order effects (VR first vs RL first) on milk portion sizes. There was a significant difference for participants who completed VR first (t_{147}=6.37; P<.001). Children poured more milk in VR (mean 258.9, SD 9153.2 g) than in RL (mean 212.5, SD 9827 g) when they completed the VR condition first. However, if the participants completed the RL condition first, there was no significant difference in milk weight between the 2 conditions.

We conducted a paired 2-sample *t* test, which shows that the mean value of cereal weight in VR (mean 54.1, SD 796.3 g) was significantly higher than the mean value of cereal weight in RL (mean 41, SD 388.6 g; t_{316}=8.16; P<.001). This shows that children served themselves more cereal in VR than what they were served by the researcher in RL. The SD for the mean value of cereal poured is much higher in the VR condition than in RL.

We conducted the paired 2-sample *t* test to understand the order effects (VR first vs RL first) for cereal. Regardless of which condition was completed first, there was a significant difference that participants served more cereal in VR than in RL. For participants who completed the VR condition first, cereal weight in VR (mean 57.2, SD 826.7 g) was significantly higher than in RL (mean 43, SD 454.7 g; t_{147}=5.71; P<.001). For participants who completed the RL condition first, cereal weight in VR (mean 51.3, SD 758 g) was significantly higher than in RL (mean 39.1, SD 323.7 g; t_{165}=5.83; P<.001).

RQ1: Food Healthiness

As per a 2-sample *t* test, the mean value of unhealthy cereal poured in RL (mean 48.4, SD 461.1 g) was significantly higher than the mean value of healthy cereal weight in RL (mean 34.2, SD 258.8 g; t_{284}=6.31; P<.001). This shows that participants served themselves more portions of unhealthy cereal in RL than the healthy cereals. Similarly, the mean value of unhealthy cereal poured in VR (mean 63.1, SD 934.1 g) was significantly higher than the mean value of healthy cereal poured in VR (mean 41.1, SD 225.2 g; t_{264}=−7.47; P<.001). This shows that participants served themselves more portions of unhealthy cereals in VR than the healthy cereals. High values of SD are due to the large variability of the food selection made by children. The healthiness of the milk had no significant effect on the serving sizes chosen by the participants in either the VR or RL conditions. These findings confirm that children's attitude toward food healthiness is similar in both VR and RL.

XSL•FO
RenderX

RQ1: Container Sizes

The weighted kappa statistics showed no agreement (κ=0.34) between the given and guessed size of bowls between VR and RL conditions. This shows that children cannot accurately match the size of the serving bowl between VR and RL. Participants made more incorrect guesses in VR first than in RL first, with a correct to incorrect distribution ratio of 59:64 and 46:26, respectively, as shown in Table 3. Chi-square test showed significant differences for small-sized bowls ($\chi^2_{1,\ 53}$=12.0; *P*=.001). This suggests that children cannot match the small-sized containers in RL correctly if the bowl is first shown in VR. No significant differences were found for medium-sized and large-sized bowls. These results suggest that children are better at matching the medium-sized containers, regardless of whether they saw them in VR or RL first.

Table . Guesses for bowl and glass based on conditions. VR[a] first means that participants were shown the bowl or glass in the VR condition first and then were asked to guess the size in the RL[b] condition.

Conditions	Small (correct/incorrect)	Medium (correct/incorrect)	Large (correct/incorrect)	Total (correct/incorrect)
Bowl VR first	9/29	37/9	13/26	59/64
Bowl RL first	12/3	21/10	13/13	46/26
Glass VR first	10/32	17/26	20/18	47/76
Glass RL first	11/10	19/9	13/8	43/27

[a]VR: virtual reality.

[b]RL: real-life.

For glass size guesses, we similarly performed weighted kappa analysis and found no agreements (κ=0.20) between the given and guessed size for the VR and RL conditions. This shows that children cannot accurately match the size of the serving glass between VR and RL. Similar to the bowl size condition, participants made more errors in guessing glass sizes in the VR-first condition than in the RL-first condition, with distribution between correct and incorrect guesses as 47:76 and 43:27, respectively. Chi-square test showed nonsignificant differences for all sizes of the glass. The results (correct-to-incorrect ratio for different sizes) in Table 3 suggest that children are better at matching different-sized glasses in VR after first seeing them in RL.

RQ2: Presence, Embodiment, and Simulator Sickness

Table 4 presents the presence and embodiment scores for 2 age groups: those aged 5 - 8 years (n=240) and 9 - 12 years (n=171). The overall scores for each factor are similar for both groups. The mean embodiment score of 3.1 suggests that participants felt a moderate or ambiguous sense of ownership and control regarding their virtual body and its location in VR. Similarly, Presence received the mean score value of 2.9 for both groups, suggesting that participants partially felt that they were present in VR. On the other hand, control received a relatively higher mean score for both groups: 3.5 for those aged 5 - 8 years and 3.6 for those aged 9 - 12 years.

We also analyzed the simulator sickness score for the 2 age groups: those aged 5 - 8 years and 9 - 12 years and found similar results. Table 5 presents a cumulative mean score of 1.10 (SD 0.32) and 1.12 (SD 0.33) for younger and older children, respectively. The low sickness scores suggest that participants did not feel any simulation sickness in VR. Some participants, however, mentioned feeling eye fatigue after the study.

XSL•FO
RenderX

Table . Presence (denoted as P) and embodiment (denoted as E) scores on the Likert scale (1 being "Not at all" to 5 being "Very much") for age groups of 5 - 8 and 9 - 12 years.

Factors	Questions	Age group (5 - 8 years), mean (SD)	Age group (9 - 12 years), mean (SD)
P: Presence	• Q1. Do you feel like you're in a kitchen while wearing the goggles? • Q2: Do you feel like you're still in the museum? • Q3: Does the kitchen feel real?	2.9 (1.1)	2.9 (1.1)
E: Body ownership	• Q4: Does the virtual hand feel like your hand? • Q5: Does the virtual body feel like your body?	2.6 (1.4)	2.5 (1.2)
E: Self-location	• Q6: Do the virtual hands appear where your hands are? • Q7: Does the virtual body appear where your body is?	3.2 (1.1)	3.4 (1.0)
E: Control	• Q8: Do the virtual hands move like you want them to?	3.5 (1.0)	3.6 (0.95)
Embodiment score (Inclusive of Q4-Q8)		3.1 (1.2)	3.2 (1.0)

Table . Sickness score on the Likert scale (1 being "Not at all" to 3 being "A lot") for the age group of 5 - 8 and 9 - 12 years.

Question ID	Question	Age group (5 - 8 years), mean (SD)	Age group (9 - 12 years), mean (SD)
S1	Do you feel sick?	1.1 (0.4)	1.1 (0.2)
S2	Does your head hurt?	1.1 (0.4)	1.1 (0.3)
S3	Do your eyes hurt?	1.1 (0.3)	1.2 (0.4)
S4	Is your stomach upset?	1.1 (0.3)	1.1 (0.3)
S5	Are you dizzy with your eyes open?	1.2 (0.4)	1.2 (0.4)
S6	Are you dizzy with your eyes closed?	1.1 (0.3)	1.1 (0.3)
S7	Do you feel like burping at all?	1.1 (0.2)	1.1 (0.4)
Sickness score (Inclusive of S1-S7)		1.1 (0.3)	1.1 (0.3)

RQ2: Repeatability

A paired 2-sample *t* test did not show any significant difference between the first and second trials for pouring milk (t_{139}=0.19; *P*=.84) and cereals (t_{133}=0.77; *P*=.44). These values could suggest that VR offers a reliable test-retest mechanism for food and drink with respect to portion size.

RQ3: Demographics

Kappa statistics were used to assess the level of agreement between VR and RL container size estimations across different demographic subgroups. The results indicated no significant agreement between the 2 conditions when analyzed by age or gender (see Table 6 for details). Similarly, no significant agreement was found between VR and RL food portion sizes across these age and gender groups.

Table . Summary of kappa statistics for container size estimation.[a]

	κ values for glass size estimation	κ values for bowl size estimation
Male	0.15 (n=108)	0.39 (n=107)
Female	0.28 (n=87)	0.26 (n=86)
5 - to 8-year-olds	0.18 (n=120)	0.33 (n=118)
9 - to 12-year-olds	0.24 (n=75)	0.35 (n=75)

[a]All κ values are less than 0.40 and hence considered small.

XSL•FO
RenderX

Qualitative Results

To better understand the quantitative findings presented in the previous section, now we present the qualitative findings that were generated from participants' responses and researchers' notes.

Higher Food Portion Sizes in VR

We found multiple reasons for higher food portion sizes in VR than in RL (RQ1). First, there is no sense of weight in VR, such as the weight of a milk can or serving bowl. The lack of weight in VR led to a "mismatch" in motor expectations. For instance, when lifting the virtual 2-liter milk can, participants often overcompensated for its expected weight, leading to exaggerated movements and accidental spilling. We included any food that landed in the bowl or glass in our final measurements because these spills were not technical errors or outliers; rather, they were a fundamental result of the participant's interactions with the VR interface. Removing them would have artificially inflated the accuracy of the VR condition.

Second, the VR food environment was found fun by participants, where they were intrigued by the novelty of the VR and/or the task presented. Participants turned the simple task of pouring food into a playful activity, where they continued pouring to overflow the containers. This is evident by high SD values for cereal and milk portion sizes in VR as compared with the RL condition (as discussed under section "RQ1: Food Portion Size"). As such, children enjoyed the freedom to interact with food in VR and tried different activities to test the virtual environment (eg, throwing containers, turning containers over to check gravity, etc). Although participants received identical instructions to serve themselves food in both conditions (VR and RL), significant variations in portion size were observed. These discrepancies likely reflect differences in how the younger cohort interpreted or retained instructions across environments.

Finally, since the children did not handle real food in RL, this pseudopouring could have created a difference. For example, participants might not want to be judged by the researcher on their intake and hence agreed on smaller portions of food in RL. Interestingly, we identified an *order effect* specifically for milk portions: participants who completed the RL condition first were more likely to replicate similar volumes in VR. This suggests that the visual transparency of the RL glassware provided a sensory anchor that guided subsequent actions in VR. In contrast, when the VR condition was performed first, the lack of haptic weight, the playful nature of the environment, and different perception of size in VR [75] might have led to significantly larger milk portions that did not align with subsequent RL servings. The RL bowl, on the other hand, was opaque; hence, the visual guidance was limited to influence participants' behaviors in VR.

Lower Presence and Embodiment in VR

In our study, presence in the VR experience and embodiment of the VR avatar received lower scores from the participants (RQ2). In the poststudy questionnaire, many participants commented that the virtual kitchen did not resemble the kitchen at their home. As said by one of the participants, "It's not my own kitchen!." Another mentioned, "The table wasn't like a kitchen table." Participants looked at how realistic the kitchen and the surroundings were, and the limitations with real-world resemblance reduced their sense of presence in the virtual kitchen. The participants strongly felt that they were in the museum while doing the VR activity. The concept of the virtual world was new to them, and they could not imagine being at 2 places—both in the real and virtual worlds simultaneously.

The low score for the control factor for 5- to 8-year-olds is in line with our observations, as younger children could not use the controllers effectively. They found it challenging to navigate the VR tasks through controller buttons and movements. On the other hand, presence received the same scores from both age groups, and self-location was found higher in older children than in the younger age group (5-8 years).

Low scores for body ownership suggest that participants did not feel embodied with the VR avatar. Participants provided comments on avatar resemblance, physicality, and fidelity. One participant mentioned in the poststudy questionnaire, "Hands do not feel real because real hands can't go through the table." Another participant commented on the nonhuman characteristics of the avatar: "The boy [avatar] did not blink. It looked like a ghost." One participant noted in the questionnaire: "The body feels like a cartoon. You can't feel the hands; you can't move the fingers." Participants were looking for sensory sensations in their hands while doing actions in VR, and perfect matching of the virtual avatar with their body characteristics (eg, skin color, hand size, body size, etc)—the absence of which reduced their ownership of the virtual body.

Discussion

Principal Findings

Our study investigated the potential of using VR to conduct food choice studies with children. We examined their perception of food characteristics (portion size and container size) and their experience of being in VR (embodiment, presence, and repeatability) and compared them with the RL interactions at a museum. We found key differences in terms of food portion sizes (RQ1) as children served themselves more food in VR than in RL settings, along with some similarities as children served themselves more portions of unhealthy food in both VR and RL. For container size mapping (RQ1), we found that children were good at matching the medium-sized bowls more accurately, regardless of whether they saw them in VR or RL first, and they are better at matching different-sized glasses in VR after first seeing them in RL. On the other hand, presence and embodiment scores (RQ2) were found low, whereas control was higher for both age groups of 5 - 8 and 9 - 12 years. Additionally, VR qualified the test-retest mechanism (RQ2) for both food and drink, as children were able to repeat the same tasks in VR. Finally, we found no influence of demographics on container size estimation or food portion size (RQ3).

Our study revealed that the effectiveness of a virtual food environment is influenced by factors related to both VR's affordances and the study's setting. These factors include the differences in handling food in VR and RL conditions, level of engagement in VR versus RL, and the limitations of VR in

XSL•FO
RenderX

supporting tactile and sensorial interactions. These differences highlight how VR, as a novel interaction medium, offers a distinct experience from RL surroundings, due to which a direct mimicry of RL interactions is neither feasible nor should be expected. The study also highlighted challenges related to ensuring data accuracy with VR and inquiring about the real and virtual worlds in the questionnaire. Consequently, our study suggests that the current form of VR is not suitable for conducting food choice studies with children. Next, we discuss our findings into 2 themes and present 4 design recommendations to guide effective VR experiences for this research context in future.

Playfulness Versus Science

Overview

Our results showed that children served themselves more food (both cereals and milk) in VR than in RL (RQ1). The SD was also much higher for VR than for RL, likely because children found the VR environment engaging and playful, wherein they dismissed the study instructions to explore the VR space (such as testing how food would overflow). Conversely, in RL, children were assisted by a researcher during the pouring tasks, which limited their opportunities for play. Despite the playful behavior in VR, we found that participants in both conditions served themselves more unhealthy food. This suggests that the differences in food-handling methods did not influence the children's overall food choices.

While extending the familiarization phase to mitigate the novelty effect may seem beneficial, it presents significant risks regarding participant fatigue and disengagement, particularly within the 5 - 12 years age demographic. Prolonged exposure to VR environments can be cognitively taxing for younger children. Consequently, we used a brief orientation to ensure that participants were proficient with the mechanics while maintaining the focus required for reliable data collection. This approach allowed us to capture more authentic behavior before the onset of boredom or cognitive exhaustion.

Prior work has shown that VR can put children in a positive mindset, even during uncomfortable medical therapy [32]. Another study [76] explored the balance between playfulness and learning, in which playfulness is a conceptualization of intrinsic motivation, suspension of reality, and internal locus of control. In our study, the lack of consequences for actions in VR amplified 2 of these aspects of playfulness: suspension of reality and internal locus of control. Children enjoyed the VR environment and the freedom of controlling their food serving without supervision increased their excitement. Research on ethical practices of involving children with VR [77] highlighted that children are easily excitable within VR worlds; hence, their exposure should be limited. We observed this effect firsthand, although our virtual environment was a kitchen and not designed to be explicitly playful. The excitement for the VR experiences was also evident by the fact that our study station was the most popular among all other stations in the study room and still children were happy to wait for their turn.

While playfulness increased the likability of VR among children, it raised challenges for using it as a scientific tool. Engagement came at the cost of accuracy and precision, diminishing the quality of the data we collected on food selection. These insights raise important questions about how to use VR for scientific studies with children, for which we provide 2 guidelines next.

Recommendation 1: Design Supervised VR Food Experiences

To achieve greater validity in food choice studies with children, we suggest developing supervised VR food experiences that include a moderator in the virtual environment. The moderator, who could be a member of the research team, could guide children through the required activities such as selecting food or containers. While some attempts have been made to develop supervised VR experiences, for example, to encourage equal participation in moderated online meetings [78], engaging children in supervised VR experiences for scientific activities has not received any attention yet.

The role of the moderator and their interactions within the environment must be carefully considered, as they should not overshadow the children's interactions in the virtual food environment. Additionally, it is also important to explore how a moderator could be embodied as an avatar (eg, humanoid or nonhumanoid) to encourage greater engagement from children of different ages. Having a moderator might lead to altered behavior, which aligns with the existing literature on the Hawthorne effect [79,80]. However, as per Slater [81], there are always a multitude of variables that could potentially influence the results, and he recommended using data triangulation—which could be employed when designing and testing supervised VR food experiences for children.

Recommendation 2: Design Gamified VR Food Experiences for Children

Building on the insights from our study that VR was a fun experience for children, we suggest developing gamified VR experiences. Recent research has used gamified VR to prepare adolescents (aged 10 - 16 years) for the rigors of magnetic resonance imaging scans [82]. By engaging in a "statue game" with a virtual panda, children practiced physical movements that mirrored the scan procedure, such as reclining on a bed in synchronization with the virtual character. This study demonstrated that interactive VR environments, combined with real-time feedback, effectively increase task compliance and minimize the involuntary motion that often compromises medical imaging. Gamified VR experiences could support children's intrinsic motivation to find play in everyday activities [76,83]. However, such experiences must be carefully designed so that the game elements encourage children to perform the required tasks more accurately and discourage behaviors that go against the study goals. Points and award systems can be used to positively reinforce children within the game. For instance, players could receive points based on how cleanly (ie, without any spillage) they perform a pouring task in VR. Consequently, gamification would not only increase children's willingness to perform the task but also enable researchers to use the virtual medium more appropriately.

XSL•FO
RenderX

"It's Not My Own Kitchen!": Presence and Embodiment

Overview

Our study revealed that children have a high threshold for realism with both their virtual avatars and the VR environment. The lack of realism created issues, as children struggled to embody the virtual avatar and did not feel present in the virtual kitchen (RQ2, Table 4). The participants' behavior in our study is typical of the Goldilocks problem [84]. In the popular children's story, Goldilocks seeks a chair and porridge that are "just right"—not too big or too small, not too hot or too cold. Similarly, our participants desired a virtual body avatar and kitchen setting that felt "just right" to them. This is an age-appropriate behavior, as children (unlike adults) engage in an internal dialogue to identify objects based on their prior knowledge (ibid.).

For embodiment, Slater [81] explained the differences in embodiment levels across age groups by looking at the biases we develop as we get older. The authors argued that as adults, our concept of "real" in VR is biased. Since adults already know the limitations of the technology, we are more likely to accept certain nonrealistic aspects of a virtual world as being "realistic" enough. Likewise, adults understand that the behavior of virtual objects may deviate from the real world, and they would still find it an acceptable interactive experience.

Our findings are in line with prior works [52,54,55], as they all highlight the importance of human-like avatars for children to feel embodied in VR. For example, Liao and colleagues [55] also reported low embodiment in a study with 6 - to 8-year-old children, as the VR experience was described as not real. However, physiological and emotional responses collected during that study suggested the children were immersed in the virtual environment. Research on the cognitive development of children [85] suggests that younger children decipher between "real" and "unreal" based on their prior perception and reasoning of their surrounding world. Our study confirmed this observation, as participants still found the VR experience very engaging despite providing low presence scores. Further studies are required to understand if and how low embodiment in VR influences the food portion sizes children serve.

In our study, control was found to be slightly higher for older children (aged 9 - 12 years) with a mean score of 3.6 (SD 0.95) as compared with the younger group (aged 5 - 8 years) with a mean score of 3.5 (SD 1.0). Control is correlated with their sense of self-location, and older children therefore also had a better sense of self-location than younger children. We found that younger children found it challenging to maneuver the VR controllers for the required tasks, whereas older children demonstrated better motor skills. We also observed that younger children were losing their balance while performing the task of pouring and tilting in VR (it is important to note that no accidents were reported, as children were seated and were in constant supervision). These findings are consistent with prior literature on movement skill development [86], which suggests that control skills are still developing in children aged 1 - 7 years, and therefore they require greater coordination to perform such tasks.

Presence was also found low in our study, potentially because of the language used in the instructions. By asking children to imagine being in their own kitchen and serving breakfast (as described under section "Study Procedure"), we unintentionally led them to compare the virtual environment with their actual home kitchen in terms of layout and food items. Moreover, participants also commented that the virtual hands did not feel real, thus contributing to low scores. Despite the low scores, children enjoyed the overall activity.

These insights raise important questions that require further investigation. For instance, is it possible for children to enjoy VR activities without being "present" in the virtual world? How important is high embodiment for children in VR? Finally, how can we design efficient tools to measure children's presence in VR? Next, we discuss 2 guidelines on how to enhance children's immersion in VR.

Recommendation 3: Design Virtual Food Selection Tasks Around Children's Motor and Perceptual Abilities

In our study, the VR experience primarily involved 2 interactions: grabbing an object (a cereal box and milk can) and tilting it onto a bowl or glass, respectively. We found that younger children (aged 5 - 8 years) struggled with these tasks because they required complex control through the VR controllers. Owing to the developing motor skills in children of various age groups, we suggest that designers should carefully develop the virtual tasks around the motor capabilities of the target age group. Additionally, designers should not try to exactly mimic the real-world interactions with food in VR as they may not be feasible. For instance, the pouring task in our study was problematic, as there is reduced weight perception in VR, which led to accidental spillage. Hence, we suggest mapping the real-world tasks to simpler actions within VR such as grabbing or moving.

Our study also indicated that children are more likely to perceive food container sizes correctly for the "medium" size (RQ1). As per Piaget's discussion [87], children younger than 11 years are known to have difficulty in estimating the object volume. For instance, such children focus only on 1 dimension of an object and tend to ignore others. Hence, they find it difficult to understand that a quantity does not change if it has been altered (eg, by stretching, shrinking, spread out, etc). Similarly, another study [75] suggests that adult participants perceived sizes very differently between VR and RL, and that in VR, they were more reliant on the familiar sizes to perceive the size of unfamiliar objects. In our study, when we asked children to pick the same bowl in VR as they had picked in RL and vice versa, children could make correct estimations only for medium size. They were able to identify all 3 sizes of glasses in VR if they had seen them first in RL. With the VR first condition, more incorrect guesses were observed. Our study suggests that apart from the cognitive inabilities and abilities of our target group, we anticipate that the virtual environment adds another layer of complexity that makes the size estimation challenging for children.

XSL•FO
RenderX

Participants guessing the "medium" size containers more accurately than other sizes could also be understood through the Goldilocks problem [84], where children selected "just the right kind" of containers for themselves. For instance, children aged 5 - 12 years are typically served in small- to medium-sized containers at home to prevent food wastage. Participants may have selected medium-sized containers due to their everyday exposure to that size. Moreover, previous works in object size perception [50,88] also indicate that object sizes in VR have an inverse correlation with virtual hand sizes, under all degrees of avatar realism. In our study, we used a single size avatar for all participants, without scaling to participants' bodies. Hence, further studies are required to understand the connection between avatar size and container size selection for children.

Our findings related to container sizes also suggest that it is more efficient for future studies to have a single size. Additionally, it was challenging to replicate the continuous variation in food amount while being poured into virtual containers in VR. Therefore, for future studies on food selection in VR, we suggest that researchers should focus on using discrete portion sizes, such as different slices of pizzas or a bowl of soup instead of involving actions such as pouring or serving food through spoons.

Recommendation 4: Design Visual Embodiment and Presence Questionnaires for Children

Our final recommendation relates to how to measure children's embodiment and presence. In our study, we found that children were unable to understand the questions related to the real and virtual world. For instance, questions related to embodying an avatar were difficult for children to interpret, as the virtual body and hands were neither their own nor very realistic. As such, children were very focused on their RL experiences and were unable to connect to an unfamiliar kitchen and unknown avatar. Our concerns with the existing questionnaires resonate with the concerns raised by other researchers [49,89], as they also noted the methodological challenges of using questionnaires and emphasized the importance of using other measures to better understand participants' perceptions of the virtual world.

Slater [81] argued that presence is not a natural phenomenon, as we do not think about it in our everyday life. Also, unlike other feelings such as anxiety, presence cannot be easily compared with our previous life experiences. Since children have fewer RL experiences than adults, their familiarity with different concepts in VR may differ. Hence, researchers should inquire about embodiment on parameters that children can easily compare with. Additionally, presence should be inquired against familiarity. For instance, questions related to familiarity with the designed experience should be asked to understand children's perspective. Designers should also build VR experiences that children are likely to be familiar with to help them develop valid connections in the virtual environment.

New tools are, therefore, required to inquire about measures of embodiment and presence of children in VR. Prior work [90] developed a visual questionnaire to understand the perceived physical literacy of children. The questionnaire shows a "bunny" character performing different physical activities. Children interpret the bunny's activities and indicate their literacy level around physical activities. Using an animal-based visual questionnaire helped children to focus on the question being asked without focusing on the finer details of the bunny character. Similarly, using animal avatars with children in VR has also received positive responses in previous research as realistic behaviors were not expected from animal characters [54]. Visual questionnaires coupled with animal characters thus present a promising direction toward measuring children's embodiment and presence within VR.

Limitations

This study has certain limitations, which might influence the generalizability of the study findings. First, there was a difference in how the VR and RL tasks were managed in our study. In the VR condition, children poured their own food, but in the RL condition, they did not. This was a necessary limitation due to the study's setting in a public museum, where the risk of spillage could impact other visitors and damage the VR equipment. Although we anticipated that the food portion sizes would be similar between both conditions, they varied. We found that this variation was not only due to the difference in pouring mechanisms but also because the novelty of VR combined with a lack of supervision within VR—which led children to treat it more as a playful activity. This highlights why our recommendation to design supervised experiences is an important consideration for future studies.

Second, children were not allowed to eat the food they served themselves due to food safety reasons such as allergies. Since the amount of food served may differ from the amount consumed, further research is needed to understand the relationship between these 2 variables. Moreover, we used an average avatar size and did not adjust the VR kitchen or its features such as container sizes to account for each child's hand size or age, which may have influenced our findings. Moreover, since eating practices vary across cultures, some may not consider cereals as a healthy option. While it is infeasible to choose a food item that satisfies every culture, our choice of food items was driven by the health star rating and sugar content mentioned on the package.

Third, due to the complexity of multiple tasks for each RQ, as well as the context of conducting research in an active public setting of a popular museum with multiple data collection stations, data loss across different tasks was an issue, causing imbalance of participant numbers across each RQ. Additionally, the playful behaviors of the participants of continuously pouring food in VR caused significant data loss for VR conditions.

Another limitation of our study is that it was conducted with an age group younger than the VR industry's recommended demographic. Although we received ethics approval and implemented safety protocols, such as limiting VR exposure and ensuring that children remained seated, the design of the headsets may have influenced the children's experience and our findings. The implication of the study could be limited to the specific hardware setup that was used, which was the Oculus Quest 2. Moreover, our participant sample may not be representative of all socioeconomic groups. Because the study was conducted at a museum with an entry fee for accompanying

XSL•FO
RenderX

adults and free access to children, the findings may not generalize to children from other socioeconomic backgrounds.

Finally, we acknowledge that the modified embodiment and simulator sickness questionnaires may have impacted the results. These changes were necessary to simplify the questions and make them suitable for child participants.

Conclusions

This paper presents insights from a controlled study conducted with 437 children (aged 5 - 12 years) at a science museum. We investigated the potential of using VR as a tool for food choice research by measuring how children perceive food portion size and container size, and how their experience in VR compares with an RL setting. Our study raised important questions about how to collect accurate data with children using VR, how to design immersive experiences for them, and how to create a valid questionnaire to measure their experience in VR. In response, we present 4 recommendations around designing supervised virtual food experiences, applying gamification to ensure data accuracy, developing virtual tasks around children's motor abilities, and designing visual questionnaires to inquire about children's immersion in VR. Despite certain limitations, our study presents the first understanding of how children perceive food in VR, how they behave in a simulated food environment, and finally how VR could serve as a potential measuring tool for measuring their eating behaviors. These insights offer foundational knowledge for future VR research with children in the domains of HCI and health, while also opening new avenues for researchers at the intersection of VR and food science.

Acknowledgments

The authors used Gemini, an artificial intelligence large language model from Google, as an editorial tool to enhance the clarity and flow of the text. This usage was strictly limited to language refinement, ensuring that all research contributions, data interpretation, and intellectual content originate entirely from the authors.

Funding

This work is funded by Scienceworks Museum and Deakin University. Associate professor TH is the recipient of an Australian Research Council Discovery Early Career Researcher Award (DE200100898) funded by the Australian Government.

Data Availability

Due to the restrictions imposed by our institutional review board and the absence of explicit participant consent for public data dissemination, the full dataset used in this study cannot be made available for public use. All data necessary to support the findings and replicate the statistical analyses are contained within the manuscript and its supplemental files. Any specific, summarized data not included in the paper may be provided by the corresponding author upon a reasonable, ethically compliant request.

Authors' Contributions

DA, TH, CGR, and SYT were involved in the conceptualization of research goals. DA, TH, and CGR were involved in the data curation. DA, TH, and MM were involved in the formal analysis of the study data. CGR and TH managed the funding for this project. TH, CGR, and SYT were involved in the investigation process and the study conduct. TH, CGR, and SYT designed the study methodology. TH and CGR were involved in the project administration. DA managed the writing – original draft. DA, TH, CGR, SYT, and MM were involved in writing – review and editing.

Conflicts of Interest

None declared.

References

1. Khot RA, Mueller FF. Human-food interaction. Found Trends Hum Comput Interact 2019 Aug 29;12(4):238-415. [doi: 10.1561/1100000074]
2. Obrist M, Comber R, Subramanian S, Piqueras-Fiszman B, Velasco C, Spence C. Temporal, affective, and embodied characteristics of taste experiences. Presented at: CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; Apr 26 to May 1, 2014. [doi: 10.1145/2556288.2557007]
3. Khot RA, Aggarwal D, Pasumarthy N. Understanding screen-based dining practices through the lens of mindful eating. Presented at: CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; Apr 29 to May 5, 2022. [doi: 10.1145/3491102.3517651]
4. Thomaz E, Essa IA, Abowd GD. Challenges and opportunities in automated detection of eating activity. In: Mobile Health: Springer International Publishing; 2017:151-174. [doi: 10.1007/978-3-319-51394-2_9]
5. Vu T, Lin F, Alshurafa N, Xu W. Wearable food intake monitoring technologies: a comprehensive review. Computers 2017;6(1):4. [doi: 10.3390/computers6010004]

XSL•FO
RenderX

6. Ferdous HS, Vetere F, Davis H, et al. Celebratory technology to orchestrate the sharing of devices and stories during family mealtimes. Presented at: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems; May 6-11, 2017. [doi: 10.1145/3025453.3025492]
7. O'Hara K, Helmes J, Sellen A, Harper R, ten Bhömer M, van den Hoven E. Food for talk: phototalk in the context of sharing a meal. Hum Comput Interact 2012;27(1-2):124-150. [doi: 10.1080/07370024.2012.656069]
8. Tai YL, Aggarwal D, Khot RA. Reconnecting with food through dining play. Presented at: CHI PLAY '20: Proceedings of the Annual Symposium on Computer-Human Interaction in PLAY; Nov 2-4, 2020; Virtual Event, Canada p. 334-348. [doi: 10.1145/3410404.3414231]
9. Arnold P, Khot RA. "You Better Eat to Survive": exploring cooperative eating in virtual reality games. Presented at: TEI '18: Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction; Mar 18-21, 2018. [doi: 10.1145/3173225.3173238]
10. Khot RA, Aggarwal D, Yi JY, Prohasky D. Guardian of the snacks: toward designing a companion for mindful snacking. Multimodality Soc 2021 Jun;1(2):153-173. [doi: 10.1177/26349795211007092]
11. Khot RA, Ng J, Aggarwal D. Crafting tangible interfaces for human digestion: unpacking the research through design prototyping journey. Presented at: TEI '22: Proceedings of the Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction; Feb 13-16, 2022. [doi: 10.1145/3490149.3502252]
12. Khot RA, Aggarwal D, Pennings R, Hjorth L. Edipulse: investigating a playful approach to self-monitoring through 3D printed chocolate treats. Presented at: CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems; May 6-11, 2017; Denver, CO p. 6593-6607. [doi: 10.1145/3025453.3025980]
13. Covaci A, Alhasan K, Loonker M, et al. No pie in the (digital) sky: co-imagining the food metaverse. Presented at: CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems; Apr 23-28, 2023. [doi: 10.1145/3544548.3581305]
14. Cheah CSL, Barman S, Vu KTT, et al. Validation of a virtual reality buffet environment to assess food selection processes among emerging adults. Appetite 2020 Oct 1;153:104741. [doi: 10.1016/j.appet.2020.104741] [Medline: 32445771]
15. Persky S, Goldring MR, Turner SA, Cohen RW, Kistler WD. Validity of assessing child feeding with virtual reality. Appetite 2018 Apr 1;123:201-207. [doi: 10.1016/j.appet.2017.12.007] [Medline: 29277518]
16. Ung CY, Menozzi M, Hartmann C, Siegrist M. Innovations in consumer research: the virtual food buffet. Food Qual Prefer 2018 Jan;63:12-17. [doi: 10.1016/j.foodqual.2017.07.007]
17. Plechatá A, Morton T, Perez-Cueto FJA, Makransky G. A randomized trial testing the effectiveness of virtual reality as a tool for pro-environmental dietary change. Sci Rep 2022 Aug 22;12(1):14315. [doi: 10.1038/s41598-022-18241-5] [Medline: 35995946]
18. Xu C, Siegrist M, Hartmann C. The application of virtual reality in food consumer behavior research: a systematic review. Trends Food Sci Technol 2021 Oct;116:533-544. [doi: 10.1016/j.tifs.2021.07.015]
19. Haycraft E, Karasouli E, Meyer C. Maternal feeding practices and children's eating behaviours: a comparison of mothers with healthy weight versus overweight/obesity. Appetite 2017 Sep 1;116:395-400. [doi: 10.1016/j.appet.2017.05.033] [Medline: 28536055]
20. Ben-Sefer E. The childhood obesity pandemic: promoting knowledge for undergraduate nursing students. Nurse Educ Pract 2009 May;9(3):159-165. [doi: 10.1016/j.nepr.2008.07.006] [Medline: 18783989]
21. Sanderson K, Patton GC, McKercher C, Dwyer T, Venn AJ. Overweight and obesity in childhood and risk of mental disorder: a 20-year cohort study. Aust N Z J Psychiatry 2011 May;45(5):384-392. [doi: 10.3109/00048674.2011.570309] [Medline: 21500955]
22. Population-based approaches to childhood obesity prevention. : World Health Organization; 2012 URL: https://apps.who.int/iris/bitstream/handle/10665/80149/9789241504782_eng.pdf [accessed 2026-06-09]
23. Di Cesare M, Sorić M, Bovet P, et al. The epidemiological burden of obesity in childhood: a worldwide epidemic requiring urgent action. BMC Med 2019 Nov 25;17(1):212. [doi: 10.1186/s12916-019-1449-8] [Medline: 31760948]
24. Nault K, Fitzpatrick M, Howard PH. Engaging youth in food environments research: the role of participation. J Hunger Environ Nutr 2010 May 28;5(2):174-190. [doi: 10.1080/19320248.2010.487020]
25. Russell CG, Burnett AJ, Lee J, Russell A, Jansen E. Measurement is the foundation of research and theory on children's eating behaviours: current issues and directions for progress. Appetite 2023 Jul 1;186:106546. [doi: 10.1016/j.appet.2023.106546] [Medline: 36958633]
26. Christensen P, Prout A. Working with ethical symmetry in social research with children. Childhood 2002 Nov;9(4):477-497. [doi: 10.1177/0907568202009004007]
27. Lakshminarayanan M. Researching children's experiences; approaches and methods. J Evid Based Soc Work 2019 Nov 2;16(6):687-690. [doi: 10.1080/26408066.2019.1676356]
28. Gutierrez RJ. Virtual reality: the expanding children's market. Journal of Gaming & Virtual Worlds 2020 Jun 1;12(2):217-219. [doi: 10.1386/jgvw_00016_5]
29. Yamada-Rice D, Mushtaq F, Woodgate A, et al. Children and virtual reality: emerging possibilities and challenges. Royal College of Art. URL: https://researchonline.rca.ac.uk/id/eprint/3553 [accessed 2023-06-11]

XSL•FO
RenderX

30. Aubrey JS, Robb MB, Bailey J, Bailenson J. Virtual reality 101: what you need to know about kids and VR. : Common Sense Media; 2018 URL: https://www.commonsensemedia.org/sites/default/files/research/report/csm_vr101_final_under5mb.pdf [accessed 2026-06-09]
31. Nilsson S, Finnström B, Kokinsky E, Enskär K. The use of virtual reality for needle-related procedural pain and distress in children and adolescents in a paediatric oncology unit. Eur J Oncol Nurs 2009 Apr;13(2):102-109. [doi: 10.1016/j.ejon.2009.01.003] [Medline: 19230769]
32. Rao DG, Havale R, Nagaraj M, et al. Assessment of efficacy of virtual reality distraction in reducing pain perception and anxiety in children aged 6-10 years: a behavioral interventional study. Int J Clin Pediatr Dent 2019;12(6):510-513. [doi: 10.5005/jp-journals-10005-1694] [Medline: 32440065]
33. Gorman D, Hoermann S, Lindeman RW, Shahri B. Using virtual reality to enhance food technology education. Int J Technol Des Educ 2022;32(3):1659-1677. [doi: 10.1007/s10798-021-09669-3] [Medline: 33976475]
34. Graeske C, Sjöberg SA. VR-technology in teaching: opportunities and challenges. Int Educ Stud 2021;14(8):76-83. [doi: 10.5539/ies.v14n8p76]
35. Makransky G, Andreasen NK, Baceviciute S, Mayer RE. Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. J Educ Psychol 2021;113(4):719-735. [doi: 10.1037/edu0000473]
36. Wang QJ, Barbosa Escobar F, Alves Da Mota P, Velasco C. Getting started with virtual reality for sensory and consumer science: current practices and future perspectives. Food Res Int 2021 Jul;145:110410. [doi: 10.1016/j.foodres.2021.110410] [Medline: 34112413]
37. Fang D, Nayga RM Jr, West GH, et al. On the use of virtual reality in mitigating hypothetical bias in choice experiments. American J Agri Economics 2021 Jan;103(1):142-161. [doi: 10.1111/ajae.12118]
38. Ouellet É, Boller B, Corriveau-Lecavalier N, Cloutier S, Belleville S. The virtual shop: a new immersive virtual reality environment and scenario for the assessment of everyday memory. J Neurosci Methods 2018 Jun 1;303:126-135. [doi: 10.1016/j.jneumeth.2018.03.010] [Medline: 29581009]
39. Allman-Farinelli M, Ijaz K, Tran H, et al. A virtual reality food court to study meal choices in youth: design and assessment of usability. JMIR Form Res 2019 Jan 9;3(1):e12456. [doi: 10.2196/12456] [Medline: 30684440]
40. Celikcan U, Bülbül A, Aslan C, et al. The virtual cafeteria: an immersive environment for interactive food portion-size education. Presented at: MHFI'18: Proceedings of the 3rd International Workshop on Multisensory Approaches to Human-Food Interaction; Oct 16, 2018. [doi: 10.1145/3279954.3279960]
41. Lombart C, Millan E, Normand JM, Verhulst A, Labbé-Pinlon B, Moreau G. Effects of physical, non-immersive virtual, and immersive virtual store environments on consumers' perceptions and purchase behavior. Comput Human Behav 2020 Sep;110:106374. [doi: 10.1016/j.chb.2020.106374]
42. Bucher T, van der Horst K, Siegrist M. The fake food buffet - a new method in nutrition behaviour research. Br J Nutr 2012 May;107(10):1553-1560. [doi: 10.1017/S000711451100465X] [Medline: 21920063]
43. Fernandes AC, Oliveira RC, Proença RPC, Curioni CC, Rodrigues VM, Fiates GMR. Influence of menu labeling on food choices in real-life settings: a systematic review. Nutr Rev 2016 Aug;74(8):534-548. [doi: 10.1093/nutrit/nuw013] [Medline: 27358442]
44. McBride CM, Persky S, Wagner LK, Faith MS, Ward DS. Effects of providing personalized feedback of child's obesity risk on mothers' food choices using a virtual reality buffet. Int J Obes (Lond) 2013 Oct;37(10):1322-1327. [doi: 10.1038/ijo.2013.87] [Medline: 23736369]
45. Smit ES, Meijers MHC, van der Laan LN. Using virtual reality to stimulate healthy and environmentally friendly food consumption among children: an interview study. Int J Environ Res Public Health 2021 Jan 26;18(3):1088. [doi: 10.3390/ijerph18031088] [Medline: 33530495]
46. Karkar A, Salahuddin T, Almaadeed N, Aljaam JM, Halabi O. A virtual reality nutrition awareness learning system for children. Presented at: 2018 IEEE Conference on e-Learning, e-Management and e-Services (IC3e); Nov 21-22, 2018. [doi: 10.1109/IC3e.2018.8632634]
47. Hoang TN, Reinoso M, Vetere F, Tanin E. Onebody: remote posture guidance system using first person view in virtual environment. Presented at: NordiCHI '16: Proceedings of the 9th Nordic Conference on Human-Computer Interaction; Oct 23-27, 2016. [doi: 10.1145/2971485.2971521]
48. Baker S, Kelly RM, Waycott J, et al. Interrogating social virtual reality as a communication medium for older adults. Proc ACM Hum-Comput Interact 2019 Nov 7;3(CSCW):1-24. [doi: 10.1145/3359251]
49. Mottelson A, Muresan A, Hornbæk K, Makransky G. A systematic review and meta-analysis of the effectiveness of body ownership illusions in virtual reality. ACM Trans Comput Hum Interact 2023 Oct 31;30(5):1-42. [doi: 10.1145/3590767]
50. Keenaghan S, Polaskova M, Thurlbeck S, Kentridge RW, Cowie D. Alice in Wonderland: the effects of body size and movement on children's size perception and body representation in virtual reality. J Exp Child Psychol 2022 Dec;224:105518. [doi: 10.1016/j.jecp.2022.105518] [Medline: 35964343]
51. de Klerk CCJM, Filippetti ML, Rigato S. The development of body representations: an associative learning account. Proc Biol Sci 2021 Apr 28;288(1949):20210070. [doi: 10.1098/rspb.2021.0070] [Medline: 33906399]

XSL•FO
RenderX

52. Weijs ML, Macartney E, Daum MM, Lenggenhager B. Development of the bodily self: effects of visuomotor synchrony and visual appearance on virtual embodiment in children and adults. J Exp Child Psychol 2021 Oct;210:105200. [doi: 10.1016/j.jecp.2021.105200] [Medline: 34116407]
53. Cowie D, Sterling S, Bremner AJ. The development of multisensory body representation and awareness continues to 10 years of age: evidence from the rubber hand illusion. J Exp Child Psychol 2016 Feb;142:230-238. [doi: 10.1016/j.jecp.2015.10.003] [Medline: 26601752]
54. Bailey JO, Schloss I. “Awesomely freaky!” The impact of type on children’s social-emotional perceptions of virtual reality characters. Presented at: CHI ’23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems; Apr 23-28, 2023; Hamburg, Germany p. 1-10. [doi: 10.1145/3544548.3581501]
55. Liao T, Jennings NA, Dell L, Collins C. Could the virtual dinosaur see you? Understanding children’s perceptions of presence and reality distinction in virtual reality environments. JVWR 2019;12(2). [doi: 10.4101/jvwr.v12i2.7361]
56. Cooper SL, Butcher LM, Scagnelli SD, et al. Australian consumers are willing to pay for the health star rating front-of-pack nutrition label. Nutrients 2020 Dec 18;12(12):3876. [doi: 10.3390/nu12123876] [Medline: 33352995]
57. For shoppers. Health Star Rating System. URL: https://www.healthstarrating.gov.au/shoppers [accessed 2025-11-24]
58. Trumbo P, Schlicker S, Yates AA, Poos M, Food and Nutrition Board of the Institute of Medicine, The National Academies. Dietary reference intakes for energy, carbohydrate, fiber, fat, fatty acids, cholesterol, protein and amino acids. J Am Diet Assoc 2002 Nov;102(11):1621-1630. [doi: 10.1016/s0002-8223(02)90346-9] [Medline: 12449285]
59. Hasni MJS, Rehman MA, Pontes N, Yaqub MZ. Health star rating labels: a systematic review and future research agenda. Food Qual Prefer 2025 Jan;122:105310. [doi: 10.1016/j.foodqual.2024.105310]
60. da Quinta N, Ríos Y, Llorente R, Naranjo AB, Baranda AB. The effect of food shape on children’s implicit and explicit emotional response. Food Qual Prefer 2023 Jul;109:104921. [doi: 10.1016/j.foodqual.2023.104921]
61. Russell A, Jansen E, Burnett AJ, Lee J, Russell CG. Children’s eating behaviours and related constructs: conceptual and theoretical foundations and their implications. Int J Behav Nutr Phys Act 2023 Feb 15;20(1):19. [doi: 10.1186/s12966-023-01407-3] [Medline: 36793039]
62. Mai R, Trendel O, Basil M. Food categorization determines whether healthier food is inferred to be tastier or less tasty. J Consum Psychol 2025 Oct;35(4):531-550. [doi: 10.1002/jcpy.1447]
63. Spanlang B, Normand JM, Borland D, et al. How to build an embodiment lab: achieving body representation illusions in virtual reality. Front Robot AI 2014;1. [doi: 10.3389/frobt.2014.00009]
64. Rogers Y. Interaction design gone wild. Interactions 2011 Jul;18(4):58-62. [doi: 10.1145/1978822.1978834]
65. Schwind V, Knierim P, Haas N, Henze N. Using presence questionnaires in virtual reality. Presented at: CHI ’19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland p. 1-12. [doi: 10.1145/3290605.3300590]
66. Gonzalez-Franco M, Peck TC. Avatar embodiment. Towards a standardized questionnaire. Front Robot AI 2018 Jun 22;5:1-9. [doi: 10.3389/frobt.2018.00074]
67. Balk SA, Bertola DB, Inman VW. Simulator Sickness Questionnaire: twenty years later. Presented at: Proceedings of the 7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design; Jun 17-20, 2013 p. 257-263. [doi: 10.17077/drivingassessment.1498]
68. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator Sickness Questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviat Psychol 1993 Jul;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
69. van Laerhoven H, van der Zaag-Loonen HJ, Derkx BHF. A comparison of Likert scale and visual analogue scales as response options in children’s questionnaires. Acta Paediatr 2004 Jun;93(6):830-835. [doi: 10.1080/08035250410026572] [Medline: 15244235]
70. Kundel HL, Polansky M. Measurement of observer agreement. Radiology 2003 Aug;228(2):303-308. [doi: 10.1148/radiol.2282011860] [Medline: 12819342]
71. Clarke V, Braun V. Thematic analysis. J Posit Psychol 2017 May 4;12(3):297-298. [doi: 10.1080/17439760.2016.1262613]
72. Creswell JW, Plano Clark VL, Gutmann ML, Hanson WE. Advanced mixed methods research designs. In: Handbook of Mixed Methods in Social and Behavioral Research: SAGE Publications; 2003:209-240.
73. Butterfield SA, Angell RM, Mason CA. Age and sex differences in object control skills by children ages 5 to 14. Percept Mot Skills 2012 Feb;114(1):261-274. [doi: 10.2466/10.11.25.PMS.114.1.261-274] [Medline: 22582694]
74. Cowie D, McKenna A, Bremner AJ, Aspell JE. The development of bodily self-consciousness: changing responses to the Full Body Illusion in childhood. Dev Sci 2018 May;21(3):e12557. [doi: 10.1111/desc.12557] [Medline: 28326654]
75. Rzepka AM, Hussey KJ, Maltz MV, Babin K, Wilcox LM, Culham JC. Familiar size affects perception differently in virtual reality and the real world. Philos Trans R Soc Lond B Biol Sci 2023 Jan 30;378(1869):20210464. [doi: 10.1098/rstb.2021.0464] [Medline: 36511414]
76. Keung CPC, Fung CKH. Pursuing quality learning experiences for young children through learning in play: how do children perceive play? Early Child Dev Care 2021 Mar 12;191(4):583-597. [doi: 10.1080/03004430.2019.1633313]
77. Southgate E, Smith SP, Scevak J. Asking ethical questions in research using immersive virtual and augmented reality technologies with children and youth. Presented at: 2017 IEEE Virtual Reality (VR); Mar 18-22, 2017; Los Angeles, CA p. 12-18. [doi: 10.1109/VR.2017.7892226]

XSL•FO
RenderX

78. Yang CL, Matsumoto K, Yu S, et al. Understanding the effect of a virtual moderator on people's perception in remote discussion using social VR. Front Virtual Real 2023;4. [doi: 10.3389/frvir.2023.1198024]
79. Sedgwick P, Greenwood N. Understanding the Hawthorne effect. BMJ 2015 Sep 4;351:h4672. [doi: 10.1136/bmj.h4672] [Medline: 26341898]
80. Wickström G, Bendix T. The "Hawthorne effect"--what did the original Hawthorne studies actually show? Scand J Work Environ Health 2000 Aug;26(4):363-367. [Medline: 10994804]
81. Slater M. How colorful was your day? Why questionnaires cannot assess presence in virtual environments. Presence (Camb) 2004 Aug 1;13(4):484-493. [doi: 10.1162/1054746041944849]
82. Yang Y, Guo M, Wu Y, et al. VR MRI training for adolescents: a comparative study of gamified VR, passive VR, 360° video, and traditional educational video. Presented at: 2025 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct); Oct 8-12, 2025. [doi: 10.1109/ISMAR-Adjunct68609.2025.00081]
83. Bodrova E, Leong DJ. The importance of being playful. Educ Leadersh 2003;60:50-53.
84. DeLoache JS, Gulliver UDH. Gulliver, goliath and goldilocks: young children and scale errors. In: Slaughter V, Brownell CA, editors. Early Development of Body Representations: Cambridge University Press; 2011:59-68. [doi: 10.1017/CBO9781139019484.005]
85. Lillard AS, Woolley JD. Grounded in reality: how children make sense of the unreal. Cogn Dev 2015;34:111-114. [doi: 10.1016/j.cogdev.2014.12.007]
86. Staples KL, MacDonald M, Zimmer C. Assessment of motor behavior among children and adolescents with autism spectrum disorder. In: International Review of Research in Developmental Disabilities: Elsevier Academic Press; 2012:179-214. [doi: 10.1016/B978-0-12-394284-5.00007-3]
87. Piaget J. Piaget's theory. In: Inhelder B, Chipman HH, Zwingmann C, editors. Piaget and His School: A Reader in Developmental Psychology: Springer; 1976:11-23. [doi: 10.1007/978-3-642-46323-5_2]
88. Ogawa N, Narumi T, Hirose M. Object size perception in immersive virtual reality: avatar realism affects the way we perceive. Presented at: 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR); Mar 18-22, 2018. [doi: 10.1109/VR.2018.8446318]
89. Slater M, Perez-Marcos D, Ehrsson HH, Sanchez-Vives MV. Towards a digital body: the virtual arm illusion. Front Hum Neurosci 2008;2:6. [doi: 10.3389/neuro.09.006.2008] [Medline: 18958207]
90. Barnett LM, Mazzoli E, Hawkins M, et al. Development of a self-report scale to assess children's perceived physical literacy. Phys Educ Sport Pedagogy 2022 Jan 2;27(1):91-116. [doi: 10.1080/17408989.2020.1849596]

Abbreviations

HCI: human-computer interaction
RL: real-life
RQ: research question
VR: virtual reality

Edited by A Choudhury; submitted 01.Dec.2025; peer-reviewed by E Hohman, Y Yang; revised version received 20.Mar.2026; accepted 22.Apr.2026; published 02.Jul.2026.

Please cite as:
Aggarwal D, Hoang T, Tan SY, Mohebbi M, Russell CG
Evaluating the Viability of Virtual Reality for Children's Food Choice Research: Comparative Mixed Methods Study
JMIR Hum Factors 2026;13:e88476
URL: https://humanfactors.jmir.org/2026/1/e88476
doi:10.2196/88476

© Deepti Aggarwal, Thuong Hoang, Sze-Yen Tan, Mohammadreza Mohebbi, Catherine G Russell. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 2.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Challenges and Benefits of Virtual Reality in Home-Based Poststroke Rehabilitation: Co-Design Qualitative Study

Trust Saidi[1], PhD; Ann Marie Hestetun-Mandrup[2,3], MSc; Nenad Pavel[4], PhD; Ingvild Kristina Hurum Rosseland[5], MBA; Kathinka Granum Selmer-Olsen[5], BSc; Nora Synnøve Mørk[5], MHS; Åse Bergheim[5], MSc; Minna Annika Pikkarainen[1,2], PhD

[1]Department of Product Design, Faculty of Technology, Art and Design, OsloMet – Oslo Metropolitan University, Oslo, Norway

[2]Department of Rehabilitation and Health Technology, Faculty of Health Sciences, OsloMet – Oslo Metropolitan University, Oslo, Norway

[3]Sunnaas Rehabilitation Hospital, Norway, Bjørnemyr, Norway

[4]Department of Product Design, Faculty of Technology, Art and Design, OsloMet – Oslo Metropolitan University, OSLO, Norway

[5]Forsterket rehabilitering Aker, Helseetaten Oslo Kommune, Oslo, Norway, Oslo Municipality, Aker Hospital, Oslo, Norway

Corresponding Author:
Trust Saidi, PhD
Department of Product Design
Faculty of Technology, Art and Design
OsloMet – Oslo Metropolitan University
PO Box 4
St. Olavs plass
Oslo, 0130
Norway
Phone: 47 48507071
Email: trust.saidi@oslomet.no

Abstract

Background: Stroke often leads to persistent impairments that limit daily functioning and psychosocial well-being. Virtual reality (VR) has emerged as a promising adjunct in stroke rehabilitation, although research has largely focused on clinical outcomes, with limited attention to user involvement and the experiences of multiple stakeholders in the design process.

Objective: This study aimed to explore the challenges and benefits of co-designing and using VR to support home-based poststroke rehabilitation.

Methods: A qualitative co-design case study was applied to gain an in-depth understanding of the challenges and benefits. Rapid co-design principles were used in developing VR prototypes delivered via head-mounted displays through 3 workshops with participants with stroke, health care professionals, and VR developers from November 2023 to May 2024. A design prototype revision was done based on feedback from the workshops. Data were collected via audio-taped co-design workshops with all participants and 10 successive semistructured interviews with health care professionals and VR developers conducted in a rehabilitation hospital. A thematic analysis was performed on transcribed recordings.

Results: In total, five main themes emerged: (1) adaptability for stroke-related impairments in home rehabilitation, (2) safety and ease of use, (3) goal orientation, (4) motivation, and (5) VR as a complementary tool. One of the primary challenges identified lies in the adaptability of VR systems for individuals with hemiparesis. Additionally, customization and safety concerns remain a complex barrier, as VR solutions must be capable of addressing a wide range of stroke-related impairments and aligning specific rehabilitation goals. VR demonstrated potential to enhance rehabilitation by simulating real-life tasks that encourage goal-oriented and motivating therapy. As a complementary tool, VR can enhance traditional rehabilitation by increasing the intensity and volume of therapy.

Conclusions: This study offers insight into how VR can be effectively integrated into rehabilitation practices. Its integration into rehabilitation requires alignment with established therapeutic principles within VR applications, such as adjustable task-specific training and meaningful outcomes tailored to individual needs, to ensure clinical relevance and user engagement. VR should complement, rather than replace, conventional therapy by increasing training intensity, reducing therapist workload, and extending rehabilitation into the home. Thoughtful co-design with stakeholders is key to creating VR tools that bridge the gap between structured clinical care and independent recovery, offering continuous support throughout the rehabilitation process.

XSL•FO
RenderX

(JMIR Hum Factors 2026;13:e78798) doi:10.2196/78798

KEYWORDS

virtual reality; rehabilitation; people after stroke; co-design; stroke

Introduction

Stroke ranks as the second leading cause of death globally and is a major contributor to adults' neurological and neuropsychological persistent impairments [1], impacting 17 million individuals annually [2]. This is a cause for concern as the demand for poststroke rehabilitation is expected to rise significantly [3], imposing considerable strains on both the quality of life of people after stroke and health care resources. Rehabilitation for people after stroke often requires a long period of rehabilitation and motor relearning [4]. Artificial intelligence (AI)–driven virtual reality (VR) solutions are expected to support individuals' self-managed rehabilitation and have the potential to effectively engage individuals in intensive, repetitive, and task-oriented activities [5]. They can serve as useful tools for monitoring progress and enhancing decision-making in the rehabilitation process for people after stroke and health care professionals [6].

Stroke can affect individuals to varying degrees, with motor impairments being the most frequent disability. These impairments can limit motor mobility and have a negative impact on individuals' physical activity levels and psychosocial well-being [7]. A European multicenter cohort study demonstrated that after 5 years, people after stroke experienced a decline in functional and motor outcomes, reverting to the levels observed at 2 months post stroke, highlighting that many individuals experience residual impairments [8]. Motor disabilities can vary from slight weakness to severe paralysis, impacting one's ability to perform everyday activities like eating, cooking, and dressing independently. It is estimated that about 75% of people after stroke initially face arm impairment, with only about 50% regaining arm function within 6 months after their stroke [9,10]. Although there are several studies on the effectiveness and use of VR, with some targeting specifically upper limb rehabilitation [11-13], they are mostly systematic reviews. The empirical studies on this topic are mainly based on the evaluation of the effects of VR in the rehabilitation process and not individual experiences [14-18]. While both the systematic reviews and evaluations on the effect of VR have enhanced our understanding of the benefits of these interventions, there is a paucity of knowledge on how these positive outcomes are achieved.

Experiencing a stroke often affects more than just physical activities and functions; it also impacts cognition and personal factors. The impact on life after stroke often leaves individuals feeling vulnerable and anxious about what lies ahead [19,20]. Cognitive challenges, including difficulties with memory, attention, and problem-solving, are common, as are emotional effects, such as mood swings, depression, and anxiety [21]. VR technologies targeting cognitive function and stress management have also been found to be effective [22]; however, a recognized gap in the literature exists regarding the effect that VR has on psychological well-being, which potentially affects cognitive rehabilitation. Therefore, greater attention is needed regarding how individuals recovering from stroke engage with VR solutions. This paper seeks to contribute to the growing body of knowledge in this area. Stroke rehabilitation is often a prolonged and challenging process, requiring long-term therapeutic intervention. A study by Johansen et al [23] examining the use of VR equipment in home settings for individuals with brain injuries highlighted the ongoing need for cognitive rehabilitation following hospital discharge. Participants in the study emphasized the importance of individualized VR interventions and noted that initial training typically occurred within hospital environments. These findings suggest that exploring the early implementation of VR in controlled clinical settings, facilitated through close collaboration with health care professionals, may offer valuable insights into the feasibility and effectiveness of VR-based rehabilitation post stroke.

The gaps identified above underscore the importance of understanding how VR technology can be both designed for and experienced by users within a safe and supportive environment, one that allows for a comprehensive exploration of its potential benefits as well as its inherent challenges. This demands a focus on the process of engaging the users rather than merely evaluating the effects. Focusing on the effects without elaborating on the development process poses the danger of oversimplifying the mechanisms through which VR contributes to recovery. Without a clear understanding of the underlying processes, it becomes challenging to optimize and innovate VR interventions to maximize their rehabilitation potential. Given the limited research on the use of immersive VR for stroke rehabilitation involving clinicians, people after stroke, and VR developers, it is crucial to integrate a multidisciplinary co-design approach to ensure that clinical VR solutions are both effective and user-centered [24-26]. By involving individuals with stroke, developers, and multidisciplinary health care professionals early in the design process, we can balance desired features with evidence-based design recommendations, enhancing the relevance and usability of these digital technologies. This approach also helps prevent the development of solutions that fail to meet actual treatment needs or pose safety risks to people after stroke when using VR solutions independently at home.

Even though VR systems promise more engaging experiences, especially when it comes to repetitive movements in poststroke rehabilitation, it is still not a go-to practice for many people after stroke and therapists. This study, therefore, aims to address the identified research gaps by exploring the challenges and benefits of co-designing and using VR to support home-based poststroke rehabilitation.

XSL•FO
RenderX

Methods

Overview

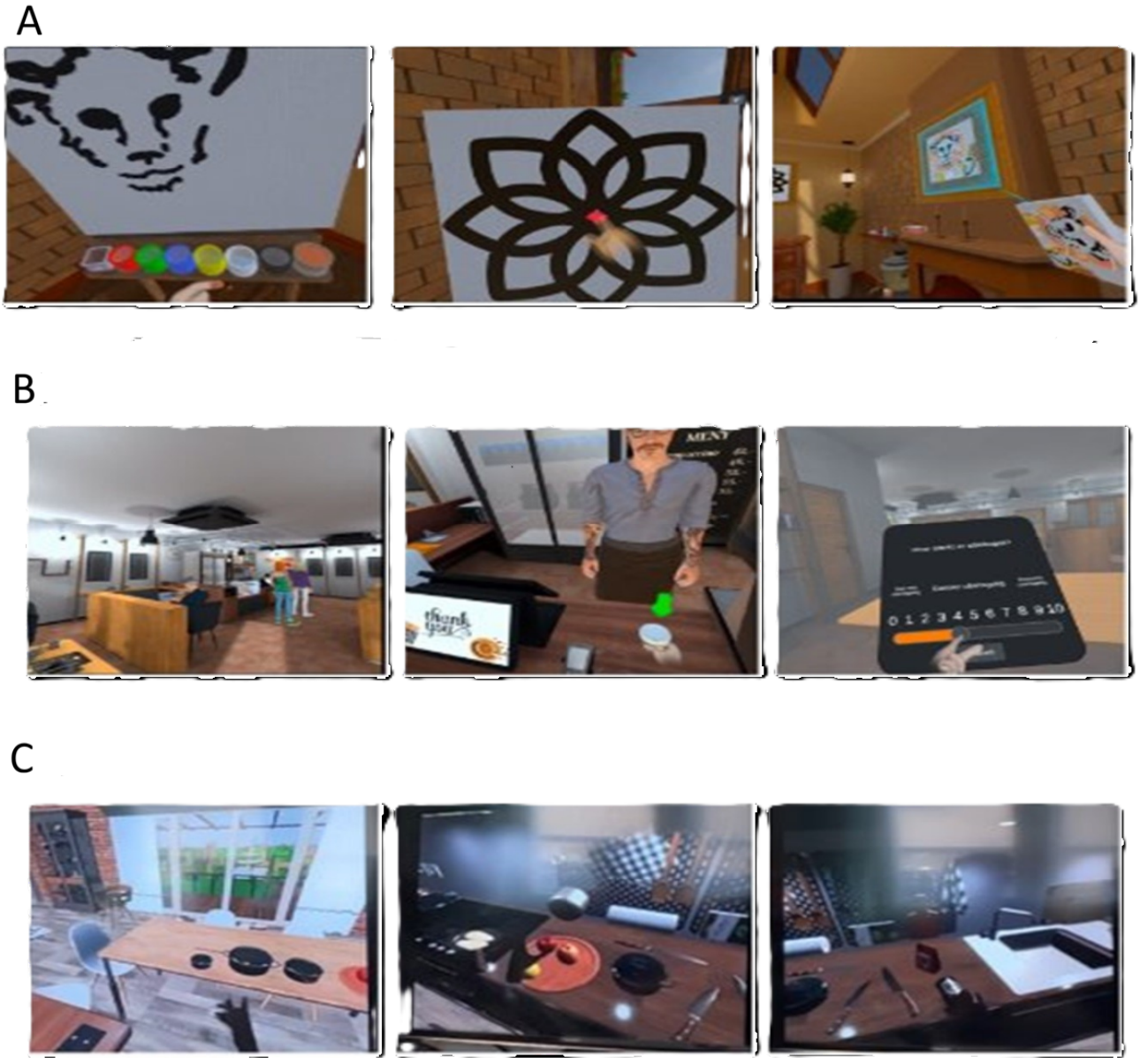
A qualitative co-design case study was conducted to gain an in-depth understanding of the real-life phenomenon within its environmental context [27]. A single case study approach was selected as it provided a robust means to explore "how" and "why" questions within a specific context, which gave the opportunity to investigate deeper causes of the phenomenon [28]. The data were collected through different sources of semistructured interviews and workshops.

Design Process Using Rapid Prototyping

We applied design principles and rapid prototyping [29] to develop VR prototypes (Figure 1) through 3 workshops with a mixture of participants held between November 2023 to May 2024 with the goal of (1) understanding the challenges and map the needs of people after stroke using an existing VR café scenario; (2) discuss possibilities and evaluate features in kitchen and a painting scenarios; and (3) test, evaluate, and refine the modified prototypes of the kitchen and painting scenarios through discussions and user feedback. The prototypes were immersive VR environments delivered via head-mounted displays (Oculus Quest 3 [Reality Labs]), featuring AI-driven interactive virtual characters, while other scenarios used scripted or noninteractive virtual elements with hand tracking.

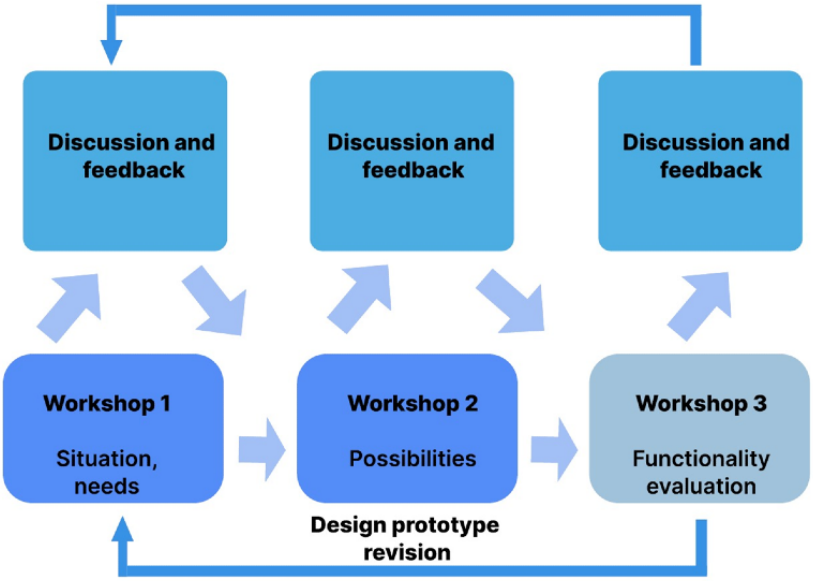
To explore the potential of immersive environments in stroke rehabilitation, 3 VR scenarios, that is, a painting studio, a café, and a kitchen (Figure 2), were designed to simulate familiar and meaningful everyday settings. These environments were purposefully developed to promote creative engagement and functional interaction, allowing participants to practice activities reflecting real-life contexts in a safe and controlled virtual space. The 3 scenarios are illustrated in Figure 1.

Figure 1. Virtual reality scenarios presented to people after stroke (n=4) during stakeholder workshops. (A) Scenario 1: Painting, (B) Scenario 2: Café, and (C) Scenario 3: Kitchen.



XSL•FO
RenderX

Figure 2. Flowchart illustrating the design workshops involving people after stroke, health care professionals, virtual reality developers, and researchers (n=16) to iteratively develop and refine a virtual reality prototype for poststroke rehabilitation.



The first scenario illustrates the painting environment, where the participants entered a living room and moved toward the drawing board, where equipment and colors were located. The participants could choose between freehand painting and tracing a preset illustration or figure. Guidelines for each activity were provided in all scenarios. The second scenario placed participants in a large café environment. They moved toward the cashier's desk, ordered their food, and paid using a tablet. Participants also had the opportunity to interact with other VR avatars in the café. The third scenario illustrates a kitchen environment, where the participants could move between the kitchen counter and the dining table. They could do different activities like setting the table, placing kitchen utensils in drawers, or putting apples in a bucket.

In the first workshop, painting and café scenarios were presented to people after stroke and health care professionals. For the subsequent workshops, only painting and kitchen scenarios were further developed, as these were identified as most relevant and useful for people after stroke undergoing home rehabilitation, particularly for enhancing physical and cognitive functions. The co-design process was led by 2 researchers (MAP and NP)—a professor and an associate professor—who brought insider knowledge about the co-designing and feature-adjusting phenomena. Furthermore, 2 additional researchers (TS and AMHM), a postdoctoral and doctoral candidate, provided an outsider view to challenge assumptions. This single case approach allowed researchers to follow the co-design development process and adaptation of VR systems to the needs and demands of people after stroke, health care professionals, and VR developers.

Each workshop began with a short introduction to the aim of the VR scenarios, followed by people after stroke testing the solutions, and finally followed by 2 discussion sessions with all stakeholders and 1 exclusively with health care professionals. Feedback from the workshops informed iterative revisions of the prototype, as illustrated in Figure 2.

Recruitment of Participants

The study was conducted in a rehabilitation hospital where participants were recruited to participate in the workshops from three distinct groups where the number of unique participants was: (1) health care professionals (n=10), (2) people after stroke (n=4), and (3) VR developers (n=2). All participants received previous written information and were invited to participate in both the workshops and follow-up interviews via phone or email. Health care professionals were recruited by the unit leader, while people after stroke were selected based on their involvement in ongoing rehabilitation programs within the hospital. The VR developers were already involved in organizing the workshops as part of their collaboration with the research team. They agreed to participate in interviews after the workshops were concluded. During the workshops, all 3 groups were asked if they were willing to participate in follow-up interviews. A total of 10 participants agreed, including health care professionals and VR developers. Additionally, 3 out of 4 participants with stroke agreed to engage in short conversations and provided feedback during VR testing. The 4 participants were recruited from a community-funded rehabilitation hospital in Norway based on the following inclusion and exclusion criteria: adults with a confirmed stroke diagnosis, speaking either English or Norwegian, and living at home in the subacute or chronic phase (more than 3 months since the last stroke) as defined by a stroke trajectory framework [30]. In addition, the participants needed to have a score of 18 or above on the Montreal Cognitive Assessment (MoCA) [31] and the ability to complete specific tasks from the Action Research Arm Test [32,33], such as pouring water from one glass to another and placing a hand on top of the head. Exclusion criteria included individuals younger than 18 years, pregnant women, and people with receptive or expressive aphasia or memory and communication impairments that hindered them from using the prototypes. The 4 participants with stroke had a balanced gender distribution and ranged in age from 53 to 64 years. The participants experienced different types of strokes; 2 experienced

XSL•FO
RenderX

ischemic strokes, while the other 2 experienced hemorrhages (1 subarachnoid and 1 brainstem). None of the participants with stroke had previous experience with VR. Additionally, several health care professionals with expertise in stroke rehabilitation (ie, physiotherapists, occupational therapists, nurses, and medical doctors) and VR developers participated. The stakeholders, consisting of people after stroke, health care professionals, and VR developers, were involved in all 3 workshops as illustrated in Table 1.

Table 1. Overview of the participants excluding the researchers involved in the workshops (n=16) and interviews (n=10).

Stakeholder	Workshop attendance	Individual interviews
Researchers	• Workshop 1: n=3 • Workshop 2: n=3 • Workshop 3: n=4 • Total number of unique participants: n=4	—[a]
People after stroke	• Workshop 1: n=3 • Workshop 2: n=3 • Workshop 3: n=3 • Total number of unique participants: n=4	—
Health care professionals	• Workshop 1: n=8 • Workshop 2: n=5 • Workshop 3: n=10 • Total number of unique participants: n=10	Number of interviewees=8 (4 occupational therapists, 2 physiotherapists, 1 medical doctor, and 1 nurse)
VR[b] technology developers	• Workshop 1: n=2 • Workshop 2: n=2 • Workshop 3: n=2 • Total number of unique participants: n=2	Number of interviewees=2 (VR developers)

[a]Not applicable.

[b]VR: virtual reality.

Data Collection

Audio recordings from each of the 3- to 4-hour co-design workshops [34] and collaboratively developed individual interviews conducted at a rehabilitation hospital served as the primary data sources for this case study. A total of 10 semistructured interviews, lasting 60-90 minutes each, were conducted and audio-recorded. Open-ended questions were used in both workshops and interviews (refer to interview guide in Multimedia Appendix 1). All meetings were conducted face-to-face. Workshop 3, being both comprehensive and advanced, built upon the outcomes of Workshops 1 and 2 and served as the primary data source for this study. Both workshops and interviews were transcribed and anonymized, with no real names or personal information retained from participants, to ensure confidentiality and adhere to ethical research standards. Researchers from Oslo Metropolitan University, led by designers with experience in design-driven innovation, user experience design, and digital prototyping (MAP and NP), mediated the workshops, while clinicians provided support to participants with stroke during the testing. Interviews and initial analysis were conducted by 3 experienced qualitative researchers (AMHM, TS, and NP) from Oslo Metropolitan University, with backgrounds in physiotherapy, science and technology studies, and computer and design science. Tasks were assigned based on relevant expertise; for instance, Norwegian-language interviews were conducted by fluent researchers, while interviews with VR developers were handled by a researcher specializing in the science, technology, and society interface.

The decision to conclude data collection after 10 interviews was guided by the concept of information power [35], which asserts that sample adequacy in qualitative research depends on the relevance and richness of the data in relation to the study's aim. Given the focused objective of gathering actionable feedback for iterative refinement of VR prototypes rather than achieving full thematic saturation, a small, purposefully selected sample was appropriate. This approach aligns with rapid prototyping methodology, which prioritizes targeted insights to inform design decisions over exhaustive theme development [36]. The study also followed participatory research principles, prioritizing a representative sample of end users to ensure relevance and applicability. The sample size (n=10) also aligns with Creswell's [37] recommended range of 5-25 participants for phenomenological studies. Finally, triangulation [38] using audio recordings from stakeholder workshops and individual interviews enhanced the credibility and depth of the findings.

Data Analysis

The data were transcribed verbatim using the transcription software Whisper (OpenAI) and manually checked for accuracy. Thematic analysis (TA) [39] was used to generate in-depth insights from participants' discussions. A codebook-based TA approach, involving multiple coders (AMHM, TS, and NP), was used to enhance the credibility and consistency of the findings [40]. The primary analysis was conducted by 3 researchers, all of whom were experienced in conducting qualitative research and applying TA as an analytical method. Further analysis was conducted collaboratively by all authors of the study and validated by health care professionals to ensure

XSL•FO
RenderX

the study addressed both theoretical and clinical aspects. The analysis followed Braun and Clarke's 6-phase framework, beginning with familiarization through repeated readings of the data to gain a comprehensive understanding. Subsequently, inductive and descriptive codes were applied, which were then grouped into preliminary themes. These initial themes were collaboratively reviewed and refined by the research team to ensure consensus about their accuracy and reliability. The processes of defining, naming the themes, and reporting the findings were conducted collectively.

To illustrate the development of the theme "safety and ease of use" using Braun and Clarke's 6-phase framework, we present an example from the analysis process. In the familiarization phase, repeated readings of transcripts revealed frequent concerns about physical safety, navigation, and psychological comfort in VR. Initial coding grouped statements about uncertainty, environmental awareness, and support needs under relevant codes, which were later collated into the broader theme. In the reviewing phase, we ensured this theme was coherent and differentiated from other related themes like "adaptability for stroke-related impairments," which focused more on task customization. "Safety and ease of use" emphasized risk minimization and user confidence. During theme definition and naming, we finalized it to reflect participants' emphasis on secure, intuitive, and supportive VR environments. Final reporting included selecting illustrative quotes to demonstrate how safety concerns informed both design recommendations and user engagement.

Rigor

Rigor was obtained through frequent debriefing and discussions among coders during data analysis and collaborative theme development within the research team. The qualitative data obtained from audio recordings of the workshops and individual interviews enabled triangulation and cross-references between findings [40], which strengthened the credibility of the results. The theoretical flexibility that TA offers was well-suited to accommodate the diverse perspectives and knowledge production of the multidisciplinary research team [41], which in this study included specialists in allied health, computer science, and design.

Ethical Considerations

The Norwegian National Research Ethics Committee for Medical and Health Research assessed the study and determined the need for ethical approval. They declared that this study's focus on health service research fell outside the scope of the Health Research Act § 2 and therefore did not require approval (ref 651236). Instead, this study was approved by the Norwegian Agency for Shared Services in Education and Research (ref 857865) to ensure compliance with privacy protection regulations. Participants received both oral and written information, and informed consent was obtained before the workshops and individual interviews. To protect the privacy of the participants, the study data were deidentified by using pseudonyms for all participants. Participants did not receive any financial or material compensation for their participation.

Results

Overview

Different participants with stroke, multidisciplinary health care professionals, VR developers, and researchers were present and participated in all workshops. There was a variety among the stroke individuals who were engaging in their first, second, or third workshop with the VR scenarios. Although engaging with avatars and objects in the café scenario was engaging, the stakeholders believed it did not target the rehabilitation principles. In the second workshop, this was amended by engaging in the kitchen and painting scenario. In the third workshop, these 2 prototypes were further developed by gamifying certain arm movements, targeting the affected arm in particular. The following results are based on the participants' needs and key challenges and benefits that were generated in workshops 1 and 2, and applied in workshop 3, which form the bases of this study, together with retrospective individual interviews with health care professionals. The findings generated five themes: (1) adaptability to accommodate stroke-related impairments in home rehabilitation, (2) safety and ease of use as fundamental in VR stroke rehabilitation, (3) goal orientation, (4) motivation, and (5) VR as a complementary tool.

Adaptability to Accommodate Stroke-Related Impairments in Home Rehabilitation

Participants viewed VR as a promising tool for poststroke home rehabilitation, offering interactive environments to aid motor recovery. However, they stressed that its usability depended on adapting to stroke-related impairments, such as accommodating both right- and left-handed hemiparetic users, as the first workshop primarily targeted right-hand users.

Customization was highlighted as crucial to address diverse needs, including issues like putting on VR glasses, gripping tasks, and aligning activities with rehabilitation principles. Participants stressed the importance of personalizing rehabilitation, as each person after a stroke faces unique challenges. Effective poststroke rehabilitation requires balancing engagement and endurance, as cognitive load and fatigue often limit people after a stroke. VR therapy demands motor and mental skills, making it vital to tailor intensity and duration to each individual's ability. It was noted that VR could overwhelm people after stroke, limiting its effectiveness.

> *When it comes to effectiveness, dosage and repetitions, VR will take more mental capacity because you use cognitive functions alongside motor function. Many people with stroke have fatigue; some can only do it for 10 minutes before their battery is low.* [Karla]

The potential for AI-powered VR solutions to enhance customization was discussed, with many participants emphasizing the rehabilitation goal of fostering independence and regaining agency. However, participants cautioned that VR was not suitable for everyone. Discrepancies between VR command techniques of gripping and actual abilities for people after stroke emerged during testing:

XSL•FO
RenderX

Therapist: You can also try to lift the kettle and the red bag there.

Person after stroke, first-time user: I try with the left hand first.

Person after stroke, first-time user: Just press through, sort of?

VR developer: Yes. Use the whole thing. I had to practice a bit with gripping, because I thought I could grip with my fingers.

VR developer: You have to use your whole hand, sort of, to grip.

Therapist: Then you try to open your hand, and then... Like that, yes!

Person after stroke, first-time user: Incredible.

Participants also noted errors between VR scenarios and real-world tasks. One participant described challenges with gripping:

> *You can pick up things and put them in a drawer, but it's hard to grip them. Then it kind of falls and suddenly pops up again...You haven't managed to do what you intended...Maybe the tasks are too difficult or not well adapted.* [Jane]

Additional challenges included the difficulty of putting on heavy headsets for those with reduced arm function or configuring the VR system. One participant emphasized the need for customized adjustments to accommodate for stroke-related individual differences and goals:

> *I believe that those with severe cognitive impairments or difficulties with visual mapping might find VR too difficult initially. If we could personalize this, because one person's goal could be the opposite to others.* [Sue]

Participants recognized that VR required a certain level of cognitive function. Practicing new strategies in advanced tasks like kitchen training posed greater challenges than simpler tasks focused on errorless learning, such as dressing. Familiar settings were suggested to enhance engagement and skill transfer from virtual practice to real-world tasks. While cognitive impairments such as visual mapping and adjusting to new surroundings could cause difficulties, VR's potential to shift focus away from disability and reduce learned nonuse was noted. One participant explained:

> *When you enter that room, you slightly forget what is affected. And when you don't see that the arm isn't working, perhaps it will engage a bit more.* [Betty]

Kitchen training was widely regarded as a familiar and widely used element in rehabilitation. Participants noted that a VR kitchen could provide additional practice opportunities, as kitchens are a universally relevant and accessible setting for most people after stroke. A participant emphasized the importance of adjustable environments to accommodate individual needs.

"*You can put things high and low. You have the ability to increase or decrease the height of your entire kitchen, the cabinet, bench, whatever, so you can lower it or increase it for adjusting the difficulty level for your needs but also making it easier to use or potentially putting it to the exact same height as you have everything at home.* [John]

He also stressed the need for customizable difficulty levels, such as easy, medium, and hard, while acknowledging that further customizations were necessary to address individual differences. Many participants agreed that a one-size-fits-all approach would not work, but segmenting recovery stages into acute, subacute, and chronic phases, or tailoring tasks based on gripping ability, was proposed as a viable strategy.

Safety and Ease of Use as Fundamentals in VR Stroke Rehabilitation

Ensuring safety was a fundamental consideration in VR-based stroke rehabilitation. Many people after stroke face mobility impairments, balance issues, and cognitive challenges, increasing the risk of falls or disorientation in virtual environments. During the workshops, one health care professional highlighted their obligation to people after stroke:

> *We cannot tell them they can do it alone if we are not sure that it is safe.* [Jane]

Participants noted that a secure VR environment could encourage engagement and push physical limits, enhancing rehabilitation progress. However, some participants warned that VR might create a false sense of safety. As one participant with stroke explained:

> *Now I can't see where my legs are or anything. It's an uncertainty. I feel like going over to hold onto the table, but that won't work. Now I'm trying to do as I do at home and park the walker next to the kitchen door.* [Person after stroke, second-time user]

The transition from institutional rehabilitation to home-based therapy was identified as a critical phase, often accompanied by anxiety and uncertainty. One participant noted that VR could reduce anxiety and prevent early readmission to health care facilities:

> *We have to make sure they go home without as much anxiety, because when they do, they might just return early to the healthcare system due to a fall or stress from something minor. With the help of virtual reality, we could avoid this because we've tested it already at "home", but at the institution.* [Tom]

Participants emphasized the need to carefully consider the difference between VR tasks and real-world activities. He further explained:

> *That's going to be challenging, since there is a clear difference between how difficult something will be in VR versus in the real world, like cutting vegetables with a knife.* [Tom]

Some participants noted that successfully completing tasks in VR but failing in real life could lead to disappointment, especially when tasks required a higher degree of task complexity.

One highlighted the importance of minimizing unnecessary movement in home-based VR rehabilitation:

XSL•FO
RenderX

> *If someone is going to use it alone, I would limit tasks where you have to walk far, because I wouldn't consider that safe.* [Jane]

Similarly, Kate pointed out that while real-life practice is ideal, VR enables people after stroke to safely train in everyday activities like walking, empowering them to actively engage in their recovery.

Beyond physical safety, participants also discussed the psychological benefits of VR, particularly its ability to reduce the fear of failure, which often prevents people after stroke from participating in rehabilitation activities. Ben underscored the importance of structured support for VR in both institutional and home settings:

> *Using it in our hospital department would be great because they can use it in their rooms with instructions to be seated during activities. And it can be used in the later afternoon when they're done with daily [therapy-assisted] rehabilitation training. At home, they would be instructed to sit in a chair or a sofa with a space around them in case they get excited and get up.* [Ben]

The focus on sitting exercises typically placed occupational therapists in charge of VR rehabilitation due to traditional task distribution. However, health care professionals raised concerns about the safety and feasibility of using the VR system without therapist assistance, questioning whether users could operate and administer the system without assistance. Technological advancement, such as developing hybrid and mixed reality VR rooms, were proposed as potential solutions to address safety concerns. One participant with stroke explained:

> *I would need the ability to sense the outline of it [the walker], because it is safer when I know where it is. Otherwise, I could have practiced in the real kitchen, so it was fairly similar.* [Person after stroke, second-time user]

Furthermore, the presence of a caregiver or therapist was seen as essential for building user confidence in using VR, as indicated by Maud:

> *It's very important to have someone who watches what you and [make sure] you do it carefully. If you fall, there's someone with you, and you feel secure. Our people with stroke have belts on, and the physio holds the belts from the back. Maybe these people with stroke can't walk without a stick, but with that support, they feel very secure.* [Maud]

One proposed solution, which came up during workshops, was to differentiate between safer, low-scale tasks for home use and more challenging tasks, such as standing or increased reaching activities, to be conducted during follow-up sessions at the hospital.

Goal Orientation

The participants highlighted that a goal-oriented approach was essential in poststroke rehabilitation, as it helped people stay motivated and engaged after stroke. VR was perceived as a platform to structure rehabilitation programs around individualized goals, allowing people after stroke to track progress and experience achievement. The participants found that incorporating meaningful, measurable objectives encouraged consistent participation and reinforced the benefits of repetitive practice. Observing progress and perceiving tangible improvements increased commitment to rehabilitation exercises. Several participants stressed the importance of person-centered VR rehabilitation, where goals align with daily needs and personal aspirations. Karla noted:

> *The person with stroke's goal is everything in the rehabilitation. If someone wants to dress independently, we have to focus on that. The person with stroke is in the driver's seat.* [Karla]

Participants emphasized that VR technology should focus on customization in goal setting to enable people after stroke to work on relevant activities, like regaining independence in dressing, cooking, or mobility. To maximize effectiveness, VR interventions needed a clear focus.

> *I think it's really important if you use this in a rehabilitation program, you need to specify your target. Is it cognition, arm movement, or balance?* [Tom]

However, limitations were noted where VR could not adapt to all needs. During the workshops, a health care professional explained:

> *Some patients can only use this movement—just this—but others need to grip something. Then, the patients need resistance, and this [VR] won't improve their function like the real activity. If they are going to move the paint, there is some weight in the paint, and you dont get transferability. To see improvement at home, they must do it similarly.* [Physiotherapist, Workshop 3]

Participants suggested that providing multiple training options would enhance adherence and long-term commitment to therapy. Effective VR rehabilitation required targeting more tasks that mirror real-life activities to develop skills essential for independent living after discharge, such as putting up a painting, washing the dishes, or putting on clothes. It was important that the chores or activities used in rehabilitation served a specific therapeutic purpose. Kate pointed out:

> *An activity like painting, you don't want to be a better painter, but you want your arm to get better. That's the method.* [Kate]

From a therapist's perspective, this highlights how everyday tasks can be repurposed as meaningful rehabilitation goals, where the focus is on physical recovery rather than mastering the activity itself. A blended approach, where people after stroke first practiced activities in VR and then transferred skills to real-life settings, was suggested to reinforce motor learning and recovery. Repeating engagement in virtual tasks before real-life performance helped internalize movements, build confidence, and strengthen neural connections essential for motor function.

However, not all participants' goals were covered by VR prototypes. One participant with stroke noted:

XSL•FO
RenderX

> *There is no gait training, at least not intentionally. I would have to be on a treadmill, walking into the woods. This is for the arm, though.* [Person after stroke, second-time user]

Tracking progress in VR was seen as a powerful motivator for people after stroke. Participants explained that observing measurable improvements encouraged continued training and inspired confidence to tackle more demanding tasks. They emphasized that VR should serve as a stepping stone, bridging the gap between virtual practice and real-life application. However, there was a need for a clear focus on whether VR training aimed for improvement or sustainability. Gamification elements, like points or rewards, were considered ineffective unless the real-world relevance of the tasks was clear for people after stroke:

> *It's not necessarily the case that you get motivated by getting extra points, if you don't quite understand why you're doing this.* [Jane]

Ensuring VR-based rehabilitation included familiar, everyday tasks enabled explicit goals in a more engaging, functional, and transferable home environment. Therefore, a variety of VR rehabilitation activities, such as hanging clothes in the laundry room or other meaningful daily activities, were considered valuable.

Motivation

The participants emphasized that motivation played a crucial role in the effectiveness of VR rehabilitation for people post stroke, ensuring long-term engagement with their rehabilitation program. Tom noted the importance of regular follow-up sessions for maintaining motivation.

> *We can see when they were very motivated here. They go home, and there are only challenges, no solutions. But when they return for follow-up sessions, they're extremely motivated.* [Tom]

Motivated individuals were more likely to push themselves and put in the effort required to achieve their rehabilitation goals, potentially leading to better motor outcomes and quality of life. People with stroke noted that the VR scenarios stimulated learning and movement, enabling them to train without being consciously aware of it.

> *Person after stroke, third-time user: Now I have used the entire room, I see.*
>
> *Therapist: Now we have followed you around the entire room.*
>
> *Person after stroke, third-time user: Today's workout.*
>
> *VR developer: Shall we see how many points there have been. It sounded like it was beeping a lot.*

Karla highlighted how competition could drive people after stroke to improve their performance.

> *It's motivating when it's like a competition. You did this last time, now try to reach this one for the progression of function.* [Karla]

Setting and achieving progressively challenging goals allowed people after stroke to experience significant functional gains, reinforcing their overall commitment to recovery and reaching rehabilitation outcomes. One participant noted the extrinsic motivation provided by VR and its effects on the brain.

> *Because when you do that, your brain tells you it's very useful, because you get points.* [Person after stroke, second-time user]

While the rehabilitation process could be challenging and frustrating, motivation was seen as a driver to overcome the physical and emotional barriers, persisting through difficulties, and maintaining a positive attitude toward recovery. Many participants emphasized the role of human support in maintaining motivation, as Ben explained:

> *If you're not motivated or very depressed, your life is over. We won't do anything with you. You need humans to train with you.* [Ben]

Health care professionals' emotional and psychological encouragement was regarded as crucial in helping people after stroke stay engaged. Motivated individuals were more likely to communicate their preferences and needs, allowing for a more personalized and active rehabilitation experience.

In addition, motivation was closely linked to psychological well-being. Engaging in meaningful and enjoyable activities improved mood, reduced anxiety, and enhanced overall mental health, which was crucial for recovery. Sue underscored the role of curiosity in maintaining motivation.

> *It's more like an opportunity to explore and be curious. If you manage to get curious, you have less capability to be afraid or be sceptical.* [Sue]

Building confidence was another key factor in motivating people after stroke to embrace new challenges and persist with VR. One suggested strategy to foster confidence was visual progress tracking.

> *If you could implement that, for example, in painting stations, storing your paintings so you could see at the beginning I was able to make a line that was so big, but after two weeks I'm able to make a line that's this big, could be a motivational tool.* [Sue]

Seeing tangible improvements over time reinforced the belief of people after stroke in their abilities, motivating continued rehabilitation efforts. Jane emphasized the importance of simplicity and ease of use for maintaining motivation.

> *For people with stroke to use it themselves, they must have high motivation. It must be simple enough to master quickly, not something they struggle with and give up in the end. It must be meaningful and beneficial for them.* [Jane]

Others suggested integrating VR with existing digital training programs, such as Exorlive, to offer a self-training option to therapy-assisted activity training. Ben highlighted the difference in outcomes between motivated and unmotivated individuals.

> *If you're a person who wants to do better, you have a very clear goal. You just do that with the help of others or sometime without any help. But if you're a person like, oh, now this is my life, I'm going to live*

XSL•FO
RenderX

> *with it, and you're not motivated to do anything, you might just sit there and let it go day by day.* [Ben]

Those with clear goals and motivation were likely to strive for improvement, while those without motivation might become complacent.

Finally, John emphasized the importance of creating varied and personalized scenarios to maintain motivation, stating:

> *To motivate people, it would be more interesting to create different environments, such as a bedroom where you can dress yourself or a garage where you can tinker and build with your hands.* [John]

By providing scenarios aligned with the specific goals and interests of people after strokes, the rehabilitation process becomes more rewarding and engaging, fostering long-term motivation.

VR as a Complementary Tool

VR was seen as a valuable complement to stroke rehabilitation, enhancing training intensity and offering structured exercises in both clinical and home settings. While not a replacement for traditional therapy, it effectively supported recovery by supplementing existing approaches. Given the limited window for optimal recovery, its role in maximizing progress was emphasized. Mary highlighted VR's contribution to intensive training.

> *Our institution specializes in giving intensive training. I don't think VR could match that, but as an additional supplement it would be amazing. Together we could help people with stroke recover more quickly. Especially for people with stroke, there is a time period to get as good as they can get.* [Mary]

This highlights that VR should complement, not replace, therapist-led rehabilitation by enhancing therapy intensity during the critical recovery phase. Additionally, participants proposed implementing validated clinical measurements, goal assessment scales, or kinematic data to monitor outcomes and progress within the VR system during the workshops. Betty noted the alignment between the painting scenario and existing cognitive tests, such as MoCA, suggesting VR's potential for cognitive diagnostics in addition to general executive functions:

> *Because we have a test where we follow a line with a pen from number to number, or from numbers to letters. That kind of cognitive test. So, both for cognition and physical aspects, it [VR] is, if I may say so, very useful.* [Betty]

Individuals with stroke face unique challenges, requiring personalized rehabilitation plans tailored to their specific needs. VR adaptability offers a promising tool for delivering customized training that complements traditional therapy, ensuring alignment with each person's circumstances. VR was seen as a tool to bridge this gap by providing structured exercises that encourage movement and engagement.

> *It's another intervention in addition to what we do. I think it would increase the activity for the people with stroke to use their arm more than they otherwise would, if they have such an aid in addition to a therapist.* [Jane]

During workshops, VR was seen as a controlled environment where people after stroke could practice real-life tasks safely before transitioning to real-world applications. Jane highlighted this advantage:

> *We've discussed it being a supplement to training and in a safer setting for the people with stroke. It will not be instead of, but it may increase the activity between each session. Someone has concrete things to practice until next time, then practice in a real kitchen situation when someone is present to secure the situation.* [Jane]

Combining home-based VR training with in-person therapy was considered an ideal model for stroke rehabilitation. Participants, like Kate, emphasized that while VR enabled frequent, self-paced training at home, regular sessions with a therapist were essential to maintain motivation. This hybrid approach demonstrated VR's potential to extend rehabilitation beyond clinical settings while preserving professional oversight. Balancing self-practice with therapist support was considered crucial. However, as Betty noted, the attitude and behavior of staff introducing VR could significantly influence patient engagement.

> *Let's say the therapist has zero interest in VR. That could easily rub off on the patient. (..) I think it makes a big difference who introduces it to the patients.* [Betty]

As stroke rehabilitation requires extensive repetition to regain lost motor functions, VR was seen as an efficient way to deliver high-volume training, ensuring people after stroke get the practice needed for meaningful improvement. One participant explicitly noted they lacked solutions targeting volume training:

> *The advantage of VR is that you can do high-volume training efficiently. In rehabilitation of people with stroke, you need to spend hundreds of training hours to achieve some really big goals.* [John]

VR was seen as a possible way to make repetitive training more accessible and help people commit to the long-term rehabilitation process necessary for significant recovery. While VR was valuable, it was perceived as a stepping stone for real-life training. During the second workshop, an occupational therapist participant emphasized the limitations of VR in meeting diverse patient needs, noting that some individuals require functional resistance and weight-bearing activities to achieve transferability to real-life tasks:

> *Most people understand that VR is a supplementary tool along the way that bridges the gap, making it easier, but they know they need real life [activities] because that's where they are anyway.*

These results show that VR should not replace traditional therapy but serve as a tool to facilitate a smoother transition to functional activities.

XSL•FO
RenderX

Summary of the Key Findings

The integration of VR into stroke home rehabilitation presents both significant challenges and promising benefits. One of the primary challenges lies in the adaptability of VR systems for individuals with hemiparesis, as these systems must accommodate both right- and left-handed users. This requires considerable technical resources and expertise to ensure that the technology can be effectively tailored to the diverse needs of people after stroke. Additionally, customization remains a complex barrier, as VR solutions must be capable of addressing a wide range of stroke-related impairments and aligning with specific rehabilitation goals, which can vary greatly from person to person. Another critical concern is the balance between engagement and endurance. While VR has the potential to make rehabilitation more stimulating, it is essential to calibrate the intensity and duration of exercises to prevent cognitive overload and fatigue, which are common among individuals with stroke. Usability also poses a significant hurdle; many people after stroke experience limited motor function, which can make it difficult to interact with VR headsets and systems. Furthermore, challenges in visual mapping and task comprehension can hinder the effectiveness of VR-based interventions.

Despite these challenges, VR offers several compelling benefits for enhancing stroke rehabilitation at home. The integration of AI into VR platforms can enable highly personalized rehabilitation experiences, adapting in real time to the user's progress and specific needs. This level of customization can significantly enhance the efficacy of home-based therapy. Moreover, VR has the potential to increase motivation and engagement through gamification, progress tracking, and the simulation of meaningful, real-life tasks that resonate with users. VR also provides a controlled and safe environment for practicing daily activities, which can help reduce anxiety and better prepare people after stroke for real-world challenges. As a supplementary tool, VR can enhance traditional rehabilitation by increasing the intensity and volume of therapy. This aligns with 10 key principles of experience-dependent neuroplasticity, where factors such as specificity, high-repetition, intensity of stimulation, timing, task-specific training, auditory stimuli, and behavioral experience are essential for optimizing recovery outcomes after brain damage [42]. Textbox 1 provides an overview of the main challenges and potential benefits associated with using virtual reality in poststroke rehabilitation. It highlights critical considerations for implementation and the opportunities VR offers to enhance recovery outcomes.

Textbox 1. Summary of the key challenges and benefits in the use of VR for rehabilitation of people after stroke.

Key challenges in virtual reality (VR) usage for people after stroke rehabilitation

- Adaptability for hemiparesis
- VR systems need to accommodate both right- and left-handed users, which requires significant resources and skill to adapt to various people after stroke needs.
- Customization barriers
- VR technology must address individual stroke-related impairments and specific rehabilitation goals, which can be complex and demanding.
- Balance between engagement and endurance
- Tailoring the intensity and duration of VR exercises to avoid overwhelming people after strokes is crucial, especially considering cognitive load and fatigue.
- Usability concerns
- VR headsets and systems can be challenging for people after strokes with limited motor function, and difficulties in visual mapping and understanding tasks.

Key benefits in VR usage for people after stroke rehabilitation

- Enhanced customization with artificial intelligence
- Artificial intelligence–powered VR solutions can provide personalized rehabilitation experiences, address individual needs, and enhance home-based therapy.
- Motivation and engagement
- VR can make rehabilitation more engaging through gamification, tracking progress, and providing meaningful, real-life tasks that motivate people after strokes.
- Controlled and safe environment
- VR offers a controlled environment for practicing real-life tasks safely, reducing anxiety and preparing people with strokes for real-world challenges.
- Supplement to intensive training
- As a complementary tool, VR can increase therapy intensity, providing high-volume, repetitive training efficiently to maximize recovery outcomes.

XSL•FO
RenderX

Discussion

Principal Findings

This study explored the challenges and benefits of co-designing and using VR to support home-based poststroke rehabilitation, with a particular emphasis on user involvement and experience. The TA revealed 5 key themes that offer insight into how VR can be effectively integrated into rehabilitation practices. Adaptability emerged as a central concern, highlighting the need for VR systems to accommodate a wide range of stroke-related impairments, including motor, cognitive, and sensory limitations. Participants emphasized the importance of customizable interfaces and adjustable task difficulty to ensure accessibility and relevance. Safety and ease of use were also critical, with users expressing a preference for intuitive designs that minimize physical strain and cognitive overload, especially in unsupervised home settings. The theme of goal orientation underscored the value of VR in facilitating task-specific training that aligns with personal rehabilitation goals, thereby enhancing the perceived relevance and purpose of exercises. Motivation was closely linked to the immersive and interactive nature of VR, which participants found to be more engaging than traditional methods, potentially supporting persistent participation over time. Finally, the role of VR as a complementary tool reflected a shared understanding that VR should not replace conventional therapy, but rather augment it, offering additional benefits for practice and reinforcement in a flexible, home-based format.

The study has revealed that one of the primary benefits lies in the adaptability and individualization of VR environments to meet the diverse needs of people after stroke. As noted in the study, many people after stroke experience hemiparesis, necessitating VR systems that accommodate both right- and left-affected users. It was noted that the final prototype involving the kitchen scenario factored this in by rewarding participants for using their affected arm. The ability to individualize VR experiences to address specific activities impaired by stroke, such as initially setting up the VR system and further conducting activities, such as reaching and gripping objects, is critical for effective rehabilitation. With appropriate individualization, VR systems have demonstrated significant effectiveness in improving various aspects of stroke rehabilitation, including upper limb function, cognitive abilities, and balance [25,43,44].

However, to bridge the gaps between the recovery process in stroke rehabilitation, technology, and clinical practice, an awareness, exploration, experimentation, and evaluation framework ensuring the solutions are engaging, accessible, accountable, and adaptable is advocated [45]. A common challenge with commercially available gaming technology, such as VR, is that even though it provides immersive, engaging, and tailored therapy (accessibility and engagement), it is not easily adapted for people after stroke, especially those experiencing limited movement or cognition. Based on our assessment, we argue that the complexity of the devices makes them difficult to adjust, posing a challenge for people after stroke to secure them properly and customize them to their individual needs. Although customization was frequently proposed to meet the individual needs of users, this highlights a critical tension. In practice, effective customization demands considerable resources to adequately and efficiently respond to the diverse and complex needs of individuals living with stroke. As indicated by one study, reducing the complexity to one working application on the goggles can limit potential errors [23]. In this study, some individuals required additional assistance in adjusting the headset or coping with cognitive impairments. There is a paradox between the ideal of customization and the reality of implementing it effectively. Achieving this may involve a balance between providing customized opportunities and still making it feasible to implement and ensure its ease of use by categorizing similar users into groups.

VR has the benefit of providing multiple training environments that target individuals by engaging them in meaningful and enjoyable activities, as particularly found in the kitchen scenario. Functional exercises, such as washing dishes, dressing, and brushing teeth, allow people after stroke to develop skills essential for independent living after discharge [46]. This aligns with one study [47], in which people after stroke found different activities in everyday life easier to perform since starting their VR training, such as opening a drawer or applying toothpaste to a toothbrush. Enriched environments are essential in leveraging VR for stroke rehabilitation, as they offer diverse and stimulating approaches to problem-solving and skill development [48]. Some overarching principles that guide an enriched environment approach are complexity, variety, and novelty of the environment as well as targeting underlying needs [49]. In our study, the VR scenarios targeted all principles, but ideas for scaffolding complexity and variety tailored for each individual were proposed.

By simulating real-life scenarios and interactive challenges, VR provides people after stroke with engaging and repetitive practice, which is crucial for neuroplasticity and motor recovery [50]. Additionally, the adaptability of VR allows therapists to modify the learning environment to match individual needs, for example, by adjusting the levels of difficulty, thereby fostering a more immersive and intensive experience that enhances engagement and accelerates rehabilitation progress. This flexibility not only improves functional recovery but also increases motivation and adherence to therapy, making VR a promising tool in modern rehabilitation strategies [51]. In a study by Gustavsson et al [47], VR provided a feeling of being in a different world where users felt they were reaching higher and moving faster. However, VR environments can deceive the mind into believing they are a real-life environment. This illusion is a cause for concern, as the immersive nature of VR can make the brain perceive the simulated surroundings as authentic, thereby inducing a false sense of safety and achievement [52].

Given the challenging nature of rehabilitation for individuals who have experienced a stroke, VR plays a critical role in motivating them by ensuring their active participation and long-term effort in the recovery process. Engaged people after stroke are more likely to adhere to their therapy schedules and participate actively in their rehabilitation exercises. Rehabilitation can be challenging and frustrating, and motivation

XSL•FO
RenderX

helps people after stroke overcome these barriers, persist through difficulties, and maintain a positive attitude toward their recovery [25]. It drives users to push through the physical and emotional challenges of rehabilitation, helping them to stay committed to their therapy even when progress seems slow or when they encounter setbacks. By fostering a sense of achievement and progress, motivation can transform the rehabilitation experience from a daunting task into a rewarding journey.

However, maintaining motivation can be difficult, especially for people after stroke facing mental health challenges, such as depression or anxiety. These conditions can significantly impact their willingness to engage in rehabilitation activities. While we saw that the use of gamification elements, such as points and rewards, served as incentives to motivate people after stroke, they work only if people after stroke understand the real-world relevance of their tasks. Without this understanding, the motivational impact of these gamification elements may be significantly diminished. This highlights the need to design VR programs that closely mimic real-world scenarios to maximize their effectiveness.

Another benefit of VR is that it creates a controlled and secure environment that encourages people after stroke to engage more actively in rehabilitation by reducing the fear of injury, thereby promoting better recovery outcomes when they feel safe enough to push their physical limits. However, this is contingent on structured supervision and support, which are essential to mitigating risks and enhancing user confidence, particularly given the mobility impairments, balance issues, and cognitive challenges faced by people after stroke, which increase the risk of falls or disorientation [25,44]. The importance of therapist involvement in tutorials and adjusting difficulty levels when engaging with VR is highlighted in multiple studies [23,47]. This is particularly important for individuals recovering from stroke, as prolonged exposure to VR may induce motion sickness or dizziness. Additionally, those with impairments in executive functioning and problem-solving challenges observed during our workshops may find it especially difficult to navigate and adapt to such technologies. To maximize the benefits of VR therapy, careful system design, real-time monitoring, and gradual progression in exercise intensity are essential to ensuring a safe and effective rehabilitation experience [53].

While support and guidance in the use of VR emerged as indispensable, this could present challenges in terms of manpower due to the decreasing number of health care professionals. The presence of therapists significantly motivated many individuals with stroke to continue engaging with VR activities. However, this reliance on external motivation could become problematic once they are discharged home, where self-motivation is essential for ongoing rehabilitation [54]. It is not feasible to maintain this level of support when individuals with stroke transition to their home setting. Many people with brain injuries are dependent on help from their families when using digital technologies, such as VR, at home [23,54].

The use of VR presented an advantage in facilitating a goal-oriented approach by enhancing consistent participation and reinforcing the benefits of repetitive practice. The application of motor learning principles in VR design is advocated to enhance rehabilitation outcomes [25]. This includes progressively challenging tasks, real-time performance feedback, and multimodal feedback [25,43]. The development process through the workshops led to improvements in VR content in line with standardized stroke guidelines, such as goal- and task-specific activities and quantifying activities. However, it emerged from the study that the challenge lies in ensuring that VR interventions are carefully designed with clear and specific goals, whether to improve cognitive function, enhance arm movement, or restore balance. It is important that the rehabilitation goals are aligned with the daily needs and personal aspirations of the people after stroke. This requires a deep understanding of each person's lifestyle, preferences, and long-term objectives. Achieving this alignment is challenging, as the goals can change over time, requiring continuous assessment and adjustment of the VR activities to ensure they remain relevant and motivating.

Despite the positive effects associated with measuring progress in VR, the process can be difficult [55]. This difficulty arises from the need to ensure that the metrics used in VR accurately reflect real-world abilities and improvements. While many VR games are primarily designed for entertainment rather than rehabilitation, the importance of involving users in the co-design of VR applications is increasingly recognized [25].

The use of VR in this study revealed that it is beneficial as a complementary tool in stroke rehabilitation, enhancing user engagement, increasing training intensity, and providing structured exercises in both clinical and home environments. While VR cannot replace traditional rehabilitation methods, it can serve as a powerful addition that supplements existing therapy approaches. Previous studies suggest that integrating specific VR technologies with traditional rehabilitation approaches can lead to greater improvements in motor function and activity levels in people after stroke compared with using conventional rehabilitation methods alone [56-58]. For example, VR can provide repetitive and varied exercises that target specific motor skills, helping people after stroke to practice and refine their movements in a controlled environment, thereby freeing therapists from the mundane tasks of high-volume training.

A potential benefit could also be, as found in a study [59], that VR demonstrated significantly greater therapeutic effects than conventional training in improving upper limb function, as assessed by the Fugl-Meyer and Action Arm Research Tests [32]. Following the stroke trajectory, VR could be implemented during the hospital stay, as a prepractice and supplement to real-life activities. In the home setting, it would need to be scaled down to accommodate safety precautions, but has the possibility to be further scaffolded when returning for follow-up at the hospital in cooperation with health care professionals. A significant challenge in using VR for rehabilitation is ensuring its effective integration with traditional therapy while maintaining professional oversight as rehabilitation extends beyond clinical settings.

Successful integration is essential, as it necessitates a seamless combination of VR-based and conventional approaches to

XSL•FO
RenderX

optimize therapeutic outcomes. Professional oversight remains crucial for monitoring patient progress, adjusting interventions, and providing ongoing support tasks that become increasingly difficult to maintain once people transition to home-based rehabilitation after stroke.

Implications for Practice

Our findings suggest that VR could be a valuable supplement to traditional rehabilitation for specific stroke groups, but it cannot replace the essential contact with the therapist, clinical reasoning, or the assessment of activity limitations and opportunities. We propose that VR could be particularly beneficial for a younger population with stroke who lack sufficient access to home-based rehabilitation and struggle with motivation for self-training. Additionally, VR may help maintain function in individuals in the chronic phase after a stroke, where repetition and high-quality training are crucial to prevent functional decline. However, it is important to acknowledge that VR cannot fully replicate training in real-life, everyday activities. For VR solutions to be effective, they must be carefully tailored to the individual's functional level and ease of use. This tailoring process, however, can be resource-intensive in an already strained health care system.

In light of this, we raise important questions about who will truly benefit from VR, given the wide range of outcomes following a stroke, and whether, in some cases, it may be more appropriate to prioritize traditional therapy. Finally, we emphasize the importance of training for relatives and caregivers, as their role may be critical to the successful use of VR at home.

Limitations and Future Work

The limitations of this study primarily stem from the use of a qualitative single case study design [27], which provides an in-depth understanding of the real-life phenomenon but may limit the generalizability of the findings across broader contexts. The reliance on rapid prototyping and TA, while effective for capturing detailed insights into VR usage in poststroke rehabilitation, may not fully address the variability of experiences and needs among a wider population of people after stroke. Also, the rapid evolution of VR technology means that some of the results will soon be outdated, limiting the applicability of the findings to current and future developments in the field [55,60].

Future work should aim to expand the scope of research by incorporating multiple case studies across diverse settings to enhance the generalizability of the results. This could involve a larger and more varied participant pool to capture a broader range of experiences and needs. Future work should focus on developing integrated, individualized therapy solutions that seamlessly combine arm therapy, cognitive therapy, and mental support within a single solution, addressing the multifaceted rehabilitation needs, such as cognitive needs or physical therapy needs, of individuals with stroke as described in Gkintoni et al [61].

This requires implementing design methodologies that facilitate co-creation with diverse stakeholders, including health care professionals, VR developers, and people after stroke with varying needs, ensuring that the solutions are both inclusive and effective. Efforts should also be directed toward creating a smooth transition from hospital settings to home environments for people after stroke, as described by Pourliaka et al [62], using VR to extend rehabilitation beyond clinical settings while maintaining professional oversight.

Furthermore, the development of AI-powered solutions that account for the personalized goals, needs, and situations of people after stroke is crucial, enabling tailored interventions that adapt to individual progress and challenges. For instance, the system could select the most appropriate exercises based on how individuals perform each exercise [63]. One option is to incorporate validated assessments, such as the short-form of MoCA [31,64], to evaluate patient situations accurately and guide therapy adjustments.

Designing solutions that not only excite individuals with stroke, but also genuinely facilitate improvements in their therapy goals and daily life challenges is vital. This involves crafting engaging VR environments that motivate consistent participation while ensuring real-world relevance and measurable benefits in rehabilitation outcomes. Further exploration of potential scenarios that can adapt to the individual requirements of home rehabilitation over time is also essential.

Additionally, more validation studies, for example, pilot trials or feasibility studies to evaluate the preliminary effects and randomized controlled trials to evaluate actual VR solution outcomes, are urgently needed [54]. In addition, standardized usability and comfort measures, such as the System Usability Scale and the Virtual Reality Sickness Questionnaire, were not used. Their inclusion could have offered more structured and quantifiable assessments of user experience, particularly regarding interface usability and potential adverse effects like motion sickness. Future work should consider integrating these validated instruments to complement qualitative feedback and provide a more comprehensive evaluation of user interaction and comfort.

Conclusion

This study underscores the transformative potential of VR in advancing stroke rehabilitation. VR provides a dynamic and interactive platform that can be customized to support the individualized trajectories of recovery, accommodating the physical limitations, cognitive profile, and personal goals of people after stroke. Its core strength lies in its adaptability by offering immersive simulations of real-world tasks that foster motivation and engagement through meaningful and goal-directed experiences. However, the successful application of VR in rehabilitation requires more than technological advancement. It needs to be integrated thoughtfully with established rehabilitation principles, including task-specific training, measurable and ecologically valid outcomes, as well as alignment with the evolving needs of individuals. A critical insight from this study is the necessity of embedding clear therapeutic intent into VR interventions. Whether the objective is motor recovery, cognitive enhancement, or balance restoration, each activity should serve a purpose that is clinically relevant and meaningful to the user's daily life. This demands a comprehensive understanding of individual motivations,

XSL•FO
RenderX

challenges, and aspirations through involving people after stroke in the design process. It is through participatory co-design that developers can ensure that VR experiences are both engaging and clinically appropriate. As an emerging technology, VR should not be considered a replacement for traditional therapy, but rather a powerful adjunct that can increase training intensity, reduce the burden on therapists, and extend rehabilitation into home settings, making therapy more accessible and continuous. When thoughtfully designed and implemented, VR can bridge the gap between structured clinical care and independent recovery, providing consistent support throughout the rehabilitation journey.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Data Availability

The dataset is derived from interviews in Norwegian and is not publicly available due to consent restrictions and data collection approval. Parts of the data can be made available from the corresponding author upon request.

Authors' Contributions

Conceptualization: NP, MAP
Methodology: NP, MAP, TS
Investigation: TS, AMHM, NP, IKHR, KGSO, NSM, ÅB
Data curation: TS, AMHM
Formal analysis: All authors
Project administration: TS, NP, MAP
Resources: IKHR, KGSO, NSM, ÅB
Validation: All authors
Visualization: NP, AMHM
Writing – original draft: TS, AMHM
Writing – review and editing: TS, AMHM, IKHR, KGSO, NSM, ÅB, NP, MAP

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview and workshop guide.
[DOCX File , 98 KB - humanfactors_v13i1e78798_app1.docx]

References

1. GBD 2019 Stroke Collaborators. Global, regional, and national burden of stroke and its risk factors, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Neurol 2021 Oct;20(10):795-820 [FREE Full text] [doi: 10.1016/S1474-4422(21)00252-0] [Medline: 34487721]
2. Fugazzaro S, Denti M, Accogli M, Costi S, Pagliacci D, Calugi S, On Behalf Of Look After Yourself Project. Self-management in stroke survivors: development and implementation of the Look after Yourself (LAY) intervention. Int J Environ Res Public Health 2021 May 31;18(11):1-14 [FREE Full text] [doi: 10.3390/ijerph18115925] [Medline: 34072998]
3. Saini V, Guada L, Yavagal DR. Global epidemiology of stroke and access to acute ischemic stroke interventions. Neurology 2021;97(20 Suppl 2):S6-S16. [doi: 10.1212/WNL.0000000000012781] [Medline: 34785599]
4. Kwakkel G, Stinear C, Essers B, Munoz-Novoa M, Branscheidt M, Cabanas-Valdés R, et al. Motor rehabilitation after stroke: European Stroke Organisation (ESO) consensus-based definition and guiding framework. Eur Stroke J 2023 Dec;8(4):880-894 [FREE Full text] [doi: 10.1177/23969873231191304] [Medline: 37548025]
5. Bui J, Luauté J, Farnè A. Enhancing upper limb rehabilitation of stroke patients with virtual reality: a mini review. Front. Virtual Real 2021 Nov 8;2:595771. [doi: 10.3389/frvir.2021.595771]
6. Dias P, Silva R, Amorim P, Lains J, Roque E, Pereira ISF, et al. Using virtual reality to increase motivation in poststroke rehabilitation. IEEE Comput. Grap. Appl 2019;39(1):64-70. [doi: 10.1109/mcg.2018.2875630]
7. Sahely A, Giles D, Sintler C, Soundy A, Rosewilliam S. Self-management interventions to improve mobility after stroke: an integrative review. Disabil Rehabil 2023 Jan;45(1):9-26 [FREE Full text] [doi: 10.1080/09638288.2022.2028019] [Medline: 35068313]

XSL•FO
RenderX

8. Meyer S, Verheyden G, Brinkmann N, Dejaeger E, De Weerdt W, Feys H, et al. Functional and motor outcome 5 years after stroke is equivalent to outcome at 2 months: follow-up of the collaborative evaluation of rehabilitation in stroke across Europe. Stroke 2015;46(6):1613-1619. [doi: 10.1161/strokeaha.115.009421]
9. Lieshout ECCV, van de Port IG, Dijkhuizen RM, Visser-Meily JMA. Does upper limb strength play a prominent role in health-related quality of life in stroke patients discharged from inpatient rehabilitation? Top Stroke Rehabil 2020 Oct;27(7):525-533 [FREE Full text] [doi: 10.1080/10749357.2020.1738662] [Medline: 32148183]
10. Schnabel S, van Wijck F, Kidd L. Supporting stroke survivors to meet their personal rehabilitation needs in community-based arm rehabilitation: development of initial programme theories to explore what may work for whom, how and under what circumstances. Front Neurol 2023;14:1089547 [FREE Full text] [doi: 10.3389/fneur.2023.1089547] [Medline: 37332992]
11. Mekbib DB, Han J, Zhang L, Fang S, Jiang H, Zhu J, et al. Virtual reality therapy for upper limb rehabilitation in patients with stroke: a meta-analysis of randomized clinical trials. Brain Inj 2020 Mar 20;34(4):456-465. [doi: 10.1080/02699052.2020.1725126] [Medline: 32064964]
12. Domínguez-Téllez P, Moral-Muñoz JA, Salazar A, Casado-Fernández E, Lucena-Antón D. Game-based virtual reality interventions to improve upper limb motor function and quality of life after stroke: systematic review and meta-analysis. Games Health J 2020 Feb;9(1):1-10 [FREE Full text] [doi: 10.1089/g4h.2019.0043] [Medline: 32027185]
13. Henderson A, Korner-Bitensky N, Levin M. Virtual reality in stroke rehabilitation: a systematic review of its effectiveness for upper limb motor recovery. Top Stroke Rehabil 2007;14(2):52-61. [doi: 10.1310/tsr1402-52] [Medline: 17517575]
14. Kang D, Park J, Eun S. Home-based virtual reality exergame program after stroke rehabilitation for patients with stroke: a study protocol for a multicenter, randomized controlled trial. Life (Basel) 2023;13(12):2256 [FREE Full text] [doi: 10.3390/life13122256] [Medline: 38137857]
15. Clark WE, Sivan M, O'Connor RJ. Evaluating the use of robotic and virtual reality rehabilitation technologies to improve function in stroke survivors: a narrative review. J Rehabil Assist Technol Eng 2019;6:2055668319863557 [FREE Full text] [doi: 10.1177/2055668319863557] [Medline: 31763052]
16. Park D, Lee D, Lee K, Lee G. Effects of virtual reality training using Xbox Kinect on motor function in stroke survivors: a preliminary study. J Stroke Cerebrovasc Dis 2017;26(10):2313-2319. [doi: 10.1016/j.jstrokecerebrovasdis.2017.05.019] [Medline: 28606661]
17. Sin H, Lee G. Additional virtual reality training using Xbox Kinect in stroke survivors with hemiplegia. Am J Phys Med Rehabil 2013;92(10):871-880. [doi: 10.1097/PHM.0b013e3182a38e40] [Medline: 24051993]
18. Moan ME, Vonstad EK, Su X, Vereijken B, Solbjør M, Skjæret-Maroni N. Experiences of stroke survivors and clinicians with a fully immersive virtual reality treadmill exergame for stroke rehabilitation: a qualitative pilot study. Front Aging Neurosci 2021;13:735251 [FREE Full text] [doi: 10.3389/fnagi.2021.735251] [Medline: 34795576]
19. DiCarlo JA, Gheihman G, Lin DJ. Reimagining stroke rehabilitation and recovery across the care continuum: results from a design-thinking workshop to identify challenges and propose solutions. Archives of Physical Medicine and Rehabilitation 2021;102(8):1645-1657. [doi: 10.1016/j.apmr.2021.01.074] [Medline: 33556351]
20. Erikson A, Ranner M, Guidetti S, von Koch L. In search of self after stroke: a longitudinal qualitative study in the context of client-centred rehabilitation. Int J Qual Stud Health Well-being 2023;18(1):2282513 [FREE Full text] [doi: 10.1080/17482631.2023.2282513] [Medline: 38010156]
21. Bucki B, Spitz E, Baumann M. Emotional and social repercussions of stroke on patient-family caregiver dyads: analysis of diverging attitudes and profiles of the differing dyads. PLoS One 2019;14(4):e0215425 [FREE Full text] [doi: 10.1371/journal.pone.0215425] [Medline: 31013289]
22. Maggio M, Latella D, Maresca G, Sciarrone F, Manuli A, Naro A, et al. Virtual reality and cognitive rehabilitation in people with stroke: an overview. J Neurosci Nurs 2019;51(2):101-105. [doi: 10.1097/JNN.0000000000000423] [Medline: 30649091]
23. Johansen T, Matre M, Tornås S, Løvstad M, Ponsford JL, Olsen A, et al. I thought it would be difficult, but this is actually something I can do - experiences with virtual reality-based cognitive training in persons with TBI. Ann Med 2025;57(1):2490218 [FREE Full text] [doi: 10.1080/07853890.2025.2490218] [Medline: 40219760]
24. Bally E, Cheng D, van Grieken A, Ferri Sanz M, Zanutto O, Carroll A, et al. Patients' perspectives regarding digital health technology to support self-management and improve integrated stroke care: qualitative interview study. J Med Internet Res 2023;25:e42556 [FREE Full text] [doi: 10.2196/42556] [Medline: 37014677]
25. Brassel S, Power E, Campbell A, Brunner M, Togher L. Recommendations for the design and implementation of virtual reality for acquired brain injury rehabilitation: systematic review. J Med Internet Res 2021;23(7):e26344 [FREE Full text] [doi: 10.2196/26344] [Medline: 34328434]
26. Salim H, Lee PY, Sharif-Ghazali S, Cheong AT, Wong J, Young I, RESPIRE Collaboration. Developing an asthma self-management intervention through a web-based design workshop for people with limited health literacy: user-centered design approach. J Med Internet Res 2021;23(9):e26434 [FREE Full text] [doi: 10.2196/26434] [Medline: 34499039]
27. Yin R. Case Study Research: Design and Methods. Thousand Oaks, California: Sage; 2009.
28. Ridder H. The theory contribution of case study research designs. Bus Res 2017;10(2):281-305. [doi: 10.1007/s40685-017-0045-z]
29. Pereira JF, Santos J, Bertanha M. Rapid prototyping and health: an ergonomic footrest to improve the venous return. Product 2020;18(2):196-203. [doi: 10.4322/pmd.2020.031]

XSL•FO
RenderX

30. Bernhardt J, Hayward KS, Kwakkel G, Ward NS, Wolf SL, Borschmann K, et al. Agreed definitions and a shared vision for new standards in stroke recovery research: the stroke recovery and rehabilitation roundtable taskforce. Int J Stroke 2017;12(5):444-450 [FREE Full text] [doi: 10.1177/1747493017711816] [Medline: 28697708]
31. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
32. Platz T, Pinkowski C, van Wijck F, Kim I, di Bella P, Johnson G. Reliability and validity of arm function assessment with standardized guidelines for the Fugl-Meyer test, action research arm test and box and block test: a multicentre study. Clin Rehabil 2005;19(4):404-411. [doi: 10.1191/0269215505cr832oa] [Medline: 15929509]
33. Persson HC, Alt Murphy M, Danielsson A, Lundgren-Nilsson, Sunnerhagen KS. A cohort study investigating a simple, early assessment to predict upper extremity function after stroke - a part of the SALGOT study. BMC Neurol 2015;15:92 [FREE Full text] [doi: 10.1186/s12883-015-0349-6] [Medline: 26084397]
34. Steen M, Manschot M, De Koning N. Benefits of co-design in service design projects. International Journal of Design 2011;5(2):53-60 [FREE Full text]
35. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
36. Younas A, Masih Y, Sundus A. Alternatives to ‘saturation’ for greater transparency in reporting of sample size decision-making in qualitative research. Evid Based Nurs 2025;28(2):77-80. [doi: 10.1136/ebnurs-2025-104292]
37. Creswell J. Qualitative Inquiry and Research Design: Choosing Among Five Traditions. Thousand Oaks: SAGE Publications; 1998.
38. Meydan C. The role of triangulation in qualitative research: converging perspectives. In: Principles of Conducting Qualitative Research in Multicultural Settings. Hershey, Pennsylvania: IGI Global; 2024:98-129.
39. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2008 Jul 21;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
40. Braun V, Clarke V. Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. Couns and Psychother Res 2020;21(1):37-47. [doi: 10.1002/capr.12360]
41. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014;9:26152 [FREE Full text] [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]
42. Kleim JA, Jones TA. Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. J Speech Lang Hear Res 2008;51(1):S225-S239. [doi: 10.1044/1092-4388(2008/018)] [Medline: 18230848]
43. Xu S, Xu Y, Wen R, Wang J, Qiu Y, Chan CC. Virtual reality enhanced exercise training in upper limb function of patients with stroke: meta-analytic study. J Med Internet Res 2025;27:e66802 [FREE Full text] [doi: 10.2196/66802] [Medline: 39969977]
44. Maheta B, Kraft A, Interrante N, Fereydooni S, Bailenson J, Beams B, Keny, et al. Using virtual reality to improve outcomes related to quality of life among older adults with serious illnesses: systematic review of randomized controlled trials. J Med Internet Res 2025;27:e54452. [doi: 10.2196/54452] [Medline: 40009834]
45. Jayasree-Krishnan V, Ghosh S, Palumbo A, Kapila V, Raghavan P. Developing a framework for designing and deploying technology-assisted rehabilitation after stroke: a qualitative study. Am J Phys Med Rehabil 2021;100(8):774-779. [doi: 10.1097/PHM.0000000000001634] [Medline: 33141773]
46. National Clinical Guideline Centre (UK). Planning and delivering stroke rehabilitation. In: Stroke Rehabilitation: Long Term Rehabilitation After Stroke [internet]. UK: Royal College of Physicians (UK); 2013.
47. Gustavsson M, Kjörk EK, Erhardsson M, Alt Murphy M. Virtual reality gaming in rehabilitation after stroke - user experiences and perceptions. Disabil Rehabil 2022;44(22):6759-6765 [FREE Full text] [doi: 10.1080/09638288.2021.1972351] [Medline: 34465269]
48. Pereira F, Bermúdez I Badia S, Jorge C, Cameirão MS. The use of game modes to promote engagement and social involvement in multi-user serious games: a within-person randomized trial with stroke survivors. J Neuroeng Rehabil 2021;18(1):62 [FREE Full text] [doi: 10.1186/s12984-021-00853-z] [Medline: 33853609]
49. Oppici L, Bērziņa G, Hestetun-Mandrup AM, Løvstad M, Opheim A, Pacheco MM, PEER-HOMEcare consortium, et al. A scoping review of preclinical environmental enrichment protocols in models of poststroke to set the foundations for translating the paradigm to clinical settings. Transl Stroke Res 2025;16(5):1850-1873. [doi: 10.1007/s12975-025-01335-3] [Medline: 39913056]
50. Khan A, Imam YZ, Muneer M, Al Jerdi S, Gill SK. Virtual reality in stroke recovery: a meta-review of systematic reviews. Bioelectron Med 2024;10(1):23. [doi: 10.1186/s42234-024-00150-9]
51. Levin MF. Can virtual reality offer enriched environments for rehabilitation? Expert Rev Neurother 2011;11(2):153-155. [doi: 10.1586/ern.10.201] [Medline: 21306202]
52. Sokołowska B. Being in virtual reality and its influence on brain health-an overview of benefits, limitations and prospects. Brain Sci 2024;14(1):72 [FREE Full text] [doi: 10.3390/brainsci14010072] [Medline: 38248287]

XSL•FO
RenderX

53. Ceradini M, Losanno E, Micera S, Bandini A, Orlandi S. Immersive VR for upper-extremity rehabilitation in patients with neurological disorders: a scoping review. J Neuroeng Rehabil 2024;21(1):75 [FREE Full text] [doi: 10.1186/s12984-024-01367-0] [Medline: 38734690]
54. Hestetun-Mandrup AM, Hamre C, Lund A, Martinsen ACT, He H, Pikkarainen M. Exploring people with stroke's perceptions of digital technologies in post-stroke rehabilitation - a qualitative study. Disabil Rehabil 2025:1-10 [FREE Full text] [doi: 10.1080/09638288.2025.2504615] [Medline: 40382688]
55. Hestetun-Mandrup AM, Toh ZA, Oh HX, He H, Martinsen ACT, Pikkarainen M. Effectiveness of digital home rehabilitation and supervision for stroke survivors: a systematic review and meta-analysis. Digit Health 2024;10:20552076241256861 [FREE Full text] [doi: 10.1177/20552076241256861] [Medline: 38832099]
56. Olana DD, Abessa TG, Lamba D, Triccas LT, Bonnechere B. Effect of virtual reality-based upper limb training on activity of daily living and quality of life among stroke survivors: a systematic review and meta-analysis. J Neuroeng Rehabil 2025;22(1):92 [FREE Full text] [doi: 10.1186/s12984-025-01603-1] [Medline: 40269877]
57. Rodríguez-Hernández M, Polonio-López B, Corregidor-Sánchez AI, Martín-Conty JL, Mohedano-Moriano A, Criado-Álvarez JJ. Can specific virtual reality combined with conventional rehabilitation improve poststroke hand motor function? A randomized clinical trial. J Neuroeng Rehabil 2023;20(1):38 [FREE Full text] [doi: 10.1186/s12984-023-01170-3] [Medline: 37016408]
58. Dixit, Phalswal U, Kalal N, Srivastava SP. Effectiveness of virtual reality-supported exercise therapy in improving upper extremity function and activities of daily living among patients after stroke: a systematic review of randomized control trials. Osong Public Health Res Perspect 2024;15(3):189-200 [FREE Full text] [doi: 10.24171/j.phrp.2023.0148] [Medline: 38988022]
59. Soleimani M, Ghazisaeedi M, Heydari S. The efficacy of virtual reality for upper limb rehabilitation in stroke patients: a systematic review and meta-analysis. BMC Med Inform Decis Mak 2024 May 24;24(1):135 [FREE Full text] [doi: 10.1186/s12911-024-02534-y] [Medline: 38790042]
60. Rodrigues P, Quaresma C, Costa M, Luz F, Fonseca MM. Virtual reality-based telerehabilitation for upper limb recovery post-stroke: a systematic review of design principles, monitoring, safety, and engagement strategies. arXiv:2501.06899 2025:00 [FREE Full text] [doi: 10.48550/arXiv.2501.06899]
61. Gkintoni E, Vassilopoulos SP, Nikolaou G. Next-generation cognitive-behavioral therapy for depression: integrating digital tools, teletherapy, and personalization for enhanced mental health outcomes. Medicina (Kaunas) 2025;61(3):431 [FREE Full text] [doi: 10.3390/medicina61030431] [Medline: 40142242]
62. Pourliaka T, Tsiakiri A, Proios H. Exploring discharge preparedness of patients with stroke: a qualitative descriptive study. Creat Nurs 2025;31(4):440-450. [doi: 10.1177/10784535251320005] [Medline: 40017407]
63. Martínez-Cid S, Vallejo D, Herrera V, Schez-Sobrino S, Castro-Schez JJ, Albusac JA. Explainable AI-driven decision support system for personalizing rehabilitation routines in stroke recovery. Prog Artif Intell 2025:1-23. [doi: 10.1007/s13748-024-00357-6]
64. McDicken JA, Elliott E, Blayney G, Makin S, Ali M, Larner AJ, VISTA-Cognition Collaborators. Accuracy of the short-form Montreal Cognitive Assessment: systematic review and validation. Int J Geriatr Psychiatry 2019 Oct;34(10):1515-1525 [FREE Full text] [doi: 10.1002/gps.5162] [Medline: 31243810]

Abbreviations

AI: artificial intelligence
MoCA: Montreal Cognitive Assessment
TA: thematic analysis
VR: virtual reality

Edited by A Kushniruk; submitted 10.Jun.2025; peer-reviewed by D Hobbs, A Glaessel; comments to author 06.Aug.2025; revised version received 28.Aug.2025; accepted 25.Nov.2025; published 15.Jan.2026.

Please cite as:
Saidi T, Hestetun-Mandrup AM, Pavel N, Rosseland IKH, Selmer-Olsen KG, Mørk NS, Bergheim Å, Pikkarainen MA
Challenges and Benefits of Virtual Reality in Home-Based Poststroke Rehabilitation: Co-Design Qualitative Study
JMIR Hum Factors 2026;13:e78798
URL: https://humanfactors.jmir.org/2026/1/e78798
doi:10.2196/78798
PMID:

XSL•FO
RenderX

©Trust Saidi, Ann Marie Hestetun-Mandrup, Nenad Pavel, Ingvild Kristina Hurum Rosseland, Kathinka Granum Selmer-Olsen, Nora Synnøve Mørk, Åse Bergheim, Minna Annika Pikkarainen. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 15.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Multimodal Virtual Reality Assessment of Medication Effects in Attention-Deficit/Hyperactivity Disorder and Its Distinction From Depression: Cross-Sectional Study

Laura Asché[1*], MSc; Julian Pakos[1*], MSc; Hannah Schrage[1], MD; Johanna Schuder[1], MD; Luisa Jung[1], MSc; Dario Sanchez[1], MSc; Benjamin Selaskowski[1], PhD; Annika Wiebe[1], PhD; Alexandra Philipsen[1], Prof Dr Med; Niclas Braun[1], PhD

Department of Psychiatry and Psychotherapy, University Hospital Bonn, University of Bonn, Bonn, North Rhine-Westphalia, Germany
[*]these authors contributed equally

Corresponding Author:
Niclas Braun, PhD
Department of Psychiatry and Psychotherapy
University Hospital Bonn
University of Bonn
Venusberg-Campus 1
Bonn, North Rhine-Westphalia, 53127
Germany
Phone: 49 228 31374
Email: niclas.braun@ukbonn.de

Abstract

Background: Over the past 2 decades, virtual reality–based neuropsychological tasks have gained traction as tools for objectively assessing symptoms of attention-deficit/hyperactivity disorder (ADHD), offering enhanced ecological validity by simulating naturalistic environments. To complement realistic settings with an ecologically valid task, we recently developed the virtual email sorting task (VEST), which immerses participants into an office environment where they sort emails while being exposed to distractors.

Objective: This study examined for the first time the VEST's sensitivity to medication effects and its specificity in differentiating ADHD from other psychiatric disorders that share overlapping cognitive symptoms, such as major depressive disorder (MDD).

Methods: A total of 23 unmedicated individuals with ADHD, 23 medicated individuals with ADHD, and 16 unmedicated individuals with MDD completed the VEST. During alternating distractor phases (DP) and nondistractor phases (NDP), we recorded the participants' task performance; head, torso, and leg actigraphy; eye movements; and brain activity using functional near-infrared spectroscopy (fNIRS), and subjective symptom ratings. Correlational analyses of main objective and subjective task-related parameters were computed. Data were analyzed using mixed-design ANOVAs.

Results: Processing time variability increased over time in participants with MDD and unmedicated ADHD as indicated by a group × block interaction (P=.04; η^2_p=0.10), while a group × phase interaction (P=.009; η^2_p=0.15) revealed that medicated participants with ADHD showed an increase during DP compared to NDP. Moreover, both ADHD groups exhibited increased head movements during DP compared to NDP (trend group × phase interaction: P=.09; η^2_p=0.08), an effect not observed in the MDD group, and higher rotation during DP in unmedicated individuals with ADHD (P<.001; η^2_p=0.23). Also, scores in 3 out of 4 subjective symptom intensity ratings of inattention, impulsivity, and emotional dysregulation were higher in at least 1 of the ADHD groups compared to the MDD group. No significant group differences were found in actigraphy measures of the arm and torso, fNIRS brain activity, or eye-tracking data. Regarding correlational analyses, inattention was correlated to off-task gaze (r=0.28; P=.03), hyperactivity with mean processing time (r=0.33; P=.01) and head movement (r=0.35; P=.006), and impulsivity with error rate (r=0.35; P=.006), and various significant correlations between objective parameters were found.

Conclusions: Our findings highlight the potential of the VEST to differentiate between ADHD and MDD, as well as to detect medication-related effects within ADHD. The results underscore the value of multimodal and ecological assessment approaches with distractors in the evaluation of attentional and behavioral symptoms. The VEST may offer a standardized way to investigate

XSL•FO
RenderX

complex behavior in mental disorders in research settings and, potentially, in clinical practice. However, further studies with greater statistical power are needed to confirm these findings.

Trial Registration: German Clinical Trials Register DRKS00031259; https://www.drks.de/search/de/trial/DRKS00031259/details

(JMIR Hum Factors 2026;13:e85351) doi:10.2196/85351

KEYWORDS

actigraphy; adult ADHD; depression; eye tracking; hyperactivity; impulsivity; inattention; medication; virtual reality; diagnosis

Introduction

For attention-deficit/hyperactivity disorder (ADHD), a neurodevelopmental disorder with an estimated prevalence of 5% in children and 2.58% in adulthood [1,2], the search for assessment tools to objectively assess its core symptoms of inattention, impulsivity, and hyperactivity [3] has been ongoing. At present, to assess adult ADHD, clinical interviews and self-report questionnaires are primarily used [4,5]. While these methods provide the highest diagnostic accuracy to date, they are influenced by various confounding factors, such as varying introspective skills and recall bias [6,7]. To overcome these limitations, researchers have explored various measures such as neuropsychological tests [8] and biomarkers [9]. Although such measures have successfully distinguished individuals with ADHD from healthy controls (HC) at the group level, they still lack sufficient clinical precision for accurate diagnoses at the individual level [8,10,11]. Moreover, studies often focus only on one or two domains such as performance or brain activity [12,13], and therefore may fail to encompass the full complexity of the disorder. Hence, to further increase the accuracy of the objective assessment of symptoms of ADHD and to fully depict the disorder, the search for suitable tools has been ongoing.

In recent years, virtual reality (VR)–based assessments have gained increasing popularity in the assessment of mental disorders in general [14], and especially in ADHD [15]. As outlined in more detail in previous works [16], VR-based assessments offer several advantages over traditional neuropsychological tasks. First, VR setups provide immersive features, such as the presentation of complex 3D environments that participants can interact with in a highly standardized manner [17]. Second, VR environments, such as virtual classrooms, are considered to be more ecologically valid than the experimental setups typically used in conventional neuropsychological tests [17]. Third, to examine the ADHD symptom of increased distractibility [3], realistic auditory and/or visual distractors can be incorporated into VR scenarios, allowing researchers to study their impact on various outcome measures. Fourth, VR can be integrated with a variety of physiological measures, including brain imaging, eye tracking, and actigraphy, enabling a multimodal assessment covering both behavioral and physiological indicators of ADHD [18,19].

One of the first VR-based assessment paradigms making use of these advantages for adults with ADHD was the virtual seminar room (VSR) [16,20], which was developed by our research group and inspired by a similar paradigm for children with ADHD, the virtual classroom [21]. In both paradigms, participants have to perform a continuous performance task (CPT) projected to the front of the virtual environment on a whiteboard or chalkboard, while the surrounding level of distraction is systematically varied by phases of high distraction (distractor phases [DP]) and low distraction (nondistractor phases [NDP]). Using the CPT, a neuropsychologically validated task for measuring sustained attention and impulsivity, in which participants are required to respond to target stimuli (eg, letters) and suppress responses to nontarget stimuli [22], was integrated into an ecologically valid classroom setting [21].

The multimodal VSR allowed for the measurement of task performance, brain activity via electroencephalography and functional near-infrared spectroscopy (fNIRS), head actigraphy, eye movements, and subjective symptom intensity. Differences were found, for instance, in CPT performance and actigraphy between individuals with ADHD and HC [20]. Based on separate training and test sets generated across multiple VSR studies [20,23], participants could be classified as ADHD or HC with 81% accuracy [18]. However, while the VSR represents a successful first step toward a more ecologically valid VR-based assessment of adult ADHD, within this scenario, it continues to rely on the highly abstract CPT.

To address this, we developed the virtual office room [24]. In this paradigm, the virtual environment is displayed in an open space office room in which the participants sit at a desk in front of a virtual desktop and perform a virtual email sorting task (VEST). In contrast to the CPT applied in the VSR, the VEST follows a more function-led than construct-driven approach, as the task is designed to simulate everyday goal-directed behavior within a realistic work context to increase its generalizability to real-world behavior [17]. Hence, more realistic stimuli material and task design are used than in the CPT. Moreover, as in the VSR, participants are exposed to alternating DP and NDP. In a first study using the VEST [24], we found a higher PTV in the ADHD group than in the HC, and a higher increase in head movement from NDP to DP in ADHD.

Beyond the differentiation between adults with and without ADHD, VR-based neuropsychological tests may also assist in assessing the effects of pharmacological treatments or other therapeutic interventions. This could, for instance, help in determining whether a treatment should be continued or adjusted, and which specific symptoms are being positively or insufficiently impacted. Up to now, the effect of medication has been investigated in 3 studies in children/adolescents using VR-based neuropsychological tests. All studies applied a virtual classroom scenario and reported lower reaction time variability [25-27], lower omission errors [25,27], as well as lower reaction times, commission errors, and decreased head movements [25] in medicated participants compared to unmedicated participants. In adults, we found a trend effect between medicated and

XSL•FO
RenderX

unmedicated participants on omission errors in the VSR [20]. In summary, while VR-based neuropsychological tests have shown promising initial results in detecting medication effects, further research is needed, especially in adult populations.

To evaluate the potential of the VEST for ADHD in adulthood, it is crucial to also examine its capacity to differentiate individuals with ADHD from individuals with other psychiatric disorders that can involve similar cognitive impairments. As a first step in this approach, we decided to test a sample with major depressive disorder (MDD) in comparison to ADHD. In addition to the generally high comorbidity of the two disorders, they overlap in some cognitive domains [4,28], which is why it is of interest to see whether the multimodal procedure can be used to find parameter characteristics that are specific to ADHD. As of now, there have only been VR-unrelated studies investigating differences in neurocognitive function between individuals with ADHD and individuals with MDD. Potvin et al [29], for example, report similar cognitive performances of both groups in various measures of subjective and objective cognitive function, including working memory and processing speed, with lower performances than HC. Moreover, in a study by Paucke et al [30], individuals with MDD, with ADHD, and with both disorders, showed impairments in sustained attention compared to HC. In a study adopting the group design of Paucke et al [30] with a notably higher sample size, Van Hal et al [31] found impairments in attention and other executive functions in all clinical groups. With regard to actigraphy, while a plethora of studies have investigated movement in daily life in individuals with MDD [32], to the best of our knowledge, no studies have applied actigraphy measures during the performance of neuropsychological tasks. In general, MDD is associated with psychomotor disturbances such as psychomotor agitation and retardation [33], and individuals with MDD show lower activity than HC [32]. Considering differences in eye movement behavior in MDD compared to HC, studies have shown an increased maintenance of gaze on dysphoric and decreased maintenance of gaze on positive stimuli [34], in addition to slower reaction times in pro- and antisaccade tasks [35]. As of now, however, no studies have investigated the influence of distracting stimuli on task performance in MDD. When examining brain activity using fNIRS, the dorsolateral prefrontal cortex (dlPFC) appears to be of particular interest in ADHD, where reduced oxyhemoglobin (oxy-Hb) concentrations have been observed during inhibition tasks [36]. Similarly, studies in individuals with depression have also reported differences in activation levels in this brain region compared to HC during verbal fluency tasks [37]. Although both clinical groups show dlPFC hypoactivation across different cognitive tasks, it remains of interest to explore whether, particularly the component of distractor inhibition, reveals a pattern that is specific to ADHD compared to MDD and if so, how it is altered through medication.

The aim of this study was to investigate the VEST's sensitivity to medication effects and its potential to differentiate individuals with ADHD from those with other psychiatric disorders. To this end, we compared 3 groups: unmedicated individuals with ADHD, medicated individuals with ADHD, and unmedicated individuals with MDD. Specifically, we examined performance on the VEST, eye movement behavior, brain activity measured via fNIRS, motor activity assessed through actigraphy (head, torso, nondominant arm, and opposite leg to nondominant arm), and self-reported symptom intensity. In addition to group comparisons, we investigated distractibility and time-on-task effects.

Based on the existing literature, we hypothesized that both ADHD groups would exhibit greater head and body movement than the MDD group.

Based on the heterogeneity and limited number of prior studies, we also formulated two research questions:

1. Can the VEST differentiate on any outcome parameter between medicated and unmedicated individuals with ADHD?
2. Can the VEST differentiate on any outcome parameter between unmedicated individuals with ADHD and unmedicated individuals with MDD?

Methods

Ethical Considerations

The study received approval from the ethics committee of the Faculty of Medicine at the University of Bonn (protocol number 105/23) and was preregistered on July 20, 2023, in the German Clinical Trials Register (DRKS00031259). All participants were informed in advance about the study's objective and procedures, the use of technical equipment, and possible risks of participation such as motion sickness, and they provided written consent before taking part. Personal data such as names and addresses were recorded only on paper documents, including the consent form and the reimbursement form. A study ID was used for all data collection. Participants received a compensation of €40 (US $47.5) for full participation. Individuals who had to be excluded after the first session or who dropped out received prorated compensation.

Participants

A total of 63 individuals participated in the study, including 24 medicated adults with ADHD, 23 currently unmedicated adults with ADHD, and 16 unmedicated adults with MDD. The participants were recruited in various ways: direct approach of suitable individuals at the ADHD outpatient clinic of the University Hospital Bonn, advertisements on the internal website, posting flyers in outpatient clinics, writing to self-help groups, and a press release followed by a radio report. Participants could take part in the study if they fulfilled either ADHD or MDD criteria but no other severe psychiatric disorder, such as psychosis or severe addiction disorder. Participants were not screened for comorbid autism spectrum disorder, as autism spectrum disorder did not constitute an exclusion criterion and its prevalence in the population with ADHD is markedly lower in adults compared to children and adolescents [38]. Furthermore, participants had to be aged between 18 and 50 years, have normal or corrected-to-normal vision, and no history of epilepsy or other severe neurological disorders, pregnancy, or skin irritation on the head.

XSL•FO
RenderX

General Procedure

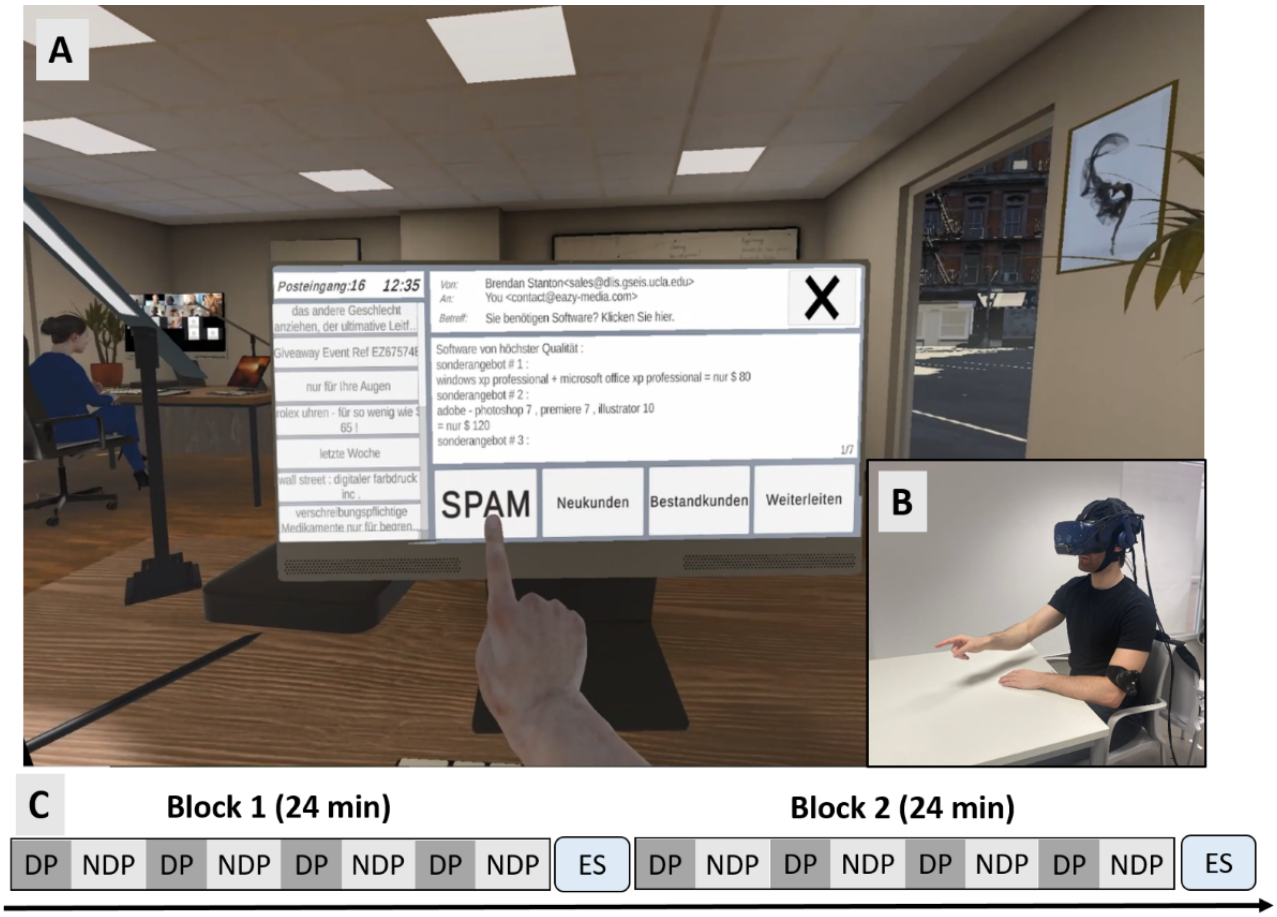
The study consisted of 2 separate appointments. During the first appointment, participants underwent a clinical assessment after providing informed consent. The assessment had 2 main objectives. First, to confirm the diagnoses of ADHD or MDD, depending on which group the participant should be assigned to, and to rule out the respective other condition. Second, to identify potential comorbidities that signified as an exclusion criterion. The assessment lasted approximately 1-2 hours and was conducted either in person at the psychiatry department of the University Hospital of Bonn or via a video call using the online platform Red Connect [39]. Two clinical interviews were administered: the Diagnostic Brief Interview for Mental Disorders (Mini-DIPS) [40] to screen for 17 mental disorders and the Integrated Diagnosis of ADHD in Adulthood [41] to confirm or rule out the diagnosis of ADHD. Both interviews are based on the criteria of the *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders* [Fifth Edition]) and *ICD-10* (*International Statistical Classification of Diseases, 10th Revision*) and were conducted in German. After the clinical interviews, participants completed various self-rating scales via the online survey tool SoSci Survey [42], including the Patient Health Questionnaire-9 (PHQ-9) [43], and ADHD Self-Rating Scale (ADHS-SB) [44]. Additionally, a self-constructed questionnaire was administered to collect various sociodemographic data and the medication status. If no exclusion criteria applied during the initial assessment, participants were invited to the second appointment. During this session, the actual VR experiment took place in the VR laboratory of the University Hospital of Bonn. Once prepared, 3 motion trackers were affixed to the participants. Next, they were seated, and preparations for fNIRS recording commenced. Prior to the start of data acquisition, participants were fitted with the head-mounted display (HMD), and the VR environment was launched. Within the VR, a brief eye-tracking calibration was conducted, followed by a 30-second acclimation period to allow participants to familiarize themselves with the virtual setting. After that, participants received task instructions and completed a short practice block, before the actual experiment started. This consisted of 2 blocks, each lasting 24 minutes. Symptom assessments were conducted following both blocks, and the Virtual Reality Sickness Questionnaire (VRSQ) [45] was administered at the conclusion of the session. Throughout the experiment, the software Lab Streaming Layer (Christian Kothe) [46] was used to record and synchronize all data streams, including task performance, fNIRS, actigraphy, and eye-tracking data. Following completion of the experiment, all measurement devices were removed, and participants were debriefed and dismissed. Since the same study supervisor conducted both the diagnostic assessment and the VR session, the study did not include experimenter blinding.

Apparatus and Virtual Environment

The VR scenario was presented via the HTC VIVE Pro Eye HMD (HTC Corporation). The VR scenario (Figure 1) was created using Unity 3D (version 2019.4.11f1; Unity Technologies) and coded in C# [24]. To enable controller-free interaction with the virtual environment, a Leap Motion Controller (Ultraleap) was used to track participants' hand movements and online transfer them to a virtual 3D hand model ("Leap Motion Realistic Hands," Unity Asset Store). The VR environment, entered via the HMD, depicted an open-plan office equipped with typical office supplies, furniture, and virtual colleagues (Figure 1; for further specifications of hardware and software, see Multimedia Appendix 1 [47,48]). In front of the participant, the VEST was displayed on a virtual computer screen.

XSL•FO
RenderX

Figure 1. Technical setup, virtual environment, and experimental design. (A) First-person view of the virtual email sorting task in a virtual office. (B) Participant during the experimental session taking place in the virtual reality laboratory of the University Hospital Bonn. (C) Experimental timeline. DP: distractor phase; ES: experience sampling; NDP: nondistractor phase.



VEST

During the experiment, participants had to perform the VEST, a task that we developed and piloted in a previous study [24]). Compared to conventional CPTs, which often involve abstract stimuli, the VEST uses stimuli that mirror real-world office activities. In this task, participants assumed the role of a secretary in a marketing agency, where they were required to categorize incoming emails into 1/4 predefined categories: “spam,” “new customer,” “existing customer,” and “forward.” To ensure comparability with classical CPT paradigms, the task was structured such that 65% (389/599) of emails belonged to the spam category, making it the default response. Emails classified as “new customer” and “existing customer” each comprised 15% (90/599) of the total emails, while 5% (30/599) of emails fell into the “forward” category. Emails classified as “spam” were sourced from the Enron Spam Dataset (“Enron Spam Dataset,” 2015) but translated into German before usage. Emails for the remaining categories were composed by members of our research group and written from the perspective of a fictitious company or private individual and included both a subject line and a request relevant to a marketing agency. The emails contained clear identifying features corresponding to their respective categories, ensuring that each email had only one correct classification. In total, 599 mails were used in the VEST and presented in a randomized order.

The emails were displayed on a virtual computer screen in front of the participants (Figure 1). The screen simulated a touch interface, allowing participants to interact with it by using their virtual hands to select the appropriate category for each email (Figure 1). Emails did not expire after a certain time limit, so participants performed the task at their own pace—once an email was categorized, the next email would automatically open. Each block consisted of 4 DP and 4 NDP, each lasting 3 minutes and alternating. Consequently, over the course of the entire experiment, participants experienced 8 DP and 8 NDP. During each DP, 6 visual, auditory, or audiovisual distractors were presented in succession, each lasting 30 seconds. These distractors were randomly selected from a pool of 48 potential distractors. The order of the first phase in each block (ie, whether it was a DP or NDP) was counterbalanced across participants.

To assess the participants’ performance on the VEST, 3 distinct parameters were separately extracted for each DP and NDP: the average processing time and PTV (calculated by using the coefficient of variation) per email as measures of vigilance, and the error rate of misclassified emails as an indicator of inattention and impulsivity. For the statistical analysis, the mean values of each parameter were calculated across all DP and NDP per block.

Experience Sampling

To evaluate participants’ subjective symptom experience, a VR-embedded user interface controlled via gesture input was

XSL•FO
RenderX

presented immediately before block 1, between the 2 blocks, and at the end of block 2. For each block, participants were asked to rate their momentary levels of inattention, impulsivity, and hyperactivity [18,20,24], as well as their emotional dysregulation. The items were rated on a 7-point Likert scale, ranging from –3 to +3. For statistical analysis, only the subjective symptom reports collected after block 1 and block 2 were used. Following the final subjective symptom report, the VRSQ was presented on the same interface and Likert scale. The VRSQ includes 8 items assessing potential physical discomfort from VR exposure, from which a total score is calculated [45].

Actigraphy Recording and Analyses

During task execution in the VR environment, the movements of the participant's head, torso, arm, and leg were recorded. Head movement data were acquired via the HMD, with its position and rotation continuously tracked in the 3D space defined by the base stations. The movements and rotation of the torso, arm, and leg were captured using HTC Vive trackers (HTC Corporation), each of which was also tracked in the 3D space relative to its own dedicated base station. The placement of the trackers depended on the participants' handedness, that is, for right-handed participants, 1 tracker was attached to the left upper arm, 1 around the torso, and 1 to the right ankle. For left-handed participants, the position of the ankle and upper arm were reversed. In line with our previous studies [16,20,24], all motion data were processed using a uniform analytical approach. First, the data were down sampled to ~10 Hz. Subsequently, the Euclidean distance between successive data points was calculated to quantify the change in position and rotation over time. The mean changes per task phase were then computed to summarize movement dynamics.

Eye Tracking Recording and Analyses

To track the participant's horizontal and vertical eye movements, the Tobii eye tracker (Tobii Technology) built into the HTC Vive Pro Eye was used (for further specifications of hardware and software, see Multimedia Appendix 1). Here, we tracked the participant's duration of momentary gaze focus on the canvas on which the task was displayed on and the duration spent looking away from the canvas (off-task gaze) [24]. The momentary gaze focus time spent off-task was calculated for each NDP and DP.

fNIRS Recording and Analyses

As in previous studies [16,20,24], the participants' fNIRS brain activity was recorded during task performance by the NIRSport 2 system (NIRx Medical Technologies) and acquisition software Aurora fNIRS (NIRx Medical Technologies). We focused on oxy-HB in the dlPFC tracked by 2 channels (Fc1-Fc3, Fc2-Fc4; see also [20]), taking previous studies on ADHD using fNIRS [36] and spatial constraints into consideration (for further specifications of hardware and software used for recording and preprocessing, see Multimedia Appendix 1). For following statistical analyses, DP and NDP were first split into six 30-second units, equivalent to the length of distractor intervals in DP. Next, block averages across all units per phase were determined, and mean concentration of oxy-Hb was calculated for each phase. Lastly, each NDP was subtracted from its corresponding DP, and mean DP-NDP contrasts across all DP-NDP pairs were determined for each participant.

Missing Data Handling

One participant was not able to conclude the experiment due to motion sickness in the second block and was excluded from analysis. For the VEST performance and head actigraphy, all remaining 62 datasets were complete and included; for eye tracking analysis, the sample was reduced to 61 datasets. Due to intermittent data transfer issues with the body trackers, some datasets were incomplete. If at least 3 out of 4 phases were successfully recorded, the phase average for each block was calculated using the 3 available values. However, datasets with more than 1 missing phase were excluded from the analysis. Using this approach, 58 datasets were included in the analysis of arm movements and 45 datasets in the analysis of torso movements. For leg movements, only 35 datasets remained, which led to the exclusion of this parameter from the analysis. Regarding the fNIRS data, 33 datasets were excluded for the left dlPFC and 39 for the right dlPFC due to insufficient data quality, leading to a reduced sample of n=29 and n=23, respectively. Since fNIRS was the only method used for analyzing brain activity in this study, the remaining datasets were still analyzed despite the decreased sample size. Multimedia Appendix 2 [49] presents descriptive summaries of missing data by modality and binary logistic regression analyses assessing whether missingness was associated with group, gender, and age.

Statistical Analysis

For the majority of parameters of interest (eg, error rate; processing time; PTV; head, torso, and arm rotation and position; and off-task gaze), separate 3×2×2 mixed-design ANOVAs were conducted, with group as a between-subjects factor and block and phase as within-subject factors. For the experience sampling data, a 3×2 mixed-design ANOVA was performed with group and block as factors. Analyses of fNIRS and VRSQ data were conducted using 1-way ANOVAs with group as the factor. Furthermore, we conducted an exploratory correlational analysis of the objective parameters and subjective, task-related symptom intensity. For this purpose, we used aggregated values across phases and blocks. Because of the exploratory nature of the correlational analysis, we decided not to correct for multiple testing. Statistical significance was set at $\alpha<.05$ and α between .05 and .10 was considered as a statistical trend level for all calculations.

Due to right-skewed distributions, several variables (eg, error rate, off-task gaze, and actigraphy measures for arm and torso) were subjected to square root transformations prior to analysis. Inattention scores exhibited left skewness; thus, these data were first mirrored and then square root transformed. Homogeneity of variances was assessed using Levene test, with $P>.05$ indicating acceptable homogeneity for all variables except inattention, for which homogeneity was not achieved despite transformation.

To control type I error inflation due to multiple testing, Bonferroni corrections were applied to the main effect of the

XSL•FO
RenderX

3-level group factor in the ANOVA, to interaction terms, and to all post hoc group comparisons. Effect sizes were interpreted according to Cohen [50] with partial eta squared (η^2_p) values of 0.01, 0.06, and 0.14 representing small, medium, and large effects, respectively for ANOVA results and Cohen *d* with values of 0.2, 0.5, and 0.8 representing small, medium, and large effects for post hoc comparisons.

Results

All key ANOVA results are summarized in Multimedia Appendix 3.

Sample

The sample consisted of 46 participants with ADHD, equally divided between the medicated and unmedicated groups, and 16 individuals with MDD. The 3 groups did not significantly differ in terms of gender, age, handedness, or level of education. Regarding self-reported clinical measures, the groups only differed on those that were specific to ADHD. Further demographic details can be found in Table 1.

XSL•FO
RenderX

Table 1. Demographic and clinical characterization of the 3 groups. *P* values indicate statistical group comparisons (1-way ANOVA or chi-square test).

Characteristics	Medicated ADHD[a] (n=23)	Unmedicated ADHD (n=23)	MDD[b] (n=16)	*P* value
Demographics				
Age (years), mean (SD; range)	29.65 (7.30; 22-44)	34.74 (10.59; 19-50)	31.75 (8.48; 20-50)	.16
Gender, n (%)				.07
Men	9 (39.1)	9 (39.1)	2 (12.5)	
Women	14 (60.9)	14 (60.9)	12 (75)	
Nonbinary	N/A[c]	N/A	2 (12.5)	
Handedness, n				.56
Right	21	22	14	
Left	1	1	2	
Both	1	N/A	N/A	
Education, n				.50
Lower secondary education	4	6	6	
Higher secondary education	11	12	5	
University degree	8	8	5	
IDA-R[d] subscore, mean (SD)				
Inattention	18.43 (5.46)	20.96 (3.32)	8.56 (4.93)	<.001[e]
Impulsivity	6.78 (3.32)	8.00 (3.77)	2.44 (2.40)	<.001[e]
Hyperactivity	7.96 (3.20)	9.78 (4.13)	3.94 (2.89)	<.001[e]
ADHS-SB[f] subscore, mean (SD)				
Inattention	14.74 (4.96)	17.35 (4.42)	6.81 (4.42)	<.001[e]
Impulsivity	5.22 (3.68)	6.48 (3.26)	1.94 (1.84)	<.001[g]
Hyperactivity	7.30 (3.27)	8.96 (4.61)	4.56 (2.94)	.003[e]
ADHD medication				
None, n (%)	N/A	23 (100)	16 (100)	N/A
Methylphenidate, n (%)	10 (43.48)	N/A	N/A	N/A
Lisdexamfetamine, n (%)	8 (34.78)	N/A	N/A	N/A
Nonstimulant, n (%)	3 (13.04)	N/A	N/A	N/A
Other, n (%)	2 (8.70)	N/A	N/A	N/A
PHQ-9[h], mean (SD)	7.78 (5.57)	8.61 (4.95)	10.50 (7.40)	.37
BSL[i], mean (SD)				
BSL-23	38.65 (15.92)	46.91 (14.91)	43.88 (21.13)	.26
BSL-VA[j]	59.96 (23.31)	60.13 (22.41)	56.07 (28.55)	.87
BSL-11	12.13 (1.82)	13.13 (2.65)	12.63 (1.96)	.31
WHO-QOL-BREF[k], mean (SD)				
Physical health	68.63 (15.71)	69.72 (13.27)	69.20 (14.34)	.97
Psychological health	57.61 (17.21)	56.52 (15.83)	52.34 (21.51)	.65
Social relationships	64.13 (20.17)	60.87 (22.39)	63.02 (19.94)	.87
Environmental health	75.82 (14.14)	73.10 (13.18)	71.68 (12.58)	.61

XSL•FO
RenderX

[a]ADHD: attention-deficit/hyperactivity disorder.

[b]MDD: major depressive disorder.

[c]N/A: not applicable.

[d]IDA-R: Integrated Diagnosis of ADHD in Adulthood.

[e]Post hoc comparisons revealed a significant difference between both ADHD groups and MDD, but not between the ADHD groups.

[f]ADHS-SB: ADHD Self-Rating Scale.

[g]Post hoc comparisons revealed a significant difference only between the unmedicated ADHD group and the MDD group.

[h]PHQ-9: Patient Health Questionnaire-9.

[i]BSL: Borderline Symptom List.

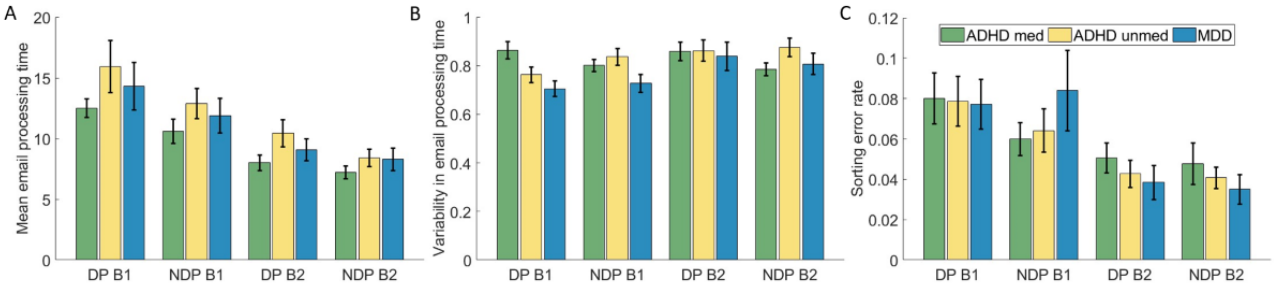
[j]BSL-VA: Borderline Symptom List-Visual Analogue Scale.

[k]WHO-QOL-BREF: WHO Quality of Life Short Version.

VEST Performance

Performance data of the VEST are presented in Figure 2. The analysis of mean processing time revealed significant main effects of block ($F_{1,59}$=68.22; P<.001; η^2_p=0.54) and phase ($F_{1,59}$=23.99; P<.001; η^2_p=0.29), but not of group ($F_{2,59}$=1.49; P=.24). Mean processing time was higher in block 1 than block 2 (mean difference [MDiff]=4.48, 95% CI 3.40-5.57) and higher in DP than NDP (MDiff=1.86, 95% CI 1.10-2.61). In addition, there was a significant interaction of block × phase ($F_{1,59}$=4.00; P=.05; η^2_p=0.06), which was not further interpreted.

Figure 2. Results of the virtual email sorting task. (A) Mean processing time in seconds. (B) Processing time variability. (C) Error rate. Error bars indicate the standard error of the mean. ADHD: attention-deficit/hyperactivity disorder; B1: block 1; B2: block 2; DP: distractor phase; MDD: major depressive disorder; med: medicated; NDP: nondistractor phase; unmed: unmedicated.



For PTV, the analysis yielded a significant main effect of block ($F_{1,59}$=8.39; P=.005; η^2_p=0.12), indicating a higher PTV in block 2 than block 1 (MDiff=–0.05, 95% CI –0.09 to –0.02). There were no significant main effects of phase ($F_{1,59}$=0.46; P=.50) and group ($F_{2,59}$=1.18; P=.32). However, the analysis revealed significant interactions of block × group ($F_{2,59}$=3.35; P=.04; η^2_p=0.10), phase × group ($F_{2,59}$=5.09; P=.009; η^2_p=0.15) and a trend for block × phase ($F_{1,59}$=3.79; P=.06; η^2_p=0.06), which was not further interpreted. In the pairwise comparisons, we found a significantly higher PTV in block 2 than 1 in the unmedicated ADHD group (P=.03; d=0.39; MDiff=–0.07, 95% CI –0.13 to –0.01) and in the MDD group (P=.005; d=0.95; MDiff=–0.11, 95% CI –0.18 to –0.03), but not in the medicated ADHD population (P=.73). For phase × group, pairwise comparisons revealed a significant difference between DP and NDP in the medicated ADHD group (P=.02; d=0.78) with higher PTV in DP than NDP (MDiff=0.07, 95% CI 0.02-0.12), while for the unmedicated ADHD group (P=.09) and the MDD group (P=.88) there was no significant difference.

The analysis of square root–transformed error rates yielded a significant main effect of phase ($F_{1,59}$=41.12; P<.001; η^2_p=0.41), indicating a higher error rate in DP than NDP (MDiff=0.02, 95% CI 0.01-0.02). Apart from that, no significant main effects of block ($F_{1,59}$=1.79; P=.19) or group ($F_{2,59}$=0.04; P=.96) or significant interactions were found.

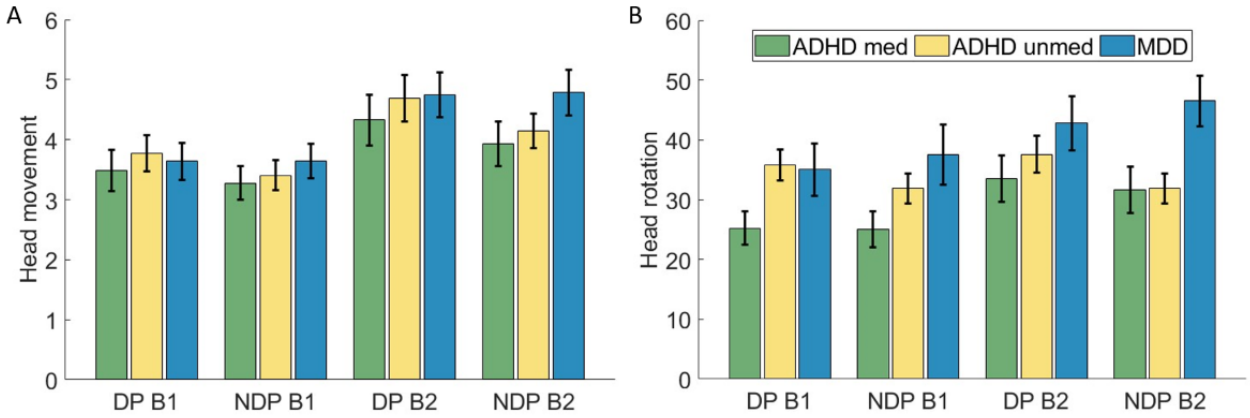
Actigraphy Results

Head

Results of head actigraphy analyses are depicted in Figure 3. The ANOVA for head movement revealed significant effects of block ($F_{1,59}$=60.12; P<.001; η^2_p=0.51) and phase ($F_{1,59}$=23.99; P=.005; η^2_p=0.13), but not of group ($F_{2,59}$=0.48; P=.62). Movement was greater in block 2 compared to block 1 (MDiff=–0.90, 95% CI –1.13 to –0.67), and higher in the DP condition than in NDP (MDiff=0.25, 95% CI 0.08-0.42). A trend toward a phase × group interaction was also observed ($F_{2,59}$=2.56; P=.09; η^2_p=0.08). Pairwise comparisons indicated significantly increased movement in DP compared to NDP for both ADHD groups (medicated: P=.03; d=0.42; MDiff=0.30, 95% CI 0.03-0.59; unmedicated: P=.001; d=0.66; MDiff=0.46, 95% CI 0.19-0.73), whereas no significant difference was found in the MDD group (P=.89).

XSL•FO
RenderX

Figure 3. Results of head actigraphy analysis. (A) Head movement. (B) Head rotation. Error bars indicate the standard error of the mean. ADHD: attention-deficit/hyperactivity disorder; B1: block 1; B2: block 2; DP: distractor phase; MDD: major depressive disorder; med: medicated; NDP: nondistractor phase; unmed: unmedicated.



For head rotation, a significant main effect of block emerged ($F_{1,59}$=17.33; P<.001; η^2_p=0.23), with greater rotation in block 2 than block 1 (MDiff=–6.25, 95% CI –9.25 to –3.25). Also, there was a trend toward a group effect ($F_{2,59}$=3.08; P=.05; η^2_p=0.10), though no significant effect of phase ($F_{1,59}$=0.20; P=.66). Pairwise comparisons showed a trend for a difference between medicated individuals with ADHD and those with MDD (P=.05; d=0.72; MDiff=11.63, 95% CI –0.08 to 23.34). Additionally, a significant phase × group interaction was detected ($F_{2,59}$=8.81; P<.001; η^2_p=0.23). Here, pairwise comparisons revealed significantly higher rotation in DP than NDP for the unmedicated ADHD group (P=.003; d=0.56; MDiff=2.83, 95% CI 1.00-4.65). By contrast, the reverse was observed for the MDD group (P=.006; d=0.65; MDiff=–3.10, 95% CI –5.29 to –0.91), and no significant difference was observed for the medicated ADHD group (P=.27).

Arms

Analysis of the square root-transformed arm movement and rotation data showed a significant block effect for both, arm movements ($F_{1,55}$=38.93; P<.001; η^2_p=0.41) and arm rotations ($F_{1,55}$=13.29; P<.001; η^2_p=0.20), with greater activity levels in block 2 than 1 (movement: MDiff=–0.30, 95% CI –0.40 to –0.21; rotation: MDiff=–0.30, 95% CI –0.46 to –0.13). No significant main effects were found for phase or group in either case (movement: phase with $F_{1,55}$=0.02; P=.90, group with $F_{1,55}$=0.53; P=.59; rotation: phase with $F_{1,55}$=0.31; P=.58, group with $F_{1,55}$=1.73; P=.19). For arm rotation, an additional interaction between block and phase was observed ($F_{1,55}$=5.41; P=.02; η^2_p=0.09), but not further explored.

Torso

For square root–transformed torso movement, the ANOVA revealed a significant main effect of block ($F_{1,42}$=14.73; P<.001; η^2_p=0.26) indicating increased movement in block 2 compared to block 1 (MDiff=–0.15, 95% CI –0.24 to –0.07). No significant main effects of phase ($F_{1,42}$=0.01; P=.91) or group ($F_{1,42}$=0.32; P=.73) were observed.

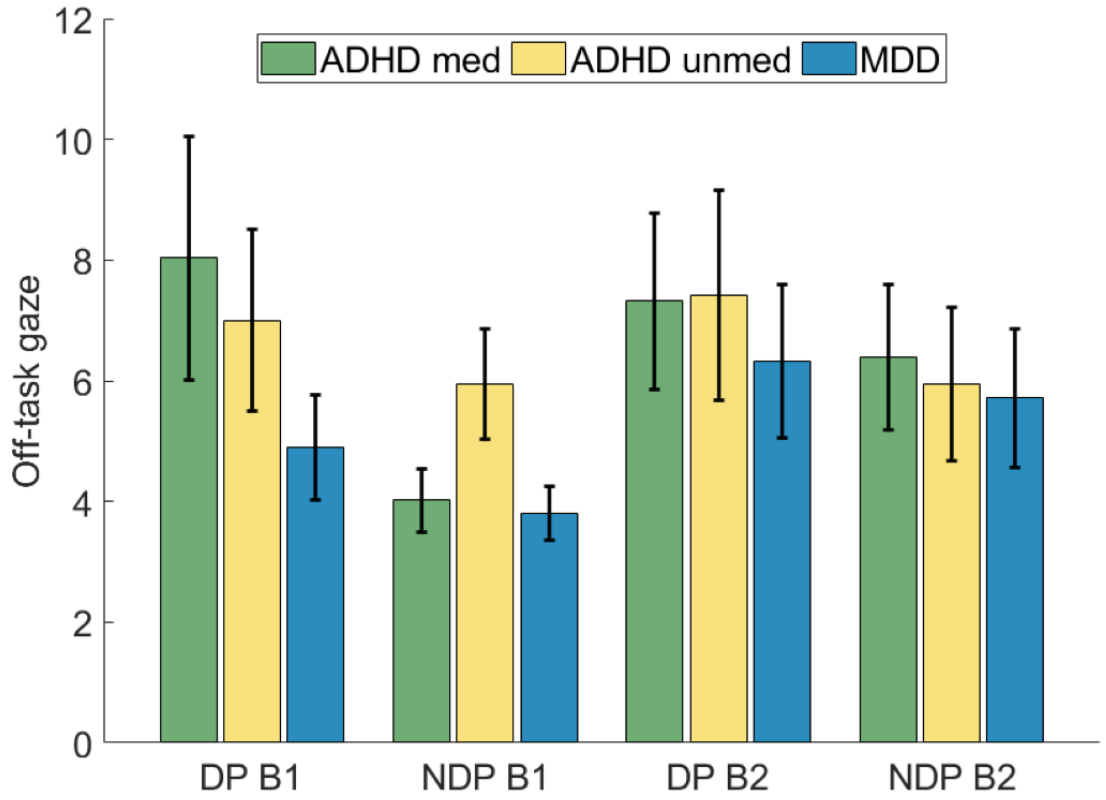
For rotation, the ANOVA revealed a significant main effect of phase ($F_{1,42}$=4.30; P=.04; η^2_p=0.09), with greater rotation observed during the DP than NDP (MDiff=0.10, 95% CI 0.00-2.00). No significant effects of block ($F_{1,42}$=0.56; P=.46), group ($F_{1,42}$=0.68; P=.51), or their interactions were found.

Eye Tracking

Eye tracking data are depicted in Figure 4. The ANOVA for square root–transformed off task gaze data revealed a significant main effect of phase ($F_{1,58}$=4.72; P=.03; η^2_p=0.075), in that across groups, there was a higher off task gaze in DP than NDP (MDiff=0.31, 95% CI 0.02-0.59). Apart from that, no main effect of block ($F_{1,58}$=0.50; P=.48) or group ($F_{1,58}$=1.16; P=.32) and no significant interactions were observed.

XSL•FO
RenderX

Figure 4. Results of the eye tracking analysis (off-task gaze time). Bar plot depicts logarithmically transformed data. Error bars indicate the standard error of the mean. ADHD: attention-deficit/hyperactivity disorder; B1: block 1; B2: block 2; DP: distractor phase; MDD: major depressive disorder; med: medicated; NDP: nondistractor phase; unmed: unmedicated.



fNIRS

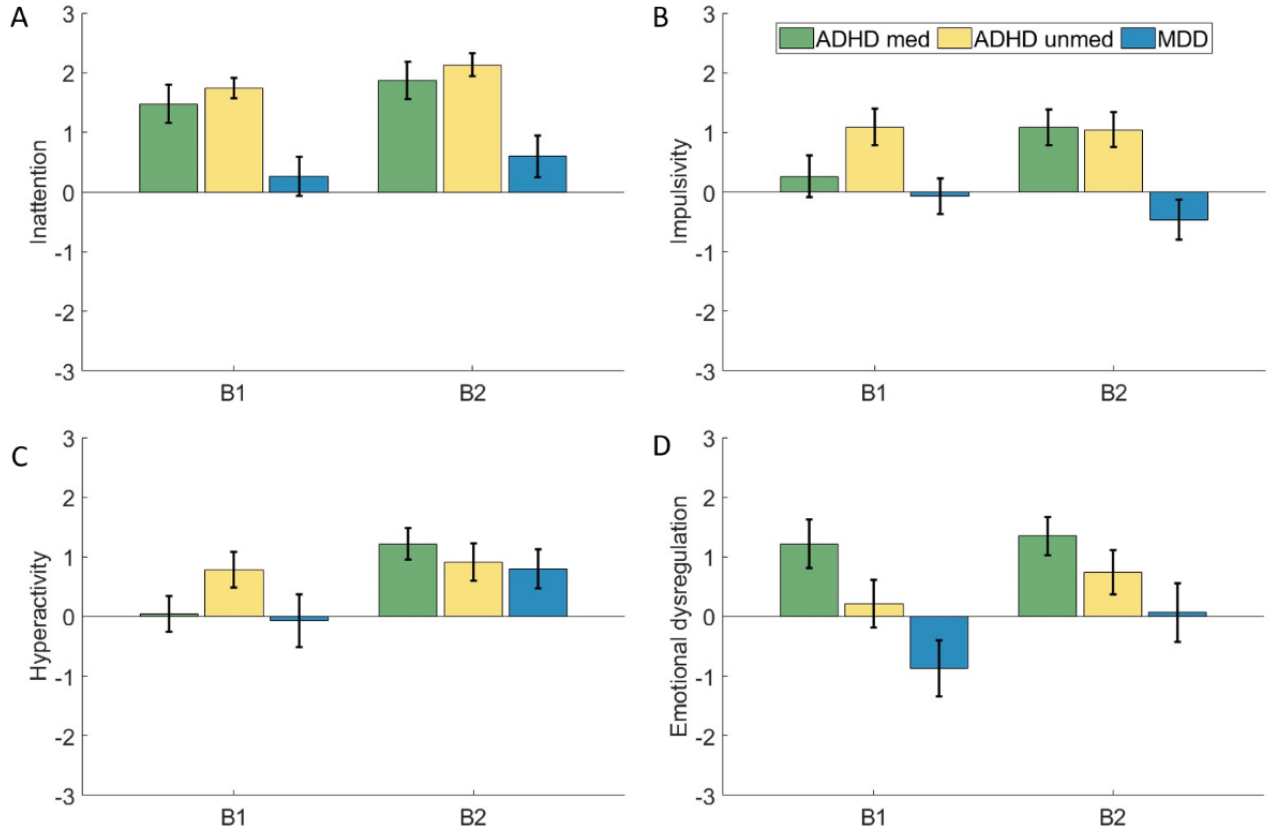
One-way ANOVAs revealed no significant group differences in DP-NDP oxy-Hb contrast on either channel of interest with $F_{1,28}$=0.29; P=.73; η^2_p=0.02 for the left dlPFC and $F_{1,22}$=0.06; P=.94; η^2_p=0.07 for the right dlPFC.

Experience Sampling

Results of the participants' reported ADHD symptom intensity during and after task performance are depicted in Figure 5. The ANOVA of self-reported square root–transformed inattention intensity revealed a significant main effect of block ($F_{1,58}$=9.52; P=.003; η^2_p=0.14), with higher scores in block 2 than 1 (MDiff=–0.09, 95% CI –0.03 to –0.15). A significant main effect of group also emerged ($F_{1,58}$=9.28; P<.001; η^2_p=0.24). Pairwise comparisons showed significantly higher inattention scores in both medicated (P=.007; d=0.91; MDiff=0.27, 95% CI 0.48-0.06) and unmedicated (P<.001; d=1.68; MDiff=0.36, 95% CI 0.57-0.15) participants with ADHD compared to those with MDD, while the ADHD groups did not differ significantly (P=.69). No significant block × group interaction was found.

XSL•FO
RenderX

Figure 5. Results of experience sampling. (A) Inattention rating after block 1 and block 2. (B) Impulsivity rating after block 1 (B1) and block 2 (B2). (C) Hyperactivity rating after B1 and B2. (D) Emotional dysregulation rating after B1 and B2. Error bars indicate the standard error of the mean. ADHD: attention-deficit/hyperactivity disorder; MDD: major depressive disorder; med: medicated; unmed: unmedicated.



Regarding subjective impulsivity, a significant main effect of group was found ($F_{2,58}$=6.16; P=.004; η^2_p=0.18). Post hoc comparisons indicated that unmedicated participants with ADHD reported significantly higher impulsivity than participants with MDD (P=.003; d=1.22; MDiff=1.40, 95% CI 0.42-2.38). Additionally, there was a trend (P=.10; d=0.73; MDiff=0.88, 95% CI –0.11 to 1.86) for higher impulsivity scores in medicated participants with ADHD than in participants with MDD, while the ADHD groups did not differ significantly (P=.44). No significant main effect of block was found ($F_{1,58}$=0.38; P=.54), nor was there a significant interaction between group and block.

For self-reported hyperactivity, a significant main effect of block was observed ($F_{1,58}$=16.87; P<.001; η^2_p=0.22), indicating higher hyperactivity levels in block 2 compared to block 1 (MDiff=–0.72, 95% CI –1.08 to –0.37). There was no significant main effect of group ($F_{2,58}$=0.54; P=.59). However, a trend-level interaction between block and group emerged ($F_{2,58}$=2.53; P=.09; η^2_p=0.08). Pairwise comparisons revealed significant increases in hyperactivity from block 1 to block 2 in the medicated ADHD group (P<.001; d=0.68; MDiff=–1.09, 95% CI –1.65 to –0.52) and MDD group (P=.02; d=0.70; MDiff=–0.87, 95% CI –1.56 to –0.17), whereas hyperactivity levels remained stable across blocks in the unmedicated ADHD group (P=.44).

For emotional dysregulation, ANOVA revealed a significant main effect of block ($F_{1,58}$=9.20; P=.004; η^2_p=0.14), with higher scores observed after block 2 compared to block 1 (MDiff=–0.53, 95% CI –0.88 to –0.18). A significant main effect of group also emerged ($F_{1,58}$=3.74; P=.03; η^2_p=0.11). Post hoc comparisons indicated that the medicated ADHD group reported significantly higher emotional dysregulation than the MDD group (P=.03; d=0.94; MDiff=1.55, 95% CI 0.15-2.96). No significant differences were found between the two ADHD groups (P=.86) or between the unmedicated ADHD and MDD groups (P=.24).

VR Sickness

Across groups, mean VRSQ scores were in a low positive range (mean 0.31, SD 1.23), indicating a low level of motion sickness during the experiment. The conducted 1-way ANOVA revealed no significant differences between groups ($F_{2,60}$=1.04; P=.36).

Correlation Analyses Between Modalities

Core findings of the correlational analysis are presented in Table 2, whereas results of all computed correlations with respective sample sizes are available in Multimedia Appendix 4. Regarding associations between task-related symptom intensity and psychophysiological markers, inattention was correlated with off-task gaze (r=0.28; P=.03). Hyperactivity correlated with both mean error rate (r=0.34; P=.008) and mean processing time (r=0.33; P=.01), as well as with head movement (r=0.35; P=.006) and head rotation (r=0.30; P=.02). Analyses further

XSL•FO
RenderX

showed a correlation between impulsivity and mean error rate (r=0.35; P=.006), and between emotional dysregulation and off-task gaze (r=0.29; P=.02). For the correlational analysis between the objective measures, we found, among other things, a significant correlation between off-task gaze and mean PTV (r=0.33; P=.01), head movement (r=0.43; P<.001), and arm movement (r=0.40; P=.002). Additionally, to its significant correlation to off-task gaze, mean PTV was also found to be correlated to head movement (r=0.35; P=.006) and rotation (r=0.26; P=.04). Furthermore, we also found a significant association between head rotation, mean processing time (r=0.30; P=.02), and head movement (r=0.51; P<.001).

Table 2. Correlation analysis among the study variables.

Variable	Mean PTV[a]	Mean error	Mean PT[b]	Off-task gaze	Head movement	Head rotation	Impulsivity	Hyperactivity	Inattention	Emotional dysregulation
Mean PTV										
r	—[c]									
P value	—									
Mean error										
r	0.079	—								
P value	.54	—								
Mean PT										
r	0.174	0.069	—							
P value	.18	.60	—							
Off-task gaze										
r	0.325	0.251	0.056	—						
P value	.01	.05	.67	—						
Head movement										
r	0.348	0.080	0.236	0.423	—					
P value	.006	.54	.07	<.001	—					
Head rotation										
r	0.262	0.038	0.297	0.069	0.505	—				
P value	.04	.77	.02	.60	<.001	—				
Impulsivity										
r	0.115	0.346	-0.066	0.181	–0.006	0.004	—			
P value	.38	.006	.62	.16	.96	.97	—			
Hyperactivity										
r	0.202	0.339	0.325	0.071	0.345	0.298	0.224	—		
P value	.12	.008	.01	.59	.006	.02	.08	—		
Inattention										
r	0.176	0.186	0.006	0.280	0.161	–0.079	0.500	0.377	—	
P value	.18	.15	.97	.03	.21	.54	<.001	.003	—	
Emotional dysregulation										
r	–0.03	0.147	–0.215	0.294	0.047	–0.047	0.395	–0.061	0.308	—
P value	.82	.26	.10	.02	.72	.72	.002	.64	.02	—

[a]PTV: processing time variability.

[b]PT: processing time.

[c]Not applicable.

XSL•FO
RenderX

Discussion

Principal Findings

In this study, we applied our virtual office room scenario with its VEST to a sample of medicated participants with ADHD, unmedicated participants with ADHD, and unmedicated participants with MDD. Using a multimodal approach, we explored the paradigm's potential to detect differences in various outcomes across the 3 groups, extending previous research on VR-based assessment methods [20,24]. While we expected more movement in the ADHD groups than in the MDD group, only partial evidence supported this hypothesis. Unmedicated individuals with ADHD showed a distraction-induced increase in head rotation, whereas those with MDD showed a decrease in rotation during distraction. A trend-level effect also indicated more head movement during distraction in both ADHD groups compared to the MDD group. Additionally, regarding our research questions, we found an increase in PTV in unmedicated individuals with ADHD and in those with MDD, but not in medicated individuals with ADHD. Furthermore, we found an increase in PTV during distraction in medicated individuals with ADHD. At least one of the ADHD groups also reported higher task-related levels of inattention, hyperactivity, and emotional dysregulation than the MDD group. No differences between groups were found in eye movement, brain activity in the dlPFC as measured via fNIRS, and movement of the arms and torso. These findings indicate differences between the investigated groups in some of the parameters measured, and distinct responses to distractors and time-on-task, which will be examined in more detail in the following discussion.

Investigating task performance, we found a significant increase in PTV over time in participants with unmedicated ADHD and participants with unmedicated MDD, but not in medicated participants with ADHD. This finding indicates a time-on-task decline of sustained attentional performance or vigilance in unmedicated ADHD and MDD. This result aligns partially with previous research using VR-based CPTs in children, reporting lower reaction time variability in medicated compared to unmedicated individuals with ADHD [25-27]. Moreover, a meta-analysis by Kofler et al [51] found significant effects of stimulant treatment on reaction time variability, whereas no consistent reaction time variability differences were found between adults with ADHD and other clinical populations, including MDD. Our results indicate that medication helps maintain attentional focus over time in ADHD and that unmedicated individuals with ADHD or MDD show a comparable decline in attentional performance.

Additionally, medicated participants with ADHD showed a significantly higher PTV in DP than NDP. Although the effect was observed only in the medicated ADHD group, this finding suggests that the task is sensitive to distractor-specific influences on performance variability. The absence of the effect in the unmedicated ADHD group remains unclear. One possible explanation concerns the measure of variability; while the coefficient of variation is a common metric for reaction time variability, some argue that controlling for mean reaction time may be inappropriate, particularly in ADHD populations [52]. In this study, while not significantly different, unmedicated individuals with ADHD showed descriptively higher mean processing time than the medicated ADHD and the MDD group. Additionally, it is unclear to what extent processing time for emails is comparable to reaction times measured in tasks such as CPTs, which may also suggest the need for an alternative measure of variability.

Of note, no group differences emerged in mean processing time and error rates. This means that we were unable to use the VEST to map the comparatively slower information processing of individuals with MDD that has been observed in other neuropsychological tasks [53]. The error rate, which can be considered a measure of impulsivity, also showed no group differences. This may be due to the task itself, as the stimulus remains visible until a response is made rather than disappearing after a brief interval. It is also possible that, alongside the well-established impulsivity in ADHD, individuals with MDD exhibit impulsivity problems that reduce group differences, which has previously been shown [54]. However, it should be noted that the previous study likewise found no differences in error rates between ADHD and HC [24].

In addition, we observed a longer processing time in DP compared to NDP across groups. While this finding suggests that the experimental induction of distractibility was generally effective in observing reduced attentional performance across all groups, the distractors did not appear to elicit symptom-specific patterns of distractibility that would differentiate between diagnostic groups or between medicated and unmedicated participants.

We hypothesized that both ADHD groups would exhibit increased movement compared to the MDD group. This hypothesis is based on prior research that suggested that the overall activity level of individuals with MDD is decreased compared to those with ADHD. In our task, we measured specific body movements while sitting during task performance, which participants with ADHD and MDD did not differ in. Hence, differences might be found in overall motoric activity but not necessarily in physical activity during cognitive tasks. Nonetheless, the hypothesis was partially supported by a trend-level increase of movement during DP in both ADHD groups, but not in the MDD group. Similarly, we previously found a more pronounced increase in head movements from NDP to DP in ADHD compared to HC [24]. This pattern may reflect core symptomatology of ADHD, particularly hyperactivity and distractibility. If replicated, this differential movement response to environmental distractors may serve as a clinically relevant marker to distinguish between ADHD and MDD populations. In line with this, unmedicated participants with ADHD also displayed a significantly greater head rotation in DP than NDP. This finding further supports the potential use of actigraphy-derived motion parameters as objective, ecologically valid markers for differential diagnosis. By following up the phase and group interaction, we also found a higher head rotation in NDP compared to DP in participants with MDD. While seemingly counterintuitive at first, this finding might be due to the paradoxical effects of distractors. Prior research has presented evidence that distractors can sometimes facilitate performance [55,56], depending on the

XSL•FO
RenderX

task characteristics and properties of the distractors, and the specific population explored, including MDD and ADHD [57,58]. Given the unclear directionality and inconsistency with baseline expectations, further research is needed to clarify the mechanisms behind this effect.

The increase in movement of the head, arms, and torso across groups is in line with our previously reported increase in head movement over time in the same paradigm in HC and individuals with ADHD [24]. Increased body movement during sustained attention tasks may reflect a compensatory mechanism to maintain attention [59,60] and can be viewed as an adaptation to the demands of prolonged sustained attention, such as decreasing arousal or accumulating fatigue.

As pertains eye tracking, we found a reduced focus on the task in phases with distractions compared to phases without distractions, but no group differences. Regarding ADHD medication effects, this null finding aligns with our previous study [20], which also revealed no significant differences in eye movement patterns between these 2 subgroups. We also did not find any differences between ADHD and MDD, indicating that all groups spent a similar amount of time looking at the task. Nonetheless, it remains possible that subtle effects exist but were not detected due to limited statistical power resulting from relatively small sample sizes.

In the fNIRS analysis, we did not observe any group differences in the DP-NDP contrast in the dlPFC. This may partly be due to the loss of data caused by technical issues, reducing the analyzable sample on both channels of interest to 29 and 23, respectively. In contrast to traditional inhibitory control tasks, where reduced activation in the dlPFC is consistently reported in individuals with ADHD compared to HC [36], we found no difference between medicated and unmedicated individuals with ADHD. It is also possible that the type of inhibitory control required in classical tasks is different from inhibiting distractors, or alternatively, that the distractors used in our study were not strong enough to elicit such effects [24].

Pertaining to self-reports of task-related symptom experience, both ADHD groups reported higher levels of task-related inattention and impulsivity compared to the MDD group. Assuming that our PTV measure partly captures inattention, this contrasts with our finding that unmedicated individuals with ADHD or MDD showed greater PTV increases than medicated individuals with ADHD. Hence, there appears to be a partial discrepancy between behavioral and self-reported measures, which could mean that they capture different aspects of attentional dysfunction or be attributed to a greater sensitivity of PTV as an index of inattention. For impulsivity, although elevated levels have also been reported in MDD [54], the higher scores observed in the unmedicated ADHD group and, at a trend level, in the medicated ADHD group are consistent with the classification of impulsivity as a core symptom of ADHD. In terms of emotional dysregulation, only the medicated ADHD group reported higher scores than the MDD group, whereas both ADHD groups showed comparable levels to each other. Difficulties in emotion regulation are frequently reported in association with ADHD [61], and in our sample, medication use does not appear to have exerted a beneficial effect, with descriptively higher scores in the medicated group. The MDD group showed the lowest scores, which differed significantly from those of the medicated ADHD group. Because participants in our MDD group were affected by mild to moderate depressive symptoms that did not currently require medication, it is conceivable that they were able to use emotion-regulation strategies such as acceptance. Nevertheless, previous research has shown that individuals with MDD, even when remitted, still tend to suppress emotions, which could also manifest in lower scores [62]. Importantly, emotional dysregulation is a construct that includes various dimensions such as emotional reactivity and intense experience of emotion, but also decreased emotional awareness [63]. Therefore, it is possible that the item used to assess task-related emotional dysregulation might have only encapsulated one facet of emotional dysregulation. Also, it should be noted that, irrespective of absolute levels, all 3 groups reported values in the positive range of emotional dysregulation after the second block.

No group differences emerged in self-reported hyperactivity, which contrasts with the distractor-induced differences detected via head actigraphy, but aligns with the null results from actigraphy-derived torso and nondominant arm movements as discussed above.

Lastly, we conducted an exploratory correlational analysis of the task-related symptom intensity and the objective parameters we measured. Two key points emerged: first, self-reported, task-related intensity of inattention, hyperactivity, impulsivity, and emotional dysregulation were each correlated with different objective measures. For instance, inattention and emotional dysregulation were associated with higher levels of off-task gaze; impulsivity was associated with a higher error rate; and hyperactivity was associated with increased head movement, higher error rate, and longer processing time. Although exploratory, this pattern might suggest that a multimodal assessment combining task performance, eye movements, and body movements (particularly head movement) may capture distinct symptom dimensions of ADHD. Second, off-task gaze and head movement and rotation showed a correlation with mean PTV, and, in the case of head rotation, with mean processing time. While these findings present a first glance into the potential relationships of different behavioral modalities during the VEST, the causal relationship of these parameters needs further investigation.

Limitations and Future Directions

While the VEST was specifically designed to increase ecological validity compared to traditional neuropsychological tasks by applying a function-led design approach, it would be of interest to investigate the specific cognitive domains it draws upon [24]. Therefore, future studies should examine its associations with established (neuropsychological) measures, for example, attention and reading comprehension. Additionally, although the VEST is designed to simulate a realistic setting and task, a systematic investigation of its ecological validity, meaning the correspondence between behaviors exhibited in the virtual environment and in real-world settings [64], is still needed. While it would be very challenging to assess eye movement behavior, brain activity, and performance measures during the

XSL•FO
RenderX

participants' work to evaluate the equivalency of behavior in the virtual office room to real-world behavior, several alternative approaches could be applied. For example, cognitive functioning at work could be assessed using questionnaires [65,66]. Moreover, office-based participants could be asked to wear a mobile eye tracker and an actigraph during work sessions, while ambient noise levels would be simultaneously measured to assess auditory distractions. This would allow investigation of the relationship between distractor-related shifts of attention and movements in the virtual office room and real life.

Further, although the total sample comprised 62 participants, the sizes of the diagnostic subgroups, particularly the MDD group, were limited. As indicated by a sensitivity analysis, our sample consisting of 63 participants provided 80% power to detect large main effects (η^2_p=0.14, as calculated via G*Power, Heinrich Heine University Düsseldorf [67], based on the harmonic mean group size). Consequently, small- to medium-sized differences between the investigated population may not have been detected. Future studies with larger and more balanced samples are necessary to replicate and extend these results. For example, to reliably detect medium main effects, at least 50 participants per group would be needed. Furthermore, while the MDD diagnosis of participants in the MDD group was confirmed using the structured clinical interview Mini-DIPS, some MDD participants only reported mild depressive symptoms in the PHQ-9 [43]. Therefore, it is unclear whether the results can be generalized to individuals with MDD with elevated symptom intensity.

While we provide an analysis across multiple levels of behaviors by group, examining the influence of time-on-task and distractors, and conduct an exploratory correlational analysis, more in-depth integrative approaches combining these measures could further enhance our understanding of behavior in ADHD. For example, a machine learning approach similar to our analysis in Wiebe et al [18] could integrate the various parameters to classify between medicated versus unmedicated individuals with ADHD, and individuals with ADHD versus HC if combined with the dataset reported in Pakos et al [24]. Additionally, a more thorough analysis of distractor-related movement and gaze behavior, and their interplay during task errors, could illuminate potential underlying causes of said errors.

Moreover, while assessing the VEST's discriminant validity in differentiating between ADHD and MDD represents an important step toward evaluating its potential use in a clinical context, future research might also explore the overlap and divergence of symptom profiles in individuals with comorbid ADHD and MDD. This could further inform the refinement and diagnostic use of VR-based assessments like the VEST. Consequently, although the actual clinical applicability of the VEST still requires further investigation, it is conceivable that it could be used, for example, as a complementary tool in differential diagnostics or for monitoring symptom severity or symptom changes associated with medication. Its administration is straightforward, and even if the space or financial resources for a fully equipped VR laboratory are not available, it would be possible to use the application on a smaller scale, for example, on a smartphone with an inexpensive VR headset alternative.

Conclusion

This study examined, for the first time, the potential of our recently developed multimodal VEST to distinguish individuals with ADHD and MDD, and to detect medication-related effects in ADHD. An increase in PTV over time was observed in unmedicated ADHD and MDD participants, but not in those with medicated ADHD, suggesting a stabilizing effect of medication on attentional capacities. Moreover, medicated individuals with ADHD showed an increase in head rotation during distractions, and, on a trend level, both medicated and unmedicated individuals with ADHD showed increased distractor-associated head movements, an effect not observed in the MDD group. This finding underscores the potential of actigraphy in combination with distracting events during task performance as a tool for differentiating ADHD from other clinical populations. Additionally, exploratory correlational analyses showed correlations between various subjective and objective, as well as psychophysiological markers. While we also observed higher scores in task-related inattention, impulsivity, and emotional dysregulation in at least 1 ADHD group compared to MDD, no differences were found in gaze behavior and brain activity measured by fNIRS. To solidify the results and be able to reliably detect potential small size differences between the groups, and verify the task's potential as an assessment tool with potential applications in research and a clinical context, higher-powered studies are needed.

Acknowledgments

We sincerely thank all participants for generously giving their time to this study. This work was funded by the iBehave Network and supported by the Open Access Publication Fund of the University of Bonn. OpenAI's GPT-4o was used solely for language editing (grammar, phrasing, clarity, and flow). It was not used for study design and conception, methodology, analysis, interpretation, or generation of scientific content. All scientific content and decisions were made exclusively by the authors.

Funding

This work was supported by the Open Access Publication Fund of the University of Bonn. LA, JP, and LJ received funding given to AP and NB by the iBehave Network, which is sponsored by the Ministry of Culture and Science of the State of North Rhine-Westphalia. The funder had no involvement in the study design, data collection, analysis, interpretation, or the writing of the manuscript.

XSL•FO
RenderX

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Study conception and design: LA, JP, AW, DS, AP, NB
Data collection: JS, HS, LA
Data analysis: LA, JP, AW, NB
Original paper draft: LA, JP, NB
Review and editing: AW, BS, LJ, AP, NB
Supervision: AP, NB
All authors approved the submitted version of this paper.

Conflicts of Interest

NB and AP received funding over the Nordrhein-Westfalen network iBehave, by which JP's, LA's and LJ's positions are funded. AP and NB received funding from MEDICE Arzneimittel Pütter GmbH & Co KG, by which AW's PhD position was funded. NB received funding from BONFOR, by which NB and BS were partly funded. LA's and JP's positions are partially funded by Neurocast for work on a mobile sensing study unrelated to this manuscript. JP received compensation from MEDICE Arzneimittel Pütter GmbH & Co KG for conducting a workshop. AP received funding from the German Federal Ministry of Education and Research, by which BS's PhD position was partly funded. AP has received further funding by the German Federal Ministry of Education and Research, Horizon2020, and DFG, North Rhine Westphalia, and Medice Arzneimittel Pütter GmbH & Co KG; she has been on advisory boards for Takeda, Medice, and Boehringer; and delivered lectures sponsored by Medice, Takeda. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Further specifications of hardware and software.
[DOCX File , 16 KB - humanfactors_v13i1e85351_app1.docx]

Multimedia Appendix 2
Missing data of each modality.
[DOCX File , 19 KB - humanfactors_v13i1e85351_app2.docx]

Multimedia Appendix 3
Tabular summaries of ANOVA key results.
[DOCX File , 16 KB - humanfactors_v13i1e85351_app3.docx]

Multimedia Appendix 4
Correlational analysis.
[DOCX File , 19 KB - humanfactors_v13i1e85351_app4.docx]

References

1. Sayal K, Prasad V, Daley D, Ford T, Coghill D. ADHD in children and young people: prevalence, care pathways, and service provision. Lancet Psychiatry 2018;5(2):175-186. [doi: 10.1016/S2215-0366(17)30167-0] [Medline: 29033005]
2. Song P, Zha M, Yang Q, Zhang Y, Li X, Rudan I. The prevalence of adult attention-deficit hyperactivity disorder: a global systematic review and meta-analysis. J Glob Health 2021;11:04009 [FREE Full text] [doi: 10.7189/jogh.11.04009] [Medline: 33692893]
3. Diagnostic and Statistical Manual of Mental Disorders. Washington, DC: American Psychiatric Association; 2013.
4. Cortese S, Bellgrove MA, Brikell I, Franke B, Goodman DW, Hartman CA, et al. Attention-deficit/hyperactivity disorder (ADHD) in adults: evidence base, uncertainties and controversies. World Psychiatry 2025;24(3):347-371 [FREE Full text] [doi: 10.1002/wps.21374] [Medline: 40948064]
5. Kooij JJS, Bijlenga D, Salerno L, Jaeschke R, Bitter I, Balázs J, et al. Updated European consensus statement on diagnosis and treatment of adult ADHD. Eur Psychiatry 2019;56(1):14-34 [FREE Full text] [doi: 10.1016/j.eurpsy.2018.11.001] [Medline: 30453134]
6. Montag C, Rumpf HJ. The potential of digital phenotyping and mobile sensing for psycho-diagnostics of internet use disorders. Curr Addict Rep 2021;8(3):422-430 [FREE Full text] [doi: 10.1007/s40429-021-00376-6] [Medline: 34258147]
7. van Bennekom MJ, de Koning PP, Denys D. Virtual reality objectifies the diagnosis of psychiatric disorders: a literature review. Front Psychiatry 2017;8:163 [FREE Full text] [doi: 10.3389/fpsyt.2017.00163] [Medline: 28928677]

XSL•FO
RenderX

8. Lange KW, Hauser J, Lange KM, Makulska-Gertruda E, Takano T, Takeuchi Y, et al. Utility of cognitive neuropsychological assessment in attention-deficit/hyperactivity disorder. Atten Defic Hyperact Disord 2014;6(4):241-248. [doi: 10.1007/s12402-014-0132-3] [Medline: 24639037]
9. Adamou M, Fullen T, Jones SL. EEG for diagnosis of adult ADHD: a systematic review with narrative analysis. Front Psychiatry 2020;11:871 [FREE Full text] [doi: 10.3389/fpsyt.2020.00871] [Medline: 33192633]
10. Kaiser A, Aggensteiner PM, Baumeister S, Holz NE, Banaschewski T, Brandeis D. Earlier versus later cognitive event-related potentials (ERPs) in attention-deficit/hyperactivity disorder (ADHD): a meta-analysis. Neurosci Biobehav Rev 2020;112:117-134 [FREE Full text] [doi: 10.1016/j.neubiorev.2020.01.019] [Medline: 31991190]
11. Parlatini V, Bellato A, Gabellone A, Margari L, Marzulli L, Matera E, et al. A state-of-the-art overview of candidate diagnostic biomarkers for attention-deficit/hyperactivity disorder (ADHD). Expert Rev Mol Diagn 2024;24(4):259-271. [doi: 10.1080/14737159.2024.2333277] [Medline: 38506617]
12. Chen H, Yang Y, Odisho D, Wu S, Yi C, Oliver BG. Can biomarkers be used to diagnose attention deficit hyperactivity disorder? Front Psychiatry 2023;14:1026616 [FREE Full text] [doi: 10.3389/fpsyt.2023.1026616] [Medline: 36970271]
13. Salazar de Pablo G, Iniesta R, Bellato A, Caye A, Dobrosavljevic M, Parlatini V, et al. Individualized prediction models in ADHD: a systematic review and meta-regression. Mol Psychiatry 2024;29(12):3865-3873. [doi: 10.1038/s41380-024-02606-5] [Medline: 38783054]
14. Wiebe A, Kannen K, Selaskowski B, Mehren A, Thöne AK, Pramme L, et al. Virtual reality in the diagnostic and therapy for mental disorders: a systematic review. Clin Psychol Rev 2022;98:102213 [FREE Full text] [doi: 10.1016/j.cpr.2022.102213] [Medline: 36356351]
15. Goharinejad S, Goharinejad S, Hajesmaeel-Gohari S, Bahaadinbeigy K. The usefulness of virtual, augmented, and mixed reality technologies in the diagnosis and treatment of attention deficit hyperactivity disorder in children: an overview of relevant studies. BMC Psychiatry 2022;22(1):4 [FREE Full text] [doi: 10.1186/s12888-021-03632-1] [Medline: 34983446]
16. Wiebe A, Kannen K, Li M, Aslan B, Anders D, Selaskowski B, et al. Multimodal virtual reality-based assessment of adult ADHD: a feasibility study in healthy subjects. Assessment 2023;30(5):1435-1453. [doi: 10.1177/10731911221089193]
17. Pieri L, Tosi G, Romano D. Virtual reality technology in neuropsychological testing: a systematic review. J Neuropsychol 2023;17(2):382-399. [doi: 10.1111/jnp.12304] [Medline: 36624041]
18. Wiebe A, Selaskowski B, Paskin M, Asché L, Pakos J, Aslan B, et al. Virtual reality-assisted prediction of adult ADHD based on eye tracking, EEG, actigraphy and behavioral indices: a machine learning analysis of independent training and test samples. Transl Psychiatry 2024;14(1):508 [FREE Full text] [doi: 10.1038/s41398-024-03217-y] [Medline: 39741130]
19. Zhang KF, Yeh SC, Hsiao-Kuang Wu E, Xu X, Tsai HJ, Chen CC. Fusion of multi-task neurophysiological data to enhance the detection of attention-deficit/hyperactivity disorder. IEEE J Transl Eng Health Med 2024;12:668-674. [doi: 10.1109/JTEHM.2024.3435553] [Medline: 39464626]
20. Wiebe A, Aslan B, Brockmann C, Lepartz A, Dudek D, Kannen K, et al. Multimodal assessment of adult attention-deficit hyperactivity disorder: a controlled virtual seminar room study. Clin Psychol Psychother 2023;30(5):1111-1129. [doi: 10.1002/cpp.2863] [Medline: 37209018]
21. Parsons TD, Duffield T, Asbee J. A comparison of virtual reality classroom continuous performance tests to traditional continuous performance tests in delineating ADHD: a meta-analysis. Neuropsychol Rev 2019;29(3):338-356. [doi: 10.1007/s11065-019-09407-6]
22. Rosvold HE, Mirsky AF, Sarason I, Bransome ED, Beck LH. A continuous performance test of brain damage. J Consult Psychol 1956;20(5):343-350. [doi: 10.1037/h0043220]
23. Selaskowski B, Asché LM, Wiebe A, Kannen K, Aslan B, Gerding TM, et al. Gaze-based attention refocusing training in virtual reality for adult attention-deficit/hyperactivity disorder. BMC Psychiatry 2023;23(1):74 [FREE Full text] [doi: 10.1186/s12888-023-04551-z] [Medline: 36703134]
24. Pakos J, Asché L, Popova A, Heine C, Sanchez D, Aslan B, et al. Evaluation of a virtual reality-based office room for a multimodal assessment of adult ADHD. BMC Digit Health 2025;3(1):71. [doi: 10.1186/s44247-025-00209-w]
25. Díaz-Orueta U, Garcia-López C, Crespo-Eguílaz N, Sánchez-Carpintero R, Climent G, Narbona J. AULA virtual reality test as an attention measure: convergent validity with Conners' Continuous Performance Test. Child Neuropsychol 2014;20(3):328-342. [doi: 10.1080/09297049.2013.792332] [Medline: 23638628]
26. Mühlberger A, Jekel K, Probst T, Schecklmann M, Conzelmann A, Andreatta M, et al. The influence of methylphenidate on hyperactivity and attention deficits in children with ADHD: a virtual classroom test. J Atten Disord 2020;24(2):277-289. [doi: 10.1177/1087054716647480] [Medline: 27178061]
27. Pollak Y, Shomaly HB, Weiss PL, Rizzo AA, Gross-Tsur V. Methylphenidate effect in children with ADHD can be measured by an ecologically valid continuous performance test embedded in virtual reality. CNS Spectr 2010;15(2):125-130. [doi: 10.1017/s109285290002736x] [Medline: 20414157]
28. Goodman D. Adult ADHD and comorbid depressive disorders: diagnostic challenges and treatment options. CNS Spectr 2009;14(6):5-7. [doi: 10.1017/s1092852900024810] [Medline: 19773716]
29. Potvin S, Charbonneau G, Juster RP, Purdon S, Tourjman SV. Self-evaluation and objective assessment of cognition in major depression and attention deficit disorder: implications for clinical practice. Compr Psychiatry 2016;70:53-64. [doi: 10.1016/j.comppsych.2016.06.004] [Medline: 27624423]

XSL•FO
RenderX

30. Paucke M, Stibbe T, Huang J, Strauss M. Differentiation of ADHD and depression based on cognitive performance. J Atten Disord 2021;25(7):920-932. [doi: 10.1177/1087054719865780] [Medline: 31409195]
31. van Hal R, Geurts D, van Eijndhoven P, Kist J, Collard RM, Tendolkar I, et al. A transdiagnostic view on MDD and ADHD: shared cognitive characteristics? J Psychiatr Res 2023;165:315-324. [doi: 10.1016/j.jpsychires.2023.07.028] [Medline: 37556964]
32. Wüthrich F, Nabb CB, Mittal VA, Shankman SA, Walther S. Actigraphically measured psychomotor slowing in depression: systematic review and meta-analysis. Psychol Med 2022;52(7):1208-1221. [doi: 10.1017/s0033291722000903]
33. Lohr JB, May T, Caligiuri MP. Quantitative assessment of motor abnormalities in untreated patients with major depressive disorder. J Affect Disord 2013;146(1):84-90 [FREE Full text] [doi: 10.1016/j.jad.2012.08.043] [Medline: 22985485]
34. Suslow T, Hußlack A, Kersting A, Bodenschatz CM. Attentional biases to emotional information in clinical depression: a systematic and meta-analytic review of eye tracking findings. J Affect Disord 2020;274:632-642. [doi: 10.1016/j.jad.2020.05.140] [Medline: 32663997]
35. Carvalho N, Laurent E, Noiret N, Chopard G, Haffen E, Bennabi D, et al. Eye movement in unipolar and bipolar depression: a systematic review of the literature. Front Psychol 2015;6:1809 [FREE Full text] [doi: 10.3389/fpsyg.2015.01809] [Medline: 26696915]
36. Hou L, Yang J, Xu L, Peng J, Joyce Law CY, Chen T. Activation of brain regions associated with working memory and inhibitory control in patients with attention-deficit/hyperactivity disorder in functional near-infrared spectroscopy: a systematic review. Curr Med Imaging 2023;19(8):865-873 [FREE Full text] [doi: 10.2174/1573405618666220822101019] [Medline: 35996256]
37. Yeung MK, Lin J. Probing depression, schizophrenia, and other psychiatric disorders using fNIRS and the verbal fluency test: a systematic review and meta-analysis. J Psychiatr Res 2021;140:416-435. [doi: 10.1016/j.jpsychires.2021.06.015] [Medline: 34146793]
38. Supekar K, Iyer T, Menon V. The influence of sex and age on prevalence rates of comorbid conditions in autism. Autism Res 2017;10(5):778-789. [doi: 10.1002/aur.1741] [Medline: 28188687]
39. RED connect [Internet]. RED Medical Systems GmbH. 2022. URL: https://www.redmedical.de/red-connect-videosprechstunde/ [accessed 2022-02-04]
40. Margraf J, Cwik JC, Pflug V, Schneider S. Strukturierte klinische interviews zur Erfassung psychischer störungen über die lebensspanne: Gütekriterien und weiterentwicklungen der DIPS-Verfahren. Z Für Klin Psychol Psychother 2017;46(3):176-186. [doi: 10.1026/1616-3443/a000430]
41. Retz W, Retz-Junginger P, Römer K, Rösler M. Standardisierte skalen zur strukturierten diagnostik der ADHS im erwachsenenalter. Fortschr Neurol Psychiatr 2013;81(7):381-389. [doi: 10.1055/s-0033-1335740] [Medline: 23856943]
42. SoSci Survey [Internet]. Leiner DJ. 2022. URL: https://www.soscisurvey.de [accessed 2022-02-04]
43. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x]
44. Rösler M, Retz W, Retz-Junginger P, Thome J, Supprian T, Nissen T, et al. Instrumente zur diagnostik der aufmerksamkeitsdefizit-/hyperaktivitätsstörung (ADHS) im erwachsenenalter: selbstbeurteilungsskala (ADHS-SB) und diagnosecheckliste (ADHS-DC). Nervenarzt 2004;75(9):888-895. [doi: 10.1007/s00115-003-1622-2] [Medline: 15378249]
45. Kim HK, Park J, Choi Y, Choe M. Virtual Reality Sickness Questionnaire (VRSQ): motion sickness measurement index in a virtual reality environment. Appl Ergon 2018;69:66-73. [doi: 10.1016/j.apergo.2017.12.016] [Medline: 29477332]
46. Kothe C, Shirazi SY, Stenner T, Medine D, Boulay C, Grivich MI, et al. The lab streaming layer for synchronized multimodal recording. Imaging Neurosci 2025;3:IMAG.a.136 [FREE Full text] [doi: 10.1101/2024.02.13.580071] [Medline: 38405712]
47. Huppert TJ, Diamond SG, Franceschini MA, Boas DA. HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. Appl Opt 2009;48(10):D280-D298 [FREE Full text] [doi: 10.1364/ao.48.00d280] [Medline: 19340120]
48. Montero Hernandez S, Pollonini L. NIRSplot: a tool for quality assessment of fNIRS scans. 2020 Presented at: Biophotonics Congress: Biomedical Optics 2020 (Translational, Microscopy, OCT, OTS, BRAIN); 2020 April 20-23; Washington, DC. [doi: 10.1364/BRAIN.2020.BM2C.5]
49. Heymans MW, Twisk JW. Handling missing data in clinical research. J Clin Epidemiol 2022;151:185-188 [FREE Full text] [doi: 10.1016/j.jclinepi.2022.08.016] [Medline: 36150546]
50. Cohen J. Statistical Power Analysis for the Behavioral Sciences [Internet]. Oxfordshire, UK: Routledge; 2025.
51. Kofler MJ, Rapport MD, Sarver DE, Raiker JS, Orban SA, Friedman LM, et al. Reaction time variability in ADHD: a meta-analytic review of 319 studies. Clin Psychol Rev 2013;33(6):795-811. [doi: 10.1016/j.cpr.2013.06.001] [Medline: 23872284]
52. Klein C, Wendling K, Huettner P, Ruder H, Peper M. Intra-subject variability in attention-deficit hyperactivity disorder. Biol Psychiatry 2006;60(10):1088-1097. [doi: 10.1016/j.biopsych.2006.04.003] [Medline: 16806097]
53. Nuño L, Gómez-Benito J, Carmona VR, Pino O. A systematic review of executive function and information processing speed in major depression disorder. Brain Sci 2021;11(2):147. [doi: 10.3390/brainsci11020147] [Medline: 33499360]
54. Fields SA, Schueler J, Arthur KM, Harris B. The role of impulsivity in major depression: a systematic review. Curr Behav Neurosci Rep 2021;8(2):38-50. [doi: 10.1007/s40473-021-00231-y]

XSL•FO
RenderX

55. Hemond C, Brown RM, Robertson EM. A distraction can impair or enhance motor performance. J Neurosci 2010;30(2):650-654. [doi: 10.1523/jneurosci.4592-09.2010]
56. Sanders GS, Baron RS. The motivating effects of distraction on task performance. J Pers Soc Psychol 1975;32(6):956-963. [doi: 10.1037//0022-3514.32.6.956]
57. van Mourik R, Oosterlaan J, Heslenfeld DJ, Konig CE, Sergeant JA. When distraction is not distracting: a behavioral and ERP study on distraction in ADHD. Clin Neurophysiol 2007;118(8):1855-1865. [doi: 10.1016/j.clinph.2007.05.007] [Medline: 17576093]
58. Schnabel R, Kydd R. Neuropsychological assessment of distractibility in mild traumatic brain injury and depression. Clin Neuropsychol 2012;26(5):769-789. [doi: 10.1080/13854046.2012.693541] [Medline: 22694212]
59. Farley J, Risko EF, Kingstone A. Everyday attention and lecture retention: the effects of time, fidgeting, and mind wandering. Front Psychol 2013;4:619. [doi: 10.3389/fpsyg.2013.00619] [Medline: 24065933]
60. Sarver DE, Rapport MD, Kofler MJ, Raiker JS, Friedman LM. Hyperactivity in attention-deficit/hyperactivity disorder (ADHD): impairing deficit or compensatory behavior? J Abnorm Child Psychol 2015;43(7):1219-1232. [doi: 10.1007/s10802-015-0011-1] [Medline: 25863472]
61. Soler-Gutiérrez AM, Pérez-González JC, Mayas J. Evidence of emotion dysregulation as a core symptom of adult ADHD: a systematic review. PLoS One 2023;18(1):e0280131 [FREE Full text] [doi: 10.1371/journal.pone.0280131] [Medline: 36608036]
62. Visted E, Vøllestad J, Nielsen MB, Schanche E. Emotion regulation in current and remitted depression: a systematic review and meta-analysis. Front Psychol 2018;9:756 [FREE Full text] [doi: 10.3389/fpsyg.2018.00756] [Medline: 29867700]
63. D'Agostino A, Covanti S, Rossi Monti M, Starcevic V. Reconsidering emotion dysregulation. Psychiatr Q 2017;88(4):807-825. [doi: 10.1007/s11126-017-9499-6] [Medline: 28194549]
64. Kothgassner OD, Felnhofer A. Does virtual reality help to cut the Gordian knot between ecological validity and experimental control? Ann Int Commun Assoc 2020;44(3):210-218. [doi: 10.1080/23808985.2020.1792790]
65. Domensino AF, Aarts E, Visser-Meily JMA, Spikman JM, van Heugten C. Development and content validity of the Cognition in Daily Life Scale (CDL). Neuropsychol Rehabil 2025;35(2):382-407 [FREE Full text] [doi: 10.1080/09602011.2024.2343149] [Medline: 38656293]
66. Heikkinen AL, Hänninen T, Kuikka P, Akila R, Savolainen A, Valtonen T, et al. The Cognitive Function at Work Questionnaire (CFWQ): a new scale for measuring cognitive complaints in occupational population. Appl Neuropsychol Adult 2023;30(6):649-660. [doi: 10.1080/23279095.2021.1970553] [Medline: 34482772]
67. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007;39(2):175-191. [doi: 10.3758/bf03193146]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ADHS-SB: ADHD Self-Rating Scale
CPT: continuous performance task
dlPFC: dorsolateral prefrontal cortex
DP: distractor phase
DSM-5: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
fNIRS: functional near-infrared spectroscopy
HC: healthy controls
HMD: head-mounted display
ICD-10: International Statistical Classification of Diseases, 10th Revision
MDD: major depressive disorder
MDiff: mean difference
Mini-DIPS: Diagnostic Brief Interview for Mental Disorders
NDP: nondistractor phase
Oxy-Hb: oxyhemoglobin
PHQ-9: Patient Health Questionnaire-9
PTV: processing time variability
VEST: virtual email sorting task
VR: virtual reality
VRSQ: Virtual Reality Sickness Questionnaire
VSR: virtual seminar room

XSL•FO
RenderX

Edited by A Kushniruk; submitted 06.Oct.2025; peer-reviewed by O Ademola-Aliu, JL Sun, D Lei; comments to author 27.Oct.2025; revised version received 17.Dec.2025; accepted 19.Jan.2026; published 02.Mar.2026.

Please cite as:
Asché L, Pakos J, Schrage H, Schuder J, Jung L, Sanchez D, Selaskowski B, Wiebe A, Philipsen A, Braun N
Multimodal Virtual Reality Assessment of Medication Effects in Attention-Deficit/Hyperactivity Disorder and Its Distinction From Depression: Cross-Sectional Study
JMIR Hum Factors 2026;13:e85351
URL: https://humanfactors.jmir.org/2026/1/e85351
doi:10.2196/85351
PMID:

©Laura Asché, Julian Pakos, Hannah Schrage, Johanna Schuder, Luisa Jung, Dario Sanchez, Benjamin Selaskowski, Annika Wiebe, Alexandra Philipsen, Niclas Braun. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 02.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Development and Validation of the Large Language Model Privacy Concern Assessment Questionnaire: Psychometric Evaluation Study on a Brief Instrument for Assessing User Privacy Concerns With Large Language Model Applications

Leming Zhou, PhD; Hongtao Wang, MS; Yi Xu, MS; I Made Agus Setiawan, PhD; Haomin Hu, PhD; Andi Saptono, PhD; Yong K Choi, PhD, MPH; Bambang Parmanto, PhD

Department of Health Information Management, School of Health and Rehabilitation Sciences, University of Pittsburgh, 3396 Fifth Avenue, Pittsburgh, PA, United States

Corresponding Author:
Leming Zhou, PhD
Department of Health Information Management, School of Health and Rehabilitation Sciences, University of Pittsburgh, 3396 Fifth Avenue, Pittsburgh, PA, United States

Abstract

Background: Privacy concerns represent a primary barrier to the adoption of large language model (LLM) applications, particularly in sensitive domains such as health care. Although a growing body of survey-based research has examined user privacy concerns with LLMs, no standardized instrument exists for measuring this construct.

Objective: This study aimed to develop and validate a concise, standardized questionnaire for assessing user privacy concerns with LLM-based systems, suitable for use by researchers and practitioners such as AI application developers and health care professionals.

Methods: A literature review was conducted to identify privacy-related domains relevant to LLM use. A draft questionnaire was developed and subjected to content validity review by 8 experts. The finalized instrument was administered via Qualtrics to adults aged 18 years or older who were aware of LLM-based applications, recruited through the University of Pittsburgh, LinkedIn, and the Pitt+Me research participant registry between April 2025 and July 2026. Internal consistency reliability and factor structure were assessed, with factor structure examined in a split-sample design, in which exploratory factor analysis (EFA) was conducted on a random one-third of the sample and confirmatory factor analysis (CFA) on the holdout two-thirds.

Results: All questionnaire items achieved acceptable content validity (item-level content validity index >0.80; average scale-level content validity index=0.964). A total of 603 responses were collected. After preprocessing, the final sample comprised 361 participants. Most participants were younger than 35 years of age (198/361, 54.8%), female (230/361, 63.7%), and White (285/361, 78.9%); 88.1% (318/361) held at least a bachelor's degree. The 6-item questionnaire showed high internal consistency (Cronbach α=0.90; McDonald ω=0.90). EFA (n=121) supported a single dominant factor with all standardized loadings ≥0.70. CFA (n=240) showed excellent fit (comparative fit index=0.996; Tucker-Lewis index=0.993; root mean square error of approximation=0.046; standardized root mean square residual=0.057). A 2-factor model separating the transparency- and control-related items fit significantly better than the 1-factor model ($\Delta\chi^2_1$=8.9; P=.003), but the 2 factors were correlated at r=0.83, indicating closely related facets. Because the Fornell-Larcker criterion was met for the transparency and control subscale but not for data use and risk salience, the 6-item total score remains the recommended primary metric.

Conclusions: The LLM privacy concern assessment is a concise, psychometrically sound 6-item instrument that measures privacy concern with LLM-based systems via 2 correlated subdimensions (transparency and control; data use and risk salience) summarized by a single total score. It enables standardized measurement across studies and practical privacy assessment in real-world LLM applications.

(JMIR Hum Factors 2026;13:e102910) doi:10.2196/102910

KEYWORDS

privacy; questionnaire; psychometric analysis; large language models; scale validation

XSL•FO
RenderX

Introduction

Background

Large language models (LLMs) have seen rapid and widespread adoption across domains ranging from general-purpose information retrieval to specialized health care consultations. As users increasingly engage with these systems for sensitive tasks, including health queries, financial decisions, and personal problem-solving, privacy concerns have emerged as a primary barrier to adoption and sustained use [1-3]. These concerns are well-founded because user-entered data often include highly sensitive information about health, finances, and personal relationships that could be used in the next round of LLM training. Unlike traditional regulated health platforms, most general-purpose LLMs currently operate outside established comprehensive data protection frameworks, leaving users with limited assurances regarding how their information is handled [4].

Information privacy concern has been investigated since the 1990s through a succession of instruments, each linked to the development of information technology and responding to the data practices of its era [5]. The Concern for Information Privacy (CFIP) scale of Smith et al [6] framed privacy concern as apprehension about organizational data handling, with 4 domains addressing collection, unauthorized secondary use, improper access, and errors. The Internet Users' Information Privacy Concerns (IUIPC) scale of Malhotra et al [7] recasts concern as an appraisal of the fairness of a data exchange, drawing on social contract theory and measures collection, control, and awareness, while the Mobile Users' Information Privacy Concerns (MUIPC) scale of Xu et al [8] draws on communication privacy management theory and measures perceived surveillance, perceived intrusion, and secondary use. Later instruments pursued either integration, nesting several first-order domains under a higher-order general factor, as in Hong and Thong's Internet Privacy Concerns (IPC) scale [9], or reducing the construct to a small number of platform-specific domains, as in the scale of Sheehan and Hoy [10] for online consumers and in the scale of Wang et al [11] for short-form video platforms. Across this lineage, a core recurs, namely, collection, awareness, and use or reuse of data, while the remaining domains and the dimensional structure itself shift with the referent, from organizational record-keeping to the commercial web and then to mobile and social media platforms [6-11], and each shift left the preceding generation of items a poorer fit than it had been. That fragility is not only conceptual but linguistic. Item-interpretation studies show that widely used privacy statements often measure more than 1 construct and are read differently by different populations [12,13], and re-examinations on independent data have repeatedly led to respecification of published factor structures and revealed validity and reliability shortfalls [14,15]. This lineage demonstrates that privacy concern instruments are context-bound. Each of these instruments was created because a shift in technology, referent, theory, or population made prior items ill-fitting. For any new instrument, the dimensional structure, item wording, and cross-context validity must be established rather than assumed.

LLM-based applications meet that same test on more than 1 dimension, which is the basis for developing a new instrument rather than adopting an existing one. First, the referent itself differs, because prior scales address an organization, website, or platform, whereas many users often interact with an LLM as a conversational partner and disclose information in natural language without a clear idea about how their inputs are stored and reused [16]. Second, the data practices also differ. User inputs may be retained for model training, and LLMs can memorize and later reproduce fragments of their training data [17] or infer unstated personal attributes such as employment, location, income, and sex from ordinary text, which is an inferential privacy risk with no close analogue in form-based web interaction [18]. The LLM provider ecosystem is also opaque and multiparty, so retention, sharing, and reuse of data are largely unobservable to the user. Finally, because item meaning does not transfer cleanly across referents [12,13], even a shared construct such as data "collection" is unlikely to be measurement-invariant when the collecting entity is a generative AI model rather than a website.

A first wave of survey studies has begun to measure users' privacy concerns with LLMs and AI chatbots, several of them in health contexts, but each has operationalized privacy through a distinct theoretical lens and for a distinct purpose. Choi et al [1] assessed disclosure willingness and its antecedents across 3 countries using items derived from the antecedents, privacy concerns, and outcomes (APCO) macromodel [5]; Liu et al [2] used privacy concern as a predictor within a behavioral model grounded in uncertainty reduction theory [19]; Liu et al [3] measured privacy awareness and expectation gaps using information flows informed by contextual integrity [20]; and Tran et al [4] used factorial vignettes to assess the contextual conditions under which data sharing is perceived as appropriate. Although each of these instruments provides theoretically rich insights within its specific framework, none was designed to yield a generalizable, stand-alone privacy concern score that could be compared across studies or applied by practitioners to a new LLM system.

What is missing is a validated, purpose-built measure of user privacy concern toward LLMs that is not tied to any single behavioral theory or downstream outcome, and that a researcher or LLM application developer can apply directly to assess privacy concern toward a specific system. This gap motivates the development of a concise, referent-specific instrument that preserves the field's replicated conceptual core, reanchors it to the LLM context, and extends it to concerns unique to conversational AI.

Objectives

The goal of this study is to develop and validate a concise, standardized instrument for assessing user privacy concerns specific to LLM-based systems. The instrument is intended to serve both researchers who need a reusable scale enabling cross-study comparison and practitioners who wish to conduct privacy assessments of specific LLM applications. Development followed a multistage process including item creation informed by the literature, expert content validation, and psychometric evaluation in a general adult sample.

XSL•FO
RenderX

Methods

Survey Development

Two authors in the study team (LZ and HW) conducted a literature review to identify privacy domains relevant to LLM use, including technology vendor trust, informed consent, data handling practices, user control, and transparency. Based on this review, a draft questionnaire was developed. A literature search was performed on Google Scholar, PubMed, and Scopus using keywords ["privacy concern" AND (questionnaire OR instrument OR scale) AND (internet OR web OR e-commerce OR "mobile app" OR "short form video" OR "Artificial Intelligence" OR AI)] without date restrictions. LZ and HW worked together and selected articles on questionnaire development and the ones that have been widely adopted in other studies. Some articles were identified from the references of selected articles. We ignored articles that simply applied these questionnaires to measure privacy concern or presented a language translation of them. The domains and questionnaire items from the 6 selected articles [6-11] were extracted and compared against the uniqueness of privacy concerns with LLMs (eg, the opacity of model vendors, new ways of providing personal data, and new ways of reusing user data).

Candidate domains were identified from the comparative synthesis of the 6 foundational instruments and the LLM privacy literature, following recommended practice of grounding a construct's content in theory and prior measurement before item generation [21]. We retained the domains with the strongest cross-instrument replication (data use, awareness, and control) since they were still highly relevant to privacy concerns with LLM applications, and added domains warranted by the distinctive data practices of LLMs (transparency for trust, risk salience, and onward third-party sharing), plus a global-concern domain to support an overall index.

Theoretically, these candidate domains are consistent with the APCO framing of privacy concern as a central antecedent of privacy outcomes [5], with social contract accounts of fairness and control in data exchange [7,22], and with communication privacy management theory's emphasis on control over disclosure boundaries [8]. All of these are strained in the conversational, inference-capable LLM setting. Grounding the domains in this measurement lineage rather than in any single behavioral theory is deliberate because recent LLM questionnaires built on APCO, uncertainty reduction theory, and contextual integrity each measured a different facet of concern and produced noncomparable scores [1-4], so this new questionnaire specifies its content at the construct level to remain usable across those frameworks.

The initial working draft included more than 1 candidate item for 7 domains; through internal discussion, and before the independent content-validity review, the pool was reduced to 1 item per domain. Wording followed the lessons of the interpretation-and-validity literature [12,13]: items are short and adopt a single consistent perspective, that is, the person's own concern or valuation, rather than mixing concern-perception with expectation phrasing [9]. Each item names a referent ("LLMs" or "LLM-based applications") so that responses attach to the intended entity rather than to websites or platforms generally, directly addressing the referent-mismatch problem the validity studies document. Response formats use ordinal Likert-type intensity anchors appropriate to each stem (eg, degree of concern, degree of importance, and frequency).

The instrument is intended as a brief privacy-concern index, scored as a single composite and composed of 1 item per narrowly defined domain, rather than a multi-item-per-factor multidimensional battery. This is a deliberate design choice, not a claim that privacy concern is psychologically unidimensional. Where a construct or domain is concrete and singular, single-item measures can perform comparably to multi-item scales and are defensible on both predictive validity and practical grounds [23-25]. A one-item-per-domain design maximizes content breadth across the privacy space while minimizing respondent burden and the wording-induced interpretation problems that multi-item scales compound [12,13]. The trade-offs, no within-domain internal-consistency estimate and greater sensitivity to item wording, are mitigated by the validation plan below and by the deliberate, theory-anchored specification of each domain. We acknowledge these limitations in the Discussion section, and we regard the instrument as preliminary. To avoid a terminological ambiguity, we note that the measurement model estimated in this study is reflective: each item is treated as an effect indicator of a common latent privacy-concern factor, which justifies the use of factor-analytic and internal-consistency evidence [26].

Responses were measured on a 5-point Likert scale with anchors tailored to each item's stem: importance items were anchored 1=extremely important to 5=not at all important; concern items were anchored 1=extremely concerned to 5=not at all concerned; and the frequency item was anchored 1=always to 5=never. The use of heterogeneous anchors (importance, concern, and frequency) was intentional to match the conceptual content of each item; the subsequent factor analysis tested whether items cohere as a single scale despite the differing anchors. All 3 anchor sets run in the same direction, from the strongest to the weakest degree of concern, so lower raw scores indicate stronger privacy concern, and item scores should be reverse-coded before a composite is formed if higher scores are to denote greater concern. The composite is a unit-weighted sum or mean of the 6 raw item responses, which assumes that the 5 ordinal positions are metrically equivalent across the 3 anchor types. The present data cannot verify that assumption. The items are known only to share a common 5-category format and common direction. We retained unit weighting rather than standardizing items to *z* scores for 2 reasons: unit-weighted composites are generally as predictively useful as differentially weighted ones and more stable across samples [27]; *z* standardization would make the score dependent on the composition of the sample in which it is computed, which is unsuitable for the benchmarking use for which the instrument is intended. Two questions are therefore left open: whether the 3 anchor types function equivalently across respondents, and whether unit-weighted and standardized scoring lead to the same conclusions. Both are discussed in the Limitations and Future Work section.

Validation followed established scale-development guidance [21,28]. Content validity of the questionnaire was evaluated by

XSL•FO
RenderX

a panel of 8 independent subject-matter experts. To preserve the independence of the appraisal, the panel comprised only experts who were not involved in item construction. They were recruited via research team members' recommendations. Their backgrounds were computer science, information science, and user experience research, and all had extensive experience in questionnaire design and evaluation. Each expert independently rated every item on 2 dimensions, relevance to the target construct (user privacy concern with LLM applications) and clarity of wording, using the 4-point ordinal scale recommended by Lynn [29]: 1=not relevant or clear, 2=somewhat, 3=quite, and 4=highly relevant or clear. Ratings of 3 or 4 were recoded as endorsement and ratings of 1 or 2 as nonendorsement [29,30]. Items for which the majority of reviewers assigned a relevance rating of 1 or 2 were removed. Items receiving low clarity ratings were retained but revised through team consensus during a structured review meeting. For each item, the item-level content validity index (I-CVI) was computed as the proportion of experts who endorsed the item. Because the I-CVI=A/N does not adjust for chance agreement, a modified kappa (κ*) was also calculated by correcting each I-CVI for the probability of chance agreement, $P_C=[N!/(A!(N-A)!)](0.5^N)$, where N is the number of experts, and A is the number endorsing the item; $\kappa^*=(\text{I-CVI}-P_C)/(1-P_C)$ was interpreted as excellent (>0.74), good (0.60-0.74), or fair (0.40-0.59) [31]. At the scale level, both the averaging method (scale-level content validity index [S-CVI]/Ave and the mean of the item I-CVIs) and the universal-agreement (UA) method (S-CVI/UA, the proportion of items endorsed by all experts) were reported [30]. An I-CVI of at least 0.78 for panels of 6 or more experts and an S-CVI/Ave of at least 0.90 were adopted a priori as evidence of satisfactory content validity [29-31]. Structural validity was assessed with exploratory factor analysis (EFA) followed by confirmatory factor analysis (CFA), evaluated against conventional fit criteria [32]. Reliability was reported as internal consistency and composite reliability where applicable. Further details are provided in the Statistical Analysis section.

The panel's ratings are one component of the content validity evidence rather than the whole of it. The theoretical grounding of the candidate domains and the mapping of each item to its domain and to the prior constructs from which that domain derives are reported in Table 1. Because the panel comprised 8 reviewers recruited on the recommendation of research team members rather than by independent nomination, its endorsement does not by itself establish content validity. The size of the panel is also small. Both are noted in the Limitations and Future Work section.

Table . Mapping between proposed questionnaire domains and constructs in prior instruments.

LLM-PCA[a] item	Domain	Antecedent constructs in prior instruments	Primary lineage source
G1. How concerned are you about the ways in which LLMs[b] use your personal data?	Data use (secondary/onward use)	Unauthorized secondary use, information use, use concerns	[6-9,11]
G2. How important is it for you to be informed about how LLMs use your data?	Awareness (notification)	Awareness of collection, awareness	[7,9,10]
G3. How much concern do you have for the privacy of your data when using applications or services that utilize LLMs?	Global privacy concern	Overall or general privacy-concern factor	[7,9]
G4. How important is transparency about data usage to your trust in LLM-based applications?	Transparency (trust-relevant)	Awareness, link to trusting beliefs	[7,9]
G5. How important is it for you to have control over how LLMs use your data?	Control	Control	[7,9]
G6. How often do you think about the potential privacy risks when using LLM-based applications?	Risk salience	Link to risk beliefs	[9]
G7. How important is it for you to know about the specific companies or organizations with whom your data is shared?	Third-party or onward sharing	Unauthorized secondary use, improper access, perceived surveillance, third-party sharing within use concerns	[6,8,9,11]

[a]LLM-PCA: large language model privacy concern assessment.

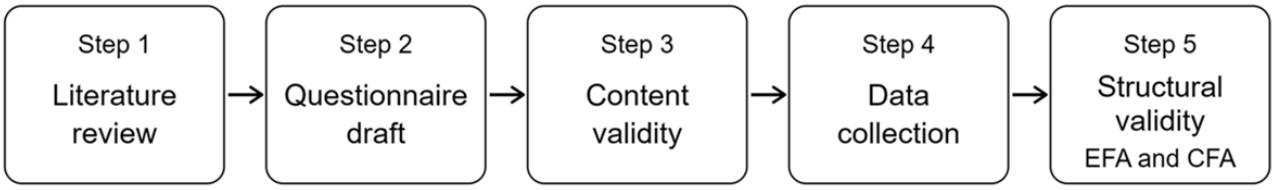
[b]LLM: large language model.

The revised instrument was then deployed on Qualtrics (Qualtrics International) for an internal pilot test conducted with the full research team. The pilot assessed the question-and-response logic, item flow, and platform implementation. Issues identified during the pilot were resolved prior to participant recruitment. In addition to the privacy-concern items, the survey collected demographic information, including age, race, sex, education level, marital

XSL•FO
RenderX

status, and annual income, to assess potential biases in respondents' answers, plus one screening question to determine whether potential participants had heard of LLMs. If they answered no or did not provide a response to this screening question, they would not be allowed to answer any questions on privacy concerns with LLMs. Responses from individuals unfamiliar with LLMs would not meaningfully reflect privacy concerns about them. The entire procedure is shown in Figure 1.

Figure 1. Flowchart of questionnaire development and validation. CFA: confirmatory factor analysis; EFA: exploratory factor analysis.



Participant Recruitment

Participants were eligible if they were adults (aged ≥18 years), understood English, and had awareness of LLM-based applications. Recruitment was conducted through three channels: (1) bulk email distributed to a random sample of students, faculty, and staff affiliated with the University of Pittsburgh; (2) a posting on LinkedIn; and (3) a posting on Pitt+Me, a research participant registry managed by the University of Pittsburgh. Interested individuals accessed the survey via a link directing them to the Qualtrics platform. When the university's electronic bulk-mail delivery service received a request for an announcement, it randomly selected the targeted audience (faculty, students, and staff in this case) and distributed the announcement. We used the service and made 2 announcements about the study in April and May 2025. In total, the message was delivered to several thousand faculty, students, and staff at Pitt. Study participation was fully voluntary, and we did not pay any compensation to the study participants. The same recruitment message was posted on the website of Pitt+Me in April, and anyone who visited the website in the study period (April 2025 to July 2026) could see the recruitment message and follow the link to answer the questions in the questionnaire.

Ethical Considerations

This study was approved by the University of Pittsburgh Human Research Protection Office (protocol ID: STUDY24110036). Participation was entirely voluntary; participants could withdraw at any time and request removal of their data without penalty.

Data Preprocessing

Prior to analysis, the raw dataset was cleaned using the following sequential criteria. First, records collected during the internal pilot phase were excluded. Second, responses that were entirely blank and responses that contained only demographic answers with no answers to any privacy concern item were removed as nonparticipating records because these entries could not provide any information for the privacy concern questionnaire. Third, participants who indicated that they had never heard of LLMs were excluded, as responses from individuals with no awareness of the technology were considered uninformative for the study's objectives. They were not allowed to answer any further questions in the questionnaire. Fourth, participants who did not answer any privacy concern–related questions were excluded since their data would not be useful for this specific study. These 4 steps removed records that should not be included in the dataset. The fifth step was a decision about how to handle missing data. We chose to only keep records with complete answers to those privacy concern questionnaire items. We adopted complete-case analysis because no imputation approach is free of assumptions, and imputed values could alter the item distributions on which the psychometric analyses depend. Records missing demographic data only were retained since they would not impact the evaluation of the large language model privacy concern assessment (LLM-PCA) questionnaire itself. Finally, administrative columns (eg, response ID, timestamp, and language) were removed from the dataset.

To evaluate whether restricting analyses to complete cases introduced systematic bias, we compared respondents retained for complete-case analysis with those excluded due to partial item nonresponse across all study variables. Continuous variables were compared using independent samples *t* tests, and categorical variables using Pearson chi-square tests of independence. A 2-sided α of .05 was applied. Because these tests were intended to detect potential nonresponse bias rather than to establish statistical equivalence, nonsignificant differences were interpreted as consistent with, but not confirmation of, ignorable missingness.

Statistical Analysis

Descriptive statistics were computed for all demographic variables, with frequencies and percentages reported. Item-level descriptive statistics (mean, SD, count of respondents to each option, and percentage) were computed for each item in the privacy concern questionnaire prior to factor analysis. Psychometric analyses were then conducted on the privacy concern items. Internal consistency reliability was assessed using Cronbach α and McDonald ω, with values ≥0.70 considered acceptable [33,34].

To examine the dimensionality and measurement properties of the scale, we used a split-sample cross-validation design: the complete-case sample was randomly partitioned into 2 parts, with roughly one-third used for EFA and the remaining two-thirds reserved for CFA. Because the total sample was modest for a split-sample design, we treat the EFA as exploratory and hypothesis-generating; confirmatory replication in a larger and independent sample remains necessary, and the reported CFA fit should be interpreted accordingly [32,34].

Because the items were measured on a 5-point Likert scale, we computed a polychoric correlation matrix to characterize

XSL•FO
RenderX

interitem association. Polychoric correlations model an underlying continuous latent response for each ordered-categorical indicator and are the principled measure of association for Likert items. Sampling adequacy was evaluated using the Kaiser-Meyer-Olkin (KMO) measure and Bartlett test of sphericity, both computed from the polychoric matrix. The number of factors to retain was determined by parallel analysis with 500 iterations of resampled data, using the 95th percentile of simulated eigenvalues as the cutoff [35]; Kaiser's eigenvalue-greater-than-one criterion was reported as a secondary index. Factor extraction used maximum likelihood (ML) on the polychoric correlation matrix; when more than one factor was retained, promax (oblique) rotation was applied. ML extraction on a polychoric matrix assumes multivariate normality of the underlying continuous responses, an assumption we adopted as reasonable for ordinal Likert indicators.

A CFA model with all retained items loading on a single latent privacy concern factor was fitted to the CFA subsample using diagonally weighted least squares (DWLS), the appropriate estimator for ordered-categorical Likert indicators. The latent factor variance was fixed to 1.0 to identify the model, allowing all factor loadings to be freely estimated.

Model fit was evaluated at both the global and local levels because satisfactory global fit does not guarantee that a model reproduces every element of the observed covariance structure. Global fit was assessed using the model chi-square statistic with its degrees of freedom and associated *P* value, the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA) with its 90% CI, the standardized root mean square residual (SRMR), goodness-of-fit index (GFI), and adjusted goodness-of-fit index (AGFI). The GFI and AGFI are reported for completeness only and are not used to support any conclusion about fit. Conventional cutoffs for acceptable fit were CFI and TLI ≥0.90 (≥0.95 for good fit), RMSEA ≤0.08 (≤0.06 for good fit), and SRMR ≤0.08 (≤0.05 for good fit). These values were treated as descriptive heuristics and not as decision thresholds. They were derived under normal-theory ML with continuous data [32], and simulation work shows that diagonally weighted and unweighted least squares applied to polychoric matrices yield systematically smaller RMSEA and larger CFI and TLI values than ML, so that applying the conventional cutoffs under these estimators tends to understate misfit [36]. Because the model chi-square test has limited power to detect misspecification at moderate sample sizes and the RMSEA is estimated with appreciable sampling variability, we additionally evaluated close fit using the test of close fit (PCLOSE [*P* value for the test of close fit]), which assesses the null hypothesis that the population RMSEA is no greater than .05. Because the model was estimated with DWLS on the polychoric correlation matrix, particular interpretive weight was given to the SRMR, which is the most robust and least biased of the common fit indices under this estimator and ordered-categorical data. The RMSEA was interpreted together with the width of its CI. Local fit was examined through the matrix of residual correlations, computed as the difference between the observed polychoric correlations and the correlations implied by the fitted model. Individual residual correlations were inspected to identify item pairs whose association the model failed to reproduce; absolute residual correlations greater than 0.10 were treated as indicators of localized strain warranting further examination [37]. The magnitude and pattern of the largest residuals were reported to characterize the nature of any local misfit, and their substantive interpretation was considered alongside the global indices in evaluating the adequacy of the model. Global fit was therefore interpreted jointly with this local fit evidence and with the nested model comparison, rather than by asking whether any single index cleared a cutoff.

Composite reliability (ω) was reported as an indicator of internal consistency, and the average variance extracted (AVE) as an indicator of convergent validity. Where a model with more than 1 factor was retained, discriminant validity between factors was assessed using the Fornell-Larcker criterion [38], which requires the AVE for a factor to exceed the squared correlation between factors, and the heterotrait-monotrait ratio of correlations (HTMT)<0.90 because values of HTMT above 0.90 indicate that factors are not empirically distinct. Reliability and AVE were computed for the full item set and for the indicators defining each factor in any multifactor solution. Composite reliability was computed as $\omega=(\Sigma\lambda)^2/[(\Sigma\lambda)^2+\Sigma(1-\lambda^2)]$, and AVE as $AVE=(\Sigma\lambda^2)/k$, where λ is the standardized loading of indicator i, and k is the number of indicators. Conventional thresholds of ω≥0.80 and AVE ≥0.50 were used [34,38].

To assess sensitivity to the random partition, the full procedure, including the item-refinement step, exploratory analysis, and confirmatory analysis, was repeated under 5 random partitions, with all item-reduction decisions made using the exploratory subsample only.

Analyses were conducted in SPSS (version 31; IBM Corp) and Python (version 3.13; Python Software Foundation) using the *pyreadstat* package for SPSS file input, *factor_analyzer* for EFA, and *semopy* for CFA [39]. Path diagrams were first visualized using semopy.semplot and postprocessed for publication.

Sample Size Considerations

Minimum sample size was estimated during participant recruitment from the general rule of thumb of 10 respondents per item for factor analysis. No formal a priori power analysis was conducted, so recruitment remained open for the duration of the fielding period, and the achieved sample was determined by the response to the 3 recruitment channels. The benchmarks set out below are the a priori criteria, against which the planned analyses were judged to be adequately powered. Whether the achieved subsamples met those benchmarks and whether the data satisfied the conditions the benchmarks assume are addressed in the Results and Limitations and Future Work sections.

For the EFA, the required sample size is not fixed by absolute N or by the subject-to-variable ratio alone but depends primarily on the level of communality and the degree to which factors are overdetermined [40]. When communalities are high and a single factor is well defined by multiple indicators, population factors can be recovered accurately even in small samples. de Winter et al [41] similarly demonstrated that well-conditioned data,

XSL•FO
RenderX

characterized by high loadings, few factors, and an adequate number of indicators per factor, yield reliable EFA solutions at sample sizes well below conventional minimums. Neither condition can be verified before the data are collected, so they were treated as criteria to be checked afterward rather than as a basis for setting a target. The a priori benchmark applied was therefore the subject-to-item ratio, for which 10:1 is the common minimum and 20:1 is at the upper end of published recommendations [42]. For a 6-item scale, these ratios correspond to 60 to 120 respondents in the exploratory subsample.

For the CFA, sample size adequacy is commonly evaluated using the ratio of cases to estimated parameters (the N:q rule), for which a ratio of 20:1 is considered ideal and 10:1 an acceptable minimum [37]. A 6-item single-factor specification estimates 12 free parameters (1 loading and 1 residual variance per item), and 1 more for a 2-factor specification, so the rule implies a minimum of 120 to 130 and an ideal of 240 to 260 cases in the confirmatory subsample. Simulation evidence indicates that the requirement for simple confirmatory factor models depends strongly on loading magnitude, with stronger loadings requiring fewer cases [43]; because loading magnitude is itself estimated from the data, that consideration could only be applied after estimation.

In total, if there are at least 180 respondents, both EFA and CFA can be performed. To achieve a more reliable result, the desired sample size is 360.

Results

Candidate Items in the New Questionnaire

The 7 candidate items were written so that each targets a distinct, narrowly defined privacy domain drawn from the synthesis mentioned earlier. The domains retain the field's replicated core (data use, awareness, and control) and extend it to LLM-salient concerns (transparency-for-trust, risk salience, and onward third-party sharing). Table 1 maps each item to its intended domain and to the antecedent constructs in the 6 foundational instruments.

The instrument is deliberately anchored to the most replicated regions of the construct while remaining fit for its new referent. G3 provides a global concern item analogous to overall privacy-concern factors. G2 (awareness) and G1 (data use) instantiate the 2 domains that recur most consistently across the lineage, giving the scale continuity with CFIP, IUIPC, MUIPC, IPC, and the short-form video scale. G5 (control) preserves the control domain, introduced in IUIPC and carried forward in IPC. For LLMs, the domain is especially consequential, because control is exercised through concrete and far-reaching choices: whether user inputs may be used to train future models, and whether conversation history persists across sessions.

The remaining 3 items (G4, G6, and G7) extend these core domains to data practices specific to LLMs. The comparative synthesis of the 6 prior instruments showed that none of them covers these concerns, so each item adds coverage rather than duplicating a domain already represented. G4 reframes awareness as transparency in the service of trust, operationalizing the concern-trust relationship that Hong and Thong [9] established as part of privacy concern's nomological network and that is acute for opaque, multiparty model providers. G6 (risk salience) captures how frequently the risk is cognitively present, a facet motivated by the demonstrated but often underrecognized LLM risks of memorization and inference [17,18] and by evidence that users disclose to conversational agents under flawed mental models [16]. G7 (onward sharing) sharpens the secondary-use and improper-access concerns of CFIP and IPC into the specific, LLM-relevant question of which third parties receive user data, reflecting the provider ecosystem described earlier. Two items (G2 and G4) intentionally draw on the same awareness or transparency region from complementary angles, notification versus transparency-for-trust, broadening coverage of that domain within a unidimensional design. Framing all 7 items as facets of a single privacy-concern construct, rather than as instruments of one behavioral theory, directly answers the fragmentation in recent LLM survey work, that is, APCO-, uncertainty reduction-, and contextual integrity–based measures each captured a different slice of concern and could not be compared [1-4], whereas the LLM-PCA is designed to be theory-agnostic at the point of measurement.

Content Validity

All 7 items demonstrated strong content validity on the relevance dimension (Table S1 in Multimedia Appendix 1). Five items (G2 to G6) were endorsed as relevant by all 8 experts (I-CVI=1.000), and the remaining 2 items (G1 and G7) were each endorsed by 7 of the 8 experts (I-CVI=0.875). Every item thus exceeded the 0.78 retention criterion, as well as the more stringent 0.83 threshold recommended for panels of 6 to 8 experts [29]. Chance-corrected agreement was uniformly excellent (modified $\kappa^* \geq 0.871$), confirming that the observed endorsement was not attributable to chance [31].

At the scale level, the averaging method yielded an S-CVI/Ave of 0.964, exceeding the 0.90 benchmark for satisfactory content validity. The more conservative UA index was S-CVI/UA=0.714, reflecting the 2 items (G1 and G7) on which a single expert each assigned a rating below 3.

Clarity ratings were also favorable, with item-level indices ranging from 0.750 to 1.000 and an S-CVI/Ave of 0.875. One item (G2) fell below the endorsement threshold on clarity alone (I-CVI=0.750, 2 dissenting experts) while retaining unanimous relevance endorsement. It was retained without modification, although its wording may warrant minor refinement in a future revision.

On the basis of these relevance results, none of the 7 candidate items were removed or reworded, and all were administered to study participants on Qualtrics. The relevance criteria were met by every item, which is the basis for that decision. The one shortfall not acted on was the clarity rating for G2 where I-CVI<0.78. Comprehension of wording is better assessed by cognitive interviewing than by expert rating. Its wording is therefore identified as a known weakness of the current version. It needs to be reassessed using cognitive interviewing in future revisions. The entire questionnaire used in this study is available in Multimedia Appendix 2.

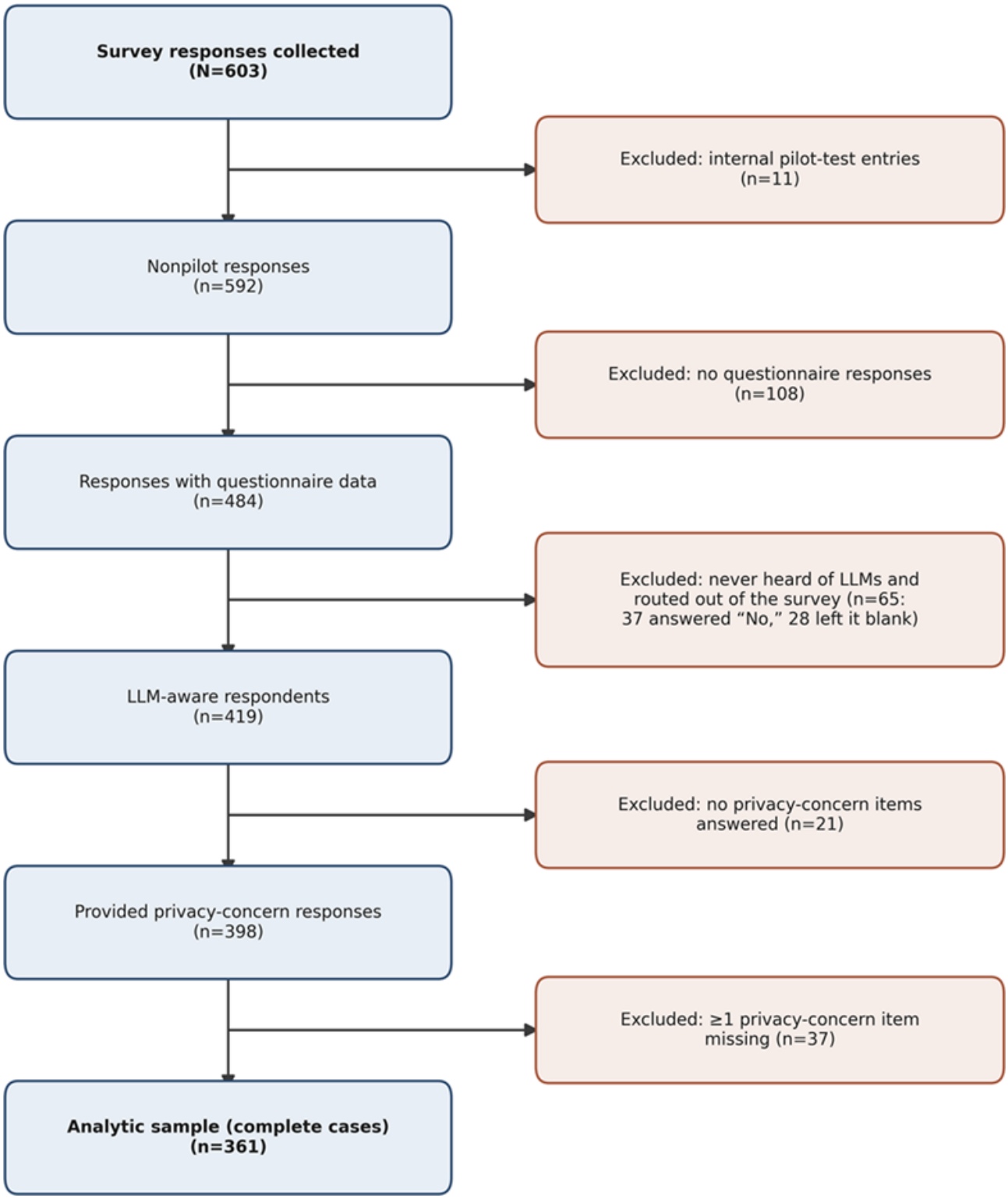
XSL•FO
RenderX

Participant Characteristics

A total of 603 survey responses were collected. We first removed 11 responses generated during internal pilot testing, leaving 592 responses. Of these, 108 contained no questionnaire data (no answer to any questions) and were excluded, yielding 484 responses. A further 65 respondents reported no awareness of LLMs (37 answered “No” and 28 left the screening item blank) and were routed out of the survey without completing the remaining items; these were excluded, leaving 419 respondents. In the next step, 21 respondents who did not answer any privacy concern survey questions were excluded from the dataset, leaving 398 eligible respondents. Of the 398 eligible respondents, 361 (90.7%) provided complete data on all analysis variables and 37 (9.3%) had partial item nonresponse. Complete and incomplete respondents did not differ significantly on any measured characteristic (Table S2 in Multimedia Appendix 1). Therefore, 37 respondents who omitted 1 or more privacy concern survey items were excluded to retain complete cases only. The resulting analytic sample comprised 361 respondents. This procedure is shown in Figure 2.

XSL•FO
RenderX

Figure 2. Flowchart of data preprocessing. LLM: large language model.



Of the 361 participants, more than half were younger than 35 years (n=198, 54.8%), most identified as White (n=285, 78.9%), nearly two-thirds were female (n=230, 63.7%), and 88.1% (n=318) held at least a bachelor's degree. Distributions for all demographic variables are presented in Table 2. Overall, the sample was predominantly White, highly educated, and female, which should be considered when interpreting the generalizability of the factor structure and score distributions.

XSL•FO
RenderX

Table . Demographics of study participants (N=361).

Demographics	Values, n (%)
Age group (years) (n=342)	
18 - 24	141 (39.1)
25 - 34	57 (15.8)
35 - 44	44 (12.2)
45 - 54	44 (12.2)
55 - 64	26 (7.2)
≥65	30 (8.3)
Race (n=358)	
Asian or Pacific Islander	31 (8.6)
Black or African American	12 (3.3)
Hispanic or Latino	12 (3.3)
White	285 (78.9)
Mixed race	15 (4.2)
Other	3 (0.8)
Sex (n=359)	
Male	109 (30.2)
Female	230 (63.7)
Nonbinary or prefer not to answer	20 (5.5)
Education level (n=360)	
High school or less	6 (1.7)
Some college or associate degree	34 (9.4)
Bachelor’s degree	117 (32.4)
Master’s degree	102 (28.3)
Professional degree (eg, MD and JD)	24 (6.6)
Doctorate (PhD and EdD)	75 (20.8)
Other	2 (0.6)
Marital status (n=358)	
Single or never married	153 (42.4)
Married or domestic partnership	175 (48.5)
Separated	3 (0.8)
Divorced	18 (5)
Widowed	9 (2.5)
Employment status (n=357)	
Employed for wages (full-time)	242 (67)
Employed for wages (part-time)	23 (6.4)
Self-employed	10 (2.8)
Out of work	9 (2.5)
Homemaker	2 (0.6)
Student	47 (13)
Retired	21 (5.8)
Unable to work	3 (0.8)
Annual income (n=358)	

XSL•FO
RenderX

Demographics	Values, n (%)
Less than US $20,000	31 (8.6)
Between US $20,000 and $50,000	86 (23.8)
Between US $50,000 and $80,000	91 (25.2)
Between US $80,000 and $110,000	55 (15.2)
Between US $110,000 and $140,000	30 (8.3)
Between US $140,000 and $180,000	15 (4.2)
Between US $180,000 and $240,000	12 (3.3)
Between US $240,000 and $300,000	7 (1.9)
More than US $300,000	3 (0.8)
Prefer not to answer	28 (7.8)

Item Descriptive Statistics

Item-level descriptive statistics for the 7 privacy concern items are presented in Table 3, including mean, SD, frequency of participants choosing each option in each question, and anchor type (concern, importance, or frequency). Items were rated on a Likert scale from 1 (extremely important, extremely concerned, or always) to 5 (not at all important, not at all concerned, or never). Therefore, lower values indicate stronger endorsement. Scoring, including the reverse-coding step, is specified in the Methods section. Item means ranged from 1.94 (G5) to 2.51 (G3) on the 1 - 5 Likert scale, indicating that, on average, participants endorsed moderate-to-high levels of privacy concern (with lower values indicating stronger endorsement). Item-level SDs were between 0.97 and 1.18. Because item dispersion was narrow and similar across items, a unit-weighted composite and one formed from standardized item scores would differ only slightly in this sample. All items used the full 1 - 5 response range. Skewness and kurtosis were calculated as well but not included in the table. Skewness ranged from 0.18 to 0.85 (maximum |skew|=0.85), and kurtosis ranged from −0.94 to 0.04 (maximum |kurtosis|=0.94), all well within conventional bounds (|skew|<2, |kurtosis|<7) for treating items as approximately continuous. Nevertheless, given the 5-point Likert response format, we used polychoric correlations for EFA and DWLS estimation for CFA throughout.

Table . Descriptive statistics and answer distribution for the 7 privacy concern items (N=361)[a].

Item	Mean (SD)	1, n (%)	2, n (%)	3, n (%)	4, n (%)	5, n (%)	Anchor type
G1	2.50 (1.14)	91 (25.2)	86 (23.8)	109 (30.2)	63 (17.5)	12 (3.3)	Concern
G2	2.05 (0.97)	122 (33.8)	131 (36.3)	79 (21.9)	25 (6.9)	4 (1.1)	Importance
G3	2.51 (1.18)	98 (27.1)	76 (21.1)	107 (29.6)	66 (18.3)	14 (3.9)	Concern
G4	2.07 (1.00)	125 (34.5)	122 (33.8)	85 (23.5)	22 (6.1)	7 (1.9)	Importance
G5	1.94 (0.99)	151 (41.8)	113 (31.3)	70 (19.4)	22 (6.1)	5 (1.4)	Importance
G6	2.36 (1.08)	89 (24.7)	120 (33.2)	94 (26)	48 (13.3)	10 (2.8)	Frequency
G7	2.37 (1.01)	74 (20.5)	139 (38.5)	94 (26)	48 (13.3)	6 (1.7)	Importance

[a]Data in columns 1 - 5 indicate the frequency of participants choosing one particular option as the answer to each question.

Instrument Refinement

The questionnaire was initially developed with 7 items, 1 for each privacy-concern domain. Before finalizing the scale, we examined the interitem polychoric correlation matrix and the fit of the 7-item unidimensional model on the confirmatory subsample (n=240). This examination revealed that items G1 ("How concerned are you about the ways in which LLMs use your personal data?") and G3 ("How much concern do you have for the privacy of your data when using applications or services that use LLMs?") were nearly collinear, with a polychoric correlation of 0.91. This indicated that both items assess a respondent's privacy concern with data use, and the strong correlation indicated that they functioned as effectively interchangeable indicators.

To determine which item to remove, we compared 2 reduced 6-item models, one omitting G1 and one omitting G3. Removing G3 produced uniformly good fit (χ^2_9=13.6; *P*=.14; RMSEA=0.046; SRMR=0.057), whereas removing G1 produced only marginally acceptable fit (χ^2_9=17.4; *P*=.04; RMSEA=0.062; SRMR=0.065). In the G1-omitted model, the retained item G3 was the weakest indicator (standardized λ=0.74; communality=0.47), whereas in the G3-omitted model, all retained items loaded more strongly. Because removing G3 yielded superior model fit and stronger indicators, we retained G1 and removed G3. The resulting 6-item scale (G1, G2, G4, G5, G6, and G7) was carried forward for all analyses reported below. Removal of an empirically redundant, conceptually overlapping item is a standard step in scale development; retaining both would have overweighted a single domain and

XSL•FO
RenderX

degraded model fit without adding measurement information. As reported in the Stability Across Alternative Sample Splits section, this removal decision was reproduced in 5 independent exploratory subsamples without consulting the confirmatory subsample.

Reliability

Cronbach α for this 6-item questionnaire was 0.90. Because α increases with both the number of items and the average interitem correlation, this value should be interpreted as reflecting high interitem redundancy in a short scale rather than as independent evidence of measurement quality [33]; McDonald ω, computed from the CFA (ω=0.90), is reported below as a complementary estimate.

EFA

EFA was conducted on the randomly selected EFA subsample (n=121). ML extraction was applied to the polychoric correlation matrix. The polychoric correlations (Table S3 in Multimedia Appendix 1) demonstrated strong but nonredundant interitem association, with off-diagonal correlations ranging from 0.45 (G6 vs G7) to 0.79 (G2 vs G5); removing G3 eliminated the near-collinear pair present in the original 7-item pool. The matrix was well-suited to factor analysis since the overall KMO measure of sampling adequacy was 0.87, in the "good" range, with per-item KMO values from 0.84 (G5) to 0.92 (G7), all above the 0.60 minimum. Bartlett test of sphericity was significant (χ^2_{15}=477.4; P<.001), confirming that the correlation matrix departed substantially from an identity matrix.

Parallel analysis based on 500 simulated datasets supported a 1-factor solution in the exploratory subsample. The first observed eigenvalue (4.17) substantially exceeded the corresponding 95th-percentile random eigenvalue (1.44), whereas the second observed eigenvalue (0.65) fell below its random benchmark (1.25). Kaiser's eigenvalue-greater-than-one criterion likewise retained a single factor. The ratio of the first to the second eigenvalue (6.46) further supported essential unidimensionality. The first eigenvalue accounted for 69.6% of total variance, and after ML extraction, the single-factor solution accounted for 63.6% of common variance. Standardized loadings on the retained factor (Table 4) were uniformly substantial in absolute value, ranging from 0.71 (G7) to 0.88 (G5), and all 6 items exceeded the stricter conventional λ≥0.70 threshold for strong indicator quality. The items with the weakest loadings (G6 and G7) may be candidates for rewording or replacement in future work. Item communalities ranged from 0.50 (G7) to 0.78 (G5). The pattern of eigenvalues, the parallel-analysis decision, and the uniformly high single-factor loadings provided convergent evidence for a dominant general factor in the exploratory subsample and supported fitting the planned single-factor CFA model on the holdout subsample.

Table . Standardized factor loadings from the exploratory factor analysis (EFA; n=121) and from the confirmatory factor analysis (CFA; n=240) for the 6-item large language model privacy concern assessment.

Item	EFA λ[a]	EFA h^2	CFA λ[b]	CFA R^2	Residual[c]
G1	0.77	0.60	0.79[d]	0.62	0.38
G2	0.85	0.72	0.85[d]	0.73	0.27
G4	0.82	0.67	0.83[d]	0.68	0.32
G5	0.88	0.78	0.81[d]	0.65	0.35
G6	0.74	0.55	0.69[d]	0.47	0.53
G7	0.71	0.50	0.69[d]	0.48	0.52

[a]Maximum likelihood extraction on the polychoric correlation matrix.

[b]Diagonally weighted least square estimation on the polychoric correlation matrix, with the latent factor variance fixed to 1.0.

[c]Standardized residual variance.

[d]P<.001.

CFA

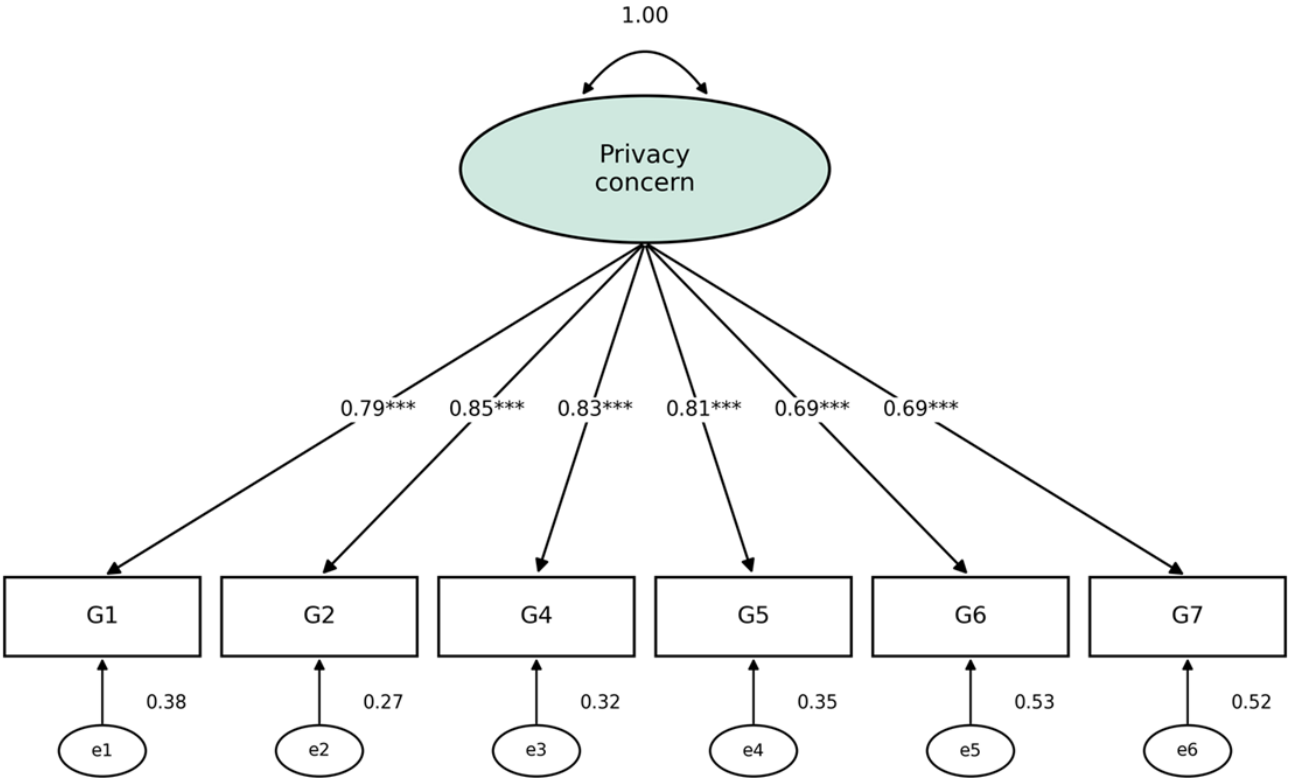
Global Fit

The unidimensional 6-item CFA model was fitted to the holdout subsample (n=240) using DWLS estimation. The model exhibited excellent global fit: χ^2_9=13.6; P=.14; CFI=0.996; TLI=0.993; RMSEA=0.046, 90% CI 0.000-0.093; SRMR=0.057. The nonsignificant model chi-square and CFI/TLI values approaching unity indicated that the unidimensional 6-item model reproduced the observed interitem associations closely. The RMSEA point estimate (0.046) fell within the range conventionally considered to reflect close fit (≤0.06), and the test of close fit was clearly nonsignificant (PCLOSE=.50, testing H : RMSEA≤0.05). SRMR (0.057) fell well within the range for acceptable fit. GFI and AGFI were strong (GFI=0.99; AGFI=0.98). All 6 standardized factor loadings were positive, substantial, and statistically significant (P<.001), with values of 0.79 (G1), 0.85 (G2), 0.83 (G4), 0.81 (G5), 0.69 (G6), and 0.69 (G7). Standardized loadings, R^2 values, and standardized residual variances are presented in Table 4. In total, 4 of the 6 items exceeded the conventional λ≥0.70 threshold for strong indicator quality, and the remaining 2 (G6 and G7, both λ=0.69) fell just below it while remaining well above the minimum acceptable value of 0.50. Item R^2 values ranged from 0.47 (G6) to 0.73 (G2), indicating that the latent privacy concern factor accounted for between 47% and 73% of the variance in each

XSL•FO
RenderX

item. Figure 3 shows the path diagram used in the CFA for the unidimensional model with 6 items.

Figure 3. Confirmatory factor analysis of the 6-item large language model-privacy concern assessment questionnaire. Standardized factor loadings are shown on the paths from the latent factor to each indicator; standardized residual variances are shown next to each error term (e1-e6). The variance of privacy concern was fixed at 1.00 for identification of the standardized solution. All loadings significant at ***P<.001.



Composite reliability was excellent (ω=0.90) above the conventional 0.80 threshold for very good internal consistency, and the AVE was 0.61, exceeding the 0.50 Fornell-Larcker convergent-validity threshold. Taken together, these results support the internal consistency and convergent validity of the refined LLM-PCA at the level of the full 6-item set. The dimensional structure is addressed in the Comparison of 1-Factor and 2-Factor Models section.

Local Fit

Because strong global fit does not guarantee adequate fit at the level of individual item pairs, we examined the residual correlation matrix (Table S4 in Multimedia Appendix 1). Residuals are the differences between observed polychoric correlations and model-implied correlations. Most residuals were small, but 4 of the 15 off-diagonal residual correlations exceeded the 0.10 threshold: G1-G6 (residual=0.21), G2-G5 (residual=0.18), G4-G5 (residual=0.15), and G2-G4 (residual=0.14). The root mean square residual (computed on the polychoric matrix) across all off-diagonal elements was 0.098. In total, 3 of the 4 elevated residuals involved the transparency- and control-oriented items G2 (being informed about data use), G4 (transparency supporting trust), and G5 (control over data use), indicating that these conceptually adjacent items share covariance beyond what the general privacy-concern factor accounts for. The remaining elevated residual, between G1 (data use) and G6 (frequency of thinking about privacy risks), reflects a smaller secondary association. These residuals indicate that although a strictly unidimensional model provides excellent global fit and a robust SRMR (computed on the Pearson covariance), it does not fully reproduce the local covariance among the transparency- and control-related items.

Comparison of 1-Factor and 2-Factor Models

Because the residual analysis indicated localized covariance among the transparency- and control-related items, we compared the unidimensional model (M1) with a 2-factor alternative (M2) on the same CFA subsample (n=240). The 2-factor model assigned the transparency- and control-oriented items to 1 factor (G2, G4, and G5) and the remaining items to a second factor (G1, G6, and G7), with the 2 factors allowed to correlate. Both models were estimated with DWLS on the polychoric correlation matrix. Fit statistics for the 2 models are presented in Table 5.

XSL•FO
RenderX

Table . Comparison of 1-factor and 2-factor confirmatory models (confirmatory factor analysis subsample, n=240).

Model	χ^2 (*df*)	*P* value	CFI[a]	TLI[b]	RMSEA[c] (90% CI)	SRMR[d]	AIC[e]	BIC[f]
One-factor (M1)	13.6 (9)	.14	0.996	0.993	0.046 (0.000-0.093)	0.057	23.21	64.97
Two-factor (M2)	4.7 (8)	.79	1.00	1.00	0.000 (0.000-0.050)	0.032	25.74	70.99

[a]CFI: comparative fit index.

[b]TLI: Tucker-Lewis index.

[c]RMSEA: root mean square error of approximation.

[d]SRMR: standardized root mean square residual.

[e]AIC: Akaike information criterion.

[f]BIC: Bayesian information criterion.

The 2-factor model fits better than the 1-factor model on every index examined: (χ^2_8=4.7; *P*=.79 vs χ^2_9=13.6; *P*=.14; CFI and TLI of 1.00 vs 0.996 and 0.993; RMSEA=0.000, 90% CI 0.000-0.050 vs 0.046, 90% CI 0.000-0.093; SRMR=0.032 vs 0.057). The nested chi-square difference test was significant ($\Delta\chi^2_1$=8.9; *P*=.003), consistent with the localized residual covariance identified earlier. Therefore, the 2-factor specification is the better-fitting model on these data, and the improvement is not marginal. Table 5 also reports the Akaike information criterion and the Bayesian information criterion, both of which favor the 1-factor model. These criteria are likelihood-based, and because the models were estimated by DWLS rather than by ML, the values are derived from the fit function rather than a true likelihood and are not interpretable on their usual scale. They are reported for completeness, and the nested chi-square difference test was used for the model comparison.

We therefore retain the 2-factor solution as the measurement model for the LLM-PCA. The first factor is defined by the transparency- and control-related items (G2, being informed about data use; G4, transparency supporting trust; and G5, control over data use) and is labeled transparency and control. The second is defined by the remaining items (G1, concern about data use; G6, frequency of thinking about privacy risks; and G7, onward third-party sharing) and is labeled data use and risk salience. Standardized loadings in the 2-factor solution ranged from 0.85 to 0.90 on the first factor and from 0.72 to 0.85 on the second.

One property of this solution affects how it should be used. The 2 latent factors correlated at *r*=0.83; therefore, they share approximately 69% of their variance. The evidence on discriminant validity is mixed. The HTMT was 0.85 (<0.90), which indicates that these 2 factors are empirically distinct. The Fornell-Larcker criterion was met for transparency and control, whose AVE (0.77) exceeds the squared factor correlation (0.69), but not for data use and risk salience, whose AVE (0.59) falls below it. In other words, the 2 factors are separable but not fully distinct. The subscale reliability was acceptable (transparency and control: α=0.91; ω=0.91; data use and risk salience: α=0.81; ω=0.81). For these reasons, the 6-item total score remains the recommended primary metric for applied use, and the subscale scores, each a unit-weighted mean of its 3 items, should be reported as secondary, facet-level information.

Stability Across Alternative Sample Splits

To verify that the reported results were not an artifact of the particular random partition, the entire split-sample procedure (item refinement, EFA, and CFA) was repeated under 5 different random partitions, holding the one-third exploratory, two-thirds confirmatory allocation fixed. The findings were stable across all 5 splits. The near-collinearity of the 2 general-concern items exceeded the 0.85 threshold in every exploratory subsample, and the reduced-model comparison, conducted on the exploratory subsample without consulting the confirmatory subsample, indicated removal of G3 in all 5 cases. Parallel analysis retained a single factor in every exploratory subsample (first-to-second eigenvalue ratio=4.3 - 7.6; KMO=0.82-0.87; minimum exploratory loading=0.57-0.71). The confirmatory model showed acceptable-to-excellent global fit on every holdout subsample (χ^2_9=7.7 - 15.8; all *P*≥.07; CFI≥0.993; RMSEA≤0.056; SRMR≤0.064; PCLOSE≥.36), with composite reliability ω=0.90-0.91 and AVE=0.59-0.64. The localized residual strain among the transparency- and control-related items replicated in every split (4-5 residual correlations exceeding 0.10 per split), and G6 and G7 were the weakest indicators in each confirmatory solution (standardized λ=0.67-0.75), consistent with the pattern reported earlier.

Discussion

Principal Findings

This study developed and validated a concise 6-item instrument, the LLM-PCA, for assessing users' privacy concerns regarding LLM-based systems. There are 4 major findings in this study. First, the 7 candidate domains could be reduced to 6 items without loss of content coverage, because 2 general concern items proved to be near-interchangeable indicators. Therefore, one of them was removed from the instrument. Second, the final instrument demonstrated high internal consistency and a clearly dominant general factor in the exploratory subsample. Third, a confirmatory model fitted to an independent holdout subsample reproduced the observed interitem associations closely, and the whole procedure replicated across 5 alternative random partitions of the sample, indicating that the findings are not an artifact of one split. Fourth, the covariance among the transparency- and control-related items is not fully explained by a single general factor. A 2-factor model fit the data better

XSL•FO
RenderX

than a 1-factor model and was retained as the measurement model, separating a transparency and control facet from a data use and risk salience facet. The 2 factors met one of the 2 conventional discriminant validity criteria but not the other, so the instrument is offered as a brief 2-facet measure whose total score remains the recommended primary metric for privacy concern with LLM applications, with facet scores providing secondary detail. To our knowledge, it is among the first peer-reviewed instruments developed specifically for this referent, addressing a gap in the growing body of survey-based research in this area.

Comparison With Prior Work

As described in the Introduction section, prior LLM privacy surveys each used a bespoke instrument built for a different theoretical framework and a different outcome; therefore, the concern levels they report cannot be placed on a common metric with LLM-PCA scores. The comparison drawn here is one of structure and measurement properties rather than of score levels. We do not claim that the instrument derives from, or validates, any single theoretical framework; establishing its theoretical grounding and construct validity will require further work. Table S5 in Multimedia Appendix 1 summarizes the comparison of the LLM-PCA with prior privacy concern instruments and recent LLM privacy studies.

The factor structure observed here should be read against prior work. Earlier general privacy concern scales, such as those developed for organizational, e-commerce, and short-form video platform contexts, have often yielded multifactor solutions capturing dimensions such as collection, unauthorized access, secondary use, and awareness [6,7,10,11]. The present results are consistent with that literature. A 2-factor solution separating transparency and control items from data use and risk salience items fit better than a single-factor solution, and those 2 facets correspond closely to the control and awareness dimensions identified in earlier instruments [7,9]. What the present design cannot do is resolve those facets into well-identified factors, because each domain is represented by a single item and each candidate factor is therefore defined by only 3 indicators. Reporting a single composite score is a decision about parsimony and applied usability at this scale length, not a finding that users hold an undifferentiated view of privacy. Whether privacy concern with LLMs is best represented by 1 general factor or by 2 correlated facets is a question for a longer instrument with multiple indicators per domain. The possibility that an early stage of public understanding contributes to the high correlation between the 2 facets is consistent with findings from Choi et al [1], who found that 66% of their sample had limited knowledge of available privacy protection options in LLMs, and from Liu et al [3], who observed a systematic underestimation of health data risk and a gap between users' stated privacy concerns and their actual expectations. As public understanding of LLM data practices matures, multidimensional solutions may emerge in future validation studies involving more experienced user populations.

Implications

The validated instrument offers practical utility for both researchers and practitioners. For researchers, the availability of a standardized 6-item scale makes it feasible to incorporate privacy concern assessment as a routine component of LLM user studies without imposing significant respondent burden. Its single composite score, with demonstrated reliability, enables meaningful cross-study comparisons, a capability that has been largely absent from the existing literature. For practitioners and AI application developers, the scale provides a concise, validated tool for benchmarking user privacy concern levels before and after system design changes, privacy policy updates, or new feature deployments, supporting data-driven evaluation of user-facing privacy interventions.

The scale's focus on LLM-specific concerns, including data use practices, transparency, user control, and organizational data sharing, makes it particularly relevant for the growing number of commercial and enterprise LLM deployments in health care, education, legal services, and other sensitive domains. In health care contexts, where prior work has shown that privacy concerns can shape users' engagement with AI-assisted consultation [1-3], this instrument could support both predeployment risk assessment and ongoing postdeployment monitoring of user trust. For example, health care organizations deploying LLM-based tools such as clinical decision support systems, patient-facing chatbots, or electronic health record–integrated assistants could administer the scale prior to deployment to establish a baseline of user privacy concern, and again following deployment to detect shifts in trust associated with system changes or privacy policy updates. Because the instrument yields a single interpretable score, it is practical for routine use in clinical and operational settings without requiring psychometric expertise. As a preliminary interpretive aid pending formal norming, and after reverse-coding so that higher scores denote greater concern, mean item scores in the upper part of the scale (roughly ≥4 of 5) may be read as indicating high privacy concern, scores near the midpoint as moderate concern, and scores in the lower part (roughly ≤2) as low concern. These provisional bands are illustrative only and should be calibrated against representative normative samples before being used to guide decisions.

Limitations and Future Work

Several limitations should be considered when interpreting findings reported in this study. First, the sample skewed toward younger, White, highly educated, and predominantly female participants, characteristics that likely reflect the early-adopter demographic of LLM technology. Because privacy attitudes toward AI systems may vary with age, educational background, digital literacy, cultural background, and prior technology exposure, the generalizability of both the scale's factor structure and the observed levels of concern to more diverse populations cannot be assumed. In particular, the overrepresentation of highly educated individuals may mean that the scale's factor structure reflects a relatively privacy-literate population. Replication studies with demographically diverse and community-based samples are needed to evaluate measurement invariance across groups. Because the sex item combined nonbinary respondents with those who preferred not to answer into one category, the response set was less inclusive than current best practice and did not allow sex-diverse participants to be identified separately; future versions should offer distinct

XSL•FO
RenderX

nonbinary and self-describe options in addition to a prefer-not-to-answer choice. Because most respondents identified as White and held at least a bachelor degree, the score distributions reported here should not be treated as normative values for the US adult population, and the provisional interpretive bands suggested in the Implications section should not be applied to more diverse populations until they have been recalibrated. Differential item functioning across education, age, and racial or ethnic groups is untested, so the items may not carry the same meaning among respondents less familiar with LLM data practices, which would affect the factor structure as well as the observed levels of concern.

Second, the questionnaire was validated as a measure of general privacy concerns with LLMs rather than concerns tied to a specific application or use context. While this generality is a strength for cross-study comparability, it means that the instrument's criterion validity, its ability to predict behaviorally relevant outcomes such as disclosure willingness, trust, or system abandonment, remains to be established. Future research should administer the scale alongside behavioral outcome measures and within the context of specific LLM applications, such as AI-assisted clinical decision support or consumer-facing chatbots, to evaluate its predictive utility across different use cases. More specifically, convergent and criterion validity could be established through several complementary approaches. Concurrent validity could be assessed by administering the LLM-PCA alongside established multi-item privacy-concern instruments, such as the CFIP scale, the IUIPC scale, the MUIPC scale, and Hong and Thong's integrated model, within the same sample; moderate-to-strong positive correlations would support convergent validity, whereas weaker associations with conceptually adjacent but distinct constructs, such as general institutional trust, technology acceptance, or privacy self-efficacy, would provide discriminant evidence and help situate the instrument within its nomological network. Criterion validity could then be evaluated against behaviorally relevant outcomes assessed by self-report or, preferably, by observation: the willingness to disclose personal and sensitive (eg, health) information to an LLM, the amount and type of information actually disclosed in logged interactions, the adoption and continued use versus abandonment of an LLM application, and the uptake of privacy-protective options such as opting out of model training, deleting conversation history, or anonymizing inputs. Known-groups comparisons would provide a further line of evidence, with higher concern scores expected among respondents who have previously experienced a privacy breach, who routinely handle sensitive data, or who work in highly regulated professions, such as health care. Because concurrent designs can establish only association, prospective designs in which baseline LLM-PCA scores are used to predict subsequently observed disclosure or adoption behavior would be needed to demonstrate predictive validity, and vignette-based experiments that manipulate contextual factors, for example, the recipient of the data, the purpose of processing, or the presence of consent, would test whether scores respond to theoretically meaningful changes in the information flow.

Third, the study relied on convenience sampling via a single university and its affiliated recruitment platforms. This limits the representativeness of the sample with respect to occupation, geographic region, and general population awareness of LLMs. Future validation studies should target broader, population-based samples.

Fourth, although both subsamples met the a priori benchmarks set out in the Methods section, the split-sample cross-validation design produced an EFA subsample of n=121 and a CFA subsample of n=240. These subsample sizes are modest in absolute terms but adequate for this initial validation: the EFA subsample satisfied the conditions under which small-sample factor recovery is reliable (high communalities and a strongly overdetermined single factor, with KMO=0.87), and the CFA subsample was at or close to the ideal 20:1 ratio of cases to estimated parameters, depending on which specification is counted. The evaluation is post hoc: confirming after the fact that the favorable condition held is weaker evidence than a prospective power analysis would have provided. Nonetheless, the modest absolute sizes may have reduced the precision of factor-loading estimates and the power of fit indices to detect localized misspecification. The 1- versus 2-factor comparison is the analysis most affected: with 3 indicators per factor at this subsample size, it can show that the 2-factor model fits better, but not how the 2 facets behave in relation to external criteria. Replication of the factor structure in independent and larger samples, ideally drawn from distinct subgroups to permit measurement-invariance testing, remains an important next step before the scale is treated as a fully validated instrument.

Fifth, the instrument uses heterogeneous response anchors across items (importance, concern, and frequency) on a common 1 - 5 scale. While this is appropriate for the conceptual content of each item, it assumes that respondents treat position 3 on each anchor type as functionally equivalent. Future work should evaluate whether converting all items to a single anchor type (eg, agreement with statements) yields comparable psychometric properties and potentially simplifies scoring. In addition, composite scores are formed by unit-weighted summation of raw responses without prior standardization, which assumes that the 5 ordinal positions are metrically equivalent across the 3 anchor types. That assumption was not tested here, and the reasons for retaining unit weighting are set out in the Methods section. Future work should compare unit-weighted with standardized scoring in the same sample and report whether substantive conclusions differ.

Sixth, the items were pretested only within the study team and did not undergo cognitive interviewing with lay respondents before fielding. Several items use relatively abstract phrasing and a comparatively high reading level, and the awareness screening item may not have been interpreted uniformly by lay respondents. Because these issues bear directly on content validity, the instrument should be regarded as preliminary, and future work should incorporate cognitive pretesting and readability assessment in community samples. Relatedly, the expert review cannot substitute for the pretesting that was not done. The panel comprised 8 reviewers identified through research team recommendation, which meets the recommended panel size but is small and not independently constituted, and it assessed clarity by inspection rather than by observing how respondents interpreted the items. The obtained clarity index

XSL•FO
RenderX

for G2 fell below the 0.78 item-level criterion because 2 of the 8 reviewers rated below the endorsement threshold despite unanimous endorsement of its relevance. It was retained unchanged, so the wording the expert panel flagged has not been tested on lay respondents. This is a known weakness of the current version.

Seventh, temporal stability of the instrument was not assessed. Because internal consistency cannot establish reliability for single-item domains, test-retest reliability over an appropriate interval is an important form of evidence that remains to be collected in a future study.

Eighth, all data were self-reported, so common method bias cannot be ruled out and associations involving the construct may be inflated relative to multimethod designs [44].

In addition, the items with the weakest loadings (G6 and G7 0.74 and 0.71, respectively) may be candidates for rewording or replacement in future work. Because each domain is represented by a single item, future multi-item per domain development could add parallel indicators for the risk-salience (G6) and onward third-party-sharing (G7) domains, or revise their current wording, to improve their measurement precision. More generally, the one-item-per-domain design carries costs. It yields no within-domain reliability estimate, so no domain score can be corrected for measurement error. It makes each domain maximally sensitive to the wording of a single stem, which is precisely the vulnerability the interpretation literature documents [12,13]. It precludes tests of within-domain measurement invariance across groups. It also limits how well any latent facet can be identified, as the 1- versus 2-factor comparison illustrates. These costs were accepted in exchange for content breadth and brevity, and they are the main reasons the instrument is presented as a brief screening and benchmarking measure rather than as a substitute for a full multidimensional battery.

Conclusions

This study addressed a recognized gap in the LLM privacy research literature by developing and preliminarily validating a concise 6-item instrument for assessing user privacy concerns with LLM-based systems. The scale demonstrated high internal consistency, and a comparison of 1- and 2-factor specifications supported a 2-factor structure comprising a transparency and control facet and a data use and risk salience facet. Because the 2 facets are highly correlated and only partly distinct, the 6-item total score is the recommended primary metric, with facet scores providing secondary detail. Unlike prior instruments in this domain, which were developed within specific theoretical frameworks for narrow research purposes, this LLM-PCA instrument is designed for broad applicability, enabling cross-study comparisons and practical privacy assessments by researchers and AI application developers alike. Future work should prioritize replication in demographically diverse samples, evaluation of criterion validity against behavioral outcomes, and administration within specific LLM application contexts to further establish the instrument's utility and boundary conditions.

Acknowledgments

The authors wish to thank all participants for their time and efforts in answering these survey questions. The authors also thank the editor and the reviewers for their insightful comments, which significantly improved the quality of this manuscript. The authors declare the use of generative AI (GAI) in the research and writing process. The following tasks were delegated to GAI tools under full human supervision: Python code generation for data analysis and data visualization and proofreading and editing on the manuscript drafted by the authors. The GAI tool used was Claude Opus 4.8 (Anthropic). Responsibility for the final manuscript lies entirely with the authors. GAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Funding

The authors declared no financial support was received for this work.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: LZ, BP
Methodology: LZ, YKC, BP
Data curation: HW, YX, IMAS, HH, AS
Investigation: LZ, HW, YX
Data analysis: LZ, YX
Validation: IMAS, HH, AS
Supervision: LZ, BP
Writing—original draft: LZ, BP
Writing—review and editing: LZ, HW, YX, IMAS, HH, AS, YKC, BP

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Tables S1-S5 reporting content validity, nonresponse, correlation, and model-fit results supporting the development and validation of the large language model privacy concern assessment.
[DOCX File, 36 KB - humanfactors_v13i1e102910_app1.docx]

Multimedia Appendix 2
A copy of the complete questionnaire.
[PDF File, 148 KB - humanfactors_v13i1e102910_app2.pdf]

References

1. Choi Y, Zhunis A, Dong W, et al. Privacy disclosure to large language models: a large-scale study on awareness, benefits, and concerns in health contexts across three countries. Comput Hum Behav Rep 2025 Dec;20:100841. [doi: 10.1016/j.chbr.2025.100841]
2. Liu Z, Zou W, Lin C. Exploring the influence of privacy concerns, AI literacy, and perceived health stigma on AI chatbot use in healthcare: an uncertainty reduction approach. Patient Educ Couns 2025 Nov;140:109271. [doi: 10.1016/j.pec.2025.109271] [Medline: 40712403]
3. Liu Z, Hu L, Zhou T, Tang Y, Cai Z. Prevalence overshadows concerns? Understanding Chinese users' privacy awareness and expectations towards LLM-based healthcare consultation. 2025 Presented at: 2025 IEEE Symposium on Security and Privacy (SP); May 12-15, 2025. [doi: 10.1109/SP61157.2025.00092]
4. Tran S, Lu H, Slaughter I, et al. Understanding privacy norms around LLM-based chatbots: a contextual integrity perspective. Presented at: Proceedings of the Eighth AAAI/ACM Conference on AI, Ethics, and Society (AIES-25); Oct 20-22, 2025. [doi: 10.1609/aies.v8i3.36735]
5. Smith HJ, Dinev T, Xu H. Information privacy research: an interdisciplinary review1. MIS Q 2011 Dec 1;35(4):989-1015. [doi: 10.2307/41409970]
6. Smith HJ, Milberg SJ, Burke SJ. Information privacy: measuring individuals' concerns about organizational practices1. MIS Q 1996 Jun 1;20(2):167-196. [doi: 10.2307/249477]
7. Malhotra NK, Kim SS, Agarwal J. Internet Users' Information Privacy Concerns (IUIPC): the construct, the scale, and a causal model. Inf Syst Res 2004 Dec;15(4):336-355. [doi: 10.1287/isre.1040.0032]
8. Xu H, Gupta S, Rosson MB, Carroll JM. Measuring mobile users' concerns for information privacy. Presented at: 33rd International Conference on Information Systems (ICIS 2012); Dec 16-19, 2012 URL: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1108&context=icis2012 [accessed 2026-10-11]
9. Hong W, Thong JYL. Internet privacy concerns: an integrated conceptualization and four empirical studies1. MIS Q 2013 Mar 1;37(1):275-298. [doi: 10.25300/MISQ/2013/37.1.12]
10. Sheehan KB, Hoy MG. Dimensions of privacy concern among online consumers. J Public Policy Mark 2000 Apr;19(1):62-73. [doi: 10.1509/jppm.19.1.62.16949]
11. Wang Q, Zhang W, Wang H. Privacy concerns toward short-form video platforms: scale development and validation. Front Psychol 2022;13:954964. [doi: 10.3389/fpsyg.2022.954964]
12. Colnago J, Cranor LF, Acquisti A, Stanton KH. Is it a concern or a preference? An investigation into the ability of privacy scales to capture and distinguish granular privacy constructs. Presented at: Proceedings of the Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022); Aug 8-9, 2022 URL: https://www.usenix.org/system/files/soups2022-colnago.pdf [accessed 2026-10-11]
13. Farzand H, Abdelwahab Gaballah S, Macdonald S, Khamis M, Marky K. Revisiting privacy scales: an investigation into the ability of privacy scales to capture and distinguish granular privacy constructs. Inf Comput Secur 2025;34(4):521-545. [doi: 10.1108/ICS-11-2024-0289]
14. Groß T. Validity and reliability of the scale Internet Users' Information Privacy Concerns (IUIPC). Proc Priv Enh Technol 2021 Apr 1;2021(2):235-258. [doi: 10.2478/popets-2021-0026]
15. Stewart KA, Segars AH. An empirical examination of the concern for information privacy instrument. Inf Syst Res 2002 Mar;13(1):36-49. [doi: 10.1287/isre.13.1.36.97]
16. Zhang Z, Jia M, Lee H, et al. "It's a fair game", or is it? Examining how users navigate disclosure risks and benefits when using LLM-based conversational agents. Presented at: CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems; May 11-16, 2024. [doi: 10.1145/3613904.3642385]
17. Carlini N, Tramer F, Wallace E, et al. Extracting training data from large language models. Presented at: 30th USENIX Security Symposium (USENIX Security 21); Aug 11-13, 2021 URL: https://www.usenix.org/system/files/sec21-carlini-extracting.pdf [accessed 2026-10-11]

XSL•FO
RenderX

18. Staab R, Vero M, Balunović M, Vechev M. Beyond memorization: violating privacy via inference with large language models. Presented at: Proceedings of ICLR 2024 (The Twelfth International Conference on Learning Representations); May 7-11, 2024 URL: https://proceedings.iclr.cc/paper_files/paper/2024/file/9028b8a3ca98f58e373f0c1497a17448-Paper-Conference.pdf [accessed 2026-10-11]
19. Berger CR, Calabrese RJ. Some explorations in initial interaction and beyond: toward a developmental theory of interpersonal communication. Human Comm Res 1975 Jan;1(2):99-112. [doi: 10.1111/j.1468-2958.1975.tb00258.x]
20. Nissenbaum H. Privacy as contextual integrity. Wash L Rev 2004;79(1):119-157 [FREE Full text]
21. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Front Public Health 2018;6:149. [doi: 10.3389/fpubh.2018.00149]
22. Donaldson T, Dunfee TW. Toward a unified conception of business ethics: integrative social contracts theory. Acad Manag Rev 1994 Apr;19(2):252. [doi: 10.2307/258705]
23. Diamantopoulos A, Sarstedt M, Fuchs C, Wilczynski P, Kaiser S. Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. J Acad Mark Sci 2012 May;40(3):434-449. [doi: 10.1007/s11747-011-0300-3]
24. Bergkvist L, Rossiter JR. The predictive validity of multiple-item versus single-item measures of the same constructs. J Mark Res 2007 May;44(2):175-184. [doi: 10.1509/jmkr.44.2.175]
25. Wanous JP, Reichers AE, Hudy MJ. Overall job satisfaction: how good are single-item measures? J Appl Psychol 1997 Apr;82(2):247-252. [doi: 10.1037/0021-9010.82.2.247] [Medline: 9109282]
26. Jarvis CB, MacKenzie SB, Podsakoff PM. A critical review of construct indicators and measurement model misspecification in marketing and consumer research. J Consum Res 2003 Sep;30(2):199-218. [doi: 10.1086/376806]
27. Bobko P, Roth PL, Buster MA. The usefulness of unit weights in creating composite scores: a literature review, application to content validity, and meta-analysis. Organ Res Methods 2007;10(4):689-709. [doi: 10.1177/1094428106294734]
28. Morgado FFR, Meireles JFF, Neves CM, Amaral ACS, Ferreira MEC. Scale development: ten main limitations and recommendations to improve future research practices. Psicol Refl Crít 2018 Jan;30(1):3. [doi: 10.1186/s41155-016-0057-1]
29. Lynn MR. Determination and quantification of content validity. Nurs Res 1986;35(6):382-385. [Medline: 3640358]
30. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. Res Nurs Health 2006 Oct;29(5):489-497. [doi: 10.1002/nur.20147] [Medline: 16977646]
31. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health 2007 Aug;30(4):459-467. [doi: 10.1002/nur.20199] [Medline: 17654487]
32. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
33. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. Res Sci Educ 2018 Dec;48:1273-1296. [doi: 10.1007/s11165-016-9602-2]
34. Nunnally JC, Bernstein IH. Psychometric Theory, 3rd edition: McGraw Hill; 1994.
35. HORN JL. A rationale and test for the number of factors in factor analysis. Psychometrika 1965 Jun;30(2):179-185. [doi: 10.1007/BF02289447] [Medline: 14306381]
36. Xia Y, Yang Y. RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: the story they tell depends on the estimation methods. Behav Res Methods 2019 Feb;51(1):409-428. [doi: 10.3758/s13428-018-1055-2] [Medline: 29869222]
37. Kline RB. Principles and Practice of Structural Equation Modeling: Guilford Press; 2023:494.
38. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 1981 Feb;18(1):39-50. [doi: 10.1177/002224378101800104]
39. Igolkina AA, Meshcheryakov G. semopy: a Python package for structural equation modeling. Struct Equ Model 2020 Nov 1;27(6):952-963. [doi: 10.1080/10705511.2019.1704289]
40. MacCallum RC, Widaman KF, Zhang S, Hong S. Sample size in factor analysis. Psychol Methods 1999;4(1):84-99. [doi: 10.1037/1082-989X.4.1.84]
41. de Winter JCF, Dodou D, Wieringa PA. Exploratory factor analysis with small sample sizes. Multivariate Behav Res 2009;44(2):147-181. [doi: 10.1080/00273170902794206] [Medline: 26754265]
42. Costello AB, Osborne JW. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Pract Assess Res Eval 2005;10(7):1-9. [doi: 10.7275/jyj1-4868]
43. Wolf EJ, Harrington KM, Clark SL, Miller MW. Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. Educ Psychol Meas 2013 Dec;76(6):913-934. [doi: 10.1177/0013164413495237] [Medline: 25705052]
44. Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]

XSL•FO
RenderX

Abbreviations

AGFI: adjusted goodness-of-fit index
APCO: antecedents, privacy concerns, and outcomes
AVE: average variance extracted
CFA: confirmatory factor analysis
CFI: comparative fit index
CFIP: Concern for Information Privacy
CVI: content validity index
DWLS: diagonally weighted least squares
EFA: exploratory factor analysis
GFI: goodness-of-fit index
HTMT: heterotrait-monotrait ratio of correlations
I-CVI: item-level content validity index
IPC: Internet Privacy Concerns
IUIPC: Internet Users' Information Privacy Concerns
KMO: Kaiser-Meyer-Olkin
LLM: large language model
LLM-PCA: large language model privacy concern assessment
ML: maximum likelihood
MUIPC: Mobile Users' Information Privacy Concerns
PCLOSE: *P* value for the test of close fit
RMSEA: root mean square error of approximation
S-CVI: scale-level content validity index
SRMR: standardized root mean square residual
TLI: Tucker-Lewis index
UA: universal-agreement

Edited by S Law; submitted 29.May.2026; peer-reviewed by FNK Şahin, H Mahmoodi, P Lialiou, T Groß; revised version received 12.Sep.2026; accepted 16.Sep.2026; published 21.Oct.2026.

Please cite as:
Zhou L, Wang H, Xu Y, Setiawan IMA, Hu H, Saptono A, Choi YK, Parmanto B
Development and Validation of the Large Language Model Privacy Concern Assessment Questionnaire: Psychometric Evaluation Study on a Brief Instrument for Assessing User Privacy Concerns With Large Language Model Applications
JMIR Hum Factors 2026;13:e102910
URL: https://humanfactors.jmir.org/2026/1/e102910
doi:10.2196/102910

© Leming Zhou, Hongtao Wang, Yi Xu, I Made Agus Setiawan, Haomin Hu, Andi Saptono, Yong K Choi, Bambang Parmanto. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 21.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mode Effects on Self-Reported Health Across Web-Based, Interviewer-Administered, and Nurse-Administered Modes: Population-Based Study

Yunlong Liang, PhD; Tarek Al Baghal, PhD; Jonathan Burton, PhD; Meena Kumari, PhD

Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, United Kingdom

Corresponding Author:
Meena Kumari, PhD
Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, United Kingdom

Abstract

Background: Probability-based surveys remain essential to epidemiology, but expanding mixed-mode data collection, which combines web self-completion and interviewer-administered surveys, raises concerns that administration mode and interviewer presence may influence health reporting. These concerns warrant further investigation of interviewer-related biases and mode differences in health reporting.

Objective: This study assessed whether responses to the 12-item General Health Questionnaire (GHQ-12) and 12-item Short Form Health Survey (SF-12), both self-completed, differed by interviewer presence and role: online without an interviewer, during a social interviewer visit, or during a nurse visit. It also assessed mode effects on doctor-diagnosed conditions recorded by interviewers in computer-assisted personal interviewing (CAPI) and when self-completed online. We further examined possible variation by sex and age.

Methods: The analysis used randomized mode allocation embedded within wave 12 (2019) of the UK Household Longitudinal Study Innovation Panel. After applying relevant exclusions, the final analytic sample comprised 1111 households (1779 participants) across the 3 randomly assigned survey modes: web-based self-completion, interviewer-administered CAPI, or nurse-administered CAPI. The sample was predominantly White British, with a broadly balanced sex distribution and participants from diverse age, employment, education, and household income groups. Participants did not always complete the survey in their randomly assigned mode. Effects were estimated using a two-stage residual inclusion instrumental variable approach, with randomized mode as an instrument for actual mode, to address confounding arising from nonrandom selection into actual mode.

Results: No statistically significant differences were observed across self-completion contexts for the GHQ-12 or the SF-12 mental and physical component scores. We found no statistically significant evidence of mode effects on doctor-diagnosed conditions. For example, no statistically significant mode effects in psychiatric conditions were identified across the comparisons: nurse relative to social interviewer (odds ratio [OR] 0.72, 95% CI 0.40 - 1.29; *P*=.53), web relative to social interviewer (OR 0.45, 95% CI 0.17 - 1.18; *P*=.14), and web relative to nurse (OR 0.63, 95% CI 0.27 - 1.49; *P*=.59). There was little evidence overall that mode-related differences varied by age or sex.

Conclusions: Our findings may support embedding self-completion modules within interviewer-administered surveys for the collection of SF-12 and GHQ-12. Meanwhile, potential mode-related differences should be considered when designing mixed-mode surveys and interpreting self-reported health data.

(JMIR Hum Factors 2026;13:e89896) doi:10.2196/89896

KEYWORDS

mode effects; surveys and questionnaires; computer-assisted personal interviewing; web-based surveys; social desirability bias; interviewer effects; mental health; population-based study

Introduction

Probability-based surveys have been a cornerstone of epidemiology and public health, providing reliable population-based estimates that enable the monitoring of population health trends to inform evidence-based policy [1]. Even amid the rise of alternative data sources, they remain essential to ensuring representativeness [2,3] and to capturing the social, psychological, behavioral, and biological determinants that shape health. Increasing emphasis on cost-efficiency and biomarker integration is accelerating the diversification of health survey administration. These modes include face-to-face interviews conducted by trained social interviewers, web-based self-administered surveys [4], and, where health is a focus, interviews conducted by nurses who can undertake standardized biospecimen collection and physical

XSL•FO
RenderX

measurements [5]. The use of these different modes allows surveys to accommodate varying data-collection requirements.

However, the use of different data-collection modes within surveys may affect responses and introduce bias in health estimates [6]. Mode effects refer to systematic differences in survey responses that arise from the mode of data collection and its associated conditions [7]. Different mechanisms may underlie these effects. The presence of an interviewer may increase socially desirable responding, whereby respondents report in ways they perceive to be more socially acceptable [8], while interviewer probing, clarification, and feedback may also affect how respondents understand and report health conditions [9,10]. The former mechanism may be particularly relevant to sensitive or subjective scales. For example, previous studies with random assignment to survey mode found that interviewer-administered surveys yielded more favorable health ratings than self-administered surveys [11-15]. To reduce mode effects arising from social desirability bias, several national surveys embed self-completion within face-to-face computer-assisted personal interviewing (CAPI) for sensitive items [16-18]. However, the context in which self-completion responses are provided may differ across modes. Interviewer presence may shape reporting even when respondents complete items themselves by influencing perceived privacy, thereby contributing to differences between self-completion embedded within interviewer-present CAPI and web self-completion [10]. For example, one study found that respondents completing the 12-Item General Health Questionnaire (GHQ-12) in the presence of an interviewer reported better mental health than those completing it without an interviewer present [19]. Similarly, among self-administered modes, less interaction with interviewers was associated with higher reporting of common mental disorder symptoms [20]. However, these findings may reflect between-sample differences or evidence from specific populations, limiting their generalizability.

For the reporting of specific chronic conditions, however, studies have primarily compared interviewer-administered with self-administered reporting. Any observed differences in reporting may therefore reflect not only social desirability bias but also interviewer-specific features, such as probing and clarification. Several studies found higher prevalence estimates in self-administered surveys than in interviewer-administered surveys for a mental health problem [21], longstanding limiting illness, and current bodily pain lasting more than 6 months [22], possibly reflecting social desirability bias. One randomized study also found that most health problems were reported more frequently in a self-administered postal survey than in an interviewer-administered survey [23]. However, existing studies have been restricted to older adults or young people [21,23], with one of these studies additionally constrained by a small sample size [21], whereas one study was unable to separate mode differences from temporal differences [22]. Consequently, the extent to which survey mode independently affects the reporting of different individual health conditions in the broader adult population remains unclear. In contrast, interviewer probing and clarification may facilitate reporting of chronic conditions. For example, previous reports of higher prevalence of longstanding illness [9], bronchial asthma, and chronic bronchitis [24] in interviewer-administered surveys may reflect interviewer probing and clarification. However, other studies reported no mode differences in the number of chronic conditions, the prevalence of longstanding illness [19,25], or the prevalence of bronchial asthma and chronic bronchitis [9]. Differences in findings across studies may also reflect population selectivity, between-sample differences, or residual selection effects. Thus, evidence supporting interviewer facilitation as a mechanism underlying mode effects remains heterogeneous.

In addition, evidence suggests that interviewers' professional roles can influence self-reported health [10], although the direction of these effects is inconsistent. Respondents interviewed by physicians or nurses have been reported to disclose fewer psychiatric symptoms [26] and fewer health problems than those interviewed by lay interviewers or completing self-administered questionnaires [23]. Conversely, higher pain intensity has been reported in some nurse-administered interviews [27]. However, these findings derive largely from specific subgroups or highly selected populations, limiting their generalizability. Meanwhile, whether interviewer effects associated with the professional role of nurses spill over into the self-administered modules embedded within CAPI surveys remains unclear. Furthermore, sex and age may modify the effects of survey mode on health reporting, but these differences have received limited attention. Sex-specific patterns may partly reflect variation in concerns about mental health stigma and disclosure. The greater perceived privacy of self-administered web surveys may facilitate the reporting of sensitive health information, particularly among men [21,28]. Age differences may reflect older adults' stronger emphasis on socioemotional goals and greater reluctance to disclose negative states in interpersonal settings. Compared with less interpersonal modes, surveys involving greater interviewer contact have elicited fewer reports of adverse psychosocial functioning among older adults, with this difference becoming more pronounced with increasing age [29].

Therefore, this study aimed to examine whether health measures completed by respondents themselves are reported differently across contexts that differ in interviewer presence and roles, and whether the reporting of doctor-diagnosed chronic conditions differs between web self-completion, social interviewer-administered CAPI, and nurse-administered CAPI. We then examined whether these associations varied by sex and age groups in secondary analyses. Specifically, we asked the following questions:

- Although the GHQ-12 and SF-12 were self-completed in all 3 conditions, do participants report mental health and functioning differently across self-completion contexts involving no interviewer, a social interviewer, or a nurse?
- Does the reporting of diagnosed health conditions differ between web self-completion, social interviewer–administered CAPI, and nurse-administered CAPI?
- Do these associations differ by sex and age group?

XSL•FO
RenderX

Methods

Data

The UK Household Longitudinal Study Innovation Panel (UKHLS-IP) is a national probability sample that covers the household population of Great Britain, specifically designed to support methodological development and experimental research in longitudinal survey design [30]. The UKHLS-IP began in 2008, with a sampling design that used a stratified, geographically clustered probability approach yielding 2760 addresses [31], and was refreshed at waves 4 (2011), 7 (2014), 10 (2017), and 11 (2018) [32]

The Innovation Panel wave 12 (IP12) issued sample comprised 2401 households: 1957 households participated as recontacts from IP11, and 444 returning households from earlier waves. Before fieldwork, these 2401 households were randomly assigned to one of three modes: (1) web-based self-completion, (2) face-to-face CAPI by social interviewers, or (3) nurse-led in-home CAPI [33]. The sample flow is described in the *Results* section.

Outcomes

The following health outcomes were chosen. Mental health was assessed using the GHQ-12 and SF-12 Mental Component Summary (MCS) scores. In all modes, respondents self-completed the scales, either as computer-assisted self-interviewing (CASI) modules within social-interviewer or nurse CAPI visits, or independently online. During the CAPI interview, the script instructed the social interviewer or nurse to pass the laptop to the respondent at the CASI section, which the respondent then completed independently. The interviewer or nurse generally remained in the room but attended to their own tasks and did not administer the questions or interact with the respondent during CASI completion, except when assistance was needed because of sight, reading, or language difficulties.

Each item in the GHQ-12 was rated on a 4-point Likert scale, yielding a total GHQ-12 score ranging from 0 (“no psychological distress”) to 36 (“severe psychological distress”), with higher scores indicating greater levels of distress [34]. The Physical Component Summary (PCS) score from the SF-12 was used to assess physical health. SF-12 MCS and SF-12 PCS scores were computed using the norm-based scoring algorithm [35]. Both scores were derived by the UKHLS team. We reverse-coded the scores such that higher values represented poorer physical and mental functioning, which allowed for effective interpretation and comparability.

In addition, we derived 6 binary indicators to capture major categories of chronic conditions, based on self-reported doctor diagnoses at IP12. Respondents were shown a list of specific conditions, on a showcard in the face-to-face modes and on screen in the web mode, and asked whether a doctor or other health professional had ever told them that they had any of these conditions. The measures, therefore, reflected doctor-diagnosed conditions reported up to the IP12 interview. These questions were ordinarily asked at a respondent’s first full individual interview. At IP12, however, both new entrants to the survey and continuing respondents were asked these questions. To improve statistical reliability and mitigate data sparsity, chronic conditions were grouped into six categories: (1) cancer, defined as any malignant neoplasm; (2) cardiovascular conditions, including congestive heart failure, coronary heart disease, angina, myocardial infarction, stroke, or high blood pressure; (3) metabolic–endocrine disorders, comprising diabetes or hypothyroidism; (4) musculoskeletal disorders, including arthritis; (5) psychiatric conditions, indicating whether respondents reported any doctor-diagnosed psychiatric disorder; and (6) respiratory conditions, including asthma, emphysema, chronic bronchitis, or chronic obstructive pulmonary disease. For each category, a value of 1 was assigned if any relevant condition was reported, and zero otherwise. The specific component conditions included in each major category, along with the number of participants reporting each condition, were described in Table S1 in Multimedia Appendix 1.

Survey Mode

The primary exposure was the actual survey mode completed by each respondent, categorized as web-based self-completion, interviewer-administered, or nurse-administered interviews. For the GHQ-12 and SF-12, all items were self-completed by respondents, including within the CASI module of CAPI interviews; for these outcomes, the exposure therefore represents the context of self-completion, namely online without an interviewer present, during a social interviewer visit, or during a nurse visit, rather than interviewer administration of the items. For doctor-diagnosed chronic conditions, the exposure reflects both survey mode and item administration, as conditions were self-reported in the web questionnaire but were elicited using a condition list in the CAPI interviews.

Covariates

Based on prior studies, we adjusted for a comprehensive set of contextual, sociodemographic, and survey-related variables to account for potential confounding [36-38]. Sex was coded as male or female. Gender identity and intersex variations were not available in the analytic data. Age was categorized into 6 groups: 16 - 25, 26 - 35, 36 - 45, 46 - 55, 56 - 65, and ≥66 years. However, because the prevalence of chronic conditions was very low at younger ages and increased sharply from the mid-50s onward, we used a binary age indicator (<56 vs ≥56 y) in models with chronic disease as the outcome to avoid sparse-data bias and to ensure sufficient events within each category for stable estimation. Ethnicity was classified as White British and other ethnicities. Marital status was grouped into married, never married, and separated/divorced/widowed. Employment status was categorized as employed (including self-employed, unpaid family workers, and apprentices), unemployed and inactive (including students, individuals unable to work due to health conditions/disabilities, and family caregivers), and retired [39]. Education was grouped into degree or equivalent, General Certificate of Secondary Education (GCSE) and A-levels or equivalent, and no qualifications or other. Monthly household income was divided into quintiles: (0 - 1969.13], (1969.13 - 2956.40], (2956.40 - 4083.32], (4083.32 - 5756.04], and (5756.04 - 20,000]. Household child status was based on the number of children, coded as none and ≥1 child.

XSL•FO
RenderX

The first participation wave was indicated by the wave of entry into the study: original sample in 2008, and refreshment samples in 2011, 2014, 2017, and 2018. Interview mode in IP11 was classified into 3 categories: interviewer-administered, web self-completion, and absence or proxy. Due to the small number of proxy respondents (n=12), they were combined with the absence of respondents.

Geographic regions included England, Wales, and Scotland. Urban-rural classification was based on population density, categorized as urban or rural areas.

Statistical Analysis

For descriptive analyses, we summarized sample characteristics using survey-design–weighted means with 95% CIs for continuous variables and weighted proportions with 95% CIs for categorical variables, based on the imputed dataset. All summaries were stratified by survey mode.

For the main analyses, we used a two-stage residual inclusion (2SRI) approach within the instrumental variable framework, with the initial random assignment used as the instrument for the survey mode actually completed [40-42]. To be valid, instrumental variables must satisfy 4 identification assumptions: relevance, independence, exclusion restriction, and monotonicity (Table 1; more details can be found in Multimedia Appendix 1). We used this approach because some respondents did not complete the survey in their assigned mode; consequently, the mode actually completed was not entirely random and may have been subject to selection. Additionally, analyzing by initial assignment preserves randomization but can be diluted by mode switching, while analyses based on the completed mode could be confounded by differences between those who did and did not switch modes. The instrumental variable approach may reduce these concerns by using the difference in completed mode that was created by randomization. Because that difference originates from randomization rather than from participants' own choices, it provides a basis for estimating the effect of completing that mode. Noncompliance and crossover are therefore incorporated through the extent to which assignment actually changed mode completion.

Table . Summary of instrumental variable assumptions and assessment.

Assumption	Assumption description	Assumption assessment	Conclusion and handling
Instrument relevance	Random assignment should be strongly associated with the survey mode actually completed.	First-stage models showed strong associations between assigned mode and completed mode. Detailed estimates and tests are reported in the Multimedia Appendix 1.	This assumption was satisfied.
Conditional independence	Random assignment should be independent of factors that may affect the outcomes, after accounting for measured covariates.	Assignment was randomized before fieldwork. Moderate covariate imbalances were observed (Table S2 in Multimedia Appendix 1), but the asymmetric switching rules operated only postrandomization and therefore do not undermine the initial assignment mechanism.	We used a conditional version of this assumption and adjusted for all prespecified covariates in both stages of the model.
Exclusion restriction	Random assignment should affect outcomes only through the survey mode actually completed.	Random assignment occurred before respondents had any survey exposure, making a direct effect of assignment on self-reported health through pathways other than the exposure unlikely.	This assumption was considered plausible, although it cannot be directly tested. Covariate adjustment was used to reduce potential alternative pathways.
Monotonicity	Assignment to a given mode should not reduce the probability of completing the interview in that mode.	Observed asymmetric switching does not, by itself, violate monotonicity. Defier status depends on treatment uptake under both assignment states and therefore cannot be inferred from a one-sided switch. Moreover, nurse-to-web switching is not evidence of defiance when nurse administration was unavailable under the comparison assignments.	This assumption was considered plausible but is not directly testable at the individual level.

In the first stage, we fitted a multinomial logistic regression model to predict respondents' actual survey completion modes, using randomized mode assignment as the instrumental variable. The first-stage model estimated an allocation arm-specific vector of probabilities for the three actual completion modes. The model adjusted for a set of covariates, including demographic characteristics, household composition, socioeconomic position, first participation wave, interview mode in IP11, and geographic indicators. By design, nurse completion did not occur among participants allocated to the web or social interviewer arms, and social interviewer completion did not occur in the nurse arm. Consistent with these switching restrictions, the fitted probabilities for mode transitions that did not occur were effectively zero. For each respondent, we then calculated

XSL•FO
RenderX

mode-specific residuals as the difference between indicators of actual completion mode and the corresponding predicted probabilities from the first-stage model.

In the second stage, the residuals for the modes were included alongside the observed survey mode and covariates in regression models estimating the associations with key outcomes. These outcomes included the total score of GHQ-12, the PCS and MCS scores from the SF-12, and a set of binary indicators for health conditions (cancer, cardiovascular conditions, metabolic-endocrine conditions, musculoskeletal disorders, psychiatric conditions, and respiratory conditions). We fitted outcome‐specific regression models: Gaussian linear models for continuous outcomes (GHQ-12 score, SF-12 PCS, and MCS scores) and logistic regression models for each binary chronic condition. All previously described covariates were included in both stages of the 2SRI analysis [43]. To avoid overstating clinical relevance, we a priori adopted the lower bound of these ranges (PCS=1.8; MCS=1.5) as our threshold for clinical importance [44]. For the GHQ-12, where no consensus threshold exists, we adopted 0.58 points on the Likert-scored scale, based on prior work, as a conservative benchmark to distinguish trivial from nontrivial differences at the group level [45]. For the GHQ-12 and SF-12 outcomes, we additionally reported standardized effect sizes. For binary outcomes, we prespecified a minimum clinically important risk difference of 5 percentage points for each chronic condition.

To address missingness, we applied multiple imputation using a random forest approach. Compared to parametric approaches, random forests are capable of capturing complex nonlinear relationships and higher-order interactions, and exhibit greater robustness to outliers and model misspecification [46]. The imputation model included outcomes and all covariates used in both the main analysis and the instrumental variable model, ensuring congeniality [47]. In addition, given the complex multistage survey design, the sampling weights, primary sampling units and stratification variables were incorporated into the imputation to preserve the underlying survey structure and mitigate potential bias arising from neglecting design features [48]. In addition, we performed imputation separately within strata of the exposure, defined by the observed survey mode, allowing the covariate and outcome distributions and their associations to differ flexibly between exposure groups and reducing bias when exposure-confounder interactions are present [49]. A total of 50 iterations per imputation were run to facilitate convergence. Given that approximately 15% of the data were incomplete, we generated 15 imputed datasets (Table S3 in Multimedia Appendix 1).

In each dataset, a 2SRI model was independently estimated to quantify exposure effects. To account for the complex survey design, we specified a clustered and stratified design and incorporated survey weights. Design-based standard errors were used in all analyses. The weight compensates for unequal selection probabilities introduced at each sampling stage and for nonresponse in IP12, ensuring that the weighted respondent sample remains representative of the UK adult population living in households at the time of fieldwork [32]. The final estimates and SEs were pooled using Rubin's rules to obtain overall effect estimates and CIs.

All statistical analyses were performed in RStudio. Two‐sided P values <.05 with a Bonferroni correction applied were considered statistically significant. Pairwise contrasts for the primary exposure (survey modes) and all interaction tests were adjusted for multiple comparisons using the Bonferroni method. The corresponding 95% CIs were Bonferroni-adjusted.

Interaction analyses were performed using a similar 2SRI approach in the imputed data accounting for complex survey design. The first-stage multinomial model for mode assignment was estimated in the full sample, after which residuals were included in second-stage models with survey-mode-by-sex and survey-mode-by-age-group interaction terms, respectively. To account for multiple comparisons, Bonferroni correction was applied across the subgroup-specific pairwise contrasts for each outcome.

As a supplementary analysis, we examined the total number of chronic-condition categories reported by each participant, ranging from 0 to 6. Because this outcome was a count variable, we used Poisson regression to compare the number of chronic-condition categories across survey modes, otherwise following the same modeling approach as the main chronic-condition analyses. Results are presented as incidence rate ratios (IRRs) with 95% CIs. As a sensitivity analysis, we repeated the main analyses after excluding participants whose completed survey mode differed from their originally allocated mode.

Additionally, we conducted a post hoc power analysis to assess the ability of the mode-specific sample sizes to detect a benchmark mode effect reported in previous literature for long-standing illness [9].

Ethical Considerations

The UKHLS received ethical approval from the University of Essex Ethics Committee. For waves prior to IP12, ethical approval for the IP was granted as part of the approvals for the UKHLS main survey data collection. From IP12 onward, separate ethical approvals have been obtained for each wave of the IP. IP12 received approval from the National Health Service East of England-Essex Research Ethics Committee (reference 19/EE/0146). Additional ethical approval was not required for the secondary data analysis conducted in this study.

Results

Following the fieldwork process, households were excluded due to noncontact, refusal, or other reasons, resulting in 1408 responding households: 442, 495, and 471 households in the social interviewer, web, and nurse groups, respectively. Further individual-level exclusions were then applied for refusals, proxy or incomplete interviews, incapacity, and children/adolescents. A small number of telephone-mode respondents were excluded from the analysis to avoid potential auditory-mode effects. In addition, respondents with a sampling weight of zero were excluded because such cases do not contribute to population estimates under this survey design [50]. The final analytic sample included 438 households (652 individuals) in the social interviewer group, 330 households (587 individuals) in the web group, and 343 households (540 individuals) in the nurse group,

XSL•FO
RenderX

resulting in a sample of 1111 households and 1779 individuals. The sample flow diagram is provided in Figure 1.

Figure 1. Flow diagram for the study sample.

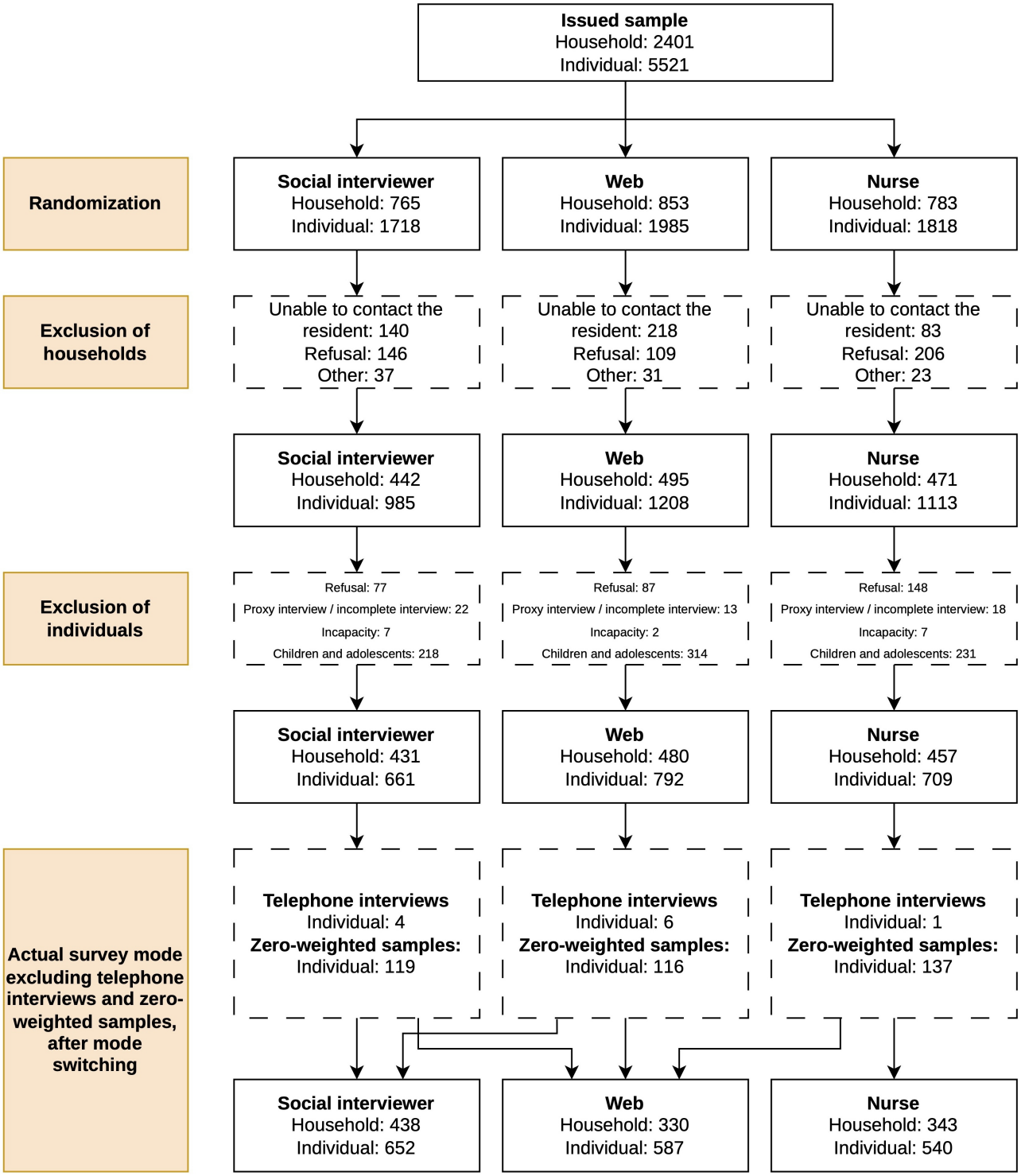


Table S4 in Multimedia Appendix 1 summarizes sample characteristics from weighted imputed datasets by actual survey mode (n=1779). In the imputed, survey-weighted sample, respondents in the web-based self-completion group were physically healthier (SF-12 PCS score 22.12, 95% CI 21.32 - 22.91) than those in either the interviewer-administered (SF-12 PCS score 26.54, 95% CI 25.44 - 27.64) or nurse-administered (SF-12 PCS score 24.71, 95% CI 23.50 - 25.91) CAPI groups. Web respondents also had a lower prevalence of cardiovascular conditions (unweighted 115/585; imputed, weighted 16.94%, 95% CI 13.21 - 20.68) and respiratory conditions (63/585, 11.46%, 95% CI 7.83 - 15.10) than did respondents in the social interviewer group (cardiovascular conditions: 187/651; 26.19%, 95% CI 22.41 - 29.96; respiratory conditions: 124/651; 19.47%, 95% CI 15.67 - 23.27), and fewer musculoskeletal disorders (74/585; 11.20%, 95% CI 8.37 - 14.04) than nurse-administered respondents (102/540; 17.77%, 95% CI 14.16 - 21.38).

XSL•FO
RenderX

Web respondents were younger, with more participants aged 16 to 25 years than in the social-interviewer group, and fewer aged 66 years or older. Compared with social-interviewer respondents, those in the web mode were generally more advantaged as they were less often separated, divorced, or widowed, less likely to have no formal qualifications, more likely to be employed, and less likely to be in the lowest income category. In addition, web respondents were significantly less likely to be separated, divorced, or widowed compared with respondents in the nurse mode. Web-mode respondents were less likely to have previously used interviewer-administered modes than respondents in the nurse mode or social-interviewer respondents, and were more likely to have previously used web self-completion.

Table 2 presents mode effect estimates for GHQ-12, the SF-12 MCS, and the SF-12 PCS, derived from a 2SRI approach that incorporates complex survey design features and multiple imputation for missingness. Across all mode comparisons, none of the estimated differences reached statistical significance. For GHQ-12, estimated differences ranged from −0.27 to 0.25 (effect sizes, −0.05 to 0.04; all *P*>.99); for SF-12 MCS, from −0.36 to 0.32 (effect sizes, −0.04 to 0.03; all *P*>.99); and for SF-12 PCS, from −1.02 to −0.36 (effect sizes, −0.10 to −0.03; *P*=.81 to >.99). Full pairwise estimates and 95% CIs are reported in Table 2. The findings of the complete case analyses were similar (Table S5 in Multimedia Appendix 1).

Table . Summary of results from the main analyses for 12-item General Health Questionnaire (GHQ-12), 12-item Short Form Mental Component Summary (SF-12 MCS), and 12-item Short Form Physical Component Summary (SF-12 PCS).

Outcomes	Estimate (95% CI)	Effect size	*P* value
GHQ-12			
Nurse (reference: social interviewer)	–0.27 (–1.38 to 0.84)	–0.05	>.99
Web (reference: nurse)	0.25 (–1.23 to 1.73)	0.04	>.99
Web (reference: social interviewer)	–0.02 (–1.97 to 1.94)	–0.01	>.99
SF-12 MCS			
Nurse (reference: social interviewer)	–0.36 (–2.14 to 1.42)	–0.04	>.99
Web (reference: nurse)	0.32 (–1.94 to 2.58)	0.03	>.99
Web (reference: social interviewer)	–0.04 (–2.99 to 2.91)	–0.01	>.99
SF-12 PCS			
Nurse (reference: social interviewer)	–0.36 (–2.25 to 1.53)	–0.03	>.99
Web (reference: nurse)	–0.66 (–2.20 to 0.88)	–0.06	.91
Web (reference: social interviewer)	–1.02 (–3.23 to 1.20)	–0.10	.81

Table 3 presents adjusted odds ratios (ORs; 95% CIs) from 2SRI logistic models that incorporate complex survey design and multiple imputation. For nurse vs social interviewer administration, ORs ranged from 0.72 to 1.56 across condition categories, with no consistent direction. In contrast, ORs for web vs nurse administration were consistently below 1, ranging from 0.54 to 0.90, corresponding to approximately 10% to 46% lower odds of reporting a doctor-diagnosed condition. Similarly, for web versus social interviewer administration, ORs were below 1 for 5 of the 6 condition categories (range 0.41 - 0.86), corresponding to approximately 14% to 59% lower odds. Musculoskeletal disorders were the only exception to this pattern for web vs social interviewer administration (OR 1.15, 95% CI 0.52 - 2.57). However, none of these pairwise comparisons reached statistical significance, and all 95% CIs included 1. The findings of the complete case analyses were similar (Table S6 in Multimedia Appendix 1), except for respiratory conditions when comparing respondents in the social interviewer (*P*=.005) mode with respondents in the web self-completion mode.

XSL•FO
RenderX

Table . Summary of results from the main analysis for cancer, cardiovascular conditions, metabolic-endocrine conditions, musculoskeletal disorders, psychiatric conditions, and respiratory conditions.

Outcomes	Odds ratio (95% CI)	*P* value
Cancer		
Nurse (reference: social interviewer)	1.23 (0.62-2.43)	>.99
Web (reference: nurse)	0.62 (0.24-1.63)	.72
Web (reference: social interviewer)	0.77 (0.25-2.35)	>.99
Cardiovascular conditions		
Nurse (reference: Social interviewer)	1.06 (0.65-1.71)	>.99
Web (reference: nurse)	0.64 (0.34-1.19)	.26
Web (reference: social interviewer)	0.68 (0.35-1.31)	.47
Metabolic-endocrine conditions		
Nurse (reference: social interviewer)	0.95 (0.52-1.75)	>.99
Web (reference: nurse)	0.90 (0.45-1.80)	>.99
Web (reference: social interviewer)	0.86 (0.36-2.04)	>.99
Musculoskeletal disorders		
Nurse (reference: social interviewer)	1.56 (0.89-2.72)	.17
Web (reference: nurse)	0.74 (0.43-1.27)	.54
Web (reference: social interviewer)	1.15 (0.52-2.57)	>.99
Psychiatric conditions		
Nurse (reference: social interviewer)	0.72 (0.40-1.29)	.53
Web (reference: nurse)	0.63 (0.27-1.49)	.59
Web (reference: social interviewer)	0.45 (0.17-1.18)	.14
Respiratory conditions		
Nurse (reference: social interviewer)	0.77 (0.44-1.36)	.82
Web (reference: nurse)	0.54 (0.27-1.09)	.11
Web (reference: social interviewer)	0.41 (0.17-1.03)	.06

Tables S7-S10 in Multimedia Appendix 1 present the results of moderation analyses stratified by sex and age group. There were no statistically significant differences in SF-12 or GHQ-12 scores between male and female individuals (Table S7 in Multimedia Appendix 1; overall interaction *P*>.99, except for GHQ-12, *P*=.25). The overall interaction between age group and survey mode was statistically significant for the SF-12 PCS score (*P*=.04) but none of the individual interaction contrasts reached statistical significance. The corresponding interactions were not statistically significant for the GHQ-12 or SF-12 MCS scores (Table S8 in Multimedia Appendix 1; *P*>.99).

For chronic conditions, there was no evidence of an overall interaction between sex and survey mode (all *P*>.99, except for musculoskeletal disorders, *P*=.25). Nevertheless, among men, musculoskeletal conditions were reported more frequently in nurse interviews than in social interviewer CAPI (OR 2.43, 95% CI 1.14 - 5.20; *P*=.01) and in web self-completion mode (OR 0.43, 95% CI 0.19 - 0.97; *P*=.04; reference: nurse mode; Table S9 in Multimedia Appendix 1). The corresponding risk differences exceeded 5% (Figure S1 in Multimedia Appendix 1). In contrast, among women, the effect estimates were closer to the null, with ORs of 1.15 (95% CI 0.52 - 2.54) for nurse vs social interviewer, 1.10 (95% CI 0.51 - 2.35) for web vs nurse, and 1.26 (95% CI 0.47 - 3.41) for web vs social interviewer. For respiratory conditions, the direction of associations was consistent in men and women, with all pairwise ORs below 1, although the reductions were more pronounced among men. Among men, web self-completion was associated with lower odds than social interviewer CAPI (OR 0.32, 95% CI 0.10 - 1.00; *P*=.049) and nurse CAPI (OR 0.45, 95% CI 0.20 - 1.00; *P*=.05). Among women, the corresponding ORs were 0.53 and 0.64, respectively, with confidence intervals including 1. The overall interaction between age and survey modes was not statistically significant for any chronic condition (all *P*>.99, except for respiratory conditions, *P*=.86). Respiratory conditions were reported significantly less often via web-based self-completion among respondents aged 56 years and older compared to social interviewer CAPI (OR 0.25, 95% CI 0.09 - 0.71; *P*=.003) and nurse CAPI (OR 0.32, 95% CI 0.13 - 0.79; *P*=.005; Table S10 in Multimedia Appendix 1), with mode differences exceeding 5% (Figure S2 in Multimedia Appendix 1). However, the direction of associations was consistent across age groups; among respondents aged less than 56 years, the corresponding ORs for web versus social

XSL•FO
RenderX

interviewer and web versus nurse were 0.53 and 0.70, respectively, although their confidence intervals included 1.

The results of the supplementary analysis of the number of chronic-condition categories are shown in Table S11 in Multimedia Appendix 1. Participants completing the survey via the web reported fewer chronic conditions than those assessed by nurses (IRR 0.72, 95% CI, 0.54 - 0.96; *P*=.02). No significant differences were observed between nurse and social-interviewer modes (IRR 1.00, 95% CI, 0.83 - 1.20; *P*>.99) or between web and social-interviewer modes (IRR 0.72, 95% CI, 0.50 - 1.02; *P*=.07).

In sensitivity analyses excluding participants who completed a survey mode that differed from their originally allocated mode (Table S12 in Multimedia Appendix 1), findings were consistent with the main analyses. The only difference was that web mode (vs social interviewer) was associated with lower odds of respiratory conditions (OR 0.49; 95% CI, 0.26 - 0.94; *P*=.03), whereas this association was marginal (*P*=.06) in the main analysis.

Discussion

In this nationally representative study in Great Britain, we did not find clear evidence of differences in self-completed GHQ-12 or SF-12 scores across modes that varied in interviewer presence and role. We found no statistically significant evidence of mode-related differences in reports of doctor-diagnosed mental health and other conditions. However, the tendency for web self-completion to yield lower reporting of chronic conditions indicates that mode-related differences cannot be excluded. There was little evidence overall that mode-related differences varied by age or sex. Although some subgroup contrasts reached statistical significance, respiratory-disease estimates were broadly consistent across strata, while the musculoskeletal contrast within the male subgroup should be interpreted cautiously.

Although previous studies suggest that responses to self-completion instruments may vary according to the interview context [19,20], our study did not find clear evidence of differences in the self-completed GHQ-12 and the mental and physical components of the self-completed SF-12 across modes that varied in interviewer presence and role. Notably, a previous study found differences in GHQ-12 responses even though the GHQ-12 was self-completed by respondents in both interviewer-administered and self-completion survey modes, suggesting that contextual factors such as interviewer presence may still influence responses [19]. In contrast, the limited differences observed in our study may partly reflect the prior survey experience of the social interviewers and the study-specific briefing social interviewers and nurses received on fieldwork and interview procedures, together with CASI procedures that reduced direct interviewer involvement and protected the confidentiality of completed responses [51]. This interpretation is also consistent with recent evidence from a randomized study, which suggested that differences in the privacy and confidentiality afforded by self-completion modes may influence disclosure of mental health and other sensitive information [20]. Although our point estimates for GHQ-12 and SF-12 MCS were slightly more favorable in the nurse mode than in the social interviewer or web modes, the differences were negligible and not statistically significant. Differences between web and social interviewer modes were also minimal. These findings provide little evidence that either the presence of an interviewer during CAPI or the interviewer's professional role introduced systematic bias in responses to the self-completed GHQ-12 or SF-12 instruments. Furthermore, our findings may support the use of appropriately implemented CASI within CAPI to minimize the mode effects previously observed when health scales were administered directly by interviewers [11,13,15,19,24]. This approach is widely adopted in many national surveys; for example, the 2023-2024 Adult Psychiatric Morbidity Survey in England similarly adopted a self-completion section within CAPI interviews to mitigate social desirability bias in sensitive item responses [18]. Overall, these findings may support the use of self-administered approaches, whether web-based or embedded as CASI within interviewer-administered surveys, particularly for administering sensitive health-related scales.

We also did not find statistically significant evidence of mode effects for doctor-diagnosed psychiatric conditions, similar to findings from previous research based on diagnostic algorithm measures [24]. A possible explanation is that factual questions with simple categorical response options may be less susceptible to social desirability bias than multi-item or more subjective mental health measures [9,24]. For other doctor-diagnosed chronic conditions, the reporting was also broadly similar across web self-completion and interviewer-administered modes, with no statistically significant evidence of mode effects. No consistent pattern emerged when comparing outcomes between nurse interviewers and lay interviewers. However, descriptive prevalence estimates tended to be lower for web self-completion, and adjusted OR estimates were generally in the direction of lower reporting for web self-completion, although the 95% CIs were wide and included the null. A supplementary analysis of the number of chronic-condition categories also suggested that web respondents reported fewer condition categories than those interviewed by nurses. The estimate was similar but less precise when web respondents were compared with social-interviewer respondents. This pattern is broadly consistent with previous research suggesting a tendency toward lower reporting of physical health conditions in web self-completion than in interviewer-administered modes [24]. One possible explanation is that interviewer probing and clarification facilitate the elicitation of chronic conditions that might otherwise go unreported in self-administered questionnaires [9]. Overall, although no statistically significant evidence of mode effects was observed for individual doctor-diagnosed chronic-condition categories, our findings provide further context for ongoing discussion of this potential mechanism. Mode-related differences in the overall reporting of doctor-diagnosed chronic conditions therefore cannot be excluded.

When the analysis was restricted to participants who did not undergo a mode switch, the association between web self-completion and respiratory conditions, relative to interviewer administration, strengthened from marginal to statistically significant (*P*=.03, compared with *P*=.06 in the

XSL•FO
RenderX

main analysis), with respondents completing the survey online reporting fewer respiratory conditions. Previous evidence likewise suggests lower reporting of respiratory conditions in self-administered questionnaires [24], although findings across studies have not been consistent [9]. However, the overall pattern of point estimates for the other health conditions remained similar to that observed in the main analysis, with generally lower reported prevalence under web self-completion than interviewer administration, although these differences were not statistically significant. Thus, the finding for respiratory conditions may reflect a broader mode-related pattern in health reporting rather than an effect specific to respiratory conditions.

No significant overall interactions of self-completion contexts with age or sex were observed for GHQ-12 or SF-12, except for age differences in SF-12 PCS. However, as the age-stratified estimates were not statistically significant and direct supporting evidence is limited, this finding should be interpreted with caution. The overall interaction tests did not provide clear evidence that mode effects on chronic conditions varied by age or sex. Although individual contrasts reached statistical significance among adults aged 56 years or older and men, respiratory conditions showed a broadly consistent pattern across age and sex strata. This consistency suggests that the observed differences may reflect a general mode-related reporting pattern rather than clear effect modification by age or sex. Among men, musculoskeletal conditions were reported more frequently in the nurse-administered mode than in social interviewer or web self-completion modes, possibly because the nurse-administered context may facilitate reporting of symptoms that might otherwise be normalized or downplayed [52]. Given the nonsignificant overall interactions and the limited statistical power of the present study, the subgroup pattern should be regarded as exploratory.

This study offers several key advantages over prior research. First, we leveraged an existing randomized allocation of survey modes within a probability-based sample and applied an instrumental-variable approach to address potential selection bias in mode choice, thereby strengthening the generalizability and robustness of our results. Although our focus was on mode effects on measurement, we noted that self-selection can also operate through sample composition as a mode-related process. The instrumental-variable design helped mitigate concerns about such nonrandom mode choice. Second, inclusion of both standardized health scales and doctor-diagnosed chronic conditions allowed us to examine mode-related differences across different types of health indicators and administration conditions. Finally, by comparing the social interviewer-administered mode, web-based self-administration, and a nurse-administered mode, we offer additional insights into how nurse-administered CAPI may be associated with differences in self-reported health.

This study has several limitations. First, although our sample included participants from diverse racial and ethnic backgrounds, the small sample sizes of minority groups necessitated aggregation, limiting our ability to examine survey‐mode effects on self‐reported health outcomes within specific subgroups. Second, reports of doctor‐diagnosed chronic diseases were sparse and therefore combined by disease family, hindering further analysis of condition‐specific reporting patterns. Nevertheless, the effect estimates were generally in the direction of lower odds of reporting conditions with web self-completion than with interviewer-administered modes, but the 95% CIs were wide. Statistical power to detect mode-related differences was limited; post hoc power ranged from 57.2% to 61.1% across the 3 pairwise comparisons. Third, this study only assessed the association between survey mode and self‐reported health and did not directly assess disease-specific stigma. The study therefore could not assess whether the extent or direction of mode differences varied according to the perceived sensitivity or stigma associated with particular conditions. Finally, the asymmetric reallocation process may have resulted in residual confounding of the estimated effect of mode received. If unmeasured respondent characteristics such as internet access, literacy, health, or engagement with the study influenced which initially assigned mode was refused, the asymmetric reallocation structure could have differentially concentrated respondents with those characteristics in particular observed mode groups. Consequently, the estimated effect of mode received may partly reflect these asymmetry-induced compositional differences. However, sensitivity analyses restricted to participants who adhered to their original randomized mode allocation produced broadly similar findings, suggesting that the main results were unlikely to be primarily driven by the asymmetric reallocation of participants across survey modes. Moreover, the observed covariate imbalances between randomized groups indicate that the independence assumption cannot be regarded as fully satisfied on the basis of randomization alone. The instrumental variable approach therefore relied on conditional independence after adjustment for measured covariates. Residual confounding due to unmeasured factors cannot be excluded and remains a limitation.

Overall, this study provides important insights for the implementation of population-based health surveys. First, the absence of differences in GHQ-12 and SF-12 scores across self-completion contexts may support the use of CASI within face-to-face interviews for collecting these measures. Second, no statistically significant mode effects were observed for psychiatric or other doctor-diagnosed conditions. However, given the tendency for web self-completion to yield lower reporting of chronic conditions, mode-related differences, potentially reflecting interviewer probing and clarification, cannot be excluded and warrant further investigation. Finally, there was no clear evidence that mode-related reporting differed by age or sex; the musculoskeletal finding in men remains exploratory given limited power.

Acknowledgments

During the preparation of this work, ChatGPT-5 (OpenAI) was used to refine the paper's grammar and readability.

XSL•FO
RenderX

Funding

YL, TAB, JB, MK, and data collection for this study were funded by Economic and Social Research Council (ES/N00812X/1 and ES/T002611/1). The funders had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. Understanding Society: UK Household Longitudinal Study is an initiative funded by the Economic and Social Research Council (ES/S007253/1).

Data Availability

The dataset is available at UK Data Service [53]. R code supporting the main analyses is available on GitHub [54].

Authors' Contributions

Conceptualization: YL, MK
Data curation: YL
Formal analysis: YL
Methodology: YL, TAB, JB, MK
Visualization: YL
Writing – original draft: YL, MK
Writing – review & editing: YL, TAB, JB, MK

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.
[DOCX File, 651 KB - humanfactors_v13i1e89896_app1.docx]

References

1. Lynn P. How to design and implement surveys that are fit for purpose. Nat Hum Behav 2025 Dec;9(12):2412-2414. [doi: 10.1038/s41562-025-02317-z] [Medline: 41034565]
2. van Alten S, Domingue BW, Faul J, Galama T, Marees AT. Reweighting UK Biobank corrects for pervasive selection bias due to volunteering. Int J Epidemiol 2024 Apr 11;53(3):dyae054. [doi: 10.1093/ije/dyae054] [Medline: 38715336]
3. Marker DA, Hilton C, Zelko J, et al. Real world data versus probability surveys for estimating health conditions at the state level. J Surv Stat Methodol 2024 Nov;12(5):1515-1530. [doi: 10.1093/jssam/smae036] [Medline: 39781537]
4. Windle J, Rolfe J. Comparing responses from internet and paper-based collection methods in more complex stated preference environmental valuation surveys. Econ Anal Policy 2011 Mar;41(1):83-97. [doi: 10.1016/S0313-5926(11)50006-2]
5. Sakshaug JW, Ofstedal MB, Guyer H, Beebe TJ. The collection of biospecimens in health surveys. In: Johnson TP, editor. Handbook of Health Survey Methods: John Wiley & Sons, Ltd; 2014:383-419.
6. Kibuchi E, Sturgis P, Durrant GB, Maslovskaya O. The efficacy of propensity score matching for separating selection and measurement effects across different survey modes. J Surv Stat Methodol 2024 Jun;12(3):764-789. [doi: 10.1093/jssam/smae017]
7. Jans M. Mode effects. In: Lavrakas PJ, editor. Encyclopedia of Survey Research Methods: SAGE Publications; 2008:476-480.
8. Zager Kocjan G, Lavtar D, Sočan G. The effects of survey mode on self-reported psychological functioning: measurement invariance and latent mean comparison across face-to-face and web modes. Behav Res Methods 2023 Apr;55(3):1226-1243. [doi: 10.3758/s13428-022-01867-8] [Medline: 35618987]
9. Christensen AI, Ekholm O, Glümer C, Juel K. Effect of survey mode on response patterns: comparison of face-to-face and self-administered modes in health surveys. Eur J Public Health 2014 Apr;24(2):327-332. [doi: 10.1093/eurpub/ckt067] [Medline: 23766339]
10. West BT, Blom AG. Explaining interviewer effects: a research synthesis. J Surv Stat Methodol 2017 Jun;5(2):175-211. [doi: 10.1093/jssam/smw024]
11. Cernat A, Couper MP, Ofstedal MB. Estimation of mode effects in the Health and Retirement Study using measurement models. J Surv Stat Methodol 2016 Dec;4(4):501-524. [doi: 10.1093/jssam/smw021] [Medline: 28612033]
12. Buskirk TD, Stein KD. Telephone vs. mail survey gives different SF-36 quality-of-life scores among cancer survivors. J Clin Epidemiol 2008 Oct;61(10):1049-1055. [doi: 10.1016/j.jclinepi.2007.11.012] [Medline: 18538997]
13. Hays RD, Kim S, Spritzer KL, et al. Effects of mode and order of administration on generic health-related quality of life scores. Value Health 2009 Sep;12(6):1035-1039. [doi: 10.1111/j.1524-4733.2009.00566.x] [Medline: 19473334]
14. Clarke PS, Bao Y. Estimating mode effects from a sequential mixed-mode experiment using structural moment models. Ann Appl Stat 2022 Sep;16(3):1563-1585. [doi: 10.1214/21-AOAS1557]

XSL•FO
RenderX

15. Ravens-Sieberer U, Erhart M, Wetzel R, Krügel A, Brambosch A. Phone respondents reported less mental health problems whereas mail interviewee gave higher physical health ratings. J Clin Epidemiol 2008 Oct;61(10):1056-1060. [doi: 10.1016/j.jclinepi.2007.12.003] [Medline: 18538990]
16. National survey on drug use and health: summary of methodological studies, 1971–2014. : Substance Abuse and Mental Health Services Administration (US); 2014 URL: https://www.ncbi.nlm.nih.gov/books/NBK519735/pdf/Bookshelf_NBK519735.pdf [accessed 2026-09-29]
17. Erens B, Phelps A, Clifton S, et al. Methodology of the third British National Survey of Sexual Attitudes and Lifestyles (Natsal-3). Sex Transm Infect 2014 Mar;90(2):84-89. [doi: 10.1136/sextrans-2013-051359] [Medline: 24277881]
18. Adult psychiatric morbidity survey: survey of mental health and wellbeing, England 2023 to 2024: methods. NHS England. 2024. URL: https://digital.nhs.uk/data-and-information/publications/statistical/adult-psychiatric-morbidity-survey/survey-of-mental-health-and-wellbeing-england-2023-24/methods [accessed 2025-08-12]
19. Tipping S, Hope S, Pickering K, Erens B, Roth MA, Mindell JS. The effect of mode and context on survey results: analysis of data from the Health Survey for England 2006 and the Boost Survey for London. BMC Med Res Methodol 2010 Sep 27;10(1):84. [doi: 10.1186/1471-2288-10-84] [Medline: 20868528]
20. Simms V, Kanengoni B, Chingono R, et al. Effect of data-collection method on reporting of common mental disorder symptoms and intimate partner violence in Zimbabwe: a cluster-randomized trial. Int J Epidemiol 2026 Jan 2;55(1):dyaf221. [doi: 10.1093/ije/dyaf221] [Medline: 41533677]
21. Milton AC, Ellis LA, Davenport TA, Burns JM, Hickie IB. Comparison of self-reported telephone interviewing and web-based survey responses: findings from the second Australian Young and Well National Survey. JMIR Ment Health 2017 Sep 26;4(3):e37. [doi: 10.2196/mental.8222] [Medline: 28951382]
22. Tagseth M, Sund ER, Hallman GTA, et al. May telephone surveys provide reliable public health surveillance data for municipalities? Mode effects differ between categories of questions. The HUNT Study, Norway. Nor Epidemiol 2019;28(1-2). [doi: 10.5324/nje.v28i1-2.3057]
23. Smeeth L, Fletcher AE, Stirling S, et al. Randomised comparison of three methods of administering a screening questionnaire to elderly people: findings from the MRC trial of the assessment and management of older people in the community. BMJ 2001 Dec 15;323(7326):1403-1407. [doi: 10.1136/bmj.323.7326.1403] [Medline: 11744565]
24. Hoebel J, von der Lippe E, Lange C, Ziese T. Mode differences in a mixed-mode health interview survey among adults. Arch Public Health 2014;72(1):46. [doi: 10.1186/2049-3258-72-46] [Medline: 25810913]
25. Greene J, Speizer H, Wiitala W. Telephone and web: mixed-mode challenge. Health Serv Res 2008 Feb;43(1 Pt 1):230-248. [doi: 10.1111/j.1475-6773.2007.00747.x] [Medline: 18211527]
26. Riessman CK. Interviewer effects in psychiatric epidemiology: a study of medical and lay interviewers and their impact on reported symptoms. Am J Public Health 1979 May;69(5):485-491. [doi: 10.2105/ajph.69.5.485] [Medline: 434280]
27. Kutschar P, Osterbrink J, Weichbold M. Interviewer effects in a survey examining pain intensity and pain interference in nursing home residents. Age Ageing 2022 Feb 2;51(2):afac008. [doi: 10.1093/ageing/afac008] [Medline: 35180286]
28. Gnambs T, Kaspar K. Disclosure of sensitive behaviors across self-administered survey modes: a meta-analysis. Behav Res Methods 2015 Dec;47(4):1237-1259. [doi: 10.3758/s13428-014-0533-4] [Medline: 25410404]
29. Luong G, Charles ST, Rook KS, Reynolds CA, Gatz M. Age differences and longitudinal change in the effects of data collection mode on self-reports of psychosocial functioning. Psychol Aging 2015 Mar;30(1):106-119. [doi: 10.1037/a0038502] [Medline: 25528065]
30. Understanding Society: Innovation Panel, waves 1-16, 2008-2023. : University of Essex, Institute for Social and Economic Research; 2024 URL: https://www.understandingsociety.ac.uk/wp-content/uploads/working-papers/2024-11.pdf [accessed 2026-09-29]
31. Lynn P. Sample design for Understanding Society. : Institute for Social and Economic Research, University of Essex; 2009 Jan URL: https://www.understandingsociety.ac.uk/wp-content/uploads/working-papers/2009-01.pdf [accessed 2026-09-29]
32. Understanding Society - the UK household longitudinal study, innovation panel, waves 1-16, user manual. : Institute for Social and Economic Research, University of Essex; 2024 URL: https://doc.ukdataservice.ac.uk/doc/7332/mrdoc/pdf/7332_ip_waves_1-16_user_guide_oct_2024.pdf? [accessed 2025-07-20]
33. Al Baghal T, Benzeval M, Burton J, Crossley TF, Kumari M, Rajatileka S. Collection of biomarkers using nurses, interviewers, and participants: the design of IP12. : Institute for Social and Economic Research, University of Essex; 2021 Jun URL: https://www.understandingsociety.ac.uk/wp-content/uploads/working-papers/2021-06.pdf [accessed 2025-07-21]
34. Goldberg DP, Gater R, Sartorius N, et al. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. Psychol Med 1997 Jan;27(1):191-197. [doi: 10.1017/s0033291796004242] [Medline: 9122299]
35. Ware JE. How to Score Version 2 of the SF-12 Health Survey (with a Supplement Documenting Version 1): Quality Metric Incorporated; 2002.
36. Benzeval M, Andrayas A, Mazza J, et al. Does the feedback of blood results in observational studies influence response and consent? A randomised study of the Understanding Society Innovation Panel. BMC Med Res Methodol 2023 Jun 7;23(1):134. [doi: 10.1186/s12874-023-01948-y] [Medline: 37280544]

XSL•FO
RenderX

37. Al Baghal T, Burton J, Crossley TF, Benzeval M, Kumari M. How different mixed-mode data collection approaches impact response rates and provision of biomeasure samples. Public Opin Q 2025;89(2):245-269. [doi: 10.1093/poq/nfaf022] [Medline: 40852323]
38. Pierce M, McManus S, Hope H, et al. Mental health responses to the COVID-19 pandemic: a latent class trajectory analysis using longitudinal UK data. Lancet Psychiatry 2021 Jul;8(7):610-619. [doi: 10.1016/S2215-0366(21)00151-6] [Medline: 33965057]
39. Resolution concerning statistics of work, employment and labour underutilization (including amendments). : International Labour Organization; 2023 URL: https://www.ilo.org/sites/default/files/wcmsp5/groups/public/%40dgreports/%40stat/documents/normativeinstrument/wcms_230304.pdf [accessed 2026-09-09]
40. Basu A, Coe NB, Chapman CG. 2SLS versus 2SRI: appropriate methods for rare outcomes and/or rare exposures. Health Econ 2018 Jun;27(6):937-955. [doi: 10.1002/hec.3647] [Medline: 29577493]
41. Baiocchi M, Cheng J, Small DS. Tutorial in biostatistics: instrumental variable methods for causal inference. Stat Med 2014 Jun 15;33(13):2297-2340. [doi: 10.1002/sim.6128] [Medline: 24599889]
42. Terza JV, Basu A, Rathouz PJ. Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. J Health Econ 2008 May;27(3):531-543. [doi: 10.1016/j.jhealeco.2007.09.009] [Medline: 18192044]
43. Walker V, Sanderson E, Levin MG, Damraurer SM, Feeney T, Davies NM. Reading and conducting instrumental variable studies: guide, glossary, and checklist. BMJ 2024 Oct 14;387:e078093. [doi: 10.1136/bmj-2023-078093] [Medline: 39401839]
44. Clement ND, Weir D, Holland J, Gerrand C, Deehan DJ. Meaningful changes in the Short Form 12 physical and mental summary scores after total knee arthroplasty. Knee 2019 Aug;26(4):861-868. [doi: 10.1016/j.knee.2019.04.018] [Medline: 31109846]
45. Ohno S, Takahashi K, Inoue A, et al. Smallest detectable change and test-retest reliability of a self-reported outcome measure: results of the Center for Epidemiologic Studies Depression Scale, General Self-Efficacy Scale, and 12-item General Health Questionnaire. J Eval Clin Pract 2017 Dec;23(6):1348-1354. [doi: 10.1111/jep.12795] [Medline: 28758322]
46. Shah AD, Bartlett JW, Carpenter J, Nicholas O, Hemingway H. Comparison of random forest and parametric imputation models for imputing missing data using MICE: a CALIBER study. Am J Epidemiol 2014 Mar 15;179(6):764-774. [doi: 10.1093/aje/kwt312] [Medline: 24589914]
47. Bartlett JW, Hughes RA. Bootstrap inference for multiple imputation under uncongeniality and misspecification. Stat Methods Med Res 2020 Dec;29(12):3533-3546. [doi: 10.1177/0962280220932189] [Medline: 32605503]
48. Kalpourtzi N, Carpenter JR, Touloumi G. Handling missing values in surveys with complex study design: a simulation study. J Surv Stat Methodol 2024;12(1):105-129. [doi: 10.1093/jssam/smac039]
49. Zhang J, Dashti SG, Carlin JB, Lee KJ, Moreno-Betancur M. Should multiple imputation be stratified by exposure group when estimating causal effects via outcome regression in observational studies? BMC Med Res Methodol 2023 Feb 16;23(1):42. [doi: 10.1186/s12874-023-01843-6] [Medline: 36797679]
50. Kaminska O, Lynn P. Weighting and sample representation: frequently asked questions. : Institute for Social and Economic Research, University of Essex; 2019 URL: https://www.understandingsociety.ac.uk/wp-content/uploads/documentation/user-guides/6614-user-guide-weighting-faqs.pdf [accessed 2026-10-07]
51. De Leeuw ED. To mix or not to mix data collection modes in surveys. J Off Stat 2005;21(5):233-255 [FREE Full text]
52. Pedulla R, Glugosh J, Jeyaseelan N, et al. Associations of gender role and pain in musculoskeletal disorders: a mixed-methods systematic review. J Pain 2024 Dec;25(12):104644. [doi: 10.1016/j.jpain.2024.104644] [Medline: 39084479]
53. Understanding Society: Innovation Panel, waves 1-15, 2008-2022. UK Data Service. URL: http://doi.org/10.5255/UKDA-SN-6849-15 [accessed 2026-09-29]
54. Researchhealth6/modeeffects_health. GitHub. URL: https://github.com/researchhealth6/Modeeffects_health [accessed 2026-10-08]

Abbreviations

2SRI: two-stage residual inclusion
CAPI: computer-assisted personal interviewing
CASI: computer-assisted self-interviewing
GCSE: General Certificate of Secondary Education
GHQ-12: 12-item General Health Questionnaire
IP: Innovation Panel
IRR: incidence rate ratio
MCS: Mental Component Summary
OR: odds ratio
PCS: Physical Component Summary
SF-12: 12-item Short Form Health Survey
UKHLS: UK Household Longitudinal Study

XSL•FO
RenderX

Edited by S Law; submitted 18.Dec.2025; peer-reviewed by B Rodgers, JF Yap; revised version received 13.Sep.2026; accepted 17.Sep.2026; published 16.Oct.2026.

Please cite as:
Liang Y, Al Baghal T, Burton J, Kumari M
Mode Effects on Self-Reported Health Across Web-Based, Interviewer-Administered, and Nurse-Administered Modes: Population-Based Study
JMIR Hum Factors 2026;13:e89896
URL: https://humanfactors.jmir.org/2026/1/e89896
doi:10.2196/89896

© Yunlong Liang, Tarek Al Baghal, Jonathan Burton, Meena Kumari. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 16.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Japanese Health Information Technology Usability Evaluation Scale for Sexually Transmitted Infection–Related Chatbots: Development and Psychometric Validation Study

Tomoko Hato[1], MD; Hirono Ishikawa[1], PhD; Kense Todo[2], MEng; Keisuke Harada[3], MEng; Atsushi Yoshikawa[2], DEng; Yoshiharu Fukuda[1], MD, PhD; Rebecca Schnall[4], RN, MPH, PhD

[1]Graduate School of Public Health, Teikyo University, 2-11-1 Kaga, Itabashi-ku, Tokyo, Japan

[2]College of Informatics, Kanto Gakuin University, Kanagawa, Japan

[3]School of Computing, Institute of Science Tokyo, Tokyo, Japan

[4]School of Nursing, Columbia University, New York, NY, United States

Corresponding Author:
Hirono Ishikawa, PhD
Graduate School of Public Health, Teikyo University, 2-11-1 Kaga, Itabashi-ku, Tokyo, Japan

Abstract

Background: The rapid expansion of mobile technology has accelerated the integration of health applications and conversational AI into clinical and public health practices. To ensure these tools are effective and sustainable, usability evaluations and early user engagement during development are essential. The Health Information Technology Usability Evaluation Scale (Health-ITUES) is a validated and flexible usability assessment instrument that is available in multiple languages and applicable across diverse contexts. However, a Japanese version of this scale has not yet been developed.

Objective: This study aimed to translate and validate a Japanese version of the Health-ITUES, customized for a sexually transmitted infection (STI)–related chatbot, and to support the usability assessment of emerging mobile health tools in Japan.

Methods: We developed a Japanese version of the Health-ITUES using a chatbot under development as a consultation tool for young women regarding STIs. First, the original scale was customized to reflect the chatbot's specific purpose and intended usage context. Following established translation guidelines, we conducted forward translation from English to Japanese, back translation, expert review, and reconciliation. We then evaluated the reliability and validity of the Japanese version in a sample of 301 young women.

Results: The Japanese version of the Health-ITUES demonstrated high internal consistency (Cronbach α=0.85 - 0.98). Confirmatory factor analysis supported acceptable construct validity (root mean square error of approximation is 0.10, comparative fit index>0.90). Additionally, the Health-ITUES scores showed strong correlations with satisfaction and usage intention for the tool (r=0.779 and 0.797, respectively).

Conclusions: The Japanese version of the Health-ITUES provides initial evidence of reliability and validity in an STI-related scenario among young women and may facilitate more rigorous usability evaluations of mHealth and conversational AI tools in Japan.

(JMIR Hum Factors 2026;13:e90483) doi:10.2196/90483

KEYWORDS

mobile health; mHealth; chatbot; usability; translation; validation; sexually transmitted infection; STI; artificial intelligence; conversational AI

Introduction

In recent years, the widespread adoption of smartphones and tablets has led to a significant increase in their daily usage. According to the Ministry of Internal Affairs and Communications' 2024 survey on information and communication usage trends, the individual smartphone ownership rate in Japan has reached 80.5%, with over 90% penetration among individuals aged 13 to 69 years [1]. As information and communication technologies continue to evolve, mobile health (mHealth), which uses smartphones and portable devices for health care, has been rapidly adopted worldwide. mHealth is already applied across various domains, such as health management, patient education, and disease prevention, and is expected to improve health care accessibility and support self-management practices [2-4].

However, several mHealth technologies have become publicly available without sufficient validation of their design,

XSL•FO
RenderX

development, or evaluation processes [5]. Although acknowledging the potential of mobile-based interventions in strengthening health systems, the World Health Organization (WHO) emphasizes the need for careful assessment of their acceptability, effectiveness, feasibility, and equity in its guidelines on digital health interventions [6]. The effectiveness and sustainability of mHealth interventions depend not only on technical performance but also on user-centered evaluations of usability and acceptability [7]. Accordingly, evaluating mHealth technologies requires a more comprehensive approach beyond measuring clinical outcomes alone. It is essential to assess their usability and acceptability from the user's perspective and incorporate user feedback from the early stages of development. In recent years, the emergence of conversational AI systems, such as ChatGPT, has further expanded the scope of mHealth. Although these systems still face challenges related to information reliability and data security, they offer the potential to facilitate interactive bidirectional communication with users, thereby contributing to behavioral change, health education, and self-management support [8,9]. Previous studies have shown that, by demonstrating empathy and a sense of familiarity, chatbots can improve health outcomes in stigmatized areas, such as sexually transmitted infections (STIs) and mental health [10-13]. However, their effectiveness may be limited if they fail to recognize user characteristics, expectations, or levels of understanding [14,15]. Even when large language model–based conversational AIs, such as ChatGPT, provide highly accurate responses, the quality of user interaction can significantly influence the ability to extract relevant information. In some cases, these interactions may even lead to misinterpretations or hinder users' health-related decision-making [16]. Therefore, usability evaluation has become a critical component in the design and assessment of such systems.

Usability is defined as the extent to which users can effectively and efficiently achieve specific goals with satisfaction. Several usability assessment tools have been developed to evaluate these aspects [17]. Early examples include the System Usability Scale (SUS) [18] and the Post-Study System Usabiity Questionnaire (PSSUQ) [19], which were originally designed for general IT systems. However, prior research has indicated that general-purpose usability measures may not adequately capture the specific requirements of health care contexts [20]. Tools such as the Mobile App Rating Scale (MARS/uMARS) [21,22], the mHealth App Usability Questionnaire (MAUQ), and the Health Information Technology Usability Evaluation Scale (Health-ITUES) [23] have been developed for mHealth-specific applications.

Health-ITUES was originally developed using a web-based communication system designed for managing nursing staff schedules [23]. Since then, it has been applied to evaluate the usability of various health technologies, including mobile applications for individuals living with HIV, apps for managing dysmenorrhea, and integrated care systems for older adults [24-27]. The Health-ITUES is one of the few instruments whose psychometric properties have been validated in accordance with the COSMIN (Consensus-Based Standards for the Selection of Health Measurement Instruments) guidelines [17]. Unlike general usability scales (eg, SUS), the Health-ITUES is designed specifically for health information technologies and is grounded in the Health IT Usability Evaluation Model (Health-ITUEM) framework, enabling the evaluation of clinically meaningful usability constructs [23,27]. It maintains conceptual comparability across different contexts while allowing item-level customization to suit specific tasks and objectives within health care systems [28,29].

To date, the Health-ITUES has been translated and validated in several languages, such as Korean [25] and Chinese [26], demonstrating its cross-cultural applicability. However, no validated Japanese version is currently available. We selected Health-ITUES for this study because its health-IT–oriented design, validated psychometric robustness, and flexibility for context-specific item adaptation make it particularly well suited for evaluating an STI-related chatbot scenario targeting young women [25,27]. By developing and validating a Japanese version tailored to STI-related consultations, this study aimed to provide a rigorous and contextually relevant tool for evaluating the usability of emerging mHealth and conversational AI tools in Japan.

Methods

Modification of the Health-ITUES

The Health-ITUES consists of 20 items and can be customized according to the characteristics of the device and the intended use. Originally, it was developed based on the Technology Acceptance Model (TAM) to evaluate the usability of a web-based scheduling system for nursing staff [23]. Subsequently, Schnall et al [27] further validated the scale using a self-management application for individuals living with HIV, focusing on the evaluation of self-management mHealth applications.

In this study, we customized the Health-ITUES for young women experiencing anxiety related to STIs, based on the validated versions used among HIV-positive individuals in the United States and women with menstrual discomfort in South Korea. Prior to initiating the translation and adaptation process, we obtained permission from the original author to develop a Japanese version of the Health-ITUES.

Japanese Version of the Health-ITUES

The translation process adhered to established guidelines for scale translation [30] and followed the sequence as described in the following sections.

Forward Translation and Reconciliation

Two researchers (a physician and a health communication specialist) independently translated the original English version into Japanese. Subsequently, they discussed and reconciled their translations. A health communication research group, comprising physicians and health psychology experts, reviewed the preliminary Japanese version to ensure clarity, cultural appropriateness, and linguistic accuracy.

Back Translation

Two professional medical translators independently translated the Japanese version of the questionnaire back into English.

XSL•FO
RenderX

The translations were then compared to confirm consistency and accuracy.

Back-Translation Review

The authors reviewed the back-translated version by retranslating it into Japanese and comparing it with the initial Japanese draft to identify discrepancies.

Original Author Review

After completing both the forward and backward translations, the original author reviewed the English version of the Japanese Health-ITUES to confirm its content validity.

Cognitive Debriefing

We conducted cognitive debriefing interviews to assess the clarity and contextual appropriateness of the Japanese version. Linguistic validation guidelines recommend a small sample from the target population, with a Japanese guideline suggesting approximately 5 to 8 participants [30-32]. In this study, 6 female students studying at a university without a medical background were interviewed using the Japanese Health-ITUES. They were asked whether any items were difficult to understand or caused hesitation when responding.

Reliability and Validity

We conducted an online survey using a chatbot under development by the authors to evaluate the reliability and validity of the Japanese version of the Health-ITUES. The chatbot was designed to alleviate anxiety related to STIs among young women and to guide them toward appropriate consultation services and testing resources. Internal consistency reliability was assessed using the Cronbach α coefficient. Construct validity was examined through confirmatory factor analysis (CFA), and criterion-related validity was evaluated by examining correlations with user satisfaction and intention to use.

Participants

Participants were women aged 18 to 29 years registered with a Japanese research agency (Asmarq Inc). According to national infectious disease surveillance data [33], reported syphilis cases—one of the major STIs—among women in their 20s have increased substantially since 2020, making this demographic particularly relevant for the chatbot's intended use. Eligibility criteria included the ability to communicate in Japanese, access the survey URL using personal computers or smartphones, and have had at least one prior sexual experience. Individuals who were health care professionals, medical students, or already engaged with STI-related medical services or consultation centers were excluded from the study.

Sample Size

To ensure adequate statistical power in scale validation studies, a commonly used rule of thumb is to recruit approximately 10 participants per item [32,34]. Given that the Health-ITUES consists of 20 items, we estimated the minimum required sample size to be 200 participants. Assuming a 20% dropout rate, the target sample size was set at 250 participants.

Procedure

Participants were recruited through a research agency. The agency randomly selected 57,185 women within the target age range from its registered panel and sent email invitations to participate in the study. Only those who provided informed consent via an embedded survey link were directed to the main questionnaire. The survey was terminated once the number of responses exceeded the target sample size. Data were collected between June 18, 2025, and June 20, 2025. A total of 301 valid responses were obtained.

Participants were informed via a web-based survey interface that the study involved interacting with a chatbot based on a hypothetical scenario (ie, their partner suddenly informed them of symptoms suggestive of an STI). Instructions on how to use the chatbot were also provided. Initially, participants completed prechatbot questions on age, region of residence, occupation, educational background, relationship status, frequency of conversational AI use, and digital health literacy [35]. They then accessed the chatbot interface through a URL embedded in the survey and interacted with it according to the scenario. After the interaction, participants returned to the survey and completed postchatbot questions assessing satisfaction, intention to use, and the Japanese version of the Health-ITUES.

Survey Items

The survey included the following items:

- Participant demographics: the demographics consisted of age, region of residence, educational background, occupation, and relationship status.
- Frequency of conversational AI use: items were rated on a 4-point scale ranging from "never" to "almost daily."
- Digital health literacy: this was measured using 15 items from the Japanese version of the Digital Health Literacy Instrument [35], excluding the subscales for operational and navigational skills.
- Postchatbot satisfaction: items were rated on a 5-point scale ranging from "very satisfied" to "not at all satisfied."
- Intention to use: this item was assessed with the question, "Would you consider using this chatbot if you were experiencing anxiety or concerns about STIs?" and rated on a 5-point scale ranging from "strongly agree" to "strongly disagree."
- Usability: usability was evaluated using the 20-item Japanese version of the Health-ITUES (Table 1).

XSL•FO
RenderX

Table . Health-ITUES[a] [27] and Japanese Health-ITUES.

Items in the Health-ITUES [27]	Items in the Japanese version of Health-ITUES
Impact	
1. I think mVIP[b] would be a positive addition for persons living with HIV.	1. I think this chatbot is helpful for young women who have concerns or symptoms related to sexually transmitted infections.
2. I think mVIP would improve the quality of life of persons living with HIV.	2. I think this chatbot will improve the quality of life of young women who have concerns or symptoms related to sexually transmitted infections.
3. mVIP is an important part of meeting my information needs related to symptom self-management.	3. This chatbot plays an important role in helping me get the information I need about concerns and symptoms related to sexually transmitted infections.
Perceived usefulness	
4. Using mVIP makes it easier to self-manage my HIV-related symptoms.	4. Using this chatbot makes it easier for me to get information about sexually transmitted infections and ease my concerns about them.
5. Using mVIP enables me to self-manage my HIV-related symptoms more quickly.	5. Using this chatbot allows me to get information about sexually transmitted infections faster and ease my concerns about them.
6. Using mVIP makes it more likely that I can self-manage my HIV-related symptoms.	6. Using this chatbot increases the likelihood of me getting information about sexually transmitted infections and easing my concerns about them.
7. Using mVIP is useful for self-management of HIV-related symptoms.	7. Using this chatbot is helpful in getting information about sexually transmitted infections and easing my concerns about them.
8. I think mVIP presents a more equitable process for self-management of HIV-related symptoms.	8. I think this chatbot equally helps everyone get information about sexually transmitted infections and ease their concerns about them.
9. I am satisfied with mVIP for self-management of HIV-related symptoms.	9. I am happy to use this chatbot to get information about sexually transmitted infections and ease my concerns about them.
10. I self-manage my HIV-related symptoms in a timely manner because of mVIP.	10. Thanks to this chatbot, I will be able to get information about sexually transmitted infections and ease my concerns about them at the right time.
11. Using mVIP increases my ability to self-manage my HIV-related symptoms.	11. Using this chatbot improves my ability to get information about sexually transmitted infections and ease my concerns about them.
12. I am able to self-manage my HIV-related symptoms whenever I use mVIP.	12. Using this chatbot allows me to get information about sexually transmitted infections and ease my concerns about them at any time.
Perceived ease of use	
13. I am comfortable with my ability to use mVIP.	13. I can feel comfortable using this chatbot.
14. Learning to operate mVIP is easy for me.	14. I can easily learn how to use this chatbot.
15. It is easy for me to become skillful at using mVIP.	15. It is easy for me to master using this chatbot.
16. I find mVIP easy to use.	16. I think this chatbot is easy to use.
17. I can always remember how to log on to and use mVIP.	17. I remember how to access this chatbot and how to use it.
User control	
18. mVIP gives error messages that clearly tell me how to fix problems.	18. This chatbot displays error messages that clearly indicate how to resolve any problems with it.
19. Whenever I make a mistake using mVIP, I can recover easily and quickly.	19. If I make a mistake on how to use this chatbot while using it, I can easily undo it.
20. The information (such as online help, on-screen messages, and other documentation) provided with mVIP is clear.	20. The information provided by this chatbot (online help, messages, etc) is easy to understand.

[a]Health-ITUES: Health Information Technology Usability Evaluation Scale.

[b]mVIP: HIV self-management app.

Statistical Analysis

We calculated the mean scores and SDs for the overall Japanese version of the Health-ITUES and its subscales (impact, perceived usefulness, ease of use, and user control). Ceiling and floor effects (acceptable if <15% at maximum and minimum scores) were examined. We assessed reliability through internal consistency, using the Cronbach α coefficient and composite reliability (CR). Convergent validity was evaluated based on the average variance extracted (AVE). To assess construct validity, a CFA was conducted according to the original 4-factor model. Model fit was evaluated using fit indices, including the

XSL•FO
RenderX

Goodness-of-Fit Index (GFI), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). Criterion-related validity was assessed by examining the correlation coefficients between Health-ITUES scores and conceptually related variables, namely user satisfaction and intention to use the chatbot. All statistical tests were conducted at a 5% significance level using IBM SPSS Statistics version 29.0 and AMOS version 31.0.

Ethical Considerations

This study was approved by the Research Ethics Committee for human participants at Teikyo University (approval numbers: Tei-Rin 24 - 114, dated January 30, 2024, and Tei-Rin 25 - 006, dated May 8, 2025). Participants provided informed consent via an embedded survey link. Participants received a token of appreciation, equivalent to 200 Japanese yen (1 Japanese yen=US $0.006 as of May 8, 2025) in points, for their participation in the study.

Results

Pilot Test

We conducted a pilot test with 6 female university students aged 18 to 22 years, recruited through convenience sampling. All participants were native Japanese speakers and nonmedical majors. They completed the Japanese version of the Health-ITUES and were subsequently interviewed regarding the clarity of language, ease of response, and confidence in their responses. One participant commented, “There were no particular difficulties in responding, but because the chatbot was accessed via a URL rather than through a login, it might be better to revise that wording.” Based on this feedback, we revised item 17 to clarify that the chatbot was accessed via a URL in this study.

Demographic Characteristics

Table 2 presents the demographic characteristics of the 301 participants. Their mean age was 26.3 (SD 2.1) years. Although more than half of the participants were company employees, the sample included a diverse range of occupations, such as students, homemakers, and part-time workers. The most common frequency of conversational AI use was “occasionally” (n=109, 36.2%), followed by “never,” “rarely,” and “almost daily,” each accounting for approximately 20% of responses. When “occasionally” and “almost daily” users are combined, approximately 60% (n=177) of participants reported using conversational AI to some extent, which is consistent with national data indicating that around 50% of young people in Japan use such technologies [36].

XSL•FO
RenderX

Table . Participant characteristics (N=301).

Characteristics	Values
Age (y), mean (SD)	26.36 (2.12)
Digital Health Literacy Instrument, mean (SD)	2.83 (0.44)
Region of residence, n (%)	
Hokkaido	9 (3.0)
Tohoku	14 (4.7)
Kanto	129 (42.9)
Chuubu	39 (13.0)
Kinki	62 (20.6)
Chuugoku/Shikoku	25 (8.3)
Kyushuu/Okinawa	23 (7.6)
Occupation, n (%)	
Company employee or government worker	181 (60.1)
Self-employed	4 (1.3)
Homemaker	33 (11.0)
Part-time worker or temporary worker	47 (15.6)
Student	9 (3.0)
Other	27 (9.0)
Educational background, n (%)	
Junior high school	7 (2.3)
High school	79 (26.2)
Vocational school	36 (12.0)
Junior college	14 (4.7)
College or university	153 (50.8)
Graduate school or higher	9 (3.0)
Other	3 (1.0)
Frequency of conversational AI use, n (%)	
Never	62 (20.6)
Rarely	62 (20.6)
Occasionally	109 (36.2)
Almost daily	68 (22.6)
Relationship status, n (%)	
Has a specific partner	241 (80.1)
Has a nonspecific partner	15 (5.0)
No partner	43 (14.3)
Other	2 (0.7)

The average score on the 15-item Digital Health Literacy Scale (DHLS) was 2.8 (SD 0.4). This result aligns closely with the findings from a study of Japanese university students, which reported an average score of 2.9 (SD 0.4) on the 21-item DHLS [37]. These findings indicate that the study sample is broadly representative of young Japanese individuals in terms of digital health literacy.

Overall Scores

The Health-ITUES consists of 20 items across 4 subscales (impact: 3 items, perceived usefulness: 9 items, ease of use: 5 items, and user control: 3 items), each rated on a 5-point Likert scale (1="strongly disagree" to 5="strongly agree"). Satisfaction was assessed on a 5-point scale (1="not satisfied at all" to 5="very satisfied"), along with intention to use (1="not at all willing to use" to 5="very willing to use"). The overall mean

XSL•FO
RenderX

score for the Japanese version of Health-ITUES was 3.4 (SD 0.9), whereas user satisfaction was 3.1 (SD 1.2) and intention to use was 3.2 (SD 1.2; Table 3). Among the subscales, user control showed slightly lower mean scores, although no substantial differences were observed. No ceiling or floor effects were detected.

Table . Mean scale and internal scale consistency scores (N=301).

Score	Mean (SD)	Cronbach α	CR[a]	AVE[b]
Health-ITUES[c] (overall)	3.427 (0.890)	0.976	—[d]	—
Health-ITUES impact	3.517 (1.012)	0.953	0.948	0.860
Health-ITUES usefulness	3.450 (0.959)	0.976	0.974	0.806
Health-ITUES ease of use	3.443 (0.980)	0.922	0.908	0.667
Health-ITUES control	3.241 (0.960)	0.849	0.830	0.624
User satisfaction	3.080 (1.169)	—	—	—
Intention to use	3.159 (1.209)	—	—	—

[a]CR: composite reliability.

[b]AVE: average variance extracted.

[c]Health-ITUES: Health Information Technology Usability Evaluation Scale.

[d]Not applicable.

Internal Consistency Reliability

The internal consistency and convergent validity were evaluated, with Cronbach α of more than 0.8, CR exceeding 0.70, and AVE exceeding 0.50. As shown in Table 3, Cronbach α coefficients for the subscales ranged from 0.85 to 0.98, indicating high reliability beyond the commonly accepted cutoff of 0.80. The impact and perceived usefulness subscales had particularly high α values (>0.95). The overall Cronbach α for all 20 items was 0.976, which may be attributed to the large number of items in the usefulness subscale and high intersubscale correlations. Descriptive statistics of the 20 items (mean, SD, skewness, kurtosis, and item-total correlations) are provided in Multimedia Appendix 1.

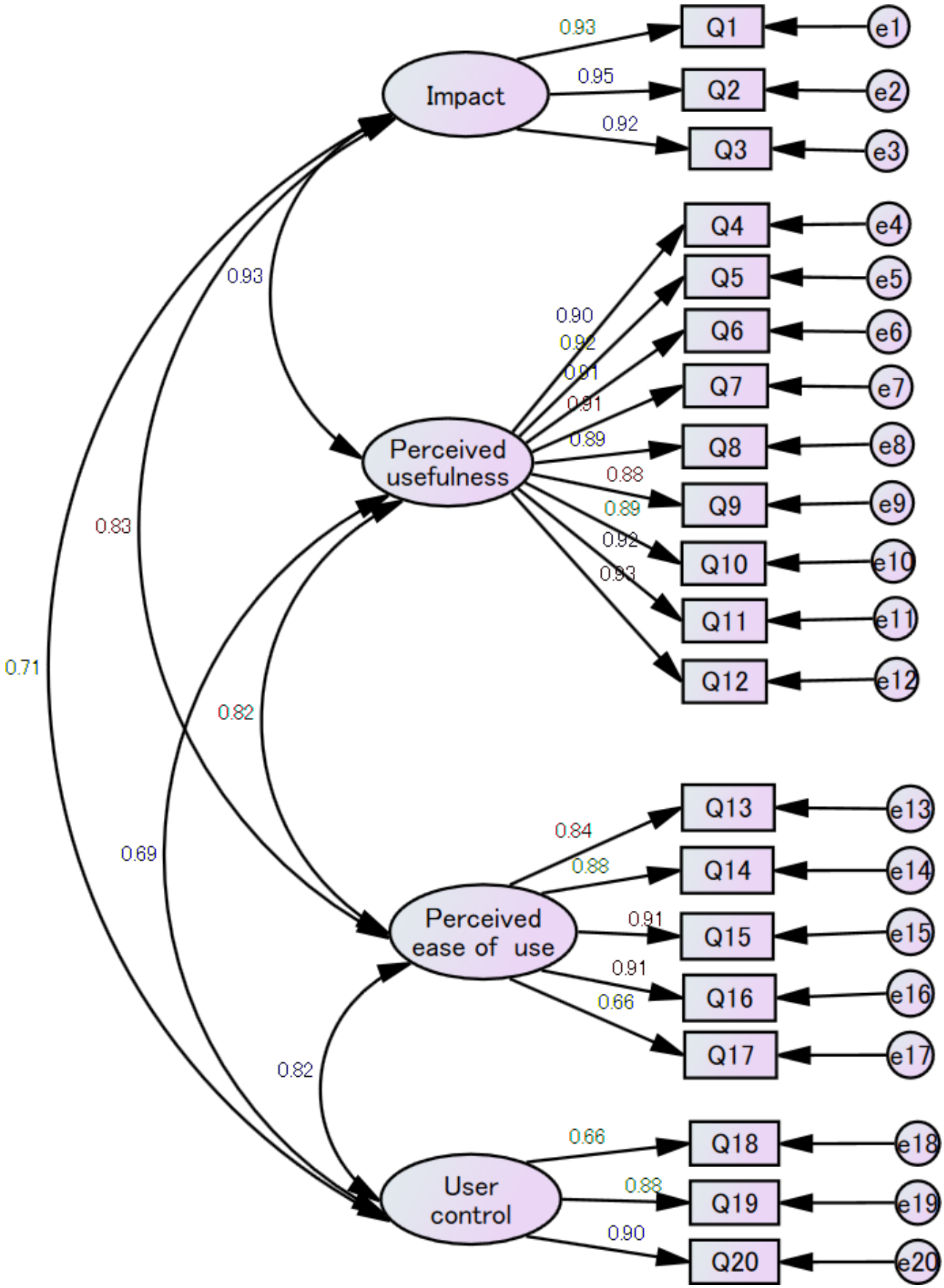
Construct Validity

As the conceptual validity of the original Health-ITUES had already been established, we performed a CFA using maximum likelihood estimation based on the original 4-factor structure. The analysis provided standardized factor loadings to assess the strength of the relationship between items and factors, along with model fit indices.

The path diagram is shown in Figure 1, and Table 4 presents the model fit indices. The CFA results indicated an adequate model fit (CFI=0.94, TLI=0.93, and SRMR=0.032). However, the GFI was relatively low (0.8), and the RMSEA was 0.100, representing the upper limit of the acceptable range [26].

XSL•FO
RenderX

Figure 1. Confirmatory factor analysis path diagram of the Japanese Health-ITUES: Health Information Technology Usability Evaluation Scale.



XSL•FO
RenderX

Table . Confirmatory factor analysis model fit indices for the Health-ITUES[a].

Fit index	CFI[b]	TLI[c]	RMSEA[d]	SRMR[e]	GFI[f]	Chi-square/*df*
Value	0.935	0.925	0.100	0.0319	0.799	660.92/164
Recommended threshold (desirable)	≥0.90 (≥0.95)	≥0.90 (≥0.95)	≤0.08 (≤0.05)	≤0.08 (≤0.05)	≥0.90	<3.00

[a]Health-ITUES: Health Information Technology Usability Evaluation Scale.

[b]CFI: Comparative Fit Index.

[c]TLI: Tucker-Lewis Index.

[d]RMSEA: root mean square error of approximation.

[e]SRMR: standardized root mean square residual.

[f]GFI: Goodness-of-Fit Index.

Intersubscale correlations ranged from 0.69 to 0.93, indicating strong relationships. In particular, perceived usefulness showed a higher correlation with impact, which is consistent with the findings of previous research [25,26,29].

We examined criterion-related validity by calculating the correlation coefficients between Health-ITUES scores and conceptually related variables, namely user satisfaction and intention to use. The correlations were 0.779 and 0.797, respectively, indicating a strong positive association (P<.001).

Discussion

Principal Findings

The emergence of generative AI has expanded the potential of mHealth. Chatbots are increasingly recognized as promising tools for supporting STI prevention and sexual health education among young people, and their role is expected to grow [11,38-41]. At the same time, usability—including user satisfaction and ease of use—remains critical for promoting continued engagement and facilitating behavior change [39,40].

The instrument selected for this study—the Health-ITUES—was developed as a user-centered evaluation scale based on the Health-ITUEM, which itself was derived from the widely used TAM and specifically adapted for health care technologies [23,28].

This study developed a Japanese version of the Health-ITUES and examined its reliability and validity using a chatbot currently under development as a consultation tool for young women with concerns about STIs. Because the original scale was designed to evaluate an app for adults living with HIV, we adapted the item wording to align with the chatbot's purpose and functions rather than conducting a direct translation.

In this study, we confirmed that the 4-factor structure of the Health-ITUES—impact, usefulness, ease of use, and user control—was largely retained in the context of a Japanese STI-related chatbot.

The translation procedures followed established guidelines, including forward translation by 2 experts in medicine and health communication, review by a multidisciplinary team of physicians and health communication experts, and a back-translation performed by a medical translation specialist. Cognitive debriefing and a final review by one of the authors, who validated the original Health-ITUES, ensured the cultural appropriateness and conceptual consistency of the translated version.

The customized Health-ITUES demonstrated adequate reliability and convergent validity for evaluating the usability of the STI-related scenario chatbot, and the main model-fit indices were satisfactory. In the CFA, although the RMSEA was slightly higher, the overall model fit was still considered acceptable because the other indices indicated a generally adequate level of fit [26]. Strong positive correlations between Health-ITUES scores and both user satisfaction and intention to use further supported criterion validity. These findings indicate that the Health-ITUES supports item-level customization while remaining suitable for multilingual translation and cultural adaptation.

The psychometric evaluation indicated very high internal consistency for the impact and usefulness subscales (Cronbach α>0.95). Although such values reflect strong reliability, they also suggest that some items may capture overlapping aspects of the same construct [42]. This interpretation is supported by the relatively strong interitem correlations observed within these subscales (r>0.85). Although these patterns suggest potential item redundancy and warrant future examination of item reduction or refinement, it is important to note that the Health-ITUES was intentionally designed to allow item-level customization while preserving conceptual comparability across contexts. The higher correlation between impact and perceived usefulness was also observed in previous studies [25,26,29]. Consistent with this, we retained all 20 items in the Japanese version without modifications or deletions.

From a theoretical perspective, this study reinforces the importance of usability as a multidimensional construct in digital health and demonstrates that the 4-factor structure of the Health-ITUES remains stable even when adapted to conversational AI contexts. Methodologically, this study highlights the feasibility of combining item-level customization with rigorous cross-cultural translation procedures while preserving conceptual equivalence. This approach provides a model for adapting usability instruments to rapidly evolving digital health technologies. To date, no other Japanese usability scale has demonstrated conceptual comparability while allowing flexible item-level customization tailored to specific use cases. The Japanese version of the Health-ITUES is expected to serve

XSL•FO
RenderX

as a valuable tool for assessing the usability of mHealth and conversational AI applications as their use expands.

The overall mean Health-ITUES score in this study was 3.4 (SD 0.9), which was lower than that reported in previous studies [24,25]. Several contextual factors may explain this difference. Our evaluation used a chatbot that is still in the development phase, which may have influenced perceptions of quality and usability. In contrast, earlier studies assessed mature, widely used applications developed for long-term engagement by specific patient populations, such as medication adherence support for individuals with HIV [27] or symptom management for dysmenorrhea [25], which may have already achieved high user acceptance. Moreover, this study involved a single-use interaction based on a hypothetical scenario rather than a real-world health context, which may have contributed to the lower scores.

At the same time, these findings highlight a practical advantage of the Health-ITUES: it can be used not only to evaluate relatively mature digital health tools but also to assess prototypes and guide iterative improvement during early-stage development. Future evaluations of the same chatbot using the Health-ITUES under conditions of continued use may help clarify how perceived usability evolves as interactions between users and conversational AI systems accumulate over time. Based on user feedback, we are currently developing an improved version of the chatbot by enhancing its frequently asked question responses and refining error messages. Incorporating the Health-ITUES from the early stages of development may support user-centered improvements and provide an objective basis for monitoring usability changes over time.

From a practical standpoint, the availability of a validated Japanese version of the Health-ITUES enables more systematic and standardized usability evaluations of mHealth and conversational AI tools in Japan.

A key strength of this study was its robust sample size and the diversity of participants in terms of region and occupation. The sample size of 301 participants in our study exceeded the commonly used 10 times per item rule [32,34]. The mean score on the 15-item DHLS (2.8, SD 0.4) aligned with that of previous Japanese studies [37], and the distribution of conversational AI use showed no marked skew, supporting the appropriateness of the sample for usability assessment in this context.

Limitations

This study had some limitations. First, participants were not actual patients or individuals currently experiencing health concerns; instead, we conducted the usability evaluation under a hypothetical scenario. Therefore, the results may not fully reflect real-world usage, and future research should assess its usability in clinical or personal health contexts. Second, the study population was limited to young women. Although we selected this demographic to align with the intended application of the mHealth chatbot, it is important to examine whether the Health-ITUES remains a valid usability assessment tool across other user groups in Japan, such as adult men, older adults, and individuals with chronic conditions, to enhance its generalizability.

Conclusions

This study presented the first Japanese version of the Health-ITUES and provides initial evidence of its reliability and validity in the context of an STI-related chatbot. To our knowledge, no other Japanese usability scale for mHealth combines validated psychometric properties with the flexibility to be customized according to specific use cases. By preserving conceptual equivalence while enabling standardized, user-centered usability evaluation, the Japanese version of the Health-ITUES addresses a critical gap in digital health assessment in Japan. As conversational AI and mHealth applications continue to expand, the Japanese version of the Health-ITUES may serve as a valuable tool for supporting the development, evaluation, and implementation of effective and user-centered digital health interventions.

Acknowledgments

Microsoft Copilot, an AI tool developed by Microsoft, was used to support language editing and translation of the manuscript. The authors reviewed and verified all AI-assisted outputs to ensure accuracy and appropriateness.

Funding

This study was supported by research funding from the Japan Health Academy under the 2024 Public Grant for Health Knowledge and Education and JSPS KAKENHI (grant 25K02857).

Data Availability

The dataset analyzed in this study is not publicly available due to ethical considerations, but it can be obtained from the corresponding author upon reasonable request.

Authors' Contributions

HI conceived and supervised the study. TH conducted the study, collected and analyzed the data, interpreted the findings, and drafted the manuscript. YF contributed to the study design and provided guidance throughout the project. KT and KH contributed to the development of the chatbot and assisted with data collection. AY oversaw the overall chatbot development and provided

XSL•FO
RenderX

advice on the analysis and interpretation of the study results. RS, the original developer of the scale, confirmed the construct validity and provided advice on the discussion and wording of the manuscript. All authors reviewed, revised, and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The descriptive statistics of the 20 items (mean, SD, skewness, kurtosis, and item-total correlations).
[DOCX File, 16 KB - humanfactors_v13i1e90483_app1.docx]

References

1. Summary of the 2024 survey on telecommunications usage [report in japanese]. : Ministry of Internal Affairs and Communications; 2025 URL: https://www.soumu.go.jp/johotsusintokei/statistics/data/250530_1.pdf [accessed 2025-11-05]
2. Free C, Phillips G, Watson L, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLOS Med 2013;10(1):e1001363. [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
3. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using health and well-being apps for behavior change: a systematic search and rating of apps. JMIR mHealth uHealth 2019 Jul 4;7(7):e11926. [doi: 10.2196/11926] [Medline: 31274112]
4. Nilsen W, Kumar S, Shar A, et al. Advancing the science of mHealth. J Health Commun 2012;17 Suppl 1:5-10. [doi: 10.1080/10810730.2012.677394] [Medline: 22548593]
5. Modave F, Bian J, Leavitt T, Bromwell J, Harris Iii C, Vincent H. Low quality of free coaching apps with respect to the American College of Sports Medicine guidelines: a review of current mobile apps. JMIR mHealth uHealth 2015 Jul 24;3(3):e77. [doi: 10.2196/mhealth.4669] [Medline: 26209109]
6. Recommendations on digital interventions for health system strengthening—evidence and recommendations. : World Health Organization; 2019 URL: https://iris.who.int/server/api/core/bitstreams/0d2bf963-9eaa-4076-be82-9c54190093fc/content [accessed 2025-11-05]
7. Giebel GD, Speckemeier C, Abels C, et al. Problems and barriers related to the use of digital health applications: protocol for a scoping review. JMIR Res Protoc 2022 Apr 21;11(4):e32702. [doi: 10.2196/32702] [Medline: 35451979]
8. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence-based chatbots for promoting health behavioral changes: systematic review. J Med Internet Res 2023 Feb 24;25:e40789. [doi: 10.2196/40789] [Medline: 36826990]
9. Islam MR, Urmi TJ, Mosharrafa RA, Rahman MS, Kadir MF. Role of ChatGPT in health science and research: a correspondence addressing potential application. Health Sci Rep 2023 Oct;6(10):e1625. [doi: 10.1002/hsr2.1625] [Medline: 37841943]
10. Latt PM, Aung ET, Htaik K, et al. Evaluation of artificial intelligence (AI) chatbots for providing sexual health information: a consensus study using real-world clinical queries. BMC Public Health 2025 May 15;25(1):1788. [doi: 10.1186/s12889-025-22933-8] [Medline: 40375254]
11. Bailey J, Mann S, Wayal S, et al. Sexual health promotion for young people delivered via digital media: a scoping review. Public Health Res 2015;3(13):1-120. [doi: 10.3310/phr03130] [Medline: 26583166]
12. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019;5:2055207619871808. [doi: 10.1177/2055207619871808] [Medline: 31467682]
13. Siddals S, Torous J, Coxon A. “It happened to be the perfect thing”: experiences of generative AI chatbots for mental health. Npj Ment Health Res 2024 Oct 27;3(1):48. [doi: 10.1038/s44184-024-00097-4] [Medline: 39465310]
14. Li J, Dada A, Puladi B, Kleesiek J, Egger J. ChatGPT in healthcare: a taxonomy and systematic review. Comput Methods Programs Biomed 2024 Mar;245:108013. [doi: 10.1016/j.cmpb.2024.108013] [Medline: 38262126]
15. Su Z, Figueiredo MC, Jo J, Zheng K, Chen Y. Analyzing description, user understanding and expectations of AI in mobile health applications. AMIA Annu Symp Proc 2021;2020:1170-1179. [Medline: 33936493]
16. Bean AM, Payne R, Parsons G, et al. Clinical knowledge in LLMs does not translate to human interactions. arXiv. Preprint posted online on Apr 26, 2025. [doi: 10.48550/arXiv.2504.18919]
17. Muro-Culebras A, Escriche-Escuder A, Martin-Martin J, et al. Tools for evaluating the content, efficacy, and usability of mobile health apps according to the consensus-based standards for the selection of health measurement instruments: systematic review. JMIR mHealth uHealth 2021 Dec 1;9(12):e15433. [doi: 10.2196/15433] [Medline: 34855618]
18. Brooke J. SUS: a retrospective. J Usabil Stud 2013;8(2):29-40 [FREE Full text]
19. Lewis JR. Psychometric evaluation of the PSSUQ using data from five years of usability studies. International Journal of Human–Computer Interaction 2002;14(3-4):463-488. [doi: 10.1080/10447318.2002.9669130]
20. de Dios López A, Real J, Meza C, et al. Validation of a questionnaire to assess the usability of and user experience with mobile health applications. Healthcare (Basel) 2024 Nov 21;12(23):2328. [doi: 10.3390/healthcare12232328] [Medline: 39684949]

XSL•FO
RenderX

21. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR mHealth uHealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
22. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR mHealth uHealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849] [Medline: 27287964]
23. Yen PY, Wantland D, Bakken S. Development of a customizable Health IT Usability Evaluation Scale. AMIA Annu Symp Proc 2010 Nov 13;2010:917-921. [Medline: 21347112]
24. Loh KP, Liu J, Ganzhorn S, Sanabria G, Schnall R. Establishing a usability cut-point for the Health Information Technology Usability Evaluation Scale (Health-ITUES). Int J Med Inform 2022 Apr;160:104713. [doi: 10.1016/j.ijmedinf.2022.104713] [Medline: 35144102]
25. Lee J, Schnall R. Validity and reliability of the Korean version of the Health Information Technology Usability Evaluation Scale: psychometric evaluation. JMIR Med Inform 2022 Jan 24;10(1):e28621. [doi: 10.2196/28621] [Medline: 35072630]
26. Guo R, Zheng Z, Yang F, Wu Y. Translation, cross-cultural adaptation, and psychometric validation of the Health Information Technology Usability Evaluation Scale in China: instrument validation study. J Med Internet Res 2025 May 2;27:e67948. [doi: 10.2196/67948] [Medline: 40315454]
27. Schnall R, Cho H, Liu J. Health Information Technology Usability Evaluation Scale (Health-ITUES) for usability assessment of mobile health technology: validation study. JMIR mHealth uHealth 2018 Jan 5;6(1):e4. [doi: 10.2196/mhealth.8851] [Medline: 29305343]
28. Brown W, Yen PY, Rojas M, Schnall R. Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013 Dec;46(6):1080-1087. [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
29. Yen PY, Sousa KH, Bakken S. Examining construct and predictive validity of the Health-IT Usability Evaluation Scale: confirmatory factor analysis and structural equation modeling results. J Am Med Inform Assoc 2014 Oct;21(e2):e241-e248. [doi: 10.1136/amiajnl-2013-001811] [Medline: 24567081]
30. Naoko I. Improving the methodological quality of the translational cultural adaptation process for patient-reported outcome measures [Article in Japanese]. Jpn J Behav Ther Res 2015;41(2):117-125. [doi: 10.24468/jjbt.41.2_117]
31. Miyaji T, Iioka Y, Kuroda Y, et al. Japanese translation and linguistic validation of the US National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). J Patient Rep Outcomes 2017;1(1):8. [doi: 10.1186/s41687-017-0012-7] [Medline: 29757296]
32. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. J Eval Clin Pract 2011 Apr;17(2):268-274. [doi: 10.1111/j.1365-2753.2010.01434.x] [Medline: 20874835]
33. Syphilis [Article in Japanese]. Ministry of Health, Labour and Welfare (MHLW). URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/kenkou/kekkaku-kansenshou/seikansenshou/syphilis.html [accessed 2026-04-09]
34. Anthoine E, Moret L, Regnault A, Sébille V, Hardouin JB. Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. Health Qual Life Outcomes 2014 Dec 9;12:176. [doi: 10.1186/s12955-014-0176-2] [Medline: 25492701]
35. Miyawaki R, Kato M, Kawamura Y, Ishikawa H, Oka K. Developing a Japanese version of the Digital Health Literacy Instrument [Article in Japanese]. Nihon Koshu Eisei Zasshi 2024 Jan 26;71(1):3-14. [doi: 10.11236/jph.23-021] [Medline: 37673595]
36. NRC daily tracking: survey on generative AI, September 2025 [Report in Japanese]. : Nippon Research Center (NRC); 2025 URL: https://www.nrc.co.jp/report/251009.html [accessed 2025-10-23]
37. Ishikawa H, Miyawaki R, Kato M, Muilenburg JL, Tomar YA, Kawamura Y. Digital health literacy and trust in health information sources: a comparative study of university students in Japan, the United States, and India. SSM Popul Health 2025;31:101844. [doi: 10.1016/j.ssmph.2025.101844] [Medline: 40746760]
38. Crutzen R, Peters GJY, Portugal SD, Fisser EM, Grolleman JJ. An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. J Adolesc Health 2011 May;48(5):514-519. [doi: 10.1016/j.jadohealth.2010.09.002] [Medline: 21501812]
39. Balaji D, He L, Giani S, Bosse T, Wiers R, de Bruijn GJ. Effectiveness and acceptability of conversational agents for sexual health promotion: a systematic review and meta-analysis. Sex Health 2022 Oct;19(5):391-405. [doi: 10.1071/SH22016] [Medline: 35863761]
40. Mills R, Mangone ER, Lesh N, Mohan D, Baraitser P. Chatbots to improve sexual and reproductive health: realist synthesis. J Med Internet Res 2023 Aug 9;25:e46761. [doi: 10.2196/46761] [Medline: 37556194]
41. Sanabria G, Greene KY, Tran JT, et al. "A great way to start the conversation": evidence for the use of an adolescent mental health chatbot navigator for youth at risk of HIV and other STIs. J Technol Behav Sci 2023 May 11;8:382-391. [doi: 10.1007/s41347-023-00315-4] [Medline: 37362063]
42. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011 Jun 27;2:53-55. [doi: 10.5116/ijme.4dfb.8dfd] [Medline: 28029643]

XSL•FO
RenderX

Abbreviations

AVE: average variance extracted
CFA: confirmatory factor analysis
CFI: Comparative Fit Index
COSMIN: Consensus-Based Standards for the Selection of Health Measurement Instruments
CR: composite reliability
DHLS: Digital Health Literacy Scale
GFI: Goodness-of-Fit Index
Health-ITUEM: Health IT Usability Evaluation Model
Health-ITUES: Health Information Technology Usability Evaluation Scale
MARS: Mobile App Rating Scale
MAUQ: mHealth App Usability Questionnaire
mHealth: mobile health
PSSUQ: Post-Study System Usabiity Questionnaire
RMSEA: root mean square error of approximation
SRMR: standardized root mean square residual
STI: sexually transmitted infection
SUS: System Usability Scale
TAM: technology acceptance model
TLI: Tucker-Lewis Index
WHO: World Health Organization

Edited by A Kushniruk; submitted 01.Jan.2026; peer-reviewed by L Salahuddin; revised version received 03.Jun.2026; accepted 23.Jun.2026; published 23.Jul.2026.

Please cite as:
Hato T, Ishikawa H, Todo K, Harada K, Yoshikawa A, Fukuda Y, Schnall R
Japanese Health Information Technology Usability Evaluation Scale for Sexually Transmitted Infection–Related Chatbots: Development and Psychometric Validation Study
JMIR Hum Factors 2026;13:e90483
URL: https://humanfactors.jmir.org/2026/1/e90483
doi:10.2196/90483

© Tomoko Hato, Hirono Ishikawa, Kense Todo, Keisuke Harada, Atsushi Yoshikawa, Yoshiharu Fukuda, Rebecca Schnall. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 23.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Monitoring Health Status: Development and Preliminary Validation of a Personal Health Index Using the International Classification of Functioning, Disability and Health

Ilkka Rautiainen[1], PhD; Lauri Parviainen[2], MD; Veera Jakoaho[2], MSc; Sami Äyrämö[1,3], PhD; Jukka-Pekka Kauppi[1], PhD

[1]Faculty of Information Technology, University of Jyväskylä, PO Box 35, Jyväskylä, Central Finland, Finland

[2]David Health Solutions Ltd., Helsinki, Finland

[3]Wellbeing Services County of Central Finland, Jyväskylä, Finland

Corresponding Author:
Ilkka Rautiainen, PhD
Faculty of Information Technology, University of Jyväskylä, PO Box 35, Jyväskylä, Central Finland, Finland

Abstract

Background: Effective health monitoring is essential for personalized care and comprehensive health assessment. Personal health indices and profiles offer a concise summary of an individual's overall health, supporting both clinical decision-making and self-management. However, global standardization remains challenging due to diverse practices and data formats across countries.

Objective: This study aimed to present a novel model for computing a personal health index and health profile using the International Classification of Functioning, Disability and Health (ICF) framework. The model was designed to handle incomplete and heterogeneous datasets and aimed to provide standardized, interpretable health metrics.

Methods: We developed a recursive algorithm that calculates the health index based on the hierarchical structure of the ICF, using all available measurements. The model incorporates time decay and linkage reliability to weight input data. Preliminary validation was conducted on data from 505 individuals, using statistical correlation analyses with self-assessed health measures (EuroQol Visual Analogue Scale and pain ratings), and a sensitivity analysis was performed to assess model robustness.

Results: The computed health index showed moderate positive correlations with EuroQol Visual Analogue Scale scores (all $P<.001$) and negative correlations with maximum pain trajectories, supporting its validity. Sensitivity analysis confirmed predictable behavior in response to input changes, and the model demonstrated resilience to missing data.

Conclusions: The proposed model offers a flexible and scientifically grounded approach to computing personal health indices and profiles within the ICF framework. It enables the integration of diverse health data sources and supports the visual representation for clinical and personal use. This model has potential applications in health monitoring, rehabilitation planning, and machine learning–based health informatics.

(JMIR Hum Factors 2026;13:e84802) doi:10.2196/84802

KEYWORDS

personal health index; International Classification of Functioning, Disability and Health; ICF; health informatics; data integration; health data standardization; longitudinal health monitoring

Introduction

Background

Effectively describing and monitoring a person's health is crucial for tailoring appropriate treatments. Comprehensive information about a person's health enables targeted interventions and provides metrics that emphasize overall health status rather than isolated symptoms.

A personal health index aims to condense a person's overall health into a single numerical value [1]. While the index seeks to summarize health in a singular number, a health profile represents health status through a set of scores [1]. These indices or profiles serve diverse purposes, offering a quick overview of an individual's health status, beneficial for both health care professionals and individuals themselves. This systematic and comprehensible monitoring approach facilitates a broad perspective on health [1].

Given the multifaceted nature of health, summarizing it into a single value or profile requires a framework that can accommodate diverse and interrelated factors. A singular value from a health index can offer a quick overview of a person's

XSL•FO
RenderX

overall health, which is particularly useful in both clinical and self-monitoring contexts [1]. Additionally, exploring various aspects of health offered by a health profile is essential for a more detailed understanding of an individual's health.

Capturing an individual's overall health concisely can be challenging due to the multifaceted factors influencing health and its changes. Prevention, evaluation, treatment planning, and rehabilitation, especially in cases such as low back pain, necessitate consideration of psychosocial and social aspects [2]. It is, therefore, necessary to systematically describe all relevant aspects affecting a person's health in a structured and standardized manner, presenting this information in an easily interpretable format.

Global standardization of health status assessment procedures for health index construction is challenging due to differing standards and practices across countries. Variations in treatment procedures, driven by diverse practices among countries, clinics, and therapists, hinder straightforward concatenation of datasets from separate clinics, especially with the use of different health questionnaires and languages.

Existing health indices often rely on fixed questionnaires or population-level parameters, limiting their adaptability and comparability across contexts. To address these limitations, a universally accepted framework is needed—one that can integrate heterogeneous data and support individualized assessment. In this study, we propose using the International Classification of Functioning, Disability and Health (ICF) [3], developed by the World Health Organization (WHO), as the foundation for constructing a personal health index. The primary goal of the ICF is to establish a standardized language and framework for describing health and health-related states [4]. It serves as a complement to the *International Classification of Diseases* (ICD) diagnosis classification system [4].

The ICF integrates 2 conceptual paradigms of disability—the medical and social paradigms. In the medical paradigm, disability is viewed as a health condition requiring medical care, often stemming from diseases, trauma, or other health conditions. On the other hand, the social paradigm considers disability as a socially created problem necessitating a political response. The ICF, by merging these paradigms, adopts a *biopsychosocial* approach, synthesizing biological, individual, and social aspects of health [4,5].

By leveraging the ICF's comprehensive structure, our model aims to translate this biopsychosocial perspective into a quantifiable health index that can be used in both clinical and research settings. This model has the potential to shift from the current disease-based care model toward a *healthy aging* approach. The WHO's concept of intrinsic capacity reflects a similar emphasis on monitoring functional trajectories across key domains to enable proactive, personalized interventions [6,7]. This approach aims to support a proactive, functioning-based way of tracking individual trajectories [3]. However, despite the clear advantages of the ICF, its practical implementation in clinical practice has proven challenging in various studies [8-11]. We argue that the health index framework introduced in this study also holds promise for simplifying the use of the ICF.

Unlike data-driven models that rely on statistical learning from large datasets, such as machine learning (ML) models that require training and may be prone to overfitting, the proposed health index is computed using a knowledge-driven, deterministic algorithm. This approach is grounded in the hierarchical structure of the ICF and applies predefined rules to aggregate available ICF-coded measurements. It operates independently of any training data or parameter estimation, ensuring consistency, interpretability, and robustness across diverse datasets. As a result, the model can be applied to any ICF-compatible data without modification, making it suitable for scalable and standardized health assessment.

Using the ICF framework involves converting original measurements to new variables in *ICF space* using accepted linking procedures. Throughout the paper, these variables are referred to as ICF codes according to ICF terminology. This approach facilitates the standardization of potentially heterogeneous datasets from different individuals into the same data space.

The health index is then recursively computed from the ICF tree structure using available measurements. Importantly, this computation does not require measurements from every node of the ICF tree, ensuring robustness to missing values. Additionally, if measurements do not cover all relevant health aspects, these gaps can be presented with the health index and health profile, and they can optimally be addressed in future data collection. Once the overall health index value is computed, it becomes possible to explore a health profile—that is, values for each ICF code separately. This approach provides insights into specific sectors of functioning, disabilities, and contextual factors affecting health. By adopting this comprehensive approach, we obtain a nuanced picture of an individual's health, highlighting specific areas that may need improvement. By leveraging the health index, health profile, and the ICF, we can construct a clinical tool that facilitates cross-country data comparison and enables the development of a universal health metric.

In the context of ML, the proposed health index provides a framework for integrating datasets from diverse sources, thereby streamlining data preparation. The importance of maintaining uniform variables across all individuals cannot be overstated, as it is crucial for the development of robust ML models. A lack of standardization can result in smaller subsets, which may compromise the effectiveness of predictive models. Furthermore, the health index can serve as an optimization target in ML, facilitating the use of standardized datasets for predicting individual health outcomes. This could, for example, assist in the selection of personalized rehabilitation pathways. To our knowledge, this study marks the first usage of the ICF framework in forming a personal health index. In the following sections, we present the theoretical foundation, algorithmic implementation, validation results, and implications for clinical practice.

Prior Work

Numerous health indices have been proposed in the literature. This section provides an overview of existing health indices and profiles, with an emphasis on research conducted in the

XSL•FO
RenderX

2000s. A literature search using Google Scholar was conducted in 2023 to identify models that explicitly used the term "health index" within this time frame. The focus was on individual-level health assessment studies, ensuring that the review reflects approaches relevant to personal health rather than population-level metrics.

One of the most straightforward methods for computing a personal health index involves the direct summation of responses to health-related questions. Kubik et al [12] constructed their personal health index by asking 6 questions on a Likert-type scale, ranging from 1 (strongly disagree) to 4 (strongly agree). These questions encompassed an individual's perception of overall health, dietary patterns, and level of physical activity. The resulting health index score ranged from a minimum of 6 to a maximum of 24. Similarly, Frank et al [13] used a comparable summation method by querying participants about their smoking, drinking, exercise routines, and dietary habits to formulate a personal health index. Another simple approach, as demonstrated by Gallup [14], involved a personal health index derived from 5 yes-or-no questions concerning self-perceptions of health. These questions addressed general health problems, feelings of well-restedness, presence of physical discomfort, experiences of worry, and feelings of sadness. The resulting index was computed as the mean of all valid responses multiplied by 100, making the index range from 0 to 100.

Meijer et al [15] introduced a health index with an aim to be internationally comparable. Experts identified 25 variables measuring health and functional ability for inclusion in the index. The included variables consisted of reports of limitations with 10 mobility-related activities, arm and fine motor functions (such as walking, sitting, and reaching), 6 severe limitations with activities of daily living (such as dressing, bathing, and eating), and 7 less disabling problems in instrumental activities of daily living (such as preparing hot meals, making phone calls, and taking medication). Additionally, self-reported health and grip strength were incorporated into the index. The computation of this health index involved fitting the data with a special case of the LISCOMP model, integrating factor analysis and regression models. The health indices were reported to be comparable across different countries but were gender-specific. The primary focus of the study was centered more on examining the population-level health indices across various countries rather than discussing individual-level health indices.

Yi et al [16] developed 4 distinct health indices—namely, cardiovascular, stress, obesity, and management—derived from responses to specific questionnaires and objective measurements. These indices were constructed using weights assigned to each questionnaire item by experts in the field. The cardiovascular and stress indices were computed from a balanced combination of questionnaire responses and objective measurements, whereas the obesity index relied solely on measured data. The management index was formulated from questionnaire responses focusing on lifestyle and dietary habits. To consolidate these individual indices into an integrated health measure, the researchers weighted each element by the inverse of its corresponding standard deviation and summed the values together to create a final integrated health index.

Kohn [17] used multiple correspondence analysis (MCA) to construct a health index. MCA has similarities with principal component analysis (PCA) but is more suitable for discrete variables. The first dimension of MCA was reported to capture 80.4% of the principal inertia, corresponding to the variance in PCA. To facilitate interpretation, the analysis primarily focused on this first dimension, which represented the health index in this context. The range of observed continuous values for the health index was from 1 to 9, with higher values denoting better health. The selected domain of variables encompassed self-assessed health, a 36-point index evaluating mental and emotional health, reported health problems, disability status, a categorical number of accidents, and current smoking status. The suggested index allowed some flexibility for the index user, offering choices such as determining which domains to include when applying MCA for the index creation process.

Poterba et al [18] used PCA in creating the health index. The PCA for this purpose has been used since at least the 1980s [19,20]. In their approach, they used the first principal component as their health index. To aid the presentation, the "raw" health scores were converted into percentile scores for each respondent at different ages. The PCA was computed based on responses to 27 health-related questions, including activities of daily living (eg, difficulties in walking, lifting, or sitting), medical history (eg, experiences with stroke, diabetes, or cancer), self-reported health, BMI, and other relevant information.

Kim et al [21] suggested a toilet-based system for collecting data, such as pulse, blood pressure, oxygen saturation, BMI, electrocardiogram readings, and other vital signals. These data were used to construct 5 separate health indices—namely, heart, blood, fitness, muscle, and mental health index. Each index was derived by summing the respective normalized values together. Total health index represented the combination of all the 5 health indices, although its computation was not described in detail.

Chen et al [22] developed MyPHI, a method for constructing a personal health index. Using soft-label optimization, they created a unique prediction model for each disease category, handling infrequent and sparse data, and prioritizing recent health records. The method outperformed linear support vector machine and logistic regression models in their comparative analysis. Data from geriatric medical examinations, including patient profiling, laboratory tests, physical and external examinations, and mental health questionnaires, were used. Disease categories, derived from a cause of death dataset based on the ICD diagnosis classification, included lung, heart, cerebrovascular diseases, diabetes, stomach, colon, liver, pancreas, septicemia, and hypertension. The method generated a vector of scores reflecting personal health risks across disease categories. According to McDowell's [1] definition, the output could be described as a health profile or a disease index, given its focus on predicting disease-specific risks.

Lai et al [23] introduced a novel personal health index based on the *technique for order preference by similarity to ideal solution*, an established ranking method from 1981, coupled with an independent entropy weighting approach. Additionally, the study used tensor decomposition to address problems

XSL•FO
RenderX

stemming from missing data in health examination records. The authors highlighted the necessity for a more holistic approach to health assessment, acknowledging the multitude of variables collected during health examinations and difficulty in evaluating individual health status through only a small subset of these variables. They, however, selected 9 key health indicators in collaboration with domain experts, forming the basis of their health index. These indicators included systolic and diastolic blood pressure; BMI; total, high-density lipoprotein, and low-density lipoprotein cholesterol levels; fasting blood glucose; triglyceride levels; and thyrotropin. Each indicator held almost equal weight in the final computation of the health index.

The main characteristics of previous health indices and our index are summarized in Table 1. In the table, "*Individual adaptability*" refers to whether the model adapts to individual-specific data rather than relying solely on population-level parameters or fixed weights. A "*Yes*" indicates that the model can compute a personalized health index based on the available data for each individual. In some models (eg, MyPHI), this requires retraining or reconfiguring the model for each disease category or data type.

Table . Summary comparison of previous health indices and our index.

Study	Method	Data type	Individual adaptability	Supports missing data	Independent of training data
Kubik et al [12]	Summation of lifestyle questions	Questionnaire	No	No	Yes
Frank et al [13]	Summation of lifestyle questions	Questionnaire	No	No	Yes
Gallup [14]	Mean of yes or no questions	Questionnaire	No	No	Yes
Meijer et al [15]	LISCOMP model (factor analysis+regression)	Questionnaire+measurements	No	Yes	No
Yi et al [16]	Expert-weighted composite indices	Questionnaire+measurements	No	No	Yes
Kohn [17]	Multiple correspondence analysis	Questionnaire	No	Limited	No
Poterba et al [18]	Principal component analysis	Questionnaire	No	No	No
Kim et al [21]	Summation of normalized sensor data	Sensor measurements	No	No	No
Chen et al [22]	Machine learning with soft-label optimization	Clinical records+Questionnaire	No	Yes	No
Lai et al [23]	TOPSIS[a]+entropy weighting+tensor decomposition	Health examination records	No	Yes	No
This study	Recursive computation from ICF[b] structure	Any ICF-compatible data	Yes	Yes	Yes

[a]TOPSIS: technique for order preference by similarity to ideal solution.

[b]ICF: International Classification of Functioning, Disability and Health.

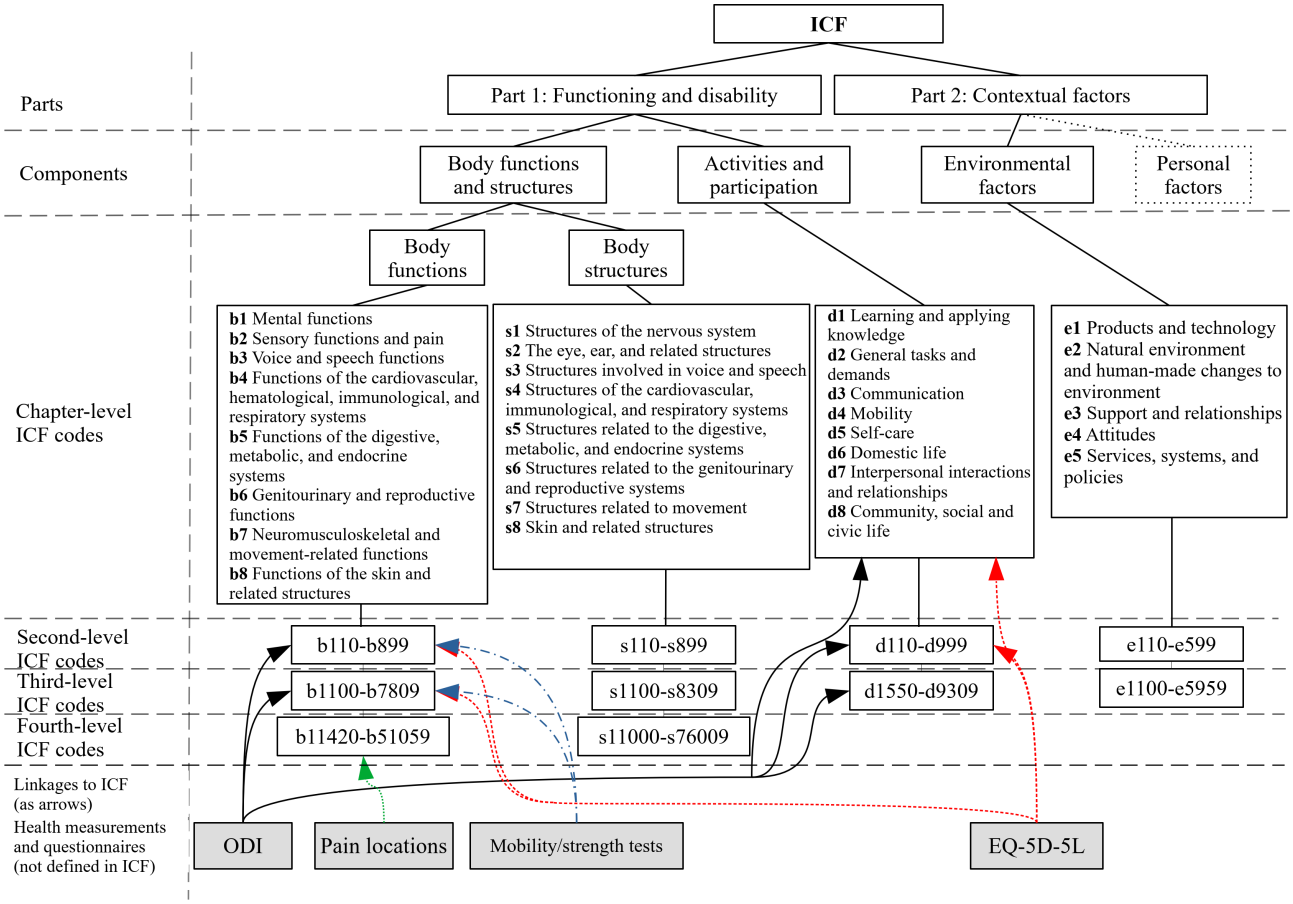
Despite the variety of existing health indices, several limitations persist: lack of standardization and limited adaptability to heterogeneous data. Existing indices require retraining and/or manual configuration to accommodate new data types or disease categories. In contrast, the ICF-based personal health index proposed in this study is designed to automatically adapt to the available data for each individual without modifying the model, making it both flexible and standardized, and well-suited for scalable clinical use.

International Classification of Functioning, Disability and Health

Figure 1 provides an overview of the ICF structure. The 4 health measurements or questionnaires used and link to the ICF during the development and preliminary validation are depicted at the bottom of the figure, although they are not part of the ICF itself. Further discussion on these elements is presented in the section "Mapping data to ICF codes."

XSL•FO
RenderX

Figure 1. The hierarchical arrangement of the International Classification of Functioning, Disability and Health (ICF), compiled from data available in [3], along with the health measurements and questionnaires used during validation and their respective connections to the ICF.



The ICF classifies health and health-related problems into 2 main parts: *functioning and disability* and *contextual factors*. These are further divided into 4 main *components: body functions* (*b*) (physiological and psychological functions of body systems), *body structures* (*s*) (anatomical parts of the body), *activities and participation* (*d*) (execution of tasks and involvement in life situations), and *environmental factors* (*e*) (physical, social, and attitudinal environment). Although *personal factors* (background of an individual's life) are part of the structure, they cannot currently be classified using the ICF [4].

The main components of the ICF encompass over 1400 subcategories distributed across 4 hierarchical levels. These categories exhibit a nested structure, where broader categories incorporate more detailed subcategories under the parent category. For instance, the ICF code *The eye, ear and related structures* (*s2*) in the *body structures* component (chapter 2) encompasses separate categories detailing structures such as the eye socket, eyeball, and surrounding areas [4].

The *chapter-level* ICF code is denoted by appending a single-digit number after the component *b*, *s*, *d*, or *e*, such as *b2* referring to the ICF code *sensory functions and pain*. The *second-level* ICF code uses a 3-digit number (eg, *b280* for *sensation of pain*). Moving further, the *third-level* ICF code consists of 4 numbers (eg, *b2801* for *pain in body part*), and the *fourth-level* code expands to 5 numbers (eg, *b28013* for *pain in back*) [4,5].

For classification completeness, an ICF code is accompanied by one or more *qualifiers*. These single-digit numbers signify attributes such as the magnitude of the level of health or the severity of the problem. A dot (.) serves as a separator between the ICF code and the qualifier(s). For instance, a complete ICF code with a qualifier could be *b280.1*, indicating a mild or slight *sensation of pain* [4].

All possible ICF codes support the first qualifier, although its application varies slightly among different main constructs. In *body functions* and *body structures* components, the first qualifier serves as a *generic qualifier*, signifying the extent or magnitude of an impairment. The generic qualifier can have values 0 (*no problem*), 1 (*mild problem*), 2 (*moderate problem*), 3 (*severe problem*), or 4 (*complete problem*). In the *activities and participation* component, the first generic qualifier indicates *performance*, representing a problem in the person's current environment [4,5].

For *environmental factors*, the generic qualifier can be either a *barrier* or a *facilitator*, allowing positive contributors to be recorded. The negative scale is used similarly to the constructs mentioned earlier for barriers, whereas facilitators are indicated using a plus sign (+) as an ICF code or qualifier separator instead of a dot. For instance, ICF code *e145+2* indicates that products for education are a moderate facilitator, distinct from *e145.2*, which signifies that the products of education are a moderate barrier [4,5]. It is worth noting that facilitators are not part of our current health index computation model.

XSL•FO
RenderX

In addition to the first qualifier, additional qualifiers can be defined for *body structures* and *activities and participation*. For example, the nature of impairment can be described using the second qualifier in *body structures* (eg, *s7300.32* indicates a partial absence of the upper extremity) [4,5]. However, as our current computation model uses only the first qualifier, these additional qualifiers are not discussed further in this paper. Throughout the remainder of this paper, the term *qualifier* will exclusively refer to the first qualifier. In our proposed model, we introduce a minor addition to the ICF, which requires 2 attributes for the qualifier, specifically *linkage reliability* and *time elapsed* from the latest valid measurement. These attributes are discussed in the following sections.

Methods

Overview

In this section, we present the materials and methods used for computing the health index. Initially, we introduce the structure of the model used in the computation, outlining its fundamental components. Subsequently, we describe the overarching process involved in deriving the health index from the model. Following this, we delve into a more comprehensive explanation of the model elements, providing further insights into the ICF code–level computations.

Materials

Data Collection and Linkage to ICF

For the initial evaluation of the health index, we used data collected from a singular David Health Solutions clinic, located in Helsinki, Finland, spanning from 2013 to 2019. David Health Solutions is a private health care organization with a global network of clinics, specializing in rehabilitation services. The dataset consisted of 505 individuals (mean age 48, SD 18 y; minimum 12 y, maximum 87 y; women: n=259; and men: n=246) undergoing rehabilitation treatment for diverse problems such as back, neck, hip, knee, shoulder, general health, and other unspecified reasons. Our original dataset included the following questionnaire and measurement datasets:

- Oswestry Disability Index [24]
- A generic health questionnaire EQ-5D-5L (EuroQol 5-dimensional, 5-level) [25]
- Mobility and maximal isometric strength tests using various spine concept rehabilitation machines
- Numeric pain rating scale for pain responses in different body parts [26]

A comprehensive overview of these questionnaires and measurements can be found in Multimedia Appendix 1. It is important to emphasize that we do not claim this to be the definitive optimal dataset; rather, it serves as an illustrative example of potential index usage. An essential preprocessing step in constructing the health index involves converting, or linking, variables from these original datasets to new variables known as ICF codes. The linkage of original variables to ICF codes is described in the following section.

Mapping Data to ICF Codes

The ICF linkage guidelines [27,28] advocate for the training of 2 medical professionals to establish linkages between the original data sources and the qualifiers of the ICF codes. During the linking process, 2 experts will independently create linkage information for each new data source. If their independent linkages match, they are accepted as final. In cases where discrepancies arise, a third expert opinion is sought to determine the definitive linkage.

Given the potential coexistence of scientifically validated and self-made (ie, expert derived but not formally validated) ICF linkages in the data, it is crucial to define separate linkage reliability values for different linkage types. Scientifically validated linkages, typically established through consensus procedures outlined in ICF-linking rules, can be assigned higher reliability scores, while nonvalidated linkages are downweighted to reflect their informal nature. Here, $r \in [0,1]$ represents the *linkage reliability* designated for the corresponding source. The index user must specify the r value for every available source before computing the index. For instance, independently validated linkages may be assigned a maximum reliability of 1, whereas self-made linkages could be considered less reliable, resulting in r values typically set to less than 1.

For our dataset, formally validated ICF linkages do not exist, and therefore, we defined ICF linkages ourselves. While this led to lower linkage reliability when compared with formally validated linkages, this is not a limitation in this study because our focus is on the presentation of the new health index concept instead of proposing an optimal dataset for the health index computation. In fact, self-made linkages highlight the fact that the proposed health index concept is generic because it can be applied to heterogeneous datasets without requirements for the initial source and form of the data.

Details of the linkage for each questionnaire and measurement dataset are provided in Multimedia Appendix 1. For the EQ-5D-5L and Oswestry Disability Index questionnaires, data were available for 111 (22%) and 147 (29%) out of 505 individuals, respectively. Pain responses were accessible for 348 (69%) individuals, and mobility or strength test results were available for 420 (83%) individuals. Additionally, EQ-VAS, self-assessed health status on a scale from 0 to 100, was answered by 168 (33%) individuals.

The most frequently available ICF code in the dataset was *b780* (*sensations related to muscles and movement functions*), accessible for 420 (83%) individuals out of 505. The next 3 most common ICF codes available were *b7305* (*power of muscles of the trunk*), *b7355* (*tone of muscles of the trunk*), and *b7401* (*endurance of muscle groups*), all of which were available for 388 (77%) individuals. Codes related to pain localization (b28010, b28013, b28014, and b28015) were each available for 348 (69%) individuals, while b7302 was available for 321 (64%). Moderate availability was observed for b7300, b7350, and b7400, each present for 261 (52%) individuals. Codes b280, d450, and d5 were each available for 182 (36%) individuals. Several additional codes—b1340, d4103, d4104, d430, d470, and d910—were each available for 147 (29%) individuals, while b640 and d7702 were available for 136 (27%). The least

XSL•FO
RenderX

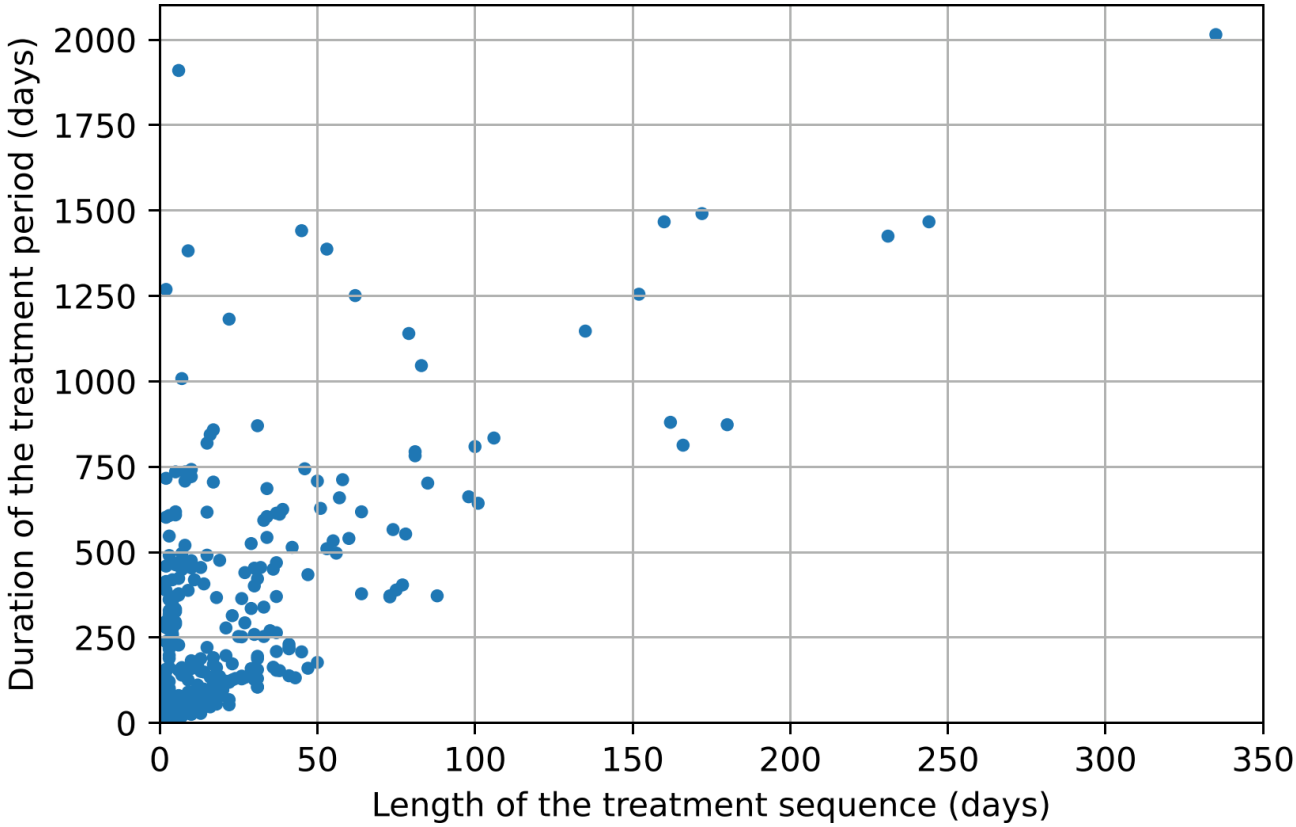
frequently observed codes (b152, b1528, d230, d455, d510, and d540) were each available for 111 (22%) individuals. To be included in the above list, there had to be at least one measurement available for the specified ICF code.

Overview of Collected Data

There are 2 important aspects to explore in the data. First, the *duration* of a treatment period is defined as the number of full calendar days between the initial and final day of the treatment. When treatment occurs only once, the duration is considered 0. Second, we can assess the frequency or the number of treatment days during the treatment period, referred to as the *length* of the treatment sequence. For example, if an individual undergoes treatment for 3 weeks with 4 clinic visits, the duration of the treatment period would be 21 days, and the length of the treatment sequence would be 4.

The median duration of the treatment period in the dataset was 69 days (IQR 0-301). Among individuals with nonzero duration, the median was 221 days (IQR 75-604). The median length of the treatment sequence was 3 days (IQR 1-16). Further exploration of this property is presented in Figure 2. The distribution of the length of the treatment sequence was as follows. A total of 148 (29%) out of 505 individuals had a sequence length of 1 day, indicating that for a substantial proportion of participants, data were available for only a single day. Furthermore, 61 (12%) individuals had 2 days, 44 (9%) had 3 days, and 26 (5%) had 4 days. Longer sequences were observed as follows: 71 (14%) individuals had 5 to 10 days, 48 (10%) had 11 to 20 days, 57 (11%) had 21 to 40 days, 31 (6%) had 41 to 80 days, 12 (2%) had 81 to 160 days, and 7 (1%) had 161 to 335 days.

Figure 2. The duration of the treatment period and the length of the treatment sequence for each individual are presented. Most individuals are clustered in the lower-left corner, indicating relatively short treatments with few individual treatment days.



Ethical Considerations

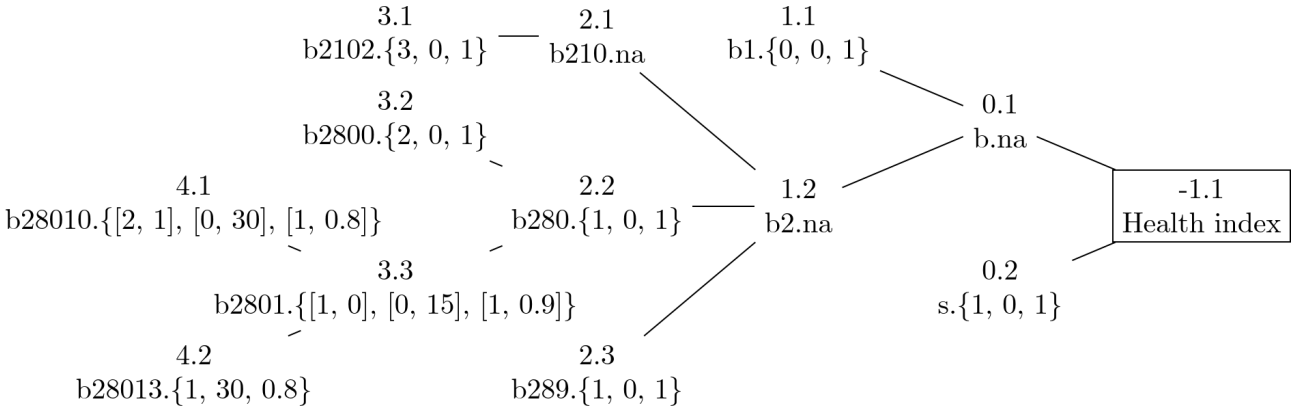
This study used previously collected and deidentified data from individuals who had provided informed consent for the use of their data in future research. The research did not involve any new interventions, physical or psychological risks, or other elements requiring ethical review as defined by the Finnish National Board on Research Integrity (TENK) guidelines on ethical review in human sciences [29]. Therefore, a separate ethical review was not required. The study was conducted in accordance with the principles of responsible conduct of research and data protection regulations. Personal data were processed in compliance with the EU General Data Protection Regulation, and deidentification was performed using procedures consistent with the Health Insurance Portability and Accountability Act Safe Harbor Method to minimize reidentification risk.

Model Architecture and Components

The proposed model structure aligns with the hierarchical organization of the ICF. Rather than incorporating all potential ICF codes, the model operates on whichever codes are available in a given dataset. In this study, we use the ICF codes listed in Figure 3 solely as an illustrative example of the model's usage; the model itself is designed to be broadly applicable to any ICF-compatible data.

XSL•FO
RenderX

Figure 3. An artificial example of a tree structure containing International Classification of Functioning, Disability and Health (ICF) codes. Each ICF code is represented by 2 rows of information. The first row includes the level number and the ICF code number within that level, separated by a dot. The format used is level.index, where level corresponds to the ICF level (with level 4 being the deepest), and index represents the per-level running index. The ICF codes are listed in alphabetical order at each level, and the running index follows this order. The level labeled as 0 refers to the ICF top components (b, d, e, and s). Level 1 corresponds to the chapter-level ICF codes, level 2 to the second-level ICF codes, and so forth. The second ICF code row indicates the ICF code and 3 attributes associated with it—namely, ICF qualifier x, time elapsed from measurement (TE) (see Multimedia Appendix B in Multimedia Appendix 2 for details on time weighting), and reliability of the linkage r (see section "Mapping Data to ICF Codes" for details on item linkage). The format is ICF_code.{qualifier, TE, reliability}. When a qualifier is not available for the ICF code, na is indicated after the dot. For the sake of simplicity, measurements from the d and e components are not incorporated in the current example, despite being supported.



Within this restricted set of ICF codes, some codes may further be absent—meaning they were not measured for all individuals. Consequently, our model includes both observed codes (those with measurements) and empty codes (those without measurements). Notably, an ICF code within the model can be observed through multiple measurements. This scenario arises when multiple original data variables are linked to the same ICF code. Additionally, a single ICF code may have multiple measurements if linked to data recorded at different time points. Throughout this study, we refer to all linked measurements as *qualifiers*, consistent with ICF terminology.

For illustrative purposes, Figure 3 provides an example of a simplified tree structure containing ICF codes. As previously mentioned, multiple original measurements can be linked to the same ICF code, resulting in the observation of that ICF code through multiple qualifiers.

In Figure 3, we observe 2 ICF codes, each associated with 2 qualifiers, both linked to the same ICF code at the third and fourth levels. The first qualifier in the ICF code *b28010* (4.1) has a value of 2 (the first number inside the square brackets). Additionally, it has attributes—namely, time_elapsed=0 and reliability=1. The second qualifier, originating from a different data source, also corresponds to the same ICF code. It has a value of 1 (the second number inside the square brackets) and attributes—again, time_elapsed=30 and reliability=0.8. Furthermore, empty ICF codes are also included in Figure 3 for illustrative purposes.

It is important to recognize that the example in Figure 3 is simplified and does not incorporate data from the ICF components related to *activities and participation* (*d*) or *environmental factors* (*e*), although these components are indeed supported within the index. Furthermore, it is not common practice to assign measurements directly at the component level. However, in this example, we have made an exception with a qualifier assigned to the *body structures* component.

Health Index Computation

The core concept behind our proposed approach is to systematically propagate information from the deepest nodes to the root of the tree. This process culminates in a single numerical metric known as the health index. Alternatively, if a health profile with multiple scores is desired, these metrics can be directly extracted from any level preceding the root level. We initiate the computation procedure at the leaf nodes of the tree. For instance, consider Figure 3, where we begin with ICF code *b28010* (4.1), then proceed to *b28013* (4.2), and subsequently to the next level represented by ICF code *b2102* (3.1), and so forth. This iterative computation continues until all ICF codes have been accounted for.

To compute the health index, we use a recursive algorithm that traverses the hierarchical ICF structure from the deepest nodes upward. At each level, the algorithm aggregates measurement values (qualifiers) using time decay and linkage reliability weights. This process continues until a single health index value is produced at the root level. The final computed value represents the unscaled health index, which we subsequently scale to a range of 0 to 100 to obtain the final health index. Detailed steps for this computation are outlined in the pseudocode presented in Figure 4, with the relevant equations provided following the algorithm. To streamline the algorithm presentation, some details that solely focus on performance optimization have been omitted. For example, given that the fourth level represents the deepest layer in the ICF framework, there is no practical need to verify child ICF codes at this level.

XSL•FO
RenderX

Figure 4. Health index computation algorithm. ICF: International Classification of Functioning, Disability and Health.

Algorithm 1: Health index computation.

```
Data: ICF code tree with qualifiers
Result: Health index
1  initialize variable level with the deepest ICF code level in the tree
2  while the health index has not been computed do
3      index = 1
4      while there are ICF codes at the level do
5          if there are child ICF codes for the ICF code level.index then
6              compute value of the ICF code x_{level.index} (Eqs. 2, 3, 4)
7              compute time weighting α_{level.index} (Eq. 5)
8              compute linkage reliability r_{level.index} (Eq. 5)
9              foreach child ICF code of level.index do
10                 remove all qualifiers s if they exist
11             index = index + 1
12     level = level - 1
13     if level < −1 then
14         scaled_health_index = HI(x_{−1.1}) (Eq. 1)
15         return scaled_health_index
```

In essence, the algorithm ensures that all available data—whether recent or older or direct or indirect—is systematically weighted and aggregated. By processing the tree level by level, the model maintains consistency and robustness, even when data are missing or unevenly distributed. In the algorithm, the input data consist of an ICF code tree with qualifiers, similar to the one illustrated in Figure 3. The desired output is the health index. The procedure begins by initializing the variable *level* (line number 1) based on the deepest available ICF code level within the data. For instance, in the data represented in Figure 3, there are qualifiers associated with 2 ICF codes at level 4 (4.1 and 4.2). Consequently, the *level* variable would be initialized as 4 in this case. The lowest possible value for *level* is −1, corresponding to the root of the tree. At this root level, the final health index is computed.

The purpose of the outer *while* loop (lines 2 - 15) is to sequentially process the ICF codes within the tree until the root is reached, at which point the health index is computed and returned. Within this loop, the variable *index* is initially set (line 3). The *index* serves as a per-level running counter, resetting to 1 whenever the tree level changes. For instance, in the processing order of ICF codes for the tree depicted in Figure 3, the sequence is 4.1, 4.2, 3.1, 3.2, 3.3, and so forth.

The inner *while* loop (lines 4 - 11) serves the purpose of iteratively performing necessary computations until all the ICF codes at the current *level* have been processed. The first *if* statement (line 5) checks whether child ICF codes are available for the currently examined ICF code, denoted as *level.index*. For instance, in Figure 3, the child ICF codes of *b280* are *b2800* and *b2801*. When child ICF codes are available for a given ICF code, the computation process begins. In lines 6 to 8, we compute the value of the ICF code, denoted as $x_{level.index}$, along with the time weighting $\alpha_{level.index}$ and the linkage reliability $r_{level.index}$. These computations are elaborated in detail in "Computing ICF code values and reliability."

Subsequently, in lines 9 and 10, we remove all qualifiers from the child ICF codes associated with *level.index*. The purpose of this step is to ensure that these values are not reused in subsequent computations. Instead, we rely on the values computed during lines 6 to 8 as we continue the computation process. Additionally, the per-level running index (*index*) is incremented after handling each ICF code (line 11), whereas the level index (*level*) is decremented once all ICF codes at that level have been processed (line 12).

When the level index (*level*) drops below −1, it signifies that we have reached the root of the tree (line 13). At this point, we have computed the raw health index, denoted as $x_{-1.1}$, according to the ICF code numbering scheme depicted in Figure 3. However, this raw value requires transformation to enhance its interpretability. Our desired presentation for the final health index is an integer ranging from 0 (indicating the worst health) to 100 (representing the best health). To achieve this, we invert the source and target values: whereas a raw health index value of 0 corresponds to optimal health, the same value within the target range signifies the poorest health condition.

As higher qualifier values indicate worse health, we invert the raw score to ensure that a higher health index corresponds to better health. This makes the final index more intuitive for clinical interpretation. The transformation is applied as follows, resulting in the final health index denoted as HI (line 14):

$$HI(x_{-1.1}) = \mathrm{nint}\left(100 - 100 \times \frac{x_{-1.1} - \min(x_{-1.1})}{\max(x_{-1.1}) - \min(x_{-1.1})}\right) \quad (1)$$

where nint represents the nearest integer function, x-1.1 is the raw value to be transformed, and min(x-1.1) and max(x-1.1) correspond to the minimum and maximum theoretical or actual

XSL•FO
RenderX

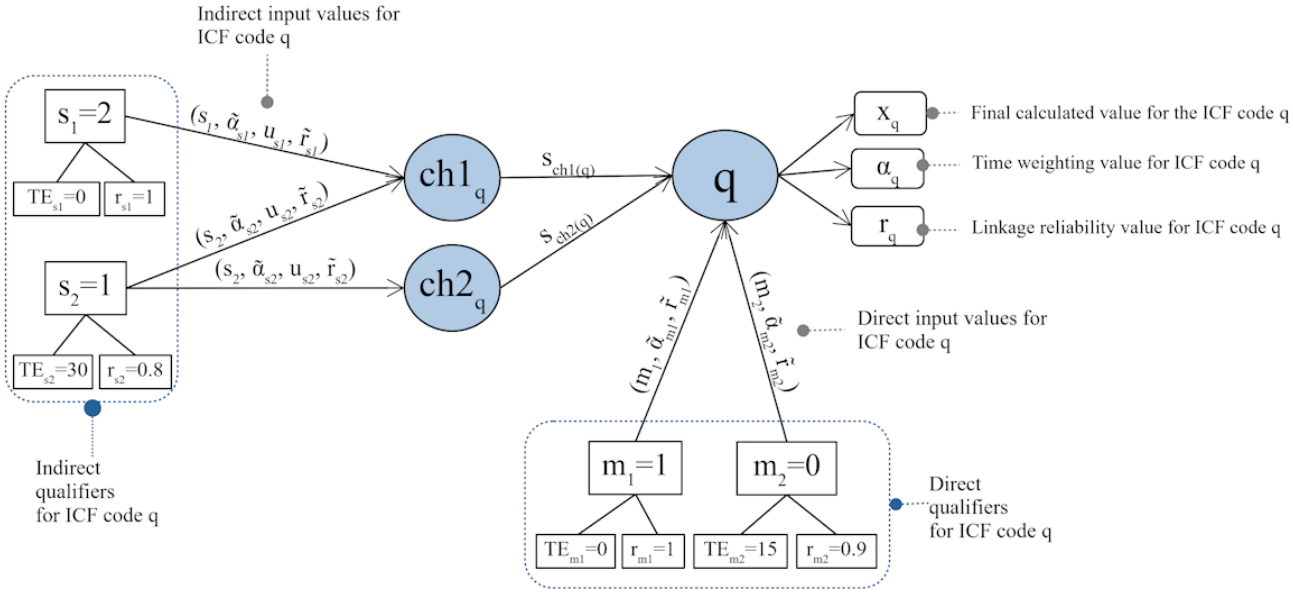
raw values. Given that we know the theoretical minimum (0) and maximum (4) values for the raw health index, our primary approach is to directly scale and invert the raw values to fit our target range. Specifically, we define min(x-1.1)=0 and max(x-1.1)=4. The resulting value after this transformation (line 15) represents the final health index.

Subsequently, once the health index has been computed, the tree is reset to its original state. This entails removing all previously computed values from the ICF codes and restoring all qualifiers to their initial positions. This procedure ensures that any future computations do not inadvertently rely on previously computed values.

Detailed Examination of Components

In Figure 5, we closely examine the components within our tree example. Specifically, we focus on 3 ICF codes: *b28010*, *b28013*, and *b2801*, extracted from Figure 3 for a comprehensive analysis. Each ICF code represents a specific health-related concept.

Figure 5. An International Classification of Functioning, Disability and Health (ICF) code–level example of the health index computation. In this figure, we have a viewpoint of an individual ICF code marked as q. Its child ICF codes are marked here as $ch1_q$ and $ch2_q$.



Let us denote the ICF code whose output value we compute as *q*. In Figure 5, this corresponds to node 3.3 (*b2801*) from Figure 3. Additionally, *q* has 2 child ICF codes: ch1q (equivalent to ICF code 4.1, *b28010*) and ch2q (equivalent to ICF code 4.2, *b28013*). These 3 codes are highlighted in blue color in Figure 5. Measurements mapped to these ICF codes are also shown in Figure 5 and are denoted as m1,m2,s1,s2∈ . Owing to the ICF mapping performed as a preliminary step, values of these measurements are strictly between the range 0≤m1,m2,s1,s2≤4, encompassing the range of qualifiers in the ICF. Separate notation is used for direct (m_i) and indirect (s_j) measurements from the point of view of the code of interest (*q*) to highlight the different roles of the measurements in the health index computation. The origin of the measurements is not important in the context of this example. We also illustrate that a single measurement (see S_2 in Figure 5) can contribute to multiple ICF codes. The qualifiers play a crucial role in linking to specific ICF codes, as indicated by arrows. In this illustrative example, they are as follows:

1. A *moderate problem* (qualifier s1=2) is recorded for ICF code *pain in head and neck* (*b28010*/ch1q). The measurement is made on the latest available day in the treatment period, as indicated by TEs1=0, and its linkage reliability is defined as the strongest possible (rs1=1).
2. A *slight problem* (qualifier s2=1) is recorded for ICF codes *pain in head and neck* (*b28010*/ch2q) and *pain in back* (*b28013*/ch2q). The measurement was made 30 days ago from the perspective of the latest available measurement, as indicated by TEs2=30, and its linkage reliability is defined as (rs2=0.8).
3. A *slight problem* (qualifier m1=1) is recorded for ICF code *pain in body part* (*b2801*/*q*). The measurement is made on the latest available day in the treatment period, as indicated by TEm1=0, and its linkage reliability is defined as the strongest possible (rm1=1).
4. A *no problem* (qualifier m2=0) is recorded for ICF code *pain in body part* (*b2801*/*q*). The measurement was made 15 days ago from the perspective of the latest available measurement, as indicated by TEm2=15, and its linkage reliability is defined as (rm2=0.9).

The different components are marked in Figure 5:

- *Direct qualifiers (D) for ICF code q*: m1 and m2 are the *direct qualifiers* for ICF code *q*. In other words, they are directly linked to *q*. TEm1 is the age of a qualifier measured in full days, and rm1 denotes the reliability of the linkage.
- *Direct input values for ICF code q*: These values (mi, α~mi, r~mi) are associated with ICF code *q*. While the qualifier mi can be directly attached to the ICF code, other values must be computed based on the available data.
- *Indirect qualifiers (I) for ICF code q*: This component represents the 2 available qualifiers, s1 and s2, for the 2 child ICF codes ch1q and ch2q. Qualifiers s1 and s2 serve as the *indirect* qualifiers for ICF code *q*. They are linked to *q* through its child ICF codes. The time elapsed value

XSL•FO
RenderX

for s1 is denoted by TEs1, and its source linkage reliability is given by rs1.

- *Indirect input values for ICF code q*: These values (sj, α~sj, usj, r~sj) are associated with the child ICF codes. They are used to compute the values sch1(q) and sch2(q) for the child ICF codes.

There are 3 outputs in the diagram:

- *Final computed value for the ICF code q*: Denoted as xq, this represents the ultimate value associated with ICF code *q*. It comprises both *direct* and *indirect* qualifiers. Specifically, xq is the weighted average of all available qualifiers: s1, s2, m1, and m2. In this case, it would be computed as follows: xq=(β~m1Dm1+β~m2Dm2+sch1(q)+sch2(q))/(β~m1Dm1+β~m2Dm2+β~s1I+β~s2I+β~s2I) where β terms refer to weighting coefficients. Details regarding the equation are explained in the following section.
- *Time weighting value for q*: Denoted as αq, this value represents the weighted mean of all available time weighting α values, as defined in Equation 5.
- *Linkage reliability value for q*: Represented by rq, this value is the weighted mean of all available *r* values, also defined in Equation 5.

Computing ICF Code Values and Reliability

In the previous section, we explained the computation of ICF code values using a practical example for clarity. Now, in this section, we delve into a broader discussion on the computation of ICF code values and their associated reliability measures.

Determining ICF Code Values

The value of the ICF code *q* is defined as follows:

(2)xq=(∑i∈Dβ~miDmi+∑k∈chqβ~xkIxk)/(∑i∈Dβ~miDmi+∑k∈chqβ~xkI)

This definition consists of 4 separate elements:

1. Function f is a weighting function that can be selected to either give emphasis to higher qualifiers (exponential weighting) in data or to highlight lower qualifiers (logarithmic weighting). It is also possible to use no weighting at all (linear weighting). The different weighting functions are presented in more detail in Multimedia Appendix 2.
2. The first sum term ∑i∈ Dβ~miDmi represents the weighted sum of qualifiers mi directly linked to the ICF code *q*. Here, *D* denotes a set of *direct qualifiers*—that is, those qualifiers that are directly linked to *q*. Additionally, we introduce a total weighting term βD for direct qualifiers:

   (3)βmiD=α~mir~mi

   In this expression, α and *r* play essential roles. The first element, α, serves as a time weighting value that accounts for the age of the corresponding qualifier. Generally, as a qualifier ages, its influence diminishes in the computation. Further details on time weighting are discussed in Multimedia Appendix 2. The second element, *r*, corresponds to the linkage reliability defined in "Mapping data to ICF codes." Notably, in Equation 2 and Equation 3, the terms β, α, and *r* carry the tilde notation (β~,α~,r~). This notation signifies that normalized values are used instead of raw values. A standard normalization procedure ensures comparability across different values. Refer to the Multimedia Appendix 2 on normalization for the 2 relevant equations.
3. The second sum term ∑k∈ chq∑j∈ kβ~sjIsj represents the weighted sum of qualifiers sj that are indirectly linked to the ICF code *q* via its child ICF codes. Here, chq refers to a child ICF code of *q* that possesses either a qualifier or a computed value. Consequently, each child ICF code of *q* contributes to the overall computation. The inner sum iterates through all the qualifiers associated with the child ICF code.

   Here, *I* denotes a set of *indirect qualifiers*. We introduce a total weighting term for indirect qualifiers, denoted as βI. This term is a slightly modified version of the weighting term shown in Equation 3, defined as follows:

   (4)βsjI=α~sjr~sjusj

   where usj quantifies the *uniqueness of source* for the qualifier sj. Specifically, it is computed based on the number of linkages between sj and child ICF codes of *q*. It is defined as usj=1/zj,zj∈ N+, where zj represents the total linkages involving sj and child ICF codes of *q*. The purpose of introducing *u* is to adjust the qualifier weighting when the same source contributes to values for 2 or more ICF codes under the same parent ICF code. The terms α and *r* have the same purpose as in Equation 3.
4. The third sum term ∑k∈ chqβ~xkIxk mirrors the second sum term discussed earlier. However, there is a crucial distinction: it operates on the computed values (*x*) rather than the qualifiers (*s*). Notably, the second sum term becomes unnecessary here, as each child ICF code corresponds to a single computed value. Furthermore, the concept of uniqueness of source does not apply in this context. Consequently, we define β~xkI similarly to Equation 3, as β~xkI=α~xkr~xk.

Assessing ICF Code Reliability

There are 2 values, αq (equivalent to αlevel.index in Figure 4 algorithm, line 7)—the weighted mean of time weighting α—and rq (equivalent to rlevel.index on line 8)—the weighted mean of linkage reliability *r*—that are computed for xq. They are defined almost identically to xqin Equation 2:

(5)αq=∑i∈ Dβ~miDαmi+∑k∈ chq∑j∈ kβ~sjIαsj+∑k∈ chqβ~xkIαxk

For computing rq, we replace the α term in the equation with *r*. Here, the β~ terms determine the weightings for each α or *r* component. These values, αq and rq, are attributes to the xq value and are intrinsically associated with it.

Results

Overview

In this section, we present the results of applying the proposed health index model to the David Health Solutions dataset. While some methodological details are included to support interpretation, the focus is on the outcomes of the validation analyses. The computed health indices for all individuals in the dataset underwent preliminary validation by comparing the results with self-reported EuroQol Visual Analogue Scale (EQ-VAS) responses and maximum pain ratings.

XSL•FO
RenderX

Two validation groups were established. The first group had a treatment period duration of at least 90 days, whereas the second group adhered to a 30-day limit. In the first group, the length of the treatment sequence (as described in "Overview of Collected Data") was a minimum of 10 days. For the 30-day group, the treatment sequence length was at least 5 days. Individuals could be assigned to both groups, provided that the day and qualifier preconditions were met.

To explore the effect of different time weightings, we computed Pearson correlation coefficients for various time decay constants (γ). Multimedia Multimedia Appendix 2 provides a detailed definition of γ. The chosen values for γ were as follows:

- $\cdot \gamma 1=(1/20)1/30\approx 0.905$ represented a very heavy time decay, where a 30-day-old qualifier contributed only 5% of its original time weight.
- $\cdot \gamma 2=(1/3)1/30\approx 0.964$ represented our estimated preliminary potential time decay value in a real-life scenario. Practically, this value implies that a measurement made 30 days ago was given one-third of its original time weight.
- $\cdot \gamma 3=1$ signified no time decay at all.

The algorithm was implemented using the Python programming language [30]. Statistical analyses were conducted using SciPy [31], and figures were generated with Matplotlib [32].

Comparing Self-Assessed Health Status Responses With Health Index

First, the health index was computed without emphasizing either lower or higher qualifiers. In Table 2, the EQ-VAS response was compared with the health index, and Pearson correlations are presented for both groups. We also explored the influence of changing the time decay constant (γ). The results predominantly reveal moderate positive correlations between the EQ-VAS response and the computed health index. The 90-day group consisted of 84 individuals, with a total of 125 EQ-VAS responses recorded. While most individuals had only one recorded EQ-VAS response, some provided 2, 3, or even 4 responses—all of which were used in the computations. The 30-day group included 115 individuals, with a total of 159 EQ-VAS responses.

Table . Pearson correlations for the EQ-VAS[a] response versus health index[b].

γ	30-d group	90-d group
(120)1/30	0.690	0.642
(13)1/30	0.700	0.659
1	0.654	0.599

[a]EQ-VAS: EuroQol Visual Analogue Scale.

[b]All reported correlations are statistically significant (P<.001).

Comparing Maximum Pain Responses With Health Index

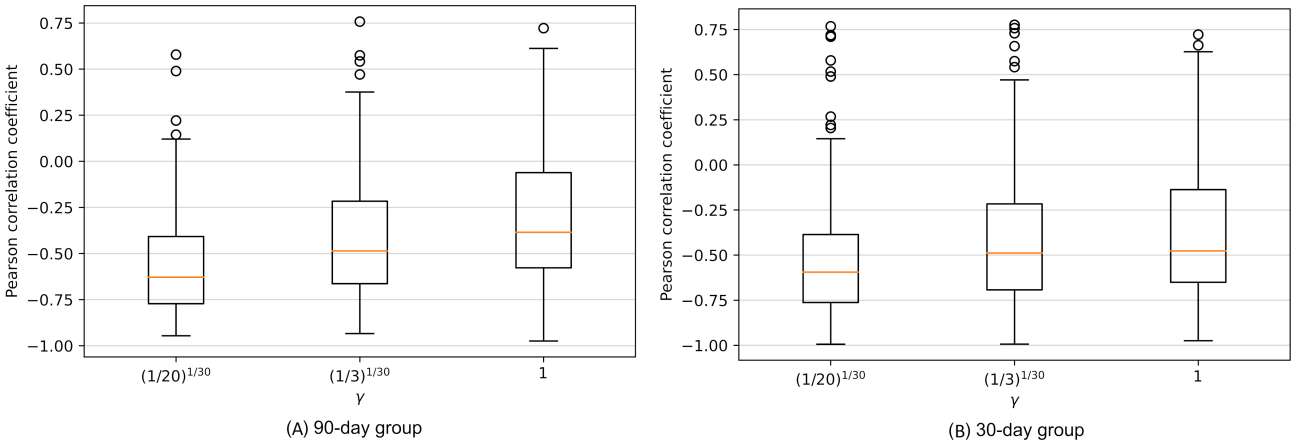
In the context of maximum pain versus the health index, we individually computed the Pearson correlation for each person between the maximum pain trajectory and the health index trajectory. We define maximum pain as the single highest value among the 4 self-assessed pain responses. Subsequently, a maximum pain trajectory was formed for each individual. As this information was also used in computing the health index, there was inevitably at least some correlation between the 2 values. However, this comparison provides a useful second perspective for model validation.

To prevent data from incorrectly showing significance, we applied the Bonferroni correction in the maximum pain versus health index trajectory correlation computations. Specifically, the α level was divided by n.

Figure 6 presents the distribution of correlation values for the 2 groups. Notably, a slight upward trend in median correlation is visible in both groups: as γ increases, the median correlation approaches zero. In all cases, the correlation was negative, varying approximately from low to moderate. There were small differences in the n values for the different groups. In some configurations, the health index value remained constant for certain individuals, making correlation computations impossible. These persons were omitted from the computations.

XSL•FO
RenderX

Figure 6. Boxplots illustrating Pearson correlations between the health index and maximum pain trajectories. (A) 90-d group, where the correlation medians for the 3 time decay groups (n=133) are −0.628, −0.486, and −0.385. (B) 30-d group, with correlation medians for the 3 time decay groups at −0.594, −0.489 (n=184), and −0.477 (n=182).



Examining the portions of significant correlations, we observed that for the 90-day group, Bonferroni-corrected significant (P<.05) portions were 55% for γ1, 41% for γ2, and 31% for γ3. In the 30-day group, the similar portions were 41% for γ1, 29% for γ2, and 23% for γ3. These findings suggest that as the length of the treatment sequence and duration of the treatment period increase, more significant correlations emerge.

In Table 3, we further examine the effect of time series length by binning the maximum pain versus health index trajectories of individuals into 3 bins based on the length of the treatment sequence. The table presents both the 90-day group and the 30-day group, separated by a comma. These results also indicate that as more data points are included in the trajectory, the significant portion increases. Additionally, in Figure 7, 2 example individuals and their maximum pain versus health index trajectories are presented.

Table . Correlation statistics between maximum pain and health index, using a 3-bin approach based on treatment sequence length[a].

Bin[b]	Day ranges	Bonferroni significant (P<.05) portions[c]	Median correlations	Subset n[d]
1	[10, 26], [5, 15]	18.8%, 7.9%	−0.385 (IQR −0.702 to −0.082),−0.569 (IQR −0.815 to −0.261)	48, 63
2	[27, 42], [16, 32]	37.5%, 18.3%	−0.508 (IQR −0.695 to −0.327),−0.376 (IQR −0.577 to −0.074)	40, 60
3	[43, 325], [33, 325]	68.9%, 65.6%	−0.509 (IQR −0.636 to −0.312),−0.511 (IQR −0.670 to −0.299)	45, 61

[a]The time decay value was defined as γ_2.

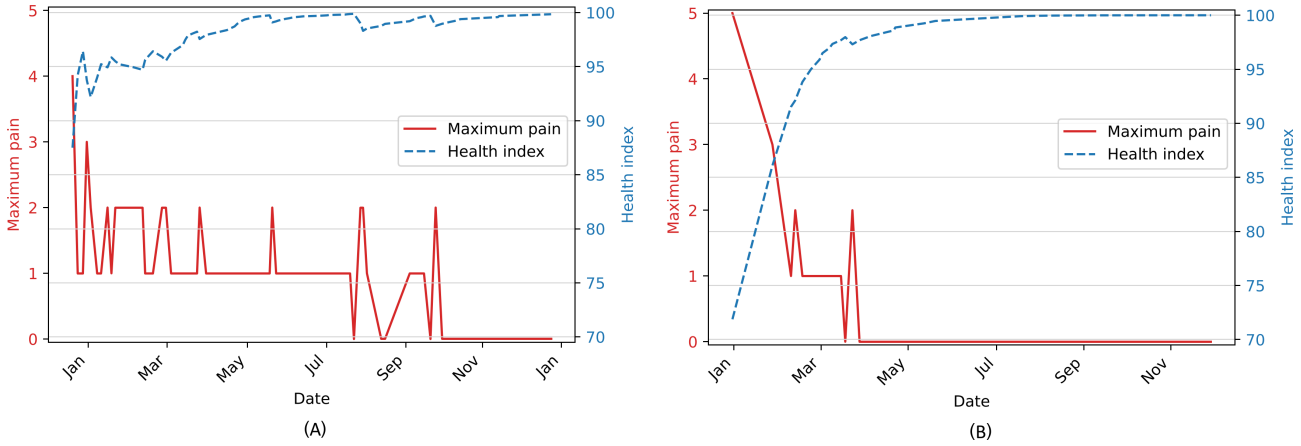
[b]The 3 bins are approximately equal in size (tertiles). For each bin, we specify the range of treatment sequence lengths included. For instance, bin 1 encompasses individuals with treatment sequences lasting from 10 to 26 days (in the 90-d group) or from 5 to 15 d (in the 30-d group).

[c]Proportion of correlations that remain significant after Bonferroni correction at a 0.05 significance level.

[d]Number of individuals included.

XSL•FO
RenderX

Figure 7. The maximum pain versus health index trajectories for 2 example individuals. The first case (A) demonstrates a moderate negative correlation with a treatment sequence length of 40 d. The second case (B) represents an extreme scenario with a very high negative correlation with a treatment sequence length of 28 d. Both cases share the same time decay value, γ_2. Notably, the Bonferroni-corrected correlations were statistically significant (P<.001).



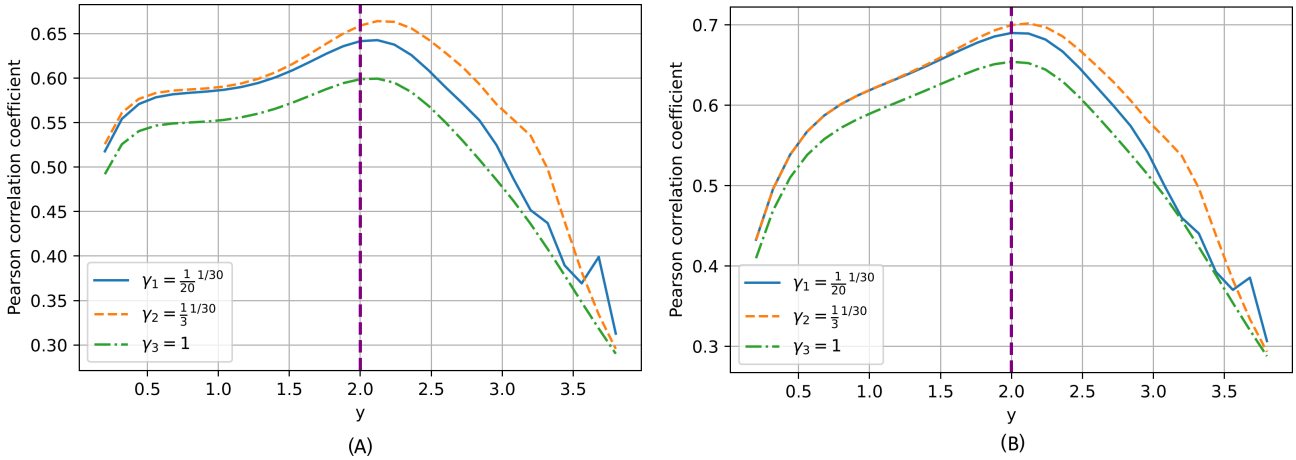
Impact of Nonlinear Weighting

In this section, we explore the impact of qualifier weighting on correlations. The model allows for the calibration of qualifier weighting by adjusting the parameter y. The weighting can operate under three distinct functions: (1) when $y \in (0,2)$, the weighting is exponential, which implies that higher qualifier values carry more weight; (2) when $y \in (2,4)$, the weighting is logarithmic, indicating that lower qualifier values are given more weight; and (3) when y=2, the weighting is linear, meaning that all qualifier values are treated equally, without any specific weighting. For details on value weighting, please refer to Multimedia Appendix 2.

In this study, we examined a total of 31 values for the y parameter. These values were selected to be evenly spaced on both sides of the linear case (y=2), ranging from y=0.2 to y=3.8 with intervals of 0.2. The results of the Pearson correlation between the health index and EQ-VAS response are depicted in Figure 8. For the 90-day group, the highest correlations for all 3 γ values were observed at y=2.12, with correlation values of 0.643, 0.664, and 0.599, respectively. In contrast, for the 30-day group, the highest correlations were observed at the linear weighting point (y=2). The correlations at this point are identical to those presented in Table 2.

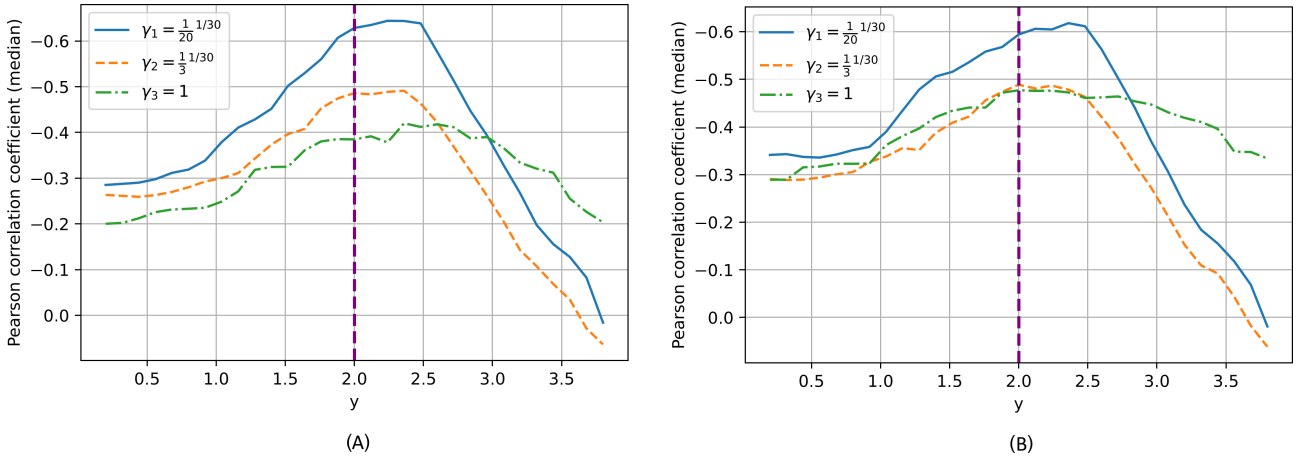
In the context of maximum pain versus the health index, the median correlations for both groups are depicted in Figure 9. For the 90-day group, the 3 γ values exhibited the highest negative correlations of −0.644 (at y=2.24), −0.491, and −0.420 (both at y=2.36), respectively. Conversely, the 30-day group demonstrated correlations of −0.618 (at y=2.36), −0.489, and −0.477 (both at y=2).

Figure 8. Pearson correlations between the health index and EQ-VAS response for 3 time decay constants. (A) 90-d group (n=125) and (B) the 30-d group (n=159).



XSL•FO
RenderX

Figure 9. Median Pearson correlations between the health index and maximum pain for 3 time decay constants. (A) shows the 90-d group (γ_1, γ_2: n=133, γ_3: n=132, 133) and (B) the 30-d group (γ_1, γ_2: n=184, γ_3: n=181, 182, 183).



Fundamental Sensitivity Analysis

The objective of conducting a sensitivity analysis is to understand how variations in input parameters influence the outcome of a model, in this case, the health index [33]. Our aim is to verify that the model behaves predictably and logically when modifications are made to the input variables. For instance, an increase in a single qualifier, while other qualifiers remain constant, should result in a decrease in the health index.

The ICF structure consists of over 1600 distinct ICF codes, each potentially having qualifiers. We use a simple data sample, as shown previously in Figure 3, to investigate the model's response to a straightforward scenario: an increase or decrease in qualifier. In this scenario, we do not consider the reliability value (r) and assume it to be 1 for all qualifiers.

We target nodes that possess one or more qualifiers for value changes, with a particular emphasis on the qualifier and its impact on the final health index. For each targeted node, we explore all possible qualifiers ranging from 0 to 4, while maintaining constant values in other nodes, as illustrated in Figure 3. This setup produces 45 distinct health index outcomes, including the default.

The outcomes of the fundamental sensitivity analysis are detailed in Table 4. Throughout this analysis, the health index exhibited anticipated behavior. It is evident that ICF codes closer to the root (ie, the health index) of the tree exert a greater influence on the health index. Conversely, alterations in the most remote parts of the tree exert minimal influence on the overall health index. An additional characteristic worth noting is the time weighting. For example, alterations in the most recent qualifier, denoted as $b28010_1$, exert a greater influence on the outcome compared to modifications in the qualifier from 30 days ago, represented as $b28010_2$.

On the basis of the empirical patterns in this dataset, we propose 2 preliminary default parameters for future use: a moderate time decay value (γ=(13)1/30) and a linear weighting function (y=2). These settings offered stable behavior and consistent correlations across both EQ-VAS and pain trajectories. They should be revisited when applying the model to other populations or more extensive ICF datasets.

Table . Outcomes of the health index from the sensitivity analysis. ICF codes b2801 (3.3) and b28010 (4.1) have 2 recorded qualifiers, each marked with a subscript of 1 or 2 to indicate their temporal sequence, with 3.3_1 denoting the most recent qualifier. All other qualifiers were consistently fixed to the values shown in Figure 3. The time weighting, γ, was fixed at (1/3)1/30

Qualifier[a]	S 0.2	b1 1.1	b280 2.2	b289 2.3	b2102 3.1	b2800 3.2	$b2801_1$ 3.3_1	$b2801_2$ 3.3_2	$b28010_1$ 4.1_1	$b28010_2$ 4.1_2	b28013 4.2
0	85.04	73.71[b]	74.59	75.34	82.74	75.46	74.09	73.71[b]	74.48	73.84	74.09
1	73.71[b]	67.94	73.71[b]	73.71[b]	79.73	74.59	73.71[b]	73.49	74.09	73.71[b]	73.71[b]
2	62.38	62.17[c]	72.83	72.08	76.72	73.71[b]	73.32	73.27	73.71[b]	73.58	73.32
3	51.05	56.40	71.95	70.44	73.71[b]	72.83	72.94	73.04	73.32	73.45	72.94
4	39.72	50.63	71.08	68.81	70.70	71.95	72.56	72.82	72.94	73.32	72.56

[a]Qualifier value (0 - 4) used for each ICF code, ranging from 0.2 to 4.2 (refer to Figure 3 for ICF code identification),

[b]Default configurations.

[c]For instance, when the qualifier for ICF code b1 (1.1) is 2, the corresponding health index is 62.17.

XSL•FO
RenderX

Discussion

Principal Results and Comparison With Prior Work

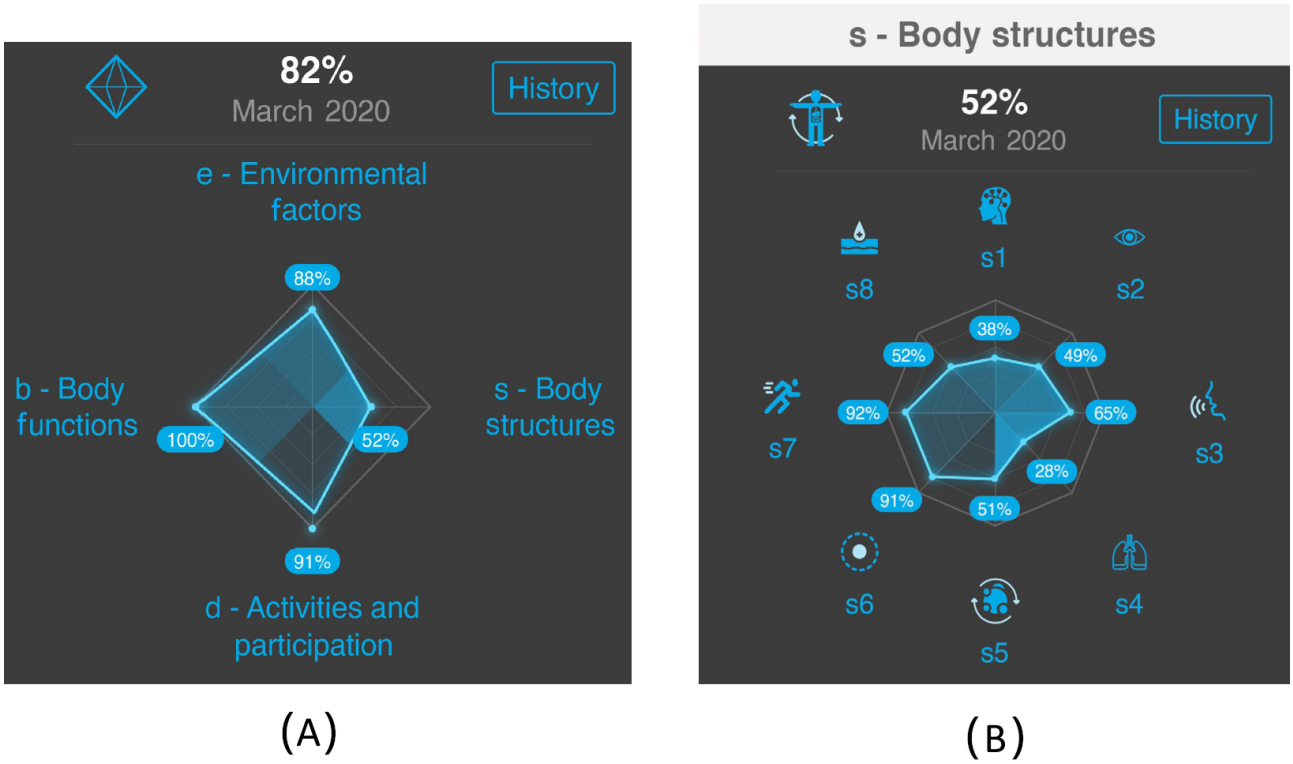
In clinical practice, the health index could serve as a decision support tool for rehabilitation planning, chronic disease management, and progress tracking. For example, a physiotherapist could use the health index to track a patient's recovery from back pain, identifying which functional domains are improving and which require further intervention. By providing a standardized and interpretable metric, it enables clinicians to quickly assess patient status, identify areas needing intervention, and monitor changes over time. The health profile further supports personalized care by highlighting specific domains of functioning and disability.

This study demonstrated that the proposed ICF-based health index correlates meaningfully with self-reported health measures. Moderate positive correlations with EQ-VAS and negative correlations with maximum pain trajectories support the index's validity. Sensitivity analyses confirmed that the model responds predictably to changes in input qualifiers, with recent measurements exerting greater influence. The model also proved robust to missing data, successfully computing health indices across varying levels of data completeness. Furthermore, the impact of time decay and qualifier weighting parameters was systematically explored, revealing optimal configurations that enhance correlation strength.

From the existing research discussed in "Prior Work," it is evident that the assessment of overall health status and the computation of health indices have followed varied approaches, using a diverse set of variables. Our approach to computing the health index and health profile is quite different from the methodologies discussed there. Below, we highlight significant divergences and commonalities between past methodologies and our approach.

First, our proposed model exhibits a generic nature: it does not assume a predetermined set of variables for utilization. Instead, the model is based on the established ICF framework, enabling its adaptability to any dataset encoded in the ICF-compatible format. However, to ensure the derived health index comprehensively reflects overall health status, the set of ICF codes—forming the foundation of computation—should evidently encompass all aspects relevant to determining the health of the individual. This selection of ICF codes may vary individually; for instance, collecting specific information on low back pain might not contribute substantially if the individual does not experience this condition. Nevertheless, our proposed model maintains the capability to compute the health index across a spectrum, ranging from scenarios with very sparse data (where only one ICF code is available, for instance) to the theoretical extreme where all ICF codes are accessible. Deriving the health index from diverse datasets offers a unique advantage. However, this can potentially become a drawback if the characteristics of the data are not incorporated into the interpretation of the results. Visual inspection of the health index and health profile, along with information on how much data has been collected (called *coverage*), as exemplified in Figure 10, can aid in this process.

Figure 10. Preliminary graphical presentation of an individual's health index and health profile. A high-level view is shown in panel A, while panel B explores the health profile further in one of the components. In panel A, the overall health index, exemplified here as 82%, is displayed at the top. Additionally, the 4 main components of the International Classification of Functioning, Disability and Health are presented. These components constitute a basic-level health profile. Panel B provides a more detailed view into the body structures component, showcasing the chapter-level values and their respective coverages.



XSL•FO
RenderX

Furthermore, it is essential to highlight here that we do not prescribe any specific ICF sets for application; our preliminary validation using collected data represents just one potential usage. Notably, extensive research exists that defines ICF core sets outlining fundamental ICF codes crucial for various aspects related to functioning, disability, and health. For example, Cieza et al [34] defined a minimal generic set of domains for describing functioning and health for adults. Additionally, more specific core sets have been introduced, eg, for chronic conditions [35], including low back pain, diabetes mellitus, obesity, depression, and others.

Second, our proposed model operates independently from existing data—meaning that it does not rely on training data, coefficients, or predefined major model parameters. Instead, its core parameters stem directly from the hierarchical structure of the ICF, acting as weightings for different levels within this structure. Thus, one can compute the health index for an individual if there is some ICF-coded data available for that person. Moreover, the health index computed for an individual remains entirely distinct and unaffected by the health indices of others. In this aspect, our approach shares similarities with simpler summation-based methods (see, eg, [12-14]).

It is important to note the absence of a definitive *ground truth* for overall health status—there exists no singular, reliable response variable that could serve as a benchmark for validating any health index. In one study using ML [22], cause of death for specific diseases was used as the ground truth. Notably, our approach refrains from using ML elements, in part due to the absence of a definitive response variable for overall health. Furthermore, our aim does not involve predicting risks for specific diseases; instead, the model focuses on providing a concise estimation of health status. We argue that our model enables, through the use of ICF, a more holistic view on health, since by the WHO definition health is “not merely the absence of disease and infirmity,” but “is a state of complete physical, mental and social well-being” [36]. Moreover, aspects such as daily life activities, mental health, and social interactions best encapsulate most people’s intuitive notion of health [34,36], all of which the ICF is adept at capturing.

Finally, our proposed model is designed to handle the longitudinal aspect of infrequent health records and missing data fluently. These are important characteristics of health data [37]. However, only 2 studies (see [22,23]) considered these problems in their approaches. Nonetheless, we argue that our proposed methodology stands out for its simplicity and use of the inherent hierarchy within the ICF framework. Our model operates without expecting identical variables for different individuals. As the only prerequisite is to have the data in the ICF-compatible format, in fact, the whole concept of missing data is eliminated from the model’s perspective. Each individual’s data are processed independently, rendering the model robust against common health data problems. Consequently, the model generates output for any dataset meeting minimal requirements. Our objective is to give precedence to recent qualifiers in the computation of the health index, while still considering past qualifiers to a certain extent. This approach not only ensures a comprehensive view of the individual’s health status but also allows both experts and individuals to track the progression of various health-related trajectories over time.

Overall, existing research aligns with several aspects of our study objectives, notably in acknowledging disparities in data across various countries [15] and addressing challenges related to sparse, infrequent data with missing values [22]. However, existing health indices and profiles in previous studies primarily rely on predefined attributes for computing the final health index.

Limitations

The ICF framework empowers its users to delineate the health and functioning of an individual in a highly detailed manner, beyond the first generic qualifier. In the *body structures* construct, the second qualifier can be used to indicate the nature of the change in the respective body structure, whereas the third qualifier specifies the location, for instance, left or right [4,38]. Moreover, in the *activities and participation* construct, the second generic qualifier (capacity) is used to denote limitations without assistance [4]. The proposed model, however, uses only the first generic qualifier, thereby not leveraging any additional information potentially provided by these other qualifiers.

In the context of the ICF’s *environmental factors* construct, the first qualifier can be used to signify either the positive effects (facilitators) or the negative effects (barriers) of the environment [4]. The current model is only equipped to handle barriers; hence, the potential positive impact of possible facilitators on the health index is not incorporated into the proposed model.

Qualifiers 8 (*not specified*) and 9 (*not applicable*) are not accommodated by the proposed model. At present, the values are treated as ordinal, and these nominal values cannot be used. This limitation might lead to a loss of some information, as these qualifiers are used in the ICF system to record information that cannot be captured using the [0,4] value range. Qualifier 8 should be applied when there is a problem, but the severity of that problem is unknown. Additionally, qualifier 9 is typically used in a situation when the use of the category is not appropriate for the individual [38]. These qualifiers have also been used in practical applications [39].

Taken together, these constraints mean that the current implementation captures only a portion of the information the ICF framework can offer. Expanding the model to incorporate additional qualifiers and environmental facilitators would allow it to reflect a broader range of clinically relevant information while keeping the computational framework unchanged.

A potential limitation concerning the practical usability of the model is that sparse data might provide a misleading impression about a person’s overall health. However, a basic data coverage metric can be used to assess the comprehensiveness of the data used and thereby also the reliability of the index outcome. Enhancing the coverage metric is also a key area for future improvements. The total coverage value could potentially be defined as a fraction representing the number of chapter-level codes that have associated measurements. To achieve a complete coverage of 100%, there would need to be a minimum of one measurement under each chapter-level code within the *body structures* component’s *s1* to *s8*, as well as all other chapter-level

XSL•FO
RenderX

codes across the remaining 3 main components. In addition, as the ICF does not currently contain the classification of the personal factors, such as an individual's age, they cannot be used in the current computation process. Moreover, the inclusion of personal factors has been identified as relevant and useful for the application of the ICF in different settings [40].

The dataset used for the preliminary validation of the health index is admittedly suboptimal, lacking the ability to encapsulate certain elements that define functioning. This study does not explicitly endorse the use of this particular set of ICF-coded data. Instead, we advocate for the use of datasets that offer a more comprehensive representation of all facets encompassed in the ICF for future applications. The existing ICF core sets can serve as a potential foundation for identifying the essential data required to collect for each specific condition. One possible solution would be to collect data that adheres to at least the minimal generic set proposed by Cieza et al [34], ensuring a baseline level of coverage across applications. An additional limitation in the preliminary validation dataset is that several variables were linked to ICF codes using expert-derived rather than formally validated linkages. Although these are downweighted using the linkage reliability parameter, formally validated linkages would further strengthen the results. The methodological framework itself is unaffected, but future applications may benefit from datasets where validated linkages are available. Incorporating formally validated or multirater ICF-linking procedures would improve reliability and external validity.

Another minor concern regarding the validation process is the current lack of differentiation between individuals under the age of 18 years and adults. The few individuals under 18 are incorporated into the larger group without any distinct consideration.

Conclusions

We assert that the ICF framework is a solid foundation for developing a health index. As a globally recognized and standardized classification system, it provides a common language and conceptual basis for describing comprehensively factors influencing an individual's health and functioning. Since its inception, the ICF has found extensive application in a variety of fields. These include, but are not limited to, clinical practice, policy formulation, social policy, and educational sectors [41]. Furthermore, it has become an indispensable tool in population health research projects worldwide [42].

An added advantage of the ICF is its compatibility with various questionnaires used to assess treatment response. These questionnaires have been linked to ICF codes, enabling the conversion of questionnaire responses into ICF code qualifiers. Many of these linkages have been scientifically validated, and comprehensive guides are available to establish new linkages for emerging datasets [28,38].

The proposed methodology consolidates information from various heterogeneous international sources, including interviews or anamnesis, validated questionnaires, physical tests, cognitive tests, etc, into a single, easily accessible location. This compatibility and consolidation process is crucial for advancing AI solutions. The training of ML algorithms requires identical variables from all individuals analyzed. Without dataset homogenization, only subsets of individuals and/or variables can be analyzed concurrently, reducing the predictive power of the trained models.

A single numerical value can be a useful tool for visualizing the individual's condition before and after treatment and for tracking rehabilitation progress. However, to provide a more comprehensive view of an individual's situation, we also aim to delve into the main categories at a more detailed level, rather than inferring a person's functional capacity solely based on a single numerical value. Consequently, our proposed method is adaptable for computing a health profile. Within the ICF context, a basic health profile would likely include separate scores for each of the 4 ICF components: *body functions*, *body structures*, *activities and participation*, and *environmental factors*. These components can be further expanded for a more detailed examination of the elements within each category, providing comprehensive information about the individual.

Our goal is to make the health index and health profile highly illustrative, catering to both professionals and individuals undergoing treatment. If extensive data are recorded for an individual in the ICF, it can be challenging to assess and quickly understand the individual's situation. The health index and health profile of an individual can also be presented visually, as shown in Figure 10, providing a concise overview of the individual's condition.

Such visualizations can be integrated into electronic health records (EHRs) or patient portals, allowing both clinicians and patients to engage with the data. For clinicians, this supports efficient triage and treatment planning; for patients, it enhances understanding and motivation by making health status tangible and actionable. As the index is based on ICF codes, it can be integrated into existing EHR systems and used across institutions and countries, supporting interoperability and standardized health assessment.

The health profile allows for a deeper evaluation of specific domains of functioning, which can be particularly motivating by highlighting areas of improvement and those requiring attention. Previously, extracting and presenting such information was labor-intensive and often difficult for patients to interpret. The proposed model simplifies this process and improves accessibility. Additionally, a data coverage metric can help assess the completeness of the health profile, guiding clinicians in identifying gaps in assessment. In Figure 10, coverage is visually represented using transparency: low transparency indicates high data availability, while high transparency signals limited coverage. This visual cue supports informed decision-making and encourages targeted data collection to improve the reliability of the health index.

In current clinical practice, not all potential ICF subcategories are used. However, we continue to strive for a holistic understanding of an individual's functional capacity. This is achieved through a combination of questionnaires, interviews, and tests, allowing us to gather extensive information, even if some ICF subcategories remain unfilled. For instance, when data from specific areas such as strength tests are entered into

XSL•FO
RenderX

the system, it automatically assigns the appropriate ICF codes based on the results. Our proposed health index framework serves as a tool to standardize this process, thereby improving the efficiency and consistency of data collection and analysis in clinical settings.

Future Work

Our ongoing efforts to enhance the index involve addressing the existing limitations outlined earlier. Specifically, we aim to improve the support for qualifiers beyond the initial qualifier and implement mechanisms to alert index users about insufficient data. Additionally, a step forward would be conducting a clinical validation of the index. Whereas the statistical analysis presented in this study provides an initial glimpse into the index's functionality, a comprehensive clinical assessment will be needed to validate its practical utility.

A concrete example of the index usage is in the field of rehabilitation, where an iterative cycle based on the ICF, known as the Rehab-Cycle, has been developed for streamlined rehabilitation management [43]. This cycle describes different phases of rehabilitation: assessment, assignment, intervention, and evaluation. The use of the health index facilitates the generation of a compact numerical measure for both the preintervention assessment and postintervention evaluation in clinical practice. Furthermore, this index can assist in enhancing resource allocation and selecting appropriate interventions during the assignment and intervention phases. As a result, it allows for a more thorough evaluation of rehabilitation effectiveness and enables comparisons across various countries and rehabilitation methodologies.

Beyond rehabilitation, the health index has potential applications in broader clinical contexts. In primary care, it could support routine health assessments by providing a standardized snapshot of patient functioning. In geriatrics, it may help monitor age-related decline and guide interventions for healthy aging. Although intrinsic capacity (IC) was not directly assessed in this study, our approach aligns closely with the IC framework, and future work may explore how the proposed health index could support IC-oriented monitoring and proactive care models as outlined in recent WHO digital health initiatives. In mental health, the index could be used to track functional outcomes and support personalized treatment planning. Integration into EHRs or patient-facing platforms could enable real-time monitoring, enhance clinician-patient communication, and support shared decision-making.

With the prospective incorporation of ML techniques, the health index or health profile scores could serve as optimization targets within predictive models. This strategy allows for a comprehensive improvement of an individual's health, rather than concentrating on a single health parameter. For instance, it could assist in determining the optimal rehabilitation pathway or treatment strategy for an individual. Additionally, the standardized nature of the index supports cross-institutional and international data harmonization, which is essential for scalable AI applications in health informatics.

Acknowledgments

We gratefully acknowledge the anonymous reviewers for their valuable comments on the initial version of the article.

Funding

IR's work was funded by David Health Solutions Ltd., Jenny and Antti Wihuri Foundation (grant 00180312, 00200302, and 00210295), and Academy of Finland (decision 356158).

Authors' Contributions

IR contributed to methodology, software, validation, formal analysis, investigation, writing of the original draft, and visualization. LP contributed to conceptualization, methodology, and data curation. VJ contributed to data curation, reviewing and editing, and project administration. SÄ contributed to conceptualization, methodology, reviewing and editing, and supervision. JPK contributed to conceptualization, methodology, software, reviewing and editing, visualization, and supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details on data linkages.
[DOCX File, 23 KB - humanfactors_v13i1e84802_app1.docx]

Multimedia Appendix 2
Details on defining weightings and normalization.
[DOCX File, 109 KB - humanfactors_v13i1e84802_app2.docx]

References

1. McDowell I. Measuring Health: A Guide to Rating Scales and Questionnaires: Oxford University Press; 2006.

XSL•FO
RenderX

2. Fava GA, Sonino N. Psychosomatic medicine: emerging trends and perspectives. Psychother Psychosom 2000;69(4):184-197. [doi: 10.1159/000012393] [Medline: 10867586]
3. Cesari M, Araujo de Carvalho I, Amuthavalli Thiyagarajan J, et al. Evidence for the domains supporting the construct of intrinsic capacity. J Gerontol A Biol Sci Med Sci 2018 Nov 10;73(12):1653-1660. [doi: 10.1093/gerona/gly011] [Medline: 29408961]
4. World Health Organization International Classification of Functioning, Disability and Health: ICF: World Health Organization; 2001.
5. ICF beginner's guide: towards a common language for functioning, disability and health. World Health Organization. 2002. URL: https://www.who.int/publications/m/item/icf-beginner-s-guide-towards-a-common-language-for-functioning-disability-and-health [accessed 2026-06-01]
6. Beyene MB, Visvanathan R, Amare AT. Intrinsic capacity and its biological basis: a scoping review. J Frailty Aging 2024;13(3):193-202. [doi: 10.14283/jfa.2024.30] [Medline: 39082762]
7. Xuan Z, Niu Y, Wang Y, et al. Perspectives and preferences on developing a digital human system for intrinsic capacity monitoring underpinned by World Health Organization's integrated care for older people guideline: qualitative study. J Med Internet Res 2025 Nov 11;27:e76222. [doi: 10.2196/76222] [Medline: 41217817]
8. Maini M, Nocentini U, Prevedini A, Giardini A, Muscolo E. An Italian experience in the ICF implementation in rehabilitation: preliminary theoretical and practical considerations. Disabil Rehabil 2008;30(15):1146-1152. [doi: 10.1080/09638280701478397] [Medline: 19230223]
9. Scheuringer M, Kirchberger I, Boldt C, et al. Identification of problems in individuals with spinal cord injury from the health professional perspective using the ICF: a worldwide expert survey. Spinal Cord 2010 Jul;48(7):529-536. [doi: 10.1038/sc.2009.176] [Medline: 20065988]
10. Wiegand NM, Belting J, Fekete C, Gutenbrunner C, Reinhardt JD. All talk, no action?: the global diffusion and clinical implementation of the international classification of functioning, disability, and health. Am J Phys Med Rehabil 2012 Jul;91(7):550-560. [doi: 10.1097/PHM.0b013e31825597e5] [Medline: 22561387]
11. Pernambuco AP, Lana RDC, Polese JC. Knowledge and use of the ICF in clinical practice by physiotherapists and occupational therapists of Minas Gerais. Fisioter Pesqui 2018 Jun;25(2):134-142. [doi: 10.1590/1809-2950/16765225022018]
12. Kubik MY, Lytle LA, Hannan PJ, Story M, Perry CL. Food-related beliefs, eating behavior, and classroom food practices of middle school teachers. J Sch Health 2002 Oct;72(8):339-345. [doi: 10.1111/j.1746-1561.2002.tb07921.x] [Medline: 12389375]
13. Frank E, Carrera JS, Elon L, Hertzberg VS. Predictors of US medical students' prevention counseling practices. Prev Med 2007 Jan;44(1):76-81. [doi: 10.1016/j.ypmed.2006.07.018] [Medline: 16978687]
14. Worldwide research methodology and codebook. : Gallup, Inc; 2023 URL: https://www.unicef.org/innocenti/media/3346/file/Gallup-World-Poll-Methodology-2023.pdf [accessed 2026-06-10]
15. Meijer E, Kapteyn A, Andreyeva T. Internationally comparable health indices. Health Econ 2011 May;20(5):600-619. [doi: 10.1002/hec.1620] [Medline: 20572201]
16. Yi SI, So BR, Lee CS, et al. Classification of health grade using bio-check unit and health index. Journal of Biomechanical Science and Engineering 2011;6(3):148-159. [doi: 10.1299/jbse.6.148]
17. Kohn JL. What is health? A multiple correspondence health index. Eastern Econ J 2012 Feb;38(2):223-250. [doi: 10.1057/eej.2011.5]
18. Poterba JM, Venti SF, Wise DA. Health, education, and the post-retirement evolution of household assets. National Bureau of Economic Research. 2013. [doi: 10.3386/w18695]
19. van der Gaag J, Wolfe BL. Estimating demand for medical care: health as a critical factor for adults and children. In: Econometrics of Health Care: Springer; 1991:31-58. [doi: 10.1007/978-94-009-2051-4_3]
20. Wagstaff A. The demand for health: some new empirical evidence. J Health Econ 1986 Sep;5(3):195-233. [doi: 10.1016/0167-6296(86)90015-9] [Medline: 10279033]
21. Kim MK, Jung HT, Kim SD, La HJ. A personal health index system with IoT devices. Presented at: 2016 IEEE International Conference on Mobile Services (MS); Jun 27 to Jul 2, 2016; San Francisco, CA p. 174-177. [doi: 10.1109/MobServ.2016.34]
22. Chen L, Li X, Yang Y, et al. Personal health indexing based on medical examinations: a data mining approach. Decis Support Syst 2016 Jan;81:54-65. [doi: 10.1016/j.dss.2015.10.008]
23. Lai G, Yu D, Zhang S, Wei Z, Sun X. Personal health index based on residential health examination. In: Yang X, Wang CD, MdS I, Zhang Z, editors. Advanced Data Mining and Applications (ADMA 2020): Springer International Publishing; 2020:569-581. [doi: 10.1007/978-3-030-65390-3_43]
24. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. Physiotherapy 1980 Aug;66(8):271-273. [Medline: 6450426]
25. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec;20(10):1727-1736. [doi: 10.1007/s11136-011-9903-x] [Medline: 21479777]
26. Williamson A, Hoggart B. Pain: a review of three commonly used pain rating scales. J Clin Nurs 2005 Aug;14(7):798-804. [doi: 10.1111/j.1365-2702.2005.01121.x] [Medline: 16000093]

XSL•FO
RenderX

27. Cieza A, Brockow T, Ewert T, et al. Linking health-status measurements to the International Classification of Functioning, Disability and Health. J Rehabil Med 2002 Sep;34(5):205-210. [doi: 10.1080/165019702760279189] [Medline: 12392234]
28. Cieza A, Fayed N, Bickenbach J, Prodinger B. Refinements of the ICF linking rules to strengthen their potential for establishing comparability of health information. Disabil Rehabil 2019 Mar;41(5):574-583. [doi: 10.3109/09638288.2016.1145258] [Medline: 26984720]
29. The ethical principles of research with human participants and ethical review in the human sciences in Finland: Finnish National Board on Research Integrity TENK Guidelines 2019. : Finnish National Board on Research Integrity TENK; 2019 URL: https://tenk.fi/sites/default/files/2021-01/Ethical_review_in_human_sciences_2020.pdf [accessed 2026-06-08]
30. Python programming language, version 3. Python Software Foundation. URL: https://www.python.org [accessed 2026-06-08]
31. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020 Mar;17(3):261-272. [doi: 10.1038/s41592-019-0686-2] [Medline: 32015543]
32. Hunter JD. Matplotlib: a 2D graphics environment. Comput Sci Eng 2007;9(3):90-95. [doi: 10.1109/MCSE.2007.55]
33. Tosin M, Côrtes AMA, Cunha A. A tutorial on Sobol' global sensitivity analysis applied to biological models. In: N C, Santos MTD, Lopes FJP, editors. Networks in Systems Biology: Applications for Disease Modeling: Springer International Publishing; 2020:93-118. [doi: 10.1007/978-3-030-51862-2_6]
34. Cieza A, Oberhauser C, Bickenbach J, Chatterji S, Stucki G. Towards a minimal generic set of domains of functioning and health. BMC Public Health 2014 Mar 3;14(1):1-9. [doi: 10.1186/1471-2458-14-218] [Medline: 24588794]
35. Cieza A, Ewert T, Ustün TB, Chatterji S, Kostanjsek N, Stucki G. Development of ICF Core Sets for patients with chronic conditions. J Rehabil Med 2004 Jul(44 Suppl):9-11. [doi: 10.1080/16501960410015353] [Medline: 15370742]
36. Salomon JA, Mathers CD, Chatterji S, Sadana R, Üstün TB, Murray CJ. Quantifying individual levels of health: definitions, concepts and measurement issues. In: Murray CJL, Evans DB, editors. Health Systems Performance Assessment: Debates, Methods and Empiricism: World Health Organization; 2003:301-318.
37. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JPA. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. J Am Med Inform Assoc 2017 Jan;24(1):198-208. [doi: 10.1093/jamia/ocw042] [Medline: 27189013]
38. How to use the ICF - a practical manual for using the International Classification of Functioning, Disability and Health. World Health Organization. URL: https://www.who.int/publications/m/item/how-to-use-the-icf---a-practical-manual-for-using-the-international-classification-of-functioning-disability-and-health [accessed 2026-06-01]
39. Negri L, Spoladore D, Fossati M, et al. Proposal for an ICF-based methodology to foster the return to work of persons with disability. Work 2023;74(2):649-662. [doi: 10.3233/WOR-211226] [Medline: 36278385]
40. Geyh S, Peter C, Müller R, et al. The personal factors of the International Classification of Functioning, Disability and Health in the literature – a systematic review and content analysis. Disabil Rehabil 2011;33(13-14):1089-1102. [doi: 10.3109/09638288.2010.523104] [Medline: 20925452]
41. Leonardi M, Lee H, Kostanjsek N, et al. 20 years of ICF-International Classification of Functioning, Disability and Health: uses and applications around the world. Int J Environ Res Public Health 2022 Sep 8;19(18):11321. [doi: 10.3390/ijerph191811321] [Medline: 36141593]
42. Madden RH, Bundy A. The ICF has made a difference to functioning and disability measurement and statistics. Disabil Rehabil 2019 Jun;41(12):1450-1462. [doi: 10.1080/09638288.2018.1431812] [Medline: 29433362]
43. Rauch A, Cieza A, Stucki G. How to apply the International Classification of Functioning, Disability and Health (ICF) for rehabilitation management in clinical practice. Eur J Phys Rehabil Med 2008 Sep;44(3):329-342. [Medline: 18762742]

Abbreviations

EHR: electronic health record
EQ-VAS: EuroQol Visual Analogue Scale
IC: intrinsic capacity
ICD: International Classification of Diseases
ICF: International Classification of Functioning, Disability and Health
MCA: multiple correspondence analysis
ML: machine learning
PCA: principal component analysis
WHO: World Health Organization

XSL•FO
RenderX

Edited by A Kushniruk; submitted 08.Oct.2025; peer-reviewed by N Yung, Y Bobrova; revised version received 09.Apr.2026; accepted 04.May.2026; published 02.Jul.2026.

Please cite as:
Rautiainen I, Parviainen L, Jakoaho V, Äyrämö S, Kauppi JP
Monitoring Health Status: Development and Preliminary Validation of a Personal Health Index Using the International Classification of Functioning, Disability and Health
JMIR Hum Factors 2026;13:e84802
URL: https://humanfactors.jmir.org/2026/1/e84802
doi:10.2196/84802

© Ilkka Rautiainen, Lauri Parviainen, Veera Jakoaho, Sami Äyrämö, Jukka-Pekka Kauppi. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 2.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Development and Psychometric Validation of a Usability Instrument Based on ISO 25010 for Electronic Health Record Systems in Peruvian Health Care Settings: Methodological Study

David Mamani-Pari[*], MSc; Alex Froilan Peñaloza-Hualpa[*], Bach; Milton Edward Humpiri-Flores[*], Dr; Eder Gutierrez-Quispe, Mg; Jorge Eddy Otazu-Luque, Mg

Escuela Profesional de Ingeniería de Sistemas, Facultad de Ingeniería y Arquitectura, Universidad Peruana Unión, Carretera Arequipa Km-6, Juliaca, Peru

[*]these authors contributed equally

Corresponding Author:
David Mamani-Pari, MSc
Escuela Profesional de Ingeniería de Sistemas, Facultad de Ingeniería y Arquitectura, Universidad Peruana Unión, Carretera Arequipa Km-6, Juliaca, Peru

Abstract

Background: The usability of electronic health records influences patient safety, clinical efficiency, and the sustainable digital transformation of health care. While tools such as the System Usability Scale exist, few are based on structured software quality models or align with local regulatory frameworks, particularly in low- and middle-income countries.

Objective: This study aims to develop and psychometrically validate an instrument to assess electronic health record usability, based on the ISO 25010 (International Organization for Standardization) quality model, and in accordance with the Peruvian Technical Health Standard (Norma Técnica de Salud [NTS]).

Methods: A 3-stage methodological study was conducted following COSMIN (Consensus-based Standards for the Selection of Health Measurement Instruments) guidelines. First, a 23-item instrument was developed, grouped into 7 usability dimensions based on ISO 25010. Second, content validity was assessed using a modified Delphi method with 10 experts, using Aiken's V coefficient. Third, structural validity and reliability were evaluated in 115 participants (physicians, nurses, and administrative staff) using confirmatory factor analysis and Cronbach α. Additionally, the module's functional compliance with the NTS was verified.

Results: The instrument demonstrated excellent psychometric properties among 115 participants (53 physicians, 48 nurses, and 14 administrative staff); high content validity (Aiken's V=0.95; 95% CI 0.84 - 0.98), excellent internal consistency (α=.968), and optimal factor fit (Comparative Fit Index=1.000; Tucker-Lewis Index=1.001). Overall usability was perceived as very high by 94 of 115 users (81.7%), and high by 21 of 115 users (18.3%), with no perceptions at lower levels. The module met 28 of 35 requirements (80%) of the Peruvian Technical Health Standard. Administrative staff reported higher satisfaction (13/14, 93% rated it as very high) compared with clinical professionals (81/101, 80% rated it as very high), particularly in the dimensions of error protection and operability.

Conclusions: This study provides the first fully psychometrically validated instrument in Spanish based on ISO 25010, System Usability Scale, and other theoretical foundations for evaluating the usability of information systems. The results highlight the importance of considering both international standards and local regulatory requirements (NTS) in the design of digital health systems. The instrument is applicable in Spanish-speaking contexts and can serve as a reference for future usability evaluations in the region as well as in low- and middle-income countries.

(JMIR Hum Factors 2026;13:e81377) doi:10.2196/81377

KEYWORDS

usability scale; human factors; electronic health record; ISO 25010; Peruvian health system; usability evaluation

XSL•FO
RenderX

Introduction

Digitization and Usability in Electronic Health Record Systems

Electronic health record (EHR) systems have become essential tools for improving the quality of medical care, reducing errors, and optimizing clinical processes [1]. However, their effectiveness depends critically on their usability, particularly in high-workload health care settings such as Peru, where health care professionals face time and resource constraints [2]. From the patient's perspective, studies have also been conducted on EHRs, highlighting the physician's central role within the workflow. Evidence shows that physicians who use EHRs to exchange information with their patients, develop care plans, and provide other services tend to report higher satisfaction with computerized systems [3]. On the other hand, there are also studies revealing that the adoption of health information systems has led to medical errors attributed to technological failures, although there are certain perceptual discrepancies between clinical and information technology (IT) professionals regarding their causes. While IT experts attribute errors to technical failures such as poor interfaces and lack of training, physicians point to contextual and human factors such as workload, familiarity with new systems, and inadequate processes [4]; it is important to note that physicians work under very tight schedules and require IT tools that truly provide support; therefore, there are various challenges to be addressed in different parts of the world depending on culture, technological capacity, socioeconomic status, government policies, and other factors; thus, it remains a subject of study for various researchers [5-9].

Context of the Peruvian Health Care System

Similarly, Peru faces challenges; given that the existence of fragmented medical records or the inability to locate a medical record—even within the same Healthcare Service Provider Institution (Instituciones Prestadoras de Servicios de Salud [IPRESS])—creates problems of information fragmentation and wasted man-hours in human resources due to the need to repeatedly fill out a new medical record, which hinders the provision of comprehensive, quality health care due to the lack of important information from previous medical records; this same issue is mentioned in the Peruvian government's Technical Implementation Document for the National Registry of Electronic Medical Records (Registro Nacional de Historias Clínicas Electrónicas [RENHICE]) in section 6.3, which states the following: "In our country, almost all medical records are handwritten. If a patient visits an IPRESS for the first time, a new medical record is opened, which involves recording data (administrative and clinical) on forms that are part of the medical record. On the other hand, when the patient returns for a subsequent visit, the staff cannot locate their manual medical record (Historia Clínica Manual [HCM]); therefore, they must re-enter the data. Likewise, the procedure is repeated when a patient changes their address, changes IPRESS facilities, or is located anywhere else in the country. Currently, a patient has as many medical records as the number of IPRESS facilities they have visited, and they may have more than one record at the same IPRESS facility. In response to this problem, the government has established the Technical Health Standard (Norma Técnica de Salud [NTS]) for Medical Record Management; however, this cannot be resolved without an adequate computer system to assist in managing medical records within a specific IPRESS—a system that can be adapted to the institution's specific needs while adhering to the Peruvian government's NTS, given that it applies to all IPRESSes, whether public, private, or mixed.

Regulatory Framework: NTS and the Need for Regulatory Alignment

Therefore, there is a need for a Clinical Record Management Information System that incorporates the criteria or considerations of the National Technical Standard (NTS) for Clinical Record Management; this system must meet usability criteria so that users are satisfied when using this tool. According to [10], the ease of use of a system can reduce errors, thereby contributing to efficiency and allowing physicians to spend more time with patients; hence the importance of evaluating usability, since studies cited in the research indicate that physicians experience usability issues, particularly when working with EHRs, which result in lengthy training periods and lost productivity. It is worth noting that usability could mean the difference between life and death; furthermore, it is argued that usability issues arise in medical devices, procedures, and even diagnostic tools; therefore, there is a need to conduct a study on the evaluation of the impact of NTS on the implementation of the Clinical Records Module in relation to its usability.

Usability Gaps and Human Factors in EHRs

According to studies, what is now known as electronic medical records (EMR) has been in use since the 1980s, primarily in developed countries. Subsequent studies were conducted on their use, as it was perceived that physicians were resistant to change; consequently, it was observed that physicians adopt the use of these systems once they perceive their utility for their patients; thus, it was concluded that the key lies in ensuring that the software tools are easy to use and highly beneficial for the medical team [11]. The evaluation of EMR management tools is almost continuous, as it positively or negatively affects the care provided to patients by physicians. Therefore, they report having surveyed users—primarily physicians and nurses who interact directly with the tool in question—and found that while the tool facilitates greater data capture, it also impacts medical staff time, as they spend more time at the computer, which in some ways reduces the time available for patient care; consequently, they concluded that the tool has minimal impact on patient care. On the other hand, they showed minimal satisfaction with the tool; however, there was resistance to returning to handwritten records [12]. Another study also notes a minimal impact from the implementation of EMR systems, and the implementation of EMR management software in health care facilities is increasingly widespread [13].

Cultural diversity, economic conditions, and other factors influence the adoption of technological alternatives; therefore, it is recommended to have a plan that coordinates the entire EMR implementation process, considering the current

XSL•FO
RenderX

organizational context, given that some countries face challenges, particularly those that are underdeveloped or developing [14]. This implies restructuring the curricula for training health care professionals, as well as providing EMR training to all staff at the institution [15]. On the other hand, studies are also being conducted on the impact of the EHR on patients by allowing them access to their records, thereby enhancing safety in medication management and providing greater control over medication [16].

The need for EHR management platforms in the health care systems of various countries is critical, as, despite the challenges, they are cost-effective to maintain because they enable faster searches, reducing the time and cost of locating patient records; consequently, they are being widely adopted worldwide [17]. These platforms can be classified into several types, such as government, commercial, organizational, or multi-organizational; however, it is important to highlight certain key characteristics they must meet, such as data interoperability, displaying the appropriate level of detail in information, and compliance with health care privacy regulations; the aforementioned aspects still pose a challenge in terms of compliance [18].

The transition to the EHR has been a fundamental advancement in the medical field, as it allows for the consolidation of clinical information, improves evidence-based decision-making, enhances access to information, and facilitates multidisciplinary collaboration. However, the implementation of EHRs has revealed significant challenges regarding efficiency, reliability, and the potential increase in medical-legal risk; that is, it is not a perfect solution; as medical malpractice lawsuits involving EHRs highlight documentation errors caused by various factors such as information fragmentation, access failures during system outages, issues with electronic data transfer, ineffective alerts, and a lack of integration among hospital systems (interoperability); which can compromise patient care and expose providers to legal liability. Therefore, the correct completion of documentation in the EHR is of utmost importance; errors in data entry generally occur due to the use of prefilled templates or the copying of unverified data; this necessitates continuous improvement of EHR systems to optimize their functionality, as well as the establishment of protocols to ensure that records are complete, accurate, and faithfully reflect the clinical assessment. In this way, it ensures that EHR platforms are not merely tools for storing information, but rather a comprehensive system supporting medical care that considers medical-legal aspects and patient safety [19]; hence the importance of EHR systems adhering to health sector regulations. It is also worth noting that there is a study evaluating the suitability of EHR data using a framework called ISPOR (International Society for Pharmacoeconomics and Outcomes Research) to ensure the relevance and validity of the information generated by EHRs; this framework includes 2 components: data delineation and data suitability for the purpose; each component involves a series of checklists [20].

Apart from the challenges mentioned above, interoperability is a critical factor for the success of health information systems, as it enables the secure and efficient exchange of patient data among different health care providers; however, there are many challenges with the integration and the automatic and secure exchange of information among the various EHR health care systems [21]. Consequently, organizations often face challenges in achieving successful technology implementation; therefore, it is of vital importance to conduct a comprehensive analysis of the opinions of internal and external users regarding EHR software decisions; this can help improve and optimize EHR information systems and reduce technology acceptance gaps [22].

The objective of this research was to design and psychometrically validate a usability evaluation instrument based on the ISO/IEC 25010 (International Organization for Standardization/International Electrotechnical Commission) standard for EHR systems in the Peruvian context. Specifically, the study sought to: (1) integrate the usability dimensions of ISO 25010 with the System Usability Scale (SUS) and other relevant theoretical foundations, constructing a comprehensive instrument in Spanish; (2) validate the instrument through expert judgment using Aiken's V coefficient and confirm its psychometric properties through confirmatory factor analysis (CFA) and reliability analysis; and (3) evaluate the usability of an EHR module implemented in accordance with the Peruvian NTS, identifying the degree of regulatory compliance and differences in the perception of usability among different types of health care professionals (physicians, nurses, and administrative staff). This study aims to fill the gap identified in the literature regarding the lack of standardized instruments in Spanish, with psychometric validations that integrate international quality standards with local health sector regulations, providing a validated tool that can be replicated in similar contexts in Latin America, such as in low- and middle-income countries.

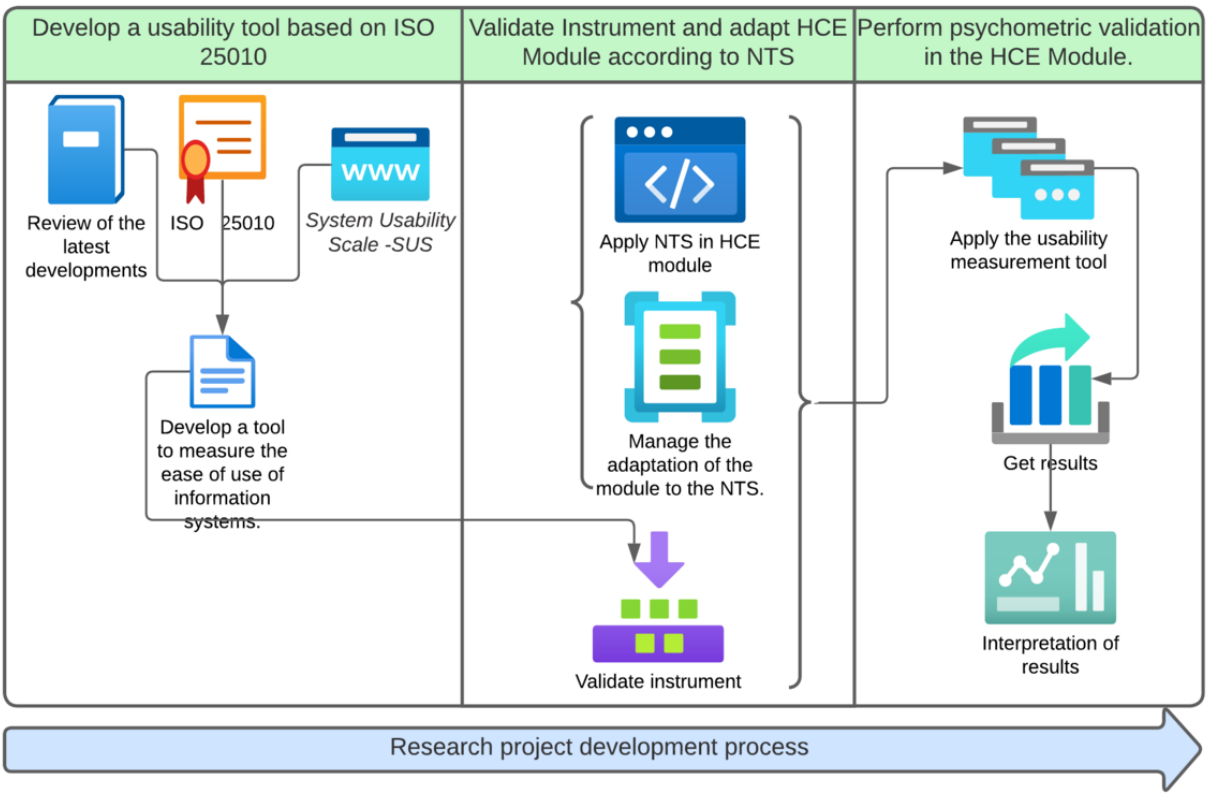
Methods

Study Design

This study followed a 3-stage sequential instrument development and validation methodology, aligned with the COSMIN (Consensus-based Standards for the Selection of Health Measurement Instruments) guidelines for psychometric studies in health. The first stage (theoretical development) integrated the usability attributes of ISO 25010, SUS, and other theoretical foundations on usability. The second stage (content validation) used the modified Delphi method with 10 experts, in addition to aligning the functionality of the electronic health record (Historia Clínica Electrónica [HCE]) module with the functional requirements of the Peruvian Technical Health Standard (NTS), while the third stage (structural validation and reliability) included the application of the instrument to 115 end users; this process is shown in Figure 1.

XSL•FO
RenderX

Figure 1. Research development methodology. HCE: Electronic Health Record (Historia Clínica Electrónica); ISO: International Organization for Standardization; NTS: Norma Técnica de Salud (Technical Health Standard); SUS: System Usability Scale.



Participation and Context

For content validation, a total of 10 experts were deliberately recruited, distributed as follows: 5 specialists in areas related to software development, 3 researchers with experience in systems evaluation, and 2 professionals from the health care sector. This selection aimed to incorporate the usability quality attribute, in accordance with the guidelines established in the ISO/IEC 25010 standard and the SUS questionnaire. This methodological approach has been previously used in similar studies [23,24].

For structural validation and reliability (psychometric), the study obtained an institutional authorization letter and involved 115 workers from a hospital network in the Puno region, distributed across the following professional categories: physicians (53/115, 46%), nurses (48/115, 42%), and administrative staff (14/115, 12%), all with at least 6 months of experience using the EHR module. The sample size was determined in accordance with the recommendation of 5 participants per item of the instrument (23 items → n=115), meeting the methodological criteria required for the application of CFA.

Likewise, it should be noted that ISO/IEC 25010:2011, which replaces the previous ISO/IEC 9126-1:2001, defines 2 models: a quality-in-use model and a product quality model, each with its corresponding characteristics. Regarding the in-use quality model, the aim is to quantify the “usability” of the application strictly related to the end user of the product [25].

Development of the Instrument

The initial questionnaire consisted of 23 items organized into 7 dimensions, which were based on the quality model of the ISO/IEC 25010 standard: (1) functional adequacy, (2) learnability, (3) operability, (4) protection against user errors, (5) esthetic presentation, (6) accessibility [26], and (7) user satisfaction. Theoretical support regarding instruments and techniques for measuring “usability” was sought, and the decision was made to develop a Spanish-language instrument based on ISO/IEC 25010:2011 and supported by the SUS and other instruments addressing the subject in question. Subsequently, the modified Delphi technique [27,28] was followed, since the traditional technique was not strictly adhered to; however, given the nature of an instrument already theoretically grounded, supported by a review of the state of the art, at least 2 rounds were necessary to obtain adequate feedback from 10 expert judges who were selected as described in the previous subsection.

Content Validity

The experts evaluated each item based on 3 criteria (relevance, appropriateness, and clarity) using a 5-point Likert scale (where 1 represents the minimum value and 5 represents the maximum value). The results were analyzed using Aiken’s V coefficient, with values ≥0.80 considered acceptable. All items exceeded this threshold (average V=0.95), although some items regarding accessibility required reformulation after the first round of evaluation. The Delphi process was concluded after 2 rounds

XSL•FO
RenderX

upon reaching stability in the responses (change <5% between rounds).

Likewise, during this stage, improvements were made to the EMRs module, prioritizing compliance with the Peruvian State Technical Health Standard; a compliance schedule was developed for task management; and finally, a checklist was created comparing the requirements outlined in the NTS with the functionalities included in the EHR module to assess the level or degree of implementation of the requirements within the module; Subsequently, functional tests were conducted prior to the system going live and its subsequent use by system users, such as doctors, nurses, and professionals from other areas.

Structural Validity and Reliability

Data Collection

The usability instrument was administered virtually via Google Forms, in accordance with the recommendations of the CHERRIES checklist for electronic survey reporting [29]. Participation was voluntary, and informed consent was obtained digitally prior to the survey. The survey was disseminated internally to health care professionals within the hospital network and administrative staff through corporate communication channels. No monetary incentives were provided. The survey consisted of 23 mandatory questions, and incomplete surveys were excluded. To prevent multiple submissions, controlled distribution through institutional corporate channels served as a restriction mechanism. Email addresses were not collected, nor were IP addresses logged, to maintain participant anonymity. A total of 115 complete responses were included in the final analysis. The response rate could not be calculated due to the institutional distribution methods. Data were exported directly from Google Forms to a secure database and analyzed using Jamovi software.

Statistical Analysis

The instrument was administered via an electronic form (Google Forms) during October-November 2024. The data were analyzed using Jamovi 2.3.28 [30]. Finally, construct validity and reliability were assessed using Cronbach α and CFA [31,32]; the latter was conducted following the standard recommendation of using between 5 and 10 participants per item or at least 100 participants [33]. In this study, we opted for 5 participants per item for CFA, given that the study population met this requirement at the facility where the case study was conducted.

Ethical Considerations

This study was not subject to full review by the ethics committee, as it meets the requirements for exemption under applicable Peruvian national regulations. Specifically, according to Ministerial Resolution No. 233-2020-MINSA, which approves the “Ethical Considerations for Health Research Involving Human Subjects” issued by the Peruvian Ministry of Health (Ministerio de Salud [MINSA]), survey-based studies that are anonymous and noninterventional are classified as low-risk or risk-free and, therefore, do not require full approval by the ethics committee. Furthermore, the study complied with Peru’s Law No. 29733 (Personal Data Protection Law) and its Regulatory Decree D.S. No. 003-2013-JUS, which guarantees the confidentiality and anonymity of participants’ data. Likewise, prior to data collection, written institutional authorization was obtained from the American Clinic of Juliaca (January 18, 2024). Informed consent was also included directly in the web-based survey instrument. Participants were informed that their responses would be collected anonymously and used exclusively for research purposes, and that their continued participation constituted implied consent. We confirm that this study falls under category (2): it was not submitted for ethical review, as it met the exemption criteria due to its anonymous, noninterventional, and survey-based nature, in accordance with the aforementioned national policy. Participants did not receive any financial or other compensation for their participation in this study.

Results

Instrument Developed

Following a review of the state of the art and validation by expert judges in 2 rounds, the Spanish-language instrument based on ISO 25010 was validated, as shown in Multimedia Appendix 1 [34-41].

Content Validity Results

This instrument was the result of an evaluation by 10 expert judges who assigned a rating on a scale of 1 to 5, where 1 represents the minimum value, and 5 represents the maximum value; these results were processed using Aiken’s V statistic to assess the content validity of the aforementioned instrument with a 95% confidence level, corresponding to 1.96.

In accordance with the previous paragraph, favorable results were obtained, with the average Aiken’s V coefficient yielding a value of 0.95, indicating that the instrument has very good content validity; these results can be verified item by item in Multimedia Appendix 2, where the mean score is 4.8 and the SD is 0.37; the lower confidence interval limit is 0.84 and the upper confidence interval limit is 0.98; these results are consequently positive for the proposed instrument, which is intended to evaluate the usability of a computer tool.

Similarly, the Aiken V results were grouped by dimension, and these also showed favorable results, with the lowest score in the accessibility dimension (an average of 0.93) and the highest score in the operability dimension (an average of 0.97), as shown in Table 1.

XSL•FO
RenderX

Table . Aiken V results by dimension.

Dimension	Recognition of suitability	Learnability	Operability	User error protection	User interface aesthetics	Accessibility	Satisfaction	General
Relevance	0.958	0.931	0.983	0.96	0.93	0.93	0.95	0.949
Relevance	0.967	0.938	0.958	0.94	0.95	0.94	0.94	0.948
Clarity	0.950	0.969	0.975	0.97	0.95	0.93	0.95	0.957
TOTAL	0.958	0.946	0.972	0.956	0.944	0.933	0.947	0.951

Additionally, the medical records module has been adapted to meet the requirements of the Peruvian State Technical Health Standard (NTS); first, a list of requirements based on the NTS was compiled; subsequently, the corresponding adjustments were made to the EHR module; this was managed through a task schedule; and finally, a checklist was created comparing the requirements implemented according to the NTS with the functionalities of the module in question; resulting in an 80% implementation of the requirements, with 20% remaining pending based on the total number of requirements; these were classified according to the following options: YES (Implemented), NO (Not Implemented), and Partial (Partially Implemented), with scores of 1, 0, and 0.5, respectively. The aforementioned results can be found in Multimedia Appendix 3.

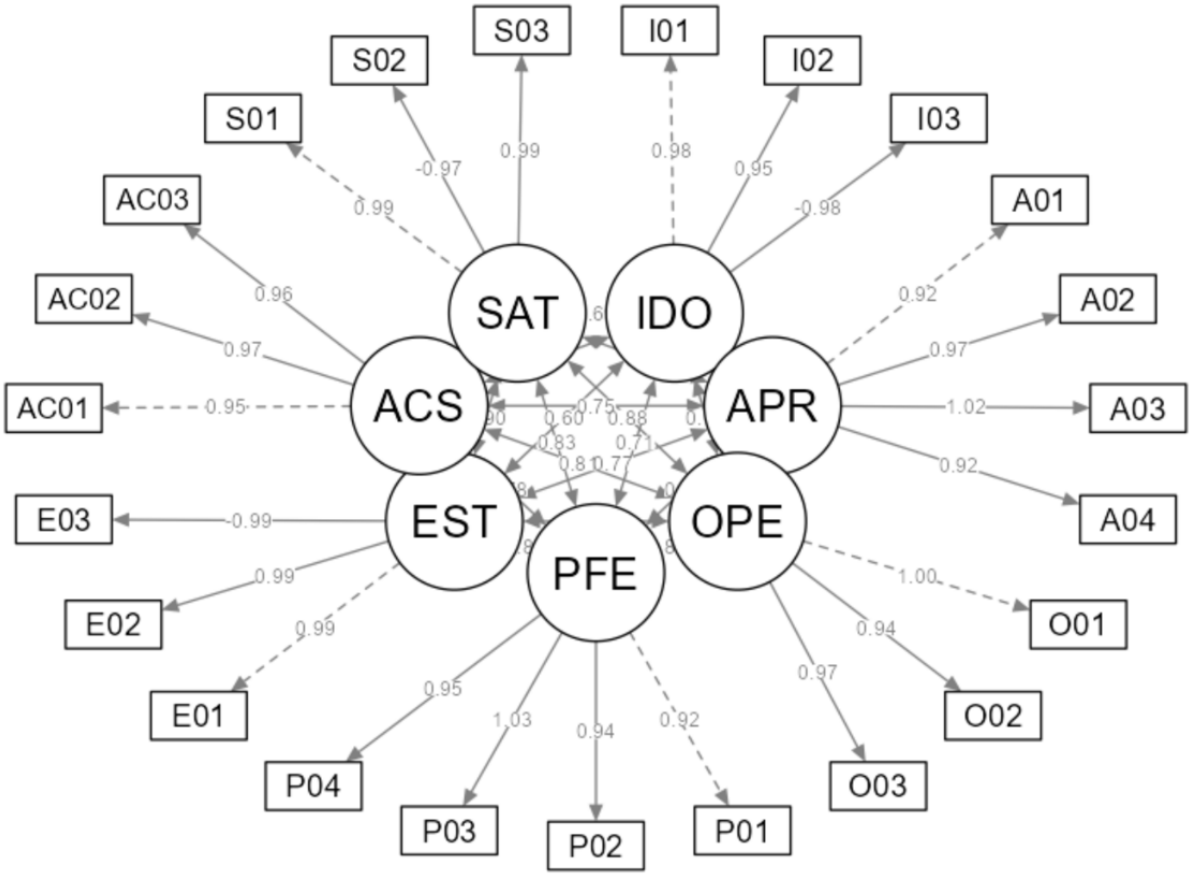
Psychometric Evaluation Results

The results of the pilot usability evaluation study with the developed instrument allowed for CFA, a statistical technique used to verify that the data from a model fit a predefined theoretical structure. Jamovi was used as the statistical analysis tool, yielding favorable results for the instrument in question: the CFI (Comparative Fit Index) has a value of 1, which measures model fit, and the scaled value is 0.994, indicating a perfect fit and a very good fit, respectively; likewise, the Tucker-Lewis Index has a value of 1.001 and the scaling of 0.993, which shows an excellent fit in both cases; similarly, this is the case for the Bentler-Bonett Norm-Normed Fit Index, where the model is 1.001 and the scaled value is 0.993, both indicating an excellent fit and a very robust model; on the other hand, the Relative Noncentrality Index measures relative noncentrality fit; the result for the model is 1.001 and the scaled value is 0.994, both of which also suggest an excellent fit; As for the NFI (Bentler-Bonett Normed Fit Index), which compares the model to the null model, the values are favorable, with the model at 0.997 and the scaled value at 0.980, indicating a base model with an excellent fit; Regarding the Bollen's Relative Fit Index, the model corresponds to 0.997 and the scaled value to 0.976, similarly indicating a good fit; these results can also be observed in Multimedia Appendix 4.

The results of the proposed usability measurement instrument model align with the predefined theoretical construct, as can also be seen in Figure 2.

XSL•FO
RenderX

Figure 2. Usability evaluation model. ACS: Accesibilidad (Accessibility); APR: Aprendizabilidad (Learnability); EST: Estética (Esthetics); IDO: Idoneidad (Appropriateness); OPE: Operatividad (Operability); PFE: Protección Frente a Errores (Error Protection); SAT: Satisfacción (Satisfaction).



Additionally, the instrument's reliability was calculated using Cronbach α based on the results, yielding a value of .968, as shown in the results processed by Jamovi (Multimedia Appendix 5). This demonstrates high internal consistency and a high correlation among the items in the measurement of the construct.

Finally, the results of applying the instrument developed in accordance with ISO 25010 were obtained; for this purpose, the same principles proposed by John Brook in the development of the scales were considered to subsequently generate the results by dimension and in general. Although there are several techniques for processing SUS results [42], it was deemed appropriate to work with the rating scale technique, given that most existing instruments in other areas use the rating scale technique; therefore, the score ranges shown in Table 2 were considered to obtain the results.

Table . Scales for evaluating usability and its dimensions.

Dimension	Very low	Low	Moderate	High	Very high
Recognition of suitability (3-15)	3 - 5	6 - 8	9 - 11	12 - 13	14 - 15
Learnability (4-20)	4 - 7	8 - 11	12 - 14	15 - 17	18 - 20
Operability (3-15)	3 - 5	6 - 8	9 - 11	12 - 13	14 - 15
User error protection (4-20)	4 - 7	8 - 11	12 - 14	15 - 17	18 - 20
User Interface Esthetics (3-15)	3 - 5	6 - 8	9 - 11	12 - 13	14 - 15
Accessibility (3-15)	3 - 5	6 - 8	9 - 11	12 - 13	14 - 15
Satisfaction (3-15)	3 - 5	6 - 8	9 - 11	12 - 13	14 - 15
Overall result					
Usability (23-115)	23 - 42	43 - 61	62 - 79	80 - 97	98 - 115

XSL•FO
RenderX

Based on the above rating scale table, the following descriptive statistics were obtained regarding the results for each dimension of the Usability Factor assessment tool for computer products, as shown in Multimedia Appendix 6.

Regarding the overall usability factor score based on the developed instrument, the following descriptive statistics were obtained: mean 4.82, median 5, mode 5.00, sum 554, SD 0.388, IQR 0.00, minimum 4, and maximum 5. These results indicate a high overall usability factor score.

Referring to the frequency table of the overall result, it can be seen that 81.7% of users perceive usability to be very high, and 18.3% of users perceive it to be high, thus demonstrating that the EHRs Module used by the institution in the case study meets the usability criteria for a software product, as shown in Table 3.

However, descriptive differences were found in usability scores among professional groups, with administrative staff obtaining slightly higher scores than clinical staff, as shown in Multimedia Appendix 7. No inferential statistical comparisons were performed.

Table . Frequency of the usability factor.

Usability	Counts	Percentage of total (%)	Cumulative percentage (%)
High	21	18.3	18.3
Very high	94	81.7	100.0

Discussion

Main Findings

This study successfully developed and validated a usability evaluation instrument based on ISO 25010 with excellent psychometric properties (Aiken's V=0.95, Cronbach α=0.968, CFI=1.000). Application of the instrument revealed that 94 of 115 users (81.7%) perceived the usability of the EHR module as very high, while the module met 28 of 35 requirements established by the Peruvian NTS (80%). Descriptive differences in usability perception were identified between clinical and administrative professionals. Highlighting three main findings: (1) Good results in psychometric properties, (2) positive perception of usability with slight differences between professional areas, and finally, partial compliance with the Peruvian NTS.

Beyond local application, this work supports the global conversation on the digital transformation of health care. The World Health Organization's Digital Health Strategy highlights the need for contextualized evaluation frameworks to ensure that digital health systems are safe, effective, and people-centered. In this vein, usability is not merely a software feature but a factor that influences clinical performance, patient safety, and professional well-being. Poor usability of RHRs has been linked globally to increased cognitive load, documentation burden, and burnout among clinical staff. By combining the usability criteria of the ISO 25010 standard with local regulatory compliance requirements, this work proposes a framework to evaluate not only the system's functionality but also the alignment between technological design and actual clinical workflows. However, most validated usability tools have been developed in high-income contexts and may not fully represent the sociotechnical realities of low- and middle-income countries, where infrastructure limitations, fragmented medical records, and evolving regulatory frameworks pose unique challenges. This instrument addresses this gap by operationalizing usability within a dual framework: international quality standards and national health regulations. This approach transforms usability scores into policy-driven digital health governance, making usability evaluation a tool for strengthening the health system, rather than merely a postimplementation metric.

Comparison With Other Research

Following a systematic review, similar studies have been identified in European countries; these place significant emphasis on measuring the acceptance of hospital HIS (Hospital Information Systems) [43] through models such as the technology acceptance model and unified theory of acceptance and use of technology, where nurses are the users with the highest ratings and training is identified as one of the determining factors. On the other hand [44], a study reveals that both the technology acceptance model and unified theory of acceptance and use of technology remain valid models for predicting the acceptance of HIS by health care professionals in hospital settings. Likewise [45-47], provided qualitative and mixed methods evidence regarding the perceptions of internal users regarding HIS in European hospitals. Meanwhile, in the Latin American context, there is a scarcity of studies related to the usability or acceptance of hospital HIS; likewise [48,49], confirm that, in general, there is significantly less scientific output in digital health due to structural reasons, given that scientific output is concentrated more in high-income countries [50,51]; therefore, after an exhaustive search, only 2 studies [52,53] were identified that were closest to implementation in the Latin American context; one was implemented in Mexico and the other in Brazil; neither evaluated acceptance factors. However, a meta-analysis study [54] conducted across multiple international contexts on professional burnout related to the use of EHRs indicates that insufficient acceptance of the system has negative consequences for the well-being of health care personnel. Hence the importance of evaluating internal users' acceptance of hospital information systems, as was done in this study; where the perceived level of the MHC's (Módulo de Historias Clínicas [Medical Records Module]) usability is very high at 81.7% and high at 18.3%; whereas in other studies [55] relatively related to usability evaluation, results were higher than those obtained in this study, with a score of 83.1; however, a direct comparison is not possible since the calculation method in this study is based on percentages, whereas the cited study on OpenEDC evaluation uses numerical scores; the cited study

XSL•FO
RenderX

involved 16 users, while our case study had a population of 115 users.

On the other hand, there is also another study [56] related to metadata management that evaluates the usability of a metadata repository called CoMetaR; 12 individuals participated in the study, in which 3 modules were evaluated, obtaining scores of 81, 81, and 72, respectively. For this, they used the SUS, which is one of the most widely used instruments for usability evaluation.

Regarding the subject of the study [57], the authors report that they conducted a systematic review in which they observed, across several primary studies, the existence of technical and social challenges in achieving quality in health information systems; according to the studies examined, various tools were used to assess software quality, including SUS, After-Scenario Questionnaire, and Nielsen's heuristics, among others; however, only 3 out of 17 studies included the processes of ISO/IEC 25010 and ISO 9126; thus, they noted that there is no precise description of how to translate these standards into practice; that is, there is a lack of success stories, case studies, and other materials describing lessons learned regarding the application of quality standards. In line with the statements of the cited authors, this study addressed the development of a usability evaluation scale based on ISO 25010 to serve as a reference for future work related to the quality of software products.

In another systematic review [58] regarding the evaluation of usability in telemedicine tools, the authors found that many studies did not use standardized instruments, which hindered an adequate description of the usability of telemedicine tools; although it is reported that most of the studies included in the systematic review demonstrate good usability, this is not specified in percentage or numerical terms, which would allow for comparisons or the identification of improvements; on the other hand, they also note that most studies use patients as the study population for usability and very rarely internal users; even here, the situation differs in the European context or in high-income countries compared with Latin American countries such as Peru. In line with these findings and in contrast to them, this study evaluated the usability of the MHC among internal users using a validated instrument based on ISO 25010, presenting statistical data on content validity, reliability, and CFA (structural validity) to serve as a foundation for future research.

Likewise, another systematic review study [59] on usability evaluation in mobile apps for physical rehabilitation reports that there are a variety of approaches and instruments for such evaluation; that many of them lack a theoretical foundation and psychometric properties regarding usability evaluation instruments; given that they found a large number of custom-developed instruments, such as hybrid instruments or those adapted from existing ones, the latter being limited by sample size, which impairs their validity; this highlights the need for alternative approaches to increase flexibility while ensuring the psychometric properties of the instruments; on the other hand, regarding the theoretical foundation of usability, they found gaps in the connection with theoretical usability models such as the ISO standards or similar ones; in contrast to the gaps perceived and reported in systematic review research studies. This research study has sought to address the reported gaps by developing an evaluation instrument based on ISO 25010, while also incorporating its corresponding psychometric properties, although it has limitations regarding cross-cultural validity that could be addressed in future studies.

It has been observed that most systematic review studies have limitations that prevent an adequate analysis; since existing research in the literature lacks standards or guidelines to facilitate an adequate analysis of the usability of IT products, as reported by [60] in their study evaluating the usability of portable devices and complementary health applications; there, they found a variety of instruments used to assess usability, yet one that outnumbered other instruments was the SUS; despite this, only 16 of the 68 studies included in the study used any statistical analysis, and only one-third of the studies explicitly reported the measurement of usability or perceived usability; however, they had difficulty distinguishing the evaluation of the wearable device from that of the accompanying mobile app; thus highlighting the lack of standardized frameworks that allow for the proper evaluation of the usability of wearable devices and complementary apps; demonstrating the need for further research to bridge the existing gap between usability and user experience.

Limitations

Although this study provides a robust instrument, some limitations suggest avenues for future research: the sample was limited to an urban hospital network, so validation in rural areas with less technological infrastructure is recommended. Likewise, it would be valuable to: (1) evaluate the impact of improving specific dimensions (eg, accessibility) on real clinical indicators (eg, reduction in medical errors), and (2) develop implementation guidelines that link usability results to user-centered design protocols. These lines of work could bridge the gap between evaluation and best practices in EHR systems, transforming usability from an abstract concept into a measurable driver of health care quality and patient safety in the region. Despite the limitations, the instrument provides comprehensive psychometric validation based on ISO 25010, paving the way for future research on cross-cultural validation and application in diverse contexts.

Conclusions

This study developed and validated an instrument for evaluating the usability of information systems, which is aligned with the dimensions of ISO 25010. The results demonstrated excellent psychometric properties, with a near-perfect factor fit (CFI=1.000, Tucker-Lewis Index=1.001) and exceptional reliability (Cronbach α=0.968). The application of the instrument in a real-world setting revealed that 81.7% of users perceived usability as "very high," validating not only the tool itself but also the importance of considering local regulatory frameworks in the design of digital health systems, such as Peru's NTS. These findings represent a significant advance over generic instruments such as the SUS, by offering systematic alignment with ISO 25010, and, on the other hand, the alignment of information systems with local legislation (NTS).

XSL•FO
RenderX

The research identified certain differences in the perception of usability according to professional roles: while administrative staff reported higher satisfaction (93% “very high”), doctors and nurses gave more critical ratings (80%), particularly in dimensions such as “error protection” and “operability.” These findings align with the alignment or adaptation of information systems to the NTS, but provide specific solutions for the Peruvian context. The main practical implications include: (1) the need to redesign workflows for critical clinical tasks, (2) prioritizing ongoing training focused on usability, and (3) establishing periodic evaluations using this instrument as part of the continuous improvement processes for EHR systems. Additionally, the study revealed that 20% of the NTS requirements (primarily interoperability) were not implemented, highlighting urgent areas for technological development.

Acknowledgments

The authors would like to thank the hospital network in the Puno region for facilitating access to health care personnel (doctors, nurses, and administrative staff) and for their valuable participation in the validation of the instrument. We would also like to thank the 10 expert judges who participated in the content validation process using the modified Delphi technique for their time and specialized contributions, which helped improve the quality of the assessment instrument.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. Article processing charges will be covered by Universidad Peruana Unión, Juliaca, Peru.

Authors' Contributions

All authors contributed equally to this study. DM-P and AP-H participated in the conceptualization of the study, development of the assessment instrument, data collection and analysis, as well as manuscript drafting and psychometric validation of the instrument. MEH-F contributed to the methodological validation, review of the study design, and critical review of the intellectual content. EG-Q and JO-L participated in the statistical analysis and interpretation of the results. All authors reviewed and approved the last version of the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Instrument according to ISO 25010.
[DOCX File, 40 KB - humanfactors_v13i1e81377_app1.docx]

Multimedia Appendix 2
Content validity results by experts.
[PDF File, 195 KB - humanfactors_v13i1e81377_app2.pdf]

Multimedia Appendix 3
Requirements implementation checklist according to NTS (Norma Técnica de Salud (Technical Health Standard).
[PDF File, 113 KB - humanfactors_v13i1e81377_app3.pdf]

Multimedia Appendix 4
Results of Jamovi statistical analysis evaluation.
[PNG File, 72 KB - humanfactors_v13i1e81377_app4.png]

Multimedia Appendix 5
Cronbach alpha results.
[PNG File, 17 KB - humanfactors_v13i1e81377_app5.png]

Multimedia Appendix 6
Descriptive results of the usability dimensions.
[PNG File, 49 KB - humanfactors_v13i1e81377_app6.png]

Multimedia Appendix 7

XSL•FO
RenderX

Results of MHC usability perception by profession.
[PNG File, 74 KB - humanfactors_v13i1e81377_app7.png]

References

1. Thakur S, Gupta B, Mathur U, Bansal D. Ultimate usability and user experience in electronic health records: designing user-friendly features for ehrs. 2023 Presented at: 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS); Trichy, India p. 1726-1731. [doi: 10.1109/ICAISS58487.2023.10250723]
2. Park S, Marquard J, Austin R, Pieczkiewicz D, Delaney C. Usability of electronic health records from nurses' perspectives: a systematic review. 2022 Presented at: 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI); Rochester, MN, USA p. 511-512. [doi: 10.1109/ICHI54592.2022.00091]
3. Elkefi S, Asan O. Exploring doctors' satisfaction with electronic health records and their features: a nationwide survey study. 2022 Presented at: 2022 IEEE International Symposium on Systems Engineering (ISSE); Oct 24-26, 2022. [doi: 10.1109/ISSE54508.2022.10005506]
4. Ndabu T, Mulgund P, Sharman R, Singh R. Perceptual gaps between clinicians and technologists on health information technology-related errors in hospitals: observational study. JMIR Hum Factors 2021 Feb 5;8(1):e21884. [doi: 10.2196/21884] [Medline: 33544089]
5. Rani BV, Singh P. A survey on electronic health records (EHRS): challenges and solutions. 2022 Presented at: 2022 6th International Conference on Computing Methodologies and Communication (ICCMC); Mar 29-31, 2022; Erode, India p. 655-658. [doi: 10.1109/ICCMC53470.2022.9754040]
6. Moholkar K, Sayyad A, Mogle S, Prasad S, Raut P, Jayabhaye S. Comprehensive analysis of electronic health records: introducing SWASTHA - a novel model. 2024 Presented at: 2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS); Jul 10-12, 2024; Coimbatore, India p. 1629-1636. [doi: 10.1109/ICSCSS60660.2024.10625324]
7. Linhares CDG, Lima DM, Ponciano JR, et al. ClinicalPath: a visualization tool to improve the evaluation of electronic health records in clinical decision-making. IEEE Trans Visual Comput Graphics 2023;29(10):4031-4046. [doi: 10.1109/TVCG.2022.3175626]
8. Clotet R, Hernandez E, Huerta M, Rivas D. Used scenarios to validate an electronic health records synchronization ecosystem. 2020 Presented at: 2020 IEEE ANDESCON. [doi: 10.1109/ANDESCON50619.2020.9272198]
9. Liu Y, Qin S, Zhang Z, Shao W. Compound density networks for risk prediction using electronic health records. 2022 Presented at: 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); Dec 6-8, 2022; Las Vegas, NV, USA p. 1078-1085. [doi: 10.1109/BIBM55620.2022.9995587]
10. Acuña R, López RL. Usability evaluation of an electronic health record (EHR) system in the outpatient department of the Hospital de Clínicas: a case study.: LACCEI, Inc; 2020. URL: https://laccei.org/LACCEI2020-VirtualEdition/work_in_progress/WP235.pdf [accessed 2024-01-14]
11. Bleich HL, Slack WV. Reflections on electronic medical records: when doctors will use them and when they will not. Int J Med Inform 2010 Jan;79(1):1-4. [doi: 10.1016/j.ijmedinf.2009.10.002] [Medline: 19939731]
12. Claret PG, Sebbanne M, Bobbia X, et al. First medical contact and physicians' opinion after the implementation of an electronic record system. Am J Emerg Med 2012 Sep;30(7):1235-1240. [doi: 10.1016/j.ajem.2011.06.014] [Medline: 22030205]
13. Tall JM, Hurd M, Gifford T. Minimal impact of an electronic medical records system. Am J Emerg Med 2015 May;33(5):663-666. [doi: 10.1016/j.ajem.2015.02.022] [Medline: 25744149]
14. Mashoka RJ, Murray B, George U, et al. Implementation of electronic medical records at an emergency medicine department in Tanzania: the information technology perspective. Afr J Emerg Med 2019 Dec;9(4):165-171. [doi: 10.1016/j.afjem.2019.07.002] [Medline: 31890478]
15. Samadbeik M, Fatehi F, Braunstein M, et al. Education and training on electronic medical records (EMRs) for health care professionals and students: a scoping review. Int J Med Inform 2020 Oct;142:104238. [doi: 10.1016/j.ijmedinf.2020.104238] [Medline: 32828034]
16. Assiri G. The impact of patient access to their electronic health record on medication management safety: a narrative review. Saudi Pharm J 2022 Mar;30(3):185-194. [doi: 10.1016/j.jsps.2022.01.001] [Medline: 35498224]
17. Nahvijou A, Esmaeeli E, Kalaghchi B, Sheikhtaheri A, Zendehdel K. Using electronic health record system to establish a national patient's registry : lessons learned from the cancer registry in Iran. Int J Med Inform 2023 Dec;180:105245. [doi: 10.1016/j.ijmedinf.2023.105245] [Medline: 37864948]
18. Azarm M, Meehan R, Kuziemsky C. Towards a universal patient-centric health record sharing platform. Health Policy Technol 2023 Dec;12(4):100819. [doi: 10.1016/j.hlpt.2023.100819]
19. Vukmir RB. Medicolegal aspects of documentation and the electronic health record. Medicina Clínica (English Edition) 2024 Apr;162(8):e9-e14. [doi: 10.1016/j.medcle.2024.01.004]
20. Fleurence RL, Kent S, Adamson B, et al. Assessing real-world data from electronic health records for health technology assessment: the suitability checklist: a good practices report of an ISPOR task force. Value Health 2024 Jun;27(6):692-701. [doi: 10.1016/j.jval.2024.01.019]

XSL•FO
RenderX

21. Fennelly O, Moroney D, Doyle M, Eustace-Cook J, Hughes M. Key interoperability factors for patient portals and electronic health records: a scoping review. Int J Med Inform 2024 Mar;183:105335. [doi: 10.1016/j.ijmedinf.2023.105335] [Medline: 38266425]
22. Sreejith R, Sinimole KR. User-centric evaluation of EHR software through NLP-driven investigation: implications for product development and user experience. J Open Innov Technol Mark Complex 2024 Mar;10(1):100206. [doi: 10.1016/j.joitmc.2023.100206]
23. Pratama AA, Mutiara AB. Software quality analysis for halodoc application using ISO 25010:2011. IJACSA 2021;12(8):383-392. [doi: 10.14569/IJACSA.2021.0120844]
24. Barletta VS, Caivano D, Colizzi L, Dimauro G, Piattini M. Clinical-chatbot AHP evaluation based on "quality in use" of ISO/IEC 25010. Int J Med Inform 2023 Feb;170:104951. [doi: 10.1016/j.ijmedinf.2022.104951] [Medline: 36525800]
25. Bernardo S, Orviz P, David M, et al. Software quality assurance as a service: encompassing the quality assessment of software and services. Future Gener Comput Syst 2024 Jul;156:254-268. [doi: 10.1016/j.future.2024.03.024]
26. Forçan LR, Abe JM, Lima LA, Nascimento SS. Questionnaire model for paraconsistent quality assessment of software developed in salesforce. In: IFIP Advances in Information and Communication Technology: Springer; 2020, Vol. 591:333. [doi: 10.1007/978-3-030-57993-7_38]
27. Jandhyala R. Delphi, non-RAND modified Delphi, RAND/UCLA appropriateness method and a novel group awareness and consensus methodology for consensus measurement: a systematic literature review. Curr Med Res Opin 2020 Nov;36(11):1873-1887. [doi: 10.1080/03007995.2020.1816946] [Medline: 32866051]
28. Fernandes CS. A reflection on the use of the Delphi technique in nursing. Texto Contexto - Enferm 2024;33:e20230227. [doi: 10.1590/1980-265x-tce-2023-0227es]
29. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet e-surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34. [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
30. The jamovi project. jamovi - open statistical software for the desktop and cloud. 2022. URL: https://www.jamovi.org/ [accessed 2025-07-26]
31. Larrosa-Dominguez M, Reverté-Villarroya S, Bernadó-Llambrich N, Sauras-Colón E, Zaragoza-Brunet J. Stroke coaching scale-11 items: construction and psychometric validation. Revista Científica de la Sociedad de Enfermería Neurológica (English ed) 2023 Jan;57:22-30. [doi: 10.1016/j.sedeng.2022.06.001]
32. Renghea A, Hernandez-Iglesias S, Cuevas-Budhart MA, et al. Internal validity and reliability of an instrument for evaluating quality of nursing care services for patients, nurses, and family members. J Healthc Qual Res 2024;39(6):381-390. [doi: 10.1016/j.jhqr.2024.07.003] [Medline: 39060137]
33. Llagostera-Reverter I, Luna-Aleixós D, Valero-Chillerón MJ, González-Chordá VM. Development and validation of meta-measurement instruments: a methodological approach. Enfermería Clínica (English Edition) 2024 Jul;34(4):322-329. [doi: 10.1016/j.enfcle.2024.07.007]
34. International organization for standardization (ISO). ISO/IEC 25010:2011(en), systems and software engineering — systems and software quality requirements and evaluation (SQuaRE) — system and software quality models. 2011. URL: https://www.iso.org/obp/ui/#iso:std:iso-iec:25010:ed-1:v1:en [accessed 2024-01-16]
35. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Hum Comput Interact 1995 Jan;7(1):57-78. [doi: 10.1080/10447319509526110]
36. Sevilla-Gonzalez MDR, Moreno Loaeza L, Lazaro-Carrera LS, et al. Spanish version of the system usability scale for the assessment of electronic tools: development and validation. JMIR Hum Factors 2020 Dec 16;7(4):e21161. [doi: 10.2196/21161] [Medline: 33325828]
37. Lewis JR. The System Usability Scale: past, present, and future. International Journal of Human–Computer Interaction 2018 Jul 3;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
38. Nielsen J. 10 usability heuristics for user interface design. 2020. URL: https://www.nngroup.com/articles/ten-usability-heuristics/ [accessed 2024-01-16]
39. Thamilarasan Y, Raja Ikram RR, Osman M, Salahuddin L, Bujeri WYW, Kanchymalay K. Enhanced system usability scale using the software quality standard approach. Eng Technol Appl Sci Res 2023;13(5):11779-11784. [doi: 10.48084/etasr.5971]
40. Chanchí EG, Gómez Álvarez MC, Yesid Campo WM. Proposal of an inspection tool according to Nielsen's usability attributes. Revista Ibérica de Sistemas e Tecnologias de Informação 2020 Feb:448-460 [FREE Full text]
41. Zen M, Burny N, Vanderdonckt J. A quality model-based approach for measuring user interface aesthetics with grace. Proc ACM Hum-Comput Interact 2023 Jun 14;7(EICS):1-47. [doi: 10.1145/3593224]
42. Tullis T, Albert B. Self-reported metrics. In: Measuring the User Experience 2013:121-161. [doi: 10.1016/B978-0-12-415781-1.00006-6]
43. Wu Y, Wu M, Wang C, Lin J, Liu J, Liu S. Evaluating the prevalence of burnout among health care professionals related to electronic health record use: systematic review and meta-analysis. JMIR Med Inform 2024 Jun 12;12(1):e54811. [doi: 10.2196/54811] [Medline: 38865188]
44. Barbalho IMP, Fernandes F, Barros DMS, et al. Electronic health records in Brazil: prospects and technological challenges. Front Public Health 2022;10:963841. [doi: 10.3389/fpubh.2022.963841] [Medline: 36408021]

XSL•FO
RenderX

45. Salas-Lumbreras PL, Ayuzo-del Valle NC, Lozano-Ulloa AG, et al. Implementing an electronic health record system in a wide health care hospital in Latin America from a medical perspective. Sal Jal 2025;12(2):86-98. [doi: 10.35366/121797]
46. Mohammad AAS, Alzyoud M, Samara EIM, et al. Adoption of electronic health records: a bibliometric analysis. In: Hannoon A, Mahmood A, editors. Intelligence-Driven Circular Economy: Regeneration towards Sustainability and Social Responsibility: Springer Nature Switzerland; 2025, Vol. 1:301-314. [doi: 10.1007/978-3-031-73899-9_24]
47. Derecho KC, Cafino R, Aquino-Cafino SL, et al. Technology adoption of electronic medical records in developing economies: a systematic review on physicians' perspective. Digit Health 2024;10:20552076231224605. [doi: 10.1177/20552076231224605] [Medline: 38222081]
48. Justo N, Espinoza MA, Ratto B, et al. Real-world evidence in healthcare decision making: global trends and case studies from Latin America. Value Health 2019 Jun;22(6):739-749. [doi: 10.1016/j.jval.2019.01.014]
49. Vinadé Chagas ME, de Oliveira Laguna Silva G, Ricardo Fernandes G, et al. The evolution of digital health: a global, Latin American, and Brazilian bibliometric analysis. Front Digit Health 2025;7:1582719. [doi: 10.3389/fdgth.2025.1582719] [Medline: 40520219]
50. Corbett S, Maruthu R, Saab MM, Lehane E. Nurses' perceptions of facilitators and barriers to their acceptance of electronic health records: a mixed-method systematic review. J Clin Nurs 2025 Aug;34(8):3085-3100. [doi: 10.1111/jocn.17736] [Medline: 40103193]
51. Griesser A, Mzoughi M, Bidmon S, Cherif E. How do opt-in versus opt-out settings nudge patients toward electronic health record adoption? an exploratory study of facilitators and barriers in Austria and France. BMC Health Serv Res 2024 Apr 8;24(1):439. [doi: 10.1186/s12913-024-10929-w] [Medline: 38589922]
52. Griesser A, Bidmon S. A holistic view of facilitators and barriers of electronic health records usage from different perspectives: a qualitative content analysis approach. HIM J 2024 Sep;53(3):227-236. [doi: 10.1177/18333583231178611]
53. Rouidi M, Elouadi AE, Hamdoune A, Choujtani K, Chati A. TAM-UTAUT and the acceptance of remote healthcare technologies by healthcare professionals: a systematic review. Informatics in Medicine Unlocked 2022;32:101008. [doi: 10.1016/j.imu.2022.101008]
54. Tetik G, Türkeli S, Pinar S, Tarim M. Health information systems with technology acceptance model approach: a systematic review. Int J Med Inform 2024 Oct;190:105556. [doi: 10.1016/j.ijmedinf.2024.105556] [Medline: 39053345]
55. Greulich L, Hegselmann S, Dugas M. An open-source, standard-compliant, and mobile electronic data capture system for medical research (OpenEDC): design and evaluation study. JMIR Med Inform 2021 Nov 19;9(11):e29176. [doi: 10.2196/29176] [Medline: 34806987]
56. Stöhr MR, Günther A, Majeed RW. The collaborative metadata repository (CoMetaR) web app: quantitative and qualitative usability evaluation. JMIR Med Inform 2021 Nov 29;9(11):e30308. [doi: 10.2196/30308] [Medline: 34847059]
57. Noël R, Taramasco C, Márquez G. Standards, processes, and tools used to evaluate the quality of health information systems: systematic literature review. J Med Internet Res 2022 Mar 8;24(3):e26577. [doi: 10.2196/26577] [Medline: 35258469]
58. Gonçalves RL, Pagano AS, Reis ZSN, et al. Usability of telehealth systems for noncommunicable diseases in primary care from the COVID-19 pandemic onward: systematic review. J Med Internet Res 2023 Mar 16;25(1):e44209. [doi: 10.2196/44209] [Medline: 36787223]
59. Hach S, Alder G, Stavric V, Taylor D, Signal N. Usability assessment methods for mobile apps for physical rehabilitation: umbrella review. JMIR Mhealth Uhealth 2023;12(1):e49449. [doi: 10.2196/49449]
60. Moorthy P, Weinert L, Schüttler C, et al. Attributes, methods, and frameworks used to evaluate wearables and their companion mHealth apps: scoping review. JMIR Mhealth Uhealth 2024 Apr 5;12(1):e52179. [doi: 10.2196/52179] [Medline: 38578671]

Abbreviations

CFA: confirmatory factor analysis
CFI: Comparative Fit Index
COSMIN: Consensus-based Standards for the Selection of Health Measurement Instruments
EHR: electronic health record
EMR: electronic medical record
HCE: Historia Clínica Electrónica (Electronic Health Record)
HCM: Historia Clínica Manual (manual medical record)
HIS: Hospital Information Systems
IEC: International Electrotechnical Commission
IPRESS: Instituciones Prestadoras de Servicios de Salud (Healthcare Service Provider)
ISO: International Organization for Standardization
ISPOR: International Society for Pharmacoeconomics and Outcomes Research
IT: information technology
MHC: Módulo de Historias Clínicas (Medical Records Module)
MINSA: Ministerio de Salud (Peruvian Ministry of Health)

XSL•FO
RenderX

NFI: Normed Fit Index
NTS: Norma Técnica de Salud (Technical Health Standard)
RENHICE: Registro Nacional de Historias Clínicas Electrónicas (National Registry of Electronic Medical Records)
RHC: Medical Record Registry
SUS: System Usability Scale

Edited by A Kushniruk; submitted 27.Jul.2025; peer-reviewed by A Hidki; revised version received 24.Mar.2026; accepted 13.Apr.2026; published 22.May.2026.

Please cite as:
Mamani-Pari D, Peñaloza-Hualpa AF, Humpiri-Flores ME, Gutierrez-Quispe E, Otazu-Luque JE
Development and Psychometric Validation of a Usability Instrument Based on ISO 25010 for Electronic Health Record Systems in Peruvian Health Care Settings: Methodological Study
JMIR Hum Factors 2026;13:e81377
URL: https://humanfactors.jmir.org/2026/1/e81377
doi:10.2196/81377

© David Mamani-Pari, Alex Froilan Peñaloza-Hualpa, Milton Edward Humpiri-Flores, Eder Gutierrez-Quispe, Jorge Eddy Otazu-Luque. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 22.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

“Helpful in Thinking Through the Question”: Redesigning the Patient Health Questionnaire-8 for Depression Assessment to Enhance Patient Engagement and Understanding

Hannah Dorris[1], BA; Jenna Sheldon[2]; Courtney Sepucha[2]; Emmanuel Kasuti Makau[3]; Pape Sow Traoré[3], MS; Elizabeth Murnane[2], PhD; Nicholas Jacobson[1], PhD

[1]Center for Technology and Behavioral Health, Dartmouth College, 46 Centerra Parkway, EverGreen Building, Lebanon, NH, United States

[2]Thayer School of Engineering, Dartmouth College, Hanover, NH, United States

[3]Department of Computer Science, Dartmouth College, Hanover, NH, United States

Corresponding Author:
Hannah Dorris, BA
Center for Technology and Behavioral Health, Dartmouth College, 46 Centerra Parkway, EverGreen Building, Lebanon, NH, United States

Abstract

Background: Major depressive disorder (MDD) is a prevalent and impairing mental health condition characterized by persistent low mood and diminished energy. Self-report instruments, such as the Patient Health Questionnaire-8 (PHQ-8), are frequently used in clinical and research settings for depression assessment.

Objective: We developed, redesigned, and tested 2 alternative versions of the PHQ-8 that incorporate visual elements and reframed statements aimed at improving acceptability and comprehension.

Methods: In a mixed methods study, 20 participants with high subclinical to moderately severe depressive symptoms provided feedback after completing 3 versions: the standard PHQ-8 form, our version augmented with visual aids, and our version that combined visual aids with rephrased item statements for greater clarity and engagement.

Results: Eighty percent (16/20) of participants preferred one of the augmented PHQ-8 versions over the standard questionnaire (P=.006).

Conclusions: Participants reported that the augmented versions improved their understanding of survey items, reduced response burden, and increased engagement, suggesting that visual and linguistic design modifications may enhance the patient experience in depression screening.

(JMIR Hum Factors 2026;13:e76230) doi:10.2196/76230

KEYWORDS

Patient Health Questionnaire-8; PHQ-8; survey design; depression; human-computer interaction; user experience

Introduction

Background

Clinicians and researchers distribute health questionnaires every day across a variety of media, such as phones, tablets, computers, or paper forms in clinical waiting rooms. Although these instruments are ubiquitous, patients often report challenges related to engagement and ease of use, motivating design research aimed at improving respondent experience and potentially supporting more consistent reporting behaviors [1].

Psychiatrists and psychologists often heavily rely on the Patient Health Questionnaire (PHQ)-8 (or PHQ-9, which includes an additional item assessing suicidal ideation) to assess major depressive disorder (MDD) severity [2-8]. In this paper, we aimed to explore the redesign of the PHQ-8 by enhancing question framing and language and adding metaphor-based visualizations. Specifically, we created 2 new questionnaires to address limitations of the standard PHQ-8 version, one augmented with visual aids and another that combined visual aids with rephrased item statements to provide greater clarity and engagement. We tested these 2 new versions with participants with depressive symptoms to evaluate the effectiveness of the new designs in terms of patient experience.

We aimed to examine the following research questions:

- How do our redesigned questionnaires, augmented with visuals and other elements, effectively address the limitations of the current PHQ-8?
- Do the augmented PHQ-8 questionnaires show acceptability when tested?

XSL•FO
RenderX

- What contributes to patient preferences for the standard or the augmented PHQ-8 questionnaires?

While design adaptations cannot eliminate all limitations common to mental health screening tools, they may improve user experience, comprehension, and perceived engagement.

Related Work

Overview

Our research is situated at the intersection of the clinical and human-computer interaction literatures. Specifically, the following subsections review how conventional questionnaires often fail to adequately engage patients and accurately capture subjective mental health experiences, the influence of mood symptoms and perceptions of stigma on self-reporting practices, and design strategies for enhancing patient engagement and understanding.

Challenges in Self-Report Questionnaires and Known PHQ-8 Biases

We first broadly examine issues with participant burden and engagement in filling out patient questionnaires before examining research on the PHQ-8 specifically. Self-report questionnaires collect important subjective data and are the gold standard for understanding a patient's condition, particularly for mental health conditions. However, questionnaires can be difficult and burdensome to fill out, especially when a person is busy [9,10]. Research shows that patients generally have low questionnaire response rates, provide careless responses, and struggle with health questionnaire literacy [11-17]. These challenges reflect the general properties of self-report instruments rather than limitations unique to the PHQ-8.

Prior work across diverse clinical populations, including hospice and youth mental health settings, has shown that patients often perceive standardized questionnaires as impersonal or insufficiently reflective of their lived experiences, which can reduce engagement [18,19]. Although these findings are not specific to depression screening, they illustrate broader challenges in designing self-report tools that feel meaningful, comprehensible, and engaging to respondents. These findings highlight a widespread need among patients in mental health, hospice, and general care for improved questionnaire design that minimizes boredom and fosters a greater sense of being understood.

Prior research has demonstrated that the PHQ-8 is subject to several areas of bias that differ from broad research on survey engagement. The main areas highlighted in prior research are recall bias, social desirability bias, and context-dependent comprehension issues. First, in exploring recall bias, recent research has demonstrated that PHQ-8 results have "modest peak-end recall bias" when tested in that the most intense and proximate emotion disproportionately affects the recall and reporting of symptoms [20]. These findings were further examined in another study that demonstrated fluctuations in patients' ability to recall symptoms, although the recall bias was not statistically significant [21]. PHQ-8 responses may also be influenced by social desirability bias.

Although social desirability bias and recall bias both impact the PHQ-8, comprehension issues are another element that is critical to examine. For example, a recent study found that when tested in another context, there were incorrect interpretations of half of PHQ-8 items and, specifically, issues in interpreting the time scale aspect of the rating [22,23]. Such misunderstandings may partially contribute to discrepancies occasionally observed between clinician-rated and self-rated depression severity [24]. Overall, many of these challenges (eg, recall bias, response burden, and social desirability) are common across screening instruments and modes of administration; thus, design improvements may help mitigate but not fully eliminate them.

Mood Questionnaires Need to Take Into Account Common Symptoms

Anhedonia is a common symptom in MDD, where patients experience a loss of interest, motivation, engagement, and inability to experience pleasure. Patients often describe anhedonia as involving not only reduced interest but also a broader flattening of emotion and diminished capacity for joy [25,26]. Therefore, when questionnaires are boring and unengaging, they can be particularly challenging for such patients to complete, compromising response rates and accuracy.

Another major component of depressive symptoms relates to perceived stigma. Stigma is known to interfere, at minimum, with how individuals choose to self-identify in having a mental illness, such as perceived need and help-seeking behavior [27,28]. Research has begun to address minimizing stigma-based bias in questionnaire design, particularly with respect to reducing social desirability bias [29]. Overall, considering 2 commonly occurring depression symptoms are anhedonia and stigma, it is important to work to understand ways to further increase participant engagement and reduce social desirability bias in the PHQ-8 survey for patients with MDD. These factors underscore the importance of designing depression assessments that minimize perceived judgment and support honest self-disclosure.

The PHQ-8 Might Have a Familiarity Bias

The PHQ-8 is the most popular questionnaire for assessing depression [30-34]. Recent studies have begun to untangle the relationship between reaction time and PHQ-8 results in individuals with depression, those with depression not seeking treatment, and those with depression seeking treatment. There was a notable difference in reaction time, as well as other observed psychometric properties, in patients with MDD completing questions 3, 7, 8, and 9 of the PHQ-8 more quickly between those who sought treatment and those who did not. These differences indicate the potential for familiarity bias to affect specific questions of the PHQ-8, yet no large empirical study has tested this claim further [35].

Designing for Patient Engagement and Understanding

We hypothesize that reasons for the PHQ-8 questionnaire failing to accurately capture MDD symptoms may be due to the instrument's structure and presentation of questions, particularly in terms of how both its visuals and texts could better promote engagement and motivation.

XSL•FO
RenderX

Visual-based design methods, such as interactive formats and adaptive questions, have begun to be implemented in surveys to enhance patient engagement. Broadly, there has been a rise in using design methods to encourage patient acceptance and understanding across a number of digital platforms, such as conversational agents, collaborative care, vital monitoring, and online cognitive behavioral therapy training [36-42]. For example, a chronic pain survey was developed in which patients select facial expressions to self-report pain levels. The study found high validity in this measure and that patients reported lower response burden and increased response motivation [36].

Other visual designs have explored the use of photos to measure affect, particularly the Photographic Affect Meter interface that asked respondents to select from a variety of pictures to choose the image that best represented their current mood. Evaluation of the Photographic Affect Meter instrument has shown strong validity and ease of use [43]. A similar photo-based questionnaire has been created to measure current pain levels [36]. Collectively, these studies demonstrate that visual and interactive elements can improve usability, reduce response burden, and increase engagement across various health contexts [36,43,44].

Visual designs based on the connection of emojis and mood have also been implemented. Studies on emoji-based tools, such as the EmojiGrid and the Emoji Current Mood and Experience Scale, highlight the potential of visual surveys to be both engaging and valid across a variety of settings [45,46]. However, emojis may be less effective in capturing nuanced emotional states, particularly among individuals experiencing high emotional intensity [47].

Many other studies have found that visual interface elements increase perceived usability and enjoyment and help patients feel more relaxed and happier [48,49]. Furthermore, research finds that increased visual appeal can lead to increased trustworthiness [50]. Considering that patients with mental health concerns often face a sense of stigma when completing assessments [18], visual elements may be an important way to build engagement and trust and elicit more honest responses. Our research, therefore, aims to overcome PHQ-8 challenges related to boredom, stigma, and inaccurate responses by exploring the design of a questionnaire that incorporates visual elements to enhance respondent user experience and engagement.

In addition to visual design, the phrasing and structure of questionnaire items also influence respondent engagement and comprehension. Text-based design is equally important, as the language and framing of items can substantially influence survey experience and response quality. In particular, it is helpful to frame survey items as a question rather than as a statement, as the PHQ-8 currently does [51]. Such framings can also impact how comfortable respondents feel in sharing negative responses [52,53]. Open-ended questions are another effective technique to encourage patient comfort and honesty, as such formats allow for more authentic feedback [54]. A hybrid approach of including a Likert scale with an open-ended question can enhance patient comfort and engagement even without the need for artificial intelligence or large language models to appropriately scale the response [54]. Overall, prior work indicates that incorporating language that is compassionate and open-ended can help encourage respondent engagement and honesty [55-57]. We therefore experimented with modifying the PHQ assessment to evaluate whether such question framings could enhance engagement.

Despite extensive research on self-report burden, visual communication, and item design across health questionnaires, relatively little work has examined how these principles can be applied to improve the PHQ-8 itself. Existing studies reveal important limitations such as recall biases, social desirability influences, and comprehension challenges, but these issues are typically reported at the level of general self-report methodology rather than addressed through targeted redesign of PHQ-8 items or layout. Furthermore, while visual and interactive elements have shown promise in enhancing engagement in other symptom-tracking tools, their effects on depression screening instruments remain largely unexplored, and prior work has rarely assessed whether such design changes preserve psychometric properties. Taken together, this literature highlights a clear gap: the need for systematically designed and evaluated adaptations of the PHQ-8 that incorporate visual elements and improved item phrasing to support patient engagement and understanding without compromising validity. Our study addresses this gap by developing and testing 2 redesigned PHQ-8 formats that integrate these design principles.

Methods

Participants

Participants were recruited from a larger Dartmouth study that examined digital treatments for anxiety and depression. Participants were recruited from across the United States via Google and Meta Ads. Individuals had previously completed the PHQ-8 and General Anxiety Disorder-7 screening questionnaires and were eligible for recontact if they were aged between 40 and 65 years, lived in the United States, and demonstrated high subclinical to moderately severe depressive symptoms (PHQ-8≥8). From this pool, we recruited 20 participants (women: n=16, 80%; and men: n=4, 20%). Their mean PHQ-8 score at screening was 14.84 (SD 5.18), which is approximately 5 points above the clinical threshold of 10. Three participants scored below the clinical threshold of 10.

Data Collection

Each participant completed a semistructured interview via Zoom lasting 30 to 60 minutes. Interview topics included condition acceptance, social support, mindset, stigma, and technology preferences. Specifically, the questions most salient to this study were as follows: At what age did your depression begin? How many years have you been living with depression? Are you currently seeking treatment? Is there any treatment or medication you've done to help your condition in the past?

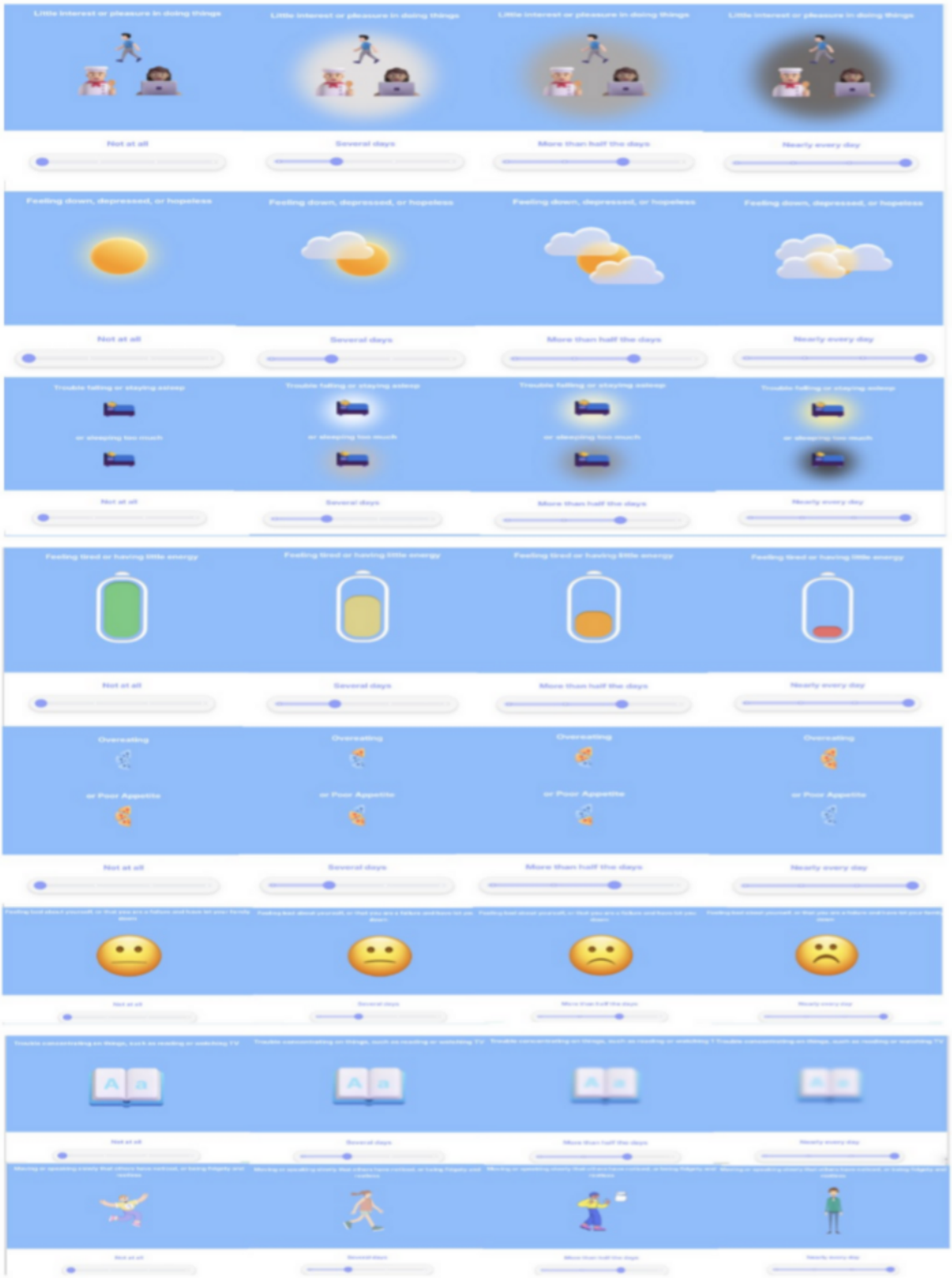
Following the interview, participants were given 2 PHQ-8, as shown in Figure 1 (we did not include the 9th question regarding suicidality, as we had already screened for this, as described earlier). The "Visual PHQ-8" (Figure 1) used the same statements of the standard PHQ-8 accompanied by

XSL•FO
RenderX

visualizations. The “Reframed PHQ-8” (Figure 2) modified the assessment statements to open-ended questions, while using the same visualizations from our Visual PHQ-8. We rephrased all the PHQ-8 items to be open-ended and more conversational. All the questions besides item 1 reflect concrete feelings or behaviors, whereas item 1 reflects a motivational state (“little interest or pleasure in doing things”). Using a “how” structure helps preserve that motivational-state meaning, while a “do you...” phrasing would incorrectly frame the item as a feeling or behavior rather than a motivational state. After completing each questionnaire, we asked participants about their experience in filling out the questionnaire and whether there was anything they would change about it. Finally, we asked participants if they had a preference among the various PHQ versions (standard, visual, and reframed) and to elaborate on such reactions. To enable us to examine links between survey preference and psychological status, participants were then given the following questionnaires: Interpersonal Social Support to capture dimensions of social support (eg, tangible, appraisal, and belonging) [58,59]; Trait Hope Scale [60,61], which is known to predict positive outcomes and resiliency; Kind of Person Scale [62,63] and Dweck Growth Mindset Scale [64] to assess growth mindset and reactions to adversity; and the Stigma Scale for Chronic Illnesses to assess psychological distress and patient performance [65].

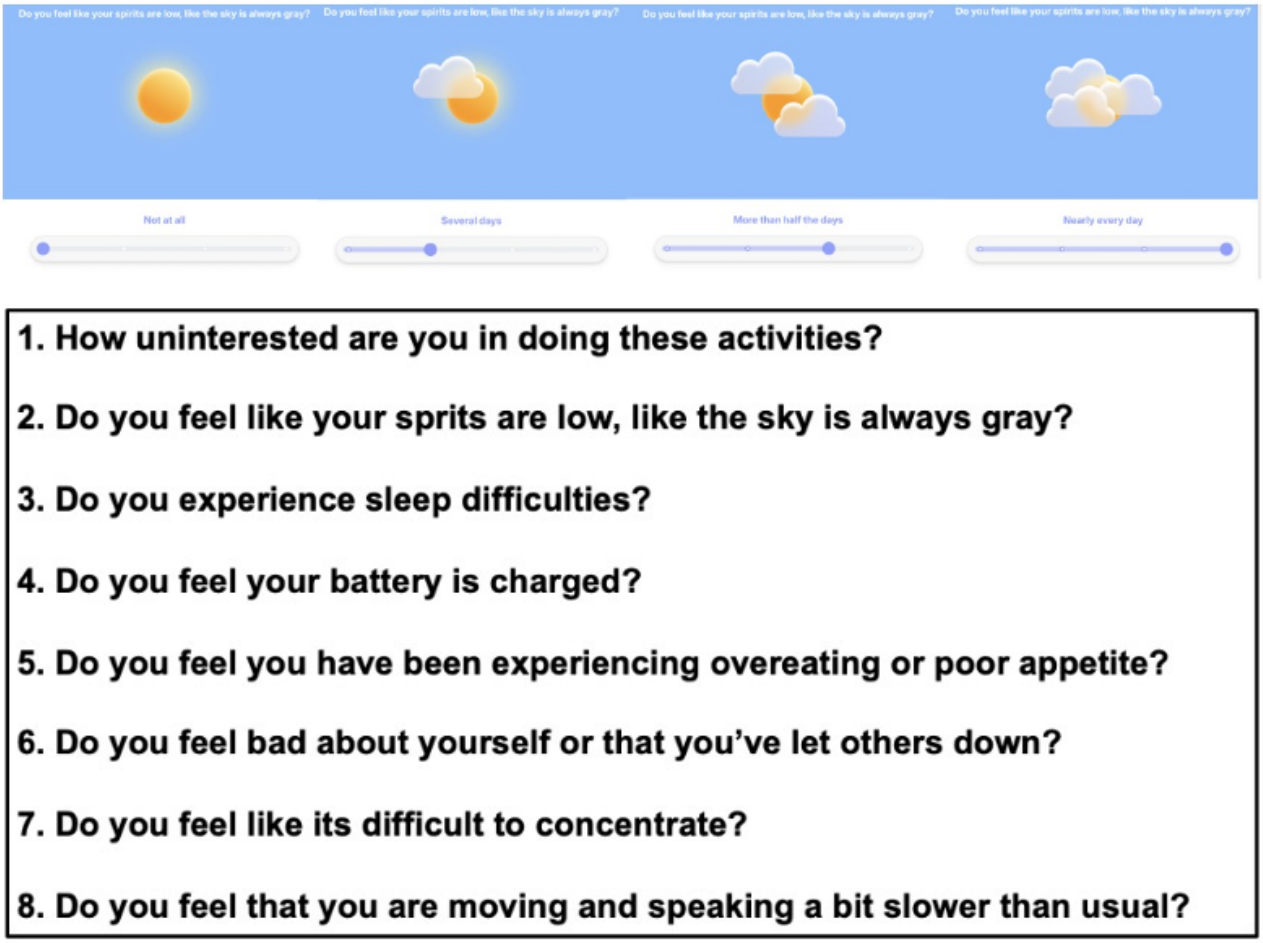
XSL•FO
RenderX

Figure 1. Our visual Patient Health Questionnaire-8 design that augments the standard questionnaire with visual aids.



XSL•FO
RenderX

Figure 2. Our reframed Patient Health Questionnaire-8 design that both augments the standard questionnaire with visual aids and also rephrases survey items.



Data Analysis

Interviews were transcribed and coded using an inductive qualitative approach. Two researchers independently reviewed all transcripts, developed an initial codebook, and resolved discrepancies through consensus. Themes were organized around depression history, current treatment status (ie, whether an individual was actively seeking therapy or medication), and questionnaire preferences.

We conducted 1-tailed exact binomial tests to evaluate whether the proportion of participants preferring augmented PHQ versions significantly exceeded chance (50%). In addition, we analyzed participants' responses to each of the PHQ-8 questionnaires. Specifically, we computed means and SDs for each questionnaire group and conducted intraclass correlation coefficients (ICCs) between the questionnaires. Finally, we analyzed responses to the Interpersonal Social Support [58,59], Trait Hope Scale [60,61], Kind of Person Scale [62,63], Dweck Growth Mindset Scale [64], and the Stigma Scale for Chronic Illnesses [65] to identify associations between those responses and participants' preferences among the different PHQ-8 versions.

Ethical Considerations

The study conducted at Dartmouth was approved by the Committee for Protection of Human Subjects under study 32782. All activities in this study were conducted under the approved protocol and complied with institutional and federal guidelines for research on human participants. All participants gave informed consent and were compensated ($20 per hour) at the completion of the study. We attest to maintaining privacy and confidentiality of all research subjects' data.

Results

Overview

Our data showed that 80% (16/20) of participants preferred one of the redesigned PHQ-8 formats over the standard version. This preference for an augmented version of the PHQ-8 was statistically significant under a 1-tailed exact binomial test (proportion: 0.80, 95% CI 0.56-0.94; P=.006), indicating an above-chance likelihood that a participant prefers a redesigned PHQ-8 (Figure 1) compared to a traditional PHQ-8 (Figure 3, Tables 1 and 2). Only 20% (proportion: 0.20, 95% CI 0.06-0.44) of participants preferred the standard PHQ-8, indicating a minority preference for the nonaugmented version.

Participants who preferred the redesigned questionnaire with visualizations (PHQ-8 mean 12.50, SD 5.34) had higher PHQ-8 scores than participants who preferred the traditional statements (PHQ-8 mean 9.75, SD 2.99). Additionally, the trend followed that those who preferred visualizations over statements had higher stigma scores (mean 23.19, SD 9.63) and lower hope

XSL•FO
RenderX

scores (mean 40.94, SD 8.64) than those who preferred statements (stigma: mean 21.25, SD 10.11; and hope: mean 51, SD 6.58).

Figure 3. The Patient Health Questionnaire-8 in standard form.

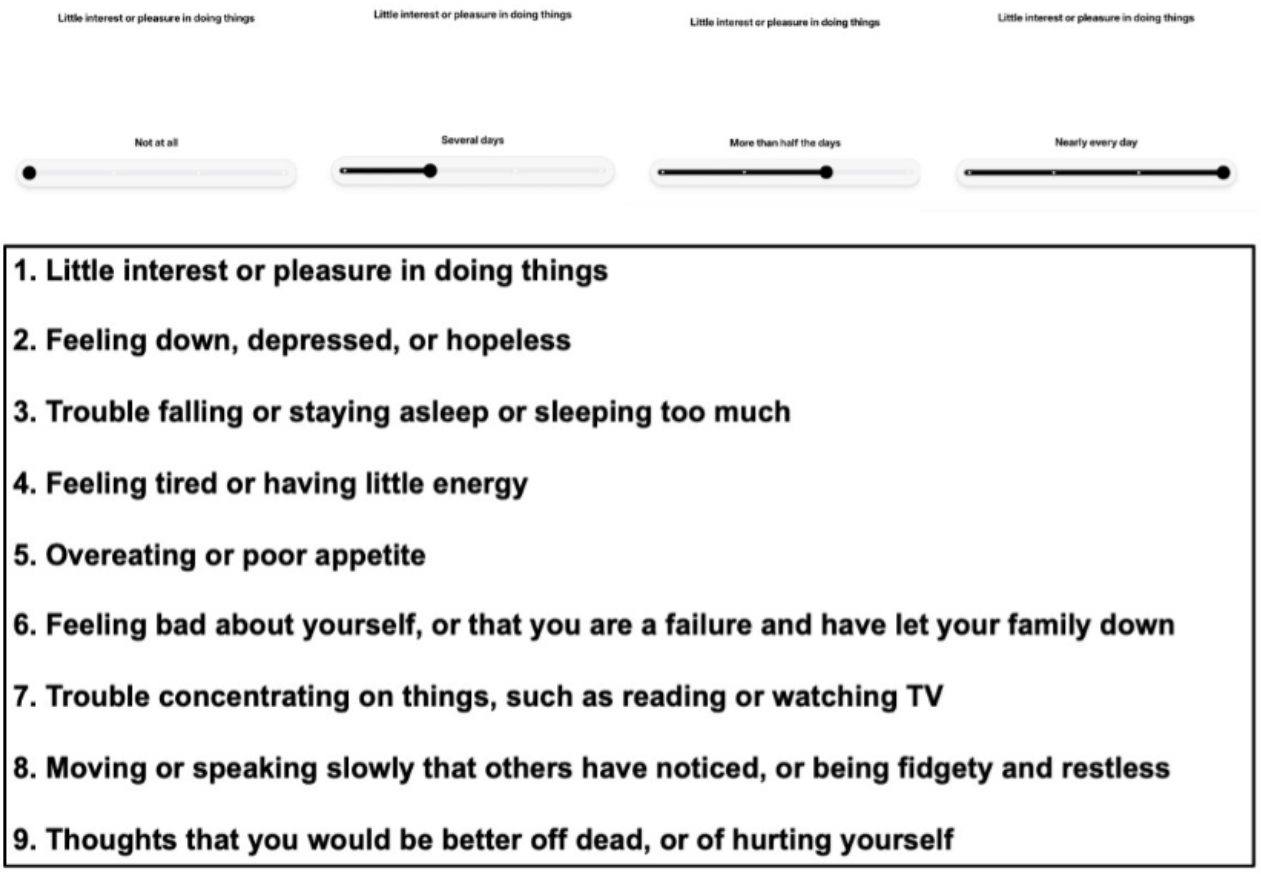


Table . Patient Health Questionnaire (PHQ) responses, summarized according to survey type preference.

Participants' favorite questionnaire	Count (N=20), n (%)	PHQ-8: statements (mean 11.95), mean (SD)	PHQ-8: statements and visualizations (mean 11.00), mean (SD)	PHQ-8: statements and questions (mean 10.60), mean (SD)	Trait hope scale (mean 42.95), mean (SD)	Stigma scale (mean 22.80), mean (SD)	Growth mindset (mean 10.11), mean (SD)	Kind of person (mean 28.50), mean
Prefer statements	4 (20)	9.75 (2.99)	6.75 (4.86)	6.75 (2.06)	51.00 (6.58)	21.25 (10.11)	10.50 (6.24)	28.00 (4.08)
Prefer questions and visualizations	7 (35)	11.29 (5.53)	10.71 (5.02)	10.43 (4.58)	38.14 (12.13)	25.29 (9.93)	11.14 (5.01)	27.43 (7.21)
Prefer statements and visualizations	9 (45)	13.44 (5.32)	13.11 (5.75)	12.44 (6.48)	43.11 (4.17)	21.56 (9.66)	9.00 (3.38)	29.56 (3.78)
Prefer either questions or statements and visualizations	16 (80)	12.50 (5.34)	12.06 (5.41)	11.56 (5.64)	40.94 (8.64)	23.19 (9.63)	10.00 (4.21)	28.63 (5.44)

XSL•FO
RenderX

Table . Response per question for the 3 Patient Health Questionnaire-8 (PHQ-8) versions.

Questionnaire type	Anhedonia, mean (SD)	Hopelessness, mean (SD)	Sleep, mean (SD)	Fatigue, mean (SD)	Appetite, mean (SD)	Self-esteem, mean (SD)	Concentration, mean (SD)	Movement, mean (SD)
PHQ-8: statements	1.75 (1.00)	1.85 (1.03)	2.00 (0.97)	2.10 (0.91)	1.15 (1.12)	1.25 (0.85)	1.35 (1.05)	0.55 (0.84)
PHQ-8: statements+visualizations score	1.45 (1.09)	1.55 (1.05)	2.05 (0.94)	1.95 (1.05)	1.00 (1.12)	1.10 (0.64)	1.35 (1.14)	0.50 (0.83)
PHQ-8: questions+visualizations score	1.30 (0.97)	1.40 (1.05)	2.00 (1.17)	1.55 (1.32)	1.00 (1.08)	1.42 (0.96)	1.40 (1.14)	0.55 (1.00)

We next discussed the differences between the augmented versions (ie, questions and visualizations vs statements and visualizations) descriptively, as the primary statistical analysis evaluated whether participants preferred an augmented PHQ-8 compared to the standard version. We only use descriptives in discussing the differences between the augmented versions, as both augmented conditions incorporated visualizations and the study design did not include a condition with just rephrased questions, limiting the ability to generalize and analyze the effect of question phrasing independently. We now specifically examine the adapted versions, where 45% (9/20) of participants (proportion: 0.45, 95% CI 0.23-0.68) preferred the visual-only PHQ-8, and 35% (7/20) of participants (proportion: 0.35, 95% CI 0.15-0.59) preferred the version that combined visuals with reframed questions. The 95% CI for each of the augmented versions demonstrates a similar spread of preference for the questions and visualizations and statements and visualizations, which could be further studied in future experiments.

Participants Who Preferred Visualizations

Participants who preferred the addition of visualizations (Figure 1) strongly emphasized that they felt more engaged and experienced more pleasure while taking the questionnaire. For example:

> *It was more engaging and fun...I really liked that. I liked having the visual.* [Participant 18]
>
> *I did find it more fun.* [Participant 19]
>
> *I liked the little emojis. That was sort of fun.* [Participant 17]

Importantly, we identified that the visualizations not only were more engaging but also prompted more reflective responding. For example, participants noted that:

> *It would make me stop and think. About things that I generally don't.* [Participant 1]
>
> *I think some of the visuals were helpful in thinking through the question.* [Participant 11]

Both these participants noted that having the visual helped them pause and think more deeply about the question.

Participants Who Preferred Visualizations and Questions

Participants who had a preference for the questionnaire that included both visualizations and reframed questions (Figure 2) explained that this design helped them feel both cared about and understood. Furthermore, participants discussed that the wording felt less negative:

> *I liked the way that they were worded. Like they just seem softer than the normal [instrument phrasings].* [Participant 6]

Similarly, other participants expressed that the reframed questions felt less negative and confrontational, and more compassionate and genuine in checking in on a person's status. For example:

> *The questions were phrased differently, and I liked that. It didn't feel negative. It was more like, "oh, we would like to know this," rather than "we know you're like this, just admit it."* [Participant 9]

This theme related to a sense of being cared about and understood was further elaborated by 2 other participants:

> *Yeah, I like these questions better...They speak to you a little bit more.* [Participant 12]
>
> *It felt more understanding, and it felt like whoever was at the other end of the survey really wanted to know to help, rather than really wanting to know because it will be useful to them.* [Participant 14]

The second common theme was that participants felt that the reframed questions enabled them to be more thoughtful in responses. For example:

> *They made me think more... It's getting at the same information, but I like the way it was worded. It is a more thoughtful way, I think.* [Participant 6]

Another participant echoed this statement, noting that:

> *It made me think because it doesn't phrase it the same way that the doctors do. I had to think about what are you asking.* [Participant 9]

The ability of the redesigned questions to be more thought-provoking may partly stem from providing a break from the repetitive nature of the traditional PHQ-8 statements:

> *I'm so used to those questions I just kind of speed through them.* [Participant 6]
>
> *...they get so repetitive. I've done where you track it; it just gets really like, "okay, I was depressed yesterday, I'm depressed today...* [Participant 9]

In this way, participants emphasized that the redesigned questions helped overcome the repetitive nature of the PHQ-8

XSL•FO
RenderX

and, in turn, helped them to be more thoughtful in their responses.

Participants also explained the redesigned questions and visuals also supported better item comprehension:

> *The questions were a lot better than the statements...Sometimes with the statements, it's difficult to put together what it actually means.* [Participant 16]

> *The cloud and gray sky [visualization and question], I was like, "oh, I can imagine that.*" [Participant 19]

Overall, while these participants explained that they felt that responses to the traditional PHQ-8 were "difficult to put together," they expressed that new visualizations helped them "imagine" the statements better.

Participants Who Preferred Statements

Despite most participants preferring redesigned versions of the PHQ-8, 20% (4/20) of participants had a preference for keeping the PHQ-8 the same with no addition of a visualization or reframed question. We observed 2 main themes for these individuals. First, they found the addition of the visualizations to be hard to relate to. For example:

> *The visualizations are distracting because it's somebody else's picture.* [Participant 3]

This participant and another also noted their distaste for the aesthetics of the visualizations, referring to them as childish:

> *they seemed kind of child-like.* [Participant 3]

> *I think the text is enough. [Better than the] stupid smiley face chart they have at the hospitals and doctors' offices.* [Participant 3]

Another theme that we observed was that the visualizations were seen as unnecessarily lengthening the self-report process or generally not necessary given how familiar participants were with the standard PHQ-8:

> *Honestly, I don't really pay attention to the visuals, so I don't think they're necessary.* [Participant 3]

They later explained that just having the statement was sufficient and efficient:

> *it just makes it shorter, sweeter, faster, less to look at, and less to concentrate on.* [Participant 2]

Participant 20 felt similarly as she had regularly taken the PHQ-8 after attending therapy and psychiatry sessions for over 20 years and therefore preferred the traditional format she was so accustomed to.

Potential Visualization Improvements

We also asked participants to provide feedback on visualizations and share any ideas for potential improvements. Some feedback about the self-report interface would be easily addressable; for example, tweaking the digital survey to automatically shift to the next question could help respondents more smoothly transition from question to question:

> *Not the visuals, but I wish I didn't have to press next next next so many times.* [Participant 13]

Other feedback applied to specific PHQ-8 questions. For example, in question 7 ("Trouble concentrating on things, such as reading the newspaper or watching television"), our design increasingly blurred out the visual as the scale increased, and one participant (participant 5) expressed that this design was difficult to see given vision issues. Another participant emphasized that they desired a consistent runner emoji for question 8 (participant 17), and another participant requested questions 3 and 5 to be separated out (participant 12).

Survey Preference and Patient Histories

We observed a tendency for participants who had been more recently diagnosed to have a higher preference for a redesigned PHQ-8 questionnaire. More specifically, participants who had been diagnosed earlier preferred statements (average time since diagnosis 23, SD 17 y), whereas participants who had been diagnosed later preferred visualizations (average time since diagnosis 18, SD 11 y), and the most recently diagnosed participants preferred visualizations and questions (average time since diagnosis 16, SD 7 y). While this effect is not statistically significant when tested with a one-way ANOVA, it suggests that there is a potential trend where individuals with depression may become accustomed to a specific format over time, whereas those more recently diagnosed individuals may face challenges engaging with these conventional self-report instruments. Of participants who preferred the statements format (n=4), 50% (n=2) were not seeking therapy, 25% (n=1) were in therapy and taking antidepressants, and 25% (n=1) were taking antidepressants but not in therapy. On the other hand, of participants who preferred the statements and visualizations format (n=9), 22% (n=2) were not seeking therapy, 33% (n=3) were in therapy and taking antidepressants, 22% (n=2) were in therapy but not taking antidepressants, and 11% (n=1) were taking antidepressants but not in therapy. All participants who preferred the question and visualization (n=7) format of the survey were actively in therapy, and 86% (n=6) were also taking antidepressants.

Scoring of PHQ-8 Surveys Compared

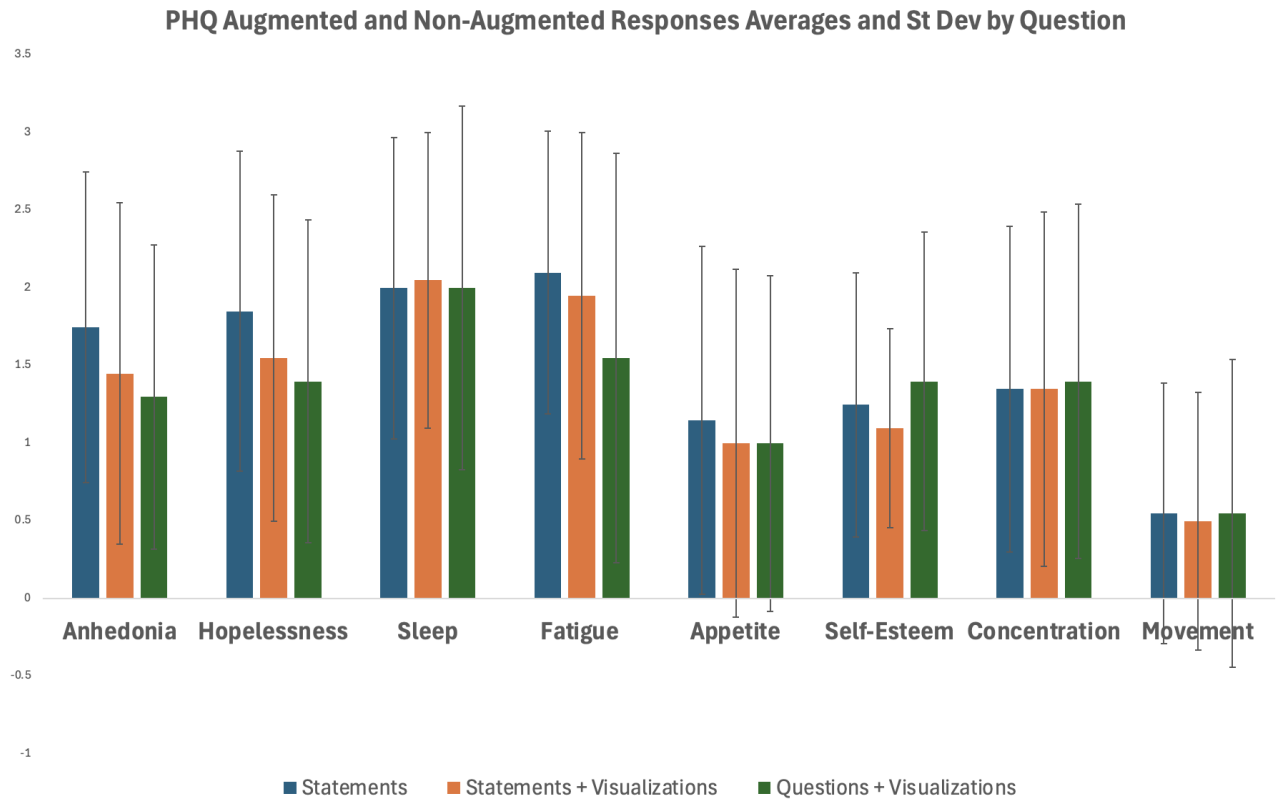
The PHQ-8 surveys included 8 questions, all on the same scaling of 0 to 3 with a maximum score of 24. In our sample, the average score of the PHQ-8 in standard form (Figure 3) was 11.95 (SD 5.03), 11 (SD 5.62) for our visual PHQ-8 design that augments the standard questionnaire with visual aids (Figure 1), and 10.6 (SD 5.45) for our reframed PHQ-8 (Figure 2). The total score for each PHQ-8 questionnaire showed ICCs that were highly consistent across the 3 surveys (average raters ICC=0.96, 95% CI 0.93-0.98; $P<.001$).

Individual Questions of PHQ-8 Surveys Compared

Overall, the movement, sleep, concentration, and appetite responses between the 3 versions stayed fairly consistent with the maximum change in averages at 0.15. However, questions related to more visible and stigmatized symptoms seemed to show more variability between the 3 versions, where anhedonia, hopelessness, fatigue, and self-esteem had still low but higher variability maximum change in averages at 0.55 between versions (Figure 4).

XSL•FO
RenderX

Figure 4. Each of the 3 versions of the Patient Health Questionnaire (PHQ)-8's individual question responses compared to each other.



The ICC showed high statistical significance across all 8 questions. For anhedonia and fatigue, the ICC estimates ranged from 0.64 to 0.70 (P<.001), whereas for hopelessness, sleep, appetite, concentration, and movement, the ICC estimates ranged from 0.88 to 0.96 (P<.001). Furthermore, all 8 questions show moderate to high correspondence with the traditional PHQ-8 and statistical significance. This correspondence helps to indicate that the augmented versions are able to effectively measure the psychometric properties assessed in the traditional version of the PHQ-8.

Discussion

Principal Findings

This study demonstrates that thoughtfully redesigning PHQ-8 items through the addition of visual elements and conversational reframing may improve people's subjective experience of completing depression screeners. Across our sample, most participants reported that the redesigned versions felt more engaging, supportive, or easier to interpret compared to the standard PHQ-8. While these adaptations do not change what is being measured, they appear to influence how users approach the task, with some describing the redesigned items as prompting more reflection or making the questions feel more personally meaningful. These findings align with broader work showing that visual and linguistic framing can shape how people make sense of self-report assessments. Specifically, we explored augmented versions of the PHQ-8 instruments, incorporating visualizations and reframed statements aimed at improving respondent engagement and comprehension. Specifically, we examined participant feedback after completing 3 versions of the PHQ-8: the standard form, a version augmented with visual aids, and a version that combined visual aids with item questions rephrased for greater clarity and compassion. We next reflect on the acceptability and overall experience of each of these versions.

Overwhelming Majority Enjoyed Augmented PHQ-8 Versions

We observed that 80% (16/20) of participants had a significantly greater preference for either augmented version of the PHQ-8 compared to the standard version of the instrument. For the 45% (9/20) of participants who most preferred the version augmented only with visuals, we observed themes related to how these additions promoted higher engagement and thoughtfulness in responses. Although augmented designs for health questionnaires have yet to be tested in the context of depression, previous work that tested the addition of visuals in patient populations such as chronic pain and affect found similar results, suggesting that visualizations may be broadly beneficial. Indeed, the addition of visuals is known to induce greater trustworthiness [50], reduce response burden, and increase response motivation [43,66]. Such findings were mirrored in this study, where patients reported feeling more engaged and thoughtful when using the augmented PHQ-8 with visualizations. Additionally, incorporating visuals into the questionnaires may elicit greater self-reflection, which could influence how participants articulate their internal states in their responses.

For the 35% (7/20) of patients who most preferred the addition of visuals and new question framings, we established a few main themes based on their feedback, which emphasized that these designs helped respondents feel more thoughtful, cared for, and understood. Such findings resonate with prior work

XSL•FO
RenderX

where patients were more comfortable, engaged, and authentic in responding to questions rather than statements [51,54-57].

Overall, our findings highlight that survey items augmented with visuals and revised phrasings help patients be more thoughtful while responding, feel more cared for and understood, and generally experience less response burden. An interesting future study, now that this redesign has been piloted, would be to integrate the feedback from our study and test the 3 versions with new patients who are not already so used to the standard format (ie, no change management needed).

The Minority of Patients Who Enjoyed the Standard Version

Examining the 20% (4/20) of patients who preferred the standard version of the PHQ-8, these participants highlighted that they found the visualizations and question changes to be distracting and that these augmentations lengthened the time it took for them to finish the questionnaire. Their feedback indicated that such reactions may stem from participants feeling a lack of connection to the visual and linguistic changes (eg, not resonating with the metaphor of energy as a battery visualization). Another potential factor we saw was a repetition and familiarity bias, where participants prefer surveys that they are more familiar with [67-69], given many of these participants had been diagnosed earlier and had been using the standard PHQ-8 for a longer number of years. Finally, these participants pointed out how the augmented questionnaire was longer and/or involved more content to consume. While our goal with the augmented questionnaires was to increase engagement and response motivation with these additional visual and textual elements [51,54-57], this feedback is a good reminder that such designs could be perceived as more burdensome or time-consuming for some individuals.

Patient Preferences in Relation to Demographic Characteristics and Survey Features

We first examined the impressions of those participants who preferred the PHQ-8 augmented with visual aids. These individuals displayed common characteristics in that they were diagnosed with MDD slightly later, on average, than those who preferred the standard statement-based survey. Additionally, more of these participants were actively seeking therapy. As mentioned, we suspect this may be due to less of a repetition bias or familiarity bias that may affect respondents who have a longer history of using the PHQ-8 as part of treatment [67-69]. Furthermore, therapy is known to not only help patients cope with their condition but also to develop greater condition acceptance and both somatic and emotional awareness [70-72]. Participants who are actively seeking therapy may therefore prefer an augmented survey that greater reflects their subjective awareness about condition on both cognitive, emotional, and physical levels.

Further exploring links between patient profiles and survey preferences, we observed a clear trend where participants with a lower depression score, higher trait hope scale, and lower stigma preferred the traditional PHQ-8 statements over the augmented versions. Said another way, as a participant's levels of depression and stigma increase and hope decreases, there appears to be a stronger preference for augmented PHQ-8 designs (Table 1). We believe this suggests the standard PHQ-8 may be better suited for someone with a more mild condition and that as condition severity worsens, the addition of visualizations and linguistic changes allows for the questionnaire to remain approachable while avoiding perceptions of stigma or judgment [28]. For example, this sense of judgment or marginalization can be seen in an aforementioned quote: "we know you're like this, just admit it." Given that prior research has demonstrated the impact that stigma and anhedonia have on survey responses [27-29] and that the current state of the PHQ-8 can fall short in accurately assessing MDD [3,7], it seems that a respondent's levels of perceived stigma are a key characteristic to understand when determining whether an augmented questionnaire may be more appropriate to deliver.

Survey Responses

The primary goal of this study was to determine if a PHQ augmented with visual and linguistic elements was acceptable and engaging for individuals with depression. Encouragingly, we found that 80% of participants did enjoy and prefer an augmented questionnaire. We also found that scores were reliable and accurate based on ICCs across all 3 questionnaires (2 augmented questionnaires and the 1 standard PHQ-8 questionnaire). Examining specific questionnaire items, we found that the movement, sleep, concentration, and appetite responses among the 3 versions had the highest agreement, whereas the more visible and stigmatized symptoms of anhedonia, hopelessness, fatigue, and self-esteem had moderate agreement compared to the same items with the traditional PHQ-8. Specifically, participants reported lower scores for these stigmatized symptoms on the PHQ-8 version augmented with visuals. We had hypothesized that our visuals would reduce shame and stigma, leading to higher scores. However, we found that the stigmatized symptom item scores reduced. We hypothesize that the addition of the visuals may have unintentionally softened or reframed the symptoms, which may have led to participants reporting symptoms with less severity. Another potential explanation for the score decrease is that individuals with higher stigma may underreport their symptoms with the addition of the visualizations. Therefore, future work is needed to examine this pattern and unpack how visualizations influence shame, stigma, and self-reporting.

Limitations and Improvements for Future Research

This study was able to demonstrate broad acceptability and engagement with the redesigned questionnaires; however, it had several limitations that need to be addressed in future studies. First, the sample size was limited to only 20 participants, which limits the generalizability of the study. The study was able to demonstrate that participants engaged with the new material. In future studies, we could further examine test-retest reliability and conduct a large-scale validation study. Additionally, participants completed the questionnaires after an interview about their depression symptoms and symptom management. Although participants' opinions on the visualizations should likely not have been influenced by this interview, future studies without an interview prior could better control for any possible effect. The study is limited by

XSL•FO
RenderX

demonstrating that participants preferred the visual format; however, this study did not test whether this preference accurately corresponds to improved internal validity or accuracy of PHQ-8 ratings. The variation of scores across the formats, although small, indicates that visual elements may influence responses, and therefore, future work would need to ensure psychometric validity of the new measures.

Our sample primarily relied on middle-aged adults (aged 40 - 65 y), which could have impacted the relationship with how participants interpreted and connected with the visual aids. Further testing across other age groups could determine if the effects from our study are reproducible age-independent. Additionally, in future studies, increasing the sample size could discern the impact of age and help-seeking behaviors on participants' views on each of the questionnaires and ensure test-retest reliability. One immediate next step is to explore a variety of additional questionnaire refinements based on participant feedback, including, as discussed previously, smoother transitions between questions, keeping visual designs sensitive to differences in people's visual abilities, using consistent emojis across item scales, and separating out questions that currently mix multiple assessments into one. In these future studies, we also intend to add a fourth survey option, which is to use the same open-ended questions tested in our study but without the addition of the visualizations. The questionnaire will retain its same length yet provide more empathetic and thought-provoking questions. It is necessary to examine more repeated and longitudinal testing of the redesigned instruments to assess their reliability and their user experience over time and in real-world self-report scenarios. Additionally, although the questionnaires had response options that were frequency-based, it is possible that some participants interpreted the visual aids as symptom intensity rather than frequency. It is possible that this juxtaposition could have influenced the interpretation of specific survey items. We intend to further explore how participants interpreted the addition of visuals and if they shifted from frequency to intensity of symptoms in future studies.

Conclusions

In exploring 2 new PHQ-8 designs that augmented the standard version with visuals as well as reframed item language, we found that participants overwhelmingly supported these augmented PHQ-8 versions. Participants expressed that they felt more engaged, understood, and thoughtful in making self-reports, suggesting that the augmented versions may better capture subjective experience and ultimately provide a more accurate diagnostic assessment. Furthermore, our findings highlight how integrating visual and linguistic modifications into depression screening tools can enhance patient experience without compromising the instrument's fundamental validity. Overall, this study provides design strategies and preliminary evidence in how to update widely used self-report instruments to be more patient centered.

Funding

This study was funded in part by the Dartmouth Cancer Center, the National Cancer Institute (grant P30CA023108), the National Institute on Drug Abuse (grant P30DA029926), the National Institute of Mental Health (grant R01MH123482), and Dartmouth College.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

NJ has received a research grant from Boehringer Ingelheim as principal investigator (although there is no salary support or direct payments to NJ, it supports research personnel and data collection).

References

1. Glenwright BG, Simmich J, Cottrell M, et al. Facilitators and barriers to implementing electronic patient-reported outcome and experience measures in a health care setting: a systematic review. J Patient Rep Outcomes 2023 Feb 14;7(1):13. [doi: 10.1186/s41687-023-00554-2] [Medline: 36786914]
2. Baas KD, Cramer AOJ, Koeter MWJ, van de Lisdonk EH, van Weert HC, Schene AH. Measurement invariance with respect to ethnicity of the Patient Health Questionnaire-9 (PHQ-9). J Affect Disord 2011 Mar;129(1-3):229-235. [doi: 10.1016/j.jad.2010.08.026] [Medline: 20888647]
3. Costantini L, Pasquarella C, Odone A, et al. Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): a systematic review. J Affect Disord 2021 Jan 15;279:473-483. [doi: 10.1016/j.jad.2020.09.131] [Medline: 33126078]
4. Kato TA, Hashimoto R, Hayakawa K, et al. Multidimensional anatomy of 'modern type depression' in Japan: a proposal for a different diagnostic approach to depression beyond the DSM - 5. Psychiatry Clin Neurosci 2016 Jan;70(1):7-23. [doi: 10.1111/pcn.12360] [Medline: 26350304]
5. Lamela D, Soreira C, Matos P, Morais A. Systematic review of the factor structure and measurement invariance of the Patient Health Questionnaire-9 (PHQ-9) and validation of the Portuguese version in community settings. J Affect Disord 2020 Nov 1;276:220-233. [doi: 10.1016/j.jad.2020.06.066] [Medline: 32697702]

XSL•FO
RenderX

6. Patel JS, Oh Y, Rand KL, et al. Measurement invariance of the Patient Health Questionnaire-9 (PHQ-9) depression screener in U.S. adults across sex, race/ethnicity, and education level: NHANES 2005-2016. Depress Anxiety 2019 Sep;36(9):813-823. [doi: 10.1002/da.22940] [Medline: 31356710]
7. Titov N, Andersson G. Using brief measures to identify depression and other mental disorders: a challenge for research and clinical practice. Internet Interv 2022 Apr;28:100450. [doi: 10.1016/j.invent.2021.100450] [Medline: 35646605]
8. Villarreal-Zegarra D, Copez-Lonzoy A, Bernabé-Ortiz A, Melendez-Torres GJ, Bazo-Alvarez JC. Valid group comparisons can be made with the Patient Health Questionnaire (PHQ-9): a measurement invariance study across groups by demographic characteristics. PLOS ONE 2019;14(9):e0221717. [doi: 10.1371/journal.pone.0221717]
9. Kaner EF, Haighton CA, McAvoy BR. "So much post, so busy with practice--so, no time!": a telephone survey of general practitioners' reasons for not participating in postal questionnaire surveys. Br J Gen Pract 1998 Mar;48(428):1067-1069 [FREE Full text] [Medline: 9624749]
10. Le A, Han BH, Palamar JJ. When national drug surveys "take too long": an examination of who is at risk for survey fatigue. Drug Alcohol Depend 2021 Aug 1;225:108769. [doi: 10.1016/j.drugalcdep.2021.108769] [Medline: 34049103]
11. Aarø LE, Fismen AS, Wold B, et al. Nordic adolescents responding to demanding survey scales in boring contexts: examining straightlining. J Adolesc 2022 Aug;94(6):829-843. [doi: 10.1002/jad.12066] [Medline: 35719057]
12. Brower CK. Too long and too boring: the effects of survey length and interest on careless responding. Master's thesis. Dayton, OH, USA: Wright State University; 2018 URL: https://etd.ohiolink.edu/acprod/odb_etd/etd/r/1501/10?clear=10&p10_accession_num=wright1524614764948287 [accessed 2025-04-23]
13. Bryant MD, Schoenberg ED, Johnson TV, Goodman M, Owen-Smith A, Master VA. Multimedia version of a standard medical questionnaire improves patient understanding across all literacy levels. J Urol 2009 Sep;182(3):1120-1125. [doi: 10.1016/j.juro.2009.05.027] [Medline: 19625036]
14. Dewaele JM. Online questionnaires. In: The Palgrave Handbook of Applied Linguistics Research Methodology: Palgrave Macmillan; 2018:269-286. [doi: 10.1057/978-1-137-59900-1_13]
15. Meier M, Martarelli CS, Wolff W. Is boredom a source of noise and/or a confound in behavioral science research? Humanit Soc Sci Commun 2024;11(1):10. [doi: 10.1057/s41599-024-02851-7]
16. Povoroznyuk R, Dzerovych N, Povoroznyuk V. A new voice: translating medical questionnaires. J World Lang 2016 May 3;3(2):139-159. [doi: 10.1080/21698252.2016.1262220]
17. Anhang Price R, Quigley DD, Hargraves JL, et al. A systematic review of strategies to enhance response rates and representativeness of patient experience surveys. Med Care 2022;60(12):910-918. [doi: 10.1097/MLR.0000000000001784]
18. Barbic S, Brooks E, Lassak N, et al. "It cannot be boring!": developing a measure of function for young adults accessing integrated youth services. J Patient Rep Outcomes 2022 Sep 3;6(1):92. [doi: 10.1186/s41687-022-00491-6] [Medline: 36057736]
19. McGrath P, Moore A, McNaught M, Palmer G, Greene A, Atkinson D. Another form to fill in! Clients' reflections on the hospice use of questionnaires. Support Care Cancer 2005 Sep;13(9):691-701. [doi: 10.1007/s00520-005-0795-6] [Medline: 15834741]
20. Horwitz AG, Zhao Z, Sen S. Peak-end bias in retrospective recall of depressive symptoms on the PHQ-9. Psychol Assess 2023 Apr;35(4):378-381. [doi: 10.1037/pas0001219] [Medline: 36757996]
21. Baryshnikov I, Aledavood T, Rosenström T, et al. Relationship between daily rated depression symptom severity and the retrospective self-report on PHQ-9: a prospective ecological momentary assessment study on 80 psychiatric outpatients. J Affect Disord 2023 Mar 1;324:170-174. [doi: 10.1016/j.jad.2022.12.127] [Medline: 36586594]
22. Velloza J, Njoroge J, Ngure K, et al. Cognitive testing of the PHQ-9 for depression screening among pregnant and postpartum women in Kenya. BMC Psychiatry 2020 Dec;20(1):31. [doi: 10.1186/s12888-020-2435-6]
23. Zotova N, Watnick D, Ajeh RA, et al. Understanding depression and the PHQ-9 items among people living with HIV: a multiple methods qualitative study in Yaoundé, Cameroon. SSM Ment Health 2024 Dec;6:100353. [doi: 10.1016/j.ssmmh.2024.100353] [Medline: 39749043]
24. Ma S, Tang Y, Wang X, Xiang Y, Luo X. Discrepancies between self-rated depression and observer-rated depression: the influence of perspective on assessment. J Affect Disord 2021;293:834-839. [doi: 10.1016/s0165-0327(99)00156-1] [Medline: 10940445]
25. Watson R, Harvey K, McCabe C, Reynolds S. Understanding anhedonia: a qualitative study exploring loss of interest and pleasure in adolescent depression. Eur Child Adolesc Psychiatry 2020 Apr;29(4):489-499. [doi: 10.1007/s00787-019-01364-y]
26. Wong S, Le GH, Phan L, et al. Effects of anhedonia on health-related quality of life and functional outcomes in major depressive disorder: a systematic review and meta-analysis. J Affect Disord 2024 Jul 1;356:684-698. [doi: 10.1016/j.jad.2024.04.086] [Medline: 38657767]
27. Davern J, O'Donnell AT. Stigma predicts health-related quality of life impairment, psychological distress, and somatic symptoms in acne sufferers. PLOS ONE 2018;13(9):e0205009. [doi: 10.1371/journal.pone.0205009] [Medline: 30265724]
28. Hahm S, Muehlan H, Stolzenburg S, Tomczyk S, Schmidt S, Schomerus G. How stigma interferes with symptom awareness: discrepancy between objective and subjective cognitive performance in currently untreated persons with mental health problems. Stigma Health 2020;5(2):146-157 [FREE Full text] [doi: 10.1037/sah0000184]

XSL•FO
RenderX

29. McGhie-Fraser B, McLoughlin C, Lucassen P, et al. Measuring persistent somatic symptom related stigmatisation: development of the Persistent Somatic Symptom Stigma scale for Healthcare Professionals (PSSS-HCP). J Psychosom Res 2024 Jun;181:111689. [doi: 10.1016/j.jpsychores.2024.111689] [Medline: 38704347]
30. Bianchi R, Verkuilen J, Toker S, et al. Is the PHQ-9 a unidimensional measure of depression? A 58,272-participant study. Psychol Assess 2022 Jun;34(6):595-603. [doi: 10.1037/pas0001124] [Medline: 35357877]
31. Ell K, Unützer J, Aranda M, Sanchez K, Lee PJ. Routine PHQ-9 depression screening in home health care: depression prevalence, clinical and treatment characteristics, and screening implementation. Home Health Care Serv Q 2006 Jan 24;24(4):1-19. [doi: 10.1300/J027v24n04_01]
32. Junaid K, Akram I, Daood M, Khan A. Validity of the patient health questionnaires (PHQ-2 and PHQ-9) for screening depression among human immunodeficiency virus patients in Lahore, Pakistan. World J Virol 2022 Sep 25;11(5):341-351. [doi: 10.5501/wjv.v11.i5.341] [Medline: 36188737]
33. Levis B, Benedetti A, Thombs BD, DEPRESsion Screening Data (DEPRESSD) Collaboration. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ 2019 Apr 9;365:l1476. [doi: 10.1136/bmj.l1476] [Medline: 30967483]
34. Maurer DM, Raymond TJ, Davis BN. Depression: screening and diagnosis. Am Fam Physician 2018 Oct 15;98(8):508-515 [FREE Full text] [Medline: 30277728]
35. Kanraweekultana N, Waijanya S, Promrit N, Korsanan A, Poolphol S. Depression analysis using behavior tracking and facial expression recognition during PHQ-9 assessment. ICIC Express Letters 2023;17:303-314. [doi: 10.24507/icicel.17.03.303]
36. Adams P, Murnane EL, Elfenbein M, Wethington E, Gay G. Supporting the self-management of chronic pain conditions with tailored momentary self-assessments. : Association for Computing Machinery; 2017 May 2 Presented at: CHI '17; May 6-11, 2017; Denver Colorado USA. [doi: 10.1145/3025453.3025832]
37. Eagle T, Blau C, Bales S, Desai N, Li V, Whittaker S. "I don't know what you mean by 'I am anxious'": a new method for evaluating conversational agent responses to standardized mental health inputs for anxiety and depression. ACM Trans Interact Intell Syst 2022 Jun 30;12(2):1-23. [doi: 10.1145/3488057]
38. Liem WW, Lattie EG, Taple BJ, et al. Improving collaborative management of multiple mental and physical health conditions: a qualitative inquiry into designing technology-enabled services for eliciting patients' values. Proc ACM Hum Comput Interact 2024 Nov;8(CSCW2):30. [doi: 10.1145/3687000] [Medline: 39822336]
39. Van Steijn F, Sogancioglu G, Kaya H. Text-based interpretable depression severity modeling via symptom predictions. : Association for Computing Machinery; 2022 Presented at: Proceedings of the 2022 International Conference on Multimodal Interaction (ICMI '22); Nov 7 to Jul 11, 2022; Bengaluru India p. 139-147. [doi: 10.1145/3536221.3556579]
40. Thieme A, Hanratty M, Lyons M, et al. Designing human-centered AI for mental health: developing clinically relevant applications for online CBT treatment. ACM Trans Comput-Hum Interact 2023 Apr 30;30(2):1-50. [doi: 10.1145/3564752]
41. Vossen W, Szymanski M, Verbert K. The effect of personalizing a psychotherapy conversational agent on therapeutic bond and usage intentions. 2024 Presented at: Proceedings of the 29th International Conference on Intelligent User Interfaces (IUI '24); Mar 18-21, 2024; Greenville SC USA p. 761-771. [doi: 10.1145/3640543.3645195]
42. Yahya Z, Faisal MR, Nugroho RA, et al. Design of application framework for vital sign monitoring and remote doctor consultation. : Association for Computing Machinery; 2023 Presented at: Proceedings of the 8th International Conference on Sustainable Information Engineering and Technology (SIET '23); Badung, Bali Indonesia p. 473-480. [doi: 10.1145/3626641.3626942]
43. Pollak JP, Adams P, Gay G. PAM: a photographic affect meter for frequent, in situ measurement of affect. 2011 Presented at: Proceedings of the International Conference on Human Factors in Computing Systems (CHI '11); May 7-11, 2011; Vancouver BC Canada p. 725-734. [doi: 10.1145/1978942.1979047]
44. Adams P, Rabbi M, Rahman T, et al. Towards personal stress informatics: comparing minimally invasive techniques for measuring daily stress in the wild. 2014 Presented at: Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth '14); May 20, 2026 to May 23, 2024; Oldenburg, Germany p. 72-79. [doi: 10.4108/icst.pervasivehealth.2014.254959]
45. Davies J, McKenna M, Denner K, Bayley J, Morgan M. The emoji current mood and experience scale: the development and initial validation of an ultra-brief, literacy independent measure of psychological health. J Ment Health 2024 Apr;33(2):218-226. [doi: 10.1080/09638237.2022.2069694] [Medline: 35535711]
46. Toet A, van Erp JBF. The EmojiGrid as a tool to assess experienced and perceived emotions. Psych 2019;1(1):469-481. [doi: 10.3390/psych1010036]
47. Kaye LK, Schweiger CR. Are emoji valid indicators of in-the-moment mood? Comput Human Behav 2023 Nov;148:107916. [doi: 10.1016/j.chb.2023.107916]
48. Jo H, Park DH. Affordance, usefulness, enjoyment, and aesthetics in sustaining virtual reality engagement. Sci Rep 2023 Sep 12;13(1):15097. [doi: 10.1038/s41598-023-42113-1] [Medline: 37699934]
49. Tse MMY, Ng SSM, Bai X, et al. An exploration of the use of visually appealing contexts in a pain management program. Eur Geriatr Med 2020 Oct;11(5):821-827. [doi: 10.1007/s41999-020-00339-6] [Medline: 32564342]

XSL•FO
RenderX

50. Lindgaard G, Dudek C, Sen D, Sumegi L, Noonan P. An exploration of relations between visual appeal, trustworthiness and perceived usability of homepages. ACM Trans Comput-Hum Interact 2011 Apr;18(1):1-30. [doi: 10.1145/1959022.1959023]
51. Ward HO, Kibble S, Mehta G, et al. How asking patients a simple question enhances care at the bedside: medical students as agents of quality improvement. Perm J 2013;17(4):27-31. [doi: 10.7812/TPP/13-028] [Medline: 24361017]
52. Fisher KA, Smith KM, Gallagher TH, Huang JC, Borton JC, Mazor KM. We want to know: patient comfort speaking up about breakdowns in care and patient experience. BMJ Qual Saf 2019 Mar;28(3):190-197. [doi: 10.1136/bmjqs-2018-008159] [Medline: 30269059]
53. Kwame A, Petrucka PM. A literature-based study of patient-centered care and communication in nurse-patient interactions: barriers, facilitators, and the way forward. BMC Nurs 2021 Sep 3;20(1):158. [doi: 10.1186/s12912-021-00684-2] [Medline: 34479560]
54. Rouder J, Saucier O, Kinder R, Jans M. What to do with all those open-ended responses? Data visualization techniques for survey researchers. Surv Pract 2021;14(1):1-9. [doi: 10.29115/SP-2021-0008]
55. Writing survey questions. Pew Research Center. 2023. URL: https://www.pewresearch.org/writing-survey-questions/ [accessed 2025-04-23]
56. Duan N, Alegria M, Canino G, McGuire TG, Takeuchi D. Survey conditioning in self-reported mental health service use: randomized comparison of alternative instrument formats. Health Serv Res 2007;42:890-906. [doi: 10.1111/j.14756773.2006.00618.x]
57. Sinkowitz-Cochran RL. Survey design: to ask or not to ask? That is the question. Clin Infect Dis 2013 Apr;56(8):1159-1164. [doi: 10.1093/cid/cit005] [Medline: 23315318]
58. Cohen S, Hoberman HM. Positive events and social supports as buffers of life change stress. J Applied Social Pyschol 1983 Apr;13(2):99-125 [FREE Full text] [doi: 10.1111/j.1559-1816.1983.tb02325.x]
59. Cohen S, Mermelstein R, Kamarck T, Hoberman HM. Measuring the functional components of social support. In: Sarason IG, Sarason BR, editors. Social Support: Theory, Research, and Applications: Martinus Niijhoff; 1985:73-94 URL: https://www.slu.edu/medicine/family-medicine/-pdf/interpersonal-support-eval.pdf [accessed 2025-04-23] [doi: 10.1007/978-94-009-5115-0_5]
60. Smedema SM, Pfaller J, Moser E, Tu WM, Chan F. Measurement structure of the trait hope scale in persons with spinal cord injury: a confirmatory factor analysis. Rehabilitation Research Policy and Education 2013;27(3):206-212. [doi: 10.1891/2168-6653.27.3.206]
61. Snyder CR, Harris C, Anderson JR, et al. The will and the ways: development and validation of an individual-differences measure of hope. J Pers Soc Psychol 1991 Apr;60(4):570-585. [doi: 10.1037//0022-3514.60.4.570] [Medline: 2037968]
62. Dweck CS. Kind of person' implicit theory scale. SPARQtools. 1999. URL: http://sparqtools.org/mobility-measure/kind-of-person-implicit-theory-scale/ [accessed 2025-04-23]
63. Dweck CS, Chiu C, Hong Y. Implicit theory scales. APA PsycNet. 1995. URL: https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft49459-000 [accessed 2025-04-23]
64. Dweck CS. Growth mindset scale. SPARQtools. 1999. URL: https://sparqtools.org/mobility-measure/growth-mindset-scale/ [accessed 2025-04-23]
65. Molina Y, Choi SW, Cella D, Rao D. The stigma scale for chronic illnesses 8-item version (SSCI-8): development, validation and use across neurological conditions. Int J Behav Med 2013 Sep;20(3):450-460. [doi: 10.1007/s12529-012-9243-4] [Medline: 22639392]
66. Tsonos D, Kouroupetroglou G. A methodology for the extraction of reader's emotional state triggered from text typography. In: Tools in Artificial Intelligence: National and Kapodistrian University of Athens; 2008:1-15 URL: https://www.intechopen.com/chapters/5301 [accessed 2025-04-23] [doi: 10.5772/6071]
67. Chew SH, Ebstein RP, Zhong S. Ambiguity aversion and familiarity bias: evidence from behavioral and gene association studies. J Risk Uncertain 2012 Feb;44(1):1-18. [doi: 10.1007/s11166-011-9134-0]
68. Hassan A, Barber SJ. The effects of repetition frequency on the illusory truth effect. Cogn Research 2021 Dec;6(1):38. [doi: 10.1186/s41235-021-00301-5]
69. De Vries A, Erasmus PD, Gerber C. The familiar versus the unfamiliar: familiarity bias amongst individual investors. Acta Commercii 2017;17(1) [FREE Full text] [doi: 10.4102/ac.v17i1.366]
70. Brani O, Hefferon K, Lomas T, Ivtzan I, Painter J. The impact of body awareness on subjective wellbeing: the role of mindfulness. Intern Body Psychother J 2014;13(Spring 2014):19-30 [FREE Full text]
71. Dindo L, Van Liew JR, Arch JJ. Acceptance and commitment therapy: a transdiagnostic behavioral intervention for mental health and medical conditions. Neurotherapeutics 2017 Jul;14(3):546-553. [doi: 10.1007/s13311-017-0521-3] [Medline: 28271287]
72. Lane RD. Measurement of levels of emotional awareness before, during, and after psychotherapy. J Psychother Integr 2023;33(2):141-159. [doi: 10.1037/int0000297]

XSL•FO
RenderX

Abbreviations

ICC: intraclass correlation coefficient
MDD: major depressive disorder
PHQ-8: Patient Health Questionnaire-8

Edited by A Kushniruk; submitted 24.Apr.2025; peer-reviewed by K Otani, M Klee, R Marshall; revised version received 08.Jan.2026; accepted 19.Jan.2026; published 21.Apr.2026.

Please cite as:
Dorris H, Sheldon J, Sepucha C, Makau EK, Traoré PS, Murnane E, Jacobson N
"Helpful in Thinking Through the Question": Redesigning the Patient Health Questionnaire-8 for Depression Assessment to Enhance Patient Engagement and Understanding
JMIR Hum Factors 2026;13:e76230
URL: https://humanfactors.jmir.org/2026/1/e76230
doi:10.2196/76230

© Hannah Dorris, Jenna Sheldon, Courtney Sepucha, Emmanuel Kasuti Makau, Pape Sow Traoré, Elizabeth Murnane, Nicholas Jacobson. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 21.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mode Effects Between Telephone and Web Interviews in the Post-COVID-19 Questionnaire Survey CoVerlauf: Exploratory Study

Paula Sofia Herrera-Espejel[1,2], MPP; Hermann Pohlabeln[3], PhD; Lisa Kühne[4], MSc; Stefan Rach[1,2], PhD

[1]Department of Epidemiological Methods and Etiological Research, Leibniz Institute for Prevention Research and Epidemiology - BIPS, Achterstr. 30, Bremen, Germany

[2]Leibniz ScienceCampus Digital Public Health, Bremen, Germany

[3]Department of Statistical Methods in Epidemiology, Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany

[4]SOCIUM Research Center on Inequality and Social Policy, Department of Health, Long-Term Care and Pensions, University of Bremen, Bremen, Germany

Corresponding Author:
Stefan Rach, PhD
Department of Epidemiological Methods and Etiological Research, Leibniz Institute for Prevention Research and Epidemiology - BIPS, Achterstr. 30, Bremen, Germany

Abstract

Background: Concurrent mixed mode designs for data collection may introduce mode-specific biases. This study investigated mode effects between computer-assisted website interviewing (CAWI) and computer-assisted telephone interviewing (CATI) in the concurrent mixed mode study CoVerlauf.

Objective: This secondary analysis of previously published results explored associations between the choice of interview mode and socioeconomic characteristics, the number of reported post-COVID-19 symptoms, and response patterns in symptoms collected via multiple-choice (MC) and free-text items.

Methods: Associations were calculated using multiple logistic regression analyses, mixed effects logistic regression analyses, and gamma generalized linear models. Selection and measurement effects in symptom counts were analyzed with the back-door method.

Results: Among 1779 participants, 77.8% (1384/1779) chose CAWI, and 22.2% (395/1779) chose CATI. Odds of selecting CATI were lower with a higher education (odds ratio [OR] 0.46, 95% CI 0.34 to 0.63), for those represented by another person (OR 0.28, 95% CI 0.13 to 0.56), and with each additional symptom experienced at the time of the interview collected from MC (OR 0.86, 95% CI 0.77 to 0.96). Compared with those aged 50 years to 59 years, the odds increased with age (60 - 69 years: OR 2.26, 95% CI 1.53 to 3.36; 70 - 79 years: OR 4.98, 95% CI 3.12 to 7.97; ≥80 years: OR 17.29, 95% CI 9.34 to 33.11) and each additional symptom experienced at infection collected via MC (OR 1.09, 95% CI 1.04 to 1.15) and free-text items (OR 1.94, 95% CI 1.61 to 2.37). Odds of responding to free-text items were higher with CATI (OR 5.22, 95% CI 4.25 to 6.42), for women (OR 1.57, 95% CI 1.30 to 1.89), for those with ≥3 pre-existing conditions (OR 1.44, 95% CI 1.07 to 1.96), with missing information on educational attainment (OR 1.95, 95% CI 1.11 to 3.42), and for each additional MC symptom (OR 1.15, 95% CI 1.12 to 1.19). Odds of collecting additional symptoms from free-text items were lower with CATI (OR 0.58, 95% CI 0.41 to 0.82) but increased with each additional MC symptom (OR 1.08, 95% CI 1.02 to 1.14), ≥2 pre-existing conditions (2: OR 1.8, 95% CI 1.02 to 3.00; ≥3: OR 2.0 95% CI 1.18 to 3.43), and ages 40 years to 49 years (OR 2.0, 95% CI 1.13 to 3.55). Selection effects were found for the number of symptoms at infection collected via MC (difference –0.63, 95% CI –0.95 to –0.31) as were measurement effects for symptoms collected via MC at infection (difference 0.97, 95% CI 0.47 to 1.5), via free text at infection (difference 0.40, 95% CI 0.29 to 0.52), and via free text at the time of the interview (difference 0.15, 95% CI 0.05 to 0.25).

Conclusions: Our results suggest that interviews via CATI resulted in higher numbers of reported symptoms collected via MC and free text and that both selection and measurement effects contributed to this increase. However, no evidence was found that the interview mode affected the overall findings of the CoVerlauf study.

(JMIR Hum Factors 2026;13:e80631) doi:10.2196/80631

KEYWORDS

concurrent mixed mode; data collection; CATI; CAWI; population-based; CoVerlauf; post-COVID-19; self-selection effect; mode measurement effect

XSL•FO
RenderX

Introduction

Background

Researchers often offer study participants the choice among different interview modes (eg, online, telephone, in person, or paper and pencil) with the purpose of broadening participation [1]. Computer-assisted website interviewing (CAWI) and computer-assisted telephone interviewing (CATI) are two modes commonly used as data collection methods in population-based studies. CAWI is a self-administered, web-based reporting method in which respondents are provided with a link to an online questionnaire that they generally complete on their own and on an internet-enabled device of their preference [2]. CATI, by contrast, is an interviewer-administered method, where trained interviewers serve as intermediaries to ask each question aloud and record responses using a combination of telephone communication and computer-guided technology [3]. Although CAWI mainly uses visual communication channels handled by the participants themselves, CATI uses aural communication and relies on the presence of the interviewers to capture responses.

The use of different interview modes can influence how study data are collected. On the one hand, preferences for interview modes can be significantly associated with selection effects: Participants picking the same mode may also share similar socioeconomic and demographic characteristics, including their use of digital technology [4]. For example, younger age groups from higher socioeconomic backgrounds and urban areas might be more inclined to use digital communication methods, while older age groups or those who habitually stay at home (eg, due to unemployment) might prefer analog ones [1] for exchanging information [5-8]. On the other hand, the use of different interview modes may also lead to measurement effects: systematic variations in how participants respond to identical questions caused by mode-specific intrinsic features (see [9,10]). For instance, interview modes involving a direct interaction with interviewers are subject to different kinds and degrees of information bias than self-administered modes (eg, diagnostic bias or social desirability bias). Collectively, population-based studies may face a situation where participants with similar characteristics choose the same interview mode, which, in turn, subjects them to data collection procedures that induce systematically different response patterns.

Existing research on the effects of mode-specific differences between CAWI and CATI (eg, visual vs aural communication channels and absence vs presence of interviewer, respectively) on response frequency and content has yielded mixed results. From one side, research suggests that CAWI respondents may be more likely to complete online questionnaires with minimal time and cognitive effort, providing shorter, if not incomplete, answers than with other modes [11]. From another side, there is evidence that CAWI may lead to more sincere and extensive answers because it allows participants to respond more calmly, at one's own pace, and with a higher sense of privacy [7]. Correspondingly, the odds of social desirability bias (ie, inclination to adjust answers to conform with anticipated expectations and norms) may vary between interview modes when interrogated about sensitive health topics. For instance, research reports lower odds of social desirability when using CAWI for answers concerning HIV testing [12], whereas higher odds are reported when using CATI for answers about symptoms of infectious diseases, mental health, or substance abuse due to the presence of an interviewer [13-15].

At the same time, research on interviewer effects has indicated that the nature of the spoken interaction in CATI results in a higher number of responses than in self-administered modes, such as CAWI [16]. Furthermore, related research shows mixed results on the effect of the different modes depending on whether the questions are with or without predefined response categories (ie, multiple-choice vs free-text items). Although one meta-analysis found that item response behavior was similar between question formats across all methods [17], a more recent study suggested that CATI respondents systematically selected more options within multiple-choice items, resulting in higher item response proportions to certain questions than with CAWI [18]. Similarly, another study concluded that CAWI questionnaires resulted in fewer free-text items overall compared with CATI [11].

Aside from frequency of response, another line of research has also pointed to between-mode differences in the content of answers. For example, the content of free-text items regarding satisfaction with particular topics in CAWI may be more likely to contain negative feedback than that in CATI [19]. Past studies have also demonstrated that free-text items are particularly prone to interviewer and recall effects in spoken interactions, potentially resulting in differential reporting between CATI and CAWI [16,20,21]. The literature also suggests that, although CAWI might induce primacy effects (eg, inclination to select the first options read from a list of options) [22,23], it is unclear whether CATI is more susceptible to either primacy or recency effects (eg, inclination to select the last options presented from the same list read aloud) [13].

These findings are reiterated in medical and epidemiological literature specifically studying the effect of interviewer- and self-administered mode choices and question format on response patterns. For example, Milton and colleagues [24] found that, related to sensitive mental health and well-being topics, the reporting of symptoms by young Australians (aged 16 years to 25 years) is dependent on whether the interview is self-administered. Overall, they observed a higher disclosure of such high-sensitive items in online-based surveys. Similarly, Zager Kocjan and colleagues [25] highlighted that merging data collected from items asking on psychological functioning from different modes into a single dataset may result in structural noninvariance with significant mean differences between these two types of modes. Namely, the authors observed a greater tendency for respondents to report fewer depressive symptoms when interviewers were directly involved in data collection, potentially resulting in the underestimation of the actual prevalence or severity of symptoms compared with responses collected anonymously through an online portal. The auditory versus visual nature of the interview is also found to have systematic differential response order effects on how respondents report sensitive symptoms [26]. For example, Cernat and colleagues [26] found that items collected from the Center

XSL•FO
RenderX

for Epidemiologic Studies Depression scale are sensitive to primacy and recency effects depending on the interview mode. Although CAWI respondents tended to select items from the beginning of the lists, CATI respondents were more likely to choose those at the end of the list.

Research Objective

This study investigated the potential effects of using different interviewing modes (CATI vs CAWI) on participant response patterns to questions about symptoms experienced in the CoVerlauf study. Conducted in Germany in 2021, the CoVerlauf study provided evidence on the odds of experiencing a post-COVID-19 condition after contracting a verified SARS-CoV-2 infection [27]. The study concluded that more than 12% of participants (203/1612, 12.6%) who reported at least one symptom at the time of their infection experienced symptoms of a post-COVID-19 condition 3 months after the infection (ie, ongoing symptoms of fatigue, breathing difficulties, and cognitive complaints following their initial infection). Since all enrolled participants could select either CAWI or CATI to complete the study questionnaire, mode effects might have impacted data collection and findings. In particular, the choice of the interviewing mode may have not only been associated with the participants' characteristics and health status but also influenced how participants interacted with the multiple-choice and free-text items beyond their shared characteristics.

This study serves two main research objectives: (1) to determine whether there is an established link between the participants' choice of the interview mode (CAWI vs CATI); their health, socioeconomic, and demographic characteristics; and the number of reported post-COVID-19 symptoms; and (2) to identify whether the free-text items regarding if participants experienced "any other symptom" aside from those listed in the multiple-choice items systematically differed between the two mode groups.

Methods

Ethical Considerations

All procedures in the CoVerlauf study were performed in accordance with the ethical standards of the institutional or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The CoVerlauf study was approved by the Ethics committee of Bremen University (reference number 2020 - 05-EILV). Written consent was obtained from all participants in the study or the legal guardians of those aged younger than 18 years. Participation was voluntary and without any monetary or nonmonetary compensation. The article does not include personal information, identifying details, or identifiable features of research participants.

Study Design

This study used existing data from 1779 individuals with a verified SARS-CoV-2 infection aged between 1 year and 98 years from the City Municipality of Bremen, Germany, who participated in the CoVerlauf study [27]. A detailed description of the study design and the recruitment protocol as well as the questionnaire in German and in English can be found in [27]. Briefly summarized, CoVerlauf explored how even mild cases of COVID-19 could also be associated with a post-COVID-19 condition, defined as reporting at the time of the participants' interview 3 months after the infection at least two of the following symptoms: fatigue, a cognitive disorder, or breathing difficulties. The study used a questionnaire survey to collect self-reported participant data on the symptoms they experienced at 2 different time points. Family members or caretakers could also answer the questionnaire in the place of persons unable to answer the questionnaire by themselves (eg, minors, older adults, or incapacitated persons). Persons were invited via land mail by the Public Health Department of the City Municipality of Bremen (Germany) if they had either tested positive for SARS-CoV-2 or been diagnosed with COVID-19 by a physician between March 1, 2020, and January 31, 2021. One reminder letter was sent out 12 weeks to 16 weeks after the initial invitation letter. Enrolled participants were first contacted by field staff to determine their preferred interviewing mode for completing the questionnaires (CATI or CAWI; concurrent mixed mode design [28]). CATI interviews were conducted by study nurses who all had multiple years of experience in recruitment and data collection for epidemiologic population-based studies and were trained in conducting standardized interviews. To record participants' answers, interviewers used a special version of the online questionnaire that, in addition to the questions, included instructions and transitional statements to be read out during the interviews [2]. CAWI questionnaires were self-administered by the participants.

A first set of questions asked participants to recall the symptoms they experienced at the time of their infection and the first weeks afterward (hereafter, referred to as "time of infection"). A second set of questions asked participants about their long-term symptoms at the time of the interview, which was conducted at least 90 days after their diagnosis (hereafter, "time of interview"). Two question formats were used to collect the information on symptoms experienced at both time points. First, participants were provided with a multiple-choice list of 13 common COVID-19 infection symptoms (common at the time of questionnaire inception): fatigue, respiratory problems, diarrhea, fever, joint pain, scratchy throat, cough, headache, sniffles, chills, nausea, loss of taste, and loss of smell. Second, a free-text item offered the opportunity to provide additional symptoms.

Free-text items were subjected to a word frequency analysis that resulted in the classification of 15 additional symptoms: "hair loss," "tightness of chest," "circulatory problems," "cognitive symptoms," "dermatological skin symptoms," "eye symptoms," "gastrointestinal symptoms," "heart problems," "loss of appetite," "mental difficulties," "nose & ear symptoms," "sensory skin symptoms," "sleep difficulties," and "flushing & sweats." The free-text items were then subject to an independent screening, coding, and consensus procedure by 2 reviewers. The study also gathered data on the participants' weight and height, pre-existing conditions, and socioeconomic status. Additionally, the duration required to complete the questionnaire was automatically logged by the CAWI and CATI software and measured in minutes for each respondent. In 4 instances, the

XSL•FO
RenderX

interview duration was longer than 1000 minutes, indicating that the interview was interrupted for more than 16 hours (CATI: 3; CAWI: 1). These outliers were excluded for the creation of figures (maximum duration after removal: 110 min) and replaced with imputed values using linear regression imputation with the MICE (multivariate imputation by chained equations) algorithm [29] for the regression analysis. A sensitivity analysis, which involved repeating the regression as a complete-case analysis excluding the 4 outliers, revealed no differences (Multimedia Appendix 1).

Statistical Methods

This study first examined the interviewing mode preference (CATI vs CAWI) of participants to the CoVerlauf study. A multiple logistic regression analysis was conducted to assess the relationship between the choice of mode and the number of symptoms collected from answers to the multiple-choice and the free-text items at the time of infection and interview separately. The corresponding regression model (equation 1 in Multimedia Appendix 2) included sociodemographic and health indicators as explanatory variables, namely sex (female vs male), age categorized into 8 age groups (0 - 17, 18 - 29, 30 - 39, 40 - 49, 50 - 59, 60 - 69, 70 - 79, ≥80 years), educational attainment based on the International Standard Classification of Education (ISCED) scale (low: 1, 2; medium: 3, 4; high: 5, 6), weight status categorized according to BMI and calculated from self-reported body weight and height (normal: BMI <25 kg/m²; overweight: BMI 25 - 30 kg/m²; obese: BMI ≥30 kg/m²), and the number of pre-existing conditions (0, 1, 2, vs ≥3). The model also included a dichotomous variable "self-reported" ("self" vs "proxy") to differentiate whether the participant self-reported their answers or had another person answer on their behalf. As in the original publication [27], missing values in the variables for weight status and educational attainment were coded as a separate category "Missing."

Furthermore, the study investigated the role of the mode in response patterns to the 2 free-text items on additional symptoms through 2 separate mixed effects multiple logistic regression analyses (equations 2 and 3 in Multimedia Appendix 2). For theses analyses, data were restructured into the long format, where participants were represented with 2 rows: one for their responses at the time of infection and another for those at the time of interview. As such, the 2 models were conducted on a dataset with twice the number of observations than the number of participants considered in each of the analyses. The first model examined the association between the mode (CATI or CAWI) and answering free-text items ("response" vs "no response"). The second model focused on the subset of individuals who answered these questions, assessing how the interview mode influenced the probability of identifying and collecting at least one or more additional symptoms (+1 symptom) compared with none (0 symptoms). Both analyses included the aforementioned sociodemographic and health variables. To account for potential dependencies within individuals, both models incorporated random effects for responses from the same individuals (ie, ID) and time point of the reported symptoms (ie, time of infection and time of interview).

A gamma generalized linear regression model with log link function was used to analyze the outcome variable of overall interview duration (in minutes) and estimate time differences between the 2 mode groups (equation 4 in Multimedia Appendix 2). The choice of the model was based on the overall right-skewed distribution and because the outcome variable is continuous.

The back-door method was used to disentangle selection and measurement effects in the observed mode effects on symptom counts [30,31]. Derived from causal inference theory, this method consists of comparing answers to one of the interview modes (principal mode) to counterfactual answers generated from other modes (alternative modes) that assume the respondents had used the principal mode. To create counterfactual answers, propensity score weights are calculated from a set of variables that are assumed to be insensitive to the choice of the interview mode (eg, age). Outcome differences between the unweighted and weighted answers in the alternative mode are assumed to capture selection effects, whereas differences between the weighted alternative mode and the principal mode quantify measurement effects. For the current analysis, CAWI was selected as the principal mode, and a set of counterfactual answers $CATI_W$ was created using propensity score weights, which were calculated using the interview mode as the outcome variable and sex, age category, educational attainment (low and medium vs high), and number of pre-existing conditions (0, 1, and 2 vs ≥3) as independent variables. Mean differences and 95% CIs were calculated for each of the 4 symptom counts (ie, from the multiple-choice and free-text items at time of infection and interview). In addition, we calculated standardized mean differences (SMDs) with 95% CIs to make the effects comparable across outcome variables.

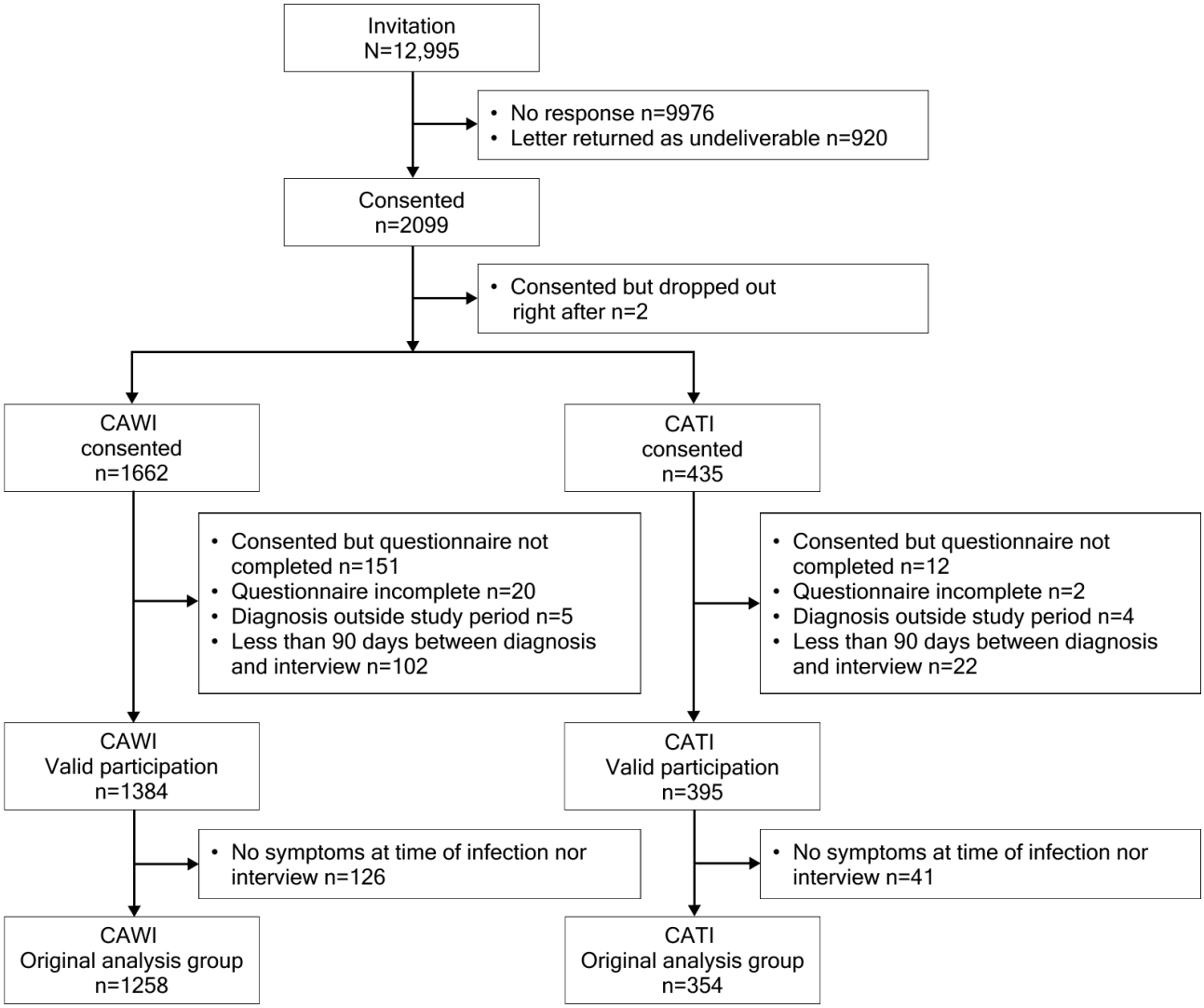
All statistical analyses were executed using R version 4.3.3. The logistic regression model was executed using the stats package [32], while the mixed effect logistic regression models used the lme4 package in R [33]. Propensity score weights were calculated using the package WeightIt [34], and SMDs were calculated using the package smd [35]. Results for all regression models with interview mode as outcome are reported as odds ratios (OR) with 95% CIs, and results from the model with duration as outcome are reported as mean ratios (MRs) with 95% CIs. Subgroup differences were assessed with Pearson χ^2 tests with Yates continuity correction.

Results

Among all enrolled participants (n=1779), 77.8% (1384/1779) used CAWI, and 22.2% (395/1779) used CATI to complete the CoVerlauf questionnaire (Figure 1).

XSL•FO
RenderX

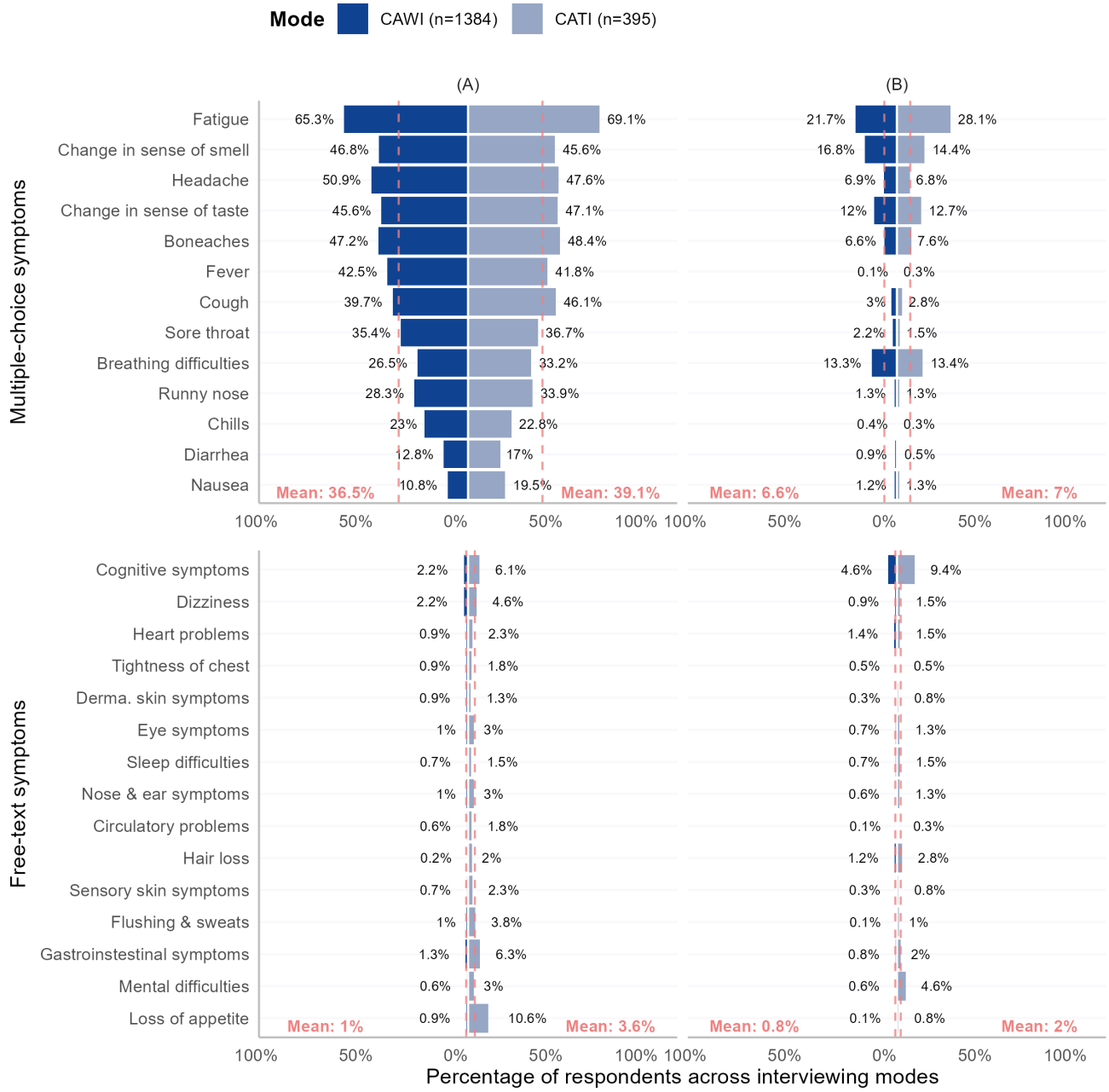
Figure 1. Flowchart of study participants across interview modes of computer-assisted website interviewing (CAWI) and computer-assisted telephone interviewing (CATI).



Compared with CAWI participants, the percentages of participants reporting symptoms was consistently higher with CATI at both time points and across the 13 multiple-choice and 15 free-text items (Figure 2).

XSL•FO
RenderX

Figure 2. Comparison of percentage of participants reporting post-COVID-19 symptoms at the (A) time of infection and (B) time of interview in multiple-choice and free-text items across interviewing modes. CATI: computer-assisted telephone interviewing; CAWI: computer-assisted website interviewing.



The multiple logistic regression analysis assessing associations between sample characteristics and the choice of CATI against CAWI revealed significant differences across most age groups, educational ISCED attainment categories, and the number of symptoms submitted at both the time of infection and interview for both question formats (Table 1). The odds of choosing CATI were lower for participants with a high education (OR 0.46, 95% CI 0.34 - 0.63) relative to those with an upper-secondary education level. In comparison with participants 50 years to 59 years old, participants became more likely to choose CATI with increasing age (60 - 69 years: OR 2.26, 95% CI 1.53 - 3.36; 70 - 79 years: OR 4.98, 95% CI 3.12 - 7.97; ≥80 years: OR 17.29, 95% CI 9.34 - 33.11). Those in age groups younger than 50 years to 59 years old were increasingly less likely to choose CATI, although ORs significantly differing from 1 were only observed in the age group 18 years to 29 years (OR 0.56, 95% CI 0.33 - 0.93). Participants who had their questionnaires completed by another person (ie, proxy respondents) were overall less likely to use CATI (OR 0.28, 95% CI 0.13 - 0.56) than those who responded for themselves. Note that participants in the lowest age group were 1.6 times more likely to use CATI (OR 1.63, CI 0.80 - 3.32), which might be partly explained by the high proportion of their questionnaires completed by a proxy respondent (Multimedia Appendix 3).

XSL•FO
RenderX

Table . Sample characteristics by interview mode and odds ratios for selecting computer-assisted telephone interviewing (CATI) over computer-assisted website interviewing (CAWI) from multiple logistic regression analysis.

Predictors	Interview mode (n=1779), n (%)		Odds ratio (95% CI)
	CATI (n=395)	CAWI (n=1384)	
N_{Sx}[a] infection MC[b]	5.09 (3.40)[c]	4.75 (2.93)[c]	1.09 (1.04 - 1.15)
N_{Sx} infection free-text[d]	0.534 (0.899)[c]	0.150 (0.540)[c]	1.94 (1.61 - 2.37)
N_{Sx} interview MC	0.909 (1.26)[c]	0.862 (1.38)[c]	0.86 (0.77 - 0.96)
N_{Sx} interview free-text	0.299 (0.745)[c]	0.126 (0.443)[c]	1.21 (0.94 - 1.54)
Sex			
Male	170 (43)	590 (42.6)	1 (reference)
Female	225 (57)	794 (57.4)	0.85 (0.65 - 1.10)
Age group (years)			
0 - 17	25 (6.3)	102 (7.4)	1.63 (0.80 - 3.32)
18 - 29	26 (6.6)	211 (15.2)	0.56 (0.33 - 0.93)
30 - 39	28 (7.1)	222 (16)	0.72 (0.43 - 1.19)
40 - 49	39 (9.9)	234 (16.9)	0.89 (0.57 - 1.39)
50 - 59	76 (19.2)	339 (24.5)	1 (reference)
60 - 69	78 (19.7)	181 (13.1)	2.26 (1.53 - 3.36)
70 - 79	62 (15.7)	73 (5.3)	4.98 (3.12 - 7.97)
≥80	61 (15.4)	22 (1.6)	17.29 (9.34 - 33.11)
Weight status according to BMI categories			
Normal	183 (46.3)	698 (50.4)	1 (reference)
Overweight	116 (29.4)	437 (31.6)	0.81 (0.59 - 1.11)
Obese	95 (24.1)	242 (17.5)	1.10 (0.76 - 1.57)
Missing	1 (0.3)	7 (0.5)	0.58 (0.03 - 3.52)
Number of pre-existing conditions			
0	146 (37)	784 (56.6)	1 (reference)
1	110 (27.8)	310 (22.4)	1.24 (0.89 - 1.72)
2	64 (16.2)	156 (11.3)	1.21 (0.80 - 1.83)
≥3	75 (19)	134 (9.7)	1.06 (0.68 - 1.63)
Education (ISCED[e])			
Low (1, 2)	63 (15.9)	155 (11.2)	1.48 (0.92 - 2.35)
Medium (3, 4)	238 (60.3)	635 (45.9)	1 (reference)
High (5, 6)	87 (22)	543 (39.2)	0.46 (0.34 - 0.63)
Missing	7 (1.8)	51 (3.7)	0.46 (0.16 - 1.14)
Proxy versus self			
Self	381 (96.5)	1307 (94.4)	1 (reference)
Proxy	14 (3.5)	77 (5.6)	0.28 (0.13 - 0.56)

[a]N_{Sx}: number of symptoms.

[b]MC: multiple-choice item.

[c]Mean (SD).

[d]Free-text item.

[e]ISCED: International Standard Classification of Education.

XSL•FO
RenderX

The odds of selecting CATI increased by 9% and almost doubled for each additional symptom collected from the multiple-choice (OR 1.09, 95% CI 1.04 - 1.15) and free-text (OR 1.94, 95% CI 1.61 - 2.37) items inquiring about symptoms experienced at the time of infection. The odds of selecting CATI were 14% lower for each additional symptom collected from the multiple-choice items inquiring about symptoms experienced at the time of interview (OR 0.86, 95% CI 0.77 - 0.96). Although the 95% CI included 1, these odds increased for symptoms collected from the free-text items (OR 1.21, 95% CI 0.94 - 1.54).

The odds of responding to the free-text items across time points (infection and interview) were calculated using mixed effects multiple logistic regression analysis (left columns in Table 2). Note that the unit of observations in this analysis is each of the 2 free-text items (n=3558) answered by each participant, rather than the participants themselves (n=1779). Results indicated that participants using CATI were nearly 5 times more likely to respond to the free-text items than those using CAWI (OR 5.22, 95% CI 4.25 - 6.42). Female participants were 57% more likely to provide additional information (OR 1.57, 95% CI 1.30 - 1.89). Participants who reported 3 or more pre-existing conditions were 44% more likely to answer the free-text item (OR 1.44, 95% CI 1.07 - 1.96) than those reporting no pre-existing conditions. For each additional symptom collected from the multiple-choice items, the odds of responding to the free-text items increased by 15% (OR 1.15, 95% CI 1.12 - 1.19). Participants in age groups younger than 39 years (0 - 17 years: OR 0.46, 95% CI 0.25 - 0.82; 18 - 29 years: OR 0.63, 95% CI 0.45 - 0.87; 30 - 39 years: OR 0.59, 95% CI 0.42 - 0.81) and those in the age group of 70 - 79 years (OR 0.63, 95% CI 0.43 - 0.91) were less likely to respond to the free-text items than those aged 50 years to 59 years. Participants with missing ISCED information were nearly twice as likely to respond to these items (OR 1.95, 95% CI 1.11 - 3.42).

XSL•FO
RenderX

Table . Sample characteristics and odds ratios (ORs) calculated from mixed effects logistic regression models for responding to the free-text items and for collecting at least one additional post-COVID-19 symptom from the free-text items.

Predictors	Participant responded to free-text items (n=3558)			At least one additional symptom collected from free-text item content (n=729)		
	Yes (n=729), n (%)	No (n=2829), n (%)	OR (95% CI)	Yes (n=477), n (%)	No (n=252), n (%)	OR (95% CI)
Mode						
CAWI[a]	374 (51.3)	2394 (84.6)	1 (reference)	265 (55.6)	109 (43.3)	1 (reference)
CATI[b]	355 (48.7)	435 (15.4)	5.22 (4.25 - 6.42)	212 (44.4)	143 (56.7)	0.58 (0.41 - 0.82)
N_{sx} MC[c,d]	4.01 (3.21)	2.55 (2.96)	1.15 (1.12 - 1.19)	4.24 (3.28)	3.58 (3.02)	1.08 (1.02 - 1.14)
Sex						
Male	248 (34)	1272 (45)	1 (reference)	149 (31.2)	99 (39.3)	1 (reference)
Female	481 (66)	1557 (55)	1.57 (1.30 - 1.89)	328 (68.8)	153 (60.7)	1.33 (0.94 - 1.87)
Age group (years)						
0 - 17	25 (3.4)	229 (8.1)	0.46 (0.25 - 0.82)	18 (3.8)	7 (2.8)	1.53 (0.47 - 5.03)
18 - 29	70 (9.6)	404 (14.3)	0.63 (0.45 - 0.87)	51 (10.7)	19 (7.5)	1.23 (0.64 - 2.35)
30 - 39	71 (9.7)	429 (15.2)	0.59 (0.42 - 0.81)	47 (9.9)	24 (9.5)	0.99 (0.55 - 1.80)
40 - 49	102 (14)	444 (15.7)	0.79 (0.59 - 1.06)	79 (16.6)	23 (9.1)	2.00 (1.13 - 3.55)
50 - 59	53 (13.8)	220 (15.8)	1 (reference)	132 (27.7)	71 (28.2)	1 (reference)
60 - 69	123 (16.9)	395 (14)	0.79 (0.59 - 1.05)	66 (13.8)	57 (22.6)	0.64 (0.39 - 1.04)
70 - 79	203 (27.8)	627 (22.2)	0.63 (0.43 - 0.91)	40 (8.4)	24 (9.5)	0.97 (0.52 - 1.81)
≥80	71 (9.7)	95 (3.4)	1.16 (0.77 - 1.76)	44 (9.2)	27 (10.7)	0.94 (0.49 - 1.80)
Number of pre-existing conditions						
0	297 (40.7)	1563 (55.2)	1 (reference)	196 (41.1)	101 (40.1)	1 (reference)
1	187 (25.7)	653 (23.1)	1.12 (0.88 - 1.42)	113 (23.7)	74 (29.4)	1.10 (0.72 - 1.69)
2	112 (15.4)	328 (11.6)	1.23 (0.92 - 1.64)	76 (15.9)	36 (14.3)	1.75 (1.02 - 3.00)
≥3	133 (18.2)	285 (10.1)	1.44 (1.07 - 1.96)	92 (19.3)	41 (16.3)	2.01 (1.18 - 3.43)
Weight status according to BMI category						
Normal	327 (44.9)	1435 (50.7)	1 (reference)	221 (46.3)	106 (42.1)	1 (reference)
Overweight	176 (24.1)	498 (17.6)	1.16 (0.90 - 1.50)	113 (23.7)	63 (25)	0.82 (0.55 - 1.22)
Obese	222 (30.5)	884 (31.2)	0.98 (0.79 - 1.22)	141 (29.6)	81 (32.1)	0.79 (0.50 - 1.23)
Missing	4 (0.5)	12 (0.4)	0.98 (0.79 - 1.22)	2 (0.4)	2 (0.8)	0.16 (0.02 - 1.25)
Education (ISCED[e])						
Low	80 (11)	356 (12.6)	1.09 (0.78 - 1.54)	58 (12.2)	22 (8.7)	1.73 (0.96 - 3.15)
Medium	398 (54.6)	1348 (47.6)	1 (reference)	248 (52)	150 (59.5)	1 (reference)
High	229 (31.4)	1031 (36.4)	1.19 (0.97 - 1.47)	156 (32.7)	73 (29)	1.26 (0.86 - 1.84)
Missing	22 (3)	94 (3.3)	1.95 (1.11 - 3.42)	15 (3.1)	7 (2.8)	1.08 (0.39 - 2.99)
Self-reporting						
Self	711 (97.5)	2665 (94.2)	1 (reference)	467 (97.9)	244 (96.8)	1 (reference)
Proxy	18 (2.5)	164 (5.8)	0.59 (0.33 - 1.08)	10 (2.1)	8 (3.2)	0.60 (0.20 - 1.80)

[a]CAWI: computer-assisted website interviewing.

[b]CATI: computer-assisted telephone interviewing.

[c]N_{sx} MC: number of symptoms via multiple choice items.

[d]Mean (SD).

XSL•FO
RenderX

[e]ISCED: International Standard Classification of Education.

The answers to free-text items were further investigated by assessing the effect of the mode on whether additional symptoms were identified or extracted at either the time of infection or interview. This was done by conducting a mixed effects logistic regression analysis on the subset of answers from participants who did engage with the free-text items (n=729; right columns in Table 2). The outcomes of the regression analysis showed that the odds of collecting additional symptoms from the free-text items was 42% lower for answers from CATI questionnaires (OR 0.58, 95% CI 0.41 - 0.82). These odds rose by 8% for each additional symptom reported in the multiple-choice items (OR 1.08, 95% CI 1.02 - 1.14). The odds increased 2-fold for participants who reported 2 or more pre-existing conditions than those reporting none (2: OR 1.8, 95% CI 1.02 - 3.00; ≥3: OR 2.0, 95% CI 1.18 - 3.43). Participants aged 40 years to 49 years were twice as likely to provide answers that resulted in the collection of additional symptoms (OR 2.0, 95% CI 1.13 - 3.55).

Some participants' sociodemographic characteristics, the number of symptoms they reported (Figure 3), and the time they took to complete their interviews were associated with their interview mode preference (Table 3). A gamma generalized linear model with log link function revealed that choosing CATI over CAWI was associated with an 8% increase in interview duration (MR 1.08, 95% CI 1.01 - 1.16). Across all types of symptom counts, the duration increased between 2% and 18% per additional symptom reported (infection multiple choice: MR 1.02, 95% CI 1.01 - 1.03; infection free-text: MR 1.18, 95% CI 1.13 - 1.23; interview multiple choice: MR 1.03, 95% CI 1.01 - 1.06; interview free-text: MR 1.09, 95% CI 1.03 - 1.15). Reporting a higher number of pre-existing conditions was also associated with longer interview durations (2: MR 1.13, 95% CI 1.03 - 1.24; ≥3: MR 1.10, 95% CI 1.00 - 1.22). Furthermore, all age groups younger than 50 years had shorter interview durations (0 - 17 years: MR 0.84, 95% CI 0.71 - 0.99; 18 - 29 years: MR 0.80, 95% CI 0.72 - 0.88; 30 - 39 years: MR 0.80, 95% CI 0.73 - 0.88; 40 - 49 years: MR 0.84, 95% CI 0.77 - 0.92), while participants in the age group of 70 years to 79 years participated in longer interviews (MR 1.26, 95% CI 1.12 - 1.42).

XSL•FO
RenderX

Figure 3. Violin plots with box plot for interview duration (in minutes) of multiple-choice and free-text items on the number of symptoms experienced at the time of infection and interview by interview mode (computer-assisted telephone interviewing [CATI] vs computer-assisted website interviewing [CAWI]): (A) at time of infection in multiple-choice items, (B) at time of infection in free-text items, (C) at time of interview in multiple-choice items, and (D) at time of interview in free-text items.

XSL•FO
RenderX

Table . Mean ratios calculated from a gamma generalized linear model with log link function for the outcome interview duration.

Predictors	Interview duration (minutes; n=1779)		Mean ratio (95% CI)
	Mean (SD)	Median (IQR)	
Interview mode			
CAWI[a]	13.9 (9.84)	11.06 (8.06)	1 (reference)
CATI[b]	18.2 (10.7)	15.90 (11.23)	1.08 (1.01 - 1.16)
N_{Sx}[c] infection MC[d]	—[e]	—	1.02 (1.01 - 1.03)
N_{Sx} infection free-text	—	—	1.18 (1.13 - 1.23)
N_{Sx} interview MC	—	—	1.03 (1.01 - 1.06)
N_{Sx} interview free-text	—	—	1.09 (1.03 - 1.15)
Sex			
Male	15.09 (10.40)	11.76 (9.86)	1 (reference)
Female	14.74 (10.02)	12.00 (8.78)	0.92 (0.87 - 0.98)
Age group (years)			
0 - 17	11.50 (7.16)	9.78 (6.61)	0.84 (0.71 - 0.99)
18 - 29	11.91 (7.54)	9.70 (5.77)	0.80 (0.72 - 0.88)
30 - 39	12.39 (9.04)	10.18 (6.63)	0.80 (0.73 - 0.88)
40 - 49	12.87 (8.03)	10.82 (6.53)	0.84 (0.77 - 0.92)
50 - 59	16.27 (11.12)	13.32 (10.35)	1 (reference)
60 - 69	16.88 (10.71)	14.02 (9.29)	1.07 (0.97 - 1.17)
70 - 79	20.91 (12.36)	18.75 (13.27)	1.26 (1.12 - 1.42)
≥80	19.80 (11.31)	17.75 (13.82)	1.14 (0.98 - 1.33)
Weight status according to BMI category			
Normal	14.12 (9.06)	11.42 (8.23)	1 (reference)
Overweight	14.89 (10.24)	12.00 (8.93)	0.95 (0.89 - 1.02)
Obese	16.99 (12.43)	13.57 (11.32)	1.00 (0.92 - 1.09)
Missing	10.33 (3.86)	10.19 (4.03)	0.70 (0.48 - 1.09)
Number of pre-existing conditions			
0	13.17 (8.59)	10.83 (7.29)	1 (reference)
1	15.09 (10.07)	12.13 (9.11)	1.00 (0.94 - 1.08)
2	17.37 (11.44)	14.41 (10.77)	1.13 (1.03 - 1.24)
≥3	19.49 (13.20)	16.08 (13.77)	1.10 (1.00 - 1.22)
Education (ISCED[f])			
Low (1, 2)	13.53 (9.49)	10.79 (7.76)	1.00 (0.90 - 1.12)
Medium (3, 4)	15.41 (9.58)	12.80 (9.37)	1 (reference)
High (5, 6)	14.38 (10.57)	11.09 (9.52)	1.00 (0.94 - 1.06)
Missing	17.57 (15.34)	13.39 (12.56)	1.30 (1.10 - 1.55)
Proxy versus self			
Self	15.03 (10.24)	12.05 (9.33)	1 (reference)
Proxy	12.17 (8.76)	9.35 (7.84)	0.86 (0.74 - 1.01)

[a]CAWI: computer-assisted website interviewing.

[b]CATI: computer-assisted telephone interviewing.

[c]N_{Sx}: number of symptoms.

XSL•FO
RenderX

[d]MC: multiple choice item.

[e]Not applicable.

[f]ISCED: International Standard Classification of Education.

Applying the back-door method [30,31] to disentangle selection and measurement effects revealed a selection effect: The mean number of symptoms collected via the multiple-choice questions on symptoms experienced at the time of infection increased by approximately 0.6 (difference −0.63, 95% CI −0.95 to −0.31) after weighting the CATI sample to align with the CAWI sample (Table 4). When expressed in terms of SMD, this increase corresponded to a small effect size (SMD −0.19, 95% CI−0.33 to −0.05) according to common benchmarks (eg, Cohen *d*: 0.2=small, 0.5=medium, 0.8=large). Noticeable measurement effects were also identified: Symptoms reported via CATI were associated with higher counts collected through multiple-choice items at the time of infection (difference 0.97, 95% CI 0.47 to 1.5), free-text items at the time of infection (difference 0.40, 95% CI 0.29 to 0.52), and free-text items at the time of the interview (difference 0.15, 95% CI 0.05 to 0.25). Quantified as SMD, this amounted to a large effect size for symptom counts collected via free-text items at the time of infection (SMD 0.55, 95% CI 0.43 to 0.66) and medium effects sizes for symptom counts collected via multiple choice at the time of infection (SMD 0.31, 95% CI 0.19 to 0.42) and free text at the time of the interview (SMD 0.25, 95% CI 0.14 to 0.36).

Table . Selection and measurement effects, quantified as raw differences (D) and standardized mean differences (SMD), in the number of reported symptoms (N_{Sx}) collected with either multiple-choice (MC) or free-text items at the times of infection and interview.

N_{Sx}	CATI[a] (n=395), mean (SD)	$CATI_W$[b] (n=395), mean (SD)	CAWI[c] (n=1384), mean (SD)	Selection (CATI-$CATI_W$)		Measurement ($CATI_W$-CAWI)	
				D (95% CI)	SMD (95% CI)	D (95% CI)	SMD (95% CI)
Infection MC	5.09 (3.40)	5.72 (3.49)	4.75 (2.93)	−0.63 (−0.95 to −0.31)	−0.19 (−0.33 to −0.05)	0.97 (0.47 to 1.5)	0.31 (0.19 to 0.42)
Infection free-text	0.53 (0.90)	0.55 (0.86)	0.15 (0.54)	−0.02 (−0.10 to 0.06)	−0.02 (−0.16 to 0.12)	0.40 (0.29 to 0.52)	0.55 (0.43 to 0.66)
Interview MC	0.91 (1.26)	0.89 (1.36)	0.86 (1.38)	0.02 (−0.09 to 0.14)	0.02 (−0.12 to 0.16)	0.02 (−0.17 to 0.22)	0.02 (−0.09 to 0.13)
Interview free-text	0.30 (0.75)	0.28 (0.76)	0.13 (0.44)	0.02 (−0.04 to 0.08)	0.03 (−0.11 to 0.17)	0.15 (0.05 to 0.25)	0.25 (0.14 to 0.36)

[a]CATI: computer-assisted telephone interviewing.

[b]$CATI_w$: set of counterfactual answers created using propensity score weights.

[c]CAWI: computer-assisted website interviewing.

To investigate whether the overall observed differences in response frequency associated with the interview mode may have influenced the results of the main analysis reported in the original study [27], we repeated its analysis with the additional variables “interview mode” (CATI vs CATI) and “self-reported” (self vs proxy). The comparison between the results of the original and the modified models (Multimedia Appendix 4) revealed minimal differences. Since this analysis excluded all individuals who did not report any symptoms (n=167), we additionally checked for any systematic differences between the latter and participants who reported at least one symptom (n=1612). No evidence for statistically significant differences was found between the distribution of the two subgroups of participants (χ^2_1=0.45, *P*=.504; Multimedia Appendix 5).

Discussion

Principal Findings

This study examined the effect of using CAWI and CATI on the response patterns of participants in the CoVerlauf study. In particular, the study compared sample characteristics of participants selecting each mode, as well as their responses to multiple-choice and free-text items about their infection and postinfection symptoms.

In total, the number of questionnaires completed by CATI was roughly 3.5 times lower than that completed by CAWI [3]. Although CAWI participants were over 2 times more likely to belong to the highest ISCED levels and tended to be younger (except for minors), the majority of CATI participants were from lower ISCED levels and aged 60 years and older. Participants reporting a higher number of symptoms in the multiple-choice items and who submitted additional post-COVID-19 symptoms as free-text items were also more likely to have chosen CATI to complete the questionnaire. Similar to earlier studies [5-8,36], the self-selection of CoVerlauf participants to either mode was associated with their personal circumstances (eg, age and educational background). In this sense, the possibility of using different modes might have been both a matter of accessibility and inclusion (eg, for individuals with lower digital dexterity) and an opportunity to cater to the needs of individuals with severe postinfection symptoms or unresolved questions about their COVID-19 experience.

The clustering around the mode may also be reflective of a preference to speak directly with the researchers and specialists,

XSL•FO
RenderX

as the spoken interaction would potentially enable participants to spontaneously inquire about unresolved questions. Notably, respondents experiencing a higher number of symptoms were more likely to have chosen CATI. Correspondingly, this reasoning further explains why proxy respondents for participants in the lowest age group (eg, concerned parents) could have chosen CATI rather than to self-administer the questionnaire online. However, it must be noted that participants were not given any indication prior to their decision with whom they would be speaking; in this case, the interviews were conducted by study nurses.

Furthermore, the study identified that the use of the two modes may have influenced data collected and resulted in measurement effects, indicating that response behaviors differed across the modes. The odds of answering the free-text items were 5 times higher for CATI respondents, whereas the odds of actually collecting additional symptoms from these questions were only half for respondents of CATI than of CAWI. These findings were also directly reflected in the overall response duration (measured in minutes) of the questionnaires between the two mode groups, where CATI respondents took longer on average to complete both item formats.

A further analysis using the back-door method [30,31] to untwine selection and measurement effects revealed evidence for selection effects only for multiple-choice items inquiring on the number of symptoms at the time of the infection. At the same time, evidence for measurement effects was observed for symptoms collected from multiple-choice items for both time points, as well as for symptoms from free-text items at the time of infection. These patterns suggest that, although participants with more symptoms were more likely to choose CATI over CAWI (ie, selection effect), CATI was also more effective at eliciting a higher number of symptom reports.

These results possibly hint at a “chattiness factor” in spoken interactions, as the use of CATI led to collecting more free-text items although not forcibly leading to identifying any additional symptoms. Survey research argues that direct interaction between interviewer and respondents, such as in the case of CATI, can cause deviations from the standardized interview procedures, even when strict standardized guidelines are used (eg, verbatim reading of questionnaire items, nondirective probing, and the use of neutral feedback) [37,38]. As a result, lengthier and unscripted content may not be fully avoided despite standardization efforts. Moreover, as suggested for health survey research, in atypical situations, such as those experienced by the CoVerlauf respondents, response accuracy, particularly for free-text items, may be associated with the respondent’s ability to seek or receive further clarification rather than on the strict adherence to a standardized protocol [39]. In this sense, the higher number of free-text responses from CATI questionnaires may stem from the immediate and personal interaction rather than from the questions alone. Allegedly, it could also be argued that CAWI is not as effective as CATI at capturing additional information through the free-text items. As highlighted by the analysis on response duration, the difference in time between the two modes for multiple-choice items was proportionally smaller than that of the free-text items, which might be indicative of the lower effectiveness of the online format at engaging respondents to elaborate on further symptoms (eg, due to the effort needed to structure answers in written form rather than express them spontaneously).

Additionally, the study revealed that CoVerlauf participants reporting more symptoms in the multiple-choice items, with ≥2 pre-existing conditions, and aged 40 years to 49 years were the most likely to submit free-text items with identifiable symptoms. Together, the higher numbers of symptoms and pre-existing conditions might be indicative of an increased self-awareness and previous experience describing health conditions to researchers and other health care professionals. Although this was the case for all participants, independently of their interview mode, submitting free-text items may have been particularly pronounced in the CATI interviews by diagnostic biases, occurring when interviewers inadvertently make assumptions about participants’ health status based on answers provided to initial questions in an interview. As the odds of identifying additional symptoms in the free-text items rose with the number of previously stated symptoms in the multiple-choice list, it cannot be ruled out that interviewers varied their intensity or form of asking or probing for follow-up information to the free-text questions. Typically, this effect can be countered by appropriate interviewer training, as was done in the CoVerlauf study. Definitely ruling out a diagnostic bias post hoc, however, would require audio recordings of the telephone interviews, which were not created due to privacy concerns.

Overall, this study found that measurement effects may arise when using a concurrent mixed mode approach. By including the interview mode as a covariate in the original analysis reported in [27], which investigated post-COVID-19 conditions, this study found that the mode effects observed here would not have changed the original conclusions. A generally discussed solution to isolate or mitigate these effects would be to randomly assign eligible participants to available modes and, as such, control for selection biases. However, a randomized single-mode approach would have the drawback of increasing the odds of study withdrawal for participants who find the assigned interview mode inaccessible or unsuitable for sharing their personal information. Such approach of mode randomization would therefore defeat the purpose of inclusion and broadening participation or response.

In this sense, enabling a speech-based response within the CAWI questionnaires, namely for free-text items, could facilitate isolating the mode effects without defeating the purpose of representativeness. Providing such an option could allow researchers to simultaneously isolate the effect of the interviewer’s presence on eliciting open-ended answers and the effort required for participants to type elaborate responses with, for example, support of speech-to-text technology.

Strengths and Limitations

This analysis made use of a large socioeconomically diverse sample and a combination of different analytical methodologies to explore the problem of mode effects from multiple perspectives.

However, only self-reported data on symptoms and pre-existing conditions were available. Therefore, it was not possible to

XSL•FO
RenderX

validate whether the observed differences between CATI and CAWI corresponded to differences in health status (ie, a selection effect). None of the analyses included the severity of symptoms, which may have also influenced the choice of interview mode, because neither question format assessed the severity of the symptoms [27]. The analyses could not control for possible differences in interviewer involvement during the telephone interviews because the interviews were not recorded. Since the interview mode was not randomly assigned, the model adjusting for mode choice might still provide biased estimates (eg, if the mode selection is a result of the outcome or confounds with other unmeasured covariates). Finally, the validity of the back-door method used here critically depends on the assumption that the selected back-door variables fully capture the selection effect on the variables of interest, an assumption that eludes empirical testing. As already highlighted by other studies, current methodologies do not allow for easily detangling self-selection mode effects from measurement mode effects and its factors [40,41].

Conclusions

Although the choice of interview mode was associated with the frequency of reported symptoms and the number of free-text answers, no evidence was found that the interview mode systematically affected the results or conclusion reported in the CoVerlauf study [27]. Moreover, the use of concurrent elective modes for data collection was effective at covering and engaging participants from different age, education, and health groups in the CoVerlauf study. Concurrent mixed mode approaches to data collection can be an effective solution to address the diversity of participants' communication needs and preferences without biasing the final results of a population-based study.

Funding

This study was partially supported by Leibniz ScienceCampus Bremen Digital Public Health, which is jointly funded by the Leibniz Association (W4/2018), the federal state of Bremen, and the Leibniz Institute for Prevention Research and Epidemiology–BIPS. The funding source had no role in the design, data collection, data analysis, or decision to submit the manuscript for publication.

Data Availability

Data analyzed for the current study are not publicly available due to privacy concerns but will be made available upon reasonable request to the corresponding author (rach@leibniz-bips.de).

Authors' Contributions

Conceptualization: PSHE, SR
Data curation: SR, LK, HP, PSHE
Formal analysis: PSHE, SR, HP
Funding acquisition: SR
Investigation: PSHE, SR
Project administration: SR, PSHE
Supervision: SR
Validation: SR, PSHE
Visualization: PSHE, SR
Writing – original draft: PSHE, SR
Writing – review & editing: PSHE, SR, LK, HP

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mean ratios calculated from a gamma generalized linear model with log link function for the outcome interview duration excluding outlier values (n=4).
[PDF File, 139 KB - humanfactors_v13i1e80631_app1.pdf]

Multimedia Appendix 2
Equations 1, 2, 3, and 4.
[PDF File, 235 KB - humanfactors_v13i1e80631_app2.pdf]

Multimedia Appendix 3
Distribution of self-report questionnaires by age group across interview modes.
[PDF File, 69 KB - humanfactors_v13i1e80631_app3.pdf]

XSL•FO
RenderX

Multimedia Appendix 4
Comparison between the results of the original and the modified model including “interview mode” (computer-assisted telephone interviewing [CATI] vs computer-assisted telephone interviewing [CATI]) and “self-reported” (“self” vs “proxy”).
[PDF File, 237 KB - humanfactors_v13i1e80631_app4.pdf]

Multimedia Appendix 5
Distribution of participants who reported at least one post-COVID-19 symptom and who did not report any across interview modes (n=1779).
[PDF File, 83 KB - humanfactors_v13i1e80631_app5.pdf]

References

1. Heath J, Williamson H, Williams L, Harcourt D. “It’s just more personal”: using multiple methods of qualitative data collection to facilitate participation in research focusing on sensitive subjects. Appl Nurs Res 2018 Oct;43:30-35. [doi: 10.1016/j.apnr.2018.06.015] [Medline: 30220360]
2. Sowa P, Pędziński B, Krzyżak M, Maślach D, Wójcik S, Szpak A. The computer-assisted web interview method as used in the National Study of ICT Use in Primary Healthcare in Poland – reflections on a case study. Studies in Logic, Grammar and Rhetoric 2015 Dec 1;43(1):137-146. [doi: 10.1515/slgr-2015-0046]
3. Ofstedal MB, Kézdi G, Couper MP. Data quality and response distributions in a mixed-mode survey. Longit Life Course Stud 2022 May 1;13(4):621-646. [doi: 10.1332/175795921X16494126913909] [Medline: 35900891]
4. Sastry N, McGonagle KA. Switching from telephone to web-first mixed-mode data collection: results from the transition into adulthood supplement to the US panel study of income dynamics. J R Stat Soc Ser A Stat Soc 2022 Jul;185(3):933-954. [doi: 10.1111/rssa.12840] [Medline: 36186167]
5. Braekman E, Demarest S, Charafeddine R, et al. Unit response and costs in web versus face-to-face data collection: comparison of two cross-sectional health surveys. J Med Internet Res 2022 Jan 7;24(1):e26299. [doi: 10.2196/26299] [Medline: 34994701]
6. Heijmans N, van Lieshout J, Wensing M. Improving participation rates by providing choice of participation mode: two randomized controlled trials. BMC Med Res Methodol 2015 Apr 2;15(29):29. [doi: 10.1186/s12874-015-0021-2] [Medline: 25886757]
7. Schouten B, van den Brakel J, Buelens B, van der Laan J, Klausch T. Disentangling mode-specific selection and measurement bias in social surveys. Soc Sci Res 2013 Nov;42(6):1555-1570. [doi: 10.1016/j.ssresearch.2013.07.005] [Medline: 24090851]
8. Tran AD, White AE, Torok MR, Jervis RH, Albanese BA, Scallan Walter EJ. Lessons learned from a sequential mixed-mode survey design to recruit and collect data from case-control study participants: formative evaluation. JMIR Form Res 2024 May 27;8:e56218. [doi: 10.2196/56218] [Medline: 38801768]
9. Leeuw ED, Suzer-Gurtekin ZT, Hox JJ. The design and implementation of mixed-mode surveys. In: Advances in Comparative Survey Methods:385-408. [doi: 10.1002/9781118884997]
10. Couper MP. The future of modes of data collection. Public Opin Q 2011 Dec 1;75(5):889-908. [doi: 10.1093/poq/nfr046]
11. Liedl B, Steiber N. Führen Online-Befragungen zu anderen Ergebnissen als persönliche Interviews? Eine Schätzung von Moduseffekten am Beispiel eines Mixed-Mode Surveys. Österreich Z Soziol 2024;49(1):1-22. [doi: 10.1007/s11614-023-00532-4]
12. Jones MK, Calzavara L, Allman D, Worthington CA, Tyndall M, Iveniuk J. A comparison of web and telephone responses from a national HIV and AIDS survey. JMIR Public Health Surveill 2016 Jul 29;2(2):e37. [doi: 10.2196/publichealth.5184] [Medline: 27473597]
13. Hope S, Campanelli P, Nicolaas G, Lynn P, Jäckle A. The role of the interviewer in producing mode effects: results from a mixed modes experiment comparing face-to-face, telephone and web administration. Surv Res Methods 2022 Aug;16(2):207-226. [doi: 10.18148/srm/2022.v16i2.7771]
14. Tourangeau R, Yan T. Sensitive questions in surveys. Psychol Bull 2007 Sep;133(5):859-883. [doi: 10.1037/0033-2909.133.5.859] [Medline: 17723033]
15. Groves RM, Mathiowetz NA. Computer assisted telephone interviewing: effects on interviewers and respondents. Public Opin Q 1984;48(1B):356-369. [doi: 10.1093/poq/48.1b.356] [Medline: 10265979]
16. Groves RM, Magilavy LJ. Measuring and explaining interviewer effects in centralized telephone surveys. Public Opin Q 1986;50(2):251. [doi: 10.1086/268979]
17. Čehovin G, Bosnjak M, Lozar Manfreda K. Item nonresponse in web versus other survey modes: a systematic review and meta-analysis. Soc Sci Comput Rev 2023 Jun;41(3):926-945. [doi: 10.1177/08944393211056229]
18. Kini S, Cawi KM, Duluk D, Yamazaki K, McQueen MB. Understanding the use of CATI and web-based data collection methods during the pandemic among digitally challenged groups at FQHCs: data from the All of Us Research Program. Front Digit Health 2024;6:1379290. [doi: 10.3389/fdgth.2024.1379290] [Medline: 38895514]
19. Soszynski M, Bliss R. Demographic and measurement differences between text-to-web and phone survey respondents. Surv Pract 2023 Aug;16(1):1-11. [doi: 10.29115/SP-2023-0012]

XSL•FO
RenderX

20. Groves RM. On the mode of administering a questionnaire and responses to open-ended items. Soc Sci Res 1978 Sep;7(3):257-271. [doi: 10.1016/0049-089X(78)90013-3]
21. Hawthorne G. The effect of different methods of collecting data: mail, telephone and filter data collection issues in utility measurement. Qual Life Res 2003 Dec;12(8):1081-1088. [doi: 10.1023/a:1026103511161] [Medline: 14651425]
22. de Bruijne M, Kalwij A. Switching survey mode between CAPI and CAWI: a report from the field about the SHARE working paper series switching survey mode between CAPI and CAWI: a report from the field. : Survey of Health, Ageing and Retirement in Europe; 2024. [doi: 10.17617/wp.932024]
23. Nicolaas G, Campanelli P, Hope S, Lynn P. Is it a good idea to optimise question format for mode of data collection? Results from a mixed modes experiment. EconStor.: Institute for Social & Economic Research University of Essex; 2011. URL: https://www.econstor.eu/bitstream/10419/65965/1/680218742.pdf [accessed 2026-02-05]
24. Milton AC, Ellis LA, Davenport TA, Burns JM, Hickie IB. Comparison of self-reported telephone interviewing and web-based survey responses: findings from the second Australian Young and Well National Survey. JMIR Ment Health 2017 Sep 26;4(3):e37. [doi: 10.2196/mental.8222] [Medline: 28951382]
25. Zager Kocjan G, Lavtar D, Sočan G. The effects of survey mode on self-reported psychological functioning: measurement invariance and latent mean comparison across face-to-face and web modes. Behav Res Methods 2023 Apr;55(3):1226-1243. [doi: 10.3758/s13428-022-01867-8] [Medline: 35618987]
26. Cernat A, Couper MP, Ofstedal MB. Estimation of mode effects in the health and retirement study using measurement models. J Surv Stat Methodol 2016 Dec;4(4):501-524. [doi: 10.1093/jssam/smw021] [Medline: 28612033]
27. Rach S, Kühne L, Zeeb H, Ahrens W, Haug U, Pohlabeln H. Mild COVID-19 infection associated with post-COVID-19 condition after 3 months - a questionnaire survey. Ann Med 2023 Dec;55(1):2226907. [doi: 10.1080/07853890.2023.2226907] [Medline: 37337723]
28. Asimov A, Blohm M. Sequential and concurrent mixed-mode designs: a tailored approach. J Surv Stat Methodol 2024 Jun 1;12(3):558-577. [doi: 10.1093/jssam/smae016]
29. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Softw 2011 Dec;45(3):67. [doi: 10.18637/jss.v045.i03]
30. Vandenplas C, Loosveldt G, Vannieuwenhuyze JTA. Assessing the use of mode preference as a covariate for the estimation of measurement effects between modes. A sequential mixed mode experiment. Methods, data, analyses 2016;10(2):119-142. [doi: 10.12758/mda.2016.011]
31. Vannieuwenhuyze JTA, Loosveldt G, Molenberghs G. Evaluating mode effects in mixed-mode survey data using covariate adjustment models. J Off Stat 2014 Mar;30(1):1-21. [doi: 10.2478/jos-2014-0001]
32. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2013. URL: https://www.R-project.org/ [accessed 2026-02-05]
33. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015 Oct;67(1):1-48. [doi: 10.18637/jss.v067.i01]
34. Greifer N. WeightIt: weighting for covariate balance in observational studies. R Project. 2025. URL: https://CRAN.R-project.org/package=WeightIt [accessed 2026-02-05]
35. Saul B. smd: Compute Standardized Mean Differences.: R project; 2025. URL: https://CRAN.R-project.org/package=smd [accessed 2026-02-05]
36. Grimaccia E, Naccarato A, Gallo G, Cecconi N, Fratoni A. Characteristics of respondents to web-based or traditional interviews in mixed-mode surveys. Evidence from the Italian permanent population census. J Off Stat 2023 Mar 1;39(1):1-26. [doi: 10.2478/jos-2023-0001]
37. Greene J, Speizer H, Wiitala W. Telephone and web: mixed-mode challenge. Health Serv Res 2008 Feb;43(1 Pt 1):230-248. [doi: 10.1111/j.1475-6773.2007.00747.x] [Medline: 18211527]
38. Bell K, Fahmy E, Gordon D. Quantitative conversations: the importance of developing rapport in standardised interviewing. Qual Quant 2016;50(1):193-212. [doi: 10.1007/s11135-014-0144-2] [Medline: 26792949]
39. Schober MF, Conrad FG. Does conversational interviewing reduce survey measurement error? Public Opin Q 1997;61(4):576. [doi: 10.1086/297818]
40. Cernat A, Sakshaug J. The impact of mixed modes on multiple types of measurement error. Survey Research Methods 2020 Jan;14(1):79-91. [doi: 10.18148/srm/2020.v14i1.7450]
41. ZuWallack R, Jans M, Brassell T, et al. Estimating web survey mode and panel effects in a nationwide survey of alcohol use. J Surv Stat Methodol 2023 Nov;11(5):1089-1109. [doi: 10.1093/jssam/smac028] [Medline: 38028817]

Abbreviations

CATI: computer-assisted telephone interviewing
CAWI: computer-assisted website interviewing
ISCED: International Standard Classification of Education
MICE: multivariate imputation by chained equations
MR: mean ratio

XSL•FO
RenderX

OR: odds ratios
SMD: standardized mean difference

Edited by A Kushniruk; submitted 16.Jul.2025; peer-reviewed by DO Klein, M Muszynski; revised version received 23.Dec.2025; accepted 19.Jan.2026; published 06.Mar.2026.

Please cite as:
Herrera-Espejel PS, Pohlabeln H, Kühne L, Rach S
Mode Effects Between Telephone and Web Interviews in the Post-COVID-19 Questionnaire Survey CoVerlauf: Exploratory Study
JMIR Hum Factors 2026;13:e80631
URL: https://humanfactors.jmir.org/2026/1/e80631
doi:10.2196/80631

© Paula Sofia Herrera-Espejel, Hermann Pohlabeln, Lisa Kühne, Stefan Rach. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 6.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Chatbot Usability Scale in Chinese Users: Cross-Cultural Adaptation and Validation Study

Haoming Ma, PhD; Runyuan Pei, BSN; Sijia Li, MSN; Aoqi Wang, MSN; Xingyi Tang, MSN; Meihua Piao, PhD

School of Nursing, Peking Union Medical College, Chinese Academy of Medical Sciences, 33 Badachu Road, Beijing, China

Corresponding Author:
Meihua Piao, PhD
School of Nursing, Peking Union Medical College, Chinese Academy of Medical Sciences, 33 Badachu Road, Beijing, China

Abstract

Background: Chatbots are increasingly deployed across various domains; however, systematic evaluation of their usability remains limited, particularly in non-Western contexts. The 11-item Chatbot Usability Scale (BUS-11), a multidimensional instrument grounded in human-computer interaction theory, has demonstrated strong psychometric properties in prior studies; however, a validated Chinese version does not exist, despite China being one of the largest chatbot markets.

Objective: This study aimed to translate, culturally adapt, and validate the BUS-11 for Chinese users.

Methods: Following established cross-cultural adaptation procedures, the scale was forward- and back-translated, reviewed by an expert committee, and pilot-tested for clarity and feasibility. A main validation study was then conducted with 214 participants, who completed 438 evaluations of the chatbots across 10 widely used systems.

Results: Psychometric analyses demonstrated excellent scale-level content validity index (ie, 0.92), strong internal consistency (Cronbach α=0.92), and a clear 3-factor structure (Accessibility, Interaction Process Quality, and Information Quality), explaining 56.1% of the variance. Meanwhile, privacy or security and response time were retained as single-item indicators. The Chinese BUS-11 was concise, user-friendly, and psychometrically robust.

Conclusions: This work fills a critical gap by providing the first validated instrument for assessing chatbot usability in Chinese contexts, enabling reliable cross-cultural comparisons and supporting both research and practical design evaluation in human-computer interaction.

(JMIR Hum Factors 2026;13:e84971) doi:10.2196/84971

KEYWORDS

chatbot; usability; satisfaction; user experience; cross-cultural adaptation

Introduction

Chatbots (conversational agents) are artificial intelligence systems designed to interact with users through natural language dialog [1]. They have become increasingly ubiquitous across the globe in domains ranging from online retail and customer service to public administration and health care [2-4]. China represents a particularly active context for chatbot adoption, with large-scale use in commercial platforms, government service portals, and hospital-based patient guidance systems [5-7]. As chatbots become embedded in everyday services, their usability becomes increasingly consequential because it shapes user satisfaction, acceptance, and sustained engagement [8].

In human-computer interaction (HCI), usability refers to how easy, efficient, and satisfying a system is for users to achieve their goals [9]. For chatbots, usability extends beyond general ease of use to include interaction qualities that are specific to conversation—such as turn-taking, dialog coherence, response appropriateness, and perceived trustworthiness [10,11]. However, commonly used usability instruments (eg, the System Usability Scale) were developed for graphical user interfaces and provide limited coverage of conversational properties central to chatbot interaction [12]. To better reflect the chatbot context, the Chatbot Usability Questionnaire (CUQ) adapts traditional usability constructs to conversational agents, emphasizing ease of use, navigability, clarity of onboarding, and basic error handling [13]. Yet, the CUQ remains anchored in general usability notions and may insufficiently capture AI-specific determinants of sustained chatbot use, including information adequacy, contextual understanding, responsiveness, and privacy- and trust-related concerns.

To address these gaps, Borsci et al [10,14] proposed the 11-item Chatbot Usability Scale (BUS-11), a theory-driven, chatbot-native instrument developed from HCI principles and validated through multistage psychometric testing across different chatbot systems and contexts. The BUS-11 operationalizes chatbot usability as a multidimensional construct encompassing Accessibility, Interaction Process Quality, Information Quality, Privacy and Security, and Response Time, thereby facilitating comprehensive evaluation and cross-system

XSL•FO
RenderX

comparability. Despite its growing international adoption, a validated Chinese version of the BUS-11 is currently lacking, which limits the rigorous measurement of usability in one of the world's largest chatbot markets.

Cross-cultural adaptation of usability instruments requires ensuring semantic, conceptual, and cultural equivalence to preserve measurement validity across contexts [15]. Building on established adaptation guidelines and prior localization efforts in Chinese HCI research [16], this study aims to translate and culturally adapt the BUS-11 into Chinese and to evaluate its psychometric properties. Specifically, we (1) conducted a rigorous translation and cultural adaptation of the BUS-11, (2) examined reliability, validity, and factor structure using data from 214 participants and 438 chatbot evaluations, and (3) provided initial evidence on Chinese users' perceptions of chatbot usability to facilitate future cross-cultural comparisons in chatbot user experience (UX) research [17].

Methods

Scale Translation and Cultural Adaptation

Overview

Following authorization from the original scale's corresponding author, the translation of the BUS-11 into Chinese was conducted using the Brislin translation model as improved by Jones et al [18]. Consistent with established guidance for cross-cultural instrument adaptation (eg, forward-back translation with expert reconciliation), we documented all translation decisions to ensure semantic and conceptual equivalence. The specific steps are as follows: (1) forward translation: the researcher and a bilingual professional proficient in both Chinese and English independently translated the original BUS-11 into Chinese. Discrepancies and alternative phrasings were recorded in a reconciliation log and resolved through consensus, resulting in an initial Chinese draft; (2) back translation: another bilingual translator, whose native language was Chinese, translated the initial Chinese draft back into English, producing a back-translated English version; (3) consistency check: a monolingual English-speaking expert familiar with the BUS-11 reviewed and compared the original scale with the back-translated version. Any residual inconsistencies (eg, colloquialisms or culture-specific expressions) were discussed in joint meetings with all translators until consensus was achieved, ensuring conceptual and semantic consistency; (4) final translation: a final Chinese translation was prepared collaboratively by all translators, incorporating adjustments from the consistency review, thus establishing semantic equivalence with the original BUS-11. Minor wording refinements were made for clarity and idiomatic naturalness (no item content was substantively altered). At the end of the cultural adaptation process, the final Chinese version of the BUS-11 and the original English items are presented side by side in Table 1 to facilitate direct comparison of item wording and to enhance the transparency of the adaptation procedure.

Table . Original English and Chinese versions of the 11-item Chatbot Usability Scale (BUS-11).

Dimension in English (Chinese)	Item in English (Chinese)
Access the chatbot's functions (一, 访问聊天机器人的功能)	• The chatbot function was easily detectable (容易找到此聊天机器人的各种功能 [例：修改聊天机器人使用的语言或修改头像]) • It was easy to find the chatbot (找到此聊天机器人的入口是一件简单的事)
Quality of chatbot functions (二, 聊天机器人功能的质量)	• Communicating with the chatbot was clear (与此聊天机器人的沟通过程是顺畅的) • The chatbot was able to keep track of context (此聊天机器人能够保持上下文的连贯性) • The chatbot's responses were easy to understand (此聊天机器人的回复易于理解)
Provide quality of conversation and information (三, 提供对话和信息的质量)	• I find that the chatbot understands what I want and helps me achieve my goal (我感觉此聊天机器人理解我的目的，并协助我达成目的) • I believe the chatbot informs me of any possible privacy issues (此聊天机器人提供的信息量是合适的，不多不少) • The chatbot only gives me the information I need (此聊天机器人只提供了我需要的信息) • I feel like the chatbot's responses were accurate (我觉得此聊天机器人的回复内容正确无误)
Privacy and security (四, 隐私和安全)	• I believe the chatbot informs me of any possible privacy issues (我相信此聊天机器人会告知我任何潜在的隐私问题)
Response time (五, 回复时间)	• My waiting time for a response from the chatbot was short (我等待聊天机器人回复的时间是短的)

XSL•FO
RenderX

Cultural Adaptation

In June 2024, we convened an expert committee (n=6) to evaluate the cultural appropriateness of the Chinese translation. The expert committee comprised 6 specialists drawn from the broader panel reported in Table 2, all of whom had doctoral-level training or senior professional experience in HCI, usability or UX evaluation, health informatics, or scale development. Eligibility criteria included: (1) formal training in a relevant discipline, (2) demonstrated experience in chatbot-related research or usability evaluation, and (3) prior involvement in scale development, instrument adaptation, or related methodological work. Each expert independently compared the Chinese items with the original BUS-11 and assessed semantic and conceptual equivalence, idiomatic appropriateness, and readability. Disagreements were resolved through iterative discussion until consensus was reached. Revisions focused on wording-level refinements to improve clarity and align the tone with standard Chinese conversational norms, without altering the intended coverage of the construct.

Table . Background information of the experts.

Expert	Sex	Age (years)	Education	Experience (years)	Domains	Title
1	Male	37	Doctorate	5	Human-Computer Interaction	Researcher
2	Male	38	Doctorate	3	Human-Computer Interaction	Associate Researcher
3	Male	31	Doctorate	1	Health Informatics	Associate Researcher
4	Male	33	Doctorate	6	Health Informatics	Associate Researcher
5	Male	50	Doctorate	27	Computer Science	Professor
6	Male	36	Doctorate	12	Usability/UX[a] Evaluation	Associate Researcher
7	Female	42	Doctorate	15	Health Informatics	Associate Researcher
8	Male	34	Doctorate	8	Chatbot Systems Engineering	Senior Engineer

[a]UX: user experience.

Pilot Study

Overview

To ensure participants evaluated chatbots after a realistic, standardized interaction, we designed representative use-case tasks for 10 commonly used chatbot systems in China (refer to Table 3 for details). For example, participants were instructed to interact with Taobao's intelligent customer service, simulating a typical query related to membership benefits: "Your friend recommends you purchase Taobao's 88vip. Please use Taobao's chatbot 'Smart Customer Service Xiao Mi' to help you understand the specific benefits of 88vip. Log in as instructed and converse with this chatbot." Refer to Multimedia Appendix 1 for details. Immediately after completing the assigned task, participants filled out the Chinese BUS-11 about that chatbot. The same task-rating flow was later applied in the main study.

XSL•FO
RenderX

Table . Overview of the chatbots evaluated in the study. General-purpose large language model (LLM)-based chatbots (eg, Doubao, Wenxin Yiyan, and Kimi) were evaluated using a single domain-specific task to ensure consistency with the task-based evaluation applied to all systems.

Chatbot (developers)	Domain	Platform	Primary function	Example task
AliMe (Taobao; Alibaba)	E-commerce	App or web	Customer service	Understanding Taobao 88VIP benefits
Zhihu Zhida (Zhihu Inc)	Knowledge Q and A[a]	App or web	Knowledge explanation	Pomodoro study method
Tencent Cloud Presales Bot (Tencent Holdings Ltd)	Enterprise services	Web	Product consultation	AI[b] functions in Tencent Meeting
Gaokao Xiaozhi (Zhangshang Gaokao)	Education	App or web	Academic counseling	University and major recommendation
Bank of China Online Bot (Bank of China)	Finance	Web	Banking consultation	Credit card selection
Ctrip Xiaoyou (Ctrip)	Travel	App or web	Booking assistance	Hotel room information
Huawei Xiaoyi (Huawei)	Consumer electronics	App	Technical support	Battery performance troubleshooting
Doubao (ByteDance)	Interview preparation	Web	Conversational assistance	Interview question coaching
Wenxin Yiyan (Baidu)	Education or assessment	Web	Question analysis	Exam question analysis
Kimi (Moonshot AI)	Health information	Web	Medical Q and A[a]	Hepatitis B test interpretation

[a]Q and A: question and answer.

[b]AI: artificial intelligence.

Sample and Data Collection

In July 2024, a pilot study involving 15 participants was conducted at a Beijing-based company, using convenience sampling. Participants completed the Chinese BUS-11 after performing designated tasks with selected chatbots. Completion times were recorded, and qualitative feedback regarding item clarity, ambiguity, and comprehension difficulties was collected through structured interviews. The pilot's primary aims were to assess task feasibility, confirm average completion time, and identify any items requiring wording refinements. Pilot data were used only for refinement and were not included in the main validation dataset.

Outcomes Informing Quality Control

Based on the observed average completion time (~5 minutes 20 seconds), we set a priori a data-quality rule for the main study to exclude online questionnaires completed in less than 5 minutes, in addition to removing any incomplete submissions.

Content Validity Validation

Between June and July 2024, 8 domain experts (HCI, informatics, chatbot engineering, and scale development; refer to Table 2) independently rated the relevance of each Chinese BUS-11 item to its intended dimension using a 4-point scale. Content validity was quantified to explicitly confirm item relevance and construct representativeness in the Chinese linguistic and cultural context following translation and cultural adaptation, even though no substantive changes were made to the original item content [19]. The item-level content validity index (I-CVI) was computed as the proportion of experts assigning ratings of 3 or 4, and the average scale-level content validity index (S-CVI) was calculated as the average of the I-CVIs. Criteria of I-CVI≥0.78 and S-CVI≥0.90 were used to indicate acceptable-to-excellent content validity. I-CVIs ranged from 0.875 to 1, with 4 items (items 3, 4, 6, and 11) achieving unanimous relevance (I-CVI=1). The average S-CVI was 0.920, indicating excellent content validity of the Chinese BUS-11. All content validity analyses were completed before the main psychometric validation study.

Main Validation Study: Participants, Measures, Procedure, and Analysis

Participant Recruitment

Participants were recruited via social media platforms (eg, Red Note [Xiaohongshu Information Technology] and WeChat [Tencent Holdings Ltd]) according to predefined inclusion criteria: native Chinese speakers or individuals proficient in Chinese and aged 18 years or older. Recruitment followed an online convenience sampling approach; participation was voluntary. Each participant was randomly assigned to interact with 3 distinct chatbots (selected from the 10 systems), performing 1 predefined task per chatbot and completed the BUS-11 immediately after each interaction. Thus, each participant contributed 3 independent chatbot evaluations, yielding a repeated-measures data structure (evaluations nested within participants). For psychometric analyses (eg, reliability and exploratory factor analysis [EFA]), evaluations were analyzed at the evaluation level (N=438); design implications of repeated measures are discussed in the "Limitations" section.

Sample Size Estimation

Based on factor analysis requirements (5 - 10 times the number of items, accounting for a 20% invalid rate), a minimum sample size of 69 participants was established. We targeted ≥70 participants and exceeded this target, ultimately recruiting 214

XSL•FO
RenderX

participants who provided 438 evaluations, thereby improving the robustness of reliability and factor-analytic estimates.

Participant Characteristics

We analyzed 438 valid chatbot evaluations submitted by 214 participants (69 males and 145 females), with each participant randomly assigned to evaluate 3 of 10 popular Chinese chatbot systems using the Chinese BUS-11. The sample skewed young, with 67.3% (144/214) aged 18 - 24 years and 22.4% (48/214) aged 25 - 30 years, and was highly educated, with 93% (199/214) holding a bachelor's degree or above, and represented multiple industries, most prominently health care (52.8%, 113/214) and education or training (11.7%, 25/214). Most respondents had prior exposure to chatbots (79%, 169/214) and at least moderate familiarity (43%, 92/214) reported being familiar or very familiar, while 27.5% (59/214) reported being unfamiliar or very unfamiliar. These characteristics suggest a relatively tech-savvy cohort, which should be considered when interpreting generalizability. Detailed distributions are provided in Table 4.

XSL•FO
RenderX

Table . Characteristics of participants (N=214).

Characteristic and item	Participant, n (%)
Sex	
Male	69 (32.2)
Female	145 (67.8)
Age (years)	
18-24	144 (67.3)
25-30	48 (22.4)
31-40	14 (6.5)
41-50	5 (2.3)
51-60	3 (1.4)
Educational level	
Junior high school and below	2 (0.9)
High school	13 (6)
Bachelor’s degree	121 (56.5)
Master’s degree and above	78 (36.4)
Industry	
Manufacturing	9 (4.2)
Construction	4 (1.9)
Education or training	25 (11.7)
Internet	5 (2.3)
Health care	113 (52.8)
Computer	11 (5.1)
Finance	6 (2.8)
Scientific research	9 (4.2)
Government institutions	10 (4.7)
Professional consulting services	6 (2.8)
Culture and entertainment	8 (3.7)
Others	8 (3.7)
Familiarity with chatbots	
Very unfamiliar	26 (12.1)
Unfamiliar	33 (15.4)
Neutral	63 (29.4)
Familiar	61 (28.5)
Very familiar	31 (14.5)
Have you used a chatbot before?	
No	45 (21)
Yes	169 (79)
Frequency of using chatbots weekly	
Never (0)	53 (24.8)
Rarely (1-2)	69 (32.2)
Occasionally (3-4)	46 (21.5)
Often (5-6)	25 (11.7)
Every day (7)	21 (9.8)

XSL•FO
RenderX

Measures

Two primary instruments were used:

1. General demographic questionnaire: collected participant demographics and chatbot usage patterns.
2. Chinese BUS-11: a culturally adapted 11-item scale assessing chatbot usability across 5 dimensions. Because 2 BUS-11 dimensions ("Privacy and Security" and "Response Time") are single-item constructs in the original instrument, they were retained in the survey but treated as single-item indicators in factor-analytic planning (refer to "Data Analysis" section).

Data Collection

Data were collected electronically via Tencent Questionnaire. Participants completed tasks on their own devices and then accessed the BUS-11 through a survey link. Strict criteria ensured data quality, and questionnaires that were not fully completed were directly discarded. Based on pilot timing, online questionnaires completed in less than 5 minutes were excluded to reduce insufficient-effort responses. In the pilot experiment, the research team measured the time taken to complete the questionnaire, with an average duration of 5 minutes and 20 seconds. The finalized main study dataset comprised 438 valid chatbot evaluations.

Data Analysis

Data were exported to Microsoft Excel, verified, and analyzed using IBM SPSS Statistics (version 26.0). Reliability was evaluated using Cronbach α coefficients for the overall scale and for each of its multi-item dimensions. Structural validity was assessed via EFA after sampling-adequacy checks (Kaiser-Meyer-Olkin [KMO] index and Bartlett test of sphericity). We used a principal components extraction with the Kaiser criterion (eigenvalues >1.0) and scree plot inspection to determine the number of factors and applied varimax rotation to aid interpretability. Standard criteria guided item retention: primary factor loadings of ≥0.40, absence of substantial cross-loadings (≥0.30), and conceptual interpretability of the resulting factor solution. Salient loadings were defined as those with values of ≥0.40. In line with the instrument's structure, the EFA focused on the 9 items related to multi-item dimensions; the 2 single-item dimensions ("Privacy and Security" and "Response Time") were not included in the EFA but were retained as single-item indicators in descriptive analyses. $P<.05$ was considered statistically significant for inferential tests (eg, Bartlett test of sphericity). We note that principal component analysis (PCA) is commonly used in early-stage exploratory work for data reduction and structure discovery, and future work may complement this with common-factor methods (eg, principal axis factoring) and confirmatory factor analysis (CFA) once larger samples are available.

Ethical Considerations

This study was approved by the Ethics Committee of the School of Nursing, Peking Union Medical College (approval no PUMCSON-2023-10). All participants provided electronic informed consent before participation. Participation was voluntary, and participants could withdraw at any time without penalty. No personally identifying information was collected in the analytic dataset. Data were stored securely and analyzed in deidentified form.

Results

Scale Reliability and Validity

Structural Validity

Before factor extraction, data suitability was confirmed (KMO=0.743; Bartlett $\chi^2_{36}=628.185$; $P<.001$) because 2 BUS-11 dimensions are single-item constructs ("Privacy and Security" and "Response Time") and they were not included in the EFA. The analysis, therefore, used the remaining 9 items (refer to Figure 1 for the scree plot).

In this study, PCA and varimax orthogonal rotation were used to extract common factors. Nine items (items 10 and 11 were excluded from the EFA due to having only 1 item each) were analyzed and combined with the scree plot (Figure 1), and 3 common factors with eigenvalues greater than 1.0 were ultimately extracted, accounting for a cumulative variance contribution rate of 56.107%. The results of the EFA for this experiment are shown in Table 5.

XSL•FO
RenderX

Figure 1. Scree plot of the factors.

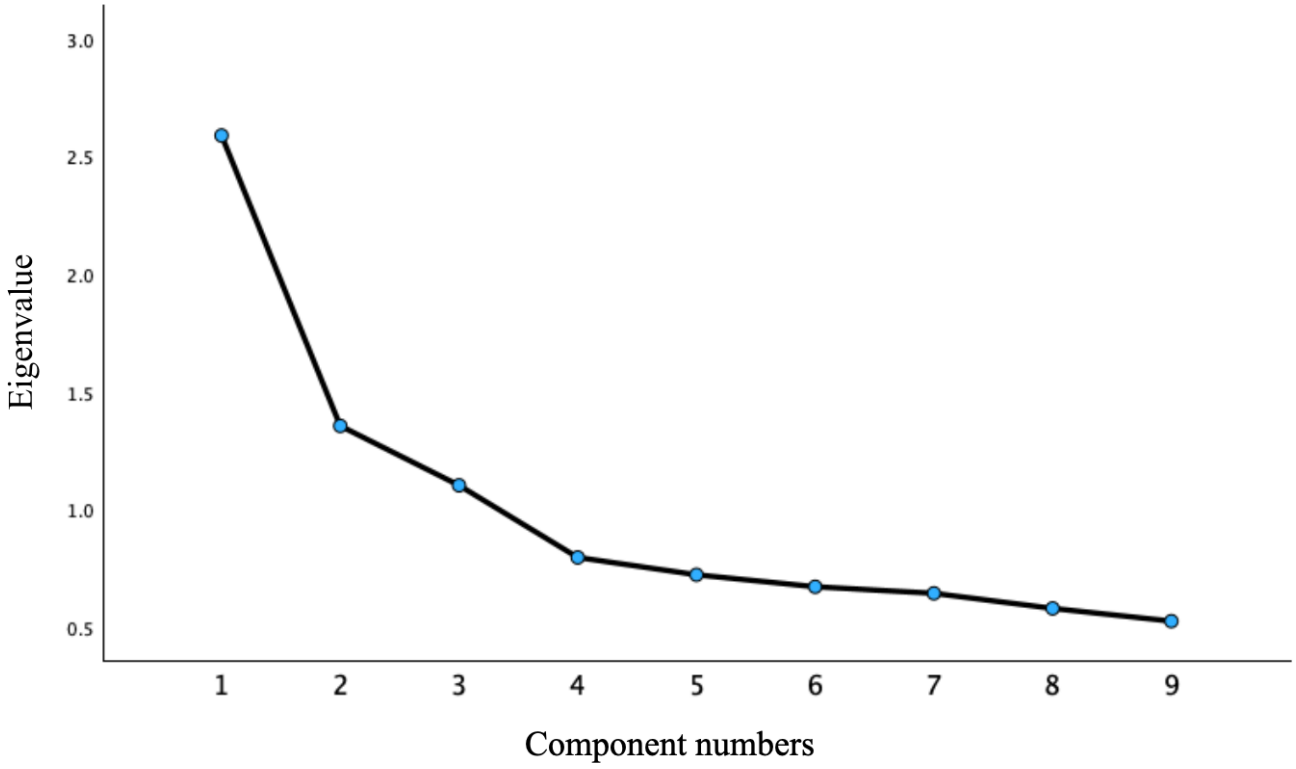


Table . Exploratory factor analysis (EFA) loadings (3-factor solution) for the Chinese 11-item Chatbot Usability Scale (BUS-11).

Item	Factor		
	1	2	3
Access the chatbot's functions			
Finding the entry to this chatbot is an easy task	0.840	—[a]	—
It is easy to find the various functions of this chatbot (eg, uploading a file, analyzing a file, changing chat avatar, and so on).	0.760	—	—
Quality of chatbot functions			
The communication process with this chatbot is smooth	—	0.757	—
This chatbot is able to maintain context coherence	—	0.759	—
The responses from this chatbot are easy to understand	—	0.716	—
Provide quality conversation and information			
I feel this chatbot understands my intentions and assists me in achieving the purpose of this chat	—	—	0.683
The information provided by this chatbot is appropriate, neither too much nor too little	—	—	0.627
This chatbot does not provide information relevant to my needs	—	—	0.629
I find the responses of this chatbot to be accurate and reliable	—	—	0.739

[a]Not available.

XSL•FO
RenderX

EFA with varimax rotation revealed a clear 3-factor structure for the Chinese version of BUS-11 (Table 5). The 3 factors explained 15.2%, 19.6%, and 21.3% of the total variance, respectively, accounting for a cumulative 56.1% of the variance. Factor 1 (Accessibility) consisted of 2 items reflecting the ease of locating and using chatbot functions, with strong loadings (0.840 and 0.760). Factor 2 (Interaction Process Quality) consisted of 3 items that described conversational smoothness, coherence, and comprehensibility, with loadings ranging from 0.716 to 0.759. Factor 3 (Information Quality) consisted of 4 items related to understanding user intent, the appropriateness of information, the absence of irrelevant content, and the reliability of responses, with loadings ranging from 0.627 to 0.739.

Only factors extracted by the eigenvalue >1 criterion are shown. Item domain labels reflect the original BUS-11 conceptual domains and are provided for reference.

Internal Consistency (Reliability)

The overall internal consistency of the Chinese BUS-11 was excellent (Cronbach α=0.92). By factor, reliability was acceptable to excellent, as indicated by Accessibility (Cronbach α=0.79), Interaction Process Quality (Cronbach α=0.91), and Information Quality (Cronbach α=0.93). Reliability was not computed for the single-item dimensions ("Privacy and Security" and "Response Time"). Item-level diagnostics (Table 6) showed corrected item-total correlations ranging from 0.453 to 0.819, indicating that all items were positively associated with the overall scale score. Cronbach α (if an item was deleted) ranged from 0.91 to 0.93, suggesting that removing any single item would not meaningfully improve internal consistency.

Table . Mean, SD, corrected item-total correlation, and Cronbach α (if an item is deleted) for the 11-item Chatbot Usability Scale (BUS-11).

Item	Values, mean (SD)	Corrected item-total correlation	Cronbach α (if an item is deleted)
1	3.79 (1.187)	0.539	0.93
2	3.57 (1.059)	0.616	0.92
3	3.77 (0.952)	0.819	0.91
4	3.77 (0.923)	0.785	0.91
5	3.79 (0.946)	0.816	0.91
6	3.66 (0.982)	0.817	0.91
7	3.57 (0.979)	0.785	0.91
8	3.53 (0.96)	0.771	0.91
9	3.55 (0.992)	0.798	0.91
10	2.80 (0.998)	0.453	0.93
11	4.08 (0.866)	0.469	0.93

User Feedback on Item Clarity and Acceptability

Overall, participants reported that the Chinese BUS-11 was easy to understand and straightforward to complete. Most respondents indicated that the items were clear and adequately reflected their experiences with chatbot interactions. The average completion time was approximately 5 minutes, suggesting that the scale imposed a minimal response burden and was suitable for use in both research and applied settings without causing respondent fatigue.

While feedback was generally positive, one recurrent issue was identified. Participants provided optional open-ended comments on item clarity and completion experience. Comments were reviewed and summarized into recurring issues. Three themes emerged: (1) ambiguity in interpreting the privacy/security item, with some participants indicating uncertainty about what constitutes "risk disclosure" in chatbot interactions; (2) context-dependence of judgments, where respondents noted that perceived usability varied by task type and chatbot domain; and (3) overall clarity and low burden, with many respondents reporting that the items were easy to understand and could be completed quickly. These feedback themes informed our interpretation of the psychometric findings and priorities for future refinement.

Sensitivity Analysis Across Sociodemographic Variables

To examine the robustness of the Chinese BUS-11 across user subgroups, we conducted sensitivity analyses by gender, age, and prior experience with chatbots (refer to Table 7). No significant differences in overall BUS-11 scores were observed between male and female participants (*P*=.12), nor across major age groups (*P*=.84), suggesting stable performance across these demographic characteristics. In contrast, participants with prior chatbot experience reported slightly higher BUS-11 scores than those without such experience (*P*=.03). This pattern indicates that the scale is sensitive to meaningful differences in user familiarity with chatbot interaction, while remaining robust across core sociodemographic variables.

XSL•FO
RenderX

Table . Sensitivity analysis of the 11-item Chatbot Usability Scale (BUS-11) scores across sociodemographic variables.

Variable and group	Value, mean (SD)	Statistical test	*P* value
Sex		*t* test	.12
Male	3.52 (0.80)		
Female	3.63 (0.69)		
Age (years)		One-way ANOVA	.84
18 - 24	3.55 (0.87)		
25 - 30	3.49 (0.95)		
31 - 40	3.30 (0.67)		
41 - 50	3.93 (0.67)		
Prior chatbot experience		*t* test	.03
Yes	3.64 (0.71)		
No	3.46 (0.78)		

Discussion

Rigorous Localization and Cultural Adaptation

This study localized the BUS-11 for the Chinese context, providing a standardized instrument for evaluating chatbot usability in China. Given BUS-11's theory-driven and multidimensional design, its adaptation is particularly relevant for contemporary conversational systems. Following established cross-cultural adaptation practices, we applied a forward-backward translation workflow (inspired by the Brislin model) combined with iterative expert committee review to ensure semantic and conceptual equivalence [20]. The expert panel (HCI/human factors–oriented) focused on pragmatic clarity, idiomatic appropriateness, and cultural relevance, resulting in a Chinese BUS-11 that remained closely aligned with the original without substantive item revisions.

Notably, the absence of major item-level changes should not be interpreted as evidence that culture is irrelevant to chatbot usability. Instead, cultural effects may operate more subtly, shaping how users interpret and prioritize usability facets (eg, privacy expectations, responsiveness norms, and conversational politeness) rather than necessitating overt rewording. Such differences may be reflected in response patterns and the observed factorial organization, underscoring the importance of future work using multigroup approaches (eg, measurement invariance testing) and direct cross-cultural comparisons to examine whether BUS-11 functions equivalently across cultural groups and whether certain dimensions show differential salience in Chinese users. In addition, consistent documentation of translation decisions improves transparency and supports reproducibility for subsequent adaptation work and comparative HCI research [20]. Given the rapid expansion of chatbot deployment across service domains [21], a psychometrically grounded Chinese BUS-11 can facilitate rigorous UX benchmarking in China and enable more meaningful integration of Chinese user evidence into the broader HCI literature [22].

Participant Sample and Representativeness

Our sample selection aimed for diversity in gender, age, education, and industry background, resulting in a broad cross-section of Chinese chatbot users. In practice, however, the achieved sample was skewed toward younger, highly educated adults. Most participants were aged 18 - 30 years, and more than 90% held at least a bachelor's degree. While this profile likely reflects the early adopters of new technologies [23] and those most engaged with contemporary chatbot applications, it also introduces some limitations on generalizability. Younger and tech-savvy users may have different usability perceptions than older or less technologically experienced individuals. Prior research in human factors has noted that older adults can face unique usability challenges and may interact with technology in distinct ways [24]. Our current sample underrepresents such groups, as well as users from rural areas or lower educational backgrounds, which means the usability impressions captured by the scale might not fully reflect the broader population of potential chatbot users. On the positive side, the concentration of experienced users in our sample could heighten sensitivity to subtle differences in chatbot interaction quality—these users are likely familiar with a variety of interfaces and may provide discerning feedback. The inclusion of participants from a range of industries (eg, education, health care, internet or information technology, and so on) is another strength, as it suggests the scale items were interpreted meaningfully across different usage contexts. Indeed, the BUS-11 was designed to be domain-agnostic, and our results tentatively support its applicability in multiple sectors. However, some fields (eg, finance and construction) were represented by relatively few respondents, so conclusions about those domains should be drawn cautiously. Likewise, because recruitment relied on convenience sampling (primarily in a few urban regions), there may be regional or cultural subgroup differences within China that our study did not capture. In summary, our sample provides a solid initial test of the scale among active chatbot users, but it is not fully representative of all user segments. Future research should broaden the sampling frame to include older adults, less frequent technology users, and a wider geographic distribution to ensure the Chinese BUS-11 is robustly validated for the general population. Broadening the participant base in this way would enhance the scale's universality and address the western, educated, industrialized, rich, and democratic bias analog in the Chinese context, thus

XSL•FO
RenderX

strengthening confidence that the tool works well for all target users, not just the young and educated subset.

Reliability, Validity, and Key Psychometric Findings

Overview

We conducted a comprehensive psychometric evaluation of the Chinese BUS-11, including assessments of content validity, construct (factorial) validity, internal consistency reliability, and user feedback on item clarity and practical usability. Overall, the findings provide strong evidence that the localized scale retains the scientific integrity of the original instrument. Here, we discuss each set of findings in turn, relating them to established benchmarks and prior research, and interpret their implications for the scale's use in HCI and human factors studies.

Content Validity

By consulting a panel of 8 experts to review the relevance of each item, we established that the content validity of the Chinese BUS-11 is high. The I-CVI for every item was very strong, with 4 items (Items 3, 4, 6, and 11) achieving an I-CVI of 1 (ie, unanimously deemed highly relevant by all experts). The remaining items also scored well above conventional acceptability thresholds [25]. Given that an I-CVI≥0.78 is often considered the minimum when 6 or more experts are involved, our results indicate that each item is considered clearly representative of the intended construct by domain specialists. The S-CVI was 0.920 (both by the averaging method and the universal agreement method), exceeding the common cutoff of 0.90 for excellent content coverage of a construct. These indices confirm that the BUS-11 (Chinese version) has comprehensive content coverage and the items collectively cover the important facets of chatbot usability without obvious omissions. This outcome was facilitated by the diverse expertise of our review panel, which included professionals in HCI, usability engineering, and systems engineering. Their varied perspectives helped ensure that the items were not only translated correctly but also conceptually appropriate for Chinese users. In practical terms, the high content validity suggests that Chinese BUS-11 can confidently be used to gauge chatbot usability aspects as intended—such as efficiency, clarity of answers, and user comfort—without missing key elements. The rigorous content validation process, in line with recommended scale development procedures, provides a solid foundation for the subsequent construct validation. It also offers reassurance to practitioners that the instrument has face validity in the local context (the items make sense to experts and presumably to end-users as well). We believe this strong content validity will help drive adoption of the scale in both research and industry evaluations, as stakeholders can trust that the instrument measures what it is supposed to measure (ie, salient user experience factors for chatbots).

Structural Validity (EFA)

To examine the underlying factor structure of the Chinese BUS-11, we carried out an EFA on the survey responses. Preliminary checks confirmed that our data were suitable for factor analysis: the KMO measure of sampling adequacy was 0.743, which is above the usual threshold of 0.60 and can be considered "middling" to "good" [26], and the Bartlett test of sphericity was highly significant, indicating sufficient interitem correlations for extracting latent factors. Using principal component extraction with varimax rotation, we identified a 3-factor solution that best fit the data. These 3 extracted common factors had eigenvalues greater than 1 and together accounted for about 56.1% of the total variance in users' responses. A cumulative variance above approximately 50% is acceptable in behavioral research, given the complexity of HCI measures, so 56% indicates that the scale captures a substantial portion of the usability perception variance in our sample. Each factor showed a clear thematic grouping of items, reflecting distinct dimensions of the chatbot user experience. Notably, 2 specific usability aspects (ie, "Privacy and Security" and "Response Time") were each originally measured by a single item in the BUS-11. As single-item facets, they could not load onto multi-item factors in the EFA and were thus analyzed separately (we examined their scores independently rather than including them in the factor structure). The remaining nine items clustered into three coherent factors: (1) Accessibility (eg, ease of accessing and using the chatbot, such as simple interface and low effort to start the interaction), (2) Interaction Quality (covering the smoothness, coherence, and understandability of the conversational exchange), and (3) Information Quality (assessing the relevance, clarity, and correctness of the chatbot's responses, without unnecessary redundancy). This structure is broadly consistent with the original conceptual framework of the BUS scale, which encompassed multiple dimensions of chatbot usability [10], but our findings suggest a somewhat more consolidated factor model in the Chinese context.

In the original development of the BUS, a 5-factor model (BUS-15) was proposed, which included separate dimensions for privacy and response speed alongside factors analogous to those we found [10]. Our EFA results imply that Chinese users may not distinguish as many separate categories, instead perceiving a more integrated set of usability factors. Attributes related to the conversational process and functional outcome appear to intertwine; users who find the interaction process smooth also tend to perceive the information provided as high quality, suggesting an underlying general perception of "interaction effectiveness." Interestingly, a recent re-examination of BUS-11 in a large multichatbot dataset also identified a simpler factor structure than initially theorized [27]. Our findings align with that perspective, reinforcing the idea that a parsimonious model (fewer and broader factors) can sufficiently capture chatbot usability evaluations. It is worth noting that we chose to retain the "Privacy and Security" and "Response Time" items as standalone indicators in the instrument, due to their importance in usability (users do care about privacy and speed) and their presence in the original scale. However, from a psychometric standpoint, single-item factors are suboptimal because they do not allow for internal consistency reliability estimation and cannot capture the breadth of a construct. Best practices in scale development typically recommend having at least 2 - 3 items per factor to achieve adequate reliability and construct representation [28]. Therefore, one implication is that future refinement of the Chinese BUS might involve expanding these facets—adding a couple of items to better gauge the privacy/security aspect of chatbot use (eg, covering data

XSL•FO
RenderX

protection, user consent, and so on) and the responsiveness aspect (eg, not just speed but also perceived responsiveness or promptness). For now, including those single items ensures that our localized scale does not lose any content relative to BUS-11, and they provide useful standalone measures (eg, designers might specifically want to see ratings of privacy transparency). In summary, the EFA supports the structural validity of the Chinese BUS-11 by confirming that its items coherently measure multiple distinct dimensions of usability. The identified 3-factor structure captures the major themes of chatbot usability experience in our sample. This structure provides an empirical basis for scoring or interpreting the scale (eg, computing subscale scores for Accessibility, Interaction Quality, and Information Quality) and sets the stage for future confirmatory testing to verify whether this structure generalizes to other samples or holds under more stringent statistical criteria.

User Feedback on Item Clarity

The observed variability in interpreting the privacy/security item suggests that privacy-related usability judgments may be more sensitive to users' privacy literacy and to how risk information is surfaced in routine interactions, consistent with prior privacy research indicating limited user engagement with privacy policies and disclosures [29,30]. From a measurement perspective, this supports retaining privacy/security as a relevant usability facet while motivating future refinement (eg, adding additional items or more concrete wording) and the use of cognitive interviewing to improve interpretability across user subgroups [31]. In addition, the brief completion time (≈3 - 4 minutes) suggests that the Chinese BUS-11 imposes low respondent burden, comparable to widely used usability instruments and suitable for rapid assessment in applied settings [32,33].

Limitations

While this study yielded encouraging results for the Chinese adaptation of BUS-11, several limitations must be acknowledged to contextualize the findings and guide future work. First, the sample, as discussed, was not fully representative of all user demographics. The geographic distribution of participants was relatively narrow (with many users from a few major cities), and certain age groups (especially adults aged older than 50 years) and occupations were underrepresented. The current sample underrepresents such groups, as well as users from rural areas or lower educational backgrounds, which means the usability impressions captured by the scale might not fully reflect the broader population of potential chatbot users. Indeed, the BUS-11 was designed to be domain-agnostic, and the results tentatively support its applicability in multiple sectors. However, some fields (eg, finance and construction) were represented by relatively few respondents, so conclusions about those domains should be drawn cautiously. Likewise, because recruitment relied on convenience and snowball sampling (primarily in a few urban regions), there may be regional or cultural subgroup differences within China that this study did not capture. This means the current validation is most directly applicable to younger, educated Chinese users in urban settings. Caution should be used in generalizing the psychometric results to, say, older adults or rural users, who might interact with chatbots differently or have different usability concerns. Expanding the sample in future studies will be important to verify that the scale maintains its reliability and validity across a broader swath of the Chinese population. Additionally, larger and more heterogeneous samples would allow for analyses of measurement invariance—checking whether the scale functions equivalently across subgroups such as older vs younger users, or high-frequency vs low-frequency chatbot users [34]. In addition to representativeness concerns, the overall sample size should be considered a methodological limitation. Although the study included 438 chatbot evaluations, these ratings were provided by 214 participants, which may be modest for robust factor-analytic procedures. According to the International Test Commission Guidelines for Translating and Adapting Tests [35], stable and generalizable factor solutions are typically supported by larger participant samples. The relatively limited number of participants in this study may therefore constrain the stability and generalizability of the extracted factor structure and other psychometric estimates. Future studies with larger and more diverse samples are needed to further confirm the robustness of the Chinese BUS-11's factor structure and measurement properties.

Second, certain BUS-11 dimensions in the current form rely on a single item (notably the "Privacy and Security" and "Response Time" facets). Single-item measures, while convenient, provide limited psychometric information. They can neither capture the full breadth of a construct nor estimate internal consistency for that construct. The strong performance of the overall scale notwithstanding, the precision and reliability of those facets could be improved. Future research should consider enriching these dimensions by developing and testing additional items. For example, the privacy facet might be expanded with items addressing whether the chatbot clearly asks permission for data usage or whether it provides settings for privacy control. The response time facet might include an item on whether the chatbot's speed meets user expectations or if delays ever cause frustration. Of course, any new items would themselves need to undergo translation and validation. The trade-off is that adding items could slightly increase completion time, but as long as the total remains reasonable (eg, <2 minutes increase), it would likely be worth it for a more robust scale. In short, the current findings highlight the need to bolster those single-item constructs for a more rigorous measurement model [28].

Third, this study primarily evaluated the scale in general usage scenarios (common chatbot systems for generic tasks) and did not deeply examine how the scale performs in specific industry contexts or use-case scenarios. It remains an open question whether users in specialized domains—such as health care, education, finance, or customer service—interpret and prioritize the BUS-11 items similarly. Different contexts may put extra emphasis on certain usability aspects; for instance, in health care chatbot applications, issues such as empathy, trust, and accuracy of information might be even more critical [36], whereas in e-commerce chatbots, speed and transactional clarity could dominate user satisfaction. The inability to test such variations is a limitation. It is possible that the BUS-11 might require minor tweaks or weighting adjustments to be optimally effective in certain domains, or it may prove to be robust as

XSL•FO
RenderX

is—we simply do not have the data yet to say. Future validations should include a variety of chatbot types and domains to ensure the scale's content validity and structure hold in those settings, or to highlight any domain-specific gaps that could be addressed by additional context-specific items.

Additionally, this study did not use external, established usability instruments to assess convergent validity. Although the BUS-11 is conceptually grounded in usability theory, the absence of a comparator scale—such as the System Usability Scale or other widely used usability measures—limits the ability to empirically examine whether the Chinese BUS-11 correlates appropriately with validated measures of related constructs. As a result, the current validity evidence is primarily restricted to internal structure and reliability and does not allow for a direct assessment of alignment with existing usability benchmarks. Future studies should consider incorporating one or more established usability scales to evaluate convergent (and potentially discriminant) validity, thereby strengthening the overall validity framework and situating the Chinese BUS-11 more clearly within the broader landscape of usability measurement tools. Moreover, test-retest reliability was not assessed due to the cross-sectional design. Therefore, the temporal stability of the Chinese BUS-11 could not be directly evaluated. Future studies should administer the scale to the same users across time points to examine score stability when the usability of a chatbot system remains unchanged.

Finally, the psychometric analysis relied on EFA due to the novelty of applying the BUS-11 in Chinese. We have not yet conducted a CFA on a separate sample, which would be the next step to confirm whether the 3-factor model (plus the 2 single-item factors) holds under stricter modeling conditions. A CFA would allow us to test the hypothesized factor structure (possibly including correlated factors or a higher-order factor of overall usability) and to adjust for measurement errors [37]. It would also enable a formal comparison of alternative models—for example, whether a structure with 5 distinct factors (if privacy and speed were expanded) fits significantly better than a 3-factor structure, or whether a single-factor model (treating usability as unidimensional) is decisively inferior to a multifactor model. Additionally, with a large enough sample, multigroup CFA could be used to test measurement invariance across key groups (eg, male vs female, younger vs older, and different regions), ensuring the scale operates equivalently [34]. Because the current sample size was modest for CFA and was collected as a single group, we reserve these confirmatory analyses for future work. In summary, the use of EFA was appropriate for initial validation; however, further confirmation via CFA is needed to solidify the factor structure and verify that the Chinese BUS-11 meets the standards of psychometric validity expected for a widely applicable research instrument.

The exploratory factor analytic results should be interpreted considering the factor extraction approach used in this study. We note that PCA is commonly applied in early-stage exploratory research for data reduction and initial structure discovery; however, PCA does not explicitly model latent constructs. As such, while appropriate for an initial examination of the scale structure, this approach provides only preliminary evidence regarding the underlying factor model. Future research with larger samples may complement these findings by applying common-factor methods (eg, principal axis factoring) and conducting CFA to more rigorously test the latent structure of the Chinese BUS-11.

Conclusion

This study delivers a validated Chinese adaptation of BUS-11 with excellent internal consistency, strong expert-rated content validity, and a clear 3-factor structure for multi-item facets (Accessibility, Interaction Process Quality, and Information Quality), while preserving the privacy and speed facets conceptually. The instrument fills a measurement gap for chatbot UX evaluation in China and provides researchers and practitioners with a concise, psychometrically sound tool suitable for laboratory and field contexts alike. Future work should establish temporal stability, convergent validity, and confirmatory structure and expand coverage to underrepresented user groups and domain-specific scenarios to further strengthen its utility and generalizability.

Funding

This study was funded by the Non-Profit Central Research Institute Fund of the Chinese Academy of Medical Sciences (grant no 2023-RC320-01).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to privacy and confidentiality considerations but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Standardized user tasks applied across 10 popular chatbot systems in China to ensure consistency and ecological validity in usability evaluations with the Chinese 11-item Chatbot Usability Scale (BUS-11).
[DOCX File, 21 KB - humanfactors_v13i1e84971_app1.docx]

XSL•FO
RenderX

References

1. Nagarhalli TP, Vaze V, Rana NK. A review of current trends in the development of chatbot systems. In: Rana N, editor. Presented at: 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS); Mar 6-7, 2020. [doi: 10.1109/ICACCS48705.2020.9074420]
2. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019;5:2055207619871808. [doi: 10.1177/2055207619871808] [Medline: 31467682]
3. Androutsopoulou A, Karacapilidis N, Loukis E, Charalabidis Y. Transforming the communication between citizens and government through AI-guided chatbots. Gov Inf Q 2019 Apr;36(2):358-367. [doi: 10.1016/j.giq.2018.10.001]
4. Landim A, Pereira AM, Vieira T, et al. Chatbot design approaches for fashion e-commerce: an interdisciplinary review. Int J Fash Des Technol Educ 2022 May 4;15(2):200-210. [doi: 10.1080/17543266.2021.1990417]
5. Li FL, Qiu M, Chen H, Wang X, Gao X. AliMe Assist: an intelligent assistant for creating an innovative e-commerce experience. In: Huang J, editor. 2017 Presented at: Proceedings of the 2017 ACM Conference on Information and Knowledge Management; Nov 6-10, 2017 URL: https://www.researchgate.net/publication/320885511_AliMe_Assist_An_Intelligent_Assistant_for_Creating_an_Innovative_E-commerce_Experience [accessed 2026-03-27]
6. Wang Y, Zhang N, Zhao X. Understanding the determinants in the different government AI adoption stages: evidence of local government chatbots in China. Soc Sci Comput Rev 2022 Apr;40(2):534-554. [doi: 10.1177/0894439320980132]
7. Shan Y, Ji M, Xie W, et al. Use of health care chatbots among young people in China during the omicron wave of COVID-19: evaluation of the user experience of and satisfaction with the technology. JMIR Hum Factors 2022 Jun 9;9(2):e36831. [doi: 10.2196/36831] [Medline: 35576058]
8. Silva GRS, Canedo ED. Towards user-centric guidelines for chatbot conversational design. Int J Hum Comput 2024 Jan 17;40(2):98-120. [doi: 10.1080/10447318.2022.2118244]
9. Hornbæk K. Current practice in measuring usability: challenges to usability studies and research. Int J Hum Comput Stud 2006 Feb;64(2):79-102. [doi: 10.1016/j.ijhcs.2005.06.002]
10. Borsci S, Malizia A, Schmettow M, et al. The Chatbot Usability Scale: the design and pilot of a usability scale for interaction with AI-based conversational agents. Pers Ubiquit Comput 2022 Feb;26(1):95-119. [doi: 10.1007/s00779-021-01582-9]
11. Ren R, Zapata M, Castro JW, Dieste O, Acuna ST. Experimentation for chatbot usability evaluation: a secondary study. IEEE Access 2022;10:12430-12464. [doi: 10.1109/ACCESS.2022.3145323]
12. Ren R, Castro JW, Acuña ST, de Lara J. Evaluation techniques for chatbot usability: a systematic mapping study. Int J Soft Eng Knowl Eng 2019 Nov;29(11n12):1673-1702. [doi: 10.1142/S0218194019400163]
13. Holmes S. Towards validating a chatbot usability scale. In: International Conference on Human-Computer Interaction: Springer. [doi: 10.1007/978-3-031-35708-4_24]
14. Borsci S, Prati E, Malizia A, Schmettow M, Chamberlain A, Federici S. Ciao AI: the Italian adaptation and validation of the Chatbot Usability Scale. Pers Ubiquit Comput 2023 Dec;27(6):2161-2170. [doi: 10.1007/s00779-023-01731-2]
15. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000 Dec 15;25(24):3186-3191. [doi: 10.1097/00007632-200012150-00014] [Medline: 11124735]
16. Wang Y, Lei T, Liu X. Chinese System Usability Scale: translation, revision, psychological measurement. Int J Hum-Comput Int 2020 Jun 14;36(10):953-963. [doi: 10.1080/10447318.2019.1700644]
17. Bolpagni M, Gabrielli S. Development of a comprehensive evaluation scale for LLM-powered counseling chatbots (CES-LCC) using the edelphi method. Public Health and Healthcare. Preprint posted online on 2025. [doi: 10.20944/preprints202501.1621.v1]
18. Jones PS, Lee JW, Phillips LR, Zhang XE, Jaceldo KB. An adaptation of Brislin's translation model for cross-cultural research. Nurs Res 2001;50(5):300-304. [doi: 10.1097/00006199-200109000-00008] [Medline: 11570715]
19. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018 May;27(5):1147-1157. [doi: 10.1007/s11136-018-1798-3] [Medline: 29435801]
20. Brislin RW. Back-translation for cross-cultural research. J Cross Cult Psychol 1970 Sep;1(3):185-216. [doi: 10.1177/135910457000100301]
21. Brandtzaeg PB, Følstad A. Chatbots: changing user needs and motivations. Interactions 2018;25(5):38-43 [FREE Full text] [doi: 10.1145/3236669]
22. Norman D. The Design of Everyday Things: Hachette UK; 2013.
23. Rogers EM, Singhal A, Quinlan MM. Diffusion of innovations. In: An Integrated Approach to Communication Theory and Research: Routledge; 2014:432-448 URL: https://www.taylorfrancis.com/chapters/edit/10.4324/9780203887011-36/diffusion-innovations-everett-rogers-arvind-singhal-margaret-quinlan [accessed 2026-03-27]
24. Charness N, Boot WR. Aging and information technology use: potential and barriers. Curr Dir Psychol Sci 2009;18(5):253-258. [doi: 10.1111/j.1467-8721.2009.01647.x]
25. Lynn MR. Determination and quantification of content validity. Nurs Res 1986;35(6):382-385. [Medline: 3640358]
26. Kaiser HF. An index of factorial simplicity. Psychometrika 1974 Mar;39(1):31-36. [doi: 10.1007/BF02291575]

XSL•FO
RenderX

27. Borsci S, Schmettow M. Re-examining the chatBot Usability Scale (BUS-11) to assess user experience with customer relationship management chatbots. Pers Ubiquit Comput 2024 Dec;28(6):1033-1044. [doi: 10.1007/s00779-024-01834-4]
28. Worthington RL, Whittaker TA. Scale development research: a content analysis and recommendations for best practices. Couns Psychol 2006;34(6):806-838. [doi: 10.1177/0011000006288127]
29. McDonald AM, Cranor LF. The cost of reading privacy policies. Isjlp 2008;4:543.
30. Baruh L, Secinti E, Cemalcilar Z. Online privacy concerns and privacy management: a meta-analytical review. J Commun 2017 Feb;67(1):26-53. [doi: 10.1111/jcom.12276]
31. Sweller J. Cognitive load during problem solving: effects on learning. Cogn Sci 1988 Apr;12(2):257-285. [doi: 10.1207/s15516709cog1202_4]
32. Brooke J. SUS: a “quick and dirty” usability scale. In: Usability Evaluation In Industry: CRC Press; 1996:4-7 URL: http://taylorfrancis.com/chapters/edit/10.1201/9781498710411-35/sus-quick-dirty-usability-scale-john-brooke [accessed 2026-03-27]
33. Lewis CC, Fretwell CE, Ryan J, Parham JB. Faculty use of established and emerging technologies in higher education: a unified theory of acceptance and use of technology perspective. IJHE 2013;2(2):22-34. [doi: 10.5430/ijhe.v2n2p22]
34. Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. Organ Res Methods 2000 Jan;3(1):4-70. [doi: 10.1177/109442810031002]
35. Hernández A, Hidalgo MD, Hambleton RK, Gómez-Benito J. International Test Commission guidelines for test adaptation: a criterion checklist. Psicothema 2020 Aug;32(3):390-398. [doi: 10.7334/psicothema2019.306] [Medline: 32711675]
36. Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 1;25(9):1248-1258. [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
37. Brown T. Confirmatory Factor Analysis for Applied Research, 2nd edition: The Guilford Press; 2015. URL: https://psycnet.apa.org/record/2015-10560-000 [accessed 2026-03-27]

Abbreviations

BUS-11: 11-item Chatbot Usability Scale
CFA: confirmatory factor analysis
CUQ: Chatbot Usability Questionnaire
EFA: exploratory factor analysis
HCI: human-computer interaction
I-CVI: item-level content validity index
KMO: Kaiser-Meyer-Olkin
PCA: principal component analysis
S-CVI: scale-level content validity index
UX: user experience

Edited by A Kushniruk; submitted 28.Sep.2025; peer-reviewed by G Gronier, J Brooke; revised version received 10.Jan.2026; accepted 19.Feb.2026; published 06.Apr.2026.

Please cite as:
Ma H, Pei R, Li S, Wang A, Tang X, Piao M
Chatbot Usability Scale in Chinese Users: Cross-Cultural Adaptation and Validation Study
JMIR Hum Factors 2026;13:e84971
URL: https://humanfactors.jmir.org/2026/1/e84971
doi:10.2196/84971

© Haoming Ma, Runyuan Pei, Sijia Li, Aoqi Wang, Xingyi Tang, Meihua Piao. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 6.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Web-Based Application for Cognitive and Functional Assessments in Dementia Screening: Mixed Methods, User-Centered Development Approach

Nattaporn Piyaamornpan[1,2,3*], MD; Suwat Srisuwannanukorn[4], MD; Kosit Tangthamrongthanawat[5], MD; Pornsawan Mekhasingharak[6], MD; Chatchawan Rattanabannakit[1], MD; Saowalak Hunnangkul[7], PhD; Natthamon Wongkom[1], BA; Vorapun Senanarong[1*], MD, FRCP

[1]Division of Neurology, Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, 2 Wang Lang Rd, Siriraj, Bangkok Noi, Bangkok, Thailand

[2]Department of Medicine, Phatthalung Hospital, Phatthalung, Thailand

[3]Division of Neurology, Department of Internal Medicine, Tha Sala Hospital, Nakhon Si Thammarat, Thailand

[4]Division of Neurology, Department of Medicine, Faculty of Medicine Vajira Hospital, Navamindradhiraj University, Bangkok, Thailand

[5]Department of Internal Medicine, Nakhon Phanom Hospital, Nakhon Phanom, Thailand

[6]Regional Health Promotion Center 2 Phitsanulok, Phitsanulok, Thailand

[7]Department of Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

[*]these authors contributed equally

Corresponding Author:
Vorapun Senanarong, MD, FRCP
Division of Neurology, Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, 2 Wang Lang Rd, Siriraj, Bangkok Noi, Bangkok, Thailand

Abstract

Background: Digital health technologies offer new opportunities for cognitive screening and monitoring among older adults. In Thailand, where dementia prevalence is rising, accessible web-based cognitive tools remain limited despite their potential to facilitate early detection and community-based assessment. Understanding usability and validity is critical to ensure successful implementation in real-world contexts.

Objective: This study aimed to develop and validate a web-based application, *Healthy Brain Test*, for cognitive and functional assessments in dementia screening among older Thai adults. Specific objectives were to (1) design user-centered cognitive modules covering key cognitive domains and (2) evaluate correlations between the web-based assessments and conventional clinical tools to determine diagnostic cutoffs for cognitive impairment.

Methods: We designed *Healthy Brain Test* as a self-administered web application suitable for older users and their caregivers. The platform includes digital versions of the Thai Mental State Examination (e-TMSE), a clock drawing test, and a category verbal fluency test, along with electronic versions of the short form of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE-16) and cognitive instrumental activities of daily living (IADLs). Participants completed both web-based and paper-based assessments. Correlations between modalities were analyzed, and receiver operating characteristic (ROC) curves were generated to determine sensitivity and specificity. Data were analyzed using SPSS for Windows, version 30.0 (IBM Corp) and MedCalc Statistical Software (MedCalc Software Ltd).

Results: A total of 198 older adults participated (women: 137/198, 69.2%; median age 69.4 years), with 57.1% (113/198) having more than 6 years of education. Of the 198 participants, 44 were diagnosed with major neurocognitive disorder, 58 were diagnosed with mild neurocognitive disorder, and 96 were cognitively normal. The e-TMSE showed strong agreement with the traditional TMSE (r=0.837; P<.001). Category verbal fluency, IQCODE-16, and IADL modules also demonstrated significant correlations (P<.001). The e-TMSE achieved an area under the ROC curve of 0.84 (bootstrapped 95% CI 0.78 - 0.89); a cutoff ≤23 provided 88.6% sensitivity and 70.1% specificity for identifying major neurocognitive disorder. Participants reported high ease of use and engagement during pilot testing.

Conclusions: *Healthy Brain Test* demonstrated strong validity and usability as a web-based cognitive and functional assessment platform for dementia screening. Its integration of established cognitive measures into a digital interface enables remote, accessible, and user-friendly evaluation for older adults and caregivers. Future research should assess long-term feasibility, user adherence, and integration with clinical workflows to support large-scale screening initiatives.

XSL•FO
RenderX

(JMIR Hum Factors 2026;13:e85454) doi:10.2196/85454

KEYWORDS

Alzheimer disease; cognitive screening; dementia; functional assessment; neurocognitive disorders; neuropsychology; web application

Introduction

Cognitive performance often declines with normal healthy aging [1], with measurable deficits in episodic memory, working memory, verbal fluency, and executive control processes such as visual attention and set-shifting [2]. These age-related changes do not inevitably progress to dementia. However, dementia remains a significant public health challenge in Thailand, and prevalence rises sharply with age. A national survey reported that 3.3% of Thai adults aged 60 years and older have some form of dementia, increasing from 1% at ages 60 years to 64 years to 31.3% among those aged 90 years and older [3].

Digital technologies have advanced cognitive screening, offering potential benefits for clinical treatment, therapeutic interventions, and behavioral research. Web-based applications now extend beyond screening to support self-management, caregiver education, cognitive rehabilitation, and remote clinical monitoring. Recent studies indicate that such applications can contribute across the dementia care continuum [4-8].

In Thailand, web-based dementia care applications are still early in development. *Dementia U-Care*, designed to help community health workers screen and monitor cognitive data, has demonstrated feasibility and positive user experiences [9]. Vorasut and Tantatsanawong [10] introduced smartphone-based screening tools using the Thai version of the Mini-Mental State Examination (MMSE; *MMSE-Thai 2002*) with text-to-speech and audio recording. More recently, Yangyuensathaporn et al [11] validated the Application-Based Cognitive Screening Test for older Thai adults, reporting 92.9% sensitivity and 70% specificity.

Standard neuropsychological instruments, including the MMSE and Montreal Cognitive Assessment (MoCA), have long served as benchmarks for detecting early cognitive impairment. The MMSE has a sensitivity of 87% to 97% and specificity of 70% to 82% depending on the population and selected cutoff points [12]. The MoCA, designed to be more sensitive to mild cognitive impairment, typically achieves about 90% sensitivity and approximately 87% specificity [13].

Given population aging and the demand for accessible screening, there is a need to develop localized, culturally appropriate digital tools. Therefore, we designed *Healthy Brain Test*, a test battery usable via a web-based application on tablet computers and mobile phones. Our objectives were to develop a web-based cognitive and functional assessment platform aimed at dementia screening in older Thai adults and to design diverse cognitive modules targeting different cognitive domains. We also sought to evaluate correlations between the web-based assessments and traditional tools.

Methods

Study Population

Participants older than 40 years were prospectively and purposefully recruited from the Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand. Recruitment aimed to evaluate cognitive and functional performance for dementia screening using a dedicated web-based application. All participants had either relatives or long-term caregivers who could provide relevant information on cognitive status and daily functioning at present and over the past 10 years.

Inclusion and Exclusion Criteria

The inclusion criteria were provision of informed consent after full explanation of study objectives and procedures; adequate literacy, vision, and hearing; and ability to read and comprehend Thai. The exclusion criteria were acute confusion, impaired decision-making capacity due to cognitive or medical conditions, and severe physical or psychiatric illness.

Diagnostic Classification

The diagnoses of major and mild neurocognitive disorders were established according to the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision* (*DSM-V-TR*) criteria [14]. Cognitively healthy control participants were recruited from relatives or spouses of patients and community volunteers. These individuals had no reported cognitive decline, demonstrated normal performance on cognitive assessments, and maintained independence in daily functioning.

Ethical Considerations

All participants provided written informed consent before enrollment. The study protocol was approved by the Human Research Protection Unit, Faculty of Medicine Siriraj Hospital, Mahidol University (COA No. Si 982/2021), and conducted in accordance with the Declaration of Helsinki. All participant data were kept confidential and anonymized in accordance with the approved ethical protocol. Participants in Phase I received 100 (US $3.20) for their participation, while participants in Phase II did not receive any financial compensation.

Study Design

This study used a user-centered design approach to develop and validate a web-based cognitive and functional screening application for the early detection of dementia. Development proceeded through 3 consecutive phases involving collaboration among computer scientists, health informaticians, psychologists, and neurologists. All phases were guided by clinical relevance, usability, and data security considerations.

Phase I: Prototype Development and Usability Testing

In Phase I, we developed a prototype of the *Healthy Brain Test* web-based application and iteratively refined it to optimize

XSL•FO
RenderX

usability. The application was designed for Android-based systems to ensure widespread accessibility and compatibility with mobile devices commonly used in community and clinical settings. Usability testing sessions with volunteer participants generated feedback that informed successive iterations of the user interface. The interface comprised 2 sections. The home page introduced the application, provided general usage instructions, and presented an electronic informed consent form. Users were required to review and agree to the consent before accessing the assessment modules. The assessment section included 5 digital cognitive and functional tasks adapted from established neuropsychological assessments. Each task was preceded by a brief, standardized instructional message, and the design enhanced interactivity while reducing reliance on motor skills. Initial trials were conducted with volunteer participants to ensure content comprehension, ease of use, and engagement.

Phase II: The Pilot Study

A pilot study at Siriraj Hospital, Thailand, evaluated the validity of the *Healthy Brain Test* web-based cognitive screening application against traditional paper-and-pencil assessments. A total of 77 participants were enrolled (50 women and 27 men). Participants were categorized into 3 groups: 21 with major neurocognitive disorder (mild to moderate dementia), 23 with mild cognitive impairment, and 33 cognitively normal control participants. Each participant completed both conventional and digital cognitive assessments. The traditional paper-and-pencil assessments included the standard Thai Mental State Examination (TMSE) and the clock drawing test (CDT), administered under the supervision of trained clinicians. After the paper-and-pencil tests, participants were guided through the web-based *Healthy Brain Test* application. The full assessment session lasted approximately 1 hour to 2 hours depending on participant performance.

Phase III: Practical Use and Data Collection

In the final phase, we implemented the *Healthy Brain Test* web-based application in real-world settings across 5 health centers in Thailand. The sites were Siriraj Hospital, Vajira Hospital, Phatthalung Hospital, Phitsanulok Health Center 2, and Nakhon Phanom Hospital. This phase aimed to assess practical use in community-based and clinical environments and to facilitate the early detection of cognitive impairment and dementia. The application was accessible via health care professionals or self-assessment under observation, and participants could complete the digital screening autonomously or with minimal assistance. This approach supported broader reach and feasibility across diverse settings, including rural areas. We enrolled 198 participants (137 women and 61 men). Based on clinical evaluation, participants were categorized as follows: 44 with major neurocognitive disorder, 58 with mild cognitive impairment, and 96 with normal cognition. Digital screening assessed cognitive and functional performance using 5 standardized tasks embedded in the application. We also collected demographic and educational data and stored them on secure cloud servers, ensuring confidentiality and compliance with data protection standards.

The Healthy Brain Test Web-Based Application

Healthy Brain Test is a web-based digital cognitive and functional screening tool that integrates 5 core neuropsychological tasks digitally adapted from validated conventional tests widely used to assess cognitive function. The battery included the following 5 tasks: a slightly modified electronic TMSE (e-TMSE); the CDT implemented as the Clox test; and a category verbal fluency test (VFT). The remaining tasks were the short form of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE-16) and cognitive instrumental activities of daily living (cognitive IADLs). Illustrative screenshots of the assessment modules are shown in Figure 1.

XSL•FO
RenderX

Figure 1. Healthy brain test module overview illustrating the list of 5 modules: electronic Thai Mental State Examination (e-TMSE), the Clox test (digitized clock drawing), category verbal fluency test (animal naming), short form of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE-16), and cognitive instrumental activities of daily living (cognitive IADLs).



Electronic Thai Mental State Examination

The e-TMSE, a digital adaptation of the standard paper-and-pencil TMSE [15], was designed to assess cognitive function through an interactive web-based interface. The traditional TMSE consists of 11 items covering 6 cognitive domains: orientation, registration, attention and calculation, recall, language, and visuospatial abilities. Each component was implemented in the digital platform, and the original maximum total score of 30 points was preserved. Participants used touch to choose from on-screen options or to handwrite responses for language items. Samples of the e-TMSE test screens are shown in Figures 2 and 3.

XSL•FO
RenderX

Figure 2. Electronic Thai Mental State Examination (e-TMSE) sample screens showing orientation and registration, which use touch-based entry.

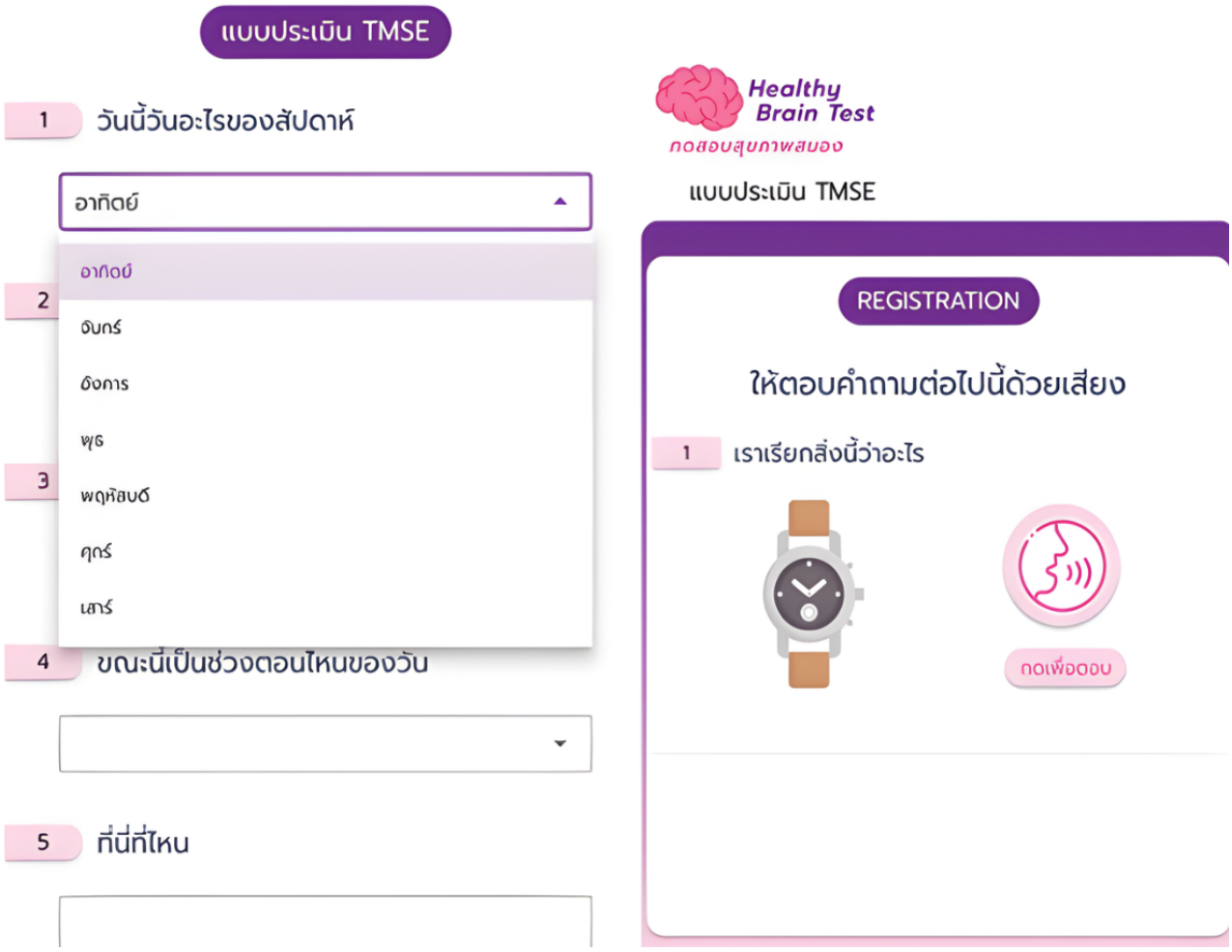
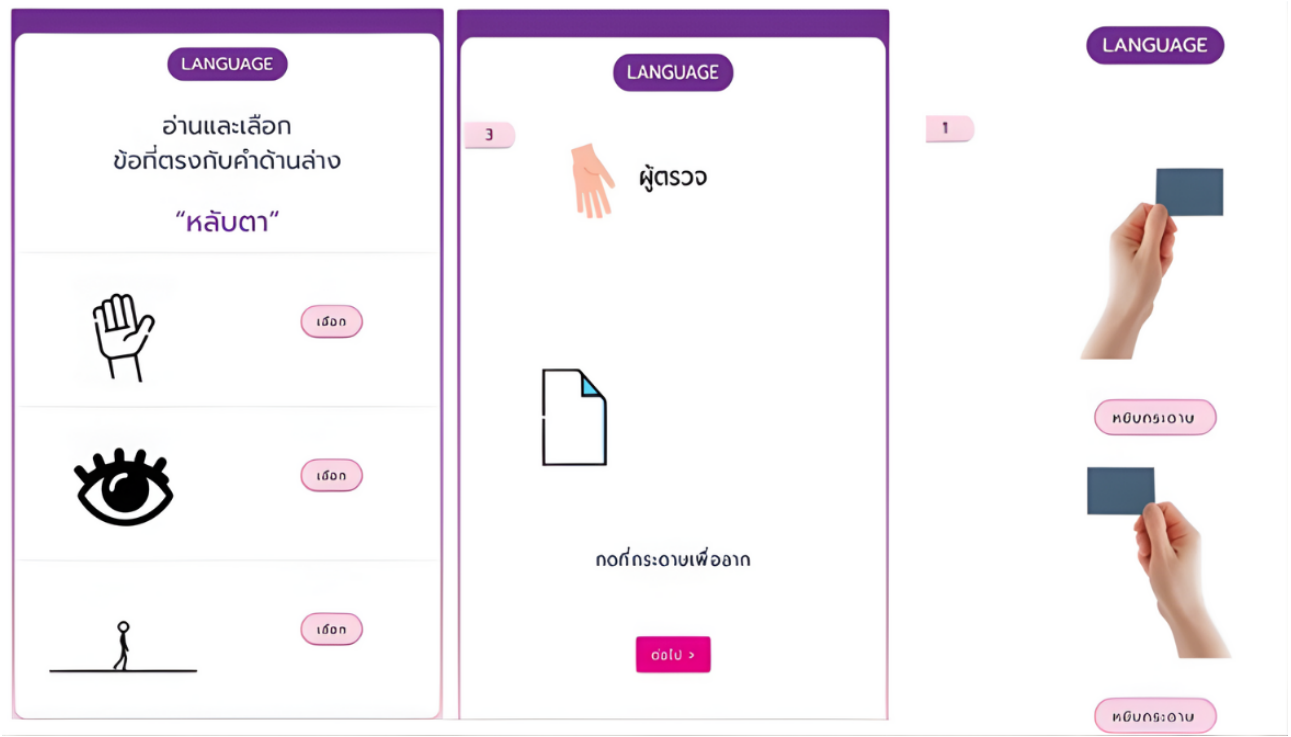


Figure 3. Electronic Thai Mental State Examination (e-TMSE) sample screens showing the language items, which allow on-screen handwriting and choice selection.

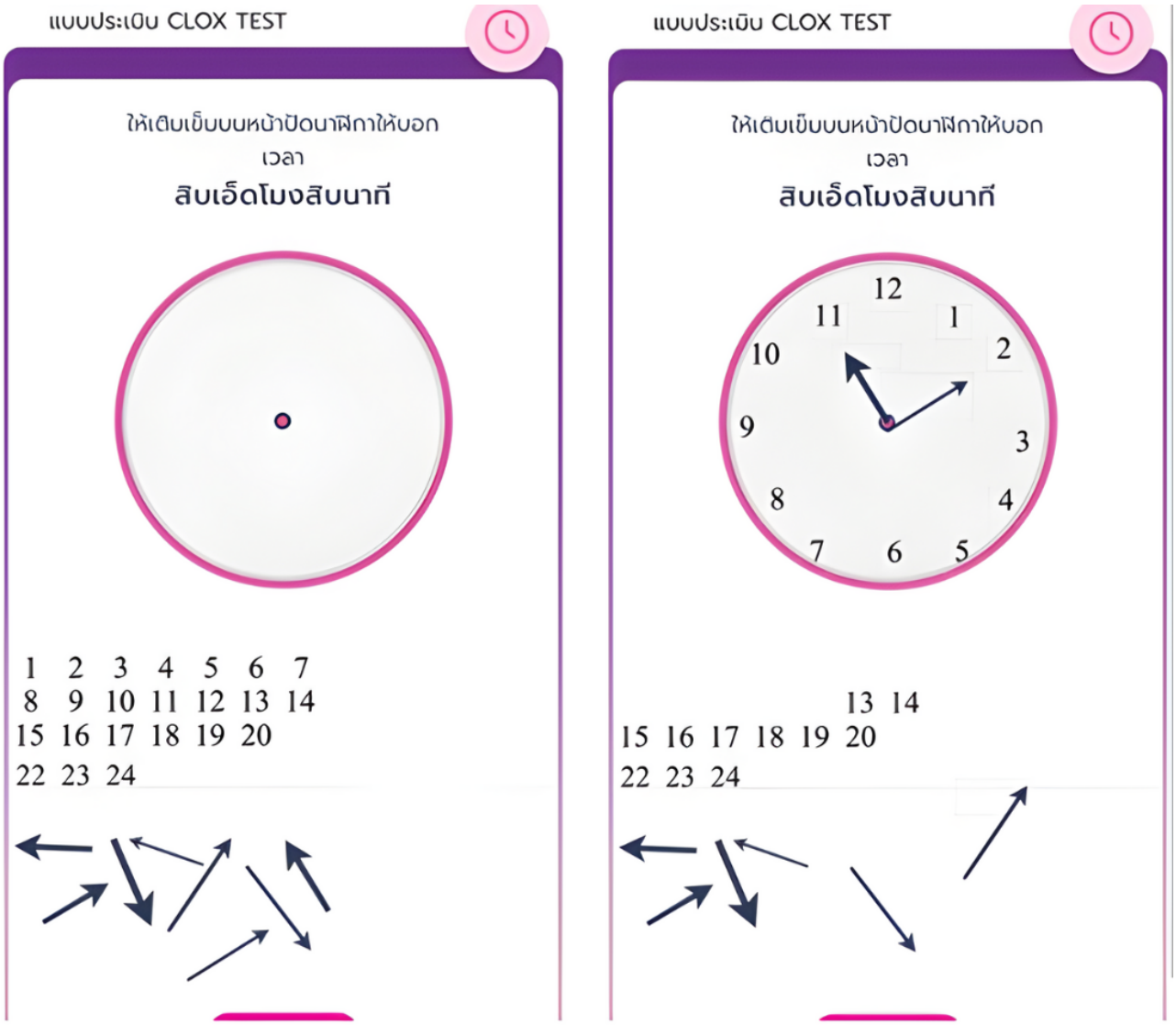


XSL•FO
RenderX

Clock Drawing Test

The CDT is widely used to assess visuospatial ability and executive function and has attracted interest for early cognitive impairment screening [16]. In this study, a digitized CDT variant, the Clox test, was implemented within the *Healthy Brain Test* application. Participants drew a clock showing a specified time on a touch-based interface. The test used a 10-point scale, where 0 indicated severe impairment and 10 represented normal performance. Illustrative screenshots are shown in Figure 4.

Figure 4. Digitized clock drawing (the clox test) sample screens showing how participants draw a clock showing a specified time on a touch-based canvas.

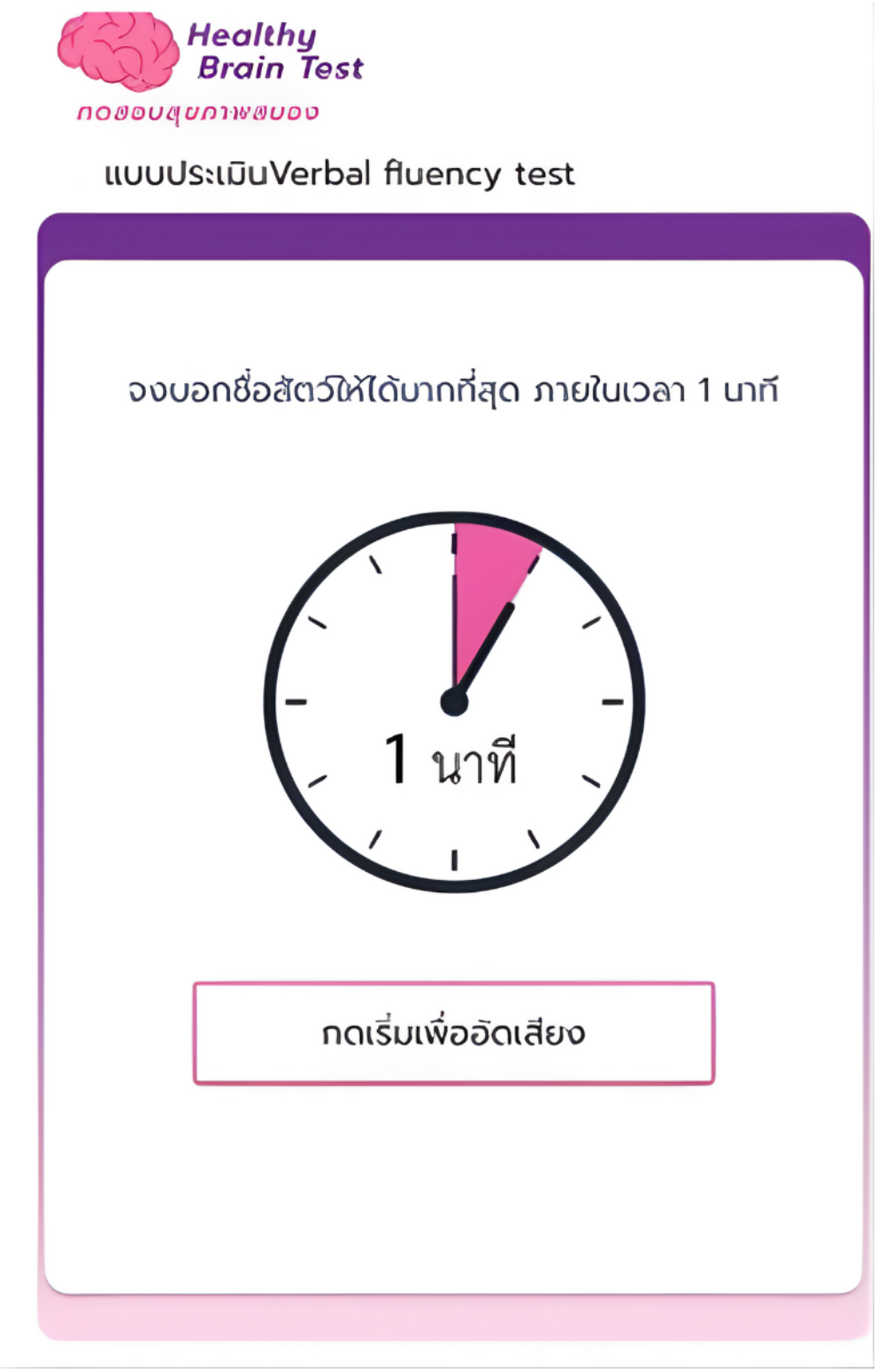


Category VFT

The category (semantic) VFT is a well-established neuropsychological measure that assesses executive function and semantic memory and shows high sensitivity for detecting cognitive impairment, including mild cognitive impairment and dementia [17,18]. In this task, participants generate as many words as possible from a specified category within a limited time, typically 1 minute. In this study, we used the animal naming category and instructed participants to name as many animals as possible within 1 minute. Illustrative screenshots are shown in Figure 5.

XSL•FO
RenderX

Figure 5. Category verbal fluency test (animal naming) sample screens, in which participants name as many animals as possible in 1 minute. The interface provides standardized instructions and timed entry.



Short Form of the Informant Questionnaire on Cognitive Decline in the Elderly

The short form of the IQCODE consists of 16 items that assess informants' perceptions of a participant's cognitive decline by comparing current performance with performance 10 years earlier. In this study, informants were relatives, long-term paid caregivers, or neighbors who had been providing some care. They rated change on a 5-point scale from 1 (much improved) to 5 (much worse), and the score was the mean of the 16 items [19].

Cognitive IADLs

The cognitive IADL module in *Healthy Brain Test* is derived from the validated Thai Activities of Daily Living (ADL) scale [20]. The instrument includes 6 basic ADL items and 7 cognitive

XSL•FO
RenderX

IADL items, and it evaluates the capacity to perform essential daily tasks that require physical function and cognitive integration. The Thai ADL measure has shown strong psychometric properties, including high test-retest reliability and robust construct validity. It also shows strong correlations with the TMSE, Clinical Dementia Rating, Barthel Index, and Functional Activities Questionnaire.

Measurements

Baseline Assessment

Trained physicians and neurologists conducted structured interviews to obtain baseline clinical and demographic data, including age, sex, educational attainment, and cognitive status.

Statistical Analysis

We calculated the required sample size using an estimated dementia prevalence of 20% in older Thai adults based on national guidelines [21]. With 95% confidence and a 5% margin of error, we applied the standard prevalence formula:

$$n=Z2\times P\times(1-P)d2$$

where n is the required sample size, Z is the 95% confidence Z-score (1.96), P is the expected prevalence (0.20), and d is the margin of error (0.05). This yielded a minimum sample size of 246 (5% margin of error) participants.

Although our planned sample size was 246 participants, the final sample comprised 198 participants. Precision is therefore slightly lower than planned; 95% CIs are reported for key estimates. For example, a 20% prevalence has a 95% margin of error of about 5.6% at n=198 versus 5.0% at n=246. However, if the web application for cognitive screening is good enough, the results should still be satisfactory. Moreover, in Thailand, medical services at general hospitals in the peripheral areas of the country are overloaded, causing investigators to be unable to complete sufficient case collection for this study. Data analysis was performed using SPSS version 30 (IBM Corp) and MedCalc Statistical Software (MedCalc Software Ltd). Descriptive statistics summarized baseline demographic and clinical characteristics. To assess validity, we computed Spearman correlation coefficients between digital test scores and conventional paper-and-pencil scores (TMSE and the CDT) because the data were not normally distributed. We considered $P<.05$ statistically significant. We also conducted receiver operating characteristic (ROC) curve analyses to evaluate the application's ability to distinguish cognitively impaired individuals from cognitively normal control participants. We calculated the area under the curve (AUC) for each digital test domain to quantify overall diagnostic performance.

Results

Phase II: Participant Demographics and Initial Validation

This phase evaluated concordance and diagnostic accuracy between traditional paper-and-pencil assessments and web-based application tests. A total of 77 participants were categorized into 3 groups: those with dementia (n=21), those with mild cognitive impairment (n=23), and cognitively normal control participants (n=33). The dementia group had a significantly higher mean age than the other groups (P=.003). All cognitive and functional measures differed significantly between groups (all P<.001). The dementia group had the lowest performance on paper-and-pencil TMSE, e-TMSE, clock drawing, and verbal fluency, and they had the highest ADL, cognitive IADL, and IQCODE-16 scores, indicating greater functional impairment and cognitive decline. Demographic characteristics for all 77 participants are shown in Table 1.

XSL•FO
RenderX

Table . Participant demographics and clinical characteristics, initial phase (n=77).

Participant characteristics	Dementia (n=21)	MCI[a] (n=23)	Normal cognition (n=33)	Total (n=77)	*P* value[b]
Age (years), median (IQR)	73.0 (69.5-79.0)	69.0 (63.0-75.0)	66.0 (56.5-72.0)	69.0 (63.0-75.0)	.003
Female, n (%)	14 (66.7)	15 (65.2)	22 (66.7)	51 (66.2)	.99
Formal education, n (%)					
Primary	13 (61.9)	6 (26.1)	5 (15.2)	24 (31.2)	
Secondary	1 (4.8)	4 (17.4)	4 (12.1)	9 (11.7)	
Vocational	3 (14.3)	3 (13)	5 (15.2)	11 (14.3)	
Bachelor's degree	3 (14.3)	8 (34.8)	15 (45.5)	26 (33.8)	
Postgraduate	1 (4.8)	2 (8.7)	4 (12.1)	7 (9.1)	
e-TMSE[c], median (IQR)	21.0 (16.5-23.0)	25.0 (24.0-28.0)	28.0 (25.5-29.0)	25.0 (23.0-28.0)	<.001
Clox[d] test, median (IQR)	7.0 (1.5-8.0)	10.0 (9.0-10.0)	10.0 (9.5-10.0)	10.0 (7.0-10.0)	<.001
Verbal fluency test, mean (SD)	11.11 (4.99)	17.61 (6.80)	19.12 (4.61)	16.48 (6.35)	<.001
Basic ADLs[e], median (IQR)	7.0 (3.5-9.0)	0.0 (0.0-1.0)	0.0 (0.0-0.0)	0.0 (0.0-2.5)	<.001
Cognitive IADLs[f], median (IQR)	6.0 (3.5-8.5)	0.0 (0.0-1.0)	0.0 (0.0-0.0)	0.0 (0.0-2.0)	<.001
IQCODE-16[g], median (IQR)	4.0 (3.8-4.6)	3.4 (3.2-3.7)	3.1 (3.0-3.3)	3.3 (3.1-3.8)	<.001
TMSE[h], median (IQR)	22.0 (21.0-25.0)	28.0 (26.0-29.0)	28.0 (27.0-29.5)	27.0 (24.0-29.0)	<.001
Clock drawing test (paper), median (IQR)	6.0 (5.0-10.0)	10.0 (10.0-10.0)	10.0 (9.3-10.0)	10.0 (7.0-10.0)	<.001

[a]MCI: mild cognitive impairment.

[b]Between-group comparisons.

[c]e-TMSE: electronic Thai Mental State Examination.

[d]Clox: digitized clock drawing.

[e]ADLs: activities of daily living.

[f]IADLs: instrumental activities of daily living.

[g]IQCODE-16: 16-item Informant Questionnaire on Cognitive Decline in the Elderly.

[h]TMSE: Thai Mental State Examination.

Digital and Standard Cognitive Screening Tests

The paper-and-pencil TMSE and e-TMSE had a strong, statistically significant correlation (Spearman ρ=0.837, *P*<.001), indicating that the digital version reflects the traditional assessment. The mean paper-and-pencil TMSE score was 26.04 (SD 3.87; median 27), while the mean e-TMSE score was 24.91 (SD 4.18; median 25). Subdomain analysis showed comparable performance across most cognitive components. The score distribution indicated a population predominantly comprising cognitively intact individuals and those with mild cognitive impairment. Details are provided in Table 2.

XSL•FO
RenderX

Table . Total and domain scores on the paper-and-pencil Thai Mental State Examination (TMSE) versus those of the electronic TMSE (e-TMSE).

TMSE components	Paper-and-pencil TMSE score, mean (SD; median)	e-TMSE score, mean (SD; median)	Correlation, Spearman ρ[a]	*P* value
Total TMSE score	26.04 (3.87; 27)	24.91 (4.18; 25)	0.837	.001
Orientation	5.11 (1.21; 6)	5.11 (1.21; 6)	0.818	.001
Registration	2.96 (0.19; 3)	2.88 (0.32; 3)	−0.102	.39
Attention	4.79 (0.92; 5)	4.75 (0.96; 5)	0.689	.001
Calculation	2.20 (0.97; 2)	2.03 (1.00; 2)	0.993	.001
Language	9.58 (1.01; 9)	8.88 (1.28; 9)	0.533	.001
Recall	2.80 (0.40; 3)	2.68 (0.58; 3)	0.691	.001

[a]Correlation of the total scores between versions: Spearman ρ=0.837 (*P*<.001).

Phase III: Full Study Implementation

A total of 198 participants were classified into 3 diagnostic groups: dementia (n=44), mild cognitive impairment (n=58), and cognitively normal (n=96). This total was lower than the planned 246 from the prevalence-based calculation. The median age differed significantly across the groups (*P*<.001). Participants with dementia were older (median 73, IQR 69.25 - 79 years) and had the highest proportion with only a primary education (30/44, 68.2%). Cognitive and functional measures differed significantly across all assessments (*P*<.001). Median e-TMSE scores were lowest in the dementia group (18.5, IQR 12.0 - 22.0) and highest in the cognitively normal group (26.0, IQR 23.0 - 28.0). The scores for the Clox test and the category VFT showed a similar pattern. Table 3 presents the baseline participant characteristics for the full implementation sample.

XSL•FO
RenderX

Table . Baseline demographics and clinical characteristics, full implementation sample (n=198).

Participant characteristics	Dementia (n=44)	MCI[a] (n=58)	Normal cognition (n=96)	Total (n=198)	*P* value
Age (years), median (IQR)	73 (69.25-79)	69.5 (62.75-75)	65.5 (52-74)	69.5 (59-76)	<.001
Female, n (%)	32 (72.7)	36 (62.1)	69 (71.9)	137 (69.2)	.38
Formal education level, n (%)					
Primary	30 (68.2)	19 (32.8)	36 (37.5)	85 (42.9)	
Secondary	4 (9.1)	8 (13.8)	13 (13.5)	25 (12.6)	
Vocational	3 (6.8)	5 (8.6)	9 (9.4)	17 (8.6)	
Bachelor's degree	5 (11.4)	22 (37.9)	30 (31.3)	57 (28.8)	
Postgraduate/graduate	2 (4.5)	4 (6.9)	8 (8.3)	14 (7.1)	
e-TMSE[b], median (IQR)	18.5 (12.0-22.0)	25.0 (23.0-28.0)	26.0 (23.0-28.0)	24 (20-28)	<.001
Clox[c] test, median (IQR)	3.0 (0.0-8.0)	9.0 (7.0-10.0)	10.0 (8.0-10.0)	9 (6-10)	<.001
Verbal fluency test, median (IQR)	7.0 (5.0-12.8)	14.0 (9.0-19.0)	17.5 (13.0-17.5)	15 (9-19)	<.001
Basic ADLs[d], median (IQR)	8.0 (3.0-14.8)	0.0 (0.0-2.0)	0.0 (0.0-0.0)	0 (0-3)	<.001
Cognitive IADLs[e], median (IQR)	8.0 (3.0-12.0)	0.0 (0.0-2.0)	0.0 (0.0-0.0)	0 (0-2)	<.001
IQCODE-16[f], median (IQR)	4.3 (3.9-4.9)	3.3 (3.1-3.8)	3.3 (3.0-3.3)	3.4 (3.1-3.9)	<.001

[a]MCI: mild cognitive impairment.

[b]e-TMSE: electronic Thai Mental State Examination.

[c]Clox: digitized clock drawing.

[d]ADLs: activities of daily living.

[e]IADLs: instrumental activities of daily living.

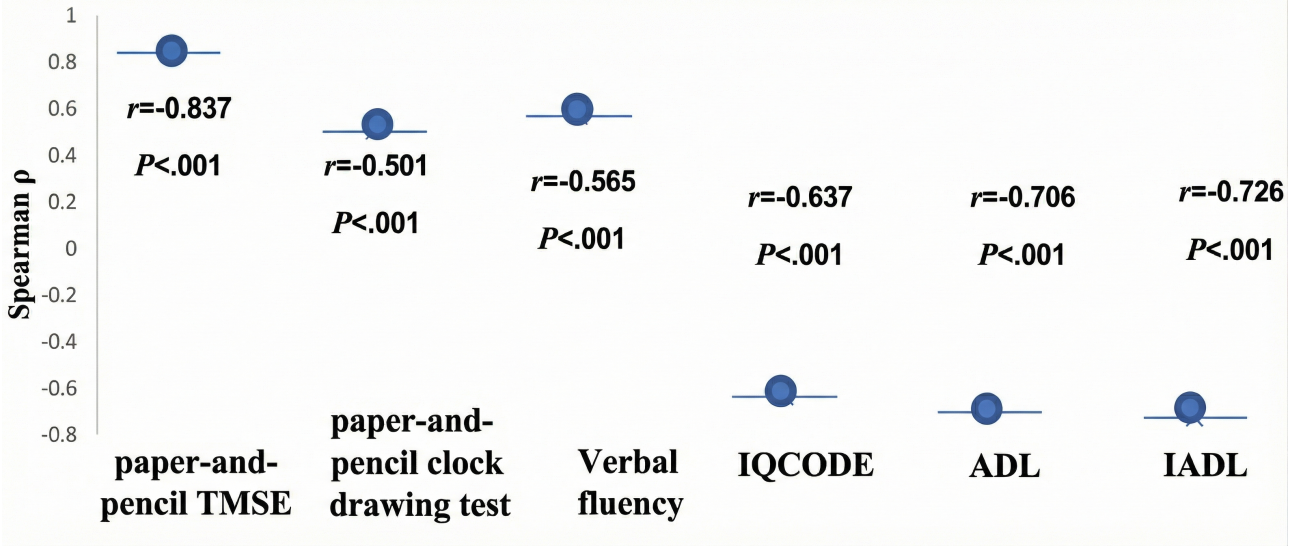
[f]IQCODE-16: 16-item Informant Questionnaire on Cognitive Decline in the Elderly.

Digital Cognitive Test: e-TMSE and Cognitive Assessment

Figure 6 illustrates the correlation matrix showing associations between the e-TMSE and related cognitive and functional assessments. Positive correlations were observed between the e-TMSE and the paper-and-pencil CDT (r=0.501) and the category VFT (r=0.565), indicating convergence with other cognitive measures. In contrast, negative correlations were found between the e-TMSE and informant-reported or functional measures, including the IQCODE-16 (r=−0.637) and the ADL score (r=−0.706). These patterns reflect inverse relationships between cognitive performance and functional independence.

XSL•FO
RenderX

Figure 6. Correlation matrix showing the correlations between the electronic Thai Mental State Examination (e-TMSE) and other cognitive and functional measures, including the Clox test as the paper-and-pencil clock drawing test, the category verbal fluency test, 16-item Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE-16), activities of daily living (ADLs), and cognitive instrumental activities of daily living (IADLs).

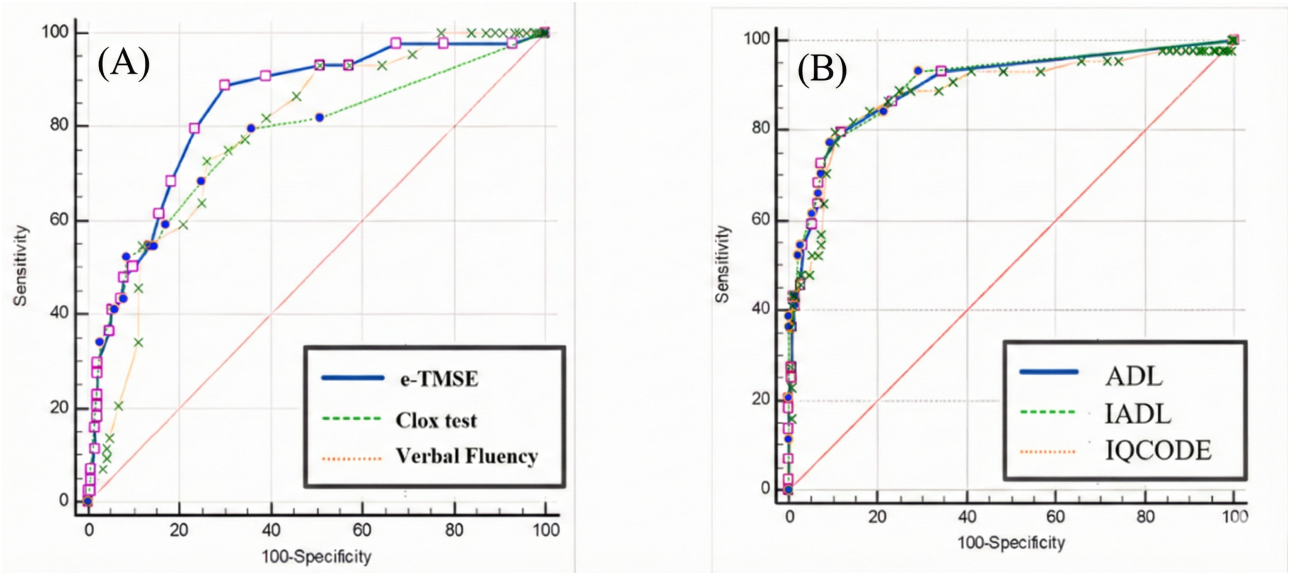


Analysis of Cognitive and Functional Assessments for Dementia Screening

We conducted ROC analyses to evaluate the web-based application's diagnostic accuracy for distinguishing individuals with normal cognition from those with major neurocognitive disorder. Figure 7 shows the ROC curves, and Table 4 summarizes the performance metrics. The AUC values were as follows: e-TMSE, 0.842; Clox test, 0.774; category VFT, 0.785; IQCODE-16, 0.884; ADL, 0.901; and cognitive IADLs, 0.906. Among these, cognitive IADLs and ADL had the highest discriminatory power. The e-TMSE yielded an optimal cutoff score of ≤23, with a sensitivity of 88.64% and a specificity of 70.13%. The details are provided in Table 4.

Of the 198 participants, 24 (12.1%) were 80 years old or older (9 dementia, 3 mild cognitive impairment, 12 controls); 16 (66.7%) of these 24 individuals had only a primary education. Approximately 5% to 10% of the total cohort (primarily older adults and those with lower education) required assistance but completed the assessment satisfactorily with medical supervision. Errors were most common in drawing tasks for which finger or stylus use was advised. Dropout rates were consistent across sites (5% - 10%). Satisfaction scores (Table S1 in Multimedia Appendix 1) and domain variations that were likely attributable to sociodemographic factors (Tables S2-S3 in Multimedia Appendix 1) are provided in Multimedia Appendix 1.

Figure 7. Receiver operating characteristic (ROC) curves for digital measures distinguishing major neurocognitive disorder: (A) electronic Thai Mental State Examination (e-TMSE), Clox test, and category verbal fluency test and (B) activites of daily living (ADLs), cognitive instrumental activities of daily living (IADLs), and 16-item Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE-16). The area under the curve (AUC) values, 95% CIs, and operating thresholds appear in Table 4.



XSL•FO
RenderX

Table . Diagnostic performance of *Healthy Brain Test* digital measures for major neurocognitive disorder.

Test	Maximum score	AUROC[a] (95% CI)	Optimal cutoff	Sensitivity (%)	Specificity (%)	Positive LR[b]	Negative LR
e-TMSE[c]	30	0.842 (0.784–0.890)	≤23	88.64	70.13	2.97	0.16
Clox[d] test	10	0.774 (0.710-0.830)	≤7	68.17	75.32	2.76	0.42
Verbal fluency test	—[e]	0.785 (0.721-0.840)	≤11	70.73	74.03	2.8	0.37
IQCODE-16[f]	—	0.884 (0.831-0.925)	>3.562	88.64	75.32	3.59	0.15
Basic ADLs[g]	—	0.901 (0.850-0.939)	>3	72.73	92.86	10.18	0.29
Cognitive IADLs[h]	—	0.906 (0.857-0.943)	>1	84.09	75.32	3.59	0.15

[a]AUROC: area under the receiver operating characteristic curve.

[b]LR: likelihood ratio.

[c]e-TMSE: electronic Thai Mental State Examination.

[d]Clox: digitalized clock drawing.

[e]Not applicable.

[f]IQCODE-16: 16-item Informant Questionnaire on Cognitive Decline in the Elderly.

[g]ADLs: activities of daily living.

[h]IADLs: instrumental activities of daily living.

Discussion

Principal Findings

Web-based cognitive screening applications are a promising approach for early detection of cognitive impairment, particularly in resource-limited settings such as Thailand. We evaluated the feasibility and diagnostic accuracy of digitizing established neuropsychological assessments—the TMSE, clock drawing, verbal fluency, IQCODE-16, and ADL functional evaluations—into a web-based format. Our findings suggest that these web-based tools can reliably identify cognitive impairment and provide screening performance comparable to traditional paper-and-pencil methods.

The TMSE, culturally adapted from the MMSE, is widely used in Thailand as a standard cognitive screening tool for dementia. It evaluates orientation, attention, memory, language, and visuospatial abilities, and prior validation studies reported 85% to 90% sensitivity and 80% to 90% specificity for detecting dementia in Thai populations [15].

This study developed and evaluated a web-based TMSE adaptation (e-TMSE) that preserved the original assessment structure and scoring criteria. Using a diagnostic cutoff of ≤23, the e-TMSE achieved an AUC of 0.84 (95% CI 0.78 - 0.89), indicating a high level of discriminative ability for identifying dementia. Sensitivity and specificity were 88.64% and 70.13%, respectively, and were comparable to those of the traditional TMSE. In addition, e-TMSE scores correlated significantly with paper-and-pencil TMSE scores (r=0.837, P<.001), supporting the concurrent validity of the digital format. Subdomain analyses showed comparable diagnostic performance across most cognitive components, although slight discrepancies were noted in registration and language. These discrepancies may reflect user interaction with the digital interface, particularly among participants with limited touchscreen experience.

The CDT is a brief, effective screening tool for visuoconstructional skills and executive function, which are commonly impaired in dementia. Prior studies reported 76% to 85% sensitivity and 81% to 86% specificity for detecting dementia [22]. In our web-based adaptation, the Clox test achieved an AUC of 0.77 (95% CI 0.71 - 0.83), with a cutoff of ≤7; sensitivity was 68.17%, and specificity was 75.32%. Although these values are somewhat lower than some traditional CDT benchmarks, they indicate moderate diagnostic accuracy and support digital screening utility. Performance differences may reflect technological literacy, touchscreen familiarity, and fine motor skills, which can influence task execution and response accuracy in older adults.

The VFT assesses executive function and lexical retrieval. Our implementation used the category (semantic) variant and resulted in an AUC of 0.79 (95% CI 0.72 - 0.84) with a cutoff of ≤11. Sensitivity was 70.73%, and specificity was 74.03%, indicating moderate diagnostic accuracy. Semantic fluency, particularly animal naming, is especially useful for early detection of cognitive impairment. Monsch et al [23] reported 100% sensitivity and 92.5% specificity for semantic fluency in distinguishing patients with Alzheimer disease from control participants who had normal cognition. Similarly, a meta-analysis by Henry et al [24] found that semantic fluency declines early in Alzheimer disease and vascular cognitive impairment, reinforcing its role as an early diagnostic marker. Furthermore, studies with individuals with mild cognitive impairment have shown that semantic fluency can be a sensitive

XSL•FO
RenderX

and efficient screening tool for early stages of dementia [18,25]. In addition to its diagnostic value, the VFT is simple to administer, cost-effective, and readily adaptable to digital formats. These features make it suitable for large-scale screening and for use in primary care or community-based settings.

The IQCODE provides complementary dementia screening by capturing caregiver-reported longitudinal cognitive change and by circumventing patient insight deficits, which are common in dementia. Previous research reported 80% to 90% sensitivity and specificity for distinguishing dementia from normal aging [26,27]. In our study, the short-form IQCODE (IQCODE-16) showed a significant negative correlation with e-TMSE scores (r=−0.637, P<.001), consistent with earlier IQCODE and TMSE findings (r=−0.679, P<.001) [19]. This consistency supports the concurrent validity of the IQCODE-16 as a reliable screening instrument. IQCODE-16 achieved an AUC of 0.88 (95% CI 0.83 - 0.93), with a cutoff of >3.562. Sensitivity was 88.64%, and specificity was 75.32%, with a marginally higher AUC than e-TMSE (AUC=0.84). These findings highlight clinical relevance, particularly in community-based or resource-limited settings where cultural, educational, or language barriers may affect performance-based testing.

Functional assessments, especially ADL, provide essential context regarding the real-world impact of cognitive decline on daily functioning. Although not diagnostic in isolation, declines in ADL—particularly cognitive IADLs—are predictive of dementia progression. Our cognitive IADL component achieved an AUC of 0.91 (95% CI 0.86 - 0.94), with a cutoff of >1; sensitivity was 84.09%, and specificity was 75.32%. The ADL measure achieved an AUC of 0.90 (95% CI 0.85 - 0.94), with a cutoff of >3; sensitivity was 72.73%, and specificity was 92.86%.

Adapting neuropsychological and functional assessments to web-based platforms, such as *Healthy Brain Test*, represents meaningful progress in dementia screening. These platforms can expand access in resource-limited regions and facilitate earlier detection of cognitive decline. In Thailand, geographic and socioeconomic disparities limit access to specialized care, and a culturally adapted, validated digital tool may help bridge screening gaps in community-based settings. In our findings, dropouts were most common among the oldest-old and those with lower education levels, primarily due to unfamiliarity with digital technology. Issues also arose from misaligned user expectations; for instance, participants often wished to erase specific lines during drawing tasks, whereas the application only permitted clearing the entire page to redraw. Additional factors contributing to dropout included poor user interface design (eg, small font sizes, login difficulties) and physical health limitations (eg, visual impairment, hand tremors). Our study recruited individuals with diverse baseline characteristics regarding age, education, and diagnostic status (dementia, mild cognitive impairment, and normal controls). It is well-established that economic conditions differ regionally in Thailand; generally, the central, southern, and northern regions are wealthier and report better health outcomes than the northeast. These socioeconomic and cultural variations may explain some of the observed differences in cognitive performance, both within the normal group and across the entire cohort at each study site.

The global adoption of mobile and web-based cognitive screening tools has expanded, offering promising alternatives to traditional paper-and-pencil assessments across clinical and community settings. The commercially available *BrainTest* reported 71% sensitivity and 90% specificity compared with comprehensive neuropsychological assessments [28]. Similarly, *M-CogScore*, which includes memory, attention, and executive tasks, achieved an AUC of 0.85 (95% CI 0.80 - 0.91) and outperformed the MMSE-2 (AUC=0.78, 95% CI 0.72 - 0.84) for distinguishing individuals with normal cognition from those with impairment [29]. In parallel, mobile cognitive screening tools designed for smartphones and tablets provide additional advantages over traditional modalities [30].

Nevertheless, important limitations persist. Previous studies [30,31] noted that most self-administered electronic tools lack rigorous psychometric validation. Practical barriers include touchscreen usability challenges among older adults with limited digital experience, as well as broader issues of digital literacy. In addition, digital tools may be less able to detect subtle behavioral, emotional, or motivational cues that inform clinical assessment. These limitations are especially pronounced in resource-limited settings, where disparities in technology access and health services can further hinder effectiveness.

In Thailand, electronic cognitive assessments have been used for over a decade. A 2018 validation study of the computerized TMSE [32] demonstrated a strong correlation with the standard pen-and-paper TMSE. The percentage of agreement between the two methods was 92.5%, with a Kappa coefficient of 0.85 (P<.001). The duration of the standard pen-and-paper TMSE was slightly shorter than that of the computerized TMSE (7.31 min vs 7.97 min; 95% CI −1.159 to −0.175; P=.09). More recently, a study on the Thai electronic MoCA [33] reported a high correlation for total scores, with a concordance correlation coefficient of 0.919 and a mean difference of −0.204 (95% CI −6.311 to 5.904). All cognitive subdomain scores showed moderate to high concordance correlation coefficient values exceeding 0.4. Furthermore, functional assessment is crucial for determining the stage of cognitive performance. Therefore, it is essential to evaluate both cognitive function and ADLs concurrently. ADLs can be self-reported or observed by family members who can provide feedback to clinicians. In the context of Thai culture, where extended families are common, family members play a vital role in observing daily functioning and providing this necessary feedback.

Thailand's cultural, linguistic, and socioeconomic factors require careful consideration when implementing web-based cognitive screening tools. Our study showed that *Healthy Brain Test*, a web-based cognitive assessment application, is clinically useful for the early detection of dementia. The tool demonstrated good psychometric performance and was validated across diverse Thai populations in multiple regions, strengthening the applicability of the findings.

Several limitations warrant consideration. The sample comprised mostly individuals with normal cognition, mild cognitive impairment, and mild dementia, with limited evaluation of

XSL•FO
RenderX

performance in moderate to severe dementia. Most notably, the final sample size was lower than the planned target of 246, which may have reduced the precision of estimates. However, a 20% prevalence of dementia with a 95% CI and with a margin of error of 5.6% led to the sample size of 198 participants in our study. Assessments were conducted under professional supervision to ensure understanding in the local Thai pronunciation, which may not reflect real-world, unsupervised use. However, caregivers can possibly play this role by helping individual older adults at home. Further limitations of our study primarily concern digital literacy and accessibility. First, a lack of digital experience was evident; the majority of older Thai participants were unfamiliar with drawing on touchscreens (eg, tablets or smartphones). However, they were able to perform tasks correctly after receiving instruction. In the context of unsupervised community testing, co-resident children or visiting family members can play a crucial role in encouraging older adults to adopt new technologies. For example, the Thai medical system successfully implemented telemedicine during the COVID-19 era, and hospitals continue to use it for patients in remote areas. A key factor in this success is the assistance provided by younger family members during assessments. Second, regarding physical and sensory impairments, Thailand's universal health coverage scheme allows individuals with vision or hearing problems to receive fully reimbursed treatments, such as cataract extraction with lens implantation and hearing aids. Third, regarding the need for technological support, older adults with cognitive impairment may require caregivers to set up or assist with web-based assessments. Despite this support, 5% to 10% of our cohort were unable to complete the assessment. Fourth, socioeconomic factors such as cost, income level, and access to reliable internet and devices may pose barriers. To address this, the Thai government has set a goal to establish free internet access zones in every village, and digital devices imported from China are becoming increasingly affordable. Last, technological anxiety remains a common issue among older adults.

Although our study included a relatively small sample size, it was conducted across 4 distinct provinces: Bangkok (Siriraj and Vajira Hospitals), Phitsanulok (Northern Thailand), Nakhon Phanom (Northeastern Thailand, bordering the Mekong River and Laos), and Phatthalung (Southern Thailand). We recently signed a Memorandum of Understanding with the Department of Medical Services, Ministry of Public Health, to deploy this web-based cognitive screening test nationwide over the next 3 years. Based on the insights gained from this widespread implementation, we anticipate developing a second iteration or an abbreviated version of the assessment.

Future research should evaluate *Healthy Brain Test* in unsupervised community-based settings, explore integration with digital health infrastructure, and enable longitudinal tracking of cognitive change. Implementation should be supported by health care professionals to ensure appropriate interpretation of results and timely referral for further evaluation or intervention.

Conclusion

This study demonstrates that the *Healthy Brain Test* web-based application is valid, efficient, and accessible for early detection of cognitive impairment. The battery comprises the following assessments: e-TMSE, CDT, category VFT, IQCODE-16, and ADL. The digital test scores correlated strongly with traditional paper-and-pencil assessments, supporting the reliability and clinical applicability of our web-based cognitive screening application. This simple, rapid tool is well-suited for identifying cognitive impairment in primary care and community-based settings. Further studies should validate performance in unsupervised and diverse populations and support integration into routine cognitive health screening protocols.

Acknowledgments

The authors thank all participants and caregivers for their contribution.

Funding

This study was partially supported by the Siriraj Foundation, Bangkok, Thailand.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request. Identifiable data are not publicly available due to privacy restrictions.

Authors' Contributions

VS conceptualized and designed the study; acquired funding; supervised the project; and led project administration, visualization, and manuscript review and editing. NW contributed to data curation, methodology development, supervision, validation, and supported formal analysis and project administration. SH conducted the formal analysis with support from NW. NP, SS, KT, and PM carried out the investigation and contributed to project administration and visualization. CR provided resources and participated in visualization and manuscript review. NP drafted the original manuscript under the supervision of VS. All authors reviewed, revised, and approved the final version of the manuscript.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Satisfaction scores and site-specific findings.
[DOCX File, 31 KB - humanfactors_v13i1e85454_app1.docx]

References

1. Jagust W. Vulnerable neural systems and the borderland of brain aging and neurodegeneration. Neuron 2013 Jan 23;77(2):219-234. [doi: 10.1016/j.neuron.2013.01.002] [Medline: 23352159]
2. Oh H, Madison C, Haight TJ, Markley C, Jagust WJ. Effects of age and β-amyloid on cognitive changes in normal elderly people. Neurobiol Aging 2012 Dec;33(12):2746-2755. [doi: 10.1016/j.neurobiolaging.2012.02.008] [Medline: 22429886]
3. Jitapunkul S, Kunanusont C, Phoolcharoen W, Suriyawongpaisal P. Prevalence estimation of dementia among Thai elderly: a national survey. J Med Assoc Thai 2001 Apr;84(4):461-467. [Medline: 11460954]
4. Lampit A, Hallock H, Valenzuela M. Computerized cognitive training in cognitively healthy older adults: a systematic review and meta-analysis of effect modifiers. PLoS Med 2014 Nov;11(11):e1001756. [doi: 10.1371/journal.pmed.1001756] [Medline: 25405755]
5. Jelcic N, Agostini M, Meneghello F, et al. Feasibility and efficacy of cognitive telerehabilitation in early Alzheimer's disease: a pilot study. Clin Interv Aging 2014;9:1605-1611. [doi: 10.2147/CIA.S68145] [Medline: 25284993]
6. Dodge HH, Zhu J, Mattek NC, et al. Web‐enabled conversational interactions as a method to improve cognitive functions: results of a 6‐week randomized controlled trial. A&D Transl Res & Clin Interv 2015 Jun;1(1):1-12. [doi: 10.1016/j.trci.2015.01.001]
7. Lee AR, Csipke E, Yates L, et al. A web-based self-management app for living well with dementia: user-centered development study. JMIR Hum Factors 2023 Feb 24;10:e40785. [doi: 10.2196/40785] [Medline: 36826978]
8. Tyack C, Camic PM. Touchscreen interventions and the well-being of people with dementia and caregivers: a systematic review. Int Psychogeriatr 2017 Aug;29(8):1261-1280. [doi: 10.1017/S1041610217000667] [Medline: 28446258]
9. Watanyulertsakul E, Khanthong P, Deeka B. Dementia U-Care: comprehensive cognitive screening application for seniors. ECTI-CIT Transactions 2024;18(2):207-221 [FREE Full text] [doi: 10.37936/ecti-cit.2024182.255205]
10. Vorasut S, Tantatsanawong P. A prototype development of dementia screening on smart phone. Journal of the Thai Medical Informatics Association 2023;1:27-35 [FREE Full text]
11. Yangyuensathaporn B, Chansaengpetch S, Jongsawadipatana A, Muangpaisan W. Validation of an application-based cognitive screening test for older Thai adults. Dement Geriatr Cogn Disord 2025;54(4):214-225. [doi: 10.1159/000543309] [Medline: 39746342]
12. Creavin ST, Wisniewski S, Noel-Storr AH, et al. Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. Cochrane Database Syst Rev 2016 Jan 13;2016(1):CD011145. [doi: 10.1002/14651858.CD011145.pub2] [Medline: 26760674]
13. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
14. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (5th ed., text rev.): American Psychiatric Association Publishing; 2022.
15. Train the Brain Forum Committee (Thailand). Thai Mental State Examination (TMSE). Siriraj Hosp Gaz 1993;45(6):359-374 [FREE Full text]
16. Aprahamian I, Martinelli JE, Neri AL, Yassuda MS. The Clock Drawing Test: a review of its accuracy in screening for dementia. Dement Neuropsychol 2009;3(2):74-81. [doi: 10.1590/S1980-57642009DN30200002] [Medline: 29213615]
17. Henderson SK, Peterson KA, Patterson K, Lambon Ralph MA, Rowe JB. Verbal fluency tests assess global cognitive status but have limited diagnostic differentiation: evidence from a large-scale examination of six neurodegenerative diseases. Brain Commun 2023;5(2):fcad042. [doi: 10.1093/braincomms/fcad042] [Medline: 36910418]
18. McDonnell M, Dill L, Panos S, et al. Verbal fluency as a screening tool for mild cognitive impairment. Int Psychogeriatr 2020 Sep;32(9):1055-1062. [doi: 10.1017/S1041610219000644] [Medline: 31258101]
19. Senanarong V, Assavisaraporn S, Sivasiriyanonds N, et al. The IQCODE: an alternative screening test for dementia for low educated Thai elderly. J Med Assoc Thai 2001 May;84(5):648-655. [Medline: 11560213]
20. Senanarong V, Harnphadungkit K, Prayoonwiwat N, et al. A new measurement of activities of daily living for Thai elderly with dementia. Int Psychogeriatr 2003 Jun;15(2):135-148. [doi: 10.1017/s1041610203008822] [Medline: 14620072]
21. Department of Medical Services (Thailand), Prasat Neurological Institute. Clinical Practice Guidelines: Dementia: Prasat Neurological Institute; 2014. URL: https://thaineuronurse.com/wp-content/uploads/2022/CNPG/10- .pdf
22. Shulman KI, Pushkar Gold D, Cohen CA, Zucchero CA. Clock‐drawing and dementia in the community: a longitudinal study. Int J Geriat Psychiatry 1993 Jun;8(6):487-496 [FREE Full text] [doi: 10.1002/gps.930080606]
23. Monsch AU, Bondi MW, Butters N, Salmon DP, Katzman R, Thal LJ. Comparisons of verbal fluency tasks in the detection of dementia of the Alzheimer type. Arch Neurol 1992 Dec;49(12):1253-1258. [doi: 10.1001/archneur.1992.00530360051017] [Medline: 1449404]

XSL•FO
RenderX

24. Henry JD, Crawford JR, Phillips LH. Verbal fluency performance in dementia of the Alzheimer's type: a meta-analysis. Neuropsychologia 2004;42(9):1212-1222. [doi: 10.1016/j.neuropsychologia.2004.02.001] [Medline: 15178173]
25. Cintoli S, Favilli L, Morganti R, Siciliano G, Ceravolo R, Tognoni G. Verbal fluency patterns associated with the amnestic conversion from mild cognitive impairment to dementia. Sci Rep 2024 Jan 23;14(1):2029. [doi: 10.1038/s41598-024-52562-x] [Medline: 38263387]
26. Jorm AF. The Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): a review. Int Psychogeriatr 2004 Sep;16(3):275-293. [doi: 10.1017/s1041610204000390] [Medline: 15559753]
27. Quinn TJ, Fearon P, Noel-Storr AH, Young C, McShane R, Stott DJ. Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) for the detection of dementia within community dwelling populations. Cochrane Database Syst Rev 2021 Jul 19;7(7):CD010079. [doi: 10.1002/14651858.CD010079.pub3] [Medline: 34278562]
28. Scharre DW, Chang SI, Nagaraja HN, Vrettos NE, Bornstein RA. Digitally translated Self-Administered Gerocognitive Examination (eSAGE): relationship with its validated paper version, neuropsychological evaluations, and clinical assessments. Alzheimers Res Ther 2017 Jun 27;9(1):44. [doi: 10.1186/s13195-017-0269-3] [Medline: 28655351]
29. Alim-Marvasti A, Kuleindiren N, Harvey K, et al. Validation of a rapid remote digital test for impaired cognition using clinical dementia rating and mini-mental state examination: an observational research study. Front Digit Health 2022;4:1029810. [doi: 10.3389/fdgth.2022.1029810] [Medline: 36620187]
30. Zorluoglu G, Kamasak ME, Tavacioglu L, Ozanar PO. A mobile application for cognitive screening of dementia. Comput Methods Programs Biomed 2015 Feb;118(2):252-262. [doi: 10.1016/j.cmpb.2014.11.004] [Medline: 25481217]
31. Charalambous AP, Pye A, Yeung WK, et al. Tools for app- and web-based self-testing of cognitive impairment: systematic search and evaluation. J Med Internet Res 2020 Jan 17;22(1):e14551. [doi: 10.2196/14551] [Medline: 31951218]
32. Dharmasaroja PA, Ratanakorn D, Nidhinandana S, Charernboon T. Comparison of computerized and standard cognitive test in Thai memory clinic. J Neurosci Rural Pract 2018;9(1):140-142. [doi: 10.4103/jnrp.jnrp_373_17] [Medline: 29456359]
33. Hemrungrojn S, Amrapala A, Kasemsantitham A, et al. The validity of the Thai electronic version of the Montreal Cognitive Assessment (MoCA). Alzheimer's & Dementia 2023 Dec;19(S20) [FREE Full text] [doi: 10.1002/alz.079254]

Abbreviations

ADL: activities of daily living
AUC: area under the curve
CDT: clock drawing test
DSM-V: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
e-TMSE: electronic Thai Mental State Examination
IADL: instrumental activities of daily living
IQCODE: Informant Questionnaire on Cognitive Decline in the Elderly
MMSE: Mini-Mental State Examination
MoCA: Montreal Cognitive Assessment
ROC: receiver operating characteristic
TMSE: Thai Mental State Examination
VFT: verbal fluency test

Edited by A Kushniruk; submitted 08.Oct.2025; peer-reviewed by H Subramaniam, J Odeniya; accepted 02.Jan.2026; published 19.Feb.2026.

Please cite as:
Piyaamornpan N, Srisuwannanukorn S, Tangthamrongthanawat K, Mekhasingharak P, Rattanabannakit C, Hunnangkul S, Wongkom N, Senanarong V
Web-Based Application for Cognitive and Functional Assessments in Dementia Screening: Mixed Methods, User-Centered Development Approach
JMIR Hum Factors 2026;13:e85454
URL: https://humanfactors.jmir.org/2026/1/e85454
doi:10.2196/85454

© Nattaporn Piyaamornpan, Suwat Srisuwannanukorn, Kosit Tangthamrongthanawat, Pornsawan Mekhasingharak, Chatchawan Rattanabannakit, Saowalak Hunnangkul, Natthamon Wongkom, Vorapun Senanarong. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 19.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete

XSL•FO
RenderX

bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Identifying Biomarkers Using a Portable, Home-Based Eye-Tracking System to Predict Short-Term Visual Fatigue Deterioration: Prospective Observational Feasibility Study

Fan Song[1,2*], MD; Guangyu Li[3*], MS; Jian Zhang[4], MS; Zhen Tian[1,2], MD; Mingge Li[5], MD; Mengying Lai[1], MS; Mingguang He[1,2,6], MD, PhD; Yanxian Chen[1,2], MD, PhD

[1]School of Optometry, Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[2]Research Centre for SHARP Vision, Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[3]Lab of Brain-inspired Algorithm and Model, Guangdong Institute of Intelligence Science and Technology, Zhuhai, Guangdong, China

[4]State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, Guangdong, China

[5]Department of Ophthalmology, Shenzhen Nanshan People's Hospital, Shenzhen, Guangdong, China

[6]The Center for Eye and Vision Research, Hong Kong, China

[*]these authors contributed equally

Corresponding Author:
Yanxian Chen, MD, PhD
School of Optometry
Hong Kong Polytechnic University
11 Yuk Choi Rd, Hung Hom
Hong Kong,
China (Hong Kong)
Phone: 852 2766 6111
Email: yan-xian.chen@polyu.edu.hk

Abstract

Background: The escalating prevalence of screen-related eye fatigue has become a health burden in the digital era worldwide, yet routine monitoring relies largely on subjective reports. This underscores the urgent need for clinically applicable, objective diagnostic solutions. Ocular metrics provide an objective method to assess computer vision syndrome, or asthenopia.

Objective: This study aimed to develop and evaluate an integrated at-home system for predicting short-term deteriorated asthenopia using objective ocular metrics. This system classifies the short-term risk level for practical monitoring and automatically generates a session report that summarizes metrics to complement symptom-based evaluation.

Methods: We developed EyeFatigue Tracker, an integrated at-home system delivered via a desktop app, comprising a head-mounted device to record binocular infrared eye videos, a deep learning (DL) model to extract ocular metrics, and a machine learning (ML) classifier to estimate asthenopia risk. The DL model, trained on an in-house dataset, segments the palpebral fissure, pupil, and iris from recorded videos to derive ocular metrics. To build the prediction model, participants were recruited to complete a 1-hour computer gameplay session. Changes in the Computer Vision Syndrome Questionnaire (CVS-Q) scores served as the primary outcome measure to classify participants into deteriorated and nondeteriorated asthenopia groups. Metrics showing significant between-group differences were used as inputs for four ML models, including support vector machine (SVM), decision tree, extreme gradient boosting (XGBoost), and random forest, to identify deteriorated asthenopia. Model performance was evaluated with fivefold cross-validation.

Results: This study enrolled 38 participants aged 19-31 (mean 24.8, SD 3.11) years. Following visual tasks, participants' CVS-Q scores were higher compared to baseline values (mean 9.21, SD 4.57, vs mean 6.76, SD 3.76; $P<.001$). Alongside the critical flicker fusion frequency (CFF), nine key features were selected as predictive indicators, with the top five reflecting fissure length variability (variance, coefficient of variation, and SD), average blink duration, and pupil size variability (coefficient of variation). Most ML models exhibited high discriminative ability, with the random forest achieving the best overall performance (mean accuracy 0.720, SD 0.035; mean area under the receiver operating characteristic curve 0.850, 95% CI 0.830-0.860).

Conclusions: The findings highlight the potential of objective indicators in identifying individuals at risk for asthenopia following computer gameplay. The ML models using ocular biomarkers identified in this study achieved plausible discriminative ability

XSL•FO
RenderX

in detecting deteriorated asthenopia. EyeFatigue Tracker functions as an integrated, at-home system that produces a risk level prediction and a concise session report, supporting early detection and informing preventive care in real-world settings.

(JMIR Hum Factors 2026;13:e84479) doi:10.2196/84479

KEYWORDS

eye tracker; asthenopia; blink; eye movement; prediction; home monitoring; pupil dynamics; visual task

Introduction

Background

With advancements in technology and the widespread use of digital devices, complaints about asthenopia, or computer vision syndrome, have rapidly increased [1]. The worldwide prevalence of asthenopia has reached 60 million, with an additional 1 million new cases reported annually [2]. Asthenopia encompasses a range of subjective symptoms, including eye fatigue, eye swelling, dryness, blurred vision, and orbital pain [3,4], often along with musculoskeletal discomfort, such as neck pain [5]; neuropsychological symptoms, such as headaches and dizziness [6]; and general malaise [7]. Although typically transient, the symptoms can be recurrent and persistent [8]. These symptoms affect not only visual comfort but also overall health and work productivity [2]. The increasing prevalence of visual fatigue has thus become a significant burden on both individuals and society, emphasizing the urgent need for effective diagnostic and management strategies.

Existing Assessment Methods and Research Gap

Despite its growing impact, there is no unified standard for diagnosing asthenopia. Existing assessment approaches can broadly be categorized into subjective symptom questionnaires and objective physiological or ocular video–based methods. Most studies rely on subjective questionnaires, with the Computer Vision Syndrome Questionnaire (CVS-Q) being a widely used tool for assessing the frequency and severity of asthenopia symptoms [3]. Such questionnaires are free to use in large populations, but participants still need to complete multiple items and calculate total scores. In routine or remote use, they may also have questions about specific items that cannot be answered promptly by eye care professionals, which makes these tools less practical for frequent at-home monitoring. In addition, objective physiological measures, such as the critical flicker fusion frequency (CFF), have been explored [9,10]. Although the CFF offers a quantifiable approach, its reliability in assessing asthenopia and its potential as a surrogate measure remain controversial [11]. Other objective metrics, such as blinking patterns, eye movement metrics, accommodative function, and pupil dynamics, offer additional quantitative insights into asthenopia [12-15]. These physiological metrics can provide reproducible endpoints but usually require dedicated clinical equipment and in-clinic testing, which reduces their feasibility for routine, at-home monitoring. For example, accommodative function reflects the eye's ability to focus but requires specialized equipment and clinical expertise, limiting its practicality. Similarly, eye-tracking devices, such as the Tobii system, can capture eye movement data, including fixation and saccade metrics [16]. However, these devices are often costly and require sophisticated analysis to interpret the data accurately [17]. More recently, some groups have combined blink, pupil, and eye movement features into portable platforms, such as smartphone-based eye movement–recording apps and virtual reality (VR) headsets with integrated eye tracking, to assess visual behavior and visual fatigue in everyday digital tasks [18,19]. Beyond these approaches, video-based deep learning (DL) methods have been used to recognize ocular anatomy in other domains, such as cataract surgery and biometrics [20-22]. Previous studies have used explicit image feature extraction steps to improve pupil and iris recognition performance in cataract surgery videos or eye gaze–tracking tasks [23-25]. Clinical studies have also applied objective pupillometry to quantify how cataract surgery affects pupil size and constriction dynamics [26]. Overall, these systems show that it is possible to automatically extract eye structure information from videos in controlled, task-specific settings, but they were not developed to monitor visual fatigue or short-term changes in symptoms during everyday screen use. In the context of visual fatigue, DL methods have been used to extract blinking patterns (eg, frequency and duration) and pupil dynamics (eg, accommodation speed) from video data to assess visual fatigue [14,27]. Although these approaches show promise, no system has yet succeeded in integrating these multiple objective metrics into a portable, convenient, and cost-effective solution for the comprehensive evaluation of asthenopia.

Objectives and Contributions of This Study

To address these challenges, we proposed a novel portable, at-home system that uses DL algorithms to simultaneously measure blinking patterns, eye movement metrics, and pupil parameters. By integrating these metrics, the device may provide a more comprehensive and objective assessment of asthenopia. Specifically, this study (1) developed an integrated at-home system that records binocular infrared (IR) eye videos and derives anatomy-aware ocular metrics, (2) identified ocular biomarkers associated with short-term deteriorated asthenopia using a standardized computer gameplay protocol, and (3) evaluated multiple machine learning (ML) models to classify at-risk individuals based on these biomarkers. This approach aims to offer a practical and accessible tool for identifying individuals prone to visual fatigue in real-world settings.

Methods

Ethical Considerations

Ethical approval was obtained before the study began from the Institutional Review Board (IRB) of the Hong Kong Polytechnic University (reference #HSEARS20240425002). The study adhered to the tenets of the Declaration of Helsinki. Written informed consent was obtained from all participants prior to study participation, and participants were informed that they

XSL•FO
RenderX

could withdraw from the study at any time without a penalty. All study data were deidentified before analysis and were stored securely to protect participant privacy and confidentiality. Participants received HK $350 (~US $45) as compensation for their participation. Potentially identifiable participant images were included only where scientifically necessary, and separate written consent for publication was obtained from the individuals concerned.

Data Collection

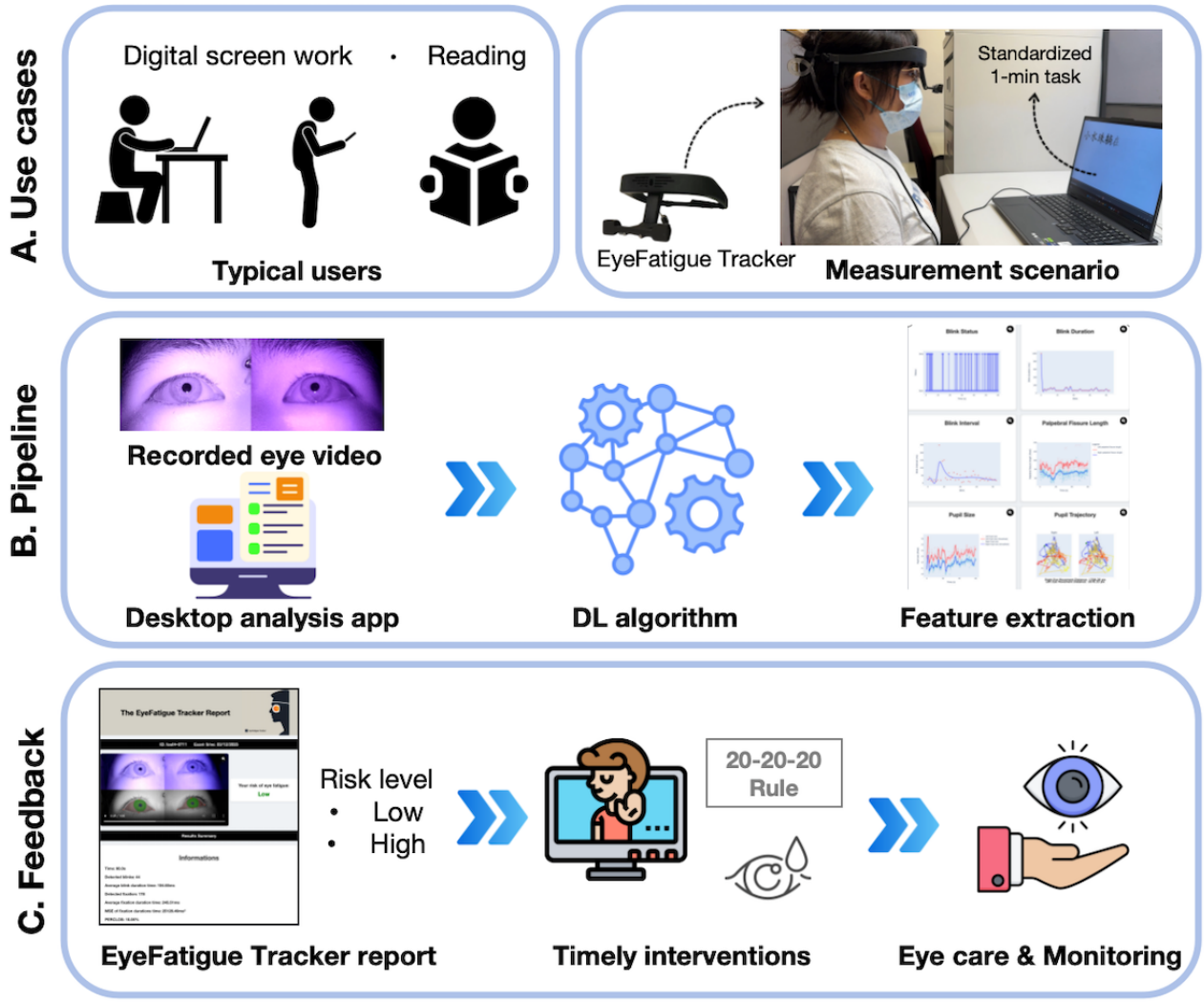
This prospective observational study was conducted between July and August 2024. Participants aged 16-35 years were included. Exclusion criteria encompassed individuals with the presence of diseases that may cause eye pain or headaches, such as glaucoma, keratitis, strabismus, conjunctivitis, iridocyclitis, ocular trauma, or other self-reported conditions; color vision deficiency; illiteracy; or cognitive dysfunction.

Sociodemographic information and general health history were collected from the participants. Following this, each participant engaged in four consecutive 15-minute computer games on a laptop (2560×1600 resolution, 240 Hz refresh rate) for a total of 60 minutes of continuous screen use. The games consisted of simple but engaging tasks (maze navigation [28], block stacking [29], obstacle avoidance [30], and sports-themed minigames [31]) to maintain attention, while providing sustained visual load (see Figure S1 in Multimedia Appendix 1). The total duration and task design were chosen to create short-term, measurable changes in visual fatigue, while preserving compliance, in line with prior work showing that approximately 1 hour of continuous digital device use can significantly aggravate visual fatigue [32]. Each participant first completed a 1-minute text-reading task (Multimedia Appendix 2), then performed the four game sessions, and immediately completed the same 1-minute text-reading task. During each text-reading task, the participant wore the EyeFatigue Tracker device, which recorded their ocular videos. The task consisted of three consecutive 20-second segments with progressively reduced contrast, smaller font size, and more complex text, increasing in difficulty from easy to moderate to difficult in a fixed order, so that we could examine how ocular metrics changed as visual and cognitive load increased within a brief standardized trial. This stepwise change in contrast, font size, and text complexity is consistent with prior work showing that smaller text, lower contrast, and more demanding reading conditions can increase visual fatigue and mental workload in digital reading tasks [33,34]. For analysis, these three segments approximately corresponded to the frame ranges 200-600 (easy), 600-1200 (moderate), and 1200-1800 (difficult) after discarding the first 200 frames.

Data collected included habitual visual acuity (HVA), tear film break-up time (TBUT), CVS-Q scores [3], Stanford Sleepiness Scale (SSS) scores [35], Karolinska Sleepiness Scale (KSS) scores [36], Fatigue Assessment Scale (FAS) scores [37], CFF values, and ocular metrics (blinking patterns, eye movement metrics, and pupil parameters). The CVS-Q served as the primary subjective measure of asthenopia and was used to define deteriorated and nondeteriorated asthenopia groups. The SSS and the KSS captured current and recent sleepiness, and the FAS assessed overall fatigue, allowing us to examine whether changes in ocular metrics are specifically related to visual fatigue rather than general tiredness. The CFF was included as a classical psychophysical index of visual fatigue to compare with the proposed ocular video–based metrics. The HVA was measured by the Early Treatment Diabetic Retinopathy Study (ETDRS) chart at 4 m. The TBUT was recorded as the time interval between the last blink and the first tear film dry spot following fluorescein instillation, using a slit lamp. The severity of asthenopia, current sleepiness, recent (last 10 minutes) sleepiness, and overall fatigue were assessed using the CVS-Q, SSS, KSS, and FAS questionnaires, respectively, with higher scores indicating greater severity. CFF values were measured using Handy Flicker HF-III (Neitz Instrument), and the average value of ascending and descending thresholds were recorded. A decrease in CFF values reflects increased visual fatigue. Ocular objective metrics were extracted from the 1-minute video recorded by the EyeFatigue Tracker device. The overall experimental setting and visual task timeline are illustrated in Figure 1A (measurement scenario).

XSL•FO
RenderX

Figure 1. Overall flowchart of the EyeFatigue Tracker system and study workflow. (A) Typical users and standardized measurement scenario during a 1-minute reading task with the head-mounted EyeFatigue Tracker. (B) Analysis pipeline: Infrared eye videos are processed by a mask R-CNN–based segmentation model to obtain masks of the palpebral fissure, iris, and pupil. From these masks, objective ocular metrics are extracted, including blink parameters, pupil dynamics, and eye movement metrics. Statistically significant metrics are then used as inputs to ML classifiers (SVM, decision tree, random forest, and XGBoost) to estimate the short-term risk for deteriorated asthenopia (low vs high). (C) Feedback module: The desktop app displays the predicted risk level, provides a brief session report summarizing key ocular metrics, and can be used for home-based eye care, longitudinal self-monitoring, and telemedicine follow-up. DL: deep learning; ML: machine learning; mask R-CNN: mask region-based convolutional neural network; SVM: support vector machine; XGBoost: extreme gradient boosting.



Proposed EyeFatigue Tracker System

System Overview

Figure 1 presents the EyeFatigue Tracker workflow used to estimate the risk of asthenopia deterioration. The EyeFatigue Tracker system classifies each 1-minute recording into a risk level (low/high) for asthenopia deterioration during at-home use. The device consists of a lightweight head-mounted binocular IR camera module, paired with a desktop analysis app, mounted on an adjustable plastic frame, which can be worn over habitual spectacles at a normal working distance from a computer screen. It connects to a standard laptop via a single USB cable and does not require a chinrest or dedicated laboratory setup, making it feasible for use in home and office environments. The hardware was assembled from off-the-shelf components to keep material costs low. Two IR cameras were used to capture videos of both eyes at a frame rate of 30 frames per second (fps). We used 940 nm near-infrared (NIR) illumination, together with an 850 nm long-pass IR filter (cut-on at 850 nm) mounted in front of the camera lens to suppress visible light from the display/ambient illumination and reduce glare-related artifacts. This rate was chosen as a practical trade-off between temporal resolution and computational cost: Blinks are brief (about 100-150 ms), so markedly lower frame rates (eg, 10-15 fps) may miss the full closing and reopening phases and distort blink duration, blink interval, and percentage of eyelid closure (PERCLOS), whereas much higher frame rates would substantially increase data size and processing time without a clear benefit for our target task. Participants wore EyeFatigue Tracker while performing a 1-minute text-reading task, during which a video containing 1800 frames (1 minute) was recorded for feature extraction. IR ocular videos were recorded at 3840×1080 pixels and 30 fps, and frames were downsampled to 480×135 pixels for training and 640×180 pixels for inference in the DL models.

XSL•FO
RenderX

Deep Learning–Based Eye Anatomy Recognition

We used two nonoverlapping datasets. Dataset A is an in-house eye video dataset from 100 subjects with different eye states (open, half-open, closed, left-right scanning), and we selected 1000 labeled eye images and used them exclusively to train and validate the segmentation model. Dataset B is an independent cohort recruited in this study (see the *Data Collection* section) and was used for feature extraction and predictive modeling.

Instance segmentation of ocular anatomy was implemented in Detectron2 [38] with a mask region-based convolutional neural network (mask R-CNN; residual network [ResNet]-50 backbone + feature pyramid network [FPN]), which we selected after a preliminary comparison with a U-Net–based model showed that the mask R-CNN produced more stable and accurate masks for small ocular structures (pupil boundary and palpebral fissure), while remaining feasible to train and run on a standard laptop. Prior to segmentation, IR frames underwent lightweight preprocessing to suppress overexposure and normalize intensity. Annotations were stored in Common Objects in Context (COCO)–style format and registered as Detectron2-compatible datasets. Segmentation outputs were grouped into three anatomical structures: palpebral fissure, pupil, and iris. All recordings were visually inspected, and videos with severe artifacts or segmentation failures were excluded to ensure that subsequent analyses primarily reflected true ocular behavior. The model was initialized with weights pretrained on the COCO instance segmentation task and then fine-tuned on Dataset A only (3000 iterations, batch size=2, learning rate=2.5×10^{4}). Due to graphics processing unit (GPU) memory constraints and the small-scale nature of the dataset, we used a batch size of two images per iteration, which is standard for a mask R-CNN in Detectron2 and provided stable convergence for this task. Training/validation/test splits were made at the participant level to avoid subject-level leakage. This model enables precise differentiation of various eye regions. We reported the Dice similarity coefficient (Dice) and intersection over union (IoU) in two forms: macro (unweighted mean across classes) and micro (global aggregate). Per structure metrics were macroaveraged across left and right eyes, whereas the microaveraged performance was obtained by aggregating all structures and images before computing the Dice, IoU, and COCO-style average precision (AP)/average recall (AR) using the Detectron2 COCOEvaluator. One-to-one instance matching was performed using the Hungarian assignment at IoU≥0.5. We also reported COCO metrics via Detectron2's evaluator, including the AP and AR at AP averaged from IoU 0.50 to 0.95 (AP@0.50-0.95), AP at IoU 0.50 (AP@0.50), and AR averaged from IoU 0.50 to 0.95 (AR@0.50-0.95). For clinical interpretability, metrics were also summarized after aggregating to eye (palpebral fissure), iris, and pupil.

Assessment Metric Computation and Risk Prediction

Based on the segmentation results, we extracted blink times, blink duration, blink interval, PERCLOS, fissure length, fixation number, fixation duration, eye movement distance, and pupil constriction speed. The average blink duration represented the mean length of each blink, while the blink interval referred to the duration between two consecutive blinks. Blink times were measured as the total number of blinks per minute. PERCLOS was defined as the proportion of time during which the eyes remained at least 80% closed. The fixation number represented the number of fixations per minute, while the average fixation duration represented the average length of each fixation. The eye movement distance was computed from the pupil centroid trajectory extracted from the pupil segmentation masks. For each frame, the pupil center was estimated, and the total movement was calculated as the cumulative Euclidean distance between consecutive centroids. Frames in which the pupil was not visible (eg, during eye closure) were automatically excluded from the calculation. The pupil constriction speed indicated the rate of pupil size change during constriction and dilation. To minimize the influence of the initial eye-opening movements in the video, the variations of ocular metrics were analyzed between frames 200 and 1800. For variability features, we calculated the SD, variance, and coefficient of variation of pupil size and fissure length within each frame range (200-600, 600-1200, 1200-1800) and across all frames (200-1800) separately for the pre- and posttask recordings. The difference in each variability metric was defined as posttask minus pretask and compared between groups. The corresponding change features were then calculated as the difference between post- and pretask summary statistics. Point-wise subtraction of raw signals was not performed.

Objective metrics were compared between the deteriorated and nondeteriorated asthenopia groups, defined using changes in CVS-Q scores. Participants were classified as having deteriorated asthenopia if their posttask CVS-Q score increased by ≥3 points compared to baseline; all others were assigned to the nondeteriorated asthenopia group. Statistically significant features were selected as inputs for the four ML models for predicting short-term deteriorated asthenopia. Statistical feature filtering (P<.05) was used to define candidate ocular biomarkers and was not re-estimated within each cross-validation fold; this design choice may introduce selection bias and was therefore noted as a limitation. This feature-filtering step was performed on the full dataset and served to define a fixed set of input features for subsequent ML analyses, rather than as a data-driven feature selection step within the model training pipeline. These ML models include both classical classifiers (SVM and decision tree) and modern tree-based ensemble methods (random forest and XGBoost), which have shown strong and robust performance in time-series prediction tasks in previous studies [39-41]. To improve robustness and mitigate variability due to dataset partitioning, an 80/20 training-test split was first applied, with the test set fully held out from all training procedures. Model training and evaluation were then conducted on the training set using stratified fivefold cross-validation, and performance was assessed using the area under the receiver operating characteristic curve (AUROC). ML models were trained on statistically significant features (P<.05) after z-score normalization, with synthetic minority oversampling technique (SMOTE) oversampling applied exclusively within the training data of each cross-validation fold. Before oversampling, the dataset contained 42 fatigued and 34 nonfatigued samples; after SMOTE, the training data in each fold were balanced to 42 samples per class, while validation and test sets retained the original class distribution. Hyperparameters for the SVM,

XSL•FO
RenderX

decision tree, random forest, and XGBoost were optimized using stratified fivefold grid search cross-validation on the training data. Hyperparameter tuning and cross-validation–based performance estimation were performed on the training set, while an independent test set was used only for final evaluation. Details of the model-specific evaluation parameters are provided in Table S1 in Multimedia Appendix 1. The final SVM used a radial basis function (RBF) kernel with principal component analysis (PCA) retaining 95% variance. Tree-based models (decision tree, random forest, XGBoost) were trained using their best-performing parameter combinations identified through the grid search. Performance metrics were derived from stratified fivefold cross-validation on the training set, while the independent test set was used only for an additional one-time validation of the selected models. After obtaining the AUROC for each fold, we computed the mean AUROC across all five folds to provide a more stable and generalized estimate of the model's discriminatory ability. Feature importance was assessed using the model with the best performance, where the relative contribution of each feature to predicting deteriorated asthenopia was quantified based on its importance scores. Finally, we integrated these functions into a desktop app: users can complete a 1-miute visual task and use the app to assess their risk level. The graphical user interface is designed for nonexpert users and provides fully automated processing with no need for manual parameter tuning, which supports convenient at-home use. All data processing, feature extraction, and model training were performed in Python using Detectron2 for segmentation and scikit-learn for ML analyses. The analysis pipeline ran on a Lenovo ThinkPad laptop (Intel i9-14900HX, 32 GB RAM, NVIDIA GeForce RTX 4060 Laptop GPU), achieving an end-to-end processing time of ~4 minutes per session, while segmentation models were trained on an NVIDIA A100 GPU server.

Data Analysis

The primary outcome measure was the presence of deteriorated asthenopia, identified by an increase of ≥3 points in the CVS-Q score following visual tasks. The sample size was calculated using G*Power 3.1 software [42], targeting a statistical power of 0.90 and a significance level (α) of .05, with an expected effect size (Cohen f=0.25). To account for a potential 20% dropout rate, the final sample size was adjusted to 38 participants. Data were presented as mean (SD) or mean (95% CI) for continuous variables with a normal distribution or as median (IQR) for nonnormally distributed variables. The Shapiro-Wilk test was used to assess data normality [43]. Paired *t* tests were used for normally distributed data [44]. Nonparametric tests, including the Mann-Whitney U test, McNemar's test, and Wilcoxon signed-rank test, were used to handle nonnormal or ordinal data [45]. The Wilcoxon signed-rank and McNemar's tests were used to measure within-participant changes in visual fatigue metrics over time points. The Mann-Whitney U test compared demographic characteristics, baseline characteristics, and changes in extracted objective metrics in subgroup analyses. All tests were two-tailed, with α=.05 considered statistically significant. The analysis was performed using Stata18.0 (StataCorp) [46].

Results

Participant Details

A total of 38 participants completed the study. Their ages ranged from 19 to 31 (mean 24.89, SD 3.11) years, and 19 (50%) were female. The mean TBUT was slightly shorter in the right eye than in the left eye (mean 6.48, SD 2.90, seconds vs mean 7.18, SD 2.72, seconds). The mean spherical equivalent was −2.23 (SD 2.30) diopters in the right eye and −1.93 (SD 2.32) diopters in the left eye. More than half of the participants (n=58%) had baseline asthenopia (CVS-Q score≥6). Participants reported a mean daily screen exposure of 9.45 (SD 3.37) hours. Their detailed baseline characteristics are summarized in Table 1.

XSL•FO
RenderX

Table 1. Baseline characteristics of participants (N=38).

Characteristics	Value
Age (years)	
Mean (SD)	24.89 (3.11)
Median (IQR)	26 (19-31)
Females, n (%)	19 (50)
TBUT[a] (seconds), mean (SD)	
Right eye	6.48 (2.90)
Left eye	7.18 (2.72)
Spherical equivalent (diopters), mean (SD)	
Right eye	–2.23 (2.30)
Left eye	–1.93 (2.32)
Asthenopia[b], n (%)	22 (58)
Screen exposure (hours/day), mean (SD)	9.45 (3.37)

[a]TBUT: tear film break-up time.

[b]Asthenopia was determined by a Computer Vision Syndrome Questionnaire (CVS-Q) score of ≥6.

Segmentation Performance of the EyeFatigue Tracker System

The model achieved an overall mean Dice of 93.3 (SD 0.4, 95% CI 92.9-93.7) and an overall mean IoU of 87.4 (SD 0.8, 95% CI 86.7-88.1). Among individual structures, the pupil showed the highest overlap, with a Dice of 94.0 (95% CI 93.7-94.4) and an IoU of 88.8 (95% CI 88.2-89.5), followed by the iris with a Dice of 86.6 (95% CI 85.1-88.0). The palpebral fissure was the most challenging, with a Dice of 75.0 (95% CI 71.4-78.6) and an IoU of 60.2 (95% CI 55.5-64.9). COCO metrics were strong at the overall level, with a mean AP@0.50 of 99.3 (SD 0.4), a mean AP@0.50-0.95 of 80.6 (SD 1.1), and a mean AR@0.50-0.95 of 84.7 (SD 1.0). At the structure level, the iris yielded the highest AP@0.50-0.95 of 87.4, and all structures maintained an AR@0.50-0.95 of >80%. The 95% CIs were narrow across metrics, indicating stable performance across folds. This pattern is consistent with the image characteristics: the pupil has a regular circular shape and strong contrast against the surrounding iris in IR videos, whereas the iris boundary and especially the palpebral fissure are less well defined, frequently affected by eyelashes and partial eyelid closure, and therefore inherently more difficult to delineate. Table 2 summarizes all objective metrics, and Figure 2 shows representative overlays for each structure and the final composite. Additional examples of segmentation performance under different eye-closure conditions are provided in Figure S2 in Multimedia Appendix 1.

Table 2. Segmentation performance by structure[a].

Structure	Dice[b], mean (SD)	IoU[c], mean (SD)	AP[d]@0.50, mean (SD)	AP@0.50-0.95, mean (SD)	AR[e]@0.50-0.95, mean (SD)
Pupil	94.0 (0.4)	88.8 (0.8)	99.2 (0.5)	75.9 (1.5)	82.1 (1.4)
Iris	86.6 (1.7)	76.4 (2.6)	99.7 (0.3)	87.4 (0.8)	90.7 (0.6)
Palpebral fissure	75.0 (4.1)	60.2 (5.3)	99.2 (1.0)	78.3 (1.3)	81.2 (1.4)
Overall[f]	93.3 (0.4)	87.4 (0.8)	99.3 (0.4)	80.6 (1.1)	84.7 (1.0)

[a]Per structure values are macroaveraged across left and right eyes.

[b]Dice: Dice similarity coefficient.

[c]IoU: intersection over union.

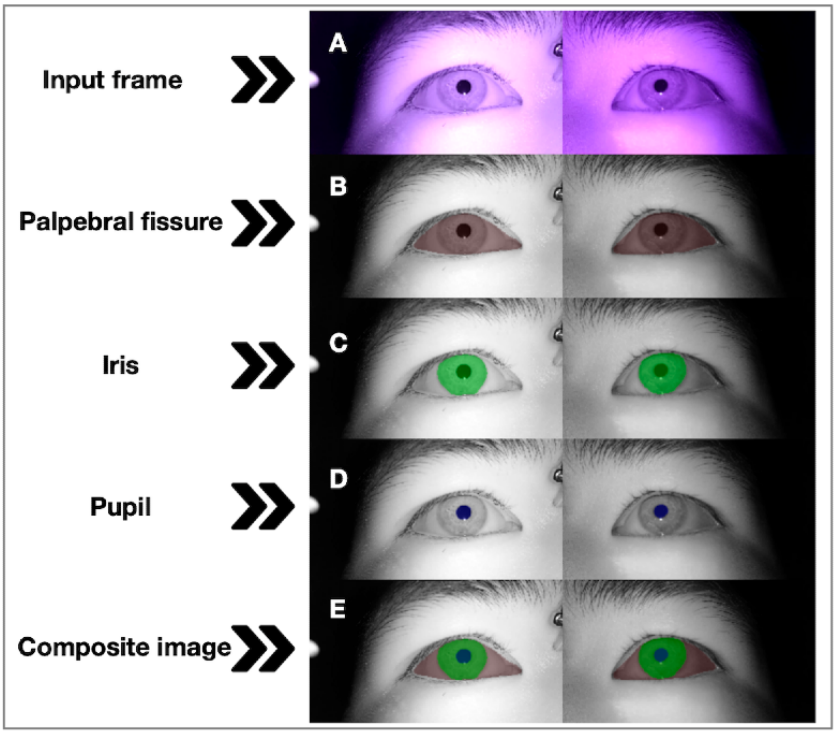
[d]AP: average precision.

[e]AR: average recall.

[f]Microaveraging across all classes.

XSL•FO
RenderX

Figure 2. Sample outputs of the segmentation system. (A) Input frame, (B) palpebral fissure mask (red), (C) iris mask (green), (D) pupil mask (blue), and (E) final composite overlay of all masks.



Comparison of Differences in Objective Metrics Before and After Visual Tasks

In addition to objective ocular metrics, subjective CVS-Q scores also showed significant worsening after visual tasks. Specifically, participants' CVS-Q scores were higher after visual tasks compared to baseline values (mean 9.21, SD 4.57, vs mean 6.76, SD 3.76, $P<.001$). The prevalence of participants in the higher-severity group (CVS-Q scores=13-18) increased from 3 (8%) of 38 participants at baseline to 13 (34%) participants, while those without asthenopia (CVS-Q scores<6) decreased from 16 (42%) of 38 participants to 12 (32%) participants. In addition, SSS, KSS, and FAS levels all slightly increased after visual tasks (SSS: median 1, IQR 0-2; KSS: median 1, IQR 0-2; FAS: median 1, IQR 0-4). Notably, the number of participants reporting the most common complaint, dryness, rose by 4 (95% CI –6% to 27%), while those reporting excessive blinking saw a substantial increase of 15 (95% CI 21%-58%). Together, these results show that the task led to more pronounced short-term aggravation of visual fatigue during the protocol (Table S2 in Multimedia Appendix 1). Changes in these ocular metrics (ie, blink patterns, eye movement metrics, and pupil parameters) and CFF values before and after visual tasks are presented in Table 3. All differences were calculated as post- minus pretask values, so negative numbers indicated a decrease after visual tasks. The deteriorated asthenopia group exhibited a marked reduction in the average blink duration (median –4.76, IQR –49.44 to 6.74, ms). Moreover, the nondeteriorated asthenopia group demonstrated an increase in blink times (median 4, IQR 0-8). Notably, these differences were statistically significant between the two groups. Figure 3 illustrates the changes in the average pupil size and palpebral fissure length over an 1800-frame interval before and after visual tasks. The deteriorated asthenopia group showed an increase in both parameters, while these metrics remained relatively stable in the nondeteriorated asthenopia group.

XSL•FO
RenderX

Figure 3. Changes in palpebral fissure length and pupil size before and after visual tasks. The x axis represents 1800 fpsframes per minute, while the y axis shows the average difference in pupil size values (A) and fissure length (B) before and after visual tasks. The blue line represents participants with deteriorated asthenopia, and the orange line represents participants without deteriorated asthenopia. Values below 0 indicate that the posttask value is smaller than the pretask value. For each time point, values were first computed per participant and averaged within each group for the pre- and posttask recordings. The posttask-pretask difference was then calculated as the difference between the two group-averaged trajectories. These curves are provided for visualization only and were not used for statistical feature extraction or ML model input. fps: frames per minute; ML: machine learning.

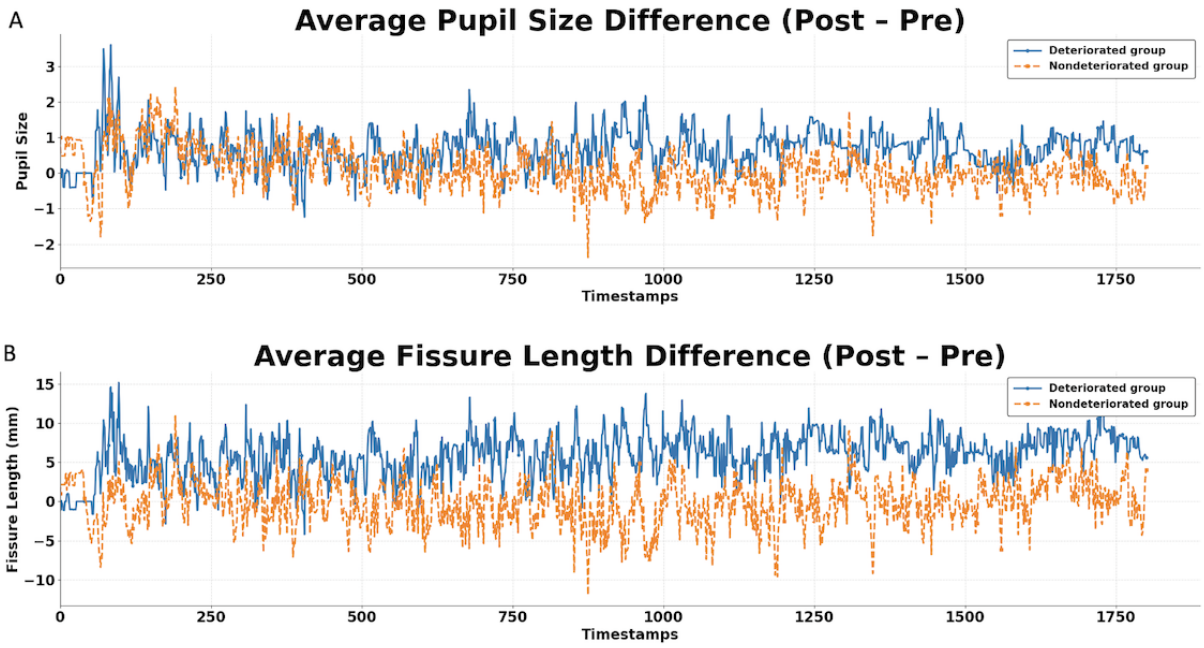


Table 3. Comparison of differences in blink patterns, eye movement metrics, pupil parameters, and CFF[a] values before and after visual tasks in participants with and without deteriorated asthenopia.[b]

Parameter difference	With deteriorated asthenopia	Without deteriorated asthenopia	*P* value
Blink times (per minute), median (IQR)	0 (–3 to 1)	4 (0 to 8)	.02
Average blink duration (ms), median (IQR)	–4.76 (–49.44 to 6.74)	12.21 (0 to 33.74)	.04
PERCLOS[c] (%), median (IQR)	–0.17 (–1.11 to 0.84)	0.56 (–0.83 to 2.78)	.10
Fixation number (per minute), median (IQR)	–5 (–14 to 10)	–7 (14 to 4)	.58
Average fixation duration (ms), median (IQR)	44.88 (–14.02 to 82.15)	–6.42 (–32.03 to 45.64)	.22
Average eye movement distance[d] (pixels), median (IQR)	36.49 (–21.60 to 75.03)	0.90 (–57.49 to 42.24)	.22
Average pupil diameter[d] (pixels), mean (95% CI)	0.15 (–0.15 to 0.45)	0.33 (–0.00 to 0.67)	.21
Maximum speed of pupil constriction[d] (pixels/second), median (IQR)	0.61 (–3.65 to 9.40)	0.77 (–8.27 to 9.23)	.71
Average speed of pupil constriction[d] (pixels/second), median (IQR)	–0.32 (–1.00 to 1.41)	0.14 (–0.96 to 0.54)	.53
CFF (Hz), mean (95% CI)	–2.97 (–3.96 to –1.98)	–2.14 (–2.96 to –1.32)	.09

[a]CFF: critical flicker fusion frequency.

[b]All values are expressed as post- minus pretask; negative values denote a reduction after visual tasks.

[c]PERCLOS: percentage of eyelid closure.

[d]Average value of both eyes.

Variability of Objective Metrics During the Reading Task

Tables 4-6 provide an overview of the changes in variability metrics, including pupil size, fissure length, blink duration, and blink interval, across the 1800-frame interval during the text-reading task in both groups, corresponding to the three increasing difficulty levels of the reading segments. Statistically significant differences were observed in several parameters. In the nondeteriorated asthenopia group, participants exhibited

XSL•FO
RenderX

significantly greater variability in pupil size across the 600-1200–frame interval, as indicated by a higher variance (median 0.64, IQR –0.27 to 1.81, vs median –0.22, IQR –1.25 to 0.68; *P*=.005) and coefficient of variation (median 3.77, IQR –3.78 to 10.48, vs median –3.31, IQR –9.62 to 0.61; *P*=.002). Similarly, fissure length variability was significantly higher in this frame interval in the nondeteriorated asthenopia group, with a higher SD (median 1.22, IQR –0.21 to 4.24, vs median –0.12, IQR –2.01 to 0.58; *P*=.007), variance (median 16.61, IQR –3.31 to 72.81, vs median –1.27, IQR –25.00 to 7.09; *P*=.003), and coefficient of variation (median 3.72, IQR 0.06-10.24, vs median –2.10, IQR –9.09 to 0.23; *P*<.001). In addition, across all frame intervals, the coefficient of variation for fissure length was significantly higher in the nondeteriorated asthenopia group (median –0.88, IQR –7.23 to 0.92, vs median 1.85, IQR –1.71 to 6.27; *P*=.006). No significant differences were found in blink duration or blink interval variability metrics (*P*>.05).

Table 4. Comparison of differences in variability metrics of pupil size before and after visual tasks in participants with and without deteriorated asthenopia.

Difference in variability metric and frame interval	With deteriorated asthenopia	Without deteriorated asthenopia	*P* value
SD, median (IQR)			
200-600	0.23 (–0.10 to 0.59)	0.16 (–0.53 to 0.62)	.2
600-1200	–0.11 (–0.46 to 0.30)	0.34 (–0.25 to 0.64)	.007
1200-1800	–0.03 (–0.29 to 0.25)	0.06 (–0.23 to 0.31)	.35
All	–0.01 (–0.17 to 0.19)	0.26 (–0.22 to 0.45)	.16
Variance, median (IQR)			
200-600	0.65 (–0.09 to 1.71)	0.39 (–1.36 to 2.03)	.21
600-1200	–0.22 (–1.25 to 0.68)	0.64 (–0.27 to 1.81)	.005
1200-1800	–0.02 (–0.60 to 0.57)	0.11 (–0.73 to 0.66)	.45
All	–0.02 (–0.20 to 0.45)	0.68 (–0.38 to 1.66)	.11
Coefficient of variation, median (IQR)			
200-600	2.86 (–4.20 to 6.84)	–0.23 (–10.25 to 9.11)	.21
600-1200	–3.31 (–9.62 to 0.61)	3.77 (–3.78 to 10.48)	.002
1200-1800	–1.98 (–7.78 to 1.91)	–0.22 (–4.08 to 3.15)	.23
All	–0.70 (–5.23 to 2.03)	2.18 (–3.22 to 5.59)	.07

Table 5. Comparison of differences in variability metrics of fissure length before and after visual tasks in participants with and without deteriorated asthenopia.

Difference in variability metric and frame interval	With deteriorated asthenopia	Without deteriorated asthenopia	*P* value
SD, median (IQR)			
200-600	0.96 (–0.13 to 2.65)	0.66 (–2.17 to 2.17)	.16
600-1200	–0.12 (–2.01 to 0.58)	1.22 (–0.21 to 4.24)	.00
1200-1800	0.13 (–1.25 to 2.51)	0.43 (–0.78 to 2.28)	.29
All	0.11 (–0.48 to 1.30)	1.01 (–0.88 to 2.61)	.28
Variance, median (IQR)			
200-600	16.10 (–2.18 to 39.78)	9.06 (–25.55 to 27.46)	.17
600-1200	–1.27 (–25.00 to 7.09)	16.61 (–3.31 to 72.81)	.003
1200-1800	1.57 (–11.38 to 23.01)	7.04 (–8.60 to 20.56)	.43
All	1.47 (–5.50 to 14.17)	13.71 (–11.46 to 39.05)	.18
Coefficient of variation, median (IQR)			
200-600	1.48 (–6.97 to 7.10)	1.84 (–8.67 to 8.17)	.96
600-1200	–2.10 (–9.09 to 0.23)	3.72 (0.06 to 10.24)	<.001
1200-1800	–2.79 (–9.74 to 2.38)	0.41 (–4.22 to 4.07)	.08
All	–0.88 (–7.23 to 0.92)	1.85 (–1.71 to 6.27)	.006

XSL•FO
RenderX

Table 6. Comparison of differences in variability metrics of blink duration and blink interval before and after visual tasks in participants with and without deteriorated asthenopia in all frame intervals.

Parameter and differences in variability metric	With deteriorated asthenopia	Without deteriorated asthenopia	*P* value
Blink duration			
SD, median (IQR)	–12.72 (–16.86 to 2.60)	–1.35 (–63.27 to 10.81)	.51
Variance, median (IQR)	–876.83 (–36720.3 to 92.59)	–61.73 (–6313.49 to 1000)	.14
Coefficient of variation, median (IQR)	–11.65 (–70.49 to 0.46)	–5.02 (–34.84 to 8.11)	.26
Blink interval			
SD, median (IQR)	–6.59 (–34.41 to 61.37)	–13.61 (–46.09 to 13.98)	.55
Variance, median (IQR)	–644.01 (–6785.52 to 13087.04)	–1833.61 (–5870.93 to 2526.15)	.63
Coefficient of variation, median (IQR)	–2.47 (–10.63 to 6.95)	–2.40 (–14.71 to 16.04]	.96

Risk Prediction for Deteriorated Asthenopia

As shown in Table 7, model performance was evaluated using accuracy, AUROC, the F_1-score, and balanced accuracy on the independent test set. Four ML models were built using ten selected features: average blink duration, coefficient of variation of fissure length (600-1200 frames), coefficient of variation of fissure length (200-1800 frames), variance of fissure length (600-1200 frames), SD of fissure length (600-1200 frames), coefficient of variation of pupil size (600-1200 frames), blink times, SD of pupil size (600-1200 frames), variance of pupil size (600-1200 frames), and the CFF. Among the four models, the random forest demonstrated the best overall performance, achieving the highest mean AUROC of 0.850 (95% CI 0.830-0.860), together with the highest accuracy (mean 0.720, SD 0.035), F_1-score (mean 0.771, SD 0.032), and balanced accuracy (mean 0.698, SD 0.042). XGBoost also achieved a high mean AUROC of 0.820 (95% CI 0.800-0.830), whereas the decision tree showed moderate performance across all metrics. In contrast, the SVM exhibited limited discriminative ability, with a balanced accuracy close to the chance level. Table 8 presents the contributions of each feature to the prediction of deteriorated asthenopia using the random forest. The importance scores indicate the relative influence of each feature on the model's predictive performance, with the average blink duration (score=0.194) and coefficient of variation of fissure length (600-1200 frames, score=0.119) being the most impactful features. Other variability-related features of fissure length and pupil size also contributed nonnegligibly to the model's predictions, whereas blink times showed a relatively lower contribution. Figure 4 shows the risk report generated by the deployed system. It predicts the deteriorated risk level (high/low), together with the summary metrics and time-series plots of objective metrics. In the prototype, the predicted risk level (low/high) is intended as the main output for lay users, whereas the detailed numerical and time-series indices are primarily provided for clinicians and researchers; future versions will include more user-friendly explanations and simplified summaries of these metrics. Enlarged versions of the individual plots are provided in Figure S3 in Multimedia Appendix 1.

Table 7. Prediction performance for detecting deteriorated asthenopia using ML[a] models.

Model	Accuracy, mean (SD)	F_1-score, mean (SD)	Balanced accuracy, mean (SD)	Mean AUROC[b] (95% CI)
Decision tree	0.688 (0.046)	0.726 (0.039)	0.682 (0.051)	0.75 (0.73, 0.77)
Random forest	0.720 (0.035)	0.771 (0.032)	0.698 (0.042)	0.85 (0.83, 0.86)
SVM[c]	0.555 (0.039)	0.692 (0.037)	0.511 (0.031)	0.59 (0.56, 0.63)
XGBoost[d]	0.677 (0.048)	0.741 (0.041)	0.661 (0.051)	0.82 (0.80, 0.83)

[a]ML: machine learning.

[b]AUROC: area under the receiver operating characteristic curve.

[c]SVM: support vector machine.

[d]XGBoost: extreme gradient boosting.

XSL•FO
RenderX

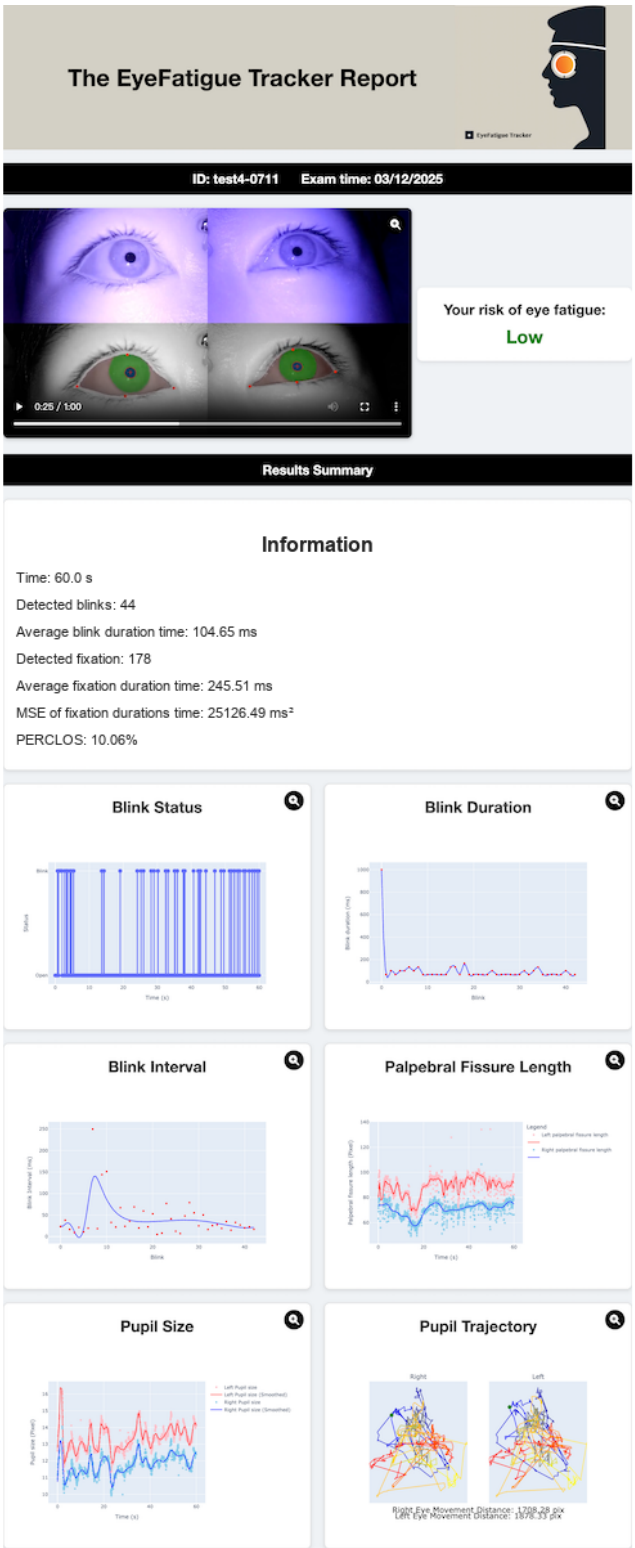
Table 8. Feature contributions to deteriorated asthenopia prediction in the random forest.

Rank	Feature	Importance score
1	Average blink duration	0.194
2	Coefficient of variation of fissure length (600-1200)	0.119
3	Coefficient of variation of fissure length (200-1800)	0.101
4	Variance of fissure length (600-1200)	0.096
5	SD of fissure length (600-1200)	0.096
6	Coefficient of variation of pupil size (600-1200)	0.087
7	CFF[a]	0.079
8	Variance of pupil size (600-1200)	0.078
9	SD of pupil size (600-1200)	0.077
10	Blink times	0.072

[a]CFF: critical flicker fusion frequency.

XSL•FO
RenderX

Figure 4. Example report generated by the EyeFatigue Tracker system. The app summarizes changes in ocular parameters during a single 1-minute standardized reading task. The panel on the top shows the IR ocular video with segmentation overlays, and the panel on the right displays the predicted risk level for eye fatigue. The summary block lists key metrics: total time, blink count, mean blink duration, fixation count, mean fixation duration, mean square error of fixation duration, and PERCLOS. The plots depict blink status, blink duration and interval, palpebral fissure length, pupil size, and pupil trajectory for both eyes across the 1-minute task. Values are illustrative. IR: infrared; PERCLOS: percentage of eye closure.



Discussion

Principal Findings

In this study, we developed and validated EyeFatigue Tracker, an at-home system that classifies the short-term risk for asthenopia deterioration using a standardized 1-minute reading task. The segmentation module performed reliably (overall Dice=93.3% and IoU=87.4%), providing stable masks for downstream analysis. Using these masks, the device objectively quantified ocular metrics related to eye fatigue, including blink patterns, eye movements, and pupil variations. Features

XSL•FO
RenderX

reflecting the intratask variability of palpebral fissure length and pupil size, together with blink measures and the CFF, were most informative for prediction. Across the four ML models, the random forest demonstrated the most balanced and robust overall performance on the test split, achieving the highest F_1-score (mean 0.771, SD 0.032), with a mean accuracy of 0.720 (SD 0.035) and a mean AUROC of 0.85 (95% CI 0.830-0.86). Taken together, these results suggest that brief, home-collected ocular signals can support practical monitoring of asthenopia in daily digital use and may offer new insights into objective biomarkers of visual fatigue.

Blink Behavior as a Compensatory Mechanism

Previous studies have reported a significant association between blink patterns and asthenopia [47,48]. In this study, the average blink duration and blink times emerged as key predictors of deteriorated asthenopia in the random forest, exhibiting notable differences between the deteriorated and nondeteriorated asthenopia groups. Specifically, the average blink duration increased in the nondeteriorated asthenopia group but remained unchanged in the deteriorated asthenopia group, consistent with prior research suggesting that prolonged blink duration serves as a compensatory mechanism to alleviate visual discomfort [27,49]. This may indicate that individuals experiencing deteriorated asthenopia fail to exhibit an adaptive increase in blink duration, potentially leading to inadequate tear film replenishment and worsened eye strain. Additionally, blink times also contributed to the prediction of deteriorated asthenopia, further supporting the idea that increased blink frequency is often a compensatory response to increasing eye fatigue [50]. These findings indicate that individuals experiencing deteriorated asthenopia may be less able to implement compensatory blink behavior effectively, exacerbating their visual discomfort and strain.

Ocular Variability as a Biomarker of Deteriorated Asthenopia

Ocular instability, reflected in fluctuations of fissure length (fissure height) and pupil size, emerged as a key indicator of deteriorated asthenopia. Among all features, the coefficient of variation of fissure length (600-1200 frames) ranked among the most influential predictors (0.119), together with its coefficient of variation in a later interval (200-1800 frames; 0.101), as well as its variance (0.096) and SD (0.096). Similarly, the coefficient of variation of pupil size (600-1200 frames; 0.087) also played a notable role in the prediction model. Notably, variability changes were higher in the nondeteriorated asthenopia group, suggesting an active and functional neuromuscular response to counteract fatigue. In contrast, the deteriorated asthenopia group exhibited lower variability changes, which may indicate fatigue-induced rigidity, where the eyelid and pupillary muscles fail to adjust dynamically, leading to sustained strain and discomfort. This distinction supports the idea that fissure height and pupil size variability reflect an individual's ability to adapt to visual fatigue.

Previous research suggests that eyelid position influences tear film stability, which may contribute to asthenopia [51]. Additionally, a variability change of fissure length may reflect the orbicularis oculi muscle's response to visual fatigue [52]. Cen et al [53] introduced pupil size as the most significant feature for SVM-based visual comfort prediction. Hopkinson [54] suggested that measuring oscillations in pupillary diameter is one of the simplest methods for assessing visual fatigue. Similarly, Geacintov and Peavler [55] observed that oscillation patterns in pupil size increase with visual fatigue, further supporting its relevance as a fatigue indicator. The variability metrics of pupil size during the specified intervals may reflect underlying oscillatory patterns associated with fatigue, aligning with prior evidence of its sensitivity to fatigue-related changes. These findings suggest that these metrics may therefore serve as potential biomarkers for distinguishing between individuals prone to deteriorated asthenopia [56]. Interestingly, variability metrics within the 600-1200–frame interval were more predictive than those spanning the entire reading task (200-1800 frames). This can be attributed to the task design, which progressively increased in difficulty across three stages. In the first stage (200-600 frames), the task was relatively easy, minimizing visual strain and limiting fluctuations of fissure length and pupil size. In the final stage (1200-1800 frames), the task became highly demanding, potentially triggering compensatory fixation strategies that suppressed variability in both groups. In contrast, the middle stage (600-1200 frames) represented a transition from low-to-moderate difficulty, requiring sustained attention and accommodation. This highlights the 600-1200–frame interval as a critical window where ocular instability becomes most pronounced, making it a sensitive marker for detecting individuals prone to deteriorated asthenopia.

Role of CFF and Fixation Parameters

Although no statistically significant difference in the CFF was observed between the deteriorated and nondeteriorated asthenopia groups, its inclusion in the predictive model was based on its recognized role in assessing eye fatigue susceptibility [10]. The CFF has long been established as a psychophysical marker of visual fatigue, as it reflects temporal resolution thresholds in the visual system that deteriorate under asthenopia-inducing conditions [57,58]. Although its individual contribution to the model was lower than most other ocular variability metrics, the CFF remains a valuable input, complementing blink- and pupil-related parameters to improve prediction accuracy.

In contrast, fixation parameters showed limited utility in distinguishing between the two groups in this study. Although Wang et al [13] reported that the average fixation duration increases and fixation numbers decrease as eye fatigue worsens, our findings aligned with their observations in both groups. Nevertheless, these changes were not statistically significant enough to differentiate the deteriorated and nondeteriorated asthenopia groups. One possible explanation is that visual fatigue did not deteriorate considerably during the computer gameplay process, reducing the sensitivity of fixation parameters as discriminative indicators in this context.

Clinical and Practical Implications of the EyeFatigue Tracker

Building on these observations, our findings contribute to advancing the diagnostics and prediction of deteriorated

XSL•FO
RenderX

asthenopia by addressing key limitations in previous studies. Previous research has often relied on head-mounted devices to analyze single-eye metrics, such as pupil accommodation speed [14], blink frequency [15], and eye closure duration [14], or has relied on eye trackers to extract pupil size and eye movement parameters, which are typically unable to directly capture blink pattern data [13]. In contrast, our EyeFatigue Tracker captures binocular ocular videos and a segmentation-driven pipeline to derive an integrated set of metrics for at-home use. Its portability and cost-effectiveness further enhance its potential for pragmatic monitoring in daily digital use. From a health care perspective, this at-home, session-based assessment could help detect early worsening of visual fatigue in high-risk users and provide objective measures to evaluate simple interventions. In telemedicine, the automatically generated reports can be shared with eye care professionals to complement questionnaires and guide follow-up. It may also be applied in clinical settings for the detection and follow-up of asthenopia, providing objective metrics alongside subjective assessments. Moreover, by introducing within-task variability measures as candidate indicators of deterioration, this study suggests potential gains in predictive performance and diagnostic utility.

Limitations and Future Directions

This study has several limitations that require consideration. First, the relatively small sample size may increase the risk of false-negative results when comparing the objective metrics between the deteriorated and nondeteriorated asthenopia groups and limits the generalizability of our findings. Future studies should recruit larger and more diverse populations to enhance statistical power and broader applicability. In addition, statistical feature filtering ($P<.05$) was used to define candidate ocular biomarkers and was not re-estimated within each cross-validation fold; this design choice may introduce selection bias and lead to optimistic performance estimates. Moreover, performance showed nonnegligible variability across cross-validation folds, underscoring uncertainty under the limited sample size and the need for external validation in independent cohorts. Second, no severe asthenopia cases were observed during the experiment. Including participants with severe symptoms in future studies would provide deeper insights into the relationship between objective metrics and severe visual fatigue. Third, the cohort was studied at a single site under a standardized protocol; subsequent research could incorporate diverse clinical or real-world settings (eg, workplaces with prolonged screen use or clinical contexts involving pre- and postsurgical patients) to better validate the predictive value of objective metrics in broader contexts. Fourth, future technical work could aim to improve the segmentation of challenging structures, such as the palpebral fissure. In addition, although a simple overexposure suppression step improved segmentation stability in our dataset, future versions of the system could integrate more advanced illumination enhancement methods to further increase robustness under challenging at-home lighting conditions. Fifth, we did not systematically test different video frame rates or learning rate schedules; future studies should compare alternative frame rates and apply learning rate–scheduling strategies (eg, step decay or plateau-based reduction) to further optimize the balance between accuracy and computational cost.

Conclusion

This study developed an at-home system that estimates the short-term risk for asthenopia deterioration after computer use and highlights objective biomarkers that may support earlier detection and prevention. Our findings indicate that brief, home-collected signals can enable practical monitoring and add objective evidence to symptom-based assessment. These conclusions should be interpreted in the light of the methodological and study limitations, including the modest sample size, the single-site design, and the use of short-term changes in CVS-Q scores as the reference outcome. Future work should validate generalizability in larger and more diverse cohorts, assess longitudinal stability in routine use, and evaluate performance in clinically severe cases.

Acknowledgments

We thank the InnoHK, a major initiative of the Hong Kong Special Administrative Region Government, for providing valuable support. The research work described in this paper was conducted in the Jockey Club Science, Technology, Engineering, and Mathematics Lab of Innovative Light Therapy for Eye Diseases, one of the labs funded by the Hong Kong Jockey Club Charities Trust. We also acknowledge the support of the Hong Kong Polytechnic University–Rohto Centre of Research Excellence for Eye Care.

Funding

This study was supported by the Hong Kong Polytechnic University–Rohto Centre of Research Excellence for Eye Care (P0046333) and the Start-up Fund for Research Assistant Professors under the Strategic Hiring Scheme (P0048638), with the sponsor or funding organization not being involved in the study’s design or implementation.

Authors' Contributions

YC and MH conceived the study and edited and reviewed the paper. FS, ML, and ML collected data. GL and ZT built the deep learning model. FS and GL conducted the literature search. FS, GL, YC, and JZ analyzed the data. FS and YC contributed to key data interpretation. FS wrote the manuscript. All authors have critically revised the manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data collection workflow and representative task materials, additional segmentation examples under different eye-closure conditions, enlarged views of the six analytic panels in the EyeFatigue Tracker risk report, parameter search grid, and comparison of subjective metrics before and after the visual tasks.
[DOCX File , 901 KB - humanfactors_v13i1e84479_app1.docx]

Multimedia Appendix 2
Standardized 1-minute text-reading task materials used before and after the game sessions.
[MP4 File (MP4 Video), 578 KB - humanfactors_v13i1e84479_app2.mp4]

References

1. Jaiswal S, Asper L, Long J, Lee A, Harrison K, Golebiowski B. Ocular and visual discomfort associated with smartphones, tablets and computers: what we do and do not know. Clin Exp Optom 2019 Sep;102(5):463-477 [FREE Full text] [doi: 10.1111/cxo.12851] [Medline: 30663136]
2. Ranasinghe P, Wathurapatha WS, Perera YS, Lamabadusuriya DA, Kulatunga S, Jayawardana N, et al. Computer vision syndrome among computer office workers in a developing country: an evaluation of prevalence and risk factors. BMC Res Notes 2016 Mar 09;9:150 [FREE Full text] [doi: 10.1186/s13104-016-1962-1] [Medline: 26956624]
3. Seguí MM, Cabrero-García J, Crespo A, Verdú J, Ronda E. A reliable and valid questionnaire was developed to measure computer vision syndrome at the workplace. J Clin Epidemiol 2015 Jun;68(6):662-673 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.01.015] [Medline: 25744132]
4. Ding Y, Guan H, Du K, Zhang Y, Wang Z, Shi Y. Asthenopia prevalence and vision impairment severity among students attending online classes in low-income areas of western China during the COVID-19 pandemic. Hong Kong Med J 2023 Apr 21;29(2):150-157 [FREE Full text] [doi: 10.12809/hkmj219864] [Medline: 37088700]
5. Demirayak B, Yılmaz Tugan B, Toprak M, Çinik R. Digital eye strain and its associated factors in children during the COVID-19 pandemic. Indian J Ophthalmol 2022 Mar;70(3):988-992 [FREE Full text] [doi: 10.4103/ijo.IJO_1920_21] [Medline: 35225557]
6. Agarwal R, Tripathi A, Khan IA, Agarwal M. Effect of increased screen time on eyes during COVID-19 pandemic. J Family Med Prim Care 2022 Jul;11(7):3642-3647 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_2219_21] [Medline: 36387628]
7. Mohan A, Sen P, Shah C, Jain E, Jain S. Prevalence and risk factor assessment of digital eye strain among children using online e-learning during the COVID-19 pandemic: digital eye strain among kids (DESK study-1). Indian J Ophthalmol 2021 Jan;69(1):140-144 [FREE Full text] [doi: 10.4103/ijo.IJO_2535_20] [Medline: 33323599]
8. Duan H, Yan W. Visual fatigue a comprehensive review of mechanisms of occurrence, animal model design and nutritional intervention strategies. Crit Rev Food Sci Nutr 2025 Dec 28;65(9):1631-1655 [FREE Full text] [doi: 10.1080/10408398.2023.2298789] [Medline: 38153314]
9. Ide T, Toda I, Miki E, Tsubota K. Effect of blue light-reducing eye glasses on critical flicker frequency. Asia Pac J Ophthalmol 2015;4(2):80-85 [FREE Full text] [doi: 10.1097/apo.0000000000000069]
10. Lin JB, Gerratt BW, Bassi CJ, Apte RS. Short-wavelength light-blocking eyeglasses attenuate symptoms of eye fatigue. Invest Ophthalmol Vis Sci 2017 Jan 01;58(1):442-447 [FREE Full text] [doi: 10.1167/iovs.16-20663] [Medline: 28118668]
11. Singh S, Downie LE, Anderson AJ. Is critical flicker-fusion frequency a valid measure of visual fatigue? A post-hoc analysis of a double-masked randomised controlled trial. Ophthalmic Physiol Opt 2023 Mar 23;43(2):176-182 [FREE Full text] [doi: 10.1111/opo.13073] [Medline: 36416367]
12. Liu Z, Zhang K, Gao S, Yang J, Qiu W. Correlation between eye movements and asthenopia: a prospective observational study. J Clin Med 2022 Nov 28;11(23):7043 [FREE Full text] [doi: 10.3390/jcm11237043] [Medline: 36498619]
13. Wang Y, Zhai G, Zhou S, Chen S, Min X, Gao Z, et al. Eye fatigue assessment using unobtrusive eye tracker. IEEE Access 2018;6:55948-55962 [FREE Full text] [doi: 10.1109/access.2018.2869624]
14. Kim T, Lee EC. Experimental verification of objective visual fatigue measurement based on accurate pupil detection of infrared eye image and multi-feature analysis. Sensors (Basel) 2020 Aug 26;20(17):4814 [FREE Full text] [doi: 10.3390/s20174814] [Medline: 32858920]
15. Divjak M, Bischof H. Eye blink based fatigue detection for prevention of computer vision syndrome. MVA 2009:1-4 [FREE Full text]
16. Tobii Pro AB. Tobii Pro Lab (Version 1.241). Danderyd, Sweden: Tobii AB; Mar 2024.
17. Miranda AM, Nunes-Pereira EJ, Baskaran K, Macedo AF. Eye movements, convergence distance and pupil-size when reading from smartphone, computer, print and tablet. Scand J Optom Vis Sci 2018 Jul 22;11(1):1-5 [FREE Full text] [doi: 10.5384/sjovs.vol11i1p1-5]

XSL•FO
RenderX

18. Parker TM, Badihian S, Hassoon A, Saber Tehrani AS, Farrell N, Newman-Toker DE, et al. Eye and head movement recordings using smartphones for telemedicine applications: measurements of accuracy and precision. Front Neurol 2022;13:789581 [FREE Full text] [doi: 10.3389/fneur.2022.789581] [Medline: 35370913]
19. Wang Y, Zhai G, Chen S, Min X, Gao Z, Song X. Assessment of eye fatigue caused by head-mounted displays using eye-tracking. Biomed Eng Online 2019 Nov 15;18(1):111 [FREE Full text] [doi: 10.1186/s12938-019-0731-5] [Medline: 31729983]
20. Giap BD, Srinivasan K, Mahmoud O, Ballouz D, Lustre J, Likosky K, et al. A computational framework for intraoperative pupil analysis in cataract surgery. Ophthalmol Sci 2025;5(1):100597 [FREE Full text] [doi: 10.1016/j.xops.2024.100597] [Medline: 39435136]
21. Giap BD, Ballouz D, Srinivasan K, Lustre J, Likosky K, Mahmoud O, et al. CatSkill: artificial intelligence-based metrics for the assessment of surgical skill level from intraoperative cataract surgery video recordings. Ophthalmol Sci 2025;5(4):100764 [FREE Full text] [doi: 10.1016/j.xops.2025.100764] [Medline: 40385240]
22. Dhir L, Habib NE, Monro DM, Rakshit S. Effect of cataract surgery and pupil dilation on iris pattern recognition for personal authentication. Eye (Lond) 2010 Jun;24(6):1006-1010 [FREE Full text] [doi: 10.1038/eye.2009.275] [Medline: 19911017]
23. Mohsin H, Abdullah SH. Pupil detection algorithm based on feature extraction for eye gaze. 2017 Presented at: 2017 6th International Conference on Information and Communication Technology and Accessibility (ICTA); December 19-21, 2017; Muscat, Oman. [doi: 10.1109/icta.2017.8336048]
24. Giap BD, Srinivasan K, Mahmoud O, Mian SI, Tannen BL, Nallasamy N. Adaptive tensor-based feature extraction for pupil segmentation in cataract surgery. IEEE J Biomed Health Inform 2024 Mar;28(3):1599-1610 [FREE Full text] [doi: 10.1109/JBHI.2023.3345837] [Medline: 38127596]
25. Giap BD, Srinivasan K, Mahmoud O, Mian SI, Tannen BL, Nallasamy N. Tensor-based feature extraction for pupil recognition in cataract surgery. Annu Int Conf IEEE Eng Med Biol Soc 2023 Jul;2023:1-4 [FREE Full text] [doi: 10.1109/EMBC40787.2023.10340785] [Medline: 38082579]
26. Ba-Ali S, Lund-Andersen H, Brøndsted AE. Cataract surgery affects the pupil size and pupil constrictions, but not the late post-illumination pupil response. Acta Ophthalmol 2017 May;95(3):e252-e253 [FREE Full text] [doi: 10.1111/aos.13291] [Medline: 27778446]
27. Yin Z, Liu B, Hao D, Yang L, Feng Y. Evaluation of VDT-induced visual fatigue by automatic detection of blink features. Sensors (Basel) 2022 Jan 25;22(3):916 [FREE Full text] [doi: 10.3390/s22030916] [Medline: 35161662]
28. Mighty maze. Maze Toys. URL: https://maze.toys/mazes/mighty/ [accessed 2024-08-01]
29. React tetris. Chvin. URL: https://chvin.github.io/react-tetris/ [accessed 2024-08-01]
30. Game over. 12wave. URL: https://12wave.com/gameover [accessed 2024-08-01]
31. Search playground: Paris 2024. Google. URL: https://searchplayground.google/intl/en/paris-2024/ [accessed 2024-08-01]
32. Kim DJ, Lim C, Gu N, Park CY. Visual fatigue induced by viewing a tablet computer with a high-resolution display. Korean J Ophthalmol 2017 Oct;31(5):388-393 [FREE Full text] [doi: 10.3341/kjo.2016.0095] [Medline: 28914003]
33. Wolffsohn JS, Lingham G, Downie LE, Huntjens B, Inomata T, Jivraj S, et al. TFOS lifestyle: impact of the digital environment on the ocular surface. Ocul Surf 2023 Apr;28:213-252 [FREE Full text] [doi: 10.1016/j.jtos.2023.04.004] [Medline: 37062428]
34. Feis A, Lallensack A, Pallante E, Nielsen M, Demarco N, Vasudevan B. Reading eye movements performance on iPad vs print using a visagraph. J Eye Mov Res 2021 Sep 14;14(2):12 [FREE Full text] [doi: 10.16910/jemr.14.2.6] [Medline: 34729132]
35. Hoddes E. The history and use of the Stanford Sleepiness Scale. Psychophysiology 1971;9:150. [doi: 10.1037/t07116-000]
36. Miley A, Kecklund G, Åkerstedt T. Comparing two versions of the Karolinska Sleepiness Scale (KSS). Sleep Biol Rhythms 2016 Feb 01;14(3):257-260 [FREE Full text] [doi: 10.1007/s41105-016-0048-8] [Medline: 28781579]
37. De Vries J, Michielsen H, Van Heck GL, Drent M. Measuring fatigue in sarcoidosis: the Fatigue Assessment Scale (FAS). Br J Health Psychol 2004 Sep 16;9(Pt 3):279-291 [FREE Full text] [doi: 10.1348/1359107041557048] [Medline: 15296678]
38. Wu F, Chiariglione M, Gao X. Automated optic disc and cup segmentation for glaucoma detection from fundus images using the detectron2's Mask R-CNN. 2022 Presented at: 2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT); October 20-22, 2022; Ankara, Turkey p. 567-570. [doi: 10.1109/ismsit56059.2022.9932660]
39. Başakın EE, Ekmekcioğlu ?, Çıtakoğlu H, Özger M. A new insight to the wind speed forecasting: robust multi-stage ensemble soft computing approach based on pre-processing uncertainty assessment. Neural Comput Applic 2021 Aug 30;34(1):783-812 [FREE Full text] [doi: 10.1007/s00521-021-06424-6]
40. Zouzou Y, Citakoglu H. General and regional cross-station assessment of machine learning models for estimating reference evapotranspiration. Acta Geophys 2022 Oct 17;71(2):927-947 [FREE Full text] [doi: 10.1007/s11600-022-00939-9]
41. Citakoglu H, Aktürk G, Demir V. Hybrid machine learning for drought prediction at multiple time scales: a case study of Ağrı station, Türkiye. Acta Geophys 2025 Mar 05;73(2):1643-1677 [FREE Full text] [doi: 10.1007/s11600-024-01501-5]
42. Faul F, Erdfelder E, Buchner A, Lang A. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160 [FREE Full text] [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]

XSL•FO
RenderX

43. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. Int J Endocrinol Metab 2012;10(2):486-489 [FREE Full text] [doi: 10.5812/ijem.3505] [Medline: 23843808]
44. Whitley E, Ball J. Statistics review 5: comparison of means. Crit Care 2002 Oct;6(5):424-428 [FREE Full text] [doi: 10.1186/cc1548] [Medline: 12398782]
45. Whitley E, Ball J. Statistics review 6: nonparametric methods. Crit Care 2002 Dec;6(6):509-513 [FREE Full text] [doi: 10.1186/cc1820] [Medline: 12493072]
46. StataCorp. Stata Statistical Software: Release 18. College Station, TX: StataCorp LLC; 2023.
47. Kim D, Choi S, Choi J, Shin H, Sohn K. Visual fatigue monitoring system based on eye-movement and eye-blink detection. Proc SPIE 2011 Feb 11;7863:786303 [FREE Full text] [doi: 10.1117/12.873354]
48. Wang Y, Zhang L, Fang Z. Eye fatigue detection through machine learning based on single channel electrooculography. Algorithms 2022 Mar 03;15(3):84 [FREE Full text] [doi: 10.3390/a15030084]
49. Kasprzak HT, Licznerski TJ. Influence of the characteristics of tear film break-up on the point spread function of an eye model. Proc SPIE 1999 Jul 14;3820:353091 [FREE Full text] [doi: 10.1117/12.353091]
50. Li S, Hao D, Liu B, Yin Z, Yang L, Yu J. Evaluation of eyestrain with vertical electrooculogram. Comput Methods Programs Biomed 2021 Sep;208:106171 [FREE Full text] [doi: 10.1016/j.cmpb.2021.106171] [Medline: 34102561]
51. Pansell T, Porsblad M, Abdi S. The effect of vertical gaze position on ocular tear film stability. Clin Exp Optom 2007 May 15;90(3):176-181 [FREE Full text] [doi: 10.1111/j.1444-0938.2007.00136.x] [Medline: 17425763]
52. Gowrisankaran S, Sheedy JE, Hayes JR. Eyelid squint response to asthenopia-inducing conditions. Optom Vis Sci 2007 Jul;84(7):611-619 [FREE Full text] [doi: 10.1097/OPX.0b013e3180dc99be] [Medline: 17632310]
53. Cen L, Choi J, Yao X, Gil Y, Narayanan S, Pentz M. A personal visual comfort model: predict individual's visual comfort using occupant eye pupil size and machine learning. IOP Conf Ser: Mater Sci Eng 2019 Sep 01;609(4):042097 [FREE Full text] [doi: 10.1088/1757-899x/609/4/042097]
54. Hopkinson RG. Glare discomfort and pupil diameter. J Opt Soc Am 1956 Aug 01;46(8):649-656 [FREE Full text] [doi: 10.1364/josa.46.000649] [Medline: 13346426]
55. Geacintov T, Peavler WS. Pupillography in industrial fatigue assessment. J Appl Psychol 1974 Apr;59(2):213-216 [FREE Full text] [doi: 10.1037/h0036529]
56. Lin Y, Fotios S, Wei M, Liu Y, Guo W, Sun Y. Eye movement and pupil size constriction under discomfort glare. Invest Ophthalmol Vis Sci 2015 Jan 29;56(3):1649-1656 [FREE Full text] [doi: 10.1167/iovs.14-15963] [Medline: 25634984]
57. Rajabi-Vardanjani H, Habibi E, Pourabdian S, Dehghan H, Maracy MR. Designing and validation a visual fatigue questionnaire for video display terminals operators. Int J Prev Med 2014 Jul;5(7):841-848 [FREE Full text] [Medline: 25104995]
58. Eisen-Enosh A, Farah N, Burgansky-Eliash Z, Polat U, Mandel Y. Evaluation of critical flicker-fusion frequency measurement methods for the investigation of visual temporal resolution. Sci Rep 2017 Nov 15;7(1):15621 [FREE Full text] [doi: 10.1038/s41598-017-15034-z] [Medline: 29142231]

Abbreviations

AP: average precision
AR: average recall
AUROC: area under the receiver operating characteristic curve
CFF: critical flicker fusion frequency
COCO: Common Objects in Context
CVS-Q: Computer Vision Syndrome Questionnaire
Dice: Dice similarity coefficient
DL: deep learning
FAS: Fatigue Assessment Scale
fps: frames per second
GPU: graphics processing unit
HVA: habitual visual acuity
IoU: intersection over union
IR: infrared
KSS: Karolinska Sleepiness Scale
mark R-CNN: mask region-based convolutional neural network
ML: machine learning
PERCLOS: percentage of eye closure
SSS: Stanford Sleepiness Scale
SVM: support vector machine
TBUT: tear film break-up time
XGBoost: extreme gradient boosting

XSL•FO
RenderX

Edited by A Kushniruk, A Bamgboje-Ayodele; submitted 20.Sep.2025; peer-reviewed by BD Giap, X Zhang, H Citakoglu; comments to author 17.Nov.2025; accepted 21.Feb.2026; published 28.Apr.2026.

Please cite as:
Song F, Li G, Zhang J, Tian Z, Li M, Lai M, He M, Chen Y
Identifying Biomarkers Using a Portable, Home-Based Eye-Tracking System to Predict Short-Term Visual Fatigue Deterioration: Prospective Observational Feasibility Study
JMIR Hum Factors 2026;13:e84479
URL: https://humanfactors.jmir.org/2026/1/e84479
doi:10.2196/84479
PMID:

©Fan Song, Guangyu Li, Jian Zhang, Zhen Tian, Mingge Li, Mengying Lai, Mingguang He, Yanxian Chen. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 28.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Proximal and Long-Term Participants' Perspectives on Approach Bias Modification for Smoking Cessation: Qualitative Study

Alla Machulska[1], PhD; Johanna Bück[1], MSc; Esra Sünkel[1], MSc; Marie Neubert[1], PhD; Tim Klucken[1], Prof Dr

Clinical Psychology and Psychotherapy, Department of Psychology, University of Siegen, Siegen, North Rhine-Westphalia, Germany

Corresponding Author:
Alla Machulska, PhD
Clinical Psychology and Psychotherapy
Department of Psychology
University of Siegen
Obergraben 23
Siegen, North Rhine-Westphalia, D-57072
Germany
Phone: 49 01745402367
Email: alla.machulska@uni-siegen.de

Abstract

Background: Approach bias modification (ApBM) aims to target maladaptive approach tendencies toward substance-related cues and has increasingly been examined as an adjunctive intervention for substance use disorders, including nicotine use. However, although participants' subjective experiences of ApBM are likely to influence both its effectiveness and successful implementation, they have rarely been systematically investigated.

Objective: This study aimed to explore both proximate and long-term subjective training experiences reported by participants from 2 previously conducted randomized controlled trials of ApBM for smoking cessation. Training experiences were explored across 2 delivery modalities (virtual reality and smartphone app) and compared with those reported in matched sham training conditions.

Methods: Participants were invited to provide open-ended feedback immediately following the intervention and up to 7 weeks after study entry (proximate feedback), as well as again 4 years later (long-term feedback). Thematic analysis was conducted to identify core themes, complemented by exploratory frequency and descriptive quantitative analyses comparing ApBM and sham conditions.

Results: In total, 87 (ApBM: n=45; sham: n=42) of the 178 participants included in the randomized controlled trials provided proximate (7-week follow-up) feedback, and 63 (ApBM: n=33; sham: n=30) provided long-term (4-year follow-up) feedback; 104 participants contributed feedback at least once. Four overarching themes were identified, with the fourth emerging only in long-term reflections: (1) perceived treatment effects, spanning beneficial changes and a perceived lack of effects; (2) mechanisms of action, including presumed working mechanisms and impeding factors; (3) feedback on the training and study experience; and (4) attribution of effects to training-specific or external factors. Overall, participants engaged extensively in reflecting on potential mechanisms underlying the training, their relation to therapeutic outcomes, and opportunities for improving future interventions. Exploratory frequency analyses indicated that participants in the ApBM group more often suggested improvements to the training, whereas participants in the sham condition more frequently reported impeding factors, particularly difficulties understanding the training rationale. Quantitative training evaluations were generally more favorable for ApBM than for sham training, as reflected in a significantly higher overall evaluation score. In particular, ApBM received higher ratings for its perceived effectiveness in reducing smoking and its applicability to everyday situations.

Conclusions: This study provides novel insights into how ApBM for smoking is experienced both shortly after training and years later. To our knowledge, this is the first study in the context of cognitive bias modification, including ApBM, to examine subjective experiences over a multiyear time frame. The findings highlight the importance of participants' understanding of the training rationale and underscore the value of incorporating participants' perspectives into the development and evaluation of ApBM interventions.

(JMIR Hum Factors 2026;13:e93238) doi:10.2196/93238

XSL•FO
RenderX

KEYWORDS

approach bias modification; cognitive bias modification; long-term outcomes; mobile app; qualitative analysis; smoking cessation; subjective experiences; virtual reality

Introduction

Cigarette smoking is highly addictive and remains prevalent despite its well-documented health risks. Globally, 1.18 billion people report regular smoking [1], with approximately half meeting diagnostic criteria for nicotine dependence [2]. Although the majority of individuals who smoke express an intention to quit [3], access to smoking cessation treatment remains limited to a small proportion of those in need. According to the World Health Organization [4], only about 30% have access to comprehensive smoking cessation services. Moreover, even among individuals who receive treatment, relapse rates remain substantially high [5,6]. Consequently, the development of feasible and easily accessible interventions that can complement and improve existing treatments represents a key priority not only in nicotine and tobacco research but also in public health.

One promising approach in this regard is the integration of therapeutic interventions into digital formats. In the context of smoking cessation, a growing number of mobile and computerized interventions have been developed, with evidence suggesting that incorporating cognitive behavioral therapy principles, such as self-reflection and cognitive restructuring, can support smoking cessation compared with no or minimal support (see Chu et al [7] for a recent meta-analysis). However, given the persistently high relapse rates even following treatment [5,7], it is likely that, beyond reflective processes targeted in traditional psychotherapeutic approaches, more automatic, implicit cognitive processes contribute to the maintenance of smoking behavior.

Over the past 2 decades, a range of computer-based training interventions has been developed to modify such implicit processes implicated in maladaptive substance use (for reviews, see Kakoschke et al [8] and Vrijsen et al [9]). Collectively, these interventions are referred to as cognitive bias modification (CBM). Whereas cognitive bias assessment tasks require equally frequent responses to substance-related and substance-unrelated stimuli (eg, equally often approaching and avoiding smoking and nonsmoking pictures), CBM training paradigms use systematic stimulus-response contingencies (eg, always avoiding smoking and always approaching nonsmoking pictures) to counteract the bias (for a recent review, see Vrijsen et al [9]). Hence, unlike traditional psychotherapeutic approaches, which primarily target reflective processes to facilitate behavioral change, CBM aims to directly alter more automatic, implicit processes through repeated training. To date, several CBM variants have been developed that differ in the specific cognitive processes they aim to modify. For example, attention bias modification (ABM) aims to train attentional allocation away from substance-related cues and toward more adaptive alternatives, using paradigms such as spatial cueing or gaze-contingent training. CBM for interpretations (CBM-I) targets the tendency to interpret ambiguous stimuli in a disorder-congruent way by consistently reinforcing benign or substance-unrelated interpretations. Approach bias modification (ApBM), on the other hand, repeatedly trains individuals to make avoidance movements in response to substance-related cues and approach movements toward neutral or goal-consistent cues [9]. Indeed, the strongest evidence for CBM's clinical effectiveness comes from ApBM in alcohol use disorder (AUD), particularly when applied as an adjunct to psychological therapy in structured clinical settings (see Pan et al [10] for a recent meta-analysis and Wittekind et al [11] for a review). Several randomized controlled trials (RCTs) have demonstrated reductions in relapse rates of up to 10% at 1-year follow-up [12-16], which has led to the inclusion of ApBM in national treatment guidelines for AUD in Germany [17] as well as Australia [18].

Since then, this approach has been extended to other consumption-related behaviors, including nicotine use [19-23]. Although several studies have reported promising effects, such as increases in the number of abstinent days [19] or reductions in daily cigarette consumption following training [20,21], null findings have also been observed [23], and overall effect sizes remain modest (for a meta-analysis, see Boffo et al [24]). While multiple factors may contribute to this inconsistency, including differences in training design, participant characteristics, and outcome variables, participants' subjective experiences of the training are likely to play an important yet underexplored role in ApBM effectiveness. This is particularly relevant given that multiple training sessions are required to induce reliable and lasting changes in cognitive biases [12,25], which may place a considerable burden on participants and affect adherence. Consequently, training acceptance and motivation are critical determinants of successful CBM implementation [26]. Yet, systematic qualitative investigation of participant experiences with CBM, particularly ApBM for smoking, remains limited. Evidence from the few available studies suggests that CBM tasks are often experienced as monotonous and repetitive [27] and that participants frequently report difficulties in understanding the relevance of the training for their behavioral goals [28]. Such perceptions are likely to undermine training motivation and engagement and may therefore partially account for the inconsistent clinical outcomes observed in the literature.

This notion is supported by evidence from related CBM domains. For example, in the context of ABM for social anxiety, one study showed that patients' perceptions of the training predicted both bias reduction and subsequent symptom improvement [29]. Systematic research on subjective experiences with CBM in general, and ApBM in particular, remains scarce in the context of substance use, although a small number of studies suggest that participants' perceptions may be important in this domain as well. For example, in an implementation study on ApBM for AUD, Schenkel et al [30] assessed training evaluation among alcohol dependent inpatients and found that although the paradigm was rated predominantly positive, most patients (60%) did not believe that the training was effective in reducing relapse rates, and only a minority

XSL•FO
RenderX

(29%) perceived a transfer to everyday behavior. Extending these findings, a qualitative study investigating a combined ABM and ApBM intervention for problematic gaming behavior reported that negative training experiences, including perceptions of excessive duration and tediousness, might have accounted for extremely high attrition rates of up to 90% [31]. Thus, participants' perceptions of CBM procedures are likely to contribute to both training uptake and effectiveness, underscoring the need for more systematic investigation of these perceptions and for strategies to improve them.

In line with this objective, recent advances in training delivery have aimed to enhance training acceptance by embedding ApBM in novel digital formats, including virtual reality (VR) and smartphone apps (for a review, see Forman et al [26]). These approaches are assumed to increase engagement and perceived relevance, as training can be conducted in immersive or consumption-related contexts. Supporting this notion, 2 proof-of-principle trials in the context of anxiety translated the principles of ABM [32] and CBM-I [33] into VR environments and reported high levels of enjoyment, motivation, flow, and presence. However, both studies were limited by small sample sizes, single-session training designs, and relatively short follow-up periods of up to 6 weeks. Besides, a mixed methods study combining internet-based ApBM and CBM-I for hazardous alcohol use and social anxiety reported overall satisfaction, but a need for a more compelling training rationale [34]. Similarly, a feasibility study on an app-based ApBM intervention for problem drinking found that in addition to positive feedback regarding usability and intuitiveness of app design, participants criticized the intervention as repetitive and insufficiently personalized [35]. In the context of ApBM for smoking, 2 larger RCTs were recently conducted by our research group, both including multiple training sessions and follow-up assessments. These studies demonstrated that ApBM delivered via VR [36] or a smartphone app [37] led to greater smoking reduction compared to respective control conditions. To date, however, little is known about how these novel delivery formats are perceived by individuals motivated to quit smoking, and whether common challenges associated with CBM, such as perceived monotony or a lack of meaningfulness, remain an issue of concern. Moreover, there are no long-term studies examining how such novel training approaches are perceived over extended periods of time or the extent to which they are retrospectively regarded as useful in the longer-term process of behavior change. This is particularly important, as research on psychotherapy outcomes indicates that patients' perceptions of treatment may evolve over time, with important implications for sustained behavioral change and well-being. For example, Malkomsen et al [38] demonstrated that patients who were initially uncertain about the benefits of their psychotherapeutic treatment reported more positive views 3 years later. Such shifts may be particularly relevant given evidence that therapy perceptions influence whether learned skills are applied after treatment termination [39].

Against this background, this study examined subjective training experiences reported by participants from 2 RCTs that used VR-based [36] and app-based [37] ApBM for smoking cessation. The aims of the study were threefold: (1) to identify themes reflecting participants' experiences with the training, (2) to examine how participants evaluate novel delivery formats of ApBM for smoking, and (3) to explore whether and how training perceptions change or remain stable over the long term. More specifically, regarding the first aim, we explored the extent to which emergent themes correspond to those identified in previous qualitative studies on CBM and psychotherapy more broadly, as well as whether study-specific themes arise in the present context. Regarding the second aim, we examined whether digitally delivered ApBM are associated with engagement and motivation and/or whether these novel formats introduce distinct challenges (eg, technical issues). Regarding the third aim, we investigated whether long-term reflections become more general over time or whether participants report additional or qualitatively different insights compared to their immediate evaluations. To answer these research questions, participants were invited to provide open-ended feedback immediately following the intervention and again approximately 4 years after study completion. Thematic analysis was used to identify core themes associated with the training principles. To contextualize and complement participants' subjective accounts, descriptive information on smoking-related outcomes and quantitative training evaluations has been additionally assessed.

Methods

Ethical Considerations

This trial was approved by the local ethics committee of the University of Siegen (reference numbers ER_16_2018 and ER_18_2018) and was conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines. Participation was voluntary, and participants were able to withdraw their consent for participation at any time. Participants did not receive financial compensation for their involvement in the research.

To ensure privacy and confidentiality, all study data were stored on secure servers of the University of Siegen and were accessible only to authorized members of the study team. Participants were assigned a unique identification number, and all research data were stored and analyzed using these codes. A separate password-protected file linking participant names to identification numbers was maintained exclusively for study administration and follow-up purposes and was stored separately from the research data.

Trial Procedure

This study examines subjective experiences with either a VR-based or app-based ApBM for individuals motivated to quit smoking. Quantitative outcomes related to training effects and clinical efficacy of these interventions have been reported previously (see Machulska et al [36,37]). For all participants, a brief smoking cessation intervention constituted treatment as usual (TAU), followed by a 14-day training program delivered either in VR or via a smartphone app. Participants randomized to the experimental group (EG) were trained to respond to smoking-related cues with avoidance movements and to alternative cues with approach movements. Participants in the control group (CG) received a sham training in which no

XSL•FO
RenderX

systematic contingency between stimulus content and response movement was implemented.

In the app-based ApBM trial [37], an additional control condition was implemented, which received the TAU but was not given access to the training app during the main study period (baseline to short-term follow up). Consequently, feedback from this group is not presented in this analysis.

To explore participants' subjective experiences, open-ended feedback was collected immediately after the training period (posttest), 7 weeks after baseline (short-term follow-up), and approximately 4 years after baseline (long-term follow-up). Notably, the posttest and 7-week follow-up assessments were both conducted within the active study phase of the original RCTs. Accordingly, both time points are considered to reflect participants' proximate experiences and evaluations, rather than fully naturalistic, longer-term reflections. Based on this notion and to increase the amount of available qualitative data per time point, feedback from the posttest and short-term follow-up assessments was combined. Accordingly, these analyses distinguish between proximate feedback (posttest and 7-week follow-up) and long-term feedback (4-year follow-up).

Posttest and short-term follow-up assessments were conducted on-site at the laboratory. After completing the quantitative questionnaire battery, participants were invited to provide open-ended feedback in written form via the survey platform LimeSurvey. At the long-term follow-up, feedback was collected remotely. To maximize response rates, 2 complementary approaches were used. First, participants were contacted by telephone and, following a brief assessment of current smoking behavior, were invited to share open-ended feedback regarding the training and/or their overall study experience. In these cases, study assistants conducting the interviews documented participants' responses verbatim. Second, all participants received an email invitation to complete an online survey, which also included an open-ended response field to allow for more general reflections on study participation.

Participants

All participants took part in 1 of 2 RCTs investigating the effectiveness of a VR-based [36] or app-based [37] ApBM for smoking. Eligibility criteria for those RCTs included being at least 18 years of age, smoking a minimum of 6 cigarettes per day for at least 6 months prior to study entry, and reporting an intention to quit smoking within the subsequent 6 months. Exclusion criteria comprised a current substance use disorder other than nicotine, current psychiatric illness, insufficient proficiency in German, uncorrected visual or auditory impairments, or dyschromatopsia. In addition, participants who dropped out prior to the first training session or failed to become aware of the training contingencies in the EG were excluded from data analyses. This resulted in a total sample of 178 participants (n_{EG}=87; n_{CG}=91), of whom 96 received the VR-based training and 82 the app-based training. Overall, 104 of the 178 participants provided feedback at the proximate or long-term assessment. Of these, 54 had been randomized to ApBM training (n_{VR}=35; $n_{App=}$19) and 50 to sham training (n_{VR}=29; $n_{App=}$31). No differences were observed between participants who provided feedback and those who did not with regard to condition (χ^2_1=0.94; *P*=.34). However, participants in the VR trial were more likely to provide feedback than those in the app trial (χ^2_1=5.83; *P*=.02). The sample was evenly distributed by gender (52/104, 50% female), with age ranging from 22 to 67 years (mean 48.65, SD 11.73).

Interventions

The interventions used in the 2 initial RCTs are briefly described below. Please refer to the original studies [36,37] for more detailed information.

TAU: Brief Smoking Cessation Intervention

Prior to the baseline assessment, all participants took part in a brief smoking cessation intervention (TAU), administered by a trained psychologist. The session was conducted in small groups of up to 4 individuals and lasted about 90-120 minutes. The following topics were covered by the intervention: psychoeducation, including reflective and impulsive determinants of smoking maintenance vs cessation, motivational support for smoking cessation, instructions to self-monitor daily smoking, and bibliotherapy through the provision of a self-help book.

Approach Bias Training Paradigm

Overview

Each training variant was based on the traditional joystick training, in which different pictures are presented consecutively on a computer screen, and participants are required to perform push (avoidance) or pull (approach) movements by means of a joystick attached to the computer. Upon a push movement, images shrink in size, creating a sense of avoidance, whereas upon a pull movement, images grow, creating a sense of approaching (see Rinck and Becker [40], for a description of the original task). This training principle was translated into 2 novel delivery modalities: a VR format in the first RCT [36], where participants executed push and pull movements with their own arms that were visible in the virtual environment, and a smartphone app in the second RCT [37], where participants performed upward (avoidance) and downward (approach) swiping movements on a touchscreen. All participants received the same cover story stating that the training aimed to target automatic processes involved in smoking by strengthening more controlled responses to smoking-related stimuli. In addition, it was communicated that this approach was expected to enhance self-control over smoking behavior in everyday contexts.

VR Training

The virtual environment was situated within an office setting, in which smoking-related and smoking-unrelated objects consecutively appeared on a virtual desk. Objects were either blue- or red-bordered. An indirect task instruction was used as participants were instructed to ignore object content and respond to border color only (ie, push red-bordered objects and pull blue-bordered objects). In the EG, all smoking-related objects were bordered in red, which required an avoidance movement, whereas all smoking-unrelated objects were bordered in blue, which required an approach movement. Hence, a 100%-contingency between object content and response

XSL•FO
RenderX

movement was used. In the CG, participants performed a sham training version that closely resembled the VR-ApBM condition, except that there was no contingency between object content and response format. Each session took approximately 10-15 minutes to complete.

Smartphone App Training

Smoking-related and smoking-unrelated pictures appeared consequently at the center of a mobile screen. Pictures were either rotated 3° to the left or 3° to the right. An indirect task instruction was used as participants were instructed to ignore picture content and respond to image rotation only (ie, swipe-up pictures rotated to the right and swipe-down pictures rotated to the left). In the EG, all smoking-related pictures were rotated to the right and had to be swiped up (avoidance), whereas all smoking-unrelated pictures were rotated to the left and had to be swiped down (approach). As with the original ApBM training, a zooming feature was introduced to create a visual sense of avoiding vs approaching the image. In the CG, participants performed a sham training version that closely resembled the app-ApBM condition, except that no contingency between picture content and response format was introduced. Each session took approximately 5 minutes to complete.

Assessments

Qualitative Data

Participants were invited to provide open-ended feedback in response to the question: “Do you have any feedback or comments about the training?” Responses were assessed at 3 time points: online immediately after the training interval (posttest) and 7 weeks after baseline (short-term follow-up), which were combined to represent proximate feedback, as well as online and via telephone approximately 4 years after baseline (long-term follow-up), which was combined to represent long-term feedback.

Quantitative Data

Overview

To aid interpretation of the qualitative findings, basic smoking-related and training-specific variables assessed at proximate and long-term time points are summarized below. For a comprehensive description of all measures administered in the original trials, please refer to the published study protocols [41,42] and original reports [36,37].

Smoking-Related Data

In addition to demographic information (ie, age and sex), participants provided smoking-related information, including (1) the duration of nicotine use (in years), (2) 7-day point abstinence, (3) number of daily smoked cigarettes, and (4) degree of nicotine dependence as measured with the Fagerström Test for Nicotine Dependence (FTND) [43] (German version [44]), with scores ranging between 0 (no/very weak dependence) and 10 (very strong dependence).

Training-Related Data

The number of completed training sessions was recorded either by study assistants (VR-based training) or automatically within the training app (app-based training). At posttest, participants additionally completed a brief training evaluation questionnaire consisting of 9 dichotomous items (0=no, 1=yes) assessing perceived ease (“The training was easy to perform.”), motivation (“It was difficult to motivate myself to complete the training.” [reversely coded]), clarity (“The training setup was clear.”), and subjective efficacy (“The training is beneficial in reducing smoking.”) of the training. Item responses were summed to yield a total evaluation score ranging from 0 to 9, with higher values indicating more positive evaluations of the training.

Data Preparation and Planned Analyses

To explore participants’ perspectives and potential group differences in subjective training experience, we conducted a thematic analysis of the proximate and long-term open-ended responses. Three independent coders (AM, ES, and JB) followed an inductive, stepwise procedure for psychological thematic analysis according to Braun and Clarke [45,46]. All coders first reviewed the initial 20% of responses and generated codes from the data. They then met to compare codes, resolve discrepancies, and agree on a coding scheme. The remaining responses were coded using this scheme. Codes were subsequently grouped into higher-order subthemes and organized into overarching themes. Finally, the coding framework and theme assignments were reviewed by all coders and adjusted where necessary. Because the thematic analysis was exploratory, no a priori hypotheses were specified. Quotations presented in the Results were translated from German to English and checked for translation accuracy. To complement the qualitative analysis, we additionally quantified how often each theme and subtheme was mentioned and explored whether frequencies differed between the EG and CG using chi-square tests. Prior to quantification, the coders convened to ensure consistent application of the final coding framework across all participants. This procedure was undertaken to guarantee the reliability and validity of the theme assignment for the frequency-based analyses.

In addition, quantitative data regarding demographic variables, smoking-related behavior, and training-specific information are presented descriptively (means, SDs, and proportions) across conditions and time points to provide contextual information for the qualitative thematic analysis.

Results

Participants’ Characteristics

Participants’ characteristics with regard to demographic and smoking-related data are provided in Tables 1 and 2. As can be seen, the sample consisted of long-term smoking individuals who had been smoking for almost 30 years. The average consumption at baseline was approximately 19 (SD 6.59) cigarettes per day, with a moderate nicotine dependence (FTND score: mean 4.85, SD 2.12). Although descriptively, participants in the CG completed slightly more training sessions than those in the EG (CG: mean 9.1, SD 6.53 vs EG: mean 7.7, SD 5.27), short-term smoking-related outcomes improved to a similar extent in both groups. At the 4-year follow-up, descriptive differences between groups were observed for abstinence (30/48, 63% in the EG vs 13/38, 34% in the CG), daily cigarette consumption (EG: mean 6.19, SD 10.11 vs CG: mean 9.26, SD

XSL•FO
RenderX

9.36), and degree of nicotine dependence (EG: FTND: mean 1.29, SD 2.45 vs CG: mean 2.41, SD 2.80). Please note that these values are reported for contextual reasons only. Formal statistical analyses of smoking-related outcomes are presented in the primary RCT reports [36,37], with additional long-term analyses reported elsewhere [47].

Table 1. Sample sizes for overall participants and study groups across assessment time points.[a]

Variable	Overall (n=104), n	EG[b] (ApBM[c]; n=54), n	CG[d] (sham; n=50), n
Baseline	104	54	50
Posttest	100	51	49
Short-term (7-week) follow-up[e]	87	45	42
Long-term (4-year) follow-up	86	48	38

[a]Sample sizes were identical across outcome measures (abstinence and daily cigarette consumption).

[b]EG: experimental group.

[c]ApBM: approach bias modification.

[d]CG: control group.

[e]For nicotine dependence (Fagerström Test for Nicotine Dependence [43]) at the 7-week follow-up, data were available for 82 participants (EG=44 and CG=38) because scores could not be obtained from 5 participants.

Table 2. Demographic characteristics and smoking-related outcomes for the overall sample and study groups.

Variable	Overall	EG[a] (ApBM[b])	CG[c] (sham)
Demographic characteristics			
Age (years), mean (SD)	48.65 (11.73)	48.35 (12.98)	48.98 (10.33)
Gender (female), n (%)	52 (50)	26 (48)	26 (52)
Duration of nicotine use (years), mean (SD)	28.97 (12.08)	28.54 (12.85)	29.44 (11.29)
Number of training sessions, mean (SD)	8.36 (5.92)	7.70 (5.27)	9.06 (6.53)
Abstinence, n (%)			
Baseline	0 (0)	0 (0)	0 (0)
Posttest	20 (20)	12 (24)	8 (16)
Short-term (7-week) follow-up	21 (24)	11 (24)	10 (24)
Long-term (4-year) follow-up	43 (50)	30 (63)	13 (34)
Daily cigarette consumption, mean (SD)			
Baseline	18.93 (6.59)	19.69 (6.98)	18.11 (6.10)
Posttest	10.25 (7.82)	9.88 (8.45)	10.63 (7.17)
Short-term (7-week) follow-up	9.13 (6.86)	9.23 (6.98)	9.02 (6.81)
Long-term (4-year) follow-up	7.54 (9.85)	6.19 (10.11)	9.26 (9.36)
Nicotine dependence (FTND[d] score), mean (SD)			
Baseline	4.85 (2.12)	4.91 (2.33)	4.78 (1.89)
Posttest	3.04 (2.62)	3.00 (2.87)	3.08 (2.35)
Short-term (7-week) follow-up	2.45 (2.36)	2.52 (2.49)	2.37 (2.24)
Long-term (4-year) follow-up	1.78 (2.66)	1.29 (2.45)	2.41 (2.80)

[a]EG: experimental group.

[b]ApBM: approach bias modification.

[c]CG: control group.

[d]FTND: Fagerström Test for Nicotine Dependence [43]; scale: 0-10.

Quantitative Training Evaluation

Table 3 summarizes participants' quantitative responses to the training evaluation questionnaire at posttest. Overall, participants were satisfied with the training. The vast majority of participants reported that the training was clearly structured (97/99, 98%) and easy to perform (93/99, 94%), and most (83/99, 84%) indicated that they would recommend it for further use.

XSL•FO
RenderX

Approximately three-quarters of participants perceived the training as helpful in reducing cigarette smoking, and 68% (67/99) reported no difficulties in motivating themselves to complete the sessions. Lower levels of agreement were observed for items referring to broader effects: only 37% (37/99) of participants endorsed the view that the training may contribute to abstinence, and 47% (47/99) believed it can be applied to everyday situations.

Although the overall pattern was similar across conditions, participants in the EG evaluated the training more positively than those in the CG, as indicated in a significantly higher training evaluation sum score (t_{97}=3.08; *P*=.003). Specifically, participants in the EG indicated to have noticed more overall positive effects of the training (EG: 37/50, 74% vs CG: 24/49, 49%; χ^2_1=6.55; *P*=.01). In addition, a greater proportion of participants in the EG believed that the training was beneficial in reducing smoking (42/50, 84% vs 30/49, 61%; χ^2_1=6.47; *P*=.01) and could be transferred to everyday situations (30/50, 60% vs 17/49, 35%; χ^2_1=6.34; *P*=.01).

The training perceptions and between-group differences suggested by these quantitative findings are examined in greater depth in the qualitative analyses below.

Table 3. Training evaluations at posttest.[a]

Item	Overall	EG[b] (ApBM[c])	CG[d] (sham)	*P* value
The training setup was clear, n (%)	97 (98)	50 (100)	47 (96)	.15
The training was easy to perform, n (%)	93 (94)	47 (94)	46 (94)	.98
I have noticed positive effects from the training, n (%)	61 (62)	37 (74)	24 (49)	.01
I have made progress in my reaction times during the training, n (%)	65 (66)	36 (72)	29 (59)	.18
The training is beneficial in quitting smoking, n (%)	37 (37)	22 (44)	15 (31)	.17
The training is beneficial in reducing smoking, n (%)	72 (73)	42 (84)	30 (61)	.01
Training can be applied to everyday situations, n (%)	47 (47)	30 (60)	17 (35)	.01
It was difficult to motivate myself to complete the training, n (%)	32 (32)	14 (28)	18 (37)	.35
The training should continue to be used, n (%)	83 (84)	44 (88)	39 (80)	.26
Sum Score (0-9)[e], mean (SD)	6.28 (2.03)	6.88 (1.93)	5.67 (1.96)	.003

[a]Group differences regarding proportions were analyzed using chi-square tests; group differences in the sum score were analyzed using a 2-sample *t* tests. All *P* values are 2-tailed.

[b]EG: experimental group.

[c]ApBM: approach bias modification.

[d]CG: control group.

[e]This row indicates the sum score (0-9) for all items answered with "yes" (coded as 1), except for the item "It was difficult to motivate myself to complete the training," which was reversely coded.

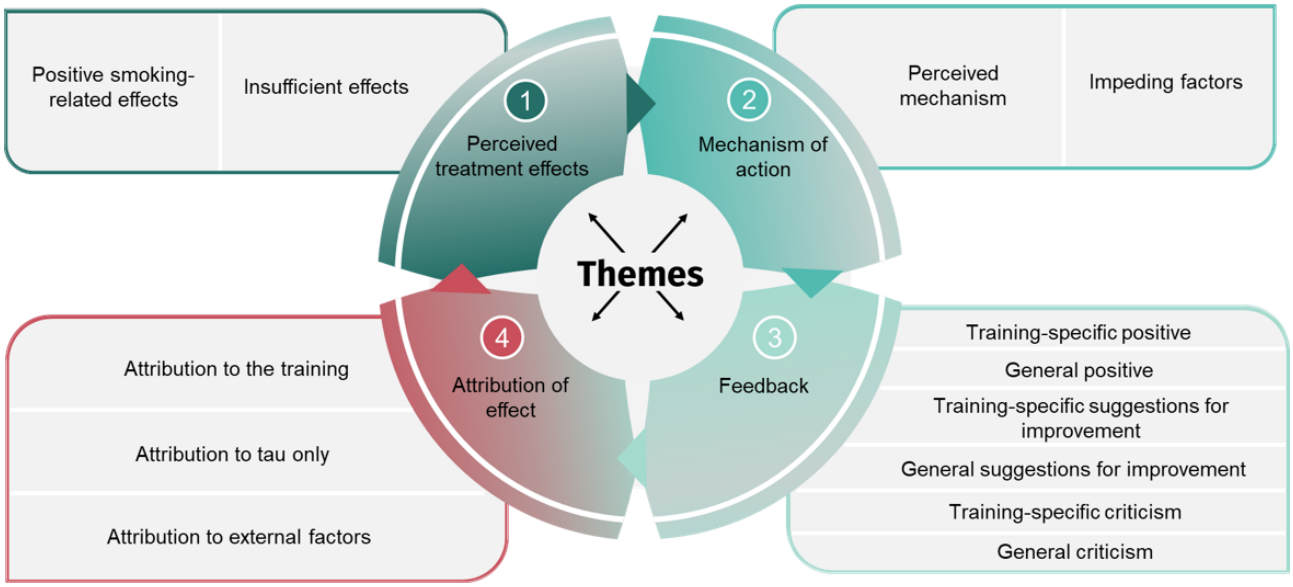
Thematic Analysis of Participants' Feedback on the Training

Overview

Overall, the analysis yielded 4 overarching themes: *perceived treatment effects*, *mechanisms of action*, *feedback on the intervention*, and *attribution of effect*, each comprising several subthemes. Whereas the first 3 themes appeared at both time points, the fourth theme (attribution of effect) was only evident in the 4-year follow-up assessment. A comprehensive overview of all themes and subthemes is presented in Figure 1.

XSL•FO
RenderX

Figure 1. Overview of the thematic analysis.



Participants' Proximate Feedback

Overview

In total, 87 participants provided proximate feedback (posttest and/or 7-week follow-up). Of these, 45 were randomized to the EG, and 42 were in the CG. Example quotations are provided within the corresponding sections. In addition, each quote is accompanied by the participant ID, along with the intervention type (V=VR-based training; A=app-based training) and group assignment (1=EG; 0=CG). The analysis yielded the following overarching themes: *perceived treatment effects*, *mechanisms of action*, and *feedback on the intervention* (separate thematic analyses of the 4-year follow-up feedback yielded comparable results, with no meaningful differences in the thematic structure or content).

Theme 1: Perceived Treatment Effects

Overview

Participants described a broad range of perceived effects related to their participation, spanning positive smoking-related changes and reports of limited or absent effects.

Subtheme 1: Beneficial Smoking-Related Effects

Participants reported a variety of beneficial effects following the training. For instance, a notable number of participants described a reduced urge to smoke:

> *During the training, I did not feel any desire to smoke. I think that an intensive use of this training could indeed be beneficial and supportive.* [134V1; post]
>
> *This is great, I no longer have any desire to smoke.* [101V1; FU]

These subjective craving reductions were often accompanied by reports of a decreased daily cigarette consumption:

> *In general, I have noticed progress after training. I haven't completely cut down on smoking yet, but I have significantly reduced my consumption.* [89V1; post]
>
> *Everything is going well; it has helped me to reduce consumption by half so far.* [100V1; FU]

Moreover, some participants reported having ceased smoking following training:

> *I managed to quit smoking, certainly with the help of the VR training.* [7V1; post]

Others provided nonspecific accounts of beneficial training outcomes. These participants noted general improvements attributed to the training, for example:

> *So far, it has been helpful and enjoyable.* [100V1; post]

Subtheme 2: Insufficient Effects

Alongside the positive effects outlined above, not all participants consistently reported beneficial or sustained outcomes from the study or the training. For instance, some noted only temporary improvements:

> *After the second session, I experienced a temporarily reduction in my urge to smoke. Unfortunately, this effect did not persist in the following sessions. My urge to smoke and smoking behavior have not changed.* [50V1; FU]

Similarly, some participants noted that reductions observed during the training phase subsequently dissipated:

> *During the two weeks of training, smoking behavior changed significantly, but after the training program ended, the effect wore off, mainly due to stressors in the personal and professional environment.* [64A0; FU]

Moreover, a subset of participants reported experiencing no discernible effects:

> *Unfortunately, it had no effect whatsoever on my smoking habits.* [44V0; post]
>
> *I couldn't notice any change in myself.* [42V0; post]
>
> *Unfortunately, the training didn't do anything for me at all.* [129V0; post]

XSL•FO
RenderX

To summarize, participants experienced a diverse array of smoking-related outcomes in relation to their study participation. While some participants reported a reduction in cravings, smoking, or complete abstinence, others exhibited only transient changes or no change at all.

Theme 2: Mechanisms of Action

Overview

This overarching theme comprises participants' reflections on 2 distinct yet interconnected aspects of the training. First, it encompasses participants' insights into how the training may have exerted its effects on smoking behavior. Second, it addresses participants' perceptions of factors that may have limited or impeded the intended impact of the training.

Subtheme 1: Perceived Mechanisms of Action

Participants expressed a number of assumptions regarding the underlying mechanisms of action through which the training might have worked. Notably, some participants described processes that were closely aligned with the theoretical principles of ApBM. In particular, 1 participant attributed the training's effects to the contingency between smoking-related stimuli and the push movement, stating:

> *In my opinion, it helps you think about other things than SMOKING. It has already helped me, and in stressful situations I was able to simply push smoking away.* [53A1; post]

Consistently, several participants reported that the training helped them to engage in more reflective smoking behavior or that the training disrupted habitual patterns:

> *Through training and the reason for training, I perceive the topic of smoking and when to smoke differently.* [36V0; post]
>
> *I thought that it worked well to break routines, especially when used specifically during the urge to smoke a cigarette.* [25A1; post]

While the aforementioned participants alluded to mechanisms of action in close alignment with the theoretical underpinnings of ApBM, other participants referred to more general mechanisms of action. For instance, some described that the training provided a helpful distraction from smoking:

> *It's simply a good substitute for cigarettes at the beginning.* [53A0; FU]
>
> *The training is very helpful and shows me that I can become a non-smoker by 'distracting' myself from smoking objects or cravings.* [48V0; FU]

Moreover, for a subset of participants, simply dedicating time to the intervention was perceived as motivating. Notably, all of these participants took part in the VR-based training, where training sessions took place in a laboratory setting as opposed to the app-based training, which was conducted remotely.

> *Just going there had a positive effect on me.* [15V0; FU]
>
> *Overall, it was quite time-consuming to attend the 15-minute training sessions (travel time), but I also think that the considerable effort involved helped to motivate me to quit smoking. So, what seemed unpleasant and almost time-consuming at first was certainly what made it easier for me to quit in my case.* [91V1; post]

Lastly, some participants did not explicitly articulate assumptions about specific mechanisms of action, but expressed general confidence in the intervention:

> *The VR training certainly serves a purpose, even if I don't recognize it.* [53V0; post]
>
> *Since I am not qualified to assess it neurologically, I trust that it [the training] helps with smoking cessation and remain curious.* [3A0; post]

Subtheme 2: Impeding Factors

Concurrently, participants identified several factors that impeded the training's impact. These limiting factors refer to potential barriers to the implementation of ApBM within novel technological environments such as VR and app-based formats. Some factors perceived as limiting the training's effectiveness were directly related to the training design itself. For instance, a number of participants indicated that the absence of a contingency between smoking-related stimuli and push movements limited the training's impact on behavior. Notably, these participants were predominantly enrolled in the sham groups:

> *If I could just push all smoking-related objects away, the effect would perhaps be better?* [1A0; post]
>
> *I didn't understand why I had to pull almost all the smoking-related objects toward me during the app training. In my opinion, it would have been better if I had pushed all the objects related to smoking away from me.* [133A0; FU]

Similar concerns were reflected in notes from participants in the EG, who perceived the approach bias assessment at posttest to measure training-specific bias change as a barrier that limited the intervention's effectiveness:

> *After the last training session, the bias assessment from the pretest had to be repeated in order to evaluate reaction times. This involved pulling smoke objects towards me again. That ruined a lot of things. I found the effect very strong. In hindsight, I should have refused, following my first impulse.* [79V1; post]

In addition, many participants reported not fully understanding the purpose of the training, with some explicitly attributing this confusion to the lack of stimulus-response contingencies:

> *Since there was no need to consistently move smoke objects away from oneself or towards oneself, the purpose of the app was unclear.* [3A0; post]
>
> *Many images are pulled in one moment and pushed away the next. You'd think that makes little sense.* [136A0; post]
>
> *I didn't understand the point of the exercises, because I don't see how that's supposed to help.* [138A0; FU]

Likewise, some participants noted that the indirect instruction to respond to a content-irrelevant feature diverted their attention

XSL•FO
RenderX

from the stimulus content, which was associated with a lack of contingency awareness and, presumably, with a diminished training effect:

> *During the app training, I never saw the objects, only how the image was tilted.* [107A0; post]
>
> *I didn't care if they [the stimuli] were smoke objects... I only paid attention to red and blue [border color].* [86V0; post]

In addition to impeding factors that were directly related to the training design, some participants referred to barriers in relation to aspects of the training implementation. For instance, technical difficulties were frequently reported, with some participants also describing these issues as diminishing their motivation:

> *Unfortunately, my mobile phone freezes and the images cannot be swiped away quickly enough. This reduces my motivation.* [27A1; post]

Such technical difficulties may provide an explanation for the insufficient or inconsistent training effects experienced by some participants. Finally, a number of participants identified everyday life factors as barriers to training adherence. For instance, external stressors were reported to have an impact on the ability of participants to complete training sessions:

> *I work in a three-shift system. If I haven't done it [the training] in the morning before the afternoon shift, I tend to forget it.* [84V1; post]
>
> *Unfortunately, due to time constraints, I only completed the app training once. However, I noticed that my craving to smoke was suddenly reduced afterwards.* [135A0; post]
>
> *In my opinion, the planning/coordination was difficult for working people to implement due to the specified appointed times, etc.* [42V0; FU]

Closely related to this barrier, 1 participant observed that the training exhibited only limited transferability to everyday life:

> *In everyday life, "habits" still need to be broken.* [48V0; post]

In conclusion, participants indicated divergent assumptions regarding the mechanisms of action, spanning from more specific and theoretically grounded processes, such as actively pushing smoking-related cues away, to more general mechanisms, such as motivational effects and a general confidence in training effects. Notably, these specific assumptions were also reflected in perceived barriers, with participants in the sham condition reporting a desire to avoid smoking-related cues and some participants in the ApBM condition deploring reduced training effectiveness due to the posttest approach bias assessment, which removed the stimulus-response contingency. Moreover, technical problems and daily hassles were associated with limited training implementation and adherence to the training regimen. As accounts of factors impeding the training's efficacy were given frequently, participants' feedback on the intervention may provide insight into how to address such barriers in future applications of ApBM.

Theme 3: Feedback on the Intervention

Overview

This theme encompasses participants' feedback on both the study and the training. Subthemes comprise positive feedback, suggestions for improvement, and criticism, which were all divided into both training-specific reflections as well as general study experiences.

Subtheme 1: Training-Specific Positive Feedback

Participants reported training-specific positive feedback, including overall satisfaction with the training, enjoyment, favorable impressions of the stimuli, ease of use, and a desire to continue engaging with the training. Moreover, certain participants emphasized the efficacy of the training in conjunction with the TAU prior to randomization:

> *Certainly not sufficient on its own for immediate smoking cessation, but definitely useful as a supplement to the [self-help] book, for example.* [20A0; post]

In addition, some participants underscored the training's potential applicability to other domains:

> *The training is very helpful and shows me that I can become a non-smoker by 'distracting' myself from smoking objects or cravings. Yes, this training can also be applied to many other areas of life. Thank you very much for this.* [48V0; FU]

Subtheme 2: General Positive Feedback

Many participants additionally expressed satisfaction with the study, reporting gratitude and enjoyment. Furthermore, participants advocated for the continuation of the research:

> *Everything was perfect and very well organized. Thank you very much.* [101V1; post]
>
> *I really enjoyed it and it helped me reduce my cravings. Thank you very much!* [26V0; post]

Subtheme 3: Training-Specific Suggestions for Improvement

Beyond the positive feedback, participants provided ideas on several training-specific improvements. These included suggestions for technical enhancements, the use of direct instructions (ie, content-relevant feature training), and incorporating a larger variation in training scenarios and/or stimulus material:

> *Since the objects in the VR training are color-coded (red/blue), I find myself automatically looking only at the colors and not at the objects themselves. I think that if it were possible in terms of programming, I would stop assigning colors to the objects after the third training session so that you only have to focus on the object without the color dictating where the object has to go. I think that would be better for the effect of reacting and acting.* [34V1; post]
>
> *I would like to see more training sessions, possibly in different environments.* [4V1; FU]

XSL•FO
RenderX

Additionally, participants recommended decreasing the number of trials per session while increasing the number of training sessions:

> *The duration is a bit too long in some cases, but I would have liked to use it more often.* [95A0; FU]
>
> *Personally, I think the training with 200 images takes too long. [...]. I'd prefer the option of completing the training 2 times with 100 images instead.* [112A1; post]
>
> *I think 6 VR sessions are not enough.* [76V1; FU]

Consistently, participants also expressed a preference for an extended training interval:

> *The number of VR training sessions would have to take place at very regular intervals and over a longer period of time, then I would imagine it having an effect.* [65V0; FU]

Furthermore, participants consistently articulated the need for a better comprehension of the theoretical underpinnings of the training:

> *I would like to continue the app training and learn more about the background and scientific findings, ie, the specific goals and behavioral changes achieved through the training.* [145A1; FU]

Additionally, participants offered a plethora of recommendations regarding specific training application and/or implementation. For instance, some participants in the VR study assumed the training to be more effective when performed at home:

> *It would certainly be more effective if it took place more often, or if you could do it from the comfort of your own home.* [66V1; post]
>
> *VR training should be carried out daily, possibly programmed into an app.* [26V0; FU]

Other participants recommended integrating the training with additional interventions, such as:

> *I would also have liked to see more variety, eg, a stay in disgusting, dirty smoking rooms compared to a fresh flower meadow; [...]. Adding smells would certainly be successful.* [13V1; post]

Finally, participants put forth the notion of applying the training in specific situations and contexts. For instance, they noted its high use as a substitute for a cigarette:

> *It would be great if I could use the training more often to replace a cigarette; else, I tried to manage it without the app.* [95A0; post]

Subtheme 4: General Suggestions for Improvement

Participants additionally offered general suggestions to enhance the training's impact on behavior. For instance, they expressed a desire for greater interaction and exchange with both fellow training participants and therapists:

> *Possibly [introduce] more training sessions and also discussions about [smoking] cessation* [46V1; post]
>
> *I would have liked personal conversations.* [155A; FU]
>
> *It would be beneficial to introduce group sessions to the VR training so that participants can exchange ideas with others. This helps to motivate.* [49V1; FU]

Subtheme 5: Training-Specific Criticism

Some participants also reported training-specific criticisms, highlighting its monotony and questioning its effectiveness as a stand-alone intervention:

> *I have the impression that the exercises should be modified with each additional training session, as in some cases they simply become monotonous.* [27V1; post]
>
> *After an initial learning phase, boredom sets in. A little more additional information would be desirable. Perhaps something encouraging, such as a counter for different things that improve over the course of quitting smoking, eg, how much CO2/tar/nicotine you have not inhaled/absorbed. An anonymous chat with like-minded people would also be nice* [136A0; post]

Subtheme 6: General Criticism

In addition, some participants expressed general criticism:

> *I cannot imagine that a long-term smoker would be motivated to quit smoking through VR training. At least not me, so I stopped after the first session!* [33V1; post]

Taken together, participants offered both positive feedback, including enjoyment and ease of use, and recommendations for improving the training. These recommendations included a higher training dosage, the implementation of direct task instruction, and the adaptation of the training for different contexts and situations. Some participants also voiced critiques of the study and training, particularly with respect to the effectiveness of the training as a stand-alone intervention.

Differences in perceived outcomes attributed to the training may reflect variations in how participants understood and experienced the training's underlying mechanisms, as well as how these were shaped by training-specific features and individual or contextual factors. In particular, the stimulus-response contingency appears to play a central role in perceived meaningfulness and training efficacy, as several participants who experienced beneficial outcomes explicitly attributed these improvements to the stimulus-response contingency. Conversely, participants in the sham conditions linked insufficient effects to the absence of such a contingency. Furthermore, the mode of delivery also emerged as relevant: participants in the VR study described the training as motivating, enjoyable, and interesting, which aligns with the rationale for implementing ApBM in immersive environments. Likewise, participants in the app study appreciated the pictorial stimuli and expressed a wish to use the app more frequently in daily contexts. Building on this, participants' descriptions of impeding factors and their suggestions for improvement provide further insights into why some individuals did not experience beneficial outcomes. Several participants reported technical problems during training, indicating that the implementation of technically innovative ApBM still needs to be refined. Others advocated for direct task instructions as well as a higher training variety

XSL•FO
RenderX

and availability. The latter may be linked to reports of external stressors that diminished adherence to the training, thereby limiting its potential effectiveness. Finally, participants also recommended combining ApBM with other established interventions, assuming that they may benefit from multimodal approaches.

To complement the qualitative analysis, we quantified how often each theme was mentioned and explored whether frequencies differed between the EG and CG. These additional analyses are displayed in Table 4. Please note that due to the open response format, absence of mention does not indicate absence of experience. Inferential statistics should therefore be interpreted with caution and are used here only for exploratory purposes.

Table 4. Frequencies of reported themes and subthemes by group.[a]

Themes and subthemes	Immediate feedback: posttest and 7-week follow-up				Long-term feedback: 4-year follow-up			
	Overall, n	EG[b] (ApBM[c]), n (%)	CG[d] (sham), n (%)	*P* value	Overall, n	EG (ApBM), n (%)	CG (sham), n (%)	*P* value
Feedback provided	87	45 (29)	42 (24)	N/A[e]	63	33 (26)	30 (19)	N/A
Theme 1: perceived treatment effects								
Beneficial smoking-related effects	25	17 (13)	8 (5)	.05	33	21 (19)	12 (9)	.06
Insufficient effects	13	7 (5)	6 (5)	.87	15	6 (4)	9 (5)	.27
Theme 2: mechanisms of action								
Perceived mechanisms of action	24	14 (7)	10 (6)	.45	9	6 (6)	3 (3)	—[f]
Impeding factors	53	19 (10)	34 (16)	<.001	9	5 (4)	4 (4)	—
Theme 3: feedback on the intervention								
Training-specific positive feedback	19	11 (4)	8 (5)	.54	12	8 (6)	4 (2)	.27
General positive feedback	10	7 (5)	3 (2)	—	28	18 (12)	11 (7)	.16
Training-specific suggestions for improvement	54	35 (22)	19 (10)	.002	2	0	2 (1)	—
General suggestions for improvement	4	2 (2)	2 (0)	—	6	3 (3)	3 (2)	—
Training-specific criticism	6	4 (4)	2 (0)	—	2	2 (2)	0	—
General criticism	1	1 (2)	0	—	1	0	1 (1)	—
Theme 4: attribution of the effect								
Attribution on the intervention	—	—	—	—	7	3 (2)	4 (3)	—
Attribution on treatment as usual only	—	—	—	—	4	1 (1)	3 (2)	—
Attribution on external factors	—	—	—	—	14	7 (7)	7 (5)	.84

[a]The number of mentions in relation to the virtual reality trial is shown in brackets. Differences between groups were analyzed using chi-square tests.

[b]EG: experimental group.

[c]ApBM: approach bias modification.

[d]CG: control group.

[e]N/A: not applicable.

[f]In cases where the expected cell frequencies fell below 5, no statistical tests were conducted.

As shown in Table 4, participants in the CG reported impeding factors significantly more often (χ^2_1=13.69; *P*<.001). A closer inspection of these comments indicated that most of these hindering factors centered on complaints about the lack of stimulus-response contingency and the perceived lack of meaningfulness of the training. Conversely, participants in the EG provided more suggestions for improving the training (χ^2_1=9.77; *P*=.002). Several interpretations may account for this pattern. For example, given that no group differences emerged in positive, training-specific feedback, the higher frequency of improvement suggestions in the ApBM groups may reflect a stronger cognitive engagement with the training procedures. Alternatively, it is possible that the more coherent and meaningful training experience in the EG facilitated participants' ability to generate concrete suggestions for refinement, whereas the relative opacity of the sham condition may have limited such reflections. However, these interpretations remain speculative, as the study was not designed to systematically probe the underlying reasons for this difference. No other group differences reached statistical significance.

XSL•FO
RenderX

Participants' Long-Term Feedback

Overview

In total, 63 participants provided open-ended comments at the final 4-year follow-up, of whom 33 were in the EG, and 30 were in the CG. The thematic structure largely mirrored the patterns observed for proximate feedback, encompassing *perceived treatment effects*, *mechanisms of action*, and *feedback on the intervention*. Each of these themes again comprised several subthemes comparable to those identified in the proximate feedback. Due to this substantial resemblance, detailed descriptions of the 3 recurring themes at the 4-year follow-up will not be presented here but are provided in the Multimedia Appendix 1. Notably, an additional overarching theme, *attribution of effects*, emerged uniquely at the 4-year follow-up, reflecting participants' retrospective reflections on what they believed to have contributed to long-term changes. A comprehensive overview of all themes, subthemes, and their interrelations is presented in Figure 1.

Theme 4: Attribution of Effect

Overview

Responses at the 4-year follow-up revealed an additional theme related to participants' attribution of treatment effects. Whereas earlier comments focused primarily on describing perceived changes, participants at the long-term follow-up engaged in a more elaborated meaning-making process by attributing their smoking-related outcomes—whether positive, limited, or absent—to specific causal sources. These attributions provide insight into how individuals retrospectively interpreted and contextualized their behavior change over an extended period.

Subtheme 1: Attribution to the Intervention

Several participants explicitly credited the training for their improvements, highlighting mechanisms or elements of the intervention they perceived as effective in supporting abstinence or reducing cigarette consumption:

> *Keep going. I am grateful for the app because it helped me to quit [smoking].* [53A1; online assessment]

> *Taking part in the study was very important for me, even though it took a long time until I finally managed to break away from nicotine. Taking part in the study also made me realize how harmful nicotine use is, which ultimately played a key role in my decision to quit smoking altogether. I would like to express my sincere thanks once again.* [133V1; online assessment]

While some participants referred broadly to the intervention as a whole, others explicitly attributed their outcomes to the combination of the training and TAU:

> *Everything was great; the book helped the most; the app was more of a supportive tool.* [133A0; telephone assessment]

Subtheme 2: Attribution to TAU Only

Some participants in the sham condition attributed their outcomes primarily to the support and guidance provided through TAU, namely, the bibliotherapy that was prescribed to all participants prior to randomization:

> *In retrospect, the book was very helpful; [I] only read it after the intervention had ended and then stopped.* [57A0; telephone assessment]

> *Because of the intervention, I succeeded in becoming a non-smoker. The book "Endlich Nichtraucher" [The Easy Way to Quit Smoking] contributed to this. The book was additionally handed out at the beginning of the study.* [68V0; online assessment]

Subtheme 3: Attribution to External Factors

A subset of participants attributed their smoking-related changes to factors outside the study context. These external attributions encompassed a range of factors that participants perceived as more decisive than the training itself. Medical issues were frequently cited as strong motivators for cessation:

> *[I had] a lung cancer surgery. Since then, I haven't touched a cigarette.* [115A0; telephone assessment]

> *It was a very good experience, which unfortunately did not work for me, but I still managed it without any aids, because a serious illness that was related to smoking was the reason; and it's just a matter of mindset and absolute willpower.* [100V1; online assessment]

In addition, some participants emphasized significant life circumstances such as the birth of a grandchild or the wishes of significant others as meaningful catalysts for change:

> *It didn't work because of that [the study], but only later due to an illness and the birth of my grandchild.* [63V0; telephone assessment]

Some participants reported relying on other interventions, including additional cessation aids or self-initiated strategies outside the study protocol:

> *I did not smoke during the study, but unfortunately, I started again two weeks later. I made another attempt in mid-2023 with Fumexan smoking cessation therapy (injections into the ear) and it has worked quite well so far.* [8V1; online assessment]

Finally, 1 participant pointed to reasons unrelated to the training, suggesting that personal decisions or timing were more influential in determining their outcome than the intervention:

> *[I have] quit independently of the study 1.5 years ago.* [93V0; telephone assessment]

In summary, participants expressed a wide range of attributions regarding their long-term smoking-related outcomes. While some linked beneficial effects to the training and/or TAU, others pointed to additional or different factors outside the study context. Such attribution patterns underscore the complexity of smoking cessation processes and the varied meaning-making practices through which participants explained change long after the intervention had ended.

To quantify how often each theme was mentioned across groups, counts of reported themes and subthemes for the 4-year follow-up are also presented in Table 4. Similar to the proximate

XSL•FO
RenderX

feedback, participants in the EG appeared to report more beneficial smoking-related outcomes; however, this difference did not reach statistical significance (χ^2=3.52; *P*=.06). No group differences emerged for the remaining themes, including effect attribution (P_s>.16).

Discussion

Principal Findings

Overview

This study sought to elucidate participants' subjective experiences with 2 newly developed ApBM programs for smoking cessation, delivered either via VR or a smartphone app. In addition, the study examined factors perceived to contribute to beneficial or insufficient training effects, as well as retrospective reflections on the training that were obtained 4 years after study inclusion, alongside proximate feedback collected up to 7 weeks after the intervention. The thematic analysis identified four overarching themes, with the fourth emerging only in the long-term follow-up: (1) *perceived treatment effects*, ranging from positive smoking-related changes to a perceived lack of effects; (2) *mechanisms of action*, encompassing reflections on presumed working mechanisms as well as factors perceived as impeding effectiveness; (3) *feedback* on the training and the broader study experience; and (4) *attribution of effect*, referring to participants' attempts to attribute behavioral change to training-specific or external influences.

Core Themes in Participants' Experiences: Links to Prior Research and Novel Insights

With regard to our first research question, the overarching themes identified in this study are broadly comparable to those reported in previous qualitative studies investigating participants' subjective perceptions of app-based psychotherapeutic interventions. For instance, Tudor-Sfetea et al [48] examined participants' perceptions of a smoking cessation app based on cognitive behavioral therapy principles and identified themes such as "*application effects*," comparable to the present theme of "*perceived treatment effects*," as well as "*application content*" related to style, quality, and engagement, which partially overlaps with the current theme of "*mechanisms of action*." At the same time, some themes appear to be specific to the present investigation, particularly "*attribution of effects*." This difference may be attributable to the short-term assessment focus (1 week) in Tudor-Sfetea et al [48], compared to the additional long-term follow-up in this study.

More specifically, in the context of CBM, Kuckertz et al [29] investigated participants' perceptions of 2 computer-based ABM interventions for anxiety and identified themes including "*perceived symptom change*," "*mechanisms of action*," "*acceptability and feedback*," and "*barriers and facilitators to engagement*." Similarly, Prior et al [34] asked adults with lived experience of hazardous alcohol use and social anxiety to evaluate a combined web-based ApBM and CBM-I intervention and revealed themes related to "anticipated clinical value," "motivation and engagement," and "perceived usefulness." Thus, it appears that similar aspects are reported across different CBM training programs, which are not solely specific to ApBM or the training delivery method used. Consistent with other studies, our findings indicate that participants engage in complex reflection on the intervention that is difficult to capture through quantitative assessment methods alone. In the context of psychotherapy research, such treatment expectancies and their perceived potential to contribute meaningfully to improvement constitute a well-established common factor to treatment efficacy [49] and have been proven to predict treatment outcome [50]. In this trial, this dynamic was most clearly reflected in Theme 1 (*perceived treatment effects*) and Theme 2 (*mechanisms of action*). Participants reported a range of perceived outcomes following the training, including reductions in craving, decreases in cigarette consumption, and complete abstinence. Importantly, these perceived outcomes were closely connected with reflections on how the training was thought to work. Anticipated mechanisms of action included the existence of a stimulus-response contingency (ie, smoke-push trials), the disruption of automated action tendencies, and the strengthening of reflective processes related to smoking behavior. While these reflections are broadly consistent with both dual-process accounts of addiction [51] and the theoretical rationale underlying ApBM [11,13], they should be interpreted as subjective explanations rather than evidence of actual mechanisms, as findings in this regard remain mixed. For example, training-specific bias change or mediation effects have not been consistently observed (for a review and a meta-analysis, see Boffo et al [24]), suggesting that additional processes, potentially not accessible to subjective reflection, may also contribute to ApBM effects and remain to be further elucidated in CBM research.

By contrast, a subset of participants in the sham training groups reported difficulties in making sense of the task, especially due to the absence of a clear stimulus-response contingency. This lack of perceived meaningfulness was often accompanied by frustration and, in some cases, explicitly linked to insufficient training effects. Such accounts may also help explain the group differences observed in quantitative training evaluation ratings.

Germane to this, several participants expressed a desire for a clearer and more coherent theoretical integration of the training rationale (see Theme 3: *feedback*). In this study, the brief smoking cessation intervention (TAU) explained the impact of reflective and impulsive processes on smoking behavior, thereby introducing the rationale underlying the training as a means to target the latter. In addition, a cover story was used to convey the basic logic of the training task. Thus, participants were provided with basic explanations of the training's purpose, which is consistent with recommendations emerging from prior qualitative research (eg, [34]). Nevertheless, these findings suggest that this level of explanation may not be sufficient and that a more sustained integration of the training rationale is warranted. In this instance, Kuckertz et al [29] compared a traditional, minimally explained ABM intervention with an "enriched" version that included more detailed explanations of the training's theoretical background. Contrary to expectations, however, qualitative analyses revealed that participants in both

XSL•FO
RenderX

conditions perceived the explanations as insufficient. Rather than receiving a single explanation at the outset, participants expressed a preference for repeated reminders of the training rationale and opportunities to discuss the training with clinicians during the program. This resembles established psychotherapeutic practices, in which treatment rationales are revisited and linked to patients' everyday experiences and goals, and suggests a potential avenue for enhancing the perceived credibility of CBM interventions.

Perceptions of Digital Delivery Formats

With regard to the second research question, another important issue emerging from this study concerns participants' perceptions of the novel delivery formats used for ApBM, namely VR and smartphone apps. While a few participants described the training as repetitive or suggested that it could be improved through greater stimulus variety, which is consistent with earlier reports on CBM task perceptions [27], others reported high levels of motivation and engagement. At the same time, several participants noted technical difficulties, which in some cases undermined motivation and disrupted training adherence. These findings highlight that the implementation of novel technologies may entail trade-offs, particularly when technological infrastructures are still under development. Importantly, prior CBM research suggests that more technologically advanced training approaches do not necessarily translate into better training perceptions. For instance, studies incorporating gamification elements [15], personalized stimuli [31], or adaptive designs [34] have reported mixed results with regard to user engagement and acceptability. Accordingly, efforts to enrich ApBM interventions should be undertaken cautiously, ensuring technical feasibility, and should not come at the expense of other critical factors such as participant support and the integration of the training into a broader theoretical or therapeutic framework (see the Core Themes in Participants' Experiences: Links to Prior Research and Novel Insights section above).

Correspondence Between Short- and Long-Term Feedback

Regarding the third research question, comparisons between proximate and long-term feedback indicated that participants' perceptions of the training were largely stable over time, as the 3 overarching themes identified at short-term follow-up were still evident 4 years later. Notably, a fourth theme emerged at the long-term follow-up concerning the *attribution* of beneficial outcomes, most commonly abstinence. Over time, participants appeared to develop a need to make sense of their behavior change, attributing it either to the training itself or to broader life circumstances. These attributions were almost evenly distributed, with approximately half of the participants explicitly emphasizing the training and/or the overall study experience as the primary reason for abstinence, while the other half referred to health- or family-related factors. Given that attribution was not explicitly probed, it is noteworthy that some participants spontaneously cited the training as causal even after such a prolonged period. As attributing change to internal or treatment-related rather than external factors is considered important for sustaining therapeutic gains [52,53], future research should systematically examine attribution processes in CBM and explore how training-related attributions can be fostered, for example, by providing and repeatedly reinforcing a theoretically and empirically grounded training rationale aligned with participants' goals (see the Core Themes in Participants' Experiences: Links to Prior Research and Novel Insights section above).

Principal Contributions

This study contributes to the literature in several ways. First, it expands prior qualitative work on CBM by examining participants' training experiences over an extended follow-up, spanning several years. Such a long time frame is rare in both CBM and psychotherapy research. Moreover, a relatively high response rate was achieved overall (104/178 participants provided feedback at least 1 assessment time point). In addition to the dual-contact approach used at follow-up (ie, telephone and online contact; see Methods), features of the original RCTs may have supported sustained engagement, including the predominantly in-person study procedures, which likely facilitated personal contact, rapport building, and peer interaction. The somewhat higher response rate in the VR-based ApBM condition further suggests that the immersive, on-site training format may have enhanced participant engagement. Moreover, the trials were explicitly framed as investigations of novel digital smoking cessation interventions, with an emphasis on both short- and long-term experiences communicated during the active study phase, which may have increased commitment to participate in the long-term follow-ups.

Second, although the overarching themes broadly align with prior qualitative work on CBM and related digital interventions, these findings provide additional insight into how different training conditions are experienced. Notably, participants in the sham condition frequently reported difficulties understanding the training rationale and, in some cases, explicitly criticized the absence of a stimulus-response contingency. This finding is particularly relevant for CBM research, as it suggests that blinding may be challenging when control conditions closely resemble the active training but lack its core mechanism. Such control conditions may not function as neutral comparators but could, in some cases, evoke frustration or reduced credibility, potentially influencing outcomes. At the same time, sham training was not consistently evaluated more negatively in this sample, and prior meta-analytic evidence indicates that sham interventions may still produce meaningful effects, particularly at higher training dosages [24]. Together, these findings highlight the ongoing challenge of identifying appropriate control conditions in CBM research and suggest that alternative comparator designs may warrant further investigation.

Third, these findings suggest that participants' perceptions of ApBM remain relatively stable over time, while also giving rise to retrospective processes of sense-making. In particular, participants reflected on whether sustained behavior change could be attributed to the intervention or to external factors. Due to the relatively small number of spontaneous attribution-related statements, it was not possible to examine predictors of attribution or their relationship with long-term

XSL•FO
RenderX

outcomes. This represents an important avenue for future research.

Limitations and Future Directions

Alongside notable strengths, several limitations warrant consideration, which highlight relevant directions for future CBM research. First, the relatively high overall response rate may reflect both sustained motivation for smoking cessation and a willingness to contribute to research in this area. Hence, findings may not fully generalize to less intrinsically motivated or less engaged populations. While ApBM is often assumed to be more effective in motivated individuals or clinical samples, systematic evidence on the role of motivation in CBM outcomes is still limited. Future work should, therefore, examine whether motivation moderates ApBM effects or whether motivational processes can themselves be targeted by CBM principles, an issue that remains largely unresolved in the current literature.

At the same time, perspectives from a notable subset of the sample remained unavailable, especially at the 4-year follow-up. Exploratory analyses comparing participants who provided feedback at any time point with those who did not on key demographic, intervention-related, and smoking-related variables indicated largely comparable groups. However, selective attrition and survivorship bias cannot be fully excluded. For example, it remains possible that participants who provided long-term feedback were those for whom the intervention was more salient (ie, who were particularly satisfied or dissatisfied), which may have influenced both the qualitative themes and the interpretation of descriptive outcome data. Relating to this, although participants were encouraged to provide both positive and negative feedback, and the thematic analysis identified a range of critical next to favorable perspectives, responses may still have been influenced by demand effects. Hence, incorporating multimodal assessments beyond self-report, such as behavioral indicators or momentary assessments, may be a fruitful direction for future research to reduce possible bias and enhance the validity of findings.

Third, the thematic analysis was based on unprompted, open-ended responses. As participants decided independently which aspects of their experience to report, the absence of a particular theme cannot be equated with the absence of the corresponding experience. This limitation may be particularly relevant for the theme of *attribution of effects*, which emerged only at the 4-year follow-up. Moreover, the open-ended question used to elicit feedback was phrased in broad terms. While some participants provided in-depth reflections on the training and its potential mechanisms, others offered more general evaluations, resulting in considerable variability in the specificity of responses. Future research could therefore benefit from targeted qualitative or mixed methods approaches—for instance, semistructured interviews—which would allow for a more systematic examination of attribution processes over time.

Implications for ApBM Task Design and Implementation

These findings offer important implications for the design and implementation of future CBM interventions. These relate primarily to (1) the communication of the training rationale and management of expectancy effects, (2) the optimal dosage and contextual embedding of the training, and (3) specific design features of the training tasks.

First, participants' requests for a stronger theoretical integration of the training, combined with greater therapeutic support, highlight the importance of how the training rationale is communicated. While many CBM trials deliberately minimize expectancy effects to enhance internal validity (ie, by providing limited information about the training's rationale or theoretical background), implementation-oriented studies aiming to maximize clinical benefit may need to adopt a different strategy [54]. In clinical and medical settings, it is well established that providing patients with clear information about an intervention, along with conveying confidence and therapeutic rationale, can meaningfully influence outcomes [55]. Accordingly, greater emphasis on external validity in CBM research, including ApBM, may be warranted, particularly when interventions are delivered to clinical or affected samples. Such an approach may not only enhance training efficacy but may also be in better accord with participants' expectations and preferences. Future work may, therefore, incorporate more transparent and evidence-informed rationales, for example, by referencing prior findings on ApBM effectiveness, to enhance engagement, perceived training relevance, and training-related effect attributions.

Second, participants frequently suggested increasing training dosage and embedding the intervention more strongly within supportive contexts, such as peer exchange. These suggestions are in line with both dual process accounts of addiction [51], which stress the importance of targeting both reflective and impulsive processes to contribute to sustained behavioral change, as well as empirical findings that show that multisession CBM as an add-on to regular treatment tends to be more effective than single-session or stand-alone trainings [24,25]. At the same time, the number and spacing of training sessions remain to be determined. In the context of ApBM for AUD, one study suggests that 6 training sessions may be optimal in terms of learning effects [25], which informed the design of the present interventions, particularly the VR-based training. However, further research is needed to establish optimal training dosage for ApBM in the context of smoking cessation. Furthermore, participant suggestions regarding context-sensitive training use (eg, during craving or before or after quitting attempts) align with ongoing developments in the field [22,56] and further highlight the potential value of involving participants in the iterative development of CBM interventions (see Prior et al [34], for an encouraging example of a participatory codevelopment approach for an ApBM + CBM-I online intervention). Taken together, task design decisions should integrate evidence-based mechanisms of ApBM with participants' experiential insights regarding how the intervention is perceived to work, as optimizing outcomes may depend on aligning these complementary perspectives.

Third, participants' feedback also pointed to specific design features of CBM tasks. In particular, some participants reported that the indirect task instruction (ie, react to a content-irrelevant feature) prevented them from paying attention to the stimulus content. As a result, they suggested that direct task instructions

XSL•FO
RenderX

(ie, approach-avoid movements explicitly based on stimulus content) might enhance training effectiveness. Traditionally, CBM paradigms rely on indirect instructions, requiring participants to ignore stimulus content and respond instead to content-irrelevant features. Empirical evidence directly comparing instruction types is limited. Yet, one study contrasted content-relevant (emotional) vs content-irrelevant (color-based) instructions and showed that emotional stimulus features were still processed when being task-irrelevant [57], underscoring the main assumptions of CBM [50,58] and suggesting that indirect instructions are not inherently inferior despite participants' reservations. Importantly, such findings also highlight that participant accounts can inform the refinement of intervention design, but that proposed adaptations require systematic empirical testing. At the same time, evidence suggests that training effects under indirect instructions may be enhanced by increasing participants' awareness of the stimulus content [57]. Such findings show that participant-informed design adaptations, such as increasing the salience of contingencies without abandoning indirect paradigms, may represent promising avenues for future task adjustments but require systematic testing.

Conclusions

To conclude, this trial demonstrates that participants' subjective experiences constitute a central dimension of how ApBM for smoking is perceived and engaged with. By assessing both proximate (up to 7-week follow-up) and long-term (4-year follow-up) perspectives, the findings highlight that behavior change is embedded in an ongoing process of meaning-making that extends beyond the duration of the training. Future research should continue to place emphasis on participants' perceptions and incorporate them into the development of ApBM protocols by adopting participatory and patient-centered approaches. Such efforts may help to enhance both the effectiveness and acceptability of the intervention by better aligning training procedures with participants' needs and experiences.

Acknowledgments

We thank Luka Koep, Sabine Stähler, Nik Hulsmans, Lukas Flöter, Verena Born, Carolin Kraus, Daphne Heidecke, and Nicole Nagy, who assisted with data collection. We would also like to express our sincere gratitude to all smoking individuals who have set themselves the goal of quitting smoking and participated in this extensive study. Machine tools were used for refinement and to improve the clarity of language.

Funding

This study receives funding from the German Ministry of Culture and Science of the Land of North Rhine-Westphalia (grant 005-1706-0006). The funding source was not involved in study design, the collection, analysis, and interpretation of data, the writing of the report, or the decision to submit the paper for publication.

Data Availability

The dataset used and analyzed during this study will be made available at Open Science Framework [59].

Authors' Contributions

Conceptualization: AM, TK
Data curation: AM, MN
Formal analysis: AM, ES, JB
Writing – original draft: AM, ES, JB
Writing – review & editing: TK, MN
All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed information on the first three overarching themes—(1) perceived treatment effects, (2) mechanisms of action, and (3) feedback on the intervention—based on 4 participants' long-term (4-year) follow-up responses.
[PDF File (Adobe PDF File), 275 KB - humanfactors_v13i1e93238_app1.pdf]

References

1. Dai X, Gakidou E, Lopez AD. Evolution of the global smoking epidemic over the past half century: strengthening the evidence base for policy action. Tob Control 2022;31(2):129-137. [doi: 10.1136/tobaccocontrol-2021-056535] [Medline: 35241576]

XSL•FO
RenderX

2. Hughes JR, Helzer JE, Lindberg SA. Prevalence of DSM/ICD-defined nicotine dependence. Drug Alcohol Depend 2006;85(2):91-102. [doi: 10.1016/j.drugalcdep.2006.04.004] [Medline: 16704909]
3. VanFrank B, Malarcher A, Cornelius ME, Schecter A, Jamal A, Tynan M. Adult smoking cessation - United States, 2022. MMWR Morb Mortal Wkly Rep 2024;73(29):633-641. [doi: 10.15585/mmwr.mm7329a1] [Medline: 39052529]
4. WHO report on the global tobacco epidemic, 2019: offer help to quit tobacco use: executive summary. World Health Organization. 2019. URL: https://www.who.int/publications/i/item/WHO-NMH-PND-2019.5 [accessed 2026-05-29]
5. Hajek P, Stead LF, West R, Jarvis M, Lancaster T. Relapse prevention interventions for smoking cessation. Cochrane Database Syst Rev 2009(1):CD003999. [doi: 10.1002/14651858.CD003999.pub3] [Medline: 19160228]
6. Zhou X, Nonnemaker J, Sherrill B, Gilsenan AW, Coste F, West R. Attempts to quit smoking and relapse: factors associated with success or failure from the ATTEMPT cohort study. Addict Behav 2009;34(4):365-373. [doi: 10.1016/j.addbeh.2008.11.013] [Medline: 19097706]
7. Chu S, Feng L, Jing H, Zhang D, Li J, Meng X, et al. Efficacy of smartphone apps used alone or with traditional interventions for smoking cessation: a systematic review and meta-analysis. BMJ Evid Based Med 2026:bmjebm-2025-113971. [doi: 10.1136/bmjebm-2025-113971] [Medline: 41529951]
8. Kakoschke N, Kemps E, Tiggemann M. Approach bias modification training and consumption: a review of the literature. Addict Behav 2017;64:21-28. [doi: 10.1016/j.addbeh.2016.08.007] [Medline: 27538198]
9. Vrijsen JN, Grafton B, Koster EH, Lau J, Wittekind CE, Bar-Haim Y, et al. Towards implementation of cognitive bias modification in mental health care: state of the science, best practices, and ways forward. Behav Res Ther 2024;179:104557. [doi: 10.1016/j.brat.2024.104557] [Medline: 38797055]
10. Pan T, Zhao X, Bartoš F, Larsen H, Manning V, Boffo M, et al. Cognitive bias modification in alcohol use disorder and problematic drinking: a revised and updated IPD Bayesian meta-analysis. Clin Psychol Rev 2026 Mar;124:102709 [FREE Full text] [doi: 10.1016/j.cpr.2026.102709] [Medline: 41653812]
11. Wittekind CE, Rinck M, Reinout W. Cognitive bias modification as add-on to the treatment of substance use disorders. In: The Sage Handbook of Addiction Psychology. Thousand Oaks, CA: SAGE Publications; 2025.
12. Eberl C, Wiers RW, Pawelczack S, Rinck M, Becker ES, Lindenmeyer J. Approach bias modification in alcohol dependence: do clinical effects replicate and for whom does it work best? Dev Cogn Neurosci 2013;4:38-51 [FREE Full text] [doi: 10.1016/j.dcn.2012.11.002] [Medline: 23218805]
13. Loijen A, Vrijsen JN, Egger JI, Becker ES, Rinck M. Biased approach-avoidance tendencies in psychopathology: A systematic review of their assessment and modification. Clin Psychol Rev 2020;77:101825. [doi: 10.1016/j.cpr.2020.101825] [Medline: 32143108]
14. Rinck M, Wiers RW, Becker ES, Lindenmeyer J. Relapse prevention in abstinent alcoholics by cognitive bias modification: clinical effects of combining approach bias modification and attention bias modification. J Consult Clin Psychol 2018;86(12):1005-1016. [doi: 10.1037/ccp0000321] [Medline: 30507226]
15. Salemink E, de Jong SRC, Notebaert L, MacLeod C, Van Bockstaele B. Gamification of cognitive bias modification for interpretations in anxiety increases training engagement and enjoyment. J Behav Ther Exp Psychiatry 2022;76:101727. [doi: 10.1016/j.jbtep.2022.101727] [Medline: 35217211]
16. Wiers RW, Eberl C, Rinck M, Becker ES, Lindenmeyer J. Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. Psychol Sci 2011;22(4):490-497. [doi: 10.1177/0956797611400615] [Medline: 21389338]
17. Kiefer F, Hoffmann S, Petersen KU. S3-Leitlinie Screening, Diagnose und Behandlung alkoholbezogener Störungen. Berlin: Springer; 2022.
18. Haber PS, Riordan BC, Winter DT, Barrett L, Saunders J, Hides L, et al. New Australian guidelines for the treatment of alcohol problems: an overview of recommendations. Med J Aust 2021;215 Suppl 7:S3-S32. [doi: 10.5694/mja2.51254] [Medline: 34601742]
19. Baird S, Rinck M, Rosenfield D, Davis M, Fisher J, Becker E, et al. Reducing approach bias to achieve smoking cessation: a pilot randomized placebo-controlled trial. Cogn Ther Res 2017;41(4):662-670 [FREE Full text] [doi: 10.1007/s10608-017-9835-z]
20. Machulska A, Zlomuzica A, Rinck M, Assion HJ, Margraf J. Approach bias modification in inpatient psychiatric smokers. J Psychiatr Res 2016;76:44-51. [doi: 10.1016/j.jpsychires.2015.11.015] [Medline: 26874269]
21. Machulska A, Rinck M, Klucken T, Kleinke K, Wunder JC, Remeniuk O, et al. "Push it!" or "Hold it!"? A comparison of nicotine-avoidance training and nicotine-inhibition training in smokers motivated to quit. Psychopharmacology (Berl) 2022;239(1):105-121 [FREE Full text] [doi: 10.1007/s00213-021-06058-5] [Medline: 35013762]
22. Wen S, Larsen H, Wiers RW. A pilot study on approach bias modification in smoking cessation: activating personalized alternative activities for smoking in the context of increased craving. Int J Behav Med 2022;29(4):480-493 [FREE Full text] [doi: 10.1007/s12529-021-10033-x] [Medline: 34697780]
23. Wittekind CE, Reibert E, Takano K, Ehring T, Pogarell O, Rüther T. Approach-avoidance modification as an add-on in smoking cessation: a randomized-controlled study. Behav Res Ther 2019;114:35-43. [doi: 10.1016/j.brat.2018.12.004] [Medline: 30716613]

XSL•FO
RenderX

24. Boffo M, Zerhouni O, Gronau QF, van Beek RJJ, Nikolaou K, Marsman M, et al. Cognitive bias modification for behavior change in alcohol and smoking addiction: Bayesian meta-analysis of individual participant data. Neuropsychol Rev 2019;29(1):52-78. [doi: 10.1007/s11065-018-9386-4] [Medline: 30644025]
25. Eberl C, Wiers RW, Pawelczack S, Rinck M, Becker ES, Lindenmeyer J. Implementation of approach bias re-training in alcoholism-how many sessions are needed? Alcohol Clin Exp Res 2014;38(2):587-594. [doi: 10.1111/acer.12281] [Medline: 24164417]
26. Forman EM, Goldstein SP, Flack D, Evans BC, Manasse SM, Dochat C. Promising technological innovations in cognitive training to treat eating-related behavior. Appetite 2018;124:68-77 [FREE Full text] [doi: 10.1016/j.appet.2017.04.011] [Medline: 28414042]
27. Beard C, Weisberg RB, Primack J. Socially anxious primary care patients' attitudes toward cognitive bias modification (CBM): a qualitative study. Behav Cogn Psychother 2012;40(5):618-633 [FREE Full text] [doi: 10.1017/S1352465811000671] [Medline: 22127022]
28. Schoenmakers TM, de Bruin M, Lux IFM, Goertz AG, Van Kerkhof DHAT, Wiers RW. Clinical effectiveness of attentional bias modification training in abstinent alcoholic patients. Drug Alcohol Depend 2010;109(1-3):30-36. [doi: 10.1016/j.drugalcdep.2009.11.022] [Medline: 20064698]
29. Kuckertz JM, Schofield CA, Clerkin EM, Primack J, Boettcher H, Weisberg RB, et al. Attentional bias modification for social anxiety disorder: what do patients think and why does it matter? Behav Cogn Psychother 2019 Jan;47(1):16-38 [FREE Full text] [doi: 10.1017/S1352465818000231] [Medline: 29729676]
30. Schenkel EJ, Rinck M, Wiers RW, Becker ES, Muhlig S, Schoeneck R, et al. Implementing approach-bias modification as add-on to varieties of clinical treatment for alcohol use disorders: results of a multicenter RCT. Eur Addict Res 2024;30(2):94-102. [doi: 10.1159/000537811] [Medline: 38503273]
31. Snippe L, Boffo M, Galvin H, Willemen R, Pronk T, Dom G, et al. How we lost 90% of participants on a bad bet: results from a pilot randomized controlled trial on cognitive bias modification in problem gamblers. J Gambl Stud 2024;40(2):521-554. [doi: 10.1007/s10899-023-10263-6] [Medline: 38006537]
32. Urech A, Krieger T, Chesham A, Mast FW, Berger T. Virtual reality-based attention bias modification training for social anxiety: a feasibility and proof of concept study. Front Psychiatry 2015;6:154. [doi: 10.3389/fpsyt.2015.00154] [Medline: 26578986]
33. Otkhmezuri B, Boffo M, Siriaraya P, Matsangidou M, Wiers RW, Mackintosh B, et al. Believing is seeing: a proof-of-concept semiexperimental study on using mobile virtual reality to boost the effects of interpretation bias modification for anxiety. JMIR Ment Health 2019;6(2):e11517 [FREE Full text] [doi: 10.2196/11517] [Medline: 30789353]
34. Prior K, Salemink E, Wiers RW, Teachman BA, Piggott M, Newton NC, et al. Acceptability and co-development of an online cognitive bias modification intervention for emerging adults with hazardous alcohol use and social anxiety: a mixed methods study. Alcohol Clin Exp Res 2020;44(11):2283-2297. [doi: 10.1111/acer.14452] [Medline: 33146919]
35. Laurens MC, Pieterse ME, Brusse-Keizer M, Salemink E, Ben Allouch S, Bohlmeijer ET, et al. Alcohol avoidance training as a mobile app for problem drinkers: longitudinal feasibility study. JMIR Mhealth Uhealth 2020;8(4):e16217. [doi: 10.2196/16217] [Medline: 32286235]
36. Machulska A, Eiler TJ, Kleinke K, Grünewald A, Brück R, Jahn K, et al. Approach bias retraining through virtual reality in smokers willing to quit smoking: a randomized-controlled study. Behav Res Ther 2021;141:103858. [doi: 10.1016/j.brat.2021.103858] [Medline: 33862407]
37. Machulska A, Eiler TJ, Haßler B, Kleinke K, Brück R, Jahn K, et al. Mobile phone-based approach bias retraining for smokers seeking abstinence: a randomized-controlled study. Int J Ment Health Addiction 2023;22(6):4126-4147 [FREE Full text] [doi: 10.1007/s11469-023-01107-w]
38. Malkomsen A, Røssberg JI, Dammen T, Wilberg T, Løvgren A, Ulberg R, et al. "It takes time to see the whole picture": patients' views on improvement in cognitive behavioral therapy and psychodynamic therapy after three years. Front Psychiatry 2024;15:1342950. [doi: 10.3389/fpsyt.2024.1342950] [Medline: 38559399]
39. French LRM, Thomas L, Campbell J, Kuyken W, Lewis G, Williams C, et al. Individuals' long term use of cognitive behavioural skills to manage their depression: a qualitative study. Behav Cogn Psychother 2017;45(1):46-57 [FREE Full text] [doi: 10.1017/S1352465816000382] [Medline: 27629570]
40. Rinck M, Becker ES. Approach and avoidance in fear of spiders. J Behav Ther Exp Psychiatry 2007;38(2):105-120. [doi: 10.1016/j.jbtep.2006.10.001] [Medline: 17126289]
41. Machulska A, Kleinke K, Eiler TJ, Grünewald A, Brück R, Jahn K, et al. Retraining automatic action tendencies for smoking using mobile phone-based approach-avoidance bias training: A study protocol for a randomized controlled study. Trials 2019;20(1):720. [doi: 10.1186/s13063-019-3835-0] [Medline: 31831080]
42. Machulska A, Eiler TJ, Grünewald A, Brück R, Jahn K, Niehaves B, et al. Promoting smoking abstinence in smokers willing to quit smoking through virtual reality-approach bias retraining: a study protocol for a randomized controlled trial. Trials 2020;21(1):227. [doi: 10.1186/s13063-020-4098-5] [Medline: 32102685]
43. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991;86(9):1119-1127. [doi: 10.1111/j.1360-0443.1991.tb01879.x] [Medline: 1932883]

XSL•FO
RenderX

44. Bleich S, Havemann-Reinecke U, Kornhuber J. Fagerström-Test für Nikotinabhängigkeit. Göttingen: Beltz Test; 2002.
45. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2008;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
46. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc and Health 2019 Jun 13;11(4):589-597. [doi: 10.1080/2159676x.2019.1628806]
47. Machulska A, Kleinke K, Neubert M, Sünkel E, Bück J, Klucken T. Long-term outcomes of approach-bias-modification for smoking cessation: a six-month and four-year follow-up study. Behav Res Ther 2026 May 26;203:105090 [FREE Full text] [doi: 10.1016/j.brat.2026.105090] [Medline: 42235414]
48. Tudor-Sfetea C, Rabee R, Najim M, Amin N, Chadha M, Jain M, et al. Evaluation of two mobile health apps in the context of smoking cessation: qualitative study of cognitive behavioral therapy (CBT) versus non-CBT-based digital solutions. JMIR Mhealth Uhealth 2018;6(4):e98 [FREE Full text] [doi: 10.2196/mhealth.9405] [Medline: 29669708]
49. Constantino MJ. Believing is seeing: an evolving research program on patients' psychotherapy expectations. Psychother Res 2012;22(2):127-138. [doi: 10.1080/10503307.2012.663512] [Medline: 22469008]
50. Greenberg RP, Constantino MJ, Bruce N. Are patient expectations still relevant for psychotherapy process and outcome? Clin Psychol Rev 2006;26(6):657-678. [doi: 10.1016/j.cpr.2005.03.002] [Medline: 15908088]
51. Wiers RW, Bartholow BD, van den Wildenberg E, Thush C, Engels RCME, Sher KJ, et al. Automatic and controlled processes and the development of addictive behaviors in adolescents: a review and a model. Pharmacol Biochem Behav 2007;86(2):263-283. [doi: 10.1016/j.pbb.2006.09.021] [Medline: 17116324]
52. Moradveisi L, Huibers MJH, Arntz A. The influence of patients' attributions of the immediate effects of treatment of depression on long-term effectiveness of behavioural activation and antidepressant medication. Behav Res Ther 2015;69:83-92. [doi: 10.1016/j.brat.2015.04.007] [Medline: 25898343]
53. Schaumberg K, Kuerbis A, Morgenstern J, Muench F. Attributions of change and self-efficacy in a randomized controlled trial of medication and psychotherapy for problem drinking. Behav Ther 2013 Mar;44(1):88-99 [FREE Full text] [doi: 10.1016/j.beth.2012.07.001] [Medline: 23312429]
54. Blackwell SE, Woud ML, MacLeod C. A question of control? Examining the role of control conditions in experimental psychopathology using the example of cognitive bias modification research. Span J Psychol 2017;20:E54. [doi: 10.1017/sjp.2017.41] [Medline: 29072159]
55. Rief W, Glombiewski JA. Erwartungsfokussierte psychotherapeutische interventionen (EFPI). Verhaltenstherapie 2016;26(1):47-54 [FREE Full text] [doi: 10.1159/000442374]
56. Wiers RW, Zerhouni O, den Uyl T, Boffo M. Cognitive bias modification in the treatment of addiction. In: Cognition and Addiction: A Researcher's Guide From Mechanisms Towards Interventions. Cambridge, MA: Academic Press; 2020:231-241.
57. Luo X, Rinck M, Bekkering H, Becker ES. The processing of task-irrelevant emotion and colour in the Approach-Avoidance Task. Cogn Emot 2019 May;33(3):548-562 [FREE Full text] [doi: 10.1080/02699931.2018.1497951] [Medline: 29985106]
58. Gladwin TE, Figner B, Crone EA, Wiers RW. Addiction, adolescence, and the integration of control and motivation. Dev Cogn Neurosci 2011;1(4):364-376 [FREE Full text] [doi: 10.1016/j.dcn.2011.06.008] [Medline: 22436562]
59. Qualitative analysis of proximal and long-term participants' perspectives on approach bias modification for smoking cessation. Open Science Framework. URL: https://osf.io/k275x/overview?view_only=ffec9e7d496f46838a9d9b067905297a [accessed 2026-06-02]

Abbreviations

ABM: attention bias modification
ApBM: approach bias modification
AUD: alcohol use disorder
CBM: cognitive bias modification
CBM-I: cognitive bias modification for interpretation
CG: control group
EG: experimental group
FTND: Fagerström Test for Nicotine Dependence
RCT: randomized controlled trial
TAU: treatment as usual
VR: virtual reality

XSL•FO
RenderX

Edited by A Parush; submitted 10.Feb.2026; peer-reviewed by A Eisingerich, T Pan; comments to author 26.Mar.2026; revised version received 09.May.2026; accepted 11.May.2026; published 22.Jun.2026.

Please cite as:
Machulska A, Bück J, Sünkel E, Neubert M, Klucken T
Proximal and Long-Term Participants' Perspectives on Approach Bias Modification for Smoking Cessation: Qualitative Study
JMIR Hum Factors 2026;13:e93238
URL: https://humanfactors.jmir.org/2026/1/e93238
doi:10.2196/93238
PMID:

©Alla Machulska, Johanna Bück, Esra Sünkel, Marie Neubert, Tim Klucken. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 22.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluating a Digital Mental Health Tool for Implementation Into New Zealand's Integrated Primary Mental Health and Addictions Service: Usability Study

Vincent Allen[1], MAppPsy; Danielle Lottridge[2*], PhD; Karolina Stasiak[1*], PhD

[1]Department of Psychological Medicine, Faculty of Medical and Health Sciences, University of Auckland, Building 501, 85 Park Road, Grafton, Auckland, New Zealand

[2]School of Computer Science, Faculty of Science, University of Auckland, Auckland, New Zealand

[*]these authors contributed equally

Corresponding Author:
Vincent Allen, MAppPsy
Department of Psychological Medicine, Faculty of Medical and Health Sciences, University of Auckland, Building 501, 85 Park Road, Grafton, Auckland, New Zealand

Abstract

Background: Rising global demand for mental health services strains traditional care models, a trend evident in New Zealand, where the Integrated Primary Mental Health and Addictions (IPMHA) model was introduced to improve access to care. While the IPMHA model shows promise, significant service-delivery challenges undermine its scalability. Digital mental health tools (DMHTs) present an opportunity for digital optimization, yet their effectiveness is often limited by low practitioner adoption, a persistent implementation barrier. To ensure these tools are impactful, a user- and context-centered DMHT design approach may help mitigate practitioner adoption barriers and create solutions that can be seamlessly integrated into clinical workflows.

Objective: Reporting on the Test stage of the 5-stage Design Thinking framework, this study aimed to evaluate the usability and acceptability of a DMHT software prototype intended to support health improvement practitioners working within New Zealand's IPMHA model.

Methods: Five health improvement practitioners from a single primary health organization participated in semistructured usability interviews. Data were collected using a think-aloud protocol during mock clinical sessions and analyzed using affinity diagramming to identify key software feature requirements necessary to promote usability, adoption, and workflow integration.

Results: Feedback was obtained for all software minimum viable product features. While practitioners found the clinical support features valuable, 2 system-level requirements were identified as prerequisites for adoption. The first was administrative optimization: the DMHT must reduce workload by automating tasks, such as clinical note entry, data reporting, and psychometric scoring. The second was seamless integration with existing clinic patient management software to eliminate double-handling of data and resolve IT-related workflow frustrations.

Conclusions: For a DMHT to be successfully adopted by IPMHA practitioners, it must primarily function to solve existing administrative and workflow inefficiencies. Clinical support features, such as the provision of therapeutic tools and exercises, though helpful, are secondary to the tool's ability to be a practical, efficient, and fully integrated component of daily practice. These findings underscore the value of user- and context-centered design in uncovering the pragmatic, systems-level needs of end users in a complex primary care service-delivery environment.

(JMIR Hum Factors 2026;13:e84412) doi:10.2196/84412

KEYWORDS

digital mental health; usability testing; technology adoption; primary health care; integrated care scalability; integrated primary mental health and addictions model; health care workflow optimization; human-centered design; more-than-human design; context-sensitive design; user-centered design

Introduction

Globally, the prevalence of mental illness is rising, with approximately 1 billion people living with some form of mental illness in 2019, most commonly depression or anxiety [1]. The COVID-19 pandemic had a significant and well-documented impact on global mental health, with initial estimates from the World Health Organization showing a 26% increase in anxiety disorders and a 28% increase in major depressive disorders in 2020 alone [2], placing additional strain on already overstretched mental health services. Despite growing demand, access to adequate care remains limited due to chronic underinvestment

XSL•FO
RenderX

in mental health services and ongoing workforce shortages [1]. For decades, international health bodies have highlighted the growing disparity between the demand for mental health support and the available resources, a crisis that has only worsened over time [3]. This treatment gap, reaching as high as 90% in some nations, underscores a persistent global failure to prioritize mental health, leaving public health systems overwhelmed and unable to provide necessary care for a substantial portion of the population [1].

This international trend is evident in New Zealand, where there has been a sharp increase in psychological distress across the population in recent years [4,5]. Demand for mental health services continues to grow, placing significant strain on an already underfunded and resource-limited public health system [6]. This issue is particularly acute among young people, women, Māori, and Pacific peoples, who are disproportionately affected and experience higher rates of psychological distress. Furthermore, socioeconomic status is a significant factor, with individuals in more deprived areas facing a greater likelihood of mental health challenges, compounded by unequal access to care [5].

To address this significant gap in service access, the New Zealand government introduced the Integrated Primary Mental Health and Addictions (IPMHA) model in 2017. This integrated care initiative, based on the primary care behavioral health approach [7], aims to provide accessible, publicly funded care for those with mild-to-moderate mental health and behavioral issues within primary care settings. Following a successful pilot, the IPMHA model was implemented nationwide and, as of mid-2024, was available to over 70% of the enrolled primary care population [8]. The service is delivered by clinicians known as health improvement practitioners (HIPs) who offer brief, 30-minute sessions using focused acceptance and commitment therapy (fACT) [9]. While designed to be delivered within a single session, the model allows for follow-up appointments as needed, with primary care physicians maintaining overall responsibility for patient care [9]. A step-by-step breakdown of a typical fACT session is available in Multimedia Appendix 1. Evaluations have shown that the IPMHA model has improved access to mental health support, particularly for underserved groups, by reducing wait times and minimizing financial barriers [9].

Despite these successes, the IPMHA model faces significant implementation challenges. Workforce shortages and inconsistent application across different regions can lead to unpredictable care quality and availability, and practitioners often struggle to maintain fidelity to the brief, single-session intervention model [9]. While the model has improved mental health service access in New Zealand, the operational efficiency and potential for scalability of the IPMHA model are undermined by these service-delivery challenges.

Digital mental health tools (DMHTs), from self-help apps to technologies that support practitioner-led care, are increasingly seen as a solution to the global mental health treatment gap [10,11]. Although commonly associated with self-guided therapy, the primary promise of these tools is their potential for integration into existing health care services [12-14]. Specifically, embedding a custom-designed digital tool within a service that faces delivery challenges, such as the IPMHA model, could enhance the service's effectiveness and scalability [9,10,15,16].

While digital tools can be effective alone, their impact is greatest when combined with practitioner support [10,11,14,17]. This suggests that digital mental health solutions should be viewed not only as a way to replace in-person services but also to augment them, thereby improving the efficiency and reach of existing, evidence-based, human-delivered care.

Despite the potential of digital mental health tools, real-world implementation is hindered by low adoption among practitioners [16,18-20]. Several recent systematic reviews show that clinicians are less likely to adopt tools that are difficult to use, increase workload, disrupt workflows, or are poorly integrated with existing organizational systems [18,20-23]. Rather than being conceptualized as stand-alone products, digital tools for integration into existing services should be designed as integral components of the service they are intended to support. Adopting a context-specific, user-centered software development methodology is necessary to ensure that these tools provide practical value while also integrating seamlessly into clinical workflows and organizational systems [10,13,15,16,24].

This study reports on the final phase of a larger research project that used the 5-stage Design Thinking framework [25,26] to develop a DMHT to enhance the IPMHA model. The Design Thinking methodology was selected for this project due to its iterative user-centered approach and proven efficacy in creating effective digital solutions for complex health care service-delivery problems [27,28].

The aim of this study was to evaluate the usability and acceptability of the software minimum viable product (MVP) developed through this process, with a focus on integration into existing primary care systems and workflows. Specifically, the study sought to evaluate the usefulness and usability of the software's features and to identify whether these features support practitioner adoption and effective integration into the IPMHA service-delivery context.

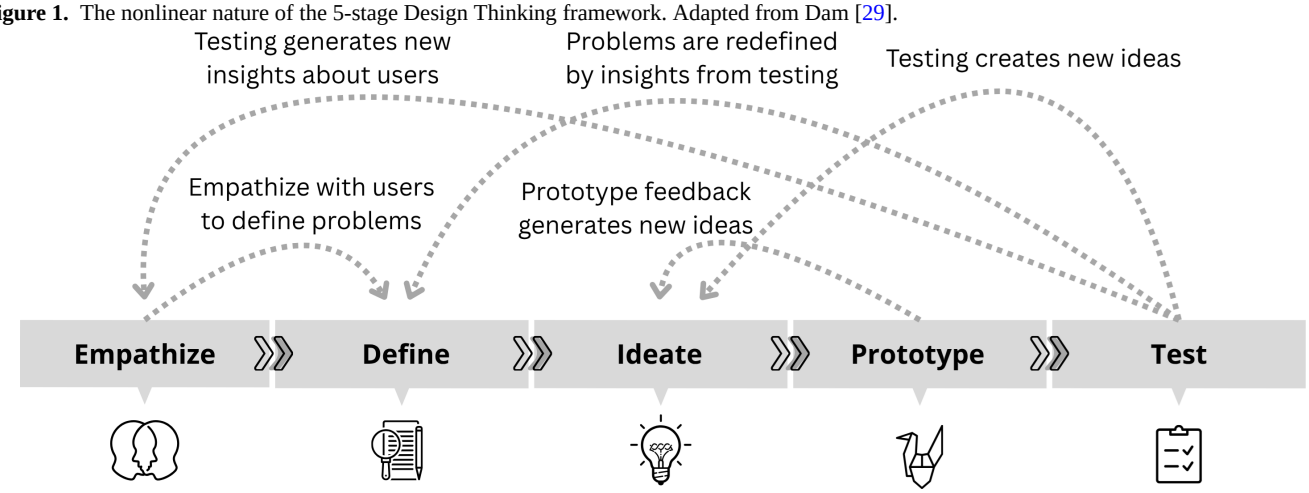
Methods

Research Design and Framework

This study reports on the Test stage of the 5-stage Design Thinking framework (Figure 1). The preceding 4 stages, described below, informed the development of the software MVP evaluated in this study.

XSL•FO
RenderX

Figure 1. The nonlinear nature of the 5-stage Design Thinking framework. Adapted from Dam [29].



1. Empathize: The project began with an exploration of HIPs' experiences working within the IPMHA model through interviews and a workforce survey. This process uncovered a variety of service-delivery challenges, many of which could be addressed by a purpose-built DMHT [30].
2. Define: Service delivery challenges identified during the Empathize stage were translated into user stories and problem statements, ensuring that the project remained focused on validated, contextually relevant, real-world problems.
3. Ideate: The research team generated a broad range of potential software solutions and features based on the problem statements synthesized during the Define stage.
4. Prototype: Software interface prototypes were developed from promising Ideate stage design concepts. Iterative feedback from HIPs and other IPMHA stakeholders shaped the evolution of low-fidelity pen-and-paper sketches into functional wireframe prototypes. The functional wireframes were created using the Figma software package [31]. The culmination of the Prototype stage was the development of a functional software MVP, built using the Flutter programming language [32] for use on an Android tablet.

The purpose of this software MVP was to address the various administrative and workflow challenges reported by HIPs and enhance the delivery of the brief fACT intervention package. To provide a general overview of the software MVP's functionality, Multimedia Appendices 2-4 present screenshots and interface flows for several key features. While this is not an exhaustive list of the software's functions, these examples highlight several core capabilities. Multimedia Appendix 2 depicts the user flow for starting a session, which includes selecting a client from the integrated electronic health record, reviewing prior session information, starting a new session, taking notes, and using a model fidelity support tool. Multimedia Appendix 3 details the creation of a wellness plan, and Multimedia Appendix 4 demonstrates the in-session experiential exercise feature.

Participants and Recruitment

Participants were HIPs working within the IPMHA model in New Zealand primary care clinics. All participants were employed by the same primary health organization (PHO) to ensure consistency in organizational systems and processes. HIPs were recruited through an electronic flyer distributed internally within the participating PHO. Detailed information outlining the study's purpose and procedures was provided to all prospective participants.

A total of 5 HIPs completed the usability interviews. Small-group user testing is considered appropriate for early-stage software evaluation with relatively homogeneous user groups, where the primary goal is to collect qualitative insights rather than achieve statistical validation. Feedback from 5 participants can uncover approximately 85% of major usability issues, as common problems typically become evident within the first few sessions, with additional participants offering diminishing returns [33-35]. Given the homogeneity of the target user group and the focused, deductive approach to data collection and analysis, small-group testing with 5 participants was deemed appropriate for this study.

Participants were eligible for inclusion in the study if they were aged 18 years or older, were able to provide informed consent, and were currently employed as an HIP within the participating PHO.

Data Collection

Data were collected through semistructured, one-on-one interviews designed to elicit in-depth feedback on the usability and perceived usefulness of the software MVP. All interviews were conducted in person by the primary investigator. Participants were informed that the interviews would take between 1 and 2 hours.

During the interview sessions, participants' interactions with the software MVP, running on a tablet device, were captured using the device's integrated screen recording functionality. This was synchronized with an audio recording of the interview to contextualize participants' comments about usability and usefulness in relation to their actions within the software. The average interview duration was 76 (SD 23.49) minutes, with sessions ranging from 49 to 117 minutes, meeting the recommended duration for a broad usability analysis [35].

Interview Procedure

The interview process was structured into 4 phases.

XSL•FO
RenderX

1. Preamble and software orientation: The interviewer introduced the participant to the overall software interface, guiding them through its key features. This phase aimed to familiarize participants with the software and to begin identifying any immediate usability issues. Participants were encouraged to verbalize their thoughts and observations while interacting with the software, following the think-aloud [36] data collection method.
2. Mock fACT session: Participants engaged in a role-play exercise in which they assumed the role of a clinician delivering a typical fACT session using the software. The interviewer acted as the client. Both parties were encouraged to remain "in character" as much as possible to simulate the intended clinical use of the software. Throughout the session, participants continued to verbalize their thoughts and critically reflect on the real-world usability and relevance of different software features.
3. Final think-aloud exploration: Following the mock session, participants were invited to freely navigate the software, revisit specific features, and reflect on their experiences. They were encouraged to comment on the ease of use, functionality, and integration of the software within their existing workflow and organizational processes.
4. Blue-sky brainstorming: The interview concluded with an open-ended brainstorming session, during which participants were invited to share any additional thoughts or suggestions about the software. They were encouraged to propose ideas for new features, suggest modifications to existing features, explore how various features could be integrated within their workflows, and reflect on which features would be essential for addressing service delivery challenges and leading to sustainable engagement with the software.

Figure 2 shows the setting for the usability interviews. The participant (substituted here with an actor) is engaging with the software via a tablet and keyboard. The interviewer, who also acted as the "client" during the mock fACT session, sat across from the participant to simulate the clinical context of how the software would likely be used.

During the mock fACT session portion of the interview, participants were instructed to conduct a typical session as they would with a regular client, using the software MVP where appropriate to supplement their standard workflow. This involved the participant using the software to load a client (played by the interviewer) from a simulated clinic database, review client data, take clinical notes, use experiential exercises, create a wellness plan, and commit all notes and session data to a simulated integrated electronic health record upon conclusion.

Figure 2. Recreation of the interview setting, showing how the participant engaged with the software minimum viable product (MVP) in a simulated clinical context.



Data Analysis

The data analysis methodology chosen for this study was affinity diagramming [37], a qualitative technique commonly used in software development and user experience design [38]. This approach allowed for the systematic exploration of participant feedback concerning the usability and acceptability of the software MVP, particularly in relation to the usefulness, importance, and workflow integration potential of the various software features. The analysis began with a thorough reading of the interview transcripts to gain a high-level overview of the dataset. Each transcript was reviewed alongside the

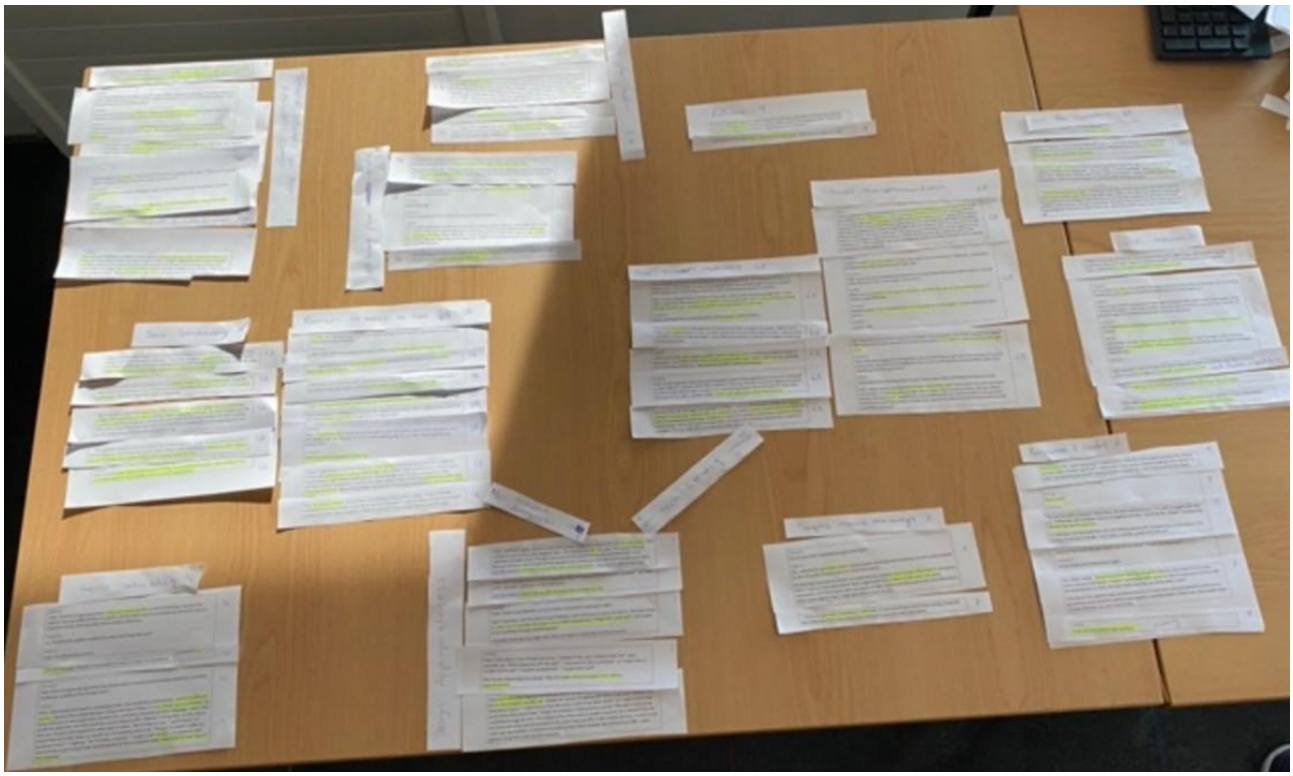
XSL•FO
RenderX

synchronized screen recordings showing participant engagement with the software. Viewing the screen capture alongside participant feedback facilitated a deep understanding of the context surrounding participants' comments, linking their verbal feedback directly to their interactions with the software.

The data collection process was structured to elicit feedback on specific features of the software, and the software feature set was used as the initial deductive framework for organizing the data. This iterative process involved sorting the interview data into different feature categories and coding relevant interview excerpts (codes) within each category. This included the creation of new categories for any emergent feature-related themes that were identified through the initial reading of the transcripts.

Once the data were categorized and coded, each feature category within the framework was independently analyzed using affinity diagramming to visually organize and group related codes (Figure 3). Related interview excerpts, now represented by these codes, were clustered together into "affinities" These clusters represented overarching themes related to that specific feature. After clustering the codes, each cluster was examined to identify the underlying theme or key message, providing a higher-level summary of user feedback for each feature.

Figure 3. Collaborative affinity diagramming using interview transcript excerpts.



Following the initial feature-driven analysis, an additional analysis of the entire dataset was conducted using the same affinity diagramming methodology. The objective was to identify which features or overall functionalities of the software HIPs identified as being the most important for facilitating effective implementation into workflows and supporting sustained long-term engagement with the software.

The construction of the feature-driven framework, refinement of that framework, transcript coding, and subsequent theme identification were conducted collaboratively by the principal researcher (VA) and a trained research assistant. The Nvivo-14 software package assisted with data organization and tracking, whereas printed copies of the transcripts were used for performing the collaborative affinity diagramming activities.

Collectively, these analyses sought to answer key questions regarding the usefulness of specific features and functionalities, the perceived importance of those features and functionalities, and the potential for integration into the HIP workflows and broader organizational systems. The findings are detailed in the *Results* section, providing a comprehensive understanding of the software MVP's usability and acceptability from the perspective of its intended users.

Ethical Considerations

This study was approved by the Auckland Health Research Ethics Committee (reference AH21908). Informed consent was obtained from all participants. Participants received a voucher valued at $50 NZD (US $29.50) for each hour of interview time. During the transcription process, any information relating to participant identity was removed, with a unique participant identifier code assigned to each interview script to ensure confidentiality.

Results

Critical Features and Functionalities to Support Practitioner Adoption and Workflow Integration

A primary goal of this research was to identify which software features HIPs deemed most important for successfully integrating the tool into their workflows. This subject was a central focus during the usability interviews. The functionalities

XSL•FO
RenderX

that HIPs specified as necessary to support the adoption and effective implementation of the software MVP are detailed in Textbox 1 and elaborated upon in the description that follows.

Participants indicated that their willingness to adopt the software is contingent on its ability to solve existing service-delivery problems and demonstrably reduce their workload. If the tool fails to do this, it will likely be treated as an optional accessory rather than an integral part of their workflow.

The most critical area identified for workflow improvement was the reduction in administrative burden. Key areas included reducing double-handling of data, automating psychometric collection and scoring, and streamlining the process for sharing resources with clients. HIPs reported that these administrative frustrations were compounded by broader IT issues. They described frustrations in dealing with workflow friction caused by software incompatibilities, the need to switch between different devices, and managing multiple logins for various systems. A key benefit of the new software, they emphasized, is its ability to eliminate these technical hurdles.

Textbox 1. Features and functionalities necessary to support implementation of the software.

1. Administrative optimization and time-saving
2. Seamless integration with existing practice management software
3. Solving IT and workflow frustrations
4. Must solve problems and reduce workload, not add tasks to an already busy workflow

To solve these interconnected issues, participants consistently stated that seamless integration with existing clinic patient management software was essential. They expressed a strong need for the software to automatically synchronize data, which would minimize manual data entry and ensure a smooth, unified workflow. This was seen as particularly important for practitioners who work across different clinics and PHOs. Ultimately, HIPs indicated that without a high level of systems integration to solve their core administrative and IT problems, they would be unlikely to use the software.

Usability and Usefulness of Specific Software Features

Participant feedback on specific features is presented in Table 1, organized by software feature or functionality. Feedback included feature validation, constructive critiques, and suggestions for modification. The summary in Table 1 is elaborated upon in the subsequent descriptions, and a detailed point-by-point breakdown is available in Multimedia Appendix 5.

XSL•FO
RenderX

Table . Practitioner feedback on specific features and overarching functionalities of the software minimum viable product (MVP).

Software feature	Feature validation	Critique and change requests
Experiential exercises and recordings.	• Useful for early career and fatigued HIPs[a]. • They are well suited for brief sessions. • Easy to integrate into existing workflow.	• Additional customizability for different client needs. • Adding visual elements may improve engagement with recordings.
Love Work Play Health tool.	• Enhances model fidelity. • Provides a useful reminder for early career HIPs.	• The design should encourage flexible use.
Model fidelity checklist.	• Enhances model fidelity. • Useful for both novice and experienced HIPs.	• Would benefit from key information being more accessible. • Would benefit from better integration into workflow.
Session history review.	• Improved efficiency compared to clinic electronic health record. • Provides easy access to patient-specific information and reminders.	—[b]
Automation of external data reporting.	• Reduces double-handling of data. • Automating psychometric scoring and reporting. • Structuring notes to align with reporting needs.	—
Psychometric recording and analysis.	• Expanding psychometric options for clinical use. • Enhancing psychometric administration in session. • Improving workflow and integration with note-taking.	—
Psychometric review.	• Visualizing change over time. • Supporting client conversations.	—
Session timer.	• Time management and session flow.	—
Values elicitation and recording.	—	• Efficiency and collaboration. • Language and accessibility issues. • Integration with contextual interview and note-taking.
Wellness plan creation.	—	• Wellness plan creation should be flexible. • Terminology concerns. • Usability and interface design. • Inclusion of a support person.
Clinical note and session data recording.	• Reducing administrative burden.	• Integrating note-taking with model fidelity tools and overall session flow. • Improving access to notes and supporting personal reflections. • Automating data recording and reporting.
Integration of the software with existing clinic IT systems.	• Integration with clinic patient management software is essential. • Reducing administrative burden and double-handling of data. • IT simplification across clinics and PHOs[c]. • Supporting model fidelity and reliable data collection.	• Securely and efficiently sharing resources with clients. • Ensuring data security and clinic buy-in.

XSL•FO
RenderX

Software feature	Feature validation	Critique and change requests
Postsession support for clients.	• Providing clients with session takeaways. • Collecting postsession outcome data.	• Balance between ongoing support and managing HIP workload. • Personalizing and adding value to session summaries. • Overcoming administrative challenges. • Format of takeaway resources must be flexible around client needs.

[a]HIP: health improvement practitioner.

[b]Not applicable.

[c]PHO: primary health organization.

Experiential Exercises and Recordings

Participants felt that having a library of prerecorded experiential exercises was a valuable asset, particularly for early-career or fatigued HIPs who might need support in delivering them consistently. The shorter exercises, such as mindful breathing, were seen as especially well suited for the brief intervention model. HIPs appreciated that having these resources integrated directly into the software was more efficient than searching for content online. Furthermore, the ability to use editable scripts allowed for customization in pacing and duration to better meet individual client needs, such as during moments of panic. To further improve client engagement, participants suggested adding visual elements or animations to the recordings and valued the ability to send recommended exercises to clients as part of a session summary.

Love Work Play Health Tool

The Love Work Play Health tool was seen as a valuable feature for enhancing fidelity to the fACT model. Participants reported that it provides a structured guide to ensure a thorough contextual interview is conducted, prompting HIPs to explore key life domains and client values. This was considered particularly useful as a reminder for newer practitioners. However, HIPs stressed the importance of flexibility, noting that it would be impractical to ask every prompt in a single session. They recommended including guidance that clarifies the tool is a flexible aid rather than a rigid script to be followed verbatim.

Model Fidelity Checklist

Participants indicated that a digital Model Fidelity Checklist was a practical tool for ensuring correct fACT delivery, serving as a step-by-step guide for novices and a helpful refresher for more experienced HIPs. Many practitioners already use printed reference materials, so integrating a digital checklist directly into the session workflow, for instance, within a note-taking template, was seen as a significant improvement. To enhance usability, participants suggested making key information more accessible through on-screen prompts or expandable info buttons. They also proposed that embedding the checklist directly into the session notes would encourage its active use and could even help streamline external data reporting.

Session History Review

The ability to easily review a client's session history was highlighted as a significant efficiency gain. Participants described the current process of reviewing past notes in clinic patient management software as slow and cumbersome. The proposed feature, which collates previous session notes, allowed HIPs to quickly prepare for appointments, access important client-specific information, and recall personal details that aid in building rapport. Moreover, HIPs often leave themselves reminders in session notes, and easy access to this history would ensure important follow-up tasks are not overlooked.

Automation of External Data Reporting

Automating external data reporting was considered an essential time-saving feature by participants. They expressed significant frustration with the "double-handling" of data, where they must reenter the same information for both clinical notes and separate reporting systems. The software's ability to capture information once and automatically transmit it was seen as a key function that would significantly reduce administrative workload. Likewise, automating the scoring and reporting of psychometric measures would not only save time but also improve model fidelity. Participants suggested that structuring note-taking templates to align with reporting requirements would further streamline this process.

Psychometric Recording and Analysis

Participants appreciated the software's flexibility in offering a range of psychometric tools beyond the 2 mandated ones, noting that practitioners often use other measures, such as the 10-item Kessler Psychological Distress Scale Kessler-10 or 9-item Patient Health Questionnaire, which are more familiar to referring physicians. The ability to administer these assessments collaboratively, either by reading them aloud or having clients complete them, was valued, and the automatic scoring feature was seen as a major time-saver that allows for the immediate discussion of results in-session. HIPs also stressed the need for a seamless workflow, allowing them to pause an assessment to explore a client's response and skip questions if necessary.

Psychometric Review

The ability to visually track a client's psychometric scores over time was viewed as a highly valuable feature. Participants explained that seeing these changes at a glance helps them monitor progress and identify areas needing further attention. They also noted that presenting these data visually on a single

XSL•FO
RenderX

screen would significantly improve their workflow when reviewing results with clients. This direct comparison of past and present scores facilitates meaningful conversations about progress and helps collaboratively identify new goals.

Session Timer

A session timer was seen as a simple but effective tool for managing session flow and adhering to tight schedules. Participants noted that having a visible timer benefits both the practitioner and the client. For the HIP, it ensures all components of the fACT model can be covered without feeling rushed. For the client, being able to view the remaining session time during collaborative tasks could help maintain focus.

Values Elicitation and Recording

Participants highlighted the need for a more efficient and accessible way to elicit client values, as traditional methods, such as card sorts, can be time-consuming. They suggested that the term "values" can be confusing for clients and recommended using simpler, more relatable language, such as asking, "What is important to you?" While values often emerge naturally during the contextual interview, HIPs expressed a desire for a designated, collaborative space within the software to seamlessly record them as they are identified.

Wellness Plan Creation

Flexibility was the key theme for feedback on the wellness plan feature. Participants emphasized that creating a plan should accommodate diverse client needs, with optional fields for elements, such as start dates or reminders. They also raised concerns about terminology, suggesting that words such as "likeliness" or "confidence" are more suitable than "willingness," and "support person" is preferable to the informal term "buddy." From a usability perspective, they recommended a more user-friendly interface for setting dates and frequencies. The ability to include a support person was seen as beneficial for client accountability, but it was stressed that clear information about data privacy is essential. Finally, providing a printable summary is crucial for clients who may not engage with digital resources.

Clinical Note and Session Data Recording

The primary value of this feature lies in its potential to significantly reduce administrative burden. Participants emphasized that eliminating the "double-handling" of data by entering notes once and having them automatically sync with the clinic's patient management software was a critical function. They also desired a workflow where fidelity tools are integrated directly into the note-taking process, using structured templates that mirror the natural progression of a session. Quick access to past notes, a separate space for personal reflections, and the ability to attach nontext materials were also identified as valuable enhancements. To ensure proper data management, participants suggested reminders for obtaining client consent and automatic submission of draft notes after a set period to prevent lost records.

Integration of the Software With Existing Clinic IT Systems

Across all interviews, seamless integration with existing clinic IT systems, particularly with patient management software, was identified as an essential, nonnegotiable requirement for adoption. Participants stated that the software must solve current administrative problems, not create new ones. By automating data entry, synchronizing notes and psychometrics, and reducing the "IT friction" of multiple logins and incompatible systems, the software would significantly reduce workload and improve job satisfaction. This integration is key to ensuring consistent data collection and allowing HIPs to securely and efficiently share resources with clients.

Postsession Support for Clients

Participants valued the software's ability to extend support beyond the session. They highlighted the importance of providing clients with tangible takeaways, such as wellness plans and experiential exercises, in a format that suits their needs, whether digitally via email or as a printed hard copy. The ability to personalize these summaries with key insights or links to external resources was seen as a way to enhance client engagement. While collecting postsession outcome data via a client-facing app was considered useful, HIPs cautioned that a balance must be struck to avoid overwhelming clients with excessive notifications. Streamlining the process of sending these resources was seen as a key benefit that could overcome common administrative challenges within clinic settings.

Discussion

Principal Results

The key objectives of this study were to evaluate the usability and acceptability of the software MVP and identify which features and functionalities were critical for facilitating effective integration into the IPMHA service-delivery context. The analysis revealed 2 nonnegotiable prerequisites for the software's adoption into HIP workflows: significant administrative optimization and seamless integration with existing IT systems. Participants' willingness to use the tool was primarily contingent on its ability to reduce their workload by automating tasks, such as note-taking, data reporting, and psychometric scoring. This need for efficiency was directly linked to the second prerequisite: robust integration with existing patient management software to synchronize data, eliminate duplicate entry, and resolve workflow friction. Without these foundational capabilities, participants indicated the software would likely be viewed as a workload-increasing add-on rather than an integral component of their clinical practice.

Beyond the overarching need for efficiency and integration, several individual features were highlighted for their potential benefits. The inclusion of experiential exercise recordings was seen as a valuable resource, particularly for newer or fatigued HIPs. The Love Work Play Health and model fidelity checklist tools were acknowledged for their potential to support HIPs with in-session training, promote adherence to the fACT model, and reinforce best practices in session delivery. The session history review feature was praised for offering a more efficient

XSL•FO
RenderX

way to access information from previous client interactions. Finally, the psychometric recording and analysis tools were appreciated for expanding the available assessment options and streamlining their administration within sessions.

These findings highlight that a successful digital tool designed to support HIPs working within the IPMHA model must not only offer features to assist with clinical practice but also seamlessly integrate with existing systems and significantly reduce the administrative burden for HIPs. Effectively addressing these administrative optimization challenges will likely be the primary determinant of practitioner engagement, sustained use, and the tool's impact on IPMHA model scalability.

Implications for Design

The feedback from HIPs regarding the importance of administrative optimization and seamless workflow integration led to a redesign of the software's backend as well as several interface changes. These changes were informed by specific usability feedback. For instance, the call for better integration of clinical note-taking with other in-session tools prompted the redesign of the note-taking tool to create a more unified workflow that matched the clinical note format typically used in primary care services. Other notable changes included a dedicated resource upload feature to facilitate sending PDFs and worksheets to clients. HIPs also stressed that they needed to retain the ability to easily print resources for less-tech-capable clients, which prompted the integration of one-click print functionality for the experiential exercise scripts, wellness plan summaries, and session summaries. Most of the model fidelity and fACT support tools were retained in the same format, with minor changes in response to feedback—for example, the addition of a "support person summary screen" in the wellness plan to enhance transparency regarding what information would be shared.

It is important to view these design changes not as a final step but as part of the ongoing iterative Design Thinking process. Our team is now applying feedback from the usability study to refine the software for clinical implementation. Most of the requested features involve backend modifications to ensure seamless integration with existing clinical systems, requiring only minor changes to the user interface. These backend updates will be incorporated into our next MVP, which will then be tested in a clinical implementation trial to validate its effectiveness and acceptability in a real-world primary care setting.

Strengths and Limitations

This study has several key strengths. The methodological approach, combining semistructured interviews with a think-aloud protocol and a mock fACT session, yielded rich, in-depth qualitative data. This captured both high-level reflections on the software's features and direct, in-the-moment usability feedback. A particular strength of this study was its use of a simulated clinical environment. Having the participant actively use the software MVP in a realistic service-delivery context helped uncover usability issues that are often missed by less hands-on methods, such as simple feature walkthrough interviews or focus groups. Furthermore, this study demonstrates the usefulness of rapid, small-sample usability testing in generating rich and actionable insights, highlighting the efficiency of this low-resource approach for iterative development cycles.

By focusing on a homogeneous user group (HIPs from a single PHO), we controlled for organizational variables. This approach provided a clearer understanding of the software's usability in a specific service context and generated actionable data to guide the MVP's iterative development for use with this specific group of prospective users.

However, the study also has notable limitations. The most significant limitation is that the usability analysis was not based on real-world clinical usage of the software. Consequently, the findings reflect practitioners' perceived usability and anticipated usefulness based on a simulated clinical environment, which may not fully correspond to their experiences during actual, day-to-day clinical practice. The small sample size of 5, while appropriate for uncovering major usability issues, limits the generalizability of these findings. Similarly, the recruitment of participants from a single PHO implies that the results may not be generalizable to HIPs working in different organizational or geographical contexts. Finally, there is a potential for social desirability bias, as participants may have been inclined to provide positive feedback to the interviewer, who was also the primary investigator involved in the development of the software.

Comparisons With Prior Research

To situate the findings of this study within the broader academic literature, we compared the software features and functionalities prioritized by HIPs against the key barriers to practitioner digital adoption identified in prior research. A comprehensive synthesis of 6 major systematic reviews [10,18,19,21-23] reveals a variety of common practitioner digital adoption barriers. Across different health care contexts, a core set of three challenges consistently emerges: (1) the fear that digital tools will increase practitioner workload and disrupt workflows; (2) inadequate training and support with digital tools; and (3) concerns regarding data privacy and security.

The most significant theme to emerge from our study, the demand for administrative optimization, strongly validates what the literature identifies as the most significant adoption barrier: the fear that digital tools will increase practitioner workload [10,18,23,24]. HIPs stated that their adoption of the software hinges on its ability to "solve problems and reduce workload, not add tasks to an already busy workflow." Specifically, participants valued features that automate clinical notes and data reporting to eliminate the "double-handling" of data, as well as those that automate psychometric scoring to save time. These findings confirm that for IPMHA practitioners, like many other health care professionals, adoption depends on a digital tool demonstrably reducing, rather than increasing, their existing administrative burden.

A common workload-related adoption barrier identified in the literature is system interoperability [10,16,18,19,22,24]. Our second key finding, the necessity of seamless integration with

XSL•FO
RenderX

existing practice management software, directly addresses the well-documented barrier of poor system interoperability. HIPs were unequivocal that without robust integration, the software would likely remain unused because it would create, rather than solve, administrative problems. This desire for a high level of interoperability was also reflected in their wish for the DMHT to solve common "IT and workflow frustrations," such as managing multiple logins and system incompatibilities. These findings confirm that, similar to other digital tools designed for health care implementation [39,40], a DMHT's value is contingent not on its clinical features alone, but on its ability to integrate seamlessly into existing workflows and organizational systems.

Beyond these primary themes, participants' concerns also echoed other established barriers. The frustrations HIPs expressed with workflow issues are symptomatic of an environment lacking integrated support systems, aligning with the consensus that effective adoption requires continuous, context-specific training and readily available technical support. Furthermore, the core barrier of privacy and security was also present in our findings. HIPs stressed the importance of "ensuring data security and clinic buy-in" and designing the system with low-risk data handling as a primary consideration. This concern extended to client-facing features, with HIPs noting the importance of clarifying what information a "support person" could access. This confirms that even when a tool offers clear workflow benefits, adoption will be hindered if it does not first establish a foundation of trust and security for both practitioners and the clients they serve.

Interestingly, whereas prior research focusing on mental health tools has highlighted a fear that technology could depersonalize the therapeutic relationship [10,18], our findings reveal a more nuanced perspective from the IPMHA workforce. Participants identified several ways the software could enhance rather than detract from relational aspects of care. For instance, the Session History Review feature was valued for enabling HIPs to recall personal client details to build rapport in subsequent sessions. Likewise, psychometric tools were seen not as impersonal data collection exercises but as a way to facilitate collaborative conversations with clients about their progress. This suggests that for this workforce, when a tool is designed to support, rather than replace, clinical judgment and interaction, it can be perceived as an asset to the therapeutic alliance.

In conclusion, the findings from this study strongly validate well-established evidence from the broader literature. The primary concerns voiced by HIPs—the fear of increased administrative burden and the absolute necessity of seamless workflow and patient management software integration—directly mirror the complex, sociotechnical barriers to digital tool adoption consistently reported across various other health care contexts. This underscores a crucial takeaway: the path to successful implementation of DMHTs is paved by solving the pragmatic, operational challenges that practitioners face. For a digital tool to be adopted by the IPMHA workforce, its value must be demonstrated through its ability to function as a trusted, efficient, and well-integrated component of the complex clinical environment it seeks to support.

Conclusions and Implications for Future Research

The central message from this study is that, for a digital tool to be adopted by HIPs, its clinical use is secondary to its ability to solve their most pressing administrative and workflow challenges. While clinical support features were valued, they are not the drivers of adoption. Instead, practitioners require a tool that functions as a practical and efficient component of their daily practice, capable of reducing their administrative load through task automation and seamless integration with existing patient management software. These findings suggest that the successful implementation of any DMHT in this context is contingent on its capacity to address the foundational, operational realities of primary care service delivery.

These findings illustrate the value of collaborative and user-centered research methodologies for DMHT development. By grounding the overarching research project in the lived experience of HIPs through a user- and context-centered design framework, this study was able to uncover the true determinants of digital adoption for this group of practitioners. A traditional, top-down, researcher-led approach to digital tool development, which might prioritize theoretical clinical advancements or novel functionalities, would have been at high risk of missing the pragmatic, foundational needs of the end users. Without this direct engagement with end users, the importance of the DMHT in solving administrative and workflow frustrations could have easily been overlooked, leading to the development of a tool that, while perhaps innovative, would ultimately fail to gain traction in the demanding environment of primary care. This underscores that for digital mental health tools to be effective and engaging, their development must be driven by the actual, expressed needs of the people who will be using them.

The findings of this study provide valuable early-stage validation of the required software features and functionalities that a DMHT must possess to effectively support IPMHA practitioners operating within New Zealand primary care. Now that the proposed feature set has been validated, future research will focus on further refinement of the software MVP into a clinically implementable DMHT. This development will prioritize the core administrative and workflow features that were validated in this study, with a particular focus on robust patient management software integration and the automation of practitioner data management tasks. A clinical implementation trial will be conducted in which IPMHA practitioners use the DMHT in their daily practice. The purpose of this trial will be to investigate the tool's impact on administrative efficiency, assess its usability and feasibility in a live clinical environment, and determine its efficacy in improving client outcomes and enhancing the overall scalability of the IPMHA model.

XSL•FO
RenderX

Funding

This work was supported by the Ember Korowai Takitini Mental Health and Addiction Innovation Grant and the Health Research Council New Zealand Health Delivery Research Activation Grant. No generative artificial intelligence tools were used in the preparation of this manuscript.

Data Availability

The qualitative interview data are not publicly available to protect participant privacy. Aggregated data supporting the findings are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: VA, KS, DL
Data curation: VA
Formal analysis: VA
Investigation: VA
Project administration: VA
Resources: VA
Supervision: KS, DL
Writing – original draft: VA
Writing – review and editing: VA, KS, DL
All authors have read and agreed to the published version of the manuscript.
KS and DL contributed equally to this work as senior authors.

Conflicts of Interest

VA has a financial interest in Lighthouse Healthtech (a social enterprise company) created to develop and implement the digital tool referenced in this research paper. The other authors declare no conflicts of interest.

Multimedia Appendix 1
Structure of a focused acceptance and commitment therapy (fACT) session.
[DOCX File, 16 KB - humanfactors_v13i1e84412_app1.docx]

Multimedia Appendix 2
Screenshots of the software minimum viable product (MVP) interface showing electronic health record integration and client selection, session and psychometric review, session start, focused acceptance and commitment therapy (fACT)–specific clinical conversation prompts, and the clinical notepad.
[PNG File, 1159 KB - humanfactors_v13i1e84412_app2.png]

Multimedia Appendix 3
Screenshots of the software minimum viable product (MVP) interface showing wellness plan creation, including setting automated plan reminders for the client.
[PNG File, 1155 KB - humanfactors_v13i1e84412_app3.png]

Multimedia Appendix 4
Screenshots of the software minimum viable product (MVP) interface showing in-session experiential exercise feature, including voice actor selection, and the exercise script view.
[PNG File, 1129 KB - humanfactors_v13i1e84412_app4.png]

Multimedia Appendix 5
Detailed feature-specific user feedback.
[DOCX File, 3780 KB - humanfactors_v13i1e84412_app5.docx]

References

1. World mental health report: transforming mental health for all. : World Health Organization; 2022 URL: https://iris.who.int/server/api/core/bitstreams/40e5a13a-fe50-4efa-b56d-6e8cf00d5bfa/content [accessed 2025-08-20]
2. Mental health and COVID-19: early evidence of the pandemic's impact: scientific brief. : World Health Organization; 2022 URL: https://iris.who.int/server/api/core/bitstreams/b516621b-1d79-4f5c-a36d-9149b28f6576/content [accessed 2025-08-20]

XSL•FO
RenderX

3. The world health report: 2001: mental health: new understanding, new hope. : World Health Organization; 2001 URL: https://iris.who.int/server/api/core/bitstreams/d685bbd3-d7bf-4b07-be2f-982f483d4491/content [accessed 2025-08-20]
4. Annual update of key results 2019/20: New Zealand Health Survey. : Ministry of Health; 2020 URL: https://www.health.govt.nz/publications/annual-update-of-key-results-201920-new-zealand-health-survey [accessed 2025-08-20]
5. Annual update of key results 2023/24: New Zealand Health Survey. : Ministry of Health; 2024 URL: https://www.health.govt.nz/publications/annual-update-of-key-results-202324-new-zealand-health-survey [accessed 2025-08-20]
6. Aotearoa New Zealand's mental health services and addiction services: the monitoring and advocacy report of the Mental Health Commissioner. : Health and Disability Commissioner (HDC), New Zealand; 2020 URL: https://www.hdc.org.nz/media/zjugnstx/hdc-aotearoa-new-zealands-mental-health-services-and-addiction-services-2020.pdf [accessed 2025-08-20]
7. Robinson PJ, Reiter JT. Behavioral Consultation and Primary Care: A Guide to Integrating Services, 2nd edition: Springer; 2015.
8. Integrated primary mental health and addiction (IPMHA) services coverage and activity.: Health New Zealand (Te Whatu Ora); 2025. URL: https://www.wellbeingsupport.health.nz/assets/IPMHA-coverage-and-activity-data-July-2024.xlsx
9. Malatest International. Final evaluation report: implementation of the integrated primary mental health and addiction services. : Ministry of Health, New Zealand; 2022 URL: https://www.wellbeingsupport.health.nz/assets/IPMHA-evaluation-report-March-2022-Malatest-International.pdf [accessed 2025-08-25]
10. Eisner E, Faulkner S, Allan S, et al. Barriers and facilitators of user engagement with digital mental health interventions for people with psychosis or bipolar disorder: systematic review and best-fit framework synthesis. JMIR Ment Health 2025 Jan 20;12(1):e65246. [doi: 10.2196/65246] [Medline: 39832352]
11. Philippe TJ, Sikder N, Jackson A, et al. Digital health interventions for delivery of mental health care: systematic and comprehensive meta-review. JMIR Ment Health 2022 May 12;9(5):e35159. [doi: 10.2196/35159] [Medline: 35551058]
12. Espel-Huynh HM, Wang L, Lattie EG, et al. Clinical effects of asynchronous provider-guided practice sessions during blended care therapy for anxiety and depression: pragmatic retrospective cohort study. J Med Internet Res 2024 Oct 18;26(1):e60502. [doi: 10.2196/60502] [Medline: 39422996]
13. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three problems with current digital mental health research . . . and three things we can do about them. Psychiatr Serv 2017 May 1;68(5):427-429. [doi: 10.1176/appi.ps.201600541] [Medline: 28412890]
14. Moshe I, Terhorst Y, Philippi P, et al. Digital interventions for the treatment of depression: a meta-analytic review. Psychol Bull 2021 Aug;147(8):749-786. [doi: 10.1037/bul0000334] [Medline: 34898233]
15. Mohr DC, Silverman AL, Youn SJ, et al. Digital mental health treatment implementation playbook: successful practices from implementation experiences in American healthcare organizations. Front Digit Health 2025;7:1509387. [doi: 10.3389/fdgth.2025.1509387] [Medline: 40007642]
16. Huang K, Tulu B, Johnson S, Davis-Martin R. Implementation barriers to digital mental health screening: a qualitative study of CAT-MH implementations. Presented at: Proceedings of the 59th Hawaii International Conference on System Sciences; Jan 6-9, 2026. [doi: 10.24251/HICSS.2026.419]
17. Leung C, Pei J, Hudec K, Shams F, Munthali R, Vigo D. The effects of nonclinician guidance on effectiveness and process outcomes in digital mental health interventions: systematic review and meta-analysis. J Med Internet Res 2022 Jun 15;24(6):e36004. [doi: 10.2196/36004] [Medline: 35511463]
18. Berardi C, Antonini M, Jordan Z, Wechtler H, Paolucci F, Hinwood M. Barriers and facilitators to the implementation of digital technologies in mental health systems: a qualitative systematic review to inform a policy framework. BMC Health Serv Res 2024 Feb 26;24(1):243. [doi: 10.1186/s12913-023-10536-1] [Medline: 38408938]
19. Borges do Nascimento IJ, Abdulazeem H, Vasanthan LT, et al. Barriers and facilitators to utilizing digital health technologies by healthcare professionals. NPJ Digit Med 2023 Sep 18;6(1):161. [doi: 10.1038/s41746-023-00899-4] [Medline: 37723240]
20. Torous J, Linardon J, Goldberg SB, et al. The evolving field of digital mental health: current evidence and implementation issues for smartphone apps, generative artificial intelligence, and virtual reality. World Psychiatry 2025 Jun;24(2):156-174. [doi: 10.1002/wps.21299] [Medline: 40371757]
21. Gagnon MP, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]
22. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR mHealth uHealth 2020 Feb 20;8(2):e15935. [doi: 10.2196/15935] [Medline: 32130167]
23. Zakerabasali S, Ayyoubzadeh SM, Baniasadi T, Yazdani A, Abhari S. Mobile health technology and healthcare providers: systemic barriers to adoption. Healthc Inform Res 2021 Oct;27(4):267-278. [doi: 10.4258/hir.2021.27.4.267] [Medline: 34788907]
24. Coren MA, Lindhiem O, Angus AR, Toevs EK, Radovic A. Provider perspectives on implementing an enhanced digital screening for adolescent depression and suicidality: qualitative study. JMIR Form Res 2025 Apr 10;9:e67624. [doi: 10.2196/67624] [Medline: 40209027]
25. Brown T. Design thinking. Harv Bus Rev 2008 Jun;86(6):84-92. [Medline: 18605031]

XSL•FO
RenderX

26. Brown T. Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation: Harper Collins; 2009.
27. Altman M, Huang TTK, Breland JY. Design thinking in health care. Prev Chronic Dis 2018 Sep 27;15:E117. [doi: 10.5888/pcd15.180128] [Medline: 30264690]
28. Oliveira M, Zancul E, Fleury AL. Design thinking as an approach for innovation in healthcare: systematic review and research avenues. BMJ Innov 2021 Apr;7(2):491-498. [doi: 10.1136/bmjinnov-2020-000428]
29. Dam RF. The 5 stages in the design thinking process. IxDF - Interaction Design Foundation. URL: https://ixdf.org/literature/article/5-stages-in-the-design-thinking-process [accessed 2026-05-26]
30. Allen V, Lottridge D, Merry S, Stasiak K. Building a digital tool to support focused acceptance and commitment therapy practitioners in New Zealand primary care: a qualitative exploration of user needs to guide software feature development. J Contextual Behav Sci 2024 Apr;32:100762. [doi: 10.1016/j.jcbs.2024.100762]
31. Figma. URL: https://www.figma.com [accessed 2025-09-02]
32. Flutter. URL: https://flutter.dev [accessed 2025-09-02]
33. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. Presented at: Proceedings of the INTERCHI '93 (Conference on Human Factors in Computing Systems); Apr 24-29, 1993. [doi: 10.1145/169059.169166]
34. Virzi RA. Refining the test phase of usability evaluation: how many subjects is enough? Hum Factors 1992 Aug;34(4):457-468. [doi: 10.1177/001872089203400407]
35. Goodman E, Kuniavsky M, Moed A. Usability tests. In: Goodman E, Kuniavsky M, Moed A, editors. Observing the User Experience, 2nd edition: Morgan Kaufmann; 2012:273-326. [doi: 10.1016/B978-0-12-384869-7.00011-5]
36. Boren T, Ramey J. Thinking aloud: reconciling theory and practice. IEEE Trans Profess Commun 2000;43(3):261-278. [doi: 10.1109/47.867942]
37. Scupin R. The KJ method: a technique for analyzing data derived from Japanese ethnology. Hum Organ 1997 Jun;56(2):233-237. [doi: 10.17730/humo.56.2.x335923511444655]
38. Mohamedally D, Zaphiris P. Categorization constructionist assessment with software-based affinity diagramming. Int J Hum Comput Interact 2009 Jan 14;25(1):22-48. [doi: 10.1080/10447310802546690]
39. Jones EK, Banks A, Melton GB, Porta CM, Tignanelli CJ. Barriers to and facilitators for acceptance of comprehensive clinical decision support system-driven care maps for patients with thoracic trauma: interview study among health care providers and nurses. JMIR Hum Factors 2022 Mar 16;9(1):e29019. [doi: 10.2196/29019] [Medline: 35293873]
40. Blecker S, Pandya R, Stork S, et al. Interruptive versus noninterruptive clinical decision support: usability study. JMIR Hum Factors 2019 Apr 17;6(2):e12469. [doi: 10.2196/12469] [Medline: 30994460]

Abbreviations

DMHT: digital mental health tool
fACT: focused acceptance and commitment therapy
HIP: health improvement practitioner
IPMHA: Integrated Primary Mental Health and Addictions
MVP: minimum viable product
PHO: primary health organization

Edited by A Kushniruk; submitted 19.Sep.2025; peer-reviewed by D Bradford, T Ungar; revised version received 13.Jan.2026; accepted 09.Feb.2026; published 04.Jun.2026.

Please cite as:
Allen V, Lottridge D, Stasiak K
Evaluating a Digital Mental Health Tool for Implementation Into New Zealand's Integrated Primary Mental Health and Addictions Service: Usability Study
JMIR Hum Factors 2026;13:e84412
URL: https://humanfactors.jmir.org/2026/1/e84412
doi:10.2196/84412

© Vincent Allen, Danielle Lottridge, Karolina Stasiak. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 4.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Association Between Web-Based Patient Portal Messaging and Disease Activity in Inflammatory Bowel Disease: Retrospective Pilot Cohort Study

Pavan Vemulakonda[1], MD; Shriansh Singh[2], BS; Liza Khutsishvili[2], BS; Anushka Deogaonkar[3], MBBS; Valerie S Stark[1], MD, MPH; Marie L Borum[3], MD, EdD, MPH

[1]Department of Internal Medicine and Primary Care, George Washington University Hospital, 2150 Pennsylvania Avenue NW, Washington, DC, United States

[2]School of Medicine and Health Sciences, George Washington University, Washington, DC, United States

[3]Department of Gastroenterology and Liver Diseases, George Washington University Hospital, Washington, DC, United States

Corresponding Author:
Pavan Vemulakonda, MD
Department of Internal Medicine and Primary Care, George Washington University Hospital, 2150 Pennsylvania Avenue NW, Washington, DC, United States

Abstract

Background: Inflammatory bowel disease (IBD) requires ongoing monitoring and effective communication between patients and health care providers. Web-based patient portals may enhance engagement, but their association with disease activity remains unclear.

Objective: We aimed to evaluate the association between patient portal messaging frequency and disease activity among patients with IBD.

Methods: We conducted a retrospective cohort study using electronic health record data from an urban academic center. Patients with Crohn disease, ulcerative colitis, or indeterminate colitis were included. Portal messaging frequency over 2 years (July 2023-July 2025) was analyzed. Associations with demographic variables and disease activity were assessed using χ^2 tests (P<.05 was considered significant).

Results: Among 442 patients, 347 (78.5%) sent at least one portal message. Patients sent a mean of 5.27 (SD 7.15) messages over 2 years. High messaging frequency was associated with active disease (P=.047) and age group (P<.001). Patients with active disease were more likely to send ≥16 messages (n=14, 12.5% vs n=18, 6%). No significant associations were found with sex or race.

Conclusions: Higher portal messaging frequency was associated with active disease in IBD. Patient portals may serve as a tool for disease monitoring and early intervention.

(JMIR Hum Factors 2026;13:e87518) doi:10.2196/87518

KEYWORDS

patient portals; digital health; electronic messaging; electronic health records; telehealth

Introduction

Inflammatory bowel disease (IBD), encompassing Crohn disease and ulcerative colitis, is a chronic, relapsing inflammatory disorder of the gastrointestinal tract requiring lifelong management. The global burden of IBD has increased substantially, with rising incidence in both Western and newly industrialized countries, reflecting environmental and lifestyle changes [1]. Achieving sustained remission remains challenging due to unpredictable disease flares and complications.

Contemporary IBD management emphasizes a treat-to-target approach, focusing on tight disease control and early detection of inflammation using clinical, endoscopic, and biomarker-based parameters [2]. However, implementation of this strategy is often limited by delays in identifying symptom exacerbations between clinic visits.

Effective disease control depends not only on pharmacologic therapy but also on timely communication between patients and health care providers. Breakdowns in communication during transitions or between visits have been associated with adverse outcomes, including increased hospitalizations and medical errors [3]. In IBD, early patient-reported symptoms frequently precede objective disease worsening, underscoring the need for accessible communication systems.

XSL•FO
RenderX

Digital health technologies, particularly electronic health record (EHR)–integrated patient portals, have emerged as important tools to facilitate patient engagement and intervisit communication [4]. These platforms allow secure messaging, access to laboratory results, and remote monitoring. Their use expanded significantly during and after the COVID-19 pandemic, accelerating adoption of telemedicine and asynchronous care models [5].

In other chronic diseases, such as diabetes and hypertension, patient portal use has been associated with improved adherence, better disease control, and enhanced patient satisfaction [6-8]. Systematic reviews suggest that patient portals can improve health care processes and some clinical outcomes, although effects vary across settings [6].

However, evidence regarding portal use in IBD remains limited and inconsistent. Studies examining digital health interventions in IBD populations are heterogeneous, with variation in study design, definitions of engagement, and outcome measures. A study by Reich et al [9] demonstrated an association between portal messaging and improved quality of life in IBD patients but did not assess disease activity directly.

Importantly, the relationship between portal messaging behavior and objective disease activity remains unclear. Increased messaging may reflect worsening disease, proactive patient engagement, or both. Additionally, demographic patterns of portal use, including age, sex, and race, have not been consistently characterized in IBD populations.

There is limited evidence evaluating whether patient-initiated portal messaging correlates with disease activity in IBD and whether it can serve as a surrogate marker for clinical worsening. This study aims to evaluate the association between web-based patient portal messaging frequency and disease activity among individuals with IBD while also characterizing demographic patterns of portal use.

Methods

Study Overview

A retrospective cohort study was conducted using EHR data from an academic medical center in Washington, DC. Patients with a confirmed diagnosis of Crohn disease, ulcerative colitis, or indeterminate colitis who had at least one outpatient gastroenterology visit during the study period were included. Patients without follow-up within the preceding 3 years were excluded.

Data Collection and Variables

Demographic variables including age, sex, and race were extracted from the EHR. Clinical data were obtained from physician documentation, endoscopy reports, and patient-reported symptoms recorded in clinic notes.

Definition of Disease Activity

Disease activity was categorized as active or in remission based on a composite assessment of clinical documentation. Patients were classified as having active disease if the last clinical note indicated ongoing symptoms, disease flare, or escalation of therapy. Patients without such documentation were classified as being in remission or having unknown disease activity.

Because standardized disease activity indices were not consistently available in the retrospective chart review, the study used our definitions of clinical remission and active disease based on physician documentation and treatment escalation. We recognize that this definition is not a universally accepted measure of remission and may introduce variability in classification.

Portal Messaging Measurement

The EHR includes a web-based patient portal through which patients can send messages to their health care providers. The number of patient-initiated messages sent to gastroenterology care providers over a 2-year period (July 2023 to July 2025) was recorded. Messaging frequency was categorized into predefined groups (0, 1 - 5, 6 - 10, 11 - 15, 16 - 20, and >21 messages) based on the distribution of messaging frequency within the cohort.

Statistical Analysis

Descriptive statistics were used to summarize patient characteristics and messaging patterns. Associations between portal messaging frequency and demographic variables (age, sex, and race) as well as clinical variables (disease activity and IBD subtype) were analyzed using χ^2 tests of association. Statistical significance was defined as $P<.05$.

Ethical Considerations

This study was approved and deemed exempt by the institutional review board at George Washington University. Informed consent was waived due to the study carrying no more than minimal risk to participants. All data were deidentified and kept in password-protected spreadsheets. Study participants did not receive any compensation.

Results

Clinical Characteristics

A total of 442 patients with IBD were included in the analysis. Of these, 182 (41.2%) had ulcerative colitis, 249 (56.3%) had Crohn disease, and 11 (3%) had indeterminate colitis. Regarding race, 196 (44.3%) patients were White, 161 (36.4%) were Black, 7 (2%) were Asian, 28 (6%) were of other races, and 50 (11%) were of unknown race. A total of 253 (57.2%) patients were female and 189 (42.8%) were male. Patients had an in-person or telemedicine visit with a gastroenterology care provider on average 3.2 times during the 2-year study period.

Portal Messaging Patterns

Overall, 347 of 442 patients (78.5%) sent at least one portal message during the 2-year study period. Patients sent an average of 5.27 messages over this time frame. A total of 201 (45.5%) patients sent 1 to 5 messages, 87 (20%) sent 6 to 10 messages, and 59 (13%) sent 11 or more messages. The distribution of messaging frequency is presented in Table 1.

XSL•FO
RenderX

Table . Number of patient-initiated portal messages in patients with inflammatory bowel disease (N=442).

Number of patient-initiated portal messages within the previous 2 years	Patients, n
0	95
1 to 5	201
6 to 10	87
11 to 15	27
16 to 20	14
More than 21	18

Association With Age

Portal messaging frequency varied significantly across age age groups (P<.001). Only 16 of 136 (11.8%) patients aged 40 to 60 years sent ≥6 messages over the previous 2 years, in comparison to 64 of 190 (33.7%) patients aged <40 years and 39 of 116 (33.6%) patients aged >60 years (P<.001; χ^2_1=20.6 and P<.001; χ^2_1=17.5). The distribution of messaging frequency by age is presented in Table 2.

Table . Number of patient-initiated portal messages within the previous 2 years among patients with inflammatory bowel disease by age (N=442).

	Patients by age group (years), n					
Messages, n	<20	20-29	30-39	40-49	50-59	60-69
0	1	12	22	18	10	32
1 to 5	2	39	50	56	36	45
6 to 10	1	13	26	10	2	23
11 to 15	0	5	6	2	0	6
16 to 20	0	4	5	0	1	2
More than 21	0	4	0	1	0	8

Association With Disease Activity

Messaging intensity also differed by disease activity (P=.047; χ^2_1=3.9), with 14 of 126 (11.1%) patients with active disease sending 16 or more portal messages in the last 2 years in comparison to just 18 of 316 (6%) patients in remission or with unknown disease activity. The distribution of messaging frequency by IBD disease activity is presented in Table 3.

Table . Number of patient-initiated portal messages within the previous 2 years among patients with inflammatory bowel disease by disease activity (N=442).

Messages, n	Patients in remission, n	Patients with active disease, n	Patients with unknown disease activity, n
0	57	24	14
1 to 5	134	55	12
6 to 10	58	24	5
11 to 15	18	9	0
16 to 20	6	7	1
More than 21	11	7	0

Association With Sex and Race

No significant associations were observed between messaging intensity and sex, race, or IBD subtype.

Discussion

Principal Findings

This study demonstrates that higher patient portal messaging intensity is associated with active disease in individuals with IBD. Additionally, a bimodal distribution of portal engagement was observed, with higher messaging frequency among younger and older patients compared to middle-aged patients. These findings suggest that patient-initiated portal messaging may serve as a behavioral marker of disease activity and engagement with care.

Comparison With Prior Work

These findings are consistent with prior studies demonstrating increasing use of digital health tools for chronic disease management and the growing role of patient portals in facilitating communication and engagement [4,6]. Previous

XSL•FO
RenderX

work has shown associations between portal use and improved patient-reported outcomes, including quality of life in IBD populations [9]. However, most prior studies have not directly examined disease activity. This study extends the existing literature by demonstrating a measurable association between portal messaging behavior and clinical disease status.

The observed bimodal age distribution aligns with findings from prior studies in both IBD and other chronic disease populations. Increased engagement among younger individuals may reflect greater familiarity with digital technologies, whereas higher engagement among older patients may be driven by increased health care needs and comorbidity burden [6,7].

An important observation in this study is the lower level of engagement among middle-aged patients (aged 40 - 60 years), which contrasts with the higher engagement seen in both younger and older groups. This finding may reflect competing professional and personal responsibilities that limit engagement with health care systems in this population. Alternatively, it may indicate differences in digital health adoption or perceived need for care. While not the primary focus of the study, this pattern highlights a potential gap in digital engagement that warrants further investigation.

Additionally, although higher messaging intensity was associated with active disease, the directionality of this relationship cannot be established. Increased messaging may reflect reactive communication in response to worsening symptoms, proactive engagement in disease management, or a combination of both.

It is also possible that a subset of portal messages reflected symptoms attributable to overlapping irritable bowel syndrome (IBS) rather than IBD activity alone. Because message content was not systematically reviewed, differentiation between IBS-related and IBD-related symptoms was not possible with the current study design. This represents an important limitation and an area for future investigation.

Clinical Implications

These findings suggest that portal messaging may serve as a low-cost and scalable indicator of disease activity in IBD. Increased messaging frequency could potentially be used to identify patients at higher risk of disease exacerbation and to prompt earlier clinical intervention.

From a patient-centered perspective, high message volume may also indicate unmet clinical needs or inefficiencies in communication workflows. For example, some patients with active disease required more than 16 messages over a 2-year period, which may suggest opportunities for improved triage systems, nurse-led protocols, remote symptom monitoring tools, or earlier clinical follow-up to address concerns more efficiently.

Furthermore, the identification of lower engagement among middle-aged patients highlights an opportunity for targeted interventions aimed at improving digital participation and ensuring equitable access to care across age groups [4,6].

Portal use may also reflect patient preference for asynchronous communication compared with telephone-based communication. Factors contributing to portal use may include convenience, faster documentation of concerns, accessibility outside clinic hours, and integration of laboratory and medication records into the EHR.

Limitations

This study has several limitations. First, as a single-center retrospective study, the findings may not be generalizable to other health care settings with different patient populations or care delivery models. Second, disease activity was determined based on clinical documentation rather than standardized indices, which may introduce variability in classification. Third, the content of portal messages was not analyzed, limiting the ability to distinguish between clinically relevant and administrative communication. Fourth, patients who did not activate the web-based patient portal were included in the study, likely impacting the distribution of portal messages. Finally, the temporal relationship between messaging and disease flares was not assessed, precluding conclusions about causality.

Additionally, a baseline comparison of portal use rates among non-IBD gastroenterology populations was not available for this study, limiting contextual interpretation of whether the observed messaging frequencies were unique to patients with IBD or reflected broader portal use trends in gastroenterology practice.

Future research should focus on multicenter studies to validate these findings across diverse populations. Prospective studies examining the temporal relationship between portal messaging and disease activity could help clarify causality. Additionally, qualitative analysis of message content may provide insight into the types of communication that are most predictive of clinical deterioration. Interventional studies aimed at enhancing portal engagement, particularly among underrepresented groups such as middle-aged patients, may further improve patient outcomes.

Conclusions

In conclusion, patient portal messaging frequency is associated with disease activity in individuals with IBD. This finding suggests that digital communication patterns may serve as an important adjunct for disease monitoring and patient-centered care. Leveraging patient-portal data may enhance early detection of disease exacerbations and improve overall clinical management in IBD.

Given the exploratory nature of this work and the study's retrospective design, these findings should be interpreted as hypothesis-generating and supportive of future prospective validation studies.

Funding

This research received no specific funding from any funding agency in the public, commercial, or not-for-profit sectors.

XSL•FO
RenderX

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

References

1. Ng SC, Shi HY, Hamidi N, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet 2017 Dec 23;390(10114):2769-2778. [doi: 10.1016/S0140-6736(17)32448-0] [Medline: 29050646]
2. Turner D, Ricciuto A, Lewis A, et al. STRIDE-II: an update on the Selecting Therapeutic Targets in Inflammatory Bowel Disease (STRIDE) initiative of the International Organization for the Study of IBD (IOIBD): determining therapeutic goals for treat-to-target strategies in IBD. Gastroenterology 2021 Apr;160(5):1570-1583. [doi: 10.1053/j.gastro.2020.12.031] [Medline: 33359090]
3. Kripalani S, LeFevre F, Phillips CO, Williams MV, Basaviah P, Baker DW. Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. JAMA 2007 Feb 28;297(8):831-841. [doi: 10.1001/jama.297.8.831] [Medline: 17327525]
4. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148. [doi: 10.2196/jmir.4255] [Medline: 26104044]
5. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: evidence from the field. J Am Med Inform Assoc 2020 Jul 1;27(7):1132-1135. [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
6. Goldzweig CL, Orshansky G, Paige NM, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
7. Lyles CR, Sarkar U, Ralston JD, et al. Patient-provider communication and trust in relation to use of an online patient portal among diabetes patients: the Diabetes and Aging Study. J Am Med Inform Assoc 2013;20(6):1128-1131. [doi: 10.1136/amiajnl-2012-001567] [Medline: 23676243]
8. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012 Nov 26;14(6):e162. [doi: 10.2196/jmir.2238] [Medline: 23183044]
9. Reich J, Canakis A, Shankar D, et al. The use of an EHR patient portal (Mychart-Epic) in patients with inflammatory bowel disease. Crohns Colitis 360 2019 Oct 1;1(3):otz039. [doi: 10.1093/crocol/otz039]

Abbreviations

EHR: electronic health record
IBD: inflammatory bowel disease
IBS: irritable bowel syndrome

Edited by A Kushniruk; submitted 10.Nov.2025; peer-reviewed by K Sandberg, S Mandal; revised version received 14.May.2026; accepted 29.May.2026; published 24.Jun.2026.

Please cite as:
Vemulakonda P, Singh S, Khutsishvili L, Deogaonkar A, Stark VS, Borum ML
Association Between Web-Based Patient Portal Messaging and Disease Activity in Inflammatory Bowel Disease: Retrospective Pilot Cohort Study
JMIR Hum Factors 2026;13:e87518
URL: https://humanfactors.jmir.org/2026/1/e87518
doi:10.2196/87518

© Pavan Srihari Vemulakonda, Shriansh Singh, Liza Khutsishvili, Anushka Deogaonkar, Valerie S Stark, Marie L Borum. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 24.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Multicenter Usability Evaluation and Co-Development of a Digital Decision-Support Tool for Labor Triage: Mixed Methods Study

Mariana Tome[1], MD; Xavier Laurent[2], PhD; Kristiyan Georgiev[1], BSc; John Tolladay[1], PhD; Sarah Collins[1], BSc (Hons); Deborah Hedgecott[1], BSc (Hons); Lyuba V Bozhilova[1], MMath, DPhil; Jane E Hirst[3], MBChB PhD; Lawrence Impey[1], MBBS, FRCOG, PhD; Antoniya Georgieva[1], PhD

[1]Nuffield Department of Women's & Reproductive Health, Oxford Labour Monitoring, University of Oxford, Nuffield Department of Women's & Reproductive Health, University of Oxford, Oxford, England, United Kingdom

[2]Artificial Intelligence (AI) Competency Centre, University of Oxford, Oxford, England, United Kingdom

[3]The George Institute for Global Health, School of Public Health, Imperial College London, London, England, United Kingdom

Corresponding Author:
Antoniya Georgieva, PhD
Nuffield Department of Women's & Reproductive Health, Oxford Labour Monitoring, University of Oxford, Nuffield Department of Women's & Reproductive Health, University of Oxford, Oxford, England, United Kingdom

Abstract

Background: Digital decision-support tools for labor care remain limited, with few technologies successfully addressing the complex, time-sensitive decisions required during labor triage. Fit4Labour is a clinician-facing, data-driven research tool, currently under development, that combines computerized cardiotocography interpretation with maternal and fetal risk factors to generate individualized risk scores at labor onset. Its primary aim is to support clinicians in identifying fetuses who may require closer monitoring or expedited delivery, while simultaneously providing reassurance in low-risk cases. By promoting consistent communication and timely escalation of care, the Fit4Labour tool seeks to strengthen clinical decision-making. Understanding and addressing usability and implementation barriers will be critical to its adoption in clinical practice.

Objective: This study aims to evaluate whether a digitally co-developed labor decision-support tool (Fit4Labour) maintains usability and implementation readiness across NHS hospitals with differing clinical contexts.

Methods: We conducted a convergent parallel mixed methods study in 3 United Kingdom hospitals (December 2022 to May 2025). Phase 1 involved iterative co-development with midwives and doctors at Oxford University Hospitals NHS Foundation Trust; Phase 2 validated the locked version at Birmingham Women's and Children's NHS Foundation Trust and Buckinghamshire Healthcare NHS Trust. Participants completed scenario-based usability sessions evaluated with the System Usability Scale (SUS) and Single Ease Question (SEQ), and task completion time, followed by focus groups and interviews analyzed thematically.

Results: Twenty-six health care professionals participated: 12 in co-development (7 midwives, 5 doctors) and 14 in validation (8 midwives, 6 doctors) phases. During co-development at Oxford, the tool met the "excellent" usability threshold (mean SUS 82.1, SD 12.3), indicating readiness for the validation phase. The locked version (v4.0) independently met the "excellent" threshold at both validation sites (combined mean SUS 85.8, SD 10.2; Birmingham 80.7, SD 10.8; Buckinghamshire 90.8, SD 7.2). Task completion times were comparable across validation sites (Birmingham 10.3, SD 1.6 min; Buckinghamshire 9.2, SD 1.9 min), while SEQ scores were consistently high across all scenarios (mean 6.1/7, SD 0.8). Thematic analysis identified 12 themes within 3 domains: clinical integration and workflow, technology adoption and implementation, and patient safety and decision-making. Participants described the Fit4Labour tool as a supportive tool, "like a co-pilot," improving confidence in decisions with the potential to aid triage assessment. Perceived limitations included an incomplete risk factor profile and the need for minor technical adjustments or integration with existing hospital systems.

Conclusions: Through systematic co-development, the Fit4Labour tool met the established usability benchmark at 2 independent NHS hospitals with markedly different clinical contexts. Clinicians viewed the tool as a supportive aid providing a shared language for risk communication and enhanced decision-making while preserving clinical autonomy. Whether these usability findings translate to improved clinical outcomes in real-world practice requires prospective evaluation.

(JMIR Hum Factors 2026;13:e87589) doi:10.2196/87589

KEYWORDS

cardiotocography; fetal monitoring; usability testing; clinical decision support systems; labor management; health care technology; co-development methodology; maternity care; digital health; labor triage; user-centered design; implementation; data-based tools in health care

XSL•FO
RenderX

Introduction

Overview

Health care systems worldwide are undergoing rapid digital transformation, with clinical decision-support tools increasingly recognized as valuable tools for improving patient safety and care quality [1-5]. This transformation is particularly important in maternity care, where time-sensitive decisions during labor have profound consequences for both maternal and neonatal outcomes [6-8]. Yet, successful implementation depends not only on algorithmic accuracy, but also on user-centered design: tools must present information clearly, fit seamlessly into workflows, and support rather than disrupt decision-making [9-13].

Implementation Challenge in Health Care Technology

Traditional development models in health care technologies often involve limited user engagement until late validation stages, resulting in tools that perform well in controlled settings ("work as imagined") but fail in the reality of busy clinical environments ("work as done") [14-16]. Insufficient investment of time and resources in co-design and usability testing remains a major barrier to adoption. Systematic reviews consistently demonstrate that poor usability is linked to low uptake, which in turn can pose patient safety risks [17-19].

Maternity care poses additional unique challenges as clinicians often manage multiple patients simultaneously with diverse risk profiles, while maintaining high safety standards [20,21]. Front-line clinical staff (eg, midwives and junior doctors) in the United Kingdom (UK) National Health System (NHS) are often key decision-makers in triage, and hold unique insights into workflow barriers; however, their perspectives are frequently underrepresented during the design of new technologies [22,23].

Challenges in Labor Care

Fetal monitoring can be performed using intermittent auscultation or cardiotocography. Cardiotocography, a graphical record of fetal heart rate and uterine contractions over time, remains the standard tool for monitoring during labor [8,24]. However, cardiotocography interpretation demands specialist expertise, and yet, the same trace can be classified differently, which may lead to suboptimal decisions [25-27].

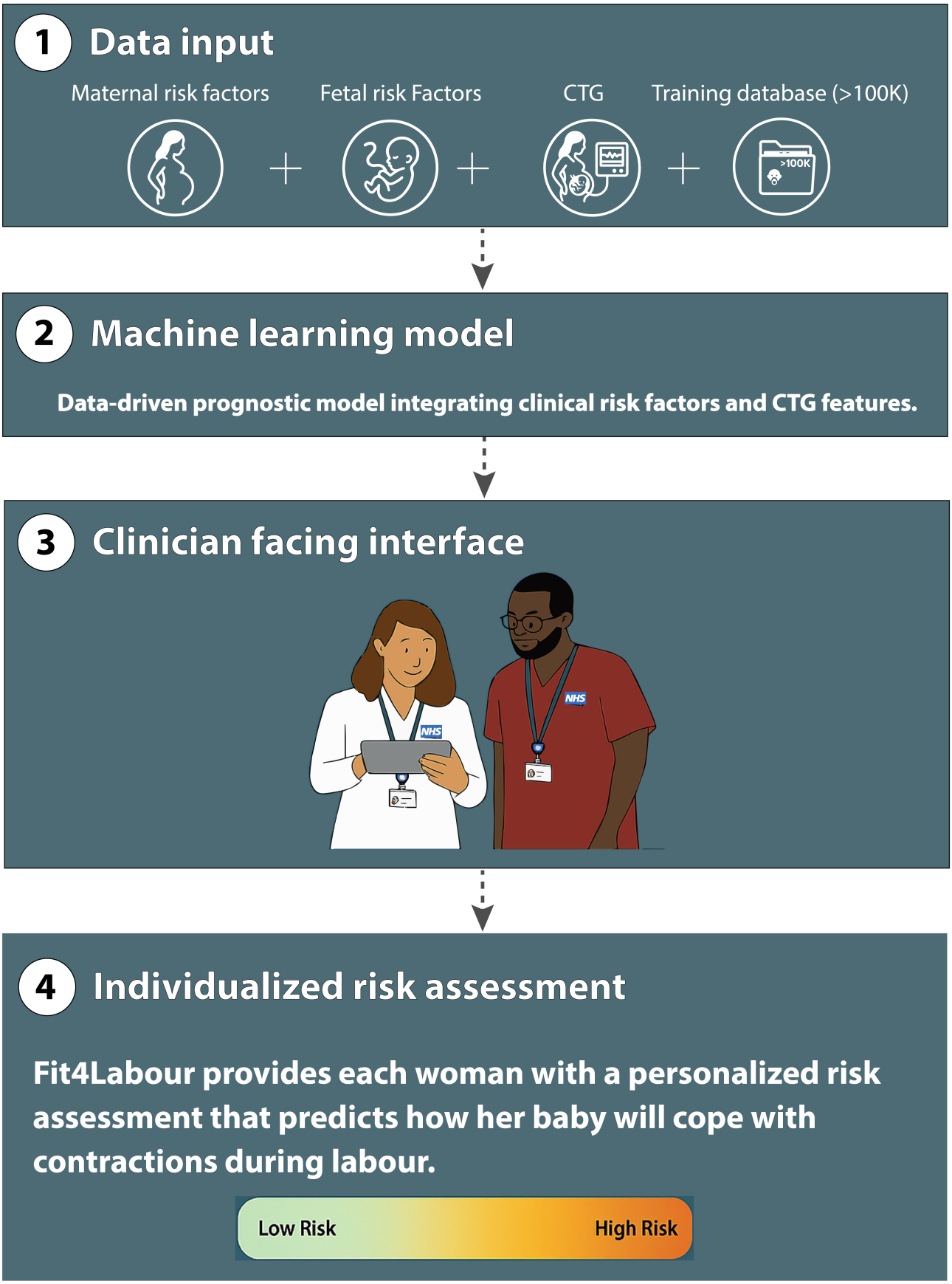
Labor triage involves the rapid assessment of maternal and fetal well-being at the onset of labor to determine appropriate monitoring, the appropriate place for birth, and whether urgent intervention (such as expedited delivery) may be required [22,24,28]. These early decisions shape the entire labor care pathway. Published investigations repeatedly identified that failures during intrapartum monitoring are associated with delays in both decision-making and escalation of care [29,30]. At the same time, modern maternity care increasingly emphasizes shared decision-making, requiring clinicians to communicate individualized risk information effectively to support women's autonomy [7,24,31].

The Fit4Labour Tool Innovation

The Fit4Labour tool (University of Oxford) was designed as a clinician-facing digital system to address these challenges by combining maternal and fetal risk factors (eg, maternal age, diabetes, and meconium-stained liquor) with computerized cardiotocography analysis to generate individualized risk scores indicating the likelihood of fetal compromise during labor (Figure 1). Unlike earlier computerized cardiotocography systems that rely mainly on human pattern recognition, the Fit4Labour tool applies a prognostic modeling approach that integrates multiple clinical risk factors with cardiotocography data to produce standardized, evidence-based predictions of adverse outcomes at delivery [32,33]. Developed by the Oxford Labor Monitoring Group through multidisciplinary, user-centered co-design involving clinicians, engineers, and patient representatives, the tool draws on a dataset of over 147,000 cardiotocographs linked to birth outcomes and prioritizes intuitive use and clear risk communication to support clinicians' intrapartum decision-making [34].

XSL•FO
RenderX

Figure 1. The Fit4Labour tool clinical workflow pathway from initial risk assessment to generation of the risk report. The diagram illustrates how the tool integrates into existing maternity workflows, combining maternal and fetal risk factors with computerized analysis of the cardiotocography trace to support triage, escalation, and communication of risk. CTG: cardiotocography.



This multicenter convergent parallel mixed methods study aimed to evaluate whether the Fit4Labour tool, developed through

XSL•FO
RenderX

intensive co-development at a single NHS hospital, maintained usability and implementation readiness across 2 NHS hospitals with differing clinical contexts.

Methods

Study Design and Conceptual Framework

We conducted a multicenter convergent parallel mixed methods study over a two-and-a-half-year period (December 2022-May 2025) to assess the usability and implementation readiness of the Fit4Labour decision-support tool. Quantitative usability metrics were combined with qualitative user experience data, analyzed in parallel and then systematically integrated. This approach allowed us to explore both the measurable performance and clinicians' lived experience, providing a comprehensive understanding of usability and potential implementation barriers.

The study was guided by human factors engineering principles and we used 3 theoretical approaches (Figure 2). First, user-centered design entails involving midwives and doctors throughout development [9]. We designed the tool based on user feedback, with the aim of aligning its features to clinicians' cognitive workflows and practical needs. Second, Distributed Cognition recognized that clinical decision-making is shared across people, tools such as checklists or cardiotocography traces, rather than occurring solely in individual minds [35]. The Fit4Labour tool was designed to provide both midwives and doctors with the same risk information, creating a shared risk score that supports team coordination and collective decision-making. Third, Sociotechnical Systems emphasized that technology operates within complex systems of people, processes, and environments [12,36]. This framework guided our assessment of whether participating hospitals had the necessary infrastructure, culture, and organizational support to implement the Fit4Labour tool effectively.

The study was conducted in 2 phases. Phase 1 was an iterative co-development process at Oxford University Hospitals NHS Foundation Trust (OUH) (referred to as Oxford throughout the paper), where the tool was refined through multiple cycles of usability testing and direct user feedback. Phase 2 involved multisite validation at Birmingham Women's and Children's NHS Foundation Trust (referred to as Birmingham throughout the paper) and Buckinghamshire Healthcare NHS Trust (referred to as Buckinghamshire), where the locked prototype version of the tool was evaluated to assess its usability and transferability across different organizational contexts.

XSL•FO
RenderX

Figure 2. Conceptual framework and study design showing the Fit4Labour tool version iterations. SEIPS: Systems Engineering Initiative for Patient Safety. SEQ: Single Ease Question; SUS: System Usability Scale.

XSL•FO
RenderX

Study Sites

Hospitals were intentionally selected to capture diverse maternity services and populations within the UK National Health Service. OUH (Oxford) is a large university teaching hospital with a well-established research infrastructure and strong links to the University of Oxford, serving both urban and rural populations. BWC (Birmingham) is a specialist teaching hospital providing high-volume maternity services to an ethnically diverse population, while BHT (Buckinghamshire) is a medium-sized district general hospital broadly representative of typical NHS maternity care, reflecting the realities of routine clinical practice. Together, these sites enabled evaluation across varied demographics, clinical practices, and infrastructure levels, thereby enhancing the generalizability within the NHS hospitals and ethnically diverse populations.

The Fit4Labour Tool

The clinician-facing Fit4Labour tool is a web-based application that runs on tablet devices or computers and can communicate with electronic maternity record systems. It integrates maternal and fetal risk factors with computerized cardiotocography analysis to produce color-coded risk categories (green=low, amber=moderate, red=high) accompanied by a numerical score indicating the risk of severe fetal compromise. Severe compromise is defined as a composite of intrapartum stillbirth, neonatal death (<28 days), neonatal encephalopathy (including hypoxic-ischemic encephalopathy), or neonatal seizures.

The tool displayed time-related changes in risk that updated in real time as the cardiotocography recording progressed, alongside detailed computerized analyses of individual cardiotocography parameters including baseline heart rate variability, accelerations, and decelerations. Reports were generated at 20- and 60-minute intervals and included integrated note-taking and record-storage features to support clinical documentation, handover continuity, and audit processes.

Study Objectives

The primary objective was to evaluate usability and acceptability of the Fit4Labour tool across 3 NHS sites using quantitative metrics (System Usability Scale, Single Ease Question, task completion times) and qualitative assessment of clinician perceptions. Secondary objectives were to explore organizational factors influencing transferability, examine how midwives and doctors perceived and interacted with the tool, assess its perceived potential to enhance patient care, and identify organizational and technical barriers that may impact adoption.

Participants and Recruitment

Participants comprised qualified midwives and doctors currently working in the selected maternity services, with direct patient care responsibilities in labor triage or delivery settings, and the need for and experience in cardiotocography interpretation as part of their clinical role. Department managers distributed information sheets to recruit participants. Three factors determined sample size: Nielsen's guidance for interface evaluation (5 - 8 participants per group typically uncover 80% - 85% of usability issues in interface evaluation studies, a pattern following the diminishing returns where each additional participant identifies fewer new problems), thematic saturation requirements for qualitative analysis, and practical constraints of multisite recruitment [37]. Using a relatively small number of samples, our convergent mixed methods design prioritized depth of insight over statistical power, consistent with exploratory implementation research [38,39]. All participants received a small honorarium (Amazon voucher) for their time.

Procedures

Between December 2022 and January 2024, the co-development phase took place in Oxford with participants testing successive prototypes in structured usability sessions. Feedback was sought on the user interface, layout, workflow, data entry, and risk visualization. Development followed a user-centered, multidisciplinary approach with iterative refinement over 4 versions at Oxford. Each version incorporated feedback from midwives and doctors, addressing user interface design, navigation, and risk visualization. Detailed version histories are provided in Table S3 in Multimedia Appendix 1. All design modifications were systematically logged, including user feedback prompting the change, implementation details, and impact assessment on overall functionality. Once usability goals, defined by high ease-of-use ratings, absence of major usability issues, and thematic saturation in qualitative feedback were achieved, the version was locked for validation testing. The locked version represented a stable version with no further interface or algorithmic modifications permitted during the validation phase. From November 2024 to May 2025, the locked version underwent prospective validation at Birmingham and Buckinghamshire. Participants followed the same protocol used at Oxford.

Usability Testing Sessions

Sessions lasted between 60 and 120 minutes and were co-facilitated by a clinical researcher (MT) and an experimental psychologist (XL) with extensive experience in usability testing. All sessions were conducted in dedicated rooms to minimize interruptions and ensure a consistent testing environment.

Testing was done in 2 separate stages (Figure S1 in Multimedia Appendix 1). In the task-based evaluation, participants first viewed a standardized 5-minute introduction video explaining the tool's purpose. The video did not include instructions on how to use the tool, as we sought to evaluate the user's intuitive navigation (Multimedia Appendix 1 video). The users were then given the tool and were asked to enter data for 4 progressive clinical scenarios, simulating common clinic triage tasks. This included baseline clinical risk factor entry, report generation, note-taking, risk factor updates, and advanced report features (Table S1 in Multimedia Appendix 1). Each scenario lasted up to 5 minutes, with facilitators observing silently and recording task completion times, errors, and usability barriers. If a task could not be completed within the allocated time, participants were instructed to move on to the next task.

After completing each scenario, participants were asked to rate the task's difficulty using the Single Ease Question (SEQ), a 7-point Likert scale (1=very difficult, 7=very easy). At the end of all tasks, they subsequently completed the System Usability Scale (SUS), a 10-item questionnaire generating an overall

XSL•FO
RenderX

usability score from 0 to 100 (Figure S2 and S3 in Multimedia Appendix 1).

The second stage was the qualitative exploration, which involved profession-specific focus groups or, where necessary, individual interviews. Discussions followed a structured topic guide exploring ease of use, clinical relevance, workflow integration, training needs, communication implications, and impact on clinical autonomy. All sessions were audio-recorded and transcribed verbatim.

Outcomes

Quantitative outcomes included SUS scores: values below 68 indicating below-average usability, between 68 and 80 indicating good usability, and above 80 indicating excellent usability [40,41]. SEQ ratings were used to assess ease of use following each clinical scenario. Lastly, task performance parameters were evaluated through completion times, error rates, and variability across participants. Qualitative outcomes captured themes relating to clinical integration, technology adoption, patient safety, and decision-making, drawing on data from focus groups and interviews.

Analysis

Quantitative Analysis

Descriptive statistics were reported as mean (SD) or median (IQR) according to the data distribution that was assessed by the Shapiro-Wilk test. No inferential statistical testing was performed. This was a deliberate choice as the primary aim of this study was to establish readiness and understand patterns of usability performance within each phase, rather than test population-level hypotheses. Moreover, phase 1 and phase involved different participants, tool versions, and clinical sites, precluding between-phase comparison being attributed to a single factor. The sample size was deemed sufficient for thematic saturation in the qualitative data. Data analysis and graphical representation were performed using the R software (version 4.5.3; R Core Team).

Qualitative Analysis

Transcripts were analyzed inductively using Braun and Clarke's 6-phase framework [42,43]. Three researchers (MT, XL, and DH) independently coded the transcripts, with consensus meetings to refine the coding framework and resolve any discrepancies. Themes were iteratively developed and mapped to usability and implementation domains. Saturation was defined as the absence of new themes across 2 consecutive coding rounds. NVivo software (Lumivero) was used for coding and data management.

Mixed Methods Integration

Quantitative and qualitative findings were integrated using a convergent parallel design [38]. Independent analyses were brought together in joint displays, with patterns categorized as convergence (alignment of findings), complementarity (mutual enrichment), or dissonance (contradictory insights). This approach enables a richer understanding of usability and implementation than either method alone.

Ethical Considerations

This usability study was conducted independently of, but in parallel with, the Fit4Labour technology evaluation and proof-of-concept study (ethics reference: 22/SC/0097). As this study involved health care professionals evaluating a prototype tool in a simulated setting, and did not involve patients, patient data, or clinical decision-making, formal research ethics committee review was determined not to be required. This was confirmed and agreed with the principal investigator at each of the 3 participating sites prior to recruitment.The study was registered with, and formally endorsed by, the research governance team at each of the three participating sites prior to recruitment, and this determination was further confirmed and agreed with the principal investigator at each site. All participants provided informed consent, recorded at the start of each audio-recorded session, after receiving detailed information regarding the study objectives and procedures, including their right to withdraw at any time without giving a reason. No identifiable patient data were collected or used in this study, and audio recordings and transcripts were stored securely with access restricted to the research team. Participants received a small honorarium (an Amazon voucher) in recognition of their time. Funding for the usability testing was provided by a National Institute for Health and Care Research (NIHR) Invention for Innovation (i4i) Product Development Award (NIHR202117); the funder had no role in study design, data collection, analysis, or interpretation.

Results

Participant Characteristics

Twenty-six health care professionals were recruited across 3 NHS sites: 15 midwives (58%) and 11 doctors (42%). Twelve participants (7 midwives and 5 doctors) contributed to Phase 1 co-development at Oxford, and 14 participants (8 midwives and 6 doctors) to Phase 2 validation at Birmingham and Buckinghamshire. Most participants (22/26, 85%) were women. Participants represented a wide range of seniority, from junior staff to consultant midwives and obstetricians, broadly reflecting the professional mix of UK maternity care (Table S2 in Multimedia Appendix 1). Although 25 participants completed the scenario-based usability testing, 26 took part in the interviews.

The Fit4Labour Tool Developments

The co-development phase at Oxford involved 3 iterative sessions, with each systematically incorporating users' feedback and testing progressively refined versions of the Fit4Labour tool (Figure 2). The first session with midwives (n=4) evaluated version 1.0, followed by a second midwives' session (n=3) a few months later testing version 2.0, and finally, version 3.0 was tested by doctors (n=5). This led to the final locked version 4.0, which was then tested in the 2 validation sites (Birmingham and Buckinghamshire). Detailed systematic documentation of design modifications is provided in Table S3 in Multimedia Appendix 1 and examples of user interface improvements in Figure S5 in Multimedia Appendix 1.

XSL•FO
RenderX

Quantitative Findings

System Usability Scale (SUS)

During the co-development at Oxford, the initial version (1.0) had a mean SUS of 77.5 (SD 15.1; rated as “good”), and by version 3.0 this had reached 83.5 (SD 9.1; rated as “excellent”). The co-development phase using the locked version (v4.0) had an overall mean SUS of 82.1 (SD 12.3) (Table 1, Figure 3), which indicated that the threshold for the locked version had been met (SUS>80).

Table . System Usability Scale scores by study phase and site. Learnability scores derived from System Usability Scale items 4 and 10.

Phase, site, and professional group	Participants, n	Version tested	SUS[b] score, mean (SD)	Range	Learnability, mean (SD)	Ratings[c]
Co-development						
Oxford						
1st Midwives session	4	v1.0	77.5 (15.1)	62.5 - 95	81.2 (16.1)	Good
2nd Midwives session	3	v2.0	85.8 (13.7)	70 - 95	79.2 (19.1)	Excellent
Doctors	5	v3.0	83.5 (9.1)	72.5 - 97.5	87.5 (12.5)	Excellent
Co-development total	12	v1.0 - 3.0	82.1 (12.3)	62.5 - 97.5	83.3 (14.4)	Excellent
Validation						
Birmingham						
Midwives	4	v4.0	76.3 (11.4)	65 - 90	62.5 (22.8)	Good
Doctors	3		86.7 (6.6)	77.5 - 92.5	75 (33.1)	Excellent
Site Subtotal	**7**		80.7 (10.8)	65 - 92.5	67.9 (25.9)	Excellent
Buckinghamshire						
Midwives	3	v4.0	94.2 (5.1)	87.5 - 100	95.8 (7.2)	Excellent
Doctors	3		87.5 (8.2)	77.5 - 97.5	95.8 (7.2)	Excellent
Site subtotal	6		90.8 (7.2)	77.5 - 100	95.8 (6.5)	Excellent
Validation total	13	v4.0	85.8 (10.2)	65 - 100	80.8 (23.7)	Excellent
Overall study (all participants[d])	25	v1.0 - 4.0	83.8 (11.3)	62.5 - 100	82 (19.5)	Excellent

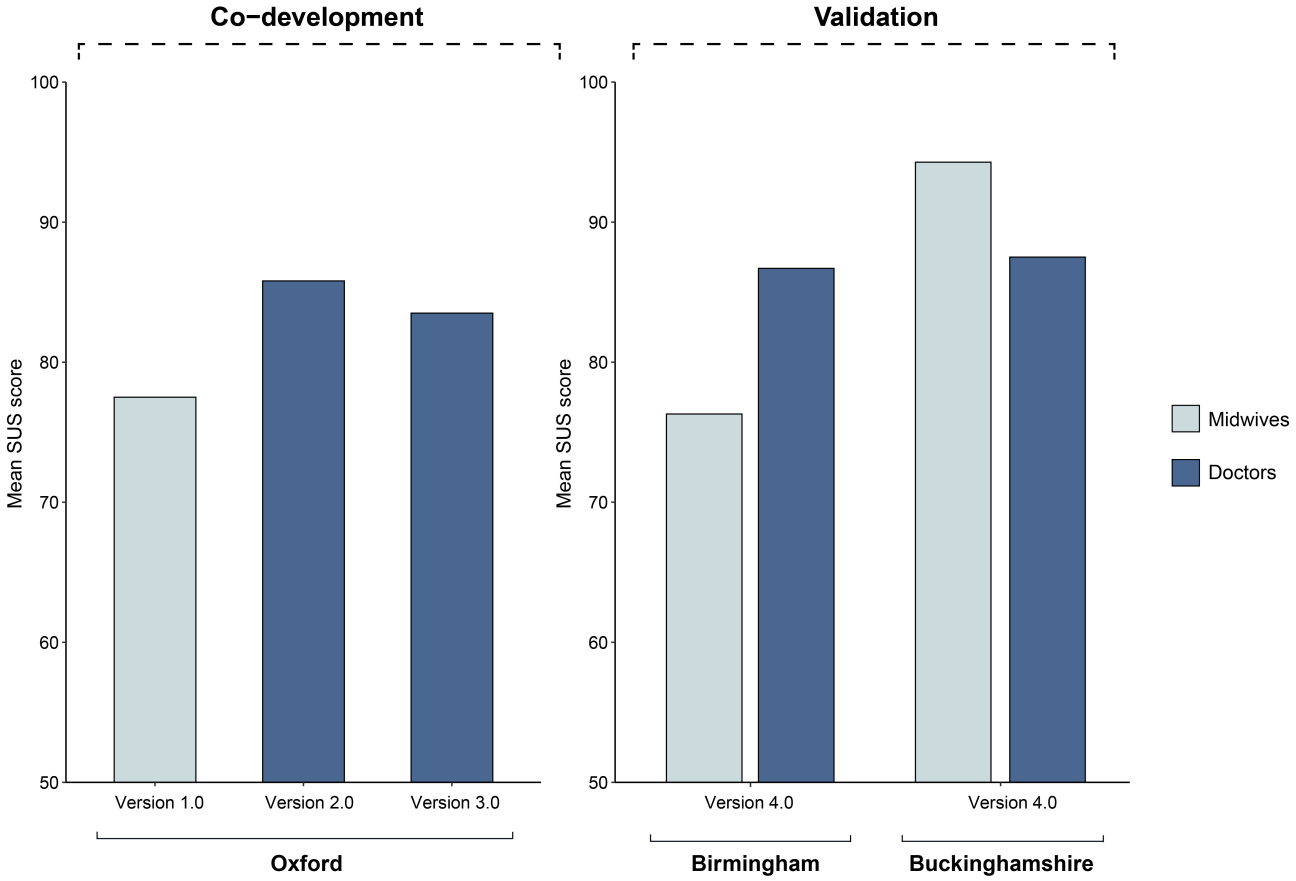
[b]SUS: System Usability Scale.

[c]Interpretation thresholds: <68=below average, 68–80=good, >80=excellent.

[d]One Birmingham midwife observed scenarios only and participated solely in focus groups, resulting in n=26 total participants but n=25 for SUS assessments.

XSL•FO
RenderX

Figure 3. SUS scores by study phase and site. During co-development, scores reflect iterative refinement across successive versions. During validation, the locked version (v4.0) independently met the "excellent" threshold at both sites. SUS: System Usability Scale.



The final locked version 4.0 independently met the "excellent" usability threshold during validation testing at 2 distinct sites, with a combined mean SUS score of 85.8 (SD 10.2), but with some intersite variability (Birmingham 80.7, SD 10.8 vs Buckinghamshire 90.8, SD 7.2). The SD of 10.2 within this phase reflects a relatively narrow spread of individual scores, suggesting a more predictable and uniform user experience across participants and sites.

Learnability scores (SUS items 4 and 10) showed an overall mean of 82 (SD 19.5). Co-development mean scores were 83.3 (SD 14.4) (Oxford) and validation mean scores were 80.8 (SD 23.7). Buckinghamshire showed a mean of 95.8 (SD 6.5), while Birmingham showed greater variability at a mean of 67.9 (SD 25.9). Professional group differences were modest (doctors: mean 87.5, SD 12.5; midwives: mean 80.4, SD 15.9 at Oxford; Table 1).

Single Ease Question (SEQ)

Task difficulty ratings demonstrated that participants perceived most scenarios as being "very easy" across all versions (mean SEQ 5.8 - 6.3 for Scenarios 1 - 3) (Table 2, Figure 4). Scenario 4, covering advanced functions, showed more variable ratings during co-development (v1.0 mean SEQ 5, SD 1.4; v2.0 mean SEQ 4.7, SD 1.2; and v3.0 mean SEQ 5.6, SD 0.7). In the validation phase, the locked version received SEQ ratings of mean 5.9 (SD 0.9) (Birmingham) and mean 6.3 (SD 0.8) (Buckinghamshire) for Scenario 4, indicating that advanced functions were perceived as easy to use. Overall, both validation sites demonstrated comparable ratings across all scenarios (Birmingham SEQ 6.2 vs Buckinghamshire SEQ 6.3), suggesting that version 4.0 had consistent perceived ease of use across scenarios.

XSL•FO
RenderX

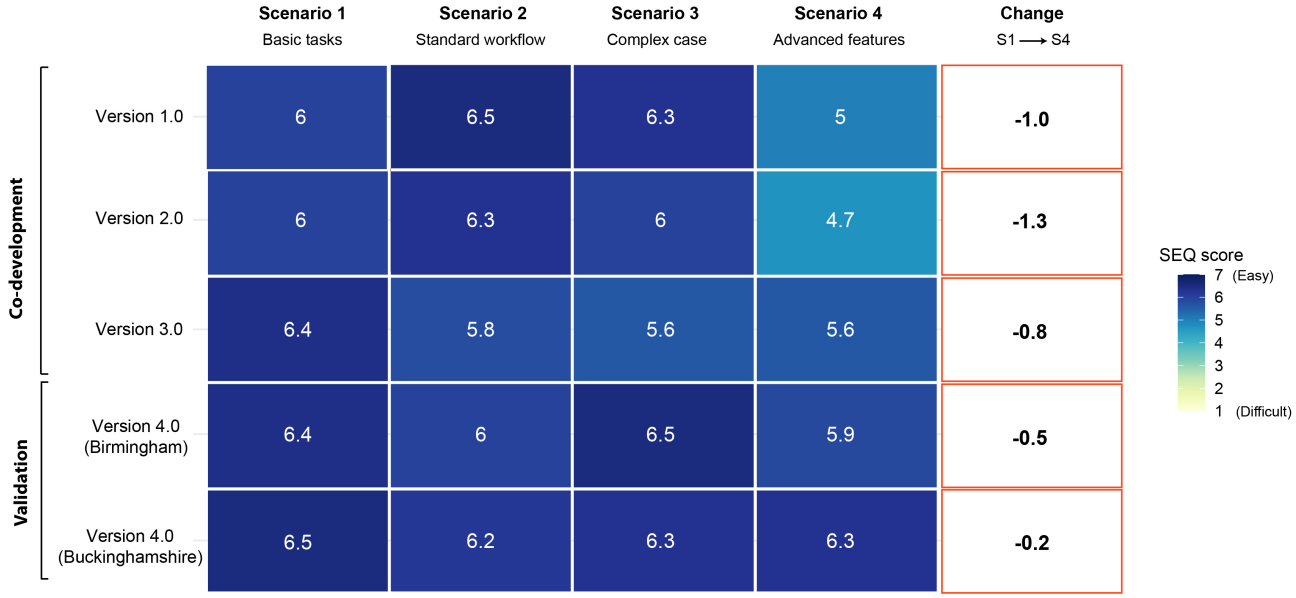
Table . Single ease question scores by study phase and scenario. Single Easy Question Scale: 1=Very Difficult, 7=Very Easy.

Phase, site, and group	Participants, n	Fit4Labour version tested	Mean SEQ[a] (SD)				
			Scenario 1	Scenario 2	Scenario 3	Scenario 4	Total
Co-development							
Oxford							
1st Midwives	4	v1.0	6 (0.8)	6.5 (0.6)	6.3 (0.5)	5.0 (1.4)	5.9
2nd Midwives	3	v2.0	6 (0)	6.3 (0.6)	6 (1)	4.7 (1.2)	5.8
Doctors	5	v3.0	6.4 (0.5)	5.8 (0.8)	5.6 (0.9)	5.6 (0.7)	5.9
Validation							
Birmingham							
Combined	7	v4.0	6.4 (0.7)	6.0 (0.9)	6.5 (0.5)	5.9 (0.9)	6.2
Buckinghamshire							
Combined	6	v4.0	6.5 (0.5)	6.2 (0.8)	6.3 (0.8)	6.3 (0.8)	6.3
Overall[b]	25	v4.0	6.3 (0.6)	6.1 (0.8)	6.1 (0.7)	5.7 (1)	6.1

[a]SEQ: Single Ease Question.

[b]One Birmingham midwife observed scenarios only and participated solely in focus groups, resulting in n=26 total participants but n=25 for SEQ assessments.

Figure 4. Heat map single ease question scores by study phase and scenario. Color-coded visualization of ease-of-use ratings (1=very difficult; 7=very easy) across scenarios. SEQ: Single Ease Question.



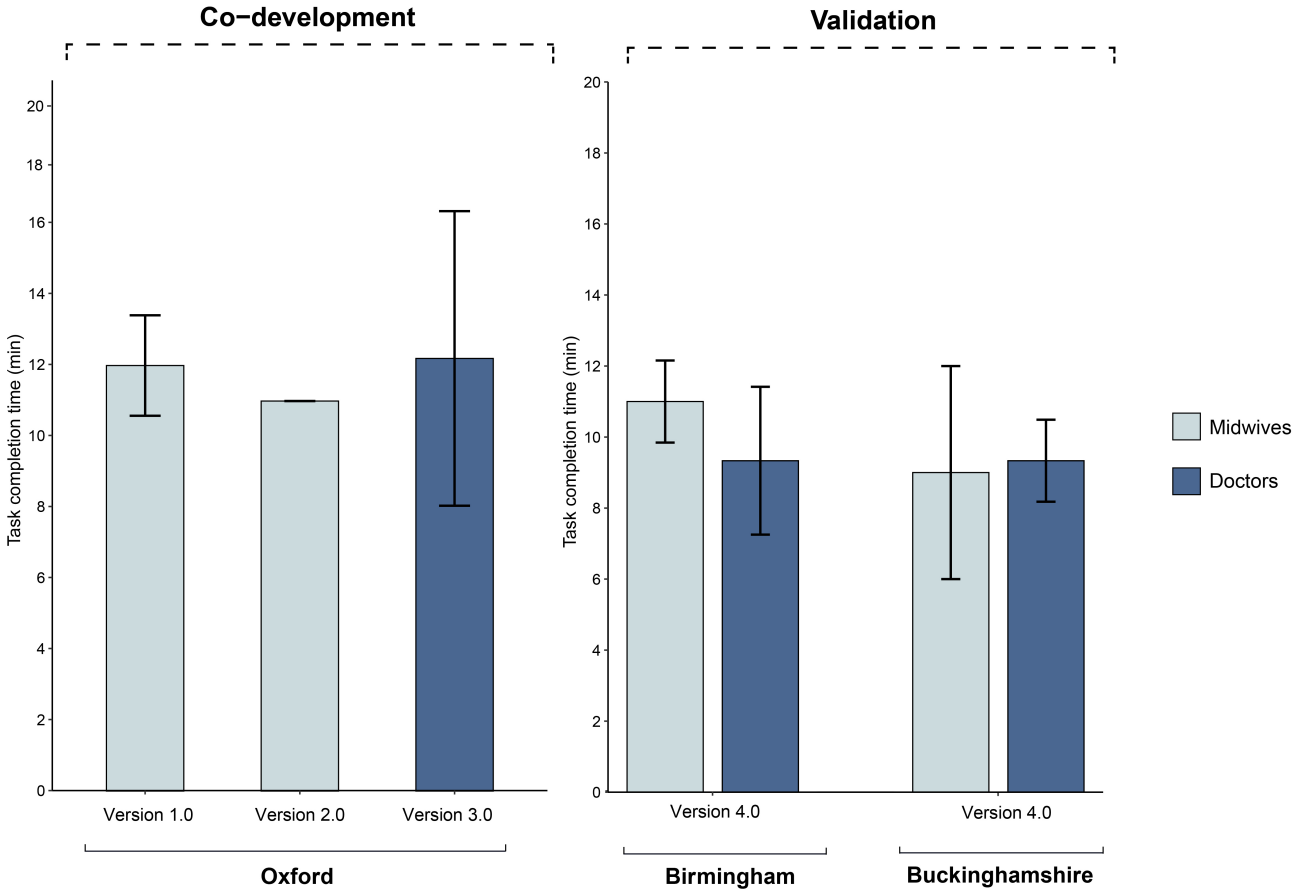
Task Efficiency and Errors

Cumulative task completion times in the co-development phase in Oxford ranged from 7 to 18 minutes, with a mean of 11.8 (SD 2.5) reflecting expected variation during refinements across successive prototype versions and iterative designs (Table S4 in Multimedia Appendix 1).

During the validation phase, the locked version 4.0 was completed in a mean of 9.8 (SD 1.8) minutes overall (Figure 5). Despite testing in different clinical environments and with different digital infrastructure, both validation sites had comparable completion times (Birmingham 10.3, SD 1.6 min and Buckinghamshire 9.2, SD 1.9 min), suggesting consistent and transferable usability of the locked version across sites.

XSL•FO
RenderX

Figure 5. Task completion times (minutes) for all 4 simulated clinical scenarios completed sequentially during usability testing across study phases and sites. Each bar represents the total time taken to complete the entire set of 4 scenarios, not the time required to use the Fit4Labour tool in real clinical practice for 1 patient.



Professional Group Comparisons

Usability was comparable between midwives and doctors. Doctors reported slightly higher SUS scores (mean 87.1, SD 8.2 vs mean 83, SD 13.8) and faster completion times (9.3 versus 10.1 min), but differences were modest. Midwives had higher variability in both. SEQ ratings were also very similar, indicating intuitive usability across professional groups.

Qualitative Findings

Thematic analysis identified 12 themes grouped within 3 domains: clinical integration and workflow, technology adoption and implementation, and patient safety and decision-making (Table 3). All thematic summaries and supporting quotes are presented in the Supplementary Analysis 1 - 3 and Figure S4 in Multimedia Appendix 1.

XSL•FO
RenderX

Table . Thematic framework showing the 3 domains and 12 themes identified through qualitative analysis across study sites. Summary of the 12 themes organized under 3 domains: clinical integration and workflow, technology adoption and implementation, and patient safety and decision-making.

Domain and theme number	Theme name	Description
Clinical integration and workflow		
1	Risk factors and clinical comprehensiveness	Need for more risk factor options (eg, IUGR[a], reduced fetal movements) and complex combinations to improve accuracy and avoid underestimating risk.
2	Timeline and workflow integration	Report timing options (20 versus 60 min) and workflow challenges including room availability, resources, and staffing constraints.
3	Screening and triage effectiveness	Value in objective screening and triage with color-coded risk stratification (green, red, amber) for prioritizing care.
4	Escalation support for midwives	Enhanced midwife-doctor communication, confidence building for junior staff, and structured escalation protocols.
Technology adoption and implementation		
5	Training and digital confidence	Digital skills vary; success linked to training champions, mandatory sessions, and understanding tool rationale.
6	Ease of use	Intuitive design, straightforward interface, easy demographic input, and successful co-development outcomes.
7	Technology infrastructure barriers	Internet connectivity, Wi-Fi issues, device charging, CTG[b] compatibility, and hardware availability.
8	Digital overload and change management	Too many tools and software can overwhelm; resistance to change needs addressing through buy-in engagement strategies.
Patient safety and decision-making		
9	Decision-making safety net	"Co-pilot" for clinical support, especially valuable in gray areas of CTG interpretation and safety net for junior staff.
10	Risk communication and intervention requests	Support for informed patient counseling with individualized risk information and transparent communication strategies. Concerns about sharing risk information could increase patient requests for interventions.
11	Clinical autonomy and professional responsibility	Tool should support and not replace clinical judgment; concerns around medico-legal risk and deskilling.
12	System-wide implementation and organizational adoption	Requires strong implementation planning, organizational readiness, and consistent departmental adoption and unified protocols.

[a]IUGR: intrauterine growth restriction.

[b]CTG: cardiotocography.

Clinical Integration and Workflow Domain

Across all sites, participants described the Fit4Labour tool as a valuable screening and triage aid, with color-coded risk stratification enabling rapid and intuitive prioritization. One Oxford doctor explained that the colors "acts as a powerful shortcut… one of the fastest and easiest ways to express risk."

The tool was also seen as an effective escalation support mechanism, particularly for junior midwives who valued having objective evidence when calling senior staff: "You feel more confident having documentation that says you need to seek urgent medical advice" (Midwife, Oxford). Similarly, a Buckinghamshire midwife highlighted that the outputs facilitated communication by noting that "the Fit4Labour tool is one language…It helps put into words when you are not happy with a trace."

XSL•FO
RenderX

However, limitations were also noted, including an incomplete risk factor profile, such as the absence of reduced fetal movements, small-for-gestational-age, previous stillbirth, and antepartum hemorrhage. Clinicians felt that these omissions could undermine accuracy and provide false reassurance, with one Oxford doctor commenting that "missing risk factors gives an impression of being very accurate and specific but may not be if you don't input certain combinations." Views on workflow integration varied, with some suggesting that 60-minute reports were impractical in fast-turnaround triage but potentially feasible in induction bays or antenatal wards.

Technology Adoption and Implementation Domain

Ease of use was widely recognized by all participants and was commonly described as being "very intuitive and straightforward." A midwife at Oxford noted its accessibility even when fatigued, explaining that "it works for tired brains." Some participants with visual impairments or neurodivergent profiles, including those who were color-blind or dyslexic, commented positively on accessibility, with one noting, "I can read it really clearly, I'm dyslexic."

Nonetheless, a common theme was that structured training was essential given variation in digital literacy. A Birmingham doctor observed that "people that struggle with technology will still struggle because their confidence with IT is low … it's not that this is difficult to use." Suggested strategies included champion-led training and mandatory education programs. "People are always anxious with new technology, it needs mandatory training to use it properly, more reassurance to not miss things before using it" (Oxford Midwife).

Perceived implementation challenges extended to both infrastructure and organizational culture. Participants raised concerns about digital overload and imposed software changes in the NHS together with the need for senior buy-in, with one participant in Birmingham noting that "nobody likes change… you'll need a lot of buy-in from senior people... it'll snowball, just like iPhones when they came in.." Another common topic was infrastructure constraints, such as poor Wi-Fi, delays in IT support, missing chargers, and limited hardware, which was cited in Birmingham ("Wi-Fi doesn't work, and you can't see the CTGs"). In contrast, Buckinghamshire participants reported implementation would likely be easier as the hospital was already paperless.

Patient Safety and Decision-Making Domain

A prominent theme was the "co-pilot in a plane" analogy, first raised at Oxford, framing the tool as a supportive safety net rather than a replacement for clinical judgment, with a Birmingham doctor emphasizing that "the physician still has the ultimate decision-making." The Fit4Labour tool was seen as particularly reassuring for junior staff and when presented with complex or ambiguous cases.

Participants valued the Fit4Labour tool to facilitate risk communication with patients, providing a "common language" for explaining probabilities and counseling women requesting care outside guidelines: "We are not informing patients of their real risks because we don't know, but this is what Fit4Labour can bring" (Doctor, Oxford). Some clinicians cautioned that transparent risk data might increase cesarian requests "for the price of honest communication of risk," while others felt it would reduce paternalism and enhance women's autonomy. Buckinghamshire participants emphasized its role in countering misinformation, particularly the perception that cesarian birth is inherently safer: "Patients think c-section is the safer option, often misled by social media; the tool can help address this and decrease requests."

Concerns regarding over-reliance and deskilling were also raised, with one participant warning that "it can deskill people… you will take shortcuts after a while." *Moreover,* participants highlighted the need for clear protocols to ensure the tool complements rather than replaces expertise. Perspectives evolved across sites, shifting from early skepticism about being "told what to do by a machine" to recognition of the tool as "an important piece of the jigsaw." Birmingham participants stressed balanced integration, underscoring that "your clinical judgment is always going to be important—see the patient beyond the technology." They also noted potential to rebuild trust in maternity care, with one participant observing that "maternity is in the public eye because of serious incidents and poor care. Social media doesn't help. Trust is broken. Fit4Labour can decrease this."

Mixed Methods Integration: Joint Display Analysis

Integration of quantitative and qualitative findings revealed strong convergence, with both complementary and contrasting insights (Table 4).

Convergence (direct alignment) was evident as high SUS and SEQ scores aligned with qualitative reports of intuitive design and ease of use, while completion times paralleled clinicians' descriptions of smoother workflows, which may translate to more efficient practice.

Complementarity (mutual enhancement) also emerged, with quantitative improvements in task completion times complementing qualitative concerns about resource constraints and highlighting both usability and systemic challenges. Quantitative differences between midwives and doctors were minor, but differences became evident for qualitative themes around escalation support and risk communication.

Dissonance (contrasting perspectives) appeared in 2 key areas: while learnability scores were high, participants still requested structured training, indicating that initial perceived ease of use of the Fit4Labour tool does not eliminate the need for formal instruction regarding its clinical application. Similarly, strong usability ratings contrasted with concerns about incomplete or "missing" risk factor coverage—highlighting that usability alone does not guarantee clinical comprehensiveness and further understanding of the limitations of the tool is needed.

XSL•FO
RenderX

Table . Mixed methods integration summary: convergence, complementarity, and dissonance across findings. Joint display illustrating alignment between quantitative and qualitative results, highlighting usability, workflow efficiency, and training needs.

Integration pattern and key findings	Evidence
Convergence	
Excellent usability	Validation phase SUS[a] scores (mean 85.8, SD10.2) independently met the "excellent" threshold and aligns with qualitative feedback stating both “very intuitive” and “easy to use.”
Co-development progression	SUS scores across co-development versions reflect iterative refinement reaching the excellent threshold by v3.0, consistent with participants’ growing familiarity with the tool’s ’co-pilot’ function.
Advanced feature complexity	More complex scenarios (eg, 4) had lowest scores (mean 5.7, SD 1.0), which aligns with concerns regarding the need for “structured and mandatory training.”
Complementarity	
Workflow integration	Within the validation phase, task completion times of 9.8 minutes were broadly feasible.
Performance predictability	Narrow SD (1.8 min) within the validation phase supported workflow consistency needs for ’rapid turnover in triage.
Professional group needs	Quantitative differences between midwives and doctors complemented qualitative emphasis on escalation support.
Infrastructure feasibility	Fast completion times (9.8 min) supported feasibility despite “Wi-Fi,” “charging,” and “compatibility” concerns.
Dissonance	
Learning versus training	High learnability scores (89.6%) contrasted with participants' expressed need for extensive training, reflecting the complexity of using the tool in clinical practice
Performance versus comprehensiveness	Excellent usability scores contrasted with concerns regarding “limited risk factors profile.”

[a]SUS: System Usability Scale.

Discussion

Main Findings

This multicenter mixed methods study demonstrated that the Fit4Labour tool developed through intensive co-design at a single NHS hospital independently met the “excellent” usability threshold (SUS >80) in simulated testing conditions at both validation sites, with consistent task completion times across 2 hospitals with different organizational contexts and digital maturity. Birmingham midwives reported lower usability scores (mean 76.3, SD 11.4), though still within the ’good’ range. This likely reflects the more critical evaluation standards based on their extensive clinical experience with existing triage systems, which strengthens the robustness of our validation findings.

Qualitative data reinforced these findings: participants described the Fit4Labour as a helpful triage and escalation tool, particularly for junior staff and as a shared “one language” for risk communication across professional boundaries. The spontaneous “co-pilot” analogy indicated that clinicians viewed the tool as supportive rather than substitutive, preserving autonomy while easing cognitive load. This preference for augmentation over automation mirrors findings from emergency triage settings, where professionals and patients alike favor hybrid digital-human approaches over fully automated solutions [44]. Although midwives and doctors have distinct roles, both groups reported that the interface met their needs, suggesting that a single design can accommodate diverse professional requirements in simulated testing conditions.

Clinicians also identified areas for improvement, including incomplete risk factor profile coverage, infrastructure barriers that may complicate seamless integration, and the need for structured training to support consistent use.

Comparison With Prior Work

Earlier evaluations of computerized cardiotocography systems in randomized controlled trials have shown limited benefit, with no clear improvements in neonatal outcomes or reductions in unnecessary interventions [45,46]. Systematic reviews suggest that for decision-support tools to make a real difference, they must be interpretable, reproducible, and trusted by clinicians, rather than simply replicating the same inconsistencies seen in human interpretation [45,47,48]. Fernandes et al [49], in a systematic review of intelligent clinical decision support systems for emergency triage, found that many such systems lack a formal implementation phase and that validated systems consistently improved clinician decision-making; this study directly addresses this gap by combining algorithmic development with structured multicenter usability evaluation.

The Fit4Labour tool represents a promising evolution in this field. Unlike earlier systems based purely on human

XSL•FO
RenderX

cardiotocography pattern recognition, it is a prognostic model trained on data from a large, multicenter clinical database, allowing it to produce standardized, evidence-based risk scores [32,33]. By providing clear, data-driven outputs, clinicians perceived the tool as having the potential to reduce variation in cardiotocography interpretation and to support less experienced staff, while complementing rather than replacing senior clinical judgment, perceptions that align with the tool's design intent but require prospective clinical evaluation to confirm.

These advantages are particularly relevant in the current maternity safety landscape. The UK Healthcare Safety Investigation Branch recently highlighted how unpredictable workloads and a constant need to juggle tasks in labor wards can compromise situational awareness and delay recognition of fetal compromise [50]. Participants in this study felt that the Fit4Labour tool could help maintain an overview of the full clinical picture, reduce cognitive load, and minimize the risk of overlooking important patient information. By alerting users when data is missing or incomplete, the tool also aligns with human-factors principles of error-tolerant design [51].

Similarly, national inquiries such as Ockenden and MBRRACE have underscored the role of poor communication and delayed escalation in adverse outcomes [29,30]. These findings resonate with observations from our study, where participants described using vague, non-specific expressions over the phone (eg, "I'm not happy with this trace") when escalating concerns about cardiotocography traces. Clinicians suggested that a numeric risk score could help reduce subjectivity in intrapartum communication, though this needs to be confirmed in adequately powered prospective studies.

Previous evaluations of computerized cardiotocography systems have reported variable user acceptance due to unclear risk calculations and scoring methods [45,52,53]. In contrast, participants found the Fit4Labour tool intuitive and transparent, which likely contributed to its positive usability scores. They also expressed interest in greater clarity regarding how individual maternal and fetal factors influence the overall risk score, including the relative weight of each contributing variable. Showing how these factors, together with cardiotocography interpretation and population data, feed into the score could help clinicians better understand and trust the tool's output, rather than viewing it as a "black box."

Finally, participants valued the unified interface that supported both midwives and doctors. Our findings indicate that a shared, role-sensitive interface can accommodate diverse professional needs while promoting interprofessional communication. This aligns with human-factors literature advocating shared mental models to bridge doctor–midwife silos, reduce fears of hierarchical barriers when escalating, and ensure complete, consistent communication in high-risk settings [54,55]. The slightly lower scores observed among senior midwives at Birmingham are consistent with evidence from emergency department settings, where senior physicians have been shown to report greater frustration with health IT systems than residents, attributed to higher expectations shaped by prior workflows. [56]

Human Factors Innovation: Cognitive Support and Communication

Intrapartum care involves complex decisions under uncertainty, where cognitive load, fatigue, and time pressure can compromise situational awareness. Participants described the Fit4Labour tool as "working for tired brains," highlighting its perceived value under conditions of fatigue, multitasking, and workload pressures, particularly relevant in maternity care. The design also appeared to support progressive disclosure: junior staff could rely on simple, color-coded outputs for rapid decision-making, while senior clinicians valued the ability to review detailed cardiotocography metrics and contributing risk factors. This layering helped ensure that transparency and autonomy were maintained without overloading users. In addition, when information was incomplete or omitted, the tool prompted users to add missing data, helping to reduce errors.

Clinical Integration and Implementation Considerations

In our analysis of over 147,000 births, the Fit4Labour algorithm identified about 1 in 3 babies who later showed signs of severe compromise, using only the first 20 minutes of cardiotocography at the start of labor [34]. Many of these babies were not recognized until 8 - 10 hours later, when emergency intervention was needed, suggesting that harm experienced during the birth process might be curtailed or even prevented. Although these results come from retrospective cohort analysis rather than real-time use, participants suggested that making risk visible, readily available, and quantifiable could support earlier and more consistent decision-making, especially when early cardiotocography changes are subtle. With maternity-related litigation in the UK now exceeding £37 billion (more than the annual cost of running maternity services), there is an urgent need for preventive, evidence-based innovations that reduce avoidable harm and help rebuild trust in care [57, 58].

Participants across all sites emphasized that structured clinician education remains essential for effective and consistent implementation. Training should go beyond button-clicking or navigation (interface use), focusing on supporting decisions and helping clinicians interpret risk, and communicating effectively with women. This approach ensures that the Fit4Labour *tool* outputs are translated into safe, evidence-based care rather than reducing decision-making to a "tick-box" exercise.

Interestingly, over nearly 2 and a half years of sessions, a noticeable shift in participants' views emerged, reflecting the broader uptake of digital technologies and AI in health care, with early anxieties about "being told what to do by a machine" evolving into recognition of a novel digital tool as a potentially "important piece of the puzzle" that enhances, rather than constrains, professional autonomy. This gradual shift also highlights increasing confidence in technology and the potential of digital tools to support decision-making and improve escalation pathways [59]. This trajectory is consistent with Longoni et al.'s framework, in which resistance to medical artificial intelligence is primarily driven by "uniqueness neglect," the concern that algorithms cannot account for individual patient characteristics. Fit4Labour tool's transparent,

XSL•FO
RenderX

interpretable outputs and preservation of clinical autonomy directly address this concern. [60]

Organizational readiness is likely to play a pivotal role in adoption. Digitally mature sites, such as Buckinghamshire, noted that the Fit4Labour tool could be easily integrated into existing workflows, while others cited unreliable Wi-Fi, limited computer access, poor and delayed IT support, and competing software as major barriers. These findings echo 'Consolidated Framework for Implementation Research (CFIR) v2.0,' emphasizing that infrastructure, culture, and implementation support, through training, local champions, and change management, are as important as usability for successful adoption [61].

Equity considerations also featured prominently in group discussions. Clinicians acknowledged the steps taken during co-development to promote inclusiveness in the user interface. These included design choices tested for accessibility among color-blind and dyslexic users, adaptations for staff wearing gloves when using touch tablets, such as larger buttons, and adjustments to font and button sizes to accommodate differences in visual acuity and hand size. However, participants also highlighted areas for further improvement. The current version is available only in English, which may create barriers for clinicians when explaining risk to non–English-speaking patients. Participants noted that using an interpreter or relying on a family member for translation, particularly over the phone, can make it more difficult to use the tool effectively in real time. Cultural factors may also influence how risk is communicated and perceived, underscoring the need for continued research and iterative refinement. Finally, participants noted that ethnic minority groups were not specifically represented and are likely underrepresented in the dataset used to refine the Fit4Labour algorithm, which could limit its generalizability. Ensuring inclusive design and representative data will be essential to prevent the reinforcement of existing inequities in maternity care [62].

Practical Implication for Digital Health Tools Development

Digital health projects often allocate significant resources to local customization (often between 15% - 20% of implementation budgets) [63-65]. In this study, the Fit4Labour tool maintained high usability across 3 hospitals with different organizational contexts and IT infrastructures when tested in the same version. This observation raises a practical question of whether intensive co-development at the outset might reduce the need for postdeployment customization. However, establishing this would require prospective studies comparing implementation costs and outcomes for tools developed with versus without intensive participatory design.

Future Research

Our team will conduct a regulated clinical evaluation of the Fit4Labour tool across multiple NHS sites to assess real-world safety, usability, and workflow integration, in line with Medicines and Healthcare products Regulatory Agency requirements for medical device approval [66]. This study will focus on outcomes such as escalation timeliness, reliability of communication, handover quality, and patient safety indicators.

We will also examine factors influencing sustained use and implementation, including the impact of structured training, local clinical champions, and integration into existing hospital systems. In parallel, our team will build on early health economic modeling to assess whether the usability patterns observed in this study translate into measurable cost savings and better clinical outcomes.

Beyond our own work, external validation in non-NHS health care systems is an important next step to determine whether the tool's core design principles transfer to other diverse health care contexts.

Finally, we will continue to engage directly with women, families, and PPI panels to explore how different ways of presenting risk information affect perception, autonomy, and trust. Embedding patient perspectives throughout these evaluations will help ensure that digital tools strengthen shared decision-making and enhance both safety and the experience of care.

Strengths and Limitations

This study has several strengths. The multicenter design, spanning 3 NHS hospitals with different organizational contexts and levels of digital maturity, provided a robust test of both usability and scalability across diverse settings. The mixed methods approach enabled convergence of quantitative usability metrics with qualitative insights into workflow, communication, and professional practice. Importantly, the locked-version (v4.0) validation offered a more rigorous test of scalability across diverse settings than is usually attempted in health care technology evaluations.

Our study also has important limitations. Usability testing was conducted in structured environments rather than during real-time clinical care, which may have increased observed task efficiency. However, as all sites used the same testing protocol, the relative comparisons between validation sites remain valid, even if absolute values might differ in clinical practice. Recruitment through departmental managers could have favored digitally confident clinicians, although efforts were made to capture a range of roles and experience. The involvement of members of the development team in evaluation may also have introduced bias, but this was mitigated by standardized protocols and independent analysis. The relatively small sample size (n=26) is consistent with Nielsen's guidance for usability studies and similar mixed methods implementation studies. No inferential statistical testing was performed, which is consistent with the exploratory descriptive aims of this work. Moreover, phase 1 and phase 2 differed simultaneously in tool version, participants, and clinical site, making between-phase inferential comparison inappropriate. Finally, the study was confined to NHS hospitals in the United Kingdom, which limits direct generalizability to health care systems with different funding models, workforce structures, technological infrastructures, or cultural norms. However, the 3 sites were selected to capture meaningful variation in organizational context and digital maturity, and the human-factors frameworks and usability

XSL•FO
RenderX

instruments applied are internationally validated. The core challenges addressed, variable cardiotocography interpretation, escalation delays, and risk communication, are not unique to the NHS and have been documented in maternity systems internationally. Future evaluation in non-NHS settings will be essential to establish broader applicability.

Conclusions

This study demonstrates that systematic co-development, grounded in human-factors principles, can produce clinician-facing decision-support tools that independently meet established usability benchmarks across NHS hospitals with different organizational contexts and digital maturity. The locked version (v4.0) showed high usability at both validation sites (SUS mean 80.7, SD 10.8 Birmingham vs SUS mean 90.8, SD 7.2 Buckinghamshire). Clinicians described the Fit4Labour tool as a supportive, like a "co-pilot" rather than a replacement for clinical judgment, and that it provided a shared language for risk communication with potential to strengthen confidence in escalation decisions. Other factors, such as organizational readiness, structured training, and adequate infrastructure support, will be important enablers of adoption. Whether usability in simulated settings translates to improved clinical outcomes in real-world practice, and whether these findings generalize beyond NHS settings, requires prospective evaluation.

Acknowledgments

We thank the Patient and Public Involvement panel from the Oxford Labor Monitoring Group for their advisory role throughout this study. We are grateful to all participating health care professionals, including midwives and doctors across the 3 participating hospitals within the National Health Service (NHS) in England, who provided invaluable feedback during system testing. Special thanks to the clinical managers, research midwives, and fetal monitoring midwives who facilitated this multisite study through their organizational support and coordination of testing sessions. We also acknowledge Dr. Sandra Morales Rios, researcher at the University of Oxford, for contribution to qualitative data coding and interrater reliability assessment.

Funding

This study was supported by a National Institute for Health and Care Research (NIHR) Invention for Innovation (i4i) Product Development Award (NIHR202117). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Authors' Contributions

MT and XL designed and conducted the usability testing sessions and focus groups, extracted quantitative data, and performed qualitative coding analysis. MT conducted the statistical analysis of usability metrics and completion time data. DH contributed to qualitative data coding and provided interrater reliability validation. KG and JT served as the tool developers and software engineers, implementing iterative design modifications based on user feedback and providing technical support throughout the co-development process to enhance the Fit4Labour system. AG conceived the original Fit4Labour tool concept and provided overall oversight and strategic guidance for the project. All authors made substantial contributions to manuscript preparation, critically reviewed the content, and provided final approval for publication.

Conflicts of Interest

All authors are or have been members of the Oxford Labor Monitoring Group at the University of Oxford, which conducts research on intrapartum fetal monitoring and decision-support tools. This study involved usability testing and evaluation of the Fit4Labour tool, which may be commercialized in the future through a planned company called Safer Birth. No commercial entity provided funding for this research, and the potential for commercialization did not influence the study design, data analysis, or interpretation. All results are reported transparently, regardless of their implications for future commercialization. JH is supported by a UKRI Future Leaders Fellowship (MR/Y033833/1).

Multimedia Appendix 1
Video demonstration of the Fit4Labour tool showing clinical workflow and user interface.
[DOCX File, 1071 KB - humanfactors_v13i1e87589_app1.docx]

References

1. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. JAMA 2018 Dec 4;320(21):2199-2200. [doi: 10.1001/jama.2018.17163] [Medline: 30398550]
2. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
3. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019 Apr 4;380(14):1347-1358. [doi: 10.1056/NEJMra1814259] [Medline: 30943338]

XSL•FO
RenderX

4. Lehne M, Sass J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. NPJ Digit Med 2019;2(1):79. [doi: 10.1038/s41746-019-0158-1] [Medline: 31453374]
5. Global Strategy on Digital Health 2020-2025: World Health Organization; 2021.
6. Ayres - de - Campos D, Spong CY, Chandraharan E, FIGO Intrapartum Fetal Monitoring Expert Consensus Panel. FIGO consensus guidelines on intrapartum fetal monitoring: cardiotocography. Intl J Gynecology & Obste 2015 Oct;131(1):13-24. [doi: 10.1016/j.ijgo.2015.06.020]
7. WHO Recommendations: Intrapartum Care for a Positive Childbirth Experience: World Health Organization; 2018.
8. Fetal monitoring in labour [NG229]. National Institute for Health and Care Excellence (NICE). 2022. URL: https://www.nice.org.uk/guidance/ng229 [accessed 2026-07-11]
9. Norman DA, editor. User Centered System Design: New Perspectives on Human-Computer Interaction: CRC Press; 1986. [doi: 10.1201/9780367807320]
10. Bates DW, Kuperman GJ, Wang S, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. J Am Med Inform Assoc 2003;10(6):523-530. [doi: 10.1197/jamia.M1370] [Medline: 12925543]
11. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 2005 Apr 2;330(7494):765. [doi: 10.1136/bmj.38398.500764.8F] [Medline: 15767266]
12. Carayon P, Schoofs Hundt A, Karsh BT, et al. Work system design for patient safety: the SEIPS model. Qual Saf Health Care 2006 Dec;15 Suppl 1(Suppl 1):i50-i58. [doi: 10.1136/qshc.2005.015842] [Medline: 17142610]
13. Blandford A, Furniss D, Vincent C. Patient safety and interactive medical devices: realigning work as imagined and work as done. Clin Risk 2014 Sep;20(5):107-110. [doi: 10.1177/1356262214556550] [Medline: 25866466]
14. Johnson CM, Johnson TR, Zhang J. A user-centered framework for redesigning health care interfaces. J Biomed Inform 2005 Feb;38(1):75-87. [doi: 10.1016/j.jbi.2004.11.005] [Medline: 15694887]
15. Cresswell KM, Bates DW, Sheikh A. Ten key considerations for the successful implementation and adoption of large-scale health information technology. J Am Med Inform Assoc 2013 Jun;20(e1):e9-e13. [doi: 10.1136/amiajnl-2013-001684] [Medline: 23599226]
16. van Gemert-Pijnen JL. Implementation of health technology: directions for research and practice. Front Digit Health 2022;4:1030194. [doi: 10.3389/fdgth.2022.1030194] [Medline: 36440459]
17. Sittig DF, Singh H. Defining health information technology-related errors: new developments since to err is human. Arch Intern Med 2011 Jul 25;171(14):1281-1284. [doi: 10.1001/archinternmed.2011.327] [Medline: 21788544]
18. Yusof MM, Takeda T, Shimai Y, Mihara N, Matsumura Y. Evaluating health information systems-related errors using the human, organization, process, technology-fit (HOPT-fit) framework. Health Informatics J 2024;30(2):14604582241252763. [doi: 10.1177/14604582241252763] [Medline: 38805345]
19. Cahill M, Cleary BJ, Cullinan S. The influence of electronic health record design on usability and medication safety: systematic review. BMC Health Serv Res 2025 Jan 6;25(1):31. [doi: 10.1186/s12913-024-12060-2] [Medline: 39762805]
20. Creswell L, Lindow BJ, Lindow SW, et al. A retrospective observational study of labour ward work Intensity: the challenge of maternity staffing. Eur J Obstetrics & Gynecol Reprod Biol 2023 Jul;286:90-94. [doi: 10.1016/j.ejogrb.2023.05.021]
21. Turner L, Ball J, Meredith P, Kitson-Reynolds E, Griffiths P. The association between midwifery staffing and reported harmful incidents: a cross-sectional analysis of routinely collected data. BMC Health Serv Res ;24(1). [doi: 10.1186/s12913-024-10812-8]
22. Maternity triage (Good Practice Paper No. 17). : RCOG; 2023 URL: https://www.rcog.org.uk/guidance/browse-all-guidance/good-practice-papers/maternity-triage-good-practice-paper-no-17/ [accessed 2026-07-11]
23. Park S, Marquard J. Nurse-centered co-design of an electronic health record nursing summary. Human Factors in Healthcare 2024 Jun;5:100065. [doi: 10.1016/j.hfh.2024.100065]
24. Intrapartum care [NG235]. National Institute for Health and Care Excellence (NICE). 2023. URL: https://www.nice.org.uk/guidance/ng235 [accessed 2026-07-11]
25. Lovers A, Daumer M, Frasch MG, et al. Advancements in fetal heart rate monitoring: a report on opportunities and strategic initiatives for better intrapartum care. BJOG 2025 Jun;132(7):853-866. [doi: 10.1111/1471-0528.18097] [Medline: 39971749]
26. Rei M, Tavares S, Pinto P, et al. Interobserver agreement in CTG interpretation using the 2015 FIGO guidelines for intrapartum fetal monitoring. Eur J Obst Gynecol Reprod Biol 2016 Oct;205:27-31. [doi: 10.1016/j.ejogrb.2016.08.017]
27. Neri S, Ramirez Zegarra R, Dininno M, Di Pasquo E, Tagliaferri S, Ghi T. Interobserver agreement of intrapartum cardiotocography interpretation by midwives using current FIGO and physiology-based guidelines. J Mater Fetal Neonatal Med 2024 Jan 2;37(1). [doi: 10.1080/14767058.2024.2425758]
28. Friedman AM, Campbell ML, Kline CR, Wiesner S, D'Alton ME, Shields LE. Implementing obstetric early warning systems. AJP Rep 2018 Apr;8(2):e79-e84. [doi: 10.1055/s-0038-1641569] [Medline: 29686937]
29. Ockenden D. Findings, conclusions and essential actions from the independent review of maternity services at the shrewsbury and telford hospital NHS trust. : Department of Health and Social Care; 2022 URL: https://www.gov.uk/government/publications/final-report-of-the-ockenden-review/ockenden-review-summary-of-findings-conclusions-and-essential-actions [accessed 2026-07-11]

XSL•FO
RenderX

30. Knight M, Bunch K, Tuffnell D, et al. Saving Lives, Improving Mothers' Care-Lessons Learned to Inform Maternity Care from the UK and Ireland Confidential Enquiries into Maternal Deaths and Morbidity 2017-19: MBRRACE-UK; 2021.
31. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). Soc Sci Med 1997 Mar;44(5):681-692. [doi: 10.1016/s0277-9536(96)00221-3] [Medline: 9032835]
32. INFANT Collaborative Group. Computerised interpretation of fetal heart rate during labour (INFANT): a randomised controlled trial. Lancet 2017 Apr 29;389(10080):1719-1729. [doi: 10.1016/S0140-6736(17)30568-8] [Medline: 28341515]
33. Ayres-de-Campos D, Sousa P, Costa A, Bernardes J. Omniview-SisPorto 3.5 - a central fetal monitoring station with online alerts based on computerized cardiotocogram+ST event analysis. J Perinat Med 2008;36(3):260-264. [doi: 10.1515/JPM.2008.030] [Medline: 18576938]
34. Bozhilova LV, Ugwumadu A, Chong HP, et al. A prognostic model for use at labour onset to estimate the risk of severe fetal compromise at birth: development and validation with over 145,000 electronic health records. Lancet Digital Health (forthcoming) 2026.
35. Hollan J, Hutchins E, Kirsh D. Distributed cognition: toward a new foundation for human-computer interaction research. ACM Transactions on Computer-Human 2000;Interaction7:174-196. [doi: 10.1145/353485.353487]
36. Holden RJ, Carayon P, Gurses AP, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 2013;56(11):1669-1686. [doi: 10.1080/00140139.2013.838643] [Medline: 24088063]
37. Nielsen J. Usability Engineering: Morgan Kaufmann; 1994.
38. Creswell JW, Plano Clark VL. Designing and Conducting Mixed Methods Research: Sage Publications; 2017.
39. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015 Sep;42(5):533-544. [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
40. Brooke J. SUS -- a quick and dirty usability scale. In: Usability Evaluation In Industry: CRC Press; 1996:189-194. [doi: 10.1201/9781498710411-35]
41. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability 2009:114-123 [FREE Full text]
42. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
43. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exer Health 2019 Aug 8;11(4):589-597. [doi: 10.1080/2159676X.2019.1628806]
44. Morlotti C, Cattaneo M, Paleari S, Manelli F, Locati F. The digitalization of emergency department triage: the perspectives of health professionals and patients. BMC Health Serv Res 1406;24(1). [doi: 10.1186/s12913-024-11862-8]
45. Ben M'Barek I, Jauvion G, Ceccaldi PF. Computerized cardiotocography analysis during labor – a state‐of‐the‐art review. Acta Obstet Gynecol Scand 2023 Feb;102(2):130-137. [doi: 10.1111/aogs.14498] [Medline: 36541016]
46. Tsipoura A, Giaxi P, Sarantaki A, Gourounti K. Conventional cardiotocography versus computerized CTG analysis and perinatal outcomes: a systematic review. Maedica (Bucur) 2023 Sep;18(3):483-489. [doi: 10.26574/maedica.2023.18.3.483] [Medline: 38023753]
47. O'Sullivan ME, Considine EC, O'Riordan M, Marnane WP, Rennie JM, Boylan GB. Challenges of developing robust AI for intrapartum fetal heart rate monitoring. Front Artif Intell 2021;4:765210. [doi: 10.3389/frai.2021.765210] [Medline: 34765970]
48. Balayla J, Shrem G. Use of artificial intelligence (AI) in the interpretation of intrapartum fetal heart rate (FHR) tracings: a systematic review and meta-analysis. Arch Gynecol Obstet 2019 Jul;300(1):7-14. [doi: 10.1007/s00404-019-05151-7] [Medline: 31053949]
49. Fernandes M, Vieira SM, Leite F, Palos C, Finkelstein S, Sousa JMC. Clinical decision support systems for triage in the emergency department using intelligent systems: a review. Artif Intell Med 2020 Jan;102:101762. [doi: 10.1016/j.artmed.2019.101762] [Medline: 31980099]
50. Investigation report: delays to intrapartum intervention once fetal compromise is suspected. : Healthcare Safety Investigation Branch; 2022 URL: https://www.hssib.org.uk/patient-safety-investigations/delays-to-intrapartum-intervention-once-fetal-compromise-is-suspected/investigation-report/ [accessed 2026-07-11]
51. Carayon P, Xie A, Kianfar S. Human factors and ergonomics as a patient safety practice. BMJ Qual Saf 2014 Mar;23(3):196-205. [doi: 10.1136/bmjqs-2013-001812] [Medline: 23813211]
52. Dlugatch R, Georgieva A, Kerasidou A. AI-driven decision support systems and epistemic reliance: a qualitative study on obstetricians' and midwives' perspectives on integrating AI-driven CTG into clinical decision making. BMC Med Ethics 2024;25(1). [doi: 10.1186/s12910-023-00990-1]
53. Aeberhard JL, Radan AP, Delgado-Gonzalo R, et al. Artificial intelligence and machine learning in cardiotocography: a scoping review. European Journal of Obstetrics & Gynecology and Reproductive Biology 2023 Feb;281:54-62. [doi: 10.1016/j.ejogrb.2022.12.008]
54. McComb S, Simpson V. The concept of shared mental models in healthcare collaboration. J Adv Nurs 2014 Jul;70(7):1479-1488. [doi: 10.1111/jan.12307] [Medline: 24237202]

XSL•FO
RenderX

55. Wu AW. Reaching common ground: the role of shared mental models in patient safety. Journal of Patient Safety and Risk Management 2018 Oct;23(5):183-184. [doi: 10.1177/2516043518805326]
56. Khairat S, Burke G, Archambault H, Schwartz T, Larson J, Ratwani RM. Perceived burden of EHRs on physicians at different stages of their career. Appl Clin Inform 2018 Apr;9(2):336-347. [doi: 10.1055/s-0038-1648222] [Medline: 29768634]
57. NHS resolution annual report and accounts 2024 to 2025. : NHS Resolution; 2025 URL: https://www.gov.uk/government/publications/nhs-resolution-annual-report-and-accounts-2024-to-2025 [accessed 2026-07-11]
58. NHS facing 'absolutely shocking' £27bn bill for maternity failings in England. The Guardian. 2025 Jul 20. URL: https://www.theguardian.com/society/2025/jul/20/nhs-facing-absolutely-shocking-27bn-bill-for-maternity-failings-in-england [accessed 2026-07-11]
59. Huo W, Yuan X, Li X, Luo W, Xie J, Shi B. Increasing acceptance of medical AI: the role of medical staff participation in AI development. Int J Med Inform 2023 Jul;175:105073. [doi: 10.1016/j.ijmedinf.2023.105073] [Medline: 37119693]
60. Longoni C, Bonezzi A, Morewedge CK. Resistance to medical artificial intelligence. J Consum Res 2019 Dec 1;46(4):629-650. [doi: 10.1093/jcr/ucz013]
61. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated consolidated framework for implementation research based on user feedback. Implementation Sci 2022;17(1). [doi: 10.1186/s13012-022-01245-0]
62. Green TL, Zapata JY, Brown HW, Hagiwara N. Rethinking bias to achieve maternal health equity: changing organizations, not just individuals. Obstet Gynecol 2021 May 1;137(5):935-940. [doi: 10.1097/AOG.0000000000004363] [Medline: 33831936]
63. Donovan T, Abell B, Fernando M, McPhail SM, Carter HE. Implementation costs of hospital-based computerised decision support systems: a systematic review. Implementation Sci 2023;18(1). [doi: 10.1186/s13012-023-01261-8]
64. Khan ZA, Kidholm K, Pedersen SA, et al. Developing a program costs checklist of digital health interventions: a scoping review and empirical case study. Pharmacoeconomics 2024 Jun;42(6):663-678. [doi: 10.1007/s40273-024-01366-y] [Medline: 38530596]
65. What's the real cost of EHR implementation for healthcare providers in 2026? Sprypt. 2024. URL: https://www.sprypt.com/blog/guide-to-the-cost-of-ehr-implementation-for-healthcare-providers [accessed 2026-07-11]
66. Clinical investigations of medical devices – guidance for manufacturers. UK Government. 2025. URL: https://www.gov.uk/guidance/notify-mhra-about-a-clinical-investigation-for-a-medical-device [accessed 2026-07-11]

Abbreviations

BHT: Buckinghamshire Healthcare NHS Trust
BWC: Birmingham Women's and Children's NHS Foundation Trust
NHS: National Health Service
NIHR: National Institute for Health and Care Research
OUH: Oxford University Hospitals NHS Foundation Trust
PPI: patient and public involvement
SEIPS: Systems Engineering Initiative for Patient Safety
SEQ: Single Ease Question
SUS: System Usability Scale
UK: United Kingdom

Edited by A Kushniruk; submitted 12.Nov.2025; peer-reviewed by C Morlotti; revised version received 25.May.2026; accepted 23.Jun.2026; published 23.Jul.2026.

Please cite as:
Tome M, Laurent X, Georgiev K, Tolladay J, Collins S, Hedgecott D, Bozhilova LV, Hirst JE, Impey L, Georgieva A
Multicenter Usability Evaluation and Co-Development of a Digital Decision-Support Tool for Labor Triage: Mixed Methods Study
JMIR Hum Factors 2026;13:e87589
URL: https://humanfactors.jmir.org/2026/1/e87589
doi:10.2196/87589

© Mariana Tome, Xavier Laurent, Kristiyan Georgiev, John Tolladay, Sarah Collins, Deborah Hedgecott, Lyuba V Bozhilova, Jane E Hirst, Lawrence Impey, Antoniya Georgieva. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 23.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information,

XSL•FO
RenderX

a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Barriers to Population Data Use in Overdose Fatality Reviews: Observational and Interview Study of Dashboard Deployment

Amey Salvi[1], MS; Cynthia Holladay[2], MPA, CCRP; Allyson L Dir[3], PhD; Bradley Ray[4], PhD; Matthew C Aalsma[5], PhD; Khairi Reda[6], PhD

[1]School of Informatics, Computing, and Engineering, Indiana University Indianapolis, Indianapolis, IN, United States

[2]Department of Pediatrics, Indiana University School of Medicine, Indianapolis, IN, United States

[3]Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, United States

[4]RTI International, Research Triangle Park, NC, United States

[5]Department of Pediatrics, Children's Health Services Research, Indiana University School of Medicine, Indianapolis, IN, United States

[6]Department of Computer Science, University of Illinois Chicago, 850 W Taylor St, CDRLC 5409, Chicago, IL, United States

Corresponding Author:
Khairi Reda, PhD
Department of Computer Science, University of Illinois Chicago, 850 W Taylor St, CDRLC 5409, Chicago, IL, United States

Abstract

Background: Overdose fatality reviews (OFRs) are an important public health tool for developing local overdose prevention strategies by reviewing individual overdose cases. While this approach offers a rich, contextual understanding of drug overdose factors in communities, it examines a small number of cases, providing limited insight into broader population-level risk patterns. To complement OFRs, we developed a real-time dashboard that visualizes trends about 5 key "touchpoints" (ie, interactions with medical and justice services preceding overdose). We then trained local OFR teams to use this dashboard to identify prevention opportunities.

Objective: This study examines the integration of population-level data into OFR practices, as well as the tensions that emerge between OFRs' traditional case-driven review processes and the statistical, population-level analysis typical in public health. We analyze how and when the dashboard was used in meetings, determine the extent to which the data informed recommendations, and identify barriers that prevented the broad uptake of this intervention.

Methods: We observed 26 OFR meetings across 11 counties in Indiana over 10 months, from November 2024 to September 2025, during which teams conducted case reviews and developed recommendations. We documented instances of dashboard use as well as "missed opportunities," in which relevant population-level data were available but not incorporated into the discussion. We also conducted semistructured interviews with OFR team members (n=7) to understand their perceptions of the dashboard, including its usefulness, usability, and adoption barriers.

Results: Despite its intended role, the dashboard was rarely integrated into OFR meetings; it was used only 10 times, compared to 114 missed opportunities in which relevant data could have informed discussions. Interviews revealed that this limited uptake was not solely due to usability barriers but reflected a deeper tension between 2 distinct analytic approaches. OFR teams prioritized narrative-driven case reviews that were grounded in empathy, local knowledge, and lived experience. This approach seemed at odds with the population-level visualizations shown in the dashboard, which required statistical abstraction and interpretation. Other barriers identified included limited time and resources, staff turnover, and varying levels of data fluency, which made it difficult for teams to confidently interpret the dashboard despite training.

Conclusions: The results highlight a tension between case-based and data-driven approaches to overdose prevention. These approaches are grounded in different workflows, values, and motivations, making it challenging for OFR teams to maintain their traditional, empathetic review practices while incorporating population-level trends. Our findings indicate the need for data tools that bridge these approaches, such as visualizations that connect aggregate patterns to individual cases. The results also underscore the need for additional support and training for teams, such as dedicated data specialists who interpret population-level trends and provide insights to augment team discussions and inform prevention strategies.

(JMIR Hum Factors 2026;13:e98864) doi:10.2196/98864

KEYWORDS

dashboard; visualization; public health; overdose; overdose prevention; recommendations; observation study; fatality review; prevention; interview; touchpoints; substance abuse

XSL•FO
RenderX

Introduction

The United States has seen a dramatic rise in drug overdoses over the past two decades, from 6.1 to 32.4 deaths per 100,000 people between 2001 and 2021 [1]. Although overdose fatalities have declined recently, the number remains high [2,3]. The response to this epidemic is further complicated by variations in risk factors across regions [4] and subpopulations [5]. Addressing this crisis thus requires data- and evidence-informed interventions tailored to the needs and challenges of local communities.

To encourage effective local overdose prevention, the Centers for Disease Control and Prevention has promoted local overdose fatality review (OFR) teams, first implemented in Maryland in early 2014 [6]. These teams unite public health, law enforcement, and emergency response stakeholders to review overdose cases. By analyzing decedents' histories and their interactions with services, teams can identify risks and gaps and recommend targeted prevention strategies [7]. Although some OFR teams have generated effective recommendations [6], this review model typically draws insights from only a handful of cases (2 - 3 per month or quarter) [8]. Such reviews provide rich qualitative insights into system and service gaps in local communities but may lead to recommendations that overlook broader overdose risk factors. Although some teams supplement their case reviews with quantitative data, these analyses are often based on small sample sizes, offering only a partial picture. This limitation presents an opportunity to enhance the OFR model by incorporating population-level data through a real-time dashboard, enabling more comprehensive, data-driven recommendations.

To address this opportunity, we partnered with the Indiana state government starting in September 2022 to develop and deploy a dashboard that provides OFR teams with real-time data about touchpoints—events and services individuals interact with *before* experiencing a fatal overdose [9,10]. The dashboard provides statistics about key touchpoints, including interactions with emergency and medical services and justice systems [11]. To introduce local OFR teams to the dashboard and provide guidance on using its data, we invited participating teams to a 2-day training workshop focused on population-level data use and evidence-based overdose prevention.

While dashboards have found success in public health settings [12-15] and are extensively used to monitor overdose trends [16-18], their use by OFR teams has received little attention. Even when team members receive training on integrating the dashboard, it represents a new, potentially fundamental change to the OFR process. Teams can also be highly diverse, including many stakeholders with no formal training or experience in data analysis. Thus, it is important to understand how data-driven approaches to overdose prevention are adopted within OFRs, including potential barriers to their uptake.

The goal of this study is to evaluate how well OFR teams can integrate a population-level touchpoints dashboard into their existing practices and whether they can inform their recommendations using touchpoint data alongside traditional case reviews. Specifically, we ask: (1) How frequently do teams consult the dashboard during OFR meetings? (2) To what extent do team members draw insights from dashboard data to make actionable recommendations for preventing overdoses? (3) What challenges do OFR teams face in effectively incorporating the dashboard into their existing workflows? To answer these questions, we observed OFR team meetings and analyzed their dashboard use, as well as any missed opportunities. We then followed up with interviews with team members to better understand their perceptions of the dashboard's utility and the barriers they faced in fully using it.

Methods

Study Setting

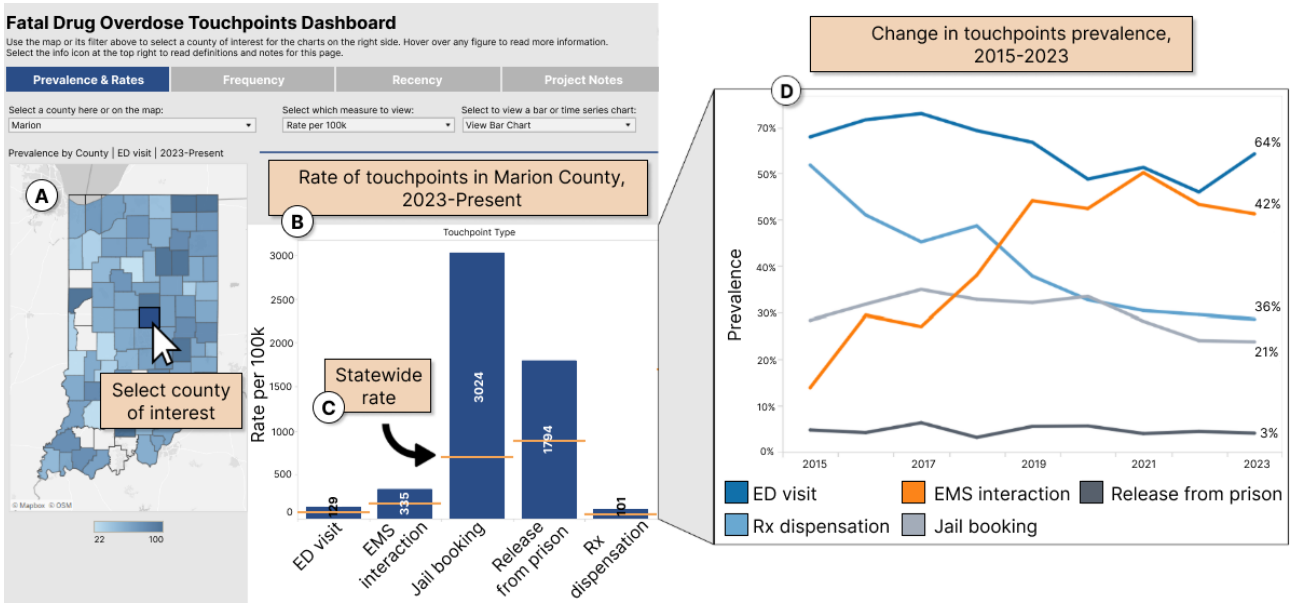
The study took place in the state of Indiana as part of the FORTRESS (Fatal Overdose Review Teams: Research to Enhance Surveillance Systems) project. Indiana has encouraged and expanded the use of local OFR teams in recent years, with teams organized at the county level. Most teams convene to review cases in a single county, but some teams oversee multiple jurisdictions, particularly in rural, low-population areas of the state. Organizational support and resources for OFR teams are provided by the Indiana Department of Health (IDOH).

The touchpoints dashboard used in this study was designed and developed through a user-centered design process that included formative focus groups with OFR experts to identify data needs, followed by an iterative design and feedback process. The expert group included experienced OFR practitioners as well as early developers of the OFR model from across the United States. The dashboard uses a record-linking process, integrating state mortality and administrative datasets to assemble a history of decedents' interactions with touchpoints and covers all overdose fatalities in the state between 2015 and the present. The focus on visualizing touchpoints that precede overdoses, as opposed to showing overdose trends [16-18], is meant to encourage a prevention-oriented mindset, consistent with the goal of OFRs. Five touchpoints are included in the dashboard: visits to the emergency department (ED), encounters with emergency medical services (EMS), dispensation of prescription drugs (Rx), jail bookings, and releases from prison. The dashboard supplies aggregate statistics displaying decedent data in the 12 months preceding their fatal overdose. These include the prevalence of these touchpoints (eg, percentage of decedents who had visited the ED in the year preceding their fatal overdose), and the rate of overdose among all touchpoint users (eg, the fraction of prison releasees who went on to experience a fatal overdose in the 12 months following their release). These two metrics help teams estimate both the potential reach of interventions in their community (prevalence) and the relative risk associated with specific touchpoints (rate). Additional measures provided by the dashboard include touchpoint frequency (average number of encounters) and the mean time between the last touchpoint interaction and death. Crucially, the dashboard, first deployed in 2023, breaks down all statistics by county, allowing teams to view data in their locality. Figure 1 illustrates two views from the dashboard. Full details can be found in Salvi et al [11]. The dashboard data are continuously

XSL•FO
RenderX

refreshed in (near) real time, approximately every 1 to 2 weeks for most touchpoints.

Figure 1. The FORTRESS (Fatal Overdose Review Teams: Research to Enhance Surveillance System) dashboard [19] shows overdose touchpoints in Indiana at the county level. The dashboard provides data on 5 touchpoints: interactions with the emergency health services (ED and EMS), legal systems (jails and prisons), and the dispensation of scheduled drugs. Visualizations illustrate the spatial distribution of touchpoints in the state (A), the rate of touchpoints in the selected county (B) versus the state average (C). The dashboard can also show how the prevalence and rates of touchpoints have changed over time, both within counties and across the whole state (D). ED: emergency department; EMS: emergency medical services; Rx: medical prescription.



Recruitment and Implementation Strategy

We recruited OFR teams for the study by reaching out to all active teams within the state. The list of active teams and their contacts was provided by IDOH. At the start of recruitment in February 2022, Indiana had nearly 2 dozen active OFR teams, with 5 more teams expected to begin meeting within the year. Of these teams, 19 responded to our recruitment and provided consent and letters of support to participate in the study. The recruited teams represented a variety of counties within the state, including both rural and urban communities. Recruited OFR teams were organized and introduced to the study using a stepped-wedge cluster design [20], in which 3 cohorts received training in successive waves (in October 2024, May 2025, and October 2025). Cohort 1 included 6 OFR teams, Cohort 2 included 8, and Cohort 3 included 5. Each training session lasted 2 days and covered dashboard functionality and data interpretation, evidence-based prevention practices, and strategies for developing actionable recommendations. Two members from each team, including the facilitator, were invited to participate in the training.

This study includes only observational and interview data from Cohorts 1 and 2, as Cohort 3 had not yet been observed at the time of writing. We sampled 11 OFR teams from the first 2 cohorts: 5 from Cohort 1 and 6 from Cohort 2. Teams were selected based on meeting schedules and the observer's availability. The observer attempted to maximize the number of meetings observed by attending as many meetings as feasible within the study period. All 11 teams observed had completed the 2-day training. Observations for Cohort 1 teams were conducted between November 2024 and September 2025, and for Cohort 2 teams between July 2025 and September 2025.

Following observations, we recruited 7 OFR team members for one-on-one interviews via email. We identified prospective interviewees through OFR team mailing lists and sent recruitment messages that included the purpose, duration, compensation, and a formal information sheet. We reached out to 9 members, 7 of whom ultimately responded and agreed to participate in the interview. Recruitment prioritized facilitators, given their depth of experience with the OFR process and their key role in conducting the meetings, although participation was not limited to them.

Data Collection

Meetings primarily covered overdose cases, with some also covering suicide cases (which we did not analyze), lasting 2 hours on average. A member of the research team (the first author) attended all meetings as an observer and documented the discussions by taking notes on dashboard use and missed opportunities. The same researcher then conducted semistructured follow-up interviews with 7 members from different OFR teams. Interviews represented diverse professional backgrounds and varying levels of experience with data. Interviewees included a registered nurse (RN), a police social worker (PSW), a nonprofit organization director, a behavioral health data analyst, a medication-assisted treatment (MAT) specialist, and OFR facilitators (see Table 1). The interviews were conducted over Zoom and lasted 20 to 30 minutes. Interview questions focused on participants' experiences with data, their perceptions of dashboard usability and utility, and the barriers they faced to its use (see Multimedia Appendix 1). This paper reports on insights from both observational studies and interviews [21,22]. This combination has been proven useful [23-25], particularly for evaluating dashboard use and utility for making data-driven decisions [26].

XSL•FO
RenderX

Table . Summary of interview participants' backgrounds and data experience.

Participant ID	Role/background	Extent of data experience	Served as OFR[a] facilitator?
P1	Nonprofit organization director	Research assistant with data analysis experience during undergraduate and graduate studies	Yes
P2	Project manager	No data tool training apart from FORTRESS[b]	No
P3	Behavioral health data analyst	No data tool training apart from FORTRESS	Yes
P4	OD2A[c] contractor	Previous experience with similar dashboards	Yes
P5	MAT[d] specialist	OFR training in addition to FORTRESS training	Yes
P6	Registered nurse	Graduate degree in statistics	Yes
P7	Police social worker	Extensive experience in data (collection and analysis)	Yes

[a]OFR: overdose fatality review.

[b]FORTRESS: Fatal Overdose Review Teams: Research to Enhance Surveillance System.

[c]OD2A: Overdose Data to Action.

[d]MAT: medication-assisted treatment.

Data Analysis

During the observations, we recorded instances in which the team's discussions and case reviews referenced 1 of the 5 touchpoints on the dashboard. Specifically, we noted: (1) dashboard use, where teams consulted the dashboard, for example, by looking up the prevalence or rates of touchpoints, (2) missed opportunities, where the discussion revolved around touchpoints for which there were data in the dashboard but where teams did not make an attempt to consult those statistics, and (3) nondashboard touchpoints (eg, housing-related) that the team discussed but for which there were no quantitative data. In addition to case discussions, we also analyzed the team's recommendations, focusing on dashboard use and missed opportunities. To analyze meeting data, we developed a codebook through an iterative coding process. We started with an initial set of codes that covered touchpoints referenced during the meetings, dashboard uses, and recommendations made. The codebook was refined iteratively and expanded to identify more specific types of dashboard use and missed opportunities. Meeting notes and observations were entered into an Excel sheet and assigned appropriate codes.

For the interview data, we used thematic analysis to identify recurring themes [27]. The interviews were video-recorded and transcribed. Interview questions suggested preliminary thematic categories related to dashboard usability, how it complements case reviews, challenges, and future improvements. We supplemented these with an open coding process to identify and tag additional topics that emerged from participants' responses. We then iteratively refined the codes, developed a codebook, and grouped the final codes into 3 broad themes relating to (1) the extent of dashboard use during and outside meetings, (2) advantages of the dashboard, and (3) challenges in using and interpreting the dashboard. See Checklist 1 for the completed Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist for the study design and analysis.

Ethical Considerations

This study was reviewed and approved by the Indiana University Institutional Review Board (15922). Participants in the interviews received an information sheet explaining the study goals and procedures before agreeing to take part. Participants received a US $40 gift card as compensation for the interviews. The dashboard displays only aggregate, population-level visualizations. No individual records were displayed in the dashboard to preserve anonymity.

Results

Dashboard Usage and Missed Opportunities

We first present our observations from OFR meetings on dashboard use and nonuse. Each OFR meeting typically lasted 2 hours, during which teams discussed 2 to 3 decedent cases. We observed a total of 48 overdose cases being discussed. The case reviews covered decedents' demographic backgrounds, obituaries, toxicology reports, medical histories, prescriptions for controlled substances and other medications, justice involvement, and next-of-kin interviews. Based on these, a timeline was constructed for each decedent that highlights the major events and touchpoints in their lives. The team discussed the case, identified service gaps, and recommended programmatic changes or policies to prevent future overdoses. Out of the 26 meetings we observed, the dashboard was used in only 2 meetings, where it was referenced either during case reviews or for making recommendations. Some meetings (10/26, 38.46%) also included a discussion of the county's overdose trends and counts, typically using historical data from prior years obtained from the coroner's office. However, none of this

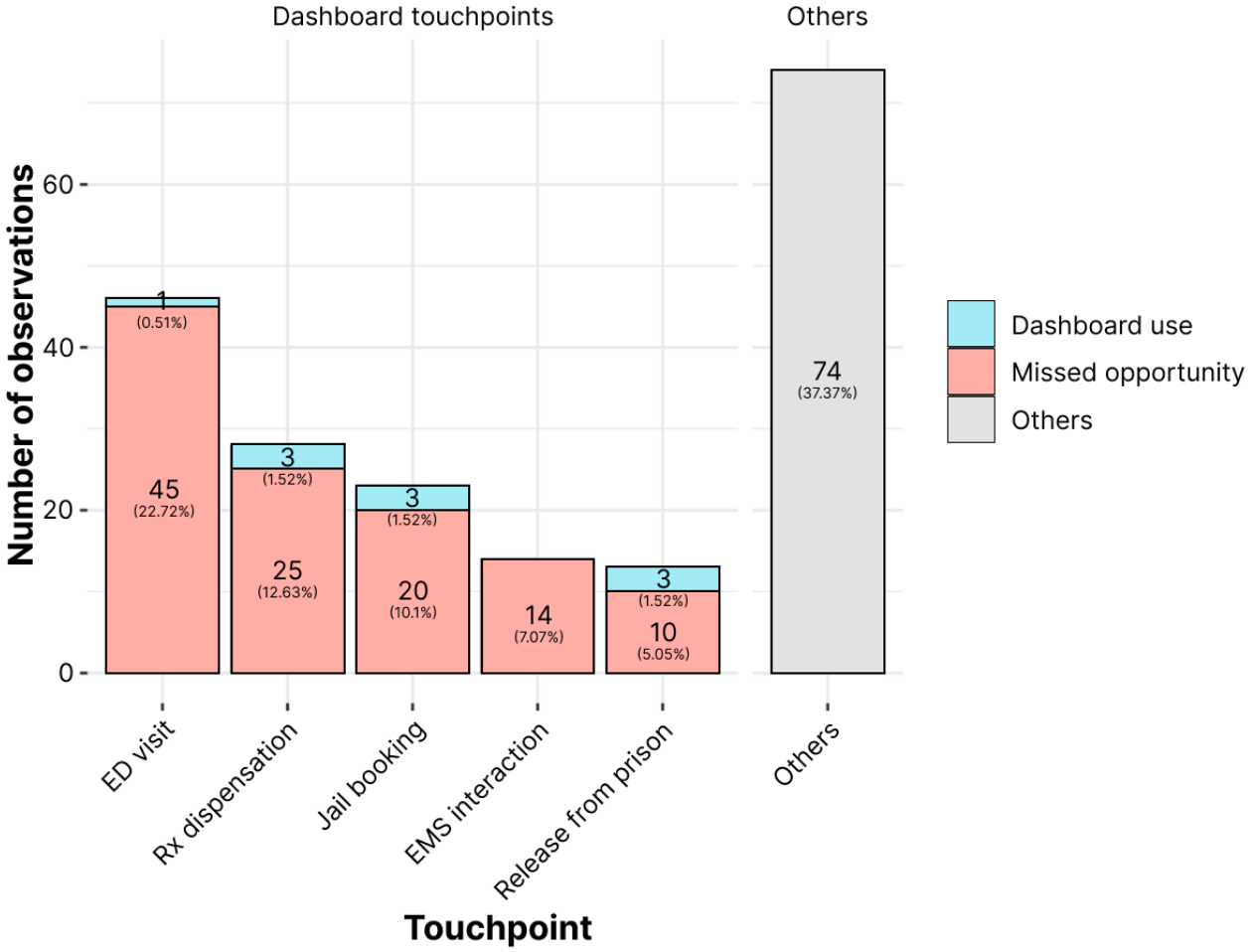
XSL•FO
RenderX

data overlapped with or provided detailed touchpoint metrics like those available in the FORTRESS dashboard.

To identify where the dashboard had been—or could have been—used, we coded all touchpoint occurrences in cases where the dashboard provides data. Recall that the dashboard provides data about 5 touchpoints. The most frequently occurring touchpoint was an ED visit, which occurred in 37.1% (46/124) of cases. This was followed by Rx dispensation (28/124, 22.58%), jail bookings (23/124, 18.55%), EMS interactions (14/124, 11.29%), and releases from prison (13/124, 10.48%).

In addition, we identified 74 other touchpoints (ie, interactions with systems such as mental health support, temporary housing, and system enhancements) for which the dashboard does not provide data. When OFR teams encountered 1 of the 5 touchpoints in a case, they used the dashboard 10 times across all observed meetings. By contrast, we identified 114 missed opportunities in which cases included touchpoints for which the dashboard provides data, but the information was not looked up or used during the discussion. Figure 2 illustrates the frequency of data use and missed opportunities across the 5 touchpoints.

Figure 2. Frequency of touchpoints from all observed meetings. ED: emergency department; EMS: emergency medical services; Rx: medical prescription.



Although limited, the use of the dashboard during meetings ranged from simple data readouts to contextual interpretation of the data and, in some cases, to generating actionable recommendations. At a basic level, there were 2 instances in which teams simply read data aloud when a case included 1 of the 5 touchpoints. For example, in one meeting, the facilitator distributed printouts showing the prevalence and rate of fatal overdoses associated with the ED touchpoint, noting that their county had a higher prevalence (73% of decedents had visited the ED in the year prior to death, compared to 64% statewide). This data, however, were not taken up further in the discussion. In a more interpretive use pattern, there were 4 instances in which teams engaged with the data by drawing on local knowledge to make sense of population-level trends in their community. For example, in one meeting, the team noted that individuals released from prison had a higher rate of fatal overdose in their county (1067 per 100,000 releases, compared to 704 statewide). This prompted a discussion of this elevated risk, with team members attributing it to factors such as lack of housing, limited support services, and barriers like suspended driver's licenses upon release. At the highest level of use, we found 4 instances in which teams translated observed patterns into recommendations. In one case, a team noted a higher prevalence of Rx dispensation for scheduled drugs among decedents in their county (56% vs 45% statewide), and identified a gap in what they described as a "transactional" system of care. They recommended educational initiatives and training for pharmacists to better support individuals in situations such as when medications are unavailable. In another instance, after observing a higher prevalence of jail bookings in their county (37% vs 25% statewide), a team recommended educating

XSL•FO
RenderX

incarcerated individuals about changes in their drug tolerance following release.

Overall, dashboard use was rare during OFR meetings: teams used it only 10 times. Most of our 114 observations were *missed opportunities*, where teams did not consult available population-level dashboard data during discussions or recommendations about touchpoints in decedents' cases. For example, in one meeting, the team reviewed a case where the decedent had a jail booking, 4 ED visits, and a prescription refill in the 12 months prior to their fatal overdose, yet the discussion centered on childhood factors and substance use disorder history. By contrast, the dashboard indicated that the county had a substantially elevated overdose rate following jail bookings (992 per 100,000 vs 469 statewide) and a slightly elevated Rx dispensation prevalence (40% vs 35% statewide). These data could have supported recommendations targeting more immediate intervention points, such as jails or pharmacies. In another meeting, despite a decedent having 18 ED visits and an EMS interaction in the prior year, the team focused on the decedent's failure to use peer recovery programs rather than on the high-risk touchpoints themselves. The dashboard, by contrast, showed above-average county rates for both ED visits and EMS interactions and indicated that 25% of decedents with these touchpoints fatally overdosed within 2 weeks. These 2 examples illustrate missed opportunities to use dashboard data to identify and prioritize touchpoints based on their risk (rate), reach (prevalence), and immediacy (recency).

Barriers to Integrating Population-Level Data in OFRs

Overview

We now discuss themes that emerged from interviews with OFR team members, including barriers to integration and future opportunities for data-driven prevention. Interviews shed light on several reasons behind the limited use of the dashboard during meetings. Reported barriers included tensions between case-based review and population-level analysis, varying levels of data literacy among team members, and limited time and resources.

Tension Between Case-Based Review and Population-Level Analysis

Participants highlighted a key distinction between OFRs, which focus on the detailed examination of a select few decedent cases, and the population-level data analysis provided by the dashboard. They contrasted OFRs' case-specific, narrative-driven analysis with the broader, aggregate approach typical in academic and public health settings. This contrast implies that OFR members apply mental models and analytical processes that differ from how the data are presented in the dashboard:

> *People we work with don't think or talk about data in the same way, so we don't think about what a quartile or an odds ratio is. At least not in the way it's supposed to. And that's some of the potential disconnect between something that is research–based and fact–based, and how we think of it in academia versus the random person on the street that will easily say, that's 5 times more likely. And that's not actually what they mean, but when it's used in that academic setting, it can get a little bit lost in translation.* [P1, nonprofit director]

Another aspect of the tension between approaches is the inherently empathetic nature of OFR case reviews, which is central to how teams understand cases. Team discussions focus on decedents' lived experiences, using that perspective to identify the challenges they faced and the factors that may have contributed to their overdose. In some instances, team members even recognized the individuals being discussed or knew their families. This empathy and attention to individuals can feel at odds with the statistical abstraction of population-level data:

> *What gives people motivation to work on top of their day-to-day workload on a goal is, their hearts, right? So, we just looked at the life and death of this individual we need to change X, Y, and Z, and these are the numbers to support it. I think that's where data usually falls short. Because we were looking at this qualitative section of this life; details that the numbers are missing... that's the SOFR [Suicide and Overdose Fatality Review] magic right there.* [P5, MAT specialist]

Limited Data Literacy and Tacit Knowledge

All participants highlighted limited data literacy as a key barrier, making it difficult for them to confidently draw appropriate inferences from the dashboard. They pointed out that the OFR teams comprise members from diverse backgrounds, many of whom do not possess data analysis skills:

> *Most of the room aren't data people. They want to focus on the case, what happened, and how can we make this better for our county. When people ask for data, they want to know, but in a way that they can digest it. So, if we get too into numbers and aggregate, we lose people.* [P5]

Participants noted that this limited experience raised a concern that the data could be misinterpreted, even with tooltip-style annotations and training videos included with the dashboard:

> *We had a disagreement of what this actually means, and how do we actually interpret it? Because it has lots of data, that's just really easy to get wrong.* [P1, nonprofit director]

Participants also noted how interpreting the dashboard can be particularly challenging in small counties with lower populations:

> *You have to really explain rate per 100,000... we're working in rural areas where we have, 8 [overdoses] total in a year, but then we're looking at numbers in the 50s and 60s, and so they're like, wait, what's happening? So, it isn't something that someone could just pick up and understand.* [P5, MAT specialist]

Participants also noted that many OFR members have a qualitative understanding of the key touchpoints in their community and do not feel the need to consult a dashboard to confirm their knowledge with raw numbers.

XSL•FO
RenderX

> *We talk commonly in our meetings, [that] person was released from jail and died of an overdose 24 hours later. That's common, we know that a lot of people do that, but we don't put a specific number on it. We all see those issues happening every day, and we don't need to say, 1 in 10 people getting released will experience this. Because of the work, we know the parameter without being able to put a specific number on it. There's definitely value to know the actual statistics so we can keep educating the community and justifying the cause. But it just takes more time.* [P7, PSW]

Perceived Data Stagnation and Insufficient Granularity

Although the dashboard updates in near real time, participants noted that its data rarely changed between meetings, as touchpoint prevalences and rates tend to remain stable over several months within most communities. This apparent stagnation appears to reduce the incentive to consult it regularly:

> *We don't need [the dashboard] all the time... ultimately, with this data, I doubt they're changing drastically month to month. Once we've identified what our patterns are, that's probably where it's staying. And so, I think we use it when we need it, but when we need it is not a weekly or even usually a monthly thing.* [P1, nonprofit director]

> *It's cumbersome to look at the dashboard at each meeting, because we're not going to see a lot of change in the numbers. So, I try to space [usage] out [so teams] not looking at the same graph, the same numbers, the same data each time.* [P5, MAT specialist]

One participant stated that they never used the dashboard because the data were not granular enough for their needs. While the dashboard breaks down touchpoints by county, consistent with how OFR teams are organized, larger urban counties tend to dedicate meetings that focus on risks for specific demographic groups or locales (eg, veterans, African American males over 50, overdoses at motels or hotels). This level of detail is not available in the dashboard:

> *It [the dashboard] kind of presents some general ideas and directions for us to go, but typically our meeting themes are more granular and more subsect from the dashboard.* [P4, Overdose Data to Action contractor]

Impact of Staff Turnover and Resource Constraints

Participants noted staff turnover and lack of resources as barriers to consistently incorporating the dashboard into meetings. They cited members leaving the team, facilitators having to manage multiple county meetings, and leadership and funding changes:

> *We had built out a "here's how we could use it, and here's what we should use it for". That got derailed when my staff person left. It needed somebody that wasn't me to carry [the dashboard implementation] out... With some of the staff turnover, we have used it less.* [P1, nonprofit director]

> *IDOH covers the OFR teams, and there has been so much structural change over the last few years as far as leadership, it is difficult [for facilitators] to figure out who's in charge, who do I go to [for directions].* [P6, RN]

Participants also highlighted the high workload and the lack of time and resources as additional barriers:

> *...people are pretty worn out. It was a hard time already doing all of this work, and now with all the funding cuts and the upheaval of nonprofits, they're not quite sure which direction to go, so FORTRESS is a great dashboard, but sometimes it's just one more thing [to learn and use].* [P6]

Importantly, one participant felt that dedicating a portion of the meeting to analyzing the dashboard data would, in effect, reduce the time available for case review, discussions, and/or recommendation making:

> *I think one is time constraint. Do we do it at the beginning, middle, or have specific meetings? Every meeting feels different, the cases flow differently. It also depends on everyone's vibe. We do two cases in two hours, and we are usually left with very little time to discuss recommendations.* [P5, MAT specialist]

Opportunities for Leveraging Population-Level Data in Prevention

Overview

Despite the obstacles to integrating the dashboard into OFR practices, participants highlighted several value points, both in and outside meetings, particularly in making local touchpoint data more accessible and serving as a resource for advocating policy changes.

Increasing Engagement Through Data

Participants noted that the dashboard can provide additional perspectives during meetings, particularly for touchpoints where representative stakeholders are absent:

> *The biggest thing is who's not at the table that can speak to that data intimately. We don't have EMS services at our table, so when we're going through recommendations, we don't have that viewpoint. So, when we're reflective on FORTRESS data, we can kind of talk to that.* [P6, RN]

Conversely, it was noted that the dashboard can help identify prevention and improvement opportunities, especially when stakeholders who can understand the data and influence change at specific touchpoints are present at the meeting:

> *If there are people in the room that have that ability to have impacts... having somebody in the room who can see the data, understand the data, and change something in my office. That's where I think it's [dashboard] most impactful.* [P1, nonprofit director]

Access to Local, Timely Data

Even without full integration of the dashboard into OFR practices, team members recognized a key benefit: centralized

XSL•FO
RenderX

access to comprehensive, timely, local-level data that may otherwise be difficult to obtain. Participants highlighted the value of the dashboard's high-level county data, particularly for smaller counties with limited access to hospital and prison information:

> *We're in a smaller community... data that the FORTRESS dashboard uses isn't usually data we have great access to like, Department of Corrections and EMS data. So it provides that guidance.* [P1, nonprofit director]

Participants also cited the dashboard's consistent updates as a major advantage (weekly for most data), even though they might not necessarily be needed during the meetings but for external needs such as pushing policies:

> *What makes the dashboard really nice is that it is current data. I think a lot of boots on the ground in the community, or at least entities like us need up-to-date data at the drop of a hat for an XYZ reason.* [P1, nonprofit director]

They noted the convenience of having all relevant data compiled in one place, explaining that this made it easier for teams to quickly spot trends and patterns that might have otherwise gone unnoticed:

> *[Looking at] the prevalence, we were kind of like, whoa, I had no idea it was like this. So, we always hear about it, but to actually see the data, it was really eye-opening.* [P3, behavioral analyst]

From Data to Action

Participants' comments suggested that the dashboard has the potential to achieve its intended purpose. Three participants noted that it can help address limitations of the case-based review process by encouraging teams to consider how prevention opportunities might apply more broadly, beyond individual cases. As one participant explained:

> *In our county, there are people coming out of prison, and we don't have a lot of information, so I think that that would be a good opportunity to bring [the dashboard] up. What can we do, maybe not for this specific case, but for this population in general, how can we reach them?* [P3]

Five participants noted that the dashboard can be a valuable tool for translating data into actionable insights, particularly for informing policy and programmatic decisions, although they recognized that realizing these opportunities often depends on external constraints, such as funding:

> *The dashboard could really be useful if we get a state-level OFR partner group, where we were going to try to present things to policymakers, [eg,] being able to show we need to give a lot more support to our justice-involved individuals. That is a really clear takeaway from this dashboard... nice to have that kind of data there and visualized.* [P4]

> *When I was looking through the dashboard, I could think, we have 10 different initiatives that would be good at addressing what the data was saying. I think the question is always funding, right? So, we can look at the dashboard and say, I can envision XYZ, but we need the dollars to do it.* [P7]

Other Dashboard Uses

Participants reported ways to use the dashboard that we did not observe directly, such as prioritizing recommendations:

> *We had a recommendations meeting, that's how we decide what recommendations to move forward on. We decided to focus on jail discharges, because that was huge... [and] this is a recommendation we can reach the most people with. I think it has been really impactful, because we had a lot of recommendations, and we were kind of like, where do we start?* [P3, behavioral analyst]

Participants also reported use outside of OFR meetings to provide the data to support for certain recommendations and initiatives:

> *I had pushed for having Narcan and test strips in our lobby and there was a lot of pushback against it. Through FORTRESS, [for] the first time I've been able to point to a statistic and say, 'this is why it's so important that we have this in the lobby.'* [P7, PSW]

Discussion

Principal Findings

OFR teams are an increasingly important public health intervention for preventing overdoses through locally tailored strategies. Traditionally, these teams conduct case reviews for a small number of decedents to identify system gaps, risk factors, and corresponding prevention opportunities within their communities. The FORTRESS dashboard was designed to augment this case-based review process by incorporating population-level data on touchpoints that precede overdoses. This study sought to examine the extent to which teams integrated the dashboard into their workflows, identify barriers to its use, and highlight opportunities for more effective integration.

Over a 10-month period, we observed 11 OFR teams in 26 meetings and found limited integration of the dashboard into those meetings. Specifically, teams consulted the dashboard in only 8% (2/26) of instances where it could have informed a case discussion or recommendation. In those few instances, dashboard use ranged from simple data lookup to informing concrete recommendations, suggesting that it can support both basic and more advanced analytical needs. However, overall uptake was limited, as teams frequently discussed touchpoints for which the dashboard provided relevant data but did not appear to use it. In effect, we observed many more missed opportunities to incorporate population-level insights into teams' discussions than actual uses. Interviews with team members uncovered several barriers to this limited uptake, as well as opportunities to improve integration.

Barriers and Opportunities

Interviews suggest that OFR teams have a strong preference for traditional case-based discussion. The OFR review process is

XSL•FO
RenderX

inherently qualitative, relying on narratives to reconstruct the decedents' timelines and to understand key events, gaps, and factors that may have contributed to their overdoses. Teams also routinely draw on next-of-kin interviews, often attempting to interpret cases from the decedent's perspective. This approach is deeply contextual, empathetic, and centered on individual lived experiences. Yet, this mode of sense-making can seem at odds with population-level analysis, in which individual experiences are abstracted into aggregate statistical measures. As a result, the case-based model for OFRs, while effective for generating contextual insights, may inherently discourage engagement with quantitative data, keeping teams focused on individual cases rather than broader risk patterns.

The current dashboard design, which separates population-level data from case-specific details, may have limited its use. Pisani et al [28] showed that data decoupling can increase OFR teams' cognitive load and recommended integrating population-level data with case-level modeling tools within a single dashboard. Our findings support this approach but further indicate that visualizations should directly merge aggregate data with anonymized case details. For instance, a timeline view could show a decedent's touchpoints while embedding population-level metrics within the same view. In such a design, individual touchpoints could be shown as points along a timeline, with their size reflecting prevalence among county residents (eg, larger points indicating more common occurrences) and color representing overdose risk (eg, deeper red indicating higher rates). By linking individual trajectories with broader population-level data in a single view, such visualizations could help teams better identify population-level patterns during discussions. This approach may also align better with existing OFR workflows, which focus on case narratives, making adoption more likely, as tools that match established practices are readily accepted [29].

Another consistently cited barrier to adoption is the limited data and visualization literacy among team members. OFR teams include stakeholders from diverse fields, such as social work, coroner services, and law enforcement, with varying levels of data expertise. Thus, difficulty in interpreting population-level data remains a key factor limiting dashboard use. These challenges are consistent with evaluations of public health dashboards, which indicate that users often struggle to interpret visualizations without support [13,30-32]. However, previous studies show that visualization skills can be effectively learned through video and interactive tutorials [33] and that users with low data literacy benefit from storytelling or narrative elements included in the dashboard [34]. In response, we developed additional training materials in the form of short videos that demonstrate how to interpret the dashboard and apply its data to make actionable recommendations.

Barriers appear to extend beyond data literacy. Some participants described the dashboard as an added burden alongside existing case-review responsibilities, leading to reluctance to engage with it. Others felt they already had sufficient knowledge of local risk factors and did not see the need to consult the dashboard, further limiting its uptake. One potential solution is to designate a dedicated data analyst within OFR teams to explore population-level data, interpret trends, and communicate key findings to the group. This approach allows other team members to focus on case discussion and recommendation-making rather than data interpretation [35,36]. Some teams that we observed have already adopted a similar model by appointing a "data specialist" to review external data sources. Traditionally, this role has involved working with tools such as REDCap databases, which provide limited quantitative information on reviewed cases. Expanding this role to include dashboard use, supported by additional training, could enable teams to better leverage it. This approach may help address both data literacy challenges and time constraints, possibly facilitating more effective integration.

A notable finding from our observations was the frequent missed opportunities to leverage the dashboard during meetings. Often, teams encountered touchpoints for which relevant data were available but did not consult the dashboard to inform their discussions or recommendations. This pattern indicates a need for more immediate support in identifying when and how to apply the dashboard. To address this, we began experimenting with a real-time feedback approach in which research team members observe OFR meetings and provide prompt, specific feedback afterward—highlighting missed opportunities and proposing ways to integrate dashboard data into the discussion. This approach builds on prior work showing that real-time feedback can help users overcome visualization literacy barriers [37] and that timely audit-and-feedback encourages change [38,39]. By clearly linking case discussions to relevant data features, this strategy may enhance teams' use of population-level data.

While our observations showed limited uptake by OFR teams, we believe that providing near real-time data on overdose touchpoints can still be valuable to other localities or states. However, to increase the chance of use during meetings, the dashboard can benefit from several modifications, including innovative visualization designs that combine population-level data with case data, such that teams can more readily integrate the two into their review process. Another potential opportunity lies in the use of large language models to provide a more user-friendly chat interface for accessing and interpreting the data provided by the dashboard [40-42]. In particular, large language models can also provide a bridge between case narratives and aggregate data, helping team members leverage both analysis modalities to make actionable recommendations [43,44]. Additionally, improving training for OFR team members on data use and incentivizing evidence- and data-driven recommendations could encourage further adoption.

Limitations

This study has several limitations that need to be considered. While we observed dashboard use during OFR meetings, we do not know how often or in what ways the dashboard is used outside those meetings. Therefore, we cannot fully determine how much it indirectly influences decision- and recommendation-making. Interview comments indeed suggest that the dashboard may have been used in other ways. Moreover, not all OFR members at the meeting were trained in dashboard use. Interviews suggest that some of the members who attended our 2-day training workshop may have left the team before our

XSL•FO
RenderX

observations. Staff turnover could thus limit dashboard use during meetings. Additionally, only a subset of OFR teams in Indiana was observed, so the discussions and practices observed may not represent all OFR teams. Similarly, the interview findings were based on 7 participants, whose experiences and perspectives may differ from those of other OFR members not included in the study. The limited sample may have biased the themes identified in the study. Still, this study highlights a need for ongoing support and training to help OFR teams generate data-informed prevention strategies.

Conclusions

This study examined whether OFR teams can effectively integrate a population-level touchpoints dashboard into their meetings to support data-driven recommendations. Our observations suggest that, while the dashboard can help teams generate useful insights and inform recommendations, its use during meetings was infrequent. Interviews with team members point to a tension between the case-based review model of OFRs and the population-level, data-driven approach afforded by the dashboard, with members struggling to incorporate the latter into their existing practices. Team members also cited limited data literacy, as well as time and resource constraints, as barriers to its use. These findings suggest a need to design data tools that better bridge the gap between case reviews and quantitative, epidemiologic analyses—for example, through visualizations that integrate individual case information with population-level metrics. Providing additional support and training for dedicated "data specialists" within teams may also help them leverage data and generate evidence-based recommendations, potentially leading to stronger prevention strategies at the local level.

Acknowledgments

The authors thank the Indiana Management Performance Hub, the Indiana Department of Health, the Department of Homeland Security, the Indiana State Police, the Professional Licensing Agency, the Department of Correction, and the Governor's Drug Prevention, Treatment, and Enforcement Office for providing data for use in the dashboard and for their feedback. We are also grateful to the overdose fatality review (OFR) team members whose participation made this study possible. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health or participating organizations. A generative artificial intelligence tool, Grammarly, was used for proofreading and minor language refinement. All content was reviewed and verified by the authors, who take full responsibility for the accuracy and integrity of the manuscript.

Funding

This research was funded by the National Institute on Drug Abuse/National Institutes of Health Helping to End Addiction Long-term Initiative through the Data 2 Action program under award numbers R61DA057660 and R33DA057660 (2022-2027) in the amount of US $4,901,551.

Data Availability

The video recordings and interview data generated and analyzed during this study are not publicly available to protect participants' confidentiality. The deidentified transcripts of the interviews can be provided from the corresponding author on reasonable request. The notes and analyses of observations of OFR meetings conducted are not publicly available due to the confidentiality agreements signed with all OFR teams. The data that support the findings of this study are available in the National Addiction & Health Data Archive Program [45].

Authors' Contributions

AS conducted the observations and interviews, analyzed the data, and wrote the manuscript. CH analyzed the data and edited the manuscript. ALD designed the study and edited the manuscript. MCA, BR, and KR conceptualized the research, designed the study, and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide for semistructured interviews.
[DOCX File, 19 KB - humanfactors_v13i1e98864_app1.docx]

Checklist 1
COREQ checklist.
[PDF File, 494 KB - humanfactors_v13i1e98864_app2.pdf]

References

XSL•FO
RenderX

1. Spencer MR, Warner M, Cisewski JA, et al. Estimates of drug overdose deaths involving fentanyl, methamphetamine, cocaine, heroin, and oxycodone: United States, 2021. : Centers for Disease Control and Prevention (CDC), National Center for Health Statistics; 2023 URL: https://www.cdc.gov/nchs/data/vsrr/vsrr027.pdf [accessed 2026-07-10]
2. CDC reports nearly 24% decline in U.S. drug overdose deaths. US Centers for Disease Control and Prevention (CDC). 2025. URL: https://www.cdc.gov/media/releases/2025/2025-cdc-reports-decline-in-us-drug-overdose-deaths.html [accessed 2026-07-10]
3. Understanding the opioid overdose epidemic. US Centers for Disease Control and Prevention (CDC). 2025. URL: https://www.cdc.gov/overdose-prevention/about/understanding-the-opioid-overdose-epidemic.html [accessed 2026-07-10]
4. Forati AM, Ghose R, Mantsch JR. Examining opioid overdose deaths across communities defined by racial composition: a multiscale geographically weighted regression approach. J Urban Health 2021 Aug;98(4):551-562. [doi: 10.1007/s11524-021-00554-x] [Medline: 34231120]
5. Choe K, Zinn E, Lu K, Hoang D, Yang LH. Impact of COVID-19 pandemic on chronic pain and opioid use in marginalized populations: a scoping review. Front Public Health 2023;11:1046683. [doi: 10.3389/fpubh.2023.1046683] [Medline: 37139395]
6. Rebbert-Franklin K, Haas E, Singal P, et al. Development of Maryland local overdose fatality review teams: a localized, interdisciplinary approach to combat the growing problem of drug overdose deaths. Health Promot Pract 2016 Jul;17(4):596-600. [doi: 10.1177/1524839916632549] [Medline: 27091609]
7. Costello E, Cruz K. Evaluation profile for overdose fatality reviews. : Centers for Disease Control and Prevention (CDC); 2022 URL: https://www.cdc.gov/overdose-resources/pdf/OD2A_EvalProfile_OverdoseFatalityReviews_508.pdf [accessed 2026-07-10]
8. Haas E, Truong C, Bartolomei-Hill L, Baier M, Bazron B, Rebbert-Franklin K. Local overdose fatality review team recommendations for overdose death prevention. Health Promot Pract 2019 Jul;20(4):553-564. [doi: 10.1177/1524839918797617] [Medline: 30205737]
9. Larochelle MR, Bernstein R, Bernson D, et al. Touchpoints - opportunities to predict and prevent opioid overdose: a cohort study. Drug Alcohol Depend 2019 Nov 1;204:107537. [doi: 10.1016/j.drugalcdep.2019.06.039] [Medline: 31521956]
10. Ray B, Christian K, Bailey T, et al. Antecedents of fatal overdose in an adult cohort identified through administrative record linkage in Indiana, 2015-2022. Drug Alcohol Depend 2023 Jun 1;247:109891. [doi: 10.1016/j.drugalcdep.2023.109891] [Medline: 37120921]
11. Salvi A, Gillenwater LA, Cockrum BP, et al. Development of a real-time dashboard for overdose touchpoints: user-centered design approach. JMIR Hum Factors 2024 Jun 11;11:e57239. [doi: 10.2196/57239] [Medline: 38861717]
12. Bach B, Freeman E, Abdul-Rahman A, et al. Dashboard design patterns. IEEE Trans Vis Comput Graph 2023 Jan;29(1):342-352. [doi: 10.1109/TVCG.2022.3209448] [Medline: 36155447]
13. Fareed N, Olvera RG, Wang Y, et al. Lessons learned from developing dashboards to support decision-making for community opioid response by community stakeholders: mixed methods and multisite study. JMIR Hum Factors 2024 Sep 9;11:e51525. [doi: 10.2196/51525] [Medline: 39250216]
14. Hoots BE. Opioid overdose surveillance: improving data to inform action. Public Health Rep 2021;136(1_suppl):5S-8S. [doi: 10.1177/00333549211020275] [Medline: 34726970]
15. Yedinak J, Krieger MS, Joseph R, et al. Public health dashboards in overdose prevention: the Rhode Island approach to public health data literacy, partnerships, and action. J Med Internet Res 2024 Feb 12;26:e51671. [doi: 10.2196/51671] [Medline: 38345849]
16. Anderson J, Demeter N, Pasquires M, Wirtz S. Using the CA opioid overdose surveillance dashboard to track opioid overdose deaths. Online J Public Health Inform 2019;11(1):e62570. [doi: 10.5210/ojphi.v11i1.9938]
17. Safranek CW, Feitzinger L, Joyner AKC, et al. Visualizing opioid-use variation in a pediatric perioperative dashboard. Appl Clin Inform 2022 Mar;13(2):370-379. [doi: 10.1055/s-0042-1744387] [Medline: 35322398]
18. Wu E, Villani J, Davis A, et al. Community dashboards to support data-informed decision-making in the HEALing communities study. Drug Alcohol Depend 2020 Dec 1;217:108331. [doi: 10.1016/j.drugalcdep.2020.108331] [Medline: 33070058]
19. FORTRESS dashboard. URL: https://indianatouchpoints.org [accessed 2026-07-13]
20. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemp Clin Trials 2007 Feb;28(2):182-191. [doi: 10.1016/j.cct.2006.05.007] [Medline: 16829207]
21. Adeoye‐Olatunde OA, Olenik NL. Research and scholarly methods: semi‐structured interviews. J Am Coll Clin Pharm 2021 Oct;4(10):1358-1367. [doi: 10.1002/jac5.1441]
22. Kallio H, Pietilä AM, Johnson M, Kangasniemi M. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. J Adv Nurs 2016 Dec;72(12):2954-2965. [doi: 10.1111/jan.13031] [Medline: 27221824]
23. Harvey SA. Observe before you leap: why observation provides critical insights for formative research and intervention design that you'll never get from focus groups, interviews, or KAP surveys. Glob Health Sci Pract 2018 Jun 29;6(2):299-316. [doi: 10.9745/GHSP-D-17-00328] [Medline: 29794000]

XSL•FO
RenderX

24. Horton J, Macve R, Struyven G. Qualitative research: experiences in using semi-structured interviews. In: The Real Life Guide to Accounting Research: Elsevier; 2004:339-357. [doi: 10.1016/B978-008043972-3/50022-0]
25. Ligthelm RJ, Borzì V, Gumprecht J, Kawamori R, Wenying Y, Valensi P. Importance of observational studies in clinical practice. Clin Ther 2007 Jun;29(6 Pt 1):1284-1292. [doi: 10.1016/j.clinthera.2007.07.004] [Medline: 18036390]
26. Dolan JG, Veazie PJ, Russ AJ. Development and initial evaluation of a treatment decision dashboard. BMC Med Inform Decis Mak 2013 Apr 21;13(1):51. [doi: 10.1186/1472-6947-13-51] [Medline: 23601912]
27. Joffe H. Thematic analysis. In: Harper D, Thompson AR, editors. Qualitative Research Methods in Mental Health and Psychotherapy: A Guide for Students and Practitioners: John Wiley & Sons, Ltd; 2011:209-223. [doi: 10.1002/9781119973249]
28. Pisani M, Oguss MK, Dickson-Gomez J, et al. Designing a substance misuse data dashboard for overdose fatality review teams: user-centered design approach. JMIR Hum Factors 2026 Jan 15;13(1):e79407. [doi: 10.2196/79407] [Medline: 41538509]
29. Tory M, Bartram L, Fiore-Gartland B, Crisan A. Finding their data voice: practices and challenges of dashboard users. IEEE Comput Grap Appl 2023;43(1):22-36. [doi: 10.1109/MCG.2021.3136545]
30. Ansari B. Evaluating the usability of public health data dashboards as information sources for professionals and the public: findings from a case study with domain experts. Health Info Libr J 2025 Jun;42(2):148-161. [doi: 10.1111/hir.12532] [Medline: 38717119]
31. Ansari B, Martin EG. Integrating human-centered design in public health data dashboards: lessons from the development of a data dashboard of sexually transmitted infections in New York State. J Am Med Inform Assoc 2024 Jan 18;31(2):298-305. [doi: 10.1093/jamia/ocad102] [Medline: 37330670]
32. Concannon D, Herbst K, Manley E. Developing a data dashboard framework for population health surveillance: widening access to clinical trial findings. JMIR Form Res 2019 Apr 4;3(2):e11342. [doi: 10.2196/11342] [Medline: 30946016]
33. Kwon BC, Lee B. A comparative evaluation on online learning approaches using parallel coordinate visualization. Presented at: CHI '16: the 2016 CHI Conference on Human Factors in Computing Systems; May 7-12, 2016. [doi: 10.1145/2858036.2858101]
34. Pozdniakov S, Martinez-Maldonado R, Tsai YS, Echeverria V, Srivastava N, Gasevic D. How do teachers use dashboards enhanced with data storytelling elements according to their data visualisation literacy skills? Presented at: LAK2023: LAK23: 13th International Learning Analytics and Knowledge Conference; Mar 13-17, 2023. [doi: 10.1145/3576050.3576063]
35. Fischer L, Ellingson K, McCormick K, Sinkowitz-Cochran R. The role of the public health analyst in the delivery of technical assistance to state health departments for healthcare-associated infection prevention. Med Care 2014 Feb;52(2 Suppl 1):S54-S59. [doi: 10.1097/MLR.0000000000000055] [Medline: 24430267]
36. Windt J, Taylor D, Little D, Sporer BC. Making everyone's job easier. How do data scientists fit as a critical member of integrated support teams? Br J Sports Med 2021 Jan;55(2):73-75. [doi: 10.1136/bjsports-2020-102938] [Medline: 32816794]
37. Boy J. Engaging the people to look beyond the surface of online information visualizations of open data [Dissertation]. : Telecom ParisTech; 2015 URL: https://inria.hal.science/tel-01225585 [accessed 2026-07-10]
38. Hysong SJ, Teal CR, Khan MJ, Haidet P. Improving quality of care through improved audit and feedback. Implement Sci 2012 May 18;7(1):45. [doi: 10.1186/1748-5908-7-45] [Medline: 22607640]
39. Patel S, Rajkomar A, Harrison JD, et al. Next-generation audit and feedback for inpatient quality improvement using electronic health record data: a cluster randomised controlled trial. BMJ Qual Saf 2018 Sep;27(9):691-699. [doi: 10.1136/bmjqs-2017-007393] [Medline: 29507124]
40. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence-based chatbots for promoting health behavioral changes: systematic review. J Med Internet Res 2023 Feb 24;25:e40789. [doi: 10.2196/40789] [Medline: 36826990]
41. Branda F, Stella M, Ceccarelli C, et al. The role of AI-based chatbots in public health emergencies: a narrative review. Future Internet 2025;17(4):145. [doi: 10.3390/fi17040145]
42. Wilson L, Marasoiu M. The development and use of chatbots in public health: scoping review. JMIR Hum Factors 2022 Oct 5;9(4):e35882. [doi: 10.2196/35882] [Medline: 36197708]
43. Wang X, Wu Z, Huang W, et al. VIS+AI: integrating visualization with artificial intelligence for efficient data analysis. Front Comput Sci 2023 Dec;17(6):176709. [doi: 10.1007/s11704-023-2691-y]
44. Wu A, Wang Y, Shu X, et al. AI4VIS: survey on artificial intelligence approaches for data visualization. IEEE Trans Visual Comput Graphics 2021;28(12):5049-5070. [doi: 10.1109/TVCG.2021.3099002]
45. Deposit data. National Addiction & Health Data Archive Program (NAHDAP). URL: https://www.icpsr.umich.edu/sites/nahdap/deposit-data [accessed 2026-07-10]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
ED: emergency department
EMS: emergency medical services

XSL•FO
RenderX

FORTRESS: Fatal Overdose Review Teams: Research to Enhance Surveillance System
IDOH: Indiana Department of Health
MAT: medication-assisted treatment
OFR: overdose fatality review
PSW: police social worker
RN: registered nurse
Rx: medical prescription

Edited by A Kushniruk, S Law; submitted 22.Apr.2026; peer-reviewed by C Noel; revised version received 20.Jun.2026; accepted 23.Jun.2026; published 27.Jul.2026.

Please cite as:
Salvi A, Holladay C, Dir AL, Ray B, Aalsma MC, Reda K
Barriers to Population Data Use in Overdose Fatality Reviews: Observational and Interview Study of Dashboard Deployment
JMIR Hum Factors 2026;13:e98864
URL: https://humanfactors.jmir.org/2026/1/e98864
doi:10.2196/98864

© Amey Salvi, Cynthia Holladay, Allyson L Dir, Bradley Ray, Matthew C Aalsma, Khairi Reda. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 27.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Sentence-Level Provenance for AI Medical Record Summarization in a Click-to-Inspect Interface: Formative Usability Evaluation

Andrew Parambath[1*], MD, MBA, MEd; Giordana Pulpo[2*], MPS; Vince Hartman[2], MS

[1]Stanford Medicine, 900 Welch Road, Suite 350, Palo Alto, CA, United States

[2]Abstractive Health, New York, NY, United States

[*]these authors contributed equally

Corresponding Author:
Andrew Parambath, MD, MBA, MEd
Stanford Medicine, 900 Welch Road, Suite 350, Palo Alto, CA, United States

Abstract

Background: Large language models can generate fluent summaries of longitudinal medical records, but in high-stakes clinical settings, verification burden remains a barrier to trust. Existing provenance mechanisms such as document-level citations and section references often require manual search within long, fragmented notes, limiting their usefulness during time-constrained workflows for clinicians.

Objective: This study aimed to design and evaluate a sentence-level provenance interface ("click to inspect") that enables rapid verification of AI-generated longitudinal medical record summaries at the level of individual statements.

Methods: Between November 2023 and January 2024, we conducted a formative usability study using remotely moderated usability sessions via Zoom to evaluate a web-based sentence-level provenance interface for AI-generated longitudinal medical record summaries. A convenience sample of clinicians was recruited through email outreach to academic and professional networks across the United States. Formative usability testing was conducted with 46 clinician interactions using synthetic longitudinal patient charts. Participants included medical students, residents, and attending physicians across multiple specialties, including internal medicine, dermatology, radiology, plastic surgery, anesthesiology, interventional radiology, obstetrics and gynecology, and family medicine. Usability was assessed using the System Usability Scale and net promoter score, alongside qualitative feedback.

Results: Clinicians reported high usability (mean System Usability Scale score 86.25, SD 7.77; 95% CI 83.96 - 88.54 from 46 participants) and a positive overall experience (net promoter score of 35; 22/46, 47.8% promoters; 18/46, 39.1% passives; and 6/46, 13% detractors). Participants described rapid access to supporting evidence as critical for trust calibration during first-pass chart review. Qualitative feedback identified friction in traditional citation-based interfaces and supported sentence-level inspectability as a low-friction verification mechanism.

Conclusions: Sentence-level provenance transforms AI-generated summaries from static narratives into interactive verification tools. An approach that enables rapid, selective inspection of individual claims during longitudinal chart review may reduce verification burden and support calibrated reliance in high-risk clinical contexts.

(JMIR Hum Factors 2026;13:e95644) doi:10.2196/95644

KEYWORDS

medical informatics; usability testing; clinical decision support systems; artificial intelligence; AI; natural language processing; electronic health records; human-centered design; health IT; trust calibration

Introduction

Background

Large language models can generate fluent summaries of long clinical text, yet in high-risk domains such as health care, trust depends not only on accuracy but also on verifiability [1,2]. This is especially true for longitudinal medical record summaries, which condense years of fragmented documentation into a single narrative [3]. Unlike encounter summaries written immediately after care [4,5], longitudinal summaries often serve as a clinician's first exposure to an unfamiliar patient's history.

Without prior mental models, clinicians may struggle to distinguish rare but true details from model error and may find it difficult to trust these outputs [6,7]. The burden of verification in time-constrained environments associated with locating supporting evidence within long and redundant notes remains a central barrier to adoption [2,8].

Most deployed clinical summarization systems have focused on encounter-based summaries such as visit notes, discharge

XSL•FO
RenderX

summaries, and emergency department documentation, which are generated immediately after a clinical interaction [4,5,9,10]. In these settings, clinicians already possess a mental model of the patient from direct participation in care, and summaries primarily support recall, documentation, and communication [11]. In contrast, longitudinal medical record summaries synthesize years of fragmented documentation across multiple encounters and specialties into a single narrative [3,12]. This creates a distinct cognitive challenge because clinicians must determine whether individual summarized statements are supported by the underlying medical record while simultaneously constructing an understanding of the patient's history. As a result, longitudinal summarization requires not only accurate summaries but also efficient mechanisms for verifying individual clinical claims during first-pass review [2,8].

Prior Work

Existing provenance approaches commonly rely on document-level citations, section references, or phrase-level highlights [9-12]. While familiar from academic writing, these mechanisms often require manual search within cited documents. In longitudinal clinical documentation, relevant evidence may be embedded within multipage notes, making targeted verification inefficient.

Recent AI-assisted clinical information retrieval and summarization systems, including Perplexity, OpenEvidence, and Microsoft Copilot, have incorporated provenance through document-level citations, section references, or highlighted supporting snippets to improve transparency [13-15]. Similarly, human-centered explainable AI research has emphasized that explanations should be designed to support user decision-making and integrate naturally into existing workflows rather than simply increasing model transparency [16,17]. However, even when provenance is available, clinicians must still navigate lengthy clinical notes and determine whether the cited evidence adequately supports individual summarized claims, a process that can interrupt reading flow and increase cognitive burden during time-constrained chart review [18-21].

Although these approaches improve transparency, little work has specifically examined provenance interfaces designed to support rapid verification of individual clinical statements during longitudinal chart review. This represents an important gap in the literature because clinicians reviewing unfamiliar patients must efficiently distinguish supported clinical information from potential model errors while simultaneously building an understanding of the patient's history, yet few provenance interfaces have been designed specifically to support this verification workflow.

Study Objectives

We designed and evaluated a sentence-level provenance interaction model ("click to inspect") that makes every summary sentence directly inspectable in its source context. We conducted formative usability testing to assess whether this interaction reduces friction during verification and supports calibrated trust.

The objective of this study was to evaluate the usability of a sentence-level provenance interface for AI-generated longitudinal medical record summaries. Specifically, we aimed to (1) evaluate the perceived usability of the interface using the System Usability Scale (SUS); (2) assess clinicians' overall experience using the net promoter score (NPS); and (3) characterize clinician perceptions of verification workflow, verification burden, and trust through qualitative usability interviews. Collectively, these aims sought to determine whether sentence-level provenance represents a usable and practical approach for supporting rapid verification of AI-generated longitudinal medical record summaries during clinical chart review.

Methods

Overview of Study Design

This study was a formative usability evaluation of a web-based sentence-level provenance interface for AI-generated longitudinal medical record summaries. The system was evaluated using synthetic patient charts designed to simulate common longitudinal documentation patterns encountered in clinical practice.

The purpose of this study was to evaluate the usability of the sentence-level provenance interface rather than the accuracy or clinical performance of the underlying summarization system. The AI-generated summaries used during testing were iteratively refined through structured clinician feedback to ensure that they reflected realistic longitudinal clinical documentation and supported evaluation of the verification workflow. Accordingly, the summaries served as a vehicle for evaluating the provenance interface and were not used to assess the performance of a specific language model or summarization approach.

The Tool and Interface

Participants interacted with a web-based sentence-level provenance interface presenting AI-generated longitudinal medical record summaries alongside a source note viewer. Every sentence within the summary was interactive. When a participant selected a summary sentence, the interface opened the originating clinical note in a side-by-side view, automatically scrolled to the corresponding source sentence, and highlighted the matched text in context. This interaction was designed to support a rapid verification workflow by allowing clinicians to inspect supporting evidence without manually searching through source documentation (Figures 1-3).

XSL•FO
RenderX

Figure 1. Click-to-inspect summaries in use. A clinician clicks on a sentence in the longitudinal AI summary (left), and the system opens the originating note (right), automatically scrolling to and highlighting the matched supporting sentence in context.

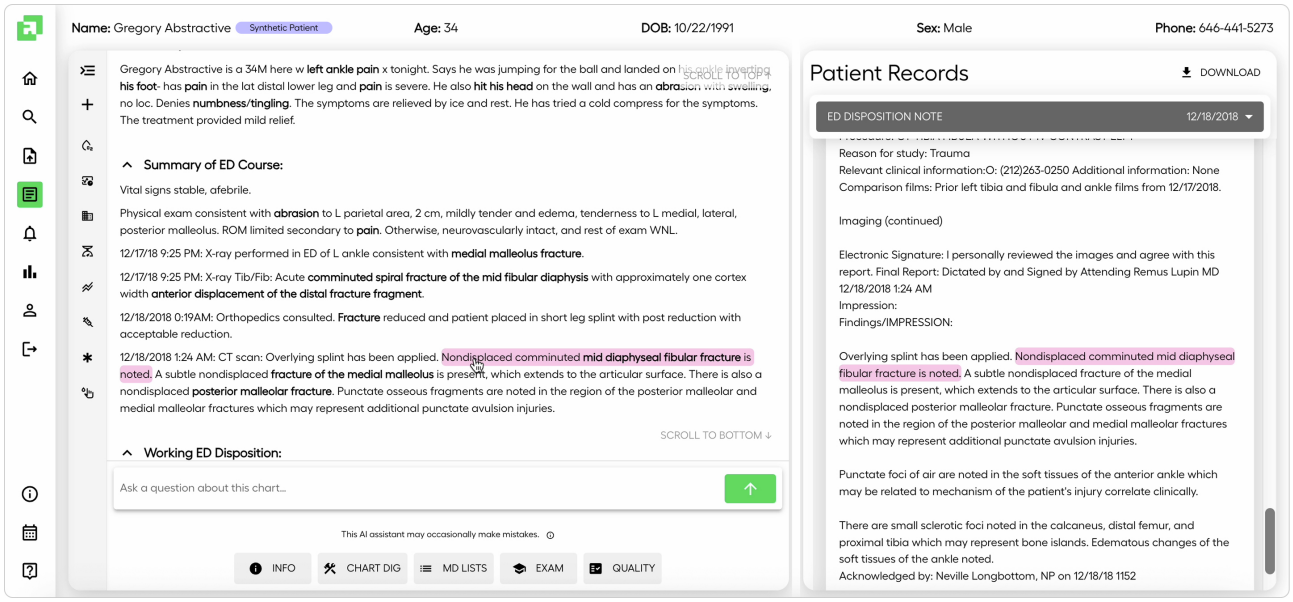
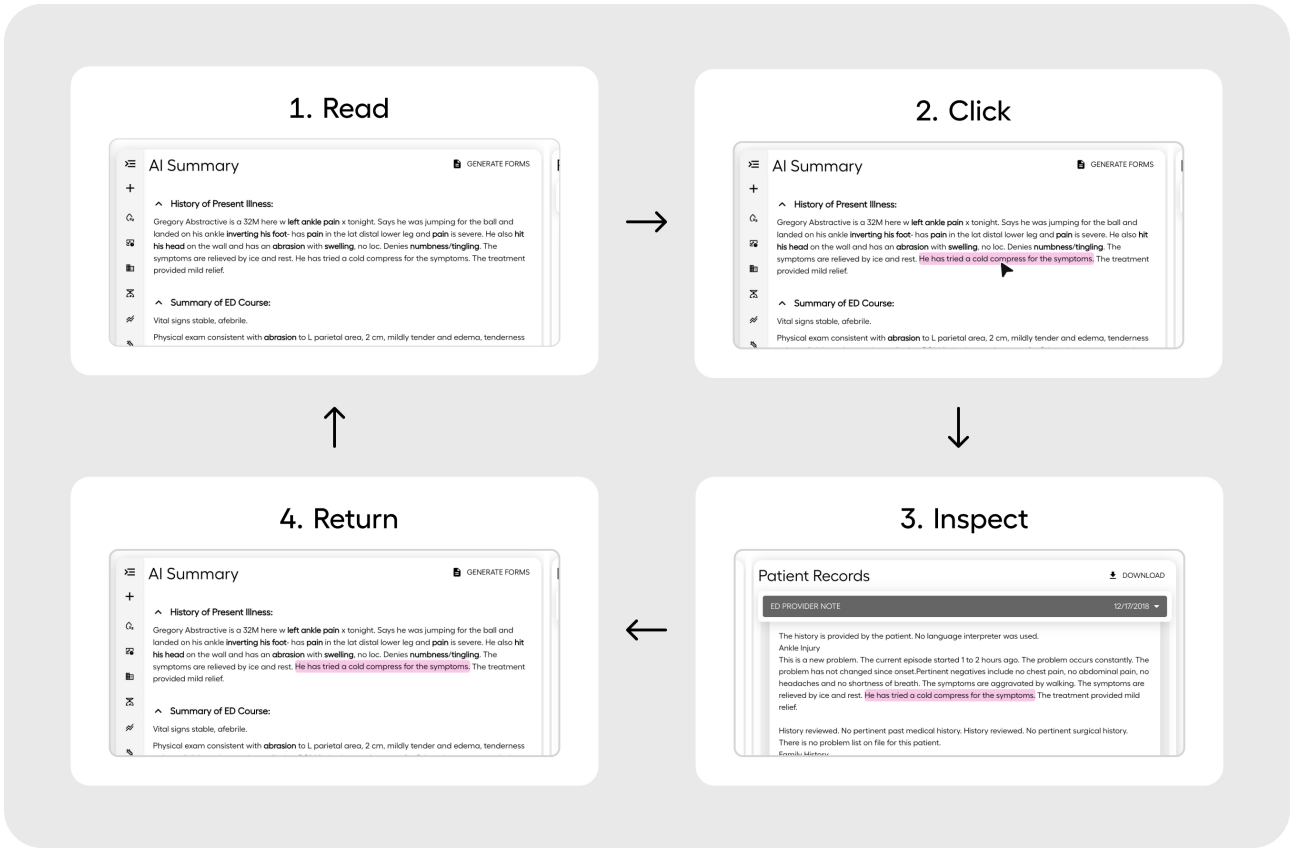
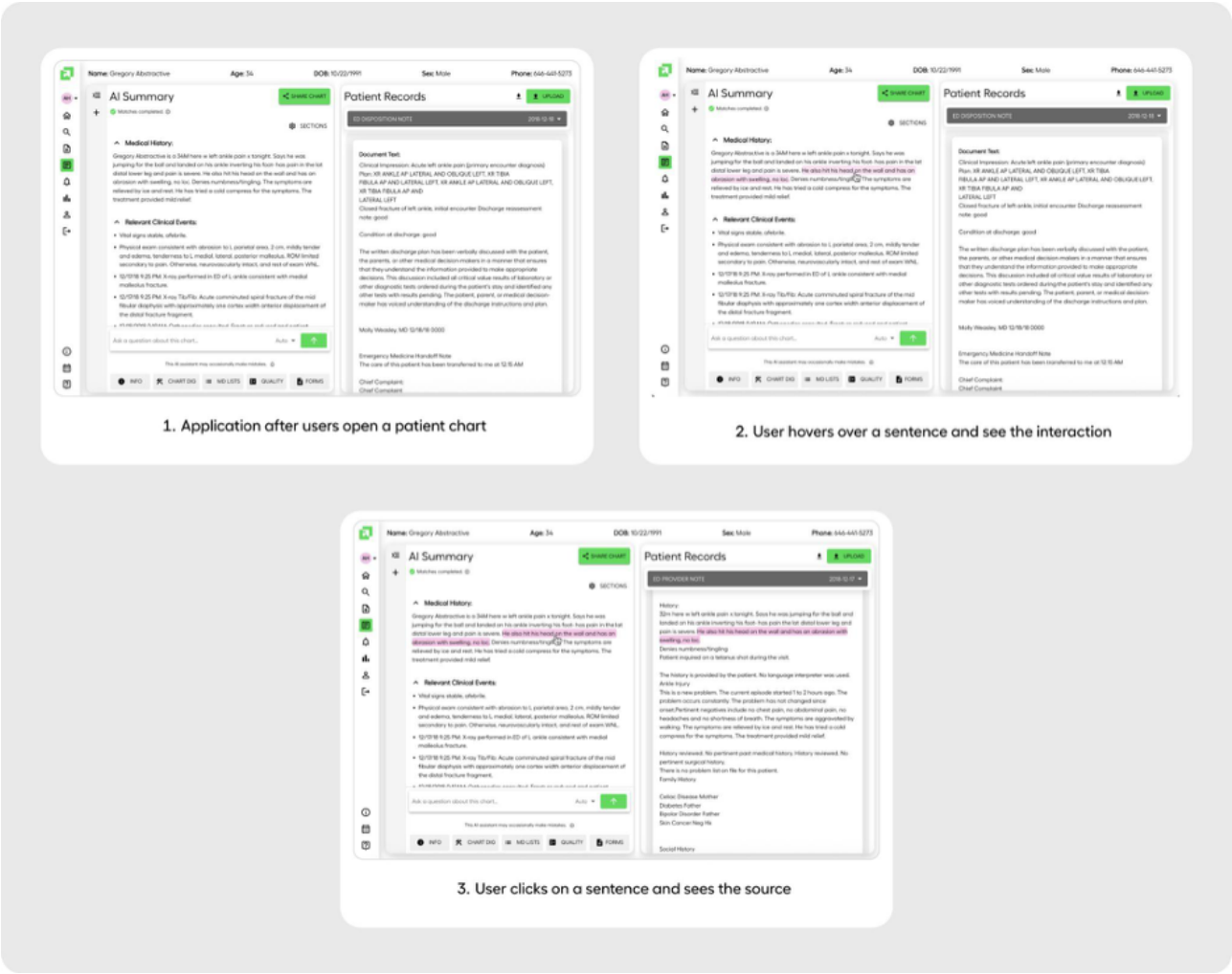


Figure 2. The click-to-inspect loop, where selecting a summary sentence opens the originating note, scrolls to the matched supporting sentence, and highlights it in context for rapid verification and return to reading.



XSL•FO
RenderX

Figure 3. The general overview of the workflow that a user would go through when navigating the application.



Two synthetic patient charts were developed for this study. The first represented an acute emergency department encounter involving an ankle injury, diagnostic imaging, and specialty consultation. The second represented a longitudinal outpatient history spanning multiple years of multispecialty documentation. Both charts were designed to reflect common longitudinal documentation patterns encountered in clinical practice, including fragmented histories, cross-specialty documentation, and varying levels of clinical complexity. Synthetic charts were authored and reviewed by board-certified clinicians to ensure clinical realism while avoiding the use of real patient information.

Sentence-level provenance links were generated using embedding-based semantic similarity between summary sentences and candidate source note sentences. Sentence embeddings were generated using the transformer-based sentence embedding model all-mpnet-base-v2 (Hugging Face). Before embedding, source clinical notes were converted from their structured representation into plain text and segmented into sentence-level units. Text was initially partitioned using line breaks, runs of 2 or more white space characters, bullet symbols, and dash space separators. Each resulting text segment was trimmed of surrounding white space and further segmented into individual sentences using the Natural Language Toolkit (Team NLTK) Punkt sentence tokenizer. Empty segments were discarded. No abbreviation expansion, spelling correction, or clinical terminology normalization was performed before embedding.

Cosine similarity was calculated between each summary sentence and all candidate sentences within the originating clinical note. For each summary sentence, the highest-scoring source sentence was selected as the provenance reference. A cosine similarity threshold of 0.70 was used to categorize provenance matches. Summary sentences whose highest similarity score was 0.70 or more were categorized as verified, whereas those with scores below 0.70 were categorized as unverified while remaining clickable within the interface. The threshold was established during iterative internal clinical review of provenance quality to identify a value that consistently distinguished clinically meaningful sentence matches from weaker semantic associations. This represented a clinician-informed calibration process rather than a formally validated statistical threshold.

The provenance interface used a one-to-one sentence mapping. Each summary sentence was linked to a single originating clinical note and a single highest-scoring source sentence within that note. The underlying summarization pipeline preserved the originating note associated with each generated summary sentence and did not generate individual summary sentences by combining information across multiple source notes.

XSL•FO
RenderX

Although a summary sentence could synthesize information from multiple sentences within the same note, the interface intentionally displayed only the highest-scoring supporting sentence. This design was selected to provide a single, consistent point of verification while minimizing visual complexity during chart review. Findings from the usability evaluation supported this approach as participants consistently described a single salient provenance location as intuitive and sufficient for verification during longitudinal chart review.

Participant Sampling and Recruiting

Clinicians were recruited between November 2023 and January 2024 through email outreach to medical students, resident physicians, and attending physicians using academic institutional mailing lists, residency program listservs, and professional networks across multiple health systems in the United States. Recruitment emails described the study as a usability evaluation of a clinical summarization interface and invited interested participants to complete a brief interest form. Participation was voluntary. Eligible participants were required to be actively engaged in clinical training or practice. No additional exclusion criteria were applied.

Usability Testing Procedures

Moderated usability sessions were conducted remotely via Zoom (Zoom Video Communications) between November 2023 and January 2024. Each session lasted approximately 30 minutes and followed a semistructured usability interview guide to ensure consistency across participants (Multimedia Appendix 1). The guide consisted of standardized clinical scenarios; predefined task-based usability assessments; and open-ended interview questions designed to evaluate clinician workflow, verification behavior, and overall usability of the sentence-level provenance interface.

During each session, participants interacted with the web-based interface while reviewing AI-generated longitudinal patient summaries. Participants were encouraged to use a think-aloud protocol, verbalizing their reasoning as they reviewed summaries and selectively inspected supporting source material. Moderators used nondirective follow-up prompts to clarify why participants chose to verify particular statements and how the interface influenced their verification process while allowing for flexibility to explore emerging usability issues. Participants were not required to inspect a predetermined number of statements, allowing for observation of natural verification behavior during clinical chart review. All sessions were audio and video recorded and transcribed for subsequent qualitative analysis.

Data Collection and Analysis

Following interaction with the interface, participants completed the SUS and an NPS questionnaire. The SUS is a validated 10-item instrument measuring perceived usability on a scale from 0 to 100. The NPS was assessed using the standard likelihood to recommend question (0-10), with respondents categorized as promoters (9-10), passives (7-8), and detractors (0-6).

In addition to quantitative usability measures, qualitative data were collected during these remotely moderated usability testing sessions via Zoom using a think-aloud session, in which participants described their verification strategies, perceived workflow friction, trust in AI-generated summaries, and overall experience using the interface during the Zoom session.

Descriptive statistics were calculated for quantitative usability outcomes. Mean SUS scores, SDs, and 95% CIs were calculated across participants. The NPS was calculated using the standard formula of the percentage of promoters minus the percentage of detractors.

Audio recordings were transcribed and independently reviewed by 2 researchers. Qualitative data were analyzed using an inductive thematic analysis approach to identify recurring themes related to verification workflow, verification burden, trust calibration, and integration into clinical practice. Themes were iteratively refined through discussion until consensus was achieved.

Ethical Considerations

This study was reviewed and approved by Pearl IRB (Indianapolis, Indiana, United States; 2025-0079) under an expedited review process and was determined to involve minimal risk. All study procedures were conducted in accordance with the approved protocol. Participants provided verbal informed consent before taking part. To protect participant privacy and confidentiality, all usability sessions were audio and video recorded solely for research purposes, securely stored in accordance with the approved study protocol, and analyzed by authorized study personnel. Synthetic patient charts were used throughout the study to ensure that no real patient information or protected health information was accessed or disclosed during usability testing. Participants received a US $25 electronic gift card upon completion of the usability session.

Results

Recruitment invitations were distributed through multiple overlapping institutional mailing lists, residency program listservs, and professional networks. As a result, the exact number of clinicians who received the invitation could not be determined because individual recipients may have received invitations through more than one distribution channel. During the recruitment period, 107 clinicians expressed interest in participating in the study, of whom 46 (43%) completed a moderated usability session and were included in the final analysis (Table 1).

XSL•FO
RenderX

Table . Characteristics of study participants (N=46).

Characteristics	Values
Age range (y)	20 - 60
Clinical specialty groups represented, n (%)	
Internal medicine	11 (23.9)
Surgical specialties	10 (21.7)
Radiology	4 (8.7)
Dermatology	3 (6.5)
Neurology	3 (6.5)
Ophthalmology	3 (6.5)
Emergency medicine	2 (4.3)
Family medicine	2 (4.3)
Obstetrics and gynecology	2 (4.3)
Pediatrics	2 (4.3)
Hematology and oncology	2 (4.3)
Anesthesiology	1 (2.2)
Psychiatry	1 (2.2)

Recruitment concluded when participant feedback demonstrated recurring and consistent observations regarding workflow, verification behavior, and overall usability. At that point, the sample size was considered sufficient for the formative usability evaluation because additional recruitment was unlikely to yield substantially new usability insights.

The mean SUS score was 86.25 (SD 7.77; 95% CI 83.96 - 88.54), indicating high perceived usability of the click-to-inspect interface.

The NPS was 35, calculated using the standard method (percentage of promoters minus percentage of detractors on the 0 - 10 NPS scale). Of the 46 participants, 22 (47.8%) were classified as promoters (scores of 9 - 10), 18 (39.1%) were classified as passives (scores of 7 - 8), and 6 (13%) were classified as detractors (scores of 0 - 6).

During the moderated think-aloud sessions, participants frequently described rapid access to supporting evidence as critical to their willingness to rely on the summary during initial review. Clinicians reported that being able to verify individual statements with a single click reduced cognitive load compared with document-level citation systems that require manual searching within lengthy notes. Participants noted that traditional provenance mechanisms often identify the source document but do not eliminate the effort required to locate the exact supporting passage within that document.

Sentence-level inspectability was consistently described as predictable and low friction. Participants reported that the absence of persistent citation markers or confidence indicators preserved reading flow while maintaining immediate access to supporting evidence when needed. Participants also emphasized that the ability to selectively verify clinically important or unexpected statements aligned well with real-world chart review workflows.

Representative participant comments included the following:

> *Being able to click directly on the sentence and see the exact line in the note makes it much easier to trust the summary.*
>
> *Normally I have to scroll through the whole note to find where something came from. This saves a lot of time.*
>
> *I like that I can verify the statements that matter without breaking my reading flow.*

Discussion

Principal Findings

This formative usability evaluation found that a sentence-level provenance interface for AI-generated longitudinal medical record summaries demonstrated high perceived usability among clinicians, with a mean SUS score of 86.25 (SD 7.77) and an NPS of 35. Consistent with the objectives of this study, clinicians reported that the click-to-inspect interaction reduced the effort required to verify AI-generated clinical statements while preserving reading flow during longitudinal chart review. Participants consistently described rapid access to supporting source material as enabling selective verification of clinically important statements without interrupting their workflow. These findings support the usability of sentence-level provenance as a verification interface independent of the underlying summarization model and suggest that embedding provenance directly within individual summary statements may reduce verification burden while supporting calibrated trust in AI-generated longitudinal summaries.

Our findings suggest that the value of provenance extends beyond simply identifying the source of summarized information. Participants reported that the interface made it easier to verify individual clinical statements while maintaining

XSL•FO
RenderX

focus on understanding the patient's history. Instead of navigating through lengthy clinical notes to locate supporting evidence, clinicians could inspect the source material only when they felt that additional verification was necessary. This workflow appeared to reduce the effort associated with verification and allowed clinicians to remain focused on interpreting the summary rather than searching for supporting documentation.

These findings are consistent with those of prior work in explainable AI showing that transparency mechanisms are most effective when they support user decision-making within existing workflows rather than simply exposing model outputs [16,17]. Existing AI-assisted clinical summarization and information retrieval systems commonly provide document-level citations, section references, or highlighted supporting snippets [13-15,22-25]. Although these approaches improve transparency, clinicians must still determine whether the cited material adequately supports the summarized clinical claim while navigating lengthy and fragmented documentation. Our findings suggest that sentence-level provenance may reduce this burden by making supporting evidence immediately accessible at the point where verification is needed.

The findings are also consistent with principles from human factors research. Clinical reasoning is typically performed at the level of individual findings, diagnoses, medication changes, and clinical events rather than entire documents. Participants consistently reported that they wanted the ability to verify only selected statements rather than every sentence within a summary. Clinicians naturally focused on information that was clinically important, unexpected, or likely to influence decision-making. The interface supported this workflow by allowing participants to selectively inspect supporting evidence without interrupting their review of the remainder of the summary. Participants also noted that the absence of persistent citation markers or confidence scores preserved reading flow while maintaining immediate access to supporting evidence when desired. Together, these observations suggest that minimizing visual complexity while preserving rapid inspectability may represent an important design principle for future AI-assisted clinical documentation systems.

Several limitations should be considered when interpreting these findings. First, this was a formative usability evaluation conducted using synthetic patient charts. Although synthetic data enable controlled testing without privacy concerns, they may not fully capture the variability, ambiguity, and complexity of real-world longitudinal clinical documentation. Second, although participants represented a broad range of clinical specialties, recruitment took place through academic institutional mailing lists, residency program listservs, and professional networks, which may limit the generalizability of these findings to other clinical settings. Third, this study evaluated perceived usability rather than objective performance. Outcomes such as time to verification, verification accuracy, detection of hallucinated or unsupported statements, and downstream effects on clinical decision-making were not measured. Accordingly, this study cannot determine whether sentence-level provenance improves clinicians' ability to identify summarization errors or unsupported clinical claims. Finally, provenance mapping relied on embedding-based semantic similarity, and its robustness across heterogeneous clinical documentation warrants further evaluation.

Future studies should evaluate sentence-level provenance in real-world clinical environments using production electronic health record data and AI-generated longitudinal summaries. Important outcome measures include verification time, verification accuracy, detection of unsupported or hallucinated statements, changes in information-seeking behavior, and downstream effects on clinical decision-making. Comparative studies should also evaluate sentence-level provenance against conventional document navigation and document-level citation interfaces to determine whether it improves verification efficiency, clinical reasoning, and trust calibration. Additional work is needed to understand how alternative provenance presentation strategies, including confidence scores and prioritization cues, influence verification behavior and clinician reliance on AI-generated summaries.

This study suggests that sentence-level provenance is a feasible and well-accepted approach for supporting verification of AI-generated longitudinal medical record summaries. As generative AI becomes increasingly integrated into electronic health record workflows, the design of verification interfaces may become as important as improvements in model performance. Supporting efficient, low-friction verification has the potential to improve clinician confidence, promote appropriately calibrated trust, and facilitate the safe integration of AI-generated longitudinal summaries into routine clinical practice.

Conclusions

Sentence-level provenance represents a human-centered approach to improving verifiability in AI-generated longitudinal medical record summaries. By making every summary sentence directly inspectable within its source context, the click-to-inspect interaction embeds verification into the act of reading rather than positioning it as an external search task. In formative usability testing, clinicians described the interface as intuitive and compatible with time-constrained workflows and reported high perceived usability. Although further empirical validation is needed, these findings suggest that designing provenance at the level of individual clinical statements may better align AI summarization tools with real-world clinical practice and support calibrated reliance in high-risk settings.

Acknowledgments

The authors acknowledge Caroline Reiner for contributions to the sentence-matching algorithm and user experience interviews. Generative AI (ChatGPT based on GPT-4; OpenAI) was used solely to assist with spelling and grammar corrections during

XSL•FO
RenderX

manuscript preparation. The authors reviewed and edited all AI-assisted revisions and take full responsibility for the final content of the manuscript.

Funding

This work was supported by the US National Science Foundation Small Business Innovation Research Phase II award. The views expressed are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Data Availability

Synthetic patient charts were used in this study. The datasets generated or analyzed during this study are not publicly available due to study constraints but are available from the corresponding author on reasonable request. Synthetic patient charts were used in this study. The clinical sentence-matching simulation can also be made available for demonstration upon reasonable request. Due to proprietary software components, the underlying code is not publicly available.

Authors' Contributions

GP contributed to the conceptualization and design of the study, development of the click-to-inspect interface, data collection, data interpretation, and drafting of the manuscript. AP contributed to manuscript drafting, critical revision for important intellectual content, and interpretation of the findings. VH contributed to the conceptualization of the project, study design, manuscript drafting, and critical revision for important intellectual content. All authors reviewed and approved the final version of the manuscript and met the International Committee of Medical Journal Editors criteria for authorship.

Conflicts of Interest

GP and VH are cofounders of Abstractive Health and hold equity interests in the company. AP also holds an equity interest in Abstractive Health. Because the sentence-level provenance interface evaluated in this study was developed by Abstractive Health, these financial interests could potentially benefit from favorable study findings. The authors designed, conducted, analyzed, and reported the study with this potential conflict disclosed.

Multimedia Appendix 1
Interview guide for moderators of usability testing sessions.
[PDF File, 8 KB - humanfactors_v13i1e95644_app1.pdf]

References

1. Van Veen D, Van Uden C, Blankemeier L, et al. Adapted large language models can outperform medical experts in clinical text summarization. Nat Med 2024 Apr;30(4):1134-1142. [doi: 10.1038/s41591-024-02855-5] [Medline: 38413730]
2. Goodman KE, Yi PH, Morgan DJ. AI-generated clinical summaries require more than accuracy. JAMA 2024 Feb 27;331(8):637-638. [doi: 10.1001/jama.2024.0555] [Medline: 38285439]
3. Pivovarov R, Elhadad N. Automated methods for the summarization of electronic health records. J Am Med Inform Assoc 2015 Sep;22(5):938-947. [doi: 10.1093/jamia/ocv032] [Medline: 25882031]
4. Hartman VC, Bapat SS, Weiner MG, Navi BB, Sholle ET, Campion TR Jr. A method to automate the discharge summary hospital course for neurology patients. J Am Med Inform Assoc 2023 Nov 17;30(12):1995-2003. [doi: 10.1093/jamia/ocad177] [Medline: 37639624]
5. Williams CY, Bains J, Tang T, et al. Evaluating large language models for drafting emergency department encounter summaries. PLOS Digit Health 2025 Jun;4(6):e0000899. [doi: 10.1371/journal.pdig.0000899] [Medline: 40526634]
6. Lu Z, Yin M. Human reliance on machine learning models when performance feedback is limited: heuristics and risks. In: CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems: Association for Computing Machinery; 2021. [doi: 10.1145/3411764.3445562]
7. Rosenbacke R, Melhus Å, McKee M, Stuckler D. How explainable artificial intelligence can increase or decrease clinicians' trust in AI applications in health care: systematic review. JMIR AI 2024 Oct 30;3:e53207. [doi: 10.2196/53207] [Medline: 39476365]
8. Asgari E, Montaña-Brown N, Dubois M, et al. A framework to assess clinical safety and hallucination rates of LLMs for medical text summarisation. NPJ Digit Med 2025 May 13;8(1):274. [doi: 10.1038/s41746-025-01670-7] [Medline: 40360677]
9. Silberlust J, Solanki P, Stevens ER, et al. Artificial intelligence-generated encounter summaries: early insights from ambulatory clinicians at a large academic health system. JAMIA Open 2025;8(5):ooaf096. [doi: 10.1093/jamiaopen/ooaf096] [Medline: 40904519]
10. Krishna K, Khosla S, Bigham J, Lipton ZC. Generating SOAP notes from doctor-patient conversations using modular summarization techniques. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: Association for Computational Linguistics; 2021:4958-4972. [doi: 10.18653/v1/2021.acl-long.384]

XSL•FO
RenderX

11. Feblowitz JC, Wright A, Singh H, Samal L, Sittig DF. Summarization of clinical information: a conceptual model. J Biomed Inform 2011 Aug;44(4):688-699. [doi: 10.1016/j.jbi.2011.03.008] [Medline: 21440086]
12. Hirsch JS, Tanenbaum JS, Lipsky Gorman S, et al. HARVEST, a longitudinal patient record summarizer. J Am Med Inform Assoc 2015 Mar;22(2):263-274. [doi: 10.1136/amiajnl-2014-002945] [Medline: 25352564]
13. Perplexity. URL: https://perplexity.ai/ [accessed 2026-02-01]
14. OpenEvidence. URL: https://openevidence.com/ [accessed 2026-02-01]
15. Microsoft Copilot. URL: https://copilot.microsoft.com/ [accessed 2026-02-01]
16. Liao QV, Varshney KR. Human-centered explainable AI (XAI): from algorithms to user experiences. arXiv. Preprint posted online on Oct 20, 2021. [doi: 10.48550/arXiv.2110.10790]
17. Zhang Y, Liao QV, Bellamy RK. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In: FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency: Association for Computing Machinery; 2020:295-305. [doi: 10.1145/3351095.3372852]
18. Overhage JM, McCallie DJ. Physician time spent using the electronic health record during outpatient encounters: a descriptive study. Ann Intern Med 2020 Feb 4;172(3):169-174. [doi: 10.7326/M18-3684] [Medline: 31931523]
19. Downing NL, Bates DW, Longhurst CA. Physician burnout in the electronic health record era: are we ignoring the real cause? Ann Intern Med 2018 Jul 3;169(1):50-51. [doi: 10.7326/M18-0139] [Medline: 29801050]
20. Arndt BG, Beasley JW, Watkinson MD, et al. Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. Ann Fam Med 2017 Sep;15(5):419-426. [doi: 10.1370/afm.2121] [Medline: 28893811]
21. Kambhamettu H, Metaxa D, Johnson K, Head A. Explainable notes: examining how to unlock meaning in medical notes with interactivity and artificial intelligence. In: CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems: Association for Computing Machinery; 2024:1-19. [doi: 10.1145/3613904.3642573]
22. Kim SS, Vaughan JW, Liao QV, Lombrozo T, Russakovsky O. Fostering appropriate reliance on large language models: the role of explanations, sources, and inconsistencies. In: CHI '25: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems: Association for Computing Machinery; 2025:1-19. [doi: 10.1145/3706598.3714020]
23. Ding Y, Facciani M, Joyce E, et al. Citations and trust in LLM generated responses. Proc AAAI Conf Artif Intell 2025;39(22):23787-23795. [doi: 10.1609/aaai.v39i22.34550]
24. Hoque MN, Mashiat T, Ghai B, et al. The HaLLMark effect: supporting provenance and transparent use of large language models in writing with interactive visualization. In: CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems: Association for Computing Machinery; 2024:1-15. [doi: 10.1145/3613904.3641895]
25. Kambhamettu H, Flores J, Head A. Traceable texts and their effects: a study of summary-source links in AI-generated summaries. In: CHI EA '25: Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems: Association for Computing Machinery; 2025:1-7. [doi: 10.1145/3706599.3719830]

Abbreviations

NPS: net promoter score
SUS: System Usability Scale

Edited by A Kushniruk, M Balcarras, S Law; submitted 18.Mar.2026; peer-reviewed by I Koulas, R Jonker; revised version received 27.Jul.2026; accepted 30.Jul.2026; published 08.Sep.2026.

Please cite as:
Parambath A, Pulpo G, Hartman V
Sentence-Level Provenance for AI Medical Record Summarization in a Click-to-Inspect Interface: Formative Usability Evaluation
JMIR Hum Factors 2026;13:e95644
URL: https://humanfactors.jmir.org/2026/1/e95644
doi:10.2196/95644

© Andrew Parambath, Giordana Pulpo, Vince Hartman. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 8.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Understanding Delivery and Engagement With a Digital Self-Management Intervention for Chronic Obstructive Pulmonary Disease Across Two Clinical Settings: Qualitative Study

Martin Ruddock[1,2], BSc, MA, MSc; Alison Blythin[3], RGN, MRes; Bethany Cliffe[4], BSc, PhD; Lucy Yardley[1,5,6], PhD; James Dodd[7], PhD; Rachel Williams[8], MSc; Katherine Bradbury[1], PhD; Tom Wilkinson[2,9], PhD; Ben Ainsworth[1], PhD

[1]Faculty of Environmental and Life Sciences, School of Psychology, University of Southampton, University Road, Southampton, England, United Kingdom

[2]NIHR Southampton Biomedical Research Centre, Tremona Road, Southampton General Hospital, Southampton, England, United Kingdom

[3]my mhealth Limited, London, United Kingdom

[4]University of Bristol, Bristol, England, United Kingdom

[5]School of Psychological Science, University of Bristol, Bristol, England, United Kingdom

[6]NIHR Applied Research Collaboration West, Bristol, England, United Kingdom

[7]Faculty of Medicine, Academic Respiratory Unit, University of Bristol, Bristol, England, United Kingdom

[8]Liskeard Community Hospital, Respiratory Team, Cornwall Partnership NHS Foundation Trust, Liskeard, United Kingdom

[9]Faculty of Medicine, Clinical and Experimental Sciences, University of Southampton, Southampton, England, United Kingdom

Corresponding Author:
Martin Ruddock, BSc, MA, MSc
Faculty of Environmental and Life Sciences, School of Psychology, University of Southampton, University Road, Southampton, England, United Kingdom

Abstract

Background: Adherence to chronic obstructive pulmonary disease (COPD) self-management plans can improve health outcomes, yet access to pulmonary rehabilitation remains limited. Digital self-management interventions, such as *myCOPD*, offer a potential means to extend support across care pathways. However, little is known about how these tools are delivered in routine practice or how delivery influences patient engagement across different clinical settings.

Objective: This study explored perceived barriers and facilitators to the delivery of and engagement with *myCOPD* by examining the perspectives of both patients and health care professionals (HCPs) across 2 National Health Service settings.

Methods: This qualitative study was conducted between November 2023 and April 2024 across 2 contrasting clinical pathways: community pulmonary rehabilitation and a hospital discharge service following acute exacerbation of COPD. Semistructured interviews were conducted with patients using *myCOPD* and HCPs involved in its delivery. Topic guides covered health background, digital technology use, and experiences of engaging with the intervention. A convenience sampling approach, with elements of purposive sampling, ensured representation across settings. Clinical data were obtained from health records with consent, and usage data were provided by my mhealth Ltd. Data were analyzed using abductive thematic analysis informed by the Medical Research Council's process evaluation framework.

Results: Thirty interviews were completed (16 patients and 14 HCPs). Patients described how multimorbidity, perceived digital ability, and socioeconomic context shaped their engagement with *myCOPD*, operating as "layers of vulnerability" that influenced whether COPD self-management could be prioritized. HCPs highlighted organizational and workflow constraints that shaped how the intervention was introduced and supported. Delivery approaches differed markedly between settings: community teams adopted proactive, relationship-based support that enabled iterative discussion, tailored recommendations, and continuity, whereas hospital teams described passive, onboarding-focused delivery shaped by time pressures, staffing constraints, and discharge priorities. These contrasting delivery models influenced engagement patterns, with proactive support associated with more *multifaceted engagement* (characterized as routine and sustained) and passive onboarding associated with *single-aspect engagement* (characterized as short-term and transactional). Participants also identified perceived barriers (eg, usability concerns, limited integration with in-person care) and perceived benefits (eg, improved inhaler technique, increased confidence, lifestyle adjustments).

Conclusions: Delivery context and local workflows play a central role in shaping how patients engage with digital self-management tools. Proactive, relational support appears to facilitate deeper and more sustained engagement, whereas passive onboarding may limit the intervention's potential. Implementation strategies should align digital tools with local capacities,

XSL•FO
RenderX

staffing structures, and patient characteristics, attending to layered vulnerabilities such as multimorbidity and digital confidence. Further work is needed to understand how proactive delivery models can be resourced and integrated within routine COPD care.

(JMIR Hum Factors 2026;13:e88108) doi:10.2196/88108

KEYWORDS

digital health interventions; engagement; delivery; digital health; real world; evaluation; qualitative

Introduction

Chronic obstructive pulmonary disease (COPD) affects an estimated 3 million people in the United Kingdom and contributes to approximately 30,000 deaths annually, many of which are preventable [1,2]. Exacerbations are clinically significant events that increase the risk of subsequent exacerbations and hospital readmissions, particularly within the first 3 months [3,4]. Effective self-management, including symptom recognition, inhaler technique, and timely action, can reduce exacerbations and improve quality of life [5,6]. However, access to pulmonary rehabilitation (PR), a core component of COPD self-management, remains limited, with fewer than one-third of eligible patients accessing PR within 90 days of referral [7]. These persistent gaps highlight the need for scalable approaches to support self-management across care pathways.

Digital self-management interventions offer a potential means to extend PR and provide ongoing support, yet evidence for their effectiveness remains mixed [8-10]. While some studies report improvements in inhaler technique, symptom monitoring, and exercise adherence, others highlight challenges related to uptake, sustained engagement, and integration into routine care. A recurring issue is that digital interventions are often implemented without sufficient understanding of how they fit within existing clinical workflows or how frontline teams introduce and support their use [11]. Implementation research emphasizes that digital tools are not inherently effective; their impact depends on how they are delivered, supported, and embedded within organizational contexts. Factors such as staffing capacity, digital readiness, professional roles, and local service priorities can shape adoption, fidelity, and engagement [12].

Engagement with digital interventions is increasingly conceptualized as an implementation outcome influenced by both individual and structural determinants. COPD disproportionately affects older adults and people experiencing socioeconomic disadvantage, who may also face barriers related to digital access, confidence, and multimorbidity [13,14]. These overlapping factors can shape how individuals navigate digital tools and the extent to which they are able, or supported, to engage with them [15,16]. Understanding how patient characteristics intersect with service-level contexts is therefore essential for ensuring equitable benefit. An intersectional approach, attending to how multiple characteristics combine to produce advantage or disadvantage, can illuminate how patients navigate digital tools within the constraints of multimorbidity, digital ability, and socioeconomic context [11,17,18]. The concept of "layers of vulnerability" proposed by Luna [19,20] is particularly relevant for understanding how vulnerabilities emerge from the interaction between individuals and their health care environments, rather than from individual characteristics alone.

myCOPD is a widely adopted digital self-management intervention recommended by the NICE Early Value Assessments for PR and COPD self-management [21]. Despite its deployment across National Health Services (NHSs), limited evidence exists on how *myCOPD* is implemented in real-world settings, how delivery varies across clinical pathways, or whether these variations influence patient engagement. Existing evaluations have focused primarily on clinical outcomes or usability, with less attention to the organizational and relational processes that shape engagement in practice [22-25].

To address this gap, we examined the implementation of *myCOPD* across 2 contrasting NHS settings: community PR and a hospital discharge pathway following an acute exacerbation of COPD. Guided by the Medical Research Council's (MRC's) process evaluation framework [26], we explored how contextual factors, delivery approaches, and patient characteristics interact to shape engagement trajectories. This study aimed to identify implementation factors influencing the delivery of and engagement with *myCOPD* across 2 NHS clinical settings and explore how patient characteristics and structural factors interact to shape engagement trajectories, informed by the concept of "layers of vulnerability."

By analyzing implementation processes across settings, this study contributes to our understanding of how digital self-management interventions can be more effectively and equitably integrated into routine COPD care. These aims informed the sampling strategy, interview topic guides, and abductive analytic approach.

Methods

Study Settings

The study was conducted across 2 contrasting NHS clinical settings:

First, the hospital discharge pathway (Bristol), where *myCOPD* was introduced as part of the acute exacerbation of COPD discharge bundle. Eligible patients were identified during follow-up clinics or virtual ward reviews within 6 weeks of hospital discharge. Delivery was primarily supported by digital health champions (DHCs) due to staffing pressures and workflow constraints.

Second, the community PR (Cornwall), where *myCOPD* was integrated into routine PR delivery. A digital health advisor (from my mhealth Ltd) and a respiratory care coordinator supported registration, activation, and ongoing use. Clinicians used the *myCOPD* dashboard to monitor progress and tailor self-management discussions.

XSL•FO
RenderX

These settings differed in purpose, workflow, and available resources, providing a natural comparison of delivery approaches.

Intervention: *myCOPD*

myCOPD is a digital self-management app providing educational modules, inhaler-technique videos, PR content, symptom tracking, and action-plan support. Clinicians can view patient-reported data through a dashboard to support remote monitoring and personalized care. The intervention content was identical across sites.

Participants, Recruitment, and Sampling

Patients were recruited from the PROPEL (A Pragmatic Real-World Multicenter Observational Research Study to Explore the Clinical and Health Economic Impact of *myCOPD*) study cohort. During their baseline PROPEL visit, all eligible participants were informed about the qualitative component and invited to opt in or out. Those who opted in were recontacted to arrange an interview after their scheduled 3-month follow-up visit, timed to ensure adequate exposure to the intervention. The eligibility criteria at each site are outlined in Textbox 1.

Textbox 1. Eligibility criteria across settings.

Setting 1 (Bristol)

- **Inclusion criteria**
  - Adult patients aged >18 years and able to provide informed consent
  - Clinical diagnosis of chronic obstructive pulmonary disease (COPD)
  - Admitted to hospital with a primary diagnosis of acute exacerbation of COPD (AECOPD)
  - Assessed in a follow-up clinic or a virtual ward within 6 weeks of an AECOPD
- **Exclusion criteria**
  - Patients aged <18 years
  - No clinical diagnosis of COPD
  - End-of-life care or palliative care
  - Unable to provide informed consent

Setting 2 (Cornwall)

- **Inclusion criteria**
  - Adult patients aged >18 years and able to provide informed consent
  - Clinical diagnosis of COPD, deemed suitable for referral to pulmonary rehabilitation by the local clinical team
  - Motivated or willing to take part
- **Exclusion criteria**
  - Unstable angina
  - Myocardial infarction within 6 weeks
  - Uncontrolled cardiac arrhythmias
  - Unstable hypertension
  - Severe cognitive impairment
  - Locomotor or other severe medical conditions preventing the patient from safe participation in group activities
  - Unable to provide informed consent
  - Any condition deemed by the principal investigator to make the participant unsuitable for the study

A convenience sampling approach was used, with elements of purposive sampling to ensure representation across both clinical settings. Health care professionals (HCPs) involved in delivering *myCOPD* were identified by site leads and approached via email. A purposive sampling strategy ensured variation in professional role (DHCs, nurses, and physiotherapists) and setting (community vs hospital).

Data Collection

Semistructured interviews were conducted via Microsoft Teams (audio or video, depending on the technological accessibility of participants). Topic guides were tailored for patients and HCPs and refined iteratively by experienced qualitative researchers. Interview topics included experience of COPD, use of digital technology, engagement with *myCOPD*, and perceived

XSL•FO
RenderX

barriers and facilitators. Example questions are presented in the full topic guides in Multimedia Appendix 1.

With informed consent, clinical data (eg, COPD severity, years since diagnosis, long-term conditions) were obtained from participants' health records held by NHS clinical teams. Similarly, with consent, *myCOPD* usage data (eg, logins, module completion, and PR access) were obtained directly from my mhealth Ltd, which provided anonymized usage summaries for participants enrolled in the PROPEL study.

Procedure

At the start of each interview, participants received a verbal recap of study aims, confidentiality, and their right to withdraw. One interviewer was present at each interview (MR conducted all patient interviews, and HCP interviews were split between BC and MR). MR (a PhD researcher in health psychology with training in qualitative interviewing and prior experience as a patient and public involvement [PPI] officer) conducted all patient interviews and facilitated PPI sessions. BC (an experienced qualitative researcher) conducted HCP interviews. Their positionalities—MR's long-standing ties to the local community and BC's research background—were acknowledged and reflexively considered throughout the analysis. Interviews were recorded, transcribed verbatim, checked for accuracy, and imported into NVivo 14 (Lumivero) for analysis. All participants received financial compensation (payments of £75 [US $100] for HCPs; vouchers of £25 [US $30] for patients). Finally, the data presented here include CAT scores that were taken at baseline by clinical research staff.

Analysis

Coding was conducted by 2 researchers (BC and MR) for HCP interviews and by a researcher (MR) for patient interviews. HCP and patient transcripts were coded independently. During theme development, we compared the coded data to identify shared patterns, at which point the analytical process merged into a unified set of themes. Analyzing the transcripts together captured the relational aspects of self-management and reflected an overlap in coding and candidate themes across groups. While both groups discussed a particular theme, typically one group described the theme in more detail; for example, HCPs and patients both described the importance of multimorbidity, but patients described in greater detail how it shaped engagement.

Researcher positionalities were addressed through structured reflexive practices. BC and MR kept journals during data collection and coding, revisiting these reflections during theme development to check for assumptions and bias. MR also held regular debrief meetings with senior qualitative researchers (BA, KB, and LY) to discuss how background experiences might shape interpretation. These discussions helped mitigate potential limitations of single coding and promoted consistency and rigor. MR also used analytic memos and diagramming to track decision-making and consider alternative interpretations. Together, these practices provided a systematic reflexive process that helped ensure interpretations remained grounded in the data.

Coding followed an abductive thematic analysis [27] workflow that combined the MRC process evaluation framework [26] with inductive code generation. Transcripts were first read repeatedly to support familiarization. A deductive coding structure based on MRC domains (context, implementation, engagement, mechanisms of impact) [26] served as the overarching structure, and data were coded line-by-line according to the relevant categories. Alongside this, inductive codes were generated to capture participant-led insights that did not fit the MRC framework. The coding framework was refined iteratively, with codes merged, repositioned, or expanded as patterns became clearer. Coded data were then reviewed within and across domains to identify candidate themes, with the MRC structure providing the overarching scaffold and inductive codes supplying the detailed content. Themes were developed through team discussion, diagramming, and matrix mapping to examine relationships across settings and between patient and HCP accounts. Reporting of qualitative methods and findings was informed by the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist (Checklist 1). This iterative movement between empirical data and implementation theory supported the identification of mechanisms, such as how delivery approaches shape engagement trajectories, that are central to understanding implementation and align with calls for qualitative methods [28].

PPI

PPI activities were integrated throughout the PROPEL study and used to refine qualitative findings. The PPI group for the work outlined here consisted of Windrush Generation Elders (pioneering individuals who immigrated to the United Kingdom from Caribbean countries and other Commonwealth nations between 1948 and 1971). They all had experience of living with or caring for someone with asthma or COPD. This group highlighted the importance of intersectionality before data collection and therefore supported this focus. One-to-one sessions were held with stakeholders (professionals based in the NHS and at the digital provider) who were experienced in delivering *myCOPD*. Work with the community-based group adopted a participatory approach to activities (facilitated by MR). These activities were designed to assess the relevance, clarity, and real-world applicability of emerging themes. Feedback informed the refinement and interpretation of themes.

The PPI group was not told the results at the beginning of the workshop (Multimedia Appendix 2 provides further detail using GRIPP2 [Guidance for Reporting Involvement of Patients and the Public, version 2] form); they were asked to individually draw journey maps of their experiences with COPD (in some cases, asthma). We then discussed their maps, with the chance for them to respond to each other and adapt their own maps. Finally, the results were introduced, and we discussed factors that overlapped with the results and those that were different, reasons for differences, and the importance of each factor. Consent was not obtained for PPI activities, and therefore these do not feature as a facet in the "Results" section. PPI highlighted that the results were relevant, how they related to lived experience, and a few aspects that were missing from the results (eg, the role of carers).

XSL•FO
RenderX

Ethical Considerations

Ethical approval was granted by the NHS Research Ethics Committee (REC reference: 23/SC/0155) and the University of Southampton (ERGO: 84797). All participants provided informed consent. Clinical and usage data were anonymized or deidentified before analysis. Confidentiality was maintained throughout, and participants received financial compensation for their time.

The qualitative interview data generated and analyzed during this study are not publicly available due to confidentiality agreements and the absence of participant consent for data sharing.

Results

Descriptive Data

Thirty interviews were conducted, 16 with patients and 14 with HCPs. Patient interviews lasted 14 to 35 (mean 23.7, SD 6.3) minutes, and HCP interviews lasted 23 to 57 (mean 36.8, SD 7.9) minutes. Sixteen of the 37 patients who initially agreed completed interviews; there was a range of reasons for nonparticipation: no response to invitations (n=15), unavailable contact details (n=2), technical failures (n=2), or withdrawal (n=2). Patient characteristics are provided in Table 1, and HCP characteristics are provided in Table 2. Participant usage data (Table 3) indicated higher engagement among Cornwall patients, who had a longer period since activation (107 - 148 days) and more frequent logins. Across both sites, 81.25% (13/16) accessed educational modules and 68.75% (11/16) accessed PR content.

Table . Patient sample characteristics.

Site	Number, n	Age (y), mean (SD)	COPD[a] severity, n			Females, n	White patients, n	Years since diagnosis, mean (SD)	Long-term conditions, n				Long-term conditions, mean (SD)	Lives alone, n	Smokers, n
			Mild	Moderate	Severe				0	1	2	3+			
Bristol	8	68.38 (5.32)	3	2	3	3	8	8.63 (5.13)	1	1	2	4	2.38 (1.41)	3	2
Cornwall	8	66.50 (7.95)	5	3	0	3	8	8.38 (8.11)	1	2	1	4	3.25 (2.76)	1	3
Total	16	67.44 (6.60)	8	7	3	6	16	8.50 (6.55)	2	3	3	8	2.81 (2.17)	4	5

[a]COPD: chronic obstructive pulmonary disease.

Table . Summary of HCPs[a] interviewed[b].

Site	Number, n	White HCPs, n	Digital health champion, n	Nurse or physiotherapist, n
Bristol	8	8	3	5
Cornwall	6	5	1	5
Total	14	13	4	10

[a]HCPs: health care professionals.

[b]One digital health champion also worked as a senior physiotherapist (Cornwall), and another had previously worked as a nurse (Bristol).

Table . Summary of usage data.

Site	Number, n	CAT[a] score at baseline, mean (SD)	Days since activation, mean (SD)	Number of login days, mean (SD)	Patients accessing education modules, n (%)	Education completion, mean (SD)	Patients accessing PR[b] modules, n (%)	PR modules accessed, mean (SD)
Bristol	8	18.29 (6.78)	168.75 (88.06)	49 (69.10)	6 (75.0)	51.6 (44.5)	5 (62.5)	2.8 (1.66)
Cornwall	8	18.09 (8.12)	148.63 (48.29)	57.88 (41.26)	7 (87.5)	58.6 (41.0)	6 (75)	2.5 (1.72)
Total	16	18.19 (7.22)	158.69 (69.39)	53.44 (55.17)	13 (81.3)	52.4 (41.3)	11 (68.75)	2.65 (1.28)

[a]The COPD Assessment Test (CAT) is designed to measure the impact of chronic obstructive pulmonary disease (COPD) on a person's life and assess changes over time: 0 - 9: low impact; 10 - 20: medium impact; 21 - 30: high impact; and 31 - 40: very high impact.

[b]PR: pulmonary rehabilitation.

XSL•FO
RenderX

Overview of Thematic Analysis

Five themes were developed, each with subthemes (Table 4). Read sequentially, the themes describe a pathway from patient characteristics to delivery context, engagement patterns, perceived barriers, and perceived benefits. This structure highlights how individual and structural factors interact to shape engagement with *myCOPD*. As mentioned in the "Methods" section, the analysis of the transcripts was often complementary between HCPs and patients. For example, both HCPs and patients could describe the health care setting as affecting engagement with *myCOPD*, though HCPs typically provided more detail.

Table . Summary of themes and subthemes.

Theme	Subthemes	Interplay between HCP[a] and patient accounts
Patient characteristics related to engagement	• Perceived digital ability • Negotiating multimorbidity • Socioeconomic status	• Complementary accounts, patients provided more detail
Important differences in health care setting	• Perceived purpose of health service • Organizational-structural factors • Approaches to delivery of *myCOPD*	• Complementary accounts, HCPs provided more detail
Engagement with *myCOPD*	• Single-aspect engagement • Multifaceted engagement	• Complementary accounts, patients provided more detail

Perceived barriers to engagement	• User perceptions of *myCOPD*
	• External barriers
		• Complementary accounts
		  • Users provided more detail
		  • HCPs provided more detail

Perceived benefits of engagement	• Targeted behavior change • Wider lifestyle adjustments	• Complementary accounts, patients provided more detail

[a]HCP: health care professional.

Patient Characteristics Related to Engagement

Three characteristics (multimorbidity, perceived digital ability, and socioeconomic context) operated as "layers of vulnerability" influencing engagement. Importantly, it was not the presence of a characteristic alone but how patients interpreted and negotiated that characteristic that shaped their engagement trajectory.

Negotiating Multimorbidity

The impact of living with multiple long-term conditions was commonly discussed. HCPs and patients often recognized the uncertainty caused by negotiating multimorbidity.

> *It's very difficult, when you've got another illness, you can't really put it all on one thing.* [PAT06, hospital patient]

Other health conditions potentially inhibit engagement with *myCOPD*. For example, positive management of another condition could affect quality of life to such an extent that COPD was not currently prioritized (eg, weight loss). In other instances, the presentation of other health conditions seemed to compete for prioritization.

> *I've got asthma, COPD, the heart problem, and I'm being treated for bladder cancer. Unfortunately, it's [COPD is] like number three [in terms of priorities].* [PAT03, hospital patient]

These accounts illustrate how multimorbidity among patients formed a core layer of vulnerability, shaping whether COPD self-management could be prioritized and how patients navigated competing health demands.

Perceived Digital Ability

Patients' perceptions of their own digital ability also impacted engagement with *myCOPD*. Some participants expressed a level of digital inhibition that prevented engagement. Patients with conditions such as dyslexia could also perceive engagement as overwhelming.

> *I'm afraid of it...I don't know what I'm doing half the time. I'll get confused. I don't want to put the wrong answer...it worries me.* [PAT05, community patient]

A common contradiction was observed among patients between their perceived digital abilities and their reported digital capabilities. The following quote is from a patient who described their use of digital technology as "not good" and "very limited."

> *[I have] an iPhone 13 mini...I go on Internet, social media occasionally...check football scores...I use Google quite a lot. I'll check the weather and watch videos...Occasionally I'll check the news. [And I use] WhatsApp...I use the NHS app to order prescriptions...I find that easier to do [with] the NHS app, probably easier than trying to get to see someone.* [PAT08, hospital patient]

Perceived digital ability operated as another layer of vulnerability, influencing not only whether patients felt able to use *myCOPD* but also how confidently they interpreted and acted on its content.

Socioeconomic Status

Socioeconomic status (SES) was rarely discussed directly, and a marker of SES was not collected as part of the PROPEL study,

XSL•FO
RenderX

but there were instances where economic resources appeared to influence engagement with *myCOPD*. The following extract highlights a participant of high SES who was less invested in the app because they were able to access support for self-management from a private health care team.

> *I couldn't keep waiting for NHS responses, so I went privately to see a respiratory consultant, and he put me onto an ENT consultant who is very good. And they also put me onto a respiratory physio, and she was very good. And I've obviously taken hints from them of what to do. And yeah, I think I've got most things under control fairly well.* [PAT09, community patient]

Socioeconomic context, specifically having the financial ability to seek private health services, presents a potential layer of vulnerability. Such differences in financial resources may influence the perceived value of engaging with *myCOPD*.

Important Differences in the Health Care Setting

Approaches to Delivering myCOPD

This theme highlights factors relevant to how HCPs approached the delivery of *myCOPD*. Two subthemes (the perceived purpose of the health service and organizational-structural factors) appeared to influence the third subtheme (approaches to the delivery of *myCOPD*). Given the distinct approaches to delivering *myCOPD*, the health care setting is conceptualized as a layer of vulnerability that can hinder or promote engagement with *myCOPD*. These approaches also influenced how clinical teams used DHCs.

Community Setting

Among community professionals, the general perception was that the purpose of their service was to provide and improve COPD self-management regimens over several weeks. Community professionals often described discussing relevant self-management strategies with patients and using *myCOPD* as an initial step toward engagement.

> *[We] try and find out which is the best route initially [for the patient to] start engaging with the app...I suggest looking at chest clearance, or if they're not taking inhalers properly then we'd look at the inhaler technique...trying to find something that they engage with well and get them to concentrate on that initially.* [HCP12, community-based HCP]

Between contacts, community professionals described checking patient usage reports in *myCOPD*. At in-person visits, professionals would check patient responses to, and understanding of, accessed content and then develop self-management strategies by signposting patients to further relevant content. In this context, DHCs were used for brief check-ins with patients and to address technical issues. This proactive approach was often described as a clear and purposeful process that used *myCOPD* as a medium to demonstrate and discuss tailored self-management techniques.

> *I do like [myCOPD] because you can see how well the patients are doing with that symptom tracker...you can have a look at your patients and say "oh, he's not good today. It might be worth giving him a phone call," [to ask the patient] "have you spoken to a GP, have you taken action on this?"* [HCP11, community-based HCP]

Hospital Setting

In contrast, hospital professionals often described a service that focused on recovery from exacerbation in preparation for discharge. Organizational-structural factors appeared to influence the approach to delivering *myCOPD*. For example, hospital-based professionals perceived internet connectivity, understaffing, and patient expectations (of being discharged) as adversely affecting how *myCOPD* was delivered.

> *There's quite a big workload for my colleagues, [it's] quite a big hospital. If there was maybe more employees to go through the app...but I don't know if we've got enough resources...to thoroughly implement it.* [HCP7, hospital-based HCP]

These contextual issues induced an approach to delivery that was passive, fundamentally reducing engagement to onboarding. This was commonly expressed as "selling" *myCOPD*. This passive approach seemed to be an embedded process in which clinical professionals avoided discussing the app with patients; anything related to the app was deemed the work of the DHC.

> *We ask the Band 4 [DHC] to see the patients for the myCOPD app, you know introduce [it and] phone them again afterwards.* [HCP1, hospital-based HCP]

Engagement With *myCOPD*

This theme describes how patients perceived engagement with *myCOPD*, a perception seemingly affected by contextual factors described in previous themes. Two subthemes are used to describe different types of engagement. The contrast in engagement across the 2 sites highlights that engagement itself is a layer of vulnerability rather than a neutral process, one that does not rely solely on design and individual capacity.

Single-Aspect Engagement

Single-aspect engagement describes a focused type of engagement, characterized by accessing specific parts of the app, as if they were stand-alone components (eg, breathing exercises, inhaler technique, and physical activity). Single-aspect engagement was often described in terms that were transactional and short-term, patients expected to achieve clear benefits from use. Both sets of patients, community and hospital, described single-aspect engagement.

> *[Regarding inhaler technique] I didn't realize I was doing so much wrong, and it does make a difference if you do it properly...[also] some of the techniques for getting your breath back [I] didn't know anything about that. Again, that's something new that I've learnt.* [PAT06, hospital patient]

Multifaceted Engagement

Multifaceted engagement describes a more comprehensive use of *myCOPD*, often characterized by a sense of purpose and routine. Single-aspect engagement remains present, but usage does not need to fulfill a specific function.

XSL•FO
RenderX

> *It's a normal daily thing. You know, I do it in the morning to record when I take my medicines and steps, and I do it in the evening to record any exercise I've done.* [PAT04, community patient]

Multifaceted engagement seemed to stem from proactive communication and management between patients and professionals. Hospital HCPs tended to check and prompt usage, whereas community HCPs tended to check and prompt meaningful engagement. Here, interviews with HCPs help to understand the different types of engagement described by patients.

> *I see if they're actually using it [myCOPD], and if not, I do tend to send them a little message...You know, "Please remember to use this" or "I have noticed you haven't activated it" or "I've noticed you haven't used it in a couple of weeks."* [HCP13, hospital HCP]

> *I'll use the app to check how the patients doing and then motivate them on these follow up calls, to motivate them how to continuously use [myCOPD] because the whole point is to make sure they're getting that PR support after an exacerbation. So, to ensure they don't get unwell and really focus.* [HCP2, community-based HCP]

Perceived Barriers to Engagement

Overview

Perceived barriers to engagement seem to be affected by the previous themes and are categorized into 2 subthemes. User perceptions of *myCOPD* describe gaps between the actual provision of *myCOPD* and users' understanding of that provision. External barriers are perceived issues that are beyond the direct remit of *myCOPD* (eg, fit with in-person care and the degree to which *myCOPD* is embedded within services). Misunderstandings about app functionality and uncertainty about data fidelity interact with existing layers of vulnerability, shaping how confidently patients and clinicians rely on the tool.

User Perceptions of myCOPD: App Functionality

While many participants (both patients and HCPs) described *myCOPD* as easy to use, some participants made initial judgments based on whether they perceived the app as intuitive. For example, some participants passed judgment based on their understanding of a specific function being absent (eg, accessibility functions) and not realizing that the desired functionality was offered.

> *Patients who cannot read, [there is a] low reading age in Cornwall, or [have] eyesight problems, or have dyslexia may struggle as the questionnaires and tabs are all word-based. Perhaps a speaker icon with the questions and tabs narrated would benefit patients.* [HCP11, community HCP]

This perceived functionality was described by some participants when navigating *myCOPD*. These perceptions could have been overcome through discussion with technical support. Similarly, some patients described burdensome expectations to complete "lengthy" physical exercise sessions, even though *myCOPD* typically builds tolerance, ranging from 8 to 40 minutes, depending on user preferences.

> *I've just been through the exercises and it's an hour's worth of exercises...the whole bundle of exercises and watching all the videos and doing the exercise is close on an hour.* [PAT09, community patient]

User Perceptions of myCOPD: Trust in Patient Usage Reports

Community-based patients and professionals articulated the importance of building trust in patient self-reports provided in *myCOPD*. This trust would encourage professionals to tailor treatment effectively. Simply providing reports without an explanation of their underlying fidelity left room for suspicion. For example, the Activity Diary in *myCOPD* has a calendar option that reports the number of minutes and the Rating of Perceived Exertion achieved during exercise, a design aimed at increasing fidelity.

> *[Usage reports] tell you that patients have done their PR, but actually they opened the tile, looked at the first page and then ticked that they've done all the exercise. You get that clinician's sixth sense, that they might not have actually done the exercise.* [HCP9, community-based HCP]

External Barriers: Contrasting Perceptions of Fit With In-Person Care

Patients did not expect HCPs to check their *myCOPD* activity but also felt unable to increase their engagement in conversations during face-to-face consultations. As outlined above, community professionals frequently described using *myCOPD* to better target their professional expertise. In contrast, hospital professionals often preferred in-person care to using *myCOPD*.

> *We see them after the acute phase, they've had an exacerbation, and they're in hospital and I suppose generally that's always gonna be better...It's better to see anyone face-to-face.* [HCP7, hospital-based HCP]

External Barriers: Embedding myCOPD Into Routine Practice

Professionals frequently mentioned that *myCOPD* was not embedded enough within services. Embedding sometimes meant having a critical mass of users to be seen as routine practice; at other times, it meant buy-in from frontline staff and managers.

> *I think it needs to be enforced by the whole team whereby we say [to patients], "look, when you're going home, we need to see this information and it's gonna be used in the clinic." I think then...adherence would be better.* [HCP6, hospital-based HCP]

Community HCPs were proactively using *myCOPD* but also highlighted the need to provide a dedicated space to support patients with engagement during existing face-to-face consultations.

> *If it [myCOPD] was part of pulmonary rehab and we did make it more approachable, [it would] make them*

XSL•FO
RenderX

> *[patients] feel like they're all doing it together, they're all in the same boat [as COPD patients]...a bit of support over the next few weeks, then you'll be good to run with it.* [HCP3, community-based HCP]

Where *myCOPD* was not well integrated into routine care, organizational constraints compounded existing layers of vulnerability, limiting opportunities for patients to develop sustained engagement.

Perceived Benefits of Engagement

The perceived benefits of engagement with *myCOPD* appear to be affected by the first 3 themes. The first subtheme captures direct benefits associated with *myCOPD*. For HCPs, this meant positive alignment with existing care guidelines; for patients, this meant that the app was a tangible resource for improving specific abilities. The second subtheme describes wider benefits of *myCOPD*.

Targeted Behavior Change

Targeted behavior change describes how *single-aspect engagement* was perceived to improve specific abilities (eg, breathing through an exacerbation, improved inhaler technique, and increased physical activity). This association often led to descriptions of improved self-confidence. Overall, patients described being able to use *myCOPD* to address and improve specific behaviors that concerned them.

> *It's [chest clearance on myCOPD] helped...that's one thing, you can hear it now. I'll get like a frog in the throat and in the chest...everybody thinks I've got cough or cold all the time, but hang on, just got to clean myself [performs chest clearance].* [PAT03, hospital patient]

> *[I've got] more confidence, really, confidence in managing my condition...When I used to walk uphill, to do the horses, [I] used to stop to take a breath or take my inhaler and now I just keep going but I slow down...I realise that I don't have to stop, and I don't have to rush...[I'm] kind of not worried about losing my breath because I can control it.* [PAT07, community patient]

Wider Lifestyle Adjustments

This subtheme describes benefits associated with *multifaceted engagement*. This seemed to be a longer-term type of engagement with benefits yet to materialize. This type of benefit was only associated with patients from community settings and was typically developed through a latent process of analysis. Short transcript extracts do not adequately explain this subtheme; however, there was one example that demonstrated the potential benefits of wider lifestyle adjustments.

One patient noticed increasing breathlessness over several weeks and did not understand why. The patient eventually accessed the reporting functions in *myCOPD* and, because they had been recording exercises and symptoms daily, were able to spot a pattern in their breathlessness. Consequently, they were able to deduce that the breathlessness was a result of chopping wood and adjusted their behavior to avoid symptoms by taking a preventative inhaler before chopping wood or delaying the activity until later in the day.

Discussion

Principal Findings

This study explored how *myCOPD* was delivered and used across 2 contrasting NHS settings and how contextual, organizational, and individual factors shaped engagement. Across both sites, patients described how multimorbidity, perceived digital ability, and socioeconomic context influenced their capacity to engage with the intervention. These characteristics did not operate in isolation; rather, they interacted with local delivery practices to shape engagement trajectories.

A key finding was the marked difference in delivery approaches between settings. Community PR HCPs adopted a proactive, iterative model of support using *myCOPD* to structure conversations, reinforce self-management strategies, and monitor progress. In contrast, hospital HCPs described a more passive, onboarding-focused approach shaped by workflow pressures, limited staffing, and the prioritization of discharge. These delivery styles appeared to influence whether patients engaged in single-aspect engagement (targeted, short-term usage) or multifaceted engagement (routine usage).

Both HCPs and patients identified barriers and facilitators to engagement. Barriers included usability concerns, competing health priorities, limited digital confidence, and challenges integrating the app within in-person care. Facilitators included personalized support from clinicians, clear explanations of app features, and opportunities to discuss content during follow-up contacts. PPI contributors confirmed the relevance of these findings, emphasizing the importance of timely support, clear communication about the purpose of the app, and recognition of the challenges posed by multimorbidity and limited digital confidence.

A key strength of this study is the inclusion of both patients and HCPs across 2 contrasting NHS pathways, enabling a nuanced understanding of how delivery context shapes engagement. The use of abductive thematic analysis allowed us to integrate theoretical insights with inductively generated themes, and PPI involvement strengthened the relevance and clarity of the findings. Collectively, these findings highlight that engagement with *myCOPD* is not simply a function of individual motivation or app design but emerges from the interaction between patient characteristics and the delivery environment, an insight that aligns with and extends implementation theories emphasizing context-mechanism interactions [27].

Comparison With Prior Work

Our findings align with implementation research demonstrating that digital interventions are not inherently effective; their impact depends on how they are introduced, supported, and embedded within clinical workflows [11,26]. Prior COPD digital evaluations [22-25] similarly reported high initial interest in inhaler videos and symptom tracking but noted rapid declines in use. Our findings extend this work by showing why this decline occurs, specifically, that passive onboarding in hospital settings limits opportunities for relational reinforcement.

XSL•FO
RenderX

More specifically, our observation that proactive, relationship-based support in community PR settings facilitated sustained engagement echoes the work of Pinnock et al [5,6] on supported self-management and aligns with the theorization of implementation mechanisms by Lewis et al [28], which emphasizes the importance of relational reinforcement and iterative feedback loops. Our data provide concrete examples of how these mechanisms operate within COPD pathways, illustrating how clinicians used *myCOPD* to scaffold conversations, tailor recommendations, and maintain continuity, practices that have been theorized but rarely described empirically in COPD digital implementation studies.

Conversely, the passive onboarding approach observed in hospital settings reflects challenges documented in acute care environments, where digital tools are often introduced during periods of high patient burden and limited staff capacity [10,14]. This is consistent with the findings of O'Connor et al [29], which indicate that workflow pressures and resource constraints in acute care limit opportunities for relational support and follow-up. Our findings extend this literature by showing how these constraints shape not only uptake but also the depth and trajectory of digital engagement, with hospital patients more likely to adopt single-aspect, short-term use.

Our findings also contribute to emerging evidence on digital inequalities. Multimorbidity, perceived digital ability, and socioeconomic context shaped how participants navigated *myCOPD*, consistent with the literature highlighting the intersection of health and digital disparities [13-16]. Informed by the concept of "layers of vulnerability" proposed by Luna [19,20], our analysis suggests that these vulnerabilities are dynamic and relational, shaped not only by individual circumstances but also by the support available within clinical settings. This was evident in patients who described shifting their self-management priorities in response to exacerbations or competing long-term conditions, relying on family members or clinicians to navigate digital tasks when confidence fluctuated, and adjusting their engagement depending on whether services provided opportunities for discussion, troubleshooting, or reinforcement.

In doing so, our study provides empirical grounding for the conceptualization of vulnerability as layered rather than categorical, as proposed by Luna [19,20], demonstrating how multimorbidity, digital confidence, and socioeconomic constraints accumulate and interact within real-world COPD pathways. This complements the call for intersectional approaches to digital health disparity research made by Husain et al [11] and aligns with the argument of Veinot et al [30] that digital inequalities are structurally produced rather than individually determined. Our findings also resonate with the findings of Greenhalgh et al [31], specifically the finding that multimorbidity often complicates digital adoption, particularly when tools are not embedded within supportive relational contexts.

Limitations

Although this study offers important insights into how delivery context shapes engagement with *myCOPD*, several limitations should be acknowledged when interpreting the findings. First, participants were recruited from within the PROPEL study cohort, which may limit transferability to settings where *myCOPD* is introduced outside of a research context. Second, interviews were conducted remotely, which may have influenced participation among individuals with limited digital access. Third, while we examined differences across 2 settings, the sample size did not allow for detailed subgroup analysis (eg, by digital literacy, severity, or SES). Finally, usage data provided by my mhealth Ltd offered helpful context but did not capture qualitative nuances of engagement. These limitations underscore the need for future work that examines delivery models across a wider range of organizational contexts and includes more granular measures of digital literacy and multimorbidity burden, as recommended by Venoit et al [30].

Implications for Practice and Implementation

Our findings highlight the importance of aligning digital self-management interventions with local workflows and capacities. Proactive, iterative support (such as checking usage data, discussing content during follow-up, and tailoring recommendations) appears to facilitate more meaningful engagement. Services adopting *myCOPD* may benefit from clarifying roles (eg, between clinicians and DHCs), ensuring adequate staffing, and embedding digital support within routine contacts. These implications arise directly from the contrast observed in our study: PR clinicians were able to provide iterative support because roles were clear and time was protected, whereas hospital clinicians lacked these structural conditions. More specifically, our data suggest that proactive delivery models may require explicit resourcing, including protected time for clinicians to review usage data and integrate digital discussions into routine care. This aligns with the argument of Greenhalgh et al [32] that digital health implementation depends on relational work and organizational readiness. Hospital settings, where workflow pressures limit such opportunities, may benefit from dedicated DHCs or structured onboarding protocols that extend beyond discharge.

At the patient level, recognizing the influence of multimorbidity, digital confidence, and socioeconomic context is essential for equitable implementation. Tailored onboarding, opportunities for hands-on support, and clear communication about the purpose and benefits of the app may help address barriers. These findings reinforce the value of attending to layers of vulnerability when designing and delivering digital self-management support. In practice, this may involve prioritizing hands-on demonstrations for patients with low digital confidence, offering multimorbidity-sensitive guidance that helps patients identify which app features align with their immediate priorities and ensuring that digital support is available beyond initial onboarding. These approaches are consistent with the suggestions of Vorrink et al [33] to improve digital self-management interventions.

For policy and commissioning, these findings suggest that digital interventions should not be implemented as stand-alone tools but as components of broader self-management pathways that account for organizational readiness and patient diversity. Commissioners may therefore need to incorporate workflow assessments, staffing considerations, and digital literacy support

XSL•FO
RenderX

into procurement and implementation plans, ensuring that digital tools are embedded within coherent, adequately resourced pathways rather than added as isolated components. This aligns with Normalization Process Theory as outlined by Murray et al [34], which emphasizes the importance of integration within existing workflows.

Implications for Research

Future research should examine how different models of delivery influence long-term engagement and clinical outcomes, including whether proactive support can be feasibly scaled. Further work is also needed to explore how digital interventions can be tailored for individuals with multimorbidity, low digital confidence, or limited access to technology. Mixed methods evaluations that integrate qualitative insights with detailed usage analytics may help identify mechanisms of engagement and inform more equitable implementation strategies.

Our findings point to several specific avenues for investigation, including comparative studies of proactive versus passive delivery models, evaluations of DHC roles within COPD pathways, and intersectional analyses that examine how layers of vulnerability shape engagement trajectories over time. Integrating relational measures of support with usage analytics may be particularly valuable for identifying mechanisms that drive sustained engagement, as suggested by the work of Lewis et al [28] on implementation mechanisms.

Conclusions

Delivery context and local workflows play a central role in shaping how patients engage with *myCOPD*. Proactive, relationship-based support appears to facilitate more integrated and sustained engagement, whereas passive onboarding may limit the intervention's potential. Considering patient characteristics, particularly multimorbidity and perceived digital ability, alongside organizational capacities may support more equitable and effective implementation of digital self-management tools in COPD care. The concept of layers of vulnerability provides a useful lens for understanding how individual and structural factors combine to shape engagement trajectories.

By illustrating how delivery models interact with patient characteristics to shape engagement, this study contributes to a more nuanced understanding of digital self-management implementation and highlights the need for context-sensitive, equity-oriented approaches to digital health in COPD care.

Acknowledgments

We would like to thank all individuals who participated in patient and public involvement activities for this study. We would also like to thank the clinical and research support provided by Cornwall Partnership NHS Foundation Trust, North Bristol NHS Trust, and University Hospitals Bristol & Weston NHS Foundation Trust. No generative AI was used in the preparation of this manuscript. We would also like to acknowledge our research collaborator Unity Insights Limited for running statistical analysis.

Funding

MR is completing a PhD jointly funded by the National Institute for Health and Care Research Southampton Biomedical Research Centre and my mhealth Ltd. This work was commissioned and funded by SBRI Healthcare. SBRI Healthcare is an Accelerated Access Collaborative (AAC) initiative, in partnership with the Academic Health Science Networks (AHSNs). The views expressed in the publication are those of the author(s) and not necessarily those of SBRI Healthcare or its stakeholders.

Data Availability

The qualitative interview data generated and analyzed during this study are not publicly available due to confidentiality agreements and the lack of participant consent for data sharing.

Authors' Contributions

MR conceived and conducted the data analysis, collected the data, and drafted and revised the manuscript. BA conceptualized the overarching study, provided supervision, and commented on manuscript drafts. KB provided supervision and commented on manuscript drafts. TW secured funding for the project, conceptualized the overarching study, provided supervision, and commented on manuscript drafts. LY provided supervision and commented on manuscript drafts. AB commented on manuscript drafts and was responsible for delivering the wider PROPEL (A Pragmatic Real-World Multicenter Observational Research Study to Explore the Clinical and Health Economic Impact of *myCOPD*) study. BC conducted initial patient and public involvement work and contributed to data collection. JD and RW served as site leads and commented on manuscript drafts.

Conflicts of Interest

MR completed this work as part of a PhD jointly funded by the Southampton National Institute for Health and Care Research Biomedical Research Centre (NIHR BRC) and my mhealth Ltd. AB is employed by my mhealth Ltd. TW is the cofounder, shareholder, and director of my mhealth Ltd. The views expressed are those of the authors and not necessarily those of the Southampton NIHR BRC or my mhealth Ltd. BC was a research fellow on this study, based at University of Southampton, and moved to the University of Westminster. All other authors declared no conflicts of interest.

XSL•FO
RenderX

Multimedia Appendix 1
Topic guide.
[DOCX File, 28 KB - humanfactors_v13i1e88108_app1.docx]

Multimedia Appendix 2
GRIPP2 form.
[DOCX File, 19 KB - humanfactors_v13i1e88108_app2.docx]

Checklist 1
COREQ checklist.
[PDF File, 493 KB - humanfactors_v13i1e88108_app3.pdf]

References

1. Bloom CI, Slaich B, Morales DR, Smeeth L, Stone P, Quint JK. Low uptake of palliative care for COPD patients within primary care in the UK. Eur Respir J 2018 Feb;51(2):1701879. [doi: 10.1183/13993003.01879-2017] [Medline: 29444916]
2. Whittaker H, Rothnie KJ, Quint JK. Cause-specific mortality in COPD subpopulations: a cohort study of 339 647 people in England. Thorax 2024 Feb 15;79(3):202-208. [doi: 10.1136/thorax-2022-219320] [Medline: 37328279]
3. Hurst JR, Quint JK, Stone RA, Silove Y, Youde J, Roberts CM. National clinical audit for hospitalised exacerbations of COPD. ERJ Open Res 2020 Jul;6(3):00208-02020. [doi: 10.1183/23120541.00208-2020] [Medline: 32984418]
4. Whittaker H, Rubino A, Müllerová H, et al. Frequency and severity of exacerbations of COPD associated with future risk of exacerbations and mortality: a UK routine health care data study. Int J Chron Obstruct Pulmon Dis 2022;17(null):427-437. [doi: 10.2147/COPD.S346591] [Medline: 35264849]
5. Pinnock H, Epiphaniou E, Pearce G, et al. Implementing supported self-management for asthma: a systematic review and suggested hierarchy of evidence of implementation studies. BMC Med 2015 Jun 1;13(1):127. [doi: 10.1186/s12916-015-0361-0] [Medline: 26032941]
6. Pinnock H, Parke HL, Panagioti M, et al. Systematic meta-review of supported self-management for asthma: a healthcare perspective. BMC Med 2017 Mar 17;15(1):64. [doi: 10.1186/s12916-017-0823-7] [Medline: 28302126]
7. Breathing well: an assessment of respiratory care in England and Wales. : Royal College of Physicians; 2024 URL: https://www.rcp.ac.uk/media/afbjck5p/nrap_breathing-well-report-final-version-2.pdf [accessed 2026-09-01]
8. Bosnic-Anticevich S, Bakerly ND, Chrystyn H, Hew M, van der Palen J. Advancing digital solutions to overcome longstanding barriers in asthma and COPD management. Patient Prefer Adherence 2023;17:259-272. [doi: 10.2147/PPA.S385857] [Medline: 36741814]
9. Scott IA, Scuffham P, Gupta D, Harch TM, Borchi J, Richards B. Going digital: a narrative overview of the effects, quality and utility of mobile apps in chronic disease self-management. Aust Health Rev 2020 Feb;44(1):62-82. [doi: 10.1071/AH18064] [Medline: 30419185]
10. Shaw G, Whelan ME, Armitage LC, Roberts N, Farmer AJ. Are COPD self-management mobile applications effective? A systematic review and meta-analysis. NPJ Prim Care Respir Med 2020 Apr 1;30(1):11. [doi: 10.1038/s41533-020-0167-1] [Medline: 32238810]
11. Husain L, Greenhalgh T, Hughes G, Finlay T, Wherton J. Desperately seeking intersectionality in digital health disparity research: narrative review to inform a richer theorization of multiple disadvantage. J Med Internet Res 2022 Dec 7;24(12):e42358. [doi: 10.2196/42358] [Medline: 36383632]
12. Eiselt AK, Kirkendall S, Xiong E, Langner D, Goldfarb M. Achieving clinically meaningful outcomes in digital health: a six-step, cyclical precision engagement framework (ENGAGE). Front Digit Health 2025;7:1713334. [doi: 10.3389/fdgth.2025.1713334] [Medline: 41608160]
13. Shatto JA, Stickland MK, Soril LJJ. Variations in COPD health care access and outcomes: a rapid review. Chronic Obstr Pulm Dis 2024 Mar 26;11(2):229-246. [doi: 10.15326/jcopdf.2023.0441] [Medline: 38241509]
14. Dunn J, Coravos A, Fanarjian M, Ginsburg GS, Steinhubl SR. Remote digital health technologies for improving the care of people with respiratory disorders. Lancet Digit Health 2024 Apr;6(4):e291-e298. [doi: 10.1016/S2589-7500(23)00248-0] [Medline: 38402128]
15. Hoffman DL, Novak TP, Schlosser A. The evolution of the digital divide: how gaps in internet access may impact electronic commerce. J Comput Mediat Commun 2000;5(3). [doi: 10.1111/j.1083-6101.2000.tb00341.x]
16. Internet access – households and individuals, Great Britain: 2018. : Office for National Statistics; 2019 URL: https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2018 [accessed 2026-09-01]
17. Figueroa CA, Luo T, Aguilera A, Lyles CR. The need for feminist intersectionality in digital health. Lancet Digit Health 2021 Aug;3(8):e526-e533. [doi: 10.1016/S2589-7500(21)00118-7] [Medline: 34325855]
18. Porroche-Escudero A, Popay J. The Health Inequalities Assessment Toolkit: supporting integration of equity into applied health research. J Public Health (Oxf) 2021 Sep 22;43(3):567-572. [doi: 10.1093/pubmed/fdaa047] [Medline: 32323722]

XSL•FO
RenderX

19. Luna F. Identifying and evaluating layers of vulnerability - a way forward. Dev World Bioeth 2019 Jun;19(2):86-95. [doi: 10.1111/dewb.12206] [Medline: 30058768]
20. Luna F. Elucidating the concept of vulnerability: layers not labels. Int J Fem Approaches Bioeth 2009 Mar;2(1):121-139. [doi: 10.3138/ijfab.2.1.121]
21. Evidence generation plan for digital technologies to support self-management of asthma: early-use assessment. : National Institute for Health and Care Excellence (NICE); 2026 URL: https://www.nice.org.uk/guidance/htg778/resources/evidence-generation-plan-for-digital-technologies-to-support-selfmanagement-of-asthma-earlyuse-assessment-pdf-20311922088133 [accessed 2026-09-01]
22. Bourne S, DeVos R, North M, et al. Online versus face-to-face pulmonary rehabilitation for patients with chronic obstructive pulmonary disease: randomised controlled trial. BMJ Open 2017 Jul 17;7(7):e014580. [doi: 10.1136/bmjopen-2016-014580] [Medline: 28716786]
23. Crooks MG, Elkes J, Storrar W, et al. Evidence generation for the clinical impact of myCOPD in patients with mild, moderate and newly diagnosed COPD: a randomised controlled trial. ERJ Open Res 2020 Oct;6(4):00460-2020. [doi: 10.1183/23120541.00460-2020] [Medline: 33263052]
24. Davies H, Chappell M, Wang Y, et al. myCOPD app for managing chronic obstructive pulmonary disease: a NICE medical technology guidance for a digital health technology. Appl Health Econ Health Policy 2023 Sep;21(5):689-700. [doi: 10.1007/s40258-023-00811-x] [Medline: 37246188]
25. North M, Bourne S, Green B, et al. A randomised controlled feasibility trial of E-health application supported care vs usual care after exacerbation of COPD: the RESCUE trial. NPJ Digit Med 2020;3(145):145. [doi: 10.1038/s41746-020-00347-7] [Medline: 33145441]
26. Moore GF, Audrey S, Barker M, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015 Mar 19;350(mar19 6):h1258-h1258. [doi: 10.1136/bmj.h1258] [Medline: 25791983]
27. May C, Finch T. Implementing, embedding, and integrating practices: an outline of normalization process theory. Sociology 2009 Jun;43(3):535-554. [doi: 10.1177/0038038509103208]
28. Lewis CC, Boyd MR, Walsh-Bailey C, et al. A systematic review of empirical studies examining mechanisms of implementation in health. Implement Sci 2020 Apr 16;15(1):21. [doi: 10.1186/s13012-020-00983-3] [Medline: 32299461]
29. O'Connor L, Wang B, Ye Z, et al. Evaluation of an integrated digital and mobile intervention for COPD exacerbation. medRxiv. Preprint posted online on Feb 14, 2025. [doi: 10.1101/2025.02.13.25322246] [Medline: 39990547]
30. Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018 Aug 1;25(8):1080-1088. [doi: 10.1093/jamia/ocy052] [Medline: 29788380]
31. Greenhalgh T, Wherton J, Papoutsi C, et al. Analysing the role of complexity in explaining the fortunes of technology programmes: empirical application of the NASSS framework. BMC Med 2018 May 14;16(1):66. [doi: 10.1186/s12916-018-1050-6] [Medline: 29754584]
32. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367. [doi: 10.2196/jmir.8775] [Medline: 29092808]
33. Vorrink S, Huisman C, Kort H, Troosters T, Lammers JW. Perceptions of patients with chronic obstructive pulmonary disease and their physiotherapists regarding the use of an eHealth intervention. JMIR Hum Factors 2017 Sep 19;4(3):e20. [doi: 10.2196/humanfactors.7196] [Medline: 28928110]
34. Murray E, Treweek S, Pope C, et al. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. BMC Med 2010 Oct 20;8(1):63. [doi: 10.1186/1741-7015-8-63] [Medline: 20961442]

Abbreviations

COPD: chronic obstructive pulmonary disease
COREQ: Consolidated Criteria for Reporting Qualitative Research
DHC: digital health champion
GRIPP2: Guidance for Reporting Involvement of Patients and the Public, version 2
HCP: health care professional
MRC: Medical Research Council
NHS: National Health Service
PPI: patient and public involvement
PR: pulmonary rehabilitation
PROPEL: A Pragmatic Real-World Multicenter Observational Research Study to Explore the Clinical and Health Economic Impact of *myCOPD*
SES: socioeconomic status

XSL•FO
RenderX

Edited by S Law; submitted 20.Nov.2025; peer-reviewed by C Bryant, T Yang; revised version received 21.Aug.2026; accepted 24.Aug.2026; published 21.Sep.2026.

Please cite as:
Ruddock M, Blythin A, Cliffe B, Yardley L, Dodd J, Williams R, Bradbury K, Wilkinson T, Ainsworth B
Understanding Delivery and Engagement With a Digital Self-Management Intervention for Chronic Obstructive Pulmonary Disease Across Two Clinical Settings: Qualitative Study
JMIR Hum Factors 2026;13:e88108
URL: https://humanfactors.jmir.org/2026/1/e88108
doi:10.2196/88108

© Martin Ruddock, Alison Blythin, Bethany Cliffe, Lucy Yardley, James Dodd, Rachel Williams, Katherine Bradbury, Tom Wilkinson, Ben Ainsworth. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 21.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Physicians' Complex Relationships With Health Information Technology and Burnout: Cross-Sectional Survey Study

Mavis Jones[1], PhD; Timothy Jason[1], MSc; Dara Liu[2], BSc; Jeannette Comeau[2], MD; Chandi Chandrasena[1], MD; Noni E MacDonald[2], MD; Janice E Graham[2], PhD

[1]OntarioMD, 150 Bloor Street, Toronto, ON, Canada

[2]Department of Pediatrics, Dalhousie University, 5849 University Avenue, Halifax, NS, Canada

Corresponding Author:
Janice E Graham, PhD
Department of Pediatrics, Dalhousie University, 5849 University Avenue, Halifax, NS, Canada

Abstract

Background: Health information technology (HIT), while designed to improve practice efficiency and patient care, can contribute to physician burnout when designed without clinical practice in mind. Administrative and clinical demands now require physicians to spend more time on HIT, a burden that contributes to burnout and affects time with patients.

Objective: This study had two objectives: (1) to examine how physician perceptions of specific HIT functions may alleviate or contribute to physician burnout within the differing health care landscapes of Ontario and Nova Scotia, with particular attention given to physician perceptions of HIT associated with administrative burdens, interoperability, and system integration, and (2) to explore physician experiences and perceptions of HIT as potential factors affecting physician burnout, with the goal of delivering findings (eg, a model) that could be applied to improve physicians' HIT experience.

Methods: We designed an exploratory mixed methods study that deployed a cross-sectional survey in 2 Canadian provinces: Ontario and Nova Scotia. Centralized clinician management software (Ontario) and medical association networks (Nova Scotia) were used to recruit from an estimated 35,341 (Ontario) and 2809 (Nova Scotia) physicians. The survey was distributed between February and April 2024. Nonphysician clinicians, clinic staff, and non-HIT users were excluded, resulting in 1245 Ontario and 136 Nova Scotia physician HIT-user respondents. For both the Ontario and Nova Scotia samples, descriptive analyses of quantitative survey items were compared, and open-text responses were subjected to qualitative coding for themes. Subgroup differences in HIT-related burnout were analyzed in the larger Ontario sample.

Results: Common experiences were apparent despite differences in the samples, provincial health systems, and available HIT. "Managing communications related to patient care" and "inputting data into your EMR" were among the top 3 administrative burdens. While "logging in and out of technology platforms" ranked higher in Nova Scotia as an administrative burden, the related theme of integration and interoperability was prominent in both samples. In the Ontario sample, perceptions of HIT, quality of support, and hours worked per week accounted for over half of the variation in self-reported burnout.

Conclusions: While physicians appreciate the advantages of HIT for patient care, they also experience an overwhelming administrative and documentation burden, as well as disjointedness across data platforms, which contributes to their burnout. Greater ongoing involvement by end users in the design and usability of these technologies, along with improved standardization and interoperability, would reduce these burdens while maintaining the benefits of digital health systems.

(JMIR Hum Factors 2026;13:e93266) doi:10.2196/93266

KEYWORDS

burnout; digital health; EMR; physician wellness; stress; health care workforce; administrative burden; electronic medical record

Introduction

Physician burnout remains a system-wide problem with consequences for physicians and health care delivery in Canada. The Canadian Medical Association's (CMA) 2025 National Physician Health Survey found that 46% of physicians (n=3310) reported symptoms of burnout [1] (high levels of at least one indicator of depersonalization or emotional exhaustion using the Maslach Burnout Inventory [2]). The CMA study connects burnout and related health outcomes to an escalation in administrative burden: over the 3 iterations of their survey—2017, 2021, and 2025—there has been a significant increase in the proportion of respondents reporting spending an excessive or moderately high amount of time on their electronic medical record (EMR) on administrative tasks outside of regular hours, from 36% to 49% in 2021 and 64% in 2025. Here and elsewhere, it has been found that the burden is particularly high in family medicine [3]. One study reported that this group spends

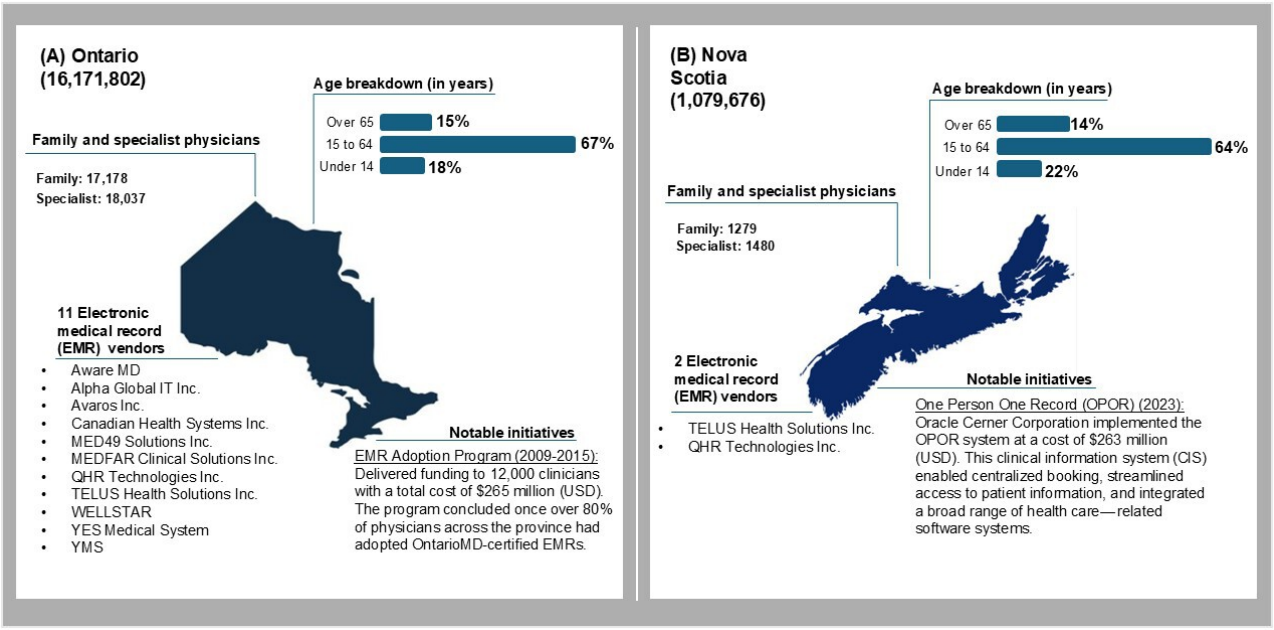
XSL•FO
RenderX

an average of 19 hours a week on administrative tasks [4]. This has an impact on access to primary care when family doctors experience administrative demands competing for patient-facing time and/or reduce (or leave) clinical practice due to burnout.

While workload, administrative burdens, and insufficient time with patients are neknown contributors to physician burnout [5,6], health information technology (HIT)—encompassing the electronic systems designed to collect, share, and manage patient information—has introduced new dynamics [7-9]. Paradoxically, the technological solutions intended to support a physician's daily work can also add to their workday burden, as they adapt to a tsunami of expensive, new, and largely unregulated HIT platforms while having a minimal role in their design and implementation [10,11].

In Canada, while federal initiatives shape the overall approach to health nationwide, provinces are responsible for (and fund) health care delivery. This leads to differences in models for care delivery that can affect physician workload—for example, in Ontario, a primary care physician in a Family Health Team may have access to more resources than a primary care physician working outside that model. Provinces are also responsible for decisions affecting the technologies clinicians use to manage patient data and to interact with others in the provincial health system, including the implementation of (provincial) system-wide initiatives. Our study focuses on 2 settings—the Canadian provinces of Ontario and Nova Scotia—to explore physicians' HIT experiences, including potential contributors to burnout in health care providers, their use of HIT, and notable HIT initiatives (Figure 1) [12-16].

Figure 1. Comparative context. (A) Ontario and (B) Nova Scotia.



While previously relying on 1 of 2 EMRs on the market in the province, Nova Scotia physicians only recently (December 6, 2025) began the implementation of One Person One Record (OPOR) [17], a clinical information system designed by Oracle Cerner to address centralized booking, access to patient information, and integration across institutions and software systems. In contrast, community-based Ontario physicians have had greater choices available to them via certification and adoption programs beginning in 2009 [18], allowing physicians to choose from a list of certified offerings meeting a minimum standard. Physicians working in both community and hospital settings rely on hospital information systems (commonly referred to as HIS) such as EPIC. Both provinces feature multiple platforms for transmitting reports from hospitals, laboratories, and diagnostic facilities to a patient's electronic record. In addition, physicians may have access to add-ons, only some of which may be integrated (ie, accessible directly from the EMR without having to log out and log in to a different platform with a different username and password).

The relative maturity of the HIT landscape in Ontario is a consequence of provincial government investment, which includes change management support via a team of field advisors and peer support (ie, physicians and other clinical staff) who connect physicians with resources, provide guidance, and offer other change management support for adopting and implementing new technologies (physicians are free to choose a noncertified EMR outside of this system). Government-funded programs are delivered by OntarioMD (OMD, a subsidiary of the Ontario Medical Association). The closest analog to OMD in Nova Scotia would be the E-Health division of the provincial medical association, Doctors Nova Scotia, which provides physician support, advice, and guidance on these technologies.

We designed a mixed methods study to explore physicians' perceptions of HIT and their relationship to their burnout. As described, Ontario and Nova Scotia are not directly comparable, given their differences in size, HIT infrastructure, and historical support for HIT adoption. We hypothesized that, despite these differences in maturity, we may see commonalities across Ontario and Nova Scotia physicians in terms of the benefit they derive from HIT, their expectations vs the reality of their experience, their usability challenges, and the ways in which the day-to-day world of clinical practice is affected by HIT.

XSL•FO
RenderX

Through this exploratory, contextual analysis, we sought to learn which aspects of HIT are perceived by physicians as valuable and which they perceive as contributing to their burnout; the balance between the 2 shapes their overall experience. Our secondary objective, given sufficient data, was to run tests of significance, which could contribute to a model for understanding the factors that contribute to tech-related burnout.

Methods

Survey Instrument

Between 2018 and 2024, OMD conducted an annual survey of clinicians as an evaluative mechanism (internal purposes only) to provide insight into the impact of physician-facing initiatives and to inform future priorities. Within this survey, there was a section on challenges, with optional open-text questions that were reviewed and coded for themes. The cumulative findings suggested that a focused interrogation of the impact of technology on physicians was warranted.

We designed and conducted a self-administered electronic questionnaire on HIT-related burnout, including potential factors, guided by (1) a literature review on HIT-related factors contributing to physician burnout, (2) a review of qualitative responses to annual OMD surveys on challenges with HIT, and (3) team discussions of evidence to date. The questionnaire was piloted by the Nova Scotia E-Health Committee, comprising 9 physicians and 2 medical learners, and then refined with support from OMD-affiliated physicians. The final 25-item instrument consisted of 23 mandatory closed-ended and 2 open-ended questions. The first question asked whether respondents were currently using HIT in their primary practice; those who responded "no" were excluded from subsequent questions on HIT. Three potentially sensitive questions in the "About You" section included "prefer not to answer" as an option. Two versions containing the same content specified regional consent and identification details (Multimedia Appendix 1). Microsoft Dynamics 365 Customer Voice was used to develop and distribute both questionnaires separately via links embedded in email and on social media. Respondents could complete the questionnaire at their own pace and edit previous responses, if necessary. No incentives were provided for survey completion. Survey design is reported according to the CHERRIES checklist [19].

Sampling and Distribution

The questionnaire was launched on February 14, 2024, with an in-field period until April 5, 2024. In Nova Scotia, distribution used nonprobability voluntary response sampling [20] via Dalhousie Faculty of Medicine (FoM) communications, the website, social media, and department newsletters. The Doctors Nova Scotia e-newsletter and discussion forum and the *Nova Scotia Health (NSH)* newsletter also distributed the e-questionnaire and reminders. In Ontario, the questionnaire was distributed using OMD's customer relationship management software to an estimated sampling frame of between 24,000 and 31,000 Ontario physicians and other clinical staff. Reminder emails to complete the questionnaire were sent at regular intervals. Additionally, social media posts were shared throughout the distribution period on behalf of both OMD and the Ontario Medical Association. Unique identifiers in the form of College of Physicians and Surgeons of Ontario registration numbers precluded duplicate responses. Eligibility criteria of "current use of HIT" and "family physician" or "specialist" were applied to ensure a valid sample of respondents according to study objectives.

Analysis

Quantitative

Using Microsoft Power Query and Power BI, Ontario and Nova Scotia data were cleaned and organized separately due to the qualitative and methodological distinctiveness of the 2 samples, thereby addressing potential methodological, statistical, and interpretive errors [21]. Structured data models and tables were derived using step-by-step queries for export into the R Studio environment (version 2024.04.2+764 [2024.04.2+764]; Posit Software, PBC). After consulting the relevant literature [22,23], we determined that there was insufficient power in the Nova Scotia sample for significance testing due to the bias inherent in small sample sizes. We then limited our comparative analysis to descriptive data. For numeric data, measures of central tendency (mean) and dispersion (SD and range) were calculated. For categorical data, frequency counts or percentages were calculated.

Statistical tests of reliability and internal item consistency were conducted on the Ontario perceptions scale data to evaluate psychometric properties, as well as the potential to calculate an overall score on perceptions. The perceptions scale data showed a standardized Cronbach α criterion of 0.854 and moderate intercorrelations (Table 1). A polychoric factor analysis with equamax rotation yielded a 2-factor solution, which could be described as *positive perceptions about HIT* (ie, "easier," "communication," "revolutionize," "access data," "quality care," and "patient coordination") and *negative perceptions about HIT* (ie, "negative impact," "reason for job stress," and "prevent patient care"). To derive an overall numeric score for perceptions as a proxy measure of attitude toward HIT, each item's Likert-scale categories were converted into numeric equivalents, with scores of 1 for "Strongly Disagree" and 5 for "Strongly Agree." The 3 *negative perceptions about HIT* items were reverse scored. The resulting overall perception score ranged from 9 to 45, with higher scores indicating more favorable attitudes toward HIT.

XSL•FO
RenderX

Table . Intercorrelations matrix for perceptions scale items.

	Access data	Communication	Patient coordination	Easier	Negative impact	Prevent patient care	Quality care	Reason for job stress	Revolutionize
Access data	—	0.380[a]	0.396[a]	0.341[a]	–0.287[a]	–0.191[a]	0.383[a]	–0.223[a]	0.266[a]
Communication	0.380[a]	—	0.397[a]	0.399[a]	–0.372[a]	–0.238[a]	0.398[a]	–0.321[a]	0.309[a]
Patient coordination	0.396[a]	0.397[a]	—	0.426[a]	–0.368[a]	–0.226[a]	0.503[a]	–0.259[a]	0.298[a]
Easier	0.341[a]	0.399[a]	0.426[a]	—	–0.545[a]	–0.425[a]	0.542[a]	–0.505[a]	0.430[a]
Negative impact	–0.287[a]	–0.372[a]	–0.368[a]	–0.545[a]	—	0.601[a]	–0.475[a]	0.658[a]	–0.353[a]
Prevent patient care	–0.191[a]	–0.238[a]	–0.226[a]	–0.425[a]	0.601[a]	—	–0.332[a]	0.591[a]	–0.217[a]
Quality care	0.383[a]	0.398[a]	0.503[a]	0.542[a]	–0.475[a]	–0.332[a]	—	–0.372[a]	0.455[a]
Reason for job stress	–0.223[a]	–0.321[a]	–0.259[a]	–0.505[a]	0.658[a]	0.591[a]	–0.372[a]	—	–0.264[a]
Revolutionize	0.266[a]	0.309[a]	0.298[a]	0.430[a]	–0.353[a]	–0.217[a]	0.455[a]	–0.264[a]	—

[a]$P<.0001$.

Ontario numeric data were assessed for evidence of normal distributions and outliers. Outliers were identified by examining boxplots and normal curves, as well as via statistical tests such as the Rosner test. Subject matter experts were consulted to determine plausible upper and lower limits for a range of values. These included the 3 physician coauthors and 2 OMD colleagues with decades of experience working with clinical practices on the use of HIT in its various forms. Outliers identified in the following variables were replaced with the median of the distribution: “hours per week” (6 values); “hours per week on admin” (10 values); “extra hours per week on admin” (8 values); “% of time spent on admin” (4 values); “% on patient care” (1 value); “% burnout from tech” (1 value); “age” (3 values); and “years of practice” (3 values).

Respondents assessed their level of tech-related burnout on a scale of 0 (no burnout) to 10 (extreme burnout). Respondents were asked to rank 10 administrative tasks (validated through a systematic literature review, consultation, and the results of internal qualitative research) in order from greatest to least burden. The percentage of respondents who selected a task as one of the top 3 most burdensome tasks was derived by summing the counts of first, second, and third place responses for each administrative task and then dividing the sums by the number of respondents. Using robust, brute-force, rank-aggregation statistical procedures, the final list of tasks from most to least burdensome was derived.

Due to the type of data and observed nonnormal distributions, nonparametric tests of group differences were selected to better understand tech-related burnout in relation to Ontario sample characteristics. Specifically, tests of group median differences were conducted using Kruskal-Wallis (KW) rank-sum tests for independent samples, as well as follow-up Mann-Whitney *U*-tests, where relevant. To determine statistical associations among numeric data, Pearson product-moment or point-biserial correlation tests were conducted.

Multiple regression analysis was used to identify potential factors associated with self-reported tech-related burnout. For this analysis, tech-related burnout was treated as a continuous numeric outcome variable [24,25]. First, potential factors and covariates were identified. Then, nonparametric tests were conducted systematically for each potential factor or covariate in relation to tech-related burnout to assess statistical significance and effect sizes. Monotonic effects of ordinal-scale variables were assessed using Bayesian generalized nonlinear multilevel modeling methods. Surviving factors were entered systematically into forward and backward stepwise regression tests until the best-fitting parsimonious model was yielded. Due to the large sample size, power, and potential clinical significance of the findings, a stringent alpha criterion of $P=.001$ was used to determine statistical significance.

Qualitative

We included open text questions (response required) to collect respondents’ reasoning behind their answers. The main open-text question was “How does health information technology contribute to or help alleviate your burnout?” We reviewed and coded categories and compared emerging themes in the quantitative data and from our literature review on physician burnout [26]. As the responses were in a long-answer format, it was possible to code multiple references within a single response. A total of 43 responses were omitted from the coding report for being either (1) uncodable (eg, a dash, a dot, a question mark, or random letters), (2) some version of “no idea,” or (3) difficult to determine what was meant by the respondent (eg, a phrase that appeared unrelated to the question), even following consultation with other members of the research team. Each team coded separately, led by a coder familiar with their jurisdiction to facilitate interpretation of context-specific references in the open text. Two coders were used for the Ontario sample and 1 for the Nova Scotia sample. While this meant that each team generated its own codebook, our intent

XSL•FO
RenderX

was to capture the perspectives of respondents within their regional contexts—given differences in population size, digital maturity, etc—to see whether any areas of commonality emerged despite those differences. After initial coding, codebooks were iteratively refined within the research team, and coding was revisited with discussion where ambiguity or disagreement emerged. Following completion of coding, salient codes were reviewed in detail for common themes across samples.

Ethical Considerations

The Nova Scotia questionnaire was approved by the Research Ethics Board at Dalhousie University (Halifax, Nova Scotia, Canada, file number: 2023 - 6993). As OMD operates under both a provincial government and professional association structure without an institutional ethics board, the Ontario survey was deployed with informed consent language (in consultation with the Dalhousie team and OMD's legal team) in the survey header, asking participants to read the consent text, which also cites OMD's privacy policy on information sharing [27]. Participants were asked to confirm they had read the text and that they agreed to participate. Emailed invitations to complete the survey also used language regarding the voluntary nature of participation.

Results

Quantitative Findings

Because the objective of the study was to explore *physician* experiences with HIT, 138 Ontario and 20 Nova Scotia nonphysician clinicians and clinic staff, as well as 24 Ontario and 4 Nova Scotia non-HIT physician users, were excluded, resulting in 1245 Ontario and 136 Nova Scotia physician HIT user respondents. Data on clinician and clinical characteristics are reported in Table 2. Significant differences in self-reported tech-related burnout by characteristic are noted for the Ontario sample only.

XSL•FO
RenderX

Table . Physician and clinical characteristics.

Characteristics	Ontario (N=1245)	Nova Scotia (N=136)
Practice type, n (%)		
Single choice		
Hospital-based	119 (9.6)	66 (48.5)
Group	812 (65.2)	51 (37.5)
Solo	250 (20.1)	12 (8.8)
Other	64 (5.1)	7 (5.1)
Fee structure		
Single choice		
Clinical or academic funding	79 (6.3)	74 (54.4)
Fee-for-service	491 (39.4)	22 (16.2)
Salaried	133 (10.7)	22 (16.2)
Other (eg, capitation models)	542 (43.5)	18 (13.2)
Clinician type[a]		
Single choice		
Family physician	911 (73.2)	53 (39.0)
Specialist	334 (26.8)	83 (61.0)
Gender[b]		
Single choice		
Woman	649 (52.1)	84 (61.8)
Man	559 (44.9)	44 (32.4)
Other	37 (3.0)	8 (5.9)
Age (y), mean (range)	52 (27-79)	50.4 (30-81)
Years of practice (y), mean (range)	22.4 (0-60)	19.5 (1-52)
Racial or ethnic identity, n (%)		
Multiple choice		
East Asian	92 (7.4)	5 (3.7)
South Asian	147 (11.8)	4 (2.9)
Black	21 (1.7)	2 (1.5)
White	744 (59.8)	104 (76.5)
Middle Eastern	95 (7.6)	2 (1.5)
Southeast Asian	13 (1.0)	2 (1.5)
Latin American	15 (1.2)	4 (2.9)
Indigenous	8 (0.6)	—
Prefer not to answer	47 (3.8)	9 (6.6)
Other	63 (5.1)	4 (2.9)
Perceived proficiency using HIT		
Single choice		
High	290 (23.3)	31 (22.8)
Above average	475 (38.2)	43 (31.6)
Average	431 (34.6)	53 (39.0)
Below average	43 (3.5)	5 (3.7)

XSL•FO
RenderX

Characteristics	Ontario (N=1245)	Nova Scotia (N=136)
Low	6 (0.5)	4 (2.9)

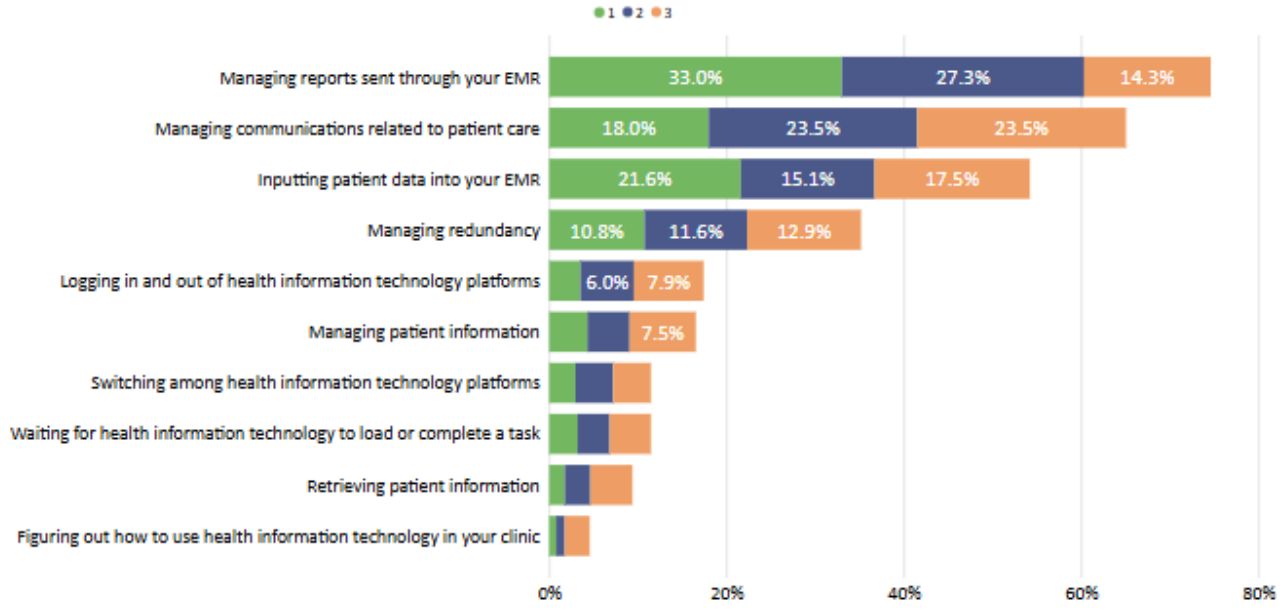
[a]*P*<.001.

[b]Due to small group numbers, the categories “nonbinary” and “prefer not to answer” were grouped with “other” responses.

Top Burdensome Administrative Tasks

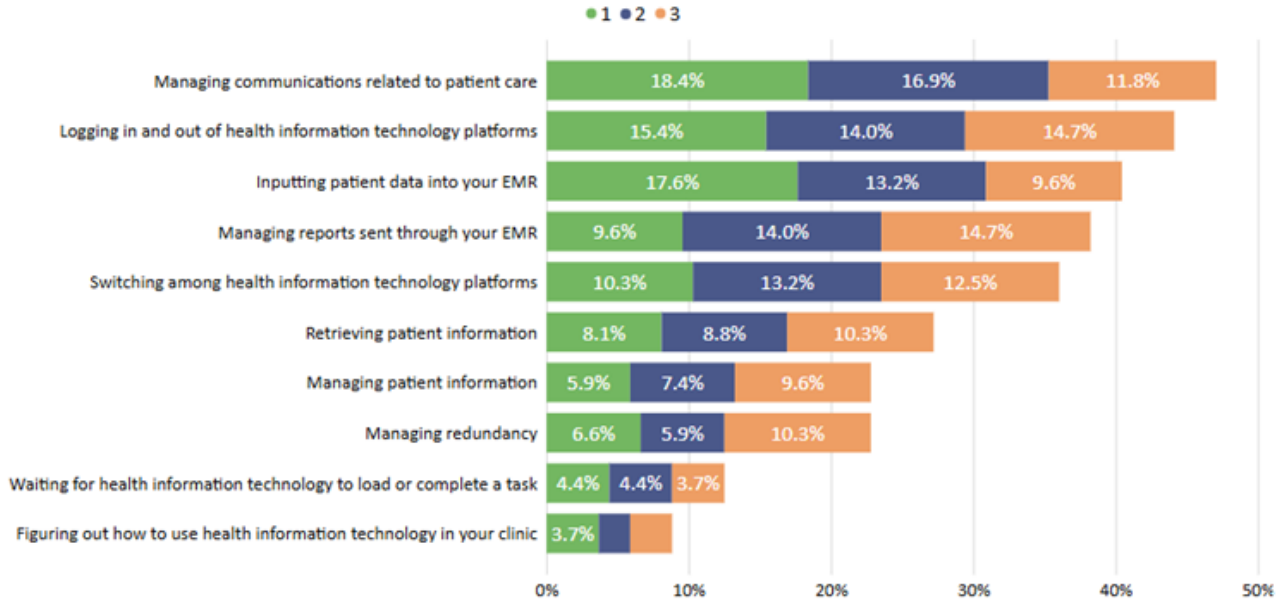
Figures 2 and 3 show the order from most to least burdensome administrative tasks, with bars representing the percentages of the top 3 burdens. Over half of the Ontario sample ranked the following administrative tasks as a top 3 burden: managing reports sent through your EMR (929/1245, 74.6%), managing communications related to patient care (810/1245, 65.1%), and inputting patient data into your EMR (675/1245, 54.2%). The top 3 burdens for the Nova Scotia sample were managing communications related to patient care (n=72, 47.1%), logging in and out of health information technology platforms (60/136, 44.1%), and inputting patient data into your EMR (55/136, 40.4%). For Ontario specialists, the top 3 most burdensome tasks were managing communications related to patient care, inputting patient data into your EMR, and managing reports sent through your EMR. For Nova Scotia specialists, the top 3 most burdensome tasks were logging in and out of HIT platforms, managing communications related to patient care, and retrieving patient information. For both Ontario and Nova Scotia family physicians, the top 3 most burdensome tasks were managing reports sent through your EMR, managing communications related to patient care, and inputting patient data into your EMR.

Figure 2. Top 3 most burdensome tasks among Ontario physicians (N=1245).



XSL•FO
RenderX

Figure 3. Top 3 most burdensome tasks among Nova Scotia physicians (N=136).



Perceptions and Motivations of HIT

Table 3 provides data on perceptions of HIT, as well as motivations for using HIT. For the Ontario sample, significant differences in burnout levels across the agreement categories for certain perception and motivation items are noted. Most Ontario and Nova Scotia physicians agreed with the following beliefs regarding HIT: HIT facilitates coordination of patient care (978/1245, 78.6% in Ontario; 82/136, 60.3% in Nova Scotia) and HIT improves the quality of patient care (819/1245, 65.8% in Ontario; 76/136, 55.9% in Nova Scotia). Most Ontario physicians agreed that HIT makes it convenient for me to access data (861/1245, 69.2%), while over half of Nova Scotia physicians agreed that HIT will revolutionize the future of medicine (82/136, 60.3%). In terms of motivations for using HIT, the most frequently cited reason for using HIT among Ontario physicians was that they either believed HIT would help them provide better patient care or believed HIT would be of value to them, both internally driven sources of motivation. Among Nova Scotia physicians, the most frequently cited reason for using HIT was that they were required to use it, an externally driven source of motivation.

XSL•FO
RenderX

Table . Perceptions of HIT and motivations for using HIT among physicians.

Survey item and category	Ontario (N=1245), n (%)	Nova Scotia (N=136), n (%)
HIT facilitates communication and coordination of patient care with my administrative staff (if applicable)[a]		
Strongly agree	255 (20.5)	26 (19.1)
Agree	723 (58.1)	56 (41.2)
Neither agree nor disagree	192 (15.4)	32 (23.5)
Disagree	51 (4.1)	17 (12.5)
Strongly disagree	24 (1.9)	5 (3.7)
HIT increases ease of communication with other health professionals[a]		
Strongly agree	143 (11.5)	12 (8.8)
Agree	582 (46.7)	59 (43.4)
Neither agree nor disagree	301 (24.2)	41 (30.1)
Disagree	180 (14.5)	18 (13.2)
Strongly disagree	39 (3.1)	6 (4.4)
HIT improves the quality of patient care[a]		
Strongly agree	152 (12.2)	8 (5.9)
Agree	667 (53.6)	68 (50.0)
Neither agree nor disagree	306 (24.6)	45 (33.1)
Disagree	80 (6.4)	11 (8.1)
Strongly disagree	40 (3.2)	4 (2.9)
HIT is a major reason for my overall job stress[a]		
Strongly agree	250 (20.1)	19 (14.0)
Agree	399 (32.0)	46 (33.8)
Neither agree nor disagree	319 (25.6)	39 (28.7)
Disagree	224 (18.0)	24 (17.6)
Strongly disagree	53 (4.3)	8 (5.9)
HIT makes it convenient for me to access all data I need from different sources for patient care[a]		
Strongly agree	245 (19.7)	14 (10.3)
Agree	616 (49.5)	57 (41.9)
Neither agree nor disagree	163 (13.1)	24 (17.6)
Disagree	160 (12.9)	27 (19.9)
Strongly disagree	61 (4.9)	14 (10.3)
HIT makes the things I want to accomplish easier to get done[a]		
Agree	579 (46.5)	56 (41.2)
Neither agree nor disagree	315 (25.3)	35 (25.7)
Disagree	172 (13.8)	25 (18.4)
Strongly disagree	61 (4.9)	11 (8.1)
HIT negatively impacts my practice[a]		
Strongly agree	100 (8.0)	6 (4.4)
Agree	225 (18.1)	24 (17.6)
Neither agree nor disagree	411 (33.0)	55 (40.4)
Disagree	389 (31.2)	37 (27.2)
Strongly disagree	120 (9.6)	14 (10.3)

XSL•FO
RenderX

Survey item and category	Ontario (N=1245), n (%)	Nova Scotia (N=136), n (%)
HIT prevents me from spending more time with my patients[a]		
Strongly agree	292 (23.5)	14 (10.3)
Agree	370 (29.7)	54 (39.7)
Neither agree nor disagree	269 (21.6)	30 (22.1)
Disagree	259 (20.8)	29 (21.3)
Strongly disagree	55 (4.4)	9 (6.6)
HIT will revolutionize the future of medicine[a]		
Strongly agree	194 (15.6)	14 (10.3)
Agree	488 (39.2)	68 (50.0)
Neither agree nor disagree	370 (29.7)	38 (27.9)
Disagree	145 (11.6)	13 (9.6)
Strongly disagree	48 (3.9)	3 (2.2)
Motivations[b]		
I am required to use HIT at my practice.	673 (54.1)	104 (76.5)
I believed HIT would help me provide better patient care.	867 (69.6)	78 (57.4)
My practice was already using HIT when I joined.	420 (33.7)	57 (41.9)
I believed HIT would be of value to me.	749 (60.2)	51 (37.5)
I believed that the technology would be easy for me to use.	549 (44.1)	41 (30.1)
It would make my practice appear attractive to potential applicants and patients.	186 (14.9)	16 (11.8)
I expected it would be interesting to use HIT at my practice.	187 (15.0)	11 (8.1)

[a]*P*<.001.

[b]This was a multiple-response option, so percentages will not add to 100%.

Physicians' Typical Work Week

Table 4 shows descriptive statistics on physicians' work week. Significant associations with technology-related burnout were noted for the Ontario sample. Ontario and Nova Scotia physicians in our sample reported an average work week of 37.4 (SD 14.0) and 39.1 (SD 14.5) hours, respectively, about half of which was spent on direct patient care (50.2%, SD 16.4) in the Ontario sample. Around 26% and 20% of Ontario and Nova Scotia physicians' time was spent on administration related to patient care. Ontario and Nova Scotia physicians worked an average of 8.6 (SD 7.1) and 4.8 (SD 5.7) additional hours per week, respectively, offsite on administration related to patient care.

Table . Physicians' typical work week.

Survey item and category	Ontario (N=1245), mean (SD; range)	Nova Scotia (N=136), mean (SD; range)
Hours worked per week at the practice[a]	37.4 (14.0; 0-90)	39.1 (14.5; 8-80)
Hours worked per week on administration related to patient care[a]	14.0 (9.1; 0-50)	10.5 (9.4; 0-60)
Hours worked per week outside the practice on administration related to patient care[a]	8.6 (7.1; 0-35)	4.8 (5.7; 0-30)
% of time at practice spent on direct patient care	50.2 (16.4; 0-100)	53.0 (28.6; 0-100)
% of time at practice spent on administration tasks related to patient care[a]	25.7 (16.1; 0-100)	25.7 (16.1; 0-100)

[a]*P*<.001.

XSL•FO
RenderX

Quality of HIT Support

In terms of HIT support, over 40% in Ontario (539/1245, 43.3%) received support from their technology vendor; over half in Nova Scotia (69/136, 50.7%) reported support from their on-site technical support. In both cases, these forms of support were deemed at least acceptable (75.4% for Ontario and 79.7% for Nova Scotia; Table 5). For the Ontario sample, significant differences by tech-related burnout are noted.

Table . Sources and quality of support among physicians.

Source of support and quality of support[a]	Ontario, n (%)	Nova Scotia, n (%)
On-site tech support	445 (35.7)	69 (50.7)
Excellent	43 (9.7)	6 (8.7)
Very good	157 (35.3)	14 (20.3)
Acceptable	158 (35.5)	35 (50.7)
Fair	61 (13.7)	11 (15.9)
Poor	26 (5.8)	3 (4.3)
Technology vendor	539 (43.3)	26 (19.1)
Excellent	36 (6.7)	—
Very good	140 (26.0)	6 (23.1)
Acceptable	230 (42.7)	14 (53.8)
Fair	82 (15.2)	4 (15.4)
Poor	51 (9.5)	2 (7.7)
Other	261 (21.0)	41 (30.1)

[a]P<.001.

Tech-Related Burnout Levels

In the Ontario and Nova Scotia samples, the median level of self-reported tech-related burnout was 7 and 6, respectively; levels with the highest frequency counts were 7 or 8 in the Ontario sample and 5, 6, or 7 in the Nova Scotia sample. A little over half of the Ontario sample reported a tech-related burnout level of at least 7, while almost two-thirds reported a tech-related burnout level of at least 6. In the Nova Scotia sample, most respondents reported a tech-related burnout level of at least 5. Table 6 shows the findings from an additional question on tech-related burnout.

Table . Percentage of overall work-related burnout due to HIT.

Province	Work-related burnout score, mean (SD; range)
Ontario (N=1245)	42.9 (26.0; 0-100)
Nova Scotia (N=136)	37.0 (25.6; 0-99)

A series of KW rank sum tests on the Ontario data, given its larger sample size, yielded significant results for support quality (KW χ_4^2=90.6; P<.001), proficiency (KW χ_4^2=25.0; P<.001), and clinician type (KW χ_1^2=21.8; P<.001). In this sample, higher quality support and greater proficiency yielded lower levels of self-reported tech-related burnout, whereas lower-quality support and greater (less) proficiency yielded lower (higher) levels of self-reported tech-related burnout. Specialists also tended to report lower tech-related burnout. More unfavorable or favorable perceptions of HIT yielded higher or lower levels of self-reported tech-related burnout. Specifically, agreeing that HIT is the reason for job stress (KW χ_4^2=523.98; P<.001), prevents patient care (KW χ_4^2=305.6; P<.001), or negatively impacts their practice (KW χ_4^2=360.8; P<.001) yielded significantly higher levels of self-reported tech-related burnout. Conversely, all other items yielded statistically significant differences in median self-reported tech-related burnout, with more favorable views of perception linked to lower levels of self-reported tech-related burnout. Pearson product moment correlation tests also yielded a significant correlation between the overall score on perceptions and tech-related burnout (r=−0.56; P<.001). Hours per week at practice (r=0.19; P<.001), hours per week (r=0.33; P<.001) or percent of time (r=0.21; P<.001) spent on administration related to patient care, and extra hours per week spent on administration related to patient care (r=0.34; P<.001) yielded weak-to-moderate positive correlations with self-reported tech-related burnout.

Systematic stepwise regression tests yielded a robust, well-fitting model explaining about 60% of the variation in self-reported tech-related burnout (R^2=0.60; F(7,1237)=266.515; P<.001). Table 7 shows the performance of individual factors in the regression model. Percentage burnout tech (ie, the respondent's estimated percentage that technology contributes to their burnout) was subsequently removed due to difficulties in interpretation, which resulted in a model accounting for about

XSL•FO
RenderX

half of the variation in self-reported tech-related burnout, with 6 stable, meaningful, and potentially actionable factors. All ordinal factors evidenced monotonic, near-linear trends. Follow-up statistical tests were conducted to confirm the regression model. Neither multicollinearity (ie, through examination of variance inflation factors) nor autocorrelation among residuals (ie, through the Durbin-Watson test) was observed. A Bonferroni test for outliers indicated no studentized outliers (P<.05).

Table . Regression factor coefficients.

Predictor	Coefficients (SE)	*t* test (*df*)[a]	*P* value
Intercept	1.841 (0.333)	5.534 (138)	<.001
Communication	–0.202 (0.050)	–4.021 (138)	<.001
Prevent patient care	0.227 (0.048)	4.695 (138)	<.001
Reason job stress	0.661 (0.059)	11.277 (138)	<.001
Support quality	–0.194 (0.046)	–4.186 (138)	<.001
Hours per week	0.017 (0.003)	4.903 (138)	<.001
Hours per week extra	0.052 (0.007)	7.529 (138)	<.001

[a]Two-tailed *t* test.

Qualitative Findings

Both coding teams began by assigning responses to the main qualitative question, “How does health information technology contribute to or help to alleviate your burnout?” into 2 themes: negative (corresponding to “contributes to”) and positive (corresponding to “alleviates”). With the larger Ontario sample, QSR NVivo software was used to facilitate coding, and automated sentiment analysis was attempted but was found to be inaccurate upon review—with frequent misattributions of sentiment due to expressive ambiguity or sarcasm, for example. As described in the “Methods” section, blank or difficult-to-interpret responses were excluded from coding.

Using this approach, 146 references were coded as negative and 31 as positive in the Nova Scotia sample, with 864 coded as negative and 657 as positive in the Ontario sample. Each coding team then proceeded with open coding for themes that emerged under each of the 2 larger categories. Codebooks were developed for each province’s sample with codenames and definitions. Once this step was completed, the 2 coding teams compared codebooks to identify similarities across themes—with a focus on aligning code definitions rather than codenames.

From this we discovered that the 2 most common positive themes in both groups were aligned, although in reverse order. The first of these in the Ontario sample dealt with HIT’s capacity to facilitate data management and accessibility (coded as “data management”—168 references; in the Nova Scotia sample coded as “accessibility”—7 references, the second most common). The second refers to administrative efficiency (coded as “regular task completion”—112 references; in the Nova Scotia sample coded as “efficiency”—11 references, the most common).

The considerably larger set of references for the negative statements also offered greater complexity and some differences between the samples. The highest frequency of negative statements in the Ontario sample referred to report management. The second most common theme in the Ontario sample and the most common in the Nova Scotia sample were aligned: the challenges of systems that “don’t talk to each other,” that is, do not interact easily (some variation of which we saw frequently), so that accessing all the necessary patient data requires logging in and out of multiple unconnected platforms. These responses often cited the interruptive nature of recalling different login information and completing dual-factor authentication. These types of references were initially coded as “interoperability and integration” in the Ontario sample and “fragmented” in the Nova Scotia sample.

Otherwise, there were additional differences in the remaining negative codes. Table 8 shows the 4 most common negative subthemes coded in each sample, including the codename, definition, number of references for each code, and a sample quote.

XSL•FO
RenderX

Table . Qualitative coding and sample responses: negative experiences.

Code	Definition	Count, n	Sample quote
Ontario (n=684)			
Managing reports	Duplicate, redundant, irrelevant, too long, labeled incorrectly, or poorly formatted reports.	177	“I have to read all reports that come in, even the useless ones. Waste of my time.”
Interoperability and integration	Systems don’t speak to one another; cognitive overload from switching between platforms and authentication modes.	120	“Lots of platforms promising to improve my workflow, but each requires a unique login and password.”
Work-life balance	EMR availability 24/7 leads to expectation to access beyond work hours.	111	"Never get a break from checking, even when on vacation.”
Referrals	Too many different forms, and specialists reject if forms are not filled correctly; no centralized system to pull information.	102	"Different institutions have different referral platforms and logins.”
Nova Scotia (n=146)			
Interoperability and integration	Fragmented systems require logging in and out of each system and systems do not communicate between one another.	30	"You need to search multiple different techs to find all the data needed.”
Accessibility	Data cannot be found; uploading delays; stored in the wrong place; difficulty navigating.	17	“Limited access to some information locations or difficulty finding specific documentation.”
Time	Time-consuming, slower to use HIT than paper.	17	“Extra hours for admin tasks leave me tired and use up brain energy.”
Usability	HIT systems are not user-friendly, not intuitive, too difficult to use, and technologically challenging.	17	“Implementation with our EMR was so non-user-friendly that we had to deimplement it and revert to phone calls.”

Discussion

Interpretation

Despite notable differences between the provinces in HIT maturity, complexity, adoption rates, practice models, and physician characteristics, common themes emerged surrounding HIT involvement in burnout. While physicians identified many positive aspects of HIT, our findings show that most physicians self-assessed their tech-related burnout level as high: 6 out of 10 in Nova Scotia and 7 out of 10 in Ontario. On average, Ontario respondents reported that 43% of their overall burnout is tech-related, while Nova Scotia respondents reported 37%. Ontario respondents reported greater tech-related burnout and attributed a higher percentage of their overall burnout to technology. Over a third of burnout in our respondents’ self-reports is attributable to HIT.

The top-ranked administrative burdens share commonalities across provinces. Nova Scotia respondents identified managing communications as their primary burden; in Ontario, this ranked second after managing reports due to the effects of nonstandardized implementation of Health Report Manager (HRM) in hospitals (a long-standing issue, which is in the process of being addressed in the province). Both samples ranked inputting patient data into their EMR in third place and logging in and out of technology platforms in their top 5 (second in Nova Scotia and fifth in Ontario). This issue emerged strongly in the qualitative material, which revealed frustrations with logging in and out of non-EMR-integrated platforms for referrals, data portals, etc, which all require unique logins, passwords, and dual-factor authentication. Responses suggest that physicians experience these as both adding to cognitive burden and taking time away from patient care. In both samples, ongoing change initiatives—the provincial rollout of OPOR in Nova Scotia and both HRM improvement initiatives [28] and the Primary Care Action Plan in Ontario [29]—are likely to have an impact on the future experience of administrative burdens and perhaps shift these rankings. Other initiatives at the national level to improve vendor attention to interoperability and standards for HIT solutions brought to market may also have an impact on the physician experience. Bill S-5, the *Connected Care for Canadians Act*, would require HIT vendors to make the technology they sell to clinics interoperable, with penalties for “data blocking.” [30-32] Further, a national effort to standardize data content nationwide is underway with the Pan-Canadian Health Data Content Framework [33]. In the meantime, expensive HIT tools are outpacing our health system’s capacity to adapt. For example, auto-population of forms, patient data integration, and AI assistance are not uniformly regulated or implemented across Canadian practices and provinces [34].

XSL•FO
RenderX

One of the strengths of our approach is the elicitation of candid physician comments via the open text survey questions and our subsequent analysis. This has allowed us to see more layers of the physician experience, especially the complex "love/hate" relationship physicians have with their HIT. In the qualitative material, we found commonality across the Ontario and Nova Scotia samples in terms of the expressed positives of HIT, particularly for data management and administrative efficiency. These align with the reasons HIT was developed and implemented, particularly where government investment is involved. We note with interest that while both samples showed common issues with interoperability and integration (in line with the quantitative results), there were differences in how respondents expressed their other challenges with HIT. In the Ontario sample, we found high counts of references to specific issues, for example, incoming document management and referrals; this was not the case in the Nova Scotia sample, where "usability" was a common theme. This may be due to differences in the samples (ie, higher counts of specific mentions in the larger Ontario sample). We also suggest that the sample from Ontario's more mature HIT environment may have progressed beyond general usability concerns to complaints about more specific challenges.

Perceptions about HIT showed some agreement with positive statements about technology. This tended to be stronger in Ontario, perhaps due to that province being a more seasoned HIT landscape than Nova Scotia and some anticipatory concern in Nova Scotia about the province's upcoming move from a largely paper-based system to OPOR. In both samples, more than half of respondents agreed with positive statements about the impact of HIT on patient care and data access. However, about half of each sample also strongly agreed that HIT prevents spending more time with patients. Indeed, both samples reported spending about half their time on direct patient care, with a notable proportion on administration (approximately 20% in Nova Scotia and 25% in Ontario). These particular (types of) perceptions are conceptually similar to tech-related burnout; therefore, one would expect robust associations. However, findings were moderate and somewhat unclear, which may indicate that perceptions are mediated or moderated by other yet unstudied attributes. Many open text responses contained some version of "I spend more time with the computer than with my patients" related to documentation and accountability expectations that increase screen-related tasks and dominate the patient encounter. These negative perceptions associated with administrative burdens compete with the perceived benefits to patient care. This aligns with findings elsewhere regarding the contribution of daily frustrations, task, and workload burden to burnout [4,7] and underscores many physicians' complex relationships with HIT.

HIT is (or should be) a highly modifiable contributor to burnout; improving physicians' experience of HIT could decrease some burnout overall. Physicians may see the benefits of HIT for patient care in principle; however, a physician's ability to yield optimal value from HIT depends on usability, ease of switching between platforms and nonstandardized forms and referrals, and lessening administrative burden, all of which influence time spent on patient care. HIT co-designed with physicians [35] could lead to interoperability and usability standards that better address the needs of a variety of different practice types, physician characteristics, and proficiency levels—with positive effects system-wide. A focus on human factors in the design of HIT can offset the usability challenges that may be unintentionally introduced by technology vendors in disregarding the realities of clinical practice [36-38]. An example of co-design supporting interoperability exists in the Belize Health Information System implementation, where Canadian and Belizean health professionals worked with community input for design and deployment [39,40]. The result was a system that was easy to use and reduced administrative tasks, including searches and redundancies—while also generating low development and maintenance costs and improving patient outcomes. Successful implementation of HIT relies on the involvement of end-user health care professionals at the outset to better translate the realities of clinical practice workflow to technological solutions.

Future Directions

The OPOR initiative presents an opportunity for a prestudy and poststudy in the Nova Scotia context, similar to other preimplementation or postimplementation studies in this area—particularly to investigate the impact on dedicated patient care time [41]. In Ontario, government-led initiatives such as the Primary Care Action Plan [29] intend to address several areas of HIT that contribute to administrative burden; plans to standardize implementation of HRM in hospitals will also address a large contributor to unwieldy inboxes [28]. A follow-up questionnaire pursuant to implemented changes could explore whether these have had an impact on administrative burden and burnout. There is also an opportunity to further investigate differences between specialists and primary care physicians. While the Nova Scotia sample had a higher representation of specialists than the Ontario sample, the area of specialty was primarily pediatrics, where physicians often act as de facto family physicians, providing incremental care over years to the same families. As we did find lower levels of self-reported tech burnout among specialists compared to family physicians, capturing the difference between "true" specialists (those who provide secondary rather than primary care) and family physicians may provide insights from the greater contrast in practice style. Although there are similarities across specialists and family physicians in terms of most burdensome tasks, the qualitative analysis in our study suggests that the reality of administrative burden in clinical practice may differ between the 2 groups. In turn, this may be contributing to tensions (eg, primary care physicians express frustration with nonstandardized referral forms and the extra administrative burden these forms generate).

Limitations

These findings are preliminary and exploratory and provide a foundation on which to further explore tech-related burnout. We recognize as a limitation that our measure is subjective, based on self-report, and not a validated objective measure. These findings are based on perceptions and associations between self-reported items. Because we used multiple avenues to promote the questionnaire, respondents were not prevented

XSL•FO
RenderX

from accessing the questionnaire more than once; however, we controlled for duplicate responses via unique identifiers. Because we used various sampling methods and exclusion criteria, it is difficult to discern the sampling frame and calculate response rates. The sample is unlikely to be representative. There are limitations to generalizability associated with the small Nova Scotia sample and lack of variability in characteristics that preclude certain statistical tests, hinder precision, and potentially introduce bias, including overestimation of effect sizes. Certain response biases (eg, respondents with higher levels of burnout may have been more motivated to participate), common in this type of research, may have influenced the analysis. We attempted to address related biases in our analysis and reporting. Because findings primarily reflect physician experiences within 2 Canadian provincial health systems, they may not reflect or represent experiences in other health care settings. Although measured on a scale following best practices and adapted from similar scales noted in the literature, our tech-related burnout scale has not yet been validated and is subject to interpretation. Although yielding a significant predictive model, it is difficult to ascertain what an increase of a burnout level of "5" to "6" entails clinically. Qualitative research may help to describe these types of scales. The inductive qualitative coding approach, in which we created separate codebooks for Ontario and Nova Scotia, posed limitations to the comparability of that coding. Further, thematic coding may be subject to bias due to the small number of coders involved (2 for Ontario, 1 for Nova Scotia). Furthermore, this analysis should be viewed as exploratory as we continue to understand tech-related burnout and further refine our approach. We note evidence of validity via the strong correlation yielded between our burnout scale and "Percent burnout from technology." Furthermore, some of the regression factors were measured on scales despite evidencing monotonic, near-linear trends from Bayesian generalized nonlinear multilevel modeling tests. Finally, self-report questionnaires are subjective, so we did not secure objective measures of either burnout or time usage. We followed best practice approaches such as item checks of respondents' attention and piloted ease of questionnaire completion. Our findings are best interpreted as reflecting respondent perceptions and experiences.

Conclusion

Our study provides a broad picture of the physician experience of HIT in the 2 settings we studied. Physician needs may change depending on their location, practice type, specialty, scope of practice, and day-to-day schedule. HIT should have the flexibility to accommodate these needs and be tailored to the people it serves. For this to happen, physicians need to be deeply involved, including in the design of programs to simplify clinical workflows, for example, the use of AI in clinical decision support. Physicians expressed a need to spend more time with patients than with HIT. The current plethora of largely unregulated vendors for an increasingly diverse set of platforms to address the needs of clinical practice, and the lack of full patient-centered integration, may result in high costs for less user-friendly systems. This study reinforces the complexity of HIT systems and their impact on administrative burden. Ongoing involvement of physicians—who actively use these systems—in HIT design and implementation is essential.

This study contributes practical relevance to the growing literature examining the relationship between technology and burnout [7,8,42-45] with analysis and interpretation of physician perspectives on their complex relationships with HIT. HIT may facilitate patient data collection, management, and use, but it can also add cognitive and administrative burdens to physicians' everyday work life. HIT is not inherently protective or harmful; rather, physician experience depends heavily on usability, interoperability, workflow alignment, and meaningful physician involvement in design and implementation to address their clinical realities.

Acknowledgments

The authors thank Christina Jones for her support in the analysis and production of this paper. The authors would also like to thank the physicians who participated in this study for their time and candid responses. No GenAI tools were used in the preparation of the manuscript.

Funding

This study was supported by the Canadian Institutes of Health Research (WI2 179934) and the Dalhousie Faculty of Medicine Osman Fund for Research in Medicine. OntarioMD is a wholly owned subsidiary of the Ontario Medical Association and is operationally funded by a transfer payment agreement from Ontario Health.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

MJ and JEG conceptualized and designed the study. TJ designed the questionnaire and conducted the statistical analysis for the quantitative data. DL contributed to the recruitment in Nova Scotia, data acquisition, and analysis. JC, CC, and NEM contributed clinical insights and assisted with the sample recruitment strategy. All authors critically reviewed the work for important intellectual content. All authors contributed to the drafting, scholarly content, and interpretation of data and provided final approval of the version submitted. All authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work have been appropriately investigated and resolved.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey instrument.
[DOCX File, 30 KB - humanfactors_v13i1e93266_app1.docx]

References

1. National Physician Health Survey 2025. : Canadian Medical Association (CMA); 2026 URL: https://digitallibrary.cma.ca/viewer?file=%2Fmedia%2FDigital_Library_PDF%2F2026-CMA-2025-National-Physician-Health-survey-EN.pdf#page=1 [accessed 2026-04-24]
2. Maslach C, Jackson SE. The measurement of experienced burnout. J Organ Behavior 1981 Apr;2(2):99-113. [doi: 10.1002/job.4030020205]
3. Healing the healers: system-level solutions to physician burnout. : Ontario Medical Association; 2021 URL: https://www.oma.org/siteassets/oma/media/pagetree/advocacy/issues/burnout/burnout-paper.pdf [accessed 2026-04-24]
4. A profession in crisis: the survival of family medicine in Ontario. : Ontario College of Family Physicians (OCFP); 2023 URL: https://ontariofamilyphysicians.ca/wp-content/uploads/2023/09/ocfp_member_survey_report_2023_05.pdf [accessed 2026-04-24]
5. Peccoralo LA, Kaplan CA, Pietrzak RH, Charney DS, Ripp JA. The impact of time spent on the electronic health record after work and of clerical work on burnout among clinical faculty. J Am Med Inform Assoc 2021 Apr 23;28(5):938-947. [doi: 10.1093/jamia/ocaa349] [Medline: 33550392]
6. Physician burnout survey. : Doctors Nova Scotia; 2017 URL: https://doctorsns.com/news-events/surveys/physician-burnout [accessed 2023-03-30]
7. Shanafelt TD, Dyrbye LN, Sinsky C, et al. Relationship between clerical burden and characteristics of the electronic environment with physician burnout and professional satisfaction. Mayo Clin Proc 2016 Jul;91(7):836-848. [doi: 10.1016/j.mayocp.2016.05.007] [Medline: 27313121]
8. West CP, Dyrbye LN, Shanafelt TD. Physician burnout: contributors, consequences and solutions. J Intern Med 2018 Jun;283(6):516-529. [doi: 10.1111/joim.12752] [Medline: 29505159]
9. Chang F, Gupta N. Progress in electronic medical record adoption in Canada. Can Fam Physician 2015 Dec;61(12):1076-1084. [Medline: 27035020]
10. Kroth PJ, Morioka-Douglas N, Veres S, et al. The electronic elephant in the room: physicians and the electronic health record. JAMIA Open 2018 Jul;1(1):49-56. [doi: 10.1093/jamiaopen/ooy016] [Medline: 31093606]
11. Lee MS, Nambudiri VE. Electronic consultations and clinician burnout: An antidote to our emotional pandemic? J Am Med Inform Assoc 2021 Apr 23;28(5):1038-1041. [doi: 10.1093/jamia/ocaa300] [Medline: 33313867]
12. Population estimates, quarterly. Statistics Canada, Government of Canada. 2018 Jun. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000901 [accessed 2025-03-20]
13. Yang J. Number of family medicine and general practice physicians in Canada as of 2024, by province*. Statista. 2025 Mar. URL: https://www.statista.com/statistics/831118/canada-family-general-practitioners-by-province/ [accessed 2025-03-20]
14. Certified EMR offerings. OntarioMD. URL: https://www.ontariomd.ca/emr-certification/omd-certified-emr-offerings/ontario-omd-certified-emr-offerings [accessed 2025-10-16]
15. Electronic medical records and integrated solutions - electronic medical records certification. Government of Nova Scotia. URL: https://novascotia.ca/dhw/ehealth/eis/emr-certification.asp [accessed 2025-10-16]
16. E-health. Doctors Nova Scotia. URL: https://doctorsns.com/contract-and-support/e-health [accessed 2025-03-16]
17. One person one record (OPOR). Nova Scotia Health. URL: https://physicians.nshealth.ca/topics/nova-scotia-provider-digital-health-toolkit/one-person-one-record-opor [accessed 2026-04-25]
18. EHealth Ontario - OntarioMD submission. : OntarioMD; 2016 Nov URL: https://files.ontario.ca/19._ontario_md.pdf [accessed 2026-09-02]
19. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34. [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
20. Burns KEA, Duffett M, Kho ME, et al. A guide for the design and conduct of self-administered surveys of clinicians. Can Med Assoc J 2008 Jul 29;179(3):245-252. [doi: 10.1503/cmaj.080372] [Medline: 18663204]
21. Mayen MSA, Nisha SN, Afrin S, Ahammed T, Chowdhury MAB, Uddin MJ. Evaluating the current methodological practices and issues in existing literature in pooling complex surveys: a systematic review. BMC Med Res Methodol 2024 Nov 13;24(1):279. [doi: 10.1186/s12874-024-02400-5] [Medline: 39538131]
22. Judijanto L. Difficulties and practical issues in statistical analysis in research with small sample size: a review. Open Access J Sci 2026;9(1):132-144. [doi: 10.15406/oajs.2026.09.00297]
23. Jaki T, Kim M, Lamont A, et al. The effects of sample size on the estimation of regression mixture models. Educ Psychol Meas 2019 Apr;79(2):358-384. [doi: 10.1177/0013164418791673] [Medline: 30911197]

XSL•FO
RenderX

24. Robitzsch A. Why ordinal variables can (almost) always be treated as continuous variables: clarifying assumptions of robust continuous and ordinal factor analysis estimation methods. Front Educ 2020 Oct 8;5:589965. [doi: 10.3389/feduc.2020.589965]
25. Pasta D. Learning when to be discrete: continuous vs categorical predictors. In: SAS Global Forum 2009: SAS Institute Inc; 2009. URL: https://support.sas.com/resources/papers/proceedings09/248-2009.pdf [accessed 2026-02-09]
26. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
27. Privacy policy. OntarioMD. URL: https://www.ontariomd.ca/privacy-policy [accessed 2025-02-09]
28. HRM improvement recommendations advisory committee. OntarioMD. URL: https://www.ontariomd.ca/pages/hrm-improvement-recommendations-advisory-committee.aspx [accessed 2025-01-25]
29. Ontario's primary care action plan: connecting every person in Ontario to primary care. : Ministry of Ontario; 2025 URL: https://www.ontario.ca/page/ontarios-primary-care-action-plan-january-2025 [accessed 2026-04-24]
30. Health Canada. The government of Canada introduces legislation to build a more connected health care system. Government of Canada. 2026 Feb. URL: https://www.canada.ca/en/health-canada/news/2026/02/the-government-of-canada-introduces-legislation-to-build-a-more-connected-health-care-system.html [accessed 2026-08-22]
31. El Sabawy D, Feldman J, Pinto AD. The Connected Care for Canadians Act: an important step toward interoperability of health data. CMAJ 2024 Dec 8;196(42):E1385-E1388. [doi: 10.1503/cmaj.241123] [Medline: 39653400]
32. Bill s-5: an act respecting the interoperability of health information technology and to prohibit data blocking by health information technology vendors. Parliament of Canada. 2025. URL: https://www.parl.ca/legisinfo/en/bill/45-1/s-5 [accessed 2026-04-26]
33. Pan-Canadian health data content framework — data content standard, version 2, september 2025. : Canadian Institute for Health Information (CIHI); 2025 URL: https://www.cihi.ca/sites/default/files/document/pan-canadian-health-data-content-framework-data-content-standard-en.pdf [accessed 2026-04-26]
34. Canadian Medical Association (CMA). 2021 national survey of Canadian physicians: quantitative market research report. : Canada Health Infoway; 2021 URL: https://www.infoway-inforoute.ca/en/component/edocman/3935-2021-national-survey-of-canadian-physicians/view-document?Itemid=101 [accessed 2026-04-24]
35. Miller AP, Ross K, Brown EM, Chandrasena C, Bhattacharya O. Interoperability of health data: the importance of including physicians. CMAJ 2025 Feb 23;197(7):E192. [doi: 10.1503/cmaj.151780-l] [Medline: 39993817]
36. Or C, Wong K, Tong E, Sek A. Private primary care physicians' perspectives on factors affecting the adoption of electronic medical records: a qualitative pre-implementation study. Work 2014;48(4):529-538. [doi: 10.3233/WOR-131808] [Medline: 24346272]
37. Karsh BT, Holden RJ, Or CKL. Human factors and ergonomics of health information technology implementation. In: Carayon P, editor. Handbook of Human Factors and Ergonomics in Health Care and Patient Safety, 2nd edition: CRC Press; 2012:249-264 URL: https://hub.hku.hk/handle/10722/185234 [accessed 2026-08-22]
38. Carayon P, Hoonakker P. Human factors and usability for health information technology: old and new challenges. Yearb Med Inform 2019 Aug;28(1):71-77. [doi: 10.1055/s-0039-1677907] [Medline: 31419818]
39. Graven M, Allen P, Smith I, MacDonald NE. Decline in mortality with the Belize Integrated Patient-Centred Country Wide Health Information System (BHIS) with embedded program management. Int J Med Inform 2013 Oct;82(10):954-963. [doi: 10.1016/j.ijmedinf.2013.06.003] [Medline: 23834838]
40. Graven M, Allen P, Allen VM, Smith I, MacDonald NE. Achievement of the 2015 Millennium Development Goal 5A (Maternal Mortality) by Belize in 2011. J Obstet Gynaecol Can 2012 Oct;34(10):913-916. [doi: 10.1016/S1701-2163(16)35404-4] [Medline: 23067946]
41. Joukes E, Abu-Hanna A, Cornet R, de Keizer NF. Time spent on dedicated patient care and documentation tasks before and after the introduction of a structured and standardized electronic health record. Appl Clin Inform 2018 Jan;9(1):46-53. [doi: 10.1055/s-0037-1615747] [Medline: 29342479]
42. Gardner RL, Cooper E, Haskell J, et al. Physician stress and burnout: the impact of health information technology. J Am Med Inform Assoc 2019 Feb 1;26(2):106-114. [doi: 10.1093/jamia/ocy145] [Medline: 30517663]
43. Jendly M, Santschi V, Tancredi S, von Wyl V, Chiolero A. Primary care physician digital health profile and burnout: an international cross-sectional study. Eur J Public Health 2025 Dec 1;35(6):1162-1169. [doi: 10.1093/eurpub/ckaf106] [Medline: 40663776]
44. Li H, Dance E, Poonja Z, Aguilar LS, Colmers-Gray I. Agreement between the Maslach Burnout Inventory and the Copenhagen Burnout Inventory among emergency physicians and trainees. Acad Emerg Med 2024 Dec;31(12):1243-1255. [doi: 10.1111/acem.14994] [Medline: 39082466]
45. Murphy DR, Satterly T, Giardina TD, Sittig DF, Singh H. Practicing clinicians' recommendations to reduce burden from the electronic health record inbox: a mixed-methods study. J Gen Intern Med 2019 Sep;34(9):1825-1832. [doi: 10.1007/s11606-019-05112-5] [Medline: 31292905]

XSL•FO
RenderX

Abbreviations

CMA: Canadian Medical Association
DNS: Doctors Nova Scotia
FOM: Faculty of Medicine
HIS: hospital information system
HIT: Health information technology
HRM: Health Report Manager
KW: Kruskal-Wallis
NSH: Nova Scotia Health
OMA: Ontario Medical Association
OMD: OntarioMD
OPOR: One Person One Record

Edited by C LeRouge; submitted 13.Feb.2026; peer-reviewed by C Or, P Haessner; revised version received 10.Jul.2026; accepted 10.Aug.2026; published 23.Sep.2026.

Please cite as:
Jones M, Jason T, Liu D, Comeau J, Chandrasena C, MacDonald NE, Graham JE
Physicians' Complex Relationships With Health Information Technology and Burnout: Cross-Sectional Survey Study
JMIR Hum Factors 2026;13:e93266
URL: https://humanfactors.jmir.org/2026/1/e93266
doi:10.2196/93266

© Mavis Jones, Timothy Jason, Dara Liu, Jeannette Comeau, Chandi Chandrasena, Noni E. MacDonald, Janice E. Graham. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 23.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

From Prototype to Practice: Mixed Methods Study of Human-Centered Design and Usability Evaluation of Augmented Reality for Complex Oncology Surgical Workflows

Maria Matsangidou[1], BA, MA, MA, PhD; Mikolaj R Kowal[2], MBChB; Jakob Yianni[3], BSc, MSc; Maria Krini[4], BA, MSc; Samir Pathak[5], BSc, MSc, MD, FRCS; Katie Gordon[6], BSc; Malcolm A Luker[7], BSc, PhD; Deborah D Stocken[6], MB BCh, MD; David Jayne[2], MB BCh, MD; Constantinos S Pattichis[1,8], HTI, BSc, MSc, PhD

[1]CYENS Centre of Excellence, Lellou Demetriades, Plateia Dimarchou 1, Nicosia, Lefkosia, Cyprus

[2]Leeds Institute of Medical Research, University of Leeds, Leeds, United Kingdom

[3]Institute of Cognitive Neuroscience, University College London, London, United Kingdom

[4]The Cyprus Association of Cancer Patients and Friends, Nicosia, Cyprus

[5]Leeds Teaching Hospitals NHS Trust, Leeds, United Kingdom

[6]Leeds Clinical Trials Research Unit, University of Leeds, Leeds, United Kingdom

[7]HoloCare Ltd, Esher, Surrey, United Kingdom

[8]Department of Computer Science, University of Cyprus, Nicosia, Cyprus

Corresponding Author:
Maria Matsangidou, BA, MA, MA, PhD
CYENS Centre of Excellence, Lellou Demetriades, Plateia Dimarchou 1, Nicosia, Lefkosia, Cyprus

Abstract

Background: Hepato-pancreato-biliary (HPB) oncology involves complex surgical planning and multidisciplinary decision-making that present significant cognitive, spatial, and communicative challenges. While augmented reality (AR) technologies are emerging as valuable tools in surgical planning, most existing systems are developed in a technologically deterministic manner, often without consideration of patients' lived experiences or the collaborative needs of multidisciplinary clinical teams. This limits their integration into complex workflows and their potential for meaningful clinical impact.

Objective: This study aimed to co-design and evaluate a human-centered AR system that supports spatial understanding, emotional sensitivity, and collaborative decision-making in HPB surgical oncology.

Methods: This mixed methods human-centered design study was conducted between February 2024 and September 2025 and comprised 2 phases. In Phase 1, convenience sampling was used to recruit 14 oncology health care professionals (surgeons, oncologists, and psycho-oncology specialists) and 10 patients with cancer from multidisciplinary oncology care settings to participate in co-design workshops and semistructured interviews exploring the cognitive, communicative, emotional, and workflow requirements for an AR system supporting HPB surgical oncology. Qualitative data were analyzed using reflexive thematic analysis and informed iterative prototype development. In Phase 2, the resulting AR prototype was evaluated through usability testing with 10 hepatobiliary clinicians using simulated multidisciplinary tumor board planning scenarios. Usability was assessed using the System Usability Scale (SUS), supplemented by observational data and qualitative feedback.

Results: The co-design process identified intersecting spatial, cognitive, emotional, and communicative challenges in HPB surgical oncology. These findings informed the iterative development of an AR prototype that functioned as a shared cognitive and communicative workspace. The system enabled collaborative manipulation of 3D anatomical models, emotionally calibrated visualizations, and seamless integration with existing planning workflows. The prototype achieved a mean SUS score of 67.25 (SD 7.95), indicating acceptable usability. In simulated multidisciplinary tumor board scenarios, clinicians reported improved spatial understanding, reduced cognitive load, and enhanced interdisciplinary communication. Patients with cancer participating in the co-design phase valued the ability to engage with their own anatomy in an emotionally manageable way and reported that interactive visualization improved their understanding and sense of agency. Overall, the findings suggest that AR systems designed with human-centered methods can function not only as visualization tools but also as boundary objects that support communication and distributed cognition in complex cancer care settings.

Conclusions: A human-centered co-design process can guide the development of AR systems that extend beyond technical novelty to become emotionally attuned, cognitively supportive, and workflow-compatible tools in surgical oncology. By foregrounding the needs of both patients and professionals, AR can function as a boundary object that facilitates collaborative decision-making, empathetic communication, and multidisciplinary coordination in complex oncology care. This study provides

XSL•FO
RenderX

actionable design recommendations for developing AR technologies that are meaningfully integrated into complex clinical contexts.

(JMIR Hum Factors 2026;13:e86725) doi:10.2196/86725

KEYWORDS

augmented reality; human-centered design; surgical oncology; co-design; multidisciplinary care; patient-clinician communication

Introduction

Augmented reality (AR) overlays digital information onto the real-world environment, enabling users to interact with patient-specific 3D anatomical information while maintaining awareness of their physical surroundings [1-3]. In surgery, AR has emerged as a promising technology for improving visualization, spatial understanding, and interpretation of complex medical imaging, allowing clinicians to manipulate and explore 3D reconstructions derived from computed tomography (CT) and magnetic resonance imaging (MRI) scans more intuitively than conventional 2D image review [4-6].

This capability is particularly relevant in hepato-pancreato-biliary (HPB) surgery, a specialty focused on diseases of the liver, pancreas, and biliary system, where intricate anatomy and close relationships between tumors, vascular structures, and bile ducts present significant challenges for surgical planning [7]. Although contrast-enhanced CT and MRI provide detailed anatomical information, surgeons must mentally reconstruct 3D spatial relationships from 2D images, making preoperative planning cognitively demanding and increasing the risk of anatomical localization and trajectory planning errors [8,9].

These challenges have driven growing interest in AR to improve surgical planning and intraoperative navigation. Achieving clear resection margins is fundamental in surgical oncology, yet accurate intraoperative tumor localization remains challenging, particularly in anatomically complex and deformable organs such as the liver [10-12]. Conventional image-guided techniques often provide limited real-time spatial guidance. AR addresses this limitation by superimposing 3D anatomical and pathological information directly onto the surgical field, improving spatial awareness and supporting tumor localization, resection planning, and intraoperative navigation [13,14].

In HPB surgery, AR has improved visualization of tumors in relation to vascular and biliary anatomy, enhancing spatial perception during complex liver resections and supporting intraoperative decision-making through real-time anatomical overlays [13,15,16]. In colorectal liver metastasis surgery, AR-assisted navigation enabled rapid registration and successful resection of previously undetectable lesions while achieving clear surgical margins [16]. AR has also improved anatomical orientation during transthoracic minimally invasive liver resections, supporting greater surgical precision [17,18].

Despite these advances, AR remains limited by tissue deformation and organ movement, which can reduce registration accuracy during liver and pancreatic surgery [16]. Broader challenges include integration with existing imaging systems, interoperability across platforms, and incorporation into surgical workflows. Addressing these issues through robust registration methods and user-centered system design will be essential for wider clinical adoption [19].

Within the field of human-computer interaction (HCI), there is increasing recognition that the effective integration of AR into health care requires more than technical sophistication; it must also address the pragmatic needs, cognitive constraints, and interaction preferences of diverse user groups. Successful adoption of AR in clinical practice hinges on understanding how technology interacts with human behavior and complex health care environments [20]. Clinical technologies that are poorly integrated into health care workflows can increase cognitive workload, disrupt clinical processes, and reduce operational efficiency, highlighting the importance of designing systems that align with routine clinical practice [19,21].

Human-centered design approaches, particularly participatory and co-design methods, have therefore become increasingly important in the development of clinical AR systems. These approaches emphasize iterative development, contextual evaluation, and continuous involvement of end users, ensuring that systems reflect the perspectives of both health care professionals and patients [22]. In oncology, where clinical decision-making is complex and communication is central to patient care, co-design helps identify workflow requirements, usability barriers, and communication needs that may otherwise remain overlooked [23,24]. Furthermore, involving patients throughout the design process can improve usability, trust, and acceptance while supporting shared decision-making and more patient-centered care [25].

Designing AR systems for surgical oncology also requires careful consideration of the clinical environment. Interfaces must support intuitive interaction in sterile, time-critical settings while minimizing cognitive load and physical discomfort [26]. Persistent challenges, including gesture occlusion, lighting interference, and fatigue associated with head-mounted displays, continue to affect usability, although multimodal interaction combining gesture, voice, and gaze offers opportunities to improve user control and workflow integration [27].

Despite the growing application of AR in HPB surgery, the systematic incorporation of both health care professionals' and patients' perspectives throughout the design and evaluation of AR systems remains limited. Consequently, important workflow, communication, and usability requirements may not be fully addressed during system development. To address this gap, this study used a structured, iterative human-centered design approach to develop and evaluate an AR platform for preoperative planning and patient communication in HPB surgery. The aims of this study are to investigate how participatory co-design with health care professionals and

XSL•FO
RenderX

patients can inform the development of clinically relevant AR technologies, evaluate the platform's usability and integration within realistic clinical workflows, and derive design principles to support the future development and implementation of human-centered AR systems for surgical oncology.

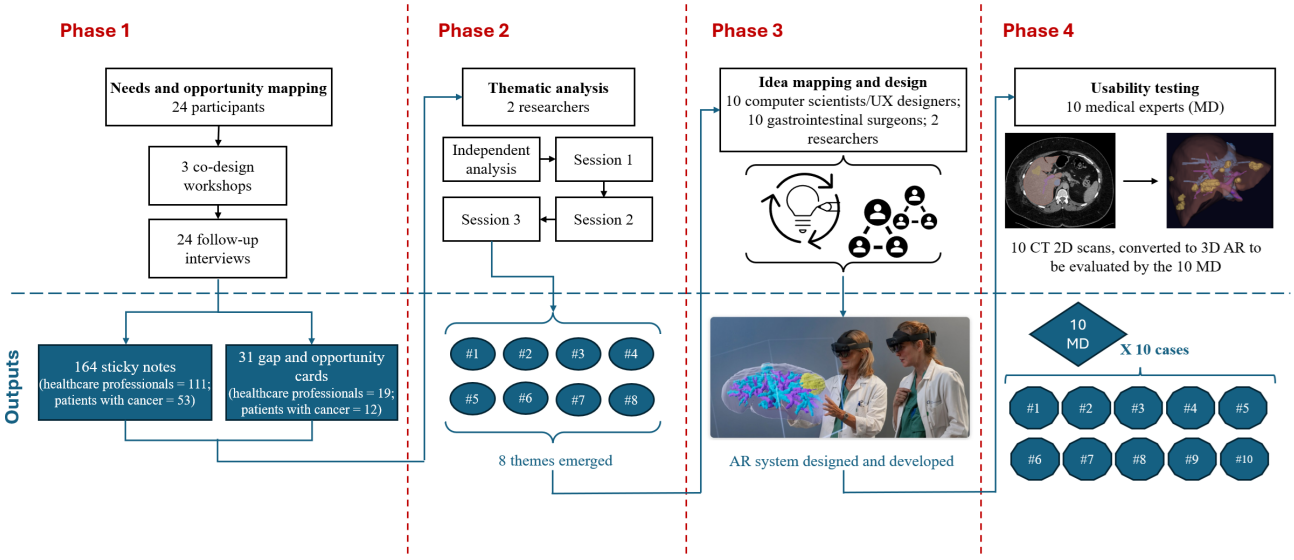
Methods

Study Design

This study integrates participatory design, immersive visualization, and clinical cocreation to explore how AR can support surgical oncology workflows, particularly in liver and pancreatic care. Grounded in principles of human-centered design [28] and participatory co-design [21], it investigates how interactive 3D anatomical models can enhance spatial understanding, interdisciplinary coordination, and communication in high-stakes surgical environments. The AR system was developed through a collaborative, iterative process within a complex sociotechnical health care ecosystem. Consistent with sociotechnical frameworks for health IT [29] and ergonomics-informed design [30], health care professionals and patients were engaged as active partners throughout the research process.

The study comprised 2 interconnected phases. Phase 1 focused on participatory co-design and iterative prototype development, while Phase 2 evaluated the usability of the resulting AR prototype in simulated multidisciplinary planning scenarios. The qualitative component of Phase 1 was guided by an interpretivist paradigm, recognizing that technology use and clinical decision-making are shaped by users' experiences and the contexts in which they occur. Semistructured interviews and co-design activities were therefore used to explore the cognitive, emotional, communicative, and workflow needs of health care professionals and patients with cancer. Figure 1 summarizes the overall study workflow.

Figure 1. To generate the catalog of ideas and guide the design of the augmented reality surgical oncology system, we followed 4 phases. In Phase 1, human-centered designers gathered data through co-design workshops and follow-up interviews. In Phase 2, we analyzed all collected materials and identified key cognitive, emotional, and communicative needs, as well as systemic gaps and opportunities in current practice. In Phase 3, we used these findings and participant-generated ideas to design and develop the augmented reality system. In Phase 4, we conducted usability testing with 10 medical experts specializing in hepatobiliary cancer treatment. AR: augmented reality; CT: computed tomography; MD: doctor of medicine; UX: user experience.



Phase 1: Co-Design and Development

Recruitment and Sampling

Phase 1 used convenience sampling to recruit participants with direct experience relevant to HPB oncology care, surgical planning, psychosocial cancer support, and patient communication. Recruitment was iterative, with participants recruited until the research team determined that the dataset provided sufficient breadth and depth of perspectives to inform the co-design process and no substantially new design requirements or experiential insights were emerging from successive workshops and interviews.

Recruitment for Phase 1 commenced in February 2024 through the project partner organizations, Leeds Teaching Hospitals NHS Trust, and the Cyprus Association of Cancer Patients and Friends. Separate invitation emails tailored to the respective participant groups were distributed via the organizations' professional mailing lists for health care professionals and patient mailing lists for patients with cancer. Health care professionals, including surgeons, oncologists, and clinical or counseling psychologists, as well as patients with cancer with relevant experience of surgical oncology care, who met the study eligibility criteria were invited to participate. Individuals who expressed interest by responding to the invitation email were subsequently contacted by the research team, provided with detailed information about the study, and invited to participate in the co-design workshops and semistructured interviews. Participation was entirely voluntary and initiated by individuals who responded to the invitation.

A total of 24 participants were enrolled in Phase 1, including oncology patients and a diverse cohort of clinical professionals specialized in oncology who were involved in the design and preliminary prototyping of the AR system. The clinical cohort (n=14) represented a wide range of roles across the oncology

XSL•FO
RenderX

care spectrum, including medical doctors specializing in oncology and/or general surgery, including cancer surgery (n=9), as well as clinical or counseling psychologists with specific expertise in cancer-related psychological support (n=5). Participants in this group ranged in age from 29 to 46 (mean 36.2, SD 5.6) years, and the gender distribution included 8 male and 6 female professionals. Their perspectives were considered particularly valuable in identifying cognitive, emotional, communicative, and workflow-related challenges associated with current cancer care practices.

The patient cohort (n=10) was previously treated for hepatobiliary, pancreatic, or other types of cancer surgery, including 4 who had specifically undergone hepatobiliary or pancreatic procedures within the previous 6 to 18 months. Their ages ranged from 35 to 55 (mean 49.8, SD 8.6) years and included male (n=7) and female (n=3) participants. All had direct experience with surgical consultations, preoperative imaging, and multidisciplinary cancer care pathways, positioning them to reflect on the clarity, accessibility, and emotional impact of both conventional and AR-enhanced communication approaches.

Procedure

The co-design activities were conducted across multidisciplinary oncology settings in Cyprus and the United Kingdom as part of the HoloSurge project. The workshops were facilitated by MM and M Krini, researchers with expertise in human-centered design, qualitative research, and stakeholder engagement. Recruitment commenced in February 2024, and the co-design workshops were conducted between March and May 2024. A total of 5 workshops were held, comprising 2 face-to-face workshops (with 4 and 5 participants, respectively) and 3 online workshops (5 participants per workshop). Participants were allocated to either the face-to-face or online workshops according to their geographical location, availability, and personal preference. This approach maximized participation while ensuring that discussions reflected authentic clinical and patient experiences across diverse health care contexts.

The co-design activities followed a structured design thinking approach [31] consisting of 4 sequential activities (refer to Multimedia Appendix 1 for the cocreative workshop materials). First, participants completed empathy mapping exercises to reflect on the thoughts, emotions, concerns, communication needs, and challenges experienced during surgical oncology care and multidisciplinary decision-making. Second, facilitated group reflection sessions encouraged participants to discuss and compare their experiences, identify shortcomings of current clinical workflows, and highlight unmet needs. Third, collaborative ideation activities enabled participants to propose and refine ideas for AR-supported functionalities, interfaces, and interaction techniques that could address the identified challenges. Finally, feature prioritization exercises required participants to evaluate the proposed ideas, discuss their feasibility and clinical value, and collectively identify the features considered most important for inclusion in the prototype.

Participants were encouraged to discuss challenges associated with current surgical planning workflows, communication practices, emotional support, multidisciplinary collaboration, and patient understanding of imaging data and treatment plans. Throughout the sessions, facilitators documented structured observational notes capturing participant interactions, nonverbal responses, group dynamics, and emergent discussion points. Workshop activities were supported by structured facilitation guides, including empathy mapping templates, collaborative ideation exercises, and feature prioritization matrices, while facilitators used a standardized observation template to document interactions and contextual observations. Workshop discussions were not audio-recorded. Instead, detailed facilitator field notes, structured observation templates, and participant-generated artifacts (eg, empathy maps and feature prioritization matrices) were collected during each workshop and used as qualitative data sources during analysis.

Following the workshops, MM and M Krini conducted semistructured interviews to explore participants' reflections and experiences in greater depth. The interviews were conducted between March and May 2024, following completion of each workshop, and were held either face-to-face or via Microsoft Teams, according to participants' geographical location and preference. Interviews were guided by a semistructured interview schedule comprising open-ended questions exploring participants' experiences of current communication challenges, expectations regarding AR integration into oncology workflows, emotional responses to anatomical visualization, and desired system functionalities (refer to Multimedia Appendix 1 for the interview guides). As expected within an iterative human-centered design process, the emphasis of the interview questions was refined between successive sessions to explore emerging themes and design requirements identified during earlier workshops while maintaining consistency with the overall study objectives. All interviews were audio-recorded with participants' consent and transcribed verbatim prior to analysis.

Data Analysis

Qualitative data sources included workshop discussions, semistructured interviews, facilitator observational notes, and participant-generated artifacts, including empathy maps and feature prioritization matrices. Qualitative data were analyzed using reflexive thematic analysis (RTA) [32]. This analytic approach was selected because of its suitability for examining experiential, emotional, and interpretive dimensions across diverse stakeholder groups while acknowledging the active role of researchers in knowledge generation. Two members of the multidisciplinary research team (MM and M Krini) independently coded the qualitative data, drawing on their complementary expertise in HCI, oncology care, psychology, and clinical communication. The analysis involved iterative familiarization with the dataset, generation of initial codes, development and refinement of themes, and interpretive synthesis. Following independent coding, the researchers engaged in reflexive discussions to compare interpretations, explore alternative explanations, and iteratively refine the developing themes. Any differences in interpretation were resolved through discussion, with themes further reviewed by the wider multidisciplinary research team to ensure that they reflected the breadth and depth of participants' perspectives rather than to achieve coding agreement.

XSL•FO
RenderX

The critical incident technique [33] was additionally used to identify moments of heightened salience during workshops and interviews, including expressions of cognitive burden, emotional distress, uncertainty, communication breakdowns, or shifts in group consensus. These incidents were used to support interpretation of key experiential and interactional challenges relevant to AR system design.

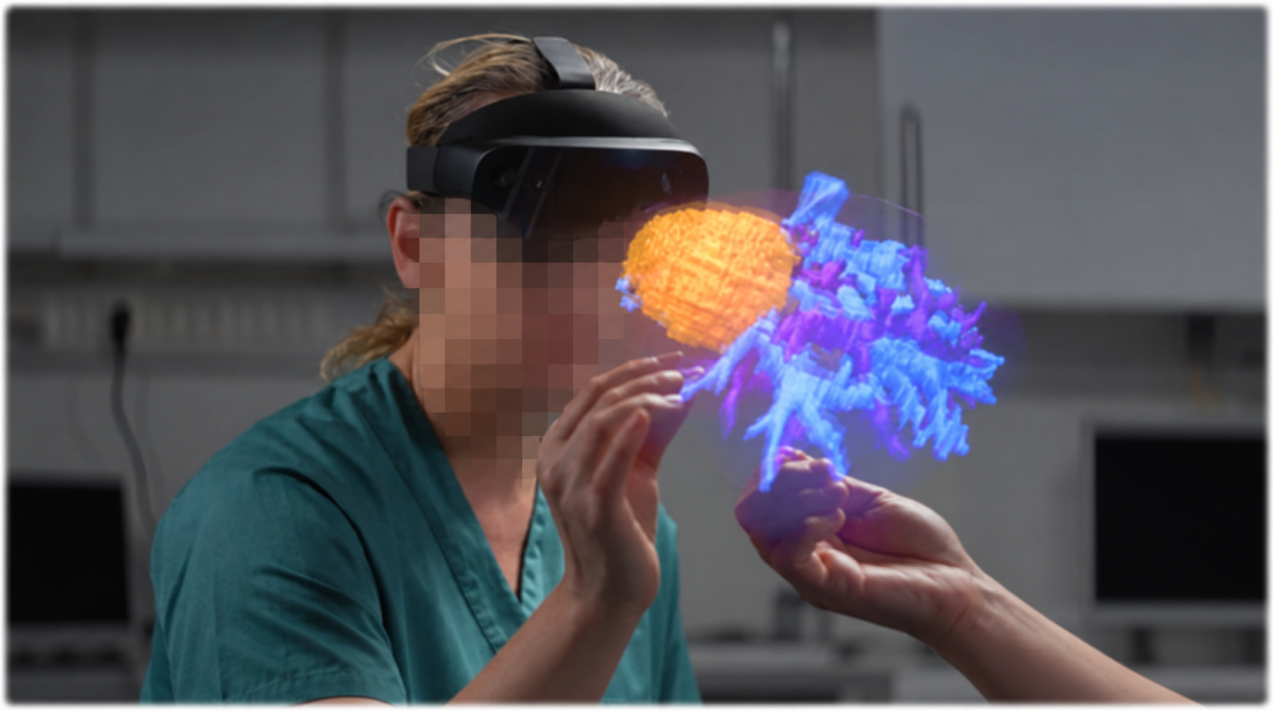
The multidisciplinary research team comprised researchers with expertise in HCI, oncology care, psychology, and clinical communication. As researchers involved in the co-design and evaluation of the AR system, we acknowledge that our disciplinary backgrounds and prior experience could influence data collection and interpretation. Reflexive discussions were therefore conducted throughout the analytical process to critically examine assumptions, identify potential disciplinary or experiential biases, and consider alternative interpretations. Credibility was enhanced through triangulation across multiple stakeholder groups (health care professionals and patients with cancer) and complementary data sources, including co-design workshops, semistructured interviews, facilitator observational notes, and participant-generated artifacts. Confirmability was supported through iterative refinement of themes and regular multidisciplinary team discussions throughout the analytical process.

Phase 2: Usability Evaluation

System Design and Development

The AR prototype was developed using a human-centered design approach informed by the findings of the co-design workshops (Phase 1). The design prioritized accurate spatial visualization, intuitive sterile-compatible interaction, and clear anatomical representation to support HPB surgical planning and multidisciplinary collaboration. Figure 2 illustrates the AR system in use during a simulated surgical planning session.

Figure 2. Depicts the augmented reality system in use during surgical planning.



The system was developed in Unity 3D using the Microsoft Mixed Reality Toolkit and deployed on the Microsoft HoloLens 2, an optical see-through head-mounted display featuring spatial mapping, hand tracking, gaze tracking, voice commands, and 6 degrees-of-freedom tracking. Patient-specific 3D anatomical models were generated from contrast-enhanced CT (portal venous phase) DICOM datasets using the HoloCare Studio 2 segmentation pipeline and integrated into the Unity-based HoloSurge application.

The prototype enabled health care professionals to inspect holographic anatomical reconstructions from multiple viewpoints, manipulate models through midair hand gestures (rotation, scaling, and repositioning), isolate anatomical structures (eg, tumors and vessels), and use contextual annotations and anatomical labels during collaborative surgical planning. Color coding, visual hierarchy, and gesture-based interaction were incorporated to improve interpretability and reduce cognitive workload.

The interface was iteratively refined through design walkthroughs with surgeons and radiologists, leading to improvements in interaction stability, anatomical labeling, and overall usability. During the second phase, participants individually used a single HoloLens 2 while other participants observed the mirrored headset view on an external display. Although the prototype was designed primarily for health care professionals in multidisciplinary planning, its architecture supports future multiuser and patient-facing collaborative visualization. To ensure consistent evaluation conditions, all sessions were conducted offline, with pilot testing confirming smooth real-time rendering and interaction performance.

XSL•FO
RenderX

Recruitment and Sampling

Phase 2 focused on evaluating the usability of the AR system within simulated multidisciplinary preoperative planning scenarios involving colorectal liver metastasis (CRLM) cases, a condition in which colorectal cancer spreads to the liver. Convenience sampling was used to recruit health care professionals with direct expertise in hepatobiliary cancer treatment and multidisciplinary tumor board decision-making, as these participants represented the intended end users of the system and were best positioned to evaluate its clinical usability and perceived utility within realistic preoperative planning scenarios. Recruitment continued until a multidisciplinary sample representing the key clinical roles involved in hepatobiliary surgical planning had been achieved, and the usability evaluation generated sufficient feedback to inform iterative refinement of the prototype.

Eligible clinicians were recruited through Leeds Teaching Hospitals NHS Trust using professional and institutional mailing lists. Invitation emails describing the study were circulated to health care professionals involved in hepatobiliary cancer care and multidisciplinary tumor board meetings. Clinicians who expressed interest were subsequently contacted by the research team, provided with detailed study information, and invited to participate in the usability evaluation. Participation was voluntary, and all participants provided informed consent prior to enrollment.

Ten medical experts participated in this evaluation, each with a professional background directly related to hepatobiliary cancer treatment. This included consultant surgeons (n=5), 2 with fewer than 5 years of consultant experience and 3 with more than 5 years in senior roles, surgical trainees (n=3), a radiology trainee (n=1), and one consultant radiologist (n=1) with more than 5 years of specialization in hepatobiliary imaging.

Participants were male (9/10, 90%) and female (1/10, 10%) clinicians. All were affiliated with high-volume surgical centers and regularly contributed to tumor board discussions and operative planning sessions. While all participants were highly experienced in their respective domains, none had previous exposure to head-mounted AR systems prior to the evaluation, which allowed for a neutral assessment of usability and perceived utility. Sessions were conducted over a 4-week period and held in person at Leeds Teaching Hospitals NHS Trust, Leeds, United Kingdom, with one session held each week.

Procedure

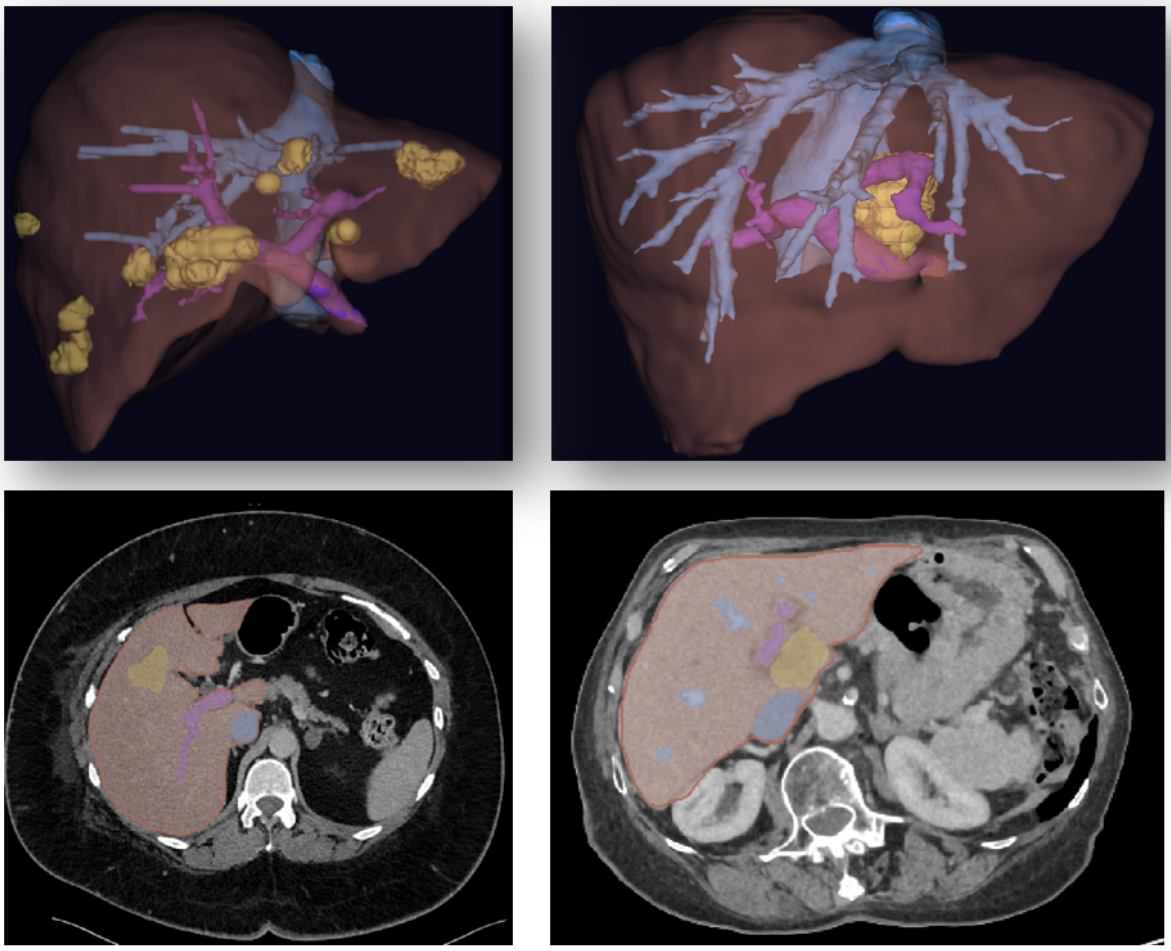
The usability evaluation was conducted in person at the Leeds Teaching Hospitals NHS Trust, United Kingdom, within a clinical simulation environment designed to closely reflect multidisciplinary tumor board planning and preoperative decision-making in hepatobiliary oncology. This setting was selected to maximize ecological validity by enabling participants to interact with the AR system under conditions representative of routine clinical practice while maintaining a controlled environment for usability evaluation.

Participants first received a brief orientation to the AR hardware, interface, and interaction techniques prior to beginning the evaluation tasks. During each session, participants reviewed approximately 2 - 3 clinical cases derived from portal venous phase CT datasets and explored interactive 3D holographic anatomical reconstructions using gesture-based interaction.

Each clinician was introduced to 2 or more patient cases selected from a total dataset of 10 CRLM cases that had been processed into interactive 3D anatomical models. Using the Microsoft HoloLens 2 headset, participants manipulated holographic structures through midair gestures, including rotation, scaling, repositioning, and spatial exploration of tumors, vascular anatomy, and surrounding liver structures to support clinical interpretation and surgical planning. These evaluations occurred in settings that closely mirrored multidisciplinary planning meetings, also known as tumor board meetings, where specialists collaboratively discuss cancer treatment decisions and clinical simulation environments. Examples of the cases are displayed in Figure 3.

XSL•FO
RenderX

Figure 3. Images from liver surgery planning, 2D images of computed tomography (CT) scans with anatomical annotations and 3D reconstructions of colorectal liver metastases (CRLMs) using portal venous phase CT images (blue: hepatic veins; purple: portal veins; yellow: CRLM). (A) Ten small CRLMs threatening future liver remnant; (B) segment 1 CRLM in close proximity to hepatic veins and portal veins.



Sessions lasted up to 20 minutes and were facilitated by members of the research team with expertise in surgical devices and technology assessment. A standardized usability evaluation protocol was followed for all participants. Facilitators documented interaction behaviors, usability challenges, technical issues, and emergent feedback using structured observational notes. The usability protocol, observational procedures, and System Usability Scale (SUS) questionnaire remained unchanged throughout the study to ensure consistency across all evaluation sessions.

SUS Administration and Analysis

Following completion of the AR interaction tasks, participants completed the SUS [34], a validated 10-item questionnaire widely used to assess perceived usability of digital systems and health care technologies. The questionnaire was administered in paper format immediately following each usability evaluation session and was completed independently by participants in the presence of a member of the research team, who was available to clarify procedural questions but did not influence participants' responses. Participants rated each statement on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). SUS scores were calculated according to the standard scoring procedure, producing a total usability score ranging from 0 to 100, with higher scores indicating greater perceived usability. Scores for early-stage prototypes often range between 60 and 70 [35].

Ethical Considerations

Ethical approval was obtained from the National Bioethics Committee (reference number: EEBK EP 2024/56) for Phase 1. All participants provided written informed consent prior to participation. Participation was voluntary, and participants could withdraw at any time without consequence. The Information Governance Team of the Leeds Teaching Hospitals NHS Trust, Leeds, United Kingdom, approved the service evaluation in Phase 2. The AR system used existing imaging data that had been acquired as part of routine clinical care. Clinician participation in Phase 2 was voluntary, and participants could withdraw at any time without consequence.

All collected data were anonymized and deidentified prior to analysis. Participant information was securely stored on password-protected institutional systems accessible only to the research team. No personally identifiable information was included in the analysis or reporting of findings. Participants

XSL•FO
RenderX

did not receive financial compensation for participation in either study.

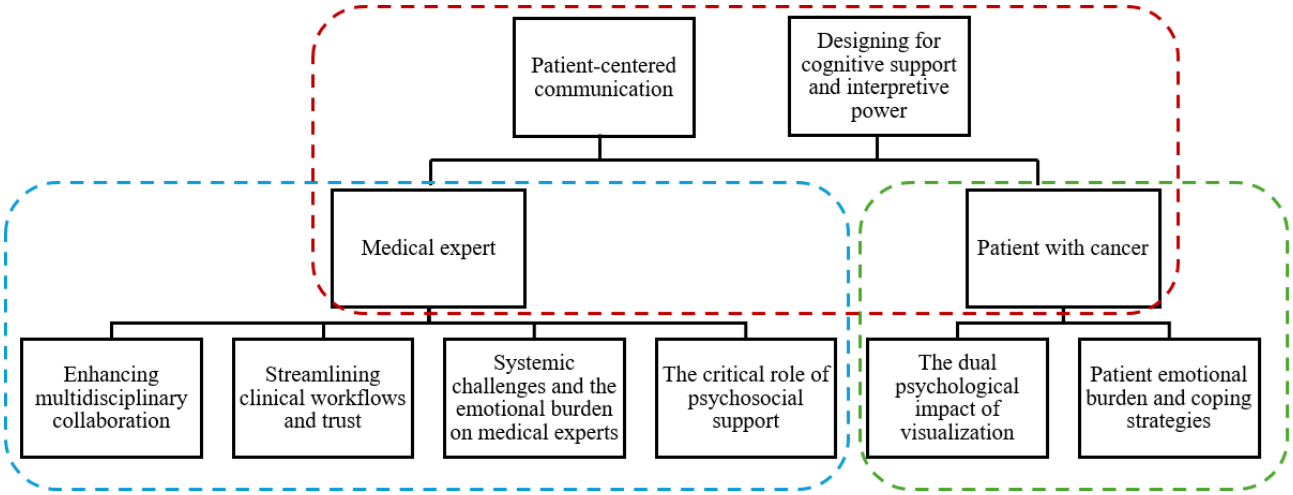
Results

Phase 1: Co-Design and Development

Overview

Through a structured co-design process involving 14 oncology professionals and 10 patients with cancer, a thematic analysis of qualitative insights identified 8 core themes that informed the development of the AR system. These themes reflect diverse yet intersecting perspectives and needs, shaping a user-centered framework for AR in cancer care. Two themes—(1) patient-centered communication and (2) designing for cognitive support and interpretive power—emerged from both groups, reflecting shared priorities in improving communication and understanding within cancer care. From the medical professionals' side, four additional themes emerged: (3) enhancing multidisciplinary collaboration**,** (4) streamlining clinical workflows and trust, (5) systemic challenges and the emotional burden on medical experts**,** and (6) the critical role of psychosocial support. These themes reflect health care professionals' core needs for efficiency, clarity, emotional sustainability, and seamless teamwork in high-pressure clinical environments. On the patient side, two unique themes were identified: (7) the dual psychological impact of visualization and (8) patient emotional burden and coping strategies. These capture patients' fundamental needs for emotional safety, autonomy, and accessible understanding throughout their cancer journey. Figure 4 visually summarizes these 8 themes and clearly indicates their sources, whether from health care professionals, patients, or both, highlighting the integrative and human-centered foundation of the AR system's design.

Figure 4. Schematic representation of the 8 core themes, illustrating the unique and shared needs of oncology professionals and patients with cancer that informed the human-centered design of the augmented reality system.



Patient-Centered Communication

Our findings reveal significant communicative and emotional challenges that patients face when trying to understand their diagnoses and proposed surgical plans. Many reported difficulties translating 2D imaging data, such as CT and MRI scans, into a meaningful spatial understanding of their anatomy, an issue well-documented in the literature as a cognitively demanding task for nonexperts [8]. One (patient 2, interview) patient reflected, "The doctor would show me the scans, and I'd nod along, but I really had no idea what I was looking at."

Patients frequently described pretending to understand explanations despite remaining uncertain, while clinicians similarly acknowledged that patients often avoid asking questions or admitting confusion (patient 5, workshop; surgeon 4, workshop). These accounts reveal an epistemic asymmetry within clinical encounters that limits genuine shared decision-making.

Patients also described frustration with existing communication practices, reporting that explanations were often unclear, inconsistent, or lacked sufficient emotional support (patient 8, workshop). These findings are consistent with previous research demonstrating that health care professionals and patients with cancer frequently fail to discuss treatment goals and clinical end points in sufficient depth, leaving many patients uncertain about the purpose and expected outcomes of treatment [36,37]. Psychologists further emphasized that effective communication extends beyond providing technical information and should also foster psychological safety and confidence throughout the decision-making process (counseling psychologist 2, workshop).

The co-design workshops identified AR as a potential solution for bridging this communication gap by providing a shared visual reference that supports both understanding and dialogue. Patients explained that unclear explanations often discouraged them from asking follow-up questions because of embarrassment or uncertainty. This aligns with evidence that many patients with cancer continue to have unanswered questions following clinical consultations [38], while psychological distress can further reduce comprehension, particularly during discussions of difficult prognoses [39]. Integrating such insights into prognostic discussions could not only improve mental well-being, but also empower patients to make informed, value-aligned decisions [39].

During the workshop, both patients and clinicians strongly endorsed the use of AR to facilitate clearer, more equitable communication. All medical experts in the study (14/14) agreed

XSL•FO
RenderX

that 3D models would enhance patient understanding, and nearly all patients (9/10) said such visual explanations would be "very useful" (patients 1‐5 and 7‐10, workshop) for explaining procedures. This is consistent with previous studies highlighting AR's promise in improving physician-patient communication [3]. As one surgeon (surgeon 6, workshop) stated, "Simple diagrams aren't enough, we need better visuals to explain surgical procedures."

Our findings extend this evidence by showing how AR can redistribute interpretive power, transforming patients from passive recipients into active participants in their care. One patient (patient 9, workshop) expressed this desire plainly, "I want to be treated like a person, not just a case."

These preferences align with human-centered design principles that advocate for patient-focused communication tools [25]. By enabling patients and health care professionals to manipulate and explore the same holographic model together, the AR system transforms 1-way explanations into active, collaborative dialogues.

Designing for Cognitive Support and Interpretive Power

Across both patient and health care professional groups, participants consistently emphasized the need for tools that reduce cognitive load in complex, high-pressure situations. Consistent with previous research [13], they viewed AR as having strong potential to support spatial reasoning and scenario planning, provided that the system is intuitive, modular, and context-aware. As one surgeon (surgeon 3, interview) explained, "In the operating room, I don't have time to figure things out. If it's not immediate and obvious, I won't use it, no matter how impressive it looks."

These findings highlight that usability and cognitive efficiency are prerequisites for successful clinical adoption.

Patients expressed similar expectations from a nonexpert perspective, emphasizing that technology should simplify rather than complicate their understanding of their condition and treatment (patient 6, workshop). Participants described becoming overwhelmed by overly technical explanations and stressed the importance of clear, accessible communication supported by appropriate guidance (patient 6, workshop). One patient (patient 2, interview) summarized this need by stating, "I want to understand what's going on, but if it's too technical or overwhelming, I just shut down."

These findings suggest that cognitive accessibility should be a fundamental design principle, supporting both patient understanding and clinical decision-making [26].

Rather than functioning solely as a visualizer, the AR system should act as an interpretive companion, one that helps users assess options, understand implications, and engage in meaningful deliberation. Based on this finding, AR systems should be developed with this in mind, incorporating interactive features that enable comparative reasoning, gesture-driven spatial manipulation, and adjustable interface layers that reveal or hide detail according to cognitive demand [27]. This approach supports a broader reconceptualization of AR, positioning it not just as a tool for seeing, but as a platform for thinking and communicating within complex clinical ecosystems. In this way, the system becomes a cognitive framework that enhances both clinical reasoning and patient understanding [26].

Enhancing Multidisciplinary Collaboration

While individual understanding is essential, as highlighted in the previous theme, this cognitive support becomes even more critical in team-based care environments. Clinical professionals, including surgeons, oncologists, and mental health specialists, identified significant cognitive load and communication barriers in existing multidisciplinary team meetings. They described the mental challenge of reconstructing anatomical relationships from 2D imaging, a well-recognized difficulty in hepatobiliary surgery [7]. As one surgeon (surgeon 1, interview) explained, "Current meetings are often a series of people looking at a flat screen and trying to mentally reconstruct the anatomy. It's time-consuming and leaves room for miscommunication." As an oncologist (oncologist 3, workshop) explained, "Overall, participants viewed multidisciplinary collaboration as fundamental for improving communication, reducing ambiguity, and supporting coordinated clinical decision-making."

These communication challenges extended beyond surgical teams. Psychologists described difficulties understanding treatment plans and surgical pathways, limiting their ability to respond to patients' concerns or provide individualized psychosocial support. One counseling psychologist (counseling psychologist 1, interview) noted, "Patients sometimes want to talk about the procedure and express fears about whether the tumor was completely removed. I want to support them in that moment, but I often don't understand the surgical pathway well enough to respond meaningfully."

Participants further explained that organizational barriers, including bureaucracy and limited integration into clinical discussions (surgeon 1 and 2, workshop; oncologist 2, workshop), frequently restricted psychologists' involvement (patient 4 and 9, workshops; clinical psychologist 2, workshop), despite evidence psychological support is essential for addressing anxiety, uncertainty, and fear of recurrence, all of which influence treatment adherence and quality of life [40,41].

Across professional groups, participants identified the need for tools that facilitate shared understanding rather than discipline-specific communication. Rather than simply visualizing anatomy, they envisioned AR as a collaborative platform capable of making complex surgical information accessible to all members of the care team, including psychosocial professionals. As one psychologist (clinical psychologist 1, interview) explained, "We're often the ones patients turn to after the medical explanation, and it's difficult to offer meaningful reassurance when we ourselves don't fully understand the treatment plan or the rationale behind it."

Consistent with previous research demonstrating AR's value in improving spatial understanding during oncologic surgery [13], participants suggested that shared interactive 3D visualizations could enhance multidisciplinary communication, strengthen collaborative decision-making, and improve continuity of patient care.

XSL•FO
RenderX

These user-driven insights informed the AR system's core functionalities, including collaborative viewing and gesture-based interaction. Given that fragmented communication is a known contributor to medical error, AR has the potential to serve as a safety-enhancing platform for aligning interprofessional perspectives. This extends existing research on AR in surgical planning [16] by positioning it not only as a technical tool, but as a collaborative artifact for real-time interdisciplinary dialogue, where all team members can simultaneously view and interact with the same 3D model.

Streamlining Clinical Workflows and Building Trust

Medical experts emphasized that successful adoption of AR depends on its seamless integration into existing clinical workflows. Participants consistently highlighted the need for systems that are intuitive, efficient, and compatible with sterile surgical environments, where time pressure and cognitive demands are high (oncologist 1, workshop; surgeon 5, workshop). These findings align with previous research demonstrating that AR adoption is facilitated by technologies that accommodate the realities of health care practice, support intuitive interaction under sterile conditions, and minimize cognitive load [19,26]. Beyond efficient visualization, participants envisioned AR as an intelligent workflow assistant capable of providing context-sensitive support throughout surgical procedures. As one surgeon (surgeon 4, interview) explained, "A helpful system should track where I am in the procedure and respond intelligently, not just display anatomy."

This perspective reflects growing interest in adaptive AR systems that interpret procedural context and deliver tailored guidance to support clinical decision-making during surgery [19].

Trust emerged as another critical requirement for clinical adoption. Participants emphasized that confidence in AR depends not only on the accuracy of anatomical visualizations but also on transparency regarding the system's limitations. Concerns regarding model registration errors and false precision, particularly when operating on deformable tissues [16], highlighted the need for interfaces that clearly communicate uncertainty. As one oncologist (oncologist 2, interview) stated, "If I don't know where the model stops being accurate, I won't use it to make a decision."

These findings directly informed the inclusion of confidence overlays and provenance panels that communicate source data, precision limits, and uncertainty, consistent with recommendations for improving clinician trust in AR-based surgical systems [19].

Participants also emphasized that technological trust should complement, rather than replace, interpersonal trust between clinicians and patients. As one surgeon (surgeon 3, workshop) noted, "The patient's trust must be earned. No technology can replace that."

This underscores that AR should be viewed as a support tool rather than a substitute for the expertise, empathy, and interpersonal connection provided by health care professionals. The success of AR in surgical contexts therefore depends not only on its technical accuracy, but also on its seamless integration into existing workflows, its ability to adapt to procedural stages, and its role in reinforcing, rather than replacing, the clinician-patient relationship.

Systemic Challenges and the Emotional Burden on Medical Experts

Medical professionals consistently described systemic inefficiencies and ethical pressures as significant challenges affecting both patient care and professional well-being. Participants highlighted concerns related to rising costs, administrative burden, resource limitations, and broader health care structures that constrained the delivery of patient-centered care (oncologist 2, workshop; surgeon 1, workshop). As one participant (surgeon 6, workshop) remarked, "Care is top notch, but it's expensive and riddled with systemic inefficiencies."

These challenges extend beyond the clinical setting into the structural realities of the health care system.

Similar concerns were expressed across professional groups, including psychologists (counseling psychologist 2, workshop), indicating that emotional fatigue and fear of error are shared experiences throughout multidisciplinary cancer care (oncologist 1, 2, and 3, workshop; surgeon 3, workshop). These findings extend existing usability literature by demonstrating that technologies intended for clinical practice should address not only functional and cognitive demands but also the emotional context in which decisions are made [21]. Participants therefore envisioned AR as more than a technical visualization tool, suggesting features such as reflective checkpoints, private annotations, and postoperative debriefing functions that could support clinical reflection and psychological resilience following complex decisions.

Participants also emphasized that future technologies should help alleviate systemic challenges while preserving the human aspects of care. They envisioned intelligent systems capable of supporting personalized treatment planning, anticipating complications, improving workflow efficiency, and assessing patient understanding without replacing clinician judgment or interpersonal relationships (surgeon 5, workshop; oncologist 2, workshop). As one surgeon (surgeon 3, workshop) explained, "...an AI for personalizing treatment and predicting complications, but not for replacing human connection."

Such perspectives position AR and AI as supportive partners in care that can reduce workload, facilitate communication, and strengthen decision-making while maintaining the central role of human expertise and compassion [19].

The Critical Role of Psychosocial Support

Psychologists consistently emphasized that effective cancer care extends beyond medical treatment, requiring a holistic approach that actively addresses patients' psychological and emotional needs. Participants viewed psychosocial support as a core component of cancer care rather than an adjunct to clinical treatment, highlighting that emotional processing, individualized care, and psychoeducation are fundamental to helping patients navigate their illness (clinical psychologist 2, workshop). As one psychologist (counseling psychologist 1, workshop) explained, "It's about creating space for emotional processing."

XSL•FO
RenderX

This perspective aligns with previous research demonstrating that successful implementation of clinical technologies requires careful consideration of users' cognitive needs, emotional state, and interaction preferences [20].

Participants further stressed that psychosocial care should be tailored to each patient's emotional readiness and should evolve throughout the cancer journey, from diagnosis to rehabilitation (clinical psychologist 2, workshop). Rather than adopting standardized approaches, psychologists described adjusting communication and therapeutic support according to patients' knowledge, concerns, and capacity to process information (clinical psychologist 3, workshop). This patient-centered perspective reflects broader human-centered design principles that promote shared decision-making by recognizing individual needs and preferences [25].

Psychologists also described the considerable emotional burden associated with their professional role. Participants reported experiencing guilt, frustration, and fear when time pressures, organizational barriers, or limited involvement in clinical decision-making prevented them from providing the level of support they considered necessary (clinical psychologist 2 and 3, workshop; counseling psychologist 2, workshop). One participant (counseling psychologist 1, workshop) reflected, "It's hard to handle emotionally when patients pass away without understanding their situation."

These concerns suggest that emotional strain is influenced not only by patient distress but also by systemic barriers that limit meaningful psychosocial engagement, consistent with evidence showing that poorly designed clinical systems can amplify professional workload and emotional fatigue [21].

Consequently, participants viewed technology as a means of strengthening rather than replacing the human aspects of cancer care. They emphasized the need for tools that facilitate communication, improve patient understanding, and create more opportunities for meaningful clinician-patient interactions. As one psychologist stated, "enhance, not replace, the human side of care" (counseling psychologist 2, workshop). Participants also highlighted the value of patient-facing communication tools that promote understanding and engagement throughout the care pathway (clinical psychologist 1; counseling psychologist 1 and 2, workshop). These insights reinforce the importance of designing technologies that support collaborative, empathetic, and patient-centered care while preserving the central role of human relationships in clinical practice [23].

In the context of the proposed AR system, these findings informed the inclusion of features that support psychosocial care, including shared visualizations for patient-clinician discussions, accessible explanations of procedures, and communication tools that adapt to patients' emotional readiness. Rather than functioning solely as a cognitive aid for clinical decision-making, the proposed AR system was envisioned as a platform that also strengthens trust, communication, and empathetic engagement throughout the cancer journey.

The Dual Psychological Impact of Visualization

Patient participants expressed markedly different emotional responses to visualizing their internal anatomy. For some, gaining visual clarity fostered a sense of empowerment; for others, confronting the graphic reality of their condition was distressing. As one participant (patient 6, workshop) explained, "I wouldn't want to see much detail; sometimes ignorance is bliss."

In contrast, another (patient 4, interview) reflected, "If I had seen the tumor and where they would cut, I would've asked more questions. I was scared because I didn't know."

These contrasting perspectives demonstrate that patients differ not in whether they value visual information, but in the amount, timing, and manner in which it is presented.

These perspectives highlight the need for visualizations to be optional, personalized, and emotionally calibrated. The contrasting views informed our design principle of optionality and personalization. AR interfaces for patients must enable modulation of exposure, staged delivery of information, and support for diverse communication preferences (eg, visual, verbal, or simplified icon-based formats). This approach aligns with best practices in inclusive interaction design, which emphasize supporting collaborative and equitable decision-making [22,23].

Participants also highlighted the importance of being able to explore holographic visualizations at their own pace within a psychologically supportive environment (patient 1, interview), a concept reflected in affective scaffolding [42]. These findings informed several key design elements, including step-by-step walkthroughs, interactive annotation tools, adjustable levels of visual detail, and customizable emotional tone settings. Addressing emotional alongside informational needs is essential, as successful implementation of AR depends not only on technical sophistication but also on accommodating users' cognitive constraints and interaction preferences [20]. By embedding emotional sensitivity into the interface, the proposed AR system supports trust, understanding, and patient autonomy while reducing the risk of emotional overload.

Patient Emotional Burden and Coping Strategies

Patient participants described cancer as a profound disruption to their daily lives and personal identities, creating an ongoing struggle to regain a sense of normalcy while adapting to life after diagnosis. Many expressed a desire to return to life before cancer, free from fear and uncertainty. As one participant (patient 2, interview) explained, "...to feel normal again, like before. To be healthy, and not afraid."

This aspiration is consistent with previous literature, which shows that many patients seek to "get back to normal" following treatment, although the meaning of normality varies between individuals, ranging from restoring previous routines to adapting to a "new normal" shaped by their cancer experience [43,44].

Fear of recurrence emerged as a persistent but often privately managed concern. Participants described coping by suppressing these fears, accepting cancer as part of their identity, or prioritizing the emotional well-being of their families over their own. One patient reflected, "I need to be strong for my family" (patient 6, interview). These findings suggest that many patients rely on self-management and informal support networks while

XSL•FO
RenderX

avoiding open discussion of their emotional struggles. Such coping strategies are consistent with literature describing self-reliance, fear of stigma, and concerns about burdening others as barriers to seeking psychological support [45-47].

These experiences reflect the concept of biographical disruption, where a cancer diagnosis threatens an individual's identity, mortality, and assumptions about the future, creating an existential rupture [48]. Addressing such disruption requires systems that recognize not only the clinical aspects of care but also the emotional and psychological burdens patients carry. Involving patients in the design process is essential to surface these often-unspoken needs and to ensure that technology acknowledges the full scope of their lived experience [25].

Patients also highlighted the importance of technologies that provide emotional support without increasing distress. Rather than replacing human care, they envisioned systems that improve understanding, facilitate communication, and connect patients with appropriate support when needed. As one participant (patient 8, workshop) stated, "Technology could help, especially if it connects me to care when needed."

Alongside these challenges, some participants also expressed gratitude for surviving cancer and for the care they had received, illustrating that resilience and appreciation coexisted with ongoing uncertainty (patient 10, interview). Participants emphasized that visual explanations, clearer communication, peer support, and access to psychological services could all contribute to reducing anxiety and improving understanding throughout the treatment journey.

These findings support a human-centered approach in which technology complements, rather than replaces, existing models of care by facilitating collaborative and equitable decision-making [23]. They also reinforce evidence that identity- and existence-related concerns remain insufficiently addressed during routine cancer care [49]. Consequently, participants identified several design priorities for the proposed AR system, including adjustable levels of visual detail, patient-paced interactions, and integrated access to psychosocial resources. Together, these features position AR as part of a broader support ecosystem that addresses both the informational and emotional challenges associated with living with and beyond cancer.

Phase 2: Usability Evaluation

The average SUS score across participants was 67.25 (SD 7.95; n=10), indicating moderate perceived usability. Qualitative feedback provided further context for these findings. Participants consistently identified the AR system as valuable for improving spatial understanding during complex surgical planning and facilitating shared understanding among multidisciplinary teams. As one surgeon (surgeon 2, interview) explained, "Being able to rotate and inspect the anatomy in 3D makes it much easier to understand the tumor's relationship to the vessels compared to looking at flat CT images."

These findings suggest that the system's principal strength lies in supporting visualization and collaborative clinical reasoning. Participants also identified several opportunities for improvement, including clearer anatomical labeling, more streamlined interaction workflows, and faster navigation between patient cases to better support high-volume multidisciplinary meetings.

Discussion

Principal Findings

This study investigated how participatory co-design with health care professionals and patients with cancer can inform the development of a human-centered AR platform for HPB surgical oncology, evaluated its usability within simulated multidisciplinary planning scenarios, and derived design principles to support the future development of AR systems for surgical oncology. The co-design process identified key cognitive, communicative, emotional, and workflow-related requirements that directly informed prototype development. The resulting AR platform demonstrated encouraging usability for an early-stage prototype while identifying key priorities for further refinement, including gesture recognition stability, anatomical labeling clarity, and workflow responsiveness. These findings provide evidence that participatory co-design can support the development of clinically relevant AR technologies that are better aligned with multidisciplinary workflows and patient-centered communication.

Our findings reinforce previous evidence demonstrating that AR can improve spatial understanding, facilitate surgical planning, and support intraoperative navigation through interactive 3D visualization of patient-specific anatomy [13]. However, unlike much of the existing literature, which has primarily focused on technical feasibility, our study adopted a human-centered approach that positioned health care professionals and patients as active contributors throughout the design process. The co-design process demonstrated that clinically relevant AR systems must address not only technical requirements but also cognitive, communicative, and psychosocial considerations that influence successful implementation within routine multidisciplinary oncology care. It is worth mentioning that, although conducted within HPB oncology, many of the design principles identified, including shared visualization, collaborative decision-making, workflow integration, and patient-centered communication, may also be transferable to other image-guided surgical specialties.

Health care professionals reported that immersive 3D models reduced the cognitive effort required to mentally reconstruct anatomy from conventional CT images, thereby supporting multidisciplinary tumor board discussions and surgical planning. Similarly, patients described improved understanding of their anatomy and planned procedures through intuitive spatial visualization. These findings are consistent with previous research showing that interpreting complex anatomical relationships from 2D imaging places considerable cognitive demands on clinicians, particularly in time-pressured environments [7,8]. Our findings suggest that AR may function as a shared cognitive workspace that facilitates communication, promotes shared understanding, and supports collaborative clinical decision-making, consistent with theories of distributed cognition in health care [50].

XSL•FO
RenderX

Importantly, our findings also suggest that AR may function as a boundary object within multidisciplinary oncology care. A shared holographic model provided a common point of reference that enabled different professional groups, including surgeons, radiologists, and psychologists, to interpret the same patient-specific anatomy according to their respective clinical perspectives while maintaining a shared understanding of the case. Participants emphasized that an AR shared visualization could improve multidisciplinary discussions, facilitate communication with patients, and better integrate psychosocial considerations into treatment planning. These findings extend the role of AR beyond a visualization tool, positioning it as a collaborative platform capable of supporting both technical decision-making and patient-centered cancer care.

The co-design process further highlighted the importance of emotional adaptation in patient-facing AR systems. While many participants viewed anatomical visualization as reassuring and empowering, others described feelings of anxiety or discomfort when confronted with detailed representations of their disease. These findings are consistent with previous studies of visual risk communication in oncology, which emphasize that the presentation of medical information can influence emotional responses and patient understanding [51]. Our findings therefore suggest that successful implementation of AR in oncology requires not only accurate anatomical visualization but also flexible approaches that allow information to be tailored according to patient preferences, emotional readiness, and clinical context.

Finally, to address the second aim of this study, the usability evaluation demonstrated acceptable usability (SUS score=67.25), providing encouraging evidence that the prototype can support clinically representative multidisciplinary planning workflows while identifying priorities for further refinement [52]. Qualitative feedback identified gesture recognition stability, clearer anatomical labeling, reduced loading times, and improved workflow responsiveness as the principal priorities for future refinement. These findings align with previous evidence demonstrating that workflow compatibility and intuitive interaction are critical determinants of successful implementation of surgical technologies in routine clinical practice [27].

Limitations and Future Directions

Several limitations should be considered when interpreting these findings. First, the study focused specifically on HPB oncology, and although many of the identified design requirements may be transferable to other image-guided surgical specialties, additional validation across broader clinical contexts is required. Second, the relatively small participant sample, particularly during the usability evaluation phase, may limit the generalizability of the findings. In addition, the prototype was evaluated within simulated multidisciplinary planning environments rather than routine clinical workflows, and the evaluation focused primarily on usability and user perceptions rather than objective measures of clinical performance. Consequently, future studies should examine the impact of AR on surgical planning accuracy, multidisciplinary decision-making, workflow efficiency, and patient outcomes in real-world clinical settings.

Future research should evaluate the platform in larger and more diverse clinical populations while continuing to refine interaction stability, workflow responsiveness, adaptive visualization, and integration with hospital information systems to support routine clinical adoption. Although the co-design process involved health care professionals and patients, future work should also include additional stakeholders, such as family caregivers, specialist nurses, allied health professionals, and health care administrators, to better capture the wider multidisciplinary cancer care ecosystem. Finally, future research should investigate emotionally adaptive visualization strategies that tailor information according to patient readiness, preferences, and clinical context while addressing broader implementation challenges, including privacy, data governance, and secure integration within health care systems.

Conclusions

This study demonstrates that the successful implementation of AR in surgical oncology depends not only on accurate anatomical visualization but also on effective integration within clinical workflows, multidisciplinary collaboration, and patient-centered communication. Through participatory co-design involving health care professionals and patients with cancer, we developed and evaluated a human-centered AR platform while deriving design principles to guide the future development and implementation of clinically relevant AR systems. These findings suggest that immersive AR technologies can function as collaborative cognitive tools that support shared understanding, improve communication, and facilitate multidisciplinary decision-making. Although further real-world clinical validation is required, this work provides a strong foundation for the future development and implementation of human-centered AR systems capable of supporting more collaborative, efficient, and patient-centered cancer care.

Acknowledgments

During the preparation of this manuscript, the authors used OpenAI's ChatGPT (GPT-5.5) to support language refinement and improve the clarity of the writing. The authors carefully reviewed and edited all AI-generated content.

Funding

This project has received funding from the European Union’s Horizon Research and Innovation Program and from the European Health and Digital Executive Agency under agreement number 101137233.

XSL•FO
RenderX

Data Availability

The datasets generated and/or analyzed during this study are not publicly available due to the inclusion of clinical imaging data and the potential risk of participant identification. Deidentified data may be made available from the corresponding author upon reasonable request and subject to institutional and ethical approval requirements.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Clinical terminology guide, semistructured interview guides, and cocreative workshops materials.
[DOCX File, 201 KB - humanfactors_v13i1e86725_app1.docx]

References

1. Barsom EZ, Graafland M, Schijven MP. Systematic review on the effectiveness of augmented reality applications in medical training. Surg Endosc 2016 Oct;30(10):4174-4183. [doi: 10.1007/s00464-016-4800-6] [Medline: 26905573]
2. Vaughan-Nichols SJ. Augmented reality: no longer a novelty? Computer (Long Beach Calif) 2010;42(12):19-22. [doi: 10.1109/MC.2009.380]
3. Jud L, Fotouhi J, Andronic O, et al. Applicability of augmented reality in orthopedic surgery - a systematic review. BMC Musculoskelet Disord 2020 Feb 15;21(1):103. [doi: 10.1186/s12891-020-3110-2] [Medline: 32061248]
4. Ren XM, Xie T, Jin X, Liu L, Liu L, Zhang M. Precision surgery in the era of 3D visualization, AR/VR, and 3D printing: current applications and future directions. Front Med 2026;13:1688748. [doi: 10.3389/fmed.2026.1688748]
5. Dwyer T, Marriott K, Isenberg T, et al. Immersive analytics: an introduction. In: In Immersive Analytics: Springer International Publishing; 2018:1-23. [doi: 10.1007/978-3-030-01388-2_1]
6. Pio L, Kassir R, La Barbera G, et al. 3D imaging contribution in pediatric surgical oncology: a multi-stakeholder assessment study. Sci Rep 2026 Mar 19;16(1):14264. [doi: 10.1038/s41598-026-44543-z] [Medline: 41857121]
7. Tang R, Ma LF, Rong ZX, et al. Augmented reality technology for preoperative planning and intraoperative navigation during hepatobiliary surgery: a review of current methods. Hepatobiliary Pancreat Dis Int 2018 Apr;17(2):101-112. [doi: 10.1016/j.hbpd.2018.02.002] [Medline: 29567047]
8. Hegarty M, Keehner M, Cohen C, Montello DR, Lippa Y. The role of spatial cognition in medicine: applications for selecting and training professionals. In: In Applied Spatial Cognition: Psychology Press; 2007:285-316. [doi: 10.4324/9781003064350-11]
9. Qadrie Z, Maqbool M, Dar MA, Qadir A. Navigating challenges and maximizing potential: handling complications and constraints in minimally invasive surgery. Open Health 2025 Oct 6;6(1). [doi: 10.1515/ohe-2025-0059]
10. Orosco RK, Tapia VJ, Califano JA, et al. Positive surgical margins in the 10 most common solid cancers. Sci Rep 2018 Apr 9;8(1):5686. [doi: 10.1038/s41598-018-23403-5] [Medline: 29632347]
11. Coutu B, Ryan E, Christensen D, et al. Positive margins matter regardless of subsequent resection findings. Oral Oncol 2022 May;128:105850. [doi: 10.1016/j.oraloncology.2022.105850] [Medline: 35436713]
12. Gilg S, Sandström P, Rizell M, et al. The impact of post-hepatectomy liver failure on mortality: a population-based study. Scand J Gastroenterol 2018;53(10-11):1335-1339. [doi: 10.1080/00365521.2018.1501604] [Medline: 30345846]
13. Giannone F, Felli E, Cherkaoui Z, Mascagni P, Pessaux P. Augmented reality and image-guided robotic liver surgery. Cancers (Basel) 2021 Dec 14;13(24):6268. [doi: 10.3390/cancers13246268] [Medline: 34944887]
14. Chan HHL, Sahovaler A, Daly MJ, et al. Projected cutting guides using an augmented reality system to improve surgical margins in maxillectomies: a preclinical study. Oral Oncol 2022 Apr;127:105775. [doi: 10.1016/j.oraloncology.2022.105775] [Medline: 35203002]
15. Kowal MR, Ibrahim M, Mihaljević AL, Kron P, Lodge P. Technological advances in pre-operative planning. J Clin Med 2025 Jul 30;14(15):5385. [doi: 10.3390/jcm14155385] [Medline: 40807006]
16. Malhotra S, Halabi O, Dakua SP, Padhan J, Paul S, Palliyali W. Augmented reality in surgical navigation: a review of evaluation and validation metrics. Appl Sci 2023;13(3):1629. [doi: 10.3390/app13031629]
17. Ntourakis D, Memeo R, Soler L, Marescaux J, Mutter D, Pessaux P. Augmented reality guidance for the resection of missing colorectal liver metastases: an initial experience. World J Surg 2016 Feb;40(2):419-426. [doi: 10.1007/s00268-015-3229-8] [Medline: 26316112]
18. Hallet J, Soler L, Diana M, et al. Trans-thoracic minimally invasive liver resection guided by augmented reality. J Am Coll Surg 2015 May;220(5):e55-e60. [doi: 10.1016/j.jamcollsurg.2014.12.053] [Medline: 25840539]
19. Douglas DB, Wilke CA, Gibson JD, Boone JM, Wintermark M. Augmented reality: advances in diagnostic imaging. Multimodal Technol Interact 2017;1(4):29. [doi: 10.3390/mti1040029]
20. Carayon P, Wetterneck TB, Rivera-Rodriguez AJ, et al. Human factors systems approach to healthcare quality and patient safety. Appl Ergon 2014 Jan;45(1):14-25. [doi: 10.1016/j.apergo.2013.04.023] [Medline: 23845724]

XSL•FO
RenderX

21. Kaihlanen AM, Gluschkoff K, Hyppönen H, et al. The associations of electronic health record usability and user age with stress and cognitive failures among Finnish registered nurses: cross-sectional study. JMIR Med Inform 2020 Nov 18;8(11):e23623. [doi: 10.2196/23623] [Medline: 33206050]
22. Bødker S. When second wave HCI meets third wave challenges. Presented at: NordiCHI '06: 4th Nordic conference on Human-computer interaction: changing roles; Oct 14-18, 2006; Oslo, Norway p. 1-8 URL: https://dl.acm.org/doi/proceedings/10.1145/1182475 [accessed 2026-09-16] [doi: 10.1145/1182475.1182476]
23. Duffy A, Boroumandzad N, Sherman AL, Christie G, Riadi I, Moreno S. Examining challenges to co-design digital health interventions with end users: systematic review. J Med Internet Res 2025 Mar 14;27:e50178. [doi: 10.2196/50178] [Medline: 40085834]
24. Matsangidou M, P. Kassianos A, Papaioannou D, et al. Virtual painkillers: designing accessible virtual reality experiences for helping cancer patients manage pain at home. Presented at: CHI '22: Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems; Apr 29 to May 5, 2022; New Orleans, LA, USA p. 1-9 URL: https://dl.acm.org/doi/proceedings/10.1145/3491101 [accessed 2026-09-16] [doi: 10.1145/3491101.3503562]
25. Jacob C, Bourke S, Heuss S. From testers to cocreators-the value of and approaches to successful patient engagement in the development of eHealth solutions: qualitative expert interview study. JMIR Hum Factors 2022 Oct 6;9(4):e41481. [doi: 10.2196/41481] [Medline: 36102548]
26. Endsley TC, Sprehn KA, Brill RM, Ryan KJ, Vincent EC, Martin JM. Augmented Reality Design Heuristics: Designing for Dynamic Interaction: Sage Publications; 2017. [doi: 10.1177/1541931213602007]
27. Harders M, Bianchi G, Knoerlein B. Multimodal augmented reality in medicine. In: Universal Access in Human-Computer Interaction: Springer; 2007:652-658. [doi: 10.1007/978-3-540-73281-5_70]
28. Norman D, Donald A. The Design of Everyday Things: Basic Books; 2002.
29. Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Qual Saf Health Care 2010 Oct;19 Suppl 3(Suppl 3):i68-i74. [doi: 10.1136/qshc.2010.042085] [Medline: 20959322]
30. Carayon P, Wooldridge A, Hoonakker P, Hundt AS, Kelly MM. SEIPS 3.0: Human-centered design of the patient journey for patient safety. Appl Ergon 2020 Apr;84:103033. [doi: 10.1016/j.apergo.2019.103033] [Medline: 31987516]
31. Altman M, Huang TTK, Breland JY. Design thinking in health care. Prev Chronic Dis 2018 Sep 27;15:E117. [doi: 10.5888/pcd15.180128] [Medline: 30264690]
32. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
33. Flanagan JC. The critical incident technique. Psychol Bull 1954 Jul;51(4):327-358. [doi: 10.1037/h0061470] [Medline: 13177800]
34. Brooke J. System Usability Scale (SUS): A Quick-and-Dirty Method of System Evaluation User Information: Digital Equipment Co Ltd; 1986, Vol. 43:1-7 URL: https://hell.meiert.org/core/pdf/sus.pdf [accessed 2026-09-24]
35. Kortum PT, Bangor A. Usability ratings for everyday products measured with the System Usability Scale. Int J Hum Comput Interact 2013 Jan;29(2):67-76. [doi: 10.1080/10447318.2012.681221]
36. Boudewyns V, Southwell BG, DeFrank JT, et al. Patients' understanding of oncology clinical endpoints: a literature review. Patient Educ Couns 2020 Sep;103(9):1724-1735. [doi: 10.1016/j.pec.2020.03.018] [Medline: 32273145]
37. Marta GN, Del Nero LG, Marta GN, et al. Treatment priorities in oncology: do we want to live longer or better? Clinics (Sao Paulo) 2014 Aug;69(8):509-514. [doi: 10.6061/clinics/2014(08)02] [Medline: 25141108]
38. Barnett MM. Does it hurt to know the worst?--psychological morbidity, information preferences and understanding of prognosis in patients with advanced cancer. Psychooncology 2006 Jan;15(1):44-55. [doi: 10.1002/pon.921] [Medline: 15750997]
39. Derry HM, Reid MC, Prigerson HG. Advanced cancer patients' understanding of prognostic information: applying insights from psychological research. Cancer Med 2019 Aug;8(9):4081-4088. [doi: 10.1002/cam4.2331] [Medline: 31199597]
40. Fereidouni Z, Dehghan Abnavi S, Ghanbari Z, et al. The impact of cancer on mental health and the importance of supportive services. Galen Med J 2024;13:e3327. [doi: 10.31661/gmj.v13i.3327] [Medline: 39224547]
41. Zacharia M, Karekla M. The role of psychologists and psychological approaches in cancer care. In: Kassianos AP, editor. In Handbook of Quality of Life in Cancer: Springer International Publishing; 2022:311-337. [doi: 10.1007/978-3-030-84702-9_19]
42. Byrne EA. Affective scaffolding and chronic illness. Philos Psychol 2024 May 18;37(4):921-946. [doi: 10.1080/09515089.2022.2147422]
43. Costanzo ES, Lutgendorf SK, Mattes ML, et al. Adjusting to life after treatment: distress and quality of life following treatment for breast cancer. Br J Cancer 2007 Dec 17;97(12):1625-1631. [doi: 10.1038/sj.bjc.6604091] [Medline: 18000503]
44. Baker P, Beesley H, Fletcher I, Ablett J, Holcombe C, Salmon P. "Getting back to normal" or "a new type of normal"? A qualitative study of patients' responses to the existential threat of cancer. Eur J Cancer Care (Engl) 2016 Jan;25(1):180-189. [doi: 10.1111/ecc.12274] [Medline: 25523386]
45. Andrykowski MA, Burris JL. Use of formal and informal mental health resources by cancer survivors: differences between rural and nonrural survivors and a preliminary test of the theory of planned behavior. Psychooncology 2010 Nov;19(11):1148-1155. [doi: 10.1002/pon.1669] [Medline: 20017114]

XSL•FO
RenderX

46. Fuller J, Edwards J, Procter N, Moss J. How definition of mental health problems can influence help seeking in rural and remote communities. Aust J Rural Health 2000 Jun;8(3):148-153. [doi: 10.1046/j.1440-1584.2000.00303.x] [Medline: 11249402]
47. Jennings KS, Cheung JH, Britt TW, et al. How are perceived stigma, self-stigma, and self-reliance related to treatment-seeking? A three-path model. Psychiatr Rehabil J 2015 Jun;38(2):109-116. [doi: 10.1037/prj0000138] [Medline: 25844914]
48. Blinderman CD, Cherny NI. Existential issues do not necessarily result in existential suffering: lessons from cancer patients in Israel. Palliat Med 2005 Jul;19(5):371-380. [doi: 10.1191/0269216305pm1038oa] [Medline: 16111060]
49. Wexler ID, Corn BW. An existential approach to oncology: meeting the needs of our patients. Curr Opin Support Palliat Care 2012 Jun;6(2):275-279. [doi: 10.1097/SPC.0b013e3283530e25] [Medline: 22488165]
50. Hazlehurst B, Gorman PN, McMullen CK. Distributed cognition: an alternative model of cognition for medical informatics. Int J Med Inform 2008 Apr;77(4):226-234. [doi: 10.1016/j.ijmedinf.2007.04.008] [Medline: 17556014]
51. Arthur T, Phillips A, Franks Z, Pearce S. Redesigning informed consent: enhancing risk communication through a novel risk visualization tool. J Curr Surg 2015;5(2-3):160-164. [doi: 10.14740/jcs275w]
52. Sauro J. Measuring usability with the System Usability Scale (SUS). MeasuringU. 2025. URL: https://measuringu.com/sus/ [accessed 2026-09-16]

Abbreviations

AR: augmented reality
CRLM: colorectal liver metastasis
CT: computed tomography
HCI: human-computer interaction
HPB: hepato-pancreato-biliary
MRI: magnetic resonance imaging
NHS: National Health Service
RTA: reflexive thematic analysis
SUS: System Usability Scale

Edited by S Law; submitted 29.Oct.2025; peer-reviewed by CW Lau, T Nilsson; revised version received 27.Jul.2026; accepted 27.Jul.2026; published 30.Sep.2026.

Please cite as:
Matsangidou M, Kowal MR, Yianni J, Krini M, Pathak S, Gordon K, Luker MA, Stocken DD, Jayne D, Pattichis CS
From Prototype to Practice: Mixed Methods Study of Human-Centered Design and Usability Evaluation of Augmented Reality for Complex Oncology Surgical Workflows
JMIR Hum Factors 2026;13:e86725
URL: https://humanfactors.jmir.org/2026/1/e86725
doi:10.2196/86725

© Maria Matsangidou, Mikolaj R Kowal, Jakob Yianni, Maria Krini, Samir Pathak, Katie Gordon, Malcolm A Luker, Deborah D Stocken, David Jayne, Constantinos S Pattichis. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 30.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Privacy Fact Sheets for Mitigating Disease-Related Privacy Concerns and Facilitating Equal Access to the Electronic Health Record: Randomized Controlled Trial

Niklas von Kalckreuth, MSc; Markus A Feufel, MSc, PhD

Division of Ergonomics, Department of Psychology and Ergonomics, Technische Universität Berlin, Straße des 17. Juni 135, Berlin, Germany

Corresponding Author:
Niklas von Kalckreuth, MSc
Division of Ergonomics, Department of Psychology and Ergonomics, Technische Universität Berlin, Straße des 17. Juni 135, Berlin, Germany

Abstract

Background: The German electronic health record (EHR) aims to enhance patient care and reduce costs, but users often worry about data privacy and security. To mitigate disease-related privacy concerns, for instance, surrounding stigmatized diseases, we test the effect of privacy fact sheets (PFSs)—a concise but comprehensive transparency feature designed to increase users' perceived control over their data—on increasing EHR use in a simulated online study.

Objective: The study aimed to investigate whether displaying a PFS shortly before upload decisions must be made mitigates disease-related privacy concerns and makes uploads more likely.

Methods: In an online survey study, 393 German participants from the recruitment platform Prolific were asked to interact with a randomly assigned medical report that varied systematically in terms of disease-related stigma (high vs low) and time course (TC; acute vs chronic). They were then asked to decide whether to upload the report to an EHR click dummy, while we systematically varied the presentation of privacy information (PFS vs no PFS). Participants were randomly (single-blinded) assigned to one of the 2×2×2 conditions (stigma, TC, privacy information).

Results: All 393 participants were randomly assigned to one of the following groups: low, acute, no PFS (n=52, 13.2%); low, chronic, no PFS (n=45, 11.5%); high, acute, no PFS (n=46, 11.7%); high, chronic, no PFS (n=55, 14%); low, acute, PFS (n=44, 11.2%); low, chronic, PFS (n=41, 10.4%); high, acute, PFS (n=56, 14.2%); and high, chronic, PFS (n=54, 13.7%). The results show that, in general, upload behavior is negatively influenced by disease-related stigma (odds ratio [OR] 0.130; $P<.001$) and positively influenced when a PFS is given (OR 4.527; $P<.001$). This increase was particularly pronounced for stigmatized diseases (OR 5.952; $P=.006$), but independent of the TC of the diseases.

Conclusions: Our results demonstrate that PFSs may help to increase EHR uploads in people interacting with a realistic EHR click dummy, by mitigating privacy concerns in scenarios involving stigmatized diseases. Results further indicate that a PFS is mainly relevant and effective for people with increased privacy risk perceptions, whereas they neither benefit nor hurt others. Thus, implementing PFSs may increase the likelihood that users who perceive high privacy risks when confronted with sensitive or stigmatized health information decide to upload their data to the EHR, ultimately increasing digital health equity.

Trial Registration: Deutsches Register Klinischer Studien DRKS00033652; https://drks.de/search/de/trial/DRKS00033652

(JMIR Hum Factors 2026;13:e71124) doi:10.2196/71124

KEYWORDS

adoption; disease characteristics; EHR; electronic health record; mHealth; mobile health; stigma; time course; transparency; transparency feature; privacy concern

Introduction

Background

The electronic health record (EHR) is one key element in the digital transformation of health care systems because it allows patients' health data (eg, diagnoses, therapies, vaccinations, and medication plans) to be readily documented, exchanged, and viewed by various stakeholders [1-6]. The resulting network of care providers can make patient treatment more effective, safer, and faster across institutions [7]. For instance, preexisting conditions, intolerances, and medication plans can be taken into account during diagnosis and treatment to prevent adverse medication interactions, duplicate diagnoses, overtreatment, and undertreatment [5]. Also, it is hoped that physicians will spend less time on obtaining patients' medical history thanks to the EHR, which they could devote to actual patient treatment [8]. In Germany, approximately 90% of residents are covered

XSL•FO
RenderX

by statutory health insurance and thus received an EHR account by default in January 2025, unless they opted out [9-11]. From October 1, 2025, health care providers are required to use the EHR infrastructure for documentation and data exchange [11]. A prerequisite for the potential of EHRs to realize is user engagement. Specifically, the Patient Data Protection Act mandates that patients maintain sole control of their data, allowing them to decide which information is stored in the EHR, who has access to it, and which data are to be deleted [5]. Consequently, the success of the EHR hinges on patients' actual use of the technology. But national and international studies indicate that many patients remain skeptical toward the EHR, mainly due to concerns arising from limited trust in both data security (ie, the technical and organizational measures that protect personal data) and data privacy (ie, the rights and rules governing how personal data is collected, used, and shared) [8,12-15]. In Germany, these concerns were intensified due to reported security issues, which were addressed and resolved during the rollout [16,17]. However, concerns about EHR use are not static in nature. Our previous studies suggest that participants adapt their attitudes toward EHR dynamically. For instance, we could show that the perceived risks and benefits of using EHR are related to disease-specific privacy concerns, such as the stigma and the time course (TC) of diseases, that is, the more permanent and risky nature of data associated with chronic conditions [18-20]. Previous studies have also shown that existing communication strategies, for instance, by health insurers or the Federal Ministry of Health, have not been sufficient to effectively convey the core functions and data protection mechanisms of the German EHR [13,14,21]. Against this background, there is a need for concise and comprehensible communication strategies that can transparently explain data privacy and data security measures, thereby strengthening user trust in the EHR. In this study, we investigate the effect of a privacy fact sheet (PFS)—a concise but comprehensive transparency feature—on increasing EHR use and, specifically, to what extent the PFS can mitigate disease-related privacy concerns and increase the upload of medical reports to the EHR.

Prior Work

"Notice and choice" is the most widely used framework for ensuring data privacy worldwide [22,23]. As its name suggests, it consists of 2 components: privacy notices and privacy choices. Whereas *privacy notices* explain how personal data are collected, processed, and shared with third parties, *privacy choices* give users control over various aspects of these practices, including the decision to start and terminate them [22]. Various studies indicate that, if informed by privacy notices, users are empowered to choose IT systems that match their preferences, typically those with high data security and privacy standards, and avoid less secure ones [24,25]. But the current formats used for privacy notices, most commonly *privacy policies*, tend to provide rather detailed information and often use legal jargon [26-30], which aims to maximize legal protection of IT providers rather than to transparently inform users [31]. Research has shown that overly lengthy and complex privacy policies may ultimately serve as a "red flag," leading users to lose trust in the provider, if not to discontinue technology use altogether [32,33]. Consequently, concise, easy-to-understand privacy notices are a prerequisite for users to adopt digital health technologies such as the EHR [33-35].

In contrast to full-blown privacy policies, the shorter *transparency features* have been shown to be an effective type of privacy notice, because they provide a brief and easy-to-understand overview of data privacy and data security measures and are meant to inform rather than to provide legal assurance [36,37]. Recent work also shows that such transparency features can decrease privacy concerns and increase EHR acceptance by strengthening users' perceived control over personal data—a construct that describes the extent to which users feel able to understand, oversee, and control how their personal data are handled and that plays a central role in privacy calculus models [7,38-40]. Stronger perceptions of control are known to reduce privacy concerns and foster trust in digital health and online technologies [7,38,39,41].

Empirical studies across domains suggest that transparency features may influence user behavior. For instance, a study in the eCommerce domain demonstrated that displaying a transparency feature positively influences purchase numbers [37]. But increased use does not (only) depend on the format of the privacy notice; it is also influenced by the contents it provides, including the efficacy of the mentioned data protection measures and privacy choices, and its timing, that is, when the privacy notice is given to users [42]. At the same time, studies in other digital contexts have reported mixed results regarding the behavioral impact of privacy notices and transparency features, suggesting that their effectiveness may depend on contextual factors such as perceived privacy risk or data sensitivity [36,43]. Our previous studies have shown that a transparency feature with a concise but comprehensive summary of all relevant contents—which we refer to as a PFS—positively influences EHR use when given shortly before the upload process [40,44]. In addition, we could show that a patient-centered framing of these contents that specifies what users can do to control the EHR and emphasizes their perceived control over personal data (eg, you can control all of your data) has the biggest effect on EHR adoption [44].

As stated at the outset, in another line of studies, we have shown that privacy concerns, intention to use the EHR, and upload behavior are influenced by the characteristics of diseases, in particular by disease-related stigma and TC [18-20]. Disease-specific stigma has been shown to have an inhibiting influence on upload behavior and to increase the risks people perceive when they are asked to upload information related to these diseases to the EHR [18,19]. Conversely, the TC of diseases (ie, whether diseases are chronic rather than acute) tends to increase both privacy concerns and intention to use the EHR. That is, patients with chronic conditions recognize a greater value in using the EHR but have heightened privacy concerns when it comes to uploading chronic conditions to the EHR [20]. In this study, we aim to merge these two lines of studies to validate and extend the positive effect of a patient-framed PFS on users' decision to upload diseases to the EHR click dummy when disease-specific privacy concerns are systematically varied.

XSL•FO
RenderX

Aim of This Research and Approach

In this study, we test whether displaying a patient-framed PFS, developed in our previous studies [40,44], shortly before the decision to upload a medical report must be made increases the likelihood that users upload medical reports to the EHR for diseases that vary along 2 dimensions: TC and disease-specific stigma. After describing the methods and results, we discuss the implications, reflect on the study's limitations, and conclude with a reflection on the objective of this study.

Methods

Ethical Considerations

This study was approved by the Ethics Committee of the Department of Psychology and Ergonomics at Technische Universität Berlin (tracking number: AWB_KAL_1_230206_Erweiterungsantrag). The study is registered as a randomized controlled trial at Deutsches Register Klinischer Studien (DRKS00033652). Participants volunteered to participate in the survey, and written informed consent was required to participate. On the first page of the survey, participants were told about the experimenter, the study purpose, what data were to be collected during the study, and where and for how long they would be stored. Also, participants had the possibility to download a PDF with the study information. Hence, participants were informed about the duration of the survey (approximately 8 min) as well as the compensation for participation. All data were collected and stored in an anonymized form. No directly identifying personal information was collected. Data were processed confidentially and used exclusively for research purposes.

Participants

The online study was conducted between April 15, 2024, and May 16, 2024. Based on an a priori power analysis for a logistic regression using G*Power (version 3.1.9.7) with disease-related stigma (high vs low), TC (acute vs chronic), and privacy information (PFS vs no PFS) as binomial distributed predictors, a false positive rate α of .05, a power β of .80, an estimated odds ratio (OR) for the predictor with the smallest expected effect size (TC) of 1.7 (derived from the prestudy with n=80 participants), and a probability of the outcome (upload decision) under the null hypothesis of 0.5, reflecting a conservative assumption, we aimed for a sample size of 363 participants. To ensure this target was met, we oversampled participants by 30%, resulting in a total sample of 471 individuals. Oversampling accounted for potential exclusions due to failed attention checks, study dropouts, self-reported invalid data (approximately 20%), as indicated in preliminary studies [40,44], and prior medical histories with the diseases used in the study (approximately 10%, based on prior findings) [18]. Individuals 18 years and older residing in Germany were allowed to participate in the study, as the content and questions of the study were designed to fit the context of the German EHR. Another prerequisite was that participants had no personal previous experience (own illness) with the diseases mentioned in the medical reports we used for this study, as the handling of stigmatized diseases by affected persons is different from that of unaffected persons [45]. Sampling was conducted through Prolific, a crowdsourcing platform used to recruit participants for online surveys and experiments, known for its diverse participant pool and high data quality [46]. Participation was compensated with 1.78€ (US $2.09) for 8 minutes, which corresponds to the German minimum wage. The mean value of the processing time was 8:47 minutes (SD 3:57 min), and the median was 8 minutes. A total of 471 individuals participated in the study.

Design

The experimental design replicates the disease-related manipulations from our previous study [18] and extends them by testing whether transparent privacy information can reduce disease-related privacy concerns and influence upload behavior. This approach also extends prior privacy calculus–based work by focusing on actual behavior rather than intentions, as behavioral measures provide stronger ecological validity and better capture situational influences on privacy decisions [47,48]. We used a 2×2×2 between-subject study design with the 3 independent variables (IVs): stigmatization potential (SP), TC, and privacy information. Each participant was assigned to one unique combination of these conditions. As in preliminary studies, SP (high vs low) and TC (acute vs chronic) were manipulated by displaying the diagnoses of a disease with the respective characteristics [18,19]. Additionally, privacy information (PFS vs no PFS) was manipulated by either displaying a PFS during the upload process or not. In preliminary studies, participants associate disease-related stigma with high risks [18,19], and consequences could arise in areas related to personal lifestyle, occupation, and social life if medical findings became known [19,45,49,50]. Furthermore, previous studies show that participants perceive the upload of diseases with a chronic TC as more beneficial than the upload of acute diseases [8,19,20]. Participants were randomly (single-blinded) assigned to one of the conditions in parallel (simple randomization, ratio: 1:1:1:1:1:1:1:1) using LimeSurvey's built-in "rand" function. The randomization process was fully automated within LimeSurvey, ensuring allocation concealment throughout data collection. Neither the participants nor the researchers conducting the data collection were aware of the assigned condition at any point prior to or during data collection. The dependent variable was the decision to upload the medical report, that is, whether participants were willing to upload the medical findings to the EHR [18,19,40]. Reporting of this randomized controlled trial followed the CONSORT-eHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines; a completed CONSORT-eHEALTH checklist is provided in Checklist 1.

Materials

Following a common practice in technology acceptance studies [51,52], we used a case vignette to represent a typical situation in which an EHR app may be used. In particular, the case vignette depicted a situation where the participant has recently started using an EHR app and is now faced with the decision to upload a medical finding to their EHR (see Multimedia Appendix 1). Additionally, the disease/injury was described in lay terms with 1 to 3 sentences (see Multimedia Appendix 2). The stimuli used in the study were realistic but specially created

XSL•FO
RenderX

for the purpose of the study. The medical reports were provided by hospitals and a medical association. To make the reports appear as realistic as possible, they were edited on the official document heads of these institutions. This was done with the permission of the institutions concerned. In selecting the diseases, both the related stigma and their TC were systematically varied. Disease-related stigma covered different risks for professional and social life, such as tests for sexually transmitted diseases (ie, gonorrhea and HIV) [53-56] and fractures or rheumatoid arthritis as diseases with low stigma. To reflect different TCs, diseases were divided according to an acute TC (eg, wrist fracture and gonorrhea) and a chronic one (eg, rheumatoid arthritis and HIV). Furthermore, diseases were selected to occur regardless of age, meaning they can affect individuals across different age groups, so that they would be perceived as realistic diseases by an age-diverse sample. Table 1 shows the diseases used as stimuli, categorized by level of perceived stigma and TC.

Table . Diseases used as stimuli, categorized by SP[a] and TC[b].

SP and TC	Acute	Chronic
Low	Fractured wrist	Rheumatoid arthritis
High	Gonorrhea	HIV infection

[a]SP: stigmatization potential.

[b]TC: time course.

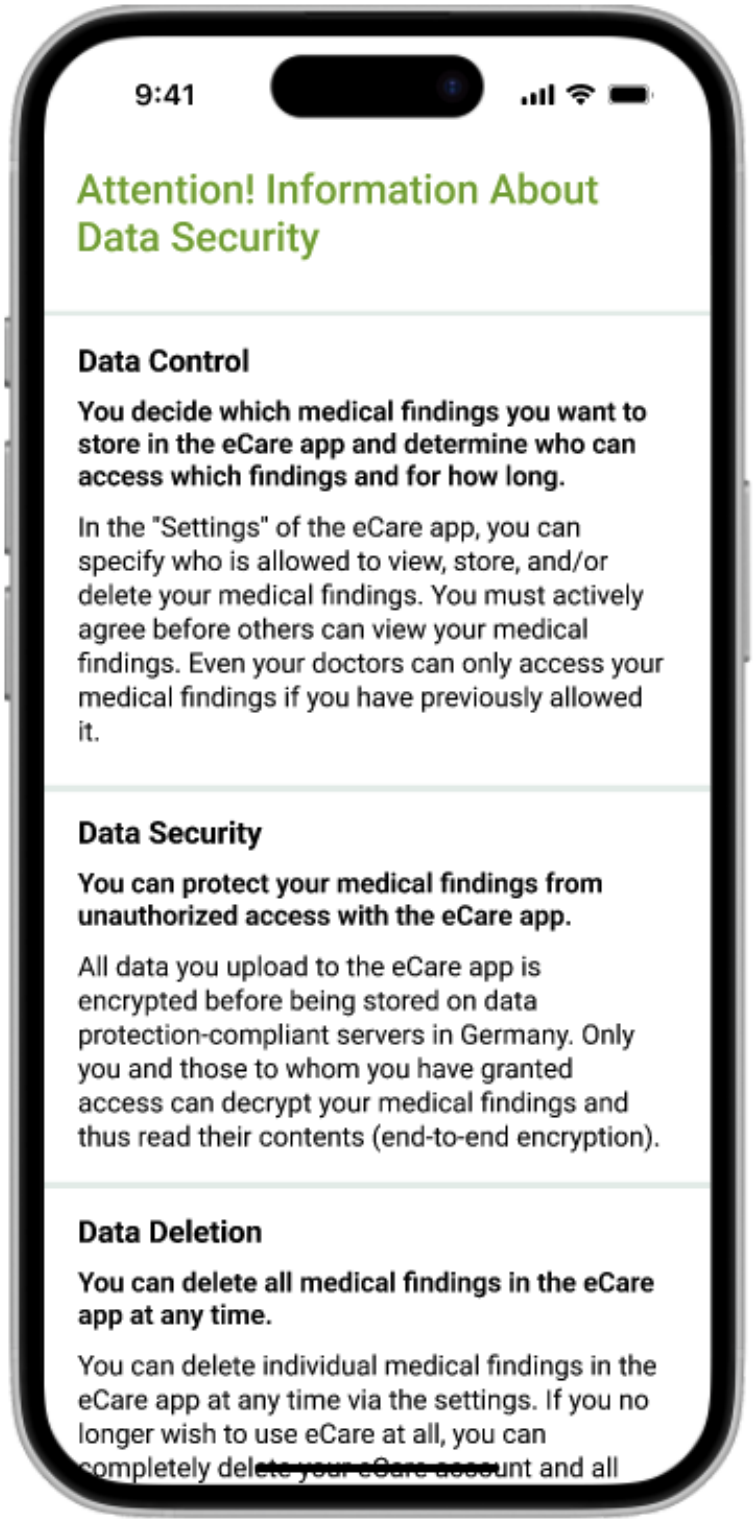
As in a previous study [19], an interactive prototype (a so-called click dummy) was used, which we created after the mobile EHR app of a German health insurance company (the BARMER) using software for interface design (FIGMA). This prototype allows for a realistic interaction with an EHR. Specifically, the prototype gave participants the ability to upload findings, grant or revoke permissions to view findings, and create medication plans. Only the "Upload findings" function was used in this study.

We used the most effective PFS that we identified based on preliminary studies [40,44], which was marked by a concise but comprehensive content and a patient-centered framing, that is, a description of what the EHR allows its users to do to control their data (eg, you can control all of your data) rather than what it does for them (eg, the EHR keeps all of your data safe). Figure 1 shows the English version of the PFS used. The English translation of the full text can be found in the Multimedia Appendix 3.

We used LimeSurvey (version 3.28.66+230719) to create and conduct a 9-page online survey. The EHR prototype was embedded into the survey using iFrame. LimeSurvey software was used to ensure that all questions had to be answered to complete the study and receive the compensation. As in previous studies, we tested the effect of the IVs by querying the perceived risk and perceived benefit of uploading findings to the EHR using validated items [18-20]. Also, we assumed that people perceived more risk when the SP was high and more benefit when the TC was chronic [19,20]. Perceived risk and perceived benefit were measured using a 7-point Likert scale ranging from 1 ("Strongly disagree") to 7 ("Strongly agree"). The decision to upload the finding was measured using a validated dichotomous item (yes/no) [18,19].

XSL•FO
RenderX

Figure 1. Privacy fact sheets used in the study.



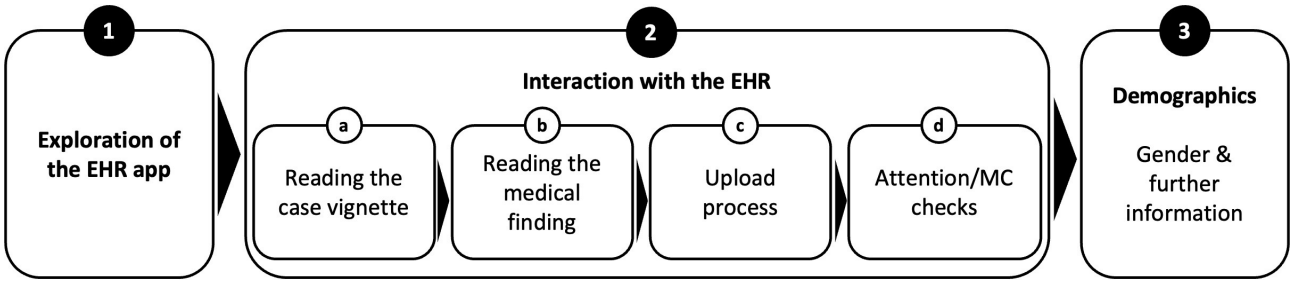
Procedure

The study procedure is shown in Figure 2. The survey consisted of 3 parts. After giving their written informed consent according to the Declaration of Helsinki of the World Medical Association, (1) participants had several minutes to interact with the EHR prototype. (2a) Participants then read a randomly selected case vignette addressing the use of the EHR in the context of uploading a medical finding. (2b) Additionally, the participants read the medical findings of the respective illness (low or high SP and acute or chronic TC, depending on the experimental group), as well as a brief description of the respective disease. (2c) Afterwards, as part of the upload process, the participants were asked to select the medical finding for upload. Depending on the experimental group, either a PFS was displayed before the disease could be selected or not. Participants then decided whether they wanted to upload the report to their EHR. (2d) After uploading, participants who were shown a PFS were asked

XSL•FO
RenderX

a question about the content of the texts to ensure that the texts were read (attention check), and all participants were asked about the perceived privacy risks and benefits of uploading the report (manipulation check). (3) The survey was completed with the collection of demographic characteristics (age, gender, education level, and experience with mobile health [mHealth] apps) as control variables, as well as the opportunity for participants to declare their responses invalid due to lack of care in processing them (see Multimedia Appendix 4 for the questionnaire).

Figure 2. Overview of the study design. EHR: electronic health record; MC: manipulation check.



Hypotheses

As mentioned above, we hypothesize that diseases with high stigma would result in a high perceived risk and a chronic TC in a high perceived benefit. Hence, we hypothesize that the upload decision is negatively influenced by high disease stigma (H1) and positively influenced by a chronic TC (H2). Based on previous studies, we also assume that a PFS will generally increase upload behavior compared to the no PFS condition (H3). More specifically, regarding the aim of our study, if a PFS can mitigate disease-related concerns, we hypothesize that showing a PFS mitigates the negative influence of disease-related stigma on the upload decision (H4). Furthermore, we hypothesize that the positive influence of a chronic TC on the upload decision will be enhanced by the presence of a PFS, as it enhances perceived benefits related to long-term health management (H5). Textbox 1 provides an overview of the hypotheses regarding the IVs.

Textbox 1. Overview of the hypotheses.

The following were the hypotheses regarding the independent variables:

- H1: The number of uploads of medical findings to the electronic health record (EHR) is lower for diseases that are stigmatized compared to those that are nonstigmatized.
- H2: The number of uploads of medical findings to the EHR is higher for chronic diseases compared to acute diseases.
- H3: The number of uploads of medical findings to the EHR is higher when transparency regarding data privacy and security is high compared to when it is low.
- H4: The increase in the number of uploads when showing a privacy fact sheet (PFS) is higher for stigmatized diseases than for nonstigmatized diseases.
- H5: The increase in the number of uploads when showing a PFS is higher for chronic diseases than for acute diseases.

Analyses

We cleaned and analyzed the data using RStudio (version 2023.09.1+494). The analysis regarding the manipulation checks of perceived privacy risks and benefits was performed using *t* tests, a statistical method used to compare the means of 2 groups. The influence of the IVs (disease-specific stigma, TC, displaying a PFS) and the interaction effects between stigma and the display of a PFS, as well as between TC and the display of a PFS on the upload decision, were tested using multiple logistic regression with dummy coding, a method used to model the probability of a binary outcome based on one or more predictor variables.

We also included a robustness check of the results regarding the upload decision. To control for potential influences of demographic and interindividual variables that could bias coefficients and *P* values, we used multiple logistic regression. To not bias *P* values as a result of controlling, we only included variables in the model that have been shown to have a causal relationship with the IVs (ie, causal confounders): age, education level, and experience with the technical system [38,57,58]. *P* values were adjusted for multiple testing using the Benjamini-Hochberg procedure [59].

Results

Survey Characteristics

A total of 471 observations were collected. A total of 78 (16.5%) records were excluded, of which 70 (14.9%) were excluded because of incomplete questionnaires, 4 (0.85%) because participants failed the attention check, and 4 (0.85%) because responses were marked as invalid by participants. A sample of 393 observations (156 female participants, 231 male participants, and 6 with no information) was used for further analysis. Figure 3 shows the participation and distribution process according to the guidelines of the CONSORT (Consolidated Standards of Reporting Trials) statement [60].

XSL•FO
RenderX

Figure 3. CONSORT (Consolidated Standards of Reporting Trials) flow chart. PFS: privacy fact sheet; SP: stigma potential; TC: time course.

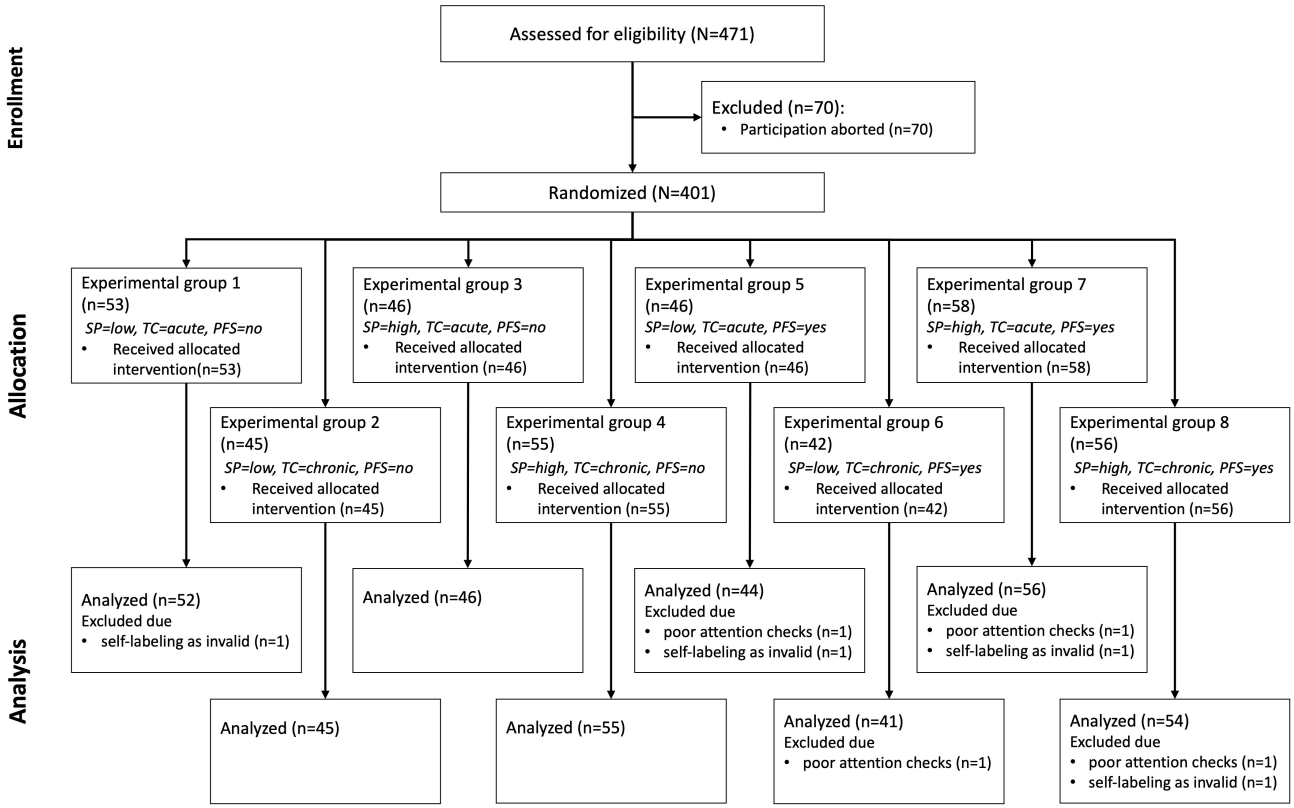


Table 2 summarizes the demographic characteristics of the entire sample. The demographic characteristics of the subsamples for each experimental group are shown in the Multimedia Appendix 5.

Table . Demographic data of the sample (n=393).

Demographic characteristic	Respondents
Age (y), mean (SD)	31.67 (9.94)
Sex, n (%)	
Female	156 (39.7)
Male	231 (58.8)
No answer	6 (1.5)
Education, n (%)	
No degree	11 (2.8)
High school/vocational education	179 (45.5)
Bachelor	102 (26.0)
Master	90 (22.9)
PhD	11 (2.8)
Experience with mHealth[a] apps, n (%)	
No use	226 (57.5)
Regular use	167 (42.5)

[a]mHealth: mobile health.

Risk and Benefit Perception

Similar to the preliminary studies [18,19], risk and benefit perception of uploading served as a manipulation check to test the validity of our manipulation (ie, the medical reports) with respect to the perception of risk (stigma) and benefit (TC). As expected, uploading medical findings of stigmatized diseases was perceived as riskier than those of nonstigmatized diseases (low: mean 3.88, SD 1.68; high: mean 5.15, SD 1.6; t_{391}=7.648; P<.001). Consequently, we assume that our risk manipulation was successful. There was no significant difference in the perceived benefit regarding the TC of the disease (acute: mean

XSL•FO
RenderX

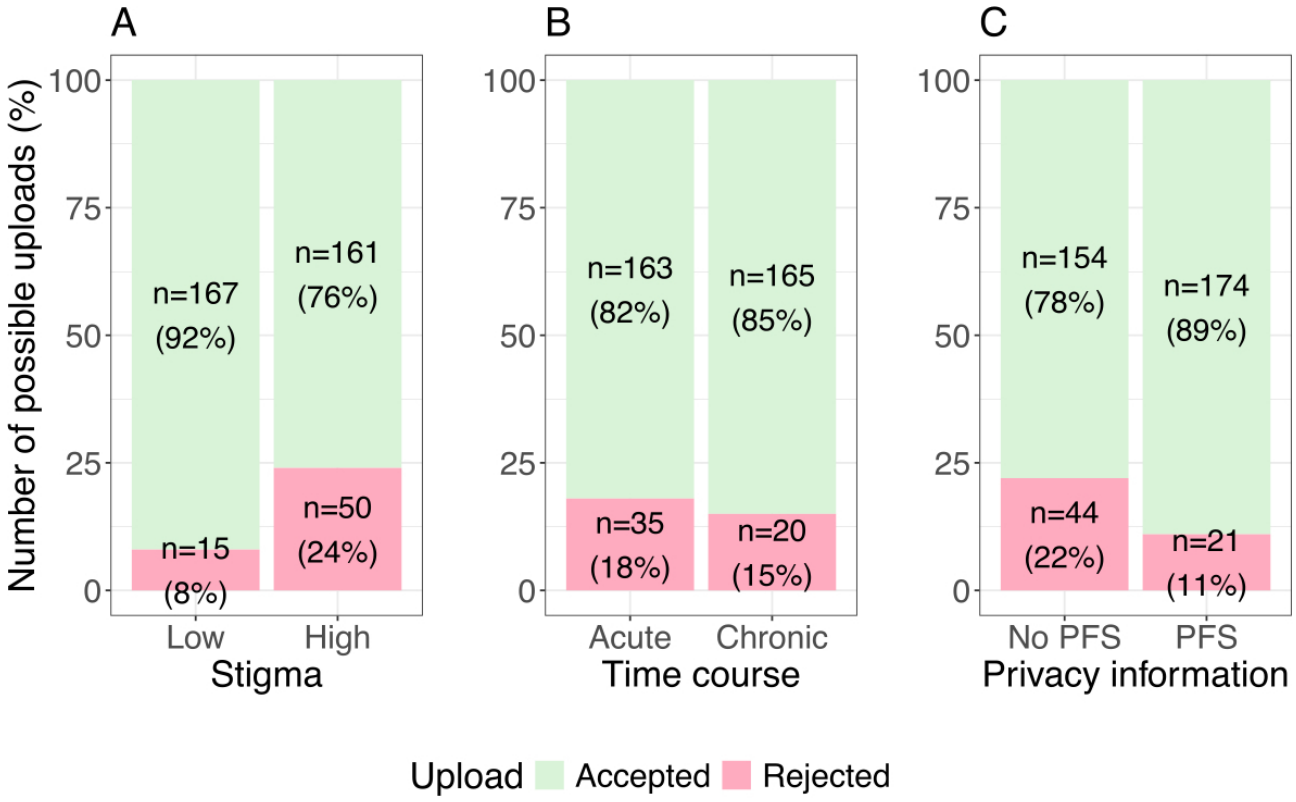
5.63, SD 1.41; chronic: mean 5.72, SD 1.16; t_{391}=0.703; *P*=.483). Consequently, we assume that our benefit manipulation was not successful.

Additionally, we analyzed the benefit perception in relation to stigma and the risk perception in relation to TC, even though these were not part of the initial manipulation checks. Uploading medical findings of nonstigmatized diseases was perceived as more beneficial than those of stigmatized diseases (low: mean 5.84, SD 1.16; high: mean 5.54, SD 1.39; t_{391}=2.345; *P*=.02). Furthermore, uploading medical findings of chronic diseases into the EHR was perceived as riskier than those of acute diseases (acute: mean 4.31, SD 1.83; chronic: mean 4.82, SD 1.64; t_{391}=2.893; *P*=.004).

Upload Behavior

Upload behavior was negatively associated with disease-related stigma (*z*=4.568; *P*<.001), thus supporting H1. Specifically, when stigma was high, it was more than seven times less likely that the report was uploaded to the EHR (76.3%, 161/211) than when stigma was low (91.8%, 167/182). TC of the disease was not associated with the decision to upload a report (z=0.877; *P*=.38). Consequently, H2 is rejected. The PFS was positively associated with the decision to upload a medical report to the EHR (*z*=3.298; *P*<.001), supporting H3. When a PFS was given, participants were more than 4 times as likely to upload the diagnosis to their EHR (89.2%, 174/195) than when a PFS was not given (77.7%, 154/198). The absolute number of uploads is shown in Figure 4 as a function of the IVs disease-related stigma, TC, and PFS vs no PFS.

Figure 4. Number of uploads to the electronic health record as a function of disease-related (A) stigma, (B) time course, and (C) privacy information. PFS: privacy fact sheet.



We also tested for interaction effects between stigma and privacy information, as well as between TC and privacy information, to explore potential moderating effects. The interaction between stigma and privacy information was significant (*z*=2.734; *P*=.006), indicating that the increase in the number of uploads when showing a PFS is higher for stigmatized diseases than for nonstigmatized diseases, thus supporting H4 (see Figure 5A). In contrast, the interaction between TC and privacy information was not significant (*z*=0.094; *P*=.92), suggesting that displaying a PFS did not differentially impact the upload decision based on whether the disease was acute or chronic (see Figure 5B). Consequently, H5 is rejected. The summary of the results of the logistic regression is shown in Table 3.

XSL•FO
RenderX

Figure 5. Number of uploads to the electronic health record as interaction between stigma and privacy information (A) and time course and privacy information (B). PFS: privacy fact sheet.

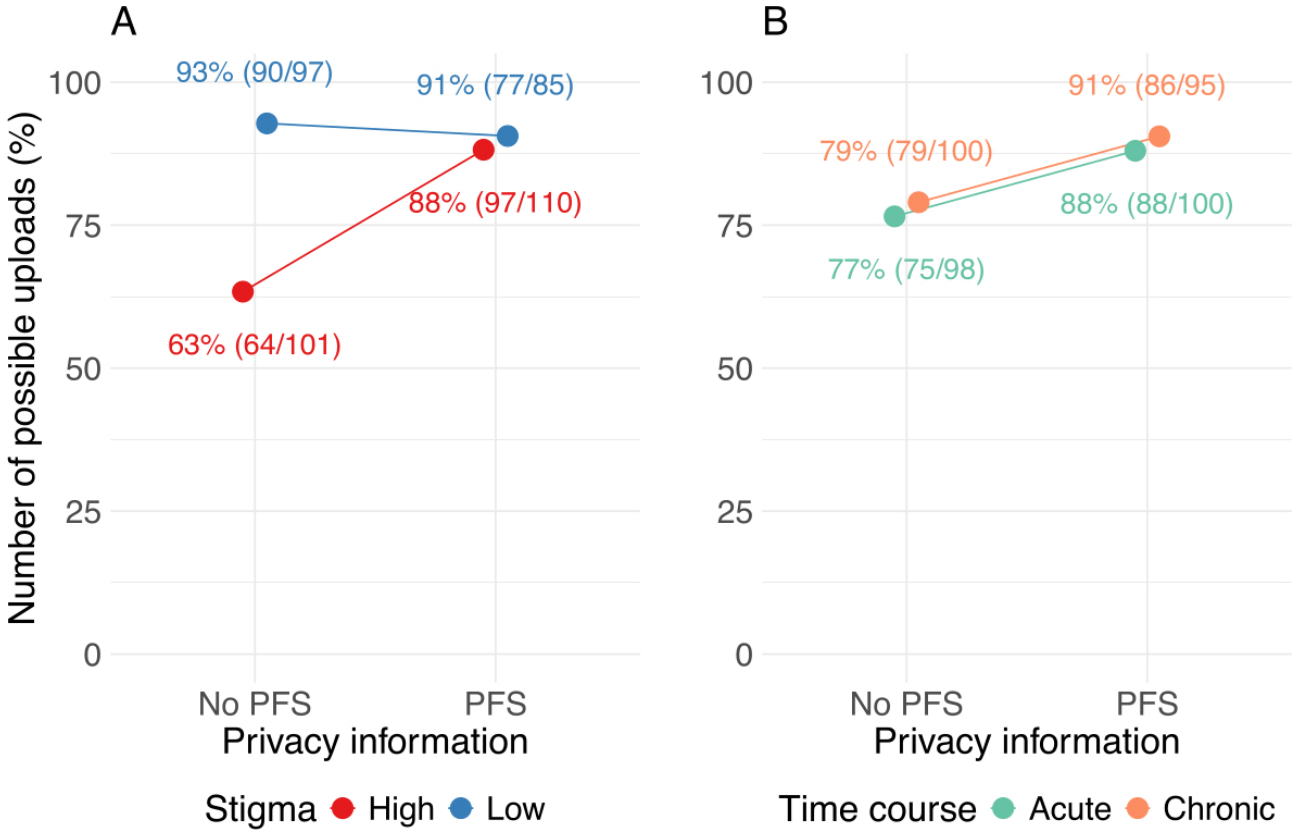


Table . Results of the logistic regression.[a]

Variable	z value	*P* value	ORs[b] and 95% CIs		
			Lower	OR	Upper
Stigma (high vs low)	4.568	<.001	0.050	0.130	0.296
TC[c] (acute vs chronic)	0.877	.38	0.672	1.382	2.869
Privacy information (PFS[d] vs no PFS)	3.298	<.001	1.888	4.527	11.490
Stigma*privacy information	2.734	.006	1.661	5.952	21.739
TC*privacy information	0.094	.92	0.325	1.058	3.390

[a]R^2=0.107 (Hosmer-Lemeshow), 0.297 (Nagelkerke), 0.284 (Cox-Snell). Model χ^2_5=37.68; *P*<.001.

[b]OR: odds ratio.

[c]TC: time course.

[d]PFS: privacy fact sheet.

Robustness Check

When controlling for interindividual variables (age, gender, education, and mHealth experience), the effects of stigma (*z*=4.820; *P*<.001) and information transparency (*z*=3.548; *P*<.001) and their interaction (*z*=3.086; *P*=.002) remained robust. Age had a negative effect on the upload behavior (*z*=2.531; OR 0.965, 95% CI 0.939 - 0.992; *P*=.01). With an increase in age, users were less likely to upload medical findings into their EHR. The other control variables did not influence the upload behavior.

Discussion

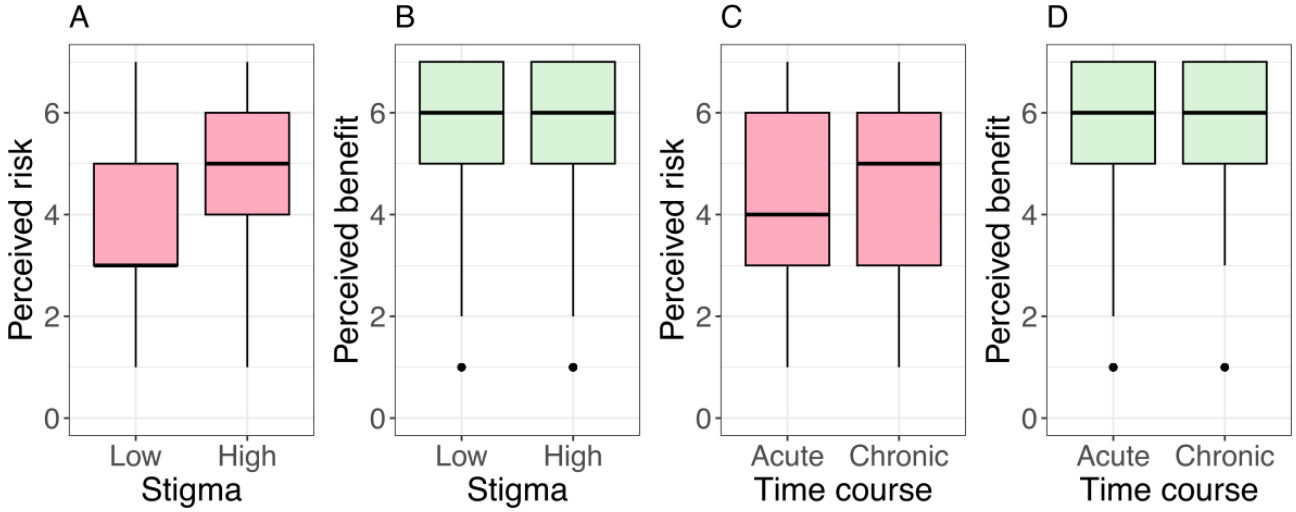
Principal Findings

The results of our study show that the decision to upload an individual medical report to an EHR click dummy is influenced by disease-related stigma as well as by privacy notices, that is, concise but comprehensive information about data privacy choices and security measures in the form of PFS. As in our preliminary studies [18,19], uploading diseases with high stigma was associated with increased privacy risk perceptions compared

XSL•FO
RenderX

to diseases with low stigma (see Figure 6A), which increased the likelihood of rejecting uploads for stigmatized diseases 6 fold compared to nonstigmatized diseases (see Figure 4A). This finding is surprising given a generally high rating of potential benefits of uploading reports to the EHR (see Figure 6B D) and an overall high willingness to upload medical findings to the EHR. In addition to the nonsignificant manipulation check regarding perceived benefits, uploads did not vary with the TC of the disease (see Figure 4).

Figure 6. Perceived risk as a function of disease-related (A) stigma and (C) time course, and perceived benefit as a function of (B) stigma and (D) time course. The horizontal line in the box represents the median.



Furthermore, a PFS positively influenced the decision to upload. When a PFS was displayed, the likelihood of uploading medical findings to the EHR was more than 3 times higher than when it was not given (see Figure 4C). This is in line with the findings of various studies showing that effective communication of data privacy choices and security information enables people to make informed decisions, thereby reducing general privacy concerns and increasing the use of EHRs [7,24,33,35,61,62]. A plausible explanation for this effect is that a PFS increases users' perceived control over personal data—allowing them to understand and oversee how their health data are handled—which in turn reduces perceived risks and facilitates the decision to upload their data [7,38,40].

Displaying a PFS did not influence the decision to upload medical findings for nonstigmatized diseases, as nearly all nonstigmatized medical reports were uploaded regardless of whether a PFS was given (see Figure 5A). In contrast, for stigmatized diseases, the PFS significantly increased the likelihood of uploads (see Figure 5B). This suggests that showing a PFS shortly before a decision to upload medical findings to the EHR must be made not only is effective in mitigating general privacy concerns but also helps to reduce specific fears related to stigmatized diseases and increase upload decisions [36,37,40].

More generally, studies in nonmedical domains, involving low-risk scenarios such as a shopping assistant [43] and an event finder [36], showed that transparency features or the transparency of privacy policies had no effect on behavior, for instance, on the decision to access the location [43] or the intention to disclose personal data to the event finder [36]. Our findings help to explain these differing findings by showing that the relevance of transparent privacy notes is mainly contingent upon the level of perceived risk associated with the data. In low-risk scenarios, such as nonstigmatized diseases, privacy concerns are typically low, which means that transparency features cannot meaningfully change the decision because there are no substantial concerns to alleviate. However, in high-risk scenarios, such as those involving stigmatized health conditions, privacy concerns are more likely to arise. In these situations, transparency features can help strengthen users' perceived control, reducing perceived risks and thereby increasing acceptance. This highlights the importance of situational context for transparency measures to matter. Transparent information about data privacy and security is not necessary in low-risk situations (although it does not hurt behavioral outcomes), but it becomes crucial for decision making in high-risk contexts, such as the handling of sensitive health data in EHRs.

Implications

The opportunities offered by implementing transparency features in the EHR should be considered by health care stakeholders. Transparency features can not only reduce general privacy concerns but can also address situational concerns triggered by disease-related stigma [18]. Thus, transparency features can ultimately help to ensure equal access to EHRs, even for users who suffer from stigmatized diseases, thereby promoting health equity [49,50,63]. This way, more patients get a chance to benefit from the EHR and, as their illnesses, allergies, and medications can be considered for future diagnostics and therapies, receive better and more targeted treatment.

Limitations and Future Directions

There are several limitations in our study, which need to be considered in subsequent studies. While our manipulation checks for perceived risk (related to stigma) were successful, the manipulation of perceived benefit (related to TC) was not. This may be due to the between-subjects design of our study. In a previous within-subjects design, where participants evaluated

XSL•FO
RenderX

both acute and chronic reports, the TC significantly impacted upload behavior [19]. It seems that participants, when comparing multiple conditions, can better discern when uploading is more or less beneficial. In our study, however, participants may have perceived the benefits of uploading as uniformly high, regardless of TC, leading to a diminished ability to detect differences.

It is clear that the adoption and approval of data-gathering technologies are strongly influenced by cultural differences [64]. In comparison to other European nations, the German population exhibits a heightened level of caution regarding the use of personal information online [65]. Given that in this study, data collection was conducted solely with residents of Germany, future studies should validate the applicability of these findings in other countries.

We deliberately excluded participants who already had a medical history with the diseases addressed in the stimuli to avoid bias in their responses. Individuals living with a stigmatized disease are more cautious to disclose the information, especially if the disease is not immediately apparent [45,66]. The question arises to what extent the behavior of stigmatized individuals can be simulated under experimental conditions. To further strengthen the validity and generalizability of our results, a follow-up study should examine the perspective of already affected individuals.

Although we captured actual click behavior, the upload decision occurred in a simulated EHR environment using fictitious diagnoses and did not involve participants' actual health data. As such, the behavior measured in this study reflects a behavioral proxy rather than real-world EHR upload behavior. Future work should examine the robustness of these findings and investigate how patients behave when real data, real accounts, and real consequences are involved.

Moreover, this was a survey study with limited immersion despite the use of an interactive click dummy. In a follow-up study, researchers could collaborate with health insurers to gather real-world data on upload behavior with a real EHR and an integrated transparency feature as used for this study. Conversely, our study faced limitations due to uncontrolled conditions like participants' locations and potential distractions, as participants completed the questionnaire online. Future research could validate our findings through a laboratory study, ensuring a more controlled environment.

Another limitation is that the distribution of our sample in terms of gender, age, and level of education does not correspond to that of the average German population [67,68]. In particular, the level of education of our sample was above average. Although we were unable to detect any effects of the control variables, gender and level of education, in the analysis, the results of this study should be validated with a more representative sample in the future.

Conclusions

Our results show that although general upload rates to the EHR are high, stigmatized diseases—even if simulated—negatively affect simulated upload behavior. However, displaying a transparency feature in the form of a PFS increases the likelihood that people upload stigmatized health data when interacting with an EHR click dummy by mitigating privacy concerns. Our findings indicate that the role of transparency features is contingent upon the level of perceived risk associated with the data to be uploaded. When the perceived risk is low, users do not need detailed privacy information to trust the technology and upload their data. However, when uploads involve sensitive data and are seen as risky, users consider privacy information and modify their upload behavior based on the information provided, potentially because transparency features help strengthen users' perceived control in such high-risk situations. Implementing transparency features in EHRs may thus help to ensure that users who perceive high privacy risks when confronted with sensitive health information are not excluded from the benefits of these systems due to privacy concerns, thereby promoting digital health equity.

Acknowledgments

The authors acknowledge support from the German Research Foundation and the Open Access Publication Fund of TU Berlin. They also thank the Evangelisches Studienwerk Villigst and the German Federal Ministry of Education and Research, which provided the doctoral scholarship (NVK) without which this research would not have been possible. They extend their gratitude to all participants of the study for their invaluable contributions.

Funding

This research was supported by the German Research Foundation and the Open Access Publication Fund of Technische Universität Berlin, which covered the publication fees. The doctoral research of NVK was supported by the Evangelisches Studienwerk Villigst and the German Federal Ministry of Education and Research.

Data Availability

The dataset analyzed in this study is available upon reasonable request. Requests to access these datasets should be directed to niklas.vkalckreuth@tu-berlin.de.

Authors' Contributions

NVK researched the literature and conceived the study under the supervision of MAF. NVK was involved in protocol development, gaining ethical approval, participant recruitment, and data analysis. NVK wrote the first draft of the manuscript. All authors

XSL•FO
RenderX

reviewed and edited the manuscript and approved the final version of the manuscript. No generative artificial intelligence tools were used in the conceptualization, data analysis, interpretation, or drafting of the scientific content of this manuscript. Only standard language-editing tools were applied.

Conflicts of Interest

NvK is a member of the JMIR Human Factors editorial board. He was not involved in the editorial or peer-review process of this manuscript. No other conflicts of interest are declared.

Multimedia Appendix 1
Case vignettes.
[PDF File, 96 KB - humanfactors_v13i1e71124_app1.pdf]

Multimedia Appendix 2
Disease descriptions.
[PDF File, 92 KB - humanfactors_v13i1e71124_app2.pdf]

Multimedia Appendix 3
Privacy fact sheets.
[PDF File, 88 KB - humanfactors_v13i1e71124_app3.pdf]

Multimedia Appendix 4
Questionnaire.
[PDF File, 77 KB - humanfactors_v13i1e71124_app4.pdf]

Multimedia Appendix 5
Demographic data of all groups.
[PDF File, 79 KB - humanfactors_v13i1e71124_app5.pdf]

Checklist 1
CONSORT-EHEALTH (v 1.6.1) checklist.
[PDF File, 1352 KB - humanfactors_v13i1e71124_app6.pdf]

References

1. Haleem A, Javaid M, Pratap Singh R, Suman R. Medical 4.0 technologies for healthcare: features, capabilities, and applications. Internet of Things and Cyber-Physical Systems 2022;2:12-30. [doi: 10.1016/j.iotcps.2022.04.001]
2. Hubmann M, Pätzmann-Sietas B, Morbach H. Telemedizin und digitale akte – wo stehen wir?: chancen und herausforderungen bei der umsetzung in klinik- und praxisalltag [Article in German]. Monatsschr Kinderheilkd 2021 Aug;169(8):711-716. [doi: 10.1007/s00112-021-01241-6]
3. Kraus S, Schiavone F, Pluzhnikova A, Invernizzi AC. Digital transformation in healthcare: analyzing the current state-of-research. J Bus Res 2021 Feb;123:557-567. [doi: 10.1016/j.jbusres.2020.10.030]
4. Bertram N, Püschner F, Gonçalves ASO, Binder S, Amelung VE. Einführung einer elektronischen patientenakte in deutschland vor dem hintergrund der internationalen erfahrungen. In: Klauber J, Geraedts M, Friedrich J, Wasem J, editors. Krankenhaus-Report 2019: Das Digitale Krankenhaus [Book in German]: Springer Berlin Heidelberg; 2019:3-16. [doi: 10.1007/978-3-662-58225-1_1]
5. Baron von Osthoff M, Watzlaw-Schmidt U, Lehmann T, Hübner J. Patientengruppenspezifische datenhoheitsbedürfnisse und akzeptanz der elektronischen patientenakte [Article in German]. Bundesgesundheitsbl 2022 Nov;65(11):1197-1203. [doi: 10.1007/s00103-022-03589-w]
6. Die elektronische patientenakte (epa) für alle [Article in German]. Bundesministerium für Gesundheit. 2021. URL: https://www.bundesgesundheitsministerium.de/elektronische-patientenakte.html [accessed 2026-01-03]
7. Dinev T, Albano V, Xu H, D'Atri A, Hart P. Individuals' attitudes towards electronic health records: a privacy calculus perspective. In: Gupta A, Patel VL, Greenes RA, editors. Advances in Healthcare Informatics and Analytics: Springer International Publishing; 2016:19-50. [doi: 10.1007/978-3-319-23294-2_2]
8. Kus K, Kajüter P, Arlinghaus T, Teuteberg F. Die elektronische Patientenakte als zentraler bestandteil der digitalen transformation im Deutschen gesundheitswesen – eine analyse von akzeptanzfaktoren aus patientensicht [Article in German]. HMD 2022 Dec;59(6):1577-1593. [doi: 10.1365/s40702-022-00921-5]
9. Verband der Ersatzkassen. Daten zum gesundheitswesen: versicherte [Article in German]. vdek. 2024. URL: https://www.vdek.com/presse/daten/b_versicherte.html [accessed 2026-01-03]

XSL•FO
RenderX

10. Lauterbach: elektronische patientenakte ab ende 2024 für alle verbindlich [Article in German]. Bundesministerium für Gesundheit. 2023. URL: https://www.bundesgesundheitsministerium.de/presse/interviews/interview/fas-030324-elektronische-patientenakte.html [accessed 2026-01-03]
11. Die elektronische patientenakte (epa) erleichtert den alltag [Article in German]. Bundesministerium für Gesundheit. 2025. URL: https://www.bundesgesundheitsministerium.de/epa-na-sicher/ [accessed 2026-01-03]
12. Hoerbst A, Kohl CD, Knaup P, Ammenwerth E. Attitudes and behaviors related to the introduction of electronic health records among Austrian and German citizens. Int J Med Inform 2010 Feb;79(2):81-89. [doi: 10.1016/j.ijmedinf.2009.11.002] [Medline: 20031482]
13. Kröner S, Schreiweis B, Strotbaum V, Brandl LC, Pobiruchin M, Wiesner M. Consumer perspectives on the national electronic health record and barriers to its adoption in Germany: does health policy require a change in communication? BMC Health Serv Res 2025 Jan 7;25(1):33. [doi: 10.1186/s12913-024-12175-6] [Medline: 39773635]
14. Panskus R, Bochum RU, Ninow M, Fahl S, Marky K, Bochum RU. Privacy mental models of electronic health records: a German case study. Presented at: Proceedings of the Nineteenth Symposium on Usable Privacy and Security; Aug 7-8, 2023 URL: https://www.usenix.org/system/files/soups2023-panskus.pdf [accessed 2026-01-03]
15. Papoutsi C, Reed JE, Marston C, Lewis R, Majeed A, Bell D. Patient and public views about the security and privacy of Electronic Health Records (EHRs) in the UK: results from a mixed methods study. BMC Med Inform Decis Mak 2015 Oct 14;15(1):86. [doi: 10.1186/s12911-015-0202-2] [Medline: 26466787]
16. “Konnte bisher noch nie gehackt werden”: die elektronische patientenakte kommt - jetzt für alle! 2024 [Article in German]. Media.ccc.de. URL: https://media.ccc.de/v/38c3-konnte-bisher-noch-nie-gehackt-werden-die-elektronische-patientenakte-kommt-jetzt-fr-alle [accessed 2026-01-03]
17. Stellungnahme zum CCC-vortrag zur epa für alle [Article in German]. Gematik. 2024 Dec 27. URL: https://www.gematik.de/newsroom/news-detail/aktuelles-stellungnahme-zum-ccc-vortrag-zur-epa-fuer-alle [accessed 2026-01-03]
18. von Kalckreuth N, Feufel MA. Influence of disease-related stigma on patients’ decisions to upload medical reports to the German electronic health record: randomized controlled trial. JMIR Hum Factors 2024 Apr 10;11:e52625. [doi: 10.2196/52625] [Medline: 38598271]
19. von Kalckreuth N, Prümper AM, Feufel MA. The influence of health data on the use of the electronic health record (EHR) – a mixed methods approach. Presented at: Americas Conference on Information Systems (AMCIS 2023) Proceedings: Diving into Uncharted Waters; Aug 10-12, 2023 URL: https://aisel.aisnet.org/amcis2023/sig_health/sig_health/2/ [accessed 2026-01-03]
20. von Kalckreuth N, Feufel MA. Disease characteristics influence the privacy calculus to adopt electronic health records: a survey study in Germany. Digit Health 2024;10:20552076241274245. [doi: 10.1177/20552076241274245] [Medline: 39247096]
21. Rau E, Tischendorf T, Mitzscherlich B. Implementation of the electronic health record in the German healthcare system: an assessment of the current status and future development perspectives considering the potentials of health data utilisation by representatives of different stakeholder groups. Front Health Serv 2024;4:1370759. [doi: 10.3389/frhs.2024.1370759] [Medline: 38800500]
22. Feng Y, Yao Y, Sadeh N. A design space for privacy choices: towards meaningful privacy control in the internet of things. Presented at: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems; May 8-13, 2021; Yokohama, Japan p. 1-16. [doi: 10.1145/3411764.3445148]
23. Sloan RH, Warner R. Beyond notice and choice: privacy, norms, and consent. SSRN Journal 2013;14(2). [doi: 10.2139/ssrn.2239099]
24. Kelley PG, Cranor LF, Sadeh N. Privacy as part of the app decision-making process. : Association for Computing Machinery; 2013 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; Apr 27 to May 2, 2013; Paris France p. 3393-3402. [doi: 10.1145/2470654.2466466]
25. Schaub F, Könings B, Weber M. Context-adaptive privacy: leveraging context awareness to support privacy decision making. IEEE Pervasive Comput 2015 Jan;14(1):34-43. [doi: 10.1109/MPRV.2015.5]
26. Meier Y, Schäwel J, Krämer NC. The shorter the better? Effects of privacy policy length on online privacy decision-making. MaC 2020 Jun 23;8(2):291-301. [doi: 10.17645/mac.v8i2.2846]
27. Fabian B, Ermakova T, Lentz T. Large-scale readability analysis of privacy policies. Presented at: Proceedings of the International Conference on Web Intelligence; Aug 23-26, 2017; Leipzig, Germany p. 18-25. [doi: 10.1145/3106426.3106427]
28. McDonald AM, Cranor LF. The cost of reading privacy policies. I/S: J Law Policy Inf Soc 2008;4(3):543-568 [FREE Full text]
29. Reidenberg JR, Breaux T, Cranor LF, et al. Disagreeable privacy policies: mismatches between meaning and users understanding. SSRN J 2015;30. [doi: 10.2139/ssrn.2418297]
30. Obar JA, Oeldorf-Hirsch A. The biggest lie on the Internet: ignoring the privacy policies and terms of service policies of social networking services. Inf Commun Soc 2020 Jan 2;23(1):128-147. [doi: 10.1080/1369118X.2018.1486870]
31. Braghin C, Cimato S, Della Libera A. Are mHealth apps secure: case study. Presented at: 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC); Jul 23-27, 2018; Tokyo, Japan p. 335-340. [doi: 10.1109/COMPSAC.2018.10253]

XSL•FO
RenderX

32. van Haasteren A, Gille F, Fadda M, Vayena E. Development of the mHealth App Trustworthiness checklist. Digit Health 2019;5:2055207619886463. [doi: 10.1177/2055207619886463] [Medline: 31803490]
33. Kulyk O, Milanovic K, Pitt J. Does my smart device provider care about my privacy? Investigating trust factors and user attitudes in IoT systems. Presented at: Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society; Oct 25-29, 2020; Tallinn, Estonia p. 1-12. [doi: 10.1145/3419249.3420108]
34. Krasnova H, Spiekermann S, Koroleva K, Hildebrand T. Online social networks: why we disclose. J Inf Technol 2010 Jun;25(2):109-125. [doi: 10.1057/jit.2010.6]
35. Schaub F, Balebako R, Durity AL, Cranor LF. A design space for effective privacy notices. In: Selinger E, Polonetsky J, Tene O, editors. The Cambridge Handbook of Consumer Privacy, 1st edition: Cambridge University Press; 2018:365-393. [doi: 10.1017/9781316831960.021]
36. Karwatzki S, Dytynko O, Trenz M, Veit D. Beyond the personalization–privacy paradox: privacy valuation, transparency features, and service personalization. J Manage Inf Syst 2017 Apr 3;34(2):369-400. [doi: 10.1080/07421222.2017.1334467]
37. Tsai JY, Egelman S, Cranor L, Acquisti A. The effect of online privacy information on purchasing behavior: an experimental study. Inf Syst Res 2011 Jun;22(2):254-268. [doi: 10.1287/isre.1090.0260]
38. von Kalckreuth N, Feufel MA. Extending the privacy calculus to the mHealth domain: survey study on the intention to use mHealth apps in Germany. JMIR Hum Factors 2023 Aug 16;10:e45503. [doi: 10.2196/45503] [Medline: 37585259]
39. Kim B, Kim D. Understanding the key antecedents of users' disclosing behaviors on social networking sites: the privacy paradox. Sustainability 2020 Jun 24;12(12):5163. [doi: 10.3390/su12125163]
40. von Kalckreuth N, Kopka M, Appel J, Feufel MA. Unlocking the potential of the electronic health record - the influence of transparency features. Presented at: Proceedings of the 32nd European Conference on Information Systems (ECIS 2024); Jun 13-19, 2024; Paphos, Cyprus p. 1-15 URL: https://aisel.aisnet.org/ecis2024/track18_healthit/track18_healthit/5/ [accessed 2026-01-03]
41. von Kalckreuth N, Feufel MA. Disclosure of health data – conceptualizing the intention to use wearables as an extended privacy calculus. Presented at: Proceedings of the Americas Conference on Information Systems (AMCIS 2021); Aug 9-13, 2021; Montreal, Canada p. 1-6 URL: https://aisel.aisnet.org/amcis2021/sig_hci/sig_hci/6/ [accessed 2026-01-03]
42. Adjerid I, Acquisti A, Brandimarte L, Loewenstein G. Sleights of privacy: framing, disclosures, and the limits of transparency. Presented at: Proceedings of the Ninth Symposium on Usable Privacy and Security; Jul 24-26, 2013; Newcastle, United Kingdom p. 1-11. [doi: 10.1145/2501604.2501613]
43. Betzing JH, Tietz M, vom Brocke J, Becker J. The impact of transparency on mobile privacy decision making. Electron Markets 2020 Sep;30(3):607-625. [doi: 10.1007/s12525-019-00332-3]
44. von Kalckreuth N, Feufel MA. Enhancing uploads of health data in the electronic health record-the role of framing and length of privacy information: a survey study in Germany. Digit Health 2025;11:20552076251351197. [doi: 10.1177/20552076251351197] [Medline: 40755954]
45. Goffman E. Stigma: Notes on the Management of Spoiled Identity: Prentice-Hall, Inc; 1963. URL: https://cdn.penguin.co.uk/dam-assets/books/9780241548011/9780241548011-sample.pdf [accessed 2026-01-03]
46. Peer E, Brandimarte L, Samat S, Acquisti A. Beyond the turk: alternative platforms for crowdsourcing behavioral research. J Exp Soc Psychol 2017 May;70:153-163. [doi: 10.1016/j.jesp.2017.01.006]
47. Keith MJ, Thompson SC, Hale J, Lowry PB, Greer C. Information disclosure on mobile devices: re-examining privacy calculus with actual user behavior. Int J Hum Comput Stud 2013 Dec;71(12):1163-1173. [doi: 10.1016/j.ijhcs.2013.08.016]
48. Keith MJ, Clark A, Masters T, Wigington C. What makes health data privacy calculus unique? Separating probability from impact. Presented at: Proceedings of the 55th Hawaii International Conference on System Sciences; Jan 4-7, 2022; Maui, HI p. 4836-4845. [doi: 10.24251/HICSS.2022.589]
49. Stablein T, Hall JL, Pervis C, Anthony DL. Negotiating stigma in health care: disclosure and the role of electronic health records. Health Sociol Rev 2015 Sep 2;24(3):227-241. [doi: 10.1080/14461242.2015.1078218]
50. Link BG, Phelan JC. Conceptualizing stigma. Annu Rev Sociol 2001 Aug;27(1):363-385 [FREE Full text] [doi: 10.1146/annurev.soc.27.1.363]
51. Cheng YH, Yeh YJ. Exploring radio frequency identification technology's application in international distribution centers and adoption rate forecasting. Technol Forecast Soc Change 2011 May;78(4):661-673. [doi: 10.1016/j.techfore.2010.10.003]
52. Lancelot Miltgen C, Popovič A, Oliveira T. Determinants of end-user acceptance of biometrics: integrating the "Big 3" of technology acceptance with privacy context. Decis Support Syst 2013 Dec;56:103-114. [doi: 10.1016/j.dss.2013.05.010]
53. Bickford J, Barton SE, Mandalia S. Chronic genital herpes and disclosure.... the influence of stigma. Int J STD AIDS 2007 Sep;18(9):589-592. [doi: 10.1258/095646207781568484] [Medline: 17784999]
54. Zacks S, Beavers K, Theodore D, et al. Social stigmatization and hepatitis C virus infection. J Clin Gastroenterol 2006 Mar;40(3):220-224. [doi: 10.1097/00004836-200603000-00009] [Medline: 16633123]
55. Berger BE, Ferrans CE, Lashley FR. Measuring stigma in people with HIV: psychometric assessment of the HIV stigma scale. Res Nurs Health 2001 Dec;24(6):518-529. [doi: 10.1002/nur.10011] [Medline: 11746080]
56. Zarei N, Joulaei H, Darabi E, Fararouei M. Stigmatized attitude of healthcare providers: a barrier for delivering health services to HIV positive patients. Int J Community Based Nurs Midwifery 2015 Oct;3(4):292-300. [Medline: 26448956]

XSL•FO
RenderX

57. Schomakers EM, Lidynia C, Vervier LS, Calero Valdez A, Ziefle M. Applying an extended UTAUT2 model to explain user acceptance of lifestyle and therapy mobile health apps: survey study. JMIR Mhealth Uhealth 2022 Jan 18;10(1):e27095. [doi: 10.2196/27095] [Medline: 35040801]
58. Uncovska M, Freitag B, Meister S, Fehring L. Patient acceptance of prescribed and fully reimbursed mHealth apps in Germany: an UTAUT2-based online survey study. J Med Syst 2023 Jan 27;47(1):14. [doi: 10.1007/s10916-023-01910-x] [Medline: 36705853]
59. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 1995 Jan 1;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
60. Cuschieri S. The CONSORT statement. Saudi J Anaesth 2019 Apr;13(Suppl 1):S27-S30. [doi: 10.4103/sja.SJA_559_18] [Medline: 30930716]
61. Cherif E, Bezaz N, Mzoughi M. Do personal health concerns and trust in healthcare providers mitigate privacy concerns? Effects on patients' intention to share personal health data on electronic health records. Soc Sci Med 2021 Aug;283:114146. [doi: 10.1016/j.socscimed.2021.114146] [Medline: 34242891]
62. Kroll T, Stieglitz S. Digital nudging and privacy: improving decisions about self-disclosure in social networks. Behav Inf Technol 2021 Jan 2;40(1):1-19. [doi: 10.1080/0144929X.2019.1584644]
63. Hatzenbuehler ML, Phelan JC, Link BG. Stigma as a fundamental cause of population health inequalities. Am J Public Health 2013 May;103(5):813-821. [doi: 10.2105/AJPH.2012.301069] [Medline: 23488505]
64. Krasnova H, Veltri NF, Günther O. Self-disclosure and privacy calculus on social networking sites: the role of culture: intercultural dynamics of privacy calculus. Bus Inf Syst Eng 2012 Jun;4(3):127-135. [doi: 10.1007/s12599-012-0216-6]
65. Custers B, Dechesne F, Sears AM, Tani T, van der Hof S. A comparison of data protection legislation and policies across the EU. Comput Law Secur Rev 2018 Apr;34(2):234-243. [doi: 10.1016/j.clsr.2017.09.001]
66. Weiss MG, Ramakrishna J, Somma D. Health-related stigma: rethinking concepts and interventions. Psychol Health Med 2006 Aug;11(3):277-287. [doi: 10.1080/13548500600595053] [Medline: 17130065]
67. Bildungsstand [Article in German]. Statistisches Bundesamt. 2020. URL: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Bildungsstand/_inhalt.html [accessed 2026-01-03]
68. Bevölkerungsstand [Article in German]. Statistisches Bundesamt. 2022. URL: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsstand/_inhalt.html [accessed 2026-01-03]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
EHR: electronic health record
IV: independent variable
mHealth: mobile health
OR : odds ratio
PFS: privacy fact sheet
SP: stigmatization potential
TC: time course

Edited by J Sarvestan; submitted 10.Jan.2025; peer-reviewed by M Greve, Y Senathirajah; revised version received 17.Nov.2025; accepted 18.Nov.2025; published 15.Jan.2026.

Please cite as:
von Kalckreuth N, Feufel MA
Privacy Fact Sheets for Mitigating Disease-Related Privacy Concerns and Facilitating Equal Access to the Electronic Health Record: Randomized Controlled Trial
JMIR Hum Factors 2026;13:e71124
URL: https://humanfactors.jmir.org/2026/1/e71124
doi:10.2196/71124

© Niklas von Kalckreuth, Markus A Feufel. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 15.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information,

XSL•FO
RenderX

a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring Medical Information Needs and Accessibility in Swedish Dental Care by Analysis of Documentation Workflows and Electronic Dental Records in Dalarna: Sociotechnical Qualitative Study

Sahid Hasan Rahim, MSc, DDS; Nadia Davoody, MSc, PhD; Stefano Bonacina, MSc, PhD

Health Informatics Centre, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Tomtebodavägen 18 A, Stockholm, Sweden

Corresponding Author:
Stefano Bonacina, MSc, PhD
Health Informatics Centre, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Tomtebodavägen 18 A, Stockholm, Sweden

Abstract

Background: Despite growing evidence demonstrating the connection between oral and systemic health, medical and dental care remain institutionally divided. A significant consequence of this division is the lack of information sharing, which is particularly problematic in dental care, where knowing patients' medical information is crucial for providing safe and effective treatments. This separation poses additional challenges in Swedish regions with limited resources, such as Dalarna, where dental care practices would benefit from improved access to relevant medical information in their electronic dental record (EDR) systems.

Objective: This study aimed to explore how current documentation workflows and EDR systems support the medical information needs within dental care practices in Dalarna and consider what influence direct access to medical information could have.

Methods: The study adopted an exploratory-descriptive qualitative approach. Semi-structured interviews were conducted with dental practitioners working in general dental practices. Data collection followed a sociotechnical framework, and thematic analysis was performed to identify key medical information needs, as well as current workflow and system limitations. Conceptual models were developed to reflect these findings.

Results: Eighteen dental practitioners were interviewed. The identified medical information needs included specific types of medical conditions, pharmacological information, treatment history, and laboratory values. Furthermore, dental practitioners highlighted substantial challenges in existing documentation workflows and the EDR system. Proposed conceptual models demonstrated how integrating EDR systems with the Swedish National Patient Overview ("Nationell Patientöversikt") via National Service Platform ("Nationell Tjänsteplattform") could streamline workflows and enhance information accessibility.

Conclusions: The findings show a clear need to improve medical information accessibility in dental care. A solution is to facilitate interoperability and align digital infrastructure with the identified needs. The proposed recommendations offer a feasible starting point for improving medical information access in Swedish dental care, particularly in resource-constrained regions like Dalarna.

(JMIR Hum Factors 2026;13:e82691) doi:10.2196/82691

KEYWORDS

dental informatics; dental care; health information exchange; health information interoperability; health information systems; medical informatics

Introduction

Medical Information in Dental Care

Comprehensive medical information is essential in dental care, as inaccurate or incomplete patient medical history assessments can lead to adverse treatment outcomes [1]. In addition to patient safety, the absence of accurate medical information exposes dental professionals to potential harm, as many procedures are invasive, carrying risks of occupational hazards such as cross-contamination and disease transmission [2]. The literature highlights several types of medical information as particularly relevant to dental care. Medical history, especially patients' diagnoses and current medications, is consistently prioritized [3-5]. Information about allergies and adverse drug reactions is also essential, as overlooking these may lead to severe or even life-threatening scenarios [3,4,6]. Additionally, details such as medical treatment plans, hospitalization history, and surgical history are considered highly valuable, as they support improved

XSL•FO
RenderX

care coordination and more effective treatment planning [3-5]. Together, these categories form the foundation of the medical information that dental professionals rely on to deliver care. Unfortunately, challenges arise due to inconsistencies in the documentation and communication of such information, particularly as it is often self-reported by patients [7]. This raises the question of whether health information technology can possibly address this gap and support more reliable access to critical medical information in dental care.

Information Technology for Record-Keeping

In dental care, patient information is primarily managed through electronic dental records (EDRs), whereas medical care relies on electronic health records (EHRs) [8-10]. The advent of such information systems has facilitated the process of recording patient information in data format, thereby giving professionals in both dental and medical care a more structured and streamlined approach to record-keeping [9,10]. This also includes the process of enabling information exchange.

Interoperability

To enable the exchange of patient information between EHR and EDR systems, interoperability is required [11]. Interoperability refers to the ability of different systems to exchange, process, and use patient data effectively, and it is commonly described across four levels. The technical level involves a shared technical infrastructure that allows systems to communicate. The syntactic level requires data to be structured in a standardized format for processing. The semantic level ensures that data are interpretable and meaningful across systems. Finally, the organizational level emphasizes that interoperability at the technical, syntactic, and semantic levels depends on alignment with regulatory, administrative, and policy frameworks, including agreements that enable data exchange while ensuring compliance with laws and regulations.

Current Dental-Medical Interoperability Efforts

The literature identifies two main approaches to achieving interoperability between dental and medical care systems. Health information exchange (HIE) serves as an intermediary to facilitate data exchange between different systems [3,12-14]. Integrated EHR-EDR (IEHR), on the other hand, operates on a single-system architecture [14]. This system has become widely adopted by organizations offering integrated dental-medical care services [15].

While both IEHR and HIE aim to facilitate information exchange between dental and medical care, their feasibility depends on organizational size, infrastructure, and financial resources. IEHR can provide data conformity and reliability, but may be financially and technically restrictive for smaller organizations [4,7,14]. In contrast, HIE can offer a more flexible and cost-effective solution, though usability barriers may impede its adoption [3,12-14].

Understanding how these approaches are implemented within a specific national context is essential for assessing their real-world applicability. Here, Sweden offers a particularly instructive case due to its publicly funded health care system, high digital maturity, and ongoing efforts to advance HIE [16,17]. In fact, Sweden's broader aim is to become a global leader in health care digitalization by 2025, a vision that includes both medical and dental care [18].

The Swedish Interoperability Strategy

Sweden is a country divided into 21 administrative regions, each governing its publicly funded medical and dental care services autonomously [19]. Due to this decentralized administrative structure, challenges have arisen in ensuring the accessibility and exchange of patient data across regions and organizations [17,20]. Hence, the Swedish government enacted the Integrated Health and Social Care Documentation Act (2022:913), establishing a legal basis for the sharing of patient information across health care services [21]. This legal foundation was followed by the implementation of nationwide digital infrastructure, the national patient overview (nationell patientöversikt; NPÖ), and the national service platform (nationell tjänsteplattform; NTjP) [22-24]. NPÖ is a nationwide digital platform primarily provided to public health care providers, allowing authorized health care professionals to retrieve patient data across organizations and regions [23]. NTjP refers to the technical infrastructure that facilitates the exchange of patient data in a standardized and secure manner [22,24]. This concept is similar to HIE, as it acts as an intermediary layer between different systems in organizations and regions.

Currently, the digital infrastructure does not support information exchange between dental and medical care services [25]. It may pose particular challenges in regions where access to both services is already constrained, particularly in underserved or rural regions [26-29]. To illustrate the relevance and urgency of such an approach, the region of Dalarna provides a meaningful context for further discussion.

The Dalarna Region

One of the most pressing issues in Dalarna is the shortage of dental professionals, with the lowest dentist-to-population ratio in Sweden [30]. Due to its vast land area, patients in remote regions may need to travel long distances to reach a general dental practice (GDP). In addition, Dalarna has an aging population, with nearly 25% of residents aged 65 years or older, compared to the national average of 19% [31]. This demographic trend places further strain on an already limited dental workforce, as older adults often require more frequent and complex dental and medical care [32]. Facilitating access to medical information in dental care in Dalarna could therefore support more effective and higher-quality care outcomes for patients.

Problem Description

Sweden's digital infrastructure, the NPÖ and the NTjP, support information exchange between health care information systems across regions and organizations [22-24]. However, these platforms do not currently provide access to medical information in dental care [25]. This lack of integration presents a missed opportunity to enhance both patient safety and efficiency of dental workflows, especially in rural regions where access to medical and dental services is limited [26-29]. The Dalarna region exemplifies this challenge, facing one of the most severe shortages of dental professionals in the country. This shortage

XSL•FO
RenderX

forces many residents to travel long distances for dental care [30]. Yet, addressing this gap does not necessarily require extensive structural or financial reforms. Rather, a pragmatic path forward may lie in leveraging existing infrastructure to enable dental professionals to access relevant medical information directly through their EDR systems. However, without a clear understanding of what specific medical information is needed and how it fits into dental workflows, such interoperability efforts risk being ineffective or even disruptive. Although existing literature reports the types of medical information relevant to dental care [3-5], it lacks specificity. Therefore, it is crucial to identify the specific medical information needs of dental professionals and examine how these can be supported within current workflows, systems, and organizational structures. This necessitates a sociotechnical approach that accounts for both technical components and end-users [33,34], primarily dental professionals.

Aim

This study aims to explore how current documentation workflows and EDR systems in dental care practices in the Dalarna region in Sweden support or limit the fulfillment of the medical information needs, and to examine the potential impact of having direct access to accurate and relevant medical information within EDR systems.

Methods

Research Design

This study adopted an exploratory-descriptive qualitative research approach to allow for a comprehensive yet structured and contextualized examination of the experiences, challenges, and needs on the subject matter [35].

Study Setting

The study was decided to be conducted in GDPs operated under the Public Dental Service Dalarna (PDSD), which is administered by the regional authority Region Dalarna and is responsible for providing dental care services within the region [36]. Consequently, private dental clinics or organizations and specialist clinics were excluded.

Participants and Sampling

The primary participants consisted of dental practitioners, as they are directly responsible for patient care [37]. Only dental practitioners having their main professional duties within GDPs operated by PDSD were considered. Purposive sampling was used to identify participants who were able to provide detailed, precise, and reliable information aligned with the study's aim [38]. Conversely, other dental professionals were excluded.

Data Collection

Interviews were deemed the most appropriate data collection method as they offered an opportunity for participants to share their perspectives in-depth related to the subject matter [39]. The interviews followed a semi-structured format to ensure a balance between systematic inquiry and flexibility. The interviews needed to capture the sociotechnical aspects relevant to the subject matter. To achieve this, an interview guide was designed following a sociotechnical framework. The sociotechnical framework defined by Sittig et al [40] provides a comprehensive lens for evaluating health information technology by examining the interaction between information systems and the social context in which they operate. It comprises 8 dimensions, where for this study, 6 dimensions were chosen based on their alignment with the research aim [40]: (1) clinical content to identify the specific medical information considered important and what should be accessible in EDR systems; (2) people to explore experiences and challenges in obtaining medical information from patients, the medical domain, or other sources; (3) human-computer interface to assess the usability of the current EDR system regarding accessibility and documentation of medical information; (4) workflow and communication to understand how medical information is obtained and documented within current workflows, focusing on the patient and EDR system; (5) internal organizational policies, procedures, and culture to examine existing policies on documentation and retrieval of medical information within PDSD; and (6) external rules, regulations, and pressures to explore the awareness of guidelines or requirements for documenting medical information in dental care. These dimensions informed the development of the interview guide in Swedish and English (Multimedia Appendix 1).

Assessment and Pilot Testing

Following development, the interview guide was reviewed by the co-authors, SB, ND, and the Head of Division at PDSD, to confirm its suitability. The guide was subsequently pilot-tested, which confirmed that the questions elicited comprehensive responses. Individuals involved in the review and pilot testing were excluded from the final study to avoid diluting the results. A bilingual consent form in both Swedish and English was developed in parallel (Multimedia Appendix 2).

Recruitment and Interview Sessions

The head of division at PDSD facilitated recruitment by distributing interview invitations to GDPs. Interviews were conducted digitally via Microsoft Teams between March 7 and July 28, 2025. Digital interviews were chosen to accommodate participants from various locations and minimize travel-related costs. All interviews were audio-recorded and transcribed using Microsoft Teams' built-in tool. Supplementary notes were taken during each session to capture contextual and paralinguistic cues. Transcripts were carefully reviewed, compiled, and returned to participants to confirm accuracy. Interviewing concluded when data saturation was reached, meaning no new information emerged, and the data were sufficiently rich to address the study's research questions [41].

Data Analysis

Thematic Analysis

Thematic analysis was conducted to identify and interpret patterns within the interview data, guided by the sociotechnical framework from Sittig et al [40]. The analysis followed Braun and Clarke's 6-phase approach [42]. Transcripts were read repeatedly to gain an in-depth understanding of the data, with audio recordings revisited when meaning was unclear. Excerpts

XSL•FO
RenderX

were then collated into potential themes, which were reviewed and refined to ensure they accurately represented the data. Themes were organized under the following columns: sociotechnical dimension, participant (anonymized), excerpt, subtheme, and theme. The entire analysis was iterative, with frequent revisions of subthemes and themes to ensure a valid and coherent representation of the findings [43].

Conceptual Modeling

Following the analysis, conceptual models were developed to illustrate current and proposed workflows for accessing medical information in dental care. These models considered integration with existing Swedish infrastructures. Unified Modeling Language (UML) was used due to its use in health system planning and stakeholder communication [44,45]. Three UML diagrams were created. Activity diagrams were used to show current and improved workflows for obtaining and documenting medical information in dental care. Use case diagrams were developed to present expected system functionalities derived from the interview data. Finally, a sequence diagram was created to illustrate how medical information could be exchanged from regional or national infrastructures to local EDR systems.

Ethical Considerations

The study was carried out in Sweden. According to the Swedish Ethical Review Act (SFS 2003:460) [46] and guidance from the Swedish Ethical Review Authority [47], the type of research presented in this article does not require formal ethical approval, as it does not involve sensitive personal data as defined by the European Union General Data Protection Regulation (EU 2016/679) [48]. Nonetheless, we emphasize that all ethical standards were strictly followed in accordance with relevant legislation, the Declaration of Helsinki [49]. Participants were fully informed about the study's purpose, their rights, and data handling procedures. Informed consent was obtained both in writing and verbally. Participation was voluntary, and withdrawal was permitted at any time. Anonymity was guaranteed, and participants were informed that findings would be publicly disseminated.

Results

Descriptive Overview of the Interview Sample

In total, 18 interviews were conducted, after which no new information emerged beyond the 11th interview. All participants reported using the T4 EDR system, developed by Carestream [50]. Fifteen participants identified as female and three as male. Key characteristics of the participants are presented in Table 1. All participants treated diverse groups of patients and performed a broad range of general dental treatments, where 8 participants identified oral surgery as their main focus. All participants primarily used free text.

XSL•FO
RenderX

Table . Key characteristics of the participants.

Participants	Age (year range)	Work experience overall (year range)	Work experience Dalarna (year range)	Scope of practice
P1	40 - 45	10 - 15	10 - 15	General Dentistry and Oral Surgery
P2	25 - 30	0 - 5	0 - 5	General Dentistry
P3	25 - 30	0 - 5	0 - 5	General Dentistry
P4	25 - 30	0 - 5	0 - 5	General Dentistry, Pediatric Dentistry, and Orthodontics
P5	25 - 30	0 - 5	0 - 5	General Dentistry and Oral Surgery
P6	25 - 30	0 - 5	0 - 5	General Dentistry
P7	45 - 50	20 - 25	10 - 15	General Dentistry, Oral Surgery, Pediatric Dentistry, and Orthodontics
P8	25 - 30	0 - 5	0 - 5	General Dentistry, Oral Surgery, and Prosthodontics
P9	45 - 50	25 - 30	5 - 10	General Dentistry and Oral Surgery
P10	25 - 30	0 - 5	0 - 5	General Dentistry and Oral Surgery
P11	25 - 30	0 - 5	0 - 5	General Dentistry, Oral Surgery, and Pediatric Dentistry
P12	20 - 25	0 - 5	0 - 5	General Dentistry
P13	25 - 30	0 - 5	0 - 5	General Dentistry
P14	35 - 40	5 - 10	0 - 5	General Dentistry, Oral Surgery, and Prosthodontics
P15	35 - 40	0 - 5	0 - 5	General Dentistry
P16	25 - 30	0 - 5	0 - 5	General Dentistry
P17	35 - 40	5 - 10	5 - 10	General Dentistry and Community Dentistry
P18	25 - 30	0 - 5	0 - 5	General Dentistry

The Specific Medical Information Needs – Clinical Content

Under the sociotechnical dimension of "clinical content," the interview data revealed three overarching categories of medical information needs: "conditions, pharmacological information, treatment history and laboratory values." Each category contains specific types of information that participants deemed essential for providing dental care (Table 2).

XSL•FO
RenderX

Table . Specific medical information identified within the sociotechnical dimension "clinical content."

Themes	Participants
Conditions	
Oncological diagnosis	P3, P5, P7, P8, P9, P11, P12, P13, P15, P16, P17, P18
Psychiatric and neurological disorders	P2, P4, P7, P11, P12, P14, P18
Cardiovascular diseases	P2, P3, P5, P8, P13, P14, P15, P16, P17, P18
Hematologic disorder	P1, P15, P17
Hepatic diseases	P3
Pulmonary diseases	P3, P17
Renal diseases	P3, P17, P18
Rheumatism	P3
Diabetes	P7, P14
Sleep apnea	P9
Infectious disease	P11, P17
Allergies in general	P2, P12, P14, P15, P18
Pharmacological information	
Penicillin allergy	P11, P14, P16
Substance abuse	P11
Comprehensive medication list	P1, P3, P4, P6, P10, P12, P15
Anticoagulants	P1, P2, P5, P6, P7, P8, P13, P14, P16, P17, P18
Antiresorptive agents	P1, P2, P3, P6, P7, P8, P9, P10, P11, P12, P13, P14, P16, P17, P18
Psychiatric medications	P2, P4, P11, P12, P16, P17, P18
Treatment history and laboratory values	
Ongoing oncological treatment	P5, P7, P8, P9, P11, P12, P16, P17
Surgical procedures	P1, P5, P10, P13, P18
Blood coagulation marker	P8, P11, P14

Conditions

Among conditions, oncological diagnoses and cardiovascular diseases were the most frequently emphasized because of the associated risks when providing invasive treatments. For patients with oncological diagnoses, participants highlighted that treatment planning and risk management are directly influenced by the potential effects of cancer on oral healing and immune response. Similarly, participants reported that patients with cardiovascular diseases require special attention due to systemic risks associated with invasive procedures:

> *Knowing if a patient has cancer or has had it in the past can influence the treatment we provide in terms of healing outcomes […] healing can be delayed, and there is a greater risk for them to get infected.* [P9]

> *To be aware of patients with cardiovascular diseases, where the choice of anesthesia matters. There is also a bleeding risk, and some may need antibiotic prophylaxis.* [P14]

Psychiatric and neurological disorders were also commonly mentioned due to their impact on patient cooperation and communication:

> *Neurodivergent conditions matter, particularly when working with children. It influences how you communicate and treat them.* [P4]

Allergies in general were also deemed important, as considerations need to be taken in relation to the materials used in dental care:

> *Some patients have sensitivities to certain materials, so we need to be careful. This is particularly important to know in [dental] prosthetic treatments, where you are dealing with crowns or removable dentures.* [P14]

Additionally, infectious diseases were highlighted due to their relevance in maintaining a safe clinical environment for both patients and dental professionals:

XSL•FO
RenderX

> *Being aware of any infectious diseases is important, not only for our own safety but also to prevent cross-contamination between patients.* [P11]

Other conditions were brought up by individual participants. Though each was mentioned less frequently, they reflect the complexity of the oral-systemic health correlation in the context of coordinating dental care:

> *Patients often have several different diagnoses, such as heart problems, kidney issues, liver disease, or lung conditions, and each of those can have a real impact on oral health. Sometimes it is not clear how it all is related [for the patient], but it definitely affects how we manage them chairside.* [P3]

> *For patients with diabetes, wound healing is affected, and there is a higher risk of infection and periodontal disease.* [P14]

Pharmacological Information

The pharmacological information theme comprises medications that participants regarded as critical, primarily due to the procedural risks associated with specific drugs. The most frequently mentioned medication category was antiresorptive agents and anticoagulants due to complications associated with dental treatments:

> *If we miss that someone has been on multiple blood thinners or bisphosphonates, and they do not mention it because they're no longer taking them or do not think it is important [to mention], it can lead to problems like excessive bleeding or delayed wound healing.* [P13]

Additionally, several participants expressed the importance of knowing whether patients are taking psychiatric medications, since that would increase their understanding of patient behavior and anxiety-induced responses during care:

> *I would actually like to see more information about psychiatric medications and prescriptions. I feel like we often miss that type of medication. We could better tailor our care for these patients, particularly in cases where these medications, combined with anesthetics or sedatives, affect how a patient responds to treatment, like increased anxiety.* [P2]

Penicillin allergies and substance abuse were also mentioned, as there were concerns about prescribing antibiotics or analgesics, particularly due to risks of adverse reactions or potential misuse by patients:

> *Someone who lived in another region started coming to clinics here in Dalarna asking for morphine. We did not know about the substance abuse at first, but then we received a warning from the local pharmacy saying we should not prescribe anything to her. […] If the patient is allergic to penicillin, we need to know so we do not prescribe it.* [P11]

Treatment History and Laboratory Values

In treatment history and laboratory values, participants frequently considered ongoing oncological treatment as essential information for treatment planning and to prevent complications:

> *If they [patients] undergo radiation therapy [for cancer], it is really important that we know about it. Not just that they had treatment, but also which region was irradiated. That kind of detail affects how we approach even routine dental procedures.* [P9]

Surgical procedures were also referenced since such information is used to assess potential risks and determine whether treatment should be delayed:

> *If someone has had an operation, we need to know, because they may have to receive antibiotic prophylaxis. I may also decide to postpone treatments.* [P18]

Likewise, information on blood coagulation markers was commonly sought to support decisions around the timing of treatments for patients and to evaluate their bleeding risks:

> *For patients on warfarin, having access to their International Normalized Ratio value is essential. It helps me determine whether I can proceed with treatment immediately or if I need to postpone or take extra precautions.* [P8]

Current Workflow: People

The sociotechnical dimension of people focused on identifying who the participants interact with in relation to obtaining medical information. As summarized in Table 3, it revealed that medical information is predominantly derived from patient reporting, proxy reporting, or through medical consultation. In general, the findings demonstrated how human factors influence the workflow, suggesting the need for medical information to be made more accessible.

XSL•FO
RenderX

Table . Sources for obtaining medical information, categorized under the sociotechnical dimension of "People".

Themes	Participants
Patient reporting	
Verbal recall	P1, P2, P3, P4, P6, P8, P9, P12, P14, P15, P16, P17, P18
Medication list	P5, P7, P8, P10, P11, P13, P15, P16, P17, P18
Online patient portal	P2, P8, P10, P11, P13, P18
Incomplete or inaccurate reporting	P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P13, P14, P15, P16, P17, P18
Proxy reporting	
Companions	P1, P2, P17
Care home	P1, P2, P5, P12
Direct relatives	P5, P11, P17
Challenges in proxy reporting	P1, P2, P11, P12
Medical consultations	
Primary care	P1, P2, P4, P5, P6, P8, P10, P11, P13, P15, P17
Hematology	P5, P7, P9, P12
Cardiology	P9, P11, P18
Oncology	P1, P9
Oral and Maxillofacial Surgery	P1, P10
Nephrology	P9
Coordination limitations	P1, P2, P4, P5, P6, P7, P8, P10, P13, P14, P15

Patient Reporting

Participants reported that medical information was often obtained through verbal recall, medication lists, or by asking patients to access their online patient portal. Verbal recall relied on patients remembering their medical history; medication lists were paper documents brought in by patients; and the online patient portal was used to verify details when verbal reports were insufficient, or no list was available:

> *Medical history is mostly obtained from the patients themselves verbally.* [P4]

> *Sometimes I ask the patient to log in to their online patient portal. The elderly do not know how, so I usually rely on what they say, or ask them to bring a medication list later.* [P8]

A recurring issue was incomplete or inaccurate reporting, mainly due to patients forgetting diagnoses or medications, particularly amongst the elderly or those without a caregiver:

> *Older patients may not know their diagnoses or medications. Also, some patients deliberately leave things out. Just asking is not always enough.* [P4]

Participants also noted that patients often unintentionally omitted relevant details or assumed they had nothing to report:

> *Some patients say that they do not take any medications, but then later it turns out that they take seven!* [P3]

Some also faced barriers when patients were unwilling or unable to disclose their history due to communication breakdowns:

> *Sometimes there are communication difficulties, particularly with foreign patients who may struggle with the [Swedish] language. Some patients are also unwilling to share their medical history.* [P14]

Proxy Reporting

There are situations where medical information is obtained through proxy reporting, which in this case refers to: direct relatives, companions, or care home personnel. Companions refer to those physically accompanying the patient. There were challenges presented here as well:

> *I remember one patient who arrived with a care home worker who did not know the patient's medical history. The patient had a neurodegenerative condition, so we were unable to obtain the information directly. We tried contacting the care home, but there was no nurse available at the time. Eventually, we received a phone number for a direct relative, who was able to provide us with the necessary information.* [P2]

Medical Consultations

Participants described engaging in medical consultations to obtain relevant information or seek advice on complex cases. Primary care units were the most frequent point of contact, although secondary care was also involved, depending on the patient's condition:

> *I would estimate that I contact primary care units about 1–3 times per month. Occasionally, I call specific doctors directly to confirm details if I suspect*

XSL•FO
RenderX

> *the information provided by the patient is inaccurate or incomplete.* [P4]

> *I contact medical departments several times a month. Hematology for chemotherapy patients, oncology for treatment schedules, as well as nephrology and cardiology for patient-specific matters.* [P9]

However, participants often encountered coordination limitations, such as delayed or absent responses, which hindered decision-making and delayed treatments:

> *The main issue with calling medical professionals is that they are hard to reach. You often have to call multiple times, and that takes time. This often leads to postponing treatments, even for emergency cases. I cannot proceed if I am not sure about a patient's medical history.* [P4]

Some reported being redirected or denied help, further complicating access to essential information:

> *I once called a primary care unit, and they responded strangely. They asked how I had gotten their number. Even after I explained that I was a dental practitioner and provided the necessary patient details, they refused to share any information.* [P5]

To cope with these challenges, some participants began asking patients to coordinate directly with their medical providers:

> *I once tried to contact a pediatrician, but it was very difficult to reach them. They had many patients, and it took nearly a year to get a phone appointment. Since then, I have let the patients handle the contact themselves. […] They [patients] ask their doctor, for example, whether they need antibiotic prophylaxis or if they should stop their blood thinner before an extraction, and then they get back to us with the answer.* [P14]

Current Workflow: Human-Computer Interface

The sociotechnical dimension of the human-computer interface examined the interaction with the EDR system when retrieving, documenting, and managing medical information (Table 4). Overall, participants emphasized the interface's critical role in the efficient and reliable handling of medical information, revealing both challenges and opportunities for improving usability.

Table . Aspects of current interface design and expressed design needs, categorized under the sociotechnical dimension of "human-computer interface."

Themes	Participants
Free text input	
Medical history section	P1, P4, P8, P9, P12, P14, P15, P16
Progress notes	P2, P3
Adaptive documentation behavior	P5, P6, P7, P10, P11, P13, P17, P18
Current interface limitations	
Information reliability	P1, P3, P12, P13, P17, P18
Information visibility	P2, P7, P9, P10, P15, P17
Manual data entry	P4, P6, P11, P12, P13, P14, P16, P17, P18
Design needs	
Annotations	P5, P6, P18
Improved information overview	P6, P15, P16, P17, P18
Confirm review	P6
Cross-disciplinary substance abuse alert	P10, P11
Integrated external data accessibility	P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P12, P13, P14, P15, P16, P17, P18

Free Text Input

All participants relied solely on free text input when documenting medical information in the EDR. This was done either in the medical history section (a structured form with predefined fields) or in progress notes (where other dental care activities are recorded).

> *There is a medical history section form that covers common illnesses like asthma, osteoporosis, cancer, diabetes, cardiovascular disease, liver and kidney diseases, rheumatism, mental health issues, history of radiation therapy, bleeding disorders, allergies, etc. If yes to any question, we ask them what medications they take and type it in.* [P14]

Some participants preferred documenting directly in progress notes due to time constraints and navigational hurdles:

> *In the EDR, we have a medical history section to fill in information. […] However, due to time constraints, I do not have time to go through everything, so I type in the progress notes instead.* [P2]

Most adopted an adaptive documentation behavior, using both sections to reinforce critical information:

XSL•FO
RenderX

> *I use the dedicated medical history section for documenting medical information. If there is something important, like bisphosphonates, I also type it in the progress notes.* [P6]

Current Interface Limitations

The category current interface limitations revealed several usability issues. One major concern was information reliability, where participants questioned the accuracy and completeness of previously entered medical information in the EDR system:

> *The problem with the current system is that we do not know if the medical information is recorded accurately or not by previous colleagues, or even if it is reliable.* [P1]

Another issue was information visibility, where inconsistent documentation across sections made it difficult to locate medical details:

> *If medical information is buried in the progress notes, it takes time to find. Time is also wasted when it is missing or outdated.* [P7]

Manual data entry was also a major limitation. Participants described it as time-consuming and prone to errors, especially when documenting long medication lists:

> *When there is a lot of medical information, entering it manually is a burden because it takes time, which could have been spent with the patient. Also, correctly spelling the names under time pressure can lead to mistakes.* [P17]

Design Needs

The findings in design needs emphasized the need for better system integration, interface enhancements, and improved visibility of patient data. All participants expressed a strong preference for integrated external data accessibility, where medical information should be directly accessible within the EDR:

> *Compared to EHR systems, which display all medications and diagnoses, EDR systems fall short. […] Our work is also time-pressured, which is why an integrated solution providing access to medical information would save time. […] Even if in another program, that is still better than nothing.* [P5]

Some also emphasized the need for access control, ensuring that only authorized personnel could view sensitive data:

> *Ideally, all relevant medical information should be retrieved in a controlled process to safeguard patients' privacy.* [P15]

Additional interface improvements suggested included:

> *Annotations: If there was a built-in description of each entry, that would be incredibly helpful. Like a pop-up suggestion that says what a particular drug is for.* [P5]
>
> *Improved Information Overview: An overview of the [medical] information is preferable, but it must be possible to filter it, so the overview stays clear and focused.* [P17]
>
> *Confirm Review Feature: It would be best if there was a function where you could sign to confirm that you have reviewed the medical history. […] to ensure accountability.* [P6]
>
> *Substance Abuse Alerts: We had a patient […] who seeks emergency care only to obtain narcotic medications. […] In EHR systems, this patient has a warning that flags up for that reason. I would like to see such functionality integrated into the EDR.* [P10]

Current Workflow: Workflow and Communication

Workflow and communication focused on broader clinical activities involving interactions with both the EDR system and patients in the context of managing medical information (Table 5). Overall, the findings indicate that workflow efficiency and patient safety are strongly linked to the availability and clarity of medical information. When such information is incomplete or difficult to retrieve, it disrupts clinical routines and compromises care.

Table . Challenges and limitations of the current workflow, as well as perceived benefits of direct medical information access, categorized under the sociotechnical dimension of "workflow and communication."

Themes	Participants
Current documentation workflow	
Time-consuming	P1, P2, P5, P6, P9, P10, P11, P13, P14, P15, P16, P17, P18
Delay treatments	P1, P2, P3, P4, P5, P6, P8, P11, P13, P16
In-treatment hazards	P1, P4, P7, P8, P13, P18
Patient frustration	P8, P10
Direct information accessibility	
Clinical time management	P1, P2, P4, P5, P7, P8, P9, P10, P11, P12, P13, P14, P16, P17, P18
Improved treatment planning	P3, P7, P8, P17
Cognitive load	P3, P12
Better decision-making	P4, P6, P7, P12, P13, P14, P15, P17, P18

XSL•FO
RenderX

Current Documentation Workflow

Participants overwhelmingly described the current documentation workflow as time-consuming, marked by frequent disruptions when obtaining and recording medical information:

> *Asking patients about their medical history, then reading and documenting it accurately, takes up valuable time from the appointment.* [P18]

These issues often led participants to delay treatments, which often involved canceling or rescheduling appointments:

> *"Sometimes, I cannot proceed with the treatment if I do not have the patient's full medical history. I have to explain to the patient that I cannot do anything at that moment because it would pose a risk to their health.* [P2]

In some instances, essential medical information would emerge during procedures, posing potential in-treatment hazards**:**

> *One time, a patient told me he was healthy, and just before the [tooth] extraction, he casually mentioned […] a serious condition. […] We had to stop the procedure immediately.* [P7]

Finally, repeated appointments caused by information gaps contributed to patient frustration:

> *I had to postpone the treatment […] The coordination turned out to be difficult, and the patient became increasingly frustrated.* [P8]

Direct Information Accessibility

Having direct information accessibility was consistently identified as a key improvement. The most frequently perceived benefit was enhanced clinical time management, as direct access would reduce administrative tasks and allow more focus on patient care:

> *"It would be incredibly helpful to have access to patients' full medical history. Then, I would not need to ask the patients because I would be able to look it up myself directly. That way, I could plan treatments faster instead of calling around for additional information.* [P2]

Others highlighted how improved treatment planning would be possible by enabling better scheduling and avoiding unnecessary visits:

> *If I had direct access, I would be able to plan better. Like, if a patient takes a certain medication, I would know that they need to come within 6 months before their next dose.* [P3]

Also, better decision-making was noted as a distinct benefit. Several participants stressed that it would reduce uncertainty, improve efficiency, and enhance patient safety:

> *Having direct access to medical information would speed up decision-making and let patients get treated faster. I would also feel more secure making the right call.* [P6]

A final point was the reduction of patients' cognitive load, as they would no longer need to recall or bring documentation themselves:

> *Patients would not have to recall or remember to bring documents, which would offload them of responsibilities.* [P3]

Current Workflow: Internal Organizational Policies, Procedures, and Culture

The findings from internal organizational policies, procedures, and culture suggest a need for stronger organizational support, more equitable access to medical information, and support for consistency in information management practices (Table 6). These are in relation to managing medical information within PDSD, but also in the context of Region Dalarna.

Table . Perspectives of organizational and procedural aspects categorized into the sociotechnical dimension of "internal organizational policies, procedures, and culture."

Themes	Participants
Current documentation practices	
Existing routines	P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P12, P13, P14, P15, P16, P17, P18
Inconsistencies	P2, P5, P7, P9, P11, P18
System access limitation	
Inaccessible medication list platform	P3, P4, P5, P6, P8, P11, P12, P13, P14, P16, P17, P18
EHR[a] access inequities	P5, P7, P9, P10, P11, P15, P17
Organizational support	
Organizational inaction	P1, P9, P15
Devalued	P11, P17

[a]EHR: electronic health record.

XSL•FO
RenderX

Current Documentation Practices

In the category of current documentation practices, it was widely acknowledged that there are existing routines in place. The existing routines involved checking and updating information for every appointment, as well as reminding patients to bring their medication lists:

> *We must always complete and review the patient's medical history before starting treatment, in order to check whether anything has changed since their last visit.* [P3]

> *The routine is that, when booking an appointment, the receptionist tells patients to bring their medication list.* [P9]

It also involved using a dedicated medical history section of EDR. Yet many participants noted inconsistencies in how these were applied, for example, documenting in progress notes or not updating medical information at all:

> *We have a dedicated medical history section that patients fill out for every visit. Though some colleagues only record information in the progress notes instead of using it. That makes things difficult because you then have to dig through the entire record to find relevant information.* [P5]

System Access Limitations

An initiative mentioned by participants was "Förskrivningskollen" [51], an online platform enabling authorized users to view patients' prescribed medications. However, it was described to be inaccessible due to technical issues or not provide enough information:

> *There is a platform that we technically have access to […]. It is called 'Förskrivningskollen.' While it shows prescriptions, it does not indicate what the patient is actually taking or the reasons behind those prescriptions.* [P4]

The second barrier concerned EHR access inequities. While hospital-based or specialized dental clinics within PDSD have access to EHR systems, GDPs do not. Participants viewed this as an unjustified discrepancy:

> *Dental practitioners working in hospital settings have access to EHR systems […] I asked around why we do not have the same access in GDPs, but no one could give me a clear answer.* [P9]

Organizational Support

Organizational support reflects how participants perceived the level of support provided for managing medical information in dental care, where organizational inaction refers to the lack of follow-through on identified issues, even after concerns were raised, and devalued captures the sense of not being prioritized by the broader organization:

> *People often do not realize that dental treatments also depend on knowing about diseases and medications. I think many decision-makers do not understand our professional needs. We are frequently treated like the poor cousins of medical professionals, but what we do has a big impact. We do not want the information out of curiosity. We need it to work safely and effectively.* [P11]

> *Dentists need access to relevant medical information, and physicians also need greater knowledge of how diseases and medications affect oral health. More collaboration is needed, because it is ultimately about the patient.* [P17]

Current Workflow: External Rules, Regulations, and Pressures

The last sociotechnical dimension examined was external rules, regulations, and pressures. What can be derived from Table 7 is a need for clearer guidance and communication regarding applicable regulations in dental care regarding the subject matter.

Table . Overview of the categorization under the sociotechnical dimension of "external rules, regulations, and pressures."

Themes	Participants
External regulations	
Awareness	P1, P2, P3, P5, P6, P10, P11, P14, P16
Standard practice	P4, P7, P8, P9, P12, P15, P17, P18
Patient consent	P5, P6, P11, P12, P14, P17, P18

External Regulations

The results show that there is uncertainty about external regulations guiding the management of medical information in dental care. While many participants demonstrated awareness of existing regulations, several assumed that obtaining and documenting medical information is standard practice:

> *I do not know of any rules that tell us explicitly to record medical history. I think it is obvious that we need to.* [P9]

> *National guidelines require us to document patients' medical information.* [P16]

Furthermore, the role of patient consent was perceived as crucial for retrieving medical information from both internal and external sources:

> *Yes, the Patient Data Act states that dental practitioners and hygienists may access medical information, but only what is necessary for [providing] dental care. However, if we want to know patients' medical history, we need their consent.* [P14]

Current Workflow: Conceptual Model

A UML activity diagram was developed to illustrate the current workflow for managing medical information in dental care based

XSL•FO
RenderX

on the interview data (Multimedia Appendix 3). It outlines the key steps involved in acquiring, updating, and documenting medical information in the EDR system before proceeding with treatments. The diagram includes multiple actors, including receptionist, patient, dental practitioner, accompanying proxy, proxy, and medical care unit, each with defined responsibilities represented within separate partitions.

The workflow begins with the receptionist scheduling the appointment and reminding the patient to bring their medication list. Upon arrival, the dental practitioner checks the EDR to determine whether the patient is new or returning. For returning patients, the medical history section and progress notes are reviewed to assess whether updates are needed. If updates are required, or the patient is new, the dental practitioner checks if an accompanied proxy is present and able to provide the necessary information. If not, alternative steps may involve contacting another proxy or a relevant Medical Care Unit.

If no proxy is involved, the dental practitioner evaluates whether the patient can provide sufficient information via verbal recall, a medication list, or their online patient portal. If these sources are insufficient, attempts are made to contact a proxy or medical care unit, with the option to postpone the appointment.

Once the necessary information is obtained, it is documented in the EDR, either in the medical history section or progress notes. The dental practitioner may seek further medical consultation if needed. Once all relevant information is confirmed and documented, treatment can proceed.

Potential Workflow: Conceptual Model

A UML activity diagram of a proposed workflow reflects potential improvements if medical information were made accessible in the EDR system (Multimedia Appendix 4). This revised workflow removes several steps previously required to manage information gaps and instead introduces a more linear progression toward care provision.

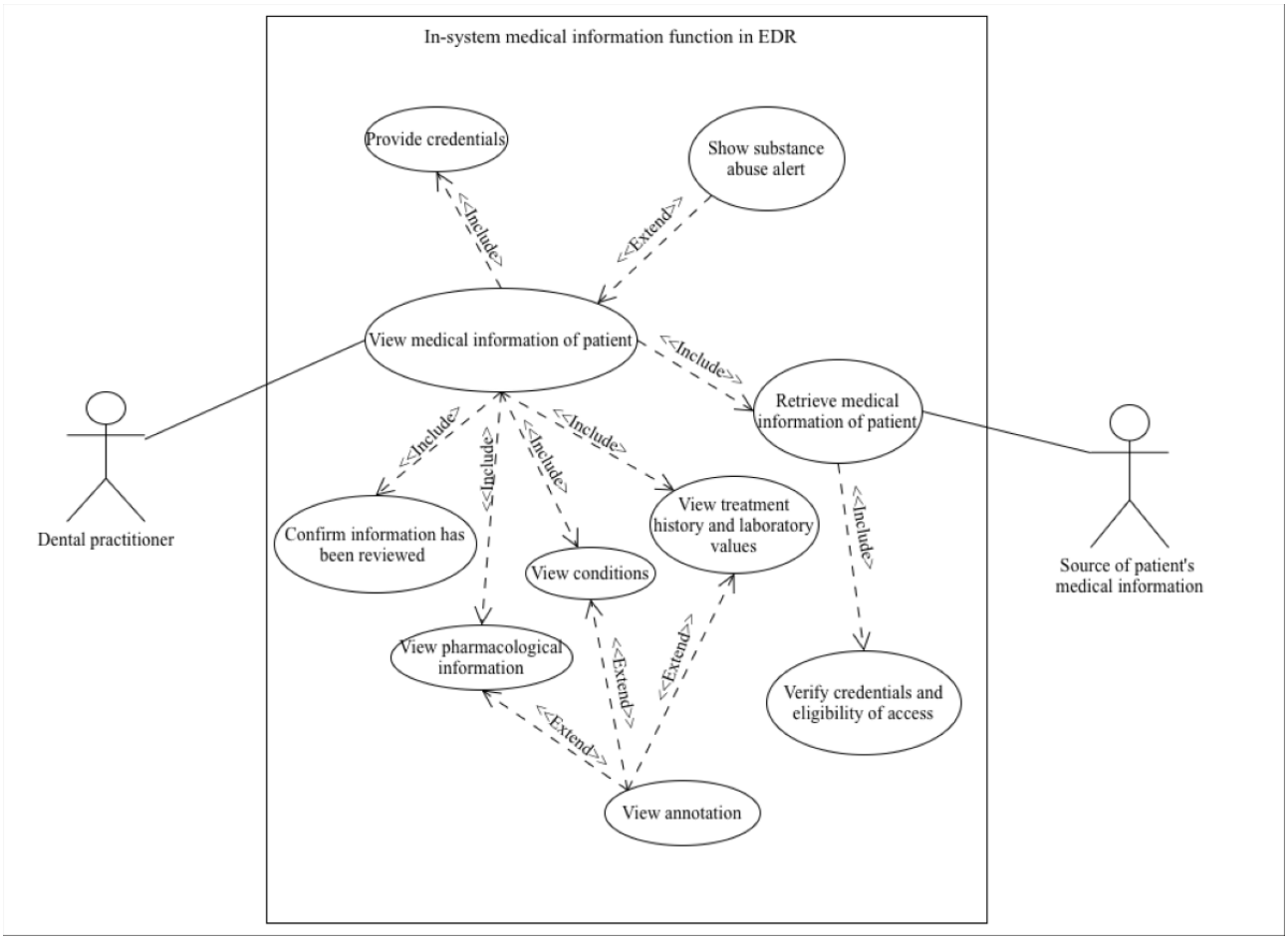
The dental practitioner begins by accessing and checking the medical information of the patient in their EDR system, with the receptionist not having to remind the patient to bring their medication list. Furthermore, there is no reliance on an accompanied proxy, proxy, and medical care unit to obtain medical information. Only if further clarification or consultancy is needed does the process escalate to contacting the medical care unit.

Use Case Related to the Potential Workflow

A UML use case diagram (Figure 1) was developed to represent conceptual expectations and functional needs described by participants regarding an EDR system with access to medical information. This model also assumes that medical information is accessible via a regional database or NPÖ through NTjP, referred to as the source of the patient's medical information in the diagram. It is important to emphasize that the use case diagram does not depict a user interface. Rather, it consolidates qualitative insights into conceptual recommendations. System-level specifications and exceptions are therefore not included.

XSL•FO
RenderX

Figure 1. A use case with an electronic dental record system that involves providing access to medical information. EDR: electronic dental record.



Below follows the use case, with its corresponding process flow detailed in Table 8**:**

1. Use case name to view the medical information of a patient.
2. Description to describe the steps required for a dental practitioner to securely retrieve and review a patient's medical information via an in-system function in the EDR system.
3. Alternative flow to ensure controlled data access, step 5 includes an alternative flow where, if the forwarded credentials are invalid or belong to a user without authorization, the data request is not processed.
4. Preconditions: the patient has given consent for a dental practitioner to access and view information, and the dental practitioner is directly involved in their care.
5. Postconditions: the patient's medical information has been securely accessed and reviewed by the dental practitioner. The system has logged the confirmation of the reviewed information.

XSL•FO
RenderX

Table . Normal flow of events for accessing medical information.

Steps	Actors	Actions
1	Dental practitioner	Initiates action to view the patient's medical information using an in-built function within the EDR[a] system
2	In-system function	Prompts user authentication to verify access rights
3	Dental practitioner	Provides credentials
4	In-system function	Sends a data access request to the external system, including credentials
5	Source of patient's medical information	Validates the access request and verifies that the credentials are valid and authorized to access medical information
6	Source of patient's medical information	Retrieves and returns relevant medical data
7	In-system function	Displays relevant data in the conditions, pharmacological information, and treatment history and laboratory values sections
8	Dental practitioner	Reviews the provided information
9	Dental practitioner	Confirms that the information has been reviewed
10	In-system function	Logs the confirmation
11	Not applicable	End of use case

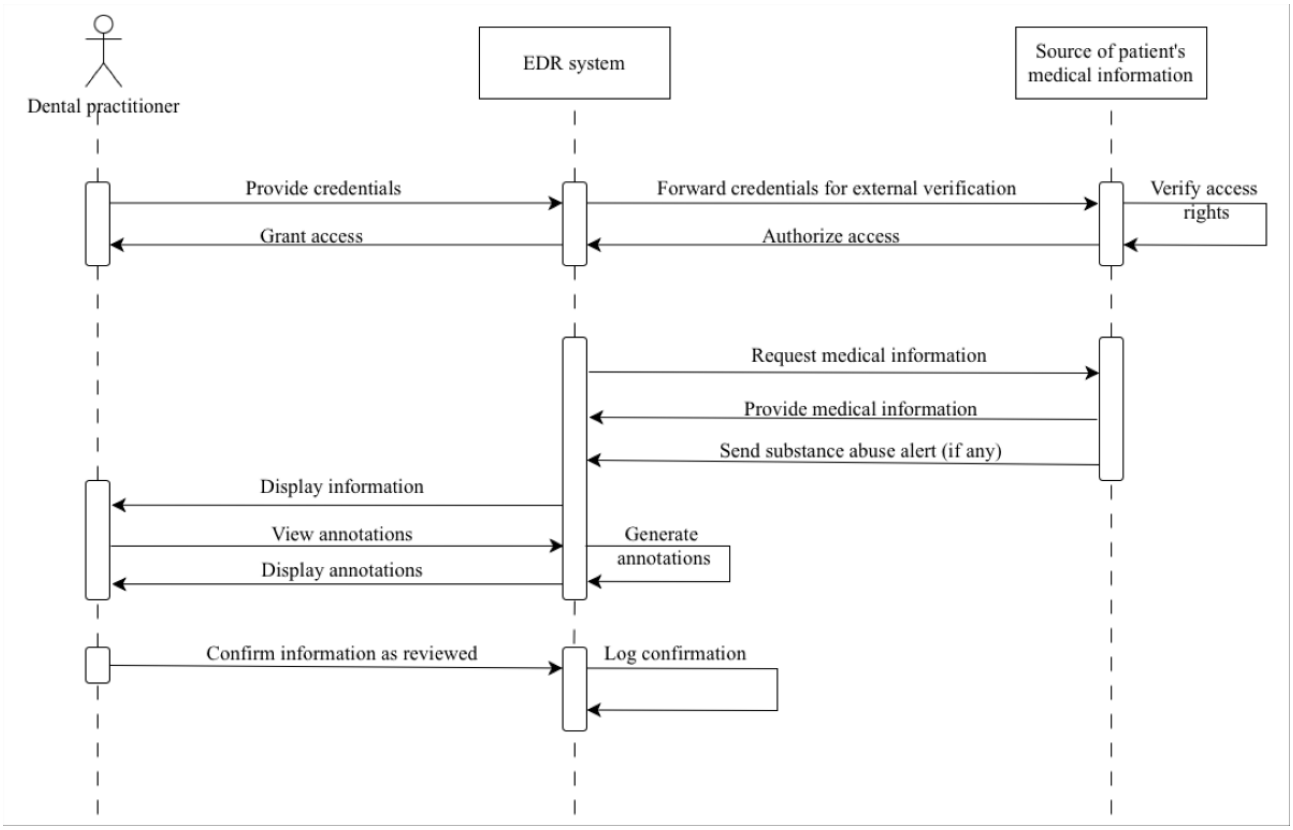
[a]EDR: electronic dental record.

Sequential Flow of Accessing Medical Information in Potential Workflow

As shown in Figure 2, the UML sequence diagram depicts the interaction between the dental practitioner, the EDR system, and the source of patient's medical information. After the dental practitioner provides credentials to the EDR system, a verification request is sent to the external source. Upon successful authorization, the EDR system sends a request and receives the patient's medical information as a message. This may include substance abuse alerts, if applicable. The EDR system then displays the retrieved data to the dental practitioner, who may also choose to view system-generated annotations. Finally, the dental practitioner confirms that the information has been reviewed, which is logged in the system.

XSL•FO
RenderX

Figure 2. The sequence of processes in facilitating medical information. EDR: electronic dental record.



Discussion

Principal Results

Identified Medical Information Needs

Overall, the interviews consistently highlighted the need for accessible medical information to support safe, timely, and informed dental care provision. The findings predominantly aligned with what is reported in prior literature [3-5]. However, this study provided additional granularity regarding the specific medical information needs and their clinical relevance. Participants emphasized 4 main categories: conditions, pharmacological information, treatment history, and laboratory values.

Medical conditions such as cardiovascular disease, cancer, and psychiatric disorders were repeatedly cited due to their implications for procedural risk and treatment planning. Pharmacological needs centered on antiresorptive agents and anticoagulants, given their link to complications such as bleeding and osteonecrosis. Substance abuse history was also flagged as critical, particularly in the context of narcotic misuse. Allergies were a concern, especially penicillin allergy, due to its prescribing relevance. Moreover, relevant medical treatment history and laboratory values were also considered essential for planning and safety.

Identified Workflow Challenges

At GDPs in Dalarna, the current workflows are hindered by several challenges. These include a reliance on patients or third parties, which often leads to incomplete information gathering. Additionally, documentation workflows are fragmented, characterized by frequent interruptions and manual redundancy. The use of EDR for documenting medical information is inconsistent, and there is limited access to supportive information systems. These challenges represent a sociotechnical gap, where systems exist to store medical information but are not designed to meet the information needs or workflow constraints of the participants. In other words, the mere availability of medical information in EDR systems is insufficient, as usability, reliability, and contextual alignment were regarded as critical elements.

Workflow Barriers and Limitations

The findings underscore a misalignment between the current system design, organizational priorities, and clinical expectancy. Participants described relying heavily on patient self-reports, proxy input, or contacting external medical care units to obtain information. Patient self-reports were often unreliable, especially with elderly patients. Proxy or medical consultations introduced additional delays and communication gaps.

Despite the existence of routines, widespread inconsistencies in how medical histories are recorded were reported. Participants believed that having direct access to relevant medical information via the EDR would enhance treatment planning and reduce delays. However, they felt unsupported by the organization**,** citing failed efforts such as the online prescription check platform “Förskrivningskollen.”

The EDR system itself relies heavily on manual input, causing duplicated entries between the medical history section and progress notes. Manual data entry was perceived as burdensome, error-prone, and inefficient. Inconsistent documentation practices also impaired information reliability and visibility,

XSL•FO
RenderX

requiring participants to spend additional time navigating and interpreting records.

There was also uncertainty about legal permissions for accessing medical records, despite the existence of the Integrated Health and Social Care Documentation Act (2022:913) and explicit clarifications from Swedish authorities [21,52]. This points to organizational and educational shortcomings rather than legal barriers. The forthcoming European health data space regulation may help address such ambiguity at the European Union level [53].

Proposed Improvements for Information Access

Enabling direct access to medical information via the EDR could streamline workflows by reducing reliance on manual data retrieval. This would improve time management, decision-making, and care coordination. A feasible path forward involves leveraging Sweden's national interoperability infrastructure. By connecting EDR systems to NPÖ through NTjP, authorized dental professionals could retrieve relevant medical data across institutional boundaries [22-24].

Addressing Interoperability Levels

To implement this vision, interoperability must be achieved at four levels [11]: (1) technical, which must be supported through NTjP; (2) syntactic, which requires EDR systems to align with NTjP service contracts; (3) semantic, to provide a view-only use of external medical information, due to current dental records being largely unstructured; and (4) organizational, which requires commitment from vendors and stakeholders to align with interoperability frameworks and recognize dental professionals' medical information access needs.

Toward a User-Centered Design Approach

Interface inefficiencies and inconsistent documentation stem not only from system limitations but also from design misalignment. Poorly designed systems, even with access to the right data, can obstruct rather than support care [3,13]. Given the high cognitive load and time pressure in dental care [54], systems should minimize user burden. Preferences expressed by participants underscore the need for user-centered EDR design regarding medical information.

Implications for Scalability and Policy Considerations

In highly urbanized or well-resourced Swedish regions, the proposed integration model would require significant adaptations to ensure scalability. These environments typically involve higher patient volumes, greater clinical complexity, and a more diverse ecosystem of EHR and EDR systems. To operate effectively under such conditions, the model must support real-time data synchronization, load balancing, and robust multisystem interoperability through standardized APIs, for example, applying the Health Level Seven Fast Healthcare Interoperability Resources standard and a modular system architecture. Larger regions may also necessitate granular, role-based access controls and comprehensive audit trails to uphold security and regulatory compliance. Scalable infrastructure, potentially leveraging cloud-based or hybrid deployment models, will be essential to accommodate periods with high system workload.

Beyond technical enhancements, successful large-scale implementation will depend on policy and infrastructural measures. These include enforcing national interoperability and terminology standards (eg, Fast Healthcare Interoperability Resources and Systematized Nomenclature of Medicine Clinical Terms), updating the Swedish Patient Data Act to explicitly enable cross-sector data sharing between dental and medical care while maintaining General Data Protection Regulation compliance, and allocating targeted funding to modernize legacy dental systems and support workforce training. Together, these considerations enhance the workflow's scalability and support its potential application in settings that differ from the region studied.

Challenges of Integrating Artificial Intelligence–Driven Dental Imaging and Analysis Into EDR

Beyond the information needs described in this study, the rapidly expanding use of specialized dental software, such as artificial intelligence (AI)-driven digital smile design systems [55] and video-based smile analysis tools like Dynasmile [56], introduces additional challenges for integrating diverse data types into EDRs. These applications generate complex non-textual outputs, including high-resolution images, videos, and AI-derived esthetic or diagnostic metrics, which are increasingly used for treatment planning. Recent reviews of AI-based digital smile design indicate promising improvements in patient and clinician satisfaction, as well as esthetic outcomes [57]. Integrating such data into EDRs requires overcoming significant interoperability barriers. For instance, digital imaging data must comply with standards like Digital Imaging and Communications in Medicine to ensure consistent storage, retrieval, and portability across systems [58]. Similarly, AI-derived outputs (eg, landmark coordinates and esthetic scores) need semantic mapping to standardized clinical concepts to ensure meaningful linkage with diagnostic and treatment data. Furthermore, exchanging large multimedia files through NTjP raises practical considerations regarding storage, capacity, bandwidth, consent, and privacy compliance [59].

Acknowledging these challenges underscores the broader implications of our findings and highlights the need for future research on incorporating advanced dental imaging and AI-generated data into EDR workflows without compromising usability, interoperability, or patient confidentiality.

Limitations

Certain limitations need to be declared. First, the sample consisted of 18 dental practitioners from PDSD-operated GDPs in Dalarna. Although the generalizability was limited due to the limited number of participants and the homogeneity of the sample, with most participants having fewer than five years of professional experience and working in similar clinical settings, thematic saturation was achieved in this study. Qualitative studies, besides providing an in-depth understanding of contextualized experience, may also support theoretical generalizability or, more appropriately for qualitative research, transferability when rich contextual information is provided [60]. This study supports the transferability of the findings to other contexts with similar characteristics. However, to strengthen the generalizability of these findings, future studies

XSL•FO
RenderX

should include a broader and more diverse sample. A quantitative or mixed-methods study would also be valuable to validate and generalize the medical information needs identified here [61].

Future research could use quantitative methods to validate and prioritize the information needs identified in this study across a broader population of dental practitioners. Large-scale surveys using structured instruments (eg, Likert-scale items) could assess both the perceived importance and frequency of specific information needs, providing estimates of their prevalence and variability at regional or national levels. Additionally, a Delphi approach incorporating iterative rounds of quantitative scoring could facilitate structured expert consensus and systematic ranking of priorities.

Conclusions

This study identified a clear need to improve access to medical information within EDR systems in GDPs in Dalarna. While medical information needs such as diagnoses, medications, and allergies are widely acknowledged in the literature, this study contributed detailed insights into the types of information most relevant to dental care. Current documentation practices were found to be fragmented, inefficient, and overly reliant on patients and third parties. Usability limitations in the EDR interface further hinder information access and consistency. Collectively, these issues reflect a sociotechnical misalignment between user needs and system capabilities. To address these gaps, the study proposes leveraging Sweden's existing interoperability infrastructure, particularly by integrating EDR systems with NPÖ via NTjP. However, achieving this requires not only technical alignment but also organizational commitment, regulatory clarity, and a user-centered system design.

Future research should validate these findings at scale and explore system design and policy solutions to ensure that medical information can be securely, efficiently, and meaningfully integrated into information systems used in dental care.

Acknowledgments

The authors want to thank the Public Dental Service Dalarna (PDSD) for their cooperation, and to all the participants who took the time to contribute to this study. It needs to be declared that the main author is a licensed dental practitioner employed by PDSD. However, employment at PDSD did not affect the research process, as no funding was received. Interactions with PDSD were solely limited to recruitment coordination.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Authors' Contributions

SHR conducted the data collection and carried out the initial analysis, which was subsequently reviewed and refined in collaboration with ND and SB.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[PDF File, 124 KB - humanfactors_v13i1e82691_app1.pdf]

Multimedia Appendix 2
Consent form in Swedish and English.
[PDF File, 116 KB - humanfactors_v13i1e82691_app2.pdf]

Multimedia Appendix 3
Unified modeling language activity diagram of the current workflow in managing medical information.
[PNG File, 666 KB - humanfactors_v13i1e82691_app3.png]

Multimedia Appendix 4
Unified modeling language activity of an improved workflow if medical information were made accessible directly in the electronic dental record.
[PNG File, 229 KB - humanfactors_v13i1e82691_app4.png]

References

XSL•FO
RenderX

1. Schmalz G, Brauer L, Haak R, Ziebolz D. Evaluation of a concept to classify anamnesis-related risk of complications and oral diseases in patients attending the clinical course in dental education. BMC Oral Health 2023 Aug 29;23(1):609. [doi: 10.1186/s12903-023-03343-x] [Medline: 37644402]
2. Harrel SK, Molinari J. Aerosols and splatter in dentistry: a brief review of the literature and infection control implications. J Am Dent Assoc 2004 Apr;135(4):429-437. [doi: 10.14219/jada.archive.2004.0207] [Medline: 15127864]
3. Li S, Felix Gomez GG, Xu H, Rajapuri AS, Dixon BE, Thyvalikakath T. Dentists' information needs and opinions on accessing patient information via health information exchange: survey study. JMIR Form Res 2024 Jan 11;8:e51200. [doi: 10.2196/51200] [Medline: 38206667]
4. Simon L, Obadan-Udoh E, Yansane AI, et al. Improving oral-systemic healthcare through the interoperability of electronic medical and dental records: an exploratory study. Appl Clin Inform 2019 May;10(3):367-376. [doi: 10.1055/s-0039-1688832] [Medline: 31141831]
5. Alanazi A, Alghamdi G, Aldosari B. Informational needs for dental-oriented electronic health records from dentists' perspectives. Healthcare (Basel) 2023 Jan 14;11(2):266. [doi: 10.3390/healthcare11020266] [Medline: 36673634]
6. Lisiecka MZ. Allergic reactions in dental practice: classification of medicines, mechanisms of action, and clinical manifestations. Clin Rev Allergy Immunol 2025 Feb 15;68(1):17. [doi: 10.1007/s12016-025-09032-7] [Medline: 39954171]
7. Adibi S, Li M, Salazar N, et al. Medical and dental electronic health record reporting discrepancies in integrated patient care. JDR Clin Trans Res 2020 Jul;5(3):278-283. [doi: 10.1177/2380084419879387] [Medline: 31560579]
8. Charangowda BK. Dental records: an overview. J Forensic Dent Sci 2010 Jan;2(1):5-10. [doi: 10.4103/0974-2948.71050] [Medline: 21189983]
9. Evans RS. Electronic health records: then, now, and in the future. Yearb Med Inform 2016 May 20;Suppl 1(Suppl 1):S48-S61. [doi: 10.15265/IYS-2016-s006] [Medline: 27199197]
10. Heid DW, Chasteen J, Forrey AW. The electronic oral health record. J Contemp Dent Pract 2002 Feb 15;3(1):43-54. [Medline: 12167912]
11. Lehne M, Sass J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. NPJ Digit Med 2019;2:79. [doi: 10.1038/s41746-019-0158-1] [Medline: 31453374]
12. Holmgren AJ, Adler-Milstein J. Health information exchange in us hospitals: the current landscape and a path to improved information sharing. J Hosp Med 2017 Mar;12(3):193-198. [doi: 10.12788/jhm.2704] [Medline: 28272599]
13. Taylor HL, Apathy NC, Vest JR. Health information exchange use during dental visits. AMIA Annu Symp Proc 2020;2020:1210-1219. [Medline: 33936497]
14. Li S, Rajapuri AS, Felix Gomez GG, Schleyer T, Mendonca EA, Thyvalikakath TP. How do dental clinicians obtain up-to-date patient medical histories? modeling strengths, drawbacks, and proposals for improvements. Front Digit Health 2022;4:847080. [doi: 10.3389/fdgth.2022.847080] [Medline: 35419556]
15. Atchison KA, Weintraub JA, Rozier RG. Bridging the dental-medical divide: case studies integrating oral health care and primary health care. J Am Dent Assoc 2018 Oct;149(10):850-858. [doi: 10.1016/j.adaj.2018.05.030] [Medline: 30057150]
16. About the Swedish healthcare system. Socialstyrelsen. 2019. URL: https://www.socialstyrelsen.se/en/about-us/healthcare-for-visitors-to-sweden/about-the-swedish-healthcare-system/ [accessed 2025-02-01]
17. Uppföljning Vision e-hälsa 2025 – Rapport avseende år 2021 (Report in Swedish). : E-hälsomyndigheten; 2022 URL: https://www.ehalsomyndigheten.se/globalassets/ehm/3_om-oss/rapporter/uppfoljning-vision-e-halsa-2025-rapport-avseende-2021.pdf [accessed 2026-01-27]
18. Follow-Up Vision for eHealth 2025 - Report on the Year 2022. : Swedish eHealth Agency; 2023 URL: https://www.ehalsomyndigheten.se/siteassets/ehm/3_om-ehalsomyndigheten/rapporter/rapporter_regeringsuppdrag/follow-up-vision-for-ehealth-2025-report-on-the-year-2022.pdf [accessed 2026-01-27]
19. Swedish Institute. Swedish government. Power for the people! This is how Sweden is governed. URL: https://sweden.se/life/democracy/swedish-government [accessed 2025-02-12]
20. Nilsson P, Stjernquist A, Janlöv N. Fragmented health and social care in Sweden - a theoretical framework that describes the disparate needs for coordination for different patient and user groups. Int J Integr Care 2016;16(6):348. [doi: 10.5334/ijic.2896]
21. The Government Offices' legal databases [Web page in Swedish]. Regeringskansliet. 2022. URL: http://rkrattsbaser.gov.se/sfst?bet=2022:913 [accessed 2025-02-14]
22. Sammanhållen journalföring – Möjligheter till digital informationsförsörjning på hälsodataområdet. : E-hälsomyndigheten; 2022 URL: https://www.ehalsomyndigheten.se/globalassets/ehm/3_om-oss/rapporter/slutrapportering-av-uppdraget-att-foresla-hur-sammanhallen-journalforing-kan-nyttjas-i-storre-utstrackning.pdf [accessed 2025-02-14]
23. Nationell patientöversikt – NPö [Web page in Swedish]. Inera. URL: https://www.inera.se/tjanster/alla-tjanster-a-o/npo---nationell-patientoversikt/ [accessed 2025-02-14]
24. Nationella tjänsteplattformen. Inera. URL: https://www.inera.se/tjanster/arkitektur-och-digital-infrastruktur/nationella-tjansteplattformen/ [accessed 2025-02-14]
25. Långt kvar till gemensamt IT-system [Web page in Swedish]. Tandläkartidningen. 2017. URL: https://www.tandlakartidningen.se/nyhet/langt-kvar-till-gemensamt-it-system/ [accessed 2025-03-22]

XSL•FO
RenderX

26. Atchison KA, Weintraub JA. Integrating oral health and primary care in the changing health care landscape. N C Med J 2017;78(6):406-409. [doi: 10.18043/ncm.78.6.406] [Medline: 29203606]
27. MacNeil RLM, Hilario H, Ryan MM, Glurich I, Nycz GR, Acharya A. The case for integrated oral and primary medical health care delivery: marshfield clinic health system. J Dent Educ 2020 Aug;84(8):924-931. [doi: 10.1002/jdd.12289] [Medline: 32638376]
28. Shimpi N, Glurich I, Panny A, Acharya A. Knowledgeability, attitude, and practice behaviors of primary care providers toward managing patients' oral health care in medical practice: Wisconsin statewide survey. J Am Dent Assoc 2019 Oct;150(10):863-872. [doi: 10.1016/j.adaj.2019.05.020] [Medline: 31446976]
29. Harnagea H, Couturier Y, Shrivastava R, et al. Barriers and facilitators in the integration of oral health into primary care: a scoping review. BMJ Open 2017 Sep 25;7(9):e016078. [doi: 10.1136/bmjopen-2017-016078] [Medline: 28951405]
30. Unika siffror: här är tandläkarbristen värst. Tandläkartidningen. 2023. URL: https://www.tandlakartidningen.se/nyhet/unika-siffror-har-ar-tandlakarbristen-varst/ [accessed 2025-02-22]
31. Dalarna. Tillväxtverket. URL: https://tillvaxtverket.se/tillvaxtverket/statistikochanalys/statistikomregionalutveckling/lansuppdeladstatistik/dalarna.1593.html?chartCollection=6#svid12_843740f184cc4e52fa14d4 [accessed 2025-02-22]
32. Janto M, Iurcov R, Daina CM, et al. Oral health among elderly, impact on life quality, access of elderly patients to oral health services and methods to improve oral health: a narrative review. J Pers Med 2022 Feb 28;12(3):372. [doi: 10.3390/jpm12030372] [Medline: 35330372]
33. Alami H, Lehoux P, Gagnon MP, Fortin JP, Fleet R, Ag Ahmed MA. Rethinking the electronic health record through the quadruple aim: time to align its value with the health system. BMC Med Inform Decis Mak 2020 Feb 17;20(1):32. [doi: 10.1186/s12911-020-1048-9] [Medline: 32066432]
34. Baxter G, Sommerville I. Socio-technical systems: from design methods to systems engineering. Interact Comput 2011 Jan;23(1):4-17. [doi: 10.1016/j.intcom.2010.07.003]
35. Hunter DJ, McCallum J, Howes D. Defining exploratory-descriptive qualitative (EDQ) research and considering its application to healthcare. Proceedings of the 6th Annual Worldwide Nursing Conference 2018 2018;4(1) [FREE Full text]
36. Tandvård genom hela livet [Web page in Swedish]. Region Dalarna. URL: https://www.regiondalarna.se/verksamhet/tandvard/ [accessed 2025-03-22]
37. Ordell S, Unell L, Söderfeldt B. An analysis of present dental professions in Sweden. Swed Dent J 2006;30(4):155-164. [Medline: 17243443]
38. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015 Sep;42(5):533-544. [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
39. DeJonckheere M, Vaughn LM. Semistructured interviewing in primary care research: a balance of relationship and rigour. Fam Med Community Health 2019;7(2):e000057. [doi: 10.1136/fmch-2018-000057] [Medline: 32148704]
40. Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Qual Saf Health Care 2010 Oct;19 Suppl 3(Suppl 3):i68-i74. [doi: 10.1136/qshc.2010.042085] [Medline: 20959322]
41. Saunders B, Sim J, Kingstone T, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907. [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
42. Vaismoradi M, Turunen H, Bondas T. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. Nurs Health Sci 2013 Sep;15(3):398-405. [doi: 10.1111/nhs.12048] [Medline: 23480423]
43. Neale J. Iterative categorization (IC): a systematic technique for analysing qualitative data. Addiction 2016 Jun;111(6):1096-1106. [doi: 10.1111/add.13314] [Medline: 26806155]
44. Ciccozzi F, Malavolta I, Selic B. Execution of UML models: a systematic review of research and practice. Softw Syst Model 2019 Jun;18(3):2313-2360. [doi: 10.1007/s10270-018-0675-4]
45. Aggarwal V. The application of the unified modeling language in object-oriented analysis of healthcare information systems. J Med Syst 2002 Oct;26(5):383-397. [doi: 10.1023/a:1016449031753] [Medline: 12182204]
46. Lag om etikprövning av forskning som avser människor [Web page in Swedish]. SVERIGES RIKSDAG. URL: https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/lag-2003460-om-etikprovning-av-forskning-som_sfs-2003-460/ [accessed 2025-08-05]
47. What the Act says [Web page in Swedish]. Etikprövningsmyndigheten. URL: https://etikprovningsmyndigheten.se/en/what-the-act-says/ [accessed 2025-02-14]
48. European Parliament and Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Off J Eur Union 2016;L119:1-88 [FREE Full text]
49. Association WM. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
50. Om oss [Web page in Swedish]. Carestream dental. URL: https://www.t4academy.se/pages/om-oss [accessed 2025-04-05]
51. Förskrivningskollen [Web page in Swedish]. E-hälsomyndigheten. URL: https://www.ehalsomyndigheten.se/yrkesverksam/vard/forskrivningskollen/ [accessed 2025-03-20]

XSL•FO
RenderX

52. Kartläggning av hinder för samverkan mellan tandvård och hälso- och sjukvård – slutrapport 2019 [Web page in Swedish]. Socialstyrelsen. 2019. URL: https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/ovrigt/2019-10-6428.pdf [accessed 2025-02-14]
53. European Health Data Space Regulation (EHDS). European Commission. URL: https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space-regulation-ehds_en [accessed 2025-05-13]
54. Pontes CC, Stanley K, Molayem S. Understanding the dental profession's stress burden: prevalence and implications. Compend Contin Educ Dent 2024 May;45(5):236-241. [Medline: 38900461]
55. Khaled Ali GA, Abdelbadei KA, Essam M, Ali GAK, et al. Smile.AI: a deep learning system for digital smile design. : IEEE Presented at: 2023 Eleventh International Conference on Intelligent Computing and Information Systems (ICICIS); Nov 21-23, 2023; Cairo, Egypt p. 568-573. [doi: 10.1109/ICICIS58388.2023.10391194]
56. Chen K, Qiu L, Xie X, Bai Y. Dynasmile: video-based smile analysis software in orthodontics. SoftwareX 2025 Feb;29:102004. [doi: 10.1016/j.softx.2024.102004] [Medline: 40438663]
57. Saini RS, Kaur K, Gurumurthy V, et al. Impact of artificial intelligence-based digital smile design on patient and clinician satisfaction and facial esthetic outcomes: a systematic review and meta-analysis. Digit Health 2025;11. [doi: 10.1177/20552076251388392] [Medline: 41122427]
58. Farman AG. Use and implication of the DICOM standard in dentistry. Dent Clin North Am 2002 Jul;46(3):565-573. [doi: 10.1016/s0011-8532(02)00003-4] [Medline: 12222098]
59. Liu TY, Lee KH, Mukundan A, Karmakar R, Dhiman H, Wang HC. AI in dentistry: innovations, ethical considerations, and integration barriers. Bioengineering (Basel) 2025 Aug 29;12(9):41007172. [doi: 10.3390/bioengineering12090928] [Medline: 41007172]
60. Carminati L. Generalizability in qualitative research: a tale of two traditions. Qual Health Res 2018 Nov;28(13):2094-2101. [doi: 10.1177/1049732318788379] [Medline: 30043686]
61. Polit DF, Beck CT. Generalization in quantitative and qualitative research: myths and strategies. Int J Nurs Stud 2010 Nov;47(11):1451-1458. [doi: 10.1016/j.ijnurstu.2010.06.004] [Medline: 20598692]

Abbreviations

AI: artificial intelligence
EDR: electronic dental record
EHR: electronic health record
GDP: general dental practice
HIE: health information exchange
IEHR: integrated EHR-EDR
NPÖ: nationell patientöversikt
NTjP: nationell tjänsteplattform
PDSD: Public Dental Service Dalarna
UML: unified modeling language

Edited by A Kushniruk; submitted 20.Aug.2025; peer-reviewed by K Chen, S Li; revised version received 05.Dec.2025; accepted 14.Dec.2025; published 29.Jan.2026.

Please cite as:
Rahim SH, Davoody N, Bonacina S
Exploring Medical Information Needs and Accessibility in Swedish Dental Care by Analysis of Documentation Workflows and Electronic Dental Records in Dalarna: Sociotechnical Qualitative Study
JMIR Hum Factors 2026;13:e82691
URL: https://humanfactors.jmir.org/2026/1/e82691
doi:10.2196/82691

© Sahid Hasan Rahim, Nadia Davoody, Stefano Bonacina. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 29.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Factors Influencing User Satisfaction in Accessing Health Data: Cross-Sectional Survey of United Kingdom Adults

Maria Xenou[1], BSc, MSc; Omar Al-Ubaydli[2], BSc, MSc, PhD

[1]UCL Institute of Health Informatics, University College London, 222 Euston Rd, London, London, United Kingdom

[2]Department of Economics and Mercatus Center, George Mason University, 4400 University Drive, Fairfax, VA, United States

Corresponding Author:
Maria Xenou, BSc, MSc
UCL Institute of Health Informatics, University College London, 222 Euston Rd, London, London, United Kingdom

Abstract

Background: While patient health records (PHRs) are becoming ubiquitous, nationwide evidence on the drivers of user satisfaction in the United Kingdom remains scarce.

Objective: This study aimed to quantify determinants of PHR user satisfaction in a nationally representative sample of United Kingdom adults and to contrast real-world experiences with hypothetical expectations among individuals without PHR exposure.

Methods: We conducted a cross-sectional online survey in March 2022 using stratified quota sampling (eg, age, sex, ethnicity) through the Prolific platform. Of 1001 respondents, 533 (53%) were female (mean age 41, SD13) years, 468 (47%) reported previous PHR use (“experienced” cohort), and 533 (53%) respondents did not (“hypothetical” cohort). Primary outcomes were five satisfaction items (5-point Likert), overall PHR evaluation (0 - 100), and stated/anticipated PHR functions. Two-sample *t*-tests with unequal variances examined between-group differences; multivariate analysis of variance (MANOVA) assessed demographic equivalence. Significance was set at $P<.05$.

Results: Respondents with PHR experience rated their record easy-to-use in 79% (370/468) of cases versus an anticipated 93% (498/533) among nonusers ($P<.001$). Experienced users rated appointment-scheduling time-saving at 68% (319/468) compared with 72% (387/533) of nonusers ($P=.17$). Mean overall evaluation scores were 71.0 (SD 19.5) versus 74.2 (SD 18.6) ($t=2.25$, $P=.02$). MANOVA confirmed demographic balance (Wilks $\Lambda=.99$, $F_{3}=1.48$, $P=.18$). Preferred functions across cohorts were viewing personal health information (experience 81%, hypothetical 90%) and lab results (50% vs 76%). Users lacking experience consistently over-estimated future use of carer-related functions (eg, children’s data: 36% vs 12%, $P<.001$).

Conclusions: Adults in the United Kingdom value PHRs primarily for personal data access and scheduling convenience. Nonusers displayed optimistic expectations about carer-related features and breadth of functionality, indicating an information gap. Enhancing transparency about current capabilities and expanding features that facilitate caregiving could increase satisfaction and adoption.

(JMIR Hum Factors 2026;13:e75935) doi:10.2196/75935

KEYWORDS

patient health record; user satisfaction; national survey; digital health; United Kingdom

Introduction

Background

Electronic access to health information is integral to patient-centered care [1]. In the United Kingdom, 73% of adults go online first for health information [2], and the National Health Service (NHS) aims to universalize digital data access under the ethos of “no decision about me without me” [3]. People may view clinical data through provider-tethered electronic medical records (EMRs) or untethered patient health records (PHRs) [4-6]. PHRs allow patients to *access, manage, and share* their complete longitudinal health data independently of any single institution [6]. Evidence suggests PHRs can reduce costs and improve outcomes by empowering self-care [4,7,8], yet utilization and satisfaction remain variable [9].

The NHS app, launched nationally in 2019, now exceeds 22 million registered users and represents most United Kingdom adults’ first exposure to a PHR ecosystem [10,11]. Previous research shows that activated patients can achieve higher levels of self-care and satisfaction by sharing all clinical notes with their clinical team, reducing their clinical workload, and improving the health information exchange [10,11]. A recent Canadian study surveyed national PHR users and highlighted time savings, reduced clinic visits, and high satisfaction [12]. They concluded that PHR users showed positive patient satisfaction, a measured decrease of appointment rates, and a decrease in their usage of the Canadian health system.

XSL•FO
RenderX

Comparable United Kingdom-wide data—including insights from individuals without PHR experience—are lacking. Understanding expectations and perceived utility among both groups is critical as PHR functions expand.

In the United Kingdom, digital access to personal health data has expanded rapidly. Approximately 73% of adults use the internet as their main health information source [13-15]. Despite the growing adoption of such tools, research on PHR user satisfaction in the United Kingdom remains limited and often focuses on small, nonrepresentative samples [16].

Objectives

We sought to (1) identify determinants of PHR user satisfaction in a nationally representative sample of United Kingdom adults [17] and (2) compare perceptions between adults with real-world PHR experience and those expressing hypothetical expectations.

Methods

Design and Setting

We explicitly identify the study as a *descriptive, cross-sectional survey*. Data were collected online between 23 - 25 March 2022. Informed electronic consent was obtained.

Participants and Sampling

Because the survey assessed attitudes toward PHRs, it was essential to provide respondents with a clear understanding of what a PHR is. Accordingly, participants were shown a lay definition adapted from the Canadian study [12]. Approximately half of respondents had real experience with a PHR—most commonly the NHS app, which enables access to medical records, vaccination status, appointments, and prescriptions [18-20]. Other PHR systems integrated with the NHS App were also represented among users.

Participants were recruited via Prolific Academic [21], which maintains a United Kingdom participant panel. Using Prolific's *nationally representative* option, we enforced quota stratification on age (six census bands), sex (male/female), and ethnicity (five ONS categories) to mirror 2021 United Kingdom Census proportions. Inclusion criteria were age ≥ 18 years and United Kingdom residence. A total of 1001 adults satisfied the quotas, resulting in margin-of-error±3 percentage-points at the 95% confidence level.

In addition to age, sex, and ethnicity, no further stratification variables (such as education level or household income) were available within the nationally representative Prolific panel profiles at the time of recruitment. Quotas were therefore restricted to Office for National Statistics census categories for age, sex, and ethnicity. The resulting sample distributions across these strata are reported in Multimedia Appendix 1.

Survey Instrument

The questionnaire replicated the Canadian instrument [12] with United Kingdom contextual modifications (eg, "accident and emergency" replacing "emergency room"). Two parallel versions were administered: an experience-based version for respondents confirming prior PHR use and a hypothetical version with isomorphic wording framed as "If you had a PHR…." Sections covered (1) user satisfaction (5 items), (2) utility (12 items), (3) impact, and (4) demographics. The full instrument is now provided as a Survey Instrument Multimedia Appendix 2 .

Statistical Analysis

Categorical variables (eg, gender, age group) were converted to dummy indicators. No missing data were observed. Between-group differences in satisfaction, utility, and impact outcomes were examined using independent two-sample t-tests with unequal variances (Multimedia Appendix 1). A multivariate analysis of variance (MANOVA) was conducted solely to assess demographic equivalence between respondents with and without prior PHR experience, using gender and age-band indicators as joint dependent variables. MANOVA was not used for outcome analysis.

All statistical tests used two-tailed p-values, with $P<.05$ considered significant. MANOVA assessed demographic equivalence across seven covariates (gender + six age bands). All analyses used STATA (version 17.0; StataCorp LLC), with $\alpha=.05$.

Ethical Considerations

The study is classified as a service evaluation per NHS REC guidance and is therefore exempt from formal NHS ethics review [22].

Results

Participant Characteristics

Data collection occurred between March 25, and April 5, 2022. A total of 1001 adults completed the survey (completion rate=99%). Table 1 compares sample demographics with 2021 Census benchmarks; deviations were <5 percentage points for all strata except adults aged ≥ 60 years (60 - 69: −6 pp;≥70: −8 pp).

XSL•FO
RenderX

Table . Participant demographics by sample [experienced-based versus hypothetical].

Characteristics	Experience-based sample (n=468), n (%)	Hypothetical sample (n=533), n (%)	Nationally representative sample (2021), %
Gender			
Female	247 (53)	255 (48)	51
Male	217 (46)	263 (49)	49
Other	4 (1)	15 (3)	—[a]
Age			
18 - 29	107 (23)	93 (17)	18
30 - 39	93 (20)	100 (19)	17
40 - 49	82 (18)	86 (16)	16
50 - 59	86 (18)	114 (21)	18
60 - 69	85 (18)	116 (22)	14
70+	15 (3)	24 (5)	17
Comfort with computers			
Completely comfortable	11 (2)	15 (3)	—
Comfortable	3 (1)	3 (1)	—
Neutral	9 (2)	22 (4)	—
Uncomfortable	112 (24)	185 (35)	—
Completely uncomfortable	333 (71)	308 (58)	—

[a]Not available.

A MANOVA assessing joint differences in sex and age distribution between respondents with and without prior PHR experience showed no statistically significant multivariate differences (Wilks' Λ=0.99; $F_{,4}$ =1.48; *P*=.18), indicating demographic balance between cohorts.

Satisfaction, Impact and Utility

Among participants with PHR experience, 79% (370/468) agreed their PHR was easy to use and 68% (319/468) reported it saved time scheduling appointments. For those without experience, 99% (528/533) stated they would use a PHR to access personal health information and 72% (387/533) expected time savings. Ease of communication with care providers was reported by 60% (283/468) of experienced users versus 75% (398/533) of nonusers (*P*=.01).

PHR use reportedly avoided a clinic visit for 42% and an Accident and Emergency visit for 4% of experienced users. Hypothetical avoidance estimates were 32% and 9% among nonusers.

Impact on Healthcare Use

Forty-two per cent (200/468) of experienced users indicated their PHR helped avoid at least one clinic visit, and 4% (19/468) avoided an emergency-room visit. Among nonusers, 32% (193/533) anticipated avoiding a clinic visit and 9% (47/533) an emergency visit.

Preferred Features and Cost Savings

Experienced users most valued: viewing their own health information 81% (377/468); viewing lab results 50% (233/468); and scheduling appointments 47% (220/468). Nonusers prioritized: viewing health information 90% (479/533); lab results 76% (405/533); and appointment scheduling 58% (309/533). Cost savings most often related to time off work (40% (187/468) vs 57% (303/533), petrol/gasoline (30% (140/468) vs 52% (278/533), and parking (21% (98/468) vs 34% (181/533).

Experience-Based vs Hypothetical Comparisons

Table 2 presents mean differences across 35 survey items. Nonusers significantly over-estimated future use for caregiving data access (eg, child records: Δ =+24 pp, *P*<.001). Despite these discrepancies, overall evaluation scores were similar (71 vs 74, *P*=.02). Specifically, three main differences emerged:

1. Nonusers over-estimated the extent to which they would use PHRs for family members' data (eg, children: 36% vs 12%, *P*=.01).
2. They over-estimated available functionality, such as preclinic forms (55% vs 18%, *P*=.01).
3. Despite this, overall satisfaction scores were comparable (74 vs 71 out of 100, *P*=.01).

XSL•FO
RenderX

Table . Participant views on PHRs (experienced-based versus hypothetical).

Participant responses	Experience-based sample (n=468)	Hypothetical sample (n=533)
Requested improvement (scale 0 to 4), mean (SD)		—[a]
Less cumbersome login authentication	2.6 (1.0)	—
More data about my health	3.1 (0.8)	—
Access to lab and diagnostic imaging results	3.1 (0.8)	—
Access to data that would allow avoiding care visit	3.1 (0.8)	—
Question answered via secure message	2.9 (0.9)	—
General convenience and less waiting in clinic	3.2 (0.8)	—
Most preferred features, n (%)	n=468	n=533
Ease of communication with family's care providers	68 (15)	165 (31)
Ease of scheduling appointments	220 (47)	309 (58)
Like to view own lab results	233 (50)	405 (76)
Like to view own health information	377 (81)	479 (90)
Like to view health information of carees	57 (12)	180 (34)
Like to fill out pre-appointment clinical questionnaire	83 (18)	291 (55)
Like to record blood pressure for care providers	31 (7)	151 (28)
Like to record glucose level for care providers	5 (1)	105 (20)
Other	17 (4)	16 (3)
Factors reported in self-reported avoidance of clinic visit, n (%)	n=468	n=533
Imaging results discussed by secure message	85 (18)	—
Question asked by message was answered	117 (25)	—
Other	20 (4)	—
Cost saving by avoiding clinic visit, n (%)	n=468	n= 533
Petrol (gasoline)	140 (30)	278 (52)
Time off work	187 (40)	3063 (57)
Getting childcare	32 (7)	70 (13)
Parking	98 (21)	181 (34)
Taxi	29 (6)	78 (15)
Other	28 (6)	65 (12)

[a]Not available.

Table 3 shows the sample mean for both groups across the different variables, along with the significance level from the aforementioned 2-sample *t*-test. There are three main findings associated with these results.

XSL•FO
RenderX

Table . Sample means by group [experienced-based versus hypothetical] with significance level of 2-sample t-test.

Variable	Hypothetical	Experience-based	Significance
Use PHR[a] to access my health information	99%	96%	1%
Use PHR to access children's health information	36%	12%	1%
Use PHR to access care health information	32%	9%	1%
Use PHR to access someone else's health information	2%	4%	None
PHRs ease communication with family care providers	31%	15%	1%
PHRs ease scheduling appointments for self/family	58%	47%	1%
PHRs enable viewing own lab results	76%	50%	1%
PHRs enable viewing own health information	90%	81%	1%
PHRs enable viewing care health information	34%	12%	1%
PHRs enable filling out pre-appt clinical questionnaire	55%	18%	1%
PHRs enable recording BP for care providers	28%	7%	1%
PHRs enable recording glucose for care providers	20%	1%	1%
PHRs enable something else	3%	4%	None
PHRs are easy to use	76%	41%	1%
PHRs save time when scheduling appointments	87%	76%	10%
PHRs allow for more convenient communication with care providers	89%	61%	1%
PHR helpfulness rating 0 - 100	74	71	5%
PHR benefit depends on volume of health data	1.00	1.01	None
PHR login frequency depends on volume of health data	0.47	0.72	1%
PHR benefit depends on promptness of data	0.96	0.93	None
PHR login frequency depends on promptness of data	0.58	0.71	5%
PHR has enabled self/family avoiding ER visit	9%	4%	1%
PHR has not enabled self/family avoiding clinic visit	64%	57%	5%
PHR allowing clinic visit avoidance saved petrol	52%	30%	1%
PHR allowing clinic visit avoidance saved work leave	57%	40%	1%
PHR allowing clinic visit avoidance saved childcare	13%	7%	1%

XSL•FO
RenderX

Variable	Hypothetical	Experience-based	Significance
PHR allowing clinic visit avoidance saved parking	34%	21%	1%
PHR allowing clinic visit avoidance saved taxi	15%	6%	1%
PHR allowing clinic visit avoidance saved other	12%	6%	1%
PHR access from desktop/laptop	76%	51%	1%
PHR access from smartphone	64%	62%	None
PHR access from tablet	25%	12%	1%
PHR access no preference	6%	1%	1%

[a]PHR: patient health record.

Discussion

Principal Findings

This study examined factors influencing user satisfaction with PHRs among a nationally representative sample of adults in the United Kingdom. Two groups were compared—those with PHR experience and those without—to explore both lived and hypothetical perspectives on digital health record use. Demographically, the groups were comparable, indicating that observed differences in satisfaction and expectations primarily reflect PHR exposure rather than underlying population differences.

Overall, adults in the United Kingdom expressed strong support for access to their personal health information. Experienced users reported that PHRs were generally easy to use, time-saving, and useful for managing appointments, while nonusers anticipated similar benefits if they had access. These findings are consistent with evidence from other contexts demonstrating that PHR adoption enhances satisfaction, engagement, and communication with health care providers [4,6,12,13]. The results also mirror those from Canada [12], where PHR access was associated with improved efficiency and reduced appointment rates. However, users in the United Kingdom rated usability slightly lower than Canadian respondents, which may reflect differences in the maturity of the platforms, user training, and integration across healthcare systems [2,5,20].

Comparison With Prior Work

Our results align with prior literature indicating that PHR usability, data accessibility, and perceived time savings are central determinants of satisfaction [3,4,6,17]. Participants with PHR experience valued the ability to view their own health and laboratory information, schedule appointments, and communicate with care providers—features frequently cited as key enablers of digital engagement [6,9,16]. These findings correspond with previous systematic reviews highlighting that patient portals improve convenience and empower users to manage their own health [4,5,7]. In contrast, our respondents rated United Kingdom systems slightly less favorably than those in comparable studies from Canada or the United States, suggesting ongoing challenges related to user interface design and interoperability within the NHS infrastructure [2,15,20].

Participants without prior experience tended to overestimate potential benefits and functionalities, a trend documented in earlier research on digital health optimism and hypothetical bias [10,15,16]. Woods et al [10] and Cunningham et al [16] found similar patterns of expectation inflation among new users of health portals, often due to limited understanding of actual system capabilities. Such insights underscore the importance of managing user expectations during onboarding and designing transparent communication strategies to clarify what PHRs can—and cannot—do.

In line with other studies, both experienced and hypothetical users viewed PHRs as tools that could reduce unnecessary clinical visits and associated costs such as travel, parking, and time off work [12,13,17]. This perception supports previous evidence that digital access to records can improve service efficiency and convenience for patients [4,7,9]. However, real-world data often indicate more modest effects, emphasizing the need for longitudinal studies assessing actual behavioral and economic outcomes [4,5].

Those with PHR experience expressed a desire for expanded functionality, greater interoperability, and simplified authentication—issues also raised by Turner et al [17] and Schneider et al [8], who highlighted the importance of usability and integration with existing clinical workflows. Improved personalisation, data richness, and accessibility were repeatedly mentioned, echoing prior calls for more inclusive and user-centered portal design [3,6,9,23].

From a policy perspective, the comparison between users and nonusers contributes new insights to the literature. While most previous studies focus solely on active users, our inclusion of a nationally representative sample allows for broader generalization. This approach complements earlier research that used local or condition-specific samples [8,16,17]. As digital transformation accelerates across the NHS, understanding perceptions among both users and nonusers is crucial for equitable and scalable implementation [15,18,20].

Finally, our findings align with recent studies demonstrating that both system and patient factors strongly predict portal use and satisfaction. Agrawal et al [24] reported that usability,

XSL•FO
RenderX

system reliability, and patient demographics influence electronic medical record engagement. Ndabu et al [23] observed that access type and age group significantly affect portal breadth of use, highlighting digital literacy and inclusivity as enduring challenges. Moreover, Whittemore et al [25] showed that targeted interventions can increase patient portal use among adults with type 2 diabetes, particularly when support is localized and tailored to patient needs. Collectively, these studies reinforce our conclusion that user experience, accessibility, and system design remain the primary drivers of satisfaction and long-term adoption.

Strengths and Limitations

Strengths include national representativeness and parallel assessment of users and nonusers. This is among the first nationally representative analyses of PHR satisfaction in the United Kingdom. It uniquely contrasts real and hypothetical users, offering insights into both actual experiences and potential barriers to uptake.

In terms of the limitations, first, the study's cross-sectional design limits causal inference. Second, older age groups (60 - 69 and ≥ 70 y) were under-represented in the sample, which may affect generalizability, particularly regarding digital literacy and accessibility. Third, hypothetical responses may reflect "hypothetical bias," where participants without prior experience imagine more positive outcomes. Fourth, participation was voluntary, raising the possibility of self-selection bias toward more digitally engaged individuals. Finally, all measures were self-reported, and experience was not verified independently. Future research could employ longitudinal designs and link survey data with usage metrics to validate self-reports.

Second, education level and household income were not available as stratification or analytic variables within the nationally representative panel profiles used for recruitment. As socioeconomic status may influence digital access, health literacy, and engagement with patient-facing technologies, future studies should incorporate education and income measures where feasible to further refine equity-related insights.

Conclusions

United Kingdom adults—both users and nonusers—demonstrate a strong interest in accessing their health data electronically. Experienced users report time and cost savings, while nonusers anticipate similar benefits. However, nonusers' expectations often exceed current system capabilities, highlighting opportunities for continued system improvement and patient education. Ongoing evaluation of the NHS App and similar platforms will be crucial to achieving equitable, efficient, and user-centered digital health access in the United Kingdom.

Acknowledgments

We thank all survey participants. No external funding was received.

Funding

The authors declared no financial support was received for this work.

Conflicts of Interest

OA is a board member and shareholder of Patients Know Best; direct family members are also shareholders.

Multimedia Appendix 1
Complete between-group comparison tables (*t* tests).
[DOCX File, 38 KB - humanfactors_v13i1e75935_app1.docx]

Multimedia Appendix 2
Survey instrument.
[DOCX File, 32 KB - humanfactors_v13i1e75935_app2.docx]

References

1. Canada Health Infoway. Patient access to health information. URL: https://insights.infoway-inforoute.ca/2022-patient-access-to-health-information/ [accessed 2022-02-01]
2. Rigby M, Georgiou A, Hyppönen H, et al. Patient portals as a means of information and communication technology support to patient- centric care coordination - the missing evidence and the challenges of evaluation. A joint contribution of IMIA WG EVAL and EFMI WG EVAL. Yearb Med Inform 2015 Aug 13;10(1):148-159. [doi: 10.15265/IY-2015-007] [Medline: 26123909]
3. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. What do we know about developing patient portals? A systematic literature review. J Am Med Inform Assoc 2016 Apr;23(e1):e162-e168. [doi: 10.1093/jamia/ocv114] [Medline: 26335985]
4. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012 Nov 26;14(6):e162. [doi: 10.2196/jmir.2238] [Medline: 23183044]

XSL•FO
RenderX

5. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. How outcomes are achieved through patient portals: a realist review. J Am Med Inform Assoc 2014;21(4):751-757. [doi: 10.1136/amiajnl-2013-002501] [Medline: 24503882]
6. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148. [doi: 10.2196/jmir.4255] [Medline: 26104044]
7. Risling T, Martinez J, Young J, Thorp-Froslie N. Evaluating patient empowerment in association with eHealth technology: scoping review. J Med Internet Res 2017 Sep 29;19(9):e329. [doi: 10.2196/jmir.7809] [Medline: 28963090]
8. Schneider H, Hill S, Blandford A. Patients know best: qualitative study on how families use patient-controlled personal health records. J Med Internet Res 2016 Feb 24;18(2):e43. [doi: 10.2196/jmir.4652] [Medline: 26912201]
9. Mold F, de Lusignan S. Patients' online access to their primary care electronic health records and linked online services: implications for research and practice. J Pers Med 2015 Dec 4;5(4):452-469. [doi: 10.3390/jpm5040452] [Medline: 26690225]
10. Woods SS, Schwartz E, Tuepker A, et al. Patient experiences with full electronic access to health records and clinical notes through the My HealtheVet Personal Health Record Pilot: qualitative study. J Med Internet Res 2013 Mar 27;15(3):e65. [doi: 10.2196/jmir.2356] [Medline: 23535584]
11. Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. J Am Med Inform Assoc 2011;18(4):515-522. [doi: 10.1136/amiajnl-2011-000105] [Medline: 21672914]
12. Graham TAD, Ali S, Avdagovska M, Ballermann M. Effects of a web-based patient portal on patient satisfaction and missed appointment rates: survey study. J Med Internet Res 2020 May 19;22(5):e17955. [doi: 10.2196/17955] [Medline: 32427109]
13. Goldzweig CL, Orshansky G, Paige NM, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
14. Dutton WH, Blank G. Next Generation Users: The Internet in Britain: Oxford Internet Institute; 2011. URL: https://oxis.oii.ox.ac.uk/blog/highlights-from-the-oxis-2011-report-launch/ [accessed 2022-01-12] [doi: 10.2139/ssrn.1960655]
15. United Kingdom Department of Health. The Power of Information: Putting all of us in control of the health and care information we need. 2012. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213689/dh_134205.pdf [accessed 2022-01-20]
16. Cunningham S, Pokrajac A, Allardice B, Brillante M, Wilson L, Wake D. My Diabetes My Way: clinical outcomes impact and user experiences for an electronic personal health record for diabetes. Future Healthc J 2019 Mar;6(Suppl 1):13. [doi: 10.7861/futurehosp.6-1-s13] [Medline: 31363538]
17. Turner K, Clary A, Hong YR, Alishahi Tabriz A, Shea CM. Patient portal barriers and group differences: cross-sectional national survey study. J Med Internet Res 2020 Sep 17;22(9):e18870. [doi: 10.2196/18870] [Medline: 32940620]
18. NHS application. URL: https://digital.nhs.uk/services/nhs-app#introducing-the-nhs-app [accessed 2022-01-20]
19. NHS app. URL: https://digital.nhs.uk/news/2021/nhs-app-turns-three-with-22-million-users [accessed 2022-01-20]
20. Tewolde S, Costelloe C, Powell J, et al. An observational study of uptake and adoption of the NHS app in england. Health Informatics. Preprint posted online on 2022. [doi: 10.1101/2022.03.16.22272200]
21. Prolific Academic Ltd. Explore our participant pool demographics. URL: https://www.prolific.co/#audience [accessed 2022-02-02]
22. Medical Research Couci. Do I need an NHS REC review?. URL: http://www.hra-decisiontools.org.uk/ethics/ [accessed 2022-02-02]
23. Ndabu T, Agrawal L, Sharman R. The role of access type and age group in the breadth of use of patient portals: observational study. J Med Internet Res 2022 Dec 27;24(12):e41972. [doi: 10.2196/41972] [Medline: 36574284]
24. Agrawal L, Ndabu T, Mulgund P, Sharman R. Factors affecting the extent of patients' electronic medical record use: an empirical study focusing on system and patient characteristics. J Med Internet Res 2021 Oct 28;23(10):e30637. [doi: 10.2196/30637] [Medline: 34709181]
25. Whittemore R, Jeon S, Akyirem S, et al. Multilevel intervention to increase patient portal use in adults with type 2 diabetes who access health care at community health centers: single arm, pre-post pilot study. JMIR Form Res 2025 Mar 25;9:e67293. [doi: 10.2196/67293] [Medline: 40131327]

Abbreviations

EMR: electronic medical record
MANOVA: multivariate analysis of variance
NHS: National Health Service
PHR: patient health record

XSL•FO
RenderX

Edited by D Gooch; submitted 13.Apr.2025; peer-reviewed by I Mircheva, L Agrawal; revised version received 10.Feb.2026; accepted 10.Feb.2026; published 06.Apr.2026.

Please cite as:
Xenou M, Al-Ubaydli O
Factors Influencing User Satisfaction in Accessing Health Data: Cross-Sectional Survey of United Kingdom Adults
JMIR Hum Factors 2026;13:e75935
URL: https://humanfactors.jmir.org/2026/1/e75935
doi:10.2196/75935

© Maria Xenou, Omar Al-Ubaydli. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 6.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Real-World Performance of a New Online Eye Symptom Triage Tool (Eye+Dot) in an Emergency Eye Clinic: Mixed Methods Evaluation Study

Louise Allen[1,2], MBBS, MD, MA (Cantab); Jen O Lim[3], BA, MB BChir; Fengyi Liu[3], BA; Nimesha Alex[1], MBBS, MS, MCH, FRCOphth

[1]Department of Ophthalmology, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Research Campus, Hills Road, Cambridge, United Kingdom

[2]University of Cambridge, Cambridge, United Kingdom

[3]School of Clinical Medicine, University of Cambridge, Cambridge, United Kingdom

Corresponding Author:
Louise Allen, MBBS, MD, MA (Cantab)
Department of Ophthalmology, Cambridge University Hospitals NHS Foundation Trust, Cambridge Biomedical Research Campus, Hills Road, Cambridge, United Kingdom

Abstract

Background: Previous studies indicate that 37% to 92% of patients presenting to hospital emergency eye clinics (EECs) could be seen in commissioned community optometrist enhanced service schemes (ESSs), reducing pressure on hospital services and moving eye care into the community. Digital triage tools may have the potential to support effective triage and use of ESSs.

Objective: This study sought to evaluate the effectiveness of a recently developed online symptom triage tool in real-world EEC practice and assess its usability and acceptance by patients.

Methods: This was a prospective, real-world study comparing automated dispositions suggested by the eye+dot online triage tool with nursing triage, using ophthalmologist dispositions and subsequent clinical findings as the reference standard. Patients aged 13 years and older were sent an SMS text message invitation to use eye+dot prior to their scheduled EEC appointment. Age, time required to complete the symptom questionnaire, and acceptability (using an in-application visual Likert scale) were also recorded. The accuracy of the different triage methods at assigning patients to eye assessments within 24 hours, 48 to 72 hours, or a week based on symptom urgency was compared. Eye+dot's accuracy in identifying patient suitability for ESSs was analyzed.

Results: A total of 282 eligible patients with a mean age of 53.14 (SD 19.8; range 13-92) years were included. The mean eye+dot test duration was 5.6 (SD 2.4) minutes, with 80.4% (168/209) of patients rating the test as good or excellent. For high-acuity symptomatology (defined via retrospective ophthalmologist recommendation for assessment within 24 hours), sensitivity was similar between eye+dot and nurse triage (58/76, 76.3%; 95% CI 65.2%-85.3% and 65/76, 85.5%; 95% CI 75.6%-92.5%, respectively; P=.19). However, eye+dot specificity was significantly higher (132/206, 64.1%; 95% CI 57.1%-70.6% vs 47/206, 22.8%; 95% CI 17.3%-29.2%; P<.001). Of 224 urgent nursing dispositions, 203 (90.6%) patients were downrated (considered suitable for nonurgent assessment) by ophthalmologist assignment, and 93 (41.5%) were downrated by eye+dot. In total, 90.8% (256/282) of the patients were considered suitable for ESS assessment through eye+dot triage. A total of 56.6% (145/256) of these patients subsequently received only medical advice without specialist investigations or management at their EEC visit, suggesting suitability for community eye care.

Conclusions: Eye+dot–automated triage has similar sensitivity but superior specificity to nursing triage for identifying high-acuity symptomatology. This small study suggests its potential to improve patient scheduling in EECs throughout the working week and improve the use of community services. A larger study is planned to establish the utility of the triage tool and develop an implementation model to scale up and spread the technology.

(JMIR Hum Factors 2026;13:e81343) doi:10.2196/81343

KEYWORDS

triage; triage methods; mobile apps; computer-assisted decision-making; decision tree; emergency service; eye diseases; optometry; mobile health; digital health

XSL•FO
RenderX

Introduction

The demand for urgent eye care in the United Kingdom continues to rise, adding to capacity pressures in hospital emergency services. New-patient eye care attendance is estimated at 20 to 30 per 1000 person-years, and ophthalmic emergencies make up 1.4% to 6% of emergency department (ED) attendances [1]. Primary care triage staff may tend to be risk averse in referring patients to hospital services, and studies of emergency eye clinic (EEC) attendance demonstrate that 37% to 92% of referred patients could have been seen in community enhanced service schemes (ESSs), including minor eye condition services or community urgent eye services, or scheduled outpatient eye clinics [2-5]. Redirecting suitable patients from hospital to community care is a key component of the Royal College of Ophthalmologists emergency eye care commissioning guidelines and the UK National Health Service (NHS) 10-year plan [6-8].

Triage support charts are used in many EECs to rate symptom acuity and guide urgency disposition for less experienced triage staff. Given the pressures on emergency services, interest in digital support tools is increasing. Active public-facing UK websites that signpost patients to eye care services include EuroEyes, NHS inform, and the NHS 111 Wales Eye Problems Symptom Checker tool. New clinician-facing triage support digital platforms, such as DemDx and ASSORT (Artificial Intelligence–Based Symptom Stratification in Ophthalmology for Rapid Triage), use machine learning and large language models to support triage decisions. Eye+dot is a new online triage tool that effectively functions as a digital symptom acuity chart using Royal College of Ophthalmologists triage urgency recommendations for use by patients at the direction of the triage provider. Patients complete an online multiple-choice questionnaire, which then provides the triaging service or clinician with a symptom report, suggestions as to the urgency with which the patient should be seen, and which eye care service provider would be suitable.

The objectives of this study were to evaluate the effectiveness of the eye+dot triage tool in real-world EEC practice and assess its usability and acceptance by patients.

Methods

Eye+Dot Triage Tool

Eye+dot has been designed for evaluation of recent-onset (within 1 week) ophthalmic symptoms and is suitable for patients aged 13 years and over. To undergo the test, patients use a web link contained in an invitation sent via email or SMS text message. Where eye+dot has been integrated, the invitation can be sent using an order from the organization's electronic patient record (EPR). The test consists of a branching logic questionnaire where each response determines the next question or question sequence. Each patient is asked a maximum of approximately 25 questions covering symptomatology, ophthalmic history, and medication use. On completion, the symptom report and suggested disposition can be viewed by the triage provider using a link either sent via email or available in the EPR. The suggested disposition includes both urgency (requiring assessment within the same day, 24 hours, 48-72 hours, or a week) and the most suitable eye care service (ED or EEC, hospital or community urgent eye service, private optometry, or pharmacy). The patient does not see the report or suggested triage disposition. However, if their responses warrant it (eg, a history of recent sight-threatening injury, recent sudden loss of vision, or symptoms of stroke), the patient is fast-tracked to the end of the questionnaire and receives a link to online NHS advice, including first aid and need for immediate ED attendance. Other than age, eye+dot does not collect any personal data, and the results must be linked to the patient by the triage provider (or their EPR) using the unique order reference provided when the invitation is sent.

The eye+dot web application (Cambridge Medical Innovation Ltd) has been developed and tested over several sequential proof-of-concept and validation studies based on simulated and real-life scenarios in 3 hospital trusts of the Cambridgeshire and Peterborough Integrated Care Service and, most recently, in minor eye condition service optometry practices 9. The initial process involved the development of multiple-choice questions, with the question set, branching logic structure and triage disposition algorithm being iteratively improved over a series of patient feedback studies since the application was first developed in 2021.

The triage tool uses only self-reported symptoms and history, and the disposition algorithm is based on the scoring system used by accepted published chart-based guides (Figure 1 [7]). Integration with an EPR (Epic) enables clinicians to order and retrieve results directly through the patient record. The application is currently in clinical use in the NHS.

XSL•FO
RenderX

Figure 1. Symptom triage support chart (adapted from the Royal College of Ophthalmologists emergency eye care guidelines) [7]. CL: contact lens; ED: emergency department; EEC: emergency eye clinic; MECS: minor eye condition service

Primary problem	ED or EEC same day High acuity—emergency	EEC or MECS within 24 hours High acuity—urgent	Hospital or MECS 48-72 hours	Community service within the week
Trauma	• Chemical injury • High-velocity injury • Severe blunt trauma • Lid laceration	As in emergency presentations but onset >24 hours ago	Low-velocity injury such as that caused by fingernail or wind-blown foreign body	Other low risk injury
Red or painful eye	• Severe pain rated ≥3 out of 5 with systemic symptoms • Generalized lid swelling with systemic symptoms	• Severe pain rated ≥3 out of 5 without systemic symptoms • Pain rated ≥2 out of 5 in a CL wearer • Increasing pain within 2 weeks of an intra-ocular procedure • Photophobia to room lighting • Halos in vision • Red eye with impaired vision and unable to read large print	• Pain rated ≤2 in CL wearer • Red eye with near-normal vision and able to read normal print	• Gritty or itchy pink or red eye with no reduction in vision and no CL history • Eyelid cysts or localised lid swelling and able to open eye fully
Painless visual disturbance	• Acute-onset floaters (too many to count) with reduced vision (unable to read large print), or visual field defect • Acute-onset diplopia with systemic symptoms	• Acute onset of >10 floaters with risk factors such as high myopia or vitreoretinal history • Painless reduction of vision within 2 weeks of an intraocular procedure • Acute-onset double vision without systemic symptoms	• Low-risk floaters (<10 floaters, no sudden onset, no visual deterioration and no ophthalmic history) • Photopsia in isolation • Low-risk blurred vision – able to read titles of print and gradual onset • Subacute-onset binocular diplopia	Gradual or >1-week onset of <10 floaters without visual deterioration
Painless visual deterioration	Acute deterioration in vision (counting fingers or worse) or visual field loss within 24 hours Deterioration in vision (unable to read large print) associated with: • systemic symptoms, • halos, or • photophobia to normal room lighting	Acute reduction in vision >24 hours and unable to read large print	Subacute visual deterioration for <1 week and unable to read large print	Gradual reduction in vision for ≥1 week and unable to read normal print
External eye	Unable to open eyelids and systemically unwell	Swelling of whole lid with erythema and limited lid opening	Swelling of whole lid with erythema	Localized swelling

Study Design

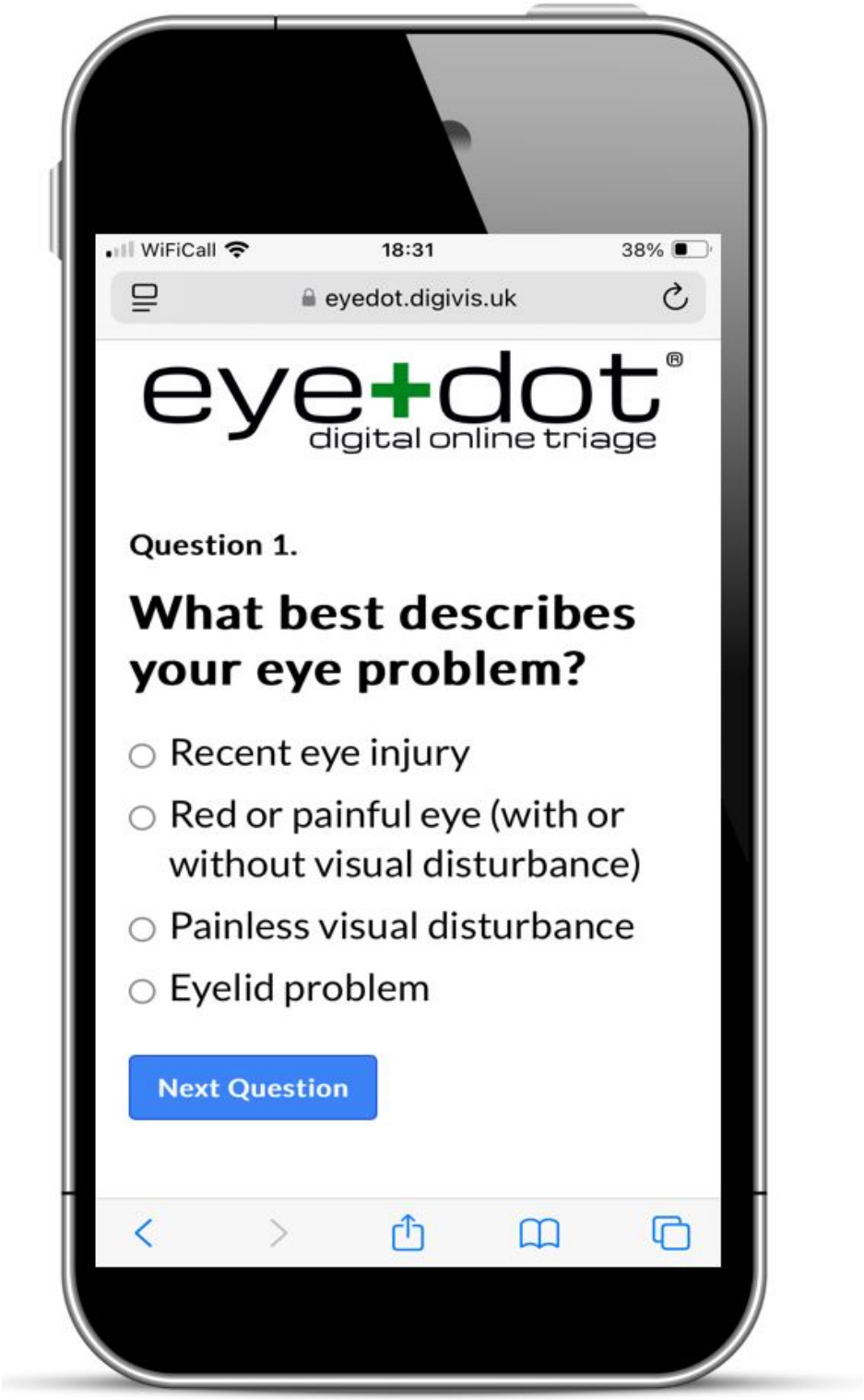
This was a prospective, single-center, real-world evaluation study that compared eye+dot triage dispositions with those generated by the standard nursing triage process. Independent triage decisions made by ophthalmologists based on their interpretation of the eye+dot symptom report were used as the standard of reference. Subsequent tests, diagnosis, and management at the scheduled EEC appointment were documented and reviewed.

The EEC at our NHS trust does not offer walk-in appointments, although patients already under the care of the trust's ophthalmology department can self-refer to this EEC. More commonly, patients are referred from primary care services via email or phone call. The specialist nurse triage team phones the patients, assesses their history, and triages them to either a scheduled EEC session or out-of-hours assessment by the on-call team. Currently, no patients are referred to local ESS optometry practices by nursing triage.

Over a 3-month period, patients aged 13 years and older with a documented mobile phone number who had scheduled appointments in EEC clinics were sent an eye+dot invitation via SMS text message prior to attendance. Patients accessing the eye+dot questionnaire were also asked to accept a General Data Protection Regulation–compliant data use notice. Data collected from the eye+dot system included the selected primary concern on the initial filter question (Figure 2), the compiled symptom report, suggested disposition, test duration, and visual Likert-scale score of acceptability (optional element of the eye+dot questionnaire; Figure 3).

XSL•FO
RenderX

Figure 2. Initial eye+dot symptom filter question as seen by the patient.



XSL•FO
RenderX

Figure 3. Visual Likert scale presented to patients after completion of the eye+dot questionnaire.

XSL•FO
RenderX

At the end of the 3-month period, records of patients who completed eye+dot triage were reviewed. Tertiary referrals, inpatient referrals, and patients referred with known diagnoses were excluded from analysis. For the remaining patients, the urgency assigned at nursing triage was documented and classified; this was defined as urgent for patients assigned for assessment on the same day (A) or within 24 hours (B), semiurgent for patients assigned for assessment within 48 to 72 hours (C), and lower urgency for patients assigned for assessment within the week (D). For the purposes of our analysis, patients were considered suitable for community care if no imaging or other diagnostic tests, prescription medication, or surgical procedures were required at the time of the EEC appointment.

To provide a reference standard, 2 ophthalmologists (a consultant and year 7 resident physician) masked to the other triage suggestions or decisions reviewed the eye+dot ophthalmic symptom report and graded the urgency using the A-to-D urgency grading criteria described above. In cases of disagreements between ophthalmologist reviewers, the higher-urgency disposition was used as the reference standard. Where a categorization of urgent by both ophthalmologists was not matched by the eye+dot and/or nursing dispositions, patient case notes were reviewed to determine whether the undertriage by eye+dot or nurses might have affected clinical outcomes.

Ethical Considerations

This service evaluation study was approved following ethical consideration by the Quality and Safety Committee at Cambridge University Hospitals NHS Foundation Trust and was conducted in accordance with the tenets of the Declaration of Helsinki. As an anonymized service evaluation study not involving randomization or deviation from the current standard of clinical care, this study was exempt from the requirement for an explicit informed consent process and national ethics committee approval (PRN13310) [10]. Patients could choose not to participate without any impact on their EEC appointment or care. The eye+dot software application replicates a chart-based symptom scoring chart and, thus, does not meet the criteria for classification as a medical device under the UK Medical Devices Regulations based on Medicines and Healthcare Products Regulatory Agency guidance [11,12].

Statistical Analysis

Patients' ages, test durations, and Likert-scale scores were analyzed using descriptive statistics. The frequency distributions of triage dispositions were compared between raters. Interrater agreement was analyzed using the percentage of agreement and quadratically weighted Cohen κ as the magnitude of disagreement in triage urgency categories is likely nonlinearly related to the resulting clinical risk for most patients. A priori interpretation of κ values was as follows: values of 0 or lower indicated no agreement, values of 0.01 to 0.20 indicated none to slight agreement, values of 0.21 to 0.40 indicated fair agreement, values of 0.41 to 0.60 indicated moderate agreement, values of 0.61 to 0.80 indicated substantial agreement, and values of 0.81 to 1.00 indicated almost perfect agreement [13]. The accuracy of eye+dot and nurse triage dispositions at identifying cases triaged as urgent (criterion A or B) by at least one ophthalmologist reviewer was calculated, and the difference in sensitivity and specificity for urgent cases was evaluated using the McNemar test with continuity correction. The number and proportion of lower-urgency patients (criterion C or D) correctly identified by eye+dot as being suitable for community ESSs was also assessed. Feedback from patients provided through a visual Likert scale (Figure 3) was converted into numerical values from 1 to 5 for analysis.

Results

Overview

Of 533 patients sent eye+dot questionnaire links, 319 (59.8%) started the questionnaire, and 315 (98.7%) of these patients completed it. In total, 33 (6.2%) patients were excluded from the subsequent analysis: of whom 17 (51.5%) already had a diagnosis at the time their appointment was scheduled, 5 (15.2%) had incomplete documentation, 10 (30.3%) did not attend, and 1 (3.0%) attended the EEC for imaging only. Therefore, full datasets from 282 (52.9% of total 533) eligible patients were available for analysis (Figure 4). Among these patients, red or painful eye and painless visual disturbance were the most common presenting concerns (Table 1).

XSL•FO
RenderX

Figure 4. Study flowchart.

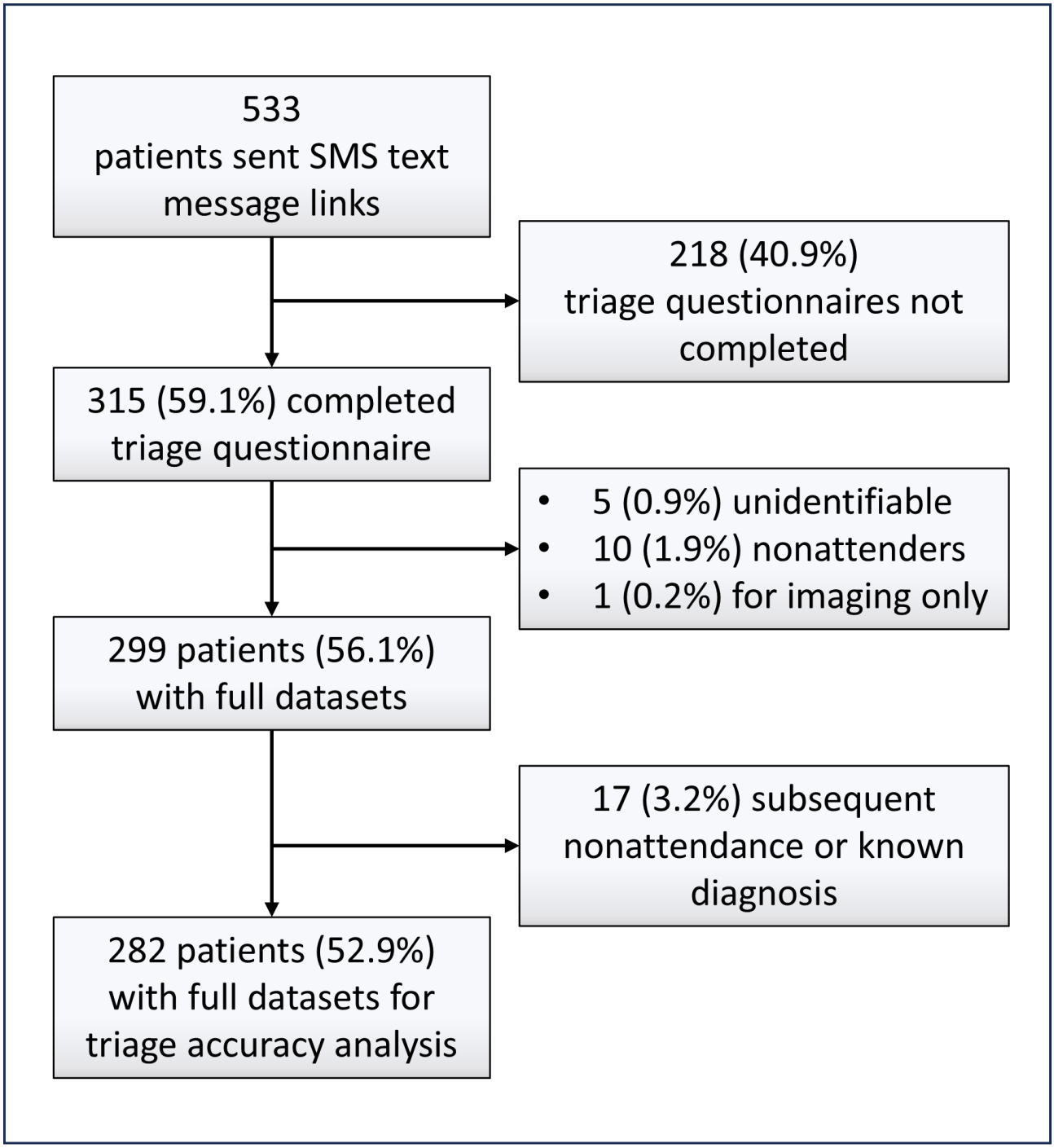


Table . Patients included for analysis by presenting concern based on response to the first question of the eye+dot questionnaire (Figure 2; N=282).

Presenting concern	Patients, n (%)
Red or painful eye	141 (50.0)
Painless visual disturbance	116 (41.1)
Eye injury	18 (6.4)
Eyelid problem	7 (2.5)

Age of Patients and Eye+Dot Acceptability

The mean ages of the patients completing eye+dot triage compared with those who chose not to use it were similar: 53.14 (SD 19.8; range 13-92) years as opposed to 53.2 (SD 19.8; range 13-94) years, respectively.

A total of 80.4% (168/209) of the patients who gave voluntary usability feedback rated the application with a 4 or 5 (with 5

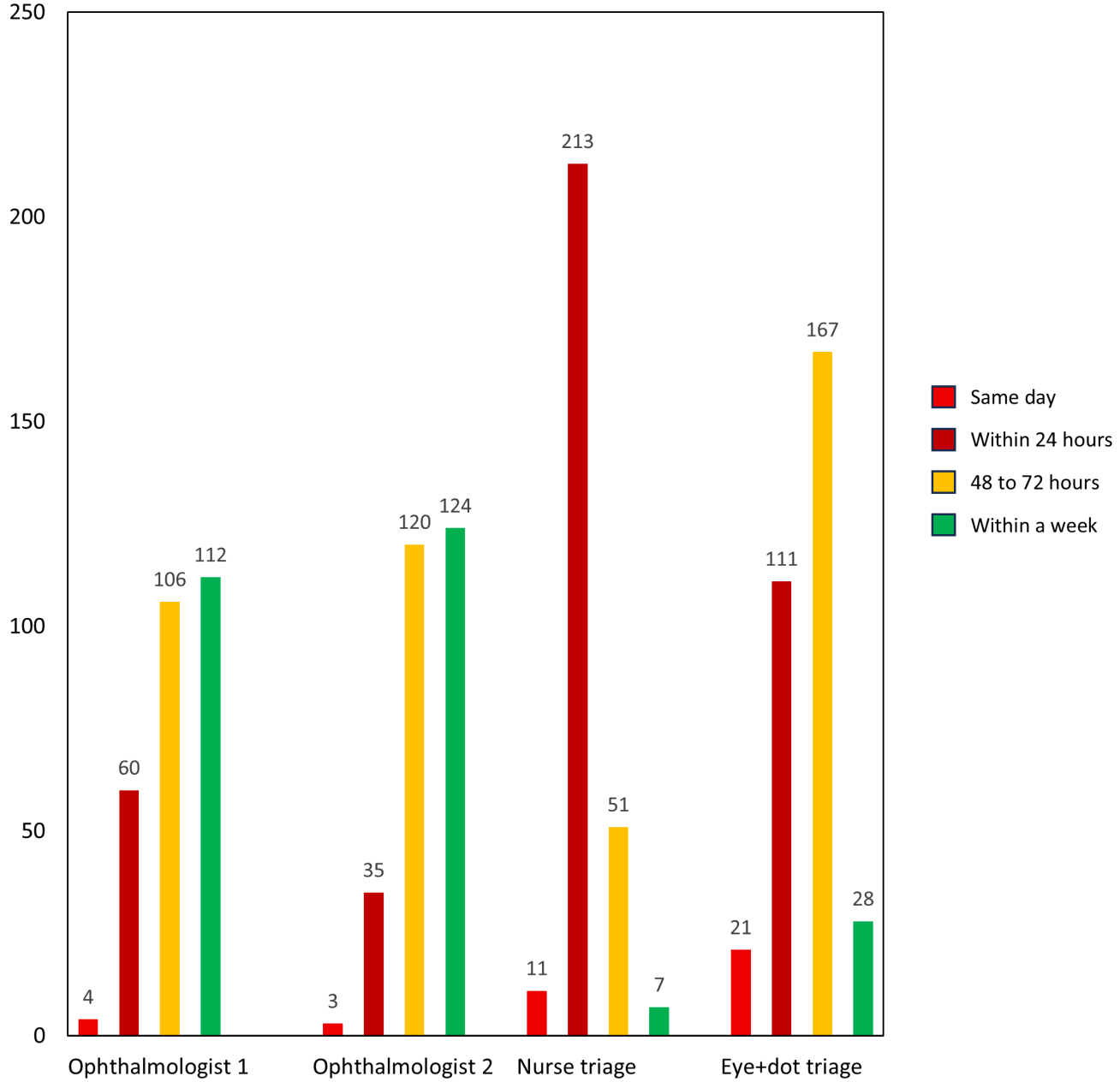
XSL•FO
RenderX

representing the most positive rating) on the visual Likert scale. The mean Likert-scale score was 4.21 (SD 0.82; range 1-5). Mean test duration was 5.6 minutes (SD 2.4 minutes; range 9 seconds-20.7 minutes).

Interrater Agreement on Urgency

The frequency distribution of urgency assignments for each rater (grades A-D) is shown in Figure 5.

Figure 5. Frequency distribution of rater urgency gradings



There was fair agreement (κ=0.22) between the 2 ophthalmologists and between ophthalmologist 1 and eye+dot (κ=0.38), but otherwise, there was only slight or almost no agreement (κ=0.01-0.2) between other raters (Table 2). Eye+dot overestimated the urgency compared with the combined ophthalmologist grading in 36.5% (103/282) of the patients and underestimated it in 1.4% (4/282); nursing triage overestimated the urgency in 59.6% (168/282) of the patients and underestimated it in 3.2% (9/282).

XSL•FO
RenderX

Table . Interrater agreements for urgency gradings A to D and percentage of agreement for urgent grades (A or B).

Raters	Cohen κ (95% CI)	Agreement with urgent grading (%)
Ophthalmologist 1 vs ophthalmologist 2	0.22 (0.10 to 0.34)	78.4
Ophthalmologist 1 vs nurse triage	0.06 (0.01 to 0.11)	37.6
Ophthalmologist 1 vs eye+dot	0.38 (0.30 to 0.47)	71.3
Ophthalmologist 2 vs nurse triage	0.01 (–0.04 to 0.06)	29.1
Ophthalmologist 2 vs eye+dot	0.05 (–0.04 to 0.13)	56.5
Eye+dot vs nurse triage	0.03 (–0.06 to 0.11)	47.9

Accuracy in Identifying Urgent Cases

Of the total 282 patients, 224 (79.4%) and 132 (46.8%) were assigned an urgent triage disposition (A or B) through nursing and eye+dot symptom triage, respectively, compared with 76 (27%) who were assigned an urgent triage disposition by at least one ophthalmologist. There was no significant difference (P=.19) in the sensitivity of nursing and eye+dot triage in identifying urgent cases as defined by ophthalmologist review, but there was a significant difference (P<.001) in specificity (Table 3). The percentage of agreement of nursing and eye+dot triage with cases graded as urgent by one or both ophthalmologists was 39.7% and 67.4%, respectively.

Table . Sensitivity and specificity of nurse and eye+dot triage for identifying urgent cases (compared with the reference-standard grades of A or B assigned by one or both ophthalmologist reviewers). *P* values were calculated using the McNemar test (N=282).

	Nurse triage vs reference-standard urgent grades (%; 95% CI)	Eye+dot triage vs reference-standard urgent grades (%; 95% CI)	*P* value
Sensitivity	85.5 (75.6-92.5)	76.3 (65.2-85.3)	.19
Specificity	22.8 (17.3-29.2)	64.1 (57.1-70.6)	<.001

Clinical Presentations and Outcomes of Urgent Cases

Further analysis was undertaken in the subset of 7.4% (21/282) of patients whom both ophthalmologists classified as urgent. In total, 19% (4/21) of these patients had been undertriaged by nursing or eye+dot triage. A total of 4.8% (1/21) of the patients' triage decisions were uprated by the ophthalmologists from both the nurse and eye+dot dispositions of C (assessment within 48 hours). This patient had added "I have a foreign object in my upper lid which is now protruding over my eyelashes and it's now blocking my vision in my right eye" in the free-text entry, which was visible to the ophthalmologists on the eye+dot report. However, upon review, this patient's injury had been sustained 1 year before, and he had a retained glass foreign body in the eyelid since that time; this history had been elicited during nursing triage.

Nursing triage was uprated from C to B in a further 14.3% (3/21) of the patients: one had a resolving corneal abrasion, the second had suspected migrainous aura with associated temporal tenderness, and the third had ophthalmic shingles with corneal involvement.

Of the 224 urgent nursing triage dispositions, 203 (90.6%) patients were downrated by ophthalmologist review, and 93 (41.5%) were downrated by eye+dot.

Service Disposition

In total, 46.5% (131/282) of the patients who attended the EEC did not require hospital diagnostic or medical management and, thus, were considered to have been suitable for community care through ESSs.

Eye+dot suggested suitability for ESSs in 90.8% (256/282) of the patients. A total of 56.6% (145/256) of these patients required assessment and advice only during their subsequent EEC visit and may have been suitable for community optometrist care. The other 43.4% (111/256) of the patients received diagnostic imaging or medical management in the EEC and may potentially have needed subsequent referral to hospital eye services (HESs) had they been seen in ESSs initially.

Discussion

Principal Findings

Triage systems should ideally have both high sensitivity and high specificity to ensure timely access to care and efficient use of resources. Triage accuracy particularly depends on factors such as clinical experience, evaluation time, and the use of support tools or frameworks [14,15].

A recent large retrospective study at Moorfields Eye Hospital ED reported nurse triage sensitivity of 96.4% and specificity of 25.1% for identifying emergency cases among patients attending the EEC [16].

This real-world validation study found similar results for nurse-led triage in relation to patients assigned an urgent grading, but the eye+dot triage support tool, although it had similar sensitivity (P=.19), showed significantly higher specificity for emergency cases than nursing triage (P<.001).

These results suggest that use of eye+dot could have safely downrated the triage category in 41.5% (93/224) of patients assigned as urgent via nurse triage, potentially enabling redistribution of patient load throughout the working week and redirecting 51.4% (145/282) of patients from a planned EEC

XSL•FO
RenderX

attendance to a community ESS appointment without the risk of a subsequent rebound referral back to the HES.

Comparison of Interrater Agreement to That in Prior Studies

It is notable that the interrater agreement in this study was poor between most triage modes (optometrists, eye+dot, and ophthalmologist reviewers), and was only fair between the 2 ophthalmologists (κ=0.22, 95% CI 0.10-0.34) and between eye+dot and the consultant ophthalmologist (κ=0.38, 95% CI 0.30-0.47). This stands in contrast to interrater agreements reported in other studies using digital triage tools. A real-world validation study at Scheie Eye Institute, University of Pennsylvania, reported excellent interrater agreement between their online symptom-based triage tool and reference standard (κ=0.91). However, this may be because prioritization categories were in wider, more discrete urgency bands of "emergency/urgent (same day)," "semi-urgent 1-4 weeks," and "nonurgent/elective" compared with the relatively narrow urgency categories used in this study [17]. Studies using large language models have also claimed relatively high interrater agreement (κ=0.773-0.751) for their models using simulated clinical scenarios, but this may not reflect the complex and sometimes inconsistent symptom reporting often encountered in the real world [18-20]. A recent pilot study of the GPT-4–powered ASSORT triage tool in 51 patients presenting in an EEC with concerns requiring prompt evaluation showed moderate interrater agreement between ASSORT and the triaging ophthalmologist (κ=0.54) with overestimation of severity [21]. In that cohort, 56.8% of patients were considered urgent by the ophthalmologist compared with the 27% (76/282) in our study population of nurse-triaged scheduled attenders, suggesting that the populations may not be comparable.

Utility of the Triage Tool as an Aid for Redirecting Patients to Community Services

Using the eye+dot triage tool, all patients with newly reported recent-onset symptoms would have had a professional eye assessment within the week, but the urgent workload would have been more evenly distributed, with most patients (256/282, 90.8%) potentially being assessed in the community ESSs first.

Of these, 43.4% (111/256) may have required diagnostic imaging or medical management, necessitating a subsequent rebound referral to HESs, although this is likely to be an overestimation as many ESS providers now have diagnostic imaging and prescribing optometrists available. A 2016 study reported that only 18.9% of 2123 patients attending ESSs in a South London area required a subsequent referral to HESs [22].

This potential benefit of redirecting patients to community ESSs before hospital assessment cannot be fully realized by some other triage tools which require the input of nurse-elicited symptoms and signs. Examples include DemDx, the Alphabetical Triage Score for Ophthalmology, and the Rome Eye Scoring System for Urgency and Emergency [16,23,24].

Usability Feedback and Potential Workflow Issues

A total of 59.8% (319/533) of the patients who were sent invitations clicked on the eye+dot link in the SMS text message, which we consider to represent a reasonable openness to engaging with the new technology, given that they had received no prior notification to expect this. Providing advance information regarding eye+dot at the time of initial contact might have increased uptake. The eye+dot application was generally well accepted and was rated within the top 2 out of 5 possible Likert scale options by 80.4% (168/209) of the patients in this study who responded to the request for feedback at the end of the questionnaire. All had already been scheduled for an EEC appointment, which may have influenced their acceptance, and it is possible that reported usability of the tool may decrease if the information collected is used to redirect patients to community services. A related potential issue, which requires evaluation in practice, is whether patients who had already been referred to the EEC by their primary care team would be resistant to being redirected to community eye care providers based on this triage process. Both ESSs and EECs are free-to-access NHS services, but patient resistance to triage advice based on financial cost may be a necessary factor to consider in other countries or contexts. NHS England is developing single point of access call centers for patients with recent-onset symptoms to be triaged and directed to a suitable service. This may be the ideal point to implement specialist digital triage tools such as eye+dot to prevent the patient from being bounced between health care providers.

There was no difference in age between those who responded to the questionnaire and those who did not; however, it is possible that older patients without mobile phone access may have been underrepresented. Systematic reviews of patient experience of primary care online triage describe the importance of issues of digital and literacy inequality, sight impairment, concerns regarding privacy, and distrust of online platforms. Some patients find that the convenience and speed of online triage platforms outweigh their transactional and impersonal nature, and for others, it is quite the reverse [25,26]. Access to the technology for the visually impaired may be considered a particular limitation for an eye symptom triage tool, although this is mitigated by the ease of responding to the eye+dot questionnaire over the telephone. None of the patients in this study cohort were identified as visually impaired, possibly because patients with preexisting eye conditions are more likely to have direct lines of access to specialist care and be less likely to use a primary triage tool. For staff, although there may be concerns that greater use of digital triage tools could lead to deskilling or even job losses, given the current staffing shortages and ever-increasing workload, this seems unlikely. Both they and patients may benefit from the advantages and support of an automated questionnaire that can be completed in the patients' own time and records important red-flag and negative symptomatology that can be documented in the EPR.

Conclusions

This real-world study demonstrates that eye+dot triage has similar sensitivity to but significantly higher specificity than nursing triage in assessing urgency for patients scheduled for EEC appointments. An implementation study is planned to determine its utility in safely downtriaging and redirecting appropriate cases to community eye care services and the effect

XSL•FO
RenderX

this has on ED and EEC service efficiency and to develop a model for the spread of the technology.

Acknowledgments

The authors are grateful to Alexandra Mirza-Davies for supporting data collection and analysis and to the staff of the emergency eye clinic at Addenbrooke's Hospital for their support with the study.

Funding

This research was unfunded but supported by the National Institute for Health and Care Research Cambridge Biomedical Research Centre (NIHR203312). The views expressed are those of the authors and not necessarily those of the National Institute for Health and Care Research or the Department of Health and Social Care.

Data Availability

The datasets generated or analyzed during this study are available in the Open Science Framework [27].

Authors' Contributions

LA and JOL wrote the manuscript with input from the other authors. LA and FL were responsible for data collection. JOL carried out the statistical analysis. LA and NA reviewed patient reports to determine optimum triage dispositions.

Conflicts of Interest

LA is the developer of the eye+dot web application and unsalaried director of Cambridge Medical Innovation Ltd, the manufacturer of eye+dot. All other authors declare no other conflicts of interest.

References

1. Emergency eye care in hospital eye units and secondary care. The Royal College of Ophthalmologists. 2024. URL: https://www.rcophth.ac.uk/wp-content/uploads/2021/01/Emergency-eye-care-in-hospital-eye-units-and-secondary-care-Copy.pdf [accessed 2026-03-15]
2. Hau S, Ioannidis A, Masaoutis P, Verma S. Patterns of ophthalmological complaints presenting to a dedicated ophthalmic Accident & Emergency department: inappropriate use and patients' perspective. Emerg Med J 2008 Nov;25(11):740-744. [doi: 10.1136/emj.2007.057604] [Medline: 18955608]
3. MacIsaac JC, Naroo SA, Rumney NJ. Analysis of UK eye casualty presentations. Clin Exp Optom 2022 May;105(4):428-434. [doi: 10.1080/08164622.2021.1937949] [Medline: 34315329]
4. Rehan SM, Morris DS, Pedlar L, Sheen N, Shirodkar AL. Ophthalmic emergencies presenting to the emergency department at the University Hospital of Wales, Cardiff, UK. Clin Exp Optom 2020 Nov;103(6):895-901. [doi: 10.1111/cxo.13050] [Medline: 32066198]
5. Vernon SA. Analysis of all new cases seen in a busy regional centre ophthalmic casualty department during 24-week period. J R Soc Med 1983 Apr;76(4):279-282. [doi: 10.1177/014107688307600408] [Medline: 6601713]
6. The NHS Long Term Plan. National Health Service. 2019. URL: https://webarchive.nationalarchives.gov.uk/ukgwa/20230418155402/https:/www.longtermplan.nhs.uk/publication/nhs-long-term-plan/ [accessed 2026-03-30]
7. Emergency eye care commissioning guidance. The Royal College of Ophthalmologists. 2020. URL: https://www.rcophth.ac.uk/wp-content/uploads/2021/12/Emergency-Eye-Care-Commissioning-Guidance.pdf [accessed 2026-03-15]
8. 10 year health plan for England: fit for the future. United Kingdom Government, Department of Health and Social Care. 2025. URL: https://www.gov.uk/government/publications/10-year-health-plan-for-england-fit-for-the-future [accessed 2026-03-15]
9. Lim JO, Farrell S, Newlands SJ, Allen L. Real-world evaluation of the eye+dot online triage support tool in community optometry practices: mixed methods evaluation study. JMIR Hum Factors 2026 Mar 16;13:e77278. [doi: 10.2196/77278] [Medline: 41838883]
10. UK Policy Framework for health and social care research. National Health Service Health Research Authority. URL: https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/uk-policy-framework-health-social-care-research/uk-policy-framework-health-and-social-care-research/ [accessed 2026-03-30]
11. The Medical Devices Regulations 2002. Legislation.gov.uk. 2002. URL: https://www.legislation.gov.uk/uksi/2002/618/contents [accessed 2026-03-30]
12. Medical devices: software applications. Medicines & Healthcare products Regulatory Agency, United Kingdom Government. 2014. URL: https://www.gov.uk/government/publications/medical-devices-software-applications-apps [accessed 2026-03-30]
13. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282. [Medline: 23092060]
14. Ouellet S, Gallani MC, Fontaine G, et al. Strategies to improve the quality of nurse triage in emergency departments: a systematic review. Int Emerg Nurs 2025 Aug;81:101639. [doi: 10.1016/j.ienj.2025.101639] [Medline: 40532318]

XSL•FO
RenderX

15. Seo YH, Lee K, Jang K. Factors influencing the classification accuracy of triage nurses in emergency department: analysis of triage nurses' characteristics. BMC Nurs 2024 Oct 17;23(1):764. [doi: 10.1186/s12912-024-02334-9] [Medline: 39420318]
16. Brandao-de-Resende C, Melo M, Lee E, et al. A machine learning system to optimise triage in an adult ophthalmic emergency department: a model development and validation study. EClinicalMedicine 2023 Dec 2;66:102331. [doi: 10.1016/j.eclinm.2023.102331] [Medline: 38089860]
17. Meer E, Ramakrishnan MS, Whitehead G, Leri D, Rosin R, VanderBeek B. Validation of an automated symptom-based triage tool in ophthalmology. Appl Clin Inform 2023 May;14(3):448-454. [doi: 10.1055/a-2065-4613] [Medline: 36990454]
18. Schumacher I, Ferro Desideri L, Bühler VM, et al. Performance analysis of an emergency triage system in ophthalmology using a customized chatbot. Digit Health 2025 May 11;11:20552076251320298. [doi: 10.1177/20552076251320298] [Medline: 40357425]
19. Tsui JC, Wong MB, Kim BJ, et al. Appropriateness of ophthalmic symptoms triage by a popular online artificial intelligence chatbot. Eye (Lond) 2023 Dec;37(17):3692-3693. [doi: 10.1038/s41433-023-02556-2] [Medline: 37120656]
20. Waisberg E, Ong J, Zaman N, et al. GPT-4 for triaging ophthalmic symptoms. Eye (Lond) 2023 Dec;37(18):3874-3875. [doi: 10.1038/s41433-023-02595-9] [Medline: 37231187]
21. Xompero C, Rossi L, Amoroso F, et al. Pilot study of ASSORT (AI-based Symptom Stratification in Ophthalmology for Rapid Triage): a triage tool for ophthalmic emergencies. AJO Int 2025 Oct;2(3):100159. [doi: 10.1016/j.ajoint.2025.100159]
22. Konstantakopoulou E, Edgar DF, Harper RA, et al. Evaluation of a minor eye conditions scheme delivered by community optometrists. BMJ Open 2016 Aug 10;6(8):e011832. [doi: 10.1136/bmjopen-2016-011832] [Medline: 27515757]
23. D'Oria F, Bordinone MA, Rizzo T, et al. Validation of a new system for triage of ophthalmic emergencies: the alphabetical triage score for ophthalmology (ATSO). Int Ophthalmol 2020 Sep;40(9):2291-2296. [doi: 10.1007/s10792-020-01413-5] [Medline: 32415656]
24. Rossi T, Boccassini B, Iossa M, Mutolo MG, Lesnoni G, Mutolo PA. Triaging and coding ophthalmic emergency: the Rome Eye Scoring System for Urgency and Emergency (RESCUE): a pilot study of 1,000 eye-dedicated emergency room patients. Eur J Ophthalmol 2007;17(3):413-417. [doi: 10.1177/112067210701700324] [Medline: 17534826]
25. Roberts C, George J, Jenkins J. Patient experiences of UK primary care online triage and consultation platforms during COVID-19: a systematic review. medRxiv. Preprint posted online on Jan 10, 2024. [doi: 10.1101/2024.01.09.24301062]
26. Wallace W, Chan C, Chidambaram S, et al. The diagnostic and triage accuracy of digital and online symptom checker tools: a systematic review. NPJ Digit Med 2022 Aug 17;5(1):118. [doi: 10.1038/s41746-022-00667-w] [Medline: 35977992]
27. Allen L. Eye+dot use in emergency eye clinic. Open Science Framework. URL: https://osf.io/q2387/overview [accessed 2026-04-08]

Abbreviations

ASSORT: Artificial Intelligence–Based Symptom Stratification in Ophthalmology for Rapid Triage
ED: emergency department
EEC: emergency eye clinic
EPR: electronic patient record
ESS: enhanced service scheme
HES: hospital eye service
NHS: National Health Service

Edited by A Kushniruk; submitted 26.Jul.2025; peer-reviewed by G Hadimani, Z Zandesh; revised version received 10.Feb.2026; accepted 21.Feb.2026; published 14.Apr.2026.

Please cite as:
Allen L, Lim JO, Liu F, Alex N
Real-World Performance of a New Online Eye Symptom Triage Tool (Eye+Dot) in an Emergency Eye Clinic: Mixed Methods Evaluation Study
JMIR Hum Factors 2026;13:e81343
URL: https://humanfactors.jmir.org/2026/1/e81343
doi:10.2196/81343

© Louise E Allen, Jen O Lim, Fengyi Liu, Nimesha Alex. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 14.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic

XSL•FO
RenderX

information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluating a Shared Decision Support Tool for Pediatric Cardiopulmonary Arrest: Mixed Methods Usability Study

Ana Rajic[1,2], MS; Sharleen Kayne Olanka[1,2], MS; Marco Generelli[1,2], MS; Jennifer Davidson[3,4], RNCS; Yiqun Lin[4], MD, MHSc, PhD; Ryan Kang[5], MSc; Kangsoo Kim[5], PhD; Pierre-Louis Rebours[6]; Marc Ibrahim[6]; Donovan Duncan[7], MD; Sergio Manzano[2,8], Prof Dr; Adam Cheng[3,4], Prof Dr; Alexandre De Masi[8], PhD; Johan N Siebert[2,8*], MD, PD; Frederic Ehrler[6,8*], PhD

[1]Educational Technologies and Learning Sciences (TECFA), Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland
[2]Department of Pediatric Emergency Medicine, Geneva Children's Hospital, Geneva University Hospitals, 47 Avenue de la Roseraie, Geneva, Switzerland
[3]Departments of Pediatrics and Emergency Medicine, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada
[4]KidSIM-ASPIRE Simulation Research Program, Alberta Children's Hospital, University of Calgary, Calgary, AB, Canada
[5]Department of Electrical and Software Engineering, Schulich School of Engineering, University of Calgary, Calgary, AB, Canada
[6]Division of Computer Sciences, Diagnostic Department, Geneva University Hospitals, Geneva, Switzerland
[7]Pediatric Intensive Care Unit, Alberta Children's Hospital, University of Calgary, Calgary, AB, Canada
[8]Faculty of Medicine, University of Geneva, Geneva, Switzerland
[*]these authors contributed equally

Corresponding Author:
Johan N Siebert, MD, PD
Department of Pediatric Emergency Medicine, Geneva Children's Hospital, Geneva University Hospitals, 47 Avenue de la Roseraie, Geneva, Switzerland

Abstract

Background: Effective team communication is critical in pediatric cardiopulmonary arrest management, where delays or miscommunication can jeopardize survival. TeamScreen, a web-based interface displayed on a large screen, was developed to enhance cardiopulmonary resuscitation (CPR) by providing real-time visualization of clinical data and resuscitation steps aligned with the American Heart Association pediatric advanced life support algorithms.

Objective: This study evaluated the usability of the TeamScreen Figma prototype, evaluating how efficiently and accurately experienced emergency physicians and nurses retrieved critical information during a simulated pediatric in-hospital cardiac arrest scenario. Although no strict time constraints were imposed, participants were instructed to perform the tasks as spontaneously and as quickly as possible.

Methods: Usability testing involved 20 pediatric emergency physicians and nurses with varied CPR experience. Participants performed 21 information retrieval tasks within a simulated pediatric cardiac arrest scenario (shockable rhythm). The data collected included audio-video recordings via the think-aloud method and participant responses to the Post-Study System Usability Questionnaire (PSSUQ) version 3 and a posttest survey. Effectiveness, efficiency, and satisfaction were measured by task completion rates, time-on-task metrics, and PSSUQ scores, respectively. Think-aloud data were analyzed for usability issues using Nielsen Norman Group's rating scale and Bastien and Scapin's ergonomic criteria.

Results: Five physicians and 15 nurses achieved a mean task success rate of 81.19% (SD 16.87%), with a mean completion time of 8.13 (SD 7.07) seconds, calculated across all 21 tasks and all participants. PSSUQ scores reflected high satisfaction (mean 2.40 [SD 1.24] of 7.00; the lower the better), notably for information clarity and system utility. Qualitative analyses identified 16 usability issues, of which 5 were deemed major, primarily involving information visibility, navigation, and density, highlighting areas for interface and workflow enhancement.

Conclusions: The usability evaluation confirmed TeamScreen's potential to improve real-time information access during pediatric CPR, with high task success and satisfaction scores supporting its role in aiding decision-making. Challenges with visibility, navigation, and information density require further refinement. These findings will guide improvements and inform the design of multicenter trials to assess TeamScreen's efficacy in simulation-based resuscitation settings.

(JMIR Hum Factors 2026;13:e78736) doi:10.2196/78736

XSL•FO
RenderX

KEYWORDS

cardiopulmonary resuscitation; pediatric emergency medicine; decision support systems; clinical; digital health; user-computer interface; health information technology; medical informatics applications; usability testing; simulation training

Introduction

Context and Problem Statement

In-hospital pediatric cardiac arrests (IHCA), although rare, require immediate, coordinated response from resuscitation teams under intense pressure to provide guideline-compliant clinical care [1]. Standardized protocols, such as the American Heart Association pediatric advanced life support (PALS) guidelines, emphasize uniform responses [2], yet adherence remains suboptimal, with frequent deviations from recommended timing of defibrillation and drug administration, errors in dosing, and omissions of key steps in both real and simulated pediatric cardiac arrest scenarios [3-6]. These deviations are strongly influenced by human factors such as high cognitive load, stress, impaired decision-making, communication breakdowns, and challenges in maintaining team situational awareness and a shared mental model during complex emergencies [4]. These lapses compromise quality of care and survival rates [7]. As a result, there is increasing interest in tools that support team cognition and real-time guideline adherence during pediatric cardiopulmonary resuscitation (CPR).

Digital cognitive aids and decision support systems have emerged as promising approaches to mitigate these problems. In adult and mixed emergency settings, cognitive aids and electronic decision support tools have been associated with fewer errors and better adherence to resuscitation algorithms in simulated IHCA [8-11]. In pediatrics, few are purpose-built for real-time use in resuscitation [12]. Tablet-based apps were developed specifically to support pediatric cardiac arrest management and showed good usability in high-fidelity simulation with reduced deviations from guideline recommendations [13-15]. A recent systematic review of cognitive aids in resuscitation similarly concluded that cognitive aids tend to reduce deviations from guidelines and improve resuscitation performance in simulated settings across adult and pediatric populations [16]. However, most existing systems are designed as single-user, handheld interfaces that primarily support the team leader, rather than functioning as shared, team-centered displays that make time-critical information visible to the entire resuscitation team in real time.

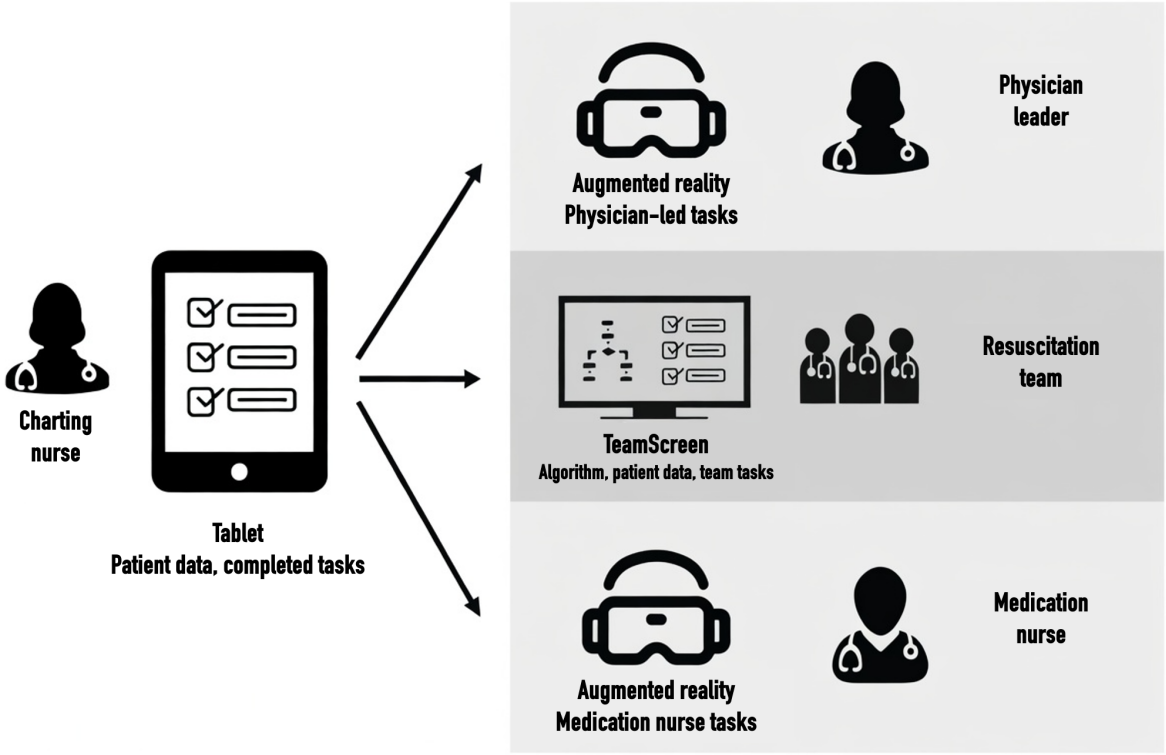
Only a few technological solutions have been conceived as large situation displays to enhance team communication and situational awareness during emergency resuscitations. Pilot work in emergency departments has suggested that dedicated situation displays can facilitate teamwork, communication, and perceived situational awareness among resuscitation teams [17,18]. More recently, an interactive, large-screen clinical decision display designed for team-wide viewing has been evaluated and shown to improve adherence to guideline time intervals in simulated advanced cardiac life support [8]. Despite these advances, pediatric-specific evidence for team-centered digital decision support during IHCA remains limited, and there is a need for tools explicitly designed around the information needs and workflows of pediatric resuscitation teams.

To address these limitations, we developed Interconnected and Focused Mobile Applications on Patient Care Environment (InterFACE), a suite of interconnected digital tools (Figure 1) designed to enhance team coordination and PALS adherence during pediatric CPR [19,20]. Building on an earlier prototype [15] that demonstrated improvements in defibrillation timing and adherence to resuscitation algorithms but also highlighted risks of cognitive overload, we refined the system into an updated concept. This system integrates role-specific, interconnected digital tools—TeamScreen, Guiding Pad (a tablet app), and augmented reality (AR) headsets [21]—to standardize care, reduce variability, and enhance CPR quality [22]. TeamScreen provides real-time resuscitation oversight, with shared access to drug and shock dose calculations and PALS workflows to support the entire team [20]. Developed iteratively with multidisciplinary input from 2 pediatric emergency departments, it addresses prior shortcomings and aligns with calls for innovative resuscitation aids [16].

The design and evaluation of InterFACE-AR followed human factors and usability engineering recommendations for high-risk medical technologies. Regulatory guidance from the US Food and Drug Administration on applying human factors and usability engineering to medical devices emphasizes structured processes, identifying critical tasks, conducting formative evaluations, and performing summative usability testing to minimize use-related risks [23]. Usability research in health information technology further supports the use of task-based testing and think-aloud methods as core, user-centered approaches to uncover interaction problems in complex interfaces [24,25]. In this context, this study focuses specifically on evaluating the usability of the updated TeamScreen interface during a simulated pediatric IHCA scenario, as an essential step before broader assessments of its impact on team performance and clinical outcomes.

XSL•FO
RenderX

Figure 1. Components of Interconnected and Focused Mobile Applications on Patient Care Environment–augmented reality (InterFACE-AR). The system integrates multiple digital tools to support pediatric resuscitation teams. The Guiding Pad (tablet), used by the charting nurse, captures patient data and completed tasks, distributing real-time information to role-specific interfaces. The TeamScreen (web-based app displayed on a large screen) provides a centralized overview of the resuscitation algorithm, patient data, and team tasks for the entire resuscitation team. Augmented reality (AR) headsets (HoloLens 2) deliver targeted, real-time guidance to the physician leader (for physician-led tasks) and the medication nurse (for medication preparation and administration).



Objective

This study aimed to evaluate the usability of the TeamScreen app in supporting resuscitation teams during pediatric IHCA management. By assessing the prototype's strengths and limitations, we sought to identify potential usability challenges and propose targeted improvements. The insights gathered were intended to inform the refinement of TeamScreen ahead of its full-scale clinical deployment and to contribute to the broader development of digital tools aimed at enhancing team performance and adherence to resuscitation guidelines.

Methods

Study Design

The usability of the TeamScreen app was assessed through a scenario-based information retrieval evaluation, combining both quantitative and qualitative methods. Multitask quantitative and qualitative usability metrics were used and are described in detail in subsequent sections. Usability was defined, according to ISO 9241‐210, as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [26]. The usability of a digital tool can be assessed by how effectively and completely users are able to perform key tasks aligned with its core functionalities. In contrast, poor usability often results in reduced task efficiency, incomplete goal achievement, or even abandonment of the technology [27]. Effectiveness was measured by task completion rates. Efficiency was evaluated through time-on-task metrics. User satisfaction was assessed using the Post-Study System Usability Questionnaire (PSSUQ) version 3 [28] and a custom posttest evaluation questionnaire. In addition, think-aloud discussions were recorded verbatim and analyzed to extract key user insights and perceptions throughout the sessions.

TeamScreen App Overview: A Core Component of InterFACE-AR

TeamScreen is a web-based app designed to centralize and display, in real time, critical information related to the patient's condition and resuscitation progress to support adherence to PALS algorithms during CPR. Displayed on a 75-inch wall-mounted screen in the resuscitation room, it provides resuscitation teams with a clear, shared overview of clinical status and required actions, aligned with PALS algorithms. This dynamic interface offers a visual overview of key data such as patient information (eg, weight, pulse presence, cardiac rhythm, and ventilation status), intervention timelines, upcoming clinical tasks, and medication dosing details without requiring direct user interaction. Instead, it relies on input from the charting nurse via the Guiding Pad, a tablet app that dynamically records interventions tailored to the patient's status. TeamScreen is part of the broader InterFACE-AR suite [22], named for its AR foundation, which also includes AR headsets (HoloLens 2, Microsoft Corp), worn by the team leader and medication nurse [21]. These deliver role-specific guidance via mixed reality, an

XSL•FO
RenderX

advanced form of AR that—unlike standard overlays—enables interactive 3D holograms anchored in the environment [29], to support clinical decision-making and medication preparation. The usability testing approach mirrored this role distribution, reflecting real-world dynamics where the Guiding Pad feeds data to TeamScreen for team-wide visibility. These tools collectively aim to improve situational awareness, communication, and adherence to PALS guidelines during IHCA, ensuring more efficient and guideline-compliant interventions. The TeamScreen prototype used in this study replicated all core functionalities of the current InterFACE-AR TeamScreen module, including the algorithm flow, task lists, timers, and weight-based drug and shock displays, but it was operated as a stand-alone simulated interface that was not technically connected to the AR headsets or live clinical systems. The Guiding Pad and AR headset components will be described and evaluated in separate studies.

Participants and Setting

Inclusion Criteria and Recruitment

The inclusion criteria comprised physicians and nurses working in the Pediatric Emergency Department of the Geneva University Hospitals in Switzerland; fluent in French at a C1 level; and with prior experience in pediatric resuscitation, gained through real clinical practice, simulation training, or both. Individuals involved in the development of the InterFACE project were excluded from the study to avoid bias.

Sample size was established at 20 participants. This number is consistent with established guidelines for usability testing in high-risk, domain-specific systems, where 15 to 20 participants are generally considered sufficient to identify the majority of critical usability issues. This recommendation is supported by prior work in the field, particularly in the context of summative evaluations of medical technologies [30]. The evaluation framework focused on interactions between the user, the task, and the system, intentionally excluding the user's actual clinical environment from the assessment [31]. Recruitment was carried out by 2 senior attending physicians in pediatric emergency medicine (SM and JNS).

Ethical Considerations

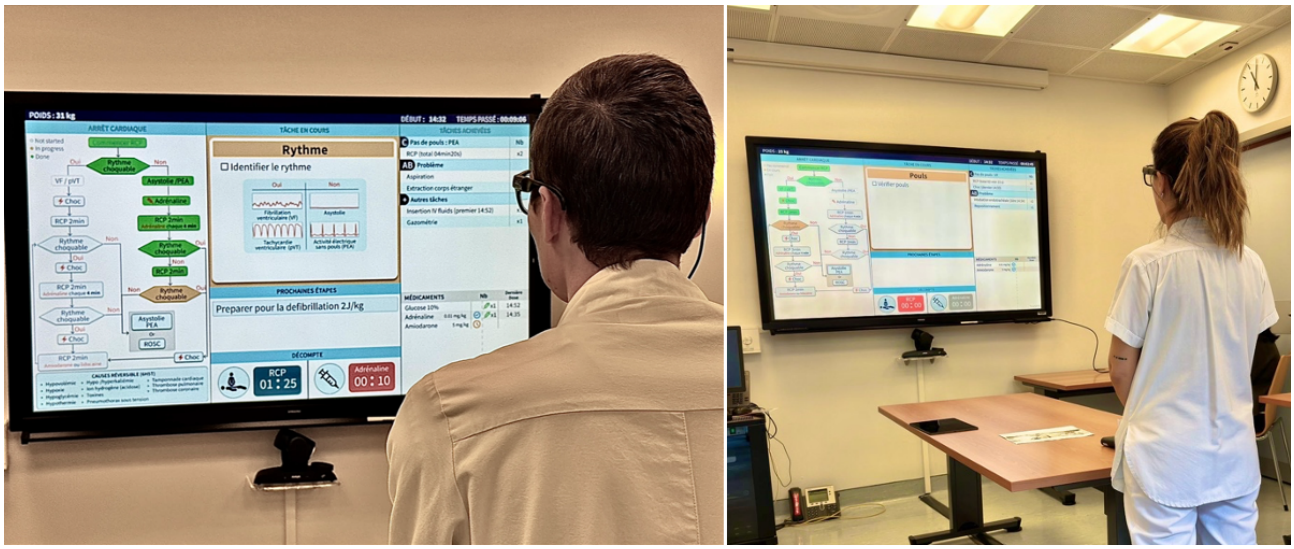
The study received ethical approval from the Geneva Cantonal Research Ethics Committee (reference Req-2023 - 00162). All participants received an information sheet detailing the study's objectives and provided written informed consent before participating. Participants could opt out at any time without providing a reason and without any consequences. Participants received no compensation for their participation. All data were anonymized to ensure confidentiality. Explicit consent was obtained from participants to use their image for publication.

Test Setting

The usability evaluation was conducted in a conference room, equipped with a prototype of the TeamScreen system displayed on a 75-inch wall-mounted Samsung Neo QLED screen (model QE75QN90). This standardized setup was designed to simulate real clinical environment and enable technical data collection. Participants stood 2.7 meters from the screen, with a table placed in front of it to simulate a resuscitation bed (Figure 2). This conference room layout (screen position, distance, and table simulating the resuscitation bed) was chosen to approximate a pediatric emergency department resuscitation bay, with the TeamScreen acting as the central "wall monitor" around which the simulated scenario unfolded. Testing sessions were conducted by 3 examiners (ADM, SO, and AR), each responsible for different aspects of the evaluation. SO operated the Guiding Pad connected to the TeamScreen and acted as the charting nurse, simulating real-time input of clinical actions to drive the system forward in alignment with the usability tasks presented to participants. AR facilitated the sessions, administered instructions, and guided the participants through the PALS algorithm, questionnaires, and posttest survey. All verbal responses and think-aloud reflections were recorded using lapel microphones. Data collection tools included a demographic questionnaire, the PSSUQ version 3, and a custom posttest survey, all administered digitally via Microsoft Forms, with responses stored in real time.

XSL•FO
RenderX

Figure 2. Usability test setting with TeamScreen. Two health care professionals interact with TeamScreen, displayed on a 75-inch wall-mounted screen, during the usability testing session. The participants stand at a standardized distance while engaging in scenario-based tasks to assess system usability. The interface provides real-time guidance on pediatric advanced life support resuscitation algorithms. The test environment simulates a clinical setting, with technical equipment used for data collection and analysis.



Test Procedure

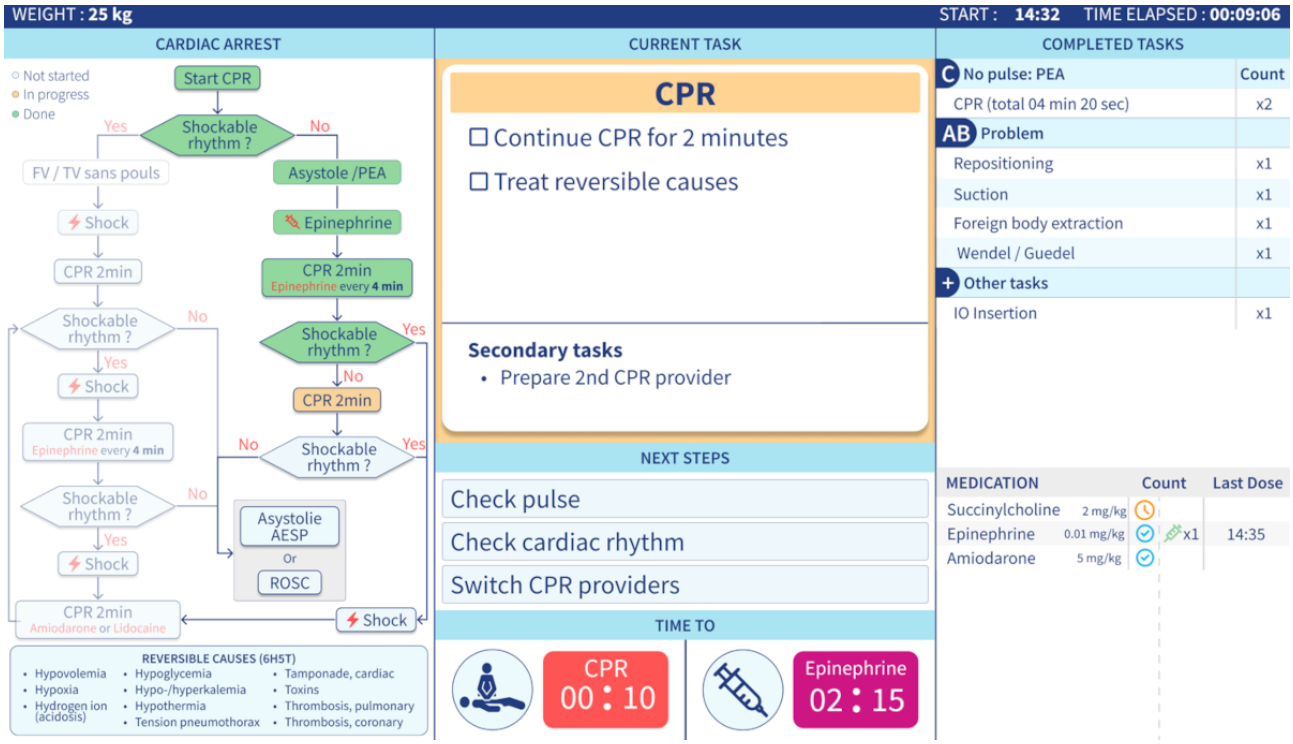
Procedure

Each session involved the same screen-based simulated pediatric resuscitation scenario, featuring a child aged 8 years (25 kg) who had sustained an electrification injury and subsequently developed a witnessed IHCA in an emergency department resuscitation bay. The initial rhythm was ventricular fibrillation/pulseless ventricular tachycardia, and the scenario was structured to follow the typical progression of a case of shockable IHCA requiring high-quality CPR, repeated rhythm checks, defibrillation, and drug administration. No manikin was used, and participants were not required to perform physical life-saving interventions; the focus was exclusively on interacting with the TeamScreen. Using TeamScreen as the shared "wall display" for the resuscitation, participants were required to complete a series of information retrieval tasks, covering key aspects of CPR resuscitation such as pulse assessment (no pulse), defibrillation, medication management (epinephrine, amiodarone, and succinylcholine as an intubation drug option), and cardiac rhythm identification (ventricular fibrillation/pulseless ventricular tachycardia). The scenario included several critical steps that had to be completed before return of spontaneous circulation (ROSC) could be achieved, such as recognizing the absence of a pulse, initiating chest compressions, performing defibrillation, and delivering the appropriate medications at the correct timing. The objective was to evaluate participants' ability to locate and interpret critical information displayed on the TeamScreen interface without prior training. The usability testing followed a standardized protocol:

1. Participant onboarding: upon arrival, participants were provided with an information sheet, and they signed an informed consent form. The study's purpose and procedure were explained, after which the participants completed a demographic questionnaire assessing their experience and expertise in pediatric resuscitation.
2. Equipment setup: participants were fitted with Tobii Pro Glasses 3 (Tobii AB) for eye-tracking analysis, and lapel microphones were used to record verbal feedback throughout the session.
3. Familiarization: clinicians who had been involved in the conception or technical development of the InterFACE-AR system were excluded. The remaining participants had not previously used the TeamScreen prototype and were exposed to it only through a short, standardized briefing immediately before the usability session. They were briefly presented with a static TeamScreen prototype (Figure 3) via FIGMA (Figma Inc) and asked to describe their observations. This phase served to ease their interaction with the interface before engaging in task execution.
4. Scenario execution and task completion: participants proceeded with the simulated IHCA scenario, performing 21 sequential information retrieval tasks (Table 1) using the think-aloud method [32]. Each task corresponded to a TeamScreen display, with additional information progressively introduced as the session advanced. These tasks were designed to assess comprehension and eye navigation on the interface. Participants were asked to verbalize their thought processes while retrieving information, guided by 1 examiner (AR), who intervened only if a participant struggled to reach a final answer. Task success, completion time, and time required for task presentation were recorded.
5. Posttest debriefing: upon completion, participants filled out the PSSUQ version 3; the responses were used to assess their satisfaction with TeamScreen. A semistructured interview was conducted to gather qualitative feedback regarding usability, strengths, and areas for improvement.
6. Closing and follow-up: participants were thanked for their time, informed about the next steps of the project, and invited to share any additional comments or suggestions.

XSL•FO
RenderX

Figure 3. TeamScreen prototype display during a simulated pediatric in-hospital cardiac arrest, showing the cardiopulmonary resuscitation (CPR) algorithm (left) and the current and upcoming tasks (center), with timer boxes for CPR cycles and epinephrine administration displayed at the bottom. The right-hand panel displays completed tasks and the medication list with weight-based dosing. Succinylcholine is included as one of several selectable intubation drugs to test navigation and selection within the interface and should not be interpreted as a recommendation for its use during ongoing cardiac arrest.



Scenario and Tasks

The IHCA scenario was standardized and tailored to each participant's professional role. Pediatric emergency physicians were assigned the role of team leader, and nurses were assigned the role of medication nurse. The scenario was co-developed by 2 pediatric emergency physicians (SM and JNS) and designed in parallel with 21 information retrieval tasks (Table 1) predefined by 3 examiners (ADM, SO, and AR), ensuring alignment between task execution and the corresponding TeamScreen displays.

Within the broader InterFACE-AR concept, the complete system is intended to provide individualized, role-specific decision support to the team leader and medication nurse through optically see-through AR headsets and to display a dynamic road map for patient care on TeamScreen, both components being coordinated via the Guiding Pad tablet app. This human factors study focuses solely on the usability of the TeamScreen app across a series of information retrieval tasks (Table 1). In the full system, weight-based medication doses are automatically calculated from the patient's weight and displayed as final numeric values on the AR headset of the medication nurse. In this TeamScreen-only prototype, dose-related panels (eg, tasks 8 and 16) primarily served to test whether participants could correctly locate where this information could be accessed rather than to verify the arithmetic itself. Similarly, the medication menu including succinylcholine (step 19) among other intubation drugs was configured to assess navigation and selection within the interface rather than as a recommendation for succinylcholine use during IHCA. An updated TeamScreen view reflecting the current implementation is provided in Multimedia Appendix 1.

Participants individually performed the 21 information retrieval tasks while progressing through a scripted walk-through simulated pediatric IHCA scenario, with clinical information and TeamScreen prototype screens presented in a structured, stepwise fashion. Each of the 20 participants completed the full scenario once, resulting in 20 standardized runs of the case.

XSL•FO
RenderX

Table . List of information retrieval tasks.

No.	Task name	Information retrieval task	Expected answer
1	Define weight I	Locate the weight entry area	Locate the weight entry area; no weight displayed[a]
2	Pulse check	Does the patient have a pulse?	Not indicated; it is in progress
3	Resuscitation time	How long has the resuscitation been going on?	15 s
4	Define weight II	What is the patient's weight?	25 kg
5	CPR[b] status	What is the status of CPR?	CPR has started; CPR was done once; CPR was performed for 10 s
6	Cardiac rhythm	What is the cardiac rhythm?	VF[c]; VF/pVT[d]
7	Shocks given	How many shocks were delivered?	1 (algorithm or completed tasks)
8	Medication status I	Can any medication be administered?	Yes (epinephrine)
9	CPR time	How much CPR time is left to do?	1 min
10	Secondary tasks	Which secondary task is proposed?	Prepare the second person for CPR
11	Next step	What is the next step?	Check the pulse; identify the cardiac rhythm; switch the CPR provider
12	A/B[e] problem assessment	Which tasks have been carried out to resolve the A/B problems?	ETT[f] intubation and repositioning
13	Intubation time	At what time did the intubation take place?	HH:MM[g]
14	Tasks in progress	Which tasks need to be done now?	Resume CPR for 2 min; administer epinephrine 0.01 mg/kg (0.1 mL/kg)
15	Shock time	At what time was the shock delivered?	HH:MM
16	Medication status II	Which medication was administered?	Epinephrine
17	Epinephrine time	How long until epinephrine must be administered again?	3 min 25 s
18	"Other tasks" section status	A team member placed an intraosseous line. Where is this information displayed?	Under "Completed Tasks"/"Other tasks"
19	Medication status	Which medication is being prepared?	Succinylcholine
20	Reversible causes	In the "Tasks in progress" area, the message "Treat the reversible causes" appeared. Where can you find help to recall the list of reversible causes?	Below the algorithm
21	Patient status: general understanding	What is the patient's current condition?	No longer in cardiac arrest; moved to ROSC[h]; ABCDE[i] assessment

[a]Success was defined as correctly identifying the relevant interface element rather than as giving a numeric answer.

[b]CPR: cardiopulmonary resuscitation.

[c]VF: ventricular fibrillation.

[d]pVT: pulseless ventricular tachycardia.

[e]A/B: airway and breathing.

[f]ETT: endotracheal tube.

[g]HH:MM: hours and minutes.

[h]ROSC: return of spontaneous circulation.

[i]ABCDE: airway, breathing, circulation, disability, exposure.

XSL•FO
RenderX

Outcome Measures

Quantitative Data

Performance Metrics

Participant performance was evaluated using 2 key efficiency metrics:

- Task success rate: tasks were classified as successful if the participant provided the expected answer (Table 1), either immediately or after a correction, as long as the final response was correct. Incorrect or incomplete responses were categorized as task failures.
- Time on task: This metric measured the average time (in seconds) taken to complete each task, starting from the end of the examiner's instructions until the participant provided a final answer (Table 1). Since participants had no prior training with TeamScreen, response time was not used as a determinant of task success.

User Satisfaction

User satisfaction was measured after completion of the tasks using the PSSUQ version 3 [28], a 16-item instrument rated on a 7-point Likert scale (Multimedia Appendix 2), rated from 1 ("strongly agree") to 7 ("strongly disagree"), in addition to a "not applicable" (N/A) option. Items 7 and 8, related to error messages and error recovery, were not applicable in this study because the TeamScreen prototype did not implement system error messaging or recovery workflows. These items were therefore marked "not applicable" and omitted from score calculations, in accordance with published PSSUQ scoring guidance [28,33], with overall scores computed as the mean of the remaining answered items.

The PSSUQ yields an overall usability score and scores on 3 subscales—system usefulness (items 1 - 6), information quality (items 7 - 12), and interface quality (items 13 - 15)—with lower scores indicating greater satisfaction in each category. Based on 21 studies including 210 participants, the benchmark means reported by Lewis [34] provided the reference values for interpreting the PSSUQ scores: overall 2.82, system usefulness 2.80, information quality 3.02, and interface quality 2.49. This comparison with published benchmark values was purely descriptive. No inferential statistical testing was performed against these reference data. Better performance and satisfaction are reflected in lower PSSUQ scores (closer to 1).

Qualitative Data

Error Analysis

All usability challenges encountered during task completion were systematically recorded and categorized using ergonomic criteria for human-computer interface evaluation [35]. Each usability issue was assigned a severity rating based on Nielsen Norman Group's rating scale [36], assessing its impact on user performance and experience. Severity level 3 indicates a major usability problem that significantly hinders task execution, forcing users to find alternative solutions and requiring urgent corrective action. A severity level 2 represents a minor usability issue that may cause frustration but does not prevent task completion, warranting moderate priority for resolution. A severity level 1 is considered a cosmetic issue, affecting only aesthetics, with no impact on usability, requiring resolution only if time allows. Finally, a severity level of 0 is considered as not usability related. Findings from the severity assessment informed targeted design refinements for TeamScreen's next iteration. The identified issues were addressed in a recently conducted multicenter simulation study, for which a manuscript is currently in preparation.

Think-Aloud Verbal Feedback

During testing, participants were encouraged to verbalize their thought processes while interacting with TeamScreen. These verbal reflections were analyzed to identify usability issues, inform content and design refinements, and improve system design for better real-world applicability.

Posttest Survey Feedback

To complement quantitative findings, qualitative feedback was collected via a standardized set of posttest interview questions (Table 2), ensuring consistency in data collection.

Responses provided a holistic understanding of user experience. Findings were systematically analyzed to refine TeamScreen's interface and functionality, ensuring alignment with user expectations and clinical workflows.

XSL•FO
RenderX

Table . Posttest survey.

Question	Type of question	Choice provided
Display understanding		
1. What is your overall impression of the display?	Open-ended question	—[a]
2. What did you find easy to understand in this display?	Open-ended question	—
3. What did you find difficult to understand in this display?	Open-ended question	—
4. What do you think of the readability of the information (eg, font size, colors, etc)?	Open-ended question	—
5. If you had to rearrange the entire TeamScreen display, what ideas or suggestions for improvement would you propose?	Open-ended question	—
Completed tasks		
6. What difficulties did you encounter while using the display, and why?	Open-ended question	—
7. What aspects of the display helped you complete the tasks (information retrieval)?	Open-ended question	—
8. What aspects of the display hindered you from completing the tasks (information retrieval)?	Open-ended question	—
Display preferences		
9. What do you think of the distribution of information in the “Algorithm” section?	Open-ended question	—
10. What do you think of the distribution of information in the “Current Task” section?	Open-ended question	—
11. To what extent did you find the display of secondary tasks helpful regarding patient status?	Likert-scale question	• Very useful • Useful • Neutral • Somewhat useful • Not useful at all
12. What do you think of the distribution of information in the “Next Steps” section?	Open-ended question	—
12a. How many “next steps” would you ideally want to see? One, two, three, or more?	Single-choice question	• One step • Two steps • Three steps • Other: _____
13. What do you think of the distribution of information in the “Countdown” section?	Open-ended question	—
14. What do you think of the distribution of information in the “Completed Tasks” section?	Open-ended question	—
15. What do you think of the distribution of information in the “Medications” section?	Open-ended question	—
15a. What do you think these three icons represent in the “Medications” section?	Closed-ended question	—

[a]Not applicable.

Data Collection and Analysis

Data Collection and Confidentiality

To ensure data confidentiality, all collected materials, including recordings, online questionnaires, and consent documents, were coded. Recordings captured only participants’ screen view, eye movements, and voice, ensuring privacy. When verbatim quotes were used in qualitative data analysis or reporting, only the assigned identifier was displayed. Each usability testing session was video- and audio-recorded for retrospective analysis of TeamScreen’s usability. Two researchers (SO and ADM) independently reviewed task success rates and completion times.

XSL•FO
RenderX

All PSSUQ scores were collected immediately after each test session and transcribed into Microsoft Excel spreadsheets for analysis. Half of the video recordings were independently assessed by SO and ADM to validate task success rates and completion times. Quantitative data collected during task execution underwent statistical analysis, and results were anonymized for reporting (see the *Results* section). During data collection and analysis, only 3 researchers (AR, SO, and ADM) had access to the audio and video recordings.

Data Analysis

Descriptive statistics were used to summarize continuous variables, including task success rates, time-on-task metrics, and PSSUQ scores. Frequency counts and percentages were used to analyze categorical variables, such as the types of usability errors. Additionally, a correlation analysis of task failure rates was performed to explore potential dependencies between tasks, indicating whether tasks shared common cognitive demands or interface elements. Quantitative data, including task durations and success rates, were analyzed using the *pandas* library in Python (version 3.12; Python Software Foundation) to identify trends and compute averages. Verbal feedback from the think-aloud sessions was audio-recorded, transcribed verbatim, and analyzed using a structured content analysis approach. Two researchers (AR and SO) independently reviewed the transcripts, segmented them into meaning units corresponding to distinct interface problems, and coded each unit using Bastien and Scapin's [35] ergonomic criteria for human-computer interaction (guidance, workload, explicit Ccntrol, adaptability, error management, consistency, significance of codes, and compatibility). These criteria were chosen because they provide a well-established, task-oriented framework for classifying usability issues in interactive systems [37] and have been widely used in the evaluation of medical and health information technology interfaces [26,38]. Coded segments were then compared between both researchers. Discrepancies were resolved through discussion, and similar issues were grouped into overarching usability problems, which were mapped to specific interface sections (header, algorithm, completed tasks, medications). For each distinct usability issue, we calculated its frequency (number of participants who mentioned it) and assigned a severity rating using Nielsen Norman Group's rating scale ranging from 0 to 4 (0=none, 4=catastrophic). Graphical representations were generated using *matplotlib*, an open-source Python library [39].

Results

Participant Characteristics

The usability tests were conducted in mid-September 2024. A total of 20 health care professionals participated in the study, of which 14 (70%) were women. The sample included 15 nurses (75%) and 5 physicians (25%). Of the 20 participants, 10 (50%) had completed PALS certification, including 3 of 5 physicians (60%) and 7 of 15 nurses (46.67%). Notably, 8 of 15 nurses (53.33%) had no PALS certification. The participants had a median experience of 6.00 (IQR 2.00-11.75) years in pediatric emergency care. All had prior exposure to real pediatric resuscitations, with the number of cases ranging from 2 to 100 and 30 cases being the most frequently reported value. Similarly, all had experience with simulated resuscitations, with the reported number of cases ranging widely, most commonly around 20 simulations.

Participants' self-reported familiarity with digital devices varied: of 20 participants, 10 (50%) described themselves as "comfortable" using tablets and smartphones, whereas 6 (30%) reported being "very comfortable." The remaining participants were either "neutral" (3/20, 15%) or "slightly uncomfortable" (1/20, 5%) with digital tools.

Usability Evaluation

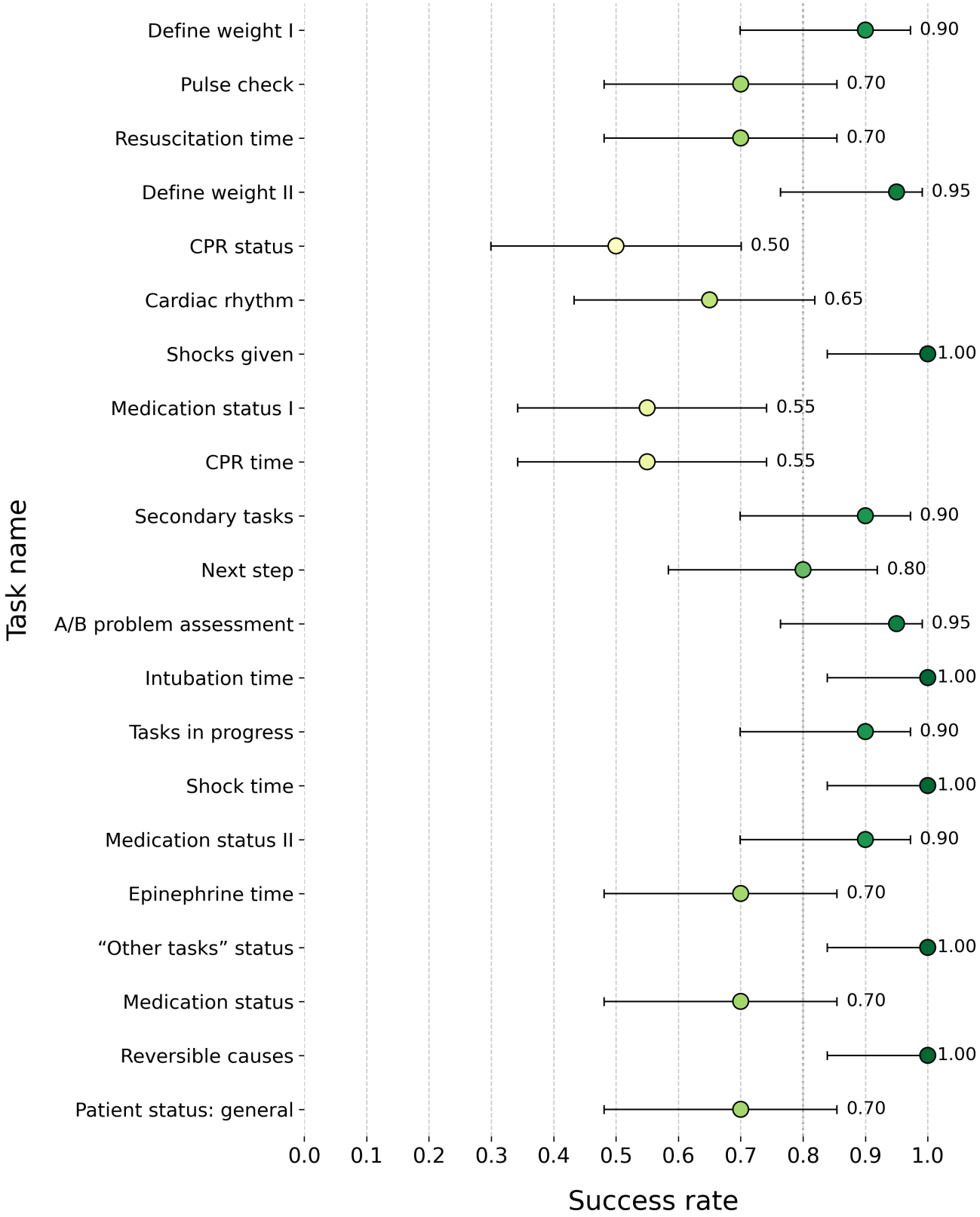
Quantitative Evaluation

Task Success Rate

Task success rates varied significantly across the usability assessment (χ^2_{20}=74.59, *P*<.001; Figure 4), ranging from 50% to 100%, with an overall success rate of 81.19% (SD 16.87%). Highest success rates (100%) were achieved for tasks 7 (*Shocks given*), 13 (*Intubation time*), 15 (*Shock time*), 18 (*"Other tasks" section status*), and 20 (*Reversible causes*), indicating that they were consistently completed without errors. The lowest success rates were achieved for task 5 (*CPR status*; 50%), followed by tasks 8 and 9 (55% each), suggesting that these tasks were more cognitively demanding or required greater user adaptation.

XSL•FO
RenderX

Figure 4. Mean (95% CI) success rate by information retrieval task. The color gradient indicates the task success rate, ranging from light yellow (lower success rates) to dark green (higher success rates). A/B, airway and breathing; CPR: cardiopulmonary resuscitation.

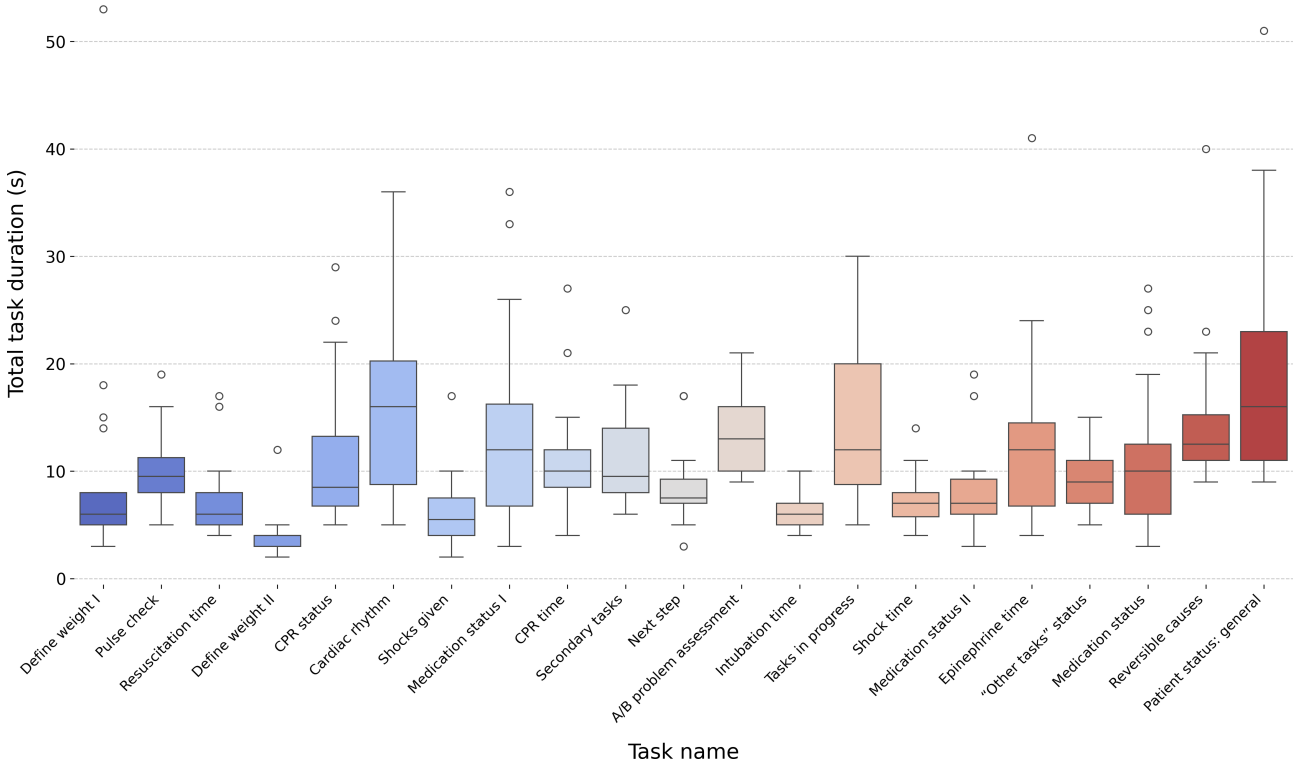


Time on Task

Task completion times varied significantly (Kruskal-Wallis H_{20}=154.99, P<.001; Figure 5), reflecting differences in task complexity. On average, participants spent 170.80 (SD 44.70) seconds completing all 21 tasks, corresponding to a mean completion time of 8.13 (SD 7.07) seconds per task across all participants. Task 21 (*Patient status: general understanding*) required the longest duration, with an average completion time of 18.45 (SD 12.19) seconds, whereas task 4 (*Define weight II*) was completed the fastest, averaging 2.35 (SD 1.90) seconds.

XSL•FO
RenderX

Figure 5. Box plot illustrating the distribution of task completion times across all 21 tasks performed during the usability study. Each box represents the IQR, with the median completion time indicated by the horizontal line inside the box. Whiskers extend to 1.5 times the IQR, and outliers are displayed as individual circles. Task 21 (Patient status: general understanding) exhibited the longest mean duration, suggesting higher complexity or additional steps required for completion, while task 4 (Define weight II) had the shortest duration, indicating a simpler information retrieval process. A/B: airway and breathing; CPR: cardiopulmonary resuscitation.

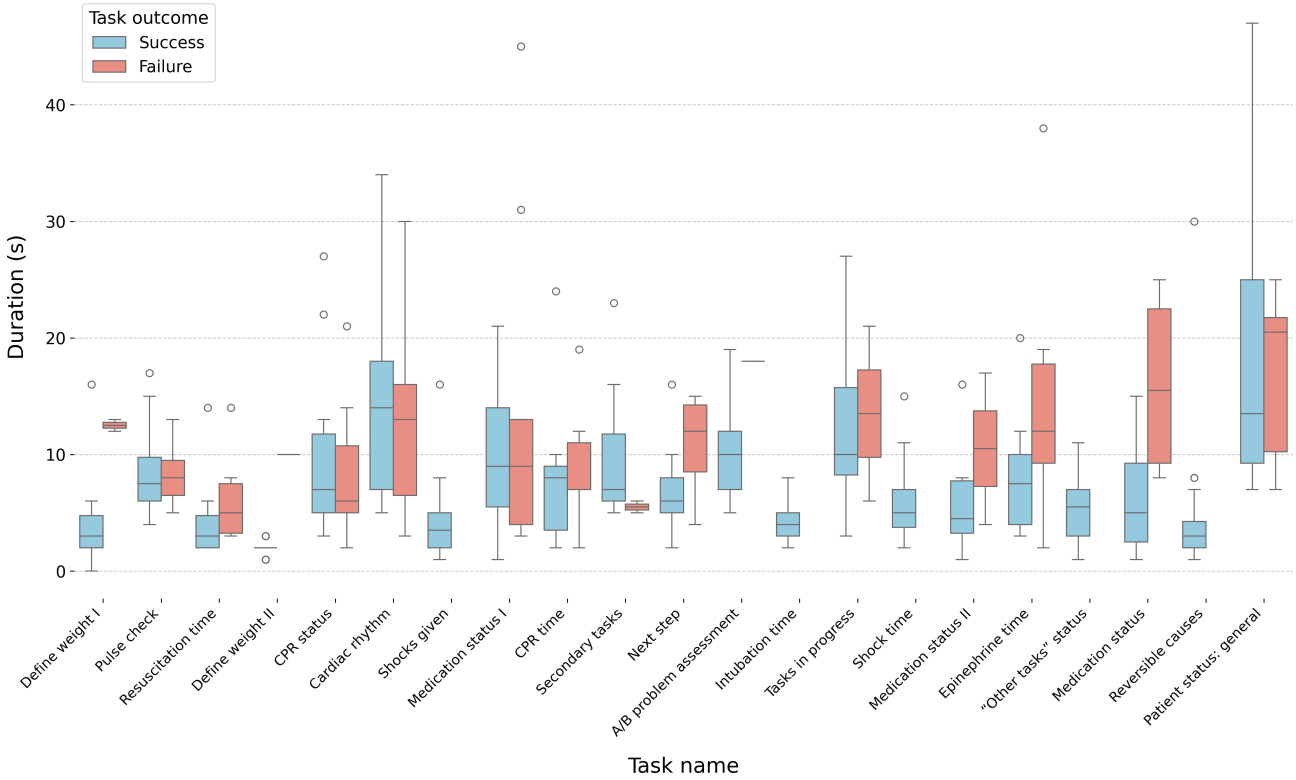


Average Task Duration by Success Versus Failure

Analysis of task completion times based on success or failure revealed a notable trend (Figure 6). In most cases, failed attempts were associated with significantly longer durations, suggesting that unsuccessful users required multiple attempts or encountered execution challenges. For example, in task 1 (*Define weight I*), failed attempts averaged 12.50 (SD 0.71) seconds, whereas successful attempts took only 3.78 (SD 3.42) seconds (P=.048), highlighting the impact of task difficulty on completion time. However, in task 5 (*CPR status*), the completion times for successful attempts (mean 10.00, SD 8.25 seconds) and failed attempts (mean 8.30, SD 5.81 seconds) were nearly identical (P=.73), indicating that participants invested consistent effort regardless of outcome.

XSL•FO
RenderX

Figure 6. Box plot comparing task completion times between successful (blue) and failed (red) attempts across all 21 tasks. Each box represents the IQR, with the median duration indicated by the horizontal line. Whiskers extend to 1.5 times the IQR, and outliers are shown as individual circles. Overall, failed attempts generally took longer to complete, indicating challenges in task execution or repeated attempts (eg, task 1 [*Define weight I*]: failures averaged 12.5 s vs 3.77 s for successful attempts). However, for task 5 (*CPR status*), success and failure durations were nearly identical, suggesting that participants required similar effort regardless of the outcome. A/B: airway and breathing; CPR: cardiopulmonary resuscitation.

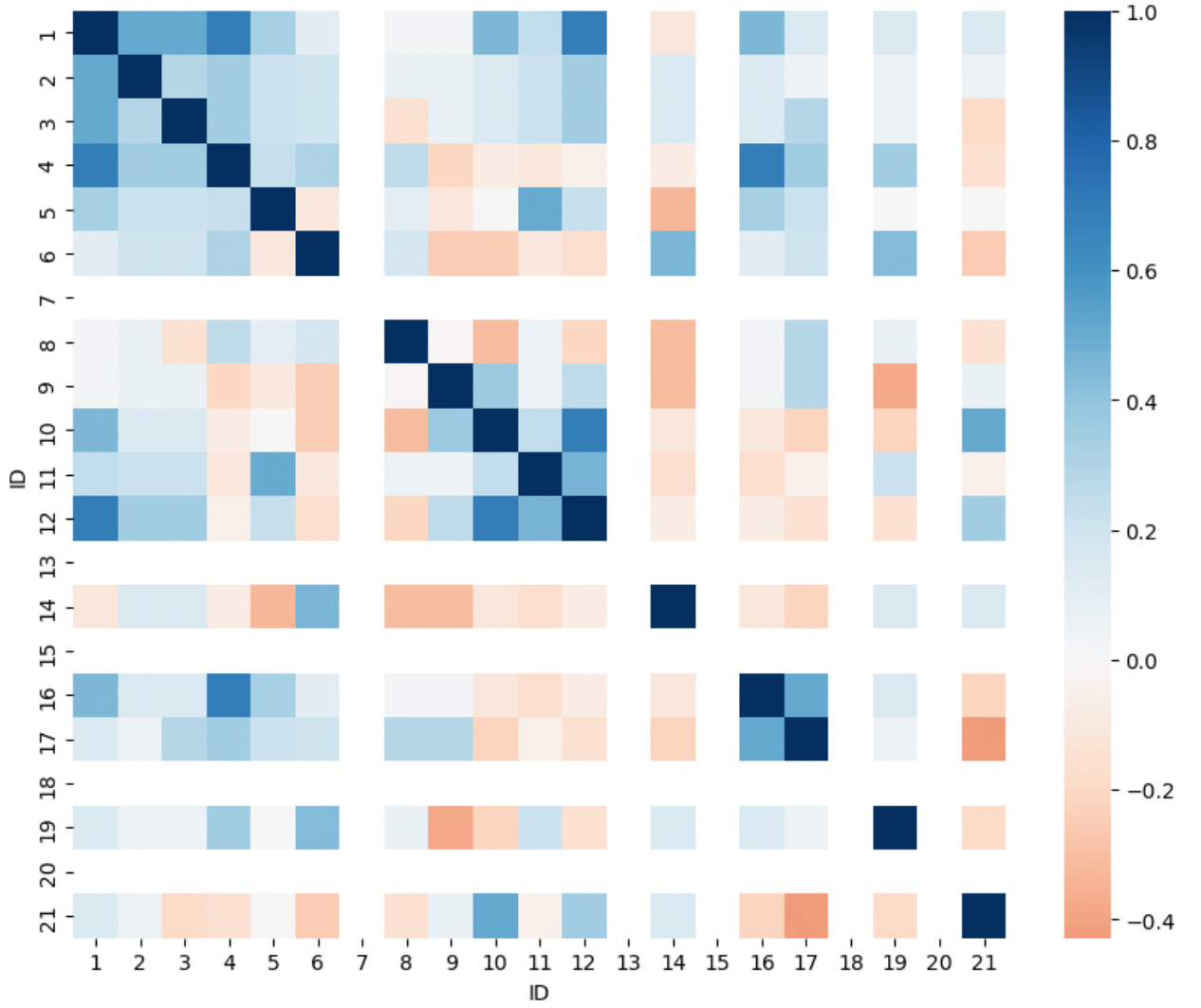


Correlation of Failure Rates Among Tasks

A correlation analysis of failure rates across tasks revealed potential dependencies, indicating that certain tasks share underlying cognitive demands or interface elements (Figure 7). Tasks with similar failure rates suggest common user difficulties, likely due to overlapping knowledge requirements or proximity within the interface. Several tasks exhibited strong positive correlations, suggesting that errors in one task were predictive of errors in another. For example, tasks 1 (*Define weight I*) and 4 (*Define weight II*) were highly correlated with tasks 12 (*A/B problem assessment*) and 16 (*Medication status II*) (Pearson correlation of per-participant binary success: r=0.69, P<.001), as all of these required users to track updated patient data, such as weight changes, completed airway and breathing interventions, and medication details. Likewise, task 4 (*Define weight II*) and task 16 (*Medication status II*) showed a strong correlation (r=0.69, P<.001), likely reflecting the dependency between understanding the patient's weight and correctly administering weight-based medication. Task 10 (*Secondary tasks*) was significantly correlated with task 12 (*A/B problem assessment*; r=0.69, P<.001), suggesting that both tasks relied on the ability to monitor procedure logs or timeline view. Finally, tasks 16 (*Medication status II*) and 17 (*Epinephrine time*) exhibited a strong correlation (r=0.51, P=.02), as both involved understanding epinephrine administration; participants who struggled with one were likely to struggle with the other, highlighting potential cognitive or interface-related dependencies. Overall, these suggest that shared interface elements or related data points contributed to error clustering, emphasizing the need for enhanced visual cues or streamlined navigation to support information retrieval across interdependent tasks.

XSL•FO
RenderX

Figure 7. Heat map visualizing the correlation of failure rates across 21 usability tasks. Darker blue regions indicate strong positive correlations. Conversely, light red areas represent negative correlations, where failure in one task was less predictive of failure in another.



Qualitative Evaluation

Frequent Errors

Analysis of task failures showed that task 5 (*CPR status*) had the highest failure frequency, accounting for 12.50% (10/80) of failed attempts. This was closely followed by tasks 8 (*Medication status I*) and 9 (*CPR time*) (9/80 each, 11.25%). Task 6 (*Cardiac rhythm*) accounted for 10% (8/80) of failures, and tasks 2, 3, 17, and 19 each accounted for 7.50% (6/80) of failures. These trends indicate that certain tasks posed consistent challenges across participants, potentially due to interface complexity, cognitive load, or unclear task presentation. Table 3 provides a detailed breakdown of failed task occurrences across the 21 tasks, considering 80 total opportunities across all tasks.

XSL•FO
RenderX

Table . Task failure frequency (N[a]=80).

Task number	Content	Failure frequency, n[b] (%)
5	What is the status of CPR[c]?	10 (12.50)
8	Can any medication be administered?	9 (11.25)
9	How much CPR time is left to do?	9 (11.25)
6	What is the cardiac rhythm?	8 (10)
17	How long until epinephrine must be administered again?	6 (7.50)
2	Does the patient have a pulse?	6 (7.50)
19	Which medication is being prepared?	6 (7.50)
3	How long has the resuscitation been going on?	6 (7.50)
21	What is the patient's current condition?	5 (6.25)
11	What is the next step?	4 (5)
16	Which medication was administered?	3 (3.75)
14	Which tasks need to be done now?	2 (2.50)
1	What is the patient's weight?	2 (2.50)
10	Which secondary task is proposed to you?	2 (2.50)
12	Which tasks have been performed to address the A/B[d] problems?	1 (1.25)
4	What is the patient's weight?	1 (1.25)

[a]N: total number of opportunities.

[b]n: number of failed attempts.

[c]CPR: cardiopulmonary resuscitation.

[d]A/B: airway and breathing.

Encountered Usability Issues

A total of 16 usability issues were identified across different sections of the TeamScreen interface. These issues were analyzed using the heuristics of Bastien and Scapin [35] and classified by severity levels according to the Nielsen Norman Group's rating scale [36] (Table 4).

XSL•FO
RenderX

Table . Classification of usability issues identified by participants (N=20).

Usability issues (description and section)	Frequency, n	Bastien and Scapin's heuristics	Problem severity[a]
General issues			
Excessive information displayed: multiple panels (eg, algorithm, completed tasks, and medications) with many elements visible at once made it difficult to identify the most relevant information for the current resuscitation step (general screen layout).	10	*Workload*: brevity, information density	3
Confusing time notation of CPR[b] *timer:* the CPR timer label did not clearly indicate whether the value represents total arrest duration, current cycle length, or time since last intervention, leading to misinterpretation of the timer.	2	*Significance of codes*	2
Header section			
Location visibility of patient weight: patient weight was located in the upper banner but not visually highlighted; several participants reported difficulty locating it quickly when asked to verify the weight.	5	*Guidance*: legibility	2
Font size of patient weight: the font size of patient weight in the header was too small to be read comfortably from the standard viewing distance of 2.7 m.	3	*Guidance*: legibility	1
Lack of current time display: the current clock time was not displayed anywhere on TeamScreen, forcing users to rely on external clocks to track absolute time during the scenario.	1	*Adaptability*	2
Algorithm section			
Poor legend visibility: the legend explaining the algorithm and color codes was small, making it difficult for users to remember or quickly decode the meaning of icons or colors.	4	*Guidance*: legibility	2
Difficulty understanding transitions between algorithm branches: transitions between algorithm branches (for example, from shockable to nonshockable pathways or post-ROSC[c] care) caused uncertainty about which branch to follow next.	6	*Guidance*: prompting	3
Visibility of patient status: patient's current state (eg, "cardiac arrest") was not clearly highlighted near the active algorithm branch, which reduced situational awareness.	3	*Compatibility*	3

XSL•FO
RenderX

Usability issues (description and section)	Frequency, n	Bastien and Scapin's heuristics	Problem severity[a]
Font size of 6H5T[d] text: the text listing reversible causes (6H5T) was presented in a small font and dense format, which participants found hard to read and use during the scenario.	3	*Guidance*: legibility	1
Completed tasks section			
Overloaded with information: the list of completed tasks showed all previous actions with detailed timestamps, resulting in a visually dense list that was difficult to scan for the most recent or most important actions.	5	*Workload*: brevity, information density	3
Poor organization of task display in sections C[e], A/B[f]	3	*Compatibility*	2
Insufficient separation between subsections	3	*Guidance*: legibility	2
Counterintuitive time notation	3	*Significance of codes*	2
Medications section			
Ambiguous icons: icons used to represent medication-related actions (ie, in preparation, ready to administer, and administered) were not self-explanatory without accompanying text labels, potentially leading to misinterpretation of their meaning.	15	*Significance of codes*	3
Missing calculated total doses/volumes: the medications panel did not display the final calculated total dose or volume for each drug, requiring users to mentally calculate weight-based dosage, which increased cognitive load and potential for dosing error.	5	*Compatibility*	2
Unclear listed medication doses	3	*Compatibility*	2

[a]Nielsen Norman severity score: 4=catastrophic, 3=major, 2= minor, 1=cosmetic, 0=none.

[b]CPR: cardiopulmonary resuscitation.

[c]ROSC: return of spontaneous circulation.

[d]6H5T: hypovolemia, hypoxia, hydrogen ion (acidosis), hypokalemia/hyperkalemia, hypoglycemia, hypothermia, tension pneumothorax, tamponade, toxins, thrombosis (pulmonary), thrombosis (coronary).

[e]C: circulation.

[f]A/B: airway and breathing.

The primary usability issues were general challenges rather than being confined to specific interface sections. These included cognitive overload due to excessive information displayed information; difficulty understanding task requirements, increasing mental workload; and medication-related confusion, particularly in interpreting drug administration icons. The most frequently reported issue was the lack of clarity in medication icons, making it difficult for participants to distinguish between medications in preparation, ready for administration, or already administered. Other issues, although present, were encountered by fewer than one-third of participants.

Overall, the study identified key usability challenges related to information visibility, navigation, and comprehension. The most critical usability issues (severity 3 - 4) included medication icon confusion (most frequently reported), excessive information density leading to cognitive overload.

User Satisfaction

PSSUQ Scores

Participants provided positive feedback on screen readability, with a mean PSSUQ score of 2.40 (SD 1.24) indicating high usability. However, they suggested improvements in information hierarchy and task simplification due to perceived cognitive

XSL•FO
RenderX

overload. Overall, PSSUQ scores were lower (better usability) than the PSSUQ version 3 benchmark (Table 5), confirming high system usefulness and interface quality.

Scores closer to 1 indicate higher usability. TeamScreen scores were descriptively lower than the published PSSUQ benchmark means, suggesting that participants found the system intuitive and effective, although information presentation still offers room for optimization.

Table . Comparison of TeamScreen PSSUQ[a] scores versus the PSSUQ version 3 (V3) benchmark.

	TeamScreen score, mean (SD)	PSSUQ V3 benchmark mean
Overall score	2.40 (1.24)	2.82
System usefulness (SysUse)	2.44 (1.22)	2.80
Information quality (InfoQual)	2.41 (1.20)	3.02
Interface quality (IntQual)	2.20 (1.41)	2.49

[a]PSSUQ: Post-Study System Usability Questionnaire.

Participant Verbal Feedback

Think-aloud verbalizations during task execution, combined with responses from the posttask debriefing survey, provided valuable contextual information that enriched the interpretation of the PSSUQ scores and offered deeper insights into users' experiences and expectations regarding the interface.

Overall, participants expressed mixed feelings about their experience with the TeamScreen interface. In terms of strengths, 8 of 20 (40%) participants highlighted the clarity and organization of the information, quality of the display, and good readability (in terms of font size and color). Several participants particularly appreciated the guidance provided through the algorithm section (12/20, 60%) and the associated color coding (8/20, 40%)—green indicating "finished" and orange indicating "in progress"—as well as the "Completed Tasks" section (9/20, 45%), which enhanced their situational awareness.

Regarding areas for improvement, 9 of 20 (45%) participants reported misunderstanding difficulties, particularly related to the middle section of the screen, not directly attributable to usability issues. Several users (10/20, 50%) reported that the amount of information presented was overwhelming, contributing to increased cognitive load. The "Current Task" area was also frequently overlooked, as 9 of 20 (45%) participants felt that the relevant information was already accessible within the algorithm panel on the left side of the screen, making it redundant. Furthermore, several users suggested making a clearer distinction between "Current Task" and "Next Steps" to better support cognitive processing under pressure. These insights emphasize the need for improvements in information hierarchy and cognitive load reduction in critical usage scenarios. Finally, participants indicated a need for greater clarity in the "Medication section" (5/20, 25%), specifically concerning the meaning of the 3 icons representing the medication status (in preparation, ready for administration, administered). However, it is important to acknowledge that many of these reported issues were, according to the participants, largely attributable to their lack of prior exposure to the app. Participants indicated that with appropriate training, the interface would become significantly more intuitive, enabling faster and more accurate information retrieval during real-time pediatric cardiac arrest scenarios, as well as more effective adherence to American Heart Association guidelines.

Modifications Based on Usability Findings

Following the usability analysis, several targeted refinements were implemented, each addressing a specific usability issue summarized in Table 4. First, limited visibility and ambiguous interpretation of the CPR timer led us to relocate the main timer to the top of the screen and clarify its label to ensure continuous and unambiguous visibility throughout the resuscitation process. Second, difficulties in reading and using the reversible causes (6H5T; hypovolemia, hypoxia, hydrogen ion [acidosis], hypokalemia/hyperkalemia, hypoglycemia, hypothermia, tension pneumothorax, tamponade, toxins, thrombosis [pulmonary], thrombosis [coronary]) prompted an increase in font size and a permanent display to provide immediate access to correction measures. Third, issues related to information overload and poor readability of the list of completed tasks motivated a restyling of the task list, replacing square markers with bullet points, and secondary tasks were repositioned higher in the interface to make them more noticeable and actionable. Fourth, in response to concerns that timers were easy to overlook as they approached zero, countdown timers were enhanced to blink when nearing expiry, drawing the team's attention and reducing the risk of delayed interventions. Finally, frequent misinterpretation of medication icons and confusion around drug administration stages led to the introduction of 3 distinct, clearly labeled icons indicating whether a drug was in preparation, prepared, or already administered. Collectively, these modifications aimed to alleviate information overload, improve navigation, and highlight essential cues in high - stress pediatric resuscitation scenarios.

Discussion

Key Findings

TeamScreen introduces a new generation of context-aware clinical displays aimed at reinforcing shared cognition and temporal coordination among resuscitation teams rather than replacing clinical judgment. This single-center, mixed methods study evaluating its usability during simulated IHCA management yielded several salient insights.

With an overall task success rate of 81%, most participants effectively located and interpreted essential clinical information under simulated urgency, despite challenges in early pediatric

XSL•FO
RenderX

IHCA response [40]. Tasks 7 (*Shocks given*), 13 (*Intubation time*), 15 (*Shock time*), 18 (*"Other tasks" section status*), and 20 (*Reversible causes*), covering shock delivery logs, intubation timestamps, and reversible - causes guidance, achieved 100% success, suggesting that these interface elements were well designed and intuitive. Conversely, tasks 5 (*CPR status*), 8 (*Medication status I*), and 9 (*CPR time*) related to tracking ongoing CPR and upcoming interventions (eg, epinephrine administration) exhibited the highest failure rates. Participants who struggled with these tasks frequently cited uncertainty regarding the location of real-time information, highlighting a need for improved visual hierarchy in high-density interface areas, in line with cognitive load theory [41].

Task completion times demonstrated substantial variability across the 21 tasks, reflecting heterogeneity in task complexity and cognitive processing demands. Task 21 (*Patient status: general understanding*) exhibited the longest mean completion time, indicating that tasks involving a broader synthesis of patient information were more time-consuming. Conversely, task 4 (*Define weight II*) was completed most rapidly, suggesting minimal cognitive load and straightforward retrieval. This efficiency may also be attributed to task 4 representing the second occurrence of a previous task, that is, task 1 (*Define weight I*), whereby participants were already familiar with the location of the relevant information on the display.

Correlation analysis showed errors clustering around related data points, such as medication timing and dosage, underscoring the importance of consistent navigation cues and harmonized data presentation across the algorithm view, timeline, and "Completed Tasks" sections. This aligns with evidence that protocol adherence significantly improves IHCA outcomes [42], emphasizing TeamScreen's potential to enhance care quality through better design. The PSSUQ scores indicated high user satisfaction, with participants valuing TeamScreen's consolidation of vital parameters, current tasks, and next - step recommendations into a central, readable display. These findings suggest that TeamScreen excels where clarity is prioritized (eg, timestamps and shock logs) but requires refinement in areas with visual clutter, icon ambiguity, or subtle task distinctions. Given varying roles in IHCA response where nurses often manage documentation and physicians lead decisions [43], future studies should explore whether task success or perceived usability differs across these groups.

Comparison With Literature

TeamScreen shares similarities with real-time dashboard systems aggregating patient vitals, medication records, and intervention timelines onto shared displays, as seen in prior work improving clinicians' access to critical data and reducing cognitive load [44]. However, unlike passive aggregators, TeamScreen integrates real-time visualization with interactive PALS-aligned algorithmic guidance, supporting both team situational awareness and decision-making during IHCA. This dual functionality addresses a gap in traditional dashboards, which often lack structured support for resuscitation guidelines. For instance, Siebert et al [45] demonstrated that AR glasses improved defibrillation dosing adherence compared to PALS pocket cards, although without broader workflow integration. A recent systematic review by Cheng et al [46] highlights the potential of AR and virtual reality to enhance resuscitation skills but notes inconsistent superiority over traditional methods, underscoring the need for integrated solutions such as TeamScreen. TeamScreen extends this concept by offering a centralized, algorithm-synchronized interface, positioning it within an emerging class of digital tools designed to optimize shared cognition and temporal alignment in high-acuity care.

Essential TeamScreen Enhancements

The usability evaluation of TeamScreen has identified critical areas for future refinement, and the following enhancements will be implemented to optimize its support for resuscitation teams. The information hierarchy will be restructured to minimize visual overload, consolidating related elements such as medication timing and dosage into a distinct, clearly labeled panel, guided by cognitive load principles [41]. This adjustment could streamline data retrieval and lessen cognitive strain during time-sensitive tasks. Dynamic visual cues, such as color shifts or blinking countdowns for medication intervals, could prompt timely actions, provided they adhere to ergonomic principles to avoid distraction in high - stress settings. Functionally, adaptive role-based prompts tied to PALS algorithms could guide users step-by-step, improving coordination and adherence. Standardizing labels, icons, and visual styles (eg, font sizes, colors, and legends) would enhance consistency and reduce the likelihood of misinterpretation, particularly for the 6H5T mnemonic and streamlined task lists. Finally, given its intuitive design, brief, role - specific onboarding is advised—especially for users with varying digital fluency—to ensure efficient navigation in multitasking, high-stakes scenarios such as CPR. Paired with these design refinements and enhancements, targeted training could maximize TeamScreen's potential to support resuscitation teams effectively.

Study Limitations

This study has several limitations. Testing TeamScreen in a simulated environment without direct user interaction limits conclusions about real-world PALS algorithms adherence or patient outcomes. The heightened stress of a real-life IHCA situation might exacerbate usability challenges not fully captured in this study. Additionally, evaluating TeamScreen isolation (ie, without AR Headsets or Guiding Pad) potentially reduces ecological validity, as the full InterFACE-AR suite's interplay was not assessed. This may also not fully capture the usability of the complete system. However, the Guiding Pad and AR headset modules have been evaluated in separate component-level usability studies, and the integrated InterFACE-AR system has recently been tested in a multicenter randomized controlled trial in a high-fidelity simulated setting. Third, the passive display design, reliant on a charting nurse's input via the Guiding Pad, aligns with its intended use but restricted participants' ability to explore the interface freely, potentially skewing engagement. Fourth, the study sample size and various clinical backgrounds of participants further limit generalizability. Fifth, our evaluation was supposedly conducted in a dedicated emergency department resuscitation bay using a fixed, wall-mounted TeamScreen display, which may limit generalizability to ward-based pediatric IHCA where mobile

XSL•FO
RenderX

configurations of TeamScreen and the Guiding Pad would be required and should be evaluated in future work. Finally, while cognitive overload was frequently noted in feedback, this dimension remained unquantified owing to the absence of a validated instrument such as the NASA-Task Load Index (NASA-TLX) [47]. This should be addressed in future studies.

Future Directions

Several of the usability issues identified in this study point to general design principles that are relevant for other digital tools in acute care. Information that is crucial for safe care (such as timers, current patient status, and key tasks) should be clearly visible on a shared team display. The amount of information on screen needs to be limited and well structured to avoid overload, and icons or color codes must be immediately understandable, even under stress. These lessons can guide the design of other real-time decision support systems that aim to support teamwork rather than only individual users.

The next phase of our program consists of 2 multicenter, simulation randomized controlled trials in 2025. The first, conducted in the second quarter of 2025, has now been completed, and the second, initiated in the fourth quarter of 2025, is currently underway, integrating TeamScreen with AR Headsets and Guiding Pad to evaluate clinical decision‐making, teamwork, communication, and patient safety outcomes across diverse hospital settings. Structured, role-specific onboarding will mitigate the learning curve effects observed in this study, offering a clearer performance baseline under realistic pressures. Because adherence to advanced life support guidelines is associated with better IHCA outcomes [42], metrics such as drug administration timing, error rates, and PALS adherence will provide deeper insights into how digital support influences care quality. Role differences (eg, nurses' documentation vs physicians' leadership [43]) will also be examined to further adapt the interfaces. Cognitive workload will be quantified using tools such as the NASA-TLX [47] to address qualitative feedback about overload and to test whether the system genuinely reduces cognitive burden.

Future iterations could incorporate artificial intelligence–driven predictive analytics to anticipate tasks (eg, rhythm reassessment), further reducing manual tracking demands. Qualitative feedback and high usability scores suggest that TeamScreen may alleviate cognitive load, although formal workload assessments are needed. Future work also needs to address more practical questions that apply to many medical digital tools, including feasibility of deployment, interoperability with existing hospital systems, and implementation costs, for widespread adoption. Integration with electronic health records, automated monitoring data capture, and tailored role-specific interfaces represent key translational goals to support clinical use of InterFACE-AR and similar decision support systems.

Conclusion

This study provides a structured usability evaluation of TeamScreen, a shared digital interface designed to assist pediatric resuscitation teams during IHCA, with its clinical impact still to be validated. Using a mixed methods approach, we identified its strengths and limitations in a simulated setting. Results revealed an 81% task success rate and high user satisfaction, confirming TeamScreen's ability to deliver accessible, interpretable clinical data under time pressure. However, issues with information hierarchy and task clarity, evidenced by correlated failures in navigation and updates, highlighted areas for refinement. The combined analyses of error categorization, verbal feedback, and questionnaire responses provided insights about the prototype's strengths and areas requiring refinement. This comprehensive evaluation not only identified critical usability challenges but also validated the potential of the TeamScreen prototype to optimize pediatric resuscitation workflows.

Funding

This study was funded by grant support provided by the Private Foundation of the Geneva University Hospitals.

Data Availability

The datasets generated during this study are available from the corresponding author on reasonable request and after ethical approval.

Conflicts of Interest

JNS received research grant support from the Geneva University Hospitals (HUG) Private Foundation for work related to this project. SM received research grant support from the Geneva Hospitals Private Foundation for work related to this project.

Multimedia Appendix 1
Updated TeamScreen prototype display during a simulated pediatric in-hospital cardiac arrest (Ventricular Fibrillation scenario), showing the cardiopulmonary resuscitation (CPR) algorithm and reversible causes (left); timers and task lists for current, secondary and next actions (center); and the "Completed Tasks" panel with a medications table that presents the final weight-based doses and administration times visible to the entire team.
[PNG File, 532 KB - humanfactors_v13i1e78736_app1.png]

Multimedia Appendix 2
Post-Study System Usability Questionnaire (PSSUQ) statements.

XSL•FO
RenderX

[DOCX File, 18 KB - humanfactors_v13i1e78736_app2.docx]

References

1. Giri S, Gyeltshen D, Wangchuk N, et al. Improving role allocation for cardiopulmonary resuscitation (CPR) in the emergency department: a quality improvement project. BMJ Open Qual 2024 Sep 30;13(3):e002870. [doi: 10.1136/bmjoq-2024-002870] [Medline: 39349305]
2. Topjian AA, Raymond TT, Atkins D, et al. Part 4: pediatric basic and advanced life support: 2020 American Heart Association guidelines for cardiopulmonary resuscitation and emergency cardiovascular care. Circulation 2020 Oct 20;142(16_suppl_2):S469-S523. [doi: 10.1161/CIR.0000000000000901] [Medline: 33081526]
3. Hunt EA, Vera K, Diener-West M, et al. Delays and errors in cardiopulmonary resuscitation and defibrillation by pediatric residents during simulated cardiopulmonary arrests. Resuscitation 2009 Jul;80(7):819-825. [doi: 10.1016/j.resuscitation.2009.03.020] [Medline: 19423210]
4. Corazza F, Stritoni V, Martinolli F, et al. Adherence to guideline recommendations in the management of pediatric cardiac arrest: a multicentre observational simulation-based study. Eur J Emerg Med 2022 Aug 1;29(4):271-278. [doi: 10.1097/MEJ.0000000000000923] [Medline: 35404331]
5. Auerbach M, Brown L, Whitfill T, et al. Adherence to pediatric cardiac arrest guidelines across a spectrum of fifty emergency departments: a prospective, in situ, simulation‐based study. Acad Emerg Med 2018 Dec;25(12):1396-1408. [doi: 10.1111/acem.13564] [Medline: 30194902]
6. Siebert JN, Ehrler F, Combescure C, et al. A mobile device app to reduce time to drug delivery and medication errors during simulated pediatric cardiopulmonary resuscitation: a randomized controlled trial. J Med Internet Res 2017 Feb 1;19(2):e31. [doi: 10.2196/jmir.7005] [Medline: 28148473]
7. Wolfe HA, Morgan RW, Zhang B, et al. Deviations from AHA guidelines during pediatric cardiopulmonary resuscitation are associated with decreased event survival. Resuscitation 2020 Apr;149:89-99. [doi: 10.1016/j.resuscitation.2020.01.035] [Medline: 32057946]
8. Crabb DB, Hurwitz JE, Reed AC, et al. Innovation in resuscitation: a novel clinical decision display system for advanced cardiac life support. Am J Emerg Med 2021 May;43:217-223. [doi: 10.1016/j.ajem.2020.03.007] [Medline: 32291164]
9. Field LC, McEvoy MD, Smalley JC, et al. Use of an electronic decision support tool improves management of simulated in-hospital cardiac arrest. Resuscitation 2014 Jan;85(1):138-142. [doi: 10.1016/j.resuscitation.2013.09.013] [Medline: 24056391]
10. Greig PR, Zolger D, Onwochei DN, Thurley N, Higham H, Desai N. Cognitive aids in the management of clinical emergencies: a systematic review. Anaesthesia 2023 Mar;78(3):343-355. [doi: 10.1111/anae.15939] [Medline: 36517981]
11. Grundgeiger T, Hahn F, Wurmb T, Meybohm P, Happel O. The use of a cognitive aid app supports guideline-conforming cardiopulmonary resuscitations: a randomized study in a high-fidelity simulation. Resusc Plus 2021 Sep;7:100152. [doi: 10.1016/j.resplu.2021.100152] [Medline: 34458879]
12. Corazza F, Fiorese E, Arpone M, et al. The impact of cognitive aids on resuscitation performance in in-hospital cardiac arrest scenarios: a systematic review and meta-analysis. Intern Emerg Med 2022 Oct;17(7):2143-2158. [doi: 10.1007/s11739-022-03041-6] [Medline: 36031672]
13. Corazza F, Arpone M, Tardini G, et al. Effectiveness of a novel tablet application in reducing guideline deviations during pediatric cardiac arrest: a randomized clinical trial. JAMA Netw Open 2023 Aug 1;6(8):e2327272. [doi: 10.1001/jamanetworkopen.2023.27272] [Medline: 37535352]
14. Corazza F, Snijders D, Arpone M, et al. Development and usability of a novel interactive tablet app (PediAppRREST) to support the management of pediatric cardiac arrest: pilot high-fidelity simulation-based study. JMIR Mhealth Uhealth 2020 Oct 1;8(10):e19070. [doi: 10.2196/19070] [Medline: 32788142]
15. Siebert JN, Lacroix L, Cantais A, Manzano S, Ehrler F. The impact of a tablet app on adherence to American Heart Association guidelines during simulated pediatric cardiopulmonary resuscitation: randomized controlled trial. J Med Internet Res 2020 May 27;22(5):e17792. [doi: 10.2196/17792] [Medline: 32292179]
16. Nabecker S, Nation K, Gilfoyle E, et al. Cognitive aids used in simulated resuscitation: a systematic review. Resusc Plus 2024 Sep;19:100675. [doi: 10.1016/j.resplu.2024.100675] [Medline: 38873274]
17. Parush A, Mastoras G, Bhandari A, et al. Can teamwork and situational awareness (SA) in ED resuscitations be improved with a technological cognitive aid? Design and a pilot study of a team situation display. J Biomed Inform 2017 Dec;76:154-161. [doi: 10.1016/j.jbi.2017.10.009] [Medline: 29051106]
18. Calder LA, Bhandari A, Mastoras G, et al. Healthcare providers' perceptions of a situational awareness display for emergency department resuscitation: a simulation qualitative study. Int J Qual Health Care 2018 Feb 1;30(1):16-22. [doi: 10.1093/intqhc/mzx159] [Medline: 29194491]
19. Ehrler F, Del Zotto M, Rouyer F, Weinhold T, Lovis C, Siebert J. Design of InterFACE: a tool to improve collaborative work and decision making during rescucitation. Stud Health Technol Inform 2018;255:117-121. [Medline: 30306919]
20. Ehrler F, Sahyoun C, Manzano S, et al. Impact of a shared decision-making mHealth tool on caregivers' team situational awareness, communication effectiveness, and performance during pediatric cardiopulmonary resuscitation: study protocol

XSL•FO
RenderX

of a cluster randomized controlled trial. Trials 2021 Apr 13;22(1):277. [doi: 10.1186/s13063-021-05170-3] [Medline: 33849611]
21. Kang R, Cheng A, Lin Y, et al. Perceived usability, user experience, and technology acceptance of role-specific augmented reality decision support tools for cardiac arrest resuscitation: prospective observational pilot study. JMIR XR Spatial Comput 2026;3:e72013. [doi: 10.2196/72013]
22. Ehrler F, Rajic A, De Masi A, et al. Overview of a user-centered, mixed-methods process for designing Interconnected and Focused Mobile Applications on Patient Care Environment (InterFACE): augmented-reality decision support system for pediatric resuscitation. JMIR Hum Factors 2026 Feb 13;13:e78144. [doi: 10.2196/78144] [Medline: 41687005]
23. Food and Drug Administration. Applying Human Factors and Usability Engineering to Medical Devices: Guidance for Industry and Food and Drug Administration Staff: Food and Drug Administration; 2016. URL: https://www.fda.gov/media/80481/download [accessed 2026-04-10]
24. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004 Feb;37(1):56-76. [doi: 10.1016/j.jbi.2004.01.003] [Medline: 15016386]
25. Or CK, Chan AHS. Inspection methods for usability evaluation. In: User Experience Methods and Tools in Human-Computer Interaction: CRC Press; 2024:170-192. [doi: 10.1201/9781003495161-7]
26. Rochat J, Ehrler F, Siebert JN, Ricci A, Garretas Ruiz V, Lovis C. Usability testing of a patient-centered mobile health app for supporting and guiding the pediatric emergency department patient journey: mixed methods study. JMIR Pediatr Parent 2022 Mar 15;5(1):e25540. [doi: 10.2196/25540] [Medline: 35289754]
27. Inal Y, Wake JD, Guribye F, Nordgreen T. Usability evaluations of mobile mental health technologies: systematic review. J Med Internet Res 2020 Jan 6;22(1):e15337. [doi: 10.2196/15337] [Medline: 31904579]
28. Sauro J, Lewis JR. Quantifying the User Experience: Practical Statistics for User Research: Morgan Kaufmann Publishers; 2016. [doi: 10.1016/B978-0-12-802308-2.00002-3]
29. Clarke K, Al-Mukhtar A, Alim L, et al. Using mixed reality simulation to improve junior medical trainees' preparedness to manage high-acuity trauma. BMJ Open Qual 2024 Apr 8;13(2):e002575. [doi: 10.1136/bmjoq-2023-002575] [Medline: 38589055]
30. Wiklund ME, Kendler J, Strochlic AY. Usability Testing of Medical Devices: CRC Press; 2015. [doi: 10.1201/b19082]
31. Yen PY, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012;19(3):413-422. [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
32. Richardson S, Mishuris R, O'Connell A, et al. "Think aloud" and "Near live" usability testing of two complex clinical decision support tools. Int J Med Inform 2017 Oct;106:1-8. [doi: 10.1016/j.ijmedinf.2017.06.003] [Medline: 28870378]
33. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Hum Comput Interact 1995 Jan;7(1):57-78. [doi: 10.1080/10447319509526110]
34. Lewis JR. Psychometric evaluation of the PSSUQ using data from five years of usability studies. Int J Hum Comput Interact 2002 Sep;14(3-4):463-488. [doi: 10.1080/10447318.2002.9669130]
35. Bastien JMC, Scapin DL. Ergonomic criteria for the evaluation of human-computer interfaces. INRIA. 1993 URL: https://blocnotes.iergo.fr/wp-content/uploads/2011/04/criteres.pdf [accessed 2026-04-10]
36. Nielsen J. Usability inspection methods. In: Conference Companion on Human Factors in Computing Systems: Association for Computing Machinery:413-414 URL: https://dl.acm.org/doi/pdf/10.1145/259963.260531 [accessed 2026-04-10]
37. Scapin DL, Bastien JMC. Ergonomic criteria for evaluating the ergonomic quality of interactive systems. Behav Inf Technol 1997 Jan;16(4-5):220-231. [doi: 10.1080/0144929971198806]
38. Barton HJ, Salwei ME, Rutkowski RA, et al. Evaluating the usability of an emergency department after visit summary: staged heuristic evaluation. JMIR Hum Factors 2023 Mar 9;10:e43729. [doi: 10.2196/43729] [Medline: 36892941]
39. Hunter JD. Matplotlib: a 2D graphics Environment. Comput Sci Eng 2007;9(3):90-95. [doi: 10.1109/MCSE.2007.55]
40. Hunt EA, Walker AR, Shaffner DH, Miller MR, Pronovost PJ. Simulation of in-hospital pediatric medical emergencies and cardiopulmonary arrests: highlighting the importance of the first 5 minutes. Pediatrics 2008 Jan;121(1):e34-e43. [doi: 10.1542/peds.2007-0029] [Medline: 18166542]
41. Sweller J, van Merriënboer JJG, Paas F. Cognitive architecture and instructional design: 20 years later. Educ Psychol Rev 2019 Jun;31(2):261-292. [doi: 10.1007/s10648-019-09465-5]
42. McEvoy MD, Field LC, Moore HE, Smalley JC, Nietert PJ, Scarbrough SH. The effect of adherence to ACLS protocols on survival of event in the setting of in-hospital cardiac arrest. Resuscitation 2014 Jan;85(1):82-87. [doi: 10.1016/j.resuscitation.2013.09.019] [Medline: 24103233]
43. Guetterman TC, Kellenberg JE, Krein SL, et al. Nursing roles for in-hospital cardiac arrest response: higher versus lower performing hospitals. BMJ Qual Saf 2019 Nov;28(11):916-924. [doi: 10.1136/bmjqs-2019-009487] [Medline: 31420410]
44. Schatz A, Osterhoff G, Georgi C, Joeres F, Neumuth T, Rockstroh M. Improving preparation in the emergency trauma room: the development and impact of real-time data transfer and dashboard visualization system. Int J Comput Assist Radiol Surg 2025 Feb;20(2):301-310. [doi: 10.1007/s11548-024-03256-2] [Medline: 39259482]
45. Siebert JN, Ehrler F, Gervaix A, et al. Adherence to AHA guidelines when adapted for augmented reality glasses for assisted pediatric cardiopulmonary resuscitation: a randomized controlled trial. J Med Internet Res 2017 May 29;19(5):e183. [doi: 10.2196/jmir.7379] [Medline: 28554878]

XSL•FO
RenderX

46. Cheng A, Fijacko N, Lockey A, et al. Use of augmented and virtual reality in resuscitation training: a systematic review. Resusc Plus 2024 Jun;18(100643):100643. [doi: 10.1016/j.resplu.2024.100643] [Medline: 38681058]
47. Hart SG. NASA-task load index (NASA-TLX); 20 years later. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting: Sage Publications; 2006. [doi: 10.1177/154193120605000909]

Abbreviations

6H5T: hypovolemia, hypoxia, hydrogen ion (acidosis), hypokalemia/hyperkalemia, hypoglycemia, hypothermia, tension pneumothorax, tamponade, toxins, thrombosis (pulmonary), thrombosis (coronary)
AR: augmented reality
CPR: cardiopulmonary resuscitation
IHCA: in-hospital cardiac arrest
InterFACE: Interconnected and Focused Mobile Applications on Patient Care Environment
PALS: pediatric advanced life support
PSSUQ: Post-Study System Usability Questionnaire
ROSC: return of spontaneous circulation
TLX: Task Load Index

Edited by A Kushniruk; submitted 08.Jun.2025; peer-reviewed by B Cahill, C Or, GJ Noordergraaf; revised version received 02.Dec.2025; accepted 14.Dec.2025; published 28.Apr.2026.

Please cite as:
Rajic A, Kayne Olanka S, Generelli M, Davidson J, Lin Y, Kang R, Kim K, Rebours PL, Ibrahim M, Duncan D, Manzano S, Cheng A, De Masi A, Siebert JN, Ehrler F
Evaluating a Shared Decision Support Tool for Pediatric Cardiopulmonary Arrest: Mixed Methods Usability Study
JMIR Hum Factors 2026;13:e78736
URL: https://humanfactors.jmir.org/2026/1/e78736
doi:10.2196/78736

© Ana Rajic, Sharleen Kayne Olanka, Marco Generelli, Jennifer Davidson, Yiqun Lin, Ryan Kang, Kangsoo Kim, Pierre-Louis Rebours, Marc Ibrahim, Donovan Duncan, Sergio Manzano, Adam Cheng, Alexandre De Masi, Johan N Siebert, Frederic Ehrler. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 28.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Development and Usability of a Dashboard for Quality Monitoring and Resident-Centered Care in Australian Residential Long-Term Care: Mixed-Methods Study

Ronald Dendere[1], BEng, MSc, PhD; Gillian Stockwell-Smith[1,2], PhD; Michelle Lang[1], MPSR; Murray Hargrave[1], PhD; Sara Mayfield[3], PhD; Leonard Charles Gray[1], PhD

[1]Centre for Health Services Research, Faculty of Health, Medicine & Behavioural Sciences, The University of Queensland, Level 5, Health Sciences Building, Herston Campus, Herston, Brisbane, Queensland, Australia

[2]School of Nursing and Midwifery, Griffith University, Brisbane, Australia

[3]Regis Aged Care, Camberwell, Melbourne, Australia

Corresponding Author:
Ronald Dendere, BEng, MSc, PhD
Centre for Health Services Research, Faculty of Health, Medicine & Behavioural Sciences, The University of Queensland, Level 5, Health Sciences Building, Herston Campus, Herston, Brisbane, Queensland, Australia

Abstract

Background: The Australian National Aged Care Mandatory Quality Indicator Program (QI Program) requires government-subsidized residential aged care service providers to report quarterly data on a set of quality indicators. These indicators measure provider performance across specific domains of care and are intended to support continuous quality improvement. Health care dashboards can enhance the use of indicators by presenting data in interactive and intuitive formats that enable actionable insights.

Objective: This mixed methods study aimed to develop an electronic dashboard to assist service providers' use of QI Program data to measure, track, and improve the quality of resident care.

Methods: A participatory design methodology was used to co-design and co-develop the dashboard. Initially, stakeholder participants for the co-design were identified. A combination of workshops, meetings, and email communications with co-design participants was then used to iteratively define and refine user requirements and to develop and improve the dashboard prototype. A 3-month pilot of the dashboard was conducted with a convenience sample of 30 end-users across 12 nursing homes and a post-pilot survey based on the System Usability Scale (SUS) was used to assess end-users' perceptions of the dashboard usability.

Results: The dashboard supports multiple user roles by enabling comparisons across homes and detailed views of all indicators for individual homes. A key feature is the ability to progressively view data at various levels of detail: groups of homes, individual homes, resident groups, and individual residents. The resident-level view enables more targeted, personalized care by helping staff identify and prioritize the specific indicators triggered by each resident. The average SUS score was 75.2 (SD 16.3), indicating good usability for the dashboard. Most survey respondents (12/14, 85.7%) were likely or extremely likely to recommend the dashboard to a colleague and agreed the dashboard would support the delivery of personalized care for residents. Almost all respondents (13/14) agreed or strongly agreed that the dashboard would assist with quality monitoring and improvement activities, and some pilot participants also made suggestions for incorporating the dashboard into those activities.

Conclusions: This study demonstrates the potential value of a co-designed dashboard to support the use of quality indicator data in residential aged care. Limitations of the current prototype include short pilot duration, convenience sampling, and reliance on manual quarterly data uploads, which constrain generalizability and scalability. Future work should explore long-term integration of the dashboard into routine quality improvement processes and evaluate its impact on resident outcomes and care quality over time.

(JMIR Hum Factors 2026;13:e80478) doi:10.2196/80478

KEYWORDS

digital health; health IT; usability; care quality; quality indicator; health care dashboard; co-design; co-development; nursing home; residential aged care; long-term care; continuous quality improvement

XSL•FO
RenderX

Introduction

Quality indicators have become a common strategy in Australia and globally to support monitoring and improvement of quality in residential long-term care settings, including residential aged care facilities (also known as nursing homes) [1-6]. These metrics are quantitative measures for assessing care processes, structures, and outcomes to indicate how well care providers are performing in specific domains of care [7]. Indicators can facilitate the identification of care domains where quality is unacceptable or low and can enable tracking and comparison of quality at home/facility, organizational, regional, and national levels [1,7,8]. Indicators can also help identify top-performing services and enable the setting of evidence-based benchmarks to guide quality improvement efforts across long-term care settings [9].

Several countries use indicators for public reporting and comparing the performance of service providers [6]. In Australia, under the National Aged Care Mandatory Quality Indicator Program (QI Program), all government-subsidized service providers are required to report data on 14 indicators to the federal government every quarter of the financial year [10]. The program is designed to enable both the government and service providers to identify areas for improvement [10]. Five of the indicators also form a performance area used to determine a home's overall rating in the public star rating system for residential aged care homes in Australia. Providers collect individual resident data in each home/facility using government-issued quality indicator spreadsheet templates that aggregate the data prior to submission. Typically, these data are sourced from different software, including clinical information systems, and risk and medication management software.

Quality indicators and their associated data must be understandable and actionable to fulfill their function as quality improvement tools [11,12]. Health information technologies (HITs), particularly dashboards, can be a key facilitator for this goal. Health care dashboards aggregate and analyze data about care recipients and care processes from multiple sources and present the data using graphics and other visuals to facilitate decision-making for health care professionals [13-15]. Dashboards are ubiquitous in acute care [13,16-18] but are increasingly utilized in the aged care sector in high-income countries, including Australia [19-21]. Although software products are available for collecting and reporting QI Program data, there is no published evidence of dashboards designed to support the practical use of these data for quality monitoring and continuous improvement by providers. Likewise, the scientific literature lacks evidence of dashboards developed specifically for QI Program indicators or demonstrating that providers are actively using these data to guide continuous quality improvement, as intended by the program. The QI Program's quarterly collection of individual resident data presents an opportunity for providers to identify care domains requiring attention for each resident (ie, the indicators triggered) and to implement targeted care protocols accordingly. This resident-level approach not only has the potential to enhance outcomes for residents but also to improve overall provider performance within the QI Program. A dashboard would consolidate resident data into a single, accessible view and eliminate the need for providers to manually search across disparate software systems or navigate individual spreadsheets used for QI Program reporting.

Research shows that, like other HITs, involving end-users in the design and development of dashboards helps ensure the final product meets their specific needs. This collaborative approach also fosters a sense of ownership among end-users and other stakeholders, increasing the likelihood of adoption and long-term use [13,22-24].

The aim of this study was to develop and evaluate the usability of a dashboard to enable residential long-term care providers to use their mandated QI Program data to measure, monitor and improve the quality of resident care.

Methods

Study Design, Context, and Setting

This was a mixed methods study that used different methodologies for the dashboard design and development, and evaluation of usability. The dashboard described in this study is a component of a larger project that aimed to develop various standards-based health data, digital tools, and strategies for quality improvement [25]. The industry partner was Regis Aged Care (Regis), one of Australia's largest aged care service providers [25].

Ethical Considerations

Ethical approval was provided by the University of Queensland Human Research Ethics Committee (ID: 2022/HE002362). After accepting invitations to participate in the project, participants were supplied with an information sheet summarizing the project and a consent form for signing prior to data collection. Furthermore, researchers began each workshop/interview by explaining the project aims, procedures, and ethical considerations, including participants' right to withdraw at any time. Participants were not compensated for participating in the project. All members of the research team adhered to institutional procedures governing research data storage, access, and privacy. Data were transcribed and deidentified using a commercial transcription service.

Iterative Co-Design Approach

Overview

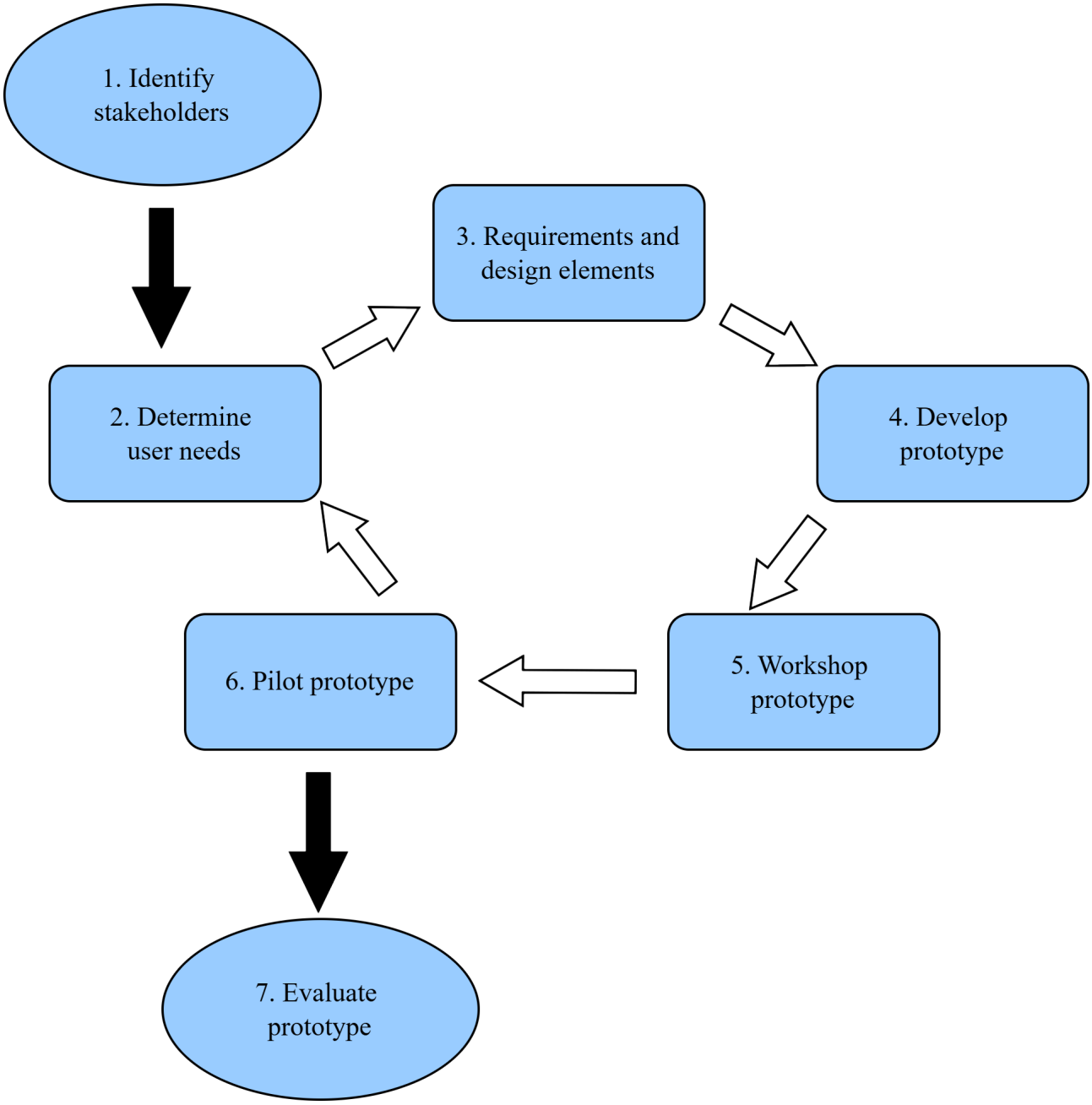
We used a participatory design methodology to develop the dashboard. The core principle of this user-centered approach is the direct involvement of people for whom a product is being designed in a collaborative co-design process [26]. This approach has been successfully used in the design of health care dashboards in general [27-29] and specifically, aged care dashboards [20,28,29].

Figure 1 summarizes the design, development, and evaluation process [27,28,30]. The process was largely iterative and entirely collaborative. Stages 1 - 3 and 5 were conducted using a combination of in-person and online workshops to accommodate participants' availability and geographical locations. Except for

XSL•FO
RenderX

the first workshop, notes and outcomes of a prior workshop were reviewed at the beginning of each workshop to play these back, and confirm agreement of outcomes and decisions. We also leveraged fortnightly project committee meetings and email communications to expedite design decisions (all members of the project committee were participants in the co-design group).

Figure 1. Co-design and evaluation stages.



Data Collection and Analysis

Workshop notes, transcripts of workshop recordings, and feedback received via email constituted the qualitative data for the project. As mentioned, the qualitative data were transcribed and deidentified using a commercial transcription service. Content analysis [31] was conducted using NVivo software [32] to derive the data sought at each stage (ie, identifying stakeholders, user needs, and requirements for the dashboard).

Identifying Stakeholders

To kickstart the project, we established a project committee consisting of the research team and 4 Regis staff members. The project committee collaboratively identified individuals and groups within Regis in executive and operational management roles for whom data aggregation and interrogation of the care, needs, and outcomes of residents formed a significant part of their responsibilities (ie, stakeholders). The Regis staff on the project committee then invited specific individuals from each group to participate in the co-design (ie, co-design participants).

Determining User Needs, Developing Requirements, and Design Elements for the Dashboard

A discussion of user needs was initiated in the first workshop by seeking out answers to three broad questions:

1. What is the purpose of the dashboard?

XSL•FO
RenderX

2. How will end-users use the dashboard?
3. What would end-users expect to achieve by using the dashboard?

Data collected during the first workshop were analyzed as described above ("Data Collection and Analysis" section) to compile a list of feature and functional requirements for the dashboard. The requirements were discussed and collaboratively refined during the second workshop. The final requirements list was translated into design elements. Design elements refer to user interface components (eg, buttons, icons, and menus), user experience patterns (eg, navigation patterns, workflows, and interaction models), layout of visuals, information architecture (organization of content) and system design (data structure and back-end logic that supports the front-end interactions/visuals).

Developing and Workshopping a Prototype

The first step in building the dashboard prototype was drawing up wireframes for each prospective view/page based on the design elements. A wireframe is a simple schematic created in the early stages of designing a digital product to help designers to communicate and co-design participants to visualize user-interfaces of digital products (eg, web pages, mobile apps, dashboards) [33,34]. We used a free, web-based diagramming tool to create the wireframes [35], which were distributed to participants via email to gather feedback and input. Design details requiring in-depth discussions were deferred for consideration during workshops to obtain consensus. This approach facilitated feedback and confirmed acceptability without requiring time-intensive coding.

Following consensus on the wireframes, development of the dashboard commenced by building the back end (ie, data model, logic, and calculations). The lead author developed the dashboard using Microsoft Power BI software (Microsoft Inc) [36], with technical and professional support from a credentialed Microsoft services provider [37]. Data for the dashboard were sourced from Regis's quarterly QI Program data. Their national quality assurance manager, who was also a co-design participant, anonymized the residents using a unique code before supplying the data to the dashboard developer. The national quality assurance manager also supplied Regis's IT team with the master spreadsheet containing the resident name and anonymization code pairs. Upon completion of the development, the dashboard file was sent to the Regis IT team who uploaded the master spreadsheet into the dashboard files prior to publishing. This approach preserved the privacy of residents while allowing individual residents to be linked between quarters and to be re-identified by Regis staff within the dashboard.

Following completion of the first prototype, and each update thereafter, we conducted online or in-person workshops to demonstrate the dashboard's features and functions to the co-design group and gather feedback. During online workshops, the lead author guided participants through the dashboard workflow. During in-person workshops, participants were given hands-on access to the dashboard to explore and discuss its features and functionality. We employed the Think-Aloud method, an established approach for user interface evaluation in digital health [38], to gather real-time feedback. Participants were asked to verbalize their thoughts, observations, and opinions as they interacted with the dashboard. A member of the research team recorded these verbal responses and observations for subsequent analysis.

Pilot and Usability Evaluation

To evaluate the dashboard usability, we conducted a pilot study designed to assess usability and perceived usefulness of the prototype, rather than to evaluate impact on care quality outcomes (which would require a longer evaluation period). Pilot participants were a convenience sample of 30 end-users in 12 nursing homes, consisting of general managers, regional general managers, clinical care managers, and clinical care specialists. Pilot participants were granted unrestricted access to the dashboard and encouraged to explore it freely. However, to facilitate structured engagement, we provided a monthly task list suggesting specific areas for exploration. Throughout the pilot, we met with participants during their routine monthly quality review meetings to monitor task progression and address user queries or technical issues.

At the conclusion of the pilot, we assessed the dashboard's usability using the System Usability Scale (SUS) [39], a widely adopted, standardized instrument for evaluating perceived product usability, including HITs [40]. It comprises 10 items rated on a 5-point Likert scale, yielding an overall usability score between 0 and 100 [41,42]. SUS scores below 60 signal poor usability while scores in the 60 - 69 range signal fair usability, 70 - 79 signal good usability, 80 - 84 signal excellent usability, and 85 - 100 signal the best achievable usability [43]. In addition to the SUS items, the survey included questions assessing the perceived usefulness of the dashboard for person-centered care, its potential to support quality improvement, and suggestions for improvements. The survey was designed and administered using the Qualtrics platform.

Results

Co-Design and Iterative Approach

Overall, we conducted 5 workshops for Stages 1 - 5: the first 2 workshops for Stages 1 - 3 and the next 3 for Stages 2 - 5 (see Figure 1). The combination of hands-on exploration and the Think-Aloud approach during in-person workshops was effective at stimulating discussion among the co-design group and facilitating development of new ideas for further enhancement of the dashboard.

Identifying Stakeholders

Overall, 9 stakeholders were identified across three broad categories reflecting the clinical care hierarchy of Regis:

1. Organization level: executive general managers, clinical governance and care committees, regional general managers, regional quality and improvement managers, regional clinical care specialist managers, quality and improvement teams, and clinical care specialists
2. Home level: general managers and clinical care managers
3. Resident level: registered nurses

Identifying stakeholders enabled the co-design group to consider their specific needs during the design of the dashboard. The co-design group consisted of the following roles (Textbox 1).

XSL•FO
RenderX

Textbox 1. Roles of the co-design group.

Research

- Informatics
- Health services and quality improvement
- Geriatric medicine
- Research nurse

Regis aged care

- Executive general manager
- Group manager
- Regional quality and improvement manager
- National quality assurance manager
- Regional clinical care specialist manager
- Quality and improvement manager

User Needs, Dashboard Requirements, and Design Elements

The final list of requirements is provided below.

1. The dashboard must display all current clinical and consumer experience indicators [10], that is, all except the "Workforce" indicator.
2. Home-level and resident-level end-users should have ability to select a specific home and period (ie, the "financial year" and "quarter") whose indicators they intend to view.
3. Organization-level end-users should be able to select a group of homes and compare their performance for any indicator.
4. The dashboard must allow benchmarking against a national or organizational statistic (eg, average, median).
5. End-users should have the ability to drill-through the data and view details for each indicator per home. Drill-through is a term for functionality in dashboards that enables users to explore data at different levels of granularity by navigating from summary-level pages to detail-level pages. For example, navigating from an overview page showing all indicators for a specific home to a page showing details for the pressure injuries indicator (distribution of pressure injury stages, pressure injuries acquired outside the home, trends, etc).
6. The drill-through function must allow the viewing of individual residents, including the resident's status for each indicator with clear indication of triggered indicators.
7. Access to resident-level data must be constrained by user credentials (eg, a registered nurse should only be able to view residents in the home in which they work).
8. The dashboard should have functionality to export visualizations/reports into PDF documents.

Design elements included drop-down lists for selecting variables to view (ie, homes, year, and quarter) and several combinations of "gauge," "card," and bar and line graph visuals in paginated views within the dashboard. Color was used extensively to convey key information visually—for example, to indicate a home's performance relative to benchmarks. Co-design participants emphasized the importance of designing the dashboard to meet the diverse needs of end-users.

Developing and Workshopping a Prototype

The final prototype incorporates 2 workflows (Figure 2) to fulfill the different needs of end-users: a common technique for tailoring health care dashboards for multiple use-cases using the same data [18,44]. The "organization-level" workflow, for users in the organization-level category (see the "Identifying Stakeholders" section in the Results), allows users to select several homes, the period, and a specific indicator for comparison. The "home-level" workflow, for users in the home- and resident-level categories, allows the selection and viewing of all indicators for a single home for a given period. Figure 3 shows the landing page for the organization-level workflow and an example of the subsequent comparison page when an indicator is selected.

XSL•FO
RenderX

Figure 2. Organization-level and home-level workflows for the dashboard. QI: quality indicator.

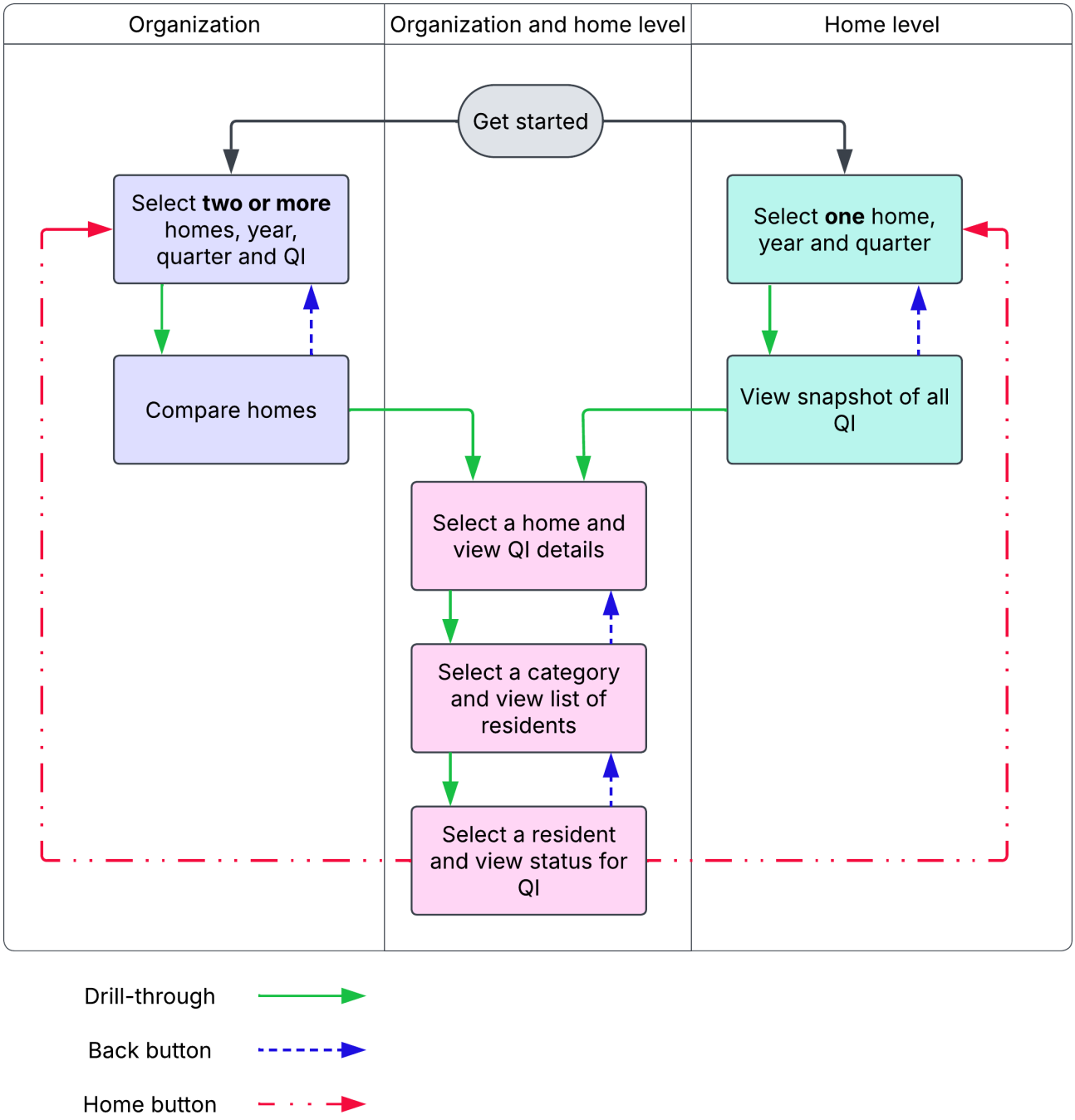
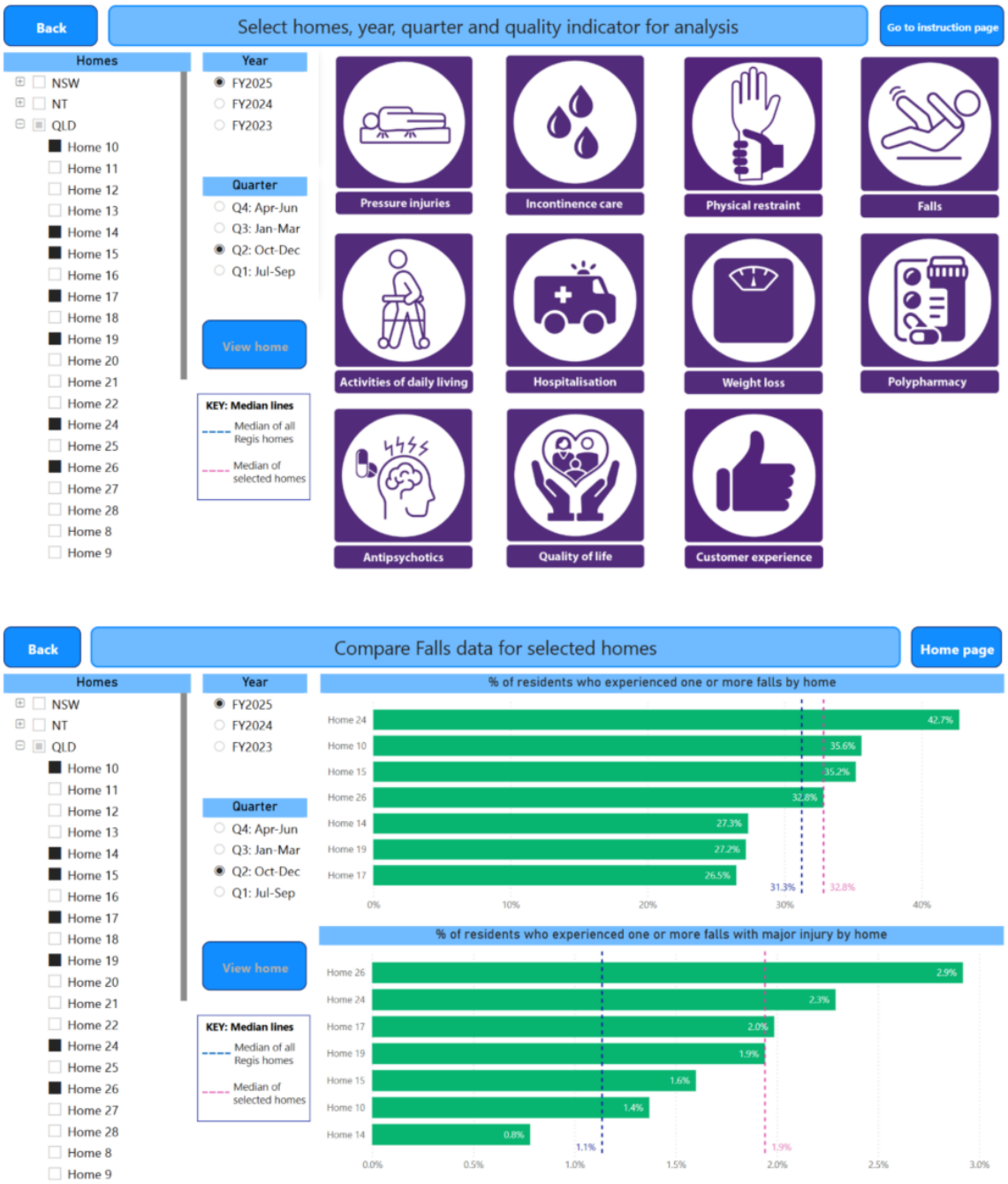


Figure 4 shows the landing page for the home-level workflow, which unlike the organization-level workflow, immediately displays an overview of all indicators after selection of a home and period. In this view, each home can be compared against a benchmark set for all homes in the organization. The benchmark was set as part of the co-design process to fulfill one of the requirements (see the "User Needs, Dashboard Requirements, and Design Elements" in the Results). It is indicated on the visual for each indicator by the dark line, and the actual value can be displayed by hovering the cursor over the line. The color scheme, also chosen during co-design, provides a visual indication of a home's performance relative to the benchmark, with the intensity of color (ie, brighter/darker) indicating the percentage increase away from benchmark, as follows: (1) green signals performance above the benchmark, (2) amber signals performance equal to the benchmark, and (3) red signals performance below the benchmark.

XSL•FO
RenderX

Figure 3. The (A) homepage for the organization-level version of the dashboard and (B) comparison of falls indicators. The names have been anonymized for illustrative purposes.



XSL•FO
RenderX

Figure 4. The homepage for the home-level version of the dashboard. The name of the selected home has been anonymized for illustrative purposes.

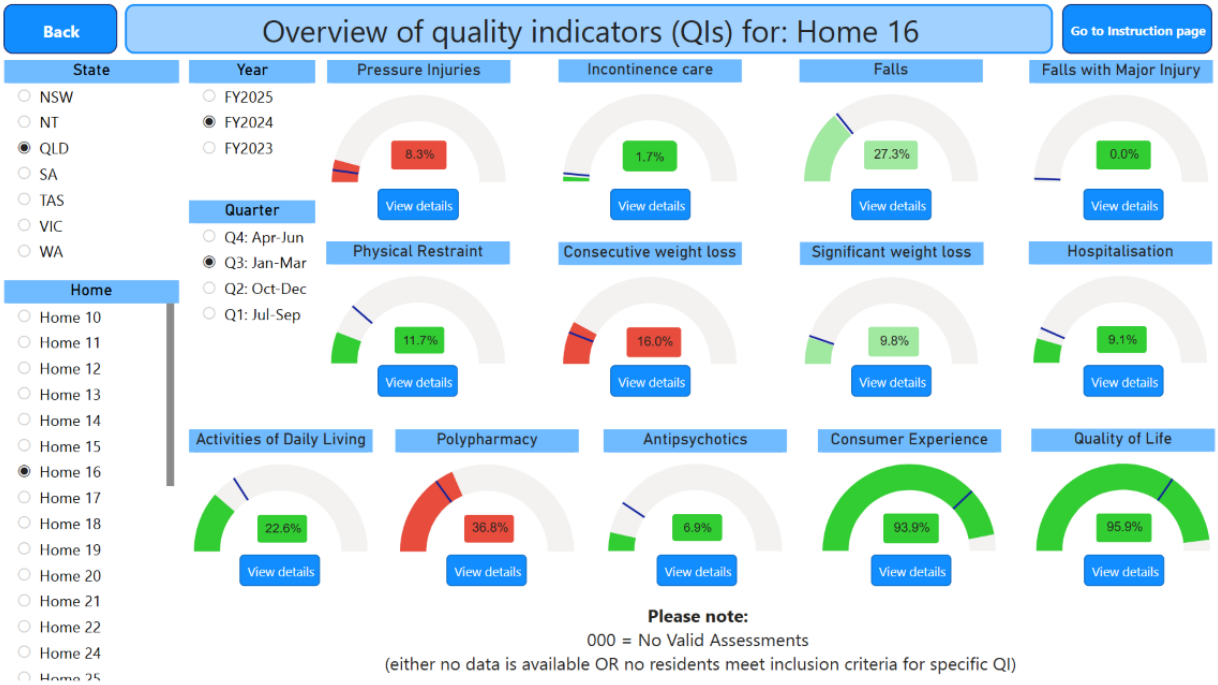
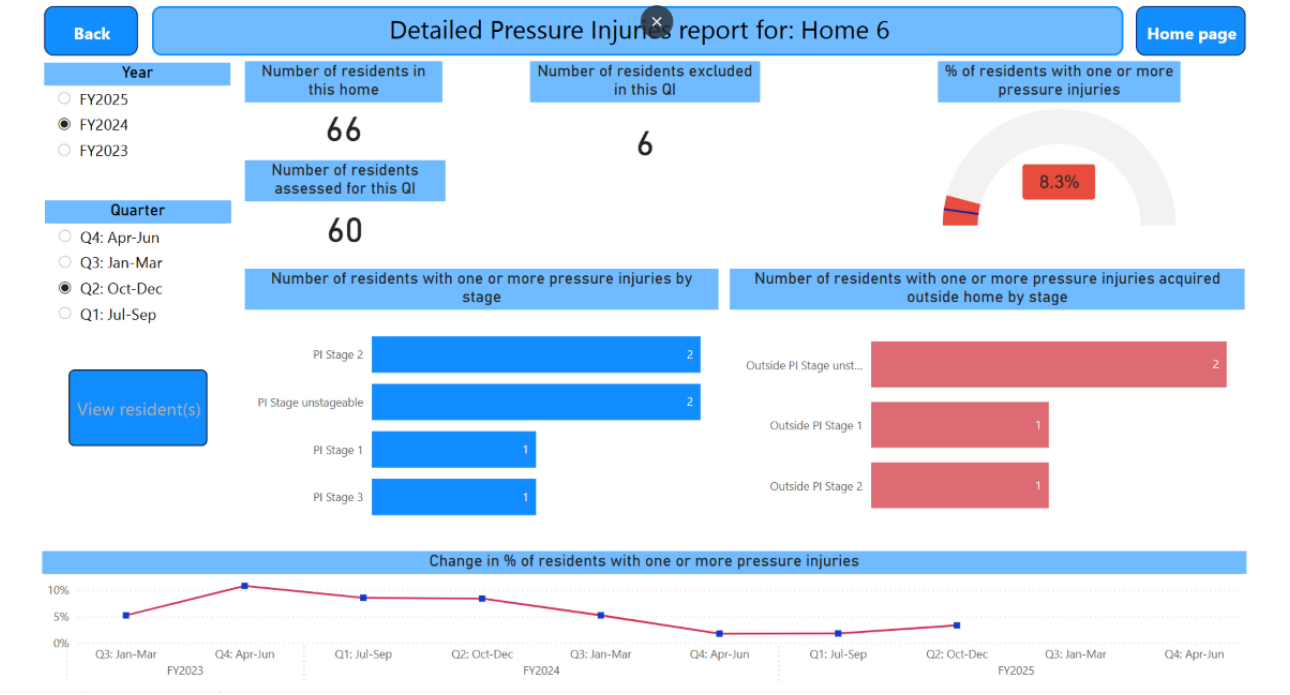


Figure 5 shows an example of a drill-through to an indicator detail page. The indicator detail page displays all the reporting items required by the QI Program for that indicator. Figure 6 shows the resident-level data where a color scheme for each quality indicator title indicates the resident's status as follows: (1) red signals that the resident has triggered the indicator (eg, the resident has a pressure injury), (2) green signals that the resident has not triggered the indicator (eg, the resident does not have a pressure injury), and (3) blue signals that the data for the indicator are not available (eg, the resident declined to be assessed for pressure injuries).

Figure 5. An example of a detailed QI (Pressure injuries) report for a specific home. The name of the home has been anonymized for illustrative purposes. QI: quality indicator.



XSL•FO
RenderX

Figure 6. An individual resident's QI status detail page. The resident's name and home have been anonymized for illustrative purposes. QI: quality indicator.



Dashboard Pilot and Evaluation

Of the 30 individuals invited to participate in the pilot, 29 participants accessed the dashboard at least once during the pilot period, and 16 opened the usability survey (53% response rate) and 14 completed the survey (87.5% completion rate). The average SUS score was 75.2 (SD 16.3), indicating end-users thought usability of the dashboard was good. Among the completed surveys, 85.7% (12/14) of respondents reported being likely or extremely likely to recommend the dashboard to a colleague. An equal proportion (12/14) believed the dashboard would support the delivery of personalized care for residents. Additionally, 92.8% (13/14) agreed or strongly agreed that the dashboard would assist with quality monitoring and improvement activities within Regis.

The drill-through to individual resident-level views was well-received, with participants noting it could help staff quickly identify and prioritize relevant care domains, thereby promoting more person-centered care. In response to a survey item on integrating the dashboard into routine workflows, pilot participants suggested it could be useful for preparing internal quarterly reviews. They also recommended refreshing the data monthly to enable exporting dashboard views for inclusion in monthly clinical quality reports. Participants offered several suggestions for further improvements. However, some of these were contradictory, reflecting differences in personal preferences. For example, while some users preferred listing homes alphabetically when comparing performance, others favored sorting by descending indicator value to highlight the best- or worst-performing homes.

Discussion

Principal Findings

This article describes the design, development, and preliminary evaluation of a dashboard for quality indicator data. To our knowledge, it is the first published study to describe a dashboard specifically designed to support the Australian QI Program in aged care. Workshops were the primary collaboration method in the co-design process, complemented by a flexible, multi-channel approach that maintained momentum by enabling the prompt resolution of minor issues between scheduled sessions.

The multi-level drill-through capability is a key feature of the dashboard, which allows users to explore data at progressively finer levels of detail. This functionality supports a broad range of users across executive, managerial, and clinical roles. By providing secure access to resident-level data, the dashboard supports data-driven personalized care, which is associated with better resident outcomes and higher satisfaction and quality of life [45]. In this way, quality indicator data can be used as an operational asset for continuous quality improvement, enabling a bottom-up approach in which targeted care at the individual resident level may cumulatively contribute to improved performance at both home and organizational levels.

The dashboard requirements that were not specific to our use-case, that is, those that can be generalized to most health care dashboards, were consistent with those in the existing literature [14]. Like participants in other co-designed health care dashboards, co-design participants required functionality for selecting specific data subsets (homes or groups of homes, years, and calendar quarters) to view/compare indicators and multi-level drill-through capability, to have the ability to benchmark homes against group statistics, and to have secure

XSL•FO
RenderX

access to resident-level data [14,19,46-48]. The design elements incorporated into the dashboard (radio buttons for navigation, color-coded visuals, line and bar charts, etc) were also consistent with existing literature [19]. The co-design and pilot participants found these design elements to be simple but powerful ways for conveying important details within the data as reported by users of other health care dashboards [19].

Study Strengths and Limitations

A key strength of this study is the participatory, iterative development approach involving stakeholders across organizational executives, managers, and nurses. This enabled the tailoring the dashboard for real-world quality monitoring workflows and to support both performance oversight and resident-level care planning. Another strength is that the pilot was conducted in an operational aged care setting, providing an initial indication of feasibility and usability in practice.

This study has several limitations. First, the pilot evaluation was conducted with a convenience sample within a single aged care provider organization, and only 14 of the 30 participants completed the post-pilot survey. This limits generalizability and introduces the possibility of response and selection bias. Second, the 3-month pilot duration may be insufficient to understand long-term adoption, workflow integration, and sustained use. Third, the evaluation focused on perceived usability and usefulness but not effectiveness; we did not directly measure changes in care quality, staff decision-making, or resident outcomes attributable to dashboard use. Fourth, the prototype relied on manual data uploads and did not include automated data pipelines, which may limit scalability and real-world implementation readiness. Fifth, the dashboard operated on QI Program data, which is collected quarterly and may be too infrequent for timely monitoring of emerging issues in resident care.

Implications for Practice and Future Development

Although the findings suggest potential for dashboards to make quality indicators more practical for monitoring care quality and supporting continuous improvement, broader implementation in practice will require technical and organizational development beyond this prototype stage. Future work should progress in 3 stages. First, implementation-ready infrastructure is needed, including a central data warehouse (or equivalent) and automated data pipelines from clinical information system sources. This would improve scalability, reduce manual workload, and enable stronger data quality controls (eg, completeness and timeliness checks). Second, dashboard functionality should be extended to support more frequent refreshes and enhanced analytics such as automated alerts and trend detection (improvements suggested by participants during this pilot). The divergent feedback on interface preferences during the pilot (eg, alphabetical vs performance-based sorting) suggests that usability in this context depends not only on simplicity, but also on flexibility. Future iterations should therefore include user-configurable interface features, such as selectable sort and display options, role-specific default views, and saved filters. These enhancements would better accommodate diverse workflows and user preferences across organizational, home-level, and clinical roles. Third, the dashboard should be evaluated in a larger, multi-site longitudinal study to measure not only usability and adoption but also the impact on quality improvement processes and resident outcomes. Organizations must establish robust governance strategies to mitigate the risk of “tunnel vision”—a phenomenon where attention becomes overly focused on the domains measured by implemented quality indicators at the expense of unmeasured aspects of care [49]. While this is a risk of using indicators for improvement more broadly [49], a dashboard with resident-level views may heighten the effect by directing clinical attention primarily toward the indicators presented on screen. Aged care organizations should also consider expanding the suite of dashboard quality indicators to complement the current QI Program and encompass a broader range of critical resident care domains.

Specific Lessons

This study holds several lessons for informatics and the aged care sector. A multi-channel collaboration approach helped us gather feedback faster than relying solely on the workshops. The aged care industry, especially in Australia, is highly regulated, and providers occasionally have to attend to short-notice site visits and requests for information from regulators, necessitating immediate and intense activity by executive and operational managers. For this reason, it was common for some co-design participants to suddenly become unavailable for scheduled co-design workshops; however, the fortnightly meeting and emails provided additional avenues and opportunities to capture feedback and contributions. While fortnightly meetings may not be feasible for others, less frequent timeframes may be sufficient (eg, monthly or bimonthly). The feedback gathered from co-design participants using the Think-Aloud technique when interacting with the dashboard during in-person workshops provided richer data than online workshops. The Think-Aloud technique, which has been successful applied in HIT co-design in aged care [50], was particularly effective at stimulating insightful discussions and generation of ideas among co-design participants.

Conclusion

In conclusion, dashboards co-designed with stakeholders have the potential to help aged care organizations operationalize quality indicator data for monitoring care quality and personalized care in residential aged care. Larger and longitudinal evaluations are required to determine their long-term impact on care quality and resident outcomes.

Funding

This research is supported by Digital Health Cooperative Research Centre Limited (DHCRC). DHCRC is funded under the Australian Commonwealth's Cooperative Research Centres (CRC) Program.

XSL•FO
RenderX

Authors' Contributions

All authors made a direct, substantial, and intellectual contribution to the conception, conduct, and write-up of this study.

Conflicts of Interest

None declared.

References

1. Frijters DHM, van der Roest HG, Carpenter IGI, et al. The calculation of quality indicators for long term care facilities in 8 countries (SHELTER project). BMC Health Serv Res 2013 Apr 15;13:138. [doi: 10.1186/1472-6963-13-138] [Medline: 23587337]
2. Loureiro N, Martin JI. Quality indicators for residential long-term care for the elderly: a scoping review. RASP 2024 Jul;12(2):1-23. [doi: 10.17583/rasp.13874]
3. Igarashi A, Eltaybani S, Takaoka M, Noguchi-Watanabe M, Yamamoto-Mitani N. Quality assurance in long-term care and development of quality indicators in Japan. Gerontol Geriatr Med 2020;6:2333721420975320. [doi: 10.1177/2333721420975320] [Medline: 35047653]
4. Inacio MC, Eshetie TC, Caughey GE, et al. Quality and safety in residential aged care: an evaluation of a national quality indicator programme. Intern Med J 2023 Nov;53(11):2073-2078. [doi: 10.1111/imj.16052] [Medline: 36878881]
5. Inacio MC, Lang C, Caughey GE, et al. The Registry of Senior Australians outcome monitoring system: quality and safety indicators for residential aged care. Int J Qual Health Care 2020 Nov 9;32(8):502-510. [doi: 10.1093/intqhc/mzaa078] [Medline: 32696038]
6. Osińska M, Favez L, Zúñiga F. Evidence for publicly reported quality indicators in residential long-term care: a systematic review. BMC Health Serv Res 2022 Nov 24;22(1):1408. [doi: 10.1186/s12913-022-08804-7] [Medline: 36424603]
7. Mainz J. Defining and classifying clinical indicators for quality improvement. Int J Qual Health Care 2003 Dec;15(6):523-530. [doi: 10.1093/intqhc/mzg081] [Medline: 14660535]
8. Zimmerman DR. Improving nursing home quality of care through outcomes data: the MDS quality indicators. Int J Geriatr Psychiatry 2003 Mar;18(3):250-257. [doi: 10.1002/gps.820] [Medline: 12642895]
9. Schwabe J, Caughey GE, Jorissen R, et al. Setting standards in residential aged care: identifying achievable benchmarks of care for long-term aged care services. Int J Qual Health Care 2024 Dec 10;36(4):mzae105. [doi: 10.1093/intqhc/mzae105] [Medline: 39562325]
10. About the QI Program. Australian Government Department of Health, Disability and Ageing. 2024. URL: https://www.health.gov.au/our-work/qi-program/about [accessed 2025-02-28]
11. Barbazza E, Klazinga NS, Kringos DS. Exploring the actionability of healthcare performance indicators for quality of care: a qualitative analysis of the literature, expert opinion and user experience. BMJ Qual Saf 2021 Dec;30(12):1010-1020. [doi: 10.1136/bmjqs-2020-011247] [Medline: 33963072]
12. Berwick DM, James B, Coye MJ. Connections between quality measurement and improvement. Med Care 2003 Jan;41(1 Suppl):I30-I38. [doi: 10.1097/00005650-200301001-00004] [Medline: 12544814]
13. Helminski D, Sussman JB, Pfeiffer PN, et al. Development, implementation, and evaluation methods for dashboards in health care: scoping review. JMIR Med Inform 2024 Dec 10;12:e59828. [doi: 10.2196/59828] [Medline: 39656991]
14. Rabiei R, Almasi S. Requirements and challenges of hospital dashboards: a systematic literature review. BMC Med Inform Decis Mak 2022 Nov 8;22(1):287. [doi: 10.1186/s12911-022-02037-8] [Medline: 36348339]
15. Stadler JG, Donlon K, Siewert JD, Franken T, Lewis NE. Improving the efficiency and ease of healthcare analysis through use of data visualization dashboards. Big Data 2016 Jun;4(2):129-135. [doi: 10.1089/big.2015.0059] [Medline: 27441717]
16. Buttigieg SC, Pace A, Rathert C. Hospital performance dashboards: a literature review. J Health Organ Manag 2017 May 15;31(3):385-406. [doi: 10.1108/JHOM-04-2017-0088] [Medline: 28686130]
17. Dowding D, Randell R, Gardner P, et al. Dashboards for improving patient care: review of the literature. Int J Med Inform 2015 Feb;84(2):87-100. [doi: 10.1016/j.ijmedinf.2014.10.001] [Medline: 25453274]
18. Young L, Vogelsmeier A. Quality dashboards in hospital settings: a systematic review with implications for nurses. J Nurs Care Qual 2024;39(2):188-194. [doi: 10.1097/NCQ.0000000000000747] [Medline: 37782907]
19. Siette J, Dodds L, Sharifi F, et al. Usability and acceptability of clinical dashboards in aged care: systematic review. JMIR Aging 2023 Jun 19;6:e42274. [doi: 10.2196/42274] [Medline: 37335599]
20. Silva SSM, Wabe N, Nguyen AD, et al. Development of a predictive dashboard with prescriptive decision support for falls prevention in residential aged care: user-centered design approach. JMIR Aging 2025 Apr 7;8:e63609. [doi: 10.2196/63609] [Medline: 40193194]
21. Ludlow K, Westbrook J, Jorgensen M, et al. Co-designing a dashboard of predictive analytics and decision support to drive care quality and client outcomes in aged care: a mixed-method study protocol. BMJ Open 2021 Aug 25;11(8):e048657. [doi: 10.1136/bmjopen-2021-048657] [Medline: 34433599]
22. Murphy DR, Savoy A, Satterly T, Sittig DF, Singh H. Dashboards for visual display of patient safety data: a systematic review. BMJ Health Care Inform 2021 Oct;28(1):e100437. [doi: 10.1136/bmjhci-2021-100437] [Medline: 34615664]

XSL•FO
RenderX

23. van Elten HJ, Sülz S, van Raaij EM, Wehrens R. Big data health care innovations: performance dashboarding as a process of collective sensemaking. J Med Internet Res 2022 Feb 22;24(2):e30201. [doi: 10.2196/30201] [Medline: 35191847]
24. van de Baan FC, Lambregts S, Bergman E, Most J, Westra D. Involving health professionals in the development of quality and safety dashboards: qualitative study. J Med Internet Res 2023 Jun 12;25:e42649. [doi: 10.2196/42649] [Medline: 37307058]
25. Aged Care Data Compare to put quality benchmarking to the test. Digital Health Cooperative Research Centre. 2023. URL: https://digitalhealthcrc.com/aged-care-data-compare-to-put-quality-benchmarking-to-the-test [accessed 2025-02-28]
26. Simonsen J, Robertson T, editors. Routledge International Handbook of Participatory Design: Routledge; 2013. [doi: 10.4324/9780203108543]
27. Ratwani RM, Fong A. "Connecting the dots": leveraging visual analytics to make sense of patient safety event reports. J Am Med Inform Assoc 2015 Mar;22(2):312-317. [doi: 10.1136/amiajnl-2014-002963] [Medline: 25336592]
28. Burningham Z, Lagha RR, Duford-Hutchinson B, et al. Developing the VA Geriatric Scholars Programs' clinical dashboards using the PDSA framework for quality improvement. Appl Clin Inform 2022 Aug;13(4):961-970. [doi: 10.1055/s-0042-1757553] [Medline: 36223868]
29. Esquer Rochin MA, Gutierrez-Garcia JO, Rosales JH, Rodriguez LF. Design and evaluation of a dashboard to support the comprehension of the progression of patients with dementia in day centers. Int J Med Inform 2021 Dec;156:104617. [doi: 10.1016/j.ijmedinf.2021.104617] [Medline: 34710725]
30. Woods L, Cummings E, Duff J, Walker K. Conceptual design and iterative development of a mHealth app by clinicians, patients and their families. Stud Health Technol Inform 2018;252:170-175. [Medline: 30040701]
31. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
32. NVivo. Lumivero. URL: https://lumivero.com/products/nvivo [accessed 2025-02-28]
33. Guilizzoni P. What is a wireframe? A guide for non-designers. Balsamiq. 2025. URL: https://balsamiq.com/learn/articles/what-are-wireframes [accessed 2025-02-28]
34. What is a wireframe? + How to create one. Cousera. URL: https://www.coursera.org/articles/wireframe [accessed 2025-02-28]
35. About draw.io. Draw.io. URL: https://www.drawio.com/about [accessed 2025-02-28]
36. Power BI. Microsoft. URL: https://www.microsoft.com/en-us/power-platform/products/power-bi [accessed 2025-02-28]
37. Agile Insights. URL: https://www.agile-insights.com.au [accessed 2025-02-28]
38. Jaspers MWM, Steen T, van den Bos C, Geenen M. The think aloud method: a guide to user interface design. Int J Med Inform 2004 Nov;73(11-12):781-795. [doi: 10.1016/j.ijmedinf.2004.08.003] [Medline: 15491929]
39. Brooke J. SUS: a "quick and dirty" usability scale. In: Usability Evaluation in Industry: CRC Press; 1996:189-194.
40. Almasi S, Bahaadinbeigy K, Ahmadi H, Sohrabei S, Rabiei R. Usability evaluation of dashboards: a systematic literature review of tools. Biomed Res Int 2023;2023:9990933. [doi: 10.1155/2023/9990933] [Medline: 36874923]
41. Lewis JR, Sauro J. The factor structure of the System Usability Scale. Presented at: First International Conference on Human Centered Design, HCD 2009; Jul 19-24, 2009. [doi: 10.1007/978-3-642-02806-9_12]
42. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
43. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4:114-123 [FREE Full text]
44. Mlaver E, Schnipper JL, Boxer RB, et al. User-centered collaborative design and development of an inpatient safety dashboard. Jt Comm J Qual Patient Saf 2017 Dec;43(12):676-685. [doi: 10.1016/j.jcjq.2017.05.010] [Medline: 29173289]
45. Millar RJ, Diehl C, Blake E, Fakeye O, Kusmaul N. Nursing home characteristics and resident quality of care outcomes: a scoping review. JLTC 2024. [doi: 10.31389/jltc.235]
46. Ghazisaeidi M, Safdari R, Torabi M, Mirzaee M, Farzi J, Goodini A. Development of performance dashboards in healthcare sector: key practical issues. Acta Inform Med 2015 Oct;23(5):317-321. [doi: 10.5455/aim.2015.23.317-321] [Medline: 26635442]
47. Randell R, Alvarado N, McVey L, et al. Requirements for a quality dashboard: lessons from national clinical audits. AMIA Annu Symp Proc 2019;2019:735-744. [Medline: 32308869]
48. van Deen WK, Cho ES, Pustolski K, et al. Involving end-users in the design of an audit and feedback intervention in the emergency department setting - a mixed methods study. BMC Health Serv Res 2019 Apr 29;19(1):270. [doi: 10.1186/s12913-019-4084-3] [Medline: 31035992]
49. Lester HE, Hannon KL, Campbell SM. Identifying unintended consequences of quality indicators: a qualitative study. BMJ Qual Saf 2011 Dec;20(12):1057-1061. [doi: 10.1136/bmjqs.2010.048371] [Medline: 21693464]
50. Cole AC, Adapa K, Khasawneh A, Richardson DR, Mazur L. Codesign approaches involving older adults in the development of electronic healthcare tools: a systematic review. BMJ Open 2022 Jul 6;12(7):e058390. [doi: 10.1136/bmjopen-2021-058390] [Medline: 35793923]

XSL•FO
RenderX

Abbreviations

HIT: health information technology
QI Program: National Aged Care Mandatory Quality Indicator Program
SUS: system usability scale

Edited by A Kushniruk; submitted 11.Jul.2025; peer-reviewed by MH Kuo, S Sohrabei; revised version received 12.Mar.2026; accepted 19.Mar.2026; published 29.Apr.2026.

Please cite as:
Dendere R, Stockwell-Smith G, Lang M, Hargrave M, Mayfield S, Gray LC
Development and Usability of a Dashboard for Quality Monitoring and Resident-Centered Care in Australian Residential Long-Term Care: Mixed-Methods Study
JMIR Hum Factors 2026;13:e80478
URL: https://humanfactors.jmir.org/2026/1/e80478
doi:10.2196/80478

© Ronald Dendere, Gillian Stockwell-Smith, Michelle Lang, Murray Hargrave, Sara Mayfield, Leonard Charles Gray. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 29.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Emotion-Adaptive Large Language Model–Driven Clinical Decision Support: User Evaluation of the Empathic Clinical Decision Support System Framework for Trust and Explainability

Tongze Zhang[1], MS; Sang Won Bae[1], PhD; Tammy Chung[2], PhD; Anind K Dey[3], PhD

[1]Department of Systems Engineering, Human-Computer Interaction and Human-Centered AI Systems Lab, AI for Healthcare Lab, Charles V. Schaefer, Jr. School of Engineering and Science, Stevens Institute of Technology, Hoboken, NJ, United States

[2]Institute for Health, Healthcare Policy and Aging Research, Rutgers Biomedical and Health Sciences, Rutgers University, Newark, NJ, United States

[3]Information School, University of Washington, Seattle, WA, United States

Corresponding Author:
Sang Won Bae, PhD
Department of Systems Engineering, Human-Computer Interaction and Human-Centered AI Systems Lab, AI for Healthcare Lab
Charles V. Schaefer, Jr. School of Engineering and Science
Stevens Institute of Technology
1 Castle Point Terrace
Hoboken, NJ, 07030
United States
Phone: 1 4122658616
Email: sbae4@stevens.edu

Related Articles:

Comment in: https://humanfactors.jmir.org/2026/1/e108778

Comment in: https://humanfactors.jmir.org/2026/1/e110739

Abstract

Background: The increasing prevalence of cannabis use has motivated researchers to develop computational behavioral models that predict usage patterns and related health impacts in naturalistic environments. However, the opaque nature of many artificial intelligence (AI) systems limits users' ability to interpret outputs and undermines trust. Existing explainable artificial intelligence techniques often remain overly technical and do not account for the confusion or frustration clinicians may experience when interpreting complex model explanations.

Objective: We propose and evaluate an empathic clinical decision support system (empathic-CDSS) that integrates large language models (LLMs) with real-time emotion recognition and explainability modules. The goal is to provide transparent, adaptive, and emotionally attuned explanations that enhance interpretability, user confidence, and trust in AI-assisted clinical decision-making.

Methods: Our empathic-CDSS integrates explainable artificial intelligence, causal inference, and affective computing within an interactive LLM-driven framework. Users' affective states were inferred using the circumplex model of affect, which characterizes emotions along 2 dimensions: valence (the degree of pleasure or displeasure) and arousal (the level of physiological activation or energy). Based on the captured emotional signals, the system dynamically adjusts the tone, structure, style, and complexity of its explanations generated by a fine-tuned LLM, enabling personalized and plain-language explanations. A total of 33 participants with diverse medical and technical backgrounds engaged with the system through guided evaluation tasks and postsession assessments to evaluate 6 dimensions of user-centered questions: usability, personalization and relevance of insights, clarity and comprehensibility, system benefits, satisfaction, and trust and reliability.

Results: Our empathic-CDSS effectively generated personalized and transparent explanations that revealed the causal reasoning behind model predictions while enhancing users' emotional engagement and trust in the system's decision logic. Continuous, affect-based feedback enabled the system to dynamically adapt explanation delivery to individual user needs. Participants reported significantly improved usability, clarity, satisfaction, and trust compared with a baseline clinical decision support system without

XSL•FO
RenderX

emotion adaptation. Improvements were also observed across additional evaluation dimensions, including personalization and perceived system benefits, supporting the feasibility and added value of integrating empathy-aware communication into AI-driven clinical decision support.

Conclusions: This study introduces a transparent, trustworthy, and emotionally adaptive framework for AI-assisted prediction and clinical decision support. By uniting causal reasoning, affective sensing, and LLM-based natural-language explanations, the empathic-CDSS offers a novel direction for developing emotionally intelligent and user-centered AI systems, with potential applications in behavioral monitoring and personalized interventions, including cannabis use and related health domains.

(JMIR Hum Factors 2026;13:e89005) doi:10.2196/89005

KEYWORDS

cannabis intoxication; cannabis-intoxicated behaviors; personalized intervention; large language models (LLMs); algorithmic decisions; transparency; healthcare AI; trustworthy AI; facial emotion recognition; mobile phone; artificial intelligence

Introduction

The integration of artificial intelligence (AI) and machine learning (ML) into health care has significantly advanced the development of clinical decision support systems (CDSSs) that can analyze complex health behaviors such as cannabis use. By combining multimodal data, including physiological, behavioral, and contextual information, AI models can provide real-time insights that assist clinicians in monitoring patient risks and treatment response. However, despite these advances, many existing CDSS systems still face challenges in terms of transparency and interpretability, and these issues remain key obstacles hindering their effective clinical application [1].

Given these limitations, explainable artificial intelligence (XAI) has been increasingly adopted to reveal how predictive models generate outputs (eg, feature-importance visualization, counterfactual explanations, and causal modeling). Yet, in practice, XAI outputs are often presented through technical plots, numeric weights, or attribution scores that clinicians without AI training may struggle to interpret. This can increase cognitive load and trigger negative affect, which paradoxically reduces trust in systems intended to improve transparency.

Despite rapid advances in XAI methods, a critical bottleneck persists in explanation delivery: how to convey model reasoning processes in practical clinical interpretation tasks in a manner that nontechnical clinicians can easily understand and use, rather than merely generating technically correct explanations. Cognitive factors such as mental workload and information complexity, as well as emotional factors such as trust, frustration, and confidence in AI recommendations, can significantly influence how clinicians interpret and rely on AI-generated insights.

Driven by the demand for more human-centered design, recent research has begun exploring how emotional and cognitive factors influence interactions between clinicians and AI systems. Several lines of research have attempted to improve human-AI interaction in CDSSs through better communication design and adaptive data presentation. For example, integrating sensor data from wearable and mobile devices has enhanced the ability of CDSSs to detect and predict patient health states in real time [1-4]. In parallel, advances in large language models (LLMs) have made it possible to process unstructured data and generate natural-language explanations [5-10]. Within the human-computer interaction community, LLM-driven conversational agents have been explored to improve medical reasoning, facilitate clinician-AI dialogue, and make AI-generated insights more accessible. However, despite their potential, most existing systems provide static, one-size-fits-all explanations that do not adapt to users' emotional or cognitive states. This limitation constrains their ability to sustain user engagement and trust, especially in emotionally demanding clinical contexts. Existing systems often treat interpretation as a fixed content generation process, underestimating health care professionals' tolerance for complexity and the dynamic shifts in emotional engagement and trust during moments of uncertainty, stress, or time pressure.

Emerging evidence from affective computing suggests that detecting and responding to users' emotions can substantially improve perceptions of system usability, reliability, and trustworthiness [11-14]. Negative emotions such as confusion or stress can reduce users' ability to comprehend complex outputs, whereas emotionally attuned interactions can promote engagement and confidence in AI systems. These dynamics are particularly relevant in health care, where clinicians must make rapid judgments under pressure and cognitive load. Complementary research in algorithmic empathy has shown that conversational LLMs can simulate empathetic dialogue. For instance, ChatGPT (OpenAI) has been observed to express empathy more often than physicians in patient-facing scenarios [15]. Nevertheless, such "synthetic empathy" can lack contextual awareness or amplify biases learned during training [16,17]. Moreover, most empathy-oriented AI systems rely solely on text-based sentiment analysis, overlooking multimodal affective signals such as valence (emotional pleasantness) and arousal (physiological activation) that form the foundation of real-time emotion modeling [18,19]. These limitations indicate a gap in developing CDSSs that dynamically adapt their communication based on clinicians' emotional states during AI interaction. To address these challenges, we developed the empathic clinical decision support system (empathic-CDSS), a novel framework that integrates XAI, affective computing, and LLMs to create an emotion-adaptive, explainable dialogue system for clinical decision-making. The system incorporates three core components: (1) a multimodal emotion-recognition module that analyzes facial behavior and linguistic sentiment to infer users' valence and arousal in real time, (2) an XAI engine that uses SHAP (Shapley Additive Explanations), causal reasoning, and

XSL•FO
RenderX

counterfactual analysis to generate transparent model explanations, and (3) an adaptive dialogue module powered by a fine-tuned LLM that dynamically adjusts the tone, complexity, and level of detail in its explanations based on the detected affective state. Through this integration, the empathic-CDSS aims to bridge the gap between technical explainability and emotional alignment, providing explanations that are both transparent and contextually sensitive to the user's experience. Based on this framework, this study aims to investigate whether such emotion-adaptive interpretability can enhance clinicians' trust, usability, and comprehension of AI-generated insights.

After developing the empathetic CDSS, we evaluated its impact on user perception and interaction experience. This study evaluates whether the empathic-CDSS enhances trust, usability, and interpretability among users with diverse clinical and technical backgrounds. Specifically, we asked: (1) How can real-time emotion recognition be integrated into an explainable dialogue system to guide adaptive responses during user-AI interaction? (2) How does emotion-adaptive explanation influence users' trust, satisfaction, and perceived reliability compared with a baseline CDSS? (3) How does the integration of personalized and real-time affective state awareness-based feedback impact the usability, interpretability, and accessibility of XAI for multidisciplinary users (clinicians and technical users)?

This study contributes to both the design and evaluation of emotion-adaptive XAI for clinical decision support. Specifically, it (1) introduces a novel framework (empathic-CDSS) that unites XAI, affective computing, and LLM-driven dialogue for empathic human-AI interaction; (2) demonstrates how valence-arousal emotion modeling can inform dynamic adaptation of AI explanations in real time; and (3) provides empirical evidence that emotion-adaptive explainability enhances trust, usability, and interpretability in clinical AI tools.

The remaining sections are organized as follows. The Methods section describes the empathic-CDSS architecture, participant recruitment, and evaluation procedure. The Results section reports quantitative and qualitative findings on usability, trust, and system satisfaction. The Discussion section interprets key results in the context of prior work and implications for empathic AI design, and the conclusions summarize contributions and directions for future research.

Methods

Study Design

We conducted a mixed-methods, within-subject experiment to evaluate the usability, trust, and explainability of two CDSS versions: (1) a baseline CDSS providing static explanations without emotion perception, and (2) an empathy-enhanced version integrating real-time emotion perception and adaptive explanation generation.

We designed and evaluated a novel CDSS named empathic-CDSS. The core functionality of empathic-CDSS integrates (1) analyses of multimodal patient data (eg, sensor data and self-report), with (2) real-time multimodal emotion recognition of the user who is using empathic-CDSS to deliver to the user, more transparent, empathic, and understandable explanations of the model's results.

For analyses of data provided by the "patient," the system is configured to analyze passive sensor data collected via smartphones and wearable devices to predict specific behavioral outcomes for the patient (eg, whether a cannabis user is in an "intoxicated" state). The system then uses multiple XAI techniques (including SHAP [20], causal analysis [21], and counterfactual reasoning) to provide explanations for these person-level predictions [22]. For further technical details on the model generation pipeline and the XAI-based interpretability results, please refer to Multimedia Appendix 1, particularly the sections titled Personalized Model Generation for Cannabis Users ("Patients") and Interpretability Analysis.

For the health care provider, or clinician or researcher who needs to interpret empathic-CDSS model outputs (ie, person-level prediction models), empathic-CDSS innovatively integrates, in real-time, multimodal emotion perception functionality as the user uses the system. Specifically, the system analyzes, in real time, users' facial expressions (extracting valence and arousal) and emotions from dialogue text entered in the system (eg, sentiment analysis), to capture users' emotional states. The system uses emotional insights from the user's facial expressions and dialogue text to dynamically adjust the style and complexity of the system's explanations of model outputs through a fine-tuned LLM, thereby providing a more empathetic and personalized interactive experience for the user (eg, the user using the system).

Participants

A total of 33 participants (P1-P33) took part in this study to represent the user who needs to interpret the empathic-CDSS model outputs. Participants were predominantly male (approximately 20/33, 70%), a majority aged 20-29 years, and most were highly educated (20 PhD students, 7 master's students, and several professionals such as physicians, nurses, and medical product managers). Participants were recruited through online advertisements and community networks. Eligibility criteria included adults aged 20 years and older. No age-based exclusion criteria were applied. Age distribution reflects voluntary enrollment during the recruitment period. Participants represented diverse disciplines, including engineering, computer science, psychology, and health care. Participants' experience in the medical field ranged from none to more than 20 years, providing a balanced mix of novice and expert perspectives for evaluating the explainable and emotion-aware AI system. This heterogeneous sample of system users was intentionally targeted to capture the interdisciplinary ecosystem in which clinical AI tools are typically developed and deployed, spanning both technically trained developers and nontechnical health care professionals. Including participants with varying levels of technical and clinical expertise allowed examination of differences across users in interpretability, emotional receptivity, and usability perceptions, thereby ensuring that the system's design and evaluation reflect real-world diversity among potential end users in clinical and decision-support contexts. Detailed demographic information about the participants is provided in Table 1.

XSL•FO
RenderX

Table 1. Participant demographics and professional background (N=33).

Characteristic and category	Values, n (%)
Gender	
Man	20 (60.6)
Woman	10 (30.3)
Prefer not to say	3 (9.1)
Age (years)	
20-29	20 (60.6)
30-39	6 (18.2)
50-60	7 (21.2)
Education level	
Bachelor's	3 (9.1)
Master's	9 (27.3)
PhD student	17 (51.5)
PhD	4 (12.1)
Occupation or discipline	
PhD or master's student (engineering, computer science, information systems, or psychology)	20 (60.6)
Medical product manager	4 (12.1)
Clinician or medical expert (physician, nurse, or counselor)	4 (12.1)
Researcher or research assistant	3 (9.1)
Other or unspecified	2 (6.1)
Experience in medical domain (years)	
None (0)	11 (33.3)
1-5	5 (15.2)
6-10	6 (18.2)
11-20	8 (24.2)
>20	3 (9.1)

Ethical Considerations

This study received approval from the Stevens Institute of Technology Institutional Review Board (2024-090 (N)) before recruitment and data collection began. Written informed consent was obtained from all participants. Participants who completed this study received a US $20 Amazon gift card as compensation.

In accordance with IRB requirements, participation was entirely voluntary. Subjects were informed of their right to withdraw from the laboratory session at any point without any negative consequences, though full completion was required for the compensation.

To protect participant privacy, all collected data (including survey responses and system logs) were anonymized immediately after collection. Personal identifiers were replaced with unique alphanumeric codes (e.g., P01, P02).

All research data were stored on secure, password-protected servers within the laboratory. Access was restricted to the primary research team to ensure that no data would be disclosed to third parties or leaked outside the institution.

Study Setting and Procedure

This study used a within-subjects design, in which each participant interacted with both versions of the system to enable direct comparison. To mitigate potential learning or order effects, the presentation order of the 2 systems was counterbalanced: approximately half of the participants were randomly assigned to use the basic system first, while the other half used the advanced system first.

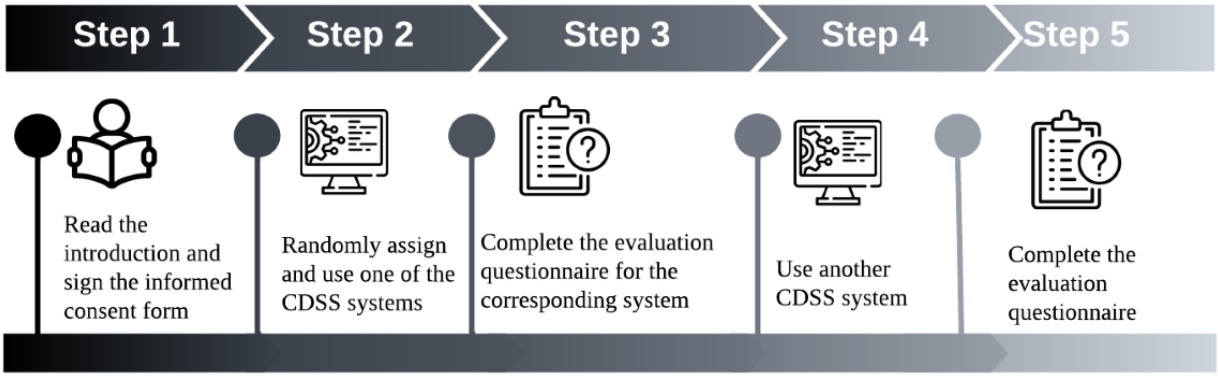
This study included 2 experimental conditions. The first was the basic-CDSS, which included the core study features, namely a personalized prediction model for cannabis intoxication and multiple XAI visualization charts (such as SHAP plots and causal diagrams). The dialogue interface was driven by a standard, not fine-tuned GPT-4 (OpenAI) model, and crucially, it did not possess any emotional perception capabilities, so its responses were purely information-oriented and did not adjust based on the user's emotional state. The second was the empathic-CDSS, which represents the complete empathic-CDSS proposed in this paper. Empathic-CDSS builds upon all the features of the basic-CDSS and integrates two core enhancement modules: (1) a multimodal emotion perception module for

XSL•FO
RenderX

real-time analysis of users' facial expressions and textual emotions; and (2) an empathy dialogue model, which has been fine-tuned on the motivational interviewing (MI) dataset [23] to generate emotion-adaptive responses.

Before any interaction with the system, participants were asked to review and complete an informed consent form that outlined this study's purpose, procedures, and their rights as participants. Participation was contingent upon signing this form. The participant then interacted with 1 of the 2 systems (randomly assigned). After completing their interaction with the system, the user filled out a comprehensive posttest questionnaire to evaluate their experience. The user then repeated the same process with the second system and filled out another evaluation questionnaire. At the end of the experiment, we asked participants to complete open-ended questions to collect qualitative feedback on the user experience for both systems. The user study procedure is illustrated in Figure 1.

Figure 1. User research tasks and questionnaire evaluation process. CDSS: clinical decision support system.



System Description: Empathic-CDSS Architecture and Processes

Basic-CDSS

The basic CDSS in this study serves as a control system, which aims to provide a standard, data-driven clinical decision support platform. The system primarily relies on traditional explainable ML analysis workflows, including structured data processing, clinical feature extraction, and interpretable presentation of model prediction results. Basic-CDSS uses generic XAI methods to explain model outputs, enabling users to understand the key factors behind predictions and use this information to inform clinical judgments. Compared to the empathic-CDSS developed in this study, basic-CDSS does not incorporate emotion recognition modules (such as text sentiment analysis or facial expression recognition) and does not use model fine-tuning with emotional strategies. Consequently, system interaction uses a standardized, nonemotion-adaptive interface presentation. Nevertheless, the system retains core analytical capabilities, including clinical data processing, model execution, and visual presentation of interpretive results.

Empathic-CDSS

The empathic-CDSS developed in this study aims to provide users with an intelligent interactive platform that integrates emotional understanding with XAI analysis. To do this, empathic-CDSS combines XAI methods with emotion-aware interaction design. The system combines multimodal emotion recognition involving text sentiment analysis and facial expression recognition with personalized, explainable ML models. Through this architecture, the system not only generates accurate and transparent analytical results but also presents them through empathetic, emotionally adaptive conversational interactions. To generate and support strategies for empathic responding during the conversational interaction, an MI dataset [23] was cleaned and used for web-based fine-tuning. This approach enhances users' trust in the system and improves their overall interaction experience.

In order to enable real-time emotion recognition for users interacting with the system and to allow ML models to process facial expressions quantitatively, the raw video data must first be converted into structured numerical features. These numerical representations enable the system to capture facial movements and emotional signals in a format suitable for computational analysis and subsequent modeling of emotion-aware interactions.

For real-time emotion recognition of the person while they are using the system, we first decoded the video frame-by-frame to convert the video into a sequence of still images. Each decoded frame was processed individually using our FacePsy [24] code to extract facial action units, head pose parameters (yaw, pitch, and roll), confidence scores for 7 emotion categories (angry, disgust, fear, happy, sad, surprise, and neutral), and 68 facial landmarks [20]. Specifically, action units are a set of features defined by the Facial Action Coding System that are used to quantitatively describe the activity of facial muscles. In addition, we extracted the confidence distribution of 7 basic emotions from the image: angry, disgust, fear, happy, sad, surprise, and neutral. These emotion confidences are inferred directly from facial features using a deep learning model, reflecting the likelihood of multiple emotions in each frame. To describe the geometry of the face, we used standard facial landmarking techniques to extract 68 facial landmarks (landmark0 to landmark67). These landmarks cover the precise locations of facial boundaries, eyebrows, eyes, nose, mouth, etc, providing a basis for capturing expression and emotion changes. The numerical combination of these features forms a high-dimensional feature vector for further analysis and

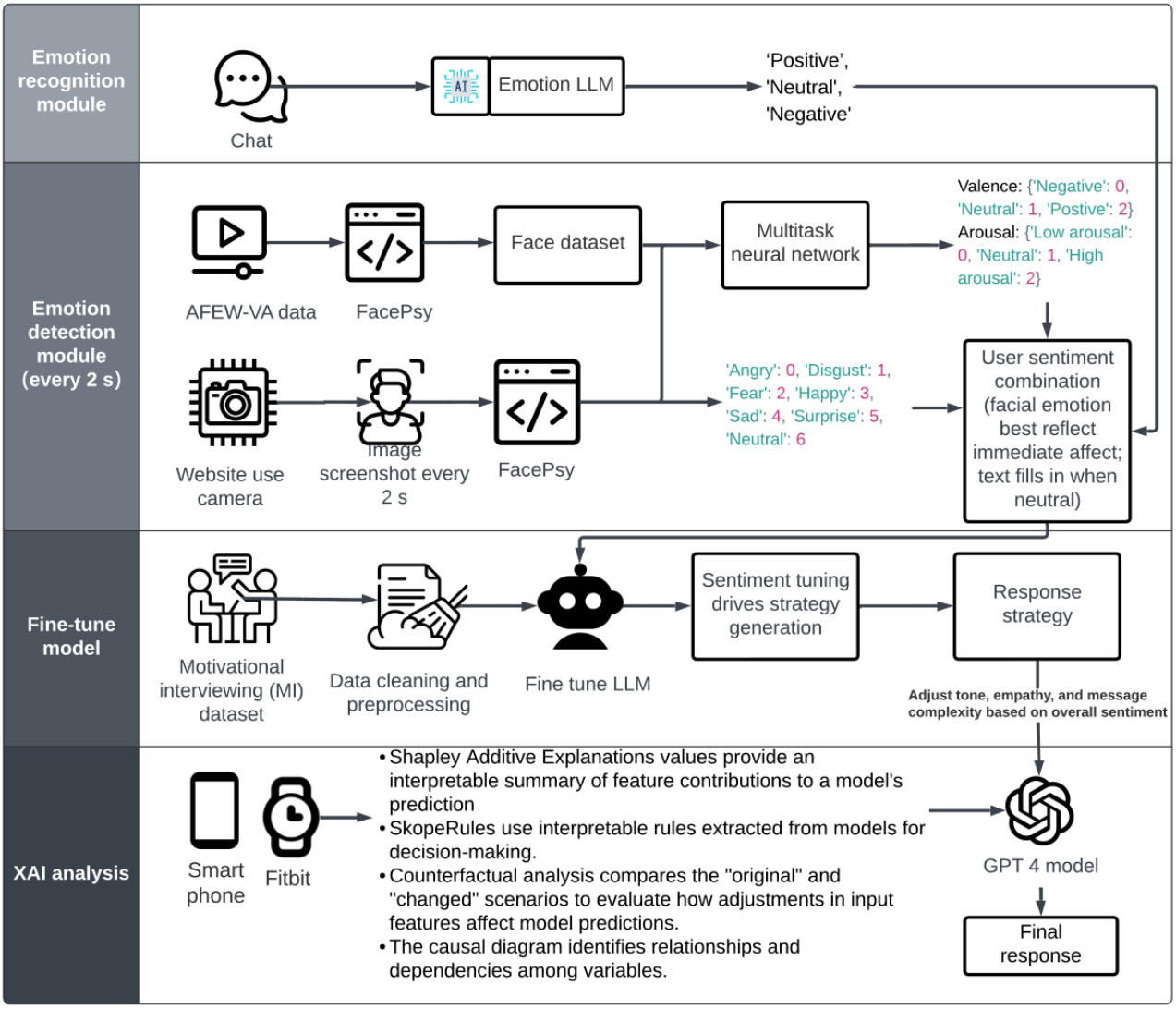
XSL•FO
RenderX

modeling. Examples of facial behavior features are provided in Multimedia Appendix 2.

Figure 2 shows the overall architecture and data processing flow of this system. Using data from smartphone and wearable device (eg, Fitbit [Google LLC]) collected in a separate study from individuals who use cannabis [25], the system combined several explanatory AI models (eg, SHAP, SkopeRules, and counterfactual analysis) and causal analysis to generate the corresponding explanatory analysis [20-22,26]. The process involves user or clinician (hereafter: "clinician") input to the system (eg, asking the system a question), AI model processing, interpretable analysis generation, and finally, feedback provided to the user.

In order to keep track of the user's interaction history and analysis results, the system is designed with a data persistence mechanism. Each user input, system-generated response, and corresponding timestamp are recorded and stored. This not only supports the system's session retention function but also provides valuable historical data for subsequent data analysis and improvement.

Figure 2. Empathic-CDSS architecture and processes. AFEW-VA: Acted Facial Expressions in the Wild – Valence-Arousal (a publicly available dataset for affective computing); CDSS: clinical decision support system; LLM: large language model; MI: motivational interviewing; SHAP: Shapley Additive Explanations; XAI: explainable artificial intelligence.



Measures

Participants completed a structured evaluation comprising 6 dimensions: system usability [6], personalization and relevance [27], clarity and comprehensibility [5], system benefits [3], overall satisfaction [9,16], and trust and reliability [17,28]. Table 2 shows the constructs for each dimension. Each item was rated on a 10-point Likert scale (1=lowest, 10=highest). Example items include "How easy was it to interact with the system using natural language?" and "How clear and understandable were the explanations provided by the system?" The complete questionnaire is available in Table 2 and Multimedia Appendix 3.

XSL•FO
RenderX

Table 2. Survey constructs and example items.

Section	Construct measured	Example item (scale range: Likert 1-10)	References
System usability	Ease of interaction, intuitiveness, and satisfaction	"How easy was it to interact with the system using natural language?"	[6]
Personalization and relevance	Perceived individualization of system insights	"To what extent did you feel the insights were personalized to patient data?"	[27]
Clarity and comprehensibility	Understandability of explanations and visuals	"How clear and understandable were the explanations provided by the system?"	[5]
System benefits	Perceived usefulness and outcome impact	"How beneficial do you believe the insights were in improving understanding of patient health behaviors?"	[3]
User satisfaction	Satisfaction	"How satisfied are you with the overall quality of the insights and recommendations provided by the system?"	[9,16]
Trust and reliability	Trust and perceived reliability	"Please rate the perceived trustworthiness of this system in meeting your needs."	[17,28]

Data Collection

Cannabis Use Data from "Patients"

We used the AWARE framework app [29] to track cannabis users' smartphone sensor data, such as GPS and accelerometer, alongside Fitbit Charge 2 data for heart rate and step count. The data were analyzed to identify patterns of cannabis use, with physiological and behavioral data segmented into 5-minute intervals. Key statistics, such as heart rate and step count, were extracted to evaluate cannabis intoxication. Events were classified into "intoxicated" and "not intoxicated" to test the system's ability to predict and interpret behavioral outcomes from sensor inputs [25]. These ground truth labels were obtained through daily ecological momentary assessment reports, where participants self-reported their current intoxication status upon receiving a prompt on their smartphone. This dataset comprises passively collected smartphone and wearable device data linked to self-reported cannabis intoxication episodes. After preprocessing, the data were segmented into annotated samples representing intoxicated and nonintoxicated states. As the data was gathered in natural settings, the class distributions reflect real-world reporting patterns rather than artificially balanced designs. To ensure model training consistency and robustness, standard preprocessing steps were implemented, including normalization, missing value handling, and participant-level modeling.

User Chat Data With the CDSS

For the chat data, user interactions, and generated responses with the CDSS were logged in Google Sheets (Google LLC). The data were structured with fields for email, timestamp, role (clinician or CDSS assistant), and content [30]. This structure allows for persistent storage of conversation history, which can be retrieved and used to maintain context in the ongoing session.

Data Analysis

Quantitative

The quantitative data for this study were derived from clinical practitioners' ratings of the systems on a 1-10 Likert scale following their experience with both the baseline-CDSS and empathic-CDSS. The rating scale encompassed multiple dimensions, including system usability, personalization and relevance, clarity of explanations, system benefits, user satisfaction, and trust and reliability. As the same group of users evaluated both systems, paired *t* tests were conducted for each questionnaire item to assess differences between the empathic-CDSS and baseline-CDSS scores. For each item, the program automatically aligned valid paired samples across both systems, calculating the mean score, SD, *t* statistic, and *P* value for both empathic-CDSS and baseline-CDSS.

Qualitative

Open-ended questions in the questionnaire were collected under the empathic-CDSS condition to supplement users' subjective experiences and perceptions of the system. These responses to open-ended questions primarily serve a descriptive purpose and summarization, aiding discussions of quantitative findings by illustrating how users interpret the system's explanatory approaches, personalized presentations, and emotion-adaptive behaviors.

ML or System Evaluation

At the natural language processing and dialogue generation level, both systems support users posing questions to the CDSS using natural language. The system first performs semantic parsing of user input, mapping natural language into retrieval requests for backend models and data stores. It extracts intent categories and slot information related to patient timeframes and metric types. Empathic-CDSS further incorporates sentiment analysis and emotion-adaptive modules. It uses a sentiment classifier based on pretrained language models to identify emotions in input text, outputting sentiment polarity and coarse-grained emotion categories (eg, neutral, confused, and concerned) as additional conditions for dialogue planning. Subsequently, the dialogue generation module integrates backend ML predictions, XAI explanations (eg, SHAP feature contribution rankings, key curves, and temporal trend charts), and sentiment analysis signals to generate responses that align with model outputs while adapting to the user's current emotional state: when detecting higher uncertainty or stress, the system uses more step-by-step, detailed explanatory structures and a more supportive tone; while using more concise, summary-style answers during neutral or positive emotional states. Baseline-CDSS uses the same cannabis intoxication prediction and XAI pipeline but presents results in a uniform,

XSL•FO
RenderX

nonemotion-adaptive format, excluding sentiment analysis or linguistic style adaptation. XAI model reliability is evaluated by separating smoking session data, with a portion used for model testing and the remaining samples for model training. Predictions are validated within retained sessions to ensure stable personalized performance, thereby enabling interpretable generation for CDSS.

Model of Emotion Valence and Arousal

We constructed an efficient model training framework for multiclassification problems with dual labels. All samples in the dataset were labeled, and the objective was to improve the training efficiency and generalization ability of the model while ensuring classification performance. Due to the large size of the training set, using all samples directly may cause the model to be disturbed by redundant information during the learning process. Therefore, we systematically designed the model training process through cross-validation (5-fold cross-validation) and feature preprocessing to improve model performance and robustness.

In the data preprocessing stage, we cleaned and converted missing values, nonnumeric fields, and coordinate-type features in the data. Coordinate information was parsed from string format into numeric vectors and expanded into separate feature columns. Subsequently, we standardized all input features to eliminate the influence of different units of measurement and ensure that the model had stable gradients and convergence speed during training.

This study used a multitask neural network (multitask neural network structure) [31] to simultaneously perform 3-class classification prediction for 2 emotion dimensions, valence (negative, neutral, and positive) and arousal (low arousal, neutral, and high arousal) [32]. This structure introduced shared hidden layers, enabling the model to share representation learning between 2 related tasks, thereby improving training efficiency and enhancing generalization ability.

The shared feature extraction part consists of 2 fully connected networks, combined with batch normalization, ReLU (rectified linear unit; an activation function commonly used in deep learning models, including facial recognition systems) activation functions, and dropout, which can effectively enhance the model's expressive ability while reducing the risk of overfitting [33]. The 2 output branches use independent fully connected layers to predict the arousal and valence labels, respectively.

During training, we used the cross-entropy loss function [34] to supervise each output branch. The final optimization objective was the weighted sum of the 2 losses, expressed as follows: $L=\lambda_1 L_{(Arousal)}+\lambda_2 L_{(Valence)}$, $L_{(Arousal)}$ and $L_{(Valence)}$ represent the cross-entropy losses for arousal and valence, respectively. This structure demonstrates robust performance and classification accuracy when handling 2 emotional label tasks simultaneously.

We used the Adam optimizer to minimize this loss function during training [35]. The model updates parameters through backpropagation and iteratively optimizes the performance of the 2 tasks. This structure demonstrates good efficiency and performance in multilabel multiclassification problems, capturing both shared information between tasks and task-specific discriminative features.

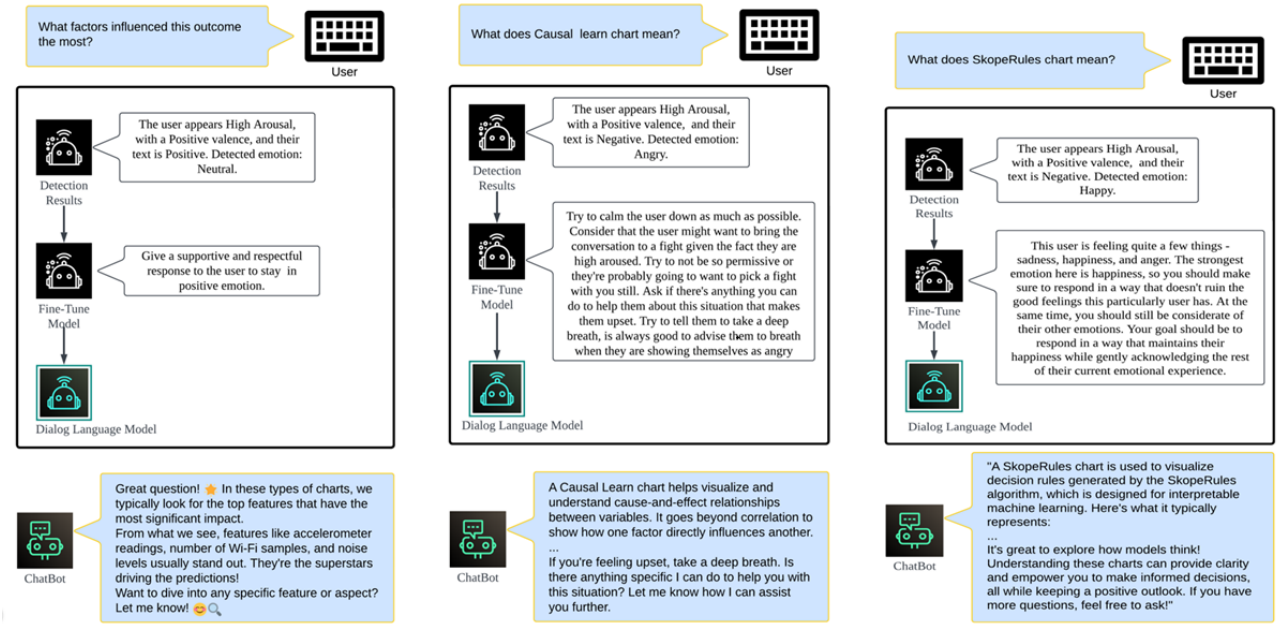
Fine-Tuning Model and Dual Language Model Architecture

To achieve more empathic and context-aware responses, the system uses a dual language model architecture that leverages both a standard GPT-4 model [36] and a fine-tuned GPT-4 model [37]. All LLM functionalities in this study, including standard GPT-4 calls and model fine-tuning, were implemented via the official application programming interfaces provided by OpenAI. This design ensures that the system combines XAI analysis with emotional context for a comprehensive and user-centered interaction. The first language model is the standard GPT-4, which is used to process the output generated by the XAI component. The model interprets the results of XAI analyses, such as feature importance, counterfactual reasoning, or causal inference graphs, and extracts key insights. The second language model is a fine-tuned GPT-4 that has been specifically trained using an MI dataset [23]. This MI dataset provides the basis for equipping the model with the ability to recognize and respond to users' emotional states. The fine-tuning process enables the model to generate responses that adhere to the principles of communicating empathy and support, based on the results of the sentiment analysis. This is crucial for building engagement into the interaction.

The fine-tuning process included preprocessing the dataset to ensure data quality and compatibility with the GPT architecture. Noise was eliminated, annotations of emotional states were aligned, and tokenization was optimized to preserve empathetic communication. These steps enabled the model to learn conversational patterns that emphasize understanding, validation, and support. Through this process, the model generates responses that match the user's psychological needs based on the emotional valence and arousal, as well as the sentiment of the user's text. The fine-tuning process focuses on enabling the model to recognize subtle emotional cues in user input and generate responses that adhere to the principles of empathy and support. By incorporating real-world patterns of conversation from the MI dataset, the system learns to address the user's concerns in a compassionate manner, which promotes engagement. This fine-tuning approach enhances the system's ability to dynamically adapt to different emotional contexts, ensuring that the user not only receives accurate explanations, but also emotionally resonant responses that make interactions more human-centered. Fine-tuned by integrating a corpus of MI data, the system now strikes a good balance between empathetic communication and clear explanations. This approach not only improves the user's experience but also reinforces the trustworthiness of the system as an XAI interaction. The system integrates multiple high-level components, including emotion recognition and natural language processing models, to achieve its core functionality. One model detects the user's emotional state through a combination of text sentiment analysis and facial expression recognition. Then, another model generates personalized and context-aware responses based on the detected emotion, as shown in Figure 3.

XSL•FO
RenderX

Figure 3. Examples of emotion detection language model.



Dialog System

The system provides an intuitive way to present complex data analyses directly to the user. Data tables are presented visually in the interface, eliminating unnecessary navigation steps. To facilitate understanding, the contents of these tables are displayed immediately without additional manipulation. The system automatically converts the analyzed data into visualizations such as SHAP plots, cause-effect plots, and rule interpretation plots. These graphical outputs greatly enhance the clarity of the information, while the images are encoded in base64 and embedded in the web page to enrich the visual experience [38]. Each table is accompanied by a detailed explanation, including the names of the features and their respective importance or influence. In addition, the system can seamlessly convert the XAI results in the data table into images and embed them in the user interface. This approach not only improves the readability of the data, but also helps the user understand the analysis results in a more intuitive way. The system interface is shown in Figure 4.

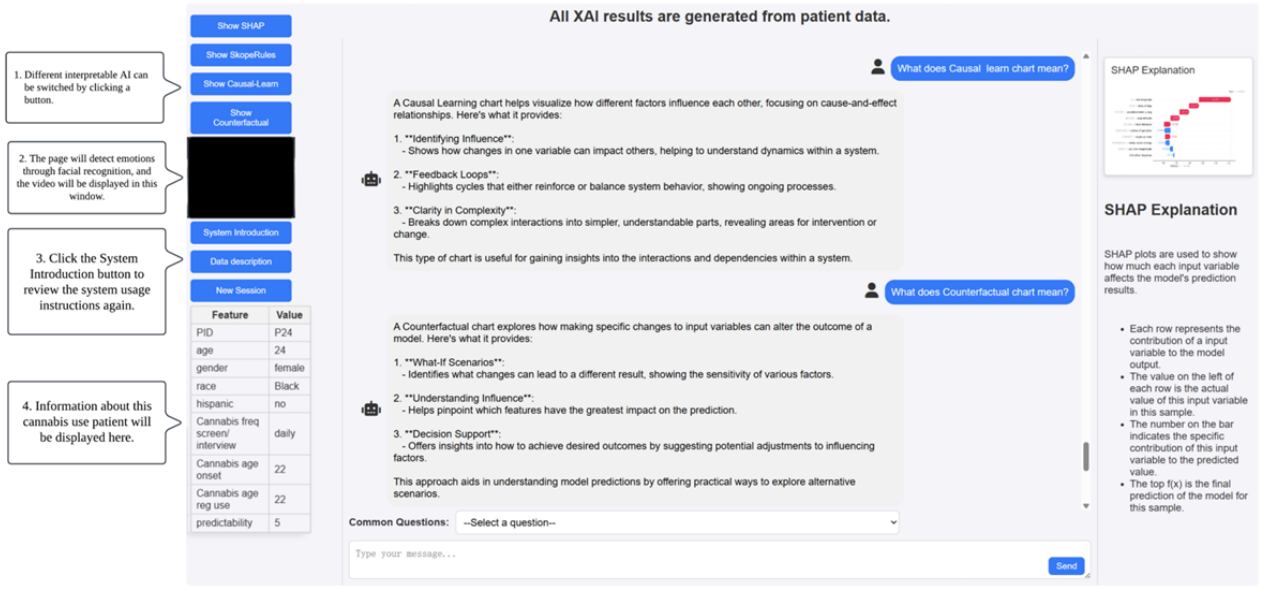
The system allows users to interact via a natural language interface, and each message is subject to a multistep sentiment analysis. The system uses 3 different mechanisms to assess the emotional state of the user. First, the sentiment of the input text is analyzed and classified as positive, negative, or neutral. At the same time, when the camera is active under the empathic-CDSS condition, facial expression recognition is used to assess the visible emotions of the user. In addition, indicators of emotional valence and arousal are calculated to capture the intensity and type of emotion, providing a comprehensive understanding of the user's emotional context.

We use a dual model architecture to process the results of 3 sentiment analysis methods (text sentiment, facial expression recognition, and value arousal indicators). The fine-tuned model only receives the sentimental results and generates preliminary analysis and corresponding response guidance based on these results. This response guidance is then passed to the GPT model, which takes the user's question and analysis results as input to generate the final response. This step-by-step approach ensures that the generated response is not only consistent with the sentiment context, but also effectively explains the XAI output in a user-friendly and sentiment-intelligent manner.

To improve the quality of the interaction, the system keeps a detailed dialog history and ensures that the conversation is consistent. The system can adapt its responses based on previous exchanges, making the interaction more personalized and meaningful. Finally, the system presents feedback through a combination of text, visuals, and emotion summaries. These outputs not only provide explanations for the model's decisions but also include explanatory images or diagrams to help the user understand the system's behavior intuitively. By integrating text sentiment analysis, facial expression recognition, and value arousal metrics, the system can respond in a user-centered manner.

XSL•FO
RenderX

Figure 4. Empathic-CDSS demonstration. AI: artificial intelligence; CDSS: clinical decision support system; PID: participant ID; SHAP: Shapley Additive Explanations; XAI: explainable artificial intelligence.



Results

Performance of Personalized Models for Predicting Cannabis Intoxication

We first assessed the performance of the personalized predictive model designed to estimate cannabis intoxication. To approximate real-world prediction conditions, we used a personalized evaluation strategy in which self-reported cannabis intoxication from selected individuals served as the prediction target. A total of 3 representative cannabis users were selected, and an individualized model was trained and evaluated for each user. For each model, 1 day of self-reported "cannabis use" was used as the test set, while all remaining historical data served as the training set.

Across all 3 users, the personalized models demonstrated strong predictive capability, achieving F_1-scores exceeding 70% in independent tests. These results indicate that the models effectively captured individual behavioral patterns and generated reliable predictions of future intoxication states. The prediction outputs and corresponding XAI explanations from these tests were subsequently incorporated into the CDSS interface presented to users during the comparative evaluation of the basic and empathic systems.

Affective Sensing Model Performance

We next evaluated the performance of the emotion-perception component of the system. Using the AFEW-VA (Acted Facial Expressions in the Wild – Valence-Arousal [a publicly available dataset for affective computing]) dataset [39], we conducted 5-fold cross-validation to assess arousal and valence prediction accuracy. As summarized in Table 3, the model achieved strong overall performance across folds. For arousal prediction, the model achieved an average F_1-score of 0.75 (SD 0.0086) and an accuracy of 88.16% (SD 0.0020). For valence prediction, the average F_1-score was 0.70 (SD 0.1173), with an accuracy of 74.39% (SD 0.0075).

To deploy the affective sensing module within the empathic-CDSS, we selected the best-performing model fold based on overall cross-validation metrics. The detailed performance of this selected fold is presented in Table 4.

As shown in Table 4, the selected model achieved an F_1-score of 0.75 with an accuracy of 88.03% for arousal prediction, and an F_1-score of 0.71 with an accuracy of 74.78% for valence prediction. These metrics confirm the effectiveness and reliability of the affective sensing module integrated into our empathic-CDSS.

Table 3. Average performance metrics of the affective sensing model.

Metrics	Arousal	Valence	Average (SD)
F_1-score	0.75	0.70	0.73 (0.035)
Precision	0.82	0.74	0.78 (0.057)
Recall	0.70	0.68	0.69 (0.014)
AUC[a]	0.93	0.87	0.90 (0.042)
Accuracy (%)	88.16	74.39	81.28 (9.75)

[a]AUC: area under the curve.

XSL•FO
RenderX

Table 4. Performance metrics of the best fold affective sensing model.

Metrics	Arousal	Valence	Average (SD)
F_1-score	0.75	0.71	0.73 (0.028)
Precision	0.80	0.74	0.77 (0.042)
Recall	0.71	0.70	0.71 (0.007)
AUC[a]	0.93	0.87	0.90 (0.042)
Accuracy (%)	88.03	74.78	81.41 (9.37)

[a]AUC: area under the curve.

Overall System Performance

We systematically evaluated our emotion-aware approach, which integrates continuous valence-arousal modeling with LLM-based dialog generation. The results indicate that the use of continuous affective representations substantially improves the system's ability to detect, interpret, and respond to subtle user sentiments during real-time interactions. Compared with traditional classification-based models, our approach enabled more contextually appropriate, emotionally consistent, and empathetic responses.

User feedback also confirmed smoother conversational flow, fewer emotionally inconsistent replies, and improved alignment with user intent. Unlike conventional affective models that depend on fixed emotion categories, our system adapts dynamically to changes in user emotional state, leading to more natural and personalized interactions.

Quantitative Comparison Between Systems

Multimedia Appendix 3 outlines the evaluation areas included in this study. To quantify differences between the empathic-CDSS and the basic CDSS, we conducted paired-sample *t* tests for all questionnaire items, reflecting within-participant comparisons between the 2 system conditions (Table 5).

XSL•FO
RenderX

Table 5. Comparative results across usability and perception metrics between the empathic- and basic-CDSS conditions. All *P* values follow JMIR formatting guidelines; *P*<.001 denotes highly significant differences at the α=.05 level.

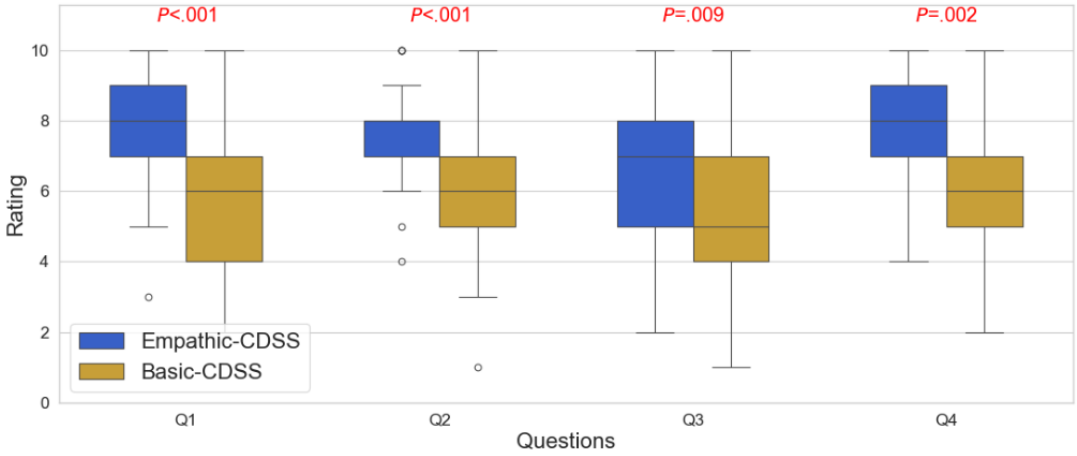
Dimension and question	Empathic mean (SD)	Basic mean (SD)	*t (32)*	*P* value
System usability				
Q1	8.00 (1.64)	6.12 (2.25)	4.178	<.001
Q2	7.73 (1.38)	5.88 (2.07)	4.245	<.001
Q3	6.67 (2.35)	5.21 (2.32)	2.784	.009
Q4	7.61 (1.71)	6.09 (2.24)	3.318	.002
Personalization and relevance				
Q1	7.76 (1.75)	6.39 (1.89)	3.090	.004
Q2	7.58 (1.46)	5.76 (1.79)	4.942	<.001
Q3	7.64 (1.41)	5.73 (1.97)	4.955	<.001
Clarity and comprehensibility				
Q1	8.09 (1.53)	6.39 (2.00)	4.430	<.001
Q2	8.03 (1.59)	6.30 (2.40)	3.713	<.001
Q3	8.24 (1.20)	6.55 (2.20)	4.033	<.001
Q4	7.36 (1.75)	5.58 (1.85)	3.835	<.001
Q5	7.15 (1.72)	6.12 (1.87)	2.489	.02
System benefits				
Q1	8.03 (1.38)	6.61 (1.78)	4.155	<.001
Q2	7.64 (1.62)	6.39 (2.22)	2.543	.02
Q3	7.73 (1.40)	6.45 (2.18)	2.950	.006
Q4	7.27 (2.15)	6.15 (2.29)	2.211	.03
User satisfaction				
Q1	7.76 (1.20)	5.85 (2.00)	4.357	<.001
Q2	7.64 (1.43)	6.06 (2.00)	3.859	<.001
Trust and reliability				
Q1	7.85 (1.20)	5.91 (1.59)	5.329	<.001
Q2	7.79 (1.17)	6.12 (1.62)	5.133	<.001

System Usability

As shown in Figure 5, the empathic-CDSS demonstrated significantly higher usability across all items. For example, users rated the ability to “interact with the system using natural language” higher for the empathic-CDSS (mean 8.00, SD 1.64) compared with the baseline system (mean 6.12, SD 2.25; *P*<.001). Significant improvements were also observed for interface intuitiveness (Q2), response time satisfaction (Q3), and accurate data interpretation (Q4; all *P*<.05). These findings suggest that incorporating emotional-perception feedback results in smoother, more natural clinician-AI interactions.

XSL•FO
RenderX

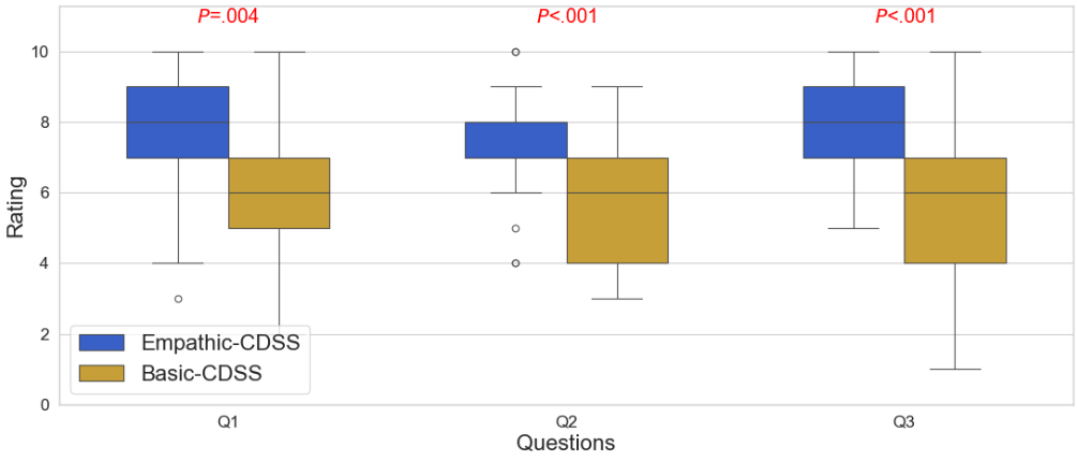
Figure 5. Comparative results across system usability. CDSS: clinical decision support system.



Personalization and Relevance of Insights

The empathic-CDSS also outperformed the baseline system in personalization and relevance (Figure 6). For example, users rated the system's ability to consider the patient's current condition significantly higher (mean 7.64, SD 1.41 vs 5.73, SD 1.97; *P*<.001). The empathic-CDSS also showed superior performance in handling missing information and referencing appropriate individual-level data (eg, Q5: 7.15 vs 6.12; *P*=.02). These results highlight improved adaptability to patient-specific information and more accurate context-aware responses.

Figure 6. Comparative results across personalization and relevance. CDSS: clinical decision support system.

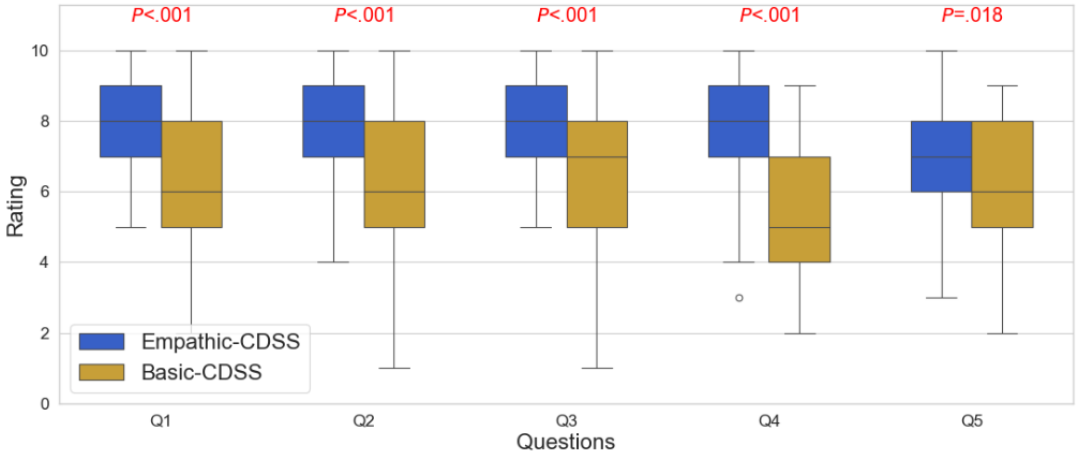


Clarity and Comprehensibility of Explanations

The empathic-CDSS achieved significantly higher scores (overall mean 7.77, SD 0.49 vs 6.19, SD 0.41) in clarity and comprehensibility compared to the baseline (Figure 7). Participants rated the clarity of explanations provided by the system significantly higher (mean 8.09, SD 1.53, vs 6.39, SD 2.00; *P*<.001). Specifically, the transition from raw technical SHAP data to LLM-generated plain-language narratives helped clinicians better grasp the causal reasoning behind the model's predictions.

XSL•FO
RenderX

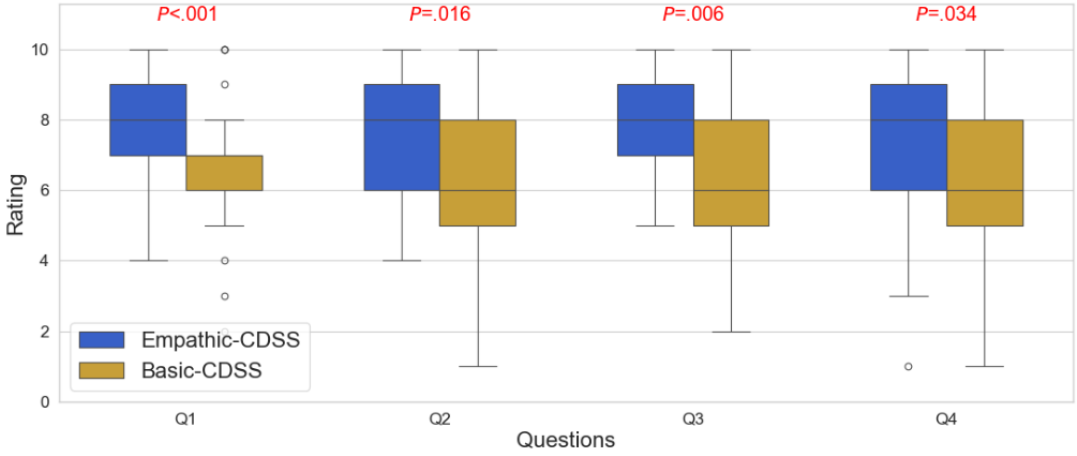
Figure 7. Comparative results across clarity and comprehensibility. CDSS: clinical decision support system.



System Benefits

As shown in Figure 8, the optimized model showed significant advantages across all 4 benefit-related metrics. Participants rated the system as more helpful for understanding patient health behaviors (Q1: mean 8.03, SD 1.38, vs 6.61, SD 1.78; *P*<.001). They were also more likely to recommend the system to others (Q3; *P*=.006). The optimized system further reduced misinformation (Q4), reflecting improvements in output accuracy and clarity.

Figure 8. Comparative results across system benefits. CDSS: clinical decision support system.



User Satisfaction, Trust, and Reliability

The empathic-CDSS consistently achieved higher satisfaction, trust, and reliability scores (Figures 9 and 10). All satisfaction-related differences were statistically significant (all *P*<.001). For example, users reported a higher likelihood of continued system use (mean 7.76, SD 1.20, vs 5.58, SD 2.00). Trust-related items also favored the empathic-CDSS, with scores above 7.7 on trustworthiness and overall reliability, compared with approximately 6.1 for the baseline model.

XSL•FO
RenderX

Figure 9. Comparative results across user satisfaction. CDSS: clinical decision support system.

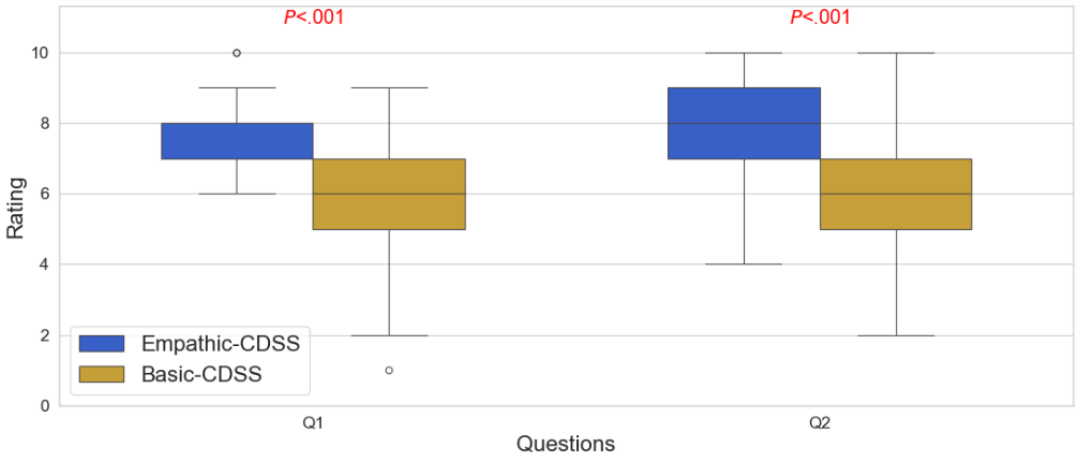
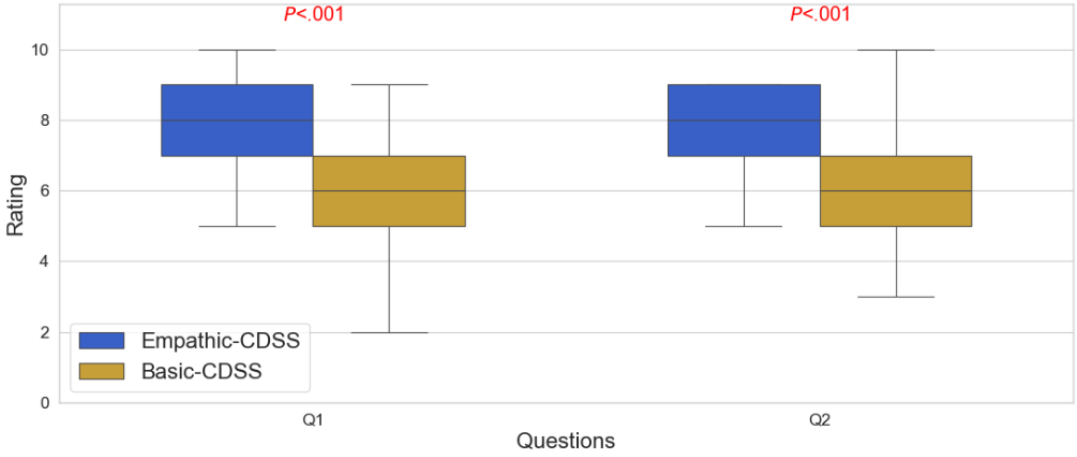


Figure 10. Comparative results across trust and reliability. CDSS: clinical decision support system.



Summary of Findings

Overall, results demonstrate that the empathic-CDSS substantially outperforms the baseline system in usability, interpretability, personalization, trustworthiness, and perceived usefulness. The combination of multimodal sentiment recognition, MI-informed fine-tuning, and XAI-driven decision support significantly enhanced user experience and increased acceptance of the system. These findings emphasize the importance of human-centered AI design and highlight the value of adaptive, affective, and context-aware interactions in improving clinical decision-support performance.

System Impact on User Emotion and Stress

To assess the systems' impact on users' emotional states, we compared self-reported emotion and stress levels before and after interaction with the basic-CDSS and the empathic-CDSS. Most participants reported neutral emotions in both conditions. A small number of mood changes were observed: after using the empathic-CDSS, 1 participant (P14) reported a shift from neutral to happiness, whereas after using the basic-CDSS, 1 participant (P21) reported a shift from happiness to neutral. This same participant showed no mood change when using the empathic-CDSS.

Stress levels also showed modest differences across conditions. Under the basic-CDSS, stress increased for several participants (for example, P21 increased from 2 to 4, and P12 increased from 4 to 6). In contrast, stress levels remained stable or decreased for most participants when using the empathic-CDSS (for example, P22 decreased from 6 to 5, and P31 decreased from 4 to 3). Overall, the basic-CDSS was associated with stress increases in 4 participants and decreases in 7 participants (mean change –0.76, SD 1.51). The empathic-CDSS produced only 1 stress increase, with a similar mean reduction (–0.84, SD 1.73), and stress remained unchanged for most participants.

Summary of Qualitative Feedback

Users provided generally positive feedback regarding the system's interactivity, natural language capabilities, and clarity of explanations. Many participants highlighted the system's ability to support conversational flexibility and deliver rapid, data-driven insights. For instance, P10 praised the system's natural dialog style, noting: "Any question can be asked, and the system appears to understand." Similarly, P11 commented that the system "can provide clear and data-driven insights based on the user's question," emphasizing trust in its analytical outputs.

Despite this positive reception, users also identified limitations related to explanation depth and interpretability. P11 remarked

XSL•FO
RenderX

that "the explanations are not very detailed," and P13 noted that "the decision path is a bit confusing," suggesting that the clarity of complex decision pathways could be enhanced. Additionally, P14 observed that "the sentiment seems to affect the output by a large margin," indicating that the weighting of the sentiment module may at times overly influence the system's responses, an important design consideration for future refinement.

Participants also suggested specific interface and usability improvements. For example, P12 recommended adjusting button colors for better visibility, while P13 expressed a desire for a more readable output structure to improve comprehension of results. Despite these suggestions, users consistently recognized the system's potential value. P11 highlighted that the empathic-CDSS's "step-by-step recommendations" offer clear guidance, and P10 suggested that the system "may assist in improving care for a patient." Regarding the XAI component, P14 stated: "The ability to understand how the model makes a decision is helpful for patients and doctors." This feedback reflects users' cautious optimism—while acknowledging areas for improvement, they recognize the importance and usefulness of transparent explanation in clinical decision support.

Discussion

Principal Findings

This study shows that the proposed empathic-CDSS, which integrates multimodal emotion perception with adaptive empathic response generation, outperformed the baseline CDSS that lacked emotion-awareness of the system's user. The empathic-CDSS produced higher scores in usability, clarity, personalization, trust, and user satisfaction. These improvements are consistent with prior research indicating that emotionally adaptive technologies can reduce cognitive load, improve comprehension, and enhance trust in AI systems [40-44].

Our findings confirm that empathetic adaptation significantly enhances system credibility, reliability, and user satisfaction. When systems recognize user emotional states, users report higher levels of system confidence—aligning with the need for more human-centered clinical AI. The integration of real-time emotion perception significantly enhances system usability, clarity, and comprehensibility while boosting the personalization and relevance of insights generated by the system. By dynamically aligning explanation complexity with user emotional states, the system successfully translates technical XAI outputs into interpretable, plain-language clinical insights.

Higher scores in clarity and comprehensibility (mean 8.09, SD 1.53, vs 6.39, SD 2.00, $P<.001$), with a large effect size, demonstrate the substantial impact of this approach on user understanding. This represents a meaningful improvement in clinicians' comprehension of model predictions rather than a marginal gain. These findings indicate that adaptive explanations with emotional perception not only enhance subjective satisfaction but also improve clinicians' ability to interpret AI predictions. By integrating causal reasoning outputs with plain-language explanations generated by LLMs, the system facilitates a clearer understanding of predictive logic and enhances the interpretability of AI-assisted decision-making. LLMs serve as a bridge, translating complex SHAP feature weights and technical causal graphs into narrative lay summaries, making them more accessible to clinicians without deep AI expertise. This shift from static technical visualizations to dynamic empathetic dialogue effectively mitigates the "black box" nature of clinical AI, thereby enhancing trust and reliability. Prior work, consistent with our observations, indicates that emotion-adaptive explanation strategies can improve user engagement, reduce frustration, and lower perceived cognitive load during complex analytical interactions [45-48].

This study's results are consistent with studies showing that empathic responses can stabilize emotional states and reduce interactional friction during demanding tasks [49]. Participants in our study frequently described the empathic-CDSS (vs basic system) as clearer, easier to engage with, and more understanding. These perceptions likely contributed to the significantly higher ratings of trustworthiness and reliability. Trust is widely recognized as a crucial factor for clinical adoption of AI systems [41,43].

Comparison With Prior Work

Most existing CDSS and XAI frameworks focus on improving algorithmic transparency while paying little attention to how explanations are communicated to clinicians or system users [50-52]. In many systems, explanation generation is treated as a static process that delivers identical information regardless of user context or emotional state [51,53]. Our findings challenge this assumption. Consistent with recent work in affective computing and empathy-aware conversational agents [40-42], our study demonstrates that dynamically adapting the communication style of explanations can significantly improve user comprehension and interaction quality.

Unlike earlier empathic-CDSS prototypes that relied on rule-based or predefined text responses, our innovative system incorporates real-time arousal and valence detection. These signals allow the LLM to adjust explanation complexity, tone, and pacing in response to subtle emotional changes during interaction. This novel feature addresses a key limitation of static XAI approaches, which do not respond to users' evolving cognitive or emotional states. Our findings also extend prior research by applying empathy-aware interaction to the problem of helping nontechnical users interpret machine-learning explanations, which represents a new and challenging scenario.

In addition, integrating XAI technology with emotion-aware LLM-driven communication enables more effective transmission of model reasoning processes to users. Real-time detection of emotional valence and arousal allows the system to dynamically adapt the level of detail, tone, and complexity of explanations to better match users' cognitive and emotional states. This real-time adaptive explanation strategy enhances the clarity and comprehensibility of model outputs, as reflected in higher user ratings and qualitative feedback, and facilitates smoother and more effective clinical decision support interactions. Furthermore, the strong performance of the affect recognition module supports the reliability of emotion-adaptive explanation delivery.

XSL•FO
RenderX

As a result, a key contribution of this work is demonstrating that the delivery method of explanations is as important as the content itself. This contribution provides a concrete example of how multimodal emotion perception, XAI techniques, and LLM-based communication can work together to improve user understanding.

Implications for LLM-Based Clinical Decision Support

This study also highlights a new role for LLMs in health care. LLMs can function as empathetic mediators between complex AI models and users. In addition to translating technical XAI outputs into natural language, the empathic-CDSS incorporates the user's emotional state into the explanation process. This advance aligns with emerging evidence that emotionally intelligent conversational agents can enhance trust, adherence, and perceived clarity even when system users lack technical expertise [42-44].

Our results point toward a new model of clinician-AI collaboration. In this model, intelligent systems do not merely present information but actively sense cognitive or emotional load and adjust the way information is delivered. This creates opportunities for more effective and supportive interactions, particularly in high-stakes clinical environments.

By improving interpretability, reducing interaction difficulty, and providing emotionally aligned explanations, empathic XAI frameworks such as empathic-CDSS may help lower barriers to adoption of AI tools in behavioral health monitoring and intervention settings. By enhancing trust, usability, and perceived usefulness, these systems facilitate greater acceptance among clinicians who may otherwise be hesitant to adopt complex AI-driven decision support tools.

Limitations and Future Work

Although this study achieved positive results, several limitations remain that need to be addressed in future work. First, the participant sample of system users consisted primarily of younger adults with high educational attainment, many of whom had limited or no real clinical experience. This demographic profile may not fully reflect practicing users who routinely engage with clinical decision support tools. Future studies should involve a more diverse user population, including physicians, nurses, and allied health professionals, in real-world clinical environments to evaluate performance, workload impact, and long-term adoption of CDSS. The age distribution in our sample was uneven, with no participants in the 40-49 years age group. Although no age-based exclusion criteria were applied, this gap likely reflects sampling variability given the modest sample size (N=33). As a result, findings may not fully generalize to midlife adults within this age range. Future studies with larger and more stratified samples are needed to examine potential age-related differences more comprehensively.

Second, the current system relies on commercial OpenAI application programming interfaces for emotion-adaptive response generation. Although effective for prototyping, this dependency raises concerns about response latency, the cost of large-scale deployment, and data privacy. Transmitting facial expressions and conversational content to external servers poses potential confidentiality risks even when anonymization is applied. To support future clinical translation, deployment using locally hosted or HIPAA (Health Insurance Portability and Accountability Act)-compliant open-source LLMs should be explored, along with on-device or edge-based emotion processing.

Third, while the emotion-perception module improved the clarity and smoothness of interactions overall, several users reported that empathic adjustments sometimes altered the specificity or focus of the explanations. For instance, some responses became overly supportive at the expense of addressing the clinical question directly. This suggests the need for finer control over the strength and timing of empathic behaviors. Future designs may include adjustable empathy levels, task-sensitive modes, or guardrails that ensure explanation fidelity is maintained.

Finally, emotional perception in the current system is based on facial arousal and valence estimation captured through the webcam. This approach may miss important cues such as vocal tone, hesitations, or physiological signals that often reflect cognitive load or frustration. Integrating multimodal data (for example, voice features or wearable sensor data) and developing reinforcement learning mechanisms that adapt emotion recognition based on user feedback could substantially improve model accuracy and the appropriateness of empathic responses.

Overall, these findings demonstrate both the potential benefits and limitations of integrating empathetic communication into AI-driven CDSSs. Key advantages include improved trust, enhanced understanding of model reasoning, and more effective user interaction. At the same time, these systems present important limitations, including potential variability in emotion recognition accuracy and the need to ensure that emotional adaptation does not compromise clinical precision or decision clarity in high-risk medical contexts. Future work should further refine these systems to better balance emotional responsiveness, interpretability, and clinical reliability.

Conclusions

This study introduced an empathic-CDSS that integrates a continuous valence-arousal emotion model with LLMs to support affect-aware interaction in clinical decision support. By using real-time emotional cues to guide response generation, the system is able to detect subtle affective shifts and deliver explanations that are more contextually appropriate and emotionally aligned with the clinician's affective state.

Our findings demonstrate the feasibility and value of combining emotion perception with LLM-based explanation generation. Participants rated the empathic-CDSS higher than the basic system across usability, clarity, personalization, trust, and satisfaction, highlighting the importance of emotion-aware communication in facilitating effective human-AI collaboration. This work suggests that the delivery and adaptation of explanations are as critical as their informational content, particularly when clinicians must interpret complex AI outputs under possible situations of high cognitive load and time pressure.

Overall, this research provides early evidence that emotionally intelligent conversational systems can enhance the clinician experience in AI-assisted decision-making. The integration of

XSL•FO
RenderX

valence-arousal modeling with LLMs offers a promising direction for future emotionally aware clinical AI tools, and it lays the groundwork for designing more supportive, intuitive, and human-centered conversational interfaces in health care.

Funding

This study was supported by the National Institute on Drug Abuse of the National Institutes of Health (U01DA056472 and R21 DA043181). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Methodological details.
[DOCX File , 22 KB - humanfactors_v13i1e89005_app1.docx]

Multimedia Appendix 2
Full facial behavior analysis feature set.
[DOCX File , 19 KB - humanfactors_v13i1e89005_app2.docx]

Multimedia Appendix 3
Survey questions.
[DOCX File , 19 KB - humanfactors_v13i1e89005_app3.docx]

References

1. Zhang T, Chung T, Dey A, Bae SW. AXAI-CDSS: an affective explainable AI-driven clinical decision support system for cannabis use. Int Conf Act Behav Comput (ABC) 2025:1-14. [doi: 10.1109/abc64332.2025.11118599]
2. Najjar R. Redefining radiology: a review of artificial intelligence integration in medical imaging. Diagnostics (Basel) 2023;13(17):2760 [FREE Full text] [doi: 10.3390/diagnostics13172760] [Medline: 37685300]
3. Bae SW, Ferreira D, Suffoletto B, Puyana JC, Kurtz R, Chung T, et al. Detecting drinking episodes in young adults using smartphone-based sensors. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Jun;1(2):1-36 [FREE Full text] [doi: 10.1145/3090051] [Medline: 35146236]
4. Zhang T, Chung T, Dey A, Bae SW. Exploring algorithmic explainability: generating explainable AI insights for personalized clinical decision support focused on cannabis intoxication in young adults. 2024 Int Conf Act Behav Comput 2024:1-15. [doi: 10.1109/abc61795.2024.10652070] [Medline: 39600343]
5. Dayanandan K, Lall B. Enabling multi-modal conversational interface for clinical imaging. CHI EA '24: Ext Abstr CHI Conf Hum Factors Comput Syst 2024:1-13. [doi: 10.1145/3613905.3650805]
6. Chen W, Cheng J, Wang L, Zhao W, Matusik W. Sensor2Text: Enabling natural language interactions for daily activity tracking using wearable sensors. Proc ACM Interact Mob Wearable Ubiquitous Technol 2024;8(4):1-26. [doi: 10.1145/3699747]
7. Ting DSW, Ong JCL, Jin L, Kabilan E, Lim G, Lim D, et al. Development and testing of a novel large language model-based clinical decision support system for medication safety in 12 clinical specialties. Research Square Preprint posted online on Mar 21, 2024. [doi: 10.21203/rs.3.rs-4023142/v1]
8. Kweon S, Kim J, Kim J, Im S. Publicly shareable clinical large language model built on synthetic clinical notes. Assoc Comput Linguist 2024:5148-5168. [doi: 10.18653/v1/2024.findings-acl.305]
9. Park YJ, Pillai A, Deng J, Guo E, Gupta M, Paget M, et al. Assessing the research landscape and clinical utility of large language models: a scoping review. BMC Med Inform Decis Mak 2024;24(1):72 [FREE Full text] [doi: 10.1186/s12911-024-02459-6] [Medline: 38475802]
10. Li J, Zhou Z, Lyu H, Wang Z. Large language models-powered clinical decision support: enhancing or replacing human expertise? Intell Med 2025;5(1):1-4. [doi: 10.1016/j.imed.2025.01.001]
11. Chen Z, Gibson J, Chiu M, Hu Q, Knight T, Meeker D. Automated empathy detection for oncology encounters. IEEE Int Conf Healthcare Inf (ICHI) 2020:1-8. [doi: 10.1109/ichi48887.2020.9374402]
12. Hirsch EM. The role of empathy in medicine: a medical student's perspective. Virtual Mentor 2007;9(6):423-427. [doi: 10.1001/virtualmentor.2007.9.6.medu1-0706]
13. Schwartz JM, George M, Rossetti SC, Dykes PC, Minshall SR, Lucas E, et al. Factors influencing clinician trust in predictive clinical decision support systems for in-hospital deterioration: qualitative descriptive study. JMIR Hum Factors 2022;9(2):e33960 [FREE Full text] [doi: 10.2196/33960] [Medline: 35550304]

XSL•FO
RenderX

14. Wang Z, Yan R, Francis S, Diaz C, Flickinger T, Lin Y, et al. Stakeholder-centric participation in large language models enhanced health systems. npj Health Syst 2025;2:22. [doi: 10.1038/s44401-025-00024-5] [Medline: 40547255]
15. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023;183(6):589-596 [FREE Full text] [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
16. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. npj Digit Med 2023;6(1):120 [FREE Full text] [doi: 10.1038/s41746-023-00873-0] [Medline: 37414860]
17. Lee JD, See KA. Trust in automation: designing for appropriate reliance. Hum Factors 2004;46(1):50-80. [doi: 10.1518/hfes.46.1.50.30392]
18. Broekens J, Hilpert B, Verberne S, Baraka K, Gebhard P, Plaat A. Fine-grained affective processing capabilities emerging from large language models. 11th Int Conf Affective Comput Intell Interact (ACII) 2023:1-8. [doi: 10.1109/acii59096.2023.10388177]
19. Mendes GA, Martins B. Quantifying valence and arousal in text with multilingual pre-trained transformers. Adv Inf Retr 2023;13980:84-100. [doi: 10.1007/978-3-031-28244-7_6]
20. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2020;2(1):56-67 [FREE Full text] [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]
21. Zheng Y, Huang B, Chen W, Ramsey J, Gong M, Cai R. Causal-learn: causal discovery in Python. J Mach Learn Res 2024;25(60):1-8 [FREE Full text]
22. Shang R, Feng KK, Shah C. Why am I not seeing it? Understanding users' needs for counterfactual explanations in everyday recommendations. Proc 2022 ACM Conf Fairness, Accountability, Transparency 2022:1330-1340. [doi: 10.1145/3531146.3533189]
23. Welivita A, Pu P. Curating a large-scale motivational interviewing dataset using peer support forums. Proc 29th Int Conf Comput Linguist 2022:3315-3330 [FREE Full text]
24. Islam R, Bae SW. FacePsy: an open-source affective mobile sensing system - analyzing facial behavior and head gesture for depression detection in naturalistic settings. Proc ACM Hum-Comput Interact 2024;8(MHCI):1-32. [doi: 10.1145/3676505]
25. Bae SW, Chung T, Zhang T, Dey AK, Islam R. Enhancing interpretable, transparent, and unobtrusive detection of acute marijuana intoxication in natural environments: Harnessing smart devices and explainable AI to empower just-in-time adaptive interventions: longitudinal observational study. JMIR AI 2025;4:e52270 [FREE Full text] [doi: 10.2196/52270] [Medline: 39746202]
26. Skope-rules: machine learning with logical rules in python. GitHub. URL: https://github.com/scikit-learn-contrib/skope-rules [accessed 2026-03-24]
27. Komiak SYX, Izak B. The effects of personalization and familiarity on trust and adoption of recommendation agents. MIS Q 2006;30(4):941-960. [doi: 10.2307/25148760]
28. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989;13(3):319-340. [doi: 10.2307/249008]
29. Ferreira D, Kostakos V, Dey AK. AWARE: mobile context instrumentation framework. Front ICT 2015;2:6. [doi: 10.3389/fict.2015.00006]
30. Kunicki ZJ, Zambrotta NS, Tate MC, Surrusco AR, Risi MM, Harlow LL. Keep your stats in the cloud! Evaluating the use of Google sheets to teach quantitative methods. J Stat Educ 2019;27(3):188-197 [FREE Full text] [doi: 10.1080/10691898.2019.1665485] [Medline: 33013153]
31. Caruana R. Multitask learning. Mach Learn 1997;28(1):41-75. [doi: 10.1023/a:1007379606734]
32. Kuppens P, Tuerlinckx F, Russell JA, Barrett LF. The relation between valence and arousal in subjective experience. Psychol Bull 2013;139(4):917-940. [doi: 10.1037/a0030811] [Medline: 23231533]
33. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Proc Int Conf Mach Learn 2015;37:448-456 [FREE Full text]
34. Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge, MA: The MIT Press; 2016.
35. Kingma DP, Ba J. Adam: a method for stochastic optimization. Int Conf Learn Representations (ICLR) 2015:1 [FREE Full text]
36. OpenAI. GPT-4 technical report. 2023. URL: https://cdn.openai.com/papers/gpt-4.pdf [accessed 2026-03-24]
37. Hu EJ, Shen Y, Wallis P, Allen-Zhu Z, Li Y, Wang S. LoRA: low-rank adaptation of large language models. Int Conf Learn Representations (ICLR) 2025 [FREE Full text]
38. Josefsson S. base64 — Base16, Base32, Base64, Base85 data encodings. Python. URL: https://docs.python.org/3/library/base64.html [accessed 2026-03-24]
39. Kossaifi J, Tzimiropoulos G, Todorovic S, Pantic M. AFEW-VA database for valence and arousal estimation in-the-wild. Image Vision Comput 2017;65:23-36. [doi: 10.1016/j.imavis.2017.02.001]
40. Picard RW. Affective Computing. Cambridge, MA: The MIT Press; 1997.
41. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res 2020 Jun 19;22(6):e15154 [FREE Full text] [doi: 10.2196/15154] [Medline: 32558657]

XSL•FO
RenderX

42. Bickmore TW, Pfeifer LM, Jack BW. Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents. In: Proceedings of the SIGCHI conference on human factors in computing systems. 2009 Presented at: CHI '09: CHI Conference on Human Factors in Computing Systems; April 4-8; Boston, MA p. 1265-1274 URL: https://doi.org/10.1145/1518701.1518891 [doi: 10.1145/1518701.1518891]
43. Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. ACM Trans Comput-Hum Interact 2005;12(2):293-327. [doi: 10.1145/1067860.1067867]
44. Liu B, Sundar SS. Should machines express sympathy and empathy? experiments with a health advice chatbot. Cyberpsychol Behav Soc Netw 2018 Oct;21(10):625-636. [doi: 10.1089/cyber.2018.0110] [Medline: 30334655]
45. Swapna M, Devarajan GG, P R, S T, Alturki N, Bashir AK. Human-centered explainable multimodal AI for personalized healthcare diagnosis in aging populations. IEEE Trans Comput Soc Syst 2025:1-11. [doi: 10.1109/tcss.2025.3623200]
46. Chandran G, Devarajan GG, T J, S MMM, Aldhyani THH, Bashir AK. Integration of neural architecture search with fuzzy deep neural network model for emotion AI in public health emergencies. IEEE Trans Comput Soc Syst 2025:1-10. [doi: 10.1109/tcss.2025.3611241]
47. Nagarajan SM, Devarajan GG, M AJ, Arockiam D, Bashir AK, Al Dabel MM. Deep multi-source visual fusion with transformer model for video content filtering. IEEE J Sel Top Signal Process 2025;19(4):613-622. [doi: 10.1109/jstsp.2025.3569446]
48. Manickam M, Devarajan GG. Cryptanalysis on the mutual authentication schemes of telecare medical information systems. 2024 3rd Int Conf Innovation Technol (INOCON) 2024:1-8. [doi: 10.1109/inocon60754.2024.10511882]
49. Medeiros L, Tibor B. An empathic agent that alleviates stress by providing support via social media. Proc 16th Conf Auton Agents MultiAgent Syst 2017:1634-1636. [doi: 10.65109/epte6334]
50. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 2020;20(1):310 [FREE Full text] [doi: 10.1186/s12911-020-01332-6] [Medline: 33256715]
51. Tonekaboni S, Joshi S, McCradden M, Goldenberg A. What clinicians want: contextualizing explainable machine learning for clinical end use. Proc Mach Learn Healthcare 2019;106:359-380 [FREE Full text]
52. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digital Health 2021;3(11):e745-e750. [doi: 10.1016/s2589-7500(21)00208-9]
53. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. NIPS'17: Proc 31st Int Conf Neural Inf Process Syst. 2017. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf [accessed 2026-04-23]

Abbreviations

AFEW-VA: Acted Facial Expressions in the Wild – Valence-Arousal (a publicly available dataset for affective computing)
AI: artificial intelligence
CDSS: clinical decision support system
Empathic-CDSS: empathic clinical decision support system
HIPAA: Health Insurance Portability and Accountability Act
LLM: large language model
MI: motivational interviewing
ML: machine learning
ReLU: rectified linear unit
SHAP: Shapley Additive Explanations
XAI: explainable artificial intelligence

Edited by Y Gong; submitted 04.Dec.2025; peer-reviewed by GG Devarajan, YK Chiam, J Bagby; comments to author 23.Feb.2026; revised version received 15.Mar.2026; accepted 16.Mar.2026; published 22.May.2026.

Please cite as:
Zhang T, Bae SW, Chung T, Dey AK
Emotion-Adaptive Large Language Model–Driven Clinical Decision Support: User Evaluation of the Empathic Clinical Decision Support System Framework for Trust and Explainability
JMIR Hum Factors 2026;13:e89005
URL: https://humanfactors.jmir.org/2026/1/e89005
doi:10.2196/89005
PMID:

XSL•FO
RenderX

©Tongze Zhang, Sang Won Bae, Tammy Chung, Anind K Dey. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 22.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Constructing a Theoretical Hybrid Care Pathway From Stakeholder Perspectives: User-Centered Design Study

Ruben Knevels[1,2,3], MSc; Jelle Schils[1], BSc; Sarah Haesevoets[4], MSc; Julie Vranken[1,2,3], MSc; Chris Burtin[4], PhD; David Ruttens[1,2,5], PhD; Femke Wouters[1,2,3], PhD

[1]Faculty of Medicine & Life Sciences, Hasselt University, Diepenbeek, Flanders, Belgium

[2]Limburg Clinical Research Center / Mobile Health Unit, Hasselt University, Diepenbeek, Flanders, Belgium

[3]Department Future Health, Ziekenhuis Oost-Limburg, Genk, Flanders, Belgium

[4]REVAL - Rehabilitation research center, Faculty of Rehabilitation Sciences, Hasselt University, Diepenbeek, Flanders, Belgium

[5]Department of Respiratory Medicine, Ziekenhuis Oost-Limburg, Genk, Flanders, Belgium

Corresponding Author:
Femke Wouters, PhD
Faculty of Medicine & Life Sciences
Hasselt University
Agoralaan
Diepenbeek, Flanders, 3590
Belgium
Phone: 32 89 80 40 32
Email: femke.wouters@uhasselt.be

Abstract

Background: Chronic obstructive pulmonary disease (COPD) is characterized by acute exacerbations that frequently result in hospitalization, highlighting the need for improved follow-up care. Hybrid care incorporates digital health strategies such as remote patient monitoring (RPM). This could offer the opportunity to detect early signs of disease worsening by remotely tracking selected health indicators. However, it is unknown how RPM can effectively be integrated into current practice.

Objective: This study aims to cocreate a theoretical hybrid COPD care pathway integrating RPM into the existing care trajectory, considering the expertise of participating care pathway stakeholders.

Methods: This 2-phase study was conducted between May and August 2025. At first, semistructured interviews were conducted with stakeholders (ie, patients and health care professionals [HCPs]) involved in COPD care to map the existing care pathway in a single tertiary care hospital in Belgium. Patients were approached during hospitalization based on predefined eligibility criteria, whereas HCPs were recruited through convenience sampling. In phase 2, patient representatives and HCPs were recruited through patient-organization networks and professional referral, respectively, using convenience sampling, and participated in separate interactive workshops. These workshops reviewed existing care pathway phases and HCP roles and explored the integration of RPM, including timing, duration, implementation, data monitoring, and follow-up responsibilities, to codevelop a theoretical hybrid care pathway. Interview and workshop data were independently reviewed by 2 researchers to extract information on care pathway elements, stakeholder roles, limitations, and proposed adaptations. Findings were organized according to the relevant care pathway phase, independently summarized and compared, with discrepancies resolved through discussion and consensus. Interview findings informed the development of an initial infographic representing the existing care pathway, which was subsequently adapted based on workshop findings to represent the proposed cocreated theoretical hybrid care pathway.

Results: In total, 10 stakeholders participated in the interviews (5 patients and 5 HCPs), and 6 stakeholders participated in the workshops (3 patient representatives and 3 HCPs). The existing COPD care pathway involves multiple stakeholders whose activities can be organized into 5 functional domains: support, treatment, medical advice, patient education, and monitoring. Stakeholders can be divided into hospital-based, community-based, or cross-setting. The workshop participants identified the need for a stakeholder dedicated to RPM activities in order to integrate RPM strategies into the care pathway. This individual could be responsible for all RPM-related activities (ie, RPM device installation, providing RPM-related education, data review, alert management, and RPM discontinuation) and could serve as the bridge between patients and physicians, ensuring close communication.

XSL•FO
RenderX

Conclusions: This study resulted in a cocreated theoretical hybrid COPD care pathway. Stakeholders highlighted the potential value of a dedicated RPM role to coordinate and deliver RPM activities as part of the COPD care pathway. Further research is needed to assess its feasibility and clinical applicability.

(JMIR Hum Factors 2026;13:e93735) doi:10.2196/93735

KEYWORDS

care pathway; chronic obstructive pulmonary disease; COPD; digital health; hybrid care; remote patient monitoring

Introduction

Chronic obstructive pulmonary disease (COPD) is a progressive respiratory condition characterized by recurrent episodes of acute symptom worsening, known as acute exacerbations of COPD (AECOPD) [1]. These are defined as "an event in the natural course of the disease characterized by a change in the patient's baseline dyspnea, cough, and/or sputum that is beyond normal day-to-day variations, is acute in onset, and may warrant a change in regular medication in a patient with underlying COPD" [1]. While approximately 18% to 33% of patients with COPD Global Initiative for Chronic Obstructive Lung Disease stage 3 or 4 are hospitalized at least once per year, rehospitalizations represent an even greater challenge [2]. A large cohort study of 15,191 patients with AECOPD reported that 35.1% of patients hospitalized for AECOPD were readmitted within 90 days [3]. AECOPD is not only a leading cause of costly emergency room (ER) visits and hospitalizations but is also associated with poor quality of life [4].

Effective management of COPD, especially during and after AECOPD events, requires coordination across multiple care settings, patient engagement, and timely intervention. Care pathways represent a potential organizational strategy for achieving these aims [5]. A care pathway can be defined as a complex intervention for the mutual decision-making and organization of a care process for a well-defined group of patients during a well-defined period [5]. Their use offers the advantages of standardizing best practice, streamlining care processes, and providing a structural framework for collaboration between care pathway stakeholders [6]. Additionally, in COPD, care pathways have shown the potential to reduce complications, readmissions, and length of stay [7]. Given that no globally established guidelines exist for structured COPD care, clinical practice differs between centers and countries [8].

Hybrid care, which combines traditional in-person care with digital health modalities such as remote patient monitoring (RPM), is being increasingly adopted across various health care domains. Evidence suggests that hybrid approaches can enhance access, patient engagement, and continuity of care, particularly in the management of chronic diseases [9-11]. In addition, they may offer a scalable solution to reduce ER visits and exacerbation-related readmissions in patients with COPD [12]. Despite the growing adoption of hybrid care models, evidence on how to effectively integrate hybrid care components into existing care pathways remains limited. Furthermore, it is unclear how these pathways should be adapted to ensure that quality, safety, and coordination are maintained [11].

The aim of this descriptive co-design study was to (1) map the current COPD care pathway, including stakeholder role delineation and perceived bottlenecks, and (2) cocreate a theoretical hybrid care pathway for patients with COPD after an AECOPD event, incorporating digital health strategies based on stakeholder perspectives.

Methods

Study Design: 2-Phase Design

A 2-phase descriptive co-design approach was used to develop a theoretical hybrid COPD care pathway incorporating digital health strategies. Semistructured interviews and interactive workshops were conducted with patients with COPD, patient representatives of COPD, and health care professionals (HCPs) involved in COPD care at a tertiary care center and its regional care network in Belgium. The study specifically targeted people who experience COPD exacerbations, defined as individuals experiencing AECOPD who required hospitalization.

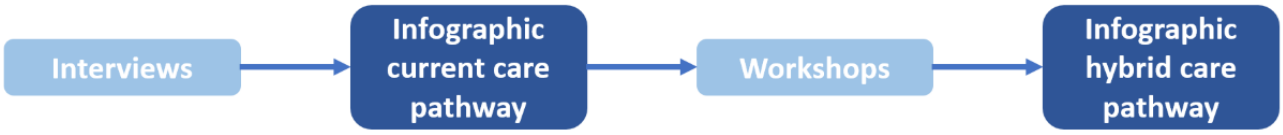
Data collection occurred in consecutive phases. Recruitment continued until sufficient information had been obtained to support the mapping of the care pathway. Information sufficiency was assessed at the level of the overall care pathway rather than within each individual stakeholder group. Recruitment continued until the interviews consistently confirmed previously identified care pathway elements and did not substantially contribute new information relevant to the study's focus. For stakeholder groups represented by multiple participants, consistency of information across interviews was considered when assessing whether additional interviews were likely to contribute new pathway elements. Groups represented by a single individual were considered specific informants because of their distinct roles and expertise in COPD care or hybrid care. In these interviews, information sufficiency could therefore not be assessed through within-group repetition. Instead, their contributions were assessed in relation to the broader pathway information obtained from other participants and their specific relevance to the pathway components with which they were involved.

Both interviews and workshops were conducted by a postdoctoral researcher (FW) with a PhD in Biomedical Sciences, accompanied by a second researcher (RK), who took field notes and digitally recorded the sessions. Both researchers had a background in digital health research. They had no direct clinical care relationship with the participants. Sessions were audio-recorded with participant consent, and recordings were destroyed following manual, clean verbatim transcription. Participant checking was not performed.

XSL•FO
RenderX

In the first phase, semistructured interviews were conducted to gather the perspectives of both patients and care pathway stakeholders on the existing COPD care pathway. Insights from these interviews informed the visual representation of the care pathway in the form of an infographic. In the second phase, this infographic was used as a guide during the workshops. A new theoretical hybrid care pathway was developed, informed by stakeholder expertise. This new hybrid care pathway was again visualized in the form of an infographic (Figure 1).

Figure 1. Study design.



This study adhered to the COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines [13]. The COREQ checklist is provided in Multimedia Appendix 1.

Phase 1: Interviews

Recruitment

Eligible interview participants included patients hospitalized due to AECOPD and HCPs involved in COPD management or RPM. Patient selection was based on predefined inclusion and exclusion criteria. All patients had a confirmed diagnosis of COPD by spirometry (Tiffeneau index <0.7) and a history of multiple AECOPD events, including hospitalization due to AECOPD within the preceding year. Patients hospitalized due to AECOPD at Ziekenhuis Oost-Limburg were screened for eligibility by DR. Eligible patients were approached in order of hospital admission. HCPs were eligible if they were actively involved in the care of patients with COPD and had at least 2 years of experience in COPD care. HCPs were identified based on their active involvement in COPD care and selected to represent different professional roles in the COPD care pathway. Potential participants were identified through professional contacts within the clinical networks and contacted by email with an invitation to participate. Recruitment therefore followed a convenience sampling approach based on the accessibility and willingness of eligible HCPs to participate. Eligible HCPs included all professionals actively involved in the care of patients with COPD (ie, pulmonologists, physiotherapists, occupational therapists, and a dedicated COPD nurse) and did not have to be associated with Ziekenhuis Oost-Limburg. To support the development of a hybrid care pathway, an additional interview was conducted with a telemonitoring nurse from Ziekenhuis Oost-Limburg who had extensive experience in implementing hybrid care in the Cardiology department.

Interview Procedures

Semistructured, face-to-face, one-on-one interviews were conducted to explore the perceptions of patients and HCPs regarding the current COPD care pathway. All interviews were carried out by the same researcher to ensure consistency. Interview guides were developed by FW by first identifying the information required from each participant group to map the existing COPD care pathway. Questions were then formulated to address the relevant experiences, roles, and perspectives of each group. Separate guides were developed for patients, physicians, occupational therapists and physiotherapists, the COPD nurse, and the RPM nurse. The guides consisted of open-ended questions covering relevant aspects of COPD care and were tailored to the specific roles and experiences of each participant group. Because the study involved RPM using digital technology, additional questions were asked to evaluate participants' digital literacy. The interview guides for patients, physicians, occupational therapists and physiotherapists, COPD nurses, and RPM nurses are provided in Multimedia Appendices 2-6, respectively.

Prior to each interview, patients were informed that the purpose of the study was to explore their experiences and perceptions of the care they received following previous hospitalizations. For HCPs, the goal was to gain insights into their professional roles and perspectives on the COPD care pathway. Interviews used open-ended questions organized into categories, including postdischarge treatment, follow-up, perceived bottlenecks, symptoms, involvement of informal caregivers, and potential areas for improvement in care. All interviews were conducted at the tertiary care center Ziekenhuis Oost-Limburg in Genk, Belgium.

Data Analysis

Interview transcripts and accompanying field notes were reviewed independently by FW and RK. Each researcher reviewed the transcripts and notes to identify information relevant to the existing COPD care pathway at Ziekenhuis Oost-Limburg, focusing on (1) the different phases of care, (2) the stakeholders involved in each phase, (3) the actions and responsibilities of each stakeholder, (4) perceived limitations or bottlenecks in the current care pathway, and (5) suggestions for improving current care processes. Formal qualitative coding was not performed; instead, relevant information was systematically extracted and organized according to the corresponding phase of the care pathway. Each researcher subsequently developed an independent summary of the identified care pathway elements and stakeholder perspectives. These summaries were compared to identify areas of agreement and discrepancy. Discrepancies were discussed by FW and RK until consensus was reached. The resulting synthesis was used to construct the initial infographic representing the existing COPD care pathway, including different phases, involved stakeholders, and their respective actions and responsibilities. Illustrative quotes were selected to reflect the identified perspectives and provide examples of the findings.

Phase 2: Workshops

Recruitment

Eligible workshop participants included Dutch-speaking representatives of patients with COPD, informal caregivers of

XSL•FO
RenderX

patients with COPD, and HCPs. Patient representatives were selected over hospitalized patients to provide perspectives informed by broader experiences with COPD and health care services, rather than experiences based solely on individual clinical status or hospitalization history. In addition, through their ongoing activities and interactions with other people living with COPD, they are exposed to a variety of patient experiences and perspectives beyond their own. Patient representatives were members of COPD vzw, a Flemish organization representing patients with COPD. Representatives were contacted by email and invited to participate in the workshop. Workshop HCPs were eligible if they were actively involved in the care of patients with COPD and had at least 2 years of experience in COPD care. HCPs were identified based on their active involvement in COPD care and selected to represent different professional roles in the COPD care pathway. Participants were invited by email and recruited using a convenience sampling approach. To maximize data richness and encourage honesty about challenges, separate workshops were organized for patient-side stakeholders and HCPs.

Workshop Procedures

Each workshop was an interactive, semistructured session designed to engage participants in collaboratively considering the implementation of digital health solutions within the current COPD care pathway. After the study objectives were introduced, participants reviewed and verified care pathway phases and HCP roles. Next, participants discussed the infographic representing the current care pathway and explored transitions toward a hybrid model. In this hybrid model, RPM was introduced by the researchers as a nonintrusive wearable device (eg, a smartwatch) capable of measuring multiple vital parameters. Specific devices, parameters, and alert protocols were not further explored. Discussion topics included role delineation, stakeholder responsibilities, the period of remote monitoring, the integration of remote monitoring into routine COPD care, and potential hurdles to implementation. The workshop guides for HCPs and patient representatives are provided in Multimedia Appendix 7.

Data Analysis

Workshop transcripts and notes were independently reviewed by FW and RK using a descriptive approach similar to that used with the interviews. The analysis focused on proposed changes to the existing care pathway and the integration of RPM and other digital health strategies, including (1) changes in stakeholder roles and responsibilities, (2) communication and follow-up processes, (3) monitoring activities, and (4) points of interaction between stakeholders. Formal qualitative coding was not performed; instead, relevant information was systematically extracted and organized according to the corresponding phase of the care pathway. FW and RK independently developed descriptive summaries of the proposed adaptations and compared their findings. Discrepancies were discussed and resolved through consensus. The agreed findings were then mapped onto the initial infographic to identify which components of the existing pathway should be retained, modified, added, or removed. This iterative process resulted in a second infographic representing the proposed theoretical hybrid COPD care pathway. Illustrative quotes were selected to reflect the identified perspectives and provide examples of the findings.

Statistical Analysis

Data processing and statistical analysis were performed using R (version 4.6.0; R Foundation for Statistical Computing, Vienna, Austria). Descriptive statistics are presented as mean (SD) or number (%), as appropriate.

Ethical Considerations

The study protocol complied with the principles of the Declaration of Helsinki and was approved by the medical ethics committee (ie, Ziekenhuis Oost-Limburg, Genk, Belgium; B3712025000001). Written informed consent was obtained from all participants prior to data collection.

Results

Participant Characteristics

Participants were recruited between May and August 2025. A total of 5 patients with COPD and 5 care pathway stakeholders were approached for the interviews. As no participants declined the invitation, 5 patients with COPD and 5 care pathway stakeholders participated in the interviews (Table 1). The interviews with patients lasted 11.8 (range 8.0-32.3) minutes, and those with stakeholders lasted 23.2 (range 15.3-45.3) minutes. For the workshops, 3 patient representatives and 3 HCPs were invited. No participants declined the invitation. Workshops with patient representatives lasted 1 hour 38 minutes, and workshops with HCPs lasted 1 hour 28 minutes. One patient representative was also active as an informal caregiver. None of the interviewees took part in the workshops. All HCPs reported daily internet use, and 4 out of 5 reported daily smartphone use, compared with 2 out of 5 and 1 out of 5 patients, respectively. Only 1 HCP received data from wearables or remote monitoring, whereas none of the patients did.

XSL•FO
RenderX

Table 1. Characteristics of the interview and workshop participants.

Characteristic	Interview		Workshop	
	Patient (n=5)	HCP[a] (n=5)	Patient representative (n=3)	HCP (n=3)
Age (years), mean (SD)	65.0 (4.1)	38.6 (14.5)	71.7 (3.1)	53.3 (14.2)
Male, n (%)	3 (60)	2 (40)	2 (66.7)	2 (66.7)
Experience[b] (years), mean (SD)	N/A[c]	6.7 (6.2)	10.7 (8.1)	24.0 (4.9)
Time since diagnosis (years), mean (SD)	7.6 (4.3)	N/A	N/A	N/A

[a]HCP: health care professional.

[b]Number of years of experience is calculated as the number of years a patient representative has participated as a representative of patients with COPD or an HCP has worked in COPD-related care.

[c]N/A: not applicable.

Care Path Overview

HCP participants indicated that the current care pathway for people who experience COPD exacerbations can be set out in general terms. However, they emphasized that individual adjustments are necessary depending on patients' medical history, pulmonary function test results, and individual illness experience.

Overall, the interviews indicated that patients may either have been formally diagnosed with COPD before their first AECOPD event or encounter the disease for the first time in the context of an AECOPD hospitalization. While mild AECOPD events can often be managed in primary care, severe cases often require hospital admission. After a period of hospitalization, the patient stabilizes and is discharged from the hospital. In the care pathway described by participants, the discharge decision is based on a clinical evaluation that includes infectious parameters, general blood work, and the need for supplemental oxygen. If necessary, the patient may be discharged with temporary supplemental oxygen at home. However, this clinically determined point of discharge did not always correspond to patients' perceived readiness to return home. Several patients described feeling insufficiently prepared for discharge, highlighting a gap between clinical stability and perceived readiness for home recovery (quote 1 in Textbox 1).

Textbox 1. Illustrative quotes from interviews and workshops with patients, patient representatives, and health care professionals.

Quote 1: patient 4 (interview)

- For me, it (being discharged) was mentally very challenging. I continually questioned myself whether I was ready for it.

Quote 2: patient representative (workshop)

- The pharmacist is actually someone who is just as close to the patient as the GP. That's where you go for your medication, and that's where your partner goes for their medication.

Quote 3: health care professional (workshop)

- Wherever a home care nurse visits, education can be provided, but we still miss a group of patients we never reach. All patients, however, visit the pharmacist. Not all physicians know that they must write a prescription for inhalation-therapy education. The pharmacist needs that prescription to be compensated.

Quote 4: home nurse (workshop)

- The required level of monitoring vital signs should be determined on an individual basis. However, a challenge is that many patients do not receive home nursing care, particularly when no technical nursing or hygiene-related support is required. In these cases, home nurses are often not involved. This creates a gap regarding patient education and support, as these activities are not reimbursed as home nursing services. From the patient's perspective, charging €10 per day for this may represent a considerable financial burden, even though a daily cost ceiling applies. At the same time, this fee would not cover the involvement of a home nurse. This creates a challenge, despite the fact that such support could be highly valuable for patient follow-up.

Quote 5: patient representative (workshop)

- When you leave the hospital, it feels like something hits you on the head. You only remember that you won't be able to manage for a while, and everything else is gone. If someone else is with you, they can put things into perspective and remember the details. But as a patient, you're told so much that you simply can't take it all in anymore.

Quote 6: pulmonologist (interview)

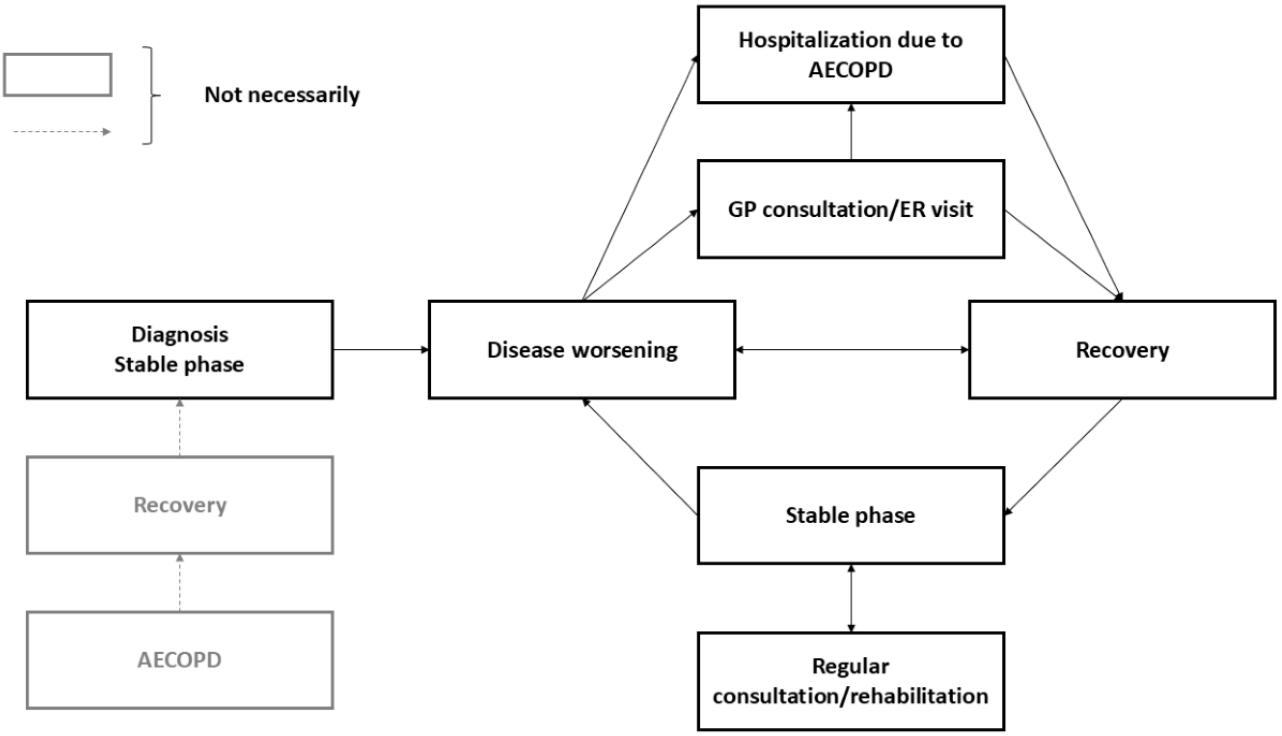
- In case of disease worsening, we always advise patients to see a physician. In most cases, this is their GP, but long waiting times can allow their condition to worsen. Having a way to monitor patients remotely, without adding excessive medical burden or cost, could be crucial.

XSL•FO
RenderX

Besides feeling unprepared, patients reported experiencing a lack of information, guidance, and follow-up after hospitalization for AECOPD. In the local care setting, discharge letters typically contain the necessary clinical details and are provided to patients or transmitted electronically to their general practitioner (GP). However, patients reported that the information they received was largely limited to their current pharmacological regimen. Moreover, they noted a lack of guidance regarding potential postdischarge symptoms, appropriate self-management strategies, and clear instructions on whom to contact in the event of clinical deterioration.

Although the exact point at which recovery begins is not always clear, 2 possible scenarios were identified during this phase. In the first, the disease suddenly worsens, often resulting in a new visit to a GP or ER, with potential hospitalization. In the second scenario, recovery progresses into a stable phase, during which patients should receive at least 1 annual outpatient follow-up that includes a pulmonology consultation and pulmonary function testing. Depending on individual needs, certain components of the care pathway may be added or omitted (Figure 2).

Figure 2. General clinical trajectory of a patient experiencing an acute exacerbation of chronic obstructive pulmonary disease (AECOPD). GP: general practitioner; ER: emergency room.



Care Pathway Stakeholders

The care pathway for patients who frequently experience COPD exacerbations can be represented by the vicious circle in the general clinical COPD trajectory (Figure 2). Across this pathway, 10 key stakeholders were identified, which can be categorized into 3 groups based on their primary setting of interaction: hospital-based, community-based, and cross-setting (Multimedia Appendix 8). The stakeholder roles described reflect responsibilities, activities, and interactions reported by study participants in relation to the existing COPD care pathway. These descriptions therefore reflect participants' experiences and perceptions of how local care was organized in practice and are not intended to define the formal scope of practice of each profession.

Hospital-Based Stakeholders

The pulmonologist is primarily responsible for patient treatment, including prescribing medication and interpreting spirometry results. Additionally, participants reported that the pulmonologist plays a critical role in patient education, discussing lifestyle modifications and pulmonary rehabilitation.

While most hospital-based interactions occur during periods of hospitalization, annual ambulatory follow-ups were reported to be recommended to assess medication efficacy and lung function progression.

The hospital nurse was mentioned as contributing both during hospitalization and in ambulatory follow-up. During hospitalization, the hospital nurse provides direct patient care and education. During outpatient visits, participants described the primary task of the nurse as performing spirometry testing.

The dietitian was mentioned as providing dietary counseling, particularly for patients experiencing unintentional weight loss or gain, thereby supporting nutritional optimization as part of COPD management.

According to interviewed participants, the social worker evaluates the patient's living environment, supports discharge planning, and offers psychosocial support when needed.

Cross-Setting Stakeholders

Although their role may appear self-evident, pharmacists were recognized by both patients and HCPs as essential contributors

XSL•FO
RenderX

to the COPD care pathway. Beyond supplying prescribed therapies, pharmacists were reported to be well-positioned to guide patients and informal caregivers in the correct use of inhalers and other respiratory treatments. Their frequent and accessible contact with patients allows them to reinforce key aspects of disease management and support adherence to therapy (quote 2 in Textbox 1). Notably, a participating HCP mentioned that, in Belgium, these educational services may be eligible for reimbursement when appropriately prescribed by the patient's treating physician (quote 3 in Textbox 1).

Another important stakeholder who can be present in both hospital-based and community-based settings is the physiotherapist. In the hospital, the role of the physiotherapist was described as focusing on patient mobilization, physiotherapy-related education, inhaler use, and general disease management. After hospital discharge, according to participants, the physiotherapist is responsible for guiding the patient through pulmonary rehabilitation.

The physiotherapist primarily focuses on restoring mobility, physical performance, and overall body function. The occupational therapist, in contrast, was described by participants as supporting patients in regaining or maintaining independence in activities of daily living (ADL). In addition to ADL training, participants mentioned that the occupational therapist may provide education on energy preservation, breathing techniques, and relaxation methods. For patients at increased risk of falls, targeted preventive interventions are also implemented.

Community-Based Stakeholders

During daily life, ongoing medical management was described by participants as being primarily coordinated by the GP. As the patient's primary point of access to health care, the GP was described as providing comprehensive and ongoing care, addressing both acute issues and long-term conditions. Participants indicated that the GP's role includes discussing lifestyle modifications, educating patients on medication use and symptom monitoring, and supporting overall disease control. Participants also highlighted that, when patients experience worsening respiratory symptoms, Belgian local practice allows the GP to perform diagnostic tests, such as a C-reactive protein test, to help identify potential infectious exacerbations.

Alongside the GP, the home care nurse was described by participants as playing a key role in supporting patients in their own homes. Within the care pathway described by participants, home care nurses were reported to provide professional nursing care, play a role in therapy- and lifestyle-related education, and assist with symptom management and overall disease management. Participants also indicated that home care nurses may monitor vital signs and other clinical indicators, enabling early detection of potential deterioration and timely intervention. However, participants noted that their involvement is generally limited to situations in which patients present with additional care needs, most commonly due to comorbid conditions that require home care nursing services (quote 4 in Textbox 1).

The final stakeholder in the COPD care pathway is the informal caregiver. Although participants noted that not all patients have access to unpaid, nonprofessional support, multiple stakeholders recognized the important role informal caregivers often play in daily life management. Participants described their responsibilities as varying according to the patient's needs and the nature of their relationship, potentially including personal care, emotional support, assistance with household tasks, and monitoring of the patient's health status. Study participants reported that the contributions of informal caregivers are often underrecognized by health care professionals. In particular, they highlighted that patients may feel overwhelmed during hospital visits and struggle to retain important information. For instance, participants described the presence of an informal caregiver or family member during the initial diagnosis, often perceived as life-altering, as helping patients better understand and remember key instructions and explanations provided by the pulmonologist (quote 5 in Textbox 1).

While most stakeholders are continuously present, certain stakeholder roles were mentioned as being embedded only as needed, such as those of the dietitian, social worker, home care nurse, and informal caregiver. Although these stakeholders play valuable roles, their involvement depends on the patient's personal circumstances, preferences, and individual care needs.

Limitations of the Current Care Pathway

Stakeholders identified several organizational and communication challenges within the current COPD pathway. These included the insufficient or time-limited involvement of physiotherapists and occupational therapists during and after hospitalization, inadequate patient and informal caregiver education, and a lack of structured follow-up at home, commonly referred to as the "black box" phase of postdischarge care. Both patients and professionals emphasized that education provided during acute hospitalization is often forgotten due to emotional strain and should therefore be repeated and reinforced during follow-up visits. In addition, physicians noted that referrals to physiotherapists and dietitians are not based on clearly defined criteria; instead, they largely depend on clinical judgment and physician preference. As a result, physiotherapy, particularly outpatient physiotherapy, is often initiated too late, even though earlier referral and intervention might be beneficial for patient outcomes. Moreover, communication between primary and secondary care has several limitations. Information is often lost or repeatedly requested, and relevant clinical details are not consistently transferred across care settings because different electronic health systems are unable to exchange data effectively. On the other hand, disease deterioration is often accompanied by patient anxiety and hesitation to seek help, sometimes reinforced by a perceived sense of failure when symptoms cannot be managed independently. This uncertainty may delay timely action, underscoring the need for objective RPM to support patients in recognizing clinically significant changes and responding appropriately (quote 6 in Textbox 1).

Development of a Hybrid COPD Care Pathway

This new theoretical hybrid care pathway focuses on patients who experience hospitalization due to AECOPD, as they place the greatest burden on the COPD care system. Stakeholders across primary and secondary care agreed that RPM strategies should be initiated and coordinated within secondary care, where clinical expertise and data interpretation capacity are

XSL•FO
RenderX

concentrated. Follow-up was considered most effective when led by a dedicated hospital-based nurse (eg, a COPD coach), who serves as the primary individual responsible for all RPM-related activities (quote 7 in Textbox 2). In addition, this individual could act as the first point of contact for patients experiencing worsening COPD-related symptoms, including those who are not enrolled in an RPM program (quote 8 in Textbox 3). RPM-related activities include installing the RPM devices, providing RPM-related education, reviewing RPM data, managing alerts, and discontinuing RPM when appropriate. While the COPD coach handles the practical aspects, close collaboration with the treating physician and regular communication with the GP remain essential. In addition to RPM management, the COPD coach can also play a key role in enhancing disease- and medication-related education and serve as a key contact person for both patients and their informal caregivers (Multimedia Appendix 8).

Textbox 2. Quote 7: health care professional (workshop).

Interviewer: Who should be responsible for reviewing the RPM-generated alerts?

Pulmonologist: Initially, this would be a nurse, who then reports back to the physicians (pulmonologists).

General practitioner: In our setting as well, especially if multiple conditions were being monitored, it would be a practice nurse who would perform the initial triage of alerts.

Home nurse: With telemonitoring, we are currently considering having an on-call nurse review the data once per day and escalate issues as needed.

Pulmonologist: I think it would be easier for our (pulmonology) department to assign someone to this task rather than shifting it to you in primary care.

General practitioner: Yes, we cannot hire someone for every pathology.

Pulmonologist: We want to reduce your workload. That already starts with a COPD exacerbation, so inclusion in RPM happens in the hospital. Therefore, it seems logical for this responsibility to remain with us (the hospital).

Textbox 3. Quote 8: health care professional (workshop).

Interviewer: Who should be the first point of contact for the patient?

Pulmonologist: A study nurse or a reference nurse, like we have in our oncology department (onco-coach). In case the patient experiences any issues, the patients contacts the nurse, who tries to solve the problem or contacts us in case of a medical issue.

Workshop participants emphasized that RPM should be considered a complementary tool rather than a replacement for direct patient contact. Although physiological parameters provide valuable insights, they do not fully capture the patient's clinical condition. Participants highlighted the importance of maintaining telephone or face-to-face interactions, as communication allows HCPs to assess additional aspects, such as the patient's breathing, speech, and overall presentation, which are essential for interpreting the clinical relevance of monitored parameters (quote 9 in Textbox 4).

Textbox 4. Quote 9: health care professional (workshop).

Pulmonologist: Is telephone consultation included somewhere in the infographic? We receive many phone calls from patients, and based on these conversations, we assess whether a patient should come in for a consultation or whether, for example, a short course of medication is appropriate if we know the patient well.

Interviewer: Could this be replaced by increased RPM?

Pulmonologist: No, home monitoring would complement this or provide an initial indication, but when you speak to a patient by phone, you also hear many relevant aspects: how they talk, how they breathe, and they can explain their situation.

Home nurse: Parameters are important, but they do not tell the whole story. With telemonitoring, when a parameter suggests that something may be wrong, it remains important to see or hear the patient directly to assess whether that value is clinically meaningful. RPM can serve as a supportive tool.

Primary care providers, particularly home care nurses, were viewed as having a supportive role, assisting with device troubleshooting or connectivity issues and communicating with the hospital nurse when necessary. As not all patients receive home care nursing, direct access to the COPD coach was considered essential.

GPs preferred to remain informed of their patients' participation in an RPM program but did not wish to interpret continuous data streams. Instead, they favored receiving periodic summary reports, with stakeholders specifically suggesting a one-page summary (quote 10 in Textbox 5) offering an overview of the patient's clinical status. HCP workshop participants reached a consensus that, within the proposed pathway, RPM data and alerts would therefore be reviewed and triaged by the designated in-hospital RPM team, with clinically relevant findings escalated to the appropriate pulmonologist (quote 7 in Textbox 2). This approach facilitates information sharing and continuity of care.

XSL•FO
RenderX

Textbox 5. Quote 10: health care professional (workshop).

I think those data should be available when the patient sits in front of you during a consultation, so you can integrate the data with everything else. When you're with the patient, you want to see it: that little graph of the oxygen saturation or breathing rate. There should already be an A4 sheet ready with those data on it.

Currently, RPM is not embedded in the daily management of patients with COPD at our care center, but it is likely to be more beneficial for those with severe disease or multiple comorbidities. However, participants indicated that patients without multiple comorbidities could also benefit from closer monitoring, representing a significant opportunity for improved care. Educating patients to enable self-monitoring is essential, and monthly follow-up for more stable patients would be desirable to ensure continuity and early detection of potential issues.

The hybrid COPD care pathway is presented as an infographic (Multimedia Appendix 8). It is organized into the main phases of the care pathway (ie, hospitalization, recovery, stable phase, and disease worsening), mirroring the circular trajectory shown in Figure 2. For each phase, the infographic summarizes the key activities of all stakeholders, grouped into 5 functional domains: support, treatment, medical advice, patient education, and monitoring. Changes from the current care pathway to the hybrid model are indicated with an asterisk.

Potential Barriers to RPM Implementation

Several potential barriers to RPM implementation were identified during the workshops, including low digital literacy, limited disease awareness, and affordability. Low digital literacy was highlighted as an important consideration in this population (quote 11 in Textbox 6). However, participants discussed that RPM could remain feasible if the system were designed to be as simple and user-friendly as possible, requiring minimal effort from patients. Limited disease awareness was also identified as a potential challenge, as some patients may be reluctant to engage in monitoring or to be continuously confronted with their disease status (quote 12 in Textbox 7). Furthermore, participants emphasized that alerts generated by the RPM system should be minimized and, where possible, should not be directly visible to or burdensome for patients.

Textbox 6. Quote 11: health care professional (workshop).

Interviewer: What are potential barriers for patients to use RPM?

Home nurse: I think the main challenge is digital illiteracy.

Pulmonologist: Yes, but perhaps patients should not have to do much themselves?

General practitioner: If it is a watch that performs many tasks automatically, this should not cause many problems.

Textbox 7. Quote 12: health care professional (workshop).

Home nurse: Another potential barrier is disease awareness. Patients may be reluctant to be monitored in this way. Similar to what we currently observe, some patients do not seek care from their general practitioner or hospital because they do not want to know how severe their disease is. RPM may confront them with their condition more directly.

Interviewer: So, they may feel that they are being confronted too much with the reality of their disease?

General practitioner and home nurse: Yes, indeed.

Discussion

Principal Findings

This study mapped the current COPD care pathway, delineated stakeholder roles, and developed a stakeholder-informed theoretical hybrid COPD care pathway incorporating RPM, while identifying key considerations for its integration into routine COPD care. Stakeholders were defined as individuals involved in the organization, delivery, or governance of COPD care. While internationally established guidelines exist for the management of COPD, there is no globally standardized COPD care pathway. Nevertheless, these guidelines recommend common principles that should underpin COPD care, including appropriate pharmacological treatment, smoking cessation, and annual vaccination. In addition, individualized assessment of nutritional status and mobility, encouragement of pulmonary rehabilitation, and evaluation of the need for home supplemental oxygen therapy should be performed [1]. However, the organization and implementation of these principles within care pathways often rely on the clinical judgment of physicians or locally developed internal hospital guidelines [8]. Consequently, substantial variation may occur between patient trajectories, depending on the country, available resources, and health system organization [8].

Although there are often no clear guidelines for COPD trajectories, evidence suggests that structured care pathways can improve care and outcomes for patients, as well as reduce unnecessary resource use and associated costs for health care providers [14,15]. The proposed theoretical hybrid care pathway is based on the COPD care pathway at a tertiary care hospital with an academic character (ie, Ziekenhuis Oost-Limburg, Genk) in Belgium and may differ from practices in other clinical institutions. Nevertheless, many of the general insights obtained remain relevant. In particular, the overarching responsibilities and the role of the COPD coach within the proposed theoretical hybrid care pathway may provide a useful starting point for consideration in other care settings and may inform the future development of similar care pathways. However, the specific

XSL•FO
RenderX

delineation of this role, including task allocation, integration within existing care teams, and organizational arrangements, is highly context- and institution-specific. As these aspects were beyond the scope of the workshops, they should be further defined and adapted by individual health care institutions during local implementation.

In the development of the hybrid care pathway, the researchers specified that RPM would be performed using a wearable device capable of measuring multiple vital parameters, without further specifications. The proposed hybrid care pathway should therefore be understood as an organizational framework for integrating RPM into COPD care rather than a device-specific protocol. It does not prescribe or evaluate specific devices, monitored parameters, alert thresholds, or escalation protocols, which should be defined and evaluated in subsequent research. Given that various hybrid RPM approaches have been evaluated in the literature, with varying results, further research should specify the components and structure of RPM in COPD [16]. Nevertheless, the integration of RPM into chronic care pathways is not a mere technical add-on but necessitates a reconfiguration of roles, responsibilities, and interactions within the care team. Previous literature on real-world implementations showed that it requires a fundamental reorganization of clinical workflows, staffing, and coordination structures [17,18]. For instance, in a multicenter European oncology program covering 33 cancer centers in France and Belgium, a remote symptom-monitoring pathway based on weekly patient-reported outcomes was successfully deployed after explicit roles were defined, alert-management workflows were established, and the RPM process was embedded into routine care using implementation science methods [19]. Similarly, in chronic heart failure care, surveys found that more than 80% of cardiology departments created dedicated RPM teams after implementing RPM, adjusting human resource allocation, and establishing new consultation workflows for handling alerts [20]. These findings support the broader concept that implementing RPM may require dedicated roles and adaptations to existing care processes.

In line with these insights, the proposed theoretical hybrid care pathway differs from the current care pathway by introducing a new stakeholder, the COPD coach, to support the integration of RPM into COPD care. While the name given to this new stakeholder is purely arbitrary, the role itself was proposed by workshop participants as a potential coordinating function for RPM-related activities. The COPD coach could be responsible for all RPM-related activities, including device installation, providing information on device use, following up on alerts, and discontinuing RPM. The COPD coach could perform these tasks in close collaboration with the treating pulmonologist and be responsible for reporting escalation outcomes to the treating GP.

Another stakeholder potentially involved in RPM activities is the informal caregiver, whose role in supporting patients throughout the disease trajectory has been repeatedly underscored by workshop participants but remains frequently underestimated [8]. Within the care pathway, the involvement of informal caregivers varies across phases of disease management but typically includes assisting with activities of daily living, medication management, emotional support, and coordination with HCPs [21]. Importantly, informal caregivers could also play an important role in the successful integration of RPM into the pathway. They may help patients operate monitoring devices, interpret and track generated health data, and liaise with the COPD care team (eg, the COPD coach) when alerts or issues arise. Empirical studies suggest that patients with engaged informal caregivers demonstrate improved symptom management, higher adherence to prescribed therapies, and more effective participation in self-management activities [22,23]. Consequently, informal caregivers act as an essential link between the patient, the RPM system, and the multidisciplinary care team, ensuring continuity, accuracy, and responsiveness throughout the proposed hybrid COPD care pathway.

While the COPD coach and informal caregiver play important roles in the proposed theoretical hybrid COPD care pathway, other stakeholders may also play significant roles. During the RPM follow-up period, initial triage and coordination are handled by the COPD coach, but the treating pulmonologist retains overall responsibility for the patient's clinical management when RPM alerts are generated. In some cases, the GP may adjust medical treatment during a scheduled visit based on RPM data trends or clinician-initiated recommendations. When an informal caregiver is absent, home care nurses may support the patient in operating RPM devices or help interpret monitoring data. They may serve as the interface between the patient, caregiver, and physician, ensuring device use and reporting. Thus, RPM becomes embedded within a multistakeholder care model in which multiple stakeholders jointly contribute to surveillance, triage, and therapeutic decision-making [17].

Although this study did not focus on a specific RPM strategy, implementing RPM requires several considerations. First, a reliable device must be used to collect patient data, which may take the form of a dedicated physical device or an application installed on the patient's smartphone. Second, data must be transmitted and stored automatically and securely, ensuring accessibility for authorized health care professionals. Where appropriate, algorithms for data processing and alert generation should be used to facilitate timely clinical decision-making. Third, the collected information should be presented in a meaningful and actionable manner for clinicians, ideally through integration with the electronic health record or other existing clinical systems [18]. Finally, patients must have sufficient digital literacy to participate effectively in an RPM program. Further research is needed to determine the level of digital literacy required for successful engagement [24].

In addition to RPM strategies, mobile health solutions can play a meaningful role in addressing educational needs. Mobile apps can provide structured, accessible, and on-demand educational content that reinforces disease knowledge, enhances understanding of treatment plans, and supports appropriate responses to changes in health status, potentially leading to reduced hospitalizations and improved quality of life [25,26]. Such digital tools may therefore complement traditional education, help overcome information loss after hospitalization, and strengthen patients' self-management capacities [27].

XSL•FO
RenderX

Future Perspectives

Future research should address several aspects related to the practical implementation of this theoretical framework. While previous research has demonstrated the feasibility and potential value of RPM using wearable devices in other care settings, further research is necessary to determine its impact on clinically relevant patient outcomes [28]. In addition, the influence of digital literacy and socioeconomic status on patients' ability to participate in RPM programs should be further explored, as these factors have previously been identified as important considerations and potential barriers to the implementation of digital health technologies in clinical care [29]. Cognitive burden should also be further investigated, particularly in relation to patients' ability and willingness to engage with RPM programs. Likewise, the influence of language proficiency and the availability and role of caregiver support should be explored, particularly among patients who may require additional support to engage with and use digital technologies.

Beyond the monitoring technology itself, future studies should further define strategies for alert management, as these are likely to influence the workload of the HCPs involved. The COPD coach role also requires further specification, including qualifications, training, workload, and patient-to-coach ratios, as well as clear procedures for alert escalation and clinical decision-making. In addition, practical and organizational considerations, including after-hours coverage, alert triage, reimbursement, and medicolegal responsibilities, need to be addressed before the role can be implemented in routine COPD care. Furthermore, the integration and communication of RPM data within existing health care networks, particularly between hospital-based and community-based care providers, should be evaluated in the context of local care structures. Finally, financial, organizational, and legal considerations require further investigation at the institutional and national levels. As these aspects are highly context-specific, they were beyond the scope of the current study.

Study Limitations

This study has several limitations. First, the sample size was relatively small, and all participants were recruited from a limited geographical region in Belgium, which may limit the transferability of the findings to other settings. Second, patients who participated in the interviews had very low daily smartphone use (1 out of 5), potentially negatively influencing their opinions on the added value of RPM. As the digital literacy of the patient representatives in the workshops is unknown, there might be bias regarding the use of this technology in this patient population. Third, the use of convenience sampling of HCPs may have introduced selection bias. Fourth, only 1 informal caregiver (who also served as a patient representative) was included in the workshops. Further research is necessary on how informal caregivers can contribute to hybrid care for patients with COPD. Fifth, RPM was introduced as a wearable device, such as a smartwatch, during the workshops, which may have influenced participant views. Despite these limitations, this study offers valuable, in-depth insights into the perspectives of key stakeholders, providing a meaningful foundation for future research on the development of hybrid care pathways.

Conclusion

This user-centered design study developed a theoretical hybrid COPD care pathway through interviews and workshops with stakeholders involved in COPD care. To facilitate the integration of RPM, stakeholders identified the need for a new role, namely, the COPD coach. This individual could be responsible for all RPM-related activities, including device installation, patient instruction, monitoring and triaging alerts, and managing the discontinuation of RPM services. A dedicated RPM role may support the coordination and integration of RPM into routine COPD care, as stakeholders highlighted that a digital tool alone may not be sufficient to achieve this. As the proposed pathway is purely a conceptual model, further research is necessary to evaluate long-term adherence in real-world clinical settings before implementation.

Acknowledgments

The authors thank all patients and health care professionals (HCPs) involved in the interviews and workshops and the delegates of the Flanders COPD association, COPD vzw, for their valuable contributions. Without their expertise, this paper could never have been written.

Funding

This research project was cofunded by the European Union (ERDF) via the Interreg Meuse-Rhine (NL-BE-DE) program (grant ID IMR6-00019).

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Generative AI Use

The authors declare that generative AI was not used in the preparation of this manuscript.

XSL•FO
RenderX

Authors' Contributions

RK executed the study and wrote the first draft of the manuscript. JS executed the study and critically revised the manuscript. SH, JV, and CB critically revised the manuscript. DR provided clinical input and critically revised the manuscript. FW executed the study, supervised the project, and wrote the first draft of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
COREQ checklist.
[PDF File (Adobe PDF File), 432 KB - humanfactors_v13i1e93735_app1.pdf]

Multimedia Appendix 2
Interview guide: patients.
[DOCX File , 5488 KB - humanfactors_v13i1e93735_app2.docx]

Multimedia Appendix 3
Interview guide: physicians.
[DOCX File , 5488 KB - humanfactors_v13i1e93735_app3.docx]

Multimedia Appendix 4
Interview guide: occupational therapists and physiotherapists.
[DOCX File , 5397 KB - humanfactors_v13i1e93735_app4.docx]

Multimedia Appendix 5
Interview guide: chronic obstructive pulmonary disease (COPD) nurses.
[DOCX File , 5396 KB - humanfactors_v13i1e93735_app5.docx]

Multimedia Appendix 6
Interview guide: remote patient monitoring (RPM) nurses.
[DOCX File , 5489 KB - humanfactors_v13i1e93735_app6.docx]

Multimedia Appendix 7
Workshop guides for healthcare practitioners and patient representatives.
[DOCX File , 175 KB - humanfactors_v13i1e93735_app7.docx]

Multimedia Appendix 8
Infographic illustrating the proposed hybrid chronic obstructive pulmonary disease (COPD) care pathway. Changes from the current care pathway are marked with an asterisk (*).
[PDF File (Adobe PDF File), 332 KB - humanfactors_v13i1e93735_app8.pdf]

References

1. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: 2026 report. Global Initiative For Chronic Obstructive Lung Disease. 2026. URL: https://goldcopd.org/2026-gold-report-and-pocket-guide/ [accessed 2026-01-20]
2. Hurst JR, Vestbo J, Anzueto A, Locantore N, Müllerova H, Tal-Singer R, et al. Susceptibility to exacerbation in chronic obstructive pulmonary disease. N Engl J Med 2010;363(12):1128-1138 [FREE Full text] [doi: 10.1056/NEJMoa0909883] [Medline: 20843247]
3. Hartl S, Lopez-Campos JL, Pozo-Rodriguez F, Castro-Acosta A, Studnicka M, Kaiser B, et al. Risk of death and readmission of hospital-admitted COPD exacerbations: European COPD Audit. Eur Respir J 2016;47(1):113-121 [FREE Full text] [doi: 10.1183/13993003.01391-2014] [Medline: 26493806]
4. Cote CG, Dordelly LJ, Celli BR. Impact of COPD exacerbations on patient-centered outcomes. Chest 2007;131(3):696-704. [doi: 10.1378/chest.06-1610] [Medline: 17356082]
5. Vanhaecht K, De Witte K, Depreitere R, Sermeus W. Clinical pathway audit tools: a systematic review. J Nurs Manag 2006;14(7):529-537. [doi: 10.1111/j.1365-2934.2006.00705.x] [Medline: 17004963]

XSL•FO
RenderX

6. Seys D, Bruyneel L, Deneckere S, Kul S, Van der Veken L, van Zelm R, et al. Better organized care via care pathways: a multicenter study. PLoS One 2017;12(7):e0180398 [FREE Full text] [doi: 10.1371/journal.pone.0180398] [Medline: 28672030]
7. Plishka CT, Rotter T, Penz ED, Hansia MR, Fraser SA, Marciniuk DD, Saskatchewan COPD CPWs Research Group. Effects of clinical pathways for COPD on patient, professional, and systems outcomes: a systematic review. Chest 2019;156(5):864-877. [doi: 10.1016/j.chest.2019.04.131] [Medline: 31150639]
8. Kayyali R, Odeh B, Frerichs I, Davies N, Perantoni E, D'arcy S, et al. COPD care delivery pathways in five European Union countries: mapping and health care professionals' perceptions. Int J Chron Obstruct Pulmon Dis 2016;11:2831-2838 [FREE Full text] [doi: 10.2147/COPD.S104136] [Medline: 27881915]
9. Toschi E, Adam A, Frimpong N, Hurlbert R, Slyne C, Laffel L, et al. Hybrid care model: combining telemedicine and office visits for diabetes management in older adults with type 1 diabetes. Med Res Arch 2024;12(9):5728. [doi: 10.18103/mra.v12i9.5728] [Medline: 39822330]
10. Zhu Y, Gu X, Xu C. Effectiveness of telemedicine systems for adults with heart failure: a meta-analysis of randomized controlled trials. Heart Fail Rev 2020;25(2):231-243 [FREE Full text] [doi: 10.1007/s10741-019-09801-5] [Medline: 31197564]
11. Connell A, Black G, Montgomery H, Martin P, Nightingale C, King D, et al. Implementation of a digitally enabled care pathway (part 2): qualitative analysis of experiences of health care professionals. J Med Internet Res 2019;21(7):e13143 [FREE Full text] [doi: 10.2196/13143] [Medline: 31368443]
12. Lu JW, Wang Y, Sun Y, Zhang Q, Yan LM, Wang Y, et al. Effectiveness of telemonitoring for reducing exacerbation occurrence in COPD patients with past exacerbation history: a systematic review and meta-analysis. Front Med (Lausanne) 2021;8:720019 [FREE Full text] [doi: 10.3389/fmed.2021.720019] [Medline: 34568376]
13. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
14. Meiwald A, Gara-Adams R, Rowlandson A, Ma Y, Watz H, Ichinose M, et al. Qualitative validation of COPD evidenced care pathways in Japan, Canada, England, and Germany: common barriers to optimal COPD care. Int J Chron Obstruct Pulmon Dis 2022;17:1507-1521 [FREE Full text] [doi: 10.2147/COPD.S360983] [Medline: 35801119]
15. Vanhaecht K, Lodewijckx C, Sermeus W, Decramer M, Deneckere S, Leigheb F, et al. Impact of a care pathway for COPD on adherence to guidelines and hospital readmission: a cluster randomized trial. Int J Chron Obstruct Pulmon Dis 2016;11:2897-2908 [FREE Full text] [doi: 10.2147/COPD.S119849] [Medline: 27920516]
16. Jang S, Kim Y, Cho W. A systematic review and meta-analysis of telemonitoring interventions on severe COPD exacerbations. Int J Environ Res Public Health 2021;18(13):6757 [FREE Full text] [doi: 10.3390/ijerph18136757] [Medline: 34201762]
17. Miranda R, Oliveira MD, Nicola P, Baptista FM, Albuquerque I. Towards a framework for implementing remote patient monitoring from an integrated care perspective: a scoping review. Int J Health Policy Manag 2023;12:7299 [FREE Full text] [doi: 10.34172/ijhpm.2023.7299] [Medline: 37579426]
18. Claggett J, Petter S, Joshi A, Ponzio T, Kirkendall E. An infrastructure framework for remote patient monitoring interventions and research. J Med Internet Res 2024;26:e51234 [FREE Full text] [doi: 10.2196/51234] [Medline: 38815263]
19. Franzoi MA, Ferreira AR, Lemaire A, Rodriguez J, Grosjean J, Ribeiro JM, Nouhaud, et al. Implementation of a remote symptom monitoring pathway in oncology care: analysis of real-world experience across 33 cancer centres in France and Belgium. Lancet Reg Health Eur 2024;44:101005 [FREE Full text] [doi: 10.1016/j.lanepe.2024.101005] [Medline: 39444707]
20. Alami S, Courouve L, Lancman G, Gomis P, Al-Hamoud G, Laurelli C, et al. Organisational impact of a remote patient monitoring system for heart failure management: the experience of 29 cardiology departments in France. Int J Environ Res Public Health 2023;20(5):4366 [FREE Full text] [doi: 10.3390/ijerph20054366] [Medline: 36901372]
21. Nakken N, Janssen DJA, van den Bogaart EHA, Wouters EFM, Franssen FME, Vercoulen JH, et al. Informal caregivers of patients with COPD: home sweet home? Eur Respir Rev 2015;24(137):498-504 [FREE Full text] [doi: 10.1183/16000617.00010114] [Medline: 26324811]
22. Trivedi RB, Bryson CL, Udris E, Au DH. The influence of informal caregivers on adherence in COPD patients. Ann Behav Med 2012;44(1):66-72. [doi: 10.1007/s12160-012-9355-8] [Medline: 22422104]
23. Cruz J, Marques A, Machado A, O'Hoski S, Goldstein R, Brooks D. Informal caregiving in COPD: a systematic review of instruments and their measurement properties. Respir Med 2017;128:13-27 [FREE Full text] [doi: 10.1016/j.rmed.2017.05.002] [Medline: 28610666]
24. Abejirinde IO, Kishimoto V, Pfisterer KJ, Uddin N, McGuire K, Brual J, et al. An equity analysis of remote patient monitoring programs unveils assumptions on digital health equity. NPJ Digit Med 2025;8(1):320 [FREE Full text] [doi: 10.1038/s41746-025-01731-x] [Medline: 40442266]
25. Glynn L, Mc Cann M, Mc Cabe C. Smartphone applications supporting self-management programme for adults with chronic obstructive pulmonary disease: a scoping review. PLOS Digit Health 2024;3(6):e0000532 [FREE Full text] [doi: 10.1371/journal.pdig.0000532] [Medline: 38870123]
26. Yang F, Wang Y, Yang C, Hu H, Xiong Z. Mobile health applications in self-management of patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis of their efficacy. BMC Pulm Med 2018;18(1):147 [FREE Full text] [doi: 10.1186/s12890-018-0671-z] [Medline: 30180835]

XSL•FO
RenderX

27. Kooij L, Vos PJE, Dijkstra A, van Harten WH. Effectiveness of a mobile health and self-management app for high-risk patients with chronic obstructive pulmonary disease in daily clinical practice: mixed methods evaluation study. JMIR Mhealth Uhealth 2021;9(2):e21977 [FREE Full text] [doi: 10.2196/21977] [Medline: 33538699]
28. Tilquin F, Le Liepvre S, Balbolia S, Pirotais M, Le Guillou Y, Cornu J, et al. Short-term prediction of COPD exacerbations based on wearable vital sign monitoring. PLOS Digit Health 2026;5(5):e0001405 [FREE Full text] [doi: 10.1371/journal.pdig.0001405] [Medline: 42207714]
29. Ginsburg GS, Picard RW, Friend SH. Key issues as wearable digital health technologies enter clinical care. N Engl J Med 2024;390(12):1118-1127. [doi: 10.1056/NEJMra2307160] [Medline: 38507754]

Abbreviations

ADL: activities of daily living
AECOPD: acute exacerbation of COPD
COPD: chronic obstructive pulmonary disease
COREQ: Consolidated Criteria for Reporting Qualitative Research
ER: emergency room
GP: general practitioner
HCP: health care professional
RPM: remote patient monitoring

Edited by S Law; submitted 18.Feb.2026; peer-reviewed by T Yang, R Izadi, WJ He; comments to author 07.Jul.2026; revised version received 15.Oct.2026; accepted 15.Oct.2026; published 30.Oct.2026.

Please cite as:
Knevels R, Schils J, Haesevoets S, Vranken J, Burtin C, Ruttens D, Wouters F
Constructing a Theoretical Hybrid Care Pathway From Stakeholder Perspectives: User-Centered Design Study
JMIR Hum Factors 2026;13:e93735
URL: https://humanfactors.jmir.org/2026/1/e93735
doi:10.2196/93735
PMID:

©Ruben Knevels, Jelle Schils, Sarah Haesevoets, Julie Vranken, Chris Burtin, David Ruttens, Femke Wouters. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 30.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Development and User Testing of an Online Platform to Deliver School-Based Inclusive Physical Education Solutions: Mixed Methods Study

Sangeetha Mohanraj[1], PhD; Laurie A Malone[2], PhD; Sunkanmi Arogbokun[3], OD; Lauren Lieberman[4], PhD; Mohanraj Thirumalai[5], PhD

[1]SHP Research Collaborative, School of Health Professions, University of Alabama at Birmingham, 701 20th Street South, Birmingham, AL, United States

[2]Department of Occupational Therapy, School of Health Professions, University of Alabama at Birmingham, Birmingham, AL, United States

[3]Rehabilitation Science, School of Health Professions, University of Alabama at Birmingham, Birmingham, AL, United States

[4]Department of Kinesiology, Sport Studies and Physical Education, SUNY Brockport, Brockport, NY, United States

[5]Division of General Internal Medicine and Population Sciences, Heersink School of Medicine, University of Alabama at Birmingham, 701 19th Street South, ALGEN 580, Birmingham, AL, United States

Corresponding Author:
Mohanraj Thirumalai, PhD
Division of General Internal Medicine and Population Sciences, Heersink School of Medicine, University of Alabama at Birmingham, 701 19th Street South, ALGEN 580, Birmingham, AL, United States

Abstract

Background: Inclusive physical education (PE) plays a vital role in promoting participation and development among students with different abilities. However, many teachers do not have adequate tools to modify PE activities to meet these diverse needs. In addition, parents are essential partners, as their involvement helps reinforce strategies and provide useful information about their children. While online platforms provide a practical way to deliver such solutions, only a few have been intentionally created to support both teachers and parents in implementing inclusive PE learning. To address this gap, we codeveloped an online platform with adapted PE experts that digitizes an inclusion screening tool and delivers personalized, evidence-based inclusion strategies and resources to teachers and parents.

Objective: This study aimed to develop an online platform that provides inclusion strategies for PE teachers and to examine how teachers and parents perceived its usability, acceptability, and overall usefulness using a mixed methods approach.

Methods: A mixed methods design was used to evaluate the platform between September 2024 and March 2025. PE teachers (n=8) and parents of children with disabilities (n=8) were recruited through convenience sampling via national social media outreach and completed individual usability testing sessions via Zoom. Usability was assessed using the System Usability Scale (SUS), the Questionnaire for User Interaction Satisfaction (QUIS), and task performance metrics. Semistructured interviews were conducted to explore user experiences. Quantitative data were analyzed descriptively, while qualitative data were analyzed using codebook thematic analysis, with findings integrated using joint display analysis.

Results: All participants successfully completed the assigned tasks, with a few instances of minor difficulty (14 total errors across 136 task attempts). Participants completed both usability testing and semistructured interviews. Usability and satisfaction scores were high across both groups (SUS: mean 85.63, SD 9.14 for teachers and mean 84.69, SD 12.50 for parents; QUIS: mean 8.52, SD 0.82 for teachers and mean 8.00, SD 1.31 for parents, out of 9). Mixed methods integration showed convergence between positive usability ratings and perceived professional value, while also identifying challenges related to navigation, terminology, learning for first-time users, and mobile interface design.

Conclusions: The online platform provides strong usability and high satisfaction among PE teachers and parents. Future work will involve improving implementation and evaluating its impact on students' participation outcomes.

(JMIR Hum Factors 2026;13:e89476) doi:10.2196/89476

KEYWORDS

students with disabilities; adapted physical activity; inclusive physical education; teacher training; professional development; user testing; usability; mixed methods

XSL•FO
RenderX

Introduction

Inclusive physical education (PE) aims to ensure that all students, regardless of ability, can participate and benefit. This is especially critical given persistent disparities in physical activity. Children with disabilities consistently engage in less physical activity than their typically developing peers and face higher rates of obesity and health-related complications [1]. National guidelines from the US Department of Health and Human Services (2018) recommend that children engage in at least 60 minutes of physical activity daily to support their physical, mental, and emotional well-being. However, a recent study in Canada found that 91% of children and adolescents with disabilities did not meet daily physical activity guidelines, compared to far lower proportions among their typically developing peers [2]. This inactivity has clear negative consequences for physical, mental, and emotional health, and underscores a critical need for inclusive PE practices that ensure equitable access to physical activity opportunities for all students during the school day.

At the same time, simply placing students with disabilities into general PE classes does not automatically yield positive outcomes. Recent research highlights that inclusive experiences depend on intentional practices and support [3]. Traditional, ableist approaches where ability is measured by performative criteria continue to marginalize students with disabilities and limit participation for all learners [4]. Haegele et al [3] surveyed 99 adapted PE specialists and identified 460 distinct barriers and 473 facilitators to inclusion, spanning seven thematic categories. Notably, they found that teacher-related factors were the most commonly cited barriers, more so than environmental or policy factors. This suggests that the attitudes, training, and strategies of PE teachers largely determine whether an "integrated" class truly feels inclusive. Other researchers have also questioned whether typical integrated PE settings, without additional supports, provide meaningful inclusion [5]. However, Universal Design for Learning (UDL) has been identified as a proactive framework for addressing learner variability in PE. By embedding UDL principles, educators can design lessons that minimize barriers for all students, rather than relying on individual modifications after challenges arise [6]. Research shows that UDL-based approaches not only support students with disabilities but also enhance engagement and outcomes for their peers without disabilities [7,8]. In short, inclusive PE is a complex challenge: it offers clear benefits for students (social interaction, improved fitness, and enhanced well-being), but achieving those benefits requires overcoming significant practical barriers in day-to-day instruction [9].

Unfortunately, many general PE teachers report feeling underprepared to effectively include students with disabilities in their classes. A lack of formal training, limited experience, and insufficient resources contribute to significant barriers in delivering high-quality, inclusive PE [10]. This is consistent with broader findings that many PE teachers feel underprepared to implement inclusive pedagogies, often due to limited exposure to effective strategies and insufficient professional development [11]. Teacher-specific practices such as offering a variety of instructional strategies, adapting equipment and rules, and collaborating with support personnel are essential for reducing environmental barriers and fostering meaningful participation [12]. Teachers frequently cite inadequate support, limited professional development opportunities, and a lack of practical tools as major obstacles to fostering inclusive environments. Alarmingly, over half (54%) of PE teachers report receiving no professional development related to inclusive practices from their school districts [10]. This issue is highlighted in research by Polat et al [13], who conducted a needs analysis with 133 teachers: they concluded that teachers' knowledge regarding special education students is often insufficient, and educators overwhelmingly desire more training and support to teach inclusively. This lack of professional preparation has been identified as a widespread issue that must be addressed to improve educational outcomes for students with disabilities [14,15], indicating a pressing need to support PE teachers through better training, resources, and peer support.

Current approaches to preparing PE teachers for inclusive practice include preservice coursework, continuing professional development workshops, consultation with adapted PE specialists, and instructional frameworks such as UDL [12,16]. While these approaches provide important foundational knowledge, access is often inconsistent, time-limited, and may not provide ongoing, just-in-time support for teachers managing real-world classroom inclusion challenges [10]. Digital platforms offer an opportunity to extend professional learning by providing scalable, accessible, and on-demand support, particularly for educators who may lack regular access to specialized training resources [13].

In response to this need, the Lieberman/Brian Inclusion Rating Scale for Physical Education (LIRSPE) was developed as a 27-item instrument designed to assess the extent to which PE teachers implement inclusive practices for all students, including those with disabilities, in their programs. The LIRSPE has demonstrated strong reliability and validity [17], making it a valuable tool for evaluating inclusive practices in PE settings. The tool was developed primarily as a practice-oriented self-assessment instrument to support educators in reflecting on and improving their inclusive teaching practices, rather than solely as a research measure. The lack of digital integration constrains data collection, limits opportunities for sharing successful strategies, and reduces its potential as a tool for ongoing professional development. Although the LIRSPE is valuable for assessing inclusive PE practices and identifying areas for improvement, its original static format does not provide embedded implementation guidance or tailored training resources. This is important given evidence that PE teachers often require practical professional development support to effectively implement inclusive practices [16,18]. As digital technologies continue to transform educational practice, there is an opportunity to adapt the LIRSPE into a more accessible and interactive format. A digital platform could enhance its use by supporting real-time feedback, facilitating resource sharing, and promoting sustained improvements in inclusive teaching practices across varied educational contexts.

Parents are important stakeholders in inclusive PE, as they provide valuable insight into their child's needs, support participation beyond the school environment, and often advocate

XSL•FO
RenderX

for appropriate inclusion and accommodations [19]. However, like PE teachers, parents often face similar barriers in accessing clear, actionable guidance for supporting inclusive PE at home and in the community. Including parents in the evaluation of the platform was therefore important to ensure that the resources and recommendations were understandable, relevant, and supportive of collaborative efforts between families and educators to promote inclusive PE participation.

To address these limitations, the present project developed a web-based platform to expand the use of the LIRSPE in supporting inclusive practices in PE. A mobile-friendly digital version of the LIRSPE, accessible via smartphones and tablets, was created, with a self-evaluation interface for teachers and para-PE teachers. In addition, the platform integrated a library of evidence-based solutions to support inclusive instruction. The design of the proposed platform was further informed by UDL, which emphasizes reducing barriers through multiple means of engagement, representation, and action and expression [7-9]. Consistent with these principles, the platform provides inclusion resources in multiple formats, including videos, downloadable documents, and interactive discussion forums, allowing users to access information in ways that align with their needs and preferences. In addition, tailored recommendations generated from users' LIRSPE responses provide individualized support, aligning the platform with the UDL goal of addressing learner variability and promoting inclusive participation [7-9]. This project followed a 2-phase design. Phase 1 involved the development of the platform, including the digitization of the LIRSPE and the integration of evidence-based inclusion resources, and phase 2 involved a usability evaluation of the platform with PE teachers and parents of children with disabilities. The present paper reports directly on phase 2. Accordingly, this study aimed to examine PE teachers' and parents' perceptions of the platform's usability, acceptability, and overall usefulness using a mixed methods approach.

Methods

Development of the System

The web-based platform was built on WordPress, chosen for its robust content management capabilities and flexibility for nontechnical users. To digitize the LIRSPE, a custom software plug-in was developed to create a dynamic survey interface. This feature adapts to the specific user, loading distinct questions for teachers, parents, or students, and using automated logic to calculate scores and assign credits immediately upon submission. Based on these scores, the system triggers the Adapted PE Decision Support Tool, delivering personalized, evidence-based inclusion solutions and recommended resources to the user. Users can also generate and download PDF summaries of their results and completion certificates directly from the dashboard.

To support ongoing professional development and peer interaction, the platform integrates a discussion forum (powered by bbPress) with automated email notifications to alert users of new replies. User access, registration, and role management are handled through a secure interface (Ultimate Member). The visual design used the Divi theme, which was customized to ensure a consistent, accessible layout across devices, enabling intuitive navigation between the survey, results, and resource libraries. Finally, a custom administrative dashboard was built to facilitate data management, enabling the research team to monitor submission dates, analyze demographic trends, and export data for reporting.

Development of Inclusion Solutions

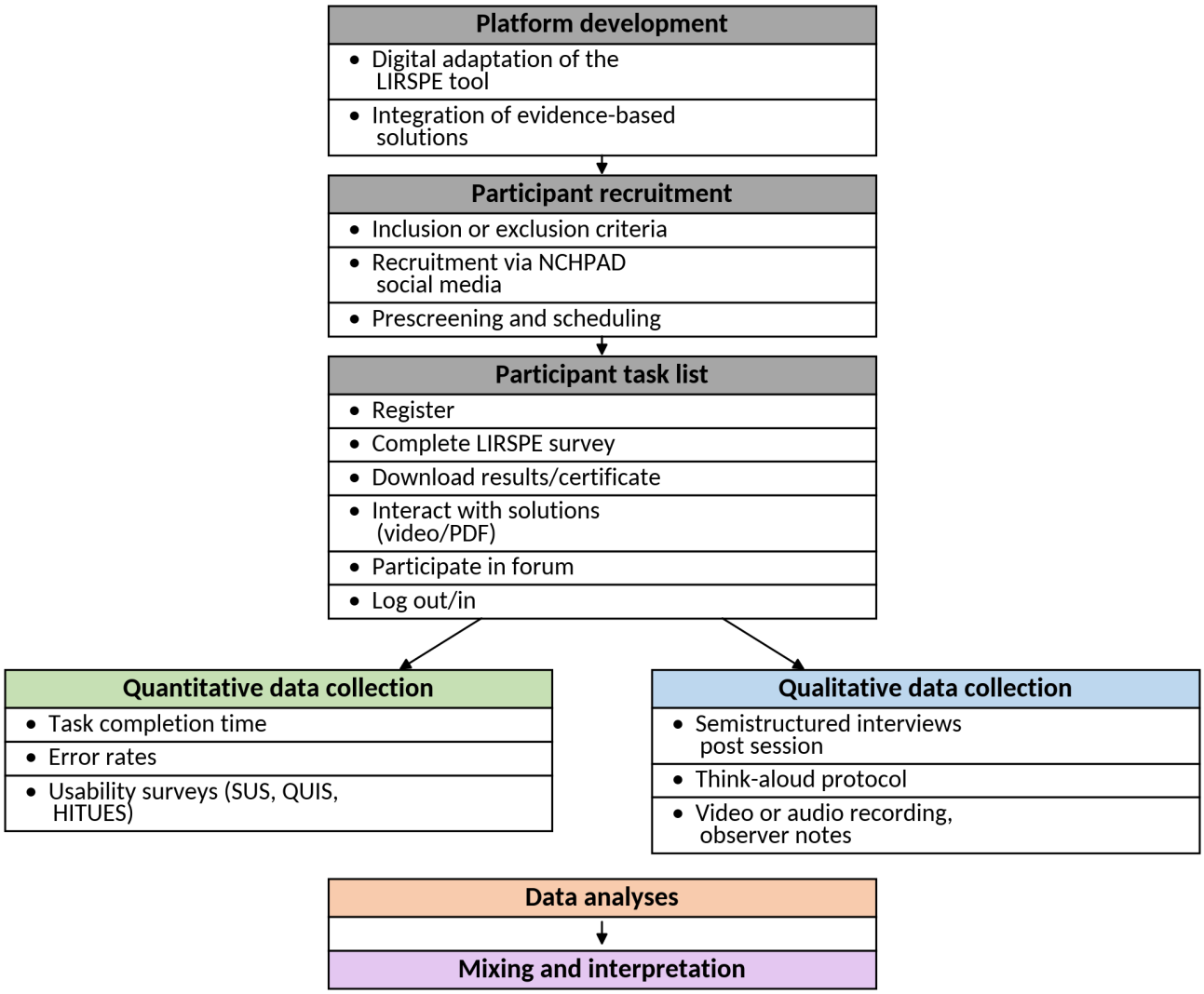
To develop the recommended solutions provided to participants upon completion of the online LIRSPE, 2 long-standing experts in adapted PE conducted a comprehensive resource review. This process involved a comprehensive search of relevant literature and online content, including YouTube instructional videos, Google searches, and materials from professional organizations such as SHAPE America and Human Kinetics. The experts also incorporated resources they had previously developed and, where gaps were identified, created new materials to address specific areas of need. The platform was designed to identify barriers to inclusive PE and provide personalized, evidence-based solutions and practical resources to improve knowledge, support decision-making, and increase confidence in implementing inclusive practices. This multisource approach ensured that the recommended solutions were both evidence-informed and practically applicable for PE teachers seeking to improve the inclusion of students with disabilities.

Mixed Methods Research Design

This study examined the strengths and weaknesses of the PE inclusion solutions platform using the International Organization for Standardization (ISO) usability criteria of effectiveness, efficiency, and satisfaction [20]. Acceptability was reflected in participants' perceptions of the platform's appropriateness and willingness to engage with or recommend its use, while usefulness referred to the perceived practical value of the platform in supporting inclusive PE decision-making and resource access. A mixed methods approach was used to provide a comprehensive assessment by integrating quantitative and qualitative insights. Specifically, we used the iterative convergent mixed methods design described by Alwashmi et al [21], which involves multiple cycles of concurrent data collection and analysis. This design allowed us to identify and address usability concerns in real time and ensured alignment between research objectives, data generation, analysis, and interpretation. The iterative process involved integrating quantitative usability metrics (task completion time, errors, and questionnaire scores) with qualitative findings from the think-aloud protocol and semistructured interviews. Following each participant session, the research team reviewed the findings, and any technical glitches identified were addressed before the next participant completed testing. More substantial refinements based on recurring feedback are described in the "Results" section. Figure 1 presents the procedural steps followed in this process [21].

XSL•FO
RenderX

Figure 1. Procedural flow of a mixed methods design for user testing of an online platform to deliver school-based physical education (PE) inclusion solutions. HITUES: Health Information Technology Usability Evaluation Scale; LIRSPE: Lieberman/Brian Inclusion Rating Scale for Physical Education; NCHPAD: The National Center on Health, Physical Activity & Disability; QUIS: Questionnaire for User Interaction Satisfaction; SUS: System Usability Scale.



Participants

The PE inclusion solutions platform is designed for use by PE teachers and parents of children with disabilities. Therefore, iterative user testing was conducted with a mix of parents and PE teachers. Participants were recruited nationally through the National Center on Health, Physical Activity & Disability (NCHPAD)'s social media site via public posts that included promotional videos and a flyer describing the study; no direct messaging or individualized outreach was used beyond these public posts. Inclusion criteria were the following: (1) aged 18 years or older, (2) fluent in English, (3) able to use a tablet computer, and (4) parent of a child with a disability or working as a PE professional. Recruitment and data collection were conducted between September 2024 and March 2025. A total of 25 individuals were reached during recruitment. Recruitment continued until the target sample of 16 eligible participants (8 PE teachers and 8 parents) was achieved, consistent with recommendations for iterative usability testing [9,22]. Eligibility screening was conducted by a member of the research team (SM). Individuals who did not meet these criteria were excluded. Eligible participants were provided with the study information and scheduled for a 1-hour individual Zoom session. Participants who did not meet the eligibility criteria or declined participation after screening were not enrolled in the study.

Ethical Considerations

The study was approved by the University of Alabama at Birmingham Institutional Review Board (IRB-300010295). All participants provided informed consent prior to participation. For the online component, oral informed consent was obtained before participants accessed the platform or completed study activities. Participation was voluntary, and all data were deidentified and stored securely to protect participant privacy and confidentiality. Participants received a US $25 gift card upon completion of the study session.

Data Collection

Data were collected through one-on-one user testing sessions with each participant conducted via Zoom. During those sessions, the researcher provided additional details about the testing process and reiterated that participation was voluntary. The session included navigating various web pages and testing features or functions. The testing session was video recorded

XSL•FO
RenderX

(computer screen only) and audio recorded. A research staff member took written notes from their observations of participants during the usability tasks. Participants were asked to complete the following tasks:

- Task 1: register as a PE professional or a parent.
- Task 2: complete the LIRSPE survey and find a solution or recommendation on how to include a child with a disability (imaginary in case they do not have one) in PE activities. Download the results copy and the certificate.
- Task 3a: download the results copy.
- Task 3b: download the certificate.
- Task 4a: from your results, check a solution that needs to be improved and comment on a video.
- Task 4b: from your results, check a solution that needs to be improved and comment on a PDF. Download or print the specific PDF.
- Task 5a: forum—post a response to the existing post "this is a test" and create a new post.
- Task 5b: forum—create a new post.
- Task 6: log out and log in.

The testing involved a "think-aloud protocol" while participants navigated the system [23]. Participants were asked to spontaneously report everything that went through their minds while performing a set of specific tasks. During testing, the system's efficiency was assessed by task completion time, and its effectiveness was evaluated based on the errors encountered while performing the tasks. System satisfaction was measured using three validated questionnaires: the System Usability Scale (SUS), a 10-item instrument with a 5-point Likert scale assessing overall usability [24]; the Questionnaire for User Interface Satisfaction (QUIS), which evaluates user satisfaction across multiple interface domains using 9-point rating scales [25]; and the Health Information Technology Usability Evaluation Scale (HITUES), a 20-item instrument with 5-point Likert response options assessing impact, usefulness, ease of use, and user control. All 3 questionnaires were self-administered electronically, with participants completing them independently via a Qualtrics survey link shared during the same Zoom session immediately after the usability tasks, rather than having items read aloud by the researcher or completed on paper. Internal consistency for these instruments has been reported as good to excellent in previous validation studies [26]. After completing the surveys, participants took part in a semistructured interview conducted by a member of the research team (SM). The interviews followed a preplanned interview guide (Multimedia Appendix 1) that explored participants' experiences, perceptions, opinions, and suggestions regarding the platform while allowing flexibility to discuss issues they considered important. Interviews lasted approximately 10 minutes. All recordings were transcribed for qualitative analysis.

Data Analysis

Qualitative data (think-aloud and interviews) were transcribed using Zoom and cleaned manually for accuracy before being imported into NVivo (Lumivero) for analysis. Data were analyzed using a codebook thematic analysis approach [27,28]. Two researchers (SA and SM) independently coded the transcripts using a shared codebook. The codebook was developed using a hybrid approach: core coding domains (platform navigation, screen layout, terminology, and learning) were derived deductively from the structured usability evaluation framework assessed by the quantitative instruments (SUS, QUIS, and HITUES) and the interview guide, while additional domains, for example, suggestions for the future, were added inductively as they emerged from the transcripts. SA and SM met iteratively to compare codes, resolve discrepancies, and reach consensus on the final coding structure. Rigor was addressed through these study-specific procedures [28]: credibility was supported by dual coding and a consensus process; transferability was supported by detailing the sample (8 PE teachers and 8 parents recruited via NCHPAD), the Zoom-based testing context, and study procedures; dependability was supported by documenting coding decisions and codebook revisions in NVivo; and confirmability was supported by SA and SM checking each code against its source (transcripts). Descriptive analysis (mean, SD and median, IQR) of quantitative data was conducted.

Iterative Platform Refinement During User Testing

Throughout user testing, the platform was refined iteratively as usability sessions progressed. After each participant session, the research team reviewed the observed usability issues, technical glitches, and participant feedback. Technical problems that affected task completion (eg, navigation errors, delayed system responses, and duplicated forum posts) were corrected before the next participant completed testing whenever feasible. This approach ensured that subsequently identified issues reflected new usability concerns rather than previously recognized technical faults.

Results

Participants Demographics

Participants in the PE teacher (n=8) and parent (n=8) groups underwent user testing. Table 1 provides the demographics of the teachers, including their age (mean 41, SD 14 y), job role, grade level taught, highest education level, state of current employment, and years of work experience. Table 2 provides the demographics of the parents of a child with a disability. Iterative modifications to the platform were made throughout testing in response to observed usability issues and participant feedback, as described in the *Methods* section and summarized in Multimedia Appendix 2.

XSL•FO
RenderX

Table . Physical education (PE) teacher demographics.

ID	Age (y)	Sex	Job role	Grade-level teaching	Highest education	Work state	Years of work experience	Device used for testing
PE1	29	Male	PE teacher in New Hampshire	Elementary	Master's degree	New Hampshire	6	Laptop
PE2	51	Female	PE teacher in New York	Elementary, middle, and high	Master's degree	Pennsylvania	20	Laptop
PE3	40	Female	Adapted PE specialist	Unified PE and sports (inclusion model in her district), and aged 3 - 18 years	Adapted PE specialist credentials, MS	California	18	Laptop
PE4	31	Female	PE teacher in South Carolina	Elementary, middle, and high	Master's degree	South Carolina	5	iPhone
PE5	40	Female	N/A[a]	N/A	Master's degree	New York	13	Laptop
PE6	32	Male	Adapted PE specialist	High school (grades 9 - 12), adapted and general PE; previously elementary and K-12	Master's degree, pre-doc	Wisconsin	7	Laptop
PE7	72	Male	Prep adapted PE teachers (undergrad and grad)	Not working with schools directly	PhD	Wisconsin	N/A	Laptop
PE8	36	Male	PE teacher	Elementary, 1-year middle school	BS PE	Alabama	8	Laptop

[a]N/A: not available.

XSL•FO
RenderX

Table . Parent demographics.

ID	Age (y)	Sex	Job status	Education	Age of the child with disability (y)	Residing state	Device used for testing
P1	52	Female	Part-time physical therapist at an outpatient pediatric clinic	DPT, clinical specialty certification	18 and 15	Maryland	Laptop
P2	54	Female	N/A[a]	Graduate degree in educational technology and leadership	17	Maryland	Laptop
P3	46	Male	Information technologist	College degree	16	Georgia	Laptop
P4	44	Female	Clinical psychologist	PhD	17	Ohio	Laptop
P5	44	Female	Occupational therapist	OT associate	12	Kentucky	Laptop
P6	42	Female	N/A	Master's degree	10	Idaho	Laptop
P7	44	Female	N/A	PhD	10	Oregon	Laptop
P8	35	Female	Medical assistant (nonprofit)	College degree	9	Florida	Laptop

[a]N/A: not available.

Multimedia Appendix 3 summarizes task completion outcomes, observed difficulties, and errors encountered by PE teachers and parents during usability testing of the PE inclusion solutions platform.

Effectiveness

Errors during task performance and task completion were used to assess system effectiveness. Overall, both PE teachers and parents were able to complete most assigned tasks with minimal errors (Multimedia Appendix 3). Among PE teachers, the majority of errors occurred during forum interactions, with 4 participants encountering difficulties when responding to existing posts and 2 participants experiencing issues when creating new posts, mainly due to navigation delays or duplicate submissions. Parents similarly completed most tasks successfully, with only isolated difficulties observed in account creation, logging in and out, and commenting on existing forum posts. These findings suggest that the platform effectively supports core tasks, with minor challenges concentrated around interactive navigation features.

Efficiency

The efficiency of teachers and parents in completing the various tasks is shown in Table 3.

Table . Time taken for completing the tasks in seconds for physical education (PE) teachers and parents[a].

Task	PE teachers, mean (SD)	PE teachers, median (IQR)	Parents, mean (SD)	Parents, median (IQR)
Account creation and login	138.75 (87.85)	113.5 (87.8-145.3)	114.63 (66.44)	92.5 (70.5-135.3)
LIRSPE[b] survey	786.00 (429.47)	805.0 (547.0-923.3)	635.88 (247.71)	636.0 (508.5-870.3)
Download result copy	22.75 (46.46)	6.5 (4.5-9.0)	37.17 (50.26)	15.0 (11.3-30.0)
Download certificate	4.25 (1.92)	4.0 (3.0-5.0)	N/A[c]	N/A
Solution PDF interaction	84.00 (54.79)	70.0 (41.5-105.5)	103.75 (123.30)	38.0 (27.8-133.5)
Video interaction	70.50 (42.54)	58.0 (36.3-86.3)	110.63 (83.15)	111.0 (56.8-140.5)
Forum: comment on existing post	56.13 (27.85)	48.5 (34.5-77.0)	84.50 (75.90)	60.5 (37.5-92.8)
Forum: create new post	81.13 (19.42)	76.5 (72.3-81.8)	63.63 (25.80)	60.0 (45.3-71.8)
Logout and login	25.63 (11.28)	24.5 (21.8-27.0)	29.88 (18.84)	22.5 (15.8-39.0)

[a]All time measured in seconds.

[b]LIRSPE: Lieberman/Brian Inclusion Rating Scale for Physical Education.

[c]N/A: not applicable.

XSL•FO
RenderX

Task completion times during usability testing were used to assess system efficiency. Table 3 presents the task completion times (in seconds) for PE teachers and parents, including both mean (SD) and median (IQR) values to facilitate comparison.

During the think-aloud process, some participants deviated from assigned tasks to provide additional feedback, which may have influenced completion times. Overall, both PE teachers and parents completed simpler tasks efficiently, while more complex tasks requiring interpretation or multiple steps, such as survey completion (PE teachers: mean 786.00, SD 429.47 s and median 805.0, IQR 547.0-923.3 s; parents: mean 635.88, SD 247.71 s and median 636.0, IQR 508.5-870.3 s) and account creation/log-in (PE teachers: mean 138.75, SD 87.85 s and median 113.5, IQR 87.8-145.3 s; parents: mean 114.63, SD 66.44 s and median 92.5, IQR 70.5-135.3 s), took longer. Parents also spent more time on video interaction (mean 110.63, SD 83.15 s and median 111.0, IQR 56.8-140.5 s), suggesting additional navigation demands.

Satisfaction

Satisfaction scores, as seen in Table 4, were assessed using 3 validated instruments: SUS, HITUES, and QUIS.

Table . Usability survey scores for teachers and parents.

Instrument	PE teachers, mean (SD)	Parents, mean (SD)
SUS[a]		
Total	85.63 (9.14)	84.69 (12.50)
HI-TUES[b]		
Total	4.60 (0.38)	4.69 (0.23)
Impact	4.83 (0.31)	4.92 (0.24)
Perceived usefulness	4.50 (0.51)	4.76 (0.35)
Perceived ease of use	4.80 (0.35)	4.78 (0.27)
User control	4.33 (0.76)	4.08 (0.79)
QUIS[c]		
Total	8.52 (0.82)	8.00 (1.31)
Overall reaction to software	8.03 (1.59)	8.13 (1.06)
Screen	8.35 (1.08)	7.80 (1.66)
Terminology and system information	8.03 (2.68)	8.10 (1.36)
Learning	9.69 (0.40)	8.63 (1.57)
System capabilities	9.06 (1.24)	6.69 (2.25)

[a]SUS: System Usability Scale.

[b]HI-TUES: Health Information Technology Usability Evaluation Scale.

[c]QUIS: Questionnaire for User Interaction Satisfaction.

PE teachers and parents reported high usability and satisfaction. SUS scores were high for both PE teachers (85.63) and parents (84.69), exceeding the benchmark score of 68. On the HITUES, PE teachers and parents reported similarly high overall scores (4.60 and 4.69, respectively), with the highest ratings in impact (4.83 and 4.92) and perceived ease of use (4.80 and 4.78), while user control received comparatively lower ratings (4.33 and 4.08). QUIS findings also reflected positive satisfaction, with overall mean scores of 8.52 for PE teachers and 8.00 for parents, particularly in learning (9.69 and 8.63), although parents reported lower ratings for system capabilities (6.69), suggesting opportunities for functional refinement.

Qualitative Findings

Qualitative findings were organized into structured usability evaluation domains derived from the study's codebook, including platform navigation, screen layout, terminology and learning factors, overall experience with the system, and suggestions for the future.

Platform Navigation

Many PE educators and guardians commended the system's intuitive nature for solution retrieval. A professional participant characterized the survey's organization as "logical and well-categorized." Nonetheless, some parents advised simplifying access to critical resources, such as survey results or video libraries, to minimize clicks. While certain parents accepted the scrolling, others deemed it excessive. Another participant suggested that "collapsible headings for grouping solutions would mitigate overwhelm." There was a prevalent request for reducing scroll length through content reorganization or the inclusion of a "jump to" feature.

Both demographic groups experienced challenges in navigating the forums. A participant noted that the platform lacked intuitiveness for posting forum inquiries. One recommendation was to reposition the "Create New Topic" button to the top of the page.

XSL•FO
RenderX

Screen Layout

A PE instructor observed that disorganized mobile layouts hindered content readability (PE7). Although many parents valued the material's organization, some raised issues about the text size in PDF formats. Recommendations for enhancement included expanding white space; increasing font size, or text enlargement and screen reading compatibility; adding visual indicators to improve readability and lessen cognitive burden; and more diverse images of kids with disabilities.

Terminology and Learning Factors

Some PE teachers valued the formal tone while parents appreciated clear language, yet others found the survey wording too complex. An educator criticized the use of advanced terminology, suggesting it alienated nonexperts: "The survey questions used '$50 words' when '$5 words' would do….it felt like it was designed by PhDs for PhDs" (PE1). A parent expressed the need for simpler language, indicating a desire for clarity. Multiple parents recommended adding definitions or tooltips to enhance understanding and accessibility.

Overall Experience and Suggestions for the Future

Overview

Parents acknowledged the system's value while expressing minor concerns. One parent rated it positively but had doubts about its correctness. Another parent strongly recommended it as transformative for advocacy. When asked for the most favorite feature, all parents mentioned the detailed solutions provided. The PE teachers regarded the platform as a valuable resource for inclusive education. Participants recognized its promise but requested enhancements for better efficiency. Recommendations for improvement included the addition of progress tracking and clearer error messaging to improve user experience. Participants also suggested the following in the future version of the system: resources to buy adaptive equipment, federal laws for each state, a checklist for para-PE teachers, and search options to find solutions without going through the survey.

Joint Display

The joint display provides a visual depiction of the integrated qualitative and quantitative (QUIS score) data (Tables 5 and 6).

Table . Joint display of quantitative and qualitative findings from the Questionnaire for User Interaction Satisfaction (QUIS) among physical education (PE) teachers.

Domain	QUIS, mean (SD)	Qualitative findings	Representative quote	Integration
Overall reaction	8.03 (1.59)	Valuable resource; most PE teachers would recommend the system.	"I'd be more inclined to use this during my prep periods. if I present at a State convention, I would probably use this resource." (PE6)	High score converges with perceived professional value and willingness to recommend.
Screen	8.35 (1.08)	Layout was generally easy to navigate, but spacing was an issue on mobile.	"Everything was labeled really nicely. The dropdown menus made a lot of sense." (PE5)	High score converges with positive desktop layout feedback.
Terminology	8.03 (2.68)	Mostly easy to understand, though some wording was viewed as too technical.	"Discussion boards used simple, precise wording anyone could understand." (PE2)	Qualitative data complements the score.
Learning	9.69 (0.40)	Mixed experiences on user-friendliness and efficient access to PE solutions.	"It's easy to find and look through things to find stuff." (PE2)	High score mostly converges with positive learning feedback, though workflow efficiency can improve.
System capabilities	9.06 (1.24)	Content was viewed as reliable; navigation was mostly easy, but forum speed was slow.	"Navigation, it was pretty smooth, pretty easy." (PE8)	High score converges with content reliability and navigation, with minor technical issues noted.

XSL•FO
RenderX

Table . Joint display of quantitative and qualitative findings from the Questionnaire for User Interaction Satisfaction (QUIS) among parents of children with disabilities.

Domain	QUIS, mean (SD)	Qualitative findings	Representative quote	Integration
Overall reaction	8.13 (1.06)	Beneficial and supportive; parents would recommend the system.	"The whole idea itself is commendable." (P3)	High score converges with perceived value and recommendation of the system.
Screen	7.80 (1.66)	Mixed reviews on layout; some found the content visually overwhelming.	"The layout is well-organized, with clear sections and headings." (P7)	Qualitative findings expand on the lower quantitative score.
Terminology	8.10 (1.36)	Mixed reviews; language was generally clear but may need simplification for some parents.	"The language used in the system is clear and understandable." (P7)	Qualitative themes expand the score.
Learning	8.63 (1.57)	Mixed experiences on user-friendliness; some users felt unsure during tasks.	"It was very user friendly, easy, not overly complicated." (P8)	High score converges with positive feedback while leaving room for improvement.
System capabilities	6.69 (2.25)	Mixed reviews on navigation, with suggestions to improve features and speed.	"After submitting a comment, it was difficult to return to the survey results; adding a back button would be helpful." (P2)	Lower score aligns with navigation and feature-related concerns.

Overall satisfaction was high for both PE teachers (8.03) and parents (8.13), a finding that converges with the universal theme of the system being a "valuable resource." This high rating is expanded by the qualitative data, indicating satisfaction is rooted in the system's mission and purpose. Furthermore, both groups found the system exceptionally quick to master (learning between 9.69 and 8.63), yet this still diverges with "mixed experiences" due to functional flaws and complex language.

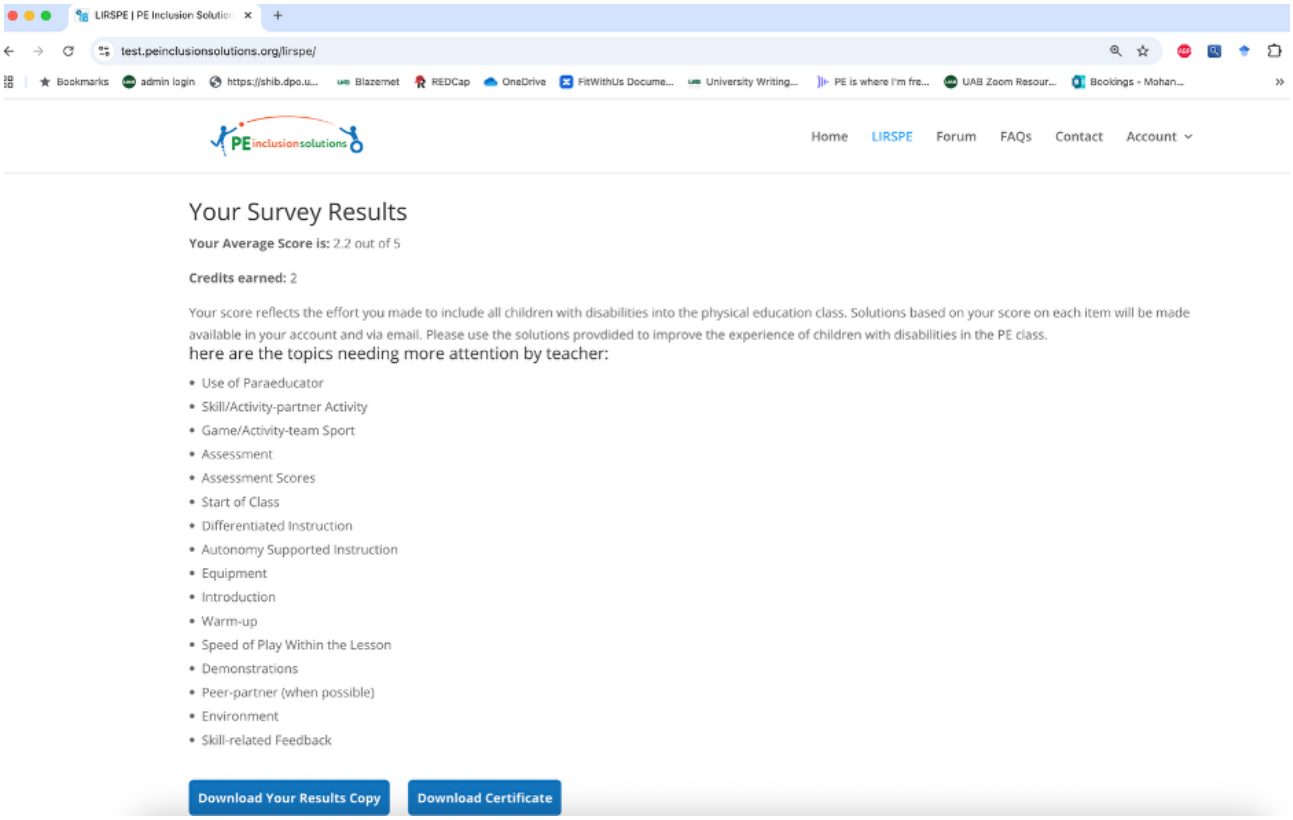
The most critical finding was the divergence in the system capabilities domain. PE teachers reported a 9.06 score for system capabilities converging with expert-level content trust. In sharp contrast, the parent group scored significantly lower (6.69), which was complemented by qualitative data detailing specific functional failures and navigation. While the PE teachers scored well (8.35) in the screen domain, the parents group scoring lower (7.80) was expanded by complaints of visual clutter that acted as a barrier for the nonexpert group. Regarding the terminologies, for PE teachers, the qualitative data complements the score, confirming that terms are generally appropriate for experts but showing inconsistency (high SD) for a few. For the parents group, the qualitative themes expanded the score, emphasizing the need for simplification and lower reading levels.

Explanation of Sections

Participants requested additional information on what to expect before starting the survey (Multimedia Appendix 4), details on what the score means to be moved closer to the score. P5 said, "When directly results show, it is confusing" (Figure 2). The explanation of scoring is provided before the start of the survey. Participants also requested an explanation for the different terms and expansion of abbreviations used in the demographic forms as different jargon is used in different states.

XSL•FO
RenderX

Figure 2. Screenshot of the Lieberman/Brian Inclusion Rating Scale for Physical Education (LIRSPE) survey results display interface. PE: physical education.



Forum

Overall, most of the participants in both groups had difficulty with the system speed while performing tasks in the forum. There was confusion in finding the add new topic button. PE1 said, “To add new topic, it takes time to scroll down and find the place to add new topic or difficult to find it.”

Demographic Form

Participants had confusion with password setup as the requirements for the password were not provided ahead. Similarly, some of them had difficulty using the DOB (date of birth) calendar option provided.

System Refinements Following User Testing

During iterative user testing, several system modifications were made to enhance platform usability (Multimedia Appendix 5); these changes are summarized in Multimedia Appendix 2.

Discussion

User Experience: Parents vs Teachers

The study findings suggest that the PE inclusion solution is usable and acceptable for both PE teachers and parents of children with disabilities. Both groups completed simple tasks with ease, while complex, multistep tasks were more difficult for both groups, particularly for parents. This difference across task types suggests that task complexity, rather than the platform’s overall design, was the main source of user difficulty. Furthermore, both groups struggled with tasks involving videos, PDFs, or the downloading of results, suggesting that navigation for resource-based activities is not intuitive. The reason for the difference between groups could be their level of familiarity with PE class activities and the terminology used in the surveys, since PE teachers work with this content regularly while parents typically do not. Some participants also deviated from the tasks, especially during the survey portion of the think-aloud process, and engaged in longer discussions, which may reflect their interest in the content rather than difficulty using the platform. Usability challenges were primarily observed in tasks involving complex navigation and access to platform resources, suggesting targeted rather than widespread issues. This pattern indicates that the platform supports core functionality but would benefit from refinement in more cognitively demanding workflows. This is consistent with established benchmarks indicating that scores above 80 reflect excellent usability.

It is important to note that the LIRSPE survey was originally developed as an assessment tool for PE teachers, with terminology and constructs tailored to professional practice [18]. When digitized for this platform, the tool was extended to parent users without a corresponding adaptation of its content or validation for this new group, and parent-specific use does not appear to have been a formal design consideration during digitization. As a result, some parents in this study expressed confusion or uncertainty regarding specific terms and survey items, which is understandable given the tool’s educator-focused design and may partly explain parents’ comparatively lower system-capability ratings. This suggests that fully meeting parent users’ needs may require a parent-adapted tool that is designed with parent input, rather than the current teacher-focused instrument. This expert-novice gap reflects a well-documented expertise reversal effect, in which design elements that support

XSL•FO
RenderX

less experienced users can be redundant or even burdensome for domain experts, and vice versa [29]. It is also consistent with the broader literature on parental involvement in physical education, which highlights that parents engage with PE content and terminology differently than PE professionals given their differing roles and familiarity with the subject [19].

The divergence between positive quantitative ratings and qualitative reports of confusion may reflect differences in what the evaluation measures captured. Standardized usability and usefulness measures assessed participants' overall perceptions of the platform, whereas qualitative feedback captured specific task-level challenges encountered during direct interaction. Participants may also have valued the platform's perceived relevance and learning potential despite difficulties with navigation or unfamiliar terminology. This discrepancy is itself a notable finding: parents' SUS scores were close to those of PE teachers and well above the accepted usability benchmark, yet qualitative data indicated that parents struggled to understand and navigate core survey content. This suggests that standardized usability scales, while useful for benchmarking overall ease of use, may not capture comprehension-level barriers faced by lay users of a tool originally designed for domain professionals. For future dual-audience digital tools, this points to a substantive design implication: usability and comprehension should be treated as distinct constructs and evaluated separately, since a high SUS score alone can mask meaningful gaps in understanding for nonexpert users.

Overall, the platform supports basic tasks but becomes less efficient when workflows require multiple steps or interpretation. Streamlining navigation and simplifying complex tasks would improve efficiency for both groups, with the largest gains for parents. The slow speed of the system could have resulted in errors, especially in the forum.

These findings are consistent with prior research showing that usability, clear navigation, and understandable terminology are important for the adoption of digital educational tools, particularly among users with different levels of expertise [13,21]. While digital platforms can improve access to training and support, design complexity and professional terminology may reduce usability for nonexpert users such as parents. This is consistent with established SUS benchmarks indicating that scores above 80 reflect excellent usability [24], suggesting the platform performs well overall despite the specific challenges noted above.

These findings also align with prior research showing that PE teachers often require additional support and training to implement inclusive practices effectively, and that educational technology adoption is influenced by usability, perceived relevance, and user confidence. The challenges observed among parent users further highlight the importance of designing inclusive digital tools that accommodate users with differing levels of expertise and familiarity with educational terminology [14,15].

This study contributes to the literature by extending LIRSPE beyond assessment into a digital platform that provides tailored inclusion solutions and practical resources. Including both PE teachers and parents also offers insight into how different stakeholder groups engage with inclusive digital tools. This platform may serve as a scalable resource for supporting inclusive PE, particularly where in-person training or specialist guidance is limited. These findings also underscore the importance of designing accessible educational technology for users with varying levels of professional expertise [13].

Future Research

Future research should evaluate the platform in larger and more diverse populations and assess its effect on inclusive PE knowledge and practice over time. Studies should also examine implementation in real-world school settings, including potential barriers to adoption, teacher and organizational readiness, scalability, sustainability, maintenance and updating of platform content, and integration into teacher professional development programs. In addition, theory-driven frameworks, such as the self-determination theory–based classification of teacher motivational behaviors proposed by Ahmadi et al [30], could be used to better understand how the platform supports autonomy, competence, and relatedness in inclusive PE practice. Mapping the platform's strategies and digital features onto these motivational behaviors may strengthen the theoretical basis and replicability of future evaluations. This approach could move research beyond usability outcomes toward examining behavioral and instructional mechanisms that influence teachers' inclusive practices and the sustained implementation of inclusive PE.

Limitations

This study has several limitations. Although the sample size was appropriate for usability testing, the findings should be interpreted as preliminary and may not be generalizable to broader populations. The inclusion of both PE teachers and parents introduced diverse user perspectives and levels of familiarity with inclusive PE concepts, which may have influenced task performance and interpretation differently. In addition, this study focused on usability, acceptability, and perceived usefulness rather than evaluating the effectiveness of the platform in improving inclusive PE knowledge or practice. Recruitment through NCHPAD's social media channels may have introduced self-selection bias, favoring individuals already engaged with adapted physical activity resources. The study also did not track long-term classroom implementation or student outcomes, and usability testing was conducted via Zoom rather than in real-world PE settings, which may limit how well the findings reflect actual use.

Conclusions

This study demonstrates that a web-based platform for delivering school-based PE inclusion solutions is both usable and acceptable to PE teachers and parents, as evidenced by high satisfaction scores and successful task completion across user groups. The integration of quantitative and qualitative user-testing data provided insights into the platform's strengths, such as its professional value and resource accessibility, as well as areas for improvement, including navigation, terminology, and support for nonexpert users.

These findings reflect users' perceptions of the platform during initial usability testing and should not be interpreted as evidence

XSL•FO
RenderX

that platform use improves teachers' knowledge, confidence, or inclusive teaching behaviors, or that it leads to improved student participation or educational outcomes. Future research is needed to examine whether sustained use of the platform produces these downstream benefits.

Acknowledgments

The authors would like to acknowledge Michelle Grenier for her contribution withto the physical education solution development. The authors would also like to acknowledge Tejaswini Chilke and Rohit Doddapani for developing the system.

As part of our commitment to transparency and ethical publishing practices, the authors disclose that generative AI were used during the preparation of this manuscript in a limited and clearly defined manner. Specifically, Claude was used to support language improvement and stylistic refinement of the manuscript text, without altering the scientific content, data interpretation, or conclusions. In addition, Claude was used to generate the procedural diagram for our mixed method. All scientific analyses, results, interpretations, and conclusions presented in this manuscript were conceived, verified, and approved entirely by the authors. The authors take full responsibility for the integrity, accuracy, and originality of the work and confirm that the use of AI tools did not replace critical scholarly judgment, data analysis, or authorship responsibilities.

Funding

This work was supported under the RERC RecTech funded by the National Institute on Disability, Independent Living, and Rehabilitation Research (grant number 90REGE0019).

Data Availability

Data will be available in deidentified format and can be obtained upon request to the authors.

Authors' Contributions

Conceptualization: SM, LAM, LL, MT
Formal analysis: SM, SA
Methodology: MT
Resources: SM, LAM, LL
Software: MT
Supervision: SM, MT
Writing – original draft: SM, LAM, SA, LL, MT
Writing – review & editing: SM, LAM, LL, SA, MT

Conflicts of Interest

None declared.

Multimedia Appendix 1
Script used to guide the participant usability testing sessions, including the introduction, background questions, assigned tasks, and posttest debriefing questions.
[DOCX File, 19 KB - humanfactors_v13i1e89476_app1.docx]

Multimedia Appendix 2
Summary of key system changes implemented in response to participant feedback and usability errors, along with subsequent testing outcomes following each refinement.
[DOCX File, 15 KB - humanfactors_v13i1e89476_app2.docx]

Multimedia Appendix 3
Additional results showing raw task performance data and completion-time descriptive statistics for physical education teachers and parents.
[DOCX File, 23 KB - humanfactors_v13i1e89476_app3.docx]

Multimedia Appendix 4
Screenshots of the platform surveys administered during user testing.
[DOCX File, 882 KB - humanfactors_v13i1e89476_app4.docx]

XSL•FO
RenderX

Multimedia Appendix 5
Screenshots of the survey interface showing the survey section with visual separation between questions and the updated and reorganized survey results section with clickable links to solutions.
[DOCX File, 915 KB - humanfactors_v13i1e89476_app5.docx]

References

1. Bandini L, Danielson M, Esposito LE, et al. Obesity in children with developmental and/or physical disabilities. Disabil Health J 2015 Jul;8(3):309-316. [doi: 10.1016/j.dhjo.2015.04.005] [Medline: 26058685]
2. MacEachern S, Forkert ND, Lemay JF, Dewey D. Physical activity participation and barriers for children and adolescents with disabilities. Intl J Disabil Dev Educ 2022 Jan 2;69(1):204-216. [doi: 10.1080/1034912X.2021.1952939]
3. Haegele JA, Wilson WJ, Zhu X, Bueche JJ, Brady E, Li C. Barriers and facilitators to inclusion in integrated physical education: adapted physical educators' perspectives. Eur Phy Educ Rev 2021 May;27(2):297-311. [doi: 10.1177/1356336X20944429]
4. Grenier M, Giese M. Ableism within adapted/physical education teacher education: implications for practice. In: Reflexivity and Change in Adaptive Physical Activity: Routledge; 2022:151-162. [doi: 10.4324/9781003196747-17]
5. Liu X, Han H, Li Z, et al. Barriers and facilitators to participation in physical activity for students with disabilities in an integrated school setting: a meta-synthesis of qualitative research evidence. Front Public Health 2025;13:1496631. [doi: 10.3389/fpubh.2025.1496631] [Medline: 40145000]
6. Holland K, Haegele JA, Zhu X, Bobzien J. "Everybody wants to be included": experiences with 'inclusive'strategies in physical education. J Dev Phys Disabil 2023;35(2):273-293. [doi: 10.1007/s10882-022-09852-x] [Medline: 35757833]
7. Molina Roldán S, Marauri J, Aubert A, Flecha R. How inclusive interactive learning environments benefit students without special needs. Front Psychol 2021;12:661427. [doi: 10.3389/fpsyg.2021.661427] [Medline: 33995221]
8. Oh-Young C, Filler J. A meta-analysis of the effects of placement on academic and social skill outcome measures of students with disabilities. Res Dev Disabil 2015 Dec;47:80-92. [doi: 10.1016/j.ridd.2015.08.014] [Medline: 26342328]
9. Aiyegbusi OL. Key methodological considerations for usability testing of electronic patient-reported outcome (ePRO) systems. Qual Life Res 2020 Feb;29(2):325-333. [doi: 10.1007/s11136-019-02329-z] [Medline: 31691202]
10. Inclusive physical education and physical activity: meeting the needs of students with disabilities—survey report. : Active Schools and Lakeshore Foundation–National Center on Health, Physical Activity and Disability (NCHPAD); 2016 URL: https://www.lakeshore.org/wp-content/uploads/2017/08/NCHPAD_Active-Schools_Survey-Report_Final.pdf [accessed 2026-09-12]
11. Grenier M, Lieberman LJ, Beach P. Training needs of educators for students with visual impairments and additional disabilities: a qualitative inquiry. Br J Vis Impair 2025 Jan;43(1):143-155. [doi: 10.1177/02646196231212740]
12. van Munster MA, Lieberman LJ, Grenier MA. Universal design for learning and differentiated instruction in physical education. Adapt Phys Activ Q 2019 Jun 27;36(3):359-377. [doi: 10.1123/apaq.2018-0145] [Medline: 31155914]
13. Polat E, Cepdibi S, Ci˙ri˙t NC, Hopcan S, Baştuğ YE. Educational technology in inclusive classrooms: assessing teacher awareness and needs. J Educ Technol Online Learn 2024;7(1):116-131. [doi: 10.31681/jetol.1355239]
14. Tarantino G, Makopoulou K, Neville RD. Inclusion of children with special educational needs and disabilities in physical education: a systematic review and meta-analysis of teachers' attitudes. Educ Res Rev 2022 Jun;36:100456. [doi: 10.1016/j.edurev.2022.100456]
15. Wilhelmsen T, Sørensen M. Inclusion of children with disabilities in physical education: a systematic review of literature from 2009 to 2015. Adapt Phys Activ Q 2017 Jul;34(3):311-337. [doi: 10.1123/apaq.2016-0017] [Medline: 28727510]
16. Lieberman LJ, Grenier M, Brian A, Arndt K. Universal Design for Learning in Physical Education: Human Kinetics; 2020.
17. Lieberman L, Brian A, Grenier M. The Lieberman–Brian Inclusion Rating Scale for Physical Education. Eur Phy Educ Rev 2019 May;25(2):341-354. [doi: 10.1177/1356336X17733595]
18. Lieberman LJ, Bean L, Grenier M. Developing a Universal Design for Learning Pedagogy: perspectives of students with and without disabilities. J Teach Phys Educ 2024;44(3):444-453. [doi: 10.1123/jtpe.2024-0015]
19. An J, Hodge SR. Exploring the meaning of parental involvement in physical education for students with developmental disabilities. Adapt Phys Activ Q 2013 Apr;30(2):147-163. [doi: 10.1123/apaq.30.2.147] [Medline: 23520244]
20. Frøkjær E, Hertzum M, Hornbæk K. Measuring usability: are effectiveness, efficiency, and satisfaction really correlated. Presented at: Proceedings of the SIGCHI conference on Human Factors in Computing Systems; Apr 1-6, 2000. [doi: 10.1145/332040.332455]
21. Alwashmi MF, Hawboldt J, Davis E, Fetters MD. The iterative convergent design for mobile health usability testing: mixed methods approach. JMIR Mhealth Uhealth 2019 Apr 26;7(4):e11656. [doi: 10.2196/11656] [Medline: 31025951]
22. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
23. Cooke L. Assessing concurrent think-aloud protocol as a usability test method: a technical communication approach. IEEE Trans Profess Commun 2010;53(3):202-215. [doi: 10.1109/TPC.2010.2052859]
24. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]

XSL•FO
RenderX

25. Harper BD, Norman KL. Improving user satisfaction: the questionnaire for user interaction satisfaction version 5.5. Presented at: Proceedings of the 1st annual mid-Atlantic human factors conference; Feb 25-26, 1993 URL: https://www.researchgate.net/profile/Kent-Norman/publication/237124434_Improving_user_satisfaction_The_questionnaire_for_user_interaction_satisfaction_version_55/links/575aadce08aed884620d8bf0/Improving-user-satisfaction-The-questionnaire-for-user-interaction-satisfaction-version-55.pdf [accessed 2026-09-12]
26. Schnall R, Cho H, Liu J. Health Information Technology Usability Evaluation Scale (Health-ITUES) for usability assessment of mobile health technology: validation study. JMIR Mhealth Uhealth 2018 Jan 5;6(1):e4. [doi: 10.2196/mhealth.8851] [Medline: 29305343]
27. Braun V, Clarke V. Thematic analysis. In: APA Handbook of Research Methods in Psychology, Vol 2 Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological: American Psychological Association; 2012:57-71. [doi: 10.1037/13620-004]
28. Boyatzis RE. Transforming Qualitative Information: Thematic Analysis and Code Development: Sage Publications, Inc; 1998. URL: https://psycnet.apa.org/record/1998-08155-000 [accessed 2026-09-12]
29. Kalyuga S. The expertise reversal effect. In: Managing Cognitive Load in Adaptive Multimedia Learning: IGI Global Scientific Publishing; 2009:58-80. [doi: 10.4018/978-1-60566-048-6.ch003]
30. Ahmadi A, Noetel M, Parker P, et al. A classification system for teachers' motivational behaviors recommended in self-determination theory interventions. J Educ Psychol 2023;115(8):1158-1176. [doi: 10.1037/edu0000783]

Abbreviations

DOB: date of birth
HITUES: Health Information Technology Usability Evaluation Scale
ISO: International Organization for Standardization
LIRSPE: Lieberman/Brian Inclusion Rating Scale for Physical Education
NCHPAD: National Center on Health, Physical Activity & Disability
PE: physical education
QUIS: Questionnaire for User Interaction Satisfaction
SUS: System Usability Scale
UDL: Universal Design for Learning

Edited by S Law; submitted 12.Dec.2025; peer-reviewed by H Tilga, KD Cocker; revised version received 07.Sep.2026; accepted 08.Sep.2026; published 07.Oct.2026.

Please cite as:
Mohanraj S, Malone LA, Arogbokun S, Lieberman L, Thirumalai M
Development and User Testing of an Online Platform to Deliver School-Based Inclusive Physical Education Solutions: Mixed Methods Study
JMIR Hum Factors 2026;13:e89476
URL: https://humanfactors.jmir.org/2026/1/e89476
doi:10.2196/89476

© Sangeetha Mohanraj, Laurie A Malone, Sunkanmi Arogbokun, Lauren Lieberman, Mohanraj Thirumalai. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 7.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Designing Coloring-Based Digital Art Therapy to Treat Alexithymia in Chinese College Students: Qualitative Study

Shuzhan Liu[1], MA; Haoyong Deng[1], MA; Jing Chen[2], PhD; Qi Wang[1], PhD

[1]College of Design and Innovation, Tongji University, Shanghai, China

[2]School of Nursing and Health Management, Shanghai University of Medicine and Health Sciences, No. 279 Zhouzhu Road, Pudong New Area, Shanghai, China

Corresponding Author:
Jing Chen, PhD
School of Nursing and Health Management, Shanghai University of Medicine and Health Sciences, No. 279 Zhouzhu Road, Pudong New Area, Shanghai, China

Abstract

Background: Alexithymia is a cognitive-emotional condition characterized by difficulties in identifying and expressing emotions. It lacks recognition among Chinese college students due to cultural norms that emphasize emotional restraint. Traditional interventions often rely on verbal expression, which poses challenges for students with alexithymia. Coloring-based digital art therapy provides a nonverbal, low-cognitive-load alternative that facilitates emotional expression through visual and interactive methods.

Objective: This study aimed to explore expert opinions on developing a coloring-based digital art therapy intervention for alexithymia among Chinese college students, develop an intervention framework, and create a digital game prototype tailored to this population.

Methods: This study was conducted in 2 phases. First, semistructured expert interviews were conducted with 10 experts to gather insights and create the intervention framework for coloring-based digital art therapy. In the second phase, a co-design process with the same experts refined the framework into a game prototype, with their feedback on design recommendations and game elements guiding the iterative development of *Coloring the Emoji*.

Results: Experts identified four core design themes: (1) visual over verbal expression, (2) creating a natural and everyday experience, (3) supporting implicit expression, and (4) ensuring psychological safety and data protection. These principles were incorporated into *Coloring the Emoji*, which guides students through color selection, emoji-based pattern selection, and free drawing to promote emotional expression. Key features of the game include the use of familiar emojis, customizable options for emotional expression, and robust privacy protections. The game is freely accessible and is well-suited for Chinese college students dealing with academic pressures and financial constraints.

Conclusions: This study presents a novel intervention framework for coloring-based digital art therapy tailored to Chinese college students with alexithymia. The game *Coloring the Emoji* provides a nonverbal, intuitive platform for emotional regulation and self-reflection, addressing the cultural and emotional challenges specific to this population. Future research should focus on user testing and explore its applicability to other cultural and age groups.

(JMIR Hum Factors 2026;13:e82128) doi:10.2196/82128

KEYWORDS

alexithymia; digital art therapy; coloring game; intervention framework; emotional regulation; ER; semistructured interview; co-design

Introduction

Background

Alexithymia is a cognitive-emotional condition [1] that affects about 10% of the general population and more than 20% of college students [2]. It results from disrupted brain circuitry in emotional processing and is characterized by deficits such as difficulty distinguishing between emotions and bodily sensations, a lack of emotional vocabulary, and reduced symbolic imagination [3]. Our previous survey of Chinese college students (n=332) revealed that 12.05% of students exhibited significant alexithymia (20-item Toronto Alexithymia Scale ≥70), with an overall average score of 59.90 (SD 8.15), which is close to the clinical threshold, suggesting a significant prevalence of alexithymia in this population.

Existing interventions primarily rely on psychotherapy approaches such as cognitive behavioral therapy (CBT) and mindfulness-based stress reduction, which aim to alleviate

XSL•FO
RenderX

symptoms by enhancing emotional awareness and expression [4]. These interventions typically require college students to verbally describe their emotions and engage in self-reflective emotional processing, which can be particularly challenging for college students with alexithymia due to their limited ability to identify emotions and express them verbally [3]. In China, where emotional restraint is culturally valued [5], reluctance to express emotions is often seen as normal social behavior rather than a clinical symptom [5], resulting in a lack of recognition of alexithymia and insufficient focus on tailored therapeutic approaches. Chinese college students face additional emotional challenges due to cultural expectations that prioritize emotional control and social conformity, making emotional expression more difficult [6]. Furthermore, practical barriers such as heavy academic workloads, geographic limitations, and financial constraints further hinder access to therapy, complicating the identification and treatment of alexithymia in this population [7,8].

Art Therapy and Its Limitations

Art therapy offers a structured and theoretically grounded approach to addressing alexithymia [9]. By involving nonverbal creative media, such as color, imagery, and movement, art therapy facilitates emotional identification and expression in a psychologically safe and cognitively manageable way [1]. This process can enhance emotional awareness, reduce avoidance, and externalize implicit emotions [10], thereby supporting long-term self-understanding and emotional regulation. By facilitating the externalization of suppressed emotions and overcoming verbal barriers, art therapy alleviates symptoms of alexithymia [11], reduces anxiety and depression [12], and enhances overall well-being [13].

However, several significant limitations continue to constrain their broader applicability. First, attitudinal barriers, such as shame and embarrassment, may hinder full engagement [14], especially among Chinese college students, for whom emotional restraint is culturally valued, and open emotional expression is often seen as a weakness [6]. Second, a lack of initiative, often driven by emotional avoidance and suppression, impedes their ability to regulate and express emotions effectively [15]. Chinese students are often taught to suppress emotions in favor of academic and social expectations, making emotional engagement more challenging. Third, time constraints, spatial limitations, and financial barriers reduce the feasibility and accessibility of interventions [16]. Chinese college students face heavy academic workloads, which limit their time for therapy. Additionally, geographic barriers and limited financial resources make it difficult for students to consistently participate in traditional therapy.

To address these challenges, digital art therapy has emerged as a promising intervention that leverages technology to enhance emotional expression and psychological healing [17]. Compared to traditional methods, it offers greater accessibility, continuity of care, adaptability for college students, and expanded modes of self-expression.

Coloring-Based Digital Art Therapy

Digital art therapy encompasses various approaches that combine creative expression with therapeutic intent. Among these, virtual reality–based art therapy offers an immersive, gravity-free environment that encourages full-body engagement and provides users with a sense of control. However, it requires costly equipment and may cause sensory discomfort, limiting its practical applicability, particularly for college students with limited finances [18]. Augmented reality–enhanced therapy, which overlays digital images onto real-world settings, is beneficial for trauma exposure and social interaction but poses challenges related to technical requirements and privacy concerns, which can hinder its widespread use [19].

Digital drawing platforms, such as Procreate or artificial intelligence–assisted tools, provide a low-pressure environment for precise, reversible emotional expression. However, they lack therapeutic framing and professional guidance, which can limit their effectiveness in addressing emotional regulation needs [20]. Therapeutic mobile apps offer a more structured approach, typically integrating mood tracking, daily prompts, and simple creative tasks. These apps are particularly accessible and well-suited for daily use, especially for college students dealing with emotional regulation challenges. However, they generally lack therapist involvement and offer limited depth of engagement, which may reduce their therapeutic effects [21]. Remote art therapy, delivered through live or asynchronous interaction, provides professional guidance and contextual sensitivity but may be hindered by technical disruptions and reduced nonverbal cues [22].

In this context, coloring-based digital art therapy builds on the accessibility and user-friendliness of digital platforms while offering unique advantages [23]. Unlike general digital drawing platforms, it integrates therapeutic guidance into its design by focusing on nonverbal emotional expression through color selection, pattern selection, and simple painting tasks [24]. This approach not only reduces cognitive load but also provides structured emotional support via predesigned templates, emotion tracking, reflective journals, and optional asynchronous therapist feedback [23]. By blending creative expression with therapeutic intent, coloring-based digital art therapy provides a safe, nonjudgmental space for users, making it particularly suitable for Chinese college students with alexithymia [25]. It helps them cope with the challenges posed by cultural emotional restraint and enables emotional exploration in a supportive environment [26].

Serious Games in Digital Art Therapy

Serious games are widely used in therapeutic contexts, designed to achieve specific goals such as emotional regulation, behavior change, or psychological healing, while also engaging users with interactive and enjoyable gameplay [27]. These games are different from traditional entertainment games, as they balance 2 key elements: achieving therapeutic outcomes and maintaining user engagement through fun and immersive experiences [28]. This combination makes serious games effective in offering users an interactive platform for emotional expression and self-reflection, which is especially beneficial for individuals with alexithymia, who may find verbal expression challenging

XSL•FO
RenderX

[29]. By incorporating structured tasks and feedback, serious games enable users to process and express emotions in a nonthreatening, low-pressure environment [27].

Among various serious games, coloring-based games stand out because of their simplicity and accessibility. These games create a low-cognitive-load environment for nonverbal emotional expression, allowing users to engage in activities such as color selection, pattern selection, and painting [24]. These tasks are specifically designed to help users externalize their emotions without the need for verbal expression, making them particularly useful for individuals with alexithymia [30]. In the case of Chinese college students, these games are especially beneficial as they provide a safe, nonjudgmental space for emotional exploration, which is crucial in a culture where emotional restraint is valued and emotional expression may be suppressed [31]. The game's structure, which involves tasks such as selecting colors, choosing patterns, and free drawing, offers a broad scope for users to express emotions creatively [32,33]. Additionally, the game settings—including visual style, real-time feedback, customizable options, and privacy measures—ensure that the intervention is engaging, accessible, and culturally sensitive [34]. These elements create a therapeutic yet enjoyable experience that respects cultural norms and encourages students to express themselves in a way that feels both safe and meaningful [35]. By integrating creative expression with therapeutic support, coloring games offer a comprehensive, user-friendly intervention that promotes emotional regulation, self-reflection, and personal growth [36].

Objectives

This study examines the use of coloring-based digital art therapy as an intervention for Chinese college students with alexithymia, focusing on how low-cognitive-load digital coloring tasks can support emotional identification and expression. The research began with expert interviews from fields including art therapy, human-centered design, digital health, and software development. These interviews provided valuable insights, which were used to create an intervention framework specifically tailored to the emotional and expressive needs of Chinese college students with alexithymia. Following the development of the framework, we engaged in a co-design phase with the same experts to refine and adapt the framework into a practical prototype. This iterative process ensured that the final game design, *Coloring the Emoji*, met both design recommendations from the previous framework and the specific needs of Chinese college students. The game, which incorporates color theory, pattern design principles, and human-computer interaction concepts, aims to facilitate nonverbal emotional exploration through intuitive color and pattern interactions. Future work will involve user testing to assess its therapeutic impact and further refine the system for broader application.

In summary, our contributions can be summarized as follows:

- An intervention framework for coloring-based digital art therapy was proposed to help Chinese college students with alexithymia better identify and describe their vague emotions, as well as effectively express and process them.
- Based on the proposed intervention framework, the authors collaborated with experts during a co-design phase to develop and refine *Coloring the Emoji*, a digital coloring game specifically designed to address the emotional needs of Chinese college students with alexithymia.

Methods

Overview

This study was conducted in 2 phases to construct the framework of coloring-based digital art therapy and to develop a coloring game for Chinese college students with alexithymia. In the first phase, semistructured expert interviews were conducted to gather professional insights and create an intervention framework for coloring-based digital art therapy. This qualitative approach helped identify key design principles and therapeutic mechanisms for the coloring game. In the second phase, we collaborated with the same experts in a co-design process to refine the framework and transform it into a game prototype. Through a series of collaborative discussions, the experts provided feedback on design recommendations and key game elements, allowing for the iterative development of *Coloring the Emoji*.

Recruitment

Purposive sampling was conducted until thematic saturation was achieved. Participants were selected through the authors' professional networks, targeting both professional experts and graduate students with relevant expertise in art therapy, digital health, human-centered design, or mobile app development. Professional experts were recruited for their extensive experience, with a minimum of 5 years in the field, as their expertise offers deeper insights and a broader understanding of the subject matter. Graduate students were chosen based on having over 5 years of professional experience in their respective fields and having participated in multiple practical projects relevant to the research topic. These graduate students were selected for their ability to provide insights closely reflecting real-world user experiences, given their engagement in daily life environments. To ensure diversity, experts were recruited from various regions of China. In addition to experts from the researchers' base in Shanghai, representing the eastern region, experts were also sought from the northern and southwestern regions to capture the regional and cultural differences across the country.

As data collection progressed, the research team continuously analyzed the interviews to assess thematic redundancy and determine when saturation had been reached. Based on this criterion, a total of 10 participants (4 graduate students and 6 professional experts) were included in the study. The inclusion criteria were as follows: (1) a master's degree or higher; (2) a minimum of 5 years of experience in art therapy, digital health, human-centered design, or mobile app development; and (3) demonstrated enthusiasm and willingness to participate. Exclusion criteria included experts who reported limited familiarity with the relevant fields or who provided low-quality responses during the initial consultation.

Procedure

We arranged semistructured interviews with each expert between April 2024 and June 2024, coordinating the time and location

XSL•FO
RenderX

in advance via email, phone, or text message. Due to geographical constraints, 2 experts were interviewed remotely via Zoom (Zoom Communications, Inc), while the remaining 8 experts were interviewed face-to-face by 2 members of the research team in a private setting. A semistructured interview guide with open-ended questions was prepared to allow each expert to share their perspectives and recommendations regarding coloring-based digital art therapy. The interview guide (Multimedia Appendix 1) was developed with 17 questions and addressed the following themes: (1) experiences in art therapy, (2) attitude toward digital art therapy, (3) suggestions for coloring-based digital art therapy, and (4) willingness of the expert to facilitate and support the follow-up study. Each interview lasted between 28 and 90 minutes. All interviews were audio- and video-recorded and verbatim transcribed within 3 days of each interview.

After the interviews, from June 2024 to July 2024, we proceeded to the second phase, engaging the same experts in multiple co-design sessions. These sessions were held remotely via Zoom and the Figma platform to refine the intervention framework and translate it into a game prototype. Experts provided feedback on the categorization of design recommendations and key game elements, ensuring the final design aligned with previous insights and therapeutic goals.

Data Analysis

Data collected from the semistructured expert interviews were analyzed using NVivo 15 (Lumivero) and an inductive approach by 2 independent researchers who were part of the research team, identifying themes related to experts' views on using digital art therapy for emotion regulation and their opinions on designing color-based interventions for Chinese college students with alexithymia. Each researcher independently reviewed the transcripts, coded the insights, and organized these codes into themes. The researchers then discussed their findings to reach a consensus on the key themes, ensuring agreement on the categorization of the data.

Ethical Considerations

The study was conducted following ethical standards and was approved by the institutional review board of the main researchers' affiliated institution (approval: TJDXSR076). Before participation, experts were informed of the study's purpose, procedures, and potential risks. Verbal informed consent was obtained when scheduling the interview via phone or email. All participants took part in this research voluntarily without any financial compensation. All participants were assured of their confidentiality, and data were anonymized to prevent identification. Personal information was securely stored and anonymized throughout the study, ensuring compliance with data protection regulations.

Results

Expert Characteristics

A total of 10 participants were involved in the study, comprising 6 professional experts and 4 graduate students with relevant expertise. The gender distribution was balanced, with 5 males and 5 females. The graduate students, aged between 24 and 27 years, included 2 students aged 24 years and 2 students aged between 25 and 27 years. All were master's degree students who had studied in their respective fields for 5 to 8 years. Their areas of expertise included art therapy (n=1), digital health (n=1), human-centered design (n=1), and mobile app development (n=1), with the students in art therapy and digital health each having more than 1 year of work experience. The experts ranged in age from 28 to 44 years, with 2 experts aged under 34 years and 4 experts aged between 34 and 44 years. All 6 experts held doctoral degrees, including 4 associate professors and 2 lecturers. The experts' years of industry experience varied: 2 experts had nearly 10 years, 3 had nearly 20 years, and 1 had more than 20 years. Their areas of expertise included art therapy (n=1), digital health (n=2), human-centered design (n=2), and mobile app development (n=1). An overview of the expert characteristics is provided in Table 1.

XSL•FO
RenderX

Table . Basic information about consulting experts (N=10).

Characteristics and classification	Participants, n
Age (y)	
24	2
34	4
44	4
Professional title	
Associate professor	4
Lecturer	2
N/A[a]	4
Education background	
Doctoral degree	6
Master's degree	4
Years of study	
5	4
Years of work	
10	2
15	1
20	3
Specialist areas	
Art therapy	2
Digital health	3
Human-centered design	3
Mobile app development	2

[a]Not available.

Intervention Framework for Coloring-Based Digital Art Therapy

The thematic analysis of expert interviews provides valuable insights into the construct of coloring-based digital art therapy for addressing alexithymia among Chinese college students.

Theme 1: Visual Instead of Verbal

Subtheme: Draw the Emotions

Most experts (7/10) believed that art therapy plays a significant role in the treatment of emotional expression disorders. They pointed out that art therapy not only helps visualize abstract emotions but also eliminates language barriers through visual expression, providing college students with a low-pressure, nonthreatening way to externalize their emotions. A smaller group (3/10) mentioned more detailed points. They expressed that color makes the process more tangible and perceptible, promoting emotional regulation and self-awareness, and helping students more intuitively recognize and process their feelings. One expert mentioned that one of her college students used warm yellow to express maternal love and the warmth of ginger soup, "As I observed the boy's coloring, he colored the 'umbrella in the rain' warm yellow, saying, 'This is the steam from my mom's ginger soup'" [E1].

Another expert also pointed out similar cases involving a student who used an orange boat to symbolize secret affection, "In one of my cases, … the student painted gray waves for 3 days, then suddenly added an orange boat, saying, 'That's the color of the headband of the girl I secretly like'" [E7].

Subtheme: Express Through Symbolism

Some experts (4/10) suggested that symbolic emotional expression is not limited to color alone, but also includes shapes, sounds, and other sensory elements, which can help college students more deeply recognize and communicate their emotions. Integrating various sensory symbols, such as visual and auditory cues, can make emotional expression more concrete and diversified. By adapting to different college students' preferences, it breaks the limitations of traditional verbal expression. One expert stated, "When the student exhibited anger, the system responded with fast music and flames appeared around the character" [E6].

Subtheme: Describe Emotional Stories in the Third Person

Experts who have had long-term communication with college students (5/10) have found that by attributing emotions to virtual characters or a third party, college students can avoid directly confronting their own emotions, which makes it easier for them to express and understand their feelings. This method provides

XSL•FO
RenderX

an indirect way to externalize emotions, making emotional expression more natural and less stressful:

> *Asking "What does the person in the picture see?" works much better than asking "How do you feel?"* [E1]

> *Every time I asked about her feelings, the girl stubbornly said, "I feel nothing," but when her cloud avatar turned gray, the system said, "Little Cloud is raining," and she cried, "It's because I want to cry."* [E9]

Subtheme: Provide Interactive Options

Experts with hands-on therapeutic experience (3/10) noted that body language, especially tactile and auditory feedback, can make emotional expression more concrete and vivid. By activating the body's instinctive responses, college students can perceive and express emotions more intuitively. This method is beneficial for those who struggle with expressing emotions, as body language provides an effective alternative, such as adjusting brush size or the intensity of movements. This sensory experience makes emotional externalization more natural and enhances the emotional regulation process:

> *A little girl I once treated expressed, "When I'm happy, I use a small brush to paint carefully, but when I'm unhappy, I use a thick brush." So, it would be better to give them different choices.* [E1]

> *It would be beneficial to incorporate some interactive features, such as adding sound effects when the student clicks a button.* [E8]

Theme 2: Create a Natural and Everyday Experience

Subtheme: Refuse to Judge

Most experts (9/10) stressed the importance of avoiding the judgment of emotional expressions, as this could lead to students feeling evaluated or restricted in their emotional expression. When students are aware that their work will be graded or categorized according to predefined emotional labels, it introduces an element of judgment, which can hinder their ability to freely express their emotions:

> *Core rule: No grading! Do not label the patient's work. Grading and labeling will make students feel that they are being evaluated, which will affect their emotional expression.* [E1]

> *Students should have the freedom to give meaning to their own emotional expressions without having fixed labels imposed by the system or the outside world.* [E5]

Subtheme: Integrate Into Life Scenes

In addition, experts (7/10) believe that the design of digital art therapy should integrate emotional expression into daily life to make it closer to students' real experiences. By combining the stress and emotions of life with artistic expression, students can release their inner emotions in a familiar situation.

> *When students face exam anxiety, some test paper-style patterns can be added to the system to help students vent their emotions through artistic expression when facing pressure and reduce the burden of stress.*

> *[E7]*

Subtheme: Establish Event Connection

More than half of the experts (6/10) mentioned that through contextual connection, students can more naturally connect emotions with specific events, which helps them understand and express emotions more easily. The design of contextual connection can provide a guide to enable students to connect emotions with real-life events and situations, making it easier for them to identify and express emotions:

> *Seeing the sunset makes the child calm, and the system learns to associate sunsets with relaxation.* [E2]

> *After watching her favorite flowers, the system links happiness with similar content.* [E3]

Subtheme: Leverage Digital Technology

Experts (3/10) emphasized that the application of digital technology can further enhance the effectiveness of digital art therapy. By using modern digital tools, students can enjoy a more fluid, intuitive, and multisensory experience of emotional expression. One expert stated, "Quick responses and smooth experiences can reduce students' anxiety caused by technical issues and help them focus on emotional expression rather than technical distractions" (E10).

Theme 3: Support Implicit Expression

Subtheme: Anonymization

Experts (4/10) emphasized that anonymity is a key feature of digital art therapy, helping students feel safe and unjudged when expressing emotions. Anonymizing the process of emotional expression allows students to explore their feelings without fear of exposure or social judgment. This feature is especially beneficial for those who experience social anxiety or feel vulnerable when discussing their emotions openly. One expert stated, "Works should provide an anonymous upload option while retaining the students' right to refuse uploading" (E10).

Subtheme: Protect Personal Privacy

Experts with a technical background (3/10) pointed out that privacy protection is another crucial issue in digital arts therapy and highlighted several methods for protecting personal data and ensuring the confidentiality of students' emotional expressions. Protecting privacy is especially important when dealing with sensitive topics such as emotional distress, mental health issues, or trauma. Students need to be assured that their personal information and emotional content will not be exposed to others without their consent:

> *Implement strict controls on who can access user data and when, ensuring only authorized people or systems can access sensitive data.* [E3]

> *Using a WeChat mini-program and bypassing campus network restrictions ensures that students can continue their coloring even if the internet is down at night.* [E6]

XSL•FO
RenderX

Theme 4: Ensure Psychological Safety and Data Protection

Subtheme: Protect User Data

Experts (10/10) agreed that data protection is a fundamental component of any digital therapy platform. Ensuring that students' personal information and emotional data are securely handled and erased when no longer needed is essential for maintaining trust and preventing exploitation. One expert stated, "When deleting data, provide a destruction certificate to ensure complete disappearance" (E6).

Subtheme: Focus on Psychological Safety

In addition to data protection, experts (7/10) emphasized that the game's design should prioritize mental safety. Ensuring that students feel emotionally safe, supported, and free from stress during the emotional expression process is essential. The design should minimize factors that could lead to psychological harm or cause unnecessary anxiety.

> *Many college students I met expressed concerns that they could not draw, had no artistic talent, and were afraid that their paintings would not look good when they first encountered coloring art therapy. Psychological counselors and digital systems should always support and protect college students' fragile hearts, encourage college students to try, and avoid any content that would frustrate college students' psychology as much as possible.* [E1]

> *…Such as detecting repeated modifications triggers a pop-up: "Do you want to take a break?"* [E3]

Based on the qualitative analysis of expert interviews, the intervention framework for coloring-based digital art therapy is summarized in Table 2.

XSL•FO
RenderX

Table . Intervention framework for coloring-based digital art therapy.

Themes and subthemes	Design recommendation
Visual instead of verbal	
Draw the emotions	• Use color to externalize emotions by allowing students to select or create colors that represent different feelings or emotional states. • This can be applied through various elements in the game, such as backgrounds, objects, or avatars.
Express through symbolism	• Enhance emotional expression with not only color but also shapes, sounds, and sensory elements, offering a more diversified and personalized way to communicate emotions.
Describe emotional stories in the third person	• Attribute emotions to virtual characters or third parties. • Allow students to avoid confronting their own feelings directly, encouraging more natural, indirect emotional externalization.
Provide interactive options	• Allow students to adjust the size, intensity, or speed of brush strokes, gestures, or movements. • Implement vibration or sound effects that correlate with the movement or emotional expression, such as changes in the brush pressure or feedback sounds based on actions.
Create a natural and everyday experience	
Refuse to judge	• Avoid grading or labeling emotional expressions.
Integrate into life scenes	• Connect emotional expression to daily life experiences. • Simulate real-life situations through themed templates or prompts that mirror common emotional triggers.
Establish event connection	• Use contextual connections (like prompts, scenarios, or selectable environmental symbols) to link emotions with specific real-life events. • Design the system to associate emotions with familiar triggers.
Leverage digital technology	• Use digital technology to create a smooth, intuitive, and multisensory emotional expression experience. • Ensure quick responses and smooth interactions.
Support implicit expression	
Anonymization	• Allows students to express emotions freely without fear of judgment or exposure through anonymity. • Provide an anonymous upload option while respecting the students' right to refuse sharing their work.
Protect personal privacy	• Implement strict access controls to ensure that only authorized people or systems can access sensitive user data. • Provide offline functionality, such as bypassing network restrictions, to ensure students can continue their emotional expression work securely, even without internet access.
Ensure psychological safety and data protection	
Protect user data	• Provide a data destruction certificate when deleting user information to ensure complete and irreversible removal.
Focus on psychological safety	• Prioritize psychological safety by creating a supportive and stress-free environment. • Implement gentle reminders or breaks to avoid overwhelming students.

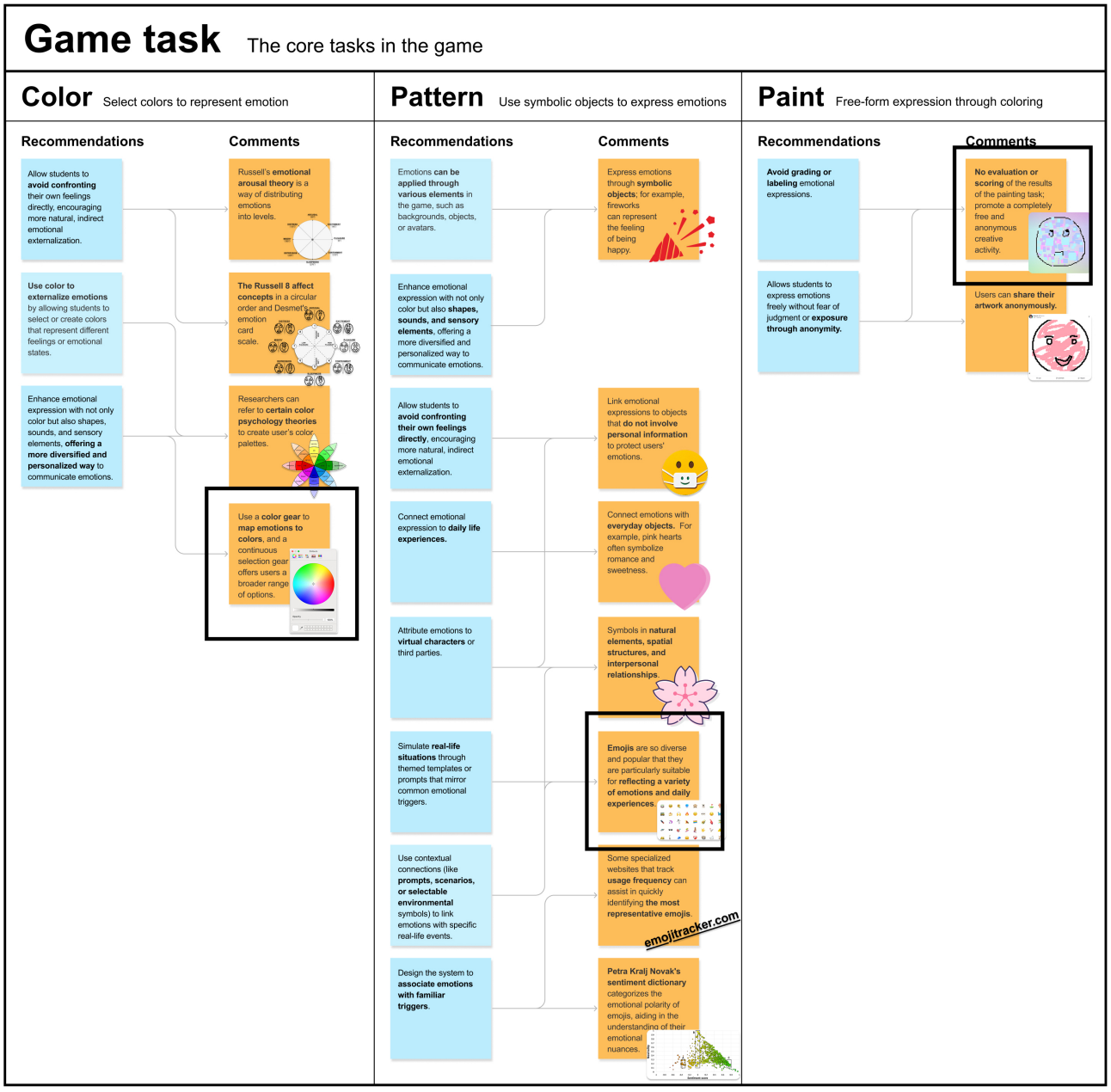
Coloring Game: *Coloring the Emoji*

Based on the insights gathered from the co-design process, *Coloring the Emoji* was developed to address the emotional and cognitive challenges faced by Chinese college students with alexithymia. During co-design sessions, experts provided feedback on how to organize and apply the design recommendations, which led to the refinement of key game

XSL•FO
RenderX

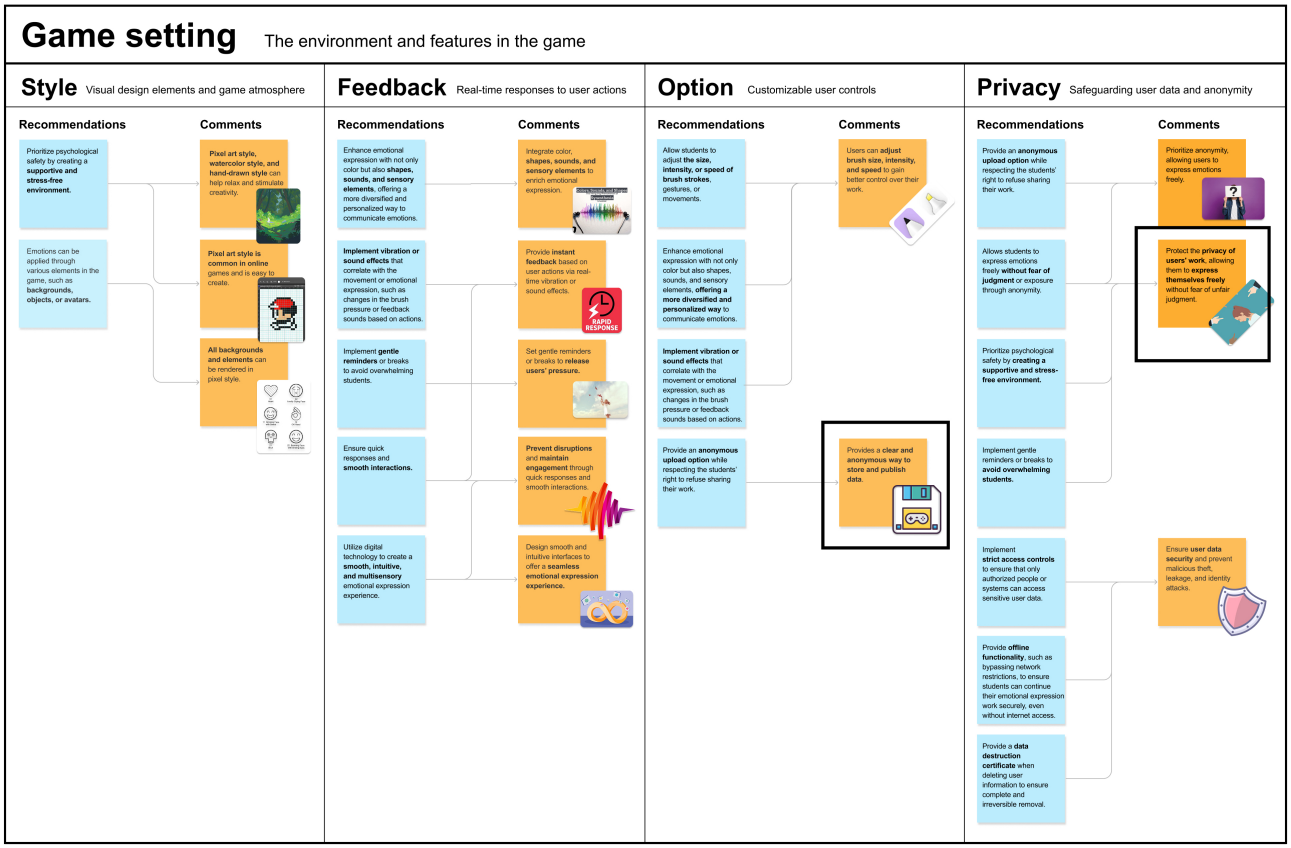
components, ensuring alignment with therapeutic goals (Figure 1 and Figure 2). The game was designed to facilitate intuitive, nonverbal emotional expression and recognition, incorporating expert insights into its core elements.

Figure 1. Game task.



XSL•FO
RenderX

Figure 2. Game setting.



The game task includes 3 key components: color, pattern, and paint [32,33]. For color, experts recommended using color to externalize emotions, allowing students to select colors that represent different feelings. Experts also suggested incorporating a continuous color selection gear to offer a broader range of options for more personalized emotional expression. This led to the creation of the *Emotional Color Gear*, which maps emotions to colors based on Russell's [37] emotional arousal theory and Desmet's [38] emotion card scale, simplifying emotions into 2 dimensions—pleasure and arousal—for easier emotional identification. In pattern, experts emphasized the use of symbolic objects and familiar symbols to help students externalize emotions indirectly. This led to the development of the *Emoji Scene Library*, which includes 53 emojis with strong emotional polarity and the highest use frequency among college students [39,40], allowing students to connect their emotions to familiar symbols and everyday experiences. The system was designed to associate emotions with common triggers, facilitating easier emotional identification [41]. For paint, experts highlighted the importance of creating a nonjudgmental space for emotional expression, with no grading or labeling of expressions, and features enabling students to share their artwork anonymously, promoting free, creative activities.

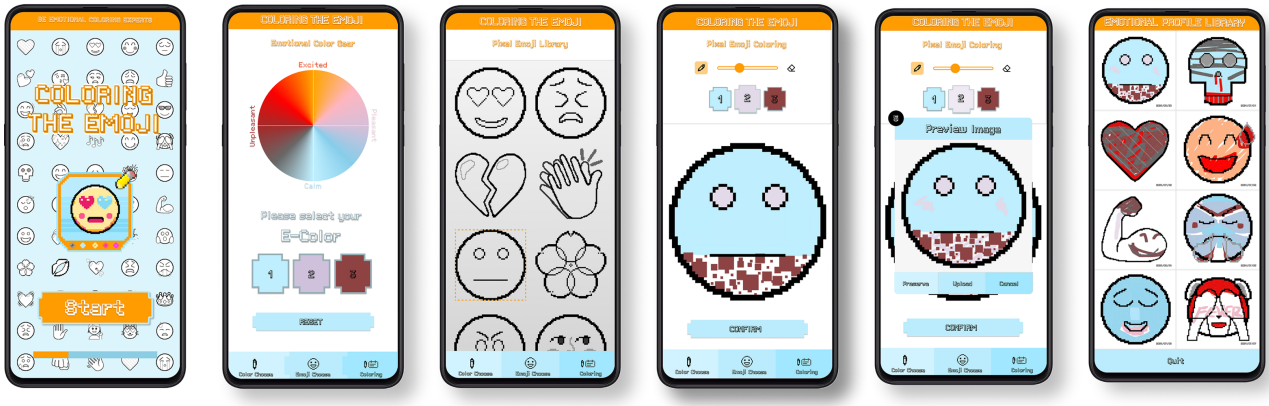
The game setting includes 4 key components: style, feedback, option, and privacy [34,35]. For style, experts recommended creating a supportive and low-pressure environment through visual elements that could help students relax and stimulate creativity. Based on this, we chose the pixel art style, which is common in online games, easy to create, and helps to create a relaxed atmosphere. In feedback, experts emphasized enhancing emotional expression not only with color but also with shapes, sounds, and sensory elements to provide instant feedback, while introducing gentle reminders and breaks to avoid overwhelming students and maintaining their engagement. For option, experts recommended allowing students to adjust brush size, intensity, and speed to better control their emotional expression, thereby enhancing engagement [42]. In privacy, experts stressed anonymity and data security, implementing strict access controls and offering the option to share artwork anonymously. Additionally, the game includes a data destruction certificate to ensure the irreversible deletion of personal data.

To ensure *Coloring the Emoji* meets the specific needs of Chinese college students with alexithymia, several design features were integrated based on expert feedback. These features address challenges such as emotional restraint, heavy academic workloads, and financial limitations. First, to accommodate emotional expression, the *Emotional Color Gear* offers a continuous selection option, allowing a broader range of color choices to represent emotions. Second, considering the popularity of emojis in Chinese digital communication, a diverse set of emojis was incorporated, reflecting a wide range of emotions and linking them to daily life experiences. Third, the game avoids grading or scoring emotional expressions, aligning with the cultural preference for privacy and reducing social pressure. Finally, robust privacy protection measures ensure students can express themselves freely without fear of judgment, with the option to share artwork anonymously. These design choices address the unique emotional challenges faced by Chinese college students, offering a culturally sensitive space for emotional exploration and self-expression.

XSL•FO
RenderX

Coloring the Emoji helps college students with alexithymia recognize and express their emotions through 3 core stages: color selection, pattern selection, and free drawing (Figure 3). In the color selection stage, students select colors to externalize and convey their feelings, thereby clearly identifying their emotions. In the pattern selection stage, students choose emojis that represent their emotions, allowing them to express their feelings indirectly and reduce emotional confrontation. Finally, the free drawing stage provides a safe, private space for unrestricted emotional expression, promoting emotional exploration and self-reflection without fear of judgment. These stages work together to offer students an intuitive, nonverbal way to engage with and understand their emotions (Multimedia Appendix 2).

Figure 3. The core interface of *Coloring the Emoji*.



Discussion

Principal Findings

This study aimed to develop a coloring-based digital art therapy intervention for treating alexithymia among Chinese college students. Through expert interviews, we gathered valuable insights that informed the creation of the intervention framework. The findings emphasized the effectiveness of using color, symbolism, and nonverbal emotional expression methods to address the challenges faced by college students with emotional processing difficulties. These insights align with existing theories highlighting the importance of visual over verbal expression in facilitating emotional regulation and expression for individuals with alexithymia [43]. Based on this framework, the *Coloring the Emoji* game was developed as a therapeutic intervention that uses coloring tasks to enhance emotional recognition and expression, offering a nonverbal, intuitive platform for self-reflection.

Intervention Framework

The intervention framework for coloring-based digital art therapy, developed through expert input, focuses on addressing the core features of alexithymia, specifically the difficulty in identifying and describing emotions. By integrating visual symbols such as colors and patterns, the framework provides a tangible method for Chinese college students to externalize their emotions, making it easier for them to communicate feelings they may struggle to verbalize [44,45]. Experts also emphasized the importance of creating a nonjudgmental and low-pressure environment for emotional expression. The design of the intervention explicitly avoids grading or labeling emotional outputs, ensuring that students can express themselves without the fear of judgment, which is particularly important for those with alexithymia who may already experience difficulties with emotional expression [46]. Furthermore, the integration of digital technology was deemed essential for creating a seamless, intuitive, and multisensory experience, allowing students to explore their emotions continuously and with reduced anxiety [47].

By combining expert insights, the intervention framework incorporates visual, symbolic, and contextual elements, making emotional exploration more relatable and personalized for students. The framework emphasizes incorporating body language and sensory feedback through adjustments to brush size and intensity, further enhancing the emotional expression process and providing a more immersive and realistic experience [42]. These principles were then directly applied in the development of *Coloring the Emoji*, creating a user-friendly, digital intervention that meets the therapeutic goals outlined in the framework.

Coloring Game

In the design of *Coloring the Emoji*, several key features were incorporated based on expert feedback to specifically address the emotional needs of Chinese college students with alexithymia. The color selection phase draws upon emotional psychology models, simplifying complex emotions into 2 dimensions—pleasure and arousal—and mapping these to specific colors [48]. This approach helps users quickly identify and select colors that best represent their emotional state, facilitating easier emotional expression. The pattern selection phase further supports emotional identification by offering a library of emojis, which were carefully selected for their strong emotional polarity [49]. Emojis are widely used in Chinese digital communication, providing students with a familiar and culturally relevant means to connect emotions with symbolic representations, thus making the emotional expression process more relatable.

The free drawing phase allows for uninhibited emotional expression, reinforcing the importance of a nonjudgmental space for self-reflection. The inclusion of interactive options, such as brush resizing and adjusting movement intensity, enhances user

XSL•FO
RenderX

engagement and allows students to express their emotions more freely [50]. These features also align with expert recommendations for integrating sensory feedback into the therapeutic process, providing a more interactive and immersive emotional exploration experience. Additionally, the game's structure supports nonverbal emotional expression, which is particularly beneficial for Chinese students, who may face cultural challenges in verbalizing their emotions.

The game also incorporates robust privacy protection measures to ensure the confidentiality of students' emotional data [51]. The option to share their artwork anonymously supports a safe, private space for emotional expression, reducing the social pressure that might otherwise inhibit participation. These privacy features, combined with the therapeutic design elements, ensure that students can engage with the game in a way that feels secure and supportive, while respecting their cultural values and emotional boundaries [52].

Limitations and Future Directions

While *Coloring the Emoji* shows promise in supporting emotional expression and regulation for Chinese college students with alexithymia, several limitations remain that need to be addressed in future research. This study focused specifically on Chinese college students, whose emotional challenges are influenced by cultural factors such as emotional restraint and social conformity. While the game design addresses these cultural nuances, its applicability to other populations remains uncertain. Future studies should explore how the game can be adapted to meet the needs of different cultural and age groups, as well as individuals with varying emotional processing difficulties. This would involve adjusting the game to accommodate diverse emotional challenges in different contexts, ensuring that the intervention remains relevant and effective across various demographics.

The current prototype requires rigorous user testing to assess its therapeutic impact and identify areas for improvement. Although expert feedback has been crucial in shaping the design, real-world testing is essential to evaluate how well the game supports emotional regulation and expression in practical settings [53,54]. Additionally, continuous collection of user data and qualitative feedback will provide insights into the game's effectiveness and adaptability to users' specific emotional needs.

Although the *Coloring the Emoji* game is beneficial for individuals with alexithymia, it may not fully replace other therapeutic methods, such as CBT or traditional art therapy [55,56]. Future research could investigate the potential integration of this coloring-based intervention with established therapeutic methods, creating a more comprehensive treatment plan. Combining digital art therapy with conventional therapies could offer a more holistic and adaptable model for treating emotional processing difficulties [57].

Conclusions

This study proposes a coloring-based digital art therapy framework developed to address alexithymia among Chinese college students. Based on this framework, we created *Coloring the Emoji*, a platform that provides an intuitive, nonverbal, and low cognitive-load method for emotional identification and expression. By integrating color, symbolism, and sensory feedback, the game offers a nonjudgmental space for anonymous self-expression, which directly addresses the cultural norms of emotional restraint and societal barriers faced by this population. The inclusion of emojis, which are widely popular in Chinese digital communication, further assists students in connecting emotions with familiar symbols. Additionally, the platform is free of charge and easily accessible, making it suitable for students facing academic pressures and financial constraints.

While the intervention shows promise, further research is needed to assess its real-world effectiveness and adaptability across diverse cultural and age groups. Future studies could explore the integration of *Coloring the Emoji* with other therapeutic approaches, such as CBT, to enhance its therapeutic impact. Further research should also examine its long-term effects, particularly in addressing co-occurring emotional issues, such as anxiety and stress.

Acknowledgments

The authors thank all the experts who participated in the interviews and co-design process for their valuable insights and contributions to this study.

Funding

The authors declared no financial support was received for this work.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

All authors contributed to the conceptualization and methodological design of the study, as well as the co-design process. SL and HD conducted the interviews, performed data analysis, wrote the first draft of the manuscript, and contributed to subsequent revisions. JC and QW provided guidance throughout the study and reviewed and edited the manuscript. All authors contributed to the final draft of the manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File, 17 KB - humanfactors_v13i1e82128_app1.docx]

Multimedia Appendix 2
Description of the user task of *Coloring the Emoji*.
[DOCX File, 57954 KB - humanfactors_v13i1e82128_app2.docx]

References

1. Miller R. The Art Therapy Process in the Context of the Psychosomatic Syndrome and Alexithymia [Master's thesis]. : Concordia University; 1989 URL: https://spectrum.library.concordia.ca/id/eprint/4434 [accessed 2026-05-11]
2. Rahimi M, Zarimeidani F, Bahreini HS, Rahmati R, Shahraki HR. Personality traits and alexithymia: a closer look at university students in a cross-sectional study. Medicine (Abingdon) 2025;104(10):e41765. [doi: 10.1097/MD.0000000000041765] [Medline: 40068073]
3. Tsubaki K, Shimizu E. Psychological treatments for alexithymia: a systematic review. Behav Sci (Basel) 2024 Dec 7;14(12):1173. [doi: 10.3390/bs14121173] [Medline: 39767314]
4. Yang E, Schamber E, Meyer RML, Gold JI. Happier Healers: randomized controlled trial of mobile mindfulness for stress management. J Altern Complement Med 2018 May;24(5):505-513. [doi: 10.1089/acm.2015.0301] [Medline: 29420050]
5. O'Loughlin JI, Cox DW, Kahn JH, Wu AD. Attachment avoidance, alexithymia, and gender: examining their associations with distress disclosure tendencies and event-specific disclosure. J Couns Psychol 2018 Jan;65(1):65-73. [doi: 10.1037/cou0000245] [Medline: 29355345]
6. Cui L, Tang G, Huang M. Expressive suppression, Confucian Zhong Yong thinking, and psychosocial adjustment among Chinese young adults. Asian J of Social Psycho 2022 Dec;25(4):715-730. [doi: 10.1111/ajsp.12529]
7. Sancar B, Aktas D. The relationship between levels of alexithymia and communication skills of nursing students. Pak J Med Sci 2019;35(2):489-494. [doi: 10.12669/pjms.35.2.604] [Medline: 31086538]
8. Bazzari AH, Bazzari FH. The correlation between alexithymia and communication skills among undergraduate pharmacy students. Heliyon 2025 Jan 15;11(1):e41402. [doi: 10.1016/j.heliyon.2024.e41402] [Medline: 39958744]
9. Lee SY, Lee HJ. The effects of group arts therapy based on emotion management training on the emotional expression, alexithymia, depression and quality of life in patients with schizophrenia. BJPsych Open 2025 Jun;11(S1):S55-S55. [doi: 10.1192/bjo.2025.10172]
10. Hong Y. "Inside Out": Using Expressive Arts Therapy to Externalize Implicit Feelings in a Chinese Origin Client with Psychosomatic Issues [Master's thesis]. : Lesley University; 2021 URL: https://digitalcommons.lesley.edu/expressive_theses/458 [accessed 2026-05-11]
11. Meijer-Degen F, Lansen J. Alexithymia—a challenge to art therapy: the story of Rita. Arts Psychother 2006;33(3):167-179. [doi: 10.1016/j.aip.2005.10.002]
12. Demi˙r V. The effect of art therapy and mindfulness based cognitive therapy program on the levels of alexithymia of individuals [Article in Turkish]. Asya Studies 2022;6(22):237-248. [doi: 10.31455/asya.1160023]
13. Cohen KL. How can Art Therapy be Utilized to Support Emotional Communication for Autistic Individuals with an Alexithymic Profile? [Master's thesis]. : Lesley University; 2025 URL: https://digitalcommons.lesley.edu/expressive_therapies_theses/64 [accessed 2026-05-11]
14. Farahani H, Watson P, Movasagh L, et al. Explainable multilayer perceptron classification model for detecting the internalized shame-based on childhood trauma, social-emotional competence, cognitive flexibility, distress tolerance, alexithymia, and age. In: Vanderheiden E, Mayer CH, editors. Shame and Ageing in a Transforming World: Springer; 2024:21-50. [doi: 10.1007/978-3-031-54585-6_2]
15. Li W, Han Y, Hu Z, Zhang Y. Childhood trauma and non-suicidal self-injury among Chinese adolescents: the chain mediating effects of alexithymia and experiential avoidance. Acta Psychol (Amst) 2024 Nov;251:104634. [doi: 10.1016/j.actpsy.2024.104634] [Medline: 39622148]
16. Schianchi I. AlexiLearn: facial expression recognition in a mobile educational tool for alexithymia and autism. Preprints. Preprint posted online on Apr 2, 2025. [doi: 10.20944/preprints202504.0165.v1]
17. Shojaei F, Osorio Torres J, Shih PC. Exploring the integration of technology in art therapy: insights from interviews with art therapists. Art Ther (Alex) 2025 Apr 3;42(2):112-118. [doi: 10.1080/07421656.2024.2383826]
18. Hacmun I, Regev D, Salomon R. The principles of art therapy in virtual reality. Front Psychol 2018;9:2082. [doi: 10.3389/fpsyg.2018.02082] [Medline: 30429813]
19. Esmael Motlagh N. The impact of AR art on social awareness: a case study. ART/icle 2024 Dec 19;4(2):196-212. [doi: 10.56590/stdarticle.1549288]

XSL•FO
RenderX

20. Zubala A, Kennell N, Hackett S. Art therapy in the digital world: an integrative review of current practice and future directions. Front Psychol 2021;12:595536. [doi: 10.3389/fpsyg.2021.600070] [Medline: 33897522]
21. Sahu RR, Raut A, Samantaray S. Technology-assisted mental healthcare: a novel approach. Presented at: Proceedings of the 10th International Conference on Frontiers in Intelligent Computing: Theory and Applications (FICTA 2022); Jun 18-19, 2022. [doi: 10.1007/978-981-19-7513-4_49]
22. Barber B, Brandoff R, Lombardi R, et al. Digital Art Therapy: Material, Methods, and Applications: Jessica Kingsley Publishers; 2016.
23. Putian G, Cho S, Jiani Z. Study of the art healing potential of coloring and drawing in video games for mental health. Int J Adv Cult Technol 2024:182-193. [doi: 10.17703/IJACT.2024.12.3.182]
24. Xi J, Gao Y, Lyu N, et al. Effect of the "Art Coloring" online coloring game on subjective well-being increase and anxiety reduction during the COVID-19 pandemic: development and evaluation. JMIR Serious Games 2022 Jul 8;10(3):e37026. [doi: 10.2196/37026] [Medline: 35575761]
25. Zhang MQ, Liu X, Huang Y. Does Mandala art improve psychological well-being in patients? A systematic review. J Integr Complement Med 2024 Jan;30(1):25-36. [doi: 10.1089/jicm.2022.0780] [Medline: 37668598]
26. Ashlock LE, Miller-Perrin C, Krumrei-Mancuso E. The effectiveness of structured coloring activities for anxiety reduction. Art Ther 2018 Oct 2;35(4):195-201. [doi: 10.1080/07421656.2018.1540823]
27. Dörner R, Göbel S, Effelsberg W, Wiemeyer J. Serious Games: Foundations, Concepts and Practice: Springer International Publishing; 2016. [doi: 10.1007/978-3-319-40612-1]
28. Caserman P, Hoffmann K, Müller P, et al. Quality criteria for serious games: serious part, game part, and balance. JMIR Serious Games 2020 Jul 24;8(3):e19037. [doi: 10.2196/19037] [Medline: 32706669]
29. Göbel S, Gutjahr M, Steinmetz R. What makes a good serious game—conceptual approach towards a metadata format for the description and evaluation of serious games. Presented at: 5th European Conference on Games Based Learning; Oct 20-21, 2011 URL: https://www.kom.tu-darmstadt.de/en/publications/GGS11-1 [accessed 2026-05-11]
30. Monaco F, Vignapiano A, Landi S, et al. Reducing state and trait anxiety through art therapy in adolescents with eating disorders: results from a pilot repeated-measures study. J Clin Med 2025 Jul 27;14(15):5298. [doi: 10.3390/jcm14155298] [Medline: 40806920]
31. Tsai W. Cultural differences in emotion expression and suppression: implications for health and well-being [Dissertation]. : University of California, Los Angeles; 2016 URL: https://escholarship.org/uc/item/8h96r3bw [accessed 2026-05-20]
32. Koo M, Chen HP, Yeh YC. Coloring activities for anxiety reduction and mood improvement in Taiwanese community-dwelling older adults: a randomized controlled study. Evid Based Complement Alternat Med 2020;2020:6964737. [doi: 10.1155/2020/6964737] [Medline: 32063986]
33. Plass JL, Homer BD, MacNamara A, et al. Emotional design for digital games for learning: the effect of expression, color, shape, and dimensionality on the affective quality of game characters. Learn Instr 2020 Dec;70:101194. [doi: 10.1016/j.learninstruc.2019.01.005]
34. Lukka L, Palva JM. The development of game-based digital mental health interventions: bridging the paradigms of health care and entertainment. JMIR Serious Games 2023 Sep 4;11:e42173. [doi: 10.2196/42173] [Medline: 37665624]
35. Mandryk RL, Birk MV. Toward game-based digital mental health interventions: player habits and preferences. J Med Internet Res 2017 Apr 20;19(4):e128. [doi: 10.2196/jmir.6906] [Medline: 28428161]
36. George A. Investigating the psychological benefits of adult colouring using a mixed method approach [Honors thesis]. : Cape Breton University; 2023 URL: https://cbu.scholaris.ca/items/75310175-e570-4337-890e-3342a5bd41b0 [accessed 2026-05-11]
37. Russell JA. A circumplex model of affect. J Pers Soc Psychol 1980;39(6):1161-1178. [doi: 10.1037/h0077714] [Medline: 7431205]
38. Desmet P. Measuring emotion: development and application of an instrument to measure emotional responses to products. In: Blythe M, Monk A, editors. Funology 2: Springer; 2018:391-404. [doi: 10.1007/978-3-319-68213-6_25]
39. Kralj Novak P, Smailović J, Sluban B, Mozetič I. Sentiment of emojis. PLoS One 2015;10(12):e0144296. [doi: 10.1371/journal.pone.0144296] [Medline: 26641093]
40. Emojitracker: realtime emoji use. 2013. URL: https://emojitracker.com [accessed 2026-05-11]
41. AV O, AO U, CC A. Encoding and decoding of emoji on social media for expression sharing among university students in Anambra State. African Journal of Social Sciences and Humanities Research 2025;8(1):24-35. [doi: 10.52589/AJSSHR-IWUY1DYS]
42. Inbar A, Shor L. The depiction of negativity: the case of the brushing hands gesture used by Hebrew speakers. In: Blanchette F, Lukyanenko C, editors. Perspectives on Negation: A views from Across the Language Sciences: De Gruyter Mouton; 2025:21-56. [doi: 10.1515/9783110762068-002]
43. Cirillo F, Spina G, Irrera M, et al. The impact of expressive arts therapy on alexithymia levels in adolescent inpatients with severe anorexia nervosa. Children (Basel) 2025 Oct 16;12(10):1394. [doi: 10.3390/children12101394] [Medline: 41153576]
44. Yang Y, Wang W, Zuo Y. A study on multimodal metaphor of emotions in emojis. Int J Multiling 2025 Jul 3;22(3):1683-1720. [doi: 10.1080/14790718.2025.2451696]

XSL•FO
RenderX

45. Hölzer M, Pokorny D, Kächele H, Luborsky L. The verbalization of emotions in the therapeutic dialogue—a correlate of therapeutic outcome? Psychother Res 1997 Jan;7(3):261-273. [doi: 10.1080/10503309712331332013]
46. Hall-Renn KE. Mindful journeys: embracing the present with non-judgmental awareness. J Creat Ment Health 2007 Dec;2(2):3-16. [doi: 10.1300/J456v02n02_02]
47. Ito T, Nomura O, Hirai K, Ushikoshi H, Saiki T. Reduced test anxiety among medical students in an extremely low-stakes examination. Med Sci Educ 2025 Oct;35(5):2647-2652. [doi: 10.1007/s40670-025-02468-8] [Medline: 41555907]
48. Schubert E. Measuring emotion continuously: validity and reliability of the two-dimensional emotion-space. Aust J Psychol 1999 Dec 1;51(3):154-165. [doi: 10.1080/00049539908255353]
49. Zhu L, Li H, He W, Hong C. What influences online reviews' perceived information quality? Electron Libr 2020 May 13;38(2):273-296. [doi: 10.1108/EL-09-2019-0208]
50. Correia S, Costa J, Estanqueiro M, Antunes MJ, Oliveira L. Emotion: development of the mind and physical expression. Presented at: 9th Conference of European Sociological Association (ESA); Sep 2-5, 2009 URL: https://www.researchgate.net/publication/228829894_Emotion_development_of_the_mind_and_physical_expression [accessed 2026-05-11]
51. Galli F, Sartor G. The mismatch between GDPR and behavioural advertising. In: Ebers M, Sein K, editors. Privacy, Data Protection and Data-Driven Technologies: Routledge; 2024:287-327. [doi: 10.4324/9781003502791-16]
52. Greenaway KH, Kalokerinos EK. Suppress for success? Exploring the contexts in which expressing positive emotion can have social costs. Eur Rev Soc Psychol 2017 Jan;28(1):134-174. [doi: 10.1080/10463283.2017.1331874]
53. Derks YP, Klaassen R, Westerhof GJ, Bohlmeijer ET, Noordzij ML. Development of an ambulatory biofeedback app to enhance emotional awareness in patients with borderline personality disorder: multicycle usability testing study. JMIR Mhealth Uhealth 2019 Oct 15;7(10):e13479. [doi: 10.2196/13479] [Medline: 31617851]
54. Or CK, Chan AHS. Inspection methods for usability evaluation. In: Stephanidis C, Salvendy G, editors. User Experience Methods and Tools in Human-Computer Interaction: CRC Press; 2024:170-192. [doi: 10.1201/9781003495161-7]
55. Connelly M, Denney DR. Regulation of emotions during experimental stress in alexithymia. J Psychosom Res 2007 Jun;62(6):649-656. [doi: 10.1016/j.jpsychores.2006.12.008] [Medline: 17540222]
56. Rufer M, Hand I, Braatz A, Alsleben H, Fricke S, Peter H. A prospective study of alexithymia in obsessive-compulsive patients treated with multimodal cognitive-behavioral therapy. Psychother Psychosom 2004;73(2):101-106. [doi: 10.1159/000075541] [Medline: 14767152]
57. Lo C. Cultural values and alexithymia. SAGE Open 2014 Oct 1;4(4):2158244014555117. [doi: 10.1177/2158244014555117]

Abbreviations

CBT: cognitive behavioral therapy

Edited by A Kushniruk; submitted 10.Aug.2025; peer-reviewed by C Or, F Burstein; revised version received 04.Jan.2026; accepted 19.Jan.2026; published 01.Jun.2026.

Please cite as:
Liu S, Deng H, Chen J, Wang Q
Designing Coloring-Based Digital Art Therapy to Treat Alexithymia in Chinese College Students: Qualitative Study
JMIR Hum Factors 2026;13:e82128
URL: https://humanfactors.jmir.org/2026/1/e82128
doi:10.2196/82128

© Shuzhan Liu, Haoyong Deng, Jing Chen, Qi Wang. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 1.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

User Experience of a Web-Based Mobile Health App Supporting Low-Income Pregnant Individuals With Diabetes: Mixed Methods Study

Sydney L Raucher[1], BA; Layna Lu[1], BA; Tazim Merchant[1], MA, MD; Elizabeth Soyemi[1], AB; Charlotte Niznik[1], APN; Rana Saber[2], MS, MSL, PMP; Chen Yeh[3], MS; Lynn M Yee[1], MPH, MD

[1]Department of Obstetrics and Gynecology, Northwestern University Feinberg School of Medicine, 250 E. Superior Street, #5-2145, Chicago, IL, United States

[2]Research Application Design and Development Core Center, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[3]Department of Preventive Medicine, Biostatistical Collaboration Center, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

Corresponding Author:
Lynn M Yee, MPH, MD
Department of Obstetrics and Gynecology, Northwestern University Feinberg School of Medicine, 250 E. Superior Street, #5-2145, Chicago, IL, United States

Abstract

Background: Diabetes mellitus management requires considerable patient self-efficacy, knowledge, and support for social determinants of health. These needs become particularly acute during pregnancy. Mobile health (mHealth) tools are a promising approach to enhance patient engagement with the health care system, education, and health promotion and may be particularly helpful during the period of rapid skills acquisition, which is a hallmark of experiencing diabetes during pregnancy. Therefore, we developed SweetMama, a web-based mHealth app designed to support and provide information to low-income pregnant individuals with gestational diabetes mellitus (GDM) or type 2 diabetes mellitus (T2DM).

Objective: This study aimed to understand the user experiences of low-income pregnant people who were randomized to use SweetMama during a feasibility trial.

Methods: This mixed methods secondary analysis of data from a feasibility randomized controlled trial (RCT) included participants randomized to SweetMama, an interactive, web-based mHealth app with multiple motivational and educational features that help reduce barriers to care, offer health education, and aim to improve diabetes self-care for low-income pregnant people. In the parent trial, English-speaking pregnant individuals with GDM or T2DM were randomized to use SweetMama during pregnancy or usual care. SweetMama users experienced an individualized curriculum from enrollment through 6 weeks postpartum. Upon exit, users completed 2 qualitative interviews (during the delivery hospitalization and at the postpartum visit) and surveys assessing standardized usability metrics. The surveys included the System Usability Scale (SUS), the Usefulness, Satisfaction, Ease of Use (USE) scale, and the mHealth App Usability Questionnaire (MAUQ) to assess usability. Qualitative data were analyzed using constant comparative techniques.

Results: Of 30 SweetMama users, 60% (n=18) had GDM, 83.3% (n=25) had publicly funded prenatal care, and the majority identified as non-Hispanic Black (n=17, 56.7%) or Hispanic (n=11, 36.7%). Scores on the SUS (median 85.0/100, IQR 70.0 - 88.8; ≥71% indicates acceptable or higher usability), USE (overall median 84.5/100, IQR 81.0 - 91.4), and MAUQ (median 84.1/100, IQR 79.0 - 91.3) indicated favorable usability assessments, particularly for the “ease of learning” domain. Qualitative interviews supported these findings: participants described the app as easy to navigate, well organized, and helpful for staying on track, citing features such as clear visual design, timely text reminders, and actionable tips. Users valued motivational elements and content specificity, while recommending increased customization and enhanced esthetics.

Conclusions: In this user experience evaluation of a web-based mHealth app for low-income pregnant individuals with diabetes, participants found the tool to be user-friendly, visually appealing, informative, and motivating. Constructive feedback for application improvement for use in a future larger trial of clinical effectiveness was collected.

Trial Registration: ClinicalTrials.gov NCT03240874; https://clinicaltrials.gov/study/NCT03240874

(JMIR Hum Factors 2026;13:e84061) doi:10.2196/84061

XSL•FO
RenderX

KEYWORDS

gestational diabetes mellitus; type 2 diabetes; mobile health; mHealth; pregnancy; usability; user experience; self-efficacy; social determinants of health

Introduction

In the wake of a growing obesity epidemic, pregnant people are more likely to develop gestational diabetes mellitus (GDM) or enter pregnancy with type 2 diabetes mellitus (T2DM) [1-8]. Both GDM and T2DM increase the risk of maternal and fetal complications, such as preeclampsia, cesarean birth, and macrosomia, among others [3,6-10]. The rise in diabetes prevalence before and during pregnancy disproportionately affects low-income pregnant people of color, highlighting the need for interventions to incorporate social determinants of health in their development to reduce the risk of adverse outcomes [2,8-18].

Although diabetes treatment can lower the risk of complications [19-21], many logistical and social barriers prevent adequate diabetes self-management [22-27]. To optimize health outcomes, pregnant individuals with diabetes must be skilled in communication, organization, health literacy, and numeracy [22,25,28]. Strong social support systems are also beneficial [25,26,29-31]. Socially vulnerable individuals may not have sufficient access to educational resources for diabetes and are at greater risk of having suboptimal social support, leading to inadequate glucose control and poor health outcomes [25,26,28-31]. Thus, there is a need for interventions that educate patients, promote the skills needed for self-management, and address socioeconomic and racial disparities for pregnant individuals with diabetes [27,29].

Mobile health (mHealth) technology has the potential to address health disparities and promote health education through tailored and literacy-appropriate content [27]. With the widespread use of smartphones, mHealth can reach many individuals and reduce burdens such as transportation to and from health care providers [32]. In addition to being accessible, mHealth interventions are relatively low-cost and capable of providing support outside of clinical encounters [32,33]. mHealth tools for diabetes in nonpregnant populations have been effective at promoting self-management [34,35]. However, for both pregnant and nonpregnant populations, there is a lack of technology centered on increasing health literacy and self-efficacy skill development for diabetes management [36]. Furthermore, the tools that do exist are rarely designed to be user-centered (eg, they tend to rely on text message- or telemedicine-based communication rather than user-friendly applications) and are infrequently targeted toward individuals at greater social disadvantage [37].

To address the lack of mHealth tools aimed at supporting diabetes self-management among low-income, racially marginalized pregnant people, we created a multiphase project called SweetMama (NCT03240874). User feedback from an earlier phase of this project, which tested a text message–based intervention to promote diabetes self-management, expressed the desire for a more interactive, expansive mHealth tool [38]. In response, we developed the SweetMama app, a motivational and educational tool aimed at addressing diabetes self-management needs disproportionately experienced by English-speaking, socially vulnerable pregnant individuals [39]. Early feedback on this tool was explored in focus groups that included low-income, racially marginalized, English-speaking pregnant people. This feedback ensured that the app was literacy-appropriate and tailored toward the needs of this population (eg, by including preferred recipes and local resources). In the next phase of the study, a randomized controlled trial (RCT) was conducted to evaluate the feasibility of SweetMama use. The purpose of this planned secondary analysis from data gathered in the feasibility RCT is to evaluate the user experience of SweetMama participants, in order to consider how user experiences inform the optimization of mHealth interventions for this unique population.

Methods

Study Overview

This is a mixed methods analysis of data collected from the SweetMama feasibility RCT, which aimed to understand the user experience with this web-based mHealth app, including usability and satisfaction. Per users' requests from an earlier phase of the project, we created a web-based mHealth app, which expanded upon our previous text message–based intervention and was an accessible, interactive option for our patient population. The home screen of SweetMama provides the app's logo as well as personalized information, including the participant's name, gestational age, upcoming appointments, and a support portal (Figures 1 and 2). SweetMama delivered literacy-appropriate weekly motivational tips and individualized goal reminders, housed in a message center for easier access (Figure 2). Goal achievement was designed to include an element of gamification, whereby users could earn trophy icons for stating they had achieved their personalized goals. A library feature offered trusted educational materials, clinic-based resources, and recipes, with the option for users to "favorite" items for later use (Figure 2). If participants were interested in receiving further information from a certain message, they could click "I would like additional tips" to receive an in-depth description related to the message (Figure 2). During the postpartum period, the SweetMama curriculum focused on reminding participants about postpartum visit attendance, while also providing resources on diet maintenance during breastfeeding, mood changes, and more. This curriculum aimed to instill self-efficacy skills in patients and ensure that they were staying on track with the goals they set during pregnancy.

Data included in this analysis were generated from qualitative interviews addressing participants' opinions and feedback on SweetMama, which was designed to provide insight into the app's usability, as well as surveys assessing standardized usability metrics. The aim of the overarching SweetMama study was to test the feasibility and acceptability of this socially tailored, educational mHealth curriculum to encourage maternal self-efficacy for diabetes management among English-speaking

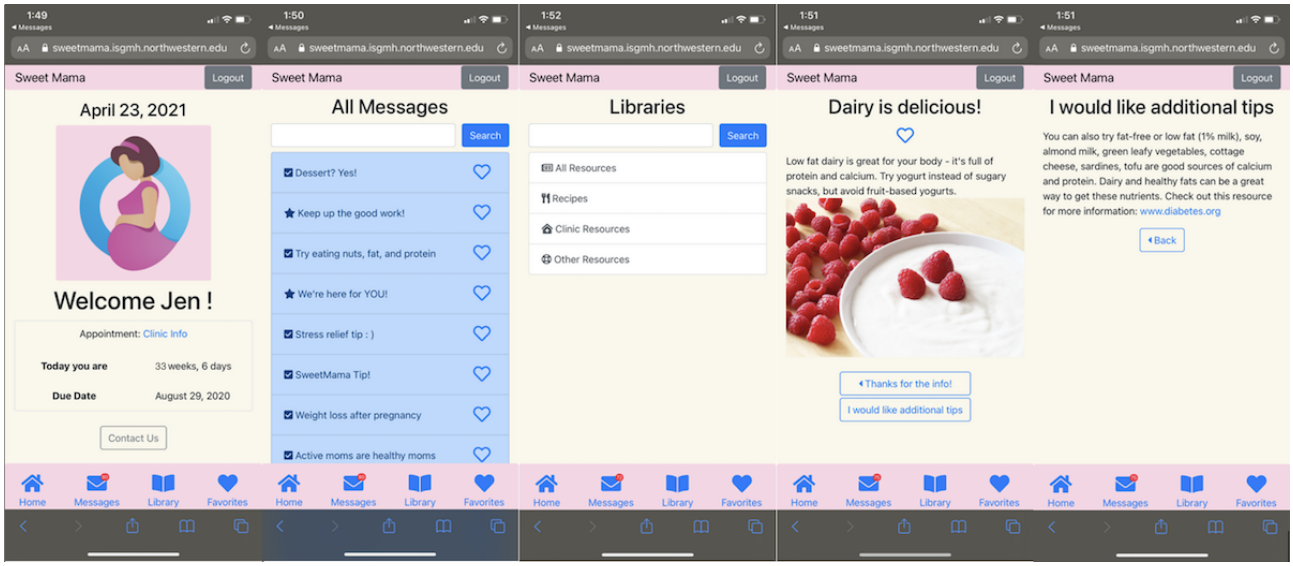
XSL•FO
RenderX

pregnant individuals. This mixed methods secondary analysis analyzed user experiences and satisfaction with SweetMama, both of which were predesignated secondary outcomes of this study assessed only in the participants randomized to use SweetMama.

Figure 1. SweetMama logo.



Figure 2. SweetMama screenshots of application features.



Recruitment

In the parent trial, research team members met with SweetMama participants at 3 separate visits: the enrollment visit (<30 wk of gestation), delivery visit, and postpartum visit. Consent was obtained at the enrollment visit. Individuals were recruited from the study site, Northwestern Medicine, at the initiation of prenatal diabetes care. Participants met the following inclusion criteria: age 18 and older, GDM or T2DM, English-speaking, confirmed intrauterine pregnancy prior to 30 weeks' gestational age, low income (defined as use of public insurance for prenatal care or household income <200% of the poverty line for family size), and access to a smartphone. This analysis focuses only on data from those who were randomized to use SweetMama. Users engaged with SweetMama throughout their pregnancy.

Quantitative Data Collection and Analysis

Users completed 3 validated, standardized assessments of usability at the time of their delivery hospitalization or within the first week after delivery. These included the System Usability Scale (SUS), the Usefulness, Satisfaction, Ease of Use (USE) scale, and the mHealth App Usability Questionnaire (MAUQ). The SUS, USE, and MAUQ were scored using standard scoring strategies for each scale, with scores normalized

XSL•FO
RenderX

to a 0 to 100 scale for each questionnaire. The SUS survey is a 10-item questionnaire using a 5-point Likert scale that provides a measure of overall user satisfaction [40,41]. A system is "acceptable" if scores are above 70, "marginal" if scores are between 51 and 70, and "not acceptable" if scores are 50 and below [41]. The USE survey (a 30-item questionnaire), which similarly assesses the usability of a system through a 7-point Likert scale, has components that address 4 domains: usefulness, ease of use, ease of learning, and satisfaction [40,42]. The MAUQ (21 items, 7-point Likert scale) assesses ease of use and satisfaction, system information arrangement, and usefulness in its version specifically for interactive mHealth tools and patients [43]. These tests are also reliable indicators of usability, acceptance, and satisfaction for evaluating mHealth resources (Cronbach α=0.91 [SUS], 0.98 [USE], 0.90 [MAUQ]) [41-43]. All scores were converted to a standard scale of 0 to 100. Descriptive summary statistics (median and IQR) were calculated. Quantitative analyses were performed using Stata v15 (StataCorp).

Qualitative Data Collection and Analysis

Two interviews were conducted to gather qualitative data about participants' experiences. The first interview occurred during the participants' delivery hospitalization or within the first week after delivery. Research personnel used an interview guide developed by the research team to ask questions and obtain feedback on areas of improvement for SweetMama through addressing three domains: (1) user experience, (2) app functionality (including features, curriculum, and preferences), and (3) user diabetes management. All interviews were conducted by a trained full-time research team member who was not involved in participants' clinical care, and who had no prior clinical relationship with participants. Participants were informed that their responses would not affect their medical care or study incentives, and they were encouraged to provide both positive and negative feedback to improve SweetMama. The second interview took place at the postpartum visit, approximately 6‐8 weeks postpartum. This qualitative interview focused on SweetMama's postpartum curriculum, with questions tailored toward feasibility, improvement, and user engagement with the app. Interviews lasted approximately 20 to 40 minutes. They were audio-recorded and transcribed using a professional service. Qualitative data were analyzed using the constant comparative method with a team-based approach [44,45]. A subset of 5 transcripts was reviewed by trained independent coders, who identified recurring concepts, referred to as subthemes, related to users' experiences with SweetMama. These subthemes were used to develop a preliminary codebook, which was discussed and refined by the full research team to ensure clarity and consistency. Two independent coders trained using Dedoose (SocioCultural Research Consultants, LLC), a qualitative coding software, then used the revised codebook to iteratively code the transcripts. The codebook continued to be updated during regular team meetings, and those changes were applied. Any discrepancies between coders were resolved on a code-by-code basis. Delivery and postpartum interviews were analyzed as a single corpus because the interview guides were designed to capture complementary aspects of a continuous user experience over time. Thematic saturation was determined through an iterative, team-based process.

Mixed Methods Integration

We used a mixed methods design in which qualitative interviews and quantitative usability surveys were collected during the same study visits and analyzed independently. Following separate analyses, results were brought together during interpretation using a narrative, side-by-side approach to examine areas of convergence and divergence. Quantitative usability scores were used to contextualize qualitative themes, and areas of alignment or mismatch were discussed during team meetings. These integrated insights informed overall conclusions on the user experience of SweetMama.

Ethical Considerations

All participants provided written informed consent prior to enrollment. Participants received up to US $100 via gift cards for their participation in the parent trial (US $25 for the enrollment visit, US $50 for the delivery visit, and US $25 for the postpartum visit). Participants who completed an interview received an additional US $10. This study was approved by the Northwestern University Institutional Review Board (IRB STU00205609).

Results

Overview

From January 2020 to October 2020, 30 participants were randomized to use SweetMama and were included in this analysis. Of these 30 participants, 18 had GDM and 12 had T2DM. The majority identified as non-Hispanic Black (n=17, 56.7%) or Hispanic (n=11, 36.7%). The majority (n=21, 70%) had completed at least some college education. All participants met the study criteria for low-income status, and 83.3% (n=25) had publicly funded prenatal care (Table 1).

XSL•FO
RenderX

Table . Participant demographics of SweetMama.

Characteristic	SweetMama users (N=30)
Age (y), mean (SD)	31.4 (4.7)
Self-reported race and ethnicity, n (%)	
Non-Hispanic white	1 (3.3)
Hispanic	11 (36.7)
Non-Hispanic Black	17 (56.7)
American Indian/Alaska Native	1 (3.3)
Medicaid, n (%)	25 (83.3)
Education, n (%)	
Some high school or less	3 (10)
High school graduate	6 (20)
Some college or technical school	9 (30)
College or technical school graduate	12 (40)
Relationship status, n (%)	
Married	14 (46.7)
Single or other relationship status	16 (53.3)
Nulliparous, n (%)	6 (20)
Current work status, n (%)	
Work full-time	11 (36.7)
Work part-time	8 (26.7)
Unemployed, student, or other employment status	11 (36.7)
Body mass index at first prenatal visit, kg/m^2 (n=29), mean (SD)	42.94 (12.4)
Diabetes diagnosis during pregnancy, n (%)	
Gestational diabetes mellitus	18 (60)
Type 2 diabetes mellitus	12 (40)
Gestational age at enrollment >20 weeks, n (%)	14 (46.7)

Quantitative Usability Metrics

At the delivery visit, 28 participants completed the surveys, and 2 participants were lost to follow-up. SUS, USE, and MAUQ scores demonstrated positive experiences with the SweetMama tool (Table 2). The median score on the SUS survey was 85.0/100 (IQR 77.0 - 88.8), and 71.4% (n=20) of participants found SweetMama to be acceptable. This generally positive feedback was corroborated by similarly high USE scores (median score 84.5/100, IQR 81.0 - 91.4), with the 4 domains of the USE scale (usefulness, ease of use, ease of learning, and satisfaction) scoring 82.1 (IQR 71.4 - 85.7), 88.3 (IQR 84.4 - 94.8), 91.1 (IQR 85.7 - 100.0), and 82.7 (IQR 73.5 - 93.9), respectively. Finally, the median MAUQ score was 84.1/100 (IQR 79.0 - 91.3). All quantitative metrics suggest favorable usability assessments.

XSL•FO
RenderX

Table . System Usability Scale (SUS), Usefulness, Satisfaction, Ease of Use (USE) scale, and mHealth App Usability Questionnaire (MAUQ) quantitative metrics (N=28).

Scale and score type	Values
SUS	
Overall score, median (IQR)	85.0 (70.0 - 88.8)
Acceptability, n (%)	
Acceptable (>70)	20 (71.4)
Marginal (51-70)	7 (25)
Not acceptable (<51)	1 (3.6)
USE scale, median (IQR)	
Overall score	84.5 (81.0 - 91.4)
Usefulness score	82.1 (71.4 - 85.7)
Ease of use score	88.3 (84.4 - 94.8)
Ease of learning score	91.1 (85.7 - 100.0)
Satisfaction score	82.7 (73.5 - 93.9)
MAUQ, median (IQR)	
Overall score	84.1 (79.0 - 91.3)

Qualitative User Experience Data

Qualitative Themes

At the delivery visit, 28 participants completed interviews, while 2 participants were lost to follow-up. At the postpartum visit, an additional 3 participants were lost to follow-up, resulting in 25 completed postpartum interviews. Most participant feedback centered on technical aspects of the application that enhanced the user experience. Analysis revealed four major themes; participants cited aspects of the (1) usability, (2) visual features, (3) content, and (4) frequency of communication in their feedback. Themes and subthemes, as well as participant-identified areas for improvement, are described later and in Table 3.

XSL•FO
RenderX

Table . Qualitative interview quotes on SweetMama user experience.

Theme and subtheme	Exemplary quotation
Usability	
Ease of use	"The app, it was easy, self-explanatory. You really didn't need anybody to explain it to you. But it was helpful in the sense that you know where to go and what you were doing, if that makes sense."
Organization	"Everything was set up in folders. I remember that. Motivation, goals. It was well organized. Easy to go back into and see what I wanted to remember or to read again."
Visual features	
Logo	"Well, I love the picture of the pregnant mommy. But mainly it, it comes directly to my, to my messages, so I was able to see when I had one from you guys. And they come, it came right up, so it was like, you know if, even if I didn't open it up right away, I could just click on the link and then still get it sent back to me. You know…that was cool."
Gamification	"I mean it had recipes and notes and little trophy award things if you were doing stuff correctly. It's like you had your own personal champion cheering you on. So yeah [I liked it]."
Content	
Tips and messages	"I think the tip messages were slightly more specific and I appreciated that. It gave me something more specific of a goal or an idea to think of just like when it's like did you sign up for WIC or this is a way to like relieve stress specifically or this is a specific way to increase your protein and your snacking and things like that. I felt like it was very tangible."
Goals	"It's still giving me information and still staying on top of what I need to be doing. So if I need to set a goal and you guys do remind me to stay on that goal…"
Recipes	"I really did like that because I know things that I thought were okay to eat, I would go on there and I would look at it and it was high in sugars so it was kind of new to me and I was like okay, I didn't know that."
Appointment reminders	"There were appointments that I would completely forget about and your text messages actually helped."
Frequency of communication	
Appropriate frequency	"[Message frequency] was just right to me…to keep you active…So you know how you get an app or something and you don't even use it you know. So like keep you active."

Usability

Themes emerged relating to two specific aspects of the app's usability: (1) ease of use and (2) app organization. SweetMama was widely appreciated for its *ease of use* by the cohort. When asked about their experience with the app, participants found SweetMama to be self-explanatory and easy to navigate. SweetMama also helped promote feelings of self-efficacy, as participants felt in control of their app experience. One participant commented: "You really didn't need anybody to explain it to you."

The *organization* of the app also helped participants with the comprehension of resources. It was easy for users to find the specific information they were seeking and to return to information they had saved or "favorited" for later review. One participant reported: "The goals are where the goals are… it's easy to get the stuff you need, information you need. You know it's easy."

Visual Features

SweetMama users favored the visual features of the app. These attributes helped promote user engagement, as they created a sense of recognition and reward when accessing and using the application. Two specific areas of positive feedback regarding visual recognition were the logo or icon and gamification aspects.

Regarding the *logo*, participants appreciated the SweetMama logo of the pregnant mother. Participants felt like they were able to relate to the picture: "It shows me that it, you know that it's about expecting mothers."

The use of a representative logo created a sentiment of belonging and support. Participants also reported that the logo was recognizable and aided in identifying information from the application.

Participants also enjoyed the addition of awards to the app. *Gamification* heightened participants' sense of accomplishment

XSL•FO
RenderX

and motivation to continue striving to achieve their goals: "It's like you had your own personal champion cheering you on."

Content

Another highlight of the SweetMama app was its content. Analysis revealed that the feedback centered around four key areas: (1) tips and messages, (2) goals, (3) recipes, and (4) appointment reminders.

First, *tips and messages* served as a guiding force, coaching participants on what they were "supposed to do." Participants liked how holistic the tips and messages were, targeting "every aspect of diabetes," and were even excited to see new content from SweetMama: "I was curious about what was the tip or what was the goal this time, you know? I was always looking forward to reading the messages."

In addition, participants noted that the specificity of the messages and the tangible nature of the tips made lifestyle goals more realistic to accomplish. Participants liked the references to specific foods to incorporate into their diets or support programs to find locally or online.

Goals set in the app also helped participants stay on track with diabetes management. Multiple participants reported that the goals motivated them to "work a little bit harder." Several participants reported how the goal-setting function and the reminders about those goals specifically helped with being proactive and motivating action: "They were inspirational, they made me want to use the app more… They made me want to set goals and complete them and then go back into the app and show that I actually did something."

Many participants reported diet as their largest struggle. Thus, they found *recipes* on SweetMama especially useful to help them navigate their new diets. Participants appreciated that SweetMama gave them guidance on diet, which helped dispel any uncertainties: "things that I thought were okay to eat, I would go on [SweetMama] and I would look at it and it was high in sugars...and I was like okay, I didn't know that." Recipes addressed issue areas such as portion sizes and the blood sugar levels of foods. Participants also appreciated the variety in recipe options, which covered meals and snacks.

Finally, the SweetMama text messaging system for *appointment reminders* was effective in helping participants remember their appointments. One participant shared: "There were appointments that I would completely forget about and your text messages actually helped."

Frequency of Communication

Participants were specifically queried about the frequency of communication and reminders within the SweetMama curriculum. Most participants found that the communication was of an appropriate intensity (3 times per wk, with additional reminders for upcoming appointments). The frequency promoted user engagement, retained user activity, and did not detract from their purpose.

Areas for Improvement

In addition to strengths, participants identified several areas for improvement. Interview guides included explicit prompts asking users to describe frustrations, dislikes, and suggested changes to encourage candid negative feedback. Suggested improvements clustered into 3 primary themes.

First, participants desired greater personalization, including the ability to tailor goals, reminders, and educational content to their specific diabetes type, treatment plan, and daily routines.

Second, users noted that the app's visual design and esthetics felt "rudimentary" and recommended enhanced graphics, more engaging layouts, and opportunities for customization (eg, personalized icons or themes) to improve user engagement.

Third, participants expressed interest in expanded content breadth, such as additional practical resources and more individualized messaging throughout pregnancy and postpartum. These findings highlight opportunities to refine SweetMama's usability in future iterations.

Discussion

Principal Findings

This study used both qualitative and quantitative methods to examine the user experience of a novel web-based mHealth app designed to augment health literacy and patient motivation and reduce logistical and social barriers for marginalized English-speaking pregnant individuals with diabetes. To our knowledge, this is among the first mixed methods evaluations of a pregnancy-focused, web-based mHealth app specifically designed to be literacy-conscious, responsive to practical barriers (eg, time constraints and food access), and intended to support individuals experiencing socioeconomic challenges. SweetMama users demonstrated high usability, including high satisfaction scores, based on quantitative assessments completed during postpartum hospitalization. These favorable usability metrics were corroborated by qualitative data; in interviews, participants cited usability, visual features, content, and frequency of communication as beneficial. Participants perceived appointment reminders and goal-tracking features as helpful for navigating competing responsibilities and time-management challenges common in settings of socioeconomic stress or low social support. Specific, actionable tips, including guidance on WIC enrollment (Women, Infant, Children—a federally funded nutrition program implemented by the Chicago Department of Public Health), affordable nutrition strategies, and recipes, were described as useful by participants facing food insecurity and limited nutrition literacy. Participants also valued the app's timely messages, motivational elements, and gamification, which contributed to perceived encouragement and support. Together, these findings illustrate how SweetMama's design features function as practical interventions for social barriers to diabetes self-management in an English-speaking, low-income pregnant population. In terms of areas of improvement, participants expressed a desire for increased personalization and esthetic appeal of the application, such as tailoring the tool's goals to patients' specific diabetes needs and allowing users to personalize the icon.

Comparison to Previous Work

These findings are consistent with the literature on mHealth interventions for patients with GDM. While various mHealth

XSL•FO
RenderX

options for GDM have been developed, only a few have evaluated the user experience of their participants. Nicholson et al [46] developed a web-based intervention, GooDMomS, which was designed to change the behavior of and educate users with GDM through informative web lessons, an online self-monitoring diary (for glucose, exercise, and diet tracking), recipes, a peer-support message board, and weekly text messages. Similar to our findings, guidance on diet was an important component for these participants. One new feature that users in GooDMomS desired was a list of suggested entrées from restaurants, rather than solely providing recipe options [46]. Personalization was also underscored in GDM literature as an important feature of an mHealth app. Duan et al [32] found that participants desired customizable features for diet and exercise goals so that they would be able to receive information about foods and workouts that were amenable to them. This individualization can help to engage patients and make them feel as though the tool is addressing their specific needs [32]. Our findings expand on existing literature by providing a mixed methods approach among a low-income, racially and ethnically diverse population of urban, American pregnant people.

These themes were also reflected in the broader type 1 diabetes mellitus (T1DM) and T2DM literature, in which several mHealth interventions have been evaluated for nonpregnant populations. In an mHealth app called Diabetes Journey, Schmidt et al [47] implemented a learner-centered curriculum for adolescents with T1DM. Consistent with our findings, the researchers found that the following factors promoted usability: interactive components, consistency in content and interaction design, variety, and interface esthetics. Furthermore, in a qualitative evidence synthesis of T2DM mHealth tools, O'Neill et al [48] found that people living with T2DM felt motivated to use mHealth tools for self-management. Similar to our findings, they favored applications that are visually appealing and incorporate features to communicate with a provider [48].

Existing literature also noted features that can be considered or expanded upon in future iterations of SweetMama. Participants in the study by Duan et al [32] emphasized that amidst the increased stress during pregnancy, being able to quickly reach out to a health care professional to ask questions would help relieve uncertainty and reduce the burden of traveling to the clinic. SweetMama did not include direct, personalized patient-to-provider communication options as patients had access to the institution's electronic medical record patient portal for such communication. Since participants were advised to use that portal for clinician communication, interviews did not directly query about this concept, but future work may benefit from incorporating a connection to the portal within an mHealth educational tool. Additionally, while participants did not explicitly mention peer support as a desired feature of SweetMama, this could serve as another aid in diabetes self-management, as users could interact with other people with similar experiences and goals [49-51]. Peer support can foster a sense of community and emotional support among pregnant individuals, which can be especially helpful for SweetMama's target audience who may lack adequate psychosocial support. Future work may aim to identify ways to connect users through a virtual SweetMama community of active users.

Strengths and Limitations

This study's strengths lie both in its population and its mixed methods approach. SweetMama was tailored specifically toward socially vulnerable people with fewer socioeconomic resources. Historically, the health concerns of socially vulnerable populations have been neglected; this study adds to the existing literature by obtaining feedback from a more diverse population regarding an emerging web-based mHealth app focused on building self-efficacy, motivation, and health literacy. These perspectives are key in building equitable mHealth interventions and can inform the development of future iterations of the tool. An additional strength is the study's use of both standardized surveys and interviews; the SUS assessed the overall usability of SweetMama, USE assessed user satisfaction and ease of learning, and MAUQ was specifically tailored to mHealth tools. When coupled with qualitative interviews and rigorous methodological analysis, we were able to glean a holistic understanding of the SweetMama user experience.

However, this study has several limitations. First, participants were recruited from a single urban academic institution and represent a relatively narrow demographic scope, which limits the generalizability of findings to pregnant people receiving care in other geographic, clinical, or socioeconomic settings. The small sample size and single recruitment location, both of which were appropriate for this phase of work, may not capture the full range of experiences among individuals with GDM or other high-risk pregnancies. For this phase of work, our sample size aligned with established norms for usability testing and mixed methods evaluations, allowed us to achieve thematic saturation in qualitative interviews, and was sufficient to detect patterns in quantitative usability scores, which showed limited variation across participants. Nonetheless, the sample lacked representation from Asian and other underrepresented groups. Additionally, all participants had to be English-speaking, despite diabetes being prevalent in Hispanic and immigrant populations, due to budgetary limitations during the intervention development process [1,2]. Future work must include a larger sample size with increased racial and ethnic representation to more fully develop and test an intervention tailored to different cultures. Another necessary next step is translating the app into different languages to suit a more linguistically diverse patient population, as well as studying the effect of SweetMama use on clinical outcomes.

The visual appeal of the app was also limited due to budgetary constraints; therefore, future work toward greater customization and enhanced graphics could enhance the appearance of SweetMama. Next, while we did not formally integrate datasets (eg, by comparing qualitative responses with quantitative scores), the qualitative and quantitative findings were used to complement one another and provide a more holistic understanding of user experiences. Due to limited variation in user scores (ie, scores were generally high), further subgroup analysis was not feasible. In addition, because participants had regular contact with the research team and received incentives for study completion, social desirability bias may have influenced both usability ratings and qualitative feedback. Finally, the current study was designed to evaluate perceived usability, satisfaction, and early user experiences with

XSL•FO
RenderX

SweetMama; sustained user engagement and clinical effectiveness were beyond its scope and should be assessed in future studies.

Future Directions

Based on our findings, several key design implications for maternal mHealth tools emerge: (1) prioritize ease of learning and low-burden access to support engagement among users with competing demands; (2) incorporate structured, organized content with actionable guidance (eg, recipes, nutrition tips, and resource navigation) to address health literacy and social needs; (3) include reminders and goal-tracking features to facilitate self-management and appointment adherence; and (4) leverage motivational framing, timely messaging, and light gamification to enhance user engagement and perceived social support. With these in mind, there are several clinical implications that can be gleaned from this work. mHealth apps like SweetMama can be highly appealing to users if they incorporate visual and knowledge-building features, while remaining focused on basic usability concepts such as ease of use and ease of learning. Applications designed with patient-centered features may support initial uptake and user satisfaction. These factors are ultimately important for sustained user engagement and improved self-management, particularly for those with greater social needs. Although SweetMama is not yet available outside of the research context, the lessons learned from this assessment can inform the development and sustainability of other interactive mHealth tools to support pregnant patients with diabetes. Our findings suggest that future maternal mHealth interventions may benefit from emphasizing low-burden interaction, actionable content, supportive reminders, and motivational features that help users integrate self-management tasks into their daily routines. For example, features such as appointment reminders, gamification, and a diverse array of recipe options, educational tips, and messaging are recommended by participants for inclusion in future designs. Additionally, new iterations of SweetMama can be created in different languages to support wider audiences and reduce health literacy and language barriers for non–English-speaking individuals. Future work may further consider whether and to what extent the experiences of users with different types of diabetes (eg, diet-treated vs medication-treated; pregestational vs gestational) may warrant alterations of interventions such as SweetMama. Further personalization can include opportunities to tailor the application based on stage of pregnancy or postpartum status, dietary preferences and restrictions, reminder frequency, and desired support resources.

In conclusion, mHealth tools are a promising avenue for GDM self-management, education, and addressing disparities. SweetMama users found this web-based mHealth app to be engaging and motivational, with its strengths lying in usability, visual features, and content. Areas for improvement included increased customization of the app's features, such as the logo and personal diabetes needs. These findings support user satisfaction and the perceived usefulness of the tool, but further research is needed to evaluate sustained engagement with the app, behavior change, and clinical outcomes. Future directions for this work also include testing SweetMama in a larger sample size, in multiple languages, and across multiple institutions. In addition, implementing new features, such as peer support and connection to patient portals, may enhance the user experience.

Acknowledgments

We would like to thank research coordinator Jenise Jackson, MPH, and research assistant Karolina Leziak, MD, for their work on the study. We would also like to thank the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R21 HD094271) for their support.

Generative AI tools were not used in the conceptualization, data curation or analysis, visualization, or original drafting of the manuscript. However, Generative AI tools were used solely for grammatical and stylistic improvements while revising the manuscript text. The authors remain fully responsible for the accuracy, originality, and integrity of all content in the manuscript.

Funding

This research was funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R21 HD094271) from August 2018 to January 2022.

Data Availability

Given the sensitive and individualized nature of qualitative data, these data will not be publicly available. Investigators wishing to access these data may contact the corresponding author. Data may be made available upon reasonable request and with evidence of appropriate approvals.

Authors' Contributions

LMY and CN contributed to conceptualization. LL, ES, and LMY handled data curation. LL, ES, and LMY conducted the formal analysis. LMY and CN managed funding acquisition. LL, ES, and LMY contributed to methodology. LMY, CN, and RS handled project administration. RS led project intervention development and provided oversight. LMY and RS managed resources. LMY, CN, and RS conducted supervision. LL, ES, and LMY managed validation. CY conducted the statistical analysis. SLR and LL contributed to visualization. SLR, TM, and LL wrote the original draft. All authors conducted review and editing.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Pu J, Zhao B, Wang EJ, et al. Racial/ethnic differences in gestational diabetes prevalence and contribution of common risk factors. Paediatr Perinat Epidemiol 2015 Sep;29(5):436-443. [doi: 10.1111/ppe.12209] [Medline: 26201385]
2. Peng TY, Ehrlich SF, Crites Y, et al. Trends and racial and ethnic disparities in the prevalence of pregestational type 1 and type 2 diabetes in Northern California: 1996-2014. Am J Obstet Gynecol 2017 Feb;216(2):177. [doi: 10.1016/j.ajog.2016.10.007] [Medline: 27751798]
3. Catalano PM. The impact of gestational diabetes and maternal obesity on the mother and her offspring. J Dev Orig Health Dis 2010 Aug;1(4):208-215. [doi: 10.1017/S2040174410000115] [Medline: 25141869]
4. Shah NS, Wang MC, Freaney PM, et al. Trends in gestational diabetes at first live birth by race and ethnicity in the US, 2011-2019. JAMA 2021 Aug 17;326(7):660-669. [doi: 10.1001/jama.2021.7217] [Medline: 34402831]
5. Zhou T, Du S, Sun D, et al. Prevalence and trends in gestational diabetes mellitus among women in the United States, 2006-2017: a population-based study. Front Endocrinol (Lausanne) 2022;13:868094. [doi: 10.3389/fendo.2022.868094] [Medline: 35733768]
6. Huet J, Beucher G, Rod A, Morello R, Dreyfus M. Joint impact of gestational diabetes and obesity on perinatal outcomes. J Gynecol Obstet Hum Reprod 2018 Nov;47(9):469-476. [doi: 10.1016/j.jogoh.2018.08.003] [Medline: 30153508]
7. Alexopoulos AS, Blair R, Peters AL. Management of preexisting diabetes in pregnancy: a review. JAMA 2019 May 14;321(18):1811-1819. [doi: 10.1001/jama.2019.4981] [Medline: 31087027]
8. Sridhar SB, Ferrara A, Ehrlich SF, Brown SD, Hedderson MM. Risk of large-for-gestational-age newborns in women with gestational diabetes by race and ethnicity and body mass index categories. Obstet Gynecol 2013 Jun;121(6):1255-1262. [doi: 10.1097/AOG.0b013e318291b15c] [Medline: 23812460]
9. American College of Obstetricians and Gynecologists Committee on Practice Bulletins—Obstetrics. ACOG Practice Bulletin No. 201: pregestational diabetes mellitus. Obstet Gynecol 2018 Dec;132(6):e228-e248. [doi: 10.1097/AOG.0000000000002960] [Medline: 30461693]
10. Guo H, Zhang Y, Li P, Zhou P, Chen LM, Li SY. Evaluating the effects of mobile health intervention on weight management, glycemic control and pregnancy outcomes in patients with gestational diabetes mellitus. J Endocrinol Invest 2019 Jun;42(6):709-714. [doi: 10.1007/s40618-018-0975-0] [Medline: 30406378]
11. Diaz-Santana MV, O'Brien KM, Park YMM, Sandler DP, Weinberg CR. Persistence of risk for type 2 diabetes after gestational diabetes mellitus. Diabetes Care 2022 Apr 1;45(4):864-870. [doi: 10.2337/dc21-1430] [Medline: 35104325]
12. Guariguata L, Linnenkamp U, Beagley J, Whiting DR, Cho NH. Global estimates of the prevalence of hyperglycaemia in pregnancy. Diabetes Res Clin Pract 2014 Feb;103(2):176-185. [doi: 10.1016/j.diabres.2013.11.003] [Medline: 24300020]
13. Schwartz N, Nachum Z, Green MS. The prevalence of gestational diabetes mellitus recurrence--effect of ethnicity and parity: a metaanalysis. Am J Obstet Gynecol 2015 Sep;213(3):310-317. [doi: 10.1016/j.ajog.2015.03.011] [Medline: 25757637]
14. Bower JK, Butler BN, Bose-Brill S, Kue J, Wassel CL. Racial/ethnic differences in diabetes screening and hyperglycemia among US women after gestational diabetes. Prev Chronic Dis 2019 Oct 24;16:E145. [doi: 10.5888/pcd16.190144] [Medline: 31651379]
15. Bellamy L, Casas JP, Hingorani AD, Williams D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. Lancet 2009 May 23;373(9677):1773-1779. [doi: 10.1016/S0140-6736(09)60731-5] [Medline: 19465232]
16. Egan AM, Enninga EAL, Alrahmani L, Weaver AL, Sarras MP, Ruano R. Recurrent gestational diabetes mellitus: a narrative review and single-center experience. J Clin Med 2021 Feb 3;10(4):569. [doi: 10.3390/jcm10040569] [Medline: 33546259]
17. Venkatesh KK, Lynch CD, Powe CE, et al. Risk of adverse pregnancy outcomes among pregnant individuals with gestational diabetes by race and ethnicity in the United States, 2014-2020. JAMA 2022 Apr 12;327(14):1356-1367. [doi: 10.1001/jama.2022.3189] [Medline: 35412565]
18. Daneshmand SS, Stortz S, Morrisey R, Faksh A. Bridging gaps and understanding disparities in gestational diabetes mellitus to improve perinatal outcomes. Diabetes Spectr 2019 Nov;32(4):317-323. [doi: 10.2337/ds19-0013] [Medline: 31798289]
19. Poolsup N, Suksomboon N, Amin M. Effect of treatment of gestational diabetes mellitus: a systematic review and meta-analysis. PLoS One 2014;9(3):e92485. [doi: 10.1371/journal.pone.0092485] [Medline: 24658089]
20. Hartling L, Dryden DM, Guthrie A, Muise M, Vandermeer B, Donovan L. Benefits and harms of treating gestational diabetes mellitus: a systematic review and meta-analysis for the U.S. Preventive Services Task Force and the National Institutes of Health Office of Medical Applications of Research. Ann Intern Med 2013 Jul 16;159(2):123-129. [doi: 10.7326/0003-4819-159-2-201307160-00661] [Medline: 23712381]
21. Lende M, Rijhsinghani A. Gestational diabetes: overview with emphasis on medical management. Int J Environ Res Public Health 2020 Dec 21;17(24):9573. [doi: 10.3390/ijerph17249573] [Medline: 33371325]
22. Yee LM, McGuire JM, Taylor SM, Niznik CM, Simon MA. Factors promoting diabetes self-care among low-income, minority pregnant women. J Perinatol 2016 Jan;36(1):13-18. [doi: 10.1038/jp.2015.134] [Medline: 26513455]

XSL•FO
RenderX

23. Yee LM, McGuire JM, Taylor SM, Niznik CM, Simon MA. “I was tired of all the sticking and poking”: identifying barriers to diabetes self-care among low-income pregnant women. J Health Care Poor Underserved 2015 Aug;26(3):926-940. [doi: 10.1353/hpu.2015.0073] [Medline: 26320923]
24. Yee LM, McGuire JM, Taylor SM, Niznik CM, Simon MA. Social and environmental barriers to nutrition therapy for diabetes management among underserved pregnant women: a qualitative analysis. J Nutr Educ Behav 2016 Mar;48(3):170-180. [doi: 10.1016/j.jneb.2015.11.003] [Medline: 26706027]
25. Nielsen KK, Kapur A, Damm P, de Courten M, Bygbjerg IC. From screening to postpartum follow-up - the determinants and barriers for gestational diabetes mellitus (GDM) services, a systematic review. BMC Pregnancy Childbirth 2014 Jan 22;14(1):41. [doi: 10.1186/1471-2393-14-41] [Medline: 24450389]
26. Carolan M. Women’s experiences of gestational diabetes self-management: a qualitative study. Midwifery 2013 Jun;29(6):637-645. [doi: 10.1016/j.midw.2012.05.013] [Medline: 22877761]
27. Anderson-Lewis C, Darville G, Mercado RE, Howell S, Di Maggio S. mHealth technology use and implications in historically underserved and minority populations in the United States: systematic literature review. JMIR mHealth uHealth 2018 Jun 18;6(6):e128. [doi: 10.2196/mhealth.8383] [Medline: 29914860]
28. Pirdehghan A, Eslahchi M, Esna-Ashari F, Borzouei S. Health literacy and diabetes control in pregnant women. J Family Med Prim Care 2020 Feb;9(2):1048-1052. [doi: 10.4103/jfmpc.jfmpc_891_19] [Medline: 32318465]
29. Muhwava LS, Murphy K, Zarowsky C, Levitt N. Perspectives on the psychological and emotional burden of having gestational diabetes amongst low-income women in Cape Town, South Africa. BMC Womens Health 2020 Oct 12;20(1):231. [doi: 10.1186/s12905-020-01093-4] [Medline: 33046050]
30. Rad GS, Bakht LA, Feizi A, Mohebi S. Importance of social support in diabetes care. J Educ Health Promot 2013;2(62):62. [doi: 10.4103/2277-9531.120864] [Medline: 24520558]
31. Miller TA, Dimatteo MR. Importance of family/social support and impact on adherence to diabetic therapy. Diabetes Metab Syndr Obes 2013 Nov 6;6(421-6):421-426. [doi: 10.2147/DMSO.S36368] [Medline: 24232691]
32. Duan B, Liu Z, Liu W, Gou B. Assessing the views and needs of people at high risk of gestational diabetes mellitus for the development of mobile health apps: descriptive qualitative study. JMIR Form Res 2022 Jul 8;6(7):e36392. [doi: 10.2196/36392] [Medline: 35802414]
33. Carrandi A, Hu Y, Karger S, et al. Systematic review on the cost and cost-effectiveness of mHealth interventions supporting women during pregnancy. Women Birth 2023 Feb;36(1):3-10. [doi: 10.1016/j.wombi.2022.03.007] [Medline: 35339412]
34. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: an overview of systematic reviews. PLoS One 2017;12(3):e0173160. [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]
35. Al-Hashmi I, Hodge F, Nandy K, Thomas E, Brecht ML. The effect of a self-efficacy-enhancing intervention on perceived self-efficacy and actual adherence to healthy behaviours among women with gestational diabetes mellitus. Sultan Qaboos Univ Med J 2018 Nov;18(4):e513-e519. [doi: 10.18295/squmj.2018.18.04.014] [Medline: 30988972]
36. Alqudah A, McMullan P, Todd A, et al. Service evaluation of diabetes management during pregnancy in a regional maternity hospital: potential scope for increased self-management and remote patient monitoring through mHealth solutions. BMC Health Serv Res 2019 Sep 13;19(1):662. [doi: 10.1186/s12913-019-4471-9] [Medline: 31514743]
37. Zahmatkeshan M, Zakerabasali S, Farjam M, Gholampour Y, Seraji M, Yazdani A. The use of mobile health interventions for gestational diabetes mellitus: a descriptive literature review. J Med Life 2021;14(2):131-141. [doi: 10.25122/jml-2020-0163] [Medline: 34104235]
38. Yee LM, Leziak K, Jackson J, et al. Patient and provider perspectives on a novel mobile health intervention for low-income pregnant women with gestational or type 2 diabetes mellitus. J Diabetes Sci Technol 2021 Sep;15(5):1121-1133. [doi: 10.1177/1932296820937347] [Medline: 32627582]
39. Yee LM, Leziak K, Jackson J, et al. SweetMama: usability assessment of a novel mobile application among low-income pregnant people to assist with diabetes management and support. Diabetes Spectr 2023 May;36(2):171-181. [doi: 10.2337/ds22-0014] [Medline: 37193207]
40. Hajesmaeel-Gohari S, Khordastan F, Fatehi F, Samzadeh H, Bahaadinbeigy K. The most used questionnaires for evaluating satisfaction, usability, acceptance, and quality outcomes of mobile health. BMC Med Inform Decis Mak 2022 Jan 27;22(1):22. [doi: 10.1186/s12911-022-01764-2] [Medline: 35081953]
41. Lewis JR. The system usability scale: past, present, and future. Int J Human Comput Inter 2018 Jul 3;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
42. Gao M, Kortum P, Oswald F. Psychometric evaluation of the USE (Usefulness, Satisfaction, and Ease of USE) questionnaire for reliability and validity. Proc Hum Factors Ergon Soc Annu Meet 2018 Sep;62(1):1414-1418. [doi: 10.1177/1541931218621322]
43. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth app usability questionnaire (MAUQ): development and validation study. JMIR mHealth uHealth 2019 Apr 11;7(4):e11500. [doi: 10.2196/11500] [Medline: 30973342]
44. Glaser BG. The constant comparative method of qualitative analysis. Soc Probl 1965 Apr;12(4):436-445. [doi: 10.1525/sp.1965.12.4.03a00070]
45. Hewitt-Taylor J. Use of constant comparative analysis in qualitative research. Nurs Stand 2001;15(42):39-42. [doi: 10.7748/ns2001.07.15.42.39.c3052] [Medline: 12212430]

XSL•FO
RenderX

46. Nicholson WK, Beckham AJ, Hatley K, et al. The Gestational Diabetes Management System (GooDMomS): development, feasibility and lessons learned from a patient-informed, web-based pregnancy and postpartum lifestyle intervention. BMC Pregnancy Childbirth 2016 Sep 21;16(1):277. [doi: 10.1186/s12884-016-1064-z] [Medline: 27654119]
47. Schmidt M, Lu J, Luo W, et al. Learning experience design of an mHealth self-management intervention for adolescents with type 1 diabetes. Educ Technol Res Dev 2022;70(6):2171-2209. [doi: 10.1007/s11423-022-10160-6] [Medline: 36278247]
48. O'Neill M, Houghton C, Crilly G, Dowling M. A qualitative evidence synthesis of users' experience of mobile health applications in the self-management of type 2 diabetes. Chronic Illn 2022 Mar;18(1):22-45. [doi: 10.1177/1742395320983877] [Medline: 33583207]
49. Alexandra Friedman M, Niznik CM, Bolden JR, Yee LM. Reciprocal peer support for post-partum patients with diabetes: a needs assessment for the Diabetes Buddy Program. J Community Health 2016 Apr;41(2):354-358. [doi: 10.1007/s10900-015-0103-4] [Medline: 26518777]
50. Safiee L, Rough DJ, Whitford H. Barriers to and facilitators of using eHealth to support gestational diabetes mellitus self-management: systematic literature review of perceptions of health care professionals and women with gestational diabetes mellitus. J Med Internet Res 2022 Oct 27;24(10):e39689. [doi: 10.2196/39689] [Medline: 36301613]
51. Ingstrup MS, Wozniak LA, Mathe N, et al. Women's experience with peer counselling and social support during a lifestyle intervention among women with a previous gestational diabetes pregnancy. Health Psychol Behav Med 2019 May 9;7(1):147-159. [doi: 10.1080/21642850.2019.1612750] [Medline: 34040844]

Abbreviations

GDM: gestational diabetes mellitus
MAUQ: mHealth App Usability Questionnaire
mHealth: mobile health
RCT: randomized controlled trial
SUS: System Usability Scale
T1DM: type 1 diabetes mellitus
T2DM: type 2 diabetes mellitus
USE: Usefulness, Satisfaction, Ease of Use
WIC: Women, Infant, Children

Edited by A Kushniruk, S Law; submitted 13.Sep.2025; peer-reviewed by A Lewkowitz, I Orimolade, V Singh; revised version received 15.Jun.2026; accepted 23.Jun.2026; published 31.Jul.2026.

Please cite as:
Raucher SL, Lu L, Merchant T, Soyemi E, Niznik C, Saber R, Yeh C, Yee LM
User Experience of a Web-Based Mobile Health App Supporting Low-Income Pregnant Individuals With Diabetes: Mixed Methods Study
JMIR Hum Factors 2026;13:e84061
URL: https://humanfactors.jmir.org/2026/1/e84061
doi:10.2196/84061

© Sydney L Raucher, Layna Lu, Tazim Merchant, Elizabeth Soyemi, Charlotte Niznik, Rana Saber, Chen Yeh, Lynn M Yee. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 31.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Novel Web-Based Technology to Promote Goal-Setting in Complex Chronic Illness: Randomized Controlled Trial

Jody Lin[1], MS, MD; Bernd Huber[2], PhD; Ofra Amir[3], PhD; Shiri Assis-Hassid[2], PhD; Sebastian Gehrmann[2], PhD; Krzysztof Gajos[2], PhD; Barbara Grosz[2], PhD; Lee Sanders[4], MPH, MD

[1]Department of Pediatrics, School of Medicine, University of Utah, 100 Mario Capecchi Drive, Salt Lake City, UT, United States

[2]Department of Computer Science, John A. Paulson School of Engineering and Applied Sciences (Computer Science), Harvard University, Cambridge, MA, United States

[3]Department of Data and Decision Sciences, Technion – Israel Institute of Technology, Haifa, Israel

[4]Department of Pediatrics, School of Medicine, Stanford University, Palo Alto, CA, United States

Corresponding Author:
Jody Lin, MS, MD
Department of Pediatrics, School of Medicine, University of Utah, 100 Mario Capecchi Drive, Salt Lake City, UT, United States

Abstract

Background: Shared goal-setting is a common feature of quality guidelines to improve care quality for children with medical complexity, but few studies have examined the efficacy of interventions designed to improve goal-setting.

Objective: This study aims to evaluate a novel internet-based tool to promote shared goal-setting (GoalKeeper) in the care of children with medical complexity.

Methods: We conducted a randomized stepped-wedge trial (intervention vs usual care) between April 1, 2019 and March 21, 2021, at primary and subspecialty care clinics at an academic medical center, including 11 medical providers (medical doctor, doctor of osteopathic medicine, or nurse practitioner). Adult parents of children with medical complexity were eligible if they were English-speaking, with home internet access, and with children with medical complexity aged younger than 12 years. Participants were clustered by provider, with providers crossing over from control to intervention at different stages. The assignment of timing for crossover was random. Control group participants received usual care without any additional interventions. Intervention group participants received a novel web-based tool called GoalKeeper for initial use during the clinic visit and subsequent continued use after the clinic visit. GoalKeeper was co-designed by parents, providers, and computer scientists to include 2 modules, meant to be used by both the parent and medical provider for each child: (1) goal elicitation, used synchronously during a clinic visit; (2) tracking, used asynchronously between visits by parents and providers of children with medical complexity. The primary outcome was quality of goal-setting assessed by the Patient Assessment of Care for Chronic Illness Care goal-setting domain at baseline (t1) and 1 month (t2). We conducted a repeated-measures mixed-effects ANOVA to evaluate between- and within-group differences over time for fixed effects (timing of intervention, intervention×time) and random effect (provider cluster).

Results: We enrolled 67 parent-child dyads (control: n=32 control; intervention: n=35). Parents had a mean age of 37.4 (SD 8.2) years, children with medical complexity with mean age 5.6 (SD 0.5) years, and 29 (44%) parents identified as Hispanic. Of the 35 intervention parents, 34 successfully used GoalKeeper during the clinical encounter with their provider. During the follow-up period, quality of goal-setting was sustained at t2 for the intervention group but declined for the control group (δ=0.03 vs −0.43; F_{49}=3.52, P=.06). Similar patterns were observed for care quality (δ=0.01 vs −0.48; F_{49}=4.28, P=.04).

Conclusions: Our study demonstrates that family-centered goal-setting may help combat the gradual decline in care quality otherwise experienced by children with medical complexity in between clinic visits. Successful use of the tool with providers in clinic suggests that digital tools are feasible interventions to change family-provider communication around family goals.

Trial Registration: ClinicalTrials.gov NCT03620071; https://clinicaltrials.gov/study/NCT03620071

(JMIR Hum Factors 2026;13:e70402) doi:10.2196/70402

KEYWORDS

children with medical complexity; digital health; family-centered care; chronic care; care coordination

XSL•FO
RenderX

Introduction

Children with medical complexity and their families face many challenges. Defined by high service needs, high resource use, and functional disability, children with medical complexity represent less than 5% of all children but a disproportionately high share of pediatric care use, coordinated across multiple health care systems, providers, and agencies [1-4]. This complex ecosystem of care hinders caregivers' and providers' abilities to have a shared understanding of care priorities to effectively address health needs as a team. Moreover, the fluctuating nature of health issues faced by children with medical complexity results in constant changes to the care team, and helping new team members be up-to-date is highly challenging. Thus, parents of children with medical complexity are less likely to report high care quality and shared decision-making during ambulatory care encounters than parents of children with noncomplex chronic conditions [5-7].

To improve care quality and shared decision-making for children with medical complexity, several national pediatric organizations have suggested the importance of a family-centered care framework centered around shared goal-setting to create individualized care plans [8,9]. While prior studies have used multidisciplinary teams for individualized care plan creation, few models exist for effective and scalable tools that facilitate shared goal-setting when prolonged visits with highly resourced teams are unavailable [10,11]. Studies of noncomplex chronic conditions in pediatrics (eg, asthma or type 1 diabetes) suggest that technology-based tools may provide efficacious ways to establish and track goals during and between ambulatory care visits [12-14]. These tools result in improved satisfaction, reduced urgent care visits, and improved outcomes. However, technology-based solutions for the care of children with medical complexity must overcome unique challenges such as team hierarchies, loosely coupled teams (eg, most providers care for children with medical complexity in diffuse teams rather than tightly coordinated team-based clinics), and asynchronous time scales among providers [15]. Prior work suggests that shared care plans for children with medical complexity require common goals to be set synchronously by the provider and caregiver, then tracked asynchronously by both but has not previously been tested [15]. It is unknown whether technology-based tools can be effective in improving the care of children with medical complexity.

In this study, we evaluate the effectiveness of a technology-based tool for eliciting and tracking shared goals in the ambulatory care of children with medical complexity. Working in a multimodal development team comprised of health care providers, behavioral scientists, computer scientists, and parents of children with medical complexity, we developed a novel, health-information technology platform, called GoalKeeper, to improve goal-concordant conversations in real-world, outpatient care settings. We hypothesized that parents of children with medical complexity exposed to GoalKeeper would report higher care quality and higher quality goal-setting when compared to parents who received usual care.

Methods

Study Design

We conducted a prospective, stepped-wedge, randomized controlled trial of intervention efficacy. The study was conducted at primary and subspecialty ambulatory clinics associated with a children's hospital at an academic medical center. In the stepped-wedge design, each provider first spent at least 3 weeks in the control condition, when research staff recruited parent-child dyads from their clinic to be in the control group, followed by crossover into the intervention group for 3 weeks. Providers crossed over at different intervals spaced 3 weeks apart (Figure S1 in Multimedia Appendix 1). The order that providers crossed over was randomized in R Statistical Software (v.3.5.3 R Core Team 2019) via RStudio (v.1.1.463 Posit team 2019) by author JL, with allocation and enrollment conducted by research assistants [16]. Each parent-child dyad served in either the control or intervention group, with the control group receiving usual care. We chose a stepped-wedge design because it offers the most feasible and effective approach to randomization of the parent-child dyads without introducing bias in the control group by changing provider behavior through exposure to the intervention [17]. Ethically, stepped-wedge designs are better suited for studies in which true equipoise has not been established for the interventions being tested [18]. Prior interviews with providers of children with medical complexity found that most providers were familiar with the practice of goal-setting and endorsed some level of goal-setting in their current practice, but structured approaches to goal-setting were not part of routine practice in these clinics [15]. For this study, we adhered to the CONSORT guidelines for reporting randomized clinical trials (see Checklist 1).

Study Population

We recruited medical providers (medical doctor, doctor of osteopathic medicine, nurse practitioner, and physician assistant) from 2 ambulatory clinics that see a high proportion of children with medical complexity—pediatric neurology and complex primary care clinics (CPCC) at the main location of a single academic health center. The health center is part of a standalone tertiary children's health system in northern California. All providers that were not trainees at these clinics were eligible for recruitment. We recruited a convenience sample of primary caregivers (eg, parents) of the patients seen by enrolled providers in person at the time of their clinic visit. We capped enrollment per provider cluster at 10 parents. Parents were eligible if they were aged 18 years or older, English-speaking, and with a child with medical complexity aged younger than 12 years. We excluded parents without home internet access or with known neurocognitive or psychiatric issues that precluded the ability to offer informed consent. We excluded older children, who may have the capacity to participate in decision-making, since the intervention was not designed for interaction with children [19]. We defined medical complexity as meeting each of 2 criteria in the 12 months prior to study enrollment: (1) ambulatory visits with at least 2 different subspecialties affiliated with the medical center and (2) medical fragility due to technology dependence (gastrostomy tube, nasogastric tube, ventriculoperitoneal shunt, or tracheostomy) or clinic visits at

XSL•FO
RenderX

CPCC. Enrollment occurred between April 1, 2019 and December 21, 2020, with follow-up through March 21, 2021. Due to institutional restrictions prompted by the COVID-19 pandemic prohibiting in-person research activities that were necessary for study activities at enrollment, recruitment was paused between March 13, 2020 and July 20, 2020.

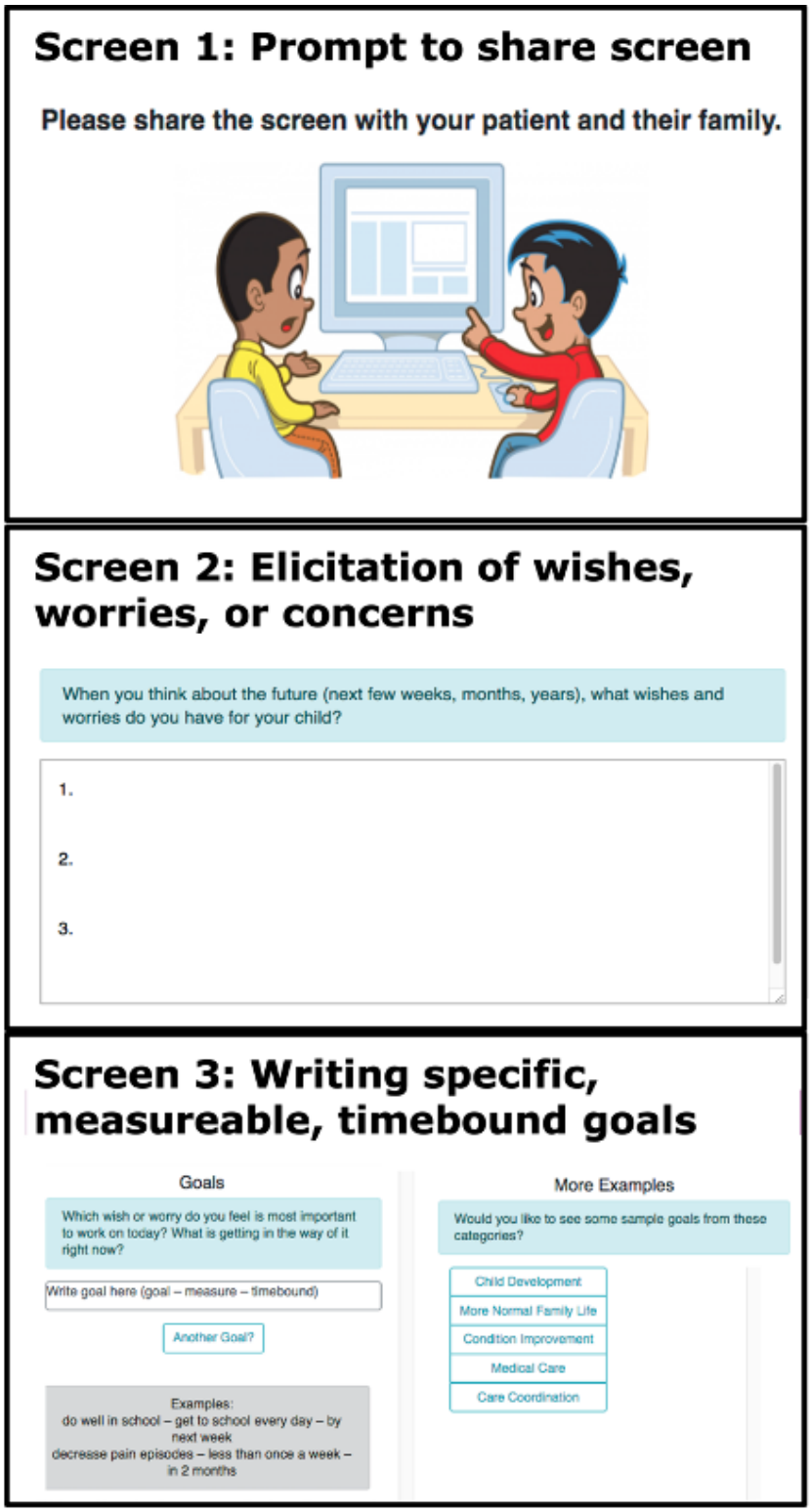
Intervention

We designed a technology-based family-centered goal-setting tool (GoalKeeper) using iterative feedback from parents and providers of children with medical complexity and informed by the challenges of technology-based tools elucidated in our prior work [15]. In the first phase of the design process, we interviewed parents and providers to develop a taxonomy of goals and effective approaches to discussing goals in an outpatient clinic visit, based on prior frameworks for goal-concordant, chronic-illness care [20-22]. In the second phase, we conducted iterative user testing with parents and providers initially via paper-based prototypes and subsequently a technology-based version. We chose to host the intervention on the internet and to function outside of the electronic health record (EHR) to allow for rapid iteration of the intervention design and ease of access to the intervention from any device (eg, computer, tablet, or smartphone). We designed GoalKeeper using responsive web technology that optimizes its use for any device. This functionality enables parents and providers to quickly input data and access GoalKeeper in any setting with internet independent of the type of device used.

GoalKeeper consists of 2 separate modules: (1) Goal Elicitation and (2) Tracking. The key module and the focus of this study is the goal elicitation module, which was designed to be used in real time in the exam room, facilitating a conversation between the parent and the medical provider during a routine, nonurgent care visit, consists of 3 components, as shown in Figure 1. The first component provides prompts to elicit the parent's "wishes/worries/concerns" about their child's health. The second component helps the parent and provider derive from these "wishes/worries/concerns" at least 1 goal that is specific, measurable, and time-bound to ensure that the goals are actionable and attainable. Specific actions were not entered into the module to reduce documentation burden based on feedback during design interviews and pretesting. The third component provides sample goals as inspiration based on a taxonomy of common children with medical complexity goals developed in the design process (Assis-Hassid S et al, unpublished data, 2021). These 3 components are synthesized on the final screen, with a "copy" button to facilitate easy transfer (pasting) of the text into the EHR. The tracking module was designed to be set up by a provider immediately after the clinical encounter. The purpose of the tracking module was to address parents' desires for symptoms tracking in the tool but was not the focus of the GoalKeeper design process given the prevalence of symptom tracking technology already in existence. The tracking module provides customizable templates that providers can assign to parents to track their progress toward the elicited goal(s). To allow for asynchronous work outside of clinic visits, parents and providers had distinct interfaces. Providers could create goals, create and assign tracking templates, and view data while parents could view set goals and view and input data into tracking templates. We placed goal elicitation in the provider interface, based on user feedback, to help flatten the parent-provider hierarchy in the clinic visit. GoalKeeper requires the medical provider to elicit and type the parent's stated goals during the clinic visit. Prior prototypes that required parents to type their goals resulted in increased parental stress and decreased medical provider engagement. In contrast to usual practice, medical providers welcomed the during-encounter goal elicitation feature as an opportunity to engage more meaningfully with parents. We optimized the tracking module for use on smartphones or tablets based on parent feedback. During the trial, further development was frozen to ensure all participants were exposed to the same version of the intervention.

XSL•FO
RenderX

Figure 1. GoalKeeper goal elicitation module.



The intervention implementation process was developed and evaluated using the Consolidated Framework for Implementation Research (CFIR) [23]. Prior to the start of the trial, we conducted stakeholder focus groups using the CFIR companion interview guide to inform implementation of GoalKeeper that resulted in the selection of 3 implementation strategies: educational materials, individual technical assistance, and automated reminders [24]. Educational materials included paper and electronic copies of an instructional manual on GoalKeeper, with parent and provider versions. Parents received a 5-minute in-person overview of GoalKeeper while providers received a 30-minute orientation, including a 5-minute educational video [25]. The study team was available for in-person and remote technical assistance but did not attend clinic visits with participants. Parents received automated weekly email reminders to track their progress on goals. Details on the implementation process and evaluation can be found in a companion study [26].

XSL•FO
RenderX

Participants in the intervention group were given GoalKeeper to use during their initial clinic visit immediately after enrolling in the study. They created a unique login and password and were emailed the webpage link to access GoalKeeper on their own device. Study team members also accessed the GoalKeeper webpage on the clinic room computer immediately prior to the start of the participant's clinic visit. Participants in the control group received usual clinical care with their provider.

Measures and Outcomes

Our randomized controlled trial is a hybrid type I effectiveness-implementation trial with a primary focus on assessing the effectiveness of the intervention but included collection of preliminary data on barriers and facilitators to implementation, reported in a separate manuscript [26,27]. Measures of primary and secondary outcomes were collected electronically immediately after the initial clinical encounter for both parents and providers following study enrollment (T1) and at 1-month post-encounter (T2). Additional implementation outcome survey measures and exit interviews were collected at 3 months post-encounter (T3) for parents and, for providers, at the end of the study period for their last enrolled parent (Figure S2 in Multimedia Appendix 1). All measures were collected via secure electronic survey tools and stored electronically via the Stanford REDCap (Research Electronic Data Capture).

The primary outcome was the difference in parent-reported quality of goal-setting between the control and intervention groups from T1 to T2, as measured by adapted questions from the goal-setting domain of the Patients Assessment of Chronic Illness Care (PACIC), a 20-item (5-point Likert scale: no/never to yes/always) validated survey measure of chronic illness care quality based on the Chronic Care Model [28]. A mean composite score is calculated for each domain and the entire survey based on relevant items. The goal-setting domain contains 5 items that cover healthy behaviors, treatment plans, and community integration. We adapted this measure for children with medical complexity because of its wide use in assessing chronic illness care quality and the lack of alternate, validated, pediatric-specific measures. Secondary outcomes measure included the PACIC subdomains for problem-solving, care coordination, and composite outcome of care quality based on all measured PACIC subdomains.

Additional secondary outcome measures included quality of shared decision-making, quality of parent-provider communication, and parenting stress. We assessed quality of shared decision-making (T1 and T2) using the National Survey for Children with Special Health Care Needs 2009 - 2010 shared decision-making domain, which creates a binary composite variable for high versus low quality shared decision-making from 4 survey questions at T1 and T2 [29]. We assessed the quality of parent-provider communication using the Communication Assessment Tool (T1), a validated 15-item (5-point Likert scale: 1-poor to 5-excellent) survey measure with mean composite outcome [30]. We assessed parenting stress using the Parenting Distress domain of the Parenting Stress Index (T1), which captures parent stress levels related to partner conflicts, social support, and life restrictions due to child rearing [31]. Scores are presented as a sum of all answers (range 12 - 60) with higher scores indicating higher stress.

Intervention Fidelity

We assessed intervention fidelity and quality in 2 ways: (1) capture of information directly from the GoalKeeper platform and (2) participant interviews at T3. Within the app, we primarily assessed number of goals set per intervention group participant but also assessed frequency of app use, and counts of data entered into each field. Through semistructured interviews, we assessed barriers and facilitators to implementation using questions from the CFIR companion interview guide (Multimedia Appendix 2), detailed in a separate manuscript [26,32].

Sample Size

We aimed to identify a moderate change in the PACIC domain of goal-setting from T1 to T2. Prior studies of the PACIC have shown in the domain of goal-setting a mean of 2.43 out of 5 (SD 1.1) [33]. We anticipated at least a δ of 0.5 (medium effect size) which has been achieved in prior studies comparing different health care practices in Switzerland [33]. With a projected attrition rate of 33%, our targeted recruitment was 60 parents (control: n=30; intervention n=30) to reach a minimum sample size of 40 parents who completed the study (control: n=20; intervention: n=20), which provides 80% power using a 2-sided comparison of α .05 to detect a δ of 0.5.

Statistical Analysis

For primary and secondary outcome measures, we used intention-to-treat analysis with repeated measures ANOVA to evaluate for between-group and within-group differences in the PACIC domain of goal-setting for intervention group versus control group while adjusting for fixed effects—timing of measurement (1: immediately after the visit and 2: 1 month after the visit) and intervention-by-time interaction—as well as random effects: provider cluster. We used a similar approach for our secondary outcomes. We conducted post hoc analyses of outcome variables with adjustment for population characteristics with statistically significant differences between subgroups. For baseline characteristics and feasibility outcome measures, we conducted descriptive analyses using Fisher exact test for binary and categorical variables and Wilcoxon sign-rank test for continuous variables. We used a statistical significance level of $P<.05$.

Ethical Considerations

This study was approved by the Stanford Single Institutional Review Board (Protocol 32161, approved 12/22/2016) and registered on clinicaltrials.gov (NCT03620071). All participants gave documented informed consent prior to enrolling in the study. All potential participants had the option to opt out of the study if they were not interested. Provider participants were approached at routine division meetings when the study team presented the study, reviewed the consent form, and gave ample time for questions. Providers could provide consent during or after the team meetings. Potential parent participants were approached in-person in the waiting room for consent and enrollment prior to their clinic visit with a participating provider. We did not use remote recruitment due to concerns of

XSL•FO
RenderX

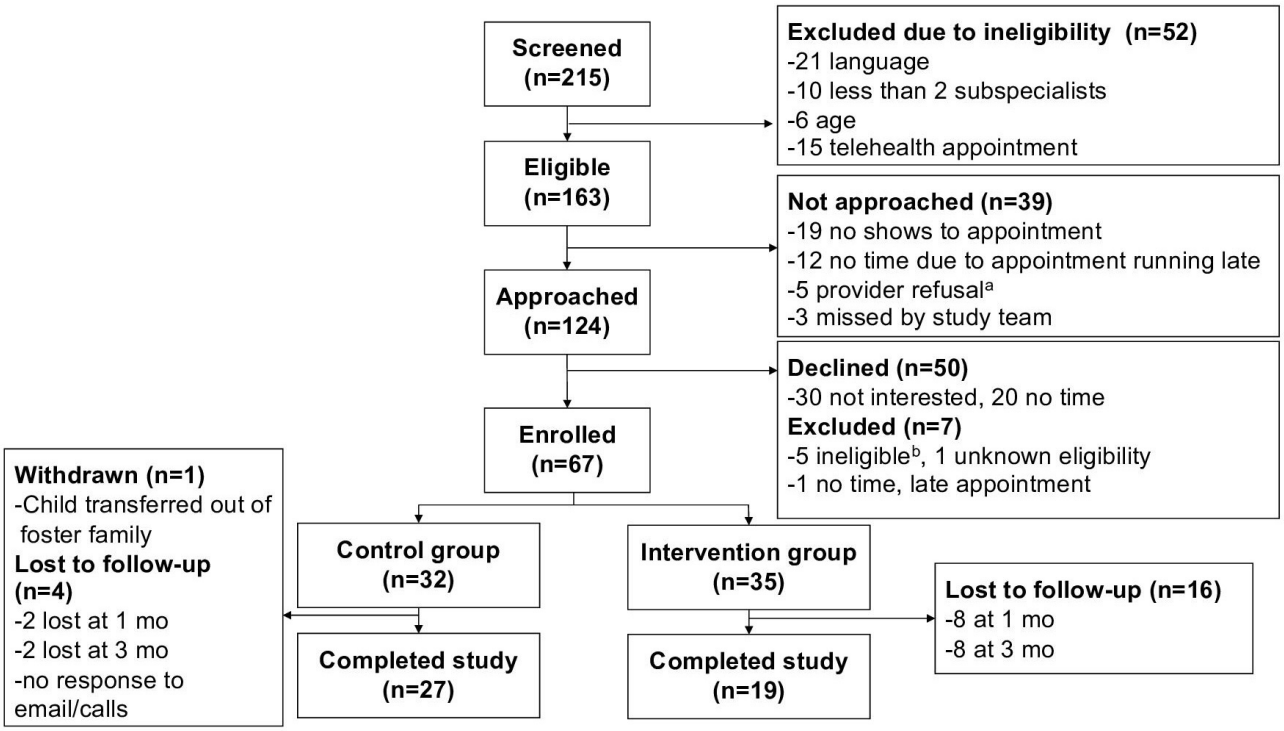
intervention fidelity as GoalKeeper was designed to be used jointly with parents and providers during a clinic visit. Study team members presented the study, reviewed the consent form, and gave ample time for questions. Consent was obtained prior to the start of the clinic visit. Each participant was assigned a randomly generated 4-digit unique identifier. The key linking the unique identifier to participant name and contact information was stored in a separate database. Parent participants received gift card remuneration after surveys at t1 and t2 (US $10 each) and interviews at t3 (US $30). Providers received a US $5 gift card remuneration per completed survey and after interview completion totaling US $50 for the entire study period.

Results

Study Population

We enrolled 67 parent-child dyads out of 124 eligible parents approached across 2 outpatient clinics and 11 medical providers. Of parent-child dyads, 32 were enrolled in the control group and 35 were enrolled in the intervention group. During the study period, 10 parent-child dyads (control: n=2; intervention: n=8) were lost to follow-up at T2, and 1 control group parent who was fostering a child with medical complexity was withdrawn due to the child being transferred out of their custody. We also had 2 providers lost to follow-up due to leaving the institution (CONSORT [Consolidated Standards of Reporting Trials] diagram in Figure 2).

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flowchart. [a]Provider rationale for refusal: parent ineligible due to underlying neurocognitive or behavioral health issue, under isolation protocol for suspected COVID-19, 1 emergent health care concern, visit changed to telehealth, patient late and had to see their non-primary provider instead. [b]Participants ineligible at visit due to: no internet or email (1), language (1), legal (1), accompanied by non-guardian/non-primary caregiver (2).



Among parent participants, the mean age was 37.4 (SD 8.2) years, 50 (76%) were female participants, 32/66 (48%) identified as Caucasian, and 25 (38%) as Hispanic. Most were highly educated, with 51 (77%) having completed at least some college. Patients had a mean age of 5.6 (SD 0.5) years and 24 (36%) were female participants. All saw multiple subspecialties (median 3 [IQR 2‐4]). Most had technology dependence (34/66, 52%) and neurodevelopmental delay (36/66, 55%). Only the ventriculoperitoneal shunt was statistically significantly different between the control (0) and intervention (4) groups. The providers were mostly medical doctors (9/11, 82%), female participants (7/11, 64%), and neurologists (6/11, 54%). Parent, child, and provider characteristics are detailed in Table 1, Table S3 in Multimedia Appendix 1, and Table S4 in Multimedia Appendix 1, respectively.

XSL•FO
RenderX

Table . Population characteristics.

Characteristic	Total, n=66	Control, n=31	Intervention, n=35	*P* value[a]
Age, mean (SD)	37.4 (8.2)	36.4 (8.7)	38.3 (7.8)	.35
Sex, n (%)				.4
Female	50 (76)	22 (71)	28 (80)	
Race/ethnicity[b], n (%)				.38
African American	2 (3)	0 (0)	2 (6)	
American Indian	2 (3)	1 (3)	1 (3)	
Asian	7 (11)	4 (13)	3 (9)	
Caucasian	32 (48)	15 (48)	17 (49)	
Hispanic	25 (38)	11 (35)	14 (40)	
Other	6 (9)	5 (16)	1 (3)	
Insurance, n (%)				.26
Public	30 (45)	11 (35)	19 (54)	
Private	31 (47)	13 (42)	18 (51)	
Other	20 (30)	13 (42)	7 (20)	
Don’t know	3 (5)	1 (3)	2 (6)	
Household size, mean (SD)	4.2 (1.3)	4.2 (1.6)	4.1 (1.0)	.76
Marital status, n (%)				.88
Single	17 (26)	7 (23)	10 (29)	
Married/living with partner	46 (69)	22 (71)	24 (69)	
Education level, n (%)				.63
Some high school	3 (2)	1 (3)	2 (6)	
High school diploma	10 (15)	6 (19)	4 (11)	
Some college	15 (23)	5 (16)	10 (29)	
College degree	22 (33)	10 (32)	12 (34)	
Advanced degree	14 (21)	7 (23)	7 (20)	

[a]Statistical significance level is $P \leq .05$.

[b]Participants could select all categories that apply.

For our primary outcome measure, quality of goal-setting, scores were lower for the intervention group at t1 when compared to the control group, 2.41 (SD 1.07) versus 2.57 (SD 1.06). At 1-month follow-up (t2), the reverse was true with scores of 2.36 (SD 1.12) for the intervention group versus 2.15 (SD 1.23) for the control group. Quality of goal-setting between t1 and t2 remained stable for the intervention group but decreased for the control group (δ=0.03 vs −0.43, F_{49}=3.52, P=.06). Similar patterns exist for other PACIC outcome measures of problem-solving, care coordination, and overall care quality (composite). For other secondary outcome measures, parenting distress at t1 was higher in the intervention group when compared to the control group (mean [SD] 27.9 [11.6] vs 22.2 [7.6]; P=.02). There were no significant group differences in parent report of quality of shared decision-making or quality of doctor-patient communication (Table 2 and Figure 3). For post hoc analyses of ventriculoperitoneal shunt, the sample size of 4 was too small for valid statistical analysis.

XSL•FO
RenderX

Table . Efficacy outcomes.

Outcome	T1		T2		*F* test (*df*)[a]	*P* value[b]
	Control, n=30	Intervention, n=35	Control, n=27	Intervention, n=27		
Quality of chronic illness care (PACIC[c]), mean (SD)						
Composite	2.62 (1.03)	2.32 (1.02)	2.12 (1.23)	2.29 (1.15)	4.28 (49)	.04
Goal-setting	2.57 (1.06)	2.41 (1.07)	2.15 (1.23)	2.36 (1.12)	3.52 (49)	.06
Problem-solving	3.03 (1.16)	2.55 (1.17)	2.55 (1.38)	2.63 (1.20)	3.53 (49)	.07
Coordination	2.27 (1.14)	2.00 (1.01)	1.67 (1.30)	1.89 (1.27)	3.33 (49)	.08
High-quality shared decision-making, n (%)	22 (73)	24 (69)	20 (74)	19 (70)	0.15 (49)	.70
Quality of communication (Communication Assessment Tool), mean (SD)	4.66 (0.78)	4.60 (0.52)	—[d]	—	—	.71
Parenting distress (Parenting Stress Index), mean (SD)	22.23 (7.66)	27.89 (11.61)	—	—	—	.02

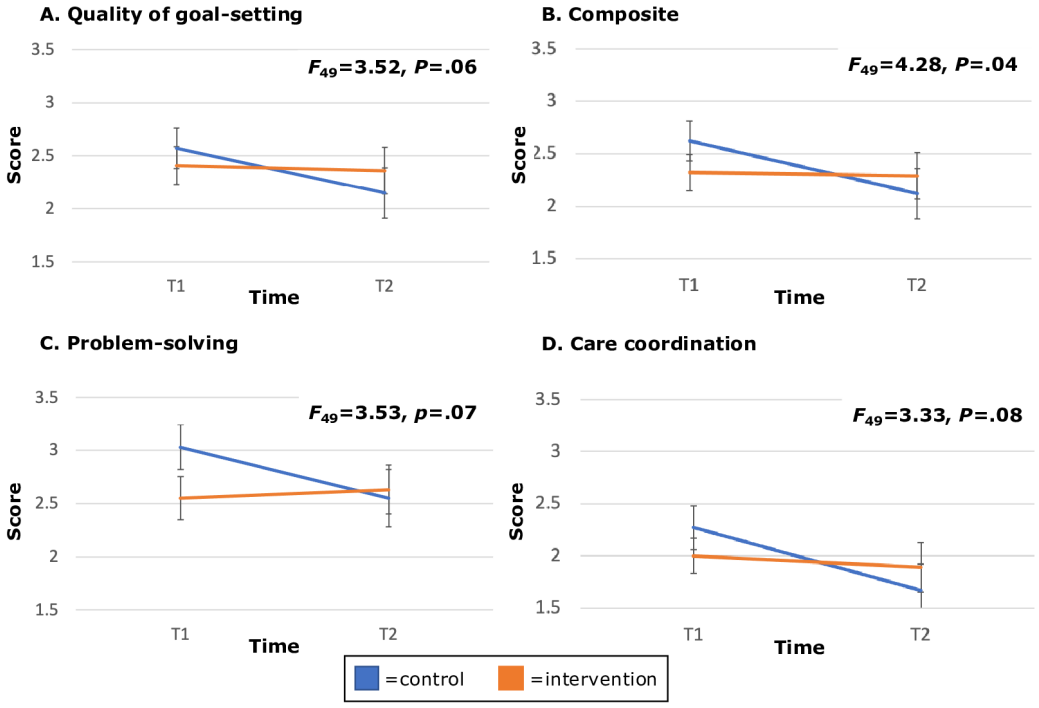
[a]*F* reported is for intervention by time.

[b]Statistical significance level is $P \leq .05$.

[c]PACIC: Patient Assessment of Care for Chronic Illness Care.

[d]not available.

Figure 3. *F* reported is for intervention by time.



Intervention Fidelity

Among the intervention group, 34 of 35 parents successfully used GoalKeeper with their provider to set goals during the clinic visit. For 1 parent, the provider decided during the visit that it was not an appropriate time to use GoalKeeper, but to preserve an intention-to-treat approach, that parent remained in the intervention group to complete the survey measures. Parents and providers set a median of 2.5 (IQR 1.5-3) goals per initial clinic visit, and providers assigned a median of 3 (IQR 3-4) tracking templates to each parent. Each provider saw a median of 2.5 (IQR 2-5) parent participants . After the initial clinic visit,

XSL•FO
RenderX

parents input information into GoalKeeper a median of 0 (IQR 0-3) times and providers viewed tracked data a median of 0.5 (IQR 0-1) times throughout the study period. A total of 19 parents (CPCC: n=8; neurology: n=11) did not enter any data into the tracking templates, while 13 parents (CPCC: n=4; neurology: n=9) entered data 1 - 10 times, and 3 parents, all from CPCC, entered data over 10 times. In total, 87 goals were set with the most common domains being development (parents: n=10; unique goals: n=16), child nutrition (parents: n=13; unique goals: n=13), and seizure control (parents: n=10; unique goals: n=10). There were no statistically significant differences in participant and patient characteristics between those who used the tracking module more extensively than those who did not. Of the 19 parents who completed the exit survey, 13 (68%) parents would recommend GoalKeeper to other parents. A total of 6 (67%) providers would recommend GoalKeeper to other providers (Table S3 in Multimedia Appendix 1).

In exit interviews, participants identified facilitators to GoalKeeper use to be the shared nature of the goal-setting activity, the tool being accessible by the internet, and the ability to set long-term goals as facilitators to tool use. Per parent participants, the main barrier for sustained tool use after the initial goal-setting clinic visit was lack of integration in the EHR and patient portal. Additional details about the interview results and the tool's feasibility and acceptability are reported in a companion manuscript [26].

Discussion

Principal Results

Our study demonstrated that the GoalKeeper intervention, a novel goal-setting platform for outpatient care of complex chronic illness, resulted in a better maintenance of overall chronic illness care quality, with no statistically significant impact on goal-setting, problem-solving, or coordination. Intervention fidelity results indicate that the sustained effect of GoalKeeper on care quality was likely the result of participant engagement with the intervention's goal elicitation module, in contrast to its tracking module, which had very low participant engagement. Furthermore, we found that the intervention is both feasible and acceptable in the context of busy outpatient care, in both primary and subspecialty care settings. Together, these results suggest that technology supports for goal-setting may be an effective way to maintain high-quality care in between clinic visits for children with medical complexity, but GoalKeeper itself needs further refinement to improve user experience. However, given the nonstatistically significant findings in individual PACIC domains, the exact mechanism for this remains unclear.

The sustained quality of care for those in the intervention group is notable compared to the decline in quality of care for the control group, even though the initial quality of care scores (PACIC composite and domain) were lower for the intervention group than the controls. One potential explanation for these initial lower scores may be that the parents in the intervention group were more stressed as indicated by the parenting distress scores. Additionally, the COVID-19 pandemic may have contributed to greater parent stress in the intervention group, since the control group was enrolled earlier in the trial, although this was not a theme found in the exit interviews. Since we collected baseline data after the encounter, another explanation may be a learning curve that participants experienced due to changing the dynamic of a clinic visit through the use of a new tool (GoalKeeper), which may cause the visit to seem of poorer quality due to the challenge of learning a new approach to care [34]. A final possible explanation is simple random variation, which justifies our choice of a randomized controlled trial design to examine intervention effects. Based on our observed participant and patient characteristics, our randomization approach appears strong and thus makes it unlikely that care quality scores were lower due to higher medical complexity or disparities in care delivery.

Comparison With Prior Work

Nonetheless, given the nonstatistically significant findings in individual PACIC domains, further research is necessary to extend and enhance the effectiveness of such platforms on care quality. Our results build on prior studies suggesting tools to improve care for children with medical complexity need to elicit and track family-centered goals [9,35]. Family engagement remains an urgent research priority for children with medical complexity, yet their providers report barriers including insufficient communication skills and supports [36,37]. Providers in our study acknowledged that they, too, struggled with the communication around goals with parents of children with medical complexity but endorsed that GoalKeeper facilitated better communication about parents' goals. Multidisciplinary care coordination programs that included family-centered goal-setting found that family-centered goals helped programs emphasize family priorities [11]. However, many approaches to family-centered goal-setting have used resource-intensive approaches, which face barriers to implementation including the scalability and replicability of programs across multiple centers. Our intervention may enable these programs to extend their reach by acting as an adjunct intervention in clinical care whereby family-centered goal-setting can be achieved without additional extended team support. Future work on tracking and adapting the goals longitudinally could further sustain engagement and lead to evaluations of how long-term care is affected by family-centered goals.

While our study demonstrates high initial adoption of GoalKeeper based on goal elicitation module use, future studies should investigate the mechanisms of effect and tools to sustain engagement, such as the effects of a more sophisticated tracking module. Other digital health interventions have initial high levels of adoption that decrease dramatically over time [38]. Prior studies have demonstrated that self-monitoring interventions may result in sustained engagement even if the initial motivation for using the app was different [39]. In our study, the design of the tracking module was not the focus of intervention development. Since other existing tools to improve care for children with medical complexity focus on symptom tracking, we wanted to avoid duplicative efforts [40,41]. Future directions should explore whether integration of the Goal Elicitation module with existing symptom trackers will improve sustained use of GoalKeeper. Sustained use of digital health

XSL•FO
RenderX

interventions after the initial adoption period has also been associated with user perceptions of usefulness, ease of use, fulfillment of user expectations, and satisfaction with the app [42]. Future evaluations should test whether these mechanisms can sustain use of our intervention in larger populations and in populations that prefer other languages.

Study findings also suggest additional design and implementation considerations for health-information technology platforms that aim to improve care for chronic conditions. Goal elicitation structures, similar to those developed for GoalKeeper, should be embedded across discrete elements of the EHR—especially the encounter note, the after-visit summary, and the patient portal. Additionally, goal-elicitation training should be integrated with provider-training modules, including simulated patient encounters, to enhance provider practice. While standalone tools could improve accessibility in community-based care, this study's findings suggest high feasibility and acceptability for integrating adaptive, chronic-illness management technology into existing workflows.

Limitations

Study findings should be interpreted in the context of a few limitations. This study was conducted in a single children's health system, which may not be reflective of the practice at other institutions caring for children with medical complexity. However, the CPCC was part of a multicenter collaboration to develop care coordination programs for children with medical complexity, so some program similarities may exist across these collaborating sites. More participants in the intervention group were lost to follow-up, which may have introduced selection bias, with those with poorer experiences with the intervention or with higher medical needs choosing not to continue to participate in the study. However, there were no statistically significant differences in parent and child characteristics between those lost to follow-up and those who completed the study. Notably, participants had to have home internet access, and while the divide between internet access is narrowing, this could have excluded populations with lower household income [43]. The majority of our study participants also had at least some college education, which is reflective of the local population. Overall, our study period was quite short. Thus, the long-term intervention use, particularly on repeat clinic visits, was not observed. The effects of long-term intervention use should be assessed with future studies.

Conclusions

Technologies that support real-time, shared goal-setting between a patient and their medical provider may serve as an adjunct to high-quality care of complex chronic illness. Informed by in-depth qualitative research and understanding of children with medical complexity-specific care needs and goals, health-technology tools like GoalKeeper are likely to be feasible and acceptable when integrated into both primary care and subspecialty care outpatient settings. Furthermore, in the context of learning health systems and emerging, generative artificial intelligence solutions, such platforms are likely to become easier to implement, scale, and study across care systems. As a result, goal-concordant care technologies, like Goalkeeper, co-designed with patients and caregivers in the context of a rigorous behavioral science framework, hold promise to improve care outcomes across a wide range of chronic and comorbid conditions.

Acknowledgments

The authors thank the patients, parents, and providers of children with medical complexity for their participation in study activities. The authors did not use generative AI for any part of this study or manuscript.

Funding

This work was supported by the National Institutes of Health, National Cancer Institute grant number R01CA204585. Dr. Lin received support from the National Institutes of Health, National Institute of Arthritis and Musculoskeletal and Skin Diseases 1K23AR079778. The funding sources had no involvement in the research activities related to this work.

Data Availability

The datasets generated or analyzed during this study are available from the senior author on reasonable request.

Authors' Contributions

JL conceptualized and designed the study, selected data inputs, assisted in interpretation of results, drafted the initial manuscript, and reviewed and revised the manuscript.
BH, OA, KG, and BG assisted in the design of the study, selected data inputs, assisted in interpretation of results, and reviewed and revised the manuscript.
SA-H and SG assisted in the design of the study and reviewed and revised the manuscript.
LS conceptualized and designed the study, selected data inputs, assisted in interpretation of results, and reviewed and revised the manuscript.
All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Supplemental figures and tables.
[DOCX File, 56 KB - humanfactors_v13i1e70402_app1.docx]

Multimedia Appendix 2
Interview guides.
[DOCX File, 22 KB - humanfactors_v13i1e70402_app2.docx]

Checklist 1
CONSORT (Consolidated Standards of Reporting Trials)-EHEALTH checklist.
[PDF File, 492 KB - humanfactors_v13i1e70402_app3.pdf]

References

1. Neff JM, Sharp VL, Muldoon J, Graham J, Myers K. Profile of medical charges for children by health status group and severity level in a Washington state health plan. Health Serv Res 2004 Feb;39(1):73-89. [doi: 10.1111/j.1475-6773.2004.00216.x] [Medline: 14965078]
2. Simon TD, Berry J, Feudtner C, et al. Children with complex chronic conditions in inpatient hospital settings in the United States. Pediatrics 2010 Oct;126(4):647-655. [doi: 10.1542/peds.2009-3266] [Medline: 20855394]
3. Cohen E, Kuo DZ, Agrawal R, et al. Children with medical complexity: an emerging population for clinical and research initiatives. Pediatrics 2011 Mar;127(3):529-538. [doi: 10.1542/peds.2010-0910] [Medline: 21339266]
4. Cohen E, Berry JG, Camacho X, Anderson G, Wodchis W, Guttmann A. Patterns and costs of health care use of children with medical complexity. Pediatrics 2012 Dec;130(6):e1463-e1470. [doi: 10.1542/peds.2012-0175] [Medline: 23184117]
5. Strickland BB, Jones JR, Newacheck PW, Bethell CD, Blumberg SJ, Kogan MD. Assessing systems quality in a changing health care environment: the 2009-10 National Survey of Children with Special Health Care Needs. Matern Child Health J 2015 Feb;19(2):353-361. [doi: 10.1007/s10995-014-1517-9] [Medline: 24912943]
6. Kuo DZ, Goudie A, Cohen E, et al. Inequities in health care needs for children with medical complexity. Health Aff (Millwood) 2014 Dec;33(12):2190-2198. [doi: 10.1377/hlthaff.2014.0273] [Medline: 25489038]
7. Lin JL, Cohen E, Sanders LM. Shared decision making among children with medical complexity: results from a population-based survey. J Pediatr 2018 Jan;192:216-222. [doi: 10.1016/j.jpeds.2017.09.001] [Medline: 29102046]
8. Council on Children with Disabilities and Medical Home Implementation Project Advisory Committee. Patient- and family-centered care coordination: a framework for integrating care for children and youth across multiple systems. Pediatrics 2014 May;133(5):e1451-e1460. [doi: 10.1542/peds.2014-0318] [Medline: 24777209]
9. McAllister JW. Achieving a shared plan of care with children and youth with special health care needs. : Lucille Packard Foundation for Children's Health; 2014 URL: https://lpfch.org/wp-content/uploads/2024/02/achieving_a_shared_plan_of_care_full.pdf [accessed 2026-01-28]
10. Kuo DZ, Berry JG, Glader L, Morin MJ, Johaningsmeir S, Gordon J. Health services and health care needs fulfilled by structured clinical programs for children with medical complexity. J Pediatr 2016 Feb;169:291-296.e1. [doi: 10.1016/j.jpeds.2015.10.010] [Medline: 26526361]
11. McAllister JW, Keehn RM, Rodgers R, Lock TM. Care coordination using a shared plan of care approach: from model to practice. J Pediatr Nurs 2018;43:88-96. [doi: 10.1016/j.pedn.2018.09.009] [Medline: 30473161]
12. Prahalad P, Tanenbaum M, Hood K, Maahs DM. Diabetes technology: improving care, improving patient-reported outcomes and preventing complications in young people with type 1 diabetes. Diabet Med 2018 Apr;35(4):419-429. [doi: 10.1111/dme.13588] [Medline: 29356074]
13. Nkoy FL, Fassl BA, Wilkins VL, et al. Ambulatory management of childhood asthma using a novel self-management application. Pediatrics 2019 Jun;143(6):e20181711. [doi: 10.1542/peds.2018-1711] [Medline: 31097465]
14. Fiks AG, Mayne SL, Karavite DJ, et al. Parent-reported outcomes of a shared decision-making portal in asthma: a practice-based RCT. Pediatrics 2015 Apr;135(4):e965-e973. [doi: 10.1542/peds.2014-3167] [Medline: 25755233]
15. Amir O, et al. From care plans to care coordination: opportunities for computer support of teamwork in complex healthcare. Presented at: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI 2015); Apr 18-23, 2015. [doi: 10.1145/2702123.2702320]
16. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. URL: https://www.R-project.org/ [accessed 2026-01-28]
17. Mdege ND, Man MS, Taylor Nee Brown CA, Torgerson DJ. Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. J Clin Epidemiol 2011 Sep;64(9):936-948. [doi: 10.1016/j.jclinepi.2010.12.003] [Medline: 21411284]
18. Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. BMC Med Res Methodol 2006 Nov 8;6:54. [doi: 10.1186/1471-2288-6-54] [Medline: 17092344]

XSL•FO
RenderX

19. Garth B, Murphy GC, Reddihough DS. Perceptions of participation: child patients with a disability in the doctor-parent-child partnership. Patient Educ Couns 2009 Jan;74(1):45-52. [doi: 10.1016/j.pec.2008.07.031] [Medline: 18760558]
20. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. JAMA 2002 Nov 20;288(19):2469-2475. [doi: 10.1001/jama.288.19.2469] [Medline: 12435261]
21. Eiser C, Morse R. A review of measures of quality of life for children with chronic illness. Arch Dis Child 2001 Mar;84(3):205-211. [doi: 10.1136/adc.84.3.205] [Medline: 11207164]
22. Wagner EH, Austin BT, Davis C, Hindmarsh M, Schaefer J, Bonomi A. Improving chronic illness care: translating evidence into action. Health Aff (Millwood) 2001;20(6):64-78. [doi: 10.1377/hlthaff.20.6.64] [Medline: 11816692]
23. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 7;4:50. [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
24. Powell BJ, Waltz TJ, Chinman MJ, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb 12;10:21. [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
25. GoalKeeper. AIM Lab Stanford vimeo page. 2018 Dec 16. URL: https://vimeo.com/306581091/2641a2e69b [accessed 2026-02-10]
26. Lin JL, Huber B, Amir O, et al. Barriers and facilitators to the implementation of family-centered technology in complex care: feasibility study. J Med Internet Res 2022 Aug 23;24(8):e30902. [doi: 10.2196/30902] [Medline: 35998021]
27. Landes SJ, McBain SA, Curran GM. An introduction to effectiveness-implementation hybrid designs. Psychiatry Res 2019 Oct;280:112513. [doi: 10.1016/j.psychres.2019.112513] [Medline: 31434011]
28. Glasgow RE, Wagner EH, Schaefer J, Mahoney LD, Reid RJ, Greene SM. Development and validation of the Patient Assessment of Chronic Illness Care (PACIC). Med Care 2005 May;43(5):436-444. [doi: 10.1097/01.mlr.0000160375.47920.8c] [Medline: 15838407]
29. 2009-2010 National Survey of Children with Special Health Care Needs. : Child and Adolescent Health Measurement Initiative, Data Resource Center for Child and Adolescent Health; 2012 URL: https://www.childhealthdata.org/learn-about-the-nsch/archive-prior-year-data-documents-and-resources [accessed 2026-01-28]
30. Makoul G, Krupat E, Chang CH. Measuring patient views of physician communication skills: development and testing of the communication assessment tool. Patient Educ Couns 2007 Aug;67(3):333-342. [doi: 10.1016/j.pec.2007.05.005] [Medline: 17574367]
31. Lee SJ, Gopalan G, Harrington D. Validation of the Parenting Stress Index-short form with minority caregivers. Res Soc Work Pract 2016 Jul;26(4):429-440. [doi: 10.1177/1049731514554854] [Medline: 27616868]
32. Qualitative data. The Consolidated Framework for Implementation Research (CFIR). URL: https://cfirguide.org/evaluation-design/qualitative-data [accessed 2026-01-28]
33. Frei A, Senn O, Huber F, et al. Congruency of diabetes care with the chronic care model in different Swiss health care organisations from the patients' perspective: a cross sectional study. Swiss Med Wkly 2014;144:w13992. [doi: 10.4414/smw.2014.13992] [Medline: 25207723]
34. Govindarajulu US, Stillo M, Goldfarb D, Matheny ME, Resnic FS. Learning curve estimation in medical devices and procedures: hierarchical modeling. Stat Med 2017 Jul 30;36(17):2764-2785. [doi: 10.1002/sim.7309] [Medline: 28470678]
35. Guerin S, Kiernan G, Courtney E, McQuillan R, Ryan K. Integration of palliative care in services for children with life-limiting neurodevelopmental disabilities and their families: a Delphi study. BMC Health Serv Res 2020 Oct 8;20(1):927. [doi: 10.1186/s12913-020-05754-w] [Medline: 33032605]
36. Coller RJ, Berry JG, Kuo DZ, et al. Health system research priorities for children and youth with special health care needs. Pediatrics 2020 Mar;145(3):e20190673. [doi: 10.1542/peds.2019-0673] [Medline: 32024751]
37. Tschudy MM, Raphael JL, Nehal US, O'Connor KG, Kowalkowski M, Stille CJ. Barriers to care coordination and medical home implementation. Pediatrics 2016 Sep;138(3):e20153458. [doi: 10.1542/peds.2015-3458] [Medline: 27507894]
38. Ancker JS, Barrón Y, Rockoff ML, et al. Use of an electronic patient portal among disadvantaged populations. J Gen Intern Med 2011 Oct;26(10):1117-1123. [doi: 10.1007/s11606-011-1749-y] [Medline: 21647748]
39. Lee K, Kwon H, Lee B, et al. Effect of self-monitoring on long-term patient engagement with mobile health applications. PLoS One 2018;13(7):e0201166. [doi: 10.1371/journal.pone.0201166] [Medline: 30048546]
40. Nkoy FL, Hofmann MG, Stone BL, et al. Information needs for designing a home monitoring system for children with medical complexity. Int J Med Inform 2019 Feb;122:7-12. [doi: 10.1016/j.ijmedinf.2018.11.011] [Medline: 30623786]
41. Desai AD, Wang G, Wignall J, et al. User-centered design of a longitudinal care plan for children with medical complexity. J Am Med Inform Assoc 2020 Dec 9;27(12):1860-1870. [doi: 10.1093/jamia/ocaa193] [Medline: 33043368]
42. Cho J. The impact of post-adoption beliefs on the continued use of health apps. Int J Med Inform 2016 Mar;87:75-83. [doi: 10.1016/j.ijmedinf.2015.12.016] [Medline: 26806714]
43. Internet, broadband fact sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/internet-broadband [accessed 2026-01-28]

XSL•FO
RenderX

Abbreviations:

CFIR: Consolidated Framework for Implementation Research
CONSORT: Consolidated Standards of Reporting Trials
CPCC: complex primary care clinics
EHR: electronic health record
PACIC: Patient Assessment of Care for Chronic Illness Care
REDCap: Research Electronic Data Capture

Edited by A Kushniruk; submitted 20.Dec.2024; peer-reviewed by C Coleman, D Poddighe; revised version received 17.Dec.2025; accepted 03.Jan.2026; published 20.Feb.2026.

Please cite as:
Lin J, Huber B, Amir O, Assis-Hassid S, Gehrmann S, Gajos K, Grosz B, Sanders L
Novel Web-Based Technology to Promote Goal-Setting in Complex Chronic Illness: Randomized Controlled Trial
JMIR Hum Factors 2026;13:e70402
URL: https://humanfactors.jmir.org/2026/1/e70402
doi:10.2196/70402

© Jody Lin, Bernd Huber, Ofra Amir, Shiri Assis-Hassid, Sebastian Gehrmann, Krzysztof Gajos, Barbara Grosz, Lee Sanders. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 20.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Digital Health Literacy and Tool Adoption in Postoperative Care in a Safety-Net Hospital Population: Mixed Methods Study

Christopher Awad[1], MBA, MD; Alexander Jakub Martinek[2], BS; Chunhao Zou[3], MS; Rosalind Byrd[4], MPH, MD; Hui Jean Ooi[4], MD; Kimberly Do[5], BS; Henry Young[1], MD; Bhavin Adhyaru[4], MD; Muhammed Idris[2], PhD; Rosa I Arriaga[3], PhD

[1]Department of Emergency Medicine, The Ohio State University Wexner Medical Center, Columbus, OH, United States

[2]Clinical Research Center, Morehouse School of Medicine, 720 Westview Drive SW, Atlanta, GA, United States

[3]Department of Computer Science, Georgia Institute of Technology, Atlanta, GA, United States

[4]Division of General Internal Medicine, School of Medicine, Emory University, Atlanta, GA, United States

[5]College of Computing, Northeastern University, Boston, MA, United States

Corresponding Author:
Alexander Jakub Martinek, BS
Clinical Research Center, Morehouse School of Medicine, 720 Westview Drive SW, Atlanta, GA, United States

Abstract

Background: Digital health tools are increasingly prevalent in postoperative care management, yet limited research exists on digital health literacy and tool adoption among safety-net hospital populations. Understanding these factors is crucial for developing effective digital health solutions for historically underserved communities.

Objective: This study aimed to evaluate digital health literacy, assess technology adoption readiness, and examine the relationship between patient-reported capabilities and demographic factors in a postoperative care context at a safety-net hospital.

Methods: We conducted a mixed methods study with 71 postoperative patients and 29 health care providers at a safety-net hospital. Participants completed a modified eHealth Literacy Scale (eHEALS) assessment and a demographic questionnaire, followed by usability testing of PocketDoc, a digital health prototype. The modified 7-item eHEALS demonstrated adequate internal consistency (Cronbach α=0.77). Qualitative data from think-aloud protocols during usability testing were collected for future analysis. This study focused on quantitative assessments of digital health literacy (using the modified eHEALS on a 5-point Likert scale) and technology adoption readiness (via usability metrics on a 10-point Likert scale) analyzed using nonparametric statistical tests. Correlations between demographic factors and digital health literacy were examined using Spearman rank-order correlation.

Results: Despite common assumptions about technology barriers in safety-net populations, 69% (49/71) of patients reported high confidence (score of ≥3 on a 5-point scale) in finding health resources online, and 61% (43/71) expressed confidence in using the internet for health-related questions. However, only 49% (35/71) felt confident in using digital resources for health decision-making. Digital health literacy scores did not correlate with age or educational level, although 79% (56/71) of patients reported ≥10 years of digital device experience. Both patients and health care providers rated PocketDoc highly for ease of use (median 10, IQR 8-10) and task intuitiveness (median 10, IQR 8-10). Patients' confidence in finding and using health resources online positively correlated with interface satisfaction (ρ=0.262-0.304 and ρ=0.010-0.027, respectively).

Conclusions: Our exploratory findings from 100 participants suggest that digital health tools may be more feasible in safety-net settings than previously considered, although the sample size and single-site design limit generalizability. However, the gap between patients' ability to find health resources (49/71, 69% confident) and their confidence in using these resources for health decision-making (35/71, 49% confident) highlights the need for targeted support in translating digital capabilities to health management skills.

(JMIR Hum Factors 2026;13:e75496) doi:10.2196/75496

KEYWORDS

patient navigation; digital health; mobile health app; postsurgical care; user experience; eHealth literacy; patient self-management; health informatics; digital intervention; human-computer interaction; usability testing; health care technology acceptance

XSL•FO
RenderX

Introduction

Postoperative care increasingly relies on digital health tools to support patient recovery, manage symptoms, and facilitate physician-patient communication [1,2]. As health care systems transition toward digital solutions, understanding patients' ability to effectively engage with these tools becomes crucial, particularly in safety-net hospital settings, where digital health literacy barriers may impact care delivery and outcomes [3]. Digital health literacy, defined as "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem" [4], plays a vital role in the successful implementation of digital health tools [2,5].

There is evidence suggesting that digital tools are effective in improving disease outcomes and health literacy [6-9]. Mobile health apps in the postoperative period have been correlated with earlier discharge, reduction of in-person follow-ups [10,11], promotion of rehabilitation [12], earlier detection of surgical complications [13], and improved communication between patients and health care professionals [14,15]. Beyond these clinical benefits, postoperative monitoring apps have shown potential to empower patients, providing autonomy over their own health and potentially improving satisfaction and recovery motivation [16-18].

However, it remains unclear what factors make these digital tools superior to human health navigators or resources available through simple internet searches [19], particularly in postoperative settings [20]. Despite the proliferation of digital health tools in postoperative care [21], limited research exists on digital health literacy and technology adoption readiness among safety-net hospital populations [22,23]. While studies demonstrate the efficacy of mobile apps in improving postoperative outcomes [24], these benefits presume a baseline level of digital health literacy that may not be universal across all patient populations, particularly in historically underserved communities [25].

In safety-net hospital settings, health literacy is shaped by factors such as educational level, language, and socioeconomic status, with lower health literacy levels linked to poorer health outcomes, reduced treatment adherence, and greater reliance on emergency care. Digital literacy, while related, is a distinct construct that poses additional barriers—particularly for older adults, non–English speakers, and individuals from lower socioeconomic backgrounds, who often face limited internet access and lack the digital skills necessary for navigating health care tools and systems. These separate constructs highlight a need for tailored tools to improve both forms of literacy [26].

The assumption that patients in safety-net hospitals face substantial barriers to digital health tool adoption has shaped both the development and implementation of digital interventions [23]. However, this assumption may not fully reflect current realities as smartphone ownership and digital technology use have become increasingly prevalent across socioeconomic groups [27,28]. As digital health tools continue to expand in postoperative care, understanding the actual digital health literacy levels and technology adoption readiness of safety-net hospital patients is essential for designing effective and equitable solutions [29].

In this study, we distinguished between 2 related but distinct constructs. *Digital health literacy* refers to patients' perceived ability to seek, evaluate, and apply online health information, whereas *technology adoption readiness* reflects patients' ability and confidence to engage with a specific digital health tool. Differentiating these constructs is critical as baseline literacy does not necessarily translate into successful interaction with or adoption of digital interventions.

In this paper, we present PocketDoc, a digital health application designed to support postoperative patient care, and use its evaluation to examine digital health literacy and technology adoption readiness in a safety-net hospital setting. We pursued 2 interrelated objectives. The primary aim was to assess digital health literacy levels among postoperative patients and health care providers. The secondary aim was to evaluate the usability of PocketDoc to determine whether a user-centered digital tool could be accessible to populations traditionally considered vulnerable to the digital divide.

These objectives are intentionally linked. Baseline digital health literacy provides essential context for interpreting usability findings, whereas usability outcomes inform whether patients are likely to adopt and meaningfully engage with digital health resources. By examining both constructs through a mixed methods approach—combining a modified eHealth Literacy Scale (eHEALS) assessment [30] with usability testing of a digital health prototype—we aimed to challenge prevailing assumptions about digital health literacy barriers and identify opportunities for more effective postoperative digital health tool implementation.

Methods

Study Design and Setting

This research was conducted at a large safety-net hospital in a Southern US state. We used a mixed methods approach combining quantitative assessments of digital health literacy with a qualitative usability testing of a digital health prototype.

Study Population and Recruitment

Participants included 71 postoperative patients and 29 health care providers. Study inclusion criteria required participants to be adults (aged ≥18 years) who had had surgery within the previous year, were caretakers of someone who had had surgery in the previous year, or provided clinical care in the postsurgical period. Participants were not excluded based on health status, gender, race, or socioeconomic status, although patients with cognitive deficits or functional impairments preventing the use of digital devices were not recruited.

Digital Health Literacy Assessment

Participants completed a modified version of the eHEALS, a validated instrument for measuring individuals' combined knowledge, comfort, and perceived skills regarding finding, evaluating, and applying online health information [30]. The modified eHEALS included 7 of the original 8 items using its standard 5-point Likert scale (1="not at all confident";

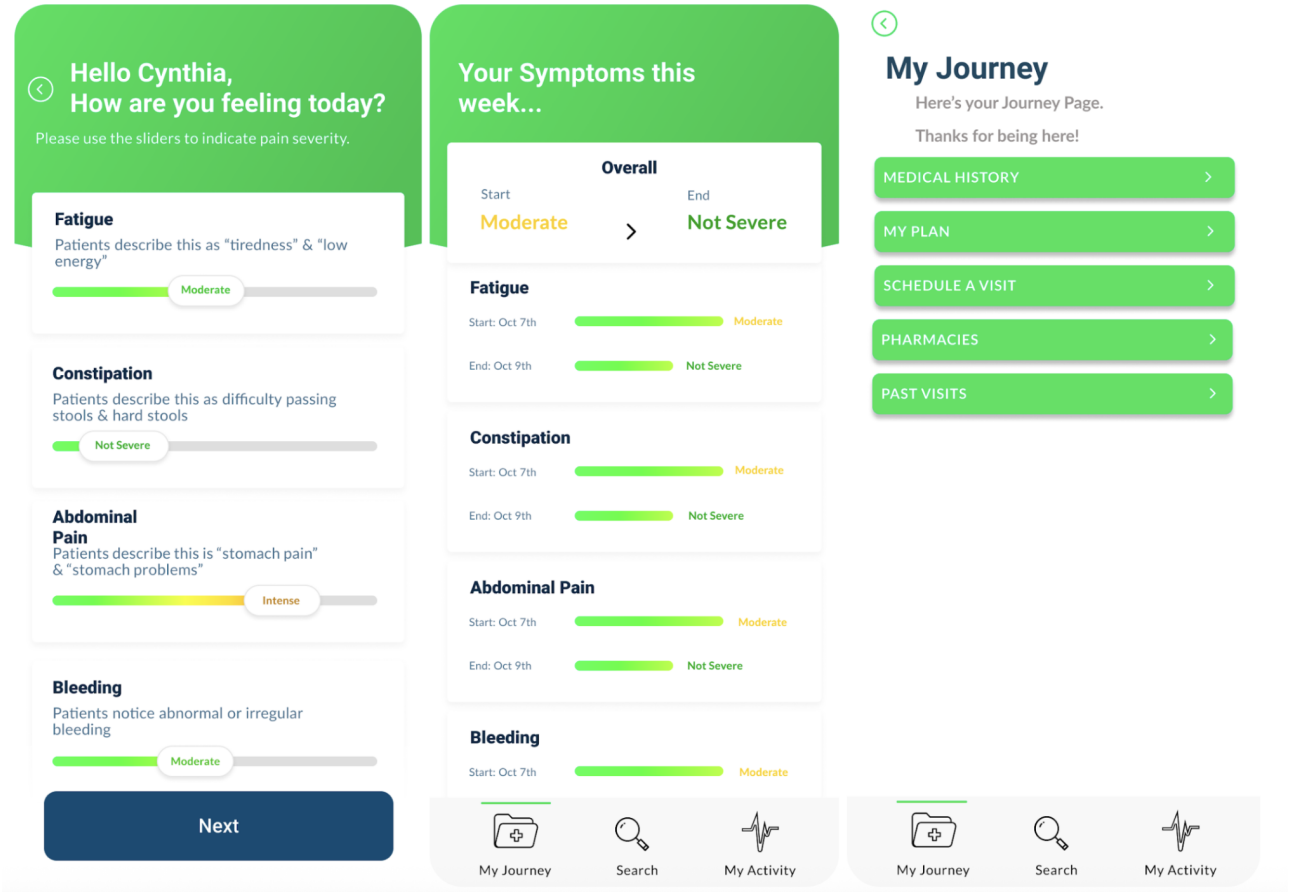
XSL•FO
RenderX

5="completely confident"). One question from the original eHEALS instrument, "I know how to use the health information I find on the internet to help me," was inadvertently omitted during survey implementation. The modified 7-item scale demonstrated adequate internal consistency among the 71 patient participants (Cronbach α=0.77). We proceeded with analysis of the collected data while acknowledging this limitation in our interpretation of the results. This assessment captured perceived digital health literacy and did not evaluate task-based interaction with a specific technology.

PocketDoc Development and Testing

PocketDoc (Figure 1) is a comprehensive digital health tool designed to support patients during the postoperative period, particularly in historically underserved populations. Key functionality includes symptom tracking, medication management with reminders, secure messaging for communication with health care providers, and educational resources tailored to specific procedures. The platform also includes appointment scheduling, follow-up management, and real-time tools for assessing symptom burden, all aimed at enhancing recovery and improving patient outcomes. These features were designed to accommodate varying levels of digital literacy, with particular attention to intuitive navigation and clear information presentation. The interface incorporated evidence-based design principles for health care applications [31], including consistent layout patterns, clear call to action elements, and simplified medical terminology. The development process included multiple rounds of iterative design and input from design professionals, business consultants, and historically underserved patients through multiple rounds of feedback.

Figure 1. Select screenshots of the introduction page and "My Journey" tab of PocketDoc.



For the usability testing and posttask surveys, responses were collected using a 10-point Likert scale (1="strongly disagree"; 10="strongly agree") to obtain granular insights into participants' experiences with the digital health prototype. Usability testing and posttask surveys were used to assess technology adoption readiness and perceived usability rather than digital health literacy.

Data Collection Procedures

Data collection occurred in 2 phases: the first phase included the demographic questionnaire, modified eHEALS assessment, and technology experience survey, and the second phase was the usability test, including standardized introductions to tasks, think-aloud protocols during task completion, and a posttask survey [32]. Sessions were conducted either in person or virtually via videoconference to accommodate participant availability, with each session lasting approximately 20 minutes.

Data Analysis

Survey results were analyzed using the Kruskal-Wallis *H* test to examine differences in digital health literacy and tool adoption across geographic categories. Post hoc Dunn-Bonferroni tests were conducted to examine rural-suburban and rural-urban comparisons in terms of digital health literacy. The Spearman rank-order correlation was used for associations between patient-reported capabilities and use of health resources, as well

XSL•FO
RenderX

as relationships between technology experience and tool adoption. Descriptive statistics were calculated for demographic and survey data. Exploratory linear regression analyses were conducted to examine whether technology experience (≥10 years vs <10 years) and overall digital health literacy were associated with patient confidence in their ability to use PocketDoc.

Ethical Considerations

This research was approved by Emory University's institutional review board in collaboration with a safety-net hospital in a southern US state (STUDY00004346). Written informed consent was obtained from all participants before taking part in the study. The informed consent process included a detailed explanation of study procedures, potential risks and benefits, the voluntary nature of participating, and the right to withdraw at any time. Participants were provided the opportunity to ask questions before providing consent. Participants received a US $20 gift card as compensation for their time. All data were deidentified before analysis.

Results

Participant Demographics

All 71 patients completed the modified eHEALS presurvey and subsequent assessments. Patient demographics are presented in Table 1. We found that there was diverse representation across age groups, with approximately half (n=33, 47%) between the ages of 18 and 34 years. Most participants came from urban (n=23, 32%) and suburban (n=43, 61%) environments, with limited rural representation (n=3, 4%). Most participants (n=41, 58%) had undergone surgery within the previous month, whereas 27% (n=19) had had surgery within the previous 6 months, and 14% (n=10) had had surgery within the previous year. Chronic conditions were reported by 58% (n=41) of the participants, indicating a population with ongoing health care needs. The patient sample was predominantly composed of Black individuals, reflecting the racial composition of the safety-net hospital population served.

Table . Demographic and clinical characteristics of postoperative patient participants (N=71).

Characteristic and category	Participants, n (%)
Survey completion	
eHEALS[a] presurvey	71 (100)
PocketDoc postsurvey	71 (100)
Time when surgery was conducted	
Within the previous month	41 (57.7)
Within the previous 6 months	19 (26.8)
Within the previous year	10 (14.1)
Chronic conditions	
Yes	41 (57.7)
No	27 (38)
Number of chronic conditions (n=41)	
1	23 (56.1)
2	12 (29.3)
3	6 (14.6)
Age group (y)	
18-24	14 (19.7)
25-34	19 (26.8)
35-44	9 (12.7)
45-54	7 (9.9)
55-64	12 (16.9)
65-74	7 (9.9)
75-84	3 (4.2)
≥85	0 (0)

[a]eHEALS: eHealth Literacy Scale.

XSL•FO
RenderX

Digital Health Literacy Levels

Analysis of the modified eHEALS survey (Table 2 and Multimedia Appendix 1) revealed that 69% (49/71) of patients reported high confidence (score of ≥3 on a 5-point scale) in finding health resources online (median 3, IQR 2-3). Furthermore, 61% (n=43) of participants expressed confidence in using the internet for health-related questions (median 3, IQR 2-3), and 63% (n=45) reported feeling capable of evaluating health resources (median 3, IQR 2-3). However, a notable drop occurred when examining confidence in health decision-making, with only 49% (n=35) feeling confident in using digital resources for making health decisions (median 2, IQR 2-3). Statistical analysis revealed that patients' confidence in finding health resources was positively correlated with interface satisfaction (ρ=0.262, 95% CI 0.03 - 0.47; *P*=.03). Similarly, patients' confidence in using these resources showed a significant positive correlation with interface satisfaction (ρ=0.304, 95% CI 0.07 - 0.52; *P*=.01).

Table . Item-level digital health literacy scores measured using the modified 7-item eHealth Literacy Scale (eHEALS).

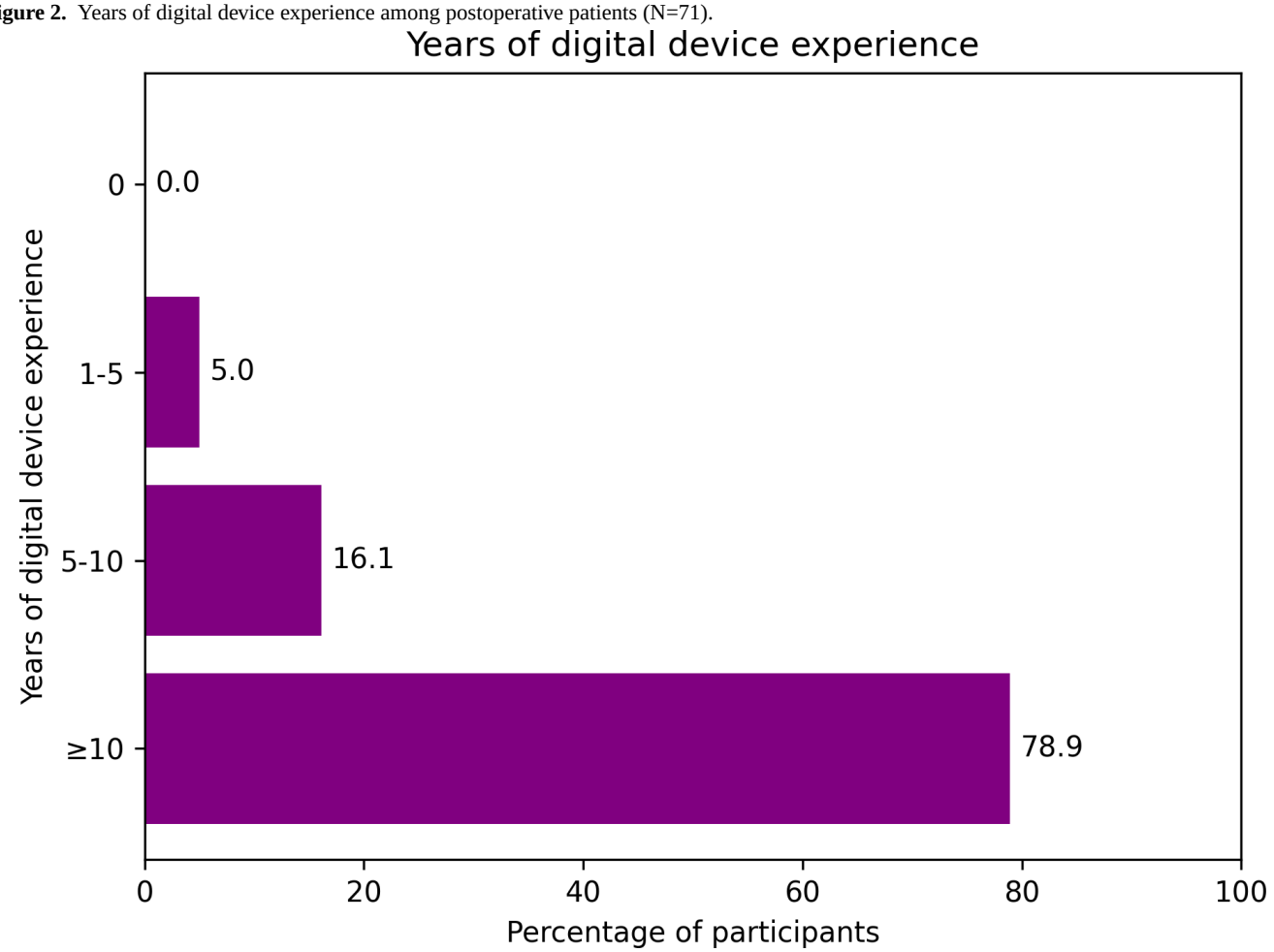
eHEALS item	Score (1-5), median (IQR)
Evaluating resources	3 (2-3)
Knowing how to find resources	3 (2-3)
Using resources	3 (2-3)
Knowing about the existence of resources	3 (2-3.5)
Identifying the quality of resources	3 (2-3.5)
Knowing where to find resources	3 (2-4)
Making decisions with resources	2 (2-3)

Technology Experience and Demographic Correlations

Analysis of technology experience (Figure 2) revealed that 79% (56/71) of the participants reported 10 or more years of digital device experience, indicating substantial familiarity with technology within the study population. Statistical analysis demonstrated no significant correlation between digital literacy and age or educational level. However, geographic location emerged as a notable factor in digital literacy levels. The Kruskal-Wallis *H* test revealed significant differences in perceptions based on living environment (effect size=0.112; *P*=.05). Urban and suburban patients reported similarly high mean confidence scores (urban: mean 9/10, SD 1.33; suburban: mean 9/10, SD 1.19), whereas rural patients reported significantly lower scores (mean 7/10, SD 0.0). Post hoc Dunn-Bonferroni tests showed differences between rural-suburban (*P*=.081) and rural-urban comparisons (*P*=.080), which did not meet the significance threshold.

XSL•FO
RenderX

Figure 2. Years of digital device experience among postoperative patients (N=71).



Tool Adoption and Use Patterns

In the usability assessment of the digital health prototype (Tables 3 and 4), participants demonstrated high levels of engagement and capability. Usability ratings across participants were high for ease of use (median 10, IQR 8-10), with task intuitiveness scoring highly at an average of 8.82 of 10 (median 10, IQR 8-10). Analysis revealed a significant correlation between prior technology exposure and perceived helpfulness of a digital health tool such as PocketDoc (ρ=0.230, 95% CI 0.01 - 0.43; *P*=.04). Participants with ≥10 years of digital device experience showed particularly strong performance in information-finding tasks (median 9, IQR 7-10).

Table . Usability and technology adoption readiness of the PocketDoc prototype among postoperative patients (N=71).

Usability metric	Survey item	Score (1-10), median (IQR)
Ease of use	"I think PocketDoc was easy to use."	10 (8-10)
Task scenario intuitiveness	"I think the Task Scenarios were intuitive to complete with PocketDoc."	10 (8-10)
Interface satisfaction	"I like the interface of PocketDoc."	9 (8-10)
Information findability	"I think my health information was presented to me in a way that was easy to find."	9 (7-10)
Information organization	"I think my health information was presented to me in a way that was well organized."	8 (7-10)
Compassionate wording	"I think my health information was presented to me in a way that was worded compassionately."	10 (9-10)
Information summary	"I think my health information was presented to me in a way that was useful for summary."	9 (7-10)
Health journey support	"I think PocketDoc would help me record my health journey after surgery."	9 (7-10)
Patient confidence	"I am confident in my ability to use PocketDoc (as demonstrated)."	10 (9-10)

XSL•FO
RenderX

Table . Internal consistency of the modified 7-item eHealth Literacy Scale (eHEALS; Cronbach α=0.77).

Item	eHEALS item wording
1	"I know what health resources are available on the Internet."
2	"I know where to find helpful health resources on the Internet."
3	"I know how to find helpful health resources on the Internet."
4	"I know how to use the Internet to answer my questions about health."
6	"I have the skills I need to evaluate the health resources I find on the Internet."
7	"I can tell high-quality health resources from low-quality health resources on the Internet."
8	"I feel confident in using information from the Internet to make health decisions."

Interestingly, participants' ability to distinguish high-quality health resources showed a significant negative correlation with ease of finding information (ρ=−0.251, 95% CI −0.46 to −0.02; *P*=.04), whereas the correlation with information organization perception was marginally significant (ρ=−0.215, 95% CI −0.42 to 0.01; *P*=.07).

In an exploratory regression analysis examining predictors of confidence in ability to use PocketDoc, the overall model was statistically significant (R^2=0.15; *P*=.004). Higher digital health literacy was independently associated with lower usability confidence (β=−0.31, 95% CI −0.50 to −0.12; *P*=.002), whereas technology experience (≥10 years vs <10 years) was not independently associated after accounting for literacy (β=0.36, 95% CI −0.19 to 0.91; *P*=.20).

Discussion

Principal Findings

Digital health literacy and technology adoption in safety-net hospital populations present a more nuanced picture than traditionally assumed. Prior research has suggested that patients in safety-net settings face barriers to digital health tool adoption due to limited access, lower digital literacy, and poor health literacy [26]. Our findings challenge the universality of these assumptions while highlighting specific areas in which targeted support may still be necessary for effective postoperative digital health implementation.

Importantly, digital health literacy and technology adoption readiness represent related but distinct constructs. The eHEALS captures patients' perceived ability to seek, evaluate, and apply online health information broadly, whereas usability findings reflect patients' interaction with a specific digital health tool within a defined clinical context. Distinguishing these constructs is critical for interpreting how baseline capabilities translate into real-world technology use.

In this study, relatively high levels of self-reported digital competency were observed, with 69% (49/71) of participants reporting high confidence in finding health resources online. This suggests that basic digital literacy barriers may be less universal among safety-net hospital patients than previously assumed [26]. These findings align with broader trends of increasing smartphone ownership and digital technology use across socioeconomic groups [27]. The high prevalence of substantial technology experience (n=56, 79% reporting ≥10 years of device use) further supports this shift.

Assessing digital health literacy alongside prototype usability provided important contextual insights. The relatively high baseline confidence in finding online health information helps frame the strong usability ratings observed for PocketDoc, including high scores for ease of use and task intuitiveness (median 10, IQR 8-10 in both cases). Together, these findings suggest that user-centered design approaches may facilitate accessibility across varying levels of digital health literacy.

A more granular pattern emerged when examining specific domains of digital health literacy. While many participants felt confident evaluating health resources (45/71, 63%) and using the internet for health-related questions (43/71, 61%), fewer reported confidence in using digital information for health decision-making (35/71, 49%). This pattern is consistent with prior work demonstrating that digital health literacy comprises multiple competencies, with decision-making often representing the greatest challenge [4,25,33,34].

The gap between general digital competency and confidence in health decision-making mirrors findings from recent studies highlighting the complexity of translating basic digital skills into applied health management [35,36]. This challenge may be particularly salient in postoperative care, where patients must interpret complex instructions and make time-sensitive decisions that carry immediate health implications [37].

Geographic differences in digital health literacy could not be robustly assessed due to the small rural sample size (n=3). Although rural participants demonstrated lower mean confidence scores than urban and suburban participants, these observations are exploratory and not suitable for inferential conclusions. Nonetheless, the observed pattern aligns with broader literature on rural-urban digital health disparities and underscores the need for future studies with adequately powered rural samples [35,38].

An additional insight emerged from the relationship between digital health literacy and usability perceptions. The negative association between critical evaluation skills and satisfaction with information organization (ρ=−0.251; *P*=.04) reflects a paradox observed in prior digital health research [39-41]. More

XSL•FO
RenderX

digitally literate users may hold higher expectations for information presentation and usability rather than experiencing reduced capability.

This interpretation is further supported by exploratory regression analyses, which suggested that higher digital health literacy was associated with lower confidence in usability. Together, these findings highlight a key design challenge: digital health tools must accommodate users with varying levels of sophistication and expectations. Adaptive interfaces that provide both simplified overviews and optional advanced features may be particularly well suited for postoperative care contexts.

While this study also collected rich qualitative data through interviews and extended usability observations, this manuscript focuses on quantitative analyses and usability themes. A forthcoming manuscript will explore qualitative findings in greater depth to further contextualize patient experiences and perspectives. This approach was chosen to maintain focus on the study's primary quantitative and usability outcomes.

Limitations

Several limitations should be considered when interpreting these findings. Our use of a modified eHEALS, which inadvertently omitted question 5 ("I know how to use the health information I find on the internet to help me"), may have affected the comprehensiveness of our digital health literacy assessment. This modification constitutes use of a nonvalidated tool, and while the modified version demonstrated internal consistency (Cronbach α=0.77), this omission limits direct comparability with other eHEALS-based studies and potentially overestimates overall literacy levels.

Our sample included limited representation from rural areas (n=3), which precludes any meaningful conclusions about geographic disparities. Any observed differences between rural and both urban and suburban participants are reported as exploratory findings only and require replication in adequately powered samples. Relatedly, post hoc power for the geographic group comparison was low; thus, this study is underpowered to detect true rural–urban and suburban differences. This limitation is relevant given the differences we observed between both urban and suburban and rural participants.

This study's reliance on self-reported measures of digital health literacy may not fully capture actual capabilities. While self-reported measures are widely used in digital health literacy research [42], they may be subject to social desirability bias and over- or underestimation of abilities. Our study may be subject to Berkson bias and self-selection bias as individuals with extremely low digital literacy or limited interest in digital tools may have been less likely to participate, and individuals with greater comfort with or interest in digital technologies may have been more likely to enroll.

Correlation analyses were exploratory and not adjusted for multiple testing, raising the possibility of type I error; therefore, the findings should be interpreted as hypothesis generating. Exploratory regression analyses were intentionally constrained to a small number of predictors due to sample size and ceiling effects in usability ratings and should not be interpreted as supporting causal inference.

Conclusions

This study advances our understanding of the evolving landscape of digital readiness among postoperative patients in safety-net settings. While participants reported substantial experience with digital devices, the data highlight a persistent challenge: translating general digital familiarity into confidence in managing personal health decisions. This gap suggests that, even when access and basic skills are present, patients may still struggle to apply digital tools meaningfully in complex clinical contexts such as postoperative care. Moreover, preliminary geographic trends point to location-based factors rather than traditional demographic variables as potential influences on confidence and engagement with digital health tools. These insights underscore the importance of designing tools that go beyond increasing access or general literacy, instead emphasizing support for decision-making and personalization based on patients' lived environments. Addressing these nuanced barriers may be key to optimizing digital health strategies and ensuring equitable postoperative support across diverse care settings.

Funding

This study was funded by a US $5000 Student Seed grant from the Emory-Georgia Tech Healthcare Innovation Program.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient-reported digital health literacy scores by domain.
[PNG File, 36 KB - humanfactors_v13i1e75496_app1.png]

References

XSL•FO
RenderX

1. Zhang J, Ge Y, Yang M, Ivers R, Webster R, Tian M. The role of digital health for post-surgery care of older patients with hip fracture: a scoping review. JMIR Preprints. Preprint posted online on Jun 10, 2021. [doi: 10.2196/preprints.31157]
2. He M, Chen M, Ji Y, Lu G. Effectiveness of smartphone app-based interventions after surgery on quality of recovery among cancer patients: a systematic review and meta-analysis. Ann Med 2024 Dec;56(1):2390167. [doi: 10.1080/07853890.2024.2390167] [Medline: 39140390]
3. Kemp E, Trigg J, Beatty L, et al. Health literacy, digital health literacy and the implementation of digital health technologies in cancer care: the need for a strategic approach. Health Promot J Austr 2021 Feb;32 Suppl 1(4):104-114. [doi: 10.1002/hpja.387] [Medline: 32681656]
4. Conard S. Best practices in digital health literacy. Int J Cardiol 2019 Oct 1;292:277-279. [doi: 10.1016/j.ijcard.2019.05.070] [Medline: 31230937]
5. Campanozzi LL, Gibelli F, Bailo P, Nittari G, Sirignano A, Ricci G. The role of digital literacy in achieving health equity in the third millennium society: a literature review. Front Public Health 2023 Feb 20;11:1109323. [doi: 10.3389/fpubh.2023.1109323] [Medline: 36891330]
6. Guhl E, Althouse AD, Pusateri AM, et al. The atrial fibrillation health literacy information technology trial: pilot trial of a mobile health app for atrial fibrillation. JMIR Cardio 2020 Sep 4;4(1):e17162. [doi: 10.2196/17162] [Medline: 32886070]
7. Hendawi R, Alian S, Li J. A smart mobile app to simplify medical documents and improve health literacy: system design and feasibility validation. JMIR Form Res 2022 Apr 1;6(4):e35069. [doi: 10.2196/35069] [Medline: 35363142]
8. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: a systematic review. J Telemed Telecare 2018 Jan;24(1):22-30. [doi: 10.1177/1357633X16673538] [Medline: 27760883]
9. Wiljer D, Shi J, Lo B, et al. Effects of a mobile and web app (thought spot) on mental health help-seeking among college and university students: randomized controlled trial. J Med Internet Res 2020 Oct 30;22(10):e20790. [doi: 10.2196/20790] [Medline: 33124984]
10. Armstrong KA, Coyte PC, Brown M, Beber B, Semple JL. Effect of home monitoring via mobile app on the number of in-person visits following ambulatory surgery: a randomized clinical trial. JAMA Surg 2017 Jul 1;152(7):622-627. [doi: 10.1001/jamasurg.2017.0111] [Medline: 28329223]
11. Lee L, Eustache J, Baldini G, et al. Enhanced recovery 2.0 - same day discharge with mobile app follow-up after minimally invasive colorectal surgery. Ann Surg 2022 Dec 1;276(6):e812-e818. [doi: 10.1097/SLA.0000000000004962] [Medline: 34091514]
12. Belarmino A, Walsh R, Alshak M, Patel N, Wu R, Hu JC. Feasibility of a mobile health application to monitor recovery and patient-reported outcomes after robot-assisted radical prostatectomy. Eur Urol Oncol 2019 Jul;2(4):425-428. [doi: 10.1016/j.euo.2018.08.016] [Medline: 31277778]
13. Hee Hwang J, Mun GH. An evolution of communication in postoperative free flap monitoring: using a smartphone and mobile messenger application. Plast Reconstr Surg 2012 Jul;130(1):125-129. [doi: 10.1097/PRS.0b013e318254b202] [Medline: 22743879]
14. Lu K, Marino NE, Russell D, et al. Use of short message service and smartphone applications in the management of surgical patients: a systematic review. Telemed J E Health 2018 Jun;24(6):406-414. [doi: 10.1089/tmj.2017.0123] [Medline: 29111887]
15. Abelson JS, Kaufman E, Symer M, Peters A, Charlson M, Yeo H. Barriers and benefits to using mobile health technology after operation: a qualitative study. Surgery 2017 Sep;162(3):605-611. [doi: 10.1016/j.surg.2017.05.007] [Medline: 28651777]
16. Semple JL, Sharpe S, Murnaghan ML, Theodoropoulos J, Metcalfe KA. Using a mobile app for monitoring post-operative quality of recovery of patients at home: a feasibility study. JMIR Mhealth Uhealth 2015 Feb 12;3(1):e18. [doi: 10.2196/mhealth.3929] [Medline: 25679749]
17. Gunter R, Fernandes-Taylor S, Mahnke A, et al. Evaluating patient usability of an image-based mobile health platform for postoperative wound monitoring. JMIR Mhealth Uhealth 2016 Sep 28;4(3):e113. [doi: 10.2196/mhealth.6023] [Medline: 27683059]
18. Temple-Oberle C, Yakaback S, Webb C, Assadzadeh GE, Nelson G. Effect of smartphone app postoperative home monitoring after oncologic surgery on quality of recovery: a randomized clinical trial. JAMA Surg 2023 Jul 1;158(7):693-699. [doi: 10.1001/jamasurg.2023.0616] [Medline: 37043216]
19. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016 Mar 10;13:35. [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
20. Spreadbury JH, Young A, Kipps CM. A comprehensive literature search of digital health technology use in neurological conditions: review of digital tools to promote self-management and support. J Med Internet Res 2022 Jul 28;24(7):e31929. [doi: 10.2196/31929] [Medline: 35900822]
21. Robinson A, Oksuz U, Slight R, Slight S, Husband A. Digital and mobile technologies to promote physical health behavior change and provide psychological support for patients undergoing elective surgery: meta-ethnography and systematic review. JMIR Mhealth Uhealth 2020 Dec 1;8(12):e19237. [doi: 10.2196/19237] [Medline: 33258787]

XSL•FO
RenderX

22. Yao R, Zhang W, Evans R, Cao G, Rui T, Shen L. Inequities in health care services caused by the adoption of digital health technologies: scoping review. J Med Internet Res 2022 Mar 21;24(3):e34144. [doi: 10.2196/34144] [Medline: 35311682]
23. Tieu L, Sarkar U, Schillinger D, et al. Barriers and facilitators to online portal use among patients and caregivers in a safety net health care system: a qualitative study. J Med Internet Res 2015 Dec 3;17(12):e275. [doi: 10.2196/jmir.4847] [Medline: 26681155]
24. Dawes AJ, Lin AY, Varghese C, Russell MM, Lin AY. Mobile health technology for remote home monitoring after surgery: a meta-analysis. Br J Surg 2021 Nov 11;108(11):1304-1314. [doi: 10.1093/bjs/znab323] [Medline: 34661649]
25. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. Int J Cardiol 2019 Oct 1;292:280-282. [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]
26. Burke J, Higgins MG, Vemuru SR, et al. Self-reported health literacy, digital literacy, and barriers to accessing care at a safety net breast surgical oncology clinic. medRxiv. Preprint posted online on Sep 23, 2024 URL: https://www.medrxiv.org/content/10.1101/2024.09.20.24314073v1 [accessed 2026-01-15]
27. Tsetsi E, Rains SA. Smartphone internet access and use: extending the digital divide and usage gap. Mob Media Commun 2017;5(3):239-255. [doi: 10.1177/2050157917708329]
28. Marler W. Mobile phones and inequality: findings, trends, and future directions. New Media Soc 2018;20(9):3498-3520. [doi: 10.1177/1461444818765154]
29. Lee WL, Lim ZJ, Tang LY, Yahya NA, Varathan KD, Ludin SM. Patients' technology readiness and eHealth literacy: implications for adoption and deployment of eHealth in the COVID-19 era and beyond. Comput Inform Nurs 2021 Nov 2;40(4):244-250. [doi: 10.1097/CIN.0000000000000854] [Medline: 34740221]
30. Norman CD, Skinner HA. eHEALS: the eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27. [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
31. Shneiderman B, Plaisant C, Cohen M, Jacobs S. Designing the User Interface: Strategies for Effective Human-Computer Interaction, 5th edition: Addison-Wesley; 2010.
32. Nielsen J. Usability Engineering: Morgan Kaufmann Publishers; 1993.
33. van der Heide I, Poureslami I, Mitic W, Shum J, Rootman I, FitzGerald JM. Health literacy in chronic disease management: a matter of interaction. J Clin Epidemiol 2018 Oct;102:134-138. [doi: 10.1016/j.jclinepi.2018.05.010] [Medline: 29793001]
34. Holmes-Rovner M, Kroll J, Schmitt N, et al. Patient satisfaction with health care decisions: the satisfaction with decision scale. Med Decis Making 1996;16(1):58-64. [doi: 10.1177/0272989X9601600114] [Medline: 8717600]
35. Arias López MD, Ong BA, Borrat Frigola X, et al. Digital literacy as a new determinant of health: a scoping review. PLOS Digit Health 2023 Oct 12;2(10):e0000279. [doi: 10.1371/journal.pdig.0000279] [Medline: 37824584]
36. Mainz A, Nitsche J, Weirauch V, Meister S. Measuring the digital competence of health professionals: scoping review. JMIR Med Educ 2024 Mar 29;10:e55737. [doi: 10.2196/55737] [Medline: 38551628]
37. Esper SA, Holder-Murray J, Meister KA, et al. A novel digital health platform with health coaches to optimize surgical patients: feasibility study at a large academic health system. JMIR Perioper Med 2024 Apr 4;7:e52125. [doi: 10.2196/52125] [Medline: 38573737]
38. Shiferaw KB, Tilahun BC, Endehabtu BF. Healthcare providers' digital competency: a cross-sectional survey in a low-income country setting. BMC Health Serv Res 2020 Nov 9;20(1):1021. [doi: 10.1186/s12913-020-05848-5] [Medline: 33168002]
39. Liu P, Yeh LL, Wang JY, Lee ST. Relationship between levels of digital health literacy based on the Taiwan digital health literacy assessment and accurate assessment of online health information: cross-sectional questionnaire study. J Med Internet Res 2020 Dec 21;22(12):e19767. [doi: 10.2196/19767] [Medline: 33106226]
40. Kinney AP, Sankaranarayanan B. Effects of patient portal use on patient satisfaction: survey and partial least squares analysis. J Med Internet Res 2021 Aug 27;23(8):e19820. [doi: 10.2196/19820] [Medline: 34448712]
41. Zhao BY, Chen MR, Lin R, Yan YJ, Li H. Influence of information anxiety on core competency of registered nurses: mediating effect of digital health literacy. BMC Nurs 2024 Sep 6;23(1):626. [doi: 10.1186/s12912-024-02275-3] [Medline: 39243073]
42. van der Vaart R, Drossaert C. Development of the digital health literacy instrument: measuring a broad spectrum of health 1.0 and health 2.0 skills. J Med Internet Res 2017 Jan 24;19(1):e27. [doi: 10.2196/jmir.6709] [Medline: 28119275]

Abbreviations

eHEALS: eHealth Literacy Scale

XSL•FO
RenderX

Edited by A Kushniruk; submitted 04.Apr.2025; peer-reviewed by B Ong, S Sahai; revised version received 20.Dec.2025; accepted 02.Jan.2026; published 23.Feb.2026.

Please cite as:
Awad C, Martinek AJ, Zou C, Byrd R, Ooi HJ, Do K, Young H, Adhyaru B, Idris M, Arriaga RI
Digital Health Literacy and Tool Adoption in Postoperative Care in a Safety-Net Hospital Population: Mixed Methods Study
JMIR Hum Factors 2026;13:e75496
URL: https://humanfactors.jmir.org/2026/1/e75496
doi:10.2196/75496

© Christopher Awad, Alexander Jakub Martinek, Chunhao Zou, Rosalind Byrd, Hui Jean Ooi, Kimberly Do, Henry Young, Bhavin Adhyaru, Muhammed Idris, Rosa I Arriaga. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 23.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluating User Engagement and Satisfaction With Digital Mental Health Interventions: Randomized Controlled Trial of a Text Messaging Program and e-Mental Health Resources

Gloria Obuobi-Donkor[1], PhD; Reham Shalaby[2], PhD, MD; Belinda Agyapong[2], MEd, PhD; Samuel Obeng Nkrumah[1], MD, MSc; Medard Kofi Adu[1], PhD; Ejemai Eboreime[1], PhD, MD; Lori Wozney[1,3], PhD; Vincent Israel Opoku Agyapong[1,2], PhD, MD

[1]Department of Psychiatry, Faculty of Medicine, Dalhousie University, 5909 Veterans Memorial Lane, 8th Floor. Abbie J. Lane Memorial Building QEII Health Sciences Centre, Halifax, NS, Canada

[2]Department of Psychiatry, Faculty of Medicine & Dentistry, University of Alberta, Edmonton, AB, Canada

[3]Mental Health and Addictions Program, IWK Health, Halifax, NS, Canada

Corresponding Author:
Vincent Israel Opoku Agyapong, PhD, MD
Department of Psychiatry, Faculty of Medicine, Dalhousie University, 5909 Veterans Memorial Lane, 8th Floor. Abbie J. Lane Memorial Building QEII Health Sciences Centre, Halifax, NS, Canada

Abstract

Background: Digital mental health tools, such as SMS text messaging and online resources, are increasingly used to support well-being. However, user satisfaction across these formats remains insufficiently explored.

Objective: The study assessed participants' engagement, perceived impact, and overall satisfaction with the Text4Support program and the e-mental health resources.

Methods: This randomized controlled study was conducted in Nova Scotia, Canada. Participants were assigned to either the Text4Support group, which received daily supportive text messages, or the control group, which received a single text message with a link to the Nova Scotia Mental Health and Addiction Program e-mental health resources. Responses to various aspects of the interventions were evaluated using a 5-point Likert scale, while overall satisfaction was measured on a scale from 0 to 10. The chi-square test and Fisher exact test were employed for data analysis.

Results: A total of 69 participants in the control group and 130 in the Text4Support group completed the satisfaction survey. The overall mean (SD) satisfaction score in the control group was 5.1 (2.3), and the overall mean (SD) satisfaction score for the Text4Support group was 7.1 (2.2). Compared to the control group, participants in the Text4Support group reported greater engagement and positive program impact. While 53.8% (70/130) of Text4Support recipients always read the messages, only 39.1% (27/69) of the control group rarely accessed the eHealth resources. When compared to the control group, participants allocated to the Text4Support group were reported to sometimes take positive action upon reading the messages (42.3% vs 33.3%). A significantly higher proportion of Text4Support users strongly agreed or agreed that the messages were supportive (81.4% vs 41.5%), positive (88.4% vs 49.2%), and helpful in coping with stress (44.2% vs 11.9%), loneliness (40.3% vs 13.4%), and improving mental well-being (51.2% vs 17.9%). In contrast, the majority of responses from the control group were largely neutral.

Conclusions: Results showed that Text4Support group participants were significantly more satisfied with the program than those receiving standard eHealth resources. This highlights that daily supportive SMS text messaging is an effective, low-cost adjunct to care delivery and mental health improvement. These findings suggest that aggregate, brief, and low-cost text-based interventions have great potential for increasing health access and engagement, particularly among traditionally disadvantaged populations with limited access to traditional services.

Trial Registration: ClinicalTrials.gov NCT05411302; https://clinicaltrials.gov/study/NCT05411302

(JMIR Hum Factors 2026;13:e76587) doi:10.2196/76587

KEYWORDS

text messaging; online resources; mental health; Nova Scotia; user satisfaction

XSL•FO
RenderX

Introduction

Psychiatric care has increasingly incorporated digital mental health interventions, which offer accessible and scalable support beyond traditional therapy and medication. Patient feedback through user satisfaction measures is vital in assessing the effectiveness of digital mental health interventions [1]. Patient experiences are gaining recognition as a key factor in enhancing health care quality alongside clinical effectiveness and safety [2,3]. These patient care experiences are included in public reporting and health performance programs [3]. Supportive SMS text messaging programs and eHealth services have become popular and convenient tools for managing mental health [4,5]. The structure and approaches of these interventions may vary in terms of user engagement and perceived effectiveness, but all aim to enhance well-being [6]. Nevertheless, satisfaction levels often vary, highlighting an area that requires more attention to optimize the effectiveness of such interventions for individuals with mental illness.

Supportive SMS text messaging programs like Text4Hope, Text4Mood, Text4PTSI, and Wellbeing offer automated text messages that provide preprogrammed positive reinforcement to users [7-9]. These coping message text programs require minimal input from users while guiding users through daily coping skills and emotional support [7-11]. Literature has shown that these programs increase feelings of connectedness and reduce mental health conditions [10]. Two randomized controlled trials (RCTs) conducted in Canada revealed that supportive text messages reduced depression and alcohol intake of subscribers [12,13].

Various studies have shown that eHealth resources such as online self-help services, psychoeducational tools, and virtual peer support groups provide an accessible way to seek help, as well as an independent role in receiving care [14,15]. These entail being able to partake in the content through specific timestamps, which ensures self-governance [16]. Nevertheless, there is little standardization regarding how users engage with eHealth interventions; it remains their prerogative to seek and employ the information furnished, and this influences overall satisfaction and perceived efficacy.

. This study aims to compare supportive SMS text messaging interventions with eHealth resources to better understand users' perceptions and feedback on the impact, engagement levels, and overall satisfaction. The study will evaluate engagement and user satisfaction with Text4Support in comparison to eHealth resources and assess the perceived relevance and impact of supportive SMS text messaging relative to these eHealth tools.

Given the importance of the perceived adequacy of digitally delivered tools as a core component of modern psychiatric intervention delivery systems to their success, it is essential to understand how users evaluate the adequacy of digitally delivered interventions. This study aims to compare supportive SMS text messaging interventions with eHealth resources to better understand users' perceptions and feedback on the impact, engagement levels, and overall satisfaction. The study will evaluate engagement and user satisfaction with Text4Support in comparison to eHealth resources and assess the perceived relevance and impact of supportive SMS text messaging relative to these eHealth tools.

The intervention, Text4Support, provides daily cognitive behavioral therapy (CBT)-informed supportive text messages for 6 months, tailored to users' primary mental health concerns and delivered as an adjunct to usual care. In contrast, the control group receives the usual care plus a single text message directing participants to freely available e-mental health tools offered through the Nova Scotia Mental Health and Addictions Program website.

Methods

Study Design

This RCT adopted a multicenter, two-arm design with a rater-blinded methodology. Participants were recruited and randomized into the Text4Support group and the control group.

Ethical Considerations

The study was approved by the Nova Scotia Health Research Ethics Board (REB File #1028174) and registered with ClinicalTrials.gov (NCT05411302). Each study participant provided informed consent, and the study was conducted according to the Declaration of Helsinki [17] and Good Clinical Practice (Canadian Guidelines) [18]. Participation was voluntary, with the option to withdraw at any time without impact on care. Participants' privacy and confidentiality were rigorously protected. All data were deidentified at the point of collection, stored, and secured. No compensation or incentives were provided to participants.

Data Collection/Intervention

The published study protocol detailed the data collection procedure [19]. In summary, data were collected via a self-administered survey powered by the Research Electronic Data Capture (REDCap) software program [20]. Participants' age, gender, income, education, housing status, relationship status, and clinical information were obtained. Data were collected between October 11, 2022 and July 31, 2024. After thoroughly informing participants about the study, we used an electronic consent form to obtain their consent.

Participants randomized to the control group received one text message with the Nova Scotia Mental Health and Addictions Program website link embedded in the message. This resource was freely accessible, and this service was provided as an add-on to participants' usual care. The website offers free, evidence-based e-mental health resources designed to address various psychiatric conditions with programs such as the early psychosis program, the eating disorder program, and recovery support [21]. Participants randomized to the Text4Support group received free daily, unilateral, CBT-based text messages tailored to their specific diagnosis as an add-on program to their usual care for 6 months. The supportive text messages were developed using evidence-based principles of CBT. Each message, limited to 160 characters, was designed to provide brief, accessible prompts that encourage adaptive coping, cognitive reframing, and emotional regulation. The content was collaboratively

XSL•FO
RenderX

created by CBT therapists, mental health professionals, and individuals with lived experience to ensure both clinical relevance and user-centered design. The underlying mechanism is grounded in the concept that regular exposure to concise, positively framed messages can reinforce helpful thinking patterns, enhance self-efficacy, and foster emotional resilience over time. Through consistent daily reinforcement, these messages aim to augment usual care by promoting psychological well-being and perceived connectedness. Participants completed the baseline surveys on enrollment via the online link. A text message with follow-up surveys, which includes the satisfaction survey, was sent to all participants at 6 weeks, 3 months, and 6 months.

Outcome Measures

The primary outcome measure was the participants' overall satisfaction with the Text4Support program or the Nova Scotia Health e-mental health online resources. Overall satisfaction was rated on an 11-point Likert scale, where zero indicated "very dissatisfied," 5 indicated "neutral," and 10 showed "very satisfied."

Secondary outcomes included the perceived impact and feedback on the Text4Support program and the Nova Scotia Health e-mental health online resources among the Text4Support and control groups. Study participants provided feedback on their satisfaction by responding to questions that assessed their engagement and use of each intervention and its perceived impact.

Engagement and use of each intervention were measured on a 5-point Likert scale (always, mostly, sometimes, rarely, and never). The perceived impact of each intervention on how participants coped with stress and loneliness, improved quality of life, overall physical and mental well-being, and the intervention's relevance, encouragement, and supportiveness were measured on a 5-point Likert scale, ranging from "strongly agree" to "strongly disagree" and including "neutral." For analysis purposes, responses were collapsed into 3 categories: strongly agree/agree, neutral, and strongly disagree/disagree.



The reliability and validity of this scale for testing satisfaction have not been examined; however, the survey was adopted in various studies to assess user satisfaction [8,10,22].

Data Analyses

Data were analyzed using SPSS for Windows (version 28; IBM Corporation) [23]. Participants' satisfaction data were presented as a continuous variable. We measured participants' overall satisfaction by rating the program from 0 to 10 (0=very dissatisfied, 5=neutral, and 10=very satisfied). The results were presented as means and SDs. Participants' engagement and use of each intervention (Text4Support/online resources) and the perceived impact of each intervention were summarized and presented as categorical variables and reported as frequency and percentages among the control group and Text4Support group. A chi-square and Fisher exact test were run to examine any differences in reporting user satisfaction, engagement, or perception of the Text4Support or online resources among the 2 groups, and a 2-tailed criterion (α<.05) was used to determine statistical differences. To handle missing data, we adopted imputation techniques, focusing on the last observation carried forward method.

Results

Figure 1 illustrates the flowchart of the participants. Eight hundred and ninety-eight patients were assessed for eligibility to enter the trial, of whom 781 eligible patients were randomized: 69 in the control group and 130 in the Text4Support group completed the satisfaction survey. The response rate in the control group was 18% (69/387), whereas the Text4Support group achieved a higher response rate of 33% (130/394).

XSL•FO
RenderX

Figure 1. Participant flowchart.

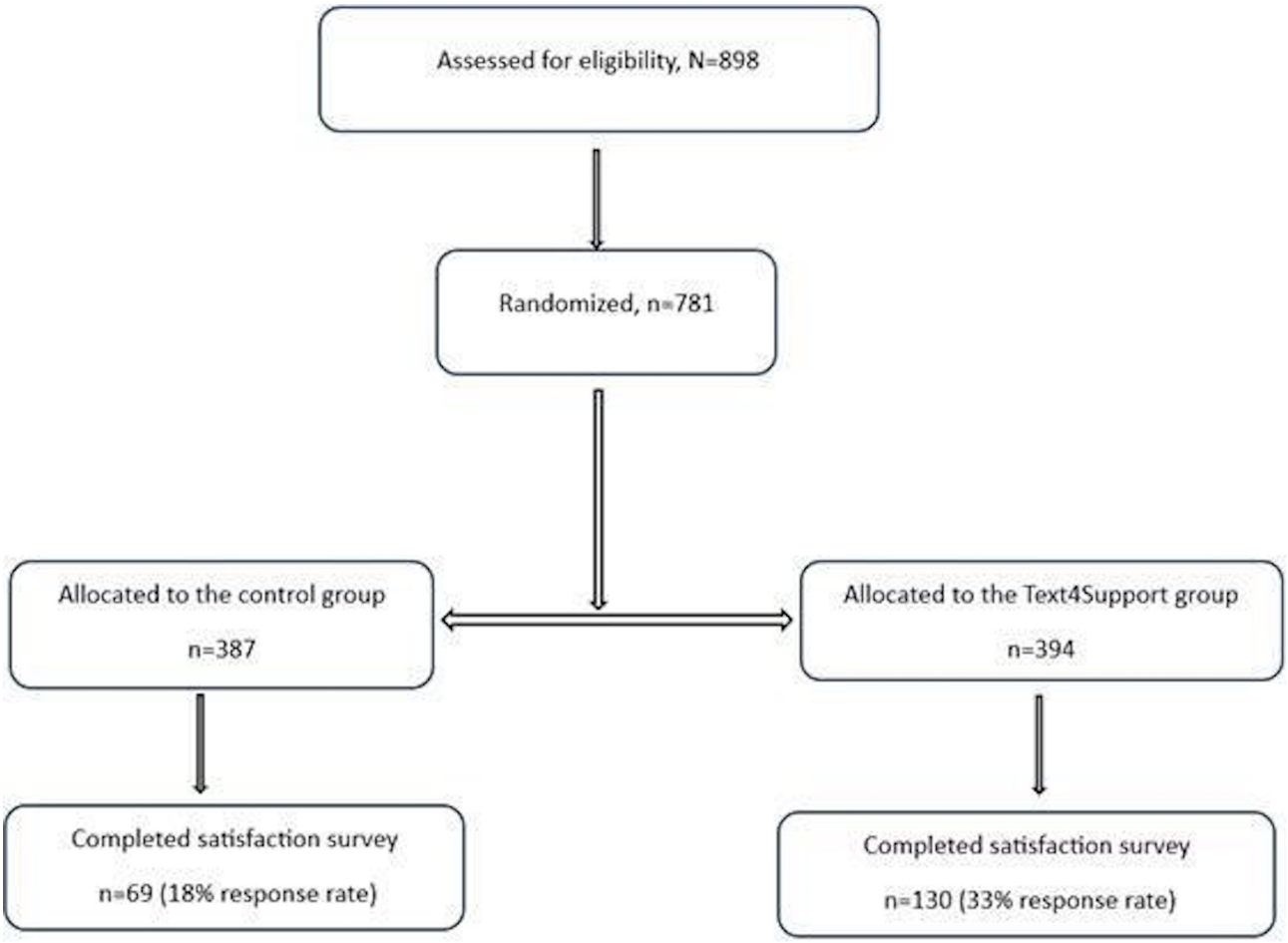


Table 1 shows the demographic characteristics of participants who completed the satisfaction survey. In total, across the 2 groups, the majority were between 26 and 60 years old (72.9%, 145/199), female (66.8%, 133/199), Caucasian (84.8%, 168/198), employed (52.3%, 103/197), and partnered/married (45.2%, 90/199) and had obtained a postsecondary education at a college or university (57.1%, 113/198). About 49% (97/199) of the study participants live in rented accommodation, 28.5% (55/193) have an income range of less than $29,590 (1 CAD≈US $0.77 at the time of the study), and 23.1% (27/117) have a primary diagnosis of depression. The chi-square analysis revealed no statistically significant differences in the demographic characteristics of participants in the Text4Support group and the control group, suggesting no differences in their demographic characteristics.

XSL•FO
RenderX

Table . Demographic characteristics of participants who completed the satisfaction survey.

Variable	Control group, n (%)	Text4Support group, n (%)	Total, n (%)	χ^2/Fisher exact *(df)*	*P* value
Age (y)					
≤25	13 (18.8)	21 (16.2)	34 (17.1)	3.1 (3)	.40
26 - 40	29 (42.0)	44 (33.8)	73 (36.7)		
41 - 60	23 (33.3)	49 (37.7)	72 (36.2)		
>60	4 (5.8)	16 (12.3)	20 (10.1)		
Sex at birth					
Male	21 (30.4)	45 (34.6)	66 (33.2)	0.4 (1)	.60
Female	48 (69.6)	85 (65.4)	133 (66.8)		
Ethnicity					
Caucasian	54 (78.3)	114 (88.4)	168 (84.8)	6.0[a]	.10
Indigenous	8 (11.6)	4 (3.1)	12 (6.1)		
Black/Hispanic	4 (5.8)	7 (5.4)	11 (5.6)		
Asian	3 (4.3)	4 (3.1)	7 (3.5)		
Employment status					
Student	7 (10.1)	11 (8.6)	18 (9.1)	2.3 (4)	.70
Employed	35 (50.7)	68 (53.1)	103 (52.3)		
Unemployed	16 (23.2)	22 (17.2)	38 (19.3)		
Retired	7 (10.1)	13 (10.2)	20 (10.2)		
Other	4 (5.8)	14 (10.9)	18 (9.1)		
Educational status					
Elementary school	2 (2.9)	2 (1.6)	4 (2.0)	1.9[a]	.80
High school	20 (29.0)	39 (30.2)	59 (29.8)		
Postsecondary (college/university)	41 (59.4)	72 (55.8)	113 (57.1)		
Postsecondary (trade school)	6 (8.7)	13(10.1)	19 (9.6)		
Other	0 (0.0)	3 (2.3)	3 (1.5)		
Relationship status					
Single	31 (44.9)	54 (42.5)	85 (42.7)	2.7[a]	.60
Partnered/married	32 (46.4)	58 (44.6)	90 (45.2)		
Separated/divorced	4 (5.8)	13 (10.0)	17 (8.5)		
Widowed	2 (2.9)	2 (1.5)	4 (2.0)		
Other	0 (0.0)	3 (2.3)	3 (1.5)		
Housing status					
Own home	19 (27.5)	43 (33.1)	62 (31.2)	1.7 (2)	.40
Rented accommodation	38 (55.1)	59 (45.4)	97 (48.7)		
Live with family/friends	12 (17.4)	28 (21.5)	40 (20.1)		
Income range (stats Canada)					
(1 CAD≈US $0.77 at the time of the study)					

XSL•FO
RenderX

Variable	Control group, n (%)	Text4Support group, n (%)	Total, n (%)	χ^2/Fisher exact *(df)*	*P* value
Less than $29,590	21 (32.3)	34 (26.6)	55 (28.5)	6.1 (5)	.30
$29,591-$59,180	22 (33.8)	28 (21.9)	50 (25.9)		
$59,181-$93,000	10 (15.4)	35 (27.3)	45 (23.3)		
$93,001-$150,000	5 (7.7)	12 (9.4)	17 (8.8)		
More than $150,000	4 (6.2)	10 (7.8)	14 (7.3)		
No income	3 (4.6)	9 (7.0)	12 (6.2)		
Primary diagnosis					
Depression	7 (15.2)	20 (28.2)	27 (23.1)	8.0[a]	.30
Anxiety	6 (13.0)	17 (23.9)	23 (19.7)		
Bipolar disorder	9 (19.6)	7 (9.9)	16 (13.7)		
Substance use disorder	2 (4.3)	3 (4.2)	5 (4.3)		
Alcohol use disorder	2 (4.3)	3 (4.2)	5 (4.3)		
Trauma-related disorders	9 (19.6)	7 (9.9)	16 (13.7)		
Psychosis	3 (6.5)	4 (5.6)	7 (6.0)		
Other	8 (17.4)	10 (14.1)	18 (15.4)		

[a]Fisher exact values.

The overall mean (SD) satisfaction score in the control group was 5.1 (2.3), and the overall mean (SD) satisfaction score for the Text4Support group was 7.1 (2.2).

Table 2 outlines participants' engagement with the Text4Support program and online mental health resources and the extent to which participants strongly agreed, agreed, were neutral, strongly disagreed, or disagreed with statements regarding these 2 resources. The majority of participants in the control group rarely read the eHealth resources (27/69, 39.1%), and 53.8% (70/130) in the Text4Support group always read the messages; 33.3% (23/69) in the control group sometimes takes a positive action after reading the online resources, and it was 42.3% (55/130) in the Text4Support group.

XSL•FO
RenderX

Table . Participants' feedback and perception of the supportive text messages and eHealth resources.

Perception and feedback	Control group, n (%)	Text4Support group, n (%)	Total, n (%)	χ^2/Fisher exact *(df)*	*P* value
Reading the messages/eHealth resources					
Always	6 (8.7)	70 (53.8)	76 (38.2)	79.8 (4)	<.001
Mostly	5 (7.2)	33 (25.4)	38 (19.1)		
Sometimes	19 (27.5)	16 (12.3)	35 (17.6)		
Rarely	27 (39.1)	6 (4.6)	33 (16.6)		
Never	12 (17.4)	5 (3.8)	17 (8.5)		
Read the messages/eHealth resources and take positive action after					
Always	3 (4.3)	13 (10.0)	16 (8.0)	15.5[a]	.003
Mostly	7 (10.1)	26 (20.0)	33 (16.6)		
Sometimes	23 (33.3)	55 (42.3)	78 (39.2)		
Rarely	21 (30.4)	28 21.5)	49 (24.6)		
Never	15 (21.7)	8 (6.2)	23 (11.6)		
The messages/eHealth resources were on the topic					
Strongly agree/agree	30 (46.2)	105 (81.4)	135 (69.6)	26.4[a]	<.001
Neutral	32 (49.2)	19 (14.7)	51 (26.3)		
Strongly disagree/disagree	3 (4.6)	5 (3.9)	8 (4.1)		
The messages/eHealth resources were on point					
Strongly agree/agree	29 (44.6)	104 (80.6)	133 (68.6)	25.7[a]	<.001
Neutral	32 (49.2)	21 (16.3)	53 (27.3)		
Strongly disagree/disagree	4 (6.2)	4 (3.1)	8 (4.1)		
The messages/eHealth resources were supportive					
Strongly agree/agree	27 (41.5)	105 (81.4)	132 (68.0)	31.7 (2)	<.001
Neutral	30 (46.2)	18 (14.0)	48 (24.7)		
Strongly disagree/disagree	8 (12.3)	6 (4.7)	14 (7.2)		
The messages/eHealth resources were positive					
Strongly agree/agree	32 (49.2)	114 (88.4)	146 (75.3)	38.0[a]	<.001
Neutral	30 (46.2)	10 (7.8)	40 (20.6)		
Strongly disagree/disagree	3 (4.6)	5 (3.9)	8 (4.1)		
The messages/eHealth resources improved my quality of life					

XSL•FO
RenderX

Perception and feedback	Control group, n (%)	Text4Support group, n (%)	Total, n (%)	χ^2/Fisher exact *(df)*	*P* value
Strongly agree/agree	8 (11.9)	47 (36.4)	55 (28.1)	13.7 (2)	.001
Neutral	43 (64.2)	55 (42.6)	98 (50.0)		
Strongly disagree/disagree	16 (23.9)	27 (20.9)	43 (21.9)		
The messages/eHealth resources improved my overall mental well-being					
Strongly agree/agree	12 (17.9)	66 (51.2)	78 (39.8)	22.6 (2)	<.001
Neutral	40 (59.7)	38 (29.5)	78 (39.8)		
Strongly disagree/disagree	15 (22.4)	25 (19.4)	40 (20.4)		
The messages/eHealth resources improved my overall physical well-being					
Strongly agree/agree	7 (10.4)	28 (21.7)	35 (17.9)	4.5 (2)	>.99
Neutral	40 (59.7)	61 (47.3)	101 (51.5)		
Strongly disagree/disagree	20 (29.9)	40 (31.0)	60 (30.6)		
The messages/eHealth resources helped me cope with loneliness					
Strongly agree/agree	9 (13.4)	52 (40.3)	61 (31.1)	16.3 (2)	<.001
Neutral	38 (56.7)	43 (33.3)	81 (41.3)		
Strongly disagree/disagree	20 (29.9)	34 (26.4)	54 (27.6)		
The messages/eHealth resources helped me cope with stress					
Strongly agree/agree	8 (11.9)	57 (44.2)	65 (33.2)	20.8 (2)	<.001
Neutral	40 (59.7)	47 (36.4)	87 (44.4)		
Strongly disagree/disagree	19 (28.4)	25 (19.4)	44 (22.4)		
The messages/eHealth resources were not relevant to my concern					
Strongly agree/agree	13 (20.0)	12 (9.3)	25 (12.9)	24.4 (2)	<.001
Neutral	33 (50.8)	31 (24.0)	64 (33.0)		
Strongly disagree/disagree	19 (29.2)	86 (66.7)	105 (54.1)		
The messages/eHealth resources were not helpful					
Strongly agree/agree	8 (12.3)	9 (7.0)	17 (8.8)	31.4 (2)	<.001
Neutral	34 (52.3)	22 (17.1)	56 (28.6)		
Strongly disagree/disagree	23 (35.4)	98 (76.0)	121 (62.4)		

XSL•FO
RenderX

Perception and feedback	Control group, n (%)	Text4Support group, n (%)	Total, n (%)	χ^2/Fisher exact *(df)*	*P* value
The messages/eHealth resources were not encouraging					
Strongly agree/agree	11 (16.9)	11 (8.5)	22 (11.3)	40.9 (2)	<.001
Neutral	33 (50.8)	17 (13.2)	50 (25.8)		
Strongly disagree/disagree	21 (32.3)	101 (78.3)	122 (62.9)		
The messages/eHealth resources were negative					
Strongly agree/agree	4 (6.2)	2 (1.6)	6 (3.1)	47.4[a]	<.001
Neutral	36 (55.4)	15 (11.6)	51 (26.3)		
Strongly disagree/disagree	25 (38.5)	112 (86.8)	137 (70.6)		

[a]Fisher exact values.

Most study participants were neutral when asked if the eHealth resources were on the topic (32/65, 49.2%), on point (32/65, 49.2%), supportive (30/65, 46.2%), helpful in coping with loneliness (38/67, 56.7%), and helpful in coping with stress or improving their physical and mental well-being (40/67, 59.7%). However, 49.2% (32/65) strongly agreed or agreed that the eHealth resources were positive. In terms of how eHealth improved quality of life, 64.2% (43/67) were neutral as well as for negativity of the online resources (36/67, 55.4%), not encouraging (33/65, 50.8%), not helpful (34/65, 52.3%), and not relevant (33/67, 50.8%). In contrast, the majority of participants in the Text4Support group strongly agreed or agreed that the messages were on the topic (105/129, 81.4%), were on point (104/129, 80.6%), supportive (105/129, 81.4%), positive (114/129, 88.8%), and helped them cope with loneliness (52/129, 40.3%), stress (57/129, 44.2%), or improve their overall mental well-being (66/129, 51.2%). In terms of how the messages improved quality of life, 42.6% (55/129) in the Text4Support group were neutral and strongly disagreed or disagreed with the messages being negative (112/129, 86.8%), not encouraging (101/129, 78.3%), not helpful (98/129, 76%), and not relevant (86/129, 66.7%). The results revealed statistically significant differences between groups, indicating that the observed effects were not due to chance and were not coincidental.

Discussion

Principal Findings

This study provides new insights into how patients perceive text message interventions, eHealth resources, and their engagement with the Text4Support program and the Nova Scotia Health online resources. Findings reveal that the overall mean (SD) satisfaction score in the control group was 5.1 (2.3), and the overall mean (SD) satisfaction score for the Text4Support group was 7.1 (2.2). Compared to the control group, participants in the Text4Support group reported greater engagement and positive program impact. While 53.8% (70/130) of Text4Support recipients always read the messages, only 8.7% (6/69) of the control group always read the eHealth resources. Participants allocated to the Text4Support group were reported to sometimes take action upon reading the messages (42.3% vs 33.3%). A significantly higher proportion of Text4Support users strongly agreed or agreed that the messages were supportive (81.4% vs 41.5%), positive (88.4% vs 49.2%), and helpful in coping with stress (44.2% vs 11.9%), loneliness (40.3% vs 13.4%), and improving mental well-being (51.2% vs 17.9%). In contrast, responses from the control group were largely neutral. The majority of respondents in the Text4Support group agreed or strongly agreed that the messages had a positive impact on them, which achieved statistical significance (P<.001) when compared with the e-mental health resource group; however, physical well-being was neutral across the 2 groups and not statistically significant (P=.1).

The response rates observed in this study provide important insights into participant engagement with digital mental health interventions. The Text4Support group demonstrated a notably higher response rate (33%) compared to the control group (18%). Although the response rates remain relatively low, raising concerns about potential attrition bias and the representativeness of the findings, the differences suggest that supportive SMS text messaging interventions may play a significant role in enhancing participant engagement and willingness to interact with mental health research and services. Improved response rates have important implications for the validity and generalizability of study findings. Higher response rates enhance the appropriateness and representativeness of the sample [24].

The findings indicate a high level of satisfaction among participants in the Text4Support group, with an average score of 7.1. Although our study participants expressed high satisfaction, other studies recorded higher satisfaction with the supportive text intervention [22,25-27]. For example, a study conducted among educators showed a mean satisfaction of about 8.5 among administrative educators [26]. Again, a cross-sectional study among 2032 participants recorded 8.55

XSL•FO
RenderX

overall mean satisfaction with the Text4Hope program [27]. Among the control group, an overall average satisfaction rate of 5.1 for online mental health resources suggests moderate user satisfaction. In terms of context and use, this measure aligns with the literature in that satisfaction with online mental health services is low to moderate [28]. As an example, one study making use of the Client Satisfaction Questionnaire-8 found that most patients reported low-to-moderate satisfaction with online psychiatric services [28].

The Text4Support program produced significant participant engagement, as 53% of participants consistently read text messages, while only 8.7% of control group members always read online resources. Outcomes from various studies have shown that interactive text-based interventions help maintain long-term therapeutic engagement, which delivers essential immediate mental health support [22,27,29,30]. A systematic review and meta-analysis found that SMS text messaging interventions for health promotion had widespread acceptance among participants and demonstrated high engagement rates throughout different health areas [29,30]. Although few participants always read the messages/eHealth resources and took action afterward, 73.8% and 83.8% of participants in the control and intervention groups, respectively, at least rarely read and took action in our study. However, the relatively low percentage of participants taking consistent action based on the messages suggests that passive consumption may not suffice for significant behavioral change. Integrating interactive components, such as bidirectional messaging, could enhance effectiveness. A scoping review of text message interventions in adolescent mental health services found that 65% of studies involved bidirectional messaging, indicating its potential to boost engagement and outcomes [31].

On the other hand, engagement with the provided eHealth resources was lower, with only a subset of participants actively utilizing them. Research indicates that eHealth interventions often struggle with uptake unless seamlessly integrated into daily routines [32]. Barriers to eHealth engagement usually stem from issues such as lack of time, perceived irrelevance, and digital literacy challenges [33]. Unlike text messages, which require minimal effort to consume, accessing eHealth resources demands additional steps that may deter some users. The brevity of text messages (limited to 160 characters) makes them easier to process, especially for individuals experiencing mental health challenges [34]. In contrast, studies on guided digital interventions suggest that providing step-by-step instructions or integrating behavioral prompts within text messages can enhance the utilization of eHealth resources [35]. One notable finding is that while 66.7% of the Text4Support group and 29.2% in the control group found the messages relevant, 9.3% and 20%, respectively, did not. The Text4Support group exhibited a notably higher percentage of perceived relevance, indicating that Text4Support was more effectively aligned with participants' mental health needs than the control condition. A significant percentage of participants, 9.3% in the Text4Support group and 20% in the control group, indicated that the messages/online resources were not pertinent to their needs. This disparity highlights the necessity of customizing digital interventions to individual users' specific needs, circumstances, and preferences. The control group's elevated percentage of irrelevance likely indicates the generic nature of nontargeted messages, while the Text4Support group's reduced percentage implies a degree of alignment with participants' needs, although there is still potential for enhancement.

Regarding perceived relevance and utility, the results show that most Text4Support participants (81.4%) found the messages relevant to the topic. In comparison, 46.2% strongly agreed or agreed with the online resource, underscoring the importance of tailored content. Personalization has been identified as a key factor in the success of text-based interventions. Literature has highlighted that those interventions allowing rapid, personalized exchanges, akin to natural conversation, were more effective in addressing individual needs [22,36].

The purpose of Text4Support was to provide emotional support through daily messages. The Text4Support group overwhelmingly found the messages supportive (81.4%) and positive (88.4%). These findings demonstrate that digital interventions with positive reinforcement significantly improve users' mental well-being [10,22,37,38]. Research indicates that such interventions can effectively develop therapeutic connections, meeting unmet needs for mental health care [10,27]. However, the debate persists on whether supportive messaging alone yields lasting psychological improvements. Studies suggest combining SMS text messaging with other treatment modalities, such as CBTs, may enhance effectiveness [39]. The supportive text messages delivered in this study were an addition to the patient's usual care, making it more effective.

The supportive text messages positively impacted mental well-being, with 51.2% of Text4Support participants reporting improvements, compared to 17.9% in the control group. Similarly, 40.3% of the intervention group reported reduced loneliness and 44.2% experienced lower stress levels. Other studies align with our findings; for example, a similar program, Wellness4Teacher, reported that 95.4% of its respondents reported that the messages were at least often positively impactful, while 93.3% reported that the messages were affirming [26]. These findings highlight the potential of digital interventions as supplementary tools for mental health support [40,41].

Our results show substantial differences in neutral responses across all groups in terms of improving the quality of life. The highest proportion of neutral participant responses was 64.2% in the control group and 42.6% in the Text4Support group, totaling 50% among the study sample. Participants who gave neutral responses were all neither for nor against the intervention having an effect on their quality of life. The slightly larger percentage of indifferent responses from the control group indicates either a lack of participation or ineffectiveness of regular care in significantly changing the participants' assessments of life quality. This is perhaps due to a more focused and structured approach, which is likely to impress participants more than those of general care. The findings align with previous research showing that participants' opinions about digital treatments vary [42,43]. For example, Proudfoot et al [43] found that while 68% of respondents indicated considerable improvements to their mental health following digital treatments,

XSL•FO
RenderX

about 25% reported neutral or unsure answers, underscoring that outcomes are determined contextually. A qualitative research study also documents instances of most respondents feeling neutral or holding dual views about digital mental health interventions [44].

Similarly, a meta-analysis of RCTs highlighted that, as a result of the differences in people's participation and expectations, neutral answers often emerge in studies involving smartphone-based mental health therapies [42].

Across groups, participants demonstrated similar disagreement percentages, with the coping of stress registering 28.4% in the eHealth resource group and 19.4% in the Text4Support group, resulting in an overall rate of 22.4%. Although digital interventions may have received more positive feedback, they were not effective for everyone, and some participants found them ineffective in stress coping [45].

However, both interventions had a weaker effect on physical well-being, with no significant differences between groups [46,47]. This finding is consistent with prior research, which has shown that text-based interventions can effectively improve mental health outcomes but have a limited influence on physical health behaviors [47,48]. Digital interventions targeting physical health require more structured behavior-change strategies, such as goal setting and reinforcement mechanisms, rather than passive messaging alone [49].

Implications for Policy, Practice, and Future Research

The findings of this study have several important implications for the design, implementation, and scaling of digital mental health interventions. First, the significantly higher satisfaction and engagement levels observed in the Text4Support group underscore the value of incorporating supportive, brief, and personalized messaging into routine mental health care. Policymakers and health care providers should consider integrating text-based interventions as a standard adjunct to traditional care, particularly for populations that may face barriers to in-person or web-based services.

Given the low engagement with static eHealth resources, health systems should prioritize low-barrier, mobile-first tools that align with users' daily routines. Moreover, to increase the effectiveness of such interventions, future policy should encourage the development of interactive, bidirectional messaging systems that foster active participation and allow for real-time support and personalization based on user feedback and needs. Tailoring content to individual needs, such as demographic or diagnostic profiles, can further enhance relevance and efficacy, as evidenced by the higher perceived relevance and positive impact of the Text4Support messages.

Although Text4Support currently relies on preprogrammed, therapist-developed messages, future iterations of supportive SMS text messaging interventions could leverage artificial intelligence, including generative artificial intelligence, to enhance message compilation, personalization, and responsiveness. Finally, given the limited impact on physical well-being, digital programs should integrate behavior change techniques (eg, goal setting, reminders, and progress tracking) when addressing physical health outcomes. Health authorities should support cross-sector collaboration between digital health developers, clinicians, and researchers to ensure interventions are both evidence-based and contextually responsive. These findings advocate for a patient-centered, adaptive approach to mental health support, where digital tools are not merely supplemental but are integral to comprehensive, accessible, and scalable care strategies.

Limitations

Our study has limitations that need to be considered when appraising the results. First, the generalizability of the results may be affected by the comparatively small sample size of participants who completed the follow-up satisfaction survey, particularly in the control group. The imbalance in group sizes (130 in the Text4Support group vs 69 in the control group) may have introduced bias or limited the statistical power to detect differences in some outcomes. The small sample size may be attributed to the online nature of the surveys; literature has reported that online distributed surveys are unlikely to keep participants in a study for follow-up assessment compared to paper-based surveys [50]. One notable limitation of this study is the relatively low survey completion rate. While this may impact the generalizability of the findings, it is consistent with patterns observed in other digital mental health and eHealth trials, which frequently report attrition rates exceeding 30% [51,52]. Again, participants' satisfaction, engagement, and perception were self-reported, which may result in recall bias since participants may overestimate their feedback. Lastly, the web-based questionnaire used to assess participant satisfaction and perception was not a validated instrument, which may reduce the reliability of the reported satisfaction rates. Nonetheless, this was one of the first RCTs examining user satisfaction among Text4Support users and Nova Scotia mental health and addiction online resources users in Nova Scotia.

Conclusions

This study's findings demonstrate the meaningful impact of personalized, supportive SMS text messaging in enhancing engagement, emotional well-being, and satisfaction among mental health service users. As mental health systems evolve to meet rising demands, digital interventions like Text4Support represent a promising, scalable solution that can complement existing care. To maximize their potential, policies must prioritize accessibility, personalization, and integration with broader care models, ensuring that digital tools meet people where they are, with the right message at the right time.

Funding

The study was funded by the Queen Elizabeth II Foundation and supported by the Global Psychological eHealth Foundation, the Department of Psychiatry at Nova Scotia Health, and Dalhousie University. The funders had no role in the study's design, execution,

XSL•FO
RenderX

data collection, analysis, interpretation, or writing. They were also not involved in reviewing, approving, or deciding to publish the results.

Data Availability

The data supporting this manuscript's conclusions will be made available upon reasonable request from the corresponding author.

Authors' Contributions

VIOA conceived, designed, and supervised the study. VIOA, RS, and GO-D conducted data analysis. The initial manuscript was drafted by GO-D. VIOA, RS, BA, SON, MKA, EE, and LW reviewed the initial draft. All the authors contributed to the study's design, revised the initial draft manuscript, and approved the final draft before submission. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

VIOA is the founder and Principal Investigator of the ResilienceNHope messaging programs and volunteers as the Board Chair of the Global Psychological eHealth Foundation, a not-for-profit organization. BA volunteers as the President and Chief Executive Officer of the Global Psychological eHealth Foundation. RS volunteers as the Program Director of the Global Psychological eHealth Foundation. VIOA, BA, and RS receive no financial or other compensation from the Global Psychological eHealth Foundation and have no financial conflicts of interest about this article. GO-D, SON, MA, EE, and LW have no commercial or financial conflict of interest.

Checklist 1
CONSORT-eHEALTH (v 1.6.1) checklist.
[PDF File, 1144 KB - humanfactors_v13i1e76587_app1.pdf]

References

1. Børge M, Yngvild A. User satisfaction with child and adolescent mental health services: the association between user satisfaction and clinical outcomes. BMC Psychiatry 2024 Apr 15;24(1):279. [doi: 10.1186/s12888-024-05715-1] [Medline: 38622603]
2. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. BMJ Open 2013 Jan 3;3(1):e001570. [doi: 10.1136/bmjopen-2012-001570] [Medline: 23293244]
3. Anhang Price R, Elliott MN, Zaslavsky AM, et al. Examining the role of patient experience surveys in measuring health care quality. Med Care Res Rev 2014 Oct;71(5):522-554. [doi: 10.1177/1077558714541480] [Medline: 25027409]
4. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013;35(4):332-338. [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
5. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
6. Aguilera A, Muñoz RF. Text messaging as an adjunct to CBT in low-income populations: a usability and feasibility pilot study. Prof Psychol Res Pr 2011 Dec 1;42(6):472-478. [doi: 10.1037/a0025499] [Medline: 25525292]
7. Obuobi-Donkor G, Shalaby R, Eboreime E, et al. Text4PTSI: a promising supportive text messaging program to mitigate psychological symptoms in public safety personnel. Int J Environ Res Public Health 2023 Feb 27;20(5):4215. [doi: 10.3390/ijerph20054215] [Medline: 36901235]
8. Agyapong VIO, Mrklas K, Juhás M, et al. Cross-sectional survey evaluating Text4Mood: mobile health program to reduce psychological treatment gap in mental healthcare in Alberta through daily supportive text messages. BMC Psychiatry 2016 Nov 8;16(1):378. [doi: 10.1186/s12888-016-1104-2] [Medline: 27821096]
9. Agyapong VIO, Hrabok M, Vuong W, et al. Changes in stress, anxiety, and depression levels of subscribers to a daily supportive text message program (Text4Hope) during the COVID-19 pandemic: cross-sectional survey study. JMIR Ment Health 2020 Dec 18;7(12):e22423. [doi: 10.2196/22423] [Medline: 33296330]
10. Agyapong B, Shalaby R, Eboreime E, et al. Subscribers' perspectives and satisfaction with the MoreGoodDays supportive text messaging program and the impact of the program on self-rated clinical measures. J Clin Med 2024 Jan 19;13(2):580. [doi: 10.3390/jcm13020580] [Medline: 38276086]
11. Kornfield R, Stamatis CA, Bhattacharjee A, et al. A text messaging intervention to support the mental health of young adults: User engagement and feedback from a field trial of an intervention prototype. Internet Interv 2023 Dec;34:100667. [doi: 10.1016/j.invent.2023.100667] [Medline: 37746639]
12. Agyapong VIO, Juhás M, Mrklas K, et al. Randomized controlled pilot trial of supportive text messaging for alcohol use disorder patients. J Subst Abuse Treat 2018 Nov;94:74-80. [doi: 10.1016/j.jsat.2018.08.014] [Medline: 30243421]

XSL•FO
RenderX

13. Agyapong VIO, Juhás M, Ohinmaa A, et al. Randomized controlled pilot trial of supportive text messages for patients with depression. BMC Psychiatry 2017 Aug 2;17(1):286. [doi: 10.1186/s12888-017-1448-2] [Medline: 28768493]
14. Berrouiguet S, Baca-García E, Brandt S, Walter M, Courtet P. Fundamentals for future mobile-Health (mHealth): a systematic review of mobile phone and web-based text messaging in mental health. J Med Internet Res 2016 Jun 10;18(6):e135. [doi: 10.2196/jmir.5066] [Medline: 27287668]
15. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11. [doi: 10.1002/wps.20083] [Medline: 24497236]
16. Karasouli E, Adams A. Assessing the evidence for e-Resources for mental health self-management: a systematic literature review. JMIR Ment Health 2014;1(1):e3. [doi: 10.2196/mental.3708] [Medline: 26543903]
17. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
18. Good clinical practices. Government of Canada. 2022. URL: https://www.canada.ca/en/health-canada/services/drugs-health-products/compliance-enforcement/good-clinical-practices.html [accessed 2025-12-24]
19. Dias RDL, Shalaby R, Agyapong B, et al. Augmenting mental health support for patients accessing different degrees of formal psychiatric care through a supportive text messaging program: protocol for a randomized controlled trial. Methods Protoc 2023 Feb 13;6(1):19. [doi: 10.3390/mps6010019] [Medline: 36827506]
20. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
21. Mental health and addictions tools. Nova Scotia Mental Health and Addictions. 2025. URL: https://mha.nshealth.ca/en/tools [accessed 2025-01-01]
22. Obuobi-Donkor G, Eboreime E, Shalaby R, et al. User satisfaction with a daily supportive text message program (Text4PTSI) for public safety personnel: longitudinal cross-sectional study. JMIR Form Res 2023 Jun 23;7:e46431. [doi: 10.2196/46431] [Medline: 37351940]
23. IBM SPSS statistics for windows, version 28.0. IBM Corp. 2021. URL: https://www.ibm.com/products/spss-statistics [accessed 2025-01-01]
24. Holtom B, Baruch Y, Aguinis H, A Ballinger G. Survey response rates: trends and a validity assessment framework. Human Relations 2022 Aug;75(8):1560-1584. [doi: 10.1177/00187267211070769]
25. Cartujano-Barrera F, Sanderson Cox L, Arana-Chicas E, et al. Feasibility and acceptability of a culturally- and linguistically-adapted smoking cessation text messaging intervention for Latino smokers. Front Public Health 2020;8:269. [doi: 10.3389/fpubh.2020.00269] [Medline: 32714891]
26. Agyapong B, Brett-MacLean P, da Luz Dias R, Wei Y, Agyapong VIO. Satisfaction with a daily supportive text messaging program aimed at promoting psychological well-being of educators: Outcomes from implementation of the Wellness4Teachers. Digit Health 2025;11:20552076251317558. [doi: 10.1177/20552076251317558] [Medline: 40904755]
27. Shalaby R, Vuong W, Hrabok M, et al. Gender differences in satisfaction with a text messaging program (Text4Hope) and anticipated receptivity to technology-based health support during the COVID-19 pandemic: cross-sectional survey study. JMIR Mhealth Uhealth 2021 Apr 15;9(4):e24184. [doi: 10.2196/24184] [Medline: 33750738]
28. Jilani AQ, Khan A, Saloni S, et al. Level of patient satisfaction with online psychiatric outdoor services. Consort Psychiatr 2023 Sep 29;4(3):23-32. [doi: 10.17816/CP5597] [Medline: 38249530]
29. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010;32(1):56-69. [doi: 10.1093/epirev/mxq004] [Medline: 20354039]
30. Head KJ, Noar SM, Iannarino NT, Grant Harrington N. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013 Nov;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
31. MacDougall S, Jerrott S, Clark S, Campbell LA, Murphy A, Wozney L. Text message interventions in adolescent mental health and addiction services: scoping review. JMIR Ment Health 2021 Jan 8;8(1):e16508. [doi: 10.2196/16508] [Medline: 33416504]
32. Free C, Phillips G, Galli L, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362. [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
33. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
34. Lim MSC, Wright C, Hellard ME. The medium and the message: fitting sound health promotion methodology into 160 characters. JMIR Mhealth Uhealth 2014 Nov 3;2(4):e40. [doi: 10.2196/mhealth.3888] [Medline: 25367387]
35. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20. [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
36. Hoermann S, McCabe KL, Milne DN, Calvo RA. Application of synchronous text-based dialogue systems in mental health interventions: systematic review. J Med Internet Res 2017 Jul 21;19(8):e267. [doi: 10.2196/jmir.7023] [Medline: 28784594]
37. Luxton DD, McCann RA, Bush NE, Mishkind MC, Reger GM. mHealth for mental health: integrating smartphone technology in behavioral healthcare. Professional Psychology: Research and Practice 2011;42(6):505-512. [doi: 10.1037/a0024485]

XSL•FO
RenderX

38. Shalaby R, Vuong W, Eboreime E, Surood S, Greenshaw AJ, Agyapong VIO. Patients' expectations and experiences with a mental health-focused supportive text messaging program: mixed methods evaluation. JMIR Form Res 2022 Jan 11;6(1):e33438. [doi: 10.2196/33438] [Medline: 35014972]
39. Senanayake B, Wickramasinghe SI, Chatfield MD, Hansen J, Edirippulige S, Smith AC. Effectiveness of text messaging interventions for the management of depression: A systematic review and meta-analysis. J Telemed Telecare 2019 Oct;25(9):513-523. [doi: 10.1177/1357633X19875852] [Medline: 31631764]
40. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13. [doi: 10.2196/jmir.1194] [Medline: 19403466]
41. Muñoz RF, Chavira DA, Himle JA, et al. Digital apothecaries: a vision for making health care interventions accessible worldwide. Mhealth 2018;4:18. [doi: 10.21037/mhealth.2018.05.04] [Medline: 30050914]
42. Firth J, Torous J, Nicholas J, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298. [doi: 10.1002/wps.20472] [Medline: 28941113]
43. Proudfoot J, Clarke J, Birch MR, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013 Nov 18;13:312. [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]
44. Lau CKY, Saad A, Camara B, Rahman D, Bolea-Alamanac B. Acceptability of digital mental health interventions for depression and anxiety: systematic review. J Med Internet Res 2024 Oct 28;26:e52609. [doi: 10.2196/52609] [Medline: 39466300]
45. Luo C, Sanger N, Singhal N, et al. A comparison of electronically-delivered and face to face cognitive behavioural therapies in depressive disorders: A systematic review and meta-analysis. EClinicalMedicine 2020 Jul;24:100442. [doi: 10.1016/j.eclinm.2020.100442] [Medline: 32775969]
46. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336. [doi: 10.1002/wps.20673] [Medline: 31496095]
47. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567. [doi: 10.2196/14567] [Medline: 31573916]
48. Smith DM, Duque L, Huffman JC, Healy BC, Celano CM. Text message interventions for physical activity: a systematic review and meta-analysis. Am J Prev Med 2020 Jan;58(1):142-151. [doi: 10.1016/j.amepre.2019.08.014] [Medline: 31759805]
49. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4. [doi: 10.2196/jmir.1376] [Medline: 20164043]
50. Daikeler J, Bošnjak M, Lozar Manfreda K. Web versus other survey modes: an updated and extended meta-analysis comparing response rates. J Surv Stat Methodol 2020 Jun 1;8(3):513-539. [doi: 10.1093/jssam/smz008]
51. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
52. Wang J, Mahe J, Huo Y, et al. Framework development for reducing attrition in digital dietary interventions: systematic review and thematic synthesis. J Med Internet Res 2024 Aug 27;26:e58735. [doi: 10.2196/58735] [Medline: 39190910]

Abbreviations

CBT: cognitive behavioral therapy
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture

Edited by A Kushniruk; submitted 27.Apr.2025; peer-reviewed by D Gratzer, TTO Kwok; revised version received 01.Dec.2025; accepted 14.Dec.2025; published 21.Jan.2026.

Please cite as:
Obuobi-Donkor G, Shalaby R, Agyapong B, Obeng Nkrumah S, Adu MK, Eboreime E, Wozney L, Agyapong VIO
Evaluating User Engagement and Satisfaction With Digital Mental Health Interventions: Randomized Controlled Trial of a Text Messaging Program and e-Mental Health Resources
JMIR Hum Factors 2026;13:e76587
URL: https://humanfactors.jmir.org/2026/1/e76587
doi:10.2196/76587

XSL•FO
RenderX

© Gloria Obuobi-Donkor, Reham Shalaby, Belinda Agyapong, Samuel Obeng Nkrumah, Medard Kofi Adu, Ejemai Eboreime, Lori Wozney, Vincent Israel Opoku Agyapong. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 21.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Designing a Substance Misuse Data Dashboard for Overdose Fatality Review Teams: User-Centered Design Approach

Marie Pisani[1], MD; Madeline K Oguss[2], MS; Julia Dickson-Gomez[3], PhD; Constance Kostelac[3], PhD; Amy Parry[3], MPH; Starr Moss[4], MS; Elizabeth Salisbury-Afshar[5], MD, MPH; Brian Patterson[6], MD, MPH; Michael Spigner[6], MD, NRP; Megan Gussick[6], MD; Alison Krautkramer[7], MS; Timothy Gruenloh[7], MS; Askar Safipour Afshar[2], MS; Preeti Gupta[2,8], MD, MPH; Anoop Mayampurath[2,7], PhD; Majid Afshar[2,7], MD, MS

[1]Department of Medicine, University of Wisconsin, Madison, WI, United States

[2]School of Medicine and Public Health, University of Wisconsin, 610 Walnut St, Suite 517, Madison, WI, United States

[3]Institute for Health and Equity, Medical College of Wisconsin, Milwaukee, WI, United States

[4]United States Department of Justice, Madison, WI, United States

[5]Department of Family Medicine and Community Health, University of Wisconsin, Madison, WI, United States

[6]BerbeeWalsh Department of Emergency Medicine, University of Wisconsin, Madison, WI, United States

[7]Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI, United States

[8]Division of Pulmonary, Critical Care, Sleep, and Allergy, University of Illinois Chicago, Chicago, IL, United States

Corresponding Author:
Majid Afshar, MD, MS
School of Medicine and Public Health, University of Wisconsin, 610 Walnut St, Suite 517, Madison, WI, United States

Abstract

Background: Overdose fatality review (OFR) is a public health process in which cases of fatal overdose are carefully reviewed to identify prevention strategies. Current OFR requires review of multiple unconnected data sources, which is a manually intensive process. The Substance Misuse Data Commons (SMDC) was created to link electronic health record data with data from local and state agencies into a single, cloud-based e-platform but does not currently have a data visualization tool.

Objective: We aimed to use human factors design principles to develop a comprehensive dashboard for the SMDC that could facilitate enhanced processes to support OFR.

Methods: We first surveyed OFR leaders in Wisconsin using the National Aeronautics and Space Administration-Task Load Index to understand the cognitive workload of 3 tasks: (1) analysis of population-level overdose trends, (2) selection and preparation of individual cases for review, and (3) abstraction of data from individual causes. We then conducted semistructured interviews to identify targets for workflow optimization. Next, we developed a prototype dashboard for evaluation using a synthetic dataset built with GPT-4. We subsequently performed iterative design sessions with heuristic evaluations and collected end-user feedback on the final prototype via a second round of semistructured interviews and targeted surveys, including the Unified Theory of Acceptance and Use of Technology and the Perceived Usefulness Questionnaire.

Results: The National Aeronautics and Space Administration-Task Load Index revealed a moderately high mental workload with the current workflow for all 3 tasks, with mean scores of 12.60 (SD 3.31), 11.90 (SD 3.57), and 12.43 (SD 5.41) for tasks 1, 2, and 3, respectively. Interviews pointed to causes including technological challenges and a reliance on manual processes. The prototype dashboard addressed these concerns by integrating multiple data sources to generate population-level visualizations and patient-level event timelines. End users reported the potential for improved efficiency and data accessibility compared to antecedent processes. The Unified Theory of Acceptance and Use of Technology results indicated the dashboard would likely be adopted if made available, with a mean of 4.07 out of 5.00 (SD 0.65). The Perceived Usefulness Questionnaire results suggested moderate usefulness for both the aggregate and individual-level data, with means of 3.61 (SD 0.82) and 3.64 (SD 0.85) out of 5.00, respectively.

Conclusions: OFR is a data-intensive process that traditionally demands substantial cognitive and manual effort, and there are multiple barriers to efficiently collecting data and presenting them for review. The dashboard offers a user-centered, informatics-based approach to streamline data aggregation and presentation, potentially enhancing the efficiency of case reviews. Implementing a dashboard that consolidates and visualizes disparate data sources has the potential to alleviate the manual workload in OFR. Ultimately, our aim is to deliver a finalized data dashboard with real-world SMDC data, giving OFR leaders additional tools to aid in their rigorous work shaping interventions to reduce overdose fatalities.

(JMIR Hum Factors 2026;13:e79407) doi:10.2196/79407

XSL•FO
RenderX

KEYWORDS

data linkage; overdose; public health informatics; qualitative research; substance use disorders

Introduction

Drug overdose death rates have been steadily rising over the past 2 decades, with the most significant annual spike occurring between 2019 and 2020, when rates surged by 31.0% in the United States [1]. To address this public health crisis, some local health departments around the country have assembled overdose fatality review (OFR) teams. OFR teams are multidisciplinary and multiagency teams with representatives from areas such as public health, safety, social services, medical examiners and coroners, emergency responders, substance use treatment providers, and other community stakeholders [2]. These teams discuss the local trends in overdose fatalities and review individual cases to identify and implement recommendations aimed at preventing overdose deaths [2]. OFR teams can help facilitate harm reduction strategies such as syringe services programs, naloxone education and training, outreach programs, and coordination of treatment services from health care settings [3]. However, many counties have yet to implement OFRs, and those that exist exhibit considerable heterogeneity in their data collection procedures, requiring substantial effort in data curation.

OFR teams in Wisconsin perform fatality reviews by integrating population- and individual-level data. These reviews involve an in-depth exploration of an individual's timeline before their death, focusing on potential opportunities for intervention. The data may come from different sources, including the medical examiners and coroners office, emergency medical services (EMS), the Department of Corrections, law enforcement, social media, and other local and state agencies. Often, obtaining these data requires collecting data on individual cases from each agency and linking across multiple sources.

The Substance Misuse Data Commons (SMDC) is a single, cloud-based data repository that links hospital electronic health record (EHR) data for patients with substance misuse to local and state agency data from EMS, the Department of Corrections, the Prescription Drug Monitoring Program, state and national death sources, statewide medical and pharmacy claims, and neighborhood-level socioeconomic data [4]. It was recently created by our team to address the issue of siloed datasets for substance misuse research, but it currently lacks an interface for data visualization [4]. Dashboards have been recognized as effective tools for visualizing public health data and facilitating disease surveillance, targeted analyses, and decision-making [5]. Additionally, a well-designed dashboard can reduce cognitive workload and improve efficiency by decreasing the amount of time spent gathering data [6]. In several states, such as Indiana and North Carolina, health departments have used dashboards of aggregate data to monitor overdose trends and uncover preventable risk factors [7,8]. While these dashboards are effective for tracking local and state-wide trends and setting case review priorities, they often lack the breadth of data and the level of detail needed by OFR teams. Currently, OFR teams in Wisconsin do not have access to the SMDC. Their data collection processes rely on manually requesting and compiling information from multiple partner agencies. The SMDC was established as a research infrastructure to enable secure multiagency data linkage, and this study represents the first effort to design a visualization interface that could make those linked data accessible and actionable for OFR teams in the future.

The objective of Phase 1 of this study was to assess the cognitive workload of OFR teams and their currently utilized data processes. The objective of Phase 2 of this study was to design a prototype data dashboard and then assess the usability and acceptability of the dashboard in a simulated study. This study introduces a novel, user-centered approach to understanding and improving the OFR process by combining cognitive workload assessment with prototype dashboard design. Unlike prior public health surveillance dashboards, our approach integrates multiagency data within a unified framework modeled on real OFR workflows and leverages synthetic data generation using large language models to enable privacy-preserving development and testing.

Methods

Phase 1: Analysis of Current Workflow

We conducted surveys and semistructured interviews with OFR leaders in Wisconsin to document the current workflow and associated cognitive workload for data collection and presentation. Participants were eligible if they took part in at least one of 3 tasks: (1) aggregating or analyzing population-level data, (2) selecting cases for case reviews, or (3) abstracting data for individual case reviews. Of the 29 counties with OFR teams, we recruited participants from 2 counties—one corresponding to the dataset currently included in our study, and the other in anticipation of future data collection for the SMDC. Recruitment occurred from organizational listservs in response to informational emails sent by local OFR leadership, followed by an invitation to the voluntary and confidential survey. Participants provided consent and were enrolled via a link to the survey in a secure, web-based Research Electronic Data Capture (REDCap; Vanderbilt University) database [9,10].

We used the National Aeronautics and Space Administration-Task Load Index (NASA-TLX) survey to assess the cognitive workload of the contemporary OFR process for each task listed above that the OFR leader participated in [11]. The NASA-TLX is a validated, multidimensional tool that assesses a task's subjective cognitive workload across 6 dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration [11]. The survey asks participants to rate each dimension on a scale between 1 and 20, with higher scores indicating greater perceived workload. For the performance dimension, lower scores correspond to higher perceived success [11]. Example items include "How mentally demanding was the task?" and "How hurried or rushed was the pace of the task?" Participants rated each applicable task they performed during the OFR process.

XSL•FO
RenderX

Domain-specific mean scores were calculated rather than a composite overall score, as the NASA-TLX dimensions are designed to capture distinct aspects of workload rather than a single latent construct. Therefore, the mean and SD for each dimension were calculated.

Survey participants were then invited to a 30-minute semistructured interview with a research team member (MP). This interview approach combined a predefined structure with flexibility for follow-up questions based on participants' responses [12]. The interview guide focused on participants' responsibilities, workflows, mental workload, key data sources, and challenges. Survey findings were shared, and participants were asked to provide insight on the results. Wisconsin OFR training and technical assistance providers and a senior qualitative research expert (JDG) reviewed the guide to ensure its relevance and rigor. Interviews were conducted between October 2023 and January 2024 with participants providing verbal informed consent through secure, virtual conference software. The audio recording was transcribed and reviewed by the interviewer.

Interviews were stored and analyzed using MAXQDA 2024 (VERBI Software). The constant comparative method was used to analyze the transcripts, and inductive coding was applied to organize the information into emergent categories [13]. The codebook was repeatedly revisited and revised during the process [13]. When no new categories emerged from the analysis of additional interviews, code saturation was determined to be met. Using the final codebook, all interviews were coded by 1 researcher. Around 3 interviews were coded by a second researcher to assess intercoder reliability (ICR), exceeding the typical 10% - 25% double-coded interviews required to establish ICR [14]. The 2 researchers then compared coded segments, and disagreements were adjudicated.

Phase 2: Data Dashboard Design and Evaluation

To mitigate data privacy concerns during dashboard development, a synthetic dataset was patterned on SDMC data [15]. The synthetic data were generated using the GPT-4 application programming interface (OpenAI) [16], with chain-of-thought instructions to create each variable from the data dictionary, similar to other best practices in prompt engineering [17]. The prompt incorporated aggregate cohort demographic descriptive statistics to preserve variable distributions within the SMDC dataset. Initial prompts were tested with 20 rows of patient data to evaluate the quality of the output before a dataset of 300 patients was created. This dataset was then scaled up to 1273 patients using YData Fabric (YData AI), a synthetic data generation software that employs generative adversarial networks to produce large volumes of data accurately replicating the statistical characteristics of the original data—in this case, the smaller synthetic dataset [18].

An initial prototype of the dashboard was developed using Microsoft Power Business Intelligence (Microsoft 2024) [19]. The dashboard was designed to emulate the workflow and highlight the priority data sources identified in Phase 1. The dashboard consisted of 3 functional components: (1) visualizations of population-level data to identify demographics and trends (Figure 1); (2) line-level data, such as individual patient timelines, to facilitate case-based reviews (Figure 2); and (3) prediction tools, including census tract-level EMS patient incident predictions, deidentified hospital note topics, and a 30-day risk score for hospital readmission or death.

We conducted 2 iterative design sessions with emergency medicine physicians, including 2 EMS medical directors, a prehospital informatician, and a clinical human factors design expert. This team provided expertise in data visualization principles and linking hospital systems with prehospital emergency services, while allowing us to reserve OFR leaders for participation in the main study. The Heuristic Evaluation Checklist for Dashboard Visualizations was used to identify and address major usability issues [20]. These sessions refined the dashboard's content, organization, and visual elements, culminating in a final prototype. After finalizing the prototype, a demonstration video was shared with participants to showcase its content, organization, and key features.

XSL•FO
RenderX

Figure 1. Substance misuse data dashboard: population-level visualizations.

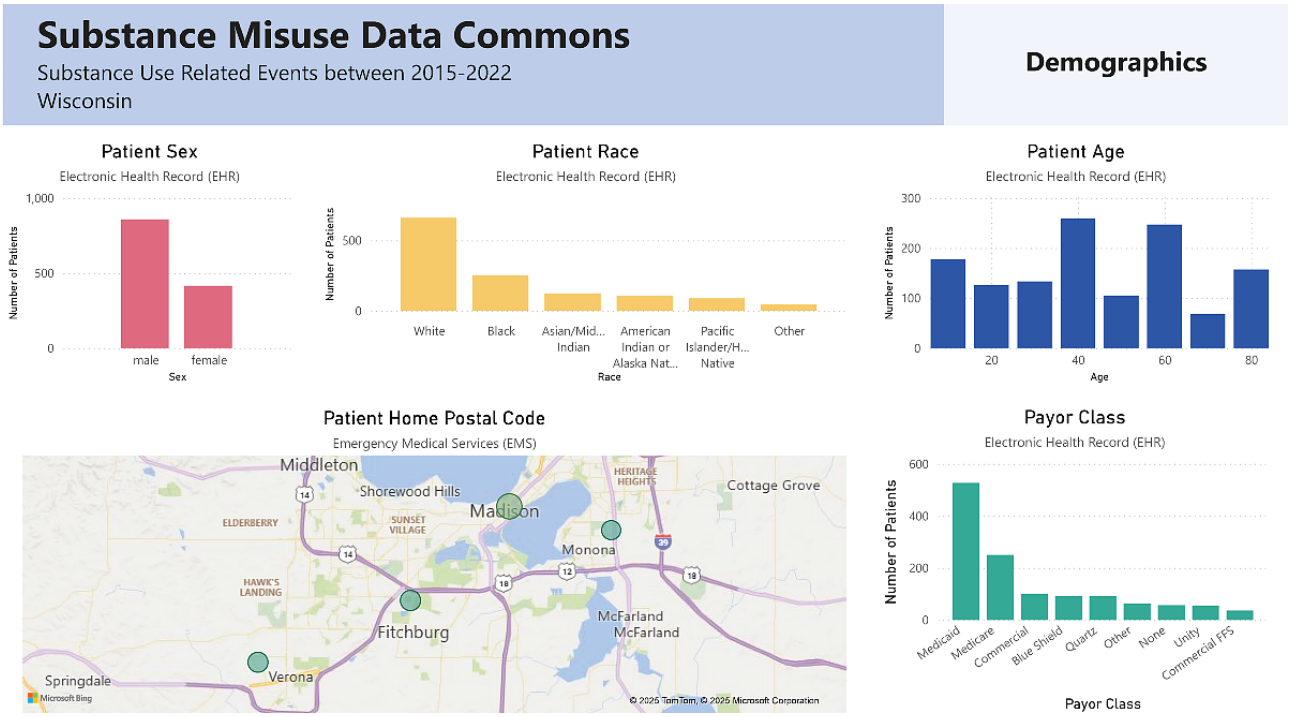
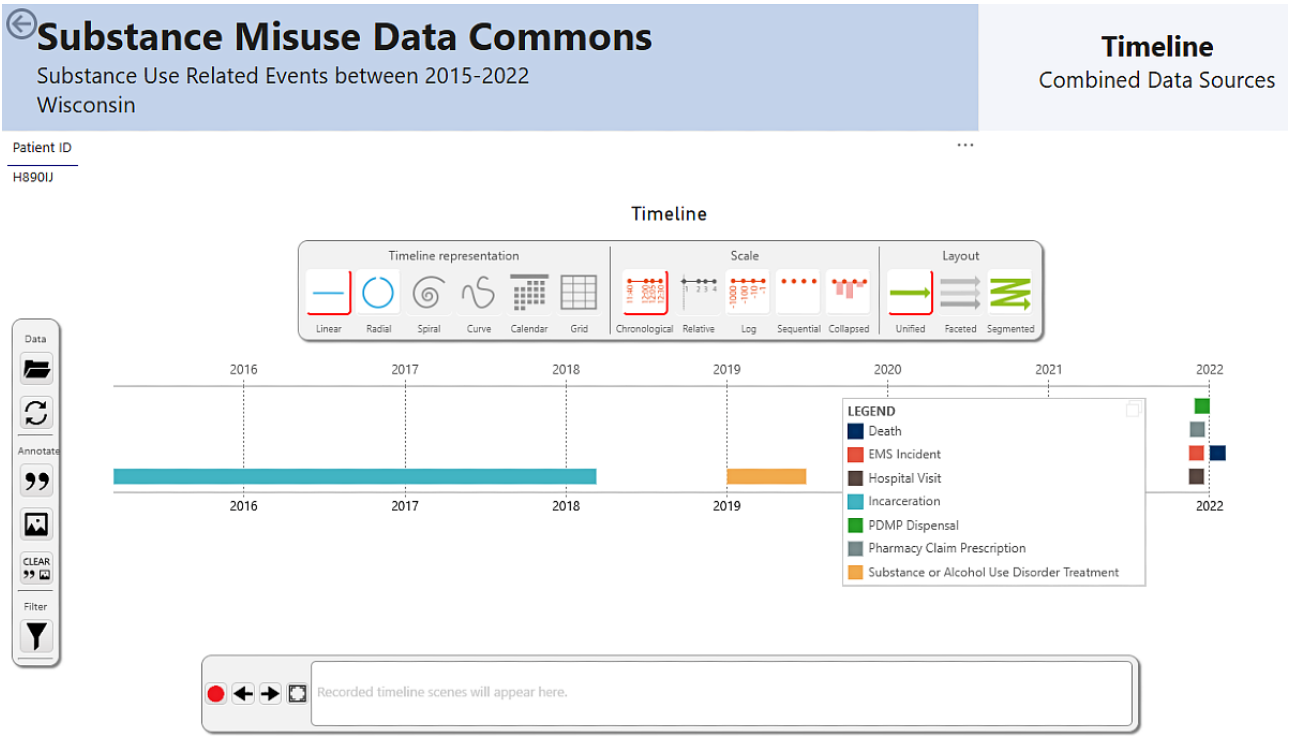


Figure 2. Substance misuse data dashboard: patient timeline visualizations.



Phase 2: Dashboard Design and Evaluation

End-user perceptions of the final prototype were assessed with semistructured interviews and surveys, using the same strategy and procedures to recruit OFR leaders as Phase 1. A total of 2 validated survey tools were used to evaluate end-user probability of adoption (Unified Theory of Acceptance and Use of Technology [UTAUT]) [21] and perception of usefulness (Perceived Usefulness Questionnaire) [22]. The interview guide explored participants' perceptions of the dashboard, conditions for use, advantages, and suggested changes. One researcher coded all 7 interviews using the final codebook, and a second researcher independently coded 3 interviews to establish ICR. Disagreements were adjudicated. The mean and SD of all survey responses were calculated.

The full set of NASA-TLX, UTAUT, and Perceived Usefulness Questionnaire questions; our interview guides and codebooks; our synthetic dataset and prompts; and our demonstration video are all viewable in our GitLab repository [23].

XSL•FO
RenderX

Ethical Considerations

The study followed the COREQ (Consolidated Criteria for Reporting Qualitative Research) reporting guidelines [24] (Checklist 1). This research was reviewed and approved by the Institutional Review Board at the University of Wisconsin, Madison (Institutional Review Board number 2023‐1091). Informed consent verbiage was included in our recruitment emails, and verbal consent was obtained during interviews. Participant email addresses were collected in the survey in order to invite the individual to participate in an interview. Survey and interview results were deidentified. No compensation was provided to participants.

Results

Phase 1: Analysis of Current Workflow

Surveys

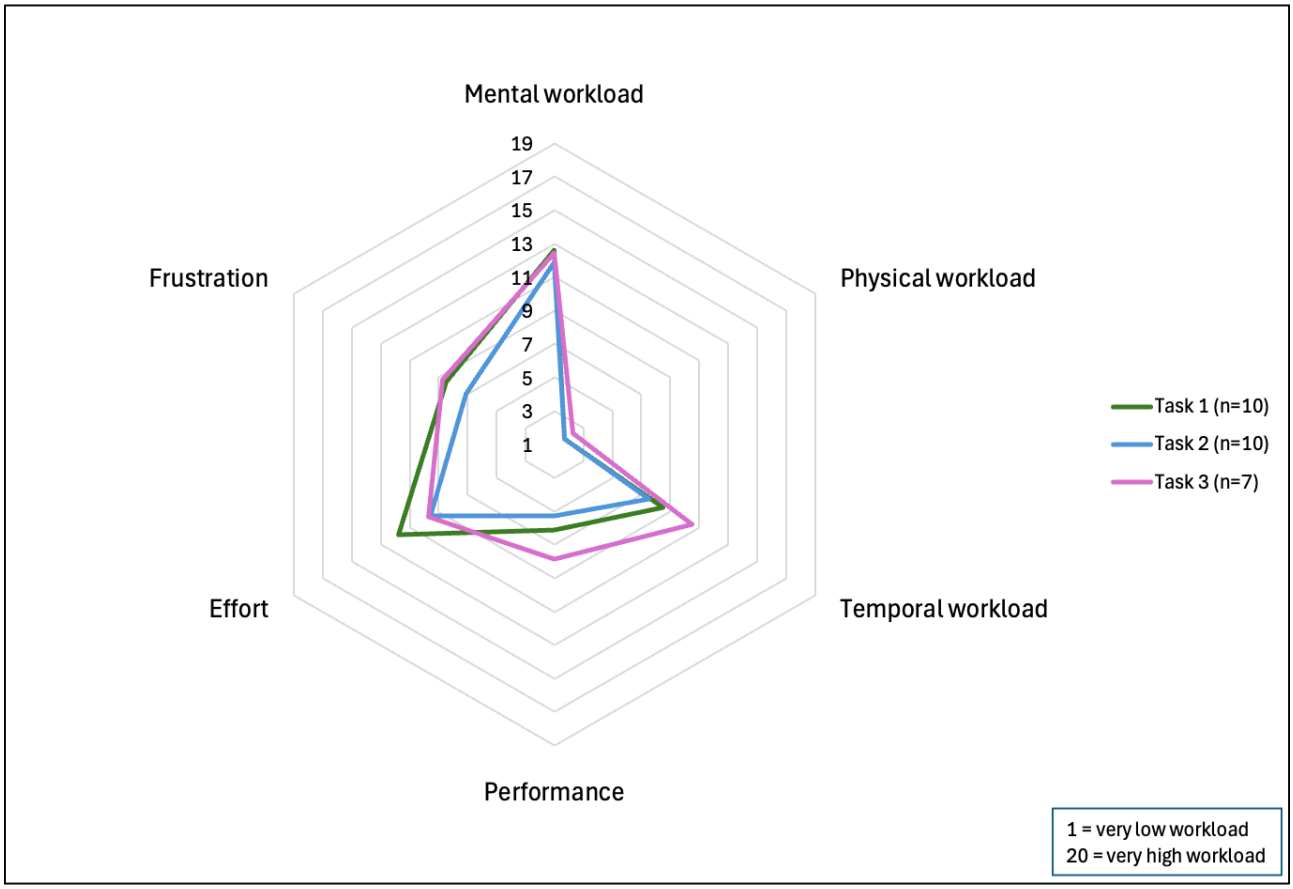
A total of 11 OFR leaders, representing both county- and state-level agencies, participated in the survey. Table 1 provides the characteristics of the participants. The survey assessed the cognitive workload of 3 distinct tasks: (1) aggregating or analyzing population-level data, (2) selecting cases for case reviews, or (3) abstracting data for individual case reviews. Tasks had unequal sample sizes due to variations in task participation among participants. High mental workload was reported across all tasks (Figure 3). The time required for each task varied among participants. Per case review period of 1-3 months depending on the team, aggregating population-level data took an average of 5.5 hours (SD 3.09), selecting cases for case review averaged 7.95 hours (SD 5.67), and abstracting data for individual case reviews required an average of 10.50 hours (SD 8.86).

Table . Phases 1 and 2 survey demographics for overdose fatality review leaders in Wisconsin, 2023‐2024.

Demographic variables	Phase 1 participants (n=11)	Phase 2 participants (n=6)
Age (y), n (%)		
20‐29	5 (45.4)	2 (33.3)
30‐39	3 (27.3)	0 (0)
40‐49	2 (18.2)	3 (50)
50‐59	1 (9.1)	1 (16.7)
Sex, n (%)		
Female	10 (90.9)	6 (100)
Race, n (%)		
White	11 (100)	6 (100)
Education, n (%)		
Technical school, vocational training, community college	1 (9.1)	0 (0)
Bachelor’s degree	1 (9.1)	1 (16.7)
Master’s degree	9 (91.8)	5 (83.3)
Sector, n (%)		
Public health	9 (81.8)	6 (100)
Government	1 (9.1)	0 (0)
Education	1 (9.1)	0 (0)

XSL•FO
RenderX

Figure 3. National Aeronautics and Space Administration-Task Load Index (NASA-TLX) survey results for overdose fatality review leaders in Wisconsin, 2023-2024.



Interviews

Key Data Sources

In total, 10 of the 11 participants completed qualitative follow-up interviews. Participants were asked to describe their data collection process, case selection methods, challenges, desires for a future state, and key data sources. Key OFR data sources highlighted by participants are presented in Table 2. Some of these sources were used for individual or population-level data only, but many were used for both. How these sources were used varied among participants. Summaries and selected quotes are highlighted below, while the full interview results and quotes are detailed in Multimedia Appendix 1.

XSL•FO
RenderX

Table . Key data sources for overdose fatality review leaders in Wisconsin, 2023 - 2024.

Category	Data contributors and resources
High-impact data contributors	1. State vital records 2. Next-of-kin interviews 3. OFR[a] agency partners
Key agency partners and contributors	1. ME's[b] or coroner's office 2. EMS[c] 3. DOC[d] 4. Law enforcement
Other data sources	1. Prescription Drug Monitoring Program 2. Electronic Surveillance System for the Early Notification of Community-Based Epidemics 3. First Watch (fire and rescue) 4. Wisconsin Statewide Health Information Network 5. Overdose Detection Mapping Application Program 6. Consolidated Court Automation Programs 7. Internet 8. News 9. Social media

[a]OFR: overdose fatality review.

[b]ME: medical examiner.

[c]EMS: emergency medical services.

[d]DOC: Department of Corrections.

Process

Participants described a typical workflow while preparing for an OFR meeting. Population-level data are analyzed to identify recent community trends. Next, representative cases are selected for review, and permission is obtained from local law enforcement to review the cases. OFR leaders then coordinate the collection of available and pertinent information on the decedents across multiple sources, which may include their own sources, agency partners, and next-of-kin interviews with the decedent's family or friends. After compiling the data, they prepare a presentation for the case review meeting, often including a timeline of the decedent's interactions with various agencies. During the sessions, OFR members, agency partners, and community representatives review the data and collaboratively brainstorm strategies for overdose prevention.

Case Selection

Participants reported several factors that influence case selection. About 50% (n=5) of the participants reported looking at demographic trends such as age, race, and sex. About 50% (n=5) of the participants reported combining multiple demographic and substance trends into a single theme and looked explicitly at decedents within that theme. For example, 1 participant stated, "Now that we're doing theme selection, we may focus on a specific drug, like, I think the next theme that we're doing is African American men between certain ages that historically used cocaine, but fentanyl was also involved in their cause of death." Around 60% (n=6) of the participants prioritize selecting cases with comprehensive data, though they noted that this is challenging due to limited data availability during the initial selection process. Other key factors influencing case selection included obtaining permission from agency partners (n=5, 50%) and ensuring cases fell within jurisdictional boundaries (n=9, 90%).

Case Data Collection and Preparation

OFR leaders reported collecting data on decedents using publicly available databases and resources provided by their health departments and from the state-level data provided. They also reported requesting data from agency partners regarding any interactions with the decedent. After data are collected, they are processed manually by the OFR leader. One participant stated, "Then those individual partners have to go in, look at the specific case and then they have to like hand-put in all of the info and then they send those to me and then I scan them into our system so that we have them electronically, and then I have to take all of those electronic copies and upload those, one question at a time into REDCap." After processing, all participants reported compiling the information into a timeline to display during case review presentations, which helped viewers to understand the decedent's story.

Reported Challenges

When asked about challenges, 90% (n=9) of the participants identified a reliance on manual processes to collect data as a significant challenge. Due to limited bandwidths, responses to data requests from agency representatives are often delayed or incomplete, which impacts the preparation of case review materials. When final requests were not fulfilled, critical data were missing from presentations. Participants also reported several technological challenges, including siloed data sources (n=4, 40%), confusing data formatting (n=7, 70%), and other technological issues (n=10, 100%), all of which impacted data collection and preparation. As a result, 90% (n=9) of the participants highlighted time pressure as a major challenge for

XSL•FO
RenderX

preparing for the OFR process. One participant explained this time pressure stating, "Lots of times for overdose fatality review, and this is true for me also, this is one part of my job, right? It's part of my FTE, it is not my full FTE, so there are directions that I'm pulled for other projects that can sometimes limit the time that I have available to work on this."

Desired Future State

Half (n=5) of the participants indicated that easier access to current data sources would be helpful, specifically mentioning simplified access as well as fewer lags to be able to identify and respond to current trends. One participant stated, "I think something that local public health and us specifically have always sort of struggled with is being able to keep up with that data…having it be a little easier for us to get local data and more quickly, that is an issue that is huge for us. We see things, and we hear about these trends, but we don't necessarily always know that that's happening until all of a sudden, it's like, hey, we're seeing this, you know, all across the county, and is that something we could have caught sooner had we been able to access that data quicker."

Participants mentioned that additional data would help them. About 70% (n=7) of the participants desired health care data, most commonly substance use disorder treatment data; however, these data are protected by federal statute, which adds complexity to accessing and sharing them. About 50% (n=5) of the participants desired criminal justice data, whether they did not have access to it or did not often receive it when requested. Participants mentioned that they worked with multiple law enforcement agencies and regularly received data from some but not others. Additionally, 2 participants mentioned that they had not been able to perform next-of-kin interviews due to barriers in setting up interviews. Other desired data sources included childhood information from school districts or Child Protective Services and input from local organizations.

About 80% (n=8) of the participants mentioned that more collaboration or support would help the OFR process. Other participants specifically mentioned that increased collaboration between health departments and local agencies would help their current workflow, create the possibility to expand their services, and address long-term sustainability of OFR.

Phase 2: Data Dashboard Design and Evaluation

Dashboard Design

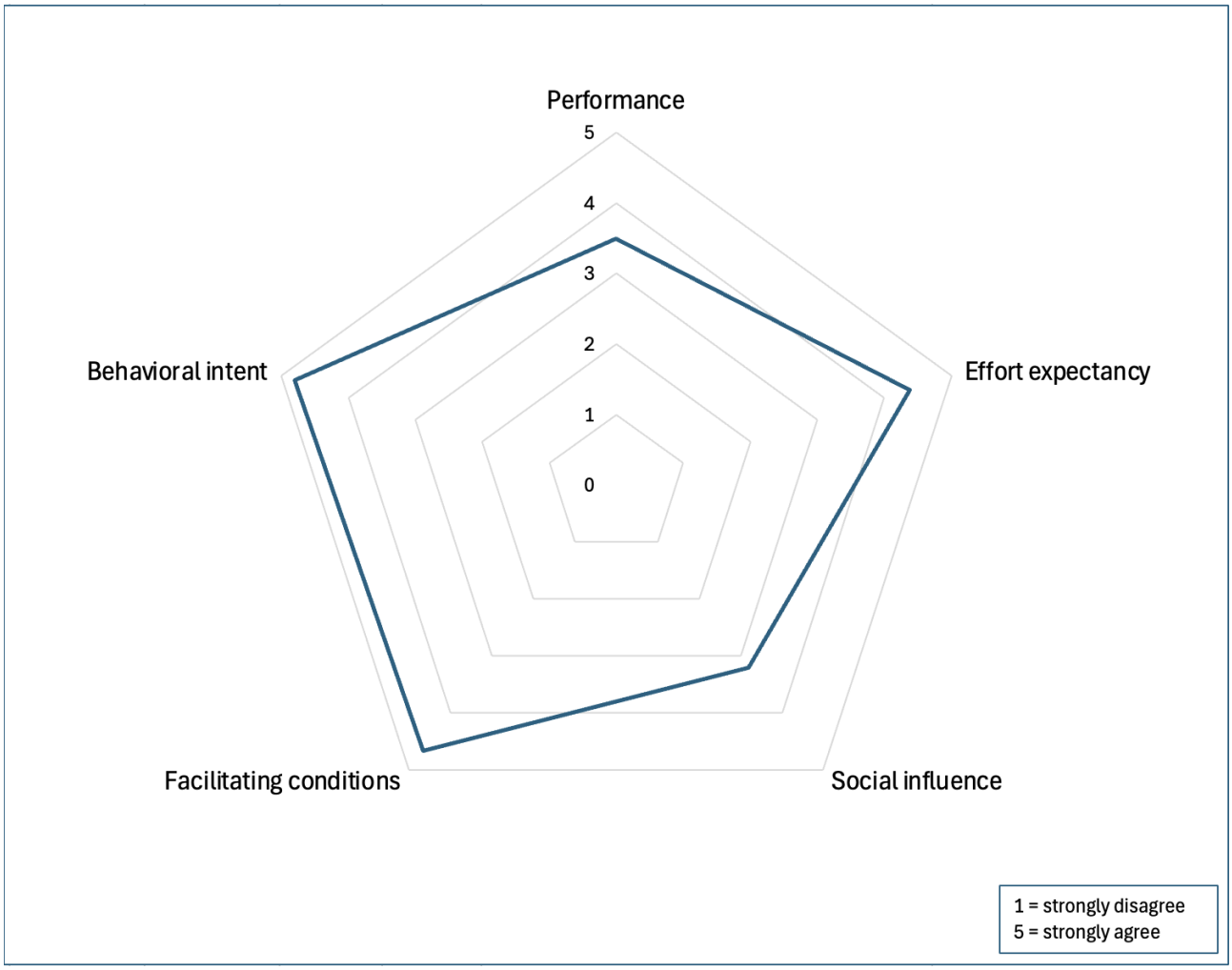
The synthetic data dashboard prototype, developed based on Phase 1 data, was refined into a high-fidelity prototype for further usability and human factors evaluation. The final prototype featured 9 theme-based aggregate data pages covering demographics, substances used, health care interactions, prehospital emergency services, social and economic factors, mortality, prescription patterns, and treatment and recovery. Select components of the line-level data and timeline are shown in Figures 1 and 2, and a full demonstration video is available in our GitLab repository [23]. The final dashboard prototype included advanced filtering capabilities, enabling users to refine data by specific time frames, substances, and death status. A total of 3 machine learning tools were integrated into the dashboard. First, the Hospital Note Topics Tool utilized latent Dirichlet allocation for topic modeling of EHR notes [25]. This tool identified prevalent themes and trends, such as patterns of substance use, health care utilization, and social determinants of health, providing users with a high-level understanding of key insights from unstructured text data. Second, the 30-Day Risk Score for Readmission and Death employed an eXtreme Gradient Boost machine learning model, which analyzed a combination of EHR notes and tabular data along with EMS and neighborhood census data to predict the likelihood of hospital readmission or death within 30 days. Third, the EMS Geographic Prediction Tool combined EMS response data with neighborhood-level census tract information to identify geographic areas at higher risk for overdose events. Additionally, the final prototype incorporated a drill-down feature, allowing users to filter patient populations by category and narrow them down to individual patients.

Surveys

A total of 6 OFR organizers participated in the Phase 2 survey, representing stakeholders from county, state, and federal agencies. Additional demographic details are provided in Table 1. All the 6 participants reported analyzing aggregate data and selecting review cases, while 4 were involved in abstracting data for case reviews. The UTAUT results indicated that the dashboard would likely be adopted if made available to participants (Figure 4), with a mean of 4.07 out of 5.00 (SD 0.65). The Perceived Usefulness Questionnaire results suggested a moderately positive perception of usefulness for the aggregate and individual-level data (Figure 5), with means of 3.61 (SD 0.82) and 3.64 (SD 0.85) out of 5.00, respectively.

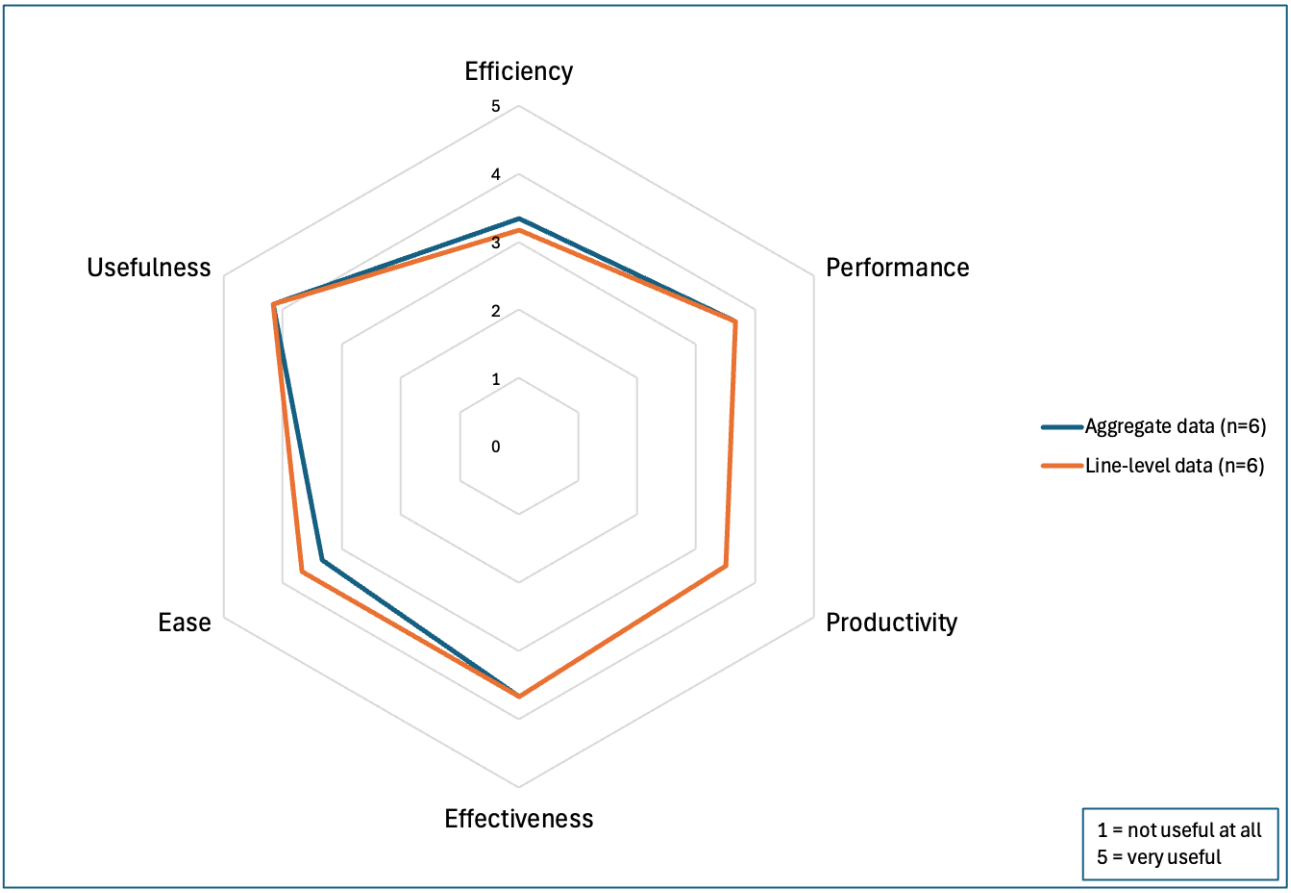
XSL•FO
RenderX

Figure 4. Unified Theory of Acceptance and Use of Technology (UTAUT) survey results for overdose fatality review leaders in Wisconsin, 2024.



XSL•FO
RenderX

Figure 5. Perceived usefulness questionnaire results for overdose fatality review leaders in Wisconsin, 2024.



Interviews

A total of 7 OFR leaders participated in semistructured interviews. Participants were specifically asked about the potential benefits of the dashboard as well as areas for improvement. The summaries of the findings are outlined below, with detailed results and quotes provided in Multimedia Appendix 2.

Benefits of the Dashboard

All the participants (n=7) highlighted improved data access as a key benefit of the dashboard. This included expanding access to currently unavailable data sources, facilitating quicker and easier access to existing data, and increasing access for less-resourced communities. Additionally, 86% (n=6) of the participants indicated that the dashboard could help to optimize workflows by reducing the manual processes and introducing time-saving features. One participant stated, “That’s helpful in just kind of like streamlining it all in one spot, because the data that I do get like from emergency department visits, I have to go through every single entry, and then like, figure it out from there. So, this is quite nice to just have it in one spot where I can look at it, write it down, we’re done.” About 71% (n=5) of the participants noted that the dashboard’s tools are especially valuable for organizational outreach efforts. The machine learning tools, particularly the EMS geographic prediction tool and the 30-day risk score for readmission or death, were highlighted as most useful for these initiatives. All the participants (n=7) reported that the dashboard appeared easy to navigate, with an intuitive structure and organization. About 71% (n=5) of the participants specifically mentioned the ability to filter patients for specific populations as a standout feature, distinguishing the dashboard from other available tools.

Areas for Improvement

Most participants suggested adding more data sources. Specific recommendations included sexual orientation and gender identity, medical comorbidities such as chronic pain, and presenting data as rates instead of counts to better represent minority group trends. Concern about the accuracy and timeliness of the data was expressed by 71% (n=5) of the participants. The SMDC cohort, which is limited to patients with hospital encounters linked to other data, excludes individuals seen only by EMS or those without EMS or hospital contact. One participant expressed this concern, stating, “I like the way that it’s laid out, but because it’s only people who are going to the ED, I don’t know that we can draw conclusions about folks that are at risk of overdose, generally speaking.” The current data were also identified as a critical need for the dashboard. Participants proposed several technological enhancements to improve the dashboard’s usability. Specifically, 43% (n=3) of the participants suggested the ability to export data from the dashboard, while 57% (n=4) recommended adding other technological features such as additional filters and hover-over tips for ease of use. Other important but less commonly cited concerns included a training or learning curve (n=3, 43%), having jurisdictional access (n=2, 29%), adding to stigma or bias (n=2, 29%), having too many years aggregated in the dataset to be reflective of current trends (n=2, 29%), and the dashboard data being deidentified and therefore not being

XSL•FO
RenderX

able to be connected to data about individuals from other sources (n=2, 29%).

Discussion

Principal Findings

This study is the first to systematically assess the cognitive workload of OFR leaders and apply those findings to the design of a user-centered data dashboard prototype. Using validated instruments (NASA-TLX, UTAUT, and Perceived Usefulness Questionnaire) and qualitative interviews, we identified substantial cognitive demands associated with data aggregation, case selection, and abstraction across multiple agencies. Guided by these findings, we developed and tested a prototype dashboard using synthetic data modeled on the SMDC to streamline data access and visualization. Participants reported improved usability, reduced time burden, and high potential for adoption in future OFR workflows.

Interpretation and Implications

Our results highlight that OFR leaders face high mental and temporal demands due to fragmented data sources and reliance on manual processes, a challenge consistent with prior studies on multiagency data integration and decision-making in public health surveillance [26]. Previous dashboards have been implemented at the state or regional level to visualize aggregate overdose trends, yet few, if any, have incorporated both population- and individual-level data or been explicitly designed around the OFR workflow. By linking health system, emergency response, and correctional datasets within a privacy-preserving synthetic environment, our approach extends beyond traditional epidemiologic dashboards to directly support operational review teams. The integration of predictive modeling tools, such as 30-day readmission and EMS geographic risk scoring, further distinguishes this work from prior descriptive dashboards by introducing prospective, data-driven decision support [7,8]. These findings align with recent public health informatics initiatives emphasizing human-centered design, cognitive workload reduction, and automation to improve decision-making efficiency [27,28].

Our SMDC data dashboard serves as a comprehensive tool for extracting, transforming, and visualizing overdose data. It was designed to refresh with recent case information from multiple sources; incorporate automation for case matching, data formatting, and quality checks; and offer easy navigation to streamline current workflows. The dashboard helps automate the current workflow challenge of siloed data sources and includes additional data that participants found valuable. It integrates key data variables from multiple agencies, with options to filter the data by important demographic and substance-related factors. The automated timeline feature compiles data from all sources in our dataset, visually representing the events leading to an overdose death. Phase 2 participants saw the dashboard's potential to reduce time pressure and reliance on manual processes. Additionally, in larger cities where reviewing every case is impractical or impossible, OFR leaders typically hand-pick cases to analyze in detail. By linking all cases and presenting the data in aggregate, this system enhances the understanding of overall trends and helps to better inform recommendations.

Many participants indicated that they would use this dashboard as an additional tool, rather than replacing their current methods. This likely explains our Phase 2 survey findings, which revealed a discrepancy between the tool's effectiveness and efficiency. This preference highlights a key concern raised in Phase 2 interviews that cannot be resolved in future dashboard versions: the deidentified nature of the SMDC data prevents integration with other data sources, such as next-of-kin interviews, a crucial part of the case review process. Other areas of improvement discussed in Phase 2 interviews can be resolved.

Phase 1 participants emphasized a desire for increased collaboration and support. Some mentioned this in the form of improved relationships between the health department and other agencies. Expanding the sectors involved in OFR processes and broadening the information available for case reviews may highlight previously unseen gaps in care. Emergency department utilization is common among those who misuse opioids and other drugs, and the number of emergency department visits is associated with an increased risk of drug overdose [29]. Our dashboard aims to establish and improve data sharing between OFR teams and health care systems, which has been identified as an important prevention strategy implementation [30].

Our findings underscore the potential of informatics-driven tools to enhance collaboration among public health, health care, and community agencies. By reducing cognitive workload and manual data handling, such tools can accelerate case review preparation, standardize data access across jurisdictions, and allow OFR teams to focus more on interpretation and prevention strategies. The use of synthetic data generated through large language models also represents an innovative method for tool development when working with sensitive, multiagency datasets. As counties nationwide expand their OFR infrastructure, the approach demonstrated here offers a transferable model for building and evaluating data dashboards that are secure, scalable, and adaptable to local contexts.

Limitations

The general process reported by participants in the interviews is mainly consistent with the process in the OFR Practitioner's Guide and the Public Health and Safety Team toolkit, which are guiding frameworks for health departments when creating OFRs and holding case review meetings [31]. However, the feedback in this study was collected from local OFR leaders in Wisconsin, and therefore, these findings may not be generalizable to other systems. Limitations inherent to the SMDC include its inclusion criteria and its deidentified nature. All data variables included in the prototype dashboard are included in the SMDC; however, the SMDC does not include all data variables important to OFR leaders when presenting cases. Finally, addressing challenges in the OFR process with new technology may not ultimately lead to better outcomes. Substance misuse is both a complex medical condition and an evolving public health issue. In order to make meaningful progress and improve outcomes, sustained collaboration across health care systems, public health agencies, and communities

XSL•FO
RenderX

will be essential to reform policies, reduce disparities, and improve medical care.

The existing OFR process is built on a thorough, team-based approach, but it includes several cognitively demanding tasks, and there are multiple challenges to timely data preparation. Increased collaboration, access to standard, centralized tools, and comprehensive data could build upon the rigorous work already being done by OFR teams in order to further augment and automate workflows to reduce manual work. We designed a user-centered data dashboard to help reduce the cognitive workloads identified from surveys and incorporate the desired data sources and workflows gathered from the interviews. Evaluative feedback indicated many potential benefits as well as some areas for improvement. This insight will guide the development of a real-time data dashboard accessible to OFR leaders in their review process.

Conclusions and Broader Implications

Despite these limitations, this study demonstrates a replicable, human-centered approach for modernizing OFR workflows through data integration and visualization. The combination of cognitive workload analysis, synthetic data modeling, and dashboard-based decision support provides a foundation for scalable OFR modernization across states. Beyond overdose prevention, this framework illustrates how human factors and data science methods can be combined to enhance other multidisciplinary public health review processes. Future work will focus on deploying the dashboard with live SMDC data, expanding data sources, and assessing real-world impacts on timeliness, data completeness, and prevention outcomes.

Funding

This work was supported by the National Institutes of Health, National Institute on Drug Abuse (R01DA051464), and the Department of Medicine of the University of Wisconsin School of Medicine and Public Health (233-AAM7146). The content is solely the responsibility of the authors and does not necessarily represent the official views of the UW-Madison Department of Medicine.

Data Availability

The full set of National Aeronautics and Space Administration-Task Load Index, Unified Theory of Acceptance and Use of Technology, and Perceived Usefulness Questionnaire questions; our interview guides and codebooks; our synthetic dataset and prompts; and our demonstration video are all viewable in our GitLab repository [23].

Authors' Contributions

Conceptualization: MKO, MP, AM, MA
Data curation: AM, MA, MKO, MP
Formal analysis: MP, AM, AP, CK, JDG, MA, MKO
Funding acquisition: AM, MA, MKO
Investigation: MKO, MP
Methodology: MP, MKO, JDG, CK, AP, ESA, BP, MS, MG, AM, MA
Project administration: AM, MA, MKO, MP
Resources: AM, AP, BP, CK, ESA, JDG, MA, MS, MG, MKO
Software: AK, AM, AP, ASA, CK, JDG, MA, MKO, PG, TG
Supervision: AM, MA, MKO
Validation: AM, AP, CK, ESA, JDG, MA, MS, MG, BP
Visualization: AK, AM, ASA, BP, MA, MG, MKO, MS, MP, PG, TG
Writing – original draft: AM, MA, MKO, MP
Writing – review & editing: MP, MKO, JDG, CK, AP, SHM, ESA, BP, MS, MG, AK, TG, PG, ASA, AM, MA

Conflicts of Interest

None declared.

Multimedia Appendix 1
Phase 1 interview results for overdose fatality review leaders in Wisconsin, 2023-2024.
[DOCX File, 21 KB - humanfactors_v13i1e79407_app1.docx]

Multimedia Appendix 2
Phase 2 interview results for overdose fatality review leaders in Wisconsin, 2024.
[DOCX File, 21 KB - humanfactors_v13i1e79407_app2.docx]

Checklist 1

XSL•FO
RenderX

COREQ checklist.
[PDF File, 378 KB - humanfactors_v13i1e79407_app3.pdf]

References

1. Spencer MR, Garnett MF, Miniño AM. Drug overdose deaths in the United States, 2002–2022. : National Center for Health Statistics; 2023 URL: https://stacks.cdc.gov/view/cdc/135849 [accessed 2024-06-20]
2. Robinson A, Christensen A, Bacon S. From the CDC: the prevention for states program: preventing opioid overdose through evidence-based intervention and innovation. J Safety Res 2019 Feb;68:231-237. [doi: 10.1016/j.jsr.2018.10.011] [Medline: 30876516]
3. Haas E, Truong C, Bartolomei-Hill L, Baier M, Bazron B, Rebbert-Franklin K. Local overdose fatality review team recommendations for overdose death prevention. Health Promot Pract 2019 Jul;20(4):553-564. [doi: 10.1177/1524839918797617] [Medline: 30205737]
4. Afshar M, Oguss M, Callaci TA, et al. Creation of a data commons for substance misuse related health research through privacy-preserving patient record linkage between hospitals and state agencies. JAMIA Open 2023 Dec;6(4):ooad092. [doi: 10.1093/jamiaopen/ooad092] [Medline: 37942470]
5. Schulze A, Brand F, Geppert J, Böl GF. Digital dashboards visualizing public health data: a systematic review. Front Public Health 2023;11:999958. [doi: 10.3389/fpubh.2023.999958] [Medline: 37213621]
6. Khairat SS, Dukkipati A, Lauria HA, Bice T, Travers D, Carson SS. The impact of visualization dashboards on quality of care and clinician satisfaction: integrative literature review. JMIR Hum Factors 2018 May 31;5(2):e22. [doi: 10.2196/humanfactors.9328] [Medline: 29853440]
7. Salvi A, Gillenwater LA, Cockrum BP, et al. Development of a real-time dashboard for overdose touchpoints: user-centered design approach. JMIR Hum Factors 2024 Jun 11;11:e57239. [doi: 10.2196/57239] [Medline: 38861717]
8. Fliss MD, Cox ME, Proescholdbell S, Patel A, Smith M. Tying overdose data to action: North Carolina's opioid and substance use action plan data dashboard. J Public Health Manag Pract 2023;29(6):831-834. [doi: 10.1097/PHH.0000000000001796] [Medline: 37498535]
9. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
10. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
11. Hart SG. NASA-task load index (NASA-TLX); 20 years later. Proc Hum Factors Ergon Soc Annu Meet 2006 Oct;50(9):904-908. [doi: 10.1177/154193120605000909]
12. Kallio H, Pietilä AM, Johnson M, Kangasniemi M. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. J Adv Nurs 2016 Dec;72(12):2954-2965. [doi: 10.1111/jan.13031] [Medline: 27221824]
13. Hewitt-Taylor J. Use of constant comparative analysis in qualitative research. Nurs Stand 2001;15(42):39-42. [doi: 10.7748/ns2001.07.15.42.39.c3052] [Medline: 12212430]
14. O'Connor C, Joffe H. Intercoder reliability in qualitative research: debates and practical guidelines. Int J Qual Methods 2020 Jan 1;19:1609406919899220. [doi: 10.1177/1609406919899220]
15. Pollack AH, Simon TD, Snyder J, Pratt W. Creating synthetic patient data to support the design and evaluation of novel health information technology. J Biomed Inform 2019 Jul;95:103201. [doi: 10.1016/j.jbi.2019.103201] [Medline: 31078659]
16. GPT-4 [technical report]. : OpenAI; 2023 URL: https://openai.com/research/gpt-4 [accessed 2025-12-10]
17. Wei J, Bosma M, Zhao VY, et al. Finetuned language models are zero-shot learners. arXiv. Preprint posted online on Sep 3, 2021. [doi: 10.48550/arXiv.2109.01652]
18. YData synthetic data for machine learning. YData. 2024. URL: https://ydata.ai/ [accessed 2024-06-19]
19. Power BI. Microsoft Corporation. 2024. URL: https://powerbi.microsoft.com/ [accessed 2025-12-08]
20. Dowding D, Merrill JA. The development of heuristics for evaluation of dashboard visualizations. Appl Clin Inform 2018 Jul;9(3):511-518. [doi: 10.1055/s-0038-1666842] [Medline: 29998455]
21. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view1. MIS Q 2003 Sep 1;27(3):425-478. [doi: 10.2307/30036540]
22. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep 1;13(3):319-340. [doi: 10.2307/249008]
23. Pisani M, Oguss M, Mayampurath A, et al. Substance misuse data commons dashboard. GitLab. 2024. URL: https://git.doit.wisc.edu/smph-public/dom/uw-icu-data-science-lab-public/substance-misuse-data-commons-dashboard [accessed 2025-12-10]
24. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
25. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003;3:993-1022. [doi: 10.5555/944919.944937]

XSL•FO
RenderX

26. Iyamu I, Gómez-Ramírez O, Xu AX, et al. Challenges in the development of digital public health interventions and mapped solutions: findings from a scoping review. Digit Health 2022;8:20552076221102255. [doi: 10.1177/20552076221102255] [Medline: 35656283]
27. Ansari B, Martin EG. Integrating human-centered design in public health data dashboards: lessons from the development of a data dashboard of sexually transmitted infections in New York State. J Am Med Inform Assoc 2024 Jan 18;31(2):298-305. [doi: 10.1093/jamia/ocad102] [Medline: 37330670]
28. Fareed N, Olvera RG, Wang Y, et al. Lessons learned from developing dashboards to support decision-making for community opioid response by community stakeholders: mixed methods and multisite study. JMIR Hum Factors 2024 Sep 9;11:e51525. [doi: 10.2196/51525] [Medline: 39250216]
29. Brady JE, DiMaggio CJ, Keyes KM, Doyle JJ, Richardson LD, Li G. Emergency department utilization and subsequent prescription drug overdose death. Ann Epidemiol 2015 Aug;25(8):613-619.e2. [doi: 10.1016/j.annepidem.2015.03.018] [Medline: 25935710]
30. Wong CA, Houry D, Cohen MK. Integrating public health and health care-protecting health as a team sport. N Engl J Med 2024 May 16;390(19):1739-1742. [doi: 10.1056/NEJMp2403274] [Medline: 38598818]
31. Rubel S, Roe S. Public health and safety team toolkit: guidance for data-driven overdose response coordination among public health, criminal justice, law enforcement, and first responders. : CDC Foundation, Centers for Disease Control and Prevention; 2022 URL: https://stacks.cdc.gov/view/cdc/127746/cdc_127746_DS1.pdf [accessed 2025-12-18]

ABBREVIATIONS

COREQ: Consolidated Criteria for Reporting Qualitative Research
EHR: electronic health record
EMS: emergency medical services
ICR: intercoder reliability
NASA-TLX: National Aeronautics and Space Administration-Task Load Index
OFR: overdose fatality review
REDCap: Research Electronic Data Capture
SMDC: Substance Misuse Data Commons
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by A Kushniruk; submitted 14.Jul.2025; peer-reviewed by AL Dir, K Schwartz; revised version received 07.Oct.2025; accepted 12.Nov.2025; published 15.Jan.2026.

Please cite as:
Pisani M, Oguss MK, Dickson-Gomez J, Kostelac C, Parry A, Moss S, Salisbury-Afshar E, Patterson B, Spigner M, Gussick M, Krautkramer A, Gruenloh T, Safipour Afshar A, Gupta P, Mayampurath A, Afshar M
Designing a Substance Misuse Data Dashboard for Overdose Fatality Review Teams: User-Centered Design Approach
JMIR Hum Factors 2026;13:e79407
URL: https://humanfactors.jmir.org/2026/1/e79407
doi:10.2196/79407

©Marie Pisani, Madeline K Oguss, Julia Dickson-Gomez, Constance Kostelac, Amy Parry, Starr Moss, Elizabeth Salisbury-Afshar, Brian Patterson, Michael Spigner, Megan Gussick, Alison Krautkramer, Timothy Gruenloh, Askar Safipour Afshar, Preeti Gupta, Anoop Mayampurath, Majid Afshar. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 15.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Navigation, Adoption, and Use of Digital Health Technologies for Irritable Bowel Syndrome Self-Management: Focus Group Study of Patient Experience and Decision-Making

Adrijana D'Silva[1*], PhD; Nicolle Hua[1*], MGenCouns; Mary V Modayil[2], PhD; Judy Seidel[1,2], PhD; Deborah A Marshall[1,3,4], PhD

[1]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[2]Primary Care Alberta, Edmonton, AB, Canada

[3]Department of Medicine, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[4]O'Brien Institute for Public Health, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[*]these authors contributed equally

Corresponding Author:
Deborah A Marshall, PhD
Department of Community Health Sciences
Cumming School of Medicine
University of Calgary
HRIC 3C60
3280 Hospital Drive NW
Calgary, AB, T2N 4Z6
Canada
Phone: 1 403 210 6377
Email: damarsha@ucalgary.ca

Abstract

Background: Irritable bowel syndrome (IBS) is a common chronic gastrointestinal disorder that impairs bowel functions and patients' overall quality of life. IBS-focused digital health technologies (DHTs), including online health resources and mobile health (mHealth) apps, have recently proliferated for patient use. However, research exploring patients' experiences with navigating, adopting, or using commercial or publicly available DHTs for IBS self-management is limited.

Objective: This study aims to explore the user experiences and decision-making of patients with IBS as they navigate, adopt, and use diverse DHTs for disease self-management.

Methods: We conducted virtual semistructured focus group interviews to explore the experiences of patients with IBS using DHTs, including their perspectives on design and features, their decision-making process in using DHTs, and recommendations for improving user experience and uptake, given the heterogeneous nature of these tools. Canada-based patients with IBS who were using or had used mHealth apps to manage symptoms were recruited through purposive sampling from previous IBS-related studies. Discussions were transcribed verbatim, and inductive thematic analysis was performed using NVivo (version 14; Lumivero). A modified version of the Expanded Unified Theory of Acceptance and Use of Technology (UTAUT2) model was applied to guide the interpretation of the dynamic relationship between the influences on participants' decisions regarding DHT use.

Results: Among the 8 participants (all female; mean age 55.3, SD 13.5 years), two themes were identified: (1) uncertainty impacts the trustworthiness of DHTs, and (2) influences that drive the decision-making process to adopt and use DHTs. The observed influences aligned with the constructs of the UTAUT2 model (performance expectancy, effort expectancy, social influence, facilitating conditions, hedonic motivation, price value, and habit), with the addition of trust and risk in participants' decision-making. Digital health literacy and patient engagement were also raised as crucial components of participants' experiences and perspectives on DHTs.

Conclusions: Findings of this study highlight the current landscape of digital health in IBS and existing gaps and challenges for patients in navigating, adopting, and using DHTs for IBS self-management. While DHTs were generally viewed positively for their value and potential, patients with IBS consider several coexisting factors and trade-offs in their decision-making. Further

XSL•FO
RenderX

investigations on the influences on and perspectives toward DHTs could enhance future development and iterations of these tools and improve patient confidence and uptake.

(JMIR Hum Factors 2026;13:e75012) doi:10.2196/75012

KEYWORDS

decision-making; digital health; health literacy; irritable bowel syndrome; mobile applications; patient experiences; patient participation; qualitative study; self-management; technology

Introduction

Digital health has revolutionized health care and disease management by leveraging technology and digital devices to improve self-management. With a growing public focus on health and well-being, more patients with chronic diseases rely on digital health, especially online health resources and mobile health (mHealth) apps [1,2], for answers, guidance, and management support.

Irritable bowel syndrome (IBS) is a disorder of the brain–gut connection that impacts patients' daily activities and well-being and is characterized by abdominal pain, bloating, altered bowel habits, and psychiatric comorbidities [3-5]. As there is no known cure for IBS, greater knowledge of the syndrome and relevant self-management options are critical in alleviating symptoms and improving quality of life [6,7]. However, IBS self-management strategies—such as lifestyle and dietary modifications and brain–gut behavior therapies [8,9]—are often coupled with challenges around accessibility, adherence, affordability, and productivity loss for patients [10-12]. Digitizing health information and support could offer alternative models of delivery and potentially address these issues by improving access to information and treatments, reducing health-related costs, and enabling and empowering patients [13-16].

In the context of IBS, digital health technologies (DHTs) encompass a diverse range of tools that support dietary, psychological, behavioral, and educational aspects of self-management. Commonly used dietary-focused tools include low fermentable oligosaccharides, disaccharides, monosaccharides, and polyols (FODMAP) diet apps, which are supported by evidence for improving symptom control and dietary adherence [17]. Psychological and brain–gut behavior apps have also demonstrated effectiveness in reducing IBS symptom severity and improving quality of life [18,19]. In addition, patients frequently use general symptom trackers, medication apps, and online educational resources of varying quality. This heterogeneity underscores the importance of understanding not only whether patients use DHTs, but also the specific domains these tools address, as each represents a difference in therapeutic mechanisms and levels of evidence within IBS care.

Our understanding of DHTs for IBS remains limited. Although several studies support the benefits of DHTs for patients with IBS [20], there is limited evidence on their experiences in a consumer context. Existing studies that qualitatively evaluate DHTs, although informative, have been limited to specific DHTs under controlled conditions [21-23] and otherwise lacked the sociodemographic context that could otherwise impact uptake and use. Other studies suggest acceptance of DHTs is determined by users' hedonic motivations, user-system fit, and perceived utility and intuitiveness of the technology [24-26]. However, it is uncertain how these influences or other priorities influence this specific patient cohort. Currently, it is difficult to assess other considerations and consequences for the real-world application of DHTs for the people with IBS. This study is an initial investigation to explore the experiences of patients with IBS in navigating, adopting, and using online health resources and mHealth apps to support self-management, and to identify key influences in their decisions to consider DHTs.

Methods

Ethical Considerations

This study was reviewed and approved by the University of Calgary Conjoint Health Research Ethics Board (REB23-1273). A qualitative study design with semistructured focus groups was used to explore the experiences and decision-making processes of patients with IBS regarding DHTs, including their adoption, use, and suggestions for improving these tools. The study methodology was reported in accordance with the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist (Multimedia Appendix 1) [27].

Positionality

The study team comprised 5 women with research experiences in health systems (AD, NH, MVM, JS, and DAM), including one (AD) with lived experiences of IBS. AD and NH, a postdoctoral fellow and research associate, respectively, were skilled in qualitative methods. AD and DAM, a health economist and health services researcher, also had experience in IBS-related research. MVM and JS, a senior scientist and scientific director, respectively, had applied research expertise and experience in primary care. All but one each have a PhD; NH has a master's degree. Information on AD's personal experience and professional background in IBS was shared with study participants prior to the focus group sessions. Some participants had an existing relationship with AD from previous research studies.

Participants

Participants were recruited through purposive sampling from previous IBS-related studies conducted by the study team. Past participants who had provided consent to be recontacted for future studies were invited by email. Convenience sampling was attempted using the study team's social media networks and the University of Calgary's research recruitment board;

XSL•FO
RenderX

while 2 individuals were recruited through this method, neither participated in the focus groups. All participants were adults with IBS living in Alberta and British Columbia, Canada, who (1) had been formally diagnosed by a health care professional and (2) had been using or had used mHealth apps to manage their symptoms. Participants who satisfied the selection criteria received an information document regarding the study and provided written consent before the focus group sessions.

The study team a priori aimed for at least 4 focus groups, with 3 participants per group, which was expected to be sufficient to reach code saturation [28-30]. While focus group size typically ranges from 6 to 8 participants, smaller groups with 3 participants were considered to allow more opportunities for engagement and account for the potentially low recruitment yield [28]. The recruitment process ended following the exhaustion of all recruitment strategies, resulting in 8 participants and 3 focus groups. Some degree of saturation was present in participants' responses through repeated codes and themes.

Data Collection

Participants completed a survey before attending the focus group session to provide details of their existing supports, demographic information, year of IBS diagnosis, symptoms, treatment history, and dietary adjustments (Multimedia Appendix 2). The survey and data were managed using Qualtrics Experience Management (Qualtrics LLC) and the University of Calgary's secured online cloud server. Survey data were analyzed and summarized using Microsoft Excel by Darren Zhang.

Focus groups were conducted virtually using Microsoft Teams. AD moderated the focus group discussions, with the support of NH as a scribe. Sessions lasted approximately 2 hours, were audio-recorded using the recording feature in Microsoft Teams, and were transcribed verbatim using the transcription service Rev (Rev.com, Inc); the transcripts were subsequently reviewed and cleaned by Darren Zhang.

Instrumentation

Participants' IBS symptoms were assessed in the survey using the IBS Symptom Severity Score (IBS-SSS); scores ranged from 0 to 500 and were categorized as mild (75-175), moderate (176-300), or severe (>300) [31]. A focus group guide was created and reviewed by the study team to support AD in facilitating semistructured sessions (Multimedia Appendix 3). The following topics were discussed in the focus groups: (1) participants' experiences navigating digital resources to self-manage IBS, (2) their experiences adopting and using DHTs to self-manage IBS, and (3) their perspectives on influences and improvements regarding the uptake of online resources and mHealth apps.

Data Analysis

Coding was performed and managed using NVivo (version 14; Lumivero). NH and AD coded the transcripts as primary and secondary coders, respectively. Because little research has explored this topic, inductive thematic analysis was used to identify and evaluate relationships, patterns, or themes, allowing data to serve as the foundation of the results [32]. Initial codes were generated to indicate recurring features within participants' responses, and related codes were then collated into groups that subsequently defined underlying themes [32]. Following thematic analysis, the findings and their relationships were visualized based on the Expanded Unified Theory of Acceptance and Use of Technology (UTAUT2) framework model [26].

Applied Framework

The UTAUT2 framework model is an extended version of the original UTAUT model to evaluate the acceptance of new commercial technologies [26,33]. With its consumer-use context, this model was chosen to provide valuable and holistic insight into patients' acceptance of DHTs, including adoption and use, for IBS self-management. The key constructs of the UTAUT2 model [26] included performance expectancy, defined as the extent to which the individual expected the performance or benefits of the technology; effort expectancy, defined as the level of effort or ease in using the technology by the individual; social influence, defined as the extent to which the individual perceived the value of others' acceptance of the technology; facilitating conditions, defined as the extent to which the individual believed that support or resources were available to facilitate the use of the technology; hedonistic motivation, defined as the extent of the individual's motivations or pleasures experienced in using the technology; price value, defined as the cost-benefit or trade-off between the perceived value or benefit of the technology and its associated monetary cost; and habit, defined as the extent to which the individual developed a habitual or perpetual use of the technology from previous experience and learning.

Also, we modified the UTAUT2 model to include "trust" and "risk" as key constructs based on our thematic findings. Although they were not included in the original UTAUT2 model [26], previous studies have recognized and incorporated them into their own models [34-36]. Based on these studies [34-36], "trust" was defined as the extent to which an individual perceived the technology or its associated qualities as likable or trustworthy, and their willingness to believe that the technology would meet their expectations, and "risk" was defined as the extent to which an individual perceived the technology to be associated with potential consequences or harm to themselves.

Additionally, DHT-related key constructs are influenced by user characteristics. In the expanded UTAUT2 model, Venkatesh et al [26] recognized how a user's demographic characteristics, specifically age, gender, and experiences or familiarity of existing technologies, moderate the impacts of the key constructs and ultimately the acceptance and use of the technology.

Results

Participant Characteristics

Of the 11 participants recruited to the study, 8 attended the focus groups, all of whom had participated in past research studies. One individual was lost to follow-up prior to the focus group sessions, and the remaining 2 could not attend the sessions due to scheduling conflicts. Participant characteristics are summarized in Table 1.

XSL•FO
RenderX

Table 1. Participants' characteristics (N=8).

Characteristics	Value
Age (years), mean (SD)	55.3 (13.5)
Gender (female), n (%)	8 (100)
Highest education, n (%)	
Postsecondary certificate or diploma below bachelor level or apprenticeship or trades certificate	1 (13)
Bachelor's degree	1 (13)
Postsecondary certificate or diploma above bachelor level	4 (50)
Master's degree	2 (25)
Identify as, n (%)	
Born outside of Canada	3 (38)
Persons with disabilities	2 (25)
2SLGBTQI+[a]	1 (13)
Racialized or visible minority	1 (13)
Indigenous	0 (0)
I have children or grandchildren aged 18 years or younger living at home	0 (0)
New to Canada (less than 5 years)	0 (0)
Other	1 (13)
None of the above	3 (38)
Years since IBS[b] diagnosis, mean (SD)	12.9 (10.6)
Types of IBS, n (%)	
Mixed	2 (25)
Constipation-dominant	5 (63)
Diarrhea-dominant	1 (13)
Unsure	0 (0)
IBS-SSS[c], mean (SD)	197.7 (84.8)
Types of digital tool used, n (%)	
Mobile health apps	7 (88)
IBS-related podcasts	4 (50)
Mobile tracking apps (eg, physical activity and daily living activities)	6 (75)
Websites	6 (75)
Support groups	1 (13)
Types of mobile health apps (n=7), n (%)	
Diet-related apps	2 (29)
IBS-specific apps	6 (86)
Mental health or emotional well-being apps	1 (14)
General gastrointestinal and health apps	2 (29)
Cognitive-behavioral therapy and/or gut hypnotherapy apps	1 (14)

[a]2SLGBTQI+: two-spirited, lesbian, gay, bisexual, transgender, queer/questioning, intersex, and others.

[b]IBS: irritable bowel syndrome.

[c]IBS-SSS: IBS Symptom Severity Score; ranges from 0 to 500, with mild (75-174), moderate (175-299), and severe (300-500) symptom severity.

All participants identified as women and had tertiary-level education. The mean age was 55.3 (SD 13.5) years, and participants had lived with an IBS diagnosis for an average of 12.9 (SD 10.6) years. Five of 8 (63%) participants reported constipation-dominant IBS, 2 (25%) reported mixed IBS, and 1 (13%) reported diarrhea-dominant IBS. The average IBS-SSS

XSL•FO
RenderX

was 197.9 (SD 84.8), indicating mild symptoms. Participants had managed their IBS symptoms using one or a combination of the following DHTs: a total of 7 (88%) had used mHealth apps, 6 (75%) used mobile tracking apps, 6 (75%) used websites, 4 (50%) used IBS-related podcasts, and 1 (13%) used support groups. Among all types of mHealth apps used (n=7), 6 participants (86%) reported having used IBS-specific apps, 2 (29%) diet-specific apps, 2 (29%) general gastrointestinal and health apps, 1 (14%) mental health and/or emotional well-being apps, and 1 (14%) cognitive-behavioral therapy and/or gut hypnotherapy apps.

Themes of Navigating, Adopting, and Using Digital Health Technologies

Two themes were identified from participants' experiences and decision-making around navigating, adopting, and using DHTs for IBS self-management: (1) uncertainty impacts the trustworthiness of DHTs, and (2) underlying influences drive the decision-making process to adopt and use DHTs.

Theme 1: Uncertainty Impacts the Trustworthiness of Digital Health Technologies

A sense of uncertainty was prevalent among participants when navigating, adopting, and using DHTs for their symptoms of IBS. While participants were generally positive toward DHTs with hope and expectations of effective self-management and symptomatic relief, the coexisting challenges and risks contributed to varying levels of distrust.

Participants expressed uncertainty toward the effectiveness, utility, and understanding of DHTs. Because of the self-directed nature of the navigation process through their mobile devices and the internet, participants were aware that not all DHTs would be effective or suitable for them, and some recounted their experiences of finding suitable DHTs through trial-and-error.

> *[. . .] As we know, it is hard to know if anything's working. It's hard to know if anything works for the symptoms that we're experiencing. So that does come into it as well, but it's not a science, it's sort of a bit of a feeling about whether you're going to continue with something and whether it's worth it.* [Participant #48]

> *I just found and tried and listened to whatever. And then if it wasn't in my wheelhouse then I just discarded.* [Participant #29]

In contrast to those with positive experiences, other participants with subpar experiences using DHTs expressed challenges and frustration in finding what they needed to manage IBS. The lack of guidance around appropriate and trustworthy DHTs left some feeling alone and overwhelmed by the amount of information online or the number of mHealth apps to consider.

> *We just kind of have to search and go online and try to look at reliable resources [. . .] I've got family help too that they're in the tech side of things, so I get some help from them from looking at different options. But basically, you do it on your own.* [Participant #37]

> *You're feeling crummy. You need a resource, you need some help, but trying to wade through everything or do a Google (search) is just so overwhelming.* [Participant #9]

Uncertainty toward the DHTs' intentions was raised, including fear of scams or malicious attacks from illegitimate websites and privacy and data exploitation from mHealth apps. Furthermore, participants discussed the accuracy of online information from the internet and their concerns about consuming IBS-related misinformation and disinformation.

> *I agree with (Participant #9) regarding credentials and if (doctors are) trying to sell stuff. I know there's some doctors out there who do that too. So yeah, just mindful of people trying to market products versus people with credentials who are trying to debunk misinformation or provide just evidence backed information.* [Participant #14]

> *I don't like even giving my credit card information or any other personal information out into the internet. You just never know who's out there and what's going to happen to it.* [Participant #37]

During these discussions, digital health literacy was raised as an essential skill and a common barrier for patients with IBS to assess, process, and understand digital health information and technologies.

> *It's not about (like) or dislike, it's always questioning what I found, is it true or not, and how to explore further to be sure that these things are true because I'm not a great researcher [. . .] because of lack of medical background. So I am always hesitant to trust this or that information. It's not disliking, but it's a concern whether it's right direction to move or not.* [Participant #4]

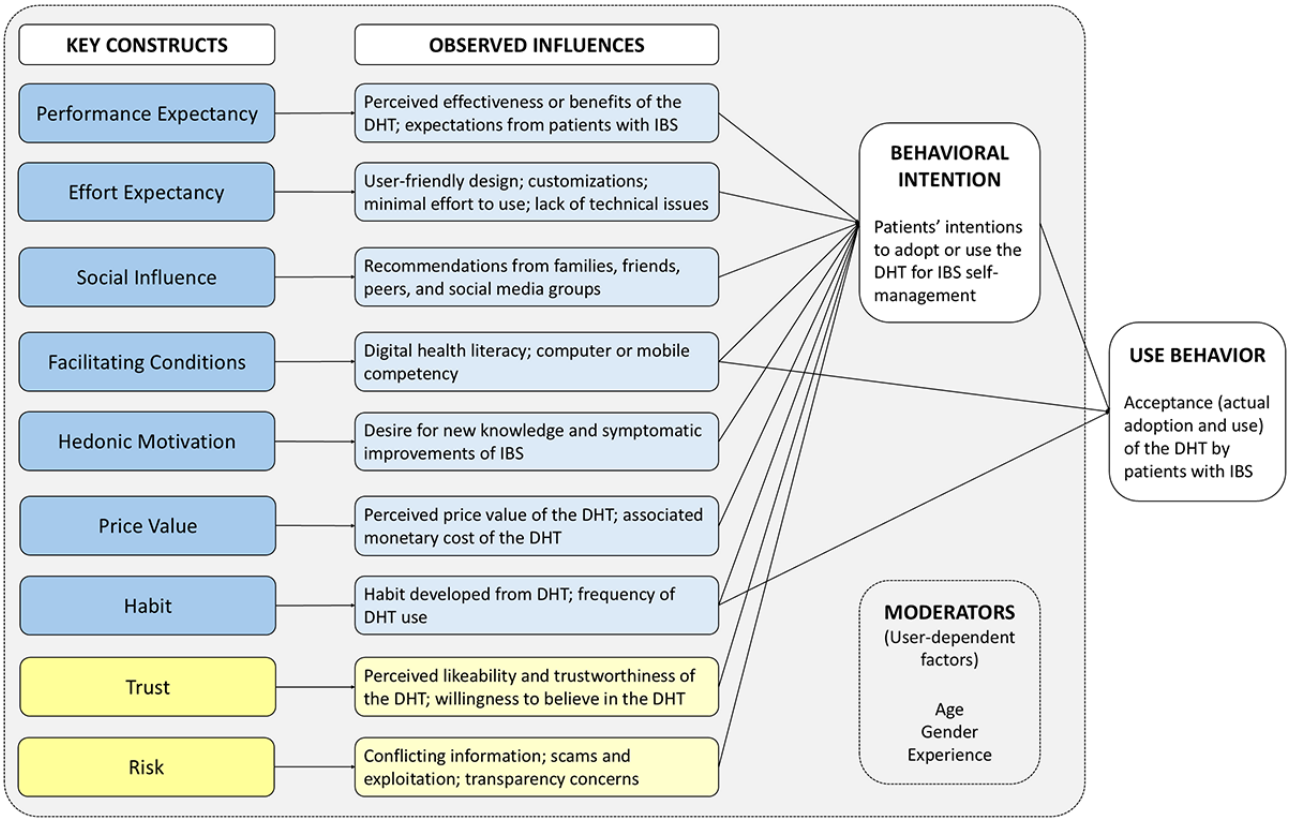
Theme 2: Influences That Drive the Decision-Making Process to Adopt and Use Digital Health Technologies

Overview

Participants alluded to several influences that contributed to their decision to use DHTs, which were contextualized using the modified UTAUT2 model (Figure 1). Applying the model to this cohort illustrates the model's relevance and the complexities of DHT acceptance among patients with IBS. The observed influences (key constructs), as proponents or opponents to their decision-making process at varying degrees, collectively impact participants' intentions and/or the actual adoption or use of DHTs. Furthermore, the influences may have been externally moderated by participants' demographics, including age, gender, and experience or familiarity with DHTs. Selected quotes are provided to illustrate the influences and their relation to key constructs from the UTAUT2 model (Multimedia Appendix 4).

XSL•FO
RenderX

Figure 1. Relationships of identified influences in the decision-making process of patients with irritable bowel syndrome (IBS) on the use and acceptance of digital health technologies (DHTs) based on the modified Expanded Unified Theory of the Use and Acceptance of Technology (UTAUT2) model. "Trust" and "risk" were not part of the original UTAUT2 model and were included as a result of the study's findings.



Performance Expectancy

Based on their prior knowledge, participants expressed varying levels of preconceived expectations toward the DHTs' performance. Aspects of participants' knowledge prior to using the DHTs included past experiences and associated emotions, user-system fit, health and disease knowledge, social influences and recommendations, and preexisting familiarity or trust toward the tool. Most participants therefore evaluated the performance of mHealth apps based on one or multiple outcomes: (1) the perceived severity and/or frequency of IBS symptoms before and after use, (2) changes to overall mood before and after use, (3) feelings toward the DHT, and (4) the ability of addressing concerns specific to the user. Online health resources, many of which served to disseminate disease knowledge and information, were evaluated differently by participants compared with apps. Participants instead focused on the accuracy, readability, accessibility, and perceived value of the information when assessing performance. One participant also noted the expectation-reality discrepancy among peers, contrasting high and unrealistic expectations of a fast, curative "fix" with DHTs designed for regular, long-term disease management.

Effort Expectancy

Most participants emphasized the importance of user-friendly DHT designs in successfully maintaining user engagement. Some described their experiences with poorly designed mHealth apps as frustrating or overwhelming (eg, rigid features, complicated functions requiring more user effort), and were subsequently discouraged from using the app further or to its fullest potential. Poor user design included features such as health monitoring functions requiring frequent data inputs or stringent timing, lacking customizable functions for personalization, and features with technical difficulties or lacking updates to address bugs. Conversely, apps with a more simplistic and seamless design reducing user effort, along with more customization capabilities, were viewed favorably. One participant considered it a priority to improve patient-centered aspects in DHT designs.

Participants unanimously agreed that patient engagement is critical for DHTs to ensure they are relevant and well-designed to address the needs of patients with IBS. Some participants considered that patients could engage in varying stages of development, such as the design and testing phases. The unanticipated burden of incorporating DHTs into their lifestyles—despite their intended use to improve patients' quality of life—further underscores the need for patient partners in the health research and technology space.

Social Influence

Many participants acknowledged their social circles as valuable support for their IBS diagnosis and management. They specified the type of support they received from their family and friends, specifically in navigating online resources, accessing the appropriate health care professionals, and recommending mHealth apps. However, some felt that certain DHT recommendations from their social circles were unhelpful or irrelevant to their circumstances because of limited knowledge or understanding of IBS. Others also were hesitant to fully embrace their support altogether, citing the social stigma around the condition.

XSL•FO
RenderX

In contrast, peer support groups were viewed favorably as a knowledgeable resource because of shared lived experiences with IBS. Most participants sought IBS support—often through social media platforms—for information regarding IBS-specific professional support, symptomatic management strategies, tools, and resources. One participant explained that peer recommendations could also help patients with IBS in decision-making about using mHealth apps by considering aspects or barriers from another perspective, such as evaluating the price relative to its value or benefits.

Facilitating Conditions

Participants' level of digital health literacy and competency with a device (eg, computer or mobile device) were identified as facilitating conditions contributing to user experiences with DHTs. Participants recognized the importance of digital health literacy in navigating, processing, and critiquing online health information. Many participants were considered to have higher digital health literacy because of their educational or professional backgrounds. Participants also acknowledged the role of computer and mobile competencies required for accepting and using DHTs, along with the significant learning curve for digital competency among older populations.

Hedonic Motivations

Most participants mentioned that their decision to adopt and use DHTs stemmed from a desire and willingness to learn about IBS and emphasized the importance of new and updated information available on these tools. Some participants had thoroughly researched and exhausted common IBS-related recommendations; DHTs that regularly provided new content were highly desirable and therefore maintained user engagement. Conversely, most participants struggled to fully engage with DHTs that lacked new content or variety, as they no longer felt the need to continue using these DHTs. Some participants also noted their interest in other IBS topics areas, such as mental health and community engagement.

Price Value

Participants frequently evaluated the cost-benefit trade-offs of DHTs—the monetary cost relative to their perceived values and benefits—in their decision-making process. Most participants first opted for services or apps that were free to download and use, but some also considered DHTs with paid business models. However, DHTs following a subscription business model were perceived by participants as the least desirable, with the accrued cost to maintain full access outweighing the potential benefits.

Habit

Participants correlated their habit or frequency of using DHTs with the DHTs' perceived utility, with greater frequency and established positive habits to be associated with symptomatic improvements and user satisfaction. Conversely, DHTs that were seldom used were deemed less useful, thus discouraging participants from using them even further.

The habit of DHT use depended on the features available, specifically reminders, and the digital platform. Some participants noted that customizable reminder functions prompted more consistent engagement from their experiences and were considered especially useful for certain DHT characteristics (eg, symptom and diet tracking and exercise programs); however, one participant mentioned that they did not need to be reminded to use their DHTs. Participants agreed that, given their frequent use of mobile devices (eg, smartphones), mHealth apps are more advantageous in maintaining user engagement. In contrast, participants perceived online websites or programs differently; they were viewed as cumbersome to locate, input login credentials, and search for the content. The integration of lifestyle with digital devices served as an antecedent factor that influenced participants' level of DHT acceptance and engagement.

Trust

In the context of the UTAUT2 model, trust in a DHT influences users' intention or willingness to use it and subsequently, whether the DHT is used for IBS (Figure 1). When navigating online health resources, participants positively associated medical or academic credibility with trust. Participants often referred to reputable medical websites (eg, Mayo Clinic and WebMD) for health information to improve their knowledge and IBS self-management. Academic and research institutions were also viewed favorably, and the involvement of a credible institution in the development of DHTs heavily weighted in favor of participants' decision-making. Some participants also acknowledged digital health literacy as a contributing factor toward trustworthiness of DHTs, as the ability to critically assess the information and sources provided greater confidence in recognizing credible websites and apps and placing trust in them.

Risk

In the acceptance and use of DHTs, participants weighed the potential risk associated with using DHTs as part of the trade-off with perceived benefits. Conflicting information, transparency concerns, and fear of financial and data exploitation were commonly raised by participants when navigating or using DHTs. For mHealth apps, concerns were primarily related to the handling and security of personal and health information. Regarding online resources, participants recognized the prevalence of conflicting information on the internet and the potential risks posed to unsuspecting users seeking clinical guidance. Participants also recounted their experiences with suspected scams, specifically websites that offered paid products and services to allegedly treat or cure IBS. One participant correlated their familiarity with the DHT and the DHTs' risk, associating less familiar DHTs with lower trustworthiness and greater risk.

Discussion

Principal Findings

Overview

Our findings highlight the importance of certain characteristics from both DHTs and users with IBS in determining intentions and success in navigating, adopting, and using DHTs, with many attributes aligning with the UTAUT2 framework [26]. Participants expressed an overarching theme of uncertainty towards DHTs in their abilities to ameliorate symptoms, which is based on their levels of and balance between perceived trust

XSL•FO
RenderX

and risk. Other relevant user-specific attributes include digital health literacy and digital competency (facilitating conditions), desire for new knowledge (hedonic motivation), habitual use of DHTs (habit), and recommendations from social circles (social influence). DHT-specific characteristics include considerations around the design and features that influence perceived effectiveness (performance expectancy), user experience (effort expectancy), and value (price value). These considerations holistically illustrate the experiences and perspectives of patients with IBS using DHTs and their underlying and multifaceted influences. Addressing barriers and leveraging facilitators is essential for developing effective DHTs, improving user experience and health outcomes, and promoting innovation within the digital health care space [37-39].

Building Trust and Addressing Uncertainties of Digital Health Technologies

Trust is a juxtaposing force to fear of uncertainty and a critical component of successful implementation, adoption, and use of DHTs. However, the unprecedented evolution of the digital health landscape poses several challenges, including limited regulations, guidelines, and a lack of robust evidence base for efficacy, safety, and impact [40,41]. Furthermore, other limitations (eg, technological, costs, and transparency concerns) add to the DHTs' "black box" nature, which could discourage users from accepting or engaging with DHTs when they are not being fully informed.

In a 2018 scoping review, trust in digital health systems was influenced by several elements, including quality, efficiency, self-efficacy, accessibility, reputation and users' recommendations, and fear of data exploitation [42]. Many of these enabling and inhibiting elements of trust align with the key constructs of the applied UTATU2 model, further highlighting the complexity and interconnectedness of patient trust in adopting and using DHTs. Participants' trust in DHTs was also compromised by their fear and the potential risk of harm through scams, malicious digital attacks, and exploitation. Similarly, Catapan et al [43] found commercial and data exploitation concerns to undermine the trustworthiness of digital health in a cohort of patients with chronic kidney disease.

Participants' hesitancy toward DHTs must also be understood within the broader digital misinformation landscape. Research has shown that IBS-related content on online platforms, such as YouTube and TikTok, is made accessible in the absence of a vetting process, and can be inaccurate, promotional, or lack scientific grounding [44,45]. Likewise, online dietary guidance for IBS often suffers from poor readability and inconsistent quality [46]. Patients with chronic, fluctuating symptoms such as IBS might be particularly vulnerable to online "cures," unregulated wellness products, or misleading health claims [47]. In this context, skepticism reflects an adaptive response to navigating oversaturated but often unreliable digital health information. Recognizing health misinformation as an external driver of uncertainty reinforces the need for credible and evidence-based DHTs that patients can trust.

Reducing the uncertainties of DHTs and improving patient trust require a large-scale, coordinated, multidisciplinary approach involving technology developers, patients, researchers, health care professionals and institutions, and government bodies. First, improving the technological systems and processes of DHTs for quality, data security, and transparency is needed, with patients and researchers involved to assess effectiveness, utility, value, and feasibility [38]. Second, improving health care system infrastructures to better respond and adapt to DHTs will further facilitate general trust in DHTs among patients and health care professionals [38]. Lastly, stronger regulatory practices and guidelines should be implemented to improve transparency by ensuring that commercial and publicly available DHTs meet efficacy, safety, and privacy standards. While governments have the capacity and means to implement laws and regulatory measures, professional health associations and bodies could also play a pivotal role in supporting health care professionals in navigating and using DHTs alongside their patients. For example, Torous and Roberts [48] argued for the American Psychiatric Association to provide digital health–related resources and guidelines to patients and clinicians. However, while advancing transparency and patient trust is crucial, it is also important to avoid overregulation, which could subsequentially limit or delay DHT innovation, adoption, and use [48].

Existing Gaps in Digital Health Literacy

Digital health literacy is a factor in participants' decision-making on adopting or using DHTs (facilitating condition) and a crucial determinant of health—especially in one's ability to navigate and critically assess health information online and leverage the benefits of effective DHTs [49]. Our sample was highly educated, which may have influenced their ability to evaluate the quality of online sources effectively without additional support. However, digital inequality is an ongoing issue; patients with lower levels of digital health literacy are more likely to struggle with navigating and using DHTs, considering that high levels of reading comprehension are expected [39,46,50]. Limited understanding of content, benefits, and risks in adopting health technologies hinders patients' ability to make informed decisions about DHTs, and by extensions, to fully trust them.

Improving digital health literacy requires a collaborative approach involving patients, health care professionals, and policymakers. Strategies include the availability of educational or digital health literacy programs [51] and adapting online information and resources to a broader audience [52]. Technology developers should also tailor their content and design based on the target users' characteristics, such as digital health literacy levels, and improve transparency around privacy and security to mitigate harm and build trust. By facilitating the conditions in which patients are equipped with the skills and confidence to effectively use DHTs, they could fully leverage these available resources.

The Financial Barrier to Digital Health Technologies

Unsurprisingly, affordability was raised as a common barrier for participants to access support, especially mHealth apps with paid models or in-app purchases. While paid apps may be perceived more positively than free apps [37], affordability, or the lack thereof, could impact patients' adherence to them [53]. In addition to the existing socioeconomic burdens of IBS [54],

XSL•FO
RenderX

patients could be especially more conscious of the financial trade-off of adopting and using cost-associated DHTs. Participants' preferences for one-time payment models over subscription models aligns with similar findings from a patient cohort with rheumatoid arthritis [55], suggesting that certain payment models for DHTs could also influence patients' decision-making process. Although Xie et al [56] suggested technology developers consider subscription plans to improve affordability and users' willingness to pay, this approach might instead deter patients with chronic diseases requiring sustained DHT use. More research is warranted to better understand the preference of patients with IBS for DHTs with certain business models to ensure accessibility and acceptance.

The Value of Patient Engagement in Digital Health Technology Development

Content and user-centered designs were important aspects of DHT acceptance and use [57]. As part of their hedonic motivations, participants valued new and varied content in DHTs and considered limited and outdated content as a major deterrent. Health information–seeking behaviors are prevalent in people with IBS [58], which could be attributed to a desire and hope for relief or a cure [47]. Furthermore, IBS is a complex disorder involving individualized brain–gut interactions and requires personalized medicine [9]. As a result, the accuracy, variety, and customizability of the content are considered invaluable for patients to explore self-management strategies from a multidisciplinary perspective [59].

User-centered design is an iterative and collaborative process that requires a foundational understanding of the target users and the context of their intended use [60]. The outcome of such design is heavily dependent on comprehensive research and observations, feedback, and recommendations from lay individuals and experts at all stages of development (eg, initial design, evaluation, and user feedback) [60,61]. Patient engagement in digital health research and development is, therefore, fundamental to constructing a successful user-centered design in DHTs for optimal efficacy and impact [61].

However, patient engagement remains limited in digital health and technology innovation [62], given existing challenges around digital health literacy, accessibility, and trust and transparency concerns [39]. Careful planning and considerations are also needed to leverage patients' inputs effectively; patients are most often involved in usability testing, where DHTs are mostly finalized which consequently leaves little flexibility to implement feedback [62]. Although there has been encouraging efforts to further engage patients with IBS in digital health research and implementation science [63], more is needed to fully ascertain the implications of DHTs and develop robust evidence specific to this population group.

Strengths and Limitations

This study is the first to explore the experiences of patients with IBS in navigating the digital landscape for self-management using the UTAUT2 model. Our findings contribute to the limited existing literature by providing insight into how DHTs can support IBS symptom management while addressing concerns related to distrust and digital literacy. We also identified key barriers and facilitators that impact patient experience and outcomes and provide recommendations for improvement based on patient perspectives.

This study also has several limitations. The transferability of our findings may be limited due to the small sample size and the presence of self-selection and sampling bias. First, our recruitment method used purposive sampling, where individuals decided whether to participate in the study. Our sample was highly educated, demonstrated high levels of digital health literacy, and showed great interest and motivation in health research and therefore was not representative of the general population with IBS. Second, all participants identified as women, which may be due to a combination of potential gender-related differences in IBS prevalence [64], mHealth app use [65], or the gender disproportion of the recruitment pool from previous IBS studies. This, however, poses the challenge of evaluating gender-based differences in this study, especially when gender is a moderating variable in both the original and expanded UTAUT models [26,33]. Third, all but one participant lived in a single Canadian province; their access or experience with health services and other non-DHT support may be different from other regions of Canada or worldwide. Lastly, coding saturation could not be fully attained in this study due to the insufficient sample size. While we have identified the most prevalent themes, having additional focus groups would further capture the themes comprehensively [30]. This would also avoid data bias and ensure the content validity, replicability, and transferability of the study [66]. Nevertheless, this study achieved its aim of building a deeper understanding and appreciation of patient experiences, perspectives, and decision-making around DHTs used for IBS that could serve as an antecedent for future studies and a guide in developing more patient-centered resources and supports.

Future Directions

Given the central role of digital literacy in our findings, future research should prioritize developing and evaluating educational or digital literacy programs specifically designed for patients with IBS. Identifying the unique barriers these patients face can guide the development of effective educational interventions. Additionally, longitudinal studies examining how enhanced digital literacy impacts the adoption and sustained use of digital tools would be valuable.

An emerging aspect in digital health is the rapidly expanding role of artificial intelligence and large language models (LLMs), such as ChatGPT, in supporting patient decision-making. Early evaluations suggest advanced LLMs can provide clinically aligned guidance, summarize complex medical concepts, and support initial health information–seeking more effectively than traditional search engines [67-69]. For patients with IBS, LLMs could serve as navigational tools that help filter misinformation, translate evidence-based information into patient-friendly language, and direct users toward credible DHTs or reputable resources. At the same time, concerns regarding bias, hallucinations, privacy, dependence on the quality and reliability of available training data and prompts, and lack of transparent regulatory oversight highlight that these tools remain imperfect and require careful integration into IBS self-management

XSL•FO
RenderX

[70-72]. Because LLM outputs are only as trustworthy as the data and sources on which they are trained, they may inadvertently reproduce inaccuracies or amplify misinformation. As AI-driven tools become more embedded in digital health, understanding how patients with IBS engage with them and how these tools can be responsibly implemented will become a critical future next step in this research. As AI-enabled tools and LLMs become increasingly integrated into consumer-health information-seeking behaviors, future work should also examine how these systems can support digital health literacy and help patients with IBS navigate an often-confusing online environment. Understanding how patients with IBS engage with these emerging technologies will be essential for designing safe and supportive digital self-management tools.

The evidence base on the feasibility and implementation of digital health technology for IBS is still lacking but necessary for real-world application. Understanding certain DHT features and attributes in patients' decision-making process could serve as a guide for future research and projects to understand and prioritize key objectives. Comprehensive economic evaluations should consider the direct and indirect costs of adopting DHTs in patients' self-management practices, as cost was considered a major barrier among participants.

Another fundamental step in this field is to improve patient engagement in health research and technology development. User-centered design was highlighted as a crucial component for the success of DHTs, and the need for greater involvement of patients in developing DHTs. Patient partners in future research and commercial development of DHTs, especially in the design and testing phases, will help ensure these tools are intuitive and meet patients' needs. Patient engagement will also be valuable in providing opportunities to quantify the significance of the identified influences in this study; this information could offer further insight into patients' decision-making process and how they weigh their considerations holistically.

Conclusion

This study highlights the complexities of the digital landscape in the context of IBS. While there is optimism toward DHTs in enhancing self-management, concerns about digital health literacy, trustworthiness, access, and cost must be addressed. With user-friendly, accessible, trusted, and cost-effective tools, the effectiveness and adoption of digital health solutions can be enhanced, and the digital health user experience optimized. It is evident that integrating DHTs in IBS self-management is a multifaceted, complex task that requires comprehensive and interdisciplinary collaboration among technology developers, health services, health professionals, and patients. More efforts are needed to address the identified barriers and leverage facilitators to improve health outcomes and quality of life for patients managing IBS or other chronic diseases.

Acknowledgments

We would like to express our gratitude to our patient research partner, Chantel Wicks, for her support and contributions to participant recruitment, and to student volunteer Darren Zhang for analyzing and summarizing the prefocus group surveys for this study.

Funding

This research was supported in full by the IMAGINE (Inflammation, Microbiome, and Alimentation: Gastro-Intestinal and Neuropsychiatric Effects), a Canadian Institutes for Health Research (CIHR) Strategy for Patient-Oriented Research (SPOR) Chronic Disease Network (funding reference number: 1715-000-001), with funding from several partners. No other financial arrangements were related to the research or assistance with manuscript preparation. DAM is supported by the Svare Chair in Health Economics, Value and Impact.

Data Availability

The datasets generated or analyzed during this study are not publicly available, as participants did not provide consent to share their individual data.

Authors' Contributions

AD and NH contributed equally as co–first authors. AD, DAM, MVM, JS, and NH conceptualized and designed the study. AD and NH collected the data. AD and NH analyzed the data. AD and NH wrote the first draft of the manuscript. All other authors provided revisions to subsequent drafts. All authors contributed to and approved the final manuscript.

Conflicts of Interest

DAM reports personal fees from the Office of Health Economics, Analytica and Novartis, and meeting expenses from ISPOR—The Professional Society for Health Economics and Outcomes Research—and Illumina. AD, MVM, JS, and NH disclose no conflicts of interest.

Multimedia Appendix 1

XSL•FO
RenderX

COREQ checklist.
[PDF File (Adobe PDF File), 481 KB - humanfactors_v13i1e75012_app1.pdf]

Multimedia Appendix 2
Pre-focus group survey of participants' characteristics, IBS severity, and digital health tools used for IBS self-management.
[DOC File , 33 KB - humanfactors_v13i1e75012_app2.doc]

Multimedia Appendix 3
Semistructured focus group interview guide.
[DOC File , 31 KB - humanfactors_v13i1e75012_app3.doc]

Multimedia Appendix 4
Participant quotes of observed influences in considering DHT adoption and use for IBS patients in relation to key constructs based on the Expanded Unified Theory of Acceptance and Use of Technology (UTAUT2) model.
[DOCX File , 30 KB - humanfactors_v13i1e75012_app4.docx]

References

1. Kamiński M, Łoniewski I, Marlicz W. "Dr. Google, I am in pain"-global internet searches associated with pain: a retrospective analysis of google trends data. Int J Environ Res Public Health 2020;17(3):954 [FREE Full text] [doi: 10.3390/ijerph17030954] [Medline: 32033087]
2. Moreno-Ligero M, Moral-Munoz JA, Salazar A, Failde I. mHealth intervention for improving pain, quality of life, and functional disability in patients with chronic pain: systematic review. JMIR Mhealth Uhealth 2023;11:e40844 [FREE Full text] [doi: 10.2196/40844] [Medline: 36729570]
3. Black CJ, Ford AC. Global burden of irritable bowel syndrome: trends, predictions and risk factors. Nat Rev Gastroenterol Hepatol 2020;17(8):473-486 [FREE Full text] [doi: 10.1038/s41575-020-0286-8] [Medline: 32296140]
4. Lacy BE, Mearin F, Chang L, Chey WD, Lembo AJ, Simren M. Bowel disorders. Gastroenterology 2016;150(6):1393-1407. [doi: 10.1053/j.gastro.2016.02.031] [Medline: 27144627]
5. Shiha MG, Aziz I. Review article: physical and psychological comorbidities associated with irritable bowel syndrome. Aliment Pharmacol Ther 2021;54 Suppl 1:S12-S23. [doi: 10.1111/apt.16589] [Medline: 34927759]
6. Cong X, Perry M, Bernier KM, Young EE, Starkweather A. Effects of self-management interventions in patients with irritable bowel syndrome: systematic review. West J Nurs Res 2018;40(11):1698-1720 [FREE Full text] [doi: 10.1177/0193945917727705] [Medline: 28854852]
7. Flik CE, van Rood YR, de Wit NJ. Systematic review: knowledge and educational needs of patients with irritable bowel syndrome. Eur J Gastroenterol Hepatol 2015;27(4):367-371. [doi: 10.1097/MEG.0000000000000252] [Medline: 25874506]
8. Moayyedi P, Andrews CN, MacQueen G, Korownyk C, Marsiglio M, Graff L. Canadian Association of Gastroenterology clinical practice guideline for the management of irritable bowel syndrome (IBS). J Can Assoc Gastroenterol 2019;2(1):6-29 [FREE Full text] [doi: 10.1093/jcag/gwy071] [Medline: 31294724]
9. Lacy BE, Pimentel M, Brenner DM, Chey WD, Keefer LA, Long MD. ACG clinical guideline: management of irritable bowel syndrome. Am J Gastroenterol 2021;116(1):17-44. [doi: 10.14309/ajg.0000000000001036] [Medline: 33315591]
10. Kinsinger SW. Cognitive-behavioral therapy for patients with irritable bowel syndrome: current insights. Psychol Res Behav Manag 2017;10:231-237 [FREE Full text] [doi: 10.2147/PRBM.S120817] [Medline: 28790872]
11. Biesiekierski JR, Manning LP, Murray HB, Vlaeyen JWS, Ljótsson B, Van Oudenhove L. Review article: exclude or expose? The paradox of conceptually opposite treatments for irritable bowel syndrome. Aliment Pharmacol Ther 2022;56(4):592-605. [doi: 10.1111/apt.17111] [Medline: 35775328]
12. Shah ED, Salwen-Deremer JK, Gibson PR, Muir JG, Eswaran S, Chey WD. Comparing costs and outcomes of treatments for irritable bowel syndrome with diarrhea: cost-benefit analysis. Clin Gastroenterol Hepatol 2022;20(1):136-144.e31. [doi: 10.1016/j.cgh.2020.09.043] [Medline: 33010413]
13. Ankersen DV, Weimers P, Burisch J. Whats "App-ening": the help of new technologies in nutrition in digestive diseases. Curr Opin Clin Nutr Metab Care 2017;20(5):426-431. [doi: 10.1097/MCO.0000000000000399] [Medline: 28768297]
14. Williams M, Barclay Y, Harper L, Marchant C, Seamark L, Hickson M. Feasibility, acceptability and cost-efficiency of using webinars to deliver first-line patient education for people with irritable bowel syndrome as part of a dietetic-led gastroenterology service in primary care. J Hum Nutr Diet 2020;33(6):758-766. [doi: 10.1111/jhn.12799] [Medline: 32816367]
15. McCrone P, Everitt H, Landau S, Little P, Bishop FL, O'Reilly G. Cost effectiveness of therapist delivered cognitive behavioural therapy and web-based self-management in irritable bowel syndrome: the ACTIB randomised trial. BMC Gastroenterol 2021;21(1):276 [FREE Full text] [doi: 10.1186/s12876-021-01848-9] [Medline: 34229619]

XSL•FO
RenderX

16. Ankersen DV, Carlsen K, Marker D, Munkholm P, Burisch J. Using eHealth strategies in delivering dietary and other therapies in patients with irritable bowel syndrome and inflammatory bowel disease. J Gastroenterol Hepatol 2017;32 Suppl 1:27-31. [doi: 10.1111/jgh.13691] [Medline: 28244677]
17. O'Keeffe M, Jansen C, Martin L, Williams M, Seamark L, Staudacher HM. Long-term impact of the low-FODMAP diet on gastrointestinal symptoms, dietary intake, patient acceptability, and healthcare utilization in irritable bowel syndrome. Neurogastroenterol Motil 2018;30(1):e13154. [doi: 10.1111/nmo.13154] [Medline: 28707437]
18. Hunt M, Miguez S, Dukas B, Onwude O, White S. Efficacy of Zemedy, a mobile digital therapeutic for the self-management of irritable bowel syndrome: crossover randomized controlled trial. JMIR Mhealth Uhealth 2021;9(5):e26152 [FREE Full text] [doi: 10.2196/26152] [Medline: 33872182]
19. Brenner DM, Ladewski AM, Kinsinger SW. Development and current state of digital therapeutics for irritable bowel syndrome. Clin Gastroenterol Hepatol 2024;22(2):222-234 [FREE Full text] [doi: 10.1016/j.cgh.2023.09.013] [Medline: 37743035]
20. D'Silva A, Hua N, Modayil MV, Seidel J, Marshall DA. Digital health interventions are effective for irritable bowel syndrome self-management: a systematic review. Dig Dis Sci 2025;70(2):644-664. [doi: 10.1007/s10620-024-08672-7] [Medline: 39402208]
21. Karkar R, Schroeder J, Epstein DA, Pina LR, Scofield J, Fogarty J. TummyTrials: a feasibility study of using self-experimentation to detect individualized food triggers. Proc SIGCHI Conf Hum Factor Comput Syst 2017;2017:6850-6863 [FREE Full text] [doi: 10.1145/3025453.3025480] [Medline: 28516175]
22. Tonkin-Crine S, Bishop FL, Ellis M, Moss-Morris R, Everitt H. Exploring patients' views of a cognitive behavioral therapy-based website for the self-management of irritable bowel syndrome symptoms. J Med Internet Res 2013;15(9):e2672 [FREE Full text] [doi: 10.2196/jmir.2672] [Medline: 24001787]
23. D'Silva A, Islam Z, Marshall DA, Vallance JK, Nasser Y, Rajagopalan V. Experiences of irritable bowel syndrome patients in a virtual yoga program: qualitative findings from a clinical trial. Dig Dis Sci 2024;69(1):169-179. [doi: 10.1007/s10620-023-08125-7] [Medline: 37914888]
24. Rai A, Chen L, Pye J, Baird A. Understanding determinants of consumer mobile health usage intentions, assimilation, and channel preferences. J Med Internet Res 2013;15(8):e2635 [FREE Full text] [doi: 10.2196/jmir.2635] [Medline: 23912839]
25. Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. JMIR Mhealth Uhealth 2019;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]
26. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Quarterly 2012;36(1):157-178. [doi: 10.2307/41410412]
27. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
28. Morgan DL. Group composition, group size, and number of groups. In: Basic and Advanced Focus Groups. Thousand Oaks: SAGE Publications, Inc; 2019:49-62.
29. Hennink MM, Kaiser BN, Weber MB. What influences saturation? Estimating sample sizes in focus group research. Qual Health Res 2019;29(10):1483-1496 [FREE Full text] [doi: 10.1177/1049732318821692] [Medline: 30628545]
30. Guest G, Namey E, McKenna K. How many focus groups are enough? Building an evidence base for nonprobability sample sizes. Field Methods 2016;29(1):3-22. [doi: 10.1177/1525822X16639015]
31. Francis CY, Morris J, Whorwell PJ. The irritable bowel severity scoring system: a simple method of monitoring irritable bowel syndrome and its progress. Aliment Pharmacol Ther 1997;11(2):395-402. [doi: 10.1046/j.1365-2036.1997.142318000.x] [Medline: 9146781]
32. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
33. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425-478. [doi: 10.2307/30036540]
34. Zhao J, Li B, Sun J, Zeng X, Zheng J. Determinants of chronic disease patients' intention to use Internet diagnosis and treatment services: based on the UTAUT2 model. Front Digit Health 2025;7:1543428 [FREE Full text] [doi: 10.3389/fdgth.2025.1543428] [Medline: 40822908]
35. Schomakers E, Lidynia C, Vervier LS, Calero Valdez A, Ziefle M. Applying an extended UTAUT2 model to explain user acceptance of lifestyle and therapy mobile health apps: survey study. JMIR Mhealth Uhealth 2022;10(1):e27095 [FREE Full text] [doi: 10.2196/27095] [Medline: 35040801]
36. Arfi WB, Nasr IB, Kondrateva G, Hikkerova L. The role of trust in intention to use the IoT in eHealth: application of the modified UTAUT in a consumer context. Technol Forecast Soc Change 2021;167:120688. [doi: 10.1016/j.techfore.2021.120688]
37. Mendiola MF, Kalnicki M, Lindenauer S. Valuable features in mobile health apps for patients and consumers: content analysis of apps and user ratings. JMIR Mhealth Uhealth 2015;3(2):e4283 [FREE Full text] [doi: 10.2196/mhealth.4283] [Medline: 25972309]

XSL•FO
RenderX

38. Mumtaz H, Riaz MH, Wajid H, Saqib M, Zeeshan MH, Khan SE, et al. Current challenges and potential solutions to the use of digital health technologies in evidence generation: a narrative review. Front Digit Health 2023;5:1203945 [FREE Full text] [doi: 10.3389/fdgth.2023.1203945] [Medline: 37840685]
39. Livieri G, Mangina E, Protopapadakis ED, Panayiotou AG. The gaps and challenges in digital health technology use as perceived by patients: a scoping review and narrative meta-synthesis. Front Digit Health 2025;7:1474956 [FREE Full text] [doi: 10.3389/fdgth.2025.1474956] [Medline: 40212901]
40. Wells C, Spry C. An Overview of Smartphone Apps. Can J Health Technol 2022;2(2). [doi: 10.51731/cjht.2022.269]
41. Sittig DF, Wright A, Coiera E, Magrabi F, Ratwani R, Bates DW. Current challenges in health information technology-related patient safety. Health Informatics J 2020;26(1):181-189 [FREE Full text] [doi: 10.1177/1460458218814893] [Medline: 30537881]
42. Adjekum A, Blasimme A, Vayena E. Elements of trust in digital health systems: scoping review. J Med Internet Res 2018;20(12):e11254 [FREE Full text] [doi: 10.2196/11254] [Medline: 30545807]
43. Catapan SDC, Taylor ML, Scuffham P, Smith AC, Kelly JT. Improving consumer trust in digital health: a mixed methods study involving people living with chronic kidney disease. Digit Health 2025;11:20552076241312440 [FREE Full text] [doi: 10.1177/20552076241312440] [Medline: 39801583]
44. Jafri FI, Ancha N, Roy A, Ahmad S, Riaz H, Hyman C. TikTok and irritable bowel syndrome (IBS): a cross-sectional study of social media content quality. J Public Health (Berl.) 2023;33(3):653-658. [doi: 10.1007/s10389-023-02062-5]
45. Başmısırlı E, Kip M, Altun H, İnanç N. Evaluating the quality and reliability of YouTube videos providing nutritional recommendations for irritable bowel syndrome. J Hum Nutr Diet 2025;38(4):e70088. [doi: 10.1111/jhn.70088] [Medline: 40611462]
46. Alfaro-Cruz L, Kaul I, Zhang Y, Shulman RJ, Chumpitazi BP. Assessment of quality and readability of internet dietary information on irritable bowel syndrome. Clin Gastroenterol Hepatol 2019;17(3):566-567 [FREE Full text] [doi: 10.1016/j.cgh.2018.05.018] [Medline: 29800724]
47. Harvey JM, Sibelli A, Chalder T, Everitt H, Moss-Morris R, Bishop FL. Desperately seeking a cure: treatment seeking and appraisal in irritable bowel syndrome. Br J Health Psychol 2018;23(3):561-579 [FREE Full text] [doi: 10.1111/bjhp.12304] [Medline: 29508539]
48. Torous J, Roberts LW. Needed innovation in digital health and smartphone applications for mental health: transparency and trust. JAMA Psychiatry 2017;74(5):437-438. [doi: 10.1001/jamapsychiatry.2017.0262] [Medline: 28384700]
49. van Kessel R, Wong BLH, Clemens T, Brand H. Digital health literacy as a super determinant of health: more than simply the sum of its parts. Internet Interv 2022;27:100500 [FREE Full text] [doi: 10.1016/j.invent.2022.100500] [Medline: 35242586]
50. Butcher CJ, Hussain W. Digital healthcare: the future. Future Healthc J 2022;9(2):113-117 [FREE Full text] [doi: 10.7861/fhj.2022-0046] [Medline: 35928188]
51. Busse TS, Nitsche J, Kernebeck S, Jux C, Weitz J, Ehlers JP, et al. Approaches to improvement of digital health literacy (eHL) in the context of person-centered care. Int J Environ Res Public Health 2022;19(14):8309 [FREE Full text] [doi: 10.3390/ijerph19148309] [Medline: 35886158]
52. Bywall KS, Norgren T, Avagnina B, Gonzalez MP, Andersson SW, IDEAHL consortium. Calling for allied efforts to strengthen digital health literacy in Sweden: perspectives of policy makers. BMC Public Health 2024;24(1):2666 [FREE Full text] [doi: 10.1186/s12889-024-20174-9] [Medline: 39350078]
53. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL. Factors influencing adherence to mHealth apps for prevention or management of noncommunicable diseases: systematic review. J Med Internet Res 2022;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]
54. Bosman MHMA, Weerts ZZRM, Snijkers JTW, Vork L, Mujagic Z, Masclee AAM. The socioeconomic impact of irritable bowel syndrome: an analysis of direct and indirect health care costs. Clin Gastroenterol Hepatol 2023;21(10):2660-2669 [FREE Full text] [doi: 10.1016/j.cgh.2023.01.017] [Medline: 36731587]
55. Chahal S, Biln N, Clarke B. Patient perspectives on a digital mobile health application for RA. Open Access Rheumatol 2021;13:63-71 [FREE Full text] [doi: 10.2147/OARRR.S296541] [Medline: 33935522]
56. Xie Z, Liu H, Or C. A discrete choice experiment to examine the factors influencing consumers' willingness to purchase health apps. Mhealth 2023;9:21. [doi: 10.21037/mhealth-22-39] [Medline: 37492118]
57. Molina-Recio G, Molina-Luque R, Jiménez-García AM, Ventura-Puertos PE, Hernández-Reyes A, Romero-Saldaña M. Proposal for the user-centered design approach for health apps based on successful experiences: integrative review. JMIR Mhealth Uhealth 2020;8(4):e14376 [FREE Full text] [doi: 10.2196/14376] [Medline: 32319965]
58. Shin A, Ballou S, Camilleri M, Xu H, Lembo A. Information- and health-care seeking behaviors in patients with irritable bowel syndrome. Clin Gastroenterol Hepatol 2020;18(12):2840-2842 [FREE Full text] [doi: 10.1016/j.cgh.2019.09.020] [Medline: 31546055]
59. Staudacher HM, Black CJ, Teasdale SB, Mikocka-Walus A, Keefer L. Irritable bowel syndrome and mental health comorbidity - approach to multidisciplinary management. Nat Rev Gastroenterol Hepatol 2023;20(9):582-596 [FREE Full text] [doi: 10.1038/s41575-023-00794-z] [Medline: 37268741]

XSL•FO
RenderX

60. Wiljén A, Chaplin JE, Crine V, Jobe W, Johnson E, Karlsson K. The development of an mHealth tool for children with long-term illness to enable person-centered communication: user-centered design approach. JMIR Pediatr Parent 2022;5(1):e30364 [FREE Full text] [doi: 10.2196/30364] [Medline: 35258466]
61. Birnbaum F, Lewis D, Rosen RK, Ranney ML. Patient engagement and the design of digital health. Acad Emerg Med 2015;22(6):754-756 [FREE Full text] [doi: 10.1111/acem.12692] [Medline: 25997375]
62. Baines R, Bradwell H, Edwards K, Stevens S, Prime S, Tredinnick-Rowe J. Meaningful patient and public involvement in digital health innovation, implementation and evaluation: a systematic review. Health Expect 2022;25(4):1232-1245 [FREE Full text] [doi: 10.1111/hex.13506] [Medline: 35526274]
63. Olson JL, Palumbo A, Desjardins C, Wicks C, Bhopa S, Cheyne K. Perspectives on the sustained engagement with digital health tools: protocol for a qualitative interview study among people living with inflammatory bowel disease or irritable bowel syndrome. BMJ Open 2024;14(11):e089220 [FREE Full text] [doi: 10.1136/bmjopen-2024-089220] [Medline: 39521469]
64. Kim YS, Kim N. Sex-gender differences in irritable bowel syndrome. J Neurogastroenterol Motil 2018;24(4):544-558 [FREE Full text] [doi: 10.5056/jnm18082] [Medline: 30347934]
65. Lim D, Norman R, Robinson S. Consumer preference to utilise a mobile health app: a stated preference experiment. PLoS One 2020;15(2):e0229546 [FREE Full text] [doi: 10.1371/journal.pone.0229546] [Medline: 32084250]
66. Fusch P, Ness L. Are we there yet? Data saturation in qualitative research. Qual Rep 2015;20(9). [doi: 10.46743/2160-3715/2015.2281]
67. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW. Large language models encode clinical knowledge. Nature 2023;620(7972):172-180 [FREE Full text] [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]
68. Bopche R, Gustad LT, Afset JE, Ehrnström B, Damås JK, Nytrø Ø. Leveraging explainable artificial intelligence for early prediction of bloodstream infections using historical electronic health records. PLOS Digit Health 2024;3(11):e0000506. [doi: 10.1371/journal.pdig.0000506] [Medline: 39541276]
69. Chen J, Ge X, Yuan C, Chen Y, Li X, Zhang X. Comparing orthodontic pre-treatment information provided by large language models. BMC Oral Health 2025;25(1):838 [FREE Full text] [doi: 10.1186/s12903-025-06246-1] [Medline: 40437500]
70. Cirillo D, Catuara-Solarz S, Morey C, Guney E, Subirats L, Mellino S. Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. NPJ Digit Med 2020;3(1):81 [FREE Full text] [doi: 10.1038/s41746-020-0288-5] [Medline: 32529043]
71. Yang S, Jing M, Wang S, Huang Z, Wang J, Kou J. Building trustworthy large language model-driven generative recommender system for healthcare decision support: A scoping review of corpus sources, customization techniques, and evaluation frameworks. Artif Intell Med 2026;171:103310. [doi: 10.1016/j.artmed.2025.103310] [Medline: 41313967]
72. Geroimenko V. Generative AI hallucinations in healthcare: a challenge for prompt engineering and creativity. In: Geroimenko V, editor. Human-Computer Creativity: Generative AI in Education, Art, and Healthcare. Switzerland: Springer Nature; 2025:321-335. [doi: 10.1007/978-3-031-86551-0_17]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
DHT: digital health technology
FODMAP: fermentable oligosaccharides, disaccharides, monosaccharides, and polyols
IBS: irritable bowel syndrome
IBS-SSS: IBS Symptom Severity Score
LLM: large language model
mHealth: mobile health
UTAUT2: Expanded Unified Theory of Acceptance and Use of Technology

Edited by A Kushniruk; submitted 26.Mar.2025; peer-reviewed by J Soldera; comments to author 16.Oct.2025; accepted 02.Jan.2026; published 02.Feb.2026.

Please cite as:
D'Silva A, Hua N, Modayil MV, Seidel J, Marshall DA
Navigation, Adoption, and Use of Digital Health Technologies for Irritable Bowel Syndrome Self-Management: Focus Group Study of Patient Experience and Decision-Making
JMIR Hum Factors 2026;13:e75012
URL: https://humanfactors.jmir.org/2026/1/e75012
doi:10.2196/75012
PMID:

XSL•FO
RenderX

©Adrijana D'Silva, Nicolle Hua, Mary V Modayil, Judy Seidel, Deborah A Marshall. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 02.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Adolescent Perceptions of an Online Safety Chatbot: Survey Study

Meriel Charles[1], PhD; James D Sauer[1], PhD; Erin Roehrer[2], PhD; Jeremy Prichard[3], PhD; Paul Watters[4], PhD; Joel Scanlan[3], PhD

[1]School of Psychological Sciences, University of Tasmania, Hobart, Australia

[2]School of Information and Communication Technology, University of Tasmania, Hobart, Australia

[3]School of Law, University of Tasmania, Hobart, Australia

[4]Cyberstronomy, Melbourne, Australia

Corresponding Author:
James D Sauer, PhD
School of Psychological Sciences
University of Tasmania
Locked bag 30
Hobart, 7001
Australia
Phone: 61 362262051
Email: jim.sauer@utas.edu.au

Abstract

Background: Adolescents face a variety of potential harms in the online environment, including exposure to distressing illegal material, cyberbullying, image-based abuse, and "sextortion." Various agencies provide on-demand helpline and information services for children and adolescents to support them with navigating online (and offline) harms.

Objective: This study examined whether a chat-based conversational agent (chatbot) might be a useful additional tool for meeting the needs of adolescents at risk from online harms. We developed a prototype chatbot—including both conversational and menu-driven user options—and evaluated users' trust in the chatbot. In this context, trust relates to perceptions of the chatbot's usability and the value of the information and support it provides.

Methods: Participants (n=224; mean age 16.8 years) interacted with the chatbots and evaluated them in terms of user trust: perceived usability and utility (ie, relevance of support resources provided).

Results: Most participants (conversational chatbot: 141/224, 63% and menu-driven chatbot: 142/224, 63%) showed a willingness to click on the chatbots' recommended support links. Participants with higher trust in the chatbots were more likely to click the links for recommended support services (with extreme evidence for large effects: δ=0.73, 95% credible interval [CrI] 0.46-1.00 and δ=0.78, 95% CrI 0.50-1.07, for the conversational and menu-driven chatbots, respectively; Bayes factor [BF_{10}]>50,000), and participants who clicked the links, compared with those who did not, reported higher rates of positive attitudes toward their decision (with extreme evidence for large effects: δ=0.87, 95% CrI 0.58-1.15 and δ=0.84, 95% CrI 0.54-1.12, for the conversational and menu-driven chatbots, respectively; BF_{10}>3,000,000). The conversational and menu-driven chatbots differed little in perceived trust or effectiveness.

Conclusions: Chatbots represent a promising additional tool to help adolescents access mental health–related support services and navigate online harms. However, establishing trust is critical.

(JMIR Hum Factors 2026;13:e71498) doi:10.2196/71498

KEYWORDS

adolescent harms; adolescent; artificial intelligence; chatbot; mental health services; online safety; trust; user-centered design

Introduction

Overview

Chat-based conversational agents—chatbots—are used in many contexts globally. Research on technology in the health care and service sector has found that the use of chatbots (compared to traditional systems, such as web interfaces or menu-based systems) can lead to a more positive user experience (ie, users feeling more comfortable and better understood) [1-4]. This is often attributed to chatbots facilitating a more conversational and interactive form of communication, making users feel more

XSL•FO
RenderX

engaged and better supported [4]. Developing these feelings of engagement and support is essential. In health care and mental health support contexts, individuals often need to share personal information with a chatbot (eg, their medical history or symptoms) to receive customized recommendations [5-8]. This process relies on users feeling secure in sharing sensitive information [2,9-12]. Effectively, the successful adoption of human-technology systems (like chatbots) relies on people's trust in these systems [3,6,13-19]. Trust in human interactions is grounded in shared emotional connections and the confidence a person has that others will prioritize their interests [19,20]. However, trust in technology and artificial intelligence (AI) systems operates differently due to the absence of genuine empathy [11,19,21]. Instead, trust in human-technology interactions is more transactional, being influenced by factors such as functionality, performance, helpfulness, predictability, and reliability [13,19]. Specifically, trust in conversational chatbots has been defined as the users' willingness to provide confidential information and to accept and act on the chatbot's recommendations [17]. In this study, we tested adolescents' feelings of trust toward, and perceptions of the utility of, 2 chatbot systems, providing proof-of-concept evidence for the potential utility of these chatbots in providing support to adolescents at risk from online harms.

Developing trust involves both cognitive and affective processes. To facilitate cognitive trust, a system must meet expectations and be perceived as useful. In contrast, affective trust is related to emotional connections, such as making a user feel secure and comfortable [22,23]. Cognitive trust, grounded in reasoning, is associated with practical aspects of an information system (ie, usage-related factors), such as functionality, ease of use, and the quality and relevance of its recommendations [16]. When users consider using new technology-based services, they can engage in a rational decision-making process, weighing potential benefits against drawbacks [22,24]. Building and reinforcing cognitive trust occurs when users experience tangible, positive outcomes (eg, when they perceive that the new system makes tasks easier). When users feel their needs are met, they are more likely to trust the system and continue using it [25-28]. Cognitive trust positively influences the acceptance and usage of technology; conversely, technology with low cognitive trust is less likely to be adopted [22,29]. Thus, developing an online safety chatbot that meets users' needs is essential for establishing and sustaining cognitive trust and for the ultimate success of the system.

Affective trust is not based solely on objective measures of performance but rather on feelings associated with security, comfort, and satisfaction [22,30]. Affective trust can help reduce the psychological distance, or perceived emotional gap, between a user and the information system [3,30-32]. Compared to basic information systems (such as a web or menu-style interface), chatbots, with their more natural interaction and personalized conversational approach, can reduce psychological distance by making users feel more comfortable and supported, encouraging users to share personal information and accept support recommendations from the chatbot [3,17,33,34]. Even adding a basic human-shaped symbol, compared to a machine-like symbol, to a chatbot has been found to enhance affective trust, reduce psychological distance between the user and the chatbot, and increase users' intentions to comply with chatbot recommendations [3].

Cognitive and affective trust often reinforce each other [22,35,36]. When a person believes a technology is useful, they become emotionally engaged with it. When they are emotionally engaged, they are more likely to view the system as useful and credible. Importantly, when users trust a chatbot, they are more likely to follow its recommendations, making trust a critical factor in predicting users' behavioral intentions, their willingness and motivation to follow advice [3,17,37-39]. A person's behavioral intention serves as an informative, though admittedly imperfect, predictor for actual behavior [40-43]. Several studies show a positive link between a stated positive intention to use an AI service (such as a chatbot) and actual usage [26,44]. Thus, investigating trust-related factors and how they relate to potential help-seeking behavior in the context of an online safety chatbot is essential for predicting how users might perceive and interact with this potentially helpful tool, as well as their willingness to accept and act on its recommendations, in a real-world setting.

The Case for an Online Safety Chatbot for Adolescents

Adolescents encounter many challenges staying safe online. The online and social media environments present a range of risks, including cyberbullying and exposure to inappropriate content, which can have serious consequences for adolescent well-being [45]. The pressures of negative online interactions can contribute to increased levels of stress, anxiety, and depression among adolescents [46,47]. The "Digital Lives of Aussie Teens" report by the e-Safety Commissioner [45] found that adolescents in Australia spend, on average, 14 hours online each week (researching topics of interest and engaging with social media, entertainment, and gaming) and, among those aged 14-17 years, just over 50% reported negative experiences in the past 6 months. Such incidents include being contacted by strangers, receiving threats and abuse, and being exposed to inappropriate or unwanted content. In addition, one-third reported negative online experiences stemming from instances of bullying that occurred at school. Although many adolescents do seek help from friends and family, a considerable portion do not share their concerns or report harmful incidents to authorities (eg, law enforcement and the school), the online platform (eg, social media platform and gaming websites), or support services (eg, helplines, mental health professionals, and counsellors). Nonetheless, a large majority (75%) of adolescents expressed a need for online safety information, with schools and reputable websites for digital well-being being their preferred sources for accessing this information. Importantly, an online safety chatbot can serve a dual purpose, not only supporting those who may have experienced harm but also reducing the risk that they may unintentionally be causing harm to others due to their limited understanding of legalities, consent, and privacy rights. By providing guidance and information on these topics, the chatbot can contribute to the promotion of responsible online behavior among adolescents. The vulnerability of adolescents to online threats, and the potentially harmful impact of such experiences on their development and mental health, highlight the need to address cyber safety

XSL•FO
RenderX

concerns among youth and ensure they have easy access to reliable information from trusted sources.

Yourtown [48], which operates Commonwealth-funded services for children across Australia, has reported that the use of on-demand helplines and websites by minors has steadily increased between 2018 and 2023. This paper suggests that chatbots could be a useful additional tool for helping address cyber safety and related mental health concerns among adolescents [49]. Integrating an online safety chatbot service into a school program could not only raise awareness of the risks associated with negative online interactions but also serve as an important resource for students seeking help. This online safety chatbot can provide direct links to information tailored to the students' concerns on issues such as cyberbullying, grooming, image-based abuse, and "sextortion," online gambling, eating disorders, and substance abuse—as well as counselling services, mental health resources, and trusted helplines. This tool could be particularly important in reaching those who may feel embarrassed or hesitant when talking to others, or who lack adequate support from family members or peers, leaving them without necessary guidance on these issues. An online safety chatbot service available on school websites could help schools bridge this gap by providing a readily accessible platform where students can seek help anonymously. By offering an additional avenue to access trustworthy information, an online safety chatbot can empower vulnerable adolescents to address their concerns proactively and confidentially and direct students to appropriate mental health support. It can also complement school and community efforts to promote online safety and well-being among young people.

Given the sensitive nature of online safety concerns, especially where adolescents may be coping with distressing emotions, a chatbot service can provide a safe and confidential space for adolescents to share their concerns [31]. This would allow the online safety chatbot to offer tailored recommendations that can be delivered empathetically and securely to help young people navigate the issues they are experiencing. This would be particularly helpful for those who may feel more comfortable asking about sensitive issues or sharing personal information to a dedicated online safety platform rather than a human [50]. However, as noted previously, the efficacy of a chatbot service is likely to be associated with users' trust in the chatbot, which, in turn, will affect how receptive users are to the chatbot's recommendations for further support.

This Study

In technology design, creating user-friendly systems is essential, particularly in sensitive information-sharing contexts [2,3,6]. When implementing an online safety chatbot in schools to help address online harms, it is important to include aspects of cognitive trust (eg, perceived utility) and affective trust (eg, feeling comfortable with its use). Several trust-related factors can influence how likely someone is to adopt and engage with a chatbot [3,16-19,26,51]. These include perceptions of the accuracy of the information the chatbot provides, how easy and user-friendly the chatbot is to interact with, and how well its suggestions align with a user's needs. The chatbot's ability to provide relevant recommendations based on the user's input, such as offering resources related to their specific concern, is important to building and maintaining trust. Additionally, feelings of trustworthiness, security, and positivity during interactions with the chatbot play a key role in impacting users' overall perception and willingness to continue using it [2,3,10-12,52]. Essentially, a user's perception of the chatbot's reliability and the more helpful and comfortable they feel with the chatbot, the more likely they may be to trust it and follow its recommendations for additional help. In this study, we included 6 measures related to trust in chatbots, tapping both cognitive and affective trust (accuracy, user-friendliness, relevant recommendations, trustworthiness, security, and positivity).

Menu-Driven and Simple Conversational Chatbot

To examine the potential for developing an online safety chatbot for a school or community setting, participants interacted with both a conversational and menu-based chatbot using a scenario-based activity (as explained further in the Methods section). A menu-driven chatbot offers users preselected options, guiding users through structured pathways based on their selections. In contrast, a conversational chatbot enables users to input questions or concerns using natural language. There are 2 main types of conversational chatbots: "simple chatbots" and "smart chatbots." Smart chatbots use advanced algorithms and AI to understand and respond to user inputs in a more flexible and natural manner. In contrast, a simple conversational chatbot, the system used in this study, simulates simple conversations, providing tailored responses and recommendations based on user textual input. They leverage advanced machine learning models to understand user input (trained on input phrases and rules) to establish user intent but then rely on predefined responses to provide answers to user input. Phrases like "Sorry, I did not understand" appear for inputs that do not match preselected intents, particularly in contexts where a less constrained chatbot may respond with unhelpful or inappropriate responses. There have been many examples in recent years of chatbots responding poorly, although none more famously than @TayAndYou [53]. The choice to implement a simple chatbot in this study was to minimize the risk to adolescents inquiring about sensitive topics.

Examining users' ratings for trust-related factors across the 2 chatbots allowed us to compare which method of interaction and design approach might better align with students' needs and preferences when seeking help navigating online safety issues. This study, being exploratory and intended as a proof of concept, investigates multiple research questions and considers alternative viewpoints. For example, students might trust a menu-driven chatbot more because it is predictable and easy to navigate with clear menu options, but this might feel less interactive and personalized compared to a conversational chatbot [54]. However, a simple conversational chatbot, though potentially more flexible and engaging, may be limited when handling complex sentences or questions. This could affect the accuracy of its responses, potentially impacting trust in the chatbot. A preference for a menu-driven chatbot or a conversational chatbot may depend on whether students prioritize a structured interface or a more conversational interaction.

XSL•FO
RenderX

Past research on technology adoption suggests perceptions of an online safety chatbot's utility may be influenced by gender-specific factors [16], highlighting the importance of considering these differences. For example, findings suggest that females tend to have a favorable view if they perceive the technology as practical, while males might need extra motivation to engage [55,56]. Thus, we examined potential gender differences in students' trust and their engagement with the chatbot's recommendations, as this may inform an online safety chatbot's marketing and design to better cater to the preferences of different gender groups, enhancing its effectiveness.

In sum, this study examined whether a chatbot might be a useful additional tool for meeting the needs of adolescents at risk from online harms. Specifically, in this proof-of-concept work, we evaluated users' (1) perceptions of efficacy for 2 types of chatbot, (2) trust in our chatbots, and (3) preferences for chatbot design features.

Research Questions

Textbox 1 outlines the study's research questions, which examine students' help-seeking behavior and their perceptions of trust and usability in chatbot systems. The questions focus on the relationships among key trust factors, behavioral intention to use chatbots, and differences between conversational and menu-driven designs. Additional areas of inquiry include user preference, personalization, anonymity, prior experience, gender differences, and common usability challenges, providing an overall framework for understanding trust and user experience in chatbot-based support.

Textbox 1. Research questions examining students' trust, usability perceptions, and behavioral intentions toward conversational and menu-driven chatbot systems.

Research questions

- How likely are students to seek help from outside their family and peer group?
- Do perceived accuracy, usefulness, ease of use, security, comfort, and credibility show a positive correlation, indicating that they may work together to contribute to trust in chatbot systems? Does this differ between conversational and menu-driven chatbots?
- Is there a relationship between a higher trust score (combined scores on perceived usefulness, ease of use, security, comfort, and credibility) and behavioral intention to use a chatbot (measured by clicking on a help link and positive feelings about the action)? Does this differ between conversational and menu-driven chatbots?
- Do conversational chatbots have a higher overall trust score than menu-driven chatbots, indicating that they are perceived as more trustworthy (based on perceived usefulness, ease of use, security, comfort, and credibility ratings)?
- Does perceived usefulness differ between conversational and menu-driven chatbots?
- In terms of perceived ease of use, is there a preference for conversational chatbots over menu-driven chatbots?
- Overall, do users prefer a conversational or menu-driven chatbot, and what factors influence users' preferences? (Qualitative)
- Is there a preference for personalization of the conversational chatbot? Is a face or name important?
- Is there a preference for strict anonymity in a chatbot system?
- How does prior experience with conversational chatbots affect users' perceptions of ease of use, comfort, and overall trust (trust score) in a conversational chatbot?
- What are the common difficulties encountered by users when using conversational and menu-driven chatbots, and how might these difficulties relate to user experience ("ease of use") and trust in the system ("trust score")? (Qualitative)
- Do trust scores differ according to user gender identity? Does gender impact a user's behavioral intention?

Methods

Ethical Considerations

All procedures were compliant with the relevant laws and institutional guidelines and have been approved by the appropriate institutional committee (March 26, 2024; project no: H0029893). Privacy rights for human subjects were observed, and informed consent was obtained prior to data collection. All data were collected anonymously and stored securely in line with the university research data governance guidelines. Participants completed the survey within their "home room" group at school, within the presence of a school staff member. A member of the research team was also present to answer questions from participants. The information sheet also provided guidance on contacting the school counselling team or 1 of 2 listed independent support lines (Beyond Blue and Headspace) if the survey caused any discomfort or raised concerns for participants.

Participants

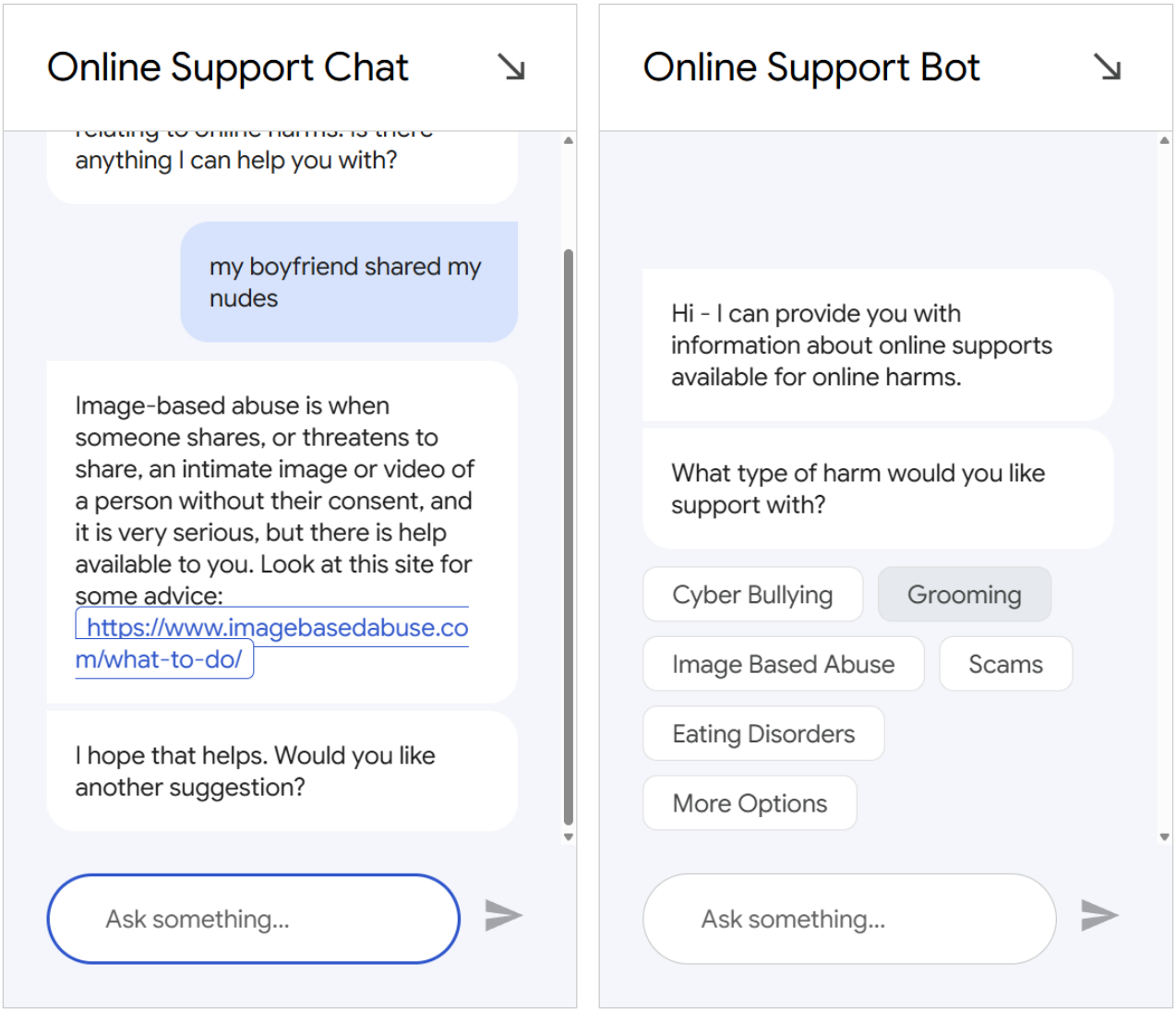
A total of 224 participants (131 female, 93 male, and 0 selecting a different identification), with a mean age of 16.8 years, completed the study. Participants were recruited from a local school, and both participant and parent or guardian information sheets were prepared and communicated to the relevant parties through the school communication system. This same system also provided a facility for parents or guardians to provide consent for participants. The school collected the list of students for whom consent was provided, and this list was later used to provide access to the survey to students who met the consent requirements.

XSL•FO
RenderX

Materials

The chatbot was developed using a co-design process, leveraging 2 focus groups at a local school that teaches grades 11 and 12. The first focus group comprised members of the pastoral care team at a school, and the second was a cohort of students. This design process is detailed in Roehrer et al [57]. The focus groups considered what harms were of greatest importance to be included, what resources were of most value, and general design considerations, including what the chatbot's opening line to users should be. The chatbot design did not attempt to mimic a young person's vocabulary or style in alignment with similarity-attraction theory [58] but aimed to use simple casual language that would be understandable at a 13-year-old reading level. The focus groups highlighted the importance of nonjudgmental advice that was factual and trustworthy. The chatbot included slang and other terms used by young people within its training data to understand inputs, but its responses did not include such terms and prioritized simple plain English responses that directed users to support resources and services. The chatbot was built on the Google Dialogflow platform [59] based on the outputs of the design requirements and in 2 variants, 1 menu-driven and 1 conversational (Figure 1). The survey was delivered to participants using the Qualtrics platform, accessed through a web link provided to them in the classroom at the time of the survey. The chatbot variants were embedded within the survey, with a randomized order of presentation of the 2 chatbot types. Following the survey, lunch was provided on the school campus for students who participated in the survey.

Figure 1. The 2 variants of the chatbot created within the project: the conversational chatbot (left) and menu-driven chatbot (right). The conversational chatbot allows the user to enter free text and engage in a conversation. The menu-driven chatbot prompts users to select response from prespecified lists.



Procedure

After reading the task instructions, participants were asked to enter their gender and age into text boxes. In addition, they were asked a single question about their general help-seeking behavior, "When something worries you, how likely are you to seek help from sources outside of your family or peer group?" (7-point scale: "very unlikely" to "very likely"). The main task consisted of 2 parts. In the first part of the task, participants engaged with the first chatbot (eg, conversational chatbot). They began by reading the first vignette, which provided context for

XSL•FO
RenderX

the interaction with the chatbot. Participants entered into a simulated "conversation" with the first interactive chatbot, assuming the role of the character in the given scenario. Based on the details shared by the participant, relevant hyperlinks to appropriate websites, such as eSafety [60] and the Butterfly Foundation [61], were embedded within the "conversations." These interactive clickable links led directly to real-world websites. For example, if the conversation related to cyberbullying, participants received a hyperlink directing them to eSafety. Additionally, the chatbot offered the option for another link to a different but relevant website. Participants received instructions at the start of the study that these hyperlinks were interactive. After engaging with the first vignette, participants proceeded to a second vignette, repeating the process (using the same interactive chatbot). After completing both interactions, participants responded to a set of 8 questions aimed at capturing their thoughts related to their experiences with the first chatbot. The following questions were asked for conversational and menu-driven chatbots following an interaction with that system and were aimed at capturing behavioral intention. Question 1: "For either vignette, did your character choose to click a support link?" (This was aimed at capturing behavioral intention). Question 2: "How did your character feel about this decision? e.g., did they feel positive about their decision, or did they have some reservation?" The participants then answered 6 questions on various aspects of trust-related factors. Participants were instructed to consider the perspective of the person seeking help while responding to these questions. The following questions related to trust with responses collected on a 5-point Likert scale, ranging from "strongly disagree" to "strongly agree": (1) "The interaction with the chatbot resulted in a recommendation for a relevant service" (perceived usefulness), (2) "Overall, I felt the chatbot provided accurate information" (perceived accuracy), (3) "The interaction with the chatbot felt user-friendly" (perceived ease of use), (4) "The interaction with the chatbot felt secure in terms of privacy" (perceived security), (5) "I felt that the interaction with the chatbot was a positive experience" (perceived comfort), and (6) "I found the chatbot to be trustworthy" (perceived credibility). Higher scores indicate an increased likelihood of trust in the chatbot. In the second part of the task, participants engaged with a different chatbot (eg, menu-driven chatbot). As in the first part, participants read 2 new vignettes, and engaged in a "conversation." Participants then answered the same set of 8 questions as in the first part. This approach allowed us to collect information on participants' engagement with the 2 distinct interactive chatbots (conversational and menu-driven), their perceptions of those interactions, and the likelihood of seeking further advice through a suggested helpline.

At the end, participants were presented with a final set of questions designed to gather their feedback and impressions directly comparing the 2 chatbot systems: Question 1: "After engaging with both systems, which did you find most helpful?" (Response options: conversational, menu-driven). Question 2: "After engaging with both systems, which did you find easier to use?" (Response options: conversational, menu-driven). Question 3: "If you wanted to find out more about support services and information related to online harms, which system would you prefer to use?" (Response options: conversational, menu-driven, neither). Question 4: "Do you believe the conversational chatbot should be personalised, for example, having a face or a name?" (Response options: yes, no, unsure). Question 5: "Do you think a chatbot service for information related to online harms should be strictly anonymous?" (Response options: yes, no, unsure). Question 6: "Before today, have you used a conversational chatbot before?" (Response options: yes, no, unsure). Question 7: "Please share your thoughts on what specific features would increase your comfort level when it comes to sharing personal information with a chatbot." Question 8: "Did you encounter any difficulties using either of the systems today?" For Questions 3, 4, 5, 7, and 8, participants were provided a text box to add any additional information.

Analysis Strategy

We used the *jamovi* [62] packages to run Bayesian analyses to examine our data. This approach provides 2 key advantages for our purposes. First, unlike null hypothesis significance testing (NHST), Bayesian analyses can quantify evidence for the null hypothesis. When a NHST approach returns a nonsignificant result, all that can be concluded is that there was no evidence of a meaningful difference between groups or association between variables. In contrast, a Bayesian approach allows researchers to draw meaningful conclusions about equivalence across conditions or the absence of associations between variables. Second, a Bayesian approach allows for targeted *t* tests and chi-square analyses without inflating the risk of Type I error [63]. When interpreting Bayes factors (BFs), we relied on Lee and Wagenmakers' [64] criteria, where evidence for the alternative (or null) hypothesis can be categorized as follows: 1-3=anecdotal, 3-10=moderate, 10-30=strong, 30-100=very strong, and >100=extreme. For Bayesian *t* tests, effect size measures are reported as δ, the population equivalent to Cohen *d* [65]. An adapted summative content analysis was performed on the survey responses, using the comments to drive the units of meaning, providing labels and then grouping labels into a final category [66]. Our analyses were largely exploratory, but the preregistration for our analysis strategy can be found here at the Open Science Framework [67].

Results

Overview

A key aim of this study was to examine aspects of trust in both a conversational and menu-driven online safety chatbot. As individuals typically require trust before they comply with recommendations [11,21], establishing and maintaining trust is an essential component to ensuring adolescents are receptive to its suggestions for related online e-safety resources.

Willingness to Seek Help From External Sources

Establishing whether young people are open to seeking help beyond their close friends and family circle is important, as this informs the need for and development of effective external support strategies. When participants were asked about the likelihood of seeking help outside their family and peer group, of the 224 participants, 39% (87/224) reported it was likely (collapsed across slightly, moderately, and very likely), 42%

XSL•FO
RenderX

(94/224) said it was unlikely (collapsed across slightly, moderately, and very unlikely), with the rest (n=43) being unsure. Thus, over one-third of students showed openness to seeking external help when needed. We measured students' willingness to follow a chatbot's recommended service by including clickable hyperlinks in their chatbot interactions and asking participants about their link-clicking behavior. We considered those who clicked on a help link might also be more likely to seek help when presented with that opportunity in real-life. When engaging with both the conversational and menu-driven chatbots, a substantial majority of the 224 participants reported they clicked on a help link (141/224, 63% and 142/224, 63% for the conversational and menu-driven chatbots, respectively). As students were no more or less likely to press on a help link via a conversational or menu-driven chatbot, this suggests that the type of interface (as they currently stand) does not influence help-seeking behavior. Although females were, numerically, more likely than males to click on help links provided by both the menu-driven (90/131, 69% vs 52/93, 56%) and conversational chatbots (87/131, 66% vs 54/93, 58%), the evidence in favor of these representing meaningful differences was anecdotal (BF_{10}<2.10).

Considering an individual's attitude, whether positive or negative, toward a behavior is important, as it can indicate the likelihood of engaging in that behavior in the future [42]. Positive feelings associated with a behavior often suggest a higher chance of repeating that behavior. Therefore, a positive feeling toward clicking a link in this study might suggest a tendency to seek assistance via a chatbot-suggested link in real-world situations. Bayesian independent samples 2-tailed *t* tests found that, for both the conversational ($BF_{10}=1.29\times10^7$) and menu-driven chatbots ($BF_{10}=3.73\times10^6$), there was extreme evidence that students who clicked on the provided link, compared with those who did not, felt more positive about their choice. For both the conversational (mean 3.50, SD 0.78 vs mean 2.76, SD 0.92; δ=0.87, 95% credible interval [CrI] 0.58-1.15) and the menu-driven chatbots (mean 3.49, SD 0.85 vs mean 2.82, SD 0.63; δ=0.84, 95% CrI 0.54-1.12), these differences exceeded the threshold for large effects. These findings suggest that choosing to click on a link while interacting with either the conversational or menu-driven chatbot is associated with a higher level of positivity regarding that decision, in contrast to choosing not to seek help via the link.

Trust and Trust-Related Factors

Trust is an important factor in users' willingness and motivation to follow advice (ie, behavioral intentions; [6,17,26]. Thus, we were interested in how trust relates to clicking on chatbot-suggested help links. We expected that higher overall trust in the chatbot should relate to a greater likelihood of clicking on a suggested help link. A Bayesian independent samples 2-tailed *t* test (1-tailed/2-tailed) showed extreme evidence (BF_{10}=87,421) for a large effect (δ=0.73, 95% CrI 0.46-1.00) such that, for the conversational chatbot, those who clicked on a help link (mean 3.68, SD 0.75) had a higher trust score (using the mean of the 6 trust-related factors) than those who did not click (mean 3.14, SD 0.65). Similarly, for the menu-driven chatbot, there was extreme evidence (BF_{10}=543,051) for a large effect (δ=0.78, 95% CrI 0.50-1.07), where those who clicked on a help link (mean 3.76, SD 0.73) had a higher trust score compared with those who did not (mean 3.22, SD 0.53). These findings are consistent with the idea that trust relates to a person's inclination to use the chatbot's suggested help features.

We compared participants' perceptions of menu-driven and conversational chatbots across various trust-related factors. As trust is a multifaceted concept, it is important to first consider the relationship between these factors and their potential collective impact on trust in chatbots. Key trust-related factors influencing chatbot adoption, which are essential for developing and maintaining trust, include the accuracy of information based on user input, user-friendliness, appropriateness of recommendations (eg, directing users to relevant help sites for specific concerns such as cyberbullying), and feelings of trustworthiness, security, and positivity toward interactions with the chatbot. The internal consistency of the trust-related factors in our study was high, with a Cronbach α coefficient of 0.88, indicating good reliability among the measured items (Tables 1 and 2). We found extreme evidence for moderate-to-strong correlations between all trust-related factors for both menu-driven and conversational chatbots, highlighting the factors' interconnectedness.

XSL•FO
RenderX

Table 1. Bayesian correlation matrix showing association between trust-related factors for the menu-driven chatbot.

Variable	Accuracy	User-friendly	Secure	Trustworthy	Positiveness	Recommended relevant
Accuracy						
r	—[a]					
BF_{10}	—					
User-friendly						
R	0.556	—				
BF_{10}	3.64×10^{16}	—				
Secure						
R	0.476	0.442	—			
BF_{10}	1.60×10^{11}	1.94×10^{9}	—			
Trustworthy						
R	0.545	0.498	0.714	—		
BF_{10}	5.85×10^{15}	3.66×10^{12}	1.13×10^{33}	—		
Positiveness						
R	0.577	0.674	0.474	0.564	—	
BF_{10}	1.85×10^{18}	8.15×10^{27}	1.27×10^{11}	1.70×10^{17}	—	
Recommended relevant						
R	0.598	0.482	0.465	0.571	0.519	—
BF_{10}	1.08×10^{20}	3.39×10^{11}	3.53×10^{10}	6.23×10^{17}	8.33×10^{13}	—

[a]Not applicable.

XSL•FO
RenderX

Table 2. Bayesian correlation matrix showing association between trust-related factors for the conversational chatbot.

Variable	Accuracy	User-friendly	Secure	Trustworthy	Positiveness	Recommended relevant
Accuracy						
R	—[a]					
BF_{10}	—					
User-friendly						
R	0.636	—				
BF_{10}	5.42×10^{23}	—				
Secure						
R	0.398	0.515	—			
BF_{10}	1.42×10^{7}	4.57×10^{13}	—			
Trustworthy						
R	0.514	0.578	0.704	—		
BF_{10}	3.84×10^{13}	2.07×10^{18}	4.48×10^{31}	—		
Positiveness						
R	0.668	0.681	0.600	0.631	—	
BF_{10}	1.35×10^{27}	4.82×10^{28}	1.93×10^{20}	1.61×10^{23}	—	
Recommended relevant						
R	0.550	0.501	0.400	0.422	0.483	—
BF_{10}	1.30×10^{16}	5.23×10^{12}	1.79×10^{7}	2.00×10^{8}	4.15×10^{11}	—

[a]Not applicable.

Table 3 displays the means and SDs for, and Bayesian 2-tailed paired samples *t* tests comparing, each trust-related factor across chatbot types. Across these comparisons, 3 key findings emerge. First, for most trust-related factors, there is evidence of equivalence across chatbot types, indicating that users did not report trusting one chatbot over the other. This evidence was strong for perceptions of security and trustworthiness, moderate for users' feelings of positiveness toward the chatbots and for the perceived relevance of the recommended information, but only anecdotal for perceptions of the accuracy of provided information. Second, there was moderate evidence that users viewed the menu-driven chatbot as more user friendly than the conversational chatbot, though the effect size only approach the cutoff to be considered small. Finally, there was anecdotal evidence suggesting that overall trust scores may have favored the menu-driven chatbot over the conversational chatbot. However, this interpretation requires 2 caveats: the effect size here was trivial, and any difference likely reflects differences in perceived user-friendliness. Together, these findings suggest that users generally reported equivalent trust for conversational and menu-driven chatbots, but that the conversational chatbot may need to be refined to improve usability.

Table 3. Mean (SD) for, and Bayesian paired samples 2-tailed t tests comparing, trust scores for the menu-driven and conversational chatbots.

Variable	Menu-driven chatbot, mean (SD)	Conversational chatbot, mean (SD)	BF_{10}[a]	δ (95% CrI[b])
Accuracy	3.54 (0.93)	3.42 (0.96)	2.46[a]	0.12 (–0.01 to 0.25)
User-friendly	3.62 (0.91)	3.45 (1.00)	3.09	0.18 (0.05 to 0.31)
Secure	3.51 (0.89)	3.48 (0.93)	10.65[a]	0.04 (–0.09 to 0.17)
Trustworthy	3.50 (0.91)	3.50 (0.94)	13.18[a]	0.01 (–0.12 to 0.14)
Positiveness	3.51 (0.88)	3.44 (0.93)	6.42[a]	0.08 (–0.05 to 0.21)
Recommended relevant	3.70 (0.90)	3.59 (1.02)	3.49[a]	0.11 (–0.02 to 0.24)
Overall trust	3.56 (0.71)	3.48 (0.76)	1.45	0.16 (0.03 to 0.29)

[a]Indicates Bayes factor (BF_{01}) expressing evidence in favor of the null/equivalence across chatbot types.

[b]CrI: credible interval (around the effect size).

XSL•FO
RenderX

When examining gender effects on overall trust, Bayesian independent samples *t* tests revealed anecdotal (BF_{01}=2.40; δ=0.19, 95% CrI –0.07 to 0.45) and moderate evidence for the null (BF_{01}=5.12; δ=0.09, 95% CrI –0.17 to 0.36) for the menu-driven and conversational chatbots, respectively. Thus, for both types of chatbot, female and male users reported similar levels of trust (mean 3.62, SD 0.63 and mean 3.48, SD 0.82, for female and male users rating the menu-driven chatbot; and mean 3.51, SD 0.70 and mean 3.43, SD 0.84, for female and male users rating the conversational chatbot).

User Preferences

Despite evidence of greater user-friendliness when considering the trust-related factors, when participants were asked directly which chatbot was easier to use, there was no difference between the conversational (118/224, 53% of participants) and the menu-driven (106/224, 47%) chatbot, with a Bayesian binomial test revealing moderate evidence in favor of the null (BF_{01}=8.70). However, when asked which chatbot was most helpful, there was anecdotal evidence supporting a preference for the conversational (131/224, 58%) over menu-driven (93/224, 42%) chatbot (BF_{10}=2.10).

Taken together, our findings suggest that although there were subtle differences in user-friendliness and overall trust between the menu-driven and conversational chatbot systems, these differences were not substantial enough to practically impact users' perceptions of trust across chatbots. Trust levels for each chatbot were consistently just above the neutral midpoint, indicating a generally positive perception. When students were explicitly asked which chatbot they would prefer to use if they were to seek help, a majority appeared to favor the conversational chatbot (92/224, 41%), over a menu-driven chatbot (74/224, 33%) or choosing neither system (58/224, 26%). However, a Bayesian multinomial test run in the *JASP* statistics package [68] provided anecdotal evidence for the null (BF_{01}=1.96), suggesting there was no statistical evidence for a preference regarding chatbot type.

Students provided 245 comments when indicating their preference for a chatbot. Slightly more commentary was provided regarding the conversational chatbot (n=88 in comparison to n=75 that specifically focused on the menu-driven chatbot), and of these 88 comments, a total of 66 were comments that supported the use and features of the conversational chatbot. These codes directly mentioned easy or easier use or interaction or understanding (n=12), being friendlier (n=1), more realistic (n=8), more relatable (n=2), more helpful (n=5), like a person/human (n=6). The codes also mentioned a greater flow of conversation (n=6) and the ability to frame the issue in their own words (n=14). Students also spoke about the conversational chatbot's ability to tailor the conversation or interaction to their specific problem (n=5). Students spoke of the perception of the conversational chatbot taking the burden of the problem (n=4) and helping them navigate to meaningful resources (n=4). Of the 75 menu-driven chatbot comments, a total of 65 were in support of the use and features of the menu-driven chatbot. These codes directly mentioned easy or easier use or interaction or understanding (n=10), being efficient (n=5), more realistic (n=8), access to resources (n=11), more helpful (n=5), quick access to information (n=3), and helpful when interacting due to a menu, prompt, or list of options (n=8). Students found the menu-driven chatbot less invasive (n=5), found that the menu prompts helped frame the issue they were facing (n=6), and preferred the option of not having to type (n=5). Concepts of "trust" were directly mentioned (n=6); however, these were either in the context of not trusting AI, "AI doesn't feel supportive or trustworthy," or in defining what a trusted person would look like ("Trusted person would be patient").

Students differed in terms of their prior experience with chatbots similar to the conversational one used in this study, with the majority of participants (134/224, 60%) reporting no previous interaction with a similar conversational chatbot, while 33% (73/224) had prior experience, and 7% (17/224) were uncertain about their past interactions. A user's degree of prior experience with similar chatbots may adversely affect participants' perceptions of the conversational chatbot regarding overall trust and user-friendliness. However, prior experience did not affect either overall trust (BF_{01}=13.04) or perceptions of user-friendliness (BF_{01}=9.35; Table 4 provides descriptive statistics).

Table 4. Mean ratings of overall trust and user-friendliness for the conversational chatbot, according to prior experience with conversational chatbots.

Measure	Response	Rating, mean (SD)
Overall trust	Yes	3.44 (0.82)
Overall trust	No	3.49 (0.75)
Overall trust	Unsure	3.52 (0.62)
User-friendliness	Yes	3.37 (1.07)
User-friendliness	No	3.47 (0.97)
User-friendliness	Unsure	3.65 (0.86)

Unsurprisingly, anonymity emerged as a significant concern among students, with 69% (155/224) expressing a preference for anonymous interactions with an online safety chatbot; only 10% (22/224) preferred nonanonymous interactions, while 21% (46/224) were undecided. These results confirm the importance of user privacy to adolescents in an e-safety context. When asked about personalizing a conversational chatbot, such as giving it a face or name, 39% (88/224) favored personalization, 34% (75/224) preferred a nonpersonalized chatbot, and 27% (61/224) were uncertain, demonstrating no significant preference for personalization of the chatbot. Thus, while some users may want a more engaging and relatable interaction with the chatbot,

XSL•FO
RenderX

others do not perceive the need to add human-like characteristics in this context.

Students directly mentioning the conversational chatbot included some concerns or issues with the conversational chatbot (n=11). These comments were focused on unhelpful responses from the chatbot (output did not make sense from input given, n=5), causing the student to hesitate in interaction (due to needing to type out the issue, n=1), perceiving to not provide advice (n=1), causing greater distress (due to misunderstanding from the chatbot, n=1), or uncomfortable (n=1), hard (n=1), and artificial (n=1) to interact with. Student comments directly mentioning the menu-driven chatbot also included concerns or issues with the menu-driven chatbot (n=6). These comments focused on the menu-driven chatbot's inability to handle or misunderstand student queries (n=3), the perception of the platform as more closed off (n=1), unhelpful (n=1), and limited in the range of topics it could offer advice on (n=1).

Of the remaining 82 comments (total comments n=245), students spoke about a negative view of turning to AI. Some students did not feel they could trust (n=3) or turn to AI (n=2), and some felt that using AI for help with online harms was inappropriate (n=8). Some students preferred to turn to people they trust (n=9) or look up the information themselves (n=5). Some students stated they did not prefer either chatbot (n=10). While others did not have a clear answer (n=2) or did not know which chatbot they preferred (n=10). However, some identified that both chatbots were appropriate/relevant resources (n=5). A few students expressed feeling confused (n=2) or frustrated (n=1) by the responses provided by either (or not specifically named) chatbots. The students identified the potential serious consequences of chatbots failing to provide correct or meaningful interaction during crises (n=2). Students directly spoke about improving the conversational chatbot (n=10), noting its limited conversation abilities and the need for trust in the interaction (n=6), as well as a more human-like approach to engage users (n=2). Additionally, students spoke about the need for the chatbot to provide advice, not just links (n=5), and felt that better options were already available (n=4). Thus, overall, students identified the importance of improving chatbots for more meaningful interactions and advice, highlighting the need for trust and a human-like approach.

Summary of Key Findings

A substantial percentage of respondents (87/224, 39%) expressed a willingness to seek help outside their family and social circle when dealing with online harms, and approximately 60% (conversational chatbot: 141/224 and menu-driven chatbot: 142/224) of respondents engaged with our chatbot systems. When dealing with chatbot systems, our results highlight 2 related issues of importance. First, trust was strongly associated with willingness to click links provided by our chatbots. Perceptions of trust were generally similar across chatbot types and for male and female participants. Second, most respondents (155/224, 70%) expressed a preference for anonymity when dealing with chatbot systems. Thus, to be effective, and capitalize on adolescents' willingness to seek external support, chatbot systems must be able to build and maintain trust with users, and it will be necessary to assure users of their continued anonymity.

Discussion

Overview

Adolescents' vulnerability to online dangers, and the associated potential negative impacts on their mental well-being, highlight the urgent need to address cyber safety issues among young people. It is essential to provide young people easy access to trustworthy information and support. Chatbots represent a promising tool in addressing cyber safety concerns, offering convenient access to established and reputable support services, and helping link students with appropriate clinical support providers. We explored whether trust in chatbot systems was influenced by the design of the chatbot. Examining students' perception of 2 chatbot systems, a menu-based and conversational chatbot, in terms of key trust-related factors could inform the development of an effective user-centered chatbot in the educational online safety context.

Trust and Chatbot Efficacy

Individuals usually need to feel trust before following recommendations [11,17-19,21]. Thus, establishing and maintaining trust in an e-safety chatbot is vital to ensure adolescents are comfortable using the system and receptive to its suggestions for online safety resources. Comparing adolescents' perceptions of trust for menu-driven and conversational chatbots allowed us to examine which interaction method and design approach better suited students' preferences. As trust is multifaceted, we evaluated how key trust-related factors (ie, perceived accuracy of information, user-friendliness, appropriateness of the chatbots' recommendations, and feelings of trustworthiness, security, and positivity toward the interaction) impacted overall trust across chatbot systems. Both chatbots were viewed more positively than negatively in terms of overall trust, indicating that they are both viable options for a trusted online safety chatbot. When students were directly asked which chatbot they would prefer for seeking help, there was some evidence for equivalence, suggesting no meaningful difference between the 2 options, despite a higher number of students endorsing the conversational chatbot over the alternative. Student comments supported the endorsement of the conversational chatbot, highlighting its "conversational nature" and "interactive" elements that simulated real-life conversations, "it was easy to communicate with and share my deepest and darkest thoughts" and "You are able to describe what you need help with in greater detail." The perceived ability for the conversational chatbot to emulate person-to-person interaction facilitated some preference for this mode of delivery, "the conversational one was easier to understand and felt like i was talking to someone rather than clicking a menu." However, there remains room for improvement based on the mean trust ratings.

Help-Seeking Behavior

Over a third of participants were open to seeking help beyond their friends and family, indicating that many young people could benefit from support provided by an online safety chatbot.

XSL•FO
RenderX

Further, there is a positive link between the intention to use tech tools, like chatbots, and their actual usage [44]. Thus, to assess users' readiness to engage with the chatbot's recommended resources, we included clickable help links to external support sites and asked users whether they clicked a link. The majority of students reported clicking a support link, with no difference between chatbot systems, indicating the interface type did not impact the likelihood of accessing recommendations for further assistance. Links were perceived by students as being helpful, with one student mentioning, "it provides useful links for kids/teens in a way that makes them feel more comfortable without seeking outside help." However, another student suggested integrating the links into the conversation, "the links should be embedded into script rather than just being a plain link, it is not a big step but it definitely makes you more likely to click on a link." There was some, albeit weak, evidence that males were less likely than females to click on help links, aligning with the well-documented gender disparity in help-seeking behavior [69]. Although participants in this study reported identifying as either male or female, it is also important to consider the perspectives of gender-diverse populations, as this at-risk group often faces extra challenges and unique concerns. Understanding specific gender-related issues that could hinder help-seeking is important for developing an inclusive online safety chatbot that caters to all young people. Unfortunately, gender-diverse populations were under-represented in our sample.

Trust was an important factor in user behavior within chatbot interactions, as the level of trust students had in the chatbot was related to the likelihood of clicking a link (engaging with a linked support service) within the chatbot interaction. Additionally, we found choosing to click on a link was related to a higher sense of positivity regarding that choice, as opposed to opting not to press a link. The underlying mechanism here can be interpreted in different ways. One possibility is that a feeling of positivity arises from the decision itself, suggesting an associated sense of satisfaction in taking a proactive step toward seeking help. Alternatively, it could reflect a preexisting state of positivity or readiness to engage with the provided resources, influencing the decision to press the help link. Further examination of these mechanisms might provide greater understanding of users' decisions to seek help and their interactions with a chatbot, particularly concerning the inclusion of links to external support resources and how they perceive these resources. This may inform strategies aimed at improving the chances of adolescents accessing online support when needed.

Design Implications: Empathy, Personalization, and Anonymity

To explore students' interest in an online safety chatbot, we used a simple conversational chatbot, which operates with predefined outputs that are matched to conversational intents. Thus, although the current results serve as a useful proof of concept, the data from this study also highlight potential areas of improvement, particularly in terms of increasing cognitive trust (through accuracy and user-friendliness) and affective trust (through empathetic responses and personalization). The simple conversational chatbot appeared to struggle at times when handling students' written input. One student expressed this frustration: "Sometimes I found it hard to specifically say the situation and get the bot to understand." An inability to handle requests effectively may have lowered trust in the chatbot's competence and would seem like a substantial issue in applied settings. As another student noted, "it just kept saying "please re-phrase that" or "thats [sic] not relevant to the issue" when it was." These issues will need to be addressed to improve user experience and effective delivery of online safety support. User-friendliness could be improved by training the underlying machine learning model on training data from adolescents to better establish the user intent. The current chatbot was primarily trained on adults, with input from only a handful of younger adults, and based on comments made during the focus groups conducted during the co-design process. Additional sampling of several hundred, or thousand, students could create a much larger training dataset to produce a more robust model of user intent (the ethics approval for this study precluded the evaluation data being used as a training dataset for the model). While a menu-driven chatbot's simple interface with clear options makes navigation easy, predictable, and reliable, its standardized interactions may reduce emotional connection with users. In contrast, a conversational chatbot can be more emotionally engaging due to its capacity for natural language and empathetic responses. The simple conversational chatbot used in this exploratory study was focused on delivering support links and had limited natural interaction and empathic capabilities. Based on feedback from students (ie, the intended users of the chatbot), it seems that there was a desire for a more empathetic and personalized chatbot. For example, one student stated:

> *The understanding of certain kid/teens phrases or slang… more of a response aside from a brief description and a link, something like what snapchat AI would respond with more care and emotion.*

Future research should explore enhancing empathetic interactions, such as using warmer language and including more reassuring phrases, as prior studies indicate the potential benefits of a strengthening the user connection with the chatbot [2,70-74]. By incorporating these improvements into future designs, we may create an engaging and informative system, leading to a positive user experience and effective delivery of online safety support.

Personalization features, such as giving a chatbot a face or name, have the potential to enhance user engagement and satisfaction by creating a more personalized and human-like experience [30,71]. However, our current findings indicate that while some users want a more engaging and relatable interaction with the chatbot, others do not see the need for human-like characteristics. One student stated, "I think that if the chatbot was personalised it would help develop more of a connection with those seeking support and as though they aren't talking to a robot" while others preferred a nonpersonalized chatbot, explaining:

> *Having a non personalised Chatbot makes the information feel more trustworthy and reliable…" and "Giving the chat bot a face would also make it*

XSL•FO
RenderX

harder to communicate your problems, I believe it should be similar to a confessional booth.

This was surprising, as studies have found including a human-like visual cue in a chatbot can improve trust and reduce the psychological distance between the user and the chatbot [3,33,71,75]. However, the opposition, or hesitation, to personalization among some participants may stem from concerns related to privacy—known as the "valley effect," where users become increasingly concerned about privacy as technology becomes more human-like in its interactions [76,77]. Thus, if personalized features were to be incorporated in future designs, there may be a need to clearly address concerns related to privacy to ensure users feel comfortable and trusting in their interactions with the chatbot.

Anonymity was a significant concern among the students, with an overwhelming preference that an online safety chatbot should be anonymous. When users feel confident that their conversations are confidential, they are more likely to engage openly and honestly with the chatbot. One student noted that "Having the chatbot clearly anonymous will encourage openness" while another explained, "Anonymity is essential for someone to share their true, unfiltered opinions on the internet, to a 'secure' bot or to another person." Many users may be dealing with sensitive issues, such as online harassment, image-based abuse, substance abuse, or mental health challenges. They may feel a sense of shame or embarrassment if this information becomes known. As one student noted, "Having this privacy could encouraged [sic] those that may be embarrassed by their situation or not willing to share when their name is attached to it." Another mentioned, "…it provides the person in need with a safe outlet whilst also giving them a reliable source that they dont [sic] have to be ashamed about." The sharing of personal information can facilitate better support and guidance tailored to their individual needs. A chatbot service can provide a safe, confidential space for adolescents to share their concerns to help them navigate their problems. On a related note, it will be vital that data security measures ensure the privacy and confidentiality of any data supplied by users accessing a chatbot system like this in applied settings. Any misuse of data provided by users would not only be extremely unethical, but would rightfully damage users' trust in the chatbot system and consequently negate its efficacy.

Limitations

Given that this is a proof-of-concept study, there are limitations to the generalizability of the results obtained. Some of these limitations relate to the number and nature of vignettes explored and to the nature of the sample tested. Obviously, a chatbot may be accessed as a source of support by individuals experiencing a wide range of personal circumstances, and our stimulus and participant sampling does not provide a comprehensive representation of population of interest (eg, our data do not speak to how user characteristics like ethnicity, gender diversity, or sexual orientation affect the actual or perceived utility of the chatbot). Moreover, although our data are promising and serve as a proof-of-concept, promising findings in controlled, simulated testing environments do not guarantee efficacy in field settings. Thus, although our data provide initial support for the utility of chatbots in supporting the needs of adolescents at risk of online harms, further work is required to establish the generalizability of these findings. Further, there are aspects of trust and of chatbot technology that require additional attention. A trust-related factor we did not examine, but that requires attention, is that trust in technology systems can be influenced by the credibility of the organization involved, and their knowledge base (ie, that provides advice and guidance) [78]. In the context of health-related services, trust in the health information provider is crucial in increasing compliance with professional advice [79,80]. If the associated organization lacks credibility, people are less likely to follow recommendations [11,21]. However, as adolescents have identified schools and reputable online safety websites as their preferred sources for accessing online safety information [45], a school- or community-based system has the potential to gain students' trust. Thus, in the context of an online safety chatbot, its association with reputable institutions like schools and government-funded organizations dedicated to online safety (eg, Butterfly Foundation, Headspace) may improve access to quality support for cyber safety issues.

Another issue is that chatbot technology continues to evolve, as large language models (LLMs) provide an opportunity to create more lifelike and realistic dialogue. These LLMs are advanced AI models trained on vast amounts of text data, which could be selected to be appropriate and realistic for an adolescent audience. They are capable of ingesting and then generating human-like text across a wide range of topics. While LLMs are not specifically chatbots, they can be used as the underlying technology for conversational agents. They excel at generating coherent and contextually relevant responses to user input, even in complex conversational scenarios. LLMs are highly versatile and can be fine-tuned for specific tasks, including chatbot applications, content generation, and language translation. Such chatbots, as with the curated set of resources used in the chatbot in this study, will require maintenance, as LLMs are only aware of resources and supports contained within their training datasets. Currency and relevancy of support provided will be an ongoing challenge for each chatbot architecture.

Conclusion

Young people understand the value of obtaining more information about online risks, not only to deal with their own negative experiences but also to support others in similar situations [45]. This information also plays a role in promoting responsible online behavior, as teens might unknowingly cause harm due to their limited understanding of legalities, consent, and privacy. An online safety chatbot has the potential to serve as a valuable supplementary cyber safety resource alongside school support systems, such as pastoral care or school psychologists, with the aim to increase accessibility for adolescents navigating online safety challenges. Integrating a chatbot into schools' digital libraries, resources, or a student portal, can promote a collaborative approach to e-safety education and support. Online safety chatbots guided by a framework of trustworthiness can support users in navigating online risks and accessing valuable resources. In designing and deploying an online safety chatbot, it will be essential to build trust to ensure users feel emotionally connected and comfortable

XSL•FO
RenderX

with its use in navigating online risks. This technology-driven approach may be particularly helpful for those who prefer digital interactions or feel apprehensive seeking help directly from others. An online safety chatbot can offer a confidential and easily accessible space where adolescents can seek guidance and access information from reputable organizations dedicated to young people and online safety.

Funding

This project was funded through the eSafety Commissioner’s Online Safety Grants Program.

Authors' Contributions

JS, JP, and ER acquired the funding from the eSafety Commissioner’s Online Safety Grants Program. JS, JP, ER, MC, JDS, and PW conceived the studies, drafted the original paper, and edited the final manuscript. JS, ER, and JDS collected the data. JDS, MC, and ER performed the analyses.

Conflicts of Interest

None declared.

References

1. Jiang K, Qin M, Li S. Chatbots in retail: how do they affect the continued use and purchase intentions of Chinese consumers? Journal of Consumer Behaviour 2022:756-772 [FREE Full text] [doi: 10.1002/cb.2034]
2. Miner AS, Milstein A, Hancock JT. Talking to machines about personal mental health problems. JAMA 2017;318(13):1217-1218. [doi: 10.1001/jama.2017.14151] [Medline: 28973225]
3. Park G, Chung J, Lee S. Human vs. machine-like representation in chatbot mental health counseling: the serial mediation of psychological distance and trust on compliance intention. Curr Psychol 2023;43(5):4352-4363. [doi: 10.1007/s12144-023-04653-7]
4. Shah J, DePietro B, D'Adamo L, Firebaugh M, Laing O, Fowler L. Development and usability testing of a chatbot to promote mental health services use among individuals with eating disorders following screening. International Journal of Eating Disorders 2022;55(9):1229-1234 [FREE Full text] [doi: 10.1002/eat.23798]
5. Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M. Perceptions and Opinions of Patients About Mental Health Chatbots: Scoping Review. J Med Internet Res 2021;23(1):e17828 [FREE Full text] [doi: 10.2196/17828] [Medline: 33439133]
6. Lattie EG, Stiles-Shields C, Graham AK. An overview of and recommendations for more accessible digital mental health services. Nat Rev Psychol 2022;1(2):87-100 [FREE Full text] [doi: 10.1038/s44159-021-00003-1] [Medline: 38515434]
7. Pereira J, Díaz Ó. Using health chatbots for behavior change: a mapping study. J Med Syst 2019;43(5):135. [doi: 10.1007/s10916-019-1237-1] [Medline: 30949846]
8. Torous J, Bucci S, Bell I, Kessing L, Faurholt-Jepsen M, Whelan P. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883]
9. Featherman M, Miyazaki A, Sprott D. Reducing online privacy risk to facilitate e‐service adoption: the influence of perceived ease of use and corporate credibility. J Serv Mark 2010;24(3):219-229. [doi: 10.1108/08876041011040622]
10. Ho A, Hancock J, Miner A. Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. J Commun 2018;68(4):712-733 [FREE Full text] [doi: 10.1093/joc/jqy026] [Medline: 30100620]
11. Liu B. In AI we trust? Effects of agency locus and transparency on uncertainty reduction in human–AI interaction. J Comput-Mediat Commun 2021;26(6):384-402. [doi: 10.1093/jcmc/zmab013]
12. Yang J, Chen Y, Por LY, Ku CS. A systematic literature review of information security in chatbots. Appl Sci 2023;13(11):6355. [doi: 10.3390/app13116355]
13. Glikson E, Woolley AW. Human trust in artificial intelligence: review of empirical research. Acad Manage Ann 2020;14(2):627-660. [doi: 10.5465/annals.2018.0057]
14. Jacovi A, Marasović A, Miller T, Goldberg Y. Formalizing trust in artificial intelligence: prerequisites, causes and goals of human trust in AI. 2021 Presented at: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT ’21); March 3-10, 2021; Virtual Event. [doi: 10.1145/3442188.3445923]
15. Kumar V. Intelligent Marketing: Employing New-Age Technologies. New Delhi: SAGE Publications India Pvt Ltd URL: https://sk.sagepub.com/book/mono/intelligent-marketing/toc [accessed 2025-12-25]
16. Ling E, Tussyadiah I, Tuomi A, Stienmetz J, Ioannou A. Factors influencing users' adoption and use of conversational agents: a systematic review. Psychol Mark 2021;38(7):1031-1051 [FREE Full text] [doi: 10.1002/mar.21491]
17. Moilanen J, van Berkel N, Visuri A, Gadiraju U, van der Maden W, Hosio S. Supporting mental health self-care discovery through a chatbot. Front Digit Health 2023;5:1034724 [FREE Full text] [doi: 10.3389/fdgth.2023.1034724] [Medline: 36960179]

XSL•FO
RenderX

18. Raiche A, Dauphinais L, Duval M, De Luca G, Rivest-Hénault D, Vaughan T. Factors influencing acceptance and trust of chatbots in juvenile offenders' risk assessment training. Front Psychol 2023;14. [doi: 10.3389/fpsyg.2023.1184016]
19. Riedl R. Is trust in artificial intelligence systems related to user personality? Review of empirical evidence and future research directions. Electron Mark 2022;32(4):2021-2051. [doi: 10.1007/s12525-022-00594-4]
20. Mayer RC, Davis JH, Schoorman FD. An integrative model of organizational trust. Acad Manage Rev 1995;20(3):709-734. [doi: 10.2307/258792]
21. Przegalinska A, Ciechanowski L, Stroz A, Gloor P, Mazurek G. In bot we trust: a new methodology of chatbot performance measures. Bus Horiz 2019;62(6):785-797 [FREE Full text] [doi: 10.1016/j.bushor.2019.08.005]
22. Komiak, Benbasat. The effects of personalization and familiarity on trust and adoption of recommendation agents. MIS Q 2006;30(4):941-960. [doi: 10.2307/25148760]
23. Özer, Zheng Y, Ren Y. Trust, trustworthiness, and information sharing in supply chains bridging China and the United States. Manag Sci 2014;60(10):2435-2460. [doi: 10.1287/mnsc.2014.1905]
24. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
25. Cardona D, Janssen A, Guhr N, Breitner M, Milde J. A matter of trust? Examination of chatbot usage in insurance business. 2021 Presented at: Proceedings of the 54th Hawaii International Conference on System Sciences; January 5-8, 2021; Honolulu. [doi: 10.24251/hicss.2021.068]
26. Gopinath K, Kasilingam D. Antecedents of intention to use chatbots in service encounters: a meta-analytic review. Int J Consum Stud 2023;47(6):2367-2395 [FREE Full text] [doi: 10.1111/ijcs.12933]
27. Lewis JD, Weigert A. Trust as a social reality. Soc Forces 1985;63(4):967-985. [doi: 10.1093/sf/63.4.967]
28. McAllister DJ. Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. Acad Manage J 1995;38(1):24-59. [doi: 10.2307/256727]
29. Qiu L, Benbasat I. Evaluating anthropomorphic product recommendation agents: a social relationship perspective to designing information systems. J Manage Inf Syst 2009;25(4):145-182. [doi: 10.2753/MIS0742-1222250405]
30. Li X, Sung Y. Anthropomorphism brings us closer: the mediating role of psychological distance in user–AI assistant interactions. Comput Hum Behav 2021;118:106680 [FREE Full text] [doi: 10.1016/j.chb.2021.106680]
31. Skjuve M, Følstad A, Fostervold K, Brandtzaeg P. My chatbot companion — a study of human-chatbot relationships. Int J Hum-Comput Stud 2021;118:102601 [FREE Full text] [doi: 10.1016/j.ijhcs.2021.102601]
32. Wang Z, Walther JB, Pingree S, Hawkins RP. Health information, credibility, homophily, and influence via the Internet: web sites versus discussion groups. Health Commun 2008;23(4):358-368. [doi: 10.1080/10410230802229738] [Medline: 18702000]
33. Cheng X, Zhang X, Cohen J, Mou J. Human vs AI: understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust and relationship norms. Inf Process Manage 2022;59(3):102940 [FREE Full text] [doi: 10.1016/j.ipm.2022.102940]
34. Pesonen JA. 'Are you OK?' students' trust in a chatbot providing support opportunities. In: Lecture Notes in Computer Science, vol 12785. Cham: Springer International Publishing; 2021:199-215.
35. Kyung N, Kwon HE. Rationally trust, but emotionally? The roles of cognitive and affective trust in laypeople's acceptance of AI for preventive care operations. Prod Oper Manag 2025. [doi: 10.1177/10591478231225891]
36. Punyatoy P. Effects of cognitive and affective trust on online customer behavior. Mark Intell Plann 2018;37(1):80-96 [FREE Full text] [doi: 10.1108/mip-02-2018-0058]
37. Chong A, Blut M, Zheng S. Factors influencing the acceptance of healthcare information technologies: a meta-analysis. Inf Manage 2022;59(3):103604 [FREE Full text] [doi: 10.1016/j.im.2022.103604]
38. Müller L, Mattke J, Maier C, Weitzel T, Graser H. Chatbot acceptance: a latent profile analysis on individuals' trust in conversational agents. 2019 Presented at: Proceedings of the 2019 on Computers and People Research Conference (SIGMIS-CPR '19); June 20-22, 2019; Nashville URL: https://doi.org/10.1145/3322385.33223922019 [doi: 10.1145/3322385.3322392]
39. Nicolaou A, McKnight D. Perceived information quality in data exchanges: effects on risk, trust, and intention to use. Inf Syst Res 2006;17(4):332-351 [FREE Full text] [doi: 10.1287/isre.1060.0103]
40. Ajzen I, Fishbein M. Attitude-behavior relations: a theoretical analysis and review of empirical research. Psychol Bull 1977;84(5):888-918. [doi: 10.1037/0033-2909.84.5.888]
41. Hung S, Ku C, Chang C. Critical factors of WAP services adoption: an empirical study. Electron Commer Res Appl 2003;2(1):42-60 [FREE Full text] [doi: 10.1016/s1567-4223(03)00008-5]
42. Schifter DE, Ajzen I. Intention, perceived control, and weight loss: an application of the theory of planned behavior. J Pers Soc Psychol 1985;49(3):843-851. [doi: 10.1037//0022-3514.49.3.843] [Medline: 4045706]
43. Wang S, Huang Y, Wang C, editors. A model of consumer perception and behavioral intention for AI service. 2020 Presented at: Proceedings of the 2020 2nd International Conference on Management Science and Industrial Engineering; April 7, 2020; Osaka URL: https://doi.org/10.1145/339673396791 [doi: 10.1145/3396743.3396791]

XSL•FO
RenderX

44. Dinh C, Park S. How to increase consumer intention to use Chatbots? An empirical analysis of hedonic and utilitarian motivations on social presence and the moderating effects of fear across generations. Electron Commer Res 2023;24(4):2427-2467. [doi: 10.1007/s10660-022-09662-5]
45. The digital lives of Aussie teens. eSafety Commissioner. 2021. URL: https://www.esafety.gov.au/research/digital-lives-of-aussie-teens#:~:text=44%25%20of%20teens%20had%20a,%2Fsocial%20groups%20(16%25): [accessed 2025-12-25]
46. Lonergan A, Moriarty A, McNicholas F, Byrne T. Cyberbullying and internet safety: a survey of child and adolescent mental health practitioners. Ir J Psychol Med 2023;40(1):43-50. [doi: 10.1017/ipm.2021.63] [Medline: 34581261]
47. Prasad S, Ait Souabni S, Anugwom G, Aneni K, Anand A, Urhi A. Anxiety and depression amongst youth as adverse effects of using social media : a review. Ann Med Surg (Lond) 2023;85(8):3974-3981 [FREE Full text] [doi: 10.1097/MS9.0000000000001066] [Medline: 37554895]
48. Yourtown. Yourtown Annual Report 2023. URL: https://publications.yourtown.com.au/2023-annual-report/page/14-15: [accessed 2025-12-25]
49. Dosovitsky G, Bunge E. Development of a chatbot for depression: adolescent perceptions and recommendations. Child Adolesc Ment Health 2023;28(1):124-127. [doi: 10.1111/camh.12627] [Medline: 36507594]
50. Ta V, Griffith C, Boatfield C, Wang X, Civitello M, Bader H. User experiences of social support from companion chatbots in everyday contexts: thematic analysis. J Med Internet Res 2020;22(3):e16235 [FREE Full text] [doi: 10.2196/16235] [Medline: 32141837]
51. Pillai R, Sivathanu B. Adoption of AI-based chatbots for hospitality and tourism. Int J Contemp Hosp Manag 2020;32(10):3199-3226 [FREE Full text] [doi: 10.1108/ijchm-04-2020-0259]
52. Jian J, Bisantz AM, Drury CG. Foundations for an empirically determined scale of trust in automated systems. Int J Cogn Ergonomics 2000;4(1):53-71 [FREE Full text] [doi: 10.1207/S15327566IJCE0401_04]
53. Zemčík T. Failure of chatbot Tay was evil, ugliness and uselessness in its nature or do we judge it through cognitive shortcuts and biases? AI Soc 2021;36(1):361-367. [doi: 10.1007/s00146-020-01053-4]
54. Nguyen Q, Sidorova A, Torres R. User interactions with chatbot interfaces vs. menu-based interfaces: an empirical study. Comput Hum Behav 2022;128:107093 [FREE Full text] [doi: 10.1016/j.chb.2021.107093]
55. Milis K, Wessa P, Poelmans S, Doom C, Bloemen E. The impact of gender on the acceptance of virtual learning environments. 2008 Presented at: Proceedings of the International Conference of Education, Research and Innovation; November 17-19, 2008; Madrid p. 1-12 URL: https://tinyurl.com/mthemap3
56. Wolpin S, Berry D, Austin-Seymour M, Bush N, Fann JR, Halpenny B. Acceptability of an electronic self-report assessment program for patients with cancer. Comput Inform Nurs 2008;26(6):332-338. [doi: 10.1097/01.ncn.0000336464.79692.6a]
57. Roehrer E, Pokawinkoon P, Watters P, Sauer JD, Scanlan J. Adolescent-centric design of an online safety chatbot. J Comput Inf Syst 2024:1-14. [doi: 10.1080/08874417.2024.2401991]
58. Feijóo-García PG, Wrenn C, Stuart J, De Siqueira AG, Lok B. Participatory design of virtual humans for mental health support among North American computer science students: voice, appearance, and the similarity-attraction effect. ACM Trans Appl Percept 2023;20(3):1-27 [FREE Full text] [doi: 10.1145/3613961]
59. Google Cloud. 2024. URL: https://cloud.google.com/dialogflow [accessed 2025-12-25]
60. Young people. eSafety Commissioner. URL: https://www.esafety.gov.au/youngpeople [accessed 2025-12-25]
61. Butterfly Foundation. URL: https://butterfly.org.au/ [accessed 2025-12-25]
62. jamovi.: The jamovi project URL: https://www.jamovi.org/2022 [accessed 2025-12-25]
63. Kelter R. Analysis of type I and II error rates of Bayesian and frequentist parametric and nonparametric two-sample hypothesis tests under preliminary assessment of normality. Comput Stat 2021;36(2):1263-1288. [doi: 10.1007/s00180-020-01034-7]
64. Lee M, Wagenmakers EJ. Bayesian Cognitive Modeling: A Practical Course. Cambridge: Cambridge University Press; 2014.
65. van Doorn J, van den Bergh D, Böhm U, Dablander F, Derks K, Draws T. The JASP guidelines for conducting and reporting a Bayesian analysis. Psychon Bull Rev 2021;28(3):813-826 [FREE Full text] [doi: 10.3758/s13423-020-01798-5] [Medline: 33037582]
66. Kleinheksel AJ, Rockich-Winston N, Tawfik H, Wyatt TR. Demystifying content analysis. Am J Pharm Educ 2020;84(1):7113 [FREE Full text] [doi: 10.5688/ajpe7113] [Medline: 32292185]
67. Open Science Framework. URL: https://osf.io/f2ejg [accessed 2025-12-29]
68. JASP.: JASP Team URL: https://www.jasp-stats.org/2024 [accessed 2025-12-25]
69. National study of mental health and wellbeing. Australian Bureau of Statistics. 2022. URL: https://www.abs.gov.au/statistics/health/mental-health/national-study-mental-health-and-wellbeing/latest-release: [accessed 2025-12-25]
70. Chaves AP, Gerosa MA. How should my chatbot interact? a survey on social characteristics in human–chatbot interaction design. Int J Hum–Comput Interact 2020;37(8):729-758. [doi: 10.1080/10447318.2020.1841438]
71. Go E, Sundar S. Humanizing chatbots: the effects of visual, identity and conversational cues on humanness perceptions. Comput Hum Behav 2019;97:304-316 [FREE Full text] [doi: 10.1016/j.chb.2019.01.020]

XSL•FO
RenderX

72. Morris R, Kouddous K, Kshirsagar R, Schueller S. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. J Med Internet Res 2018;20(6):e10148 [FREE Full text] [doi: 10.2196/preprints.10148]
73. Roy R, Naidoo V. Enhancing chatbot effectiveness: the role of anthropomorphic conversational styles and time orientation. J Bus Res 2021;126:23-34 [FREE Full text] [doi: 10.1016/j.jbusres.2020.12.051]
74. Sheehan B, Jin H, Gottlieb U. Customer service chatbots: anthropomorphism and adoption. J Bus Res 2020;115:14-24 [FREE Full text] [doi: 10.1016/j.jbusres.2020.04.030]
75. Diederich S, Brendel AB, Kolbe LM. Designing anthropomorphic enterprise conversational agents. Bus Inf Syst Eng 2020;62(3):193-209. [doi: 10.1007/s12599-020-00639-y]
76. Mori M, MacDorman K, Kageki N. The uncanny valley [from the field]. IEEE Robot Autom Mag 2012;19(2):98-100. [doi: 10.1109/MRA.2012.2192811]
77. Song SW, Shin M. Uncanny valley effects on chatbot trust, purchase intention, and adoption intention in the context of e-commerce: the moderating role of avatar familiarity. Int J Hum–Comput Interact 2022;40(2):441-456. [doi: 10.1080/10447318.2022.2121038]
78. Siau K, Wang W. Building trust in artificial intelligence, machine learning, and robotics. Cutter Bus Technol J 2018;31(2):47-53 [FREE Full text]
79. Holroyd TA, Oloko OK, Salmon DA, Omer SB, Limaye RJ. Communicating recommendations in public health emergencies: the role of public health authorities. Health Secur 2020;18(1):21-28 [FREE Full text] [doi: 10.1089/hs.2019.0073] [Medline: 32078416]
80. Whetten K, Leserman J, Whetten R, Ostermann J, Thielman N, Swartz M. Exploring lack of trust in care providers and the government as a barrier to health service use. Am J Public Health 2006;96(4):716-721. [doi: 10.2105/AJPH.2005.063255] [Medline: 16507725]

Abbreviations

AI: artificial intelligence
BF: Bayes factor
CrI: credible interval
LLM: large language model
NHST: null hypothesis significance testing

Edited by C Jacob; submitted 19.Jan.2025; peer-reviewed by PG Feijóo-García, M Chatzimina; comments to author 25.Aug.2025; revised version received 15.Dec.2025; accepted 16.Dec.2025; published 12.Feb.2026.

Please cite as:
Charles M, Sauer JD, Roehrer E, Prichard J, Watters P, Scanlan J
Adolescent Perceptions of an Online Safety Chatbot: Survey Study
JMIR Hum Factors 2026;13:e71498
URL: https://humanfactors.jmir.org/2026/1/e71498
doi:10.2196/71498
PMID:

©Meriel Charles, James D Sauer, Erin Roehrer, Jeremy Prichard, Paul Watters, Joel Scanlan. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 12.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Simulated Workflow Feasibility Evaluation of a Web-Based Periorbital Measurement Platform: Development and Usability Study

George R Nahass[1,2*], BA; Jacob van der Ende[3*], MD; Sasha Hubschman[2], MD; Benjamin Beltran[1], BS; Bhavana Kolli[2], MHA; Caitlin Berek[2], BS; James D Edmonds[2], BA; RV Paul Chan[2], MD; Pete Setabutr[2], MD; James W Larrick[3,4], MD, PhD; Darvin Yi[1,2], PhD; Ann Q Tran[2], MD

[1]Department of Biomedical Engineering, University of Illinois Chicago, Chicago, IL, United States

[2]Department of Ophthalmology, University of Illinois Chicago, Chicago, IL, United States

[3]Quina Care, San Miguel, Ecuador

[4]Panorama Research Institutes, Sunnyvale, CA, United States

[*]these authors contributed equally

Corresponding Author:
Ann Q Tran, MD
Department of Ophthalmology
University of Illinois Chicago
1855 W. Taylor Street
Chicago, IL, 60612
United States
Phone: 1 (312) 996 9120
Email: annqtran@uic.edu

Abstract

Background: Periorbital measurements such as margin to reflex distances, palpebral fissure height, and scleral show are critical in diagnosing and managing conditions like ptosis and disorders of the eyelid. However, deployment of automated periorbital measurement algorithms in structured research workflows remains limited by the lack of integrated capture and data management infrastructure.

Objective: We developed and evaluated Glorbit, a lightweight, browser-based application for automated periorbital distance measurement using artificial intelligence (AI). The objective was to evaluate end-to-end workflow feasibility of the platform under simulated, operator-run conditions.

Methods: The application integrates a DeepLabV3 segmentation model into a modular image processing pipeline with secure, site-specific Google Cloud storage, supporting local preprocessing and cloud upload through Firebase-authenticated logins. The full workflow—metadata entry, facial image capture, segmentation, and upload—was tested. After the session, the participants completed a Likert-style survey.

Results: Glorbit successfully ran on all tested platforms, including laptops, tablets, and mobile phones across major browsers. A total of 15 volunteers were enrolled in this study in which the app completed predefined workflow steps in all simulated, operator-run sessions. The segmentation model produced outputs on all images, and the average session duration was 101.7 (SD 17.5) seconds. Simulated experience scores on a 5-point Likert scale were uniformly high.

Conclusions: Glorbit is a cross-platform application that supports structured periorbital image capture and automated inference within a unified workflow. In simulated, operator-run testing, the platform demonstrated successful execution of predefined workflow steps across devices. These findings support the technical feasibility of the system as a research-oriented data collection framework and may inform future evaluations in broader research settings.

(JMIR Hum Factors 2026;13:e82859) doi:10.2196/82859

KEYWORDS

artificial intelligence; global health; deployable applications; periorbital distance; simulation

XSL•FO
RenderX

Introduction

Periorbital measurements such as margin to reflex distance 1 and 2 (MRD 1/2), palpebral fissure height, and scleral show are essential components of clinical assessment in a range of conditions, including ptosis, thyroid eye disease, and congenital craniofacial conditions [1-3]. These measurements guide surgical decision-making, track disease progression, and serve as important inputs for diagnosis [4-7]. However, manual measurement of these distances incurs substantial intergrader variability, even when trained graders performed the analysis [8,9].

Previous work has demonstrated that automated periorbital measurements from facial photographs can achieve clinically acceptable accuracy by using deep learning–based segmentation approaches [10,11]. For example, a trained DeepLabV3 model achieved mean absolute errors consistently below established intergrader variability thresholds for key measurements, including MRD1, MRD2, and intercanthal distance, with 86% of the measurements falling within intergrader thresholds. This approach also outperformed the benchmark method PeriorbitAI, successfully processing 100% of the images across diverse disease populations, including thyroid eye disease and craniofacial conditions, compared to 59%-85% success rates for PeriorbitaAI [10,12]. However, due to the lack of deployable infrastructure capable of integrating image capture, processing, and secure data management, both of these algorithms (and others) cannot be integrated into routine data collection workflows for research use.

In response to this gap, we developed Glorbit, a web-based application for artificial intelligence (AI)-based periorbital distance measurement and metadata capture. Rather than introducing a new periorbital measurement algorithm, the primary contribution of Glorbit lies in system integration or in the unification of image capture, local preprocessing, AI inference, metadata collection, and secure, site-specific data storage within a browser-based interface. This contribution is orthogonal to algorithm development, and as such, we focus exclusively on assessing workflow operability of the application rather than measurement accuracy, reliability, or readiness for clinical use. In this paper, we describe the system architecture and image processing pipeline and present a small, simulated feasibility study assessing end-to-end workflow execution and participant-perceived enrollment experience under controlled conditions.

Methods

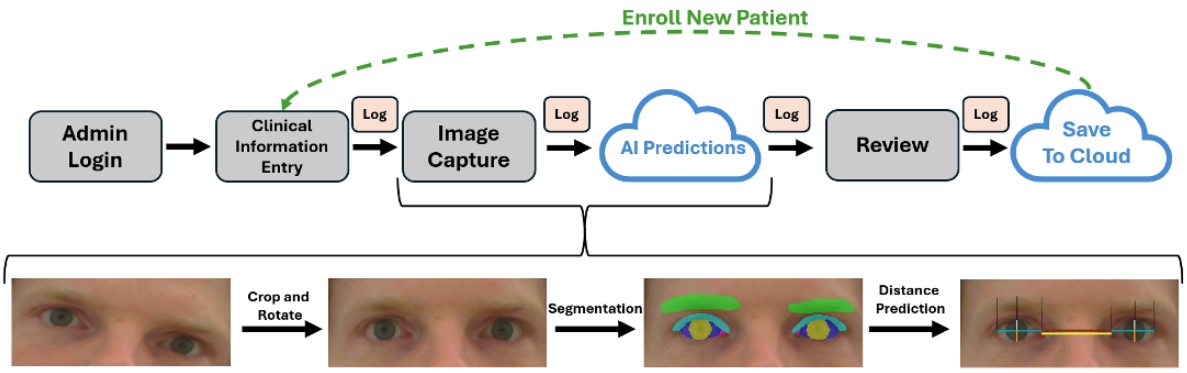
Ethical Considerations

The simulated enrollment study was designed to evaluate technical operability and workflow execution rather than performance under real world clinical constraints. No direct identifiers (eg, name, date of birth, medical record number) were collected, and each participant was assigned a deidentified patient ID. Verbal consent was obtained in accordance with an institutional review board–exempt protocol (STUDY2025-0731). All captured images, derived measurements, and metadata were deleted following data analysis. No participant images or metadata were retained in persistent cloud buckets. Participants were not compensated.

System Overview and User Flow

Glorbit is a browser-based application designed to support structured periorbital image capture and automated distance inference for research use. Upon login, authenticated users are presented with a structured form to collect brief patient history and visit metadata. This information is stored temporarily in session state within the browser. After submission, users proceed to an image capture page, where a webcam or device camera is used to capture a frontal facial image. The image is processed locally to generate segmentation and AI-derived distance measurements, which are then displayed for user review. Once confirmed, the image and associated metadata are uploaded to a secure, site-specific cloud storage location. Logging is integrated throughout the pipeline to capture system events, errors, and upload status. See Figure 1 for an overview of users' steps throughout the application.

Figure 1. Glorbit application workflow and artificial intelligence (AI) processing pipeline. The top row illustrates the user-facing app interface: a user logs in, enters clinical metadata, captures an image, reviews AI-predicted periorbital distances, and uploads results to cloud storage. Cloud-shaped nodes denote steps involving remote model inference or cloud-based data storage. All user actions, including reverse navigation, are logged. The bottom row depicts the backend image processing steps: cropping and rotation based on facial landmarks, semantic segmentation of key anatomical regions, and automated distance prediction using landmark geometry.



XSL•FO
RenderX

Image Processing Pipeline

Each captured image undergoes a predefined processing pipeline for anatomical alignment, cropping, and segmentation, as described by Nahass et al [10]. MediaPipe FaceMesh landmarks are used to rotate and align the image to a horizontal eye-level axis [13]. A region of interest is then cropped around the eyes and periorbital region. The aligned image is passed to a trained DeepLabV3 segmentation model for periorbital anatomic segmentation, followed by the geometric calculation of periorbital distances from the segmented output (Figure 1) [14]. In the event of a model failure, the app handles the error, logs the failure, and prompts the user to try again. The periorbital distance prediction model integrated into Glorbit was selected from prior work and was used without modification; evaluation of measurement accuracy, repeatability, and robustness was outside the scope of this study, which focused on platform feasibility and workflow execution.

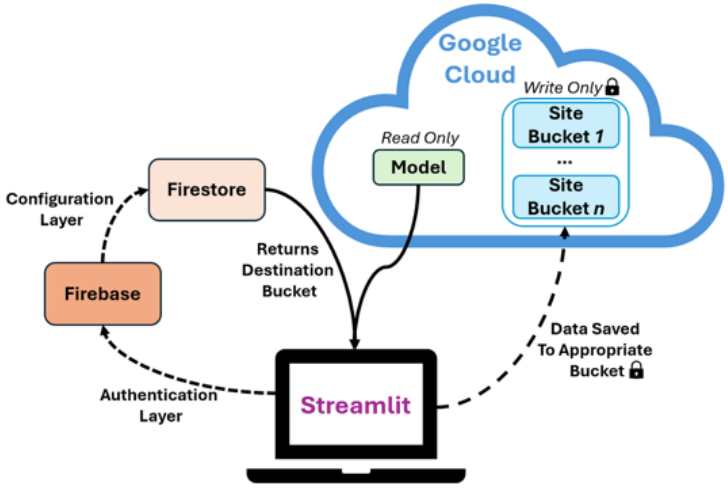
Clinical Data

Glorbit captures a structured set of clinical metadata alongside each image to support contextual analysis and downstream modeling. At minimum, users are required to input a deidentified patient ID, age, and sex to proceed with the workflow. Optional fields include ethnicity (selected from a site-customizable dropdown list), relevant comorbidities (eg, thyroid eye disease, craniofacial conditions, prior eyelid surgery), visit type (new or follow-up), and image conditions (eg, use of tape, lighting notes). Sites may also enable the entry of laboratory values when available. Supported laboratory measurements include levels of alanine aminotransferase, aspartate aminotransferase, hemoglobin A_{1c}, estimated glomerular filtration rate, thyroid-stimulating hormone, thyroid-stimulating immunoglobulin, albumin-to-creatinine ratio, and calcium, with standardized units displayed next to each input. All metadata are stored alongside the captured image and the AI-derived periorbital distances in the designated site-specific cloud storage bucket. Only the required fields must be completed to advance to image capture; all other fields are optional and may be tailored to institutional needs.

Data Handling and Storage Architecture

The app frontend is built in Streamlit and deployed as a Docker container to ensure reproducibility. Firebase Authentication manages secure login, with role-based access control and per-site credentialing. Firestore is used to assign each user's upload path based on their authenticated credentials, enabling dynamic configuration of storage destinations. This allows a single instance of the app to be deployed across multiple locations without requiring hardcoded site-specific settings or local reconfiguration. Glorbit stores the following data elements: (1) cropped facial images, (2) AI-predicted periorbital measurements, (3) session-level metadata entered by the operator, and (4) system logs generated for debugging and audit. All data are encrypted at rest and in transit, and access rules are configured through Google Cloud Identity and Access Management policies. Glorbit is designed to operate on consumer-grade hardware and can be accessed on laptops, tablets, and mobile phones with integrated webcams. A graphical schematic of data movement from the operator's perspective can be found in Figure 2.

Figure 2. System architecture for secure authentication, configuration, and cloud-based storage within the Glorbit platform. Users authenticate via Firebase, after which Firestore provides site-specific configuration, including the destination storage bucket. The local Streamlit app reads artificial intelligence model weights from a centralized, read-only location in Google Cloud and writes collected data to site-specific, write-only buckets. All data transfers are protected using encryption at rest and in transit, as indicated by lock icons. Access control is governed via Google Cloud's Identity and Access Management framework.



Security Design

The primary data elements with potential status as Protected Health Information within the Glorbit workflow are cropped images of the periorbital region. While the removal of the full face reduces identifiability, we treat these cropped images as sensitive biometric data. To protect this information, all uploads utilize technical safeguards, including encryption in transit (TLS) and at rest (AES-256) via Google Cloud's native encryption [15-17]. Each deployment site is assigned a unique Google Cloud Storage bucket with Identity and Access Management roles that can be configured to restrict access by

XSL•FO
RenderX

site. Firebase Authentication enforces user-specific logins, and Firestore assigns site-level write access dynamically. Full event and error logging is implemented for auditability. The platform architecture incorporates technical safeguards commonly required for Health Insurance Portability and Accountability Act (HIPAA)-compliant workflows; however, compliance is conditional upon the specific deployment environment. Specifically, when deployed within an institutional environment governed by a Business Associate Agreement with the cloud provider, the system provides the necessary infrastructure to maintain a HIPAA-compliant data pipeline when appropriately configured [18,19]. In the absence of a Business Associate Agreement between an institution deploying Glorbit and the cloud provider, Glorbit is not HIPAA-compliant.

Internal Testing: Cross-Platform Compatibility

Glorbit was tested by a member of the study team (GRN) across multiple operating systems (Windows, iOS, ChromeOS), browsers (Chrome, Firefox, Safari), and device types (laptop, personal computer, smartphone, tablet) to assess basic cross-platform compatibility. Testing focused on verifying that core workflow components, including authentication, metadata entry, camera access, image capture, and execution of the image processing pipeline, functioned consistently across platforms. A workflow failure was defined as the inability to successfully complete any predefined step, including authentication errors, camera access failure, image capture failure, segmentation, or inference errors, resulting in missing outputs, upload errors, or logging failures. Failures were identified through user-facing error messages hardcoded into the pipeline, as well as through a review of system logs confirming completion of each stage. In addition, controlled negative-input scenarios such as the absence of a detectable face in the camera frame were manually tested to confirm appropriate system behavior, including user-facing feedback and event logging. These checks were performed to validate control flow and error handling rather than to quantify measurement accuracy. This testing was performed under natural indoor lighting conditions with 5 trial runs on each combination.

External Testing: Survey and Simulated Enrollment Testing

To evaluate participant-perceived enrollment experience, we conducted a simulated enrollment study with 15 adult volunteers (age ≥18 years) at the Illinois Eye and Ear Infirmary. Each session was conducted by a trained study operator (GRN) using a 2021 MacBook Air in a private room under natural indoor lighting conditions. The operator followed the standard Glorbit workflow, including metadata entry and facial image capture. Failures of the workflow were assessed using the same protocol as described earlier.

Following the simulated interaction, participants completed a brief anonymous survey assessing their experience immediately following the interaction. The survey included 5 Likert-scale items (1-5) evaluating perceived intuitiveness, perceived efficiency, clarity of workflow steps, self-reported confidence in the displayed outputs, and perceived potential usefulness in a clinical context. All responses were stored without any identifiers. This survey was intended to capture participant-perceived workflow clarity and general impressions following an operator-run simulated interaction; it was not designed to assess operator usability, measurement validity, or clinical readiness. Any reported confidence in the displayed outputs following the simulated interaction reflect participant perception only and does not represent validation of measurement accuracy.

Results

App Walkthrough and Cross-Platform Compatibility

Screenshots of the core workflow of Glorbit are shown in Figure 3, illustrating the minimalistic, stepwise interface. To assess platform execution, Glorbit was internally tested across multiple browsers, operating systems, and device types. Across all tested platforms, core workflow components, including authentication, form entry, camera access, image capture, execution of the image processing pipeline, and system logging, were executed without observed workflow interruption across tested configurations.

Table 1 shows the evaluated platforms and devices. In addition, controlled negative input scenarios were manually evaluated, including cases in which no face was detectable within the camera frame. In these scenarios, the system generated appropriate user-facing feedback and recorded the event in the application logs.

XSL•FO
RenderX

Figure 3. Screenshot of the Glorbit app interface. (A) The form used to input patient metadata, including demographics and comorbidities. (B) Face alignment guide with a live camera view for guided image capture. These two steps represent the only direct patient interactions aside from the final review and submission step.

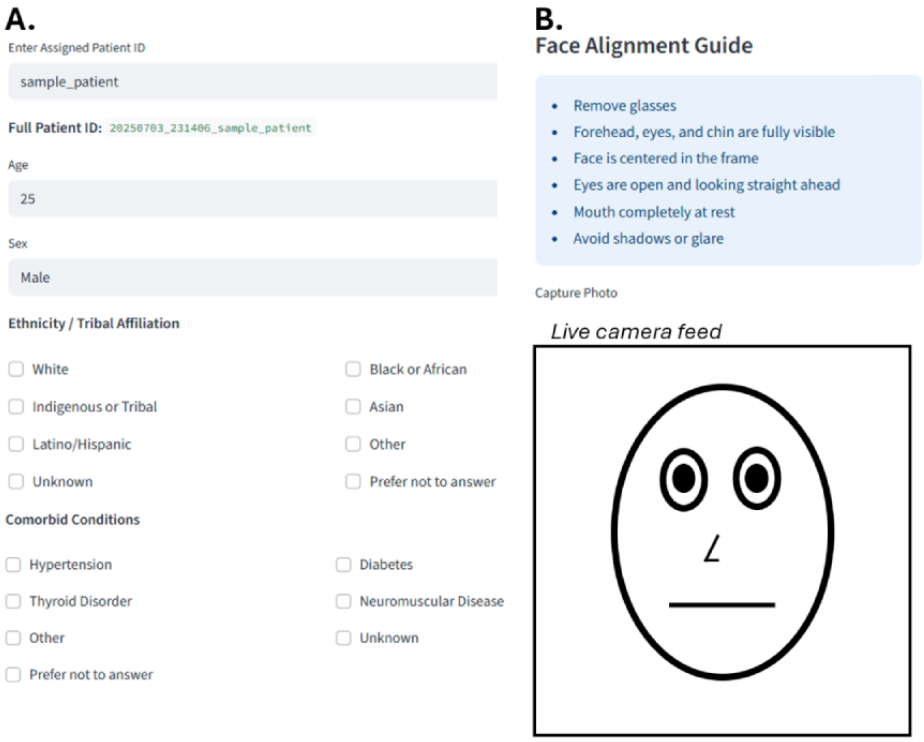


Table 1. Summary of the operating systems, device types, and camera sources that the workflow feasibility of Glorbit was evaluated on. Five execution runs were completed for each configuration.

	Operating system	Device type	Browser	Camera source
1	Windows	Desktop PC	Chrome	USB webcam
2	Windows	Desktop PC	Firefox	USB webcam
3	macOS	MacBook	Chrome	Integrated webcam
4	macOS	MacBook	Firefox	Integrated webcam
5	macOS	MacBook	Safari	Integrated webcam
6	ChromeOS	Chromebook	Chrome	Integrated webcam
7	iOS	Smartphone	Chrome	Front camera
8	iOS	iPad	Chrome	Front camera

Simulated Enrollment Performance

We enrolled 15 adult volunteers in a simulated enrollment study to assess the workflow operability of Glorbit under simulated operating conditions. Participant demographics are summarized in Table 2. The group's mean age was 32.6 (SD 11.0) years, was 67% (10/15) female, and spanned 5 self-identified ethnic groups.

In all 15 cases (100%), the app completed the predefined workflow steps (metadata entry, image capture, segmentation, and upload) in all the simulated sessions conducted by a trained operator. The segmentation model generated outputs in all cases, and no image processing or upload failures were observed under the tested conditions. The average session duration, measured from initial form entry to final upload, was 101.7 (SD 17.5) seconds on consumer-grade hardware (Figure S1 in Multimedia Appendix 1). Logging was successful in all cases, with system events and session metadata stored as expected.

XSL•FO
RenderX

Table 2. Demographic characteristics of the simulated study participants (N=15).

Characteristics	Values
Age (y)	
Mean (SD)	32.6 (11.0)
Maximum	58
Minimum	24
Median	26
Sex, n (%)	
Male	5 (33)
Female	10 (67)
Ethnicity, n (%)	
White	2 (13)
Latino/Hispanic	4 (27)
Black or African	3 (20)
Asian	5 (33)
Other	1 (7)

Participant Perception of Glorbit

Following each session, participants completed a brief anonymous postusage survey to assess their perceived experience during the simulated interaction. These responses reflect participant perceptions following an operator-run simulated interaction and do not represent an evaluation of clinician usability, workflow performance in routine practice, or validation of measurement accuracy. Average responses were high across all 5 items (Table 3).

Participants rated the app as intuitive and efficient (both 5.0, SD 0.0), with high self-reported confidence in the displayed outputs (4.9, SD 0.3) and a high perceived comfort in a hypothetical clinical context (4.9, SD 0.3). Participants also reported clear understanding of each step in the data collection workflow (4.8, SD 0.4). Across all participant responses to all questions, 93.3% (70/75) of the responses were 5, 6.7% (5/75) of the responses were 4, and no 3 or lower were observed. The implications of this ceiling effect are discussed in the Limitations section.

Table 3. Summary of participant feedback from postsession surveys in the simulated enrollment study. Each participant (N=15) rated 5 statements on a 5-point Likert scale (1=strongly disagree, 5=strongly agree).

Statement	Average rating, mean (SD)
The Glorbit app was intuitive and easy to use.	5.0 (0.0)
The process of capturing and uploading patient data was fast and efficient.	5.0 (0.0)
I would feel comfortable using this tool in a clinical setting.	4.9 (0.3)
I understood each step of the data collection process.	4.8 (0.4)
I have confidence in the displayed outputs.	4.9 (0.3)

Discussion

Principal Findings

Many periorbital distance prediction algorithms have been developed in recent years, but their utility as broader research tools remains limited by the absence of deployable infrastructure [9,12,20-22]. In such contexts, there is a need for lightweight and secure systems capable of image capture, processing, and storage that can be employed within structured data collection workflows. To meet this need, we developed an internet browser–based app, Glorbit. Glorbit does not introduce a new periorbital measurement algorithm; instead, its primary contribution lies in the integration of existing periorbital measurement models into a unified system that supports image capture, inference, metadata collection, and secure storage. Periorbital distance measurement is a property of the underlying AI model rather than the deployment platform itself, and therefore, evaluation of this was outside the scope of this study, which was designed to assess the workflow operability of Glorbit.

In a simulated study of 15 participants, Glorbit completed predefined workflow steps (metadata entry, image capture, segmentation, and upload) in all simulated sessions conducted under operator-run conditions. These findings reflect successful technical execution of the workflow under simulated conditions and should not be interpreted as evidence of field readiness or validated clinical usability. The uniformly high survey ratings reflect participant-perceived workflow clarity following an

XSL•FO
RenderX

operator-run simulated interaction and do not represent an evaluation of clinician usability, workflow performance in routine practice, or clinical utility. Glorbit provides error messages and logs all user movement in the event of a failure (which were not observed in the simulated user study). This error handling is intended to provide administrators with detailed diagnostic information for troubleshooting workflow interruptions during future testing or deployment.

Additionally, the complete end-to-end workflow duration through Glorbit was short, with a mean session time of 101.7 (SD 17.5) seconds, inclusive of clinical metadata entry and user review of AI generated outputs. For contextual reference, prior studies have reported longer durations for traditional periorbital measurement workflows, including assisted measurement using ImageJ toolkits (327, SD 116 s) and purely manual measurement (804, SD 204 s) [23]. These previously reported values were obtained under different study designs and are provided here for qualitative context rather than as a direct efficiency comparison. Additionally, these timing results were obtained during simulated, operator-run enrollments and may not generalize to routine clinical or real-world research settings.

To date, several mobile apps have been developed for facial measurement. However, these tools primarily focus on interpupillary distance and do not include integrated metadata capture or cloud upload functionality for research workflows [24,25]. Glorbit is designed to include a combination of site-specific secure cloud storage and a modular architecture that can be deployed across institutions without requiring local IT infrastructure. Table 4 summarizes selected functional characteristics of Glorbit and other facial measurement tools, emphasizing deployment features rather than measurement accuracy or clinical validation.

Table 4. Comparison of Glorbit to existing facial measurement platforms.

Platform	Cloud storage[a]	Measurement type[b]	Open source	Primary function
Glorbit	Yes	MRD[c]1, MRD2, PFH[d] (48 distances)	Yes	Integrates inference, metadata forms, and storage
MediaPipe [13]	No	Facial landmarks	Yes	Google tool for face mesh detection
OpenFace [26,27]	No	Facial landmarks	Yes	Academic tool for facial behavior analysis, can extract 68 landmarks

[a]Cloud storage reflects availability of integrated cloud upload functionality.

[b]Measurement type describes the anatomical features or distances quantified by each system.

[c]MRD: margin to reflex distance.

[d]PFH: palpebral fissure height.

In addition to serving as a storage repository for providers, the captured data could potentially be used as future training data to create patient-specific models. One of the most persistent barriers to equitable AI is poor model generalization to underrepresented populations, often due to data captured under different conditions being out of distribution to the original training set [28-32]. Glorbit is designed to facilitate ethically governed data collection across diverse sites, which may support future efforts to improve model generalization. If integrated into research workflows, structured image and metadata capture could support the development of more representative, generalizable models through real-world data acquisition and iteration. Further strengthening the motivation for large-scale collection of high quality data is the developing field of Oculomics, which aims to leverage deep learning to make predictions about systemic health from images of the eyes [33-37]. While most current Oculomics research focuses on retinal imaging, recent studies have shown that external eye photographs can also be used to predict systemic biomarkers such as alanine aminotransferase, aspartate aminotransferase, and hemoglobin A_{1c} [38-40]. With the increasing number of foundational ophthalmic deep learning systems, having specific multimodal datasets available for finetuning may support future development and evaluation of downstream predictive models.

An important consideration for potential future deployment of tools like Glorbit in clinical settings is the protection of patient privacy and compliance with institutional and national data governance requirements. While the platform utilizes cropped periorbital images to minimize identifiability, these data elements are treated as sensitive biometric information requiring protection. Consequently, deployment within a regulated clinical environment would require the establishment of a Business Associate Agreement between the deploying institution and the cloud service provider to ensure HIPAA security requirements are met. Glorbit's architecture is designed to incorporate technical safeguards (access control, end-to-end encryption, and audit logging) commonly required for HIPAA-compliant workflows when governed by such an agreement [18,19]. However, institutional oversight remains a necessary component of medical AI deployment, as the software's compliance status is conditional upon the host site's administrative and security configurations. To promote transparency and allow for local adaptation and compliance with patient protection laws on a per nation basis, Glorbit is fully open-source.

Limitations

This study represents an initial feasibility evaluation and should be interpreted in that context. The simulated enrollment involved a modest number of participants (N=15) and was designed to assess workflow operability, system stability, and platform execution rather than to formally evaluate usability using standardized instruments or to assess performance across diverse clinical users and settings. As such, measurement validity, repeatability, and robustness of the underlying periorbital distance prediction models were not assessed here. Although

XSL•FO
RenderX

the simulated enrollment cohort included participants from multiple self-identified ethnic groups, the small sample size precluded meaningful subgroup analysis as well as drawing conclusions regarding reliability in real-world settings. Taken together, this study does not assess demographic-specific measurement accuracy, robustness across lighting conditions, or device-related biases—all important factors in computer vision applications. These factors represent known risks for facial analysis systems and will require dedicated, adequately powered evaluations in future studies.

In addition, sessions were conducted by a trained operator on a convenience sample, and participant feedback reflects perceived clarity and comfort following observed system interaction. As such, these ratings should be interpreted as indicators of basic workflow acceptability rather than sensitive or discriminative measures of usability performance and may not generalize to a larger population. Additionally, survey responses demonstrated strong ceiling effects, likely reflecting the controlled, operator-run nature of the simulated enrollment sessions. Future assessments of user experience (as opposed to participants) should incorporate standardized usability instruments and comparative study designs.

Finally, operator-facing usability and workflow burden were not formally evaluated in this study and would require evaluation in clinician-operated studies. Similarly, while the system architecture was engineered to permit offline use with local data storage, these capabilities were not evaluated here. Future work should include larger, clinician-operated evaluations across multiple sites to further assess usability and performance under routine clinical conditions.

Acknowledgments

The study team would like to gratefully acknowledge Alexander Schönjahn from Quina Care. We also thank our funders.

Funding

This study was supported by an unrestricted grant from Research to Prevent Blindness, a donation from the Cless Family Foundation, and the National Institutes of Health P30 EY001792 core grant. The funding organizations had no role in the design or conduct of the study; collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Authors' Contributions

GRN conceived the study, designed the system architecture, implemented the application, conducted testing, analyzed data, and drafted the manuscript. JvdE contributed to study design, manuscript writing and editing, and served as the primary global health liaison for deployment planning. BB assisted with coding and technical implementation. SH and JDE supported user interface development, usability testing, and manuscript editing. BK and CB facilitated ethical compliance and institutional review board processes. RVPC, PS, JWL, DY, and AQT provided critical oversight, domain expertise, and manuscript review. All authors approved the final version of the manuscript.

Conflicts of Interest

SH reports stock or stock options in Horizon Surgical Systems. PS reports consultancy for Oyster Point Pharma and Viatris and is a board member of the American Society of Ophthalmic Plastic and Reconstructive Surgery; he also reports stock or stock options in Lodestone Pharmaceuticals. AQT reports grants from the Illinois Society for the Prevention of Blindness and consultancy for Genentech/Roche. No other disclosures are reported.

Multimedia Appendix 1
Distribution of end-to-end session duration. Histogram showing the distribution of total session duration for simulated Glorbit enrollment sessions (N=15), measured from initial metadata entry through image capture, artificial intelligence (AI) processing, and final review.
[PNG File , 37 KB - humanfactors_v13i1e82859_app1.png]

References

1. Conrady C, Patel B. Crouzon Syndrome. StatPearls Treasure Island (FL): StatPearls Publishing; May 25, 2025.
2. Koka K, Patel B. Ptosis Correction. StatPearls Treasure Island (FL): StatPearls Publishing; Jul 10, 2023.
3. MacLachlan C, Howland HC. Normal values and standard deviations for pupil diameter and interpupillary distance in subjects aged 1 month to 19 years. Ophthalmic Physiol Opt 2002 May;22(3):175-182. [doi: 10.1046/j.1475-1313.2002.00023.x] [Medline: 12090630]
4. Nemet AY. Accuracy of marginal reflex distance measurements in eyelid surgery. J Craniofac Surg 2015 Oct;26(7):e569-e571. [doi: 10.1097/SCS.0000000000001304] [Medline: 26468822]

XSL•FO
RenderX

5. Cruz AA, Coelho RP, Baccega A, Lucchezi MC, Souza AD, Ruiz EE. Digital image processing measurement of the upper eyelid contour in Graves disease and congenital blepharoptosis. Ophthalmology 1998 May;105(5):913-918. [doi: 10.1016/S0161-6420(98)95037-0] [Medline: 9593397]
6. Cruz AA, Lucchezi MC. Quantification of palpebral fissure shape in severe congenital blepharoptosis. Ophthalmic Plast Reconstr Surg 1999 Jul;15(4):232-235. [doi: 10.1097/00002341-199907000-00002] [Medline: 10432517]
7. Bodnar ZM, Neimkin M, Holds JB. Automated ptosis measurements from facial photographs. JAMA Ophthalmol 2016 Feb;134(2):146-150. [doi: 10.1001/jamaophthalmol.2015.4614] [Medline: 26605967]
8. Boboridis K, Assi A, Indar A, Bunce C, Tyers AG. Repeatability and reproducibility of upper eyelid measurements. Br J Ophthalmol 2001 Jan;85(1):99-101 [FREE Full text] [doi: 10.1136/bjo.85.1.99] [Medline: 11133723]
9. Hussey VM, Tao JP. Oculofacial plastic surgeon distribution by county in the United States, 2021. Orbit 2022 Dec;41(6):687-690. [doi: 10.1080/01676830.2021.1989468] [Medline: 34672850]
10. Nahass GR, Koehler E, Tomaras N, Lopez D, Cheung M, Palacios A, et al. State-of-the-art periorbital distance prediction and disease classification using periorbital features. ArXiv. Preprint posted online on May 14, 2025 2025. [doi: 10.48550/arXiv.2409.18769]
11. Nahass GR, Koehler E, Tomaras N, Lopez D, Cheung M, Palacios A, et al. Open-source periorbital segmentation dataset for ophthalmic applications. Ophthalmol Sci 2025;5(4):100757 [FREE Full text] [doi: 10.1016/j.xops.2025.100757] [Medline: 40933660]
12. Van Brummen A, Owen JP, Spaide T, Froines C, Lu R, Lacy M, et al. PeriorbitAI: artificial intelligence automation of eyelid and periorbital measurements. Am J Ophthalmol 2021 Oct;230:285-296 [FREE Full text] [doi: 10.1016/j.ajo.2021.05.007] [Medline: 34010596]
13. Kartynnik Y, Ablavatski A, Grischenko I, Grundmann M. Real-time facial surface geometry from monocular video on mobile GPUs. ArXiv. Preprint posted online on July 15, 2019 2019. [doi: 10.48550/arXiv.1907.06724]
14. Chen LC, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. ArXiv. Preprint posted online on December 5, 2017 2017. [doi: 10.48550/arXiv.1706.05587]
15. Google security overview. Google Cloud. URL: https://docs.cloud.google.com/docs/security/overview/whitepaper [accessed 2026-01-11]
16. Default encryption at rest. Google Cloud. URL: https://docs.cloud.google.com/docs/security/encryption/default-encryption [accessed 2026-01-11]
17. Encryption in transit for Google cloud. Google Cloud. URL: https://docs.cloud.google.com/docs/security/encryption-in-transit [accessed 2026-01-11]
18. Business associates. US Department of of Health and Human Services. URL: https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/business-associates/index.html [accessed 2026-01-11]
19. HIPAA compliance on Google cloud. Google Cloud. URL: https://cloud.google.com/security/compliance/hipaa [accessed 2026-01-11]
20. Rana K, Beecher M, Caltabiano C, Macri C, Zhao Y, Verjans J, et al. Artificial intelligence to automate assessment of ocular and periocular measurements. Eur J Ophthalmol 2025 Jan;35(1):346-351 [FREE Full text] [doi: 10.1177/11206721241249773] [Medline: 38710195]
21. Chen H, Tzeng S, Hsiao Y, Chen R, Hung E, Lee OK. Smartphone-based artificial intelligence-assisted prediction for eyelid measurements: algorithm development and observational validation study. JMIR Mhealth Uhealth 2021 Oct 08;9(10):e32444 [FREE Full text] [doi: 10.2196/32444] [Medline: 34538776]
22. Guo Y, Liu J, Ruan Y, Rokohl A, Hou X, Li S, et al. A novel approach quantifying the periorbital morphology: a comparison of direct, 2-dimensional, and 3-dimensional technologies. J Plast Reconstr Aesthet Surg 2021 Aug;74(8):1888-1899. [doi: 10.1016/j.bjps.2020.12.003] [Medline: 33358464]
23. Peterson JC, Nahass GR, Lasalle C, Bradley DC, Wu D, Zorra I, et al. Development and validation of a semiautomated tool for measuring periorbital distances. Ophthalmol Sci 2025;5(6):100887 [FREE Full text] [doi: 10.1016/j.xops.2025.100887] [Medline: 40917265]
24. Han KD, Jaafar M, Stoakes IM, Hoopes PC, Moshirfar M. Comparing the effectiveness of smartphone applications in the measurement of interpupillary distance. Cureus 2023 Jul;15(7):e42744 [FREE Full text] [doi: 10.7759/cureus.42744] [Medline: 37529827]
25. Singman E, Matta N, Tian J, Silbert D. The accuracy of the plusoptiX for measuring pupillary distance. Strabismus 2014 Mar;22(1):21-25. [doi: 10.3109/09273972.2013.877941] [Medline: 24564725]
26. Baltrusaitis T, Robinson P, Morency L. OpenFace: an open source facial behavior analysis toolkit. 2016 Presented at: IEEE Winter Conference on Applications of Computer Vision (WACV); March 7-10; Lake Placid, NY. [doi: 10.1109/wacv.2016.7477553]
27. Hu J, Mathur L, Liang P, Morency LP. OpenFace 3.0: a lightweight multitask system for comprehensive facial behavior analysis. ArXiv. Preprint posted online on June 3, 2025 2025. [doi: 10.48550/arXiv.2506.02891]
28. Hong Z, Yue Y, Chen Y, Cole L, Lin H, Luo Y, et al. Out-of-distribution detection in medical image analysis: a survey. ArXiv. Preprint posted online on July 3, 2024 2024. [doi: 10.48550/arXiv.2404.18279]

XSL•FO
RenderX

29. Rashidisabet H, Sethi A, Jindarak P, Edmonds J, Chan RVP, Leiderman YI, et al. Validating the generalizability of ophthalmic artificial intelligence models on real-world clinical data. Transl Vis Sci Technol 2023 Nov 01;12(11):8 [FREE Full text] [doi: 10.1167/tvst.12.11.8] [Medline: 37922149]
30. Mårtensson G, Ferreira D, Granberg T, Cavallin L, Oppedal K, Padovani A, et al. The reliability of a deep learning model in clinical out-of-distribution MRI data: a multicohort study. Med Image Anal 2020 Dec;66:101714 [FREE Full text] [doi: 10.1016/j.media.2020.101714] [Medline: 33007638]
31. Ozkan E, Boix X. Multi-domain improves classification in out-of-distribution and data-limited scenarios for medical image analysis. Sci Rep 2024 Oct 18;14(1):24412 [FREE Full text] [doi: 10.1038/s41598-024-73561-y] [Medline: 39420026]
32. Karimi D, Gholipour A. Improving calibration and out-of-distribution detection in deep models for medical image segmentation. IEEE Trans Artif Intell 2023 Apr;4(2):383-397 [FREE Full text] [doi: 10.1109/tai.2022.3159510] [Medline: 37868336]
33. Zhu Z, Wang Y, Qi Z, Hu W, Zhang X, Wagner SK, et al. Oculomics: current concepts and evidence. Prog Retin Eye Res 2025 May;106:101350 [FREE Full text] [doi: 10.1016/j.preteyeres.2025.101350] [Medline: 40049544]
34. Honavar SG. Oculomics - the eyes talk a great deal. Indian J Ophthalmol 2022 Mar;70(3):713 [FREE Full text] [doi: 10.4103/ijo.IJO_474_22] [Medline: 35225499]
35. Patterson EJ, Bounds AD, Wagner SK, Kadri-Langford R, Taylor R, Daly D. Oculomics: a crusade against the four horsemen of chronic disease. Ophthalmol Ther 2024 Jun;13(6):1427-1451 [FREE Full text] [doi: 10.1007/s40123-024-00942-x] [Medline: 38630354]
36. Suh A, Hampel G, Vinjamuri A, Ong J, Kamran SA, Waisberg E, et al. Oculomics analysis in multiple sclerosis: current ophthalmic clinical and imaging biomarkers. Eye (Lond) 2024 Oct;38(14):2701-2710. [doi: 10.1038/s41433-024-03132-y] [Medline: 38858520]
37. Zhou Y, Chia MA, Wagner SK, Ayhan MS, Williamson DJ, Struyven RR, et al. A foundation model for generalizable disease detection from retinal images. Nature 2023 Oct;622(7981):156-163 [FREE Full text] [doi: 10.1038/s41586-023-06555-x] [Medline: 37704728]
38. Babenko B, Traynis I, Chen C, Singh P, Uddin A, Cuadros J, et al. A deep learning model for novel systemic biomarkers in photographs of the external eye: a retrospective study. Lancet Digit Health 2023 May;5(5):e257-e264 [FREE Full text] [doi: 10.1016/S2589-7500(23)00022-5] [Medline: 36966118]
39. DeBuc DC. AI for identification of systemic biomarkers from external eye photos: a promising field in the oculomics revolution. Lancet Digit Health 2023 May;5(5):e249-e250 [FREE Full text] [doi: 10.1016/S2589-7500(23)00047-X] [Medline: 36966119]
40. Babenko B, Mitani A, Traynis I, Kitade N, Singh P, Maa AY, et al. Detection of signs of disease in external photographs of the eyes via deep learning. Nat Biomed Eng 2022 Dec;6(12):1370-1383 [FREE Full text] [doi: 10.1038/s41551-022-00867-5] [Medline: 35352000]

Abbreviations

AI: artificial intelligence
HIPAA: Health Insurance Portability and Accountability Act
MRD 1/2: margin reflex distance 1 and 2

Edited by A Kushniruk; submitted 22.Aug.2025; peer-reviewed by V Pulipati, L Seim; comments to author 02.Jan.2026; revised version received 16.Feb.2026; accepted 21.Feb.2026; published 17.Apr.2026.

Please cite as:
Nahass GR, Ende JVD, Hubschman S, Beltran B, Kolli B, Berek C, Edmonds JD, Chan RVP, Setabutr P, Larrick JW, Yi D, Tran AQ Simulated Workflow Feasibility Evaluation of a Web-Based Periorbital Measurement Platform: Development and Usability Study JMIR Hum Factors 2026;13:e82859
URL: https://humanfactors.jmir.org/2026/1/e82859
doi:10.2196/82859

©George R Nahass, Jacob van der Ende, Sasha Hubschman, Benjamin Beltran, Bhavana Kolli, Caitlin Berek, James D Edmonds, RV Paul Chan, Pete Setabutr, James W Larrick, Darvin Yi, Ann Q Tran. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 17.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Perceptions of Health Care Professionals on the Integration and Use of AI in Clinical Cancer Care: Interview Study

Akhona C Khumalo, MSc; Ala Sarah Alaqra, PhD

Information Systems, Karlstad Business School, Karlstad University, Universitetsgatan 2, Karlstad, Sweden

Corresponding Author:
Akhona C Khumalo, MSc
Information Systems, Karlstad Business School, Karlstad University, Universitetsgatan 2, Karlstad, Sweden

Abstract

Background: Artificial intelligence (AI) is increasingly recognized for its potential to transform cancer care. However, much of the existing evidence of its efficacy comes from controlled settings. There remains a need to complement this knowledge with insights into how AI tools are perceived and used in real-world clinical settings, as well as how their use impacts clinical practice.

Objective: This study aimed to explore key factors influencing clinicians' acceptance of AI tools and examine how AI adoption and use impact clinical workflows in cancer care.

Methods: We used purposive sampling for recruiting oncology-related health care professionals and collected data using web-based semistructured interviews to gather their perceptions. Data were thematically analyzed and interpreted through the lenses of sociotechnical systems theory and the Unified Theory of Acceptance and Use of Technology.

Results: Participants largely accept and perceive AI tools as beneficial to clinical practice. Unified Theory of Acceptance and Use of Technology constructs were reflected in our data as determinants of intention to adopt AI tools. Trust appears as an influential factor in shaping attitudes toward AI tools. Acceptance is found to both precede AI tool use and to grow following successful integration. The use of AI tools is perceived to yield operational benefits, such as reduced workload and time savings, and clinical benefits, such as increased diagnostic reliability and reduced patient recall. Minimal disruption to clinical workflows following integration of AI tools was reported for some cancer screening applications and organ-at-risk segmentation, whereas greater disruption was anticipated for 3D cancer screening. Although accountability and lack of explainability are highlighted in literature as barriers to AI adoption, participants do not view these as significant obstacles in image-based diagnostic contexts. Additionally, negative impacts, such as overreliance on AI and reduced critical review of AI results, arise in association with the use of AI tools.

Conclusions: Participants perceive AI tools to deliver benefits to clinical cancer care. However, their adoption relies on their alignment with clinical needs and seamless integration into clinical workflows. To encourage clinician acceptance, the identified concerns must be addressed. Future work should focus on training programs, co-design with clinicians, and exploration of mitigation strategies for emerging adverse effects, such as automation bias and potential skill erosion.

(JMIR Hum Factors 2026;13:e83240) doi:10.2196/83240

KEYWORDS

artificial intelligence; cancer care; oncology; clinical workflows; health care professionals; human factors; sociotechnical systems theory; technology acceptance; UTAUT; skill erosion

Introduction

Background

Amid the staggering number of cancer incidences worldwide, artificial intelligence (AI) promises a better future, improving efficacy through faster and more accurate health care decisions [1,2]. This promise can make a huge difference in oncology, where the fight against cancer places a high demand on health care resources, while also being challenged by a workforce shortage. AI is increasingly integrated into various stages of cancer care, including tumor detection, diagnosis, clinical decision-making, risk stratification and prognosis, treatment planning, drug response prediction, and clinical research [3,4]. Although AI has transformative potential in health care and an increasing number of AI products on the market, the number of clinically implemented AI tools is still relatively low [5]. In oncology, evidence of AI's clinical impact is still limited [4], and there is a lack of understanding regarding how AI is perceived, accepted, and used by clinicians in real-world settings. Juluru et al [6] draw attention to the difficulty of deploying AI tools into the workflow of clinical practice, which poses a challenge to their adoption. While different workflow models have been evaluated in controlled settings, Verma et al [7] highlight that a clear gap exists between research and clinical environments, emphasizing the need for more focus on the role

XSL•FO
RenderX

and impact of AI in clinical settings. To bridge this gap, empirical studies are needed to investigate the integration of AI tools into clinical workflows [8] and verify the benefits of AI tool use in the real world [9].

AI can be defined as the ability of machines to perform cognitive functions usually associated with the human mind, such as perceiving, reasoning, learning, interacting with the environment, problem-solving, decision-making, and even demonstrating creativity [10]. These qualities set AI technologies apart and present unique challenges and opportunities that influence user acceptance. In this study, AI tools refer to AI systems, primarily, but not exclusively, based on machine learning and deep learning techniques, designed to perform clinical tasks relevant to cancer care. The value of AI tools in practice can only be fully realized if such technologies are integrated into routine clinical workflows [11,12] and if clinicians are willing to accept and use them [13]. As prior research has shown, user acceptance plays an integral part in any technology's success [14,15]. Acceptance of AI by clinicians is crucial for its successful integration into clinical practice. AI acceptance depends on how clinicians perceive its impact on their work processes and professional environment [13]. Despite the growing body of research, few studies have examined cancer care-related firsthand perspectives. Furthermore, most prior work has emphasized technical performance, leaving a gap in understanding how AI integrates into real-world clinical workflows. Finally, while acceptance models (eg, Unified Theory of Acceptance and Use of Technology [UTAUT]) and sociotechnical perspectives (eg, sociotechnical systems [STS] theory) have been applied separately, their combined use offers a rich account of how technical, social, and organizational factors jointly influence AI adoption.

This study, therefore, explores clinicians' perceptions of the use of AI in clinical practice. It aims to identify factors that influence clinician acceptance and the successful integration of AI into routine workflows. We ask the following research questions:

1. What are the perceptions of health care professionals on the integration and use of AI in oncology clinical practice?
    a. How does the adoption and use of AI tools impact the clinical workflow in cancer care settings?
    b. What factors influence clinician acceptance of AI tools in cancer care settings?

We draw from the STS theory and the UTAUT as lenses for interpreting our data. This study contributes an empirically grounded understanding of AI acceptance and integration from the perspective of clinicians. It identifies and consolidates key factors influencing adoption while highlighting the complex interplay between organizational dynamics, human factors, technological characteristics, and environmental factors that shape AI implementation into clinical practice.

User Acceptance in AI Adoption

The goal of technology adoption is for the technology to be used to address specific needs, and effective utilization depends on user acceptance. Nonacceptance of technologies by users results in failed technology adoption efforts [16]. Electronic health records (EHRs) offer an example of how health care technologies, if not well accepted or integrated, can negatively affect clinical practice. EHR systems were initially introduced to streamline documentation and improve efficiency, yet many clinicians report that they increase administrative burden, fragment workflows, and contribute to burnout [17,18]. Studies have shown that simple tasks, such as medication renewal, may involve more than 20 steps within the EHR system, requiring clinicians to spend substantial time on manual work [17]. In some cases, physicians spend nearly half their workday on desktop medicine rather than direct patient care, with additional hours spent documenting outside of normal work hours [19]. This slows down workflows and leads to cognitive overload and clinician dissatisfaction [20]. The example with EHRs demonstrates how poorly designed or inadequately integrated technologies can become burdensome rather than supportive.

These concerns also apply to the use of AI in health care. If AI tools are not transparent or explainable, clinicians may be unable to understand or trust the decisions AI tools make, making them reluctant to adopt or rely on such systems [21-23]. This is important in clinical environments where accountability and patient safety are critical. Studies show that the lack of publicly available information on the development, validation, and ethical considerations of approved medical AI tools poses serious transparency gaps that hinder clinicians' ability to assess their safety and reliability [21]. Similarly, Marey et al [22] indicate how the lack of transparency in complex AI systems, in the case of cardiovascular imaging, poses significant barriers to clinical acceptance. The latter highlights the need for sufficient training of clinicians on AI products [22]. Furthermore, clinician acceptance of AI is strongly influenced by concerns around trust [23]. Clinicians may resist relying on AI tools in their daily practice if such tools introduce uncertainty or misalign with clinical judgment, which eventually results in reducing workflow efficiency and deterring use [13,23].

A Sociotechnical Perspective

We draw on STS theory to understand AI adoption in oncology. STS emphasizes the interdependence of social subsystems (people and organizations) and technical subsystems (infrastructure, tools, and processes), and the principle of joint optimization: outcomes are best when both subsystems are aligned [24-26]. This lens highlights how AI integration depends not only on technical performance but also on human roles, organizational dynamics, and external influences such as regulations and other environmental factors [26]. Applying STS allows us to examine AI adoption beyond technical efficacy, focusing instead on how AI reshapes clinical workflows and how clinicians, organizations, and technologies co-adapt in practice.

A Model for User Acceptance

Individual user acceptance has been studied for many years, and many frameworks and models have been developed that effectively explain the factors that influence acceptance and use. In this work, we use the UTAUT. Developed by Venkatesh et al [27], UTAUT synthesized 8 technology acceptance models including technology acceptance model (TAM), theory of

XSL•FO
RenderX

planned behavior, motivational model, the combination of TAM-theory of planned behavior, theory of reasoned action, social cognitive theory, the model of PC utilization, and innovation diffusion theory, and proved to have higher explanatory power over all 8. UTAUT has been used in many studies exploring the acceptance of various technologies, including those used within health care. As proposed by Venkatesh et al [27], UTAUT has 4 core constructs—social influence, effort expectancy (EE), performance expectancy (PE), and facilitating conditions—defined as:

- PE refers to the degree to which an individual believes that using the system will help him or her attain gains in job performance [27].
- EE is the degree of ease associated with the use of the system [27].
- Social influence (SI) is the degree to which an individual perceives that important others believe he or she should use the new system [27].
- Facilitating conditions (FC) refer to the degree to which an individual believes that organizational and technical infrastructure exists to support the use of the system [27].

Attitudes, understood as the organization of one's beliefs about an object, that incline them towards certain actions [28], are also argued as determinants in some literature sources. PE, EE, and social influence are all direct determinants of behavioral intention (BI). BI, in the context of this study, refers to a person's intent to use AI tools. BI, in turn, determines actual AI tool use. EE, social influence, and PE are moderated by additional constructs: Gender, Age, Experience, and Voluntariness. Experience pertains to the individual's experience with the new technology. Facilitating conditions, on the other hand, are a direct determinant of actual use [27]. In this study, UTAUT constructs inform our understanding of how AI acceptance takes shape within health care institutions.

Methods

Study Design

Given our aim to identify key factors that shape clinician acceptance and the impact of AI integration in clinical practice, this study used a qualitative research design and semistructured interviews to elicit perceptions and opinions on AI use in real-world settings. Semistructured interviews gained popularity for their flexibility and versatility, allowing the researcher to be structured, yet free to explore other relevant ideas that might come up during the interview [29,30]. They are well-suited to research questions aimed at addressing whether a proposed service is warranted**,** how to implement it, how it is performing, or how it can be improved [29]. The interview guide is provided in Multimedia Appendix 1.

We adopted purposive sampling to ensure the inclusion of participants with relevant expertise. The following selection criteria were applied: (1) the person works in oncology or other cancer care-related fields; (2) the work area has implemented, is evaluating AI tools for implementation, or has considered AI tools; (3) the participant is familiar with AI and its uses in cancer-related care. This approach was chosen to specifically target individuals with direct experience or informed perspectives on the use of AI in cancer care. The study included 18 health care professionals involved in cancer care and a medical AI research and development expert. Participants were recruited using professional networks, referrals from prior interviewees, and institutional platforms such as official websites, professional associations, and digital publications. Our sample primarily included professionals using, implementing, or evaluating AI tools for some cancer care-related tasks, as well as those exploring AI tools that could be suitable for use in their clinical areas. A small number of nonusers were also included. Nonusers can offer valuable insights into technology acceptance studies because these participants may elucidate on reasons for nonacceptance. It is also worth noting that some clinicians are involved in research. Table S1 in Multimedia Appendix 2 details the characteristics of the study participants.

Data Collection

Interviews were conducted online, primarily via Zoom, except for one interview using Microsoft Teams, based on a participant's preference. Interviews took an average of 42 minutes of discussion ±3 minutes for introductions. Interviews were conducted in English (n=12) and Swedish (n=7). All interview sessions were recorded using an independent recording device (Olympus digital voice recorder) to ensure audio quality and data security. Transcription and translation were carried out using our university-approved (adhering to the university's policies) tool, which provides initial automated transcription and translation called Amberscript. This was human-checked together with the recording of the interviews to make sure that the content is transcribed properly and translated. This was done by 2 double checkers to make sure of the validity of the data. We have taken additional steps to double-check with some ambiguity regarding the translation of some Swedish-related terminologies. Only one participant was concerned about being misquoted, and the transcript was approved by that participant after reviewing it. Analysis was performed in English.

Data Analysis

The data were analyzed thematically, following Braun and Clarke's [31] reflexive thematic analysis (TA). TA is applied to qualitative data to identify, analyze, and interpret patterns of meaning [32]. This is achieved systematically through the iterative process of familiarizing oneself with the data, generating codes from the data, generating initial themes, reviewing the themes, and defining and naming themes. We followed this approach, inductively generating codes and subthemes directly from the data. Since we were using a reflexive TA, saturation was not sought. As discussed by Braun and Clarke [33], the concept of saturation as a rationale for sample size does not align with reflexive TA. In line with their arguments, in our study, the following make it difficult to claim saturation: the semistructured nature of interviews, the lack of homogeneity in our sample, and the inductive nature of our analysis. Instead, we used information power [34] to judge the sufficiency of the collected data. The quality of our dialogues with the specifically purposively sampled group of participants yielded rich data that is relevant to our research aim. The

XSL•FO
RenderX

findings align with existing organizational and technology acceptance theories relevant to our research questions. Even though we performed cross-case analysis, we consider the number of participants (n=19) to provide sufficient relevant information, given that implementation and use of AI tools in clinical cancer care is at its early stages and the availability of relevant participants is still sparse.

TA was conducted using NVivo (version 14; Lumivero) and yielded 378 codes. Both authors reviewed the codes and the themes. This involved discussing the codes and deciding their suitability within allocated subthemes and later assessing for fit within the theory-based themes. The inductively generated codes were subsequently organized into higher-order themes aligned with STS theory. In this way, STS theory served as an interpretive lens, not a coding framework. Table S2 in Multimedia Appendix 3 shows how the subthemes map to STS theory components. Although this study does not quantitatively measure acceptance as is commonly done with the technology acceptance model, our results nonetheless align well with UTAUT constructs; hence, the framework was subsequently applied to discuss the results and provide deeper insights into clinician acceptance. The codebook is provided in Multimedia Appendix 4.

Researcher Reflexivity

The authors do not work in the medical domain. Both authors have a technical background; their knowledge, therefore, informed the development of interview questions and the flow of discussions and influenced the interpretation of results. To mitigate potential bias, the authors iteratively checked codes and themes and held continuous discussions.

Ethical Considerations

Ethical approval for this study was granted by the ethical adviser of Karlstad University (HS 2023/1591) before data collection. An information sheet and consent form covering both participation and interview recording were sent to participants at least 4 hours before the scheduled interview. At the start of each session, the study was introduced, and participants were allowed to ask questions before proceeding. The consent form is provided in Multimedia Appendix 5. Participants were provided a lunch voucher (≈€20 [US $19.44]) only after the interviews, as a gift in appreciation for their participation. Most of the participants accepted the voucher; however, a few declined.

Results

Study Themes

Participants were asked to share their perceptions and opinions on the adoption and use of AI in their professional context, and more broadly in cancer care. Their responses inform 6 themes formed from 28 subthemes. The main themes include General AI Adoption Perceptions to set the stage for the participants' perspectives on AI, then Social Subsystem (Organization), Social Subsystem (People), Technical Subsystem, Impacts of AI integration as Indicators of Joint Optimization, and External Systems, in line with the components of STS theory. The themes are presented in Table 1, along with the associated subthemes. An extended version of this table that includes the number of coded references and the number of participants linked to each of the subthemes is presented in Table S3 in Multimedia Appendix 6.

XSL•FO
RenderX

Table . Overview of themes and related subthemes. The letters G, SO, SP, T, J, and E serve as identifiers for the subthemes and their associated theme.

Theme (abbreviation used in subtheme numbering)	Subthemes
1. General AI[a] Adoption Perceptions (G)	• G1. AI[a] developments • G2. Limitations and complexity • G3. AI applications
2. Social Subsystem – Organization (SO)	• SO1. Organization-specific challenges to adopting AI • SO2. Facilitating factors at the organization level
3. Social Subsystem – People (SP)	• SP1. Need for AI awareness • SP2. Attitudes towards AI • SP3. Age-related perceptions • SP4. Trust • SP5. Involvement in shaping the solution • SP6. Responsibility and Accountability • SP7. Automation bias • SP8. Clinician autonomy, behavioral impacts, and interpersonal factors • SP9. Effect on jobs, skills, and competencies
4. Technical Subsystem (T)	• T1. AI potential • T2. Fit-for-task design • T3. Ease of use • T4. Challenges relating to data • T5. Interpretability and Explainability of AI tools • T6. Unsuccessful AI adoption efforts • T7. Effects of AI tools on workflows
5. Impacts of AI Integration as indicators of joint optimization (J)	• J1. Benefits to clinical practice • J2. Clinical benefits • J3. Continuity
6. External systems (E)	• E1. Regulatory influence • E2. Vendor stability • E3. AI and clinical guidelines • E4. Macro-level enablers

[a]AI: artificial intelligence.

General AI Adoption Perceptions

This theme captures the participants' general perceptions of AI in cancer care.

G1: AI Developments

Participants hold varying views about the developments in medical AI. While some perceive it as fast-moving, with a lot happening in a short space of time, others see it as moving more slowly than initially expected. As one participant states:

> *I'm a little surprised that we haven't… that we're not further along than we are in those applications. If you think back five years and try to predict where we would be, you would have thought we would have come further.* [P19]

However, the pervasiveness of AI in society is expected to increase exposure to AI tools, resulting in younger professionals who are more likely to be comfortable with AI use. Similarly, large language models are expected to make an impact in clinical care in the near future. One participant also calls for the health care industry to reconsider its approach and focus more on proactive care. In this area, AI can make a substantial impact, and patients can benefit more in the long term.

G2: Limitations and Complexity

A small number of participants express uncertainty about the future of AI and its promise to cancer care. They caution that, in addition to the performance of tools, institutions should also look at the decisions that established the gold standards upon which the tools are built. Furthermore, they argue that the complexity of health care deems AI incapable of handling health issues in the same nuanced way as clinicians. For example,:

> *What you realize is that healthcare is so enormously complex… this AI doesn't have human common sense. And a lot of what we do in healthcare is another kind of complex assessment of the best choice for a particular patient, taking into account other diseases, social situation, general status of the patient and the patient's wishes… there is so much that is weighed into treatment choices.* [P13]

Currently, AI tools support only a small part of the assessment process.

G3: AI Applications

Participants mention several applications of AI, both in use and under consideration, and these include uses in mammography, segmentation of organs at risk, screening for lung and pancreatic

XSL•FO
RenderX

cancer, surgery, genetic exams, patient note-taking, patient communication, and data consolidation tools.

Social System - Organization

Overview

This theme examines organizational-level factors that shape AI adoption. Participants discussed both challenges and facilitators to adoption at the organizational level.

SO1: Organization-Specific Challenges to Adopting AI

Participants describe the validation and security processes performed at the initial phase as laborious, especially as the same tools must be revalidated in different cities within the same region. They question the logic of repeating these processes. However, they acknowledge that defining boundaries, distinguishing between what is considered local and what is not, and determining whether validation in context is truly applicable to another may be challenging. Participants also highlight the need for continual validation rather than a one-time occurrence. They note the procurement process as another challenge, mainly due to bureaucracy. Participants also identify user buy-in as problematic, with the "staff on the floor," as one patient puts it, being the hardest to convince that the tools are to their benefit: :

> *we have done it this way always. So, why do we need to change how we do things?* [P1]

Participants also note issues such as a lack of motivation to learn and use new system features, and hesitation to adopt tools that have not been widely accepted, even though there is sufficient evidence to support them. Other reasons include the expectation of big impact solutions before they are worth considering. For example, one participant believes that AI in dermatology will only be truly worthwhile when an AI solution exists that can scan the entire body for suspicious lesions, rather than smaller tools that target a small surface area. Moreover, organizations are said to fail to provide clear guidance on what AI tools to use and how to use them, leaving employees to explore on their own.

Lack of clinician engagement can also be detrimental to AI adoption. In one case, radiologists have been dissatisfied with the implementation of an AI tool. They resist the AI tool because they had not been informed early in the adoption process. This is viewed as continuing to overshadow the perceptions of the involved radiologists, regardless of the same tool having been accepted in other institutions.

SO2: Facilitating Factors at the Organizational Level

According to the participants, several factors facilitate the adoption of AI tools at the organizational level. These factors include formal communication, informal social discussions, preparation, and technological orientation of organizational units. Formal communication early on is required to get people on board and reduce skepticism. Participants emphasize the importance of presenting concrete use cases and examples tailored to specific specialties that illustrate how the tools can be beneficial and where they would fit in the workflow. This approach of giving concrete examples is seen as more effective than discussing AI in broad, abstract terms. AI champions are said to play an important role in communication. Informal social discussions among colleagues also play a role in facilitating acceptance. According to participants, the talks in their social circles are concerned with which AI tools to use, which are best at performing specific tasks, and the positive impacts of having introduced AI tools, such as time-saving, which is reportedly very visible to staff.

Thorough preparation before introducing AI tools is another key factor. Participants mention the importance of conducting evaluations and validation and providing sound scientific evidence, characterized by robust methods, sufficiently large datasets, and clear presentation of results, both from internal unpublished reports and publications. According to one patient:

> *The most important thing spontaneously is that there is a good scientific basis and a good study with many included patients, where clear results have been shown. That is kind of the basis for being able to get skeptics to believe in AI.* [P13]

Participants also stress the need to analyze beforehand how implementation would affect people and their daily work. A good approach to preparedness is specifying suitable criteria. A participant from one of the Swedish regions notes that they take into consideration 5 key things: having a good use case, quality of the AI tool, economic viability, being well-integrated technically, and the possibility to validate it within the acquiring health care institution.

The technology orientation of organizational units, along with their tech-savvy experts, appears to be more willing to accept AI tools, though the same can also breed skepticism. Participants note that specialties like radiology and surgery, which are already tech-heavy specialties, may be more open towards trying AI tools, which extend upon what is typically used. Even when the clinicians have negative feedback about the performance of the specific tools they test, there is an impression that they are eager and even excited to try them. Participants highlight that there is ongoing technological development in their fields, and one should be prepared to explore new solutions. However, participants also recognize that not all colleagues are tech-savvy, and those colleagues with less technical experience may still be uncomfortable with the latest technologies. Interestingly, some participants state that while familiarity with technology makes it easier to accept AI, it also, in some ways, makes them more critical and want a deeper understanding of the underlying parts of how the tools work before they can fully accept them.

Social System - People

This theme explores the perceptions about human and social factors that influence AI acceptance and use, along with the broader impact of AI use on practice and the workforce.

SP1: Need for AI Awareness

Discussions with participants underscore the critical need to increase awareness amongst health care professionals about what AI entails, which tools are available for use, how they are used, and the potential unintended consequences that might arise from them. To enable clinicians to detect potential errors emanating from AI tools, participants emphasize the need to

XSL•FO
RenderX

enhance professionals' understanding of how AI tools function, as well as their limitations. One participant stressed that

> *You have to know what you are dealing with, of course… you have to know what you are working with to be able to use it in the intended way.* [P19]

Participants also emphasize that the AI tools can make mistakes and should not be relied upon without critical thinking.

SP2: Attitudes Towards AI

Participants communicate that attitudes towards AI tool use are largely positive. They use phrases like "happy we have it," "generally satisfied," "very positive," "healthy and reasonable attitude," "positive in general," and some express a desire to work with AI more. Positive attitudes also appear to be influenced by perceived benefits, as well as positive personal experiences with AI tools used for nonclinical purposes. Participants encourage the use of AI tools to support clinical tasks:

> *We need to look at it as a helper more, and I think we can work more with the patient if we embrace the AI products that are in development in our different areas.* [P1]

However, they also stress the need to maintain a critical perspective. To a lesser extent, negative feelings towards AI tools are also evident, emanating from a lack of communication, indifference, skepticism based on their personal beliefs about AI, and fear of the unknown trajectory of AI. Participants note that skepticism diminishes as people begin to see the performance and results of the AI tools.

SP3: Age-Related Perceptions

Participants express mixed views regarding age and AI acceptance. Some participants believe that younger professionals know more and have more positive attitudes towards using AI tools. However, a different view also exists, suggesting that there is no consistency regarding whether younger or older generations are more or less likely to accept AI. The participant further explains:

> *There are different reasons for different generations to be skeptical or hesitant to use these tools. So, it's not really the same reason for everyone. Some older people, they just don't trust this, um, for some younger people, it's not that much about trust. It's more about fear or confidence level; and not all of them [reasons] are invalid.* [P7]

SP4: Trust

There are mixed views regarding trust in AI tools. Participants note that the question of trustworthiness of AI tools remains, with concerns about how they can ascertain quality and reliability. Such concerns affect some professionals' willingness to test AI tools. Nonetheless, others use AI tools for supporting clinical work, even though they highlight that they would not trust them unquestioningly. For example, one says:

> *I think that it won't happen that I blindly trust this. But it will be like a help… like I'm on the right track. It will help me do a potentially better assessment.* [P8]

Participants also highlight that trust and openness towards AI tools are influenced by previous exposure and experience, scientific evidence, certification of tools, and thorough evaluation in the specific context of use. Some participants question why AI should be trusted less, given that health care professionals already trust and rely on many other technological tools. Still, others point out that health care professionals would rather place greater trust in their own and their colleagues' abilities over AI tools.

SP5: Involvement in Shaping the Solution

The degree to which individuals are involved in shaping AI solutions plays an important role in ensuring alignment with clinical goals. One participant states:

> *So, it's extremely important to work with the companies that have created the AI solution to improve the functionality of the solution. And I have regular meetings with the product developers at that company where they ask questions that they want answers to from a clinical perspective.* [P2]

Participants describe their involvement as ranging from no involvement to some form of feedback to demonstrations of proposed solutions, when change is required during the validation stage, or when institutions have begun using AI tools. While most participants report limited opportunities for health care professionals to be involved in the design of the AI tools they ultimately use, they emphasize the importance of doing so to reduce problems and anchor the solutions properly to the users' needs and expectations. They also highlight that, should they have been involved in design, they could have influenced how the tools work to overcome current shortfalls. Health care professionals can give suggestions when the tools fail to fulfill their needs. Participants believe that early engagement greatly influences how users react to the tool's actual use. They stress that health care professional involvement is necessary during the design stage, and as tools evolve. One participant says:

> *I really don't have specific evidence for what I'm saying now, but it is my experience for many other types of, um, just sort of software tools which clinicians are using. If the clinicians have been involved in the development or early testing of things, they are often quite positive towards the results they get, and they want to use the tools.* [P4]

However, not all participants agree with this view. A different view is that health care professionals need not be involved early in the design, but rather at a later stage, when the AI solution works at least semi-decently.

SP6: Responsibility and Accountability

Although a small number of participants note that it is still unclear who should be responsible and held accountable should something go wrong following an AI tool giving an incorrect result, several participants firmly believe that the responsibility lies with them as clinicians, and they are accountable to the patient. Given that AI tools are intended to support, and the doctors are responsible for verifying their results and making

XSL•FO
RenderX

the final decision, the issue of responsibility and accountability does not seem especially problematic. One supported this saying:

> *AI is very important, but still needs the radiologists to check their data. So, if an exam goes to the patient with the wrong interpretation, the error is from the radiologist. So, it's always on the radiologist, not on the AI.* [P10]

One participant states that it is the developer's responsibility to build a tool that minimizes the number of false positives, to use sufficiently large datasets, and to carry out performance measurements against established benchmarks. At the operational level, however, many of the participants do not appear to question where responsibility lies.

SP7: Automation Bias

A small number of participants have already noted adverse effects of AI use, such as automation bias amongst colleagues. This bias stems from over-reliance on AI tools, either by trusting AI tools more than their judgments or due to finding it so easy to click and move on as opposed to taking time to verify the results. One participant observed that younger professionals are more susceptible to believing and being influenced by AI-generated results. One said:

> *When the AI has pointed to something, there's a tendency among some reviewers to just buy the AI's decision instead of trusting their own assessment.* [P14]

Participants stress the importance of exercising caution, verifying results, applying their judgment, and the need for enhanced training to mitigate this problem.

SP8: Clinician Autonomy, Behavioral Impacts, and Interpersonal Factors

Participants highlight issues relating to autonomy. As they state, in some cases, clinicians simply resist algorithmic support even when algorithms work. Instead, they would like to feel that it's their knowledge that gives the correct answer. Regarding the assumption that the clinician determines what is correct, it is also noted that sometimes the AI tool detects cancer that may be difficult for the clinician, or even several clinicians, to spot in an image, thus a patient is assumed to be cancer-free when cancer is in fact present.

Some participants also fear that AI may overtake clinical work, reducing clinicians to passive bystanders and undermining their confidence. A participant emphasizes that it is important for one to acquire one's skillset first, build knowledge, and gain hands-on experience to trust one's own judgment. Only then can AI serve as a tool to enhance those skills, which is a far better approach than becoming overly reliant on AI from the outset. They also highlight how dependence on AI tools develops over time. For instance, they may start off skeptical and anxious to use AI tools, but later the opposite may occur, where they become uneasy when the exams have not also been reviewed by the AI tool. In support of this, one states that despite the ability to work without AI tools, a dependency forms as one uses them, and if removed, their absence is noticeably felt. Further to this, another warns that using AI tools daily carries the risk that clinicians will cease to question the results or consider alternative interpretations.

> *I guess the fear is always about relying on it too much.* [P12]

Additionally, participants highlight that the use of AI tools can impact interpersonal relationships. For example, replacing administrative staff may mean they lose support from them, not just for the tasks the AI tools fulfill, but also for the support tasks the administrative staff does in the periphery.

SP9: Effect on Jobs, Skills, and Competencies

Participants believe that while there may be a threat to administration staff, clinical staff may not have to worry about job loss for several years. Instead, AI tools are expected to serve as support and complement clinical practice. Although one participant, somewhat agrees to this, they still state a persistent concern regarding job impact:

> *But I still think that in the long run we may be working ourselves out, and that it's just a matter of time. But, it feels like it will take considerably longer than we initially feared.* [P14]

AI tools handle simple clinical tasks, and the participants stress that the physician's judgment remains indispensable, both for conducting assessments themselves and for overseeing the AI-assisted processes. Nonetheless, participants perceive a threat to skills that might weaken due to over-reliance on AI tools. Particularly, the younger generation of clinicians who may have more exposure to AI tools may not be getting sufficient hands-on experience. Participants fear that this will result in eroding their confidence in their professional knowledge. To address the issue, they suggest that a plan must be in place that will require a certain amount of work to be done by the clinician, without AI assistance. For example, a radiologist should retain the skill to delineate structures manually. Participants also suggest that younger physicians should redo by themselves the work that AI tools have performed to gain experiential practice. This should be viewed as a necessary part of education. They note that a shift is beginning to happen in medical education towards covering AI fundamentals and preparing medical students for professions that are AI-augmented. In addition, health care institutions provide tool-specific training, helping ensure the correct use of AI tools. Participants regard training to be straightforward and brief, and training sessions typically last about an hour or two. In some cases, multiple sessions are offered. However, attendance in training sessions is also limited by time constraints. In other cases, the AI tools used are deemed intuitive enough to use without prior training.

Technical Subsystem

This theme covers participants' perceptions of AI tools in terms of their potential, the effort associated with their use, and the importance of aligning them with specific clinical tasks.

T1: AI Potential

Apart from already experienced benefits, participants perceive AI as having significant potential across various areas of cancer care and enhancing their clinical practice. First, participants believe AI tools can help health care professionals in identifying

XSL•FO
RenderX

better treatments more quickly; by rapidly collating and analyzing information from various sources, AI can support the selection of the most suitable treatment arm for a specific condition. AI tools can also support compliance with required clinical standards of image taking, particularly for dermatological conditions. In addition, AI tools can aid in evaluating the need for care and improve the safety of medical procedures, as well as relieve radiologists of certain tasks, allowing them more time for other clinical activities. They are also expected to reduce health care costs. One participant explains the potential for better procedures, saying:

> *This is with skin cancer. It [AI tool] can help identify and make the correct diagnosis on the right tumors, and that you should operate on the right spot if it is [present], so that you avoid unnecessary excisions. Because even the most experienced dermatologists operate on four benign changes before they find a melanoma.* [P18]

Another highlighted the potential of AI tools in increasing patient support:

> *I think that can also be something where AI can be really helpful, to help us monitor patients throughout their chemotherapy journey.* [P12]

Second, participants believe AI can aid patients' understanding of their disease, facilitate understanding of scientific literature about the disease, and help patients feel more in control of their treatment. Thirdly, AI can assist health care professionals in staying up to date with clinical practices and developments. It is also expected to accelerate the translation of research findings into clinical practice.

Even with so much potential, participants point out that these tools still have limitations in handling certain functions that require human judgment. For instance, handling decisions on how to proceed with a patient's chemotherapy after years of treatment involves deep contextual understanding that only a physician can provide. Furthermore, clinicians are better at handling standardized and systematized nomenclature, an area where AI tools still fall short.

T2: Fit-to-Task Design

Participants point to the importance of clear use cases and AI tools that are designed for specific tasks that can support their work. One participant emphasizes that as there are many AI tools on the market,

> *It is very important that one is very clear on the suitability of the AI tool for its intended use, and are clear on the kind of testing they should do before they can use it.* [P19]

Participants stress that the tools must meet performance expectations and should have clearly defined limits and boundaries. Participants also note that AI tools should help them achieve specific goals such as improving the quality of care, using limited resources efficiently, and improving operational effectiveness. Without this alignment, there is no incentive to adopt such technologies.

T3: Ease of Use

Although a few participants note that some AI tools initially require more work, resulting in longer processes, they acknowledged that over time, these solutions eventually show improvement and positive outcomes. The longer processes are associated with the need to change work processes and make adjustments to ensure optimal functioning and fit within the workflow. The participants view this as a natural part of change. One states:

> *and there will certainly be a lot of work in the first year when we try to adjust so that it is right and good. So, there will certainly be a period now when things are slower for us and we will not be as productive, but that's part of it. There is not much to argue about. It's just that if you believe in it, you have to accept that it will take a while before all the new routines have settled in and become as good and effective as before.* [P2]

Several of the current AI tools are perceived as easy to understand and easy to use, and where errors occur, recovery requires little effort. One example is that if an AI-based segmentation tool returns incorrectly drawn structures, the radiologists, who possess the expertise to delineate them manually, can easily make corrections. Support is available for health care professionals when using AI tools. The support is provided by vendors, as well as IT departments within health care institutions. High availability and responsiveness when support is needed are appreciated and underscored by phrases like "all the time," "24/7 almost," "continuous support," and "when we have challenging questions... we turn to them."

T4: Challenges Relating to Data

Participants also raise data privacy and security as concerns. Particularly, the potential misuse of patient data sent to the cloud and shared with third-party companies is disconcerting.

> *That is one of the three major challenges that we have encountered. It is the technical implementation to get through data protection considerations, patient data considerations, cloud solutions and all that.* [P9]

However, others point to General Data Protection Regulation compliance and pseudonymization of sensitive data as providing the necessary security. Another ethical concern raised is the potential of AI tools to propagate bias, whether inherent, learned from training data, or intentional. Furthermore, inconsistencies in how colleagues input data can impact AI tool performance.

T5: Interpretability and Explainability of AI Tools

To encompass both interpretability and explainability, this section uses the concept of transparency. While most AI tools in clinical practice are admittedly opaque or "black-boxed," participants express differing views on the necessity of transparency. A lack of transparency is said to introduce uncertainty, as clinicians cannot discern what the model has taken into consideration to reach its conclusions. Some of the participants emphasize the need to understand how the models arrive at their decisions to clarify aspects that may have been missed during a clinical evaluation, or in the case of predictive

XSL•FO
RenderX

tools, to support the development of suitable interventions. However, participants also suggest that transparency is more critical in some contexts than others. For instance, with image interpretation, where clinicians can independently verify results by reviewing the same images, transparency may be less critical. On the other hand, some participants feel that they do not need tool transparency. For them, the tool's high level of performance, paired with supervision by a human expert, is deemed sufficient. For example, one states:

> *No, I don't need to know how it is done inside, how the AI have been calculating. But, of course, I have to compare and see. Do I agree? Is this the right decision? Is it really so suspicious.* [P19]

Others believe that the level of detail about how a model works would be beyond clinicians' comprehension. One participant points out that there is a trade-off between accuracy and explainability of AI tools, noting that improvements in one often come at the expense of the other. As a result, they may settle for partial explanations and experiment with different approaches to understand how the models achieve their results. Nonetheless, another believes that transparency of AI tools could offer tangible benefits to clinicians, providing peace of mind, opportunities for learning, and helping them clarify conclusions better for their patients.

T6: Unsuccessful AI Adoption Efforts

The following examples provide insight into how AI implementations can fail. Participants shared examples of unsuccessful AI adoption efforts that reveal technical- and implementation-related challenges. In one case, a commercial product, despite having been validated in advance by the purchasing health care institution, resulted in too many false positives once introduced into clinical practice. Another example involves an in-house development project that, when benchmarked against commercial alternatives, performed poorly. Even when technical infrastructure is in place, hurdles remain. One participant describes a locally deployed model as difficult to operate and poorly integrated into the existing pipeline, calling it "a pain" to manage. Custom-built systems are presented as producing minimal gains in comparison to larger commercial solutions. At the same time, the participant cautions that the success attributed to some AI tools, particularly in pathology, is often overstated, with many models still failing to generalize well between patient cohorts. They state:

> *So, so far, those claims were, I would say, highly exaggerated because, um, they usually I mean, maybe we haven't seen the right models, but it's like they would over analyze each particular sample, like in a particular cohort, it would work well. If you try to generalize it to other cohorts like the slides which we would have, then it would start to behave in a crazy way.* [P3]

T7: Effect of AI Tools on Workflows

Participants note that AI integration changes work routines, but its impact on workflow varies across contexts. Some require more adjustments than others. For example, one participant states that:

> *As far as finding breast cancer is concerned, it's a big challenge to change all our routines so that they work with the new way of examining breasts.* [P2]

This suggests that clinicians need to accept new ways of working, a process said to be challenging at first and often met with resistance. The participant highlights that workflow changes will be due to the additional steps required in the process of 3D breast cancer screening. Nevertheless, they expect a positive overall outcome eventually. Participants highlight that AI tools in one area of care can significantly alter the patient's pathway after an examination, and the clinicians must rethink how the quick AI responses will affect operations at different times in the day. In breast cancer screening, for example, the use of AI tools introduces a new process of reviewing images, where an AI tool reviews the images at one point, and radiologists review the images at another. There are different points and scenarios in the workflow where senior or junior radiologists get involved and perform their assessment. One simply puts it:

> *The main difference is that some exams go only to one doctor instead of two. So, that's the only difference in workflow.* [P17]

Others note that, initially, screening may take a little longer as images contain more information and AI-generated markings that highlight suspicious changes in a breast image. The workflow can also differ depending on the protocols or image modalities used. For instance, whether a computed tomography (CT) scan, magnetic resonance imaging (MRI), or another interface is used. Several participants feel that workflow changes are not drastic and have minimal impact on cognitive load. In one case, however, they note that an AI tool may perform its task satisfactorily, but a lack of integration with other technical systems may cause a workflow disruption. AI integration into the workflow must be well organized and accompanied by staff training to result in a more efficient process. After integrating AI tools in their work, participants note that evidence shows a positive synergy between good AI tools and experienced clinicians.

Impacts of AI Integration as Indicators of Joint Optimization

This theme reflects the outcome of AI tools and the social system as a jointly optimized system.

J1: Benefits to Clinical Practice

Participants note several benefits. The 2 primary benefits that clinicians note are workload reduction and time savings. Workload reduction is highlighted as one of the key goals of adopting AI in clinical settings, which are currently overwhelmed due to staff shortages.

> *Well, in Sweden, we have a lack of radiation oncologists. So, that means that we are finally able to do our jobs on time. Otherwise, we were always behind schedule. So now we are on schedule again.* [P6]

Participants point out that the introduction of AI tools reduces manual work and increases productivity. They explain that in

XSL•FO
RenderX

breast cancer screening, AI tools handle scans that would otherwise wait for the radiologist to review. Similarly, in segmentation tasks, the time to draw structures is significantly reduced by the automatic generation of structures compared with when they are done manually. A participant points out that:

> *The physicians love it because they say, oh, I've got all this time. I don't need to take this patient [for] three hours. Now I can do something else and take another patient.* [P16]

In speech recognition systems designed for taking health care records, the need for an administrative middleman is eliminated, resulting in quicker access to health care records for subsequent care. Participants also note that the use of AI results in increased access to information and frees up time, allowing them to work more with patients. One emphasizes that the use of AI tools improves current work processes and adds the ability to perform functions that are otherwise not possible for humans. While they may not fulfill all needs, AI tools are seen as making a difference in the care process. Furthermore, participants highlight that AI tools increase accuracy and reliability, can minimize the risk of oversight, and increase clinician confidence in performing certain tasks. Participants also highlight that in the long run, AI may contribute to reduced costs through paying fewer salaries and reducing the need for multidisciplinary team meetings where expert views of different professionals are corroborated to reach decisions about cancer patients; that is, their diagnosis, recommended treatment, and follow-up care. One participant states:

> *And this is a situation where I think this gathering of many different expert views… is going to be, if not replaced, at least assisted by AI, which will start to provide the same comprehensive advice. And perhaps often it will be just an oncologist then, assisted by AI, that makes the decision, ultimately together with the patients, rather than having this [multidisciplinary] conference.* [P4]

Though AI tools may be costly, the cost is expected to be offset by the benefits.

J2: Clinical Benefits

Participants note that the use of AI tools leads to better cancer detection, allowing clinicians to identify smaller cancers at an earlier stage. For example, one participant hares that when the AI tool they are implementing in breast cancer screening was evaluated,

> *They found 29% more cancers compared to when there were 2 radiologists, and they found slightly smaller cancers and clinically significant cancers at an earlier stage.* [P13]

Increased accuracy results in lowered patient recall. Furthermore, a participant highlights that the advanced 3D imaging technology used in breast screening can provide better detection of cancers in dense breast tissue. Using AI tools in segmentation tasks is reported by a participant as yielding better organ structures and missing fewer structures compared with manual delineation by physicians. Participants also observe that AI tools reduce inter-observer variation and perception errors. The greatest benefit, one says, is to the patients, accruing from early detection of disease and potential complications, as well as improved, timely treatment. A participant makes a similar assertion:

> *So, it [AI tool] helps the patient. It lowers the error in the exams.* [P10]

J3: Continuity

Some participants indicate a willingness to change their AI tools due to unmet needs, high costs, and unacceptable error rates. There is also the possibility that other tools are now available on the market that offer better features, better prices, or possibilities to run on on-site servers. A participant put it this way:

> *We would love some extra things that are not there, we are considering doing a new comparison… and see if there are new vendors… who are doing a better job.* [P6]

However, others have reached satisfaction with their existing AI tools and have no plans to look for alternatives. Even though some participants observed that performance has remained consistent since their tools were evaluated and validated, they see continual and long-term validation of the tools as necessary to ensure ongoing reliability and effectiveness. One participant also notes that the first version of nearly any technological tool in a complex system, such as health care, is often imperfect, and better versions tend to emerge over time.

External Systems

This theme reflects the influence of the external systems, that is, the environment, on the STS, and how they can enable or constrain AI integration efforts at the local level.

E1: Regulatory Influence

Participants highlight challenges related to regulations. One notes the complexity of regulating AI in medicine, particularly that models are learning from data and evolving over time, making it difficult to claim reproducibility of results. Others pointed to the lack of clear guidelines that pose a challenge to AI use. One participant states:

> *I think it's a problem to have unclear guidelines and guidelines that are difficult to follow. It sets the stage for people to try to circumvent them or ignore them.* [P9]

Even when guidelines exist, they are said to lack the detail required to give sufficient guidance. Participants perceive inconsistencies in how regulations are applied. For instance, while institutions are required to submit a notice to the Medical Products Agency to perform a prospective evaluation of an AI application, there is no equivalent regulation obligation when purchasing and deploying an AI application in clinical practice. This gap in regulation creates a loophole for the adoption of AI tools that have not undergone proper evaluation.

E2: Vendor Stability

Whether AI vendors will remain viable in the long term could pose a threat to the adoption and sustained use of AI tools. A

XSL•FO
RenderX

participant expresses concern that some vendors may not have staying power. This is a risk for health care institutions that often seek long-term partnerships with technology providers. This uncertainty prompts the preference for in-house developed solutions.

> *We can see, as a risk of course, for the moment. So many companies have popped up who provide us with solutions. Will they be here in 5 or 10 years, 20 years? You know… So, we need solutions that are reliable…* [P1]

E3: AI and Clinical Guidelines

Participants reflect on how they perceive the influence of clinical guidelines and AI tools in cancer care. They acknowledge that AI could give recommendations that are not entirely aligned with existing guidelines, but could still offer better options for some patients: a flexibility that is considered acceptable given the mutable nature of clinical guidelines. Participants note that AI has the potential to help interpret the results of mega-trials and rapidly analyze the vast number of publications that inform guideline development. However, they also feel that AI tools remain inferior to the expert panels who create guidelines, and that AI tools should be optimized to follow guidelines while retaining their potential benefits. Participants also felt that clinical guidelines should acknowledge the increasing use of AI:

> *I also think that the guidelines should mention it [the use of AI tools]… they need to acknowledge AI and give us some guidelines about when it's feasible to use it or not, and some of the drawbacks that they have.* [P15]

Participants also highlight the need for AI developers to stay up to date with guidelines to effect the necessary changes in model behavior, and it is seen as positive when vendors align their AI solutions with international guidelines and can tailor them to national guidelines.

E4: Macro-Level Enablers

As highlighted by participants, infrastructure and governance structures at the national level influence the implementation of AI tools. Technologically advanced countries that also have more centralized digital health systems, like Denmark, are viewed as better positioned to integrate AI tools in health care, as opposed to decentralized ones, like Sweden, that are perceived to have more challenges adopting AI tools. Participants point out the redundancy and cost associated with repeating the evaluation and validation of the same AI tool across different institutions and regions. They add that due to costs, not all regions use AI tools. They further suggest forming collaborations and cooperation between and within regions to promote mutual learning and optimize resources when exploring different AI solutions. One of them says:

> *I hope that in Sweden we can establish a collaboration between different regions and between private and county council-funded care, where we collaborate on evaluations, because there are so many different solutions, and it is difficult to test all of them. It would be great if we could create some forum where we have common procedures on how to evaluate and how to share our experiences.* [P2]

Discussion

Principal Findings

This study explores clinicians' perspectives and opinions on the use of AI in clinical practice, with the aim of identifying key factors that shape clinician acceptance and AI integration into everyday practice. We sought to answer the questions: (1) How does the adoption and use of AI tools impact the clinical workflow in cancer care settings? (2) What factors influence clinician acceptance of AI tools in cancer care settings? To interpret the results, we draw on STS theory and UTAUT as complementary frameworks. We find that the successful integration of AI tools into clinical practice requires organizational support, careful consideration of potential effects on individuals, and alignment between technical offerings and the requirements of clinicians. While perceived ease of use (EE), PE, social influence, and facilitating conditions influence acceptance, attitudes, and trust are also expected to influence an individual's BI and use of AI tools, suggesting an extension to the UTAUT model.

Workflow Impacts and Sociotechnical Alignment

The introduction of AI tools in clinical practice changes the way that care is delivered. Understanding their adoption requires understanding their impact on clinical workflows. In this section, we use STS theory to understand how workflow dynamics change as AI tools are introduced. Participants note that the core goals for adopting AI are improving the quality of care and making efficient use of resources. To achieve these goals, there is a need to optimize clinical workflows while recognizing the interdependence between social and technical elements that comprise the work system.

From the social subsystem perspective, conditions positioning the organization for successful AI adoption include prior preparation, the uptake of robust AI tools and effective technical integration, increasing awareness, careful consideration of impacts on people and their daily practice, as well as feedback, adaptation, and continual validation. Introducing AI tools must take into consideration the local infrastructure, clinical practice, and institutional dynamics that shape and are shaped by their use. Organizational readiness provides a foundation for successful adoption and integration and requires both the capacity to deploy tools and to develop the necessary competencies [35]. Organizations must provide training, and, importantly, ensure that staff have adequate time for training [13]. The key to organizational readiness lies in aligning human roles, organizational processes, and technical functionalities to achieve both adaptation and value realization.

Local validation helps ensure that an AI tool performs adequately within a specific clinical context. It includes rigorous assessment to establish the AI model's reliability, safety, and efficacy [36]. However, such validation efforts can be resource-intensive. This raises the need to reconsider how validation processes might be shared or collaboratively undertaken across institutions serving similar populations,

XSL•FO
RenderX

thereby reducing redundancy and making the process more efficient. Validation has previously been discussed as a single event that occurs prior to the tool's use in clinical practice [37]. This study also highlights the need for continual validation to ensure that systems remain reliable over time, which is especially important as AI models continue to learn from data and undergo adaptations. Local, recurring validation allows institutions to assess both the performance of the tools and the unintended consequences. Consequences such as workflow disruptions, cost inefficiencies, and bias may hinder long-term use [38]. These assessments ultimately inform institutional decisions regarding the continued use or abandonment of an AI tool. Beyond such high-level decisions, the integration of AI brings visible shifts in daily practice.

Integration of AI into routine clinical workflows presents a challenge to effective AI adoption in health care [6]. Clinical workflows change due to the introduction of AI tools, but integration also requires clinicians to adopt different routines, thus shaping the nature of work. Such changes require that stakeholders be adequately involved and informed early on to clearly understand the roles fulfilled by AI tools and the context-specific benefits to be derived from them. This early engagement increases stakeholder buy-in [37,39]. As our results reveal, there are different workflow impacts in different contexts. In breast cancer screening, apart from handling image analysis, the introduction of AI tools also alters the timing and sequence of clinician involvement, necessitating localized rethinking of roles and responsibilities, and prompting reallocation of expertise, such as that of junior and senior radiologists. Even when performing the same type of clinical task, institutions may adopt distinct workflows. The selection of an appropriate workflow is important, as variations in workflow design can lead to differences in clinical effectiveness and implementation success [40]. The speed of processing by AI tools produces results more quickly, which also alters expectations around when and how subsequent tasks are performed. Where segmentation of structures is concerned, AI tools perform the segmentation process that was previously handled manually by radiologists. Radiologists review the AI-generated structures and rectify those that are incorrectly produced. Thus, an asynchronous collaboration between radiologists and AI tools is formed. This collaboration results in more consistent structures and helps reduce inter-observer variability. It also maintains a closer alignment with clinical guidelines. However, participants also highlight that the impact of AI tools varies depending on imaging modality and clinical protocols used. For example, workflows differ depending on whether CT, MRI, or other imaging systems are used. For AI tools to be efficiently incorporated into health care practices, they must be designed with these practices in mind [7]. While many current AI users believe the workflow changes are not drastic, integration of AI tools can be challenging, resulting in workflow disruptions. Efforts should be made to integrate AI tools seamlessly into clinical workflows, as workflow disruption can increase cognitive effort and time for performing a task [41] and hinder adoption [8]. While several researchers highlight the importance of seamless integration into existing workflows, institutions may need to embrace structural changes and new clinical workflows [42].

In this study, we find that the majority of participants were not involved in the design of the AI tools acquired by their institutions, and that several users report that the tools did not fully address their needs. Tools developed without sufficient input from the clinicians who will use them tend to fail to deliver clinical value [43]. Better alignment can be achieved through early involvement of health care professionals in design activities, as useful AI tools ideally involve clinicians who can clearly define relevant use cases, based on their needs [7,43]. Nonetheless, feedback and adaptation, which involve health care professionals communicating with AI tool providers, help with aligning the AI tools with user needs. However, it is done retrospectively. This is important in STS theory, which emphasizes the significance of the interactions between people and technologies in the design and redesign of systems, key to the pursuit of optimization [44].

While the introduction of AI reshapes clinical routines and requires adjustment from clinicians, it also delivers tangible benefits that can improve workflow performance and clinical outcomes. Having put AI tools to use, several benefits are identified, such as time-saving, increased work efficiency, workload reduction, reduced inter-reader variation, increased tumor detection accuracy, and reduced patient recall. The benefits demonstrate that the integration of AI tools supports and enhances existing workflows, benefiting both clinicians and patients. The appreciation of AI tools despite some shortfalls attests to their utility. The realization of benefits, along with the participants' view that great human-AI synergy arises when AI tools are integrated with experienced clinical judgment, aligns with STS's understanding of co-adaptation between humans and technology.

AI tool design and integration should consider human factors [45,46]. The perceptions and opinions of health care professionals have given further insight into pressing concerns associated with AI adoption and use, such as trust, responsibility and accountability, interpretability and explainability, autonomy, and threats to jobs, skills, and competencies. We find that health care professionals are largely optimistic that the use of AI tools will yield positive returns for the health care industry overall. It must be noted that tensions exist between the perceived utility of AI tools and the commitment to maintain clinicians' personal autonomy, where fear of displacement and the desire to preserve professionals' authority have been highlighted. Autonomy extends beyond decision-making to encompass the preservation of clinical judgment, which is deeply informed by formal training, accumulated experience, and tacit knowledge acquired through practice. Verma et al [7] find that the clinicians' need to maintain control over AI is raised in relation to responsibility and trust but is driven by their service to patients. Similarly, our results show that most health care professionals do not perceive accountability as a dilemma; they take full responsibility for patient assessments even when supported by AI tools. They emphasize that AI tools are used only as support and that physicians take the final decision regarding a patient and remain responsible and accountable to the patient. The threat of job loss also does not appear to be a significant concern to most in the short term. An understanding of how AI tools function and their current limitations provides health care

XSL•FO
RenderX

professionals with assurance that widespread job displacement is not imminent but may be a concern for the future. To date, AI tools fulfill narrow-based tasks, whereas clinicians have expertise to handle diverse types of data modalities and have a holistic view of different diseases and patient situations [43]. Nonetheless, there is fear that fewer opportunities will exist for junior professionals and administrative workers. Although the threat may be distant for some, the concern about job security must be addressed [47]. Participants also foresee that skills may erode due to overreliance on AI. Increased reliance on AI tools means that they will have reduced hands-on practice, which may hinder their ability to perform certain tasks independently. Maintaining such skills is necessary if clinicians are to continue to oversee and verify AI-generated results. Measures, such as clinicians occasionally handling clinical tasks without AI assistance, are proposed. On the contrary, clinicians may have to adopt a different type of thinking, accepting that roles and responsibilities shift, and that they may no longer need to possess skills for certain tasks that AI tools execute exceptionally.

Trust in AI tools remains a complex issue. Although AI perceptions amongst clinicians appear largely positive, trust continues to be a concern for many. The novelty of AI in health care, lack of awareness, and uncertainty about AI's long-term impact contribute to issues of trust. Closely linked is the perceived threat to clinician autonomy, confidence in their own work, and impact on interpersonal relationships. Lack of explainability is also linked to diminished trust [7]. Trust or lack thereof influences the adoption of AI tools and human-AI collaboration [47,48]. We also find that over time, however, skepticism appears to lessen with increased exposure, experience, quality control, and demonstrated performance of AI tools. While a lack of trust results in low adoption, trusting AI tools uncritically leads to overreliance and automation bias. Automation bias, along with other cognitive biases, can undermine the effectiveness of human-AI collaboration that is essential for optimal clinical outcomes [48]. This underscores the need for further research into the implications of too much reliance on AI tools and approaches to mitigate it. As participants noted, clinicians may occasionally accept AI-generated outputs without critically reviewing them, such as neglecting to verify the results independently or allowing the AI tool's output to unduly influence their clinical judgment, undermining confidence in their expertise, and increasing patient risk.

Regarding the actual technology, participants view AI tools as having great potential to help improve health care delivery. Achieving this requires alignment between the design and performance of the AI tool and the individual or organizational goal to accomplish a specific task. Explainability and interpretability are well-known challenges in medical AI, touted as posing a significant barrier to the adoption of AI tools in clinical workflows [49]. Explainable AI offers transparency regarding how AI models reach their outputs, enables error analysis, and increases clinician trust [49]. Although interpretability and explainability are clearly desirable, in contexts such as image diagnostics, their lack does not appear to be a significant impediment since doctors can themselves perform the same tasks manually. This allows them to verify, accept, or refute AI-generated results. In line with this, Aung et al [9] argue that since clinicians use technologies such as CT and MRI, whose internal mechanisms they may not fully comprehend, they may also somewhat accept AI tools that lack transparency if their efficacy is proven. Nevertheless, AI tools can identify anomalies that are invisible or inexplicable to human experts, resulting in a lack of consensus between the clinicians and the AI tools. Lack of consensus is not new and can happen between clinicians. In the same way, there are protocols in place to address human-AI disparities. Still, AI as a technical subsystem presents a challenge to the transparency of tool operations, negatively impacting users' understanding and trust in a tool's given outcomes. However, the social subsystem, like in the case of physicians in image diagnostics areas, fills in this gap using their expertise, illustrating how STS relies on the interplay between human insight and technical function. It is worth noting that explainability remains a critical requirement for AI-based clinical decision support systems in oncology [50]. In fact, explainable and interpretable tools would be advantageous in many contexts, but they may not always hinder tool use.

Although data-related challenges are coded under the technical subsystem in our results, this discussion elaborates on how these issues are embedded within broader sociotechnical dynamics. One key data-related concern highlighted by participants is privacy and security, a concern also explored by Momani [51] and Gawankar et al [52]. While compliance with regulations such as the General Data Protection Regulation (EU) and Health Insurance Portability and Accountability Act (US) may address data privacy and security concerns, the other risks associated with the dependence of AI models on data are inherently sociotechnical in nature. Issues of data generation, availability, and correctness cannot easily be disentangled from the social system. Failed AI implementations also do not necessarily imply technical flaws, as inaccurate outputs may stem from an AI tool itself, poor-quality data input, or misalignment with clinical routines, contributing to suboptimal system performance. Another significant concern that arises from the AI models' reliance on data is the propagation of bias, a phenomenon also documented in literature [53,54]. The issue of model bias warrants further scrutiny. Kotter and Ranschaert [43] argue that addressing this challenge requires standardized and regulated monitoring of outcomes.

Beyond the technical and social factors influencing AI adoption are broader external environmental factors. Participants expressed concerns about current regulatory frameworks for medical AI. They emphasize that regulations fail to account for the adaptive nature of AI algorithms, challenging the reproducibility of AI results. This is underscored by Chua et al [50], who also highlighted the need for a regulatory framework that acknowledges AI's self-learning characteristic. The ambiguity and complexity of regulatory requirements, as well as inconsistent enforcement, create loopholes that may be detrimental to clinical practice. There is a need for clear guidelines that offer sufficient detail and are easy to interpret and implement in real-world health care settings. Additionally, the stability of AI vendors may significantly influence their

XSL•FO
RenderX

sustained use. Participants express concern about the longevity of AI vendors, a concern that causes uncertainty since institutions show a preference for long-term partnerships with their suppliers. Continued use of AI tools requires adaptation and technical support, which depend on reliable vendor engagement. It is essential that vendors remain operational and maintain strong relationships with the institutions using their products. Failure to do so would lead to reduced user confidence and eventual tool abandonment. Although developing in-house solutions may be an alternative, health care institutions must be cognizant of the distinct challenges associated with such an approach [37,43].

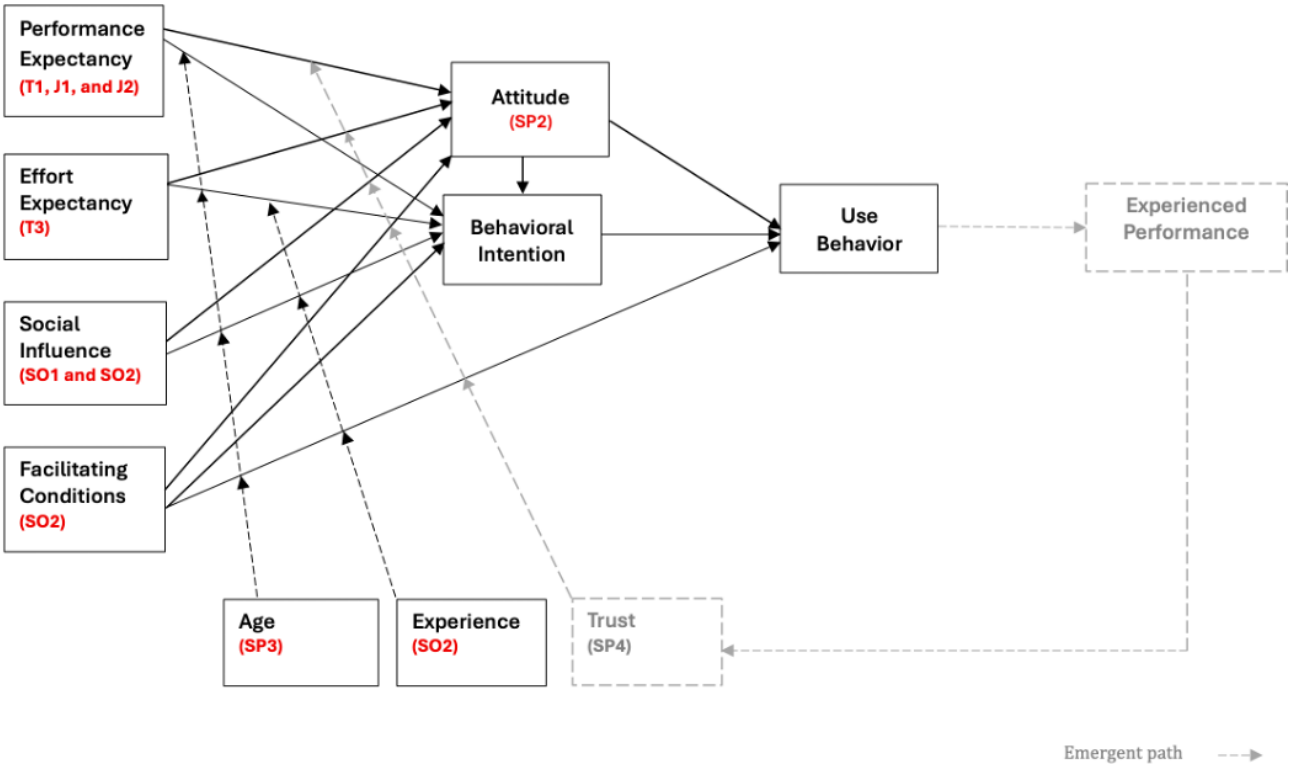
Based on the shared perspectives, it can be inferred that the discussed social aspects are shaped by the views and performance of the technical system. Clinician behaviors evolve over time in response to using AI tools, while the tools themselves adapt through feedback to better meet the needs of clinicians, illustrating a mutual shaping of the social and technical systems. The main principle of the STS theory is that systems are most effective when the social and technical subsystems are jointly optimized to support a system's objectives, such as improving the efficiency of clinical processes. To achieve this outcome, there must be a good fit between the subsystems, which facilitates seamless integration of AI tools into clinical practice. A well-optimized system requires a careful balance between tools that provide value while preserving enough clinician involvement to mitigate concerns of loss of control, which could hinder AI acceptance. Viewing AI through the lens of STS theory emphasizes the need to maintain the human-in-the-loop approach. This is particularly relevant given that our results suggest that participants perceive an optimal state of working when both an AI tool and a health care expert function together in a complementary manner.

Acceptance of AI Tools by Clinicians

Overview

To answer the question of acceptance, we turn to the UTAUT model designed for understanding user acceptance. While themes aligned with the STS lens, some of our results reflect the constructs of UTAUT. UTAUT explains factors influencing BI, which in turn influence the use of technology [27]. The STS view of adoption, covering technical capabilities, user perceptions, and social factors, aligns well with the UTAUT constructs. Our results, on the other hand, also include actual use, as some participants have already begun using AI tools, and the use of technology is a strong signal of acceptance and successful integration. Figure 1 presents the UTAUT model, depicting the supporting factors according to our results, with relevant subthemes from the TA in parentheses, for example, T3 below.

Figure 1. UTAUT diagram, with relevant subthemes between the parentheses in red (adapted from Dwivedi et al [55], which is published under Creative Commons Attribution 4.0 International License [56]). The moderators (age and experience) are as per the original UTAUT model by Venkatesh et al [27].



Effort Expectancy (T3)

Our findings show that participants perceive AI tools to be easy to use. As a result, individuals are likely to accept them [14,27]. These findings are in line with several studies identified in a recent literature review by Madan et al [8]. Indeed, some views suggest that some level of difficulty exists. Still, these views are accompanied by expectations that such challenges only happen at the initial stages after the AI tools are introduced, but tend to diminish in the long run. Results also indicate that understanding the tools does not require intense training.

XSL•FO
RenderX

Performance Expectancy (T1, J1, and J2)

The results show that AI tools are perceived to be useful. Perceived usefulness [14], also known as PE [27], is a strong predictor of BI to use a technology [27,57]. PE can be drawn from our results covering the potential for AI tools and the benefits of their use. The results show that participants expect and even experience increased performance with AI tools integrated into clinical practice. AI tools have the potential to streamline care and alleviate pressure on strained resources, provide informational support for patients, and support knowledge development. Participants report that AI tools align well with the tasks they need to perform and contribute to increased accuracy, reduced workload, reduced inter-reader variation, reduced perception errors, time savings, and efficient work processes.

Additionally, using AI tools is seen as a safeguard, likely to increase their awareness of potential human errors. These tools support clinicians' performance and benefit patients through early tumor detection, better treatment, and reduced recall. Given the overarching goal to provide high-quality care while optimizing resource use, these perceptions underscore the importance of EE in shaping positive attitudes and BI to use AI. These participants' perceptions are critical as successful adoption of technologies often depends on how individuals perceive their relevance and utility [58]. While our results indicate BI to use AI tools, the realization of the outcomes mentioned above also confirms actual usage.

Social Influence (SO1 and SO2)

Our results on social influence highlight that information exchange occurs through informal professional interactions. These discussions are positive in nature as they revolve around best-performing tools for specific use cases, as well as the visible benefits derived from the AI tools that are introduced, which are likely to influence or even reinforce their intention to use them. Since many clinicians face time constraints and huge workloads, AI tool benefits such as time-saving and workload reduction can encourage their acceptance [13]. Social influence is expected to have a more substantial effect on BI where tool use is mandated rather than voluntary [27]. Apart from facilitating formal communications, AI champions can also be great influencers.

Facilitating Conditions (SO2)

Several facilitating conditions have also been mentioned by the participants, such as validation of tools, organizational communication, and integration with other technical infrastructure. Facilitating conditions extend to ensuring access to the necessary knowledge [59]. Availability of training and support facilitates the use of AI tools and has been highlighted by participants. Proper training on how to use AI tools increases confidence and encourages acceptance of AI tools [13]. However, we also find that, in some contexts, AI tools are seen as intuitive enough to integrate and use without prior training. This may be particularly true in technically inclined specialties where the introduction of AI tools can be easily integrated into existing routines. In other contexts, however, concerns are raised regarding how well tools integrate with other existing platforms, which deters from the willingness to continue using them.

Moderators (SP3, SO2, and SP4)

Age is expected to play a role in the willingness of health care professionals to accept and use AI tools. While this is not reported in first-person terms, discussions suggest that there is a difference between how younger professionals perceive AI compared with older professionals. Age has already been proven to have a moderating effect between the first 3 constructs and BI [27] regarding technology acceptance. However, its impact in relation to AI tools should be further investigated. Exposure and experience with AI tools, even prior to joining health care institutions, are expected to positively influence the willingness to use AI tools in clinical practice. Participants expect younger professionals to have had more exposure and experience with AI tools, both in educational and personal settings. Apart from this, our study suggests that clinicians who work in tech-heavy specialties, while critical towards how AI tools perform, display less resistance towards AI use. Future research can elucidate whether the experience of technology use in general plays a role in individuals' BI to use AI tools.

Attitude (SP2)

Attitude, which had been included initially as a construct in the TAM, was not included in UTAUT after being found to be redundant as a predictor of BI. Dwivedi et al [55] reintroduced attitude to their UTAUT adaptation, finding that attitude plays an important role in acceptance and use of technology. In turn, attitude is influenced by EE, PE, social influence, and facilitating conditions. The latter is especially true in nonvoluntary settings. Our results reflect a more positive attitude than negative ones, as well as a keenness to use AI tools. According to Dwivedi et al [55], users with a positive attitude are more likely to use the technology; thus, attitude influences both BI and usage behavior. As earlier established, trust (SP4), or lack thereof, can impact whether individuals accept technology. The issue of trust is more pronounced in the context of AI tools, where functionality demands more autonomy and offers less transparency. Trust is likely to have a moderating effect between PE, EE, SI, and FC and an individual's attitude towards AI tools. We have also learned that experienced positive performance reduces skepticism and increases an individual's level of trust. Further research can elucidate the hypothesized effect of trust. A recent study [60] finds that age and experience do not significantly influence attitude.

The alignment of the data, as shared by participants, with UTAUT constructs suggests a strong foundation of acceptance of AI tools in cancer care, while also highlighting potential hindrances. The already established use of AI tools, accompanied by a broadly positive attitude, reflects acceptance and supports UTAUT's outcome of use behavior. However, continued use is contingent on consistent performance of AI tools, adaptability, and addressing unmet needs. While UTAUT helps us understand individual acceptance, its constructs clearly show the interplay between the social and the technical aspects of an STS.

XSL•FO
RenderX

PE, while seemingly technical, is not only about AI's capabilities and its usefulness, but it is also shaped by social perceptions and experiences. Performance of AI tools shapes how individuals perceive them, and perceptions of usefulness, in turn, influence the social aspect, human BI. EE may also appear to be a social aspect since it pertains to how people feel about using the technology and the effort required from them; however, it is influenced by the tool design, interface, and technical support. When AI tools are effectively integrated into clinical workflows, they are more likely to reinforce the perceptions of their usefulness. Social influence is more obviously tied to the social subsystem: people, relationships between them, and social norms. Facilitating conditions entwine the social and technical systems in that they encompass training and other factors relating to organizational preparedness, as well as technical infrastructure, interoperability, technical capability, and fit. These constructs influence sociotechnical outcomes: BI and use. This overlay between UTAUT and STS signifies how AI acceptance is shaped by the interplay between AI tools and the social system that they are intended to support. From an STS perspective, the relationship between user acceptance and the successful integration of AI tools in clinical workflows is bidirectional. Well-integrated AI tools that align with clinical tasks, provide utility, and minimize disruption can foster clinician acceptance [13,50]. However, clinician acceptance can also facilitate a smooth integration of AI tools into the clinical workflow. Openness and willingness to engage with AI tools enable learning, engagement in feedback and adaptation, and meaningful use.

Limitations

Recruiting participants with the specific profile required in this study was challenging due to their limited availability and demanding schedules. Using purposive sampling, we attempted to recruit 135 potential participants. Although efforts were made to recruit participants from a range of medical specialties within the set criteria, and without emphasis on proficiency levels, there remains a possibility of selection bias, as researchers cannot guarantee reaching all relevant individuals. Additionally, the voluntary nature of participation may have also introduced self-selection bias, whereby those with a stronger interest in AI were more likely to take part. We do not claim generalizability of these results to all oncology clinicians. However, the participants' time constraints and scarcity in research participation make the contribution of the participants especially valuable, lending it a unique strength.

Our sample included 18 health care professionals involved in cancer care, encompassing both practicing clinicians and nonclinical roles such as biomedical technicians and medical physicists, as well as one participant from AI research and development, to capture a broader range of insights related to AI adoption and use in cancer care. Although not all participants were patient-facing clinicians, those in nonclinical roles play a critical role in implementing and managing AI tools, particularly in diagnostic imaging, radiation therapy, and treatment planning.

While the sample size is adequate for qualitative analysis, it remains relatively small and is drawn from specific health care institutions, primarily located in the Nordic region. This may limit the transferability of the findings to other contexts or countries. The majority of participants are based in Sweden, a technology-forward country. Such a positive view may bias the results towards more positive perceptions of AI tools. Additionally, the study focused on contexts where image-based AI tools are used. For this reason, issues that may be unique to cancer care contexts that use different data modalities could have been missed. However, this is not expected to invalidate the results, as literature shows that the most approved AI tools in cancer care are AI-based image diagnostic tools.

We acknowledge that the term "AI tool" may be perceived as too general. However, we maintain its use as an umbrella term because participants describe a variety of AI tools and even similar tools used differently within similar clinical domains. This diversity reflects real-world variation in AI implementation and related experiences across health care settings. As with all qualitative studies, the findings are intended to offer rich, contextual insights into the phenomena of AI acceptance and adoption and not statistical generalizability. Despite these limitations, this study offers a valuable perspective into the real-world application of AI tools in cancer care from a sociotechnical viewpoint.

Conclusions, Recommendations, and Future Research Directions

Participants (primarily users/evaluators in settings where AI tools are present) generally perceive AI tools to deliver tangible benefits for both clinicians and patients. However, successful implementation requires careful integration into workflows, alignment with clinician needs, and continued validation. Applying STS theory, we offer a holistic view of how AI tools shape and are shaped by health institutions, people, and their work processes. Understanding this sociotechnical interplay is critical for AI tools that support clinicians in delivering quality care while preserving clinical expertise. Using UTAUT, we identify the determinants influencing the clinicians' intention to adopt or continue using AI tools. Our findings suggest that acceptance is not only a precursor to use but can also result from how well AI tools integrate into and support established work routines and clinical needs. Although a general positivity towards using AI tools is seen, we highlight the need for health care institutions to pay attention to the concerns that negatively impact clinicians' acceptance of AI tools, such as the need for communication early on in the project, performing an impact analysis to ascertain tool impact on people and workflows early on, increasing knowledge of AI and potential negative consequences, and carrying out continual validation.

More importantly, the newly arising negative effects of AI use, such as automation bias and potential skill erosion, must be addressed. Mitigation strategies, such as periodic non-AI proficiency runs where work should periodically be done independently and without the support of AI tools, have been suggested by participants to preserve skills. Another way is AI implementation that enforces double reading for edge cases to ensure that 2 clinicians attend complex or ambiguous cases and continue to apply and sharpen their skills.

External factors such as regulations, vendor stability, and broader infrastructure and governance at the national level can

XSL•FO
RenderX

also facilitate or hinder adoption and sustained use. Current approaches to validation are perceived as onerous. The cost, time, data, and data preparation tasks, and effort to perform local validation for every trial AI product are not sustainable and delay AI adoption. As suggested by participants, shared regional validation consortia may be a solution. This may be a reasonable approach for regions within a single country or group of countries where a similar patient population is served. Within such a consortium, validation strategies, resources, and experiences can be shared, which may reduce costs and uncertainty and result in more robust, long-term implementations.

Future research should address training needs and how to implement safeguards against unintended consequences, such as overreliance and skill erosion, as well as prioritize co-design with clinicians to ensure better tool-task alignment.

Acknowledgments

We want to thank Neshe Tuna, who assisted during the data collection process. We would also like to thank our participants for sharing their time and valuable contributions.

Funding

The authors declare no financial support was received for this work.

Data Availability

All data generated or analyzed during this study are included (in aggregated form) in this published article (and its supplementary information files).

Authors' Contributions

ACK and ASA contributed to the conceptualization, data collection, analysis, and writing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview Guide.
[PDF File, 97 KB - humanfactors_v13i1e83240_app1.pdf]

Multimedia Appendix 2
Table S1 – Participant Characteristics.
[DOCX File, 17 KB - humanfactors_v13i1e83240_app2.docx]

Multimedia Appendix 3
Table S2 – Mapping of subthemes to STS theory components.
[DOCX File, 28 KB - humanfactors_v13i1e83240_app3.docx]

Multimedia Appendix 4
Codebook.
[PDF File, 153 KB - humanfactors_v13i1e83240_app4.pdf]

Multimedia Appendix 5
Consent Form.
[PDF File, 174 KB - humanfactors_v13i1e83240_app5.pdf]

Multimedia Appendix 6
Table S3 – Themes, sub-themes, and associated codes.
[DOCX File, 23 KB - humanfactors_v13i1e83240_app6.docx]

References

1. Khalifa M, Albadawy M. AI in diagnostic imaging: revolutionising accuracy and efficiency. Computer Methods and Programs in Biomedicine Update 2024;5:100146. [doi: 10.1016/j.cmpbup.2024.100146]

XSL•FO
RenderX

2. Alowais SA, Alghamdi SS, Alsuhebany N, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ 2023 Sep 22;23(1):689. [doi: 10.1186/s12909-023-04698-z] [Medline: 37740191]
3. Tiwari A, Mishra S, Kuo TR. Current AI technologies in cancer diagnostics and treatment. Mol Cancer 2025 Jun 2;24(1):159. [doi: 10.1186/s12943-025-02369-9] [Medline: 40457408]
4. Kann BH, Hosny A, Aerts H. Artificial intelligence for clinical oncology. Cancer Cell 2021 Jul 12;39(7):916-927. [doi: 10.1016/j.ccell.2021.04.002] [Medline: 33930310]
5. Wang F, Beecy A. Implementing AI models in clinical workflows: a roadmap. BMJ Evid Based Med 2025 Sep 22;30(5):285-287. [doi: 10.1136/bmjebm-2023-112727] [Medline: 38914450]
6. Juluru K, Shih HH, Keshava Murthy KN, et al. Integrating Al algorithms into the clinical workflow. Radiol Artif Intell 2021 Nov;3(6):e210013. [doi: 10.1148/ryai.2021210013] [Medline: 34870216]
7. Verma H, Mlynar J, Schaer R, et al. Rethinking the role of AI with physicians in oncology: revealing perspectives from clinical and research workflows. 2023 Apr 19 Presented at: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems; Apr 23-28, 2023 URL: https://dl.acm.org/doi/proceedings/10.1145/3544548 [doi: 10.1145/3544548.3581506]
8. Madan Y, Perivolaris A, Adams-McGavin RC, Jung JJ. Clinician interaction with artificial intelligence systems: a narrative review. J Med Artif Intell 2025;8:22-22. [doi: 10.21037/jmai-24-279]
9. Aung YYM, Wong DCS, Ting DSW. The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare. Br Med Bull 2021 Sep 10;139(1):4-15. [doi: 10.1093/bmb/ldab016] [Medline: 34405854]
10. Rai A, Constantinides P, Sarker S. Editor's Comments: next-generation digital platforms: toward human–AI hybrids. MIS Q 2019;43(1):iii-iix. [doi: 10.25300/MISQ/2019/431E0]
11. Yin J, Ngiam KY, Teo HH. Role of artificial intelligence applications in real-life clinical practice: systematic review. J Med Internet Res 2021 Apr 22;23(4):e25759. [doi: 10.2196/25759] [Medline: 33885365]
12. Zhou S, Marklund H, Blaha O, et al. Deep learning assistance for the histopathologic diagnosis of Helicobacter pylori. Intell Based Med 2020 Nov;1-2:100004. [doi: 10.1016/j.ibmed.2020.100004]
13. Lambert SI, Madi M, Sopka S, et al. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. NPJ Digit Med 2023 Jun 10;6(1):111. [doi: 10.1038/s41746-023-00852-5] [Medline: 37301946]
14. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep 1;13(3):319-340. [doi: 10.2307/249008]
15. Bawack RE, Kala Kamdjoug JR. Adequacy of UTAUT in clinician adoption of health information systems in developing countries: the case of Cameroon. Int J Med Inform 2018 Jan;109:15-22. [doi: 10.1016/j.ijmedinf.2017.10.016] [Medline: 29195701]
16. Hwang Y, Al-Arabiat M, Shin DH. Understanding technology acceptance in a mandatory environment. Inf Dev 2016 Sep;32(4):1266-1283. [doi: 10.1177/0266666915593621]
17. Ratwani RM, Savage E, Will A, et al. A usability and safety analysis of electronic health records: a multi-center study. J Am Med Inform Assoc 2018 Sep 1;25(9):1197-1201. [doi: 10.1093/jamia/ocy088] [Medline: 29982549]
18. Sinsky C, Colligan L, Li L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. Ann Intern Med 2016 Dec 6;165(11):753-760. [doi: 10.7326/M16-0961] [Medline: 27595430]
19. Tai-Seale M, Olson CW, Li J, et al. Electronic health record logs indicate that physicians split time evenly between seeing patients and desktop medicine. Health Aff (Millwood) 2017 Apr 1;36(4):655-662. [doi: 10.1377/hlthaff.2016.0811] [Medline: 28373331]
20. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons for physicians not adopting clinical decision support systems: critical analysis. JMIR Med Inform 2018 Apr 18;6(2):e24. [doi: 10.2196/medinform.8912] [Medline: 29669706]
21. Fehr J, Citro B, Malpani R, Lippert C, Madai VI. A trustworthy AI reality-check: the lack of transparency of artificial intelligence products in healthcare. Front Digit Health 2024;6:1267290. [doi: 10.3389/fdgth.2024.1267290] [Medline: 38455991]
22. Marey A, Arjmand P, Alerab ADS, et al. Explainability, transparency and black box challenges of AI in radiology: impact on patient care in cardiovascular radiology. Egypt J Radiol Nucl Med 2024;55(1):183. [doi: 10.1186/s43055-024-01356-2]
23. Hassan M, Kushniruk A, Borycki E. Barriers to and facilitators of artificial intelligence adoption in health care: scoping review. JMIR Hum Factors 2024 Aug 29;11:e48633. [doi: 10.2196/48633] [Medline: 39207831]
24. Cooper R, Foster M. Sociotechnical systems. American Psychologist 1971;26(5):467-474. [doi: 10.1037/h0031539]
25. Walker GH, Stanton NA, Salmon PM, Jenkins DP. A review of sociotechnical systems theory: a classic concept for new command and control paradigms. Theor Issues Ergon Sci 2008 Nov;9(6):479-499. [doi: 10.1080/14639220701635470]
26. Appelbaum SH. Socio‐technical systems theory: an intervention strategy for organizational development. Manag Decis 1997 Aug;35(6):452-463. [doi: 10.1108/00251749710173823]
27. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view1. MIS Q 2003 Sep 1;27(3):425-478. [doi: 10.2307/30036540]
28. Rogers EM. Diffusion of Innovations, 4th edition: The Free Press; 1995.

XSL•FO
RenderX

29. Adeoye‐Olatunde OA, Olenik NL. Research and scholarly methods: semi‐structured interviews. J Am Coll Clin Pharm 2021 Oct;4(10):1358-1367 [FREE Full text] [doi: 10.1002/jac5.1441]
30. Fylan F. Semi-structured interviewing. In: A Handbook of Research Methods for Clinical and Health Psychology: Oxford University Press; 2005, Vol. 5:65-78. [doi: 10.1093/med:psych/9780198527565.003.0006]
31. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
32. Clarke V, Braun V. Thematic analysis. J Posit Psychol 2017 May 4;12(3):297-298. [doi: 10.1080/17439760.2016.1262613]
33. Braun V, Clarke V. To saturate or not to saturate? questioning data saturation as a useful concept for thematic analysis and sample-size rationales. Qual Res Sport Exerc Health 2021 Mar 4;13(2):201-216. [doi: 10.1080/2159676X.2019.1704846]
34. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
35. Alsheibani S, Messom C, Cheung Y. Re-thinking the competitive landscape of artificial intelligence. 2020 Jan 7 Presented at: Proceedings of the 53rd Hawaii International Conference on System Sciences; Jan 7, 2020. [doi: 10.24251/HICSS.2020.718]
36. Shah N. Validation and verification of artificial intelligence containing products across the regulated healthcare or medical device industries. IJSR 2024 Jul 5;13(7):66-71. [doi: 10.21275/ES24701081833]
37. Khumalo AC, Broden K, Bellström P, Alaqra AS. AI adoption in oncology: implementation process, barriers, and facilitators. 2024 Presented at: 47th Information Systems Research Seminar in Scandinavia (IRIS); Aug 11-14, 2024 URL: https://aisel.aisnet.org/iris2024/5/ [accessed 2025-08-01]
38. Youssef A, Pencina M, Thakur A, Zhu T, Clifton D, Shah NH. External validation of AI models in health should be replaced with recurring local validation. Nat Med 2023 Nov;29(11):2686-2687. [doi: 10.1038/s41591-023-02540-z] [Medline: 37853136]
39. Ahmad R, Kyratsis Y, Holmes A. When the user is not the chooser: learning from stakeholder involvement in technology adoption decisions in infection control. J Hosp Infect 2012 Jul;81(3):163-168. [doi: 10.1016/j.jhin.2012.04.014] [Medline: 22633278]
40. Barinov L, Jairaj A, Becker M, et al. Impact of data presentation on physician performance utilizing artificial intelligence-based computer-aided diagnosis and decision support systems. J Digit Imaging 2019 Jun;32(3):408-416. [doi: 10.1007/s10278-018-0132-5] [Medline: 30324429]
41. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3(1):17. [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
42. Lotter W, Hassett MJ, Schultz N, Kehl KL, Van Allen EM, Cerami E. Artificial intelligence in oncology: current landscape, challenges, and future directions. Cancer Discov 2024 May 1;14(5):711-726. [doi: 10.1158/2159-8290.CD-23-1199] [Medline: 38597966]
43. Kotter E, Ranschaert E. Challenges and solutions for introducing artificial intelligence (AI) in daily clinical workflow. Eur Radiol 2021 Jan;31(1):5-7. [doi: 10.1007/s00330-020-07148-2] [Medline: 32797308]
44. Abbas R, Michael K. In: Papagiannidis S, editor. Socio-Technical Theory: A Review: TheoryHub Book; 2023.
45. Xu W. Toward human-centered AI: a perspective from human-computer interaction. Interactions 2019;26(4):42-46. [doi: 10.1145/3328485]
46. Asan O, Choudhury A. Research trends in artificial intelligence applications in human factors health care: mapping review. JMIR Hum Factors 2021 Jun 18;8(2):e28236. [doi: 10.2196/28236] [Medline: 34142968]
47. Rony MKK, Parvin MR, Wahiduzzaman M, Debnath M, Bala SD, Kayesh I. "I Wonder if my Years of Training and Expertise Will be Devalued by Machines": concerns about the replacement of medical professionals by artificial intelligence. SAGE Open Nurs 2024;10:23779608241245220. [doi: 10.1177/23779608241245220] [Medline: 38596508]
48. Knop M, Weber S, Mueller M, Niehaves B. Human factors and technological characteristics influencing the interaction of medical professionals with artificial intelligence-enabled clinical decision support systems: literature review. JMIR Hum Factors 2022 Mar 24;9(1):e28639. [doi: 10.2196/28639] [Medline: 35323118]
49. Houssein EH, Gamal AM, Younis EMG, Mohamed E. Explainable artificial intelligence for medical imaging systems using deep learning: a comprehensive review. Cluster Comput 2025 Sep;28(7). [doi: 10.1007/s10586-025-05281-5]
50. Chua IS, Gaziel-Yablowitz M, Korach ZT, et al. Artificial intelligence in oncology: path to implementation. Cancer Med 2021 Jun;10(12):4138-4149. [doi: 10.1002/cam4.3935] [Medline: 33960708]
51. Momani A. Implications of artificial intelligence on health data privacy and confidentiality. arXiv. Preprint posted online on Jan 3, 2025. [doi: 10.48550/arXiv.2501.01639]
52. Gawankar S, Nair S, Pawar V, Vhatkar A, Chavan P. Patient privacy and data security in the era of AI-driven healthcare. Presented at: 2024 8th International Conference on Computing, Communication, Control and Automation (ICCUBEA); Aug 23-24, 2024. [doi: 10.1109/ICCUBEA61740.2024.10775004]
53. Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: a call for open science. Patterns (N Y) 2021 Oct 8;2(10):100347. [doi: 10.1016/j.patter.2021.100347] [Medline: 34693373]

XSL•FO
RenderX

54. Nazer LH, Zatarah R, Waldrip S, et al. Bias in artificial intelligence algorithms and recommendations for mitigation. PLOS Digit Health 2023 Jun;2(6):e0000278. [doi: 10.1371/journal.pdig.0000278] [Medline: 37347721]
55. Dwivedi YK, Rana NP, Jeyaraj A, Clement M, Williams MD. Re-examining the unified theory of acceptance and use of technology (UTAUT): towards a revised theoretical model. Inf Syst Front 2019 Jun;21(3):719-734. [doi: 10.1007/s10796-017-9774-y]
56. Attribution 4.0 international (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2026-03-28]
57. Chang A. UTAUT and UTAUT 2: a review and agenda for future research. WINNERS 2012;13(2):10. [doi: 10.21512/tw.v13i2.656]
58. Kim YJ, Choi JH, Fotso GMN. Medical professionals' adoption of AI-based medical devices: UTAUT model with trust mediation. J Open Innov Technol Mark Complex 2024 Mar;10(1):100220. [doi: 10.1016/j.joitmc.2024.100220]
59. Dingel J, Kleine AK, Cecil J, Sigl AL, Lermer E, Gaube S. Predictors of health care practitioners' intention to use AI-enabled clinical decision support systems: meta-analysis based on the unified theory of acceptance and use of technology. J Med Internet Res 2024 Aug 5;26:e57224. [doi: 10.2196/57224] [Medline: 39102675]
60. Heinrichs H, Kies A, Nagel SK, Kiessling F. Physicians' attitudes toward artificial intelligence in medicine: mixed methods survey and interview study. J Med Internet Res 2025 Aug 26;27:e74187. [doi: 10.2196/74187] [Medline: 40857713]

Abbreviations

AI: artificial intelligence
BI: behavioral intention
CT: computed tomography
EE: effort expectancy
EHR: electronic health record
FC: Facilitating Conditions
MRI: magnetic resonance imaging
PE: performance expectancy
SI: Social Influence
STS: sociotechnical systems
TA: thematic analysis
TAM: technology acceptance model
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by A Kushniruk; submitted 29.Aug.2025; peer-reviewed by L Seim, S Sankaranarayanan; revised version received 04.Feb.2026; accepted 19.Feb.2026; published 20.Apr.2026.

Please cite as:
Khumalo AC, Alaqra AS
Perceptions of Health Care Professionals on the Integration and Use of AI in Clinical Cancer Care: Interview Study
JMIR Hum Factors 2026;13:e83240
URL: https://humanfactors.jmir.org/2026/1/e83240
doi:10.2196/83240

© Akhona C Khumalo, Ala Sarah Alaqra. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 20.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Digital Health Literacy and Health Technology Readiness Among People With Epilepsy or Multiple Sclerosis: Cross-Sectional Study

Anna Vahteristo[1], MHSc, MSocSc; Virpi Jylhä[1,2], PhD; Hanna Kuusisto[1,3], MD, PhD

[1]Department of Health and Social Management, University of Eastern Finland, P.O. Box 1627, Kuopio, Finland

[2]Research Centre for Nursing Science and Social and Health Management, Wellbeing Services County of North Savo, Kuopio University Hospital, Kuopio, Finland

[3]Department of Neurology, Tampere University Hospital, Tampere, Finland

Corresponding Author:
Anna Vahteristo, MHSc, MSocSc
Department of Health and Social Management, University of Eastern Finland, P.O. Box 1627, Kuopio, Finland

Abstract

Background: Digital health services (DHS) are an increasingly integral part of health care services. Understanding users' abilities to engage with DHS is crucial to ensuring that health technology meets their needs. Assessing digital health literacy (DHL) and health technology readiness can help identify the strengths and weaknesses of DHL in different subgroups.

Objective: This study aimed to assess DHL and health technology readiness among people with epilepsy or multiple sclerosis (MS) and, accordingly, identify and categorize them into distinct subgroups or profiles. In addition, we aimed to investigate respondents' use of DHS in managing their chronic condition and differences in DHL and health technology readiness between DHS users and nonusers.

Methods: An electronic survey was distributed to people with epilepsy or MS with the help of patient organizations. The questionnaire included the Finnish version of the Readiness and Enablement Index for Health Technology. The subgroups of respondents were identified using k-means cluster analysis. Nonparametric tests were used to compare health technology readiness among identified subgroups.

Results: Respondents (N=289) had mid- to high-level scores in all the dimensions describing DHL and health technology readiness. A total of 4 distinct profiles emerged with different strengths and weaknesses in their DHL and health technology readiness. There was a significantly higher proportion of DHS users among the 2 profiles with the highest DHL, profile 1 (62/81, 76.5%) and profile 2 (59/80, 74.7%), compared with the profile with the lowest DHL, profile 4 (20/50, 40%; $P<.001$). In contrast, those with the lowest confidence in their DHL had higher emotional distress, reported lower confidence in the support from their health care providers, and had a smaller proportion of DHS users. In addition, the DHS users had significantly higher DHL levels in 6 of the 7 dimensions, as well as higher confidence in the support they received from their health care providers (mean 2.71, SD 0.72; $P=.01$) compared with nonusers (mean 2.42, SD 0.90) and in social support for health (mean 2.81, SD 0.71; $P=.02$), compared with nonusers (mean 2.54, SD 0.85).

Conclusions: Identifying subgroups with distinct profiles, characterized by different strengths and weaknesses in their DHL and health technology readiness, is crucial in ensuring the development of responsive and inclusive DHS to meet the needs of all users, particularly those requiring support in using DHS. In addition, the nonusers had lower confidence in the support they received from their health care provider than the users. Further research is needed to understand this difference.

(JMIR Hum Factors 2026;13:e85625) doi:10.2196/85625

KEYWORDS

digital health; digital health literacy; health technology readiness; epilepsy; multiple sclerosis

Introduction

Digital health interventions have proven to be clinically effective [1-9], cost-effective [2,3], and beneficial in the self-management of chronic diseases [4,10]. These interventions are closely associated with digital health literacy (DHL) [11-14], which refers to the ability to search for, understand, and evaluate health-related information from electronic sources and then use it to solve health-related problems [15]. Therefore, assessing DHL is crucial for elucidating the success of digital health interventions [16].

A high DHL level has been associated with higher education [17,18], increased use of digital health [17,19,20], and digital competence [21]. However, the relationship between DHL and chronic diseases remains complex. Some studies have noted low or inadequate DHL among patients with chronic diseases

XSL•FO
RenderX

[17,22,23], whereas others have found good levels of DHL [18,19]. Chronic diseases, including neurological disorders [24], are a growing cause of increased disability-adjusted life years and a burden on health care systems [25]. Epilepsy and multiple sclerosis (MS) are chronic preference-sensitive neurologic conditions [26] with multiple treatment options [27,28]. Treatment success depends on a combination of evidence-based medicine and patient-provided information [29] alongside individualized care plans and regular monitoring [30,31].

Both patients and health care providers have found digital health applications useful in the care and self-management of chronic conditions [32]. This trend is reflected in the increasing use of digital health technology for neurological care [33]. For example, the Finnish Neuro Registry is integrated into electronic patient records and includes a patient interface for people with MS and a digital seizure diary for those with epilepsy [34]. Similarly, electronic patient interfaces for epilepsy [35] and MS [36], as well as disease-specific digital care pathways for MS [37,38], have demonstrated promising results in improving care [35-37].

To ensure the effectiveness of digital health services (DHS), it is essential to elucidate the DHL of individuals with chronic diseases. In addition to disease-related background and sociodemographic characteristics, individuals' health technology readiness and emotional and mental well-being should be assessed to identify vulnerable subgroups [39]. It is also essential to understand more about individuals' strengths and weaknesses in terms of DHL and health technology use. Therefore, in this study, we assessed DHL and health technology readiness among people with epilepsy or MS and identified profiles of DHL and health technology readiness in this population. In addition, we investigated respondents' use of DHS in the management of their chronic condition and compared DHL and health technology readiness between the users and nonusers of DHS.

Methods

Recruitment

A cross-sectional survey was conducted between April 2022 and April 2023 to investigate DHL and health technology readiness among people with epilepsy or MS. The questionnaire was available in both Finnish and Swedish, the official languages of Finland. The survey targeted people diagnosed with epilepsy or MS, with no other eligibility criteria.

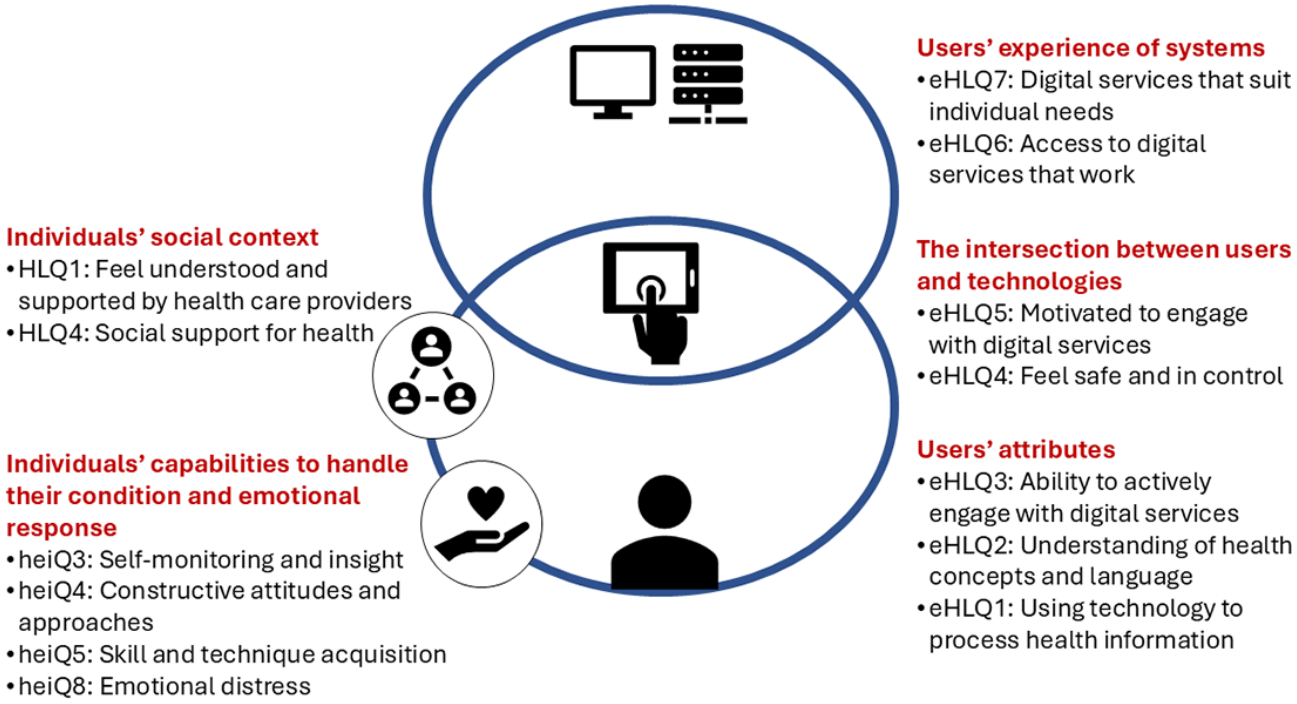
The first round of data collection began in April 2022 with an invitation to an online survey distributed by the Finnish Neuro Society to its members. The survey was closed in August 2022, when no further responses were received despite reminders. The second round involved paper questionnaires and the online survey. The latter was distributed by the Finnish Epilepsy Association to its members in November 2022 and by the Finnish Pensioners' Federation in January 2023, with the aim of reaching members diagnosed with epilepsy. The second round of data collection was closed in March 2023. The paper questionnaires were distributed to all eligible patients at the Department of Neurology of a university hospital in Finland. Additionally, they were distributed at events organized by the Finnish Epilepsy Association. The final paper questionnaires were distributed at the end of March, with a final response deadline in April 2023.

Instruments

The questionnaire was based on the Finnish version of the Readiness and Enablement Index for Health Technology (READHY-FIN) [40]. Readiness and Enablement Index for Health Technology (READHY) is an internationally validated instrument that measures levels of DHL and health technology readiness [41]. It consists of questions from 3 widely used instruments: the Health Education Impact Questionnaire [42], the Health Literacy Questionnaire [43], and the eHealth Literacy Questionnaire [44]. READHY-FIN comprises 65 items across 13 dimensions assessing 5 aspects of DHL and health technology readiness (Figure 1 [41]). These items are rated on a 4-point Likert scale ranging from 1="strongly disagree" to 4="strongly agree." The dimension of *emotional distress* was reverse-coded such that a higher score depicted less emotional distress. The score for each dimension and domain was calculated as the mean of the scores for the included items. If at least 50% of the items within a dimension were answered, the score for the dimension was manually calculated as an average of the items that were answered [41]. The READHY-FIN instrument was translated and culturally tested according to the protocol of the University of Swinburne [45] and validated in the Finnish population [40]. For the Swedish questionnaire, we used the Swedish version of READHY (READHY-SWE), which was translated by the developers of the original instrument [41].

XSL•FO
RenderX

Figure 1. Finnish version of the Readiness and Enablement Index for Health Technology [40] (adapted from Kayser et al [41], which is published under Creative Commons Attribution 4.0 International License [46]). eHLQ: eHealth Literacy Questionnaire; heiQ: Health Education Impact Questionnaire; HLQ: Health Literacy Questionnaire.



Additionally, the questionnaire included items describing the participants' sociodemographic and disease characteristics. The use of different service channels for managing their chronic condition included multiple choices. Of these, the use of DHS was defined as the use of any digital health service channel (patient interfaces or digital care pathways for MS or epilepsy) to manage their chronic condition; nonusers were defined as those who never used DHS to manage their chronic condition.

Statistical Analysis

The data analysis was performed with SPSS (version 29.0; IBM Corp). The data are presented as percentages and means. K-means clustering was used to identify profiles of respondents based on the dimensions of the READHY instrument. Cluster analysis can be used to identify subgroups or profiles of technology users based on the similarity of response patterns [47]. In previous studies using the READHY instrument, the number of clusters was 4 [41,48], 5 [39], or 6 [49]. Therefore, we tested our data using models ranging from 3 to 6 clusters. We assessed the appropriateness of the number of clusters by evaluating SDs and ANOVA for between-group differences, as well as pairwise comparisons between profiles on each READHY dimension. We chose the 4-cluster model because it produced the best between-profile variability and minimal within-profile variability.

To increase understanding of the dimensions and domains of READHY in different groups of respondents, we conducted comparisons based on respondents' characteristics, use of DHS, and the profiles identified in cluster analysis. We conducted further analyses using nonparametric Kruskal-Wallis and Mann-Whitney tests to compare different profiles and respondent characteristics, as well as to compare the characteristics of users and nonusers of DHS. The significance level was set at $P<.05$ [50]. Pairwise comparisons were performed when significant differences ($P<.05$) were indicated. Significance values were adjusted by the Bonferroni method for multiple tests to control for type I error.

Ethical Considerations

Ethics approval for the research was obtained from the Ethics Committee of Tampere University Hospital (R21057). The research complied with the World Medical Association's Declaration of Helsinki and was conducted in accordance with the principles of good research practice [51,52], the Finnish National Medical Research Act [53], and guidelines governing nonmedical research [54].

The respondents were informed that participation was voluntary and they had the opportunity to opt out at any time. Their anonymity was guaranteed, and no identifiable information was collected. Written informed consent was obtained from the respondents after they read the information sheet and privacy notice and before they started to fill out the anonymous questionnaire. Responding on behalf of a person was possible when the respondent was unable to do so themselves [54]. No compensation was provided to participants.

Results

Respondent Characteristics

The total number of participants was 289; their characteristics are summarized in Table 1. Their average age was 49.3 (SD 13.3) years, with most respondents (228/287, 79.4%) identifying as women. Two-thirds (188/289, 65.1%) of participants were diagnosed with MS, and two-thirds (188/288, 65.3%) were identified as users of DHS in managing their chronic condition.

XSL•FO
RenderX

Table . Characteristics of all respondents (N=289) and of the 4 Readiness and Enablement Index for Health Technology profiles.

Sociodemographic variables	All (N=289), n (%)	Profile 1 (n=81), n (%)	Profile 2 (n=80), n (%)	Profile 3 (n=78), n (%)	Profile 4 (n=50), n (%)
Gender (n=287)					
Woman	228 (79.4)	61 (76.3)	64 (80.0)	64 (82.1)	39 (79.6)
Man	59 (20.6)	19 (23.8)	16 (20.0)	14 (17.9)	10 (20.4)
Age groups (years; n=284)					
<30	20 (7.0)	6 (7.4)	6 (7.7)	5 (6.5)	3 (6.3)
30-39	47 (16.5)	18 (22.2)	10 (12.8)	12 (15.6)	7 (14.6)
40-54	114 (40.1)	30 (37.0)	33 (42.3)	26 (33.8)	25 (52.1)
55-64	66 (23.2)	16 (19.8)	18 (23.1)	21 (27.3)	11 (22.9)
>64	37 (13.0)	11 (13.6)	11 (14.1)	13 (16.9)	2 (4.2)
Diagnosed with chronic disease					
Epilepsy	101 (34.9)	25 (30.9)	24 (30.0)	32 (41.0)	20 (40.0)
Multiple sclerosis	188 (65.1)	56 (69.1)	56 (70.0)	46 (59.0)	30 (60.0)
Time from the diagnosis of chronic disease (years; n=287)					
<1	8 (2.8)	4 (4.9)	4 (5.0)	0 (0.0)	0 (0.0)
1-5	90 (31.4)	22 (27.2)	30 (37.5)	19 (25.0)	19 (38.0)
6-10	39 (13.6)	13 (16.0)	7 (8.8)	16 (21.1)	3 (6.0)
>10	150 (52.3)	42 (51.9)	39 (48.8)	41 (53.9)	28 (56.0)
Educational level (n=286)					
Primary and lower secondary education	22 (7.7)	6 (7.9)	5 (6.3)	5 (6.4)	6 (12.2)
Upper secondary or vocational education	88 (30.8)	16 (20.0)	21 (26.6)	26 (33.3)	25 (51.0)
Bachelor's-level degree	89 (31.1)	24 (30.0)	28 (35.4)	26 (33.3)	11 (22.4)
Master's-level or doctoral degree	87 (30.4)	34 (42.5)	25 (31.6)	21 (26.9)	7 (14.3)
Use of DHS[a] for the care of chronic disease (n=288)					
Active users (monthly or more often)	58 (20.1)	20 (24.7)	18 (22.5)	15 (19.5)	5 (10.0)
Occasional users (once or few times a year)	130 (45.1)	42 (51.9)	41 (51.3)	32 (41.6)	15 (30.0)
Nonusers	100 (34.7)	19 (23.5)	21 (26.3)	30 (40.0)	30 (60.0)

[a]DHS: digital health services.

Most respondents were evenly distributed across 3 of the 4 educational levels (Table 1): upper secondary school or vocational education (88/288, 30.8%), lower-level university degree (89/288, 31.1%), and upper-level university degree or doctoral degree (87/288, 30.4%).

The Use of Health Care Services and Digital Health for Managing Chronic Conditions

The use of various health care channels for managing respondents' chronic conditions is reported in Table 2. Face-to-face visits were the most common channel used by the respondents (271/288, 94.1%). However, nearly two-thirds (172/271, 63.5%) of those with face-to-face visits contacted their health care provider only once a year or less often.

Nearly half (91/188, 48.4%) of the users of DHS for managing their chronic condition had used DHS a few times a year, and one-quarter (47/188, 25.0%) of the users had used it monthly. There were also very active users, as 5.9% (11/188) used DHS weekly.

XSL•FO
RenderX

Table . Use of health care service channels.

	Face-to-face clinical visits (n=288), n (%)	Phone and video calls (n=289), n (%)	Digital health services (n=288), n (%)
Weekly or several times a week	3 (1.0)	1 (0.3)	11 (3.8)
Monthly	7 (2.4)	6 (2.1)	47 (16.3)
Few times a year	89 (30.9)	91 (31.5)	91 (31.6)
Once a year or less often	172 (59.7)	86 (29.8)	39 (13.5)
Not at all	17 (5.9)	105 (36.3)	100 (34.7)

DHL and Health Technology Readiness

The average level of respondents' DHL and health technology readiness, measured using the READHY dimensions, ranged from 2.61 to 3.19 (Multimedia Appendix 1). Respondents were highly confident in their *ability to actively engage with digital services* (mean 3.19, SD 0.70), *self-monitoring and insight* (mean 3.19, SD 0.42), and *constructive attitudes and approaches* (mean 3.00, SD 0.71). The dimensions in which respondents were least confident were *feeling understood and supported by health care providers* (mean 2.61, SD 0.82) and *social support for health* (mean 2.72, SD 0.76).

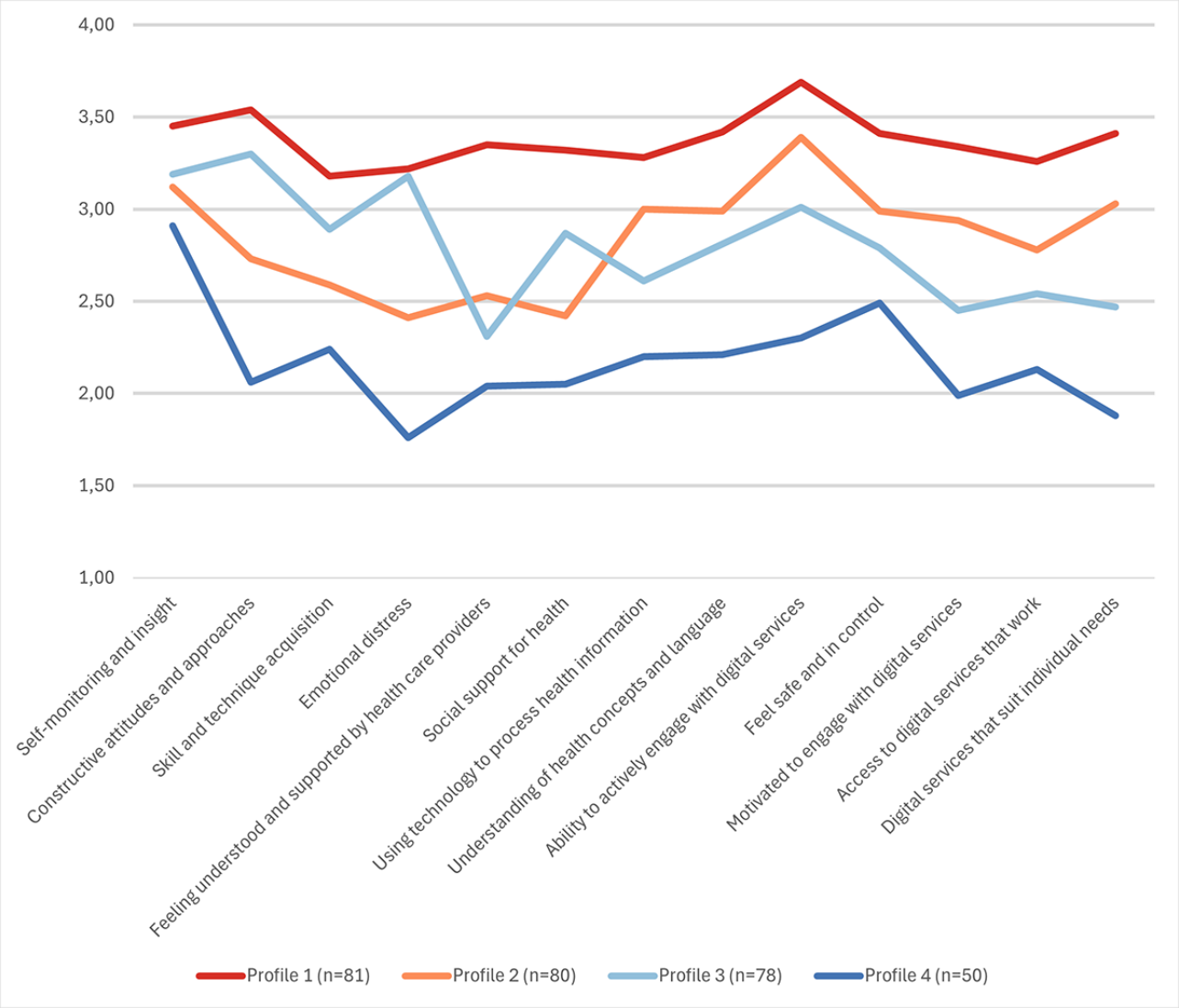
Respondents with higher educational levels had more confidence in their DHL and health technology readiness than those with lower educational levels. In pairwise comparisons, participants with master-level or doctoral degrees scored significantly higher than respondents with upper secondary school or vocational education in the domains of *users' attributes* ($P<.001$), *the intersection between users and technologies* ($P=.001$), *users' experience of systems* ($P<.001$), and *individuals' social context* ($P=.005$). Additionally, respondents with bachelor-level degrees had significantly higher scores than those with upper secondary or vocational education in the domains of *intersection between users and technologies* ($P=.02$) and *users' experience of systems* ($P=.02$).

Four READHY Profiles to Describe Subgroups of Individuals With Epilepsy or MS

We identified 4 profiles based on respondents' DHL and health technology readiness (Figure 2).

XSL•FO
RenderX

Figure 2. Four profiles of digital health literacy and health technology readiness.



Profile 1 (81/289, 28%) had a high level of DHL and health technology readiness. They scored high in all READHY dimensions, especially in the *ability to actively engage with digital services* (mean 3.69, SD 0.42) and *constructive attitudes and approaches* (mean 3.54, SD 0.44).

Profile 2 (80/289, 27.7%) had generally high and higher midlevel scores in DHL and health technology readiness. Their scores on DHL ranged between 3.39 (*ability to actively engage with digital services*) and 2.78 (*access to digital services that work*). They had high confidence in their *self-monitoring and insight* (mean 3.12, SD 0.33) but lower midlevel confidence in their abilities to handle *emotional distress* (mean 2.42, SD 0.54).

Profile 3 (78/289, 27%) had generally midlevel DHL and health technology readiness. They had midlevel scores in all DHL dimensions and in the social context. In addition, they had high-level scores in *constructive attitudes and approaches* (mean 3.30, SD 0.45) and in *self-monitoring and insight* (mean 3.19, SD 0.34).

Profile 4 (50/289, 17.3%) had low to lower midlevel DHL and health technology readiness. They did not score high on any READHY dimensions, with the lowest scores in their ability to handle *emotional distress* (mean 1.76, SD 0.51). They had lower midlevel confidence in both dimensions describing their confidence in their social context. However, they had higher midlevel scores on *self-monitoring and insight* (mean 2.91, SD 0.54).

Educational levels varied significantly among the 4 profiles (P<.001). Pairwise comparisons indicated that profile 4 had significantly lower educational levels than profiles 1 (P<.001) and 2 (P=.009).

DHL and Health Technology Readiness Among Users and Nonusers of DHS

Respondents who used DHS to manage their chronic condition reported higher levels of DHL and health technology readiness than nonusers (Table 3). The significant difference was evident in 6 of the 7 dimensions describing the respondents' DHL, as well as in both dimensions describing the respondents' social context. In addition, the difference was significant in *skill and technique acquisition*, one of the 4 dimensions describing individuals' capabilities to handle their condition and emotional response.

XSL•FO
RenderX

Table . Means of Readiness and Enablement Index for Health Technology domains among the users (n=188) and nonusers (n=100) of digital health services.

Dimensions	Users, mean (SD)	Nonusers, mean (SD)	*P* value[a]
Self-monitoring and insight	3.23 (0.36)	3.12 (0.50)	.20
Constructive attitudes and approaches	3.06 (0.66)	2.88 (0.79)	.08
Skill and technique acquisition	2.85 (0.50)	2.62 (0.66)	.002
Emotional distress	2.79 (0.72)	2.64 (0.82)	.18
Feeling understood and supported by health care providers	2.71 (0.72)	2.42 (0.90)	.01
Social support for health	2.81 (0.71)	2.54 (0.85)	.02
Using technology to process health information	2.94 (0.50)	2.62 (0.62)	<.001
Understanding of health concepts and language	3.03 (0.50)	2.73 (0.61)	<.001
Ability to actively engage with digital services	3.32 (0.61)	2.93 (0.79)	<.001
Feel safe and in control	3.04 (0.55)	2.84 (0.71)	.06
Motivated to engage with digital services	2.91 (0.58)	2.46 (0.68)	<.001
Access to digital services that work	2.84 (0.55)	2.54 (0.55)	<.001
Digital services that suit individual needs	2.95 (0.30)	2.48 (0.77)	<.001

[a]*P*<.05 is considered statistically significant.

Users of DHS were significantly younger (*P*=.03) than nonusers. This was the only statistically significant difference identified between the 2 groups in terms of their characteristics. However, we identified significant differences (*P*<.001) in the proportion of DHS users among different profiles. In pairwise comparisons, there were significantly fewer users of DHS among profile 4 compared with profile 1 (*P*<.001) and profile 2 (*P*=.001).

Discussion

Principal Findings and Comparison With Prior Research

The key finding of our study was that, although the respondents were fairly to highly confident in their overall DHL and health technology readiness, we identified 4 subgroups with distinct READHY profiles. Each profile had different strengths and weaknesses in their confidence in the 13 READHY dimensions.

Our findings align with those of previous research with the READHY instrument [48,49] and with research using other DHL instruments [17,21]. More specifically, our results indicate that among people with chronic conditions, there are differences in their DHL, suggesting that, in addition to their disease background, other factors are also associated with their DHL, which is consistent with earlier findings [21,55]. However, despite being highly confident in their ability and motivation to manage their condition and emotional response, the respondents reported lower confidence in receiving social support compared with previous research [48,49,56]. Consistent with this, Kuusisto et al [57] reported that people with MS felt that the information obtained from their health care professional was insufficient to facilitate their participation in shared decision-making regarding their care. This potential gap in patient-health care provider communication needs to be addressed to ensure that individuals with chronic conditions receive the support necessary to manage their condition. Further research is needed to determine whether this is a cultural phenomenon or a result of the health care system.

The 4 READHY profiles had different strengths and weaknesses. Profiles 1 and 2, with the highest confidence in their DHL, also had a significantly higher proportion of DHS users compared with profile 4 with the lowest level of DHL, indicating an association between DHL and the use of DHS for managing their chronic condition. However, although profile 1 had high confidence across all dimensions, profile 2 had higher *emotional distress* and lower confidence in their *skill and technique acquisition* to manage their condition and in the social support they received compared with profile 1. This implies that profile 2 could benefit from additional support for managing their health and emotional well-being. Profile 3, with midlevel scores, and profile 4, with low to lower midlevel scores on DHL dimensions and fewer users of DHS, indicated that they could represent individuals who need more support in using DHS. Additionally, profile 4—characterized by the highest emotional distress and the lowest confidence in the social support received compared with the other profiles—may represent individuals who require more support when using DHS to manage their chronic condition and who may therefore be more likely to prefer traditional face-to-face visits.

XSL•FO
RenderX

Our results also indicated that respondents with higher educational levels had significantly higher DHL, supporting the conclusion by Thorsen et al [39] that factors other than general sociodemographic characteristics and having a chronic disease are significant for the level of DHL and health technology readiness. The 4 profiles we identified increase this understanding of the factors related to DHL and health technology readiness. Identifying the various factors that support and hinder each profile's readiness to use health technology [53] is crucial for developing targeted health technology solutions [57]. Thus, although all 4 profiles reported good confidence in their ability to self-monitor their condition, promoting DHL remains necessary.

Our finding of two-thirds of the respondents having used DHS aligns with research on other chronic conditions [13,37,58]. In addition, DHS could be seen as a channel for frequent contact, as it was used several times a year, even monthly. In comparison, although most participants had face-to-face visits with their health care providers, these visits typically occurred only once or a few times per year. This finding indicates a need to identify the use cases that could be directed to DHS. Furthermore, user-centered development of DHS and the promotion of DHL may support the effective use of DHS.

Users of DHS had higher DHL and health technology readiness than nonusers. In addition, users also had significantly higher confidence in their *social support for health, feeling understood and supported by health care providers*, and *skill and technique acquisition* related to their capabilities to handle their condition. On the basis of these results, more research is required to investigate these relationships. Notably, despite the significant differences between users and nonusers of DHS, the nonusers also had reasonably good DHL and health technology readiness. This could be a result of variation in the level of digitalization among public health care providers, potentially resulting in inequality in access to and opportunities to use DHS.

Limitations and Recommendations for Future Research

This study has several limitations. First, despite the questionnaires being distributed via patient organizations and a neurology outpatient clinic, the sample size remained limited. Because the total number of individuals who received the questionnaire was unknown, it was not possible to calculate sample estimates. This may affect the generalizability of the findings. Second, because participation was voluntary, our sample may be subject to selection bias, potentially leading to an overrepresentation of DHS users. However, the questionnaires were also distributed in paper format through a neurology outpatient clinic, aiming to reach individuals who might not engage with online platforms. Third, prior computer use was not included as a background variable; thus, the association between DHS use and prior computer use could not be investigated.

In addition, this study was conducted in a single geographical area, reflecting the Finnish health care context, and caution must be exercised when generalizing the findings to other health care systems. Finally, the cross-sectional nature of the study precluded the determination of the evolution of DHL level, health technology readiness, and DHS use, necessitating additional longitudinal research. Further research is also required to elucidate whether the participants' perceptions of receiving inadequate social support regarding their health are a cultural phenomenon or a result of the health care system. Furthermore, additional research is needed on using the READHY instrument in the Finnish population and other health care systems.

Conclusions

Our key finding was the identification of 4 subgroups with distinct READHY profiles and different strengths and weaknesses in their DHL and health technology readiness. Understanding these strengths and weaknesses can help in the development of targeted interventions to encourage specific patient groups to take full advantage of DHS. However, this requires the development of processes to facilitate the identification of profiles. Additionally, adequate resourcing of different care channels is needed to meet users' needs.

We identified 2 profiles with high confidence in their DHL, both of which had a high proportion of DHS users and may possibly represent active users of various DHS. In contrast, the profiles with lower DHL could be more likely to use traditional face-to-face health care services and therefore benefit from the support in using DHS in the management of their chronic conditions. Notably, although all 4 profiles expressed generally high confidence in *self-monitoring and insight*, 3 reported low or medium confidence in *being understood and supported by health care providers*. This underlines the importance of ensuring that all individuals feel understood and supported by their health care providers, regardless of the health care channel they use. Future studies should further explore the phenomenon of not feeling supported by health care providers.

Acknowledgments

The authors would like to thank the Finnish Neuro Society, Finnish Epilepsy Association, and Finnish Pensioners' Federation, as well as the Department of Neurology of Tampere University Hospital, for their help in the dissemination of the survey.

Funding

This work was supported by the Strategic Research Council established within the Academy of Finland (grant 31213358415).

Data Availability

The data generated and/or analyzed during this study are not publicly available due to the data management plan of the project but are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mean scores of Readiness and Enablement Index for Health Technology domains and dimensions across participant profiles.
[DOCX File, 17 KB - humanfactors_v13i1e85625_app1.docx]

References

1. Eberle C, Stichling S. Clinical improvements by telemedicine interventions managing type 1 and type 2 diabetes: systematic meta-review. J Med Internet Res 2021 Feb 19;23(2):e23244. [doi: 10.2196/23244] [Medline: 33605889]
2. Del Pino R, Díez-Cirarda M, Ustarroz-Aguirre I, et al. Costs and effects of telerehabilitation in neurological and cardiological diseases: a systematic review. Front Med 2022 Nov 29;9:832229. [doi: 10.3389/fmed.2022.832229] [Medline: 36523783]
3. Gentili A, Failla G, Melnyk A, et al. The cost-effectiveness of digital health interventions: a systematic review of the literature. Front Public Health 2022 Aug 11;10:787135. [doi: 10.3389/fpubh.2022.787135] [Medline: 36033812]
4. Alaslawi H, Berrou I, Al Hamid A, Alhuwail D, Aslanpour Z. Diabetes self-management apps: systematic review of adoption determinants and future research agenda. JMIR Diabetes 2022 Jul 28;7(3):e28153. [doi: 10.2196/28153] [Medline: 35900826]
5. Eberle C, Stichling S. Telemetric interventions offer new opportunities for managing type 1 diabetes mellitus: systematic meta-review. JMIR Diabetes 2021 Mar 16;6(1):e20270. [doi: 10.2196/20270] [Medline: 33724201]
6. Kraef C, van der Meirschen M, Free C. Digital telemedicine interventions for patients with multimorbidity: a systematic review and meta-analysis. BMJ Open 2020 Oct 13;10(10):e036904. [doi: 10.1136/bmjopen-2020-036904] [Medline: 33051232]
7. Kuan PX, Chan WK, Fern Ying DK, et al. Efficacy of telemedicine for the management of cardiovascular disease: a systematic review and meta-analysis. Lancet Digit Health 2022 Sep;4(9):e676-e691. [doi: 10.1016/S2589-7500(22)00124-8] [Medline: 36028290]
8. Lee JJ, Abdul Aziz A, Chan ST, et al. Effects of mobile health interventions on health-related outcomes in older adults with type 2 diabetes: a systematic review and meta-analysis. J Diabetes 2023 Jan 17;15(1):47-57. [doi: 10.1111/1753-0407.13346] [Medline: 36649940]
9. Ma Y, Zhao C, Zhao Y, et al. Telemedicine application in patients with chronic disease: a systematic review and meta-analysis. BMC Med Inform Decis Mak 2022 Apr 19;22(1):105. [doi: 10.1186/s12911-022-01845-2] [Medline: 35440082]
10. Renzi E, Baccolini V, Migliara G, et al. The impact of eHealth interventions on the improvement of self-care in chronic patients: an overview of systematic reviews. Life (Basel) 2022 Aug 17;12(8):1253. [doi: 10.3390/life12081253] [Medline: 36013432]
11. Choi M. Factors associated with eHealth use among community dwelling older adults. Int J Nurs Pract 2022 Dec;28(6):e13092. [doi: 10.1111/ijn.13092] [Medline: 35977421]
12. Ramstad KJ, Brørs G, Pettersen TR, et al. eHealth technology use and eHealth literacy after percutaneous coronary intervention. Eur J Cardiovasc Nurs 2023 Jul 19;22(5):472-481. [doi: 10.1093/eurjcn/zvac087] [Medline: 36190843]
13. Tang J, Howell M, Lee VW, et al. Patients' perspectives, factors, and patterns of eHealth use in kidney transplant recipients. Kidney Int Rep 2023 Jan 20;8(4):727-736. [doi: 10.1016/j.ekir.2023.01.007] [Medline: 37069969]
14. Wong DK, Cheung MK. Online health information seeking and ehealth literacy among patients attending a primary care clinic in Hong Kong: a cross-sectional survey. J Med Internet Res 2019 Mar 27;21(3):e10831. [doi: 10.2196/10831] [Medline: 30916666]
15. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9. [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
16. Zangger G, Bricca A, Liaghat B, et al. Benefits and harms of digital health interventions promoting physical activity in people with chronic conditions: systematic review and meta-analysis. J Med Internet Res 2023 Jul 6;25:e46439. [doi: 10.2196/46439] [Medline: 37410534]
17. Shiferaw KB, Tilahun BC, Endehabtu BF, Gullslett MK, Mengiste SA. E-health literacy and associated factors among chronic patients in a low-income country: a cross-sectional survey. BMC Med Inform Decis Mak 2020 Aug 6;20(1):181. [doi: 10.1186/s12911-020-01202-1] [Medline: 32762745]
18. Redfern J, Coorey G, Mulley J, et al. A digital health intervention for cardiovascular disease management in primary care (CONNECT) randomized controlled trial. NPJ Digit Med 2020 Sep 10;3:117. [doi: 10.1038/s41746-020-00325-z] [Medline: 32964140]
19. Ernsting C, Stühmann LM, Dombrowski SU, Voigt-Antons JN, Kuhlmey A, Gellert P. Associations of health app use and perceived effectiveness in people with cardiovascular diseases and diabetes: population-based survey. JMIR Mhealth Uhealth 2019 Mar 28;7(3):e12179. [doi: 10.2196/12179] [Medline: 30920383]
20. Knitza J, Simon D, Lambrecht A, et al. Mobile health usage, preferences, barriers, and eHealth literacy in rheumatology: patient survey study. JMIR Mhealth Uhealth 2020 Aug 12;8(8):e19661. [doi: 10.2196/19661] [Medline: 32678796]

XSL•FO
RenderX

21. Lee J, Tak SH. Factors associated with eHealth literacy focusing on digital literacy components: a cross-sectional study of middle-aged adults in South Korea. Digit Health 2022 May 20;8:20552076221102765. [doi: 10.1177/20552076221102765] [Medline: 35615270]
22. Schrauben SJ, Appel L, Rivera E, et al. Mobile health (mHealth) technology: assessment of availability, acceptability, and use in CKD. Am J Kidney Dis 2021 Jun;77(6):941-950. [doi: 10.1053/j.ajkd.2020.10.013] [Medline: 33309860]
23. Wu Y, Wen J, Wang X, et al. Associations between e-health literacy and chronic disease self-management in older Chinese patients with chronic non-communicable diseases: a mediation analysis. BMC Public Health 2022 Nov 29;22(1):2226. [doi: 10.1186/s12889-022-14695-4] [Medline: 36447176]
24. GBD 2016 Neurology Collaborators. Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2019 May;18(5):459-480. [doi: 10.1016/S1474-4422(18)30499-X] [Medline: 30879893]
25. World health statistics 2023: monitoring health for the SDGs, sustainable development goals. World Health Organization. 2023. URL: https://www.who.int/publications/i/item/9789240074323 [accessed 2026-02-12]
26. Heesen C, Solari A. Editorial: Shared decision-making in neurology. Front Neurol 2023 Jun 6;14:1222433. [doi: 10.3389/fneur.2023.1222433] [Medline: 37346163]
27. van der Horst DE, Garvelink MM, Bos WJ, Stiggelbout AM, Pieterse AH. For which decisions is shared decision making considered appropriate? - A systematic review. Patient Educ Couns 2023 Jan;106:3-16. [doi: 10.1016/j.pec.2022.09.015] [Medline: 36220675]
28. Zhu M, Dong D, Lo HH, Wong SY, Mo PK, Sit RW. Patient preferences in the treatment of chronic musculoskeletal pain: a systematic review of discrete choice experiments. Pain 2023 Apr 1;164(4):675-689. [doi: 10.1097/j.pain.0000000000002775] [Medline: 36149784]
29. Hämäläinen P, Viitala M, Kuusisto H, Ruutiainen J, Soilu-Hänninen M. MyMS: an interface for patient-reported outcomes for Finnish individuals with multiple sclerosis. Int J MS Care 2024 Oct;26(Q4):273-280. [doi: 10.7224/1537-2073.2023-082] [Medline: 39386272]
30. Ward M, Goldman MD. Epidemiology and pathophysiology of multiple sclerosis. Continuum (Mount Lawley) 2022 Aug 1;28(4):988-1005. [doi: 10.1212/CON.0000000000001136] [Medline: 35938654]
31. Thijs RD, Surges R, O'Brien TJ, Sander JW. Epilepsy in adults. Lancet 2019 Feb 16;393(10172):689-701. [doi: 10.1016/S0140-6736(18)32596-0] [Medline: 30686584]
32. Arsad FS, Syed Soffian SS, Megat Kamaruddin PS, et al. The impact of eHealth applications in healthcare intervention: a systematic review. J Health Res 2022;37(3):178-189. [doi: 10.56808/2586-940X.1020]
33. Masanneck L, Gieseler P, Gordon WJ, Meuth SG, Stern AD. Evidence from ClinicalTrials.gov on the growth of Digital Health Technologies in neurology trials. NPJ Digit Med 2023 Feb 10;6:23. [doi: 10.1038/s41746-023-00767-1]
34. Kuusisto H, Hämäläinen P, Nurmi H, Kälviäinen R, Soilu-Hänninen M. Patient generated digital health data: an example from the Finnish Neuro Registry. In: Envisioning the Future of Health Informatics and Digital Health: IOS Press; 2025:174-178. [doi: 10.3233/SHTI250072]
35. Fitzsimons M, Power K, McCrea Z, et al. Democratizing epilepsy care: utility and usability of an electronic patient portal. Epilepsy Behav 2021 Sep;122:108197. [doi: 10.1016/j.yebeh.2021.108197] [Medline: 34273742]
36. Voigt I, Benedict M, Susky M, et al. A digital patient portal for patients with multiple sclerosis. Front Neurol 2020 May 22;11:400. [doi: 10.3389/fneur.2020.00400] [Medline: 32670174]
37. Vesinurm M, Maunula A, Olli P, et al. Effects of a digital care pathway for multiple sclerosis: observational study. JMIR Hum Factors 2024 Aug 7;11:e51872. [doi: 10.2196/51872] [Medline: 39110966]
38. Wenk J, Voigt I, Inojosa H, Schlieter H, Ziemssen T. Building digital patient pathways for the management and treatment of multiple sclerosis. Front Immunol 2024 Feb 15;15:1356436. [doi: 10.3389/fimmu.2024.1356436] [Medline: 38433832]
39. Thorsen IK, Rossen S, Glümer C, Midtgaard J, Ried-Larsen M, Kayser L. Health technology readiness profiles among Danish individuals with type 2 diabetes: cross-sectional study. J Med Internet Res 2020 Sep 15;22(9):e21195. [doi: 10.2196/21195] [Medline: 32930669]
40. Jylhä V, Turja T. Health technology readiness: the translation of READHY-FIN questionnaire. FinJeHeW 2023;15(2):219-229. [doi: 10.23996/fjhw.126867]
41. Kayser L, Rossen S, Karnoe A, et al. Development of the multidimensional readiness and enablement index for health technology (READHY) tool to measure individuals' health technology readiness: initial testing in a cancer rehabilitation setting. J Med Internet Res 2019 Feb 12;21(2):e10377. [doi: 10.2196/10377] [Medline: 30747717]
42. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 2013 Jul 16;13:658. [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]
43. Osborne RH, Elsworth GR, Whitfield K. The Health Education Impact Questionnaire (heiQ): an outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. Patient Educ Couns 2007 May;66(2):192-201. [doi: 10.1016/j.pec.2006.12.002] [Medline: 17320338]

XSL•FO
RenderX

44. Kayser L, Karnoe A, Furstrand D, et al. A multidimensional tool based on the eHealth literacy framework: development and initial validity testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36. [doi: 10.2196/jmir.8371] [Medline: 29434011]
45. Hawkins M, Osborne RH. Translation Integrity Procedure (TIP) For the Translation and Cultural Adaptation of Psychometric Questionnaires: Arts and Design, Swinburne University of Technology; 2019.
46. CC BY 4.0: attribution 4.0 international. Creative commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2026-02-24]
47. Oliveira FK, Oliveira M, Gomes AS, Queiros LM. Identifying user profiles from statistical grouping methods. J Inf Syst Eng Manag 2018;3(1):10. [doi: 10.20897/jisem.201806]
48. Rossen S, Kayser L, Vibe-Petersen J, Ried-Larsen M, Christensen JF. Technology in exercise-based cancer rehabilitation: a cross-sectional study of receptiveness and readiness for e-Health utilization in Danish cancer rehabilitation. Acta Oncol 2019 May;58(5):610-618. [doi: 10.1080/0284186X.2018.1562213] [Medline: 30698060]
49. Nielsen AS, Hanna L, Larsen BF, Appel CW, Osborne RH, Kayser L. Readiness, acceptance and use of digital patient reported outcome in an outpatient clinic. Health Informatics J 2022;28(2):14604582221106000. [doi: 10.1177/14604582221106000] [Medline: 35658693]
50. Grove SK, Gray JR, Burns N. The Practice of Nursing Research: Appraisal, Synthesis, and Generation of Evidence, 7th edition: Elsevier Health Sciences; 2012.
51. The European Code of Conduct for Research Integrity – Revised Edition: ALLEA; 2023. [doi: 10.26356/ECOC]
52. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). European Union. URL: http://data.europa.eu/eli/reg/2016/679/2016-05-04 [accessed 2025-08-24]
53. Medical Research Act. Ministry of Social Affairs and Health, Finland. URL: https://www.finlex.fi/api/media/statute-foreign-language-translation/688701/mainPdf/main.pdf?timestamp=1999-04-08T21%3A00%3A00.000Z [accessed 2026-02-12]
54. The ethical principles of research with human participants and ethical review in the human sciences in Finland. : Finnish National Board on Research Integrity TENK; 2019 URL: https://tenk.fi/sites/default/files/2021-01/Ethical_review_in_human_sciences_2020.pdf [accessed 2025-08-24]
55. Madrigal L, Escoffery C. Electronic health behaviors among US adults with chronic disease: cross-sectional survey. J Med Internet Res 2019 Mar 5;21(3):e11240. [doi: 10.2196/11240] [Medline: 30835242]
56. Terp R, Kayser L, Lindhardt T. Older patients' competence, preferences, and attitudes toward digital technology use: explorative study. JMIR Hum Factors 2021 May 14;8(2):e27005. [doi: 10.2196/27005] [Medline: 33988512]
57. Kuusisto H, Apila S, Saranto K. Information provision and quality. A pilot study on shared decision-making in multiple sclerosis. Stud Health Technol Inform 2022 Jun 29;295:179-182. [doi: 10.3233/SHTI220691] [Medline: 35773837]
58. Tosun AT, Isiklar C, Yildirim M, Coskunsu DK. e-Health literacy status of individuals with multiple sclerosis in Turkey. Telemed J E Health 2024 Apr;30(4):e1166-e1171. [doi: 10.1089/tmj.2023.0151] [Medline: 37851986]

Abbreviations

DHL: digital health literacy
DHS: digital health services
eHLF: eHealth Literacy Framework
MS: multiple sclerosis
READHY: Readiness and Enablement Index for Health Technology
READHY-FIN: Finnish version of the Readiness and Enablement Index for Health Technology
READHY-SWE: Swedish version of the Readiness and Enablement Index for Health Technology

Edited by A Kushniruk; submitted 10.Oct.2025; peer-reviewed by D Vlahovic, S Bergh; accepted 02.Jan.2026; published 05.Mar.2026.

Please cite as:
Vahteristo A, Jylhä V, Kuusisto H
Digital Health Literacy and Health Technology Readiness Among People With Epilepsy or Multiple Sclerosis: Cross-Sectional Study
JMIR Hum Factors 2026;13:e85625
URL: https://humanfactors.jmir.org/2026/1/e85625
doi:10.2196/85625

XSL•FO
RenderX

© Anna Vahteristo, Virpi Jylhä, Hanna Kuusisto. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 5.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Digital Mental Health Promotion Services for Youth: A Qualitative Study of Help-Seeking Through Mindhelper.dk

Amalie Oxholm Kusier, MA; Caroline Høier Dalsgaard, MSc; Sofie Have Hoffmann, MSc; Lau Caspar Thygesen, PhD; Anna Paldam Folker, PhD

National Institute of Public Health, The Faculty of Health Sciences, University of Southern Denmark, Studiestræde 6, Copenhagen, Denmark

Corresponding Author:
Amalie Oxholm Kusier, MA
National Institute of Public Health, The Faculty of Health Sciences, University of Southern Denmark, Studiestræde 6, Copenhagen, Denmark

Abstract

Background: Young people increasingly experience mental health challenges and often turn to the internet for support. Self-guided digital mental health promotion services have become widely used resources for youth seeking help and guidance. These platforms offer accessible, anonymous support, yet little is known about the concerns young people articulate when engaging with them.

Objective: This study aims to examine inquiries submitted to the digital letter box of Mindhelper.dk, Denmark's most widely used digital mental health promotion service. Using qualitative analysis, the study aims to identify recurring themes in young people's mental health concerns, explore gender differences in engagement, and generate insights to inform the development of more relevant and targeted digital self-help interventions.

Methods: Using an inductive thematic approach informed by a constructivist-grounded theory coding framework, this study analyzes 2523 inquiries submitted to the Mindhelper digital letter box between March 2016 and August 2023. The dataset provides unsolicited first-person accounts from young people in moments of emotional vulnerability, offering insights into how mental health concerns are articulated in naturalistic settings.

Results: The analysis identifies 17 recurring themes reflecting the mental health challenges young people seek help for. These were grouped into 3 overarching analytical categories: Social Relationships and Social Contexts, Emotional Life, and Body and Illness, with the first 2 dominating the material. Prominent themes included Sociality, Love Life, Unease, Self-doubt and Insecurity, and Seeking Support. Across genders, inquiries frequently focused on social relationships, particularly Sociality and Love Life. However, girls were markedly overrepresented among users, while only minor gender differences were observed in the distribution of themes.

Conclusions: The findings suggest that young people's mental health concerns are closely tied to everyday developmental and relational challenges rather than severe psychopathology alone. Digital letter box services may capture early expressions of distress that might not otherwise reach formal support systems. This highlights the preventive potential of such services and the value of using self-initiated digital data to inform the development of relevant digital mental health support and better understand how young people articulate and act on emerging mental health concerns.

(JMIR Hum Factors 2026;13:e91017) doi:10.2196/91017

KEYWORDS

digital mental health; youth mental health; help-seeking behavior; gender differences; digital letter box; mental health promotion; qualitative analysis

Introduction

A substantial proportion of young people in Denmark report poor mental health. This is particularly pronounced among young women, with 34.8% of women aged 16 - 24 years reporting poor mental health in the 2023 Danish National Health Survey, compared with 16.0% among men in the same age group [1]. A recent assessment of well-being in Denmark further indicates that more than 50% of young women aged 16 - 24 years report high levels of stress, compared with 31,6% among young men in the same age group [2]. This challenge is not unique to Denmark; globally, youth mental health trends indicate a substantial increase in issues such as anxiety, psychological distress, self-harm, suicide, and depressive symptoms [3,4]. According to the Lancet Psychiatry Commission, mental ill health now accounts for at least 45% of the overall burden of disease among individuals aged 10 - 24 years, and suicide remains the fourth leading cause of death among those aged 15 - 29 years worldwide [5]. These figures highlight the

XSL•FO
RenderX

significant global impact of mental health problems among young people.

Adolescence and emerging adulthood (the latter typically defined as ages 18 - 25 years) represent particularly vulnerable periods due to rapid physical, cognitive, emotional, and social development [6]. Poor mental health during these formative years can have severe long-term consequences [7,8]. For example, one study found that adolescents with poor mental well-being had significantly higher risks of mental illness, self-harm and suicide attempts, alcohol and substance abuse, interpersonal violence, and all-cause mortality than those with better mental well-being [9]. Emerging adulthood is also a critical window for the onset of mental disorders, with 63% - 75% of mental illnesses first appearing before the age of 25 years [5]. This emphasizes the need for preventive strategies during this developmental stage.

In response to these challenges, digital mental health interventions have emerged as a promising approach [5,10,11]. Digital mental health interventions refer to mental health care delivered via digital technologies, and they can take a variety of forms depending on the target population, purpose, and content of the intervention, the technology used, and the degree of human interaction involved [12]. Research indicates that such interventions have considerable potential to promote mental health, prevent distress, and support the treatment of mental illness. As young people are among the highest users of the internet and digital technologies, these interventions offer advantages such as accessibility, flexibility, and broad reach, making them well suited to address some of the limitations observed in traditional youth mental health services [5]. However, the potential of digital mental health services to promote well-being among young people is not without challenges. Evidence suggests that the level of engagement with digital interventions is critical for achieving positive mental health outcomes [13]. Yet, engagement is influenced by a range of facilitators and barriers that determine whether young people adopt and continue using these services [14].

To ensure that digital mental health support services are effective, it is essential that they are designed with content perceived as relevant and useful by young people. A systematic review by Ho et al [15] identified perceived fit and perceived usefulness as key facilitators of engagement with web-based mental health interventions for young people. Perceived usefulness was enhanced when interventions provided informative and actionable content, while content that was repetitive, overly simple, unspecific, or already known acted as barriers. Similarly, perceived fit was strengthened when interventions were age-appropriate, relevant, and culturally sensitive. Personalization—such as individualized interaction and feedback—further improved perceived fit, whereas lack of personalization and irrelevance (eg, cultural mismatch) reduced engagement. Findings from Borghouts et al [16] reinforce these conclusions, emphasizing that participants were more likely to engage when they perceived the intervention to be useful and a good fit for them. Furthermore, a review has shown that research on youth digital mental health interventions is primarily based on young women, leaving a significant gap in knowledge about how young men engage with and experience these interventions [17]. Understanding these gendered aspects is essential to ensure that digital mental health services are inclusive and meet the needs of all young people.

Building on this, this study seeks to contribute to the field by providing insight into the issues young people seek help for when experiencing mental distress. This is achieved by examining inquiries submitted to a digital letter box in Denmark's most widely used digital mental health promotion services, Mindhelper.dk. Specifically, the study aims to (1) identify recurring themes in young people's inquiries about mental health and well-being, (2) examine how gender may shape these experiences in the context of engaging with a self-guided digital platform, and (3) generate insights to inform the design of more relevant and targeted digital self-help interventions for young people. Through qualitative analysis of letter box inquiries, this study offers valuable knowledge about the mental health challenges faced by Danish youth and their engagement with digital support services.

Methods

This section presents the methodological framework of the study. To situate the analysis, we first describe the empirical context, followed by a detailed account of the data and ethical considerations. Finally, we outline the analytical strategy applied in the study.

Empirical Context

Mindhelper.dk was launched in 2016 and is now the largest public digital youth mental health–promoting service in Denmark with approximately 1 million visitors a year. Mindhelper is a self-guided service that aims to provide young people in Denmark with knowledge, advice, and tools on a wide range of mental health and well-being topics, including anxiety, depression, dating, and friendship problems, self-harm, and suicidal thoughts. The platform offers online courses, papers, videos, and guidance on where to seek help offline, as well as a digital letter box for personalized advice. In addition to its website, Mindhelper also provides support through an app with a focus on reducing stress, pressure, or anxiety.

Mindhelper primarily targets young people aged 13 - 25 years, but usage data show that the platform is also widely accessed by children younger than 13 years, young adults older than 25 years, and adults such as parents and professionals (eg, schoolteachers and student counselors). All services at Mindhelper are free, self-guided, and fully anonymous, requiring no user account.

Mindhelper was developed through a collaboration between the Centre for Digital Psychiatry (CEDIP) and 4 Danish municipalities. CEDIP oversees ongoing operations and content development, which is cocreated with young people through a youth network and a panel of 12 members who contribute to testing, feedback, and idea development. Since 2019, Mindhelper has been permanently funded by Denmark's 5 regions, and in 2025, national government funding was added.

XSL•FO
RenderX

The Data Material

This study explores young people's inquiries to the Mindhelper letter box over a period of nearly 8 years, covering inquiries submitted between 2016 and 2023. Since 2016, young people have been able to write anonymously to the digital letter box and receive a response within 10 days from psychology students employed at CEDIP. All inquiries and responses are published on Mindhelper.dk, allowing users to read and benefit from each other's experiences. The only editing occurs when personally identifiable information is included, in which case the inquiry is anonymized. There is no censorship of content. Inquiries can be accessed via free-text search using subject-specific keywords or browsed through thematic categories available on the website.

Over time, a few changes have been made to the letter box. When the letter box was launched, young people were required to state their gender (male or female) and age (13 - 20 years) when submitting inquiries based on set categories. In 2021, the gender requirement was removed, and, in summer 2024, the age category was expanded to 13 - 25 years. When submitting inquiries, young people can write up to 1200 characters, and the date and time are automatically entered.

The letter box archive represents a unique empirical dataset, offering rare access to anonymous, first-person accounts from young people written in moments of emotional vulnerability. Unlike retrospective interviews or researcher-generated data, these inquiries are unsolicited and self-initiated, providing immediate and authentic expressions of mental health concerns. This makes the material particularly valuable for understanding youth experiences of distress and well-being as they unfold. At the same time, it is important to note that, as a digitally mediated dataset, the material is influenced by the platform's affordances and constraints, including its written format, character limits, and the characteristics of its user base.

The dataset analyzed includes all inquiries submitted to the letter box between March 4, 2016, and August 7, 2023, comprising a total of 2523 inquiries. The data extraction was conducted by CEDIP and included only the inquiries from the young people–not the associated responses–to ensure that the analysis was not influenced by the content of the replies. Additionally, the analysis did not take into account Mindhelper's thematic categorization of the inquiries.

Ethical Considerations

This study is registered and approved by the University of Southern Denmark in accordance with the Data Protection Regulation and the General Data Protection Regulation (EU) 2016/679 and has additionally been approved by the University's Research Ethics Committee (approval ID 23/24293). The analyzed material consists of anonymous inquiries submitted to the letter box of a digital mental health promotion platform. No compensation was provided, as users submitted inquiries voluntarily as part of their use of the platform. All inquiries were anonymized before researcher access and were already publicly available. Although users consented to having their posts published, they were not informed that the content might later be used for research, which raises an ethical consideration about secondary use of data. We assess the risk of harm as minimal because the data were anonymized and originally shared in a public forum. Obtaining informed consent was not feasible, as submissions were anonymous.

Protection of Anonymity in Quotations

However, several measures were applied to balance the use of illustrative quotations with the protection of contributor anonymity. First, all quotations were translated from Danish to English using an interpretive, meaning-based translation process rather than a literal translation. This constituted an additional analytical transformation and reduced traceability. Second, the majority of quotations were shortened to retain analytical relevance while avoiding unnecessary detail that could increase identifiability. In addition, quotations were deliberately selected to illustrate broader themes, shared feelings, or general experiences reflected across multiple inquiries, rather than highly specific or personally identifying narratives that also appeared in the material. Third, although inquiries on the platform were publicly associated with unique identification numbers, these identifiers were deliberately omitted from all quotations and descriptions in the manuscript to prevent direct linkage to the original submissions. Overall, this approach was consistent with principles for the responsible use of publicly available, anonymized data.

Analysis Strategy

Coding Framework

The analysis followed an inductive thematic approach, informed by Charmaz's constructivist grounded theory as a pragmatic coding framework, to examine patterns of meaning across the 2523 inquiries [18-20]. While guided by constructivist grounded theory principles, the aim was not to generate a formal grounded theory or explanatory model but to use iterative coding strategies to develop analytically grounded themes and categories. Coding followed 2 main phases described by Charmaz: initial coding, involving close reading of each word, line, or segment of data to remain open to multiple possible interpretations, and focused coding, which uses the most significant or frequent initial codes to sort, synthesize, and organize larger segments of the dataset [19]. These phases were operationalized across 3 coding rounds: round 1 (initial coding), and rounds 2 and 3 (focused coding), as described in the following sections.

Given the size of the dataset, we used NVivo (Lumivero), a qualitative analysis software, to facilitate data storage, coding, and comparison [18]. To ensure feasibility, the first author (AOK) conducted a pilot coding on a small sample (n=49 inquiries), focusing on the technical aspects of accessing and organizing the material and enabling comparison across the dataset. This confirmed NVivo as a suitable tool for managing the data. The test coding did not involve engagement with the thematic content, ensuring consistency with the inductive and grounded approach applied in the study.

Coding Round 1: Initial Open Coding

All inquiries were read and coded by the first author (AOK). The aim of the initial open coding round was to identify topics, descriptions, and statements that contributed to a nuanced understanding of why the young person contacted the letter box.

XSL•FO
RenderX

Following the grounded theory principle of line-by-line coding, each relevant segment was assigned a code, and a single inquiry could contain multiple codes, as short statements or excerpts often addressed several points [19]. The coding process remained close to the data, using simple and precise codes, as outlined in the codebook (Multimedia Appendix 1). Importantly, coding was kept open, allowing new codes to be created as new perspectives emerged. In addition to assigning short code titles, a code description was developed for each code.

In the open coding round, gender, age, and the electronically registered submission time and date were specified for each inquiry. For inquiries where gender was not stated—since it was not a requirement—"unknown" was entered, unless the content of the inquiry explicitly indicated gender (eg, "I am a 16-year-old girl"), thus relying solely on explicit, self-reported statements of gender and avoiding interpretive attribution. This would have applied equally to all explicitly stated gender identities, whether binary or nonbinary.

Gender and age were also adjusted if the information explicitly provided in the inquiry differed from the predefined categories, in which case the text always took analytical precedence. Such adjustments were particularly common in relation to age, as some young people stated that they were younger than 13 years or older than 20 years. Similarly, gender categorizations were guided exclusively by how young people described themselves in the inquiry text. If a young person had explicitly stated a gender identity other than girl or boy, this would have been recorded accordingly; however, no inquiries included such explicit self-description.

Among the inquiries included in the analysis, it was not possible to identify the gender of the young person in 455 inquiries. It should be emphasized that the unspecified gender category should not be interpreted as reflecting a particular gender identity or group. Rather, it solely indicates that the young person did not disclose their gender, regardless of whether their gender identity is binary, nonbinary, or otherwise.

Coding Round 2: Focused Coding and Refinement of Codes

In the second coding round, we moved toward focused coding, aiming to synthesize and explain larger segments of data [19]. At this stage, we sought to ensure consistency in the content of the codes and comparability between the inquiries assigned to each code, as the goal was to determine the adequacy of the codes. The first author (AOK) reread all coded data within each of the 115 codes, and any inquiry that did not align with the overall content of its assigned code was recoded. Similarly, code titles and descriptions were revised when necessary to better reflect their content and maintain proximity to the data. Codes were deleted if they became irrelevant due to recoding or merged when they overlapped with other codes. Code descriptions were actively used as analytical reference points during this process to assess whether the inquiries grouped under each code consistently reflected the defined content of the code. Where discrepancies were identified, codes and their descriptions were iteratively revised to maintain conceptual coherence and proximity to the data. After reviewing each code, frequently used words were searched across the remaining material to identify any potentially overlooked inquiries. This process involved comparing data with data and data with codes, which helped refine and strengthen the coding structure. Following this focused coding round, the number of codes was reduced to 101 (for a complete overview of the codes, see Multimedia Appendix 1).

Coding Round 3: Analytical Theme Formation

In the third and final round of coding, the 101 empirically generated codes were grouped into 17 distinct analytical themes. This step marks another move away from initial coding, which fractured the data into discrete segments, toward reassembling the material into a coherent whole. The focus is on constructing an analytical narrative that brings the emerging categories together in a meaningful and connected way [19]. The first author (AOK) drafted the initial themes, which were then discussed jointly with all the study's authors. Through this iterative process, the final themes and their labeling were iteratively refined and defined. To provide an additional overview of the thematic directions in the material, the 17 themes were further grouped into 3 overarching analytical categories: Social Relationships and Social Contexts, Emotional Life, and Body and Illness (Multimedia Appendix 1).

Following the coding rounds and the identification of themes, exploratory cross-tabulations were conducted between themes, as well as between themes and gender, age, and time of submission (month and year) using NVivo; only patterns deemed analytically meaningful are reported in the results. The unit of analysis was the individual inquiry; accordingly, the reported counts represent the number of inquiries associated with each theme, rather than population-level distributions of concerns.

In the presentation of the 17 identified themes in the Results section, selected quotes are included to support the thematic analysis. These quotes were chosen based on their capacity to vividly exemplify the thematic patterns identified. The aim was to foreground excerpts that best articulate the meaning and scope of each theme, thereby supporting the analytical claims made and providing insight into how each theme is grounded in the data.

Exclusion of Inquiries

During the 3 rounds of coding, inquiries were continuously excluded. Of the 2523 inquiries, 93 were excluded, primarily because they fell outside the scope of the letter box service—for instance, inquiries from young people seeking assistance with school assignments. Inquiries submitted by parents concerning their children's well-being were also excluded, as the study focuses exclusively on inquiries submitted by young people themselves. Inquiries that were incomprehensible were likewise excluded. Additionally, 19 inquiries could not be categorized under broader themes due to their highly individual content and are therefore not included in the thematic analysis. The final thematic structure is based on 2411 inquiries.

Results

This section presents the main findings, including identified themes, intersections between themes, and gendered patterns.

XSL•FO
RenderX

Thematic Analysis

Textbox 1 presents the 17 identified themes, organized across 3 analytical categories: Social Relationships and Social Context, Emotional Life, and Body and Illness. Each theme includes an integrated analytic description that captures recurring patterns in the data. For ease of presentation, the themes will hereafter be referred to using their short title shown in italics.

As illustrated in Figure 1, young people predominantly write about themes within the overarching category Social Relationships and Social Contexts when reaching out to the letter box service. However, a notable number of inquiries also touch upon topics related to complex emotions, as reflected in the themes grouped under the category Emotional Life. Themes within the overarching category Body and Illness appear less frequently in the inquiries, with the notable exception of the theme Psychiatric Diagnoses, which stands out as more prominent.

Textbox 1. Identified themes with integrated analytic descriptions.

Social Relationships and Social Contexts

- Sociality: navigating social relationships
- Love Life: navigating romantic relationships and feelings
- Seeking Support (represents a crosscutting theme related to help-seeking behavior rather than a distinct concern or emotional issue): attempts to seek support that are often unmet
- Education and Work: experiencing educational distress and uncertainty
- Family Life: experiencing strained family relationships
- Abuse: experiences of violence and harm in relationships

Emotional Life

- Unease: experiencing persistent anxiety and worry
- Self-Doubt and Insecurity: negative self-evaluation
- Despair: experiences of hopelessness, emptiness, and loss of meaning
- Negative Emotions: low and fluctuating mood
- Exhaustion: persistent tiredness and feeling overwhelmed
- Loneliness: feeling alone and excluded from others
- Identity: struggling to form identity

Body and Illness

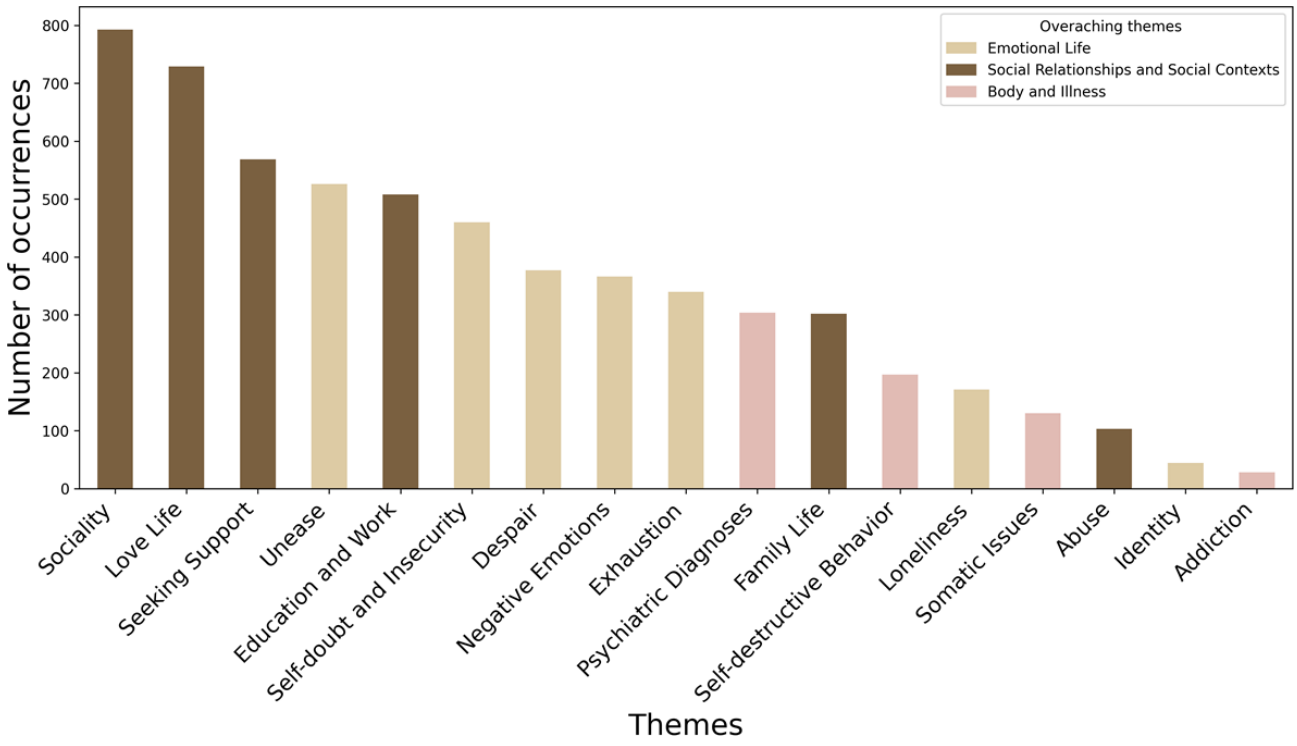
- Psychiatric Diagnoses: experiences of and seeking clarification about psychiatric diagnoses
- Self-Destructive Behavior: engaging in harmful practices
- Somatic Issues: bodily symptoms, experiences, and concerns
- Addiction: experiences and concerns related to addictive behaviors

Figure 1 shows the distribution of the 2411 inquiries across the identified themes and 3 analytical categories.

The analyses indicate that inquiries to the letter box come from children and young people aged between 8 and 28 years, with the median age across themes ranging from 15 to 17.5 years. Those writing about topics such as *Love Life*, *Psychiatric Diagnoses*, and *Somatic Issues* tended to be slightly older than individuals addressing other themes. In contrast, inquiries related to *Self-Destructive Behavior* were generally submitted by younger individuals compared with the remaining themes. Furthermore, cross-tabulations between themes and time of submission did not reveal any notable patterns.

XSL•FO
RenderX

Figure 1. The distribution of inquiries across 17 identified themes.



Social Relationships and Social Contexts

Social relationships and the social contexts that young people navigate throughout adolescence play a significant role in their inquiries to the digital letter box. Under this analytical category, we present 6 themes that represent some of the most frequently addressed topics in young people's inquiries.

With 793 inquiries, *Sociality* emerges as the most addressed theme, reflecting how young people navigate social relationships. Nearly half of the inquiries within this theme involve friendships, for example, fears of losing friends, broken friendships, conflicts, or feeling misunderstood or unwanted, as the following quote illustrates:

> *[...] I don't really talk to many people my own age, and I haven't spent time with anyone outside of school or work in the past year, except for my partner. Honestly, I don't know what to do and just feel like crying. When I do try to take the initiative to arrange something with friends or acquaintances, it never ends up happening. I generally just feel rejected and unaccepted. It's like I'm stuck in a bubble all by myself.* [Boy, aged 17 years]

Some describe dilemmas such as romantic feelings for friends, worries about friends, or comparing themselves to others. Other submissions reflect more broadly on social life—how they act in social settings, expectations to be social, or doubts about their social skills, for example:

> *I am 20 years old and do not feel that I am doing enough "typical youth things." I often feel more mature than my peers. I therefore feel that I have too few friends and do not know what to spend my time on after work... [...].* [Gender unspecified, aged 20 years]

Some young people seek advice on becoming more social, while others express a desire to withdraw. Many describe insecurity in social situations, such as fear of judgment, saying the wrong thing, or feeling awkward. Others mention shyness, introversion, or being withdrawn. A number of young people also write about concerns of close relationships—family, partners, or friends—where young people seek advice on how to support someone or cope with being close to someone with, for example, mental illness, as expressed by a 13-year-old girl: "My friend has been going through a difficult time lately, and I would like to help him with it. [...]." Finally, some write about social relationships or contexts that have had a particularly negative impact, such as encounters with bullying, experiences of betrayal and mistrust, facing cultural practices within the family, and dealing with difficulties forming close connections.

Love Life is also a central theme with 729 inquiries, highlighting how young people navigate romantic relationships and feelings. These include experiences of being in love, having a crush, or flirting with someone. One submission reads:

> *I'm a 13-year-old boy. I've met a girl—she's really sweet, but I don't know her very well. I think I might have a crush on her, but I'm not sure what that feels like since I've never had a girlfriend before. I don't know if we're a good match, because she goes to a different school than I do, and she's 12. I don't know if she likes me back, and I'm afraid of embarrassing myself. How can I tell if she likes me—can I sense it from the way she talks or looks at me? Thanks in advance for your help.* [Boy, aged 13 years]

While such experiences can be positive, they can also involve dilemmas, such as falling for a friend, someone with a different sexual orientation, or someone who does not reciprocate the feelings. Additionally, young people write about heartbreak and

XSL•FO
RenderX

ex-partners, as one shares: "I desperately need help. My boyfriend broke up with me very suddenly, which now makes me feel like my life has come to a standstill [...]" [Girl, aged 17 years]. Numerous inquiries concern challenges or dilemmas related to romantic relationships with partners including insecurity, jealousy, being apart from their partner, doubts about the relationship's future, and conflicts between romantic partners and friends or family. Few young people write about broader worries in love life, such as how to handle specific romantic situations, worries about finding love, or meeting expectations around love: "[...] I have realized that I am probably afraid of love. I am afraid of being let down. I am afraid of being hurt [...]" [Girl, aged 16 years]. Intimate experiences and sexuality are rarely mentioned, but the few inquiries that do address these topics describe issues such as sex in relationships, sexual expectations, reflections on sexual debut, and coming out or exploring one's sexual identity.

Represented by 569 inquiries, the theme *Seeking Support* points to how young people attempt to seek support but are often not met with the help they need. Rather than reflecting a specific domain of concern, this theme captures how young people navigate help-seeking processes across different types of issues. In the inquiries, young people reflect on or describe experiences, attempts, or actual interactions of seeking help or support related to worries or mental health struggles, most often involving reaching out to partners, friends, or parents. While some young people report positive experiences when reaching out to others, many instead receive limited or insufficient support. Others describe explicitly negative experiences, such as being rejected or dismissed by those they turned to, as the quote demonstrates:

> *I feel like I can't turn to anyone, and I've been feeling really bad lately. Everywhere I read, it says you should talk to someone about it, but what if you don't have anyone to talk to at all? When I try to talk to my mom, we end up arguing because she doesn't understand me and places really high expectations on me. My dad and I don't have a close relationship. [...]* [Gender unspecified, aged 15 years]

Several express reservation about seeking help due to fear of mental health stigma and the shame it generates, and a few mention lacking a support network entirely—prompting them to turn to the letter box instead. Experiences of seeking professional help, such as from general practitioners, psychologists, or services such as Headspace and BørneTelefonen (The Children's Helpline), are also shared in the inquiries:

> *[...] I have been going to a psychologist. She didn't help me, and she seemed indifferent to me and my problems just like everyone else is, indifferent...So what can I do now? [...]* [Girl, aged 15 years]

These encounters are often described as inadequate or disappointing, leading them to seek support through the Mindhelper letter box. Others ask for advice on whether or how to approach professional support. Additionally, several inquiries mention interactions with teachers—ranging from helpful conversations to unsuccessful attempts to get support. Finally, a few young people express a general difficulty in talking about emotions or opening up to others.

A total of 508 inquiries is included under the theme *Education and Work* capturing experiences of educational distress and uncertainty. Nearly half describe educational distress—often due to feelings of insecurity, loneliness, or experiences of exclusion from their class, as expressed by the 17-year-old girl: "I am terrified that I will be left out again in group work. […]." A few young people mention both supportive and difficult teacher relations. Many inquiries concern academic performance, including descriptions of pressure to achieve good grades or difficulties keeping up academically, for example: "My grades are dropping, and it really stresses me out a lot. [...]" [Girl, aged 14 years]. Challenges such as giving presentations, participating in group work, or engaging in classroom activities are also mentioned. Additionally, a few young people write about difficulties with concentration in school. Several young people express thoughts about dropping out, changing their educational path, or switching schools, indicating that youth is a time marked by transitions that often lead to worry and reflection. This is illustrated by a young person who states:

> *I need your help. I'm in a situation where I'm unsure whether I've chosen the right education or the right place for my apprenticeship. My motivation isn't what it used to be when I first started, and I no longer look forward to going to work. I don't feel safe or comfortable at my workplace the way I think I should. [...] I'm training to become a baker, but the way people speak there doesn't really match the kind of person I am. I'm quiet, shy, and introverted, and the industry is much tougher and harsher than I expected. What should I do?* [Girl, aged 19 years]

These concerns are also reflected in inquiries about choosing an education, taking a gap year, or starting at a new educational institution. A few inquiries concern work or part-time jobs, describing pressure, poor conditions, low energy, or workplace well-being. Some also seek advice on job opportunities.

Family Life is addressed in 302 inquiries, illustrating experiences of strained family relationships. Young people describe their relationship with parents or family dynamics, shaped by events such as divorce. Some describe specific conflicts or disagreements with parents, while others express general distress at home, saying that they find it difficult to be there:

> *I can't be around my parents for more than a couple of hours without us starting to argue. Not a day goes by without them making me cry. Normally, I'm a strong person, but I can't handle being torn down every day, and I no longer want to live at home. [...]* [Boy, aged 15 years]

A few mention problematic parental behavior, including dynamics marked by violence or growing up with alcohol abuse by 1 or both parents. Some inquiries also include descriptions of sibling relationships, which can be either positive or negative—for example, concerns about siblings or reflections on how sibling dynamics affect the family as a whole.

XSL•FO
RenderX

The least frequent theme within the category of Social Relationships and Social Contexts, *Abuse*, appears in 103 inquiries, pointing to experiences of violence and harm in relationships. Young people report experiences of physical, psychological, or sexual violence, as illustrated by the account of a young person:

> *[...] At home, I'm having some problems...My dad...I'm afraid of him...But I'm also very stubborn and often talk back. But he becomes very violent when things don't go his way...He yells loudly, grabs us (me and my little brother), and sometimes pushes us. [...]* [Girl, aged 16 years]

Over half concern physical abuse, including violence in romantic relationships or violence observed between parents. The inquiries also include descriptions of sexual abuse, rape, or incest. Others describe psychological or verbal abuse, and a few reflect on having inflicted psychological harm on others.

Emotional Life

A large proportion of young people's inquiries to the letter box concern their emotional life, reflected in 7 distinct themes. *Unease* emerges as the most prominent theme within the emotional category, represented by 528 inquiries, reflecting experiences of persistent anxiety and worry. Young people describe unease in various ways, often explicitly referring to anxiety or panic attacks, for example: "I have struggled with anxiety since I was 6 years old, but it has never been diagnosed. [...]" [Girl, aged 15 years]. Others report uncontrollable racing thoughts, overwhelming worries, or a tendency to overthink everything, for example, worries about the future, or intrusive fears related to death or illness, as expressed in the following excerpt from an enquiry:

> *I can't be in myself anymore, I have far too many thoughts that I can't manage. I overthink EVERYTHING, to the point where it hurts. All kinds of thoughts are spinning around in my head. It's so exhausting, and it affects my everyday life. I can't even feel happy anymore, because I'm constantly thinking about everything. [...]* [Girl, aged 16 years]

In some cases, unease is expressed through physical symptoms resembling panic attacks, such as heart palpitations, shortness of breath, sweating, or dizziness. In a limited number of inquiries, young people describe a sense of unreality, such as feeling detached from their body or being unable to distinguish between dream and reality.

With 460 inquiries, the theme of *Self-doubt and Insecurity* emerges as a central concern revealing patterns of negative self-evaluation among the young people writing to the letter box. The young people express this by thoughts of not feeling good enough, feeling worthless, or being generally unsure of oneself. For some, this is aimed at their body and appearance, with statements such as feeling ugly or fat, as expressed in the following quote:

> *I'm a 15-year-old girl writing here because I'm really unhappy with my body and appearance. Lately, I've become truly dependent on food and have been eating large amounts outside of meals—often in secret because it feels so shameful. I just feel ugly and stupid. My family sees me in a very negative light, which I guess is understandable, but it's also exhausting.* [Girl, aged 15 years]

A small number of inquiries also describe a pervasive perfectionism, rooted in high expectations of oneself and a fear of failure. In other cases, young people explicitly state that they struggle with low self-esteem or lack of confidence. Among 377 inquiries, young people express profound *Despair*, reflected in experiences of hopelessness, emptiness, and loss of meaning. Over half of the inquiries within this theme mention suicidal thoughts, while others describe a sense of detachment—likened to, for example, carrying a black hole inside.

> *I've been feeling empty for a while. Especially in the evenings, I feel like I have nothing to live for. I find it hard to concentrate at school. Sometimes I feel okay, but other times I feel absolutely terrible. Sometimes I just wish I could disappear. Of course I want to live, but I also want to feel good. I'm afraid to tell my parents, even though I know it's one of the best things I could do. I'm also scared to say it out loud, mostly because I don't have a reason to be sad. But still, I feel really awful. Is it normal to cry without having a reason?* [Gender unspecified, aged 13 years]

A recurring statement is the loss of motivation and desire, whether for everyday tasks, former interests, or pursuing education or work. Some express a wish to escape—physically by leaving everything behind or mentally by becoming invisible or disappearing. For others, it takes the form of hopelessness and the belief that nothing will ever improve.

Negative Emotions such as feeling sad, depressed, or downcast are reflected in 366 inquiries, where low and fluctuating mood frequently emerges. Many young people describe a persistent sense of emotional distress, often articulated as a lack of joy, a constant urge to cry, or crying frequently, which is exemplified by a 20-year-old young woman, who opens her inquiry by writing: "I have been feeling sad and down for quite some time. Without being able to say for sure, I'm afraid I might be developing depression. [...]."

While sadness dominates this theme, a few inquiries express anger or frustration, sometimes through descriptions of shouting or screaming, as a 17-year-old girl writes in her inquiry: "[...] I also get angry and irritated with people more quickly, and that makes me even more angry. [...]." Some young people seek advice on managing overwhelming emotions or mood swings. This theme also includes experiences of grief and anticipatory grief.

Exhaustion is a recurring theme in 340 inquiries, described as constant fatigue or feeling drained of energy. Sleep problems are also common, including difficulty falling asleep or poor sleep quality, for example: "[...] I have trouble sleeping. I often lie for several hours with my eyes closed. [...]" [Boy, aged 13 years]. Additionally, some report feeling stressed, pressured, or overwhelmed in their daily lives. These experiences are clearly illustrated in the following inquiry:

XSL•FO
RenderX

> *I feel constantly down, sad, and incredibly tired all the time. I usually bike 14 km on weekdays, but I can't cope with it, and I'm just really sad. I get stressed and angry quickly, and I find it hard to tolerate many things, which makes me easily irritated with people. What could be wrong with me? Or what can I do to feel better?* [Girl, aged 16 years]

The theme of *Loneliness* is based on 171 inquiries, with loneliness expressed both directly and indirectly in young people's accounts of feeling lonely, alone, or experiencing life as lonely. While several young people explicitly mention loneliness, others express it through statements about often being alone, as seen in the words of a 15-year-old who begins their inquiry with: "I feel very left out, wrong, and alone. [...]."

The final theme within Emotional Life concerns *Identity* and appears in 44 inquiries, where young people struggle to form a sense of identity. Most involve young people reflecting on their interests and hobbies as part of shaping their identity. One young person describes her struggle with finding purpose:

> *I just have this problem that I don't have any interests or hobbies to spend my time on. I have a boyfriend and a job, and when I'm not working or with my boyfriend, I simply don't know what to do. As a child, I tried a whole range of extracurricular activities—everything from handball to cooking and crocheting—but nothing ever caught my interest. I have no talents, and I'm neither creative, musical, nor anything like that. TV series and movies don't really appeal to me, and books can't hold my attention either. Because I don't have anything I'm passionate about or spend my time on, I feel empty inside.* [Girl, aged 19 years]

A few describe compulsive lying to close relations—sometimes, for example, about sexual experiences—which have spiraled out of control. Others express a sense of identity crisis, feeling disconnected from who they once were.

Body and Illness

While 4 themes concerning Body and Illness have been identified, their significance appears limited within the overall narrative of young people's inquiries to the letter box. However, several young people write about *Psychiatric Diagnoses*, constituting the largest theme within this category, with 304 inquiries. Thus, although the category Body and Illness itself is relatively small, the theme *Psychiatric Diagnoses* ranks as the tenth most prevalent across the 17 identified themes, as previously shown in Figure 1. This theme includes both young people's experiences with psychiatric diagnoses and their desire to clarify whether they might have one. Depression is the most frequently described condition, with young people sharing current or past experiences of living with depression, and many seeking advice and information about whether they might be experiencing symptoms. This is illustrated in the words of a young person who writes:

> *I'm unsure whether I'm just going through a rough patch or if I'm starting to develop depression. I feel really down, drained of energy, and I don't find joy in the things that usually make me happy. [...]* [Girl, aged 20 years]

In addition to depression, 13 other psychiatric diagnoses are mentioned in the inquiries, with some of the most frequently cited being eating disorders, autism, obsessive-compulsive disorder, and attention-deficit/hyperactivity disorder.

The theme of *Self-Destructive Behavior* is identified in 197 inquiries, where young people engage in harmful practices, primarily covering accounts of current or past self-harming behavior such as cutting, scratching the skin, or hitting themselves. One young person shares: "I really can't take it anymore...Self-harm and suicidal thoughts are taking up so much space, and I just feel completely awful about myself. [...]" [Girl, aged 15 years]. Additionally, several young people express having a disturbed relationship with food, which may manifest in both their thoughts and behaviors.

A total of 132 inquiries to the letter box concern *Somatic Issues* encompassing bodily symptoms, experiences, and concerns. About half describe recurring physical symptoms linked to psychological distress, such as stomach pain, headaches, physical discomfort, tension, dizziness, nausea, and heart palpitations.

> *[...] I have so many thoughts that just keep going around in circles. Thoughts like: Why do I always do everything wrong, why didn't I say anything, I'm fat, I should lose weight, etc. These thoughts also often make me feel nauseous, and I feel like I'm about to throw up just from the thought of having to eat something. [...]* [Girl, aged 15 years]

Other inquiries reflect the impact of living with physical illness or injury on mental health and well-being. Additionally, several young people ask questions about the body, including reactions to various conditions, menstruation, pregnancy, contraception, sexually transmitted infections, diet, and exercise.

The final theme related to Body and Illness is the smallest in the letter box analysis, with 28 identified inquiries concerning *Addiction* focusing on young people's experiences and concerns related to addictive behaviors. Over half involve the use of alcohol, cannabis, or harder drugs as the young person writes about in his inquiry:

> *I've lost my entire circle of friends from childhood and school to hash or because I got kicked out of school. [...] Now I mostly sit alone from morning to evening and smoke myself numb. [...] I feel alone with everything and can't really see a way out...Help plz.* [Boy, aged 19 years]

While young people do not always describe their use as addiction, the extent of substance use is central to the problems they seek help for. Others write about addiction-related issues such as dependency on gaming, exercise, or shopping.

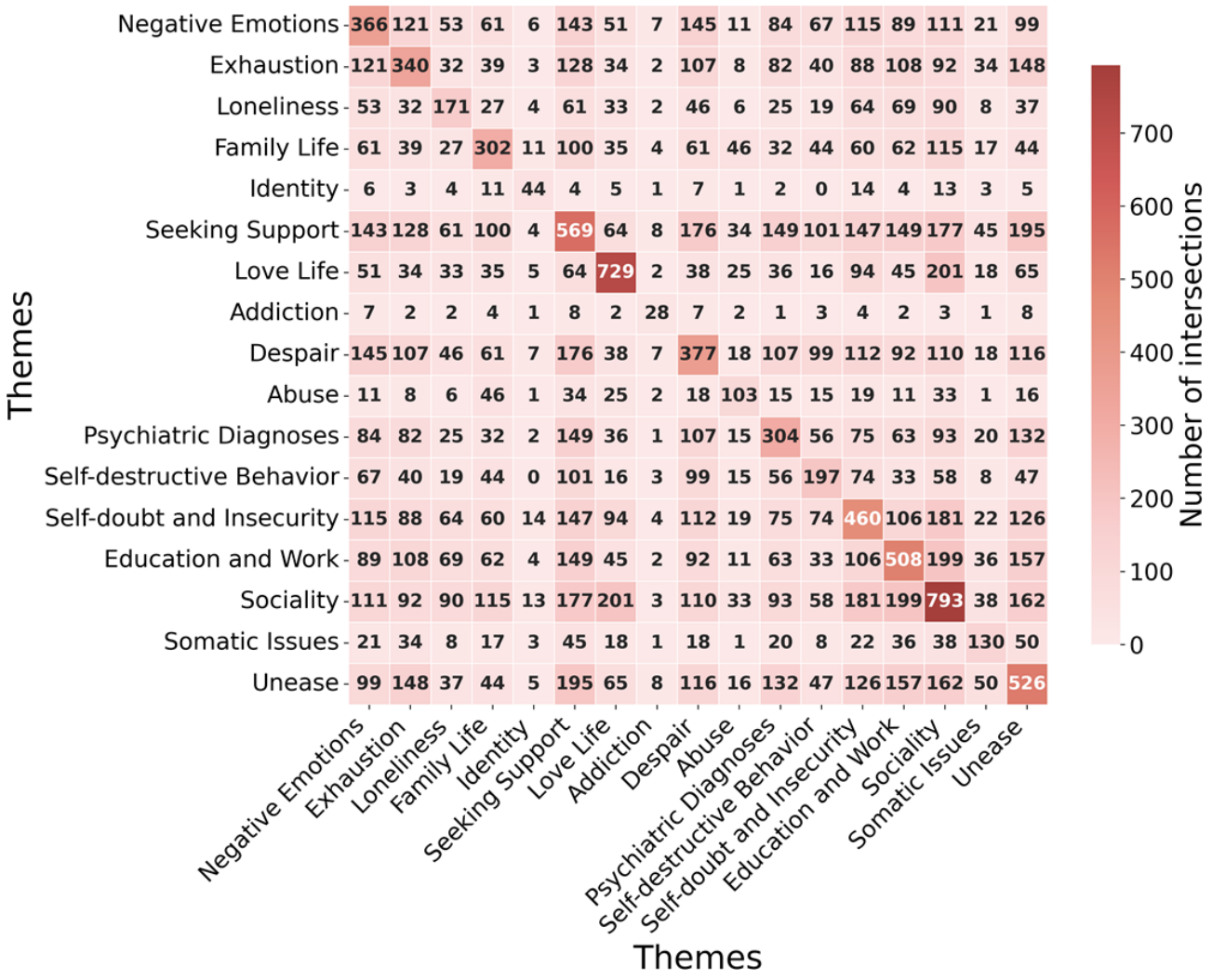
Intersection of Themes

This section explores the intersections between the 17 identified themes. The heatmap (Figure 2) shows how the themes are connected, with darker areas indicating stronger overlaps between themes. By presenting the data as a heatmap, we gain

XSL•FO
RenderX

a clearer understanding of the complexity and interrelated nature of young people's concerns.

Figure 2. Intersection of themes.



The central role of social relationships in young people's well-being is confirmed by examining the 3 most frequent combinations of 2 themes. A total of 201 inquiries from young people address both the themes *Love Life* and *Sociality*. These entries often revolve around dilemmas related to romantic feelings and friendships, as the 14-year-old girl expresses it in the title of her inquiry: "I'm in love with my friend, but she's not into girls," or issues involving partners and friends, as expressed by a 13-year-old girl who writes: "I have a friend who can't keep her hands off my boyfriend."

The combination of the themes *Sociality* and *Education and Work* is also frequently represented in young people's inquiries (n=199). These inquiries often include descriptions of social challenges within the school context and how such difficulties negatively impact their overall school well-being, as one young person puts it:

> *I'm a 14-year-old girl in 8th grade. I don't really fit in with my class, and I really want to. I feel left out during events and I'm not really part of the conversations. [...] And I really want to have more friends in my class and maybe be part of the community. But I don't know how.* [Girl, aged 14 years]

The third most common thematic combination is found in the pairing *Seeking Support* and *Unease* (n=195). A significant proportion of the inquiries within this theme combination reflect young people's experiences with anxiety, while also expressing efforts to seek professional help or posing questions about how to access such support. This is exemplified in the following inquiry:

> *I think I'm struggling with social anxiety. I've never been formally diagnosed, but I've had a few panic attacks. I generally feel very uncomfortable in large groups of people, and I break out all over my body when I'm at parties, in my school's cafeteria, or even just walking down the street. I'm fairly aware of my situation and I know what to do when I experience these "attacks" and the spots on my skin. Still, I'm tired of how much it affects my everyday life, and that's why I wanted to ask what you think I should do about it. Should I start seeing a psychologist and talk about it, even though I'm quite self-aware? Or should*

XSL•FO
RenderX

I just accept that this is how my daily life is going to be? [Girl, aged 18 years]

Furthermore, Figure 2 highlights that most inquiries to the digital letter box involve multiple themes within a single inquiry, with intersections appearing across nearly all 17 themes. Notably, the heatmap reveals that the theme *Seeking Support* is frequently addressed in conjunction with other themes, indicating that it rarely stands alone. This theme captures young people's experiences of navigating help-seeking processes in response to distress, functioning as a crosscutting dimension rather than expressing a specific concern or emotional issue, as is the case with the other 16 themes. Accordingly, it underscores the potential of a digital letter box as a space for outreach when traditional support networks feel inaccessible. The only theme combination not represented is between *Identity* and *Self-Destructive Behavior*, marked with a zero in the corresponding square to indicate that no inquiries addressed both themes.

Gender Patterns Across Themes

Table 1 presents the gender distribution across themes, highlighting a notable gender imbalance in young people's use of the letter box, with girls being overrepresented. Although this overall trend is clear, only slight variations appear when comparing gender distribution across individual themes. For instance, *Love Life* was the most frequently addressed theme among boys (105/286, 36.7%) compared with girls (499/1670, 29.9%) and those with unspecified gender (125/455, 27.5%). Conversely, *Sociality* was most prevalent among girls (567/1670, 34.0%) and the unspecified gender category (130/455, 28.6%), although it was also common among boys (96/455, 33.6%). Girls (416/1670, 24.9%) and individuals with unspecified gender (105/455, 23.1%) wrote more often about *Seeking Support* than boys (48/286, 16.8%). Similarly, for *Education and Work*, girls (366/1670, 21.9%) and those with unspecified gender (96/455, 21.1%) addressed these themes more frequently than boys (46/286, 16.1%).

Table . Gender patterns across themes.

Themes	Inquiries[a]		
	Girls, n (%)	Boys, n (%)	Unspecified gender, n (%)
Sociality	567 (34.0)	96 (33.6)	130 (28.6)
Love Life	499 (29.9)	105 (36.7)	125 (27.5)
Seeking Support	416 (24.9)	48 (16.8)	105 (23.1)
Education and Work	366 (21.9)	46 (16.1)	96 (21.1)
Family Life	205 (12.3)	30 (10.5)	67 (14.7)
Abuse	68 (4.1)	14 (4.9)	21 (4.6)
Unease	377 (22.6)	43 (15.0)	108 (23.7)
Self-Doubt and Insecurity	315 (18.9)	50 (17.5)	95 (20.9)
Despair	254 (15.2)	43 (15.0)	80 (17.6)
Negative Emotions	289 (17.3)	22 (7.7)	55 (12.1)
Exhaustion	253 (15.1)	26 (9.1)	61 (13.4)
Loneliness	119 (7.1)	18 (6.3)	34 (7.5)
Identity	28 (1.7)	9 (3.1)	7 (1.5)
Psychiatric Diagnoses	229 (13.7)	25 (8.7)	50 (11.0)
Self-Destructive Behavior	154 (9.2)	12 (4.2)	31 (6.8)
Somatic Issues	88 (5.3)	14 (4.9)	30 (6.6)
Addiction	9 (0.5)	12 (4.2)	7 (1.5)
Total	1670 (100)	286 (100)	455 (100)

[a]A single inquiry may appear under multiple themes.

Small variations were also observed across gender categories within themes related to emotional life. Young people with unspecified gender (108/455, 23.7%) and girls (377/1670, 22.6%) were more likely to address topics related to *Unease* than boys (43/286, 15.0%). Similarly, experiences with *Negative Emotions* were more frequently described by girls (289/1670, 17.3%) and those with unspecified gender (55/455, 12.1%) than by boys (22/286, 7.7%). Mentions of *Exhaustion* were also more common among girls (253/1670, 15.1%) and the unspecified gender category (61/455, 13.4%) than among boys (26/286, 9.1%).

XSL•FO
RenderX

For themes related to body and illness, the greatest variation was observed between boys and girls. Girls were considerably more likely to write about *Self-Destructive Behavior* (154/1670, 9.1%) than boys (12/286, 4.2%), while individuals with unspecified gender fell in between (31/455, 6.8%). Likewise, boys addressed subjects related to the theme *Addiction* more frequently (12/286, 4.2%) than girls (9/1670, 0.5%), with the unspecified gender category again positioned between the two (7/455, 1.5%).

Discussion

Principal Findings

This study analyzed inquiries submitted to the digital letter box of Mindhelper, Denmark's most widely used digital mental health promotion service, to identify key mental health themes and gender differences in engagement. A thematic analysis revealed 17 themes across 3 overarching categories— Social Relationships and Social Contexts, Emotional Life, and Body and Illness.

With some of the most salient themes centering on *Sociality* and *Love Life*, our findings confirm adolescence as a socially sensitive stage [6,21] and extend existing knowledge by showing that distress rarely occurs in isolation. Instead, it is deeply entangled with relational contexts such as friendships, romantic relationships, and family life [22,23]. These patterns may help inform how digital support services are structured to better reflect and respond to interconnected user needs.

A central insight emerging from this study is that digital help-seeking often occurs at the intersection between unmet needs in informal support networks and barriers to formal care. While previous research has emphasized the importance of social networks for youth help-seeking and conceptualized it as a socially embedded process [24-28], our findings suggest that some young people turn to digital platforms precisely when these networks are experienced as insufficient, inaccessible, or unresponsive. This shift carries both potential and limitations. On the one hand, digital letter boxes provide a low-threshold, anonymous entry point to support, enabling young people to articulate concerns that might otherwise remain unexpressed [29,30]. On the other hand, help-seeking outside established social relationships may risk reinforcing isolation or limiting opportunities for sustained support, particularly given the one-off and text-based nature of such services. The findings suggest that digital mental health services should be designed not only as stand-alone support but also as bridges to existing social and professional resources, with particular attention to users who lack supportive networks.

Second, our analysis highlights gendered engagement patterns without sharp divides. While social relationships are important for both boys and girls, boys more frequently address *Love Life*, whereas girls more often write about friendships, broadly social life, and themes related to Emotional Life. While girls dominate overall volume, boys' inquiries often cluster around behavioral expressions such as addiction and abuse, suggesting that they may be more likely to externalize distress through behaviors such as substance use, a pattern noted in previous research on gendered expressions of distress [31].

These patterns prompt a closer look at help-seeking behavior. Girls turning to the digital letter box for advice far more often than boys may reflect a higher prevalence of distress among Danish girls, as documented in national health surveys [5], but it may also indicate that boys are less inclined to seek help. Previous research has identified gendered barriers to help-seeking among young men [32-35]. For instance, a review of 31 studies found that 27 identified masculine attitudes—such as the need to appear strong, preference for self-management, and viewing help-seeking as weakness—as barriers to seeking support [32]. While the review covered youth help-seeking and health care access broadly, most studies focused on mental health (n=17) and sexual health (n=10). Although evidence suggests that understandings of masculinity and related behaviors influence perceptions of mental health and help-seeking [31], several scholars stress the importance of incorporating broader contextual factors when examining gender differences in help-seeking behavior. For instance, studies have shown that ethnicity, education, and age also influence help-seeking tendencies [36-38]. McKenzie et al [39] also underline the importance of adopting an intersectional perspective to better understand men's experiences of mental illness stigma, noting that factors such as homophobia and racism can shape these experiences. Moreover, gender differences in help-seeking appear to diminish at higher levels of psychological distress, which may suggest that men are less likely than women to recognize mild symptoms of mental distress [40-42]. Furthermore, research has highlighted the role of social network quality in shaping help-seeking behavior [43,44].

Building on this, the role of the digital context should be further considered in studies of gender differences in help-seeking and engagement within mental health services. Digital features such as anonymity and accessibility may shape some of the patterns observed between genders. Importantly, a review study highlights that attitudes and preferences—such as trust and distrust in health care providers, beliefs about the necessity or effectiveness of care, perceived dissimilarity with providers, preferences for self-reliance, and concerns about anonymity or other potential negative consequences—are strong predictors of help-seeking among young people [45]. This suggests that research on gendered barriers to help-seeking should consider not only the individual factors but also the broader context of service provision, including whether services are delivered digitally. In line with other scholars, we therefore point to the need for models of help-seeking and engagement that incorporate these contextual factors to gain a more comprehensive understanding of gendered digital mental health behavior and to inform the development of designs that more effectively appeal to boys [41,46,47].

Analyses such as the present one are important because they enable the systematic identification of emerging help-seeking needs among young people. These needs might otherwise go unnoticed and yet are essential for developing relevant support content. At the same time, it is necessary to consider the rapidly evolving landscape of digital mental health services driven by

XSL•FO
RenderX

technological developments. Over the course of this study, major changes occurred that may influence young people's help-seeking behavior and use of digital services such as letter boxes. In late 2022, ChatGPT was launched, and by 2025, artificial intelligence–generated content became a dominant feature of Google's search results, potentially introducing new sources for youth seeking mental health advice. Previous studies, conducted before these developments, show that text-based queries via search engines remain the most common online help-seeking approach among young people [30,48]. This reliance on search engines underscores how artificial intelligence integration may reshape the information young people encounter, creating new challenges for developers to ensure access to professional, evidence-based resources. These developments underscore the need for future research to examine their implications for youth help-seeking behavior and engagement with digital mental health platforms.

Strengths and Limitations

While this study offers valuable insight, it is important to acknowledge both its methodological strengths and limitations. A key strength lies in its unique data foundation, providing rare access to young people's real-time expressions of mental health challenges, rather than retrospective accounts. The anonymous digital context enables open engagement with sensitive and stigmatized topics, supporting more candid insights into youth experiences. Additionally, the large and temporally broad dataset allows for a comprehensive view of evolving challenges young people face.

Despite the study's strengths, several limitations should be acknowledged. Each submission to the Mindhelper letter box provides a one-time snapshot of a young person's mental health concerns, without the possibility of follow-up or contextual elaboration beyond what the individual chooses to express. Moreover, the material reflects a specific context: young people who choose to engage with this type of digital service. Consequently, the sample may be biased toward those who are willing and able to seek help through digital platforms. In addition, written communication is central on a digital platform, which means the format requires users to be able to read and write, potentially excluding youth with limited literacy skills. However, no formal writing standards are imposed, which is reflected in the data, occasionally marked by linguistic ambiguity.

Finally, attention should be drawn to the analytical process. The coding was conducted by a single researcher, which may have introduced some bias. However, the systematic and transparent approach, combined with reflexive engagement and author group involvement, supports the credibility of the analysis. In addition, changes in gender categories on Mindhelper introduced limitations for the gender-related analyses. Initially, binary categories allowed some insight into differences between boys' and girls' mental health struggles, although this approach excluded nonbinary youth and may have discouraged nonbinary self-identification. Removing gender categories entirely improved inclusivity but restricted the ability to examine how gender may shape mental health. Future research should seek methods that capture gender diversity while enabling meaningful analysis.

Conclusions

Seventeen themes emerged, reflecting the diverse concerns young people seek advice for, with social relationships and emotional experiences constituting central domains. Gender differences were observed in the extent of engagement, with girls more frequently using the digital letter box than boys, while only minor variations appeared in the distribution of themes across genders. Taken together, the findings suggest that many of the concerns young people raise are closely tied to everyday developmental and relational challenges rather than severe psychopathology alone. The digital letter box thus captures a broader spectrum of lived experience, including early signs of distress that may not otherwise reach formal systems. This highlights both the preventive potential of such services and the value of using self-initiated digital data to better understand how young people articulate and act on emerging mental health concerns.

Acknowledgments

The authors are grateful to the Mindhelper team at the Centre for Digital Psychiatry for their assistance in extracting the data. Furthermore, they would like to express their gratitude to Teresa Beata Bundgård for her assistance in refining the figures. The authors used Microsoft Copilot (GPT-5), an artificial intelligence–based language model, to assist with language editing and refinement of this manuscript, in accordance with journal guidelines. CHD was affiliated with the National Institute of Public Health at the University of Southern Denmark at the time the study was conducted and is currently affiliated with the company Capgemini.

Funding

This research was supported by an institutional grant from the University of Southern Denmark, funds from the Jascha Foundation, Danish Regions, and TrygFonden.

Data Availability

The datasets analyzed during this study are available from the corresponding author upon reasonable request. While the original inquiries are publicly accessible via the Mindhelper.dk platform, anonymized analytical materials (eg, coded data and NVivo project files) may also be shared in accordance with ethical and data protection considerations.

XSL•FO
RenderX

Authors' Contributions

Conceptualization: AOK (lead), CHD (supporting), SHH (supporting), LCT (supporting), APF (supporting)
Data curation: AOK (lead), CHD (supporting)
Formal analysis: AOK (lead), CHD (supporting)
Funding acquisition: APF (lead), LCT (equal), SHH (supporting), AOK (supporting)
Investigation: AOK
Methodology: AOK (lead), CHD (supporting), LCT (supporting), SHH (supporting), APF (supporting)
Project administration: AOK
Supervision: APF (lead), LCT (equal)
Validation: AOK (lead), CHD (supporting), SHH (supporting), LCT (supporting), APF (supporting)
Visualization: CHD
Writing – original draft: AOK
Writing – review & editing: AOK (lead), CHD (supporting), SHH (supporting), LCT (supporting), APF (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Codebook.
[DOCX File, 48 KB - humanfactors_v13i1e91017_app1.docx]

References

1. Jensen HAR, Møller SR, Jezek AH, Davidsen M, Ekholm O, Christensen AI. Danskernes sundhed 2023 [Report in German]. : Statens Institut for Folkesundhed, SDU; 2024 URL: https://www.sdu.dk/sif/-/media/images/sif/udgivelser/2025/danskernes_sundhed_2023__.pdf [accessed 2026-07-17]
2. Jensen HAR, Davidsen M, Christensen AV, Ekholm O, Christensen AI. Danskernes Sundhed—Den Nationale Sundhedsprofil 2025 [Report in German]. : Statens Institut for Folkesundhed, SDU; 2026 URL: https://www.sdu.dk/sif/-/media/images/sif/udgivelser/2026/den_nationale_sundhedsprofil_2025.pdf [accessed 2026-07-17]
3. McGorry P, Gunasiri H, Mei C, Rice S, Gao CX. The youth mental health crisis: analysis and solutions. Front Psychiatry 2024;15:1517533. [doi: 10.3389/fpsyt.2024.1517533] [Medline: 39906686]
4. Child, adolescent and youth mental health in the 21st century. : OECD; 2026 URL: https://doi.org/10.1787/1092c3cb-en [accessed 2026-07-17]
5. McGorry PD, Mei C, Dalal N, et al. The Lancet Psychiatry Commission on youth mental health. Lancet Psychiatry 2024 Sep;11(9):731-774. [doi: 10.1016/S2215-0366(24)00163-9] [Medline: 39147461]
6. Rapee RM, Oar EL, Johnco CJ, et al. Adolescent development and risk for the onset of social-emotional disorders: a review and conceptual model. Behav Res Ther 2019 Dec;123:103501. [doi: 10.1016/j.brat.2019.103501] [Medline: 31733812]
7. Johnson D, Dupuis G, Piche J, Clayborne Z, Colman I. Adult mental health outcomes of adolescent depression: a systematic review. Depress Anxiety 2018 Aug;35(8):700-716. [doi: 10.1002/da.22777] [Medline: 29878410]
8. Mental health of adolescents. World Health Organization. 2025. URL: https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health [accessed 2026-07-16]
9. Tolstrup JS, Larsen SR, Kelleher I, Cannon M, Lytken Larsen CV, Nordentoft M. Mental well-being in adolescence and eight years of follow-up for mental illness, risky behaviours, and mortality in 67,945 15-19-year-olds: a prospective cohort study. Lancet Reg Health Eur 2025 Nov;58:101435. [doi: 10.1016/j.lanepe.2025.101435] [Medline: 40989561]
10. Wright M, Reitegger F, Cela H, Papst A, Gasteiger-Klicpera B. Interventions with digital tools for mental health promotion among 11-18 year olds: a systematic review and meta-analysis. J Youth Adolesc 2023 Apr;52(4):754-779. [doi: 10.1007/s10964-023-01735-4] [Medline: 36754917]
11. Yeo G, Reich SM, Liaw NA, Chia EYM. The effect of digital mental health literacy interventions on mental health: systematic review and meta-analysis. J Med Internet Res 2024 Feb 29;26:e51268. [doi: 10.2196/51268] [Medline: 38421687]
12. Hopkin G, Coole H, Edelmann F, et al. Toward a new conceptual framework for digital mental health technologies: scoping review. JMIR Ment Health 2025 Feb 19;12:e63484. [doi: 10.2196/63484] [Medline: 39969824]
13. Gan DZQ, McGillivray L, Han J, Christensen H, Torok M. Effect of engagement with digital interventions on mental health outcomes: a systematic review and meta-analysis. Front Digit Health 2021;3:764079. [doi: 10.3389/fdgth.2021.764079] [Medline: 34806079]
14. Lehtimaki S, Martic J, Wahl B, Foster KT, Schwalbe N. Evidence on digital mental health interventions for adolescents and young people: systematic overview. JMIR Ment Health 2021 Apr 29;8(4):e25847. [doi: 10.2196/25847] [Medline: 33913817]

XSL•FO
RenderX

15. Ho TQA, Le LKD, Engel L, et al. Barriers to and facilitators of user engagement with web-based mental health interventions in young people: a systematic review. Eur Child Adolesc Psychiatry 2025 Jan;34(1):83-100. [doi: 10.1007/s00787-024-02386-x]
16. Borghouts J, Eikey E, Mark G, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387. [doi: 10.2196/24387] [Medline: 33759801]
17. Potts C, Kealy C, McNulty JM, et al. Digital mental health interventions for young people aged 16-25 years: scoping review. J Med Internet Res 2025 May 9;27:e72892. [doi: 10.2196/72892] [Medline: 40344661]
18. Patton MQ. Qualitative Research & Evaluation Methods, 3rd edition: SAGE Publications Inc; 2002.
19. Charmaz K. Constructing Grounded Theory—A Practical Guide Through Qualitative Analysis: SAGE Publications Inc; 2006.
20. Alvesson M, Sköldberg K. Reflexive Methodology—New Vistas for Qualitative Research: SAGE Publications Inc; 2009.
21. Veenstra R, Laninga-Wijnen L. The prominence of peer interactions, relationships, and networks in adolescence and early adulthood. In: Crockett LJ, Carlo G, Schulenberg JE, editors. APA Handbook of Adolescent and Young Adult Development: American Psychological Association; 2023:225-241. [doi: 10.1037/0000298-014]
22. Lin J, Guo W. The research on risk factors for adolescents' mental health. Behav Sci (Basel) 2024 Mar 22;14(4):263. [doi: 10.3390/bs14040263] [Medline: 38667059]
23. Copeland M. Embedded distress: social integration, gender, and adolescent depression. Soc Forces 2023 Mar;101(3):1396-1421. [doi: 10.1093/sf/soac034] [Medline: 36688226]
24. Aguirre Velasco A, Silva Santa Cruz I, Billings J, Jimenez M, Rowe S. What are the barriers, facilitators and interventions targeting help-seeking behaviours for common mental health problems in adolescents? A systematic review. Focus (Am Psychiatr Publ) 2025 Jan;23(1):98-118. [doi: 10.1176/appi.focus.25023003] [Medline: 39776457]
25. Ryan SM, Jorm AF, Toumbourou JW, Lubman DI. Parent and family factors associated with service use by young people with mental health problems: a systematic review. Early Interv Psychiatry 2015 Dec;9(6):433-446. [doi: 10.1111/eip.12211] [Medline: 25762206]
26. Stunden C, Zasada J, VanHeerwaarden N, et al. Help-seeking behaviors of transition-aged youth for mental health concerns: qualitative study. J Med Internet Res 2020 Oct 5;22(10):e18514. [doi: 10.2196/18514] [Medline: 33016882]
27. Radez J, Reardon T, Creswell C, Orchard F, Waite P. Adolescents' perceived barriers and facilitators to seeking and accessing professional help for anxiety and depressive disorders: a qualitative interview study. Eur Child Adolesc Psychiatry 2022 Jun;31(6):891-907. [doi: 10.1007/s00787-020-01707-0]
28. Wahdi AE, James C, Nadhira DA, Fine SL. Barriers and facilitators of seeking help for mental health challenges among adolescents across 13 countries: a qualitative investigation. J Adolesc Health 2025 Sep;77(3):421-427. [doi: 10.1016/j.jadohealth.2024.02.035] [Medline: 40908059]
29. Sit M, Elliott SA, Wright KS, Scott SD, Hartling L. Youth mental health help-seeking information needs and experiences: a thematic analysis of Reddit posts. Youth Soc 2024 Jan;56(1):24-41. [doi: 10.1177/0044118X221129642]
30. Pretorius C, Chambers D, Coyle D. Young people's online help-seeking and mental health difficulties: systematic narrative review. J Med Internet Res 2019 Nov 19;21(11):e13873. [doi: 10.2196/13873] [Medline: 31742562]
31. Mokhwelepa LW, Sumbane GO. Men's mental health matters: the impact of traditional masculinity norms on men's willingness to seek mental health support; a systematic review of literature. Am J Mens Health 2025;19(3):15579883251321670. [doi: 10.1177/15579883251321670] [Medline: 40421591]
32. Palmer R, Smith BJ, Kite J, Phongsavan P. The socio-ecological determinants of help-seeking practices and healthcare access among young men: a systematic review. Health Promot Int 2024 Apr 1;39(2):daae024. [doi: 10.1093/heapro/daae024] [Medline: 38513244]
33. Westberg KH, Nyholm M, Nygren JM, Svedberg P. Mental health problems among young people—a scoping review of help-seeking. Int J Environ Res Public Health 2022 Jan 27;19(3):1430. [doi: 10.3390/ijerph19031430] [Medline: 35162452]
34. Sheikh A, Payne-Cook C, Lisk S, Carter B, Brown JSL. Why do young men not seek help for affective mental health issues? A systematic review of perceived barriers and facilitators among adolescent boys and young men. Eur Child Adolesc Psychiatry 2025 Feb;34(2):565-583. [doi: 10.1007/s00787-024-02520-9] [Medline: 39004687]
35. Fisher K, Seidler ZE, King K, Oliffe JL, Rice SM. Men's anxiety: a systematic review. J Affect Disord 2021 Dec 1;295:688-702. [doi: 10.1016/j.jad.2021.08.136] [Medline: 34517242]
36. Chu JP, Hsieh KY, Tokars DA. Help-seeking tendencies in Asian Americans with suicidal ideation and attempts. Asian Am J Psychol 2011;2(1):25-38. [doi: 10.1037/a0023326]
37. Madsen J, Jobson L, Slewa-Younan S, et al. Arab Australian men's perspectives on mental health: exploring beliefs about causes, help-seeking attitudes, and experiences. Psychol Men Masculin 2025;26(4):610-623. [doi: 10.1037/men0000520]
38. Nystrom DA, del Campo C, Fernández-Cornejo JA. Intersectional analysis of men's masculinities and mental health-seeking behavior: a novel application of multiple correspondence analysis. Psychol Men Masculin 2024;25(1):57-70. [doi: 10.1037/men0000457]
39. McKenzie SK, Oliffe JL, Black A, Collings S. Men's experiences of mental illness stigma across the lifespan: a scoping review. Am J Mens Health 2022;16(1):15579883221074789. [doi: 10.1177/15579883221074789] [Medline: 35125015]

XSL•FO
RenderX

40. Biddle L, Gunnell D, Sharp D, Donovan JL. Factors influencing help seeking in mentally distressed young adults: a cross-sectional survey. Br J Gen Pract 2004 Apr;54(501):248-253. [Medline: 15113490]
41. Vogel DL, Heath PJ. Men, masculinities, and help-seeking patterns. In: Wong YJ, Wester SR, editors. APA Handbook of Men and Masculinities: American Psychological Association; 2016:685-707. [doi: 10.1037/14594-031]
42. Seidler ZE, Dawes AJ, Rice SM, Oliffe JL, Dhillon HM. The role of masculinity in men's help-seeking for depression: A systematic review. Clin Psychol Rev 2016 Nov;49:106-118. [doi: 10.1016/j.cpr.2016.09.002] [Medline: 27664823]
43. Vogel DL, Wade NG, Wester SR, Larson L, Hackler AH. Seeking help from a mental health professional: the influence of one's social network. J Clin Psychol 2007 Mar;63(3):233-245. [doi: 10.1002/jclp.20345] [Medline: 17211874]
44. Ishikawa A, Rickwood D, Bariola E, Bhullar N. Autonomy versus support: self-reliance and help-seeking for mental health problems in young people. Soc Psychiatry Psychiatr Epidemiol 2023 Mar;58(3):489-499. [doi: 10.1007/s00127-022-02361-4] [Medline: 36112162]
45. Huff NR, Dunderdale L, Kellogg AJ, Isbell LM. Factors related to help-seeking and service utilization for professional mental healthcare among young people: an umbrella review. Clin Psychol Rev 2024 Dec;114:102504. [doi: 10.1016/j.cpr.2024.102504] [Medline: 39395209]
46. Haavik L, Joa I, Hatloy K, Stain HJ, Langeveld J. Help seeking for mental health problems in an adolescent population: the effect of gender. J Ment Health 2019 Oct;28(5):467-474. [doi: 10.1080/09638237.2017.1340630] [Medline: 28719230]
47. Ellis LA, Collin P, Hurley PJ, Davenport TA, Burns JM, Hickie IB. Young men's attitudes and behaviour in relation to mental health and technology: implications for the development of online mental health services. BMC Psychiatry 2013 Apr 20;13(1):119-119. [doi: 10.1186/1471-244X-13-119] [Medline: 23601273]
48. Kauer SD, Mangan C, Sanci L. Do online mental health services improve help-seeking for young people? A systematic review. J Med Internet Res 2014 Mar 4;16(3):e66. [doi: 10.2196/jmir.3103] [Medline: 24594922]

Abbreviations

CEDIP: Centre for Digital Psychiatry

Edited by H Witteman; submitted 07.Jan.2026; peer-reviewed by L Lynch, Y Chen; revised version received 11.Jun.2026; accepted 22.Jun.2026; published 31.Jul.2026.

Please cite as:
Kusier AO, Dalsgaard CH, Hoffmann SH, Thygesen LC, Folker AP
Digital Mental Health Promotion Services for Youth: A Qualitative Study of Help-Seeking Through Mindhelper.dk
JMIR Hum Factors 2026;13:e91017
URL: https://humanfactors.jmir.org/2026/1/e91017
doi:10.2196/91017

© Amalie Oxholm Kusier, Caroline Høier Dalsgaard, Sofie Have Hoffmann, Lau Caspar Thygesen, Anna Paldam Folker. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 31.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Factors Associated With Health Perception and Psychological Well-Being Among Third-Age University Students: Cross-Sectional Study on the Role of Cyberchondria

Semra Gündoğdu[1*], PhD; Özlem Özgür[2], PhD; Asli Kilavuz[3*], MD, PhD; İsmail Tufan[2], Prof Dr, PhD

[1]Department of Internal Medicine Nursing, Faculty of Nursing, Akdeniz University, Antalya, Türkiye

[2]Department Of Gerontology, Faculty Of Health Sciences, Akdeniz University, Antalya, Türkiye

[3]Department of Internal Medicine, Division of Geriatrics, Faculty of Medicine, Ege University, 5th floor, Bornova, Izmir, Türkiye

[*]these authors contributed equally

Corresponding Author:
Asli Kilavuz, MD, PhD
Department of Internal Medicine, Division of Geriatrics, Faculty of Medicine, Ege University, 5th floor, Bornova, Izmir, Türkiye

Abstract

Background: With Türkiye's aging population and widespread internet usage, the excessive seeking of health information through online platforms, formally identified as cyberchondria, has emerged as a concern affecting older individuals' health perception and psychological well-being.

Objective: This study aimed to evaluate the association of cyberchondria with health perceptions and psychological well-being among older participants in a third-age university program.

Methods: This cross-sectional study included 352 participants aged ≥60 (mean 67.4, SD 4.7) years from Tazelenme University, Antalya, Türkiye, and was conducted between November and December 2024. Data were collected using the Cyberchondria Severity Scale Short Form (CSS-12), Individual Health Perception Scale, and Psychological Well-Being Scale for Older People (PWBS-OP). Statistical analyses included correlation and multiple regression analyses.

Results: The mean age of the 352 participants was 67.4 (SD 4.7, range 60-87) years, 38.6% (n=136) were first-year students, 69.6% (n=245) were men, 58.8% (n=207) were married, 48.9% (n=172) were university graduates or above, and 65.1% (n=229) had chronic diseases. Significant negative correlations were found between CSS-12 distress, compulsion, and total scores with both Individual Health Perception Scale and PWBS-OP scores (P<.05). According to the multiple linear regression analysis, the presence of chronic disease was the only significant positive predictor of higher health perception levels (B=2.16, 95% CI 1.721-3.593; P=.003), whereas factors such as age, gender, education level, and CSS-12 total scores did not demonstrate a statistically significant impact (P>.05). Psychological well-being (PWBS-OP) was significantly and positively predicted by age (B=0.22, 95% CI 0.037-0.399; P=.02), marital status (B=1.67, 95% CI 1.064-2.716; P=.002), economic status (B=2.10, 95% CI 1.729-3.464; P=.003), and the thought of having an undiagnosed disease (B=3.38, 95% CI 1.038-5.830; P=.007), whereas it was significantly and negatively predicted by medical examinations in the past year (B=−1.29, 95% CI −2.093 to 0.485; P=.002), undergoing examinations without a physician's recommendation (B=−2.04, 95% CI −3.794 to −0.290; P=.02), searching for health-related topics on the internet (B=−1.26, 95% CI −2.152 to −0.360; P=.006), and CSS-12 total scores (B=−0.13, 95% CI −0.261 to −0.060; P=.04).

Conclusions: High cyberchondria levels significantly impair older adults' psychological well-being and health perception. While demographic factors positively influence health perception, excessive internet-based health seeking deteriorates psychological well-being. Digital health literacy programs, professional online health counseling, and psychological support should be recommended to target cyberchondria risks in older populations.

(JMIR Hum Factors 2026;13:e83371) doi:10.2196/83371

KEYWORDS

cyberchondria; third-age university; health perception; psychological well-being; older adults

Introduction

In alignment with global demographic shifts, Türkiye is experiencing a significant increase in its older population. In 2024, the older population was 10.6% of Türkiye's total population and is expected to increase to 13.5% by 2030 [1]. Parallel to this demographic transition, technology has become an indispensable component of daily activities, education, and health care [2]. Modern society necessitates that older adults

XSL•FO
RenderX

engage with new information and communication technologies to maintain independence, safety, and social connectivity. Research shows that the percentage of individuals aged ≥65 years who search for health information online rose from 14.5% in 2011 to 43.6% in 2020 [3]. Consequently, internet usage among older adults in Türkiye reached 46.9% in 2024, with health-related information seeking emerging as one of the most frequent online activities [1]. While these digital tools empower older adults in health management and decision-making, the unregulated nature of online information has introduced a distinct challenge: the phenomenon of cyberchondria.

Cyberchondria is defined as the excessive and compulsive seeking of medical information online, a behavior typically driven by and resulting in heightened health anxiety [4-6].

The dissemination of misleading or incomplete information, combined with the internet's capacity to reinforce preexisting fears, creates a reinforcing cycle in which search attempts intended to provide relief instead exacerbate distress [7]. This information pollution directly undermines efforts to achieve "complete well-being," which the World Health Organization (WHO) defines not merely as the absence of disease, but as a state of total physical, mental, and social wellness [8]. Within this holistic framework, cyberchondria poses a significant threat to an individual's subjective evaluation of their own health and their overall mental equilibrium.

The subjective assessment of an individual's physical and mental state is conceptualized as "health perception." Health perception involves a complex blend of feelings and thoughts that may not always align with objective, clinically determined health status [9].

In the context of cyberchondria, intensive exposure to disease symptoms online can lead individuals to select and adapt these symptoms to their own self-image, potentially deteriorating their health perception even in the absence of clinical illness.

Beyond perception, this behavior also impacts "psychological well-being," a multidimensional construct encompassing self-acceptance, environmental mastery, and autonomy. When older adults encounter inconsistent or frightening health information, it can diminish their sense of environmental mastery and autonomy, thereby reducing their overall quality of life [7].

To mitigate these risks and promote active aging, "lifelong learning" programs have become vital interventions. In Türkiye, the 60+ Tazelenme University model provides a platform for education and social engagement that has been shown to improve life satisfaction and self-health evaluations among participants. However, while these students are encouraged to use technology for healthy aging, their vulnerability to cyberchondria and its subsequent impact on their psychological and perceptual health remains an area requiring closer investigation.

A study also showed that older participants from third-age universities improved life satisfaction, subjective well-being, and health habits [10]. Third-age university students showed improved self-care behaviors in comparison to their nonparticipant peers and reported better self-health evaluations [11]. This reveals that education and social engagement help senior citizens with their health perception and psychological wellness. However, the amount and quality of online health information, to some extent, is disruptive, especially the cyberchondria phenomenon, which manifests as excessive and anxiety-driven searches. It is likely to exacerbate the perception and mental health of older adults.

Third-age university students are an illustrative example of the so-called "active agers," and this group has a higher degree of motivation for self-care and social involvement than the average older adult. Nevertheless, the paradox of this greater digital involvement stems from the fact that, although this group, being engaged in education, ought to be the most psychologically protected, the existence of an "at-risk" phenomenon, in health-related internet searching, may, for multiple reasons, increase their exposure to the internet and, as a consequence, increase health-related misinformation, health-related anxiety, and cyberchondria. As of now, the literature has completely overlooked the question of whether the environment of lifelong education has the potential to counteract the negative psychological effects of cyberchondria. For this reason, this study aimed to provide answers to the following hypotheses:

- Hypothesis 1: Higher levels of cyberchondria are correlated to lower health perceptions of older adults participating in lifelong learning programs.
- Hypothesis 2: Higher levels of cyberchondria are correlated to lower psychological well-being of those individuals.
- Hypothesis 3: Specific sociodemographic and digital usage patterns are significant predictors of health perception and well-being within this educational context.

Given this context, understanding the relationship between internet-based health seeking and the holistic well-being of older adults is essential for developing effective digital health literacy and psychological support programs. Therefore, this study aimed to evaluate the association of cyberchondria with health perceptions and psychological well-being among older participants enrolled in a third-age university program.

Methods

Study Design

This study is a descriptive and cross-sectional study conducted at the Tazelenme University of Antalya Campus between November and December 2024. 60+ Tazelenme University is a third-age university established as a social responsibility project that has been providing 4 years of free education since 2016 with the principle of lifelong learning for individuals aged ≥60 years.

Study Sample

The research population comprised 643 Tazelenme University students enrolled in the program conducted within the scope of Akdeniz University Center for Applied Research on Aging Studies (1st grade: n=215, 33.4%; 2nd grade: n=184, 28.6%; 3rd grade: n=124, 19.3%; and 4th grade: n=120, 18.7% students) [12]. The sample size was calculated using the "Sample Size for % Frequency in a Population" module in the OpenEpi program [13] for a cross-sectional study. The population size

XSL•FO
RenderX

was taken as 643, the expected frequency or proportion was determined as 50% because no prior estimate was available, the confidence level was set at 99%, the absolute margin of error was set at 5%, and the design effect was set at 1.0 under the assumption of simple random sampling. Under these assumptions, the minimum sample size required was calculated as 327. Inclusion criteria for the research were being aged ≥60 years, being a Tazelenme University student, and volunteering to participate in the research. Noninclusion criteria for the study included being aged <60 years, not being officially enrolled in the specified university campus, and not having internet access or a valid email address, as an online survey platform was used for data collection. Exclusion criteria included surveys containing incomplete or inaccurate data and withdrawal of consent at any point during the survey process.

Data Collection Tools

Data were collected using a "Personal Information Form," "Cyberchondria Severity Scale Short Form (CSS-12)" [14], Individual Health Perception Scale (IHPS) [15,16], and the Psychological Well-Being Scale for the Older People (PWBS-OP) [17].

The data were collected between November and December 2024 using an online survey form created with Google Forms. The survey form was sent to the students' email addresses. The survey took approximately 15 minutes to complete. The research surveys were conducted in a single session.

Personal Information Form

Overview

This form was created by the researchers and included questions related to demographic information, such as age, class (1, 2, 3, or 4), gender (female or male), marital status (married or single), family type (nuclear or extended), and educational status (the educational status of the participants refers to the highest level of formal education completed before joining the third-age university program). As 60+ Tazelenme University is a social responsibility project based on the principle of lifelong learning, it accepts students from all educational backgrounds, provided they meet the age requirement (primary school, secondary school, high school, university, or higher), marital status (married, divorced, widowed, or single), socioeconomic status (income less than expenditure, income equals expenditure, and more income than expenditure), presence of health insurance (yes or no), smoking status (nonsmoker, ex-smoker, or current smoker), living arrangement (with spouse, alone, spouse and children, or children), number of physician visits in the past year (never, 1 - 3 times, 4 - 6 times, 7 - 9 times, or ≥10 times), daily internet use duration (1 - 2 h, 3-4 h, 5-6 h, or >6 h), the source for accessing health information (health care professionals, TV or radio, printed press, internet, family members, or friends).

CSS-12

The scale was developed by McElroy et al [14]. The Turkish validity and reliability study of the scale was conducted by Tuğtekin and Tuğtekin, reporting a Cronbach α coefficient of 0.92 for the overall scale [18]. In this study, the internal consistency coefficient (Cronbach α) for the total CSS-12 was calculated as an α of 0.75. The scale consists of 4 dimensions and 12 statements. The scale is a 5-point Likert-type scale (1=never, 2=rarely, 3=sometimes, 4=often, and 5=always).

A total score between 12 and 60 can be obtained from the scale. An increase in the score obtained from the scale is interpreted as an increase in the level of cyberchondria. The total cyberchondria score is calculated by summing the scores obtained from each question.

The CSS-12 consists of 4 dimensions: excessiveness (EXC), which assesses increasing or repeated health-information searching (items 1, 3, and 6); distress (DIST), which assesses anxiety or distress resulting from searching (items 4, 8, and 9); reassurance (REAS), which assesses searching attempts that lead individuals to seek medical support (items 5, 11, and 12); and compulsion (COMP), which assesses searching that may interfere with online or other aspects of daily life (items 2, 7, and 10) [18].

IHPS

The Health Perception Scale, used to assess health perception, was developed by Diamond et al [15]. The validity and reliability of the Turkish version were established by Kadıoğlu and Yıldız, who reported a Cronbach α coefficient of 0.77 for the total scale [16]. For this study sample, the internal consistency of the IHPS was found to be an α of 0.69. It is a 5-point Likert-type scale consisting of 15 items and 4 subfactors (control center, self-awareness, certainty, and importance of health). The scale consists of 6 items with positive statements (1, 5, 9, 10, 11, and 14) and 9 items with negative statements (2, 3, 4, 6, 7, 8, 12, 13, and 15). Each item is scored as follows: "1, strongly disagree"; "2, disagree"; "3, undecided"; "4, agree"; and "5, strongly agree." Negative statements are scored in reverse (5=strongly disagree, 4=disagree, 3=undecided, 2=agree, and 1=strongly agree). The scale ranges from a minimum score of 15 to a maximum score of 75. A high score indicates a high perception of health, while a low score indicates a low perception of health.

Control center (COM) subdimension aims to determine whether the individual attributes being healthy to factors outside themselves (luck, fate, or religious belief), that is, whether they gather the control center in themselves for being healthy and their confidence in themselves to change their health (items 2, 3, 4, 12, and 13).

Self-awareness (SAW) subdimension aims to determine the level of the individual's belief about whether being healthy is in their own hands regarding their self-awareness perception about exercise and proper nutrition for being healthy (items 5, 10, and 14).

Certainty (CER) subdimension aims to determine whether the individual has a definite idea about what they should do to stay healthy and be healthier (items 6, 7, 8, and 15).

Importance of health (IOH) subdimension aims to determine how much importance the individual gives to their health, to what extent they make financial sacrifices in this regard, and

XSL•FO
RenderX

whether the importance they give to health is one of the priorities in their life (items 1, 9, and 11) [16].

PWBS-OP

The scale was developed and validated for the Turkish older population by Gümüş Demir, with a reported Cronbach α of 0.89 [17]. In this study, the internal consistency coefficient for the PWBS-OP was determined to be an alpha of 0.92. PWBS-OP consists of 15 items and a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). No cutoff score has been determined for the scale. The minimum score that can be obtained from the scale is 15, and the maximum score is 75. Items 8 and 13 are scored in reverse. Higher scores indicate a higher level of well-being [17].

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki. The study was approved by the Akdeniz University Medical Scientific Research Ethics Committee (TBAEK-511, dated July 25, 2024). The necessary legal permissions were obtained from the Directorate of the Center for Applied Research on Aging to conduct the research. In addition, informed consent was obtained from all participants who volunteered to participate in this study.

Statistical Analysis

The study data were analyzed using SPSS 25.0 for Windows (SPSS Inc). For descriptive findings, categorical variables were presented as numbers and percentages, while continuous variables were presented as mean (SD) and minimum to maximum values. The Kolmogorov-Smirnov normality test was used to determine whether the data showed a normal distribution. It was determined that the data, except for the CSS-12 total score, did not conform to normal distribution. In the analysis of nonparametric data, the Mann-Whitney *U* test, Kruskal-Wallis test, and Spearman correlation were used. For variables where the difference was significant, pairwise comparisons between groups were evaluated using the post hoc Dunn test (with Bonferroni correction). Pearson correlation analysis was used in the analysis of parametric data. Factors playing a role in predicting participants' health perception and psychological well-being levels were evaluated using multiple linear regression analysis. Variables found to be statistically significant in the pairwise comparison were included in the regression model. Categorical variables were included in the regression analysis by defining one dummy variable (coded as 0 - 1) representing the number of categories in the variable. Multiple linear regression models were created using the forward selection method. The normality of the residuals obtained from the final regression model was checked using the Kolmogorov-Smirnov test. All the findings were tested at a significance level of 0.05.

Results

Characteristics of Participants

The mean age of the 352 participants was 67.4 (SD 4.7, range 60-87) years. Of the 352 participants, 136 (38.6%) were first-year students, 245 (69.6%) were men, 207 (58.8%) were married, and 172 (48.9%) were university graduates or above (Table 1). No statistically significant difference was found between the daily internet usage times of the 4 classes (*P*=.10).

In terms of health status, 229 (65.1%) participants reported living with at least one chronic disease, and 199 (56.5%) spent between 1 and 2 hours on the internet daily. Participants demonstrated a high level of engagement with online health information; 167 (47.4%) participants reported searching the internet before starting a treatment recommended by their physician. While 247 (70.2%) participants were unsure about the reliability of online health resources, official health care organization websites (n=189, 52.8%) and social networks (n=133, 37.8%) were the most preferred sources for health searches. Notably, a small but significant minority (n=15, 4.3%) reported discontinuing medical treatments based on information found online (Table 2).

XSL•FO
RenderX

Table . Sociodemographic characteristics of Tazelenme University students (N=352).

Characteristics	Values
Age (years), mean (SD)	67.4 (4.7)
Class, n (%)	
Class 1	136 (38.6)
Class 2	78 (22.2)
Class 3	99 (28.1)
Class 4	39 (11.1)
Gender, n (%)	
Female	107 (30.4)
Male	245 (69.6)
Marital status, n (%)	
Married	207 (58.8)
Divorced	70 (19.9)
Widowed	63 (17.9)
Single	12 (3.4)
Education level[a], n (%)	
Primary school	25 (7.1)
Secondary school	16 (4.5)
High school	139 (39.5)
University or higher	172 (48.9)
Smoking, n (%)	
Nonsmoker	147 (41.8)
Ex-smoker	158 (44.9)
Current smoker	47 (13.4)
Economic status, n (%)	
Income<expenditure	74 (21)
Income=expenditure	218 (61.9)
Income>expenditure	60 (17)
Living arrangement, n (%)	
Spouse	149 (42.3)
Alone	112 (31.8)
Spouse and children	55 (15.6)
Children	30 (8.5)
Relatives	6 (1.7)
Health insurance, n (%)	
Yes	345 (98)
No	7 (2)

[a]Educational status represents the participants' formal education level prior to their enrollment at Tazelenme University

XSL•FO
RenderX

Table . Medical data of participants and their health information search characteristics on the internet (N=352).

Characteristics	Values
Presence of chronic disease, n (%)	
Yes	229 (65.1)
No	123 (34.9)
Number of physician visits in the past year, n (%)	
Never	24 (6.8)
1 - 3 times	159 (45.2)
4 - 6 times	104 (29.5)
7 - 9 times	29 (8.2)
≥10 times	36 (10.2)
Daily internet use duration, n (%)	
1 - 2 hours	199 (56.5)
3 - 4 hours	114 (32.4)
5 - 6 hours	28 (8.0)
>6 hours	11 (3.1)
Unrecommended medical testing in the past year, n (%)	
Yes	117 (33.2)
No	235 (66.8)
Internet-based self-diagnosis and self-medication in the past year, n (%)	
Yes	13 (3.7)
No	339 (96.3)
Belief of having an undiagnosed illness, n (%)	
Yes	47 (13.4)
No	305 (86.6)
Frequency of searching for health topics online, n (%)	
Always	33 (9.4)
Often	52 (14.8)
Sometimes	162 (46.0)
Rarely	85 (24.1)
Never	20 (5.7)
Internet research before starting prescribed treatment, n (%)	
Yes	167 (47.4)
No	185 (52.6)
Internet-driven treatment discontinuation, n (%)	
Yes	15 (4.3)
No	337 (95.7)
Perceived reliability of online health resources, n (%)	
Not reliable	73 (20.7)
Reliable	32 (9.1)
Not sure	247 (70.2)
Internet resources used for health searches[a,b], n (%)	
Social networks	133 (37.8)
Blogs	19 (5.4)

XSL•FO
RenderX

Characteristics	Values
Forums	14 (4.0)
Personal websites	63 (17.9)
Health organization websites	186 (52.8)
Newspaper articles, news stories, or academic articles	119 (33.8)

[a]Based on a multiple-response question, participants could select more than one option.

[b]Percentages are calculated based on the total number of participants (N=352) and may exceed 100% due to multiple selections.

Score Averages of All Scales

Descriptive statistics for the assessment tools revealed a mean CSS-12 total score of 28.50 (SD 6.96), indicating the sample's general level of cyberchondria. The mean scores for health perception (IHPS) and psychological well-being (PWBS-OP) were 53.74 (SD 6.53) and 58.82 (SD 8.41), respectively. Within the subdimensions of the CSS-12, "excessiveness" (mean 8.31, SD 2.72) and "reassurance seeking" (mean 7.57, SD 2.45) were observed to have the highest mean values (Table 3).

Table . Subscale scores and total score averages of all scales.

Subscales and total scores	Mean (SD)	Minimum to maximum
CSS-12[a]-excessiveness	8.31 (2.72)	3 - 15
CSS-12-distress	7.40 (2.39)	3 - 15
CSS-12-reassurance	7.57 (2.45)	3 - 15
CSS-12-compulsion	5.21 (2.27)	3 - 15
CSS-12-total Score	28.50 (6.96)	12 - 53
IHPS[b]-COM[c]	18.01 (3.62)	8 - 25
IHPS-SAW[d]	11.05 (2.09)	5 - 15
IHPS-CER[e]	13.39 (3.11)	4 - 20
IHPS-IOH[f]	10.56 (1.94)	4 - 15
IHPS-total score	53.74 (6.53)	39 - 75
PWBS-OP-total score[g]	58.82 (8.41)	30 - 74

[a]CSS-12: Cyberchondria Severity Scale Short Form.

[b]IHPS: Individual Health Perception Scale.

[c]COM: control center.

[d]SAW: self-awareness.

[e]CER: certainty.

[f]IOH: importance of health.

[g]PWBS-OP: Psychological Well-Being Scale for the Older People.

The Relationship of Sociodemographic Characteristics, Medical Data, and Internet Usage With IHPS and PWBS-OP

Comparative analyses indicated that health perception levels were significantly higher among male participants (P=.02) and those with a high school education or above (P=.047; Table 4). Conversely, psychological well-being was found to be lower in participants who were married (P=.002), those with lower income levels (P<.001), and those who reported thinking they might have an undiagnosed illness (P<.001). The psychological well-being level of older people living with their spouse and children was found to be lower than that of those living alone (P=.02; Table 4). The analysis demonstrated no statistically significant differences in IHPS and PWBS-OP total scores across classes, smoking, and presence of health insurance, all failing to reach the formal significance threshold (P>.05).

According to the pairwise comparisons in the Dunn test with Bonferroni correction displayed, PWBS-OP total scores differed significantly only between married participants (mean rank 162.70, SD 100.96) and widowed or single participants (mean rank 210.20, SD 101.10; adjusted P=.002). No statistically significant differences were observed between married and divorced participants (adjusted P=.56) or between divorced and widowed or single participants (adjusted P=.26). Individuals who are widowed or single tend to have higher total PWBS-OP scores. The statistically significant variance in health perception based on education level identified in the overall Kruskal-Wallis analysis (P=.047) was not robust enough to be sustained during pairwise comparisons after applying the Dunn test with Bonferroni correction designed to mitigate type I error.

XSL•FO
RenderX

Nonetheless, an examination of the mean ranks indicates that older adults with a primary or middle school education level tended to exhibit lower health perception compared to those with high school (adjusted *P*=.06) and university (adjusted *P*=.06) degrees. Pairwise comparisons in the Dunn test with Bonferroni correction were performed to investigate the specific differences in PWBS-OP total scores across perceived economic status groups. The post hoc analysis revealed that participants who perceived their income as "greater than expenses" (mean rank 232.97, SD 86.05) exhibited statistically significant and substantially higher PWBS-OP total scores compared to both the "income equal to expenses" group (mean rank 170.95, SD 100.08; test statistic=−62.019; *P*<.001) and the "income less than expenses" group (mean rank 147.07, SD 101.72; test statistic=−85.892; *P*<.001). Conversely, the difference in PWBS-OP total scores between the "income equal to expenses" and "income less than expenses" groups was not statistically significant after applying the Dunn test with Bonferroni correction (test statistic=−23.873; *P*=.24). Following the significant overall Kruskal-Wallis test (*P*=.02), pairwise comparisons in the Dunn test with Bonferroni correction were performed to identify the source of variation across living arrangement groups. The post hoc analysis confirmed that participants living alone (mean rank 197.04, SD 98.23) exhibited statistically significant higher PWBS-OP total scores compared to those living with their spouse and children (mean rank 143.92, SD 97.82; test statistic=53.122; adjusted *P*=.009).

Table . Differences in IHPS[a] and PWBS-OP[b] scores according to the sociodemographic characteristics of participants.

Characteristics	IHPS total score		PWBS-OP total score	
	Mean (SD)	*P* value	Mean (SD)	*P* value
Gender		.02[c]		.15[c]
Male	55.0 (5.6)		57.9 (8.4)	
Female	53.2 (6.8)		59.2 (8.4)	
Marital status		.44[d]		.002[d]
Married	53.6 (6.3)		57.7 (8.4)	
Divorced	54.6 (6.6)		59.2 (7.9)	
Widowed or single	53.3 (7.1)		61.6 (8.5)	
Educational status		.047[d]		.05[d]
Primary or secondary school	51.2 (5.6)		60.8 (9.3)	
High school	54.2 (6.8)		58.7 (8.1)	
University or higher	54.0 (6.4)		58.5 (8.4)	
Economic status		.44[d]		<.001[d]
Income<expenditure	52.9 (6.6)		56.2 (8.9)	
Income=expenditure	54.0 (6.6)		58.4 (8.2)	
Income>expenditure	53.7 (6.4)		63.5 (6.5)	
Living arrangement		.18[d]		.02[d]
Spouse	54.4 (6.2)		58.6 (8.1)	
Alone	53.7 (6.9)		60.6 (7.9)	
Spouse and children	52.4 (6.9)		56.5 (8.3)	
Children or relatives	53.3 (6.1)		57.7 (10.2)	

[a]IHPS: Individual Health Perception Scale.

[b]PWBS-OP: Psychological Well-Being Scale for the Older People.

[c]Mann-Whitney *U* test.

[d]Kruskal-Wallis test.

Participants without chronic illnesses had significantly higher levels of perceived health (*P*=.001). PWBS-OP total score of those with chronic disease was determined to be lower than that of those that of without chronic disease (*P*=.008). The IHPS score average of older people who do not think they have an undiagnosed disease is higher than that of those who have such thoughts (*P*=.02). PWBS-OP total score of those who never went to a physician examination in the past year was found to be higher than that of those who did (*P*<.001). In the Dunn test with Bonferroni correction, older individuals who had not had a physician's examination in the past year were found to have higher PWBS-OP total scores than other groups. The Dunn test

XSL•FO
RenderX

with Bonferroni correction revealed statistically significant differences between those who visited a physician 1 to 3 times and those who visited 7 to 9 times (P=.01), between those who visited 7 to 9 times and those who did not visit at all (P=.001), between those who visited 10 times and those who did not visit at all (P=.02), and between those who did not visit at all and those who visited 4 to 6 times (P=.03). PWBS-OP total score of those who had medical tests done without a physician's recommendation in the past year was determined to be higher than those who did not have tests done (P=.002). PWBS-OP total score of those who do not think they have an undiagnosed disease was found to be higher than those who have such thoughts (P<.001). PWBS-OP total score of those who stated that they "always" research health-related topics on the internet was found to be higher than the groups giving other answers (P=.02). The Dunn test with Bonferroni correction showed that those who always search for health information online had the highest total PWBS-OP score compared to other groups. In pairwise comparisons, statistically significant differences were found between those who frequently search for health information online and those who always do so (P=.08) and between those who always do so and those who sometimes do so (P=.03; Table 5). The analysis demonstrated no statistically significant differences in IHPS and PWBS-OP total scores across daily internet use duration, internet-based self-diagnosis and self-medication in the past year, internet research before starting prescribed treatment, internet-driven treatment discontinuation, and perceived reliability of online health resources, all failing to reach the formal significance threshold (P>.05).

Table . Comparison of IHPS[a] and PWBS-OP[b] scores based on participants' medical data and internet usage habits.

Characteristics	IHPS total score		PWBS-OP total score	
	Mean (SD)	*P* value	Mean (SD)	*P* value
Presence of chronic disease		.001[c]		.008[c]
Yes	52.9 (6.4)		57.0 (8.3)	
No	55.4 (6.6)		60.4 (8.4)	
Number of physician visits in the past year		.48[d]		<.001[d]
Never	55.3 (7.4)		63.4 (7.8)	
1 - 3	53.9 (6.4)		60.1 (7.8)	
4 - 6	53.8 (6.9)		57.9 (7.8)	
7 - 9	51.8 (5.2)		54.8 (7.7)	
≥10	53.4 (6.6)		55.9 (10.9)	
Unrecommended medical testing in the past year		.31[c]		.002[c]
Yes	54.2 (6.5)		60.8 (7.0)	
No	53.5 (6.5)		57.8 (8.8)	
Belief of having an undiagnosed illness		.02[c]		<.001[c]
Yes	51.4 (5.5)		53.8 (9.5)	
No	54.1 (6.6)		59.6 (7.9)	
Frequency of searching for health topics online		.86[d]		.02[d]
Always	54.7 (9.1)		63.4 (6.6)	
Often	53.6 (6.9)		57.0 (8.1)	
Sometimes	53.4 (5.9)		58.4 (8.3)	
Rarely	53.9 (6.1)		59.4 (7.9)	
Never	54.6 (7.1)		57.2 (11.7)	

[a]IHPS: Individual Health Perception Scale.

[b]PWBS-OP: Psychological Well-Being Scale for the Older People.

[c]Mann-Whitney *U* test.

[d]Kruskal-Wallis test.

To delineate specific differences across physician visit frequency groups following the general Kruskal-Wallis analysis, pairwise comparisons with Bonferroni correction were conducted. Significant variations were observed among groups. Specifically,

XSL•FO
RenderX

participants reporting no physician visits (mean rank 234.71, SD 100.06) exhibited significantly higher PWBS-OP scores compared to those visiting 7 to 9 times (mean rank 123.33, SD 92.66; test statistic=111.381; adjusted *P*=.001), 10 or more times (mean rank 153.40, SD 108.34; test statistic=81.306; adjusted *P*=.02), and 4 to 6 times (mean rank 165.12, SD 95.99; test statistic=69.588; adjusted *P*=.03). Furthermore, participants with 1 to 3 visits (mean rank 190.08, SD 99.74) also demonstrated significantly higher scores than those with 7 to 9 visits (mean rank 123.33, SD 92.66; test statistic=66.757; adjusted *P*=.01). No statistically significant differences were found between the "none" and "1 to 3 times" groups (adjusted *P*=.45) or among the groups with 4 or more visits. Participants reporting that they "always" search for health-related topics on the internet (mean rank 229.71, SD 87.80) exhibited significantly higher PWBS-OP scores compared to those searching "frequently" (mean rank 154.14, SD 101.58; test statistic=75.568; adjusted *P*=.008) and "sometimes" (mean rank 171.25, SD 99.94; test statistic=58.459; adjusted *P*=.03). No statistically significant differences were found among other pairwise comparisons after applying the stringent Bonferroni correction, notably between the "always" group and the "never" (adjusted *P*>.99) or "rarely" (adjusted *P*=.20) groups.

Correlations Between Participants' Characteristics and CSS-12, IHPS, and PWBS-OP Scores

In the research, a negative and low-level significant correlation was found between the CSS-12 compulsion subscale and the COM subscale score of IHPS and between participants' age and the SAW subscale score of IHPS (*P*<.001 and *P*<.05, respectively). A negative and low-level significant correlation was found between CSS-12-distress, CSS-12-compulsion, and CSS-12-total score and the CER subscale score of IHPS (*P*<.001). A positive and low-level significant correlation was found between CSS-12-reassurance and CSS-12-total Score and the IOH subscale score of IHPS (*P*<.05). A positive and low-level significant correlation was found between participants' age and IHPS total score and PWBS-OP total score (*P*<.05). A negative and low-level significant correlation was found between CSS-12-distress, CSS-12-compulsion, and CSS-12-total score and IHPS total score and PWBS-OP total score (*P*<.05; Table S1 in Multimedia Appendix 1).

Multiple Linear Regression Analysis of Factors Affecting Health Perception and Psychological Well-Being in Highly Educated Older Adults Engaged in Lifelong Learning

According to the multiple linear regression analysis, the presence of chronic disease was the only significant positive predictor of higher health perception levels (*B*=2.16, 95% CI 1.721-3.593; *P*=.003), whereas factors such as age, gender, education level, and CSS-12 total scores did not demonstrate a statistically significant impact (*P*>.05). Psychological well-being (PWBS-OP) was significantly and positively predicted by age (*B*=0.22, 95% CI 0.037-0.399; *P*=.02), marital status (*B*=1.67, 95% CI 1.064-2.716; *P*=.002), economic status (*B*=2.10, 95% CI 1.729-3.464; *P*=.003), and the thought of having an undiagnosed disease (*B*=3.38, 95% CI 1.038-5.830; *P*=.007), whereas it was significantly and negatively predicted by medical examinations in the past year (*B*=−1.29, 95% CI −2.093 to 0.485; *P*=.002), undergoing examinations without a physician's recommendation (*B*=−2.04, 95% CI −3.794 to −0.290; *P*=.02), searching for health-related topics on the internet (*B*=−1.26, 95% CI −2.152 to −0.360; *P*=.006), and CSS-12 total scores (*B*=−0.13, 95% CI −0.261 to −0.060; *P*=.04; Table S2 in Multimedia Appendix 1).

Discussion

Principal Findings

In this study, increased levels of cyberchondria were found to be associated with significant decreases in both health perception and psychological well-being. The findings demonstrated that age, marital status, economic status, and the absence of undiagnosed illness had significant positive effects on psychological well-being. Particularly, the absence of undiagnosed illness positively influenced the psychological well-being of highly educated older adults engaged in lifelong learning.

Interestingly, this study observed slight but significant increases in health perception and psychological well-being as age advanced. This phenomenon has been explained in the literature by factors such as increased life experience, acceptance, and meaning-seeking among older individuals [19]. However, this finding is not consistent across all studies. For instance, Vuorre and Przybylski [20] reported in their analysis across 168 countries that the association between internet and mobile technology use and well-being did not differ significantly by age and was generally minimal. This situation may vary according to cultural factors, individual coping styles, and levels of digital awareness among highly educated older adults engaged in lifelong learning.

The significantly lower psychological well-being levels found among women compared to men are also noteworthy. Chen et al [21] indicated that older women are psychologically more vulnerable due to higher loneliness, lack of social support, and caregiving responsibilities. Similarly, Li et al [22] revealed that women developed higher anxiety when facing uncertainty in the pandemic context. Older women often serve as primary "health managers" or informal caregivers within family networks. When navigating digital health ecosystems, familial responsibility can manifest as a heavy psychological burden. Unlike men, older women tend to be more meticulous and overly cautious in online symptom checking, driven by a sense of duty to protect both their own well-being and the well-being of those they are responsible for. As web-based health resources generate highly ambiguous medical information, this extensive data searching behavior without formal clinical filtering triggers a cycle of cognitive fatigue and cyberchondria. Consequently, the low psychological well-being in older women may not only be a reflection of structural vulnerability but also a direct result of the overly cautious, digitally driven overanalyzing loop that disproportionately affects highly educated older women engaged in lifelong learning.

The negative relationship observed between worsening economic conditions and psychological well-being reveals that

XSL•FO
RenderX

financial security serves as a fundamental buffering mechanism for mental health in older adults [23]. Economic vulnerability not only increases individual stress levels but also damages well-being by limiting social participation and sense of belonging [6]. For highly educated older adults engaged in lifelong learning, quality of life is not just about food and shelter; it is about attending courses, buying books, and participating in intellectual communities. The economic crisis directly undermines this lifestyle and identity. What they are experiencing is not just economic stress, but an identity disruption. The person wants to be educated, knowledgeable, and intellectually active, but their budget does not allow it. This gap (higher education vs lower economic status) creates deep psychological unease in their personality. As educated individuals can read macroeconomic trends, inflation curves, and the future more rationally, they are more susceptible to a wave of anticipatory anxiety about the future.

In individuals with thoughts of undiagnosed illness, lower levels of health perception and psychological well-being were identified. This finding demonstrates that difficulty coping with uncertainty and diagnostic ambiguity can trigger cyberchondria behaviors and anxiety [22,24]. Highly educated older adults engaged in lifelong learning have solved lifelong problems by "reading and researching." When faced with an undiagnosed symptom, they exhibit the same reflex and turn to the internet. Highly educated older adults without medical training, however, become overwhelmed by the manipulative information pool on the internet. Owing to their high digital literacy, they conduct far deeper searches than the average user. This, instead of reducing uncertainty, leads to cognitive overload. The greatest bulwark of psychological well-being in old age is "autonomy" and "cognitive control." Consequently, within this specific cohort, the erosion of psychological well-being is not merely a consequence of fearing a physical illness but rather a reflection of a deeper existential crisis: the collapse of their cognitive autonomy and self-efficacy, which are fundamental pillars of successful aging when confronted with an unresolvable digital information maze.

This study found that more frequent physician visits in the past year, undergoing medical tests without physician recommendation in the past year, internet-based health research behavior, and cyberchondria behavior were significantly and negatively associated with the psychological well-being levels of highly educated older adults engaged in lifelong learning. The literature also indicates that excessive and inappropriate searching for health information online was associated with an increased likelihood of health anxiety and decreased psychological well-being in individuals [25]. Particularly in highly educated older adults engaged in lifelong learning, despite increased access to technological information, health anxiety and stress symptoms are associated with cyberchondria [3,7].

Intensive and complex health information online, especially in older adults with high health anxiety, is associated with an increased likelihood of worry and fear, which may be linked to impaired psychological well-being. Additionally, using information obtained from the internet without seeking expert opinion can negatively associate with health perception by causing misdiagnosis and unnecessary tests. As web-based health information is inherently ambiguous and probabilistic, this intensive data consumption without formal clinical filtration intensifies latent health anxiety rather than mitigating it [9]. The unique pathway-based insight offered by this study is that this digitally driven anxiety directly translates into concrete, uncoordinated medical overconsumption. To resolve the cognitive dissonance generated by alarmist online information, these highly educated older adults engaged in lifelong learning bypass professional triage, leading to frequent physician visitations and self-directed, unrecommended diagnostic testing. Every ambiguous test result or clinical nuance, in turn, sparks further online cross-referencing, reinforcing the cyberchondria cycle and severely deteriorating both their psychological well-being and their trust in health care systems [26]. Consequently, our findings emphasize that in highly educated older cohorts, technology use is a double-edged sword; high digital literacy, when uncoupled from the ability to filter out cyberchondria-inducing medical noise, transforms a cognitive asset into a profound source of psychological and systemic vulnerability.

The impact of CSS-12's "distress" and "compulsion" subdimensions reveals a distinct process-oriented vulnerability in highly educated older adults. While cyberchondria is widely conceptualized as a compulsive, anxiety-driven pattern [24,27], in this specific cohort, it transforms an ingrained problem-solving orientation into a maladaptive loop. Driven to resolve somatic uncertainty, these digitally literate individuals engage in exhaustive online cross-referencing. However, because they possess the cognitive capacity to analyze complex data but lack formal clinical frameworks to filter it, this ambiguous web-based information precipitates severe cognitive overload. Ultimately, this pathway constitutes an active cognitive trap where an older individual's intellectual resources and drive for autonomy paradoxically undermine their mental resilience, thereby eroding subjective well-being.

The finding that 65% (n=229) of participants managed chronic conditions and underwent multiple examinations highlights a unique process-oriented mechanism within highly educated older adults engaged in lifelong learning. While clinical status traditionally shapes subjective health perception [21,28], a chronic diagnosis in this demographic threatens cognitive autonomy, driving individuals to deploy their digital literacy for intensive online research. However, navigating inherently ambiguous web-based health data creates an overanalysis trap that triggers a compulsive cascade of medical cross-checking and heightened anxiety. This pathway-based dynamic explains why the complete absence of chronic disease history emerged as the strongest positive predictor of health perception. Spared from the somatic triggers that launch them into the exhausting digital maze, these highly educated older adults engaged in lifelong learning avoid cyberchondria-induced cognitive overload, allowing their intellectual resources to function as stable psychological assets rather than anxiety amplifiers.

Particularly in older adults with multiple chronic diseases, deterioration in both physical and psychological functioning is common. Ryff developed the "Psychological Well-being" model, which divides well-being into 6 distinct dimensions: self-acceptance, positive relationships with others, autonomy,

XSL•FO
RenderX

environmental mastery, finding purpose and meaning in life, and personal growth [29]. When psychological well-being is evaluated according to the subdimensions of Ryff's model, cyberchondria behavior can be considered to have negative effects on autonomy and environmental mastery dimensions.

The findings that nearly half (n=167, 47.4%) of the participants engaged in preconsultation internet research and a critical minority (n=15, 4.3%) discontinued medical treatment based on online information expose the behavioral realities of digital health tracking in highly educated older adults engaged in lifelong learning [30]. Evaluating these findings through a process-oriented framework reveals how an older, highly educated demographic's pursuit of medical autonomy can paradoxically manifest as clinical nonadherence. For these intellectually engaged individuals, researching symptoms prior to a physician visit is an active effort to maintain agency over age-related somatic changes. However, when highly educated individuals cross-reference complex, unverified web data with professional clinical advice, it often generates profound cognitive dissonance. When this dissonance is left unresolved, the pathway-based trajectory shifts from proactive self-care to dangerous behavioral autonomy, driving individuals to independently terminate prescribed treatments. This maladaptive pattern clarifies why an increased frequency of digital health seeking correlates negatively with psychological well-being, as unguided online searches systematically deteriorate uncertainty-coping mechanisms and amplify somatic threat perceptions [6,31]. Furthermore, this distinct nuance reconciles the apparent contradiction with macro-level data, such as Vuorre and Przybylski's [20] finding that general broadband adoption does not consistently undermine global well-being. The critical insight here is that while general digital connectivity remains benign on a macro scale [20], a targeted, pathway-based exploration reveals that anxiety-driven medical overanalysis by highly educated older adults engaged in lifelong learning presents a direct, acute threat to their psychological resilience and subjective well-being.

Evaluating technology-induced well-being through a population-specific lens reveals a critical process-oriented nuance in our highly educated older cohort: the positive link between cyberchondria's "reassurance seeking" and health importance reflects an intellectual drive for autonomy rather than pure pathology. This intersection highlights a vital pathway-based pivot; while unguided searches induce anxiety, structured information gathering enhances health self-efficacy [32], converting anxious drives into proactive self-care agency. This adaptive trajectory is mirrored by the participants' sophisticated information seeking hierarchy, which uniquely included academic literature. However, as established by Cao and Chu [33], the perceived credibility of these mixed-reliability platforms fundamentally dictates subsequent psychological well-being. The resulting cognitive friction between official medical narratives and informal digital commentary ultimately determines whether this information pathway reinforces psychological resilience or precipitates cyberchondria.

This study determined that cyberchondria behavior in highly educated older adults engaged in lifelong learning was associated with an increased likelihood of deterioration in mental and physical health conditions and was negatively associated with psychological well-being. Schuster et al [3] similarly reported that older adults with high health anxiety experienced more stress when researching health information online. Additionally, encountering unreliable information sources provides grounds for older adults to discontinue treatment or self-diagnose [3,30,34]. Thus, for highly educated older adults engaged in lifelong learning, unguided digital health seeking serves as a critical vulnerability that undermines both subjective resilience and objective health safety.

The findings regarding participants' daily internet usage duration and source preferences in this study underscore the critical importance of enhancing digital health literacy among highly educated older adults engaged in lifelong learning. Recent evidence from Xu and Starcevic [7] highlights that in the modern digital era, health seeking behavior in older adults is intricately linked to mental well-being and health-related quality of life. Improving digital health literacy is now recognized as a vital protective mechanism that shields older individuals from the adverse effects of misinformation while empowering them to make informed and accurate health decisions. Therefore, initiatives aimed at reducing cyberchondria risk must prioritize the development of advanced digital literacy to align with the contemporary technological landscape.

It is emphasized that exposure to health information available online can create difficulties in accessing accurate information and cognitive fatigue, particularly in highly educated older adults engaged in lifelong learning, and therefore, cyberchondria can have negative effects on psychological well-being. Recent research underscores that this cognitive strain fuels cyberchondria, which in turn serves as a significant predictor of decline in various dimensions of Ryff's psychological well-being model [29], including psychological growth and psychosocial stability [7]. As highly educated older adults engaged in lifelong learning struggle to reconcile conflicting online reports, their "self-acceptance" regarding their health status and "autonomy" in health-related decision-making are compromised, as they become increasingly fused with anxiety-inducing information. Ultimately, this deterioration in eudaimonic well-being dimensions significantly reduces the overall quality of life in this population.

Limitations and Strengths of the Study

This study has several limitations. First, as the research has a cross-sectional design, establishing causal relationships between variables is limited. Second, the use of an online data collection method, while a methodological convenience that facilitated a larger sample size within the university network, inherently introduces concerns regarding data reliability and selection bias. As our sample consists of highly educated older adults engaged in lifelong learning who are already active internet users, individuals with lower digital access or literacy are not represented in the data. This limits the generalizability of our findings to the broader older population in Turkey. The majority of participants being men (approximately 70%) and having high education levels makes it difficult to generalize the findings to the general older population. Furthermore, the possibility that

XSL•FO
RenderX

health care service utilization may have changed following the COVID-19 pandemic was not considered.

The most important strength of the research is addressing the relationship between cyberchondria behavior and psychological well-being in the older population. Thus, digital health seeking behavior and psychological effects in highly educated older adults engaged in lifelong learning, which have not been sufficiently examined in the literature, were determined. By using validated tools such as the CSS-12, this study provides a foundational understanding of digital health seeking behaviors and their psychological impacts on an increasingly tech-savvy older population.

Conclusion and Recommendations

In conclusion, high levels of cyberchondria behavior in highly educated older adults engaged in lifelong learning negatively associate with psychological well-being levels. Furthermore, while sociodemographic factors may have positive effects on health perception, excessive and uncontrolled internet-based health information seeking can negatively impact the psychological well-being of highly educated older adults engaged in lifelong learning. Accordingly, to reduce the information confusion and misguidance that highly educated older adults engaged in lifelong learning encounter when seeking health information online, it is recommended to expand reliable digital health literacy education, increase internet-based health counseling services by health care professionals, and develop psychological support and awareness programs targeting cyberchondria risk. Additionally, conducting longitudinal research, including older adults from different socioeconomic and cultural levels in future studies, will provide stronger evidence regarding the association between cyberchondria and psychological well-being.

Acknowledgments

We thank the 60+ Tazelenme University Akdeniz University Coordination and its students for the scientific study conducted under the leadership of İT.

The authors declare the use of generative AI (GenAI) in the research and writing process. According to the GAIDeT (Generative AI Delegation Taxonomy; 2025), the following tasks were delegated to GenAI tools under full human supervision: translation. The GenAI tool used was DeepL. Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Funding

The authors declared that this study has received no financial support.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: SG, AK, OO, İT
Data curation: SG, OO
Formal analysis: AK, SG
Investigation: SG, AK, OO and İT
Methodology: SG, AK
Project administration: AK, SG, OO and İT
Supervision: AK, İT
Writing – original draft: AK, SG
Writing – review & editing: SG, AK, OO, İT
All authors approved the final version to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional results.
[DOCX File, 20 KB - humanfactors_v13i1e83371_app1.docx]

References

1. Elderly statistics, 2024. TURKSTAT. 2025. URL: https://data.tuik.gov.tr/Bulten/Index?p=Istatistiklerle-Yaslilar-2024-54079 [accessed 2025-08-23]

XSL•FO
RenderX

2. Ozgur O, Ozgun Basibuyuk G. Third-age university students' use of gerontechnological products and their attitudes towards the use of gerontechnological products: the example of 60+ refreshment university. J Aging Long Term Care 2024;6(3):119-137. [doi: 10.51819/jaltc.2023.1108712]
3. Schuster AM, Ghaiumy Anaraky R, Cotten SR. Online health information seeking and the association with anxiety among older adults. Front Public Health 2023;11:1076571. [doi: 10.3389/fpubh.2023.1076571] [Medline: 36844827]
4. Starcevic V, Berle D, Arnáez S. Recent insights into cyberchondria. Curr Psychiatry Rep 2020 Aug 27;22(11):56. [doi: 10.1007/s11920-020-01179-8] [Medline: 32852626]
5. Ambrosini F, Truzoli R, Vismara M, Vitella D, Biolcati R. The effect of cyberchondria on anxiety, depression and quality of life during COVID-19: the mediational role of obsessive-compulsive symptoms and Internet addiction. Heliyon 2022 May;8(5):e09437. [doi: 10.1016/j.heliyon.2022.e09437] [Medline: 35600442]
6. Rahme C, Akel M, Obeid S, Hallit S. Cyberchondria severity and quality of life among Lebanese adults: the mediating role of fear of COVID-19, depression, anxiety, stress and obsessive-compulsive behavior-a structural equation model approach. BMC Psychol 2021 Oct 29;9(1):169. [doi: 10.1186/s40359-021-00674-8] [Medline: 34715930]
7. Xu RH, Starcevic V. Cyberchondria in older adults and its relationship with cognitive fusion, health-related quality of life, and mental well-being: mediation analysis. JMIR Aging 2025 May 21;8:e70302. [doi: 10.2196/70302] [Medline: 40397949]
8. Summary Reports on Proceedings Minutes and Final Acts of the International Health Conference Held in New York From 19 June to 22 July 1946: World Health Organization; 1948.
9. Demirhan H, Eke E. The effects of digitalization on health information search behavior: an investigation on cyberchondria disease. Int J Health Sociol 2021;1(1):1-18. [doi: 10.29228/ijheso.49395]
10. Zielińska-Więczkowska H. Correlations between satisfaction with life and selected personal resources among students of Universities of the Third Age. Clin Interv Aging 2017;12:1391-1399. [doi: 10.2147/CIA.S141576] [Medline: 28919725]
11. Zadworna M. Healthy aging and the University of the Third Age - health behavior and subjective health outcomes in older adults. Arch Gerontol Geriatr 2020;90:104126. [doi: 10.1016/j.archger.2020.104126] [Medline: 32512462]
12. Akdeniz University Center for Applied Research on Aging Studies. Administrative student registry data for the 60+ Tazelenme University Program [unpublished internal report]. : Akdeniz University; 2024.
13. Dean AG, Sullivan KM, Soe MM. OpenEpi: open source epidemiologic statistics for public health, version 3.01 [Archived version]. OpenEpi. URL: https://web.archive.org/web/20260814154726/https://www.openepi.com/Menu/OE_Menu.htm [accessed 2026-08-16]
14. McElroy E, Kearney M, Touhey J, Evans J, Cooke Y, Shevlin M. The CSS-12: Development and validation of a Short-Form Version of the Cyberchondria Severity Scale. Cyberpsychol Behav Soc Netw 2019 May;22(5):330-335. [doi: 10.1089/cyber.2018.0624] [Medline: 31013440]
15. Diamond JJ, Becker JA, Arenson CA, Chambers CV, Rosenthal MP. Development of a scale to measure adults' perceptions of health: preliminary findings. Journal Community Psychology 2007 Jul;35(5):557-561 [FREE Full text] [doi: 10.1002/jcop.20164]
16. Kadioğlu H, Yildiz A. Validity and reliability of Turkish version of perception of health scale. Turkiye Klinikleri J Med Sci 2012;32(1):47-53. [doi: 10.5336/medsci.2010-21761]
17. Gümüş Demir Z. Yaşlılarda Psikolojik İyi Oluş Ölçeği'nin Geliştirilmesi. Senex: Yaşlılık Çalışmaları Dergisi 2022;6(1):16-32 [FREE Full text]
18. Tuğteki˙n U, Barut Tuğteki˙n E. Siberkondri Ciddiyet Ölçeği'nin Kısa Formunun Türkçeye Uyarlanması ve Öğretmen Adaylarının Aşırı Çevrim İçi Bilgi Arama Davranışları. Anemon Muş Alparslan Üniversitesi Sosyal Bilimler Dergisi 2021;9(6):1747-1762. [doi: 10.18506/anemon.963253]
19. Carstensen LL, Shavit YZ, Barnes JT. Age advantages in emotional experience persist even under threat from the COVID-19 pandemic. Psychol Sci 2020 Nov;31(11):1374-1385. [doi: 10.1177/0956797620967261] [Medline: 33104409]
20. Vuorre M, Przybylski AK. Global well-being and mental health in the internet age. Clinical Psychological Science 2024 Sep;12(5):917-935. [doi: 10.1177/21677026231207791]
21. Chen TF, Pien LC, Fan CS, Liang KL, Chiu YW. Financial strain and social support as moderators of the relationship between living alone and depressive symptoms in older people. BMC Geriatr 2024 Aug 1;24(1):646. [doi: 10.1186/s12877-024-05237-1] [Medline: 39090539]
22. Li Y, Li J, Zhou C, et al. Unraveling the relationships among pandemic fear, cyberchondria, and alexithymia after China's exit from the zero-COVID policy: insights from a multi-center network analysis. Front Psychiatry 2024;15. [doi: 10.3389/fpsyt.2024.1489961]
23. Srivastava S, Purkayastha N, Chaurasia H, Muhammad T. Socioeconomic inequality in psychological distress among older adults in India: a decomposition analysis. BMC Psychiatry 2021 Apr 7;21(1):179. [doi: 10.1186/s12888-021-03192-4] [Medline: 33823847]
24. Nasiri M, Mohammadkhani S, Akbari M, Alilou MM. The structural model of cyberchondria based on personality traits, health-related metacognition, cognitive bias, and emotion dysregulation. Front Psychiatry 2022;13:960055. [doi: 10.3389/fpsyt.2022.960055] [Medline: 36699479]

XSL•FO
RenderX

25. Muse K, McManus F, Leung C, Meghreblian B, Williams JMG. Cyberchondriasis: fact or fiction? A preliminary examination of the relationship between health anxiety and searching for health information on the internet. J Anxiety Disord 2012 Jan;26(1):189-196. [doi: 10.1016/j.janxdis.2011.11.005] [Medline: 22137465]
26. Fergus TA, Dolan SL. Problematic internet use and internet searches for medical information: the role of health anxiety. Cyberpsychol Behav Soc Netw 2014 Dec;17(12):761-765. [doi: 10.1089/cyber.2014.0169] [Medline: 25412398]
27. Bajcar B, Babiak J. Self-esteem and cyberchondria: the mediation effects of health anxiety and obsessive–compulsive symptoms in a community sample. Curr Psychol 2021 Jun;40(6):2820-2831. [doi: 10.1007/s12144-019-00216-x]
28. Yang H, Deng Q, Geng Q, et al. Association of self-rated health with chronic disease, mental health symptom and social relationship in older people. Sci Rep 2021;11(1):14653. [doi: 10.1038/s41598-021-94318-x]
29. Ryff CD. Happiness is everything, or is it? Explorations on the meaning of psychological well-being. J Pers Soc Psychol 1989;57(6):1069-1081. [doi: 10.1037/0022-3514.57.6.1069]
30. White RW, Horvitz E. Cyberchondria: Studies of the escalation of medical concerns in Web search. ACM Trans Inf Syst 2009;27(4):1-37. [doi: 10.1145/1629096.1629101]
31. Maftei A, Holman AC. Cyberchondria during the coronavirus pandemic: the effects of neuroticism and optimism. Front Psychol 2020;11:567345. [doi: 10.3389/fpsyg.2020.567345] [Medline: 33192848]
32. Starcevic V, Berle D. Cyberchondria: towards a better understanding of excessive health-related internet use. Expert Rev Neurother 2013 Feb;13(2):205-213. [doi: 10.1586/ern.12.162] [Medline: 23368807]
33. Cao Z, Chu Y. Online health information seeking and health anxiety in the elderly: the mediating role of perceived information quality. IR 2025;30(3):267-304. [doi: 10.47989/ir30341443]
34. Boyacioğlu NE, Irmak HS, Çaynak S. Internet use for seeking health information in old age. Turkiye Klinikleri J Health Sci ;6(3):641-650. [doi: 10.5336/healthsci.2020-79104]

Abbreviations

CER: certainty
COM: control center
COMP: compulsion
CSS-12: Cyberchondria Severity Scale Short Form
DIST: distress
EXC: excessiveness
IHPS: Individual Health Perception Scale
IOH: importance of health
PWBS-OP: Psychological Well-Being Scale for Older People
REAS: reassurance
SAW: self-awareness
WHO: World Health Organization

Edited by A Kushniruk; submitted 02.Sep.2025; peer-reviewed by H Durmus, P Juarez-Rodriguez; revised version received 23.Jun.2026; accepted 22.Jul.2026; published 21.Aug.2026.

Please cite as:
Gündoğdu S, Özgür Ö, Kilavuz A, Tufan İ
Factors Associated With Health Perception and Psychological Well-Being Among Third-Age University Students: Cross-Sectional Study on the Role of Cyberchondria
JMIR Hum Factors 2026;13:e83371
URL: https://humanfactors.jmir.org/2026/1/e83371
doi:10.2196/83371

© Semra Gündoğdu, Özlem Özgür, Asli Kilavuz, İsmail Tufan. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 21.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effectiveness of a Brief Digitized Contact-Based Intervention in Improving Mental Health Stigma and Help-Seeking in Young Adults: Mixed Methods Study

Annabel Songco[1], PhD; Natalie Dahora[2], MClinPsych; Andrew Essen[2], MClinPsych; Andrew Mackinnon[1], PhD; Gemma Sicouri[1], DClinPsych, PhD; Elizabeth Pellicano[3], MPsych, PhD; Jennifer L Hudson[1], MClinPsych, PhD

[1]Black Dog Institute, UNSW Sydney, Hospital Road, Randwick, Sydney, New South Wales, Australia

[2]Department of Psychology, Macquarie University, Sydney, New South Wales, Australia

[3]Department of Clinical, Educational, and Health Psychology, University College London, London, United Kingdom

Corresponding Author:
Jennifer L Hudson, MClinPsych, PhD
Black Dog Institute, UNSW Sydney, Hospital Road, Randwick, Sydney, New South Wales, Australia

Abstract

Background: Contact-based interventions, where individuals share their lived experiences of mental health difficulties and recovery, appear effective in reducing stigma.

Objective: This study examined the effectiveness of a brief digitized contact-based intervention in reducing mental health stigma and improving help-seeking intentions in young adults. The study also sought a deeper understanding of the perspectives of young adults toward the contact-based intervention.

Methods: A mixed methods study, using quantitative and qualitative analyses, examined mental health public stigma, self-stigma, and help-seeking attitudes in undergraduate psychology students (N=328) before, immediately after, and 1 week following the contact-based intervention, compared to an active control. The intervention comprised a brief video (8 min) of a mental health consumer sharing their lived experience of mental health difficulties and their recovery. Participants were invited to participate in a subsequent semistructured interview (n=12) that further explored their perspectives on the contact-based intervention.

Results: Relative to control participants, those in the intervention group reported small improvements in mental health public stigma and help-seeking intentions using formal sources of help; however, this was not maintained at 1-week follow-up. There was limited evidence that the intervention improved self-stigma or help-seeking intentions using informal sources of help. Improvements in self-stigma were not evident following the intervention, but were observed 1 week later. Additional exploratory analyses showed that participants with high mental distress were more likely to provide negative evaluations of the intervention in terms of the helpfulness of the videos, relatability to the speaker, and showed a decrease in their desire to seek help from friends and family following the videos, compared to those with low mental health distress.

Conclusions: Although this brief digitized contact-based intervention holds promise as an easily disseminated strategy for young adults to reduce mental health public stigma and improve formal help-seeking intentions, these effects show that overall improvements in public stigma and formal help-seeking intentions were short-lived. Furthermore, the potential negative impact of the intervention on informal help-seeking intentions in those with high mental health distress suggests a need to tailor interventions for young adults with high levels of anxiety or depression. The study has important implications for designing brief contact-based interventions and what young adults would benefit from the most over time.

Trial Registration: ISRCTN Registry ISRCTN65081246; https://www.isrctn.com/ISRCTN65081246

(JMIR Hum Factors 2026;13:e74391) doi:10.2196/74391

KEYWORDS

contact-based intervention; mental health; stigma; help-seeking; young adults; lived experience

Introduction

Background

Mental health stigma is a significant deterrent to individuals with mental disorders from seeking help through health services [1]. Both public stigma (negative attitudes and stereotypes toward others or groups) and self-stigma (internalized negative attitudes) have a negative impact on help-seeking behaviors and appear to disproportionately impact youth [2]. Stigma is a key barrier to help-seeking behavior, indirectly contributing to the maintenance of untreated mental health difficulties and associated economic and social costs.

XSL•FO
RenderX

Contact-based interventions are one of the most effective strategies for improving mental health stigma [3,4]. Contact-based interventions involve either direct or indirect interaction with an individual from the respective marginalized group to allow opportunity for the disconfirmation of stereotypes and subsequent reduction of prejudicial attitudes [5]. These interventions were derived from Allport's [6] theory of intergroup contact, based on the premise that interaction of a positive nature between individuals of majority and minority groups works to reduce prejudices. The aim of these interventions is to reduce prejudicial beliefs and attitudes through interpersonal contact, which typically involves an individual with lived experience of mental health difficulties sharing and discussing their personal experience. Contact-based interventions can involve varying degrees of interpersonal engagement and activate different psychological mechanisms for attitude change. When evaluating randomized controlled trials, Corrigan and colleagues [5] found that contact-based interventions delivered a small-to-moderate reduction in stigma and were better than education interventions for reducing stigma in adults. Further meta-analyses found pure contact-based interventions to be comparable in their effectiveness with combined contact and education interventions [7], although there is some question about the long-term maintenance of effects [8].

The effectiveness of contact-based interventions does not appear to differ across modalities, such as face-to-face or video. With the increasing accessibility of technology, the growing popularity of telehealth, online services, and immersive technologies, such as virtual reality in the psychological and medical fields, there is a rising interest in the efficacy of scalable, easily implementable digital stigma reduction interventions that are capable of reaching broad audiences. Several studies have found that when compared to live contact-based interventions, interventions consisting of video interpersonal contact have proven to be just as efficacious [7,8]. A review by Janoušková and colleagues [9] showed that video interventions for young adults, which included social contact, were comparable to or better than traditional modalities for improving stigma. However, most studies focused on evaluating the effectiveness of contact-based interventions in addressing public stigma, with few studies on self-stigma, highlighting a gap in the literature [10]. In a young adult sample, one study of a prolonged intervention comprising five 2-hour workshops found that contact significantly improved self-stigma [11]. However, little is known about the effectiveness of brief, more easily disseminated interventions on self-stigma for this population. Given the role of both public and self-stigma in mitigating help-seeking behavior, these factors are crucial to appropriately evaluate and develop antistigma programs, with further research needed on the optimal duration of the intervention and effectiveness of brief interventions.

Few studies have investigated improvements to help-seeking following contact, beyond stigma improvement [12]. Encouragingly, recent studies targeting young adults have found improvements in help-seeking intentions following pure video contact and combined contact and education interventions [13,14]. Factors such as openness to seeking help and increased mental health literacy have improved help-seeking intentions and attitudes in young adults. Moreover, mental health stigma research has predominantly focused on stigma toward more severe psychiatric illnesses. For example, individuals with schizophrenia are often stigmatized and perceived as dangerous, unstable, or unpredictable, with lower likelihood of recovering from their disorder [15]. Research focused on more common mental health difficulties, such as anxiety and depression, is limited, despite being the leading cause of mental health burden in young people [16]. In a recent review, Curcio and Corboy [17] indicated that individuals with mood or anxiety disorders are perceived differently, with stigma relating to perceptions around fault and responsibility [18]. Thus, the nature and intensity of mental health stigma can differ across mental disorders and further research on contact-based interventions that target common mental health difficulties, such as anxiety and depression, is needed.

A further critical gap in the literature concerns the evaluation of moderating variables that may enhance the effectiveness of interventions. Some studies have found participant variables such as age, gender, and ethnicity to moderate the effectiveness of contact interventions [19]. Contact theory stipulates that similarity and equality between the individual of the stigmatized group and the majority group member maximizes the applicability and credibility of the intervention and resulting efficacy [20]. Indeed, Kruger and colleagues [13] found that tailoring contact intervention videos to college students by using individuals who were college students themselves increased help-seeking intentions, while the generic contact intervention did not.

Another factor to consider when tailoring contact-based interventions is an individual's previous or current experiences with mental health difficulties, which can impact subsequent help-seeking behaviors. Brennan and McGrew [21] found differences in how mental health consumers and nonconsumers perceived a contact-based intervention, with nonconsumers significantly more likely to find the intervention educational, endorse feelings of encouragement, and express beliefs that recovery is possible. Further, Clement et al [1] reported that when discussing mental health stigma, individuals with previous mental health difficulties were more likely to report feelings of shame and embarrassment and less likely to report weakness or social rejection, compared to those without previous mental health difficulties. Cerully and colleagues [22] reported that while contact-based interventions are effective for those with and without personal experiences of mental health difficulties, they are less effective for those with prior personal experiences. Thus, there is currently a gap in the literature with few studies investigating how individual characteristics, such as previous or current mental health difficulties, may impact the efficacy of the intervention.

This Study

The batyr@uni program, run by the mental health organization batyr, is a contact-based intervention aimed at reducing general mental health stigma and increasing help-seeking for university students. The intervention is delivered to university students as well as to schools, with a focus on common mental health

XSL•FO
RenderX

difficulties such as anxiety and depression. The batyr@uni program (60‐90 min in duration) can be delivered face-to-face or online and focuses on sharing the lived experience stories of individuals with mental health difficulties and their subsequent recovery via help-seeking. The lived experience stories (delivered live or through brief digital videos) provide valuable insights into how young adults can support their peers, strategies, and self-care practices to maintain good mental health, help identify signs and symptoms of mental ill health, and access support. The program involves young people from diverse backgrounds (eg, age, gender, and geographical context) and experiences to optimize relatability to audience members. During the COVID-19 pandemic, the batyr@uni program transitioned from traditional face-to-face to online video delivery.

The aim of this study was to quantitatively examine the effectiveness of one component of the digitized contact-based intervention (ie, an 8-min video of lived experience stories targeting anxiety and depression) in reducing mental health stigma and improving help-seeking intentions in young adults. We also used a qualitative approach to provide a deeper understanding of young adults' perspectives on the brief contact-based intervention. Further, the study aimed to identify potential differences in whether young adults experiencing high or low mental health distress respond differently to the intervention.

Based on previous literature, it was hypothesized that those in the intervention group would experience greater reduction in both public and self-stigma, and greater improvements in help-seeking intentions (using formal and informal sources of help), relative to the active control group, and that these effects would be maintained at 1-week follow-up. In addition, it was hypothesized that individuals experiencing high mental health distress would perceive mental health stigma differently and find the intervention more beneficial, compared to individuals with low mental health distress. For the qualitative data, we used reflexive thematic analysis to identify themes around participants' perceptions of the intervention [23].

Methods

Participants

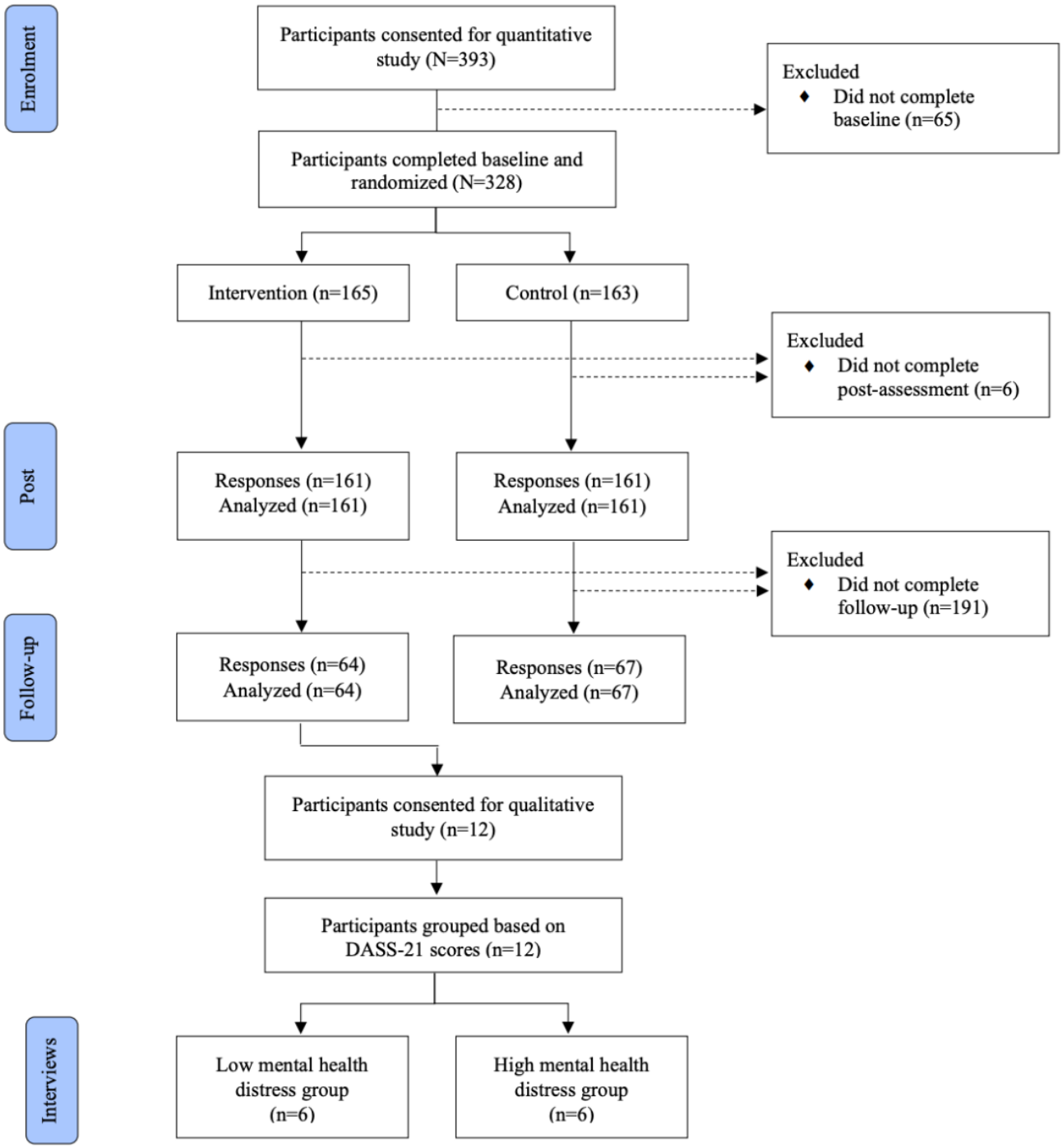
Participants were undergraduate psychology students (N=328) aged 17 to 24 (mean 18.95, SD 1.61) years. The majority of participants were female (269/328, 82%) and self-identified their ethnicity as Caucasian (173/328, 52.7%), Asian (77/328, 23.5%), self-identified (35/328, 10.7%), Middle Eastern (31/328, 9.5%), African (6/328, 1.8%), Hispanic (5/328, 1.5%), and Aboriginal and/or Torres Strait Islander (5/328, 1.5%). From this larger sample, 12 participants volunteered to participate in a subsequent qualitative study. These participants were aged 18 to 24 (mean 20.75, SD 2.70) years, and 5 identified as male, 5 as female, and 2 reported their gender as "Self-identified." Half of the participants identified as Caucasian (n=6) and half as Asian (n=6). Participants were recruited via a university research portal and received course credit in return for their participation. The study was advertised to first-year psychology undergraduate students through the university research portal and was conducted online via the LimeSurvey platform (LimeSurvey GmbH). A G-Power analysis was used to determine adequate sample size (power=.95, Cohen*d*=0.25, and α=.05).

Study Design

Quantitative data were collected at 3 time points (baseline, postintervention, and 1-wk follow-up) to evaluate the effectiveness of a brief digitized contact-based intervention (one component of the batyr@uni program) in reducing mental health stigma and improving help-seeking intentions in young adults. In addition, participants were invited to participate in a subsequent qualitative study comprising a semistructured interview relating to their perspectives and attitudes about the contact-based intervention. This mixed methods study used both a quantitative and qualitative approach to provide a deeper insight into mental health stigma in young adults (refer to Figure 1 for study flow).

XSL•FO
RenderX

Figure 1. Study flow (reason for withdrawals and noncompleters unknown due to the anonymous nature of the online study). DASS-21: Depression, Anxiety, and Stress Scale - Short Form.



Ethical Considerations

This study was approved by the Macquarie University Human Research Ethics Committee (reference no. 52021904124875). As the authors changed institutions after data collection, but before analysis could be completed, the study was also reviewed and approved by the University of New South Wales Research Ethics and Compliance Committee (reference no. iRECS7757). This study was not initially defined as a clinical trial by the Macquarie University Human Research Ethics Committee as participants were randomly allocated to receive 2 educational videos designed to change attitudes rather than biomedical or health outcomes. As a result, the study was not preregistered as a clinical trial. However, following the University of New South Wales' broader definition of a clinical trial, the study was retrospectively registered with the International Standard Randomized Controlled Trial Registry (reference no. ISRCTN65081246). Informed consent was obtained from all participants prior to participation, with the right to withdraw at any time. Participant privacy and confidentiality were maintained through deidentification of data, and participants received course credit in return for their participation.

XSL•FO
RenderX

Quantitative Study

Intervention

Participants allocated to the intervention group viewed 1 of 4 videos (8 min) of a mental health consumer sharing their lived experience of mental health difficulties, specifically relating to anxiety and depression, and their recovery. The videos provided by batyr are being delivered as part of their mental health stigma intervention, the batyr@uni program, and included a short prerecorded debrief by a batyr representative who emphasized key themes. To increase diversity of the speakers across the control and the intervention group, 4 versions of the video were used, varying the gender and perceived racial appearance of the speaker: a female speaker with a White racial appearance, a male speaker with a White racial appearance, a female speaker with a non-White racial appearance, and a male speaker with a non-White racial appearance. Racial characteristics were coded based on visual appearance only; speakers did not self-report race or ethnicity. All speakers were young adults, approximately matching the age range of the study population. Participants were randomly allocated to watch 1 of the 4 videos.

Active Control

Participants in the control group viewed 1 video (8 min) of an individual speaking about the benefits of healthy eating. The information on healthy eating was adapted from an Australian Government resource (Australian National Health and Medical Research Council, 2013), and all videos followed the same script. The individual speaker in the video for the active control group was matched with the video in the intervention group in terms of approximate age, gender, and ethnicity.

Qualitative Study

Participants were eligible to participate in the qualitative study if they had previously completed the quantitative study and were allocated to the intervention group. Participants in the intervention group were provided with an opportunity to express their interest in participating in an interview to discuss their perspectives on the brief contact-based intervention.

Measures

Demographics

Demographics such as age, gender, ethnicity, and whether participants identified as Aboriginal or Torres Strait Islander are reported in Table 1.

Table . Demographic characteristics of participants at baseline.

Variable	Intervention	Control	Total
Age, mean (SD; range)	18.91 (1.51; 17-24)	18.96 (1.67; 17-24)	18.95 (1.61; 17-24)
Sex, n (%)			
Female	143 (43.6)	126 (38.4)	269 (82)
Male	19 (5.8)	33 (10.1)	52 (15.9)
Do not want to say	3 (0.9)	0 (0)	3 (0.9)
Self-identified	0 (0)	4 (1.2)	4 (1.2)
Ethnicity, n (%)			
Caucasian	92 (28)	81 (24.7)	173 (52.7)
Asian	35 (10.7)	42 (12.8)	77 (23.5)
Self-identified	18 (5.5)	17 (5.2)	35 (10.7)
Middle Eastern	14 (4.3)	17 (5.2)	31 (9.5)
African	3 (0.9)	3 (0.9)	6 (1.8)
Hispanic	2 (0.6)	3 (0.9)	5 (1.5)
Don’t want to say	1 (0.3)	0 (0)	1 (0.3)
Aboriginal and/or Torres Strait Islander, n (%)			
No	162 (49.4)	160 (48.8)	322 (98.2)
Yes	3 (0.9)	2 (0.6)	5 (1.5)
Don’t want to say	0 (0)	1 (0.3)	1 (0.3)

Public Stigma

Mental health stigma was measured with a modified version of the Generalized Anxiety Stigma Scale (GASS)—Personal Stigma Subscale. This 10-item scale measures public stigma related to anxiety [24] and was adapted in this study to measure general mental health stigma. To achieve this, the word “anxiety” in each item was replaced with “mental health problems” to create a new Mental Health Stigma Scale (MHSS).

Participants rated each item on a 5-point scale ranging from strongly disagree to strongly agree. The total score is the sum of all items, with higher scores indicative of greater stigma toward mental health. This measure has demonstrated good internal consistency, test-retest reliability, and validity in adults

XSL•FO
RenderX

[24,25] and young adults [26]. In this study, the modified MHSS showed good reliability across time points (Cronbach α=0.84, 0.89, and 0.89) and high correlations with the GASS at each time point (ranging from *r*=0.80 to 0.90).

Self-Stigma

The Self-Stigma of Mental Illness Scale (SSOMI) measured internalized stigma of mental illness [27]. Participants rated 10 items on a 5-point scale ranging from strongly disagree to strongly agree. Items are summed to derive a total score, with higher scores indicative of greater self-stigma of mental illness. As the SSOMI uses future-tense language (ie, "I would feel..."), the questionnaire could be used with individuals who do not currently have clinical levels of a mental health diagnosis, and as a result, this is best considered as anticipated self-stigma. The SSOMI has good internal consistency, test-retest reliability, and validity in young adults [27,28]. The SSOMI in this study showed good reliability across time points (Cronbach α=0.81, 0.82, and 0.87).

Help-Seeking Intentions

The General Help-Seeking Questionnaire (GHSQ) assesses intentions to seek help [29]. Participants were asked, "if you were having a personal or emotional problem, how likely is it that you would seek help from each of the following sources?" Participants indicated how likely, on a 7-point scale (extremely unlikely to extremely likely), they would seek help from each source. The GHSQ contains two subscales: (1) mean intentions to seek formal sources of help (ie, mental health professional, phone helpline, family doctor or general practitioner [GP], and teacher or tutor) and (2) mean intentions to seek informal sources of help (ie, partner, friend, parent, relative, or family member). Higher scores indicated greater help-seeking intentions. The GHSQ has demonstrated good internal consistency, test-retest reliability, and validity in young adults [30]. In this study, the formal subscale yielded good reliability (Cronbach α=0.80, 0.81, and 0.80), while the informal subscale demonstrated low reliability (Cronbach α=0.51, 0.55, and 0.55).

Mental Health Distress

The Depression, Anxiety, and Stress Scale - Short Form (DASS-21) is a 21-item self-report measure that assesses depressive, anxious, and stress-related symptoms [31]. Participants rated each item on a 4-point scale (ie, never, sometimes, often, and almost always), indicating the extent to which each item applied to them over the past week. Total scores were summed into 3 subscales (depression, anxiety, and stress), with higher scores indicating greater distress; categorized into normal, mild, moderate, severe, and extremely severe based on the DASS manual. The DASS-21 has consistently demonstrated good reliability [25,32] and excellent internal reliability in this study (Cronbach α=0.95).

Procedure

Prior to the study, participants were provided with participant information and consent forms that included details about the quantitative study. Once informed consent was obtained, participants completed demographic questions (age, gender, and ethnicity) and baseline assessments (public stigma, self-stigma, help-seeking intentions, and current levels of mental health distress). Participants were then randomly allocated to an intervention group (ie, brief digitized contact-based intervention video) or an active control group (ie, healthy eating video) through random number generation implemented in the LimeSurvey platform. After viewing the full video, participants completed assessments immediately after the video and at 1-week follow-up.

After completing the quantitative study, participants had an opportunity to participate in a qualitative study involving an interview relating to their perspectives and attitudes about the intervention. Participants completed individual semistructured interviews with a researcher through Zoom (Zoom Communications, Inc). Each participant was asked the same open-ended questions relating to their previous relevant experiences, help-seeking attitudes, perspectives on the contact-based intervention, and thoughts about the speaker in the intervention (refer to Multimedia Appendix 1). Participants were asked additional follow-up questions to provide further clarification. All Zoom interviews (26 min on average, ranging from 17 to 32 min) were audio-recorded with the participants' permission and transcribed verbatim.

Data Analysis

Quantitative Study

This study used mixed model analysis of covariance (ANCOVA), with time (baseline, postintervention, and 1-wk follow-up) as the within-subjects factor and group (intervention or active control) as the between-subjects factor.

Despite random assignment after the completion of baseline measures and no differences in the way individuals progressed through the trial, there were substantial baseline differences between groups on outcome measures. Accordingly, to examine the effect of the intervention on the outcome measures (MHSS, SSOMI, and GHSQ), ANCOVA using baseline status on the outcome variable as a covariate was chosen as the most appropriate method to adjust for baseline imbalance. Baseline values of each outcome were centered at the whole-sample (intention-to-treat) mean at baseline, so the model parameters for group represent the effect of the intervention at the group mean. The effect of the covariate was allowed to differ postintervention and at follow-up, as was the effect of the intervention itself. Standardized effect sizes were calculated using the difference between groups at the mean and dividing it by the modeled pooled SD for the variable at each time point. Effect sizes (the strengths of the effects demonstrated) were calculated using Cohen *d* estimates of 0.2 for small effect sizes, 0.5 for medium, and 0.8 for large [33].

These analyses were repeated separately for participants with baseline total scores on the DASS-21 of 14 or higher (ie, high mental health distress group) and those with scores below 14 (ie, low mental health distress group) to explore the effectiveness of the intervention in these subgroups.

Qualitative Study

Participants (n=12) were divided into 2 groups based on their levels of current distress to identify potential qualitative differences between individuals with high mental health distress

XSL•FO
RenderX

(total scores>14 on the DASS-21) and low mental health distress (total scores<14 on the DASS-21). Groups consisted of an equal number of participants (n=6). An explanatory sequential design was used to integrate the quantitative and qualitative findings. Quantitative data were collected and analyzed first, followed by qualitative interviews with a subset of participants who volunteered to elaborate on their experiences with the intervention. This approach allowed for a deeper understanding of unexpected or complex results by examining whether participants with high mental health distress responded differently to the intervention compared to those with low mental health distress.

We followed Braun and Clarke's [23] method of reflexive thematic analysis within a contextualist framework, which emphasizes that a person's meaning-making occurs through an interplay between the individual and their environment. In so doing, we used an inductive (bottom-up) approach to identify patterned meanings within the dataset.

One lead researcher conducted all qualitative interviews and transcriptions, immersing themselves in the data. Once completed, participant transcripts were separated into 2 groups ("low mental health distress" or "high mental health distress") and read and reread, with reflexive notes (memos) made relating to notable or recurring observations or concepts. Codes were then applied to each individual transcript, line by line. These codes were discussed during regular meetings with another member of the research team, at which points some codes were modified and reapplied to the transcripts. Given the significant overlap in codes and potential themes, the groups (low and high mental health distress) were combined for subsequent analysis. To identify candidate themes and subthemes, the lead researcher grouped codes together, focusing on semantic features of the data. These initial themes and subthemes were further refined during regular discussions with the broader team, and eventually drawn together into a thematic map. Relevant data, including direct quotes from the transcripts, were collated under the respective themes and subthemes to demonstrate theme or subtheme coherence. The thematic map and the quotes supporting it were reviewed several times with the broader team prior to writing up the results. Consistent with reflexive thematic analysis methodology, we did not focus on data saturation, as we recognize that decisions about sample size and the number of data items collected are inherently subjective and context-dependent, and cannot be fully determined before analysis [23]. Instead, our sampling approach was guided by the concept of "information power" [34] in qualitative research, which considers sample size based on the study aim, sample specificity, theory, data richness, and relevance.

Results

Quantitative Findings

Primary Analyses

Public Stigma

There was a significant difference in mental health public stigma of 0.76 points between the intervention and the active control group postintervention ($F_{1,\ 319}$=8.53, 95% CI 0.25 - 1.27; P=.004). This represented a small effect size (0.18) in favor of the intervention.

Self-Stigma

There was no significant difference between the intervention and active control groups on self-stigma postintervention (modeled mean difference=0.37, SD 6.88; $F_{1,\ 319}$=0.74, 95% CI –0.47 - 1.22; P=.39). However, there was a significant group difference on self-stigma at 1-week follow-up (modeled mean difference=1.55, SD 6.92; $F_{1,\ 166.41}$=5.76, 95% CI 0.27 - 2.82; P=.02, showing a small, beneficial effect size (0.22).

Help-Seeking Intentions: Formal

There was a significant difference of 0.23 points (SD 1.58) between the intervention and the active control group on intentions to seek help from formal sources of help postintervention ($F_{1,\ 319}$=7.78, 95% CI 0.07 - 0.39; P=.006), showing a small effect size (0.15) in favor of the intervention. There was no significant group difference in intentions to seek help from formal sources from baseline to 1-week follow-up (modeled mean difference=0.24, SD 1.53; $F_{1,\ 145.23}$=2.13, 95% CI –0.08 to 0.57; P=.15).

Help-Seeking Intentions: Informal

There were no significant group differences on intentions to seek help from informal sources postintervention (modeled mean difference=–0.07, SD 1.15; $F_{1,\ 319}$=1.72, 95% CI –0.17 to 0.03; P=.19) or at 1-week follow-up (modeled mean difference=–0.20, SD 1.20; $F_{1,\ 139.40}$=2.69, 95% CI –0.44 to 0.04; P=.10).

Refer to Table 2 for means and SDs of outcome measures across time points.

XSL•FO
RenderX

Table . Group means and SDs of outcomes for the overall sample and subgroups.

	Baseline		Postintervention		Follow-up	
	n[a]	Mean (SD)	n	Mean (SD)	n	Mean (SD)
Intervention group						
Overall[b]						
Public stigma	165	2.94 (3.57)	161	2.35 (3.36)	64	2.64 (3.57)
Self-stigma	165	33.96 (6.48)	161	33.16 (6.68)	64	32.16 (7.80)
Formal help-seeking	165	3.88 (1.47)	161	3.98 (1.62)	64	3.70 (1.43)
Informal help-seeking	165	4.45 (1.14)	161	4.47 (1.19)	64	4.14 (1.23)
Subgroups[c,d]						
Public stigma[e]						
High distress	121	2.88 (3.36)	118	2.14 (3.00)	49	2.53 (3.42)
Low distress	44	3.11 (4.15)	43	2.95 (4.19)	15	3.00 (4.12)
Self-stigma[f]						
High distress	121	34.78 (5.93)	118	33.81 (6.62)	49	34.04 (6.90)
Low distress	44	31.73 (7.44)	43	31.35 (6.59)	15	26.00 (7.59)
Formal help-seeking[g]						
High distress	121	3.83 (1.45)	118	3.90 (1.59)	49	3.57 (1.38)
Low distress	44	4.00 (1.54)	43	4.23 (1.68)	49	4.13 (1.57)
Informal help-seeking[h]						
High distress	121	4.40 (1.09)	118	4.37 (1.15)	49	3.88 (1.19)
Low distress	44	4.59 (1.28)	43	4.77 (1.26)	15	4.98 (0.97)
Control group						
Overall[b]						
Public stigma	163	3.99 (4.52)	161	4.00 (4.88)	67	4.31 (4.70)
Self-stigma	163	33.22 (6.64)	161	32.90 (7.09)	67	33.21 (6.66)
Formal help-seeking	163	3.81 (1.50)	161	3.67 (1.55)	67	3.35 (1.57)
Informal help-seeking	163	4.50 (1.12)	161	4.57 (1.11)	67	4.58 (1.11)
Subgroups[c,d]						
Public stigma						
High distress	115	3.86 (4.46)	114	3.95 (4.86)	48	4.35 (4.57)
Low distress	48	4.31 (4.71)	47	4.13 (5.00)	19	4.21 (5.16)
Self-stigma						

XSL•FO
RenderX

	Baseline		Postintervention		Follow-up	
	n[a]	Mean (SD)	n	Mean (SD)	n	Mean (SD)
High distress	115	34.38 (6.71)	114	33.85 (7.22)	48	34.06 (7.05)
Low distress	48	30.44 (5.61)	47	30.60 (6.27)	19	31.05 (5.09)
Formal help-seeking						
High distress	115	3.73 (1.51)	114	3.62 (1.58)	48	3.33 (1.58)
Low distress	48	3.99 (1.48)	47	3.79 (1.48)	19	3.39 (1.60)
Informal help-seeking						
High distress	115	4.37 (1.13)	114	4.47 (1.12)	48	4.55 (1.10)
Low distress	48	4.80 (1.07)	47	4.84 (1.06)	19	4.67 (1.16)

[a]n: number of observations.

[b]Overall: overall sample.

[c]High distress: high mental health distress group (>14 on the Depression, Anxiety, and Stress Scale - DASS-21).

[d]Low distress: low mental health distress group (<14 on the Depression, Anxiety, and Stress Scale - DASS-21).

[e]Public stigma: Mental Health Public Stigma Scale (MHSS).

[f]Self-stigma: Self-stigma of Mental Illness Scale (SSOMI).

[g]Formal help-seeking: General Help-Seeking Questionnaire – formal subscale (GHSQ-formal).

[h]Informal help-seeking: General Help-Seeking Questionnaire – informal subscale (GHSQ-informal).

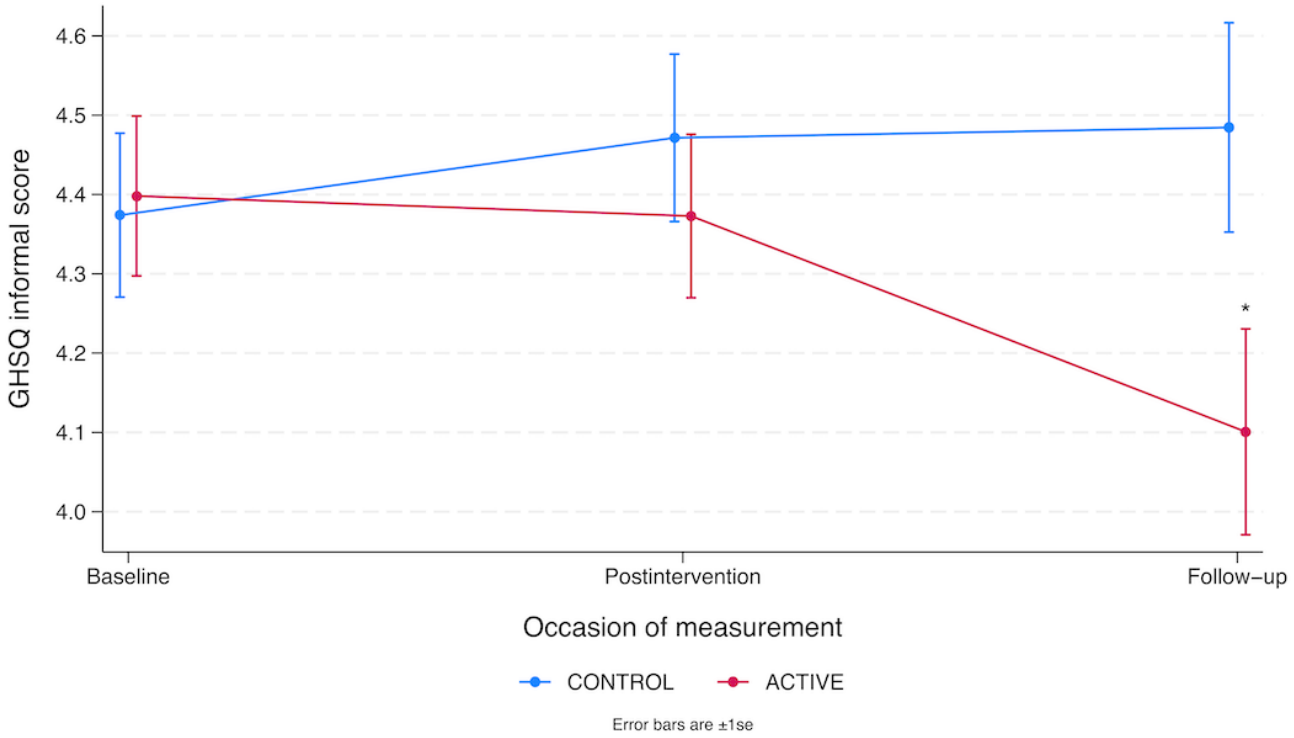
Exploratory Analyses

Mental Health Distress

Exploratory subgroup post hoc analyses examined the effectiveness of the intervention for individuals with high mental health distress and low mental health distress (Table 2). Those with high mental health distress showed reduced levels of mental health public stigma (modeled mean difference=1.00, SD 4.01; $F_{1,\ 229}$=11.70, 95% CI 0.43 - 1.58; P<.001) immediately following the intervention and were less likely to seek help using informal sources of help (eg, a partner, friend, parent, relative, or family member) at 1-week follow-up (modeled mean difference=0.40, SD 1.16; $F_{1,\ 101.7}$=7.56, 95% CI 0.11 - 0.69; P=.007). This effect of a decrease in intentions to seek help from informal sources, for individuals with high mental health distress was present at follow-up (P=.007), but not immediately following the intervention (P=.06; Figure 2). The effect of informal help-seeking was not significant in participants with low mental health distress postintervention (P=.41) and at 1-week follow-up (P=.07). Instead, participants with low mental health distress reported being more likely to seek help from formal sources of help (eg, a mental health professional, phone helpline, GP, or teacher) following the intervention (modeled mean difference=0.41, SD 1.57; $F_{1,87}$=4.41, 95% CI 0.02 - 0.79; P=.04), but this was not significant at follow-up (P=.09).

XSL•FO
RenderX

Figure 2. Effect of a significant decrease in informal help-seeking at follow-up for individuals with high mental health distress in the intervention group versus the active control group (*P=.007).



For individuals with high mental health distress, there was no significant effect of intentions to seek help from formal sources of help following the intervention (P=.06) or at 1-week follow-up (P=.54). Furthermore, there were no improvements in self-stigma for participants with high mental health distress or low mental health distress, immediately after the intervention (P=.39; P=.85) or at 1-week follow-up (P=.11; P=.18), respectively.

Qualitative Findings

Participant Perceptions of the Contact-Based Intervention

We identified two themes regarding participants' perceptions of the brief digitized contact-based intervention: (1) helpfulness of the intervention and (2) relatability to the speaker (Table 3).

Table . Themes and subthemes for participant perspectives of the contact-based intervention.

Theme	Subtheme
1. Helpfulness of the intervention	• 1.1. Helpful for the participant • 1.2. Helpful for other people • 1.3. Learning something new from the intervention
2. Relatability to the speaker	• 2.1. Understanding how the speaker would be feeling • 2.2. Descriptions of the speaker

Theme 1: Helpfulness of the Intervention

Participant responses varied, with both groups indicating mixed feedback on the helpfulness of the intervention. Participants in the high mental health distress group reported that the intervention was not particularly helpful for themselves (subtheme 1.1), with all participants in that group indicating that they did not learn anything new from the intervention (subtheme 1.3). One participant stated that "the video didn't have a huge impact" and that they "wouldn't say it was helpful" as they have "seen so many videos like that before" (subtheme 1.1). Another participant stated that the video did not increase their awareness, but served to reinforce "how cruel other people can be toward others because they're different" (subtheme 1.3). Notably, this lack of perceived benefit from the video was shared by participants in the low mental health distress group with a history of mental health treatment, who voiced similar responses that the video was not helpful "for someone like me who does know quite a bit about it," and that there was nothing in the video that "I wasn't aware of" (subtheme 1.3). Conversely, those in the low mental health distress group without a history of mental health treatment generally reported that the video was "definitely helpful," "enlightening," or "educational and inspiring" (subtheme 1.1).

While participants in the high mental health distress group reported finding the intervention not helpful for themselves, many indicated that they believed it would be helpful for other people (subtheme 1.2) without their own experiences of mental health difficulties. One participant in the high mental health distress group reported that "for some people, it could be quite helpful, somebody could see it and may be motivated to reach

XSL•FO
RenderX

out to someone" (subtheme 1.2). This view was shared by other participants in the same group, who reported that the video might help other people "find and seek some help" (subtheme 1.2). Relatedly, participants in the low mental health distress group with a history of mental health treatment reported similar thoughts that the video might be more helpful for people who "don't know much about mental health" (subtheme 1.2) or people with higher levels of stigma.

Theme 2: Relatability to the Speaker

Participants also spoke at length about the video speaker and the relatability they felt to either the speaker or their story. Importantly, participants in the high mental health distress group reported finding the speaker and their experience less relatable, although they understood how the speaker would be feeling (subtheme 2.1). Several participants in the high mental health distress group stated that while they did not relate to the speaker's experience or situation, they could relate to the general feeling of having mental health difficulties. For example, one participant stated that they "related to the speaker not in the situation he was in, but in the whole idea of how other people exclude you," while another noted that they could relate "not in terms of their circumstance, but specifically to how they were feeling" (subtheme 2.1).

Consistent with this view, one participant from the low mental health distress group with a history of mental health treatment reported a similar observation, noting that while they "felt pretty different to the speaker," they thought "they would probably understand what it feels like to be a little bit different" (subtheme 2.1). Conversely, other participants in the low mental health distress group reported greater relatability to the speaker, with many stating that they could understand the speaker's situation and relate it to something similar they had previously experienced. Extending on this, one participant in the low mental health distress group stated that they "definitely did relate" and felt "connected" to the speaker, while another participant reported feeling "very empathetic" toward what the speaker was going through (subtheme 2.1).

Furthermore, notable differences arose between groups in descriptions of the speaker (subtheme 2.2). Participants in the low mental health distress group reported feeling strongly about the speaker, often describing them as "real," "sincere," "genuine," "brave," "heartfelt," and "engaging" (subtheme 2.2). Participants in the high mental health distress group referred to the speaker as "interesting," "straightforward," or "easy to understand" (subtheme 2.2).

Discussion

Principal Findings

This study aimed to investigate the effectiveness of a brief digitized contact-based intervention, specifically targeting anxiety and depression, in reducing self-stigma and public stigma and improving help-seeking intentions in young adults. In line with our hypothesis, overall public stigma around mental health and formal help-seeking intentions appeared to improve for those in the intervention group immediately following the intervention. These effects, however, were not sustained at follow-up. Contrary to our hypothesis, the contact-based intervention had a limited impact on improving self-stigma or informal help-seeking intentions across the whole sample. Although immediate improvements in self-stigma were not observed for either group postintervention, individuals who received the intervention demonstrated small improvements in self-stigma 1 week after the intervention. Finally, quantitative and qualitative results showed differing effectiveness of the intervention, with those experiencing high mental health distress responding less positively than those with low mental health distress, contrary to predictions.

Overall, these findings suggest that watching the brief videos can potentially lead to immediate reductions in public stigma and contemporaneous increases in the intentions to seek help from formal sources of help, such as a mental health professional, phone helpline, GP, or teacher. As these effects were not maintained at 1-week follow-up, the intervention's impact on public stigma and help-seeking intentions appears to be short-lived. Conversely, a sleeper effect was observed for self-stigma, with improvements in self-stigma only emerging at follow-up. While previous reviews and meta-analyses have found contact effects to persist at short-term follow-up [7], others have not [8]. These findings may not be surprising given the intervention brevity and the small initial effect. Due to dropout, a lack of power at follow-up due to insufficient sample size may have also influenced the null findings relating to self-stigma and help-seeking from informal sources of help.

Recent meta-analyses of contact interventions yield small to moderate effects; however, the improvement in public stigma in this study was of small effect. It may be the case that the "dose" of the current intervention, at only approximately 8 minutes in duration, influenced the magnitude of the effect. Previously evaluated interventions generally range from 30 minutes to hours in duration [1,7-9]. Although the meta-analysis conducted by Morgan and colleagues [8] did not find that the duration of interventions influenced effectiveness, the review only examined studies targeting severe psychiatric stigma. One consideration in this study that may have contributed to the small effect of the brief digitized contact-based intervention on young adults is the isolated context in which the video was delivered. Typically, the lived experience stories are embedded within a wider batyr@uni program (60 - 90 min in duration), which involves a live debrief discussion on key themes and actionable takeaways from the lived experience stories. While the brief videos used in this study include a prerecorded debrief, perhaps integrating a live facilitator to draw links between the story and discussions around key mental health themes may play a greater role in the intervention's overall impact in the future. Evidence suggests that contact-based interventions are more effective when combined with psychoeducation components to improve mental health literacy [12], including approaches that explicitly target misconceptions regarding treatment-seeking [14], which may be more compelling for young people [14].

Shifting our focus to more nuanced exploratory findings within the subgroups, although both individuals with high and low mental health distress showed improvements in public stigma immediately following the intervention, the findings suggest

XSL•FO
RenderX

that individuals with high mental health distress were less likely to want to seek help from informal sources of help, such as through a partner, friend, parent, relative, or family member, 1 week later. This finding is potentially concerning in the absence of an increase in formal help-seeking intentions in this subgroup, with no significant effect of intentions to seek help from formal sources such as a mental health professional, phone helpline, family doctor, GP, or teacher. It is these formal sources of help that have a higher likelihood, although far from guaranteed, to be evidence-based compared to help that may come from friends and family [35]. In contrast, the findings suggest that individuals with low mental health distress were more likely to use formal sources of help immediately after watching the brief video intervention; however, these effects were not sustained at follow-up. Further caution is warranted when interpreting the results, due to the exploratory and underpowered nature of the subgroup analyses or may reflect the complex interactions with prior mental health experiences. Additionally, the low reliability of the informal help-seeking subscale in the GHSQ suggests that these results should be interpreted with caution. Nonetheless, given that contact-based interventions are designed to increase help-seeking attitudes, particularly for individuals experiencing mental health difficulties who may need it the most, careful consideration in the design of brief digitized interventions is needed and tailoring the videos to target these vulnerable groups and encourage help-seeking would be beneficial.

Furthermore, the sleeper effect observed in improved self-stigma for all those who received the intervention was not evident in the subgroup analyses. No intervention effects at postintervention or follow-up were observed in self-stigma for either those with high or low mental health distress. This is not surprising given the small effect size observed in the whole sample (0.2). An insufficient sample size and subsequent lack of power in the subgroups may have contributed to the null findings for self-stigma. Regardless, these effects may indicate that the current intervention is not sufficiently powerful to shift young adults' perception of their mental health. Future research should focus on replication with a larger sample and enhancing the intervention to allow for more meaningful and long-term shifts in mental health stigma.

Several key themes in the qualitative results were identified, showing differences between the subgroups on topics related to the helpfulness of the intervention and relatability to the speaker in the videos. For instance, individuals with high mental health distress tended to have more negative views about the helpfulness of the intervention and were less likely to be able to relate to the speaker's experience in the videos. On the contrary, individuals with low mental health distress reported more positive responses to the intervention and found the videos educational and helpful. These qualitative findings provide valuable insights into young adults' perceptions of the digitized contact-based intervention. While improvements in public stigma were observed in individuals with high mental health distress, they appeared to respond more negatively toward the contact-based intervention compared to those with low mental health distress. They reported that the intervention provided them with no additional information or learning, that they were familiar with the content conveyed, leaving little impact and negatively reinforced the stereotype that people can be cruel toward individuals with mental health difficulties who are perceived as different. This supports previous research indicating that contact-based interventions are less effective for individuals with prior mental health difficulties, which may be related to preexisting awareness and understanding [22]. A potential explanation may be that these individuals were less likely to see the value in seeking help from and disclosing their mental health to family members or friends following the intervention. These findings add to the growing literature indicating differences in perceptions between those with and without mental health difficulties [1,21,22].

Some participants in this study expressed their belief that the contact-based intervention would be helpful for others. Individuals with low mental health distress reported that they found the intervention helpful, educational, and insightful. A possible interpretation for this positive response may be that individuals with low mental health distress feel more connected and less alone after hearing about other people's lived experiences of mental health difficulties. This is important to consider. If contact-based interventions are less effective for individuals with high mental health distress, future research could focus on alternative interventions or tailoring the content to focus less on education and awareness and more on targeting stigma and feelings of shame and embarrassment, which may be more appropriate to increase and improve help-seeking in young adults. This aligns with previous qualitative research, where individuals with a previous mental health concern were found to identify with themes of shame and embarrassment, compared to those without a previous mental health history [1].

Another factor to consider in designing these interventions pertains to the relatability of the speaker in the contact-based intervention. While the lived experience stories delivered within the batyr@uni program are represented by individuals from diverse backgrounds and experiences, these findings suggest that young adults' relatability to the video speaker differs depending on their levels of mental health distress. Individuals with high mental health distress related less and responded less positively to the speaker. In contrast, participants with low mental health distress spoke more positively, often describing the speaker as highly relatable, genuine, and engaging. This discrepancy may contribute to differences in the effectiveness of the contact-based intervention between groups, suggesting different underlying mechanisms driving change in individuals. Perhaps individuals with high mental health distress experience unique challenges and find the more positive or uplifting aspects of the intervention to be somewhat incongruent with their current mental state. While individuals with low mental health distress may be able to connect more to the positive aspects of the speaker's journey and view the speaker more favorably, and thus are more likely to seek help from formal sources of help as indicated in the quantitative subgroup analysis. Furthermore, the videos and the speaker delivering the intervention were not piloted for likability and perceived relatability prior to the intervention, which are factors shown to influence the effectiveness of antistigma interventions and may potentially account for group differences in perceived helpfulness,

XSL•FO
RenderX

help-seeking intentions, and stigma among the low and high mental health distress groups. These are important aspects to consider when developing and evaluating the effectiveness of stigma-related interventions for young adults with mental health difficulties. Further research is needed to determine whether altering such factors may improve the efficacy of such interventions.

Strengths, Limitations, and Future Directions

This study was the first to investigate whether a brief digitized contact-based intervention (one component of the batyr@uni program) was effective in reducing public stigma, self-stigma, and help-seeking intentions in a young adult university sample using a mixed methods study design. The quantitative and qualitative analyses provide a deeper insight into young adults' perspectives of a brief digitized contact-based intervention targeting anxiety and depression. However, one limitation in this study is the way in which some constructs were measured. For instance, we did not measure actual help-seeking behavior, mental health difficulties, and the construct measurement of self-stigma was more aligned with anticipated self-stigma (ie, future-tense items may reduce sensitivity to change for participants without a current stigmatized identity), which may help explain stability in scores across time points. While help-seeking intentions predict behaviors [29], the relationship is not always strong [36]. Future research should prioritize behavioral measurement of actual help-seeking behavior to better evaluate intervention outcomes of such clinical significance.

Moreover, studies have found that factors such as openness to seeking help, increased mental health literacy, including recognition of disorder symptomology and treatment benefits, predicted help-seeking intentions within an undergraduate sample [12]. Further research is needed to determine whether additional components to contact alone, such as mental health literacy and openness to help-seeking, are needed to bridge the gap between stigma improvement and help-seeking attitudes [12].

In this study, approximately 1 in 5 participants dropped out prior to completing assessments at postintervention, and of those who did, over half did not complete the follow-up assessment. Although there were no differences at baseline or across conditions for dropout, it is difficult to determine whether any systematic factor influenced the dropout rate. As the study was completed online and anonymized, it was not possible to determine the reason for dropout. Such substantial dropout likely influenced the power to detect effects at that time point, possibly accounting for the absence of certain effects persisting over time.

A further limitation is that substantial baseline differences were identified between the intervention and control groups on outcome measures, despite randomization to condition after the completion of baseline measures. Given precision in the identical nature of the procedures between conditions and no differences in the way individuals progressed through the trial, it was concluded that these differences occurred by chance. Nevertheless, this impacted our analysis method and interpretation of results, which involved ANCOVA using baseline status on the outcome variable as a covariate to adjust for the baseline imbalance between groups.

The sample in this study limits the generalizability of findings to other similar groups of young adults. In particular, the small sample size of 12 participants in the qualitative study may not be representative of the perceptions and attitudes of the population from which they were drawn. Participants in this study were majority female undergraduate university students studying psychology, and thus, we might expect lower levels of mental health stigma and greater understanding or awareness when compared to the broader population. Relatedly, for the qualitative study, participants voluntarily expressed their interest in the study and willingness to share their perspectives. The participants involved may have been those more familiar with mental health stigma or those with strongly held opinions that they wished to discuss. Further, the brief contact intervention specifically targeted anxiety and depression in young adults; however, the nature and intensity of mental health stigma can differ across mental disorders, thus the findings may be limited to experiences with common mental disorders. Another limitation is due to the small sample of 6 participants in each group (high vs low mental health distress group). That said, the information elicited from these participants during the in-depth interviews was sufficiently rich to address the study objectives.

Despite these limitations, significant qualitative differences in this study were found among the sample of psychology students, perhaps indicating that greater differences may exist within the broader population. Future research should extend the evaluation of contact-based interventions beyond a university setting to young adults within the community and from different cultural and socioeconomic backgrounds. Social and cultural differences, as well as digital literacy factors, would likely appear more pronounced within diverse community settings and may weigh more heavily on digital intervention efficacy and moderators.

Given the benefits of a digitized and brief intervention in this study for those with low mental health distress, such as cost-effectiveness, accessibility, and convenience, more widespread dissemination of contact-based interventions appears possible. This has implications for when face-to-face dissemination is not feasible, such as government-mandated lockdowns, or for those living in rural and remote regions. Further research on the effectiveness of a digitized and brief intervention in varied populations would be useful.

Conclusions

In conclusion, this study suggests that a brief contact-based video targeting anxiety and depression holds promise as an intervention that could be widely disseminated to reduce young adults' mental health public stigma in the short term; however, further research evaluating its long-term impact is needed. The intervention had a limited effect on self-stigma. Although the intervention changed the way people viewed others with mental health difficulties, it did not change how individuals perceived their own mental health. In fact, there is some evidence that the intervention may lead young people with high mental distress to be less likely to speak to their friends and family. Together, these findings suggest the need for brief digitized contact-based interventions to be adapted to better cater to individuals with

XSL•FO
RenderX

high mental health distress. Factors such as the perceived helpfulness of the intervention and relatability to the speaker in the videos were identified as key themes to consider when developing and disseminating contact-based interventions for young adults. Several study limitations may underlie null effects following the contact intervention for improvements in self-stigma and help-seeking intentions, and limit the extent to which findings may be extrapolated, indicating a need for further research and replication, particularly in more diverse populations. Further research is needed to determine whether tailored alternate strategies or adjunct components, such as individuals' mental health literacy, prior experience with mental health difficulties, perceived likability of speakers delivering the videos, and behavioral measures of help-seeking, could improve outcomes, particularly in self-stigma and help-seeking intentions in those with high mental health distress.

Acknowledgments

The authors would like to acknowledge that this was part of a Master's research project by ND and AE who contributed to the design, evaluation, and implementation of this research trial. We would like to thank all the participants who took the time to take part in the study. We would also like to acknowledge batyr for providing access to the video resources.

Generative artificial intelligence (AI) tools were not used at any stage in the preparation of this manuscript.

Funding

The research team received $5000 from batyr to support personnel for the independent preparation of the results for publication. The authors do not receive any royalties for mental health literacy programs for youth. No other external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article. The design and execution of the research project received no funding.

Data Availability

The data generated and analyzed in this study are highly sensitive and therefore not publicly available. The study's ethical approvals limit data sharing. Recordings include identifiable information and cannot be shared. Deidentified data (excluding recordings) are available from the corresponding author, JH, upon reasonable request.

Authors' Contributions

JH, ND, AE, and EP conceived the study. ND and AE collected the data and executed the research study. AS, JH, AM, and AE analyzed and interpreted the data. AS, ND, AE, JH, and AM wrote the manuscript, and all authors provided critical feedback. All authors contributed to the approved final version of the manuscript for submission.

Conflicts of Interest

The preparation of the publication was partially supported by batyr. batyr were not involved in the design, execution or write-up of the study. batyr supplied the video materials of the intervention and was given a draft of the manuscript to review prior to submission for the purposes of ensuring the accurate description of the program. batyr met with the research team throughout the execution of the project for the purposes of updating progress.

Multimedia Appendix 1
Bespoke interview questions created for this study.
[DOCX File, 17 KB - humanfactors_v13i1e74391_app1.docx]

References

1. Clement S, Schauman O, Graham T, et al. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. Psychol Med 2015 Jan;45(1):11-27. [doi: 10.1017/S0033291714000129] [Medline: 24569086]
2. Schnyder N, Panczak R, Groth N, Schultze-Lutter F. Association between mental health-related stigma and active help-seeking: systematic review and meta-analysis. Br J Psychiatry 2017 Apr;210(4):261-268. [doi: 10.1192/bjp.bp.116.189464] [Medline: 28153928]
3. Clement S, Lassman F, Barley E, et al. Mass media interventions for reducing mental health-related stigma. Cochrane Database Syst Rev 2013 Jul 23;2013(7):CD009453. [doi: 10.1002/14651858.CD009453.pub2] [Medline: 23881731]
4. Thornicroft G, Mehta N, Clement S, et al. Evidence for effective interventions to reduce mental-health-related stigma and discrimination. Lancet 2016 Mar;387(10023):1123-1132. [doi: 10.1016/S0140-6736(15)00298-6]
5. Corrigan PW, Morris SB, Michaels PJ, Rafacz JD, Rüsch N. Challenging the public stigma of mental illness: a meta-analysis of outcome studies. Psychiatr Serv 2012 Oct;63(10):963-973. [doi: 10.1176/appi.ps.201100529] [Medline: 23032675]
6. Allport G. The Nature of Prejudice: Addison-Wesley Pub. Co; 1954.

XSL•FO
RenderX

7. Maunder RD, White FA. Intergroup contact and mental health stigma: a comparative effectiveness meta-analysis. Clin Psychol Rev 2019 Aug;72:101749. [doi: 10.1016/j.cpr.2019.101749] [Medline: 31254936]
8. Morgan AJ, Reavley NJ, Ross A, Too LS, Jorm AF. Interventions to reduce stigma towards people with severe mental illness: systematic review and meta-analysis. J Psychiatr Res 2018 Aug;103:120-133. [doi: 10.1016/j.jpsychires.2018.05.017] [Medline: 29843003]
9. Janoušková M, Tušková E, Weissová A, et al. Can video interventions be used to effectively destigmatize mental illness among young people? A systematic review. Eur psychiatr 2017;41(1):1-9. [doi: 10.1016/j.eurpsy.2016.09.008]
10. Waqas A, Malik S, Fida A, et al. Interventions to reduce stigma related to mental illnesses in educational institutes: a systematic review. Psychiatr Q 2020 Sep;91(3):887-903. [doi: 10.1007/s11126-020-09751-4] [Medline: 32372401]
11. Martínez-Hidalgo MN, Lorenzo-Sánchez E, López García JJ, Regadera JJ. Social contact as a strategy for self-stigma reduction in young adults and adolescents with mental health problems. Psychiatry Res 2018 Feb;260:443-450. [doi: 10.1016/j.psychres.2017.12.017] [Medline: 29272729]
12. Gulliver A, Griffiths KM, Christensen H, Brewer JL. A systematic review of help-seeking interventions for depression, anxiety and general psychological distress. BMC Psychiatry 2012 Jul 16;12(1):81-81. [doi: 10.1186/1471-244X-12-81] [Medline: 22799879]
13. Kruger E, Pitts SC, Denenny D, DeLuca JS, Schiffman J. Efficacy of contact intervention videos on college students' intentions toward mental health help-seeking. J Am Coll Health 2022 Oct;70(7):2033-2041. [doi: 10.1080/07448481.2020.1842417] [Medline: 33253017]
14. Shahwan S, Lau JH, Goh CMJ, et al. The potential impact of an anti-stigma intervention on mental health help-seeking attitudes among university students. BMC Psychiatry 2020 Nov 25;20(1):562. [doi: 10.1186/s12888-020-02960-y] [Medline: 33238951]
15. Robinson P, Turk D, Jilka S, Cella M. Measuring attitudes towards mental health using social media: investigating stigma and trivialisation. Soc Psychiatry Psychiatr Epidemiol 2019 Jan;54(1):51-58. [doi: 10.1007/s00127-018-1571-5] [Medline: 30069754]
16. Vos T, Lim SS, Abbafati C, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct;396(10258):1204-1222. [doi: 10.1016/S0140-6736(20)30925-9]
17. Curcio C, Corboy D. Stigma and anxiety disorders: a systematic review. Stigma Health 2020;5(2):125-137. [doi: 10.1037/sah0000183]
18. Chang CC, Wu TH, Chen CY, Lin CY. Comparing self-stigma between people with different mental disorders in Taiwan. J Nerv Ment Dis 2016 Jul;204(7):547-553. [doi: 10.1097/NMD.0000000000000537] [Medline: 27218219]
19. Wong EC, Collins RL, Cerully JL, Yu JW, Seelam R. Effects of contact-based mental illness stigma reduction programs: age, gender, and Asian, Latino, and White American differences. Soc Psychiatry Psychiatr Epidemiol 2018 Mar;53(3):299-308. [doi: 10.1007/s00127-017-1459-9] [Medline: 29196773]
20. Corrigan P, Roe D, Tsang H. Challenging the Stigma of Mental Illness: Lessons for Therapists and Advocates: John Wiley & Sons; 2011.
21. Brennan M, McGrew JH. Evaluating the effects of NAMI's consumer presentation program, In Our Own Voice. Psychiatr Rehabil J 2013 Jun;36(2):72-79. [doi: 10.1037/h0094974] [Medline: 23647146]
22. Cerully JL, Collins RL, Wong E, Seelam R, Yu J. Differential response to contact-based stigma reduction programs: perceived quality and personal experience matter. Psychiatry Res 2018 Jan;259:302-309. [doi: 10.1016/j.psychres.2017.10.027] [Medline: 29096336]
23. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019 Aug 8;11(4):589-597. [doi: 10.1080/2159676X.2019.1628806]
24. Griffiths KM, Batterham PJ, Barney L, Parsons A. The Generalised Anxiety Stigma Scale (GASS): psychometric properties in a community sample. BMC Psychiatry 2011 Nov 22;11(1):184-184. [doi: 10.1186/1471-244X-11-184] [Medline: 22108099]
25. Henry JD, Crawford JR. The short‐form version of the Depression Anxiety Stress Scales (DASS‐21): construct validity and normative data in a large non‐clinical sample. British J Clinic Psychol 2005 Jun;44(2):227-239. [doi: 10.1348/014466505X29657]
26. Busby Grant J, Bruce CP, Batterham PJ. Predictors of personal, perceived and self-stigma towards anxiety and depression. Epidemiol Psychiatr Sci 2016 Jun;25(3):247-254. [doi: 10.1017/S2045796015000220] [Medline: 25791089]
27. Tucker JR, Hammer JH, Vogel DL, Bitman RL, Wade NG, Maier EJ. Disentangling self-stigma: are mental illness and help-seeking self-stigmas different? J Couns Psychol 2013 Oct;60(4):520-531. [doi: 10.1037/a0033555] [Medline: 23815629]
28. Wallin E, Maathz P, Parling T, Hursti T. Self-stigma and the intention to seek psychological help online compared to face-to-face. J Clin Psychol 2018 Jul;74(7):1207-1218. [doi: 10.1002/jclp.22583] [Medline: 29315545]
29. Wilson CJ, Deane FP, Ciarrochi J, Rickwood D. Measuring help-seeking intentions: properties of the general help-seeking questionnaire. Can J Couns 2005;39(1):15-28 [FREE Full text]
30. Deane FP, Wilson CJ, Ciarrochi J. Suicidal ideation and help-negation: not just hopelessness or prior help. J Clin Psychol 2001 Jul;57(7):901-914. [doi: 10.1002/jclp.1058] [Medline: 11406803]

XSL•FO
RenderX

31. Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
32. Ronk FR, Korman JR, Hooke GR, Page AC. Assessing clinical significance of treatment outcomes using the DASS-21. Psychol Assess 2013 Dec;25(4):1103-1110. [doi: 10.1037/a0033100] [Medline: 23730826]
33. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: Lawrence Erlbaum Associates; 1988. [doi: 10.1049/cm.1988.0027]
34. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
35. Gandhi E, OGradey-Lee M, Jones A, Hudson JL. Receipt of evidence-based care for children and adolescents with anxiety in Australia. Aust N Z J Psychiatry 2022 Nov;56(11):1463-1476. [doi: 10.1177/00048674211068780] [Medline: 35044881]
36. Nagai S. Predictors of help‐seeking behavior: distinction between help‐seeking intentions and help‐seeking behavior. Jpn Psychol Res 2015 Oct;57(4):313-322. [doi: 10.1111/jpr.12091]

Abbreviations

ANCOVA: analysis of covariance
DASS-21: Depression, Anxiety and Stress Scale - Short Form
GASS: Generalized Anxiety Stigma Scale
GHSQ: General Help-Seeking Questionnaire
GP: general practitioner
MHSS: Mental Health Stigma Scale
SSOMI: Self-Stigma of Mental Illness Scale

Edited by A Kushniruk; submitted 25.Mar.2025; peer-reviewed by R Szekely, R Chen; revised version received 26.Mar.2026; accepted 27.Mar.2026; published 10.Jun.2026.

Please cite as:
Songco A, Dahora N, Essen A, Mackinnon A, Sicouri G, Pellicano E, Hudson JL
Effectiveness of a Brief Digitized Contact-Based Intervention in Improving Mental Health Stigma and Help-Seeking in Young Adults: Mixed Methods Study
JMIR Hum Factors 2026;13:e74391
URL: https://humanfactors.jmir.org/2026/1/e74391
doi:10.2196/74391

© Annabel Songco, Natalie Dahora, Andrew Essen, Andrew Mackinnon, Gemma Sicouri, Elizabeth Pellicano, Jennifer L Hudson. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 10.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Digital Episodic Future Thinking Intervention (Luminaut): Co-Design and Iterative Development Study

Naomi Hoffmann[1], PhD; Caitlin A Howlett[1], PhD; Kate Little[2], PhD; Ian Gwilt[3], PhD; Megan A Rebuli[1], BNutDiet; Paige G Brooker[1], PhD; Aaron Davis[2], PhD

[1]Commonwealth Scientific and Industrial Research Organisation, PO Box 10041, Adelaide, SA, Australia

[2]School of Architecture and Built Environment, Adelaide University, Adelaide, SA, Australia

[3]School of Art and Design, Adelaide University, Adelaide, SA, Australia

Corresponding Author:
Naomi Hoffmann, PhD
Commonwealth Scientific and Industrial Research Organisation, PO Box 10041, Adelaide, SA, Australia

Abstract

Background: Digital health interventions can be effective at changing behavior, but achieving long-term adherence remains a challenge. One psychological barrier to health behavior change is *future discounting*, or the tendency to prefer smaller, short-term rewards over larger, long-term rewards. Episodic Future Thinking (EFT) can disrupt future discounting and is a promising technique for improving health behavior, but such interventions have not been co-designed to address end user needs.

Objective: This study aimed to co-design an app with end users to deliver an EFT intervention aimed at promoting health behavior change in those in the prerisk phase for chronic conditions.

Methods: Community members participated in up to 2 series of face-to-face co-design workshops. A prototype of the app was reviewed, and insights were gathered to understand (1) the optimal characteristics of the app and (2) the concepts of future discounting and EFT. Themes were generated using inductive thematic analysis.

Results: Participants were South Australian adults (n=30) who were predominately affluent women (27/30, 90%) aged 25 - 44 years (mean 36.37, SD 5.65 years). Feedback generated from the first workshop series resulted in 26 suggestions of which 15 informed iterative app development. Higher-level principles were identified and categorized into 5 overarching themes: concept acceptance, triggers and barriers, personalization, gamification, and user-friendly interface.

Conclusions: This study used co-design methodology to develop an app-based EFT intervention. Ongoing engagement with end users and key stakeholders (eg, health care professionals) is needed to ensure that the app meets changing needs. Future work will aim to evaluate its effectiveness in a large-scale clinical trial.

(JMIR Hum Factors 2026;13:e74099) doi:10.2196/74099

KEYWORDS

Episodic Future Thinking; digital intervention; future discounting; health behavior change; health behaviors; co-design methods; mobile phone

Introduction

Background

Noncommunicable diseases (NCDs), including cardiovascular diseases, diabetes, cancers, and chronic respiratory diseases, account for 75% of deaths each year [1], making them the leading cause of death worldwide [2]. The main risk factors for NCDs are largely preventable and include modifiable lifestyle behaviors such as poor dietary intake, physical inactivity, alcohol consumption, and illicit drug use [3]. Therefore, devising effective lifestyle interventions to reduce NCD risk is an important research priority. The aim of this study was to develop a co-designed smartphone app to deliver a psychological intervention to improve health behaviors in individuals at risk of NCD.

Challenges in Promoting Sustained Health Behavior Change

Despite the well-established health benefits of consuming a nutritious diet and engaging in regular physical activity [4-6], many people find it challenging to adopt and maintain health-promoting behaviors [7]. Indeed, the 2022 Australian National Health Survey found that only 4.2% of Australians met both the fruit and vegetable intake recommendations, while only 23.9% met the physical activity guidelines [8]. Long-term adherence to lifestyle behavior change interventions is an ongoing challenge [7] and may be partly due to underlying psychological barriers. One such barrier is future discounting, namely, the tendency for individuals to devalue rewards as a function of delayed receipt [9], or prioritize behaviors that provide immediate gratification over longer-term benefits [10].

XSL•FO
RenderX

Higher rates of future discounting have been linked with poor health behaviors and obesity [11,12].

In the context of health behaviors, future discounting involves placing less value on the future by engaging in lifestyle behaviors that produce immediate rewards (eg, sedentary behavior and/or consuming discretionary foods high in salt, sugar, and/or fat). Furthermore, such behaviors often lead to the onset of negative, delayed health consequences (ie, chronic disease) in lieu of long-term future rewards (eg, good health). A preference for smaller, immediate rewards over larger, delayed rewards is a transdiagnostic process that has been implicated in health and disease [13]. Future discounting—or delay discounting as it is otherwise known—is typically measured via monetary or domain-specific (eg, food) choice tasks or questionnaires to assess how quickly rewards of varying amounts reduce in value over time. In the eating domain, a meta-analysis found a medium effect size (d=0.43) for the relationship between future discounting for food and monetary rewards and obesity [14]. This suggests that future discounting is a critical target for interventions aiming to modify risky lifestyle behaviors [15].

Episodic Future Thinking Interventions

Episodic Future Thinking (EFT), namely, a form of prospective thought, involves mentally envisioning oneself experiencing specific, personal, and detailed future-oriented events [16]. EFT is a type of psychological intervention that aims to disrupt one's tendency to discount the future by altering the perceived value of delayed outcomes [17]. EFT enables individuals to not only mentally experience upcoming events [18] but also consider both the immediate and extended consequences of their actions [19]. Specifically, EFT encourages individuals to consider the potential future consequences of current behaviors that have a detrimental impact on long-term health [12,20].

EFT is based on several components of interventions of behavior change technique (BCT) specified in the behavior change taxonomy [21] and Ontology [22]. These components include those related to goal-directed BCTs (ie, goal setting and comparing goals with current behavior) and increasing awareness of consequences of BCTs (ie, prompt comparative imaging of future outcomes). Nevertheless, it is important to understand why these techniques are effective by examining the mechanisms by which these intervention components change outcome behaviors [23,24]. The proposed mechanisms of action for EFT are a reduction in present bias or future discounting and an increase in long-term prospective thinking and valuation of future rewards [25,26], which fall under the domain of bodily disposition and, in particular, a mental disposition (ie, a temporal orientation toward the future) [27].

EFT interventions involve pre-experiencing positive, specific, vivid, future events that are either planned to occur or could realistically occur—a technique described as a form of "mental time travel" [28-30]. Specifically, it involves imagining *who* one is with, *what* one is doing, *where* one is, and *how* one is feeling. The process is used to generate a set of episodic cues to prompt participants to repeatedly practice visualization of such events. EFT cues have typically been generated using 2 approaches: interview-guided or survey-guided, with cue delivery methods usually involving text, audio, and/or visual formats [31].

Previous research has shown that EFT is effective for reducing future discounting rates. Three recent, large meta-analyses reported medium-sized effects in clinical and nonclinical populations [32-34]. These analyses showed that intervention effects were stronger when the imagined future events were more positive, vivid, and related to the delayed choice, as well as for those individuals with higher choice impulsivity, such as steeper discounting [33]. Specifically, EFT was more effective at promoting healthier decision-making in samples with higher choice impulsivity, which suggests that such training is more effective for people who are less likely to examine the future health consequences of their actions [35]. Thus, choice impulsivity may be a mechanism underpinning the effectiveness of EFT [36,37]. Furthermore, EFT has been shown to reduce the occurrence of a range of risky health behaviors, including cigarette smoking [38]; consuming less (approximately 300 calories) energy-dense, nutrient-poor food during an ad libitum eating task with adults and children with overweight and obesity [39,40]; and decreasing the consumption of alcoholic beverages in people with alcohol dependence [41].

Although EFT is a promising intervention technique, high variability exists in the specific methodology used. Specifically, previous studies have used various amounts of episodic cues generated, timeframes of future events, frequency of engaging with episodic cues, type of cue delivery (eg, text vs audio formats), and whether personal health goals are linked to the episodic cues. Further discrepancies include the control conditions used as comparators (eg, episodic recent thinking, standardized episodic thinking, and health information thinking) and whether the timeframe of episodic cues matches those of discounting time delays on various tasks and questionnaires used to assess such mechanisms of action [31,34]. Furthermore, studies have differed in intervention length and proximity of cue delivery to engaging in the targeted health behavior. Finally, most studies were conducted in the laboratory or face-to-face, which may warrant concerns regarding daily delivery of EFT cues to enable practice of imagery, as well as future intervention scale-up and implementation in real-world settings. While there is currently no standardized protocol for delivering EFT, it is important to address some of these limitations in developing future interventions. Specifically, it is critical to develop EFT interventions that can be scaled up for delivery in the real world and help people to make sense of and apply EFT in their everyday decisions.

More recent EFT studies have used mobile health (mHealth) and other digital technologies to deliver EFT interventions in real-world settings, with a number of studies using smartphones [15,42-44]. Nevertheless, these previous studies involved generating cues in person with a researcher, which were delivered via text or similar prompts to read daily and practice vividly imaging cues. Prior studies did not use smartphone apps to initially develop cues, nor to deliver them daily via in-built notifications. The independent use of a smartphone app also affords participants more autonomy. Studies using immersive environments such as virtual reality to deliver EFT have also emerged and have found promising results for fostering

XSL•FO
RenderX

future-oriented thinking [45,46]. Additionally, some previous studies have delivered the EFT cues via an audible format, namely, recordings of participants reading their cues [42,43,47]. A recent pilot study using a smartphone app to develop and create audio-based EFT cues reported that such delivery reduces engagement as participants felt uncomfortable listening to their own voice, and it was practically challenging to find somewhere suitable to listen to such recordings [48]. Thus, it may be more acceptable to deliver cues via written format that could be read out loud if preferred, but it is crucial to engage with end users to determine the most acceptable delivery mode. Importantly, no research to date has developed a user-designed comprehensive digital EFT intervention that enables initial cue generation and notifications to foster daily practice.

The Benefits of Co-Design

More broadly, one ongoing criticism of smartphone app development and design, including mHealth apps, is the lack of co-design methodology [49]. Co-design leverages insights from people's lived experiences to foster design-led discovery, innovative ideas, and creative problem-solving, with end users' unique perspectives, values, and needs being placed at the center of the design process [50,51]. The co-design approach also provides a way to ensure that a product, system, or service is suitable for the people for whom it is intended [52], which could mitigate some of the £100 billion (US $134 billion) per annum in research waste [53,54]. Over the last decade, repeated calls have emerged to adopt more participatory approaches to technology design to improve end user engagement [52,55] and to add more precision when tailoring content and interactions to community and individual needs.

This Study

Although previous research has used mHealth technology to deliver EFT interventions, no research to date has used co-design to create an engaging digital intervention with potential end users. Furthermore, prior research examining digitally delivered EFT interventions has focused on laboratory-based testing. More broadly, such limitations likely contribute to the lack of uptake, adoption, and scale-up of digital health interventions in real-world settings [56], which adversely impacts health outcomes [57-59]. Using co-design may help address the evidence to practice gap by enhancing adoption by end users, lowering cost and waste, and boosting behavior change by targeting the appropriate enablers [60]. Using co-design is particularly important for mHealth interventions, given it is the most direct technology-consumer link [61]. Nevertheless, high drop-off rates are common due to design issues as end users' preferences are not often considered [62]. Sustainable use of mHealth interventions could be facilitated using co-design processes [63]. Furthermore, the importance of documenting the co-design process used and how it has influenced the development of digital health intervention is necessary, given the current lack of standardized measures to assess impact [64] and the inconsistent reporting to date regarding specifics of the co-design sessions [65].

Thus, the purpose of this research was to co-design a smartphone app prototype of an EFT intervention, namely, "Luminaut." The app aims to support individuals with their health-related behavior change goals via EFT by drawing on the perspectives of adults in the "prerisk" phase for chronic health conditions. The "prerisk" phase refers to the stage before risk factors for chronic conditions emerge whereby individuals are still relatively healthy but aiming to prevent the development of chronic health conditions. Specifically, this study used an iterative co-design process to (1) understand the optimal characteristics and components (eg, content, format, and delivery) of the *Luminaut* app based on the perspectives of potential end users, and (2) gather end users' feedback based on interacting with the *Luminaut* app prototype.

Methods

Overview

This study used patient and public involvement, defined as research designed and conducted *with* or *by* members of the public rather than *to*, *about*, or *for* them [66]. Specifically, we used a co-design approach based on the British Design Council's Double Diamond Design Process—a framework that structures the design process into 4 stages: Discover, Define, Develop, and Deliver [67]. Co-design was chosen for its participatory nature, engaging users and stakeholders throughout the process [68]. The Double Diamond framework uses an iterative, user-centered process and involves collaboration between stakeholders to revisit and refine design solutions based on emerging insights and continuous feedback. Gero and Milovanovic's [69] Situated Function-Behaviour-Structure was used as an analytical framework as it provides a dynamic model for understanding how design evolves as a result of collaborative interactions, feedback, and ongoing refinement through the various stages of co-design. Using this lens, feedback, and decision-making among participants including users, researchers, and designers can help actively shape the development of the design, ensuring that it remains responsive to user needs and emerging insights.

Ethical Considerations

This study was approved by the Commonwealth Scientific and Industrial Research Organisation Human Research Ethics Committee (ID 2023_055_LR). All procedures were conducted in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Declaration of Helsinki (1975), as revised in 2000. We adhered to the GRIPP2-SF (Guidance for Reporting Involvement of Patients and Public Short Form) reporting checklist, which is used for reporting the involvement of people with lived experience in research [66]. To maintain participant privacy, the authors deidentified any personal information contained in the data and results so that individual participants cannot be identified.

Participants and Sample Size

Participants were adults aged 25-44 years. This age group was selected as individuals in this age group are considered to be in the "prerisk" phase for chronic health conditions [70] and are more likely to adopt and use health apps [71]. In line with co-design conventions, we aimed to recruit a minimum of 8 participants per co-design workshop, with no more than 20 - 25

XSL•FO
RenderX

participants in a single workshop and up to a maximum total sample size of 36. Sample size calculations for co-design research are focused on accessing the viewpoints of different stakeholder groups, the development of saturation of insights, and deep engagement regarding the research questions, rather than focusing on statistical significance [72,73]. In many generative co-design practices, the focus is on accessing tacit and latent knowledge and insights through exploratory practices, rather than on accessing explicit knowledge, behavioral data, or generalizable "facts" [74].

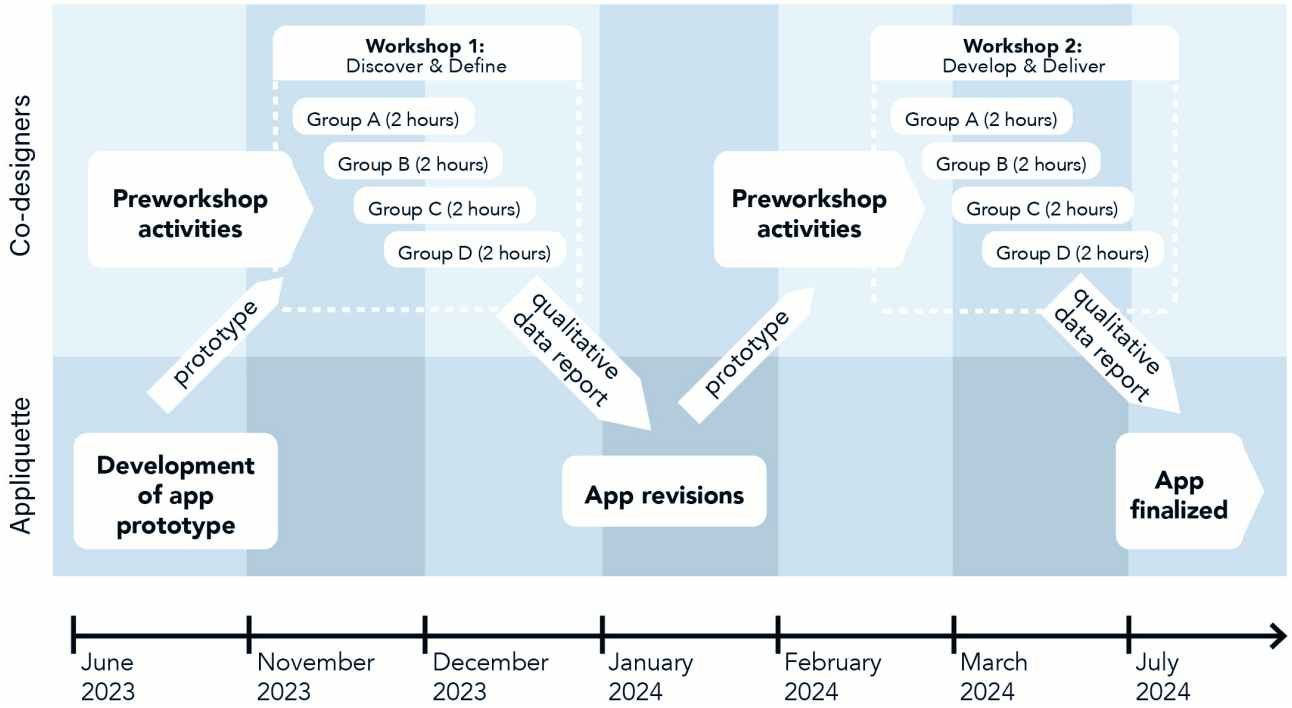
Co-Design Facilitator Team

The co-design project team comprised a research team including 2 behavioral scientists, a nutrition scientist, and a research dietitian who was also the study coordinator, along with a team of 3 design researchers with expertise in co-designing for health.

Study Design

An overview of the study design is illustrated in Figure 1. Briefly, a prototype of the Luminaut mobile app was developed by a third-party app developer (Appliquette Pty Ltd). Once a prototype of the app was developed, participants installed the app on their personal mobile devices and provided feedback via a face-to-face workshop ("workshop 1"). Revisions to the mobile app were then made and presented to participants in a second workshop ("workshop 2"). The app was revised a second and final time to reflect any additional feedback raised.

Figure 1. Schematic representation of the study design.



Development of the "Luminaut" Mobile App Prototype

Before commencing engagement with the co-design facilitators, the researchers worked directly with Appliquette to develop an app prototype as a starting point for the design challenge that translated the mechanics of the principles of Future Discounting into an app-based experience. This iterative prototyping approach is commonly used in user experience design and provided a shared starting point of (immediate) past experience for the workshop, rather than relying on having had experience with other analogous smartphone apps [75].

A schematic depiction of the app was developed using mood boards in a collaborative interface design tool, Figma, as a way for Appliquette to develop iterative prototypes and collect feedback from the researchers to satisfy the functional requirements of the app. Figma was used to cocreate and illustrate the page layout and order, content (eg, formatting and arrangement), functionality (eg, help and back buttons), images and animations, and different interface displays. The app was pilot-tested by a minimum of 2 researchers during all stages of development. The content of the app is outlined in Textbox 1. Of note, participants can be allocated to complete all or any combination of these app components but were asked to complete all components in this study.

XSL•FO
RenderX

Textbox 1. Luminaut app components and specific activities.

Health Goals

- Users selected a health domain (ie, physical activity or dietary behavior).
- Users created 3 health goals (related to the chosen health domain) that were achievable within 3 predetermined timeframes (3, 6, and 12 months).
- Users rated how achievable and important their goals were (1=very achievable or unimportant to 5=very achievable or important).
- Users who rated health goals as 3 or less were asked to regenerate a goal that was important and achievable.

Future Events

- Users described 3 future events that they have planned or anticipate could take place within 3 predetermined timeframes (3, 6, and 12 months).
- Each event had to be positive, vivid, and described using present tense, and must occur within a 24-hour timeframe.
- Users rated how enjoyable, important, exciting, and vivid each event is (1=very achievable or unimportant to 5=very achievable or important).

Episodic Cues

- Combine health goals with each positive future events to form cues delivered via daily in-built notifications (personalized).

Ecological Momentary Assessment

- Questions to capture momentary mood and emotions included
- Fatigue
- Sleepiness
- Alertness
- Energy

Procedure

Participants were recruited through paid advertisements on Facebook and Instagram (Meta). Interested individuals were directed to complete a lead generation form, which prompted them to enter their name and preferred email address. Those who expressed interest in participating were emailed a brief online questionnaire hosted in REDCap (Research Electronic Data Capture; Vanderbilt University, 2004), which included an electronic participant information sheet and consent form that outlined the study inclusion and exclusion criteria. Individuals were required to tick a checkbox to indicate electronic informed consent. Those who provided informed consent were directed to a set of questions to determine eligibility, dietary requirements, availability (ie, date and time), and preferred workshop location. Those who completed the initial questionnaire were assigned a unique identifier code, which was generated in REDCap. To be eligible to participate, individuals needed to be (1) aged between 25 and 44 years, (2) fluent in the English language, (3) have access to an internet-connected smartphone device (iOS or Android) that has the minimum required operating system (ie, iOS11 or Android 6), (4) willing and able to install a free prototype of the "Future Self 4 Health" smartphone app on their device, and (5) able to attend up to 2 face-to-face workshops held in a metropolitan area of Adelaide, South Australia. Individuals with a self-reported diagnosis of aphantasia, namely, the inability to imagine or visualize or create mental images, were not eligible to participate. Of note, the app was originally called "Future Self for Health" and was then later changed to *Luminaut* following feedback obtained during the co-design process (for more information, see Multimedia Appendix 1).

Those deemed eligible were provided a unique link to complete an online survey to collect demographic information (ie, age, gender, education level, fluent speaker of languages other than English, health conditions, postcode, and smartphone device) hosted in REDCap. Participants were then contacted by the research coordinator and sent a follow-up email containing preworkshop activities and specific details about the face-to-face workshops. Participants received an honorarium in the form of an e-gift card up to the value of Aus $100 (US $71.48) for each workshop that they attended.

Preworkshop Activities

The *Luminaut* prototype app was free for participants to download via the App Store (Apple) or Google Play Store (Android) using beta-testing links (ie, TestFlight). Prior to workshop 1, participants were emailed instructions about how to install and login to the app and information concerning their allocated workshop (ie, date, time, and venue), as well as an infographic to familiarize participants with the concept of future discounting. Participants received a link to download the app prototype and were asked to test its functionality and features over approximately a 1-week period.

Co-Design Workshops

A series of 8 workshops (4 iterations of workshop 1, and 4 iterations of workshop 2) were conducted between November 2023 and March 2024. Participation involved completing up to two, 2-hour face-to-face workshops. Co-design workshops were facilitated through Match Studio (a creative research laboratory at the University of South Australia). To increase participant diversity, a series of workshop times and locations were offered, with the aim to recruit a sample of individuals with different

XSL•FO
RenderX

levels of socioeconomic advantage. The locations chosen encompassed Local Government Areas (LGA) with both high and low rankings according to the Australian Index of Relative Socio-economic Advantage and Disadvantage (IRSAD) [76]. The two locations in South Australia were (1) the University of South Australia City West Campus, Adelaide (LGA IRSAD quintile 5); and (2) the Christie Downs Community House, Christies Downs (LGA IRSAD quintile 3) [76]. Two time slots were offered: daytime (10:30 AM to 12:30 PM) or evening (5:30 PM to 7:30 PM).

In response to participant preferences, most workshops (6/8, 75% comprising 4 daytime and 2 evening) were facilitated at the University of South Australia City West Campus as well as 2 daytime workshops at the Christie Downs Community House. The workshops used a combination of participant-led documentation and recording during the sessions. Where small group verbal sharing took place, these parts of the workshops were recorded using Dictaphones (Panasonic RR-US065, Olympus VN-731PC, and Sony R41174041), while facilitators also took field notes to aid in transcription and thematic analysis. Light refreshments were offered to participants throughout as a way of offering additional value to participants and providing an opportunity to invert power relationships between the facilitators (serving) and participants (receiving).

Workshops were informed by previous studies and iteratively by findings arising from each co-design session, in accordance with the Design Council's Double Diamond Design Process framework (ie, Discover, Define, Develop, and Deliver), which combines both divergent and convergent approaches [67]. The first workshop was designed to engage with the "Discover" and "Define" phases of the Double Diamond, using creative activities to help participants access and share tacit and latent, as well as explicit, knowledge [74]. Activities involved facilitated discussions and participant-led documentation (worksheets, low-fidelity prototypes, and card-sorting activities). A timeline of activities during workshop 1 is provided in Table 1, with a longer description of each workshop contained in Multimedia Appendix 2.

Table . Timeline of activities during workshop 1.

Time, minutes	Arrival
5	Opening remarks: Acknowledgment of Country, purpose and objectives of the workshop
15	Future discounting introduction and icebreaker
30	Activity 1: Card sort: What is and what isn't future discounting
25	Activity 2: Where and when activity stories
10	Break
30	Activity 3: App inventory exercise
30	Activity 4: Specific app test feedback
5	Wrap up
	Close

Following the completion of each instance of this workshop, insights were synthesized to clarify the app's intended function, the behaviors it aimed to support, and to refine the design objectives. Across the workshop 1 series (ie, 4 workshops), small changes were made to the length of time spent on each of the activities to respond to the experiences and interests of the co-design participants. At the completion of the 4 instances of this workshop, the data were analyzed by AD and KL using inductive thematic analysis [77]. The analysis was then translated into an interim report that was shared with the researchers who reviewed the collated feedback in collaboration with AD and KL. The researchers devised a list of changes to be made to the app. Appliquette Pty Ltd addressed and implemented the changes, which were categorized as "high," "medium," and "low" priority, and released an updated version of the app to participants.

A few weeks prior to workshop 2, participants were given the opportunity to test the revised version of the app (over approximately a 1-week period) and were asked to provide feedback and reflections via an anonymous online survey, hosted via the Qualtrics survey platform (Qualtrics). This survey was designed to prompt open-ended reflection while also asking participants to compare what they had been testing with the original prototype. Participant contributions to this survey were analyzed through an inductive thematic analysis process and were used to shape the design and focus of workshop 2, which engaged the "Develop" and "Deliver" phases of the Double Diamond process. A timeline of workshop 2 is provided in Table 2.

XSL•FO
RenderX

Table . Timeline of activities during workshop 2.

Time, minutes	Arrival
5	Opening remarks: Acknowledgment of country, purpose, and objectives of the workshop
5	Future discounting introduction: Recap from previous workshop
20	Activity 1: Explaining future discounting to a friend using storyboard template and noun project icons
10	Activity 2: Development of sorting game activity based on selected icons
(10)	Optional activity: Financial and other barriers to achieving goals. Mapping example icons and types of visualizations to financial work, family, stress, etc, barriers
10	Break
15	Activity 4: Metaphors and dashboards review Avatars, tree growing, looking through portal, adjusting a date, advent calendar, road, stepping stones/islands, seasons, graphs/abstract data visualization, list
40	Activity 5**:** Specific feedback on current app state • What milestones give a badge? • How much difference does a photo make? What if it is a stock image? (Use cards from first workshop + icon cards)
5	Wrap up Close

Thematic Analysis

Involvement in the co-design workshops allowed researchers to immerse themselves in the data. The reflexive thematic analysis framework established by Braun and Clarke [77] was used for inductive coding, namely, generating codes directly from the data rather than from preexisting categories. Following each workshop, all data, including participant-led documentation, researcher field notes, transcribed audio recordings, and survey responses, were reviewed to identify key themes and patterns and coded using NVivo (QSR International) and Microsoft software (Microsoft Inc, ie, Excel and Word).

Data from each of the 9 activities across the workshops (workshop 1 series=4 activities, workshop 2 series=5 activities) were reviewed independently by the co-design facilitators after each iteration of the workshop to ensure that the workshops were effectively contributing to the ongoing development process. The co-design facilitators made small changes, such as adjusting the time spent on different activities, to ensure that all intended explorations in the workshop were undertaken across the workshop series. Initial themes were also reviewed by the co-design facilitators to develop a rich understanding of the data following Braun and Clarke's [78] recent guidance on quality practice in reflexive thematic analysis at the end of the workshop series to develop recommendations and overarching themes.

Across the series of co-design workshops, the process of theme development involved multiple cycles of coding, categorizing, and revising. Themes were reassessed at each stage to ensure that they accurately captured the core patterns in the data, with codes merged, split, or redefined as necessary. This iterative approach enabled the identification of key themes that emerged as most relevant to the area of enquiry [77]. A final report was then generated by the co-design team for the nutrition and behavioral scientists (Multimedia Appendix 3). The report included high-level insights from the co-design process, a description of the workshop structure and activities, interim recommendations following workshop 1 series, data from the asynchronous feedback process, and data from both workshop series.

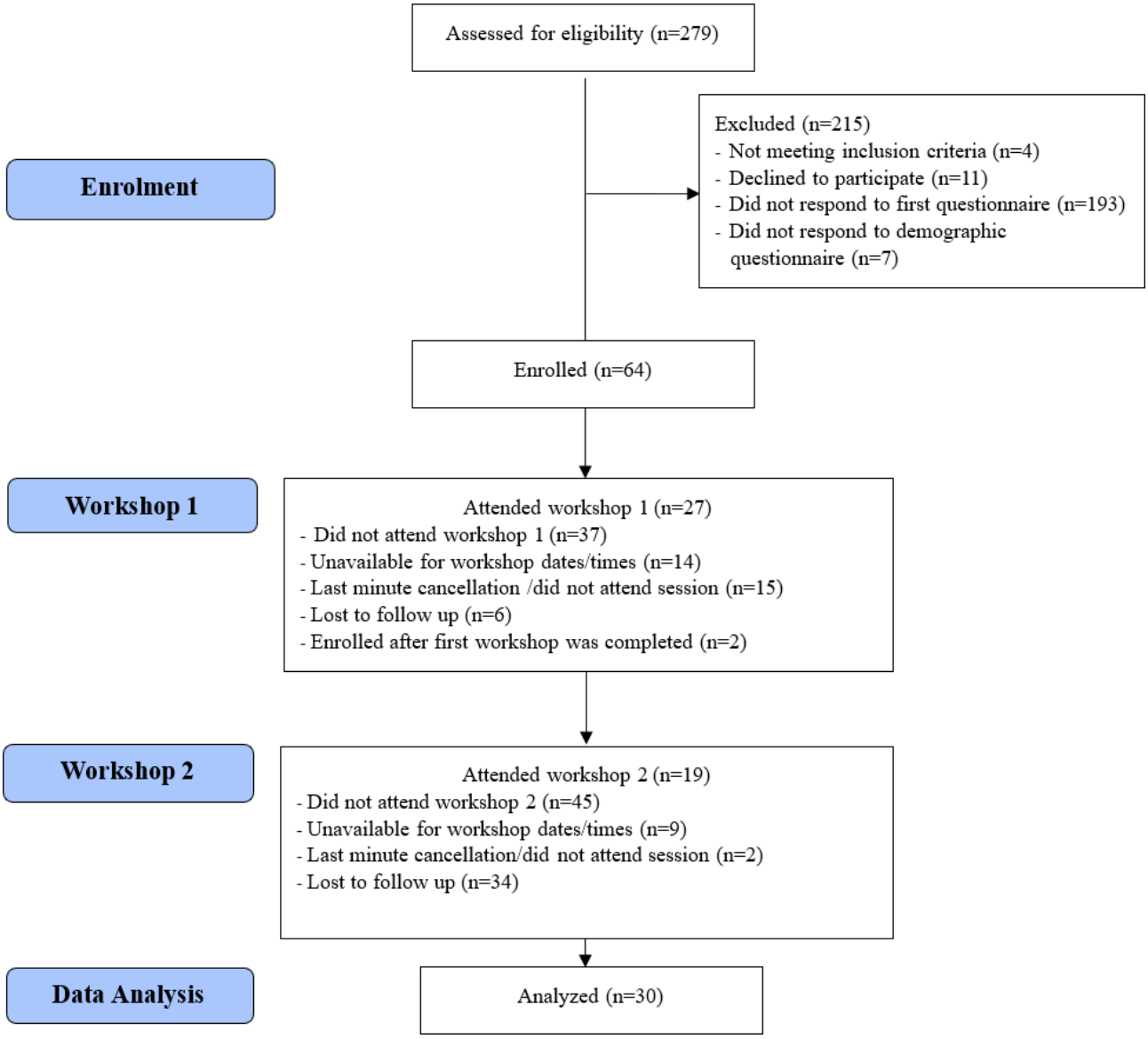
Results

Participant Recruitment

A total of 279 individuals expressed interest in participating in the study via the social media (Facebook and Instagram) advertisements. Of these, 71 individuals provided informed consent, completed the follow-up questionnaire, and met the study eligibility criteria, following which 64 (90%) individuals completed the demographic questionnaire and were enrolled into the study. Eligible participants were contacted by the research coordinator (MR) to confirm their availability to participate. A total of 30 (47%) participants attended at least 1 workshop, 16 (25%) participants completed both workshops, 11 (17%) participants completed the first workshop only, and 3 (5%) participants completed the second workshop only (Figure 2).

XSL•FO
RenderX

Figure 2. Participant flow diagram.



Sample Characteristics

The characteristics of the participants who attended workshops are shown in Table 3. Participants were mostly women aged 25 - 44 years with a bachelor or postgraduate university degree, who owned an Apple smartphone device, had relatively high socioeconomic advantage in accordance with the Australian Bureau of Statistics Socio-Economic Indexes for Areas, Australia [76], and lived with no or 1 chronic health condition. The characteristics of all participants who were initially enrolled into the study are provided in Multimedia Appendix 4.

XSL•FO
RenderX

Table . Characteristics of participants in workshops 1 and 2 as well as the total sample.

Characteristic	Workshop 1	Workshop 2	All attendees
Total, N	27	19	30
Female, n (%)	24 (88.89)	17 (89.47)	27 (90)
Age (years), mean (SD)	36.74 (5.52)	37.37 (4.96)	36.37 (5.65)
Age range (years)	25 - 44	28 - 44	25 - 44
Education, n (%)			
Secondary education (year 12)	—[a]	—	—
Certificate III and IV level	2 (7.41)	—	2 (6.67)
Advanced diploma/diploma level	1 (3.70)	—	1 (3.33)
Bachelor degree level	8 (29.63)	9 (47.37)	11 (36.67)
Graduate diploma/certificate level	7 (25.93)	4 (21.05)	7 (23.33)
Postgraduate degree level	9 (33.33)	6 (31.58)	9 (30)
IRSAD[b] quintiles, n (%)			
1 (most disadvantaged)	1 (3.70)	1 (5.26)	1 (3.33)
2	1 (3.70)	—	1 (3.33)
3	8 (29.63)	5 (26.32)	9 (30)
4	4 (14.81)	4 (21.05)	5 (16.67)
5 (most advantaged)	13 (48.15)	9 (47.37)	14 (43)
Fluent in languages other than English, n (%)	10 (37.04)	5 (26.32)	11 (36.67)
Number of chronic health conditions, n (%)			
0	11 (40.74)	10 (52.63)	14 (46.67)
1	12 (44.44)	6 (31.58)	12 (40)
2	2 (7.41)	2 (10.53)	2 (6.67)
3	2 (7.41)	1 (5.26)	2 (6.67)
Smartphone device type, n (%)			
Apple	16 (59.26)	11 (57.89)	18 (60)
Android	11 (40.74)	8 (42.11)	12 (40)

[a]Not available.

[b]IRSAD: Index of Relative Socio-economic Advantage and Disadvantage—SEIFA Postal Area (POA) [76].

Specific App Feedback From Co-Design Engagement

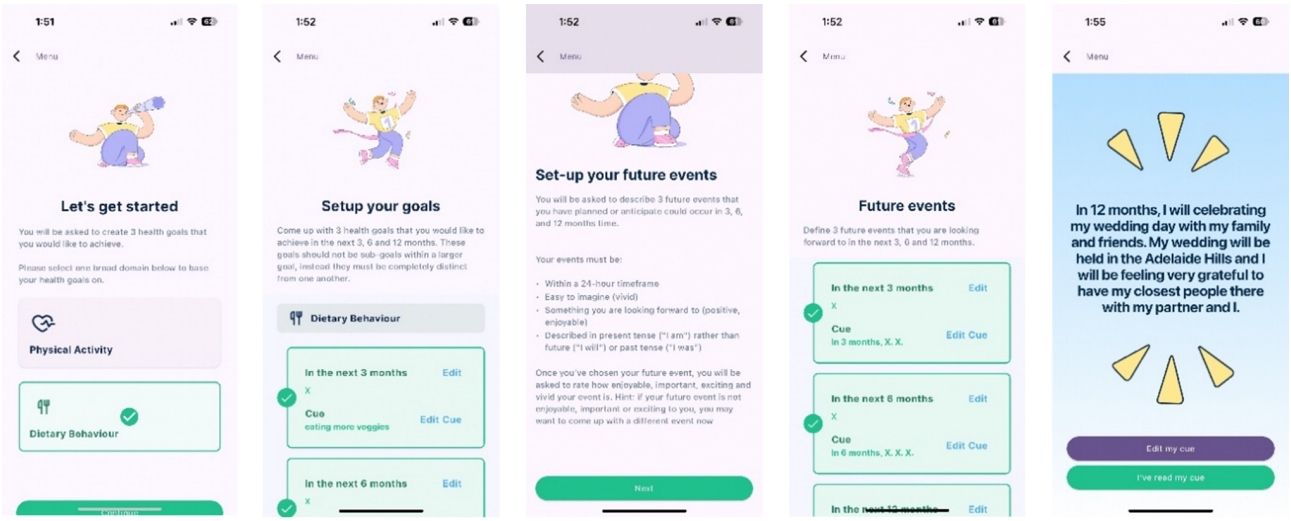
At the end of the first workshop series, 26 suggestions for changes to the *Luminaut* app were identified from the data (Multimedia Appendix 3). These suggestions ranged from high-level conceptual suggestions (eg, shifting "language to be more positive rather than the negative framing of 'future discounting'") to specific feedback (eg, "lock or add friction to reentering quiz after completion to avoid looping"). The research team recommended implementing 16 of these suggestions within the app, which were categorized as high (n=8), medium (n=6), or low (n=2) priority. A high priority suggestion included adding in clearer instructions and screenshots that showed a preview of how to navigate the app, while a medium priority suggestion included using images to represent goals (icons or stock images) and a low priority suggestion included enabling push notifications that could be customizable based on personal preferences by asking participants to nominate suitable times of day practice visual imagery using the cues. Practical considerations including time and budget were the primary reasons for not implementing all user suggestions at this stage; however, these suggestions were noted for future possible versions of the app. An additional 4 suggestions were identified as needing further exploration and bespoke activities were developed to explore these ideas in the second series of workshops. These suggestions mostly included exploring an introduction narrative or ways to introduce the concept of "future discounting" quickly, as well as opportunities for various visual metaphors to be used as a structure for the app dashboard.

During the second workshop series, participants provided further feedback on the app design (see Figures 3A-3E for screenshots of the updated app prototype). Specifically, they provided

XSL•FO
RenderX

specific feedback on the experience of using the updated prototype, while also exploring the higher-level conceptual challenges identified during the first workshop series. This dual focus enabled the co-design process to explore ideas that could reasonably be integrated and explored in this project and were appropriate to the project stage, while also contributing to strategic opportunities for future research. Key feedback on the app included that there were "too many notifications" and exploring whether sentence starters or generative artificial intelligence could be integrated to help participants create realistic, yet aspirational health goals. Facilitators reflected that there was a tension between viewing the app as a goal setting and achievement app and understanding the principles of catalyzing subtle behavior shifts away from future discounting; in effect, a microexample of the challenge of future discounting itself.

Figure 3. . Design of the Episodic Future Thinking (EFT) Luminaut smartphone app. (A) The screen where participants select a health domain (physical activity or dietary behavior); (B) the instruction screen for setting up health goals for the 3 timepoints for the selected health domain; (C) the instruction screen for setting up the future events; (D) the home screen for the 3 timepoints for the future events; and (E) an example of an EFT cue delivered to participants daily to assist in practicing visual imagery.



A **B** **C** **D** **E**

A total of 8 narrative elements were identified for use within the app. The most common was variations of "make good choices now to enjoy the benefits later," which mirrored the introductory content presented to participants. However, a novel conception also emerged with "planting a seed and nurturing a plant" providing a metaphor that is positive in both the present and future framings. Preference for navigational metaphors also followed this biophilic concept with growth and trees identified as the most preferred way of arranging information. Metaphors that were identified as inappropriate tended to separate time into discrete moments rather than a continuum, including a portal to a future time, moving between islands, and jumping seasons.

Insights Into Future EFY Intervention Design

Overview

As well as delivering specific feedback that could inform the ongoing app development, the project also enabled the development of higher-level principles and learnings that can be widely applied in similar projects and systems using EFT interventions. Five overarching themes were identified, which were each grounded in the lived experiences of participants and integral to the design decisions made during the project. The five themes were (1) Perceived Value, Understanding, and Trust in Shaping Acceptance of the EFT Concept; (2) Factors Influencing Engagement: Triggers and Barriers; (3) The Role of Personalization in Enhancing User Experience; (4) Game Mechanics as Drivers of User Motivation; and (5) The Importance of Usability in Optimizing User Experience.

Perceived Value, Understanding, and Trust in Shaping Acceptance of the EFT Concept

Concept acceptance encompassed participants' understanding and trust in the underlying premise of the app. While many participants recognized the app's potential to facilitate behavior change, there was notable confusion regarding some aspects of the intervention, including its functionalities and intended outcomes. One participant remarked that, "[It needs] some explanation or context to help know what to do (e.g., infographic)" and another requested, "...guidance about how to set a goal and how to achieve it. 'Help me figure it out.'" This underscored the necessity for clearer communication about how the app would support users in their health journeys, particularly highlighting the importance of an introductory guide to convey this information effectively.

Trust also emerged as a crucial factor with participants indicating that the app's credibility and the reliability of its recommendations were essential for their ongoing use and for considering recommending it to others. The participants sought to use an app from a brand they could trust. They also sought evidence of scientific foundations and anecdotal evidence or direct experience of successful outcomes from using the app. One participant said, "I have to see the complete product, finished, use it for a bit and see that it is working for me." Others

XSL•FO
RenderX

remarked that they would need to see "scientific evidence" or have the app "recommended by a psychologist."

The concept of EFT is grounded in scientific evidence; however, communicating this effectively is challenging due to the complexity of the psychological underpinnings, a challenge that participants identified. This theme was explored iteratively throughout the co-design process, with potential solutions collaboratively developed during the co-design sessions to enhance clarity and understanding.

Factors Influencing Engagement: Triggers and Barriers

Participants identified various triggers that could influence their consideration of using the app, such as hearing about the illness of a friend or family member, health events, or personal milestones. These triggers could serve as motivators to engage with the app, prompting users to seek support for their health journeys. Examples include, "Family member diagnosed with illness, eg, diabetes," "Bad news from the GP," and "Change in location/moving."

Barriers were also identified that could hinder their engagement. These included competing priorities in daily life ("What are all the other things that I should be doing?"), time and effort required to use the app ("Effort, mental effort"), aversion to constraint ("You don't want to be the type of person who denies themselves"), difficulty navigating unfamiliar or complex tasks ("Inability to prioritise can be a big barrier, that overwhelm feeling"), technological illiteracy ("Some people don't even know how to install an app"), and financial cost ("Healthy food is expensive"). Understanding the interplay between triggers and barriers can help inform design decisions, ensuring that the app effectively supports users in overcoming obstacles while leveraging motivating factors to encourage sustained use.

The Role of Personalization in Enhancing User Experience

The theme of personalization emerged prominently, with participants emphasizing the significance of tailoring the app experience to individual needs and preferences. Many expressed a desire for customizable features, including editable goals, the ability to change notifications (time and frequency), and options to select avatars and images that reflect their personal preferences and circumstances. For example, one participant requested the "[o]ption to switch off [notifications]. [I'm] trying hard to reduce phone usage, for health." Participants indicated that a personalized approach could foster a deeper connection to the app, which may ultimately enhance motivation and adherence. This insight informed design decisions, resulting in the inclusion of flexible options.

Game Mechanics as Drivers of User Motivation

Overview

Participants showed enthusiasm for gamification elements within the app, believing that game-like features could enhance user engagement and make the experience more enjoyable. Suggestions included reward systems, challenges, social sharing capabilities, progress tracking, and streak tracking. Participants noted that incorporating these elements could transform engagement with the app into a fun and interactive experience, fostering community support and accountability, while offering rewards that reinforced positive behavior. Integration of gamified components designed to encourage sustained interaction and facilitate progress tracking is planned for future iterations of the app. The integration of these gamification elements is expected to create a more engaging and motivating user experience.

Reward Systems

Several participants suggested implementing a reward system, where users could earn rewards for completing tasks. For example, "[I] like goal setting aspects...and small rewards in-app for progressing." Suggested rewards included frequent flyer points, financial credit, vouchers, and in-app rewards (eg, badges). One participant mentioned that the app's use of positive language feels rewarding, remarking that, "Positive language is itself a reward. Maybe you don't need [other] rewards."

Challenges

Participants expressed interest in participating in challenges. One participant stated, "I like to enter competitions."

Social Sharing

The potential for social sharing features was also highlighted. Multiple participants contributed, with one remarking that, "Connecting with friends is a big one.[it's] encouraging and inspiring to see other people, including family and friends," and another commented, "I like that functionality, where you can give someone a virtual high five. Celebrate the fact that they reached the 100-day streak or whatever."

Progress Tracking

The ability to track progress through gamified elements was noted as a key feature, to record and present information for direct feedback and tracking over time, useful for monitoring progress toward personal goals. During the discussion, participants expressed their views: one noted, "[I like to] see information at a glance…[to] check if I'm improving," and another remarked, "Tracking...a way of showing in the app, this is where I started, and this is where I got to..."

Streak Tracking

The concept of tracking streaks was particularly popular among participants. One participant stated, "Streaks are motivating."

The Importance of Usability in Optimizing User Experience

Participants were prompted and encouraged to provide feedback on their lived experiences with the app. Bugs and glitches were identified and addressed throughout the process. Participants emphasized that an easy-to-navigate interface is essential for ensuring engagement and accessibility. Feedback underscored the need for simplicity and intuitiveness in app design, highlighting the importance of clear instructions, straightforward layouts, and visually appealing design elements. By prioritizing usability, there is an opportunity to create an engaging environment that fosters positive interactions with the app.

XSL•FO
RenderX

Discussion

Principal Findings

This study involved co-designing a digital health intervention, "Luminaut," which is aimed at disrupting future discounting via EFT in individuals at risk of developing chronic health issues. The co-design approach that was used generated app-specific feedback for technical improvements and led to high-level conceptual suggestions that helped shape the Luminaut intervention. This study revealed 5 high-level conceptual suggestions that were important to the participants when engaging with digital EFT interventions such as the Luminaut app. These themes included concept acceptance, triggers and barriers, personalization, gamification, and a user-friendly interface. Overall, the co-design approach was integral to shaping the Luminaut app and it also provides insights into EFT interventions more broadly regardless of delivery mode.

The first theme identified was that *concept acceptance* of future discounting may promote participants' understanding and trust in the underlying premise of the app. The co-design process provided valuable insights into how participants receive the concept of future discounting in the context of app creation and how they would like to interact with it. Specifically, the high-level conceptual suggestions related to this theme included the overarching challenge of engaging with the concept of future discounting. The concept of future discounting was perceived as being primarily negative in nature and it was suggested that more positive framing of the concept such as "future investing" be used when describing the intervention to participants, which should be considered when delivering EFT or other psychological interventions aimed at reducing future discounting. While it was relatively straightforward to include an explanation of the term "future discounting" into subsequent versions of the Luminaut app, it prompts further consideration of lay understandings of complex psychological phenomena. These contextual insights may elude formal assessments of efficacy, such as randomized controlled trials, which are necessarily more structured and less flexible; however, when combined, they may offer complementary insights at different stages of the research. Through the open, iterative, and creative process, participants shared their evolving understanding and active engagement with the concept, rather than simply providing feedback. By sharing their learning journey, participants revealed both the process of understanding and what they already know about the future discounting concept based on their prior experiences.

In addition, it was found that there are specific *triggers and barriers* that motivate or hinder individuals' engagement with the Luminaut app. Specifically, the participants identified that the recommendation of the app by a health care professional (eg, a psychologist) would boost trust, which expands on previous co-design research indicating that participants would be more willing to use an app if it was prescribed [79]. Such findings support the idea of using such apps as a way to empower individuals who are ready to take action toward health-related behavior change, which is consistent with the findings of previous co-design work in participants living with chronic health conditions [80].

The finding that participants valued *personalization* based on individual needs and preferences aligns with prior research indicating that profiling, tailoring, and autonomy are important when engaging individuals with digital health solutions for chronic disease prevention or management [79,81,82]. Taken together, such findings are a critical reminder that there is no "one-size-fits-all" approach when it comes to health care and that creating digital technologies that align with individual needs is a promising strategy for health behavior change [83]. Indeed, an advantage of the co-design process is its ability to identify diverse collectively agreed factors that may enhance engagement with such interventions to promote health behavior change.

In addition, elements of *gamification* including reward systems, challenges, social sharing, progress tracking, and streak tracking were also mentioned as features that could elicit a more enjoyable experience while using the app. The latter finding supports previous work showing that apps of gamified features in nongame settings, particularly in digital interventions for health and chronic conditions, are growing rapidly and may enhance the perceived experiences of users [84]. Future work is needed to determine whether the inclusion of such elements has benefits beyond boosting short-term engagement as indicated in recent systematic reviews [85].

Finally, a *user-friendly interface* was encouraged to ensure engagement and accessibility. The latter speaks to the idea of ensuring that digital interventions have high levels of usability, which may be enhanced via incorporating simple features (eg, a back button) and should be systematically evaluated such as via the System Usability Scale score [86].

Learnings From the Co-Design Process

One of the key learnings from the co-design process included the need to iteratively develop the workshop content based on participant feedback. Namely, the co-design process involved not only engaging with end users to co-design the product (ie, the Luminaut app) but the research process itself. For example, the co-design workshops involved exploring how the concept of future discounting could be communicated quickly and effectively, and the development of an overarching navigational metaphor for engaging with the future. A key piece of feedback from the first series of workshops was that the way the co-design facilitators explained the concept of future discounting made the whole project "make sense." Thus, in the second series of workshops, the facilitators constructed an activity using a storyboarding template to focus on how participants might explain the concept of future discounting to their friends.

In addition, our multidisciplinary approach provided both challenges and benefits, which is common when using such approaches [87]. For example, there were discipline-based differences regarding conceptualizations of bias in research whereby co-design practitioners consider health researchers a key stakeholder to include as participants in co-design workshops. In contrast, the nutrition and behavioral health scientists viewed researcher input as a potential source of bias in such settings, given their status as "experts" in their given

XSL•FO
RenderX

field. In this case, it was deemed useful to have one of the content experts (ie, researchers) present at each workshop as an observer and to answer any content-specific questions if they arose from participants.

Strengths and Limitations

This study had several strengths and limitations. While co-design studies often focus on involving end users in the conceptualization of research questions prior to any research taking place, this study was focused on exploring how access to support for overcoming future discounting behaviors might be scaled through digital technologies. The requirement of the research funding to develop a digital intervention served as a limitation; however, the advanced prototyping that was enabled by the imposition of this constraint enabled richer explorations of how technology mediated health interactions to take place. Indeed, other similar research aimed at developing digital health interventions has also used prototypes (ie, a website) as it provides stakeholders with a tangible artifact to engage with during the iterative testing process [88]. In this study, as the app design progressed, findings from the "Develop" and "Deliver" phases were used to refine earlier stages, ensuring that the design remained responsive to user needs and stakeholder input. This iterative process ensured that the design was continuously aligned with evolving insights. Nevertheless, it should be noted that this approach does necessitate the upfront allocation of time and money to support this iterative, stakeholder-engaged process, albeit the reward is a more sustainable and appropriate intervention in the medium to long term.

Although our sample was relatively large, which was a clear strength, the sample was subject to several limitations, such as comprising mostly female participants who spoke English fluently with high levels of education. Participants also had a relatively advantaged socioeconomic status, despite attempts to promote diversity through holding workshops in disadvantaged areas (eg, in the outer southern areas of Adelaide). Thus, the Luminaut app may be more tailored toward affluent women in terms of the look, feel, and technological familiarity. Of note, the lack of a representative sample in this study is consistent with previous work and remains an ongoing challenge in health co-design research [80]. Nevertheless, others have argued that as qualitative research relies on purposive sampling methods, there is less emphasis placed on achieving statistical generalizability than there would otherwise be when using quantitative methods [89].

Another limitation is that the 2 selected workshop locations may have excluded those individuals living in rural and remote areas and/or without readily available access to technology such as smartphone devices, which may have further contributed to the "digital health divide" based on an individual's social status [90]. Indeed, individuals with low socioeconomic status as well as older adults and ethnic minorities typically have lower digital literary and more difficulty accessing digital health interventions [91]. Such accessibility issues may be addressed in future research using non–digital health intervention solutions or via a digital and/or physical asynchronous co-design process. Specifically, the concept of "low-contact co-design" involving the use of technology to replicate the use of physical workshops conducted at different times and locations has been proposed to navigate the process of co-design when physical proximity is a challenge, which provides co-designers with new opportunities and possibilities for engaging with communities [92].

An additional strength of this study was the use of co-design methodology, which generates unique, insightful contributions and progresses unknowns. Nevertheless, such methods do not have the significance of larger-scale population research that will be derived from future clinical trials aimed at examining the efficacy of the app for improving health outcomes. Co-designing with a predetermined output (eg, a smartphone app–based intervention) is challenging as it constrains the process. However, we were subject to limitations of the funding for the current project and funding models more broadly, which require precise descriptions about the format and delivery of interventions in funding applications. While this approach allowed the research team to focus on exploring digital solutions, it may have limited the ability to critically evaluate whether an app was the ideal solution for delivering this intervention. Notwithstanding, a key component of EFT training involves sending reminders or prompts to participants to practice the mental imagery exercises, which is easily facilitated via techniques such as automated smartphone notifications. Since smartphone app development remains expensive, there is an ongoing tension between deep engagement with members of the community and conducting consumer market research or beta testing, the latter of which may be more economically efficient. Notwithstanding, consumer market research would not allow unique insights into EFT intervention development and the concept of future discounting more broadly.

Finally, the inclusion of both highly experienced facilitators with expertise in co-design [68] and content experts in nutrition and behavioral science within the project team is considered a strength and undoubtedly enhanced the research process. Indeed, the use of multi-disciplinary teams with different content expertise and methodological viewpoints is deemed critical for developing digital health solutions for complex problems [93].

Implications for Future Research and Health Policy or Practice

This study provides a blueprint for other researchers to determine the feasibility of using co-design practices in developing digital health behavior change mobile apps as well as EFT interventions more broadly. Furthermore, it would be useful to determine the extent to which co-designed apps enhance effectiveness in a more formal evaluative capacity. Nevertheless, the growing use of such practices speaks to the potential for creating sustainable and personalized digital preventive health programs to target a reduction in chronic conditions, a key priority area for Australian and global health strategies [94]. Finally, the insights gained from this project will help inform future large-scale clinical trials in prerisk or at-risk groups aimed at testing the feasibility, acceptability, and efficacy of the intervention alongside ongoing usability testing. Such trials could also serve the dual purpose of highlighting the cost-effectiveness of digital health interventions via an economic

XSL•FO
RenderX

evaluation [95], which is critical for reducing the gap between research and end user needs and increasing clinical research utility [96].

Importantly, this project represents the first phase of the co-design process. Specifically, co-design uses an iterative process whereby consumers and other key stakeholders are continuously involved in refinement and development of the intervention in response to ongoing feedback and changing needs. In addition, future co-design work could involve health professionals who may be involved in facilitating behavior change or self-management of chronic conditions in individuals. Indeed, the current finding that trust could be enhanced via recommendations from credible sources (eg, psychologists), which is essential for ongoing use, supports the idea of engaging health care professionals in the co-design process. Furthermore, previous research has identified that a lack of health professional recommendations is a major barrier to self-management app use primarily due to the preparedness and capacity of such professionals [97]. Thus, such engagement would likely facilitate a better understanding of health professionals' use of digital health apps in practice to support wide-scale implementation and community evaluation in the real world.

Finally, participants identified a range of suggested improvements that could be integrated into future iterations of the app. Some of these improvements involved using additional BCTs, namely, the active ingredients of interventions that are critical for changing behavior or the key mechanisms of action [21]. These included self-monitoring of behavior (ie, tracking against health goals) and providing information about health consequences if health goals were not achieved. Future research should aim to map these suggested improvements onto the taxonomy of BCTs based on relevant health behavior change theories and a theoretical framework known as the Theoretical Domains Framework [98,99], which was developed by behavioral scientists and implementation researchers. Such mapping could help inform which combination of BCTs to include in future versions of the app based on an evaluation of potential effectiveness drawn from the existing literature alongside the existing BCTs used. Specifically, EFT uses several techniques described in the behavior change taxonomy such as goal- and planning-related strategies as well as comparing future goals with existing behaviors [21].

Conclusions

This study used a co-design process to inform the development of the "Luminaut" app, a digital EFT intervention that could potentially improve and maintain health-promoting behaviors and disrupt the psychological phenomenon known as future discounting (ie, valuing smaller, immediate rewards over larger, delayed rewards). Such interventions may lead to an enhanced adoption of healthier lifestyle changes amongst Australian adults who are currently in the prerisk or at-risk phase for chronic conditions. Using co-design methods to refine the design of digital health interventions may improve the relevance, scalability, sustainability, and engagement for end users. Engagement with digital interventions is a priority to ensure long-lasting behavior change and enable healthy ageing of the Australian population. The current project was valuable in gaining consumer perspectives about what makes the "Luminaut" app a relevant and engaging tool to promote adherence to health-related behavior change interventions and, in turn, has the potential to reduce the risk and burden of chronic conditions.

Acknowledgments

The authors wish to thank David Cox for his input and advice during the early stages of the project, and Emily Brindal for her helpful advice on this project. The authors would also like to acknowledge Appliquette for developing the Luminaut smartphone app.

Funding

This work was supported by funding from the Commonwealth Scientific and Industrial Research Organisation Precision Health Future Science Platform.

Data Availability

The data that support the findings of this study are available from the corresponding author (NH) upon reasonable request and subject to relevant human research ethics approval.

Authors' Contributions

NH, CAH, MR, KL, IG, AD, and PGB facilitated or contributed to the co-design workshops. AD, KL, and MR conducted data analysis. NH, CAH, PGB, MR, AD, and KL contributed to writing the first draft of the manuscript. All authors have contributed to, and approved, the final version of the manuscript.

Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. Some of the authors (NH, CH, MAR, and PGB) were involved in the development of the app they have been evaluating.

XSL•FO
RenderX

Multimedia Appendix 1
Description of the app naming competition.
[DOCX File, 16 KB - humanfactors_v13i1e74099_app1.docx]

Multimedia Appendix 2
Description of the workshops.
[DOCX File, 33 KB - humanfactors_v13i1e74099_app2.docx]

Multimedia Appendix 3
Project data report.
[DOCX File, 16533 KB - humanfactors_v13i1e74099_app3.docx]

Multimedia Appendix 4
Characteristics of the enrolled participants according to workshop attendance.
[DOCX File, 18 KB - humanfactors_v13i1e74099_app4.docx]

Checklist 1
GRIPP2-SF checklist.
[DOCX File, 17 KB - humanfactors_v13i1e74099_app5.docx]

References

1. Noncommunicable diseases. : World Health Organization; 2024.
2. Noncommunicable diseases country profiles 2018. : World Health Organization; 2018.
3. Budreviciute A, Damiati S, Sabir DK, et al. Management and prevention strategies for non-communicable diseases (NCDs) and their risk factors. Front Public Health 2020;8:574111. [doi: 10.3389/fpubh.2020.574111] [Medline: 33324597]
4. Anderson E, Durstine JL. Physical activity, exercise, and chronic diseases: a brief review. Sports Med Health Sci 2019 Dec;1(1):3-10. [doi: 10.1016/j.smhs.2019.08.006] [Medline: 35782456]
5. Afshin A, Sur PJ, Fay KA, et al. Health effects of dietary risks in 195 countries, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2019 May;393(10184):1958-1972. [doi: 10.1016/S0140-6736(19)30041-8]
6. Hu FB. Diet strategies for promoting healthy aging and longevity: an epidemiological perspective. J Intern Med 2024 Apr;295(4):508-531. [doi: 10.1111/joim.13728] [Medline: 37867396]
7. Middleton KR, Anton SD, Perri MG. Long-term adherence to health behavior change. Am J Lifestyle Med 2013;7(6):395-404. [doi: 10.1177/1559827613488867] [Medline: 27547170]
8. National health survey. Australian Bureau of Statistics. 2023. URL: https://www.abs.gov.au/statistics/health/health-conditions-and-risks/national-health-survey/latest-release [accessed 2026-04-20]
9. Madden GJ, Bickel WK. Impulsivity: The Behavioral and Neurological Science of Discounting: American Psychological Association; 2010. [doi: 10.1037/12069-000]
10. Lin H, Epstein LH. Living in the moment: effects of time perspective and emotional valence of episodic thinking on delay discounting. Behav Neurosci 2014 Feb;128(1):12-19. [doi: 10.1037/a0035705] [Medline: 24512061]
11. Hershfield HE. Future self-continuity: how conceptions of the future self transform intertemporal choice. Ann N Y Acad Sci 2011 Oct;1235(1):30-43. [doi: 10.1111/j.1749-6632.2011.06201.x] [Medline: 22023566]
12. Kakoschke N, Cox DN, Ryan J, et al. Disrupting future discounting: a commentary on an underutilised psychological approach for improving adherence to diet and physical activity interventions. Public Health Nutr 2023 May;26(5):1088-1093. [doi: 10.1017/S136898002200252X] [Medline: 36786324]
13. Bickel WK, Jarmolowicz DP, Mueller ET, Koffarnus MN, Gatchalian KM. Excessive discounting of delayed reinforcers as a trans-disease process contributing to addiction and other disease-related vulnerabilities: emerging evidence. Pharmacol Ther 2012 Jun;134(3):287-297. [doi: 10.1016/j.pharmthera.2012.02.004]
14. Amlung M, Petker T, Jackson J, Balodis I, MacKillop J. Steep discounting of delayed monetary and food rewards in obesity: a meta-analysis. Psychol Med 2016 Aug;46(11):2423-2434. [doi: 10.1017/S0033291716000866] [Medline: 27299672]
15. Kakoschke N, Hawker C, Castine B, de Courten B, Verdejo-Garcia A. Smartphone-based cognitive bias modification training improves healthy food choice in obesity: a pilot study. Eur Eat Disord Rev 2018 Sep;26(5):526-532. [doi: 10.1002/erv.2622] [Medline: 30003634]
16. Atance CM, O'Neill DK. Episodic future thinking. Trends Cogn Sci (Regul Ed) 2001 Dec;5(12):533-539. [doi: 10.1016/S1364-6613(00)01804-0]
17. Peters J, Büchel C. Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal-mediotemporal interactions. Neuron 2010 Apr 15;66(1):138-148. [doi: 10.1016/j.neuron.2010.03.026] [Medline: 20399735]

XSL•FO
RenderX

18. Szpunar KK, Spreng RN, Schacter DL. A taxonomy of prospection: introducing an organizational framework for future-oriented cognition. Proc Natl Acad Sci U S A 2014 Dec 30;111(52):18414-18421. [doi: 10.1073/pnas.1417144111] [Medline: 25416592]
19. Boyer P. Evolutionary economics of mental time travel? Trends Cogn Sci (Regul Ed) 2008 Jun;12(6):219-224. [doi: 10.1016/j.tics.2008.03.003]
20. Sofis MJ, Lemley SM, Jacobson NC, Budney AJ. Initial evaluation of domain-specific episodic future thinking on delay discounting and cannabis use. Exp Clin Psychopharmacol 2022 Dec;30(6):918-927. [doi: 10.1037/pha0000501] [Medline: 34096759]
21. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
22. Marques MM, Wright AJ, Corker E, et al. The Behaviour Change Technique Ontology: Transforming the Behaviour Change Technique Taxonomy v1. Wellcome Open Res 2023;8:308. [doi: 10.12688/wellcomeopenres.19363.2] [Medline: 37593567]
23. Michie S, Thomas J, Johnston M, et al. The Human Behaviour-Change Project: harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation. Implement Sci 2017 Oct 18;12(1):121. [doi: 10.1186/s13012-017-0641-5] [Medline: 29047393]
24. Nielsen L, Riddle M, King JW, et al. The NIH Science of Behavior Change Program: transforming the science through a focus on mechanisms of change. Behav Res Ther 2018 Feb;101:3-11. [doi: 10.1016/j.brat.2017.07.002] [Medline: 29110885]
25. Schacter DL, Benoit RG, Szpunar KK. Episodic future thinking: mechanisms and functions. Curr Opin Behav Sci 2017 Oct;17:41-50. [doi: 10.1016/j.cobeha.2017.06.002] [Medline: 29130061]
26. Ward AM. A critical evaluation of the validity of episodic future thinking: a clinical neuropsychology perspective. Neuropsychology 2016 Nov;30(8):887-905. [doi: 10.1037/neu0000274] [Medline: 26901169]
27. Schenk PM, Wright AJ, West R, et al. An ontology of mechanisms of action in behaviour change interventions. Wellcome Open Res 2023;8:337. [doi: 10.12688/wellcomeopenres.19489.2] [Medline: 38481854]
28. Tulving E. What Is episodic memory? Curr Dir Psychol Sci 1993 Jun;2(3):67-70. [doi: 10.1111/1467-8721.ep10770899]
29. Szpunar KK. Episodic future thought: an emerging concept. Perspect Psychol Sci 2010 Mar;5(2):142-162. [doi: 10.1177/1745691610362350] [Medline: 26162121]
30. Suddendorf T, Corballis MC. Mental time travel and the evolution of the human mind. Genet Soc Gen Psychol Monogr 1997 May;123(2):133-167. [Medline: 9204544]
31. Brown JM, Stein JS. Putting prospection into practice: methodological considerations in the use of episodic future thinking to reduce delay discounting and maladaptive health behaviors. Front Public Health 2022;10:1020171. [doi: 10.3389/fpubh.2022.1020171] [Medline: 36408004]
32. Ye JY, Ding QY, Cui JF, et al. A meta-analysis of the effects of episodic future thinking on delay discounting. Q J Exp Psychol (Hove) 2022 Oct;75(10):1876-1891. [doi: 10.1177/17470218211066282]
33. Rösch SA, Stramaccia DF, Benoit RG. Promoting farsighted decisions via episodic future thinking: a meta-analysis. J Exp Psychol Gen 2022 Jul;151(7):1606-1635. [doi: 10.1037/xge0001148] [Medline: 34843367]
34. Colton E, Connors M, Mahlberg J, Verdejo-Garcia A. Episodic future thinking improves intertemporal choice and food choice in individuals with higher weight: a systematic review and meta-analysis. Obes Rev 2024 Oct;25(10):e13801. [doi: 10.1111/obr.13801] [Medline: 39095999]
35. Benoit RG, Gilbert SJ, Burgess PW. A neural mechanism mediating the impact of episodic prospection on farsighted decisions. J Neurosci 2011 May 4;31(18):6771-6779. [doi: 10.1523/JNEUROSCI.6559-10.2011] [Medline: 21543607]
36. Peters J, Büchel C. The neural mechanisms of inter-temporal decision-making: understanding variability. Trends Cogn Sci (Regul Ed) 2011 May;15(5):227-239. [doi: 10.1016/j.tics.2011.03.002]
37. Rung JM, Madden GJ. Demand characteristics in episodic future thinking II: the role of cues and cue content in changing delay discounting. Exp Clin Psychopharmacol 2019 Oct;27(5):482-495. [doi: 10.1037/pha0000260] [Medline: 30762382]
38. Stein JS, Wilson AG, Koffarnus MN, Daniel TO, Epstein LH, Bickel WK. Unstuck in time: episodic future thinking reduces delay discounting and cigarette smoking. Psychopharmacology (Berl) 2016 Oct;233(21-22):3771-3778. [doi: 10.1007/s00213-016-4410-y] [Medline: 27553824]
39. Daniel TO, Stanton CM, Epstein LH. The future is now: reducing impulsivity and energy intake using episodic future thinking. Psychol Sci 2013 Nov 1;24(11):2339-2342. [doi: 10.1177/0956797613488780] [Medline: 24022653]
40. Daniel TO, Said M, Stanton CM, Epstein LH. Episodic future thinking reduces delay discounting and energy intake in children. Eat Behav 2015 Aug;18:20-24. [doi: 10.1016/j.eatbeh.2015.03.006] [Medline: 25863227]
41. Snider SE, LaConte SM, Bickel WK. Episodic future thinking: expansion of the temporal window in individuals with alcohol dependence. Alcohol Clin Exp Res 2016 Jul;40(7):1558-1566. [doi: 10.1111/acer.13112] [Medline: 27246691]
42. Sze YY, Daniel TO, Kilanowski CK, Collins RL, Epstein LH. Web-based and mobile delivery of an episodic future thinking intervention for overweight and obese families: a feasibility study. JMIR Mhealth Uhealth 2015 Dec 16;3(4):e97. [doi: 10.2196/mhealth.4603] [Medline: 26678959]
43. O'Neill J, Daniel TO, Epstein LH. Episodic future thinking reduces eating in a food court. Eat Behav 2016 Jan;20:9-13. [doi: 10.1016/j.eatbeh.2015.10.002] [Medline: 26562686]

XSL•FO
RenderX

44. Leahey TM, Gorin AA, Wyckoff E, et al. Episodic future thinking, delay discounting, and exercise during weight loss maintenance: the PACE trial. Health Psychol 2020 Sep;39(9):796-805. [doi: 10.1037/hea0000860] [Medline: 32833481]
45. Mertens ECA, Siezenga AM, Tettero T, van Gelder JL. A future orientation intervention delivered through a smartphone application and virtual reality: study protocol for a randomized controlled trial. BMC Psychol 2022 Dec 20;10(1):315. [doi: 10.1186/s40359-022-01025-x] [Medline: 36539846]
46. Mertens ECA, Siezenga AM, van der Schalk J, van Gelder JL. A novel smartphone-based intervention aimed at increasing future orientation via the future self: a pilot randomized controlled trial of a prototype application. Prev Sci 2024 Feb;25(2):392-405. [doi: 10.1007/s11121-023-01609-y] [Medline: 37976010]
47. Persson DR, et al. Episodic future thinking as digital micro-interventions. 2023 Presented at: 25th International Conference on Human-Computer Interaction; Jul 23-28, 2023; Copenhagen, Denmark p. 213-237. [doi: 10.1007/978-3-031-48038-6_14]
48. Persson DR, Bardram JE, Bækgaard P. Perceptions and effectiveness of episodic future thinking as digital micro-interventions based on mobile health technology. Digit Health 2024;10:20552076241245583. [doi: 10.1177/20552076241245583] [Medline: 38577315]
49. Mrklas KJ, Barber T, Campbell-Scherer D, et al. Co-design in the development of a mobile health app for the management of knee osteoarthritis by patients and physicians: qualitative study. JMIR Mhealth Uhealth 2020 Jul 10;8(7):e17893. [doi: 10.2196/17893] [Medline: 32673245]
50. Boyd H, McKernon S, Mullin B, Old A. Improving healthcare through the use of co-design. N Z Med J 2012 Jun 29;125(1357):76-87. [Medline: 22854362]
51. Davis A, Tuckey M, Gwilt I, Wallace N. Understanding co‐design practice as a process of “welldoing”. Int J Art Design Ed 2023 May;42(2):278-293. [doi: 10.1111/jade.12459]
52. Giacomin J. What is human centred design? Design J 2014 Dec;17(4):606-623. [doi: 10.2752/175630614X14056185480186]
53. Glasziou P, Chalmers I. Paul glasziou and iain chalmers: is 85% of health research really “wasted”? BMJ 2016 [FREE Full text] [doi: 10.1136/bmj.k4645]
54. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet 2009 Jul 4;374(9683):86-89. [doi: 10.1016/S0140-6736(09)60329-9] [Medline: 19525005]
55. Scariot CA, Heemann A, Padovani S. Understanding the collaborative-participatory design. Work 2012 Mar;41(S1):2701-2705. [doi: 10.3233/WOR-2012-0656-2701]
56. Frank P, Young WR. Patient and participant engagement and involvement in the co-design of digital health interventions, in remote monitoring and wearable devices in healthcare. In: Remote Monitoring and Wearable Devices in Healthcare: Springer; 2025:123-146. [doi: 10.1007/978-3-031-98897-4_7]
57. Burke LE, Ma J, Azar KMJ, et al. Current science on consumer use of mobile health for cardiovascular disease prevention: a scientific statement from the American Heart Association. Circulation 2015 Sep 22;132(12):1157-1213. [doi: 10.1161/CIR.0000000000000232] [Medline: 26271892]
58. Malloy JA, Partridge SR, Kemper JA, Braakhuis A, Roy R. Co-design of digital health interventions for young adults: protocol for a scoping review. JMIR Res Protoc 2022 Oct 24;11(10):e38635. [doi: 10.2196/38635] [Medline: 36279167]
59. Winters N, Oliver M, Langer L. Can mobile health training meet the challenge of ‘measuring better’? In: Unterhalter E, editor. Measuring the Unmeasurable in Education: Routledge; 2020:115-131. [doi: 10.4324/9780429444494-6]
60. Ferguson C, Hickman LD, Turkmani S, Breen P, Gargiulo G, Inglis SC. “Wearables only work on patients that wear them”: barriers and facilitators to the adoption of wearable cardiac monitoring technologies. Cardiovasc Digit Health J 2021 Apr;2(2):137-147. [doi: 10.1016/j.cvdhj.2021.02.001] [Medline: 35265900]
61. Tzimourta KD. Human-centered design and development in digital health: approaches, challenges, and emerging trends. Cureus 2025 Jun;17(6):e85897. [doi: 10.7759/cureus.85897] [Medline: 40548156]
62. Carrera Diaz K, Yau J, Iverson E, et al. Human-centered design approach to building a transition readiness mHealth intervention for early adolescents. J Pediatr Psychol 2025 Jan 1;50(1):106-114. [doi: 10.1093/jpepsy/jsae066] [Medline: 39172486]
63. An Q, Kelley MM, Hanners A, Yen PY. Sustainable development for mobile health apps using the human-centered design process. JMIR Form Res 2023 Aug 25;7:e45694. [doi: 10.2196/45694] [Medline: 37624639]
64. Hanrahan M, Wilson C, Keogh A, et al. How can patients shape digital medicine? A rapid review of patient and public involvement and engagement in the development of digital health technologies for neurological conditions. Expert Rev Pharmacoecon Outcomes Res 2025 Feb 7;25(2):137-154. [doi: 10.1080/14737167.2024.2410245]
65. Kilfoy A, Hsu TCC, Stockton-Powdrell C, Whelan P, Chu CH, Jibb L. An umbrella review on how digital health intervention co-design is conducted and described. NPJ Digit Med 2024 Dec 23;7(1):374. [doi: 10.1038/s41746-024-01385-1] [Medline: 39715947]
66. Staniszewska S, Brett J, Simera I, et al. GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research. BMJ :j3453. [doi: 10.1136/bmj.j3453]
67. Innovation by design: how design enables science and technology research to achieve greater impact. : British Design Council; 2015.
68. Sanders EBN, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008 Mar;4(1):5-18. [doi: 10.1080/15710880701875068]

XSL•FO
RenderX

69. Gero J, Milovanovic J. The situated function-behavior-structure co-design model. CoDesign 2021 Apr 3;17(2):211-236. [doi: 10.1080/15710882.2019.1654524]
70. Rashmi R, Mohanty SK. Examining chronic disease onset across varying age groups of Indian adults using competing risk analysis. Sci Rep 2023 Apr 10;13(1):5848. [doi: 10.1038/s41598-023-32861-5] [Medline: 37037884]
71. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125. [doi: 10.2196/jmir.5604] [Medline: 28428170]
72. Leask CF, Sandlund M, Skelton DA, et al. Framework, principles and recommendations for utilising participatory methodologies in the co-creation and evaluation of public health interventions. Res Involv Engagem 2019;5:2. [doi: 10.1186/s40900-018-0136-9] [Medline: 30652027]
73. Guest G, Bunce A, Johnson L. How many interviews are enough? An experiment with data saturation and variability. Field methods 2006;18(1):59-82. [doi: 10.1177/1525822X05279903]
74. Sanders E, Stappers PJ. Convivial Toolbox: Generative Research for the Front End of Design: BIS Publishers; 2012.
75. Pettit CJ, Glackin S, Trubka R, et al. A co-design prototyping approach for building a Precinct Planning Tool. ISPRS Ann Photogramm Remote Sens Spatial Inf Sci 2014;II-2:47-53. [doi: 10.5194/isprsannals-II-2-47-2014]
76. Socio-economic indexes for areas (SEIFA). Australian Bureau of Statistics. 2021. URL: https://www.abs.gov.au/statistics/people/people-and-communities/socio-economic-indexes-areas-seifa-australia/latest-release [accessed 2026-03-03]
77. Braun V, Clarke V, Thematic analysis. APA handbook of research methods in psychology. In: Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D, Sher KJ, editors. Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological Ed: American Psychological Association; 2012, Vol. 2. [doi: 10.1037/13620-004]
78. Braun V, Clarke V. One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qual Res Psychol 2021 Jul 3;18(3):328-352. [doi: 10.1080/14780887.2020.1769238]
79. Pienkowska A, Ang CS, Mammadova M, Mahadzir MDA, Car J. A diabetes education app for people living with type 2 diabetes: co-design study. JMIR Form Res 2023 Sep 18;7(1):e45490. [doi: 10.2196/45490] [Medline: 37721799]
80. Tay BSJ, Edney SM, Brinkworth GD, et al. Co-design of a digital dietary intervention for adults at risk of type 2 diabetes. BMC Public Health 2021 Nov 11;21(1):2071. [doi: 10.1186/s12889-021-12102-y] [Medline: 34763701]
81. Asbjørnsen RA, Hjelmesæth J, Smedsrød ML, et al. Combining persuasive system design principles and behavior change techniques in digital interventions supporting long-term weight loss maintenance: design and development of eCHANGE. JMIR Hum Factors 2022 May 27;9(2):e37372. [doi: 10.2196/37372] [Medline: 35622394]
82. Gibson I, Neubeck L, Corcoran M, et al. Development of a digital health intervention for the secondary prevention of cardiovascular disease (INTERCEPT): co-design and usability testing study. JMIR Hum Factors 2024 Oct 23;11:e63707. [doi: 10.2196/63707] [Medline: 39441626]
83. Asbjørnsen RA, Wentzel J, Smedsrød ML, et al. Identifying persuasive design principles and behavior change techniques supporting end user values and needs in eHealth interventions for long-term weight loss maintenance: qualitative study. J Med Internet Res 2020 Nov 30;22(11):e22598. [doi: 10.2196/22598] [Medline: 33252347]
84. Alahäivälä T, Oinas-Kukkonen H. Understanding persuasion contexts in health gamification: a systematic analysis of gamified health behavior change support systems literature. Int J Med Inform 2016 Dec;96:62-70. [doi: 10.1016/j.ijmedinf.2016.02.006] [Medline: 26944611]
85. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Jul;71:31-48. [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
86. Sauro J, Lewis JR. When designing usability questionnaires, does it hurt to be positive? : Association for Computing Machinery; 2011 May 7 Presented at: CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver BC Canada p. 2215-2224. [doi: 10.1145/1978942.1979266]
87. Agapie E, Haldar S, Poblete SG. Using HCI in cross-disciplinary teams: a case study of academic collaboration in HCI-Health teams in the US using a team science perspective. Proc ACM Hum-Comput Interact 2022 Nov 7;6(CSCW2):1-35. [doi: 10.1145/3555610]
88. Waddell A, Seguin JP, Wu L, et al. Leveraging implementation science in human-centred design for digital health. : Association for Computing Machinery; 2024 May 11 Presented at: CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems; May 11-16, 2024; Honolulu HI USA. [doi: 10.1145/3613904.3642161]
89. Agius SJ. Qualitative research: its value and applicability. Psychiatrist 2013 Jun;37(6):204-206. [doi: 10.1192/pb.bp.113.042770]
90. Kontos E, Blake KD, Chou WYS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172. [doi: 10.2196/jmir.3117] [Medline: 25048379]
91. Arias López MDP, Ong BA, Borrat Frigola X, et al. Digital literacy as a new determinant of health: a scoping review. PLOS Digit Health 2023;2(10):e0000279. [doi: 10.1371/journal.pdig.0000279]
92. Davis A, Gwilt I, Wallace N, Langley J. Low-contact co-design: considering more flexible spatiotemporal models for the co-design workshop. Strategic Design Res J 2021;14(1):124-137. [doi: 10.4013/sdrj.2021.141.11]

XSL•FO
RenderX

93. Marwaha JS, Landman AB, Brat GA, Dunn T, Gordon WJ. Deploying digital health tools within large, complex health systems: key considerations for adoption and implementation. NPJ Digit Med 2022 Jan 27;5(1):13. [doi: 10.1038/s41746-022-00557-1] [Medline: 35087160]
94. Murray E, Hekler EB, Andersson G, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851. [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
95. Slattery P, Saeri AK, Bragge P. Research co-design in health: a rapid overview of reviews. Health Res Policy Sys 2020 Dec;18(1):17. [doi: 10.1186/s12961-020-0528-9]
96. Ioannidis JPA. Why most clinical research is not useful. PLoS Med 2016 Jun;13(6):e1002049. [doi: 10.1371/journal.pmed.1002049] [Medline: 27328301]
97. Bradway M, Morris RL, Giordanengo A, Årsand E. How mHealth can facilitate collaboration in diabetes care: qualitative analysis of co-design workshops. BMC Health Serv Res 2020 Nov 30;20(1):1104. [doi: 10.1186/s12913-020-05955-3] [Medline: 33256732]
98. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implementation Sci 2012;7(1):1-17. [doi: 10.1186/1748-5908-7-37]
99. Atkins L, Francis J, Islam R, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. Implementation Sci 2017 Dec;12(1):1-18. [doi: 10.1186/s13012-017-0605-9]

Abbreviations

BCT: behavior change technique
EFT: Episodic Future Thinking
GRIPP2-SF: Guidance for Reporting Involvement of Patients and Public Short Form
IRSAD: Index of Relative Socio-economic Advantage and Disadvantage
LGA: Local Government Areas
mHealth: mobile health
NCD: noncommunicable disease
REDCap: Research Electronic Data Capture

Edited by A Kushniruk; submitted 17.Apr.2025; peer-reviewed by E Mertens, J Scott; revised version received 11.Mar.2026; accepted 25.Mar.2026; published 06.May.2026.

Please cite as:
Hoffmann N, Howlett CA, Little K, Gwilt I, Rebuli MA, Brooker PG, Davis A
Digital Episodic Future Thinking Intervention (Luminaut): Co-Design and Iterative Development Study
JMIR Hum Factors 2026;13:e74099
URL: https://humanfactors.jmir.org/2026/1/e74099
doi:10.2196/74099

© Naomi Kakoschke, Caitlin A Howlett, Kate Little, Ian Gwilt, Megan Rebuli, Paige G Brooker, Aaron Davis. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 6.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Young People's Perceptions of Signposting in a Digital Mental Health Helpline: Mixed Methods Analysis of Cross-Sectional Data

Angel Yee Fong Au, MSci; Anya Jan, MSc; Shyn Wei Phua, MSci; Leslie Morrison Gutman, PhD

Faculty of Brain Sciences, University College London, 1-19 Torrington Place, London, United Kingdom

Corresponding Author:
Leslie Morrison Gutman, PhD
Faculty of Brain Sciences, University College London, 1-19 Torrington Place, London, United Kingdom

Abstract

Background: Mental health problems are prevalent among young people aged 16 to 24 years. With the health care system prioritizing severe cases, most young people wait months before accessing professional support. One-to-one helplines offer alternative and accessible mental health services for young people with emotional support, psychoeducation, and signposting. Signposting empowers young people to access long-term support beyond a brief helpline session. However, young people often choose not to access the signposts. Despite its importance, there is a dearth of existing research examining signposting via digital mental health helplines for young people.

Objective: Using cross-sectional survey data from The Mix, a UK charity supporting young people aged 25 years or younger, this study conducted a mixed methods analysis of their multichannel (webchat, email, telephone, and web-based contact form) helpline survey between February 2020 and October 2023.

Methods: The analytic sample included 296 participants who collectively received 872 signposts (approximately 872/4500, 19% of signposts provided during the survey collection period), of which 822 with complete outcome data were included in the statistical models. Multinomial logistic regressions were conducted to examine whether young people's use and perceived usefulness of the signposts they received differed across modes of delivery and their demographic characteristics (gender, ethnicity, and age). Qualitative thematic analysis of 106 open-ended responses from 97 participants was also examined to illuminate why young people found signposting helpful and how it could be improved.

Results: In the overall model, which included all predictors, webchat users identifying as White, women, and aged 16 - 19 years were significantly more likely to use and find signposts helpful than to perceive them as unhelpful (odds ratios [OR] 0.28, 95% CI 0.17-0.46; *P*<.001), not intend to use them (OR 0.13, 95% CI 0.07-0.26; *P*<.001), or only plan to use them later (OR 0.29, 95% CI 0.18-0.46; *P*<.001). Thematic analysis of open-ended responses revealed that young people found the choice of signposts relevant and appreciated how signposting was integrated with emotional support. Young people also felt more hopeful after being signposted and gained both clarity and insight into the support available. However, they also noted challenges, such as feeling overwhelmed or encountering outdated signposts.

Conclusions: Given the increasing reliance on digital mental health services, ensuring that signposting remains accessible, relevant, and tailored to diverse user needs is essential. By optimizing signposting strategies, helplines can empower young people to seek appropriate long-term support, ultimately improving mental health outcomes.

(JMIR Hum Factors 2026;13:e73369) doi:10.2196/73369

KEYWORDS

internet; online; services; adolescents; youth; qualitative; survey; crisis; web-based; telephone

Introduction

Mental Health Support for Young People

Mental health problems explain 13% of the global burden of disease for 10- to 19-year-olds and are experienced by 14% of adolescents [1]. In the United Kingdom, 22% of young people aged 8 to 25 years experience a probable mental health problem [2]. While young people experiencing severe conditions, like psychosis or eating disorders, were seen within 2 to 4 weeks of referral [3], half of those with mild-moderate mental health problems waited 5 to 18 weeks for support [4]. Due to this delay, young people perceive care to be inaccessible and unresponsive, and therefore, can be discouraged from seeking help and often rereferred with more severe conditions [5,6]. Many times, even when support is offered after long waits, young people are more likely to refuse it or drop out [7].

Digital mental health interventions such as one-to-one helplines, peer support, and self-help resources are alternative sources of mental health support for young people [8]. Available across multiple modes of delivery (eg, text, email, phone calls, and

XSL•FO
RenderX

webchat), one-to-one helplines offer brief, confidential support alongside signposting to specialist services [9,10]. Signposting encourages individuals to actively access specialist services, thus empowering them to build long-term support networks and develop healthy coping strategies beyond the helpline sessions [11]. However, many young people do not follow up on signposting [12]. Furthermore, there is a dearth of research examining young people's experiences of signposting via digital mental health helplines, especially across various modes of delivery and in relation to young people's demographic characteristics, information that is essential for improving the uptake and tailoring the usefulness of this service to different groups.

To help address this gap, this study examines signposting data from The Mix's helpline, the United Kingdom's leading digital mental health support platform for under-25s, to analyze differences in young people's signposting experiences according to the mode of delivery and young people's gender, ethnicity, and age. Using mixed methods, we combined quantitative findings with qualitative insights to inform recommendations to improve signposting via digital mental health helplines for young people.

Research Background

Signposting involves a practitioner sharing resources and encouraging individuals to actively access them [12]. It requires young people to acknowledge their need for support, identify which services to access, and feel confident that those services are helpful [13]. When young people seek support, many are unaware of the services available to them [14]. This lack of awareness can act as a barrier to accessing timely and relevant support. Even when young people are aware of the services available, only a fraction actively engage and access them. To unpack how signposting may support young people beyond the immediate helpline contact, we situate this study within complementary perspectives on help-seeking and engagement. Help-seeking models (eg, Andersen's behavioral model) conceptualize service use as shaped by (1) predisposing factors (eg, demographics and beliefs), (2) enabling factors (eg, access, knowledge of services, and perceived fit), and (3) perceived need. In this framing, signposting may function as a bridge between "in-the-moment" support and longer-term care by reducing informational barriers, increasing perceived options, and supporting readiness to act. In addition, frameworks for engagement in digital and brief interventions emphasize that uptake is more likely when support is experienced as relationally safe, autonomy-supportive, and tailored to the user's context (eg, clear rationale, collaborative choice, and practical next steps). Together, these perspectives support interpreting signposting outcomes as a function of both service design (eg, mode of delivery and the interactional context) and user context (eg, disclosure, identity safety, and perceived accessibility of resources). For instance, Mustafa et al [12] found that while 79% (26/33) of health students knew about the available mental health services, only 27% (9/33) accessed them, highlighting the critical role practitioners play in bridging the gap, not only by sharing information, but also by motivating young people to engage with support services. Poor signposting and repeated referrals [15] can also push young people between irrelevant services, reducing their confidence in getting support.

These frameworks directly informed this study. Andersen's model guided our focus on predisposing factors (gender, ethnicity, and age) and enabling factors (mode of delivery) as predictors of signposting outcomes, while engagement frameworks motivated our expectation that relational and interactional features of each mode would shape whether signposting felt relevant and actionable. Together, they justify a mixed methods approach—quantitative analysis to identify which user and service characteristics predict signposting outcomes, and qualitative analysis to illuminate the mechanisms that explain those patterns.

Research examining signposting has mainly focused on primary care and school settings. For example, previous qualitative studies have examined the barriers to signposting for primary care professionals in the United Kingdom, which included a lack of formal training, decreased motivation to signpost, and time constraints [16,17]. In terms of young people's mental health, Tegethoff and colleagues [18] found that using school mental health services for signposting significantly predicted the use of external mental health services, showing the importance of schools in raising awareness of available resources. However, there is a dearth of studies examining signposting for young people's mental health in the United Kingdom. In 1 notable exception, a qualitative study found that volunteers felt they lacked skills in person-centered signposting but benefited from the extensive signposting database [19]. This study, however, only focused on volunteers of a mental health helpline for young people delivered via email.

More research is needed to examine young people's signposting experiences for their mental health, especially across different modes of delivery. Helplines can be delivered asynchronously (eg, email and web-based contact form) or in real-time (eg, webchat and phone). Young people prefer to have control over how they access support, and offering a multichannel helpline can encourage help-seeking [20]. Helplines via email or contact form encourage young people to explain their situation without time constraints, which is particularly effective for those experiencing social anxiety [21]. Webchats empower users to share their concerns via text messages and have live conversations with helpline agents. Similarly, phone calls allow users to express themselves synchronously with their voice. Despite these differences in user preferences, there is little or no information on whether young people's use and their perceptions of signposts differ across modes of delivery. Such information would provide a better understanding of whether signposting is more or less well-received and used depending on how it is delivered.

There is also no current research examining whether young people's experiences of signposting differ across their demographic characteristics. Nevertheless, past studies have found differences in young people's use of helplines across demographic groups. Most helpline users tend to be White, women [22], and older adolescents [23], with men [24] less likely to access helplines. Due to the unavailability of culturally specific services [25], South Asian and Black Caribbean

XSL•FO
RenderX

adolescents are also less likely to seek help as they worry about confidentiality [22,26] and parent surveillance or disapproval [27]. It is increasingly important to examine diversity in young people's experiences with helplines, especially as gender and ethnic minorities [28] are not usually reported in helpline evaluations [23].

This Study

Using secondary cross-sectional survey data from The Mix regarding their multichannel helpline (ie, webchat, email, web-based contact form, and phone), quantitative analysis was conducted to examine whether young people's use and perceived usefulness of the signposts they received differed across modes of delivery and their demographic variables (gender, ethnicity, and age), reflecting the predisposing and enabling factors identified in help-seeking frameworks. Qualitative thematic analysis of open-ended responses was also examined to illuminate why young people found signposting helpful and how it could be improved.

Methods

Setting

The Mix's one-to-one, anonymous helpline provides emotional support and information via email, webchat, telephone, and web-based contact form. It is delivered by volunteers who have received a 25-hour training. Volunteers signpost young people based on The Mix's signposting database "The Red Book," which compiles national and local services for diverse problem areas. Following the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist, we note this was a closed, postcontact web survey administered to recent users of The Mix's multichannel helpline (webchat, email, telephone, and web contact form) between February 2020 and October 2023.

Participants

From February 2020 to October 2023, The Mix emailed helpline users a survey link. Participation was voluntary. During this time, The Mix provided approximately 4500 signposts to users. Of the 807 participants over the age of 16 who responded to The Mix's survey, we excluded participants with missing signposting data (n=4), those who did not complete any demographic questions (n=58), those who were not signposted (n=387), and those who were signposted but did not report any outcomes (n=63), totaling 511 excluded participants. This resulted in an analytical sample of 296 participants who collectively received 872 signposts, representing approximately 19% of the signposts provided during the survey collection period. Of these, 49 participants only partially reported signposting outcomes, creating a sample of 822 signposts included in the statistical models. The characteristics of the analytical sample are presented in Table 1.

Fisher exact test revealed significant differences in mode of delivery (P<.001) between the analytical sample (n=296) and those who were not signposted (n=387), with the analytical sample showing a higher proportion of phone and contact form users, but a lower proportion of email and webchat users. Compared with those who were signposted but did not report any outcomes (n=63), the analytical sample (n=296) differed significantly according to mode of delivery (P<.001), gender (P<.001), ethnicity (P<.001), and age (P=.001). Specifically, the analytical sample had a higher proportion of users aged 16 - 18 and 19 - 21, users identifying as White, Asian, or Mixed ethnicity, and women or other genders.

Fisher tests also revealed no demographic differences between participants who provided qualitative responses and those who did not in the signposted group (mode of delivery: P=.98, gender: P=.98, ethnicity: P=.99, and age: P=.99).

XSL•FO
RenderX

Table . Demographic characteristics of the final sample (n=296).

Characteristics	n (%)
Gender	
Women	176 (59)
Men	45 (15)
Nonbinary	3 (1)
Transgender	2 (1)
Other	9 (3)
Prefer not to say	12 (4)
Missing	49 (17)
Age (y)	
16 - 19	101 (34)
20 - 21	85 (29)
22 - 25	55 (19)
Older than 26	9 (3)
Prefer not to say	2 (1)
Missing	44 (15)
Ethnicity	
White	189 (64)
Asian	23 (8)
Black	10 (3)
Mixed	12 (4)
Other	1 (0.3)
Prefer not to say	7 (2)
Missing	54 (18)
Mode of delivery	
Webchat	127 (43)
Contact form	96 (32)
Email	54 (18)
Phone	19 (6)

Procedure

From February 2020 to October 2023, all helpline users aged 16 to older than 25 years who contacted the helpline were asked to complete their user survey via email. After clicking the survey link, participants read an information sheet explaining the research aims. Informed consent was then obtained online. Only survey responses from those aged 16 or above were included in this study in line with ethical guidelines for informed consent. Upon providing informed consent, participants completed the online survey on SmartSurvey (Appendix S1 in Multimedia Appendix 1).

This study analyzed only the demographic data and questions related to signposting. The survey platform, number of pages or items, use of randomization or skip or branching, presence of completeness checks, and whether answers could be reviewed or changed before submission were not specified.

Measures

Demographic Questions

Modes of delivery included webchat, contact form, email, and phone. Gender options included women, men, nonbinary, transgender, other, and prefer not to say. Consistent with the ethnic groups in the UK Census [29], ethnicity options included 19 ethnic groups. Since at least 10 observations per predictor group are required for multinomial logistic regressions to yield reliable findings [30], ethnicity was collapsed into 6 broad groups—White, Asian, Black, Mixed, Other, and prefer not to say. Age was collapsed into 16 - 19 years (further education and transition into adulthood), 20 - 21 years (higher education and graduation), 22 - 25 years (postgraduate education and employment), older than 26 years (caregivers or professionals), and prefer not to say.

XSL•FO
RenderX

Signposting Outcomes

For the young person's use and their perception of the usefulness of each signpost, as well as their overall signposting experience, the survey included multiple-choice and open-ended questions, outlined in Table 2. In line with CHERRIES, items comprised fixed-response options plus open-ended questions on experience and improvement.

Table . Survey questions on signposting outcomes.

Question types	Questions
Multiple-choice question on the outcome for each signpost	• Did you find the below organizations or resources recommended to you useful? • Yes (Used it and found it useful) • No (Used it and found it not useful) • I haven't used it but I'm planning to do so • I haven't used it and NOT planning to do so • Not relevant
Open-ended questions for each participant	• Please tell us how the service helped you. • How can we improve? • Please let us know if you have any other comments or feedback.

Data Analysis

Quantitative Analysis

As a closed survey, the view rate is not applicable. Participation and completion rates were not reported. Analyses used the analytic samples described below. Addressing each signpost as an independent item, item-wise analysis was carried out to explore whether demographic variables (categorical) predicted young people's use or perceived usefulness of the signposts (categorical). Using the *mblogit* function from the *mclogit* package in R (R Core Team) [31], multinomial logistic regressions were conducted with only the fixed effects of mode of delivery, gender, ethnicity, and age on signposting outcomes, alongside an overall model with fixed effects of all variables. "Yes" (useful) was the reference category. Although this outcome was a repeated measure depending on the number of signposts for each participant, the model could not converge after adding the random intercept of participants. Nonconvergence may reflect the complexity of the multinomial outcome and limited sample sizes in certain categories. As a result, we reported fixed-effects models treating each signpost outcome as an item-level observation. This decision prioritized model stability and interpretability given sparse cells, but may underestimate uncertainty because outcomes are clustered within participants; findings should therefore be interpreted as exploratory associations.

Qualitative Analysis

Inductive, critical realist thematic analysis [32] was conducted. Responses relevant to signposting from the open-ended questions were first shortlisted by filtering responses from participants who were signposted, then searching for keywords (eg, "link," "resource," "signpost," "organisation," and "website"), and reading the remaining responses to include relevant ones without the keywords (eg, "I decided to book an appointment with my GP to get help").

From these shortlisted responses, 18 initial codes were developed and later organized into 4 main themes using NVivo 14 (Lumivero) [33]. Themes emerged from responses reflecting similar signposting experiences or impacts, with smaller subthemes representing recurring concepts. A codebook with 4 themes and 12 subthemes was eventually created (Appendix S2 in Multimedia Appendix 1).

With varying length and depth of responses, some were coded to more than 1 subtheme. For intercoder reliability, 25 responses were independently coded. The percentage of agreement (ie, proportion of items assigned to the same codes by both researchers) was 76%, indicating substantial agreement between the coders even when responses were mapped to multiple codes [34].

To compare differences in the qualitative feedback across demographic groups, simple sentiment analysis was also conducted by contextually reading each response and annotating it as expressing either positive or negative sentiment. Positive responses were identified as those expressing satisfaction, perceived benefits, or motivation to engage with signposts (eg, "I found a really helpful support group"). Negative responses were categorized as those highlighting barriers, dissatisfaction, or challenges (eg, "the signposted resources didn't meet my needs"). Responses were contextually read and annotated accordingly. Coding of sentiment was conducted manually, with annotations guided by an agreed-upon framework to ensure consistency.

Ethical Considerations

Ethical approval was granted by the University College London Research Ethics Committee (ethics ID: 16583/003). Participants first viewed an online information sheet and then provided informed consent before answering any items. To protect participant privacy, all survey data were fully anonymized, and no personally identifiable information was collected or retained. As compensation for their time, participants were entered into a prize draw with the chance to win a £50 (US $66.64) Amazon voucher.

XSL•FO
RenderX

Results

Overview

Due to varied levels of data missingness for demographic variables, observations with missing data were removed before analyzing each predictor (ie, participants with missing data for ethnicity were only excluded from analyses on ethnicity).

Quantitative Analysis: Signposting Outcomes

As illustrated in the Appendix S3 in Multimedia Appendix 1, individual models were first conducted for the mode of delivery (n=822), gender (n=684), ethnicity (n=669), or age (n=697) before running an overall model with all predictors included (n=669). The individual models demonstrated patterns broadly consistent with the overall model, with some notable differences. The mode of delivery model highlighted that webchat was the most effective channel for signposting outcomes, with other platforms (contact form, email, and phone) associated with a lower likelihood of participants using and finding signposts helpful. The gender model showed that participants identifying as other genders or those who preferred not to disclose their gender were less likely to find signposts helpful, with a greater likelihood of reporting irrelevance or a delayed intention to use them. The ethnicity model revealed that individuals who preferred not to disclose their ethnicity were significantly more likely to perceive signposts as unhelpful or irrelevant, a trend more pronounced than in the overall model. The age model showed that older participants (20 - 25 y) were less likely to find signposts helpful, with these effects being slightly more potent in the individual model than when controlling for all predictors in the overall model. In the overall model, which included all predictors, webchat users identifying as White, women, and aged 16 - 19 years were significantly more likely to use and find signposts helpful than to perceive them as unhelpful (odds ratio [OR] 0.28, 95% CI 0.17-0.46, P<.001), not intend to use them (OR 0.13, 95% CI 0.07-0.26, P<.001), or only plan to use them later (OR 0.29, 95% CI 0.18-0.46, P<.001). These results demonstrate the importance of both the delivery mode and participant demographics in predicting the perceived effectiveness of signposting.

Of 669 signposts in the overall model, 360 (54%) were given via webchat, 480 (72%) were shared with participants who identified as women, 521 (78%) with White participants, and 350 (52%) with 16- to 19-year-olds. Thus, webchat, women, White, and 16- to 19-year-olds were the reference categories for the multinomial logistic regressions.

A multinomial logistic regression examined whether mode of delivery, gender, ethnicity, or age predicted signposting outcomes. A significant chi-square difference test (F_{64}=181.27, P<.001) indicated that our model significantly improved upon the null. The model explained 43% of the variance (Nagelkerke R^2) with a low Akaike Information Criterion (AIC=1934) and Bayesian Information Criterion (BIC=2240).

Shown in Table 3 as the intercept, webchat users identifying as White, women, and 16- to 19-year-olds were significantly more likely to use the signposts and find them useful than perceiving them as unhelpful (OR 0.28, 95% CI 0.17-0.46; P<.001), not intending to use them (OR 0.13, 95% CI 0.07-0.26; P<.001), or only planning to use them (OR 0.29, 95% CI 0.18-0.46; P<.001). This reference group, therefore, represented the most favorable signposting outcomes in the model.

Compared with webchat users, users accessing other platforms (contact form, email, and phone) were less likely to use and find the signposts helpful. Contact form users were more likely not to plan to use the signposts (OR 4.98, 95% CI 2.29-10.83; P<.001), only plan to use them later (OR 3.95, 95% CI 1.98-7.89; P<.001), or find them irrelevant (OR 2.90, 95% CI 1.60-5.25; P<.001). Email users were more likely to only plan to use the signposts (OR 2.77, 95% CI 1.50-5.13; P=.001). Phone users were even more likely to have used the signposts but find them unhelpful (OR 4.17, 95% CI 1.58-11.00; P=.004).

Compared with women, participants identifying as other genders or preferring not to say their gender were less likely to find signposting useful. Those identifying as other genders were more likely to plan to use signposts later (OR 8.46, 95% CI 2.13-33.66; P=.002) or find them irrelevant (OR 4.49, 95% CI 1.22-16.59; P=.02). Participants who preferred not to disclose their gender were more likely to have no intention to use the signposts (OR 3.82, 95% CI 1.19-12.26; P=.03).

Compared with White participants, those who preferred not to share their ethnicity were more likely to find the signposts unhelpful (OR 45.97, 95% CI 3.56-594.15; P=.003) or irrelevant (OR 16.95, 95% CI 1.50-191.26; P=.02). The wide CIs reflect the small sample size for this group and should be interpreted with caution.

Compared with 16- to 19-year-olds, 20- to 25-year-olds were less likely to find signposts helpful. 20- to 21-year-olds were more likely to only plan to use the signposts (OR 3.13, 95% CI 1.64-5.97; P=.001) or perceive them as irrelevant (OR 1.80, 95% CI 1.01-3.20; P=.045), while 22- to 25-year-olds were more likely to not plan to use them (OR 2.92, 95% CI 1.30-6.53; P=.009) or find them irrelevant (OR 1.80, 95% CI 1.05-3.08; P=.03). This suggests engagement with signposting decreases progressively across 20- to 25-year-olds.

XSL•FO
RenderX

Table . Multinomial logistic regression on signposting outcomes and all predictors (n=669)[a].

Predictors	Useful: no		Useful: not used and no intent		Useful: not used but has intent		Useful: not relevant	
	Odds ratios (95% CI)	*P* value	Odds ratios (95% CI)	*P* value	Odds ratios (95% CI)	*P* value	Odds ratios (95% CI)	*P* value
Intercept	0.28 (0.17 - 0.46)	<.001	0.13 (0.07 - 0.26)	<.001	0.29 (0.18 - 0.46)	<.001	0.77 (0.54 - 1.10)	.15
Mode of delivery								
Contact form versus webchat	1.82 (0.78 - 4.26)	.17	4.98 (2.29 - 10.83)	<.001	3.95 (1.98 - 7.89)	<.001	2.90 (1.60 - 5.25)	<.001
Email versus webchat	1.69 (0.83 - 3.44)	.15	0.14 (0.02 - 1.05)	.06	2.77 (1.50 - 5.13)	.001	1.24 (0.72 - 2.15)	.44
Phone versus webchat	4.17 (1.58 - 11.00)	.004	0.00 (0.00-Inf)	>.99	0.83 (0.24 - 2.93)	.78	0.99 (0.38 - 2.60)	>.99
Gender								
Men versus women	0.63 (0.29 - 1.40)	.26	1.87 (0.80 - 4.36)	.15	1.35 (0.71 - 2.58)	.37	1.07 (0.61 - 1.88)	.81
Nonbinary or transgender versus women	0.00 (0.00-Inf)	>.99	0.00 (0.00-Inf)	>.99	0.34 (0.06 - 1.90)	.22	0.40 (0.11 - 1.43)	.16
Other versus women	0.00 (0.00-Inf)	>.99	0.00 (0.00-Inf)	>.99	8.46 (2.13 - 33.66)	.002	4.49 (1.22 - 16.59)	.02
Prefer not to say versus women	0.00 (0.00-Inf)	>.99	3.82 (1.19 - 12.26)	.03	0.85 (0.26 - 2.72)	.78	0.86 (0.31 - 2.40)	.77
Ethnicity								
Asian versus White	2.40 (0.92 - 6.29)	.08	0.88 (0.25 - 3.06)	.84	1.37 (0.54 - 3.47)	.50	2.05 (0.95 - 4.42)	.07
Black versus White	1.56 (0.36 - 6.75)	.55	0.82 (0.09 - 7.77)	.86	2.00 (0.59 - 6.72)	.26	1.66 (0.56 - 4.91)	.36
Mixed versus White	1.72 (0.47 - 6.31)	.42	2.07 (0.47 - 9.10)	.34	0.69 (0.18 - 2.61)	.59	2.06 (0.81 - 5.25)	.13
Other versus White	1.89 (0.00-Inf)	>.99	3.79 (0.00-Inf)	>.99	1.12 (0.00-Inf)	>.99	88660893.75 (0.00-Inf)	>.99
Prefer not to say versus White	45.97 (3.56 - 594.15)	.003	0.00 (0.00-Inf)	>.99	3.17 (0.14 - 70.30)	.47	16.95 (1.50 - 191.26)	.02
Age (y)								
20 - 21 versus 16 - 19	1.89 (0.89 - 4.00)	.10	1.96 (0.78 - 4.93)	.15	3.13 (1.64 - 5.97)	.001	1.80 (1.01 - 3.20)	.045
22 - 25 versus 16 - 19	1.73 (0.86 - 3.48)	.12	2.92 (1.30 - 6.53)	.009	1.46 (0.75 - 2.86)	.26	1.80 (1.05 - 3.08)	.03
Older than 26 versus 16 - 19	0.82 (0.15 - 4.48)	.82	0.80 (0.09 - 7.42)	.84	1.08 (0.28 - 4.15)	.91	1.00 (0.31 - 3.24)	>.99
Prefer not to say versus 16 - 19	0.00 (0.00-Inf)	>.99	0.00 (0.00-Inf)	>.99	0.00 (0.00-Inf)	>.99	0.08 (0.01 - 1.14)	.06

[a]Note. Nagelkerke R^2=0.430. Extremely large odds ratios and wide CIs (eg, 0.00 to infinity) may result from zero or small sample sizes for comparisons. Nonbinary (n=8) and transgender (n=6) were collapsed due to small sample sizes. Other ethnicities (n=3) and those who preferred not to say their age (n=4) were not collapsed due to the distinct theoretical implications, but relevant findings should be interpreted with caution.

XSL•FO
RenderX

Qualitative Analysis

Of 358 participants who were signposted to at least 1 resource, 257 participants submitted 441 responses across 3 open-ended questions. Among these, 106 responses from 97 participants relevant to signposting (eg, “Sent me some links to places to help”) were analyzed. Among 106 responses, 76 (72%) indicated positive sentiment, while 30 (28%) were negative. Furthermore, 4 main themes were identified: “choice of signposts,” “signposting experience,” “responses to signposting,” and “effectiveness of signposting.” Table 4 presents the definitions, examples, and number of responses across themes. Note that each response could be coded to multiple subthemes.

XSL•FO
RenderX

Table . Definitions, examples, and number of responses across themes and subthemes.

Theme and subtheme	Response type and examples
Theme 1: choice of signposts	Comments specific to the signposts that influence young people's perception of whether signposting was effective
Relevance (n=8)	Whether the signposts were relevant to the young people's specific concerns • "provided some websites and organizations which would help me with my issue" • "Recommend me to an organization that can help me with my bank account issue"
Personalized (n=14)	Whether the signposts were specific to the young people's demographics (eg, geographical location, age, and support needs) • "clear and useful about what would be available to someone my age" • "Make note of the personal details I sent to you and tailor the help offered with those details"
Expectation (n=5)	Whether the signposts met young people's other expectations (eg, number, features, and specific requests) • "an app would be good with some resources"
Theme 2: signposting experience	Comments specific to how signposting was completed by helpline volunteers
Integration of emotional support with signposting (n=30)	How signposting was integrated with offering emotional support by the volunteers • "They understood what I was going through and gave me organizations to look at" • "…feel a bit demoralizing…better balance between providing some level of support and recommending services"
Elaboration on signposts (n=4)	Whether sufficient details of signposts were provided to young people • "…gave detailed explanations of when and how to use them" • "…explain more about what the websites are"
Preference for additional support (n=5)	Preference for receiving additional support such as direct referrals • "Employ some experts in the field of sexual assault…" • "contact the other agency with you"
Theme 3: responses to signposting	Immediate responses to signposting during or shortly after the chat
Hopefulness and direction (n=14)	Signposting inspiring hope and a sense of direction • "…gave me hope that things will improve" • "…pointing me in the right direction of where to go"
Rejection and overwhelm (n=7)	Signposting leading to feelings of rejection or overwhelm • "…felt like most comments…was a lead-in to recommend a specific service" • "I'm struggling to get through the list"
Theme 4: effectiveness of signposting	Effects of signposting on young people's well-being after the chat has ended
Insights into resources (n=53)	Enhanced knowledge of resources for young people • "…made me aware of some local resources that I didn't know was available"
Empowerment (n=13)	Enabling young people to learn more about their situation and seek support • "…managed to get some plan figured out for myself and scheduled appointments" • "feel a lot more confident in myself"
Access to signposts (n=10)	Ease for young people to access the signposts • "helplines provided to me were either no longer available or outdated" • "helped me to access the resources I needed"

XSL•FO
RenderX

Theme and subtheme	Response type and examples
Impacts on well-being (n=5)	Long-term mental health outcomes from signposting • "…helped not to kill myself" • "received counseling to get past…a time where I felt lost"

Theme 1: Choice of Signposts

Relevance

Participants appreciated the relevance of signposts to their concerns, while some critiqued their relevance. Participants valued the volunteers' effort in selecting the most pertinent signposts: "They also took the time to make sure they directed me to the most relevant resources." However, several participants highlighted that volunteers sometimes misunderstood their needs and shared less suitable signposts. For instance, "The websites that were recommended for me […] were to give me more information on my situation, rather than what I could do next to improve."

Personalized

Some participants recognized how signposts were tailored to their age, location, or support needs, whereas others found them too general. Participants liked it when volunteers inquired about their needs before signposting ("before recommending other services, I was asked how I'd like to contact people"). However, some felt that the volunteers failed to personalize the signposts despite their efforts to communicate their needs ("Make note of the personal details I sent to you and tailor the help offered").

Expectation

Responses mentioned specific requests, such as being signposted to apps ("an app would be good with some resources").

Theme 2: Signposting Experience

Integration of Emotional Support and Signposting

Participants felt heard alongside being signposted ("felt really listened to and actually heard"), while others expressed dissatisfaction with the signposting experience. Participants appreciated being "welcomed with compassion and understanding in very quick time frames" and the nonjudgmental space to share ("gave me the space to get things off my chest"). Yet, some responses suggested that the emotional support felt scripted ("might have been [an] automatic response") and "a lead-in to recommend a specific service." One participant noted the imbalance between emotional support and signposting: "Try not to solely rely on signposting loads of organisations, as it can feel a bit demoralising to just be moved onto someone else."

Elaboration on Signposts

Some responses valued detailed explanations of signposts, while others anticipated further information. Participants appreciated how the volunteers "gave detailed explanations of when and how to use [the signposts]," while others wanted volunteers to "[explain] what each thing could do, such as explaining what counselling is and what it would benefit."

Preference for Additional Support

Responses indicated a preference for additional support, such as in-house specialists ("Employ some experts in the field of sexual assault, rape"), more partnerships ("working with more [organizations]"), or direct referrals. One participant suggested that direct referrals may be more effective than signposting because young people may not be well-informed enough to access support: "Make the referral for [the person] because they may be confused and not do it, putting themselves at more of a risk."

Theme 3: Responses to Signposting

Hopefulness and Direction

Participants illustrated that signposting inspired hope and a sense of direction. Before using the helpline, participants were desperate for support ("I was really desperate for some help and you gave it to me"). Signposting then led them to feel more hopeful about the future ("given me some hope that improvement to my situation is possible") and seek help ("I hope at least one or more [of the signposts] will be helpful"). Additionally, some acknowledged how signposting guided them to appropriate services ("point me in the right directions to organizations and resources"), and they "don't feel stuck again."

Rejection and Overwhelm

Participants conveyed feelings of rejection or overwhelm. Several participants felt dismissed to other services and regretted seeking help: "I was just being referred to other services, so I am quite unhappy that my reaching out didn't pay off as I feel like I'm not gonna get helped." Another emphasized that the excessive number of signposts demotivated them from seeking support: "I was overwhelmed by the amount of other places to look for help that I'm struggling to get through the list."

Theme 4: Effectiveness of Signposting

Effective signposting empowers young people to learn about, trust, and take action to access support [13]. Subthemes were developed to reflect these aspects.

Insights Into Resources

Many participants discovered new local resources ("[…] made me aware of some local resources that I didn't know was available") or found suitable resources ("helped me to find the resources to get the help I need").

Empowerment

Participants demonstrated how signposting empowered them to express themselves and seek help. They found their situations more manageable ("Made me feel in control of my situation") and became confident to open up ("I had the courage to speak up to someone about how I felt"). Some took immediate action to access support: "With only contacting the helpline yesterday

XSL•FO
RenderX

I have already managed to get some plan figured out for myself and scheduled appointments."

Access to Signposts

Participants critiqued that signposts were inaccessible, although 3 accessed the suitable signposts ("helped me to access the resources I needed"). Furthermore, 6 participants were unable to use The Mix's counseling services due to unavailability: "I did want telephone counselling but the referral pathway was closed due to no space." One highlighted other signposts were outdated: "helplines provided to me were either no longer available or outdated."

Impacts on Well-Being

Participants shared that the signposts improved their well-being ("I still use the app that they suggested to me, and so far it's been helping"). Signposting helped participants navigate difficult times ("received counseling to get past […] a time where I felt lost") and was even lifesaving ("It helped found myself and the right people to go. It helped not to kill myself").

Connections With Quantitative Findings

A breakdown of the participants' demographics and the sentiment of their responses is presented in Table 5. Regarding the mode of delivery, email (n=12, 63%), telephone (n=4, 67%), and contact form (n=34, 68%) showed similar levels of positive sentiment, while webchat had the highest proportion of positive sentiment (n=26, 84%). Email helpline users mentioned that signposts were given too quickly *(*"Through email, take time to work through each issue, rather than quickly offering other organisations") while others reported otherwise *(*"With the email I can take my time and follow each link"). Contact form users had mixed sentiments, with some appreciating its asynchronous delivery ("helpful in visualising and stepping away from my issue") and others dissatisfied with feeling pushed away *(*"[…] actually help me rather than referring"). A few phone helpline users commented on the relevance of signposts ("Provide more relevant support services").

XSL•FO
RenderX

Table . Demographics characteristics and sentiments of the open-ended responses (n=106).

Characteristics	Positive, n/N (%)	Negative, n/N (%)
Gender		
Women	38/55 (69)	17/55 (31)
Men	11/16 (69)	5/16 (31)
Nonbinary	2/2 (100)	0/2 (0)
Transgender	1/1 (100)	0/1 (0)
Other	3/3 (100)	0/3 (0)
Prefer not to say	2/4 (50)	2/4 (50)
Missing	19/25 (76)	6/25 (24)
Age (y)		
16 - 19	30/45 (67)	15/45 (33)
20 - 21	9/11 (82)	2/11 (18)
22 - 25	16/23 (70)	7/23 (30)
Older than 26	3/3 (100)	0/3 (0)
Prefer not to say	1/1 (100)	0/1 (0)
Missing	17/23 (74)	6/23 (26)
Ethnicity		
White	47/66 (66)	19/66 (66)
Asian	6/9 (67)	3/9 (33)
Black	0/2 (0)	2/2 (100)
Mixed	1/1 (100)	0/1 (0)
Other	0/0 (0)	0/0 (0)
Prefer not to say	2/2 (100)	0/2 (0)
Missing	20/26 (77)	6/26 (23)
Mode of delivery		
Contact form	34/50 (68)	16/50 (32)
Email	12/19 (63)	7/19 (37)
Phone	4/6 (67)	2/6 (33)
Webchat	26/31 (84)	5/31 (16)

Regarding gender and ethnicity, the same proportion of participants who identified as women (38/55, 69%) and men (11/16, 69%) reported positive sentiments, while 100% (6/6) of those identifying as nonbinary (n=2), transgender (n=1), or other genders (n=3) expressed positive sentiments. Of those who preferred not to say their gender (n=4), 2 (50%) expressed positive sentiment. White (n=47, 71%) and Asian (n=6, 67%) participants showed similar levels of positive sentiments, but there were 100% positive responses among those identifying as mixed ethnicity (n=1) or who preferred not to say (n=2). Black participants exclusively left negative responses (n=2, 100%), with one hoping for support from in-house experts in sexual assault and another commenting on access to The Mix's internal counseling services.

In terms of age, responses from 16- to 19-year-olds (n=30, 67%) and 22- to 25-year-olds (n=16, 70%) had a similar percentage of positive sentiments, compared with much higher proportions of 20- to 21-year-olds (n=9, 82%), older than 26-year-olds (n=3, 100%), and those who preferred not to say (n=1, 100%). Several responses are directly related to age. For example, a 22- to 25-year-old preferred more age-tailored signposts ("more information about […] different services that are available to young people"), while an older than 26-year-old parent appreciated the helpline ("My daughter is really struggling. Just knowing I contacted someone helped her").

Discussion

Overview

Our study revealed that participants generally perceived signposts as helpful or were at least motivated to access them after using the helpline, illustrating how signposting may empower young people to seek relevant support. Signposting outcomes varied across modes of delivery, with webchat showing the strongest association with positive outcomes, the

XSL•FO
RenderX

contact form yielding mixed results, and email or phone showing weaker associations. Similar outcomes were observed across gender, ethnicity, and age, although certain minority groups (eg, those preferring not to say their gender or ethnicity, 20- to 25-year-olds) were more likely to ignore or find signposts irrelevant.

Qualitative thematic analysis indicated that participants generally found the signposts relevant and personalized. They valued volunteers' efforts in listening actively and explaining the signposts, which appeared to inspire hope and motivate participants to access support. However, some felt bombarded with too many signposts, leading to feelings of rejection or overwhelm, potentially reducing their intention to use signposts. Those discouraged from accessing signposts often faced barriers such as outdated information, explaining why some participants may have used signposts but found them unhelpful.

Mode of Delivery

The mode of delivery was significantly associated with participants' perceptions of usefulness and their use of signposts, underlining their benefits and disadvantages for young people. Across the 4 modes, webchat was associated with the strongest positive outcomes, with participants perceiving signposts as relevant, trustworthy, and helpful. Real-time interactions with volunteers via webchats may facilitate collaborative therapeutic relationships by allowing young people to control the pace of disclosure, without the social pressure to respond instantly in phone calls. This may empower young people to share emotionally challenging experiences [35], such as suicidality, thus helping volunteers to identify the most suitable signposts. Although helpline contact is brief, relational processes, such as emotional containment, trust-building, and perceived autonomy, are central to whether signposting feels safe and usable. Our qualitative themes suggest that when young people feel heard and validated, signposting is more likely to be experienced as supportive rather than overwhelming. These relational processes are particularly important for young people at higher risk, such as those disclosing suicidality or experiencing a mental health crisis. Volunteers should be trained to prioritize risk assessment and immediate safety planning, ensuring that signposting or referrals to external services complement rather than replace immediate support.

The contact form showed weaker associations, with signposts often perceived as irrelevant. The contact form requires users to explain their situation and receive one-off support asynchronously. For this reason, the degree of personalization in signposts may be limited by the amount of information the users share [19], causing some to perceive the signposts as less tailored.

Similarly, email support was associated with less positive outcomes for young people, possibly due to its asynchronous nature, with participants not using the signposts but intending to. Limited information may affect volunteers' ability to select appropriate signposts, lowering their confidence to provide appropriate support via email [19]. Volunteers' reduced emotional connection to users who contact via email [19] may further encourage volunteers to focus on signposting rather than taking more time to understand their issues, which our findings suggest may lead to some user dissatisfaction with their signposts and email support, more generally.

Despite positive reception in the existing research [36], the phone helpline in our study mainly offered 1 signpost per user, which was often perceived as unhelpful. Potential power imbalances between volunteers and users may explain this finding. Telephone helplines enhance the authority of volunteers [35], possibly reducing young people's sense of control in the conversation, and they could agree to signposts that they may not find beneficial [20].

Integrating Findings Across Methods

Our qualitative themes help interpret why webchat was associated with more favorable signposting outcomes. Webchat had the highest proportion of positive sentiment (26/31, 84%), compared with 63% - 68% across other modes. Young people most often described signposting as helpful when it was embedded within an emotionally validating interaction (refer to Integration of Emotional Support With Signposting section under Theme 2) and when resources felt personally relevant rather than generic (refer to Personalized section under Theme 1). Webchat may support this because synchronous text-based conversation allows users to disclose at their own pace while also enabling volunteers to check understanding, collaboratively narrow options, and immediately clarify barriers to follow-through (eg, eligibility, waiting lists, and how to access), conditions that facilitate both personalization and emotional integration. Where these conditions were absent, young people described feelings of rejection or overwhelm (refer to Rejection and Overwhelm section under Theme 3), or found signposts insufficiently tailored to their demographics or circumstances (refer to Relevance section under Theme 1). This relational "fit" may be harder to achieve in asynchronous modes (email or contact form) where reduced back-and-forth can limit tailoring, and in phone-based interactions where perceived authority dynamics may reduce collaborative choice.

Gender

Signposting was generally associated with positive outcomes across genders, including women, men, nonbinary, and transgender young people. Those identifying as "other" tended to either plan to use the signposts or find them irrelevant, whereas those preferring not to disclose their gender were more likely not to use the signposts. Gender minorities, who may be hesitant to seek help from general services, could find signposting valuable for discovering lesbian, gay, bisexual, transgender, queer or questioning, and others (LGBTQ+) informed services [37], boosting their confidence in accessing support [13]. Nonetheless, as existing LGBTQ+ services may not meet the needs of all, some may find their signposts irrelevant. Participants who preferred not to share their gender might have received less relevant signposts, as volunteers had insufficient information to identify suitable services. Preferences not to disclose gender may reflect broader concerns about confidentiality and stigma, as well as uncertainty about whether the service is gender-affirming, rather than simply a lack of relevance of the signposts. While the open-ended responses seem to indicate that young people who identified as nonbinary, transgender, or other were extremely positive about The Mix's

XSL•FO
RenderX

signposting, the sample sizes for gender minorities were notably small (6 to 25 signposts, 6 to 17 participants), requiring further research to validate our findings. Nondisclosure of gender or ethnicity may reflect safety and trust considerations. Qualitative comments suggest that perceived cultural fit and confidentiality concerns shape willingness to engage with signposts.

Ethnicity

Signposting was generally associated with positive perceptions across ethnicities, although participants preferring not to say their ethnicity were more likely to find signposts unhelpful and irrelevant. For those who preferred not to disclose their ethnicity and found signposts unhelpful, volunteers may have lacked sufficient personalized information to customize their signposts. As with gender, nondisclosure of ethnicity may reflect that building trust with marginalized groups can take longer due to broader issues related to privacy, stigma, and cultural norms around help-seeking, alongside previous experiences of discrimination when seeking support, causing some to withhold their details [22,26]. In our qualitative analysis, Black participants critiqued the signposts; one mentioned limited access to counseling, and the other preferred receiving sexual assault support within The Mix rather than being signposted elsewhere, demonstrating that access to support remains a major barrier to help-seeking. Overall, the results suggest a need to regularly review and update the signposting database to provide more culturally specific and tailored services. Nonetheless, sample sizes for other ethnicities are remarkably small (3 signposts, 14 participants), so these findings should be interpreted cautiously. More broadly, nondisclosure can be understood as an equity signal. When services feel uncertain in cultural safety, young people may reduce identifiable data-sharing. Service design should therefore treat demographic "prefer not to say" responses not as missingness alone, but as potentially meaningful indicators of trust, perceived stigma, and safety.

Age

Signposting was generally helpful across age groups, yet 20- to 25-year-olds were more likely to find signposts irrelevant. 22- to 25-year-olds were more inclined not to use the signposts, while 20- to 21-year-olds would plan to use them. Our qualitative findings suggest that these results indicate that more age-tailored support is needed. For example, 20- to 21-year-olds could be signposted to student support and external services [38], while 22- to 25-year-olds may require more services focused on the transition from school to work, relationship advice, and parenting. Notably, advice-seeking users older than 25 years were often caregivers or professionals seeking advice for young people.

Limitations

There are several limitations to this study. First, since the participants were recruited by emailing helpline users the survey link, self-selection bias may have influenced our results [39]. Those having strong opinions about the helpline may have been more likely to share their feedback, hence skewing the data. Additionally, listwise deletion based on missing demographic data could have retained participants who were more comfortable or engaged with the service. Future studies should consider using stratified sampling to ensure that demographic groups of interest are sampled equally. Second, as we used survey data from a general UK-based helpline for young people, our analytic sample, 64% White (n=189) and 74% (n=221) identifying as men or women, was representative of the user demographic at The Mix, which is 82% (24,158/29,389) White and 97% (34,053/35,215) men or women. Due to the small percentages of ethnic and gender minorities, we collapsed categories for the quantitative analyses. Inflated ORs were still found, as not all minority groups have sufficient observations for each signposting outcome. Future research should focus on gender, ethnicity, and sexuality to enrich the existing findings, conducting focus groups or interviews for more in-depth qualitative analysis.

Furthermore, the outcomes measured in the survey may not be easily compared with outcomes used by other helplines, such as quantitative, self-report measures (eg, study by Law et al [40]). As both our study and other helplines tend to rely on retrospective self-report, recall bias may have affected the accuracy of participants' responses. To understand the different stages of signposting [13], additional research could split the question on signposting into 4 subquestions asking participants whether they found the signposts relevant, whether they planned to use them, whether they accessed the signposts, and whether the signposts were helpful if they used them.

Recommendations

These findings highlight the critical role of signposting in supporting young people's mental health, particularly through webchat, which emerged as the most effective mode of delivery. Given the increasing reliance on digital mental health services, ensuring that signposting remains accessible, relevant, and tailored to diverse user needs is essential. By optimizing signposting strategies, helplines can empower young people to seek appropriate long-term support, ultimately improving mental health outcomes.

Given the mixed findings of other communication channels (eg, email, contact form, and phone), enhancing volunteer training in these areas is recommended. Training should focus on balancing emotional support and signposting in asynchronous communication, as well as effectively navigating suitable signposts during phone-based interactions [13]. For webchat, volunteers should be trained to use brief collaborative checks (eg, confirming priorities, offering 2 - 3 best-fit options, and checking practical barriers such as eligibility, cost, and waitlists) so that signposting feels tailored and actionable rather than overwhelming. For email and contact-form interactions, templates that include a short, personalized rationale for each resource and clear "next-step" instructions (eg, who to contact and what to say) may improve perceived relevance and follow-through.

Additionally, periodic reviews of the signposting database are essential to ensure signposts are up to date and comprehensive. Detailed descriptions of available signposts (eg, referral requirements, waitlist lengths, and access pathways) can enhance the user experience and improve the perceived relevance of signposts [16]. Where feasible, services could also provide an

XSL•FO
RenderX

"actionable signpost pack," particularly suited to asynchronous modes, such as email and contact form (eg, 1 - 3 links plus a brief explanation of why each option fits, what to expect next, and how to access), to reduce cognitive load and increase the likelihood of uptake. Expanding the database to include additional services tailored to ethnic minorities and older youth (20- to 25-year-olds) is also critical, as these groups were identified as having unique needs that may not be adequately addressed by existing services [14]. To further support marginalized groups, services could train volunteers to introduce optional demographic questions as a tool for personalizing support rather than an administrative requirement, and to incorporate explicit affirmations of cultural safety and inclusivity at the start of interactions.

Finally, integrating feedback mechanisms into the signposting process can ensure continuous improvement and adaptability to the evolving needs of young people. As part of routine quality improvement, services could monitor signposting outcomes by mode of delivery and patterns of nondisclosure, treating increases in "not relevant" responses or high "prefer not to say" rates as potential signs to review training, cultural safety, and the signposting database. For webchat, real-time postsession feedback prompts may be the most feasible. For asynchronous modes, follow-up survey links could be embedded in email responses. Periodic evaluations of these mechanisms will allow helplines to refine their signposting strategies, ensuring they remain effective and user-centered. Strengthening these approaches will not only enhance engagement but also reinforce the role of helplines as a vital bridge between immediate support and long-term health care.

Acknowledgments

We are grateful to The Mix, especially Bohdana Dock, for their support throughout the study.

Funding

This research received no external funding.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey, codebook, and individual models.
[DOCX File, 33 KB - humanfactors_v13i1e73369_app1.docx]

Checklist 1
CHERRIES checklist.
[PDF File, 117 KB - humanfactors_v13i1e73369_app2.pdf]

References

1. Mental health of adolescents. World Health Organization. URL: https://who.int/news-room/fact-sheets/detail/adolescent-mental-health [accessed 2025-01-28]
2. Mental health of children and young people in England, 2023 - wave 4 follow-up to the 2017 survey. NHS England. 2023 Nov 21. URL: https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2023-wave-4-follow-up [accessed 2025-01-28]
3. Brief guide: waiting times for community child and adolescent mental health services. Care Quality Commission. 2020. URL: https://cqc.org.uk/sites/default/files/Brief_guide_Waiting_times_for_community_child_and_adolescent_mental_health_services.pdf [accessed 2025-01-28]
4. Edbrooke-Childs J, Deighton J. Problem severity and waiting times for young people accessing mental health services. BJPsych Open 2020 Oct 12;6(6):e118. [doi: 10.1192/bjo.2020.103] [Medline: 33040766]
5. Anderson JK, Howarth E, Vainre M, Jones PB, Humphrey A. A scoping literature review of service-level barriers for access and engagement with mental health services for children and young people. Child Youth Serv Rev 2017 Jun;77:164-176. [doi: 10.1016/j.childyouth.2017.04.017]
6. Westberg KH, Nyholm M, Nygren JM, Svedberg P. Mental health problems among young people-a scoping review of help-seeking. Int J Environ Res Public Health 2022 Jan 27;19(3):1430. [doi: 10.3390/ijerph19031430] [Medline: 35162452]
7. Westin AML, Barksdale CL, Stephan SH. The effect of waiting time on youth engagement to evidence based treatments. Community Ment Health J 2014 Feb;50(2):221-228. [doi: 10.1007/s10597-012-9585-z] [Medline: 23283487]
8. Leece S, Wilson H, Andriopoulou P. Experience of online, text-based, anonymous mental health support during COVID-19 for young women previously at risk of suicide. Couns Psychol Q 2024 Oct;37(4):495-516. [doi: 10.1080/09515070.2023.2252770]

XSL•FO
RenderX

9. Bloch S, Leydon G. Conversation analysis and telephone helplines for health and illness: a narrative review. Res Lang Soc Interact 2019 Jul 3;52(3):193-211. [doi: 10.1080/08351813.2019.1631035]
10. Fukkink RG, Bruns S, Ligtvoet R. Voices of children from around the globe; an international analysis of children's issues at child helplines. Child Soc 2016 Nov;30(6):510-519 [FREE Full text] [doi: 10.1111/chso.12150]
11. Alexander M, Hofstetter E. Somewhere to turn to: signposting in service provision. Discourse & Communication 2021 Apr;15(2):119-138. [doi: 10.1177/1750481320982091]
12. Mustafa N, Mumvuri J, Walsh P, Lonsdale P, Green J. Are you feeling stressed? Looking after the mental health of health students: resources to raise self-awareness and signpost health students to sources of self help and support. Science Talks 2023 May;6:100169. [doi: 10.1016/j.sctalk.2023.100169]
13. Hughes G, Panjwani M, Tulcidas P, Byrom NC. Student mental health: the role and experiences of academics. : King's College London; 2018 URL: https://kclpure.kcl.ac.uk/portal/en/publications/student-mental-health-the-role-and-experiences-of-academics [accessed 2026-04-30]
14. Edwards D, Carrier J, Csontos J, et al. Review: Crisis responses for children and young people - a systematic review of effectiveness, experiences and service organisation (CAMH-Crisis). Child Adolesc Ment Health 2024 Feb;29(1):70-83. [doi: 10.1111/camh.12639] [Medline: 36785997]
15. Gilmour L, Ring N, Maxwell M. Review: The views and experiences of suicidal children and young people of mental health support services: a meta-ethnography. Child Adolesc Ment Health 2019 Sep;24(3):217-229. [doi: 10.1111/camh.12328] [Medline: 32677214]
16. Brunton L, Tazzyman A, Ferguson J, Hodgson D, Nelson PA. The challenges of integrating signposting into general practice: qualitative stakeholder perspectives on care navigation and social prescribing in primary care. BMC Prim Care 2022 Apr 1;23(1):66. [doi: 10.1186/s12875-022-01669-z] [Medline: 35365072]
17. Saxild S, Wilson P, de Voss S, Overbeck G. Clinicians' experiences in signposting an online mental health resource to expectant mothers: a qualitative study. BMC Pregnancy Childbirth 2023 May 10;23(1):336. [doi: 10.1186/s12884-023-05671-w] [Medline: 37165318]
18. Tegethoff M, Stalujanis E, Belardi A, Meinlschmidt G. School mental health services: signpost for out-of-school service utilization in adolescents with mental disorders? A nationally representative United States cohort. PLoS ONE 2014;9(6):e99675. [doi: 10.1371/journal.pone.0099675] [Medline: 24911241]
19. Campagnola M, Burlibasa A, Gutman LM. Barriers and enablers to the delivery of email communication for a helpline service for young people. PEC Innov 2022 Dec;1:100059. [doi: 10.1016/j.pecinn.2022.100059] [Medline: 37213768]
20. Gibson K, Cartwright C, Kerrisk K, Campbell J, Seymour F. What young people want: a qualitative study of adolescents' priorities for engagement across psychological services. J Child Fam Stud 2016 Apr;25(4):1057-1065. [doi: 10.1007/s10826-015-0292-6]
21. Delsignore A, Rufer M, Emmerich J, Weidt S, Brühl AB, Moergeli H. E-mail support as an adjunct to cognitive-behavioral group therapy for social anxiety disorder: impact on dropout and outcome. Psychiatry Res 2016 Oct 30;244:151-158. [doi: 10.1016/j.psychres.2016.07.038] [Medline: 27479106]
22. Franks M, Medforth R. Young helpline callers and difference: exploring gender, ethnicity and sexuality in helpline access and provision. Child Fam Soc Work 2005 Feb;10(1):77-85. [doi: 10.1111/j.1365-2206.2005.00345.x]
23. Mathieu SL, Uddin R, Brady M, et al. Systematic review: the state of research into youth helplines. J Am Acad Child Adolesc Psychiatry 2021 Oct;60(10):1190-1233. [doi: 10.1016/j.jaac.2020.12.028] [Medline: 33383161]
24. Sagar-Ouriaghli I, Godfrey E, Bridge L, Meade L, Brown JSL. Improving mental health service utilization among men: a systematic review and synthesis of behavior change techniques within interventions targeting help-seeking. Am J Mens Health 2019;13(3):1557988319857009. [doi: 10.1177/1557988319857009] [Medline: 31184251]
25. Baskin C, Zijlstra G, McGrath M, et al. Community-centred interventions for improving public mental health among adults from ethnic minority populations in the UK: a scoping review. BMJ Open 2021 Apr;11(4):e041102. [doi: 10.1136/bmjopen-2020-041102]
26. Mojaverian T, Hashimoto T, Kim HS. Cultural differences in professional help seeking: a comparison of Japan and the U.S. Front Psychol 2012;3:615. [doi: 10.3389/fpsyg.2012.00615] [Medline: 23426857]
27. Gould D, Lauer L, Rolo C, Jannes C, Pennisi N. Understanding the role parents play in tennis success: a national survey of junior tennis coaches. Br J Sports Med 2006 Jul;40(7):632-636. [doi: 10.1136/bjsm.2005.024927] [Medline: 16702176]
28. Stoilova M, Livingstone S, Donovan S. Outcomes and effectiveness of children's helplines: a systematic evidence mapping. : The London School of Economics and Political Science; 2019 URL: https://lse.ac.uk/business/consulting/reports/outcomes-and-effectiveness-of-childrens-helplines [accessed 2025-01-28]
29. Ethnic group, England and Wales: Census 2021. Office for National Statistics. 2022 Nov 29. URL: https://ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/bulletins/ethnicgroupenglandandwales/census2021 [accessed 2025-01-28]
30. de Jong VMT, Eijkemans MJC, van Calster B, et al. Sample size considerations and predictive performance of multinomial logistic prediction models. Stat Med 2019 Apr 30;38(9):1601-1619. [doi: 10.1002/sim.8063] [Medline: 30614028]
31. Ihaka R, Gentleman R. R: a language for data analysis and graphics. J Comput Graph Stat 1996 Sep;5(3):299-314. [doi: 10.1080/10618600.1996.10474713]

XSL•FO
RenderX

32. Braun V, Clarke V. Thematic analysis. In: Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D, Sher KJ, editors. APA Handbook of Research Methods in Psychology, Vol 2 Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological: American Psychological Association; 2012:57-71. [doi: 10.1037/13620-004]
33. NVivo | The #1 qualitative data analysis software for 30 years. Lumivero. URL: https://lumivero.com/products/nvivo/ [accessed 2025-01-28]
34. Graham M, Milanowski A, Miller J. Measuring and promoting inter-rater agreement of teacher and principal performance ratings. : Center for Educator Compensation Reform; 2012 Feb URL: https://files.eric.ed.gov/fulltext/ED532068.pdf [accessed 2025-01-28]
35. Kit PL, Teo CT, Tan M, Park Y. Singapore children's online counselling experiences on a live chat portal. Br J Guid Counc 2019 May 4;47(3):304-316. [doi: 10.1080/03069885.2018.1485871]
36. Digital direction youth consultation report 2019. : The Participation People; 2019 URL: https://infogram.com/p2b-or-place2be-online-mental-health-research-rep-2018-19-1h174908r5jq4zj?live
37. Goldbach JT, Rhoades H, Green D, Fulginiti A, Marshal MP. Is there a need for LGBT-specific suicide crisis services? Crisis 2019 May;40(3):203-208. [doi: 10.1027/0227-5910/a000542] [Medline: 30109965]
38. Cage E, Jones E, Ryan G, Hughes G, Spanner L. Student mental health and transitions into, through and out of university: student and staff perspectives. Journal of Further and Higher Education 2021 Sep 14;45(8):1076-1089. [doi: 10.1080/0309877X.2021.1875203]
39. Bethlehem J. Selection bias in web surveys. Int Statistical Rev 2010 Aug;78(2):161-188. [doi: 10.1111/j.1751-5823.2010.00112.x]
40. Law J, Haner D, Simon A. Proof positive: evaluation of kids help phone's phone and live chat counselling services. Kids Help Phone. 2015. URL: https://kidshelpphone.ca/wp-content/uploads/Kids-Help-Phone-Proof-Positive-2015-English.pdf [accessed 2025-01-28]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
LGBTQ+: lesbian, gay, bisexual, transgender, queer or questioning, and others
OR: odds ratio

Edited by C Jacob; submitted 03.Mar.2025; peer-reviewed by E Hernandez-Encuentra, R Jansen, S Joshi; revised version received 23.Feb.2026; accepted 27.Feb.2026; published 19.May.2026.

Please cite as:
Au AYF, Jan A, Phua SW, Gutman LM
Young People's Perceptions of Signposting in a Digital Mental Health Helpline: Mixed Methods Analysis of Cross-Sectional Data
JMIR Hum Factors 2026;13:e73369
URL: https://humanfactors.jmir.org/2026/1/e73369
doi:10.2196/73369

© Angel Yee Fong Au, Anya Jan, Shyn Wei Phua, Leslie Morrison Gutman. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 19.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Identifying Skill and Usability Barriers to Digital Health Tool Use Among Older Adult Patients in US Safety Net Clinics: Mixed Methods Study

Taylor Rapson[1,2], PhD; Magaly Ramirez[1], PhD; Sandy He[2], BS; Jeanette Wong[2,3], BS; Hyunjin Cindy Kim[2,3], MPH; Isabel Luna[2,3], BA; Andersen Yang[2,3], MPH; Junhong Li[2,3], BS; Paul A Fishman[1], PhD; James D Ralston[1,4], MPH, MD; Courtney R Lyles[2,3,5*], PhD; Elaine C Khoong[2,3,6*], MS, MD

[1]Department of Health Systems and Population Health, School of Public Health, University of Washington, Seattle, WA, United States

[2]Division of General Internal Medicine at Zuckerberg San Francisco General Hospital, Department of Medicine, University of California, San Francisco, 1001 Potrero Avenue, Main Building 5, 1st floor, 1M, San Francisco, CA, United States

[3]Action Research Center for Health, Zuckerberg San Francisco General Hospital, University of California, San Francisco, San Francisco, CA, United States

[4]Kaiser Permanente Washington Health Research Institute, Seattle, WA, United States

[5]Center for Healthcare Policy and Research, University of California at Davis, Sacramento, CA, United States

[6]Division of Clinical Informatics and Digital Transformation (DoC-IT), Department of Medicine, University of California, San Francisco, San Francisco, CA, United States

[*]these authors contributed equally

Corresponding Author:
Elaine C Khoong, MS, MD
Division of General Internal Medicine at Zuckerberg San Francisco General Hospital, Department of Medicine, University of California, San Francisco, 1001 Potrero Avenue, Main Building 5, 1st floor, 1M, San Francisco, CA, United States

Abstract

Background: Despite their benefits, digital health tools often face adoption barriers because of the digital divide. Identifying the fundamental user skills required to effectively navigate these tools and the usability barriers is essential to addressing disparities in use.

Objective: This study aimed to identify the skill and usability barriers to using digital health tools.

Methods: This study included English-, Spanish-, or Cantonese-speaking patients, aged ≥50 years, who received care at an urban safety net health system in the United States. Participants completed a survey examining sociodemographic characteristics and digital health tool use and were observed and video recorded as they navigated four digital health care tasks: (1) launch a video visit, (2) visit a health website through a URL, (3) log in to the patient portal, and (4) sign up for a patient portal account. Participants who could not independently perform the tasks received additional support. Tasks were conducted in English, while instructions and additional assistance were provided in each participant's preferred language. Video recordings were thematically coded to identify the fundamental skills needed for effective digital tool use and usability barriers in the design of digital tools. We examined whether task independence was associated with participant demographics and thematic categories using Kruskal-Wallis, χ^2, and Fisher exact tests.

Results: In total, 74% (34/46), 52% (31/60), 71% (44/62), and 70% (43/61) of participants (N=64) independently completed digital tasks 1, 2, 3, and 4, respectively. Older age, minoritized races and ethnicities, non-English language preference, lower educational attainment, access to cellular data only or no internet access, and lack of a portal account were associated with a higher likelihood of requiring assistance or being unsuccessful at completing each task (P<.001, except for older age [P=.004]). The qualitative coding of video recordings identified 3, 4, and 6 categories of typing, navigation, and human-computer interaction (HCI) skills, respectively, as fundamental skills required to independently complete digital tasks. χ^2 and Fisher exact tests indicated significant associations between most typing, navigation, and HCI categories and independent task completion. We coded usability barriers as one of 6 learnability challenges or 3 operability challenges.

Conclusions: This study identified that independent use of digital health tools requires fundamental typing, navigation, or HCI skills as well as high usability of digital tools. The inclusion of 4 different digital tasks added specificity to the type of skills and usability considerations necessary to ensure accessibility of digital health tools to diverse older adults. This study underscores the need for vendors to cocreate digital health tools with historically excluded end users in mind. As health care systems expand

XSL•FO
RenderX

digital tool adoption, they must distinguish fundamental skill gaps from usability barriers, as each may require different intervention strategies.

(JMIR Hum Factors 2026;13:e78430) doi:10.2196/78430

KEYWORDS

digital skills; usability; digital equity; digital health technology; digital disparities; patient portal; eHealth; digital literacy

Introduction

Electronic health records, patient portals, telemedicine, and eHealth (often referred to as digital health tools) have increased the opportunities for individuals to engage more directly in their own care. Access to these tools enables appointment scheduling, medication refills, and communication with health care providers, reducing some of the accessibility burdens of face-to-face interactions [1]. These tools have been linked to improved chronic disease outcomes [2-7]. However, there are documented differences in their use among marginalized populations who may face individual and systemic barriers that hinder adoption [8-11]. Previous research has demonstrated different rates of digital health tool use among individuals and groups from minoritized racial and ethnic backgrounds (including those who prefer languages other than English), age groups, and genders, as well as differences in use for those who report barriers to broadband internet, devices, or data (eg, due to cost) [2,8-14]. In particular, older adults, especially those with non-English language preferences or who identify as a racial or ethnic minority, experience large disparities in digital health tool adoption because of steep disparities in device, internet, or cellular data access [9,14,15]. For example, the Pew Research Center reported that, according to a 2025 survey, although 97% of adults younger than 50 years owned a smartphone, only 78% of those 65 years and older owned a smartphone [16]. The Pew Research Center also found that while 81% of White adults in the United States had home broadband, only 71% of Black adults and 68% of Hispanic adults had it. Similarly, Black (19%) and Hispanic (28%) participants were more likely to rely on their smartphones for internet access (which may be less stable than broadband) than White (13%) or English-speaking Asian (11%) participants [16]. This digital divide may be exacerbated in health care settings that rely on digital health tools in routine practice, which is becoming increasingly common [10,13,15,17]. Studies within the past 5 years show that nearly 90% of US health care systems offered telehealth services, especially nonprofit hospitals, which have adopted telemedicine at a higher rate than other health systems [18,19].

To address these barriers for marginalized populations, there is a need for pragmatic research rooted in identifying and understanding digital literacy skills and user needs [20-23]. There is a growing body of literature documenting the self-reported digital literacy skills needed to access digital tools, yet the fundamental user skills required to effectively navigate these tools have not been well defined [23-27]. Self-reported digital literacy skills refer to a user's perceived ability or confidence to perform general digital tasks and frequency of engagement, such as confidence in sending emails or frequency of using search engines. However, these metrics fail to delineate the fundamental knowledge and skills required to engage successfully with digital tools [25-27]. Fundamental knowledge and skills, in contrast, are the observable digital abilities required to navigate and interact with *any* digital device or system that must be present to successfully complete a digital task, such as being able to navigate a web browser, knowing how to open device apps, or recognizing clickable links [25-27]. Observational data are necessary to identify the fundamental knowledge and skills that impact self-reported digital literacy. To our knowledge, no studies have evaluated the fundamental digital knowledge and skills that individuals need to use digital health tools, particularly in marginalized communities disproportionately affected by the digital divide.

It is also important to identify usability barriers—which arise from poor product design—that prevent individuals from successfully interacting with digital tools and contribute to disparities in digital health tool uptake and use [17,22,28-30]. These usability challenges can impact both patients and clinicians. Clinical care team members may not be comfortable with workflows or tools for using eHealth or offering and supporting patients in using digital health tools [31]. Poorly designed interfaces, unclear navigation paths, and accessibility issues disproportionately impact individuals with limited digital literacy, disabilities, or language barriers [11,17,29,30,32]. These usability challenges, which often arise from limited patient participation in co-designing a digital health tool, can create unnecessary friction, discourage continued engagement, and deter patients from using digital health tools to engage in continuity of care [11,30,33]. Moreover, individuals from marginalized communities are less often represented in usability studies, limiting our understanding of the specific challenges they face when navigating these tools and hindering the development of equitable interventions to build the skills essential to improve the uptake of digital health tools [21,33]. Pragmatic human-centered research rooted in identifying the usability barriers for marginalized populations, using representative sampling within health care systems outside academic medical centers, may reveal key themes to improve user experience and mitigate disparities.

To address these gaps in the literature, we conducted a mixed methods study among multilingual older adults receiving care at an urban safety net health system. The aim of this study was to assess patient characteristics associated with successful completion of digital health tasks and to identify the skill and usability barriers that prevented successful completion. Using observational data of participants interacting with real-world digital health tools, we focused on the qualitative identification of skills and usability barriers mapped to existing frameworks. We then linked these qualitatively identified skills and barriers

XSL•FO
RenderX

to the observed completion rates of digital health tasks (such as completing the process of logging in to a patient portal) to pinpoint the fundamental skills necessary to complete each digital health task independently and the usability barriers that many individuals face when navigating these tasks. We also examined the task completion rates overall and by key sociodemographic characteristics to assess disparities in task completion success.

Methods

Study Setting

This study was conducted within the San Francisco Health Network (SFHN), San Francisco County's safety net health system encompassing 14 community-based primary care clinics. The SFHN serves a racially and ethnically diverse patient population of about 60,000, most of whom are enrolled in Medicaid or citywide insurance plans. The SFHN patient population is primarily non-White (80%); notably, 45% do not identify English as their primary language (35% prefer Spanish and >10% prefer Cantonese exclusively), and more than half report barriers to accessing and using digital health technology [34,35].

Recruitment and Participation

Individuals were eligible for participation in this study if they met the following inclusion criteria: (1) receiving primary care at an SFHN clinic; (2) being aged 50 years or older; (3) speaking English, Spanish, or Cantonese as their primary language; (4) and not having a cognitive or behavioral impairment that would preclude them from providing informed consent or completing the digital tasks. We focused on adults aged 50 years or older, given the prevalence of digital barriers within this population. We focused on these 3 languages, as these are the most spoken languages among patients at SFHN, where our study was conducted. Between April 2023 and June 2024, patients at SFHN who met these inclusion criteria were recruited by the study team over the phone after receiving prior approval from their primary care providers. Study team members recruited 101 eligible patients who subsequently completed a participant questionnaire and performed the digital tasks.

Ethical Considerations

The study was approved by the UCSF institutional review board (#22 - 36493). A subsample of 64 participants provided written consent to record their progress through each digital task for this study. Participants were advised of their right to stop taking part in the study at any time. Participant questionnaires, observation notes, and video recordings were deidentified using study identification numbers during data collection. All physical documents with data were stored in a locked cabinet in a secure location. All electronic data were stored on Research Electronic Data Capture (REDCap; version 13.4.10; Vanderbilt University) and on servers accessible only to the study team. Participants were compensated US $20 for completing the questionnaire and US $40 for completing up to 4 digital tasks while being observed.

Data Collection

The participant questionnaire, administered in the participant's preferred language, included items eliciting information on participant age, sex, race, ethnicity, preferred language, comfort speaking English, and educational attainment and questions about device preference, access to a device with video capability, type of internet access, and current enrollment in the SFHN patient portal. Questionnaires were administered in person or over the phone by study staff. All digital task observations were conducted and video recorded in person. See Multimedia Appendix 1 for questions from the participant questionnaire included in this analysis.

After the participants completed the questionnaire, they were asked to use their preferred digital device, if on hand, or to choose a device provided by the research team that they felt comfortable using (either an Android smartphone, an Apple iPad, a Microsoft Windows laptop, or an Apple MacBook) to complete the digital task assessment. Participants were filmed as they attempted four digital tasks that simulated the necessary skills and actions required to use digital health technology effectively: (1) launching a video visit, (2) going to a specific health website, (3) signing in to a patient portal emulator using a dummy account, and (4) signing up for a patient portal account via the emulator. These tasks were selected due to their importance for engaging with digital health tools. While digital health tools have become more inclusive in recent years and offer the ability to switch between English and Spanish (and sometimes other languages) on the platform itself, many elements of these digital tools (such as automated text messages or notifications and prompts) are often unavailable in languages besides English; therefore, to replicate that experience, the platforms that the participants navigated to complete each task were in English only. Further details about the steps involved in each task are detailed in Multimedia Appendix 2, and specific instructions for each task are available online [40].

Participants were asked to independently complete each task by following a set of written instructions in their preferred language; if they were unsuccessful after 2 attempts, research staff provided a standardized instructional handout or video, followed by tailored, language-concordant verbal assistance if necessary. After each step of assistance, participants were asked to try the task again; if they were unsuccessful again, the research staff proceeded to the next level of assistance. Participants were deemed unsuccessful if they were unable to complete the task after receiving tailored one-on-one assistance. The level of assistance provided was logged by the research team member conducting the assessment. For accuracy, a second team member reviewed video recordings and assigned their score. If the 2 scores differed, discrepancies were addressed via team discussion.

The participant questionnaire, written task instructions, and assistance from the research staff were administered in the participant's preferred language, as determined during recruitment. The questionnaire was either verbally administered by an interviewer proficient in the participant's preferred language or completed as a paper questionnaire translated into their preferred language. Additionally, the interviewers provided

XSL•FO
RenderX

translated instructions and, if needed, delivered verbal assistance.

Analysis

Qualitative Analysis

For the qualitative analysis, we used framework analysis—a matrix-based method that allows for systematic comparison of themes across cases while retaining case context—to identify both the fundamental skills required of users and the usability barriers that prevented successful use [36]. First, we watched a subsample of video observations to gain a comprehensive understanding of the participants' experiences and behaviors across the digital tasks. This initial stage helped to identify recurrent ideas, patterns, and areas of interest (typing skills, navigation skills, and human-computer interaction skills). We then developed a thematic framework to guide our analysis by generating a priori coding categories guided by the research questions ("what are the specific and foundational skills a user needs to successfully use digital health tools?" and "what are the usability barriers of these tools?") and the existing literature about digital literacy and usability [27,32,36-38]. We categorized qualitative codes according to the level of digital literacy they reflected and distinguished them as novice or advanced, where appropriate. This categorization was informed by the digital literacy framework of van Deursen et al [27] and modified to appropriately reflect the skills pertinent to using digital health tools rather than the internet in general. This framework differentiates operational, formal, informational, and strategic digital skills, which capture the range of competencies needed to use digital tools effectively [27]. We primarily focused on operational and formal skills, such as device handling, typing, and navigating digital interfaces, which were most relevant to the tasks in our study. Novice-level skills included novice keyboard skills and novice device skills, which were observed as operational inability or skill gaps when performing a task. Advanced-level skills included advanced typing and advanced device skills—skills that encompassed strategic problem-solving abilities, such as troubleshooting errors and intuitive device understanding. We also mapped the predetermined coding categories of learnability (the capability of digital health tools to enable the user to learn their application) and operability (the capability of digital health tools to enable the user to operate and control them) to challenges associated with completing digital tasks due to usability issues [38]. We also used an inductive coding approach to allow codes to emerge from the data to limit a priori bias. Inductive and deductive coding ensured that both anticipated and novel insights were captured.

Next, we compiled and organized the codes in a structured matrix format, which enabled us to summarize and compare the nature and context of responses, while preserving the richness of the video observations. Each row represented an individual participant, and each column presented the themes associated with a specific task (Figure S1 in Multimedia Appendix 2). As a calibration step, 3 independent coders watched a sample of the videos and created a templated summary for each task in the matrix tool. This was used to assess the template and confirm that the domains were intuitive; identify any missing or incorrectly labeled domains; clarify any issues; and ensure consistency in content, style, and organization across coders. During this process, actions that supported or deviated from these predicted codes led to additional codes, iteratively expanding upon and updating the codebook to include relevant concepts (see Multimedia Appendix 3). After finalizing the codebook and template, to assess coding consistency, 3 members of the research team independently coded 25% of the recordings. After ensuring >85% agreement, the 3 members independently coded the remaining recordings.

Quantitative Analysis

For the quantitative component, survey responses were first exported from REDCap for preliminary descriptive summaries, focusing on the variables age, sex, race and ethnicity, preferred language, educational attainment, access to a device with video capability, source of internet access, and enrollment in the SFHN patient portal. These sociodemographic characteristics comprised the covariates for analysis.

Our primary quantitative outcome variable was task completion, which assessed whether a participant independently performed the task only with simple written instructions, performed the task after receiving standard assistance (ie, detailed handouts and video tutorials), performed the task after receiving tailored assistance (ie, verbal instructions and one-on-one help), or was unsuccessful. For analyses, we dichotomized this outcome, categorized as performing a task independently versus requiring any support (standard or tailored) or being unsuccessful (examples of these materials are available online) [39]. This cut point was chosen, as it represents the key distinctions between a user demonstrating fundamental skills and those requiring further assistance or training. Because of technical issues (eg, malfunctioning cameras or recording failures) and participants occasionally declining tasks that they found too difficult, the number of video recordings varied slightly across tasks, as individuals without recordings were treated as missing data and removed from the analysis for that task.

We conducted global tests of association to evaluate relationships between participant sociodemographic characteristics and task completion across all 4 tasks. Kruskal-Wallis tests were used to test the association between age—modeled as a continuous variable—and task completion. χ^2 tests of independence were used for categorical sociodemographic variables, and Fisher exact tests were used when the expected cell counts were <5. These tests evaluated whether task completion status differed by sociodemographic characteristic when considering all tasks jointly, rather than within a single task. R software (version 4.3.3; R Foundation for Statistical Computing) was used to conduct all analyses.

Comparison of Qualitative Findings and Quantitative Data

To assess the relationship between identified themes and task completion, qualitative and quantitative analyses were conducted in parallel using a convergent triangulation design in which the datasets were integrated prior to the combined analysis [40-42]. Key qualitative themes were grouped into higher-order categories to facilitate interpretation. This involved identifying

XSL•FO
RenderX

conceptual connections between codes and organizing them into thematic categories. These thematic categories were then converted into categorical variables and incorporated into the quantitative dataset. Triangulating the 2 types of data supported the identification of foundational skills necessary to use digital health tools successfully (and, conversely, the skill gaps that required interventional support). We used χ^2 tests of independence and Fisher exact tests to determine the significance of associations between each identified skill or skill gap and the task completion outcome. Fisher exact tests were used when the expected cell counts were <5.

Results

Participant Characteristics

In total, 64 participants, comprising 40 men and 24 women, provided consent for the video recording of their digital task completion efforts. The language spoken by most participants (n=42, 65.6%) was English, followed by Cantonese (n=14, 21.9%) and Spanish (n=8, 12.5%). The mean age was 61.9 (SD 7.4; range 51.0 - 83.4) years. Notably, 79.7% (51/64) of the participants identified as non-White, and 53.1% (34/64) of the participants had an educational attainment of a high school degree or less. Most participants (55/64, 85.9%) strongly preferred using their smartphone, with some preferring a desktop or laptop computer (12/64, 18.8%) or tablet (9/64, 14.1%), and only 1.6% (1/64) preferred using a simple cell phone. Moreover, 90.6% (58/64) of the participants owned a device with video capability. Of the 64 participants, 60.9% (n=39) had both Wi-Fi and unlimited data on their smartphones, whereas 23.4% (n=15) used Wi-Fi only and 10.9% (n=7) had unlimited data only. Lastly, 65.6% (42/64) of the participants were enrolled in the patient portal at SFHN. Participant characteristics are displayed in Table 1.

XSL•FO
RenderX

Table . Participant descriptives.

Characteristic	Subsample with video observations (N=64)
Age (years), mean (SD)	61.9 (7.4)
Male sex (%)	62.5
Race and ethnicity, n (%)	
Non-Hispanic White	12 (18.8)
Non-Hispanic Black or African American	10 (15.6)
Non-Hispanic Asian	21 (32.8)
Non-Hispanic other or multiple races	4 (6.3)
Hispanic or Latinx	16 (25.0)
Decline to answer	1 (1.6)
Preferred language, n (%)	
English	42 (65.6)
Spanish	8 (12.5)
Cantonese	14 (21.9)
Education, n (%)	
High school (no degree) or less	18 (28.1)
High school degree or GED[a]	16 (25.0)
Some college or associate's degree	18 (28.1)
Bachelor's degree or higher	12 (18.8)
Own device with video capability, n (%)	
Yes	58 (90.6)
No/don't know	6 (9.4)
Preferred device(s)[b], n (%)	
Smartphone	55 (85.9)
Desktop or laptop computer	12 (18.8)
Tablet	9 (14.1)
Simple cell phone	1 (1.6)
Internet access, n (%)	
Both Wi-Fi and unlimited data on phone	39 (60.9)
Wi-Fi only	15 (23.4)
Unlimited data on phone only	7 (10.9)
Neither/don't know	3 (4.7)
Enrolled in patient portal, n (%)	
Yes, at SFHN[c]	42 (65.6)
No/don't know	22 (34.4)

[a]GED: General Educational Development.

[b]The categories are not mutually exclusive.

[c]SFHN: San Francisco Health Network.

Task Completion

Task completion rates per task are displayed in Table 2. As shown in Table 2, 46 to 62 participants were filmed while completing the 4 digital tasks. Independence in task completion was the highest for task 1 (34/46, 74%) and lowest for task 2 (31/60, 52%).

Global tests of independence were used to evaluate associations between participant characteristics and task completion across all 4 tasks. Age, race and ethnicity, preferred language,

XSL•FO
RenderX

educational attainment, access to a device with video capability, source of internet access, and enrollment in the patient portal were significantly associated with task completion. Participants who were older were more likely to require assistance or be unsuccessful across tasks (P=.004). Participants who were White were significantly more likely to complete each task independently than participants of other races and ethnicities (P<.001). Similarly, preferred language was associated with task completion, with those who preferred Spanish or Cantonese more likely to require assistance or be unsuccessful than those who preferred English (P<.001). Participants with a high school education or less were also more likely to struggle to complete the tasks independently and had higher rates of being unsuccessful or requiring assistance (P<.001). Access was also significantly associated with task completion, with those who had access to a device with video capability and Wi-Fi more likely to complete tasks independently than those without video capability (P=.05) or who relied on data to access the internet (P<.001). Individuals already enrolled in the patient portal tended to be more successful and had a higher likelihood of completing each task independently than those who were not enrolled in the patient portal or were unsure (P<.001).

XSL•FO
RenderX

Table . Task completion rates for each task by key sociodemographic characteristics.

	Task 1 (n=46)		Task 2 (n=60)		Task 3 (n=62)		Task 4 (n=61)		*P* value[c]
	Indep[a]	A/U[b]	Indep	A/U	Indep	A/U	Indep	A/U	
Participants, n (%)	34 (74)	12 (26)	31 (52)	29 (48)	44 (71)	18 (29)	43 (70)	18 (30)	
Age (years), mean (SD)	60.2 (6.49)	67.2 (7.98)	60.8 (7.68)	62.7 (5.99)	61.1 (7.11)	62.3 (5.75)	60.5 (7.21)	63.5 (5.23)	.004[d]
Male sex (%)	58.82	66.67	64.52	62.07	68.18	44.44	69.77	38.89	.06
Race and ethnicity, n (%)									.001
NH[e] White	8 (23.5)	1 (8.3)	8 (25.8)	4 (13.8)	12 (27.3)	0 (0)	12 (27.9)	0 (0)	
NH Black or African American	4 (11.8)	3 (25.0)	6 (19.4)	4 (13.8)	9 (20.5)	2 (11.1)	9 (20.9)	2 (11.1)	
NH Asian	10 (29.4)	5 (41.7)	8 (25.8)	10 (34.5)	11 (25.0)	8 (44.4)	11 (25.6)	7 (38.9)	
NH other or multiple races	3 (8.8)	0 (0)	3 (9.7)	1 (3.4)	4 (9.1)	0 (0)	4 (9.3)	0 (0)	
Hispanic or Latinx	9 (26.5)	3 (25.0)	6 (19.4)	10 (34.5)	8 (18.2)	8 (44.4)	7 (16.3)	9 (50.0)	
Preferred language, n (%)									.001
English	23 (67.6)	5 (41.7)	26 (83.9)	13 (43.3)	37 (84.1)	5 (27.8)	36 (83.7)	6 (33.3)	
Spanish	4 (11.8)	2 (16.7)	0 (0)	8 (26.7)	3 (6.8)	5 (27.8)	2 (4.7)	6 (33.3)	
Cantonese	7 (20.6)	5 (41.7)	5 (16.1)	8 (30.0)	4 (9.1)	8 (44.4)	5 (11.6)	6 (33.3)	
Education, n (%)									.001
High school degree, GED[f], or less	16 (47.1)	11 (91.7)	10 (32.3)	22 (75.9)	29 (65.9)	17 (94.4)	16 (37.2)	15 (83.3)	
Some college or higher	18 (52.9)	1 (8.3)	21 (67.7)	7 (24.1)	15 (34.1)	1 (5.6)	27 (62.8)	3 (16.7)	
Own device with video capability, n (%)									.05
Yes	34 (100)	9 (75.0)	30 (96.8)	26 (89.7)	40 (90.9)	16 (88.9)	40 (93.0)	15 (83.3)	
No/don't know	0 (0)	3 (25.0)	1 (3.2)	3 (10.3)	4 (9.1)	2 (11.1)	3 (7.0)	3 (16.7)	
Internet access, n (%)									.001[g]
Both Wi-Fi and unlimited data	23 (67.6)	5 (41.7)	19 (61.3)	18 (62.1)	28 (63.6)	10 (55.6)	28 (65.1)	9 (50.0)	
Wi-Fi only	8 (23.5)	3 (25.0)	11 (35.5)	2 (6.9)	11 (25.0)	3 (16.7)	10 (23.3)	4 (22.2)	
Unlimited data on phone only	3 (8.8)	1 (8.3)	1 (3.2)	6 (20.7)	5 (11.4)	2 (11.1)	5 (11.6)	2 (11.1)	
Neither/don't know	0 (0)	3 (25.0)	0 (0)	3 (10.3)	0 (0)	3 (16.7)	0 (0)	3 (16.7)	
Enrolled in patient portal, n (%)									.001

XSL•FO
RenderX

	Task 1 (n=46)		Task 2 (n=60)		Task 3 (n=62)		Task 4 (n=61)		*P* value[c]
	Indep[a]	A/U[b]	Indep	A/U	Indep	A/U	Indep	A/U	
Yes, at SFHN[h]	28 (82.4)	3 (25.0)	27 (87.1)	13 (44.8)	33 (75.0)	8 (44.4)	34 (79.1)	7 (38.9)	
No/don't know	6 (17.6)	9 (75.0)	4 (12.9)	16 (55.2)	11 (25.0)	10 (55.6)	9 (20.9)	11 (61.1)	

[a]Indep: independent completion of task.

[b]A/U: assisted or unsuccessful completion of task.

[c]χ^2 test was used to assess differences by sociodemographic characteristic, unless otherwise specified; $P \leq .05$ was considered significant.

[d]Kruskal-Wallis test was used to assess differences by age; $P \leq .05$ was considered significant.

[e]NH: non-Hispanic.

[f]GED: General Educational Development.

[g]Fisher exact test was used to assess differences by internet access; $P \leq .05$ was considered significant.

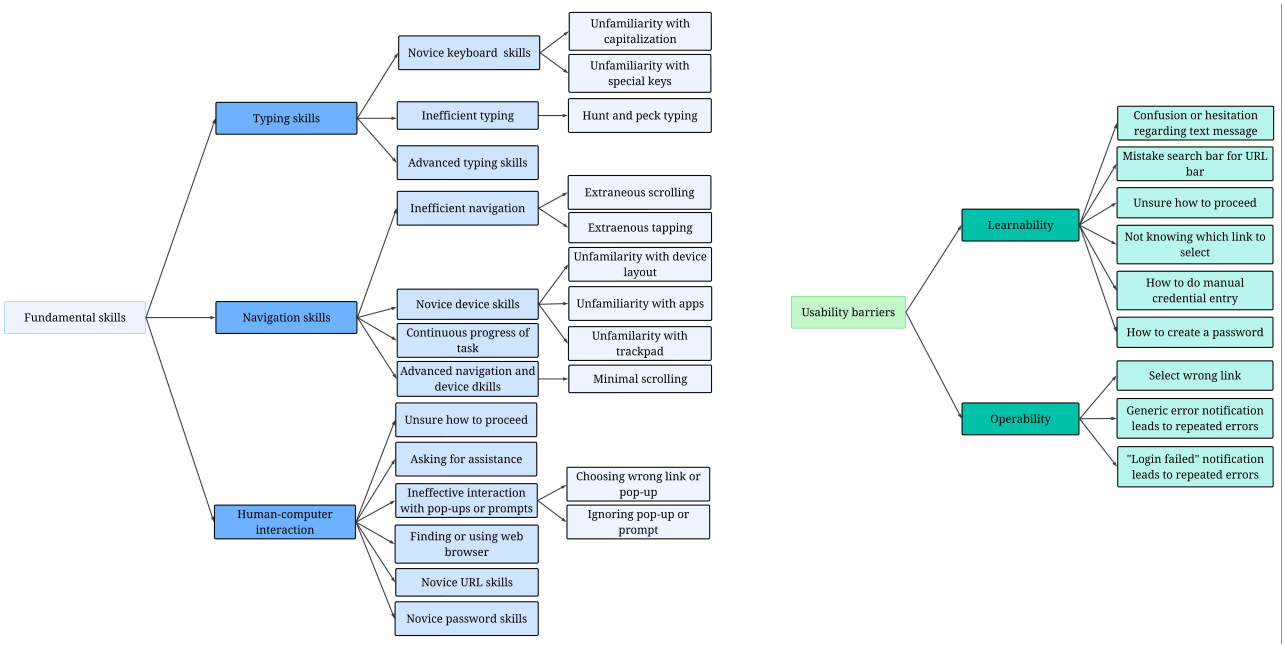
[h]SFHN: San Francisco Health Network.

Video Observations

Thematic Coding

The following findings were based on 91% agreement between the 3 coders for the first 15 participants and independent coding of the remaining videos. Thematic findings were grouped into factors, mapped to the overall dimensions of fundamental skills and usability barriers, and reorganized for interpretation. To ease analysis, we decided to separate the codes representing usability barriers from those representing fundamental skills, with several codes representing skills and usability barriers. Figure 1 shows an overview of themes generated during the observational coding process, with subcategories within each key theme, and separated by fundamental skills and usability barriers.

Figure 1. Qualitative thematic maps for fundamental skills and usability barriers.



Fundamental Skills

Video observations revealed that participants exhibited a range of digital skills from novice to advanced. The coding matrix categorized participant behaviors into key themes, including typing skills, navigation skills, and human-computer interaction. Typing skills referred to a participant's ability to navigate the keyboard (particularly capitalizing letters and finding special keys) and demonstration of inefficient typing ("hunting" and searching for letters) or advanced typing skills (proficiency with typing, capitalization, and special keys). Navigation skills comprised inefficient navigation (eg, extraneous scrolling or tapping on the device), novice device skills (eg, unfamiliarity with device layout or apps), advanced navigation and device skills (often indicated by minimal scrolling), and continuous progress of the task. Human-computer interaction (HCI) encompassed codes that demonstrated the way participants engaged or did not engage with their devices to complete the tasks, ranging from how they responded to prompts; whether they were unsure how to proceed, requested assistance, or ineffectively interacted with pop-ups or prompts; to having

XSL•FO
RenderX

trouble with specific elements of the tasks (eg, novice URL skills, trouble finding or using a web browser, or novice password skills).

A significant proportion of participants exhibited novice-level digital skills, particularly in typing and navigation. Many participants relied on inefficient typing techniques, such as hunt-and-peck typing and relying on using one finger or a stylus to locate and enter subsequent characters. The inability to appropriately capitalize letters or locate and select special characters created significant barriers to completing the tasks independently. Similarly, many participants demonstrated navigation inefficiencies. Participants exhibited trouble navigating their device with extraneous and repeated tapping or scrolling, confusion locating specific apps or phone settings, or difficulty using a touchscreen or trackpad. In contrast, participants who demonstrated success at navigating the tasks independently almost exclusively displayed advanced navigation skills and the ability to progress through the tasks continuously without breaks in performance flow.

Many individuals also struggled with more complex HCI actions beyond simply navigating their device or using the keyboard. Common observations included participants not understanding how to proceed through the task because they struggled with fundamental concepts such as recognizing common digital icons (eg, settings icon, “close window” button, home icon, or back/forward arrows) or clicking the “next” button, having trouble finding or navigating to a web browser (or conceptually understanding what it is), ineffectively interacting with task pop-ups or prompts (eg, ignoring a pop-up/prompt or choosing the wrong link), and difficulty using the web browser (eg, not knowing how to select the web browser icon to open a new window or how to proceed once the web browser was open). Individuals also demonstrated novice URL and password skills, as they struggled with accurate text entry to successfully type in a URL for a website or enter the provided password. Difficulty performing complex tasks despite receiving instructions was a common reason why participants asked for assistance, often demonstrated even before they attempted the task or after several unsuccessful attempts to complete an action.

Table 3 presents the association between typing, navigation, and HCI codes and task performance. Most notably, across tasks, keyboard skills, navigation skills, and HCI skills showed a significant association between the factor and the ability to complete each task independently. Individuals who demonstrated unfamiliarity with navigating the keyboard, particularly with how to capitalize letters or find special characters, always required assistance or were unsuccessful at completing the digital tasks (P<.001 in all cases). Inefficient navigation with extraneous scrolling or tapping of their device was strongly associated with needing assistance or unsuccessfully completing the task, particularly in task 1 (P=.001), task 2 (P<.001), and task 4 (P=.006), which required more complex device navigation to complete the task. Novice device skills, demonstrated by lack of familiarity with one’s device, whether it be the layout, finding and using apps, or navigating the trackpad, were significantly associated with requiring assistance or unsuccessfully completing the task across tasks (task 1, P<.001; task 2, P<.001; task 3, P=.004; task 4, P=.002). Typing skills (both inefficient typing and advanced typing skills) were not significantly associated with digital task completion (exact P values given in Table 3). However, we did not assess speed to completion.

The complexity of the digital tasks required several fundamental human-computer interaction skills, beyond navigating one’s device, to complete each task. Displaying confusion or hesitation with how to proceed during a task was a significant barrier for individuals during each task, ranging from not being sure how to proceed to confusion about how to navigate barriers that arise or where to put information (tasks 1, 2, and 4, P<.001; task 3, P=.02). Individuals also often asked for assistance when stuck on a task, demonstrating the complexity of progressing through a task as a barrier to independently completing a task (P<.001 in all cases). In task 2, the ability to find and use a web browser appropriately and the ability to type in a URL correctly were also significantly associated with being unsuccessful or requiring assistance to complete the task (P<.001 for both). In tasks 3 and 4, having trouble with creating a username or password (eg, not sure how to create one) was associated with needing assistance or unsuccessfully completing the task (task 3, P=.004; task 4, P=.006).

XSL•FO
RenderX

Table . Association between thematic fundamental skills codes and task completion.

	Task number	Independent completion[a], n	Assisted/unsuccessful completion[a], n	*P* value
Typing skills				
Inefficient typing (tasks 2 - 4)	2	24	24	>.99
	3	25	17	<.001[b]
	4	28	15	.26
Novice keyboard skills (tasks 2 - 4)	2	0	10	<.001[b]
	3	0	5	<.001[b]
	4	0	7	<.001[b]
Advanced typing skills (tasks 2 - 4)	2	5	0	.05[b]
	3	7	0	.17[b]
	4	8	0	.09[b]
Navigation skills				
Inefficient navigation	1	4	8	.001[b]
	2	4	23	<.001
	3	11	9	<.001[b]
	4	11	12	.006
Novice device skills	1	2	10	<.001[b]
	2	2	21	<.001
	3	2	6	.004[b]
	4	2	7	.002[b]
Advanced navigation and device skills	1	28	1	<.001[b]
	2	29	1	<.001
	3	35	4	<.001[b]
	4	33	2	<.001[b]
Continuous progress of task	1	22	0	<.001
	2	24	0	<.001
	3	31	0	<.001
	4	23	0	<.001
Human-computer interaction				
Unsure how to proceed	1	7	12	<.001
	2	0	15	<.001
	3	4	6	.02[b]
	4	5	11	<.001[b]
Asking for assistance	1	3	12	<.001[b]
	2	1	19	<.001
	3	7	13	<.001
	4	8	13	<.001

XSL•FO
RenderX

	Task number	Independent completion[a], n	Assisted/unsuccessful completion[a], n	*P* value
Ineffective interaction with pop-ups or prompts	1	3	15	.006
	2	4	11	.04[b]
	3	8	11	.002
	4	12	13	<.001[b]
Finding or using web browser (task 2 only)	2	0	13	<.001
Novice URL skills (task 2 only)	2	3	14	<.001
Novice password skills (tasks 3 and 4 only)	3	5	8	.004[b]
	4	13	13	.006

[a]The independent completion and assisted/unassisted completion columns show the number of participants displaying each qualitative code per task.
[b]Indicates Fisher exact test; χ^2 tests of independence used in all other instances. $P \leq .05$ was considered significant.

Usability Barriers

Video observations also revealed that task completion was hindered by several usability barriers. The key usability barriers across all the tasks were related to learnability and operability. Tables 4 and 5 summarize the learnability and operability issues for each task. In task 1, joining a video visit from a text message and unfamiliarity with system-generated notifications led to hesitation or errors in navigating the log-in prompt (characterized by the codes “confusion/hesitation regarding text messages” and “unsure how to proceed”), illustrating learnability barriers to joining video visits, despite the instructions in the text. In addition, individuals often did not select the video link, instead opting to choose an alternative method, such as replying to the text (despite it being a no-reply message) or attempting to join the video visit unsuccessfully. Additionally, once the participant successfully selected the link in the text message, additional pop-ups and links necessary to join the video call confused the participant, resulting in those pop-ups and links being ignored or ineffectively addressed. The operability limitation of this task was that the system did not provide clear enough cues for individuals to correctly select the link to join the video visit; therefore, individuals often did not know which link to select and instead selected the wrong link, either delaying their ability to join a video visit or preventing it altogether.

XSL•FO
RenderX

Table . Examples of learnability usability barriers by task.

Task	Thematic code	Example	Illustrative quote
1: sign in to a video visit	• Confusion or hesitation regarding text messages • Unsure how to proceed • Not knowing which link to select	• Clicking on a video call link from a text message requires a level of understanding (to click the link) that users might not have • Results in hesitating to click the link, ignoring the text, or replying to a no-reply text • Individual ignores necessary prompts for joining a video call or selects other incorrect links	• "Ok so I have to copy that link and put it on my phone?" • "So, I open it, I guess?" • "Ok but what does…main menu? What does that look like?"
2: go to a specific website	• Mistake search bar for URL bar • How to do manual credential entry • Unsure how to proceed • Not knowing which link to select	• Individual types the URL into the search bar instead of the URL bar • Individual is not sure how to proceed once they type the URL in • Individual does not select the appropriate link or button; instead, they select elsewhere	• "I don't know how to get to a website." • "If I put it up here, it doesn't come down here." [in reference to search bar] • "Now I'm stuck."
3: sign in to the patient portal	• How to do manual credential entry • Unsure how to proceed	• Individual does not display the basic level of knowledge for password or username duplication (copying the given password/username into account) • Individual stops after typing in password, without clicking enter to continue	• "Oh, it says Patient, I'm the patient!" [in response to the username being "Patient123" after entering only "123"]
4: sign up for a patient portal account	• How to create a password • Unsure how to proceed	• Individual does not display the basic level of knowledge necessary to create a password • Individual does not know how to proceed once they type in a password or username	• "Did it accept? Did it accept, or not?" • "I don't know what you mean by that." [in response to being instructed to capitalize a letter] • "Is that the password? Nothing? I have to create? Invent something?"

XSL•FO
RenderX

Table . Examples of operability usability barriers by task.

Task	Thematic code	Example	Illustrative quote
1: sign in to a video visit	• Selecting wrong link	• Individual clicks on a wrong link to join the video call	—[a]
2: go to a specific website	• Generic error notification leads to repeated errors	• Complexity and length of the URL results in errors difficult to discern and correct	• [Participant mistakenly typed a forward slash instead of a backslash]
3: sign in to the patient portal	• Sign-in process difficult to remember or master • Generic error notification leads to repeated errors	• “Login failed” notification does not indicate the error, resulting in difficulty correcting username or password errors	• "Even if I signed in now, after I walk out the door I won't know how to do it again." • "Those are the little things that make a difference." [when noticing that "s" in the password “AppleSauce” needs capitlizing]
4: sign up for a patient portal account	• Activation code disrupts sign-up process • Generic error notification leads to repeated errors	• Activation code causes confusion and leads to error making • “Login failed” notification does not indicate the error, resulting in difficulty correcting username or password errors	• “I don’t know [where the activation code is].” • “Where do I put [the activation code], right here?” • "I'm confused.” [when unable to quickly identify error]

[a]Not applicable.

Task 2 was the most difficult for participants, with learnability and operability barriers affecting the ability to complete the task of signing in to the patient portal on a web interface. Many participants displayed misinterpretation of browser suggestions and confusion regarding where to type in the URL (coded as “mistake search bar for URL bar”) and the extent to which the URL needs to be complete (categorized as “how to do manual credential entry;” many stopped before typing in the complete URL or mistakenly used autofill to create a false URL), illustrating the learnability barriers associated with this task. In addition, individuals were often unsure how to proceed, sometimes selecting irrelevant links or failing to recognize the correct link to continue to the patient portal website. Typing in a URL with multiple characters and capitalizations also proved to be a user burden, as many individuals made small mistakes that led to repeated failures, without indications of where the error was, illustrating the operability barrier in URL design. These mistakes included typos in the URL or using autofill to automatically complete the URL, despite the autofill not matching the provided URL.

Participants struggled with the password aspects of tasks 3 and 4, further exposing learnability and operability barriers. They often misunderstood the login prompts, illustrating conceptual clarity issues in password duplication or creation for digital tools (categorized as “how to do manual credential entry” and “how to create a password”). The knowledge required to understand and replicate the complexity required for password creation is too steep, resulting in learnability barriers for many users who do not possess the basic level of knowledge necessary to navigate username and password duplication or creation. Additionally, in task 4, once individuals created a password, they often did not know what to do next to proceed, illustrating the learnability issues of progressing successfully through the steps required to sign in or sign up for a patient portal account.

The operability limitation of these tasks was notifications that did not adequately notify the participant of their error, leading to repeated errors (categorized as “generic error notification leads to repeated errors” and “‘login failed’ notification leads to repeated errors”). For example, the activation codes were confusing and led to errors among participants without proper indication. When prompted that an activation code was needed, many did not know what it was, where to find it, or what to do with it. If they mistyped the activation code but not the username or password, they would receive only a “login failed” notification that did not properly indicate where the error occurred.

Discussion

Principal Results

Our study identified the fundamental skills necessary to be able to effectively navigate digital health tasks, found significant disparities in those skills among participants, and established usability issues associated with navigating digital health tools.

Although nearly universal access to digital devices has resulted in increased device ownership and personal use, only 53.1% (34/64) of participants were able to complete more than half of the tasks independently. Notably, 15.6% (10/64) were unable to complete any tasks independently, indicating a discrepancy between device ownership and ability to navigate these devices effectively, particularly for digital tasks related to health care. These data are supported by prior work assessing disparities in digital health tool use [9,11,43].

We expected that typing skills, navigation skills, and HCI skills would be associated with the ability to complete the digital tasks independently, but it was found that only a subset of these

XSL•FO
RenderX

factors were associated with task completion, which may be limited by our sample size.

Challenges in Task Completion

Nearly half of the participants were able to complete each task independently, with 56.3% (36/64) able to complete tasks 1, 3, and 4 independently. The task of typing in a URL in task 2 posed unique barriers and resulted in more participants requiring assistance or being unsuccessful. This is potentially explained by the fact that typing in a complete URL is increasingly uncommon in digital health contexts, where users typically rely on search engines (eg, searching "SFHN MyChart patient portal" in the search bar), use autocomplete based on browsing history, or access the portal through an app. This is further supported by the fact that many participants struggled to recognize the URL bar or conceptually understand what was being asked of them during task 2. Despite this, many informational handouts about the patient portal continue to direct users to type in a specific URL to access the patient portal. These materials should consider providing alternative instructions that align with common user behavior, such as using search terms or app-based access.

Disparities in Task Completion

We anticipated that there would be disparities in task completion based on age, sex, race and ethnicity, preferred language, education, access to a device with video capability, source of internet access, and patient portal enrollment. Older age, minoritized races and ethnicities, preferred languages other than English, low educational attainment, possession of a device without video capability, data as the source of internet access, and not being enrolled in the patient portal contributed significantly to gaps in digital skills. These sociodemographic characteristics were significantly associated with requiring assistance or being unsuccessful, particularly for tasks that involved multiple steps or complex engagement with digital interfaces. This suggests that individuals from these groups may be disproportionately less likely to have the skills required to feel confident and able to complete each task independently. Moreover, tasks 3 and 4 were more complex, involving reading comprehension and following multiple prompts to enter pertinent text or create a unique password. These elements (such as log-in prompts and password creation details) were in English only, which posed obvious barriers to individuals who preferred non-English languages.

Although this study is not sufficiently powered to examine the intersectional vulnerabilities of participants who experience multiple, overlapping socioeconomic disadvantages—factors that may confound their ability to complete digital tasks independently—there remains a critical need for future research to adopt an inclusive approach to participant recruitment and analysis in usability studies. Specifically, examining how compounded forms of marginalization (eg, low income, limited education, language barriers, and age-related factors) shape digital engagement is essential for uncovering the full extent of challenges faced by individuals most at risk of digital exclusion. Such efforts are key to informing the design of equitable digital health tools and implementation strategies that promote broader and more effective adoption across diverse populations [17,21,22,33,44].

Challenges in Digital Proficiency

These results illustrate that the basic foundational skills necessary for successful navigation and use of digital health tools include competency with one's device, such as knowing basic keyboard navigation and how to find appropriate apps or change settings, being able to navigate complex processes such as interacting with pop-ups and successfully navigating web browsers, and being able to proceed after each step. Without these basic foundational skills, an individual's ability to successfully use and navigate digital health tools decreases significantly. Strategies for improving digital skills to ensure that individuals have the necessary foundational skills include implementing digital competency screening tools and providing digital skills workshops and trainings that emphasize hands-on practice with common tasks such as manual text entry, device and keyboard navigation, managing settings, and interacting with complex pop-ups or prompts. Prior work from our team has demonstrated that the brief, 9-item Skills Measurement and Readiness Training for Digital Health (SMART Digital Health) Scale is reliable for measuring skills for digital health-related tasks and activities and predicting the level of support an individual may need [23].

In addition, prior studies have found that providing support such as technical support or digital education alleviates some of the digital proficiency challenges faced by many and promotes the use of digital health tools, particularly among those who face greater barriers to digital use [45-47]. In addition to providing detailed assistance and training to address the complex barriers faced by those with limited digital literacy, ongoing or repeated support may be necessary, as one-time training or technical help may not be sufficient to overcome persistent technical challenges [47].

Usability Barriers May Impact Digital Task Success

The findings highlight the critical role of learnability and operability challenges across a range of digital health tasks commonly required for digital patient engagement. Participants encountered distinct barriers during each task, verbalizing frustration and making repeated assistance requests. Many of the difficulties were often tied to misunderstandings about how to initiate or complete a task, suggesting that digital inclusion efforts should simplify interfaces, provide clear prompts, and reduce extraneous steps to mitigate the usability issues faced by many participants.

Learnability issues can lead to a range of negative user experiences, including hesitation and confusion, a greater need for assistance and additional support to complete each task, and low user confidence. Each task required a level of understanding that users might not have to complete the task, resulting in a system that does not equally support users of all backgrounds to adapt and use the system. In addition, there were operability barriers across each task. Particularly, difficulties such as clicking the wrong link, making errors without descriptive or specific notifications, or being confused about how to proceed during a task result in a system that does not provide users with

XSL•FO
RenderX

the necessary functionalities to allow users of all backgrounds to adapt and use the system. Efforts to improve the clarity of digital task prompts and processes, ranging from simplifying the URL to log in to the patient portal to clarifying how to successfully join a video visit from a text (or link) or how to engage with certain prompts, will enhance system learnability and operability and allow users to better engage with the system and remember how to use the system going forward.

These findings are consistent with prior research identifying digital navigation and account access as common friction points for users with limited digital literacy [9,11]. However, our human-centered design (HCD) approach to this study adds nuance by documenting how these challenges manifest in real time and vary by task complexity. Previous literature often examines digital competence in aggregate, but our task-specific analysis reveals that even users with some experience may struggle with unfamiliar or multistep processes, particularly when there are long task completion times or when confusing interfaces or prompts lead to repeated errors [48,49]. As adoption of digital health tools continues to increase, there are opportunities to use HCD approaches to identify usability barriers and align digital health tool design with user needs to ensure a more equitable distribution of digital health tool benefits [32]. This is particularly important, as there are numerous examples in which digital tools exacerbate inequities. HCD approaches can help uncover previously unknown health equity issues and drive efforts to address those inequities more acutely experienced by marginalized populations that are often not considered priorities in research [28-30,32,33].

Vendors Need to Prioritize Co-Design to Address Usability Barriers

The findings from this study underscore the pressing need for digital health vendors to take a more active role in addressing usability barriers and skill gaps through cocreation. While health systems increasingly rely on digital platforms to facilitate patient engagement, results demonstrate that usability challenges and complex design, often resulting from a failure to cocreate with users, are common friction points that inhibit successful digital health tool use. These barriers are not merely technical inconveniences; they are structural barriers that prevent users from accessing care and support. Prior studies have similarly identified usability challenges—such as manually entering data, following procedures to retrieve data (such as appointment details), and understanding medical terminology—as key barriers that prevent equitable use of patient portals among diverse populations [49,50].

To address usability challenges, vendors must adopt HCD principles and co-design strategies that engage diverse users throughout the design process to ensure that digital interfaces are intuitive, accessible, and responsive to users' varying levels of digital proficiency. This includes incorporating plain language, reducing unnecessary steps, improving language functionalities for non-English speakers, and reducing privacy and security complexities that act as a barrier to users with low digital literacy [50]. Vendors must not assume a baseline level of digital literacy and should instead actively test for, design against, and consult patients about learnability, operability, and task complexity barriers in these digital health tools. By treating usability and accessibility as core priorities rather than downstream implementation concerns, vendors can play a pivotal role in promoting digital equity and closing gaps in health care access. For example, strict security needs for usernames and passwords (including activation codes and other 2-factor authentication) are not "user-friendly," and while they make sense from a security perspective, they are advanced skills that many users do not understand or have experience with. Issues with complex password creation, entering an activation code, and other 2-factor authentication methods are demonstrated tension points for users from marginalized backgrounds who may have little experience in engaging with strict and complex privacy protocols. Vendors must include patient voices when weighing privacy concerns against usability to ensure that access and usability decisions do not further inequities [28,50].

In the meantime, without more usable products, health care systems need to be prepared to support patients in accessing and using existing digital health tools. As prior studies have shown, explicit identification of a "digital navigator" to support patients in using digital health tools facilitates equitable access and use of digital health tools [51,52]. Unfortunately, this technical assistance often falls to clinical care team members who are not trained to provide this type of support; as a result, patients do not receive adequate support, particularly in safety net systems [52,53]. Increasing support for digital navigator training and incorporating them as members of the clinical care team will be crucial to help overcome gaps in foundational skills and usability barriers.

Limitations

We note several limitations of our study. We recruited participants from a large urban safety net health system that serves a diverse population, but our stratified samples based on core sociodemographic characteristics are small, limiting our ability to evaluate disparities in fundamental skills by sociodemographic characteristics. Additionally, the Hawthorne effect may have resulted in participants changing their behavior while completing the digital tasks because they were being observed. While the chosen tasks are essential in accessing and using digital health technology, they may not be representative of specific usability barriers that individuals face within the patient portal app itself (such as scheduling an appointment) or may be considered by some as antiquated processes (such as using a URL to access the patient portal). We used a patient portal emulator for this study, which was a web browser–based process; this may not reflect the same usability barriers that would exist for the use of a smartphone app. However, the fundamental skill challenges that were present for many patients would impact the use of a mobile app. Moreover, despite advances in digital health tool accessibility (translation to multiple languages and apps available on most devices), there are still limitations to the ability of digital health tools to be inclusive, resulting in issues that are not strictly due to lack of skills but are also influenced by language barriers or other sociodemographic barriers. However, our results will still provide further direction for future digital inclusion studies.

XSL•FO
RenderX

Implications for Future Research

Future studies should more deeply examine usability issues of digital health tools, evaluating the tools themselves and not just tasks that are related to these tools. Future studies should also explore longitudinal improvements in digital skills following structured training interventions. Evidence has shown that provision of multimodal and audiovisual digital training is supported by patients and has been associated with improved use [45,46]. This evidence suggests that training among populations with marginalized identities can be beneficial, yet evidence is needed to assess training implications on digital skills. Additionally, assessing the impact of different instructional approaches—such as interactive tutorials versus guided assistance—could yield insights into optimizing digital literacy education.

Furthermore, there are benefits of improving digital literacy broadly, not only specific to digital health tools and digital health engagement. This study demonstrated that there are disparities in foundational skills that are not specific to digital health tools; therefore, designing collaborations with libraries or continuing education classes could serve as implementation strategies that can help mitigate the effects of these skill disparities and reduce digital exclusion.

Conclusion

By integrating the results of the fundamental skills necessary to use digital health tools and usability barriers to using these tools, this study highlights critical gaps in digital literacy, particularly in device navigation and progressing through complex tasks. This study underscores the importance of usability-driven improvements to reduce digital barriers. By addressing these challenges through targeted interventions, digital accessibility may be enhanced, empowering users to navigate digital environments more effectively. Future work should explore user-tailored interventions aimed at providing individuals with the core fundamental skills necessary to use digital health tools, iterative usability testing to improve the usability of these tools, and personalized support mechanisms to support individuals as they navigate these tools. By systematically identifying and addressing these usability challenges, researchers and developers can create more inclusive digital health solutions that promote equitable health care access and improve patient outcomes across all populations.

Funding

EK received support from the National Heart Lung and Blood Institute (K23HL157750), the National Institute on Aging of the National Institutes of Health under award number P30AG015272, and the UCSF Population Health and Health Equity Program.

Data Availability

The dataset generated and analyzed during this study cannot be shared publicly due to the nature of the data. A subset of data that can be deidentified can be shared upon reasonable request to the corresponding author, subject to institutional review board approval and applicable data governance requirements.

Authors' Contributions

Conceptualization: TR (lead), CRL (equal), EK (equal)
Data curation: TR (lead), HCK (equal)
Formal analysis: TR
Funding acquisition: EK
Investigation: TR, SH, JW, HCK, IL, AY, JL, EK, CRL
Methodology: TR (lead), MR (equal), CRL (supporting), EK (supporting)
Project administration: HCK (supporting), EK (lead)
Supervision: TR (supporting), MR (supporting), PAF (supporting), JDR (supporting), CRL (supporting), EK (lead)
Validation: TR
Visualization: TR
Writing – original draft: TR
Writing – review & editing: TR, MR, SH, JW, HCK, IL, AY, JL, PAF, JDR, CRL, EK

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey administered to participants.
[DOCX File, 35 KB - humanfactors_v13i1e78430_app1.docx]

Multimedia Appendix 2
Description of digital tasks and the coding matrix.
[DOCX File, 30 KB - humanfactors_v13i1e78430_app2.docx]

XSL•FO
RenderX

Multimedia Appendix 3
Qualitative codebook.
[DOCX File, 36 KB - humanfactors_v13i1e78430_app3.docx]

References

1. Grossman LV, Masterson Creber RM, Benda NC, Wright D, Vawdrey DK, Ancker JS. Interventions to increase patient portal use in vulnerable populations: a systematic review. J Am Med Inform Assoc 2019 Aug 1;26(8-9):855-870. [doi: 10.1093/jamia/ocz023] [Medline: 30958532]
2. Lyles C, Schillinger D, Sarkar U. Connecting the dots: health information technology expansion and health disparities. PLoS Med 2015 Jul;12(7):e1001852. [doi: 10.1371/journal.pmed.1001852] [Medline: 26172977]
3. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015 Feb 10;17(2):e44. [doi: 10.2196/jmir.3171] [Medline: 25669240]
4. Robinson SA, Netherton D, Zocchi M, Purington C, Ash AS, Shimada SL. Differences in secure messaging, self-management, and glycemic control between rural and urban patients: secondary data analysis. JMIR Diabetes 2021 Nov 19;6(4):e32320. [doi: 10.2196/32320] [Medline: 34807834]
5. Brands MR, Gouw SC, Beestrum M, Cronin RM, Fijnvandraat K, Badawy SM. Patient-centered digital health records and their effects on health outcomes: systematic review. J Med Internet Res 2022 Dec 22;24(12):e43086. [doi: 10.2196/43086] [Medline: 36548034]
6. Bretschneider MP, Kolasińska AB, Šomvárska L, Klásek J, Mareš J, Schwarz PE. Evaluation of the impact of Mobile health app vitadio in patients with type 2 diabetes: randomized controlled trial. J Med Internet Res 2025 May 9;27:e68648. [doi: 10.2196/68648] [Medline: 40344662]
7. Fuster-Casanovas A, Melero SM, Cabutto D, et al. Exploring digital health tools for depression management in primary health care: systematic review and meta-analysis. J Affect Disord 2025 Jul 15;381:494-506. [doi: 10.1016/j.jad.2025.03.187] [Medline: 40203968]
8. Ancker JS, Hafeez B, Kaushal R. Socioeconomic disparities in adoption of personal health records over time. Am J Manag Care 2016 Aug;22(8):539-540. [Medline: 27541700]
9. Khoong EC, Rivadeneira NA, Hiatt RA, Sarkar U. The use of technology for communicating with clinicians or seeking health information in a multilingual urban cohort: cross-sectional survey. J Med Internet Res 2020 Apr 6;22(4):e16951. [doi: 10.2196/16951] [Medline: 32250280]
10. Sarkar U, Karter AJ, Liu JY, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 1;18(3):318-321. [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]
11. Tieu L, Schillinger D, Sarkar U, et al. Online patient websites for electronic health record access among vulnerable populations: portals to nowhere? J Am Med Inform Assoc 2017 Apr 1;24(e1):e47-e54. [doi: 10.1093/jamia/ocw098] [Medline: 27402138]
12. Gibbons MC. A historical overview of health disparities and the potential of eHealth solutions. J Med Internet Res 2005 Oct 4;7(5):e50. [doi: 10.2196/jmir.7.5.e50] [Medline: 16403714]
13. Eberly LA, Kallan MJ, Julien HM, et al. Patient characteristics associated with telemedicine access for primary and specialty ambulatory care during the COVID-19 pandemic. JAMA Netw Open 2020 Dec 1;3(12):e2031640. [doi: 10.1001/jamanetworkopen.2020.31640] [Medline: 33372974]
14. Ryskina KL, Shultz K, Zhou Y, Lautenbach G, Brown RT. Older adults' access to primary care: gender, racial, and ethnic disparities in telemedicine. J Am Geriatr Soc 2021 Oct;69(10):2732-2740. [doi: 10.1111/jgs.17354] [Medline: 34224577]
15. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015 Feb 20;17(2):e40. [doi: 10.2196/jmir.3703] [Medline: 25707035]
16. McClain C, Bishop W. What we know about internet use, smartphone ownership and digital divides in the US. Pew Research Center. 2026 Jan 8. URL: https://www.pewresearch.org/short-reads/2026/01/08/internet-use-smartphone-ownership-digital-divides-in-u-s/ [accessed 2026-02-17]
17. Alkureishi MA, Choo ZY, Rahman A, et al. Digitally disconnected: qualitative study of patient perspectives on the digital divide and potential solutions. JMIR Hum Factors 2021 Dec 15;8(4):e33364. [doi: 10.2196/33364] [Medline: 34705664]
18. Lyles CR, Nelson EC, Frampton S, Dykes PC, Cemballi AG, Sarkar U. Using electronic health record portals to improve patient engagement: research priorities and best practices. Ann Intern Med 2020 Jun 2;172(11 Suppl):S123-S129. [doi: 10.7326/M19-0876] [Medline: 32479176]
19. Jiang JX, Ross JS, Bai G. Unveiling the adoption and barriers of telemedicine in US hospitals: a comprehensive analysis (2017–2022). J Gen Intern Med 2024 Oct;39(13):2438-2445. [doi: 10.1007/s11606-024-08853-0] [Medline: 38985409]
20. McGinty EE, Alegria M, Beidas RS, et al. The Lancet Psychiatry Commission: transforming mental health implementation research. Lancet Psychiatry 2024 May;11(5):368-396. [doi: 10.1016/S2215-0366(24)00040-3] [Medline: 38552663]
21. Richardson S, Lawrence K, Schoenthaler AM, Mann D. A framework for digital health equity. NPJ Digit Med 2022 Aug 18;5(1):119. [doi: 10.1038/s41746-022-00663-0] [Medline: 35982146]

XSL•FO
RenderX

22. Rodriguez JA, Clark CR, Bates DW. Digital health equity as a necessity in the 21st century cures act era. JAMA 2020 Jun 16;323(23):2381-2382. [doi: 10.1001/jama.2020.7858] [Medline: 32463421]
23. Tieu L, Lyles CR, Kim HC, et al. A self-report measure of digital skills needed to use digital health tools among older adults-the Skills Measurement and Readiness Training for Digital Health (SMART Digital Health) Scale. J Am Med Inform Assoc 2025 Nov 1;32(11):1674-1684. [doi: 10.1093/jamia/ocaf151] [Medline: 40971609]
24. Choi J, Choi S, Song K, et al. Everyday digital literacy questionnaire for older adults: instrument development and validation study. J Med Internet Res 2023 Dec 14;25:e51616. [doi: 10.2196/51616] [Medline: 38095999]
25. Rising KL, Guth A, Gentsch AT, et al. Development and preliminary validation of a screener for digital health readiness. JAMA Netw Open 2024 Sep 3;7(9):e2432718. [doi: 10.1001/jamanetworkopen.2024.32718] [Medline: 39254972]
26. Nelson LA, Pennings JS, Sommer EC, Popescu F, Barkin SL. A 3-item measure of digital health care literacy: development and validation study. JMIR Form Res 2022 Apr 29;6(4):e36043. [doi: 10.2196/36043] [Medline: 35486413]
27. van Deursen A, Helsper EJ, Eynon R. Measuring digital skills: from digital skills to tangible outcomes project report. : University of Twente; 2014 URL: www.oii.ox.ac.uk/research/projects/?id=112 [accessed 2026-04-16]
28. Evans L, Evans J, Pagliari C, Källander K. Scoping review: exploring the equity impact of current digital health design practices. Oxf Open Digit Health 2023 Jan 1;1:qad006. [doi: 10.1093/oodh/oqad006]
29. Ancker JS, Barrón Y, Rockoff ML, et al. Use of an electronic patient portal among disadvantaged populations. J Gen Intern Med 2011 Oct;26(10):1117-1123. [doi: 10.1007/s11606-011-1749-y] [Medline: 21647748]
30. Antonio MG, Petrovskaya O, Lau F. Is research on patient portals attuned to health equity? A scoping review. J Am Med Inform Assoc 2019 Aug 1;26(8-9):871-883. [doi: 10.1093/jamia/ocz054] [Medline: 31066893]
31. Lyles CR, Nguyen OK, Khoong EC, Aguilera A, Sarkar U. Multilevel determinants of digital health equity: a literature synthesis to advance the field. Annu Rev Public Health 2023 Apr 3;44:383-405. [doi: 10.1146/annurev-publhealth-071521-023913] [Medline: 36525960]
32. Lyon AR, Aung T, Bruzios KE, Munson S. Human-centered design to enhance implementation and impact in health. Annu Rev Public Health 2025 Apr;46(1):467-485. [doi: 10.1146/annurev-publhealth-071823-122337] [Medline: 39656968]
33. Veinot TC, Ancker JS, Cole-Lewis H, et al. Leveling up: on the potential of upstream health informatics interventions to enhance health equity. Med Care 2019 Jun;57(Suppl 6 Suppl 2):S108-S114. [doi: 10.1097/MLR.0000000000001032] [Medline: 31095048]
34. Khoong EC, Butler BA, Mesina O, et al. Patient interest in and barriers to telemedicine video visits in a multilingual urban safety-net system. J Am Med Inform Assoc 2021 Feb 15;28(2):349-353. [doi: 10.1093/jamia/ocaa234] [Medline: 33164063]
35. San Francisco Department of Public Health annual report 2022-2023. : San Francisco Department of Public Health; 2023 URL: https://www.sf.gov/sites/default/files/2024-02/FY22-23%20DPH%20Annual%20Report_0.pdf [accessed 2026-04-16]
36. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13(1):117. [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
37. Klingberg S, Stalmeijer RE, Varpio L. Using framework analysis methods for qualitative research: AMEE Guide No. 164. Med Teach 2024 May;46(5):603-610. [doi: 10.1080/0142159X.2023.2259073] [Medline: 37734451]
38. Alonso-Ríos D, Vázquez-García A, Mosqueira-Rey E, Moret-Bonillo V. Usability: a critical analysis and a taxonomy. Int J Hum Comput Interact 2009 Dec 28;26(1):53-74. [doi: 10.1080/10447310903025552]
39. Khoong E. Digital inclusion, access, and literacy. : OpenICPSR (Inter-University Consortium for Political and Social Research [Distributor]); 2026 Feb 9 URL: https://www.openicpsr.org/openicpsr/project/233082/version/V7/view?path=/openicpsr/233082/fcr:versions/V7/2.-Chinese-Assistance-Tools/Chinese_Handout_Task2_VisitHealthWebsite.docx&type=file [accessed 2025-06-18] [doi: 10.3886/E233082V7]
40. Creswell JW, Plano Clark VL. Designing and Conducting Mixed Methods Research, 3rd edition: Sage Publications; 2017. URL: https://collegepublishing.sagepub.com/products/designing-and-conducting-mixed-methods-research-3-241842 [accessed 2026-04-16]
41. Klassen AC, Creswell J, Plano Clark VL, Smith KC, Meissner HI. Best practices in mixed methods for quality of life research. Qual Life Res 2012 Apr;21(3):377-380. [doi: 10.1007/s11136-012-0122-x] [Medline: 22311251]
42. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs—principles and practices. Health Serv Res 2013 Dec;48(6 Pt 2):2134-2156. [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
43. Taha J, Sharit J, Czaja SJ. The impact of numeracy ability and technology skills on older adults' performance of health management tasks using a patient portal. J Appl Gerontol 2014 Jun;33(4):416-436. [doi: 10.1177/0733464812447283] [Medline: 24781964]
44. López L, Green AR, Tan-McGrory A, King RS, Betancourt JR. Bridging the digital divide in health care: the role of health information technology in addressing racial and ethnic disparities. Jt Comm J Qual Patient Saf 2011 Oct;37(10):437-445. [doi: 10.1016/s1553-7250(11)37055-9] [Medline: 22013816]
45. Lyles CR, Tieu L, Sarkar U, et al. A randomized trial to train vulnerable primary care patients to use a patient portal. J Am Board Fam Med 2019;32(2):248-258. [doi: 10.3122/jabfm.2019.02.180263] [Medline: 30850461]

XSL•FO
RenderX

46. Hefner JL, Sieck CJ, Walker DM. Patient and physician perspectives on training to improve communication through secure messaging: clarifying the rules of engagement. Health Care Manage Rev 2022;47(1):3-11. [doi: 10.1097/HMR.0000000000000279] [Medline: 32379081]
47. Hernandez AM, Khoong EC, Kanwar N, et al. Lessons learned from a multi-site collaborative working toward a digital health use screening tool. Front Public Health 2024;12:1421129. [doi: 10.3389/fpubh.2024.1421129] [Medline: 39040864]
48. Segall N, Saville JG, L'Engle P, et al. Usability evaluation of a personal health record. AMIA Annu Symp Proc 2011;2011:1233-1242. [Medline: 22195184]
49. Britto MT, Jimison HB, Munafo JK, Wissman J, Rogers ML, Hersh W. Usability testing finds problems for novice users of pediatric portals. J Am Med Inform Assoc 2009;16(5):660-669. [doi: 10.1197/jamia.M3154] [Medline: 19567793]
50. Sarkar U, Gourley GI, Lyles CR, et al. Usability of commercially available mobile applications for diverse patients. J Gen Intern Med 2016 Dec;31(12):1417-1426. [doi: 10.1007/s11606-016-3771-6] [Medline: 27418347]
51. Wisniewski H, Gorrindo T, Rauseo-Ricupero N, Hilty D, Torous J. The role of digital navigators in promoting clinical care and technology integration into practice. Digit Biomark 2020;4(Suppl 1):119-135. [doi: 10.1159/000510144] [Medline: 33442585]
52. Natsuhara KH, Shah ND, Ma C, et al. Bridging the digital divide with digital navigation in the California safety net: a mixed-methods study. J Gen Intern Med 2026 Mar 17. [doi: 10.1007/s11606-026-10270-4] [Medline: 41845090]
53. Kong M, Willard-Grace R, Khoong E, et al. Implementing team-based telemedicine workflows in safety-net primary care. Am J Manag Care 2024 May;30(6 Spec No):SP437-SP444. [doi: 10.37765/ajmc.2024.89550] [Medline: 38820184]

Abbreviations

HCD: human-centered design
HCI: human-computer interaction
REDCap: Research Electronic Data Capture
SFHN: San Francisco Health Network
SMART Digital Health Scale: Skills Measurement and Readiness Training for Digital Health Scale

Edited by A Kushniruk; submitted 09.Jun.2025; peer-reviewed by A Sheon, AB Garcia; revised version received 20.Feb.2026; accepted 31.Mar.2026; published 04.May.2026.

Please cite as:
Rapson T, Ramirez M, He S, Wong J, Kim HC, Luna I, Yang A, Li J, Fishman PA, Ralston JD, Lyles CR, Khoong EC
Identifying Skill and Usability Barriers to Digital Health Tool Use Among Older Adult Patients in US Safety Net Clinics: Mixed Methods Study
JMIR Hum Factors 2026;13:e78430
URL: https://humanfactors.jmir.org/2026/1/e78430
doi:10.2196/78430

© Taylor Rapson, Magaly Ramirez, Sandy He, Jeanette Wong, Hyunjin Cindy Kim, Isabel Luna, Andersen Yang, Junhong Li, Paul Fishman, James D Ralston, Courtney Lyles, Elaine C Khoong. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 4.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Cocreating a Digital Patient Preparedness Tool for International Students Accessing Primary Care in Germany: A Convergent Methods Study Within the Health CASCADE Network

Vinayak Anand Kumar[1,2], PhD; Margrit Schreier[1], Prof Dr; Maria Giné-Garriga[3], Prof Dr; Maria Ortmann[1], BS; Fatima-Zohra Belmokhtar[4], BA; Simona Grineviciute[1], BS; Tran Ngoc-Huong Quan[1], BS; Likhita Aluru[5], BA; Sonia Lippke[1,6], Prof Dr

[1]School of Business, Social and Decision Science, Constructor University, Campus Ring 1, Bremen, Germany

[2]Department of Psychology, University of Bamberg, Bamberg, Germany

[3]School of Health Sciences Blanquerna, Universitat Ramon Llull, Barcelona, Catalonia, Spain

[4]Faculty 11 Human and Health Sciences, University of Bremen, Bremen, Germany

[5]Department of Neurobiology, Northwestern University, Evanston, IL, United States

[6]Faculty of Health, Hamburg University of Applied Sciences/ Hochschule für Angewandte Wissenschaften, Hamburg, Germany

Corresponding Author:
Vinayak Anand Kumar, PhD
School of Business, Social and Decision Science, Constructor University, Campus Ring 1, Bremen, Germany

Abstract

Background: Digital health interventions (DHIs) are a potential tool to address communication challenges in primary care medicine by improving patient participation, treatment adherence, and self-management of chronic diseases. However, Germany's digital health infrastructure remains underdeveloped compared to other Organization for Economic Cooperation and Development (OECD) countries. Designing tailored, context-sensitive tools requires a deep understanding of patient needs, particularly for underserved populations navigating complex health care systems. Participatory approaches that actively involve end-users and stakeholders can help ensure that digital health tools are responsive to these needs. Behavioral science frameworks such as the Behavior Change Wheel (BCW), which incorporates the Capability, Opportunity, Motivation–Behavior (COM-B) model, can further support this process. Such technologies, developed using cocreation and behavior change frameworks, may improve health outcomes for underserved patient populations by addressing context-specific needs. This study explores how cocreation and behavioral science can inform the adaptation of a digital patient preparedness tool for international students accessing primary care in Germany.

Objective: This study aimed to gain insight into international students' experiences with primary care in Germany and explore whether adapting an existing digital patient preparedness intervention could address communication challenges and improve the standard of care. Using cocreation methods and the Behavior Change Wheel (BCW), we identified key design specifications and behavior change levers to inform tool development.

Methods: A mixed methods design was used to identify design specifications for a DHI across 4 cocreation workshops with 12 students at an international university in Germany. Quantitative data were used for descriptive insights, and qualitative data were analyzed using qualitative content analysis. Workshops were informed by the BCW and the Health CASCADE cocreation methods selector tool.

Results: Cocreators reported feeling misunderstood, anxious, and ill-informed during primary care interactions, with system-level barriers compounding communication difficulties. Despite this, many engaged in preparatory behaviors (eg, note-taking) to manage uncertainty and structure their consultations. Feedback on an existing digital intervention was mixed: while cocreators appreciated its intent, structured lesson formats were seen as too time-consuming. Cocreators preferred a concise, interactive design. Communication prompts, appointment scheduling, and personalized feedback were frequently requested features, though tool adoption was seen as contingent on addressing broader system-level frustrations.

Conclusions: International students' negative health care experiences, often stemming from unclear communication, perceived indifference, and difficulty navigating an unfamiliar medical system, may be mitigated through codesigned, personalized digital interventions. Frequently requested features, such as appointment scheduling, clinic directories, test result access, and interactive tools like chatbots, may help address barriers related to system navigation, communication, and access. This study demonstrates how cocreation methods can be integrated with the BCW to inform the development of context-sensitive digital health tools for specific target groups.

XSL•FO
RenderX

(JMIR Hum Factors 2026;13:e87519) doi:10.2196/87519

KEYWORDS

digital health; cocreation; behavior change; patient preparedness; primary care; international students; communication barriers; behavior change wheel; COM-B model; Germany; Capability, Opportunity, Motivation – Behavior Model

Introduction

Effective communication between patients and practitioners is pivotal to the quality and delivery of health care services [1]. This is especially true in primary care medicine where physicians serve as trusted advisors, functioning as the central hub for managing complex, chronic conditions and navigating care between multiple specialists. In these care spaces, effective patient-physician communication was found to improve understanding, trust, and engagement in health-related behaviors [2,3]. However, the burden on the primary care system in Germany, including physician shortages and resource constraints [4-6], has given rise to communication failures, which in turn have implications for the perceived quality of care [7,8]. These challenges are particularly pronounced for migrant populations in Germany, who are unfamiliar with the nation's health care system and clinical care methodology [9]. Communication challenges in German health care services, such as language and cultural barriers [10], have been linked to poor symptom management [11] and worsened patient outcomes [12].

Digital health interventions (DHIs) could be practical tools that address the communication challenges in primary care settings and improve overall standard of care. For example, in Norway, a digital symptom-reporting and communication intervention was shown to enhance patient-provider dialogue [11], and in Sweden, a digital portal for young diabetes patients was found to facilitate trust and dialogue between patients and health care workers [13]. Moreover, a recent systematic review found that, more broadly, digitalization in health care is generally beneficial for improving communication-related factors like patient participation, treatment adherence, and self-management in chronic care [14]. Based on the Bertelsmann Digital Health Index, Germany is far behind most European neighbors with respect to digitization in health care, being placed second to last in terms of digital readiness, policy, and data use [15]. These shortcomings stem not only from slow implementation, but also from tools that fail to address real-world needs [16]. In this context, cocreation offers a pragmatic approach to developing fit-for-purpose tools by involving patients and other end-users directly in the design of digital health tools.

Digital services can help individuals navigate health care systems [17], book appointments [18], increase access to care [19], improve patient engagement [20], and enhance disease monitoring and management [21]. DHIs may be especially useful for migrants who often face challenges in navigating the health care system, making appointments, preparing for consultations, and engaging effectively with health care providers, in addition to language and cultural barriers [22]. International students form a key subgroup within the migrant population are central to Germany's skilled migration strategy [23]. International students contribute to the economy through tuition fees, living expenses, and potential future participation in the skilled labor market if they remain in Germany after graduation [24]. Health care quality is a known factor influencing such decisions [25]. Therefore, understanding the needs of international students is not only critical to evaluating health care accessibility for future skilled migrants, but has broader economic implications for Germany [26].

The unique health care experiences and needs of the international student population in Germany have been largely unstudied, especially in primary care settings. However, some studies have been carried out exploring the health care experiences of young people in general. A cross-sectional study comparing university students in Germany and Belgium found that students in Germany were more likely to seek specialist and emergency care [27]. This pattern may indicate underusage of primary care services or difficulties navigating them. Moreover, there is evidence that individuals with a migrant background have higher odds of not having a general practitioner (GP), further suggesting that services may be underused [28]. These prior studies have highlighted communication barriers for broader migrant groups and refugees [29,30]. It is crucial to study international students as the intersection of these 2 populations to address the specific health priorities of this young, well-educated, and mobile population.

DHIs offer opportunities to personalize support and enhance communication before consultations. They also enable data capture on patient concerns, helping physicians develop more accurate patient profiles [31,32]. These tools are most effective when cocreated with end users to reflect their communication preferences, which often include clarity, empathy, and cultural sensitivity [33]. Behavioral science, particularly the Behavior Change Wheel (BCW) framework and the corresponding Capability, Opportunity, Motivation–Behavior (COM-B) model, offer a systematic way to map how digital tools can build patient capabilities, create opportunities, and motivate change [34]. The COM-B model conceptualizes behavior as arising from interactions between capability, opportunity, and motivation (COM), where capability refers to individuals' knowledge and skills, opportunity to external factors that enable or constrain behavior, and motivation to reflective and automatic processes that drive action. It is embedded within the BCW, which provides a structured process for identifying target behaviors and selecting appropriate behavior change techniques (BCT). Previous studies have demonstrated that the BCW provides a systematic approach to identifying design considerations for behavior change interventions [35]. Cocreation helps align the design of tools with real-world considerations [36] and, when combined with the BCW, can help create acceptable and context-sensitive interventions for specific target groups [37].

To address the above challenges and opportunities, this study aims to adapt an existing digital patient preparedness tool (DiPPT) using cocreation with students at an international university in Germany, guided by behavior change theory. The

XSL•FO
RenderX

cocreation workshops were structured using the PRODUCES framework [38], a cocreation framework for developing public health interventions that guides researchers through key stages of cocreation, including defining the Problem and Objective, planning the Design, identifying Users and Cocreators, and considering Evaluation and Scalability [38]. In the present study, the framework supported a systematic approach to end-user involvement, while the BCW [39] informed both the workshop topics and the derivation of intervention specifications from the workshop outputs. The workshops aimed to identify communication and preparation needs in primary care and to map behavior change levers for patient preparation using the COM-B model to inform the design of a digital intervention.

The study was guided by the following research questions which were investigated using different methods across a series of cocreation workshops:

1. How do international students prepare for conversations with physicians in German primary care settings?
2. What are the needs of international students with respect to receiving and relaying information in German primary care settings?
3. What are the waiting room experiences of international students prior to GP appointments?
4. What aspects of the preexisting DiPPT did international students see as suitable for a digital intervention to facilitate communication?
5. What aspects of the DiPPT did international students see as not suitable for a digital intervention to facilitate communication?
6. What are the needs of international students for a digital intervention to facilitate communication?

The methods used across the workshops demonstrate how cocreation can be used to identify end-user needs and translate these into design specifications for a behavior change intervention, which was a central objective of this Health CASCADE study.

Methods

Study Design

This study uses a convergent mixed methods design [40], with qualitative and quantitative data collected in parallel, to cocreate a digital health intervention for international students preparing for primary care consultations in Germany. The structure and methods of the cocreation workshops were specified in advance; building on a prior communication intervention (DiPPT) in obstetric care [41,42] and adapting it to the current target group.

Data integration occurred at the interpretation stage through narrative integration [43,44]. Quantitative survey data were used to descriptively characterize cocreators, summarize health care experiences, waiting room experiences, preparation behaviors, and evaluate ratings of existing DiPPT lessons. Qualitative data were used to explore and contextualize these experiences and behaviors in greater depth to identify needs and preferences for a redesigned digital intervention. See Table 1 for how qualitative and quantitative data were used to understand cocreators' experiences and behaviors, and inform the adaptation of DiPPT for international students accessing primary care in Germany.

XSL•FO
RenderX

Table . Overview of cocreation workshops conducted with international students accessing primary care in Germany, aligned with stages of the BCW[a] framework in a convergent mixed methods study.

BCW[a] stage	Aim	Methods
Understand the behavior	Explore the communication experiences of international students accessing primary care.	Qualitative data • Focus groups. Quantitative data • Survey on doctor-patient communication.
Identify intervention options	Evaluate the relevance of content from an existing intervention (DiPPT)[b].	Qualitative data • Empathy maps exploring patient emotions and behaviors in waiting rooms. Quantitative data • Survey on clinic waiting room experience. • Rating each communication training component from DiPPT based on brief descriptions of each lesson.
Identify intervention options	Explore needs for a digital intervention in general as well as refine and prioritize content and features from the existing DiPPT for a redesigned digital tool.	Qualitative data • Card sorting exercise of intervention ideas. Quantitative data • Rerating each communication training component from DiPPT after having used the web-app.
Identify intervention options	Co-design and discuss the structure and core features of a tailored preparedness intervention.	Qualitative data • Visioning exercises and cocreation of design charts. • Focus groups to evaluate proposed intervention. Quantitative data • Survey on demographics and health.

[a]BCW: Behavior change wheel.

[b]DiPPT: digital patient preparedness tool.

Ethical Considerations

Ethical approval for the study was granted by the Ethics Committee at Constructor University (application number: 2022_01_B). All cocreators provided informed consent prior to taking part in each workshop and were informed about the purpose of the study, the types of data collected, and their rights as cocreators. Eleven out of 12 cocreators were employed as paid contributors and compensated equally based on the number of hours they contributed to the project (US $14.36/h). One cocreator was unable to receive payment due to visa-related employment restrictions and participated voluntarily. To protect cocreator privacy and confidentiality, audio recordings and workshop materials were anonymized with disclosure controls applied to names, health conditions, institutions, countries, and locations.

Digital Patient Preparedness Tool

The original digital intervention (DiPPT) was developed to support communication between pregnant women and health care professionals in obstetric care [41,42,45,46]. The web-based tool delivered structured communication training through short lessons and exercises designed to help expectant mothers prepare for consultations and express concerns clearly. The intervention targeted safe communication behaviors during maternity care encounters and was informed by the Health Action Process Approach (HAPA) model, primarily using action planning (BCT 1.4) to encourage patients to plan for obstetric and midwife consultations ahead of time.

Recruitment and Data Collection

To ensure the intervention was grounded in real-world experiences, international students were engaged as cocreators. The following section outlines how they were selected and involved in a series of cocreation workshops designed to explore patient behaviors and collaboratively adapt a digital preparedness tool using the BCW [39]. Additional details are mentioned in Supplement 1 in Multimedia Appendix 1 that consists of a full list of tools used to collect data from cocreators.

Selecting Cocreators

Cocreators were purposively sampled from an international university in Germany, using criterion sampling [47]. Eligibility criteria included being at least 18 years old, English-speaking, and having received primary care in Germany. Recruitment took place via in-class presentations, posters, flyers, and digital listings. Interested individuals completed an eligibility screening form and participated in a brief interview to gauge motivation for taking part. Twelve cocreators were selected. Of these, 10 attended all 4 workshops and 2 attended 3. Cocreators were

XSL•FO
RenderX

relatively homogeneous with respect to demographic characteristics. All were between 19 and 22 years of age, and 9 of the 12 were female. Cocreators represented 6 academic disciplines, including psychology (n=5), chemistry and biotechnology (n=3), and computer science, economics, business administration, and medicinal chemistry (n=1 each). Eight cocreators were born in Asia or Africa, with the remaining cocreators originating from the Americas (ie, North, Central, South America) (n=2) and Europe (n=2) (Supplement 4 in Multimedia Appendix 2).

Cocreation Workshops

The aim of the cocreation workshops was to explore how international students in Germany prepare for health care consultations and to collaboratively adapt an existing digital communication intervention (DiPPT) into a preparedness tool tailored to this group's needs. To design the intervention, the BCW [39] was used to inform the structure of the workshops: to first understand patient behaviors and then identify and refine potential intervention options. Workshop methods were sourced from consultation with experts in qualitative research methods and from the cocreation methods selector tool [48], which was developed within the Health CASCADE research network to support evidence-based cocreation in public health. To structure the cocreation workshops so that they could be evaluated in scientific terms, the PRODUCES framework was used [38]. The PRODUCES framework guides cocreation of public health interventions by structuring the planning, conduct, evaluation, and reporting of participatory processes, enabling a systematic and reproducible approach to intervention development [38]. Table 1 details how relevant stages of the BCW were operationalized in terms of workshop aims and methods used.

Quantitative Data Collection

A combination of nominal, Likert-scale, and free-text items was used in the survey instruments. All survey measures were developed specifically to explore patient needs and experiences in primary care, with a particular focus on communication and waiting room experiences. As there were no instruments with established psychometric properties that aligned precisely with the research focus, the bespoke tools were deemed fit for purpose. Furthermore, the need to capture target group–specific variables, minimize respondent burden, and reduce disclosure risk further supported the use of a short, tailored instrument rather than adapting existing scales designed for different contexts. Supplement 1 in Multimedia Appendix 1 consists of the full list of survey items.

Qualitative Data Collection

A variety of qualitative methods for data collection was used in the workshops to capture cocreators' thoughts, opinions, experiences, and emotions both in primary care and concerning the requirements of a DHI designed to support international students in seeking health care. Focus groups were employed to elicit a broad range of experiences in primary care and to facilitate discussion of proposed interventions. These were supplemented by methods such as empathy maps to collect data on emotions and visioning exercises designed to stimulate creative thinking about adapting the original DHI to the primary care context.

Data Processing and Analysis

Quantitative Analysis

Survey responses were processed using R (v4.5.1; R Foundation for Statistical Computing) [49] and RStudio (Posit PBC) [50], with health conditions categorized according to *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision* (*ICD-10*) codes and countries of origin generalized to regions. Descriptive statistics and visualizations (bar charts, histograms, heatmaps) were used to summarize patterns in health status, waiting room experiences, and communication challenges; no inferential statistical analyses were conducted, as the quantitative data were solely intended to provide descriptive workshop-level insights alongside the qualitative findings. All code used to derive descriptives can be found in the Open Science Framework (OSF; Center for Open Science) repository, with the OSF Identifier: JBEAP [51].

Qualitative Analysis

Audio recordings from the focus groups were transcribed verbatim by a transcription agency. Transcripts were checked for accuracy by trained student assistants and anonymized using QualiAnon (v1.3.1; Qualiservice) [52]. Empathy maps, card sorting grids, and visioning charts were digitized using photographic records and recreated in JamBoard (Google LLC) for analysis.

Data was analyzed using structured qualitative content analysis (QCA) [53]. A coding frame was developed iteratively through deductive and inductive approaches. Initial categories were defined deductively based on the research questions and communication constructs from the original DiPPT, then expanded inductively through review of transcripts and workshop outputs. The final frame consisted of 12 parent categories, some with secondary layers, resulting in 52 unique codes. Supplement 2 in Multimedia Appendix 3 details the full coding frame.

Coding was carried out by 2 independent researchers (VAK and MS) across three iterative rounds using QualCoder (v3.5; Colin Curtain) [54]. Thematic segmentation was first applied to identify units of meaning. In each round, the 2 coders independently assigned codes to a shared subset of data (initial round: 45.95% of all segments), blinded to each other's coding decisions during the initial coding stage. After each round, discrepancies were discussed and resolved either through consensus or by refining the coding frame. This two-pronged approach ensured both coder agreement and alignment with the data. As a result, the coding frame evolved in close collaboration between VAK and MS and was refined to improve consistency, clarity, and coverage. All frame iterations were documented to create an audit trail. Reliability and validity assessments were conducted after each round, yielding a final interrater agreement of 74.55% and a "miscellaneous" coding rate of 4.2%, indicating improved frame specificity and low residual ambiguity. Regarding the reliability assessment, code-specific metrics such as Cohen κ were not used due to the large coding frame relative to the sample set, which can distort agreement estimates when

XSL•FO
RenderX

codes were infrequently used or unused [49]. Reliability and validity measures for each category are presented in Supplement 3 in Multimedia Appendix 4. Final coding was then completed by the lead researcher (VAK) using the validated frame. Outputs were extracted using custom SQL queries (accessed through OSF with the identifier: JBEAP) [51].

Results

Cocreators

Twelve international students participated in the cocreation workshops. Half reported living with a long-standing illness, including conditions related to mental health, musculoskeletal issues, and respiratory disorders. Despite this, most cocreators had visited a GP fewer than 5 times in the past year. Supplement 4 in Multimedia Appendix 2 provides detailed information about all data visualizations describing cocreators.

Understanding Experiences and Behaviors of Cocreators

Cocreators described health care as an environment in which emotional uncertainty, system-level barriers, and communication breakdowns were common. Despite these challenges, many took active steps to prepare for appointments, suggesting that patient preparedness may be a promising target behavior. Insights from the workshops are presented below and summarized in Table 2. See Supplement 5 in Multimedia Appendix 5 for all data visualizations relating to cocreator experiences and behavior when receiving care.

Table . Topics, qualitative categories, example quotations, and most frequently used codes relating to health care experiences, waiting room experiences, and patient preparedness behaviors among international students in Germany.

Topic and category	Code	Example
Experiences of cocreators		
Assigning responsibility	Primarily doctor	"I agree in the sense that it's both of their responsibilities, but I think it's mostly the health care workers" [Cocreator 2, February 6, 2024]
Needs and complaints in primary care	Dignity in care	"The doctor is feeling maybe that I'm wasting his time, and it's really not a nice feeling, like I'm being neglected" [Cocreator 5, February 6, 2024]
System challenges in Germany	Insufficient care	"Then, I feel like, okay, go on [REDACTED]'s page, look at the doctors that speak those languages, and then you look up their appointments and then they don't have more appointments." [Cocreator 4, February 6, 2024]
Emotions whilst waiting to see the doctor	Fear and anxiety	"The doctor came by and was like 'Let me have a look at your hand." He did some tests and then it was like 'Okay, we're going to get you to take an image.' Then again, I was just waiting and I didn't know whether I was supposed to tell somebody to take me to get the image or not. But then, after an hour and a half of waiting, someone came. However, there's this creepy aura" [Cocreator 2, February 6, 2024]
Thoughts while waiting to see the doctor	Anticipation or expectation	"Am I having a heart attack?" [Empathy map – Chest pain, February 9, 2024]
Behavior of cocreators		
Preparation method	Taking notes	"Usually, I write down everything I have to say to the doctor" [Cocreator 3, February 6, 2024]
Preparation location	On the way	"Then maybe you revise it like when you're on the train to refresh your mind" [Cocreator 11, February 6, 2024]
Preparation timing	Day before	"I think mostly like the day before" [Cocreator 11, February 6, 2024]
Reasons to prepare	Prompt the doctor	"I want to sometimes discuss with the doctor if this is a possible thing if this is a possible disease, or possible answer for whatever my symptoms are" [Cocreator 2, February 6, 2024]

Quantitative findings further showed that most cocreators viewed communication as a shared responsibility between physician and patient, although few felt that patients should take the lead role (Table 2 and Supplement 5 in Multimedia Appendix 5). Qualitative findings expanded on this pattern by demonstrating that cocreators generally expected physicians to

XSL•FO
RenderX

assume primary responsibility for facilitating effective communication because of their duty of care or professional obligation.

Waiting Room Experiences

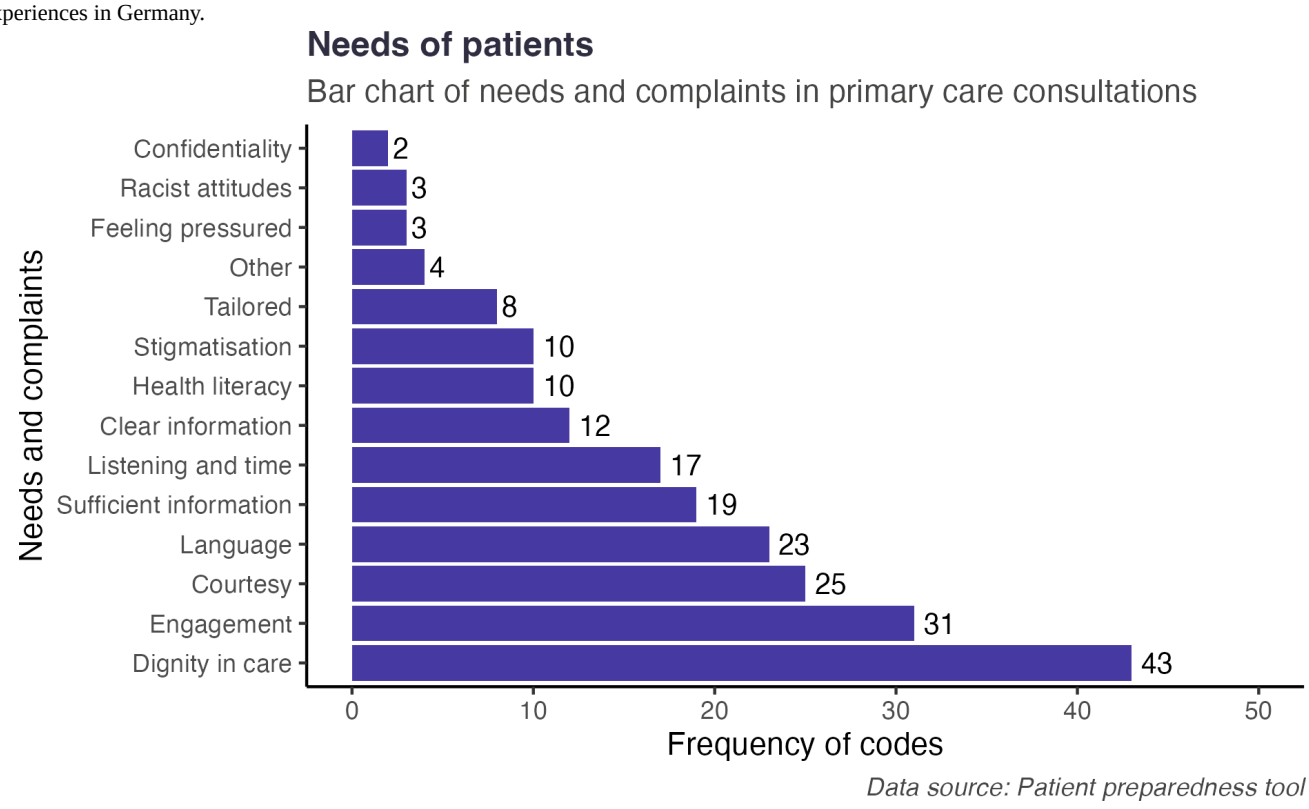
Quantitative and qualitative findings addressed different aspects of the waiting room experience. While survey data characterized where, with whom, and how long cocreators waited, qualitative findings highlighted the emotional experience of waiting. Survey data showed that the majority of cocreators arrived alone to appointments, had access to a waiting room, and waited inside the clinic. Waiting times varied considerably across cocreators; 6 waited less than 30 minutes on average, while 5 waited more than 30 minutes on average. The qualitative data showed that cocreators frequently experienced anxiety and uncertainty while waiting, often linked to anticipated diagnoses, unclear consultation expectations, or past experiences. This psychological vulnerability was one of the most frequently cited themes during empathy mapping activities.

Needs of Patients

Communication Needs and Challenges

With respect to communication challenges that patients face, qualitative and quantitative data converge. Eight cocreators reported feeling misunderstood by health care workers in the quantitative data, and 4 reported not being listened to. Qualitative data expanded upon this by specifying the perceived causes of communication difficulties. Language barriers, lack of empathy and engagement, insufficient explanations, and dismissive behavior were frequently reported communication challenges across the focus groups. Figure 1 details the frequency of all codes used for the category “needs and complaints” in health care settings. Cocreators articulated a range of communication needs, including clearer explanations, culturally sensitive interactions, emotionally validating exchanges, and information delivered in plain language. Time constraints during consultations were also identified as a barrier to effective communication.

Figure 1. Frequencies of communication-related needs and complaints reported by international students during cocreation workshops on primary care experiences in Germany.

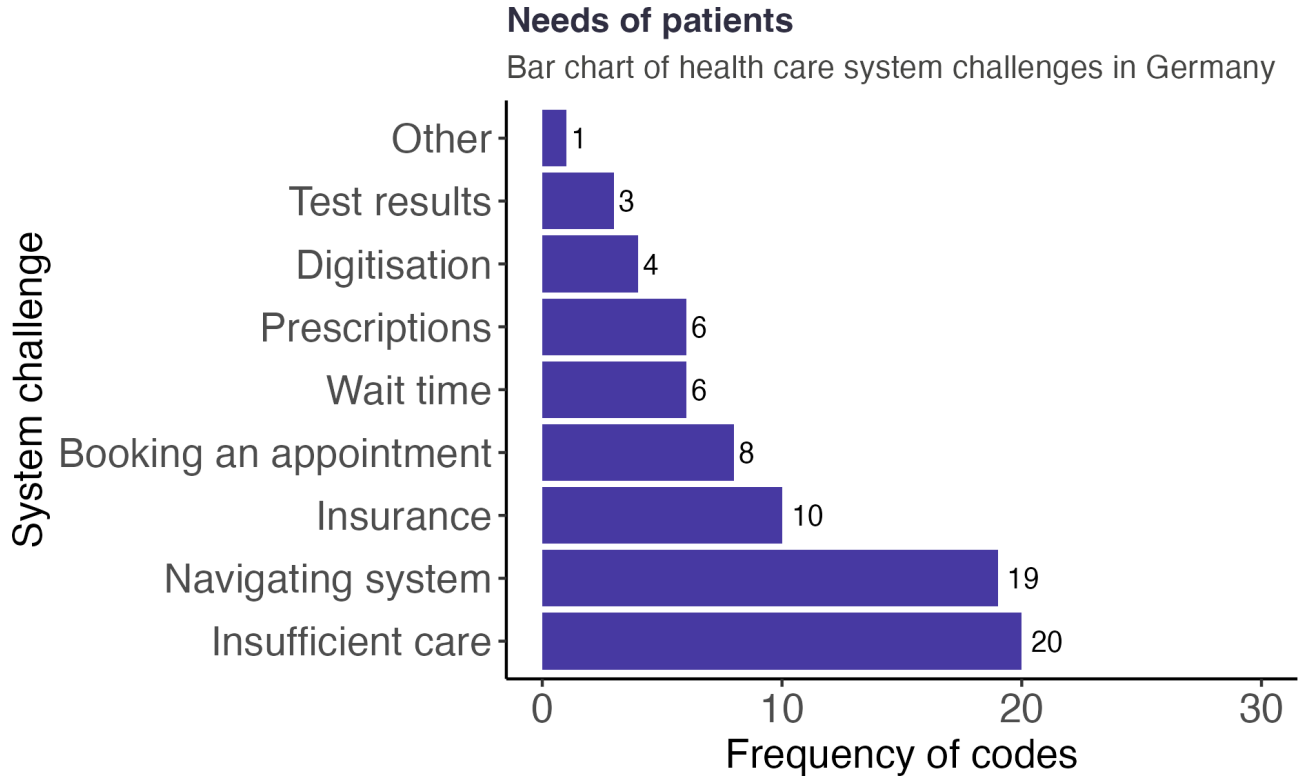


System Level Needs and Challenges

In addition to interpersonal issues, cocreators highlighted system-level needs in the focus groups such as improved access to appointments, better navigation support, and more user-friendly digital infrastructure, particularly for tasks like receiving test results. Complaints about difficulty navigating the system and inadequate digital services were among the most frequently cited challenges across the dataset. Figure 2 contains the frequency of all codes used for the category “system level challenges” in the German health care system.

XSL•FO
RenderX

Figure 2. Frequencies of system-level challenges reported by international students when navigating the German health care system during cocreation workshops.



Patient Preparedness behavior

Focus groups revealed differences in how patients prepared for their appointments. Cocreators frequently described taking notes, listing symptoms, or researching conditions the day before or on the way to the clinic. These actions were intended to structure the consultation, ensure efficiency, and psychologically prepare for the encounter. “Taking notes” was one of the most frequently coded behaviors in the dataset. Some cocreators used preparation as a way to prompt the physician, express concerns clearly, or guide the direction of the conversation.

Intervention Options

Qualitative data revealed that while the concept of a DiPPT was welcomed, the design must prioritize usability and relevance. Rather than structured training, cocreators sought lightweight, responsive features that could flexibly support them before and during primary care consults. These insights are detailed below and summarized in Table 3. Supplement 6 in Multimedia Appendix 6 outlines all data visualizations relating to cocreator needs for the digital health intervention.

XSL•FO
RenderX

Table . Qualitative categories, example quotations, and most frequently used codes relating to perceptions of the DiPPT[a] and desired features for a redesigned digital preparedness intervention.

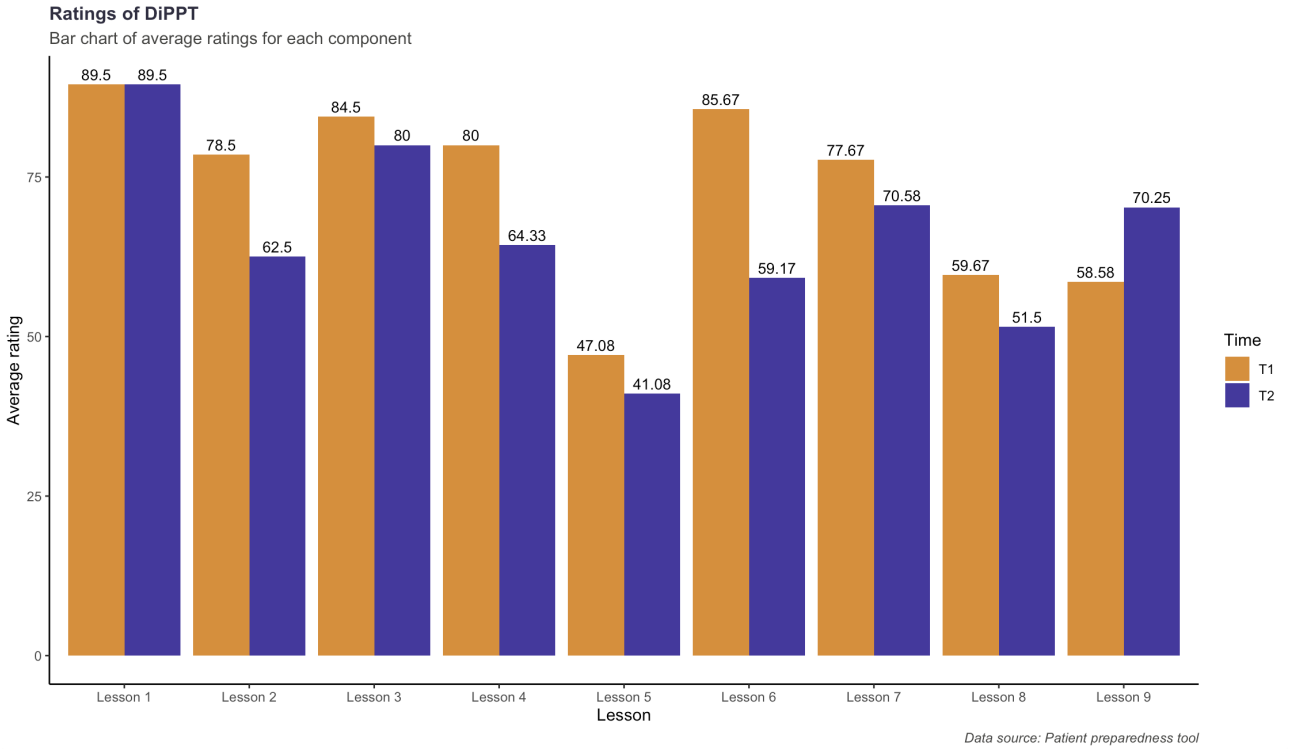
Category	Code	Example
Elements of DiPPT[a] to retain	Retain content	“Maybe some lessons could be especially useful to a lot of people.” [Cocreator 5, February 20, 2024]
Elements of DiPPT to modify	Modify content	“Lesson 3 was frustrating, get to the point.” [Visioning flip chart – Lesson 3, February 20, 2024]
Features to include in the digital health tool	Communication training	"We had problems going to the Hausarzt explaining our state, like giving an explanation to the Hausarzt about our state, our symptoms. That was the main problem. If this app kind of helps in all of this, the communication between us and the doctors, then we can be better. " [Cocreator 10, February 20, 2024]
Features to exclude in the digital health tool	Open-ended responses	“When you’re typing a long answer... you expect something back.” [Cocreator 6, February 20, 2024]
Utility of tool	Limited use	“I don’t think I would use it multiple times...maybe just once if I have a really serious appointment.” [Cocreator 7, February 20, 2024]

[a]DiPPT: Digital Patient Preparedness Tool.

Suitability of the Existing DiPPT

Quantitative data identified the lessons that cocreators rated most and least favorably, while qualitative findings help explain both the content and delivery characteristics that may have contributed to these ratings. Figure 3 presents the average ratings cocreators assigned to each lesson before using the web-app (ie, with only a description of what was covered in the lesson) and after using the web-app.

Figure 3. Average ratings of lessons from the digital patient preparedness tool (DiPPT) by international student cocreators before and after using the intervention during cocreation workshops in Germany.



With respect to intervention design, the data converge in suggesting a preference for concise, low-burden content. The highest-rated lessons were among the shortest in the intervention and contained relatively few open-ended or closed-ended exercises, while lesson 5, which had the lowest rating, contained the most open-ended activities of all lessons in the intervention. Qualitative data support this interpretation, with cocreators

XSL•FO
RenderX

expressing dissatisfaction with lengthy lessons, quizzes, open-ended activities, and the existing lesson-based format.

With respect to communication support, the data converge in suggesting that cocreators valued content that helped them overcome barriers encountered during health care interactions, whereas lower-rated lessons asked patients to assume additional responsibility beyond overcoming those barriers. The highest-rated lessons focused on communicating information clearly (Lesson 1), speaking up about concerns (Lesson 3), and preparing for consultations (Lesson 7). Qualitative findings help explain this pattern, as cocreators frequently reported difficulties obtaining sufficient and clear information, having their concerns taken seriously, being listened to, and communicating effectively with health care workers. Lesson 3 directly addressed complaints relating to dignity in care, engagement, courtesy, and feeling heard by encouraging patients to express concerns and advocate for their needs. Lessons 1 and 7 addressed needs for sufficient and clear information by supporting patients in structuring information, asking questions, and seeking clarification. Lesson 1 additionally provided strategies for communicating clearly, which may be particularly relevant for cocreators who reported language-related communication difficulties.

The data also converge in suggesting that content perceived as shifting responsibility towards patients was viewed less favorably. Two of the lowest-rated lessons focused on empathizing with the doctor (Lesson 5) and structuring information for the doctor (Lesson 6), both of which emphasized adapting patient behavior to facilitate communication with health care workers. Qualitative findings support this interpretation, as most cocreators expected physicians to take primary responsibility for communication because of their professional role and expertise.

The data further suggest that alignment with cocreators' existing behaviors and perceived needs influenced lesson ratings. Lesson 7, which received a relatively high rating, aligned closely with preparedness behaviors identified in the workshops, including taking notes, prioritizing concerns, and preparing questions before consultations. In contrast, Lesson 8, which received a relatively low rating, did not directly address the communication difficulties or health care navigation challenges raised by cocreators. Instead, it focused primarily on stress management and may therefore have been perceived as less relevant to their immediate needs.

Together, the data suggest that cocreators valued practical communication support that directly addressed challenges encountered during primary care consultations, while content perceived as burdensome, less relevant, or shifting responsibility towards patients was viewed less favorably.

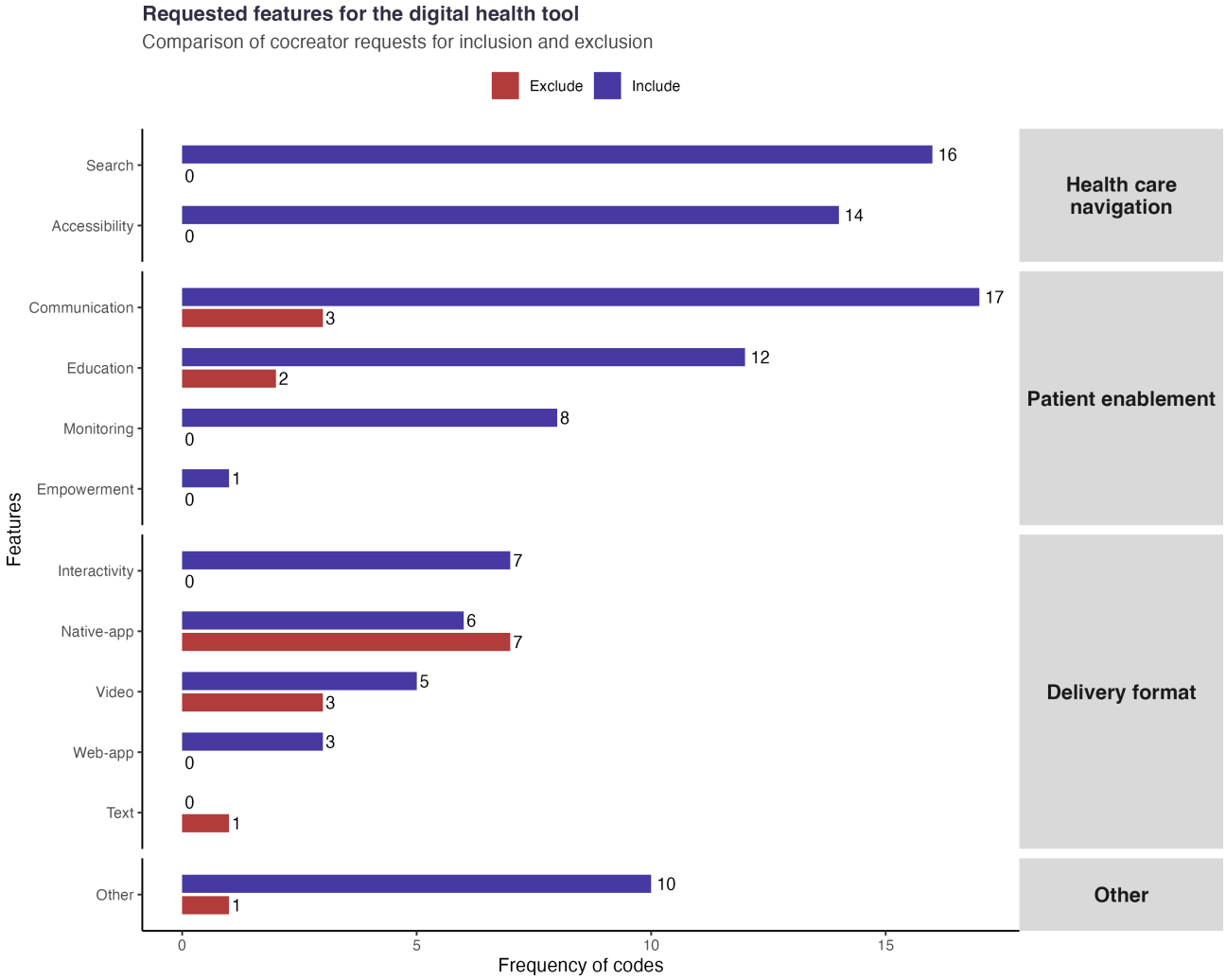
Needs for a Digital Tool

Quantitative lesson ratings indicated general acceptance of communication-related content, while qualitative data identified communication support as the most frequently requested feature. Together, the data suggest that communication support may be a useful component of a digital preparedness intervention for international students.

Qualitative data from focus groups, card sorting activities, and visioning exercises identified several additional functional needs. These included support for appointment scheduling, clinic search functions, access to patient rights and symptom information, health monitoring tools, personalized feedback, and interactive features such as chatbots. The most frequently requested features were communication prompts, scheduling functions, and interactive components. Figure 4 includes the features cocreators requested to be included and excluded from the design specifications for a digital health intervention.

XSL•FO
RenderX

Figure 4. Frequencies of features requested for inclusion and exclusion in a redesigned digital health intervention (DHI) for international students accessing primary care in Germany.



Qualitative data also highlighted reservations regarding adoption of the intervention. Many cocreators indicated that they would use such a tool only under certain conditions or not at all, particularly if it addressed only doctor-patient communication. Several cocreators suggested that the perceived value of the intervention would increase if it also supported broader health care needs, such as appointment scheduling, clinic navigation, or access to health care information.

Quantitative and qualitative data provided complementary insights into the suitability of delivering the intervention through a digital medium. While survey data indicated that most cocreators could use mobile devices while waiting for appointments, limited Wi-Fi availability and variable mobile data quality may restrict digital engagement in practice. Qualitative findings further highlighted concerns regarding the usefulness of a standalone communication app. Together, these findings suggest that successful implementation may depend on both the digital environment in which the intervention is used and the extent to which it addresses needs beyond communication support.

Discussion

Principal Findings

This study adopted a convergent mixed methods design to cocreate a DiPPT with students from an international university in Germany. To our knowledge, this study is the first to explore the health care needs and experiences of international students accessing primary care in Germany and to use these insights to cocreate a DiPPT. Quantitative and qualitative data together captured international students' health care experiences, preparation behaviors, and perceptions of a digital preparedness intervention. Integrating quantitative and qualitative findings enabled interpretation of both the prevalence and context of cocreators' experiences. Survey findings provided descriptive insight into communication perceptions, health care experiences, and waiting room conditions, while qualitative findings contextualized these patterns by exploring emotional vulnerability, preparation behaviors, and intervention preferences in greater depth. Cocreators' needs explored in the qualitative data included both interpersonal communication needs and challenges as well as broader system-level barriers related to navigating the health care system and accessing services. Importantly, these findings suggest that the health care experiences of international students may reflect the more

XSL•FO
RenderX

general experience of young adults navigating an unfamiliar health care system independently for the first time. Collectively, the data informed a design specification for a digital intervention tailored to international students in Germany while demonstrating a replicable approach for integrating cocreation methods and the BCW in digital health intervention development.

In the survey data, most cocreators agreed that communication should be a shared responsibility between physician and patient, and the majority disagreed that patients should carry the lead role. Qualitative findings indicated that despite the shared responsibility, physicians were expected to take the lead due to professional responsibility. The expectation that both parties should contribute to communication is consistent with evidence that younger people increasingly wish to participate actively in health care decisions rather than adopt a purely passive role [55,56]. Similarly, recent work on Generation Z has found that young adults value collaborative relationships with health care providers, leveraging digital fluency to become more health literate and actively taking part in decision-making [57]. However, expectations regarding who should lead the consultation may also reflect previous health care experiences in cocreators' countries of origin. Most cocreators in the present study originated from Asia and Africa, where shared decision-making often differs from the Western European and North American model. Across many South Asian settings, patients commonly value involvement while expecting physicians to provide recommendations and assume greater responsibility for directing consultations [58-60]. Evidence from African settings is more heterogeneous, but similarly suggests that clinician authority remains influential in many contexts, while patients continue to value involvement in decisions [61,62]. Taken together, these findings suggest that the pattern observed in this study may reflect the combined influence of being a young adult seeking active involvement in health care while drawing on previous experiences of health care systems in which physicians are expected to take greater responsibility for guiding clinical discussions.

Findings from the survey data further showed how many cocreators felt misunderstood or not listened to, while qualitative findings provided insight into the perceived causes of these experiences, including language barriers, insufficient explanations, lack of empathy, and dismissive behavior. Across the focus groups, cocreators reported feeling emotionally vulnerable during primary care interactions. For international students, these experiences may be compounded by the transition to a new health care environment.

Cocreators described not only communication difficulties during consultations but also uncertainty about how health care services operate in Germany. This suggests that experiences of care are influenced by both being recent migrants and navigating an unfamiliar health care system independently for the first time. Emerging adulthood is characterized by increasing responsibility for managing one's own health care, including recognizing when to seek care, arranging appointments, selecting providers, communicating symptoms, and making health care decisions without parental support [63]. Previous research has shown that young adults commonly experience uncertainty about how health care systems operate, question whether their symptoms warrant seeking care, worry about whether their concerns will be taken seriously, and often lack the health care knowledge and confidence developed through prior experience [64,65]. These challenges may be further amplified when young adults are simultaneously adapting to a new health care system with unfamiliar organizational structures, administrative processes, and communication norms. Previous studies in Hungary, Turkey, and the United States have reported similar findings: international students experienced language barriers, difficulties with appointment and insurance processes, perceived lack of dignity or insufficient care, and challenges navigating health care services [66-68]. These findings echo broader research documenting negative health care experiences among migrant groups in Germany, particularly concerning unclear communication, perceived indifference, cultural barriers and lack of dignity in care [69-72]. Complaints of dismissiveness and language barriers were among the most frequently coded responses, aligning with broader literature on migrant health care experiences in Germany [73,74].

This study confirms that in addition to provider demeanor, system-level needs and challenges also shape health care experiences of international students. While the surveys primarily characterized communication experiences and waiting room conditions, qualitative findings additionally identified broader system-level challenges, including difficulties arranging appointments, identifying suitable providers, and accessing digital infrastructure. These findings align with recent research showing that migrant groups in Germany often face structural barriers when navigating health care systems, including limited information about available services and administrative complexity [22]. While these challenges are common among migrant populations, they may have particular implications for international students, who must navigate unfamiliar health care processes independently for the first time. These findings highlight the importance of accessible, efficient, and supportive health care encounters for individuals who are unfamiliar with the health care system. Studies of digital health systems also suggest that insufficient integration of digital services and patient-facing tools within existing health care infrastructures can further limit accessibility and usability [14,75].

Cocreating an intervention with international students provided insight into how health care support may need to be tailored to this population. Cocreators preferred concise, practical, and multifunctional digital tools that could support both consultation preparation and broader health care needs.

Qualitative data also indicated that most cocreators actively prepared for consultations by taking notes and researching symptoms, with a motivation to improve communication efficiency and manage emotional stress. Preparation strategies were implemented just prior to appointments and were often used to prompt physicians during the consultation. This mirrors findings from previous studies highlighting that patients who prepare in advance are more likely to participate actively in clinical conversations [45,76]. While cocreators accepted shared responsibility for communication, they generally expected more engagement and empathy from providers, reinforcing the need for bidirectional communication efforts [77,78].

XSL•FO
RenderX

Feedback on the current DiPPT was mixed. Integrated qualitative and quantitative data suggested that perceptions of the intervention were influenced by both the content and delivery format of the lessons. The highest-rated lessons addressed challenges frequently identified by cocreators throughout the workshops, including obtaining clear information, communicating concerns, and preparing for consultations. In contrast, lower-rated lessons were perceived as more burdensome, less relevant to cocreators' immediate needs, or as placing additional responsibility on patients rather than health care providers. This latter finding converges with cocreators' views regarding communication responsibility, with cocreators generally expecting physicians to take the lead role in facilitating effective health care communication. Qualitative data further indicated dissatisfaction with lengthy lessons, quizzes, and open-ended exercises, while quantitative findings showed that the highest-rated lessons were among the shortest and least demanding components of the intervention.

Together, the data suggest that intervention acceptability may depend not only on addressing communication challenges experienced by patients, but also on presenting support in a concise, practical, and low-burden format that aligns with cocreators' expectations of the doctor–patient relationship. Similar to findings in the literature [11,45], digital interventions that incorporate personalized feedback and patient-centered support are more likely to be adopted. However, many cocreators expressed skepticism about routine tool use, emphasizing that added features were necessary to drive engagement [79,80].

Implications

The BCW was used to derive design specifications for a DiPPT tailored to international students [81]. Cocreation insights were mapped onto COM-B determinants (Capability, Opportunity, and Motivation) to identify key behavioral drivers of patient preparedness (the target behavior) among international students (the target audience). Capability-related needs included structured communication support (eg, checklists), opportunity-related factors reflected physician encouragement and system integration, and motivation-related factors emphasized interactive and multifunctional features (eg, chatbots). The relationship between these determinants and resulting intervention features is summarized in Supplement 7 in Multimedia Appendix 7.

As a potential implementation pathway, the intervention may be embedded into university orientation programs and distributed via clinic-based pamphlets, thereby normalizing the practice of patient preparation among students. Moreover, orientation programs may benefit from providing international students with clearer information about how the German health care system functions. This approach is echoed by Abdulai and colleagues [82], who emphasize the importance of considering contextual factors and leveraging existing university support structures when supporting international students. Informing students about available health care services, as well as existing tools and resources for finding clinics, booking appointments, and accessing digital health services, may help alleviate some of the navigation challenges reported by cocreators in the workshops. Taken together, these findings offer a behaviorally grounded and contextually responsive foundation for intervention development.

This study illustrates the value of combining behavioral science frameworks such as the BCW and COM-B model with systematic cocreation methodologies, including the Health CASCADE guidelines [83] and the PRODUCES framework [38]. The PRODUCES framework helped in conceptualizing the problem, communicating the objective and design of, as well as identifying the cocreators to recruit as part of the workshops, and in determining the scaling strategy to use. The BCW played a key role in shaping the sequence and focus of the workshops, facilitating the identification of end-user needs and behavior change determinants. The design of the workshops was also guided by the work of Health CASCADE, who have developed cocreation guidelines [83], to create an inclusive and transparent process and ensure that cocreators' voices were meaningfully integrated into the intervention's design. By combining these methodologies with the BCW and COM-B, the study provides a replicable and scalable roadmap for developing digital health interventions that are grounded in theory and shaped by the lived experiences of users by means of acting as a role model. Together, these frameworks support the creation of interventions that are sensitive to real-world user needs, anchored in validated behavior change methods, and designed with sustainability in mind through participatory development practices.

Feasibility

While the previous section outlines the behavioral determinants and design specifications of the intervention, questions remain regarding its feasibility and potential adoption in real-world settings. Although cocreators expressed preferences for an intervention that would support doctor-patient communication, several indicated that they would likely use such a tool only once or before particularly important medical appointments.

The data also highlighted challenges relating to the digital medium. Together, the limited connectivity and skepticism towards a standalone communication app suggest that the intervention may function primarily as a situational preparation aid rather than a tool intended for continuous engagement. This pattern aligns with broader findings in digital health research showing that initial interest in digital tools often does not necessarily translate into sustained use [84,85]. Many digital health interventions experience rapid declines in engagement following early exposure, even when users report positive perceptions of their usefulness [86,87].

Adoption of digital health tools is also influenced by the broader environment in which they are deployed. Cocreators highlighted that the perceived value of a preparedness tool would increase if it addressed system-level challenges, such as appointment scheduling, clinic navigation, or access to medical documentation. Prior research has shown that digital health interventions are more likely to be adopted when integrated into existing care infrastructures rather than functioning as standalone behavior change tools [88,89].

XSL•FO
RenderX

If the intervention is primarily used as a one-off preparation tool, its effectiveness may therefore depend on where it is embedded within existing systems, such as appointment booking platforms, university health services, or patient portals. Future work should therefore explore which digital tools and communication channels international students currently rely on when navigating health services in Germany, as this may reveal more feasible points of integration. In some contexts, a lightweight intervention embedded within existing platforms or informational resources may be more appropriate than a standalone application.

Limitations

To enhance cocreation and better understand the mutual communication between physicians and patients, future studies should consider involving physicians and clinic administrators as cocreators in the process. Additionally, the coding frame used for qualitative content analysis was limited in distinguishing between specific content, design, and functional preferences for digital tools. Future iterations should further refine coding categories with respect to specific aspects of the DiPPT that should be retained or excluded, and further delineating the miscellaneous categories for cocreator needs of a digital health intervention.

Furthermore, quantitative data were collected anonymously, preventing direct linkage with qualitative insights by person. Future designs should link cocreator data to deepen the understanding of tool acceptability among distinct user profiles.

Finally, cocreation prioritizes depth of insight over statistical representativeness, so the study involved a small sample of international students (n=12), homogeneous with respect to age (19 - 22 y), enrolled at a single English-speaking international university in Germany, limiting generalizability. Therefore, the findings presented here should not be interpreted as reflecting the needs of the broader migrant population in Germany. Moreover, the extent to which the findings are specific to international students, reflect broader migrant experiences, or arise from the intersection of student and migrant identities remains unclear. Further research is needed to disentangle these influences and examine the transferability of the findings to other populations. However, the methodological approach may be transferable to other contexts, and future research should extend this approach to diverse academic institutions, age groups, and include a wider range of cultural backgrounds.

The detailed reporting of the cocreation procedures and the transparent presentation of the coding frame provide a foundation that future researchers can use to inform the design of workshops with other migrant subgroups and settings.

Conclusion

In conclusion, this study identified key communication and system navigation challenges faced by international students in German primary care and translated these into design specifications for a tailored DiPPT. By integrating cocreation methods with the BCW and COM-B model, the study offers a foundation for future intervention development.

Acknowledgments

During the preparation of this manuscript, the authors used ChatGPT (OpenAI, GPT-5.3) to assist with improving the readability and clarity of selected text. All outputs were carefully reviewed, revised, and validated by the authors, who take full responsibility for the content of the publication.

Funding

This work was supported by the Health CASCADE network. Health CASCADE is a Marie Skłodowska-Curie Innovative Training Network funded by the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 956501.

Data Availability

The data supporting the findings of this study are not publicly available due to the sensitive nature of the qualitative data, which include personal health experiences and high-risk identifiers that could compromise co-creator confidentiality. Requests for access to anonymized excerpts or further details for academic purposes may be considered on a case-by-case basis and should be directed to the corresponding author at vinayak.anand-kumar@pm.me. Access will be granted only if deemed appropriate and in alignment with ethical approval and data protection regulations.

Authors' Contributions

Conceptualization: VAK.
Data curation: VAK, MO, FZB, SG, TNHQ.
Formal analysis: VAK, MS.
Investigation: VAK, MO, FZB, SG, TNHQ
Methodology: VAK, MS, MO, FZB, SG, TNHQ, LA.
Project administration: VAK.
Supervision: MS, MG-G, SL
Validation: MS.
Visualization: VAK.
Writing – original draft: VAK.

XSL•FO
RenderX

Writing – review & editing: VAK, MS, MG-G, SL, MO, FZB, SG, TNHQ, LA.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Tools used to collect data from cocreators.
[DOCX File, 23 KB - humanfactors_v13i1e87519_app1.docx]

Multimedia Appendix 2
Visual summaries of cocreator demographics and health status.
[DOCX File, 1352 KB - humanfactors_v13i1e87519_app2.docx]

Multimedia Appendix 3
Final coding frame used for qualitative content analysis.
[DOCX File, 35 KB - humanfactors_v13i1e87519_app3.docx]

Multimedia Appendix 4
Interrater agreement and coding frame validity metrics.
[DOCX File, 16 KB - humanfactors_v13i1e87519_app4.docx]

Multimedia Appendix 5
Visual summaries of health care experiences and preparation behaviors.
[DOCX File, 2927 KB - humanfactors_v13i1e87519_app5.docx]

Multimedia Appendix 6
Visual summaries of needs and preferences for digital health tool design.
[DOCX File, 846 KB - humanfactors_v13i1e87519_app6.docx]

Multimedia Appendix 7
Detailed specifications for a health behavior change intervention for international students in Germany based on the cocreation workshops.
[DOCX File, 21 KB - humanfactors_v13i1e87519_app7.docx]

References

1. Sharkiya SH. Quality communication can improve patient-centred health outcomes among older patients: a rapid review. BMC Health Serv Res 2023 Aug 22;23(1):886. [doi: 10.1186/s12913-023-09869-8] [Medline: 37608376]
2. Matusitz J, Spear J. Effective doctor–patient communication: an updated examination. Soc Work Public Health 2014;29(3):252-266. [doi: 10.1080/19371918.2013.776416] [Medline: 24802220]
3. Fiorini F, Granata A. Doctor-patient communication. G Ital Nefrol 2019 Apr;36(2):2019-vol2. [Medline: 30983173]
4. Falkenstein L, Eckel N, Kadel SB, Koenig J, Litaker D, Eichinger M. Service provision and utilisation in German paediatric primary care practices during public health crises: protocol of the mixed-methods COVID-19 PedCare Study. BMJ Open 2022 Oct 11;12(10):e054054. [doi: 10.1136/bmjopen-2021-054054] [Medline: 36220317]
5. Luppa M, Giersdorf J, Riedel-Heller S, Prütz F, Rommel A. Frequent attenders in the German healthcare system: determinants of high utilization of primary care services. Results from the cross-sectional German health interview and examination survey for adults (DEGS). BMC Fam Pract 2020 Jan 13;21(1):10. [doi: 10.1186/s12875-020-1082-9] [Medline: 31931727]
6. Stark S, Kluge M, Schaubroeck E, et al. “What else to say?”-Primary health care in times of COVID-19 from the perspective of German general practitioners: an exploratory analysis of the open text field in the PRICOV-19 study. PLoS ONE 2023;18(3):e0282504. [doi: 10.1371/journal.pone.0282504] [Medline: 36930662]
7. Sun N, Rau PLP. Barriers to improve physician–patient communication in a primary care setting: perspectives of Chinese physicians. Health Psychol Behav Med 2017 Jan;5(1):166-176. [doi: 10.1080/21642850.2017.1286498]
8. Rocque R, Leanza Y. A systematic review of patients’ experiences in communicating with primary care physicians: intercultural encounters and a balance between vulnerability and integrity. PLoS ONE 2015;10(10):e0139577. [doi: 10.1371/journal.pone.0139577] [Medline: 26440647]
9. Zeisler ML, Bilgic L, Schumann M, et al. Interventions to increase the reachability of migrants in Germany with health interview surveys: mixed-mode feasibility study. JMIR Form Res 2020 Apr 15;4(4):e14747. [doi: 10.2196/14747] [Medline: 32293576]

XSL•FO
RenderX

10. Olsavszky V, Bazari M, Dai TB, et al. Digital translation platform (Translatly) to overcome communication barriers in clinical care: pilot study. JMIR Form Res 2025 Mar 14;9:e63095. [doi: 10.2196/63095] [Medline: 39451122]
11. Seljelid B, Varsi C, Solberg Nes L, Øystese KA, Børøsund E. A digital patient-provider communication intervention (InvolveMe): qualitative study on the implementation preparation based on identified facilitators and barriers. J Med Internet Res 2021 Apr 8;23(4):e22399. [doi: 10.2196/22399] [Medline: 33830063]
12. Noack EM, Schulze J, Müller F. Designing an app to overcome language barriers in the delivery of emergency medical services: participatory development process. JMIR Mhealth Uhealth 2021 Apr 14;9(4):e21586. [doi: 10.2196/21586] [Medline: 33851933]
13. Nordfeldt S, Ängarne-Lindberg T, Berterö C. To use or not to use--practitioners' perceptions of an open web portal for young patients with diabetes. J Med Internet Res 2012 Nov 9;14(6):e154. [doi: 10.2196/jmir.1987] [Medline: 23137767]
14. Brands MR, Gouw SC, Beestrum M, Cronin RM, Fijnvandraat K, Badawy SM. Patient-centered digital health records and their effects on health outcomes: systematic review. J Med Internet Res 2022 Dec 22;24(12):e43086. [doi: 10.2196/43086] [Medline: 36548034]
15. Summary #smarthealthsystems: focus europe [Webpage in German]. Bertelsmann Stiftung. URL: https://www.bertelsmann-stiftung.de/en/publications/publication/did/summary-smarthealthsystems-focus-europe [accessed 2025-07-16]
16. Caumanns J. For discussion: the state of digitization of the German healthcare system. Z Evid Fortbild Qual Gesundhwes 2019 Jun;143:22-29. [doi: 10.1016/j.zefq.2019.04.002] [Medline: 31204267]
17. Doshi S, Garg S, Kakkar S, et al. Evidence-based best practices for patient engagement using digital navigation pathways: insights from over 800,000 digital prescriptions. Cardiovasc Digit Health J 2022 Aug;3(4):S6. [doi: 10.1016/j.cvdhj.2022.07.016]
18. Atherton H, Eccles A, Poltawski L, Dale J, Campbell J, Abel G. Investigating patient use and experience of online appointment booking in primary care: mixed methods study. J Med Internet Res 2024 Jul 8;26:e51931. [doi: 10.2196/51931] [Medline: 38976870]
19. Maita KC, Maniaci MJ, Haider CR, et al. The impact of digital health solutions on bridging the health care gap in rural areas: a scoping review. Perm J 2024 Sep 16;28(3):130-143. [doi: 10.7812/TPP/23.134] [Medline: 39135461]
20. Senek M, Drummond D, Pinnock H, et al. Impact of digital health on patient-provider relationships in respiratory secondary care based on qualitative and quantitative evidence: systematic review. J Med Internet Res 2025 May 30;27:e70970. [doi: 10.2196/70970] [Medline: 40446293]
21. Zhou Z, Jin D, He J, et al. Digital health platform for improving the effect of the active health management of chronic diseases in the community: mixed methods exploratory study. J Med Internet Res 2024 Nov 18;26:e50959. [doi: 10.2196/50959] [Medline: 39556830]
22. Davitian K, Noack P, Eckstein K, Hübner J, Ahmadi E. Barriers of Ukrainian refugees and migrants in accessing German healthcare. BMC Health Serv Res 2024 Sep 24;24(1):1112. [doi: 10.1186/s12913-024-11592-x] [Medline: 39317924]
23. Fachkräftestrategie der bundesregierung. Bundesministerium für Arbeit und Soziales. URL: https://www.bmas.de/DE/Service/Publikationen/Broschueren/fachkraeftestrategie-der-bundesregierung.html [accessed 2025-07-16]
24. Wolter A. Migration and higher education in germany. In: Slowey M, Schuetze HG, Zubrzycki T, editors. Inequality, Innovation and Reform in Higher Education: Challenges of Migration and Ageing Populations: Springer; 2020:39-57. [doi: 10.1007/978-3-030-28227-1_3]
25. Tran NTM, Cameron MP, Poot J. Perception of institutional quality differences and intention of migrants to return home: a case study of Vietnamese diaspora. Asia-Pac J Reg Sci 2022 Feb;6(1):213-237. [doi: 10.1007/s41685-021-00212-9]
26. OECD. International Migration Outlook 2022: OECD Publishing; 2022. [doi: 10.1787/30fe16d2-en]
27. Koop R, Kartounian H, Devroey D. Comparison of health-seeking characteristics of German and Belgian university students. J Med Life 2017;10(1):38-43. [Medline: 28255374]
28. Tillmann J, Puth MT, Frank L, Weckbecker K, Klaschik M, Münster E. Determinants of having no general practitioner in Germany and the influence of a migration background: results of the German health interview and examination survey for adults (DEGS1). BMC Health Serv Res 2018 Oct 3;18(1):755. [doi: 10.1186/s12913-018-3571-2] [Medline: 30285753]
29. Castañeda H. Illegality as risk factor: a survey of unauthorized migrant patients in a Berlin clinic. Soc Sci Med 2009 Apr;68(8):1552-1560. [doi: 10.1016/j.socscimed.2009.01.024] [Medline: 19246143]
30. Tezcan-Güntekin H. Racially discriminatory experiences among older chronically ill people of Turkish origin in Germany. Eur J Public Health 2022 Oct 21;32(Supplement_3). [doi: 10.1093/eurpub/ckac129.730]
31. Charbonneau DH, Hightower S, Katz A, et al. Smartphone apps for cancer: a content analysis of the digital health marketplace. Digit Health 2020;6:2055207620905413. [doi: 10.1177/2055207620905413] [Medline: 32110428]
32. Berdahl CT, Henreid AJ, Pevnick JM, Zheng K, Nuckols TK. Digital tools designed to obtain the history of present illness from patients: scoping review. J Med Internet Res 2022 Nov 17;24(11):e36074. [doi: 10.2196/36074] [Medline: 36394945]
33. Li J, Luo X, Cao Q, Lin Y, Xu Y, Li Q. Communication needs of cancer patients and/or caregivers: a critical literature review. J Oncol 2020;2020:7432849. [doi: 10.1155/2020/7432849] [Medline: 32454826]
34. Chauhan BF, Jeyaraman M, Mann AS, et al. Behavior change interventions and policies influencing primary healthcare professionals' practice—an overview of reviews. Implementation Sci 2017 Dec;12(1):3. [doi: 10.1186/s13012-016-0538-8]

XSL•FO
RenderX

35. Malloy JA, Partridge SR, Kemper JA, Braakhuis A, Roy R. Feasibility and preliminary efficacy of co-designed and co-created healthy lifestyle social media intervention programme the daily health coach for young women: a pilot randomised controlled trial. Nutrients 2024 Dec 18;16(24):4364. [doi: 10.3390/nu16244364] [Medline: 39770984]
36. Maenhout L, Latomme J, Cardon G, Crombez G, Van Hove G, Compernolle S. Synergizing the behavior change wheel and a cocreative approach to design a physical activity intervention for adolescents and young adults with intellectual disabilities: development study. JMIR Form Res 2024 Jan 11;8:e51693. [doi: 10.2196/51693] [Medline: 38206648]
37. Parsons K, Payne S, Codling S, Murphy M. A reflection on the co-design approach to the development of the Medway Can campaign; a whole systems approach to obesity prevention using COM-B. Health Mark Q 2024;41(1):33-49. [doi: 10.1080/07359683.2023.2235501] [Medline: 37486187]
38. Leask CF, Sandlund M, Skelton DA, et al. Framework, principles and recommendations for utilising participatory methodologies in the co-creation and evaluation of public health interventions. Res Involv Engagem 2019;5(1):2. [doi: 10.1186/s40900-018-0136-9] [Medline: 30652027]
39. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions: Silverback Publishing; 2014. URL: https://www.behaviourchangewheel.com/ [accessed 2023-10-01]
40. Creswell JW, Designing P, Research CMM. Designing and Conducting Mixed Methods Research, 3rd edition: SAGE Publications; 2017. URL: https://www.sagepub.com/shop/buy-a-book/designing-and-conducting-mixed-methods-research-3-241842 [accessed 2023-10-01]
41. Lippke S, Wienert J, Keller FM, et al. Communication and patient safety in gynecology and obstetrics - study protocol of an intervention study. BMC Health Serv Res 2019 Nov 28;19(1):908. [doi: 10.1186/s12913-019-4579-y] [Medline: 31779620]
42. Kötting L, Anand-Kumar V, Keller FM, Henschel NT, Lippke S. Effective communication supported by an app for pregnant women: quantitative longitudinal study. JMIR Hum Factors 2024 Apr 26;11:e48218. [doi: 10.2196/48218] [Medline: 38669073]
43. Plano Clark VL, Garrett AL, Leslie-Pelecky DL. Applying three strategies for integrating quantitative and qualitative databases in a mixed methods study of a nontraditional graduate education program. Field methods 2010 May;22(2):154-174. [doi: 10.1177/1525822X09357174]
44. Fetters MD, Curry LA, Creswell JW. Achieving integration in mixed methods designs-principles and practices. Health Serv Res 2013 Dec;48(6 Pt 2):2134-2156. [doi: 10.1111/1475-6773.12117] [Medline: 24279835]
45. Derksen C, Dietl JE, Haeussler FE, Steinherr Zazo M, Schmiedhofer M, Lippke S. Behavior change training for pregnant women's communication during birth: a randomized controlled trial. Applied Psych Health & Well 2023 Aug;15(3):865-883. [doi: 10.1111/aphw.12413]
46. Kötting L, Derksen C, Keller FM, Lippke S. Comparing the effectiveness of a web-based application with a digital live seminar to improve safe communication for pregnant women: 3-group partially randomized controlled trial. JMIR Pediatr Parent 2023 Jul 24;6:e44701. [doi: 10.2196/44701] [Medline: 37486755]
47. Schreier M. Sampling and generalization. In: Flick U, editor. The SAGE Handbook of Qualitative Data Collection: SAGE Publications; 2018:84-97. [doi: 10.4135/9781526416070.n6]
48. Agnello DM, Balaskas G, Steiner A, Chastin S. Methods used in co-creation within the health CASCADE co-creation database and gray literature: systematic methods overview. Interact J Med Res 2024 Nov 11;13:e59772. [doi: 10.2196/59772] [Medline: 39527793]
49. The R Project for Statistical Computing. URL: https://www.R-project.org [accessed 2026-08-29]
50. Posit. URL: https://posit.co [accessed 2024-10-13]
51. Co-creating a digital patient preparedness tool: applications in co-creation and behaviour change. Open Science Framework. URL: https://osf.io/jbeap [accessed 2025-11-10]
52. Nicolai T, Mozygemba K. User Manual: QualiAnon—Tool for the Anonymization of Text Data (Version 13): Universität Bremen; 2023. [doi: 10.26092/elib/2576]
53. Schreier M. Qualitative Content Analysis in Practice: SAGE Publications; 2012. URL: https://www.sagepub.com/shop/buy-a-book/qualitative-content-analysis-in-practice-1-234633 [accessed 2021-11-01]
54. Curtain C. QualCoder version 3.8.2. GitHub. 2023. URL: https://github.com/ccbogel/QualCoder/releases [accessed 2024-10-13]
55. Schuiteman S, Chua KP, Plegue MA, Ilyas O, Chang T. Self-management of health care among youth: implications for policies on transitions of care. J Adolesc Health 2020 May;66(5):616-622. [doi: 10.1016/j.jadohealth.2020.01.009] [Medline: 32113903]
56. Wilson KL, Flores S, Apata BO, Tasnim S, Garney WR, Ajayi KV. Understanding healthcare autonomy among adolescents and young adults in the United States: a scoping review. Front Health Serv 2025;5:1720972. [doi: 10.3389/frhs.2025.1720972] [Medline: 41601473]
57. Paquin MA. Generation Z's experience making health care decisions. : The Chicago School of Professional Psychology; 2025 URL: https://www.proquest.com/openview/2d1fc92034e80363642e22400c32f555/1.pdf?cbl=18750&diss=y&pq-origsite=gscholar [accessed 2026-08-01]

XSL•FO
RenderX

58. Ahmed S, Jafri H, Rashid Y, et al. Autonomous decision-making for antenatal screening in Pakistan: views held by women, men and health professionals in a low-middle income country. Eur J Hum Genet 2019 Jun;27(6):848-856. [doi: 10.1038/s41431-019-0353-1] [Medline: 30718884]
59. Vakil K, Desse TA, Manias E, et al. Patient-centered care experiences of first-generation, South Asian migrants with chronic diseases living in high-income, Western countries: systematic review. Patient Prefer Adherence 2023;17:281-298. [doi: 10.2147/PPA.S391340] [Medline: 36756536]
60. Gopichandran V. Shared decision making in an unequal world. Indian J Med Ethics 2022;8(2):153-154. [doi: 10.20529/IJME.2022.045]
61. Eales OO, Smith S. Do socio-economically disadvantaged patients prefer shared decision-making? S Afr Fam Pract 2021;63(1):e1-e6. [doi: 10.4102/safp.v63i1.5293]
62. Owino RS, Kituuka O, Kutyabami P, Sewankambo NK. Shared decision-making between patients and healthcare providers at rural health facilities in Eastern Uganda: an exploratory qualitative study. BMC Med Ethics 2025 Jan 27;26(1):13. [doi: 10.1186/s12910-025-01172-x] [Medline: 39871240]
63. Wright JA, Volkman JE, Leveille SG, Amante DJ. Predictors of online patient portal use among a diverse sample of emerging adults: cross-sectional survey. JMIR Form Res 2022 Feb 15;6(2):e33356. [doi: 10.2196/33356] [Medline: 35166686]
64. Viktorsson L, Törnvall E, Falk M, Wåhlin I, Yngman-Uhlin P. Young adults' needs when seeking first-line healthcare: a grounded theory design. PLoS ONE 2022;17(2):e0263963. [doi: 10.1371/journal.pone.0263963] [Medline: 35167607]
65. Varty M, Popejoy LL. A systematic review of transition readiness in youth with chronic disease. West J Nurs Res 2020 Jul;42(7):554-566. [doi: 10.1177/0193945919875470] [Medline: 31530231]
66. Atiku LYL, Adofo E. Access to healthcare: a promise fulfilled? A hoax? Or, a matter of control? A qualitative study into the experiences of international students in Hungary. Front Public Health 2025;13:1635089. [doi: 10.3389/fpubh.2025.1635089] [Medline: 41211379]
67. Tang C, Gui X, Chen Y, Magueramane M. New to a country: barriers for international students to access health services and opportunities for design. Presented at: Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 21-24, 2018. [doi: 10.1145/3240925.3240969]
68. Masai AN, Güçiz-Doğan B, Ouma PN, Nyadera IN, Ruto VK. Healthcare services utilization among international students in Ankara, Turkey: a cross-sectional study. BMC Health Serv Res 2021 Apr 7;21(1):311. [doi: 10.1186/s12913-021-06301-x] [Medline: 33827556]
69. Hogg R, Hanley J, Smith PA. Learning lessons from the analysis of patient complaints relating to staff attitudes, behaviour and communication, using the concept of emotional labour. J Clin Nurs 2018 Mar;27(5-6):e1004-e1012. [doi: 10.1111/jocn.14121] [Medline: 29052343]
70. Hadziabdic E, Higginbottom G. Complaints concerning communication reported by users of healthcare in a specific region in Sweden. J Hosp Adm 2015;4(3):39. [doi: 10.5430/jha.v4n3p39]
71. Karaduman ME, Hocaoğlu A, Sabak M, Zengi˙n S. Analysis of the complaints of the patients and their relatives to healthcare professionals. J Contemp Med Sci 2023;13(1):72-77. [doi: 10.16899/jcm.1218472]
72. Forray AI, Oltean O, Hanft-Robert S, et al. Uncovering multi-level mental healthcare barriers for migrants: a qualitative analysis across China, Germany, Netherlands, Romania, and South Africa. BMC Public Health 2024 Jun 14;24(1):1593. [doi: 10.1186/s12889-024-19046-z] [Medline: 38877460]
73. Langer T, Zapf T, Wirth S, et al. How are pediatric hospitals in North-Rhine Westfalia prepared to overcome language barriers? A pilot study exploring the structural quality of inpatient care. Gesundheitswesen 2017 Jul;79(7):535-541. [doi: 10.1055/s-0042-102349] [Medline: 27144710]
74. Priebe S, Sandhu S, Dias S, et al. Good practice in health care for migrants: views and experiences of care professionals in 16 European countries. BMC Public Health 2011 Dec;11(1):187. [doi: 10.1186/1471-2458-11-187]
75. Seljelid B, Varsi C, Solberg Nes L, Stenehjem AE, Bollerslev J, Børøsund E. Content and system development of a digital patient-provider communication tool to support shared decision making in chronic health care: InvolveMe. BMC Med Inform Decis Mak 2020 Mar 4;20(1):46. [doi: 10.1186/s12911-020-1065-8] [Medline: 32131808]
76. Kehl KL, Landrum MB, Arora NK, et al. Association of actual and preferred decision roles with patient-reported quality of care: shared decision making in cancer care. JAMA Oncol 2015 Apr;1(1):50-58. [doi: 10.1001/jamaoncol.2014.112] [Medline: 26182303]
77. Kiani FZ, Ahmadi A. Barriers and facilitating factors of communication in Iranian educational health care centers: a systematic review. Strides Dev Med Educ 2019;In Press(In Press). [doi: 10.5812/sdme.80871]
78. Singh H, Dey AK. Listen to my story: contribution of patients to their healthcare through effective communication with doctors. Health Serv Manage Res 2021 Aug;34(3):178-192. [doi: 10.1177/0951484820952308] [Medline: 32903093]
79. Almaiah MA, Jalil MA, Man M. Extending the TAM to examine the effects of quality features on mobile learning acceptance. J Comput Educ 2016 Dec;3(4):453-485. [doi: 10.1007/s40692-016-0074-1]
80. Koo C, Chung N, Nam K. Assessing the impact of intrinsic and extrinsic motivators on smart green IT device use: reference group perspectives. Int J Inf Manage 2015 Feb;35(1):64-79. [doi: 10.1016/j.ijinfomgt.2014.10.001]

XSL•FO
RenderX

81. Michie S. Implementation science: understanding behaviour change and maintenance. BMC Health Serv Res 2014 Jul;14(S2). [doi: 10.1186/1472-6963-14-S2-O9]
82. Abdulai M, Roosalu T, Wagoner B. Cultural barriers and enablers of integrating educational migrants from the Global South: the case of graduate students in Europe. Int J Educ Dev 2021 Oct;86:102479. [doi: 10.1016/j.ijedudev.2021.102479]
83. Chastin SFM, Smith N, Agnello DM, et al. Principles and attributes of evidence-based co-creation: from naïve praxis toward a trustworthy methodology - A Health CASCADE study. Public Health 2025 Nov;248:105922. [doi: 10.1016/j.puhe.2025.105922] [Medline: 40992098]
84. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
85. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 6;19(12):e402. [doi: 10.2196/jmir.8578] [Medline: 29212630]
86. Short CE, DeSmet A, Woods C, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292. [doi: 10.2196/jmir.9397] [Medline: 30446482]
87. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567. [doi: 10.2196/14567] [Medline: 31573916]
88. Zanaboni P, Fagerlund AJ. Patients' use and experiences with e-consultation and other digital health services with their general practitioner in Norway: results from an online survey. BMJ Open 2020 Jun 17;10(6):e034773. [doi: 10.1136/bmjopen-2019-034773] [Medline: 32554721]
89. Senek M, Long J, Ohlsen S, et al. Factors affecting the retention of healthcare assistants in English mental health services: a qualitative interview study. BMC Health Serv Res 2025 Apr 5;25(1):505. [doi: 10.1186/s12913-025-12665-1] [Medline: 40188051]

Abbreviations

BCT: behavior change technique
BCW: behavior change wheel
COM-B: Capability, Opportunity, Motivation – Behavior
DHI: digital health intervention
DiPPT: Digital Patient Preparedness Tool
GP: general practitioner
HAPA: health action process approach
***ICD-10*:** *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision*
OECD: Organization for Economic Cooperation and Development
OSF: Open Science Framework
QCA: qualitative content analysis

Edited by S Law; submitted 10.Nov.2025; peer-reviewed by M Katapodi, T Jimoh; revised version received 07.Aug.2026; accepted 13.Aug.2026; published 14.Sep.2026.

Please cite as:
Anand Kumar V, Schreier M, Giné-Garriga M, Ortmann M, Belmokhtar FZ, Grineviciute S, Quan TNH, Aluru L, Lippke S
Cocreating a Digital Patient Preparedness Tool for International Students Accessing Primary Care in Germany: A Convergent Methods Study Within the Health CASCADE Network
JMIR Hum Factors 2026;13:e87519
URL: https://humanfactors.jmir.org/2026/1/e87519
doi:10.2196/87519

© Vinayak Anand Kumar, Margrit Schreier, Maria Giné-Garriga, Maria Ortmann, Fatima-Zohra Belmokhtar, Simona Grineviciute, Tran Ngoc-Huong Quan, Likhita Aluru, Sonia Lippke. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 14.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Association of eHealth Literacy With Cognitive Function, Frailty, and Functional Status in Hospitalized Geriatric Patients: Cross-Sectional Observational Study

Chantal Flemm[1,2], MSc; Rainer Wirth[2], Prof Dr Med; Sarah El Bataioui[3], MSc; Diana Daubert[2], BSc; Ida Haussleiter, Dr med; Matthias Tristan Hermenau[2], BA; Jan Dieris-Hirche[3], Dr med; Judith Frederica Kuhlmann[2]; Katharina Kunde[4], MA; Ulrike Trampisch[2], Dr sportwiss; Bianca Ueberberg[5], Dr rer nat; Horst Christian Vollmar[4], Prof Dr Med; Theresa Sophie Busse[1], Prof Dr

[1]Digital Health, Department of Human Medicine, Faculty of Health, Witten/Herdecke University, Alfred-Herrhausen-Straße 50, Witten, North Rhine-Westphalia, Germany

[2]Department of Geriatric Medicine, Marien Hospital Herne, University Hospitals of the Ruhr-University of Bochum, Herne, North Rhine-Westphalia, Germany

[3]Department of Psychosomatic Medicine and Psychotherapy, LWL-Universitätsklinikum Bochum, Bochum, North Rhine-Westphalia, Germany

[4]Institute of General Practice and Family Medicine (AM RUB), Medical Faculty, Ruhr University Bochum, Bochum, North Rhine-Westphalia, Germany

[5]Department of Psychiatry and Psychotherapy, LWL-Universitätsklinikum Bochum, Bochum, North Rhine-Westphalia, Germany

Corresponding Author:
Chantal Flemm, MSc
Digital Health, Department of Human Medicine, Faculty of Health, Witten/Herdecke University, Alfred-Herrhausen-Straße 50, Witten, North Rhine-Westphalia, Germany

Abstract

Background: The rapid growth of digital health technologies has highlighted the importance of eHealth literacy, which is crucial for navigating health information and tools such as telemedicine and eHealth records. However, older adults often have reservations about digital technologies and face barriers due to cognitive, sensory, and motor impairments, which hinder their ability to effectively use digital health applications. Despite its relevance, research on eHealth literacy in older populations is limited, with a particular gap in understanding how self-assessments of eHealth literacy correlate with health characteristics.

Objective: The study seeks to explore eHealth literacy among hospitalized geriatric patients in Germany. The main objective is to evaluate the eHealth literacy level within this particular group and identify the clinical and sociodemographic factors that may influence it.

Methods: Data from this observational study were analyzed to explore characteristics influencing eHealth literacy in hospitalized geriatric patients. The study included patients aged 60 or older in the Department of Geriatric Medicine at Marien Hospital Herne and enrolled in the standardized early rehabilitation pathway. eHealth literacy was measured using the German version of the eHealth Literacy Scale (GR-eHEALS), a self-assessment tool. Sociodemographic and clinical data, including cognitive function (Montreal Cognitive Assessment [MoCA]), frailty (Clinical Frailty Scale [CFS]), and depression (Depression in Old Age Scale [DIA-S]), were collected. Statistical analyses involved descriptive statistics, *t* tests, chi-square tests, and regression analyses to examine the relationship between eHealth literacy and various factors. Sociodemographic and clinical data, including cognitive function (MoCA), frailty (CFS), and depression (DIA-S), were collected.

Results: A total of 244 participants (aged 60 - 97, mean 77.2, SD 6.9 years) were included in the analysis. Participants generally reported an eHealth literacy, with higher scores in “information seeking” (mean 3.64, SD 0.84) compared to “information appraisal” (mean 3.32, SD 0.80). Significant differences in eHealth literacy were observed by age, with older participants scoring lower on both subscales. Frailty was significantly associated with critical appraisal of online health information (B=–1.372, 95% CI –2.554 to –0.189), suggesting that individuals who are frail struggle more with evaluating health information. No significant associations were found with other sociodemographic variables, including MoCA scores, education, or gender.

Conclusions: While the findings suggest a positive self-reported ability to engage with digital health tools, the selective sample and reliance on self-assessment limit the accuracy of these results. It also emphasizes the need for tailored interventions to support older adults in improving their eHealth literacy, particularly those with higher levels of frailty. Further research, including longitudinal studies and objective assessments, is essential to better understand and address the digital health needs of the geriatric population.

Trial Registration: German Clinical Trials Register DRKS00032931; https://drks.de/search/de/trial/DRKS00032931

XSL•FO
RenderX

(JMIR Hum Factors 2026;13:e92045) doi:10.2196/92045

KEYWORDS

eHealth literacy; digital health literacy; electronic health literacy; electronic health; digital health; geriatric patients; digital inclusion; electronic health information; geriatrics

Introduction

The use of digital tools is steadily increasing in both everyday life and health care [1,2]. This includes digital health offers like electronic health records, wearable devices to track health-related parameters, telemedicine, and smart hospitals, as well as AI [3]. Digital health, or eHealth, in this context, is defined as the use of modern information and communication technologies to deliver, manage, and support health care services, health information exchange, and clinical documentation across institutional and geographic boundaries. In the context of digital health care and telemedicine, eHealth encompasses a broad range of activities including electronic health records, telemedicine, remote patient monitoring, mobile health, and digital tools for health education, disease prevention, and patient engagement [4,5]. In addition to these primary known developments, there are innovations that so far only affect small areas of daily life, such as deep learning methods, virtual reality and augmented reality, as well as digital informed consent [3].

Since the German Digital Care Act was passed in 2019, digital health applications (DiGA) have been included in the catalog of services covered by statutory health insurance in Germany and can therefore be prescribed by physicians and psychotherapists [6]. Applications that meet the requirements of the Federal Institute for Drugs and Medical Devices (Bundesinstitut für Arzneimittel und Medizinprodukte [BfArM]) with regard to various factors such as safety, quality, and data protection and undergo an appropriate approval process are then intended to support insured persons in the treatment of illnesses and the compensation of impairments [7]. This regulation ensures that DiGA provide users with evidence-based information and support [8]. Health information is increasingly available on the internet, but misinformation is widespread there as well [9]. For example, AI provides evidence-based answers to health questions but also incorrect answers [10] as well as potentially harmful information [11].

Those innovations also raise challenges regarding data privacy and security. In addition, we note that challenges arise when systems are not interoperable, and data flow is not secure. Other organizational issues may include liability and malpractice, as well as regulatory challenges. Personal and psychological barriers are also an obstacle to the use of digitalization [3].

Therefore, as health care services increasingly transition to digital arrangements, individuals are required to develop new skills to use them and participate in health care [12]. Dealing with these challenges requires a wide range of personal skills, which are described under the term eHealth literacy and include analytical competencies (traditional literacy, media literacy, and information literacy) and context-specific competencies (computer literacy, scientific literacy, and health literacy) [13].

eHealth literacy is defined by Norman and Skinner [13] as the ability to seek, find, understand, evaluate, and apply health information from electronic sources to address or solve a health problem. The continuous advancement of digitalization has led to a wide range of additions to the definition by various authors, as well as an update to Norman and Skinner's definition [14]. These definitions focus primarily on critical perspectives on information and the relevance of individual health literacy, organizational health literacy, and navigational health literacy [15]. These skills are particularly crucial in the context of digital health care and telemedicine, where patients and health care providers alike rely on digital platforms for communication, diagnosis, and treatment options [16]. The importance of eHealth literacy extends beyond just accessing information. For patients, it is essential for self-management of health conditions, particularly in chronic diseases, where patients regularly need to make informed decisions [13,17]. However, studies have shown that these users in particular pose challenges: social and structural determinants of eHealth literacy in digital health care and telemedicine include socioeconomic status (income, education, and employment), age, ethnicity, social support, language proficiency, and access to digital infrastructure (such as broadband internet and devices). Lower income, lower educational attainment, older age, minority ethnicity, and unemployment are consistently associated with lower eHealth literacy scores, which in turn limits effective engagement with telemedicine and digital health tools [18-20]. In addition, lower levels of eHealth literacy may contribute to health disparities and limit the benefits of digital health technologies [19]. Low eHealth literacy is associated with increased barriers to telemedicine adoption, especially among older adults and socially disadvantaged groups, exacerbating health disparities and limiting the reach of digital health innovations [21].

Furthermore, higher levels of eHealth literacy have been associated with better health outcomes, as individuals with greater eHealth literacy are more likely to make informed and accurate health decisions, enhancing their overall well-being [19]. Higher levels of eHealth literacy are additionally associated with improved health-promoting behavior, better self-management, and more effective use of health services [19].

This shows that the level of eHealth literacy is highly relevant, while at the same time its degree of manifestation in society varies greatly. For health care providers, understanding a patient's level of eHealth literacy is key to delivering effective telemedicine services and ensuring patients can fully engage with digital health interventions [16]. Measurement of eHealth literacy is therefore an area in which a great deal of research is being conducted.

Despite its importance, the evidence on eHealth literacy among the older population remains incomplete, particularly regarding the factors associated with individual differences in eHealth

XSL•FO
RenderX

literacy [22]. This group is often at higher risk of difficulties navigating digital health tools [23,24], yet their specific challenges and needs have not been sufficiently addressed. In addition to chronological age, frailty—a multidimensional geriatric syndrome characterized by diminished physiological reserve and increased vulnerability to adverse health outcomes—is increasingly recognized as an important determinant of digital engagement in later life [25]. Frailty frequently coexists with cognitive impairment, functional limitations, and multimorbidity, all of which may compromise an individual's ability to access, understand, and effectively use digital health technologies [25,26]. However, differences are not solely attributed to age: cognitive, sensory (eg, tactile, visual, or auditory), and motor impairments directly affect one's ability to interact with and independently use technologies [27]. Cognitive barriers include declines in memory, attention, and information processing speed, as well as mild cognitive impairment and dementia, which can limit comprehension of instructions and the ability to navigate digital platforms, even when cognitive scores are within normal limits [28]. Functional barriers are common and include visual and hearing impairments, reduced manual dexterity (eg, arthritis), and sensorimotor changes, all of which hinder basic interactions with devices such as touchscreens and reading on small screens [27]. Motivational barriers encompass low self-efficacy, resistance to change, generational skepticism, privacy concerns, and a lack of perceived benefit or relevance of digital health tools. Many older adults report frustration, confusion, and mistrust in digital platforms, which can demotivate engagement [29,30]. Additional factors such as limited digital literacy, lack of awareness, complicated application interfaces, and insufficient support networks further exacerbate these barriers [31].

Nevertheless, people with cognitive impairments also use digital devices [32]. They expect time savings, report convenience and faster access to care as perceived advantages of digital health solutions [14]. However, in order to meet these positive expectations, usage must be successful, which may be difficult due to the obstacles mentioned above. Building on the recognition that digital health solutions can offer significant benefits, it is crucial to explore how these solutions can be used by those who face cognitive challenges, such as hospitalized geriatric patients.

This study aims to address this gap by examining eHealth literacy in a sample of hospitalized geriatric patients in Germany. The primary goal is to assess the level of eHealth literacy in this specific population and identify the clinical and sociodemographic characteristics that influence it.

Hypotheses of the study were:

- Self-reported eHealth literacy does vary by age.
- Participants with lower functional status and higher frailty report lower levels of eHealth literacy scores.

Methods

Study Design and Setting

The reporting of this observational study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [33]. The presented data are from an observational study of the DiGA4Aged project (dDiGA for older people with depressive disorders). The project consists of a pilot study [34], an observational study, and a randomized controlled trial (RCT) [35]. It aims to examine the feasibility, acceptance, and clinical relevance of a DiGA for older adults with depressive disorders.

The quantitative, observational study was designed to explore clinical and sociodemographic factors associated with digital media use and DiGA engagement among hospitalized geriatric patients. For this analysis, only data regarding eHealth literacy were used. Data collection took place in the Department of Geriatric Medicine at Marien Hospital Herne, Germany, and included both inpatient care and the geriatric day clinic. Throughout the study period (January to December 2024), all consecutively admitted patients were assessed for eligibility and invited to participate. Detailed methodological information on this screening study as well as the preceding pilot study is available in a previously published joint study protocol [36]. There have been no deviations from the planned and reported procedure.

Patients' Characteristics and Recruitment

Participants qualified for inclusion if they were at least 60 years old, admitted to the Department of Geriatric Medicine at Marien Hospital Herne in Germany, enrolled in the standardized geriatric early rehabilitation pathway (rehabilitation during acute care), which requires a minimum hospital stay of 14 days, and able to provide written informed consent. Patients were included if they reported using the internet to search for health information. Exclusion criteria included an anticipated hospital stay shorter than 14 days, insufficient cognitive ability to give informed consent, or refusal to participate. There were no restrictions concerning gender or specific medical diagnoses.

Recruitment was conducted consecutively throughout the study period of the entire year 2024. Eligible patients were identified through the hospital information system and personally contacted by trained research staff within the last three days before their planned discharge.

Study participation involved a 2-step process: first, the use of routinely collected clinical data, and second, the collection of additional information through standardized questionnaires for patients consenting to the extended data acquisition. For the reported analyses, only patients who completed both steps and the questionnaire regarding eHealth literacy were included.

Measurements

Trained study staff visited all patients with a tablet computer and conducted the survey in a structured interview format. Participants provided their responses verbally, and the study staff entered the answers directly into the electronic data capture system. The Montreal Cognitive Assessment (MoCA) was

XSL•FO
RenderX

administered separately using the standard paper-based format, as some tasks (eg, drawing) require written responses. eHealth literacy was measured using the German version of the eHealth Literacy Scale (GR-eHEALS) [37], the revised GR-eHEALS developed by Norman and Skinner [13]. The GR-eHEALS was first developed including 16 items, with the first 8 items corresponding to the original eHEALS, assessing individuals' ability to find, evaluate, and apply online health information. Items 9 - 16 were based on the eHEALS-E questionnaire from Petrič et al [38]. However, validation was only successful for items 1 - 8.

Responses were recorded on a 5-point Likert scale ranging from "strongly disagree" (1) to "strongly agree" (5). The scale includes 2 validated subscales: the Information Seeking scale (items 1 - 4) and the Information Appraisal scale (items 5 - 8). Both scales were scored by summing the respective item responses to generate a total score for each subscale [37]. This paper reports only data from the original eHEALS (items 1 - 8). The information seeking scale achieved a satisfactory Cronbach α of 0.92, and the information appraisal scale achieved a satisfactory Cronbach α of 0.92 and 0.83 [37].

The GR-eHEALS was chosen since it is a validated instrument to measure eHealth literacy. However, due to the complexity of the content dimensions, the operationalization of eHealth literacy is conceptually challenging. A systematic review by Xie et al [39] also highlighted the widespread use of eHEALS as a primary assessment tool for eHealth literacy. Moreover, a review by Huang et al [40] discussed various tools for measuring eHealth literacy in older adults, including eHEALS. Their findings suggest that while these tools are valuable, there is a need to strengthen their implementation in clinical practice. Based on the positive assessment by Huang et al [40] among the existing options, eHEALS was selected for this study. While self-assessment is a valuable tool for understanding an individual's perceived competency, it has inherent limitations. Subjective measures often overestimate actual eHealth literacy, as individuals may report high confidence without possessing the necessary skills to effectively navigate, evaluate, or apply online health information in real-world scenarios [41,42].

Furthermore, sociodemographic data from the geriatric assessment conducted routinely within the first 48 hours of admission: age, sex, years of education (self-reported), MoCA (to assess cognitive function) [43], Barthel Index (to assess functional status) [44], Clinical Frailty Scale (CFS) [45,46], living situation prior to admission (alone vs with others), Depression in Old Age Scale (DIA-S, to assess depression symptoms) [47], and care degree on admission. MoCA, Barthel Index, and DIA-S were assessed both at admission and at discharge. For this analysis, discharge values were used whenever available, as they were considered to better reflect the patients' stable cognitive, functional, and psychological status following recovery from the acute illness. If a discharge assessment was unavailable (eg, because the participant declined retesting), the corresponding admission value was used. Frailty was assessed using the CFS, a 9-point clinical scale ranging from 1 (nonfrail) to 9 (severely frail), with higher scores indicating greater levels of frailty and vulnerability [45,46]. For the purpose of statistical analyses, the CFS was dichotomized into nonfrail (CFS 1 - 4) and frail (CFS 5 - 9). The assessment was performed as part of routine geriatric clinical care by trained geriatric physicians and study nurses according to standardized procedures within the department.

The care degree is part of the German long-term care insurance system that regulates reimbursement for care and reflects the individual's level of care dependency based on impairments in independence across 6 domains of daily functioning. Care degrees range from 1 (minor impairment of independence) to 5 (most severe impairment requiring extensive care).

Ethical Considerations

The DiGA4Aged project received ethical approval from the Medical Faculty of Ruhr-University Bochum on November 17, 2023 (registration number 23 - 7901). The observational study was conducted according to the Declaration of Helsinki. Data were recorded and managed using the web-based electronic data capture system REDCap (Vanderbilt University) [48].

Data Handling and Statistical Analysis

No separate a priori sample size calculation or formal power analysis was conducted. The analyzed cohort originated from the DiGA4Aged screening study, which was primarily designed to identify and recruit eligible participants for the subsequent pilot study and RCT. Consequently, the duration and target sample size of the screening study were prospectively determined by the expected number of screened geriatric patients required to achieve the planned recruitment for the intervention study, as described in the published study protocol [36]. In accordance, all eligible geriatric patients admitted during the predefined one-year recruitment period were consecutively invited to participate. The current analyses represent an exploratory analysis of this screening cohort. All questions were administered in person during the participants' hospital stay. Statistical analyses were performed with SPSS statistical software (IBM Corp, SPSS Statistics for Windows, version 29.0).

Descriptive statistics were used to characterize the study sample. Continuous variables were summarized using medians, means, and SD. Categorical variables were presented as absolute and relative frequencies (n, %). To assess whether the distribution of values in the subsample differed from the distribution observed in the overall study population, different one-sample procedures were applied depending on the scale level of the variables. For metric variables, one-sample *t* tests were conducted to determine whether the mean of the subsample significantly deviated from the mean of the total sample. For categorical variables with exactly 2 response categories, one-sample binomial tests were performed. These tests evaluated whether the observed proportions in the subsample differed from those in the overall sample. For categorical variables with more than 2 categories, one-sample chi-square goodness-of-fit tests were used. In these analyses, the observed category frequencies of the subsample were compared with the expected frequencies derived from the distribution of the full sample. This approach allowed for evaluating deviations across the entire categorical distribution rather than testing individual categories separately.

XSL•FO
RenderX

Item-level statistics for the GR-eHEALS were calculated, including medians, means, SDs, and the distribution of response categories. To test for group differences, *t* tests and ANOVA were conducted. Participants were categorized based on predefined cut-off values or the median, shown in Table 1 and Multimedia Appendix 1. For variables with established clinical thresholds, participants were divided into 2 groups according to predefined cut-off values: cognitive impairment was defined as a MoCA score <26 [43], frailty as a CFS score ≥5 [45,46], and clinically relevant depressive symptoms as a DIA-S score ≥3 [47]. For continuous variables without established cut-offs, such as age, years of education, and Barthel Index, participants were split using the median value. Sex was categorized as male and female, and living situation was grouped into living alone and living with others. Subsequently, linear regression analyses were performed to examine associations between eHealth literacy and various participant characteristics. Prior to the regression analyses, multicollinearity was assessed using variance inflation factors (VIF). In addition, residual plots were visually inspected to evaluate linearity, homoscedasticity, and normality of residuals. No major violations of the regression assumptions were identified. Frailty was entered into the regression models as a dichotomous variable based on the CFS (nonfrail: CFS 1 - 4 vs frail: CFS 5 - 9). These analyses aimed to explore how sociodemographic and clinical factors relate to eHealth literacy levels in the study population. For the remaining categorical variables, the same cut-off values or medians described above were applied. Continuous variables were included in the regressions in their original scale. As the analyses were exploratory in nature, no correction for multiple testing was applied. Therefore, all statistical tests should be interpreted as exploratory and hypothesis-generating. The overall model fit was assessed using R^2 and adjusted R^2 values.

Table . Categorization to test for group differences, *t* tests and ANOVA.

Variables	Categories
Variables with established clinical thresholds	
Montreal Cognitive Assessment.	• 1 - 25 (indication of cognitive impairment) • 26 - 30 (no indication of cognitive impairment)
Clinical Frailty Scale	• 1 - 4 (nonfrail to mildly frail) • 5 - 9 (moderately to severely frail)
Depression in Old Age Scale	• 0 - 2 (no or minimal depressive symptoms) • 3 - 10 (elevated depressive symptoms)
Continuous variables divided according to median value	
Age	• ≤77 years • >77 years
Years of education	• ≤11 years • >11 years
Barthel Index	• ≤75 • >75
Categorical variables	
Sex	• Male • Female
Living situation	• Living alone • Living with others

Results

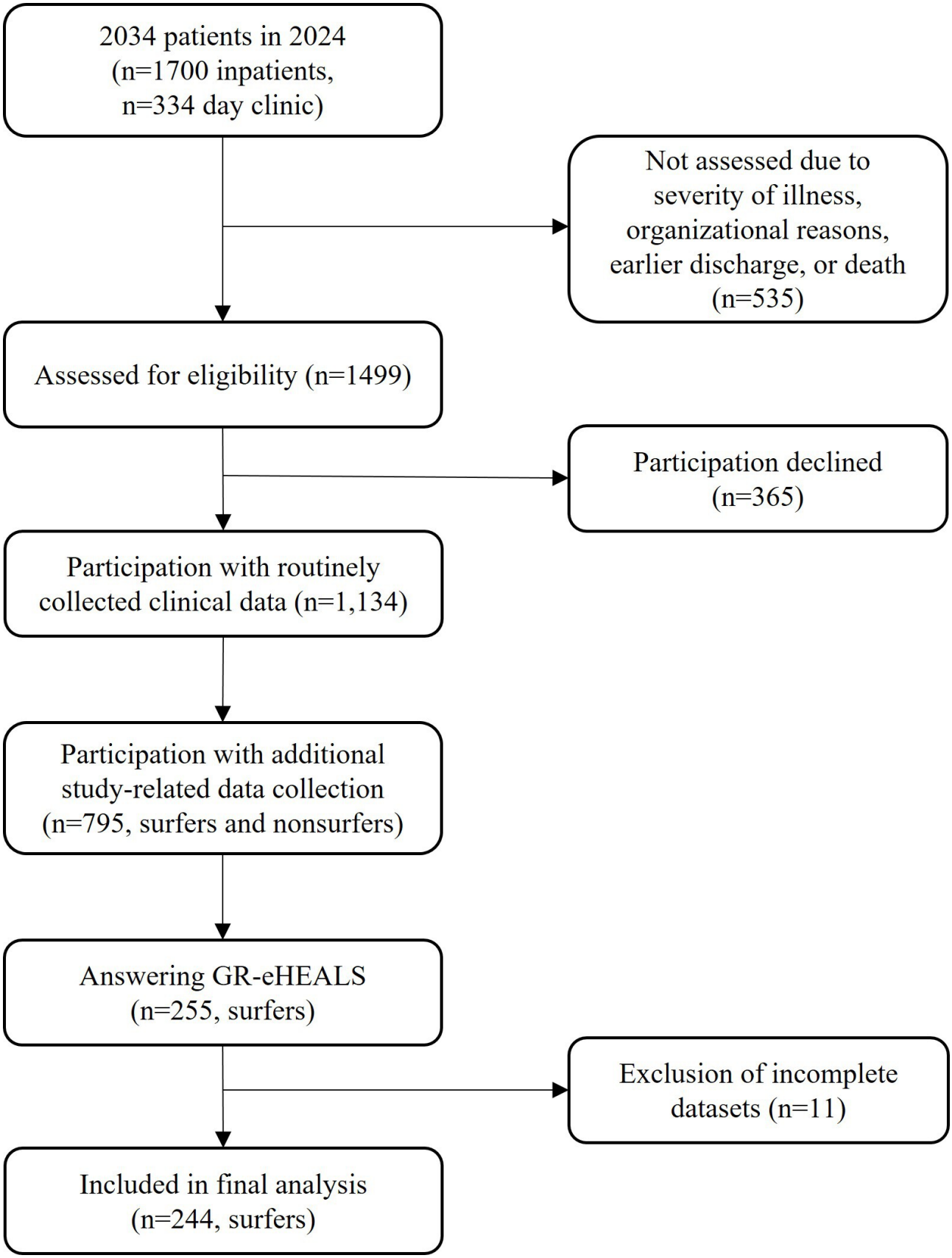
Characteristics of Study Participants

Throughout the study period, all patients fulfilling the inclusion criteria were consecutively invited to participate. The following procedure was used to identify patients eligible for inclusion: 1499 individuals were approached for the study. A total of 1134 of them gave written informed consent for the use of routinely collected clinical data for research purposes. Among these, 795 patients additionally agreed to take part in the supplementary study-related data collection. Of these 795 patients, 255 reported using the internet for health-related information and therefore answered the GR-eHEALS. This group of patients forms the study population of the reported substudy described below.

After excluding cases with incomplete responses, a total of 244 fully completed GR-eHEALS questionnaires were included in the final analyses, as illustrated in Figure 1. The characteristics of the study population for this substudy are summarized in Table 2. In total, participants' age ranged from 60 to 97 years (mean age 77.2, SD 6.9 years), with females accounting for 121 out of 244 (49.6%) of the sample. To show the relation between the subgroup of 244 patients and the entire cohort, we provide the characteristics in Table 2. In the following, we compare the subgroup to the group of nonsurfers to point out the differences.

XSL•FO
RenderX

Figure 1. Consolidated Standards of Reporting Trials diagram. GR-eHEALS: German version of the eHealth Literacy Scale.



XSL•FO
RenderX

Table . Characteristics of the entire cohort, study participants, and group comparisons.

Characteristics	Entire cohort (n=1134)	GR-eHEALS-sample (N=244)	*P* value[a]
Age, mean (SD) years	79.2 (14.8)	77.2 (6.9)	.54
Sex, n (%)			
Male	398 (35.1)	123 (50.4)	<.001
Female	736 (64.9)	121 (49.6)	.32
Years of education (self-reported), mean (SD)	10.8 (2.3)	11.9 (2.4)	.047
MoCA[b], mean (SD)	17.2 (6.2)	19.7 (6.8)	.002
Barthel Index[c], mean (SD)	69.7 (18.4)	73.3 (18.7)	<.001
Clinical Frailty Scale, mean (SD)	5.6 (1.0)	5.4 (1.1)	.13
DIA-S[d], mean (SD)	3.2 (2.6)	2.3 (2.5)	.47
Care degree on admission, n (%)			<.001
None	322 (28.4)	88 (36.1)	
1	122 (10.8)	26 (10.7)	
2	441 (38.9)	85 (34.8)	
3	224 (19.8)	40 (16.4)	
4	19 (1.7)	3 (1.2)	
5	2 (0.2)	0	
Living situation prior to admission[e], n (%)			.006
Alone	452 (56.9)	122 (50.0)	
With others (in private household)	318 (40.1)	119 (48.8)	
Nursing home	24 (3.0)	3 (1.2)	
Setting of recruitment, n (%)			
Inpatient care	958 (84.5)	198 (81.1)	.04
Geriatric day clinic	176 (15.5)	46 (18.9)	<.001

[a]Comparison of the study participants (n=244) and the population in the overall study (nonsurfers, n=890).

[b]MoCA: Montreal Cognitive Assessment.

[c]Barthel Index 0-100, with 100 indicating complete independence and 0 indicating no independence.

[d]DIA-S: Depression in Old Age Scale.

[e]Missing values in up to 64 patients were only recorded in a subsample of 794 participants due to methodological reasons

Comparison of the Substudy Participants and the Population in the Overall Study

To interpret the results in relation to the overall DiGA4Aged study and the total geriatric population, we compared the study participants reported here with the overall screening cohort of the substudy (n=1134). The overall screening cohort included both the study participants and patients who were excluded from this substudy because they did not use the internet to search for health-related information (referred to as “nonsurfers”). Although these nonsurfers did not complete the GR-eHEALS, they participated in the screening study by providing consent for the use of their routinely collected clinical data, allowing for the characterization of the study sample within the overall screening cohort. In the following and for descriptive purposes, we refer to the participants included in the substudy as “study participants” and to the nonparticipants as “nonsurfers” for simplicity. Nonsurfers are the 1134 people who participated in the routine data collection, minus the 244 people who completed the eHEALS questionnaire. This means that 890 nonsurfers are compared with 244 study participants.

When comparing the study participants with the overall screening cohort, the study participants were younger and more often male than the nonsurfers. The study participants showed better cognitive and functional status compared to the nonsurfers. Specifically, the study participants had significantly higher MoCA scores (19.7 vs 17.2; *P*=.002) and a higher Barthel Index (73.3 vs 69.7; *P*<.001) than the nonsurfers, suggesting greater independence in activities of daily living. A notable difference was observed in the distribution of care degrees on admission, with the study participants having fewer participants with a care degree of 3 or higher (*P*<.001) than the nonsurfers.

XSL•FO
RenderX

Finally, the sample of study participants also showed significant differences in living arrangements prior to admission, with more participants living with others or in a private household, and fewer living alone compared to the nonsurfers (*P*=.006). Details are provided in Table 2.

eHEALS Items and Scales

When examining eHealth literacy, the responses from participants on the GR-eHEALS offer a nuanced picture of their abilities to find, evaluate, and apply online health information. Looking at the individual items, participants report a broad range of self-perceived eHealth literacy, with responses spanning the full range of the 5-point Likert scale from "strongly disagree" (1) to "strongly agree" (5), shown in Table 3. For example, item 2, "I know how to use the internet to find answers to my health questions," received a high level of agreement, with a mean score of 3.79 (SD 0.97). However, some items, such as item 8, "I feel confident using internet information to make health decisions," rated lower and were more varied, with a broader spread of responses, especially leaning towards neutrality or disagreement, as indicated by the mean score of 2.85 (SD 1.12). The mean scores for individual items ranged up to 3.79 (item 2: "I know how to use the internet to find answers to my health questions"), with SDs varying from 0.97 to 1.17, indicating some variability in responses. The overall distribution of answers showed that most participants tended to agree or strongly agree with the statements about their eHealth literacy, with items 1‐4 (subscale Information Seeking) generally receiving higher agreement rates than items 5‐8 (subscale Information Appraisal), which assessed critical appraisal of health information.

XSL•FO
RenderX

Table . Overview of item and scale statistics (N=244)[a].

Original item in German (English translation)	Median(IQR)	Mean (SD)	Response distribution, n (%)					Sum (mean [SD])/max
			1	2	3	4	5	
Information seeking								
Ich weiß, wie ich Internetseiten mit hilfreichen Gesundheitsinformationen finden kann. (I know how to find websites with helpful health information.)	4 (4-4)	3.75 (0.98)	14 (5.7)	19 (7.8)	13 (5.3)	165 (67.6)	33 (13.5)	
Ich weiß, wie ich das Internet nutzen kann, um Antworten auf meine Gesundheitsfragen zu erhalten. (I know how to use the internet to find answers to my health questions.)	4 (4-4)	3.79 (0.97)	11 (4.5)	23 (9.4)	11 (4.5)	160 (65.6)	39 (16.0)	
Ich weiß, welche Seiten mit Gesundheitsinformationen im Internet verfügbar sind. (I know which websites with health information are available on the Internet.)	4 (2-4)	3.27 (1.17)	25 (10.2)	48 (19.7)	26 (10.7)	126 (51.6)	19 (7.8)	
Ich weiß, wo ich im Internet hilfreiche Gesundheits-informationen finden kann. (I know where to find helpful health information on the Internet.)	4 (4-4)	3.73 (1.01)	10 (4.1)	29 (11.9)	19 (7.8)	145 (59.4)	41 (16.8)	
Scale: information seeking	4 (3.25-4)	3.64 (0.84)	15 (6.1)	30 (12.3)	17 (7.0)	149 (61.1)	33 (13.5)	14.5 (3.4)/20
Information appraisal								

XSL•FO
RenderX

Original item in German (English translation)	Median(IQR)	Mean (SD)	Response distribution, n (%)					Sum (mean [SD])/max
			1	2	3	4	5	
Ich weiß Gesundheitsinformationen aus dem Internet so zu nutzen, dass sie mir weiterhelfen. (I know how to use health information from the internet in a way that helps me.)	4 (3-4)	3.46 (1.07)	11 (4.5)	45 (18.4)	39 (16.0)	119 (48.8)	30 (12.3)	
Ich bin in der Lage, Internetseiten mit Gesundheitsinformationen kritisch zu bewerten. (I am able to critically evaluate websites with health information.)	4 (3-4)	3.60 (1.01)	12 (4.9)	29 (11.9)	34 (13.9)	139 (57.0)	30 (12.3)	
Ich kann zwischen vertrauenswürdigen und fragwürdigen Internetseiten mit Gesundheitsinformationen unterscheiden. (I can distinguish between trustworthy and questionable websites with health information.)	4 (2-4)	3.39 (1.09)	14 (5.7)	49 (20.1)	32 (13.1)	125 (51.2)	24 (9.8)	
Ich fühle mich sicher darin, Informationen aus dem Internet zu nutzen, um Entscheidungen in Bezug auf meine Gesundheit zu treffen. (I feel confident using information from the internet to make decisions about my health.)	3 (2-4)	2.85 (1.12)	27 (11.1)	82 (33.6)	46 (18.9)	79 (32.4)	10 (4.1)	

XSL•FO
RenderX

Original item in German (English translation)	Median(IQR)	Mean (SD)	Response distribution, n (%)					Sum (mean [SD])/max
			1	2	3	4	5	
Scale: information appraisal	3.5 (3-4)	3.32 (0.80)	16 (6.6)	51 (20.9)	38 (15.6)	115 (47.1)	24 (9.8)	13.3 (3.2)/20
Scale: GR-eHealth Literacy (item 1 - 8)	3.6 (3.25-4)	3.48 (0.76)	16 (6.6)	40 (16.4)	28 (11.4)	132 (54.1)	28 (11.5)	27.8 (6.1)/40

[a]Response distribution: 1=strongly disagree, 2=disagree, 3=neutral, 4=agree, 5=strongly agree.

To calculate the overall eHealth literacy score, the summed scores of Items 1 - 8 (Information Seeking and Information Appraisal) were used, with a possible range from 8 to 40 points and 4 to 20 points for the subscores, respectively, divided by the number of items. The mean score for the eHealth literacy scale was 3.48 (SD 0.76); for the Information Seeking subscale (Items 1 - 4), it was 3.64 (SD 0.84), and for the Information Appraisal subscale (Items 5 - 8), it was slightly lower (mean 3.32, SD 0.80). In Table 3, absolute frequencies of sum-scores were rounded to whole persons. Due to this rounding, minimal discrepancies (+1 person, 2%) may occur. To maintain a consistent sample size of n=244 (100%), the response option with the largest proportion was reduced by one person (2%; following standard rounding rules).

To examine differences in eHealth literacy across the subscales, independent *t* tests were conducted for each scale, using predefined cut-off or median values. Significant differences were only found for age across all scales, with t_{242}=2.365; *P*=.02 for the Information Seeking scale; t_{242}=2.544; *P*=.012 for the Information Appraisal scale; and t_{242}=2.652; *P*=.009 for the whole eHealth Literacy scale. In addition, frailty was significantly associated with the Information Appraisal scale, where participants with higher frailty scores (frailty 5 - 9) demonstrated lower scores on the Information Appraisal subscale compared to those with lower frailty scores (frailty 1 - 4), with t_{242}=2.067; *P*=.040. For all other comparisons, including MoCA, DIA-S, years of education, Barthel Index, gender, care degree, and living situation, no significant differences were found across the scales.

Characteristics Influence on eHealth Literacy

To investigate the influence of various characteristics on eHealth literacy, linear regression analyses were conducted for all scales (Information Seeking, Information Appraisal, and the overall eHealth Literacy), see Table 4. Due to missing data, regression analyses were based on complete cases, resulting in 207 participants included in the regression models. For the Information Seeking scale, the regression model explained 3.2% of the variance (R^2=0.032, adjusted R^2=–0.012). No individual predictor showed a significant association with information seeking. For the Information Appraisal scale, the model explained 6.6% of the variance (R^2=0.066, adjusted R^2=0.024). Frailty emerged as significantly negatively associated with Information Appraisal (B=–1.372; *P*=.02), indicating that individuals with higher frailty scores were less likely to critically appraise online health information. However, no other characteristics showed significant association with the participants' ability to assess health information. The regression model for the overall eHealth literacy score (Items 1 - 8) explained 5.3% of the variance (R^2=0.053, adjusted R^2=0.010). Frailty showed a marginally significant negative association (B=–2.217; *P*=.05).

XSL•FO
RenderX

Table . Linear regression analysis of factors associated with eHealth literacy[a].

Characteristics	Sum score for eHealth literacy (GR-eHEALS)[b]			
	B[c]	SE	Standardized β (95% CI)	*P* value
Scale information seeking (item 1 - 4)[d]				
Age	−0.040	0.036	−0.082 (−0.111 to 0.031)	.27
Sex (male)	−0.084	0.512	−0.012 (−1.094 to 0.926)	.87
Years of education	0.067	0.106	0.047 (−0.142 to 0.276)	.53
MoCA[e]	0.069	0.668	0.007 (−1.249 to 1.388)	.92
Barthel Index	0.009	0.014	0.050 (−0.019 to 0.038)	.51
Frailty Score	−0.845	0.635	−0.097 (−2.097 to 0.407)	.19
DIA-S[f]	−0.059	0.518	−0.008 (−1.081 to 0.963)	.91
Care degree on admission	−0.143	0.224	−0.049 (−0.585 to 0.299)	.53
Living situation (alone)	−0.012	0.506	−0.002 (−1.011 to 0.986)	.98
Scale information appraisal (item 5 - 8)[g]				
Age	−0.049	0.034	−0.105 (−0.116 to 0.018)	.15
Sex (male)	−0.630	0.484	−0.096 (−1.583 to 0.324)	.19
Years of education	0.045	0.100	0.033 (−0.153 to 0.242)	.65
MoCA	0.046	0.631	0.005 (−1.199 to 1.291)	.94
Barthel Index	0.018	0.013	0.099 (−0.009 to 0.044)	.18
Frailty Score	−1.372	0.600	−0.163 (−2.554 to −0.189)	.02
DIA-S	−0.068	0.490	−0.010 (−1.034 to 0.897)	.89
Care degree on admission	−0.038	0.212	−0.013 (−0.455 to 0.380)	.86
Living situation (alone)	−0.296	0.478	−0.045 (−1.239 to 0.647)	.54
Scale eHealth Literacy (item 1 - 8)[h]				
Age	−0.088	0.065	−0.101 (−0.216 to 0.039)	.17
Sex (male)	−0.713	0.920	−0.058 (−2.529 to 1.102)	.44
Years of education	0.111	0.190	0.043 (-0.264 to 0.487)	.56
MoCA	0.115	1.201	0.007 (−2.254 to 2.484)	.92
Barthel Index	0.027	0.026	0.080 (−0.023 to 0.078)	.29
Frailty score	−2.217	1.141	−0.140 (−4.467 to 0.034)	.05
DIA-S	−0.127	0.932	−0.010 (−1.965 to 1.710)	.89
Care degree on admission	−0.181	0.403	−0.034 (−0.975 to 0.614)	.66
Living situation (alone)	−0.308	0.910	−0.025 (−2.102 to 1.486)	.74

[a]From 1.077 to 1.191, indicating no relevant multicollinearity.

[b]GR-eHEALS: German version of the eHealth Literacy Scale.

[c]B: unstandardized regression coefficient.

[d]Scale information seeking (items 1-4): R^2=0.032, adjusted R^2=–0.012, $F_{9,197}$ =0.729; *P*=.68.

[e]MoCA: Montreal Cognitive Assessment.

[f]DIA-S: Depression in Old Age Scale.

[g]Scale information appraisal (items 5-8): R^2=0.066, adjusted RR^2 = 0.024, $F_{9,197}$=1.551; *P*=.13.

[h]Scale eHealth Literacy (items 1-8): R^2=0.053, adjusted R^2=0.010, $F_{9,197}$=1.226; *P*=.28.

XSL•FO
RenderX

Discussion

Main Findings

The study aimed to explore eHealth literacy among geriatric patients, focusing on how sociodemographic and clinical factors may influence it. Patients included were 60 - 97 years old and reported using the internet to search for health-related information. The findings suggest that older adults, particularly those with higher levels of frailty, tend to have lower eHealth literacy scores, particularly in the area of information appraisal. It partially supports our hypothesis, as frailty was significantly associated with lower information appraisal scores in the multivariate analysis. However, no significant gender differences in eHealth literacy were observed, which suggests that digital socialization or gender stereotypes may not be as impactful in this cohort. Furthermore, despite expectations, age was not significantly associated with eHealth literacy in this study, which may reflect the stability of self-assessments despite fluctuating objective digital skills. Although age was significantly associated with eHealth literacy in the bivariate analyses, this association was no longer significant in the multivariate regression models. This suggests that chronological age alone cannot independently explain the differences in eHealth literacy. In our geriatric sample, frailty was significantly associated with the ability to critically evaluate online health information, suggesting that functional health status may be more relevant than chronological age. Given the close association between age and frailty, the bivariate association with age may partly reflect age-related differences in functional status rather than an independent effect of age itself. Although frailty was significantly associated with the information appraisal dimension of eHealth literacy, the regression models explained only a limited proportion of variance. This suggests that additional factors beyond the sociodemographic and clinical characteristics examined may contribute to differences in eHealth literacy among geriatric patients. Potential factors that were not assessed in this study include previous experience with digital technologies, digital skills, general health literacy, social support and attitudes, and motivation towards digital technologies. However, only 255 of 795 participants (32%) in our study used the internet to gather health-related information, demonstrating the age gap in this area. These findings indicate that a considerable proportion of hospitalized geriatric patients do not use the internet as a source of health information and may therefore not currently benefit from digital health services. Several studies have examined eHealth literacy in older populations, with varying methodologies and findings. For example, Jørgensen et al (2023) [49] reported that many hospitalized geriatric patients had limited computer habits and digital literacy, with only half regularly using computers and many unfamiliar with basic digital functions. They also identified that higher digital literacy was associated with less computer-related stress and more frequent use of the internet, particularly for health-related tasks. This aligns with our findings, which showed that participants with lower levels of frailty demonstrated higher eHealth literacy scores, especially in the area of information appraisal. However, our study extends Jørgensen et al's [49] work by emphasizing the specific challenges faced by individuals who are frail in critically appraising online health information, which was a key finding in our analysis. While their study primarily focused on general digital skills, our results highlight that eHealth literacy may not only be related to basic computer skills but may also be associated with physical and cognitive health factors. In particular, frailty was linked to lower confidence in critically appraising online health information.

Arcury et al [50] showed that higher levels of eHealth literacy, measured by eHEALS, were associated with better health knowledge and more frequent use of health information sources. It is plausible that participants who used digital health tools more frequently may have had higher eHealth literacy scores, although we cannot directly confirm this in our data.

Self-assessment may have introduced a self-report bias. A major gap in the current research is the predominance of subjective self-assessment tools, with relatively few studies using objective, performance-based measures of eHealth literacy. This limits the ability to accurately predict actual usage behavior and identify those at risk for poor engagement or adverse outcomes in digital health care and telemedicine [32-34]. Therefore, understanding the role of subjective self-assessments in the broader picture of eHealth literacy is crucial. A review by Azizan [29] showed that most of the available tools (57%) currently involve self-assessment. Such tools often require significantly less time and organizational effort, as well as evaluation time, than a performance test, which would enable more objective results. Therefore, it represents a tool that can be realistically used in practice. Since no validated German instrument for objective evaluation exists, this approach was chosen. It offers first insights into the needs of the geriatric population, since the self-assessment can show the need for support, either knowledge improvement or empowerment to enable safe digital health engagement.

With regard to self-assessment, an effect that has been described extensively in the literature may come into play: the Dunning-Kruger effect, which is characterized by 2 mechanisms: First, a lack of knowledge means that people are unaware of their own gaps in knowledge, causing them to overestimate their abilities. Second, greater competence increases awareness of how much one does not yet know, resulting in a more cautious self-assessment [51]. This phenomenon is particularly relevant in the context of eHealth literacy. Older adults may feel confident in their ability to use digital tools, even when their skills do not align with their perceptions. For instance, a person might believe they are proficient at navigating health apps or searching for medical information online, but in reality, they may struggle with critical aspects like evaluating the credibility of sources or understanding complex health data. This misalignment between perceived and actual competence often arises from the first mechanism of the Dunning-Kruger effect, lack of awareness about one's own limitations. The phenomenon suggests that subjective self-assessments might not always accurately reflect actual eHealth literacy, particularly in individuals who are frail, in whom functional limitations may affect confidence and perceived abilities to engage with digital health information. Interestingly, although no significant gender differences were found, this may be due to the complexity of

XSL•FO
RenderX

digital socialization and the fact that both men and women in our sample may have had similar levels of exposure to digital health tools. The lack of significant differences in age and self-reported eHealth literacy could be attributed to the stability of self-perception despite fluctuating digital abilities. This suggests that subjective self-assessments may not always accurately reflect actual eHealth literacy, particularly in individuals who are frail, whose functional status and cognitive abilities are compromised. For a more objective measurement, individuals could first be asked to assess their own digital literacy, which could then be verified by testing their actual abilities. By comparing self-assessments with real-world performance, a clearer understanding of their true eHealth literacy can be achieved.

Our results can also be supported in part by the use of a framework. Richardson et al (2022) [52] have developed the "Framework for digital health equity," which offers a systematic approach to making digital interventions more equitable by taking social, structural, and technological aspects into account. The framework takes into account cumulative and interactive effects across multiple domains, as well as risk and resilience factors. Among other factors, the framework states that socioeconomic status, functional status, and access to support are crucial for the successful use of digital offerings for health-related purposes. This cannot be seen in our data since there were no significant associations found with living situation. From the authors' perspective, the idea that having a reliable support system—whether family members or caregivers—can facilitate better engagement with digital health tools is, however, very convincing. In this context, Hunsaker et al [53] investigated the types of support older adults receive when using technology. They identified informal support (eg, spouses, adult children, grandchildren, and siblings) and formal support (eg, computer classes, technical support staff in stores and via remote access). However, the study also noted that a major problem in this regard is that support is often not provided immediately. This may explain the results, as there are a variety of support options available that are not exclusively tailored to the individual's living situation. This is certainly an exciting area to be further explored.

The relevance of frailty is also reflected in this framework. However, we found that frailty was significantly associated with the information appraisal dimension of eHealth literacy, an aspect that has received less attention in previous studies. One possible explanation is that the interaction between cognitive and physical limitations may reduce confidence and perceived ability to effectively navigate and critically evaluate online health information.

Outlook and Requirements

Our findings support the applicability of a validated instrument for assessing eHealth literacy in geriatric populations, reinforcing its utility for measuring digital literacy in older adults despite its limitations as a self-assessment instrument. The variability in scores across studies indicates that subjective assessments may not fully capture the complexities of digital engagement, especially when factors like frailty and cognitive function are involved. Future research should aim to combine subjective self-assessments with more objective, performance-based measures of eHealth literacy such as task-based evaluations, simulated digital health scenarios, or observed information appraisal tasks. Longitudinal studies that track changes in digital literacy over time and their impact on health outcomes would provide valuable insights into how eHealth literacy evolves with age and frailty. Future studies should investigate factors such as digital skills and general health literacy in addition to sociodemographic and clinical characteristics to develop a more comprehensive understanding of the determinants of eHealth literacy among geriatric patients. In view of the rise of AI and the increase in misinformation on the internet, as well as the difficulty of distinguishing it from true information, the skills required are also changing, merging into the meta-competence of eHealth literacy. This is also evident in the ongoing development of the revised eHEALS [14]. This digital progress will require continuous adjustments to education and operationalization.

Further exploration is also needed to evaluate the effectiveness of interventions designed to improve eHealth literacy in older adults, as they require human support, especially those with better eHealth literacy eg, as part of nursing. It would be beneficial to assess how these interventions are implemented and whether they address the specific barriers faced by older adults in digital health settings. Previously, searching the internet for health information required typing, reading, and appraisal of such information. Current developments in the area of AI are already changing these requirements fundamentally and may contribute to digital health equity in the future. However, it will be interesting to observe how the already widespread use of AI in the younger population will change the extent and depth of health-associated applications in the older population.

Strengths and Limitations

This study has several strengths, including the use of a validated instrument for assessing eHealth literacy (GR-eHEALS). Although the validity of self-measurement is limited in this case, the gold standard in the German validated version was used for self-reported digital health literacy. In addition, the study's large cohort of 244 participants strengthens the robustness of the findings within the context of geriatric patients who engage with digital health tools. The sample also includes a broad age range (60 - 97 years), which allows for the investigation of potential variability in eHealth literacy across different age groups. Despite the self-reported nature of the data, the use of the GR-eHEALS allowed for a standardized approach to assess key aspects of eHealth literacy, including information seeking and critical appraisal.

However, there are weaknesses that must be considered when interpreting the results. The reliance on self-assessment as the sole measure of eHealth literacy is a significant limitation. Self-reported data can lead to overestimations of actual eHealth literacy, as participants may perceive themselves as more capable than they are in practice [41,42]. The absence of objective measures to validate these self-assessments limits the accuracy of the findings. In addition, these analyses were restricted to participants who reported using the internet to obtain health-related information, as completion of the

XSL•FO
RenderX

GR-eHEALS requires prior engagement with digital health resources. Consequently, the study population represents a selected subgroup of digitally engaged geriatric patients rather than the entire geriatric inpatient population. This selection may have resulted in higher average eHealth literacy scores than would be expected in unselected hospitalized older adults and limits the generalizability of the findings to patients with little or no experience using digital health technologies. Furthermore, while the study included a relatively large sample, the population is not fully representative of the broader geriatric hospitalized patient group. The participants in this study were notably fitter than the average geriatric patient, as evidenced by their higher scores on cognitive and functional assessments. Therefore, the results must be interpreted with caution, as they may not reflect the experiences of the wider geriatric population, especially those with greater frailty or cognitive impairment. No adjustment for multiple comparisons was performed. Consequently, statistically significant findings should be interpreted with caution and require confirmation in future hypothesis-driven studies.

Conclusion

This study provides important insights into eHealth literacy in geriatric patients and its associated factors and underscores the importance of interventions that consider the unique challenges faced by older adults, particularly those with frailty. Future research should focus on integrating subjective and objective measures of eHealth literacy, exploring the impact of social support, and evaluating the effectiveness of current interventions aimed at enhancing digital health engagement among older populations.

Generative AI Use

No AI was used in the creation of this manuscript.

Funding

The DiGA4Aged study is funded by the InnovationsFoRUM, a funding programme of the Medical Faculty at the Ruhr University Bochum (funding number: IF-025-22). The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript.

Data Availability

The data presented in this study are available upon request from the corresponding author. The data are not publicly available.

Authors' Contributions

Conceptualization: TSB, ISH, JD-H, HCV, RW
Methodology: CF, RW
Data acquisition: DD, CF, MH, JK
Data curation: CF
Data analysis and interpretation: TSB, CF, UT, RW
Resources: RW
Visualization: CF
Supervision: TSB, RW
Project administration: CF, RW
Funding acquisition: TSB, JD-H, HCV, RW
Writing – original draft: CF
Writing – review and editing: TSB, SEB, DD, CF, ISH, MH, JD-H, JK, KK, UT, BU, HCV, RW

Conflicts of Interest

None declared.

Checklist 1
Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist.
[PDF File, 104 KB - humanfactors_v13i1e92045_app1.pdf]

References

1. Recommendations on digital interventions for health system strengthening. World Health Organization. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2026-09-13]
2. Theobald P, Herold F, Gronwald T, Müller NG. Use of digital health technologies in Germany [Article in German]. Präv Gesundheitsf 2026 Aug;21(3):414-424. [doi: 10.1007/s11553-025-01228-1]
3. Shapoo N, Shapoo N, Rehman A, Boma N. Smart healthcare: the role of digital health in modern medicine. Health Care Sci 2025 Jun;4(3):179-187. [doi: 10.1002/hcs2.70019] [Medline: 40568626]

XSL•FO
RenderX

4. Knaup P, Benning NH, Seitz MW, Eisenmann U. eHealth and clinical documentation systems. Stud Health Technol Inform 2020 Sep 25;274:174-188. [doi: 10.3233/SHTI200676] [Medline: 32990673]
5. Pagliari C, Sloan D, Gregor P, et al. What is eHealth (4): a scoping exercise to map the field. J Med Internet Res 2005 Mar 31;7(1):e9. [doi: 10.2196/jmir.7.1.e9] [Medline: 15829481]
6. Digital health applications (DiGA) [Webpage in German]. Federal Ministry of Health. URL: https://www.bundesgesundheitsministerium.de/themen/krankenversicherung/online-ratgeber-krankenversicherung/arznei-heil-und-hilfsmittel/digitale-gesundheitsanwendungen.html [accessed 2025-11-28]
7. DiGA und DiPA [Wepage in German]. Federal Institute for Drugs and Medical Devices. URL: https://www.bfarm.de/DE/Medizinprodukte/Aufgaben/DiGA-und-DiPA/_node.html [accessed 2026-10-16]
8. DiGA guidelines (as of December 10, 2025, version 3.6) [Webpage in German]. Federal Institute for Drugs and Medical Devices. URL: https://www.bfarm.de/SharedDocs/Downloads/DE/Medizinprodukte/diga_leitfaden.html [accessed 2026-09-13]
9. Rebitschek FG, Gigerenzer G. Assessing the quality of digital health services: how can informed decisions be promoted? [Article in German]. Bundesgesundheitsbl 2020 Jun;63(6):665-673. [doi: 10.1007/s00103-020-03146-3]
10. Ayers JW, Zhu Z, Poliak A, et al. Evaluating artificial intelligence responses to public health questions. JAMA Netw Open 2023 Jun 1;6(6):e2317517. [doi: 10.1001/jamanetworkopen.2023.17517] [Medline: 37285160]
11. Andrikyan W, Sametinger SM, Kosfeld F, et al. Artificial intelligence-powered chatbots in search engines: a cross-sectional study on the quality and risks of drug information for patients. BMJ Qual Saf 2025 Jan 28;34(2):100-109. [doi: 10.1136/bmjqs-2024-017476] [Medline: 39353736]
12. Hung C, Katapally TR. Assessing the role of digital literacy in accessing and utilising virtual healthcare services: a systematic review protocol. J Eval Clin Pract 2025 Feb;31(1):e14245. [doi: 10.1111/jep.14245] [Medline: 39564975]
13. Norman CD, Skinner HA. eHEALS: the eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27. [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
14. Milanti A, Norman C, Chan DNS, So WKW, Skinner H. eHealth Literacy 3.0: updating the Norman and Skinner 2006 model. J Med Internet Res 2025 Mar 11;27:e70112. [doi: 10.2196/70112] [Medline: 40068166]
15. Sinha M, Fischer F, Giehl C, et al. Relevance of eHealth Literacy (eHL) for health services research and practice - part I [Article in German]. Gesundheitswesen 2026 Feb;88(2):119-127. [doi: 10.1055/a-2663-4406]
16. Fitzpatrick PJ. Improving health literacy using the power of digital communications to achieve better health outcomes for patients and practitioners. Front Digit Health 2023;5:1264780. [doi: 10.3389/fdgth.2023.1264780] [Medline: 38046643]
17. Busse TS, Nitsche J, Kernebeck S, et al. Approaches to improvement of digital health literacy (eHL) in the context of person-centered care. Int J Environ Res Public Health 2022 Jul 7;19(14):8309. [doi: 10.3390/ijerph19148309] [Medline: 35886158]
18. Hua Z, Yuqing S, Qianwen L, Hong C. Factors influencing eHealth literacy worldwide: systematic review and meta-analysis. J Med Internet Res 2025 Mar 10;27:e50313. [doi: 10.2196/50313] [Medline: 40063939]
19. Milanti A, Chan DNS, Parut AA, So WKW. Determinants and outcomes of eHealth literacy in healthy adults: a systematic review. PLoS One 2023;18(10):e0291229. [doi: 10.1371/journal.pone.0291229] [Medline: 37792773]
20. Estrela M, Semedo G, Roque F, Ferreira PL, Herdeiro MT. Sociodemographic determinants of digital health literacy: a systematic review and meta-analysis. Int J Med Inform 2023 Sep;177:105124. [doi: 10.1016/j.ijmedinf.2023.105124] [Medline: 37329766]
21. Sainimnuan S, Preedachitkul R, Petchthai P, Paokantarakorn Y, Siriussawakul A, Srinonprasert V. Low prevalence of adequate eHealth literacy and willingness to use telemedicine among older adults: cross-sectional study from a middle-income country. J Med Internet Res 2025 Jul 1;27:e65380. [doi: 10.2196/65380] [Medline: 40591893]
22. Jiang X, Wang L, Leng Y, et al. The level of electronic health literacy among older adults: a systematic review and meta-analysis. Arch Public Health 2024 Nov 7;82(1):204. [doi: 10.1186/s13690-024-01428-9] [Medline: 39511667]
23. Heponiemi T, Kainiemi E, Virtanen L, et al. Predicting internet use and digital competence among older adults using performance tests of visual, physical, and cognitive functioning: longitudinal population-based study. J Med Internet Res 2023 May 5;25:e42287. [doi: 10.2196/42287] [Medline: 37145836]
24. Ge H, Li J, Hu H, Feng T, Wu X. Digital exclusion in older adults: a scoping review. Int J Nurs Stud 2025 Aug;168:105082. [doi: 10.1016/j.ijnurstu.2025.105082] [Medline: 40279791]
25. Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. Lancet 2013 Mar 2;381(9868):752-762. [doi: 10.1016/S0140-6736(12)62167-9] [Medline: 23395245]
26. Hoogendijk EO, Afilalo J, Ensrud KE, Kowal P, Onder G, Fried LP. Frailty: implications for clinical practice and public health. Lancet 2019 Oct 12;394(10206):1365-1375. [doi: 10.1016/S0140-6736(19)31786-6] [Medline: 31609228]
27. Reissmann M, Oswald V, Zank S, Tesch-Römer C. Digital participation in old age [Chapter in German]. In: Kaspar R, Simonson J, Tesch-Römer C, Wagner M, Zank S, editors. Old Age in Germany Studies on Health and Society, Vol 8 Springer, Berlin, Heidelberg, 8th edition: Springer; 2023. [doi: 10.1007/978-3-662-66630-2_7]
28. Çöme O, Gökdemir Ö, Bilik Sezer B, Kasapoğlu SS, Kjær NK, Güldal D. The interplay between cognitive function and digital health literacy among older adults: implications for e-health equity and accessibility. Int J Med Inform 2025 Sep;201:105934. [doi: 10.1016/j.ijmedinf.2025.105934] [Medline: 40273597]

XSL•FO
RenderX

29. Azizan A. Barriers and motivators of geriatric engagement in digital therapeutics: a narrative review. Ageing Res Rev 2026 Jan;114:102963. [doi: 10.1016/j.arr.2025.102963] [Medline: 41318100]
30. Aslan A, Mold F, van Marwijk H, Armes J. What are the determinants of older people adopting communicative e-health services: a meta-ethnography. BMC Health Serv Res 2024 Jan 11;24(1):60. [doi: 10.1186/s12913-023-10372-3] [Medline: 38212713]
31. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021 Aug 17;21(1):1556. [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]
32. Sezgin D, Hegerath-Segler FM, Christie H, et al. Recommendations for the development and use of technology to support people living with dementia and caregivers: a Delphi study. Alzheimers Dement 2025 Oct;21(10):e70755. [doi: 10.1002/alz.70755] [Medline: 41017559]
33. Cuschieri S. The STROBE guidelines. Saudi J Anaesth 2019 Apr;13(Suppl 1):S31-S34. [doi: 10.4103/sja.SJA_543_18] [Medline: 30930717]
34. Chatsatrian M, Kunde K, Bosompem J, et al. Usability evaluation of digital health applications for older people with depressive disorders: prospective observational study in a mixed methods design. JMIR Hum Factors 2025 Nov 28;12:e66271. [doi: 10.2196/66271] [Medline: 41328453]
35. Mai A, Pape M, Busse TS, et al. Implementation of a digital nurse to improve the use of digital health applications (DiGA) for older people with depressive disorders (DiGA4Aged): a randomized proof of concept study. Trials 2025 Apr 10;26(1):130. [doi: 10.1186/s13063-025-08809-7] [Medline: 40211379]
36. Giehl C, Chatsatrian M, Vollmar HC, et al. Exploring accessibility, user experience and engagement of digital media among older patients with depression: a pilot and observational screening study protocol of the DiGA4Aged study. BMJ Open 2024 Nov 7;14(11):e086779. [doi: 10.1136/bmjopen-2024-086779] [Medline: 39515866]
37. Marsall M, Engelmann G, Skoda EM, Teufel M, Bäuerle A. Measuring electronic health literacy: development, validation, and test of measurement invariance of a revised German version of the eHealth literacy scale. J Med Internet Res 2022 Feb 2;24(2):e28252. [doi: 10.2196/28252] [Medline: 35107437]
38. Petrič G, Atanasova S, Kamin T. Ill literates or illiterates? Investigating the eHealth literacy of users of online health communities. J Med Internet Res 2017 Oct 4;19(10):e331. [doi: 10.2196/jmir.7372] [Medline: 28978496]
39. Xie L, Zhang S, Xin M, Zhu M, Lu W, Mo PKH. Electronic health literacy and health-related outcomes among older adults: a systematic review. Prev Med 2022 Apr;157:106997. [doi: 10.1016/j.ypmed.2022.106997] [Medline: 35189203]
40. Huang YQ, Liu L, Goodarzi Z, Watt JA. Diagnostic accuracy of eHealth literacy measurement tools in older adults: a systematic review. BMC Geriatr 2023 Mar 29;23(1):181. [doi: 10.1186/s12877-023-03899-x] [Medline: 36978033]
41. Crocker B, Feng O, Duncan LR. Performance-based measurement of eHealth literacy: systematic scoping review. J Med Internet Res 2023 Jun 2;25:e44602. [doi: 10.2196/44602] [Medline: 37266975]
42. Wang C, Chang L, Chen X, Kong J, Qi H. eHealth literacy assessment instruments: scoping review. J Med Internet Res 2025 Aug 20;27:e66965. [doi: 10.2196/66965] [Medline: 40835422]
43. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
44. Lübke N, Meinck M, von Renteln–Kruse W. The Barthel Index in geriatrics. a context analysis for the Hamburg Classification Manual [Article in German]. Z Gerontol Geriat 2004 Aug;37(4):316-326. [doi: 10.1007/s00391-004-0233-2]
45. Rockwood K, Song X, MacKnight C, et al. A global clinical measure of fitness and frailty in elderly people. CMAJ 2005 Aug 30;173(5):489-495. [doi: 10.1503/cmaj.050051] [Medline: 16129869]
46. Singler K, Gosch M, Antwerpen L. Clinical Frailty Scale (CFS) [Report in Dutch]. : Deutsche Gesellschaft für Geriatrie e.V. (DGG); 2020 URL: https://www.dggeriatrie.de/images/Bilder/PosterDownload/200331_DGG_Plakat_A4_Clinical_Frailty_Scale_CFS.pdf [accessed 2026-10-28]
47. Heidenblut S, Zank S. Development of a new screening instrument for geriatric depression. The depression in old age scale (DIA-S) [Article in German]. Z Gerontol Geriatr 2010 Jun;43(3):170-176. [doi: 10.1007/s00391-009-0067-z] [Medline: 19760357]
48. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
49. Jørgensen BB, Gregersen M, Pallesen SH, Damsgaard EMS. Computer habits and digital literacy in geriatric patients: a survey. Digit Health 2023;9:20552076231191004. [doi: 10.1177/20552076231191004] [Medline: 37588159]
50. Arcury TA, Sandberg JC, Melius KP, et al. Older adult internet use and eHealth literacy. J Appl Gerontol 2020 Feb;39(2):141-150. [doi: 10.1177/0733464818807468] [Medline: 30353776]
51. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. J Pers Soc Psychol 1999 Dec;77(6):1121-1134. [doi: 10.1037//0022-3514.77.6.1121] [Medline: 10626367]
52. Richardson S, Lawrence K, Schoenthaler AM, Mann D. A framework for digital health equity. NPJ Digit Med 2022 Aug 18;5(1):119. [doi: 10.1038/s41746-022-00663-0] [Medline: 35982146]
53. Hunsaker A, Nguyen MH, Fuchs J, Djukaric T, Hugentobler L, Hargittai E. "He explained it to me and I also did it myself": how older adults get support with their technology uses. Socius 2019 Jan;5. [doi: 10.1177/2378023119887866]

XSL•FO
RenderX

Abbreviations

BfArM: Federal Institute for Drugs and Medical Devices
CFS: Clinical Frailty Scale
DIA-S: Depression in Old Age Scale
DiGA: Digital health application
eHEALS-R: revised eHealth Literacy Scale
GR-eHEALS: German version of the eHealth Literacy Scale
MoCA: Montreal Cognitive Assessment
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology guideline
VIF: variance inflation factors

Edited by A Bamgboje-Ayodele; submitted 23.Jan.2026; peer-reviewed by P Haessner, Y Kume; revised version received 13.Sep.2026; accepted 15.Sep.2026; published 09.Nov.2026.

Please cite as:
Flemm C, Wirth R, El Bataioui S, Daubert D, Haussleiter I, Hermenau MT, Dieris-Hirche J, Kuhlmann JF, Kunde K, Trampisch U, Ueberberg B, Vollmar HC, Busse TS
Association of eHealth Literacy With Cognitive Function, Frailty, and Functional Status in Hospitalized Geriatric Patients: Cross-Sectional Observational Study
JMIR Hum Factors 2026;13:e92045
URL: https://humanfactors.jmir.org/2026/1/e92045
doi:10.2196/92045

© Chantal Flemm, Rainer Wirth, Sarah El Bataioui, Diana Daubert, Ida Haussleiter, Matthias Tristan Hermenau, Jan Dieris-Hirche, Judith Frederica Kuhlmann, Katharina Kunde, Ulrike Trampisch, Bianca Ueberberg, Horst Christian Vollmar, Theresa Sophie Busse. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 9.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Telemedicine Adoption Among Health Care Professionals in Israeli Geriatric Medical Centers: Moderated Mediation Cross-Sectional Questionnaire Study

Gizell Green[1,2], PhD; Alena Lochmannová[3], PhD; Riki Tesler[4], MPH, PhD; Tammy Porat-Packer[2], MHA, MD

[1]Nursing Department, The Max Stern Yezreel Valley College, D.N. Emek Yezreel, Emek Yezreel, Northern District, Israel

[2]Shoham Medical Center, Pardes Hana Karkur, Haifa District, Israel

[3]Department of Crisis Management, Faculty of Security Management, Police Academy of the Czech Republic in Prague, Prague, Czech Republic

[4]Department of Health Systems Management, Faculty of Health Sciences, Ariel University, Ariel, Judea and Samaria Area, Israel

Corresponding Author:
Gizell Green, PhD
Nursing Department, The Max Stern Yezreel Valley College, D.N. Emek Yezreel, Emek Yezreel, Northern District, Israel

Abstract

Background: Telemedicine adoption among health care professionals is hindered by resistance to change (RTC), comprising routine-seeking, emotional reactivity, short-term focus, and cognitive rigidity, which interacts with perceived ease of use (PEU) and anxiety within the Technology Acceptance Model framework.

Objective: This study aimed to test a moderated mediation model in which PEU mediates the association between RTC, assessed both overall and by dimension, and intention to use telemedicine, while telemedicine-related anxiety moderates the association between PEU and intention to use. Participants were recruited from 4 governmental geriatric medical centers in Israel.

Methods: A cross-sectional study was conducted among 377 health care professionals from 4 Israeli governmental geriatric medical centers. Telemedicine use in the participating centers was limited, optional, and consisted mainly of remote patient monitoring. Data collection was conducted between March and June 2024. Data were collected using validated self-administered questionnaires measuring telemedicine-related anxiety, RTC with 4 dimensions (routine seeking, emotional reactivity, short-term focus, and cognitive rigidity), PEU, and intention to use telemedicine. Statistical analyses were performed using SPSS (version 28) and PROCESS Model 14 for moderated mediation analysis.

Results: The moderated mediation analysis revealed that RTC was negatively associated with the intention to use telemedicine through the lower PEU, with this indirect pathway being attenuated but not eliminated by high telemedicine-related anxiety. The overall indirect effect remained negative and statistically significant across all anxiety levels.

Conclusions: Health care professionals' RTC reduced their intention to use telemedicine by lowering PEU, with telemedicine-related anxiety moderating but never eliminating this relationship. These findings suggest that telemedicine implementation in geriatric settings should consider both RTC and PEU. The positive indirect association observed for cognitive rigidity was unexpected and should be replicated before specific practical recommendations are made.

(JMIR Hum Factors 2026;13:e95371) doi:10.2196/95371

KEYWORDS

telemedicine; anxiety; health personnel; geriatrics; surveys and questionnaires; technology acceptance; Israel

Introduction

Background

Telemedicine, defined as the use of telecommunications technology to provide medical care and health services remotely [1], has emerged as a transformative force in health care delivery [2,3]. This technology encompasses a variety of applications, including virtual consultations, remote patient monitoring, and mobile health apps, which allow health care providers to diagnose, treat, and manage patients without the need for an in-person visit [4]. Despite its potential to increase access to health care, improve patient outcomes, and enhance the efficiency of health care delivery as documented by Alotaibi and Federico [5], the World Health Organization [6], and Bashshur et al [7], telemedicine has faced inconsistent uptake among health care professionals due to various barriers [8,9], with resistance to change (RTC) emerging as a particularly significant obstacle within the Technology Acceptance Model (TAM) framework [10].

RTC represents a multifaceted construct encompassing routine seeking, emotional reactivity, short-term focus, and cognitive rigidity [11-13]. More specifically, it refers to the emotional

XSL•FO
RenderX

and behavioral response of individuals or groups when faced with new ideas, practices, or technologies [14-16]. This resistance can manifest as skepticism, reluctance, or outright opposition to adopting changes, often hindering the implementation of new initiatives or technologies, such as telemedicine in health care settings [17]. Oreg [13,18] developed the RTC scale, which was later validated across 17 nations [19], demonstrating its universal applicability. Understanding the dimensions of routine seeking, emotional reactivity, short-term focus, and cognitive rigidity of this construct is crucial for addressing barriers to telemedicine adoption.

The first dimension, routine seeking, reflects the tendency of individuals to prefer established habits and processes over new methods or technologies [20]. This preference often arises from a desire for stability and familiarity, leading to reluctance in embracing changes that disrupt existing routines, as noted by Kotter [21]. The second dimension, emotional reactivity, plays a crucial role as it encompasses the immediate emotional responses individuals experience when faced with change [14,15]. Emotions such as fear, anxiety, or frustration can significantly impact one's willingness to adopt new practices or technologies [22]. The third dimension, short-term focus, is characterized by the inclination to prioritize immediate outcomes and benefits over long-term gains. Individuals with a short-term perspective may resist change due to their concern for the immediate implications of new practices, often overlooking potential future advantages [17]. Finally, cognitive rigidity represents a key element, reflecting the inability or unwillingness to adapt one's thinking or behavior in response to new information or changing circumstances. This rigidity can result in a steadfast adherence to old ways of thinking, making it challenging for individuals to accept and implement change [13].

The adoption and use of telemedicine technologies have been significantly influenced by the interaction between RTC, perceived ease of use (PEU), and anxiety [17,23]. Within the TAM framework, research has shown that the PEU of technology significantly influences an individual's intention to use it, particularly in health care settings [17,24]. The TAM, originally developed by Davis [10], posits that PEU and perceived usefulness are primary determinants of technology acceptance. Recent extensions of TAM and related dual-factor models have added inhibiting factors, including RTC and technology-related anxiety, to the original focus on facilitating beliefs such as PEU [14,17,24]. When users perceive a technology as user-friendly, their likelihood of adopting it increases, as they feel more confident in navigating its features [25]. However, this relationship can be moderated by anxiety related to using technology. Individuals who experience higher levels of anxiety may struggle to engage with even the simplest digital platforms, leading to a decrease in their intention to use such technologies [26,27]. This relationship suggests that while ease of use is a critical factor in technology adoption, addressing the psychological barriers, such as anxiety, is essential for enhancing user intent [28].

Anxiety can significantly impede an individual's perception of ease of use when engaging with telemedicine technologies, as heightened feelings of apprehension may lead to a lack of confidence in navigating these platforms [26]. When health care professionals experience heightened anxiety about adopting new technologies, their perception of these technologies' ease of use tends to decline [29]. Consequently, this diminished perception of ease can adversely affect the intention to use telemedicine services, as users may avoid adopting technologies that they find intimidating or challenging to operate [30]. Furthermore, prior studies have reported an association between RTC and intention to use telemedicine [20,31] as well as an indirect association through PEU [32].

Research has also associated telemedicine-related anxiety with less favorable usability perceptions and weaker intentions to use digital technologies [26,27,29,30,33]. However, previous studies have generally examined these relationships separately, focusing on RTC and telemedicine acceptance [17,24], PEU and intention to use [23], or telemedicine-related anxiety and technology acceptance [27,33]. It is therefore unclear whether telemedicine-related anxiety changes the strength of the association between PEU and intention to use and whether the indirect association between RTC and intention to use through PEU differs across anxiety levels.

Previous studies have also tended to treat RTC as a single overall construct rather than examining its 4 dimensions separately. Although RTC comprises routine seeking, emotional reactivity, short-term focus, and cognitive rigidity [13,18,19], limited evidence is available on whether these dimensions show the same pattern in relation to telemedicine acceptance. Reliance on an overall RTC score may conceal dimension-specific associations with PEU and intention to use. Governmental GMCs provide a theoretically relevant setting for this model because telemedicine must be incorporated into established multidisciplinary workflows rather than adopted as a stand-alone technology [11,34]. In this setting, resistance may shape perceptions of ease of use, while anxiety may weaken the association between PEU and intention to use. The present study extends recent TAM and dual-factor models by examining RTC both as an overall construct and as 4 separate dimensions within 1 conditional process model. It tests whether RTC is associated with intention to use through PEU, whether this indirect association differs across anxiety levels, and whether the same pattern applies to all 4 RTC dimensions [32,35,36].

Main Research Question and Hypotheses

The main research question is as follows: Is the indirect relationship between RTC (overall and its subcategories: routine seeking, emotional reactivity, short-term focus, and cognitive rigidity) and intention to use telemedicine via the mediator ease of use moderated by anxiety among health care professionals in governmental geriatric hospitals in Israel?

Accordingly, the hypotheses are as follows:

- H1: Higher RTC, as well as higher levels of each dimension, will be associated with lower PEU of telemedicine.
- H2: Higher PEU will be associated with a stronger intention to use telemedicine.
- H3: PEU will mediate the association between RTC and intention to use telemedicine. Specifically, higher RTC will be associated with lower PEU, which, in turn, will be

XSL•FO
RenderX

associated with a weaker intention to use telemedicine. This indirect association will also be examined for 4 dimensions of RTC.

- H4: Telemedicine-related anxiety will moderate the indirect association between RTC and intention to use telemedicine through PEU. Specifically, the indirect association will become weaker as telemedicine-related anxiety increases.

Methods

This research used a cross-sectional study design to examine the phenomenon at a single point in time [37]. The study methodology adhered to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines to ensure methodological rigor and reporting transparency [38]. A completed STROBE checklist is provided in Checklist 1.

Participants

Study participants comprised health care professionals employed in four Israeli governmental GMCs: Dorot Medical Center, Fliman Medical Center, Shmuel Harofeh Medical Center, and Shoham Medical Center. The sample consisted of 377 health care professionals representing caregivers, including physicians, nurses, physiotherapists, speech-language therapists, occupational therapists, and clinical dietitians. At the time of data collection, the target population in these 4 centers totaled 1184 professionals: 146 physicians, 790 nurses, and 248 allied health professionals.

Participant inclusion criteria encompassed health care professionals (physicians, nurses, and multidisciplinary staff including physiotherapists, speech therapists, occupational therapists, and clinical dietitians) employed within the specified governmental GMCs (Dorot, Fliman, Shmuel Harofeh, and Shoham), maintaining employment tenure of at least 6 months, demonstrating Hebrew language proficiency, and providing voluntary informed consent for participation. At the time of data collection, telemedicine use in the participating centers was limited and optional and consisted mainly of remote patient monitoring. The questionnaire assessed telemedicine in general and did not distinguish among audio-only, video-based, remote-monitoring, or hybrid modalities, or between dyadic and triadic interactions. Telemedicine use was neither mandated nor actively discouraged by hospital management. Exclusion criteria consisted of health care professionals with less than 6 months of employment at the GMCs, insufficient Hebrew language competency, and declined consent for research participation. The study sought to reach the entire staff population. Visits were conducted in every department across all work shifts, and questionnaires were offered to all staff present. A total of 377 professionals completed the questionnaire, corresponding to 31.8% of the 1184 professionals employed in the 4 centers. The 377 questionnaires included in the analyses were fully completed. The sample's professional composition was similar to that of the target population (nurses: n=256, 67.9% vs physicians: n=46, 12.2% vs allied health: n=75, 19.9%).

Procedure

Data collection operations were supervised by a designated research coordinator responsible for managing the comprehensive data gathering process across all participating GMCs. The research coordinator maintained ongoing communication with field surveyors and medical center administration while providing oversight for the complete quantitative data collection protocol.

The data acquisition was conducted through self-administered questionnaires distributed by trained student surveyors from the University's School of Health Sciences. Survey implementation followed prescheduled department visits coordinated with departmental supervisors, during which staff members were invited to complete anonymous self-report questionnaires addressing the research variables. Supervisors and hospital managers were not present during questionnaire completion, and data were collected by independent academic surveyors who were not employed by the participating centers. Questionnaires were completed privately in designated rooms and were collected anonymously. Data collection was conducted between March and June 2024 across the 4 participating governmental GMCs.

Upon questionnaire completion, surveyors secured all materials in sealed envelopes and transferred them to the research coordinator for subsequent coding and data entry procedures. This systematic approach ensured data integrity and maintained participant anonymity throughout the collection process.

Instruments

The data collection for this study was carried out using a structured questionnaire that included 5 essential measurement scales designed to assess various aspects of telemedicine use and the participants' reactions.

The first measurement scale focused on telemedicine-related anxiety. Anxiety levels concerning the use of telemedicine in patient care settings were evaluated through a 4-item instrument based on the work of Venkatesh et al [27]. A sample item in this scale was as follows: "I experience concern when using telemedicine in patient care contexts." Participants' responses were captured on a 6-point scale, with 1 indicating strong disagreement and 6 indicating strong agreement. The reliability of this scale, measured by Cronbach α, was found to be α=.89, suggesting a high level of internal consistency.

The second measurement scale addressed RTC, a critical factor in understanding how health care professionals adapt to new technologies such as telemedicine. This scale was based on Oreg [18] 17-item instrument, which examines 4 distinct dimensions of change resistance.

The first dimension, routine seeking, was measured using 5 items. A sample item from this dimension was as follows: "I typically view changes (such as telemedicine implementation) negatively within my workplace." The reliability of this scale was α=.75, indicating moderate internal consistency.

The second dimension, emotional reactivity, consisted of 4 items. A sample item in this category was as follows: "Learning about significant workplace changes (like telemedicine adoption)

XSL•FO
RenderX

would likely increase my stress levels." This dimension had a reliability of α=.84, demonstrating strong internal consistency.

The third dimension, short-term focus, was assessed using 4 items. An example item was as follows: "Workplace plan modifications (including telemedicine integration) create considerable inconvenience for me." This dimension demonstrated a high level of reliability with α=.88.

The final dimension, cognitive rigidity, was evaluated using 3 items. A sample item for this dimension was as follows: "I rarely modify my opinions once formed." Initially, the reliability for this dimension was suboptimal, with α=.58. However, after removing one poorly performing item, the reliability increased to an acceptable level, with α=.75.

Each of these scales was carefully selected to provide comprehensive insight into the factors influencing health care professionals' attitudes toward telemedicine and their readiness for its integration into their work environments.

The full scale for measuring RTC showed satisfactory internal consistency (α=.86). The scale used a 6-point response format (1=strongly disagree, 6=strongly agree) for all items.

Perceptions of telemedicine's ease of use were assessed using a 6-item adaptation of Davis [10] PEU scale. On a 7-point Likert continuum (1="strongly disagree," 7="strongly agree"), respondents evaluated statements such as "Acquiring the skills needed to use telemedicine in my workplace would be straightforward for me." The scale demonstrated excellent internal consistency (α=.94). Furthermore, intentions to adopt telemedicine were measured with a 3-item behavioral-intention instrument, also derived from Davis [10]. Using the same 7-point response format, participants indicated their agreement with items like "I plan to incorporate telemedicine into my professional practice within the next year." Reliability for this measure was outstanding (α=.95).

To place these perceptions and intentions in context, a structured demographic questionnaire captured key personal and professional characteristics, including age, gender, highest educational qualification, professional role, years of health care experience, and current employment status within geriatric care.

All instruments were translated from English into Hebrew and back-translated into English, with discrepancies reviewed and resolved to ensure semantic and conceptual equivalence. The instruments were based on previously validated scales; prior to the main data collection, the Hebrew versions were pilot-tested among health care professionals working in governmental geriatric facilities to assess item clarity and cultural appropriateness, while internal consistency was evaluated in the full study sample.

Statistical Analysis

Statistical analyses were performed using the SPSS software (version 28).

Prior to analysis, all questionnaires were screened for missing values at both the item and scale levels, and no missing values were identified on any study variable. This completeness is attributable to the mode of administration: questionnaires were completed in the presence of a trained surveyor, who remained available to clarify any item that respondents found unclear and reviewed each questionnaire for unanswered items upon collection. Accordingly, all 377 cases contributed to every analysis.

Multicollinearity was assessed separately for each moderated mediation model using variance inflation factor (VIF) and tolerance statistics. VIF values below 5 and tolerance values above 0.20 were considered acceptable.

Analytical procedures encompassed descriptive statistical measures, frequency distributions, percentage calculations, internal consistency assessments, and correlation analyses. The research hypotheses were evaluated through PROCESS Model 14 (version 4.3) to examine moderated mediation pathways. Separate moderated mediation analyses were also conducted for each RTC dimension (routine seeking, emotional reactivity, short-term focus, and cognitive rigidity) to examine whether the moderated mediation pattern was consistent across the 4 dimensions. These analytical approaches enabled a comprehensive investigation of the relationships among study variables while accounting for the proposed moderated mediation effects within the theoretical model.

Ethical Considerations

Ethical approval for this investigation was obtained from the Ethics Committee of the School of Health Sciences, Ariel University, under protocol code AU-HEA-GG-20231110 - 1. Study participants received comprehensive information regarding research objectives and were provided with assurances concerning data confidentiality and participant anonymity. Participation remained entirely voluntary, with individuals retaining the right to discontinue involvement at any point without penalty or adverse consequences. All participants provided written informed consent before participation.

Data security protocols were implemented to ensure secure storage, with access restricted exclusively to authorized research personnel. Publication of findings adhered to aggregate reporting standards, eliminating any identifying information pertaining to individual participants or participating health care institutions. These measures ensured full compliance with research ethics standards and participant protection requirements.

Results

The mean age varied across professional groups: physicians (mean 40.66, SD 11.51 y), nurses (mean 42.73, SD 12.86 y), and other multidisciplinary staff (mean 37.25, SD 10.02 y), with nurses being the oldest group on average and multidisciplinary staff being the youngest. For other background characteristics, see Table 1.

XSL•FO
RenderX

Table . Background characteristics by professional groups[a].

Characteristic and categories	Nurses (n=256, 67.9%), n (%)	Physicians (n=46, 12.2%), n (%)	Multidisciplinary staff (n=75, 19.9%), n (%)	Chi-square (*df*)
Children				11.09[b] (2)
Yes	192 (75.0)	24 (52.2)	48 (64.0)	
No	64 (22.3)	22 (42.9)	27 (36.0)	
Religion				45.39[c] (8)
Jewish	108 (42.2)	15 (32.6)	55 (73.3)	
Muslim	117 (45.7)	23 (50)	16 (21.3)	
Christian	17 (6.6)	1 (2.2)	2 (2.7)	
Druze	0 (0)	0 (0)	1 (1.3)	
Other	14 (5.5)	7 (15.2)	1 (1.3)	
Level of religiosity				38.85[c] (8)
Secular	75 (29.3)	21 (45.7)	42 (56.0)	
Traditional	110 (43.0)	12 (26.1)	11 (14.7)	
Religious	50 (19.5)	6 (13.0)	18 (24.0)	
Very religious	7 (2.7)	0 (0)	2 (2.7)	
Other	14 (5.5)	7 (15.2)	2 (2.7)	
Work shifts				155.56[c] (2)
Morning	51 (19.9)	28 (59.5)	69 (93.2)	
Evening or night shifts	205 (80.1)	18 (40.5)	6 (6.8)	

[a]Percentages are calculated within professional groups.

[b]*P*<.01.

[c]*P*<.001.

Table 1 shows that most nurses (n=192, 77.7%) and multidisciplinary staff (n=48, 64.0%) had children, compared to 57.1% (n=24) of the physicians. Religious affiliation varied significantly, with multidisciplinary staff being predominantly Jewish (n=55, 73.3%), while nurses and physicians showed more diversity between Jewish and Muslim affiliations. The majority of nurses worked evening or night shifts (n=205, 84.4%), whereas most physicians (n=28, 59.5%) and particularly multidisciplinary staff (n=69, 93.2%) worked morning shifts.

Multicollinearity diagnostics indicated no problematic multicollinearity across the 5 models, with VIF values ranging from 1.064 to 1.504 and tolerance values ranging from 0.665 to 0.940.

Table 2 shows that telemedicine-related anxiety demonstrated positive correlations with RTC variables, particularly with emotional reactivity (r=0.55; P<.001) and routine seeking (r=0.51; P<.001). Conversely, PEU showed negative correlations with all resistance dimensions except cognitive rigidity, which exhibited a positive correlation (r=0.28; P<.001). The strongest relationship emerged between PEU and intention to use telemedicine (r=0.69; P<.001).

Before evaluating the research hypotheses through PROCESS Model 14 (version 4.3) to examine moderated mediation pathways, all variables were standardized (z scores) to facilitate interpretation. The index of moderated mediation for the indirect association between RTC and intention to use through ease of use was statistically significant (index=0.016, 95% CI 0.003-0.035), supporting the presence of a moderated indirect effect. Table 3 presents a detailed description of all analysis steps.

XSL•FO
RenderX

Table . Correlation matrix, means, and SDs for model variables.

Variable	1	2	3	4	5	6	7	8
Routine seeking	1.00	0.68[a]	0.70[a]	0.13[b]	0.86[a]	0.51[a]	−0.31[a]	−0.32[a]
Emotional reactivity	0.68[a]	1.00	0.74[a]	0.16[a]	0.87[a]	0.55[a]	−0.27[a]	−0.23[a]
Short-term focus	0.70[a]	0.74[a]	1.00	0.16[a]	0.88[a]	0.43[a]	−0.30[a]	−0.25[a]
Cognitive rigidity	0.13[b]	0.16[a]	0.16[a]	1.00	0.40[a]	0.10	0.28[a]	0.12[b]
Total RTC[c]	0.86[a]	0.87[a]	0.88[a]	0.40[a]	1.00	0.54[a]	−0.24[a]	−0.25[a]
Telemedicine-related anxiety	0.51[a]	0.55[a]	0.43[a]	0.10	0.54[a]	1.00	−0.31[a]	−0.31[a]
PEU[d]	−0.31[a]	−0.27[a]	−0.30[a]	0.28[a]	−0.24[a]	−0.31[a]	1.00	0.69[a]
Intention to use telemedicine	−0.32[a]	−0.23[a]	−0.25[a]	0.12[b]	−0.25[a]	−0.31[a]	0.69[a]	1.00
Mean (SD)	3.03 (1.00)	3.01 (1.11)	2.61 (1.16)	3.91 (1.15)	3.08 (0.85)	3.50 (1.39)	4.53 (1.25)	4.01 (1.46)

[a]*P*<.001.

[b]*P*<.05.

[c]RTC: resistance to change.

[d]PEU: perceived ease of use.

Table . Moderated mediation model predicting intention to use through ease of use, with anxiety as moderator.

Model components	*B* (SE; 95% CI)	*t* test (*df*)	*P* value
Predicting the mediator: ease of use			
RTC[a]	−0.24 (0.05; −0.34 to −0.14)	−4.72 (375)	<.001
Conditional effects of ease of use on intention at different anxiety levels			
Low	0.7 (0.05; 0.61 to 0.80)	14.4 (372)	<.001
Mean	0.62 (0.04; 0.54 to 0.70)	15.01 (372)	<.001
High	0.56 (0.06; 0.45 to 0.68)	9.58 (372)	<.001
Predicting the dependent variable: intention to use			
Resistance	−0.05 (0.04; −0.14 to 0.04)	−1.05 (372)	.29
Ease of use	0.63 (0.04; 0.55 to 0.71)	15.88 (372)	<.001
Conditional indirect effects at different anxiety levels			
Low	−0.17 (0.05; −0.27 to −0.08)	—[b]	—
Mean	−0.15 (0.04; −0.24 to −0.07)	—	—
High	−0.14 (0.04; −0.22 to −0.06)	—	—

[a]RTC: resistance to change.

[b]Not applicable.

RTC was negatively associated with ease of use (*B*=−0.24, SE=0.05, t_{375}=−4.72; *P*<.001), which in turn was positively associated with intention to use (*B*=0.63, SE =0.04, t_{372}=15.88; *P*<.001). The significance of the mediation effect was estimated using 95% CIs calculated based on 5000 bootstrap samples. Since zero was not included in the CIs at all levels of anxiety, the indirect relationships were significant: anxiety (indirect effect =−0.15, 95% CI −0.236 to −0.072).

The interaction between ease of use and anxiety was significant (*B*=−0.07, SE=0.03, t_{372}=−1.98; *P*=.04). The decomposition of the interaction into conditional effects of ease of use on intention at different levels of anxiety demonstrates that the positive relationship between ease of use and intention to use weakened as anxiety increased: low anxiety (*B*=0.70, SE=0.05, t_{372}=14.40; *P*<.001), mean anxiety (*B*=0.62, SE=0.04, t_{372}=15.01; *P*<.001), and high anxiety (*B*=0.56, SE=0.06, t_{372}=9.58; *P*<.001).

XSL•FO
RenderX

The conditional indirect relationships show that the negative effect of RTC on intention to use through reduced ease of use weakens as anxiety increases. However, even at high anxiety levels, the indirect relationship remains significant. Therefore, it appears that while anxiety moderates the strength of the indirect effect, it cannot eliminate it entirely.

Figure 1 illustrates how anxiety moderates the relationship between ease of use and intention to use. Three lines represent different levels of anxiety: low (1 SD below the mean), mean, and high (1 SD above the mean). The positive relationship between ease of use and intention to use is the strongest at low anxiety levels (slope =0.70) and weakest at high anxiety levels (slope =0.56), though the relationship remains positive and significant across all anxiety levels. This pattern suggests that while ease of use was consistently associated with intention to use, the magnitude of this association was smaller at higher levels of anxiety.

Figure 1. The moderating effect of anxiety on the relationship between ease of use and intention to use.

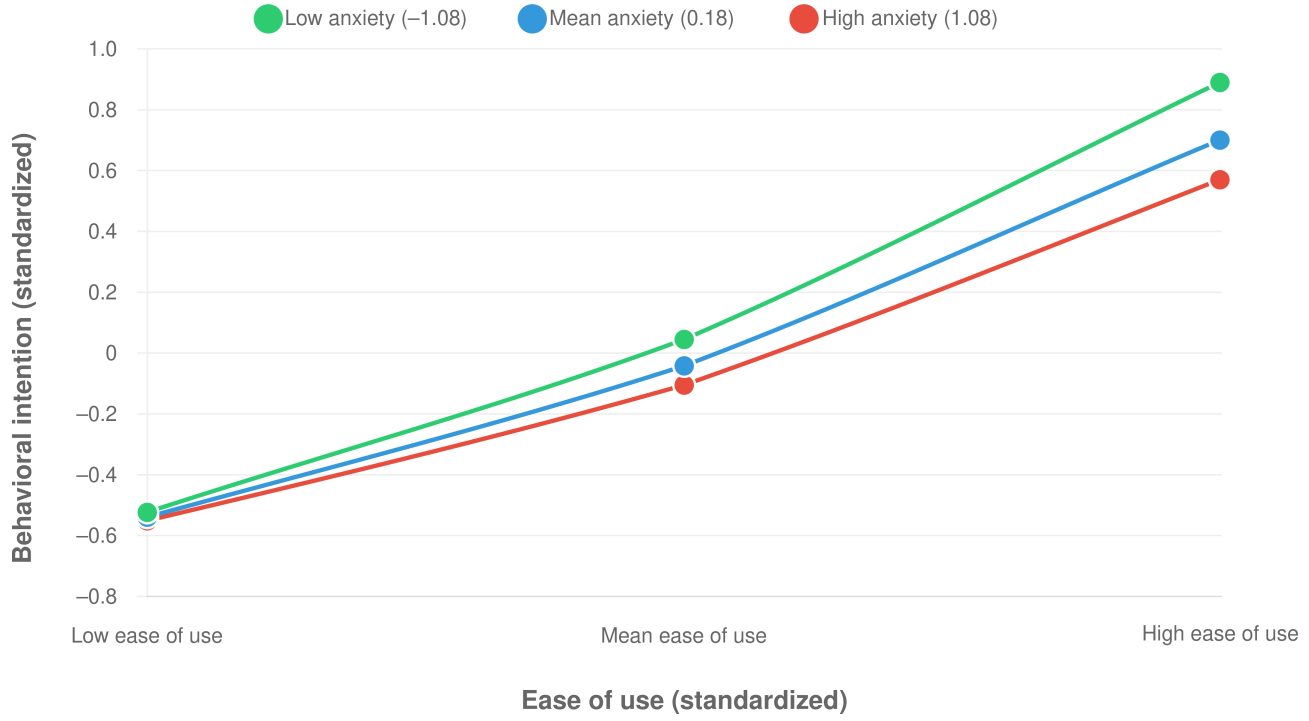
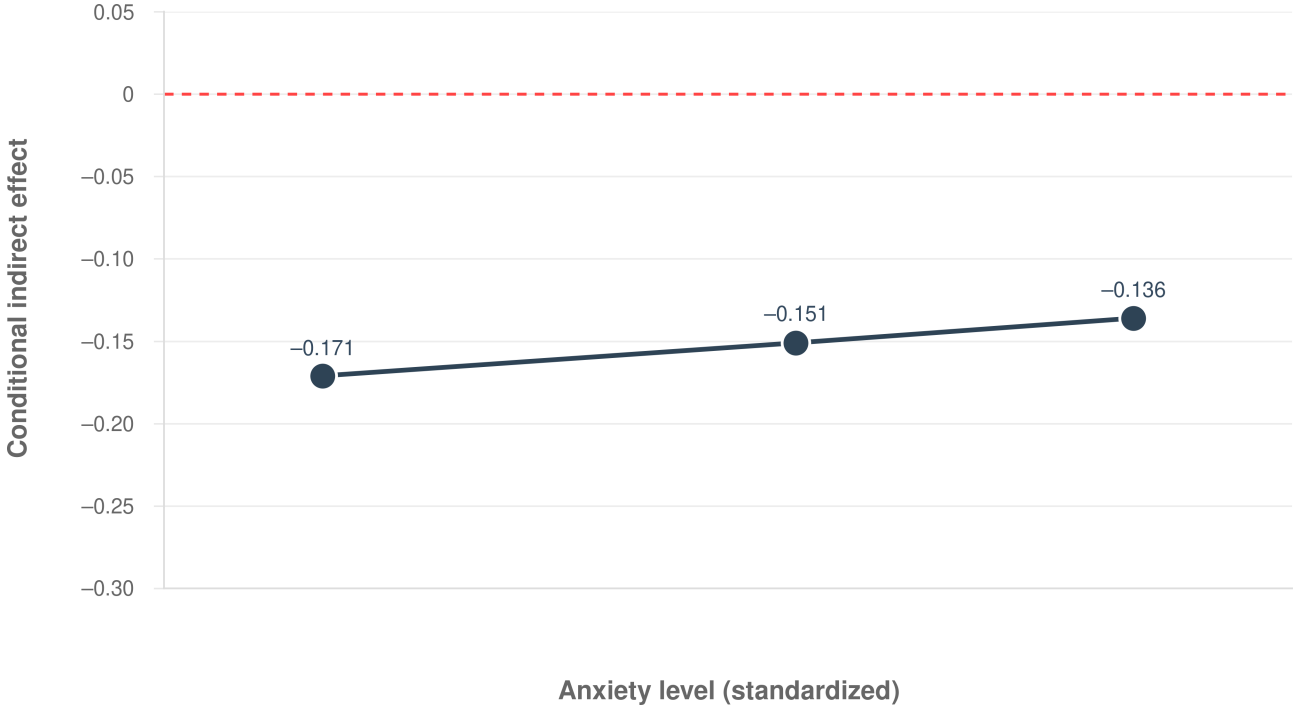


Figure 2 shows that the negative indirect association between RTC and intention to use through ease of use decreases as anxiety increases, from −0.17 at low anxiety to −0.14 at high anxiety. All indirect effects are statistically significant, as indicated by CIs that do not include zero. The pattern demonstrates that while anxiety reduces the indirect effect, it remains significant across all anxiety levels examined.

The hypothesized moderated mediation model, in which ease of use mediated the association between RTC and intention to use, with anxiety moderating the association between ease of use and intention to use, is presented in Figure 3.

XSL•FO
RenderX

Figure 2. Conditional indirect effects of resistance to change (RTC) on intention to use through ease of use at different levels of anxiety.

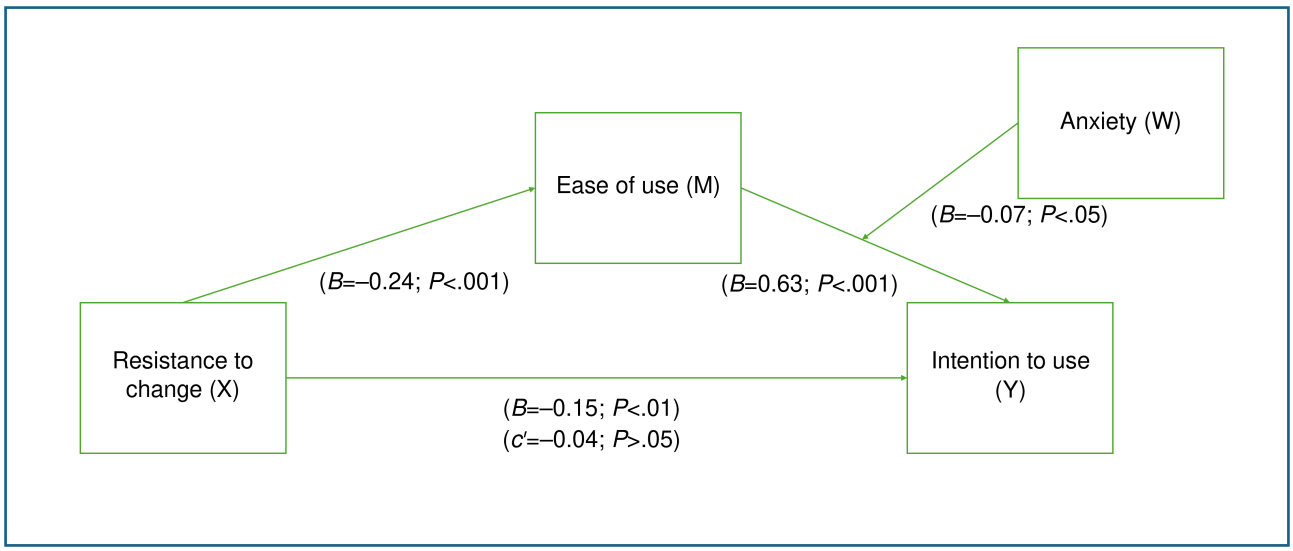


Following the main analysis, each of the 4 dimensions of RTC (routine seeking, emotional reactivity, short-term focus, and cognitive rigidity) was examined separately as an independent variable using Model 14 to test the moderated mediation hypothesis.

The index of moderated mediation for the indirect relationship between routine seeking and intention to use through ease of use, with anxiety as the moderator, was statistically significant (index =0.020, 95% CI 0.003-0.042), supporting the presence of a moderated indirect effect. Table S1 presents a detailed description of all analysis steps in Multimedia Appendix 1.

Routine seeking was negatively associated with ease of use (B=–0.32, SE=0.05, t_{375}=−6.41; P<.001), which in turn was positively associated with intention to use (B=0.62, SE=0.04, t_{372}=15.50; P<.001). The significance of the mediation effect was estimated using 95% CIs calculated based on 5000 bootstrap samples. Since zero was not included in the CIs at all levels of anxiety, the indirect relationships were significant across all anxiety levels.

Figure 3. The research model.



As presented in the results, the interaction between ease of use and anxiety was not significant (B=−0.06, SE=0.03, t_{372}=−1.83; P=.07). Although the simple interaction effect was marginally nonsignificant, Hayes [39] argues that the index of moderated mediation is the more appropriate and sensitive test for detecting moderation in the indirect pathway, as it directly examines moderation of the mediated effect through bootstrap procedures rather than relying on the traditional interaction test which may lack sufficient power to detect subtle moderation effects.

XSL•FO
RenderX

The conditional effects show that ease of use remains positively associated with intention to use across all anxiety levels: low anxiety (B=0.69, SE=0.05, t_{372}=13.84; P<.001), mean anxiety (B=0.61, SE=0.04, t_{372}=14.71; P<.001), and high anxiety (B=0.56, SE=0.06, t_{372}=9.53; P<.001). The conditional indirect relationships show that the negative effect of routine seeking on intention to use through reduced ease of use weakens as anxiety increases: low anxiety (indirect effect =−0.22, 95% CI −0.30 to −0.15), mean anxiety (indirect effect=−0.20, 95% CI −0.27 to −0.13), and high anxiety (indirect effect=−0.18, 95% CI −0.26 to −0.11). All indirect effects remained statistically significant across anxiety levels, indicating that while anxiety moderates the strength of the indirect effect, it cannot eliminate it entirely.

The index of moderated mediation for the indirect relationship between emotional reactivity and intention to use through ease of use, with anxiety as the moderator, was statistically significant (index=0.019, 95% CI 0.003-0.040), supporting the presence of a moderated indirect effect. Table 4 presents a detailed description of all analysis steps.

Table . Moderated mediation model predicting intention to use through ease of use, with anxiety as moderator.

Model component	*B* (SE; 95% CI)	*t* test (*df*)	*P* value
Predicting the mediator: ease of use			
Emotional reactivity	−0.27 (0.05; −0.38 to −0.17)	−5.38 (375)	<.001
Predicting the dependent variable: intention to use			
Emotional reactivity	0.003 (0.05; −0.09 to 0.09)	.06 (372)	.95
Ease of use	0.64 (0.04; 0.56 to 0.71)	15.87 (372)	.001
Anxiety	−0.10 (0.05; −0.19 to −0.01)	−2.09 (372)	.03
Ease of use × anxiety	−0.07 (0.03; −0.14 to −0.004)	−2.08 (372)	.03
Conditional effects of ease of use on intention at different anxiety levels			
Low	0.71 (0.05; 0.62 to 0.81)	14.51 (372)	<.001
Mean	0.62 (0.04; 0.54 to 0.71)	15.00 (372)	<.001
High	0.56 (0.06; 0.45 to 0.68)	9.55 (372)	<.001
Conditional indirect effects at different anxiety levels			
Low	−0.20 (0.05; −0.29 to −0.11)	—[a]	—
Mean	−0.17 (0.04; −0.26 to −0.09)	—	—
High	−0.15 (0.04; −0.24 to −0.08)	—	—

[a]Not applicable.

Emotional reactivity was negatively associated with ease of use (B=−0.27, SE=0.05, t_{375}=−5.38; P<.001), which in turn was positively associated with intention to use (B=0.64, SE=0.04, t_{372}=15.87; P<.001). The significance of the mediation effect was estimated using 95% CIs calculated based on 5000 bootstrap samples. Since zero was not included in the CIs at all levels of anxiety, the indirect relationships were significant across all anxiety levels.

The interaction between ease of use and anxiety was significant (B=−0.07, SE=0.03, t_{372}=−2.08; P=.04). The conditional effects show that ease of use remains positively associated with intention to use across all anxiety levels: low anxiety (B=0.71, SE=0.05, t_{372}=14.51; P<.001), mean anxiety (B=0.62, SE=0.04, t_{372}=15.00; P<.001), and high anxiety (B=0.56, SE=0.06, t_{372}=9.55; P<.001).

The conditional indirect relationships show that the negative effect of emotional reactivity on intention to use through reduced ease of use weakens as anxiety increases: low anxiety (indirect effect=−0.20, 95% CI −0.29 to −0.11), mean anxiety (indirect effect=−0.17, 95% CI −0.26 to −0.09), and high anxiety (indirect effect=−0.15, 95% CI −0.24 to −.08). All indirect effects remained statistically significant across anxiety levels, indicating that while anxiety moderates the strength of the indirect effect, it cannot eliminate it entirely.

The index of moderated mediation for the indirect relationship between short-term focus and intention to use through ease of use, with anxiety as the moderator, was statistically significant (index=0.021, 95% CI 0.004-0.044), supporting the presence of a moderated indirect effect. Table S2 in Multimedia Appendix 1 presents a detailed description of all analysis steps.

Short-term focus was negatively associated with ease of use (B=−0.30, SE=0.05, t_{375}=−5.86; P<.001), which in turn was positively associated with intention to use (B=0.64, SE=0.04, t_{372}=15.80; P<.001). The significance of the mediation effect was estimated using 95% CIs calculated based on 5000 bootstrap samples. Since zero was not included in the CIs at all levels of anxiety, the indirect relationships were significant across all anxiety levels.

The interaction between ease of use and anxiety was significant (B=−0.07, SE=0.03, t_{372}=−2.07; P=.04). The conditional effects show that ease of use remains positively associated with intention to use across all anxiety levels: low anxiety (B=0.71,

XSL•FO
RenderX

SE=0.05, t_{372}=14.24; P<.001), mean anxiety (B=0.62, SE=0.04, t_{372}=14.96; P<.001), and high anxiety (B=0.56, SE=0.06, t_{372}=9.56; P<.001). The conditional indirect relationships show that the negative effect of short-term focus on intention to use through reduced ease of use weakens as anxiety increases: low anxiety (indirect effect=−0.21, 95% CI −0.31 to −0.12), mean anxiety (indirect effect=−0.18, 95% CI −0.27 to −0.10), and high anxiety (indirect effect=−0.17, 95% CI −0.25 to −0.09). All indirect effects remained statistically significant across anxiety levels, indicating that while anxiety moderates the strength of the indirect effect, it cannot eliminate it entirely.

The index of moderated mediation for the indirect relationship between cognitive rigidity and intention to use through ease of use, with anxiety as the moderator, was statistically significant (index=−0.020, 95% CI −0.043 to −0.004), supporting the presence of a moderated indirect effect. Table 5 presents a detailed description of all analysis steps.

Table . Moderated mediation model predicting intention to use through ease of use, with anxiety as moderator.

Model component	*B* (SE; 95% CI)	*t* test (*df*)	*P* value
Predicting the mediator: ease of use			
Cognitive rigidity	0.27 (0.05; 0.17 to 0.37)	5.27 (375)	<.001
Predicting the dependent variable: intention to use			
Cognitive rigidity	−0.07 (0.04; −0.15 to 0.01)	−1.71 (372)	.08
Ease of use	0.66 (0.04; 0.58 to 0.74)	15.82 (372)	<.001
Anxiety	−0.08 (0.04; −0.16 to −0.0004)	−1.98 (372)	.04
Ease of use × anxiety	−0.08 (0.03; −0.14 to −0.01)	−2.24 (372)	.02
Conditional effects of ease of use on intention at different anxiety levels			
Low	0.74 (0.05; 0.64 to 0.84)	14.48 (372)	<.001
Mean	0.64 (0.04; 0.56 to 0.73)	15.02 (372)	<.001
High	0.58 (0.06; 0.46 to 0.69)	9.75 (372)	<.001
Conditional indirect effects at different anxiety levels			
Low	0.20 (0.04; 0.12 to 0.28)	—[a]	—
Mean	0.17 (0.04; 0.10 to 0.25)	—	—
High	0.16 (0.04; 0.09 to 0.23)	—	—

[a]Not applicable.

Cognitive rigidity was positively associated with ease of use (B=0.27, SE=0.05, t_{375}=5.27; P<.001), which in turn was positively associated with intention to use (B=0.66, SE=0.04, t_{372}=15.82; P<.001). The significance of the mediation effect was estimated using 95% CIs calculated based on 5000 bootstrap samples. Since zero was not included in the CIs at all levels of anxiety, the indirect relationships were significant across all anxiety levels.

The interaction between ease of use and anxiety was significant (B=–0.08, SE=0.03, t_{372}=–2.24; P=.02). The conditional effects show that ease of use remains positively associated with intention to use across all anxiety levels: low anxiety (B=0.74, SE=0.05, t_{372}=14.48; P<.001), mean anxiety (B=0.64, SE=0.04, t_{372}=15.02; P<.001), and high anxiety (B=0.58, SE=0.06, t_{372}=9.75; P<.001). The conditional indirect relationships show that the positive effect of cognitive rigidity on intention to use through increased ease of use weakens as anxiety increases: low anxiety (indirect effect=0.20, 95% CI 0.12-0.28), mean anxiety (indirect effect=0.17, 95% CI 0.10-0.25), and high anxiety (indirect effect=0.16, 95% CI 0.09-0.23). All indirect effects remained statistically significant across anxiety levels, indicating that while anxiety moderates the strength of the indirect effect, it cannot eliminate it entirely.

Discussion

Principal Findings

The results of the moderated-mediation analyses fit the proposed model: caregivers with stronger dispositional RTC reported weaker intention to use telemedicine, and this link was explained mainly by their lower ease of use. The indirect link was weaker when telemedicine-related anxiety was higher, but it stayed statistically significant across the whole range of anxiety observed. Since the data were collected at 1 point in time and rely on self-report, these findings describe associations and not proven cause and effect. The overall indirect effect of RTC remained negative and statistically significant across all anxiety levels. This profile dovetails with provider-side applications of the TAM, in which PEU has consistently been found to mediate associations between inhibiting traits and usage intentions among nurses, physicians, and allied professionals [40,41].

The buffering role of anxiety accords with hospital studies in which technology-related anxieties were associated with a weaker link between usability perceptions and usage intentions. Kummer et al [33], for example, found that anxieties about a sensor-based medication system suppressed the PEU-intention link among ward staff; our data show a comparable attenuation

XSL•FO
RenderX

and are consistent with PEU operating as a mediator even among highly apprehensive users. These findings add to recent TAM extensions by showing that RTC was associated with intention mainly through PEU and that the pattern differed across RTC dimensions. Routine seeking, emotional reactivity, and short-term focus followed the expected negative pathway, whereas cognitive rigidity did not. This suggests that resistance should not be treated as a single, uniform barrier in technology-acceptance models. Telemedicine-related anxiety changed the strength of the PEU-intention association but did not remove it, supporting its role as a boundary condition of the indirect pathway.

The dimension-specific results provide further detail. Routine seeking produced the largest conditional indirect effect, dropping from –0.22 at low anxiety to –0.18 at high anxiety. It is noteworthy that while the simple interaction effect for routine seeking was marginally nonsignificant, the index of moderated mediation remained significant, consistent with Hayes' [39] argument that the bootstrap-based index provides a more sensitive test for detecting moderation in indirect pathways. Prior evidence from electronic-record roll-outs shows that clinicians who prize habitual workflows are especially sensitive to any additional cognitive effort a new system entails, consistent with our finding that PEU accounted for much of the association in this group [40].

Emotional reactivity and short-term focus followed the same gradient, falling from –0.20 to –0.15 and from –0.21 to –0.17, respectively. Both dimensions capture immediate affective or workload costs, which inflate perceived difficulty during the early learning curve and thus translate into lower intention to use. Reports of caregivers abandoning teleconsultation platforms after initial frustration with interface complexity mirror this mechanism [42].

Unlike the other RTC dimensions, cognitive rigidity was positively associated with PEU (B=0.27), and the corresponding indirect association with intention to use decreased from 0.20 at low anxiety to 0.16 at high anxiety. This direction was contrary to H1 and to the usual interpretation of cognitive rigidity as a barrier to change [13]. The result should therefore be treated with caution. One possible explanation is that some telemedicine procedures follow fixed and rule-based steps that may appear straightforward to professionals who prefer structured working methods. However, the present study did not assess task-technology fit or preferences for structured workflows, and the study by Cady and Finkelstein [34] did not examine cognitive rigidity directly. The finding may also reflect characteristics of the present sample or the modified cognitive rigidity subscale. It does not show that cognitive rigidity facilitates telemedicine use and should be replicated in other samples [34]. Across the RTC dimensions, the PEU × telemedicine-related anxiety interaction was statistically significant in the emotional reactivity, short-term focus, and cognitive rigidity models, but not in the routine-seeking model. The conditional indirect associations nevertheless remained statistically significant at all examined anxiety levels. This suggests that PEU may remain relevant even among professionals reporting higher telemedicine-related anxiety. Routine seeking, emotional reactivity, and short-term focus were associated with negative indirect pathways of different magnitudes. These findings may inform future intervention research, but the present study did not evaluate any tailored strategy or its effectiveness. Because the cognitive rigidity finding was unexpected, no specific practical recommendation should be based on it at this stage.

Limitations and Future Research

The findings need to be interpreted in light of several limitations. Because the data were collected at 1 time point, the temporal order of RTC, PEU, telemedicine-related anxiety, and intention to use cannot be established. The indirect associations identified in the PROCESS analyses therefore do not demonstrate causal pathways. Longitudinal research conducted before and after telemedicine implementation would help clarify how these relationships develop over time.

All variables were measured by self-report in the same questionnaire. The results may therefore have been affected by common method bias, shared measurement variance, or socially desirable responding. The study also examined intention to use telemedicine rather than actual use. Future research could combine questionnaires with system logs or administrative records.

Participants were recruited from 4 governmental GMCs within 1 national health system, and the final sample represented 31.8% (n=377) of the eligible workforce. Although staff were approached across departments and work shifts, participation was voluntary, and those who participated may have differed from those who did not. The findings may therefore not apply to other health care settings, countries, professional groups, or telemedicine systems. In addition, telemedicine use in the participating centers was limited, optional, and mainly related to remote patient monitoring. The questionnaire did not distinguish among specific telemedicine modalities or interaction structures, so modality-specific analyses could not be conducted. The findings should therefore be interpreted primarily in relation to this context. Future studies should examine whether the observed associations differ across audio-only, video-based, remote-monitoring, and hybrid models and between dyadic and triadic interactions.

The positive association found for cognitive rigidity also requires caution. One item was removed from this subscale, and the direction of the result was contrary to the original hypothesis. This finding should be examined again in other samples and with additional assessment of cognitive rigidity.

Conclusions

This study indicates that clinicians' RTC was associated with weaker intention to use telemedicine, an association statistically accounted for by lower ease of use, and that this indirect association was smaller at higher levels of telemedicine-related anxiety while remaining significant throughout the observed range (−0.17 at low anxiety vs −0.14 at high). Routine seeking, emotional reactivity, and short-term focus were each associated with the negative pathway. Cognitive rigidity showed an unexpected positive association with PEU and a positive indirect association with intention to use. This finding should be treated as preliminary and replicated before practical implications are

XSL•FO
RenderX

drawn from it. The findings extend dual-factor approaches to technology acceptance by showing that RTC dimensions did not follow a uniform pattern and that telemedicine-related anxiety conditioned the indirect association between RTC and intention to use through PEU. The findings suggest that RTC and PEU may warrant consideration when developing strategies to support telemedicine use intentions in geriatric practice.

Acknowledgments

We wish to express our deep appreciation and gratitude to the following medical centers for their significant contribution to data collection for this research: G. Mendelsson, MD, manager of Dorot Medical Center & Staff; Inna Shugaev, MD, manager of Fliman Medical Center & Staff; Nadya Kahansky, MD, manager of Shmuel Harofeh Medical Center & Staff; and Yehonatan Hershkovits, Fliman Medical Center.

During the preparation of this manuscript, the authors used ChatGPT by OpenAI and DeepL solely for translation support and language editing, including clarity and readability checks. These tools were not used to generate scholarly content, data, analyses, interpretations, references, or conclusions. The English wording was also reviewed by a native English-speaking academic collaborator. The authors reviewed and edited all outputs and take full responsibility for the final manuscript.

Funding

This study was supported by a grant from the Israel National Institute for Health Policy Research.

Data Availability

The data supporting the findings of this study are not publicly available because they contain sensitive information that could compromise participant privacy and confidentiality. Deidentified data may be made available from the corresponding author upon reasonable request, subject to approval by the relevant institutional review procedures and applicable data protection requirements. The questionnaire used in this study is also available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: GG, AL, RT
Investigation: GG, AL, RT
Methodology: GG, AL, TP-P
Software: GG
Supervision: GG, TP-P
Validation: GG, AL, TP-P
Writing – original draft: GG
Writing – review and editing: GG, AL, TP-P, RT

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary moderated mediation analyses for the routine seeking and short-term focus dimensions of resistance to change.
[DOCX File, 18 KB - humanfactors_v13i1e95371_app1.docx]

Checklist 1
STROBE checklist.
[DOCX File, 20 KB - humanfactors_v13i1e95371_app2.docx]

References

1. Singh V, Dev V. Telemedicine adoption in India: identifying factors affecting intention to use. Int J Healthc Inf Syst Inform 2021 Oct;16(4):1-18. [doi: 10.4018/IJHISI.20211001.oa34]
2. Stoumpos AI, Kitsios F, Talias MA. Digital transformation in healthcare: technology acceptance and its applications. Int J Environ Res Public Health 2023 Feb 15;20(4):3407. [doi: 10.3390/ijerph20043407] [Medline: 36834105]
3. Bocean CG, Vărzaru AA. Health status in the era of digital transformation and sustainable economic development. BMC Health Serv Res 2025 Mar 5;25(1):343. [doi: 10.1186/s12913-025-12498-y] [Medline: 40045359]
4. Türkyılmaz S. Investigation of telemedicine services, an innovative and technology-based healthcare application, by using the extended Technology Acceptance Model (TAM2): an example from Turkey. Eur J Appl Sci 2023;11(1):594-621. [doi: 10.14738/aivp.111.14101]

XSL•FO
RenderX

5. Alotaibi YK, Federico F. The impact of health information technology on patient safety. Saudi Med J 2017 Dec;38(12):1173-1180. [doi: 10.15537/smj.2017.12.20631] [Medline: 29209664]
6. Telemedicine: opportunities and developments in Member States: report on the Second Global Survey on eHealth 2009. : World Health Organization; 2010 URL: https://iris.who.int/server/api/core/bitstreams/1d9e2557-20e2-4ab4-bdd7-648f962f866d/content [accessed 2026-09-12]
7. Bashshur RL, Howell JD, Krupinski EA, Harms KM, Bashshur N, Doarn CR. The empirical foundations of telemedicine interventions in primary care. Telemed J E Health 2016 May;22(5):342-375. [doi: 10.1089/tmj.2016.0045] [Medline: 27128779]
8. Sedláček D, Lochmannová A, Šín R, Křivková J, Kovalčinová K, Marek P. Frontline ethical burdens: a mixed-methods investigation of moral distress in emergency medical services. Bratisl Med J 2025;126(7):1463-1471. [doi: 10.1007/s44411-025-00221-6]
9. Sharon C, Hochwald IH, Green G, Lochmannová A. Technology-enriched high-fidelity simulation in nursing education: learning effectiveness and mediating model. Med Sci Educ 2026 Jul 31. [doi: 10.1007/s40670-026-02842-0]
10. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep 1;13(3):319-340. [doi: 10.2307/249008]
11. Schürmann F, Westmattelmann D, Schewe G. Factors influencing telemedicine adoption among health care professionals: qualitative interview study. JMIR Form Res 2025 Jan 27;9:e54777. [doi: 10.2196/54777] [Medline: 39869885]
12. Laumer S, Maier C, Eckhardt A. Why do they resist? An empirical analysis of an individual's personality trait resistance regarding the adoption of new information systems. In: Proceedings of the 18th European Conference on Information Systems (ECIS): Association for Information Systems (AIS) Electronic Library (AISeL); 2010. URL: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1013&context=ecis2010 [accessed 2026-09-12]
13. Oreg S. Personality, context, and resistance to organizational change. Eur J Work Organ Psychol 2006 Mar;15(1):73-101. [doi: 10.1080/13594320500451247]
14. Tsai TH, Lin WY, Chang YS, Chang PC, Lee MY. Technology anxiety and resistance to change behavioral study of a wearable cardiac warming system using an extended TAM for older adults. PLoS One 2020;15(1):e0227270. [doi: 10.1371/journal.pone.0227270] [Medline: 31929560]
15. Lochmannová A. Exploring the role of virtual reality in preparing emergency responders for mass casualty incidents. Isr J Health Policy Res 2025 Apr 9;14(1):22. [doi: 10.1186/s13584-025-00681-9] [Medline: 40205512]
16. Bednar M, Dufek P, Lochmannova A, Simon M, Bures M. Use of virtual reality for education and training of emergency rescue system for crisis situations. Proc Int Conf Educ Technol Comput 2023 Sep 26:142-147. [doi: 10.1145/3629296.3629318]
17. Kautish P, Siddiqui M, Siddiqui A, Sharma V, Alshibani SM. Technology-enabled cure and care: an application of innovation resistance theory to telemedicine apps in an emerging market context. Technol Forecast Soc Change 2023 Jul;192:122558. [doi: 10.1016/j.techfore.2023.122558]
18. Oreg S. Resistance to change: developing an individual differences measure. J Appl Psychol 2003 Aug;88(4):680-693. [doi: 10.1037/0021-9010.88.4.680] [Medline: 12940408]
19. Oreg S, Bayazit M, Vakola M, et al. Dispositional resistance to change: measurement equivalence and the link to personal values across 17 nations. J Appl Psychol 2008 Jul;93(4):935-944. [doi: 10.1037/0021-9010.93.4.935] [Medline: 18642996]
20. Verpaalen IAM, Holland RW, Ritter S, et al. Resistance to contact tracing applications: the implementation process in a social context. Comput Human Behav 2022 Sep;134:107299. [doi: 10.1016/j.chb.2022.107299]
21. Kotter JP. Leading Change: Harvard Business School Press; 1996.
22. Özdemir-Güngör D, Camgöz-Akdağ H. Examining the effects of technology anxiety and resistance to change on the acceptance of breast tumor registry system: evidence from Turkey. Technol Soc 2018 Aug;54:66-73. [doi: 10.1016/j.techsoc.2018.03.006]
23. Fahmi I. The antecedents of intention to use telemedicine. J Consum Sci 2022;7(2):97-114. [doi: 10.29244/jcs.7.2.97-114]
24. Tsai JM, Cheng MJ, Tsai HH, Hung SW, Chen YL. Acceptance and resistance of telehealth: the perspective of dual-factor concepts in technology adoption. Int J Inf Manage 2019 Dec;49:34-44. [doi: 10.1016/j.ijinfomgt.2019.03.003]
25. Safi S, Thiessen T, Schmailzl KJ. Acceptance and resistance of new digital technologies in medicine: qualitative study. JMIR Res Protoc 2018 Dec 4;7(12):e11072. [doi: 10.2196/11072] [Medline: 30514693]
26. Amarantou V, Kazakopoulou S, Chatzoudes D, Chatzoglou P. Resistance to change: an empirical investigation of its antecedents. J Organ Change Manag 2018 Apr 9;31(2):426-450. [doi: 10.1108/JOCM-05-2017-0196]
27. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003 Sep 1;27(3):425-478. [doi: 10.2307/30036540]
28. Yeow YKA, Lim WK. Resisting technology-enabled change within healthcare organization. Acad Manag Proc 2017 Aug;2017(1):15065. [doi: 10.5465/AMBPP.2017.15065abstract]
29. Vadakkemulanjanal Joseph G, Anit Thomas M. The mediation effect of technology anxiety and barriers on technology exposure to teachers' technology adoption. Proc 3rd Int Conf Intell Comput Instrum Control Technol 2022:1737-1741. [doi: 10.1109/ICICICT54557.2022.9918006]

XSL•FO
RenderX

30. Nestik T, Zhuravlev A, Patrakov E, et al. Technophobia as a cultural and psychological phenomenon. Interação Rev Ensino Pesqui Ext 2019;20(1):266-281. [doi: 10.33836/interacao.v20i1.178]
31. Haig EL, Woodcock KA. Rigidity in routines and the development of resistance to change in individuals with Prader-Willi syndrome. J Intellect Disabil Res 2017 May;61(5):488-500. [doi: 10.1111/jir.12368] [Medline: 28266087]
32. Samhan B. Patients' resistance towards health information technology: a perspective of the dual factor model of IT usage. In: Proc Hawaii Int Conf Syst Sci: IEEE Computer Society; 2025:3401-3410. [doi: 10.24251/HICSS.2017.412]
33. Kummer TF, Recker J, Bick M. Technology-induced anxiety: manifestations, cultural influences, and its effect on the adoption of sensor-based technology in German and Australian hospitals. Inf Manag 2017 Jan;54(1):73-89. [doi: 10.1016/j.im.2016.04.002]
34. Cady RG, Finkelstein SM. Task-technology fit of video telehealth for nurses in an outpatient clinic setting. Telemed J E Health 2014 Jul;20(7):633-639. [doi: 10.1089/tmj.2013.0242] [Medline: 24841219]
35. Green G, Lochmannová A, Porat-Packer T. The pathway from loneliness to chronic diseases: examining the mediating roles of depression and self-perceived health, and the moderating effect of age among older adults in Israel. J Appl Gerontol 2026 May;45(5):939-950. [doi: 10.1177/07334648251353283] [Medline: 40570249]
36. Green G. Electronic health literacy among older adults: development and psychometric validation of the Hebrew version of the electronic health literacy questionnaire. Int J Med Inform 2025 Feb;194:105691. [doi: 10.1016/j.ijmedinf.2024.105691] [Medline: 39571390]
37. Zangirolami-Raimundo J, Echeimberg JDO, Leone C. Research methodology topics: cross-sectional studies. J Hum Growth Dev 2018;28(3):356-360. [doi: 10.7322/jhgd.152198]
38. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008 Apr;61(4):344-349. [doi: 10.1016/j.jclinepi.2007.11.008] [Medline: 18313558]
39. Hayes AF. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach, 3rd edition: Guilford Press; 2022.
40. Cho Y, Kim M, Choi M. Factors associated with nurses' user resistance to change of electronic health record systems. BMC Med Inform Decis Mak 2021 Jul 17;21(1):218. [doi: 10.1186/s12911-021-01581-z] [Medline: 34273990]
41. Kung LH, Yan YH, Kung CM. Exploring telemedicine usage intention using technology acceptance model and social capital theory. Healthcare (Basel) 2024 Jun 26;12(13):1267. [doi: 10.3390/healthcare12131267] [Medline: 38998802]
42. Diel S, Doctor E, Reith R, Buck C, Eymann T. Examining supporting and constraining factors of physicians' acceptance of telemedical online consultations: a survey study. BMC Health Serv Res 2023 Oct 19;23(1):1128. [doi: 10.1186/s12913-023-10032-6] [Medline: 37858170]

Abbreviations

GMC: geriatric medical center
PEU: ease of use
RTC: resistance to change
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
TAM: Technology Acceptance Model
VIF: variance inflation factor

Edited by A Kushniruk; submitted 15.Mar.2026; peer-reviewed by A Gamus, J Křivková, M Bureš, M Zwilling; revised version received 31.Jul.2026; accepted 26.Aug.2026; published 05.Oct.2026.

Please cite as:
Green G, Lochmannová A, Tesler R, Porat-Packer T
Telemedicine Adoption Among Health Care Professionals in Israeli Geriatric Medical Centers: Moderated Mediation Cross-Sectional Questionnaire Study
JMIR Hum Factors 2026;13:e95371
URL: https://humanfactors.jmir.org/2026/1/e95371
doi:10.2196/95371

© Gizell Green, Alena Lochmannová, Riki Tesler, Tammy Porat-Packer. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 5.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic

XSL•FO
RenderX

information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring Common and Novel Actualized Affordances of Fitbit: Mixed Methods Study

Moayad Alshawmar[1], PhD; Bengisu Tulu[2], PhD; E Vance Wilson[2], PhD; Adrienne Hall-Phillips[2], PhD; Ibrahim Aljadani[2]; Emmanuel Agu[2], PhD

[1]Imam Mohammad ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia

[2]Worcester Polytechnic Institute, Worcester, MA, United States

Corresponding Author:
Moayad Alshawmar, PhD
Imam Mohammad ibn Saud Islamic University (IMSIU)
Al Thoumamah Rd
Riyadh, 11564
Saudi Arabia
Phone: 966 2585529
Email: mnalshawmar@imamu.edu.sa

Abstract

Background: Although fitness apps could promote healthier lifestyles, evidence on the effectiveness of app-based interventions remains inconsistent. Previous studies have used affordance theory to identify the factors that generate exercise-related value for users. However, many fitness app affordance studies have examined multiple fitness apps collectively, assuming similar design intentions across platforms. Moreover, most have relied on predefined affordances rather than investigating emergent or novel ones that may reveal unique user–fitness app interactions.

Objective: This study aimed to identify the common affordances actualized by Fitbit users and uncover novel affordances that emerge from their interactions with this specific app, thereby extending the understanding of how affordances contribute to user engagement and health outcomes.

Methods: We used a 2-stage mixed methods design. First, a cross-sectional web-based survey was conducted with 442 US-based Fitbit users engaging in regular exercise. The participants selected affordances from a list identified in prior literature and could report additional affordances in open-text responses. To corroborate and extend the survey findings, 15,000 user reviews were collected from the Google Play Store, of which 2674 (17.8%) comments were automatically categorized into affordance themes and 1182 (7.9%) were manually validated as relevant. Reviews were thematically classified into affordance categories via a generative pretrained transformer–based approach guided by survey-identified affordances.

Results: The survey revealed that the most frequently actualized affordances were updating (351 participants and 749 review mentions; total=1100) and reminding (319 participants and 143 mentions; total=462), underscoring Fitbit's role in tracking progress and sustaining routines. Competing (99 participants and 88 mentions; total=187) and rewards (133 participants and 32 mentions; total=165) highlighted gamification, whereas comparing (151 participants and 8 mentions; total=159) and guidance (118 participants and 25 mentions; total=143) reflected benchmarking and instructional support. Other affordances such as searching (135 participants and 2 mentions; total=137), encouraging (75 participants and 19 mentions; total=94), and watching others (68 participants and 3 mentions; total=71) were less common, whereas recognizing (58 participants and 0 mentions; total=58) and self-presentation (47 participants and 1 mention; total=50) were the least common. The novel affordances included encouraging others (14 participants and 1 mention; total=15), accountability (3 participants and 9 mentions; total=12), and self-comparison (3 participants and 5 mentions; total=8).

Conclusions: Most Fitbit users actualized updating and reminding affordances, whereas a limited number of users actualized the other affordances. Moreover, few Fitbit users actualized novel affordances that reflect self-regulation, an extension of social connection, and personal meaning. This study emphasizes that Fitbit should focus on core tracking and reminding for most users while providing optional features that foster guidance, community, accountability, and personal relevance. Designing features that facilitate the actualization of common and novel affordances may improve app effectiveness and, ultimately, the health benefits of fitness technologies.

XSL•FO
RenderX

(JMIR Hum Factors 2026;13:e85412) doi:10.2196/85412

KEYWORDS

fitness apps; exercisers; affordance; affordance actualization; exercise behaviors

Introduction

Mobile fitness apps have emerged as widely adopted digital tools intended to support individuals in pursuing healthier lifestyles [1]. Through their interactive and data-driven nature, fitness apps encourage consistent exercise, potentially enhancing both physical fitness and cognitive health while contributing to the prevention of chronic illnesses [2-4].

Although these apps hold potential for improving exercise behaviors, evidence on the effectiveness of app-based interventions remains inconsistent [5]. Several studies report minimal to no observable improvements in exercise following app-based interventions [6,7], whereas others demonstrate meaningful gains in exercise engagement and outcomes among adults [8,9]. These mixed findings indicate that the effectiveness of fitness app interventions in promoting exercise is not yet definitively established, highlighting the need to better understand the pathways through which these apps contribute to exercise improvement.

To clarify these inconsistencies, recent research has shifted from solely examining intervention health-related outcomes to investigating the factors that generate exercise-related value for users [4,10-12]. Some of these studies interpret fitness app use through the lens of technology affordance actualization, emphasizing what the app enables users to do rather than the app feature list. For example, while some fitness apps offer group-based exercise features, certain users may actualize them as sources of exercise guidance, whereas others may use them primarily for comparing themselves to other participants. This view recognizes that users may use apps in ways that either follow or depart from the purposes intended by the designers [13,14]. Thus, studying the apps' affordances can help explain the core benefits of using apps from the users' perspective despite their design differences because affordances are driven by users' goals and how users interact with the app to achieve these goals.

Prior fitness app research has identified several affordances that users actualize while interacting with app features, such as self-monitoring, exercise guidance, competition, and social comparison [10,11,15-17]. However, many of these studies have analyzed multiple fitness apps collectively, with the assumption of shared design purposes across platforms. While such breadth may provide useful generalizations for some studies' purposes, it risks overlooking meaningful differences rooted in users' interaction with specific fitness apps that differ in design, even among apps with similar functions. For example, although Fitbit (Google LLC) and MyFitnessPal (MyFitnessPal, Inc) offer similar features (eg, activity tracking), we cannot assume that they are identical or that their users behave in the same way. What each app measures, how it tracks those measurements, and how the data are presented to users can differ substantially.

Furthermore, most prior fitness app research has focused on applying the literature-uncovered affordances rather than exploring new ones [15,16,18-20]. This approach may limit opportunities to discover new affordances, particularly given that affordance theory suggests that users actualize possibilities based on their diverse goals, needs, and intentions [13,14,21]. While applying literature-identified affordances can deepen our understanding of how these affordances contribute to exercise improvement, exploring new affordances can broaden research possibilities and provide developers with clearer insights into what their apps are used for.

This study aims to address these limitations by investigating both common (fitness app literature–predefined) affordances focusing their actualization on one fitness app (Fitbit) and novel (not yet uncovered) affordances actualized by users of the app. Specifically, we aimed to answer the following research questions: (1) which of the fitness app literature affordances are Fitbit app users actualizing? (2) What are the novel affordances actualized through users' interactions with Fitbit?

This study used a 2-stage methodology to uncover the actualized affordances of the Fitbit app. First, a survey was conducted to identify both common and novel affordances experienced by users. On the basis of the survey findings, user reviews from the Google Play Store were analyzed and thematically classified via a generative pretrained transformer (GPT)–based approach guided by the affordances identified in the survey.

Methods

Study Design

This study used a cross-sectional web-based survey, and its design and findings were reported in accordance with the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) guidelines [22]. The survey investigated the common and novel affordances actualized through user interaction with a widely used fitness app, Fitbit. To complement the survey, user reviews from the Google Play Store were analyzed. The affordances identified in the survey, both common and novel, were used as thematic categories to classify review comments through a GPT-based approach. By triangulating survey data with a large corpus of user reviews, we sought to establish robust evidence of the affordances actualized by Fitbit users.

Measures

Survey

To uncover the common affordances actualized by Fitbit users, participants were provided with a list of established fitness app affordances, as shown in Table 1, and asked to select all that applied to their use of the app. Each affordance was accompanied by a short definition adapted from prior studies [11,17]. The participants could select one or multiple affordances that Fitbit enabled during their use.

XSL•FO
RenderX

Table 1. Survey items used to uncover the actualization of common fitness app affordances in the context of Fitbit.

Affordance	Definition
What did you use Fitbit for? Choose all that apply.	
Comparing	To compare my exercise performance with others
Guidance	To get guidance how to perform my exercise
Self-presentation	To present myself as a physically active person
Rewards	To receive rewards for my exercise
Recognizing	To get recognition from others for my exercise
Encouraging	To have other people encourage my exercise
Competing	To compete with others
Watching others	To keep an eye on others' ways of doing exercise
Reminding	To remind me to do an exercise activity
Updating	To update me with the status of my exercise progress
Searching	To search for exercise information

To capture novel affordances, the survey included an "other" option that allowed participants to write in additional action possibilities not included in the predefined list. These written responses were analyzed to determine whether they represented true affordances, that is, goal-directed action possibilities emerging from user-technology interaction rather than simple descriptions of Fitbit features. In doing so, we applied the 3 affordance identification principles described by Alshawmar et al [23], which are grounded in established affordance theory [13,21,24]. First, affordances must reflect a user-perceived action possibility. If a response did not contain a user goal or intended action, it was treated as a feature rather than an affordance. Second, affordances must be described in terms of *why* the user engages with the technology, not *what* the technology provides. For example, the ability to share achievements is a feature; "receiving encouragement from others" is an affordance because it reflects a purpose or goal behind using that feature. Third, multiple features may enable the same affordance. Therefore, responses were grouped at the level of user goal rather than specific app components.

All written responses were independently reviewed by 2 coders trained in affordance theory. Coding disagreements were discussed and resolved through consensus. Interrater agreement before reconciliation was 0.92, indicating high reliability. Responses that reflected affordances already included in the predefined list were classified as common and removed from this step. The remaining responses were examined for thematic similarity, which resulted in the identification of 3 novel affordances. Each novel affordance was then defined based on prior literature to ensure conceptual clarity. For example, written responses such as "to hold myself accountable" were classified as the novel affordance accountability, consistent with the definition of accountability by Dhiman et al [25] as the need to justify one's actions to oneself.

In addition, demographic information and use data (eg, age, gender, duration and frequency of use, and premium subscription) were collected to provide context for interpreting users' affordance actualization patterns.

User Review Data

To support the survey findings of common and novel affordances, we analyzed Fitbit app users' review comments. User reviews were collected from the Google Play Store via an automated web scraping tool [26]. Approximately 380,000 reviews were initially gathered spanning multiple app versions and time ranges. After preprocessing and filtering, we removed duplicates, non–English-language content, very short reviews with fewer than 6 words, and reviews retained from the past 3 years that contained the word "to" as a simple heuristic for goal-directed app use (eg, "to track progress" and "to stay updated"). Standard text-cleaning steps such as lower casing and punctuation removal were applied to prepare the data for analysis. This data preprocessing resulted in a final sample of 15,000 reviews.

For thematic classification, we developed a model based on 14 predefined affordance-based themes identified from prior literature and the survey findings. Each review was analyzed using GPT-4.1 mini (OpenAI) via an instruction-based, zero-shot, label-constrained prompt designed for thematic classification (refer to Multimedia Appendix 1 for more details about the prompt).

The prompt introduced the model to the predefined affordance categories, each clearly defined and exemplified (eg, "Comparing; comparing performance with others" and "Guidance; receiving instructions on how to exercise"). The GPT was instructed to select exactly 1 theme per review when applicable or to generate a single-word new (nonaffordance) theme if the review did not fit any predefined category. The large language model's temperature was set to 0.3 to minimize variability and ensure stable classifications across reviews [27,28].

This resulted in a dataset consisting of 12,378 reviews categorized as "nonaffordance" and 2622 reviews assigned to the predefined affordance themes.

This design encouraged the model to apply conceptual reasoning rather than keyword matching, aligning the classification with

XSL•FO
RenderX

the theoretical framework of affordance actualization (refer to the example classifications in Multimedia Appendix 2).

To ensure classification accuracy, human validation of the GPT-based thematic assignments was conducted. After the GPT classified the full dataset, a stratified random sample of 300 reviews was drawn. One author manually coded the validation sample, applying the same thematic definitions provided to the language model. To assess agreement, the author-coded labels were then compared to the GPT-generated labels. Of the 300 double-coded reviews, 255 (85%) received identical theme assignments.

Finally, the 2622 Fitbit review comments categorized into the 14 affordances were reviewed by 1 author. This resulted in 1182 comments being retained for relevance.

Participants and Recruitment Procedures

Before launching the main study, we carried out a pilot survey between June 1, 2022, and July 5, 2022. The participants were recruited through a university Sona Systems participant pool and via social media platforms (Facebook [Meta Platforms Inc] and Reddit [Reddit Inc]). In total, 403 individuals responded. After incomplete submissions, responses that failed validation checks, and cases showing inattentive patterns (such as straight-lining or unusually short completion times) were excluded, 81 usable responses remained. These participants were aged 18 to 52 years, with 318 (78.9%) participants being aged ≥25 years. Insights from this pilot, including feedback on the chosen affordances and the comments written in the "other" category, provided evidence supporting the adequacy of the survey instrument.

For the main study, we determined the required sample size via the formula by Gaskin [29]. Because social media recruitment produces a high rate of invalid responses, we opted to use Prolific (Prolific Academic Ltd), a crowdsourcing platform recognized for providing higher-quality data [30]. A prescreening survey was first distributed to 1300 participants, of whom 622 (47.8%) current adult US-based Fitbit users who engaged in exercise were identified. Of these 622 users, 506 (81.4%) completed the main survey. After applying quality control measures, including attention checks, of the 622 users, 442 (71.1%) valid responses were retained for analysis.

Ethical Considerations

Both the pilot and main studies received approval from the institutional review board (IRB-22-0351) of the Worcester Polytechnic Institute and were conducted in accordance with ethical guidelines for human subject research. The participants provided informed consent electronically at the beginning of each survey, where they were informed about the study's purpose, the voluntary nature of participation, and their right to withdraw at any time. The consent form also noted that the data would be stored securely and accessible only by the research team. All the responses were collected anonymously.

For the pilot study, student participants recruited through the university's local participant pool received 0.5 course credit as an incentive, whereas participants recruited through social media received an opportunity to enter a raffle for 1 of 10 US $10 prizes as an incentive. The distribution of incentives was managed by one of the authors. For the main study, all recruitment and distribution of incentives were managed by Prolific. The participants recruited for the prescreening survey received a flat rate of US $0.18 as an incentive. Among these participants, those who were eligible to take part in the actual study (current Fitbit users who engaged in exercise) received a flat rate of US $6 as an additional incentive.

Results

Descriptive Statistics

The main study sample demographics are shown in Table 2.

All participants were current Fitbit users with varying use patterns, as shown in Table 3.

XSL•FO
RenderX

Table 2. Demographics of the main study sample based on an online survey of US-based Fitbit users who engaged in regular exercise (N=442).

	Participants, n (%)
Gender	
Man	221 (50)
Woman	212 (47.9)
Transgender	4 (0.9)
Nonbinary	5 (1.1)
Age (y)	
18-24	31 (7)
25-34	146 (33)
35-44	137 (31)
45-54	57 (12.9)
55-64	48 (10.9)
65-74	18 (0.1)
≥75	5 (1.1)
Race	
Asian	35 (7.9)
Black	40 (9)
White	358 (80.9)
Other	9 (2.1)
Employment status	
Employed full time	309 (69.9)
Employed part time	62 (14)
Not in paid work (eg, homemaker, retired, or disabled)	40 (9)
Unemployed (and job seeking)	18 (4.1)
Other	13 (2.9)

XSL•FO
RenderX

Table 3. Use patterns of the main study sample based on an online survey of US-based Fitbit users who engaged in regular exercise (N=442).

	Participants, n (%)
Frequency of Fitbit app use	
Daily	194 (43.9)
4-6 times a week	137 (31)
2-3 times a week	93 (21)
Once a week	9 (2)
Less than once a week	9 (2)
Duration of Fitbit app use	
<1 month	0
Between 1 week and 1 month	0
1-3 months	18 (4.1)
4-6 months	40 (9)
7-9 months	22 (5)
10-12 months	35 (7.9)
>12 months	327 (74)
Duration of premium use	
Never	88 (19.9)
Up to 3 months or free trial	110 (24.9)
4-6 months	44 (10)
7-12 months	58 (13.1)
>12 months	142 (32.1)
Device used	
Phone	177 (40)
Fitbit wearable device	181 (41)
Smartwatch (such as Samsung Galaxy Watch and Apple Watch)	49 (11.1)
Laptop or desktop computer	35 (7.9)
Exercise engaged in	
Running	106 (24)
Walking	274 (62)
Biking	22 (5)
Swimming	5 (1.1)
Spinning	4 (9)
Hiking	13 (2.9)
Yoga	5 (1.1)
Other	13 (2.9)

Common Fitbit Actualized Affordances

Common affordances discovered in the fitness app literature, as presented with definitions in Table 4, were actualized by the participants in this study via Fitbit. However, as per the arguments in this study, users' actualizations may vary from one fitness app to another. When Fitbit was used, as shown in Table 5, the most frequently actualized affordances among users were updating (n=351, 79.4%; 351 participants and 749 review mentions; total=1100) and reminding (n=319, 72.2%; 319 participants and 143 review mentions; total=462). These findings indicate that users rely primarily on Fitbit to maintain awareness of their activity status and receive prompts that help sustain exercise routines. Other affordances also demonstrated meaningful engagement: competing (99, 22.4%; 99 participants and 88 review mentions; total=187) and rewards (n=133, 30.1%; 133 participants and 32 review mentions; total=165) highlight the motivational role of competition and incentives, whereas comparing (n=151, 34.16%; 151 participants and 8 review mentions; total=159) and guidance (n=118, 26.7%; 118

XSL•FO
RenderX

participants and 25 review mentions; total=143) emphasize the app's instructional and benchmarking functions.

Additional affordances were less prevalent but still notable: searching (n=135, 30.5%; 135 participants and 2 review mentions; total=137), encouraging (n=75, 16.9%; 75 participants and 19 review mentions; total=94), and watching others (n=68, 15.4%; 68 participants and 3 review mentions; total=71). Finally, recognizing (n=58, 13.1%; 58 participants and 1 review mention; total=59) and self-presentation (n=47, 10.6%; 47 participants and 4 review mentions; total=51) were the least actualized, reflecting more specialized uses such as acknowledgment and image projection. Overall, the results show that Fitbit affordances extend beyond core tracking functions to include social, motivational, and comparative features, although these were less consistently used across the user base.

Table 4. Definitions of common affordances based on the fitness app literature.

Common affordance	Definition
Comparing	Allows users to evaluate their performance relative to that of others [11,17,31]
Guidance	Offers structured instructions or exercise recommendations that help users perform activities effectively [17,31]
Self-presentation	Allows users to publicly share achievements or statuses, facilitating impression management and self-image [17]
Rewards	Provides recognition in the form of badges, points, or virtual incentives, reinforcing desired behaviors [11,17]
Recognizing	Facilitates acknowledgment or praise from others for achievements, enhancing validation and social reinforcement [11]
Encouraging	Permits users to receive motivational messages from others, facilitating social support [11]
Competing	Enables users to engage in challenges or contests, fostering motivation through social rivalry [11]
Watching others	Allows users to observe others' activities or progress [17]
Reminding	Provides timely prompts or reminders for workouts and goals, supporting consistency and habit formation [11]
Updating	Enables continuous self-tracking and progress awareness [11]
Searching	Provides tools to look up workouts, information, or content, enabling discovery and information retrieval [11]

Table 5. Common affordance actualizations uncovered through the online survey and supported by Fitbit app reviews.

Affordance	Actualization via survey (N=442), n (%)	Actualization via Fitbit reviews—relevant comments (n=2648 Fitbit review comments in total), n/N (%)	Total actualizations (survey+reviews)
Updating	351 (79.4)	749/1124 (66.6)	1100
Reminding	319 (72.2)	143/245 (58.4)	462
Competing	99 (22.4)	88/129 (68.2)	187
Rewards	133 (30.1)	32/55 (58.2)	165
Comparing	151 (34.2)	8/27 (29.6)	159
Guidance	118 (26.7)	25/143 (17.5)	143
Searching	135 (30.5)	2/18 (11.1)	137
Encouraging	75 (17)	19/475 (4)	94
Watching others	68 (15.4)	3/3 (100)	71
Recognizing	58 (13.1)	0/1 (0)	58
Self-presentation	47 (10.6)	1/25 (4)	51

Novel Actualized Affordances

The participants also selected an option in the survey to write down different affordances that they had actualized via Fitbit. This resulted in 158 written responses (affordances). Some of the affordances written were similar to those listed in the survey but were written by participants with more use specifications. For example, while one of the choices for using Fitbit in the survey was "to update me with the status of my exercise," some participants selected "others" to write "to see my progress for the day." We excluded responses related to common affordances (n=129, 81.6%) and were left with 29 responses for further analysis.

We analyzed these 29 responses based on the 3 affordance principles we followed [23]. This resulted in the identification of 3 novel affordances, as defined in Table 6 and shown with user comments and quotes from the survey in Table 7. For example, as a purpose of using Fitbit, some participants wrote "to hold myself accountable." This finding indicated that, using Fitbit, users built a sense of accountability toward exercising. In other words, Fitbit affords users the feeling of being

XSL•FO
RenderX

accountable as a motivation to accomplish their regular exercise. These 3 affordances were used as themes to be uncovered within users' comments on Fitbit.

As shown in Table 8, novel affordances included encouraging others (n=14, 3.2%; 14 participants and 1 review mention), accountability (n=3, 0.7%; 3 participants and 9 review mentions), and self-comparison (n=3, 0.7%; 3 participants and 5 review mentions).

Table 6. Definitions of novel affordances.

Affordance	Definition
Accountability	A sense of accountability for pursuing their exercise using various features such as today's activity, cheer, and friends
Self-comparing	Assessing oneself against specific criteria or standards to gain insights into one's strengths, weaknesses, areas for improvement, and overall effectiveness
Encouraging others (the opposite of receiving encouragement)	Offering support, motivation, or positive feedback to someone to boost their confidence, inspire action, or reinforce desirable behavior

Table 7. Participant and user quotes reflecting the novel affordances.

Affordance	Actualization via survey (N=442)	Actualization via Fitbit reviews (n=2648 Fitbit review comments)
Accountability	• "To hold myself accountable" • "To hold each other accountable" • "To hold myself accountable to good exercise"	• "Very easy way to check my progress and hold myself accountable to my fitness goals." • "Let us me know to move, check my blood sugar & monitors my progress. Makes me accountable!"
Self-comparing	• "To compare myself with my past performance" • "I like to see my overall trend for the week" • "Love to compare what I do week to week"	• "My favorite app! I have 10 years of step data and can look back to see how various activities have increased my steps over the years. All the data is still there through several Fitbit trackers and 3 phones." • "Do not realize how much walking I'm doing. However, it great to see I'm no longer a couch potato."
Encouraging others	• "To encourage my friends when competing" • "To encourage others' exercise" • "To give other people motivation"	• "Love the Fitbit app and all it has to offer helps me stay on track and encourage others to challenges."

Table 8. Novel affordance actualizations uncovered through the online survey and supported via Fitbit app reviews.

Affordance	Actualization via survey (N=442), n (%)	Actualization via Fitbit reviews—relevant comments (n=2648 Fitbit review comments), n/N (%)	Total actualizations (survey+reviews)
Accountability	3 (0.7)	9/112 (8)	12
Self-comparing	3 (0.7)	5/81 (6.2)	8
Encouraging others	14 (3.2)	1/1 (100)	15

Discussion

Principal Findings

Although prior studies have identified several frequently actualized affordances in fitness apps, much of this work examines multiple platforms collectively and relies largely on affordances already established in the literature. This broad approach obscures meaningful app-specific differences and constrains the discovery of novel affordances that may emerge from users' diverse goals and interaction patterns. As a result, existing research provides an incomplete account of how fitness apps generate value in real use contexts.

This study explored how Fitbit users actualized both common and novel affordances for the sake of deeper understanding of fitness app true value. The study integrated insights from survey responses and app store reviews and used established guidelines to uncover the affordances. The findings reaffirm that certain affordances, particularly updating and reminding, are central to users' engagement. These affordances align closely with prior research emphasizing self-monitoring and feedback mechanisms as key drivers of continued use in fitness app contexts [32-35]. By enabling users to track their current activity and receive prompts to sustain routines, Fitbit fulfills core behavioral support functions that encourage habit formation and long-term adherence [10,11,15-17].

While updating and reminding dominated the value that users derived from Fitbit, other affordances played meaningful roles in shaping some of the users' engagement. Competing and rewards, for instance, highlighted the motivational force of gamification, consistent with previous work demonstrating the

XSL•FO
RenderX

appeal of competitive challenges and incentive-based reinforcement in fitness technologies [36]. Similarly, comparisons and guidance emphasize the importance of reference points and instructional support, showing that users draw value not only from tracking but also from situating their performance against that of others and learning how to optimize their efforts [37]. These findings resonate with affordance actualization theory, which stresses that technologies are appropriated in ways that reflect both design intent and users' personal goals.

Although less frequent, affordances such as encouraging, watching others, recognizing, and self-presentation revealed meaningful but more specialized social and motivational dimensions of use. Encouragement and recognition, in particular, pointed to the social support embedded in fitness apps even if these affordances are not commonly actualized. Self-presentation emerged as the least actualized affordance, suggesting that most Fitbit users may not use the app primarily for impression management. These findings indicate that not all the possible affordances will be actualized by all fitness apps users. As this study showed, some affordances were barely actualized by users. These findings may contrast with those of prior work that documents high user reliance on socially oriented fitness app affordances, including competition and peer encouragement [10,18], suggesting that affordance actualization mechanisms are not uniform across fitness technologies. This could be due to their type of users or the way in which these fitness apps are designed. Consequently, the study findings suggest that affordance actualization should be examined at the level of the individual technology as even similar technologies (eg, fitness apps) demonstrate distinct actualization patterns shaped by their design configurations and user populations.

Most fitness app studies have applied common affordances [15,16,18-20], but this study revealed several novel affordances contributing to the niche literature [10,11,17,31]. The *emergent* affordances were accountability, self-comparison, and encouraging others. These affordances represent additional ways in which a subset of users interact with Fitbit. For example, accountability reflects users' desire to justify actions to themselves, aligning with broader theories of self-regulation [25]. Previous studies have discovered that comparing oneself to others is an affordance of fitness apps [11,17]. Self-comparison extends the comparison literature by shifting the frame of reference from others to one's past self, highlighting reflective, longitudinal self-evaluation [38]. These affordances illustrate the nuanced ways in which users may reinterpret features to satisfy individualized goals.

It is important to interpret the novel affordances with caution. These emergent themes were reported by a small subset of participants and, therefore, should not be considered dominant affordances of the Fitbit app. Rather, they represent preliminary evidence of user-technology relations that extend beyond commonly applied, predefined affordance taxonomies. Although low in frequency, such responses are theoretically meaningful, as affordance actualization theory acknowledges that users may perceive and enact unintended or highly individualized action possibilities depending on personal goals and situational contexts [13,21]. These emergent patterns, while not widespread, highlight opportunities for future research to validate and elaborate on these affordances using larger samples and person-centered analytical approaches.

From a design perspective, the identification of these novel affordances suggests actionable avenues for developers to support users who seek stronger behavioral reinforcement and self-referential goal alignment. For example, to better enable the accountability affordance, fitness platforms could implement a commitment contract module (eg, structured pledges such as "I commit to..." with self-defined consequences) and provide an accountability dashboard that transparently tracks completed vs missed exercise sessions. To facilitate the self-comparison affordance, apps may incorporate temporal self-evaluation features such as "Past me vs now" comparison cards, digital snapshot summaries, and personal improvement challenges that benchmark current performance against a user's historical averages (eg, past–7-, 30-, or 365-day activity baselines). Finally, to support the affordance of encouraging others, platforms can introduce encouragement-based interactive elements, including peer cheer buttons, shareable motivation templates, and encouragement badges or kudos that users can award to friends or groups to reinforce social support behaviors.

Taken together, these design examples demonstrate how specific features can be engineered to strengthen the perception and actualization of emerging affordances, particularly for users who rely on digital tools not just for tracking activity but also for sustaining routines, reflecting on personal progress, and motivating others. This reinforces the importance of examining affordances within the context of a single technology ecosystem and tailoring future affordance research to behavioral patterns, motivational drivers, and downstream outcomes that reflect real user goals rather than only surface-level feature descriptions.

Limitations and Future Research

The first limitation of this study relates to the sample composition and self-report nature of the data. Although the dataset included a robust number of respondents (N=442), the sample consists exclusively of current US-based Fitbit users, limiting generalizability to broader global populations or non-Fitbit fitness app users. Most participants were White (358/442, 80.9%), employed (371/442, 83.9%), and aged 25 to 44 years (283/442, 64%), which may underrepresent the perspectives of older adults, unemployed individuals, or ethnically diverse groups, segments that may perceive or actualize affordances differently. Additionally, while 74% (327/442) reported more than 12 months of app use and 43.9% (194/442) used Fitbit daily, exercise behavior regularity was not directly measured beyond self-identification, meaning that the group-based causal inferences about exercise consistency or app-enabled intervention needs remain interpretive rather than behavioral. Finally, only 41% (181/442) used a Fitbit wearable device, and 62% (274/442) tracked walking activities, suggesting that affordance experiences may reflect tracking-light or single-activity use patterns, potentially overlooking affordances that emerge in more intensive, multiactivity, or device-dependent exercise contexts. Therefore, future studies should replicate the model with diverse, international, and

XSL•FO
RenderX

behavior-verified samples and explore predictors and outcomes of affordance actualization to deepen explanatory insights.

The second limitation is that review comments were analyzed primarily to corroborate the survey findings rather than identify additional affordances. Future research could extend this approach by systematically examining other Fitbit affordances within user reviews, thereby offering a more comprehensive understanding of the range of affordances that a technology such as Fitbit offers to users.

The third limitation is that, although demographic and use characteristics were collected, several variables exhibited substantial distributional imbalance (eg, duration of use and exercise type), limiting the feasibility of fine-grained subgroup analyses. Future research with more balanced and behavior-verified samples is needed to examine the heterogeneity of Fitbit affordance actualizations in greater depth.

Conclusions

This study shows that Fitbit users actualize a range of common and novel affordances, highlighting how individuals engage with fitness technology in diverse and app-specific ways. By identifying common actualized and emerging affordances, the findings reinforce the importance of examining affordance actualization within a single technological context. Overall, this study advances understanding of how fitness apps create value for users and offers direction for designing features that better support sustained engagement and healthier behavior.

Acknowledgments

Artificial intelligence tools were used exclusively to improve the English language and enhance readability.

Funding

This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (grant IMSIU-DDRSP2601).

Data Availability

The datasets generated or analyzed during this study are not publicly available due to institutional review board but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Base prompt used for generative pretrained transformer–based thematic classification.
[DOCX File , 13 KB - humanfactors_v13i1e85412_app1.docx]

Multimedia Appendix 2
Sample review inputs and generative pretrained transformer outputs.
[DOCX File , 15 KB - humanfactors_v13i1e85412_app2.docx]

References

1. Coughlin S, Whitehead M, Sheats JQ, Mastromonico J, Smith S. A review of smartphone applications for promoting physical activity. Jacobs J Community Med 2016;2(1):021 [FREE Full text] [Medline: 27034992]
2. Physical activity. World Health Organization. 2024 Jun 26. URL: https://www.who.int/news-room/fact-sheets/detail/physical-activity [accessed 2026-02-01]
3. Sousa Basto P, Ferreira P. Mobile applications, physical activity, and health promotion. BMC Health Serv Res 2025 Mar 10;25(1):359 [FREE Full text] [doi: 10.1186/s12913-025-12489-z] [Medline: 40065339]
4. Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. JMIR Mhealth Uhealth 2019 Aug 29;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]
5. He Z, Hassan MA, Saiz-González P, Ryu S, Wang R, Gao Z. Smartphone app-based interventions on physical activity behaviors and psychological correlates in healthy young adults: a systematic review. PLoS One 2024;19(4):e0301088 [FREE Full text] [doi: 10.1371/journal.pone.0301088] [Medline: 38578729]
6. Sultoni K, Peralta L, Cotton W. Technology-supported physical activity course for increasing physical activity levels of university students: system development and pilot implementation. Int J Hum Mov Sports Sci 2023 Feb;11(1):44-57. [doi: 10.13189/saj.2023.110106]
7. Simons D, De Bourdeaudhuij I, Clarys P, De Cocker K, Vandelanotte C, Deforche B. Effect and process evaluation of a smartphone app to promote an active lifestyle in lower educated working young adults: cluster randomized controlled trial. JMIR Mhealth Uhealth 2018 Aug 24;6(8):e10003 [FREE Full text] [doi: 10.2196/10003] [Medline: 30143477]

XSL•FO
RenderX

8. Al-Nawaiseh HK, McIntosh WA, McKyer LJ. An-m-health intervention using smartphone app to improve physical activity in college students: a randomized controlled trial. Int J Environ Res Public Health 2022 Jun 13;19(12):7228 [FREE Full text] [doi: 10.3390/ijerph19127228] [Medline: 35742477]
9. Gabbiadini A, Greitemeyer T. Fitness mobile apps positively affect attitudes, perceived behavioral control and physical activities. J Sports Med Phys Fitness 2019 Mar;59(3):407-414. [doi: 10.23736/S0022-4707.18.08260-9] [Medline: 29619794]
10. James TL, Deane JK, Wallace L. An application of goal content theory to examine how desired exercise outcomes impact fitness technology feature set selection. Inf Syst J 2019 Jan 04;29(5):1010-1039. [doi: 10.1111/isj.12233]
11. James TL, Wallace L, Deane JK. Using organismic integration theory to explore the associations between users' exercise motivations and fitness technology feature set use. MIS Q 2019;43(1):287-312 [FREE Full text] [doi: 10.25300/MISQ/2019/14128]
12. Rockmann R, Gewald H. Individual fitness app use: the role of goal orientations and motivational affordances. In: Proceedings of the 25th Americas Conference on Information Systems. 2019 Presented at: AMCIS 2019; August 15-17, 2019; Cancún, Mexico URL: https://www.researchgate.net/publication/332726798_Individual_Fitness_App_Use_The_Role_of_Goal_Orientations_and_Motivational_Affordances
13. Markus ML, Silver MS. A foundation for the study of IT effects: a new look at desanctis and poole's concepts of structural features and spirit. J Assoc Inf Syst 2008 Oct;9(10):609-632 [FREE Full text] [doi: 10.17705/1jais.00176]
14. Strong DM, Volkoff O, Johnson SA, Pelletier LR, Tulu B, Bar-On I, et al. A theory of organization-EHR affordance actualization. J Assoc Inf Syst 2014;15(2):53-85. [doi: 10.17705/1jais.00353]
15. Alshawmar M. A review of the applications of affordance theory in mHealth app research. In: Proceedings of the Hawaii International Conference on System Sciences 2021. 2021 Presented at: HICSS-54; January 5, 2021; Kauai, HI. [doi: 10.24251/hicss.2021.437]
16. Suh A. Sustaining the use of quantified-self technology: a theoretical extension and empirical test. Asia Pac J Inf Syst 2018 Jun;28(2):114-132 [FREE Full text] [doi: 10.14329/apjis.2018.28.2.114]
17. Rockmann R, Gewald H. Activity tracking affordances: identification and instrument development. In: Proceedings of the PACIS 2018. 2018 Presented at: PACIS 2018; June 26-30, 2018; Yokohama, Japan URL: https://aisel.aisnet.org/pacis2018/232/
18. Chen X, Zhu Y, Xu X. Being there and being with them: the effects of visibility affordance of online short fitness video on users' intention to cloud fitness. Front Psychol 2024;15:1267502 [FREE Full text] [doi: 10.3389/fpsyg.2024.1267502] [Medline: 38362244]
19. Zhou X, Krishnan A, Dincelli E. Examining user engagement and use of fitness tracking technology through the lens of technology affordances. Behav Inf Technol 2021 Apr 17;41(9):2018-2033. [doi: 10.1080/0144929x.2021.1915383]
20. Suh A, Li M. How the use of mobile fitness technology influences older adults' physical and psychological well-being. Comput Hum Behav 2022 Jun;131:107205. [doi: 10.1016/j.chb.2022.107205]
21. Leonardi PM. When flexible routines meet flexible technologies: affordance, constraint, and the imbrication of human and material agencies. MIS Q 2011 May;35(1):147-167 [FREE Full text] [doi: 10.2307/23043493]
22. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
23. Alshawmar M, Tulu B, Wilson V, Hall-Phillips A. Clarifying the relationship between fitness apps' affordances and features. In: Proceedings of the 28th Americas Conference on Information Systems. 2022 Presented at: AMCIS 2022; August 10-14, 2022; Minneapolis, MN URL: https://www.researchgate.net/publication/362212690_Clarifying_the_Relationship_Between_Fitness_Apps'_Affordances_and_Features
24. Strong D. Affordance theory and how to use it in IS research. In: Galliers RD, Stein MK, editors. The Routledge Companion to Management Information Systems. London, UK: Routledge; 2017:164-178.
25. Dhiman A, Sen A, Bhardwaj P. Effect of self-accountability on self-regulatory behaviour: a quasi-experiment. J Bus Ethics 2015 Dec 24;148:79-97. [doi: 10.1007/s10551-015-2995-4]
26. Google-play-scraper 1.2.7. Python Software Foundation. URL: https://pypi.org/project/google-play-scraper/ [accessed 2026-02-01]
27. Agarwal A, Mittal K, Doyle A, Sridhar P, Wan Z, Doughty JA, et al. Understanding the role of temperature in diverse question generation by GPT-4. In: Proceedings of the 55th ACM Technical Symposium on Computer Science Education. 2024 Presented at: SIGCSE 2024; March 20-23, 2024; Portland, OR URL: https://dl.acm.org/doi/10.1145/3626253.3635608 [doi: 10.1145/3626253.3635608]
28. Renze M. The effect of sampling temperature on problem solving in large language models. In: Proceedings of the Findings of the Association for Computational Linguistics. 2024 Presented at: EMNLP 2024; Nov 12-16, 2024; Miami, FL URL: https://aclanthology.org/2024.findings-emnlp.432/ [doi: 10.18653/v1/2024.findings-emnlp.432]
29. Causal models. StatWiki. URL: https://statwiki.gaskination.com/index.php/Causal_Models [accessed 2026-02-01]
30. Peer E, Brandimarte L, Samat S, Acquisti A. Beyond the Turk: alternative platforms for crowdsourcing behavioral research. J Exper Soc Psychol 2017 May;70:153-163. [doi: 10.1016/j.jesp.2017.01.006]

XSL•FO
RenderX

31. Alshawmar M, Mombini H, Vaghefi I, Tulu B. Investigating the affordances of wellness mHealth apps. In: Proceedings of the Hawaii International Conference on System Sciences. 2021 Presented at: HICSS 2021; January 5, 2021; Kauai, HI. [doi: 10.24251/hicss.2021.462]
32. Yang Y, Koenigstorfer J. Determinants of fitness app usage and moderating impacts of education-, motivation-, and gamification-related app features on physical activity intentions: cross-sectional survey study. J Med Internet Res 2021 Jul 13;23(7):e26063 [FREE Full text] [doi: 10.2196/26063] [Medline: 34255656]
33. Henriksen A, Johannessen E, Hartvigsen G, Grimsgaard S, Hopstock LA. Consumer-based activity trackers as a tool for physical activity monitoring in epidemiological studies during the COVID-19 pandemic: development and usability study. JMIR Public Health Surveill 2021 Apr 23;7(4):e23806 [FREE Full text] [doi: 10.2196/23806] [Medline: 33843598]
34. Antoun J, Itani H, Alarab N, Elsehmawy A. The effectiveness of combining nonmobile interventions with the use of smartphone apps with various features for weight loss: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022 Apr 08;10(4):e35479 [FREE Full text] [doi: 10.2196/35479] [Medline: 35394443]
35. Bo Y, Liu QB, Tong Y. The effects of adopting mobile health and fitness apps on hospital visits: quasi-experimental study. J Med Internet Res 2023 Jul 28;25:e45681 [FREE Full text] [doi: 10.2196/45681] [Medline: 37505809]
36. Mazeas A, Duclos M, Pereira B, Chalabaev A. Evaluating the effectiveness of gamification on physical activity: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Jan 04;24(1):e26779 [FREE Full text] [doi: 10.2196/26779] [Medline: 34982715]
37. Arigo D, Brown MM, Pasko K, Suls J. Social comparison features in physical activity promotion apps: scoping meta-review. J Med Internet Res 2020 Mar 27;22(3):e15642 [FREE Full text] [doi: 10.2196/15642] [Medline: 32217499]
38. Mabe PA, West SG. Validity of self-evaluation of ability: a review and meta-analysis. J Appl Psychol 1982;67(3):280-296. [doi: 10.1037//0021-9010.67.3.280]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
GPT: generative pretrained transformer

Edited by J Sarvestan; submitted 07.Oct.2025; peer-reviewed by Y Lu, R Monachelli; comments to author 20.Nov.2025; revised version received 26.Jan.2026; accepted 28.Jan.2026; published 18.Feb.2026.

Please cite as:
Alshawmar M, Tulu B, Wilson EV, Hall-Phillips A, Aljadani I, Agu E
Exploring Common and Novel Actualized Affordances of Fitbit: Mixed Methods Study
JMIR Hum Factors 2026;13:e85412
URL: https://humanfactors.jmir.org/2026/1/e85412
doi:10.2196/85412
PMID:

©Moayad Alshawmar, Bengisu Tulu, E Vance Wilson, Adrienne Hall-Phillips, Ibrahim Aljadani, Emmanuel Agu. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 18.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Designing for Amplifying Voices—Supporting Bangladesh's Low-Income Ready-Made Garment Workers Through Technology Amid COVID-19 Pandemic: Interview Study Among Workers

Rahat Jahangir Rony[1], PhD; Nova Ahmed[2], PhD

[1]School of Computer Science and Informatics, Cardiff University, Abacws, Senghennydd Road, Cardiff, Wales, United Kingdom

[2]Department of Electrical and Computer Engineering, Design Inclusion and Access Lab (DIAL), North South University, Dhaka, Bangladesh

Corresponding Author:
Rahat Jahangir Rony, PhD
School of Computer Science and Informatics, Cardiff University, Abacws, Senghennydd Road, Cardiff, Wales, United Kingdom

Abstract

Background: The ready-made garments (RMG) industry is a crucial component of Bangladesh's economy, using over 4 million workers from low-income backgrounds who often neglect their health care needs. Historically, this sector has faced criticism for labor exploitation, unsafe working conditions, and rights violations, as highlighted by tragic accidents resulting in loss of life. Compliant factories may uphold higher labor standards, but many noncompliant factories expose workers to poor conditions, increasing health risks. The COVID-19 pandemic intensified vulnerabilities, leading to widespread factory closures and job losses, increased health risks, and left millions of workers without wages. Although the government attempted to provide some relief, it fell short in offering job security, social protection, health services, and emergency assistance. Limited access to technology due to digital literacy gaps further hinders these workers, who primarily use basic mobile phones for communication rather than accessing health or emergency services. Thus, there is a pressing need to develop a sustainable system that capitalizes on their existing technological familiarity.

Objective: Our goal was to gain a deep understanding of RMG workers' experiences, focusing on their work environments, technological interactions related to health care management, and the impacts of COVID-19 on their circumstances. This understanding aims to inform the design of a technology-based framework that is both sustainable and contextual.

Methods: In phase 1, we conducted in-person interviews with 55 RMG workers, comprising 32 female and 23 male participants from urban and suburban areas of Dhaka and Gazipur, before the pandemic. Participants were aged between 18 and 40 years. In phase 2, we reconnect with 12 phase 1 participants during the pandemic and also consulted 3 stakeholders from RMG factories via phone. Each interview, conducted in Bengali, was recorded with consent, totaling 846 minutes of discussion that were translated and transcribed. Thematic analysis was used to analyze the results.

Results: Insights gathered revealed variations in working conditions, personal experiences, perceptions of health care, lifestyle choices, and technology use tied to factory compliance. Workers at compliant factories had better health care support and used technology more effectively than those in noncompliant settings. The pandemic drastically altered the landscape for all workers, heightening health and safety concerns and shortages of emergency assistance. The RMG sector faces significant challenges, highlighting the urgent need for targeted emergency relief and health services.

Conclusions: This research examined the challenges and technology use among RMG workers during the pandemic, with a focus on health care perspectives. Based on our findings, we proposed a technology-based framework called VOICE (Virtual Outlet for Integrated Community Engagement) that connects workers with service providers through a simplified interface. This aims to assist marginalized communities during emergencies and enhance their overall well-being.

(JMIR Hum Factors 2026;13:e90675) doi:10.2196/90675

KEYWORDS

ready-made garment workers; RMG; COVID-19 pandemic; well-being; technology use; health care; Bangladesh; occupational health; digital literacy; health management; technology-based; framework

XSL•FO
RenderX

Introduction

Background

The ready-made garment (RMG) industry is a crucial sector of Bangladesh's economy, contributing over 10% to the national gross domestic product and 84% of total export earnings [1-3]. The sector employs over 4 million workers, with nearly 90% being female workers from low-income and rural backgrounds [4,5]. This aligns with Lewis's economic growth model, characterized by an unlimited labor supply driven by low labor costs [4,6]. As of 2018, the entry-level wages for workers were US $62.42 [7-10]. The government of Bangladesh raised monthly wages by about 51% that year, setting the minimum wage at $94.22 [10]. In 2023, the entry-level wage was adjusted to BDT 12,500 ($102) per month with a 9% annual increment [11], which still reflects low pay. Though this sector is economically significant, it has also been criticized for labor exploitation, unsafe working conditions, and violations of workers' rights [12-15]. Such incidents have sparked greater awareness of workers' rights, mandating that all RMG factories comply with internationally recognized standards [14,16-19]. In Bangladesh, RMG factories are divided into compliant and noncompliant factories. While some facilities follow international labor standards and provide workers with better rights and living conditions (compliant factories) [20], many (noncompliant factories) expose workers, especially female, to verbal and physical abuse, lack of sanitation, long working hours, excessive workloads, and health risks [20-24]. Such a situation impacts the health of workers, especially female workers. Accordingly, health care awareness among RMG workers remains low [25], and access to formal health services is limited due to economic constraints, stigma, and limited health information [24,26].

The health-wealth gradient in labor markets demonstrates that workforce participation sometimes depends on workers' capacity to accept positions, which is shaped by population health and social safety nets [27]. The COVID-19 pandemic intensified the vulnerabilities of this workforce. Global supply chain disruptions and order cancellations (approximately USD $2.8 billion worth of orders were withdrawn) by Western brands led to the closure of thousands of factories [28,29], leaving over a million workers temporarily unemployed without wages or severance pay [30]. While the government of Bangladesh attempted relief packages (approximately US $13,549 million), there were no safeguarding policies to ensure job and health security or social protection [31]. In addition, living in densely populated settlements, RMG workers faced heightened health risks, food insecurity, inadequate emergency aid, and emotional distress during the pandemic [32,33].

The pandemic has, in addition, underscored the growing role of technology in emergency response. In both developed and developing contexts, crisis response technologies, such as crowdsourcing platforms, helplines, and mobile health services, have improved access to timely support [34-36]. In Bangladesh, services such as 999 (general emergency) and 333 (COVID-19), as well as telehealth platforms such as Tonic (Telenor Health and Grameenphone) and Doctime (Doctime Limited), became available [37-39]. Recent work on AI-based risk assessment tools highlights critical considerations for deploying conversational health technologies in real-world settings, including the need for evidence of grounding and safety-focused evaluation [40]. However, these solutions often remain inaccessible to low-income populations due to limited digital literacy, lack of trust, and gendered disparities in technology use [41,42]. Workers in informal economies, including RMG workers, typically use basic mobile phones primarily without internet connectivity for communication rather than accessing health or emergency services [43,44].

Given this context, there was a critical need to understand how RMG workers interact with technology, manage health needs, and cope with crises in order to design meaningful, accessible digital interventions. Yet, no such research has explored the living realities of this particular industrial low-income group in developing countries, especially during global crises such as COVID-19.

To address this gap, we examined the daily experiences of RMG workers before and during the pandemic on the basis of compliant and noncompliant factories, with a focus on health perceptions and technology use. Specifically, we investigated:

- Research question 1: How do RMG workers manage their daily lives, health care needs, and technology use in their existing work environment?
- Research question 2: How did the COVID-19 pandemic affect their daily lives, health, and access to support?

Aims and Approach

This research aimed to investigate key research questions and develop technology-based design implications that can provide emergency support for ready-made garment workers and other low-income groups during times of crisis. The goal is to create a platform that amplifies their voices. However, many existing technology services are underused due to usability barriers and a lack of contextual relevance. To address this issue, we propose a technology-based framework called VOICE (Virtual Outlet for Integrated Community Engagement). This framework leverages service interface platforms to offer emergency support while ensuring that the voices of low-income communities are heard. In this paper, we report the results of in-depth interviews conducted with RMG workers in Bangladesh.

Methods

Overview

We conducted qualitative interviews in 2 phases: before the COVID-19 pandemic (phase 1) and during the pandemic (phase 2) in Bangladesh, specifically between January and April 2020.

Phase 1 occurred from January to February 2020, before the COVID-19 lockdown in Bangladesh that started at the end of March 2020 (by that time, China halted the shipping of raw materials, accessories, machinery, and spare parts; it affected the apparel and textile industries [29]). During this phase, we interviewed 55 RMG workers (32 female and 23 male workers) across 6 focus group discussions (FGDs) to gain insights into their lives and daily experiences.

XSL•FO
RenderX

In phase 2, we conducted one-on-one follow-up phone interviews with 12 RMG workers (9 female and 3 male workers) from phase 1 during the first week of April 2020, after the COVID-19 lockdown was imposed in Bangladesh. Additionally, we interviewed 3 stakeholders (including merchandisers, managers, and factory owners) connected to the financial policies of the RMG sector.

Phase 1 Method

Participant Recruitment Technique

We recruited participants from 3 areas in the Dhaka division: Mirpur, Ashulia, and Gazipur, as shown in Tables 1 and 2. Key informants (KIs), such as landlords and house managers who are well-known and trusted within the local community, assisted us in recruiting workers.

Table . Dimensions of research and details of the participants.

Phase 1: study before the COVID-19 pandemic (January 20 to February 20)						
FGD[a] groups	Participants count, n	Sex	Age range (years)	Area	Factories workers are engaged, n	Working experiences (years)
Group 1	11	Female	22 - 38	Mirpur, Dhaka, Urban	9	3 - 10
Group 2	7	Male	25 - 40	Mirpur, Dhaka, Urban	7	3 - 10
Group 3	12	Female	20 - 30	Ashulia, Dhaka, Suburban	3	3 - 5
Group 4	8	Male	18 - 30	Ashulia, Dhaka, Suburban	7	3 - 10
Group 5	9	Female	18 - 40	Gazipur Circle, Gazipur, Suburban	7	3 - 5
Group 6	8	Male	18 - 40	Gazipur Circle, Gazipur, Suburban	7	3 - 15

[a]FGD: focus group discussion.

Table . Details of the participants.

Phase 2: study during the COVID-19 pandemic (first week of April 20)						
One to one interview	Participants count, n	Sex	Age range (years)	Area	Factories workers or stakeholders are engaged, n	Working experiences (years)
RMG[a] workers	12	Male and Female	18 - 35	Mirpur, Ashulia, and Gazipur Circle, Dhaka	10	3 - 10
Factory owner	1	Female	18 - 35	Dhaka	1	8
Assistant manager (compliance)	1	Male	18 - 35	Dhaka	1	8
Merchandiser	1	Male	18 - 35	Dhaka	1	5

[a]RMG: ready-made garment.

These workers frequently communicate with KIs regarding rent-related matters. The KIs managed RMG workers in various roles, such as helpers and operators, from both compliant and noncompliant factories. The participants were related but not family, such as friends and colleagues. We ensured that male and female workers were recruited separately in each area for their comfort. The participants' age range was 18 to 40 (mean 27.37, SD 5.65) years, and they represented 40 different RMG factories, as indicated in Table 1. Among them, 31% (17/55) worked in compliant factories, while 69% (38/55) were employed in noncompliant factories. The participants were eager to share their experiences, reflecting Bangladesh's collectivist societal structure [45]. FGDs allow participants to express their beliefs, experiences, perceptions, and attitudes [46]. Each discussion format involved an interactive exchange between participants and the research team in a group setting [46,47].

Workers typically work 6 days a week and spend their weekends, particularly Fridays, attending to necessary tasks at home and outside. To accommodate their schedules, we preferred to form large focus groups of up to 12 workers at a time within the same community. This format also encouraged participants to have their voices heard [48].

Study Design and Procedure

The study aimed to gain a brief understanding of the personal lives, social aspects, relationships, and everyday workplace experiences of RMG workers. We divided the interviews into 2 segments that highlight these workers' positive and negative

XSL•FO
RenderX

insights. First, we examined the daily work systems, including workload, health issues, work environment, relationships with coworkers, and interactions with supervisors. Second, we explored the workers' personal lives, focusing on financial stability, family dynamics, friendships, sources of entertainment, technology use, social relationships, and their responsibilities regarding children.

Before each discussion, we engaged with the participants to arrange the discussion venue, prioritizing their comfort [49], and an evaluator for the qualitative study. We aimed to create an interactive zone through informal discussions. We explained the research objectives, methods, ethics, and incentives at the beginning of each discussion. We ensured participants provided consent and addressed their concerns about the topics. Once consent was granted, the lead researcher (RJR) guided the discussions, while other team members assisted with follow-up questions and took notes. The study was semistructured and conducted in Bengali, with nontechnical questions for the participants. Each FGD lasted, on average, 120 (mean 121, SD 2.19) minutes. We audio-recorded each discussion with the participants' consent.

Analysis

The interviews were conducted in Bengali, which required transcription and translation of all the interviews. A total of 726 minutes of audio was translated and transcribed into English. Being fluent in both English and Bengali, we carefully transcribed and translated the audio directly into English (forward translation). We revised the translated transcription to ensure data validity and manually analyzed it. We read and reread all the transcriptions multiple times to become thoroughly familiar with the data. We started the analysis process after completing the first FGD because the collected contents and the analysis mutually apprised each other [50]. We used thematic analysis, starting with an open-inductive coding procedure that also includes both inductive and deductive processes [51,52]. After the first phase of open coding, we generated a codebook that the authors (RJR and NA) collectively revised to group similar codes into second-order codes. The first author (RJR) reviewed all the themes multiple times, and the authors (RJR and NA) generated the themes collaboratively. Theme 1 addressed the challenges RMG workers face in noncompliant factories, significantly affecting their well-being. In contrast, theme 2 emphasized the opportunities for RMG workers in compliant factories, positively impacting their quality of life.

Ethical Considerations

The institutional review board approved the research (application: IRB-No.0501). Participants had the right to skip any questions or withdraw from the discussions at any point without losing their incentives. All recordings, notes, and data collected are securely stored in a private drive, with access granted only to the researchers. For participating in the FGDs, we offered snacks and daily grocery items worth BDT 500 (US $4 as of Sep 1, 2026) to each participant as incentives. Each KI received the same snacks, daily grocery items, and BDT 1000 (US $8 as of Sep 1, 2026) for their additional fieldwork.

Phase 2 Method

Participant Recruitment Technique and Study Design

We could reach 12 individual RMG workers (9 female and 3 male workers) from the participants of phase 1 for one-on-one phone interviews (Table 1), having collected their phone numbers during that study. The significant reduction from 55 participants in phase 1 to 12 in phase 2 reflects the extraordinary circumstances during the initial COVID-19 lockdown in Bangladesh, when participants could be reached only by phone due to a complete shutdown. This represents a subset that may not fully capture the diversity of experiences across all factory types and locations. However, they did represent 10 different factories and the same 3 geographic areas, providing some breadth.

This follow-up investigation aimed to understand the impact of the COVID-19 pandemic on these workers, including how they receive immediate support and the health and safety situations in their localities. Additionally, we formally requested that factory authorities assist us in contacting potential stakeholders in the RMG sector. Through this process, we recruited 3 participants: a factory manager, a merchandiser, and a factory owner, all from different garment factories. This allowed us to gain insight into the overall state of the RMG sectors and the future challenges workers face, buyers' perspectives, and the role of authorities in relation to COVID-19.

Study Procedure and Analysis

We conducted audio phone conversations with the workers, each call lasting around 10 minutes. In total, we had 120 minutes of conversation (mean 10.1, SD 1.40 minutes). The 3 conversations with stakeholders were approximately 30 minutes each, resulting in 90 minutes of discussion (mean 30, SD 2 minutes). We did not record any phone calls and took real-time notes during the conversations. All discussions were conducted in their native Bengali language, and the lead researchers moderated the data collection.

We used the same content analysis method as previously discussed. Through inductive open coding, we generated first-order codes, grouped similar codes into second-order codes, and ultimately developed the final theme 3, which addressed how the COVID-19 pandemic made the RMG workers' lives difficult.

Ethical Considerations

This study was conducted under the same institutional review board guidelines. We ensured the anonymity of all participants and securely stored the notes in a locked drive. No incentives were provided for participation in the phone conversations; we verbally thanked each participant after our calls.

Results

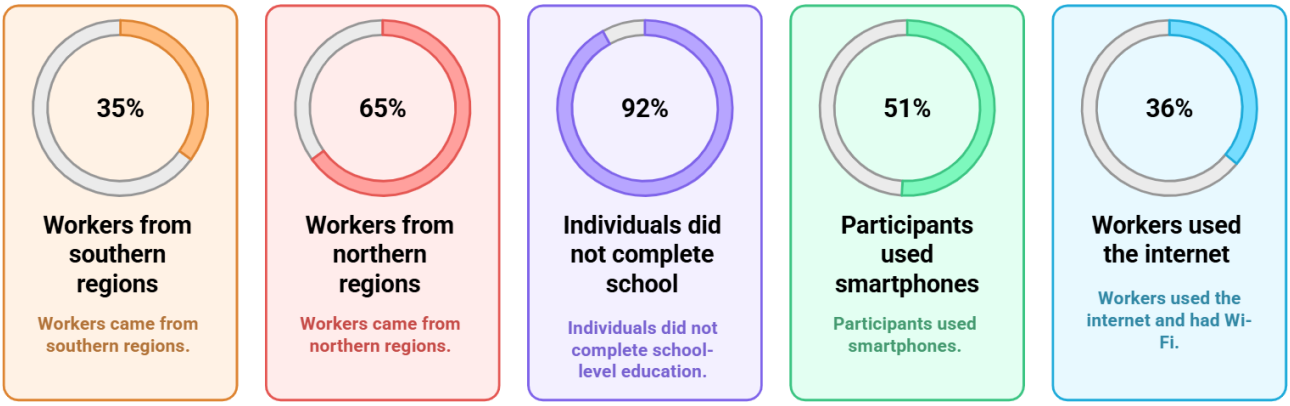
Overview

The holistic demography is shown in Figure 1. These participating workers came from the southern (19/55, 35%) and northern (36/55, 65%) regions of Bangladesh, both of which are low-income areas. Due to economic constraints, 92% (51/55)

XSL•FO
RenderX

of these individuals did not complete their school-level education. Despite their limited education, many secured early employment in RMG factories, as these facilities often prioritize hiring younger, physically fit workers to boost productivity. It was interesting since, in Bangladesh, male workers generally earn more than female workers in households [53]; however, female workers contribute significantly to family earnings, particularly in the RMG sector.

Figure 1. Participants' profile.



Additionally, 51% (28/55) of participants used smartphones, with higher use among male participants, though many were unaware of how to use them. About 36% (20/55) of workers used the internet and had Wi-Fi support at home, mostly from compliant factories. Their internet use was primarily limited to online calling apps (eg, IMO; PageBites, Inc) for communication with distant relatives and coworkers. A few male workers had social media but were inactive, while others rarely watched YouTube.

In the context of wages, at the time of phase 1 data collection (January-February 2020), the minimum wage for RMG workers was BDT 8000 (approximately US $65 as of August 17, 2026). However, a significant portion of workers, particularly in noncompliant factories, earn below the minimum wage. By the time of phase 2 (April 2020), the COVID-19 pandemic had led to widespread factory closures, leaving many workers without any wages or severance pay, compounding their existing financial vulnerability.

Theme 1: Sounds of Distress From Noncompliant Factory Workers

The working environment for employees in noncompliant factories greatly affects their well-being and lifestyles. Participants noted that noncompliant RMG factories often neglect workers' rights and health, resulting in these workers facing the harshest conditions (eg, limiting water intake or the ability to visit the toilet to increase productivity). Many female workers in noncompliant factories experience abuse. They reported that senior officers, such as line chiefs, floor managers, and supervisors, often exert pressure and punish primarily female employees. Additionally, these workers frequently do not receive their rightful salaries and remain silent about these injustices. Fearful of losing their jobs, female workers are reluctant to confront authority regarding abuse, making their situation even more challenging. Despite these hardships, they continue working due to their obligations, which impact their well-being all over. One female worker from a noncompliant factory expressed her emotional and mental state as follows:

> *Some line chiefs treat us very poorly. Supervisors are mean. I frequently feel like crying–we can't even wipe our eyes. If we cry, they will punish us by standing in front of the chamber for hours. They say this is not a place for crying; this is a workplace.* [Female worker, Ashulia, group 3]

Male participants shared that factory authorities often prefer hiring female workers over male workers because female workers tend to be quieter and less likely to protest against irregularities and misbehavior. Many female participants agreed, mentioning that there have been fewer protests in factories due to the high number of female employees, which has led to increased injustices impacting passively their well-being. A male worker from a noncompliant factory described the biased recruitment process, stating that young, attractive, and healthy workers are favored, particularly female workers. He said:

> *RMG authorities do not prefer elderly people or men. Overweight will not get a job. It's all about being young and slim. They prefer female because they are less likely to protest against mistreatment. They can pay female less.* [Male worker, Mirpur, group 2]

Because of such conditions, workers get sick and also often neglect their health and personal care to focus on their job responsibilities. They believe that their roles require them to remain healthy, but typically they do not receive adequate health care support from their factories. Additionally, authorities overlook workers' health care needs and treatment costs. Many workers reported that they often do not have proper restroom breaks due to excessive pressure, which limits their water intake throughout the day. Consequently, many are concerned about potential job loss that may arise from significant health issues. One female participant highlighted how their workload adversely affects their health:

> *We start working in the morning and continue without breaks, only pausing for lunch, then sitting again until we finish. We don't drink water or go to the restroom. Drinking water makes us feel pressured to use the washroom, which hampers production. Taking*

XSL•FO
RenderX

> *bathroom breaks is not prohibited, but we avoid them saving time.* [Female worker, Ashulia, group 3]

In addition, these workers often restrict the use of phones in the workplace. They have shared that they secretly use their phones on staircases or near restrooms to make calls during work hours. Otherwise, they will be punished. Furthermore, many do not have internet access to their home phones, as they cannot bear the additional costs of the internet for low pay. Additionally, due to heavy workload, they cannot manage time to use the phone, resulting in they cannot explore entertainment and any support options through phone. Therefore, this theme shows that noncompliant factory environments and lack of support decrease workers' quality of life.

Theme 2: Workers in Compliant Factories Lead Better Quality of Life

Participants from compliant factories unanimously highlighted the importance of timely salary payments and fair overtime compensation. Female workers reported significantly better job experiences, positively impacting their lives.

They also receive support from the factories regarding any workers' issues. One female worker expressed her feelings about increased income and teamwork at compliant facilities:

> *We do not feel overwhelmed by our workload. If someone is under too much pressure, management assigns additional workers to help. Sometimes, they implement a double process to meet targets. During shipment, we may work overtime until a maximum of 7 PM, and they compensate us properly for that. All the workers cooperate during shipment, and we are okay with it.* [Female worker, Gazipur, group 5]

Participants mentioned that compliant factories provide personal health facilities. Workers can take breaks whenever needed and access plenty of drinking water during their shifts. One female participant shared that there are medical staff at the factories, which many workers rely on due to the free services provided:

> *There is a medical team inside the factory. If someone feels ill, they prescribe medication, sometimes even provide medicines. They can be granted emergency leave or transported to a hospital by ambulance, with the company covering the expenses. Their salary will not be deducted; we depend on this support.* [Female worker, Gazipur, group 5]

Compliant factories also do not impose strict restrictions on communication, providing dedicated mobile phones on the factory floors for emergency calls, which the workers feel is good, as it helps to reduce the cost of regular personal communication. One female worker explained:

> *We do not use our personal mobile phones in the garments, but we are not bothered because several phones are available on every floor. We can use those for communication. If someone needs to reach us, they can call on those phones. We can talk even under heavy work pressure and frequently access them without any cost.* [Female worker, Gazipur, group 5]

Workers with internet access at home are typically employed in compliant factories. They usually share internet costs and live in relatively better areas than workers in noncompliant factories. This support often extends beyond shared costs, primarily provided by the house owner. Some believed that internet-based technology provided better resources for their children at home. They encourage their children to use mobile phones and the internet for educational purposes. One participant shared that he bought a computer for her daughter to use it:

> *She reads the Bengali newspaper online (on laptop). I encourage her to watch animations and tutorials on YouTube. I want to make my daughter a computer engineer.* [Male worker, Ashulia, group 4]

Therefore, the theme highlighted that compliant factory workers received significantly better support at the workplace, which encouraged them to lead a quality life.

Theme 3: COVID-19 Pandemic Makes Workers' Lives Difficult and Threatening

The COVID-19 pandemic has negatively impacted all factory workers, regardless of their compliance status. We found workers faced significant health care risks during the pandemic, as COVID-19 was highly contagious. Many workers lived in congested areas, often sharing a single floor with 6 to 8 families [54,55]. As a result, they struggled to maintain social distance, as their families shared kitchens and washrooms. Participants noted that they could not prevent their children from playing together in the evenings. This situation was consistent across all study areas. Compliant factory workers typically relied on medical support from their factories, but this support was halted due to the pandemic, raising concerns for their well-being. A female worker expressed her worries about social distancing:

> *We can't maintain social distancing because our area is very congested. If you want to do it personally, I think you can't. We have a shared kitchen. If anyone is infected here, the situation will be dangerous. We do not have health care support or food from anywhere.* [Female worker, Mirpur, Dhaka]

Participants reported that their factories closed suddenly, often without prior notice, and many did not receive their salaries. This lack of communication left them struggling to survive in the following days. They expressed the need to be financially prepared due to the uncertainty brought about by the pandemic, particularly concerning their ability to pay monthly house rent, grocery bills, and utility costs. Most of them found it difficult to save money due to low incomes, with many workers indicating that they had no savings to cover their monthly expenses. Some participants noted that they did not receive emergency relief from the private or government sectors since relief efforts often focused on the extremely poor. Other participants echoed a similar sentiment. This lack of adequate support was voiced by a male worker, who said:

> *We do not get any salary in advance. We received no relief from anywhere, not even from our garments. The situation is the same for everyone. Our living areas are not slum areas, so government and private relief efforts focus on slum areas. The government*

XSL•FO
RenderX

> *announces salaries for workers. That might be another reason we are not receiving emergency relief. We are on an unpredictable journey. We just need food and treatment now, nothing else. We only need rice, red lentils, and oil, and that's it.* [Male worker, Mirpur, Dhaka]

Accordingly, the stakeholders in this study expressed concerns about the RMG sector, which has historically faced significant losses. During the study period, they noted that the Bangladesh government announced a US $592 million incentive package to support RMG factories and their workers [31]. This support is provided to 4000 registered RMG factories by the Bangladesh Garment Manufacturers and Exporters Association, all of which are either compliant or partially compliant with regulations. However, stakeholders were worried that the thousands of unregistered and noncompliant factories and their workers would not receive any incentive support. The factories lost ongoing contracts, which added financial pressure due to their reliance on imported goods from China during the early days of the COVID-19 outbreak. One support authority representative shared his thoughts on the challenges faced by RMG workers:

> *The pandemic leads to food crises. People need food, treatment, not clothing. Our foreign income will decline, and we must support 4 million workers. This is extremely difficult.* [Merchandiser, an RMG group]

Another representative highlighted the plight of RMG workers. Approximately 4 million workers are employed in both compliant and noncompliant factories, and their wages are directly tied to the factories' income. Workers and individuals connected to the RMG industry will experience difficulties if these factories cannot sustain their operations. A management representative reinforced this sentiment:

> *The entire garment sector is at risk. Even if some factories collapse, millions of families will be affected; only Allah (God) knows what will happen!* [Assistant manager, a textile factory in Bangladesh]

Discussion

Principal Findings

We interviewed RMG workers and stakeholders from the RMG industry in Bangladesh. The interviews provided insights into the working environment, work experiences, perceptions of health care, lifestyle, and technology use, which varied depending on the type of RMG factory. This variation significantly impacted the well-being of the workers. Specifically, workers from compliant factories received essential health care support and demonstrated more effective use of technology than those in noncompliant factories. The disparities between compliant and noncompliant factory workers in our study mirror the regional disparities observed in labor market analyses [27]. As the pandemic spread globally, the circumstances for all workers changed, irrespective of their factory type. This shift profoundly affected their daily lives, raised health and safety concerns, and created a lack of emergency support.

The RMG industry faced numerous challenges, highlighting the urgent need for focused emergency relief and health care solutions for this community. Although the RMG industry has since begun to recover from the pandemic impacts, our findings reveal the community's vulnerabilities during emergencies. Based on these insights, we proposed a technology-based framework design, VOICE, to improve the lifestyle and well-being of this community, which struggles with limited exposure to technology. This initiative could also benefit other low-income communities in the future.

Requirement of Contextualized and Accessible Technology Intervention

The participants expressed that they work hard in a noncompliant work environment, but overall neglect their health issues. Although several earlier studies have highlighted the challenges faced by RMG workers. For instance, one study examined the workplace challenges faced by RMG workers in both export processing zone and nonexport processing zone factories [56,57]. While their findings were similar to those observed in noncompliant factories (theme 1), the studies did not address the conditions in compliant factories (addressed by theme 2), which is a crucial point to note, because compliant factory workers are facilitated with better support at the workplace. We cannot change the work environments of noncompliant factories, which depend on high-level stakeholders, and the National Tripartite Plan of Action on Fire Safety and Structural Integrity is constantly working to improve them [13,15,58]. We mostly consider their health problems, as previous studies found similar findings [21-24,59], but this worsened during the pandemic. They also needed emergency relief to survive. Another study analyzed the numerical impact of the COVID-19 pandemic on RMG workers [55] but did not empirically examine the situation (addressed by theme 3). Our research differs from existing research, focusing on supporting the health and emergency response of RMG workers through technological implications.

During the COVID-19 pandemic, workers lacked emergency aid, including health care support, financial assistance, and food. National policymakers prioritized monetary factors over workers' safety and health. As a result, RMG factories remained open [60], leading to confusion over the payment of outstanding employee salaries and prompting large numbers of workers to return to their factories [61]. Thousands of RMG workers walked toward factories as public transport was unavailable due to lockdown measures [60]. This situation highlighted a gap in information dissemination and created confusion about where to seek help. The prolonged COVID-19 crisis led to the spread of the Delta variant, resulting in further factory closures [62] and the need for emergency support. Later, the Omicron variant also affected RMG factories [63]. The disparities experienced during this crisis called for urgent attention. The continuous struggle of the workers, who held on to hope for a better future, was evident in their dreams. Maintaining social distance [64] was frustrating for many, as they relied on daily earnings. Some relocated to villages to manage their expenses, disregarding the risks of COVID-19 [32,33,65]. Engaging participants enabled researchers to explore ways to help communities seek assistance and support from one another. The intuitive systems designed

XSL•FO
RenderX

for Western countries require redesigning and reforming to better suit the conditions of poverty-stricken, densely populated regions like Bangladesh [66]. The challenges participants face present design opportunities to incorporate the voices of low-income and marginalized communities.

RMG workers in Bangladesh frequently struggled to access online health care services [38,67,68] due to limited access to and knowledge of technology. Toyama [69] discussed that technology integration could address existing societal problems; however, low-resource communities often require unique, simple, and effective systems that accommodate limited technology options. Connecting them with available technology poses challenges. Thus, a contextualized design approach can support users with restricted technological capabilities. That is why understanding their technological familiarity through this research was important.

Information from various contact points is widely available on social media but remains inaccessible to many communities. For instance, the recent flooding in southern Bangladesh demonstrated that obtaining information from victims was challenging, and many victims were excluded from emergency relief [70]. A point-of-contact–based service could function similarly to Facebook's service in terms of location and crisis response during emergencies [34]. One notable example is Ushahidi, a crisis-mapping organization in Kenya, Afghanistan, Haiti, and Mexico that gathers data from social media networks [35,71]. In Bangladesh, OpenStreetMap has delivered emergency support to remote regions during natural disasters [36]. In Pakistan, experts have suggested that coordination platforms could be effective tools for facilitating communication among donors and volunteers during emergency relief [72].

During the COVID-19 crisis, nongovernmental organizations and foundations in Bangladesh collaborated on a single online platform to offer nationwide support [73]. For instance, to address food and medicine shortages among low-income and marginalized communities, Building Resources Across Communities developed a map pinpointing slum areas in Bangladesh, helping donors assist at specific locations [74]. Existing emergency response systems often require internet access and advanced technology, whereas a more feasible service could leverage mobile networks across Bangladesh, enabling communities to reach out via a toll-free number. Toll-free services have been implemented multiple times in Bangladesh [37,38,67,68,73], but these efforts have not yet converged to create a platform for donors and volunteers during emergencies [72]. Our understanding of RMG workers showed that they possess mobile phones but limited skills. Therefore, the technology design should consider a simple mechanism that requires minimal new skill learning.

Design Implication as VOICE: Merging Service Interface and Service Providers

Overview

We propose a technology-based framework design named "VOICE," focusing on consolidating existing support systems into a single minimalist platform, working as a community to assist low-income and digitally low-literate groups, particularly in emergency support and health care, using their existing technology access and knowledge. While advanced AI-based health assessment tools show promise, their deployment in low-resource settings requires careful attention to evidence of grounding and safety evaluation to prevent harm from unverifiable outputs [40]. Our VOICE framework avoids these complexities by providing a simple and human-mediated interface. Table 3 shows how the empirical findings guide the design considerations for VOICE. Then, in the VOICE framework design illustrated in Figure 2, a service interface will collaborate with service providers to offer user-requested services to communities and individuals, and the communication model will be virtual.

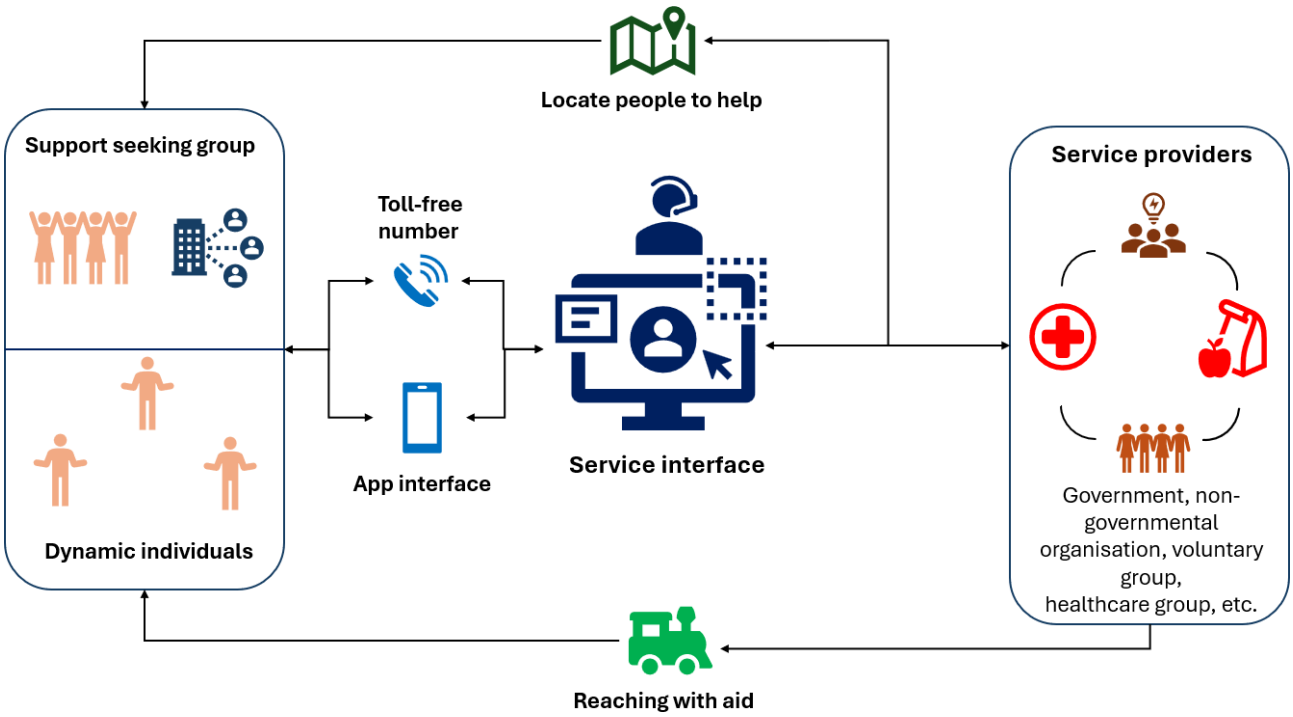
Table . Bridging empirical insights to the VOICE[a] framework design.

Empirical finding	Design implication	VOICE feature
Workers have limited digital literacy and use basic phones (theme 1)	System must function without smartphone or internet dependency	Toll-free number as primary access point
Workers lack access to health care during emergencies (themes 2 and 3)	Need for integrated health support regardless of factory type	Health care provider integration module
Workers face sudden factory closures without communication (theme 3)	Need for emergency alert systems accessible to all	Broadcast notification capability
Workers live in congested areas with shared facilities (theme 3)	Support systems must account for community-level needs	Location-based relief coordination
Workers lack knowledge of available services (themes 1 - 3)	Single, consolidated platform needed	Unified service interface aggregating existing providers
Workers experience financial constraints (theme 1)	Service must be free to access	Toll-free number with no call charges
Workers change locations or phone numbers frequently (themes 2 and 3)	System must accommodate mobile populations	Dynamic user database with flexible contact updates

[a]VOICE: Virtual Outlet for Integrated Community Engagement.

XSL•FO
RenderX

Figure 2. VOICE framework for supporting low-income groups. NGO: nongovernmental organization; VOICE: Virtual Outlet for Integrated Community Engagement.



Service Interface

The service interface will act as a platform that provides all users with a toll-free number, enabling multiple ways to access these services in their native language. This will facilitate interactions between service providers and individuals or groups seeking support. This will allow individuals to request assistance with emergency aid, health care, and related matters. For instance, if someone requests emergency assistance, the service interface can create a location-based database, as has been done in South America to map the supply chain of the offshore wind industry and send the main goods when demand is high [75]. This database can identify nearby individuals who may also need help, enabling efficient distribution of emergency relief. Similarly, if a group seeks health care assistance, the service interface can connect them remotely with a health care platform. At this point, the service interface can initiate the Relief Centre model suggested by Ozen and Krishnamurthy [76] to expand the services, which will include points of distribution, physical capacity such as volunteers, potential inventories to points of distribution, and a triage queue to check victim information before item distribution.

Even outside of crises, service providers can help individuals access the regular health care system at a marginal cost by using the databases available through the service interface. Though it is discussed that emergency medical services development in low-income countries involves understanding the unique needs of diverse communities, requires broad stakeholder involvement to maintain volunteer networks, and can be achieved through low-cost, bottom-up strategies [8], these needs need to be incorporated into the service interface and service providers. The community's mobility is essential, as workers often change locations, contact points, and phone numbers over time. Therefore, community contact points must be flexible and able to adapt to real-time needs, given the dynamic nature of these populations. Stakeholders, such as telecommunications companies, can support and facilitate the service interface model.

Service Providers

We examine the existing service providers in Bangladesh that use various interfaces inspired by best practices. For example, mapping tools are designed to connect donors with needy communities [74]. Some nongovernmental organizations and volunteer organizations contacted low-income individuals to offer assistance [73]. The Bangladeshi government sold rice at BDT 10 per kilogram to support low-income families [77]. There are also health care platforms [38,67,68] that support individuals seeking help. However, many of these services currently use a top-down approach, in which providers conduct community needs assessments. In this proposed technology-based design, we suggest that service providers and interfaces collaborate on a single platform, thereby enhancing the effectiveness of the services offered. This VOICE framework design and its possibilities could ensure that the voices of low-income communities are heard, with implications for future developments.

Addressing Digital Literacy Barriers in VOICE Design

Our findings revealed that while 51% (28/55) of participants had smartphones (themes 1 and 2), actual use was limited to basic calling apps (eg, IMO), with few accessing health services online. This digital literacy gap informed 4 core design principles in VOICE:

- No internet dependency: Workers do not need internet connectivity or data plans. The system operates through standard cellular networks in the country.

XSL•FO
RenderX

- No cost to access: Users do not incur call charges, eliminating financial barriers. This was critical, given workers' economic constraints (theme 1).
- Language accessibility: All interactions can be performed in their native language using simple, nontechnical vocabulary.
- Human backup: Live operators can be available 24/7 for users who cannot navigate automated systems, recognizing that some workers may need human assistance.

By focusing on these aspects, VOICE aims to make support accessible even to users who have never used digital services before.

Limitations and Way Forward

We have acknowledged our limitations for this research. First, we interviewed 55 RMG workers before the COVID-19 pandemic, but we could reach only 12 during the pandemic. This reduction reflects the unprecedented disruption caused by Bangladesh's national lockdown in April 2020, including transportation suspension, worker displacement, and communication barriers. This limited participation limits us from gaining insight into all the unique challenges the pandemic poses. The findings from phase 2 are presented as exploratory insights rather than statistically representative data. Similar to qualitative research, the focus is on depth and theoretical saturation rather than sample size. The data from phase 2 provided a rich and contextual understanding of the immediate impact of the pandemic on a specific group of workers. In addition, such limitations can potentially be mitigated in future emergencies through the following recruiting strategies by Kim et al [78], who recruited rural participants during the COVID-19 pandemic via local pharmacist referrals and community outreach, which helped facilitate communication and build trust and rapport between participants and researchers.

Second, during FGDs, not all participants were equally responsive because we had to arrange large groups, leading a few to remain silent. According to Simmel's [79] theory of group size, large groups result in weaker member involvement and participation. At this point, we noted nonverbal cues (eg, expressions) and brief comments from quieter participants, which were included in the analysis. Silent participants' perspectives were also partly captured through the similarities in experiences expressed by more vocal participants and the consistency of themes across groups. In future research, it is necessary to arrange small focus groups to obtain rich data, primarily from female participants [80]. In addition, social, religious, and aristocratic sectors are relevant for small interviews [79].

Third, following this study, it took approximately 3 years to achieve regular support; however, the necessary support systems during the emergency were not implemented. In future research, we can involve such RMG communities in design sessions, such as co-design or participatory design workshops, where users can cocreate or ideate the design recommendations, highlighting the practical real-world challenges and making the system inclusive [81,82].

Conclusions

This research explored the workplace environment, daily lives, challenges, personal perspectives on health care, and the use of technology by RMG workers, as well as the crisis they faced during the pandemic. We found that many of them neglect their health care needs and require more support from various sources. The complexity of issues arising from the COVID-19 pandemic is significant. The workers expressed significant concerns about job security, food availability, and health care, highlighting the gaps in support that are often overlooked by authorities and themselves. The workers' distress was echoed by the stakeholders. In response, we have proposed a technology-based conceptual framework, VOICE, that suggests workers could connect with service providers through a simple point-of-contact interface, such as a toll-free number. This setup would enable service providers to reach marginalized communities during emergencies or in regular situations, providing the support needed to improve their lives and well-being.

Acknowledgments

The authors thank all their participants for their cooperation and support. The authors extend their special thanks to Dr Kimia-tuz-Zaman from North Dakota State University, and Gazi Fokhrul Islam and Anik Saha for their contributions during this study's design and interview phases. During the preparation of this manuscript, the authors used Grammarly for the purpose of language editing and improving readability. The authors reviewed and edited the suggested text changes and take full responsibility for the final version of the manuscript.

Funding

This study was financially supported by BRAC University, Bangladesh, and logistically supported by the Design Inclusion and Access Lab at North South University.

Data Availability

The datasets generated during the study are not publicly available due to participants' confidentiality restrictions but are available from the corresponding author on reasonable academic request, subject to institutional review board approval.

XSL•FO
RenderX

Authors' Contributions

RJR and NA designed and conducted the qualitative study, with NA supervising the overall process. RJR played a significant role in writing the manuscript from start to finish, as well as supervising data analysis, and conducting the literature review. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. TradeInformation. BGMEA. URL: https://bgmea.com.bd/page/TradeInformation [accessed 2023-06-01]
2. World development indicators. World Bank Group. 2019. URL: https://databank.worldbank.org/data/source/world-development-indicators [accessed 2023-06-01]
3. Mirdha RU. Bangladesh remains the second largest apparel exporter. The Daily Star. URL: https://www.thedailystar.net/business/export/bangladesh-remains-the-second-biggest-apparel-exporter-1614856 [accessed 2026-08-24]
4. Islam MS. Ready-made garments exports earning and its contribution to economic growth in Bangladesh. GeoJournal 2021 Jun;86(3):1301-1309. [doi: 10.1007/s10708-019-10131-0]
5. Islam MS, Rakib MA, Adnan A. Ready-made garments sector of Bangladesh: its contribution and challenges towards development. J Asian Dev Stud 2016;5(2):02-12. [doi: 10.62345/jads.2016.5.2.6]
6. Lewis WA. Economic development with unlimited supplies of labour. Manch Sch 1954 May;22(2):139-191. [doi: 10.1111/j.1467-9957.1954.tb00021.x]
7. Alam MN, Alias R, TahlilAzim M. Social compliance in ready-made garment industry in Bangladesh: Challenges for HR personnel. J Appl Environ Biol Sci 2017;7(10):8-18 [FREE Full text]
8. Delaney PG, Moussally J, Wachira BW. Future directions for emergency medical services development in low- and middle-income countries. Surgery 2024 Jul;176(1):220-222. [doi: 10.1016/j.surg.2024.02.030] [Medline: 38599983]
9. Correspondent S. Tk 8,000 a month. The Daily Star. URL: https://www.thedailystar.net/business/news/bangladesh-rmg-garment-workers-minimum-salary-8000-taka-announced-1633342 [accessed 2026-08-24]
10. Ramesh Kurpad M. Made in Bangladesh: challenges to the ready-made garment industry. J Int Trade Law Policy 2014 Mar 11;13(1):80-96. [doi: 10.1108/JITLP-06-2013-0019]
11. Ahmed FZ, Greenleaf A, Sacks A. The paradox of export growth in areas of weak governance: the case of the ready made garment sector in Bangladesh. World Dev 2014 Apr;56:258-271. [doi: 10.1016/j.worlddev.2013.11.001]
12. Ahamed F. Improving social compliance in Bangladesh's ready-made garment industry. Lab Manag Dev 2013;13:1-17 [FREE Full text]
13. Ansary MA, Barua U. Workplace safety compliance of RMG industry in Bangladesh: structural assessment of RMG factory buildings. Int J Disaster Risk Reduct 2015 Dec;14:424-437. [doi: 10.1016/j.ijdrr.2015.09.008]
14. Barua U, Ansary MA. Workplace safety in Bangladesh ready-made garment sector: 3 years after the Rana Plaza collapse. Int J Occup Saf Ergon 2017 Dec;23(4):578-583. [doi: 10.1080/10803548.2016.1251150] [Medline: 27808602]
15. About us. International Labour Organization. 2020. URL: https://www.ilo.org/sites/default/files/wcmsp5/groups/public/@asia/@ro-bangkok/@ilo-dhaka/documents/sitestudiocontentelements/wcms_con_txt_dha_abo_en.pdf [accessed 2023-06-01]
16. Akhter S. Endless misery of nimble fingers: the Rana Plaza disaster. Asian J Women Stud 2014 Jan;20(1):137-147. [doi: 10.1080/12259276.2014.11666176]
17. Bolle MJ. Bangladesh apparel factory collapse: background in brief. : Congressional Research Service, the Library of Congress; 2014 URL: https://digital.library.unt.edu/ark:/67531/metadc463476/m1/1/high_res_d/R43085_2014Jan10.pdf [accessed 2026-08-13]
18. Sinkovics N, Hoque SF, Sinkovics RR. Rana Plaza collapse aftermath: are CSR compliance and auditing pressures effective? Account Audit Account J 2016 May 16;29(4):617-649. [doi: 10.1108/AAAJ-07-2015-2141]
19. Manik JA, Yardley J. Building collapse in Bangladesh leaves scores dead. The New York Times. 2013 Apr 24. URL: https://www.nytimes.com/2013/04/25/world/asia/bangladesh-building-collapse.html [accessed 2026-08-24]
20. Baral LM. Comparative study of compliant & non-compliant RMG factories in Bangladesh. Int J Eng Technol 2010;10(2):93-94 [FREE Full text]
21. Ahmed S, Raihan MZ. Health status of the female workers in the garment sector of Bangladesh. J Fac Econ Adm Sci 2014;4(1):43-58 [FREE Full text]
22. Akhter S, Rutherford S, Chu C. Sufferings in silence: violence against female workers in the ready-made garment industry in Bangladesh: a qualitative exploration. Womens Health (Lond) 2019;15:1745506519891302. [doi: 10.1177/1745506519891302] [Medline: 31856698]
23. Akhter S, Salahuddin AFM, Iqbal M, Malek A, Jahan N. Health and occupational safety for female workforce of garment industries in Bangladesh. J mech eng 2010;41(1):65-70. [doi: 10.3329/jme.v41i1.5364]

XSL•FO
RenderX

24. Steinisch M, Yusuf R, Li J, et al. Work stress: its components and its association with self-reported health outcomes in a garment factory in Bangladesh-findings from a cross-sectional study. Health Place 2013 Nov;24:123-130. [doi: 10.1016/j.healthplace.2013.09.004] [Medline: 24095949]
25. Hasan ATH, Hassan R, Khan ZR, Nuzhat E, Arefin U. Influence of socio-demographic factors on awareness of HIV/AIDS among Bangladeshi garment workers. Springerplus 2013 Dec;2(1):1-7. [doi: 10.1186/2193-1801-2-174]
26. Islam Mondal N, Islam R, Rahman O, Rahman S, Hoque N. Determinants of HIV/AIDS awareness among garments workers in Dhaka City, Bangladesh. World J AIDS 2012;02(4):312-318. [doi: 10.4236/wja.2012.24042]
27. Liu D, Shen Q, Liu J. The health-wealth gradient in labor markets: integrating health, insurance, and social metrics to predict employment density. Computation 2026;14(1):22. [doi: 10.3390/computation14010022]
28. Paton E. 'Our situation is apocalyptic': Bangladesh garment workers face ruin. The New York Times. 2020. URL: https://www.nytimes.com/2020/03/31/fashion/coronavirus-bangladesh.html?eafs_enabled=false [accessed 2023-06-01]
29. Kabir FH. Coronavirus impact: disruption in supply chain worries garment exporters. The Financial Express. 2020 Feb 14. URL: https://thefinancialexpress.com.bd/public/print/coronavirus-impact-disruption-in-supply-chain-worries-garment-exporters-1581396740 [accessed 2026-08-24]
30. Kelly A. Primark and Matalan among retailers allegedly cancelling £ 2.4 bn orders in 'catastrophic' move for Bangladesh. The Guardian. 2022 Oct 19. URL: https://www.theguardian.com/global-development/2020/apr/02/fashion-brands-cancellations-of-24bn-orders-catastrophic-for-bangladesh [accessed 2026-08-24]
31. Hossain J, Akter A, Ahmed M. The impact of covid pandemic on the garment workers in bangladesh. Fair Wear. 2021. URL: https://api.fairwear.org/wp-content/uploads/2021/07/The-impact-of-the-Covid-pandemic-on-garment-workers-with-a-specific-focus-on-women-workers.pdf [accessed 2026-08-24]
32. Kamruzzaman M, Sakib SMN. Bangladesh imposes total lockdown over COVID-19. Anadolu Ajansı. 2020 Mar 25. URL: https://www.aa.com.tr/en/asia-pacific/bangladesh-imposes-total-lockdown-over-covid-19/1778272 [accessed 2026-08-24]
33. Khan S, Siddique R, Ali A, Xue M, Nabi G. Novel coronavirus, poor quarantine, and the risk of pandemic. J Hosp Infect 2020 Apr;104(4):449-450. [doi: 10.1016/j.jhin.2020.02.002] [Medline: 32057788]
34. Crisis Response. Facebook Crisis Response. 2020. URL: https://www.facebook.com/about/crisisresponse [accessed 2023-06-01]
35. Gao H, Barbier G, Goolsby R. Harnessing the crowdsourcing power of social media for disaster relief. IEEE Intell Syst 2011;26(3):10-14. [doi: 10.1109/MIS.2011.52]
36. Latif S, Islam KMR, Khan M, Ahmed SI. OpenStreetMap for the disaster management in Bangladesh. Presented at: 2011 IEEE Conference on Open Systems; Sep 25-28, 2011. [doi: 10.1109/ICOS.2011.6079240]
37. COVID-19 dashboard for Bangladesh. COVID Dashboard. URL: https://dashboard.dghs.gov.bd/pages/covid19.php [accessed 2026-08-24]
38. Free health service TONIC launched for Grameenphone customers. Grameenphone. URL: https://www.grameenphone.com/about/media-center/press-release/free-health-service-tonic-launched-grameenphone-customers?srsltid=AfmBOorwzE8moDsl091wLLXO3_TmTKvu6vgc78p4z8Nub_8n4JLBtpbQ [accessed 2026-08-24]
39. National Emergency Service 999. International Telecommunication Union. 2020. URL: https://www.itu.int/net4/wsis/archive/stocktaking/Project/Details?projectId=1514906121 [accessed 2023-06-01]
40. Hu Y. Toward retrieval-grounded evaluation for conversational large language model-based risk assessment. JMIR AI 2026 Mar 12;5:e90759. [doi: 10.2196/90759] [Medline: 41818631]
41. Sambasivan N, Checkley G, Ahmed N, Batool A. Gender equity in technologies: considerations for design in the Global South. Interactions 2017;25(1):58-61. [doi: 10.1145/3036653]
42. Sambasivan N, Batool A, Ahmed N. "They don't leave us alone anywhere we go": gender and digital abuse in South Asia. Presented at: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019. [doi: 10.1145/3290605.3300232]
43. Sultana S, Ahmed SI, Fussell SR. "Parar-daktar Understands My Problems Better" Disentangling the challenges to designing better access to healthcare in rural Bangladesh. Proc ACM Hum-Comput Interact 2019;3(CSCW):1-27. [doi: 10.1145/3359270]
44. Sambasivan N, Checkley G, Batool A, Ahmed N, Nemer D, Gaytán-Lugo LS, et al. "Privacy is not for me, it's for those rich women": performative privacy practices on mobile phones by women in South Asia. Presented at: Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018); Aug 12-14, 2018 URL: https://www.usenix.org/system/files/conference/soups2018/soups2018-sambasivan.pdf [accessed 2026-08-13]
45. Kim HS, Sherman DK, Taylor SE. Culture and social support. Am Psychol 2008 Sep;63(6):518-526. [doi: 10.1037/0003-066X] [Medline: 18793039]
46. O.Nyumba T, Wilson K, Derrick CJ, Mukherjee N. The use of focus group discussion methodology: insights from two decades of application in conservation. Methods Ecol Evol 2018 Jan;9(1):20-32. [doi: 10.1111/2041-210X.12860]
47. Morgan DL. Focus groups. Annu Rev Sociol 1996 Aug;22(1):129-152. [doi: 10.1146/annurev.soc.22.1.129]
48. Hopkins PE. Thinking critically and creatively about focus groups. Area 2007 Dec;39(4):528-535. [doi: 10.1111/j.1475-4762.2007.00766.x]

XSL•FO
RenderX

49. Guba EG, Lincoln YS. Effective Evaluation: Improving the Usefulness of Evaluation Results through Responsive and Naturalistic Approaches: Jossey-Bass; 1981. URL: https://www.scribd.com/document/592640963/1981-guba-effectiveevaluation [accessed 2026-08-13]
50. Glaser B, Strauss A. Discovery of Grounded Theory: Strategies for Qualitative Research, 1st edition: Routledge; 1999. [doi: 10.4324/9780203793206-1]
51. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
52. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
53. Education Household Survey. : BBS Portal; 2014 URL: https://objectstorage.ap-dcc-gazipur-1.oraclecloud15.com/n/axvjbnqprylg/b/V2Ministry/o/office-bbs/2024/12/2b31744d1f674476bfdbd31e93ec5abb.pdf [accessed 2023-06-01]
54. Ahmed N, Jahangir Rony R, Tuz Zaman K. Social distancing challenges for marginal communities during COVID-19 pandemic in Bangladesh. J Biomed Anal 2020;3(2):5-14. [doi: 10.30577/jba.v3i2.45]
55. Rabi Ullah M. Impact of Covid-19 on readymade garments workers at Dhaka division of Bangladesh. Adv Soc Sci Res J 2024;11(4):210-225. [doi: 10.14738/assrj.114.16737]
56. Kabir H, Maple M, Usher K. Prevalence and risk factors of sexual harassment in the workplace by female readymade garment workers in Bangladesh. Front Public Health 2025;13:1580709. [doi: 10.3389/fpubh.2025.1580709] [Medline: 40475200]
57. Kabir H, Maple M, Islam MS, Usher K. A qualitative study of the working conditions in the readymade garment industry and the impact on workers' health and wellbeing. Environ Occup Health Pract 2022;4(1):2021-0020. [doi: 10.1539/eohp.2021-0020-OHW] [Medline: 40060310]
58. Staying engaged: A Sustainability Compact for continuous improvements in labour rights and factory safety in the Ready-Made Garment and Knitwear Industry in Bangladesh, 2013, Geneva. : Trade Europa; 2013 URL: https://www.ilo.org/sites/default/files/wcmsp5/groups/public/%40asia/%40ro-bangkok/%40ilo-dhaka/documents/genericdocument/wcms_397570.pdf [accessed 2023-06-01]
59. Akhter S, Rutherford S, Chu C. Sewing shirts with injured fingers and tears: exploring the experience of female garment workers health problems in Bangladesh. BMC Int Health Hum Rights 2019 Jan 21;19(1):2. [doi: 10.1186/s12914-019-0188-4] [Medline: 30665456]
60. Star Online Report. Coronavirus: authorities decide to keep factories open. The Daily Star. 2020. URL: https://www.thedailystar.net/coronavirus-deadly-new-threat/news/coronavirus-authorities-decide-keep-factories-open-1884043 [accessed 2026-08-24]
61. Bangladesh apparel factory workers rush back amid shutdown. New Age. 2020 Mar. URL: https://www.newagebd.net/article/103809/bangladesh-apparel [accessed 2023-06-01]
62. Post-Eid Lockdown: Jul 23-Aug 5: Offices, factories to stay shut. The Daily Star. 2021 Jul 8. URL: https://www.thedailystar.net/news/bangladesh/news/offices-factories-stay-shut-2134176 [accessed 2026-08-24]
63. RMG exports may face Omicron blow: businesses. RMG Bangladesh. 2021. URL: https://rmgbd.net/2021/12/rmg-exports-may-face-omicron-blow-businesses [accessed 2023-06-01]
64. Ebola outbreak requires urgent scale-up of response. World Health Organization. 2020. URL: https://www.who.int [accessed 2023-06-01]
65. CDC Weekly. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China [Article in Chinese]. Zhonghua Liu Xing Bing Xue Za Zhi 2020 Feb 10;41(2):145-151. [doi: 10.46234/ccdcw2020.032]
66. Saleh A. In Bangladesh, COVID-19 threatens to cause a humanitarian crisis. The European Sting. 2020 Apr 7. URL: https://europeansting.com/2020/04/07/in-bangladesh-covid-19-threatens-to-cause-a-humanitarian-crisis/ [accessed 2026-08-24]
67. Best online doctor consultation & telemedicine in Bangladesh. DocTime Healthcare Anytime Anywhere. URL: https://doctime.com.bd [accessed 2026-08-24]
68. Health Plus. ROBI. URL: https://www.robi.com.bd/en/health-plus [accessed 2026-08-24]
69. Toyama K. Rescuing social change from the cult of technology. In: Innovations in Teaching & Learning Conference Proceedings 2017, Vol. 9. URL: https://journals.gmu.edu/ITLCP/article/view/1853 [accessed 2026-09-29]
70. Alam Z. Shelter Cluster_Eastern Floods_Shelter Damage and Needs Assessment 01102024. Global Shelter Cluster. 2024 Oct 1. URL: https://www.sheltercluster.org/bangladesh-floods-2024/documents/shelter-clustereastern-floodsshelter-damage-and-needs-assessment [accessed 2026-08-24]
71. Heinzelman J, Waters C. Crowdsourcing crisis information in disaster-affected Haiti. : US Institute of Peace; 2010 URL: https://preparecenter.org/sites/default/files/crowdsourcing_crisis_information_in_disaster-affected_haiti.pdf [accessed 2026-08-13]
72. Krüger M, Ahmad SSO, Shahid S. Investigating the requirements of an online emergency response platform. 2017 Nov 16 Presented at: ICTD '17: Ninth International Conference on Information and Communication Technologies and Development; Nov 16-19, 2017. [doi: 10.1145/3136560.3136605]
73. Cholo Sobai. URL: https://choloshobai.com/ [accessed 2026-08-24]
74. BRAC. URL: https://humanitarianaidmanagement.brac.net/ [accessed 2026-08-24]

XSL•FO
RenderX

75. Siqueira F, Rodrigues T, Faria C, Campos B, Matsumoto W. Mapping Brazil and Rio de Janeiro's offshore wind energy supply chain: opportunities and challenges. Presented at: Offshore Technology Conference Brasil; Oct 28-30, 2025. [doi: 10.4043/36200-MS]
76. Ozen M, Krishnamurthy A. Evaluating relief center designs for disaster relief distribution. J Humanit Logist Supply Chain Manag 2018 Apr 3;8(1):22-48. [doi: 10.1108/JHLSCM-03-2017-0012]
77. Govt to start sale of rice at Tk 10 per kg in Dhaka. The Financial Express. 2020 Apr 04. URL: https://thefinancialexpress.com.bd/national/govt-to-start-sale-of-rice-at-tk-10-per-kg-in-dhaka-1586000315 [accessed 2026-08-24]
78. Kim NH, Wilson N, Mashburn T, et al. Lessons learned recruiting a diverse sample of rural study participants during the COVID-19 pandemic. Int J Drug Policy 2021 Nov;97:103344. [doi: 10.1016/j.drugpo.2021.103344] [Medline: 34186474]
79. Simmel G. The number of members as determining the sociological form of the group. I. Am J Sociol 1902 Jul;8(1):1-46. [doi: 10.1086/211115]
80. Toner J. Small is not too small: reflections concerning the validity of very small focus groups (VSFGs). Qual Soc Work 2009;8(2):179-192. [doi: 10.1177/1473325009103374]
81. Zamenopoulos T, Alexiou K. Co-design as collaborative research. : University of Bristol; 2018 URL: https://oro.open.ac.uk/58301/1/Co-Design_CCFoundationSeries_PUBLISHED.pdf [accessed 2026-08-13]
82. Jagtap S. Co-design with marginalised people: designers' perceptions of barriers and enablers. CoDesign 2022 Jul 3;18(3):279-302. [doi: 10.1080/15710882.2021.1883065]

Abbreviations

FGD: focus group discussion
KI: key informant
RMG: ready-made garment
VOICE: Virtual Outlet for Integrated Community Engagement

Edited by A Kushniruk; submitted 01.Jan.2026; peer-reviewed by L Herdiman, Y Hu; revised version received 26.Jun.2026; accepted 22.Jul.2026; published 02.Oct.2026.

Please cite as:
Rony RJ, Ahmed N
Designing for Amplifying Voices—Supporting Bangladesh's Low-Income Ready-Made Garment Workers Through Technology Amid COVID-19 Pandemic: Interview Study Among Workers
JMIR Hum Factors 2026;13:e90675
URL: https://humanfactors.jmir.org/2026/1/e90675
doi:10.2196/90675

© Rahat Jahangir Rony, Nova Ahmed. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 2.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluating Acceptance and Use of Digital Health Platforms for Consultation and Medication Delivery Among Men Who Have Sex With Men and Transgender People Living With HIV: Cross-Sectional Survey Informed by an Extended UTAUT Approach

Emmanuel Oluwatosin Oluokun*, PhD; Festus Fatai Adedoyin*, PhD; Huseyin Dogan*, Prof Dr; Nan Jiang*, Prof Dr

School of Computing and Engineering, Faculty of Media, Science & Technology, Bournemouth University, Fern Barrow, Poole, England, United Kingdom

*all authors contributed equally

Corresponding Author:
Emmanuel Oluwatosin Oluokun, PhD
School of Computing and Engineering, Faculty of Media, Science & Technology, Bournemouth University, Fern Barrow, Poole, England, United Kingdom

Abstract

Background: Digital health platforms can expand access to HIV care, but among men who have sex with men (MSM) and transgender people living with HIV and AIDS in Nigeria, adoption is shaped by structural stigma, criminalization, and fear of disclosure as much as by system functionality. Teleconsultation and medication-delivery platforms offer alternative pathways to care, but their acceptance within marginalized populations cannot be assumed and requires empirical investigation.

Objective: This study aimed to examine the factors influencing behavioral intention (BI) to adopt TechAids, a confidentiality-oriented digital health platform designed to support HIV consultation and service access among MSM and transgender individuals in Nigeria, using an extended Unified Theory of Acceptance and Use of Technology (UTAUT) framework.

Methods: A cross-sectional survey was conducted in Lagos, Nigeria, through HIV service delivery and outreach settings facilitated by partnering nongovernmental organizations, supplemented by online LGBTQ+ (lesbian, gay, bisexual, transgender, queer, and other sexual and gender minorities) networks. The study used nonprobability convenience sampling with prespecified eligibility criteria. Data collection was conducted from January 2025 to May 2025. The sample comprised 141 platform users and 27 health care providers (N=168). An extended UTAUT model incorporated performance expectancy (PE), effort expectancy (EE), social influence (SI), facilitating conditions (FC), trust, and perceived stigma (PS). Data were analyzed using correlation analysis and ordinary least squares (OLS) regression. Age and educational attainment were examined as potential moderating variables. Qualitative feedback was also collected to contextualize quantitative findings.

Results: FC emerged as the strongest predictor of BI to use the platform (β=.813; $P<.001$), explaining a substantial proportion of variance in adoption intention ($R^2=0.652$). Age demonstrated a modest but statistically significant positive effect on BI (β=.105; $P=.03$). Contrary to theoretical expectations, PE, EE, SI, trust, and PS did not significantly predict intention to adopt the platform. Qualitative feedback highlighted practical infrastructure-related concerns, including reliable internet access, offline functionality, and availability of technical support, as more salient than psychological or feature-based considerations.

Conclusions: In resource-constrained, highly stigmatized contexts, enabling infrastructure and practical support may outweigh cognitive, social, and attitudinal determinants of technology acceptance. Successful digital health interventions for marginalized populations therefore require privacy-conscious, user-centered design alongside sustained investment in the infrastructure that supports real-world use.

(JMIR Hum Factors 2026;13:e93129) doi:10.2196/93129

KEYWORDS

UTAUT; digital health; mHealth; HIV/AIDS; transgender; stigma; technology acceptance; mobile health; Unified Theory of Acceptance and Use of Technology

XSL•FO
RenderX

Introduction

Despite decades of progress in HIV treatment, sub-Saharan Africa continues to shoulder a disproportionate burden of infection and mortality [1]. Within this region, men who have sex with men (MSM) and transgender people are particularly vulnerable due to intersecting clinical, social, and legal factors. They experience higher HIV prevalence, reduced access to prevention and treatment services, and heightened exposure to stigma, discrimination, and violence.

Digitalization is reshaping how health services are designed and delivered, including the expansion of remote consultation, mobile health (mHealth) apps, and digitally mediated access to medicines and counseling. In HIV care, digital pathways can be particularly valuable where individuals face practical barriers (distance, time, and cost) and social barriers (fear of disclosure, stigma, and discrimination). These barriers are often intensified for MSM and transgender people in contexts where sexual and gender minorities experience marginalization, safety concerns, and reduced willingness to engage with clinic-based services. Although digital interventions promise improved continuity of care and confidentiality, adoption is not automatic; it depends on perceived value, usability, social context, enabling conditions, and trust in how sensitive information is handled [2-4]. A systematic review and meta-analysis of counseling support for HIV self-testing among MSM similarly underscores the centrality of structured counseling and linkage-to-care pathways within digital HIV interventions, reinforcing the case for embedding counseling and education functions, not only consultation scheduling, within platforms such as TechAids, developed by EO (lead author) [5].

Digital health technologies, including mHealth apps, teleconsultations, and online medication delivery, have been proposed as a way to extend care beyond physical clinic walls and offer discreet, flexible access to services. Evidence suggests that mHealth interventions can improve adherence to treatment and support self-management in chronic conditions [6], and that digital platforms can increase access to care in underserved populations [7]. However, acceptance of such tools depends on a complex interplay of perceived usefulness, usability, trust, privacy, and the structural conditions that enable or constrain use [8,9]. A recent randomized controlled trial of a chatbot-delivered HIV self-testing support intervention among MSM found it to be noninferior to human-delivered real-time support in increasing self-testing uptake and linkage to counseling, illustrating that digitally mediated, low-cost support models can match human-delivered care for key populations in other global settings [10].

The Unified Theory of Acceptance and Use of Technology (UTAUT), developed by Venkatesh et al [2], provides a well-established framework for understanding why individuals adopt or reject information systems. UTAUT has been widely applied across health care contexts, from electronic health record adoption by clinicians [11] to patient portal usage [12] and mHealth applications [13]. However, standard UTAUT constructs may inadequately capture the psychological and social complexities of HIV care, particularly for criminalized populations. HIV-related stigma, both anticipated and enacted, profoundly shapes health care-seeking behavior [14]. For MSM and transgender individuals, this stigma operates multiplicatively, intersecting with sexual orientation and gender identity stigma to create what researchers term “layered” or “syndemic” stigma [15,16].

Similarly, trust in health technologies becomes paramount when sharing sensitive information about HIV status, sexual behavior, and gender identity could expose individuals to discrimination or legal jeopardy [17,18]. McKnight et al [19] conceptualized trust in technology as comprising beliefs about reliability, functionality, and helpfulness, alongside faith in the provider’s benevolence and integrity. In stigmatized health domains, these dimensions acquire heightened salience; users must trust not only that systems will function reliably, but that data will be protected and that no harm will result from engagement.

This study examined BI to adopt TechAids, a digital platform developed specifically for discreet HIV service access in Nigeria. TechAids provides teleconsultation scheduling, medication delivery coordination, appointment reminders, and educational resources through a mobile app. The platform emphasizes privacy through end-to-end encryption, anonymous registration options, and discreet packaging for medication delivery. We extended the UTAUT framework by incorporating trust and Perceived Stigma (PS) as additional predictors, recognizing that standard technology acceptance models may inadequately represent the lived realities of marginalized populations navigating highly stigmatized health conditions.

Our research addressed three primary questions:

1. Which factors most strongly predict intention to adopt TechAids among MSM and transgender individuals living with HIV in Nigeria?
2. Do trust and PS play significant roles beyond standard UTAUT constructs in predicting adoption intention?
3. Do demographic characteristics, specifically age and education, moderate the relationships between UTAUT constructs and BI, suggesting the need for differentiated implementation strategies across user segments?

Methods

Theoretical Framework and Hypotheses

We built our conceptual model on UTAUT’s foundation while extending it to address the unique considerations of digital health platforms serving stigmatized populations. The UTAUT establishes a comprehensive framework that helps understand how users accept or reject technology solutions and how this shapes their behavioral patterns. In this work, we used the original UTAUT baseline model from Venkatesh et al [2]. This is taken up and elaborated in the context of a stigma-aware digital health context.

Performance expectancy (PE) represents the degree to which individuals believe that using TechAids will help them manage their HIV care more effectively [2]. In this context, PE encompasses beliefs about whether the platform will improve access to medications, simplify appointment scheduling, enhance

XSL•FO
RenderX

communication with health care providers, and ultimately contribute to better health outcomes. Prior research in HIV mHealth suggests that perceived utility for medication adherence and symptom monitoring strongly predicts app usage [18].

Effort expectancy (EE) captures perceived ease of use—the extent to which potential users anticipate that interacting with TechAids will be free from difficulty [2]. For populations with varying levels of digital literacy, interface simplicity and intuitive navigation become critical adoption factors [20]. Given that many MSM and transgender individuals in Nigeria access smartphones but may have limited experience with health applications, EE could substantially influence adoption intentions.

Social influence (SI) reflects the degree to which important others' opinions affect technology adoption decisions [2]. Within tight-knit MSM and transgender communities, peer recommendations and community advocate endorsements may carry particular weight [21]. However, social influence operates ambivalently in stigmatized contexts, while peer support may encourage adoption, concerns about community gossip or inadvertent disclosure might inhibit it [17].

Facilitating conditions (FC) encompasses perceptions of available resources and infrastructure support for platform use [2]. This construct includes access to smartphones and reliable internet connectivity, compatibility with existing devices, availability of technical assistance, and adequacy of bandwidth for app functionality. In sub-Saharan African contexts, where internet penetration remains uneven and data costs consume significant portions of household budgets, FC frequently emerge as critical determinants of mHealth adoption [22].

Trust represents confidence that TechAids will reliably protect sensitive information and operate in users' best interests [19]. We conceptualized trust as comprising 3 dimensions: confidence in the platform's technical security (data encryption and secure storage), faith in the organization's benevolence (commitment to user welfare), and belief in institutional integrity (honest representation of capabilities and limitations). For platforms handling HIV-related data, trust becomes foundational; users must believe their information will not be compromised, sold, or disclosed to authorities [17].

Perceived stigma (PS) captures individuals' concerns about potential discrimination, judgment, or social consequences resulting from platform use. We measured both anticipated stigma (expectations of being devalued if HIV status becomes known) and internalized stigma (negative self-perceptions related to HIV). While TechAids explicitly addresses stigma through confidentiality features, PS might still deter adoption if users worry about visible app icons, delivery package identification, or data breaches.

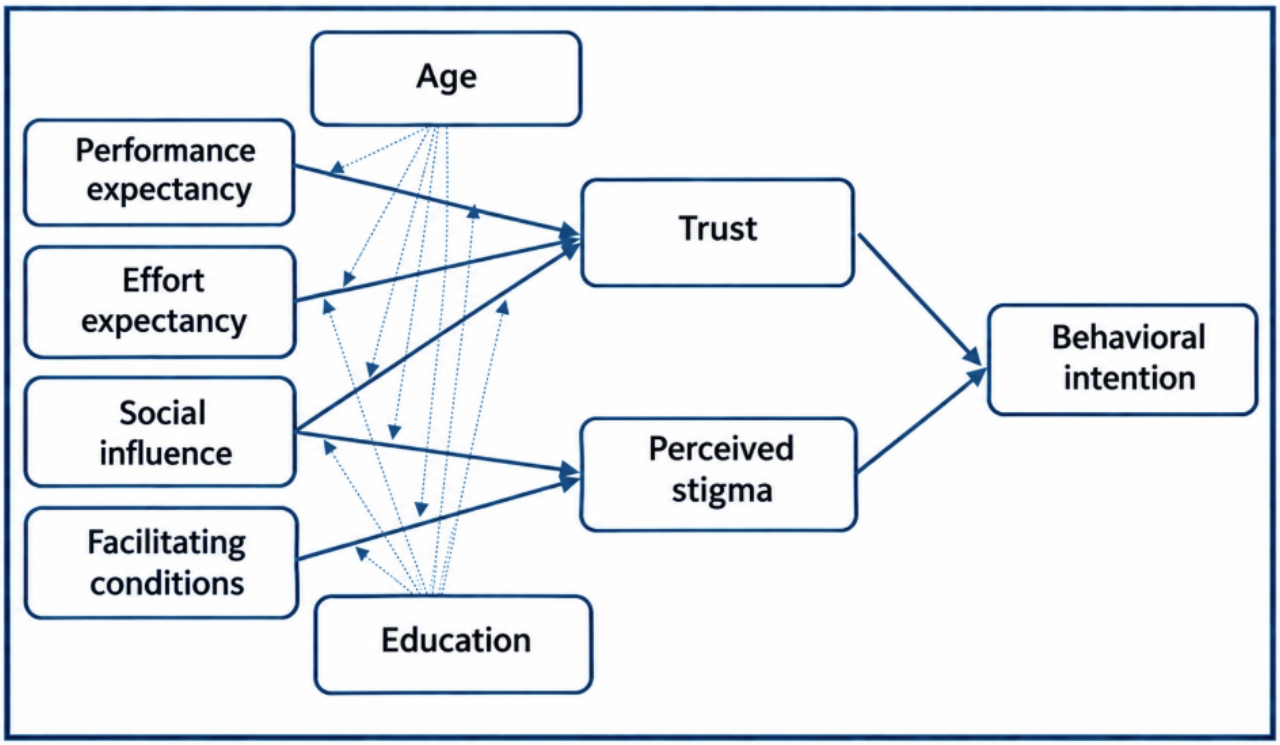
We also examined age and education as potential moderators, recognizing that technology adoption patterns often vary across demographic groups [2]. Younger individuals typically demonstrate greater digital fluency, while education level may influence both technical competence and health care information processing.

Education and counseling are also core, rather than peripheral components of HIV care, and this is reflected directly in TechAids' design: alongside teleconsultation and medication delivery, the platform provides educational resources intended to support adherence counseling and risk-reduction communication. We therefore situate PE and FC in this study not only as technology-acceptance constructs but also as proxies for whether the platform's behavioral and counseling functions are perceived as effective and accessible. This framing is consistent with the evidence that structured counseling support substantially improves linkage to care following HIV self-testing among MSM [5], reinforcing the theoretical rationale for treating educational and counseling content as integral to, rather than separate from, the technology-acceptance constructs examined here.

Figure 1 illustrates the hypothesized relationships among constructs.

XSL•FO
RenderX

Figure 1. Extended UTAUT model for TechAids adoption. Source: adapted from Venkatesh et al [2]. UTAUT: Unified Theory of Acceptance and Use of Technology.



Research Hypotheses

Research hypotheses serve as a cornerstone of empirical research. They provide a focused framework for data collection and analysis and are essential for confirming or refuting theoretical assumptions. Sekaran and Bougie [23] define a research hypothesis as a logically conjectured relationship between 2 or more variables expressed in the form of a testable statement. Developing research hypotheses through UTAUT enables examination of the theoretical connections between different constructs. They establish the direction for data collection and analysis while enabling researchers to validate or adjust the UTAUT model through empirical evidence in various contexts. Based on the specific context of TechAids, Multimedia Appendix 1 summarizes the hypotheses.

Study Design and Setting

This study adopted an explanatory mixed methods design combining quantitative survey data assessing UTAUT constructs with qualitative open-ended responses about platform preferences and concerns. The quantitative component tested an extended UTAUT model using ordinary least squares (OLS) models while the qualitative component used thematic analysis of open-ended responses to triangulate statistical findings and provide contextual interpretation. The UTAUT model created by [2] was used in the study to identify the variables influencing MSM and transgender individuals' acceptance and use of the TechAids app. According to the literature, the UTAUT model, which is used in various communities and research fields, can benefit from varying scale items and sizes. Items were organized into 7 latent constructs aligned with an extended UTAUT framework: PE, EE, SI, FC, trust, PS, and BI, alongside age and education. Constructs were operationalized using multiitem scales adapted from prior technology acceptance and e-health research. All items were rated on a 5-point Likert scale (1=strongly disagree to 5=strongly agree). Explanation was made to discuss the use of the scale to ensure a clear understanding of the items by the researchers before completing the survey.

The TechAids Platform

TechAids is a mobile-optimized digital platform developed through participatory design workshops with Nigerian MSM and transgender communities [24]. The platform offers 4 core functions: (1) teleconsultation scheduling with trained HIV health worker clinicians, (2) confidential medication delivery coordination with discreet packaging, (3) automated appointment and medication reminders, (4) educational resources about HIV treatment and wellness. Security features include end-to-end encryption for all communications and automatic session timeouts. For this study, we developed a functional prototype incorporating these features to allow participants hands-on exploration before completing surveys.

Participants and Recruitment

We used nonprobability convenience sampling in recruiting participants who were accessible through HIV clinics and community service settings facilitated by The Initiative for Equal Rights (TIERs) and the Happiest Ones Health Support & Rights Initiative (HOHSRI), together with online LGBTQ+ (lesbian, gay, bisexual, transgender, queer, and other sexual and gender minorities) networks in Lagos, Nigeria. Eligibility criteria for the user group included (1) age 18 years or older, (2) self-identification as MSM, transgender, or gender diverse, (3) confirmed HIV-positive status or high engagement in HIV prevention services, (4) ownership of or regular access to a

XSL•FO
RenderX

smartphone, and (5) ability to provide informed consent. Recruitment materials emphasized voluntary participation, assured confidentiality, and clarified that declining to participate would not affect access to any health services. Nongovernmental organizations (NGOs) serve as gatekeepers for the recruitment of participants to build trust and minimize coercion concerns.

Criteria (2) and (3) were applied disjunctively ie, participants qualified by self-identifying as MSM, transgender, or gender diverse, or by having confirmed HIV-positive status or high engagement in HIV prevention services. This explains why, as reported in Multimedia Appendix 2, most participants identified as heterosexual and HIV-negative. Health care providers were recruited separately through the same NGOs and were eligible if they currently delivered HIV-related services to MSM or transgender clients. Providers and users were analyzed within a single regression model because both were prospective adopters of the same platform and UTAUT items were worded to apply across both roles. Recruitment proceeded via NGO and clinic staff introductions during routine service contact or outreach events, supplemented by online LGBTQ+ network recruitment. Recruitment source was not recorded, precluding a breakdown by channel. This was noted as a limitation of recruitment tracking.

Data collection and survey instrument: Participants interacted with the TechAids functional prototype and then completed a structured questionnaire comprising 28 Likert-scale items covering PE, EE, SI, FC, trust, PS, and BI. The measurement items are presented in Multimedia Appendix 3. The questionnaire also included 3 open-ended questions: (1) What are the things you like about a mobile app for consultation and medication delivery? (2) What are the things you dislike about using a mobile app for consultation and medication delivery? and (3) What do you think would make TechAids App more useful for people living with HIV and AIDS? A cross-sectional survey was conducted from January 2025 to May 2025.

Ethical Considerations

The Bournemouth University Research Ethics Committee approved this study (Ethics ID 51057). Given the sensitivity of research involving criminalized populations and HIV status, we implemented enhanced confidentiality protections. All data were stored on password-protected servers without personal identifiers and transferred via secure protocols.

Informed consent was obtained from all participants prior to enrollment; consent procedures explained the study's purpose, the voluntary nature of participation, the right to withdraw at any time without affecting access to health services, and the confidentiality protections in place, and were delivered verbally or in writing according to participant preference and literacy. To mitigate risk associated with participation in a criminalized legal context, study staff member (EOO) avoided collecting or recording any information that could independently identify a participant's sexual orientation, gender identity, or HIV status outside the secure survey instrument; prototype testing sessions were conducted in private spaces arranged by partnering NGOs to minimize visibility to others; and participants were not required to disclose sensitive identity or status items, consistent with the "prefer not to say" response option used throughout the survey. Data linking participant identities to survey responses were not retained beyond the minimum required for recruitment logistics, and all analytic data were fully deidentified prior to analysis.

Data Analysis

A priori power analysis using G*Power 3.1 (developed by Franz Faul) [25] was used to determine the minimum sample size. We conducted analyses using Statistical Package for the Social Sciences (SPSS) version 28 (IBM Corp.) [26].

Estimation Approach

OLS regression was used to estimate the relationships between the independent variables and BI to use the digital health platform. The operationalization of each theoretical variable was based on a composite index which consisted in averaging the responses for their corresponding Likert-scale indicators from validated measurement scales. This procedure allowed the constructs to be treated as continuous observed values that are suitable for regression analysis. The reliabilities of the scales were evaluated by Cronbach α during model estimation to examine internal consistency between the components [27]. Descriptive statistics and a Pearson correlation analysis were carried out to examine distributions of variables and preliminary associations.

Treating averaged multiitem Likert composites as approximately continuous for OLS estimation is a widely used convention in technology-acceptance survey research, including prior UTAUT studies in similar low-resource settings [28], and is generally considered acceptable when each composite is derived from four or more reliable indicators, as is the case for all constructs here (Multimedia Appendix 3). BI, like the other UTAUT constructs, reflects an underlying latent variable rather than a strictly interval-level measure, and that ordinal regression or structural equation modeling (SEM) would provide a valuable robustness check on the OLS findings reported below, particularly given the unusually strong association observed between FC and BI. This is a methodological limitation and direction for future analysis.

Composite reliability (CR) = (∑λi)2(∑λi)2+∑(1−λi2)where (lambda) is the standardized factor loading.

Average variance extracted (AVE) = ∑λi2n, where n is the number of indicators per construct

Predictive accuracy was evaluated through the coefficient of determination (R^2), which represents the proportion of variance in BI explained by the predictor variables. The OLS regression equation was estimated using standardized coefficients to facilitate comparison of effect sizes across predictors.

RBI2=1−∑(BIobs−BIpred)2∑(BIobs−BI−obs)2

A priori power analysis conducted using G*Power (effect size f^2=0.15, significance level α=.05, statistical power=0.80, and 8 predictors) indicated a minimum required sample size of 109. The final sample size of 168 therefore exceeded this threshold, confirming adequate statistical power for detecting medium-sized effects.

XSL•FO
RenderX

Measurement of Reliability and Validity of UTAUT Constructs

Reliability analysis was conducted to assess the internal consistency of the items for each latent factor of the UTAUT model before going into the model testing. Cronbach α was applied to ensure that the scale used for rating (Likert scale) is reliable. Multimedia Appendix 3 shows the Cronbach α scores achieved for each factor. An acceptable minimum threshold for Cronbach α is suggested to be 0.7 according to [29,30].

Relationships Among Constructs

Correlation Analysis

The study used correlation analysis to evaluate the strength and direction of the linear relationship between 2 variables. Correlation analysis helps to determine which factors are most strongly associated with users' willingness to adopt and use health technologies. A positive correlation indicates that as 1 variable increases (ie, perceived usefulness), the other (ie, intention to use the app) also increases. The coefficient (ρ) value ranges from −1 to +1. Quantitative analysis techniques like correlation are essential for testing the relationships among these constructs. The Pearson correlation coefficient is used to examine the strength and direction of linear relationships between UTAUT constructs ie,

$$\rho XY = \sum(Xi - \bar{X})\sum(Yi - \bar{Y})\sum(Xi - \bar{X})2\sum(Yi - \bar{Y})2$$

where X_i= scores on construct X (eg, PE); Y_i= scores on construct Y (eg, BI); $\bar{X}$ and $\bar{Y}$ are means of X and Y respectively.

Regression Analysis

OLS regression was used to help determine which constructs significantly impact BI to use the technology. It provides insights into which factors are most influential to enhance technology adoption. Regression analysis predicts the value of a dependent variable (eg, BI) based on independent variables (eg, PE, EE, SI, FC, trust, and PS). We estimated a set of standardized OLS regression models that map directly onto the extended UTAUT structure.

The baseline regression model is specified as follows:

$$BI = \beta 0 + \beta 1 PEi + \beta 2 EEi + \beta 3 SIi + \beta 4 FCi + \beta 5 TRi + \beta 6 PSi + \beta 7 Agei + \beta 8 Edui + \varepsilon i$$

Each construct is a latent variable represented by a set of observed indicators. For example,

$$PEi = \lambda PE1 XPE1,i + \lambda PE2 XPE2,i + \lambda PE3 XPE3,i + \lambda PE4 XPE4,i + \delta PE,i$$

$$EEi = \lambda EE1 XEE1,i + \lambda EE2 XEE2,i + \lambda EE3 XEE3,i + \lambda EE4 XEE4,i + \delta EE,i$$

$$SIi = \lambda SI1 XSI1,i + \lambda SI2 XSI2,i + \lambda SI3 XSI3,i + \lambda SI4 XSI4,i + \delta SI,i$$

$$FCi = \lambda FC1 XFC1,i + \lambda FC2 XFC2,i + \lambda FC3 XFC3,i + \lambda FC4 XFC4,i + \delta FC,i$$

$$TRi = \lambda TR1 XTR1,i + \lambda TR2 XTR2,i + \lambda TR3 XTR3,i + \lambda TR4 XTR4,i + \delta TR,i$$

$$PSi = \lambda PS1 XPS1,i + \lambda PS2 XPS2,i + \lambda PS3 XPS3,i + \lambda PS4 XPS4,i + \delta PS,i$$

$$Agei = \lambda Age1 XAge1,i + \lambda Age2 XAge2,i + \lambda Age3 XAge3,i + \lambda Age4 XAge4,i + \delta PS,i$$

$$Edui = \lambda PS1 XPS1,i + \lambda PS2 XPS2,i + \lambda PS3 XPS3,i + \lambda PS4 XPS4,i + \delta PS,i$$

where λPEi, λEEi, λSIi, λFCi, λTRi, λPSi, are factor loadings, XPEi, XEEi, XSIi, XFCi, XTRi, XPSi, are measured items, and δPE, δEE, δSI, δFC, δTR, δPS represent measurement errors.

Moderation Analysis

To test moderation hypotheses H7 and H8, interaction terms were created between each UTAUT construct and both demographic moderators (age and education). Both moderators were mean-centered before creating product terms to reduce multicollinearity. Eight separate regression models were then estimated, each including the full set of main effects plus one interaction term, consistent with the estimation approach described above.

Categorization of Qualitative Data

Open-ended responses to the 3 qualitative survey questions (Table 1) were analyzed using an inductive, data-driven thematic-categorization approach, following the reflexive thematic analysis approach described by Braun and Clarke [31]. Analysis proceeded in 4 steps: (1) the lead author (EOO) read all open-ended responses in full to gain familiarity with the data; (2) initial descriptive labels were generated inductively from the responses, question by question, rather than applied from a preexisting framework; (3) labels were iteratively grouped and refined into a smaller set of categories per question as additional responses were reviewed with earlier responses rechecked against revised category definitions; and (4) the resulting categories were tabulated against illustrative response counts (Table 1). Coding was conducted by a single coder, EOO (the lead author); no second coder independently coded the data, and consequently, no formal interrater reliability statistic (eg, Cohen kappa) was calculated and no coding discrepancies required resolution through discussion or arbitration. The categorization should therefore be read as a descriptive thematic summary intended to contextualize the quantitative findings, rather than as output from an independently verified, consensus-based qualitative analysis. Independent double-coding, with formal interrater reliability assessment, is identified as a priority for future mixed methods work with this dataset. This is revisited in the Limitations section.

XSL•FO
RenderX

Table . Qualitative feedback from the survey.

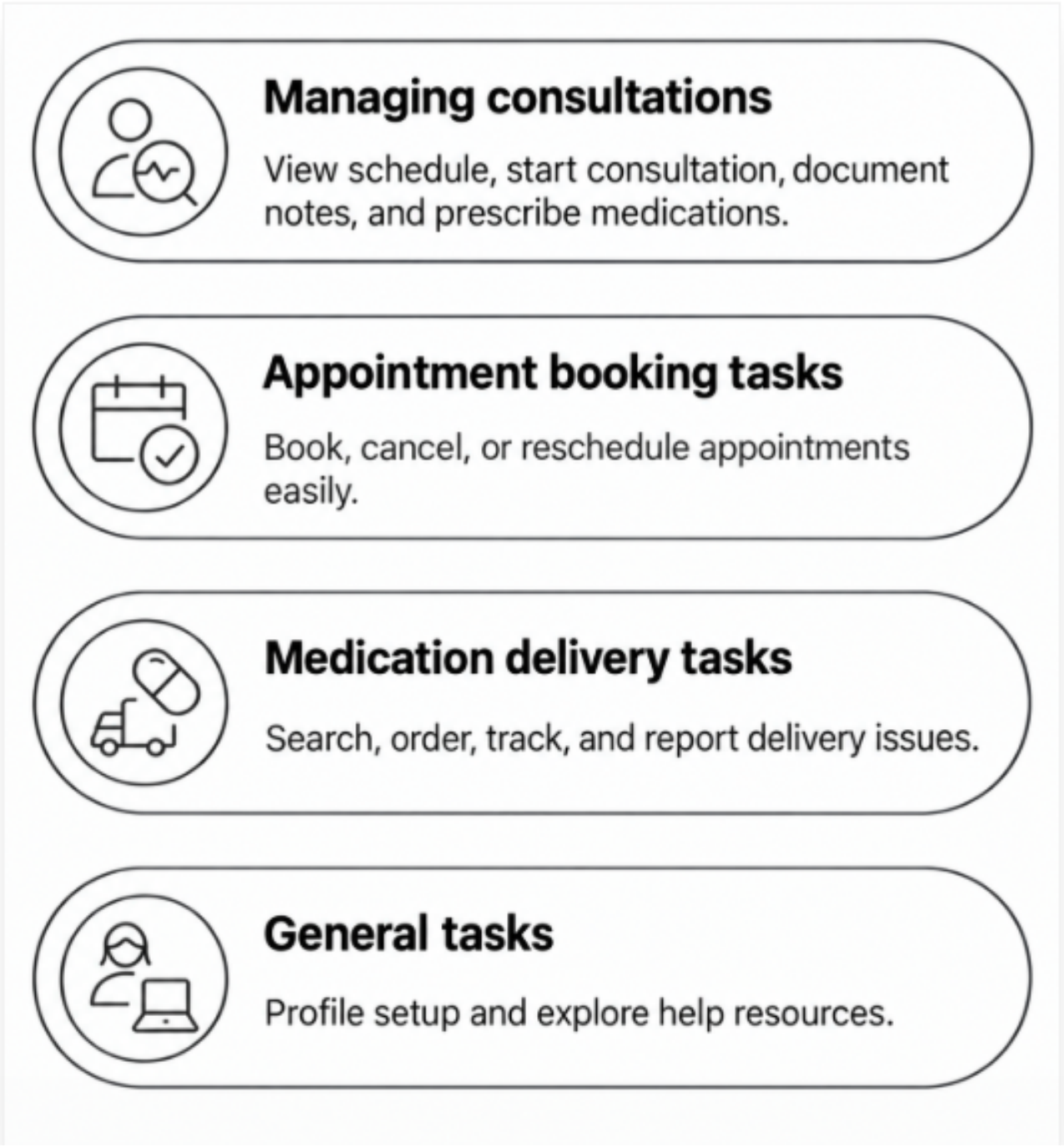
Serial number	Survey questions	Categorization of the survey responses
1.	What are the things you like about a mobile app for consultation and medication delivery?	• Access and convenience. • Personalization and control. • Privacy and security. • Medication management and delivery. • Communication and support. • Health information and education.
2.	What are the things you dislike about using a mobile app for consultation and medication delivery?	• User experience. • Technical performance and app stability. • Access to specialists and services. • Internet dependence and connectivity.
3.	What do you think would make TechAids app more useful for people living with HIV or AIDS?	• Access to educational content • Access to health care providers • End-to-end encryption • Free and fast internet • Offline functionality • Quick helpline access • Regular app update • Smart reminder

Research Participants

The study population of this research consists of 168 respondents who participated in the completion of a survey which consisted of 2 stakeholders ie, health care providers and patients. The respondents that did not complete the survey questions were expunged from the sample. On average, completion of the survey took approximately 32 minutes, which could contribute to the reason some respondents did not complete the survey. The survey focuses on evaluating the willingness to adopt the use of digital health care technologies among MSM and transgender individuals living with HIV/AIDS. The relatively small sample size in this study reflects practical constraints associated with conducting research in underserved communities [32]. The research participants were given some tasks to perform on the TechAids app's prototype. The tasks completed are highlighted in Figure 2 below.

XSL•FO
RenderX

Figure 2. Prototype design tasks.



Results

Overview

Two sets of results from the UTAUT-based survey were presented in this section. First, quantitative analysis, and especially descriptive and inferential analyses, were used to explore key factors influencing the use and acceptance of technology. Second, the qualitative data were analyzed using a categorization of ideas approach, highlighting themes and patterns in participants' perceptions and experiences. Both qualitative and quantitative methods offer a comprehensive understanding of technology adoption in HIV/AIDS care among MSM and transgender people living with HIV/AIDS.

Participant Characteristics

Multimedia Appendix 2 summarizes demographic characteristics. The final sample comprised 141 platform users (patients) and 27 health care providers (n=168). User participants averaged 31.2 (SD 4.7, range 21 - 40) years. Nearly half (n=69, 48.9%) fell in the 31 - to 40-year age bracket, while 51.1% (n=72) were aged 21 - 30 years. Educational attainment was relatively high: 34% (48/141) held master's degrees, 30.5% (43/141) bachelor's degrees, and 35.5% (50/141) diplomas or equivalent technical qualifications.

Gender identity patterns reflected the sensitivity of these questions in the Nigerian context. Among users, 31.2% (44/141) identified as male, 36.2% (51/141) as female, and 32.6%

XSL•FO
RenderX

(46/141) preferred not to disclose. Only 2 participants (2/141, 1.4%) explicitly identified as transgender, though an additional 46.8% (66/141) preferred not to answer this question, suggesting possible transgender identity among nondisclosers. Nearly half (69/141, 48.9%) declined to state sexual orientation, while 49.6% (70/141) identified as heterosexual and 1.4% (2/141) as gay or lesbian. These patterns likely reflect justified caution given Nigeria's criminalization of same-sex relationships.

Regarding HIV status, 54.6% (77/141) of users indicated HIV-negative status (likely reflecting engagement in prevention services), while 45.4% (64/141) declined to disclose, a pattern consistent with pervasive HIV stigma. Health care provider participants (n=27) averaged slightly older (19/27, 70.4% aged 31‐40 years) and showed similar reluctance to disclose personal information on sensitive items.

Reliability Analysis

Consistency of the constructs of the model was measured using Cronbach α. Internal consistency was assessed for each multiitem construct, targeting α≥0.70. Construct validity was evaluated using exploratory factor-analytic logic, together with the interconstruct correlation matrix among the 7 construct means. Factor analysis was used to investigate underlying relationships (called factors) among a set of observed variables. It helps reduce a large number of variables into fewer unobserved variables (factors) that explain the patterns in the data. This was used to generate the factor loadings for the predictors.

All constructs demonstrated acceptable to excellent internal consistency (Multimedia Appendix 3). Cronbach α values ranged from 0.701 (BI) to 0.887 (trust), well exceeding the 0.70 threshold [27]. Trust showed the highest reliability, suggesting that items measuring confidence, security, and confidentiality concerns cohered strongly. Item-total correlations within each construct ranged from 0.52 to 0.84, indicating that individual items contributed meaningfully to their respective scales. Factor loadings from exploratory factor analysis ranged from 0.668 to 0.903, confirming that items loaded strongly on their intended constructs. CR and AVE computed from single-factor loadings extracted per construct via principal axis factoring are reported alongside Cronbach α in Multimedia Appendix 3. All 7 constructs exceeded the conventional CR threshold of 0.70 (range: 0.707 to 0.889). AVE exceeded the conventional 0.50 threshold for 5 of the 7 constructs (SI=0.509, FC=0.561, trust=0.669, PS=0.572), while PE (AVE=0.429), EE (AVE=0.457), and BI (AVE=0.382) fell modestly below this threshold, indicating that convergent validity for these 3 constructs, while supported by adequate CR and Cronbach α was not fully confirmed by AVE alone. This was reported transparently as a measurement limitation in the Limitation and Future Research section rather than treating CR or alpha as sufficient on their own. A Fornell-Larcker discriminant-validity check, comparing the square root of each construct's AVE against its correlations with all other constructs (Table 2), confirmed discriminant validity for all construct pairs except FC and BI: the FC-BI correlation (r=0.802) exceeded both constructs' √AVE (0.749 and 0.618, respectively), indicating a discriminant-validity concern between these 2 constructs specifically. This finding statistically corroborates, rather than merely conceptually motivates, the caution urged elsewhere in this study regarding the magnitude of the FC–BI association and is consistent with the possibility of conceptual or item-level overlap between these 2 constructs that should be addressed through item refinement in future research. A full confirmatory factor analysis was not conducted, given sample size constraints relative to the number of constructs and items; this is noted as a limitation.

XSL•FO
RenderX

Table . Correlation matrix showing relationships between UTAUT[a] constructs and [b]BI to use the Techaids app. The extremely high [c]FC-BI[b] correlation (r=0.802) raised concerns about multicollinearity that we addressed through VI[d]F analysis in regression models.

Constructs	Age	Education	PE[e]	EE[f]	SI[g]	FC[c]	Trust	PS[h]	BI[b]
Age	1	−0.115	−0.035	0.012	0.058	−0.065	0.014	0.087	0.048
Education	−0.115	1	0.013	−0.105	−0.154	−0.016	0.068	−0.027	−0.018
PE	−0.035	0.013	1	−0.033	−0.152	0.004	0.003	0.047	−0.071
EE	0.012	−0.105	−0.033	1	0.092	0.032	0.038	0.203	0.038
SI	0.058	−0.154	−0.152	0.092	1	0.048	0.067	0.028	0.025
FC	−0.065	−0.016	0.004	0.032	0.048	1	−0.088	0.059	0.802
Trust	0.014	0.068	0.003	0.038	0.067	−0.088	1	−0.011	−0.078
PS	0.087	−0.027	0.047	0.203	0.028	0.059	−0.011	1	0.003
BI	0.048	−0.018	−0.071	0.038	0.025	0.802	−0.078	0.003	1

[a]UTAUT: Unified Theory of Acceptance and Use of Technology.

[b]BI: behavioral intention.

[c]FC: facilitating conditions.

[d]VIF: variance inflation factor.

[e]PE: performance expectancy.

[f]EE: effort expectancy.

[g]SI: social influence.

[h]PS: perceived stigma.

Correlation Analysis

Hypotheses were tested using correlation coefficient. This was used to find the linear relationship between 2 constructs. Table 2 shows the correlation coefficients between multiple variables. Each cell in the matrix reflects the strength and direction of the linear relationship between 2 variables using a value between −1 and +1.

Table 2 presents zero-order correlations among all study variables. FC showed a remarkably strong positive correlation with BI (r=0.802; P<.001), a magnitude rarely observed in technology acceptance research. This relationship was notably stronger than correlations between other UTAUT constructs and intention: PE (r=−0.071; P=.37), EE (r=0.038; P=.63), and SI (r=0.025; P=.75) showed negligible associations with intention. Trust correlated negatively with intention (r=−0.078; P=.32), while PS showed essentially zero correlation (r=0.003; P=.97).

Age demonstrated weak positive correlations with BI (r=0.048; P=.54) and PS (r=0.087; P=.26). Education showed small negative correlations with most constructs, but none reached statistical significance. The intercorrelations among UTAUT predictors were generally weak, with 1 notable exception: EE and PS correlated at r=0.203 (P=.008), suggesting that individuals who anticipated greater stigma also perceived the platform as more difficult to use.

Regression Analysis

The relationship between the independent variables X (ie, PE, EE, SI, FC, TR, PS, age, and education) and the dependent variable Y (BI) are shown in Table 3. Each coefficient quantifies how much the dependent variable is expected to increase or decrease when the corresponding independent variable increases by one unit, assuming all other variables are held constant.

Table 3 presents regression results predicting BI. The overall model explained substantial variance: R^2=0.652, $F_{8, 159}$=37.25; P<.001, indicating that our extended UTAUT model accounted for approximately 65% of variation in adoption intentions. This model was estimated on the combined sample of platform users and health care providers (n=168), as justified in Participants and Recruitment section. As a sensitivity check, we reestimated the model restricted to the (n=141) platform users only. The pattern of results closely replicated the combined-sample findings: FC remained the dominant predictor and, if anything, showed a slightly larger standardized effect in the user-only model (β=.856, t_{132}=19.37; P<.001) than in the combined sample (β=.813), and age remained a significant, modestly sized positive predictor (β=.133; P=.003 in the user-only model vs β=.105; P=.03 in the combined sample). PE, EE, SI, trust, PS, and education remained nonsignificant in both models, and the user-only model explained more variance overall (R^2=0.746 vs R^2=0.652 for the combined sample). The significance pattern across all 8 predictors was therefore identical between the combined and user-only samples (Table 4), indicating that including health care providers in the primary analysis did not materially distort the substantive conclusions of this study.

XSL•FO
RenderX

Table . Regression analysis results for UTAUT[a] constructs predicting Techaids app usage intention.

Constructs	Unstandardized β	Standardized β	SE (β)	*t* test (*df*)	Significance level	VIF[b]
Intercept	1.446	0	0.482	2.99 (159)	.003	—[c]
PE[d]	−0.081	−.071	0.053	−1.53 (159)	.13	1.029
EE[e]	0.024	.023	0.050	0.48 (159)	.63	1.064
SI[f]	−0.032	−.030	0.050	−0.63 (159)	.53	1.065
FC[g]	0.720	.813[h]	0.041	17.47[h] (159)	<.001[h]	1.020[h]
Trust	−0.007	−.008	0.043	−0.16 (159)	.87	1.022
PS[i]	−0.056	−.055	0.049	1.156 (159)	.25	1.059
Age	0.014	.105[h]	0.006	2.253[h] (159)	.03[h]	1.030[h]
Education	0.003	.005	0.023	0.108 (159)	.91	1.053

[a]UTAUT: Unified Theory of Acceptance and Use of Technology.

[b]VIF: variance inflation factor.

[c]Not applicable.

[d]PE: performance expectancy.

[e]EE: effort expectancy.

[f]SI: social influence.

[g]FC: facilitating conditions.

[h]*P*<.05.

[i]PS: perceived stigma.

Table . Comparison of standardized regression coefficients, combined sample (n=168) vs platform-users-only subsample (n=141). R^2=0.652 (combined) vs 0.746 (users-only).

Predictor	Combined β (n=168)	*P* values	Users-only β (n=141)	*P* values
PE[a]	−.071	.13	−.038	.40
EE[b]	.023	.63	.035	.45
SI[c]	−.030	.53	−.027	.55
FC[d]	.813	<.001	.856	<.001
Trust	−.008	.87	−.026	.56
PS[e]	−.055	.25	−.051	.27
Age	.105	.03	.133	.003
Education	.005	.91	−.039	.39

[a]PE: performance expectancy.

[b]EE: effort expectancy.

[c]SI: social influence.

[d]FC: facilitating conditions.

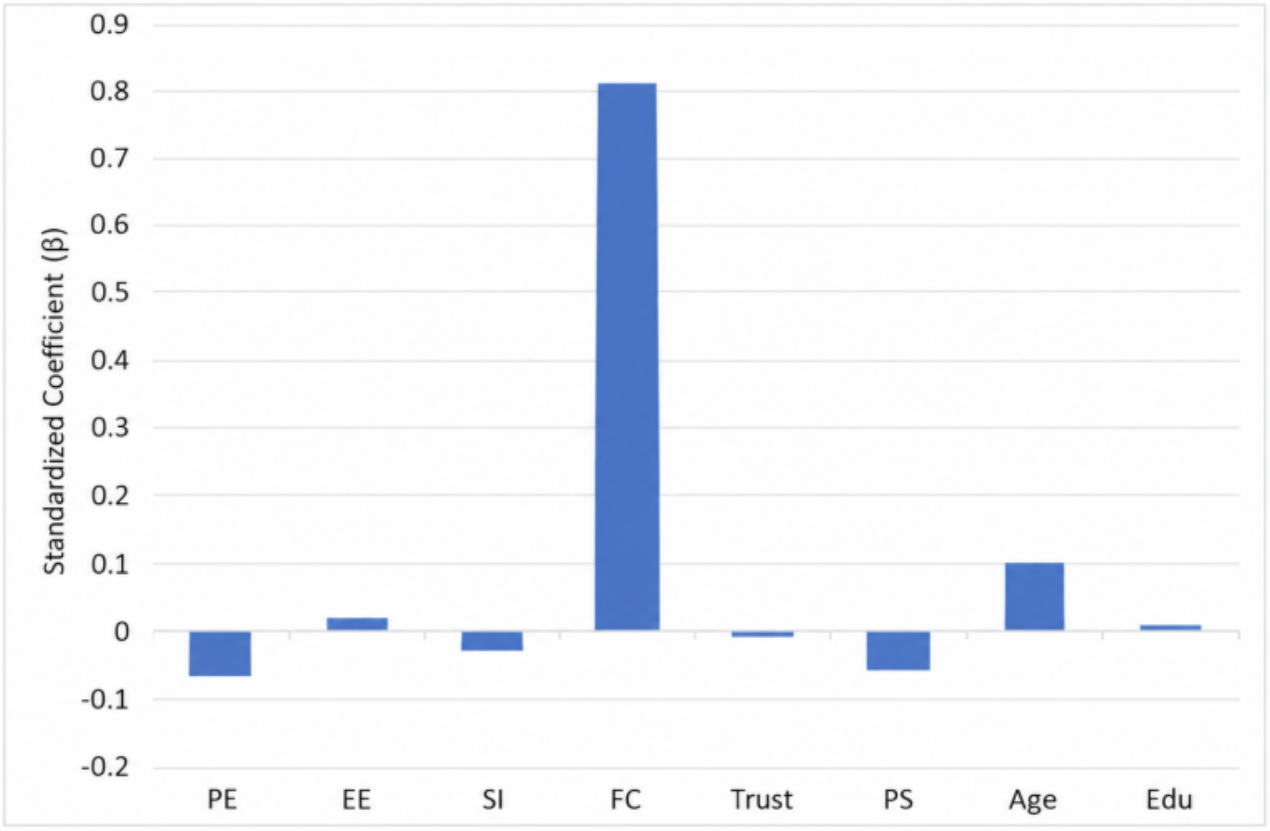
[e]PS: perceived stigma.

In the regression analysis, FC emerged as the overwhelmingly dominant predictor (β=.813; SE=.041; t_{159}=17.47; *P*<.001) (Figure 3). The standardized coefficient magnitude indicates that a one-standard-deviation increase in perceived FC corresponded to a .720 standard deviation increase in intention to adopt TechAids—an exceptionally large effect. Age showed a small but statistically significant positive effect (β=.105, SE=.006, t_{159}=2.253; *P*=.03), suggesting that older participants within our age range expressed slightly greater adoption intention.

Contrary to our hypotheses, PE (β=−.071; *P*=.13), EE (β=.023; *P*=.63), SI (β=−.030; *P*=.53), trust (β=−.008; *P*=.87), and PS (β=−.055; *P*=.25) did not significantly predict BI. Education also showed no effect (β=.005; *P*=.91).

XSL•FO
RenderX

Figure 3. Standardized beta (β) of UTAUT constructs predicting intention to use the TechAids App. EE: effort expectancy; FC: facilitating conditions; PE: performance expectancy; PS: perceived stigma; SI: social influence; UTAUT: Unified Theory of Acceptance and Use of Technology.

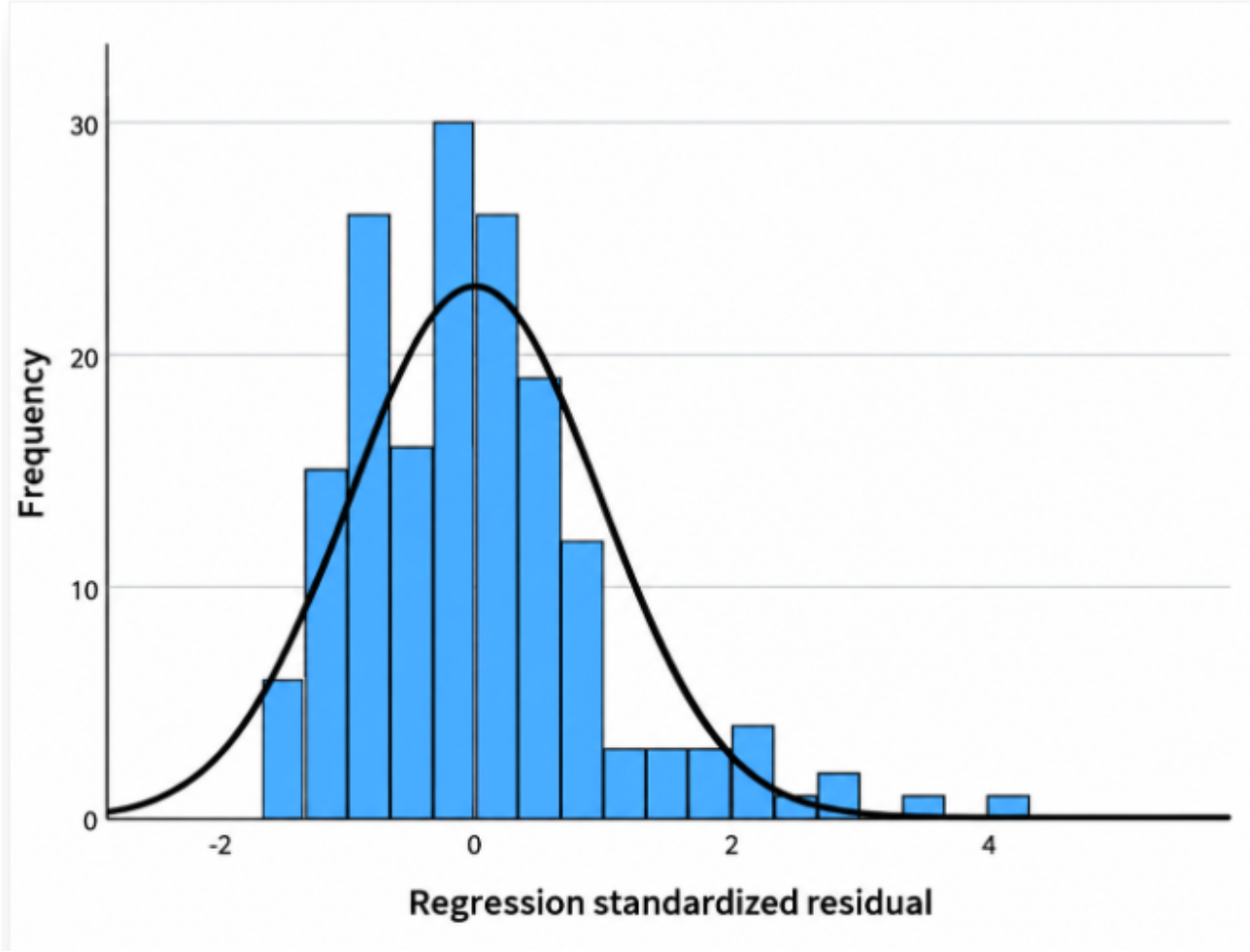


We examined variance inflation factors (VIFs) to assess multicollinearity, which indicated no cause for concern. FC showed VIF=1.02, well below conventional thresholds of 5, indicating that its strong effect was not attributable to collinearity with other predictors. All other VIF values ranged from 1.022 to 1.065 (Table 3), suggesting minimal shared variance among predictors (independent variables). Tolerance values, the reciprocal of VIF, confirmed this pattern (range 0.939 - 0.980) with all values well above the conventional concern threshold of 0.10. The Durbin-Watson statistic was 1.782, within the conventionally acceptable range of 1.5 to 2.5, indicating no evidence of autocorrelation among residuals. Case-level influence diagnostics likewise raised no concerns: the Cook distance ranged from approximately 0 to 0.140 (mean 0.006, SD 0.014) with no case exceeding the threshold of 1.0 conventionally used to flag seriously influential observations; 8 cases (4.8%) exceeded the more conservative screening threshold of 4 /n, a proportion consistent with chance variation at this sample size. Leverage (hat) values ranged up to 0.192 (mean 0.054, SD=0.026), with 8 cases (4.8%) exceeding the 2(k+1)/n screening threshold; critically, no case was flagged by both the Cook distance and leverage simultaneously, indicating that no single observation was both high-leverage and disproportionately influential on the fitted model. Together, these diagnostics indicate that the regression results are not attributable to a small number of unduly influential cases.

Residual diagnostics revealed no substantial violations of OLS assumptions. Residuals appeared approximately normally distributed, as indicated by inspection of the standardized residual histogram (Figure 4), and residuals-versus-fitted plots showed no systematic patterns suggestive of heteroscedasticity (Figure 5). While the low VIF for FC (1.020) rules out collinearity with other predictors as an explanation for its dominant effect, it does not, on its own, rule out conceptual overlap between the FC and BI item sets, or common-method variance arising from collecting all constructs via a single self-report instrument at a single timepoint. A comparison of item wording (Multimedia Appendix 3) shows the 2 construct sets to be conceptually distinct. FC items ask about resources, knowledge, compatibility, and support, while BI items ask about future usage, recommendation, and reliance, but we cannot rule out common-method bias as a contributor to the magnitude of the observed association given the single-source, single-timepoint design. We formally examined this possibility using Harman's single-factor test, entering all 28 items across the 7 constructs into an unrotated principal component analysis. Seven components had eigenvalues greater than 1.0, consistent with the intended 7-construct structure, and the first, largest unrotated factor accounted for only 16.0% of total variance, well below the 50% threshold conventionally used to flag a common-method-variance concern. This result indicates that common-method bias is unlikely to be a major contributor to the magnitude of the FC -BI association; the Fornell-Larcker discriminant-validity violation between these 2 constructs reported in the Reliability Analysis section, rather than common-method variance, is therefore the more likely explanation for the strength of this association.

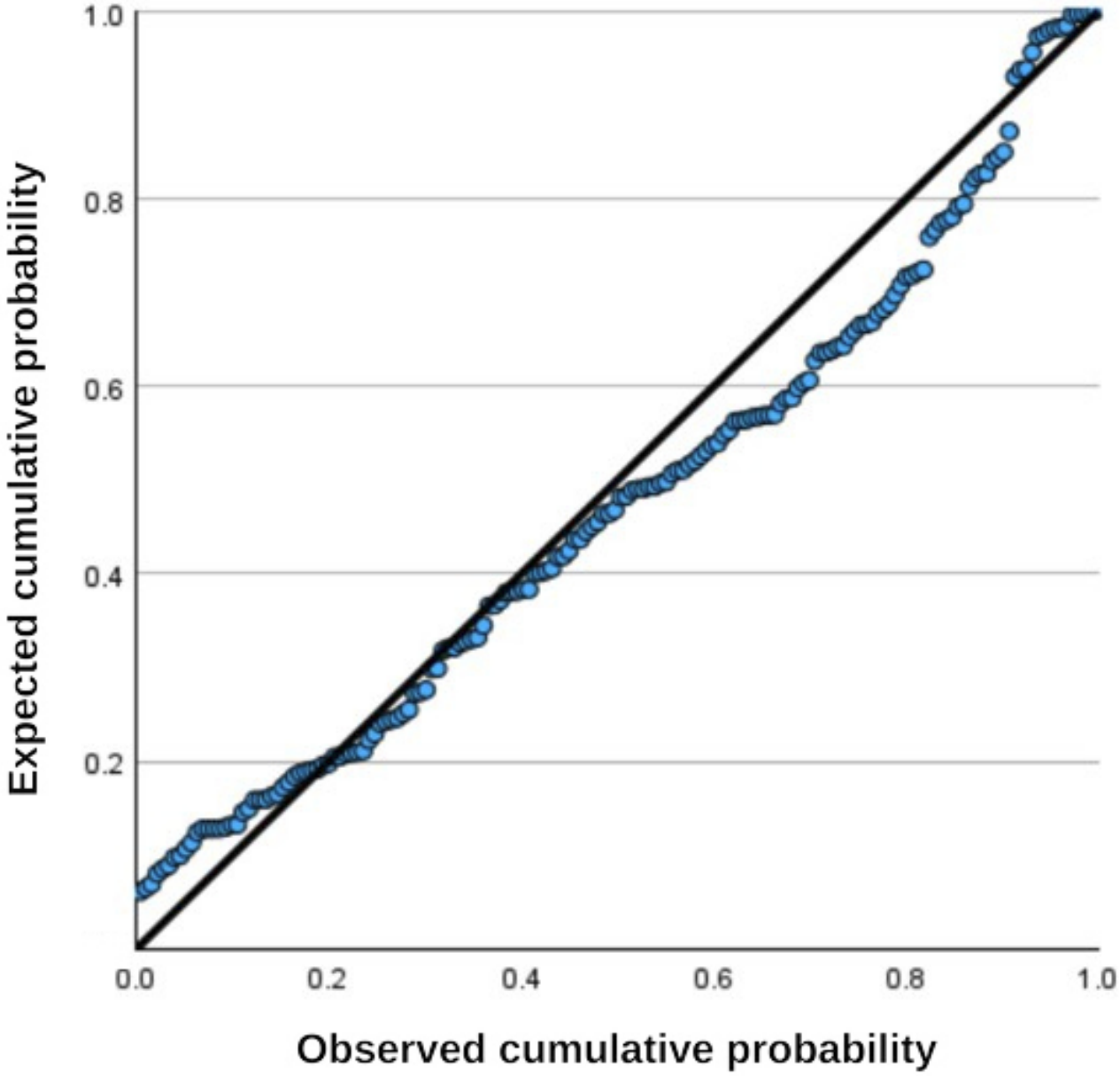
XSL•FO
RenderX

Figure 4. Histogram of standardized residuals for the OLS regression model. OLS: ordinary least squares.



XSL•FO
RenderX

Figure 5. Normal probability-probability plot of regression standardized residual.



Moderation Analysis

Using the 8 interaction models described in the Methods section, results indicated no significant moderation. All interaction coefficients had *P*-values exceeding 0.10. For example, the Age × FC interaction yielded β=−.002 (SE=.009, *P*=.83) (Table 5), suggesting that FC's strong positive effect on BI operates equivalently across younger and older participants. Similar nonsignificant patterns emerged for all other interactions. H7 and H8 were therefore not supported: neither age nor education moderated the relationship between any UTAUT construct and BI. This is distinct from the significant main effect of age on BI reported in the Regression Analysis section (β=.105; *P*=.03); that finding indicates older participants reported somewhat higher intention overall, not that age altered the strength of any UTAUT relationship, and the 2 findings should not be conflated.

XSL•FO
RenderX

Table . Interaction effects of age and education on UTAUT[a] constructs predicting [b]BI.

Interaction term	β_3	SE	*t* test (*df*)	*P* values	Interpretation
PE[c]×Age	.008	.012	0.67 (158)	.51	Not significant
EE[d]×Age	−.003	.011	−0.27 (158)	.79	Not significant
SI[e]×Age	.015	.013	1.15 (158)	.25	Not significant
FC[f] ×Age	−.002	.009	−0.22 (158)	.83	Not significant
PE[g]×Education	−.021	.031	−0.68 (158)	.50	Not significant
EE×Education	.018	.028	0.64 (158)	.52	Not significant
SI×Education	−.012	.030	−0.40 (158)	.69	Not significant
FC×Education	.007	.025	0.28 (158)	.78	Not significant

[a]UTAUT: Unified Theory of Acceptance and Use of Technology.

[b]BI: behavioral intention.

[c]PE: performance expectancy.

[d]EE: effort expectancy.

[e]SI: social influence.

[f]FC: facilitating conditions.

[g]PE: perceived stigma.

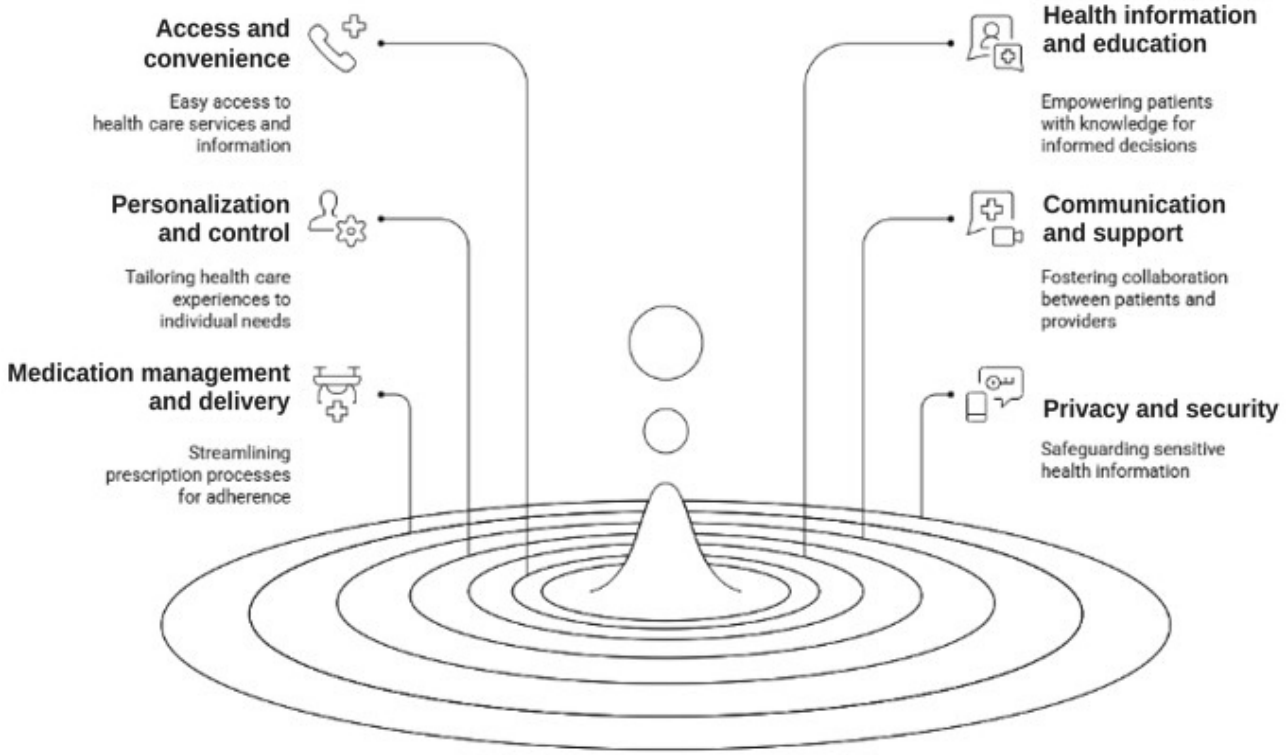
Qualitative Findings

Table 1 presents the key themes identified from the qualitative feedback, categorized using the approach described in the Categorization of Qualitative Data subsection of the Methods section. Open-ended responses were collected for 3 survey questions (Table 1). Of the total sample (n=168), 168 participants responded to Question 1 (things liked about mHealth apps); 162 participants responded to Question 2 (things disliked about mHealth apps); and 157 participants responded to Question 3 (what would make TechAids more useful for people living with HIV and AIDS). Not all participants provided feedback to every open-ended question.

Several aspects the users appreciate about using mobile apps for consultation and medication delivery were highlighted. The users value the ability to reach health care providers anytime in terms of accessibility and convenience. It will afford them the opportunity to manage care from their preferred location. Features like reminders for appointments and medication intake are highly appreciated through personalization and control on the app. Users feel secure through strong end-to-end privacy and security measures put in place. Timely alerts and easy access to prescriptions help users adhere to their medication intake and stay on track with their treatment. Users are empowered to better understand and manage their health condition through access to reliable health information and education on the app. Communication and support roles leave users feeling supported and informed, especially in managing their ongoing care needs (Figure 6).

XSL•FO
RenderX

Figure 6. Users' perceived benefits of a mobile app-based platform for consultation and medication delivery



Four key areas of dissatisfaction were identified by users of mobile apps for consultation and medication delivery. User experience during face-to-face consultation is always limited to the allotted time so as to give other opportunities to be attended to. This will only allow limited interaction time and a lack of personal connection with health care providers, leaving users feeling unheard and unsatisfied. Users submitted that poor technical performance and app instability, including frequent crashes, freezing, lagging, and delayed responses disrupt smooth usage and undermine trust in the app. Another major issue pointed out is access to specialists and services, as users often face long waiting times and difficulty finding relevant HIV and AIDS expertise. Dependence on internet connectivity was identified as a significant barrier, as the system will record failure in the use of the app without internet connectivity. The users in low-network or rural areas will find it difficult to receive timely care and support.

The survey findings suggest 8 key areas of improvement to make the mobile app more useful for MSM and transgender individuals living with HIV/AIDS in Nigeria. Users want offline functionality and free internet, which will allow them to use essential features without data or network glitches. Privacy and data security are essential as users expect their sensitive information to be well secured and protected. There is a strong demand for quick helpline support and access to health care providers at any point in time, most especially during emergencies. They also need health education and engagement through regularly updated app content and smart reminders (Figure 7).

XSL•FO
RenderX

Figure 7. Top recommendations to make TechAids app more useful (authors compilation).



Discussion

Principal Findings

This study applied an extended UTAUT model to understand acceptance and intended use of TechAids among MSM and transgender individuals in Nigeria. Using construct-level data and a detailed regression specification, we produced a robust mathematical model of BI and visualized the relative importance of different predictors. The results converge on a clear message that FC is the dominant driver of intention to adopt TechAids. When respondents perceive that they have the necessary resources, infrastructure, and support, they are highly likely to intend to use the platform; when these conditions are absent or fragile, even strong beliefs about usefulness and ease of use do not translate into intention. The nonsignificance of PE, EE, and SI contrasts with much of the prior UTAUT and mHealth literature, warrants more critical interpretation than setting alone can provide. Three contextual explanations, not mutually exclusive, are worth distinguishing. First, infrastructural dominance may be statistically suppressing the explanatory contribution of other predictors: with FC explaining the large majority of variance in intention (r=0.802 with BI), the remaining variance available for PE, EE, and SI to predict may simply be too small to detect with this sample size, even if these constructs retain genuine, smaller, theoretical importance. Second, measurement-related explanations are plausible: high nondisclosure rates on sensitive items may have restricted variability not only in trust and PS but, indirectly, in the broader pattern of responses, attenuating observed associations. Third, sample-characteristic explanations are relevant given that the analytic sample combined platform users and health care providers with potentially different bases for evaluating usefulness, ease of use, and social influence. This was addressed directly in the Regression Analysis section via a user-only sensitivity check. We do not interpret these findings as evidence that PE, EE, and SI are unimportant in principle, but rather that, in this resource-constrained and highly stigmatized setting, their influence may be empirically overshadowed by more immediate infrastructural and measurement constraints.

The prominence of FC is consistent with findings from other developing-country contexts where access to devices, connectivity, and support is uneven [28]. In such settings, abstract beliefs about usefulness can be easily overridden by concrete constraints: intermittent data, shared phones, unreliable electricity, or lack of technical help. Our FC items capture perceptions of device availability, internet access, compatibility, and support, all of which proved critical for respondents considering a digital HIV app. From an implementation perspective, this means that digital health tools cannot be divorced from their infrastructural environment. The small but significant effect of age suggests that older respondents within the 21 - 40 year range are modestly more inclined to adopt TechAids. Older participants may have longer histories of engagement with HIV care and may view digital tools as a pragmatic way to manage frequent appointments and medication refills. This finding echoes broader work on technology adoption in households, where life cycle and accumulated responsibilities often shape perceived benefits [33].

Trust and PS did not emerge as significant predictors of BI in the regression model, but this should not be taken to mean that they are unimportant. Qualitative work and broader HIV

XSL•FO
RenderX

literature emphasize that trust and stigma are central to health-seeking behavior among MSM and transgender communities [34,35]. Qualitatively, users explicitly valued "Privacy & Security" as a feature they liked about consultation and medication-delivery apps, and identified "End-to-end encryption" as a top recommendation for making TechAids more useful (Table 1), both directly relevant to trust and PS. These concerns surfaced spontaneously and prominently in open-ended responses, while the corresponding quantitative constructs were not significant predictors of intention, suggesting this is a measurement and variance issue rather than evidence that trust and stigma are genuinely unimportant to this population. Several factors plausibly explain the gap: high rates of nondisclosure on sensitive identity and HIV-status items (Participant Characteristics) likely restricted the variability available to detect trust and PS effects in the regression model; and because recruitment proceeded through trusted NGO gatekeepers under assured confidentiality conditions (Participants and Recruitment section), participants who agreed to take part may represent a subgroup with comparatively higher baseline trust and lower perceived risk than the broader key population, further compressing variance on these constructs.

For designers, implementers, and policymakers working on digital health for key populations in low-resource settings, the findings point to several priorities:

1. Invest in FC, not just app design: device support, zero, or low-rated data packages, offline functionality, and local helpdesks or digital navigators should be treated as core components of the intervention, not optional extras.
2. Design for infrastructural fragility: features such as low-bandwidth modes, delayed synchronization, SMS fallbacks, and robust error-handling can ensure continuity of care even when connectivity is intermittent.
3. Interpret the age effect cautiously: age showed a small but significant positive main effect on BI, and older participants in our sample expressed modestly higher adoption intention overall. However, moderation analysis found no evidence that age altered the strength of any UTAUT relationship. The data therefore do not support differentiated, age-targeted UTAUT-based messaging strategies, and any age-related design choices should be informed by further targeted research rather than this finding alone.

Limitation and Future Research

Beyond methodological considerations already discussed, several substantive limitations warrant acknowledgment. Our study examined a single country context (Nigeria) which may not be generalizable to other developing or developed contexts where infrastructural, cultural, or policy conditions differ. Small sample size (N=168) is another limitation of this study which may restrict the statistical power of the analysis and the ability to detect subtle relationships among the UTAUT constructs. Future research could extend this work by collecting larger, multisite samples to strengthen the robustness and combining longitudinal usage data from the deployed TechAids platform with survey measures. The cross-sectional design precludes causal inference, and BI may not reflect sustained real-world use. Analytically, OLS regression of composite Likert constructs should be supplemented by SEM or ordinal regression. AVE fell below the 0.50 threshold for 3 constructs, leaving convergent validity unconfirmed, and a discriminant-validity concern exists between FC and BI. Full Confirmatory Factor Analysis was precluded by sample size. Although Harman's test found no pervasive common-method bias, single-instrument, single-timepoint measurement cannot fully rule it out; temporally separated measurement of key constructs is recommended.

Combining platform users (n=141) and providers (n=27) in 1 model limits subgroup comparisons; a larger provider sample is needed for stratified analysis. High nondisclosure rates on gender, sexual orientation, and HIV status (33% - 49%) constrain confidence in sample representativeness. Qualitative coding by a single author (EOO) without interrater reliability checks limits robustness. Selection bias through NGO and clinic gatekeepers likely underrepresents the most marginalized individuals, while social desirability bias may affect sensitive disclosures. Finally, use of a prototype rather than the live platform may not capture real-world barriers or sustained engagement challenges.

Conclusion

This study examined factors predicting intention to adopt TechAids, a digital HIV care platform, among MSM and transgender individuals in Nigeria. Using an extended UTAUT framework, we found that FC, perceptions of internet access, device compatibility, technical knowledge, and support availability emerged as the overwhelmingly dominant predictor of adoption intentions, explaining approximately 65% of variance. Traditional UTAUT constructs (PE, EE, SI) and our theoretical extensions (trust, PS) showed no significant effects, though qualitative data suggested these factors remain conceptually important.

These findings challenge assumptions underlying much digital health discourse. Technology acceptance discussions frequently emphasize cognitive factors, usefulness, ease of use, features, treating infrastructure as merely enabling background conditions. Our results suggest that in resource-constrained settings, this hierarchy inverts: infrastructure becomes foreground, overwhelming cognitive considerations. Well-designed platforms with privacy-enhancing features may generate enthusiasm among potential users, but this enthusiasm translates into adoption only when individuals possess reliable connectivity, adequate devices, and confidence they can access technical support when needed. These findings describe a strong cross-sectional association between perceived FC and BI to adopt TechAids, not a demonstrated causal effect of infrastructure on actual platform use or downstream health outcomes. FC being strongly associated with intention does not, on its own, prove that infrastructure alone determines whether people go on to use the platform or experience improved HIV care outcomes; verifying that pathway requires the longitudinal and real-world usage data called for in Limitation and Future Research section.

For digital health implementers, these findings carry clear implications: infrastructure support must be treated as core intervention components, not optional supplements. Zero-rated

XSL•FO
RenderX

data plans, offline functionality, low-bandwidth optimization, device loan programs, and community-based technical support infrastructure should be budgeted alongside platform development. Without these enabling foundations, even the most thoughtfully designed platforms will fail to achieve meaningful adoption and sustained engagement.

Acknowledgments

We extend deep gratitude to the men who have sex with men (MSM) and transgender individuals who participated in this research, sharing their time and insights despite risks associated with disclosure in Nigeria's criminalized context. We thank the nongovernmental organizations (NGOs) that facilitated recruitment while protecting participant safety: The Initiative for Equal Rights (TIERs); and Happiest Ones Health Support & Rights Initiative (HOHSRI). We used the generative AI tools (ChatGPT and Anthropic-Claude) for language refinement, grammar editing, and visual generation of figures and diagrams. All AI-assisted outputs were critically reviewed.

Funding

The authors declared no financial support was received for this work.

Data Availability

The datasets generated and analyzed during this study are not publicly available due to the sensitive nature of data involving HIV status and criminalized populations. Deidentified data may be available from the corresponding author upon reasonable request and with appropriate ethical approval.

Authors' Contributions

Conceptualization: EOO (lead), FA (equal), HD (equal), NJ (equal)
Data curation: EOO
Editing: FA (lead), HD (equal), NJ (equal)
Formal analysis: EOO
Methodology: EOO
Reviewing: FA (lead), HD (equal), NJ (equal)
Supervision: FA (lead), HD (equal), NJ (equal)
Visualization: EOO
Writing – original draft: EOO

Conflicts of Interest

None declared.

Multimedia Appendix 1
Research hypotheses and assumptions.
[DOCX File, 16 KB - humanfactors_v13i1e93129_app1.docx]

Multimedia Appendix 2
Demographics table.
[DOCX File, 17 KB - humanfactors_v13i1e93129_app2.docx]

Multimedia Appendix 3
Reliability and average variance extracted (AVE) by construct.
[DOCX File, 22 KB - humanfactors_v13i1e93129_app3.docx]

References

1. Pedroso B, Pilatti LA, de Francisco AC, dos Santos CB. Quality of life assessment in people with HIV: analysis of the WHOQOL-HIV syntax. AIDS Care 2010 Mar;22(3):361-372. [doi: 10.1080/09540120903111502] [Medline: 20390517]
2. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view1. MIS Q 2003 Sep 1;27(3):425-478. [doi: 10.2307/30036540]
3. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172. [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]

XSL•FO
RenderX

4. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367. [doi: 10.2196/jmir.8775] [Medline: 29092808]
5. Chen S, Fang Y, Chan PSF, Kawuki J, Mo P, Wang Z. Counseling supporting HIV self-testing and linkage to care among men who have sex with men: systematic review and meta-analysis. JMIR Public Health Surveill 2024 Jan 24;10(1):e45647. [doi: 10.2196/45647] [Medline: 38265866]
6. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52. [doi: 10.2196/jmir.3951] [Medline: 25803266]
7. Kruk ME, Gage AD, Arsenault C, et al. High-quality health systems in the sustainable development goals era: time for a revolution. Lancet Glob Health 2018 Nov;6(11):e1196-e1252. [doi: 10.1016/S2214-109X(18)30386-3] [Medline: 30196093]
8. Lupton D. Critical perspectives on digital health technologies. Sociol Compass 2014 Dec;8(12):1344-1359. [doi: 10.1111/soc4.12226]
9. Greenhalgh T, Vijayaraghavan S, Wherton J, et al. Virtual online consultations: advantages and limitations (VOCAL) study. BMJ Open 2016 Jan 29;6(1):e009388. [doi: 10.1136/bmjopen-2015-009388] [Medline: 26826147]
10. Chen S, Yu FY, Fang Y, et al. Chatbot-delivered real-time support to improve HIV self-testing rates: a randomized clinical trial. JAMA Netw Open 2025 Nov 3;8(11):e2544821. [doi: 10.1001/jamanetworkopen.2025.44821] [Medline: 41284298]
11. Shibl R, Lawley M, Debuse J. Factors influencing decision support system acceptance. Decis Support Syst 2013 Jan;54(2):953-961. [doi: 10.1016/j.dss.2012.09.018]
12. Rho MJ, Choi IY, Lee J. Predictive factors of telemedicine service acceptance and behavioral intention of physicians. Int J Med Inform 2014 Aug;83(8):559-571. [doi: 10.1016/j.ijmedinf.2014.05.005] [Medline: 24961820]
13. Dwivedi YK, Rana NP, Jeyaraj A, Clement M, Williams MD. Re-examining the Unified Theory of Acceptance and Use of Technology (UTAUT): towards a revised theoretical model. Inf Syst Front 2019 Jun;21(3):719-734. [doi: 10.1007/s10796-017-9774-y]
14. Katz IT, Ryu AE, Onuegbu AG, et al. Impact of HIV‐related stigma on treatment adherence: systematic review and meta‐synthesis. J Int AIDS Soc 2013 Nov;16(3S2):18640. [doi: 10.7448/IAS.16.3.18640] [Medline: 24242258]
15. Logie CH, James L, Tharao W, Loutfy MR. HIV, gender, race, sexual orientation, and sex work: a qualitative study of intersectional stigma experienced by HIV-positive women in Ontario, Canada. PLoS Med 2011 Nov;8(11):e1001124. [doi: 10.1371/journal.pmed.1001124] [Medline: 22131907]
16. Turan B, Hatcher AM, Weiser SD, Johnson MO, Rice WS, Turan JM. Framing mechanisms linking HIV-related stigma, adherence to treatment, and health outcomes. Am J Public Health 2017 Jun;107(6):863-869. [doi: 10.2105/AJPH.2017.303744] [Medline: 28426316]
17. Bauermeister JA, Pingel ES, Jadwin-Cakmak L, et al. Acceptability and preliminary efficacy of a tailored online HIV/STI testing intervention for young men who have sex with men: the Get Connected! program. AIDS Behav 2015 Oct;19(10):1860-1874. [doi: 10.1007/s10461-015-1009-y] [Medline: 25638038]
18. Schnall R, Rojas M, Bakken S, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251. [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
19. McKnight DH, Choudhury V, Kacmar C. Developing and validating trust measures for e-commerce: an integrative typology. Inf Syst Res 2002 Sep;13(3):334-359. [doi: 10.1287/isre.13.3.334.81]
20. Chib A, van Velthoven MH, Car J. mHealth adoption in low-resource environments: a review of the use of mobile healthcare in developing countries. J Health Commun 2015;20(1):4-34. [doi: 10.1080/10810730.2013.864735] [Medline: 24673171]
21. Sullivan PS, Khosropour CM, Luisi N, et al. Bias in online recruitment and retention of racial and ethnic minority men who have sex with men. J Med Internet Res 2011 May 13;13(2):e38. [doi: 10.2196/jmir.1797] [Medline: 21571632]
22. Opoku MP, Torgbenu EL, Nketsia W, Mprah WK, Edusei AK. Living with HIV in Ghana: disclosure of status and experiences of persons receiving antiretroviral therapy. Afr J AIDS Res 2021 Oct;20(3):204-213. [doi: 10.2989/16085906.2021.1964546] [Medline: 34517791]
23. Oluokun EO, Adedoyin FF, Dogan H, Jiang N. Co-designing digital health intervention for monitoring medication and consultation among transgender people in underserved communities: collaborative approach. JMIR Hum Factors 2024 Sep 12;11(1):e45826. [doi: 10.2196/45826] [Medline: 39264700]
24. Sekaran U, Bougie R. Research Methods for Business: A Skill Building Approach, th edition: John Wiley & Sons; 2016. URL: https://librarysearch.kcl.ac.uk/discovery/fulldisplay/alma990609877790206881/44KCL_INST:44KCL_INST [accessed 2026-09-29]
25. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
26. Corp IBM. IBM SPSS Statistics for Windows, Version 280: IBM Corp; 2021. URL: https://www.ibm.com/docs/en/SSLVMB_28.0.0/pdf/IBM_SPSS_Statistics_Brief_Guide.pdf [accessed 2026-09-29]
27. Nunnally J, Bernstein IH. Psychometric Theory, 3rd edition 1994. URL: https://books.google.co.in/books/about/Psychometric_Theory_3E.html?id=_6R_f3G58JsC&redir_esc=y [accessed 2026-09-29]

XSL•FO
RenderX

28. George D, Mallery P. SPSS for Windows Step by Step: A Simple Guide and Reference, th edition 2003. URL: https://books.google.co.in/books/about/SPSS_for_Windows_Step_by_Step.html?id=KS1DPgAACAAJ&redir_esc=y [accessed 2026-09-29]
29. Field AP. Discovering Statistics Using SPSS, 2nd edition 2005. URL: https://www.scribd.com/document/408460799/Discovering-Statistics-Using-SPSS-2nd-Edition-Andy [accessed 2026-09-29] [doi: 10.1348/000709906X100611]
30. Bonevski B, Randell M, Paul C, et al. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. BMC Med Res Methodol 2014 Mar 25;14(1):42. [doi: 10.1186/1471-2288-14-42] [Medline: 24669751]
31. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
32. Gu D, Khan S, Khan IU, et al. Assessing the adoption of e-Health technology in a developing country: an extension of the UTAUT model. Sage Open 2021 Jul;11(3):21582440211027565. [doi: 10.1177/21582440211027565]
33. Brown SA, Venkatesh V. Model of adoption of technology in households: a baseline model test and extension incorporating household life cycle1. MIS Q 2005 Sep 1;29(3):399-426. [doi: 10.2307/25148690]
34. Berger BE, Ferrans CE, Lashley FR. Measuring stigma in people with HIV: psychometric assessment of the HIV stigma scale. Res Nurs Health 2001 Dec;24(6):518-529. [doi: 10.1002/nur.10011] [Medline: 11746080]
35. Oluokun EO, Adedoyin FF, Dogan H, Jiang N. Digital interventions for managing medication and health care service delivery in West Africa: systematic review. J Med Internet Res 2024 Oct 9;26:e44294. [doi: 10.2196/44294] [Medline: 39383531]

Abbreviations

AVE: average variance extracted
BI: behavioral intention
CR: composite reliability
EE: effort expectancy
FC: facilitating conditions
HOHSRI: Happiest Ones Health Support and Rights Initiative
LGBTQ+: lesbian, gay, bisexual, transgender, queer, and other sexual and gender minorities
mHealth: mobile health
MSM: men who have sex with men
NGO: nongovernmental organization
OLS: ordinary least squares
PE: performance expectancy
PS: perceived stigma
SEM: structural equation modeling
SI: social influence
SPSS: Statistical Package for the Social Sciences
UTAUT: Unified Theory of Acceptance and Use of Technology
VIF: variance inflation factor

Edited by A Kushniruk; submitted 10.Feb.2026; peer-reviewed by M Chakit, S Chen; revised version received 21.Aug.2026; accepted 21.Aug.2026; published 06.Oct.2026.

Please cite as:
Oluokun EO, Adedoyin FF, Dogan H, Jiang N
Evaluating Acceptance and Use of Digital Health Platforms for Consultation and Medication Delivery Among Men Who Have Sex With Men and Transgender People Living With HIV: Cross-Sectional Survey Informed by an Extended UTAUT Approach
JMIR Hum Factors 2026;13:e93129
URL: https://humanfactors.jmir.org/2026/1/e93129
doi:10.2196/93129

© Emmanuel Oluwatosin Oluokun, Festus Fatai Adedoyin, Huseyin Dogan, Nan Jiang. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 6.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Physician-Patient Communication and Physicians' Acceptance of a Tailored Digital Health Information Service: Quantitative Online Survey

Paula Memenga[1], MA; Elena Link[2], Prof Dr

[1]Department of Journalism and Communication Research, Hanover University of Music Drama and Media, Expo Plaza 12, Hannover, Germany

[2]Department of Communication, Johannes Gutenberg University Mainz, Mainz, Germany

Corresponding Author:
Paula Memenga, MA
Department of Journalism and Communication Research, Hanover University of Music Drama and Media, Expo Plaza 12, Hannover, Germany

Abstract

Background: Providing tailored information is an essential part of health care. However, physicians often lack time for detailed education during the consultation. An additional, tailored digital health information service (DHIS) could help physicians meet their patients' information needs regardless of time and place and extend physician-patient communication to the digital realm.

Objective: This study aimed to examine physicians' intentions to provide a DHIS to their patients and identify facilitating factors and barriers, guided by the extended unified theory of acceptance and use of technology.

Methods: A cross-sectional online survey with German physicians from various specialties was conducted in March 2022. The sample (N=364) ranged in age from 33 to 75 years (mean 53.92, SD 8.12), and the majority were male participants (31.9% [n=116] female participants). A blockwise multiple linear regression analysis was conducted to identify facilitating factors and barriers of physicians' intentions to provide a DHIS.

Results: Overall, 54.1% (n=197) of the surveyed physicians were (rather) willing to provide a tailored DHIS, 23.9% (n=87) were undecided, and 22% (n=80) were (rather) not willing to provide such a service to their patients. The overall model of a blockwise multiple linear regression analysis explained 56.8% of the variance of physicians' intentions. Perceived usefulness for job performance and patient outcomes as well as personal innovativeness was positively associated with physicians' intentions to provide a DHIS to their patients. Ease of use, social influence, facilitating conditions, price value, and habit were not associated with their intentions.

Conclusions: The perspective of the majority of surveyed physicians suggests that a tailored DHIS seems to be a promising way to provide additional health information and thus enhance face-to-face physician-patient communication. Efforts supporting the implementation of DHIS should address job performance and patient outcomes in particular. Further, physicians with a more positive attitude could serve as multipliers to increase the adoption of DHIS.

(JMIR Hum Factors 2026;13:e80620) doi:10.2196/80620

KEYWORDS

digital health information; information provision; physician-patient communication; health technology acceptance; UTAUT2; extended unified theory of acceptance and use of technology

Introduction

Digital Provision of Health Information

Since patients have a right to informed medical decision-making [1], and providing tailored information is an essential part of health care [2], physicians are required to meet their patients' needs for adequate, high-quality information [3]. However, physicians often lack the time to provide their patients with detailed information during the consultation [4,5]. Further, an increasing amount of false or misleading information, particularly prevalent during the COVID-19 pandemic [6], poses a challenge in providing accurate health care. These barriers emphasize the need to rethink physician-patient communication in general and consider new ways of providing high-quality health information to patients.

The digital age, characterized by advances in information and communication technology [7,8], offers new opportunities for information provision that could help to supplement the traditional relationship between health professionals and their patients [9]. For example, face-to-face physician-patient communication could be extended to the digital provision of health information, enabling physicians to meet their patients'

XSL•FO
RenderX

needs for high-quality information regardless of time and place. Such a digital health information service (DHIS) is also in line with global objectives of strengthening health care and improving patient education through the implementation of digital health technologies [10,11]. It can be understood as a web-based platform through which physicians can provide their patients with information tailored to their individual disease or health situation (eg, on symptoms, medical examinations, treatment options, prevention, and early detection measures) in addition to the explanations given during the consultation. Thus, patients receive tailored, high-quality information from their most trusted source of health information [12,13] that can be accessed via a computer or mobile phone, enabling them to autonomously build an adequate knowledge base for informed decision-making.

A basic requirement for realizing the potential of a DHIS in overcoming barriers to information provision and complementing routine health care is physicians' acceptance of such a service [14,15]. Existing studies on digital health interventions revealed mixed results regarding physicians' willingness to engage with digital services. On the one hand, physicians in Germany perceived great opportunities from health apps with regard to prevention, reminder, and lifestyle support features [16] and showed high use of internet-related technologies for patient support [17]. On the other hand, prior studies showed a moderate acceptance of mobile apps for disease management [18] and a rather low acceptance of eHealth interventions in inpatient routine care [19].

Based on the heterogeneous state of research, this study aimed to examine physicians' intentions to provide a DHIS to their patients as well as identify facilitating factors and barriers. The study was guided by the extended unified theory of acceptance and use of technology (UTAUT2) [20], which is a widely accepted theory for understanding physicians' adoption of digital health interventions [21,22] and informing eHealth implementation [23].

Theoretical Background

Postulates of the UTAUT2

The UTAUT2 [20], which we used as a guiding framework to examine physicians' intentions to provide a DHIS, is a common technology acceptance model that was originally developed to explain consumers' acceptance and use of innovative technologies. It can be understood as an integrative model as it unifies elements of 8 theoretically proven models from different fields of research, such as information systems research (eg, Technology Acceptance Model), sociology (eg, Innovation Diffusion Theory), and psychology (eg, theory of planned behavior). UTAUT2 is an extension of the unified theory of acceptance and use of technology (UTAUT) [24], which postulates performance expectancy, effort expectancy, social influence, and facilitating conditions as key determinants of usage intentions. UTAUT2 additionally considers hedonic motivation, price value, and habit as predictors of the intention to use a specific technology.

Previous research has shown that UTAUT and UTAUT2 are also effective in various health contexts, explaining patients' and health care providers' acceptance of health-related technologies such as teleconsultation technology [25-27], electronic health records [28-32], mobile health apps [18,33-37], or eHealth interventions [19,38].

Modeling Predictors of Physicians' Intentions to Provide a DHIS Using UTAUT2

Overview

Based on the UTAUT2 and extant research in the health context, performance expectancy, effort expectancy, social influence, facilitating conditions, price value, and habit were integrated into our research model. Hedonic motivation is conceptualized within UTAUT2 as the pleasure or enjoyment derived from technology use and has primarily been applied in consumer-oriented and leisure-related contexts [20]. In contrast, physicians' use of digital health technologies occurs within professional roles and clinical responsibilities. Consistent with prior health care-related applications of UTAUT2, we therefore excluded hedonic motivation [27,31,33]. In addition, personal innovativeness was added as an important factor to explain physicians' intentions to provide a DHIS to their patients.

Performance Expectancy

In the health context, performance expectancy [20,24] is defined as the degree to which a physician expects that using a specific digital technology provides personal benefits in the daily work life, for example saving time during office hours or improving patient communication [22]. It thus describes the perceived usefulness of the DHIS for improving physicians' job performance. Previous research has shown that performance expectancy is a key predictor of healthcare providers' intention to use digital, health-related technologies [18,19,28]. However, digital health services for tailored information provision not only aim to improve job performance but also to promote patient-centered care [10]. In line with research showing that physicians perceive digital health services' potential to support patients' health promotion and prevention [16,17], performance expectancy might also refer to the physicians' perceived usefulness for improving patient outcomes, for example, meeting information needs, supporting decision-making, and improving health and well-being. According to the theoretical assumptions and state of research outlined, we derived the first hypothesis:

H1:Perceived usefulness for job performance (H1a) and perceived usefulness for improving patient outcomes (H1b) will be positively related to physicians' intention to provide a DHIS.

Effort Expectancy

Effort expectancy [20,24] refers to the perceived ease of use associated with using a specific technology which is a well-supported facilitator of a higher intention to use a digital, health-related technology [25,28,29,37]. In the context under study, effort expectancy refers to the extent to which a physician expects to be able to use a DHIS with low effort [22]. We derived the following hypothesis:

H2: Perceived ease of use will be positively related to physicians' intention to provide a DHIS.

XSL•FO
RenderX

Social Influence

Social influence [20,24] covers subjective norms as a known driver of a variety of social and health behaviors such as using digital, health-related services [19,28,39]. Applied to our study, these codes of conduct can be defined as the extent to which a health care provider perceives that important reference groups think that he or she should use a particular technology [22]. In the medical setting, relevant reference groups could be patients as well as colleagues [19]. Accordingly, the perception that patients think one should provide a tailored DHIS for them and that colleagues think one should provide digital information to one's patients may be related to the intention to provide such a service. Thus, we assumed the following:

H3: Subjective norms related to patients (H3a) and colleagues (H3b) will be positively related to physicians' intention to provide a DHIS.

Facilitating Conditions

Facilitating conditions [20,24] refer to the available external support and resources for using digital technology. For example, the practices' or hospitals' technical infrastructure and knowledge necessary to support its implementation could influence physicians' intentions to provide a DHIS [20,22,24]. In line with former research that found a higher degree of facilitating conditions to be positively related to higher usage intentions among health care providers [25,28,29,39], we derived the following hypothesis:

H4: Facilitating conditions will be positively related to physicians' intention to provide a DHIS.

Price Value

Price value [20] refers to the monetary costs associated with using digital technology and positively affects intentions to use digital, health-related technologies in general [26,32,40]. Thus, we assumed the following:

H5: Price value will be positively related to physicians' intention to provide a DHIS.

Habit

Habit [20] describes the extent to which the use of digital technology is perceived as an automatic process based on prior experiences which is associated with higher usage intentions [37]. It refers to physicians' perceptions that they have already developed a routine or habit of using digital technologies such as a DHIS in their daily work lives. Therefore, we postulated the following hypothesis:

H6: Habit will be positively related to physicians' intention to provide a DHIS.

Personal Innovativeness

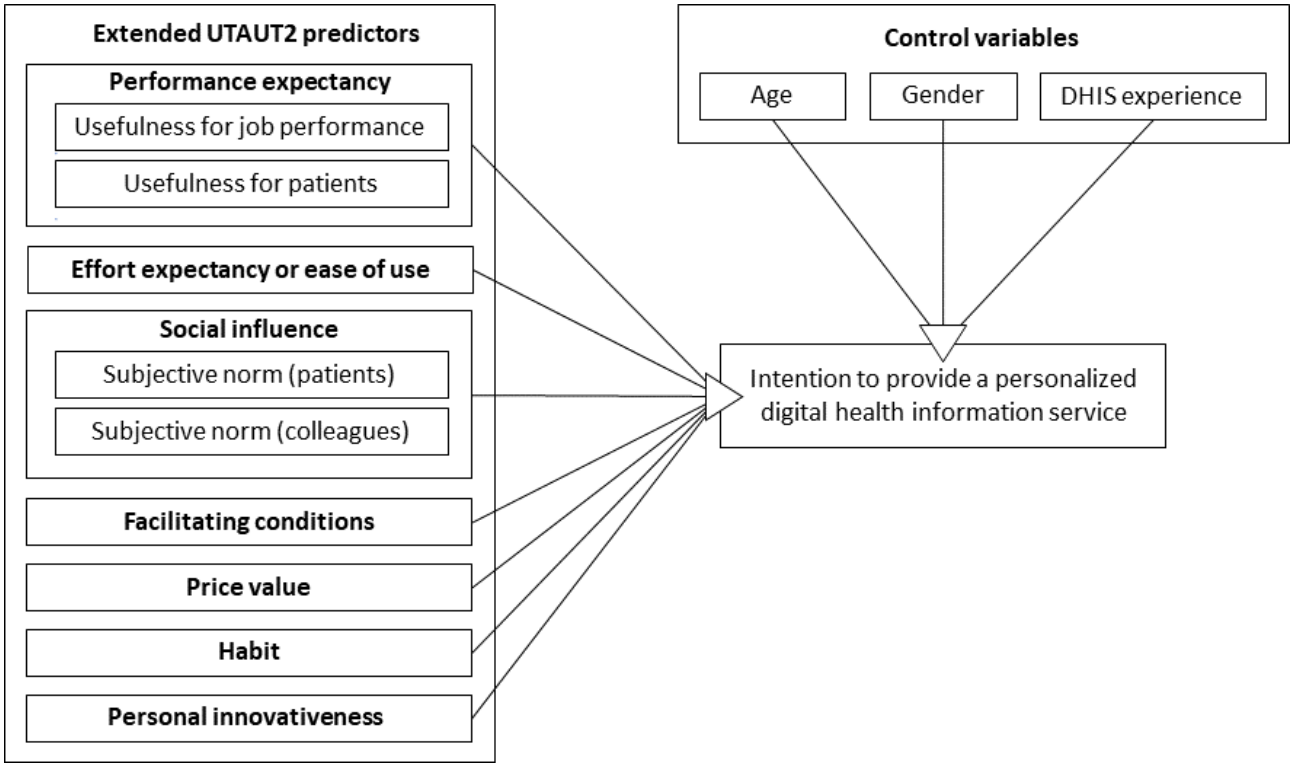
Although personal innovativeness is not included in the UTAUT2 model, first studies support that integrating it into UTAUT2 to explain patients' and health care providers' acceptance of digital, health-related technologies provides additional explanatory power [26,33,34,37,41]. Thus, it can be considered as a further important construct that requires more attention in health service acceptance and implementation research [40]. Innovativeness is a personal trait that refers to the "extent to which an individual is innovative by his or her perception of new idea or technology" [26], leading these individuals to be more likely to try innovations. In the current context, it can be postulated that physicians with higher personal innovativeness may be more willing to provide a tailored DHIS for their patients, leading to the following hypothesis:

H7: Personal innovativeness will be positively related to physicians' intention to provide a DHIS.

An overview of the proposed research model is shown in Figure 1.

XSL•FO
RenderX

Figure 1. Research model. DHIS: digital health information service; UTAUT2: extended unified theory of acceptance and use of technology.



Methods

To test our hypotheses, we conducted a cross-sectional online survey in March 2022 with physicians from various specialties (eg, general medicine, cardiology, dermatology, and gynecology). The physicians were recruited and invited to use different strategies. At first, physicians and medical associations were contacted and asked to forward the invitation to the survey. In addition, the call was published in professional journals for physicians. A commercial online access panel was used in the second step.

Measures

Intention to Provide a DHIS

To measure physicians' intentions to provide a tailored DHIS to their patients, they first received an introduction describing the service as a digital, web-based platform through which tailored health information (eg, on the basics of a disease, symptoms, risks, or treatment options) can be provided in addition to the explanations given during the consultation and accessed by patients at home via a computer or smartphone (S1 in Multimedia Appendix 1). This description was based on the features of an existing DHIS. Subsequently, they were asked whether they had provided a comparable service to their patients in the past and to rate their intentions to (continue to) provide such a service using the 3 original UTAUT2 items adapted to the current context [20] (Table S2 in Multimedia Appendix 1). The 3 items were measured on a 5-point Likert-type scale (1=*strongly disagree* to 5=*strongly agree*). The scale showed excellent internal consistency and was summarized into a mean index (α=.94, mean 3.45 [*SD* 1.17]).

Performance Expectancy: Perceived Usefulness for Job Performance and Patients

The measurement of perceived usefulness of the DHIS for job performance and improving patient outcomes was adapted from the original UTAUT2 items [20] (Table S2 in Multimedia Appendix 1). Both constructs were measured with 5 items on a 5-point Likert-type scale (1=strongly disagree to 5=strongly agree). The scales showed excellent internal consistency and were summarized into mean indices (usefulness for job performance: α=.90, mean 3.49 [SD 0.94]; usefulness for patients: α=.91, mean 3.68 [SD 0.83]).

Effort Expectancy: Perceived Ease of Use

Perceived ease of use was also measured with a modified version of the UTAUT2 items [20], using 3 items on 5-point Likert-type scales (1=strongly disagree to 5=strongly agree) (Table S2 in Multimedia Appendix 1). The mean index was calculated due to the good internal consistency (α=.86, mean 3.21 [SD 1.02]).

Social Influence: Subjective Norms

Subjective norms related to patients and colleagues were each measured with 3 items adapted from Venkatesh et al [20] and Park and Smith [42] (Table S2 in Multimedia Appendix 1). The participants assessed their responses on 5-point Likert-type scales, ranging from 1=strongly disagree to 5=strongly agree. Both scales showed acceptable internal consistencies and were transformed into mean indices (subjective norm patients: α=.76, mean 2.94 [SD 0.84]; subjective norm colleagues: α=.79, mean 2.81 [SD 0.82]).

Facilitating Conditions

Facilitating conditions were measured with the original UTAUT2 items [20] (Table S2 in Multimedia Appendix 1). The 4-item measure was rated on a 5-point Likert-type scale

XSL•FO
RenderX

(1=strongly disagree to 5=strongly agree). A mean index of high reliability was calculated (α=.74, mean 3.50 [SD 0.90]).

Price Value

Price value was measured with 3 items about physicians' requirements for the costs and incentives that might be associated with providing a digital service [20,40] (Table S2 in Multimedia Appendix 1). Participants rated their agreement that a DHIS should be free to physicians and patients on a 5-point Likert-type scale ranging from 1=strongly disagree to 5=strongly agree. However, as the scale showed poor internal consistency (α=.47), no mean index could be computed. Therefore, 1 single item that most accurately represented costs for the physician was selected for the analysis (mean 4.28 [SD 0.92]). We note that this item reflects physicians' perception of whether the service should be free rather than the traditional UTAUT2 definition of price value, which captures the trade-off between perceived benefits and costs of using a technology.

Habit

The habit of using digital technologies was measured using the original 3-item UTAUT2 measure [20] (Table S2 in Multimedia Appendix 1) and was evaluated on a 5-point Likert-type scale (1=strongly disagree to 5=strongly agree). The scale showed good internal consistency and was summarized into a mean index (α=.81, mean 3.78 [SD 0.95]).

Personal Innovativeness

Personal innovativeness was measured with 3 items that were adapted from Baudier et al [27] (Table S2 in Multimedia Appendix 1). Participants rated their responses on a 5-point Likert-type scale, ranging from 1=strongly disagree to 5=strongly agree. The scale was compressed into a mean index (α=.84, mean 3.30 [SD 1.03]).

Control Variables: Age, Gender, and DHIS Experience

Age, gender, and DHIS experience were used as control variables as these constructs are also assumed to affect technology usage intentions [20]. To measure physicians' experience with a tailored DHIS, they were asked if they had provided a comparable service to their patients in the past (no/yes; 15.4% (56/364) with DHIS experience).

Data Analysis

To identify facilitating factors and barriers of physicians' intentions to provide a tailored DHIS in line with hypotheses 1 to 7, we conducted a blockwise multiple linear regression analysis. The first block considered the control variables age, gender, and DHIS experience. In the second block, the theoretically derived factors were included.

Ethical Considerations

The study received ethical approval from the Central Ethics Committee at Leibniz University Hannover (review number: EV LUH 19/2021). All participants were informed about the content of the survey prior to their participation and provided informed consent. No personally identifiable information was collected. Participants recruited through the commercial online access panel received financial compensation from the market research company in accordance with its standard remuneration rates. There were no further compensations. All methods were performed in accordance with the Declaration of Helsinki. The study was reported in accordance with the American Psychological Association Journal Article Reporting Standards for quantitative research.

Results

The sample (N=364) ranged in age from 33 to 75 years (mean 53.92, SD 8.12) and the majority were male participants (31.9% [n=116] female participants). Overall, 97% (n=353) worked in a physician's practice and 3% (n=11) worked in a hospital or other medical facility in Germany. The majority participated via the commercial online access panel (n=329).

Overall, 54.1% (n=197) of the surveyed physicians were (rather) willing to provide a tailored DHIS, 23.9% (n=87) were undecided, and 22% (n=80) were (rather) not willing to provide such a service to their patients. The overall model of the blockwise multiple linear regression analysis explained 56.8% of the variance of the intention to provide patients with a tailored DHIS. The amount of explained variance of the theoretically derived predictors usefulness for job performance, usefulness for patients, ease of use, subjective norms, facilitating conditions, price value, habit, and personal innovativeness was 42.1%. The results are displayed in Table 1.

XSL•FO
RenderX

Table . Predictors of physicians' intentions to provide a digital health information service (N=364)[a].

Predictors	Unstandardized coefficient (B)	95% CI for B	SE B	β	R^2_{corr}	ΔR^2_{corr}
Constant	−0.34	−1.25 to 0.57	0.46			
Sociodemographics and DHIS[b] experience					0.147	0.147[c]
Age	0.00	−0.01 to 0.01	0.01	−0.02		
Gender (ref. female)	0.00	−0.18 to 0.18	0.09	0.00		
Current use (ref. yes)	0.56	0.33 to 0.80	0.12	0.17[c]		
Extended UTAUT2[d] predictors					0.568	0.421[c]
Usefulness for job performance	0.36	0.20 to 0.51	0.08	0.29[c]		
Usefulness for patients	0.28	0.10 to 0.46	0.09	0.20[e]		
Ease of use	0.08	−0.04 to 0.19	0.06	0.07		
Subjective norm (patients)	0.04	−0.09 to 0.18	0.07	0.05		
Subjective norm (colleagues)	−0.01	−0.13 to 0.11	0.06	−0.01		
Facilitating conditions	−0.02	−0.14 to 0.11	0.06	−0.01		
Price value	0.03	−0.06 to 0.13	0.05	0.03		
Habit	0.06	−0.06 to 0.18	0.06	0.05		
Personal innovativeness	0.28	0.17 to 0.40	0.06	0.25[c]		

[a]The table shows model 2 of the blockwise multiple linear regression analysis, identifying physicians' intentions of providing a tailored digital health information service to their patients. The overall model was significant ($F_{12,350}$=40.70, P<.001) and explained 56.8% of variance. All variance inflation factors <5.

[b]DHIS: digital health information service.

[c]P≤.001.

[d]UTAUT2: extended unified theory of acceptance and use of technology.

[e]P ≤.01

As assumed in H1, perceived usefulness for job performance (β=.29; P<.001) and perceived usefulness for patient outcomes (β=.20; P=.002) were positively related to physicians' intentions to provide a DHIS. Physicians who expect a tailored DHIS to be useful in their daily work life (eg, for saving time during office hours, improving patient communication) and for improving patient outcomes (eg, decision-making, empowerment, improving health) tend to be more willing to provide such a service to their patients. Thus, our results supported H1a and H1b.

H2 focused on perceived ease of using a DHIS, which we postulated to be positively related to physicians' intentions to provide such a service. However, ease of use (β=.07; P=.20) showed no significant association with providing intentions. Based on these findings, H2 had to be rejected.

Considering social influence, we postulated that subjective norms related to patients (H3a) and colleagues (H3b) as physicians' relevant reference groups are positively related to physicians' intentions to provide a DHIS. Contrary to our assumptions, both patient-related (β=.03; P=.53) and colleague-related subjective norms (β=−.01; P=.88) were not associated with physicians' intentions. Accordingly, H3a and H3b had to be rejected.

In H4, we assumed that facilitating conditions are positively related to physicians' intentions to provide their patients with tailored health information via a digital service. However, there was no relation between facilitating condition (β=−.01; P=.77) and physicians' providing intentions. Thus, H4 had to be rejected.

XSL•FO
RenderX

H5 concerned the monetary costs that are associated with the provision of a DHIS, which we assumed to be positively related to physicians' intentions to provide such a service. However, price value (β=.03; *P*=.46) did not serve as a facilitating factor of physicians' intentions. Therefore, H5 had to be rejected.

H6 focused on the habit of technology use being positively related to physicians' intentions. Our results indicated that habit (β=.05; *P*=.33) is not associated with physicians' intentions to provide a DHIS to their patients. Accordingly, H6 was rejected.

In H7, we postulated that physicians' personal innovativeness is related to higher intentions to provide a DHIS. Indeed, personal innovativeness (β=.25; *P*<.001) was related to the intention to provide such a service, indicating that a more positive attitude toward digital innovation tends to be a facilitating factor. Thus, H7 was supported.

Discussion

Principal Findings

The current study examined physicians' intentions to provide a tailored DHIS for their patients. Our findings revealed that the physicians surveyed showed moderate intentions, with the majority (n=197, 54.1%) being rather willing to provide such a service. Guided by the theoretical framework of the UTAUT2, we identified factors that are associated with their intentions. Perceived usefulness for job performance, perceived usefulness for patients, and personal innovativeness were positively associated with their intentions to provide a DHIS, while perceived ease of use, social influence, facilitating conditions, price value, and habit were not significantly associated with their intentions.

Theoretical Implications for a Model of Physicians' Intentions to Provide a DHIS

The analysis revealed that the tested model, based on the UTAUT2, accounted for 56.8% of the variance. Nevertheless, the effectiveness of this framework in explaining physicians' intentions to provide a DHIS is questionable, as only 3 out of the 9 integrated factors contributed significantly to variance explanation. Physicians' perceived usefulness for job performance and perceived usefulness for patients were most strongly associated with their intentions to provide a DHIS, which is in line with prior research on health professionals' intentions to use health-related technology in general [18,19,28].

Furthermore, the inclusion of personal innovativeness within the UTAUT2 framework demonstrated a positive association with physicians' intentions to provide a DHIS. This finding strengthens the suggestion to integrate the personal innovativeness construct into models explaining physicians' intentions to engage with digital services. Such an implication is also supported in prior studies on both patients' [33,34] and physicians' [37] acceptance of digital health technologies. As an interim conclusion, it can therefore be stated that performance expectancy, separated for job performance and patient outcomes, and personal innovativeness could serve as basic factors for a robust model.

The majority of theoretically derived factors were not associated with physicians' intentions to provide a tailored DHIS. First, perceived ease of use did not prove to be relevant, indicating that in the specific context of a tailored DHIS, its implementation might be worth any effort, as digital health information is a "common good" [10] that should be accessible to everyone [11]. In this regard, Zhai et al [39] point out that physicians usually have a high workload and therefore might not mind spending more time on familiarization with a new service that could diminish their workload. In addition, the fact that perceived ease of use is not a key determinant of physicians' intentions could be related to the sample of the online access panel, which may be more digitally savvy and therefore perceive fewer difficulties in the context of handling DHIS. The potential value of integrating perceived ease of use into a model for explaining physicians' intentions to provide a DHIS should thus be explored with a less digitally savvy sample.

Second, social influence—the perception that patients or colleagues think one should provide a DHIS—was irrelevant to the physicians' intentions. However, the mean scores of these variables were mediocre, indicating that norms regarding the provision of a DHIS might not be established and thus be less influential. Indeed, digital health services are not yet widely used in Germany [43]. In this regard, our study found that only 15.4% (n=56) of respondents had already provided a comparable service to their patients. The inclusion of social influence as a potential determinant of physicians' intention to provide a DHIS could therefore be fruitful at a later stage, when the relevant norms are more established in Germany.

In addition, facilitating conditions were not related to physicians' intentions to provide a DHIS. This finding suggests that having the necessary infrastructure such as technical resources and support might not be relevant until the actual implementation of such a service, which is in line with the first version of the UTAUT framework [24]. Moreover, physicians may already have a good infrastructure in this respect. For example, a DHIS only requires computer and internet access, which is available in the majority of German practices and hospitals. In addition, the presumably digitally savvy participants of an online access panel may also have the technical know-how to use a DHIS. Therefore, it may be worthwhile to investigate the significance of facilitating conditions within a less digitally savvy sample, while also considering the digital infrastructure of the respective medical facilities.

Furthermore, price value, as measured in this study, was not related to physicians' intentions to provide a DHIS. The item used reflected physicians' preference for the service to be free of charge, rather than the traditional UTAUT2 definition of price value, which captures the trade-off between perceived benefits and costs. While many physicians indicated that the service should be free of charge, this does not allow us to conclude the influence of actual costs on their intention to provide a DHIS. Previous studies often did not include price value when examining usage intentions of health-related technologies, as it may be less relevant in professional health care contexts [27,31,33,38]. This highlights the need for context-specific adaptations when specifying models to explain physicians' intentions to use digital health technologies.

XSL•FO
RenderX

Habit was also not related to physicians' intentions to provide a DHIS. Thus, familiarity with digital technologies in general is not relevant in the context of a DHIS. This result could be attributed to our measurement. As DHIS are not yet widely used in Germany [43], we asked about physicians' habits regarding digital technologies in general and not about the perceived routine of using a DHIS in particular. Thus, integrating habit in handling a DHIS as a determinant of physicians' intentions to provide such a service appears less meaningful until DHIS are more established in Germany.

Limitations and Resulting Tasks for Future Research

Our study is not without limitations. First, as we conducted a cross-sectional survey, it is not possible to draw causal conclusions. Second, our study is not representative of all physicians in Germany. Since we focused on physicians working in practices and did not stratify by their expertise, physicians in hospitals and other medical institutions as well as certain specialties are underrepresented in this study. Hence, a potential bias may exist, given that a DHIS may be considered more beneficial for certain disciplines. In addition, the majority of the surveyed physicians were participants in an online access panel and are possibly more digitally savvy, which does not necessarily reflect the perspective of most physicians in Germany.

Third, the physicians assessed their intentions to provide a DHIS solely based on a textual description of its features, which introduced the service in general terms and presented it as easy to use. They did not have the opportunity to try it out with their patients, and item-level responses may reflect assumptions based on the vignette rather than experience with a system. As a result, intentions reported in this study may not fully correspond to the actual provision of a DHIS. Although this approach is common in survey-based technology acceptance research, it may limit the validity of our findings. Future research should test a DHIS in practice, focusing on specific specialties to improve the quality of usage prediction by evaluating the actual use in a particular context (eg, cardiology) rather than intentions in general. Experimental or longitudinal designs could be used to allow for causal inferences and to explore the effectiveness of such a service on patient outcomes and physicians' job performance.

A fourth limitation concerns the theoretical framework applied in this study. Although UTAUT2 has been widely used in health care research, it is a consumer-oriented acceptance model that may not fully capture the professional and organizational context of physicians as users of digital health technologies. We partly addressed this limitation by adapting the model to the study context, for example, by excluding hedonic motivation. Nevertheless, future research should consider extending acceptance models to more explicitly reflect physicians' professional roles and decision-making contexts. Although hedonic motivation was excluded based on theoretical and contextual considerations, future studies could empirically examine its relevance in professional health care settings to determine whether enjoyment contributes to physicians' acceptance of DHIS.

Practical Implications

There are several practical implications that can be derived from our findings. First, when promoting a DHIS to physicians, the personal value for improving their daily work life, patient communication, and patient outcomes should be addressed. In detail, it should be explained that a DHIS can save time during the medical consultation and has great potential to support patients' decision-making by providing important, tailored health information. Second, physicians with a positive attitude toward digital innovation may support dissemination by sharing experiences and demonstrating use cases at medical congresses or other professional events, thereby increasing awareness of DHIS options. In addition, a digital information service and its potential could be presented in national medical journals, which could contribute to the establishment of social norms. Another sustainable approach to promote the provision of digital health information is the implementation of eHealth in professional education. At last, in the sense of empowerment and patient-centered care, patients could be encouraged to claim their right to be adequately informed and to directly ask their physicians for digital health information. Furthermore, the effects of providing a tailored DHIS on patient outcomes need to be examined in more detail to provide comprehensive evidence.

Conclusions

In conclusion, the perspective of the majority of surveyed physicians suggests that a tailored DHIS seems to be a promising way to provide additional health information and thus enhance face-to-face physician-patient communication. The key determinants of physicians' intentions to provide such a service—perceived usefulness for improving job performance and patient outcomes as well as physicians' personal innovativeness—underscore the importance of addressing these aspects when supporting its implementation. Future research should aim to identify additional determinants to further advance the development of a more comprehensive model and focus efforts to establish DHIS in Germany.

Acknowledgments

Parts of this work have been presented at the 73rd conference of the International Communication Association 2023 in Toronto, Canada.

Funding

This work was financially supported by the Volkswagen Foundation (grant number: 98975). The funding body did not contribute to the design of the study, the collection, analysis, and interpretation of data, or the writing of the manuscript.

XSL•FO
RenderX

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Authors' Contributions

PM and EL did the conceptualization, methodology, data collection, and data analysis of the study. EL was responsible for funding acquisition and project administration. PM wrote the main manuscript text. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Introduction of the digital health information service and survey items.
[DOCX File, 19 KB - humanfactors_v13i1e80620_app1.docx]

References

1. Hahlweg P, Bieber C, Levke Brütt A, et al. Moving towards patient-centered care and shared decision-making in Germany. Z Evid Fortbild Qual Gesundhwes 2022 Jun;171:49-57. [doi: 10.1016/j.zefq.2022.04.001] [Medline: 35595668]
2. Epstein RM, Street RL. The values and value of patient-centered care. Ann Fam Med 2011;9(2):100-103. [doi: 10.1370/afm.1239] [Medline: 21403134]
3. Southwell BG, Wood JL, Navar AM. Roles for health care professionals in addressing patient-held misinformation beyond fact correction. Am J Public Health 2020 Oct;110(S3):S288-S289. [doi: 10.2105/AJPH.2020.305729] [Medline: 33001723]
4. Weise J. How to reach a practice – information sources and communication channels for and by doctors [Book in German]. In: Matusiewicz D, Stratmann F, Wimmer J, editors. Marketing Im Gesundheitswesen: Einführung - Bestandsaufnahme – Zukunftsperspektiven [Marketing in Health Care: Introduction – Stocktaking – Future Perspectives]: Springer; 2019:433-444. [doi: 10.1007/978-3-658-20279-8_32]
5. Yarnall KSH, Pollak KI, Østbye T, Krause KM, Michener JL. Primary care: is there enough time for prevention? Am J Public Health 2003 Apr;93(4):635-641. [doi: 10.2105/ajph.93.4.635] [Medline: 12660210]
6. Zarocostas J. How to fight an infodemic. The Lancet 2020 Feb;395(10225):676. [doi: 10.1016/S0140-6736(20)30461-X]
7. Akdur G, Aydin MN, Akdur G. Adoption of mobile health apps in dietetic practice: case study of Diyetkolik. JMIR Mhealth Uhealth 2020 Oct 2;8(10):e16911. [doi: 10.2196/16911] [Medline: 33006566]
8. Mesko B, Győrffy Z. The rise of the empowered physician in the digital health era: viewpoint. J Med Internet Res 2019 Mar 26;21(3):e12490. [doi: 10.2196/12490] [Medline: 30912758]
9. Awad A, Trenfield SJ, Pollard TD, et al. Connected healthcare: Improving patient care using digital health technologies. Adv Drug Deliv Rev 2021 Nov;178:113958. [doi: 10.1016/j.addr.2021.113958] [Medline: 34478781]
10. Krist AH, Phillips R, Leykum L, Olmedo B. Digital health needs for implementing high-quality primary care: recommendations from the National Academies of Sciences, Engineering, and Medicine. J Am Med Inform Assoc 2021 Nov 25;28(12):2738-2742. [doi: 10.1093/jamia/ocab190] [Medline: 34537851]
11. Global strategy on digital health 2020-2025. : World Health Organization; 2021 URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2024-02-05]
12. Baumann E, Czerwinski F, Rosset M, Seelig M, Suhr R. Wie informieren sich die Menschen in Deutschland zum Thema Gesundheit? Erkenntnisse aus der ersten Welle von HINTS Germany [Article in German]. Bundesgesundheitsbl 2020 Sep;63(9):1151-1160. [doi: 10.1007/s00103-020-03192-x]
13. Swoboda CM, Van Hulle JM, McAlearney AS, Huerta TR. Odds of talking to healthcare providers as the initial source of healthcare information: updated cross-sectional results from the Health Information National Trends Survey (HINTS). BMC Fam Pract 2018 Aug 29;19(1):146. [doi: 10.1186/s12875-018-0805-7] [Medline: 30157770]
14. Pfeuffer N, Beyer A, Penndorf P, et al. Evaluation of a health information exchange system for geriatric health care in rural areas: development and technical acceptance study. JMIR Hum Factors 2022 Sep 15;9(3):e34568. [doi: 10.2196/34568] [Medline: 36107474]
15. Philippi P, Baumeister H, Apolinário-Hagen J, et al. Acceptance towards digital health interventions - model validation and further development of the unified theory of acceptance and use of technology. Internet Interv 2021 Dec;26:100459. [doi: 10.1016/j.invent.2021.100459] [Medline: 34603973]
16. Wangler J, Jansky M. Welchen Nutzen bringen Gesundheits-Apps für die Primärversorgung? Ergebnisse einer Befragung von Allgemeinmedizinern [Article in German]. Präv Gesundheitsf 2021 May;16(2):150-156. [doi: 10.1007/s11553-020-00797-7]
17. Kirkovits T, Schinkoethe T, Drewes C, et al. eHealth in modern patient-caregiver communication: high rate of acceptance among physicians for additional support of breast cancer patients during long-term therapy. JMIR Cancer 2016 Sep 19;2(2):e14. [doi: 10.2196/cancer.5132] [Medline: 28410173]

XSL•FO
RenderX

18. Breil B, Salewski C, Apolinário-Hagen J. Comparing the acceptance of mobile hypertension apps for disease management among patients versus clinical use among physicians: cross-sectional survey. JMIR Cardio 2022 Jan 6;6(1):e31617. [doi: 10.2196/31617] [Medline: 34989683]
19. Hennemann S, Beutel ME, Zwerenz R. Ready for eHealth? Health professionals' acceptance and adoption of eHealth interventions in inpatient routine care. J Health Commun 2017 Mar;22(3):274-284. [doi: 10.1080/10810730.2017.1284286] [Medline: 28248626]
20. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012 Mar 1;36(1):157-178. [doi: 10.2307/41410412]
21. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935. [doi: 10.2196/15935] [Medline: 32130167]
22. Li J, Talaei-Khoei A, Seale H, Ray P, Macintyre CR. Health care provider adoption of eHealth: systematic literature review. Interact J Med Res 2013 Apr 16;2(1):e7. [doi: 10.2196/ijmr.2468] [Medline: 23608679]
23. Heinsch M, Wyllie J, Carlson J, Wells H, Tickner C, Kay-Lambkin F. Theories informing eHealth implementation: systematic review and typology classification. J Med Internet Res 2021 May 31;23(5):e18500. [doi: 10.2196/18500] [Medline: 34057427]
24. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view1. MIS Q 2003 Sep 1;27(3):425-478. [doi: 10.2307/30036540]
25. Adenuga KI, Iahad NA, Miskon S. Towards reinforcing telemedicine adoption amongst clinicians in Nigeria. Int J Med Inform 2017 Aug;104:84-96. [doi: 10.1016/j.ijmedinf.2017.05.008] [Medline: 28599820]
26. Baudier P, Kondrateva G, Ammi C. The future of telemedicine cabin? The case of the French students' acceptability. Futures 2020 Sep;122:102595. [doi: 10.1016/j.futures.2020.102595] [Medline: 32834075]
27. Baudier P, Kondrateva G, Ammi C, Chang V, Schiavone F. Patients' perceptions of teleconsultation during COVID-19: a cross-national study. Technol Forecast Soc Change 2021 Feb;163:120510. [doi: 10.1016/j.techfore.2020.120510] [Medline: 33318716]
28. Ahmed MH, Bogale AD, Tilahun B, et al. Intention to use electronic medical record and its predictors among health care providers at referral hospitals, north-west Ethiopia, 2019: using unified theory of acceptance and use technology 2 (UTAUT2) model. BMC Med Inform Decis Mak 2020 Sep 3;20(1):207. [doi: 10.1186/s12911-020-01222-x] [Medline: 32883267]
29. Dimitrovski T, Bath PA, Ketikidis P, Lazuras L. Factors affecting general practitioners' readiness to accept and use an electronic health record system in the Republic of North Macedonia: a national survey of general practitioners. JMIR Med Inform 2021 Apr 5;9(4):e21109. [doi: 10.2196/21109] [Medline: 33818399]
30. Ravangard R, Kazemi Z, Abbasali SZ, Sharifian R, Monem H. Development of the UTAUT2 model to measure the acceptance of medical laboratory portals by patients in Shiraz. Electron Physician 2017 Feb;9(2):3862-3869. [doi: 10.19082/3862] [Medline: 28465819]
31. Silver RA, Subramaniam C, Stylianou A. The impact of portal satisfaction on portal use and health-seeking behavior: structural equation analysis. J Med Internet Res 2020 Mar 27;22(3):e16260. [doi: 10.2196/16260] [Medline: 32217505]
32. Tavares J, Oliveira T. Electronic health record portal adoption: a cross country analysis. BMC Med Inform Decis Mak 2017 Jul 5;17(1):97. [doi: 10.1186/s12911-017-0482-9] [Medline: 28679423]
33. Chang YT, Chao CM, Yu CW, Lin FC. Extending the utility of UTAUT2 for hospital patients' adoption of medical apps: moderating effects of e-Health literacy. mob inf syst 2021 Apr 12;2021:8882317. [doi: 10.1155/2021/8882317]
34. Huang CY, Yang MC. Empirical investigation of factors influencing consumer intention to use an artificial intelligence-powered mobile application for weight loss and health management. Telemed J E Health 2020 Oct;26(10):1240-1251. [doi: 10.1089/tmj.2019.0182] [Medline: 31971883]
35. Kim B, Lee E. What factors affect a user's intention to use fitness applications? The moderating effect of health status: a cross-sectional study. Inquiry 2022;59:469580221095826. [doi: 10.1177/00469580221095826] [Medline: 35580021]
36. Schomakers EM, Lidynia C, Vervier LS, Calero Valdez A, Ziefle M. Applying an extended UTAUT2 model to explain user acceptance of lifestyle and therapy mobile health apps: survey study. JMIR Mhealth Uhealth 2022 Jan 18;10(1):e27095. [doi: 10.2196/27095] [Medline: 35040801]
37. Sezgin E, Özkan-Yildirim S, Yildirim S. Understanding the perception towards using mHealth applications in practice: physicians' perspective. Inf Dev 2018 Mar;34(2):182-200. [doi: 10.1177/0266666916684180]
38. Koivumäki T, Pekkarinen S, Lappi M, Väisänen J, Juntunen J, Pikkarainen M. Consumer adoption of future MyData-based preventive eHealth services: an acceptance model and survey study. J Med Internet Res 2017 Dec 22;19(12):e429. [doi: 10.2196/jmir.7821] [Medline: 29273574]
39. Zhai H, Yang X, Xue J, et al. Radiation oncologists' perceptions of adopting an artificial intelligence-assisted contouring technology: model development and questionnaire study. J Med Internet Res 2021 Sep 30;23(9):e27122. [doi: 10.2196/27122] [Medline: 34591029]
40. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Dec;4(1):50. [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]

XSL•FO
RenderX

41. Okumus B, Ali F, Bilgihan A, Ozturk AB. Psychological factors influencing customers' acceptance of smartphone diet apps when ordering food at restaurants. Int J Hosp Manag 2018 Jun;72:67-77. [doi: 10.1016/j.ijhm.2018.01.001]
42. Park HS, Smith SW. Distinctiveness and influence of subjective norms, personal descriptive and injunctive norms, and societal descriptive and injunctive norms on behavioral intent: a case of two behaviors critical to organ donation. Human Comm Res 2007 Apr;33(2):194-218. [doi: 10.1111/j.1468-2958.2007.00296.x]
43. Thiel R, Deimel L, Schmidtmann D, Piesche K, Hüsing T, Rennoch J, et al. #Smarthealthsystems: digitalisation strategies in international comparison [Report in German]. : Bertelsmann Stiftung; 2018 URL: https://www.bertelsmann-stiftung.de/fileadmin/files/Projekte/Der_digitale_Patient/VV_SHS-Gesamtstudie_dt.pdf [accessed 2025-02-05]

Abbreviations

DHIS: digital health information service
UTAUT: unified theory of acceptance and use of technology
UTAUT2: extended unified theory of acceptance and use of technology

Edited by A Kushniruk; submitted 14.Jul.2025; peer-reviewed by A Alshaher, J Grosser, S Voigt-Radloff; revised version received 18.Dec.2025; accepted 02.Jan.2026; published 06.Mar.2026.

Please cite as:
Memenga P, Link E
Physician-Patient Communication and Physicians' Acceptance of a Tailored Digital Health Information Service: Quantitative Online Survey
JMIR Hum Factors 2026;13:e80620
URL: https://humanfactors.jmir.org/2026/1/e80620
doi:10.2196/80620

© Paula Memenga, Elena Link. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 6.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Near Miss Reporting and Organizational Learning in Health Care: Conceptual Framework Development Study

Mohammed As'ad, MSc, MRCEM, MRCS, MBA

Dr Sulaiman Al Habib Medical Group, Corporate QPS, Headquarters, Al Awsat Valley S, Al Olaya, Riyadh, Saudi Arabia

Corresponding Author:
Mohammed As'ad, MSc, MRCEM, MRCS, MBA
Dr Sulaiman Al Habib Medical Group, Corporate QPS, Headquarters, Al Awsat Valley S, Al Olaya, Riyadh, Saudi Arabia

Abstract

Background: Near miss events can reveal system problems before patients are harmed, but current reviews are inconsistent and often rely on simple counts that are distorted by patient volume and reporting culture. Consequently, leaders cannot tell whether a rise in reports means that safety is getting worse or that staff are reporting more, and current systems are not strong enough to clearly separate real safety risks from random variation.

Objective: This study developed a 3-level near miss framework (NM³), a conceptual framework that converts descriptive near miss data into decision-grade intelligence through a structured, evidence-based process, including baseline measurement and advanced interpretation and governance.

Methods: NM³ was developed to provide decision-grade analytics for acute inpatient hospital settings. The framework was designed as a maturity model, progressing from baseline measurement to advanced interpretation. It integrates standardized definitions, rate calculations, statistical process control, severity weighting, and learning metrics.

Results: Level 1 establishes an organizational baseline through near miss rates per 1000 patient-days and near miss–to-harm ratios monitored with control charts. Level 2 introduces domain-specific denominators and unit-level charts to detect local variation. Level 3 applies severity weighting to generate a Near Miss Index; incorporates learning yields at 90 and 180 days; and triangulates near miss trends with harm events, exposure, reporting volume, and culture measures. A synthetic example demonstrates how the framework converts raw reports into stable rates, weighted indices, and learning metrics.

Conclusions: NM³ provides a structured pathway for organizations to strengthen near miss analytics. By progressing through maturity levels, leaders can improve the interpretation of safety signals, prioritize high-consequence risks, and integrate near miss reporting into governance.

(JMIR Hum Factors 2026;13:e87846) doi:10.2196/87846

KEYWORDS

near miss; patient safety; statistical process control; safety culture; quality improvement; acute care; hospital

Introduction

Background

Near miss events, which reveal systemic hazards before harm occurs, are fundamental to proactive safety governance. Global health policies now emphasize their systematic analysis, a principle operationalized in frameworks such as the National Health Service England's Patient Safety Incident Response Framework [1-3]. Consequently, leaders must treat near miss data as a crucial leading indicator of systemic safety rather than as a secondary metric [1].

Effective safety oversight requires analyzing trends rather than isolated event counts. Best practices advocate interpreting safety indicators (eg, rates of near misses, medication errors, patient falls, or infection events) alongside workforce culture and patient experience metrics to discern meaningful signals from statistical noise, thereby strengthening governance [4,5]. However, translating this into practice is fraught with methodological challenges. Voluntary reporting systems suffer from significant underascertainment due to reporting culture, workload, and fear [6,7]. These barriers disproportionately suppress near miss submissions compared to harm incidents, distorting data and undermining valid comparisons [8-10].

Analytical choices add further uncertainty. Raw event counts are confounded by patient volume and reporting behavior, while meaningful rates require domain-aligned denominators [4]. Furthermore, overdispersion is common in health care data, meaning that conventional statistical process control (SPC) charts can produce false signals if used without adjustment [11]. Although corrective methods such as the Laney p' chart exist [12], they have not been embedded within a structured near miss analytic framework that links SPC to event classification, severity weighting, and organizational learning—a gap this study addresses.

XSL•FO
RenderX

Objectives

These challenges create a critical interpretive dilemma for leaders: does a rise in reported near misses signify deteriorating safety or an improving reporting culture? Without standardized terminology, risk-adjusted denominators, and dispersion-robust control charts, leaders risk misinterpreting these vital data, misallocating resources, and failing to address systemic vulnerabilities [1,4,13,14].

From a human factors perspective, this interpretive dilemma reflects a classic problem of decision-making under uncertainty within complex sociotechnical systems [15]. Leaders must derive meaning from incomplete, socially mediated, and variably coded safety signals. In the absence of structured analytic support, interpretation depends on informal sensemaking [16] and locally constructed mental models [17], increasing the risk of inconsistent prioritization and resource allocation.

Therefore, the objective of this study was to develop a conceptual framework that provides a structured, evidence-based methodology for near miss analysis. To achieve this, we designed a 3-level maturity model that integrates standardized terminology, SPC, severity weighting, and novel learning metrics. This structured approach aims to provide organizational leaders with reliable, decision-grade intelligence to guide and sustain safety improvement initiatives. The 3-level near miss framework (NM^3) does not treat reporting frequency as a safety outcome; rather, it provides a structured analytic pathway that transforms near miss data into interpretable, learning-oriented intelligence for governance and improvement.

Methods

Design and Approach

This study reports the development of a conceptual framework for analyzing near miss safety events, following the integrative methodology by Jabareen [18]. This method proceeds through a series of iterative phases: mapping selected data sources, reading and categorizing the data, identifying and naming concepts, deconstructing and categorizing those concepts, integrating them into a coherent framework, and validating the result through an iterative process of rethinking and synthesis. The approach constructs interrelated concepts by synthesizing theory and evidence rather than through empirical data collection. The aim was to create a structured, 3-level model for near miss analytics that offers decision-grade intelligence to support organizational safety governance. The process was exclusively analytic and integrative and did not involve human participants.

Evidence and Theoretical Sources

The framework integrates principles from peer-reviewed research and authoritative institutional guidance. To align with established international standards, its foundational definitions were derived from the World Health Organization (WHO) International Classification for Patient Safety (ICPS) and Agency for Healthcare Research and Quality (AHRQ) Common Formats for Event Reporting–Hospital (version 2.0) [19,20]. The analytical components were informed by literature on the application of SPC in health care, including tutorials on chart construction and reviews of its effectiveness [20-22]. Methods for weighting event severity were adapted from the Institute for Healthcare Improvement (IHI) Severity Assessment Code (SAC) Matrix. This tool is widely used in root cause analysis [23]. To account for known limitations in reporting data, the design was informed by recent systematic reviews that identify common barriers to near miss reporting [6,8]. Conclusively, to ensure contemporary policy relevance, the framework incorporates strategic priorities for safety measurement from key governance documents issued by the WHO and National Health Service England [1,3,19]. Only peer-reviewed publications and official institutional reports were used in the synthesis.

Framework Construction

Overview

The framework was constructed as a staged maturity model with 3 progressive levels, each representing greater analytic depth for interpreting near miss data. Construction followed a sequence of definitional alignment, operational specification, and integration into maturity levels. International taxonomies provided the definitional base, operational rules were derived from methodological literature on SPC and risk assessment, and maturity staging was informed by organizational safety models that emphasize progressive capacity building [18,19,24]. The 3 levels are summarized in Table 1, which presents their scope, core components, and associated formulas. These definitions and operational rules are further detailed and operationalized in Table 2 to ensure direct alignment between theoretical sources and the applied framework design. Together, these tables codify the design specifications of the framework and ensure reproducibility in application.

XSL•FO
RenderX

Table . Summary of the 3 levels of the decision-grade near miss framework[a].

Level	Scope and purpose	Core components	Formulas and calculations
Level 1—essential minimum (organization-wide baseline)	Establish consistent, decision-usable metrics across the whole hospital system	• Standardizes definitions (World Health Organization or Agency for Healthcare Research and Quality taxonomy: near miss, no-harm incident, unsafe condition, and incident) • Records if the event reached the patient • Captures detection mode (human, alarm, patient, and chance) • Collects system-wide numerators (counts) • Uses a single denominator: patient-days or admissions • Applies SPC[b] (u-chart) for rates	• Near miss rate=(near misses÷patient-days)×1000 • Reported near miss–to-harm ratio=reported near misses÷reported harm events • SPC=u-chart (Laney U' if overdispersed)

Level 2—targeted stratification (high-risk or high-volume areas)	Enables hot spot detection in specific clinical units or processes	• Applies domain-specific denominators
		  • ICU[c] or wards: per 1000 patient-days
		  • Pharmacy: per 10,000 doses
		  • Laboratory: per 1000 specimens
		  • Transfusion: per 1000 units
		  • Procedures: per 1000 surgeries or diagnostic studies
		• Generates unit-level SPC charts (interpreted trend over time rather than as league tables)
		• Identifies clusters and recurrent themes
			• Domain near miss rate=(near misses÷domain exposures)×constant
			• Examples
			  • ICU: (near misses÷ICU patient-days)×1000.
			  • Pharmacy: (near misses÷doses dispensed)×10,000
			• Unit-level SPC: separate u-charts by ward or domain

Level 3—advanced decision-grade analytics (mature organizations)	Prioritizes risks by consequence and integrates cultural context for leadership decision-making	• Applies IHI SAC[d] scoring
		• Creates a severity-weighted NMI[e]
		• Tracks corrective action yield
		  • Percentage implemented by 90 days
		  • Percentage sustained at 180 days
		• Triangulates with harm events, reporting volumes, exposure, and safety culture survey results
			• Severity-weighted NMI=Σ(IHI SAC scoring×event count)
			• Learning yield (90 days)=actions implemented÷near misses
			• Learning yield (180 days)=actions sustained÷actions implemented
			• Triangulation=interpretation of NMI, rates, harm events, and culture indicators together

[a]This table presents the staged maturity model used to interpret near miss data. Level 1 (essential minimum) provides an organization-wide baseline using standardized taxonomy, patient-day denominators, near miss rates, near miss–to-harm ratios, and statistical process control with Laney U' adjustment when overdispersion occurs. Level 2 (targeted stratification) adds diagnostic resolution with domain-specific denominators such as intensive care unit patient-days, medication doses, laboratory specimens, transfusion units, or procedures and generates unit-level u-charts interpreted longitudinally. Level 3 (advanced decision-grade analytics) introduces Institute for Healthcare Improvement Severity Assessment Code scoring; a Near Miss Index; learning yield metrics at 90 and 180 days; and triangulation with harm rates, exposure, reporting volume, and safety culture survey data.

[b]SPC: statistical process control.

[c]ICU: intensive care unit.

[d]IHI SAC: Institute for Healthcare Improvement Severity Assessment Code.

[e]NMI: Near Miss Index.

XSL•FO
RenderX

Table . Data elements and operational definitions used across the 3 levels of the near miss framework[a].

Data element	Operational definition	Allowed values or codes	Levels used	Source	Notes or pitfalls
Event type	Classification of an incident according to the World Health Organization or the Agency for Healthcare Research and Quality taxonomy	Near miss, no-harm incident, and harmful incident	L1, L2, and L3	[20,25]	Risk of misclassification if definitions are unclear
Reached patient	An indicator of whether the event reached the patient	Yes or no	L1, L2, and L3	[25]	Must be clearly distinguished from "no-harm incident"
Detection mode	How the event was detected	Human, alarm, patient, and chance	L1, L2, and L3	[25]	Requires training for consistent classification
Unit or domain	Location or process domain in which the event occurred	Intensive care unit, operating room, emergency department, pharmacy, laboratory, and transfusion	L2 and L3	Adapted from quality indicator conventions	Essential for stratification
Denominator type	Exposure measure aligned to the domain	Patient-days, doses, specimens, units, and procedures	L1 and L2	[21,26]	Denominator accuracy depends on data systems
Near miss rate	Standardized rate of near misses	(Near misses÷denominator)×scaling constant	L1 and L2	[21,26]	Must aggregate if denominators are too small
Near miss–to-harm ratio	Ratio of near misses to harmful incidents	Integer ratio	L1	[18,19,24]	Requires reliable coding of harm incidents
Severity score	Potential consequences if the event had not been intercepted	Minor, moderate, major, and catastrophic	L3	[23]	Interrater agreement is needed
Near Miss Index	Aggregate severity-weighted index of near misses	Σ(Institute for Healthcare Improvement Severity Assessment Code×event count)	L3	[23]	Sensitive to scoring reliability
Learning yield (90 days)	Proportion of near misses with actions implemented within 90 days	Action implemented÷near misses	L3	[27]	Requires reliable action tracking
Learning yield (180 days)	Proportion of implemented actions sustained at 180 days	Actions sustained÷actions implemented	L3	[28]	Measurement of sustainability is often weak
Triangulation indicators	Complementary safety indicators for context	Harm rates, reporting volume, and culture scores	L3	[29,30]	Must avoid simplistic correlations

[a]This table defines the data elements required for consistent measurement and interpretation of near miss events. It includes classifications, denominators, severity scoring, and learning metrics, with allowed values, levels of use, sources, and potential pitfalls. L1 (level 1: essential minimum) refers to organization-wide baseline measurement of near misses using standardized taxonomy, patient-day denominators, rates, and statistical process control (SPC) monitoring. L2 (level 2: targeted stratification) applies domain-specific denominators (eg, intensive care unit, pharmacy, laboratory, transfusion, and procedures) and unit-level SPC charts for hot spot detection. L3 (level 3: advanced decision-grade analytics) incorporates severity scoring; the Near Miss Index; learning yields at 90 and 180 days; and triangulation with harm, exposure, reporting volume, and culture indicators.

Level 1: Essential Minimum

The first level establishes a standardized foundation for organization-wide surveillance of near miss events. At this stage, events are classified using the WHO ICPS and the AHRQ's Common Formats, which define the categories of near miss, no-harm incident, and harmful incident [20,25]. Each event is recorded with an indication of whether it reached the patient and with specification of the detection mode, classified as human intervention, automated alarm, patient self-report, or chance discovery. These data elements are defined in Table 2, which provides the operational specifications required at all 3 levels.

For rate construction, the denominator is hospital patient-days, consistent with conventions used for monitoring hospital-acquired conditions such as falls and infections [26].

XSL•FO
RenderX

The calculations specified for this level, summarized in Table 1, are the near miss rate, defined as the number of near misses per 1000 patient-days, and the reported near miss–to-harm ratio, defined as the number of reported near misses divided by the number of reported harmful incidents. Temporal analysis is performed using u-charts, which accommodate variable denominators, and the Laney U' modification is applied when overdispersion is identified, ensuring stable control limits and valid interpretation of signals [21,26,31].

Level 2: Targeted Stratification

The second level adds analytic depth by applying stratification to high-risk or high-volume domains in which the opportunity for near misses is greatest. The framework specifies denominators that align with domain-specific exposures (Table 1): for intensive care, near misses per 1000 intensive care unit patient-days; in pharmacy services, near misses per 10,000 medication doses; in laboratory settings, near misses per 1000 specimens; in transfusion services, near misses per 1000 blood units; and in procedural areas, near misses per 1000 surgeries or diagnostic studies. These denominators are drawn from existing quality indicator conventions and are further defined in Table 2 to ensure consistent application across settings. Organizations must ensure these definitions are applied with high fidelity, as variations in how they are operationalized can compromise consistency and comparability. For example, a "procedure" or a "dispensed dose" can introduce measurement error.

Each domain generates unit-level u-charts that are interpreted longitudinally within the unit rather than across units, reflecting best practice in the application of SPC for quality improvement [21,26]. Where denominators fall below stability thresholds, data are aggregated to quarterly intervals to reduce small-number instability, which is a recognized limitation in attribute charting for health care data [26]. Aggregation applies primarily to low-volume domains such as transfusion or pharmacy services; higher-volume domains (eg, intensive care units or surgical units) maintain monthly analysis to preserve sensitivity to special-cause variation.

Level 3: Advanced Decision-Grade Analytics

Level 3 introduces prioritization and learning metrics. Severity weighting was adapted from the IHI SAC Matrix [23]. Each near miss is scored on a 4-point ordinal scale representing the worst reasonable outcome if not intercepted: minor, moderate, major, or catastrophic. The Near Miss Index (NMI) is calculated by summing severity scores across events in a reporting period. To capture the organizational response, specific learning metrics were defined. However, systematic reviews indicate that only a minority of reported incidents lead to implemented systemic changes, highlighting a critical gap between reporting and learning [27].

Learning yield (90 days) measures the proportion of near misses that generate an implemented corrective action within 90 days of reporting. This time frame is critical for ensuring timeliness; health care literature shows that prolonged delays between incident reporting and visible intervention can erode frontline staff engagement and undermine trust in the safety system [32].

Learning yield (180 days) assesses the proportion of those initial actions that are sustained at the 180-day mark. This follow-up period is a standard in implementation science used to differentiate a temporary, reactive solution from a change that has been successfully embedded into routine clinical practice. Evidence of sustainment should be determined through objective methods such as process audits or analysis of postimplementation performance data [28]. The 90- and 180-day intervals are operational thresholds specified by this framework to operationalize responsiveness and sustainment. Therefore, these metrics align with quality improvement literature by providing quantitative indicators of a functional learning system, measuring both its immediate responsiveness (action closure) and its ability to create lasting change (sustainment) [33]. Importantly, learning yield is a process metric that quantifies the timeliness and persistence of organizational response; it does not measure whether the implemented actions were effective at reducing harm. Demonstrating intervention effectiveness would require outcome-level studies with appropriate controls, which lies beyond the scope of a reporting-based analytic framework. Therefore, learning yield should be interpreted as evidence of a functioning learning loop rather than proof of harm reduction.

Finally, triangulation was embedded to contextualize near miss data. Framework design requires that near miss signals be interpreted alongside harmful incident rates, exposure denominators, reporting volumes, and scores from validated safety culture instruments such as the Safety Attitudes Questionnaire [29] and the Hospital Survey on Patient Safety Culture [30]. Triangulation addresses the limitation of voluntary reporting systems, in which fluctuations may reflect cultural variation more than underlying risk [8].

Ethical Considerations

This study did not involve human participants or patient-level data. All information was derived from peer-reviewed literature and publicly available institutional reports. Therefore, institutional review board approval was not required. Organizations implementing the framework with operational data should obtain local quality improvement or institutional review board determinations as appropriate.

Results

The outcome of this study is a 3-level framework for decision-grade analysis of near miss data. The framework integrates taxonomy, operational definitions, SPC, and severity weighting.

Framework Artifact

The framework begins with the capture of events through 3 complementary input channels: voluntary incident reporting systems, daily safety huddles, and targeted audits. In daily huddles, frontline staff verbally flag safety concerns and near miss events; a designated recorder enters qualifying events into the incident reporting system using the same WHO and AHRQ classification fields, ensuring a single integrated dataset. Audit-identified events, such as medication discrepancies found during pharmacy reconciliation, are likewise entered through

XSL•FO
RenderX

the reporting system. Therefore, all inputs are coded using the WHO and AHRQ classifications to ensure consistency of terms. Data elements include event type, a patient-reached flag, detection mode, and unit or domain. The information is processed through successive levels. The first level produces organization-wide rates and ratios with SPC monitoring. The second level produces domain-specific rates using aligned denominators and unit-level charts. The third level weights events by severity, computes an NMI, and links reporting to corrective actions and culture data. The flow from inputs to outputs, with feedback to learning systems, is depicted in Figure 1.

XSL•FO
RenderX

Figure 1. Framework architecture for decision-grade near miss analytics. Inputs from incident reporting systems, daily safety huddles, and targeted audits are standardized using World Health Organization (WHO) and Agency for Healthcare Research and Quality (AHRQ) definitions of near miss, no-harm incident, and harmful incident, together with event descriptors such as detection mode and whether the patient was reached. Data then progresses through 3 maturity levels. Level 1 (essential minimum) produces hospital-wide near miss rates per 1000 patient-days and near miss–to-harm ratios, monitored with u-charts (control charts for attributes) adjusted using the Laney U' adjustment method when overdispersion is present. Level 2 (targeted stratification) applies domain-specific denominators (intensive care unit [ICU] patient-days, medication doses, laboratory specimens, transfusion units, and procedures) and generates unit-level u-charts that are interpreted longitudinally. Level 3 (advanced analytics) weights near misses by potential severity using the Institute for Healthcare Improvement Severity Assessment Code (IHI SAC) scale and calculates a Near Miss Index (NMI=Σ[severity score×event count]). Near miss data are triangulated with harm rates, exposure measures, reporting volumes, and safety culture survey scores and are linked to organizational learning yields at 90 and 180 days (action implementation and sustainment). Outputs include an organizational dashboard, unit-level trend cards, and quarterly prioritization briefs.

XSL•FO
RenderX

Interpretation Matrix

The framework also serves as a structured aid for leadership interpretation. Patterns in near miss reporting require context. In some instances, near misses increase while harm decreases. This suggests stronger interception and more open reporting. In other cases, near misses increase while total reports remain flat. This indicates better detection in selected domains. A third pattern is when both near misses and overall reports increase. This is more complex and may reflect cultural shifts and emerging system strain. A fourth scenario arises when both near misses and harm decrease simultaneously, which may signal declining reporting engagement rather than genuine safety improvement. A fifth scenario, in which near miss and harm rates remain stable with no SPC signals, may indicate system equilibrium but still warrants periodic review to guard against complacency. Each of these patterns is linked to diagnostic checks such as denominators, SPC variation, harm rates, and culture scores. The matrix is presented in Table 3. It structures trends into observed patterns, diagnostic considerations, and leadership implications.

Table . Interpretation matrix for leadership decision-making[a].

Observed trend	Diagnostic checks	Possible interpretation	Leadership action
Increasing near misses while harm decreases or remains flat	Stability of denominators and SPC[b] signals	Interception and reporting culture have improved	Reinforce effective defenses and promote open reporting
Increasing near misses while overall reporting remains stable	Clustering in specific domains and detection processes in those areas	Detection has improved in selected domains	Investigate process precursors and address system vulnerabilities
Both near misses and overall reports increase	Harm rates, exposure changes, system vulnerabilities, and SPC variation	Reporting culture is improving, but system stress may be rising	Review high-consequence clusters using the Near Miss Index, apply corrective actions, and monitor sustainment
Both near misses and harm decrease	Culture survey trends, reporting volume, and staff engagement data	Reporting fatigue or disengagement rather than true safety improvement	Conduct targeted culture assessment, reinforce nonpunitive reporting, and verify harm data completeness
Both near misses and harm remain flat (no SPC signals)	Denominator stability, process changes, and culture survey trends	The system is at equilibrium, or surveillance sensitivity has plateaued	Conduct periodic review of detection processes and consider proactive audits to test reporting sensitivity

[a]This matrix presents 5 common patterns in near miss reporting and links each pattern to diagnostic checks, possible interpretations, and recommended leadership actions. The aim is to provide consistent decision guidance and reduce the risk of misinterpreting reporting variation as changes in underlying risk.

[b]SPC: statistical process control.

Application and Governance Outputs

A synthetic example was constructed to demonstrate the framework's application, using parameters detailed in Multimedia Appendix 1 [19,23,27,28,31]. In a hypothetical quarter, 50 reported harmful incidents and 10 reported near misses were modeled across 50,000 patient-days. This yields a reported harm rate of 1.0 per 1000 patient-days and a near miss rate of 0.2 per 1000 patient-days. The resulting reported near miss-to-harm ratio is 0.2:1, reflecting the common underreporting of near misses. Severity was then assigned using a weighted scale (minor=1, moderate=2, major=3, and catastrophic=5). The 10 near misses were distributed as 5 minor, 3 moderate, 1 major, and 1 catastrophic, which computes a severity-weighted NMI of 19. Organizational response was modeled over time: at 90 days, 3 corrective actions were implemented for a learning yield of 30% (3/10); at 180 days, 2 of those actions were sustained for a sustainment rate of approximately 67% (2/3). This example illustrates how the NM³ framework converts raw and even sparse reported data into standardized rates, a risk-prioritized index, and actionable governance metrics. The detailed rationale for each parameter is provided in Multimedia Appendix 1.

The framework produces outputs that can be integrated into governance processes. The first level produces a monthly organizational dashboard with standardized rates and SPC signals. The second level provides unit-level reports with stratified denominators and longitudinal trends. The third level delivers a quarterly report with weighted indices, learning yields, and triangulated indicators. Together, these outputs move near miss reporting from descriptive tallies to decision-grade intelligence that supports prioritization and oversight.

Discussion

Principal Findings

This study developed a 3-level framework to strengthen the analysis of near miss data. The framework integrates taxonomy, rate calculation, SPC, severity weighting, and learning metrics. It addresses the current challenge that near miss reporting often produces data but not reliable interpretation. By combining definitional clarity with analytic structure, it converts descriptive data into decision-grade intelligence.

The contribution of this work lies in its ability to unify 3 methodological strands. The first is a standardized classification of events, which ensures that data reflect common definitions rather than local interpretation. The second is the application

XSL•FO
RenderX

of SPC, which provides stable detection of variation over time and guards against overdispersion. The third is the use of severity weighting and learning metrics, which elevate high-consequence signals and link reporting to organizational action. Together, these strands close the gap between reporting and governance.

The NM³ framework can be understood as a decision support architecture within a health care sociotechnical system [15,34,35]. By standardizing how safety events are classified, stabilized over time, weighted by potential consequence, and interpreted in relation to exposure and culture metrics, the framework reduces cognitive ambiguity in governance decision-making. Rather than leaving interpretation to informal heuristics, NM³ provides a shared analytic structure that supports consistent situational awareness across organizational levels [36]. In this way, near miss analytics shift from a descriptive reporting exercise to a structured component of human-centered safety governance [37].

Comparison With Prior Work

Existing approaches often remain descriptive. Most incident reporting systems provide tallies and categorical breakdowns but do not resolve the dilemma of what rising or falling near miss counts mean. The WHO ICPS established a definitional foundation [25], but it does not specify analytic processes. The structure-process-outcome model by Donabedian [38] underpins quality measurement broadly, but it does not address the specific problem of near miss interpretation. More recent frameworks, such as the Value Transformation Framework in primary care [39], focus on organizational change at the system level rather than the metrics of event reporting. Therefore, this study adds a new lens that sits between operational reporting and strategic governance.

The framework also supports practical adoption. Levels are designed as a maturity pathway. Organizations at baseline can implement level 1 by adopting a taxonomy and calculating standardized rates with simple u-charts. More advanced organizations can progress to level 2, where stratification by domain exposes hot spots without generating cross-unit competition. The most mature can reach level 3, where severity weighting, NMIs, and learning yields allow prioritization of risk and assessment of system responsiveness. In this way, the framework adapts to capacity rather than demanding wholesale transformation at once. In terms of scale, level 1 analytics can be applied by any hospital with a functioning incident reporting system. For level 2 stratification, sufficient denominator volume is needed to produce stable u-charts at the unit level; where monthly event counts are very low, such as in small community hospitals, a quarterly aggregation, as specified in the framework, preserves analytic validity. Level 3 severity weighting and learning yields are most informative when the organization reports at least several dozen near misses per quarter, providing enough events for meaningful index calculation and trend detection.

For governance, the framework strengthens the link between frontline reporting and board oversight. Dashboards at level 1 provide stable high-level indicators. Stratified reports at level 2 support managers who need to identify domain-specific risks. The advanced analytics of level 3 help leaders focus on high-consequence clusters and track whether actions are sustained. This aligns with international calls for safety metrics that are not only valid but actionable.

Future work must test and refine the framework. Empirical studies should evaluate its feasibility in practice, assess the reliability of severity scoring, and compare outputs across institutions. Research should also explore how digital systems can automate rate calculation, control chart production, and NMI computation. Benchmarking across hospitals and international contexts will determine generalizability and adaptation needs. Such validation will establish whether the framework produces consistent intelligence and supports decisions that improve safety outcomes. Furthermore, research should explore the integration of qualitative data from near miss narratives with the quantitative outputs of the NM³ framework to provide richer contextual understanding.

Limitations

This framework has important limitations. It depends on voluntary reporting systems that are known to undercount events substantially; estimates suggest that voluntary systems capture as few as 5% to 10% of patient safety incidents, with near misses particularly underrepresented [7,9]. Reporting behavior is shaped by culture, workload, and fear, and these influences may distort numerators. Denominator data also present challenges. Accurate counts of patient-days, doses, or procedures are not always available or consistent across systems. Errors in denominators can lead to misleading rates.

Event classification introduces uncertainty at level 1: even with standardized WHO and AHRQ definitions, assigning the correct event type depends on individual reviewer judgment, and variability in interpretation may introduce classification error. At level 3, severity scoring presents a similar challenge; the IHI SAC score provides structure, but interrater reliability was not tested here. Training, calibration exercises, and periodic agreement checks are recommended for both event classification and severity scoring. Learning yield metrics depend on robust action tracking, which is not consistently available in many organizations.

Finally, this study is conceptual. The framework was derived from literature and policy rather than empirical testing. Its generalizability across different health care contexts remains uncertain until validation is undertaken. Pilot applications will be necessary to confirm feasibility, reliability, and value in governance. Implementation also presupposes dedicated human resources and analytic expertise. This may not be available, particularly for level 2 stratification and level 3 severity weighting in resource-constrained settings, and organizations should assess workforce readiness before progression beyond level 1.

Implications

This framework has direct implications for health care organizations, quality leaders, and researchers. For organizations, the staged design lowers the barrier to adoption by allowing hospitals to begin with standardized baseline measurement and progress to more advanced analytics as

XSL•FO
RenderX

institutional capacity grows, rather than requiring wholesale transformation at the outset.

For leadership, NM³ converts descriptive reporting into actionable intelligence that supports evidence-based governance. The tiered outputs provide a common language for safety oversight and enable boards to distinguish changes in reporting culture from genuine shifts in underlying risk, aligning with international calls for actionable patient safety metrics.

For researchers, NM³ provides a reproducible conceptual model that can be operationalized and validated across diverse health care contexts. Adoption in varied settings will help establish its role in patient safety improvement and determine the conditions under which the framework yields the greatest analytic value.

Conclusions

This study developed a 3-level framework for decision-grade interpretation of near miss data within acute hospital systems. The framework integrates taxonomy, SPC, severity weighting, and organizational learning metrics. It addresses the ambiguity that has long limited the use of near miss reports in governance.

By structuring analysis into progressive levels, the framework functions as a cognitive and organizational scaffold for interpreting safety signals in complex sociotechnical environments. It enables leaders to distinguish culture shifts from genuine changes in risk and to prioritize high-consequence signals. Adoption has the potential to strengthen human-centered safety governance and support proactive risk mitigation.

Funding

The authors declare that no financial support was received for this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Rationale for synthetic example parameters.
[DOCX File, 29 KB - humanfactors_v13i1e87846_app1.docx]

References

1. Global patient safety action plan 2021-2030. : World Health Organization; 2021 URL: https://www.who.int/publications/i/item/9789240032705 [accessed 2026-04-15]
2. Global patient safety report 2024. : World Health Organization; 2024 URL: https://www.who.int/publications/i/item/9789240095458 [accessed 2026-04-15]
3. Patient safety incident response framework. National Health Service England. URL: https://www.england.nhs.uk/patient-safety/patient-safety-insight/incident-response-framework/ [accessed 2026-04-15]
4. Health at a glance 2023: OECD indicators. Organisation for Economic Co-operation and Development. 2023. URL: https://www.oecd.org/en/publications/2023/11/health-at-a-glance-2023_e04f8239.html [accessed 2026-04-15]
5. de Bienassis K, Klazinga N. Comparative assessment of patient safety culture performance in OECD countries. : Organisation for Economic Co-operation and Development; 2024 URL: http://oecd.org/en/publications/comparative-assessment-of-patient-safety-culture-performance-in-oecd-countries_d0552328-en.html [accessed 2026-04-15]
6. Oweidat I, Al-Mugheed K, Alsenany SA, Abdelaliem SMF, Alzoubi MM. Awareness of reporting practices and barriers to incident reporting among nurses. BMC Nurs 2023 Jul 3;22(1):231. [doi: 10.1186/s12912-023-01376-9] [Medline: 37400810]
7. Pham JC, Girard T, Pronovost PJ. What to do with healthcare incident reporting systems. J Public Health Res 2013 Dec 1;2(3):e27. [doi: 10.4081/jphr.2013.e27] [Medline: 25170498]
8. Alfayez A, Althumairi A, Aljuwair M, Althukair D, Aljabri D. Factors affecting patient safety near miss reporting: a systematic review. J Adv Nurs 2026 Mar;82(3):2022-2035. [doi: 10.1111/jan.70033] [Medline: 40574456]
9. Hoops K, Pittman E, Stockwell DC. Disparities in patient safety voluntary event reporting: a scoping review. Jt Comm J Qual Patient Saf 2024 Jan;50(1):41-48. [doi: 10.1016/j.jcjq.2023.10.009] [Medline: 38057189]
10. Caspi H, Perlman Y, Westreich S. Managing near-miss reporting in hospitals: the dynamics between staff members' willingness to report and management's handling of near-miss events. Saf Sci 2023;164:106147. [doi: 10.1016/j.ssci.2023.106147]
11. Waqas M, Xu SH, Hussain S, Aslam MU. Control charts in healthcare quality monitoring: a systematic review and bibliometric analysis. Int J Qual Health Care 2024 Jul 19;36(3):mzae060. [doi: 10.1093/intqhc/mzae060] [Medline: 39018022]
12. Arafah M. Using the Laney p' control chart for monitoring COVID-19 cases in Jordan. J Healthc Eng 2022;2022:6711592. [doi: 10.1155/2022/6711592] [Medline: 36247090]
13. Woodier N, Burnett C, Sampson P, Moppett I. Patient safety near misses – still missing opportunities to learn. J Patient Saf Risk Manag 2024;29(1):47-53. [doi: 10.1177/25160435231220430]

XSL•FO
RenderX

14. Tresfon J, van Winsen R, Brunsveld-Reinders AH, Hamming J, Langeveld K. Hospital ward incidents through the eyes of nurses – a thick description on the appeal and deadlock of incident reporting systems. Saf Sci 2025 Apr;184:106728. [doi: 10.1016/j.ssci.2024.106728]
15. Carayon P, Schoofs Hundt A, Karsh BT, et al. Work system design for patient safety: the SEIPS model. Qual Saf Health Care 2006 Dec;15 Suppl 1(Suppl 1):i50-i58. [doi: 10.1136/qshc.2005.015842] [Medline: 17142610]
16. Maitlis S. The social processes of organizational sensemaking. Acad Manag J 2005 Feb;48(1):21-49. [doi: 10.5465/amj.2005.15993111]
17. Saposnik G, Redelmeier D, Ruff CC, Tobler PN. Cognitive biases associated with medical decisions: a systematic review. BMC Med Inform Decis Mak 2016 Nov 3;16(1):138. [doi: 10.1186/s12911-016-0377-1] [Medline: 27809908]
18. Jabareen Y. Building a conceptual framework: philosophy, definitions, and procedure. Int J Qual Methods 2009;8(4):49-62. [doi: 10.1177/160940690900800406]
19. Runciman W, Hibbert P, Thomson R, Van Der Schaaf T, Sherman H, Lewalle P. Towards an International Classification for Patient Safety: key concepts and terms. Int J Qual Health Care 2009 Feb;21(1):18-26. [doi: 10.1093/intqhc/mzn057] [Medline: 19147597]
20. COMMON FORMATS: hospital version 2.0. AHRQ PSO Privacy Protection Center (PSOPPC). URL: https://www.psoppc.org/common-formats/hospital-2.0 [accessed 2026-04-15]
21. Mohammed MA, Worthington P, Woodall WH. Plotting basic control charts: tutorial notes for healthcare practitioners. Qual Saf Health Care 2008 Apr;17(2):137-145. [doi: 10.1136/qshc.2004.012047] [Medline: 18385409]
22. Mohammed MA. Using statistical process control to improve the quality of health care. Qual Saf Health Care 2004 Aug;13(4):243-245. [doi: 10.1136/qhc.13.4.243] [Medline: 15289621]
23. RCA2: improving root cause analyses and actions to prevent harm. Institute for Healthcare Improvement. URL: https://www.ihi.org/library/tools/rca2-improving-root-cause-analyses-and-actions-prevent-harm [accessed 2026-04-15]
24. Vincent C, Burnett S, Carthey J. The measurement and monitoring of safety: drawing together academic evidence and practical experience to produce a framework for safety measurement and monitoring. The Health Foundation. 2013. URL: https://www.health.org.uk/sites/default/files/TheMeasurementAndMonitoringOfSafety_fullversion.pdf [accessed 2026-04-15]
25. The conceptual framework for the international classification for patient safety. : World Health Organization; 2009 URL: https://www.who.int/publications/i/item/WHO-IER-PSP-2010.2 [accessed 2026-04-15]
26. Thor J, Lundberg J, Ask J, et al. Application of statistical process control in healthcare improvement: systematic review. Qual Saf Health Care 2007 Oct;16(5):387-399. [doi: 10.1136/qshc.2006.022194] [Medline: 17913782]
27. Goekcimen K, Schwendimann R, Pfeiffer Y, Mohr G, Jaeger C, Mueller S. Addressing patient safety hazards using critical incident reporting in hospitals: a systematic review. J Patient Saf 2023 Jan 1;19(1):e1-e8. [doi: 10.1097/PTS.0000000000001072] [Medline: 35985209]
28. Proctor E, Silmere H, Raghavan R, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76. [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
29. Sexton JB, Helmreich RL, Neilands TB, et al. The Safety Attitudes Questionnaire: psychometric properties, benchmarking data, and emerging research. BMC Health Serv Res 2006 Apr 3;6:44. [doi: 10.1186/1472-6963-6-44] [Medline: 16584553]
30. Smits M, Christiaans-Dingelhoff I, Wagner C, Wal GV, Groenewegen PP. The psychometric properties of the “Hospital Survey on Patient Safety Culture” in Dutch hospitals. BMC Health Serv Res 2008 Nov 7;8:230. [doi: 10.1186/1472-6963-8-230] [Medline: 18990256]
31. Mohammed MA, Laney D. Overdispersion in health care performance data: Laney’s approach. Qual Saf Health Care 2006 Oct;15(5):383-384. [doi: 10.1136/qshc.2006.017830] [Medline: 17074879]
32. Patient safety incident reporting and learning systems: technical report and guidance. : World Health Organization; 2020 URL: https://www.who.int/publications/i/item/9789240010338 [accessed 2026-04-15]
33. Dixon-Woods M, McNicol S, Martin G. Ten challenges in improving quality in healthcare: lessons from the Health Foundation’s programme evaluations and relevant literature. BMJ Qual Saf 2012 Oct;21(10):876-884. [doi: 10.1136/bmjqs-2011-000760] [Medline: 22543475]
34. Holden RJ, Carayon P, Gurses AP, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 2013;56(11):1669-1686. [doi: 10.1080/00140139.2013.838643] [Medline: 24088063]
35. Ladell M, Yale S, Bordini BJ, Scanlon MC, Jacobson N, Papautsky EL. Why a sociotechnical framework is necessary to address diagnostic error. BMJ Qual Saf 2024 Nov 20;33(12):823-828. [doi: 10.1136/bmjqs-2024-017231]
36. Endsley MR. Toward a theory of situation awareness in dynamic systems. Hum Factors 1995;37(1):32-64. [doi: 10.1518/001872095779049543]
37. Kushniruk A, Kaufman D. Human factors and organizational issues in health informatics: review of recent developments and advances. Yearb Med Inform 2024 Aug;33(1):196-209. [doi: 10.1055/s-0044-1800744] [Medline: 40199306]
38. Donabedian A. The quality of care. How can it be assessed? JAMA 1988;260(12):1743-1748. [doi: 10.1001/jama.260.12.1743] [Medline: 3045356]

XSL•FO
RenderX

39. Teisberg E, Wallace S, O'Hara S. Defining and implementing value-based health care: a strategic framework. Acad Med 2020 May;95(5):682-685. [doi: 10.1097/ACM.0000000000003122] [Medline: 31833857]

Abbreviations

AHRQ: Agency for Healthcare Research and Quality
ICPS: International Classification for Patient Safety
IHI: Institute for Healthcare Improvement
NMI: Near Miss Index
NM³: 3-level near miss framework
SAC: Severity Assessment Code
SPC: statistical process control
WHO: World Health Organization

Edited by A Kushniruk; submitted 15.Nov.2025; peer-reviewed by A Vannucci, C Recsky; revised version received 20.Feb.2026; accepted 31.Mar.2026; published 22.Apr.2026.

Please cite as:
As'ad M
Near Miss Reporting and Organizational Learning in Health Care: Conceptual Framework Development Study
JMIR Hum Factors 2026;13:e87846
URL: https://humanfactors.jmir.org/2026/1/e87846
doi:10.2196/87846

© Mohammed As'ad. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 22.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

"To Choose a Band-Aid" - Children's Preferences for Participation in Health Care Situations: Q Methodology Study

Merja Hietanen[1,2], BN, MMSc; Maksims Kornevs[2], PhD; Catarina Nahlén Bose[1], PhD; Jayanth Raghothama[2], PhD; Sebastiaan Meijer[2], PhD

[1]Swedish Red Cross University, Hälsovägen 11 C, Huddinge, Stockholm, Sweden

[2]KTH Royal Institute of Technology, Huddinge, Stockholm, Sweden

Corresponding Author:
Merja Hietanen, BN, MMSc
Swedish Red Cross University, Hälsovägen 11 C, Huddinge, Stockholm, Sweden

Abstract

Background: Child-centered care (CCC) is standard practice in pediatrics, emphasizing the child as an individual with rights while acknowledging the child's role within the family. A key aspect of CCC is the involvement of the child in health care decisions alongside parents and professionals. Although this is a right recognized by the United Nations Convention on the Rights of the Child, it may not always be applied in practice.

Objective: The aim of this study is to explore the participation preferences of 3- to 5-year-old children in health care from both their own viewpoint (the child's perspective) and the viewpoint of their parents and health care professionals.

Methods: Q methodology was used to study preferences, comparing responses from 12 children, 14 parents, and 12 health professionals who ranked 25 statements about ways in which children could participate in health care situations. Factor analysis was used to reveal similarities and differences in views on participation preferences. The children's rankings were also analyzed separately for comparison.

Results: Analysis of rankings from children, their parents, and health care professionals identified 3 perspectives with different preferences: direct communication between the child and healthcare professionals; understanding and shared decision-making; and responsive and child-led participation. A separate analysis of children's rankings resulted in 3 perspectives: being included in and setting their own terms for participation; small choices, meaningful outcomes; and trust through familiarity and shared decision-making.

Conclusions: This study suggests that children value shared decision-making and situational control but prefer to leave major decisions to adults. It affirms that preschool-aged children can meaningfully participate in health care when given age-appropriate choices, support, and tools. Children's perspectives must be acknowledged directly rather than adults assuming their views. The findings support CCC principles and reinforce the United Nations Convention on the Rights of the Child mandate to respect children's views regarding all issues relevant to them.

(JMIR Hum Factors 2026;13:e89802) doi:10.2196/89802

KEYWORDS

children; child participation; child's perspective; health care; Q methodology

Introduction

Child-centered care (CCC) is the standard focus of care in pediatrics [1]. This concept recognizes the individual child at the center of care as a social agent with rights, but also as part of a family [1,2]. The involvement of children in their health care concerns and decisions alongside parents and health care professionals (HCPs) is central to CCC [2]. Child participation can be viewed from 2 different perspectives: the child's perspective and the child perspective [3], which are succinctly different in the person formulating the perspective. The child's perspective refers to a subjective view of a child's understanding of the world through their perceptions and experiences [4]. The child perspective looks at the same situation through the adult's perspective by trying to find the child's point of view, but it is still an adult's interpretation of the child [4]. Both perspectives must be considered when implementing CCC [5,6], including the adult's view of the child's best interest and the child's own preferences for the situation [6].

Despite its common usage, the term "participation" has different meanings in different academic and professional disciplines and settings. It can refer to anything from physical presence to activity, competence, and a sense of self [7]. Participation can also mean children's ability to form and express their opinions and influence matters that affect them [8], which is in line with Article 12 of the United Nations Convention on the Rights of

XSL•FO
RenderX

the Child (UNCRC) [9]. While the declaration does not specify an age limit for children's participation, it does state that their age and maturity should be taken into account. Their opinions should be taken seriously and not dismissed due to their age [9]. In this study, the participation of preschool- and school-aged children in health care situations is defined as deciding how to act, receiving adequate information, making decisions, as well as communicating with HCPs [10].

Previous research shows that children are competent agents who can construct their social worlds in collaboration with others [11,12]. They want to be involved in decision-making processes [13,14], yet they are often unable to participate or express their views [11,15,16]. Parents and HCPs play an important role in facilitating children's participation [10]. However, research has found that parents and HCPs limit children's opportunities to participate in health care [15,17,18], their opinions are not asked for [15], there is no agreement on the level of children's participation [11], or parents protect their children from decision-making instead of allowing shared decision-making [19]. A recent review revealed a knowledge gap concerning preschool-aged children's preferences for participation in health care situations [10]. In accordance with their rights as stated in the UNCRC [9], more research should be conducted into children's perspectives and participation in health care situations, since this area has not been sufficiently explored. Therefore, it is important to better understand how children prefer to participate in health care situations.

The Swedish National Child Health Care (CHC) program provides free, voluntary health care for children from birth to 6 years of age [20]. This study was conducted in Child Welfare Centers, which are part of the program. It explores the ways in which children prefer to participate in health care situations, focusing on preschool-aged children (3 - 5 years old). It examines the topic from 2 perspectives: the child's perspective, that is, the views of the children themselves, and from the child perspective, that is, the views of their parents and HCPs. It explores differences in views expressed by the children themselves, their parents, and HCPs. The research aims to answer the following question: what are the most important ways of participating in health care from the child's perspective and the child perspective?

Methods

Design

The design uses the mixed methods approach of Q methodology. This methodology was developed to examine similarities and differences in participants' views on a specific issue [20,21]. Participants rank statement cards that present different viewpoints on a given issue. Each individual ranking is then compared with the rankings of the other participants using factor analysis. Participants with common views are identified for the same factor [20]. Q methodology studies commonly use approximately 40 statements, but the number of statements can vary from less than 20 to more than 200 in studies with adult participants [22-24]. It is more important to cover all possible opinions on the topic in the statements than to have a fixed number of statements [23].

The Q set for this study, consisting of 25 statements for ranking, was formed based on the results of a literature review on children's participation in health care situations [10]. Initially, 95 statements were formed from the themes of the literature review in English. Some of the statements had similar content and needed to be simplified to make them more accessible to 3- to 5-year-olds. The statements were discussed among all the authors, and duplicate or overlapping statements were removed. Once all the themes from the literature review had been incorporated into the statements, the authors agreed on a set of 25 statements. These statements were the same for all participants but were adapted in their formulation and pictorial representation to reflect the 3 different perspectives from which they would respond: child, parent, and HCP (Table 1). The statements were translated into Swedish by the authors. One of the authors speaks Swedish as a first language, while another speaks English as a first language. The translations of the questions, Q statements, and, later in the data analysis phase, the transcripts were checked and verified by all the authors, who are fully fluent in both languages.

XSL•FO
RenderX

Table . Q statements.

Number	Statements for children	Statements for parents	Statements for health care professionals
1	I want the doctor and nurse to talk to me too	My child wants the doctors and nurses to talk directly to her or him	Children want me to talk directly to them
2	I want the doctor or nurse to use words which I can understand	Doctor or nurse should use words that my child will understand	I should use words the child understands
3	I want to answer questions from doctor or nurse	My child wants to answer questions from the doctor or nurse	Children want to answer questions from me
4	I want to ask questions to the doctor or nurse	My child wants to ask questions from doctor or nurse	Children want to ask me questions
5	I can give my answers by pointing or nodding	My child can use nonverbal communication instead of talking to the doctor or nurse	Children can use nonverbal communication during the visit
6	I want parents and doctor or nurse to make decisions about me	My child wants me to make decisions with doctors and nurses	Children want the parent and me to make decisions during the visit
7	I want to make decisions except for difficult issues	My child wants to make the decisions, except for the difficult issues	Children want to make decisions, except for the difficult issues
8	I have important things to say before decisions are made about me	My child has important things to say before doctor or nurse and I decide on something about her or him.	Children have important information to share before the parent and I make decisions about them
9	As I grow older, I can make decisions on things concerning me	When my child grows older, he or she can decide on matters concerning her or him	When children grow older, they can decide on matters concerning them
10	I should be asked when decisions are made about me	My child should be asked before decisions are made about her or him	Children should be asked before decisions are made about them
11	I want to decide what to do when waiting for the doctor or nurse	My child wants to decide what he or she wants to do during the waiting time for the doctor or nurse	Children want to decide what they want to do during the waiting time to doctor or nurse
12	When I have answered many questions, I want to play	My child stops answering questions when he or she feels he has participated enough	Children stop answering questions when they feel they have participated enough.
13	I will resist things I do not want to do	My child will protest when he or she does not want to participate anymore	Children protest against procedures when they do not want to participate anymore
14	I want to do something else when the doctor or nurse speaks	My child wants to concentrate on something else during the health care visit when I talk with the doctor or nurse	Children want to focus on something else during the health care visit when I discuss with the parent
15	I want to quietly listen to doctor or nurse	My child wants to listen to the doctor or nurse but wants to stay quiet during the visit	Children want to listen to me but stay quiet during the health care visit
16	I want to talk to the doctor or nurse only when my parent is present	My child wants to talk to the doctor or nurse only when I am present	Children talk to me only when the parent is present
17	I want the doctor or nurse to listen to me too	My child wants the doctor or nurse to listen to her or him also	Children want me to listen to them also
18	I want my parents to let me talk like I want with the doctor or nurse	My child wants me to let her or him speak freely with doctor or nurse	Children want to speak freely with me, without the parent interrupting
19	I feel good when I make decisions together with my parents and doctor or nurse	My child feels that she or he is participating greatly when she or he decides something together with me and doctor or nurse	Children feel that they are participating greatly when they decide something together with their parents and me
20	I want the doctor or nurse to speak to me in a nice way	My child wants the health care staff to talk to her or him in a nice way	Children want the staff to talk to them in a nice way

XSL•FO
RenderX

Number	Statements for children	Statements for parents	Statements for health care professionals
21	I want to have choices from which to choose from during my health care visit	My child wants to have choices to choose from during the health care visit	Children want to have choices to choose from during the health care visit
22	I want to think before answering questions	My child needs time before answering questions from doctor or nurse	Children want to have time to think before answering questions
23	I want to make decisions together with the adults (my parents and doctor or nurse)	My child wants to make decisions together with me and doctor or nurse	Children want to make decisions together with their parents and me
24	I like to meet the same nurse or doctor every time I come to the health care center	My child wants to meet the same staff every time we come for a visit	Children want to meet the same staff members at every health care visit
25	I want the doctor or nurse to ask me before they are going to do something to me	My child wants doctor or nurse to ask her or him before they do any procedures with her or him	Children want me to ask them before doing any procedures

Parents and HCPs were asked to respond on behalf of their child or patient, indicating their preferences. To illustrate the meaning of the statements, Microsoft Bing Copilot (version 2024 May) was used to draw pictures for the children's cards (Figure 1). The content of the images and the desired meaning were discussed jointly by the authors during the development of the images. The first author also conducted a preliminary test with a convenience sample of 3 children of the appropriate age for the study to ascertain the images conveyed the intended meaning. The final selection was then made and agreed upon by all authors.

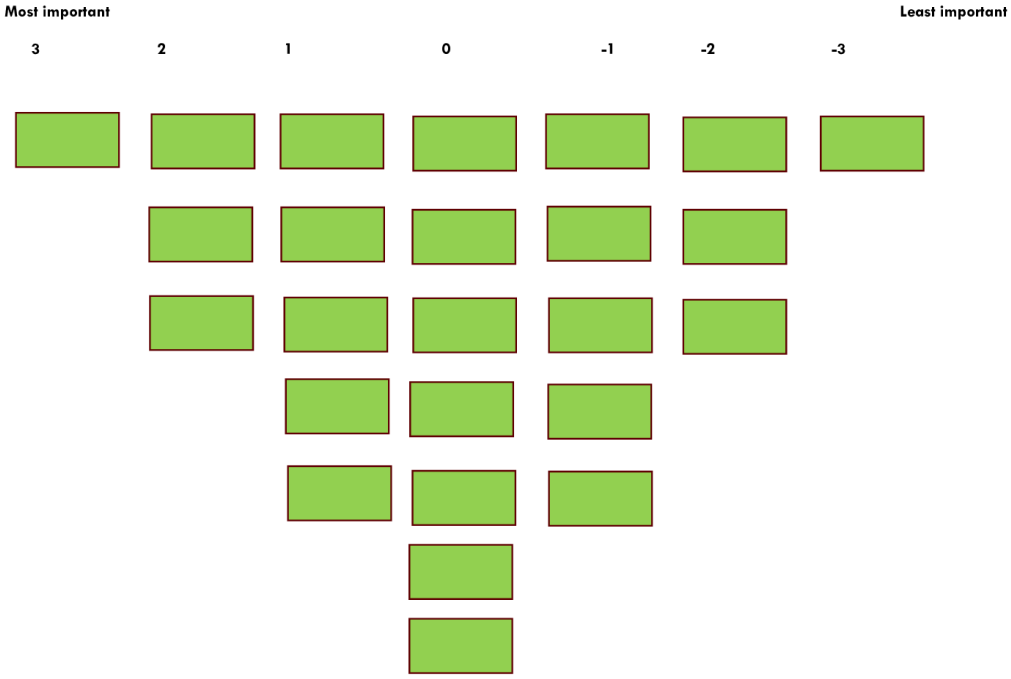
Figure 1. Q statement card for children.



The Q grid, a fixed pattern in the form of a normal distribution for ranking, was chosen to range from least important (−3) to most important (+3), with the extremes of the pattern allowing only 1 statement (Figure 2).

XSL•FO
RenderX

Figure 2. Q grid.



After the participants had ranked the statements, postsorting interviews were conducted for all participants (Textbox 1). The postsorting interviews were conducted to let the participants provide rationale behind the Q sorting and to express their feelings and opinions while conducting the Q sorting [20,25].

Textbox 1. Postsorting interview guide.

Why did you choose to rank the statements in this order?
Why do you think that statement X is more important than statement Y?
Why do you think that statement X is the most or least important?
Are there other ways of involving children that are not mentioned in the statements?

Setting

This study focuses on Child Welfare Centers in Stockholm Region, Sweden. These centers form a part of the national CHC program, which provides free, voluntary health services for children aged 0 to 6 years. The program is widely used and supports child development, parenting, and health promotion [26]. CHC nurses conduct regular health check-ups, including growth monitoring, vision, and balance tests. They also use activities such as drawing and memory games to assess motor skills, communication, and social interaction [27,28]. The goal is to support families and identify the need for further interventions for children [27].

Participants

The participants were recruited via child welfare centers. Three child welfare centers in the Stockholm region participated in the study. One of the centers took part solely through staff interviews in order to obtain the required number of staff members for the study. Nurses from the other 2 centers recruited the child and parent participants through purposive sampling. Participants had to be a child aged between 3 and 5 years, or a parent of a child aged between 3 and 5 years, or a nurse or doctor working with children in this age group in a child welfare center, and they had to be able to understand and speak Swedish. Information about the study was sent to potential participants (children and their parents) prior to their appointment. This included a simplified version with pictures for the child. If a parent agreed to participate in the study, or to allow their child to participate, the nurses contacted the first author to arrange an appointment. Some parents were also asked during the appointment if they wanted to participate in the study, as the nurses and the first author agreed on days when the first author would wait for potential participants at the child welfare center.

XSL•FO
RenderX

The required number of participants was 12 or more in each category. There were no limits on how many participants of a certain age or with a certain amount of work experience needed to be recruited. Seventeen children agreed to be interviewed. However, only 12 of them could complete the sorting task and the subsequent interview. They did not complete the sorting for various reasons: they stopped answering questions, started playing instead of continuing the sorting, or could not concentrate on the questions and directed their attention elsewhere. None of the 3-year-olds (3 persons) who agreed to be interviewed could complete the sorting task. Fourteen parents agreed to be interviewed, but 1 decided not to complete the postsorting interview. Twelve HCPs agreed to be interviewed and completed the sorting and postsorting interviews. This resulted in a convenience sample of 38 participants who were willing to take part.

Data Collection

If a parent agreed to participate in the study, the first author was introduced to them and their child by nurses. Statement ranking and postsorting interviews were conducted immediately after the child's appointment, in a waiting or appointment room at the child welfare center and away from other patients or visitors. Nurses and doctors also conducted the ranking and interview at the child welfare center when their schedules permitted.

All participants were asked to provide the following background information: the age and gender of the child, the age group and gender of the parent, and the work experience and gender of the staff member. Children and parents were also coded together to enable a comparison between how they sorted items.

The child and parent participants completed the Q sorting exercise simultaneously, sitting side by side. The first author verbally explained the sorting procedure to the participants and also provided them with written information about the study. The first author then read and showed the Q statement cards with pictures to the child, while the parent sorted their statements independently. All participants received the same Q statements but written from the different perspectives of the participants (Table 1). After sorting, the final Q grid with the statement cards was photographed for analysis. Participants were interviewed about their motivations for sorting. These postsorting interviews were audio-recorded for the adult participants. The entire ranking situation, including the sorting and postsorting interview, was audio-recorded for pediatric participants. These interviews were transcribed verbatim in Swedish and, during the analysis, translated into English by the authors.

Ethical Considerations

This study does not fall under the Swedish Ethics Review Act since the study does not collect sensitive data and is nonexperimental. However, the Swedish Ethical Review Authority has given an advisory opinion that there are no ethical objections to this study (Dnr 2024-02308-01). The study follows the ethical principles outlined in the Declaration of Helsinki [29]. Participation in the study was voluntary. Adult participants signed a consent form for their participation and an additional form if their child was participating. The children were verbally asked for their assent and were informed that they could withdraw from the study at any time. This study concerns children's preferences and does not collect any identifying information. Any such information provided by participants during postsorting interviews was omitted or anonymized during transcription. Participants were given a cinema ticket as a thank you for taking part.

Analysis

The factor analysis was carried out using the PQMethod software [30], a statistical program. First, data from the ranking of statements were entered into the PQMethod software. Centroid factor analysis, principal component factor analysis, and varimax rotation of factors were performed for each individual sorting. Then, based on the factors obtained, all participants were compared with each other. The common views on each factor were identified by distinguishing the preferred participation methods that the participants in the factor felt most strongly about. These extremes of +3, +2, −2, and −3 were then analyzed to characterize the common views of the factor. The participants associated with the factor were then identified. Several assessments were performed with a variable number of factors, and the optimal balance for most participants was identified when the number of factors was 3. This gave a good balance of most participants associated with the factors. The resulting factors included most adult participants, but only 6 out of the 12 children were identified to be associated with the factors. Therefore, the children's Q sorts were also analyzed separately using the same factor analysis method and analysis of common views characterizing the different factors. Finally, the participants' motivations for sorting from the postsorting interviews were used to corroborate and clarify their preferences for child participation.

Results

Participant Characteristics and Analysis Overview

The sample included 38 participants: 12 children, 14 parents, and 12 HCPs (see Table 2 for participant characteristics). The results of comparing the Q sorts of children, parents, and HCPs are presented as 3 factors representing the children's different preferences for participation, followed by 3 factors from by-person factor analysis of the children's Q sorts only. The Q sort values by factor are presented in Table 3. Finally, the alternative ways of participating, as suggested by the participants during the postsorting interview, are presented.

XSL•FO
RenderX

Table . Participant characteristics.

Participant group and characteristics	Factor all[a] 1	Factor all 2	Factor all 3	Factor children[b] 1	Factor children 2	Factor children 3
Children						
Age (y)						
5	1	1	1	2	2	1
4	0	2	1	1	2	2
3	0	0	0	0	0	0
Gender						
Boys	1	1	1	2	2	1
Girls	0	2	1	1	2	2
Parents						
Age range (y)						
26 - 35	4	0	2	N/A[c]	N/A	N/A
36 - 45	3	1	0	N/A	N/A	N/A
46 - 55	1	0	0	N/A	N/A	N/A
Age not given	1	0	0	N/A	N/A	N/A
Gender						
Male	3	0	1	N/A	N/A	N/A
Female	6	1	1	N/A	N/A	N/A
Health care professionals						
Work experience with children (y)						
<2	1	0	0	N/A	N/A	N/A
2 - 5	2	1	0	N/A	N/A	N/A
5 - 10	1	0	1	N/A	N/A	N/A
10 - 15	0	1	0	N/A	N/A	N/A
>15	1	3	1	N/A	N/A	N/A
Gender						
Female	5	5	2	N/A	N/A	N/A
Male	0	0	0	N/A	N/A	N/A

[a]"Factor all" refers to the factor analysis of all participants' rankings.

[b]"Factor children" refers to the factor analysis of children's rankings only.

[c]Not applicable.

XSL•FO
RenderX

Table . Factor Q sort values for each statement.

Number	Statement	Factor all[a] 1	Factor all 2	Factor all 3	Factor children[b] 1	Factor children 2	Factor children 3
1	I want the doctor and nurse to talk to me too	3	0	−2	−1	−1	−2
2	I want the doctor/nurse to use words which I can understand	2	2	0	−1	0	0
3	I want to answer questions from doctor/nurse	0	0	−2	2	1	0
4	I want to ask questions from doctor/nurse	1	−2	0	0	−1	−1
5	I can answer by pointing or nodding	0	1	2	0	0	−3
6	I want parents and doctor/nurse to take the decisions over me	−1	0	0	1	2	0
7	I want to make decisions except for difficult issues	−2	1	−1	1	2	0
8	I have important things to say before decisions are taken about me	0	1	3	0	0	0
9	As I grow older I can take decisions on things concerning me	1	−1	1	−3	1	0
10	I should be asked when decisions are made about me	0	−1	0	2	−2	0
11	I want to decide what to do when waiting for the doctor/nurse	−1	−1	−1	−1	0	1
12	When I have answered many questions I want to play	−1	−1	−1	2	1	0
13	I will resist things I do not want to do	1	0	2	0	−3	−2
14	I want to do something else when the doctor/nurse speaks	−2	0	−3	1	1	−2
15	I want to quietly listen to doctor/nurse	−3	−2	0	0	−2	0

XSL•FO
RenderX

Number	Statement	Factor all[a] 1	Factor all 2	Factor all 3	Factor children[b] 1	Factor children 2	Factor children 3
16	I want to talk to the doctor/nurse only when my parent is present	−2	−3	−1	1	−1	0
17	I want the doctor/nurse to listen to me too	2	1	0	−1	2	−1
18	I want my parents to let me talk like I want with the doctor/nurse	−1	−2	0	0	−1	−1
19	I feel good when I make decisions together with my parents and doctor/nurse	0	3	0	1	0	2
20	I want the doctor/nurse to speak to me in a nice way	2	2	0	−2	0	0
21	I want to have choices from which to choose from during my health care visit	0	0	−2	0	3	−1
22	I want to think before answering questions	0	−1	0	−1	1	−3
23	I want to make decisions together with my parents and doctor/nurse	−1	2	2	−2	0	3
24	I like to meet the same nurse or doctor every time I come to the child welfare center	1	1	0	−2	−1	2
25	I want the doctor or nurse to ask me before they are going to do something to me	1	0	2	3	-2	1

[a]"Factor all" refers to the factor analysis of all participants' rankings.

[b]"Factor children" refers to the factor analysis of children's rankings only.

This study shows that children and adults have qualitatively different views, as only 1 of the child participants ranked the statements in a similar way to the parent. The perspectives identified in the factor analysis of all the participants' rankings reveal 3 different approaches: involving children in communication during the health care visit, sharing decision-making with children, and, lastly, allowing children to set their own limits for their participation. Adult participants placed particular emphasis on HCPs' communication with children and the establishment of a trusting relationship with them in the postsorting interviews.

Factor All 1: Direct Communication Between the Child and HCPs

Nine parents, 5 HCPs, and 1 child associated with this factor. They felt strongly about HCP communication with children, with the highest-ranked statements concerning how HCPs interact with children as patients and provide children with the possibility to participate in discussions. One parent saw health

XSL•FO
RenderX

care visits as a learning opportunity for children to learn to have a conversation.

> *You should still teach the child that first the doctor has to say something and then they can say it. Not just throw out a bunch of words, but you should still try to learn to have a conversation.* [Parent 4]

Communication was considered important for building a trustworthy relationship and ensuring a positive and comfortable health care encounter for the child. A positive and respectful way of communicating with the children was seen as the basis and an essential condition for the visit, more important than the children taking part in decision-making. Meeting the same staff was not necessary, but the communication skills of the HCP were highlighted for "inviting the child to participate in the visit." Nonverbal communication was not ranked high in importance but was acknowledged during the interviews as an important means of communication for individual children. One HCP described this interpretation of signals as an important part of her communication skills.

> *…it's rather important that I as a healthcare professional can listen to them and see this and interpret their signals, do what they want and don't want, with their body language instead.* [HCP 9]

HCPs felt that most children wanted to participate in the visit. Similarly, parents belonging to this factor described their children as willing to talk and ask questions. Therefore, staying quiet and listening or using nonverbal communication (statements 15 and 5) were not important.

> *She is not so interested in being quiet [laughs]* [Parent 14]

Statements concerning decision-making and choosing from choices were mainly understood to refer to major decisions on treatment options and considered less important. Both parents and HCPs felt that children did not want such responsibility and doubted children's understanding of the consequences of decisions. Some decisions, for example, diet changes for obese children, were not seen as suitable for children. While 1 parent felt that their child with more care needs and experience would need to be more involved in decision-making, other parents and HCPs stated that children were more interested in knowing what would happen during the visit than in making decisions with adults.

> *And this idea that we should be able to make decisions during the care visit, I don't think they care about that in the slightest, they just want to know what's going to happen.* [HCP 12]

Parents and HCPs felt that decisions should be age-appropriate and aligned with children's cognitive levels. Allowing children to choose the order of procedures during a health assessment was considered valuable and within the limits of children's decision-making abilities.

> *… here he got to choose which arm you want the injection to be taken. It's the child's body and I think it's important for the child to say that this arm will feel better to me. You might have pain in some place before.* [Parent 4]

Sometimes procedures had to be done regardless of choice, making mutual understanding between the child and HCP essential. Protesting was seen as a way for children to express emotions, and while protests were respected and could delay care, they did not always change the outcome.

Factor All 2: Understanding and Shared Decision-Making

Three children, 1 parent, and 5 HCPs were associated with this factor. They felt strongly about participating in decision-making and thought it was important for children to understand and follow the events of the health assessment at the child health care center. They also did not think that the parents' presence was necessary for children to talk to the nurse or doctor. HCPs felt they needed to quickly build a relationship with the child and show that they were friendly and not scary. Using translators in the situation hindered this relationship-building, but regular contact with the same HCP helped. Another effective approach was to prepare the children at home before the visit. Like in Factor all 1, a positive and comfortable health care encounter with direct communication with children was considered the foundation for a visit, but this factor emphasized shared decision-making with parents and HCPs. This required understanding of the discussion during the visit.

> *Sometimes certain words become difficult, and she can't quite keep up.… But this thing about making decisions together. It's linked to the fact that if you understand the meaning of it and in the same way, it's important to do it together.* [Parent 12]

Shared decision-making with parents and HCPs was found to be valuable, at the appropriate level of decision-making for children. Even though children might not fully understand what it involved until it was imminent, some decisions were beyond children's capability. Participants associated with this factor felt that children could decide more as they grew older. On the other hand, relevant decision-making for children was seen as children deciding over a range of options presented, giving them a sense of control over the situation.

> *They can choose the color of the paper they draw on, for example, for their drawing. And they can choose the order in which we do things, whether we weigh them first or check their eyesight first. But I don't think the difficult decisions feel important for the child to choose. Instead, they think it's nice that someone else decides. For example, whether to take a vaccine. They can't decide that.* [HCP 6]

Factor All 3: Responsive and Child-Led Participation

Two children, 2 parents, and 2 staff members were associated with this factor. Their preferences are interpreted as the children setting limits on their participation. These limits include consulting them before procedures, sharing decision-making, the possibility of protesting, and the use of nonverbal communication. Asking the children before procedures (statement 25) was motivated by the aim of involving them when possible, allowing them to make decisions based on their age and situation. Participants stated that children would participate for as long as they were interested, after which they

XSL•FO
RenderX

would move on to playing. One parent noted that varying levels of participation meant that offering choices during health assessments was less important.

> *I thought it was like this, yes, sometimes it is important. Sometimes it's a bit varied from day to day. Sometimes it's really important to get him to speak up and tell me about something and ask me something. Sometimes let me sit in a corner.* [Parent 9]

When children grew older, they could make more decisions, but at this age, some decisions on treatments were too complex for children to make. Children's participation was vital for some procedures; therefore, a gentle approach was used, with the possibility of postponing the procedure.

> *…if you take something as concrete as checking vision. You almost have to try somehow to make the child want to do it, because otherwise you might miss vision problems and there will be problems later in school or something, or headaches...* [HCP 7]

HCPs found it important to talk directly to children, even though they felt that children rarely asked questions of them. Similarly, a parent thought that the questions should be directed at children. If the child was unwilling to answer or colored the answer with fantasy, the parent would step in.

> *Yes, but if he wants to answer, he can. And then... we had a visit to the BUM [medical center] and task: does he eat vegetables? Yes, ask him. And then he unraveled the whole vegetable stash as well.* [Parent 9]

The importance of nonverbal communication was recognized. A parent explained that their child was very shy but willing to participate in their own way.

> *…it's very much that if she can respond with body language instead, or she can be quiet if she wants to, and a lot of things like that, as she wants it. She doesn't usually, and neither do her siblings, who want so much... care whether decisions are made with them or not.* [Parent 6]

Factor Children 1: Included in and Setting Their Own Terms for Participation

Three children were associated with this factor. They felt strongly about receiving information, being consulted prior to procedures, participating in discussions, and choosing how long to participate. This gave them a sense of control within the health care situation. As with Factor all 3, the children set their own terms for participation but placed the least importance on making decisions alone or with parents and HCPs. In the following situation, the child suddenly starts playing while still answering questions from the interviewer. She understands the question in a practical way. Furthermore, it can be interpreted that decisions regarding play are important to her. Her actions demonstrate how she sets the terms for her participation, showing agency.

> ***Interviewer:*** *Okay. [child walks away from the table where the interview is being done] Are you going on the slide? Was it hard to make the cards? You turn your head. Was it easy? Mmm. Is there anything that you would like to do when you are with the doctor or nurse that was not included? Is there anything that you would like to do when you are with the doctor or nurse? Apart from going down the slide. Is there anything else you would like to do?*
>
> ***Child:*** *Swing.*
>
> ***Interviewer:*** *Sing?*
>
> ***Child:*** *Swing.*
>
> ***Interviewer:*** *While you are waiting? Or after you have been to the doctor or nurse?*
>
> ***Child:*** *Mmm.*
>
> ***Interviewer:*** *So, is it important what you do while you are waiting?*
>
> ***Child:*** *Mmm.* [Child 3]

This example presents a comfortable level of decision-making for the child. The statement 7 ("I want to decide on easy things but not difficult") illustrated with a child picking a Band-Aid reflects a familiar, practical choice.

> ***Interviewer:*** *We'll take this one which is important, first that pile. I put them on the floor. If you look at them and say what is the most important of all these. Then it's just one card. See, [child's name]? Should I read them all? This one? [reads aloud the card the child is pointing to] I want to decide on easy things but not difficult ones. Is it most important when you come to the nurse?*
>
> ***Child:*** *[inaudible] band-aid.*
>
> ***Interviewer:*** *Yes, to be able to choose band-aids?*
>
> ***Child:*** *Yes.* [Child 9]

Factor Children 2: Small Choices, Meaningful Outcomes

Four children were associated with this factor. They felt strongly about having a choice of options (statement 21) and making decisions about minor issues, while leaving major decisions to parents and HCPs. These choices could be understood as play-related, perceiving the visit as a meeting with HCPs and playtime, and choosing what to play with while waiting as meaningful decisions. Although nurses also use games and play as a method of observing children's development during health assessments, this may explain the child's confusion.

> ***Interviewer:*** *Is it good like this? Is there anything else you would like to do when you come to meet the nurse?*
>
> ***Child:*** *Read a book.*
>
> ***Interviewer:*** *Read a book, okay.*
>
> ***Child:*** *Playing with dolls.*
>
> ***Parent:*** *You want to be in the waiting room.*
>
> ***Interviewer:*** *There are lots of things you want to play with. Mmm. Okay.*
>
> ***Child:*** *Doll. Slide. Read a book.*
>
> ***Interviewer:*** *So, before the visit or after the visit?*

XSL•FO
RenderX

Child: *After the visit.* [Child 17]

The children actively ranked their preferences during the interview but struggled to motivate their choices. One child asked for help from their mother, who then interpreted the child's response.

> ***Interviewer:*** *Why did you choose to sort the statements like this? Was it a...?*
>
> ***Child:*** *Mum.*
>
> ***Interviewer:*** *Wasn't it you who decided?*
>
> ***Child:*** *No.*
>
> ***Interviewer:*** *No?*
>
> ***Parent:*** *It was you who decided all by yourself.*
>
> ***Child:*** *I don't know.*
>
> ***Parent:*** *You don't know why you chose that?*
>
> ***Child:*** *No.* [Child 1]

One parent noted that the child chose opposite preferences, likely due to interpreting the statements differently.

> *…this one sounds like he's going to choose different things when he leaves, but for me it's different choices to choose whether he wants to... start by weighing himself or measuring himself first…* [Parent 1]

In a few cases, participating in the process was more important than the result. One child took over the following postsorting interview and proudly explained her ranking but left out the details of her priorities.

> ***Child:*** *Now you got to make piles.*
>
> ***Parent:*** *I got to make piles too. It is the same as you. What I think is important [talking to the child]*
>
> ***Child:*** *What is important?*
>
> ***Parent:*** *That doctors and nurses listen to... to the child? I don't know what you chose.*
>
> ***Child:*** *I chose all the things that were not important and all the things that were important and in between.* [Parent 4 interview]

Some children associated with this factor were described by parents as talkative and curious, which matched their ranking. They ranked the statements about being quiet, being asked before procedures or decisions affecting them were made, and the possibility of protesting the lowest. These children seem accustomed to expressing their opinions and are unaware of situations in which they cannot do so verbally.

Factor Children 3: Trust Through Familiarity and Shared Decision-Making

Three children were associated with this factor. They prioritized decision-making with parents and HCPs as well as a positive and comfortable health care encounter by HCPs, similar to Factor all 2. However, unlike Factor all 2, participants in this factor considered it important to meet the same HCPs (statement 24) and did not consider it important for HCPs to talk to children as well (statement 1). This suggests the need for parents to be present and supportive during meetings with HCPs, and for children to be willing to build relationships with HCPs. In the following example, the child's wish to repeat the vaccination suggests an optimal level of shared decision-making and trust with the HCP.

> ***Interviewer:*** *Is there anything else you would like to do when you come to the nurse that is not on these cards? Is there anything you absolutely want to do when you come to the nurse?*
>
> ***Child:*** *Hmmm.*
>
> ***Interviewer:*** *That might be a difficult question because we have so many cards here.*
>
> ***Parent:*** *Exactly. Is there anything [child's name]? Is there anything you would like to do?*
>
> ***Child:*** *No.*
>
> ***Interviewer:*** *So, then you're done with it.*
>
> ***Child:*** *Syringe.*
>
> ***Interviewer:*** *What did you say?*
>
> ***Child:*** *Syringe.*
>
> ***Parent:*** *Would you like a shot [vaccine, author's remark]? You've had a shot.*
>
> ***Child:*** *Yes.*
>
> ***Sibling:*** *He really likes syringes now.*
>
> ***Interviewer:*** *Oh, so it went so well that you thought I could do that next time too. That's great. Okay. I think I'm done too. Then, I thank you for the interview.* [Child 16]

Nonverbal communication was ranked as the least important. One child who relied on nonverbal communication during the interviews ranked it low and prioritized decision-making instead. This suggests that the feeling of being involved mattered more than the way of communication, as confirmed by her parent:

> *She just wants to get... So, in her own way, be quiet, respond with thumbs up, thumbs down. Then she's usually happy.* [Parent 6]

Additional Suggested Ways of Participating

In the postsorting interview, participants were asked for other ways children could participate in health care situations. Children mentioned different toys for the waiting room, while adults recommended postponing procedures to a later time if the child was not ready. They highlighted the importance of preparatory information, which helped both children and parents make visits run smoothly. Using pictures to aid communication and allowing children to play with a toy before talking to HCPs were also found to be effective strategies. Reflecting on the visit afterward to assess the child's experience was suggested. The value of the HCP's skills and parental involvement in the process was emphasized, as praise or confirmation from the parent while the child is answering questions promoted their participation, whereas a lack of parental engagement in the situation demotivated children.

> *Or being a parent and looking at the clock all the time and just being stressed and going to work. That doesn't work either. Most often, many times it doesn't work. You have to come back again, so that the child realizes that we don't have time to be here.* [HCP 8]

XSL•FO
RenderX

HCPs also felt that they needed to be flexible and provide options for children. Despite the many obligatory examinations and decisions that children could not influence, HCPs recognized that there were always ways to involve them in the visit

> *I think that in everything we do with children there is always a choice. They can always be allowed to choose which band-aid, which arm.* [HCP 12]

Discussion

Principal Findings

The aim of this study was to understand children's preferences for participation in health care situations and to examine them from the child's perspective as well as from the child perspective. This study sheds new light on the preferences of preschool-aged children. To our knowledge, this is the first study to compare the perspectives of preschool-aged children, their parents, and health professionals on children's participation. The key findings can be summarized as follows.

First, preschool-aged children are happy to leave major decisions to adults, but they want to be in control of health care situations. Similarly, Darcy et al [31] described how children undergoing cancer treatment wanted to control how, when, and at what pace procedures were conducted, and how much they participated in them. According to Moore and Kirk [32], this type of decision, which does not alter the outcome, is considered tokenistic because the planned procedures will be carried out regardless. The current study found that this kind of control is important for preschool-aged children, giving them an opportunity to participate and set their own limits to participation. This relates to the findings by Carlsson et al [33] on how HCPs considered participation as a prerequisite for care. They tried to "make the most of the moment" by involving children at every step, for example, by letting the child choose if they wanted to sit or lie down during the examination. Decosta et al [34] found that actively involving children in the preparation of medical procedures made diabetes education more meaningful, as it engaged them in their own care in an age-appropriate way. Furthermore, positive feedback, play-based approaches, and playful communication encouraged the young (<7) children's participation. These approaches were designed to help children gradually understand and engage more with their care. They are valuable for encouraging children to participate, as the present research supports the idea that some medical examinations depend on children's participation. For example, a vision check requires the child's participation. Children's protests or nonparticipation could postpone the examination to later and result in unnecessary suffering of undetected and treatable vision problems.

Second, a supportive and present parent appears to be important to children. Harlow et al [35] define parental presence in pediatric hospital care as being physically or emotionally present with their child, reducing the child's anxiety and providing emotional support. Besides the emotional support, children seek their parents for advocacy and sense of security during hospitalization [36,37]. Another aspect is learning how to participate with the help of supportive and encouraging adults. In turn, adults need to strike a balance between protecting and encouraging participation [19], or between intervening and giving autonomy [38], in order to provide children with an opportunity to express their views [16]. This means focusing on the child at the center of the care, as outlined in the CCC principles, and giving the child a voice [1,2,16].

Third, participants in the study suggested other ways of involving children, such as providing preparatory information with visual aids or postponing the procedure to a later date. Previous studies have demonstrated the effectiveness of visual aids and digital tools in preparing children for procedures and facilitating their participation in daily discussions about their care in hospital wards [27,39,40]. Digital tools could standardize the information provided to children. A digital application could facilitate children's participation by providing an understanding of what will happen during the procedure [41], while enabling HCPs to identify areas where the child requires further information before the procedure begins. Another suggestion from the participants was to ask children about their experience of the health care situation. This idea is worth exploring further in future research. Asking children at the end of a visit about their experience would provide valuable insight and help HCPs prepare for subsequent encounters with children.

Finally, children and their parents perceive and interpret the world differently. While preschool-aged children are still developing their language skills and abstract thinking abilities [42], they are capable of processing the information presented to them [43] and actively shaping their social environment as competent participants [13,14]. The aim of this Q methodology study was to identify and describe the common perspectives of the participants regarding child participation in health care situations. The results comparing the 3 different views (children, parents, and HCPs) and motivations for the rankings suggest that the resulting perspectives are coherent. Children may not be able to justify their choices, but this should not diminish the importance of their opinions, nor should it be a reason to override them if they differ from the views of adults [16]. Children's preferences identified in this study are in line with the core principles of CCC: trust, respect, autonomy, self-determination, and decision-making regarding their care together with parents and HCPs [2]. The current study showed that parent proxy does not always provide a correct picture of the phenomenon. Therefore, the perspectives of children should be sought in research and health care practice, as well as in other disciplines, and practitioners should be made aware of this perspective, as set out in the UNCRC [9]. This document should be understood as legally binding within state parties [9], ensuring that children are able to express their views. As the children were unable to justify their rankings, it would be interesting to explore this further.

Strengths and Limitations

This study has some limitations. It revealed various perspectives on children's preferences regarding participation in health care situations. However, it does not indicate how prevalent these preferences are among preschool children. The Q statements were developed from themes identified in a literature review, simplified into age-appropriate Swedish, and supported by animations. To minimize potential misunderstandings, only

XSL•FO
RenderX

participants who were fluent in Swedish were included in the interviews and analysis. Prior to the study, the meaning of the images provided with the statements for children was discussed with a convenience sample of children. Nevertheless, misunderstandings regarding the language may still have occurred. Conducting a pilot study could have been helpful. Children and their parents were recruited by nurses at Child Welfare Centers. This may have introduced a bias toward more outgoing children, which could have influenced the study's outcomes. Appointments scheduled late in the afternoon, after a full day at preschool, may have reduced the children's focus, particularly among the 3-year-olds, making it challenging for them to complete the statement ranking task. Conducting the ranking task alongside their child may have influenced some parents' responses if they overheard the interviewer's conversation with their child. Nevertheless, most parents were highly engaged in the ranking process, as they found the number of statements substantial and did not have time to listen to their child's interview.

To our knowledge, no Q methodology studies have been conducted on children younger than 5 years. The number of statements varies greatly from less than 20 to more than 200 in Q-methodology studies [23,24]. In this study, coverage was achieved without redundancy, making additional statements unlikely to expand the viewpoint space meaningfully. Furthermore, Q sorting can be difficult, demanding, and time-consuming [20]. A more concise set reduces participant fatigue. The 25 statements in this study with child participants align with established practice while maintaining methodological rigor.

Conclusion

In this study, 4- and 5-year-olds participated enthusiastically in the ranking process. This method also enabled us to compare the children's perspectives with the efforts of parents and health professionals to understand them and, in some cases, their ability to explain the children's rankings. None of the 3-year-old participants could complete the ranking procedure. It would be interesting to explore whether 3-year-olds share the perspectives identified in this study or whether they have different views, but different methods are needed for this.

Acknowledgments

Microsoft Bing Copilot (version 2024 May) was used to draw pictures for the children's statement cards for illustration of the statements.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this study.

Data Availability

Data are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: MH, MK, CNB, JR, SM
Formal analysis: MH, MK
Investigation: MH
Methodology: MH, MK, CNB, JR, SM
Project administration: MH
Supervision: MK, CNB, JR, SM
Visualization: MH, MK
Writing – original draft: MH
Writing review & editing: MH, MK, CNB, JR, SM

Conflicts of Interest

None declared.

References

1. Coyne I, Holmström I, Söderbäck M. Centeredness in healthcare: a concept synthesis of family-centered care, person-centered care and child-centered care. J Pediatr Nurs 2018;42:45-56. [doi: 10.1016/j.pedn.2018.07.001] [Medline: 30219299]
2. Coyne I, Hallström I, Söderbäck M. Reframing the focus from a family-centred to a child-centred care approach for children's healthcare. J Child Health Care 2016 Dec;20(4):494-502. [doi: 10.1177/1367493516642744] [Medline: 27141084]
3. Nilsson S, Björkman B, Almqvist AL, et al. Children's voices--differentiating a child perspective from a child's perspective. Dev Neurorehabil 2015 Jun;18(3):162-168. [doi: 10.3109/17518423.2013.801529] [Medline: 23924164]
4. Sommer D, Pramling Samuelsson I, Hundeide K. Child Perspectives and Children's Perspectives in Theory and Practice: Springer; 2010. [doi: 10.1007/978-90-481-3316-1]

XSL•FO
RenderX

5. Halldén G. Barns inflytande över barndomen [Article in Swedish]. Sociol Forsk 2005;42(4):5-11. [doi: 10.37062/sf.42.19347]
6. Söderbäck M, Coyne I, Harder M. The importance of including both a child perspective and the child's perspective within health care settings to provide truly child-centred care. J Child Health Care 2011 Jun;15(2):99-106. [doi: 10.1177/1367493510397624] [Medline: 21685225]
7. Imms C, Adair B, Keen D, Ullenhag A, Rosenbaum P, Granlund M. "Participation": a systematic review of language, definitions, and constructs used in intervention research with children with disabilities. Dev Med Child Neurol 2016 Jan;58(1):29-38. [doi: 10.1111/dmcn.12932] [Medline: 26411643]
8. Lansdown G. Conceptual framework for measuring outcomes of adolescent participation. : UNICEF; 2018 URL: https://www.unicef.org/media/59006/file [accessed 2026-03-26]
9. Convention on the rights of the child. United Nations. 1989. URL: https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child [accessed 2026-03-26]
10. Hietanen M, Nahlén Bose C, Stålberg A, Raghothama J, Meijer S. A systematic review on children's participation in healthcare from a child's perspective. Scand J Caring Sci 2025 Dec;39(4):e70157. [doi: 10.1111/scs.70157] [Medline: 41299201]
11. Davies A, Randall D. Perceptions of children's participation in their healthcare: a critical review. Issues Compr Pediatr Nurs 2015;38(3):202-221. [doi: 10.3109/01460862.2015.1063740] [Medline: 26331450]
12. Rindstedt C, Aronsson K. Children's intent participation in a pediatric community of practice. Mind, Cult Act 2012 Oct;19(4):325-341. [doi: 10.1080/10749039.2012.663449]
13. Montreuil M, Carnevale FA. A concept analysis of children's agency within the health literature. J Child Health Care 2016 Dec;20(4):503-511. [doi: 10.1177/1367493515620914] [Medline: 26666263]
14. James A, Prout A. A new paradigm for the sociology of childhood? Provenance, promise and problems. In: Constructing and Reconstructing Childhood, 2nd edition: Routledge; 1997:7-33 URL: https://www.taylorfrancis.com/chapters/mono/10.4324/9780203362600-8/new-paradigm-sociology-childhood-provenance-promise-problems-allison-james-alan-prout [accessed 2026-03-26]
15. Coyne I. Children's participation in consultations and decision-making at health service level: a review of the literature. Int J Nurs Stud 2008 Nov;45(11):1682-1689. [doi: 10.1016/j.ijnurstu.2008.05.002] [Medline: 18706560]
16. Davies C, Fraser J, Waters D. Establishing a framework for listening to children in healthcare. J Child Health Care 2023 Jun;27(2):279-288. [doi: 10.1177/1367493519872078] [Medline: 31470733]
17. Coyne I, Gallagher P. Participation in communication and decision-making: children and young people's experiences in a hospital setting. J Clin Nurs 2011 Aug;20(15-16):2334-2343. [doi: 10.1111/j.1365-2702.2010.03582.x] [Medline: 21410571]
18. Quaye AA, Coyne I, Söderbäck M, Hallström IK. Children's active participation in decision-making processes during hospitalisation: an observational study. J Clin Nurs 2019 Dec;28(23-24):4525-4537. [doi: 10.1111/jocn.15042] [Medline: 31430412]
19. Coyne I, Harder M. Children's participation in decision-making: balancing protection with shared decision-making using a situational perspective. J Child Health Care 2011 Dec;15(4):312-319. [doi: 10.1177/1367493511406570] [Medline: 21828162]
20. Barker JH. Q-methodology: an alternative approach to research in nurse education. Nurse Educ Today 2008 Nov;28(8):917-925. [doi: 10.1016/j.nedt.2008.05.010] [Medline: 18572285]
21. Stephenson W. The inverted factor technique. Br J Psychol 1936 Apr;26(4):344-361. [doi: 10.1111/j.2044-8295.1936.tb00803.x]
22. Churruca K, Ludlow K, Wu W, et al. A scoping review of Q-methodology in healthcare research. BMC Med Res Methodol 2021 Jun 21;21(1):125. [doi: 10.1186/s12874-021-01309-7] [Medline: 34154566]
23. Duncan Millar J, Mason H, Kidd L. What is Q methodology? Evid Based Nurs 2022 Jul;25(3):77-78. [doi: 10.1136/ebnurs-2022-103568] [Medline: 35609957]
24. Hensel D, Toronto C, Lawless J, Burgess J. A scoping review of Q methodology nursing education studies. Nurse Educ Today 2022 Feb;109:105220. [doi: 10.1016/j.nedt.2021.105220] [Medline: 34902708]
25. Damio SM. Q methodology: an overview and steps to implementation. Asian J Univ Educ 2016;12:105-117 [FREE Full text]
26. Vägledning för barnhälsovården. : Socialstyrelsen; 2014 URL: https://www.socialstyrelsen.se/contentassets/093cadca5119416bbb1bd6168a75f600/2014-4-5.pdf [accessed 2026-03-26]
27. Lind BM, Wigert H, Jenholt Nolbris M, Patriksson K. Child healthcare nurses' experience of communication with 4-year-old children during their visit to the child health care center. Public Health Nurs 2024;41(5):1098-1105. [doi: 10.1111/phn.13351] [Medline: 38881097]
28. Reuter A, Johansen K. Utvecklingsuppföljning 5 år [Article in Swedish]. Rikshandboken i barnhälsovård. 2024. URL: https://www.rikshandboken-bhv.se/halsobesok/utvecklingsuppfoljning/utvecklingsuppfoljning-i-olika-aldrar/utvecklingsuppfoljning-5-ar/ [accessed 2025-10-06]
29. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]

XSL•FO
RenderX

30. Schmolck P, Atkinson J. PQMethod manual. schmolck. 2014 Mar. URL: http://schmolck.org/qmethod/pqmanual.htm [accessed 2025-04-23]
31. Darcy L, Enskär K, Björk M. Young children's experiences of living an everyday life with cancer - a three year interview study. Eur J Oncol Nurs 2019 Apr;39:1-9. [doi: 10.1016/j.ejon.2018.12.007] [Medline: 30850132]
32. Moore L, Kirk S. A literature review of children's and young people's participation in decisions relating to health care. J Clin Nurs 2010 Aug;19(15-16):2215-2225. [doi: 10.1111/j.1365-2702.2009.03161.x] [Medline: 20659201]
33. Carlsson IM, Nygren JM, Svedberg P. Patient participation, a prerequisite for care: a grounded theory study of healthcare professionals' perceptions of what participation means in a paediatric care context. Nurs Open 2017 Jan;5(1):45-52. [doi: 10.1002/nop2.106] [Medline: 29344394]
34. DeCosta P, Grabowski D, Jespersen LN, Skinner TC. Playful communication and care: exploring child-centred care of young children with type 1 diabetes through the framework of zone of proximal development. Front Clin Diabetes Healthc 2021;2:707553. [doi: 10.3389/fcdhc.2021.707553] [Medline: 36994334]
35. Harlow AB, Ledbetter L, Brandon DH. Parental presence, participation, and engagement in paediatric hospital care: a conceptual delineation. J Adv Nurs 2024 Jul;80(7):2758-2771. [doi: 10.1111/jan.15996] [Medline: 38037504]
36. DeCosta P, Skinner TC, Sørensen JL, Topperzer MK, Grabowski D. Young children's perspectives on treatment and care: a qualitative study using narrative and play-based interviewing. J Pediatr Nurs 2023;73:211-220. [doi: 10.1016/j.pedn.2023.10.014] [Medline: 37862851]
37. Ryan MJ, Lee LA, Carnevale FA, et al. Parental and family presence are essential: a qualitative study of children's lived experiences with family presence in pediatric intensive care. J Pediatr Nurs 2025;80:e228-e235. [doi: 10.1016/j.pedn.2024.12.017] [Medline: 39753454]
38. Hart RA. Children's participation: from tokenism to citizenship. : UNICEF International Child Development Centre; 1992 URL: https://participationpool.eu/wp-content/uploads/2020/05/Hart-R.-1992-Children-Participation-from-Tokenism-to-Citizenship.pdf [accessed 2026-03-26]
39. Carlsson IM, Arvidsson S, Svedberg P, et al. Creating a communication space in the healthcare context: children's perspective of using the eHealth service, Sisom. J Child Health Care 2021 Mar;25(1):31-43. [doi: 10.1177/1367493520904804] [Medline: 32048874]
40. Stålberg A, Sandberg A, Larsson T, Coyne I, Söderbäck M. Curious, thoughtful and affirmative-young children's meanings of participation in healthcare situations when using an interactive communication tool. J Clin Nurs 2018 Jan;27(1-2):235-246. [doi: 10.1111/jocn.13878] [Medline: 28514530]
41. Stålberg A. Design and redesign of the IACTA app, an interactive communication tool intended to facilitate young children's participation in healthcare situations. J Pediatr Nurs 2021;61:260-268. [doi: 10.1016/j.pedn.2021.07.015] [Medline: 34333243]
42. Pakpahan FH, Saragih M. Theory of cognitive development by Jean Piaget. J Appl Linguist 2022;2(2):55-60. [doi: 10.52622/joal.v2i2.79]
43. Woodhead M. Early childhood development: a question of rights. Int J Early Childh 2005 Dec;37(3):79-98. [doi: 10.1007/BF03168347]

Abbreviations

CCC: child-centered care
CHC: child health care
HCP: health care professional
UNCRC: United Nations Convention on the Rights of the Child

Edited by K Sellen; submitted 17.Dec.2025; peer-reviewed by J Staines, L Munongi; revised version received 17.Mar.2026; accepted 17.Mar.2026; published 21.Apr.2026.

Please cite as:
Hietanen M, Kornevs M, Nahlén Bose C, Raghothama J, Meijer S
"To Choose a Band-Aid" - Children's Preferences for Participation in Health Care Situations: Q Methodology Study
JMIR Hum Factors 2026;13:e89802
URL: https://humanfactors.jmir.org/2026/1/e89802
doi:10.2196/89802

© Merja Hietanen, Maksims Kornevs, Catarina Nahlén Bose, Jayanth Raghothama, Sebastiaan Meijer. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 21.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use,

XSL•FO
RenderX

distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Anticipated Barriers and Facilitators to the Implementation of an AI Tool Advising on the Placement of Mitigation Strategies to Improve Air Quality in the Hospital Setting: Qualitative Study

Yolanda Barrado-Martín[1], PhD; Sanjoli Mathur[2], MSc; Zarnie Khadjesari[2], PhD; Sigrún Eyrúnardóttir Clark[3], MSc; Cecilia Vindrola-Padros[3], PhD; Laurence B Lovat[4], MBBS, PhD; Fiona Stevenson[1], PhD

[1]Research Department of Primary Care & Population Health, University College London, Upper 3rd Floor, Royal Free Campus, Rowland Hill St., London, England, United Kingdom

[2]School of Health Sciences, University of East Anglia, Norwich, England, United Kingdom

[3]Rapid Research Evaluation and Appraisal Lab (RREAL), Department of Targeted Intervention, University College London, London, England, United Kingdom

[4]Intelligent Systems in Gastroenterology & Healthcare Technology Lab (InSIGHT AI), Division of Surgery and Interventional Science, University College London, London, England, United Kingdom

Corresponding Author:
Yolanda Barrado-Martín, PhD
Research Department of Primary Care & Population Health, University College London, Upper 3rd Floor, Royal Free Campus, Rowland Hill St., London, England, United Kingdom

Abstract

Background: AI is increasingly prevalent within health care; however, its implementation within the hospital setting may come with its own limitations, given the high-pressure and sensitive environment. Barriers and facilitators to the implementation of AI within hospitals have been largely explored quantitatively, and there is a scarcity of qualitative research on the implementation of AI tools to improve air quality.

Objective: This study aims to qualitatively explore, through a normalization process theory (NPT) lens, the anticipated barriers and facilitators to the successful implementation of an AI tool to guide the placement of mitigation strategies to reduce airborne transmission of infections within the hospital setting. Specifically, we sought to (1) understand the day-to-day work of key hospital staff involved in reducing airborne transmission of infections, (2) identify potential barriers to the implementation of an AI tool to guide the placement of mitigation strategies in the hospital setting, and (3) recognize necessary conditions for the successful implementation of the AI tool within the hospital setting.

Methods: We conducted ethnographic observations (n=4) and semistructured interviews (n=18) with hospital staff following preliminary one-on-one meetings and group discussions involving patients, members of the public, national leaders in ventilation working within academic and industry environments, and hospital staff (n=70). Data were collected between October 2025 and May 2026 in the United Kingdom. Interview transcripts, field notes, and summaries from preliminary wider engagement work activities were analyzed in NVivo (Lumivero; version 12) following codebook thematic analysis to facilitate confirmatory triangulation. Findings were organized according to the core constructs of NPT.

Results: Hospital staff understanding of the AI tool and their preparedness to invest time in its implementation varied across those consulted. The anticipated work required by hospital staff to successfully implement the AI tool and its recommended mitigation strategies was identified as a significant challenge for its scalability. Differences in the ways of working among trusts and the competition for limited resources (collective action), and issues associated with hospitals' physical estate (coherence) were seen as particularly problematic. The need for local champions (cognitive participation) was raised as a facilitator, as well as expert validation and the assessment of the AI tool's effectiveness (reflexive monitoring).

Conclusions: This preimplementation study suggests that the successful implementation of an AI tool to improve air quality in the hospital setting depends not only on the effectiveness demonstrated by the AI tool but also on the way trial outcomes are conveyed to hospital staff. Hospital staff anticipate an initial resistance to the AI tool's implementation. Further, the impact of the AI tool and suggested mitigation strategies in workflows, costs, and staff and patient outcomes may be decisive.

(JMIR Hum Factors 2026;13:e93420) doi:10.2196/93420

KEYWORDS

barriers and facilitators; artificial intelligence; implementation science; hospital; cross-infections; air filters; infection control; qualitative research; normalization process theory

XSL•FO
RenderX

Introduction

AI is increasing its prevalence within health care and its adoption is supported by UK health policy, such as the National Health Service (NHS) 10-year plan and the creation of AI working groups to steer the use of AI and speed up its use within the NHS [1,2]. While AI's predominant use has been in medical imaging [3], it has the potential to contribute to other areas, facilitating clinical decision-making, early diagnosis [4], and hospital documentation [5]. This aligns with the NHS ambition to shift "from analogue to digital," prioritizing digital tools, including AI-driven solutions, to promote continuous monitoring and the shift "from sickness to prevention." These shifts are intended to improve population health and the sustainability of health care [2].

AI-enabled tools have already shown potential in predicting hospital-acquired infections with high accuracy [6]. Early detection and intervention of airborne disease transmission can protect frontline workers and facilitate hospitals working at full capacity by reducing staff sickness. Our wider research team is developing and implementing an AI tool—the Artificial Intelligence Air Safety Tool (AISaT; University College London). AISaT will process data about the characteristics of a hospital room, such as its dimensions and the position of people in the room, to then produce recommendations on where best to position mitigation strategies within that room to reduce airborne infection transmission. Examples of mitigation strategies could include devices such as high-efficiency particulate air (HEPA) filters [7], or equipment such as screens, curtains [8], windows, or doors.

Despite the promise of AI in infection prevention and the potential of qualitative research to help facilitate implementation [9], few studies have focused on qualitatively exploring barriers and facilitators to implementation of AI within hospital settings [10,11]. This is known to be challenging [12], and identified barriers include the lack of trust, issues with generalizability, privacy and maintenance plans, or the previous lack of AI implementation guidelines [13], which have only been recently addressed [14]. When qualitative methods have been used, studies have mostly focused on AI looking to improve patient care directly (eg, by speeding diagnosis) [11,15].

The aim of this study was to explore, through a normalization process theory (NPT) lens [16,17], the anticipated barriers and facilitators to the successful implementation of AISaT, a tool that seeks to protect both staff and patients. This theory was chosen to map the anticipated requirements for the integration of AISaT into daily practice and facilitate trial delivery. By focusing on stakeholder experiences, including those of clinical, technical, and leadership teams, this work seeks to inform strategies that align with the NHS's preventative care vision and support the normalization of such innovations in routine hospital practice.

The study sought to answer the question: how do relevant stakeholders perceive the implementation of an AI tool to guide the placement of mitigation strategies to reduce airborne transmission of infections within the hospital setting? Our objectives were to (1) understand the day-to-day work of key hospital staff involved in reducing airborne transmission of infections, (2) identify potential barriers to the implementation of an AI tool to guide the placement of mitigation strategies in the hospital setting, and (3) recognize necessary conditions for the successful implementation of the AI tool.

Methods

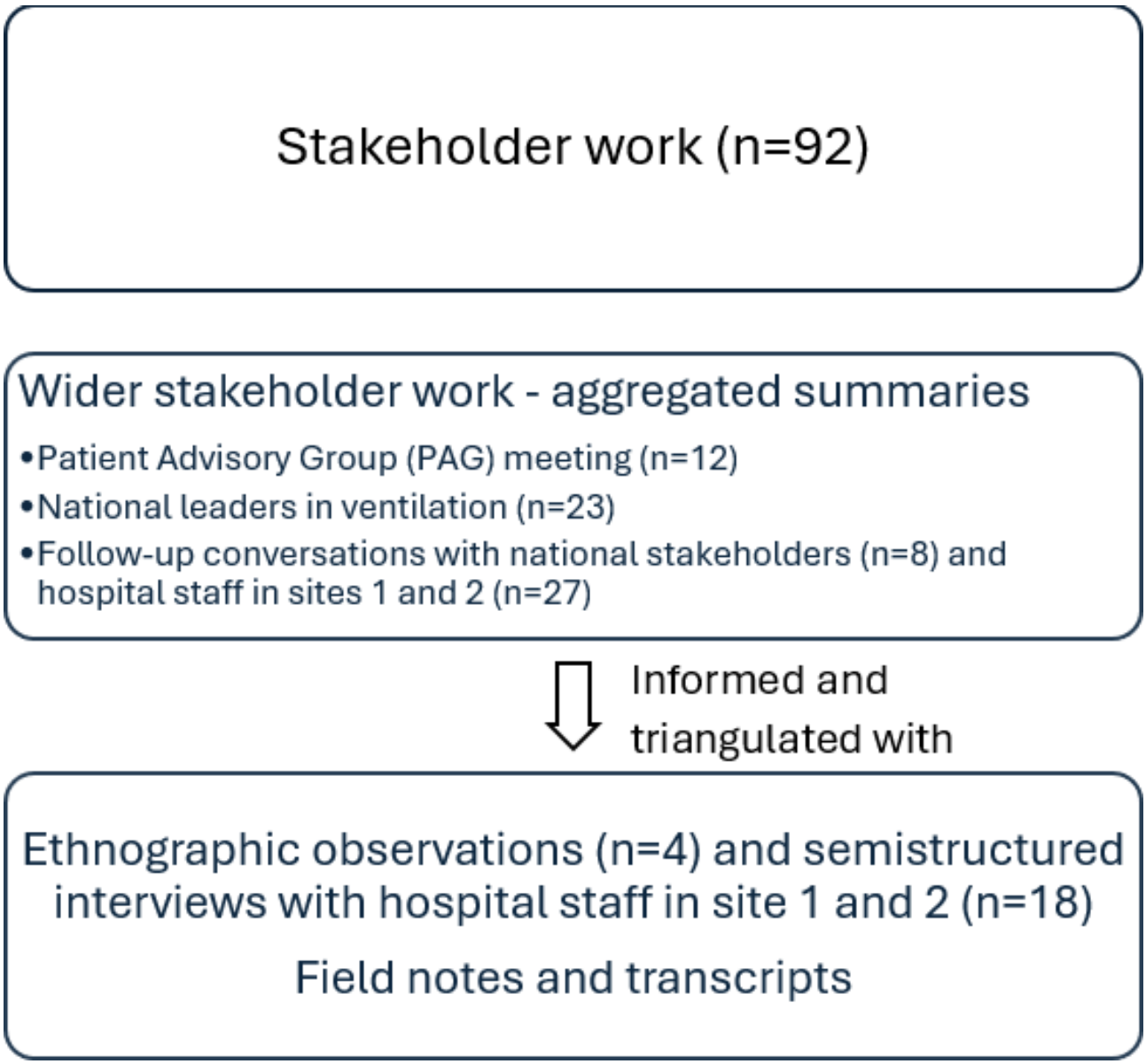
Design

Wider stakeholder engagement activities, drawing on findings from 2 systematic reviews in this area [18,19], were conducted. These comprised (1) an in-person patient advisory group (PAG) meeting, (2) an online workshop with national leaders, (3) follow-up conversations with national leaders, and (4) hospital staff based at 2 hospitals who would be involved in the subsequent clinical trial (trial registration number: ISRCTN12421486). Summaries containing data gathered from the informal conversations (wider stakeholder work) informed recruitment and topic guide development for formal data collection, while contributing to the joint analysis.

Drawing on insights from this wider stakeholder engagement, we conducted formal data collection (ethnographic observations and semistructured interviews) between October 15, 2025, and May 7, 2026, at a large teaching hospital in Greater London and a smaller teaching hospital in an urban town. The focus of this research was on the roles and responsibilities of hospital staff working to keep hospitals free from airborne disease transmission. Data collected at this stage were triangulated with aggregated summaries from the wider stakeholder engagement activities (informal conversations with hospital staff, national leaders in ventilation, and members of the public) as reflected in Figure 1.

XSL•FO
RenderX

Figure 1. Summary of types of data collected.



Wider stakeholder engagement, and particularly informal interviews, compensated for relatively less formal semistructured interviews. Ethnographic observations complemented what was being said during semistructured interviews by allowing us to understand staff's day-to-day work in ventilation. Formal semistructured interviews included questions raised in informal conversations and expanded on current practices, acceptability, and feasibility aspects. Data obtained formally and informally aligned well.

All data were collected before trialing of the AISaT system, so both informal discussions and formal data collection were based on a hypothetical future implementation of AISaT rather than on its actual use or implementation. The COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist was used to support the presentation of methods and results [20] (Checklist 1).

For context, AISaT will require the user to introduce details on room geometry, airflow (including the nature of air conditioning if present), occupancy, and mitigations used (air filter, screens, fans, and so on). AISaT then uses computational fluid dynamics (CFD)–based simulations to generate 2D and 3D representations of the entire room, with a color chart showing ranked recommendations (from highest to lowest) based on the likelihood of chosen mitigations to reduce the number of respiratory particles. AISaT may only need to be used once in rooms where occupancy remains similar over time (eg, consulting rooms), while in areas where occupancy changes, it may need to be rerun accordingly. The ultimate decision to place the mitigation strategies where AISaT recommends lies with the human operating the device, who will be provided with alternatives and will need to avoid hazardous placements.

Wider Stakeholder Engagement

The PAG was involved throughout the study. The PAG was consulted on the relevance and priority of clean air within hospitals, and their initial reactions to HEPA filters and to AISaT. Members of the group were recruited through advertisement on the National Institute for Health and Care Research (NIHR) People in Research website, charity

XSL•FO
RenderX

connections, and established patient networks. The interaction with patients was led by SM, guided by Simon Watt and FS.

National leaders were invited to the online workshop and follow-up conversations were drawn from academic, industry, and health care sectors, through connections with members of the trial team and through online searches, because of their contributions to the field of ventilation. The workshop shared the project aims, sought first reactions about the trial, and helped to identify further national leaders in the field and key hospital job roles to target in ethnographic observations and interviews. It lasted 60 minutes to maximize participation by people working in the health service in high-pressure roles. The online workshop was led by YB-M and SM, with support from FS, ZK, and SEC.

The online workshop was supplemented by follow-up conversations, which allowed stakeholders to share more detail in relation to their views. Follow-up conversations were also conducted with people who could not join the workshop, and those who were identified through snowball sampling (n=8 [YB-M: n= 7; SM: n= 1]). After collecting information during the workshop about relevant hospital roles to engage with, the research team scheduled follow-up conversations with professionals in these roles (n=27 [YB-M: n= 25; SM: n= 2]). These informal conversations clarified the contribution of different job roles in keeping the hospital free from airborne disease transmission. A draft summary of barriers and enablers identified through the PAG, national leaders' workshop, and preliminary contacts in hospitals is provided in Multimedia Appendix 1, which is the focus of a separate publication reporting on the development of an implementation strategy for rollout. Further description of the data collection process followed for wider stakeholder engagement is provided in Multimedia Appendix 2.

Ethnographic Observations and Semistructured Interviews

Hospital staff were recruited following a combination of purposive and snowball sampling techniques. Participants in diverse roles (eg, infection prevention and control [IPC], estates, and clinical) were invited to capture the views of all those likely to be involved in the implementation of Air Safety. Potential participants were approached up to 3 times by email. We completed ethnographic observations in the work environment and online semistructured interviews following a predefined internal study protocol. Researchers had no prior relationship with participants. Participants had read a participant information sheet and knew the aims of the study. The researcher introduced themselves and their role in the study and explained the aims of the semistructured interview or ethnographic observation.

Ethnographic observations were used to understand hospital staff's day-to-day involvement in keeping the hospital air free from infection while complementing and facilitating understanding of the content of interviews. Ethnographic observations were unstructured, but a minimal data collection template was used (Multimedia Appendix 3). These were conducted by female researchers trained in ethnographic research methods and with previous experience in ethnography (FS and YB-M). Hospital staff at both site 1 and site 2 (the two trial sites) were invited to participate in ethnographic observations; however, volunteers were only found for site 1. A total of 8 hours of ethnographic observation were conducted at site 1 with the 4 research participants who volunteered to be shadowed. Field notes were taken in situ and transcribed electronically as soon as possible after the ethnographic observation. Interactions with nonparticipants were not included in field notes.

Each semistructured interview was guided by an interview schedule (Multimedia Appendix 4) which had been pilot-tested in preliminary conversations with stakeholders. Two researchers with expertise in qualitative methods (YB-M and SEC) and a background in psychology and molecular medicine conducted the interviews (YB-M: n=9; SEC: n=9). The interviews were conducted iteratively, building on previous insights.

Semistructured interviews lasted on average 33.9 (SD 6.61) minutes (ranging from 25 to 49 minutes; median 33.09, IQR 28.7-38.1 minutes). All interviews were audio- or video-recorded on Microsoft Teams with automatic transcription. The researchers (YB-M and SEC) cross-checked transcripts for accuracy and pseudonymized the data. Given hospital staff's limited availability, interviews were not returned to participants to check their content, but during the interview, summarization and clarification were used instead.

Semistructured interviews concluded when thematic sufficiency had been reached, considering input both from semistructured interviews and previous informal conversations with hospital staff from the same hospitals. No interviews were repeated. Participants did not provide feedback on the findings.

Analysis

Ethnographic observations and semistructured interview transcripts were analyzed using codebook thematic analysis [21], which sits between coding reliability and reflexive approaches of thematic analysis [22], in NVivo (Lumivero; version 12) [23]. YB-M coded 2 interviews and developed a preliminary coding framework inductively that was presented and discussed with FS. The rest of the ethnographic observations and semistructured interviews, and the wider stakeholder engagement work, were coded by YB-M applying this framework. Coding was conducted iteratively, and new codes and refinements were incorporated and discussed with the team while analyzing the remaining data. Subthemes were organized according to the core constructs of NPT [16,17,24]. We selected NPT to understand how the use of AISaT would need to be embedded in the day-to-day routines within the hospital, rather than focusing on a list of elements that could influence its successful implementation. However, we have used the Consolidated Framework for Implementation Research (CFIR) in a separate publication to develop the implementation strategy for rollout and the Non-Adoption, Abandonment, Scale-Up, Spread, and Sustainability (NASSS) framework to understand how AI implementation was affected by different factors in previous research [19].

We completed the following steps: (1) familiarization with the data started during data collection and continued through successive readings of the transcripts and field notes; (2) generating initial codes involved inductive coding of the data;

XSL•FO
RenderX

(3) refining the codebook led to the organization of codes into units of meaning, still quite attached to the data and similar to what would have been topic summaries; and (4) the NPT constructs were then brought into the analysis and initial codes were reexamined, searching for more interpretive content relevant for each of the constructs.

Different data sources were analyzed together, adding depth to the analysis. Summaries of data from all informal conversations, PAG meetings, conversations, and workshops with national leaders were aggregated with an indication of the respondents supporting it. Feedback from wider stakeholder engagement was initially analyzed separately, leading to the generation of summaries and informing recruitment and topic guide development for ethnographic observations and semistructured interviews. For instance, informal conversations with stakeholders suggested different hospital staff needed to be involved in deciding on the use of AISaT. Ethnographic observations suggested that those staff would be part of committees that would review and decide whether changes or the introduction of AISaT would go ahead. Subsequent semistructured interviews tried to unpack who would be leading the implementation and cascading those decisions at the local level. The aim of this study was not to quantify responses; instead, we have indicated within the findings section the agreement or disagreement between data sources, or the origin of the data supporting the different themes.

Rigor

The workshop with national leaders was recorded to facilitate accurate note-taking by researchers (YB-M and SM) and support credibility. Likewise, semistructured interviews were recorded and transcribed to capture hospital staff's accounts accurately. Field notes were taken during each follow-up conversation with national leaders and informal conversations with hospital staff. Field notes were then summarized via a Rapid Research Evaluation and Appraisal Lab (RREAL) data collection sheet [25] that reflected areas of agreement and disagreement, as well as areas requiring further data collection. Informal conversations were not audio-recorded as they were exploratory and aimed to understand the level of involvement of different professionals and their initial reactions. Once relevant staff had been identified through informal conversations and the study had received ethical approval, semistructured interviews were conducted with further staff in relevant roles to avoid participant burden, enrich data collection, and facilitate confirmatory triangulation. This approach, together with the ethnographic observations, allowed a more diverse and prolonged engagement of the researcher with hospital staff in relevant roles. The resulting sample size of the interviews was smaller but more relevant than those involved in informal conversations. To reinforce dependability, we kept a record of every interaction with stakeholders in a spreadsheet, field notes were made during informal conversations and ethnographic observations, and interview transcripts were checked and stored together to facilitate an audit trail. An entry was added to a reflexive diary after every interaction with the stakeholders, regular discussions were held with the rest of the research team, and quotes from interviews are provided throughout the findings to ease confirmability. Finally, while the aim of this study was not to produce generalizable findings but to inform the development of AISaT and the next stages of the trial, we described in detail the types of data collected and the setting, so other researchers can gauge the contextual similarities.

The implementation research team consists of experts in digital health, implementation science, medical sociology, and qualitative research. Researchers involved in data collection and analysis were not involved in AISaT's development other than by providing our insights to members of the wider team, including engineers working on the development of AISaT. Hence, our knowledge of AISaT was limited to the information shared with participants and discussed during interviews.

Ethical Considerations

The Air Safety Trial was approved by the Health Research Authority and Health and Care Research Wales Ethics Committee (ref 25/EM/0156). All participants involved in ethnographic observations and semistructured interviews provided informed written consent.

Results

Stakeholder Characteristics

We involved 12 public contributors (through the PAG) living around the Greater London area. All have a long-term condition and experience of care in multiple hospitals. Contributors were aged between 20 and 80 years, included 8 women and 4 men, self-identifying as Black, African, Caribbean, or Black British (n=4); Asian or Asian British (n=4); and White (n=4).

A total of 58 stakeholders were involved in conversations around the project and its potential implementation: 23 were involved in a workshop with national leaders (including senior NHS staff and policymakers, industry experts, and academics), and 8 national leaders took part in follow-up conversations, such as the 27 hospital staff who were identified either using online searches or purposive sampling within the 2 clinical trial hospitals.

Ethnographic observations were conducted with 4 hospital staff and semistructured interviews with 18 staff members, including 11 male and 11 female staff from Black, African, Caribbean or Black British (n=2), Asian or Asian British (n=8), and White (n=12) ethnic backgrounds. Staff interviewed had not been observed. Of 56 hospital staff members initially approached for observation and interviews at site 1, 5 declined to participate (feeling other colleagues were better placed to respond [n=1], having no time [n=4]). A further 22 did not respond to our contact, and with another 7, communication was lost after having expressed initial interest.

The job roles of those observed or interviewed were clinical (n=11), IPC (n=3), estates (n=6), and roles within the management of the hospital (n=2). Among those interviewed who shared job-related details, they had been in post for an average of 10.4 (SD 7.32) years (min=1; max=26), and most ranked the importance of ventilation within their roles as high (min 1 – max 10): 1 or 2 (n=1), 4 (n=1), 8 (n=3), and 10 (n=4).

Hospital staff interviewed or involved in follow-up conversations as part of the wider stakeholder work described

XSL•FO
RenderX

several tasks undertaken as part of their role to reduce airborne disease transmission (summarized in Table 1). Their main role was isolating or managing patients with known airborne-transmitted infections, placing them in appropriate spaces (according to their risk level), or advising others on how to do this.

All stakeholders and participants provided their views on the hypothetical implementation of AISaT rather than during or after interacting with it. This work informed further development of AISaT prior to its testing and rollout.

Table . Hospital staff's reported roles in keeping the hospital free from airborne disease transmission.

Category	Specific tasks
Room usage and air changes	• Assessing room-use changes • Checking that required air changes per hour are being achieved within rooms
Ward and patients' oversight	• Management and placement of patients with known airborne-transmitted infections (including isolation prioritization) • Conducting ward audits • Daily review of side-room usage • Ensuring adequate testing of patients
Staff supervision	• Ensuring staff are aware of infections and alerting IPC[a] when needed • Ensuring compliance with IPC training and practices • Ensuring staff have adequate protection and providing related advice
Ventilation system management	• Ensuring the ventilation system is functioning adequately • Following Health Building Notes and Health Technical Memoranda
Liaison with other teams	• Continuous liaison with estates regarding ventilation issues • Attending multidisciplinary quarterly ventilation meetings within the trust

[a]IPC: infection prevention and control.

Main Findings

Findings are organized according to the 4 constructs of NPT: coherence, cognitive participation, collective action, and reflexive monitoring (Table 2).

XSL•FO
RenderX

Table . Summary table.

NPT[a] construct	Focus	Subthemes	Key insights	Expected enablers	Anticipated risks and barriers
Coherence	Understanding how the intervention could work in the hospital environment	• Current practices to mitigate airborne disease transmission • Shared goals • Anticipated barriers and facilitators to adoption of AISaT[b] • Anticipated requirements for AISaT's implementation	• Understanding of value varies • Organizational staff are optimistic, while clinical and IPC[c] staff are more skeptical or cautious	• Use in outpatient areas • User-friendly interface • Inclusion in national guidelines • Senior leadership buy-in	• Conflicts with fixed ventilation systems • Old NHS[d] building infrastructure • PPE[e] and handwashing seen as higher priority • Focus on aerosol particle dispersion versus air changes
Cognitive participation	Developing interest in the implementation and enhancing its relevance	• Decision-making and ownership of AISaT • Content and design of AISaT • Timing of AISaT's implementation	• Engagement depends on multidisciplinary buy-in and having designated "champions" to lead	• Multidisciplinary decision-making • Local champions to cascade information • Potential (though risky) mandated use	• Lack of staff agency or ownership • Staff burnout and workload • Financial constraints • Not seen as a top priority in comparison with patient care
Collective action	Planning for the implementation and rollout	• Identification of resources • Generalizability and sustainability	• Implementation is anticipated to be hampered by technical fragmentation and the cost of maintaining mitigation strategies	• External contractors to manage tool (reduces staff workload) • Integration with EHRs[f]	• Costs of potential implementation and maintenance of mitigation strategies • Restricted data access • Expected fragmented AI tool use within trusts • Lack of universal EHR systems
Reflexive monitoring	Planning for the evaluation and refinement	• Providing key metrics and assurances • Futureproofing AISaT	• Success is expected to be facilitated by long-term evidence and expert validation of effectiveness	• Expert validation • Strong evidence from clinical trials	• Lack of a clear plan for future refinement support • Uncertainty regarding postproject sustainability

[a]NPT: normalization process theory.

[b]AISaT: Artificial Intelligence Air Safety Tool.

[c]IPC: infection prevention and control.

[d]NHS: National Health Service.

[e]PPE: personal protective equipment.

[f]EHR: electronic health record.

Coherence (How Do Stakeholders Understand AISaT and Its Recommended Mitigation Strategies?)

Current Practices to Mitigate Airborne Disease Transmission

The most common mitigation strategy used is the building management system (or air handling units) that represents the centralized ventilation system of the building. Our ethnographic observations captured how different units provide clean air and extract used air from hospital areas. Typically, the placement of ventilation outlets is decided by architects, and any changes to these require an authorizing engineer's sign-off.

> *HEPA filters available [within ventilation outlets] all have been allocated at design stage. It could be possible to uninstall as needed, things can change, and more things can be added. The authorising*

XSL•FO
RenderX

> *engineer (AE) for ventilation would need to agree and sign off.* [Ethnographic observations, informal conversation, estates, HS104]

However, some interviewees raised concerns where portable HEPA filters have been trialed within the trust, mainly related to space constraints where they can obstruct staff's work or pose a hazard to patients.

> *if they [agitated patients] move the locker or they're able to reach that and push it, it poses quite a danger of the actual physical unit being dropped and damaged, but also it then causing harm to somebody if it falls on them.* [Semistructured interview, clinical, HS102]

Further concerns included the impact of a device, such as a portable HEPA filter, increasing room temperature and noise; competing for limited sockets; lacking evidence proving their effectiveness; taking significant time to clean the air; requiring maintenance plans and a large number of devices in wards; and their associated costs. Despite these reservations, staff who had HEPA devices within their work environment agreed they would be happy to have HEPA devices affixed to walls.

Other mitigation strategies reported included: prioritizing isolation of potentially infectious individuals, faster patient testing, improving hand hygiene, reducing the number of visitors when necessary, using personal protective equipment (PPE), constructing further isolation single or side rooms in new buildings, allowing longer cleaning times after conducting procedures with patients with certain infections, and using curtains or transparent partitioning screens. An interviewee flagged that some of this equipment is hardly being used after the COVID-19 pandemic due to limited storage and risk of falls for patients though. Also discussed in the interviews was that some spaces contain or have tested air filtration machines (such as HEPA mobile devices); however, this is not a generalized approach across the hospital.

> *As far as I know there are some units like I think dental units that have filtration machines. But it's not something that we would routinely put in all...areas such depends... on the risk assessment of what sort of procedure they're going to undertake in the room, the size of the room, obviously. What's the current air change? What's the supply and extract...within the room. So yeah, we have some mitigations like filtration machines, but they're not widely used...* [Semistructured interview, IPC, HS110]

The NHS estate is generally old and comes with its own limitations (eg, lack of ceiling space, limited ability for new infrastructure projects, and limited electrical supply and clinical space).

> *...when we build new buildings (...) there's always more and more required from an engineering point of view. (...) you end up with a situation that (...) there's just no space. (...) a good idea is to look at the ceiling heights and make sure that you've got more space. The problem is, the more space you have and the higher the building, the more expensive it is...* [Semistructured interview, management, HS101]

Natural ventilation was not generally recommended but does seem to be the only option in some older buildings. Despite this, some staff in the older buildings had to work with sealed windows due to nearby building works.

Overall, mitigation strategies were welcomed as a way of protecting staff and patients, and staff were happy with current strategies in place but did share that they would appreciate further isolating rooms and "20 air changes that would be great. (...) that was the envy that I held for other areas. But in terms of [wanting further] technology, not really" (semistructured interview, clinical, HS103).

Hospital staff also reflected on the high costs of enhancing ventilation, the limited impact of mitigation strategies on air changes, and the limited number of isolation rooms. Due to financial constraints and the need for space, sometimes hospital areas have been repurposed without following the required procedures, resulting in tension between teams.

> *...they'll just change the use of a room, but obviously there's a process to go through to do that, to make sure it's safe, but not everybody understands that, and to be honest, they think (...) it's a bit of a fuss, and they know it's going to cost them money (...) And if they don't have any money, but yet they desperately need to do something in a space because it's the only space they've got, they're just going to go ahead and do it.* [Semistructured interview, management, HS101]

Shared Goals

Hospital staff interviewed reported that the aims of AISaT were a good fit with the trust's overall goals of keeping staff and patients safe. Individuals involved in the wider engagement work raised that hospitals are perceived as risky spaces where people can catch infections, so they need to be protected, prioritizing the most vulnerable. Staff interviewed also described that the AI tool could be a source of reassurance to demonstrate the effectiveness of existing practices and support faster decision-making.

> *I think it would give everyone reassurance. Certainly, in terms of pandemic or significant heightened risk, it would be helpful, not really for the confidence of the staff, but also the safety of the staff. And it would overcome the challenge we have in not being able to change the footprint of the healthcare facility.* [Semistructured interview, clinical, HS103]

Interviewees were positive about the potential impact of AISaT on their work, as it would make their environment safer, if it did not take a lot of their time to input the data into AISaT. A couple of staff raised that AISaT could guide hospital staff but also architects and engineers to keep the hospital air clean and optimize staff capacity and space use.

> *If we had a tool like that (...) every time that you had a new space (...) you could run the whole of the design through the AI software, and it would identify where you may (...) add in things at the base build level (...)*

XSL•FO
RenderX

> *I can't see a problem with using models to help to help give you the optimum design for a room. It's a good idea.* [Semistructured interview, management, HS101]

Anticipated Barriers and Facilitators to Adoption of AISaT

Hospital staff observed and interviewed, and wider stakeholders, pointed out that new buildings still have ventilation issues, but that it may be easier to implement AISaT in these settings as they are more likely to have standardized room measurements to comply with the latest regulations. Hospital staff also raised that AISaT would be particularly valued in areas where ventilation cannot be improved systemically across the site (eg, old buildings).

> *...in an ideal world, we would have something like a built-in air conditioning system or something that was systemic across the building. But being such an old building and I think the electrical grid is the reason why we couldn't have anything more systemic built-in. So, we're having to rely on things like windows and maintaining good infection control with our isolation in our side rooms...* [Semistructured interview, clinical, HS102]

Concerns related to staff resistance were raised as:

> *it is quite tricky in a Trust to get them [staff] to accept changes like this and to get them to think it's worthwhile because they can't. They don't necessarily see the value of it.* [Semistructured interview, management, HS101]

Not seeing a clear benefit from AISaT and its recommended mitigation strategies or not experiencing an immediate benefit were anticipated to be the main barriers to staff engagement. Additional anticipated barriers included the lack of awareness and knowledge around ventilation and the impact that the positioning of mitigation strategies can have; limited digital understanding and guidance; lack of preventative mindset; and limited trust or confidence in AI in general. In contrast, some hospital staff felt the pandemic had raised staff's awareness of ventilation, and they were now actively requiring ventilation improvements within the trust. Trust in AI was also described as a potential facilitator for the adoption of AISaT, as staff reported AI is becoming more present in daily lives and at work. Hospital staff's workload could be a source of staff resistance; however, if the AI could reduce workload, it could be seen also as a facilitator.

Some hospital staff interviewed and involved in wider engagement work referred to inpatient areas or wards as areas where the potential implementation would not be a priority given the number of people working within those areas, with the associated movement, the lack of space, connectivity issues within hospitals, and the number of beds which would require several HEPA devices. Wider stakeholder engagement work involving hospital staff encouraged the use of mitigation strategies already existing within the wards, and that AISaT should consider the space available and size of any new suggested strategies. The potential implementation of AISaT in outpatient areas, isolation rooms, labs, or neonatal spaces was more acceptable.

> *...you have 6 - 7 patients in there and then some patients, they bring quite a lot of their stuff...they put a lot of things in their cupboard and we have only a small one (...) they need to consider... space (...) It's another tool [HEPA device] that they don't see any results from it. It's just there physically. (...) they're still getting sick. They're still getting infection (...) I don't think anything would eradicate that. They don't see any success to it. They're just thinking, you know, it's another thing that's occupying the space of the patient, that it's unnecessary. It's another tool that...they have to make sure it's on (...) let's be honest, it's not a priority for us to do that. It's not our job...to do that (...) it shouldn't be us who needs to check it.* [Semistructured interview, clinical, HS108]

Besides some clinicians not feeling that dealing with AISaT and the associated mitigation strategies would be part of their role, an interviewee was struggling to think in what area of the hospital the AI tool could be of use:

> *...[AISaT] it's an interesting tool... in certain areas like theaters or operating rooms I don't think they would be useful because if the air changes... will have to depend on the... centralized ventilation system, because...they're essentially critical departments...With the AI tool, I'm not sure where we can use them.* [Semistructured interview, IPC, HS110]

Stakeholders mentioned that PPE may be seen as a cheaper alternative to the mitigation strategies AISaT may offer. This was in contrast with several staff members mentioning the inconvenience of working with PPE: "It's not amazing, to be honest, if you can imagine doing quite a difficult procedure over a pronounced amount of time wearing masks and PPE is not amazing." (semistructured interview, clinical, HS112), particularly in buildings with poorly balanced temperatures.

Hospital staff did not perceive the need for further mitigation strategies in their workplace, and some failed to see the point of AISaT. However, highlighting AISaT advantages, such as staff and patient benefit and cost reductions, may be needed to make AISaT attractive to hospital staff (see the "Reflexive Monitoring" section).

> *People's time is the crux. (...) you can usually win hearts and minds from people if they can really see a benefit to a bit of effort that they might have to put in during implementation. But if they can see that this is going to reduce my staff sickness is going to mean I've got less patients getting cross infected on the wards. And I'm going to have my beds all open and not closed due to infection. I think they're massive pluses.* [Semistructured interview, management, HS111]

XSL•FO
RenderX

Anticipated Requirements for AISaT's Implementation

Hospital staff and wider stakeholders suggested using a series of strategies that could facilitate implementation and a future rollout of AISaT. The most frequently discussed strategy was gaining buy-in from senior staff (eg, the chief executive, IPC director, and authorizing engineers), board members, and champions on the ground.

> *Your best bet is to go to the director of infection control and prevention. If he was interested and if he thought it was an advantage, I would say that would be a good place to start. (...) If you could get [Chief Executive] interested, obviously that would be the best person (...) if he was interested and thought it was a good idea and it was protecting his patients.* [Semistructured interview, management, HS101]

Clear communication was recommended, with staff and patients, about what is needed from users and about AISaT complementing human decision-making. Also raised was the need to transparently report how AISaT is better than a human, how the AI works, and clarity on its benefits as well as its limitations.

Wider stakeholder work suggested that gaining support from recognized bodies would also be required, through appearance in guidelines such as NICE (National Institute for Health and Care Excellence); dissemination at national meetings (National Clinical Engineering, Ventilation Teams Meetings); and having the support of medical organizations.

Providing training on AISaT to hospital staff who would use it or work around it could facilitate its uptake and build trust in the AI tool. Ensuring all staff are involved in training could be positive, to promote inclusivity and facilitate the acceptability of AISaT. Stakeholders thought training should include ventilation, the relevance of positioning the mitigation strategies, and AISaT itself. However, one interviewee doubted the need for training as they envisaged the implementation of AISaT as a one-off interaction with an external contractor. Another participant cautioned that colleagues may not react well to training on ventilation when they are faced with poor ventilation levels, high temperatures, and are unable to open (sealed) windows or use fans due to electrical supply limitations and scarcity of sockets.

> *I think it [training] would be beneficial, but I think the response would be...interesting. (...) we've had such challenges every summer...where it's been very hot on the ward and we've had to use fans and...We're trying to as best as we can to isolate these patients who've got airborne infections to then talk about ventilation and the joys of ventilation. I think that may not be received as well.* [Semistructured interview, clinical, HS102]

Hospital staff were generally cautious and expressed concerns about the need for appropriate validation and assurances that the AI would be providing safe recommendations. This caution was not necessarily negative, as "proper use of it could really help ease our workflow (...) having it do the recommendations for you and then you just check it makes it convenient" (semistructured interview, IPC, HS113). However, some interviewees stated they could not comment on the acceptability and potential to implement AISaT without having seen it, and knowing in detail how it had been trained, or having a summary of its impact when used within the hospital environment.

Cognitive Participation (How Will Staff Engage With AISaT and Its Recommended Mitigation Strategies?)

Decision-Making and Ownership of AISaT

It was agreed that early staff engagement would be required, not only to gather their views to develop AISaT but to facilitate the decision-making process of who should take ownership of the AI tool. Like current practice, the development of a multidisciplinary working group would ensure the decision-making was not "down to a person" but to a multidisciplinary team to determine whether AISaT should be used, and the details on when and how.

First funding would need to be sought. Then, the contractor (HS104 and their team) would gather evidence and information about AISaT. Then they would present the idea to the Committee. If the Committee agrees, then they would need to ensure buy-in from IPC. Discussions between estates, medical, and IPC teams would need to happen to explore what effect this is going to have (eg, having to close an area to assess the impact of adding new HEPA filters or placing them where AISaT suggests) and would provide suggestions as to how this effect could be managed. Nine out of 10 times medical teams and IPC will say no; the area cannot be closed, and there will need to be negotiations until there is agreement between all parties. This may mean that the implementation does not occur soon, as pointed out during ethnographic observations.

A few interviewees suggested a similar process of group decision-making when choosing among the recommended mitigation strategies suggested by AISaT.

> *...the decision will not ultimately be from one person. It's always a multidisciplinary thing.You'll have to present the options to the end users, to other experts to see what is really the best option in practise.* [Semistructured interview, IPC, HS110]

> *Say it's showing that we need HEPA filters... 100% estates, finance, infection control would all want to be involved in that. And it wouldn't be...just down to the ward manager to say, yeah, we're going to buy four HEPA filters for this ward.* [Semistructured interview, clinical, HS203]

Another interviewee suggested that the implementation of AISaT should come as an imposition from IPC rather than an exercise requiring hospital staff's engagement and acceptance.

> *I think for this realistically it wouldn't so much be staff engagement and encouragement, it would be imposed. So, if it's supported, it would be supported on one ground or another, be a patient safety or financial and they would be imposed.* [Semistructured interview, clinical, HS103]

On this line, another interviewee wondered whether the strategies shared by AISaT would be recommendations

XSL•FO
RenderX

(something staff could choose to follow or not) or actual policy they would need to follow. If the latter, current policies in this area come from the trust and from IPC, so any AISaT recommendations should be aligned with those existing policies.

Despite most interviewees being in favor of a multidisciplinary effort involving IPC, clinical, estates, and operations for AISaT's implementation, none of them suggested themselves or their team would be involved in leading the implementation of a tool such as AISaT. Only one IPC staff member agreed when prompted by the researcher that IPC may be best positioned to lead on AISaT's implementation.

> *Well, it could fall under IPC... 'cause sometimes with estates, if we let them lead on it...some trusts outsource them. So, I do get the point of IPC trying to lead on this... I would agree, it's part of the things that we check and that we try to enforce as a team to ensure that the ventilation throughout the hospital is good. So yeah, I think we can take a lead on it.* [Semistructured interview, IPC, HS113]

Most hospital staff involved in interviews and wider stakeholder work placed ownership on IPC: "infection control [IPC] (...) can oversee the project because (...) when it comes to infection, they need to be involved in that" (semistructured interview, clinical, HS109), given their cross-trust presence. Contrasting this, an IPC staff member insisted this would be a role for the estates team, who have expertise in ventilation and lead critical ventilation meetings.

> *I think it should be a multidisciplinary approach on the initial stages until we all have a full understanding of...how it works in practise. But I think ultimately it will have to be estates leading on it. Obviously with our...expert input from ...clinical areas, IPC, ventilation engineers...Anything ventilation related would fall under estates responsibility.* [Semistructured interview, IPC, HS110]

> *I think the IPC would probably come to someone in estates and we would accompany them to that space, and if there's a measurement issue, we could figure that out... But they [IPC] would decide where is the best place to put any device that might impact have a clinical impact on patients or staff.* [Semistructured interview, estates, HS116]

There were mixed opinions among hospital staff involved in interviews and wider stakeholder work who felt that, as long as staffing costs were funded, then the inputting of data into AISaT could fall under estates, engineers, or IPC.

There was more agreement in the cascading down of decisions where estates and facilities teams would be the ones in charge of cleaning, maintaining, and replacing the recommended devices or equipment to mitigate airborne transmission of infections. Interviewees further raised that if the HEPA filters are considered nonmedical devices, then installation could come from estates.

Content and Design of AISaT

There was a difference in opinions on how AISaT should be designed to facilitate its uptake. Some hospital staff thought a ranking of potential mitigation strategies that are the best at preventing airborne transmission of infection should be provided. They felt that, regardless of the lack of availability within the trust, staff having a comprehensive list may help the trust push to purchase these items, especially in preparation for a future pandemic. Others suggested a ranked list should only contain mitigation strategies available within the trust as it may create anxiety among staff if they do not have access to the best rated options, or they may get "overwhelmed" (HS113) by a list of endless options.

Interviewees shared some technical factors that may help with the implementation of AISaT; this included ease of use, integrating AISaT within NHS systems (or whitelisting websites or the AI tool); compatibility with diverse devices such as computers and iPads; facilitation of automatic data entry through LiDAR software to capture room dimensions, or importing patients' data from electronic health records. Further suggestions on AISaT's functionality included its ability to perform a risk assessment for the use of physical separation equipment and an air filter (eg, HEPA device) with a prediction model to alert staff of outbreak risks. Also suggested was demonstrating through AISaT the variable cost of a mitigation strategy, which could help to choose from what is available in the market. The ethnographic work suggested it would be helpful if AISaT could keep a record of HEPA filter performance and dates of installation, and other stakeholders suggested AISaT could include reminders to staff to attend PPE fit testing. It was suggested that AISaT should account for movement and that it should consider the fact that staff may be contagious. This is particularly relevant in wards with more vulnerable patients who are usually seen by a range of staff. Two clinical hospital staff members also reflected on the need to involve clinical experts to tailor the AI tool to the needs of different areas with different ventilation needs.

> *...you need people or clinical staff that understand their local areas so you can tailor it for the area that you're using it in because each procedure each diagnosis, medical diagnosis, etcetera has a different outcome for every different area* [Semistructured interview, clinical, HS112]

Timing of AISaT Implementation

Among interviewees, many agreed that implementation of AISaT was not a priority now, but it would become a priority in future pandemics and, hence, highlighted the relevance of this research. However, about one-third of those interviewed felt AISaT is currently relevant for the prevention of infection.

> *I'd hope that you could use it every day, because I'd hope it would really help us to prevent cross infection, prevent outbreaks...So I think... for prevention, be proactively using it... I'm sure it's useful...in a pandemic or an outbreak as well. But...I'd want to use it before you got to that point. I'm completely*

XSL•FO
RenderX

> *with...try to prevent it and control it.* [Semistructured interview, management, HS111]

Collective Action (How Will Staff Work to Implement AISaT and Its Recommended Mitigation Strategies?)

Identification of Resources

Often, hospital staff were unclear about who would fund the potential implementation and rollout of AISaT and its recommendations. When hospital staff had suggestions, they mentioned:

> *...maybe NHSE can provide a sponsorship if we're looking into a bigger perspective of the whole UK.* [Semistructured interview, clinical, HS109]

However, there were also suggestions that "executives would agree to fund from a central pot because it's not one area (semistructured interview, management, HS111)" or that Integrated Care Boards (ICBs) should be able to provide funding or suggestions of possible funding. The need to develop a business plan to cost staff and time required for the implementation of AISaT and its recommendations was mooted.

Likewise, hospital staff interviewed and involved in wider stakeholder work raised concerns about the costs to maintain the mitigation strategies (eg, HEPA filters), as well as the impact of maintenance on staff's workload and capacity. They also referred to the high costs of making changes to ventilation systems and that any changes must meet the latest guidelines or Health Technical Memoranda (HTMs). Conversations reflected the items' costs (eg, AISaT and implementing its recommendations), the associated staff costs (eg, data entry into the AI tool, delivering on maintenance plans, and so on), and concerns as to whether the trust would be willing to invest in AISaT and its recommended mitigation strategies even if the trial proved its effectiveness.

> *I think with the amount of (...) maintenance involved like routine cleaning, routine maintenance, reporting of equipment that's faulty, it's already a bit of a challenge just on day-to-day with clinical items that patients use(...) It can be quite challenging on a large [busy] ward (...) and a substantial service to then add another item in there that isn't immediately impacting the day-to-day tasks or benefiting the patients in that immediately visible way. It can be a barrier.* [Semistructured interview, clinical, HS102]

As previously mentioned, an interviewee envisaged AISaT being used by an external contractor, as the rooms of interest would usually maintain their size and characteristics, and so the need to enter data into AISaT and to choose a recommended mitigation strategy was perceived as a one-off action. However, the one-off use was shared by most interviewed: "you probably only use it once, and that's probably it (semistructured interview, clinical, HS202)," who would only use it again if there was a change in the placement of objects in the room. Such implementation was seen as advantageous as it would not impact staff's capacity or require further training. In successive interviews, further hospital staff members agreed this could be a feasible option to facilitate implementation without negatively impacting staff's workload.

> *...Like a contractor. (...) Yeah, like I said... it's just another thing to think of and we already have a load of things...you know, pressured. So you...don't want to commit to something that you can't really [do].* [Semistructured interview, clinical, HS108]

Generalizability and Sustainability

Hospital staff interviewed raised that communication among staff could become a barrier to the implementation of AISaT when different members of staff work on similar issues but from different teams, such as the NHS and contractors working together. Also raised was fragmentation in how different trusts incorporate new devices, and a lack of guidelines on how to integrate new digital developments into hospitals, which would take a long time to develop. Unsurprisingly, the acquisition of funding for AISaT's implementation may also be different across trusts.

Access restrictions were also identified as barriers by hospital staff involved in wider stakeholder work regarding data on room characteristics (eg, size, height, objects in the room, and ventilation characteristics) and floor maps. Difficulties may also arise when attempting to import data, as floor maps are often in PDF format instead of in electronic formats that could be easily exported. Different hospital buildings, even within the same trust, also have different levels of data available within their records. This would, therefore, require manual data entry, which would require costing staff time.

The ethnographic work suggested that the potential implementation of AISaT and its recommended mitigation strategies would need approval from multidisciplinary committees. The implementation timing would need to be agreed, as it is not uncommon for projects to be approved but face delays in green lights for actual implementation. Within the approval process, the different management systems in different buildings (eg, privately funded or NHS-maintained) would need to be factored in, as there are differences in these processes even for buildings within the same trust. In privately funded buildings, for instance, there is the need to factor in the landlord's position regarding the implementation of AISaT and its mitigation strategies.

> *Don't forget that this building is not owned by the NHS. It's rented. They [Landlords] would need to be in agreement with it, so you'd have to bring them in if potentially you were thinking about that [installation of mitigation strategies].* [Semistructured interview, clinical, HS112]

Reflexive Monitoring (How Will People Assess the Success of AISaT and Its Recommended Mitigation Strategies?)

Providing Key Metrics and Assurances

Hospital staff interviewed referred to the need to prove the effectiveness of AISaT, addressing the 3 NHS priorities: costs, effectiveness, and implementation practicalities. Key metrics should include those of interest for IPC, such as the extent

XSL•FO
RenderX

AISaT reduces infections; whether it has an impact on staff, patients, or both; whether it leads to bed-day savings or fewer closed beds, fewer appointment cancelations (in outpatients); and reduction of outbreaks. But also the extent to which it is effective, such as the impact on air changes per hour; its energy and financial savings (eg, reduction of costs to ventilate an area). It would need to produce benefits on safety grounds but also need to lead to financial benefits.

> *...they're pushing for sustainability and they're pushing for capacity, bed capacity, patients flow and reduction of infections as such. And it would help the hospital if it's done properly, that they have seen all the areas of concerns before they actually invest into something. Because...we've seen some projects that had been invested, they've spent quite a lot of money, but it hasn't been pushed through...before the commitment to be set out there, they really need to make sure that this is... worth it to the benefit of the hospital, the patients and the staff.* [Semistructured interview, clinical, HS108]

Beyond the health economics assessment, hospital staff also anticipated the need for expert validation to explore to what extent AISaT's recommendations are aligned with gold standards, and evaluate differences between human and AISaT, to explain why AISaT is better.

> *As long as we're happy that it's...tested and validated and that there's always the double check until we've got confidence that it does actually predict. I think there's got to be that validation process continually (...) sort of accuracy tolerances built in. (...) Because it's there's a lot at stake if it takes us to the wrong decision or the wrong action.* [Semistructured interview, management, HS111]

Participants also suggested the need for live monitoring of the quality of the air in parallel to the implementation of AISaT and its recommended mitigation strategies.

> *If introducing HEPA, there would need to be a live measurement of pressure, as this will tell them when the filter is filled and needs changing.* [Ethnographic observations, notes from informal conversation, estates, HS106 and HS107]

Future-Proofing AISaT

Interviews with hospital staff and conversations with those involved in wider stakeholder engagement referred to the need to develop AISaT to be able to adapt to different technologies (eg, PC, phones, tablets, and advancements in these devices), settings (eg, different hospitals and room characteristics), and policy, as these are likely to evolve with time. When considering preparedness for the future, an individual involved in the wider stakeholder engagement work raised the need to confirm who would be responsible for refining and providing technical support to hospitals following the research trial or development phase of AISaT, showing concerns about the ongoing feasibility and ownership of AISaT.

Discussion

Principal Findings

This paper explores key stakeholders' anticipated barriers and facilitators to the implementation of AISaT to improve its acceptability at the implementation stage [11]. Our findings, organized around NPT, suggest that while AISaT and its recommended mitigation strategies are generally positively received, their wide implementation may be hindered by its associated maintenance and staff costs; the lack of perceived need and immediately observable benefits; and organizational differences within trusts and hospital buildings.

Coherence (How Do Staff Understand AISaT and Its Recommended Mitigation Strategies?)

Coherence was variable across hospital staff. While hospital staff in organizational roles anticipated AISaT's usefulness in their daily work, hospital staff in clinical and IPC roles were more divided on the value of the AISaT tool, with some reluctant to share their views on AISaT's potential prior to reviewing results from a clinical trial. This may reflect staff not seeing the clinical need required to facilitate AISaT's implementation [26]. The main risks anticipated to AISaT's implementation were that existing strategies used within hospitals, such as fixed ventilation systems, are hard to change and may be different from the strategies AISaT may recommend, which are likely to involve portable HEPA filters. Additional anticipated risks to AISaT's implementation were the potential challenges associated with HEPA devices and the lack of standardized rooms in old NHS buildings. A number of additional potential challenges have also been identified, such as PPE being seen as a more economical alternative; that handwashing and isolation rooms are being prioritized in current guidelines [27]; and that hospital staff felt improving air changes should be prioritized, whereas AISaT's current prioritization is focused on diminishing aerosol concentrations. The main expected enablers identified related to the implementation of AISaT in outpatient areas; ensuring the AISaT interface is easy to use and responsive to staff's needs; buy-in from senior staff; and its dissemination through national guidelines. In line with findings from previous research, there is a need to communicate that using AISaT does not remove human involvement from the decision-making process [11,15]. AISaT may be more likely to enhance its coherence in the hospital environment by tailoring its implementation approach to the ways of functioning of different hospitals, while providing staff with proof of the effectiveness of the tool in a way that is relevant for the metrics used within the NHS (eg, justifying cost savings and benefits for staff and patients).

Cognitive Participation (How Will Staff Engage With AISaT and Mitigation Strategies?)

Cognitive participation is anticipated to depend on multidisciplinary decision-making on the implementation of AISaT and its recommendations and having champions to cascade decisions down and encourage implementation on the ground. Anticipated barriers to implementation could be overcome by imposing the use of AISaT rather than relying on

XSL•FO
RenderX

staff's engagement. Imposition, however, would come with the inherent risk of losing the feeling of shared ownership, which would add to a lack of agency among staff, becoming a potential implementation challenge [28]. The potential implementation of AISaT was often not perceived as a priority, in a climate where staff feel overloaded, there are increasing financial constraints, and direct patient care is the most pressing matter. Such a lack of agency, particularly among those in clinical and estates roles, while raising the need to involve multidisciplinary teams may reflect the ways of working within the NHS as well as the nature of the AISaT tool falling in a confluence of responsibilities among different roles.

Collective Action (How Will Staff Work to Implement AISaT and Recommended Mitigation Strategies?)

Collective action is anticipated to be impacted by costs associated with the implementation and maintenance of AISaT and its recommended mitigation strategies; the fragmented implementation of AI tools and devices within trusts; restrictions in accessing data; and variations in approval processes within trusts and buildings. While it was suggested that AISaT could be best integrated within a comprehensive electronic health record, this comes with an added layer of difficulty since electronic health record systems are not universally used across UK hospitals, which would require further adaptations to different systems used. This is a known generalizability issue previously identified in the literature [29,30]. An expected enabler identified was the external implementation of AISaT (eg, by a contractor), which would reduce its impact on staff workload.

Reflexive Monitoring (How Will People Assess the Success of AISaT and Mitigation Strategies?)

Reflexive monitoring is anticipated to derive from expert validation and evidence on effectiveness and impact prior to the implementation of AISaT. However, previous research has also pointed out that such evaluation should continue after implementation [31]. An anticipated risk identified is the need for future refinement support that is unclear after the end of this project.

Comparison With Prior Work

Our findings are consistent with those of a recent systematic review where AI implementation within hospitals was found to be influenced by understanding and trust in AI, quantification of the benefits of the AI, external validation to facilitate generalization, costs of implementing the AI, training staff and maintaining the AI, and the availability of policies to fund and regulate AI solutions [19]. However, our findings also point toward other critical elements, such as specific environmental characteristics where the AI and mitigation strategies will be implemented (eg, old buildings with busy and crowded wards, with scarce space and overloaded hospital staff), together with the priority given to AISaT and its mitigation strategies within the existing ways of working and measures already available. These findings also align well with a recent criticism of the NHS 10-year plan that AI implementation cannot sort out the whole structural issues within the NHS such as the constrained budgets [32]. Likewise, our findings are aligned with the need to transparently report on how the AI was trained as well as the benefits associated with its use, providing relevant metrics to the end users, and acknowledging the limitations of the tool [15].

Beyond our focus on AISaT's implementation, but very closely related, difficulties in the use of HEPA devices, such as noise, which had been previously reported in the literature [33], did not preclude their hospital-wide use through ventilation outlets installed in ceilings. This is in line with HEPA filters being the most common mitigation strategy used in hospitals to reduce airborne disease transmission within papers included in a recent review [18], subsequent research [34], and aligned with estates technical guidelines [35]. Although the review reported improvements in patients' health outcomes [18], some stakeholders included in our study stressed the lack of randomized controlled trials proving HEPA devices' effectiveness.

Future Research

This work involved extensive stakeholder engagement, capturing views from the public, subject matter experts, and hospital health care professionals, which helped identify common ground, but also areas of discordance between accounts even among professionals with similar job roles. Hence, it responds to the recommendation to include stakeholders at an early stage to facilitate future acceptability of AI tools [15]. This study is not without limitations; we struggled to access the hospital environment for further ethnographic work, only being able to observe estates work within one building. Difficulties may have been aggravated by the timing of our research during flu season. Furthermore, AISaT was under development at the time of conducting this study, and hence neither researchers nor participants had a tangible idea of what the tool would look like. This may have made it difficult for hospital staff and stakeholders to provide more specific responses and imagine what its implementation could look like in practice. Our formal data collection sample was focused on hospital staff from 2 hospitals, but more so from one of them, due to low response rates and delays in getting the relevant authorizations to do research in the second research site. Hospital staff shared their reservations about the AI tool during interviews; however, there may have been a natural selection bias. Through the wider stakeholder engagement (eg, via a national leaders' workshop and follow-up conversations with national leaders), experts from around the country were involved, which helps reinforce the applicability of our findings to other hospitals with caution. Areas of concern such as impact on costs, workloads, and diverse ways of working were raised by all parties, supporting the transferability of our findings to other hospitals and health systems working under limited resources, staffing pressures, and in old buildings. This approach supported our ability to provide analytical generalization, to aid the development of the AI tool and future trial plans. Future research should involve relevant stakeholders in designing user-friendly AI tools and seeking their feedback with demonstrations of the product to visualize the end-product and think of ways to implement it in their practice.

XSL•FO
RenderX

Conclusions

Even when AISaT and its recommended mitigation strategies are not considered a priority within the hospital environment now, in a climate of tight budgets, staff shortages, and heavy workloads, staff expressed their openness to accommodate AI innovations if the effectiveness and benefits are proven and shared with them. Therefore, when planning implementation across hospitals, it will be important to frame this according to the tenets of NPT to enable understanding at the level of the structure of services as well as people's experiences to engender the flexibility that will be necessary to maximize the likelihood of success across different NHS trusts and associated hospitals.

Acknowledgments

We would like to thank all stakeholders involved in this study, from those involved in wider stakeholder engagement work (via patient advisory group meetings, the National leaders' workshop, and follow-up conversations) to those being observed and interviewed as part of this study. We would like to thank Mr Simon Watt (Public Engagement Manager) for his work in recruiting, managing, and facilitating patient advisory group meetings; Ms Anna Byrne (Air Safety Program Manager) for her work identifying and inviting National Leaders to the workshop; and the Air Safety's Project Team for their collaboration and support through this study. Authors attest that no generative AI was used in the development of this manuscript.

Funding

This study is funded by the National Institute for Health and Care Research (NIHR) under its Programme Grants for Applied Research (Ref: NIHR205439). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Data Availability

The datasets analyzed during this study are not publicly available due to their containing information that could identify participants and they did not provide consent to these data being shared outside the research team beyond anonymized quotes in publications. Data supporting the results have been included in this published article. If you would like more information about the nature of the data, please contact the corresponding author.

Authors' Contributions

Conceptualization: YB-M (lead), FS (supporting), CVP (supporting), ZK (supporting), LBL (supporting)
Data curation: YB-M
Formal analysis: YB-M (lead), SM (supporting)
Funding acquisition: LBL (lead), FS (supporting), CVP (supporting), ZK (supporting)
Investigation: YB-M (lead), FS (supporting), SM (supporting), SEC (supporting), ZK (supporting)
Methodology: FS (lead), CVP (supporting), ZK (supporting), YB-M (supporting)
Project administration: YB-M
Supervision: FS
Writing – original draft: YB-M
Writing – review & editing: YB-M, FS, SEC, LBL, ZK, SM, CVP

Conflicts of Interest

None declared.

Multimedia Appendix 1
Determinants of implementation of Artificial Intelligence Air Safety Tool identified through wider stakeholder engagement.
[DOC File, 41 KB - humanfactors_v13i1e93420_app1.doc]

Multimedia Appendix 2
Detailed data collection process followed for stakeholder involvement.
[DOCX File, 17 KB - humanfactors_v13i1e93420_app2.docx]

Multimedia Appendix 3
Ethnographic observations sheet.
[DOC File, 34 KB - humanfactors_v13i1e93420_app3.doc]

Multimedia Appendix 4
Semistructured interview topic guide.

XSL•FO
RenderX

[DOC File, 67 KB - humanfactors_v13i1e93420_app4.doc]

Checklist 1
COREQ 32-item checklist.
[DOC File, 57 KB - humanfactors_v13i1e93420_app5.doc]

References

1. New commission to help accelerate NHS use of AI. Medicines and Healthcare products Regulatory Agency. 2025. URL: https://www.gov.uk/government/news/new-commission-to-help-accelerate-nhs-use-of-ai [accessed 2026-01-15]
2. Fit for the future: the 10 year health plan for England. : Department of Health and Social Care; 2025 URL: https://assets.publishing.service.gov.uk/media/6888a0b1a11f859994409147/fit-for-the-future-10-year-health-plan-for-england.pdf [accessed 2026-10-03]
3. Liao X, Yao C, Jin F, Zhang J, Liu L. Barriers and facilitators to implementing imaging-based diagnostic artificial intelligence-assisted decision-making software in hospitals in China: a qualitative study using the updated Consolidated Framework for Implementation Research. BMJ Open 2024 Sep 10;14(9):e084398. [doi: 10.1136/bmjopen-2024-084398] [Medline: 39260855]
4. Ramsay AIG, Crellin N, Lawrence R, et al. Procurement and early deployment of artificial intelligence tools for chest diagnostics in NHS services in England: a rapid, mixed method evaluation. EClinicalMedicine 2025 Nov;89:103481. [doi: 10.1016/j.eclinm.2025.103481] [Medline: 41357329]
5. Maleki Varnosfaderani S, Forouzanfar M. The role of AI in hospitals and clinics: transforming healthcare in the 21st century. Bioengineering (Basel) 2024 Mar 29;11(4):38671759. [doi: 10.3390/bioengineering11040337] [Medline: 38671759]
6. El Arab RA, Almoosa Z, Alkhunaizi M, Abuadas FH, Somerville J. Artificial intelligence in hospital infection prevention: an integrative review. Front Public Health 2025;13:1547450. [doi: 10.3389/fpubh.2025.1547450] [Medline: 40241963]
7. Ting Wu H, Shuang Li Q, Chen Dai R, et al. Effects of air-conditioning systems in the public areas of hospitals: a scoping review. Epidemiol Infect 2021;149. [doi: 10.1017/S0950268821001990]
8. An IH, Kwak DB, Lee J, Park SH, Yook SJ. Mitigating airborne virus transmission in hospital wards: a study on effects of ventilation system, air purifier and curtains. Indoor Built Environ . [doi: 10.1177/1420326X251360217]
9. Hogg HDJ, Sendak MP, Denniston AK, Keane PA, Maniatopoulos G. Unlocking the potential of qualitative research for the implementation of artificial intelligence-enabled healthcare. J Med Artif Intell 2023;6:8-8. [doi: 10.21037/jmai-23-28]
10. Borges do Nascimento IJ, Abdulazeem H, Vasanthan LT, et al. Barriers and facilitators to utilizing digital health technologies by healthcare professionals. NPJ Digit Med 2023 Sep 18;6(1):161. [doi: 10.1038/s41746-023-00899-4] [Medline: 37723240]
11. Fazakarley CA, Breen M, Leeson P, Thompson B, Williamson V. Experiences of using artificial intelligence in healthcare: a qualitative study of UK clinician and key stakeholder perspectives. BMJ Open 2023 Dec 11;13(12):e076950. [doi: 10.1136/bmjopen-2023-076950] [Medline: 38081671]
12. Gerhold M. 10-year health plan: we need to move from ‘techno optimism’ to ‘techno realism’. The Health Foundation. 2025. URL: https://www.health.org.uk/features-and-opinion/blogs/10-year-health-plan-we-need-to-move-from-techno-optimism-to-techno [accessed 2026-01-15]
13. Kamel Rahimi A, Pienaar O, Ghadimi M, et al. Implementing AI in hospitals to achieve a learning health system: systematic review of current enablers and barriers. J Med Internet Res 2024 Aug 2;26:e49655. [doi: 10.2196/49655] [Medline: 39094106]
14. Nair M, Nygren J, Nilsen P, et al. Critical activities for successful implementation and adoption of AI in healthcare: towards a process framework for healthcare organizations. Front Digit Health 2025;7:1550459. [doi: 10.3389/fdgth.2025.1550459] [Medline: 40453810]
15. Almagharbeh WT, Alfanash HA, Alnawafleh KA, et al. Application of artificial intelligence in nursing practice: a qualitative study of Jordanian nurses’ perspectives. BMC Nurs 2025 Jan 25;24(1):92. [doi: 10.1186/s12912-024-02658-6] [Medline: 39863852]
16. May C, Finch T. Implementing, Embedding, and Integrating Practices: An Outline of Normalization Process Theory: Sage Publications; 2009:535-554. [doi: 10.1177/0038038509103208]
17. May CR, Albers B, Bracher M, et al. Translational framework for implementation evaluation and research: a normalisation process theory coding manual for qualitative research and instrument development. Implement Sci 2022 Feb 22;17(1):19. [doi: 10.1186/s13012-022-01191-x] [Medline: 35193611]
18. Brady G, Bennin F, De Koning R, et al. Interventions used to reduce infectious aerosol concentrations in hospitals-a review. EClinicalMedicine 2025 Jan;79:102990. [doi: 10.1016/j.eclinm.2024.102990] [Medline: 39802303]
19. Clark SE, Mathur S, Barrado-Martín Y, et al. An umbrella review of the facilitators and barriers to implementing artificial intelligence (AI) solutions within hospital settings: through the lens of the NASSS framework (spread, scale-up and sustainability). medRxiv. Preprint posted online on Jun 20, 2025. [doi: 10.1101/2025.06.19.25329916]

XSL•FO
RenderX

20. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
21. Oliveira G. Developing a codebook for qualitative data analysis: insights from a study on learning transfer between university and the workplace. Int J Res Method Educ 2023 May 27;46(3):300-312. [doi: 10.1080/1743727X.2022.2128745]
22. Braun V, Clarke V. Is thematic analysis used well in health psychology? A critical review of published research, with recommendations for quality practice and reporting. Health Psychol Rev 2023 Dec;17(4):695-718. [doi: 10.1080/17437199.2022.2161594] [Medline: 36656762]
23. NVivo | the best qualitative data analysis software for 30 years. Lumivero. 2018. URL: https://lumivero.com/products/nvivo/ [accessed 2026-09-23]
24. Murray E, Treweek S, Pope C, et al. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. BMC Med 2010 Oct 20;8:63. [doi: 10.1186/1741-7015-8-63] [Medline: 20961442]
25. Vindrola-Padros C, Chisnall G, Polanco N, Vera San Juan N. Iterative cycles in qualitative research: introducing the RREAL sheet as an innovative process. SSRN Journal. Preprint posted online on Jul 14, 2022. [doi: 10.2139/ssrn.4162797]
26. Leenen JPL, Hiemstra P, Ten Hoeve MM, et al. Exploring the complex nature of implementation of artificial intelligence in clinical practice: an interview study with healthcare professionals, researchers and policy and governance experts. PLOS Digit Health 2025 May;4(5):e0000847. [doi: 10.1371/journal.pdig.0000847] [Medline: 40333664]
27. Health building note 00-09: infection control in the built environment. : Department of Health; 2013 URL: https://www.england.nhs.uk/wp-content/uploads/2021/05/HBN_00-09_infection_control.pdf [accessed 2026-01-15]
28. Emond Y, Wolff A, Bloo G, et al. Complexity and involvement as implementation challenges: results from a process analysis. BMC Health Serv Res 2021 Oct 23;21(1):1149. [doi: 10.1186/s12913-021-07090-z] [Medline: 34688287]
29. Futoma J, Simons M, Panch T, Doshi-Velez F, Celi LA. The myth of generalisability in clinical research and machine learning in health care. Lancet Digit Health 2020 Sep;2(9):e489-e492. [doi: 10.1016/S2589-7500(20)30186-2] [Medline: 32864600]
30. Yang J, Soltan AAS, Clifton DA. Machine learning generalizability across healthcare settings: insights from multi-site COVID-19 screening. NPJ Digit Med 2022 Jun 7;5(1):69. [doi: 10.1038/s41746-022-00614-9] [Medline: 35672368]
31. Zabala G, Pruitt ZM, Fairbanks RJ, Ratwani R. Assessing the readiness of health care organizations for safe AI integration: perspectives from quality and safety leaders. J Patient Saf 2026 Mar 1;22(2):168-172. [doi: 10.1097/PTS.0000000000001428] [Medline: 41324142]
32. Dunn P, Mays N, Alderwick H. Dazed and confused? policy ideas behind the 10-year health plan. The Health Foundation. 2025. URL: https://www.health.org.uk/reports-and-analysis/analysis/dazed-and-confused-policy-ideas-behind-the-10-year-health-plan [accessed 2026-01-15]
33. Love C, Street A, Riddell E, et al. Acceptability of air cleaning units on inpatient wards: help for infection control or hindrance for ward occupants? J Hosp Infect 2025 Jan;155:254-256. [doi: 10.1016/j.jhin.2024.09.010] [Medline: 39341284]
34. Brock RC, Goudie RJB, Peters C, et al. Efficacy of air cleaning units for preventing SARS-CoV-2 and other hospital-acquired infections on medicine for older people wards: a quasi-experimental controlled before-and-after study. J Hosp Infect 2025 Jan;155(1-8):1-8. [doi: 10.1016/j.jhin.2024.09.017] [Medline: 39374708]
35. NHS estates technical bulletin (NETB 2023/01A): application of HEPA filter devices for air cleaning in healthcare spaces: guidance and standards. : NHS England; 2023 URL: https://www.england.nhs.uk/long-read/application-of-hepa-filter-devices-for-air-cleaning-in-healthcare-spaces-guidance-and-standards/ [accessed 2026-01-15]

Abbreviations

AISaT: Artificial Intelligence Air Safety Tool
CFD: computational fluid dynamics
CFIR: Consolidated Framework for Implementation Research
COREQ: Consolidated criteria for Reporting Qualitative Research
HEPA: high-efficiency particulate air
HTM: Health Technical Memoranda
ICB: Integrated Care Board
IPC: infection prevention and control
NASSS: Non-Adoption, Abandonment, Scale-Up, Spread and Sustainability
NHS: National Health Service
NICE: National Institute for Health and Care Excellence
NIHR: National Institute for Health and Care Research
NPT: normalization process theory
PAG: patient advisory group
PPE: personal protective equipment
RREAL: Rapid Research Evaluation and Appraisal Lab

XSL•FO
RenderX

Edited by S Law; submitted 12.Feb.2026; peer-reviewed by D Chrimes, K Khashu, Z Li; revised version received 07.Sep.2026; accepted 10.Sep.2026; published 13.Oct.2026.

Please cite as:
Barrado-Martín Y, Mathur S, Khadjesari Z, Clark SE, Vindrola-Padros C, Lovat LB, Stevenson F
Anticipated Barriers and Facilitators to the Implementation of an AI Tool Advising on the Placement of Mitigation Strategies to Improve Air Quality in the Hospital Setting: Qualitative Study
JMIR Hum Factors 2026;13:e93420
URL: https://humanfactors.jmir.org/2026/1/e93420
doi:10.2196/93420

© Yolanda Barrado-Martín, Sanjoli Mathur, Zarnie Khadjesari, Sigrún Eyrúnardóttir Clark, Cecilia Vindrola-Padros, Laurence B Lovat, Fiona Stevenson. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 13.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Correction: Auxiliary Teaching and Student Evaluation Methods Based on Facial Expression Recognition in Medical Education

Xueling Zhu[1,2], MSE; Roben A Juanatas[1], PhD

[1]College of Computing and Information Technologies, National University, 551 Mariano Fortunato Jhocson Street, Sampaloc, Manila, Philippines
[2]College of Big Data and Artificial Intelligence, Anhui Xinhua University, Hefei, Anhui, China

Corresponding Author:
Xueling Zhu, MSE
College of Computing and Information Technologies, National University, 551 Mariano Fortunato Jhocson Street, Sampaloc, Manila, Philippines

Related Article:

Correction of: https://humanfactors.jmir.org/2025/1/e72838

Abstract

(JMIR Hum Factors 2026;13:e89669) doi:10.2196/89669

In “Auxiliary Teaching and Student Evaluation Methods Based on Facial Expression Recognition in Medical Education” [1], the authors noted one error in the author affiliation list.

A second affiliation has now been added for author XZ, as follows:

College of Big Data and Artificial Intelligence, Anhui Xinhua University, Hefei, Anhui, China

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Reference

1. Zhu X, Juanatas RA. Auxiliary teaching and student evaluation methods based on facial expression recognition in medical education. JMIR Hum Factors 2025 May 22;12:e72838. [doi: 10.2196/72838] [Medline: 40402552]

Submitted 16.Dec.2025; this is a non–peer-reviewed article;accepted 16.Dec.2025; published 14.Jan.2026.

Please cite as:
Zhu X, Juanatas RA
Correction: Auxiliary Teaching and Student Evaluation Methods Based on Facial Expression Recognition in Medical Education
JMIR Hum Factors 2026;13:e89669
URL: https://humanfactors.jmir.org/2026/1/e89669
doi:10.2196/89669

© Xueling Zhu, Roben A Juanatas. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 14.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Correction: Acceptability of a Digital Care App in Patients Undergoing Hip and Knee Arthroplasty: Prospective Cohort Study

Yacine Louni[1], MCDM; Matthew Laroche[1], BCom; Abdulrhman Alnasser[2], MD; Mohammad Abuhaneya[2], MD; Eric Belzile[3], MSc; Sandhya Baskaran[3], MEd, MBA; Jennifer Mutch[2], MDCM, FRCSC; Anthony Albers[2], MDCM, FRCSC

[1]Faculty of Medicine, McGill University, Montreal, QC, Canada

[2]Division of Orthopaedic Surgery, McGill University, St Mary's Hospital, 3830 Avenue Lacombe, Montreal, QC, Canada

[3]St Mary's Hospital Research Center, St Mary's Hospital, Montreal, QC, Canada

Corresponding Author:
Anthony Albers, MDCM, FRCSC
Division of Orthopaedic Surgery, McGill University, St Mary's Hospital, 3830 Avenue Lacombe, Montreal, QC, Canada

Related Article:

Correction of: https://humanfactors.jmir.org/2026/1/e79682

(JMIR Hum Factors 2026;13:e101097) doi:10.2196/101097

In "Acceptability of a Digital Care App in Patients Undergoing Hip and Knee Arthroplasty: Prospective Cohort Study" [1], the authors noted one error.

The credentials of author Sandhya Baskaran originally appeared as:

MSc

These are now corrected to:

MEd, MBA

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Reference

1. Louni Y, Laroche M, Alnasser A, et al. Acceptability of a digital care app in patients undergoing hip and knee arthroplasty: prospective cohort study. JMIR Hum Factors 2026 Jan 27;13:e79682. [doi: 10.2196/79682] [Medline: 41592299]

Submitted 11.May.2026; this is a non–peer-reviewed article;accepted 19.May.2026; published 14.Aug.2026.

Please cite as:
Louni Y, Laroche M, Alnasser A, Abuhaneya M, Belzile E, Baskaran S, Mutch J, Albers A
Correction: Acceptability of a Digital Care App in Patients Undergoing Hip and Knee Arthroplasty: Prospective Cohort Study
JMIR Hum Factors 2026;13:e101097
URL: https://humanfactors.jmir.org/2026/1/e101097
doi:10.2196/101097

© Yacine Louni, Matthew Laroche, Abdulrhman Alnasser, Mohammad Abuhaneya, Eric Belzile, Sandhya Baskaran, Jennifer Mutch, Anthony Albers. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 14.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Risk Factors for Noninitiation and Dropout in Blended Therapy in Inpatient Psychiatric Patients: Retrospective Cohort Study

Nikita Gupta[1], Dr med; Fabian Gardin[1], Dr med; Thomas Berger[2], Prof Dr; Wolfram Kawohl[1,3,4], Prof Dr Med

[1]Clienia AG, Oetwil am See, Switzerland

[2]University of Bern, Bern, Switzerland

[3]University of Zurich, Zurich, Switzerland

[4]University of Nicosia, Nicosia, Cyprus

Corresponding Author:
Nikita Gupta, Dr med
Clienia AG
Schlösslistrasse 8
Oetwil am See, 8618
Switzerland
Phone: 41 789106034
Email: gupta.n44@gmail.com

Related Article:

Correction of: https://humanfactors.jmir.org/2026/1/e85816/

(JMIR Hum Factors 2026;13:e111245) doi:10.2196/111245

In "Risk Factors for Noninitiation and Dropout in Blended Therapy in Inpatient Psychiatric Patients: Retrospective Cohort Study" [1], the group definitions underlying the dropout analysis were described incorrectly. The Methods section and Figure 1 state that the dropout comparison was defined as "Group B1: patients who completed 100% of all assigned modules" versus "Group B2: patients who completed <100% of all assigned modules." However, the logistic regression model reported in Table 3 was in fact conducted using a different definition: patients who completed none of their assigned modules (dropout) versus patients who completed at least one assigned module (nondropout).

All statistical analyses and results reported in the article (Tables 2 and 3, including all odds ratios, CIs, and *P* values) were calculated using this correct definition and remain valid and unchanged. The error is limited to the description of the group definitions in the Methods section, Figure 1, and corresponding wording in the Results and Discussion.

The interpretation of the dropout-related findings is refined by the correction: associations reflect the likelihood of completing no assigned modules at all, rather than the likelihood of incomplete module completion, as could be inferred from the original text. The noninitiation analysis is unaffected by this correction.

First, in the Division of Patient Groups section of the Methods, the following statement was revised:

> *Group B1: patients who completed 100% of all assigned modules*
>
> *Group B2: patients who completed <100% of all assigned modules*

The corrected text now reads:

> *Dropout: patients who completed none of their assigned modules*
>
> *Nondropout: patients who completed at least one of their assigned modules*

Second, in the same section, the following statement was revised:

> *...and dropout, the incomplete use of the assigned therapeutic modules.*

The corrected text now reads:

> *...and dropout, defined as the failure to complete any of the assigned therapeutic modules after platform initiation.*

Third, in Figure 1, the lower-level box labels were revised from the following:

> *Group B1: Patients who completed 100% of the assigned modules*
>
> *Group B2: Patients who completed <100% of the assigned modules*

The corrected figure now reads:

XSL•FO
RenderX

Dropout: Patients who completed none of the assigned modules

Nondropout: Patients who completed at least one of the assigned modules

Fourth, in the Distribution of Platform Use Parameters section of the Results, the following sentence was revised:

Notably, only 5 (1.8%) out of 278 patients completed all assigned modules (group B1).

The corrected text now reads:

Notably, only 5 (1.8%) out of 278 patients completed all assigned modules.

Fifth, in the same section, the following sentence was revised:

In terms of group distribution, 94 (33.8%) patients were in group A (noninitiation group), 184 (66.2%) patients initiated the platform, of whom 179 (64.4%) patients did not complete all assigned modules (group B2).

The corrected text now reads:

In terms of group distribution, 94 (33.8%) patients were in group A (noninitiation group), and 184 (66.2%) patients initiated the platform. Among those who initiated the platform, 94 (51.1%) completed none of their assigned modules (dropout), while 90 (48.9%) completed at least one assigned module (nondropout).

Finally, in the Discussion section, the following statement was revised:

...more willing to follow therapeutic guidance and complete assigned modules due to their conscientious traits.

The corrected text now reads:

...more willing to follow therapeutic guidance and continue engaging with assigned modules due to their conscientious traits.

The correction will appear in the online version of the paper on the JMIR Publications website together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Reference

1. Gupta N, Gardin F, Berger T, Kawohl W. Risk factors for noninitiation and dropout in blended therapy in inpatient psychiatric patients: retrospective cohort study. JMIR Hum Factors 2026 Jun 29;13:e85816 [FREE Full text] [doi: 10.2196/85816] [Medline: 42372259]

Submitted 04.Sep.2026; this is a non–peer-reviewed article;accepted 11.Sep.2026; published 22.Oct.2026.

Please cite as:
Gupta N, Gardin F, Berger T, Kawohl W
Correction: Risk Factors for Noninitiation and Dropout in Blended Therapy in Inpatient Psychiatric Patients: Retrospective Cohort Study
JMIR Hum Factors 2026;13:e111245
URL: https://humanfactors.jmir.org/2026/1/e111245
doi:10.2196/111245
PMID:

©Nikita Gupta, Fabian Gardin, Thomas Berger, Wolfram Kawohl. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 22.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Conversational AI in Hereditary Cancer Care: Sociotechnical Study of Responsible Design Requirements

Dominic Lammert[1], PhD; Jacqueline Lammert[2], Dr med; Justin Hofenbitzer[2], MA; Tristan Radtke[3], Dr iur, LLM; Yingping Sun[1,4], BSc; Martin Haimerl[5], Dr Ing; Stefanie Betz[6,7], Dr; Doran Holger Scholl[1], MSc; Jürgen Pfeffer[1], Dr

[1]Chair of Computational Social Science, TUM School of Social Science and Technology, Technical University of Munich, Hochschule für Politik München, Richard-Wagner-Straße 1, Munich, Bavaria, Germany

[2]Chair of Medical Informatics, Institute for AI and Informatics in Medicine, TUM School of Medicine and Health, TUM University Hospital, Technical University of Munich, Munich, Germany

[3]Chair of Law and Regulation of the Digital Transformation, Department of Governance, TUM School of Social Sciences and Technology, Technical University of Munich, Munich, Germany

[4]School of Business, Shandong Normal University, Jinan, Shandong, China

[5]Faculty II: Engineering and Technology, Furtwangen University, Furtwangen, Baden-Wurttemberg, Germany

[6]Faculty I:, Computer Science and Applications, Furtwangen University, Furtwangen, Baden-Wurttemberg, Germany

[7]School of Engineering Sciences, Software Engineering, Lappeenranta-Lahti University of Technology (LUT), Lappeenranta, Finland

Corresponding Author:
Dominic Lammert, PhD
Chair of Computational Social Science, TUM School of Social Science and Technology, Technical University of Munich, Hochschule für Politik München, Richard-Wagner-Straße 1, Munich, Bavaria, Germany

Abstract

Background: Individuals with *BRCA1/2* germline variants face complex, preference-sensitive medical decisions under psychological distress. Structural constraints in genetic counseling create support gaps during waiting periods, prompting patients to rely on fragmented or misleading online information. Conversational AI may offer low-threshold, continuous informational support, yet methods for its responsible integration into clinical care remain unclear.

Objective: This study identifies and systematizes stakeholder-informed sociotechnical design requirements and governance conditions for a clinically bounded conversational AI system intended to support comprehension, appraisal, and appropriate use of complex hereditary cancer information in *BRCA1/2*-related care.

Methods: We conducted an exploratory, qualitative, predevelopment requirements study using 2 structured participatory workshops with interdisciplinary stakeholders (N=18). Guided exercises included stakeholder analysis, the value proposition canvas, and the sustainability awareness framework. Structured workshop outputs, field notes, and artifacts were analyzed using inductive thematic analysis.

Results: Stakeholder workshops revealed that responsible design was shaped less by consensus around standalone responsible AI principles than by three recurrent cross-stakeholder tensions: (1) accessibility and continuity of support versus clinical role boundaries, (2) simplification and emotional reassurance versus medical precision and epistemic transparency, and (3) technical scalability versus governance, data protection, and institutional accountability. These tensions translated into design requirements for constrained system authority, mandatory human oversight, context-sensitive communication, and integration into clinical and regulatory governance structures.

Conclusions: Conversational AI in this context should be explicitly constrained and embedded within clinical governance structures to be perceived as responsible. Responsible implementation requires transparent design, human oversight, integration into existing care pathways, and early alignment with German/European Union (EU) regulatory requirements concerning data protection, medical device classification, genetic data, and AI transparency.

(JMIR Hum Factors 2026;13:e97869) doi:10.2196/97869

KEYWORDS

BRCA1/2; conversational AI; responsible AI; participatory design; genetic counseling; patient decision support; breast cancer gene 1 and 2

XSL•FO
RenderX

Introduction

Overview

Individuals carrying pathogenic germline variants in *BRCA1/2* face complex, preference-sensitive medical decisions regarding risk-reducing surgeries and surveillance, often under significant psychological distress with long-term implications for physical and mental health [1,2]. Despite the importance of genetic counseling, clinical pathways often lack continuous support beyond initial disclosure, while limited consultation time constrains the ability to address informational and psychosocial needs [3].

In Germany and internationally, structural bottlenecks and prolonged waiting times for specialized second opinions create a critical informational gap during the decision-making period. Patients frequently turn to online resources, where fragmented or misleading information may distort risk perception and increase uncertainty [4-6].

Conversational agents, particularly chat and voice bots, are increasingly discussed as low-threshold tools in health care. While advances in large language models (LLMs) have improved accessibility, challenges regarding reliability and ethical governance remain [7,8]. This study does not assume that patients lack access to conversational AI through general-purpose systems. Rather, it asks what added value a purpose-built, institutionally governed, and clinically bounded system could provide in *BRCA1/2*-related care. The envisioned conversational AI system is conceptualized as a supportive tool for waiting periods that complements, rather than replaces, human expertise.

Building on responsible AI principles, this study examines how ethical considerations can be translated into concrete design requirements for sensitive clinical contexts. While prior research has focused on the accuracy and usability of medical chatbots, empirical work on how responsibility is operationalized in emotionally sensitive, preference-sensitive decision contexts remains limited. This study therefore investigates how stakeholder perspectives can inform the responsible design and future integration of conversational AI into hereditary cancer care.

Using a qualitative sociotechnical approach, the study combines stakeholder analysis [9], the value proposition canvas (VPC) [10], and the sustainability awareness framework (SusAF) [11] to examine how digital health tools can be ethically integrated into the organizational and relational context of hereditary cancer care and how they shape responsibility, knowledge mediation, and patient autonomy. This study derives actionable design implications for the responsible integration of conversational AI into clinical care pathways, grounded in interdisciplinary stakeholder perspectives.

As an early-stage, stakeholder-informed design inquiry, this study did not include direct patient participation in the initial phase. Patient involvement is addressed as a methodological limitation and as a requirement for subsequent phases of system development.

This study contributes by translating abstract principles of responsible AI into concrete, stakeholder-informed design requirements and governance conditions for conversational AI in sensitive clinical contexts. In particular, it foregrounds legal and regulatory requirements not as external constraints to be addressed after design but as constitutive design conditions shaping system boundaries, data protection, medical device classification, genetic data governance, AI transparency, and institutional accountability in the German/European Union (EU) context.

Background

Clinical Context: *BRCA1/2* Mutation Carriers

Carriers of *BRCA1/2* germline variants face substantially increased cancer risks, requiring complex decisions on prophylactic surgery and surveillance, with implications for psychological well-being, family planning, and quality of life [1,2,12-14].

Standard genetic counseling, often limited to a pretest and disclosure session, frequently fails to meet patients' longitudinal needs. Psychological distress at the time of result disclosure impairs information processing, while limited consultation time constrains clinicians' ability to address psychosocial concerns in depth [3].

In Germany, specialized second-opinion services exist through the German Consortium for Familial Breast and Ovarian Cancer (FBREK), yet structural bottlenecks result in waiting times of up to 6 to 8 weeks. This gap contributes to patient overload and delayed decision-making, particularly for non-German-speaking individuals [15-17]. International evidence confirms similar deficits; for example, a recent Facing Our Risk of Cancer Empowered (FORCE) survey (n=1261) reported substantial gaps in posttest support, particularly regarding surgical menopause and family risk communication [4].

This "support vacuum" is often filled by online resources, where fragmented, inconsistent, or misleading information can distort risk perception and increase uncertainty [5,18-20]. In *BRCA1/2*-related hereditary cancer care, these risks are amplified by the complexity of decisions involving probabilistic cancer risk, preventive surgery, surveillance, fertility implications, family communication, and psychosocial consequences. Importantly, access to digital health information does not necessarily imply the ability to interpret, evaluate, or apply that information appropriately. Patients' perceived confidence in navigating online resources may diverge from their actual health literacy or eHealth literacy, particularly when information is technical, emotionally charged, or clinically preference-sensitive. Consequently, there is a substantial unmet need for accessible, evidence-based support that strengthens comprehension, appraisal, contextualization, and appropriate use of health information in preparation for clinical decision-making.

Chat and Voice Bots as Support for *BRCA1/2* Mutation Carriers

The evolution of conversational AI began with ELIZA [21] and PARRY [22], early systems rooted in mental health applications.

XSL•FO
RenderX

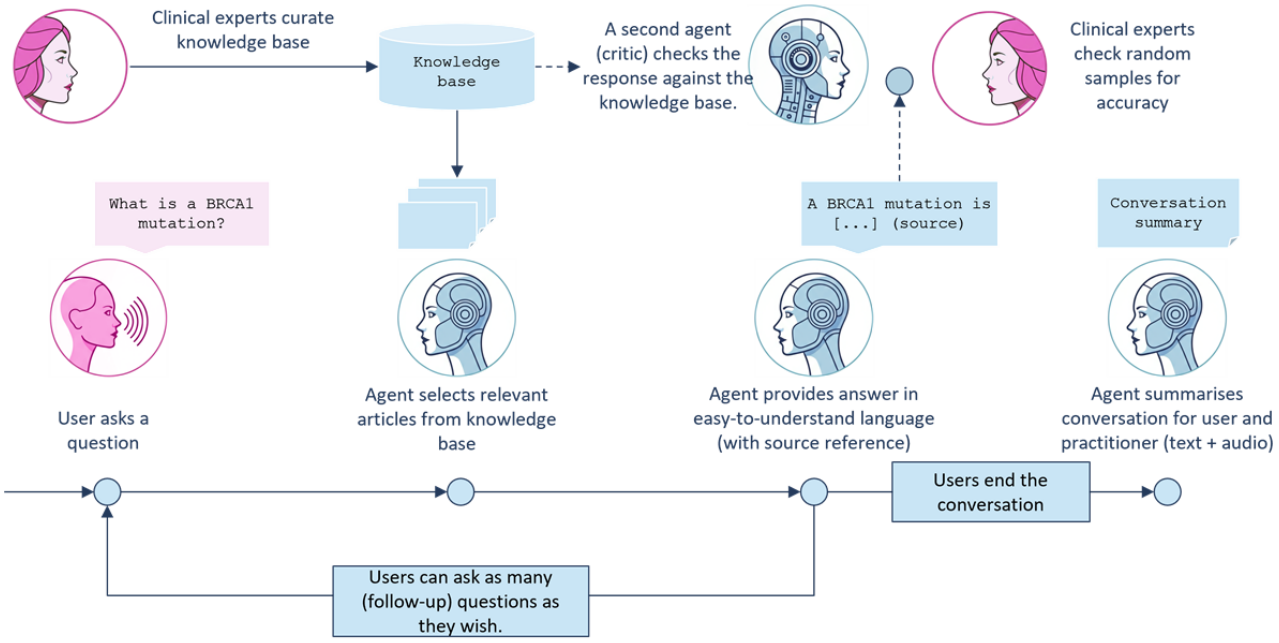
Today, medical chatbots are deployed across domains, though ethical and implementation challenges remain significant [7,8,23]. In oncology, they support screening, symptom tracking, and risk communication [24,25], with reported high user satisfaction, for example, in breast cancer education [26].

This study addresses the previously described "support vacuum" for *BRCA1/2* mutation carriers during waiting periods for clinical consultations. A conversational chat and voice bot is conceptualized as a supportive tool that provides evidence-based, interactive information to assist in preparing for clinical encounters.

At this stage, the chat and voice bot represents a conceptual system design rather than an implemented prototype. The envisioned architecture would combine an LLM with retrieval-augmented generation (RAG) to ground responses in a curated, evidence-based knowledge base. The proposed content development process would follow established standards such as the International Patient Decision Aid Standards (IPDAS) and would build on materials from the FBREK Consortium and the BRCA Network. The system is envisioned to support multilingual use and varying levels of health literacy.

In the proposed design, system integrity would depend on automated cross-referencing with the knowledge base, regular expert review, and clearly defined content governance processes. These mechanisms should be understood as design requirements derived from stakeholder input rather than as implemented or technically validated system features. Accordingly, Figure 1 illustrates the planned system architecture, not a deployed system.

Figure 1. Planned system architecture: a clinical knowledge base integrated via retrieval-augmented generation (RAG) to provide evidence-based support.



The rationale for such a purpose-built system is not that general-purpose LLMs are unavailable or incapable of conversational interaction. Rather, its added value lies in clinical and institutional governance. In contrast to general-purpose LLMs, the envisioned system would be restricted to a clearly defined *BRCA1/2*-related informational and preparatory role, grounded in a curated and expert-reviewed knowledge base, aligned with established patient decision-aid standards, and connected to escalation pathways for human support. The system is therefore conceptualized not as a more autonomous alternative to existing LLMs, but as a clinically bounded sociotechnical arrangement that limits unsupported interpretation, clarifies accountability, and supports preparation for clinical encounters.

A single example illustrates what this constraint adds. A 28-year-old carrier whose mother died of ovarian cancer at 42 years asks whether a positive BRCA1 result means she will certainly develop the disease. An unconstrained general-purpose system typically answers with cumulative lifetime incidence, approximately 44% by age 80 years [1], names risk-reducing salpingo-oophorectomy without temporal context, and closes with a generic instruction to consult a physician. Each element is accurate in isolation, yet together they equate genetic predisposition with imminence for a user in acute distress. A clinically bounded system would instead be configured to set that figure against the age-stratified incidence for her own age band, which remains an order of magnitude lower over the coming years [1]; to locate the surgical discussion in its guideline-anchored window from about 35 years of age for BRCA1 carriers, adjusted to the earliest diagnosis in her family [13,14]; and to route her to the responsible contact in her own care pathway. The difference lies not in model capability or conversational tone, but in the curated knowledge base, the response constraints, and the escalation rules for which the institution remains accountable.

Functionally, the envisioned system would provide structured summaries in text and audio formats to support clinical consultations and would include escalation pathways to human professionals for users in psychological distress.

XSL•FO
RenderX

From Ethical Principles to Design Practice: Responsible AI

Conversational agents have the potential to improve accessibility and user experience [27-29], but challenges remain regarding inclusivity, reliability, and ethical governance [30-34]. This study adopts a socioinformatics perspective, conceptualizing the system as embedded within social and organizational contexts where design choices shape care practices [35-37]. Responsible AI provides a framework to address these challenges, emphasizing transparency, fairness, and accountability, while highlighting the need to address issues such as explainability and bias [38-43].

In this study, responsible AI is understood as the design, development, and governance of AI systems in ways that are lawful, ethically aligned, technically and socially robust, transparent, accountable, privacy-preserving, and subject to meaningful human oversight. The EU High-Level Expert Group's Trustworthy AI guidance serves as the primary reference point for operationalizing this understanding, particularly through its emphasis on human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, nondiscrimination and fairness, societal and environmental well-being, and accountability [39].

These principles were used as sensitizing concepts rather than as a deductive coding scheme. Human oversight informed the analysis of role boundaries, escalation pathways, and the positioning of medical professionals as primary decision makers. Transparency and explainability guided attention to uncertainty communication, disclosure of system limitations, and prevention of automation bias. Privacy and data governance informed the analysis of genetic data sensitivity, consent, data minimization, and institutional accountability. Technical robustness and safety informed requirements for curated knowledge bases, expert review, and constrained system authority. Fairness and societal well-being informed attention to multilingual access, health literacy, eHealth literacy, accessibility, and the prevention of exclusion.

To operationalize these principles, the study draws on participatory design and value-sensitive design (VSD) to integrate stakeholder perspectives into system development [9,44,45]. Accordingly, responsible AI is treated not as a fixed checklist but as a sociotechnical design task involving the translation of general principles into concrete governance conditions, communication strategies, and role boundaries for a clinically bounded conversational AI system in *BRCA1/2*-related hereditary cancer care.

Research Gap and Research Question

Despite advances in digital health, limited evidence exists on how conversational AI can responsibly support *BRCA1/2* mutation carriers during waiting periods between genetic counseling, second opinions, and clinical decision-making. The underexplored issue is not simply whether conversational AI can provide information in genetic medicine, but how such systems can support the comprehension, appraisal, contextualization, and appropriate use of complex hereditary cancer information in emotionally charged and preference-sensitive contexts.

Existing research on medical chatbots has focused primarily on technical accuracy, usability, and user satisfaction, while less attention has been paid to how conversational AI can remain bounded by clinical governance, human oversight, validated content, escalation pathways, and responsible AI principles. In *BRCA1/2*-related care, this is particularly important because users must interpret probabilistic risk information and preventive options without mistaking informational support for clinical decision-making authority.

This study addresses this gap through a stakeholder-informed, sociotechnical perspective. It is guided by the following research question: how can stakeholder-informed design requirements for the responsible design and future integration of a conversational AI system in clinical care pathways be identified?

Methods

Overview

This study used a qualitative, practice-oriented design to identify stakeholder-informed design requirements for the responsible development of a conversational AI system in hereditary cancer care. The system served as an empirical case to examine sociotechnical and ethical dynamics. The study focused on deriving context-sensitive design requirements rather than evaluating system performance or user outcomes.

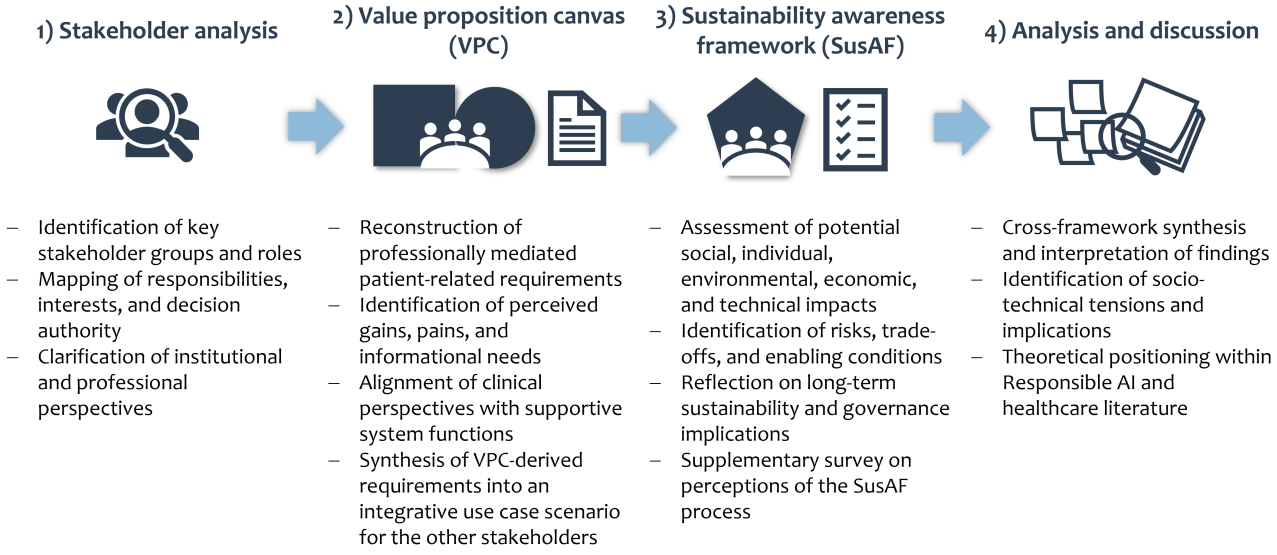
The conversational AI system examined in this study represents a conceptual design case rather than a fully implemented or deployed product. No prototype was implemented, no technical validation was performed, and no content pipeline was operationalized as part of this study.

The study followed a qualitative exploratory design informed by interpretive research principles and is reported in accordance with established standards, including the COREQ (Consolidated Criteria for Reporting Qualitative Research), to enhance transparency and rigor. In addition, relevant items from the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist were considered where applicable to enhance transparency in reporting digital health research.

The methodology integrated 3 complementary frameworks, including stakeholder analysis [9], the VPC [10], and the SusAF [11], which together structure data collection and analysis across sociotechnical dimensions (Figure 2).

XSL•FO
RenderX

Figure 2. Sequential analytical work steps applied in the study.



Stakeholder Analysis

The stakeholder analysis followed a 4-step process (identification, relevance assessment, interaction, and solution development) based on established participatory design approaches [9].

An interdisciplinary core team identified relevant stakeholders across clinical, technical, and social domains. For each group, key characteristics, including interests, perceived influence (scale 1 - 10), and conflict potential (scale 1 - 10), were assessed (Table 1).

XSL•FO
RenderX

Table . Stakeholder overview (abbreviated) with influence and conflict potential.

Stakeholder[a]	Known interests	Assumed interests	Perceived influence	Conflict potential	Present in the workshop
Patients	Clear, trustworthy info	Emotional relief; control	10	8	No
Oncologists	Preference-sensitive decisions	Time savings; clarity	10	8	Yes
Surgeons	Safe, aesthetic surgery	Manage expectations	9	7	Yes
Radiologists	Structured screening	Clarify false positive risk	9	7	Yes
Nursing staff	Coordination	Reduce cancellations	9	7	Yes
Patient advocacy	Participation; access	Strengthen autonomy	8	8	Yes
Psychologists	Psychosocial continuity	Early crisis detection	8	7	Yes
Ethicists	Ethical reflection	Influence guidelines	8	6	Yes
Hospital IT	IT security	Integration	7	5	Yes
UI[b]/UX[c] designers	User-centered design	Reduce cognitive load	7	3	Yes
Health insurers	Cost control; quality	Prevent misinformation	7	3	Yes
Disability groups	Accessibility	Avoid digital exclusion	6	4	Yes
Linguists	Cultural sensitivity	Prevent misunderstandings	6	4	Yes
Data protection experts	GDPR[d] compliance	Legal risk reduction	5	5	Yes
Media educators	Explain complexity	Improve transparency	5	3	Yes

[a]For each stakeholder group, perceived influence and conflict potential were assessed on a 1-10 scale through consensus-based discussion by the interdisciplinary core team. These scores represent qualitative assessments specific to the investigated use case rather than empirically measured quantities.

[b]UI:user interface.

[c]UX: user experience.

[d]GDPR: General Data Protection Regulation.

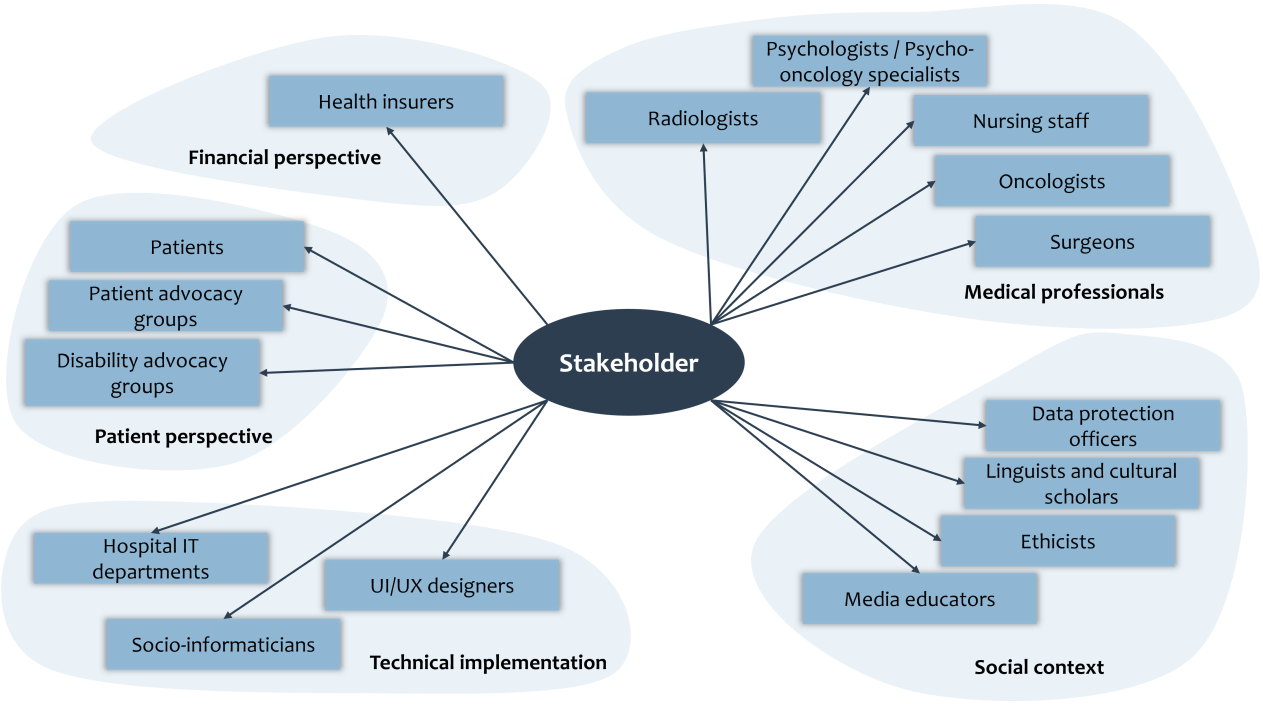
To ensure analytical focus, only stakeholders with an influence score of 5 or higher were included in participatory workshops, while additional stakeholders relevant for later implementation phases were documented but not actively involved. This threshold reflects a balance between inclusiveness and feasibility in early-stage participatory design.

While patients represent the primary end users, direct patient involvement was not included in this initial exploratory phase. This was an ethics-oriented precautionary decision within a phased predevelopment process, rather than an empirically demonstrated indication that patient involvement would have been harmful, inappropriate, or infeasible. To incorporate patient-related perspectives at this stage, psycho-oncology experts and patient advocacy organizations with experience in *BRCA1/2*-related care were included as professional stakeholders. Consequently, the findings reflect stakeholder-mediated interpretations of patient needs rather than direct patient accounts. Direct patient involvement is planned as a necessary next step to validate, refine, and potentially challenge the stakeholder-derived requirements once system boundaries, safeguards, and governance conditions have been further specified.

A stakeholder mapping was conducted to visualize relationships and cluster actors across medical, social, technical, financial, and patient-related domains (Figure 3).

XSL•FO
RenderX

Figure 3. Stakeholder mapping for the medical chat and voice bot.



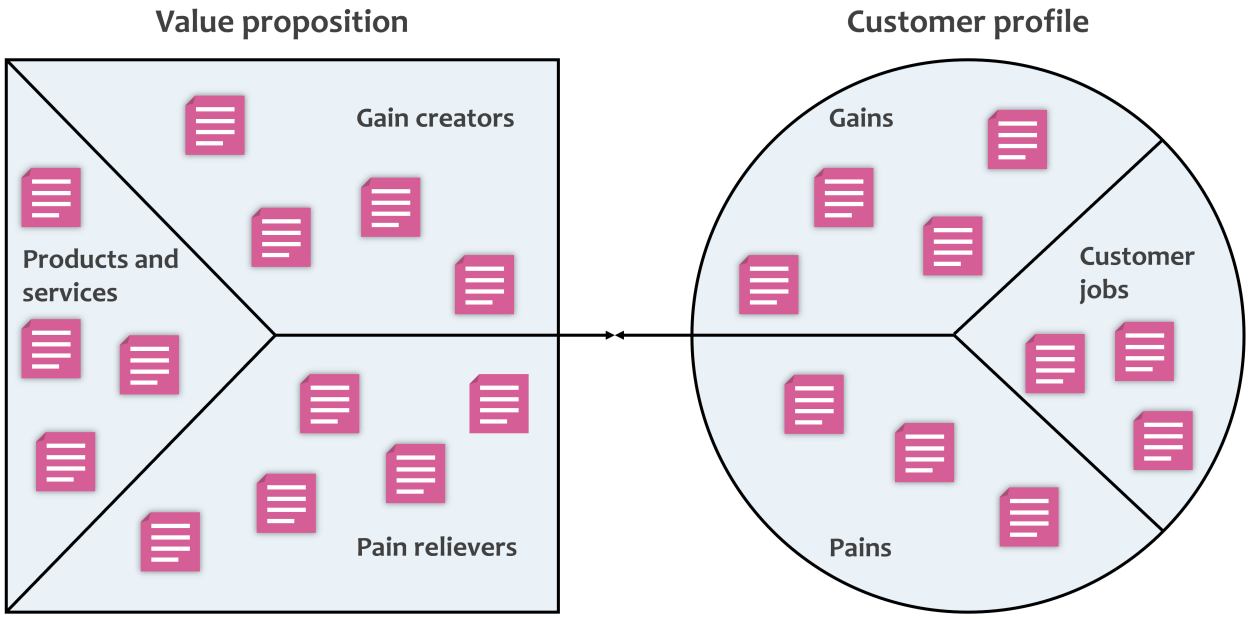
Participants were recruited using purposive sampling to ensure representation of key stakeholder groups relevant to hereditary cancer care and conversational AI governance, based on professional expertise and practical involvement.

Participatory interactions informed the identification of key tensions, synergies, and unmet needs. Data from workshop outputs, field notes, and artifacts were analyzed iteratively to derive design requirements.

VPC

The VPC [10] was used as a co-design instrument to align the needs of *BRCA1/2* mutation carriers with clinically appropriate and technically feasible system components (Figure 4). The workshop involved medical professionals from relevant disciplines, complemented by technical and socioinformatics expertise.

Figure 4. Value proposition canvas (VPC).



Within the customer segment, the workshop characterized informational needs, emotional burdens, and expectations regarding counseling. The value map translated these into system functionalities and communicative boundaries, positioning the chat and voice bot as a supportive tool rather than a decision-making authority.

The workshop resulted in a consolidated one-page user-case scenario, which served as a shared reference for subsequent

XSL•FO
RenderX

participatory sessions and supported alignment between clinical requirements and system design (see Multimedia Appendix 1).

SusAF

The SusAF was applied as a participatory instrument to examine the broader sociotechnical implications of the chat and voice bot. In contrast to the VPC workshop, the SusAF session involved all identified stakeholder groups (n=18) to capture a wide range of perspectives.

The framework structures analysis across 5 sustainability dimensions and 3 orders of effects, enabling the identification of both immediate and long-term impacts [11] (Tables 2 and 3).

Participants identified and discussed potential positive, negative, and ambivalent effects, which were subsequently prioritized using an impact matrix based on likelihood and significance (Figure 5). This process enabled the identification of key factors that informed design requirements.

Results were further structured using the sustainability awareness diagram (SusAD) to map impacts across dimensions and effect levels, supporting interdisciplinary reflection and making implicit risks explicit (Figure 6).

A final stakeholder survey confirmed the usefulness of the framework and the relevance of the identified implications for system design.

Table . Sustainability dimensions of the SusAF[a].

Dimension	Description
Social	Concerns relationships between individuals and groups, including trust, communication, participation, inclusion, and the balance of potentially conflicting interests
Individual	Addresses individual well-being, autonomy, dignity, privacy, safety, and the capacity to make informed decisions and exercise agency
Environmental	Focuses on the stewardship of natural resources, including energy consumption, material use, and ecological impacts across the system life cycle
Economic	Encompasses financial and organizational aspects such as value creation, cost structures, governance, and long-term economic viability
Technical	Relates to the longevity and evolution of the system, including maintainability, adaptability, security, scalability, and resilience over time

[a]SusAF: sustainability awareness framework.

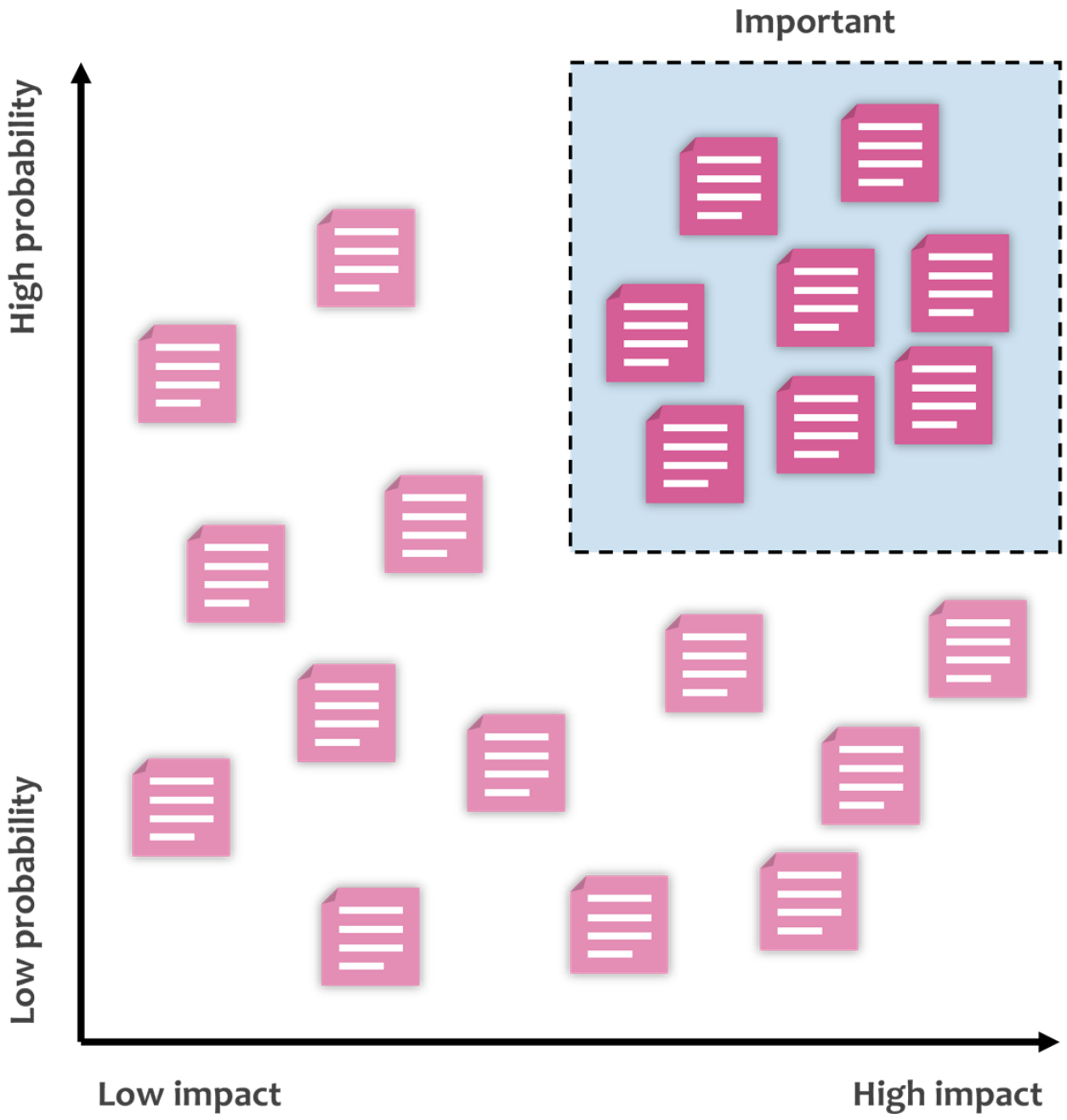
Table . Orders of sustainability effects in the SusAF[a].

Order of effect	Description
Immediate effects	Direct impacts from development, deployment, use, and disposal, including technical features and life-cycle energy consumption
Enabling effects	Indirect impacts emerging over time that induce changes in user behavior, clinical practices, or organizational routines
Structural effects	Long-term, systemic changes at the societal level influencing norms, policies, institutions, or health care infrastructures

[a]SusAF: sustainability awareness framework.

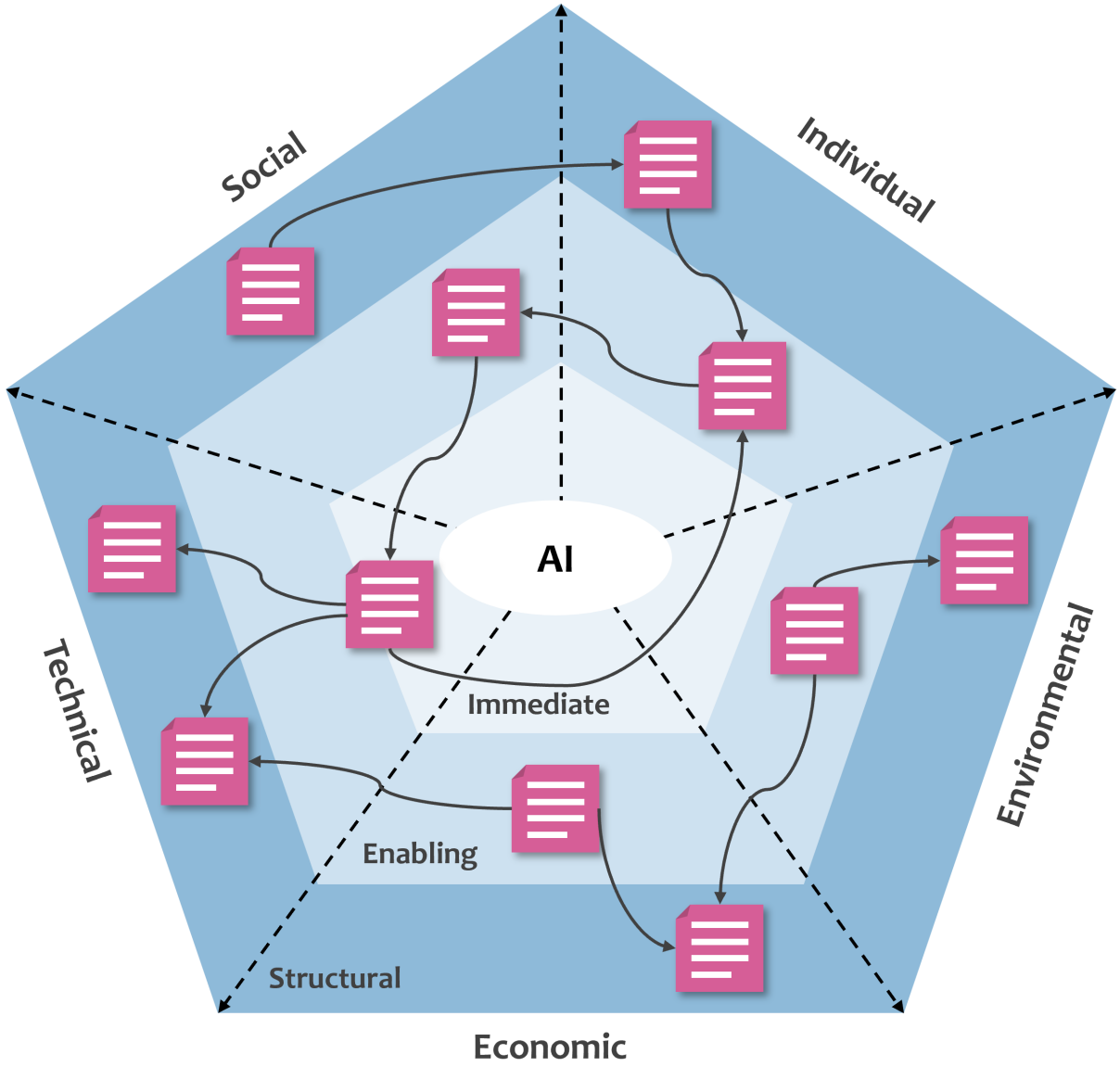
XSL•FO
RenderX

Figure 5. The weighting of impacts within the sustainability awareness framework (SusAF).



XSL•FO
RenderX

Figure 6. Sustainability awareness diagram (SusAD).



Workshop Procedure, Data Collection, and Data Analysis

A total of 2 participatory workshops were conducted between July and September 2025, each lasting approximately 4 hours. Workshop 1 focused on stakeholder analysis and the VPC and included 8 participants from clinical and technical domains. Workshop 2 applied the SusAF and involved all 18 stakeholders. Each workshop followed a structured sequence consisting of (1) introduction and contextual framing, (2) guided individual reflection, (3) small-group work in interdisciplinary subgroups, and (4) plenary discussion and consolidation of results. Workshops were moderated by 2 facilitators with backgrounds in socioinformatics and qualitative research, following a predefined facilitation guide to ensure consistency.

The workshops followed a semistructured format combining individual reflection, collaborative exercises, and framework-based activities. Participants worked both individually and in interdisciplinary groups to generate and reflect on design-relevant insights.

Across both workshops, documented materials (structured outputs, field notes, and artifacts) comprised approximately 45 pages of textual data. No audio recordings were made; discussions were documented in real time and consolidated immediately after each session. The use of multiple data sources enabled triangulation and a comprehensive understanding of stakeholder perspectives.

Data were analyzed using an inductive, codebook-supported form of thematic analysis. The analysis followed a staged procedure consisting of (1) familiarization with structured workshop outputs, field notes, and artifacts, (2) open coding of design-relevant needs, risks, tensions, and safeguards, (3) grouping of codes into preliminary categories, (4) comparison of categories across stakeholder analysis, VPC, and SusAF outputs, and (5) iterative refinement into higher-level design requirements.

XSL•FO
RenderX

The stakeholder analysis, VPC, and SusAF frameworks structured data collection and served as sensitizing lenses during interpretation, but they did not predetermine the final themes. Codes were first generated inductively from the documented workshop materials. They were then compared across the 3 frameworks to identify recurring patterns, contradictions, and cross-stakeholder tensions. For example, codes related to "role boundaries," "medical accountability," "automation bias," and "escalation to human professionals" were consolidated into the higher-level requirement of constrained system authority and mandatory human oversight. Similarly, codes concerning "information overload," "adaptive explanation," "emotional reassurance," and "uncertainty communication" informed the requirement of context-sensitive communication.

Coding was conducted independently by 2 researchers, DL and YS, on the complete set of documented materials. Differences in coding and category assignment were discussed until consensus was reached. Categories were retained as higher-level design requirements only when they were supported by multiple data sources or when contested perspectives could be reconstructed as meaningful stakeholder tensions. This procedure enabled triangulation across data sources, including field notes, structured workshop artifacts, and framework-based outputs, as well as across analytical lenses, including stakeholder analysis, VPC, and SusAF.

Workshops were not audio-recorded due to the sensitive clinical context and to foster an open discussion atmosphere. To mitigate potential loss of nuance and selective documentation, field notes were taken by multiple facilitators in parallel, compared after each workshop, and expanded immediately after each session while the discussion context was still fresh. Structured workshop outputs, including written artifacts, prioritization matrices, stakeholder maps, and framework-based templates, were treated as primary qualitative data rather than as supplementary documentation. Field notes were used to contextualize these artifacts, reconstruct stakeholder reasoning, and identify tensions that were not fully captured in the structured templates.

The research team's dual role as system designers and analysts was addressed through reflexive practices. Researchers documented assumptions and potential biases in analytic memos and revisited them throughout the coding process. Particular attention was given to disconfirming cases and tensions between stakeholder groups. Interpretations were retained only when supported by multiple data sources or when contested perspectives could be reconstructed.

The findings therefore describe stakeholder-informed design and governance requirements rather than evidence of system performance, safety, usability, or clinical effectiveness. Future work should therefore develop and evaluate a prototype, validate the reliability of the RAG-based architecture, and specify the content governance pipeline in detail.

To enhance transparency, illustrative excerpts from field notes and artifacts are reported in the Results section. These are paraphrased for readability and attributed to stakeholder groups to maintain confidentiality.

As the system is in an early conceptual stage, no real-world use data (eg, system logs, user interaction metrics, or adherence data) were collected. Instead, insights are based on structured stakeholder interactions and qualitative analysis.

Ethical Considerations

This study did not involve direct patient participation or the collection of patient-level health data. The study involved professional stakeholders who contributed expert perspectives in participatory design workshops. All participants were recruited based on their professional expertise and provided informed consent prior to participation.

Formal ethical approval was not required for this expert consultation and qualitative design study in accordance with institutional guidelines on good scientific practice at the Technical University of Munich. Nevertheless, ethical safeguards were applied throughout the research process, including voluntary participation, informed consent, confidentiality, anonymized documentation of workshop outputs, responsible data handling, and anonymized reporting of stakeholder perspectives.

No audio recordings were made. Workshop data were documented through structured artifacts, field notes, and summarized outputs without attributing statements to identifiable individuals. All study materials were stored on access-restricted institutional systems and were accessible only to the research team.

Results

The following sections present stakeholder-derived design requirements and sociotechnical tensions identified through the participatory co-design process using the VPC and SusAF.

VPC: Aligning Patient Needs With System Capabilities

The VPC workshop synthesized stakeholder-informed interpretations of the needs of *BRCA1/2* mutation carriers, structured into customer profiles (gains, pains, and customer jobs) and corresponding system responses.

The Mutation Carrier Profile: Gains, Pains, and Customer Jobs

Stakeholders described 2 central phases: understanding genetic risk and evaluating preventive options in relation to life plans.

Gains

Stakeholders emphasized the need for continuous access to reliable, evidence-based information to support decision-making. Key gains include structured action options, individualized risk information, and preparation for clinical consultations. Emotional needs include empathic communication, continuity of care, and reassurance, as well as multilingual and accessible information.

Pains

Fear emerged as a dominant factor, including anxiety related to diagnosis, surveillance, fertility, and family implications. This is compounded by information overload, time pressure, and uncertainty about preventive options. Additional challenges

XSL•FO
RenderX

include limited psychosocial support, lack of trustworthy information, language barriers, and financial concerns.

Customer Jobs

Mutation carriers aim to regain orientation and control, understand individual risk, and make informed decisions regarding prevention, surveillance, and family planning. Stakeholders highlighted the importance of trust in clinical care and access to understandable information, alongside initial skepticism toward digital tools.

Stakeholders emphasized informational fragmentation during waiting periods, as illustrated by one participant, "patients receive pieces of information from different disciplines, but no coherent picture emerges." Emotional burden was closely linked to informational uncertainty, with one expert noting that patients fear "making the wrong decision under pressure." At the same time, clinical stakeholders stressed decision quality and guideline adherence, warning that "simplifying information too much risks misinterpretation." Stakeholder perspectives differed in how needs were prioritized. Clinical stakeholders emphasized decision quality and guideline alignment, while psychosocial and advocacy stakeholders focused on emotional burden and continuity of support. Technical stakeholders highlighted feasibility, scalability, and interaction design.

The System's Value Proposition: Mechanisms for Support

The chat and voice bot was conceptualized as a supportive tool that complements clinical care.

Gain Creation

Adaptive communication tailors information to individual needs and supports preparation for clinical consultations, enhancing perceived autonomy.

Pain Relievers

The system addresses emotional distress through validation and referral to human support while maintaining clear role boundaries and providing structured, evidence-based information.

Trust and Onboarding

Stakeholders emphasized that trust depends on transparent communication, clinical integration, and clear positioning as a non-authoritative informational tool.

Information Support

The system provides continuous access to curated information on prevention, fertility, and psychosocial aspects in multilingual and accessible formats.

Boundary Management and Governance

The system incorporates safeguards against overreliance, including explicit uncertainty communication, encouragement of clinical verification, and expert-reviewed quality assurance. The system is constrained to an informational and preparatory role and includes default escalation pathways to human professionals.

Across stakeholder groups, three requirements were consistently prioritized: (1) clear communication of system limitations, (2) integration into clinical care pathways, and (3) mechanisms for human oversight and escalation. Other features were considered secondary, indicating that trust-building and governance are foundational design conditions.

User Case Scenario as an Integrative Result

A narrative user case scenario was developed to represent a mutation carrier's trajectory from diagnosis to decision-making. It translates abstract requirements into a clinical context, supports alignment on system boundaries (eg, exclusion of longitudinal monitoring and imaging diagnostics), and informs subsequent sustainability and ethical evaluations (see Multimedia Appendix 1).

Cross-Stakeholder Tensions

The analysis revealed several tensions between stakeholder groups, indicating that system design requires balancing competing priorities rather than addressing isolated needs. The most distinctive contribution of the stakeholder workshops was not the identification of responsible AI principles as such, but the reconstruction of how these principles were negotiated, prioritized, and translated into design requirements in the specific context of *BRCA1/2*-related hereditary cancer care. The 3 core requirements should therefore be understood as outcomes of cross-stakeholder tensions rather than as standalone principles.

A central tension concerned the level of system autonomy. Clinical stakeholders emphasized strict role boundaries and the positioning of the system as a nonauthoritative informational tool, whereas advocacy and psychosocial stakeholders highlighted the need for continuity and emotional support, which may increase reliance on the system.

A second tension involved the depth and complexity of information. Medical professionals favored precise and detailed communication to avoid misinterpretation, while communication-focused stakeholders emphasized simplification and progressive disclosure to prevent cognitive overload.

A third tension emerged between technical scalability and ethical safeguards. Technical stakeholders emphasized efficiency and automation, whereas ethicists and data protection experts prioritized transparency, explainability, and strict governance mechanisms.

These tensions were reflected in contrasting stakeholder statements. For example, clinical stakeholders emphasized strict system boundaries ("the system must not be perceived as giving medical advice"), whereas advocacy-oriented participants stressed the need for continuity ("patients need something that feels like ongoing support, not just information"). Similarly, clinicians highlighted the importance of precision ("details matter to avoid wrong conclusions"), while communication-focused stakeholders emphasized simplification ("too much detail overwhelms patients in this situation").

These contrasting perspectives indicate that design requirements emerge from negotiated trade-offs rather than uniform stakeholder agreement. These trade-offs informed the subsequent

XSL•FO
RenderX

interpretation of onboarding, communication design, quality management, and clinical integration as interdependent governance mechanisms rather than isolated design features.

Results From the SusAF

Stakeholders identified sustainability-related impacts across all 5 SusAF dimensions. Rather than treating these dimensions as separate domains, the analysis showed that sustainability in this context depends on whether the system can support *BRCA1/2* mutation carriers during a clinically and emotionally sensitive waiting period without displacing human care or amplifying misinterpretation.

Context-Specific Social and Individual Implications

Across the social and individual dimensions, stakeholders emphasized that 24-hour availability, multilingual access, and the possibility to ask sensitive questions in a low-threshold format could support orientation, self-efficacy, and preparation for clinical consultations. These benefits were considered particularly relevant for topics such as hereditary risk, fertility, preventive surgery, family communication, and psychosocial burden. At the same time, participants warned that decontextualized genetic information could increase anxiety, distort risk perception, or contribute to stigmatization and discrimination. The main design implication was therefore not simply to increase access to information but to provide adaptive, progressively disclosed, and clearly bounded information with transparent escalation pathways to human professionals.

Technical, Data-Related, and Governance Implications

The technical dimension was closely linked to these psychosocial and clinical concerns. Stakeholders viewed a curated knowledge base and RAG as promising mechanisms for source-based information provision, but only if combined with expert review, continuous content maintenance, robust data protection, and explicit communication of uncertainty. Automation bias was identified as a key risk, particularly because users may interpret chatbot responses as medical authority in moments of uncertainty or distress. Technical sustainability therefore depends on maintaining a clear distinction between informational preparation and clinical decision-making.

Environmental and Economic Conditions

Environmental and economic effects were discussed as conditional rather than inherent benefits of digital support. Remote access may reduce unnecessary travel or preparatory consultations, and structured information may help clinical encounters become more focused. However, these potential benefits depend on energy-efficient infrastructure, long-term maintenance capacity, sustainable financing, and regulatory compliance. Stakeholders therefore emphasized that digitalization should not be assumed to be sustainable by default; its value depends on governance, clinical integration, and continued institutional responsibility.

The SusAF analysis showed that the most relevant sustainability concerns in this use case are not generic efficiency gains but the interaction between access, psychological safety, genetic data sensitivity, clinical accountability, and long-term governance. Without clear boundaries, transparency, and integration into clinical workflows, features intended to empower users may also amplify risks such as misinterpretation, overreliance, exclusion, or inequity, underlining the dual-use character of conversational AI in health care contexts.

Stakeholder Reflections on the Process

The postworkshop evaluation (n=14/18) indicates a broadly positive reception of the SusAF process. Ten respondents rated its usefulness as high to very high, and none reported low relevance. Eight participants considered the knowledge gained to justify the time investment, while 6 participants assessed the effort-benefit ratio as moderate.

Participation influenced perspectives, with 8 respondents reporting moderate to strong changes in their understanding of medical chatbot requirements. These shifts reflected increased awareness of heterogeneous user needs and structural constraints of clinical practice, including limited time and the need for interdisciplinary coordination. Interdisciplinary exchange was identified as a key success factor, with 13 participants rating it as highly helpful.

Regarding implementation, stakeholders emphasized that the system should function as a supportive, preparatory tool rather than a standalone solution. Key requirements include transparent communication of system limitations, robust informed consent, and clear positioning of medical professionals as primary points of contact. Participants also highlighted the importance of cultural sensitivity through multilingual and accessible communication.

Discussion

The findings highlight how responsible conversational AI in hereditary cancer care depends on constraining system authority, embedding human oversight, and integrating systems into clinical care pathways.

Onboarding Process and Informed Consent

The stakeholders understood onboarding as a central sociotechnical mechanism for shaping trust and role perception within the care pathway. In the context of hereditary cancer, initial digital interactions appear closely linked to processes of trust formation and emotional safety. A human-mediated onboarding process emerges as particularly important for positioning the chat and voice bot as a supportive tool rather than a replacement for clinical expertise. Introducing the system through medical professionals may therefore contribute to reduced anxiety and clearer role expectations within the care pathway.

From an ethical perspective, informed consent can be understood as a multilayered and ongoing communicative process rather than a one-time procedural step. This includes explicitly communicating system limitations, potential inaccuracies, and the need for human verification. Such transparency is critical for mitigating automation bias and safeguarding individual autonomy, while also reducing cognitive overload and reinforcing clinical responsibility as part of sustainable system use.

XSL•FO
RenderX

The findings further indicated that onboarding should be conceptualized as adaptive and revisitable over time, particularly as patients transition between different phases of decision-making. As individuals move from initial risk assessment to surgical considerations, informational needs and emotional states evolve, requiring corresponding adjustments in how consent and system use are framed.

Linguistic Abilities and Communication Strategies

The findings indicated that linguistic configuration represented a central component of the sociotechnical integration of the chat and voice bot. It could be conceptualized along 3 interrelated dimensions, including sociolinguistic awareness, inclusivity, and strategic communication. The ability to balance an appropriate medical register with adaptive language complexity emerged as particularly critical. This includes adjusting terminology and depth of explanation to individual patient backgrounds, ranging from barrier-free communication for less experienced users to more technical discourse when appropriate. Such adaptability appeared essential for fostering resonance and preventing alienation.

Linguistic inclusivity further extended beyond translation toward a multilingual and multimodal communication architecture. Enabling interaction in a user's preferred language may strengthen accessibility and perceived conversational intimacy. At the same time, bridging the digital–physical divide through multimodal outputs, such as combined voice, text, and structured materials for clinical preparation, appeared to support continuity of care and integration into clinical workflows.

With regard to communication strategies, the findings pointed to the importance of balancing factual clarity with empathic responsiveness. While concise and action-oriented responses are necessary to avoid ambiguity and prevent misinterpretation, acknowledging emotional states remains central to user-centered interaction. In this context, epistemic transparency—including explicit communication of system limitations—emerged as a key mechanism for aligning professional credibility with responsible use.

Taken together, these observations suggest that effective communication design requires the integration of sociolinguistic sensitivity and factual precision. This integration enables the system to address users as individuals rather than as abstract clinical cases, thereby supporting both informational and emotional needs.

Finally, communication must be understood as embedded within broader sociocultural contexts. This includes sensitivity to topics such as fertility, body image, and genetic responsibility, as well as to variations in health literacy, perceptions of medical authority, and attitudes toward preventive interventions. Cultural competence, therefore, extends beyond multilingual support and represents a core condition for responsible system design.

Usability and User Experience

In the workshops, usability and user experience (UX) extended beyond interface design and constituted central conditions shaping the perceived clinical fit of the system. In the context of hereditary cancer risk, UX can be understood as a form of emotional scaffolding, particularly for users experiencing cognitive and emotional overload.

The SusAF analysis indicates that perceived autonomy is closely linked to interaction design. Features such as progressive disclosure and modular content enable users to regulate the depth and pace of information, thereby supporting agency and reducing the risk of disempowerment. In this sense, usability also appears to function as a proxy for institutional trust, as transparent interaction design and explicit cues for escalation to human professionals reinforce confidence in the system's appropriate role.

At the same time, usability must be responsive to heterogeneous user groups. This includes ensuring that voice-based interaction and accessibility features reduce barriers without oversimplifying medically complex information. Technical robustness and error transparency are equally critical, as usability-related failures—such as latency or unmasked hallucinations—carry ethical implications by potentially fostering misplaced trust.

Quality Management

Across the participatory workshops, stakeholders emphasized that quality management for medical conversational agents should be embedded within existing structures of clinical governance rather than treated as a purely technical concern. The chat and voice bot can thus be understood as a preparatory and informational tool whose legitimacy depends on clearly defined institutional responsibilities and explicit role boundaries between automated support and human medical expertise.

Accordingly, quality assurance appears to rely on clinically curated knowledge bases, expert validation processes, and established medical standards, rather than being delegated to the system itself. In this context, quality management emerges as a continuous and iterative process requiring ongoing expert review and system refinement.

With regard to accountability, responsibility was consistently located within the clinical organization rather than attributed to the system. Quality management therefore functions as a mechanism for ensuring traceability, transparency, and professional accountability, supported by clear escalation pathways to human professionals.

This governance perspective also clarifies the added value of a purpose-built system compared with general-purpose LLMs. While general-purpose systems may provide fluent and accessible responses, their use alone does not establish clinical accountability, validated content governance, institutionally defined role boundaries, or reliable escalation pathways. For *BRCA1/2* mutation carriers, these governance conditions were not peripheral safeguards but central design requirements, because users may otherwise mistake conversational fluency for medical authority in moments of uncertainty or distress, as illustrated by the risk-framing and triage example outlined above.

XSL•FO
RenderX

Environmental and Economic Trade-Offs of Digital Support

Environmental and economic considerations emerged as secondary but relevant governance concerns. In contrast to the social, individual, technical, and legal issues, these effects were less specific to hereditary cancer care, but they remain important for assessing whether conversational AI can be responsibly sustained in clinical practice.

From an environmental perspective, potential benefits arise primarily through the reduction of physical travel, as remote access to specialized information may decrease unnecessary in-person consultations and associated emissions. However, these benefits are highly dependent on system design. Energy-intensive processes, including model training, inference, and supporting infrastructure, may offset such gains if not carefully managed. Environmental outcomes thus emerge as indirect consequences of architectural and governance decisions, requiring energy-efficient infrastructures and minimized computational overhead.

Similarly, economic implications reflect a tension between potential efficiency gains and structural costs. While shifting preparatory informational tasks to the system may relieve clinical staff and enable more focused patient encounters, the development, certification, maintenance, and regulatory compliance of medical-grade AI systems represent substantial and ongoing financial commitments.

Taken together, sustainability cannot be assumed as an automatic outcome of automation. Instead, both environmental and economic viability depend on governance conditions, including transparent financing structures, institutional responsibility for long-term maintenance, and alignment with health care reimbursement systems.

Data Protection, Privacy, and Legal Requirements

The findings indicated that legal and regulatory compliance is not merely a boundary condition but a central determinant of trust, system acceptance, and responsible integration into clinical care. Data protection requirements, in particular, can be understood as core design conditions that shape system architecture, communication strategies, and governance structures.

The deployment of the chat and voice bot falls within the scope of the General Data Protection Regulation (GDPR; EU) 2016/679, as it involves the processing of personal data (Article 4(1) GDPR; cf. Court of Justice of the European Union [CJEU], Endemol Shine, C-740/22), including chat logs and genetic risk information. The institution determining the purposes and means of processing acts as the controller (Article 4(7) GDPR).

From a design perspective, data minimization and user control emerge as central requirements. An anonymization-oriented approach (Recital 26(5) GDPR) may be implemented through pseudonymized access codes. Where datasets cannot be linked to identifiable individuals by any party, the information is considered anonymized and thus falls outside the scope of the GDPR (cf. CJEU, Single Resolution Board [SRB], C-413/23 P).

Compliance with core principles such as lawfulness, fairness, and transparency (Article 5(1)(a) GDPR) must be operationalized in ways that remain accessible to users. While scientific research may benefit from specific provisions (Article 89 GDPR), deployment in clinical practice requires a valid legal basis, such as a medical contract (Article 6(1)(b) GDPR) or consent ([46], para 48). For special categories of personal data, including genetic and broadly defined health data (cf. CJEU, C-21/23), explicit consent (Article 9(2)(a) GDPR) or Article 9(2)(h) GDPR in conjunction with Section 22 of the German Bundesdatenschutzgesetz (Federal Data Protection Act; BDSG) applies, provided appropriate safeguards are ensured.

At the same time, the findings highlight that extensive legal information may itself contribute to cognitive overload, particularly in emotionally vulnerable situations. This suggests the need for layered privacy communication, combining concise summaries with access to more detailed information, in line with transparency requirements under Article 13 and 14 GDPR [47].

Concerns regarding automated decision-making further underline the importance of clearly defined system boundaries. Although the system was designed to provide informational support without producing legally binding or directly consequential decisions, maintaining this distinction is critical to avoid regulatory risks (Article 22 GDPR; see also CJEU, Schufa, C-634/21) and unintended overreliance.

Beyond data protection, stakeholders pointed to the relevance of the EU AI Act (Regulation [EU] 2024/1689), particularly regarding transparency obligations (Article 50(1)). Clear disclosure of AI interaction was therefore considered essential to support informed use and prevent misinterpretation.

The classification of the system under the Medical Device Regulation (MDR; Regulation [EU] 2017/745) remains context-dependent [48-50]. Maintaining the system's role as an informational and preparatory tool, rather than engaging in monitoring or treatment (Article 2(1) MDR) [51,52] appears critical for avoiding more stringent regulatory requirements. Current legislative developments (the European Commission's proposed COM [2025] 1023 final) [53] may affect the interaction between the MDR and the AI Act.

At the national level, the German Genetic Diagnosis Act (GenDG) requires explicit consent for processing genetic data (Section 11 GenDG). These requirements are closely linked to broader questions of liability and professional responsibility [54]. Rather than attributing responsibility to the system, accountability remains anchored in clinical institutions. In this context, technical safeguards, such as constrained system prompts and conservative response strategies, function as mechanisms to reinforce the system's supportive role and limit potential liability (Sections 280, 823 German Civil Code–BGB).

Overall, legal and regulatory requirements emerged as integral components of system design. Rather than being addressed retrospectively, they shape core design decisions, including system boundaries, communication strategies, and the positioning of human professionals as primary decision makers.

XSL•FO
RenderX

Compared with the informal use of general-purpose LLMs, an institutionally governed system would also allow clearer allocation of controller responsibilities, data minimization strategies, consent procedures, and accountability structures within the clinical setting.

Summary: Reframing Responsibility in Conversational AI for Health Care

Taken together, this sociotechnical analysis suggests that responsibility in conversational AI is not an intrinsic system property but a relational and organizationally situated outcome. Rather than arising from technical accuracy or compliance alone, it emerges from concrete design decisions, including limiting system authority, embedding escalation pathways, and ensuring transparency about uncertainty.

These findings point to a shift in emphasis from optimizing autonomous systems toward designing sociotechnical arrangements that preserve relational care and professional responsibility. More broadly, they indicate that responsibility in medical AI is less a function of model performance than of how system capabilities are deliberately positioned, limited, and governed within clinical contexts.

This study contributes stakeholder-informed insights into how principles of responsible AI can be translated into context-sensitive design considerations for clinical implementation. Rather than proposing new normative frameworks, its central contribution lies in showing how responsible AI requirements are shaped through practical negotiations between competing clinical, psychosocial, technical, and organizational priorities, including accuracy versus accessibility, emotional support versus role boundaries, and scalability versus governance.

Although situated in hereditary cancer care, the identified tensions—including those between autonomy and safety, accessibility and accuracy, and scalability and governance—are likely to extend to other domains of AI-supported, preference-sensitive decision-making in health care.

Limitations

The qualitative and exploratory design prioritizes analytical transferability over statistical generalizability. While this enables in-depth, context-sensitive insights, the findings may not fully represent all clinical settings.

The purposive sampling of stakeholders strengthens practical relevance but introduces potential selection bias. The sample predominantly reflects expert perspectives from the European, specifically German, health care context. Validation in different cultural and regulatory environments is therefore required.

Construct validity may be affected by varying interpretations of abstract concepts such as sustainability. This was addressed through the use of established frameworks (VPC and SusAF) and structured conceptual framing. Nevertheless, reactive bias remains a limitation, as participants in workshop settings may exhibit socially desirable or consensus-oriented responses. Mitigation strategies included neutral facilitation and anonymous postworkshop surveys.

Because workshops were not audio-recorded, the analysis may have missed nuances of interaction, phrasing, or disagreement that would have been available in verbatim transcripts. This limitation was mitigated through parallel field note-taking by multiple facilitators, immediate postworkshop consolidation, triangulation with structured artifacts and framework-based outputs, and consensus-based interpretation by 2 researchers.

Differences in digital literacy among participants likely influenced perceptions of AI-related risks. Although a common baseline was established through introductory explanations, such variation cannot be fully controlled in interdisciplinary settings. In addition, the interpretive subjectivity inherent in qualitative analysis may affect results, despite collaborative coding and iterative validation processes.

The absence of direct patient involvement is a substantive methodological limitation, particularly because the envisioned conversational AI system is intended to support *BRCA1/2* mutation carriers. Patient perspectives were represented indirectly through psycho-oncological and advocacy expertise, which provided clinically and experientially informed approximations of user needs but cannot substitute for direct accounts of lived experience. Future research should therefore involve mutation carriers directly under appropriate ethical and clinical safeguards to validate, refine, and potentially challenge the stakeholder-derived requirements.

The research team's dual role as system designers and analysts may have influenced problem framing and interpretation. While this enabled contextual sensitivity and iterative refinement, it also required continuous reflexive practices to mitigate potential confirmation bias.

The findings are context-dependent and may not be directly transferable to settings lacking institutional oversight, digital literacy, or clearly defined professional responsibility structures.

Although the manuscript discusses key German/EU regulatory considerations, a definitive legal classification of the system under the MDR and the EU AI Act was beyond the scope of this exploratory design study and should be examined in future work together with legal and clinical implementation experts.

Overall, the findings should be understood as context-sensitive rather than definitive, highlighting the need for longitudinal research and broader population-based studies to assess the long-term deployment of responsible AI in health care.

Summary and Conclusion

This study provides a structured account of how established responsible AI principles are negotiated and translated into stakeholder-derived design requirements for conversational AI in *BRCA1/2*-related hereditary cancer care. The findings suggest that effective implementation depends less on increasing system autonomy than on constraining and embedding AI within clinical governance structures.

Across stakeholder groups, responsible use was consistently associated with transparent communication of system limitations, clearly defined role boundaries, and mechanisms for human oversight. Against the backdrop of widely accessible general-purpose LLMs, these findings suggest that the added

XSL•FO
RenderX

value of a purpose-built system lies not in conversational access alone, but in clinically bounded governance, validated content, institutional accountability, data protection, and reliable escalation pathways. These elements emerge as central conditions for trust, appropriate use, and the prevention of over-reliance.

By combining stakeholder analysis, the VPC, and the SusAF, this study demonstrates how abstract principles of responsible AI can be translated into context-sensitive design considerations. Importantly, these considerations arise from negotiated tensions rather than from uniform stakeholder agreement.

Future research should directly involve *BRCA1/2* mutation carriers, evaluate real-world use, and examine how such systems interact with clinical workflows over time. In addition, the findings underline that systematic alignment with evolving regulatory frameworks should not be treated as a downstream compliance task but as an integral part of responsible design, particularly in relation to GDPR, MDR, GenDG, the EU AI Act, and institutional accountability in the German/EU health care context.

Acknowledgments

ChatGPT (OpenAI; model version not documented) was used during manuscript preparation. The tool was used for reference formatting, encompassing standardization of 54 references from heterogeneous source formats. All references were subsequently verified against authoritative sources. The tool was also used for language refinement: ChatGPT suggested minor reformulations of author-written content to ensure consistent use of English throughout the article. Any suggestions were reviewed, accepted, rejected, or modified by the authors. The tool was not used for literature selection, study design, data collection, analysis, interpretation, or the generation of scientific claims. All AI-assisted content was reviewed and verified by the authors, who take full responsibility for the manuscript.

Funding

This research received no external funding.

Conflicts of Interest

JL received honoraria from AstraZeneca GmbH (Germany) and Novartis Pharma GmbH (Germany).

Multimedia Appendix 1
One pager/user case scenario, “Knowledge against fear,” providing support for *BRCA1/2* carriers.
[DOCX File, 110 KB - humanfactors_v13i1e97869_app1.docx]

References

1. Kuchenbaecker KB, Hopper JL, Barnes DR, et al. Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. JAMA 2017 Jun 20;317(23):2402-2416. [doi: 10.1001/jama.2017.7112] [Medline: 28632866]
2. Zilski N, Speiser D, Bartley J, et al. Quality of life after risk-reducing salpingo-oophorectomy in women with a pathogenic BRCA variant. J Sex Med 2023 Dec 22;21(1):33-39. [doi: 10.1093/jsxmed/qdad143] [Medline: 37973412]
3. Mir ZM, Fei LYN, McKeown S, et al. Knowledge, perceptions, attitudes, and barriers pertaining to genetic literacy among surgeons: a scoping review. Can J Surg 2024 Mar 19;67(2):E118-E127. [doi: 10.1503/cjs.001523]
4. Genetic testing: results from FORCE’s 2024 survey highlighting the needs of people at high risk for breast cancer. 2024. URL: https://www.facingourrisk.org/blog/results-from-forces-2024-survey-highlighting-the-needs-of-people-at-high-risk-for-breast-cancer [accessed 2026-09-08]
5. Anderer S. Patients are turning to TikTok for health information—here’s what clinicians need to know. JAMA 2024 Apr 16;331(15):1262. [doi: 10.1001/jama.2024.1280]
6. Schaeffer D, Berens EM, Gille S, Griese L, Klinger J, de Sombre S, et al. Health literacy of the population in germany before and during the corona pandemic: results of the HLS-GER 2. Bielefeld University, Interdisciplinary Centre for Health Literacy Research. 2021 URL: https://pub.uni-bielefeld.de/record/2950305 [accessed 2026-09-08]
7. Barreda M, Cantarero-Prieto D, Coca D, et al. Transforming healthcare with chatbots: Uses and applications-a scoping review. Digit HEALTH 2025;11:20552076251319174. [doi: 10.1177/20552076251319174] [Medline: 40103640]
8. Hindelang M, Sitaru S, Zink A. Transforming health care through chatbots for medical history-taking and future directions: comprehensive systematic review. JMIR Med Inform 2024 Aug 29;12(1):e56628. [doi: 10.2196/56628] [Medline: 39207827]
9. Simonsen J, Robertson T, editors. Routledge International Handbook of Participatory Design: Routledge; 2012. [doi: 10.4324/9780203108543]
10. Osterwalder A, Pigneur Y, Bernarda G, Smith A, Papadakos T. Value Proposition Design: How to Create Products and Services Customers Want: John Wiley & Sons; 2014:1-320.

XSL•FO
RenderX

11. Becker C, Betz S, Chitchyan R, et al. Requirements: the key to sustainability. IEEE Softw 2016;33(1):56-65. [doi: 10.1109/MS.2015.158]
12. Klapdor R, Weiß C, Kuehnle E, et al. Quality of life after bilateral and contralateral prophylactic mastectomy with implant reconstruction. Breast Care (Basel) 2020 Oct;15(5):519-526. [doi: 10.1159/000505449] [Medline: 33223996]
13. Kommission ovar. AGO – die arbeitsgemeinschaft gynäkologische onkologie. 2024. URL: https://www.ago-online.de/ago-kommissionen/kommission-ovar [accessed 2026-09-08]
14. Kommission mamma. AGO – Die Arbeitsgemeinschaft Gynäkologische Onkologie. 2024. URL: https://www.ago-online.de/ago-kommissionen/kommission-mamma [accessed 2026-09-08]
15. Lux MP, Lewis K, Rider A, Niyazov A. Real-world multi-country study of BRCA1/2 mutation testing among adult women with HER2-negative advanced breast cancer. Future Oncol 2022 Mar;18(9):1089-1101. [doi: 10.2217/fon-2021-1387] [Medline: 35098723]
16. Heublein S, Baum J, Jaeger A, et al. Current treatment practices and prognostic factors in early-stage ovarian cancer-an analysis of the NOGGO/JAGO. Cancers (Basel) 2023 Mar 29;15(7):2038. [doi: 10.3390/cancers15072038] [Medline: 37046699]
17. Alves-Nogueira AC, Melo D, Carona C, Figueiredo-Dias M. The psychosocial impact of the decision to undergo risk-reducing salpingo-oophorectomy surgery in BRCA mutation carriers and the role of physician-patient communication. Curr Oncol 2023 Feb 17;30(2):2429-2440. [doi: 10.3390/curroncol30020185] [Medline: 36826146]
18. Yeung A, Ng E, Abi-Jaoude E. TikTok and attention-deficit/hyperactivity disorder: a cross-sectional study of social media content quality. Can J Psychiatry 2022 Dec;67(12):899-906. [doi: 10.1177/07067437221082854] [Medline: 35196157]
19. Loomba S, de Figueiredo A, Piatek SJ, de Graaf K, Larson HJ. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat Hum Behav 2021 Mar;5(3):337-348. [doi: 10.1038/s41562-021-01056-1] [Medline: 33547453]
20. Villa-Ruiz C, Kassamali B, Mazori DR, Min M, Cobos G, LaChance A. Overview of TikTok's most viewed dermatologic content and assessment of its reliability. J Am Acad Dermatol 2021 Jul;85(1):273-274. [doi: 10.1016/j.jaad.2020.12.028] [Medline: 33359080]
21. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. Commun ACM 1966 Jan;9(1):36-45. [doi: 10.1145/365153.365168]
22. Colby KM, Hilf FD, Weber S, Kraemer HC. Turing-like indistinguishability tests for the validation of a computer simulation of paranoid processes. Artif Intell 1972 Jan;3:199-221. [doi: 10.1016/0004-3702(72)90049-5]
23. Mohamed Jasim K, Malathi A, Bhardwaj S, Aw ECX. A systematic review of AI-based chatbot usages in healthcare services. J Health Organ Manag 2025 Jan 28;39(6):877-899. [doi: 10.1108/JHOM-12-2023-0376] [Medline: 39865955]
24. Wang A, Qian Z, Briggs L, Cole AP, Reis LO, Trinh QD. The use of chatbots in oncological care: a narrative review. Int J Gen Med 2023;16:1591-1602. [doi: 10.2147/IJGM.S408208] [Medline: 37152273]
25. Xu L, Sanders L, Li K, Chow JCL. Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review. JMIR Cancer 2021 Nov 29;7(4):e27850. [doi: 10.2196/27850] [Medline: 34847056]
26. Lin SJ, Sun CY, Chen DN, et al. Chatbots for breast cancer education: a systematic review and meta-analysis. Support Care Cancer 2025 Jan;33(1):55. [doi: 10.1007/s00520-024-09096-9]
27. Howcroft A, Bennett-Weston A, Khan A, Griffiths J, Gay S, Howick J. AI chatbots versus human healthcare professionals: a systematic review and meta-analysis of empathy in patient care. Br Med Bull 2025 Sep 22;156(1):ldaf017. [doi: 10.1093/bmb/ldaf017] [Medline: 41115171]
28. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023 Jun 1;183(6):589-596. [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
29. Choi K (, Fitzek S. User and provider experiences with health education chatbots: qualitative systematic review. JMIR Hum Factors 2025;12:e60205-e60205. [doi: 10.2196/60205]
30. Grassini E, Buzzi M, Leporini B, Vozna A. A systematic review of chatbots in inclusive healthcare: insights from the last 5 years. Univ Access Inf Soc 2025 Mar;24(1):195-203. [doi: 10.1007/s10209-024-01118-x]
31. Giuffrè M, Kresevic S, You K, et al. Systematic review: the use of large language models as medical chatbots in digestive diseases. Aliment Pharmacol Ther 2024 Jul;60(2):144-166. [doi: 10.1111/apt.18058] [Medline: 38798194]
32. Chen D, Avison K, Alnassar S, Huang RS, Raman S. Medical accuracy of artificial intelligence chatbots in oncology: a scoping review. Oncologist 2025 Apr 4;30(4):oyaf038. [doi: 10.1093/oncolo/oyaf038] [Medline: 40285677]
33. Huo B, Boyle A, Marfo N, et al. Large language models for chatbot health advice studies: a systematic review. JAMA Netw Open 2025 Feb 3;8(2):e2457879. [doi: 10.1001/jamanetworkopen.2024.57879] [Medline: 39903463]
34. Chow JCL, Li K. Large language models in medical chatbots: opportunities, challenges, and the need to address AI risks. Information 2025;16(7):549. [doi: 10.3390/info16070549]
35. Wulf V, Pipek V, Randall D, Rohde M, Schmidt K. In: Stevens G, editor. Socio-Informatics: A Practice-Based Perspective on the Design and Use of IT Artefacts: Oxford University Press; 2018:35-37.
36. Zweig KA, Krafft TD, Klingel A, Park E. Sozioinformatik: Ein Neuer Blick Auf Informatik Und Gesellschaft: Carl Hanser Verlag; 2021.

XSL•FO
RenderX

37. Chun J, Schroeder de Witt C, Elkins K. Comparative global AI regulation: policy perspectives from the EU, china, and the US. arXiv. Preprint posted online on 2024. [doi: 10.2139/ssrn.5104429]
38. Shrotriya S, Banu N, Kulkarni A, Aiyangar S. Navigating the path to ethical and responsible AI integration in health and life sciences with human and machine collaboration. J Soc Clin Data Manag 2025;5(1):1-12. [doi: 10.47912/jscdm.412]
39. European Commission, High-Level Expert Group on Artificial Intelligence. Ethics guidelines for trustworthy AI. : Publications Office of the European Union; 2019. [doi: 10.2759/22710]
40. Giovanola B, Tiribelli S. Beyond bias and discrimination: redefining the AI ethics principle of fairness in healthcare machine-learning algorithms. AI Soc 2023 Apr;38(2):549-563. [doi: 10.1007/s00146-022-01455-6]
41. Miller T. Explanation in artificial intelligence: insights from the social sciences. Artif Intell 2019 Feb;267:1-38. [doi: 10.1016/j.artint.2018.07.007]
42. Bostrom N. Superintelligence: Paths, Dangers, Strategies: Oxford University Press; 2014.
43. Floridi L, Cowls J, Beltrametti M, et al. AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds Mach 2018 Dec;28(4):689-707. [doi: 10.1007/s11023-018-9482-5]
44. Grunwald A. Technikfolgenabschätzung: Eine Einführung, 2nd edition: edition sigma, Berlin; 2010:1-346.
45. Friedman B, Hendry DG. Value Sensitive Design: Shaping Technology with Moral Imagination: MIT Press; 2019.
46. EDPB Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects. European Data Protection Board (EDPB). 2019 Oct 8 URL: https://content.next.westlaw.com/practical-law/document/Ic162f9fbeff311e9adfea82903531a62/EDPB-Guidelines-2-2019-on-the-processing-of-personal-data-under-Article-6-1-b-GDPR-in-the-context-of-the-provision-of-online-services-to-data-subjects?viewType=FullText&transitionType=Default&contextData=(sc.Default) [accessed 2026-09-19]
47. Ebner GK. Weniger ist mehr? die informationspflichten der DS-GVO – eine kritische analyse. : Nomos, Baden-Baden; 2022. [doi: 10.5771/9783748936879]
48. Burk E, Chen X, Haftenberger A, Khorasani H. Künstlich intelligente Medizinprodukte im regulatorischen Korsett aus MDR/IVDR und KI-VO – Komplikation mit Mehrwert. Medizin Produkte Recht 2025:176-185.
49. Ebers M. AI robotics in healthcare between the EU medical device regulation and the Artificial Intelligence Act: gaps and inconsistencies in the protection of patients and care recipients. Oslo Law Rev 2024;11(1):1-12. [doi: 10.18261/olr.11.1.2]
50. Hauglid MK, Mahler T. Doctor chatbot: the EU s regulatory prescription for generative medical AI. Oslo Law Rev 2023 Nov 14;10(1):1-23. [doi: 10.18261/olr.10.1.1]
51. Schulz-Große S, Genske A. Auswirkungen der neuen KI-Verordnung auf den Behandlungsalltag – Datenschutzrechtliche Anforderungen und Sorgfaltspflichten der Behandelnden beim Einsatz von künstlicher Intelligenz zu Behandlungszwecken. : Gesundheit und Pflege; 2023 URL: https://beck-online.beck.de/Dokument?vpath=bibdata%2Fzeits%2Fgup%2F2023%2Fcont%2Fgup.2023.81.1.htm&pos=6 [accessed 2026-09-19]
52. Beckers R, Kwade Z, Zanca F. The EU medical device regulation: Implications for artificial intelligence-based medical device software in medical physics. Phys Med 2021 Mar;83:1-8. [doi: 10.1016/j.ejmp.2021.02.011] [Medline: 33657513]
53. Digital omnibus on AI regulation proposal. European Commission. 2025. URL: https://digital-strategy.ec.europa.eu/en/library/digital-omnibus-ai-regulation-proposal [accessed 2026-09-08]
54. Radtke T. Die (unzureichende) private Rechtsdurchsetzung unter der KI-Verordnung. JZ 2026;81(4):127-136. [doi: 10.1628/jz-2026-0051]

Abbreviations

BDSG: Bundesdatenschutzgesetz (Federal Data Protection Act)
BGB: Bürgerliches Gesetzbuch (German Civil Code)
CJEU: Court of Justice of the European Union
COREQ: Consolidated Criteria for Reporting Qualitative Research
EU: European Union
FBREK: German Consortium for Familial Breast and Ovarian Cancer
FORCE: Facing Our Risk of Cancer Empowered
GDPR: General Data Protection Regulation
GenDG: German Genetic Diagnosis Act
IPDAS: International Patient Decision Aid Standards
LLM: large language model
MDR: Medical Device Regulation
RAG: retrieval-augmented generation
SRB: Single Resolution Board
SusAD: sustainability awareness diagram
SusAF: sustainability awareness framework
UX: user experience
VPC: value proposition canvas

XSL•FO
RenderX

VSD: value-sensitive design
WHO: World Health Organization
xAI: explainable AI

Edited by H Witteman; submitted 10.Apr.2026; peer-reviewed by AT Saeed, T Coffin; revised version received 06.Aug.2026; accepted 12.Aug.2026; published 30.Sep.2026.

Please cite as:
Lammert D, Lammert J, Hofenbitzer J, Radtke T, Sun Y, Haimerl M, Betz S, Scholl DH, Pfeffer J
Conversational AI in Hereditary Cancer Care: Sociotechnical Study of Responsible Design Requirements
JMIR Hum Factors 2026;13:e97869
URL: https://humanfactors.jmir.org/2026/1/e97869
doi:10.2196/97869

© Dominic Lammert, Jacqueline Lammert, Justin Hofenbitzer, Tristan Radtke, Yingping Sun, Martin Haimerl, Stefanie Betz, Doran Holger Scholl, Jürgen Pfeffer. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 30.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Risk Visualization With Petal-X and Continuous Glucose Monitoring to Support Person-Centered Care and Lifestyle Change in Cardiovascular Prevention: Qualitative Descriptive Study

Adrijana Svenšek[1], PhD; Diego Rojo[2], MSc; Mateja Lorber[1], PhD; Zalika Klemenc-Ketis[3,4,5], PhD; Lucija Gosak[1], PhD; Katrien Verbert[2,6], PhD; Gregor Štiglic[1,7,8,9], PhD

[1]Faculty of Health Sciences, University of Maribor, Žitna ulica 15, Maribor, Slovenia

[2]Department of Computer Science, KU Leuven, Leuven, Flanders, Belgium

[3]Community Health Centre Ljubljana, Primary Healthcare Research and Development Institute, Ljubljana, Ljubljana, Slovenia

[4]Department of Family Medicine, Faculty of Medicine, University of Maribor, Maribor, Slovenia

[5]Department of Family Medicine, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia

[6]Augment-IMEC Research Group at KU Leuven, Flanders, Belgium

[7]Susan Wakil School of Nursing and Midwifery, The University of Sydney, Sydney, New South Wales, Australia

[8]Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia

[9]Usher Institute, University of Edinburgh, Edinburgh, Scotland, United Kingdom

Corresponding Author:
Adrijana Svenšek, PhD
Faculty of Health Sciences, University of Maribor, Žitna ulica 15, Maribor, Slovenia

Abstract

Background: Cardiovascular diseases (CVDs) remain a leading cause of morbidity and mortality. In primary prevention, effective communication of individual risk and support for behavior change are crucial. Interactive risk visualizations can enhance understanding and engagement, while continuous glucose monitoring (CGM) offers immediate feedback on the physiological effects of daily choices.

Objective: This study aimed to explore how patients at increased CVD risk and family medicine nurses perceive CVD risk visualization (Petal-X) and the integration of real-time glucose monitoring, with a focus on perceived person-centered care (PCC), motivation, and lifestyle change. This manuscript reports the qualitative phase of a larger sequential mixed methods study; the quantitative randomized controlled trial is reported separately.

Methods: Using total population (census) sampling within a closed, predefined trial cohort, we invited every eligible patient who had completed the intervention arm of a parallel randomized controlled trial and every eligible registered nurse providing preventive consultations at the same single community health center in central Slovenia who had been familiarized with the intervention and with the Petal-X dashboard; 6 patients at increased CVD risk who had used the Petal-X dashboard in combination with CGM, and 6 nurses familiar with the intervention, took part, and interviews were conducted between September and December 2024. Data were collected through individual, face-to-face semistructured interviews that explored perceptions of the visualization, real-time glucose monitoring, and lifestyle change. Interviews were audio-recorded, transcribed verbatim, and analyzed by inductive qualitative content analysis within a qualitative descriptive design; coding continued until code saturation and data adequacy were judged to have been reached.

Results: The analysis produced 11 subcategories, which were grouped into 4 categories, with an overarching theme running across them: visualization and digital tools as support for PCC to enhance motivation and lifestyle change in individuals at risk of CVD. Patients and nurses reported that personalized visual displays improved understanding of risk factors, facilitated dialogue and shared decision-making, and increased patients' sense of involvement, responsibility, and control, although the 2 groups differed in emphasis: patients foregrounded personal meaning and self-regulation, whereas nurses foregrounded communication and fit with the consultation workflow. CGM was perceived as particularly motivating by linking behavior (food, activity, and stress) to real-time glucose changes, encouraging dietary and activity adjustments that some participants maintained after the intervention. Key barriers included variable digital literacy (especially among older users), fear or mistrust of technology, the attentional burden of continuous monitoring, potential information overload, workflow integration needs, and affordability or accessibility of CGM.

XSL•FO
RenderX

Conclusions: Interactive CVD risk visualization Petal-X, especially when complemented by CGM, was perceived as a meaningful support for PCC and a driver of motivation and lifestyle change in primary prevention. Implementation should address training, usability, integration into routine workflows, and equity of access.

Trial Registration: ClinicalTrials.gov NCT06226948; https://clinicaltrials.gov/study/NCT06226948

(JMIR Hum Factors 2026;13:e92101) doi:10.2196/92101

KEYWORDS

cardiovascular disease; dashboard; continuous glucose monitoring; person-centered care; lifestyle change; qualitative content analysis; registered nurses

Introduction

Primary health care in Slovenia is based on a network of 63 community health centers organized into 10 administrative regions, with approximately 900 teams of which around 200 operate privately under concession [1]. Family medicine practices offer both preventive and curative services. The preventive services are focused on individuals aged 30 years and older with no chronic disease, where registered nurses (RNs) assess the risk of developing chronic conditions [2,3]. In this process, prognostic models are used, which combine multiple predictive factors into a mathematical equation to estimate the likelihood of an outcome [4,5]. One of these is the SCORE2 (Systematic Coronary Risk Evaluation 2) model for estimation of cardiovascular disease (CVD), validated on the European population, which includes age, sex, smoking status, systolic blood pressure, total cholesterol, and high-density lipoprotein cholesterol [6]. Such models enable the early identification of high-risk individuals and guide the allocation of limited health care resources where they can achieve the most significant impact [7,8].

Nevertheless, CVDs remain the leading cause of death worldwide, accounting for over 20 million deaths each year [9]. Many individuals do not recognize how their daily habits (diet, physical activity, and smoking) directly influence their long-term CVD risk [10]. Therefore, effective communication of risk and the potential benefits of lifestyle change play a critical role in CVD prevention [11].

Visual tools are becoming an increasingly important approach in health care communication, as graphical and interactive representations of data improve understanding and foster motivation for lifestyle modification [9]. Digital health solutions that present personalized risk and the potential effects of change can help patients perceive their health status more tangibly. The growing digitalization of health care offers new possibilities for integrating real-time data with personalized risk assessment [12,13]. Continuous glucose monitoring (CGM) devices, for example, provide patients with immediate feedback on how their dietary and behavioral choices affect their blood glucose levels [14,15]. When combined with visualizations of long-term CVD risk, such as those provided by digital tools, this integration can support shared decision-making, enhance patient engagement, and promote more timely and sustainable lifestyle changes [16,17].

In this context, we examined the Petal-X visualization tool, which is based on the Petal Product Plots visualization and an explanation layer on top of the SCORE2 model. Its accuracy is comparable with the graphical score charts used in clinical practice. The developers of Petal-X conducted a study with nursing students, finding that its use improved understanding of risk factors and stimulated reflection on possible lifestyle changes [18].

The aim of our qualitative study was to explore how visualization of CVD risk, combined with the possibility of real-time blood glucose monitoring, influences the perception of person-centered care (PCC) among patients at increased risk of CVD. Furthermore, we sought to understand how RNs and patients perceive the use of these digital tools, as well as the opportunities and challenges they identify for their integration into everyday clinical practice.

Methods

Research Design

This research forms the qualitative phase of a sequential mixed methods design [19], carried out after the quantitative phase. The results of the quantitative analysis reported by Svenšek et al [20] indicated certain patterns that we wanted to explore in more detail. For this phase, we used a qualitative descriptive design [21], with data analyzed by inductive qualitative content analysis [22,23]. This approach was chosen because the aim of the study was descriptive and evaluative rather than theory-generating: we set out to characterize how patients who had completed the intervention arm, and the nurses working in the same preventive service who had been familiarized with it, experienced the Petal-X dashboard and CGM, including perceived usability, the perceived contribution to PCC and to motivation, and the barriers to implementation in routine practice. The study was not designed to develop a theory of a social or clinical process, and no theory-generating claims are made. We collected the data through individual semistructured interviews. We ensured the rigor of the qualitative study by adhering to Lincoln and Guba's [24] criteria of credibility, dependability, confirmability, and transferability. Procedures are reported in accordance with the COREQ (Consolidated Criteria for Reporting Qualitative Research) 32-item checklist [25].

Setting, Participants, and Recruitment

The study was conducted at a single community health center in central Slovenia that delivers preventive family medicine services. It was nested within a parallel randomized controlled trial (ClinicalTrials.gov NCT06226948) evaluating interactive

XSL•FO
RenderX

digital visualization counseling for patients at risk of CVD [20]. Recruitment for the interviews was open to individuals across Slovenia, although the promotional campaign and study implementation were concentrated at this center. Patients were drawn from the trial intervention arm (n=30 at baseline), in which participants had a mean age of 51.9 (SD 7.2) years, 70% (21/30) were women, and the mean baseline 10-year SCORE2 risk was 4.87 (SD 2.61) and the highest of the 4 trial arms [20].

Patients were eligible if they were aged 40 years or older, had no diagnosed CVDs, and had at least 1 CVD risk factor (elevated systolic blood pressure, elevated fasting glucose, elevated blood lipids, or elevated BMI); nurses were eligible if they provided preventive consultations at the participating center, had been familiarized with the Petal-X dashboard and with CGM-based monitoring during the study period, and had at least 5 years of professional experience as a nurse, which was verified individually for each nurse before inclusion. Eligible patients who had completed the trial intervention arm were identified from the trial records by the study team, and eligible nurses were identified through the participating center; potential participants were then contacted individually by the first author (AS), in person or by telephone, informed about the aim and the voluntary nature of the study, and, if they agreed, gave written informed consent before an interview was scheduled. Recruitment was not purposive. We used a total population (census) approach within the closed, predefined cohort created by the parent trial: every patient who had completed the trial intervention arm and every nurse at the participating center who had been familiarized with the intervention was invited, and no further selection step was applied within this eligible pool, either on the basis of participant characteristics (eg, age, sex, and baseline risk category) or on the basis of the content of their experience. The size and composition of the sample were therefore determined by the boundaries of the parent trial and by willingness to participate, rather than by deliberate selection for particular characteristics, by maximum variation, or by an emergent analytic rationale. No more than 2 interviews were conducted per week so that each interview could be analyzed before the next. The final sample comprised 6 nurses and 6 patients.

Definition of Intervention Completion and Extent of Exposure

Because eligibility for interview depended on having completed the intervention arm, we define here what completion involved in this study, independently of the report of the parent trial [20]. The intervention arm did not consist of a single encounter with the dashboard. Each patient attended a face-to-face preventive consultation, delivered by the first author (AS), an RN and PhD candidate, during which the patient's individual 10-year SCORE2 risk was calculated and explored interactively using the Petal-X dashboard (duration of consultation: 30 minutes). Each patient was then fitted with a CGM sensor, which was worn during everyday life for 10 days, with real-time access to their own glucose readings through the associated smartphone app, and attended a further study visit 6 months later, at which the glucose record and the risk profile were reviewed together. The intervention therefore comprised both a nurse-led visualization consultation and a defined period of CGM-based self-monitoring. Patients were counted as completers of the intervention arm and were therefore eligible to be approached for an interview if they attended all 3 scheduled study visits (the baseline assessment, blood sampling at 3 months, and the final assessment at 6 months), wore the CGM sensor for at least 5 of the 10 intended days, and completed the final study assessment. Patients who withdrew, who discontinued sensor wear before this threshold, or who did not attend the final assessment were not counted as completers and were not invited to interview. Of the 30 patients randomized to intervention group 1 (Petal-X visualization combined with CGM), 27 met this definition of completion and were invited, and 6 agreed to be interviewed. Interviews were conducted a median of 5 weeks after the end of each patient's participation, so that all patients described a completed rather than an ongoing experience of the intervention.

For nurses, participation did not involve delivering the intervention. All trial consultations, including the presentation of the Petal-X visualization and the application and explanation of the CGM sensor, were carried out by the first author (AS), and the participating nurses had no formal delivery role in the trial. They were RNs providing routine preventive consultations at the same center who had been familiarized with the intervention during the study period: each nurse took part in a single structured familiarization session of approximately 30 minutes, in which the Petal-X dashboard was presented and explored with example risk profiles, the rationale of the SCORE2-based visualization was explained, and the principles and practical use of CGM-based monitoring were demonstrated. Within this session, the nurses also used the dashboard themselves on example profiles, so that their appraisals were based on direct, if brief, hands-on contact with the tool rather than on observation alone. They also worked alongside the trial in the same preventive service and were therefore familiar with how it was organized and with the patients taking part. At the time of the interview, each nurse had approximately 7 months of familiarity with the tool and the intervention, which was delivered between May and December 2024, and the dashboard was displayed again at the start of the interview so that the discussion could refer to specific features of the visualization.

In this study, therefore, "completion" has a different meaning for the 2 groups. For patients, it denotes completion of the intervention arm as defined in the section Setting, Participants, and Recruitment; for nurses, it denotes completion of this familiarization and continuous presence in the setting for the whole intervention period, rather than a threshold of sessions delivered. Their accounts should accordingly be read as the informed appraisal of experienced preventive care nurses who knew the intervention and the tool well, and who work daily with the paper-based SCORE2 charts that the dashboard is intended to replace, rather than as the reflections of practitioners who had themselves delivered it in routine care; this distinction is revisited in the Limitations section.

Data Collection

Interviews were conducted by the first author (AS), a female RN holding a master's degree (MSc), who was trained in qualitative research methods and was a doctoral researcher at

XSL•FO
RenderX

the University of Maribor and teaching assistant at the time of the study. The interviewer had no prior clinical or supervisory relationship with the participants before the study commenced. Participants were informed of the researcher's professional and academic background and that the interviews formed part of a doctoral research project on digital tools for cardiovascular risk communication and prevention. The researcher's background in nursing and digital health shaped an interest in PCC and technology-supported prevention; to limit the influence of these assumptions, reflexive memos were kept, and the developing interpretation was discussed in regular peer-debriefing meetings within the research team.

We conducted interviews with RNs who had been familiarized with the Petal-X dashboard for CVD risk visualization and with the potential for real-time blood glucose monitoring during the study period; at the start of each interview, the dashboard was displayed again so that the discussion could refer to specific features of the visualization. The patients we interviewed had all taken part in the intervention arm that combined the Petal-X dashboard with real-time monitoring of blood glucose levels. We were interested in the participants' perceptions of the intervention and the use of new technologies. We used semistructured interviews to preserve certain concepts [24]. With the help of semistructured interviews, we obtained all the necessary information for the research, allowing participants as much freedom as possible in their responses. We opted for semistructured interviews because we did not want to stray from our topic and wanted to obtain an accurate description and perception of the participants' involvement in interventions and visualizations related to CVD risk. With semistructured interviews and semi–open-ended questions, we also aimed to gather information about the experiences, feelings, and advice of RNs and research participants regarding the visualization of the risk of CVD using Petal-X and the potential for real-time monitoring of blood glucose levels. The interview guides for nurses and for patients are provided in Multimedia Appendices 1 and 2.

Data Analysis

All interviews were audio-recorded, transcribed verbatim, and anonymized prior to analysis. Interviews were conducted and transcribed in Slovenian, and all coding and analysis were carried out in Slovenian to stay close to participants' original wording; the illustrative quotations presented in the Results section were subsequently translated into English by the bilingual research team and checked against the original transcripts for accuracy and meaning. Data were analyzed by inductive qualitative content analysis [22,23]. Analysis proceeded alongside data collection, so that preliminary observations could inform the emphasis given to particular questions in later interviews.

Each transcript was first read repeatedly to obtain a sense of the whole. Meaning units relevant to the aim of the study were then identified and condensed, and each condensed meaning unit was labeled with a code that stayed close to participants' own wording. The first transcripts were coded independently by 2 researchers (AS and LG), who compared their code lists and agreed on a shared coding frame; this frame was applied to the remaining transcripts and revised whenever new codes were added. Codes were compared systematically within and across the 2 participant groups and sorted, on the basis of similarity in content, into subcategories and then into categories. Finally, an overarching theme expressing the thread of meaning that ran across the categories was formulated. The developing category structure was reviewed and refined together with a senior researcher (GS) in peer-debriefing meetings until consensus was reached. Categories were refined iteratively by writing analytic memos, revisiting transcripts, and actively searching for discrepant accounts (eg, accounts of CGM worry or of limited benefit for some users). An audit trail (coding memos and versioned code lists) was maintained. To strengthen credibility, verbatim transcripts were returned to participants for review and possible correction (member checking).

Because the sample was bounded by the closed intervention arm of the parent trial, the size of the sample was fixed in advance rather than determined by the progress of the analysis. We therefore considered code saturation and the adequacy of the data for the descriptive aim of the study [26]. Toward the end of data collection, later interviews were adding few, if any, new codes, and the research team judged the material sufficient to describe the categories reported. We present this as an assessment of data adequacy rather than as a formal claim of saturation, since the number of available participants was fixed by the parent trial. In the Results section, quotes are attributed using participant identifiers to preserve anonymity: P denotes patients and N denotes RNs (eg, P1, N2). Each quotation is additionally accompanied by brief, nonidentifying contextual information about the speaker, namely, sex, age, and baseline 10-year SCORE2 risk category for patients, and years of professional experience for nurses, so that readers can interpret each quotation in relation to the individual speaking rather than only in relation to the participant group.

Qualitative Rigor

Qualitative rigor was addressed using established criteria for qualitative rigor: credibility, dependability, confirmability, and transferability. Credibility was supported by verbatim transcription, systematic comparison of codes within and across the 2 participant groups, member checking (returning transcripts to participants for review and possible correction), regular peer-debriefing meetings within the research team, and the deliberate search for and analysis of negative cases. Dependability and confirmability were supported by maintaining an audit trail of coding memos and versioned code lists, so that the development of the codes, subcategories, and categories could be traced back to the underlying data. Transferability was supported by providing a thick description of the setting, participants, and context, enabling readers to judge the applicability of the findings to other primary care settings.

Ethical Considerations

The study was approved by the Ethics Committee for Nursing Research at the Faculty of Health Sciences, University of Maribor (approval number 02/7K-2023), and the National Medical Ethics Committee of the Republic of Slovenia (approval number 0120-217/2023/4). We have also registered the research protocol of the clinical trial (ClinicalTrials.gov ID:

XSL•FO
RenderX

NCT06226948) [27]. All participants were informed that their data would be used exclusively for research purposes. All participants received written and oral information about the study and gave written informed consent before their interview, including explicit consent to audio-recording and to the use of anonymized verbatim quotations in publications. Recordings were transcribed verbatim and deidentified; direct identifiers were removed and participants are referred to only by coded identifiers (P for patients and N for nurses). Data were stored securely, were accessible only to the research team, and were used exclusively for research purposes. Participants did not receive any financial compensation for taking part in the study.

Results

Participant Characteristics

The sample comprised 6 RNs and 6 patients at increased CVD risk, all from a single community health center in central Slovenia. All patients had taken part in the intervention arm of the parallel randomized controlled trial [20] that combined the Petal-X dashboard with CGM; they were aged 40 years or older, had no diagnosed CVD, and had at least 1 CVD risk factor. Among the 6 patients interviewed, 4 were women and 2 were men, aged 46 - 67 (mean 57.8, SD 7.8) years; at baseline, 1 was at low, 2 at moderate, and 3 at high 10-year SCORE2 risk. The interviewed patients were therefore somewhat older than the intervention arm as a whole, in which the mean age was 51.9 (SD 7.2) years [20]. All 6 participating nurses were women, aged 26 - 50 (mean 33.3, SD 10.2) years, and each had at least 5 years of professional experience as a nurse and worked in preventive family medicine consultations at the participating center throughout the trial period. To avoid duplication, participant characteristics are reported in the text only; brief individual descriptors accompany each quotation so that quotations can be read in relation to the individual speaker.

The analysis produced 11 subcategories, which were grouped into 4 categories (Table 1). Running across all 4 categories was an overarching theme, formulated as “Visualization and digital tools as support for PCC to enhance motivation and lifestyle change in individuals at risk of CVD.”

XSL•FO
RenderX

Table . Display categories and subcategories with codes.

Categories and subcategories	Codes
Person-centered care	
Goal-oriented care and individual approach	Time devoted to the patient; patient-centered care; tools as a means of improving treatment; individualized approach, and treatment; daily routine; meal planning; recommendations for the use of technology; involvement of family members; faster task completion; and ease of treatment.
Involvement in treatment	Patient involvement, treatment planning, increased sense of responsibility, sense of control, active role of the patient, and quality of treatment.
Knowledge and monitoring of risk factors for the development of CVD[a]	
Monitoring of vital functions and detection of changes or deviations	Real-time monitoring of blood glucose levels, monitoring of risk factors, changes in monitored vital signs, immediate monitoring of condition, detection of changes, and faster identification of risks.
Risk factors	Tool for identifying variable risk factors, raising awareness of risk, reducing risk, visualizing risk, incorporating relevant risk factors, transparency and relevance of factors, link between data and risk, and link between food and sugar.
Understanding of impacts	The impact of risk factors on disease vulnerability, visual representation of health status, understanding vulnerability, understanding necessary changes, understanding through visualization, simplicity and clarity of visualization, and understanding lifestyle changes.
Visualization and interpretation of results and the impact of technology use on lifestyle changes	
Personalized graphical display of results	Interactivity of visualization, real-time data visualization, visualization without explanation is not always informative enough, graphical representation of data, decision-making based on visualization, visual presentation as an advantage, color visualization, connection between colors and risk, display of changes, visual type, visualization as a communication tool, facilitating conversation.
Improved understanding of the interpretation of results	Easier understanding of data; simplification of complex data, visualization as a tool for clear presentation, advantage of images over numbers, awareness and impact on results, advantage of visualization over SCORE2[b] tables, graphical representation of risk, visual displays enable faster and more accurate understanding without complicated procedures, transparency of visualization, ease of presentation for patients, and easier interpretation of results for patients.
Technology and usability	Significant impact of technology, usefulness of CGM[c] for people with diabetes, self-monitoring and prevention, disease prevention, benefits for patients, and practicality.
Motivation and lifestyle changes	Patient motivation, changes in physical activity, motivation of the older adult population, motivation to change lifestyle, dietary changes, maintaining lifestyle changes, connection with meals, increased awareness, and control over food intake.
Identification of problems and challenges and recommendations for improvement	
Identification of problems or deviations	Identifying problems, realizing deviations, recognizing patterns, linking to results, reservations about use and lack of trust in technology, patients' fear of using technology, negative interpretation of technology, and the challenge of affordability.
Opportunities for improvement and recommendations	Opportunities for improvement and emphasis on improvements, recommendation for wider implementation, link between decisions and health, potential for the future, application in other areas of health care, lack of education and knowledge and its inclusion, and improvement of visual tools.

[a]CVD: cardiovascular disease.

[b]SCORE2: Systematic Coronary Risk Evaluation 2.

[c]CGM: continuous glucose monitoring.

XSL•FO
RenderX

Overview of the Findings

The 4 categories describe complementary aspects of participants' experience rather than stages of a process. PCC describes the character of the consultation as participants experienced it: goal-oriented, individualized, and involving the patient in decisions. Knowledge and monitoring of risk factors describes what participants reported learning and noticing through Petal-X and CGM, including the recognition of deviations and the connection between everyday behavior and measurable indicators. Visualization and interpretation of results and the impact of technology use on lifestyle changes describes how the display itself was used and appraised, together with the dietary and physical activity changes that patients associated with it. Identification of problems and challenges and recommendations for improvement gathers the limitations that both groups reported, fear or mistrust of technology, the burden of continuous monitoring, variable digital literacy, information overload, and the cost of CGM, alongside their suggestions for addressing them.

Across all 4 categories, patients and nurses converged on the perceived usefulness of the visual display but differed consistently in emphasis. Patients spoke primarily about personal meaning, reassurance, and self-regulation; nurses spoke primarily about communication, the efficiency of the consultation, and the fit of the tool with existing workflows. Where the 2 groups diverged, this is made explicit in the sections that follow.

Category 1: Person-Centered Care

Goal-Oriented Care and Individual Approach

Participants described consultations as more individualized and goal-focused than usual encounters, with more time for dialogue and tailoring recommendations to daily routines (eg, meal planning, physical activity, and family support). Nurses emphasized that visual displays supported more transparent communication and helped patients identify personally salient risk factors. Interpreted within the core category, these accounts suggest that the value of the consultation lay less in the visual tool itself than in how it reorganized the encounter around the individual: by making risk personally legible, the visualization created a shared reference point that shifted the interaction from information delivery toward negotiated, goal-oriented care. Patients and nurses agreed that a consultation supported by the visualization differed from a routine preventive encounter, but they located that difference in different places. For patients, the difference was one of attention and time: the consultation was experienced as being about them rather than about a presenting complaint. One participant contrasted it directly with her usual visits, noting that "the consultation was dedicated to me. We talked about me, my lifestyle and my habits. It took more time than my usual healthcare visits, where I normally go only because of specific problems," and adding that she "felt more involved and received more information about my health" (P1; woman, aged 67 years, baseline SCORE2 risk 4%). Nurses rarely framed the change in terms of time; their emphasis fell on what the display did for the structure of the conversation, describing it as a way of making the assessment legible to the patient, in that "visualization is very useful in terms of transparency...easier to present to patients" (N2; ≥5 years of professional experience).

Read together, the 2 accounts suggest that the perceived value of the consultation lay less in the visual tool itself than in how it reorganized the encounter around the individual. By giving both parties a shared reference point, the visualization shifted the interaction away from the delivery of a risk figure and toward a negotiated discussion of daily routines, meal planning, and physical activity, which in several accounts extended to the involvement of family members. This is the sense in which participants described the care as goal-oriented: the goals were formulated during the consultation, and the display supplied the common ground on which they were agreed.

Involvement in Treatment

Both groups linked this reorganization of the consultation to a more active patient role, again with a difference of emphasis. Patients framed involvement in terms of partnership, control, and responsibility, describing a shift from recipient to interlocutor: one participant said that with this kind of care she "felt like a real conversation partner. We discussed what I could improve, and I had not experienced something like that before," and connected seeing and understanding her data in real time with a "stronger sense of control over my health" (P3; woman, aged 58 years, baseline SCORE2 risk 7.8%). Nurses described the same shift from the outside, as a lowering of the threshold for participation, since with the visual display it was "easier for patients to get involved...easier to understand" (N3; ≥5 years of professional experience). Patients in turn linked this sense of control and responsibility to their willingness to follow the recommendations agreed during the consultation. It is worth noting that none of the reservations raised in the interviews concerned the shared discussion of results itself; where patients reported burden, it related to continuous self-monitoring, described under category 4, rather than to their involvement in decisions about their care.

Category 2: Knowledge and Monitoring of Risk Factors for the Development of CVD

Monitoring Vital Functions and Detecting Deviations

Real-time monitoring (particularly CGM) was described as enabling rapid feedback and earlier recognition of deviations. Participants highlighted the importance of continuous access to data; nurses emphasized increased awareness and perceived patient independence. This pattern indicates that real-time monitoring did more than supply data; by compressing the interval between behavior and feedback, it appeared to support a sense of agency and self-regulation, which participants linked to sustained motivation rather than to a single moment of insight. Real-time monitoring, and CGM in particular, was the element of the intervention that patients raised most spontaneously. What they valued was less the data themselves than the continuous availability of feedback and the short interval between an action and its visible consequence. One participant described the sensor as a constant companion that also allowed retrospective reflection: "Those data were always on my side...when I saw that the glucose level was high, or when it increased, I was a little more careful to eat less...I could look back over the last

XSL•FO
RenderX

10 days and reflect on my state over time" (P1; woman, aged 67 years, baseline SCORE2 risk 4%). The color coding of the display was repeatedly described as doing interpretive work that numbers alone did not: "Numbers do not tell you enough, because you do not know exactly where the threatened area is. When it is visualized and turns orange or red, you experience it almost like a silent alarm, and then you approach lifestyle change more seriously" (P4; woman, aged 58 years, baseline SCORE2 risk 5.8%).

Nurses observed the same behavior from a professional standpoint but described it as a change in patients' attentiveness and independence rather than as a personal experience, reporting that "patients monitor their condition more closely...more aware" (N2; ≥5 years of professional experience). Across both groups, therefore, monitoring was described as doing more than supplying data: by compressing the interval between behavior and feedback it appeared to support self-regulation, which participants linked to sustained attention to their habits rather than to a single moment of insight. The contrast with the annual preventive checkup, in which feedback arrives once and well after the behavior it reflects, was implicit in several patient accounts.

Identifying Risk Factors

Whereas CGM was associated with day-to-day feedback, Petal-X was associated with making the composition of long-term risk visible. Patients described the dashboard as the first occasion on which their risk profile had been presented in a form they could interpret; one participant recalled being able to "see for the first time how high my risk [was]...and what that means" (P3; woman, aged 58 years, baseline SCORE2 risk 7.8%). Nurses' accounts converged with this, describing the display as making the contribution of individual modifiable factors transparent and relevant, and therefore easier to connect to a concrete recommendation within the consultation. What patients added to the nurses' account was the personal meaning of that transparency: the value they described lay in relating the displayed factors to their own everyday behavior, rather than in the precision of the estimate itself.

Understanding Impacts

Participants described visualization as enhancing comprehension of how lifestyle factors influence risk and what changes are needed. Nurses similarly described visual feedback as supporting awareness and action. Taken together, these accounts frame comprehension as a mechanism rather than an outcome: translating abstract, probabilistic risk into a concrete visual form appeared to make the link between everyday choices and long-term risk actionable, consistent with the core category that connects visualization and digital tools to motivation and lifestyle change. Both groups described comprehension as the mechanism through which the visualization worked, rather than as an end point in itself. Patients contrasted the visual display with the numerical formats they had encountered previously: "Before, when I received results only as percentages, I could not fully imagine what those numbers actually meant for me. The visualization tool allowed me to see how my everyday decisions, such as diet and physical activity, directly affected my health. It was not just a number on paper, but a clear picture of how my choices influenced my risk" (P2; man, aged 63 years, baseline SCORE2 risk 10%). Nurses described the same effect in terms of its immediacy within the encounter, observing that with the visual display "patients immediately become aware" (N2; ≥5 years of professional experience). The difference between the 2 perspectives is largely one of timescale: nurses referred to comprehension achieved during the consultation, whereas patients described an understanding that persisted afterward and informed subsequent choices. Taken together, the accounts indicate that translating an abstract, probabilistic estimate into a concrete visual form was what made the link between everyday choices and long-term risk actionable for participants.

Category 3: Visualization and Interpretation of Results and the Impact of Technology Use on Lifestyle Changes

Personalized Graphical Display and Interpretation

Interactivity was described as the feature that turned the display from a presentation into a discussion. Patients reported that reviewing the graphs together created an opening for asking questions and for proposing changes themselves: "When we reviewed the graphical displays of my results together, I could ask questions and suggest changes to my care plan. In the past I was often just a listener, but now I was truly involved in deciding about my health" (P2; man, aged 63 years, baseline SCORE2 risk 10%). Nurses valued the same interactivity for a different reason, namely, the speed with which it surfaced where change was possible; one nurse noted that the tool "very clearly illustrates deviations...opportunities for improvement" (N4; ≥5 years of professional experience). Nurses qualified this in an important respect: several observed that a visualization presented without accompanying explanation was not in itself sufficiently informative, and therefore the professional's interpretation remained necessary. That qualification locates the benefit in the combination of display and dialogue rather than in the display alone, and it is consistent with the patients' own emphasis on the conversation that the graphs made possible.

Technology and Usability

Appraisals of usability were broadly positive in both groups but were expressed in group-specific terms. Patients spoke of the technology as unproblematic in use, one stating simply that she was "very satisfied with the technology we used" (P3; woman, aged 58 years, baseline SCORE2 risk 7.8%). Nurses assessed usability against the tools they currently use, and it was in this comparison that the perceived advantage was most clearly articulated: "The table we use in the clinic is very comprehensive and time-consuming, because you have to look for each piece of data separately and compare it. With this tool, everything is together, explained and shown in one place, and it is visually simpler" (N3; ≥5 years of professional experience). The nurses' positive appraisal was nevertheless conditional. They regarded the tool as practical and promising for preventive family medicine provided that it could be fitted into existing consultation workflows and that additional support was available for patients with lower digital literacy, conditions to which they returned when discussing implementation barriers.

XSL•FO
RenderX

Motivation and Lifestyle Changes

Patients' accounts of behavior change were concrete and specific, and in several cases extended beyond the intervention period. One participant reported that he still followed certain guidelines and remained motivated, having "realised that my wellbeing has improved a lot and that I have much more energy than before" (P2; man, aged 63 years, baseline SCORE2 risk 10%). Another described a gradual rather than an abrupt change: "I changed the way I eat, especially sweets and fructose. I started to be very selective about what I eat. I also started going to the swimming pool and try to maintain walking as much as I can. The change was not sudden, but gradually I changed my diet a lot" (P4; woman, aged 58 years, baseline SCORE2 risk 5.8%). Nurses did not report outcomes of this kind, which they were not in a position to observe; their contribution was an account of why the tool might be motivating, namely, that it renders the room for improvement visible and concrete rather than abstract. Nurses were also more cautious than patients about how far this effect would reach, anticipating that motivation would be easier to elicit in younger, digitally confident patients than in older ones, a reservation examined in the next category.

Category 4: Identification of Problems and Challenges and Recommendations for Improvement

Problems and Deviations

Implementation barriers included fear or mistrust of technology, the risk of negative interpretation, usability challenges for older users, affordability (especially for CGM), and the time burden of conventional tools. The barriers described by the 2 groups were largely distinct, reflecting their different positions in relation to the intervention. Patients described difficulties arising from continuous self-monitoring. These ranged from an initial apprehension about what the results would show, "I was a little scared of reality...maybe it is a little scary to compare it with reality, but on the other hand I would say it is more positive" (P1; woman, aged 67 years, baseline SCORE2 risk 4%), to the practical and attentional burden of wearing and following the sensor: "Constant monitoring was at times somewhat burdensome and stressful, because I was paying attention to the numbers the whole time...because my work is very active and I usually do not have my phone with me, it was a burden to always watch where my phone was so as not to lose the signal with the sensor" (P2; man, aged 63 years, baseline SCORE2 risk 10%). A related concern was the fear of overlooking something (P3; woman, aged 58 years, baseline SCORE2 risk 7.8%), and one participant described the monitoring becoming compulsive before settling: "it can quickly become an obsession. When I was wearing the glucose device, during the first few days I was constantly checking it. It was almost like a teenager who cannot get away from the phone. You have to find a balance" (P4; woman, aged 58 years, baseline SCORE2 risk 5.8%). Importantly, these accounts were not offered as reasons to reject the technology; the same participants also described the monitoring as motivating, and the difficulty they identified was one of calibration rather than of value.

The nurses' concerns lay elsewhere, in feasibility and equity. They contrasted the tool favorably with the existing paper-based procedure, noting that with "SCORE2 tables, you need a lot of time and it is more likely that you make a mistake and then start over. Here it is much simpler and faster" (N5; ≥5 years of professional experience), but identified the digital divide as the principal limitation on who could benefit: "For the older population, it may be harder to motivate them to use this tool, whereas younger people, who are more digitally advanced, would probably understand more easily why it is useful" (N1; ≥5 years of professional experience). Nurses further raised the affordability and accessibility of CGM, which is not routinely reimbursed for people with no diabetes, and the risk that a patient might interpret a display negatively in the absence of professional support. Taken together, the patient-reported and nurse-reported barriers point in different directions for implementation: the former concerns how continuous monitoring is introduced and framed for the individual, whereas the latter concerns whether the approach can be delivered equitably within routine services.

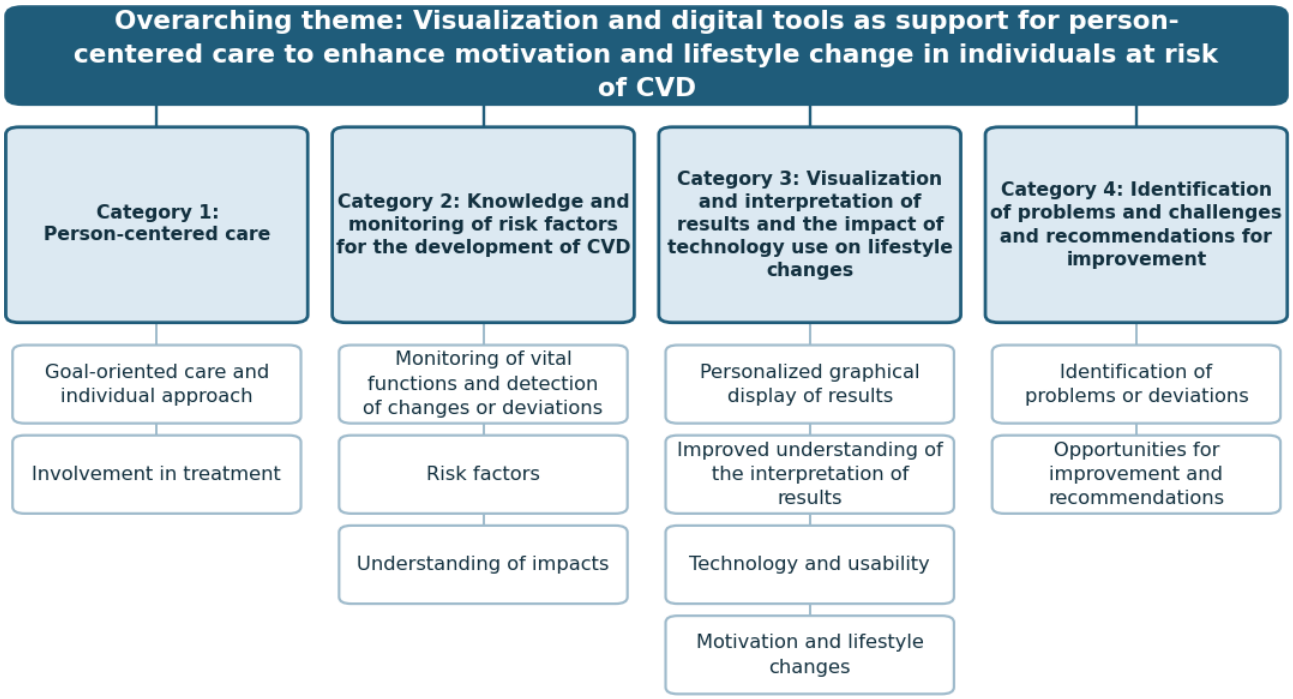
Opportunities for Improvement

Suggested improvements focused on training and support, avoiding information overload, preserving the intuitive visual metaphor (flower), and extending use to other clinical areas. The suggestions made by each group followed from the barriers they had identified. Patients proposed extending the scope of what is measured and consolidating it in one place: "Maybe in the future it would be possible to measure something else, not only glucose. It would be useful if the same sensor or application could also include movement, steps and other aspects, so that everything would be visualized in one application" (P4; woman, aged 58 years, baseline SCORE2 risk 5.8%). The nurses' recommendations were oriented toward wider adoption and the conditions for it, encompassing training and support for both staff and patients, the avoidance of information overload, and the preservation of the intuitive flower metaphor, which they regarded as central to the accessibility of the tool. Several also saw application beyond cardiovascular prevention: "I think this is a big bright spot for the future. It could also be used in other areas of health care, or in other fields, to motivate patients" (N2; ≥5 years of professional experience). Across the 2 groups, the recommendations were complementary: patients focused on the completeness and integration of the information they receive, and nurses focused on the organizational preconditions for using that information in routine practice.

Figure 1 presents a conceptual map of the categories and subcategories identified in the analysis, together with the overarching theme that runs across them. The map summarizes the descriptive structure of the findings; the layout does not imply a causal, temporal, or hierarchical sequence between the categories.

XSL•FO
RenderX

Figure 1. Conceptual map of the categories and subcategories describing how patients and nurses experienced the Petal-X visualization and continuous glucose monitoring. Categories and subcategories are shown as a descriptive structure; the layout does not imply a causal, temporal, or hierarchical sequence between categories. CVD: cardiovascular disease.



Relation to the Parallel Quantitative Study

This paper reports the qualitative phase of a sequential mixed methods study; the quantitative randomized controlled trial is reported separately. We did not statistically integrate the 2 datasets in this paper. Instead, consistent with sequential mixed methods designs, the qualitative findings are connected to the trial at the level of interpretation: they help explain the mechanisms that may underlie the trial results, namely, greater understanding, involvement, self-perceived control, and motivation supported by visualization and real-time feedback. The literature review informed the selection of technologies and risk factors incorporated into the intervention.

Discussion

Principal Findings

In this qualitative study, both nurses and patients experienced interactive CVD risk visualization (Petal-X), particularly when combined with CGM, as a meaningful support for PCC. Across 4 categories and 11 subcategories, participants described a clearer understanding of risk factors, more individualized and shared consultations, a greater sense of involvement and control, and stronger motivation for dietary and physical activity change, which some patients reported maintaining after the study. At the same time, participants and nurses highlighted important barriers, including variable digital literacy among older users, fear or mistrust of the technology, the burden of continuous self-monitoring, potential information overload, workflow and training needs, and the affordability and accessibility of CGM. Qualitative analysis of interviews with nurses and participants reveals key aspects of PCC that were identified during the introduction of new technologies in the family medicine practice. The participants expressed satisfaction with the individualized approach, which includes adapting care to the needs and wishes of individuals and actively involving patients in the treatment process.

The perceptions of nurses and participants regarding the use of cardiovascular risk visualization (Petal-X) and CGM are crucial for understanding how effectively these technologies improve management and motivation for lifestyle changes. The categories that enable us to answer the research question are the following: goal-oriented care and an individual approach, involvement in treatment, monitoring of vital functions and technology, motivation and lifestyle changes, understanding of impacts and risk factors, and opportunities for improvement. The category that most strongly organized participants' accounts was PCC.

PCC, supported by technologies that enable personalization and visualization of health data, emerged in our study as an effective strategy for encouraging lifestyle changes among patients in the family medicine setting. Such an approach not only enhances the quality of care but also fosters collaboration between patients and health care professionals, ultimately contributing to improved long-term health outcomes. These findings are consistent with previous interventions that have focused on patient education, structured follow-up and enhanced support for self-management, such as programs for disease self-management education [28,29], structured discharge plans with tailored follow-up [30], and lifestyle counseling, home visits, and telephone support led by health care professionals [31].

Our results suggest that interventions that combine education, personalized care plans, and lifestyle adaptation to the individual's cardiovascular risk could be further strengthened by incorporating visual risk displays [9,32-35], mobile apps [36-39], and additional biometric measurements. CVDs remain a leading cause of morbidity and mortality worldwide, and early

XSL•FO
RenderX

identification and engagement of at-risk individuals are crucial for effective prevention. In this context, digital tools that translate complex prognostic information into intuitive, patient-friendly formats align strongly with the principles of PCC and offer a promising route to more effective preventive strategies.

A central finding of our qualitative analysis is that goal-oriented care and an individualized, data-driven approach are key to tailoring care for patients at risk of CVD. Participants consistently reported that they experienced the consultation as more holistic, with more time for conversation and shared decision-making about their health. This tailored approach took into account specific needs, values, and goals, and included adjustments to diet, physical activity, and daily routines, often with the support of family members. New forms of visualization, such as interactive graphs and personalized analyses of risk factors, were perceived as particularly helpful. They improved communication, enhanced understanding of patients' health status, and encouraged them to participate actively in planning their care.

The Petal-X visualization dashboard, developed and used in our study, represents a concrete step toward augmenting SCORE2 with an explanation layer and operationalizing PCC in routine practice. By providing a clear, structured overview of individual CVD risk factors, Petal-X made the risk profile more transparent and easier to interpret for both patients and nurses. Visualizations increased the sense of overview and control among participants, allowing them to see how different risk factors contribute to their overall risk and what might change if they modify certain behaviors [9,27,32,34,35]. Nurses emphasized that this approach improved the quality of consultations, simplified workflows, and made it easier to involve patients in setting realistic, personalized goals, findings that echo positive outcomes reported in other technology-supported interventions [40-43].

In addition to visualization, the use of CGM, typically reserved for people with diabetes, introduced a dynamic and highly engaging element to lifestyle counseling. Participants carried the CGM device with them and were able to observe real-time changes in their glucose levels in response to food intake, physical activity, stress, and other daily situations. This immediate feedback created a direct link between behavior and physiological response, which several participants identified as a strong motivator for changing and maintaining healthier habits. Similar benefits of real-time monitoring and digital feedback loops have been documented in studies using mobile apps and wearable technologies to support self-management and lifestyle modification [28,29,36-39,41].

Overall, our qualitative findings indicate improvements in patients' understanding of their health status, awareness of CVD risk, satisfaction with the integration of technologies, and willingness to participate in planning their lifestyle changes. Participants reported an increased motivation to continue monitoring and adjusting their lifestyle even after the study ended. From the perspective of nurses, the combined use of Petal-X and CGM supported more structured, person-centered consultations and facilitated the integration of preventive interventions into everyday practice. These experiences are consistent with evidence from other technology-enhanced interventions, which have demonstrated positive outcomes when new tools are embedded in supportive, PCC models [42-44].

The Slovenian context adds an important system-level dimension to these findings. Currently, there is no standardized digital tool that integrates multiple solutions for both nurses and patients, nor is there a unified national electronic health record system that would seamlessly support such approaches. Our implementation of Petal-X and CGM within this fragmented digital environment underscores the potential of these tools to enhance care even in the absence of a comprehensive national infrastructure. At the same time, it underscores the necessity for national guidelines on integrating new technologies into health care, particularly in the areas of prevention and chronic disease management.

Finally, lifestyle modification remains one of the most important modifiable risk factors for reducing the prevalence of CVD risk, making research and innovation in this area especially valuable in Slovenia. Our results suggest that lifestyle change is closely linked to patients' perception of PCC, especially the experience of an individualized approach, personalized visualization of risk, and a holistic, respectful consultation process. The inclusion of new technologies in lifestyle interventions can increase patient motivation, enable real-time monitoring of health status, and support higher-quality PCC. Further research should explore how such tools can be scaled up, integrated into national digital ecosystems, and combined with established interventions, such as education, structured follow-up, and community-based support [28,30,31], to strengthen prevention and long-term management of CVD risk.

Future Research

Several directions for future research follow from these findings. First, larger studies across multiple primary care settings are needed to test whether the perceived benefits of risk visualization and CGM extend to more diverse populations and to evaluate how the tools can be scaled and embedded in routine workflows and in national digital health infrastructure, including electronic health record systems. Second, because CGM carries nontrivial costs, future work should include an affordability and risk-benefit analysis from a public health perspective, so that decisions about wider adoption weigh the added value of continuous monitoring against its cost and its equity implications for patients who cannot afford such devices. Third, given the focus on secondary prevention and PCC, future studies should more explicitly incorporate patients' own preferences, needs, and values, for example, through a structured "What Matters to You?" approach, so that digitally supported risk communication is aligned with what patients themselves consider important. Finally, longer follow-up is needed to determine whether the motivation and behavioral changes that participants described translate into the long-term sustainability of lifestyle change and into measurable cardiovascular risk reduction over time.

Limitations

Several limitations should be considered when interpreting these findings. First, this was a qualitative descriptive study based on

XSL•FO
RenderX

semistructured interviews with a small sample of nurses (n=6) and intervention participants (n=6) recruited by total population sampling from a closed, predefined trial cohort, which supports depth of description but limits transferability to broader primary care populations and settings. Second, although recruitment was open across Slovenia, the promotional campaign and implementation were concentrated in a single Community Health Centre in central Slovenia, so local organizational routines and context may have shaped participants' experiences and perceptions. In addition, recruitment from a single center, combined with the small sample size, resulted in a relatively homogeneous sample, which may further limit the transferability of the findings. Third, the study relied on self-reported experiences, which may be influenced by recall bias and social desirability. Finally, perceptions of Petal-X and CGM may reflect novelty effects and practical implementation constraints; importantly, participants and nurses also reported barriers such as variable digital literacy (particularly among older users), fear or mistrust of technology, potential information overload, workflow integration needs, and affordability or access limitations of CGM, which may affect scalability and equity of implementation.

A further limitation, and one that follows directly from the sampling strategy described, concerns who was eligible to be interviewed. Both samples were restricted to individuals who had completed their part of the intervention: patients who had completed the intervention arm, as defined in the Methods section, and nurses who had completed the familiarization with the intervention and remained in the preventive service throughout the trial period. Patients who withdrew, who discontinued sensor wear, or who did not attend the final assessment were by definition unavailable for interview, and their reasons for discontinuing, which may include precisely the burdens, anxieties, and usability problems that an intervention of this kind is most likely to generate, are therefore not represented in our data. Completer-only sampling plausibly selects for higher engagement with digital tools, greater tolerance of continuous self-monitoring and of the practical and attentional demands of wearing a sensor, and a degree of comfort with self-quantification that may not be typical of the wider primary prevention population. The barriers reported here, including the initial apprehension about what the results would show, the compulsive checking described by one participant, and the burden of keeping a smartphone within range of the sensor, should therefore be read as difficulties reported by people who nevertheless persisted to the end of the intervention, and they are likely to understate both the frequency and the intensity of the problems that would arise in an unselected population.

A related but distinct consideration applies to the nurses. As described in the Methods section, all trial consultations were delivered by the first author (AS), and the nurses interviewed had been familiarized with the intervention and the Petal-X dashboard rather than having used them with their own caseload in routine practice. Their appraisals of usability, workflow fit, and feasibility are therefore the anticipatory judgments of informed prospective users rather than reports of sustained delivery, and they may not capture the frictions that would emerge once such a tool is embedded in everyday consultations, nor the adaptations that practitioners make over time. Equally, the nurses who agreed to be interviewed were those who had engaged with the familiarization offered at the center, which may again select for interest in digital tools. Research in which nurses themselves deliver visualization-supported consultations over a sustained period is needed before conclusions about routine implementation can be drawn. It should also be noted that the interviewed patients were somewhat older than the intervention arm as a whole (mean 57.8 vs 51.9 years), so that the older group whose digital literacy the nurses identified as a barrier was, if anything, overrepresented among the interviewees, although these were older adults who had nonetheless completed the intervention successfully. Studies that deliberately sample noncompleters, partial users, and individuals who decline such interventions are needed to characterize the full range of responses to visualization-supported and CGM-supported cardiovascular prevention.

Conclusions

This qualitative study demonstrates that interactive CVD visualization (Petal-X), particularly when combined with real-time glucose feedback via CGM, can enhance key elements of PCC in primary prevention. Both patients and nurses described that those personalized visual displays improved comprehension of risk factors, supported shared decision-making, and increased patients' sense of involvement and control, mechanisms that were perceived as essential drivers of motivation and lifestyle change.

At the same time, implementation challenges were identified, including variable digital literacy (particularly among older users), concerns about information overload, the need for training of health care professionals, and affordability and accessibility of devices such as CGM. Future work should evaluate how these tools can be integrated into routine workflows and national digital infrastructure, and how they can be scaled in a way that preserves usability, equity, and sustained lifestyle benefits.

Acknowledgments

The authors want to thank all participants who took part in this research. The authors declare the use of generative artificial intelligence (GAI) in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GAI tools under full human supervision: translation and reformatting. The GAI tool used was Chat GPT 5.1. Responsibility for the final manuscript lies entirely with the authors. GAI tools are not listed as authors and do not bear responsibility for the outcomes.

XSL•FO
RenderX

Funding

This work was partially funded from the Slovenian Research Agency (grant numbers N3- 0307 and P2-0057) and was partially funded by the European Union's Horizon Europe research and innovation program under the grant agreement number 101159018. Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Data Availability

The qualitative data generated and analyzed during this study are not publicly available because they consist of individual interview transcripts that contain potentially identifying information, and participants did not consent to public sharing of the full transcripts. Deidentified excerpts supporting the findings are included within the paper. Further deidentified data are available from the corresponding author on reasonable request, subject to approval by the relevant ethics committee and a data-sharing agreement.

Authors' Contributions

Conceptualization: GS, AS, ML, ZK-K
Formal analysis: LG, GS, AS
Investigation: AS
Methodology: AS, KV, DR, GS
Writing – original draft: AS, KV, DR, GS
Writing – review & editing: GS, AS, ML, ZK-K, LG, DR, KV

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide for registered nurses.
[DOCX File, 16 KB - humanfactors_v13i1e92101_app1.docx]

Multimedia Appendix 2
Interview guide for patients.
[DOCX File, 16 KB - humanfactors_v13i1e92101_app2.docx]

Checklist 1
COREQ checklist.
[DOCX File, 21 KB - humanfactors_v13i1e92101_app3.docx]

References

1. Kovačec S, Klemenc-Ketiš Z, Poplas-Susič A, Kravos A. Experience and views of primary care physicians involved in reorganisation of care in family medicine practices during COVID-19 pandemic: a qualitative study from Slovenia. Eur J Gen Pract 2023 Dec;29(2):2193886. [doi: 10.1080/13814788.2023.2193886] [Medline: 37013826]
2. Maučec Zakotnik J, Čuš B, Vrbovšek S, Govc Eržen J, Medved N. Programme of integrated prevention of chronic non-communicable diseases in adults—description of the preventive pathway of the individual being assessed. In: Govc Eržen J, Petek Šter M, editors. Handbook for Family Physicians: Implementing Integrated Prevention of Chronic Non-Communicable Diseases in Model Family Medicine Practices: National Institute of Public Health; 2017:10-16 URL: http://skupajzazdravje.nijz.si/media/splet-.zdm.prirocnik.pdf
3. Petek Šter M, Barbič Žagar B. Treatment of patients with arterial hypertension in family medicine in Slovenia – assessment of the situation after the introduction of model practices. ZdravVestn 2021;90(9-10):478-487. [doi: 10.6016/ZdravVestn.3069]
4. Deng W, Wang D, Wan Y, Lai S, Ding Y, Wang X. Prediction models for major adverse cardiovascular events after percutaneous coronary intervention: a systematic review. Front Cardiovasc Med 2023;10:1287434. [doi: 10.3389/fcvm.2023.1287434] [Medline: 38259313]
5. Wolbers M, Koller MT, Witteman JCM, Steyerberg EW. Prognostic models with competing risks: methods and application to coronary risk prediction. Epidemiology 2009 Jul;20(4):555-561. [doi: 10.1097/EDE.0b013e3181a39056] [Medline: 19367167]
6. Hageman S, Pennells L, Ojeda F, et al. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. Eur Heart J 2021 Jul 1;42(25):2439-2454. [doi: 10.1093/eurheartj/ehab309]

XSL•FO
RenderX

7. Badawy M, Naing L, Johar S, et al. Evaluation of cardiovascular diseases risk calculators for CVDs prevention and management: scoping review. BMC Public Health 2022 Sep 14;22(1):1742. [doi: 10.1186/s12889-022-13944-w] [Medline: 36104666]
8. Bansal M, Kasliwal RR, Trehan N. Comparative accuracy of different risk scores in assessing cardiovascular risk in Indians: a study in patients with first myocardial infarction. Indian Heart J 2014;66(6):580-586. [doi: 10.1016/j.ihj.2014.10.399] [Medline: 25634388]
9. Svenšek A, Lorber M, Gosak L, Verbert K, Klemenc-Ketis Z, Stiglic G. The role of visualization in estimating cardiovascular disease risk: scoping review. JMIR Public Health Surveill 2024 Oct 14;10:e60128. [doi: 10.2196/60128] [Medline: 39401079]
10. Kaminsky LA, German C, Imboden M, Ozemek C, Peterman JE, Brubaker PH. The importance of healthy lifestyle behaviors in the prevention of cardiovascular disease. Prog Cardiovasc Dis 2022;70:8-15. [doi: 10.1016/j.pcad.2021.12.001] [Medline: 34922952]
11. Sebastian SA. Cardiovascular disease risk communication: strategies, impact, and future directions. Curr Probl Cardiol 2024 May;49(5):102490. [doi: 10.1016/j.cpcardiol.2024.102490] [Medline: 38417473]
12. Murthi S, Martini N, Falconer N, Scahill S. Evaluating EHR-integrated digital technologies for medication-related outcomes and health equity in hospitalised adults: a scoping review. J Med Syst 2024 Aug 23;48(1):79. [doi: 10.1007/s10916-024-02097-5] [Medline: 39174723]
13. Vallée A. Digital twin for healthcare systems. Front Digit Health 2023;5:1253050. [doi: 10.3389/fdgth.2023.1253050] [Medline: 37744683]
14. Mansour M, Saeed Darweesh M, Soltan A. Wearable devices for glucose monitoring: a review of state-of-the-art technologies and emerging trends. Alex Eng J 2024 Feb;89:224-243. [doi: 10.1016/j.aej.2024.01.021]
15. Svenšek A, Gosak L, Kopitar L, Štiglic G. Using a continuous blood glucose monitor as a form of remote patient care. : Sustainable Digital Future of Healthcare: MI'23 Professional Meeting URL: https://plus-legacy.cobiss.net/cobiss/si/sl/bib/172101123 [accessed 2026-01-23]
16. Jansen J, McKinn S, Bonner C, Muscat DM, Doust J, McCaffery K. Shared decision-making about cardiovascular disease medication in older people: a qualitative study of patient experiences in general practice. BMJ Open 2019 Mar 20;9(3):e026342. [doi: 10.1136/bmjopen-2018-026342] [Medline: 30898831]
17. Cowap L, Riley V, Grogan S, et al. "They are saying it's high, but I think it's quite low": exploring cardiovascular disease risk communication in NHS health checks through video-stimulated recall interviews with patients - a qualitative study. BMC Prim Care 2024 Apr 23;25(1):126. [doi: 10.1186/s12875-024-02357-w] [Medline: 38654245]
18. Rojo D, Lamqaddam H, Gosak L, Verbert K. Petal-X: human-centered visual explanations to improve cardiovascular risk communication. arXiv. Preprint posted online on Jun 26, 2024. [doi: 10.48550/arXiv.2406.18690]
19. Creswell JW. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches: SAGE Publications; 2014.
20. Svenšek A, Lorber M, Klemenc-Ketiš Z, Gosak L, Muršec D, Štiglic G. Interactive digital visualization counseling for lifestyle change in patients at risk of cardiovascular diseases: randomized controlled trial. JMIR Public Health Surveill 2026 Jan 22;12:e83488. [doi: 10.2196/83488] [Medline: 41569629]
21. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g] [Medline: 10940958]
22. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
23. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
24. Lincoln YS, Guba EG. Naturalistic Inquiry: SAGE Publications; 1985.
25. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
26. Hennink MM, Kaiser BN, Marconi VC. Code saturation versus meaning saturation: how many interviews are enough? Qual Health Res 2017 Mar;27(4):591-608. [doi: 10.1177/1049732316665344] [Medline: 27670770]
27. Svenšek A. Using the visualization for lifestyle change in patients at risk of cardiovascular diseases. : ClinicalTrials.gov; 2024 URL: https://clinicaltrials.gov/study/NCT06226948 [accessed 2026-10-05]
28. Fors A, Ekman I, Taft C, et al. Person-centred care after acute coronary syndrome, from hospital to primary care - a randomised controlled trial. Int J Cardiol 2015;187:693-699. [doi: 10.1016/j.ijcard.2015.03.336] [Medline: 25919754]
29. Iturralde E, Sterling SA, Uratsu CS, Mishra P, Ross TB, Grant RW. Changing results-engage and activate to enhance wellness: a randomized clinical trial to improve cardiovascular risk management. J Am Heart Assoc 2019 Dec 3;8(23):e014021. [doi: 10.1161/JAHA.119.014021] [Medline: 31787053]
30. Averbuch T, Lee SF, Zagorski B, et al. Effect of a transitional care model following hospitalization for heart failure: 3-year outcomes of the Patient-Centered Care Transitions in Heart Failure (PACT-HF) randomized controlled trial. Eur J Heart Fail 2024 Mar;26(3):652-660. [doi: 10.1002/ejhf.3134] [Medline: 38303550]

XSL•FO
RenderX

31. Fors A, Blanck E, Ali L, et al. Effects of a person-centred telephone-support in patients with chronic obstructive pulmonary disease and/or chronic heart failure - a randomized controlled trial. PLoS ONE 2018;13(8):e0203031. [doi: 10.1371/journal.pone.0203031] [Medline: 30169539]
32. Afzal S, Ghani S, Hittawe MM, et al. Visualization and visual analytics approaches for image and video datasets: a survey. ACM Trans Interact Intell Syst 2023 Mar 31;13(1):1-41. [doi: 10.1145/3576935]
33. Svenšek A, Gosak L, Lorber M, Štiglic G, Fijačko N. Review and comparative evaluation of mobile apps for cardiovascular risk estimation: usability evaluation using mHealth app usability questionnaire. JMIR Mhealth Uhealth 2025 May 8;13:e56466. [doi: 10.2196/56466] [Medline: 40341099]
34. Theis S, Mertens A, Schlick CM. Preliminary results of an exploratory study towards a general task and data model for telemedical visualizations. Presented at: Conference: 19thTriennial Congress of the International Ergonomics Association; Aug 9-14, 2015; Melbourne p. 1-6 URL: https://publications.rwth-aachen.de/record/659041 [accessed 2026-02-25]
35. Turchioe MR, Myers A, Isaac S, et al. A systematic review of patient-facing visualizations of personal health data. Appl Clin Inform 2019 Aug;10(4):751-770. [doi: 10.1055/s-0039-1697592] [Medline: 31597182]
36. Fijačko N, Masterson Creber R, Gosak L, et al. A review of mortality risk prediction models in smartphone applications. J Med Syst 2021 Nov 4;45(12):107. [doi: 10.1007/s10916-021-01776-x] [Medline: 34735603]
37. Lee C, Lee K, Lee D. Mobile healthcare applications and gamification for sustained health maintenance. Sustainability 2017;9(5):772. [doi: 10.3390/su9050772]
38. Lohnari T, Patil S, Patil S. Use of mobile applications in healthcare: a review. Int J Eng Res Gen Sci 2016;4:38-42 [FREE Full text]
39. Moungui HC, Nana-Djeunga HC, Anyiang CF, Cano M, Ruiz-Postigo JA, Carrion C. Dissemination of mobile health applications: a systematic review. JMIR Preprints. Preprint posted online on Jun 26, 2023. [doi: 10.2196/preprints.50293]
40. Bonner C, Batcup C, Fajardo M, Trevena L. Biological age calculators to motivate lifestyle change: environmental scan of online tools and evaluation of behaviour change techniques. Health Promot J Austr 2023 Feb;34(1):202-210. [doi: 10.1002/hpja.671] [Medline: 36198168]
41. Lobelo F, Kelli HM, Tejedor SC, et al. The Wild Wild West: a framework to integrate mHealth software applications and wearables to support physical activity assessment, counseling and interventions for cardiovascular disease risk reduction. Prog Cardiovasc Dis 2016;58(6):584-594. [doi: 10.1016/j.pcad.2016.02.007] [Medline: 26923067]
42. Oliver JJ, Streitz MJ, Hyams JM, et al. The HEART score as a prognostic tool for revascularization. Intern Emerg Med 2020 Jun;15(4):607-612. [doi: 10.1007/s11739-019-02206-0] [Medline: 31625076]
43. Stiglic G, Kocbek P, Fijacko N, Zitnik M, Verbert K, Cilar L. Interpretability of machine learning‐based prediction models in healthcare. WIREs Data Min Knowl 2020 Sep;10(5). [doi: 10.1002/widm.1379]
44. Bonner C, Batcup C, Cornell S, et al. Interventions using heart age for cardiovascular disease risk communication: systematic review of psychological, behavioral, and clinical effects. JMIR Cardio 2021 Nov 5;5(2):e31056. [doi: 10.2196/31056] [Medline: 34738908]

Abbreviations

CGM: continuous glucose monitoring
COREQ: Consolidated Criteria for Reporting Qualitative Research
CVD: cardiovascular disease
PCC: person-centered care
RN: registered nurse
SCORE2: Systematic Coronary Risk Evaluation 2

Edited by S Law; submitted 24.Jan.2026; peer-reviewed by HY Goh, Y Wei; revised version received 08.Sep.2026; accepted 08.Sep.2026; published 13.Oct.2026.

Please cite as:
Svenšek A, Rojo D, Lorber M, Klemenc-Ketis Z, Gosak L, Verbert K, Štiglic G
Risk Visualization With Petal-X and Continuous Glucose Monitoring to Support Person-Centered Care and Lifestyle Change in Cardiovascular Prevention: Qualitative Descriptive Study
JMIR Hum Factors 2026;13:e92101
URL: https://humanfactors.jmir.org/2026/1/e92101
doi:10.2196/92101

© Adrijana Svenšek, Diego Rojo, Mateja Lorber, Zalika Klemenc-Ketis, Lucija Gosak, Katrien Verbert, Gregor Štiglic. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 13.Oct.2026. This is an open-access article distributed under

XSL•FO
RenderX

the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Real-World Health Monitoring Using Wrist-Worn Wearables in Community-Dwelling Older Adults in China: Mixed Methods Feasibility Study

Ying Bai[1], PhD; Yi Tang[1,2], MD, PhD; Yiwei Zhao[1], MS; Lixin Ma[3], MCM; Yanjun Ma[4], PhD; Zhibin Wang[1], PhD

[1]Department of Neurology & Innovation Center for Neurological Disorders, Xuanwu Hospital, Capital Medical University, National Center for Neurological Disorders, 45 Changchun Street, Xicheng District, Beijing, China

[2]Beijing Key Laboratory of Digital Medicine for Cognitive Disorders, Beijing, China

[3]Department of Neurology, The First Hospital of Hebei Medical University, Shijiazhuang, Hebei, China

[4]Xuanwu Hospital, Capital Medical University, National Center for Neurological Disorders, Beijing, China

Corresponding Author:
Yi Tang, MD, PhD
Department of Neurology & Innovation Center for Neurological Disorders, Xuanwu Hospital, Capital Medical University, National Center for Neurological Disorders, 45 Changchun Street, Xicheng District, Beijing, China

Abstract

Background: Real-world behavioral and physiological functioning, including heart rate, daily mobility, and sleep patterns, may provide information about aging and early neurodegenerative changes that is not captured by periodic clinical visits. Wrist-worn wearable devices offer a promising approach for continuous, passive monitoring of these domains. However, evidence regarding feasibility, usability, and acceptability in real-world observational settings among community-dwelling older adults in China remains limited.

Objective: This study primarily aimed to evaluate the feasibility, usability, and acceptability of using wrist-worn wearable devices to monitor behavioral and physiological function in community-dwelling older adults under naturalistic conditions. As a secondary objective, the study descriptively characterized wearable-derived behavioral and physiological measures among participants with minimal consecutive valid monitoring.

Methods: We enrolled 486 community-dwelling older adults from the Beijing region of the China-Aging cohort, each of whom received a Huawei Band 8 commercial wrist-worn wearable device. A valid monitoring day was defined as 10 or more hours of wear time. Participants were classified into three monitoring-pattern categories based on predefined wear-time criteria: no valid monitoring day, nonconsecutive valid monitoring, and minimal consecutive valid monitoring (≥3 consecutive valid monitoring days). Demographic characteristics, health status, and cognitive and functional measures (Mini-Mental State Examination and Barthel Activities of Daily Living Index) were compared across groups. Usability, acceptability, and perceived barriers were assessed using a structured user experience survey.

Results: Of 486 participants, 224 (46.1%) achieved minimal consecutive valid monitoring, 156 (32.1%) had nonconsecutive valid monitoring, and 106 (21.8%) had no valid monitoring day. Self-rated health differed significantly across monitoring patterns (P=.005), whereas no significant differences were observed for the other measured baseline characteristics. Among participants with minimal consecutive valid monitoring, mean daily heart rate was 72.1 (SD 8.6) bpm, median daily step count was 6177 (IQR 4249 - 8872) steps/day, and sleep data were available for 211 out of 224 (94.2%), with a median sleep duration of 7.23 (IQR 6.48 - 8.04) hours. The user experience survey was completed by 106 out of 486 (21.8%) participants, with higher participation among those with more sustained monitoring. Respondents generally reported favorable usability and acceptability; common barriers included device discomfort, data synchronization issues, and battery life.

Conclusions: Wrist-worn wearable monitoring was feasible under naturalistic conditions for a substantial proportion of community-dwelling older adults in China, with generally favorable usability and acceptability among survey respondents. However, monitoring patterns varied considerably across individuals, highlighting challenges for long-term use and large-scale implementation. These findings support the potential of wearable technologies for longitudinal monitoring outside clinic-based assessments. Further research is needed to establish the stability and clinical relevance of wearable-derived behavioral measures over longer monitoring periods.

(JMIR Hum Factors 2026;13:e98592) doi:10.2196/98592

KEYWORDS

wearable devices; older adults; feasibility; usability; acceptability; digital health; China

XSL•FO
RenderX

Introduction

Population aging is a major global demographic trend that is reshaping public health priorities and increasing demands on health care systems worldwide [1]. As life expectancy increases, a growing proportion of older adults experience age-related declines in functional ability, reflecting gradual changes in physical and cognitive capacities that are essential for maintaining independent living [2]. Functional decline is strongly associated with increased morbidity, mortality, health care utilization, and societal burden [3,4].

Functional deterioration in later life rarely occurs abruptly. Accumulating evidence indicates that subtle and progressive changes in daily-life behaviors, such as alterations in walking patterns [5], physical activity levels [6], and sleep characteristics [7,8], often precede clinically detectable impairment by several years. Longitudinal and meta-analysis studies further suggest that increased variability in routine behaviors, particularly sleep and activity patterns, is associated with emerging cognitive and functional vulnerability, even in the absence of overt clinical symptoms [8-10].

Despite growing recognition of early functional change, current assessment approaches remain largely limited to episodic clinic-based assessments and self-reported measures [11]. Standard assessments typically rely on brief performance tests administered during scheduled clinical visits, retrospective questionnaires, or caregiver reports, providing only snapshots of functioning at isolated time points [11,12]. As a result, these approaches are limited in their ability to capture gradual trajectories of functional change, particularly during preclinical or early stages when deviations from individual baselines are subtle, heterogeneous, and context-dependent [12,13]. Moreover, self-reported and proxy-reported measures are vulnerable to recall bias and reporting inaccuracies, constraining sensitivity to small but clinically meaningful behavioral alterations [14,15].

Wearable sensing technologies have emerged as a promising approach to addressing these limitations. Wrist-worn devices can passively and continuously capture daily behaviors, including physical activity [16], sleep [17], and heart rate dynamics [18], using embedded accelerometers and photoplethysmography (PPG) sensors with minimal user burden [16,19]. Beyond passive data collection, accumulating evidence suggests that these continuously monitored behaviors carry prognostic significance [5,7,10].

Wearable-derived behavioral and physiological patterns show strong clinical associations in older adults. Lower daily step counts, reduced activity, and more sedentary profiles are associated with higher all-cause mortality and cardiovascular risk [4,20-22]. Similarly, sleep characteristics, including shorter duration, greater night-to-night variability, and longitudinal deterioration, are associated with subsequent cognitive impairment and neurodegenerative risk [7,8,23]. Recent work suggests that real-world behavioral measures derived from wrist-worn devices, including walking speed, activity patterns, and sleep timing, can predict incident dementia years before clinical diagnosis [5]. In particular, gait-related measures have been developed as digital biomarkers [24] and can distinguish mild Alzheimer disease from cognitively normal controls in community settings [25]. Beyond observational associations, wearable technologies are increasingly being applied in intervention studies and structured remote monitoring programs in clinical research [26-29]. Together, these findings support the potential value of wearable-derived measures for longitudinal health monitoring in older adults. However, their use depends in part on whether adequate monitoring can be achieved under real-world conditions [30,31].

However, important gaps remain in understanding real-world feasibility among older adults. Most feasibility studies have been conducted over short durations or under controlled conditions, often involving intensive researcher support, structured onboarding, frequent follow-up, or incentives [16,32]. Study samples are also frequently highly selected, excluding individuals with lower digital literacy, comorbidities, or functional limitations, thereby limiting generalizability [30,33]. While such approaches are appropriate for early-stage validation, they may not reflect feasibility under naturalistic, everyday conditions, where ongoing researcher support may be limited and users vary in their capabilities and support needs [32,34]. Additionally, feasibility data from Chinese community settings remain particularly limited, despite China's rapidly aging population and increasing policy interest in digital health monitoring for older adults.

Accordingly, this study aimed to evaluate the feasibility, usability, and acceptability of wearable monitoring among community-dwelling older adults under naturalistic conditions by examining (1) the initiation and maintenance of wearable monitoring, (2) the proportion of participants achieving the predefined minimum criterion for consecutive valid monitoring, (3) measured participant characteristics associated with different wearable monitoring patterns, and (4) usability experiences and perceived barriers reported by participants who completed the user experience survey.

Methods

Study Design

This study was conducted within the China Healthy Aging Cohort Study (China-Aging; ClinicalTrials.gov identifier: NCT06472674), an ongoing longitudinal, multicenter cohort of community-dwelling older adults across several regions in China [35]. The China-Aging study aims to investigate the prevalence and risk factors of disability and cognitive impairment and to develop predictive models by integrating traditional clinical measures with technology-derived data.

The wearable monitoring component was established within the Beijing center of the China-Aging cohort to collect longitudinal, real-world behavioral and physiological data using wrist-worn wearable devices. The present study reports the feasibility, usability, and acceptability of this monitoring approach under naturalistic conditions. A mixed methods design was used, combining quantitative wearable monitoring data with user experience survey data, including open-ended responses analyzed thematically. Baseline demographic and

XSL•FO
RenderX

clinical assessments were conducted by trained researchers as part of the China-Aging cohort procedures.

Ethical Considerations

The study was approved by the Ethics Committee of Xuanwu Hospital, Capital Medical University (approval number 2024214001), and all participants provided written informed consent prior to participation.

Participants

Participants were recruited from the Beijing region of the China-Aging cohort through the Baizhifang Community Health Centre. Eligible participants were invited to take part in this wearable monitoring study after completing baseline assessments within the cohort.

Inclusion criteria were (1) age 55 years or above, (2) residence in the local community for at least 1 year, (3) ability to independently use a wrist-worn device, (4) ownership of a compatible Android smartphone for successful device pairing, and (5) ability to install and use the Huawei Health app with regular access to Wi-Fi or mobile data for data synchronization. Exclusion criteria included (1) unstable medical conditions or acute illness at the time of recruitment, (2) severe cervical or lumbar spine disease or advanced osteoarthritis substantially limiting routine daily activity, and (3) incomplete baseline assessment.

Baseline assessments included demographic characteristics (age, sex, education level, marital status, living arrangement, care arrangement, and self-rated health status), cognitive function assessed using the Mini-Mental State Examination (MMSE) [36], and functional status assessed using the Barthel Activities of Daily Living (ADL) Index [37]. All enrolled participants demonstrated sufficient cognitive and functional capacity for independent device use at baseline.

Study Procedure

Wearable devices were distributed beginning in December 2023, with rolling enrollment following the completion of baseline assessments. Devices were offered consecutively to eligible and willing participants after the confirmation of smartphone compatibility and successful completion of device pairing, connection testing, and data transmission. Participants were not selected based on demographic characteristics, self-rated health status, or prior technology experience.

Participants received the wearable device at no cost and were instructed to wear it during daily life without restrictions, including during sleep, removing it only when necessary (eg, for charging). The monitoring start date was defined as the day on which participants completed device activation and first wore the device following study onboarding procedures. No predefined monitoring duration or mandatory wear requirements were set, and participants were allowed to discontinue device use at any time without affecting their participation in the China-Aging cohort. Monitoring continued until the last successful data synchronization or the end of November 2025, whichever occurred first.

In-person technical support was provided during initial device setup, with subsequent troubleshooting available through a WeChat group. Brief reminders were sent approximately every 2 days during the first 2 weeks to support device activation and data synchronization and to identify early technical issues. Thereafter, no proactive reminders or adherence-enhancing contact were provided, although participant-initiated technical support remained available through WeChat. This limited-support approach was intended to approximate real-world wearable use under naturalistic conditions.

Wearable Device and Data Acquisition

Participants were provided with a commercial wrist-worn wearable device (Huawei Band 8; Huawei Technologies Co) designed for the continuous monitoring of heart rate, physical activity, and sleep. The device incorporates a PPG sensor for heart rate monitoring and a triaxial accelerometer for activity tracking. Proprietary onboard algorithms generate summary measures of heart rate, step counts, and sleep parameters as implemented by the device manufacturer.

Participants paired the device with the Huawei Health app on their personal smartphones, through which wearable data were transmitted via Bluetooth and synchronized to the Huawei Health Research Platform. Data available through the research platform were pseudonymized and identified by wearable device ID. The linkage between wearable device ID and participant number was maintained separately by the research team. Data exported for analysis were stored on a secure research laptop accessible only to the study team.

Wear-Time Definition and Monitoring Criteria

Minute-level heart rate data were used as the primary indicator of wear-time, as heart rate is recorded only during direct skin contact [17,38]. Brief data gaps may occur due to transient loss of sensor contact or motion-related artifacts. Consistent with the improved nonwear detection algorithm proposed by Choi et al [39], which extended the foundational approach of Troiano et al [40], short interruptions of 1 - 2 minutes were treated as continuous wear, whereas longer gaps were considered potential nonwear periods. In the present study, nonwear was defined as intervals of 90 or above consecutive minutes without valid heart rate data, and subsequent data were treated as a new wear period. This approach was further reviewed and confirmed through consultation with Huawei technical engineers. Accordingly, daily wear-time was calculated as the sum of all valid wear periods within each calendar day.

A valid monitoring day was defined as a day with 10 or more hours of wear time, consistent with definitions used in wearable and accelerometry research [33,41]. A minimum of 3 valid monitoring days has been used in previous wearable research [33]. In the present feasibility study, we additionally required these 3 valid days to be consecutive as a pragmatic operational criterion for identifying short-term monitoring continuity under naturalistic conditions.

Based on these criteria, participants were classified into three monitoring-pattern categories:

XSL•FO
RenderX

1. No valid monitoring day: no day met the ≥10-hour valid-day criterion.
2. Nonconsecutive valid monitoring: 1 or more valid monitoring days, but no sequence of 3 or more consecutive valid monitoring days.
3. Minimal consecutive valid monitoring: 3 or more consecutive valid monitoring days.

User Experience Survey

User acceptability and usability were assessed using a self-administered questionnaire following a period of real-world wearable use. The questionnaire was developed for this study, informed by existing wearable usability instruments [32,42], and designed to capture multiple dimensions of usability, acceptability, and perceived impact in real-world use. It comprised 16 items, including 13 Likert-scale questions covering device comfort, ease of use, display readability, charging, app connectivity, data recording and accuracy, perceived lifestyle impact, and willingness for continued use and data sharing, as well as 3 open-ended questions exploring perceived benefits, barriers, and suggestions for improvement.

All Likert-scale items were rated on a 5-point scale, with higher scores indicating more positive evaluations. The questionnaire was originally developed in Chinese and translated into English for reporting. It was distributed to all participants after an initial period of device use, and responses were collected for subsequent analysis. Full details of the questionnaire are provided in Multimedia Appendix 1.

Statistical Analysis

All analyses were conducted using R (version 4.5.1) [43]. Descriptive statistics were used to summarize participant characteristics, monitoring metrics (eg, wear-time and valid days), and wearable-derived measures (eg, heart rate, physical activity, and sleep). Continuous variables are reported as mean (SD) for normally distributed data and median (IQR) for nonnormally distributed data.

Group comparisons were performed to examine potential differences in demographic, self-rated health status; cognitive function; and functional characteristics associated with wearable monitoring patterns. The distributions of continuous variables were assessed using the Shapiro-Wilk test. Continuous variables were summarized using median (IQR) and compared across monitoring-pattern groups using the Kruskal-Wallis test, with epsilon-squared (ε^2) reported as the effect size. Categorical variables were compared using Pearson chi-square tests or exact tests where appropriate, with Cramér *V* reported as the effect size. These analyses were exploratory and intended to characterize potential differences associated with monitoring patterns. Statistical significance was defined as a 2-sided $P<.05$.

For survey data, responses to Likert-scale items were summarized descriptively using median (IQR) and mean (SD). Open-ended responses were analyzed using an inductive thematic approach. Two researchers (YB and YZ) independently reviewed the responses and developed initial thematic categories related to perceived benefits, barriers, and suggested improvements. They then compared their categorizations and discussed disagreements, refining the categories until consensus was reached on the final coding and thematic structure.

To assess the robustness and potential sources of selection bias, we examined alternative monitoring thresholds and continuous monitoring measures, compared baseline characteristics between enrolled and nonenrolled participants, and compared baseline characteristics and monitoring measures between survey respondents and nonrespondents. Details of these analyses are provided in Multimedia Appendix 2.

Results

Participant Characteristics

All 710 participants approached for the wearable study owned a smartphone. Of these, 147 used a phone that was incompatible with the Huawei Health app (eg, iOS devices or unsupported Android versions), 64 did not proceed because stable data transmission could not be established during connection testing, and 13 did not enroll for other reasons that were not systematically recorded. Refusal was not recorded as a separate category and may therefore have been included among these 13 participants. Ultimately, 486 participants received a Huawei Band 8 and entered the wearable study, with wearable data available for all enrolled participants.

Baseline comparisons between the 486 enrolled and 224 nonenrolled participants showed no significant differences in age, sex, education, marital status, or self-rated health. Enrolled participants had higher MMSE scores and differed in living arrangement and care arrangement; a statistically significant but small difference was also observed in ADL. Full results are provided in Multimedia Appendix 2.

Among the 486 enrolled participants, 224 (46.1%) achieved minimal consecutive valid monitoring (≥3 consecutive valid monitoring days), 156 (32.1%) had nonconsecutive valid monitoring, and 106 (21.8%) had no valid monitoring day. Among participants with no valid monitoring day, 97 of 106 (91.5%) participants had a maximum observed daily wear time of less than 1 hour, while the remaining 9 of 106 (8.5%) participants had a maximum daily wear time of 1 to less than 5 hours. Within the nonconsecutive valid monitoring group, 119 of 156 (76.3%) participants had 1 valid monitoring day, 29 of 156 (18.6%) participants had 2 valid monitoring days, and 8 of 156 (5.1%) participants accumulated 3 or more valid monitoring days but never achieved 3 consecutively. Overall, 148 of 156 (94.9%) participants in this group had only 1 or 2 valid monitoring days. Full results are provided in Multimedia Appendix 2.

Baseline demographic, self-rated health status, cognitive, and functional characteristics across the 3 monitoring patterns are summarized in Table 1. The observed age range among enrolled participants was 59 to 86 years. Age, sex, education level, marital status, living arrangement, care arrangement, MMSE score, and Barthel ADL score did not differ significantly across monitoring patterns (all $P>.05$). Self-rated health was the only measured baseline characteristic that differed significantly across groups (Cramér $V=0.114$; $P=.005$), with better self-rated health

XSL•FO
RenderX

more frequently reported among participants with minimal consecutive valid monitoring.

XSL•FO
RenderX

Table . Baseline characteristics by wearable monitoring pattern (N=486). Effect sizes are reported as ε^2 for Kruskal-Wallis tests, and Cramér V for chi-square tests.

Characteristic	No valid monitoring day (n=106)	Nonconsecutive valid monitoring (n=156)	Minimal consecutive valid monitoring (n=224)	*P* value	Effect size
Age (years), median (IQR; range)	69 (66 - 72; 60 - 83)	68 (65 - 73; 60 - 86)	68 (65 - 71; 59 - 85)	.07	0.007[a]
Female sex, n (%)	65 (61.3)	94 (60.3)	155 (69.2)	.15[b]	0.062[c]
Education level, n (%)				.30[b]	0.040[c]
At or below junior high school	35 (33)	50 (32.1)	57 (25.4)		
High school/technical secondary	36 (34)	60 (38.5)	79 (35.3)		
At or above college/undergraduate	35 (33)	46 (29.5)	88 (39.3)		
Marital status, n (%)				.15[b,d]	0.078[c]
Married	98 (92.5)	145 (92.9)	213 (95.1)		
Not currently married[e]	8 (7.5)	11 (7.1)	11 (4.9)		
Living arrangement, n (%)				.97[b]	0.000[c]
With spouse only	76 (71.7)	107 (68.6)	151 (67.4)		
With children only	9 (8.5)	18 (11.5)	24 (10.7)		
With spouse and children	7 (6.6)	12 (7.7)	18 (8)		
With relatives/friends	12 (11.3)	15 (9.6)	28 (12.5)		
Living alone	2 (1.9)	4 (2.6)	3 (1.3)		
Care arrangement, n (%)				.97[b]	0.000[c]
No caregiver	72 (67.9)	99 (63.5)	148 (66.1)		
Spouse caregiver	32 (30.2)	51 (32.7)	67 (29.9)		
Child caregiver	2 (1.9)	6 (3.8)	7 (3.1)		
Relative/friend caregiver	—[f]	—	1 (0.4)		
Other	—	—	1 (0.4)		
Self-rated health status, n (%)				.005[b]	0.114[c]
Good or very good	18 (17)	36 (23.1)	71 (31.7)		
Fair	75 (70.8)	105 (67.3)	143 (63.8)		
Poor or very poor	13 (12.3)	15 (9.6)	10 (4.5)		
MMSE[g] score, median (IQR)	29 (27 - 29)	29 (27 - 29)	29 (27 - 30)	.09[h]	0.006[a]
Barthel ADL[i] Index, median (IQR)	100 (95 - 100)	100 (95 - 100)	100 (95 - 100)	.95[h]	0.000[a]

[a]Epsilon-squared.

[b]Chi-square test.

[c]Cramér V.

[d]Interpret with caution due to small cell counts.

[e]Includes single, divorced, and widowed.

[f]Not applicable.

XSL•FO
RenderX

[g]MMSE: Mini-Mental State Examination.

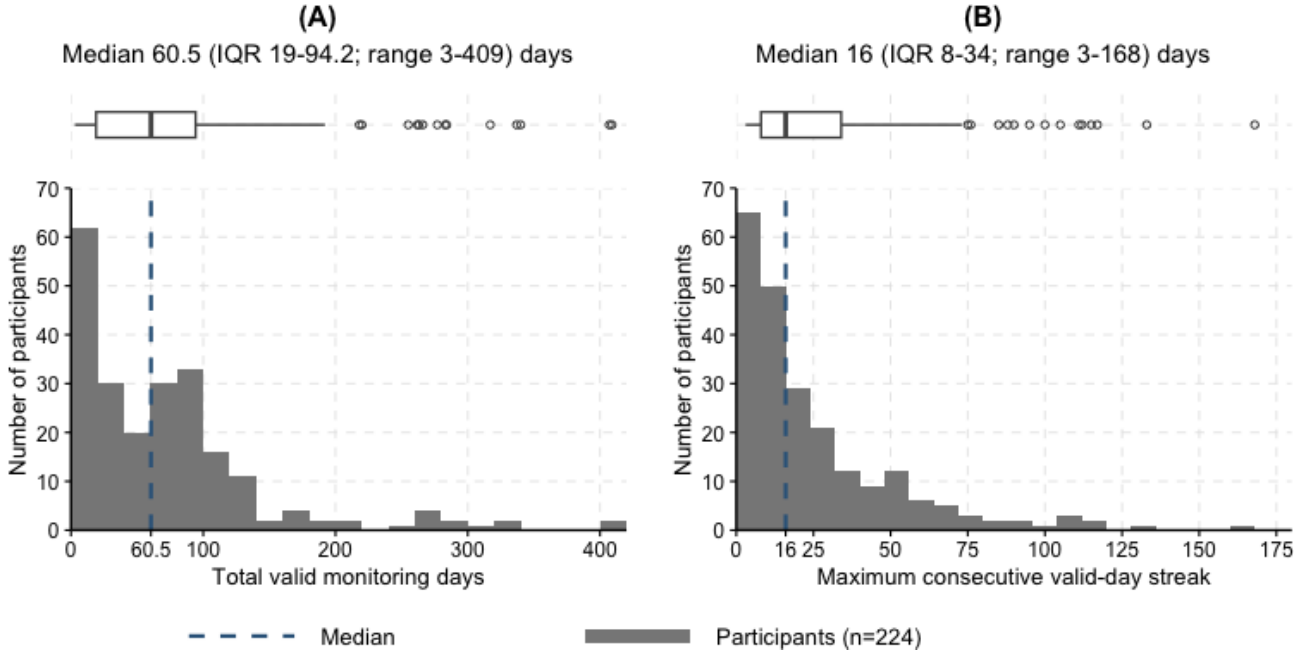
[h]Kruskal-Wallis test.

[i]ADL: Activities of Daily Living.

Wearable Monitoring Patterns and Data Availability

Among the 224 participants with minimal consecutive valid monitoring, the median total number of valid monitoring days was 60.5 (IQR 19 - 94.3; range 3 - 409), while the median maximum consecutive valid-day streak was 16 days (IQR 8 - 34; range 3 - 168), indicating substantial heterogeneity in both monitoring extent and continuity (Figure 1).

Figure 1. Monitoring extent and continuity among participants with minimal consecutive valid monitoring. (A) Distribution of total valid monitoring days. (B) Distribution of the maximum consecutive valid-day streak. Histograms show the number of participants across monitoring durations, with dashed blue lines indicating the median. Boxplots show the median, IQR, and distributional spread; circles indicate outliers.



Sensitivity analyses were conducted to examine the robustness of the findings to the primary ≥3-consecutive-valid-day criterion. Among the 380 participants with 1 or more valid monitoring days, 224 (58.9%), 211 (55.5%), 196 (51.6%), and 178 (46.8%) achieved 3 or more, 4 or more, 5 or more, and 7 or more consecutive valid monitoring days, respectively. The association between better self-rated health and consecutive valid monitoring remained directionally consistent across these alternative thresholds. Full results of the sensitivity analyses are provided in Multimedia Appendix 2.

Among participants with minimal consecutive valid monitoring, heart rate and step-count data were available for all 224 participants. Mean daily heart rate was 72.1 (SD 8.6) bpm, and median daily step count was 6177 (IQR 4249 - 8872) steps/day. Sleep data were available for 211 of 224 (94.2%) participants, with a median sleep duration of 7.23 (IQR 6.48 - 8.04) hours (Table 2).

Table . Summary of wearable-derived signal availability and characteristics among participants with minimal consecutive valid monitoring.

Signal	Participants with available data, n (%)	Summary metrics
Heart rate	224 (100)	72.1 (8.6; 49.4 - 93.1)[a] bpm
Sleep	211 (94.2)	7.23 (6.48 - 8.04; 3.22 - 10.39)[b] hours
Step count	224 (100)	6177 (4249 - 8872; 269 - 19,067)[b] steps/day

[a]Mean daily heart rate (SD; range).

[b]Median (IQR; range).

User Experience Survey

User acceptability and experience were assessed using a self-administered questionnaire completed by 106 of 486 (21.8%) participants after a period of real-world device use. Survey response rates were higher among participants with minimal consecutive valid monitoring (74/224, 33%) than among those with nonconsecutive valid monitoring (21/156, 13.5%) or no valid monitoring days (11/106, 10.4%). Although respondents and nonrespondents were broadly comparable in measured baseline characteristics (Multimedia Appendix 2), respondents demonstrated greater monitoring continuity and persistence. The survey findings therefore primarily reflect the

XSL•FO
RenderX

experiences of participants who maintained greater wearable use.

Across the 13 Likert-scale items (n=106), ratings were generally above the neutral midpoint (score=3), indicating overall favorable usability and acceptability (see Multimedia Appendix 3 for full item-level results). Median scores were 4 for most items, with IQRs ranging from 3-4 to 4-5, and mean values ranging from 3.51 to 4.22, with SDs ranging from 0.59 to 1.09. The highest rating was observed for independence of use (Q6; mean 4.22, SD 0.96), suggesting that most participants were able to operate the device without assistance. Items related to usability and data perception, including comfort, ease of use, readability, charging, connectivity, accuracy, and trust (Q1-Q8), also showed consistently positive ratings.

Perceived impact on health behavior was more varied. While participants reported increased health awareness (Q9) and greater willingness to engage in physical activity (Q10), perceived changes in daily habits (Q11) were more modest (median 3), indicating greater variability in responses. This suggests that although respondents reported greater awareness and motivation, perceived changes in daily habits were more limited. Willingness for long-term use (Q12) and data sharing with health care providers (Q13) remained high (median 4, IQR 4-4), indicating generally favorable acceptability among respondents.

For the open-ended questions, participants reported both perceived benefits and barriers to device use. The most commonly reported benefits included improved sleep monitoring (38/106, 35.8%) and activity or exercise tracking (37/106, 34.9%), followed by increased awareness of heart rate and oxygen saturation (29/106, 27.4%) and greater attention to overall health management (29/106, 27.4%). Only a small proportion (4/106, 3.8%) of participants reported no perceived benefits.

Reported barriers included device discomfort (18/106, 17%), data synchronization issues (12/106, 11.3%), and limited battery life (12/106, 11.3%). Some participants noted that removing the device for charging or showering disrupted routines and increased the likelihood of forgetting to put it back on. Others reported difficulty using functions beyond basic activity tracking. In contrast, 36 of 106 (34%) participants reported no inconvenience and were satisfied with device use.

Suggestions for improvement focused on both functionality and usability. The most common suggestions included adding additional health monitoring features (35/106, 33%), such as blood pressure or glucose tracking, simplifying device operation and guidance (17/106, 16%), and improving device comfort and battery performance (12/106, 11.3%). Approximately one quarter (27/106, 25.5%) of respondents reported no additional suggestions and were satisfied with the current device. A detailed breakdown of thematic categories is provided in Multimedia Appendix 3.

Discussion

Principal Findings

This study evaluated the real-world feasibility, usability, and acceptability of using a consumer wearable device for continuous health monitoring in community-dwelling older adults under limited-support, naturalistic conditions. The findings indicate that wearable monitoring was feasible for a substantial proportion of participants, with many achieving consecutive valid monitoring and maintaining device use beyond the initial monitoring period. At the same time, monitoring patterns varied considerably across individuals, with substantial heterogeneity in both monitoring extent and continuity, highlighting challenges for sustained real-world monitoring and large-scale implementation.

The observed monitoring patterns should be interpreted in the context of the limited-support design. Participants received support during initial device setup and brief reminders during the first 2 weeks but no proactive adherence support thereafter. This differs from previous studies using structured follow-up, training, prompting, or incentives and reporting relatively high adherence [16,32,44,45]. For example, Paolillo et al [45] reported wearable use on an average of 95% of study days during a supported 30-day protocol, while Holmqvist et al [44] reported a mean daily wear time of 21 hours during a structured 4-week study involving training and automated prompting. Although direct comparisons are limited by differences in study design and monitoring definitions, the present findings provide an estimate of wearable use under conditions with limited ongoing support. The substantial variation in monitoring extent and continuity also aligns with broader patterns of digital health use, in which sustained engagement depends on the integration of technology into everyday routines [46,47]. Together, these findings suggest that wearable feasibility is influenced by implementation context and support intensity and that supported and naturalistic studies provide complementary evidence for understanding real-world use.

Implementation Challenges and Associated Factors

From an implementation perspective, participants with nonconsecutive valid monitoring may represent an important target for improving sustained wearable use. Most participants in this group accumulated only 1 or 2 valid monitoring days, suggesting that initial device use did not consistently develop into sustained monitoring [48,49]. Practical barriers identified in the user experience survey, including data synchronization, battery management, and interruptions associated with device removal, may contribute to difficulties in incorporating wearable use into everyday routines [29,32,50]. Previous research suggests that sustained wearable use can be supported by the development of routine behaviors, such as integrating device wear and charging into established daily activities [50,51]. Additional support during the early monitoring period may therefore be one strategy for facilitating the transition from initial to more sustained use.

Participants with no valid monitoring day may represent a different implementation challenge. Although all participants in this group generated some wearable data, observed wear was

XSL•FO
RenderX

extremely limited: among the 106 participants, 97 (91.5%) participants had a maximum daily wear time of less than 1 hour, and none reached 5 hours on any observed day. This pattern suggests that the predominant challenge was establishing meaningful wearable use after initial setup, rather than wear that narrowly fell below the ≥10-hour valid-day criterion. Notably, this occurred despite successful device pairing, connection testing, and data transmission before enrollment, suggesting that successful technical onboarding alone may not be sufficient to ensure subsequent wearable use. Potential barriers may include technical and usability challenges, perceived usefulness, and difficulties integrating the device into everyday activities [29,32,52,53]. However, the specific reasons for limited monitoring in this group could not be determined because survey participation was particularly low, and barriers reported by survey respondents may not reflect the experiences of participants with little valid monitoring. Further research is therefore needed to identify the factors contributing to limited early-stage monitoring and determine whether targeted technical or usability support could facilitate continued device use.

Taken together, these findings suggest that different monitoring patterns may require different implementation strategies. Participants with some valid monitoring but limited continuity may benefit from support focused on establishing consistent monitoring routines, whereas those who do not achieve a valid monitoring day may require greater attention to barriers occurring earlier in the monitoring process. Distinguishing between these patterns may therefore help inform more targeted approaches to supporting wearable monitoring.

Self-rated health was the only measured baseline characteristic significantly associated with monitoring pattern, with better self-rated health more frequently reported among participants achieving minimal consecutive valid monitoring. This association remained directionally consistent across alternative consecutive-day thresholds, suggesting that the finding was not dependent on the primary ≥3-day criterion. Similar associations between health status and digital health use have been reported previously, with greater health burden potentially increasing the practical demands of device use [34,48,49]. Self-rated health is a multidimensional measure that may reflect aspects of physical and psychological well-being, perceived control, and health-related self-efficacy, which have also been associated with technology adoption and health behaviors [52,54]. However, the mechanisms underlying the observed association were not assessed in the present study, and the finding should therefore be interpreted as exploratory rather than causal.

The absence of significant associations between monitoring patterns and age, education, cognitive function, or functional status is also informative, suggesting that these measured characteristics did not substantially differentiate monitoring patterns within the enrolled cohort. However, these findings should be interpreted in the context of the relatively cognitively and functionally intact study population. The high and narrow range of MMSE and Barthel ADL scores limited the ability to examine the effects of greater cognitive or functional impairment. In addition, comparisons between enrolled and nonenrolled individuals identified differences in cognitive function, living arrangement, and care arrangement, indicating that selection may occur before wearable monitoring begins. Therefore, the absence of associations within the enrolled cohort should not be interpreted as evidence that cognitive or functional characteristics are not important for wearable adoption more broadly [52,53,55]. Further research including older adults with more diverse cognitive, functional, and digital capabilities is needed.

Among participants with minimal consecutive valid monitoring, wearable-derived heart rate and step-count data were available for all participants, while sleep data were available for 94.2% (n=211) of the participants. This indicates that multiple behavioral and physiological measures could be obtained among participants who achieved minimal consecutive valid monitoring. The slightly lower availability of sleep data may reflect the additional requirements of overnight wear, including charging and device removal. More broadly, these findings indicate that successful wearable monitoring does not necessarily result in equivalent data availability across monitoring domains, an important consideration when consumer wearables are used for longitudinal health monitoring.

The user experience survey also provided insight into practical aspects of device use. While some participants reported no inconvenience, others identified barriers related to comfort, data synchronization, and battery life, and 33% (35/106) of the respondents suggested additional health-monitoring functions, such as blood pressure or glucose monitoring. These findings suggest that both practical burden and perceived relevance may influence continued wearable use and that greater alignment between device functionality and users' health priorities may enhance acceptability [29,32,46,47]. However, these findings should be interpreted with caution because the respondents demonstrated greater monitoring continuity and persistence than nonrespondents, indicating that the usability, acceptability, and barrier findings disproportionately reflect the experiences of participants who maintained greater wearable use.

Strengths, Limitations, and Future Directions

The principal strength of this study lies in its community-based design under naturalistic conditions, which captured variability in wearable use with limited ongoing support. The combination of objective monitoring data and user experience responses also enabled feasibility to be examined from complementary perspectives. Several limitations should be considered. First, recruitment from an urban community setting and requirements for smartphone compatibility and successful data transmission may limit generalizability, particularly to older adults with lower digital capability or greater cognitive or functional impairment. Differences between enrolled and nonenrolled participants further indicate potential preenrollment selection. Second, the ≥3-consecutive-valid-day criterion was a pragmatic operational definition rather than a validated threshold, although sensitivity analyses using alternative thresholds and continuous monitoring measures supported the overall findings. Third, the relatively low user experience survey participation rate (106/486, 21.8%) and greater monitoring persistence among respondents may have resulted in more favorable estimates of usability and acceptability. In addition, individual-level use of technical support was not systematically recorded, limiting our ability to

XSL•FO
RenderX

assess whether support use differed across monitoring patterns or influenced monitoring continuity. Finally, the study used a single consumer wearable ecosystem with proprietary algorithms that were not independently validated against research-grade reference measures in this population. Future research should therefore examine longer-term monitoring patterns, identify barriers to sustained use, evaluate targeted technical and behavioral support, and validate wearable-derived measures against reference standards. Longitudinal associations with cognitive, functional, and health outcomes should also be investigated to establish the potential clinical relevance of these measures.

Conclusions

In conclusion, this study demonstrates that wrist-worn wearable monitoring is feasible for a substantial proportion of community-dwelling older adults in China under limited-support, naturalistic conditions, although monitoring extent and continuity varied considerably. Usability and acceptability were generally favorable among survey respondents. Future research should identify strategies to support sustained monitoring and establish the clinical relevance of wearable-derived measures for longitudinal changes in cognitive and functional health.

Acknowledgments

This study was conducted within the China Healthy Aging Cohort Study (China-Aging). The authors would like to thank all participants for their time and contribution, as well as the staff at participating community health centers for their support with recruitment and data collection. They also thank Dr Xiaolong Zhang (Beijing Anding Hospital, Capital Medical University) for his helpful comments and suggestions from digital health and psychological perspectives, which informed the interpretation of the findings and the discussion.

The authors did not use generative AI in any portion of the preparation of this manuscript.

Funding

This study was funded by Noncommunicable Chronic Diseases-National Science and Technology Major Project (2025ZD0546200) and Beijing Outstanding Young Scientist Program (JWZQ20240101023). The funders had no role in the study design; data collection, analysis, or interpretation; manuscript preparation; or the decision to submit the manuscript for publication.

Data Availability

Deidentified data supporting the findings of this study are available to qualified researchers upon reasonable request, subject to approval by the Ethics Committee of Xuanwu Hospital, Capital Medical University. Requests should be directed to the corresponding author.

Authors' Contributions

Conceptualization: YB, ZW
Data curation: YZ
Formal analysis: YB
Funding acquisition: YT
Investigation: YZ, LM, ZW
Methodology: YB (lead), YM (supporting)
Project administration: YT
Resources: LM, ZW
Supervision: YT
Visualization: YB
Writing – original draft: YB
Writing – review & editing: YB (lead), YZ (supporting), LM (supporting), ZW (supporting), YM (supporting), YT (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Wearable device usability and acceptability survey.
[PDF File, 50 KB - humanfactors_v13i1e98592_app1.pdf]

Multimedia Appendix 2
Summary of additional analyses on monitoring patterns, sensitivity analyses, and potential selection bias.
[PDF File, 192 KB - humanfactors_v13i1e98592_app2.pdf]

XSL•FO
RenderX

Multimedia Appendix 3
Summary of survey results for usability and acceptability.
[PDF File, 232 KB - humanfactors_v13i1e98592_app3.pdf]

References

1. Ageing and health. World Health Organization. 2025 Oct 1. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2026-01-27]
2. World report on ageing and health. : World Health Organization; 2015 Sep 29 URL: https://www.who.int/publications/i/item/9789241565042 [accessed 2026-01-27]
3. Newman AB. The epidemiology and societal impact of aging-related functional limitations: a looming public health crisis. J Gerontol A Biol Sci Med Sci 2023 Jun 16;78(Suppl 1):4-7. [doi: 10.1093/gerona/glad021] [Medline: 37325965]
4. Paluch AE, Bajpai S, Bassett DR, et al. Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. Lancet Public Health 2022 Mar;7(3):e219-e228. [doi: 10.1016/S2468-2667(21)00302-9] [Medline: 35247352]
5. Chan LLY, Espinoza Cerda MT, Brodie MA, Lord SR, Taylor ME. Daily-life walking speed, running duration and bedtime from wrist-worn sensors predict incident dementia: a watch walk—UK biobank study. Int Psychogeriatr 2025 Jun;37(3):100031. [doi: 10.1016/j.inpsyc.2024.100031] [Medline: 39827009]
6. Del Pozo Cruz B, Ahmadi M, Naismith SL, Stamatakis E. Association of daily step count and intensity with incident dementia in 78 430 adults living in the UK. JAMA Neurol 2022 Oct 1;79(10):1059-1063. [doi: 10.1001/jamaneurol.2022.2672] [Medline: 36066874]
7. Keil SA, Schindler AG, Wang MX, et al. Longitudinal sleep patterns and cognitive impairment in older adults. JAMA Netw Open 2023 Dec 1;6(12):e2346006. [doi: 10.1001/jamanetworkopen.2023.46006] [Medline: 38048131]
8. Xia Y, He S, Sun Z, Wang F. The association between sleep variability and cognitive function: a systematic review and meta analysis. J Affect Disord 2026 Jan 15;393(Pt B):120481. [doi: 10.1016/j.jad.2025.120481] [Medline: 41101484]
9. Chamberlin KW, Etzkorn LH, Davoudi A, et al. Interdaily variability of rest–activity patterns and risk of dementia in older adults: the ARIC Neurocognitive Study. Alzheimers Dement (N Y) 2026 Jan;12(1):e70219. [doi: 10.1002/trc2.70219] [Medline: 41626291]
10. Wang W, Wanigatunga AA, Etzkorn LH, et al. Association between circadian rest-activity rhythms and incident dementia in older adults: the atherosclerosis risk in communities study. Neurology 2026 Jan 27;106(2):e214513. [doi: 10.1212/WNL.0000000000214513] [Medline: 41461059]
11. Piau A, Wild K, Mattek N, Kaye J. Current state of digital biomarker technologies for real-life, home-based monitoring of cognitive function for mild cognitive impairment to mild Alzheimer disease and implications for clinical care: systematic review. J Med Internet Res 2019 Aug 30;21(8):e12785. [doi: 10.2196/12785] [Medline: 31471958]
12. Shephard RJ. Limits to the measurement of habitual physical activity by questionnaires. Br J Sports Med 2003 Jun;37(3):197-206. [doi: 10.1136/bjsm.37.3.197] [Medline: 12782543]
13. González DA, Gonzales MM, Jennette KJ, Soble JR, Fongang B. Cognitive screening with functional assessment improves diagnostic accuracy and attenuates bias. Alzheimers Dement (Amst) 2021 Jan;13(1):e12250. [doi: 10.1002/dad2.12250] [Medline: 34934799]
14. Jekel K, Damian M, Wattmo C, et al. Mild cognitive impairment and deficits in instrumental activities of daily living: a systematic review. Alzheimers Res Ther 2015;7(1):17. [doi: 10.1186/s13195-015-0099-0] [Medline: 25815063]
15. Kroenke K, Stump TE, Monahan PO. Agreement between older adult patient and caregiver proxy symptom reports. J Patient Rep Outcomes 2022 May 14;6(1):50. [doi: 10.1186/s41687-022-00457-8] [Medline: 35567663]
16. Teixeira E, Fonseca H, Diniz-Sousa F, et al. Wearable devices for physical activity and healthcare monitoring in elderly people: a critical review. Geriatrics (Basel) 2021 Apr 7;6(2):38. [doi: 10.3390/geriatrics6020038] [Medline: 33917104]
17. de Zambotti M, Cellini N, Goldstone A, Colrain IM, Baker FC. Wearable sleep technology in clinical and research settings. Med Sci Sports Exerc 2019 Jul;51(7):1538-1557. [doi: 10.1249/MSS.0000000000001947] [Medline: 30789439]
18. Alugubelli N, Abuissa H, Roka A. Wearable devices for remote monitoring of heart rate and heart rate variability—what we know and what is coming. Sensors (Basel) 2022 Nov 17;22(22):8903. [doi: 10.3390/s22228903] [Medline: 36433498]
19. Lu L, Zhang J, Xie Y, et al. Wearable health devices in health care: narrative systematic review. JMIR Mhealth Uhealth 2020 Nov 9;8(11):e18907. [doi: 10.2196/18907] [Medline: 33164904]
20. Ding D, Nguyen B, Nau T, et al. Daily steps and health outcomes in adults: a systematic review and dose-response meta-analysis. Lancet Public Health 2025 Aug;10(8):e668-e681. [doi: 10.1016/S2468-2667(25)00164-1] [Medline: 40713949]
21. Hamaya R, Shiroma Jr EJ, Moore CC, Buring JE, Evenson KR, Lee IM. Time- vs step-based physical activity metrics for health. JAMA Intern Med 2024 Jul 1;184(7):718-725. [doi: 10.1001/jamainternmed.2024.0892] [Medline: 38767892]
22. LaMonte MJ, LaCroix AZ, Nguyen S, et al. Accelerometer-measured physical activity, sedentary time, and heart failure risk in women aged 63 to 99 years. JAMA Cardiol 2024 Apr 1;9(4):336-345. [doi: 10.1001/jamacardio.2023.5692] [Medline: 38381446]
23. Winer JR, Mander BA, Helfrich RF, et al. Sleep as a potential biomarker of tau and β-amyloid burden in the human brain. J Neurosci 2019 Aug 7;39(32):6315-6324. [doi: 10.1523/JNEUROSCI.0503-19.2019] [Medline: 31209175]

XSL•FO
RenderX

24. Chan LLY, Choi TCM, Lord SR, Brodie MA. Development and large-scale validation of the Watch Walk wrist-worn digital gait biomarkers. Sci Rep 2022 Oct 10;12(1):16211. [doi: 10.1038/s41598-022-20327-z] [Medline: 36217013]
25. Varma VR, Ghosal R, Hillel I, et al. Continuous gait monitoring discriminates community-dwelling mild Alzheimer's disease from cognitively normal controls. Alzheimers Dement (N Y) 2021 Jan;7(1):e12131. [doi: 10.1002/trc2.12131] [Medline: 33598530]
26. Kim HJ, Hwang YJ, Park JE, Lee DY. Effectiveness of walking prescription using mobile health technology on the changes in daily steps in older adults with cognitive impairment: randomized controlled study. JMIR Aging 2025 Jun 11;8:e63081. [doi: 10.2196/63081] [Medline: 40499016]
27. Lentzen M, Vairavan S, Muurling M, et al. RADAR-AD: assessment of multiple remote monitoring technologies for early detection of Alzheimer's disease. Alzheimers Res Ther 2025 Jan 27;17(1):29. [doi: 10.1186/s13195-025-01675-0] [Medline: 39865315]
28. Muurling M, de Boer C, Kozak R, et al. Remote monitoring technologies in Alzheimer's disease: design of the RADAR-AD study. Alzheimers Res Ther 2021 Apr 23;13(1):89. [doi: 10.1186/s13195-021-00825-4] [Medline: 33892789]
29. Muurling M, de Boer C, Hinds C, et al. Feasibility and usability of remote monitoring in Alzheimer's disease. Digit Health 2024;10:20552076241238133. [doi: 10.1177/20552076241238133] [Medline: 38601188]
30. Mc Ardle R, Jabbar KA, Del Din S, et al. Using digital technology to quantify habitual physical activity in community dwellers with cognitive impairment: systematic review. J Med Internet Res 2023 May 18;25:e44352. [doi: 10.2196/44352] [Medline: 37200065]
31. Ning Y, Chen M, Yang H, Jia J. Accelerometer-measured physical activity timing with incident dementia. Alzheimers Dement 2025 Feb;21(2):e14452. [doi: 10.1002/alz.14452] [Medline: 39868637]
32. Keogh A, Dorn JF, Walsh L, Calvo F, Caulfield B. Comparing the usability and acceptability of wearable sensors among older Irish adults in a real-world context: observational study. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e15704. [doi: 10.2196/15704] [Medline: 32310149]
33. Chan A, Chan D, Lee H, Ng CC, Yeo AHL. Reporting adherence, validity and physical activity measures of wearable activity trackers in medical research: a systematic review. Int J Med Inform 2022 Apr;160:104696. [doi: 10.1016/j.ijmedinf.2022.104696] [Medline: 35121356]
34. Abouzahra M, Ghasemaghaei M. The antecedents and results of seniors' use of activity tracking wearable devices. Health Policy Technol 2020 Jun;9(2):213-217. [doi: 10.1016/j.hlpt.2019.11.002]
35. Chang J, Wu Y, Zhao Y, et al. Cohort profile: China healthy aging cohort study (China-Aging). J Alzheimers Dis 2025 Sep;107(1):348-360. [doi: 10.1177/13872877251360028] [Medline: 40660850]
36. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
37. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. Md State Med J 1965 Feb;14:61-65. [Medline: 14258950]
38. Nelson BW, Low CA, Jacobson N, Areán P, Torous J, Allen NB. Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. NPJ Digit Med 2020;3:90. [doi: 10.1038/s41746-020-0297-4] [Medline: 32613085]
39. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. Med Sci Sports Exerc 2011 Feb;43(2):357-364. [doi: 10.1249/MSS.0b013e3181ed61a3] [Medline: 20581716]
40. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
41. Sasaki JE, Júnior JH, Meneguci J, et al. Number of days required for reliably estimating physical activity and sedentary behaviour from accelerometer data in older adults. J Sports Sci 2018 Jul;36(14):1572-1577. [doi: 10.1080/02640414.2017.1403527] [Medline: 29125022]
42. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: a mixed-methods evaluation. JMIR Mhealth Uhealth 2016 Jan 27;4(1):e7. [doi: 10.2196/mhealth.4225] [Medline: 26818775]
43. The R Project for Statistical Computing. URL: https://www.r-project.org/ [accessed 2026-04-14]
44. Holmqvist S, Kaplan M, Chaturvedi R, Shou H, Giovannetti T. Longitudinal and combined smartwatch and ecological momentary assessment in racially diverse older adults: feasibility, adherence, and acceptability study. JMIR Hum Factors 2025 Apr 8;12:e69952. [doi: 10.2196/69952] [Medline: 40198914]
45. Paolillo EW, Lee SY, VandeBunte A, et al. Wearable use in an observational study among older adults: adherence, feasibility, and effects of clinicodemographic factors. Front Digit Health 2022;4:884208. [doi: 10.3389/fdgth.2022.884208] [Medline: 35754462]
46. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11. [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
47. Moore K, O'Shea E, Kenny L, et al. Older adults' experiences with using wearable devices: qualitative systematic review and meta-synthesis. JMIR Mhealth Uhealth 2021 Jun 3;9(6):e23832. [doi: 10.2196/23832] [Medline: 34081020]
48. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]

XSL•FO
RenderX

49. Yardley L, Spring BJ, Riper H, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
50. Lally P, van Jaarsveld CHM, Potts HWW, Wardle J. How are habits formed: modelling habit formation in the real world. Euro J Social Psych 2010 Oct;40(6):998-1009. [doi: 10.1002/ejsp.674]
51. Peng W, Li L, Kononova A, Cotten S, Kamp K, Bowen M. Habit formation in wearable activity tracker use among older adults: qualitative study. JMIR Mhealth Uhealth 2021 Jan 19;9(1):e22488. [doi: 10.2196/22488] [Medline: 33464216]
52. Czaja SJ, Charness N, Fisk AD, et al. Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352. [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
53. Vaportzis E, Clausen MG, Gow AJ. Older adults perceptions of technology and barriers to interacting with tablet computers: a focus group study. Front Psychol 2017 Oct 4;8:1687. [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]
54. Heart T, Kalderon E. Older adults: are they ready to adopt health-related ICT? Int J Med Inform 2013 Nov;82(11):e209-e231. [doi: 10.1016/j.ijmedinf.2011.03.002] [Medline: 21481631]
55. Mitzner TL, Boron JB, Fausset CB, et al. Older adults talk technology: technology usage and attitudes. Comput Human Behav 2010 Nov 1;26(6):1710-1721. [doi: 10.1016/j.chb.2010.06.020] [Medline: 20967133]

Abbreviations

ADL: Activities of Daily Living
MMSE: Mini-Mental State Examination
PPG: photoplethysmography

Edited by Y Gong; submitted 17.Apr.2026; peer-reviewed by B Cao, N Stasinos; revised version received 08.Sep.2026; accepted 05.Oct.2026; published 29.Oct.2026.

Please cite as:
Bai Y, Tang Y, Zhao Y, Ma L, Ma Y, Wang Z
Real-World Health Monitoring Using Wrist-Worn Wearables in Community-Dwelling Older Adults in China: Mixed Methods Feasibility Study
JMIR Hum Factors 2026;13:e98592
URL: https://humanfactors.jmir.org/2026/1/e98592
doi:10.2196/98592

© Ying Bai, Yi Tang, Yiwei Zhao, Lixin Ma, Yanjun Ma, Zhibin Wang. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 29.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Body Image and Eating Disorder Education Chatbot, JEM, in Australia and Canada: First 6-Month Real-World Survey Evaluation

Gemma Sharp[1], PhD; Sara Marini[2], MPP; Emily Tam[2], MPH; Hao Hu[1], MIT

[1]Department of Neuroscience, Monash University, 99 Commercial Road, Melbourne, Victoria, Australia

[2]National Eating Disorder Information Centre, Toronto, ON, Canada

Corresponding Author:
Gemma Sharp, PhD
Department of Neuroscience, Monash University, 99 Commercial Road, Melbourne, Victoria, Australia

Abstract

Background: Body image dissatisfaction, disordered eating, and eating disorders represent significant public health concerns; however, many affected individuals never access evidence-based support. We co-designed and developed a rule-based chatbot, JEM, which conducts conversations addressing evidence-based psychoeducation and psychotherapeutic microinterventions. We previously demonstrated the feasibility, acceptability, and preliminary satisfaction of the JEM chatbot in a research setting. However, broader satisfaction, experiences, and user-reported outcomes in real-world settings have not yet been investigated.

Objective: This study aims to conduct a real-world evaluation of the JEM chatbot in Australia and Canada, the two countries that have hosted a deployment of the chatbot to date. Specifically, we aim to explore user satisfaction and experiences with the chatbot and within-session differences in user mood and body image satisfaction when completing the chatbot's microinterventions.

Methods: Respondents were users of the JEM chatbot aged 13 to 64 years who self-selected to complete a web-based overall evaluation survey (N=230; n=122 in Australia and n=108 in Canada) over a 6-month period. This evaluation survey included user demographic characteristics, satisfaction measures, and the System Usability Scale. Respondents for the within-session pre-post analyses were JEM chatbot users who chose to complete brief web-based surveys immediately before and after completing one of the chatbot's microinterventions during the same 6-month period. Sample sizes varied across microinterventions, ranging from 75 to 276 respondents overall (Australia: n=34 - 146; Canada: n=39 - 130). These surveys included validated visual analog scales (VAS) measuring mood (anxiety, depression, happiness, confidence) and body image satisfaction (body size satisfaction, body shape satisfaction, physical attractiveness).

Results: Demographic characteristics showed that survey respondents were commonly young adult cisgender women and nonbinary individuals across Australia and Canada. Respondent satisfaction with the chatbot was high in both countries (Australia: mean 76.1, SD 22.7; Canada: mean 78.8, SD 14.3), and the usability of the chatbot was rated as "excellent" in both countries (Australia: mean 86.5, SD 16.9; Canada: mean 89.5, SD 11.6) according to the System Usability Scale. Across completed microintervention surveys, patterns of within-session pre-post ratings were broadly similar in Australia and Canada, with effect sizes generally ranging from very small to large across VAS-measured mood and body image outcomes.

Conclusions: The JEM chatbot achieved high satisfaction and usability ratings. Among respondents who completed pre-post surveys, immediate within-session differences in mood and body image ratings were observed following the completion of chatbot microinterventions. The study findings were broadly similar across Australia and Canada. These results provide evidence of user experience and within-session differences following engagement with JEM and support continued evaluation in future studies.

(JMIR Hum Factors 2026;13:e90783) doi:10.2196/90783

KEYWORDS

eating disorder; body image; mental health; chatbot; conversational agent; artificial intelligence; digital health; psychoeducation; microintervention; evaluation; survey

Introduction

Eating disorders are characterized by disruptions in eating or feeding behaviors that negatively impact an individual's physical and psychosocial functioning [1]. These disorders are present across ages, genders, and backgrounds, and their prevalence and impact are increasing [2,3]. Eating disorders are complex and involve a combination of biological, psychological, sociocultural, and environmental factors [2]. Body image concerns and body dissatisfaction have a significant association with the development of some eating disorders and extreme weight-loss behaviors [4]. Body image dissatisfaction and eating disorders are major public health concerns contributing to substantial economic and health burdens [5-7]; however, many

XSL•FO
RenderX

affected individuals cannot access psychoeducation, early intervention, or timely support.

In response to the growing demand for accessible mental health support, digital mental health interventions have increased in popularity, including artificial intelligence (AI) chatbots. Chatbots offer advantages such as accessibility, instantaneous responses, and being free of cost to users, and they can act as a pathway to more individualized services [8,9]. They can also be helpful in overcoming barriers such as mental health stigma, which is particularly common in body image distress and eating disorders, or having a lack of services available [8,9]. An expanding body of research is supporting chatbot usage and efficacy in the mental health care sector [10-13]. However, despite the rapid growth in chatbot development, relatively few have been rigorously evaluated in real-world eating disorder or body image–specific contexts, with most evidence derived from controlled or early-stage pilot studies [14]. This limits understanding of how such tools perform under naturalistic conditions, including voluntary use, more diverse user populations, and nonresearch settings. As a result, there remains limited evidence regarding engagement, acceptability, and short-term outcomes of body image– and eating disorder–focused chatbots when deployed at scale.

To date, a small number of chatbots with eating disorder– and/or body image–focused evidence-based components have been reported in the literature, including Tessa [15], Topity [16], Alex [17], an early-stage unnamed chatbot for adolescents at risk of eating disorders [18], ED ESSI [19,20], and JEM [21] (formerly KIT), which is the focus of the current study. These systems differ in their primary aims and intervention focus, spanning prevention and early intervention (Tessa, Topity), support for individuals on treatment waitlists (ED ESSI), and early-stage or partially evaluated tools (Alex). Only Tessa, Topity, and ED ESSI have demonstrated effectiveness in randomized controlled trials (RCTs), with Tessa reducing weight and shape concerns in young women [15], Topity improving body image outcomes in adolescents [16], and ED ESSI reducing eating disorder pathology and mood-related outcomes in adolescents and adults awaiting treatment [19]. Certain components of Alex have also been evaluated in controlled settings, although the optimized system has not yet been fully tested [22]. Overall, despite promising early findings, there remains a lack of large-scale real-world evaluations of eating disorder and body image–focused chatbots that combine psychoeducation with brief evidence-based skill-building interventions.

JEM is a rule-based chatbot (ie, operating using predefined conversational pathways rather than generative AI) developed using the Google Dialogflow platform [23]. The chatbot was designed to deliver evidence-based eating disorder–focused psychoeducation alongside brief skills-based microinterventions [21]. The psychoeducation component provides structured information on eating disorders, including prevalence, types, warning signs, health effects, causes, and help-seeking, while the microinterventions focus on brief cognitive behavior therapy (CBT), acceptance and commitment therapy (ACT), and mindfulness-informed exercises targeting cognitive and emotional processes. These microinterventions include education on cognitive distortions or unhelpful thinking styles (CBT), practicing detaching from unhelpful thoughts via cognitive defusion exercises (ACT), and mindful breathing. Both the psychoeducation and microintervention components are delivered using multimodal formats, including text, audio, video, and graphics.

JEM was co-designed with young people aged 13 years and older with lived experience of body image concerns and/or eating disorders, parents/carers, and health professionals [21]. The co-design process provided iterative feedback on content, tone, and usability, which informed the refinement of both the psychoeducation and microintervention components. Overall, co-design findings supported the acceptability and feasibility of JEM and informed its final structure and delivery format [21]. The efficacy of JEM was not tested in an RCT setting before wider scale deployment within Australia because the development timing coincided with the COVID-19 pandemic when demand for support was high and many health initiatives were decided to be fast-tracked [24,25].

While it can be argued that RCTs are ideal for establishing evidence of efficacy as they offer structured environments for rigorous hypothesis testing and reduce the influence of external factors, their findings may not always be generalizable [26]. Evaluating a chatbot in a real-world setting is advantageous in this regard. Naturalistic evaluations are beneficial for understanding user experiences, engagement, and outcomes as they offer increased ecological validity by enabling more natural interactions and outcomes [27-29]. Moreover, they can involve a diverse user base, allowing for a better understanding of the chatbot's usability and effectiveness across various demographics [28,29]. Additionally, real-world evaluations may capture more accurate long-term engagement, as users are not confined by the structured timeframe or artificial environment of an RCT, making it easier to evaluate the chatbot's sustainability and broader impact [26,27]. However, such designs also introduce limitations including self-selection bias, lack of control conditions, and challenges in causal inference.

Real-world evaluations of mental health–specific chatbots are seemingly quite rare [30]. For instance, a study of Wysa, a mobile app for mental resilience and well-being, involved 129 voluntary users [31]. Data were collected via anonymous in-app feedback and assessment questionnaires from users across 23 time zones. High Wysa users showed significantly greater improvement in major depression symptoms with a moderate effect than low users, with 67.7% reporting the app as helpful and encouraging. The authors noted that these results aligned with prior RCTs on conversational agents for mental well-being [32,33] and emphasized that the naturalistic design offered scalable, real-time insights into engagement and effectiveness.

A single-session "mini-course" version of the Tessa chatbot, focused on eating disorders, was studied in users prompted via social media searches [34]. This mini-course improved body image (moderate effect) and motivation to change (small effect), but the study was not fully naturalistic due to researcher screening before mini-course access. When Tessa was deployed in a real-world US setting in June 2023, it shifted from

XSL•FO
RenderX

rule-based responses to offering "off-script" dieting and weight-loss advice, which can be very harmful in eating disorder contexts. This issue was reported by users, not researchers, and attracted global attention [35,36]. Since 2023, Tessa has not been implemented in any public forums to the best of our knowledge. Overall, real-world evaluations of body image– or eating disorder–specific chatbots and mental health chatbots in general remain scarce.

The objective of the current study was to conduct a novel real-world evaluation of our body image and eating disorder education chatbot, JEM, in Australia and Canada, the 2 countries (representing 2 continents) that have hosted a deployment of the chatbot to date. Exploring both settings provided an opportunity to examine whether similar patterns of user experiences and outcomes were observed across different implementation contexts.

Given the exploratory nature of this real-world evaluation, no a priori directional hypotheses were specified. However, the microinterventions were grounded in CBT, ACT, and mindfulness approaches and were examined as within-session pre-post changes in self-reported mood and body image satisfaction following engagement with the microinterventions. The primary aim of the study was to examine user satisfaction and experiences and explore within-session pre-post changes in body image satisfaction and mood for the microinterventions. No specific eating disorder symptoms or behavioral outcomes were assessed. A secondary aim was to explore and describe any differences between the Australian and Canadian JEM deployments.

Methods

Ethical Considerations

This study was approved by the Monash University Human Research Ethics Committee (ID 26129), which encompassed research conducted in both Australia and Canada. Chatbot users were offered the opportunity to complete web-based surveys throughout their conversation with the JEM chatbot; however, the surveys were framed as completely optional, and users could ignore the survey and continue using the chatbot without consequence. When users clicked on a survey weblink, they were presented with detailed study information. Users were informed of their right to withdraw from the survey at any time without any negative impacts. After the detailed information, users were asked to consent by clicking on a button that they had read the information and consented to proceeding to the survey. For users aged 13 to 15 years, ethics-approved procedures required them to confirm that they had obtained parental or guardian consent via an online acknowledgment. Given the fully anonymous and low-risk nature of the research, this approach was deemed appropriate and proportionate by the approving committee. The lower age threshold of 13 years was selected in line with common minimum ages for digital platform access and social media use and to capture early adolescent users who may engage with online eating disorder–related content. Users aged 16 and 17, like those aged 18 and over, were considered to be sufficiently mature to consent for themselves without parent or guardian consent required according to our ethics approval. Users completed the surveys anonymously—they were not required to provide their name or any potentially identifying information at any point. No compensation was offered for completing surveys or using the chatbot.

Study Design

This study used a naturalistic mixed methods design comprising 2 complementary data structures: (1) cross-sectional user evaluation surveys and (2) within-session pre-post microintervention assessments embedded within chatbot interactions. A minimum of 34 completed responses per microintervention was used as a pragmatic threshold intended to support estimation of moderate effect sizes [37], consistent with prior work [31,34], and the exploratory nature of this real-world design [38].

Participants

Participants were individuals aged 13 years and older residing in Australia or Canada who engaged with the JEM chatbot from January 2023 to June 2023 in Australia or September 2024 to March 2025 in Canada. Geographic location was based on self-reported residence and deployment platform context within an anonymous web-based design. These time periods covered the first 6 months of JEM deployment in each country for optimal comparability, particularly in terms of the novelty of the chatbot service to the users in the respective countries [39]. In this 6-month period, system-level usage included 6862 total chatbot sessions and 39,184 interactions in the Australian deployment and 1332 sessions and 5370 interactions for the Canadian deployment according to Google Dialogflow analytics definitions [40], which are distinct from individual-level survey respondents analyzed in this study. Study participants were recruited from individuals already engaging with the JEM chatbot. All surveys (overall evaluation and microintervention surveys) were open surveys, and a self-selection sampling method was used, representing a convenience sample. No direct contact with potential participants was made by any members of the research team. Recruitment occurred in writing through the chatbot's "Provide Feedback/Feedback Survey" and "Skill Survey" conversation options, which users could select voluntarily. These options led to web-based surveys hosted by Qualtrics [41]. Surveys were administered with standard Qualtrics platform-level fraud-prevention settings applied to reduce duplicate responses (eg, cookie-based controls) [42,43], although these do not fully eliminate the possibility of repeat submissions in anonymous online settings. Only completed questionnaires were analyzed. Due to the fully anonymous design of the chatbot and surveys, individual users could not be tracked across sessions or timepoints. Therefore, engagement is reported at the session level, and survey participation represents voluntary, self-selected responses from the overall user base (see Multimedia Appendix 1 for aggregate study flow). For users who initiated the overall survey, completion rates were 80.1%, and for users who initiated microintervention surveys, completion rates ranged from 70.2% to 75.0%. There were no significant differences between the Australian and Canadian chatbot deployments for these completion rates (all *Ps*>.05).

XSL•FO
RenderX

Intervention

JEM was a web-based and rule-based chatbot powered by the Google Dialogflow platform [23]. The chatbot was available 24 hours per day, 7 days a week to users in Australia via a Monash University website (Figure 1) and in Canada via the National Eating Disorder Information Centre website (Figure 2). These deployment characteristics are reported to contextualize accessibility and real-world use; however, the primary focus of the intervention is the structure and content of the psychoeducational and microintervention pathways delivered within the chatbot. The chatbot was free of charge to access and users could converse with the chatbot as long as they wished to do so. There was no login or authentication required to use JEM, and usage was completely anonymous. Users were cautioned on these websites that the JEM chatbot was not monitored by any humans, and if they were experiencing an emergency, they should contact country-specific emergency services with contact details provided.

Figure 1. Australian version of the JEM chatbot hosted on a Monash University webpage. Users clicked on the blue speech bubble icon to start the conversation.



XSL•FO
RenderX

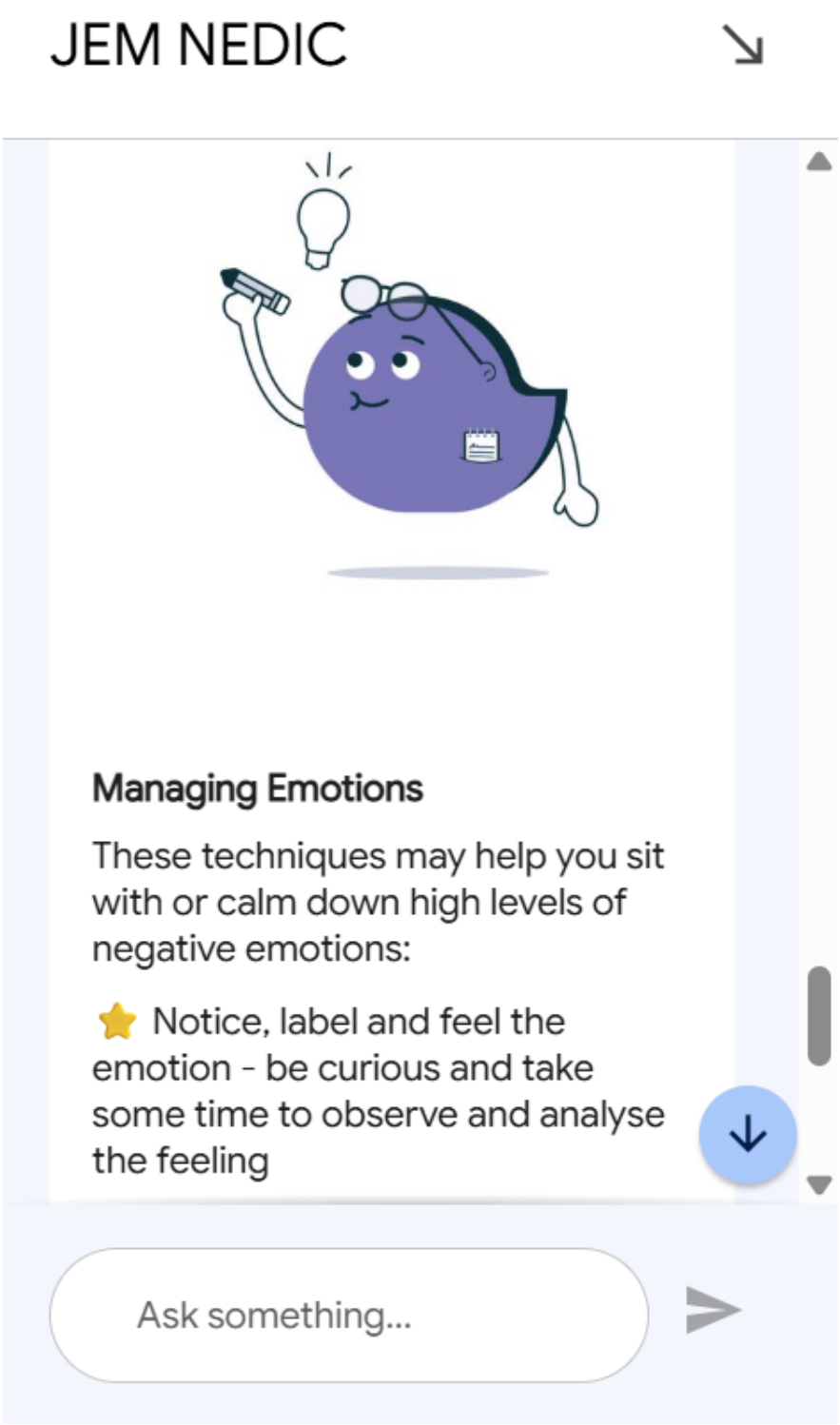
Figure 2. Canadian version of the JEM chatbot hosted on the National Eating Disorder Information Centre website. Users clicked on the purple JEM icon to start the conversation.



The chatbot's conversation content was based on evidence-based information or interventions for eating disorders, specifically psychoeducation, CBT, ACT, and mindfulness (for detailed information, see Beilharz et al [21]). Owing to the short conversational style of the JEM chatbot, the microinterventions taught by JEM were carefully selected, such as for CBT, education on cognitive distortions, for ACT, cognitive defusion exercises, and mindful breathing for mindfulness (see Figures 3 and 4 as examples). All microinterventions were estimated to take less than 5 minutes to complete. The chatbot operated through a rule-based decision-tree structure, where user inputs were matched to predefined intents that triggered specific conversational pathways. Users engaged with JEM via either free-text inputs or menu-based options, which routed them to psychoeducation content or presented a set of available microintervention scripts. The 10 microintervention scripts were discrete, standardized conversational pathways that users could actively select from a menu of available skills. Each microintervention followed a fixed sequence of prompts and responses with no within-script branching. These scripts were delivered in multimodal formats, including text, images, audio, and video. Users self-selected into microinterventions during their interaction with the chatbot, and each script could be completed within a single session or revisited multiple times. Completion was operationalized as reaching the end of the scripted conversational pathway.

XSL•FO
RenderX

Figure 3. Example conversation from the JEM chatbot for the "Managing Emotions" microintervention (Canadian version).



There were no major conceptual differences in intervention content and technical delivery between the Australian and Canadian deployments. In terms of safety, natural language processing was used for JEM to detect user-typed risk phrases such as "I want to kill myself," and "I wish I was dead," and JEM responded with 24/7 crisis support services specific to Australia or Canada. Specifically, risk detection was implemented using Google Dialogflow's intent classification and entity extraction [44] to identify free-text inputs indicative of self-harm and suicidality. When a high-risk intent or entity was detected, the system immediately triggered a predefined response flow that bypassed standard conversational pathways and directed users to country-specific 24/7 crisis support services. There was also a "Get Urgent Help Now" option for users to select where the chatbot responded with these same crisis support services. If a user was to type a prompt that the chatbot could not match to a preprogrammed answer, the chatbot responded with "I'm just a simple bot so I'm still learning how

XSL•FO
RenderX

to respond to your typed message. Try using my buttons to have a conversation with me" along with crisis support service contacts. There were no adverse events reported by users during the study period.

XSL•FO
RenderX

Figure 4. Example conversation from the JEM chatbot for the “Feeling Worthy” microintervention (Australian version).

XSL•FO
RenderX

Measures

Surveys

There were 2 types of web-based surveys, hosted by Qualtrics [41], where the survey weblinks were included within the JEM chatbot's conversation: (1) overall evaluation survey accessed via the "Provide Feedback" option and (2) skill surveys presented for every microintervention conversation (10 in total). There were no "back" options on any of the surveys for people to change their earlier responses.

Overall Evaluation Survey

The content and structure of this survey were based on our previous chatbot co-design research [20,21]. These items were designed as brief, study-specific indicators of user experience within an embedded chatbot context and were not intended as stand-alone psychometric instruments. In the first section of the survey, demographic characteristics were asked, particularly age, gender identity, LGBTIQA+ community membership (lesbian, gay, bisexual, transgender, intersex, queer, asexual people, or people otherwise diverse in gender, sexual orientation, and/or innate variations of sex characteristics), and ethnicity. Note that ethnicity was asked in a different format in Australia compared to Canada to suit the standard ethnicity demographics of the country [45]. The second section of the survey included categorical items addressing (1) reasons for using the chatbot, (2) whether the user found the information they were looking for, and (3) whether they intended to seek further support. The third section involved a single-item user satisfaction sliding scale ranging from 0 ("not at all satisfied") to 100 ("completely satisfied"). This satisfaction item was followed by an optional open-text question: "Please let us know your reason(s) for your satisfaction rating." The fourth and final section of the survey used the validated System Usability Scale (SUS) [46], which comprised 10 items that reflected various statements related to system usability (eg, "I thought the system was easy to use"). Responses were recorded on a 5-point Likert scale ranging from 1 ("strongly disagree") to 5 ("strongly agree"). The SUS score yielded a single number between 0 and 100, with higher scores representing higher perceived usability of the system. The SUS demonstrated high reliability in this study's sample (Cronbach α=0.89 for both Australian and Canadian samples).

Microintervention Survey

For every microintervention included in the JEM chatbot (10 in total, see Beilharz et al [21] for comprehensive descriptions), there was a web-based survey offered. If the user decided to complete the survey, they were asked for their current mood ratings for how happy, confident, anxious, and depressed they felt, with each being rated on a visual analog scale ranging from 0 ("not at all") to 100 ("very much"). The respondents were then asked to similarly rate for body image satisfaction on visual analog scales ranging from 0 ("not at all") to 100 ("very much"), specifically physical attractiveness, body size satisfaction, and body shape satisfaction. Note that we did not ask demographic characteristics of respondents in the microintervention surveys. Single-item visual analog scales were used to assess momentary mood and body satisfaction to minimize respondent burden and enable repeated, brief assessments within a chatbot-based design. Single-item visual analog scale measures are widely used for capturing transient, state-level affective and body image experiences and have been shown to be sensitive to within-person changes over time [47-50]. Importantly, prior research indicates that single-item measures of body satisfaction are responsive to short-term changes in both observational and online intervention contexts [51-53]. Respondents were instructed to complete the microintervention and then return to the survey immediately afterward, where they were then presented with the same mood and body image satisfaction visual analog scales again. Thus, the before and after microintervention ratings were completed within the same survey for each microintervention.

Data Analysis

The Statistical Package for Social Sciences (SPSS; version 31; IBM Corp) [54] was used for quantitative statistical analysis. Descriptive statistics were used to analyze demographic characteristics and user satisfaction and experiences. To explore if there were differences between countries, 2-tailed independent samples *t* tests were used for continuous data and Fisher's exact tests for categorical data (2-tailed). To examine within-session pre-post microintervention changes in mood and body image, Cohen *d* was calculated as a measure of effect size [37], with 95% CIs reported. Analyses were conducted at the survey or session level due to the fully anonymous design of the chatbot, which did not allow the linkage of responses across users or timepoints. Any missing data were handled with listwise deletion, acknowledging that missingness in optional, anonymous web-based surveys may not be completely at random. Given the exploratory and hypothesis-generating nature of these analyses, no formal correction for multiple testing was applied [55]. Applying such corrections in this context may have increased the risk of type 2 error and obscured potentially meaningful signals warranting further investigation. The results were therefore interpreted cautiously, acknowledging the increased risk of type 1 error associated with multiple comparisons. The open-ended qualitative text responses addressing reasons for chatbot satisfaction ratings were analyzed using an abbreviated content analysis approach [56], given the brevity of the responses (30 words maximum). GS conducted initial coding, with HH independently reviewing and refining codes. All authors agreed on the final content interpretations.

Results

User Characteristics

Across all analyses, reported Ns refer to completed survey instances (survey responses). As shown in Table 1, from the overall evaluation surveys, respondents were, on average, in their early 20s, but age distribution varied substantially, particularly in the Australian sample. The respondents predominantly identified as girls or women, followed by nonbinary individuals. Around two-thirds of respondents did not identify as LGBTIQA+; however, when they did, queer was the most common identity. There were no differences between Australian and Canadian-based respondents for any of these demographic characteristics (all *P* values >.05). The majority of respondents were European or White in Canada, and around

XSL•FO
RenderX

a quarter identified as culturally and linguistically diverse (CALD) in Australia. Owing to the different style of questioning required, ethnicity was not compared between countries.

Table . Respondent demographic characteristics for overall evaluation survey by country (N=230 total).

Demographic characteristic	Australia (n=122)	Canada (n=108)
Age (y), mean (SD; range)	21.0 (12.0; 13 - 64)	22.8 (6.8; 14 - 43)
Gender, n (%)		
Girl or woman	99 (81.1)	87 (80.6)
Boy or man	7 (5.7)	6 (5.6)
Nonbinary	14 (11.5)	12 (11.1)
Transgender woman	0 (0)	0 (0)
Transgender man	1 (0.8)	0 (0)
Prefer not to answer	1 (0.8)	3 (2.8)
Ethnicity		
Australia CALD[a], n (%)		
No	89 (73.0)	—[b]
Yes	31 (25.4)	—
Prefer not to answer	2 (1.6)	—
Canada, n (%)		
European or White	—	81 (75.0)
African or Black	—	3 (2.8)
Latin, South or Central American	—	3 (2.8)
East Asian	—	3 (2.8)
“Mixed”	—	6 (5.6)
Prefer not to answer	—	12 (11.1)
LGBTIQA+[c] n (%)		
No	81 (66.4)	72 (66.7)
Bisexual	4 (3.3)	6 (5.6)
Lesbian	5 (4.1)	3 (2.8)
Queer	17 (13.9)	15 (13.9)
Gay	10 (8.2)	6 (5.6)
Asexual	0 (0.0)	3 (2.8)
Prefer not to answer	5 (4.1)	3 (2.8)

[a]CALD: culturally and linguistically diverse.

[b]Ethnicity was assessed using country-specific response categories. Blank cells indicate categories that were not applicable to that country.

[c]LGBTIQA+: lesbian, gay, bisexual, transgender, intersex, queer, asexual people, or people otherwise diverse in gender, sexual orientation, and/or innate variations of sex characteristics.

Experiences and Satisfaction

As indicated in Table 2 from the overall evaluation surveys, the most common reasons in both countries for using JEM chatbot were that the respondents thought they may have an eating disorder, had issues finding information, and were simply curious about the chatbot. Respondents found the information they were looking for the vast majority of the time, and the majority intended to seek or maybe seek further support after using the chatbot. The overall satisfaction rating for the JEM chatbot was high on average, and the SUS rating was in the “excellent” category for both Australia and Canada [46]. There were no country-based significant differences for any of these findings (all *P* values >.05).

For the open-ended qualitative responses to reasons for the satisfaction ratings, this question was completed by a self-selected subset of respondents (Australia n=16; Canada n=12), which did not significantly differ between countries (all *P* values>.05). Given the lower response rates, responses were analyzed in aggregate rather than by country (Table 3). The response themes were mostly positive in nature, with the most

XSL•FO
RenderX

common focused on JEM chatbot providing helpful information and being easy to use.

Table . Responses to overall evaluation survey by country (N=230 total).

Measure	Australia (n=122)	Canada (n=108)
Reasons for using the chatbot, n (%)[a]		
May have an eating disorder	71 (58.2)	61 (56.5)
Trouble finding information via other sources	32 (26.2)	29 (26.9)
Curious about the chatbot	29 (23.8)	27 (25.0)
Chatbots do not judge	13 (10.7)	12 (11.1)
Not ready to speak to a real person	10 (8.2)	9 (8.3)
Find information, n (%)		
Yes	112 (91.8)	100 (92.6)
No	10 (8.2)	8 (7.4)
Seeking further support, n (%)		
Yes	79 (64.8)	77 (71.3)
Maybe	15 (12.3)	14 (13.0)
No	14 (11.5)	12 (11.1)
I already have support	14 (11.5)	5 (4.6)
Satisfaction rating, mean (SD; range)	76.1 (22.7; 20.0 - 100)	78.8 (14.3; 30.0 - 100)
SUS[b] rating, mean (SD; range)	86.5 (16.9; 20.0 - 100)	89.5 (11.6; 40.0 - 100)

[a]Percentages add up to greater than 100% as respondents could choose multiple responses.

[b]SUS: System Usability Scale.

Table . Themes for open-ended qualitative responses for reasons for satisfaction ratings by country (N=28 total).

Theme	Category	Total, n (%)[a]	Australia, n	Canada, n
Chatbot provided helpful information	Positive	18 (64.3)	10	8
Chatbot was easy to use	Positive	16 (57.1)	10	6
JEM character was fun or cute	Positive	7 (25.0)	3	4
Prefer a chatbot over a helpline	Positive	6 (21.4)	6	0
Chatbot had a friendly tone	Positive	4 (14.3)	0	4
Difficult to find the information sought	Negative	2 (7.1)	2	0
Chatbot conversation was targeted at too young an audience	Negative	1 (3.6)	0	1

[a]Percentages add up to greater than 100% as qualitative responses could be classified into multiple themes.

Changes to Mood and Body Image

Four of the 10 total microinterventions included in the JEM chatbot's conversation yielded sufficient completed survey responses (n≥34) from either country to conduct suitable within-session pre- or postchange analyses (Table 4). These 4 microinterventions were (1) Managing unhelpful thoughts (CBT and ACT strategies), (2) Managing emotions (ACT strategies), (3) Challenging beauty (CBT strategies), and (4) Feeling worthy (CBT strategies). Sample sizes for individual microintervention analyses varied and were modest for some subgroups; therefore, the findings should be interpreted with appropriate caution.

Across microinterventions, the within-session pre-post ratings were broadly similar in Australia and Canada, with effect sizes ranging from very small to large. The Managing unhelpful thoughts microintervention showed very small or small effect sizes across most measures, with moderate effects observed for anxiety. The Managing emotions microintervention showed the largest within-session effect sizes for depression, with moderate

XSL•FO
RenderX

effects observed for body size satisfaction and the remaining measures showing very small-to-small effects. The Challenging beauty microintervention showed very small-to-moderate effects across measures, with confidence and physical attractiveness showing the largest effect sizes. The Feeling worthy microintervention showed very small-to-moderate effects across measures, with confidence, physical attractiveness, body size satisfaction, and body shape satisfaction showing the largest effect sizes.

XSL•FO
RenderX

Table . Responses to mood and body image measures before and after microintervention surveys by country.

Measure	Australia			Canada		
	Before, mean (SD)	After, mean (SD)	Cohen *d* (95% CI)	Before, mean (SD)	After, mean (SD)	Cohen *d* (95% CI)
Microintervention: Managing unhelpful thoughts (n=146 for Australia and n=130 for Canada)						
Happy	28.9 (21.9)	29.4 (24.2)	0.03 (−0.14 to 0.19)	29.7 (21.7)	29.2 (23.1)	−0.03 (−0.20 to 0.14)
Confident	20.0 (20.0)	23.4 (21.8)	0.20 (0.04 to 0.37)	20.5 (19.8)	23.0 (20.7)	0.16 (−0.01 to 0.34)
Anxious	63.0 (28.7)	53.9 (30.9)	−0.45 (−0.62 to −0.23)	64.1 (28.3)	54.4 (30.4)	−0.53 (−0.70 to −0.34)
Depressed	50.4 (31.4)	45.3 (32.0)	−0.25 (−0.42 to −0.09)	51.2 (30.7)	45.7 (31.7)	−0.29 (−0.46 to −0.11)
Physically attractive	21.9 (24.5)	22.9 (24.3)	0.06 (−0.10 to 0.23)	22.1 (24.4)	22.2 (23.0)	0.00 (−0.17 to 0.17)
Satisfied body size	17.8 (24.0)	20.4 (24.6)	0.20 (0.03 to 0.37)	18.3 (24.0)	20.0 (23.2)	0.15 (−0.02 to 0.33)
Satisfied body shape	17.8 (24.5)	20.2 (26.1)	0.18 (0.01 to 0.34)	18.2 (24.6)	19.7 (25.2)	0.14 (−0.04 to 0.31)
Microintervention: Managing emotions (n=34 for Australia and n=44 for Canada)						
Happy	36.6 (21.8)	38.3 (24.8)	0.14 (−0.20 to 0.48)	35.5 (21.2)	38.9 (23.6)	0.22 (−0.09 to 0.52)
Confident	24.2 (14.8)	26.0 (19.5)	0.12 (−0.22 to 0.46)	23.6 (15.6)	27.3 (20.2)	0.22 (−0.08 to 0.53)
Anxious	65.6 (26.4)	56.2 (28.0)	−0.38 (−0.73 to −0.03)	64.5 (26.9)	53.7 (28.3)	−0.43 (−0.73 to −0.11)
Depressed	48.8 (29.9)	34.1 (24.5)	−0.78 (−1.13 to −0.38)	48.4 (29.1)	31.9 (22.8)	−0.79 (−1.12 to −0.44)
Physically attractive	22.2 (22.8)	23.5 (26.3)	0.12 (−0.22 to 0.46)	20.9 (22.5)	22.8 (25.7)	0.18 (−0.12 to 0.48)
Satisfied body size	13.1 (17.6)	17.7 (21.7)	0.43 (0.07 to 0.78)	13.1 (17.1)	18.1 (20.8)	0.51 (0.19 to 0.82)
Satisfied body shape	17.5 (21.9)	19.2 (21.8)	0.19 (−0.15 to 0.54)	18.5 (22.5)	20.1 (22.9)	0.19 (−0.11 to 0.49)
Microintervention: Challenging beauty (n=50 for Australia and n=40 for Canada)						
Happy	40.5 (30.5)	43.6 (31.4)	0.15 (−0.13 to 0.43)	38.8 (29.2)	41.5 (29.5)	0.13 (−0.19 to 0.44)
Confident	23.5 (25.9)	33.6 (29.6)	0.47 (0.18 to 0.76)	25.2 (26.4)	32.2 (29.2)	0.38 (0.06 to 0.70)
Anxious	59.4 (31.1)	47.7 (33.5)	−0.44 (−0.73 to −0.15)	60.5 (30.5)	52.9 (33.0)	−0.32 (−0.63 to 0.00)
Depressed	46.7 (33.5)	40.4 (29.1)	−0.31 (−0.59 to −0.03)	49.2 (30.6)	43.7 (29.4)	−0.25 (−0.56 to 0.07)
Physically attractive	20.4 (25.5)	29.5 (32.1)	0.44 (0.15 to 0.73)	21.0 (26.6)	28.6 (30.9)	0.42 (0.09 to 0.74)
Satisfied body size	20.6 (27.7)	27.6 (32.7)	0.33 (0.04 to 0.62)	21.2 (29.8)	26.3 (31.9)	0.26 (−0.06 to 0.58)
Satisfied body shape	23.0 (27.7)	28.3 (30.9)	0.24 (−0.05 to 0.52)	22.6 (28.7)	27.1 (29.1)	0.21 (−0.11 to 0.53)
Microintervention: Feeling worthy (n=36 for Australia and n=39 for Canada)						
Happy	28.4 (21.1)	31.9 (24.0)	0.15 (−0.18 to 0.47)	24.3 (20.0)	31.0 (25.6)	0.23 (−0.09 to 0.54)
Confident	17.4 (16.1)	28.2 (22.5)	0.46 (0.11 to 0.80)	14.5 (15.5)	28.3 (25.1)	0.50 (0.16 to 0.83)
Anxious	63.5 (27.4)	51.2 (33.0)	−0.37 (−0.71 to −0.03)	65.4 (29.3)	55.1 (34.5)	−0.27 (−0.60 to 0.05)
Depressed	61.1 (26.1)	58.4 (29.0)	−0.10 (−0.43 to 0.23)	59.5 (26.4)	58.4 (29.6)	−0.03 (−0.35 to 0.03)

XSL•FO
RenderX

Measure	Australia			Canada		
	Before, mean (SD)	After, mean (SD)	Cohen *d* (95% CI)	Before, mean (SD)	After, mean (SD)	Cohen *d* (95% CI)
Physically attractive	12.6 (16.8)	21.7 (26.0)	0.47 (0.11 to 0.81)	10.0 (11.8)	21.7 (26.9)	0.47 (0.13 to 0.80)
Satisfied body size	8.9 (12.3)	16.8 (21.4)	0.40 (0.05 to 0.72)	7.8 (11.5)	19.2 (24.7)	0.45 (0.11 to 0.78)
Satisfied body shape	10.2 (13.5)	17.4 (22.8)	0.41 (0.06 to 0.74)	11.2 (15.5)	20.7 (27.3)	0.42 (0.09 to 0.74)

Discussion

Principal Findings

This study provides one of the first real-world evaluations of a mental health–focused chatbot deployed at scale across 2 countries and continents, providing insight into chatbot use under naturalistic conditions. Overall, the findings indicated that the JEM chatbot was well accepted by respondents who completed surveys in both Australia and Canada, demonstrated excellent usability, and showed immediate within-session differences in mood and body image ratings. Satisfaction was high, information needs were largely met, and most survey respondents reported an intention to seek or consider further support after using the chatbot, suggesting that JEM may function as a supportive resource that could potentially facilitate consideration of further support.

Across microinterventions, patterns of within-session pre-post ratings were broadly similar in Australia and Canada, with effect sizes generally ranging from very small to large. While the magnitude of effect sizes varied across microinterventions and outcomes, the largest effects were typically observed for depression, anxiety, confidence, and body size satisfaction outcomes. These findings reflect immediate within-session pre-post changes and should not be interpreted as evidence of causal or sustained intervention effects.

In contrast, effect sizes for happiness ratings were generally very small across the microinterventions included in the analyses. This may reflect the brief nature of the interventions, which are designed to reduce distress and maladaptive cognitions rather than induce immediate positive affect, and is consistent with theoretical expectations of CBT and ACT informed strategies [57-59].

Gender identity beyond the binary was represented in the sample, with approximately 11% of respondents identifying as nonbinary. All study analyses included these respondents within the overall sample rather than stratifying by gender. Overall, JEM's nongendered and inclusive framing of content [21] suggests potential applicability across gender identities, although future work should directly examine differential experiences among gender diverse users, particularly their responses to psychotherapeutic microinterventions.

Comparison With Prior Work

These findings extend the very limited literature on eating disorder– and body image–specific intervention chatbots, particularly in real-world contexts. Previous RCTs have demonstrated efficacy for Tessa, Topity, and ED ESSI under controlled conditions [15,16,19], while other chatbots have been evaluated only partially or at early stages of development [17,18]. This study extends this work by demonstrating that a rule-based chatbot can be deployed with high user satisfaction and usability in naturalistic settings. Furthermore, the patterns of within-session pre-post differences observed among JEM users aligned in direction with findings reported in more controlled trials, while recognizing the differences in study design.

Compared with prior naturalistic evaluations of mental health chatbots such as Wysa [31], this study focused on a more specific population and set of outcomes. The patterns of within-session pre-post differences in anxiety, depression, and body image satisfaction observed among JEM survey respondents were mostly consistent with those reported in broader mental health chatbot evaluations. These observations are consistent with the literature suggesting that well-designed conversational agents may be associated with positive user-reported outcomes in real-world observational settings [31,60] and align with broader discussions about the promise and fragility of AI-mediated mental health care in real-world settings [61].

Importantly, the present findings contrast with concerns raised following the real-world deployment of the Tessa chatbot in 2023 in the United States, where the chatbot deviated from its rule-based design and provided dieting and weight loss advice [35]. No adverse events or harmful outputs were reported by users during the study period. JEM's rule-based architecture and co-designed content were intended to minimize such risks; however, all safety outcomes were not assessed in this study.

The very small effects on happiness from microinterventions also align with previous digital intervention research, where reductions in negative symptoms are often more readily observed than increases in positive affect, particularly following brief, single-session interventions [59,62]. Together, these findings suggest that JEM performs in a manner consistent with existing evidence while addressing a notable gap in real-world evaluations of body image– and eating disorder–specific chatbots.

Recent RCTs of generative AI-based conversational agents further suggest both the promise and the complexity of chatbot-delivered mental health interventions, reinforcing the importance of safety-focused design and careful real-world evaluation [63]. While the vast majority of qualitative feedback in this study was positive, with respondents reporting helpfulness, ease of use, and an engaging conversational style, a minority had negative feedback. More broadly, rule-based

XSL•FO
RenderX

chatbot systems are often perceived as less flexible than emerging large language model (LLM)–based approaches, which offer more naturalistic dialogue [64]. This highlights an important design trade-off between maintaining clinically safe, structured responses and meeting evolving user expectations for more adaptive and personalized conversational experiences. Future JEM iterations could explore hybrid architectures that combine rule-based safety constraints with controlled generative or retrieval-based components to enhance conversational flexibility while maintaining clinical safety and content fidelity [65].

Limitations and Future Research

Several limitations should be acknowledged for this study. First, the absence of a control group precludes causal inference and means changes cannot be definitively attributed to the chatbot alone. Second, outcomes were assessed immediately before and after individual microinterventions, limiting conclusions about the durability of effects. The immediate pre- and postintervention design may also be susceptible to demand characteristics and expectancy effects. However, differences were not uniform across outcomes or microinterventions, suggesting variability in within-session responses across measures and intervention content. Third, sample sizes varied across microinterventions, with 4 yielding sufficient data for analysis, which may have influenced the pattern of the findings based on differential engagement. Additionally, reliance on self-report measures and voluntary survey completion introduces potential response bias and means that the results reflect only a subset of users. Due to the fully anonymous design of JEM, which did not include login, authentication, or persistent user identifiers, individuals could not be tracked across sessions or over time. As a result, it was not possible to construct a participant-level CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) style flow diagram [66] tracking individuals through the chatbot content or determine precisely where attrition occurred within the chatbot journey. Engagement data are therefore reported via text and diagram at the session level, limiting the assessment of repeat usage and longitudinal engagement. Furthermore, this study did not assess changes in specific eating disorder symptoms, eating behaviors, or clinical outcomes.

Future research should examine longer-term outcomes associated with repeated or sustained chatbot use, including whether these immediate within-session differences are associated with longer-term outcomes, such as eating disorder risk or engagement with formal care. Hybrid designs combining real-world deployment with comparison conditions may help balance ecological validity with internal rigor. Further investigation into engagement patterns, personalization, and differential effects across demographic subgroups is also warranted.

Conclusions

This study demonstrated that a co-designed evidence-based chatbot, JEM, can be deployed at scale across multiple countries, achieving high usability and user satisfaction. Immediate within-session differences in mood and body image ratings were observed following the completion of the microinterventions investigated. However, further controlled research is needed, and the longer-term effects are yet to be determined. The study findings were broadly similar across Australia and Canada. These results extend the limited real-world evidence for mental health and body image– or eating disorder–specific chatbots, demonstrating that rule-based systems with co-designed content can be acceptable to users. As digital mental health interventions evolve, future research should explore how different approaches, including rule-based and generative AI chatbots, can complement each other to provide scalable, safe, and personalized support. Collectively, these findings suggest that chatbots such as JEM may represent a feasible and acceptable low-barrier resource for individuals seeking information and support related to body image and eating disorder concerns.

Acknowledgments

The authors would like to thank the JEM chatbot users in Australia and Canada for their valuable study involvement.

ChatGPT (OpenAI) was used to assist with language editing and improving clarity and flow of the manuscript. The authors take full responsibility for the content.

Funding

The authors declared no financial support was received for this work.

Data Availability

The datasets generated and analyzed during the current study are not publicly available under the participant confidentiality conditions of ethics approval from the Monash University Human Research Ethics Committee (ID 26129). The corresponding author can be contacted with queries about the data.

Authors' Contributions

Conceptualization: GS
Data curation: GS, HH
Formal analysis: GS, HH

XSL•FO
RenderX

Investigation: GS
Methodology: GS, SM, ET
Project administration: GS, SM, ET
Software: HH
Supervision: GS
Writing – original draft: GS
Writing – review and editing: GS, SM, ET, HH

Conflicts of Interest

GS is a Section Editor for JMIR Mental Health. The JEM chatbot was originally developed at Monash University and is now owned and managed by GS, including the JEM® registered trademark. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Aggregate study flow for the Australian (A) and Canadian (B) JEM chatbot deployments. Ten microinterventions were available within JEM, of which 4 generated sufficient completed surveys to meet the prespecified threshold for analysis.
[PNG File, 44 KB - humanfactors_v13i1e90783_app1.png]

References

1. Diagnostic and Statistical Manual of Mental Disorders, 5th edition: American Psychiatric Association Publishing; 2022.
2. Hay P, Aouad P, Le A, et al. Epidemiology of eating disorders: population, prevalence, disease burden and quality of life informing public policy in Australia-a rapid review. J Eat Disord 2023 Feb 15;11(1):23. [doi: 10.1186/s40337-023-00738-7] [Medline: 36793104]
3. Qian J, Wu Y, Liu F, et al. An update on the prevalence of eating disorders in the general population: a systematic review and meta-analysis. Eat Weight Disord 2022 Mar;27(2):415-428. [doi: 10.1007/s40519-021-01162-z] [Medline: 33834377]
4. Shagar PS, Harris N, Boddy J, Donovan CL. The relationship between body image concerns and weight-related behaviours of adolescents and emerging adults: a systematic review. Behav change 2017 Dec;34(4):208-252. [doi: 10.1017/bec.2018.3]
5. Deloitte Access Economics. Paying the price report. : Butterfly Foundation; 2024 URL: https://butterfly.org.au/wp-content/uploads/2024/10/deloitte-au-eco-paying-the-price-second-edition-180724-new-Oct-24.pdf? [accessed 2026-07-03]
6. Obeid N, Coelho JS, Booij L, et al. Estimating additional health and social costs in eating disorder care for young people during the COVID-19 pandemic: implications for surveillance and system transformation. J Eat Disord 2024 Apr 26;12(1):52. [doi: 10.1186/s40337-024-01003-1] [Medline: 38671494]
7. Obeid N, Silva-Roy P, Booij L, Coelho JS, Dimitropoulos G, Katzman DK. The financial and social impacts of the COVID-19 pandemic on youth with eating disorders, their families, clinicians and the mental health system: a mixed methods cost analysis. J Eat Disord 2024 Mar 29;12(1):43. [doi: 10.1186/s40337-024-00986-1] [Medline: 38553730]
8. Löchner J, Carlbring P, Schuller B, Torous J, Sander LB. Digital interventions in mental health: an overview and future perspectives. Internet Interv 2025 Jun;40:100824. [doi: 10.1016/j.invent.2025.100824] [Medline: 40330743]
9. Torous J, Linardon J, Goldberg SB, et al. The evolving field of digital mental health: current evidence and implementation issues for smartphone apps, generative artificial intelligence, and virtual reality. World Psychiatry 2025 Jun;24(2):156-174. [doi: 10.1002/wps.21299] [Medline: 40371757]
10. Haque MDR, Rubya S. An overview of chatbot-based mobile mental health apps: insights from app description and user reviews. JMIR mHealth uHealth 2023 May 22;11:e44838. [doi: 10.2196/44838] [Medline: 37213181]
11. He Y, Yang L, Qian C, et al. Conversational agent interventions for mental health problems: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2023 Apr 28;25(1):e43862. [doi: 10.2196/43862] [Medline: 37115595]
12. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464. [doi: 10.1177/0706743719828977] [Medline: 30897957]
13. Feng X, Tian L, Ho GWK, Yorke J, Hui V. The effectiveness of AI chatbots in alleviating mental distress and promoting health behaviors among adolescents and young adults: systematic review and meta-analysis. J Med Internet Res 2025 Nov 26;27:e79850. [doi: 10.2196/79850] [Medline: 41313175]
14. Fu L, Burns R, Xie Y, et al. The development and use of AI chatbots for health behavior change: scoping review. J Med Internet Res 2026 Jan 28;28:e79677. [doi: 10.2196/79677] [Medline: 41604667]
15. Fitzsimmons-Craft EE, Chan WW, Smith AC, et al. Effectiveness of a chatbot for eating disorders prevention: a randomized clinical trial. Int J Eat Disord 2022 Mar;55(3):343-353. [doi: 10.1002/eat.23662] [Medline: 35274362]
16. Matheson EL, Smith HG, Amaral ACS, et al. Using chatbot technology to improve Brazilian adolescents' body image and mental health at scale: randomized controlled trial. JMIR mHealth uHealth 2023 Jun 19;11:e39934. [doi: 10.2196/39934] [Medline: 37335604]

XSL•FO
RenderX

17. Shah J, DePietro B, D'Adamo L, et al. Development and usability testing of a chatbot to promote mental health services use among individuals with eating disorders following screening. Int J Eat Disord 2022 Sep;55(9):1229-1244. [doi: 10.1002/eat.23798] [Medline: 36056648]
18. Gullo NA, Goldberg J, Howe CP, et al. User-centered development of a chatbot for diverse adolescents at high risk for eating disorders. Int J Eat Disord 2026 Apr;59(4):724-735. [doi: 10.1002/eat.70008] [Medline: 41392306]
19. Sharp G, Dwyer B, Randhawa A, McGrath I, Hu H. The effectiveness of a chatbot single-session intervention for people on waitlists for eating disorder treatment: randomized controlled trial. J Med Internet Res 2025 May 21;27:e70874. [doi: 10.2196/70874] [Medline: 40397933]
20. Sharp G, Dwyer B, Xie J, et al. Co-design of a single session intervention chatbot for people on waitlists for eating disorder treatment: a qualitative interview and workshop study. J Eat Disord 2025 Mar 11;13(1):46. [doi: 10.1186/s40337-025-01225-x] [Medline: 40069853]
21. Beilharz F, Sukunesan S, Rossell SL, Kulkarni J, Sharp G. Development of a positive body image chatbot (KIT) with young people and parents/carers: qualitative focus group study. J Med Internet Res 2021 Jun 16;23(6):e27807. [doi: 10.2196/27807] [Medline: 34132644]
22. Fitzsimmons-Craft EE, Rackoff GN, Shah J, et al. Effects of chatbot components to facilitate mental health services use in individuals with eating disorders following online screening: an optimization randomized controlled trial. Int J Eat Disord 2024 Nov;57(11):2204-2216. [doi: 10.1002/eat.24260] [Medline: 39072846]
23. Customer experience agent studio. Google Cloud. URL: https://cloud.google.com/products/conversational-agents?hl=en [accessed 2025-12-30]
24. Hart L, Phillipou A. COVID has presented unique challenges for people with eating disorders. They'll need support beyond the pandemic. The Conversation. 2020. URL: https://theconversation.com/covid-has-presented-unique-challenges-for-people-with-eating-disorders-theyll-need-support-beyond-the-pandemic-148903 [accessed 2025-03-03]
25. National mental health and wellbeing pandemic response plan. : Australian Government, National Mental Health Commission; 2020 URL: https://www.mentalhealthcommission.gov.au/sites/default/files/2024-03/national-mental-health-and-wellbeing-pandemic-response-plan.pdf [accessed 2026-07-03]
26. Mohr DC, Schueller SM, Riley WT, et al. Trials of intervention principles: evaluation methods for evolving behavioral intervention technologies. J Med Internet Res 2015 Jul 8;17(7):e166. [doi: 10.2196/jmir.4391] [Medline: 26155878]
27. Manole A, Cârciumaru R, Brînza R, Manole F. An exploratory investigation of chatbot applications in anxiety management: a focus on personalized interventions. Information 2025;16(1):11. [doi: 10.3390/info16010011]
28. Fan X, Chao D, Zhang Z, Wang D, Li X, Tian F. Utilization of self-diagnosis health chatbots in real-world settings: case study. J Med Internet Res 2021 Jan 6;23(1):e19928. [doi: 10.2196/19928] [Medline: 33404508]
29. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019;5:2055207619871808. [doi: 10.1177/2055207619871808] [Medline: 31467682]
30. Jabir AI, Martinengo L, Lin X, Torous J, Subramaniam M, Tudor Car L. Evaluating conversational agents for mental health: scoping review of outcomes and outcome measurement instruments. J Med Internet Res 2023 Apr 19;25(9):e44548. [doi: 10.2196/44548] [Medline: 37074762]
31. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR mHealth uHealth 2018 Nov 23;6(11):e12106. [doi: 10.2196/12106] [Medline: 30470676]
32. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 6;4(2):e19. [doi: 10.2196/mental.7785] [Medline: 28588005]
33. Ly KH, Ly AM, Andersson G. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. Internet Interv 2017 Dec;10(C):39-46. [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]
34. Nemesure MD, Park C, Morris RR, et al. Evaluating change in body image concerns following a single session digital intervention. Body Image 2023 Mar;44:64-68. [doi: 10.1016/j.bodyim.2022.11.007] [Medline: 36495690]
35. Jargon J. A chatbot was designed to help prevent eating disorders. Then it gave dieting tips. The Wall Street Journal. 2023. URL: https://www.wsj.com/articles/eating-disorder-chatbot-ai-2aecb179#comments_sector [accessed 2024-12-30]
36. Sharp G, Torous J, West ML. Ethical challenges in AI approaches to eating disorders. J Med Internet Res 2023 Aug 14;25:e50696. [doi: 10.2196/50696] [Medline: 37578836]
37. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
38. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
39. Talukdar N, Yu S. Breaking the psychological distance: the effect of immersive virtual reality on perceived novelty and user satisfaction. J Strateg Mark 2024 Nov 16;32(8):1147-1171. [doi: 10.1080/0965254X.2021.1967428]
40. Analytics. Google Cloud. URL: https://docs.cloud.google.com/dialogflow/es/docs/analytics [accessed 2025-12-30]
41. Qualtrics. URL: https://www.qualtrics.com [accessed 2025-12-30]

XSL•FO
RenderX

42. Fraud detection. Qualtrics. 2026. URL: https://www.qualtrics.com/support/survey-platform/survey-module/survey-checker/fraud-detection [accessed 2026-01-01]
43. Prevent multiple submissions. Qualtrics. 2026. URL: https://www.qualtrics.com/support/survey-platform/survey-module/survey-options/survey-protection/#:~:text=Prevent%20Multiple%20Submissions,-Qip%3A%20This%20option&text=This%20setting%20works%20by%20placing,them%20to%20take%20the%20survey [accessed 2026-01-01]
44. Intent matching. Google Cloud. URL: https://docs.cloud.google.com/dialogflow/es/docs/intents-matching#:~:text=When%20an%20end%2Duser%20writes,also%20known%20as%20intent%20classification [accessed 2026-04-30]
45. Guidance on demographic questions. McMaster University—Research & Innovation. URL: https://research.mcmaster.ca/home/support-for-researchers/ethics/mcmaster-research-ethics-board-mreb/guidance-on-demographic-questions [accessed 2026-01-01]
46. Brooke J. SUS: A 'Quick and Dirty' Usability Scale. In: Jordan PW, Thomas B, Weerdmeester WA, McClelland AL, editors. Usability Evaluation in Industry: Taylor and Francis; 1996:189-194.
47. Birkeland R, Thompson JK, Herbozo S, Roehrig M, Cafri G, van den Berg P. Media exposure, mood, and body image dissatisfaction: an experimental test of person versus product priming. Body Image 2005 Mar;2(1):53-61. [doi: 10.1016/j.bodyim.2004.11.002] [Medline: 18089174]
48. Tiggemann M, Zaccardo M. Mood and body dissatisfaction visual analogue scales [database record]. APA Psyc Tests. 2015. URL: https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft47509-000 [accessed 2026-07-16]
49. Cash TF. Cognitive-behavioral perspectives on body image. In: Cash TF, Pruzinsky T, editors. Body Image: A Handbook of Theory, Research, and Clinical Practice: Guilford Press; 2002:269-276.
50. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
51. Fuller-Tyszkiewicz M, Dias S, Krug I, Richardson B, Fassnacht D. Motive- and appearance awareness-based explanations for body (dis)satisfaction following exercise in daily life. Br J Health Psychol 2018 Nov;23(4):982-999. [doi: 10.1111/bjhp.12334] [Medline: 30039516]
52. Rogers A, Fuller-Tyszkiewicz M, Lewis V, Krug I, Richardson B. A person-by-situation account of why some people more frequently engage in upward appearance comparison behaviors in everyday life. Behav Ther 2017 Jan;48(1):19-28. [doi: 10.1016/j.beth.2016.09.007] [Medline: 28077218]
53. Sonneville KR, Calzo JP, Horton NJ, Haines J, Austin SB, Field AE. Body satisfaction, weight gain and binge eating among overweight adolescent girls. Int J Obes (Lond) 2012 Jul;36(7):944-949. [doi: 10.1038/ijo.2012.68] [Medline: 22565419]
54. IBM SPSS software. IBM. URL: https://www.ibm.com/spss [accessed 2025-12-30]
55. Perneger TV. What's wrong with Bonferroni adjustments. BMJ 1998 Apr 18;316(7139):1236-1238. [doi: 10.1136/bmj.316.7139.1236] [Medline: 9553006]
56. Bengtsson M. How to plan and perform a qualitative study using content analysis. Nurs Plus Open 2016;2:8-14. [doi: 10.1016/j.npls.2016.01.001]
57. Beck JS. Cognitive Behavior Therapy: Basics and Beyond, 2nd edition: Guilford Press; 2011.
58. Hayes SC, Strosahl KD, Wilson KG. Acceptance and Commitment Therapy: The Process and Practice of Mindful Change, 2nd edition: Guilford Press; 2011.
59. Schleider JL, Weisz JR. Little treatments, promising effects? Meta-analysis of single-session interventions for youth psychiatric problems. J Am Acad Child Adolesc Psychiatry 2017 Feb;56(2):107-115. [doi: 10.1016/j.jaac.2016.11.007] [Medline: 28117056]
60. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR Ment Health 2018 Dec 13;5(4):e64. [doi: 10.2196/mental.9782] [Medline: 30545815]
61. Clegg KA. "Feasible but fragile": an inflection point for artificial intelligence in mental health care. J Med Internet Res 2025 Dec 16;27:e89202. [doi: 10.2196/89202] [Medline: 41410285]
62. Weisz JR, Kuppens S, Ng MY, et al. What five decades of research tells us about the effects of youth psychological therapy: a multilevel meta-analysis and implications for science and practice. Am Psychol 2017;72(2):79-117. [doi: 10.1037/a0040360] [Medline: 28221063]
63. Heinz MV, Mackin DM, Trudeau BM, et al. Randomized trial of a generative AI chatbot for mental health treatment. NEJM AI 2025 Mar 27;2(4). [doi: 10.1056/AIoa2400802]
64. Wang L, Wan Z, Ni C, et al. Applications and concerns of ChatGPT and other conversational large language models in health care: systematic review. J Med Internet Res 2024 Nov 7;26:e22769. [doi: 10.2196/22769] [Medline: 39509695]
65. Kocaballi AB, Sezgin E, Clark L, et al. Design and evaluation challenges of conversational agents in health care and well-being: selective review study. J Med Internet Res 2022 Nov 15;24(11):e38525. [doi: 10.2196/38525] [Medline: 36378515]

XSL•FO
RenderX

66. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

ACT: acceptance and commitment therapy
AI: artificial intelligence
CALD: culturally and linguistically diverse
CBT: cognitive behavior therapy
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
LGBTIQA+: lesbian, gay, bisexual, transgender, intersex, queer, asexual people, or people otherwise diverse in gender, sexual orientation, and/or innate variations of sex characteristics
LLM: large language model
RCT: randomized controlled trial
SUS: system usability scale

Edited by A Kushniruk, S Law; submitted 07.Jan.2026; peer-reviewed by I Pepe, M Alcaraz-Ibanez; revised version received 23.Jun.2026; accepted 25.Jun.2026; published 20.Jul.2026.

Please cite as:
Sharp G, Marini S, Tam E, Hu H
Body Image and Eating Disorder Education Chatbot, JEM, in Australia and Canada: First 6-Month Real-World Survey Evaluation
JMIR Hum Factors 2026;13:e90783
URL: https://humanfactors.jmir.org/2026/1/e90783
doi:10.2196/90783

© Gemma Sharp, Sara Marini, Emily Tam, Hao Hu. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 20.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Strategic Use of Negative Emojis in Messaging-Based Interventions for Public Health Communication on Social Media: Mixed Methods Study

Yue Luo[1], PhD; Shubin Yu[2], PhD

[1]Department of Visual Communication Design, College of Fine Arts, Guangdong Polytechnic Normal University, Guangzhou, China

[2]Department of Communication and Culture, BI Norwegian Business School, Nydalsveien 37, Oslo, Norway

Corresponding Author:
Shubin Yu, PhD
Department of Communication and Culture, BI Norwegian Business School, Nydalsveien 37, Oslo, Norway

Abstract

Background: Despite the growing importance of social media in mobile health (mHealth) communication, we lack a clear understanding of how emotional elements like emojis shape message effectiveness. Furthermore, since emojis are inherently tied to text, their impact may be highly dependent on the relevance and context of the accompanying written content.

Objective: This study aimed to investigate how the valence of emojis affects the effectiveness of mHealth messages and to determine the role of emoji-text congruence in shaping message outcomes.

Methods: A mixed-method approach was used, encompassing 3 complementary studies. First, an analysis of real-world health-related tweets (N_1=257,648) quantified social media engagement in relation to positive, negative, and absent emojis. Building on these insights, 2 controlled experiments (N_2=220; N_3=190) further explored how negative emojis influence preventative health behaviors under varying levels of text congruence.

Results: The automatic content analysis revealed that messages containing negative emojis generated significantly higher social media engagement than those with positive emojis (β=0.24, SE=0.03, $t_{257,593}$=9.59; P<.001). Subsequently, experimental findings indicated that negative emojis can effectively promote preventative health behaviors (t_{218}=−4.15; P<.001). However, messages with high negative emoji-text congruence are more persuasive in promoting preventive behavior than those with low congruence ($F_{1,186}$=5.46, η^2=.028; P<.05), highlighting the complex interplay between emotional signaling and message consistency.

Conclusions: These findings advance our theoretical understanding of emoji-based health communication and provide practical guidelines for public health organizations seeking to optimize their mHealth messaging strategies. Our results highlight the need for careful consideration of both emotional valence and message coherence when designing health communications in the digital age.

(JMIR Hum Factors 2026;13:e78824) doi:10.2196/78824

KEYWORDS

emojis; preventive behavioral intention; emotional valence; congruence; mobile health communication; social media engagement

Introduction

Background

In mobile health (mHealth), disseminating accurate, comprehensible, and persuasive health information is crucial [1]. Campaigns, whether in the form of health posts or advertisements, aim to foster positive attitudinal and behavioral shifts, which is especially important when conveying protective measures against infectious diseases such as Ebola, COVID-19, and Monkeypox, or promoting routine practices like physical examinations [1]. Given the vast reach and interactive nature of social media, public health advocates increasingly leverage these platforms to communicate vital information [2-5]. Within these digital interactions, emojis have become ubiquitous, functioning as paralinguistic cues that supplement textual content and amplify emotional expression [6-8]. Scholarly attention has thus turned to understanding how emojis support messaging-based interventions in mHealth, particularly since research indicates that their inclusion can evoke more intense emotional responses compared to text alone, potentially enhancing behavioral intentions [9-17].

While several studies have demonstrated that the presence of emojis generally benefits health communication [13,16,18,19], most have focused on their mere presence versus absence, without delving into the nuanced effects of specific emoji characteristics. Emojis are not monolithic; they vary in

XSL•FO
RenderX

emotional valence and serve as potent social cues in online interactions [6]. In the context of health communication, positive emojis (eg, ,) are conceptually defined as symbols that convey approach-oriented or supportive emotions, signaling safety or encouragement to cope [15,20]. By contrast, negative emojis (eg, ,) express avoidance-oriented, alarming, or aversive emotions, signaling risk or concern [16,17]. The emotions as social information (EASI) model provides a foundational framework by suggesting that emotional expressions influence behavior via two main pathways: a direct affective reaction (eg, emotional contagion) and a cognitive inferential process (eg, deducing meaning or intent) [21]. Despite its strengths, scholars call for more comprehensive testing of the EASI model across digital contexts to better understand the specific persuasive impact of emoji valence in health messaging [6,21]. A clear grasp of how emoji valence operates is essential for designing more effective public health interventions.

Beyond shaping attitudes and behavioral intentions, a cornerstone of health communication efficacy [22-24], the success of health messages on social media is reflected in engagement metrics. Social media engagement is the extent to which users interact with online content through behaviors such as liking, commenting, and sharing [25-27]. These metrics indicate that a message has captured audience interest, stimulated discussion, and achieved widespread dissemination. Moreover, the congruence between a message's components—specifically, the emotional tone of an emoji and the accompanying text—is a known predictor of message credibility and persuasiveness [28,29]. Incongruence between emoji valence and textual content could compromise message clarity and reduce persuasive effectiveness.

This study addresses these gaps by examining the strategic use of negative emojis in public health communication through messaging-based interventions on social media. We integrate three complementary theoretical perspectives: the EASI model [21], which explains how emotional expressions shape behavior; emotional contagion theory [30], which details how negative emotions can rapidly spread within social networks; and schema-congruity theory [31,32], which elucidates how congruence between emojis and text facilitates multimodal cognitive processing. Guided by these theoretical underpinnings, we address the following hypotheses:

Effectiveness of Emoji Valence on Social Media Engagement

The affective pathway of the EASI model posits that observed emotions can induce similar feelings in recipients, while Emotional Contagion Theory [30] indicates that negative emotions tend to spread more rapidly, serving as potent threat signals [33,34]. Previous research has shown that negative emotional cues, even when conveyed through emojis, can heighten risk perception and, in turn, drive user engagement as audiences seek, share, or discuss the information [35-37]. In the context of health communication, negative emojis act as digital "alarm signals," motivating engagement through two complementary processes: (1) individual-level contagion (unconscious adoption of fear or concern) [30] and (2) social-level diffusion (sharing messages to alert others) [38]. Critically, negative emojis exhibit stronger contagion effects than positive ones.

Hypothesis 1: compared to positive emojis, using negative emojis in health-related social media posts will generate higher social media engagement.

Effectiveness of Emoji Valence on Preventive Behavioral Intention

Previous studies have leveraged emoji valence to inform individuals' perceptions and behaviors, such as emotional responses to food [39] and intentions to receive vaccination [16]. Preventive behavioral intention refers to the deliberate motivation to adopt health-protective actions and serves as a critical precursor to actual health behavior change [16,40-42]. Persuasive messaging often leverages emotional valence to amplify risk perception and motivate action [22,43]. Negative emojis (eg, ,) serve as condensed emotional cues that trigger a "risk-as-feelings" response, making abstract health risks more tangible and activating defensive motivations [43]. This process is consistent with the affective pathway anticipated by the EASI model and the human predisposition to give priority to negative signals as survival-relevant warnings [33,35].

Hypothesis 2: Compared to positive emojis, using negative emojis in health-related social media posts will increase individuals' preventive behavioral intentions.

Moderating Role of Emoji-Text Congruence on Preventive Behavioral Intention

In messaging-based health interventions on social media, emojis are often conveyed to users alongside textual messages. Schema-congruity theory [31,32,44] suggests that messages aligning with an individual's existing schemas are processed more smoothly. People tend to have more positive attitudes toward messages that exhibit high internal congruity [32,44]. For example, in high-risk contexts, recommendations are more effective when the visual content, textual message, and information source are mutually consistent [1,16]. Erle et al [6] assert that although emojis enhance the comprehensibility of information on social media, this effect is not realized solely through perceived valence. According to the EASI model's inferential pathway, the alignment between the emotional tone of emojis and the textual narrative builds a cohesive communicative schema. Negative emojis can enhance preventive intentions when they are congruent with risk-oriented text, whereas positive emojis may backfire when they are perceived as incongruent with the seriousness of the message. Although some degree of incongruence may introduce novelty effects [45,46], in health communication, a mismatch (eg, pairing positive emojis with severe warnings or negative emojis with overly reassuring texts) could undermine message credibility and clarity [38].

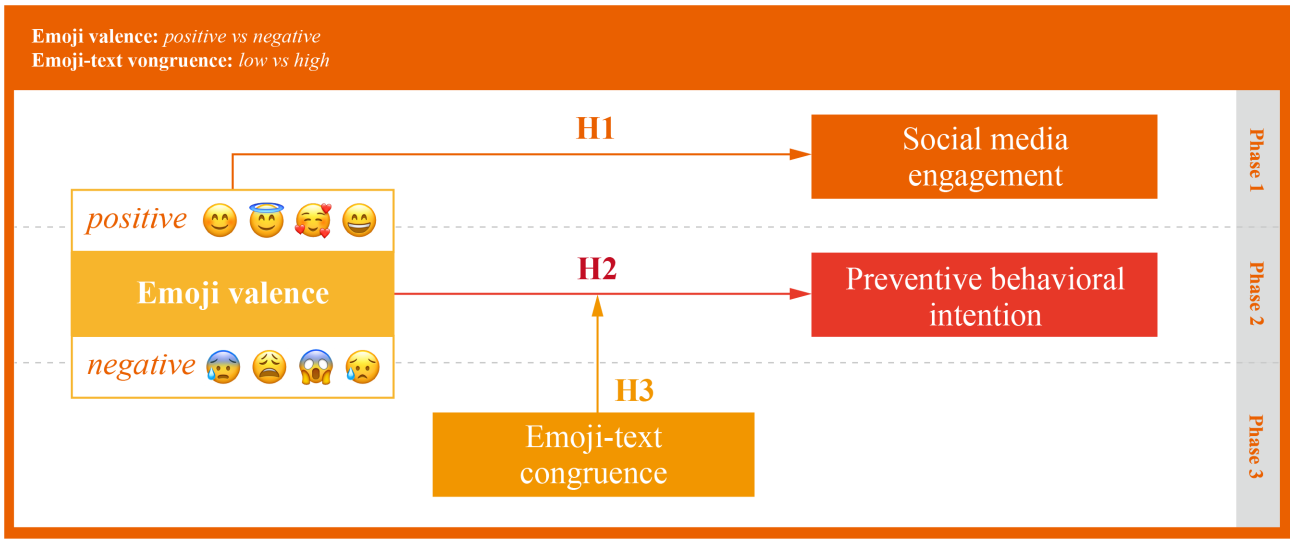
Hypothesis 3: the impact of emoji valence on preventive behavioral intentions is moderated by emoji-text congruence; specifically, negative emojis will be most effective when paired with text that is congruent in conveying risk or threat, while the

XSL•FO
RenderX

persuasive effectiveness of positive emojis may be diminished if this text conveys negative emotional framing.

The conceptual model of this study is presented in Figure 1.

Figure 1. The conceptual model and phase hypotheses.



Methods

Overview

To systematically test these hypotheses, we used a multimethod approach comprising an automated content analysis of real-world social media data and 2 online simulation experiments. To enhance the ecological validity of the study, we examined or simulated three phases of messaging-based health communication interventions across different social media platforms (Twitter, Facebook, and WeChat). These three phases cover both emergency public health crises and regular physical examination scenarios. This robust approach enables a comprehensive examination of how emoji valence and congruence impact both social media engagement and preventive behavioral intentions in health communication.

Ethical Considerations

This research received approval from the Academic Ethics Review Committee of the College of Fine Arts, Guangdong Polytechnic Normal University (approval number GPNU-GD25YYS34). Before participation, all participants were presented with an electronic informed consent form on the Credamo [47] platform describing the study purpose, voluntary nature of participation, data use, and compensation. Only those who provided digital consent were allowed to proceed. Participants were informed that they could withdraw from the survey at any time without penalty. All responses were collected anonymously (or de-identified before analysis) and analyzed in aggregate form only, and no personally identifiable information was retained by the research team. Those who completed the survey received compensation of 3 RMB (about US $0.44 as of July 16, 2026). The study complied with applicable institutional and national ethical guidelines for research involving human participants.

Phase 1

Data Collection

To examine how the emotional valence of emojis relates to public engagement with health messages, we collected 320,151 tweets from 48 prominent Twitter (now “X”) accounts specializing in health communication, including the World Health Organization, NPR Health, and NBC News Health (see Multimedia Appendix 1 for the full list). Data were retrieved using the academic research product track of the Twitter application programming interface (API) v2, which provides full-archive access to public tweets. Using the academictwitteR package in R (version 4.3; R Foundation for Statistical Computing), an automated script retrieved all tweets posted by each account between January 1, 2019, and December 31, 2021. This 3-year window was selected to capture health communication dynamics spanning the prepandemic period, the onset of COVID-19, and the ongoing pandemic, thereby enhancing temporal generalizability. For each tweet, the script extracted textual content, posting timestamp, originating account, and engagement metrics (likes, retweets, replies, and quote tweets). During data processing, 2 accounts contributing fewer than 10 tweets over the study period were excluded, yielding a final analytic sample of 46 accounts. Several preprocessing steps were applied to ensure data quality and analytic consistency. First, duplicate tweet records arising from API pagination were removed. Retweets (identified by the “RT @” prefix) were excluded, as their engagement metrics reflect the original author’s content rather than the retweeting account’s communication choices. Tweets in languages other than English were filtered out using the language metadata provided by the Twitter API, and tweets with fewer than 10 characters or empty text fields were discarded. Finally, to mitigate the influence of viral outliers on model estimates, tweets exceeding the 99th percentile on any individual engagement metric (likes, retweets, replies, or quote tweets) were trimmed. After these steps, 257,648 tweets remained for analysis (See Multimedia Appendix 2 for detailed documentation).

XSL•FO
RenderX

Sentiment Classification

The emotional valence of emojis within each tweet was classified using a three-stage automated cascade. First, each emoji character was scored using the VADER (Valence Aware Dictionary and Sentiment Reasoner) sentiment analyzer via the *VADER R* package, similar to the Emoji Sentiment Ranking [15]. When VADER could not assign a nonzero score to the raw character, the emoji's standardized Unicode name was retrieved using the emoji R package and re-scored through VADER. For the remaining unscored emojis, a keyword-matching heuristic compared the emoji name against curated lists of positive and negative semantic patterns. Emojis scoring above +0.05 were labeled positive, below −0.05 negative, and between these thresholds neutral. At the tweet level, tweets were then categorized into five groups based on the composition of their emoji labels: positive (n=16,861), negative (n=1312), neutral (n=13,470), mixed (containing emojis of differing valences; n=528), or none (no emojis present; n=225,477).

Measures and Variables

The dependent variable was social media engagement, operationalized as the sum of likes, retweets, replies, and quote tweets [26,27]. This composite score was log-transformed using log (x+1) to address the pronounced right-skew typical of engagement distributions [27]. We additionally analyzed each component—log likes, log retweets, and log replies—separately to assess consistency across engagement types. The independent variable was emoji valence, coded as a five-level categorical factor (negative, positive, neutral, mixed, or none). To strengthen internal validity, five tweet-level covariates were included: tweet character length, URL presence (binary), hashtag count, mention count, and a weekend posting indicator. Variance inflation factors ranged from 1.02 to 1.31, indicating no multicollinearity (Multimedia Appendix 2).

Analytical Approach

Because tweets are nested within accounts that vary in audience size, topical focus, and credibility, we used hierarchical linear models (HLM) with random intercepts for accounts. This approach partitions engagement variance into within-account and between-account components, preventing the conflation of emoji effects with account-level characteristics (see Multimedia Appendix 2 for model equations). The intraclass correlation coefficient (ICC) was computed from null (intercept-only) models to quantify between-account variation. We followed a sequential model-building strategy: Model 1 (M1) included a binary emoji presence indicator; Model 2 (M2) replaced this with the five-level valence factor; Model 3 (M3) added covariates to M1; and Model 4 (M4) combined the valence factor with all covariates. M4 served as the primary model for hypothesis testing. Models were estimated via restricted maximum likelihood (REML) using the *lme4* package, with Satterthwaite-approximated degrees of freedom from lmerTest. Model fit was compared using AIC and BIC (Multimedia Appendix 2).

Phase 2

Design and Procedure

This phase aimed to investigate the influence of different emoji valence in health messages on the intention to engage in preventive behaviors. An online experiment was conducted in November 2022, simulating a health message when a booster shot of the COVID-19 vaccine had been made available to the public. While ensuring the consistency of the message text, the health message introduced and persuaded participants about the vaccine booster shot by categorizing the emojis into two emotional valences (positive vs negative). Participants were sent a link to a survey in which they were invited to browse a webpage and sign an informed consent form. They were then exposed to one of two stimuli (ie, a positive emoji or a negative emoji). A piece of health information published by a fictitious health organization was displayed to the participants. Participants were asked to carefully read the information and evaluate their intention to receive the booster shot for COVID-19. Finally, the participants reported their demographic information, such as sex, age, and education.

Stimulus Material

The selection of several representative emojis was based on the Emoji Sentiment Ranking developed by Kralj Novak et al [15]. We invited four visual design professionals, including two visual designers and two design department teachers from universities. Four experts selected candidate emojis from the list of Apple-device emojis in the Emojipedia database [48]. The selection was based on three criteria: (1) the official descriptions and meanings displayed in the database; (2) the Emoji Sentiment Ranking developed by Kralj Novak et al [15]; and (3) the quality of their semantic and visual representativeness [49]. Ultimately, experts selected  and  as representatives of positive emojis, and  and  as representatives of negative emojis. After determining the four emoji representatives, we conducted a pretest for manipulation check purposes. The pretest was conducted online using Credamo, and participants were asked to evaluate whether "this message uses positive (or negative) emojis?" using a Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). We recruited 120 participants for the pretest (mean 29.39, SD 6.58), and 65 (54.2%) of them identified as women. The results of a *t* test showed that, compared to participants who saw negative emojis (mean 2.17, SD 1.77), participants in the positive emoji condition were more likely to believe that the emojis used in the information tended to express positive emotions (mean 6.20, SD 1.15, t_{118}=−14.8; P<.001). In addition, participants who received health information with negative emojis (mean 6.15, SD 1.52) were more likely to believe that the emojis were negative than those in the positive emoji condition (mean 1.67, SD 1.08, t_{118}=18.6; P<.001). These results provide convincing evidence for the manipulation, demonstrating that participants can correctly identify the emotional value of emojis.

We created a fictitious organization named "Health Online" and designed a Facebook post using Adobe Illustrator (Multimedia Appendix 3). The post was displayed in traditional Chinese characters, with the headline and content arranged

XSL•FO
RenderX

sequentially from top to bottom. The publishing unit, publishing time, and the body of the information were designed sequentially below. The content presented an overview of COVID-19 vaccine booster shots along with corresponding vaccination recommendations aimed at promoting immunization. Finally, emojis were incorporated into the messages, and the emotional valence of the emojis was manipulated to be either positive or negative. The emojis were positioned within the text, both in the middle and at the end. Apart from the manipulation of emojis, all other content remains consistent.

Participants

Phase 2 recruited 220 participants from Hong Kong, Macau, and Taiwan online via Credamo, and assigned them equally to one of the four scenarios. All participants were users of traditional Chinese characters, and 135 (61.4%) of them were women. Participants' ages ranged from 18 to 45 years old (mean 27.42, SD 6.32). Among them, 180 individuals (81.8%) hold bachelor's degrees or higher. All participants reported having prior experience reading Facebook posts.

Measurement Instrument

The scale to measure vaccination intention was adapted from Gerend and Shepherd's research [40] and was measured using a 7-point Likert scale. The first four items were anchored from 1 (strongly disagree) to 7 (strongly agree). The last two items were anchored from 1 (very unlikely) to 7 (very likely). These questions examined participants' willingness to receive the COVID-19 vaccine booster shot in the short and long term ()(Textbox 1). Higher scores indicate stronger preventive behavioral intention. We examined the reliability of the scale in this experiment (α=0.919).

Textbox 1. The reliability of the measurement.

Construct: Preventive Behavioral Intention (Adapted from Gerend & Shepherd [40])

Items:

- I will get more information about the COVID-19 vaccine booster shot
- I will consider getting a booster shot
- I will try the booster shot
- I will actually accept the booster shot
- How likely are you to accept the vaccine if hospitals or medical facilities offer the booster shot?
- Will you actively accept the vaccination if the time limit of the booster shot is every 6 months or 8 months?

Cronbach alpha: 0.919

Phase 3

Overview

The objective of study 3 was to further demonstrate the interaction effect of emoji valence and emoji-text congruence on health behavioral intentions. Study 3 is a 2 (emoji valence: positive versus negative) by 2 (emoji-text congruence: low versus high) between-subject experiments design.

Design and Procedure

Participants were informed that this study was conducted to evaluate information about physical examinations. We used scenarios related to physical examinations for the online experiment in April 2023. After signing the informed consent form, participants were provided with and read a health message about regular blood tests. The emojis embedded in the health message could be either negative or positive. Additionally, the message text was manipulated to have high or low congruence with the negative emojis while ensuring the meaning remained consistent (ie, positive emoji and low congruence, positive emoji and high congruence, negative emoji and low congruence, or negative emoji and high congruence). Participants were first prompted to assess the congruence between the emojis and the text in the health message (ie, "I feel that the emojis in the message are not consistent with the text"). Then their preventive behavioral intention was also assessed (ie, If hospitals or medical institutions offer regular check-up services, how likely am I to undergo regular physical examinations?). At the end of the experiment, participants reported their demographic information.

Stimulus Material and Participants

To enhance the generalizability of emojis and the robustness of cross-context in this study, 4 additional representatives, different from those in study 2, were selected for this phase of the experiment.  and  were chosen as positive emojis, while  and  were selected as negative emojis. Selecting additional emojis serves to strengthen external validity. The emotional valence of these emojis was ensured as they were also derived from previous studies [13,16,50]. In addition, to simulate the public's reception of health messages, we created a fictitious WeChat public account named "Health Smart Butler." The visualization of its health tweets was designed using Adobe Illustrator (Multimedia Appendix 4). All stimuli were displayed in Chinese, with components such as the title, the message publisher, message release date and time, the purpose of regular physical examination, and the role of blood tests. Next, the congruence of the body content with the negative emojis was categorized into "high" and "low" expressions. The content highlighted the benefits of regular physical examination, providing assurance for one's life. Moreover, the importance and usefulness of blood tests in these examinations were presented (specific stimulus texts are shown in the table below). To manipulate the emoji-text congruence, we placed the emoji in different locations. For the high fit with negative emojis

XSL•FO
RenderX

condition, emojis were placed after negative words such as risks, problems, lacking. In contrast, in the low fit with negative emojis condition, emojis were displayed next to neutral or positive words such as tests, healthy, crucial, testing. Please see Table 1 for detailed information.

A total of 190 online participants from mainland China were recruited through Credamo and randomly assigned to each experimental group. Participants' ages ranged from 19 to 57 years (mean 30.19, SD 6.62), with 113 (59.5%) identified as women. Among them, 174 individuals (91.6%) had completed a bachelor's degree or higher. All participants reported having prior experience reading health messages through WeChat public account posts.

Table . The text of the experimental stimulus.

Congruence	High (low) fit with negative (positive) emojis	Low (high) fit with negative (positive) emojis
Title	Regular blood tests, expose hidden health risks [emoji]	Regular blood tests [emoji], expose hidden health risks
Content	The main purpose of regular health check-ups for healthy people is to timely reveal potential health risks [emoji] and prevent the occurrence of diseases. Ignoring regular specialized check-ups may lead to early signs of diseases being overlooked, bringing unnecessary risks to your healthy life.	The main purpose of regular health check-ups for healthy individuals [emoji] is to timely reveal potential health risks and prevent the occurrence of diseases. Ignoring regular specialized check-ups may lead to early signs of diseases being overlooked, bringing unnecessary risks to your healthy life.
	In the annual health check-up, blood tests are especially crucial. Not having a blood test could mean that you are unaware of your own health condition, missing the key moment to identify potential common health problems [emoji]. During blood collection, multiple tubes are usually used for testing; missing the check-up could mean missing the chance to test your blood count, making it impossible to detect whether your body is lacking in iron, vitamins, etc [emoji] [emoji] [emoji]	In the annual health check-up, blood tests are especially crucial [emoji]. Not having a blood test could mean that you are unaware of your own health condition, missing the key moment to identify potential common health problems. During blood collection, multiple tubes are usually used for testing [emoji] [emoji] [emoji]; missing the check-up could mean missing the chance to test your blood count, making it impossible to detect whether your body is lacking in iron, vitamins, etc

Results

Phase 1

The final dataset comprised 257,648 tweets from 46 health-focused Twitter accounts. Of these, 32,171 (12.5%) contained at least one emoji: 16,861 positive (6.5% of all tweets), 13,470 neutral (5.2%), 1312 negative (0.5%), and 528 mixed-valence (0.2%). Overall, tweets received a median of 12 total engagements (mean 32.22, SD 49.96). Descriptively, tweets with negative emojis received higher mean likes (mean 39.5, SD 49.5) than tweets with positive emojis (mean 16.4, SD 36.4) or no emojis (mean 20.6, SD 32.5). Full descriptive statistics are presented in Multimedia Appendix 2.

Null-model ICCs indicated that between-account differences explained 48.5% of the variance in log total engagement, 46.1% in log likes, 45.4% in log retweets, and 28.4% in log replies, confirming the necessity of multilevel modeling.

Table 2 presents the primary hypothesis-testing model (M4), incorporating emoji valence and all covariates, for each engagement outcome. For log total engagement, tweets containing negative emojis received significantly higher engagement than tweets without emojis (β=0.25, *SE*=0.02, $t_{257,596.35}$=10.44; *P*<.001). Neutral (β=0.23, *SE*=0.01; *P*<.001) and mixed (β=0.36, *SE*=0.04; *P*<.001) emojis were also positively associated with engagement, whereas positive emojis showed no significant difference from emoji-free tweets (β=0.01, *SE*=0.01, $t_{257,610.93}$=1.52; *P*=.13). Critically, a planned contrast confirmed that negative-emoji tweets generated significantly higher engagement than positive-emoji tweets (Δβ=0.24, *SE*=0.03, $t_{257,595.11}$=9.59; *P*<.001), directly supporting H1. These patterns were consistent across engagement sub-metrics: negative emojis outperformed positive emojis in log likes (Δβ=0.17, $t_{257,595.36}$=6.70; *P*<.001), log retweets (Δβ=0.29, $t_{257,595.71}$=13.90; *P*<.001), and log replies (Δβ=0.14, $t_{257,599.22}$=9.24; *P*<.001). Among covariates, URL presence was the strongest predictor (β=0.93; *P*<.001), followed by weekend posting (β=0.11; *P*<.001), hashtag count (β=0.07; *P*<.001), and tweet length (β=0.01; *P*<.001), while mention count was negatively associated with engagement (β=−0.09; *P*<.001). The inclusion of covariates in M4 improved model fit substantially over M2 (ΔAIC=−47,903 for log total engagement), yet the negative emoji advantage persisted, indicating that the effect is not attributable to the tweet-level characteristics examined.

XSL•FO
RenderX

Table . Sequential model results for log total engagement. Reference category for emoji valence is "No Emoji." M4 is the primary hypothesis-testing model. The notable sign change for the positive-emoji coefficient from M2 (−0.243) to M4 (+0.012, n.s.) indicates that the apparent negative association of positive emojis in M2 was confounded with tweet characteristics (eg, positive-emoji tweets tended to contain fewer URLs and be shorter).

Predictor	M1: emoji presence	M2: emoji valence	M3: presence+ covariates	M4: valence + covariates
Intercept, β (SE)	1.93 (0.14)[b]	1.92 (0.14)[b]	0.51 (0.14)[b]	0.53 (0.14)[a]
Has emoji (binary), β (SE)	0.07 (0.01)[b]	—	0.12 (0.01)[b]	—
Negative, β (SE)	—	0.33 (0.023)[b]	—	0.25 (0.02)[b]
Neutral, β (SE)	—	0.40 (0.01)[b]	—	0.23 (0.01)[b]
Mixed, β (SE)	—	0.31 (0.04)[b]	—	0.36 (0.04)[b]
Positive, β (SE)	—	−0.24 (0.01)[b]	—	0.01 (0.01)
Tweet length, β (SE)	—	—	<0.01 (<0.001)[b]	<0.01 (<0.001)[b]
URL present, β (SE)	—	—	0.95 (0.01)[b]	0.93 (0.01)[b]
Hashtag count, β (SE)	—	—	0.07 (<0.01)[b]	0.07 (<0.01)[b]
Mention count, β (SE)	—	—	−0.09 (<0.01)[b]	−0.09 (<0.01)[b]
Weekend, β (SE)	—	—	0.11 (<0.01)[b]	0.11 (<0.01)[b]
σ^2 (account)	0.85	0.84	0.90	0.90
σ^2 (residual)	0.90	0.89	0.74	0.74
Akaike information criterion	705,115	701,986	654,547	654,083
Bayesian information criterion	705,157	702,059	654,641	654,208

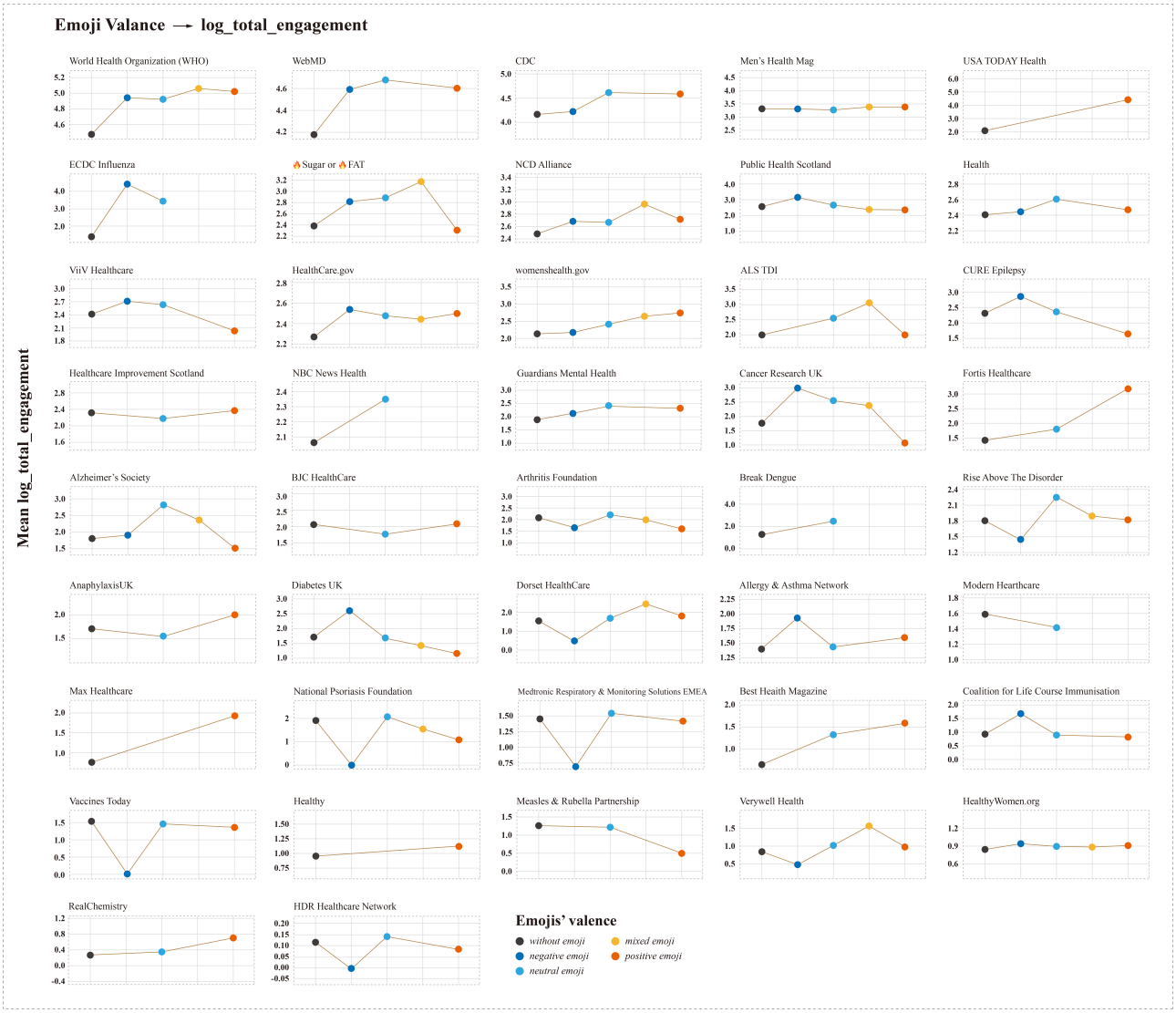
[a]*P*<.05.

[b]*P*<.001.

Although the pooled multilevel models show a clear average advantage for negative emojis, the account-specific patterns are not perfectly uniform. To address this heterogeneity, we examined per-account estimates of the negative-emoji coefficient (relative to positive or to no-emoji, depending on the model specification). Most accounts showed a positive association consistent with the pooled effect (Figure 2), but a small subset diverged from this pattern. Specifically, the negative-emoji effect was statistically indistinguishable from zero for accounts such as the "Health," "Arthritis Foundation," and "CDC," and it was directionally reversed for accounts like "WHO" and "National Psoriasis Foundation," where negative-emoji tweets were associated with lower engagement than comparable posts with positive emojis.

XSL•FO
RenderX

Figure 2. The multilevel effect of emoji valence on engagement. Four accounts were excluded from the figure due to the absence of tweets containing both negative and positive emojis.



Phase 2

A series of 2-tailed *t* tests were used to analyze the results. The result of the manipulation check showed that, compared to information with negative emojis (mean 2.32, SD 1.76), receivers perceived a post with positive emojis to contain positive emotions (mean 4.78, SD 1.75, t_{218}= −10.4; *P*<.001). Therefore, the manipulation of emojis in the experiment was successful.

We conducted another independent-sample 2-tailed *t* test to determine the effects of emoji valence on the intention to receive a booster shot. The results (Table 3) indicated a significant main effect of emoji valence on booster shot intention (t_{218}= −4.15; *P*<.001). Specifically, using negative emojis (mean 5.61, SD 0.99) increased participants' intention to receive booster shots more than using positive emojis (mean 4.88, SD 1.53). These findings support hypothesis 2 and, simultaneously, supplement the conclusions of study 1. The use of negative emojis in health messages not only enhances social media engagement but also increases people's willingness to receive booster shots.

Table . The results of *t* test.

Group	Mean (SD)	*t* test (*df*)
Preventive behavioral intention		218 (−4.15)[a]
Negative emojis (N=110)	5.61 (0.99)	
Positive emojis (N=110)	4.88 (1.53)	

[a]*P*<.001.

XSL•FO
RenderX

Phase 3

The results of a *t* test revealed that participants who received the messages with positive emojis (mean 6.34, SD 0.82) were more likely to perceive the emoji as conveying more positive emotional valence than did those in the negative emojis condition (mean 1.43, SD 0.72, $t_{186.89}$= −43.85; *P*<.001). Compared with participants who received messages containing positive emojis (mean 1.44, SD 0.77), those exposed to messages with negative emojis perceived the emotional valence communicated by the emojis as more negative (mean 6.55, SD 0.67, $t_{186.72}$=48.70; *P*<.001). Another manipulation check to assess whether emojis in messages matched the text was appropriately manipulated suggested that in messages with negative emojis, content that was perceived to be low in congruence with negative emojis (mean 4.02, SD 1.73) mismatched negative emojis more than those that were high in congruence with negative emojis (mean 3.11, SD 1.66, $F_{1, 186}$=7.62; *P*=.006). Additionally, among messages using positive emojis, content with high congruence to negative emojis (mean 2.82, SD 1.29) was perceived to have a lower match to positive emojis than content with low congruence to negative emojis (mean 3.52, SD 1.64, $F_{1, 186}$=4.81; *P*=.03). Therefore, both manipulations were successful.

We conducted a 2-way ANOVA analysis to test hypothesis 3. The results indicated that there was no significant main effect of emojis on preventive behavior intentions ($F_{1, 186}$=0.01, η^2<.001; *P*=.91), regardless of whether the health message contained positive (mean 6.03, SD 0.87) or negative emojis (mean 6.01, SD 0.87). On the other hand, the main effect of emoji-text congruence on preventive behavior intentions was also not significant ($mean_{low}$ 5.97, SD 0.87; $mean_{high}$ 6.07, SD 0.87, $F_{1, 186}$=0.58; *P*=.45). However, interestingly, a significant interaction effect on preventive behavior intentions was observed ($F_{1, 186}$=5.46, η^2=.028; *P*<.05), as shown in Table 4 and Figure 3. It showed that in health messages with negative emojis, messages with high negative emoji-text congruence (mean 6.11, SD 0.75) enhance people's intentions to engage in preventive behaviors more than those with low congruence (mean 5.83, SD 0.95). In contrast, when the emojis in the health messages are positive, participants' intention to undergo regular physical examinations became less favorable as the negative emoji-text congruence increases ($mean_{low}$ 6.22, SD 0.74 versus $mean_{high}$=5.91, SD 0.97). These findings support hypothesis 3 and, simultaneously, supplement the conclusions of studies 1 and 2. The use of negative emojis in health messages not only enhances social media engagement but also increases people's preventive behavior intentions, especially when negative emojis have high congruence with the text.

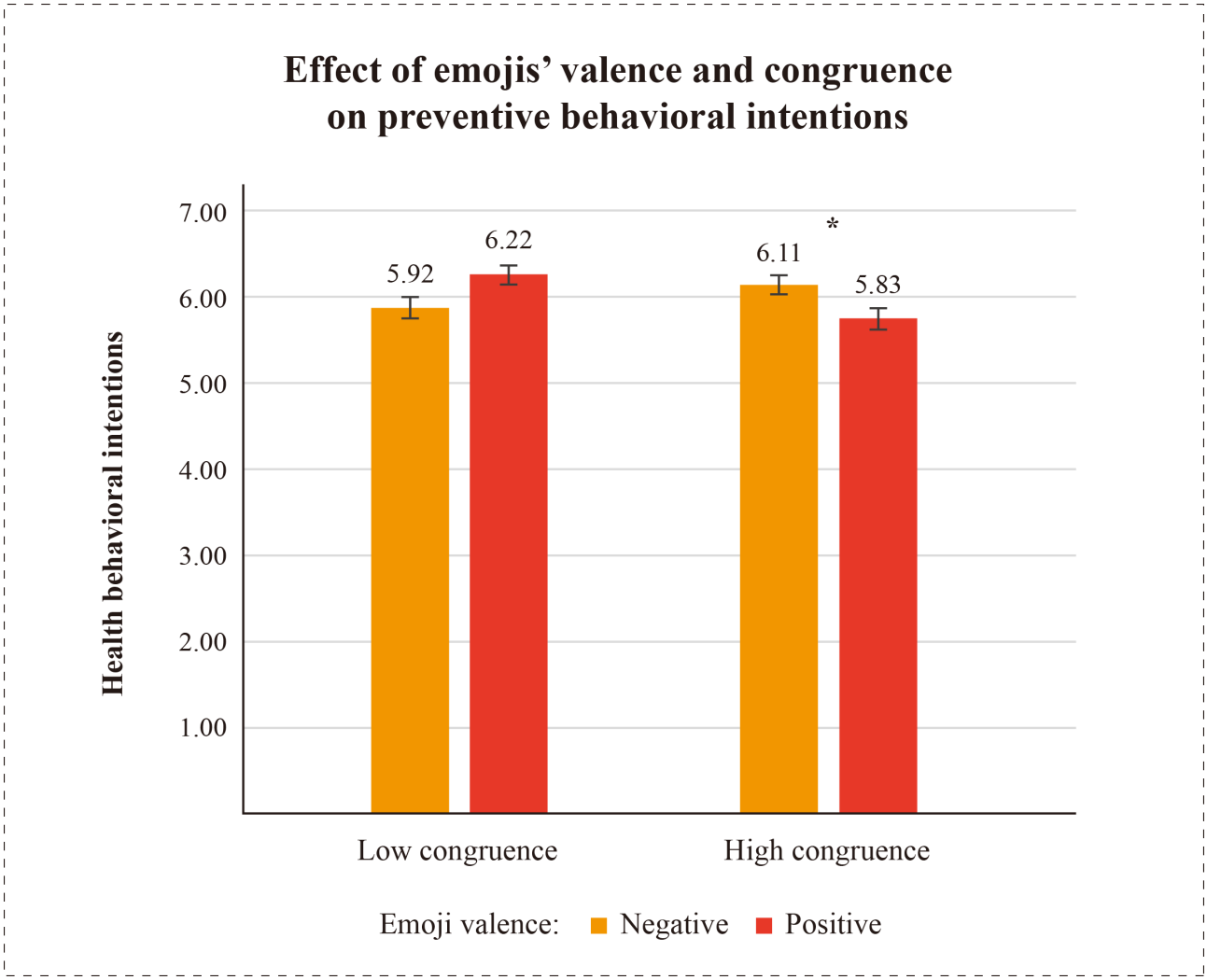
Table . Descriptive statistics of ANOVA.

Source	*F* (*df*)	*P* value	η^2
Intention to take regular physical examinations			
Emoji valence	0.01 (1, 186)	.91	<.001
Emoji-text congruence	0.58 (1, 186)	.45	.003
Emoji valence × emoji text congruence	5.46 (1, 186)	.02[a]	.028

[a]*P*<.05.

XSL•FO
RenderX

Figure 3. Interaction effect of emojis' valence and negative emoji-text congruence on preventive behavioral intentions. * *P*<.05.



Discussion

Principal Findings

This study used automated content analysis and experimental methods to examine the role of emojis in health communication. Specifically, we investigated the effects of emojis with different emotional valence on social media engagement, as well as how their congruence with text influences preventive behavioral intentions. Compared to positive emojis, the use of negative emojis in health messages generated more social media engagement and preventive behavioral intentions. Negative emojis demonstrated significantly stronger persuasive effectiveness than positive emojis or the absence of emojis in text-based mHealth interventions during the pandemic. However, this advantage depends on the congruence between the emojis and the accompanying textual message.

Comparison With Prior Work and Theoretical Contributions

We first indicate that the valence of emojis used in tweets has a significant impact on the social media engagement that tweets receive, even after accounting for the nested data structure through a random-intercept HLM. This finding was also robust when examined using an OLS model with account fixed effects and clustered standard errors. Specifically, tweets that used negative emojis received more social media engagement than tweets that used no emojis or positive emojis. Furthermore, the ICC of approximately 0.49 indicates that a substantial proportion of the variation in social media engagement is attributable to differences between Twitter accounts, underscoring the importance of accounting for this multilevel structure. Consistent with previous studies, unlike daily interactions on social media, negative visuals tend to capture individuals' attention and interest in health communication [16,35,37,51]. Users tend to interact and share health information with negative emojis, and these visuals with negative emotions evoke rich emotional experiences, including fear [52,53]. This result confirms that the information processing of emojis in health communication on social media provides users with a pathway for emotional reactions. Public health organizations can use emojis in health message design and disseminate them via social media to obtain more comments, likes, and shares, and to increase information coverage.

The findings of further experiments show that the use of negative emojis in health messages can elicit higher preventive behavioral intentions compared to positive emojis. This extends the view of previous research that the valence of emojis can guide individuals' perceptions and behavioral intentions

XSL•FO
RenderX

[16,19,39]. Compared with positive emojis, negative emojis may be more effective in motivating the public to engage in low-risk actions because they have the potential to activate heightened perceptions of risk and fear [14,16,17]. Also, this finding further demonstrates the effectiveness of using emojis for emotional contagion and health persuasion on social media [35,51]. We do not believe that the use of negative emojis in health messages is intended to immerse recipients in negative emotions. Instead, the negative emotions triggered by these emojis should be transformed by individuals into feelings of hope [52,53], thereby emphasizing the importance and necessity of preventive behaviors. This result further complements the emotional pathways provided by emojis in health communication and addresses the need, as proposed by Erle et al [6], to test the applicability of the EASI model within digital communication contexts.

In a postpandemic, regular physical examinations context, emoji valence alone is not sufficient to shape preventive behavioral intentions. The effect of emojis depends on their congruence with the textual framing. The use of negative emojis in health messages can increase preventive behavioral intentions, as demonstrated in the framework of high negative emoji-text congruence. According to the EASI model, the inferential process serves as the primary determinant of the subsequent behavioral effects of emotional expressions when the meanings of emojis are relatively vague and unclear [6]. Our findings also qualify previous work suggesting that a moderate level of incongruity in the schema can enhance evaluations by increasing novelty [45,46]. In the context of preventive health communication, the valence of negative emojis with low congruence to text presentation may appear less likely to be experienced as “interesting” and more likely to be interpreted as confusing or misaligned, thereby weakening individuals’ preventive intentions. Negative emojis in health messages with high emoji-text congruence maximize the consistency among the visual, emotional, and content elements conveyed to the user [16,19,54]. Naturally, individuals are more likely to positively and smoothly accept the emotions elicited by an object with high internal consistency, thereby translating those emotions into preventive behavior intentions. However, when negative emojis are replaced with positive ones, the effects of emotional inference become counterproductive due to the inconsistency in the schema. Negative emojis in a textual context of high negative emoji-text congruence may prompt the audience to devote more cognitive effort in processing the information, thereby increasing preventive behavior intentions. In real-world health communication, it is common for message design components to yield only small-to-moderate effect sizes because outcomes are shaped by multiple influences. However, when health messages are repeatedly delivered to large populations, even modest effects can accumulate and become highly consequential at the population level. Emojis, similar to facial expressions, provide two pathways—emotional reactions and inferential processes—in the dissemination of health information on social media platforms. More specifically, we build on the “emojis as social information” model proposed by Erle et al [6] and extend its applicability to the context of messaging-based interventions in mHealth. Emojis, as multimodal cues that convey both social information and visual semantics, interact with textual health messages to form multimodal content. This integration facilitates the interpretation, interaction, and acceptance of information by individuals on social media platforms.

Practical Implications

In addition to theoretical refinement, this study also offers several practical recommendations. First, health organizations and online health accounts are encouraged to use emojis as an interactive tool for health communication on social media. Second, using negative emojis not only increases audience engagement but also enhances individuals’ intentions to engage in preventive and health behaviors. Beyond the communication and interaction of health knowledge in daily life, the use of negative emojis is particularly effective for health messages that require strong public attention. For instance, during urgent health crises such as monkeypox or COVID-19 outbreaks, negative emojis can encourage individuals to adopt preventive behaviors like vaccination or promote the spontaneous communication of health messages. Finally, when using negative emojis in health messages, it is crucial to ensure congruence with the textual content. Properly placing emojis can maximize the persuasive power of health messages in mHealth and social media.

Limitations and Further Research

This study was deliberately designed to span multiple time points, demonstrating that emojis have influence on health communication across different phases of the pandemic and beyond, especially negative emojis. The first 2 phases were conducted in a high‐threat context (COVID-19), but the findings should not be interpreted as evidence that negative emojis are more persuasive than positive emojis across diverse health contexts. By contrast, in postpandemic settings such as regular physical examination scenarios, the persuasive impact of emoji valence depends on its congruence with the accompanying text. These results reveal the boundary conditions for generalizing beyond the pandemic period and highlight the need to explore the role of emojis across a broader range of health risks and cultural contexts. Further research should consider additional specific health contexts, such as sexual and reproductive health, depression, and anxiety. Currently, the study only compares the effects of positive and negative emojis, without examining neutral emojis and combinations of mixed valences. On social media, users often use emojis with different emotional valences to supplement their emotions and textual meanings [25,55,56]. Combinations of emojis with different emotional valences may become a direction for further research. This study relies on social media engagement and preventive behavioral intention as observed outcome indicators. The limitation highlights the need for future research to further test the applicability of the EASI model by incorporating mediating variables (eg, arousal and perceived ambiguity) and simultaneously represents an issue that the field of emoji research seeks to address. In addition, the visual stimuli used in this study were selected from the “Smileys” series of emojis. Future work should consider nonfacial expression emojis, such as gestures, occupations, and actions, to expand the applicability of emoji categories in health communication. Moreover, the

XSL•FO
RenderX

study participants focused on the Chinese-speaking region; future studies may investigate the effect of emoji valence in different cultures. Furthermore, the effect of emoji valence may also depend on different languages and texts, as well as the complexity of the information. Future studies may investigate how these contextual factors influence the effect of emoji valence on the public's perception, understanding, and attitudes.

Conclusions

This study demonstrates that emojis are not merely decorative elements in digital health communication, but persuasive multimodal cues that can shape both audience engagement and preventive behavioral intentions. Across the pandemic and post-pandemic contexts, negative emojis were found to be especially effective in attracting social media engagement and motivating preventive responses. However, their persuasive advantage is not universal; in lower-threat contexts, such as regular physical examinations, their effectiveness depends on congruence with the accompanying textual framing. These findings extend the applicability of the EASI model to mHealth and social media by showing that emojis influence health communication through both emotional reactions and inferential processing. Overall, this study highlights the strategic value of emoji use in health message design and underscores the importance of aligning visual and textual cues to maximize persuasive impact.

Acknowledgments

The study expresses its gratitude to the journal editors and reviewers.

Funding

This research was supported by the Guangdong Planning Office of Philosophy and Social Science [Grants: GD25YYS34].

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization; data curation; funding acquisition; investigation; project administration; visualization; writing – original draft; writing – review and editing: YL
Conceptualization; data curation; funding acquisition; methodology; supervision; validation; writing – original draft; writing – review and editing: SY

Conflicts of Interest

None declared.

Multimedia Appendix 1
A list of Twitter (X) accounts.
[DOCX File, 21 KB - humanfactors_v13i1e78824_app1.docx]

Multimedia Appendix 2
Phase 1.
[DOCX File, 27 KB - humanfactors_v13i1e78824_app2.docx]

Multimedia Appendix 3
Stimulus (Phase 2).
[DOCX File, 373 KB - humanfactors_v13i1e78824_app3.docx]

Multimedia Appendix 4
Stimulus (Phase 3).
[DOCX File, 417 KB - humanfactors_v13i1e78824_app4.docx]

References

1. Kaye SA, White MJ, Lewis I. The use of neurocognitive methods in assessing health communication messages: a systematic review. J Health Psychol 2017 Oct;22(12):1534-1551. [doi: 10.1177/1359105316630138] [Medline: 26908587]
2. Li YT, Chen ML, Lee HW. Health communication on social media at the early stage of the pandemic: examining health professionals' COVID-19 related tweets. Soc Sci Med 2024 Apr;347:116748. [doi: 10.1016/j.socscimed.2024.116748] [Medline: 38484456]

XSL•FO
RenderX

3. King AJ. A content analysis of visual cancer information: prevalence and use of photographs and illustrations in printed health materials. Health Commun 2015;30(7):722-731. [doi: 10.1080/10410236.2013.878778] [Medline: 25061954]
4. Heldman AB, Schindelar J, Weaver JB. Social media engagement and public health communication: implications for public health organizations being truly "Social". Public Health Rev 2013 Jun;35(1):1-18. [doi: 10.1007/BF03391698]
5. Zang S, Zhang X, Xing Y, Chen J, Lin L, Hou Z. Applications of social media and digital technologies in COVID-19 vaccination: scoping review. J Med Internet Res 2023 Feb 10;25:e40057. [doi: 10.2196/40057] [Medline: 36649235]
6. Erle TM, Schmid K, Goslar SH, Martin JD. Emojis as social information in digital communication. Emotion 2022 Oct;22(7):1529-1543. [doi: 10.1037/emo0000992] [Medline: 34351198]
7. Ko E(, Kim D, Kim G. Influence of emojis on user engagement in brand-related user generated content. Comput Human Behav 2022 Nov;136:107387. [doi: 10.1016/j.chb.2022.107387]
8. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive health communication applications for people with chronic disease. Cochrane Database Syst Rev 2005 Oct 19;2005(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]
9. Boender TS, Louis-Ferdinand N, Duschek G. Digital visual communication for public health: design proposal for a vaccinated emoji. J Med Internet Res 2022 Apr 7;24(4):e35786. [doi: 10.2196/35786] [Medline: 35389363]
10. Langdon KJ, Scherzer C, Ramsey S, Carey K, Rich J, Ranney ML. Feasibility and acceptability of a digital health intervention to promote engagement in and adherence to medication for opioid use disorder. J Subst Abuse Treat 2021 Dec;131:108538. [doi: 10.1016/j.jsat.2021.108538] [Medline: 34154869]
11. Lotfinejad N, Assadi R, Aelami MH, Pittet D. Emojis in public health and how they might be used for hand hygiene and infection prevention and control. Antimicrob Resist Infect Control 2020 Feb 10;9(1):27. [doi: 10.1186/s13756-020-0692-2] [Medline: 32041666]
12. Sick J, Monteleone E, Pierguidi L, Ares G, Spinelli S. The meaning of emoji to describe food experiences in pre-adolescents. Foods 2020 Sep 16;9(9):1307. [doi: 10.3390/foods9091307] [Medline: 32947969]
13. Yu S, Zhao L. Designing emotions for health care chatbots: text-based or icon-based approach. J Med Internet Res 2022 Dec 8;24(12):e39573. [doi: 10.2196/39573] [Medline: 36454078]
14. Helme DW, Donohew RL, Baier M, Zittleman L. A classroom-administered simulation of a television campaign on adolescent smoking: testing an activation model of information exposure. J Health Commun 2007 Jun;12(4):399-415. [doi: 10.1080/10810730701325947] [Medline: 17558790]
15. Kralj Novak P, Smailović J, Sluban B, Mozetič I. Sentiment of emojis. In: Perc M, editor. PLoS ONE 2015;10(12):e0144296. [doi: 10.1371/journal.pone.0144296] [Medline: 26641093]
16. Lin TS, Luo Y. Health persuasion through emoji: how emoji interacted with information source to predict health behaviors in COVID-19 situation. SSM Popul Health 2023 Mar;21:101343. [doi: 10.1016/j.ssmph.2023.101343] [Medline: 36712145]
17. Ray EC, Merle PF. Disgusting face, disease-ridden place?: Emoji influence on the interpretation of restaurant inspection reports. Health Commun 2021 Dec;36(14):1867-1878. [doi: 10.1080/10410236.2020.1802867] [Medline: 32806958]
18. Benito-Ostolaza JM, Echavarri R, Garcia-Prado A, Oses-Eraso N. Using visual stimuli to promote healthy snack choices among children. Soc Sci Med 2021 Feb;270:113587. [doi: 10.1016/j.socscimed.2020.113587] [Medline: 33486426]
19. Lin TS, Luo Y. Emoji and visual complexity in health information design: a moderated serial mediation model. Telematics and Informatics 2023 Nov;85:102065. [doi: 10.1016/j.tele.2023.102065]
20. Pfeifer VA, Armstrong EL, Lai VT. Do all facial emojis communicate emotion? The impact of facial emojis on perceived sender emotion and text processing. Comput Human Behav 2022 Jan;126:107016. [doi: 10.1016/j.chb.2021.107016]
21. Van Kleef GA. How emotions regulate social life: the emotions as social information (EASI) model. Curr Dir Psychol Sci 2009;18(3):184-188. [doi: 10.1111/j.1467-8721.2009.01633.x]
22. Harrington NG. Introduction to the special issue: message design in health communication research. Health Commun 2015;30(2):103-105. [doi: 10.1080/10410236.2014.974133] [Medline: 25470434]
23. Vann RJ, Tanner EC, Kizilova E. Perceived access, fear, and preventative behavior: key relationships for positive outcomes during the COVID-19 health crisis. J Consum Aff 2022;56(1):141-157. [doi: 10.1111/joca.12439] [Medline: 35603321]
24. Yıldırım M, Geçer E, Akgül Ö. The impacts of vulnerability, perceived risk, and fear on preventive behaviours against COVID-19. Psychol Health Med 2021 Jan;26(1):35-43. [doi: 10.1080/13548506.2020.1776891] [Medline: 32490689]
25. Huesch M, Chetlen A, Segel J, Schetter S. Frequencies of private mentions and sharing of mammography and breast cancer terms on Facebook: a pilot study. J Med Internet Res 2017 Jun 9;19(6):e201. [doi: 10.2196/jmir.7508] [Medline: 28600279]
26. Yost E, Zhang T, Qi R. The power of engagement: understanding active social media engagement and the impact on sales in the hospitality industry. Journal of Hospitality and Tourism Management 2021 Mar;46(1447-6770):83-95. [doi: 10.1016/j.jhtm.2020.10.008]
27. Yu S, Hu Y. When luxury brands meet China: the effect of localized celebrity endorsements in social media marketing. Journal of Retailing and Consumer Services 2020 May;54(1):102010. [doi: 10.1016/j.jretconser.2019.102010]
28. Baum M, Schäfer M, Kabst R. Modeling the impact of advertisement-image congruity on applicant attraction. Hum Resour Manage 2016 Jan;55(1):7-24. [doi: 10.1002/hrm.21652]
29. Mandler G. The structure of value: accounting for taste. In: Affect and Cognition: Psychology Press; 2014:3-36. [doi: 10.4324/9781315802756]

XSL•FO
RenderX

30. Hatfield E, Cacioppo JT, Rapson RL. Emotional contagion. Curr Dir Psychol Sci 1993 Jun;2(3):96-100. [doi: 10.1111/1467-8721.ep10770953]
31. Anderson RC, Pearson PD. A schema-theoretic view of basic processes in reading comprehension. In: Handbook of Reading Research 1984:255-291.
32. OSGOOD CE, TANNENBAUM PH. The principle of congruity in the prediction of attitude change. Psychol Rev 1955 Jan;62(1):42-55. [doi: 10.1037/h0048153] [Medline: 14357526]
33. Baumeister RF, Bratslavsky E, Finkenauer C, Vohs KD. Bad is stronger than good. Review of General Psychology 2001 Dec;5(4):323-370. [doi: 10.1037/1089-2680.5.4.323]
34. Kramer ADI, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. Proc Natl Acad Sci U S A 2014 Jun 17;111(24):8788-8790. [doi: 10.1073/pnas.1320040111] [Medline: 24889601]
35. Clarke N, Pechey E, Kosīte D, et al. Impact of health warning labels on selection and consumption of food and alcohol products: systematic review with meta-analysis. Health Psychol Rev 2021 Sep;15(3):430-453. [doi: 10.1080/17437199.2020.1780147] [Medline: 32515697]
36. Fan R, Xu K, Zhao J. An agent-based model for emotion contagion and competition in online social media. Physica A: Statistical Mechanics and its Applications 2018 Apr;495:245-259. [doi: 10.1016/j.physa.2017.12.086]
37. Brewer NT, Parada H, Hall MG, Boynton MH, Noar SM, Ribisl KM. Understanding why pictorial cigarette pack warnings increase quit attempts. Ann Behav Med 2019 Mar 1;53(3):232-243. [doi: 10.1093/abm/kay032] [Medline: 29850764]
38. Hou JR, Kankham S. More than feelings? How Facebook reaction icons affect online users' behavioral intentions toward online health rumor posts. INTR 2022 Nov 2;32(6):1978-2002. [doi: 10.1108/INTR-04-2021-0236]
39. Jaeger SR, Lee SM, Kim KO, Chheang SL, Roigard CM, Ares G. CATA and RATA questions for product-focused emotion research: five case studies using emoji questionnaires. Food Qual Prefer 2018 Sep;68:342-348. [doi: 10.1016/j.foodqual.2018.04.001]
40. Gerend MA, Shepherd JE. Predicting human papillomavirus vaccine uptake in young adult women: comparing the health belief model and theory of planned behavior. Ann Behav Med 2012 Oct;44(2):171-180. [doi: 10.1007/s12160-012-9366-5] [Medline: 22547155]
41. Guidry JPD, Carlyle KE, LaRose JG, Perrin P, Messner M, Ryan M. Using the health belief model to analyze Instagram posts about zika for public health communications. Emerg Infect Dis 2019 Jan;25(1):179-180. [doi: 10.3201/eid2501.180824] [Medline: 30561302]
42. King AJ, Lazard AJ. Advancing visual health communication research to improve infodemic response. Health Commun 2020 Dec;35(14):1723-1728. [doi: 10.1080/10410236.2020.1838094] [Medline: 33089711]
43. Slovic P, Peters E, Finucane ML, Macgregor DG. Affect, risk, and decision making. Health Psychol 2005 Jul;24(4S):S35-S40. [doi: 10.1037/0278-6133.24.4.S35] [Medline: 16045417]
44. Peracchio LA, Tybout AM. The moderating role of prior knowledge in schema-based product evaluation. J CONSUM RES 1996 Dec;23(3):177. [doi: 10.1086/209476]
45. Meyers-Levy J, Tybout AM. Schema congruity as a basis for product evaluation. J CONSUM RES 1989 Jun;16(1):39. [doi: 10.1086/209192]
46. Cheong Y, Kim K. The interplay between advertising claims and product categories in food advertising: a schema congruity perspective. Journal of Applied Communication Research 2011 Feb;39(1):55-74. [doi: 10.1080/00909882.2010.536845]
47. Credamo. URL: https://www.credamo.app [accessed 2026-07-15]
48. Emojipedia. URL: https://emojipedia.org/ [accessed 2026-07-15]
49. Stoianov D, Kemp N, Wegener S, Beyersmann E. Emojis and affective priming in visual word recognition. Cogn Emot 2025 Nov;39(7):1594-1608. [doi: 10.1080/02699931.2024.2402492] [Medline: 39370670]
50. Spinelli S, Masi C, Zoboli GP, Prescott J, Monteleone E. Emotional responses to branded and unbranded foods. Food Qual Prefer 2015 Jun;42:1-11. [doi: 10.1016/j.foodqual.2014.12.009]
51. Stonbraker S, Porras T, Schnall R. Patient preferences for visualization of longitudinal patient-reported outcomes data. J Am Med Inform Assoc 2020 Feb 1;27(2):212-224. [doi: 10.1093/jamia/ocz189] [Medline: 31670816]
52. Nabi RL, Myrick JG. Uplifting fear appeals: considering the role of hope in fear-based persuasive messages. Health Commun 2019 Apr;34(4):463-474. [doi: 10.1080/10410236.2017.1422847] [Medline: 29313717]
53. Nabi RL. Emotional flow in persuasive health messages. Health Commun 2015;30(2):114-124. [doi: 10.1080/10410236.2014.974129] [Medline: 25470436]
54. Herrando C, Constantinides E. Emotional contagion: a brief overview and future directions. Front Psychol 2021;12:712606. [doi: 10.3389/fpsyg.2021.712606] [Medline: 34335425]
55. Ricard BJ, Marsch LA, Crosier B, Hassanpour S. Exploring the utility of community-generated social media content for detecting depression: an analytical study on Instagram. J Med Internet Res 2018 Dec 6;20(12):e11817. [doi: 10.2196/11817] [Medline: 30522991]
56. Szeto MD, Barber C, Ranpariya VK, et al. Emojis and emoticons in health care and dermatology communication: narrative review. JMIR Dermatol 2022;5(3):e33851. [doi: 10.2196/33851] [Medline: 36405493]

XSL•FO
RenderX

Abbreviations

EASI: emotions as social information
ICC: intraclass correlation coefficient
mHealth: mobile health
REML: restricted maximum likelihood
VADER: Valence Aware Dictionary and Sentiment Reasoner

Edited by A Kushniruk, S Law; submitted 10.Jun.2025; peer-reviewed by M Laeeq Khan, S Shin, Y Qiu, Z Zhang; revised version received 31.May.2026; accepted 23.Jun.2026; published 31.Jul.2026.

Please cite as:
Luo Y, Yu S
Strategic Use of Negative Emojis in Messaging-Based Interventions for Public Health Communication on Social Media: Mixed Methods Study
JMIR Hum Factors 2026;13:e78824
URL: https://humanfactors.jmir.org/2026/1/e78824
doi:10.2196/78824

© Yue Luo, Shubin Yu. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 31.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effectiveness of Mobile-Delivered Exercise and Yoga Programs on Depressive Symptom Reduction in Employees: Randomized Controlled Trial

Jeong-Hyun Kim[1,2], MD, PhD; So Young Yoo[3,4,5], MD, PhD; Sohee Oh[6], PhD; Sun-Young Moon[2], MD; Kyesan Lee[7], PhD; Ki-Hyeok Nam[8,9], MSc; Heyeon Park[10], PhD; Se Hee Jung[11,12], MD, PhD; Hyo-Joon Choi[13], AA; Hyo Jung Kim[1], MSc; Seok Im Lee[4,5], MSc; Namhee Kim[4], MSc; Myoungnam Song[14], MSc; Beomjun Min[15], MD, PhD; Ji-Hye Lee[2], MSc; Jeong A Shin[1], MSc

[1]Department of Psychiatry, Seoul National University Bundang Hospital, 82 Gumi-ro 173 beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do, Republic of Korea

[10]Division of General Studies & Teaching Profession, Dongduk Women's University, Seoul, Republic of Korea

[11]Department of Rehabilitation Medicine, Seoul Metropolitan Government - Seoul National University Boramae Medical Center, Seoul, Republic of Korea

[12]Department of Rehabilitation Medicine, College of Medicine, Seoul National University, Seoul, Republic of Korea

[13]01arts Corporation, Seoul, Republic of Korea

[14]R&D Center, HYUNDAI MIB International Co., Ltd., Gunpo, Republic of Korea

[15]Seoul Chung Psychiatry Clinic, Seoul, Republic of Korea

[2]Department of Public Health Medical Services, Seoul National University Bundang Hospital, Bundang, Republic of Korea

[3]Department of Psychiatry, College of Medicine, Seoul National University, Seoul, Republic of Korea

[4]Department of Psychiatry, Seoul Metropolitan Government-Seoul National University Boramae Medical Center, Seoul, Republic of Korea

[5]Interdisciplinary Program in Cognitive Science, Seoul National University, Seoul, Republic of Korea

[6]Medical Research Collaborating Center, Seoul Metropolitan Government Seoul National University Boramae Medical Center, Seoul, Republic of Korea

[7]Electronic Engineering, College of Electronics & Information, Kyunghee University, Yongin, Republic of Korea

[8]Korea Association of Sports Culture and Industry, Seoul, Republic of Korea

[9]Misoinfotech, Gwacheon, Republic of Korea

Corresponding Author:
Jeong-Hyun Kim, MD, PhD
Department of Psychiatry, Seoul National University Bundang Hospital, 82 Gumi-ro 173 beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do, Republic of Korea

Abstract

Background: Mental health challenges such as stress and depression are prevalent among employees. Mobile health platforms that deliver exercise or yoga interventions offer a promising approach to improve mental health outcomes in this population.

Objective: This study aimed to assess the effectiveness of 12-session adaptive moderate-intensity exercise and yoga programs delivered via a motion-detecting digital platform in reducing stress and depressive symptoms among employees.

Methods: This was an unblinded, 3-arm, parallel-group, randomized controlled trial conducted at Seoul National University Bundang Hospital and Boramae Medical Center between November 2023 and January 2024. Eligible participants were full-time employees. Seventy-five participants were randomly assigned to an exercise, a yoga, or a cognitive behavioral therapy–based self-care control group using computer-generated randomization. The exercise and yoga groups engaged in motion-detecting, adaptive physical activity training, whereas the control group accessed mobile-based, self-directed stress management educational materials. The intervention was largely automated, with no individualized therapeutic guidance provided. Allocation was concealed until trial entry. All recruitment and outcome assessments were conducted in person at the hospitals. The primary outcomes were perceived stress and depressive symptoms, whereas the secondary outcomes included posttraumatic stress, insomnia severity, cognitive stress response, occupational stress, and burnout. Physiological outcomes were assessed using heart rate variability and electroencephalography. Measurements were collected at baseline, immediately after the intervention, and at 4-week follow-up. Data were analyzed using a multivariate linear model to evaluate the main effects of time, group, and time×group interactions.

Results: Of the 75 randomized participants (exercise: n=24, 32%; yoga: n=25, 33.3%; and control: n=26, 34.7%), 71 (94.7%) who completed at least 9 of the 12 sessions (≥40 min each) were included in the outcome analysis (exercise: n=21, 29.5%; yoga:

XSL•FO
RenderX

n=24, 33.8%; and control: n=26, 36.6%). For the coprimary outcomes, the group×time interaction for depressive symptoms (Patient Health Questionnaire-9) approached but did not reach the Bonferroni-corrected threshold ($F_{4,136}$=2.71; P=.03; adjusted α=.025); however, planned pairwise comparisons revealed significantly greater improvement in the yoga group compared to the control group at 4-week follow-up (β=−3.67; adjusted P<.001). For the Perceived Stress Scale, the interaction was not significant (P=.29), although a significant main effect of time (P<.001) indicated overall stress reduction across all groups. For secondary outcomes, a significant group×time interaction was found for the Cognitive Stress Responses Scale (P=.003), indicating differential trajectories of improvement. The yoga group showed a consistent linear decrease, whereas the exercise group showed immediate but less sustained gains.

Conclusions: Digitally delivered adaptive yoga programs demonstrated superior and sustained improvements in depressive symptoms and Cognitive Stress Responses Scale scores compared with the active cognitive behavioral therapy–based self-care control group. However, the exercise program showed more modest and less sustained effects, warranting further investigation using larger samples.

Trial Registration: ClinicalTrials.gov NCT06620783; https://clinicaltrials.gov/ct2/show/NCT06620783

(JMIR Hum Factors 2026;13:e80287) doi:10.2196/80287

KEYWORDS

exercise; yoga; mental health; stress; depression; employee; mobile health; digital health

Introduction

Mental health issues, including stress, depression, and anxiety, are major global health concerns, particularly among employees. Pressures in modern work environments, compounded by sedentary lifestyles, have led to an increased prevalence of psychological distress [1]. These mental health conditions not only affect individual well-being but also contribute to broader societal effects, including reduced productivity and increased health care costs [2]. Studies have shown that poor mental health among employees is associated with absenteeism, presenteeism, and lower workplace efficiency, highlighting the need for effective, accessible, and sustainable interventions [3]. Considering these challenges, there is growing interest in identifying effective, accessible, and sustainable interventions to improve mental health, particularly among the working population.

Physical exercise is associated with numerous mental health benefits, particularly in alleviating anxiety and depressive symptoms. Recent systematic reviews and meta-analyses have confirmed that physical activity can reduce psychological distress, improve mood, and enhance cognitive function [4,5]. These effects are believed to be mediated by various physiological mechanisms, including endorphin release, improved neuroplasticity, and regulation of the hypothalamic-pituitary-adrenal axis [5]. Despite these benefits, busy employees find adherence to regular exercise challenging, with time constraints and fatigue after long workdays being cited as common barriers [6].

Yoga has been shown to offer substantial mental health benefits as a mind-body intervention. Yoga integrates physical posture with breath control and mindfulness to provide a holistic approach to promoting physical and psychological well-being. Studies have suggested that yoga reduces stress and anxiety while enhancing mindfulness and relaxation [7]. Recent studies have demonstrated that yoga can significantly reduce symptoms of stress, anxiety, and depression while improving emotional regulation and social connectedness [8,9]. A 2024 randomized controlled trial (RCT) found that yoga can benefit mental health, with participants reporting improved self-compassion, extrinsic affect, and personal and communal spiritual well-being compared with control groups [10]. Moreover, yoga's flexibility allows practitioners to choose options ranging from gentle meditative practices to more intense forms, thus aligning with their time constraints and physical capabilities. This makes yoga particularly well-suited to individuals with busy schedules, potentially addressing some of the challenges associated with adherence to more physically demanding exercise routines [11].

Although physical exercise and yoga have been validated as effective mental health interventions, they offer distinct benefits and challenges, particularly for busy employees. Yoga, owing to its low impact and flexible nature, may be more accessible and sustainable for individuals with limited time. In contrast, moderate-intensity exercise boosts vitality and energy, which may be particularly beneficial for individuals experiencing fatigue or low energy [12]. However, lack of motivation, busy work schedules, and family commitments can pose significant barriers to adherence [13].

Recently, the integration of digital tools, including mobile apps, into health care has expanded the scope of health interventions. Mobile health platforms are increasingly being used to deliver exercise and wellness programs, offering convenience and personalization [14]. Mobile technology can provide access to personalized training programs at the user's convenience, overcoming traditional barriers to participation such as location, time, and access to expert instruction [15]. Wang et al [16] demonstrated that real-time feedback using smart wearable devices could improve postural control precision during balance training, suggesting that such systems enhance user engagement by providing immediate and actionable insights into movement accuracy.

This study evaluated the effects of a 12-session yoga program and an adaptive moderate-intensity physical exercise program using a mobile platform on employees' mental health outcomes compared to those of an active control group based on cognitive behavioral therapy (CBT). The primary hypothesis was that

XSL•FO
RenderX

yoga and physical exercise would positively influence mental health outcomes, such as reduced stress and depressive symptoms, compared to the control group. Furthermore, we hypothesized that these interventions induce beneficial changes in physiological markers, such as heart rate variability (HRV) and electroencephalography (EEG), and that differential effects may exist between the yoga and exercise groups on specific outcomes. By incorporating a mobile platform, we aimed to enhance the feasibility of the intervention by allowing participants to engage in programs at home and monitor their progress using real-time feedback.

Methods

Participants

The participants were enrolled via advertisements at Seoul National University Bundang Hospital and Boramae Medical Center between November 2023 and January 2024. The inclusion criteria were as follows: (1) employed adults aged 19 to 49 years; (2) individuals experiencing moderate-to-high levels of stress, defined as a Perceived Stress Scale (PSS) score of ≥14 at baseline; (3) individuals undergoing pharmacological treatment for psychiatric conditions, such as depression, anxiety disorders, or insomnia, who were eligible if their symptoms were stable and their medication dosages were not expected to change during the clinical trial; and (4) individuals who fully understood the clinical trial protocol and voluntarily consented to participate.

The exclusion criteria were as follows: (1) individuals aged <19 or >49 years; (2) those with dementia, intellectual disabilities, or other cognitive impairments; (3) individuals with seizure disorders, stroke, or other neurological conditions; (4) those with a history or current diagnosis of psychotic disorders, such as schizophrenia or bipolar disorder; (5) individuals with active suicidal ideation; (6) those using medications or having medical conditions (eg, heart or lung disease) that could affect the reliability of HRV testing; (7) individuals who received nonpharmacological psychiatric treatment or counseling within the past 6 months; and (8) individuals with physical limitations preventing them from participating in the exercise or yoga program.

The presence of psychiatric disorders was assessed using the Mini International Neuropsychiatric Interview, a short, structured psychiatric interview designed to identify a broad spectrum of psychiatric disorders as outlined in the *DSM-IV* (*Diagnostic and Statistical Manual of Mental Disorders* [Fourth Edition]), and the *ICD-10* (*International Statistical Classification of Diseases, Tenth Revision*). The Korean version of the Mini International Neuropsychiatric Interview has been shown to have good validity and reliability. The interviews, which typically took 15 to 30 minutes to complete, were conducted by 2 qualified psychologists with master's degrees who were familiar with the tool.

The total sample size was calculated a priori using the G*Power software (version 3.1.9.4; Heinrich-Heine-Universität Düsseldorf). On the basis of previous meta-analyses of digital mental health interventions [17], we anticipated a large effect size (f=0.40). To achieve a statistical power of 0.80 with a significance level (α) of .05 for a 3-group ANOVA, a minimum of 66 participants (22 per group) was required. This provided a conservative power estimate for the multivariate linear model (MLM) used in the primary analysis. After accounting for a potential dropout rate of 10%, 75 participants were recruited.

Measures and Assessments

Overview

The participants who provided consent underwent a baseline assessment. This assessment included demographic information, such as age, sex, marital status, education level, and occupation. It also included an evaluation of psychiatric diagnoses, medical history, and mental health–related questionnaires.

Changes in the PSS and Patient Health Questionnaire-9 (PHQ-9) scores for depressive symptoms were used as the primary psychological outcome measures. The secondary psychological outcomes included changes in the scores of scales assessing insomnia (Insomnia Severity Index), occupational stress (Korean Occupational Stress Scale [KOSS]), burnout (Maslach Burnout Inventory-General Survey), positive and negative affect (Positive and Negative Affect Schedule-Revised), posttraumatic stress symptoms (Posttraumatic Stress Disorder Checklist-5 [PCL-5]), interoceptive awareness (Multidimensional Assessment of Interoceptive Awareness [K-MAIA]), mindfulness (Five Facet Mindfulness Questionnaire [FFMQ]), and cognitive stress responses (Cognitive Stress Response Scale [CSRS]).

Originally developed by Cohen et al [18], the PSS is a 14-item scale; however, a modified 10-item version is commonly used [19,20]. This scale measures the degree to which individuals have perceived their lives as unpredictable, uncontrollable, or overwhelming in the past month. Higher scores indicate greater perceived stress. Total scores range from 0 to 40, with each item rated on a 5-point Likert scale, ranging from 0 (never) to 4 (very often).

PHQ-9 is a self-report depression screening tool that assesses the severity of depressive symptoms [21]. It comprises items based on 9 diagnostic criteria for major depressive disorder from DSM-IV. Total scores range from 0 to 27, with higher scores indicating more severe depressive symptoms.

Insomnia Severity Index was developed by Morin [22] to assess insomnia severity based on the diagnostic criteria of *DSM-IV* and the International Classification of Sleep Disorders. It comprises 7 items that examine the type and severity of insomnia over the past 2 weeks, satisfaction with sleep, daytime functioning, impact of sleep disturbances, and distress caused by sleep difficulties. Total scores range from 0 to 28, with higher scores indicating more severe insomnia.

The KOSS-Short Form is a shortened version of the KOSS consisting of 19 items reduced from the original 43. Each item is rated on a 4-point Likert scale, ranging from 1 (not at all) to 4 (very much). Higher scores indicate greater job stress experienced in the past month. The scale comprises 8 domains: physical environment, job demands, lack of job control, insufficient social support, job insecurity, organizational

XSL•FO
RenderX

injustice, inappropriate rewards, and work-life imbalance. It was designed for easy field application [23].

The Maslach Burnout Inventory-General Survey, developed by Schaufeli et al [24] and adapted by Shin [25], was used to measure participants' current levels of job burnout. This scale assesses 3 core aspects of workplace burnout: exhaustion, cynicism, and professional efficacy. Each item is rated on a 7-point Likert scale, ranging from 0 (strongly disagree) to 6 (strongly agree). Higher scores on the exhaustion and cynicism domains and lower scores on the professional efficacy domain indicate higher degrees of job burnout.

To measure the participants' positive and negative affect, we used the Positive and Negative Affect Schedule developed by Watson et al [26] and adapted by Park and Lee [27]. It assesses the emotions experienced over the past week using 20 items, with participants rating their affective experiences on a 5-point Likert scale.

We used the PCL-5 to measure the severity of posttraumatic stress disorder (PTSD) symptoms. The original PCL was developed by Weathers et al [28] and revised by Weathers et al [29] to align with the *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders* [Fifth Edition]) diagnostic criteria. The Korean version of the PCL-5 was validated by Lee et al [30]. The PCL-5 assesses the severity of symptoms experienced in the past month in relation to traumatic events. It comprises 20 items rated on a 5-point Likert scale, ranging from 0 (not at all) to 4 (extremely). Higher scores indicate more severe PTSD symptoms. PCL-5 includes 4 subscales: intrusion, avoidance, negative alterations in cognition and mood, and hyperarousal.

To assess participants' habitual body awareness, we used the K-MAIA, which was validated by Gim et al [31] based on the original scale developed by Mehling et al [32]. K-MAIA consists of 32 items rated on a 7-point Likert scale, ranging from 0 (not at all) to 6 (always). It comprises 6 subscales: noticing, accepting, attention regulation, mind-body connection awareness, returning to the body, and trusting. Noticing assesses awareness of positive, negative, and neutral bodily sensations by measuring sensitivity to physical sensations. Accepting evaluates the tendency to neither avoid nor divert attention from physical discomfort or the ability to not experience emotional distress from such discomfort. Attention regulation refers to the ability to control and maintain focus on bodily sensations. Mind-body connection awareness measures recognition of the connection between bodily sensations and emotions. Returning to the body refers to the ability to manage distress through attention to bodily sensations, leading to internal insight. Trusting reflects the extent to which one's body is experienced as safe and trustworthy.

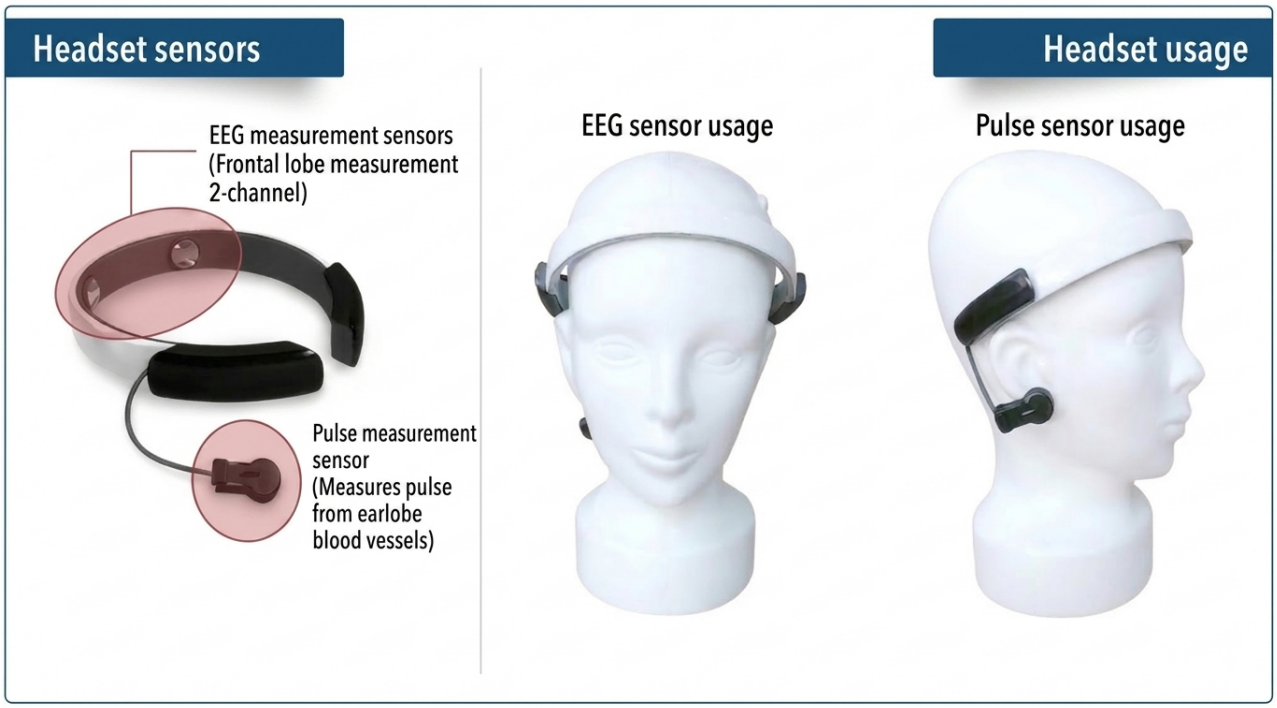
To measure participants' habitual mindfulness levels, we used the Korean version of the FFMQ (K-FFMQ) validated by Won and Kim [33] and based on the original scale developed by Baer et al [34]. K-FFMQ includes 39 items rated on a 7-point Likert scale, ranging from 1 (not at all) to 7 (always). Mindfulness is defined as observing the flow of internal and external stimuli as they occur without judgment and is considered an adaptive response that helps prevent automatic, impulsive, and maladaptive behaviors [34]. This scale comprises 5 factors. Nonreactivity refers to the ability to refrain from immediate reactions to internal experiences without becoming overwhelmed. Observing means attending to and noticing internal phenomena (eg, bodily sensations, thoughts, and emotions) and external stimuli (eg, sounds and smells). Description measures the ability to label and describe observed experiences in words. Acting with awareness refers to fully engaging in the present activity without distraction. Nonjudgment refers to refraining from making evaluations such as good or bad, right or wrong, or valuable or worthless.

To assess participants' cognitive stress responses, we used the CSRS developed by Koh and Park [35]. This scale evaluates 3 dimensions of the cognitive stress response: extreme negative, aggressive-hostile, and self-depreciative thoughts. It comprises 21 items rated on a 5-point Likert scale, ranging from 0 (not at all) to 4 (very much). The test-retest reliability ranges from 0.87 to 0.95, and the internal consistency is 0.94.

In addition, we measured physiological variables, including HRV and EEG, which are widely used as stress-related physiological biosignals [35-37]. EEG and HRV were measured using the neuroNicle FX2 (OMNI C&S, Inc) [38]. The device was equipped with 2 EEG monopolar electrodes positioned over the left and right prefrontal areas (FP1 and FP2, respectively), along with ground and reference electrodes attached to a clamp on the right earlobe (Figure 1). Additionally, one pulse wave sensor (photoplethysmograph [PPG]) was attached to the same clamp to measure HRV. EEG and PPG data were simultaneously recorded. The device was approved by the South Korean Ministry of Food and Drug Safety (authorization 16 - 4837). EEG and HRV measurements were recorded over a 5-minute period in a resting state with eyes closed, and the results were analyzed.

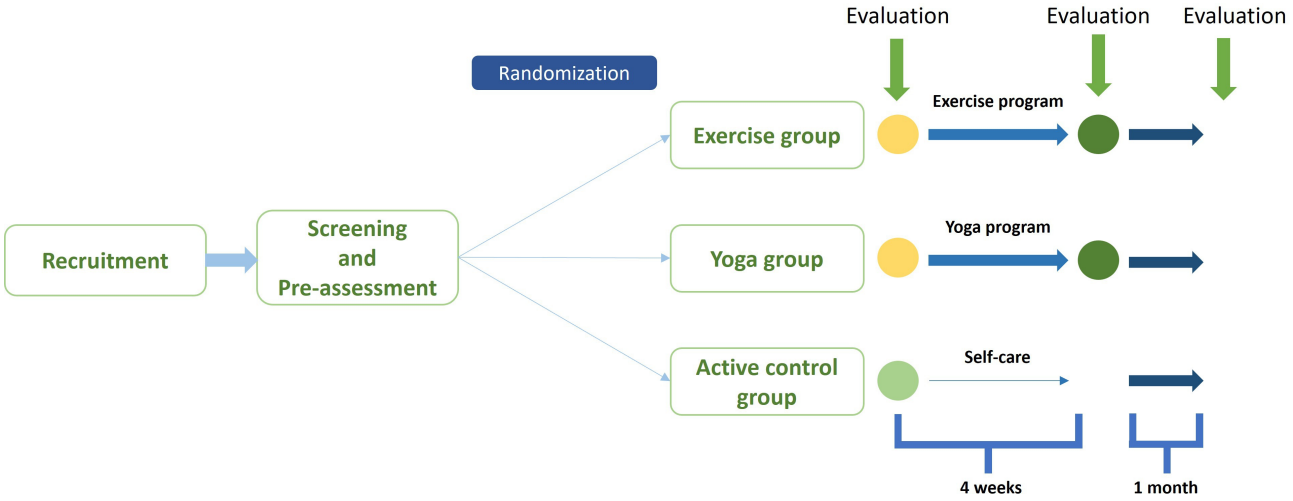
XSL•FO
RenderX

Figure 1. neuroNicle FX2 device. EEG: electroencephalography.



Throughout the study, all self-reported psychological, cognitive, and physiological indices were measured at 3 time points: during enrollment (baseline assessment), immediately after training (post-training assessment), and 4 weeks after training (1-mo follow-up assessment; Figure 2). Although psychological outcome measures (eg, PSS, PHQ-9, and CSRS) are typically validated for self-report, their administration in this study was conducted in person at the hospital during assessment visits, rather than purely online or remotely. This approach was used to ensure data quality and minimize potential technical issues or self-report bias often associated with unproctored web-based questionnaires. In addition, participants completed a structured user satisfaction and usability questionnaire, comprising four 4-point Likert-scale items and 6 closed-ended items (3 single-response and 3 multiple-response items in which up to 2 options could be selected), at their final visit (Multimedia Appendix 1).

Figure 2. Assessment process of the study.



Apparatus

The reliability of the prefrontal EEG marker from the neuroNicle FX2 device [38] was previously established by Choi et al [39], who reported no significant differences in the means of the tested variables between the prefrontal and occipital regions.

The raw data collected by the equipment were processed using TeleScan (OMNI C&S, Inc). Ocular, muscular, and other types of artifacts were detected using an automatic removal algorithm, and contaminated periods longer than 10 seconds were removed. For EEG, the band-pass filter (infinite impulse response Butterworth filters, high-pass filter: first order with fc=2.6 Hz; low-pass filter: eighth order with fc=43 Hz), with the 3 to 43 Hz range, was applied to the data. All data were digitized in continuous recording mode (5 min of recording, 250-Hz sampling rate, and 15-bit resolution). For HRV, the PPG

XSL•FO
RenderX

waveform and heartbeat time interval of 4 ms for each heartbeat were detected using the pulse wave sensor. Raw EEG and PPG data were transmitted to the host device in real time. In the host device, a fast Fourier transform was applied to calculate the power spectrum of the EEG and HRV data.

The power spectra of the θ (4 - 8 Hz), α (8 - 12 Hz), low β (12 - 15 Hz), β (15 - 20 Hz), high β (20 - 30 Hz), and γ (30 - 40 Hz) frequency bands were calculated from the EEG data. In addition, relaxation (ratio of α to high β power) and concentration (ratio of low β to θ power) indices were calculated from the EEG data. For HRV parameters, the power of the low-frequency (0.04 - 0.15 Hz) and high-frequency (0.15 - 0.4 Hz) bands was calculated. All variables were normalized according to sex and age.

Platform of Fitness and Yoga

A fitness platform was developed to enable participants to engage in exercise or yoga programs. To improve motion recognition accuracy, usability testing was conducted in 3 phases: the first phase involved office workers, the second included medical professionals, and the third targeted the general public in home settings. The evaluation focused on key performance indicators, including user interface intuitiveness, functional accessibility, system stability, and latency. The tests identified specific areas for improvement, such as menu complexity, ambiguity in instructional text, and suboptimal user inteUser Interface intuitiveness. Recognizing that these issues could increase user-entry barriers and adversely affect adoption rates, corrective refinements were implemented prior to this clinical trial. Furthermore, the testing phases revealed that environmental variables, such as background interference and camera positioning, affected motion recognition accuracy. Consequently, the recognition algorithm was further optimized, and a comprehensive user manual was developed to ensure consistent performance across diverse environments. To ensure accuracy and quality of information, this study used Redmine, an open-source project management and defect-tracking platform, to oversee the development lifecycle. White box testing methodologies were used for logical defect detection and structural verification. We used a suite of automated testing tools, such as unit, Pytest, and SonarQube, to ensure the precision of the software's internal logic. The version of the platform evaluated was the initial release (version 1.0, "frozen" during the trial period). No major content or functional changes, bug fixes, or downtimes occurred that might have influenced the study results. Institutional affiliations were not displayed on the digital intervention platforms and were disclosed only during the consent process.

Randomization and Intervention Conditions

Overview

This study used a parallel-group randomized controlled design. Participants were assigned to 3 groups by 1:1:1 block randomization (stratified by organization and sex) with randomly selected block sizes (3, 6, or 6) using REDCap (Vanderbilt University) tools hosted at the Medical Research Collaborating Center of Seoul National University Bundang Hospital. REDCap generates randomization codes with SAS (SAS Institute). Owing to the nature of the intervention, participants and intervention providers were not blinded; however, the allocation sequence was concealed from participants until they entered the trial and from the investigators until the end of the study. Participants were aware of the 3 intervention types but were not informed about the intervention of interest or the comparators.

Participants accessed the platform via mobile devices or online and followed the instructed movements displayed on the screen (Multimedia Appendix 2). Participants' joint movements were recorded and analyzed in real time during the sessions. They were also provided feedback on their exercise performance during relaxation and transitions between movements to ensure and enhance their participation and engagement. Only participants' skeletal movement data (body movement) were collected to check the accuracy of their movements, without facial recognition or screen recording. Participants' movements were tracked, and they were alerted with pop-up notifications of any deviations from the prescribed exercise or yoga movements, encouraging adherence. Motivational pop-up messages were displayed at the end of each session. Participants were instructed to access a designated internet address via a web browser using their unique ID to engage in the exercise, yoga, or CBT-based self-care program specifically developed for this study. During the study period, participants were advised to refrain from engaging in any exercise beyond the prescribed program. The program was conducted in a quiet and comfortable location of the participant's choosing. To minimize environmental influences, participants were instructed to use the same location for each session, ensuring that it was a space where they could engage in exercise or yoga without interruption. Notably, participants did not need to visit the hospital to complete the exercise or yoga sessions.

The control group participants who practiced CBT-based self-care were instructed to access a designated website through a web browser to view educational materials based on CBT. This paper covers the definition of stress, its signs and symptoms, and strategies for its management.

Participants accessed the intervention platforms free of charge using their own devices and internet connections and received 300,000 KRW (US $231) after completing the 4-week study. The intervention itself (exercise and yoga sessions) was fully automated, relying on motion detection technology to provide real-time objective feedback. Human involvement during the trial was limited to recruitment, in-person baseline or outcome assessments at the hospital, and basic technical support, as needed. This minimal human involvement was designed to reflect the potential for low-cost routine applications (see section *Discussion*). The use of the interventions was defined and measured by the completion of prescribed 12 sessions within the 4-week study period. This adherence metric was monitored using the platform's backend log file analysis. The ultimate measure of usage intensity had a high completion rate (71/75 randomized participants). During the intervention period and follow-up, no significant harmful or unintended effects were reported in any of the groups (exercise, yoga, and control).

XSL•FO
RenderX

Exercise Program

The exercise program was designed by the Korea Association of Sports Culture and Industry, an expert exercise organization involved in this study. On the basis of a literature review, the program was designed to alleviate the symptoms of conditions, such as depression, anxiety, and posttraumatic stress, by considering the intensity, frequency, duration, and type of exercise. These exercises included adaptive moderate-intensity exercises. Each session comprised warm-up, training, and cool-down phases. The program was conducted over 12 sessions, with each session lasting 50 minutes, held 3 times a week for 4 weeks.

- Warm-up (8 min): Stretching, which focuses on flexibility and balance, was incorporated to enhance joint mobility before the main exercise. This phase, which included self-stretching exercises, lasted for 8 minutes.
- Training (26 min): The main exercises focused on core strength and balance, with strength exercises being conducted for 26 minutes.
- Cool down (16 min): Designed to calm the sympathetic nervous system, this 16-minute phase included balance stabilization, static stretching, breathing exercises, mindset focus, and relaxation movements.

The protocol also incorporated expert recommendations from a researcher in the Rehabilitation Medicine Department at the Boramae Medical Center, ensuring a comprehensive approach to the exercise regimen. Multimedia Appendices 3 and 4 present the detailed protocols for the exercise program.

Yoga Program

An example of the movements in the program protocol is presented in Multimedia Appendix 5, and the schedule of the 4-week yoga program is presented in Multimedia Appendix 6.

CBT-Based Self-Care Program

The CBT-based active control group received educational materials aimed at stress management delivered via the platform through either web or mobile access. This occurred 3 times per week over 4 weeks, for a total of 12 sessions. The materials were based on CBT principles, with a focus on applying the content learned in the “Let’s Take a Moment” section to actionable behavior. This group served as an active control intervention to control for the nonspecific effects of attention, time, and digital platform access. The program was designed to encourage CBT-based self-care practices appropriate for an active control group. Multimedia Appendix 7 presents the protocol for the CBT-based stress management educational materials.

The CBT-based self-care control group was advised to refrain from exercising during the study period.

Adherence was monitored using the platform’s backend log file analysis. A minimum completion threshold of 9 of 12 sessions (≥75%), each lasting at least 40 minutes, was required for inclusion in the outcome analysis. The mean number of completed sessions did not differ significantly between groups. The quality of the exercise and yoga performance was ensured through the platform’s real-time motion detection system, which provided automated corrective feedback when movement deviations were detected.

Statistical Analysis

Demographic variables and baseline psychological measurements were analyzed using ANOVA or Kruskal-Wallis tests for continuous variables, whereas chi-square or Fisher exact tests were used for categorical variables to compare the 3 groups.

The primary analysis followed the intention-to-treat principle using an MLM for repeated measures [40,41], accommodating potential missing data points. This model evaluated the effects of group, time, and their interaction across 3 time points: baseline, post-training (4 wk), and follow-up (8 wk). A first-order autoregressive (AR(1)) correlation structure was specified to account for within-subject correlations, and baseline PSS scores were included as covariates to minimize confounding effects.

To control for the family-wise error rate (FWER), a hierarchical testing strategy was adopted for the 2 co-primary outcomes (PSS and PHQ-9). First, the group×time interaction was evaluated at a Bonferroni-corrected significance level of 0.025 (0.05/2 outcomes). Regardless of the interaction significance, prespecified planned pairwise comparisons (yoga vs control, exercise vs control, and yoga vs exercise) were performed at each post-baseline time point using a Bonferroni-adjusted alpha of 0.0083 (0.025 per 3 groups) to test our a priori hypotheses. For secondary outcomes, although treated as exploratory, a Bonferroni-adjusted alpha of 0.0167 (0.05 per 3 groups) was applied for pairwise comparisons to maintain statistical stringency.

In accordance with CONSORT (Consolidated Standards of Reporting Trials) reporting standards, results of group comparisons were reported as mean differences (β), adjusted 95% CIs, and adjusted *P* values. These CIs were calculated to align with corrected α levels (.0083 and .0167, respectively) to ensure a transparent interpretation of both statistical significance and clinical magnitude. The overall model parameters (eg, *F*, df_1, and df_2) are reported in the table.

Frontal θ power was defined as the primary physiological outcome of interest. All other EEG frequency bands (α, β, and γ) were treated as secondary or exploratory outcomes. Consequently, no formal adjustment for multiple testing was applied to these exploratory measures, and the results should be interpreted as hypothesis generating.

For the user satisfaction and usability questionnaire administered at the final visit (Multimedia Appendix 1), the four 4-point Likert-scale items were compared between the 2 active intervention groups (exercise vs yoga) using 2-tailed independent samples *t* tests, with Cohen *d* as the effect size because these items referred specifically to the assigned exercise or yoga program. The 6 closed-ended items—3 single-response items and 3 multiple-response items in which participants could select up to 2 options—were summarized as frequencies and percentages, and between-group differences in the endorsement of each response option were examined using Fisher exact tests.

XSL•FO
RenderX

These analyses were exploratory and were not adjusted for multiple comparisons.

All statistical analyses were performed using SAS (version 9.4).

Ethical Considerations

The study protocol for this RCT was approved by the institutional review boards of Seoul National University Bundang Hospital (approval B-2309-855-301) and Boramae Medical Center in South Korea (approval 30-2023-106), and the trial was registered at ClinicalTrials.gov (NCT06620783). At their initial onsite visit, all participants were provided with comprehensive written and verbal information regarding the study purpose, procedures, and interventions and subsequently provided written informed consent in person prior to participation. Participants were informed of their right to withdraw from the study at any time without penalty, and the original consent also covered secondary analyses without the need for additional consent. All data collected, including self-reported psychological, cognitive, and physiological indices, were handled with strict confidentiality; data were deidentified and stored using unique research IDs. Participants received compensation of KRW 100,000 (US $77) for their time and travel expenses at each of the 3 required evaluation visits (baseline, post-intervention, and 4-week follow-up), amounting to a maximum total compensation of KRW 300,000 (US $231).

During the intervention period and follow-up, no significant harmful or unintended effects were reported in any of the groups (exercise, yoga, and control).

Results

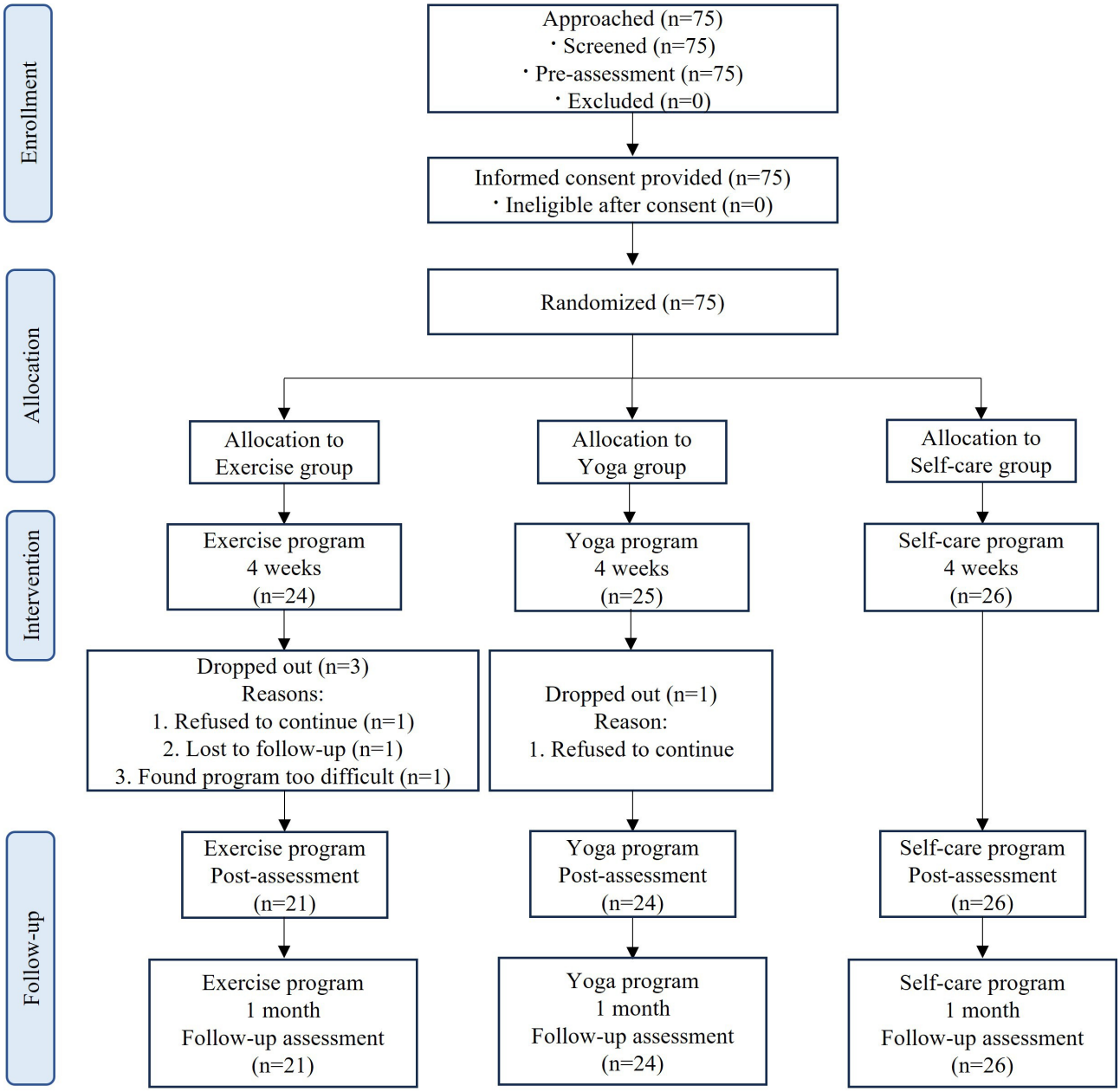
Overview

Of the 75 randomized participants (exercise, 24/75, 32%; yoga, 25/75, 33.3%; and control, 26/75, 34.7%), 71 who completed at least nine of the 12 sessions, with a minimum duration of 40 minutes per session, were included in the outcome analysis (exercise, 21/71, 29.6%; yoga, 24/71, 33.8%; and control, 26/71, 36.6%). Figure 3 presents a flow diagram of participant enrollment and retention. Specific reasons for attrition included personal time constraints (2/75, 2.67%), exercise program difficulty (1/75, 1.33%), and inability to establish follow-up contact (1/75, 1.33%). Participants' average age was 34.14 (SD 6.18) years, and 77.3% (58/75) were women. No significant group differences were observed in demographic characteristics, such as age, sex, marital status, or education level at baseline (all P>.05; Table 1).

Both the exercise and yoga interventions led to improved mental health outcomes over time. The yoga group demonstrated the most pronounced and sustained benefits, while the exercise group showed immediate gains in some cognitive measures.

XSL•FO
RenderX

Figure 3. A flow diagram of the study process.



XSL•FO
RenderX

Table . Demographic and clinical characteristics of each group at baseline.

Variables	Exercise (n=24)	Yoga (n=25)	CBT[a]-based self-care (n=26)	*P* value
Age (years), mean (SD)	34.58 (6.49)	34.60 (6.10)	34.04 (6.19)	.94
Sex (female), n (%)	18 (75)	20 (80)	20 (76.92)	.91
Marital status (married), n (%)	7 (29.17)	12 (48)	8 (30.77)	.31
Education level, n (%)				.66
Bachelor's degree or higher	19 (79.17)	22 (88)	23 (88.46)	
Job, n (%)				.69
Nurse	3 (12.50)	8 (32)	6 (23.08)	
Researcher	10 (41.67)	10 (40)	12 (46.15)	
Office worker	8 (33.33)	4 (16)	5 (19.23)	
Other	3 (12.50)	3 (12)	3 (11.54)	
PSS[b]_total, mean (SD)	21.42 (3.74)	21.88 (4.02)	19.27 (3.55)	.03
PHQ[c]_total, mean (SD)	6.63 (3.89)	8.08 (5.59)	6.88 (3.80)	.72
ISI[d]_total, mean (SD)	9.67 (5.40)	9.24 (4.87)	9.23 (4.99)	.97
KOSS-SF[e]_total, mean (SD)	42.75 (8.24)	43.36 (8.08)	43.04 (7.22)	.99
MBI-GS[f]_total, mean (SD)	47.83 (9.25)	49.60 (8.09)	48.15 (6.86)	.66
PANAS-R[g]_total, mean (SD)	43.92 (10.06)	47.04 (11.29)	45 (9.30)	.87
PCL-5[h]_total, mean (SD)	14.13 (16.16)	13.96 (14.82)	9.27 (12.94)	.31
Intrusion	3.79 (4.87)	3.44 (3.59)	2.35 (3.59)	.28
Avoidance	2.13 (2.52)	2.16 (2.03)	1.46 (2.14)	.32
Alterations in cognition and mood	4.92 (6.13)	4.92 (6.37)	3.46 (5.21)	.60
Hyperarousal	3.29 (3.78)	3.64 (5.06)	2.00 (3.25)	.39
K-MAIA[i]_total, mean (SD)	91.25 (31.30)	97.92 (31.95)	105.12 (20.11)	.20
K-FFMQ[j]_total, mean (SD)	160.79 (28.54)	163.88 (21.35)	164.46 (23.32)	.44
CSRS[k]_total, mean (SD)	39.29 (28.54)	35.68 (11.99)	32.96 (9.25)	.44
Extreme negative thoughts	6.46 (2.36)	6.44 (2.77)	5.81 (2.08)	.53
Aggressive hostile thoughts	5.38 (2.46)	4.84 (1.11)	4.73 (1.66)	.33
Self-depreciative thoughts	13.79 (6.88)	11.72 (5.65)	10.73 (3.95)	.26

[a]CBT: cognitive behavioral therapy.

[b]PSS: Perceived Stress Scale.

[c]PHQ: Patient Health Questionnaire-9.

[d]ISI: Insomnia Severity Index.

[e]KOSS-SF: Korean Occupational Stress Scale-Short Form.

[f]MBI-GS: Maslach Burnout Inventory-General Survey.

[g]PANAS-R: Positive and Negative Affect Schedule-Revised.

[h]PCL-5: Posttraumatic Stress Disorder Checklist-5.

[i]K-MAIA: Korean version of the Multidimensional Assessment of Interoceptive Awareness.

[j]K-FFMQ: Korean version of Five Facet Mindfulness Questionnaire.

[k]CSRS: Cognitive Stress Response Scale.

XSL•FO
RenderX

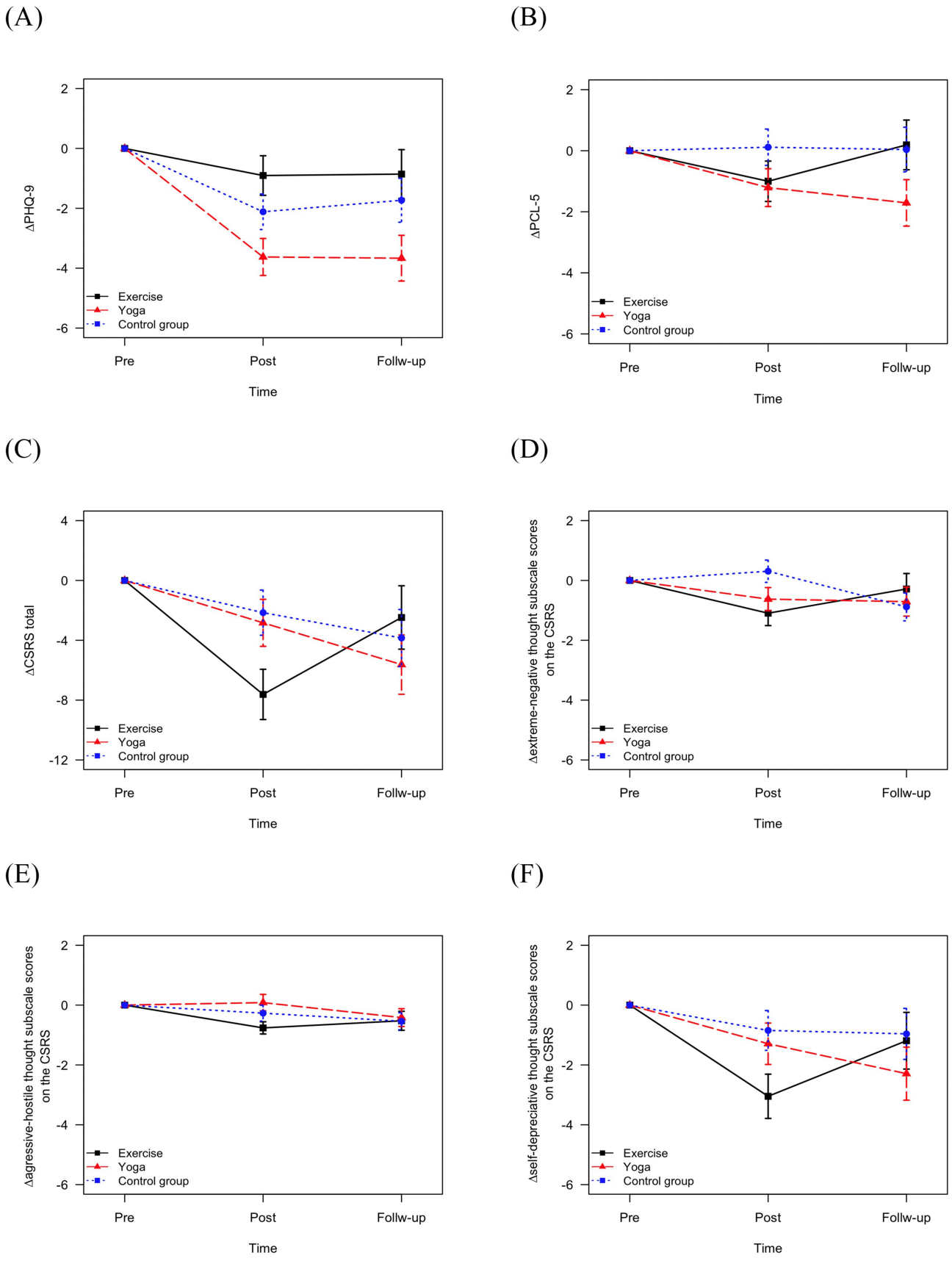
Primary Outcomes

The group×time interaction for PHQ-9 did not reach the prespecified Bonferroni-corrected significance level of 0.025 ($F_{4,136}$=2.71; P=.03). However, planned pairwise comparisons revealed a significant reduction in depressive symptoms for the yoga group compared to the control group at the 4-week follow-up (β=−3.67, 95% CI −5.64 to −1.69; adjusted P<.001). No significant differences were observed between the exercise and control groups at any time point (Figure 4A).

We also analyzed changes in another primary outcome: the PSS. The MLM showed that the group×time interaction was not statistically significant ($F_{4,136}$=1.27; P=.29), indicating that the trajectory of stress reduction did not differ significantly between the yoga, exercise, and control groups. However, a significant main effect of time was observed ($F_{2,136}$=38.56; P<.001), confirming a general reduction in perceived stress across all participants from baseline to the 4-week follow-up. The main effect of group was not significant ($F_{2,68}$=1.13; P=.33). The results, including the F statistics and P values, are summarized in Table 2, while Table 3 presents between-group differences and within-group changes in items from baseline (adjusted 95% CI).

XSL•FO
RenderX

Figure 4. Changes in scores of Patient Health Questionnaire (PHQ)-9, Posttraumatic Stress Disorder Checklist-5, and Cognitive Stress Response Scale (CSRS).



XSL•FO
RenderX

Table . Results of the multivariate linear model for repeated measures.

Outcome and effect	*F* statistics (*df*)	*P* value	Time	Least squares mean (SE)		
				Exercise	Yoga	Self-care
PSS[a]						
Group	1.13 (2, 68)	.33	Baseline	21.76 (0.77)	21.96 (0.82)	19.27 (0.68)
Time	38.56 (2, 136)	<.001	4W[b]	17.24 (0.89)	16.83 (1.20)	15.96 (0.83)
Group×time	1.27 (4, 136)	.29	8W[c]	16.71 (1.31)	16.50 (1.24)	16.35 (0.77)
PHQ[d] total						
Group	2.49 (2, 67)	.09	Baseline	6.02 (0.63)	7.66 (0.77)	7.87 (0.71)
Time	20.13 (2, 136)	<.001	4W	5.11 (0.56)	4.04 (0.46)	5.76 (0.59)
Group×time	2.71 (4, 136)	.03	8W	5.16 (0.75)	4 (0.47)	6.14 (0.53)
Hyperarousal (PCL-5[e] subscale)						
Group	0.24 (2, 67)	.79	Baseline	2.64 (0.65)	3.35 (0.81)	2.70 (0.60)
Time	1.53 (2, 136)	.22	4W	1.64 (0.51)	2.14 (0.48)	2.81 (0.53)
Group×time	2.61 (4, 136)	.04	8W	2.83 (0.76)	1.64 (0.48)	2.73 (0.56)
Extreme negative thought						
Group	0.17 (2, 67)	.85	Baseline	6.26 (0.39)	6.23 (0.46)	6.24 (0.43)
Time	3.77 (2, 136)	.03	4W	5.16 (0.35)	5.60 (0.42)	6.54 (0.43)
Group×time	4.04 (4, 136)	.004	8W	5.97 (0.42)	5.52 (0.35)	5.35 (0.42)
Aggressive hostile thought						
Group	0.26 (2, 67)	.77	Baseline	4.95 (0.27)	4.76 (0.20)	4.92 (0.32)
Time	7.44 (2, 136)	<.001	4W	4.19 (0.13)	4.84 (0.29)	4.65 (0.18)
Group×time	3.05 (4, 136)	.02	8W	4.43 (0.11)	4.34 (0.17)	4.38 (0.12)
Self-depreciative thought						
Group	1.09 (2, 67)	.34	Baseline	13.17 (1.18)	11.33 (0.95)	11.60 (0.72)
Time	38.56 (2, 136)	<.001	4W	10.13 (0.66)	10.03 (0.78)	10.75 (0.47)
Group×time	3.03 (4, 136)	.02	8W	11.98 (1.00)	9.03 (0.50)	10.64 (0.62)
CSRS[f] total						
Group	0.71 (2, 67)	.495	Baseline	37.27 (2.44)	34.44 (1.77)	35.49 (1.48)
Time	10.99 (2, 136)	<.001	4W	29.65 (1.62)	31.60 (1.72)	33.34 (1.23)
Group×time	4.28 (4, 136)	.003	8W	34.79 (2.27)	28.81 (1.09)	31.65 (1.36)

[a]PSS: Perceived Stress Scale.

[b]4W: 4 weeks.

[c]8W: 8 weeks.

[d]PHQ: Patient Health Questionnaire-9.

[e]PCL-5: Posttraumatic Stress Disorder Checklist-5.

[f]CSRS: Cognitive Stress Response Scale.

XSL•FO
RenderX

Table . Between-group changes and within-group differences in items from baseline[a].

Outcome, comparison, and time	Exercise	*P* value	Yoga	*P* value	Self-care	*P* value
PSS[b]						
Mean (SD)						
4W[c]	−4.52 (4.98)	<.001	−5.13 (4.44)	<.001	−3.31 (3.26)	<.001
8W[d]	−5.05 (6.55)	.002	−5.46 (5.06)	<.001	−2.92 (4.17)	.002
Between-group difference (95% CI)						
4W	−1.22 (−4.51 to 2.08)	.33	−1.82 (−4.73 to 1.09)	.10	Reference	—[e]
8W	−2.12 (−6.43 to 2.18)	.19	−2.54 (−5.99 to 0.92)	.05	Reference	—
Within-group change (95% CI)						
4W	−4.52 (−7.36 to −1.69)	<.001	−5.13 (−7.50 to −2.75)	<.001	−3.31 (−4.99 to −1.63)	<.001
8W	−5.05 (−8.78 to −1.31)	<.001	−5.46 (−8.16 to −2.75)	<.001	−2.92 (−5.07 to −0.78)	<.001
PHQ[f] total score						
Mean (SD)						
4W	−0.90 (2.77)	.15	−3.63 (3.57)	<.001	−2.12 (3.17)	.002
8W	−0.86 (2.74)	.17	−3.67 (3.69)	<.001	−1.73 (3.46)	.02
Between-group difference (95% CI)						
4W	1.21 (−1.06 to 3.48)	.16	−1.51 (−4.02 to 1.00)	.11	Reference	—
8W	0.87 (−1.50 to 3.25)	.33	−1.94 (−4.60 to 0.72)	.05	Reference	—
Within-group change (95% CI)						
4W	−0.90 (−2.49 to 0.68)	.13	−3.63 (−5.54 to −1.71)	<.001	−2.12 (−3.75 to −0.48)	<.001
8W	−0.86 (−2.42 to 0.71)	.14	−3.67 (−5.64 to −1.69)	<.001	−1.73 (−3.51 to 0.05)	.01
Hyperarousal (PCL-5[g])						
Mean (SD)						
4W	−1.00 (2.98)	.14	−1.21 (4.15)	.17	0.12 (3.06)	.85
8W	0.19 (2.98)	.77	−1.71 (3.39)	.02	0.04 (3.00)	.95
Between-group difference (95% CI)						
4W	−1.12 (−3.22 to 0.98)	.20	−1.32 (−3.79 to 1.14)	.20	Reference	—
8W	0.15 (−1.93 to 2.23)	.86	−1.75 (−3.90 to 0.41)	.05	Reference	—
Within-group change (95% CI)						

XSL•FO
RenderX

Outcome, comparison, and time	Exercise	*P* value	Yoga	*P* value	Self-care	*P* value
4W	−1.00 (−2.54 to 0.54)	.12	−1.21 (−3.22 to 0.80)	.15	0.12 (−1.31 to 1.54)	.85
8W	0.19 (−1.35 to 1.73)	.76	−1.71 (−3.35 to −0.06)[h]	.01	0.04 (−1.36 to 1.44)	.95
Extreme negative thought						
Mean (SD)						
4W	−1.10 (2.00)	.02	−0.63 (2.22)	.18	0.31 (1.72)	.37
8W	−0.29 (1.74)	.46	−0.71 (2.39)	.16	−0.88 (1.80)	.02
Between-group difference (95% CI)						
4W	−1.40 (−2.71 to −0.10)[h]	.01	−0.93 (−2.27 to 0.41)	.09	Reference	—
8W	0.60 (−0.63 to 1.83)	.24	0.18 (−1.25 to 1.60)	.77	Reference	—
Within-group change (95% CI)						
4W	−1.10 (−2.13 to −0.06)[h]	.01	−0.63 (−1.70 to 0.45)	.16	0.31 (−0.49 to 1.11)	.35
8W	−0.29 (−1.18 to 0.61)	.44	−0.71 (−1.86 to 0.45)	.14	−0.88 (−1.72 to −0.05)[h]	.01
Aggressive hostile Thought						
Mean (SD)						
4W	−0.76 (1.48)	.03	0.08 (1.41)	.78	−0.27 (1.89)	.47
8W	−0.52 (1.29)	.08	−0.42 (0.97)	.05	−0.54 (1.33)	.05
Between-group difference (95% CI)						
4W	−0.49 (−1.66 to 0.67)	.31	0.35 (−0.76 to 1.47)	.44	Reference	—
8W	0.01 (−0.90 to 0.93)	.97	0.12 (−0.66 to 0.90)	.71	Reference	—
Within-group change (95% CI)						
4W	−0.76 (−1.53 to 0.00)	.02	0.08 (−0.60 to 0.77)	.77	−0.27 (−1.15 to 0.61)	.46
8W	−0.52 (−1.19 to 0.14)	.06	−0.42 (−0.89 to 0.06)	.03	−0.54 (−1.16 to 0.08)	.04

[a]Statistical significance: *P* values and 95% CIs reflect alpha levels adjusted for multiple comparisons (Bonferroni correction) and baseline PSS imbalances, as specified in *Methods*.

[b]PSS: Perceived Stress Scale.

[c]4W: 4 weeks.

[d]8W: 8 weeks.

[e]Not applicable.

[f]PHQ: Patient Health Questionnaire-9.

[g]PCL-5: Posttraumatic Stress Disorder Checklist-5.

[h]*P*<adjusted α (primary outcomes: .008; secondary outcomes: .017).

XSL•FO
RenderX

Secondary Outcomes

Overview

Similarly, a significant time×group interaction effect ($F_{4,136}$=2.61; *P*=.04) was observed for hyperarousal symptoms, as measured by the PCL-5 subscale (Figure 4). Although pairwise comparisons did not show significant between-group differences at specific time points, a trend analysis indicated that the yoga group exhibited a significant linear decrease in hyperarousal symptoms over time, with an average reduction of 0.854 points per month (95% CI 0.18-1.52; *P*=.01). The exercise group showed a nonsignificant downward trend, whereas the control group showed no significant change. These findings suggest that the yoga intervention led to a continuous decrease in hyperarousal symptoms during the study period.

Cognitive stress responses, as measured by the CSRS total scores, showed a significant time×group interaction effect ($F_{4,136}$=4.28; *P*=.003; Figure 4). At the post-training assessment, the exercise group showed the most substantial reduction in total CSRS scores; however, this effect diminished at the 1-month follow-up. In contrast, both yoga and control groups demonstrated significant linear decreases in CSRS scores over time. Notably, the yoga group exhibited a more pronounced average reduction of 2.81 points per month (95% CI 1.27-4.36; *P*<.001), while the control group showed an average reduction of 1.923 points per month (95% CI 0.54-3.30; *P*=.007). The exercise group showed significant improvements immediately post-training (post vs pre: β=−7.62, 95% CI −13.18 to −2.06; P_{adj}<.01); however, some of this gain was lost by follow-up. Conversely, the yoga and control groups showed a continuous and significant decrease in CSRS scores during the study period.

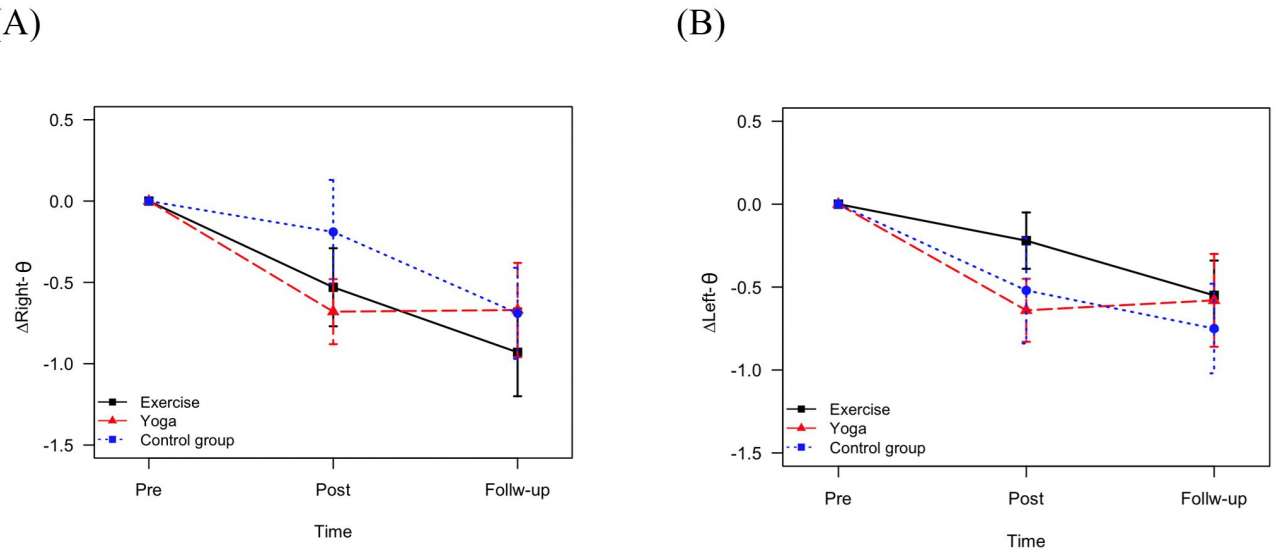
As a subscale of CSRS, extreme negative thoughts showed a significant time×group interaction effect ($F_{4,136}$=4.04; *P*=.004; Figure 4), indicating that changes in extreme negative thoughts over time differed across groups. The exercise group showed a substantial reduction in extreme negative thoughts post-training; however, this effect was not sustained at 1-month follow-up. In contrast, the control group demonstrated a significant decrease between the post-training and follow-up assessments, with an average reduction of 1.192 points (95% CI 0.42-1.97; *P*=.003). The yoga group showed no significant changes over time. These findings suggest that the exercise intervention led to an immediate decrease in extreme negative thoughts, although this effect diminished over time. Conversely, the control group experienced a delayed yet significant reduction, whereas the yoga group did not exhibit notable changes on this subscale.

Aggressive hostile thoughts also showed a significant time×group interaction effect ($F_{4,136}$=3.05; *P*=.02; Figure 4). The exercise group showed the greatest reduction in aggressive hostile thoughts post-training. The control group showed a decreasing trend over time (β=.269, 95% CI 0.02-0.52; *P*=.04). However, the yoga group showed a decreased effect after the post-training point. The self-depreciative thought subscale of the CSRS also had a significant time×group interaction effect ($F_{4,136}$=3.03; *P*=.02), with the yoga group showing the greatest sustained reduction (β=1.146, 95% CI 0.50- 1.79; *P*=.006; Figure 4). The exercise group showed a significant change immediately post-training, which was reversed after the training intervention ended.

Regarding biological markers, significant time effects were observed for both right and left EEG θ power (right θ: $F_{2,135}$=12.06; *P*<.001; and left θ: $F_{2,135}$=10.37; *P*<.001). As shown in Figure 5A, for right θ power, post-training levels were significantly lower than baseline levels (β=.45, 95% CI 0.15-0.76; *P*=.004), with a further decrease observed at the 1-month follow-up compared to baseline (β=.753, 95% CI 0.43-1.07; *P*<.001). However, the difference between post-training and follow-up θ power was not statistically significant (β=.302, 95% CI −0.04 to 0.65; *P*=.09). A comparable pattern was observed for the left θ power, as depicted in Figure 5. Post-training levels significantly decreased compared to baseline (β=.47, 95% CI 0.19-0.75; *P*=.001). This reduction remained significant at follow-up (β=.631, 95% CI 0.33-0.93; *P*<.001). No significant differences were found between the post-training and follow-up measurements (β=.16, 95% CI −0.15 to 0.48; *P*=.31). Other frequency bands, including α, β, and γ, did not show significant changes between pre- and post-training assessments. Furthermore, no significant group effects or group×time interactions were observed for any frequency band, including θ. Expanding on previous findings, no statistically significant changes were observed in the main HRV indices across time points.

XSL•FO
RenderX

Figure 5. Changes in the power of the θ frequency band of (A) the right and (B) the left electroencephalogram channel at 3 time points.



User Satisfaction and Usability

Responses to the user satisfaction and usability questionnaire are summarized in Multimedia Appendix 1. For the 4 Likert-scale items, the yoga group rated their assigned program as significantly more helpful than the exercise group for stress relief (mean 2.83, SD 0.70 vs mean 2.33, SD 0.66 on a 4-point scale; t_{43}=−2.45; *P*=.02; Cohen *d*=0.73) and for mental health management (mean 2.88, SD 0.61 vs mean 2.48, SD 0.60; t_{43}=−2.20; *P*=.03; Cohen *d*=0.66). Perceived helpfulness of the platform for overall health management (*P*=.10) and ease of use of the platform (*P*=.47) did not differ significantly between the 2 groups.

For the 6 closed-ended items, response distributions were largely comparable between the exercise and yoga groups, and no between-group differences reached significance for desired improvements (Q5), perceived inconveniences (Q6), prior and intended platform use (Q7), main reasons for using the program (Q8), or main barriers to engagement (Q10). The only significant difference was for preferred platform functions (Q9), where participants in the exercise group more frequently endorsed "referring to a variety of exercise programs" than those in the yoga group (57.1% vs 25.0%; Fisher exact *P*=.04). Descriptively, the most frequently endorsed inconvenience in both groups was a preference for compatibility across multiple devices (76.2% and 75.0% in the exercise and yoga groups, respectively), and the most frequently endorsed barrier to sustained engagement was insufficient personal motivation to continue (45.5% and 62.5%, respectively). Full item-level frequencies, percentages, and test statistics are provided in Multimedia Appendix 1.

Discussion

Principal Findings

This study aimed to assess whether a 12-session mobile-delivered exercise and yoga programs could effectively reduce stress and depressive symptoms among employees. An RCT demonstrated that both an adaptive moderate-intensity exercise program and a yoga program delivered via a digital platform significantly improved mental health outcomes among employees with moderate-to-high stress compared with a CBT-based self-care control condition. Both interventions decreased depressive symptoms, hyperarousal, and maladaptive CSRS scores. In particular, the yoga intervention showed the most pronounced and sustained benefits across a range of psychological measures. The significant group×time interactions for depressive symptoms, hyperarousal, and CSRS indicated that the trajectories of improvement differed by group, with the yoga group often showing the greatest improvement over time. This is consistent with research on the mental health benefits of mindfulness-based practices, such as yoga, which have been shown to alleviate anxiety and depressive symptoms [10]. A significant main effect of time was observed on frontal θ power; however, no significant group or group×time interaction effects were found. The exercise group also showed improvements in most outcomes, although in some cases, these were not as large or sustained as those in the yoga group. The yoga group demonstrated the greatest and most sustained improvements, whereas the exercise group showed immediate but less sustained gains in secondary cognitive measures. These results suggest differential efficacy patterns among the intervention modalities.

Furthermore, CSRS scores, particularly for extreme negative, aggressive hostile, and self-depreciative thoughts, decreased more prominently in the yoga and exercise groups than in the control group. These findings suggest that physical activity and mind-body interventions can effectively reduce cognitive distortions associated with stress. Such decreases in negative thought patterns may have contributed to the overall improvement in mental health observed in this study.

In addition to psychological outcomes, this study explored brainwave activity, focusing specifically on θ power in the prefrontal cortex, which is commonly associated with cognitive processes, such as attention and memory. We observed a significant decrease in θ power after both interventions across all groups. Elevated θ power in resting states has been associated with poor cognitive functioning [42], including deficits in attention and executive function. The overall reduction in resting frontal θ power over time may be consistent with changes in

XSL•FO
RenderX

neural markers associated with stress and arousal regulation. However, given the absence of cognitive performance measures, this interpretation is exploratory and should be viewed with caution [43]. This interpretation is supported by the results of previous studies. Thornberry et al [43] demonstrated that decreases in θ power are associated with improved spatial learning. Brzezicka et al [44] reported that reduced θ power in the dorsolateral prefrontal cortex, especially under a high cognitive load, predicts better task performance, which allows the brain to manage cognitive demands more effectively. The overall reduction in θ power observed across all groups may reflect improved arousal and attention regulation, possibly enhanced by the mindfulness components inherent in the yoga sessions. However, the absence of a significant group×time interaction precludes attributing this effect specifically to yoga.

Comparison With Previous Studies

Our results are consistent with those of a growing body of literature highlighting the mental health benefits of exercise and yoga interventions. A recent systematic review by Singh et al [45] found that physical activity interventions can significantly reduce depression, anxiety, and psychological distress in various populations. Similarly, our exercise intervention reduced depressive symptoms and stress-related outcomes. The observed improvements in the exercise group align with the findings of Arida and Teixeira-Machado [46], who noted that exercise contributes to neuroplasticity and cognitive reserve, which are crucial for brain resilience to cognitive impairment.

Our findings on cognitive resilience and stress responses complement those of Shields et al [47], who found that collegiate athletes exhibited cognitive resilience through adaptations in mood and stress responses during a competitive season, suggesting that regular physical activity supports mental health and cognitive flexibility. Furthermore, Brush et al [48] showed that 8 weeks of moderate-intensity aerobic exercise can reduce depressive symptoms and influence neural indicators of reward and cognitive control in young adults with major depression, indicating potential neurobiological pathways for exercise effects and paralleling our EEG findings.

Bouthillier and Asmundson [49], Tseng et al [50], and Soteriades et al [51] demonstrated the efficacy of physical exercise (including aerobic and resistance training) in reducing anxiety, PTSD symptoms, and occupational stress in patients with anxiety disorders, middle-aged adults with poor sleep, and firefighters, respectively. By including physiological stress measures (HRV and EEG), our study provides additional evidence that such interventions can induce favorable changes in stress-related biomarkers, although our HRV results did not reach significance, possibly because of the relatively short intervention period.

Previous studies conducted on trauma-exposed and high-stress populations have supported the benefits of combining mind-body practices with exercise. Goldstein et al [52] reported that an integrative exercise program for veterans with PTSD reduced symptoms, whereas Rosenbaum et al [53] observed that an exercise program tailored to police officers with PTSD improved psychological outcomes. Similarly, our yoga program, which included mindfulness and relaxation components, produced notable improvements in hyperarousal and negative thought patterns, reflecting outcomes particularly relevant to stress and trauma-related conditions. McKeon et al [54] also showed that a mental health–informed physical activity intervention delivered via social media was feasible for first responders, suggesting that digital delivery can be effective, a finding echoed by the high adherence and engagement observed via our digital platform.

Our results on physical activity and stress align with epidemiological findings such as those of LeardMann et al [55], who observed an association between regular vigorous exercise and a lower incidence of PTSD symptoms in military personnel. In our study, participants in the exercise and yoga groups engaged in structured physical activity and showed corresponding decreases in stress and hyperarousal, supporting the protective and ameliorative effects of exercise on stress-related outcomes. Schuch et al [4] and Mikkelsen et al [5] found that physical exercise reduces anxiety and depression, likely through endorphin release and enhanced neuroplasticity. The decreases in depressive symptoms and improvements in hypervigilance and cognitive responses observed in our study align with these physiological mechanisms.

Our findings emphasize the importance of exercise intensity and enjoyment. Paolucci et al [56] noted that both high-intensity interval training and moderate continuous training could reduce depression and inflammation, although intensity matters and high intensity may increase perceived stress. In the exercise program, we chose moderate-intensity exercise and observed mental health benefits without evidence of increased stress, suggesting that a tolerable intensity level is crucial for psychological benefits. Similarly, the yoga program, which was generally low-to-moderate intensity, yielded strong benefits, consistent with the idea that the optimal intensity for mental health improvements might not necessarily be maximal.

The general improvements in mental health observed in this study are supported by large-scale data, such as those of Chekroud et al [57], who found that individuals who exercise have a significantly lower mental health burden. In that study, all exercise types were associated with better mental health, with decreases in mental health burden ranging from 12% to 22% compared to nonexercisers. We observed robust improvements in the exercise and yoga groups, reinforcing the idea that regular physical activity, whether traditional exercise or yoga, can substantially improve mental well-being. Notably, Chekroud et al [57] also found that certain group-based exercises and mindfulness practices (eg, yoga and tai chi) had larger effects than some other forms of activity, a trend mirrored in our study in which the yoga group often experienced equal or greater benefits than the conventional exercise group.

Our study adds to existing evidence on the efficacy of yoga in occupational contexts. Rhodes et al [58] demonstrated the benefits of yoga in women with treatment-resistant PTSD, and Shohani et al [59] demonstrated its positive effects on stress, anxiety, and depression in women. Jeter et al [60] found that a yoga program for police trainees improved mood and stress, whereas Weathers-Meyer et al [61] reported that firefighters who participated in an 8-week yoga program experienced

XSL•FO
RenderX

reduced PTSD symptoms, negative affect, and trait anxiety. Similarly, our yoga program improved mood and stress (depressive symptoms and CSRS). Fang and Li [62] found that nurses who regularly practiced yoga reported better sleep quality and lower work stress, highlighting the broad benefits of yoga.

The comparable effectiveness of the CBT-based self-care control and yoga intervention groups may reflect the active psychological mechanisms embedded in both interventions. Self-guided CBT and psychoeducation, even in brief, have significantly reduced depressive symptoms in occupational populations [63]. This suggests that the control group did not serve as a passive comparator but instead offered evidence-based psychological tools with known antidepressant efficacy. Furthermore, stress regulation and mindfulness-based practices in yoga may overlap with CBT techniques through shared mechanisms of emotion regulation, attentional control, and cognitive reappraisal [64]. These parallels could account for the similar symptom reduction observed in both groups.

In contrast, the relatively weaker performance of the exercise group may be attributable to several pragmatic and methodological factors. First, the intervention was delivered remotely, without live supervision. Prior meta-analyses suggest that structured and professionally supervised exercise interventions tend to be associated with greater improvements in depressive symptoms compared to structured formats [65]. Second, real-world barriers, such as fatigue after long workdays and time constraints, may adversely affect engagement with exercise sessions [66]. Finally, the 4-week intervention duration may not have been sufficient to produce robust antidepressant effects from exercise [67], whereas the cognitive and emotional benefits of CBT-based self-care or yoga may manifest more quickly.

The user satisfaction findings converged with the primary outcome results. The yoga group, which showed the most pronounced and sustained improvements in depressive symptoms and cognitive stress responses, also rated their program as significantly more helpful for both stress relief and mental health management than the exercise group. This concordance between subjective perceptions of benefit and objective changes in psychological outcomes supports the internal consistency of our findings and suggests that the superior efficacy of the yoga intervention was accompanied by a correspondingly favorable user experience. The usability feedback also carries practical implications for future platform development. Across both groups, the most frequently endorsed request was for compatibility across multiple devices, and the most frequently reported barrier to sustained engagement was insufficient personal motivation to continue. These observations indicate that broadening device support and incorporating motivation-enhancing features (eg, progress reports, reminders, and social or gamified elements) may further improve adherence and engagement when such digitally delivered programs are scaled to routine workplace settings.

Suggestions and Limitations

These findings have important clinical implications. First, yoga is a viable and accessible intervention, and exercise may complement it in reducing workplace stress and improving employee mental health. Given that these programs were delivered via a mobile platform, our study supports the potential of digital health solutions to reach employees with time or accessibility constraints. Employers and health professionals should consider incorporating such programs into workplace wellness initiatives to address stress and mental health proactively.

Second, the distinct advantages of yoga suggest that mind-body exercises may be particularly useful for individuals who struggle with stress-related cognitive distortions or heightened arousal. However, the exercise program showed meaningful improvements in vitality and energy, consistent with findings indicating that exercise boosts energy levels. A combined approach that offers both traditional exercises and yoga can cater to personal preferences and target a broader range of outcomes.

Despite these promising results, this study had several limitations. First, the relatively short intervention period (12 sessions over 4 wk) may limit the generalizability of the findings. Although this study demonstrated short-term improvements in mental health outcomes, it remains unclear whether these benefits can be sustained in the long term. Future studies should include longer follow-up periods to assess the durability of the observed effects. Second, our sample size (N=75, three groups) was calculated to detect large effects (f=0.40). A larger sample size would increase statistical power. Third, the study relied on self-reported measures of psychological outcomes, which could be subject to response bias. Objective measures of stress, such as cortisol levels, were not obtained in this study but could strengthen future research. Fourth, although we attempted to control for participant behavior outside the interventions (eg, advising against extra exercise), we cannot fully rule out external factors influencing the outcomes (eg, changes in work stress or personal life events during the study). Fifth, the control group received a CBT-based self-care program that, while intended to be minimal, might have provided some benefits (eg, stress management education). This may have diluted between-group differences. Nevertheless, the inclusion of basic psychoeducation in the control condition made our findings more applicable to real-world settings in which doing nothing is often unethical or impractical. More importantly, by comparing the interventions to active digital control, the study successfully isolated the differential impact of physical and mind-body components (yoga and exercise) from the nonspecific effects of general engagement and digital access. The observed significant group×time interactions validate that the specific mechanisms of the active interventions produced differential trajectories of mental health improvement compared to CBT-based self-care control. Sixth, the homogeneity of the sample—predominantly female, young, urban, and well-educated employees—may limit the applicability of the findings to other populations, such as men, older adults, rural workers, and those with lower digital literacy. The Korean cultural context may also influence intervention uptake. Seventh, the stringent exclusion criteria ensured participant safety but may limit applicability to employees with comorbid psychiatric or physical conditions. Eighth, several elements of this RCT may differ from routine application.

XSL•FO
RenderX

In-person assessments and follow-ups conducted at the hospital provided a level of accountability and reduced loss to follow-up, which might not exist in purely remote routine applications. Additionally, the use of unique IDs and concealed allocations, when necessary, would be streamlined in a real-world scenario. The omission of in-person assessments in a routine setting might affect retention rates, underscoring the need for future studies on effective remote engagement strategies (eg, gamification or peer support). Ninth, EEG analyses were exploratory, and we did not apply formal corrections for multiple comparisons across frequency bands. As such, the observed time-related reductions in θ power should be considered hypothesis generating and require replication in future studies with preregistered EEG outcomes and appropriate controls for multiple testing. Finally, although the study successfully used a digital platform to deliver interventions, technical issues or user proficiency may have influenced engagement and outcomes. In addition, the analyses of the user satisfaction and usability questionnaire were exploratory and were not corrected for multiple comparisons; given the small subgroup sizes, these results should be interpreted with caution and regarded as hypothesis generating.

Despite these limitations, this study highlighted several unmet needs and areas for future research. First, further investigations are required to explore the long-term effects of digital mental health interventions, particularly in diverse workplaces. Understanding how different populations engage with digital platforms and whether the benefits of yoga and exercise persist beyond the initial intervention period is crucial for upscaling these programs.

Future research should focus on refining the digital delivery of mental health interventions. For example, incorporating gamification or social support elements may motivate users to adhere to the programs. Finally, studies comparing these interventions across different psychological distress levels are required. Although this study primarily focused on individuals with moderate-to-high stress, future research should investigate whether these findings hold true for individuals with lower baseline stress or severe mental health conditions. This could help tailor interventions for specific mental health needs, ensuring that the most appropriate treatment is provided.

Conclusions

This study provides evidence that digitally delivered adaptive yoga programs demonstrate superior and sustained improvements in depressive symptoms and CSRS compared to an active CBT-based self-care control. However, the exercise program showed more modest and less sustained effects, which warrant further investigation using larger samples. The use of a motion-detecting digital platform demonstrated high feasibility and engagement, suggesting that mobile health interventions can overcome common barriers to exercise (eg, time and accessibility) in employee populations. Our findings support the integration of regular physical activity, including mind-body practices such as yoga, into workplace wellness programs to promote mental well-being. Large-scale studies with longer follow-up periods are warranted to confirm these results, examine the long-term maintenance of benefits, and evaluate the cost-effectiveness of such interventions. Employers and health care providers should consider leveraging digital health platforms to deliver structured mind-body programs, particularly yoga, to reduce workplace stress and enhance overall mental health among employees; however, the role of exercise warrants further investigation.

Acknowledgments

We would like to thank Nankyung Lee, a professional yoga teacher, for her invaluable assistance in designing the yoga program. We also extend our gratitude to the participants for their commitment and time during the trial. The authors declare the use of generative AI (GenAI) in the research and writing process. According to the GAIDeT (Generative AI Delegation Taxonomy; 2025), the following tasks are delegated to GenAI tools under full human supervision: literature search and systematization, proofreading, and editing. The GenAI tool used was ChatGPT 4.5. Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

Funding

This research was supported by the Culture, Sports, and Tourism R&D Program through the Korea Creative Content Agency funded by the Ministry of Culture, Sports, and Tourism in 2022 (Project Name: K-Total Force Fitness Sports Care Service Process Development, Project R-2022 - 020114).

Data Availability

The deidentified datasets generated and analyzed in this study are available from the corresponding author upon reasonable request. Access will be granted following a review of the research proposal by the study team to ensure that the data use aligns with the original consent and ethical approval. No data were deposited in a public repository due to restrictions under the institutional review board approval concerning sensitive health information.

Authors' Contributions

Conceptualization: SYY, JHK, KL
Data curation: HJK, SIL, NK, JHL, JAS
Formal analysis: SO, SYY, JHK

XSL•FO
RenderX

Funding acquisition: KL, MS, SYY, JHK
Investigation: SYM, HJK, SIL, NK, JHL, JAS
Methodology: SYY, JHK, KHN, HP, SHJ
Project administration: SYY, JHK, NK, JAS
Resources: SYY, JHK, SHJ
Software: KL, HJC, MS
Supervision: SYY, JHK, SO
Validation: KL, HJC, MS
Visualization: KHN, HP, SO
Writing – original draft: SYY, JHK, SO, KHN, HP
Writing – review & editing: SYY, JHK, SO

Conflicts of Interest

The intervention platform (fitness platform) was developed specifically for this research by MS from Hyundai MIB International and KL from Kyung Hee University. Ownership of the software is that of both Hyundai MIB International and KL.

Multimedia Appendix 1
User satisfaction and usability questionnaire items and item-level results.
[DOCX File, 25 KB - humanfactors_v13i1e80287_app1.docx]

Multimedia Appendix 2
The screenshots of the exercise, yoga, and control platform.
[DOCX File, 857 KB - humanfactors_v13i1e80287_app2.docx]

Multimedia Appendix 3
Schedule for the 4-week adaptive moderate-intensity exercise program.
[DOCX File, 25 KB - humanfactors_v13i1e80287_app3.docx]

Multimedia Appendix 4
Examples of movements included in the exercise program protocol.
[DOCX File, 268 KB - humanfactors_v13i1e80287_app4.docx]

Multimedia Appendix 5
Example of movements in the yoga program.
[DOCX File, 90 KB - humanfactors_v13i1e80287_app5.docx]

Multimedia Appendix 6
Schedule of the 4-week yoga program.
[DOCX File, 20 KB - humanfactors_v13i1e80287_app6.docx]

Multimedia Appendix 7
Schedule of the 4-week self-care program based on cognitive behavioral therapy.
[DOCX File, 18 KB - humanfactors_v13i1e80287_app7.docx]

Checklist 1
CONSORT-eHEALTH checklist.
[PDF File, 368 KB - humanfactors_v13i1e80287_app8.pdf]

References

1. Mental health: strengthening our response. World Health Organization. 2018. URL: https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response [accessed 2026-08-09]
2. OECD. Health at a Glance 2021: OECD Indicators: OECD Publishing; 2021. [doi: 10.1787/ae3016b9-en]
3. Mental health and employers: the case for investment-pandemic and beyond. : Deloitte; 2022 URL: https://www.deloitte.com/uk/en/services/consulting/analysis/mental-health-and-employers-the-case-for-investment.html [accessed 2026-08-09]
4. Schuch FB, Vancampfort D, Firth J, et al. Physical activity and incident depression: a meta-analysis of prospective cohort studies. AJP 2018 Jul 1;175(7):631-648. [doi: 10.1176/appi.ajp.2018.17111194]

XSL•FO
RenderX

5. Mikkelsen K, Stojanovska L, Polenakovic M, Bosevski M, Apostolopoulos V. Exercise and mental health. Maturitas 2017 Dec;106:48-56. [doi: 10.1016/j.maturitas.2017.09.003]
6. Koh YS, Asharani PV, Devi F, et al. A cross-sectional study on the perceived barriers to physical activity and their associations with domain-specific physical activity and sedentary behaviour. BMC Public Health 2022 May 26;22(1):1051. [doi: 10.1186/s12889-022-13431-2] [Medline: 35614402]
7. Cramer H, Lauche R, Langhorst J, Dobos G. Yoga for depression: a systematic review and meta-analysis. Depress Anxiety 2013 Nov;30(11):1068-1083. [doi: 10.1002/da.22166] [Medline: 23922209]
8. Kanchibhotla D, Harsora P, Subramanian S. Influence of yogic breathing in increasing social connectedness among Indian adults. Acta Psychol (Amst) 2024 Mar;243:104164. [doi: 10.1016/j.actpsy.2024.104164]
9. Wu Y, Yan D, Yang J. Effectiveness of yoga for major depressive disorder: a systematic review and meta-analysis. Front Psychiatry 2023;14:1138205. [doi: 10.3389/fpsyt.2023.1138205] [Medline: 37032928]
10. Brandão T, Martins I, Torres A, Remondes-Costa S. Effect of online Kundalini Yoga mental health of university students during Covid-19 pandemic: a randomized controlled trial. J Health Psychol 2024 May;29(6):567-580. [doi: 10.1177/13591053231220710] [Medline: 38230561]
11. Ross A, Thomas S. The health benefits of yoga and exercise: a review of comparison studies. J Altern Complement Med 2010 Jan;16(1):3-12. [doi: 10.1089/acm.2009.0044] [Medline: 20105062]
12. Wender CLA, Manninen M, O'Connor PJ. The effect of chronic exercise on energy and fatigue states: a systematic review and meta-analysis of randomized trials. Front Psychol 2022;13:907637. [doi: 10.3389/fpsyg.2022.907637] [Medline: 35726269]
13. Paguntalan JC, Gregoski M. Physical activity barriers and motivators among high-risk employees. Work 2016 Nov 22;55(3):515-524. [doi: 10.3233/WOR-162424] [Medline: 27792026]
14. Oba T, Takano K, Katahira K, Kimura K. Use patterns of smartphone apps and wearable devices supporting physical activity and exercise: large-scale cross-sectional survey. JMIR Mhealth Uhealth 2023 Nov 22;11:e49148. [doi: 10.2196/49148] [Medline: 37997790]
15. Fabbrizio A, Fucarino A, Cantoia M, et al. Smart devices for health and wellness applied to tele-exercise: an overview of new trends and technologies such as IoT and AI. Health Care (Don Mills) 2023;11(12):1805. [doi: 10.3390/healthcare11121805]
16. Wang IL, Wang LI, Liu Y, Su Y, Yao S, Ho CS. Application of real-time visual feedback system in balance training of the center of pressure with smart wearable devices. Int J Environ Res Public Health 2021 Sep 13;18(18):9637. [doi: 10.3390/ijerph18189637] [Medline: 34574560]
17. Pauley D, Cuijpers P, Papola D, Miguel C, Karyotaki E. Two decades of digital interventions for anxiety disorders: a systematic review and meta-analysis of treatment effectiveness. Psychol Med 2023 Jan;53(2):567-579. [doi: 10.1017/S0033291721001999] [Medline: 34047264]
18. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
19. Taylor JM. Psychometric analysis of the Ten-Item Perceived Stress Scale. Psychol Assess 2015;27(1):90-101. [doi: 10.1037/a0038100]
20. Lee J, Shin C, Ko Y, et al. The reliability and validity studies of the Korean Version of the Perceived Stress Scale. Psychosom Med 2012;20(2):127-134 [FREE Full text]
21. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
22. Morin CM. Insomnia: Psychological Assessment and Management: Guilford Press; 1993.
23. Song H, Kim HR, Kim I, et al. Occupational stress (KOSS®19): scale development and validation in the Korean context. Ann Occup Environ Med 2025;37:e12. [doi: 10.35371/aoem.2025.37.e12]
24. Schaufeli WB, Leiter MP, Maslach C, Jackson SE. The Maslach Burnout Inventory-Test Manual, 3rd edition: Consulting Psychologists Press; 1996.
25. Shin K. The Maslach Burnout Inventory-General Survey (MBI-GS): an application in South Korea. Korea J Ind Organ Psychol 2003;16(3):1-17 [FREE Full text]
26. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988 Jun;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
27. Park HS, Lee JM. A validation study of Korean version of PANAS-revised. kjp 2016 Dec 31;35(4):617. [doi: 10.22257/kjp.2016.12.35.4.617]
28. Weathers FW, Litz BT, Herman DS, et al. The PTSD checklist (PCL): reliability, validity, and diagnostic utility. 1993 Presented at: 9th Annual Convention of the International Society for Traumatic Stress Studies; Oct 24-27, 1993 URL: https://www.researchgate.net/publication/291448760_The_PTSD_Checklist_PCL_Reliability_validity_and_diagnostic_utility [accessed 2026-08-14]
29. Weathers FW, Litz BT, Keane TM, et al. The PTSD checklist for DSM-5 (PCL-5). : National Center for PTSD; 2013 URL: https://www.ptsd.va.gov/ [accessed 2026-08-05]

XSL•FO
RenderX

30. Lee DH, Gu M, Kwon W, Kim S. A study on reliability and validity of the Korean Version of PCL-5 (Posttraumatic Stress Disorder Checklist for DSM-5) for adults. kjcp 2020 May 31;32(2):559-582. [doi: 10.23844/kjcp.2020.05.32.2.559]
31. Gim WS, Sim KL, Cho OK. Korean Multidimensional Assessment of Interoceptive Awareness (K-MAIA): development and validation. Korean J Str Res 2016 Sep 30;24(3):177-192. [doi: 10.17547/kjsr.2016.24.3.177]
32. Mehling WE, Price C, Daubenmier JJ, Acree M, Bartmess E, Stewart A. The Multidimensional Assessment of Interoceptive Awareness (MAIA). PLoS ONE 2012;7(11):e48230. [doi: 10.1371/journal.pone.0048230] [Medline: 23133619]
33. Won D, Kim K. Validation of the Korean version of Five-Factor Mindfulness Questionnaire. Korea J Health Psychol 2006;11(4):297-317 [FREE Full text]
34. Baer RA, Smith GT, Hopkins J, Krietemeyer J, Toney L. Using self-report assessment methods to explore facets of mindfulness. Assessment 2006 Mar;13(1):27-45. [doi: 10.1177/1073191105283504] [Medline: 16443717]
35. Koh KB, Park JK. Development of the Cognitive Stress Response Scale. J Korea Neuropsychiatr Assoc 2004;43(3):320-328 [FREE Full text]
36. Kim HG, Cheon EJ, Bai DS, Lee YH, Koo BH. Stress and heart rate variability: a meta-analysis and review of the literature. Psychiatry Investig 2018 Mar;15(3):235-245. [doi: 10.30773/pi.2017.08.17]
37. Hou X, Liu Y, Sourina O, Tan YRE, Wang L, Mueller-Wittig W. EEG based stress monitoring. 2015 Presented at: 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC); Oct 9-12, 2015; Hong Kong, China p. 3110-3115. [doi: 10.1109/SMC.2015.540]
38. Neuronicle FX2. OMNIFIT. URL: https://www.omnifit.co.kr/mindcare [accessed 2024-09-21]
39. Choi J, Ku B, You YG, et al. Resting-state prefrontal EEG biomarkers in correlation with MMSE scores in elderly individuals. Sci Rep 2019 Jul 18;9(1):10468. [doi: 10.1038/s41598-019-46789-2]
40. Laird NM, Lange NT, Stram DO. Maximum likelihood computations with repeated measures: application of the EM algorithm. J Am Stat Assoc 1987 Mar;82(397):97-105. [doi: 10.1080/01621459.1987.10478395]
41. Diggle PJ. An approach to the analysis of repeated measurements. Biometrics 1988 Dec;44(4):959-971. [Medline: 3233259]
42. Tan E, Troller-Renfree SV, Morales S, et al. Theta activity and cognitive functioning: integrating evidence from resting-state and task-related developmental electroencephalography (EEG) research. Dev Cogn Neurosci 2024 Jun;67:101404. [doi: 10.1016/j.dcn.2024.101404] [Medline: 38852382]
43. Thornberry C, Caffrey M, Commins S. Theta oscillatory power decreases in humans are associated with spatial learning in a virtual water maze task. Eur J Neurosci 2023 Dec;58(11):4341-4356. [doi: 10.1111/ejn.16185] [Medline: 37957526]
44. Brzezicka A, Kamiński J, Reed CM, Chung JM, Mamelak AN, Rutishauser U. Working memory load-related theta power decreases in dorsolateral prefrontal cortex predict individual differences in performance. J Cogn Neurosci 2019 Sep;31(9):1290-1307. [doi: 10.1162/jocn_a_01417] [Medline: 31037988]
45. Singh B, Olds T, Curtis R, et al. Effectiveness of physical activity interventions for improving depression, anxiety and distress: an overview of systematic reviews. Br J Sports Med 2023 Sep;57(18):1203-1209. [doi: 10.1136/bjsports-2022-106195] [Medline: 36796860]
46. Arida RM, Teixeira-Machado L. The contribution of physical exercise to brain resilience. Front Behav Neurosci 2020;14:626769. [doi: 10.3389/fnbeh.2020.626769] [Medline: 33584215]
47. Shields MR, Brooks MA, Koltyn KF, Kim JS, Cook DB. Cognitive resilience and psychological responses across a collegiate rowing season. Med Sci Sports Exerc 2017 Nov;49(11):2276-2285. [doi: 10.1249/MSS.0000000000001363] [Medline: 28682806]
48. Brush CJ, Hajcak G, Bocchine AJ, et al. A randomized trial of aerobic exercise for major depression: examining neural indicators of reward and cognitive control as predictors and treatment targets. Psychol Med 2022 Apr;52(5):893-903. [doi: 10.1017/S0033291720002573] [Medline: 32838817]
49. LeBouthillier DM, Asmundson GJG. The efficacy of aerobic exercise and resistance training as transdiagnostic interventions for anxiety-related disorders and constructs: a randomized controlled trial. J Anxiety Disord 2017 Dec;52:43-52. [doi: 10.1016/j.janxdis.2017.09.005] [Medline: 29049901]
50. Tseng TH, Chen HC, Wang LY, Chien MY. Effects of exercise training on sleep quality and heart rate variability in middle-aged and older adults with poor sleep quality: a randomized controlled trial. J Clin Sleep Med 2020 Sep 15;16(9):1483-1492. [doi: 10.5664/jcsm.8560] [Medline: 32394890]
51. Soteriades ES, Vogazianos P, Tozzi F, et al. Exercise and occupational stress among firefighters. Int J Environ Res Public Health 2022 Apr 20;19(9):4986. [doi: 10.3390/ijerph19094986] [Medline: 35564381]
52. Goldstein LA, Mehling WE, Metzler TJ, et al. Veterans group exercise: a randomized pilot trial of an integrative Exercise program for veterans with posttraumatic stress. J Affect Disord 2018 Feb;227:345-352. [doi: 10.1016/j.jad.2017.11.002] [Medline: 29145076]
53. Rosenbaum S, Stierli M, McCullagh S, et al. An open trial of the RECONNECT exercise program for NSW Police Officers with posttraumatic stress disorder or psychological injury. Health Promot J Austr 2022 Jan;33(1):28-33. [doi: 10.1002/hpja.406] [Medline: 33463832]
54. McKeon G, Steel Z, Wells R, et al. A mental health-informed physical activity intervention for first responders and their partners delivered using facebook: mixed methods pilot study. JMIR Form Res 2021 Apr 22;5(4):e23432. [doi: 10.2196/23432] [Medline: 33885376]

XSL•FO
RenderX

55. LeardMann CA, Kelton ML, Smith B, et al. Prospectively assessed posttraumatic stress disorder and associated physical activity. Public Health Rep 2011;126(3):371-383. [doi: 10.1177/003335491112600311] [Medline: 21553666]
56. Paolucci EM, Loukov D, Bowdish DME, Heisz JJ. Exercise reduces depression and inflammation but intensity matters. Biol Psychol 2018 Mar;133:79-84. [doi: 10.1016/j.biopsycho.2018.01.015] [Medline: 29408464]
57. Chekroud SR, Gueorguieva R, Zheutlin AB, et al. Association between physical exercise and mental health in 1·2 million individuals in the USA between 2011 and 2015: a cross-sectional study. Lancet Psychiatry 2018 Sep;5(9):739-746. [doi: 10.1016/S2215-0366(18)30227-X]
58. Rhodes A, Spinazzola J, van der Kolk B. Yoga for adult women with chronic ptsd: a long-term follow-up study. J Altern Complement Med 2016 Mar;22(3):189-196. [doi: 10.1089/acm.2014.0407] [Medline: 26863321]
59. Shohani M, Badfar G, Nasirkandy MP, et al. The effect of yoga on stress, anxiety, and depression in women. Int J Prev Med 2018;9(1):21. [doi: 10.4103/ijpvm.IJPVM_242_16] [Medline: 29541436]
60. Jeter P, Cronin S, Khalsa SB. Evaluation of the benefits of a Kripalu Yoga Program for police academy trainees: a pilot study. Int J Yoga Therap 2013 Jan 1;23(1):24-30. [doi: 10.17761/ijyt.23.1.3x94511x3u47n0q5]
61. Weathers-Meyer AJ, Lowe AC, McGrew SJ, et al. A pilot feasibility trial of mind–body tactical training for firefighters: evaluation of a yoga-based transdiagnostic program. Mindfulness (N Y) 2024 Aug;15(8):1972-1986. [doi: 10.1007/s12671-024-02419-1]
62. Fang R, Li X. A regular yoga intervention for staff nurse sleep quality and work stress: a randomised controlled trial. J Clin Nurs 2015 Dec;24(23-24):3374-3379. [doi: 10.1111/jocn.12983] [Medline: 26478577]
63. Ebert DD, Heber E, Berking M, et al. Self-guided internet-based and mobile-based stress management for employees: results of a randomised controlled trial. Occup Environ Med 2016 May;73(5):315-323. [doi: 10.1136/oemed-2015-103269] [Medline: 26884049]
64. Hölzel BK, Lazar SW, Gard T, Schuman-Olivier Z, Vago DR, Ott U. How does mindfulness meditation work? Proposing mechanisms of action from a conceptual and neural perspective. Perspect Psychol Sci 2011 Nov;6(6):537-559. [doi: 10.1177/1745691611419671] [Medline: 26168376]
65. Schuch FB, Vancampfort D, Richards J, Rosenbaum S, Ward PB, Stubbs B. Exercise as a treatment for depression: a meta-analysis adjusting for publication bias. J Psychiatr Res 2016 Jun;77:42-51. [doi: 10.1016/j.jpsychires.2016.02.023] [Medline: 26978184]
66. Abdin S, Welch RK, Byron-Daniel J, Meyrick J. The effectiveness of physical activity interventions in improving well-being across office-based workplace settings: a systematic review. Public Health 2018 Jul;160:70-76. [doi: 10.1016/j.puhe.2018.03.029] [Medline: 29751224]
67. Liu L, Xin X, Wang H, Zhang Y. Effectiveness of physical exercise on mental health among university students: a systematic review and meta-analysis. Front Psychol 2025;16:1612408. [doi: 10.3389/fpsyg.2025.1612408]

Abbreviations

CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
CSRS: Cognitive Stress Response Scale
***DSM-5*:** *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition)
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition
EEG: electroencephalography
FFMQ: Five Facet Mindfulness Questionnaire
FWER: family-wise error rate
HRV: heart rate variability
***ICD-10*:** *International Statistical Classification of Diseases, Tenth Revision*
K-FFMQ: Korean version of the Five Facet Mindfulness Questionnaire
K-MAIA: Korean version of the Multidimensional Assessment of Interoceptive Awareness
KOSS: Korean Occupational Stress Scale
MLM: multivariate linear model
PCL-5: Posttraumatic Stress Disorder Checklist-5
PHQ-9: Patient Health Questionnaire-9
PPG: photoplethysmograph
PSS: Perceived Stress Scale
PTSD: posttraumatic stress disorder
RCT: randomized controlled trial

XSL•FO
RenderX

Edited by A Kushniruk; submitted 10.Jul.2025; peer-reviewed by R Endrighi, U Kirk; revised version received 17.Jun.2026; accepted 21.Jul.2026; published 01.Sep.2026.

Please cite as:
Kim JH, Yoo SY, Oh S, Moon SY, Lee K, Nam KH, Park H, Jung SH, Choi HJ, Kim HJ, Lee SI, Kim N, Song M, Min B, Lee JH, Shin JA
Effectiveness of Mobile-Delivered Exercise and Yoga Programs on Depressive Symptom Reduction in Employees: Randomized Controlled Trial
JMIR Hum Factors 2026;13:e80287
URL: https://humanfactors.jmir.org/2026/1/e80287
doi:10.2196/80287

© Jeong-Hyun Kim, So Young Yoo, Sohee Oh, Sun-Young Moon, Kyesan Lee, Ki-Hyeok Nam, Heyeon Park, Se Hee Jung, Hyo-Joon Choi, Hyo Jung Kim, Seok Im Lee, Namhee Kim, Myoungnam Song, Beomjun Min, Ji-Hye Lee, Jeong A Shin. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 1.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluating a Dual Digital Cognitive Behavioral Therapy and Health and Wellness Coaching Intervention for Anxiety and Depression: Single-Arm Pilot Study

Alex Kirk[1], PhD; Christopher S King[1], BA; Robert Gallop[2], PhD; Brian D Doan[1], PhD; Lorelei Simpson Rowe[3], PhD; Laura E Randa[1], MPA

[1]Toivoa, Inc, 591 Collaboration Way, Newark, DE, United States

[2]Department of Mathematics, West Chester University, West Chester, PA, United States

[3]Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN, United States

Corresponding Author:
Alex Kirk, PhD
Toivoa, Inc, 591 Collaboration Way, Newark, DE, United States

Abstract

Background: Anxiety and depressive disorders remain highly prevalent and insufficiently treated, with many individuals experiencing persistent or untreated symptoms, limited access to evidence-based care, or insufficient support between clinical encounters. Adults with disabilities represent a particularly underserved subpopulation, often facing compounded barriers to mental health care and higher rates of anxiety and depression. Digital therapeutics offer a scalable opportunity to address these gaps by extending structured, evidence-based interventions beyond traditional care settings.

Objective: This pilot study evaluated Rauha, a novel digital therapeutic created by Toivoa Inc, that integrates cognitive behavioral therapy (CBT)–based modules with live weekly sessions led by a National Board-Certified Health and Wellness Coach (NBC-HWC), delivering structured, smartphone-based psychoeducation and interactive therapeutic exercises combined with personalized mental health coaching to support behavior change.

Methods: Thirteen adults with mobility and/or hearing disabilities and clinically elevated anxiety and/or depression were enrolled in a single-arm, within-participants design. Participants completed 8 weeks of CBT modules delivered via smartphone, accompanied by synchronous virtual mental health coaching. Anxiety and depression were assessed using the Hamilton Anxiety and Hamilton Depression Rating Scales, respectively, at baseline, post treatment, and at the 4-week follow-up.

Results: Mean reductions were significant for both anxiety (−13.05, SD 2.51; $P<.001$) and depression (−12.83, SD 1.55; $P<.001$), exceeding thresholds for clinical significance and sustained through follow-up. Post treatment, 84.6% (11/13) of participants showed clinically significant improvement in both anxiety and depression. At follow-up, 76.9% (10/13) and 92.3% (12/13) of participants showed clinically significant improvement in anxiety and depression, respectively. Between baseline and follow-up time points, these reductions corresponded to mean shifts from moderate to mild anxiety on the Hamilton Anxiety Rating Scale and from moderate to mild/nondepressed on the Hamilton Depression Rating Scale. Participants reported strongly favorable acceptability, experience, and usability ratings for the Rauha treatment program, demonstrating 100% treatment retention and an average replay rate of 5.5 for personalized smartphone content.

Conclusions: The findings suggest that a combined digital CBT and NBC-HWC approach can yield clinically meaningful and durable symptom reductions in depression and anxiety, coupled with high user acceptability and engagement, for adults with disabilities. These findings provide preliminary evidence supporting Rauha as a scalable, evidence-informed mental health intervention with the strong potential to improve access and address key barriers to care.

(JMIR Hum Factors 2026;13:e92448) doi:10.2196/92448

KEYWORDS

digital therapeutic; digital cognitive behavioral therapy; virtual mental health; health and wellness coaching; anxiety; depression; disability

XSL•FO
RenderX

Introduction

Digital Interventions for Adults with Disabilities

About 18.2% and 21.4% of adults reported symptoms of anxiety and depression, respectively [1]. These conditions are often undertreated, partly due to many obstacles prohibiting treatment initiation and engagement, resulting in a substantial unmet need for effective and accessible mental health interventions [2]. Among underserved populations are adults with physical disabilities or functional limitations, who comprise a large proportion of the US population, with over 40% of those aged 45 to 64 years reporting such limitations, a 21% increase in prevalence from 2002 to 2016 [3]. This growing population faces compounded barriers to care, including limited access to evidence-based therapies, highlighting the critical need for scalable interventions tailored to both mental health and functional challenges. Additionally, this population is at higher risk of mental health difficulties than those without a disability, with recent data showing that 32.9% of those with disabilities report significant mental distress as compared with 7.2% of those without disabilities [4].

Unfortunately, individuals with disabilities and functional impairments also face significant barriers in accessing adequate treatment, including limited treatment options, a lack of adequate training among providers concerning the unique needs of those with disabilities, and limited transportation capabilities [5,6]. Further worsening this public health problem is a severe shortage of mental health providers. On average, within the United States, only 26.4% of all mental health care needs are being met based on provider volume and availability [7]. These provider shortages strongly contribute to the 2023 finding that 46% of US adults with a mental disorder did not receive care for their condition [8].

Fortunately, evidence suggests that these obstacles are surmountable. For example, research has shown that cognitive behavioral therapy (CBT) and mindfulness-based therapies are strong approaches for this population [9-11]. Furthermore, virtual mental health interventions for those with disabilities can address obstacles related to mobility or transportation [12,13], reduce wait times, and increase overall access [14-16]. Self-guided CBT has shown significant promise in treating adults with mild-to-severe mental health symptoms without requiring a licensed provider [17,18].

In response to the rapid expansion of digital health technologies, the digital therapeutics (DTx) category has emerged as a distinct class of evidence-based interventions. Although definitions vary, DTx is largely characterized as a digital, evidence-based medical intervention delivered directly to patients to prevent, treat, or manage a disease or disorder using validated clinical frameworks [19]. As it relates to mental health, DTx and closely related evidence-based digital interventions demonstrate significant promise in reducing symptom severity across multiple conditions [20-24]. However, there are notable limitations, including high rates of attrition [25], often attributed to symptom severity beyond the scope of self-guided DTx capabilities, technological limitations, and insufficient personalization or support [26,27]. Importantly, evidence suggests that attrition can be reduced and adherence improved through the inclusion of human support elements that enhance accountability, personalization, and timely response to changes in treatment needs [28,29]. However, despite their promise, many existing digital platforms remain inadequately accessible for individuals with disabilities, inadvertently reproducing the very barriers they aim to reduce [26,30]. As such, designing DTx in alignment with established accessibility standards, such as the Web Content Accessibility Guidelines (WCAG) 2.2 [31], is therefore not a peripheral design consideration but a foundational requirement for equitable delivery of evidence-based care.

Beyond digital solutions, another step in expanding mental health care has been the significant increase in health and wellness coaches, such as National Board-Certified Health and Wellness Coaches (NBC-HWC). Health and wellness coaches are professionals who support patients in making behavioral changes according to self-chosen health goals and needs [32]. In terms of efficacy, a recent meta-analysis supports the use of health and wellness coaches as a stand-alone behavioral intervention in improving mental health outcomes among those with medical and psychological conditions [33]. In addition, adding a health coach to self-guided CBT significantly improves mental health outcomes over and above self-guided CBT alone [34]. At a systemic level, the use of health and wellness coaches is a promising avenue for improving flexible access to mental health care, given that these sessions are significantly shorter than traditional psychotherapy, typically around 30 minutes [32,35], allowing for more patients to receive care per single coach when compared to licensed mental health providers.

In summary, although adults with disabilities face significant mental health concerns and obstacles to accessing adequate care, ongoing innovations in technology, treatment delivery, and types of care show promising results in better reaching and treating this underserved population.

Development of the Rauha Mental Health Program

Rauha is a smartphone-based DTx developed to deliver flexible, accessible, and evidence-based mental health care for adults with anxiety-related and depression-related symptoms, including those with functional or sensory limitations. The Rauha DTx is a Health Insurance Portability and Accountability Act (HIPAA)–compliant software, certified with WCAG 2.2 AA accessibility standards. This certification ensures usability for individuals with functional, sensory, or other accessibility challenges, as only 5.2% of the top 1 million websites pass WCAG accessibility testing [36]. Per WCAG principles, Rauha is perceivable, operable, understandable, and robust [31]. Therapeutic content is delivered across text, audio, and video, with captions, transcripts, and high-contrast scalable text to support perceivability. The platform allows full keyboard navigation and works with assistive technologies, avoiding reliance on time-sensitive interactions. Content is written in plain language and presented in a consistent, structured format. The system was tested across devices and browsers and designed to work with common accessibility tools.

The clinical program of Rauha consists of 2 integrated components: (1) asynchronous, self-guided, interactive modules

XSL•FO
RenderX

and exercises rooted in standard CBT [37], delivered digitally via recorded sessions by a licensed psychologist, allowing users to listen, review, and engage with content multiple times at their convenience; and (2) weekly live virtual coaching sessions with an NBC-HWC focused on implementing evidence-based behavior change strategies to support patients' chosen goals. The use of NBC-HWCs rather than licensed therapists was intentional to support scalability, expand access to care, and align the live treatment component with a structured behavior change model that complements the asynchronous CBT content.

This clinical program is meant to serve as an intermediary treatment option, balancing the benefits of both self-guided, smartphone-based CBT programs and traditional one-on-one care offered by clinicians (Table 1). Traditional psychotherapy is often constrained by provider availability and extended wait times [38,39], whereas self-guided CBT offers immediate access but may be limited by engagement and adherence challenges [40], as well as rarely, if ever, meeting WCAG standards [41]. Hybrid models that combine digital content with human support can address these limitations by maintaining scalability while incorporating structured support, which has been associated with improved engagement and more efficient use of clinical resources [42].

Table . Comparison of cognitive behavioral therapy (CBT) delivery models and access characteristics.

Feature	CBT delivery model		
	Self-guided CBT applications	Traditional provider care	Rauha CBT model
Clinician involvement	Little to none	Direct	Psychologist-developed and guided
Delivery format	Asynchronous	Synchronous	Combined
Wait times	Immediate	48-day average	3-day average
Scalability	High	Low	High
Consistency of care	Standardized content	Clinician-dependent	Standardized content with personalized coaching
Meets WCAG[a] standards	Rare	Not applicable	Yes
Flexible to patient needs	Limited	High	High
Clinical standards	Variable	High	High
Cost efficiency[b]	Low cost, variable engagement	High cost, resource-intensive	Moderate cost, scalable support

[a]WCAG: Web Content Accessibility Guidelines.

[b]Cost efficiency reflects relative resource utilization and scalability rather than absolute cost.

Study Objectives

The central aim of this study was to evaluate Rauha in the treatment of moderate-to-severe symptoms of anxiety and depression in adults with a hearing and/or mobility-related disability. We hypothesized that there would be (1) at least 80% treatment retention among those who initiated the intervention and (2) a statistically and clinically significant reduction (CSR) in anxiety and depressive symptoms between baseline and the end of the treatment period through a 4-week follow-up. As an exploratory aim, we assessed user acceptability, experience, and usability of the Rauha program.

Methods

Study Design

This study used a single-arm, within-participants design to measure the clinical impact of Rauha on symptoms of anxiety and depression, as measured by the Hamilton Anxiety Rating Scale (HAM-A) [43] and Hamilton Depression Rating Scale (HAM-D) [44], respectively. Patients were assessed with the HAM-A and HAM-D at baseline (preintervention), immediately after completing treatment (post intervention), and 4 weeks after completing treatment (follow-up). Participation from baseline to follow-up lasted approximately 12 weeks.

Participants

Participants were recruited through approved study flyers at Endeavors Programs in San Antonio, Texas, as well as through other approved recruitment channels, including clinical and community referrals. Inclusion criteria for participation were a self-reported mobility or hearing disability, age 22 years or older, a minimum score of 10 on the Patient Health Questionnaire-9 (PHQ-9) [45] or Generalized Anxiety Disorder-7 (GAD-7) [46], access to a stable internet connection via Wi-Fi or mobile data, access to an iPhone or iPad (optionally, an iPhone could be provided by the research team), English fluency, resident of the United States for the duration of the study, no changes in psychotropic medication or psychotherapy treatment in the 30 days before study entry, and comprehension of and willingness to sign informed consent and complete all study procedures.

Exclusion criteria were a diagnosis of psychotic or bipolar disorder, active participation in another treatment trial, nontobacco substance use disorder within the past 12 months, any suicidality beyond passive ideation in the past 12 months, and currently pregnant or planning to become pregnant during the treatment period.

XSL•FO
RenderX

Ethical Considerations

This study was approved by the Advarra Institutional Review Board (Protocol 79652). All participants provided informed consent prior to participation in any study-related activities. Participation was voluntary, and participants were informed that they could withdraw from the study at any time without penalty. Participant privacy and confidentiality were protected throughout the study. All study data were collected, transmitted, and stored within a HIPAA-compliant digital platform and secure data management dashboard with restricted access limited to authorized study personnel. Appropriate administrative, technical, and physical safeguards were implemented to protect participant information. Data were deidentified prior to analysis and reporting. Participants received US $125 in compensation for their participation in the study in accordance with the institutional review board–approved protocol and were not contingent upon study outcomes.

Materials

Patient Health Questionnaire-9

The PHQ-9 is a 9-item self-report measure corresponding to the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-5) [47] criteria for major depressive disorder and is used as an inclusion criterion in this study [45]. Each item is rated on a 4-point Likert scale ranging from 0 (“not at all”) to 3 (“nearly every day”), yielding total scores ranging from 0 to 27, where higher scores reflect more severe levels of depressive symptoms. The PHQ-9 has established cutoff scores of 5, 10, 15, and 20, representing mild, moderate, moderately severe, and severe depression, respectively. It has demonstrated good internal consistency (α=.85) [45].

Generalized Anxiety Disorder-7

The GAD-7 is a 7-item self-report measure of anxiety symptoms corresponding to the DSM-5 [47] criteria for GAD, used as an inclusion criterion in this study [46]. Each item is rated on a 4-point Likert scale ranging from 0 (“not at all”) to 3 (“nearly every day”), yielding total scores ranging from 0 to 21, with higher scores indicating greater anxiety symptoms. The GAD-7 has established cutoff scores of 5, 10, and 15, representing mild, moderate, and severe anxiety, respectively. It has demonstrated good internal consistency (α=.89) [48].

Columbia Suicide Severity Rating Scale

The Columbia Suicide Severity Rating Scale (C-SSRS) is a 6-item suicide risk screening tool that assesses suicidal ideation and behavior and is used as an exclusionary criterion in this study [49]. It was administered in an interview format, first at baseline and then at the start of each coaching session. The baseline C-SSRS assessed past-month suicidality (with the exception of item 6), whereas the subsequent weekly C-SSRS was used to assess any suicidality between each coaching session. The C-SSRS has demonstrated excellent internal consistency (α=.94) [49].

Hamilton Anxiety Rating Scale

The HAM-A is a structured interview for examiners to rate anxiety severity, and it was used as a primary outcome measure in this study [43]. It assesses anxiety symptoms based on 14 items, each rated from 0 (“not present”) to 4 (“severe”), with scores ranging from 0 to 56. It uses established cutoff scores, where a score of 7 or lower indicates no or minimal anxiety, 8 indicates mild anxiety, 15 indicates moderate anxiety, and 24 or higher indicates severe anxiety [50]. The HAM-A has demonstrated good internal consistency (α=.89) [51].

Hamilton Depression Rating Scale

The HAM-D is a structured interview for examiners to rate depression severity, and it was used as an outcome measure in this study [44]. It assesses depression symptoms based on 17 items, which range from 0 (“not present”) to 4 (“severe”) with total scores ranging from 0 to 54. Among these 17 items, 8 items are rated on a 0 to 2 scale and 9 items are rated on a 0 to 4 scale. In terms of established cutoff scores, a score of 7 or lower suggests no depression, 8 indicates mild depression, 17 indicates moderate depression, and 24 or greater indicates severe depression [52]. The HAM-D has demonstrated good internal consistency (α=.81–.89) [53].

Participant-Reported Surveys

Three exploratory surveys comprising 35 questions were created to assess the acceptability, usability, and user experience of the Rauha DTx.

User Acceptability Survey

This 10-item scale measured overall satisfaction, perceived value, and personal relevance of the Rauha DTx, including how useful and worthwhile participants found its content and structure. Items were rated on a 5-point Likert scale ranging from 1 (“strongly disagree”) to 5 (“strongly agree”). Example items include “My participation on this app was worth the time” and “I am satisfied with the help I received on the app.”

System Usability Survey

This 10-item scale measured users’ perceptions of ease of use, efficiency, and learnability of the Rauha DTx, evaluating whether participants found the interface intuitive and consistent. Items were rated on a 5-point Likert scale ranging from 1 (“strongly disagree”) to 5 (“strongly agree”). Example items include “I think that I would like to use this system frequently” and “I feel very confident using this app.”

User Experience Survey

This 15-item scale measured qualitative aspects of user experience within the Rauha program, including the perceived helpfulness of program features, engagement with multimedia content, and support from mental health coaches. It captured both functional and interpersonal satisfaction with the DTx and coaching components. Items were rated on a 5-point Likert scale ranging from 1 (“strongly disagree”) to 5 (“strongly agree”). Example items include “My mental health coach was critical to my completion of the program” and “I found the mental health coaches to be a useful part of the program.”

Procedures

Recruitment and Screening

Adults interested in participating in the study first initiated contact with study personnel at Endeavors Programs. All

XSL•FO
RenderX

interested adults then preconsented to an initial screening with the GAD-7 and PHQ-9 to establish appropriateness for study inclusion based on anxiety and/or depressive symptoms, respectively, and a self-reported disability. Those who met these initial eligibility criteria were then invited to the study consent process, where the full study details were outlined, including verification of remaining inclusion and exclusion criteria. After signing the consent form, participants were scheduled to complete the baseline HAM-A and HAM-D. Importantly, the HAM-A and HAM-D were administered by licensed clinicians who received the same training to standardize assessment procedures and scoring to maximize interrater reliability. These clinicians were independent from the intervention team, and each participant was evaluated by the same clinician at baseline, post treatment, and follow-up to ensure consistency in assessment and minimize variability due to differences in raters across time points. Upon completion of the structured interviews, participants were invited to download the Rauha smartphone DTx, assigned a coach, and booked for their virtual coaching orientation appointment. This orientation appointment was not part of the treatment but rather was used as an introduction to assist with registering the participant on the Rauha DTx and to help prepare the participant for the treatment itself. Registration for Rauha included a brief disability and demographic screening, a program introduction video by a licensed psychologist, and a coach-led onboarding session to orient participants to the program.

Treatment

The Rauha program targeted both anxiety and depressive symptoms using an 8-week intervention, combining the smartphone-based, self-guided component with weekly sessions led by an NBC-HWC and implementing evidence-based behavior change strategies. The selection of an 8-week intervention period was informed by prior work on both traditional CBT and digitally delivered CBT, as well as practical considerations related to the structure of the DTx and coaching model. CBT is typically delivered as a structured, time-limited intervention, with standard protocols often spanning on the order of 6 to 12 weeks depending on the target condition and setting [54,55]. Internet-based and guided CBT programs follow a similar pattern, with many interventions delivered within a defined multiweek window and a substantial proportion clustering around approximately 7 to 9 weeks [56,57]. An 8-week duration was therefore chosen as a balance between providing sufficient time for skill acquisition and consolidation while maintaining a format that is feasible and consistent with established CBT and digital intervention models.

Each week, the CBT-based smartphone program focused on a new, overarching topic comprising multiple lessons that fulfill that topic (Multimedia Appendix 1). These lessons consist of psychoeducational videos featuring a licensed psychologist, animated videos illustrating therapeutic concepts from an example patient perspective, infographics, and brief reading lessons to teach psychoeducation and various CBT concepts, as well as interactive quizzes and exercises that allow participants to apply the skills being taught. Each weekly digital topic is designed to take approximately 60 minutes to complete and can be divided into smaller segments, allowing participants to engage with the material according to their schedule and preferences.

In addition to the 60 minutes spent in the smartphone program, participants also met with an assigned NBC-HWC for 30 minutes each week. In accordance with board certification training, coaches used a client-centered approach to support participants in identifying mental health–related goals, using behavior change strategies consistent with motivational interviewing, behavioral activation, values identification, and solution-focused behavioral therapy. At the start of each coaching session, coaches administered the C-SSRS to assess suicidality and ensure participant safety. A safety escalation protocol was put in place in the event a participant endorsed any item on the C-SSRS beyond item 2, which would entail a same-day warm transfer back to their referring clinician for further risk management and appropriate treatment planning in the event of nonimminent suicidal concerns. In the case of imminent suicidal intent, emergency services would be activated, followed by notifying the original referring provider for follow-up once immediate safety was confirmed. Following the initial safety assessment, each session proceeded according to a semistructured approach that entailed (1) a brief mood check-in, (2) review of digital CBT content according to that week, (3) reviewing and reinforcing past skills (eg, behavioral activation, values clarification), (4) setting goals for the upcoming week, and then (5) identifying and troubleshooting any barriers to continued skills practice or weekly goals. While this semistructured approach was standardized across participants, coaches were permitted to tailor certain aspects of coaching according to individual needs, so long as it aligned with the framework of the CBT program. Specifically, this tailoring was restricted to the examples used to illustrate CBT concepts, pacing of skill learning based on participant comprehension or engagement, and goal refinement based on individual readiness and unique contextual barriers. NBC-HWC coaching is standardized to support engagement, goal attainment, and real-world application of CBT skills, without providing psychotherapy or modifying CBT content.

Steps were taken to ensure coaching fidelity both before participant enrollment and throughout active intervention, trained by a licensed clinician. Before participant enrollment, coaches were required to complete standardized training in the intervention protocol, undergo ongoing supervision, and adhere to rigorous quality assurance processes, ensuring consistent and reproducible delivery across all participants. This systematic approach was used in order to provide evidence-based support with high fidelity, minimizing between-participants variability and maximizing confidence in outcomes.

Upon completion of the intervention, anxiety and depressive symptoms were assessed using the clinician-administered HAM-A and HAM-D, respectively. These assessments were then administered at 4 weeks post intervention (ie, follow-up) to evaluate the durability of symptom changes. Importantly, even after the 8-week coaching component ended, participants retained access to the smartphone-based CBT program through the 4-week follow-up period, allowing them to revisit any clinical content or exercises.

XSL•FO
RenderX

Statistical Approach

Clinical Outcome Analysis

Regarding the first hypothesis, the retention rate was calculated as the proportion of participants who completed the full 8-week intervention to include all smartphone-based CBT modules and all 8 coaching sessions. All analyses related to the second hypothesis used the sample of participants who initiated treatment. Clinical measures are summarized with means and SDs for continuous measures and percentages for categorical and binary measures. Mixed-effects modeling (MEM) [58] was used for analyses focusing on symptom change. MEM accommodates the within-participant correlation over time for repeated measures, which was fit through a compound symmetry design for the variance-covariance matrix. MEM does not require balanced data per person, where missed assessments are excluded from the analysis. Time was measured discretely by assessment point corresponding to baseline, post treatment, and follow-up and treated as a fixed effect in the MEM. Linear contrasts of the model parameterization allowed us to estimate change from baseline to post treatment corresponding to the intervention effect, as well as change from baseline through follow-up, which allowed us to quantify the durability of the intervention effect. We used partial eta as an indication of within-person change. Following Cohen conventions [59], partial eta squared values of 0.1, 0.3, and 0.5 represent small, medium, and large effect sizes, respectively. Jacobson and Truax [60] criteria informed CSR calculations. Percentages with 95% CIs using the Clopper-Pearson exact CI formulas were used to calculate the proportion of participants achieving a CSR. Additionally, we calculated percent change in symptoms ranging from 10% to 90% for both anxiety and depression with 95% CIs.

Scoring and Analysis of Participant-Reported Surveys

Responses to the user acceptability, usability, and experience surveys were summarized descriptively at both the item and scale levels. For each item, we calculated the proportion of participants selecting each Likert-scale response option, with primary emphasis on the percentage endorsing “agree” or “strongly agree” as an indicator of a positive response. At the scale level, we computed for each participant the percentage of items endorsed as “agree” or “strongly agree,” and then averaged these values across participants to generate overall acceptability, usability, and user experience scores (reported as percentages with SDs). This approach was used given the exploratory nature of the measures and small sample size, prioritizing depth of user perspectives.

Replay Rate

We examined the replay rate as an exploratory indicator of engagement using platform analytics. The system passively tracks user interactions at the lesson level, including instances where participants returned to previously completed smartphone-based CBT content. A replay event was defined as any access of a completed lesson beyond the initial viewing. For each participant, replay events were summed across the 8-week intervention, and the average replay rate was calculated across participants.

Power Analysis

Prior to the start of the study, preplanned power calculations were based on a 2-tailed test with an α level of .05. Using a subgroup of 72 participants with disabilities reported in previous work [61], baseline HAM-D and HAM-A SDs were 9.1 and 5.5, respectively. Using these SDs with our treatment sample, there was 82.9% and 87.1% power to detect reductions of 8 and 8.5 points, respectively, on the HAM-D. Similarly, there was 78.8% and 85.6% power to detect reductions of 4.5 and 5, respectively, on the HAM-A. As such, the proposed data analysis is proof-of-concept with the goal to serve as initial evidence for a future larger study. With the exploratory nature of our investigation, corrections for multiple comparisons are generally not needed [62]. As such, the level of significance was set at .05 for all analyses.

Results

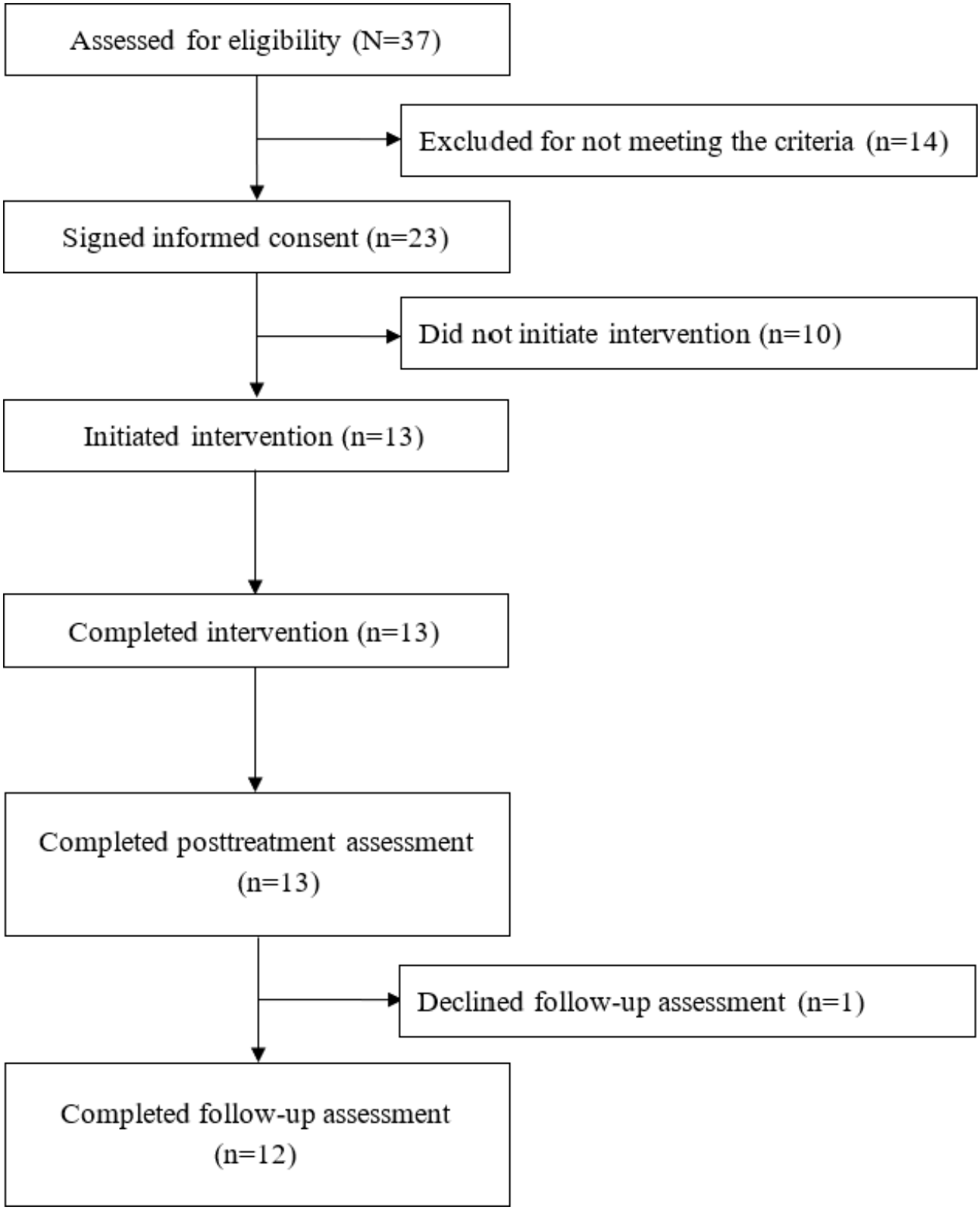
Study Flow and Retention

A total of 37 individuals were assessed for eligibility, of whom 14 were excluded for not meeting inclusion criteria. Twenty-three participants provided informed consent. Of these, 10 did not initiate the intervention: 1 declined to participate and did not complete baseline HAM-A or HAM-D assessments, and 9 elected not to continue due to preference for live clinician therapy, time-commitment requirements, discomfort with recorded sessions, or inability or unwillingness to complete the intervention on an iOS device. The remaining 13 participants initiated and completed the intervention, with no attrition during treatment, and all 13 completed the post treatment assessment. At follow-up, 12 participants completed the follow-up assessment, with 1 participant declining follow-up participation. Participant flow of the study is shown in Figure 1 (CONSORT [Consolidated Standards of Reporting Trials] diagram), detailing screening, enrollment, intervention initiation and completion, post treatment assessment, and follow-up assessment.

The final analytic sample consisted of 13 adults, of whom 53.8% (n=7) were female participants and 46.2% (n=6) were male participants. Reporting of age and race or ethnicity was optional; 12 of 13 participants reported age and 9 of 13 reported race or ethnicity. Among those reporting age, the mean age was 34.42 (SD 13.51) years. Of the 9 participants reporting race or ethnicity, 5 identified as White, 2 as two or more races, 1 as Hispanic, and 1 as Black or of African descent. With respect to disability status, 6 participants reported a mobility-related disability, 2 reported a hearing-related disability, and 5 reported both mobility-related and hearing-related disabilities. Based on baseline HAM-A and HAM-D scores, 10 participants exhibited clinically elevated levels of both anxiety and depression, 2 exhibited elevated depression only, and 1 participant did not demonstrate clinically elevated anxiety or depression despite meeting inclusion criteria. No participants endorsed any suicidal symptoms.

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of participant progress through the study.



Primary Outcomes

Table 2 provides descriptive statistics for HAM-D and HAM-A at our 3 assessment points.

XSL•FO
RenderX

Table . Clinical descriptives by assessment time point (N=13).

Assessment	HAM-D[a], mean (SD)	HAM-A[b], mean (SD)
Baseline	19.92 (5.14)	21.69 (9.35)
Post treatment	7.54 (6.63)	9.85 (8.9)
Follow-up	7.17 (5.36)	9 (7.01)

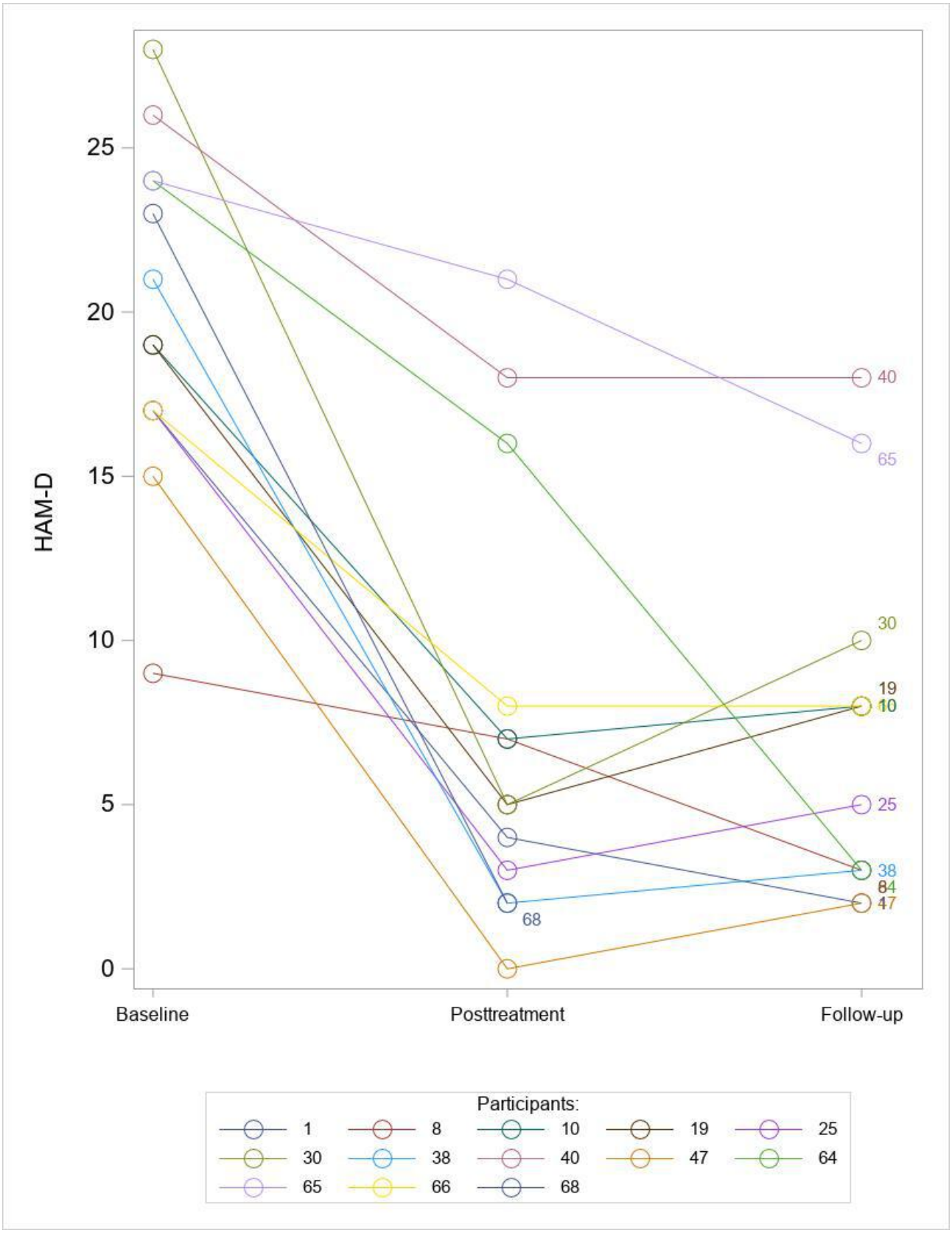
[a]HAM-D: Hamilton Depression Rating Scale.

[b]HAM-A: Hamilton Anxiety Rating Scale.

Individual spaghetti plots are illustrated in Figure 2 for HAM-D and in Figure 3 for HAM-A. Both figures illustrate 2 major findings in our study: (1) the majority of participants improved over the observational period and (2) the completeness of the data, where all participants except 1 (ID=68, who did not provide follow-up assessment) completed all 3 assessments. Figure 4 illustrates the mean profile over time for both anxiety (HAM-A) and depression (HAM-D).

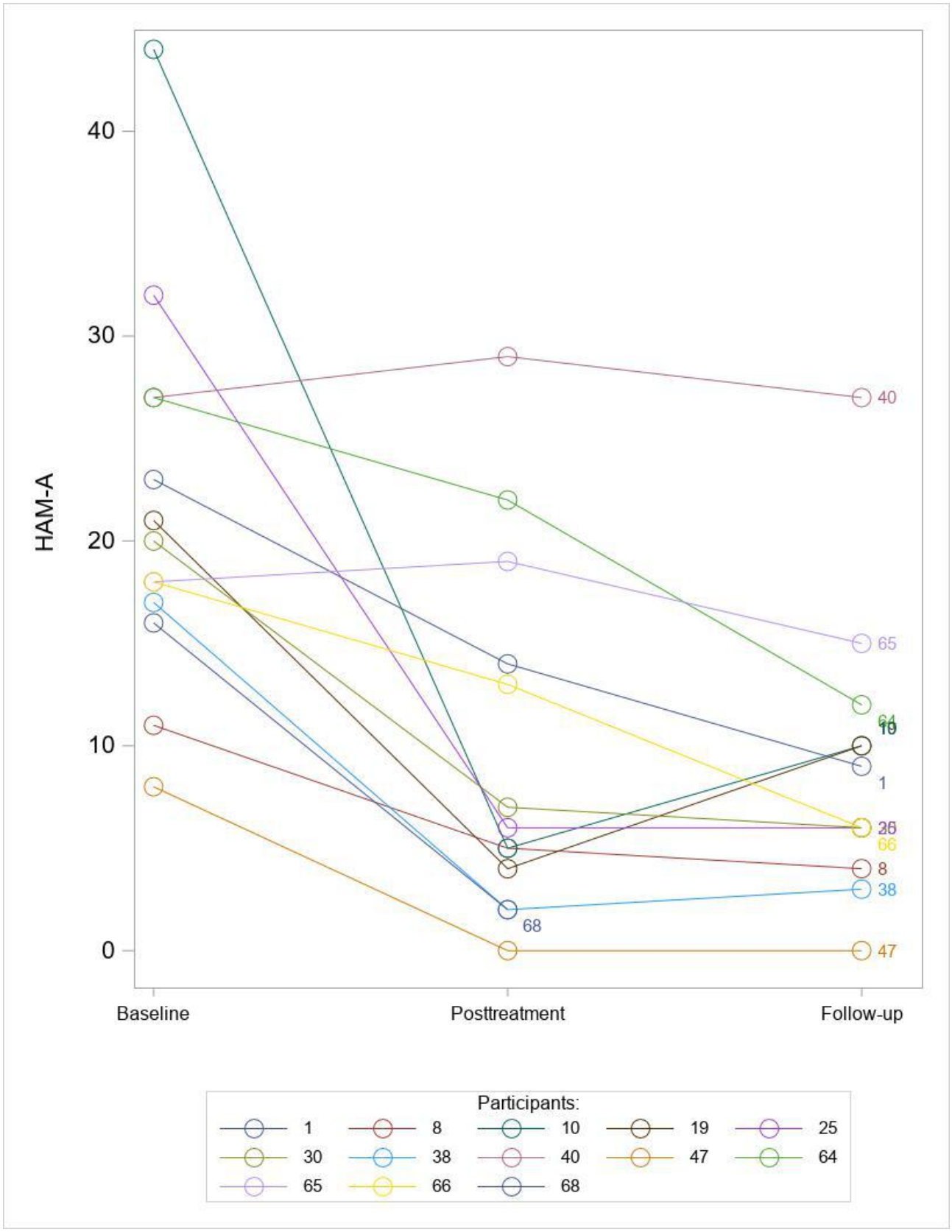
XSL•FO
RenderX

Figure 2. Individual profiles measured using the Hamilton Depression Rating Scale (HAM-D).



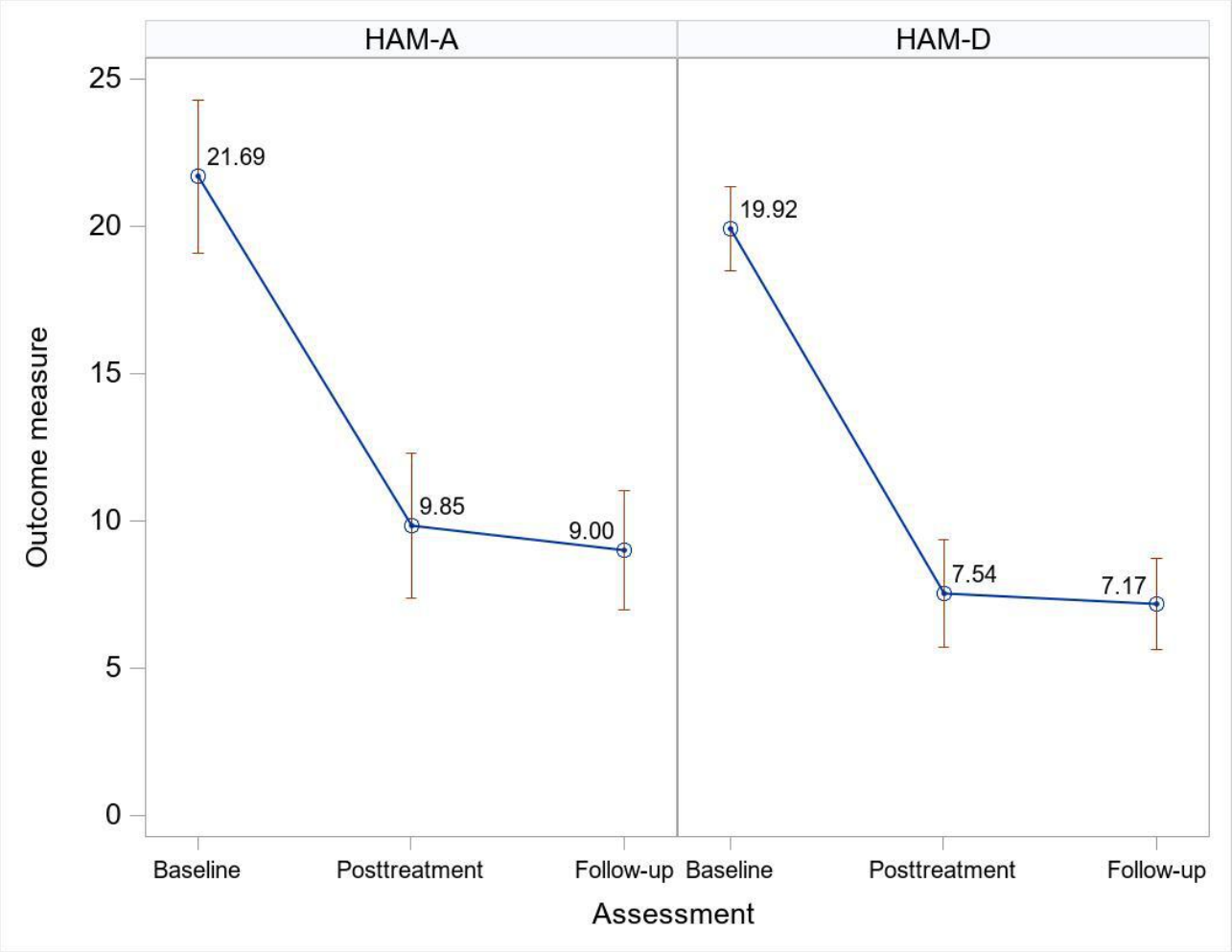
XSL•FO
RenderX

Figure 3. Individual profiles measured using the Hamilton Anxiety Rating Scale (HAM-A).



XSL•FO
RenderX

Figure 4. Mean profile for Hamilton Anxiety (HAM-A) and Hamilton Depression (HAM-D).



With respect to anxiety, we observed a significant average reduction of 11.84 (SE 2.44) from baseline (mean 21.69, SD 9.35) to post treatment (mean 9.85, SD 8.90; t_{13}=4.85, P<.001). From post treatment (mean 9.85, SD 8.90) to follow-up (mean 9.00, SD 7.01), we observed a nonsignificant reduction of 1.24 (SE 2.51; t_{13}=0.48, P=.64). Cumulatively from baseline (mean 21.69, SD 9.35) to follow-up (mean 9.00, SD 7.01) for anxiety, we observed a significant average reduction of 13.05 (SE 2.51; t_{13}=5.21, P<.001), corresponding to a mean shift from moderate to mild anxiety between baseline and follow-up. Partial η estimates quantifying the size of the effect for the change from baseline to post treatment and follow-up were η^2_p=0.80 (95% CI 0.48-0.90) and η^2_p=0.82 (95% CI 0.52-0.91), respectively, corresponding to large effects. Similarly, for depression, we observed a significant average reduction of 12.39 (SE 1.50) from baseline (mean 19.92, SD 5.14) to post treatment (mean 7.54, SD 6.63; t_{13}=8.24, P<.001). From post treatment (mean 7.54, SD 6.63) to follow-up (mean 7.17, SD 5.36), we observed a nonsignificant reduction of 0.44 (SE 1.55; t_{13}=0.29, P=.78). Cumulatively from baseline (mean 19.92, SD 5.14) to follow-up (mean 7.17, SD 5.36), we observed a significant average reduction of 12.83 (SE 1.55; t_{13}=8.30, P<.001), corresponding to a mean shift from moderate to mild/nondepressed between baseline and follow-up. Partial η estimates quantifying the size of the effect for the change from baseline to post treatment and follow-up were η^2_p=0.92 (95% CI 0.76-0.95) and η^2_p=0.92 (95% CI 0.76-0.96), respectively, corresponding to large effects.

Based on published research, we define CSR as a 4-unit change in both HAM-A [63] and HAM-D [64]. Using all 13 participants, with nonresponse corresponding to not achieving CSR, at post treatment, 84.6% (95% CI 54.5%-98.0%) achieved CSRs for both anxiety and depression. At follow-up, 76.9% (95% CI 46.2%-95.0%) achieved a CSR in anxiety and 92.3% (95% CI 74.0%-99.0%) in depression. Interestingly, 1 participant (ID=8) started the study without an elevation in the HAM-A or HAM-D but still displayed a CSR in both. As CSR can depend on initial measurement score, the percentage of participants achieving a symptom reduction ranging from 10% to 90% (in 10% increments) from baseline to post treatment and baseline to follow-up are shown in Tables 3 and 4, respectively.

XSL•FO
RenderX

Table . Percentage of participants reaching percentage thresholds of symptom reduction from baseline to post treatment (N=13).[a]

Symptom reduction[a]	HAM-D[b], percentage of participants (95% CI[c])	HAM-A[d], percentage of participants (95% CI[c])
10% reduction	100 (75.3 - 100)	84.6 (54.6 - 98.1)
20% reduction	92.3 (64.0 - 99.8)	76.9 (46.2 - 95.0)
30% reduction	84.6 (54.6 - 98.1)	69.2 (38.6 - 90.9)
40% reduction	69.2 (38.6 - 90.9)	61.5 (31.6 - 86.1)
50% reduction	69.2 (38.6 - 90.9)	61.5 (31.6 - 86.1)
60% reduction	61.5 (31.6 - 86.1)	53.8 (25.1-80.8)
70% reduction	53.8 (25.1 - 80.8)	46.2 (19.2-74.9)
80% reduction	38.5 (13.9 - 68.4)	46.2 (19.2-74.9)
90% reduction	23.1 (5.04 - 53.8)	7.69 (0.19-36.0)

[a]Thresholds of symptom reduction from baseline to post treatment in 10% increments.

[b]HAM-D: Hamilton Depression Rating Scale.

[c]95% CIs are Clopper-Pearson exact confidence bounds.

[d]HAM-A: Hamilton Anxiety Rating Scale.

Table . Percentage of participants reaching percentage thresholds of symptom reduction from baseline to follow-up (N=12).[a]

Symptom reduction[b]	HAM-D[c], percentage of participants (95% CI)	HAM-A[d], percentage of participants (95% CI)
10% reduction	100 (75.3-100)	91.70 (61.5-99.8)
20% reduction	100 (75.3-100)	83.30 (51.6-97.9)
30% reduction	100 (75.3-100)	83.30 (51.6-97.9)
40% reduction	83.30 (51.6-97.9)	83.30 (51.6-97.9)
50% reduction	83.30 (51.6-97.9)	83.30 (51.6-97.9)
60% reduction	58.30 (27.7-84.8)	66.70 (34.9-90.1)
70% reduction	41.70 (15.2-72.3)	41.70 (15.2-72.3)
80% reduction	33.30 (9.92-65.1)	25.00 (5.49-57.2)
90% reduction	0.00 (0.00-26.5)	8.33 (0.21-38.5)

[a]One participant was excluded due to not providing follow-up data. 95% CIs are Clopper-Pearson exact confidence bounds.

[b]Thresholds of symptom reduction from baseline to 4-week follow-up in 10% increments.

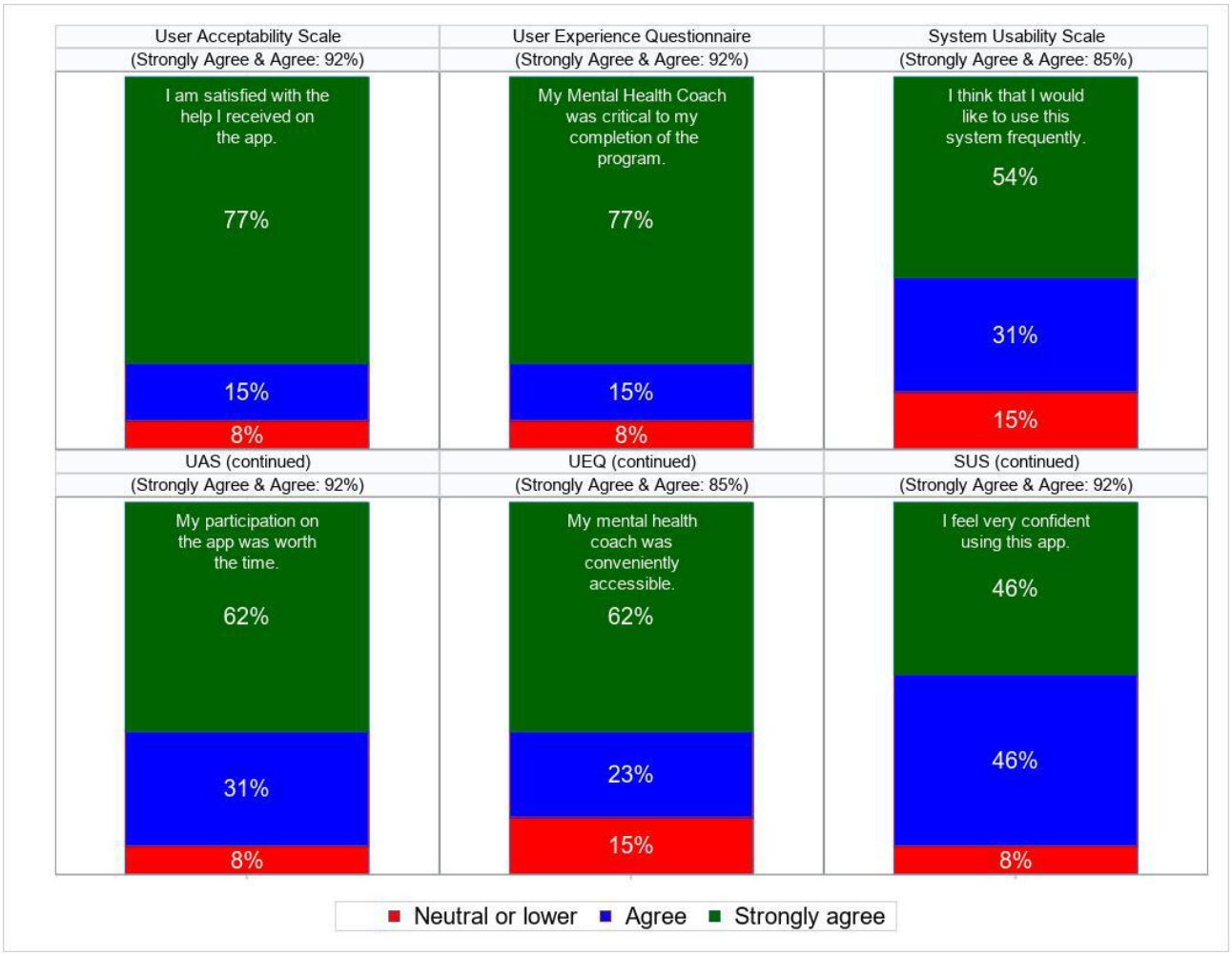
[c]HAM-D: Hamilton Depression Rating Scale.

[d]HAM-A: Hamilton Anxiety Rating scale.

Figure 5 summarizes the acceptability, experience, and usability of the Rauha smartphone platform. Among participants who agreed or strongly agreed based on direct DTx experience, 85.6% (SD 10.4%) reported positive acceptability, 87.7% (SD 8.3%) reported positive experience, and 85.0% (SD 6.3%) reported positive usability. The most highly endorsed items (>90%) highlight the aspects of the program driving the positive user experience: overall satisfaction with the treatment program, relevance of treatment strategies to daily life, the central role of mental health coaches in treatment success, use of video and text delivery within Rauha, ease of understanding digital content, and confidence in using the platform. Importantly, there was no negative feedback on any item across these self-report surveys and no accessibility issues, given the WCAG 2.2 AA standards of the smartphone program. These findings underscore that participants valued not just general satisfaction but specific program components that supported engagement, skill application, and confidence in using Rauha effectively.

XSL•FO
RenderX

Figure 5. Ratings of acceptability, user experience, and usability of the Rauha program. UAS: User Acceptability Scale; UEQ: User Experience Questionnaire; SUS: System Usability Scale.



Regarding replay rate, participants revisited previously completed CBT lessons multiple times over the course of the intervention, with an average replay rate of 5.5. Replay behavior was derived from platform usage analytics, with some modules showing higher repeat access based on user interaction patterns (eg, repeated entry and time spent within lessons). In the context of consistently high acceptability, experience, and usability ratings, this pattern of re-engagement is more consistent with voluntary reuse of content than with difficulty understanding the material.

Discussion

Principal Findings

The central aim of this study was to evaluate Rauha—a mental health DTx intervention combining an asynchronous, CBT-based treatment with a synchronous NBC-HWC implementing behavior change strategies—in treating symptoms of moderate-to-severe anxiety and depression in adults with hearing and/or mobility-related disabilities. We hypothesized that (1) there would be a treatment retention rate of at least 80%, and (2) there would be a CSR in internalizing symptoms from pretreatment to post treatment and through follow-up. The data supported both hypotheses, with a retention rate of 100% and a statistically and clinically significant decrease in both anxiety and depression symptoms on average that was sustained during the 4-week follow-up. Furthermore, this average decrease in symptoms far exceeded the minimum threshold for clinical significance by a magnitude of 3, such that participants had even more benefit than expected.

This is the first known study to assess the treatment effect of a WCAG 2.2 AA–certified platform-aligned DTx combined with a weekly, virtual NBC-HWC to treat moderate to severe symptoms of anxiety and depression in adults with disabilities. In addition, as noted previously, 1 participant with subclinical baseline levels of both anxiety and depression nevertheless achieved a CSR despite a generally lower capacity for symptom change among those with lower baseline severity [9]. Given these results, this study offers strong preliminary data to support this novel approach to treatment. In addition, the 100% retention rate also speaks to the acceptability and engagement of the treatment itself, given that treatment retention is often an issue both in research [65] and clinical [66] settings. These findings extend the overall evidence of CBT-based interventions to less-used treatment modalities while also being applied to an underserved clinical population, adults with disabilities and functional limitations. The clinically significant outcomes from this study are either commensurate with or exceed many previous outcome studies, including internet-based CBT [18,67,68], traditional CBT delivered as weekly individual

XSL•FO
RenderX

therapy [69-71], combined internet-based and individual face-to-face CBT [70,72-74], and CBT combined with pharmacotherapy [75].

At a systems level, and as it relates to barriers prohibiting access to adequate treatment, both of the Rauha treatment components can increase overall access to care, as self-guided treatment does not require time from a provider for a patient to receive care, and there are over 12,000 NBC-HWCs that can be considered part of the otherwise limited mental health resources to drastically expand treatment options [76]. Thus, CBT and other evidence-based approaches can be highly scaled with this approach without sacrificing clinical impact.

Strengths and Limitations

This study has several notable strengths. First, it evaluated a novel, digitally delivered CBT intervention targeting an underserved population with nuanced physical and mental health needs, addressing a critical gap in accessible mental health care. These findings extend our understanding of the potential for nontraditional treatment modalities and suggest that CBT can remain clinically impactful even when delivered through a smartphone-based DTx. Second, the use of clinical significance, in addition to statistical significance, provides a richer assessment of treatment impact, highlighting the clinical meaningfulness of Rauha for patients. Third, the inclusion of a 4-week postintervention follow-up allowed for a preliminary assessment of the durability of symptom improvements. While this time frame provides early support for the maintenance of treatment effects, longer follow-up periods are needed to fully characterize sustained outcomes in anxiety and depression. Interestingly, there was an increase in clinically significant improvement in depression between post treatment and follow-up, which may be attributable to participants having continued access to the smartphone-based CBT program to revisit relevant clinical content, though the robustness of such a finding would require further research. Finally, assessing user acceptability, experience, and usability offers greater insight into the participant's experience with the DTx, demonstrating highly positive views regarding the ease of use, DTx content, and coaching sessions. Together, these elements underscore the potential of Rauha as a scalable, accessible, and evidence-informed treatment.

This study has several limitations that should be considered when interpreting the findings. First, the sample size was small, which limits statistical power, the precision of estimates, and the generalizability of findings. Accordingly, the observed large effect sizes should be interpreted with some caution and considered preliminary support for proof-of-concept requiring replication in higher-powered clinical studies. Second, the participant population was limited to individuals with hearing and mobility disabilities, and researchers allowed certain demographic disclosures to be optional for participant comfort. Future efforts could expand representation across a broader range of functional and accessibility challenges and require demographic reporting, which would enhance generalizability. Third, the study did not include a control condition, which limits formal causal inference such that results may reflect uncontrolled factors (eg, natural symptom fluctuation, placebo effects, regression to the mean, or untracked changes in medication during the intervention). Given this, future studies incorporating a control condition are planned to rigorously assess causal treatment effects. Fourth, participants were recruited in part through a mental health clinic, which may introduce selection bias, as the individuals referred may differ from the broader population in terms of treatment readiness, clinical profiles, and motivation to engage in care. Future studies can address this by using more diverse recruitment strategies to strengthen generalizability. Finally, because Rauha combines a self-guided digital component with NBC-HWC support, it is not possible to attribute the observed outcomes to either component alone. That is, the present design does not allow for attribution of effects to either the digital CBT component or the coaching component independently. Future studies using component analyses or factorial designs are needed to isolate the mechanisms of change and optimize the intervention itself by identifying the most impactful factors. Collectively, these limitations highlight the need for continued research while underscoring the promising potential of Rauha to improve outcomes in this population.

Conclusions

To our knowledge, this is the first study to evaluate a hybrid digital CBT intervention—combining self-guided, smartphone-delivered CBT with live weekly sessions led by a NBC-HWC—for adults with disabilities experiencing anxiety and/or depression. The findings provide preliminary support for clinical benefit of Rauha, high engagement, accessibility, and scalability, while identifying important directions for future research. Importantly, this work supports the potential of a coordinated digital CBT and coaching model to address persistent barriers to care and deliver flexible, clinically meaningful support to populations historically underserved in mental health services. Overall, Rauha represents a promising, evidence-informed approach that extends the reach of CBT and highlights the value of integrating DTx with coaching to improve mental health outcomes among adults with functional and accessibility challenges.

Acknowledgments

The authors declare that generative artificial intelligence tools were not used in the research and writing process.

Funding

This study was funded internally by Toivoa Inc.

XSL•FO
RenderX

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Jill Palmer, DSW, LCSW, served as the site principal investigator at Endeavors Programs, San Antonio, Texas, overseeing the study conduct.

Conflicts of Interest

LER is the founder and CEO of Toivoa Inc (developer of Rauha), BDD is the cofounder of Toivoa Inc, and AK and CSK are paid employees of Toivoa Inc.

Multimedia Appendix 1
Rauha 8-week cognitive behavioral therapy digital therapeutic program.
[DOCX File, 22 KB - humanfactors_v13i1e92448_app1.docx]

References

1. Terlizzi EP, Zablotsky B. Symptoms of anxiety and depression among adults: United States, 2019 and 2022. Natl Health Stat Report 2024 Nov 7(213):CS353885. [doi: 10.15620/cdc/64018] [Medline: 39591466]
2. Ramachandiran AK, Gunning F, Santillana M, et al. Sociodemographic disparities, healthcare system trust, and social support in mental health treatment among U.S. adults with depressive or anxiety symptoms. J Mood Anxiety Disord 2026 Mar;13:100166. [doi: 10.1016/j.xjmad.2026.100166]
3. Zajacova A, Montez JK. Explaining the increasing disability prevalence among mid-life US adults, 2002 to 2016. Soc Sci Med 2018 Aug;211:1-8. [doi: 10.1016/j.socscimed.2018.05.041] [Medline: 29864712]
4. Cree RA, Okoro CA, Zack MM, Carbone E. Frequent mental distress among adults, by disability status, disability type, and selected characteristics—United States, 2018. MMWR Morb Mortal Wkly Rep 2020 Sep 11;69(36):1238-1243. [doi: 10.15585/mmwr.mm6936a2] [Medline: 32914770]
5. Gréaux M, Moro MF, Kamenov K, Russell AM, Barrett D, Cieza A. Health equity for persons with disabilities: a global scoping review on barriers and interventions in healthcare services. Int J Equity Health 2023 Nov 13;22(1):236. [doi: 10.1186/s12939-023-02035-w] [Medline: 37957602]
6. Manning RB, Cipollina R, Lowe SR, et al. Barriers to mental health service use among people with disabilities during the COVID-19 pandemic. Rehabil Psychol 2023 Nov;68(4):351-361. [doi: 10.1037/rep0000512] [Medline: 37470994]
7. Mental Health Care Health Professional Shortage Areas (HPSAs). Kaiser Family Foundation (KFF). URL: https://www.kff.org/other-health/state-indicator/mental-health-care-health-professional-shortage-areas-hpsas/ [accessed 2025-09-08]
8. Key Substance Use and Mental Health Indicators in the United States: Results from the 2024 National Survey on Drug Use and Health. : Center for Behavioral Health Statistics and Quality (CBHSQ), Substance Abuse and Mental Health Services Administration (SAMHSA), U.S. Department of Health and Human Services; 2025 URL: https://www.samhsa.gov/data/report/2023-nsduh-annual-national-report [accessed 2025-09-08]
9. Mohr DC, Hart S, Vella L. Reduction in disability in a randomized controlled trial of telephone-administered cognitive-behavioral therapy. Health Psychol 2007 Sep;26(5):554-563. [doi: 10.1037/0278-6133.26.5.554] [Medline: 17845107]
10. Steinberg R, Posa S, Pattni N, et al. Psychosocial group therapy interventions for patients with physical disabilities: a scoping review of implementation considerations. Rehabil Psychol 2023 Aug;68(3):235-260. [doi: 10.1037/rep0000491] [Medline: 37079820]
11. Trott M, Koblitz A, Pardhan S. Cognitive behavioural therapy versus other interventions on mental health in people with sensory impairments: a systematic review and meta-analysis. J Psychosom Res 2025 Feb;189:111998. [doi: 10.1016/j.jpsychores.2024.111998] [Medline: 39647352]
12. Tan Y, Liang X, Ming W, et al. The efficacy of telemedicine interventions on quality of life and depression in individuals with spinal cord injury: a systematic review and meta-analysis. Front Psychiatry 2025;16:1434376. [doi: 10.3389/fpsyt.2025.1434376] [Medline: 39916741]
13. Friedman C, VanPuymbrouck L. Telehealth use by persons with disabilities during the COVID-19 pandemic. Int J Telerehabil 2021;13(2):e6402. [doi: 10.5195/ijt.2021.6402] [Medline: 35646237]
14. Jacobs JC, Blonigen DM, Kimerling R, et al. Increasing mental health care access, continuity, and efficiency for veterans through telehealth with video tablets. Psychiatr Serv 2019 Nov 1;70(11):976-982. [doi: 10.1176/appi.ps.201900104] [Medline: 31378193]
15. Pullyblank K, Krupa N, Scribani M, Chapman A, Kern M, Brunner W. Examining differences in time to appointment and no-show rates between rural telehealth users and non-users. Front Digit Health 2024;6(1264893):1264893. [doi: 10.3389/fdgth.2024.1264893] [Medline: 38343906]

XSL•FO
RenderX

16. Sun CF, Correll CU, Trestman RL, et al. Low availability, long wait times, and high geographic disparity of psychiatric outpatient care in the US. Gen Hosp Psychiatry 2023;84:12-17. [doi: 10.1016/j.genhosppsych.2023.05.012] [Medline: 37290263]
17. Karyotaki E, Efthimiou O, Miguel C, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry 2021 Apr 1;78(4):361-371. [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
18. Bress JN, Falk A, Schier MM, et al. Efficacy of a mobile app-based intervention for young adults with anxiety disorders: a randomized clinical trial. JAMA Netw Open 2024 Aug 1;7(8):e2428372. [doi: 10.1001/jamanetworkopen.2024.28372] [Medline: 39163044]
19. Wang C, Lee C, Shin H. Digital therapeutics from bench to bedside. NPJ Digit Med 2023 Mar 10;6(1):38. [doi: 10.1038/s41746-023-00777-z] [Medline: 36899073]
20. Comtois KA, Mata-Greve F, Johnson M, Pullmann MD, Mosser B, Arean P. Effectiveness of mental health apps for distress during COVID-19 in US unemployed and essential workers: remote pragmatic randomized clinical trial. JMIR mHealth uHealth 2022 Nov 7;10(11):e41689. [doi: 10.2196/41689] [Medline: 36191176]
21. Zainal NH, Soh CP, Van Doren N, Benjet C. Do the effects of internet-delivered cognitive-behavioral therapy (i-CBT) last after a year and beyond? A meta-analysis of 154 randomized controlled trials (RCTs). Clin Psychol Rev 2024 Dec;114:102518. [doi: 10.1016/j.cpr.2024.102518] [Medline: 39579466]
22. Etzelmueller A, Vis C, Karyotaki E, et al. Effects of internet-based cognitive behavioral therapy in routine care for adults in treatment for depression and anxiety: systematic review and meta-analysis. J Med Internet Res 2020 Aug 31;22(8):e18100. [doi: 10.2196/18100] [Medline: 32865497]
23. Rothman B, Slomkowski M, Speier A, et al. A digital therapeutic (CT-152) as adjunct to antidepressant medication: a phase 3 randomized controlled trial (the Mirai study). J Affect Disord 2025 Nov 1;388:119409. [doi: 10.1016/j.jad.2025.119409] [Medline: 40378969]
24. Rehman A, Jamil T, Baloch SK, Rehman MSU. Digital therapeutics in mental health care: a critical evaluation of mobile applications and online interventions for anxiety and depression. Rev Appl Manag Soc Sci 2024 Nov 9;7(4):359-372. [doi: 10.47067/ramss.v7i4.386]
25. Nwosu A, Boardman S, Husain MM, Doraiswamy PM. Digital therapeutics for mental health: is attrition the Achilles heel? Front Psychiatry 2022;13:900615. [doi: 10.3389/fpsyt.2022.900615] [Medline: 35982936]
26. Borghouts J, Eikey E, Mark G, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387. [doi: 10.2196/24387] [Medline: 33759801]
27. Ciharova M, Cuijpers P, Amanvermez Y, et al. Use of tailoring features and reasons for dropout in a guided internet-based transdiagnostic individually-tailored cognitive behavioral therapy for symptoms of depression and/or anxiety in college students. Internet Interv 2023;34:100646. [doi: 10.1016/j.invent.2023.100646] [Medline: 38099094]
28. Seittu HAM, Falk T, Bhatnagar K, Saarni SE. Therapists' role in patient adherence to internet-based cognitive behavioral therapy: qualitative study. J Med Internet Res 2025 Sep 10;27(1):e71852. [doi: 10.2196/71852] [Medline: 40929730]
29. Brotherdale R, Berry K, Bucci S. A qualitative study exploring the digital therapeutic alliance with fully automated smartphone apps. Digit Health 2024;10:20552076241277712. [doi: 10.1177/20552076241277712] [Medline: 39687527]
30. Global report on health equity for persons with disabilities. : World Health Organization; 2022 URL: https://www.who.int/publications/i/item/9789240063600 [accessed 2026-06-05]
31. Web Content Accessibility Guidelines (WCAG) 2.2. World Wide Web Consortium (W3C). URL: https://www.w3.org/TR/WCAG22/ [accessed 2025-12-11]
32. Wolever RQ, Simmons LA, Sforzo GA, et al. A systematic review of the literature on health and wellness coaching: defining a key behavioral intervention in healthcare. Glob Adv Health Med 2013 Jul;2(4):38-57. [doi: 10.7453/gahmj.2013.042] [Medline: 24416684]
33. Boehmer KR, Álvarez-Villalobos NA, Barakat S, et al. The impact of health and wellness coaching on patient-important outcomes in chronic illness care: a systematic review and meta-analysis. Patient Educ Couns 2023 Dec;117:107975. [doi: 10.1016/j.pec.2023.107975] [Medline: 37738790]
34. Mohr DC, Duffecy J, Ho J, et al. A randomized controlled trial evaluating a manualized TeleCoaching protocol for improving adherence to a web-based intervention for the treatment of depression. PLoS One 2013;8(8):e70086. [doi: 10.1371/journal.pone.0070086] [Medline: 23990896]
35. Gierisch JM, Hughes JM, Edelman D, et al. The effectiveness of health coaching. : Veterans Health Administration, Department of Veterans Affairs, Health Services Research & Development Service (HSR&D), Evidence-based Synthesis Program (ESP); 2017 URL: https://www.hsrd.research.va.gov/publications/esp/health-coaching-REPORT.pdf [accessed 2026-06-05]
36. The WebAIM Million—the 2025 report on the accessibility of the top 1,000,000 home pages. WebAIM. URL: https://webaim.org/projects/million/2025 [accessed 2026-01-19]
37. Beck JS. Cognitive Behavior Therapy: Basics and Beyond, 3rd edition: The Guilford Press; 2020.
38. CCBHC Impact Report. : National Council for Mental Wellbeing; 2024 URL: https://www.thenationalcouncil.org/resources/2024-ccbhc-impact-report/ [accessed 2025-09-08]

XSL•FO
RenderX

39. Psychologists reaching their limits as patients present with worsening symptoms year after year. : American Psychological Association; 2023 URL: https://www.apa.org/pubs/reports/practitioner/2023-practitioner-pulse-survey.pdf [accessed 2026-04-29]
40. Andersson G, Titov N. Advantages and limitations of internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11. [doi: 10.1002/wps.20083] [Medline: 24497236]
41. Bunyi J, Ringland KE, Schueller SM. Accessibility and digital mental health: considerations for more accessible and equitable mental health apps. Front Digit Health 2021;3:742196. [doi: 10.3389/fdgth.2021.742196] [Medline: 34713206]
42. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30. [doi: 10.2196/jmir.1602] [Medline: 21393123]
43. Hamilton M. The assessment of anxiety states by rating. Br J Med Psychol 1959;32(1):50-55. [doi: 10.1111/j.2044-8341.1959.tb00467.x] [Medline: 13638508]
44. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960 Feb;23(1):56-62. [doi: 10.1136/jnnp.23.1.56] [Medline: 14399272]
45. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
46. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
47. Diagnostic and Statistical Manual of Mental Disorders, 5th edition: American Psychiatric Association; 2013. [doi: 10.1176/appi.books.9780890425596]
48. Löwe B, Decker O, Müller S, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
49. Posner K, Brown GK, Stanley B, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry 2011 Dec;168(12):1266-1277. [doi: 10.1176/appi.ajp.2011.10111704] [Medline: 22193671]
50. Matza LS, Morlock R, Sexton C, Malley K, Feltner D. Identifying HAM-A cutoffs for mild, moderate, and severe generalized anxiety disorder. Int J Methods Psychiatr Res 2010 Dec;19(4):223-232. [doi: 10.1002/mpr.323] [Medline: 20718076]
51. Kummer A, Cardoso F, Teixeira AL. Generalized anxiety disorder and the Hamilton Anxiety Rating Scale in Parkinson's disease. Arq Neuropsiquiatr 2010 Aug;68(4):495-501. [doi: 10.1590/s0004-282x2010000400005] [Medline: 20730299]
52. Zimmerman M, Martinez JH, Young D, Chelminski I, Dalrymple K. Severity classification on the Hamilton Depression Rating Scale. J Affect Disord 2013 Sep 5;150(2):384-388. [doi: 10.1016/j.jad.2013.04.028] [Medline: 23759278]
53. Carmody TJ, Rush AJ, Bernstein I, et al. The Montgomery Asberg and the Hamilton ratings of depression: a comparison of measures. Eur Neuropsychopharmacol 2006 Dec;16(8):601-611. [doi: 10.1016/j.euroneuro.2006.04.008] [Medline: 16769204]
54. van Beugen S, Ferwerda M, Hoeve D, et al. Internet-based cognitive behavioral therapy for patients with chronic somatic conditions: a meta-analytic review. J Med Internet Res 2014 Mar 27;16(3):e88. [doi: 10.2196/jmir.2777] [Medline: 24675372]
55. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. World Psychiatry 2019 Feb;18(1):20-28. [doi: 10.1002/wps.20610] [Medline: 30600624]
56. Mehta S, Peynenburg VA, Hadjistavropoulos HD. Internet-delivered cognitive behaviour therapy for chronic health conditions: a systematic review and meta-analysis. J Behav Med 2019 Apr;42(2):169-187. [doi: 10.1007/s10865-018-9984-x] [Medline: 30387008]
57. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
58. Molenberghs G, Verbeke G. Linear Mixed Models for Longitudinal Data: Springer; 2000. [doi: 10.1007/978-1-4419-0300-6]
59. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: Routledge; 2013. [doi: 10.4324/9780203771587]
60. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [doi: 10.1037//0022-006x.59.1.12] [Medline: 2002127]
61. DeRubeis RJ, Hollon SD, Amsterdam JD, et al. Cognitive therapy vs medications in the treatment of moderate to severe depression. Arch Gen Psychiatry 2005 Apr;62(4):409-416. [doi: 10.1001/archpsyc.62.4.409] [Medline: 15809408]
62. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology 1990 Jan;1(1):43-46. [Medline: 2081237]
63. Fan JQ, Lu WJ, Tan WQ, et al. Effectiveness of acupuncture for anxiety among patients with Parkinson disease: a randomized clinical trial. JAMA Netw Open 2022 Sep 1;5(9):e2232133. [doi: 10.1001/jamanetworkopen.2022.32133] [Medline: 36129711]
64. Rush AJ, South C, Jain S, et al. Clinically significant changes in the 17- and 6-item Hamilton Rating Scales for depression: a STAR*D report. Neuropsychiatr Dis Treat 2021;17:2333-2345. [doi: 10.2147/NDT.S305331] [Medline: 34295161]
65. Cooper AA, Conklin LR. Dropout from individual psychotherapy for major depression: a meta-analysis of randomized clinical trials. Clin Psychol Rev 2015 Aug;40:57-65. [doi: 10.1016/j.cpr.2015.05.001] [Medline: 26067572]
66. Henzen A, Moeglin C, Giannakopoulos P, Sentissi O. Determinants of dropout in a community-based mental health crisis centre. BMC Psychiatry 2016 Apr 19;16(1):111. [doi: 10.1186/s12888-016-0819-4] [Medline: 27095462]

XSL•FO
RenderX

67. Hallgren M, Helgadóttir B, Herring MP, et al. Exercise and internet-based cognitive-behavioural therapy for depression: multicentre randomised controlled trial with 12-month follow-up. Br J Psychiatry 2016 Nov;209(5):414-420. [doi: 10.1192/bjp.bp.115.177576] [Medline: 27609813]
68. Richards D, Enrique A, Eilert N, et al. A pragmatic randomized waitlist-controlled effectiveness and cost-effectiveness trial of digital interventions for depression and anxiety. NPJ Digit Med 2020;3(1):85. [doi: 10.1038/s41746-020-0293-8] [Medline: 32566763]
69. Dobkin RD, Menza M, Allen LA, et al. Cognitive-behavioral therapy for depression in Parkinson's disease: a randomized, controlled trial. Am J Psychiatry 2011 Oct;168(10):1066-1074. [doi: 10.1176/appi.ajp.2011.10111669] [Medline: 21676990]
70. Mathiasen K, Andersen TE, Lichtenstein MB, et al. The clinical effectiveness of blended cognitive behavioral therapy compared with face-to-face cognitive behavioral therapy for adult depression: randomized controlled noninferiority trial. J Med Internet Res 2022 Sep 7;24(9):e36577. [doi: 10.2196/36577] [Medline: 36069798]
71. Richards DA, Rhodes S, Ekers D, et al. Cost and Outcome of BehaviouRal Activation (COBRA): a randomised controlled trial of behavioural activation versus cognitive-behavioural therapy for depression. Health Technol Assess 2017 Aug;21(46):1-366. [doi: 10.3310/hta21460] [Medline: 28857042]
72. Nakao S, Nakagawa A, Oguchi Y, et al. Web-based cognitive behavioral therapy blended with face-to-face sessions for major depression: randomized controlled trial. J Med Internet Res 2018 Sep 21;20(9):e10743. [doi: 10.2196/10743] [Medline: 30249583]
73. Kemmeren LL, van Schaik A, Draisma S, Kleiboer A, Riper H, Smit JH. Effectiveness of blended cognitive behavioral therapy versus treatment as usual for depression in routine specialized mental healthcare: E-COMPARED trial in the Netherlands. Cogn Ther Res 2023;47(3):386-398. [doi: 10.1007/s10608-023-10363-y]
74. Lungu A, Jun JJ, Azarmanesh O, Leykin Y, Chen CEJ. Blended care-cognitive behavioral therapy for depression and anxiety in real-world settings: pragmatic retrospective study. J Med Internet Res 2020 Jul 6;22(7):e18723. [doi: 10.2196/18723] [Medline: 32628120]
75. Nakagawa A, Mitsuda D, Sado M, et al. Effectiveness of supplementary cognitive-behavioral therapy for pharmacotherapy-resistant depression: a randomized controlled trial. J Clin Psychiatry 2017;78(8):1126-1135. [doi: 10.4088/JCP.15m10511] [Medline: 28252882]
76. NBHWC welcomes new executive director and board chair to lead the future of health and wellness coaching. PR Newswire. URL: https://www.prnewswire.com/news-releases/nbhwc-welcomes-new-executive-director-and-board-chair-to-lead-the-future-of-health-and-wellness-coaching-302457952.html [accessed 2025-12-11]

Abbreviations

C-SSRS: Columbia Suicide Severity Rating Scale
CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
CSR: clinically significant reduction
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
DTx: digital therapeutic
GAD-7: Generalized Anxiety Disorder-7
HAM-A: Hamilton Anxiety Rating Scale
HAM-D: Hamilton Depression Rating Scale
HIPAA: Health Insurance Portability and Accountability Act
MEM: mixed-effects modeling
NBC-HWC: National Board-Certified Health and Wellness Coach
PHQ-9: Patient Health Questionnaire-9
WCAG: Web Content Accessibility Guidelines

Edited by D Gooch; submitted 29.Jan.2026; peer-reviewed by A Nair, C Br; revised version received 01.May.2026; accepted 27.May.2026; published 25.Jun.2026.

Please cite as:
Kirk A, King CS, Gallop R, Doan BD, Simpson Rowe L, Randa LE
Evaluating a Dual Digital Cognitive Behavioral Therapy and Health and Wellness Coaching Intervention for Anxiety and Depression: Single-Arm Pilot Study
JMIR Hum Factors 2026;13:e92448
URL: https://humanfactors.jmir.org/2026/1/e92448
doi:10.2196/92448

XSL•FO
RenderX

© Alex Kirk, Christopher S King, Robert Gallop, Brian D Doan, Lorelei Simpson Rowe, Laura E Randa. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 25.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Patient Experiences of Digital Technology Use in Interstitial Lung Disease (PRODIGY-ILD Study): Qualitative Study Using Reflexive Thematic Analysis

Emer Gunne[1], MBChB, MSc; Alessandro Franciosi[1,2], PhD; Cormac McCarthy[1,2], PhD; Michael P Keane[2,3], MD; Peter Doran[1,2], PhD; Sinead Holden[4], PhD

[1]Clinical Research Centre, School of Medicine, University College Dublin, Elm Park, Dublin, Ireland

[2]St. Vincent's University Hospital, Dublin, Leinster, Ireland

[3]Belfield, College of Health and Agricultural Science, University College Dublin, Dublin, Leinster, Ireland

[4]Belfield, School of Public Health, Physiotherapy & Sports Science, University College Dublin, Dublin, Leinster, Ireland

Corresponding Author:
Emer Gunne, MBChB, MSc
Clinical Research Centre, School of Medicine, University College Dublin, Elm Park, Dublin, Ireland

Abstract

Background: Digital health technology enables collection of continuous physiological and behavioral data from participants in clinical trials. This supports hybrid trial designs, potentially reducing clinic visits and participant burden for patient monitoring. Interstitial lung disease (ILD) is characterized by an unpredictable clinical course, creating a need for new treatments and more sensitive approaches to assessing treatment effectiveness, disease progression, and clinically meaningful trial end points.

Objective: This study aimed to explore the experiences of individuals with ILD using digital tools in a clinical study to inform digital health–enabled clinical research.

Methods: This qualitative study was conducted within the PRODIGY-ILD (Predicting Outcomes using Digital Technology in Interstitial Lung Disease) cohort, a prospective observational study using wearable devices and electronic patient-reported outcome measures for a planned 3 years of longitudinal monitoring. Participants were recruited from a specialist outpatient ILD clinic. A topic guide was developed iteratively, and individual semistructured interviews were conducted remotely via Zoom (Zoom Video Communications, Inc) and/or telephone, audio-recorded, and transcribed. Data were analyzed using reflexive thematic analysis with NVivo software (Lumivero).

Results: Fifteen of the final 20 participants recruited to the PRODIGY-ILD study consented and completed interviews. Four key themes were identified, highlighting how trust, digital literacy, participant-initiated engagement with data, and illness burden shape sustained participation in digital health–enabled clinical research: (1) trust and altruism override data concerns: confidence in researchers' data handling and a desire to contribute enabled data sharing; (2) navigating digital tools: friction and flexibility: digital literacy, usability, and device compatibility varied, but participants were able to use workarounds to maintain engagement; (3) participant-initiated engagement with wearable data: participants moved from passive to active engagement, in many cases integrating devices into daily routines; and (4) life-limiting illness as a constraint on digital trial participation: managing symptoms and severe comorbidities reduces motivation and engagement with study technology.

Conclusions: Despite participants' motivations to contribute data to research, engagement was shaped by usability, participant-initiated engagement, and the constraints of living with chronic illness. There is a need for patient-centered design, tailored support, and flexible trial procedures to optimize adherence in digital health–enabled clinical research.

(JMIR Hum Factors 2026;13:e91533) doi:10.2196/91533

KEYWORDS

digital technology; interstitial lung disease; remote monitoring; clinical trials; qualitative thematic analysis

Introduction

Interstitial lung diseases (ILDs) are a heterogeneous group of chronic respiratory conditions characterized by significant symptom burden and progressive declines in lung function typical of fibrotic phenotypes [1]. The clinical course of ILD is often unpredictable, with periods of stability punctuated by episodes of deterioration that can be difficult to identify early using traditional, clinic-based measures [1]. Recent studies on the feasibility and acceptability of home monitoring to provide real-world data for patients have been promising [2-9]. Advances in digital technology, including smartphone apps and wearables,

XSL•FO
RenderX

offer opportunities to collect continuous and timely data on physical activity, symptoms, and physiological parameters to augment outpatient clinical data and to monitor participants in clinical trials [10].

Although digital monitoring in ILD has shown promise, a mixed methods study by Althobiani et al [11] suggested that patients may experience both reassurance and burden when engaging with self-management using digital monitoring, with engagement shaped by symptom severity, treatment effects, and uncertainty around how monitoring data are interpreted or acted upon clinically. However, existing studies have largely assessed the acceptability of digital monitoring using cross-sectional surveys [11] or short-term evaluations; little is known about how patients experience sustained engagement with these technologies over time. This study addresses this gap by examining the lived experience of participating in a longitudinal digital monitoring program for ILD, capturing how engagement evolves across differing clinical trajectories. Understanding how patients experience, perceive, and adapt to digital monitoring in real-world contexts is essential to ensure that these technologies are acceptable, feasible, and responsive to patient needs, facilitating sustained engagement and data quality.

PRODIGY-ILD [12] (Predicting Outcomes using Digital Technology in Interstitial Lung Disease) is a prospective cohort study designed to examine whether remotely collected electronic patient-reported outcome measures (ePROMs) and wearable-derived physiological data can identify early indicators of disease trajectory in people living with ILD, for a planned 3-year follow-up. In this study, participants completed regular symptom-based ePROMs, including weekly and monthly measures, and wore a smartwatch that passively captured behavioral and physiological data throughout follow-up, as described in the study protocol [12]. The primary quantitative analysis focus was on identifying candidate digital biomarkers and understanding temporal patterns in symptoms and physiology across the disease course. What remains underexplored is how participants live with, interpret, and adapt to digital monitoring over time, particularly when technologies are embedded within life-limiting illness. This qualitative substudy, nested within the PRODIGY-ILD study, aimed to explore participant experiences, attitudes, and perceptions of using the MyCap app (Vanderbilt University Medical Center), the companion app to REDCap (Vanderbilt University) research software for ePROMs and the Apple Watch wearable device, to inform the development of future digital health design for clinical trials [13,14].

Methods

Design and Theoretical Approach

Overview

This study was designed as a qualitative study with individual interviews. A qualitative design with individual interviews was selected to capture the depth, nuance, and lived experience of participants in a one-to-one setting, enabling a rich and contextualized understanding that cannot be fully represented through quantitative measures alone. The COREQ (Consolidated Criteria for Reporting Qualitative Research) [15] 32-item checklist (Checklist 1) informed the study design and was used for transparency and completeness of reporting interviews.

Positioning

This study was informed by an experiential critical realist stance [16], recognizing that while ILD and associated physiological changes represent real underlying phenomena, participants' experiences of symptoms and digital health technologies are subjectively perceived and socially mediated. Knowledge was therefore understood as being generated through participants' accounts of lived experience, coconstructed through interaction with the researcher and shaped by the analytical framework applied.

This epistemological position has important analytic implications, particularly given the researcher's prior involvement in the wider PRODIGY-ILD study. Rather than treating prior relationships as a source of bias to be eliminated, they were understood as a contextual feature of the research process that could shape both data generation and interpretation. Familiarity between participants and the researcher may have influenced what participants chose to disclose, how experiences were framed, and the assumptions embedded within accounts.

Reflexive thematic analysis was therefore used, explicitly acknowledging the active role of the researcher in meaning-making. Reflexivity was operationalized through ongoing critical self-reflection throughout data collection and analysis, including the use of reflexive memo-writing to document assumptions, emotional responses, and evolving interpretations, and regular revisiting of analytic decisions to consider how the researcher's positionality and prior knowledge may have shaped theme development. Themes are understood as analytic constructions rather than emergent truths, and interpretations were iteratively refined through engagement with the data and discussion within the research team, ensuring that themes remained grounded in participants' accounts while recognizing the interpretive nature of qualitative analysis.

Development of Interview Guide

The interview guide was developed through a comprehensive process that drew upon existing literature, team discussions, and pilot testing with 2 participants. The approach was iterative, allowing for continual refinement to ensure the guide was both relevant to the study objectives and responsive to participant feedback. Key topic areas included onboarding, usability, perceived usefulness, impact on daily life, privacy concerns, and suggestions for improvement. The guide consisted of open-ended questions, supplemented with targeted prompts, to facilitate in-depth exploration of participant experiences. Further details of the interview template and question structure are provided in Multimedia Appendix 1.

Participant Recruitment and Sampling

Participants were recruited from the existing PRODIGY-ILD study cohort (n=20 final number onboarded between April 30 and November 22, 2024) attending the specialist ILD outpatient clinic. The inclusion criteria were as follows: (1) age ≥18 years,

XSL•FO
RenderX

(2) diagnosis of idiopathic pulmonary fibrosis or progressive pulmonary fibrosis, and (3) ability to provide written informed consent. The exclusion criteria were as follows: (1) patients who are unwilling to wear a smart watch for the duration of the study and (2) cognitive impairment or inability to understand and follow instructions that would limit the patient's understanding of the project or the measurement.

A purposive sampling approach was used, whereby individuals were invited to participate based on their direct experience of living with ILD and engagement with wearable and app use as part of the parent study. All eligible participants were sent a text message and a hard-copy patient information leaflet and informed consent form to inform participants of the qualitative substudy (Multimedia Appendix 2). Reasons for nonparticipation included death, withdrawal from the parent study, being too unwell, or personal reasons.

Data Collection

One-on-one semistructured interviews were conducted between July 16 and August 20, 2025, using the Zoom online platform (Zoom Video Communications, Inc) to conduct and record the semistructured interview using Zoom's autotranscription function. Three participants were unable to log on to Zoom and were interviewed by phone, the audio of which was recorded and autotranscribed by Zoom. Patient A's wife was present and Patient B's son was present during the interviews; however, both answered all questions independently. Participants were interviewed once. Interviews lasted between 18 and 51 minutes and were recorded with permission. Notes were taken after each interview to document details and initial impressions. Transcripts were checked against the recordings for accuracy after each interview. All transcripts were anonymized according to COREQ. Member checking was not performed to reduce participation burden.

Data Analysis

Reflexive thematic analysis was conducted in line with Braun and Clarke's 6-phase approach [17,18], using an inductive approach, allowing patterns of meaning to emerge from participants' accounts, such that theme development was not constrained by predefined categories.

Thematic analysis proceeded through the following stages, once audiochecked anonymized transcripts were imported into NVivo qualitative data analysis software (Table 1).

Table . Practical workflow.

Step	Activity	Tool or platform
Interview and record	Semistructured interviews	Zoom and/or phone call
Transcription	Auto + manual cleaning	Word
Coding and theme generation	Manual and NVivo and Word	NVivo and Word
Reflection and reporting	Summarize and contextualize	Word and Word tables

Researchers familiarized themselves with the transcripts by reading them multiple times. They then generated initial codes, which included usability domains such as onboarding and reminders, as well as inductive codes that emerged directly from the data. These codes were subsequently grouped into candidate themes, which were reviewed and refined iteratively to ensure coherence and distinctiveness, reflecting the principles of reflexive thematic analysis. Analytic credibility was supported through the maintenance of a detailed audit trail documenting iterative theme development. Regular discussions among the research team provided a forum for reflexive dialogue, in which interpretations were questioned, alternative readings considered, and analytic decisions refined, recognizing that themes represent informed analytic judgment rather than objective consensus. Consideration was also given to the inclusion of deviant cases and contrasting viewpoints, with divergent perspectives actively sought and incorporated to ensure a comprehensive understanding of participants' experiences.

Once the themes were finalized, they were defined and named, and the findings were reported with illustrative participant quotes to provide context and depth.

NVivo was used for parent and child coding; the coding structure is detailed in Multimedia Appendix 3. Themes are presented alongside tables of relevant participant quotes, and divergent perspectives are highlighted to reflect a range of experiences. The clinical trial context is also provided to aid the interpretation of themes within the broader study framework.

Research Team and Reflexivity

Interviews were conducted by EG pursuing a PhD in digital technology for clinical trials at the time of the study. All participants were known to the researcher through their enrollment in the PRODIGY-ILD study prior to the interviews. The interviewer introduced herself in the context of her research role and explained that the interviews sought to explore participants' experiences and perspectives on wearable technology use and apps for patient-reported outcomes. Participants were aware of the overarching study aims but were not provided with the specific interview questions in advance.

This prior familiarity between the interviewer and participants may have fostered a sense of trust and openness, encouraging participants to share their honest experiences and perspectives more freely. Additionally, the interviewer's background and expertise in clinical trials and digital health may have shaped the direction of the interviews, with her insights and probing questions encouraging deeper exploration of topics related to digital technology use in health care. Overall, these factors likely contributed to a data collection process that was more attuned to the study setting, while also necessitating reflexivity to account for potential influences on participants' responses.

XSL•FO
RenderX

Ethical Considerations

Full ethics approval was granted for this research study by the St. Vincent's University Hospital Research Ethics Committee (reference number: RS23-023). Explicit consent was sought from participants for the collection and processing of their data. Data processing agreements are in place to ensure that personal data are processed as is necessary to achieve the objective of the health research and to ensure that data shall not be processed in such a way that might cause damage or distress to the participants. St. Vincent's University Hospital and University College Dublin are joint data controllers for the study. This study complies with the General Data Protection Regulation. Study results are being presented at medical conferences and disseminated via peer-reviewed journals.

Results

Overview

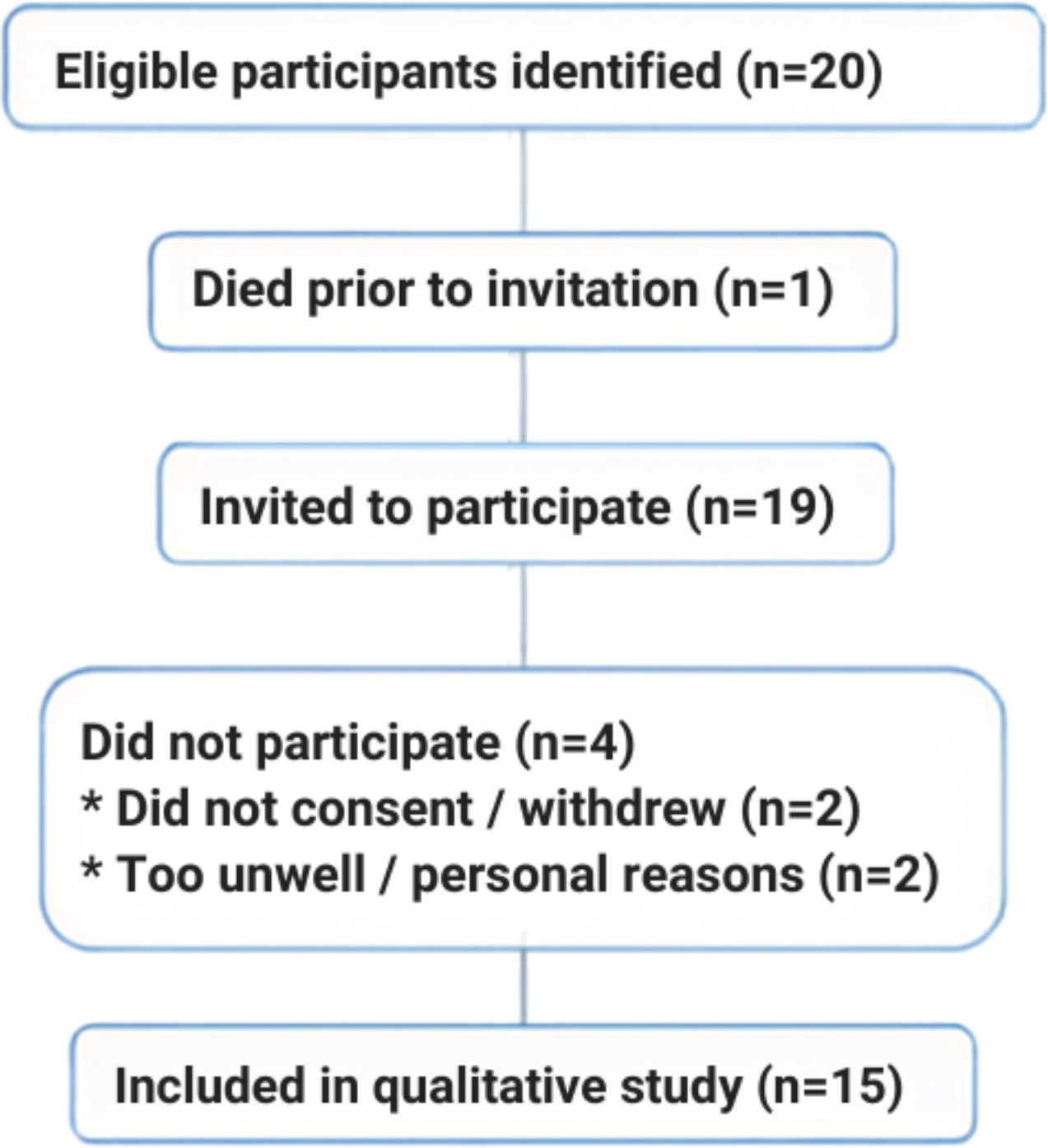
In total, 15 participants completed the qualitative study (Figure 1).

Demographics of study participants are presented in Table 2. Participants were 80% (12/15) male, 86% (13/15) Caucasian, with a mean age of 62 (SD 16.3) years. This demographic is very similar to the overall PRODIGY-ILD dataset, where participants were 80% (16/20) male, 90% (18/20) Caucasian, with a mean age of 62.5 (SD 14.8) years. Patient-reported outcome measure (PROM) completion was high over the first 12 months (overall adherence >99%). Smartwatch wear time was similarly high, with a median wear time of approximately 96% (IQR 95.9%-99.5%) over 12 months, indicating sustained engagement with the digital monitoring protocol (Multimedia Appendix 4).

Our findings produced 4 main themes with subthemes (Figure 2). NVivo parent and child codes for each theme are available in Multimedia Appendix 3.

XSL•FO
RenderX

Figure 1. Predicting Outcomes using Digital Technology in Interstitial Lung Disease qualitative study recruitment flowchart.



XSL•FO
RenderX

Table . Participant demographics (N=15).

Variable	Value
Age (years), mean (SD)	62 (16.3)
Gender, n (%)	
Men	12 (80)
Women	3 (20)
Ethnicity, n (%)	
Caucasian	13 (86)
Asian	1 (7)
Mixed	1 (7)
ILD[a] subtype, n (%)	
IPF[b]	4 (27)
CPFE[c]	3 (20)
CTD-ILD[d]	3 (20)
HP[e]	3 (20)
Fibrotic sarcoidosis	2 (13)
Time since diagnosis (years), median (IQR)	4.5 (2.0 - 8.0)
Supplemental oxygen, n (%)	4 (40)
Comorbidities, n (%)	
Cardiovascular	11 (73)
Metabolic	9 (60)
Respiratory	4 (27)
Gastrointestinal	4 (27)
Rheumatological	3 (20)
Malignancy	1 (7)

[a]ILD: interstitial lung disease.

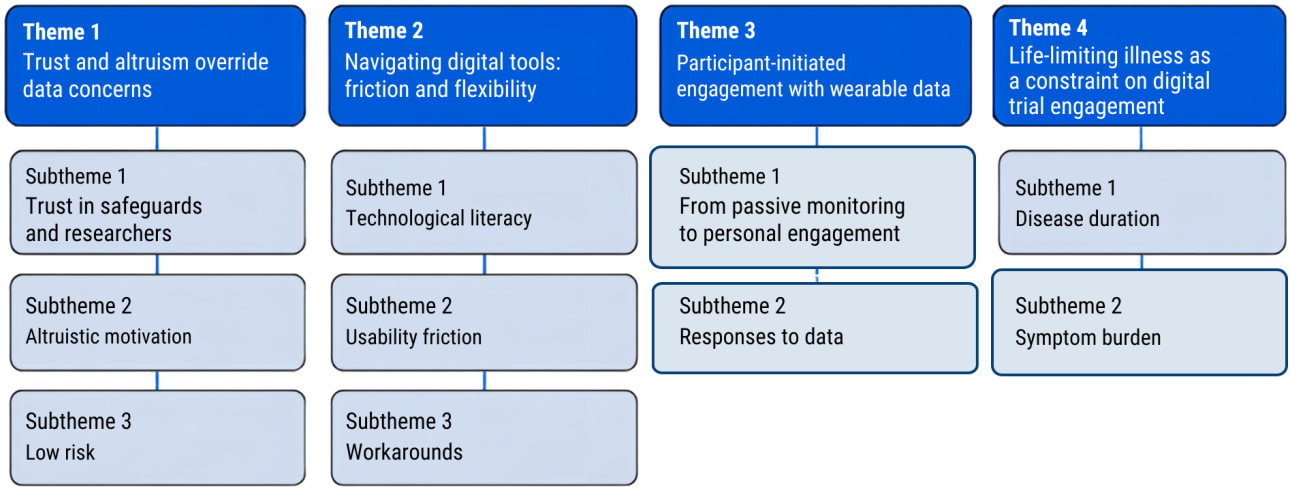
[b]IPF: idiopathic pulmonary fibrosis.

[c]CPFE: combined pulmonary fibrosis and emphysema.

[d]CTD-ILD: connective tissue disease-associated interstitial lung disease.

[e]HP: hypersensitivity pneumonitis.

Figure 2. Themes and subthemes from qualitative analysis.



XSL•FO
RenderX

Theme 1: Trust and Altruism Override Data Concerns

Reflexive Thematic Concept

In the context of digital surveillance within clinical trials, participants' decisions to share personal health data are not solely determined by individual privacy concerns, but are influenced by confidence in research teams, ethical frameworks, and the belief that their contributions have meaningful impact. This analysis synthesizes 3 core subthemes: trust in safeguards and researchers, altruistic motivation, and low perceived sensitivity of data.

Trust in Safeguards and Researchers

Participants described why they felt comfortable sharing their data, highlighting the importance of transparent communication in building trust. One patient noted:

> *...I wasn't worried, because...I knew from talking to yourself, it's going to be people who require that information…and are going to use it. To see if they can make improvements.* [Patient C]

One patient backed this up by saying:

> *As I say, you know, I mean…there's not enough trust in this world…if you're doing something like this…then you trust the people that are doing it are gonna be…with all the information.* [Patient D]

These sentiments underscore how personal interactions with researchers can directly influence participants' willingness to share sensitive data.

Participants accepted data sharing within the "safe" frame of health research in a University Hospital environment with project approval by the Ethics committee. One patient commented, "...obviously, you know, you've got fairly strong ethics...to consider..." (Patient E). One patient said, "I just thought….It's [name of university], it's a…public body, you got to trust people, don't you?" (Patient F). This participant's willingness to share data was rooted in the institution's reputation.

These findings suggest that clear communication regarding clinical trial governance, such as data security and institutional trustworthiness, may be critical for participant recruitment and retention in clinical trials.

Altruistic Motivation

Participants' belief that their data could help others promoted their willingness to share. Participants expressed a variety of altruistic motivations for sharing their health data (Table 3). The responses suggest that participants viewed their data sharing as a way to directly support others facing similar health challenges, demonstrating a strong sense of empathy and desire to make a positive impact. Participants appeared to see their involvement as an opportunity to advance medical research and improve therapeutic options, reflecting a commitment to the broader scientific community. In addition, participants viewed their data sharing as a way to create a legacy of improved care for others, reflecting a strong sense of social responsibility and a desire to contribute to the well-being of future generations.

XSL•FO
RenderX

Table . Participant altruistic motivations for sharing health data.

Altruistic motivation	Illustrative participant quotations
Helping others directly	• Patient F: "...with the watch..., if it's going to help somebody else [will] wear it.…If it helps collect data, to find,…trends and stuff like that, then that's good, isn't it?" • Patient G: "Oh I have no problem with that, Oh, God, no not a problem at all. If it helps others. If it helps with, you know what I'm going through, and probably will be going through worse, I suppose. If it helps others I have no problem at all. No problem. No. No why, why would it bother me at all? No. As I say, if it helps someone else, it's a great thing." • Patient C: "Didn't bother me…it doesn't, as I say, it's not something that I get excited about. I mean I have a progressive illness.…If I can help, by doing, we'll say, this, then, take as much information as you require. Do you know, that's my attitude…"
Contributing to research and advancing knowledge	• Patient H: "Oh, fine! If it could be of benefit, that's great! I mean.…Because I find it convenient, and I like it.…And mainly what I like, too, is the idea that you are...you're doing this study." • Patient E: "...and if it'll help you know, to develop, you know, some additional therapies or monitoring with them. That's a very good thing. So no, I have no issue with it at all." • Patient J: "I'm glad that it is being collected,…If we can help, alleviate problems, for other people isn't that…only good. If it contributes to knowledge about this disease. I'm in favour of that. Yes." • Patient I: "I would say that they definitely should consider. Joining it, because really, like, interesting.…Because you could help, like, them to see what's actually going on and make, you know, new medicine or something. To…to help people with the same condition."
Supporting future generations	• Patient D: "Well, no, I mean, it's…if the research…leaves…a better understanding.…And things like that. And it could help future generations, like…" • Patient K: "I'd say you're giving information now that in the future, people could use, hopefully. For the betterment of life, I suppose. For other people's health. I'd say, we're here today, and we get all our medicines through other people doing what we're doing. But they did it years ago. And…they've improved our lives. So why shouldn't we pay back and improve? Maybe our grandchildren's life, or whatever. With all this data, all this information, and they correlate it all together. You know well, it's amazing, what you folk can do with it. I haven't a clue, but, you know."

Overall, the range of altruistic motivations expressed by participants, whether helping others directly, advancing research, or supporting future generations, highlights an overarching willingness to share health data for the greater good. These authentic participant voices reinforce the importance of recognizing and honoring these motivations in the design and communication of health research studies.

Low Risk

Participants described the health data collected by digital devices (such as steps or heart rate) as nonsensitive and therefore were not concerned about who might access it. Comparisons to bank details were common: for instance, one participant who felt largely unconcerned about sharing this kind of information explained, "I can't see what…anybody could gain out of it, like, it's not like you're giving them your bank details and things like that." This sentiment was given by another participant, who clarified, "And again, if I mean, it's not information. That's you know. It's not my bank account details." Such comparisons to the sensitivity of banking information emphasize the distinction between health data and sharing perceived high-risk financial details.

For these individuals, the relatively harmless nature of physical activity or biometric data reduced perceived risks; as one participant put it, "It's useless," suggesting that the information collected was not "usable by somebody else." Another participant described their cautious approach to financial data, stating:

> *No....The less people to have access to my bank details, the better. I mean, I'll get some apps that come up, and it says, oh, like, I want you to pay this or pay that, and I just…I just…delete them straight away, like...*

These views illustrate how perceptions of sensitivity are shaped by the type of data in question.

However, one participant admitted some hypothetical unease:

> *I'm not sure…what I would think…if I heard somebody was hacking my devices and taking my*

XSL•FO
RenderX

health data. Because…I'm not sure if there's anything personal in what the watch is collecting. Sorry, sorry. When I say personal, I mean usable by somebody else.

This response highlights the variability in how participants assess the sensitivity of their data and that, for some, concerns about unauthorized access linger even when the data seem relatively harmless.

This response may reflect varying levels of digital literacy or previous experiences with data breaches, suggesting that educational interventions could influence perceived sensitivity. For many, a low perceived sensitivity facilitated trust and willingness to share, while for one, residual concerns persisted. These findings indicate that researchers must remain attentive to the different perceptions of data sensitivity and tailor engagement strategies accordingly.

Taken together, these findings suggest that participant engagement with digital health trials can be supported by building interpersonal trust, transparent ethical safeguards, and clearly framing the perceived value and safety of the data being collected. Researchers should consider how trial communications, onboarding, and technology framing can promote a sense of contribution without harm.

Theme 2: Navigating Digital Tools: Friction and Flexibility

Reflexive Thematic Concept

Participants' experiences with wearable technology and its accompanying app were strongly shaped by individual attitudes, digital literacy, device usability, and workaround strategies. While some approached the technology with confidence and curiosity, "I suppose I tried it because I love gadgets…I was never not going to have a smart watch," others encountered hesitancy and anxiety when facing practical challenges such as charging, syncing, or app lag. The type of smartphone (Android vs Apple) and the need to use unfamiliar or multiple devices created additional friction. One patient shared:

I haven't used apple things before, so…to be honest, I had…I can hardly make head or tail of the…iPhone, to be honest with you. [Patient J]

Technological Literacy

Literacy played a key role in shaping participants' engagement with the digital tools. Individuals who were more comfortable with digital devices generally described the technology as intuitive, unobtrusive, and even enjoyable to use. One patient described the watch as being, "a brilliant piece of technology,…I found it so helpful" (Patient C). For these participants, some cited age (Patient L) as an advantage with fewer barriers. Others described overcoming initial hesitations or lack of interest (Patient I).

In contrast, participants with lower technological confidence described themselves as "technophobes" and often relied on family or informal caregivers for setup and troubleshooting. As one patient described, "my daughter set it up for me…knew what the phone was all about. So, yeah, I got used to it pretty quick" (Patient K). Some participants acknowledged a shift in attitude as the perceived usefulness of the tools became clearer, suggesting that skepticism was not fixed but context-dependent (Patient J).

The varying levels of technological confidence, experiences, and attitudes (Textbox 1) among participants highlight the need for tailored onboarding processes and ongoing support.

Textbox 1. Participant experiences with technology.

Positive experiences

- Patient L: "To be honest, it was fine for me. I suppose maybe my age helped that. If I'd been older, obviously, it would have been more of a struggle."
- Patient E: "I'm perfectly happy with the technology. You know, to me, it's….As technology goes, it's very easy."
- Patient I: "I just never really think about it at all, or had the idea to actually use it. But now, since I started using it, I find it really more interesting. How it can track everything."
- Patient C: "Give it a chance. Use it, because you will find it useful. The technology isn't scary, and it works. And I found it incredibly helpful in my day-to-day activities."

Challenges and scepticism

- Patient F: "I was always a bit of a technophobe. I tended to think they were a bit of a...a bit some of the things a bit gimmicky, some of the stuff."
- Patient K: "It was awkward for me because, as I said, I'm totally illiterate when it comes to computers and whatever."
- Patient M: "I am one of these people that, when I got my mobile phone, I got it for me to use. Not for other people to contact me."
- Patient J: "I'm not a big fan of technology. I'm not really a fan of it at all, like,….Although I'm very used to the Apple Watch now. I'm a bit of a troglodyte."

Usability Friction

Everyday issues such as remembering to wear the device after showering, keeping it charged, app lag, syncing problems, and confusion over duplicate user profiles frequently disrupted initial adherence. One patient noted:

The only time I don't wear it is when I take a shower...occasionally I forget to put it back on and again I would defend myself with my age, my issues, and being woolly headed sometimes. [Patient E]

One patient commented:

XSL•FO
RenderX

> *my biggest problem was charging when I was having a shower and remembering to put it back on…until you got used to it, you'd forget the odd day to put it back on.* [Patient F]

One patient appreciated aspects such as device comfort and battery life:

> *It's very comfy, very usable…the battery lasts for ages. It's good, I was a bit worried about it, but it lasts a day or two.* [Patient F]

Technical issues with the app were also a concern; some struggled with questionnaires and app responsiveness, as one patient described:

> *I do find it a bit laggy at times...you do the first one, and then if you're kind of too quick doing the second one, it doesn't take, and you have to go back in and do it.* [Patient L]

One patient described problems with confusion over user profiles:

> *Obviously, I think some days I've forgotten to do the questionnaires, although there's been a couple of times when that's happened, where I'm nearly convinced that I did answer the questionnaires, and then I'm wondering, did I somehow use the wrong profile.* [Patient E]

Participants expressed the need for more flexibility in device and platform choices to better fit their routines and preferences. One patient described:

> *An Android user, like me, who kept two separate phones, wasn't integrating it into their lives properly.…To use them properly, you have to be able to integrate them into your life, and they become not just a data collector...* [Patient M]

These technical challenges highlight the importance of user-centered, flexible digital health designs.

Workarounds

To overcome initial technological barriers, participants used social support and developed routines. One patient gets family support:

> *I would need some assistance, yes, because I'm not...I'm not too good on this technology....Yes well [my son] goes over things with me there that are on....he comes out every week....He looks over things, and tells me about this or that....He's well up on all of that.* [Patient B]

More confident users like Patient E independently sought solutions:

> *so like I say, it just took me a while to you know, to work around the format of the watch, because again, that was different to the watch I'd had before. But I mean, it's, you know, it's a graphical interface. It's not that difficult to you know, to work through. And you can always Google it if you need to. So yeah, no, I don't think I had too many issues with that.*

One patient described developing routines to efficiently complete required tasks:

> *Because it tends to be just the two questions, the breathing and the coughing. You know, it only takes 30 seconds to do it. So I tend, as soon as I get the message come through, I'll tend to do it straight away. But the, the one where it could be, sort of, 19 questions, or whatever. You know, I tend to sort of sit down for an hour, and then go through them properly, and...try and be as accurate as I can. No, it's fairly straightforward, and, you know, there's nothing complicated about it.* [Patient D]

One patient took a pragmatic approach to device notifications:

> *...when I'm on my bike. It kind of pings the watch. And it says something like. I see you're working out. Or your cycling, or something like that. And, uh, I tend to ignore it.…I just click it back to what it was.…I'm just…wearing it as a watch.* [Patient F]

In terms of charging routines, one patient describes:

> *my biggest problem was charging when I was having a shower and remembering to put it back on,….[Now I]...put it back on straight away. You know, as soon as you come out and dry your hair.* [Patient L]

These varied adaptation strategies demonstrated resilience and resourcefulness, regardless of initial technological proficiency.

Workarounds, whether social (family assistance) or individual (routine-building), illustrate the importance of both personal and environmental resources in successful adoption. Informal caregiver involvement becomes an unplanned implementation mechanism; trials may need to recognize and resource this hidden support network.

Feedback on app reminders was mixed: some found them helpful for adherence, others considered them intrusive (on bank holidays), highlighting the importance of customizable engagement strategies.

Designing trials that acknowledge the desire for more flexibility in device and platform choices highlights the importance of user-centered design. Allowing participants to use their own devices (Bring Your Own Devices [BYOD]) and offering customizable engagement strategies can improve trial inclusivity and reduce the burden on participants.

Theme 3: Participant-Initiated Engagement With Wearable Data

Reflexive Thematic Concept

Although the trial design intended passive data collection, participants frequently engaged actively with their wearable data. For many, the device became embedded into daily routines, supporting self-monitoring, emotional reassurance, and in some cases behavior change, with the majority (14/15) expressing intention to continue using the wearable. Participant-initiated engagement revealed both empowering and distressing impacts, underscoring the personal meanings participants assigned to digital data.

XSL•FO
RenderX

From Passive Monitoring to Personal Engagement

Passive data collection can become active self-monitoring, altering behavior or anxiety levels. The engagement with wearable data spanned a spectrum from enthusiastic daily use to more passive or intermittent interaction, with most (11/15) individuals noting specific health benefits and increased self-efficacy.

For some, the device remained a passive tool, with limited personal interaction, as one patient reflected:

> *I'd just go back to my old watch, so…I'm sorry, yeah, it's not…I should be gushing about it, but I'm not…,* [Patient F]

This illustrates that when data collection is entirely passive and lacks user interaction, engagement may remain minimal.

In contrast, others found significant reassurance and utility in the ability to retrospectively review their health data. One patient shared:

> *...I find it incredibly useful, and it helps me, and it also helps the fact that…the readings are stored on the iPhone. And you can go back and look through them. Plus, it averages out your results over a week, a month, 6 months. And you can start to see if you're…going up, down, or you're remaining stable? So, from that point of view, it's absolutely super.* [Patient C]

This highlights how long-term data tracking and the ability to observe trends can provide participants with a sense of control and reassurance.

For some, the wearable device offered emotional support and comfort. One expressed, "I use it I feel it…it is a comfort to have it," suggesting that the device became more than a data collector—it provided a sense of security (Patient B). One patient went further, describing the device as integral to her sense of self: "If I don't wear it, is it my….It's…it's my…myself is not complete" (Patient N). This sentiment underscores how, for certain individuals, the wearable became an extension of their identity.

Curiosity and discovery also motivated ongoing engagement for some participants. One patient remarked:

> *I find it, like, really interesting, because I never really thought of. How much an apple watch can actually do in the health app? And I just find it really interesting.* [Patient I]

This sense of discovery fostered a positive attitude towards the technology and encouraged regular use.

Others described how the device became embedded in their daily routines, supporting proactive self-monitoring and health management. One patient explained:

> *I would generally look at it in the morning for the oxygen overnight, and then at night, I would check everything, so I would check heart rate for the day. Oxygen saturation, steps walked,…Oh, in the morning, I would check…sleep and oxygen saturation,...then in the evening, everything.* [Patient L]

This routine use demonstrates how wearables can facilitate regular health checks and empower participants to take an active role in managing their condition.

Collectively, these narratives illustrate the diverse ways in which participants engaged with wearable technology—ranging from limited interaction to deep integration into daily life. The emotional and practical significance of these devices varied, but for many, they provided reassurance, comfort, curiosity, and a means to monitor and manage their health more effectively. These insights highlight the importance of considering individual preferences and experiences when designing digital health interventions for people living with chronic conditions.

Responses to Data

Participants described both comfort and anxiety in response to the data. For some, it provided reassurance and control; for others, it increased worry, especially when data confirmed decline. Some participants changed behaviors based on data while others did not.

Participants described how wearable device data influenced their understanding and management of their health. For example, one patient reflected:

> *It's great, because you notice, right, I'm not imagining it. It is there, and it's something, like I was able to say. Like, weirdly enough. Whatever the hell it was. Tuesdays were terrible for AFib. All my episodes have either been on a Monday night, or Tuesday morning.* [Patient C]

This demonstrates how device data enabled validation of symptoms and identification of patterns, enhancing self-understanding.

One patient shared:

> *Yes, it tells me how many hours I slept. Which was a good thing, because then I could justify why I was tired.* [Patient K]

Here, the objective data supported self-justification and explanation of symptoms, reinforcing the device's utility in daily life.

One patient explained:

> *I can sit down and relax and watch it and see if the heart rate starts to go back down. Then I get settled, more settled then you know.* [Patient G]

This illustrates how real-time feedback from the device was used to guide self-regulation and coping strategies.

One patient noted:

> *Yeah. I tried to look at it almost, like, every single day. It's kind of like a routine I like to see if it's, like, normal, high, or low.* [Patient I]

For Patient I and others, daily engagement with wearable metrics became a habitual part of life.

XSL•FO
RenderX

However, not all experiences were reassuring. One patient said:

> *It definitely didn't reassure me, and it made me the opposite, because I could...I knew I was...I knew I was starting to struggle more, and I could see, probably, it was progressing, and I suppose the stats show that, you know, when I, when I checked them, I could see the decline and everything.* [Patient L]

For some, data indicating deterioration contributed to increased anxiety, showing the nuanced emotional impact of continuous tracking.

A patient described:

> *Because I have got used to it now...I'm watching the…you know, steps and that. Like, the Fitbit aspect of it. Again, it's not something I'd be obsessive about, but it's just interesting to see.* [Patient J]

This suggests that integrated tracking features can foster ongoing, non-intrusive engagement without leading to obsession.

Finally, one patient commented:

> *I don't have any little goals or anything, I just...observe what's happening with it, and that's….Okay, no, it's convenient, and….Yes, I realize that it's…that it's there for the statistical records, and that you're interested in the nighttime heart rates and all these things. Of course, I'm interested in it, too. I have a personal interest in it.* [Patient H]

This reflects how some participants engaged with data in a more observational, less goal-oriented manner, highlighting the diversity of usage styles.

The wearable device transitioned from a passive study requirement to a routine aspect of daily life. This integration facilitated regular health checks, promoted new routines, and, for some, prompted behavior change, an unintended consequence. This illustrates how wearables were embedded into morning and evening routines, supporting self-monitoring and proactive health management. Continuous monitoring can have both reassuring and anxiety-inducing effects on participants. Providing clear communication about the purpose and benefits of data collection, along with emotional support resources, can help mitigate negative emotional impacts.

Theme 4: Life-Limiting Illness as a Constraint on Digital Trial Participation

Reflexive Thematic Concept

Participants' engagement with digital tools was shaped by the length of time they have had the diagnosis and ongoing demands of managing chronic and progressive illness. Coping with fatigue, breathlessness, medication routines, and oxygen setup already demands much of patients before introducing additional digital devices.

Disease Duration Might Foster Routine Management or Digital Fatigue

Participants reported a wide range of diagnostic journeys (Multimedia Appendix 5), from recent confirmations to decades-long histories of illness. Some described protracted periods of misdiagnosis or uncertainty, while others experienced abrupt transitions from health to chronic disease management.

Longer disease duration may foster routine management strategies but could also lead to "digital fatigue," a diminished motivation to engage with new technologies due to the cumulative demands of ongoing self-monitoring. Conversely, those with recent diagnoses sometimes reported uncertainty and a steeper learning curve in adopting digital tools. Notably, a dropout from the study who did not consent to be interviewed had been diagnosed a year previously, citing anxiety with using digital tools. In both cases, the context of diagnosis duration shaped the readiness and capacity for digital trial participation.

Symptom Burden May Shape Willingness or Capacity to Use Technology

Participants described how illness-imposed restrictions fundamentally altered their daily routines, social activities, and independence:

> *If I want to do anything, I've got to think oxygen. If I don't, I'll just get caught short. And I'm like a fish out of water, I'm floundering.*

Reliance on supplemental oxygen was a defining feature for many, shaping both physical mobility and daily planning: "I wake up in the morning, and before I go to the toilet or do anything, then the oxygen has to go on...." The logistics of managing oxygen supplies and equipment often required advance planning and adaptation, leaving less flexibility for unscheduled or spontaneous engagement with digital trial tools. One participant had so much to take with them for a trip they didn't bring the study phone:

> *I went up to [the city] Saturday, stayed overnight, came back yesterday evening. I didn't bring my phone with me, you know? So that's the only thing…be inclined to forget…the phone.*

Another spoke about having enough to do when faced with throat cancer to focus on wearing the watch and charging it, "No, I was more concerned about the cancer in my throat. I didn't think of the watch. You know?" This participant's watch wear time completely stopped when faced with cancer.

Cough varied in frequency and severity, with some experiencing persistent symptoms and others only occasional bouts, while for others it seemed irrelevant:

> *…I'm keeping in mind what I've been putting in on the…on the MyCap as well, and I know that. Most of the time, it's been…pretty stable, like….The last two weeks now...my…cough has been getting a little bit worse, because today now, when I'm coughing. It's actually painful in my chest. Now, I don't know whether I have picked up an infection or something. But, as I say, I'll sort of wait and see.*
>
> *All the questionnaires that you've been giving me about coughing, I never understood them….But since about the 2nd of [Month] I am now getting coughing fits....*

XSL•FO
RenderX

> *I get two…questionnaires a week. And one of them is about my cough…I don't really have a cough…doesn't seem that relevant to me*

Fatigue was a prominent and often unpredictable barrier, affecting both physical stamina and cognitive function: "I have a lot of fatigue….I'm sometimes woolly headed," "There's just no energy. I've no energy at all," "I get tired very quickly…it varies from day to day, I mean, some mornings I get up and I feel fine....But then there are other days I wake up, and I've just got nothing in the tank." Fatigue can be a cause of delayed interactions with digital trial tools. The first dropout from the study who declined to be interviewed had been vague in responses to follow-up consistently being late to fill out questionnaires. Perhaps this was due to mental fatigue affecting their focus or articulation. The cognitive effects of fatigue make complex or multistep digital tasks more challenging, underscoring the need for accessible, low-burden digital designs in future trials. These limitations often led to prioritization of essential tasks, with discretionary activities, including digital trial engagement becoming secondary. For some, the cognitive and logistical effort required to interact with apps or devices was an additional burden layered atop daily challenges.

Participants' experiences of life-limiting illness fundamentally shape their engagement with digital technology study tools. Core symptoms—fatigue, breathlessness, cough—and the practical demands of managing comorbidities and oxygen create capacity and motivation challenges. For example, the participant with throat cancer vividly shows how comorbidities directly interrupted study tasks. Understanding this context is essential for designing digital interventions that are feasible, empathetic, and responsive to the lived experiences of people with chronic, life-limiting conditions. To address these challenges, digital interventions should allow flexible participation schedules and low-effort digital interfaces, ensuring that technology adapts to the needs and capacities of participants rather than imposing additional burdens.

Discussion

Principal Findings

Participants described core themes influencing engagement with digital health technologies, including trust, digital literacy, emotional responses to data, and illness-related constraints. This qualitative substudy provides an in-depth examination of how individuals living with fibrotic ILD experience sustained participation in a digitally enabled clinical research program. While digital monitoring is increasingly promoted as a means to enhance data granularity and reduce reliance on clinic-based assessments, our findings demonstrate that engagement with such technologies is not passive or uniform. Instead, participation is actively shaped by trust, digital literacy, emotional responses to data, and the constraints imposed by living with a life-limiting illness.

Continuous digital monitoring has the potential to increase granularity of outcome measurement; however, it may also introduce measurement reactivity, whereby participants alter behavior or reporting in response to being observed [19]. In the context of this study, participants had continuous access to wearable-derived data via the Apple Health app (Apple Inc), enabling interaction with trial data beyond the boundaries of the study protocol. While prior work has focused on feasibility and validation of digital end points, our findings highlight how access to consumer platforms such as Apple Health enables participant-initiated use of trial data, introducing behavioral adaptation and potential cointervention [20-22]. These processes may influence activity patterns, symptom reporting, and disease perception, with implications for the validity and interpretation of trial outcomes. Recognizing and characterizing these mechanisms is essential for the design and interpretation of future digital and hybrid trials [23].

Participants interviewed between 8 and 14 months into the planned 3-year study found digital monitoring acceptable, which aligns with emerging evidence showing high feasibility of home monitoring in ILD and other chronic respiratory diseases [24,25]. High PROM adherence (>99%), alongside high smartwatch wear times (median 96%, IQR 95.9%-99.5) in the first 12 months of PRODIGY-ILD, reinforces the feasibility of sustained digital data collection, an essential requirement for longitudinal outcomes. For clinical trials, this suggests that integrating wearables and regular ePROMs is likely to be acceptable to most participants, supporting their inclusion as trial end points or exploratory biomarkers. However, the context provided by qualitative interviews also highlights that certain factors such as comorbidities can limit willingness or capacity to engage with digital tools over time. For some participants, illness-related challenges led to reduced engagement or prioritization of digital trial activities, underscoring the importance of designing flexible, low-burden interventions that adapt to fluctuating participant capacity and lived experience. This suggests that while overall feasibility is high, successful implementation requires ongoing attention to accessibility, participant support, and the minimization of additional burdens.

The qualitative analysis of participant experiences in the PRODIGY-ILD study reveals a number of implications for the design and conduct of clinical trials that incorporate digital monitoring tools. Our 4 themes: trust and altruism, navigating digital tools, adoption of technology, and the impact of life-limiting illness offer insights into how clinical trials can be refined to improve recruitment, adherence, data quality, and participant well-being.

Participants' willingness to share personal health data was significantly shaped by their trust in research teams, confidence in ethical safeguards, and a strong sense of altruism. Transparent communication about data governance and the institutional reputation of research settings were central to fostering trust. For clinical trials, this highlights the need for clear onboarding processes that articulate data security, privacy protections, and the ethical framework underpinning the study [26,27]. Recognizing and explicitly honoring participants' altruistic motivations, such as helping others, contributing to research, and supporting future generations, can further strengthen engagement and retention [28,29]. Trial communications should frame participation as a meaningful contribution to the advancement of care, while addressing any residual privacy concerns through ongoing dialogue [30,31].

XSL•FO
RenderX

Participant experiences with digital tools varied widely according to technological literacy, device compatibility, and everyday usability. Those less familiar with technology relied on informal support from family or caregivers, consistent with prior findings that digital literacy strongly shapes engagement and retention in remote studies [32]. Usability friction, such as charging, syncing issues, and app lag, caused frustration and has similarly been recognized as a key barrier to sustained adherence in digital health interventions [33-35]. To mitigate these barriers, clinical trials should incorporate stratified onboarding and support mechanisms tailored to different levels of digital literacy [33-35]. Flexible approaches, such as allowing participants to use their own devices and providing customizable engagement strategies, can reduce burden and improve inclusivity [36]. Recognizing the hidden labor of informal caregivers and resourcing this support where possible may be crucial for sustained adherence, especially among older or less tech-savvy populations [32,35].

The study revealed that participants often engaged actively with their wearable data, leading to both empowering and distressing impacts. While many found reassurance and motivation through self-monitoring, others experienced heightened anxiety when confronted with evidence of disease progression. These findings highlight the importance of developing communication strategies that provide support and emotional resources to help participants understand their data. Trials should carefully consider how much real-time data visibility is appropriate, balancing the benefits of participant engagement against the risk of unintended psychological harm. Structured feedback mechanisms and clear explanations of the purpose and limitations of digital end points should be used to safeguard participant well-being and trial integrity [20,25].

The demands of managing progressive illness, fatigue, breathlessness, medication regimens, and comorbidities significantly shaped participants' capacity and motivation to engage with digital trial tools [37-39]. For participants with severe comorbidities, the burden of treatment can mean trial activities are deprioritized [37,39]. To accommodate these realities, clinical trials must design digital interventions that are flexible, low-burden, and adaptive to fluctuating participant capacity [22,26]. Features such as adjustable participation schedules, simple interfaces, and options to pause or modify engagement can make trials more feasible and empathetic. Understanding the lived experience of chronic illness should inform all stages of digital trial development, from feasibility testing to ongoing support [16].

Limitations

There are limitations that should be considered when interpreting the findings of this study. First, the sample size was small (n=15), reflecting the exploratory nature of the work and the intensive, interview-based methodology. While this enabled rich, in-depth accounts, the perspectives captured may not represent the broader population of individuals living with fibrotic ILD. Two participants who dropped out of the study did not consent to being interviewed, which may bias findings toward acceptability of technology use. Participants were also drawn from a single specialist ILD service, which may limit transferability to settings with different patient demographics, service structures, or technology access. Second, several participants relied on informal caregiver support to manage digital trial tasks, such as device setup, troubleshooting, and data entry, which may not be replicable for individuals without such assistance. The role of caregivers, although integral for some participants' engagement, introduces variability that may influence the feasibility and acceptability of digital tools in more socially or digitally isolated groups. Elements of response bias may have been introduced, as participants could have been influenced by their existing relationship with the researcher, potentially leading them to provide responses they perceived as desirable or expected. Finally, the heterogeneous clinical course of fibrotic ILD, many of which are progressive and may be life-limiting, may have influenced participants' engagement with the study. Participants described a strong desire to understand their condition, reflecting the scarcity of effective treatments and the uncertainty surrounding disease trajectory. This sense of urgency and investment may have enhanced adherence to digital tools and PROMs beyond what might be expected in conditions with a more stable or treatable course, limiting generalizability to other patient populations.

Conclusion

Taken together, these findings extend existing literature on the feasibility and acceptability of digital monitoring in ILD by demonstrating that sustained engagement is a dynamic, emotionally mediated process embedded within the broader context of illness, identity, and daily life. While overall adherence in PRODIGY-ILD was high, qualitative insights reveal the effort, adaptation, and negotiation that underpin this success. Embedding qualitative research within digital trials provides critical insight into how technologies are actually used, interpreted, and experienced and is essential for understanding the mechanisms through which digital end points are generated.

In conclusion, successful integration of digital technologies into clinical trials requires more than technical feasibility. It demands a participant-centered approach that builds trust, acknowledges altruistic motivations, accommodates variability in digital literacy, recognizes the role of informal caregivers, and addresses the emotional and practical realities of living with chronic, life-limiting illness. By illuminating these processes, this study contributes to the methodological foundations of digital health–enabled research and offers actionable insights to inform the design, conduct, and interpretation of future digital and hybrid clinical trials.

Acknowledgments

The authors gratefully acknowledge the support of the Health Research Board – Trials Methodology Research Network (HRB-TMRN), Galway, Ireland, and the UCD Clinical Research Centre for their guidance and support throughout this study. We thank Lindsay Brown, advanced nurse practitioner, for her valuable advice on patient-reported outcome measure (PROM)

XSL•FO
RenderX

selection and her ongoing clinical input during the study. We also acknowledge Dylan Keegan for his contribution to the development of the Participant Information Leaflet and Informed Consent Form. We are grateful to the Irish Lung Fibrosis Association Patient and Public Involvement (PPI) contributors, Nicola Cassidy, Paula Jenkins, and Robert Hurley, whose feedback informed the design of the study website. Finally, we thank Patrick McKay of Advantage Point Creative Consultancy for the design and development of the study website.

Funding

Grant support for this research study was provided by the Health Research Board – Trials Methodology Research Network (HRB-TMRN). The funding body had no role in the design of the study protocol and was not involved in the collection, analysis, and interpretation of data or manuscript preparation. UCD Clinical Research Centre provided digital device funding.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the sensitive nature of qualitative interview data and the potential risk of participant reidentification in a rare disease population, but they are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: SH, PD, EG
Data curation: EG, SH
Formal analysis: EG, SH
Funding acquisition: PD, SH
Investigation: EG
Methodology: EG, SH
Project administration: EG, SH, PD
Resources: SH, PD, CMC, MPK
Supervision: SH, PD, CMC
Validation: SH, PD
Visualization: EG, SH
Writing – original draft: EG, SH
Writing – review & editing: EG, SH, PD, CMC, MPK, AF

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview script.
[DOCX File, 20 KB - humanfactors_v13i1e91533_app1.docx]

Multimedia Appendix 2
Consent for qualitative substudy.
[DOCX File, 17 KB - humanfactors_v13i1e91533_app2.docx]

Multimedia Appendix 3
NVivo parent child codes for each theme.
[DOCX File, 20 KB - humanfactors_v13i1e91533_app3.docx]

Multimedia Appendix 4
Smartwatch weartime.
[DOCX File, 14 KB - humanfactors_v13i1e91533_app4.docx]

Multimedia Appendix 5
Participant reported disease duration.
[DOCX File, 15 KB - humanfactors_v13i1e91533_app5.docx]

Checklist 1
COREQ checklist.

XSL•FO
RenderX

[PDF File, 114 KB - humanfactors_v13i1e91533_app6.pdf]

References

1. Raghu G, Remy-Jardin M, Richeldi L, et al. Idiopathic pulmonary fibrosis (an update) and progressive pulmonary fibrosis in adults: an official ATS/ERS/JRS/ALAT clinical practice guideline. Am J Respir Crit Care Med 2022 May 1;205(9):e18-e47. [doi: 10.1164/rccm.202202-0399ST] [Medline: 35486072]
2. Althobiani M, Alqahtani JS, Hurst JR, Russell AM, Porter J. Telehealth for patients with interstitial lung diseases (ILD): results of an international survey of clinicians. BMJ Open Respir Res 2021 Dec;8(1):e001088. [doi: 10.1136/bmjresp-2021-001088] [Medline: 34969772]
3. Althobiani MA, Evans RA, Alqahtani JS, et al. Home monitoring of physiology and symptoms to detect interstitial lung disease exacerbations and progression: a systematic review. ERJ Open Res 2021 Oct;7(4):00441-2021. [doi: 10.1183/23120541.00441-2021] [Medline: 34938799]
4. Moor CC, Mostard RLM, Grutters JC, et al. Home monitoring in patients with idiopathic pulmonary fibrosis. A randomized controlled trial. Am J Respir Crit Care Med 2020 Aug 1;202(3):393-401. [doi: 10.1164/rccm.202002-0328OC] [Medline: 32325005]
5. Noth I, Cottin V, Chaudhuri N, et al. Home spirometry in patients with idiopathic pulmonary fibrosis: data from the INMARK trial. Eur Respir J 2021 Jul;58(1):2001518. [doi: 10.1183/13993003.01518-2020] [Medline: 33419890]
6. Wijsenbeek MS, Bendstrup E, Valenzuela C, et al. Disease behaviour during the peri-diagnostic period in patients with suspected interstitial lung disease: the STARLINER study. Adv Ther 2021 Jul;38(7):4040-4056. [doi: 10.1007/s12325-021-01790-y] [Medline: 34117601]
7. Wuyts WA, Wijsenbeek M, Bondue B, et al. Idiopathic pulmonary fibrosis: best practice in monitoring and managing a relentless fibrotic disease. Respiration 2020;99(1):73-82. [doi: 10.1159/000504763] [Medline: 31830755]
8. Maher TM, Schiffman C, Kreuter M, et al. A review of the challenges, learnings and future directions of home handheld spirometry in interstitial lung disease. Respir Res 2022 Nov 11;23(1):307. [doi: 10.1186/s12931-022-02221-4] [Medline: 36369156]
9. Wijsenbeek MS, Moor CC, Johannson KA, et al. Home monitoring in interstitial lung diseases. Lancet Respir Med 2023 Jan;11(1):97-110. [doi: 10.1016/S2213-2600(22)00228-4] [Medline: 36206780]
10. Apple Heart Study: assessment of wristwatch-based photoplethysmography to identify cardiac arrhythmias. ClinicalTrials.gov. URL: https://clinicaltrials.gov/study/NCT03335800 [accessed 2026-07-23]
11. Althobiani MA, Shuttleworth R, Conway J, et al. Supporting self-management for patients with interstitial lung diseases: utility and acceptability of digital devices. PLOS Digit Health 2024 Jan;3(1):e0000318. [doi: 10.1371/journal.pdig.0000318] [Medline: 38190384]
12. Gunne E, Holden S, Franciosi AN, Keane M, McCarthy C, Doran P. Predicting outcomes using DIGital TechnologY in patients with Interstitial Lung Disease (PRODIGY-ILD): protocol for a prospective cohort study. BMJ Open 2025 Apr 28;15(4):e088271. [doi: 10.1136/bmjopen-2024-088271] [Medline: 40295125]
13. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
14. Harris PA, Swafford J, Serdoz ES, et al. MyCap: a flexible and configurable platform for mobilizing the participant voice. JAMIA Open 2022 Jul;5(2):ooac047. [doi: 10.1093/jamiaopen/ooac047] [Medline: 35673353]
15. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
16. Harrison Ginsberg K, Babbott K, Serlachius A. Exploring participants' experiences of digital health interventions with qualitative methods: guidance for researchers. J Med Internet Res 2024 Nov 28;26:e62761. [doi: 10.2196/62761] [Medline: 39607999]
17. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019;11(4):589-597. [doi: 10.1080/2159676X.2019.1628806]
18. Braun V, Clarke V. Toward good practice in thematic analysis: avoiding common problems and be(com)ing a knowing researcher. Int J Transgend Health 2023;24(1):1-6. [doi: 10.1080/26895269.2022.2129597] [Medline: 36713144]
19. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J Clin Epidemiol 2014 Mar;67(3):267-277. [doi: 10.1016/j.jclinepi.2013.08.015] [Medline: 24275499]
20. Digital health technologies for remote data acquisition in clinical investigations. US Food and Drug Administration. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/digital-health-technologies-remote-data-acquisition-clinical-investigations [accessed 2026-01-16]
21. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. NPJ Digit Med 2019;2(1):14. [doi: 10.1038/s41746-019-0090-4] [Medline: 30868107]
22. Manta C, Patrick-Lake B, Goldsack JC. Digital measures that matter to patients: a framework to guide the selection and development of digital measures of health. Digit Biomark 2020;4(3):69-77. [doi: 10.1159/000509725] [Medline: 33083687]

XSL•FO
RenderX

23. CTI considerations for advancing the use of digital technologies for data capture & improved clinical trials. : Clinical Trials Transformation Initiative; 2021 URL: https://ctti-clinicaltrials.org/wp-content/uploads/2021/06/CTTI_Digital_Health_Technologies_Recs.pdf [accessed 2026-07-21]
24. Moor CC, van Leuven SI, Wijsenbeek MS, Vonk MC. Feasibility of online home spirometry in systemic sclerosis-associated interstitial lung disease: a pilot study. Rheumatology (Oxford) 2021 May 14;60(5):2467-2471. [doi: 10.1093/rheumatology/keaa607] [Medline: 33212511]
25. Nakshbandi G, Moor CC, Johannson KA, Maher TM, Kreuter M, Wijsenbeek MS. Worldwide experiences and opinions of healthcare providers on eHealth for patients with interstitial lung diseases in the COVID-19 era. ERJ Open Res 2021 Jul;7(3):00405-2021. [doi: 10.1183/23120541.00405-2021] [Medline: 34471631]
26. Nebeker C, Harlow J, Espinoza Giacinto R, Orozco-Linares R, Bloss CS, Weibel N. Ethical and regulatory challenges of research using pervasive sensing and other emerging technologies: IRB perspectives. AJOB Empir Bioeth 2017;8(4):266-276. [doi: 10.1080/23294515.2017.1403980] [Medline: 29125425]
27. Sharon T. When digital health meets digital capitalism, how many common goods are at stake? Big Data Soc 2018 Jul;5(2). [doi: 10.1177/2053951718819032]
28. McCann SK, Campbell MK, Entwistle VA. Reasons for participating in randomised controlled trials: conditional altruism and considerations for self. Trials 2010 Mar 22;11:31. [doi: 10.1186/1745-6215-11-31] [Medline: 20307273]
29. Houghton C, Dowling M, Meskell P, et al. Factors that impact on recruitment to randomised trials in health care: a qualitative evidence synthesis. Cochrane Database Syst Rev 2020 Oct 7;10(10):MR000045. [doi: 10.1002/14651858.MR000045.pub2] [Medline: 33026107]
30. Rosa C, Marsch LA, Winstanley EL, Brunner M, Campbell ANC. Using digital technologies in clinical trials: current and future applications. Contemp Clin Trials 2021 Jan;100:106219. [doi: 10.1016/j.cct.2020.106219] [Medline: 33212293]
31. Bietz MJ, Bloss CS, Calvert S, et al. Opportunities and challenges in the use of personal health data for health research. J Am Med Inform Assoc 2016 Apr;23(e1):e42-e48. [doi: 10.1093/jamia/ocv118] [Medline: 26335984]
32. Jimenez J, del Rio A, Berman AN, Grande M. Personalizing digital health: adapting health technology systems to meet the needs of different older populations. Healthcare (Basel) 2023 Jul 26;11(15):2140. [doi: 10.3390/healthcare11152140] [Medline: 37570380]
33. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848. [doi: 10.2196/18848] [Medline: 32213476]
34. Whitelaw S, Pellegrini DM, Mamas MA, Cowie M, Van Spall HGC. Barriers and facilitators of the uptake of digital health technology in cardiovascular care: a systematic scoping review. Eur Heart J Digit Health 2021 Mar;2(1):62-74. [doi: 10.1093/ehjdh/ztab005] [Medline: 34048508]
35. Cajita MI, Hodgson NA, Lam KW, Yoo S, Han HR. Facilitators of and barriers to mHealth adoption in older adults with heart failure. Comput Inform Nurs 2018 Aug;36(8):376-382. [doi: 10.1097/CIN.0000000000000442] [Medline: 29742549]
36. Demanuele C, Lokker C, Jhaveri K, et al. Considerations for conducting bring your own “device” (BYOD) clinical studies. Digit Biomark 2022;6(2):47-60. [doi: 10.1159/000525080] [Medline: 35949223]
37. Daniore P, Nittas V, Haag C, Bernard J, Gonzenbach R, von Wyl V. From wearable sensor data to digital biomarker development: ten lessons learned and a framework proposal. NPJ Digit Med 2024 Jun 18;7(1):161. [doi: 10.1038/s41746-024-01151-3] [Medline: 38890529]
38. May CR, Eton DT, Boehmer K, et al. Rethinking the patient: using Burden of Treatment Theory to understand the changing dynamics of illness. BMC Health Serv Res 2014 Jun 26;14:281. [doi: 10.1186/1472-6963-14-281] [Medline: 24969758]
39. Eton DT, Ramalho de Oliveira D, Egginton JS, et al. Building a measurement framework of burden of treatment in complex patients with chronic conditions: a qualitative study. Patient Relat Outcome Meas 2012;3:39-49. [doi: 10.2147/PROM.S34681] [Medline: 23185121]

Abbreviations

BYOD: Bring Your Own Devices
COREQ: Consolidated Criteria for Reporting Qualitative Research
ePROM: electronic patient-reported outcome measures
ILD: interstitial lung disease
PRODIGY-ILD: Predicting Outcomes using Digital Technology in Interstitial Lung Disease
PROM: patient-reported outcome measure

XSL•FO
RenderX

Edited by K Adapa; submitted 16.Jan.2026; peer-reviewed by J Edwards, K Moor; revised version received 03.Jun.2026; accepted 20.Jun.2026; published 25.Aug.2026.

Please cite as:
Gunne E, Franciosi A, McCarthy C, Keane MP, Doran P, Holden S
Patient Experiences of Digital Technology Use in Interstitial Lung Disease (PRODIGY-ILD Study): Qualitative Study Using Reflexive Thematic Analysis
JMIR Hum Factors 2026;13:e91533
URL: https://humanfactors.jmir.org/2026/1/e91533
doi:10.2196/91533

© Emer Gunne, Alessandro Franciosi, Cormac McCarthy, Michael P Keane, Peter Doran, Sinead Holden. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 25.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluation of the Feasibility and Acceptability of Perfect Fit, a Virtual Coach–Based mHealth Intervention for Smoking Cessation and Physical Activity in Adults: Mixed Methods Study

Milon H M van Vliet[1,2], MSc, PhD; Eline Meijer[1,2], MSc, PhD; Nele Albers[3,4], MSc, PhD; Kristell M Penfornis[5], MSc, PhD; Walter Baccinelli[6], MSc, PhD; Bouke L Scheltinga[7], MSc, PhD; Roxy A van Eersel[1,2], MSc; Niels H Chavannes[1,2], MD, PhD; Anke Versluis[1,2], MSc, PhD; The Perfect Fit Consortium[8,9]

[1]Department of Public Health and Primary Care, Leiden University Medical Center, PO Box 9600, Leiden, The Netherlands

[2]National eHealth Living Lab, Leiden University Medical Center, Leiden, The Netherlands

[3]Department of Intelligent Systems, Delft University of Technology, Delft, The Netherlands

[4]Department of Communication and Cognition, Tilburg University, Tilburg, The Netherlands

[5]Department of Psychology, Unit Health, Medical, and Neuropsychology, Leiden University, Leiden, The Netherlands

[6]Netherlands eScience Center, Amsterdam, The Netherlands

[7]Biomedical Signals and Systems, University of Twente, Enschede, The Netherlands

[8]See Acknowledgments

[9]

Corresponding Author:
Milon H M van Vliet, MSc, PhD
Department of Public Health and Primary Care, Leiden University Medical Center, PO Box 9600, Leiden, The Netherlands

Abstract

Background: Mobile health (mHealth) interventions with virtual coaches offer scalable and potentially cost-effective solutions for health behavior change. However, these interventions commonly present challenges, such as limited personalization and insufficient grounding in evidence-based strategies. Perfect Fit (PF; Perfect Fit consortium), a personalized mHealth intervention with a text-based virtual coach, supports adults in quitting smoking and becoming more physically active. By combining innovative techniques, including sensor technology, end user involvement, and evidence-based strategies, PF aims to address common challenges faced by mHealth interventions, including those with virtual coaches.

Objective: The study primarily investigated the feasibility and acceptability of PF. The secondary aim was to explore associations between sociodemographic, smoking-, and physical activity–related characteristics and the feasibility and acceptability outcomes. The third aim was to evaluate the feasibility of conducting the research study.

Methods: A single-arm, pre-post, mixed methods study was conducted in the Netherlands with 100 adults who smoked. The intervention lasted approximately 16 weeks. Data were collected at baseline, during the intervention, and postintervention (4 months). Quantitative data included usage data and self-report questionnaires on feasibility, acceptability, and baseline characteristics. Qualitative data were gathered through postintervention semistructured interviews. Analyses included descriptive and inferential analyses, as well as the framework approach for the qualitative data.

Results: PF usage varied considerably across participants (n=87). The mean satisfaction rating was 2.79 (SD 0.73; scale range 1 - 4), and perceived usability had a median score of 67.50 (range 12.50 - 87.50; scoring range 0 - 100), indicating OK-to-good usability. The mean virtual coach acceptance rating was –0.27 (SD 1.30; scale range −3 to 3; n=77). Higher PF usage was associated with greater satisfaction, usability, and coach acceptance (all $P \leq .004$). Frequent connection issues with the smartwatch were a disruptive factor. Qualitative findings (n=12) provided in-depth insights into PF's feasibility and acceptability, encompassing both positive and negative experiences. For instance, some participants valued the virtual coach for its anonymity, low-threshold access, and the sense of control it offered, while others preferred a human coach for greater accountability. Suggested improvements included more varied content and enhanced adaptability of the coach to users' input and personal situations. Exploratory analyses suggested that high PF users were older than moderate ($P=.01$) and low PF users ($P=.05$). Importantly, PF was perceived as similarly feasible and acceptable across socioeconomic groups ($P>.05$), aligning with one of the project's goals. Finally, research procedures and recruitment strategies proved feasible.

Conclusions: PF shows potential as an accessible and inclusive strategy for multiple health behavior changes, contributing to public health. Findings highlight areas for improvement and can guide the future development of virtual coach interventions.

Trial Registration: ClinicalTrials.gov NCT06095999; https://clinicaltrials.gov/study/NCT06095999

XSL•FO
RenderX

International Registered Report Identifier (IRRID): RR2-10.1177/20552076241300020

(JMIR Hum Factors 2026;13:e83456) doi:10.2196/83456

KEYWORDS

Physical activity; smoking cessation; virtual coach; conversational agent; chatbot; mHealth intervention; feasibility; acceptability; mixed methods

Introduction

The development of mobile health (mHealth) interventions has increased noticeably over the past 30 years [1,2], offering promising solutions in terms of reach, accessibility [3], and scalability [2,4]. A systematic review of reviews has shown that mHealth interventions can support individuals in adopting healthier behaviors [5]. This is especially relevant for high-risk behaviors such as smoking and insufficient physical activity (PA), which are major contributors to chronic diseases and premature mortality [6-8] and the primary targets of the intervention evaluated in this study. In 2021, approximately 17% of the global adult population smoked [8], and in 2022, 31% did not meet PA recommendations. Moreover, unhealthy behaviors often co-occur [9,10], especially among individuals with lower literacy or in vulnerable positions. These individuals also tend to experience poorer health outcomes and reduced life expectancy [6,11-14]. Given the high prevalence and health burden of smoking and insufficient PA, these behaviors are key public health concerns [6].

Encouraging smoking cessation and PA promotion simultaneously may be particularly beneficial, as they can reinforce each other [15-17]. PA can help reduce nicotine cravings [15] and alleviate withdrawal symptoms [16], while quitting smoking improves fitness, potentially encouraging more PA [17]. However, sustained behavior change is difficult to achieve without effective support. Studies show that the likelihood of successfully quitting smoking is around 3 times higher with professional support compared to no support [13,18]. Similarly, PA interventions incorporating behavioral support elements (eg, feedback on performance) can effectively promote both the initiation and maintenance of PA [19]. This highlights the need for effective support to facilitate lasting behavior change.

Due to the ongoing health care crisis, professional support is increasingly limited. Health care demand exceeds supply [20], and mHealth interventions could help alleviate this burden by providing scalable [2,21], remote [4], and potentially cost-effective [5] smoking cessation and PA promotion support [5]. They can empower individuals to take greater control of their health, improve health outcomes, and potentially reduce health care use [22]. Additionally, mHealth interventions may lower barriers related to fear of stigmatizing interactions when seeking support, often experienced by people who smoke [21]. Finally, they enable just-in-time adaptive interventions, which adjust delivery based on the time of day and integrate into users' daily lives [4,23]. Hence, mHealth interventions are seen as a promising way to meet growing health care demands.

Although promising, mHealth interventions also have disadvantages. Compared to human support, they are often experienced as less personal and tailored [24] and can evoke a reduced sense of personal accountability [21], which can contribute to low user engagement and adherence [4]. Integrating virtual coaches into mHealth interventions may help overcome these limitations by simulating human support while maintaining digital delivery advantages. Virtual coaches are AI conversational agents that mimic human interactions through text, speech, or both [1,25,26]. They offer interactive, personalized support, which may enhance adherence and engagement compared to more static mHealth interventions [27]. Reviews of virtual coach interventions for smoking cessation [21,25] and PA promotion [1] reported positive outcomes regarding effectiveness, acceptability, and user experience. However, these findings should be interpreted with caution due to study heterogeneity and underpowered studies. While these findings highlight the potential of virtual coaches, further research is needed to address key challenges and knowledge gaps.

The literature on mHealth interventions with virtual coaches identifies key challenges, including a lack of theory-based development; limited and mixed evidence on long-term engagement (eg, due to nonstandardized measurement methods) [1,4,25,28]; and digital inclusion problems [2,20,21,24]. Several strategies have been proposed to address these challenges. First, it is crucial to base interventions on behavior change techniques [29], theories, or models [1,4,14]. Second, personalizing content and timing may increase relevance [27], a sense of autonomy [25], and engagement [4,27,28]. In PA interventions, wearables can further support this personalization by enabling real-time feedback [30]. Third, prioritizing simple and user-friendly content can increase engagement [4,14,31]. Fourth, incorporating relational strategies, such as motivational interviewing [28] or emoji [3] to mimic human-like interactions, can support long-term behavior change [25,32]. Finally, interdisciplinary collaboration [4,21] and early end user involvement in development [14,20,22,31,33] can improve research and intervention quality. Incorporating these strategies is vital, as low engagement hinders the feasibility, acceptability, and, ultimately, effectiveness of interventions [25,27,28].

In addition to these challenges, the literature highlights several knowledge gaps, including limited research on the feasibility and acceptability of virtual coaches [1,5]. While effectiveness is crucial, long-term adoption also depends on factors such as ease of use, trust in the technology, and user satisfaction, making it essential to understand user needs and experiences [24-26,33,34]. To address this knowledge gap, qualitative and mixed methods designs can help to explore facilitators, barriers, and limitations [14,26,33]. Another knowledge gap is the limited

XSL•FO
RenderX

understanding of who benefits most from virtual coaches and under which circumstances [1,25,26]. Such insights could support tailored intervention development [4], improve feasibility and acceptability across diverse populations [26], and help identify target groups [5]. Particular attention is needed for individuals with lower digital skills, eHealth literacy [20,21,24], and socioeconomic positions (SEP) [1,2,14]. Although mHealth interventions have the potential for broad scalability [2], many remain inaccessible to these populations [20,22,31]. To avoid exacerbating health disparities [20-22], virtual coaches should be designed with accessibility in mind [2,14]. This includes early end user involvement [14,20,22,31,33,35] and tailoring interventions to users' digital skills and eHealth literacy [20,22]. It is also key to examine how individual characteristics, such as age, SEP, and eHealth literacy, influence feasibility and acceptability to identify who benefits most [33].

With an interdisciplinary consortium, we developed Perfect Fit (PF; Perfect Fit consortium) [36], a smartphone-based mHealth intervention with a virtual coach that provides real-time, personalized, text-based, and visual feedback to support smoking cessation and PA promotion. PF was specifically designed to leverage the assumed synergy between these behaviors. It also addresses common challenges and knowledge gaps in mHealth interventions, including those with virtual coaches. For instance, it includes behavior change techniques (eg, goal-setting) [29] commonly used in smoking cessation [37] and PA promotion [38], as well as behavioral theories, namely identity theories [39-41] and the Relapse Prevention Model [42]. Additionally, PF incorporates sensor-based personalized feedback using a smartwatch and was developed iteratively with end users [43].

This study presents the results of the real-world evaluation of PF, using a single-arm, pre-post, mixed methods design. The first aim is to investigate the feasibility and acceptability of PF. The second aim is to explore associations between baseline characteristics (ie, sociodemographic-, smoking-, and PA-related characteristics) and the feasibility and acceptability outcomes. This will provide insight into factors influencing intervention success and PF's suitability for different individuals. The third aim is to investigate the feasibility of conducting the study itself. Many mHealth studies face high dropout rates [21] and difficulties in recruiting participants, particularly those with lower SEP [14]. Examining these factors can inform effective recruitment and retention strategies for future research. By using a mixed methods approach, we gain insight into PF's overall potential, participant experiences, and contextual factors. This will help identify strengths and areas for improvement, ultimately informing future development and evaluation.

Methods

Study Design

This paper reports on a larger study conducted in the Netherlands, in which PF was evaluated using a single-arm, pre-post, convergent mixed methods design [43]. This paper focuses on the feasibility and acceptability of PF, as well as the feasibility of conducting the study itself. The data relevant to this paper were collected between August 2023 and June 2024. Findings on short- and long-term preliminary effectiveness will be reported elsewhere. This study is reported in line with the Good Reporting of A Mixed Methods Study (GRAMMS) checklist [44] (Checklist 1).

To assess the feasibility and acceptability of PF and the study itself, data were collected at baseline, during the intervention (ie, user log data and sensor data), and at postintervention (4 months after baseline). The quantitative component included (1) self-report questionnaires on baseline characteristics, feasibility and acceptability of PF, and virtual coach acceptance; (2) app-generated usage data; and (3) data on study procedures documented in a researcher's log. Self-report questionnaires comprised both self-developed and adapted items, as well as validated instruments, including the eHealth Literacy Questionnaire (eHLQ) [45,46], the Fagerström Test for Nicotine Dependence (FTND) [47], the Godin-Shephard Leisure-Time PA questionnaire (GSLTPAQ) [48,49], and the System Usability Scale (SUS) [50]. The qualitative component consisted of semistructured interviews exploring the feasibility and acceptability of (1) PF, (2) the virtual coach, and (3) the study procedures. The study was approved by the scientific committee of the Department of Public Health and Primary Care at Leiden University Medical Center (approval number: WSC-2023 - 24). Full study details are available in the published protocol [43]. Deviations from the protocol are reported below.

Patient and Public Involvement

Patient and public involvement (PPI) activities were conducted during the development of PF and the execution of this study. Multimedia Appendix 1 provides a description of the PPI aims, methods, results, and reflections, reported according to the revised version of the Guidance for Reporting Involvement of Patients and the Public (GRIPP2) short-form checklist [51] to ensure transparency and comprehensiveness. Additional details on PPI activities can be found in the published article on the development of PF [36].

Participants and Recruitment

Participants were recruited using a combination of online and offline strategies. Examples include social media posts and advertisements, newsletters or websites of affiliated organizations (eg, health care insurers), a news item in a local newspaper, and outreach via the networks of the PF research team and advisory panel. Recruitment started mid-August 2023, and the final participant was included on February 5, 2024. Sample size estimation is discussed in the study protocol [43].

Participants were eligible if they (1) were aged ≥18 years, (2) smoked daily, (3) intended to quit smoking within 6 weeks, (4) could walk pain-free, (5) could understand and read Dutch (at least B1 level), and (6) owned a smartphone with Internet access. To examine PF in high-risk individuals, at least 50% of participants were required to be at higher cardiovascular risk according to the Dutch general practice guideline [52]; that is, women aged ≥55 years or men aged ≥50 years. Additionally, at least 75% of participants had to reside in the Leiden region. Exclusion criteria included (1) enrollment in smoking cessation treatment at intervention start at the time of eligibility screening, (2) major lower extremity surgery in the past year, (3) use of

XSL•FO
RenderX

antipsychotics or diagnosis of a serious psychiatric illness (eg, schizophrenia/psychosis or major depression), and (4) pregnancy.

For the qualitative interviews, we aimed to recruit around 10‐15 participants to achieve data saturation [53]. A heterogeneous subsample was sought, with variation primarily in PF usage level, as well as in gender, age, SEP, and success in changing health behavior.

Intervention

PF is an evidence-based, personalized mHealth intervention with a virtual coach that supports individuals in quitting smoking and increasing PA. The intervention was expected to last around 16 weeks but could be adjusted to individual needs. The virtual coach, "Sam," guided users through 3 phases, including a preparation phase, an execution phase, and a closing dialog. Communication consisted of chat messages enriched with emoji, images, and animated informational videos, and users received links to relevant external sources. The coaching system allowed for both system- and user-initiated conversations, with a mix of constrained (preset responses) and unconstrained (free-text) input. A smartwatch (Garmin Forerunner 55; Garmin Ltd) [54] measured PA, enabling the coach to provide personalized PA goals and feedback. Users interacted with the coach via a smartphone app with chat functionality (ie, the NiceDay app; NiceDay Healthcare Nederland BV; [55], originally developed for remote therapy), and users installed 2 additional apps to enable a connection between the smartwatch and the coaching system.

In the preparation phase, users were encouraged to complete activities and dialogs that prepared them for quitting smoking and increasing PA. These dialogs covered topics such as medication and nicotine replacement therapy and self-monitoring of current behavior. Users were also guided in selecting a quit date and formulating a specific, measurable, achievable, relevant, and time-bound (SMART) long-term PA goal through the coach-initiated goal-setting dialog. After this dialog, the coach provided daily short-term PA goals (ie, step goals) based on previous PA measured via the smartwatch, along with feedback on whether the goal was met. The self-selected quit date (within a predefined range) then marked the start of the execution phase, during which the coach provided support through activities facilitating behavior change and weekly reflections on progress. The closing dialog was the final conversation in which the coach encouraged users to review their progress and achievements and develop a relapse prevention plan. PF included 21 coach-initiated core components considered important for supporting behavior change and forming the overall intervention structure. These consisted of 7 preparation phase dialogs or videos, the introduction video for the execution phase, the weekly reflection dialog occurring 12 times, and the closing dialog. In addition, users could initiate optional components at any time, including dialogs (eg, a high-risk situation and relapse dialog for difficult moments) and 25 short activities (eg, positive self-talk healthy eating tips), which could be repeated as often as desired to enhance engagement and personalization. Further details on the development of PF [36], the intervention components and features [43], the technical architecture [56], and the open-source code of the system [57] are described elsewhere.

Procedures

After consent, an onboarding procedure was initiated. Participants received the smartwatch and information materials. Specifically, participants received a video explaining the study and intervention, an installation booklet with instructions on app installation and smartwatch-coach connection, and an information booklet summarizing key instructions and tips for interacting with the coach. The latter also served as a workbook, offering space for notes. After receiving the materials, participants had a meeting with a researcher (via video call or telephone) to verify app installation, coach connection, and smartwatch use. During the intervention, participants could contact the research team via email or phone for technical support. A detailed participant timeline is described in the study protocol [43].

Quantitative Data Collection

Participants received a link to the online baseline questionnaire (T0) during the week of the onboarding meeting and the postintervention questionnaire (T1) 16 weeks later, at the expected end of the intervention. To maximize completion, researchers reminded participants via email and phone calls. Participants could keep the smartwatch if they completed at least 80% of the questionnaires of the larger study.

Qualitative Data Collection

A subsample of participants who had given prior consent to be contacted for an interview was invited at T1. Interviews were conducted online or in person (at Leiden University Medical Center), lasted approximately one hour, and were carried out by 2 medicine master's students trained in interviewing. In order to prevent social desirability bias, the interview schedule consistently focused on both positive and negative aspects of PF in general and its specific components.

Outcomes

Only measures relevant to the present paper are reported below. The interviews conducted at T1 were used to assess both primary and secondary qualitative outcomes (eg, experiences with PF and research participation). Full details are described in the protocol [43].

Baseline Variables

The following variables were assessed at baseline (T0) and were used to describe the sample and explore associations with the primary outcomes.

Participant Characteristics

Background characteristics were assessed using self-developed items and included gender, age, educational level (as a measure for SEP) [58], and the presence of any physical or mental (chronic) conditions.

eHealth Literacy

Assessed using 5 of the 7 scales of the Dutch eHLQ: (1) using technology to process health information (Cronbach α=0.77; McDonald ω=0.81); (2) understanding of health concepts and

XSL•FO
RenderX

language (Cronbach α=0.77; McDonald ω=0.84); (3) ability to actively engage with digital services (Cronbach α=0.83; McDonald ω=0.88); (4) feel safe and in control (Cronbach α=0.83; McDonald ω=0.87); and (5) are motivated to engage with digital services (Cronbach α=0.81; McDonald ω=0.85) [45,46]. The remaining 2 scales were not included, as they were not considered relevant to our study. Each scale consists of 5 items rated on a 4-point scale from 1 (strongly disagree) to 4 (strongly agree).

Intention to Quit Smoking

Assessed using a single item, "Which of the following plans best applies to you? I intend to…" [59]. Answer categories were (1) quit within the next month, (2) quit between 1 and 6 months from now, (3) quit sometime in the future, beyond 6 months, or (4) not planning to quit.

Smoking Behavior and Physical Nicotine Dependence

Assessed using a Dutch translation of the 6-item FTND (eg, "Do you smoke more in the morning than during the rest of the day?") [47,60]. Total scores were calculated according to standard FTND scoring, resulting in a total score ranging from 0 to 10, with higher scores indicating higher nicotine dependence.

Using E-Cigarettes

Participants were asked, using self-developed items, whether they used e-cigarettes. If they responded "yes," they were subsequently asked whether they used e-cigarettes with or without nicotine.

Intention to Become Sufficiently Physically Active

Assessed using a single item, "Which of the following plans best applies to you? I intend to…" [59]. Answer categories were (1) become sufficiently physically active within the next month, (2) become sufficiently physically active between 1 and 6 months from now, (3) become sufficiently physically active sometime in the future, beyond 6 months, or (4) not planning to become sufficiently physically active. Before this question, the definition of PA (ie, exercise, but also physical activities like walking to the supermarket) and the World Health Organization (WHO)–recommended PA guidelines [7] were explained.

Self-Reported Level of PA

Assessed using a Dutch translation of the 3-item GSLTPAQ [48,49]. Participants were asked to report, on average, how many times during a typical week they engaged in more than 15 minutes of (1) strenuous exercise (eg, running and soccer), (2) moderate exercise (eg, brisk walking and badminton), or (3) mild exercise (eg, easy walking yoga). Example activities for each intensity level were provided. Total scores were calculated according to standard GSLTPAQ scoring. A total score of 24 or more was defined as "active," a score between 14 and 23 as "moderately active," and a score below 14 as "insufficiently active" [48].

Feasibility and Acceptability of PF and the Virtual Coach

To investigate the feasibility and acceptability of PF, the following primary outcomes were assessed [43].

PF Usage (During Intervention and T1)

Objective intervention usage data collected through the app throughout the intervention duration included (1) the number of days used (ie, from the first to the last recorded activity in the app), and (2) the number of completed coach-initiated core intervention components (ie, 21 in total). Components were only logged as completed if finished in full, providing a conservative estimate of usage. As part of the T1 questionnaire, participants also self-reported how frequently they interacted with the virtual coach during PF using a self-developed item.

A deviation from the study protocol [43] was that we did not analyze the self-reported intervention completion item. Due to the personalized intervention duration and content, the definition of "completion" was somewhat subjective. Therefore, app usage data, which objectively tracked completed components, was considered more reliable and also correlated significantly with self-reported completion.

Satisfaction With PF (T1)

Assessed using 2 self-developed items that were rated on Likert scales from 1 to 4 (Spearman-Brown r=0.75, indicating sufficient overlap to allow averaging the scores on the 2 items): "How satisfied are you with the amount of support you received?" and "Would you use the Perfect Fit program again, if needed?" Higher average scores reflected greater satisfaction.

Usability of PF (T1)

Assessed using the 10-item SUS [50] (Cronbach α=0.84; McDonald ω=0.90). Responses were given on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree; eg, "I think the Perfect Fit program is unnecessarily complex"). Standard SUS-scoring was used, resulting in total scores ranging from 0 to 100, with higher scores indicating greater usability.

Acceptance of the Virtual Coach (T1)

Assessed with 6 items adapted from Provoost et al [61] covering satisfaction, usability, willingness to continue interaction, relationship, preference for a human coach versus coach Sam, and adherence to advice from coach Sam [62] (Cronbach α=0.81; McDonald ω=0.90). Items were scored on a 7-point Likert scale from −3 to 3 (eg, "How satisfied were you with Coach Sam?"). Item 2 was reverse-scored, so that higher average scores indicated more positive attitudes.

Qualitative Data From Semistructured Interviews (T1)

Qualitative feedback regarding participants' experiences with PF, the virtual coach, and the smartwatch complemented the quantitative data (see protocol for the interview protocol [43]).

Sensor Data Issues (During Intervention)

During the study, we observed frequent issues with the connection between the smartwatch and the virtual coach. These issues caused the coach to communicate inaccurate step counts and goals and triggered repeated notifications from the coach

XSL•FO
RenderX

to the user about connection issues. The extent and nature of the sensor data issues were investigated, as they may have negatively impacted PF's feasibility and acceptability. This was analyzed as an exploratory outcome and was an addition to the original protocol [43].

Study Feasibility

The following secondary outcomes were used to assess the feasibility of conducting the study: (1) recruitment, response, and consent rates, recorded by the researchers in a participant screening and inclusion log throughout the study; (2) recruitment strategies, assessed at T0 via a self-report item asking participants how they were informed about the study (eg, via social media); (3) study adherence, monitored by the researchers from onboarding until T1 (ie, those who withdrew before completing onboarding were categorized as withdrawals rather than dropouts); and (4) qualitative data from semistructured interviews at T1, capturing participants' experiences with research participation (see protocol for the interview protocol [43]).

Data Analysis

Quantitative

Most data preparation and analyses were performed in SPSS (version 29.0; IBM Corp), and RStudio (version 2024.04.2+764; PBC) was used for calculating McDonald omega and for preparing and analyzing PF usage and step count data. Details on data preparation, including age and SEP coding, winsorizing [63] of extreme outlying PA values, dichotomization of variables for exploratory analyses, and sensor data processing are provided in Multimedia Appendix 2.

Descriptive statistics (eg, mean, SD, and frequencies) were used to summarize baseline characteristics of the study sample. Chi-square tests, independent samples *t* tests, and Mann-Whitney *U* tests were used for dropout analyses.

To address the primary and tertiary study aims—investigating the feasibility and acceptability of PF, acceptance of the virtual coach, and feasibility of conducting the study—descriptive analyses were conducted.

To address the secondary aim—exploring associations between baseline characteristics and the feasibility and acceptability outcomes—exploratory one-way ANOVAs, Kruskal-Wallis tests, Mann-Whitney *U* tests, chi-square tests, and Pearson and Spearman correlations were performed. Post hoc pairwise comparisons were conducted with Tukey honestly significant difference (HSD) corrections for ANOVAs and Bonferroni corrections for Kruskal-Wallis tests.

For the exploratory sensor data issue analyses, descriptive statistics were used. Associations between days with step count data and weekday, phone operating system, and PF usage group were explored with a Pearson chi-square test, a Mann-Whitney *U* test, and a Kruskal-Wallis test.

Qualitative

All semistructured interviews were audio-recorded, pseudonymized, and transcribed verbatim. Qualitative data were analyzed using the framework approach [64,65] in ATLAS.ti (version 23.2.3.27778; ATLAS.ti Scientific Software Development GmbH). The first transcript was independently coded by the 2 master's students who had conducted the interviews and by MHMvV, resulting in an initial coding scheme. The students then independently coded and discussed the second and third transcripts, refining the coding scheme when discrepancies arose. The remaining transcripts were each coded by one student, with the coding scheme further adjusted as needed. All revisions were discussed with MHMvV, RAvE, and EM. Data interpretation and theme generation were carried out by MHMvV in collaboration with EM and AV and were guided by the research aims, with consideration of the corresponding quantitative outcomes. Interview quotations were labeled with participant identifiers according to PF usage category (L=low, M=moderate, and H=high) and an individual interview number. Quotations were translated from Dutch to English with care to preserve the original meaning.

Data Triangulation

The analysis followed a convergent mixed methods approach. Quantitative and qualitative data were first analyzed separately using applicable analytical methods. Qualitative codes were derived from the interview data and organized within categories aligned with the study aims and quantitative outcomes. Subsequently, findings from both data sources were compared and integrated for each outcome. Results are therefore presented per outcome, combining quantitative results with qualitative findings to illustrate convergence and divergence and to provide additional context to the interpretation of the data.

Ethical Considerations

This study has been approved by the scientific committee of the Department of Public Health and Primary Care at Leiden University Medical Center (approval number: WSC-2023 - 24). The Medical Research Ethics Committee Leiden, The Hague, and Delft reviewed the study proposal and provided a declaration of no objection, indicating the research does not fall under the Dutch Medical Research with Human Subjects Law (nWMO; approval number: N23.045 METC-LDD). Interested individuals received a digital information letter and a link to the online screening questionnaire (hosted in secure software, Castor electronic data capture) [66]. Eligible participants were included after signing an online informed consent form and were informed that they could withdraw from the study at any time for any reason. Verbal informed consent was obtained and audio-recorded before each interview. Participants received a €25 (based on the conversion rate of 1.0786 between August 2023 and June 2024, this amount is equivalent to approximately US $26.97) gift voucher for the interview.

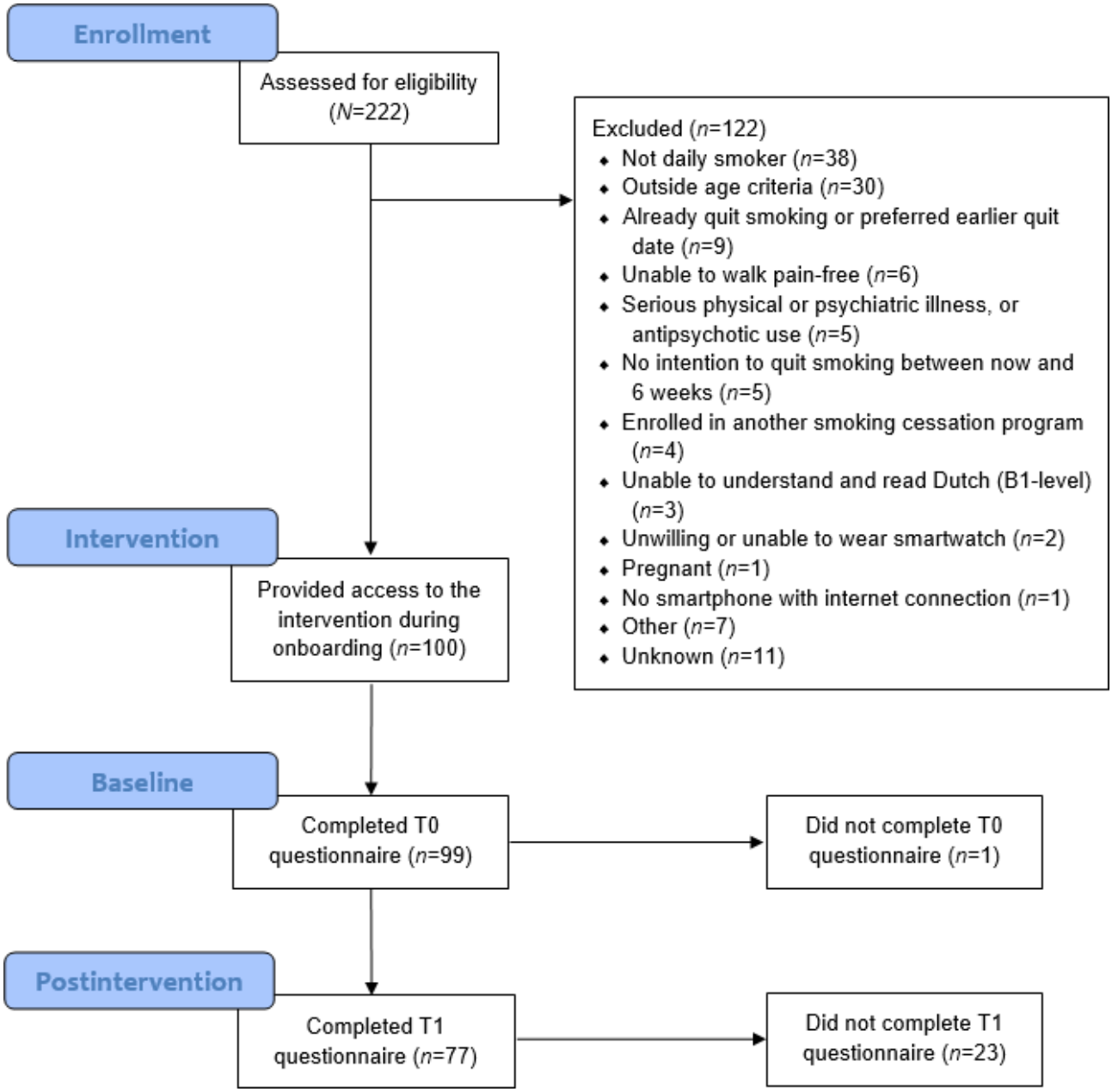
Results

Sample Characteristics

The participant flow diagram is presented in Figure 1 [67]. One hundred participants were given access to PF during onboarding. One participant dropped out before completing the baseline questionnaire, resulting in 99 participants who started PF and completed the baseline questionnaire.

XSL•FO
RenderX

Figure 1. Adapted CONSORT (Consolidated Standards of Reporting Trials) participant flow diagram (adapted from Hopewell et al [67]).



Baseline sociodemographic, smoking- and PA-related characteristics for the total sample (n=99) and T1 completers (n=77) are presented in Table 1. Of the 99 participants, 58.6% (n=58) were female, and the mean age was 51.96 (SD 14.08) years, ranging from 19 to 86 years. The distribution of SEP was comparable to the general Dutch population [68]. Median scores on the 5 eHealth literacy scales ranged from 2.80 to 3.00 on the 4-point scale. Regarding smoking-related characteristics, the majority of participants (n=77, 77.8%) intended to quit smoking within one month. At baseline, participants smoked an average of 15 cigarettes per day and reported low-to-moderate nicotine dependence [47,69]. Regarding PA-related characteristics, 54.5% (n=54) of the participants intended to be sufficiently physically active within one month. The median GSLTPAQ score suggests that at least half of the participants were already sufficiently active at baseline. However, the GSLTPAQ data were highly skewed and included extreme outliers; therefore, absolute values should be interpreted with caution. Dropout analyses showed that none of the baseline characteristics were significantly associated with study adherence at T1.

A total of 12 participants actively withdrew (via email or phone) from both the intervention and the study (ie, full dropouts), and 7 participants discontinued the intervention but continued completing study questionnaires. Reasons for dropout included: (mental) health problems (n=4), persistent technical issues with the smartwatch and/or app (n=4), dissatisfaction with the smartwatch functionality or intervention content (n=4), personal circumstances (n=1), a preference for human support over virtual coaching (n=1), smoking relapse and stress experienced due to the PA component of the intervention (n=1), having quit smoking and therefore no longer needing the intervention (n=1), and unknown reasons (n=3).

XSL•FO
RenderX

Table . Baseline characteristics of the total sample and of postintervention completers.

Baseline characteristics	Baseline (T0) completers (total sample), n=99	Postintervention (T1) completers, n=77
Gender, n (%)		
Female	58 (58.6)	42 (54.5)
Male	40 (40.4)	34 (44.2)
Other	1 (1.0)	1 (1.3)
Age (years), mean (SD)	51.96 (14.08)	51.29 (13.71)
SEP[a], n (%)		
Low	20 (20.2)	18 (23.4)
Middle	44 (44.4)	31 (40.3)
High	34 (34.3)	27 (35.1)
Prefer not to say	1 (1.0)	1 (1.3)
Physical or mental (chronic) conditions, n (%)		
No condition	69 (69.7)	56 (72.7)
Physical condition/conditions	21 (21.2)	15 (19.5)
Mental condition/conditions	5 (5.1)	3 (3.9)
Both physical and mental condition/conditions	4 (4.0)	3 (3.9)
eHealth literacy (eHLQ)[b], median (range)		
Scale 1: using technology to process health information	2.80 (1.80 - 4.00)	3.00 (1.80 - 4.00)
Scale 2: understanding of health concepts and language	3.00 (1.60 - 4.00)	3.00 (2.00 - 4.00)
Scale 3: ability to actively engage with digital services	3.00 (1.80 - 4.00)	3.00 (1.80 - 4.00)
Scale 4: feel safe and in control	3.00 (1.60 - 4.00)	3.00 (1.60 - 4.00)
Scale 5: motivated to engage with digital services	3.00 (1.60 - 4.00)	3.00 (1.80 - 4.00)
Intention to quit smoking, n (%)		
Between now and 1 month	77 (77.8)	60 (77.9)
Between 1 and 6 months	22 (22.2)	17 (22.1)
In the future, but not within 6 months	0 (0.0)	0 (0.0)
Number of cigarettes smoked a day, median (range)	15.00 (2.00 - 60.00)	16.00 (2.00 - 60.00)
Nicotine dependence (FTND)[c], mean (SD)	4.40 (2.23)	4.38 (2.27)
Using e-cigarettes with nicotine, n (%)		
No	90 (90.9)	69 (89.6)
Yes	9 (9.1)	8 (10.4)
Intention to become sufficiently physically active, n (%)		
Between now and 1 month	54 (54.5)	41 (53.2)
Between 1 and 6 months	43 (43.4)	35 (45.5)
In the future, but not within 6 months	2 (2.0)	1 (1.3)
Never	0 (0.0)	0 (0.0)
Level of PA[d] (GSLTPAQ)[e], median (range)	24.00 (0.00 - 79.00)	25.00 (0.00 - 79.00)

[a]SEP: socioeconomic position.

[b]eHLQ: eHealth Literacy Questionnaire.

XSL•FO
RenderX

[c]FTND: Fagerström Test for Nicotine Dependence.

[d]PA: physical activity.

[e]GSLTPAQ: Godin-Shephard Leisure-Time Physical Activity Questionnaire.

Qualitative Interview Sample

Twelve participants took part in the postintervention individual interviews, forming a heterogeneous sample, including 3 low PF users, 5 moderate users, and 4 high users. Seven of the 12 participants were male, the average age was 54.8 (range 35 - 77) years, and SEP levels varied (see Table S1 in Multimedia Appendix 2 for detailed characteristics).

Primary Outcomes

Feasibility and Acceptability PF

Usage of PF

Twelve of the 99 participants had missing PF usage data, likely due to dropping out before entering their participant code in the app or entering it incorrectly, which prevented linkage between their app usage data and questionnaire responses.

Participants with available PF usage data (n=87) completed on average 54.7% of the 21 core intervention components (mean 11.49, SD 5.89). Six participants completed all components. The median PF usage duration was 107 days (range 4 - 202), which closely approximates the expected intervention period of 16 weeks (ie, 112 days). Participants completed a median of 4 optional short PF activities (range 0 - 46, as all 25 available activities could be repeated). Twelve participants completed none. Participants' self-reported frequency of weekly coach interaction is presented in Multimedia Appendix 2. A significant positive correlation was found between the number of completed core components and optional activities (Spearman ρ=0.638; P<.001). For subsequent exploratory analyses, usage was categorized into 3 groups (low, moderate, and high) based on the number of core components completed. See Table S2 in Multimedia Appendix 2 for details on this categorization and the corresponding distribution of optional activities completed within each group.

Exploratory analyses (n=77) revealed significant differences between PF usage groups in the other primary outcomes: satisfaction with PF ($F_{2, 74}$=6.09; P=.004; η^2=0.14), usability of PF (H_2=11.16; P=.004), and acceptance of the virtual coach ($F_{2,74}$=8.59; P<.001; η^2=0.19). Post hoc tests showed that high PF users reported significantly higher scores than moderate users on all 3 outcomes. Additionally, high users reported significantly higher scores on the acceptance of the coach than low users.

Qualitative data provided insight into reasons for lower usage of PF. Some participants reported that, after a while, they found PF or the virtual coach too intense, repetitive, or even irritating, which led to lower usage:

> *In the last few weeks, I used Coach Sam a bit less because I found it quite intense.* [M2, participant who was still smoking and increased PA at T1]

Others expressed a growing desire for autonomy, leading them to reduce their interaction with the coach as they gained confidence in managing their behavior change independently:

> *Over time, I noticed I got used to it […] and started looking at it less. In the beginning, I used it a lot, but gradually I started thinking. 'I'll do this myself now.'* [H2, abstinent and decreased PA at T1.]

Satisfaction With and Usability of PF

At postintervention, the mean satisfaction rating was 2.79 (SD 0.73; n=77) on a scale from 1 to 4. Perceived usability had a median score of 67.50 (range 12.50 - 87.50), indicating OK-to-good usability [70].

Qualitative interview data revealed that participants expressed both positive and negative experiences regarding satisfaction with and usability of PF, which are summarized in Table 2. Some features were experienced differently by participants, such as perceptions of how enjoyable and motivating PF was and the perceived variation in short optional activities. Issues with the connection between the smartwatch and the coach were often mentioned and likely impacted PF's feasibility and usability (see Results, Exploratory analyses section, "Sensor data issues").

XSL•FO
RenderX

Table . Perceived strengths and points for improvement of Perfect Fit, with illustrative quotations from the qualitative interviews, grouped under descriptive labels reflecting recurring topics in participants' responses.

Strength	Point for improvement
General experience	
Useful, comprehensive, and informative • "I really think it's a good product you've created, and I genuinely believe it's useful and effective for people seeking that kind of support. I do hope that […] such a product could become available on the market for smoking cessation and increasing physical activity." (H1[a], abstinent and decreased PA[b] at T1[c])	—[d]
Enjoyable[e] • "When it worked properly [without technical issues], I enjoyed using Perfect Fit." (H3, abstinent and increased PA at T1)	Not sufficiently enjoyable/engaging • "It worked, so I thought it was a good program. But I didn't find it very engaging. I wasn't looking forward to it like, 'Oh great, I get to use the program again soon', or anything." (M5, abstinent and increased PA at T1)
Motivating • "It's a good initiative. I think it stimulates people. […] Those videos were good, also the questions that were asked and the occasional tips you got. […] Like, 'What is this really doing to you?'. It actually says what you already know. We all know it, but we push it aside, and those things were confronting again. Like, 'Oh right, that could indeed be a consequence' – but it was told in a pleasant and friendly way." (H2, abstinent and decreased PA at T1)	Not sufficiently motivating • "In general, I found it very interesting because you're really working on your own development. You really start thinking about it. And you actually do move more and smoke less. Except at the end of the program, because maybe you become a bit more complacent – or maybe coach Sam just didn't provide the right trigger. I thought that was a pity. But I do think it's very personal." (M2, smoking and increased PA at T1) • "It didn't motivate me enough to quit smoking." (L1, smoking and increased PA at T1)
Personalization	
Timing: on-demand support and 24/7 availability • "Of course, you have people around you who support you when you're trying to quit smoking. But Sam was there 24/7. So even late at night or early in the morning, I could reach out to Sam." (M3, abstinent and increased PA at T1)	—
Content: personalized step goals • Interviewer: "Which aspects of the program did you like?" P: "Definitely the step goal. You want to reach it, so that's motivating when you get it. I used it a lot." (H4, abstinent and increased PA at T1)	—
Content: variety in short optional activities • "What I liked about the program was the variety – you didn't keep getting the same thing, but you could choose. I enjoyed doing the relaxation exercises and watching the information videos." (H3, abstinent and increased PA at T1)	Content: insufficient variety in short optional activities • "I found them a bit underwhelming [referring to the PF activities], because they didn't really add anything for me. […] They didn't shift my mindset. […] You know, you've been at it for so long, and you already know everything so well – that smoking is bad for you. You've already tried quitting a thousand times." (M2, smoking and increased PA at T1) • "Maybe expand the activities. […] For example, once you've completed an activity, you could move on to a next step – like, 'you've finished this level, now you can continue to the next level' or something like that." (H1, abstinent and decreased PA at T1)
—	Content: insufficient variety during execution phase • "The week after, I just got exactly the same questions again, and I would have preferred if the tone or the way the questions were asked had been a bit different. So that you'd feel like you were progressing further in the program together with Sam." (M3, abstinent and increased PA at T1)
Technical aspects and usability	

XSL•FO
RenderX

Strength	Point for improvement
Clear and easy to use • "It wasn't disappointing. I mean, I'm obviously not 20 anymore, so I didn't grow up with apps and computers. I don't think I'm completely incapable, but it's not my hobby. So it's always a bit nerve-wracking, but it's quite manageable, even if you're less skilled." (H3, abstinent and increased PA at T1)	Complicated installation and app integration • "The installation was technically challenging. It was difficult that you needed to link the Garmin Connect app with the NiceDay app and also to add Sammy as a chatbot. That was a bit complicated and tricky, and it should be simpler." (L1, smoking and increased PA at T1)
—	Technical issues • "Except when it malfunctioned, I thought: here we go again. 'I haven't heard anything for a while,' Sam said. Then I thought, well, it's not that long, Sam, but yeah, you can't really express that (laughing)." (H3, abstinent and increased PA at T1)
—	Smartwatch connection issues • "I contacted you [the researchers] again at some point because the step counter wasn't working. I couldn't get it to work. And eventually, I sort of gave up – like, 'Never mind, I don't want to be a bother.' If I'd had a mid-program evaluation talk [with a researcher], I might have brought it up again. I would've really appreciated that." (M5, abstinent and increased PA at T1)
Responsive technical support and troubleshooting • "And you [the researchers] were always very accessible, which makes a difference. If something doesn't run smoothly, it's easy to get in touch. […] It's just a quick email and you get an immediate reply, so that makes things less frustrating when something goes wrong." (H1, abstinent and decreased PA at T1) • "A few times, I saw that zero steps were recorded, and I wondered how that could be. Then I checked the booklet [the Perfect Fit paper manual]: 'Oh right, every time something happens or if you put your phone in battery-saving mode, you have to reopen the app.'" (H2, abstinent and decreased PA at T1)	—

[a]Interview ID is based on Perfect Fit usage: low (L), moderate users (M), and high (H) users.

[b]PA: physical activity.

[c]T1: postintervention.

[d]Not applicable.

[e]Quotes placed opposite each other illustrate opposing experiences reported by different participants; not every strength has a corresponding point for improvement, and not every point for improvement corresponds to a strength.

None of the 12 interview participants reported feeling uncomfortable or unsafe when sharing personal information during the intervention. Some noted that the questions were not particularly sensitive and/or attributed their sense of safety to the study's affiliation with Leiden University Medical Center, an academic hospital in the Netherlands:

> *Yes, I felt safe sharing information. Although I'm not exactly sure why – maybe because it was still in the research phase and the data was handled confidentially for a good cause. […] And it wasn't that much personal information anyway.* [L3, smoking and increased PA at T1]

Acceptance of the Virtual Coach

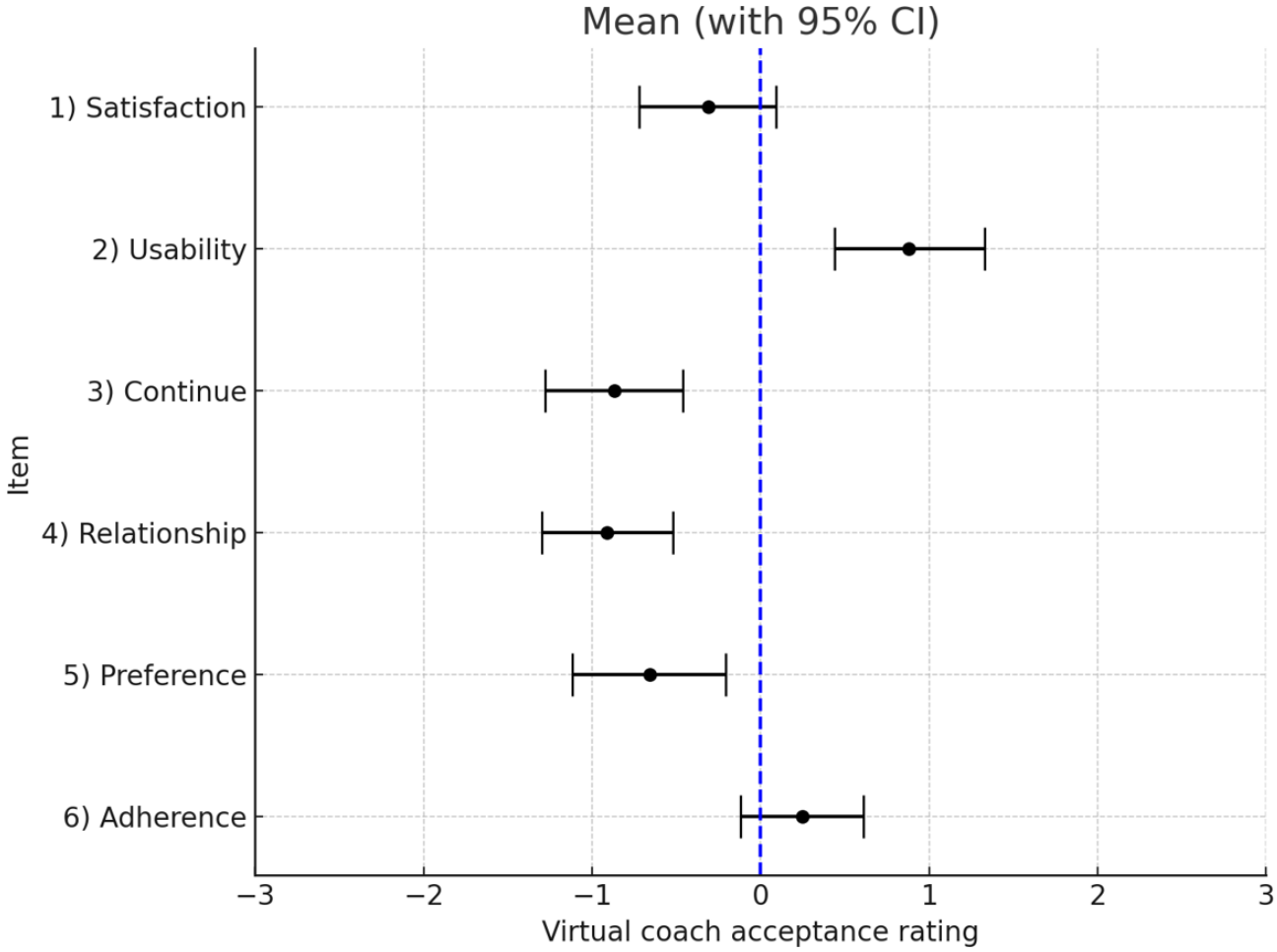
At postintervention, the mean virtual coach's acceptance rating was −0.27 (SD 1.30; n=77) on a scale from −3 to 3. Figure 2 shows the mean scores and 95% CIs for each of the 6 items assessing coach acceptance. The item on usability received the highest rating, whereas the coach-user relationship was rated the lowest.

Qualitative findings provided context and depth to the quantitative results, for instance, regarding the moderately negative appraisal of the virtual coach, "I think the setup [of PF] is really great. […] But the virtual coach – yeah, I actually liked that the least." (L1, smoking and increased PA at T1). Participants frequently described the coach's communication as repetitive and static in style and content. Many expressed a desire for more flexibility, suggesting the coach should be able to adapt to personal situations (eg, being ill and unable to exercise) or user input (eg, by reflecting on user messages or tailoring content accordingly).

> *Probably because he [the virtual coach] didn't learn from my experiences. […] At one point, there was an exercise where you had to note why things were going well or not. I typed it in my notes or the app, but it wasn't read or analyzed. […] So that doesn't help. Maybe it's useful for a bit of self-reflection, but […] that coach didn't learn anything from me.* [L3, smoking and increased PA at T1]

XSL•FO
RenderX

Figure 2. Mean scores (95% CIs) for the 6 virtual coach acceptance items (with varying response labels, eg, difficult-easy), with higher scores indicating more positive attitudes toward the virtual coach.



Still, several participants expressed appreciation for the support provided by the coach and its friendly, nonjudgmental, and empathetic tone:

> *Not judgmental and really encouraging when you'd done something well. Like occasionally checking in: 'How are you feeling?', 'How's it going?', ' Any difficult moments?', 'No?', ' Great job.'* [H1, abstinent and decreased PA at T1]

Participants held differing views on specific features of the coach. For example, the frequency of notifications and the chatbot format were appreciated by some but criticized by others. The fact that Sam was a chatbot, rather than a human coach, was seen as a benefit by some participants, as it offered a sense of anonymity, nonjudgment, and autonomy. In contrast, others perceived it less favorably, noting that it reduced their sense of personal accountability and placed full responsibility on themselves:

> *Of course it's a computer, but I actually liked that. […] Constantly talking to a real person – 'Why did you smoke?' – that can get frustrating. Like: I don't know, just drop it. And with this program, you could choose what you wanted and how you wanted to receive support.* [H3, abstinent and increased PA at T1]

> *You know, a lot of things in the app are quite noncommittal, and I wouldn't mind a stricter Sam. […] Less 'you may', and more 'you must'.* [M1, smoking and increased PA at T1]

These different experiences of participants appeared to influence their perceived bond with coach Sam. While the above-mentioned positive aspects seemed to contribute to a sense of support, several participants felt that Sam being a computer made it difficult—if not impossible—to form a real connection, unlike with a human coach. At the same time, participants noted that developing a human-like bond was not necessary to find Sam's support helpful or to have a positive experience with PF, "What did I think of coach Sam? […] A nice computer, right? (laughs)" (H2, abstinent and decreased PA at T1).

Exploratory Analyses

Sensor Data Issues

Given the unexpectedly frequent issues with the connection between the smartwatch and the virtual coaching system (see Methods, Outcomes section, "Sensor data issues"), the extent and nature of these issues were explored post hoc.

Among the 87 participants with available PF usage data, the median percentage of days with available step count data during the intervention period (ie, from first to last activity in the PF

XSL•FO
RenderX

app) was only 43.1% (range 0.0% - 98.3%), indicating substantial missing data. Several participants mentioned smartwatch connection issues during contact with the research team. These issues occurred despite participants wearing the smartwatch, suggesting that the missing data were primarily due to technical issues. Supporting this, exploratory analyses showed that the missing step count data were not associated with specific days of the week (χ^2_6=5.97; *P*=.43; Cramér *V*=0.03), suggesting that missing data were not due to patterned nonwear (eg, on weekends) but more likely due to technical issues. This is further supported by the significant association between operating system and the percentage of days with step count data (*U*=670.00; *Z*=−2.14; *P*=.03; *r*=−0.23). Specifically, Android users (n=51, 58.6%; median rank=48.86) had significantly more days with data than iPhone users (n=36, 41.4%; median rank=37.11), possibly reflecting connectivity issues specific to the operating system. In addition, there were significant differences between PF usage groups in the percentage of days with step count data (H_2=24.29; *P*<.001). Post hoc comparisons showed that participants in the high-usage group (median 85.6%, range 0.0% - 98.3%) had significantly more days with step count data than those in the moderate-usage (median 26.8%, range 0.0% - 88.8%; *P*<.001) and low-usage groups (median 30.8%, range 0.0% - 97.8%; *P*=.003). These findings could indicate that persistent connection issues might have contributed to lower usage. This interpretation is supported by qualitative data: some participants reported feeling demotivated or even discontinuing their use of PF due to unresolved connection issues between the smartwatch and coach (also see the quotation accompanying the point for improvement "Smartwatch connection issues" in Table 2):

> *I sent several emails about it, but it just kept being difficult that Sam didn't register my steps. And that was actually demotivating – when Sam would say, 'you've walked zero steps today.' Then I'd try again [to fix it], and sometimes it worked, sometimes it didn't. I still can't put my finger on it. But yes, that did demotivate me at times.* [M3, abstinent and increased PA at T1]

Although potential causes of the sensor connection issues were identified, many remained difficult to trace. Troubleshooting strategies had limited and often temporary effects. Technical lessons learned and recommendations for future research are provided in Multimedia Appendix 3.

Associations

To explore associations between baseline characteristics and PF usage, satisfaction, usability, and coach acceptance, we conducted exploratory quantitative analyses, complemented by qualitative data that provided contextual insight. As PF was specifically designed to be suitable for individuals with lower SEP and eHealth literacy, results related to these variables and significant associations with other variables (*P*<.05) are highlighted here. Full details and additional nonsignificant results are provided in Table S3 in Multimedia Appendix 2.

Age differed significantly between PF usage groups. Post hoc comparisons showed that participants in the high-usage group (mean 58.39, SD 12.79) were older than those in the low-usage (mean 50.30, SD 13.95; *P*=.05) and moderate-usage groups (mean 48.47, SD 13.85; *P*=.01). This contrasts with some of the interview data, in which older participants expressed that they expected PF to be easier to use for younger individuals, often linking this to digital skills:

> *When I typed something and Sam replied, it happened very quickly, and every time I had to scroll back to read it. But as soon as he typed something new, it jumped up again. I found that really annoying. […] Younger people probably find that easier, they're also faster than me.* [M4, 61 years]

Digital literacy was also mentioned more generally as influencing ease of use, particularly in dealing with errors, such as the smartwatch connection issues:

> *You had to first open and then close the app again to get a connection. Luckily, I'm quite tech-savvy, so I sorted it out quickly. But I can imagine that being very frustrating if you're not.* [M1, 48 years]

However, a few older participants with lower digital skills noted that they still found PF manageable (see quotation accompanying the strength "Clear and easy to use" in Table 2).

In addition to age, quantitative analyses showed that the eHealth literacy scale 4—"Feel safe and in control"—was weakly, positively correlated with the perceived usability of PF. This suggests that participants who felt more ownership over their data and perceived it as secure reported higher usability of PF.

No significant associations were found in the quantitative analyses between SEP, eHealth literacy subscales (except for scale 4), and any of the primary outcomes. This suggests that PF may have been perceived as similarly feasible and acceptable across these groups. Interestingly, a trend-level difference in satisfaction with PF was observed across SEP groups, with somewhat higher satisfaction among participants with low SEP (mean 3.03, SD 0.67) compared to middle (mean 2.90, SD 0.70) and high SEP participants (mean 2.56, SD 0.71). However, a post hoc power analysis using the largest observed effect size for SEP and eHealth literacy associations (*f*=0.30) indicated only about 61% power with the current sample size. This likely limits the ability to detect significant differences.

Secondary Outcomes

Recruitment, Response, Consent Rates, and Recruitment Strategies

Recruitment took place over approximately 6 months. In total, 361 individuals expressed interest in the study, of whom 222 were screened for eligibility, resulting in a response rate of 0.61. The consent rate among those eligible was 0.83 (see Figure 1). These rates indicate both the feasibility of the recruitment strategies and substantial initial interest in PF. The target sample size of 100 was reached before all eligible individuals could be enrolled, further suggesting promising user interest in a virtual coach-based mHealth intervention for smoking cessation and PA promotion.

XSL•FO
RenderX

Among the 100 participants enrolled, recruitment was most effective through newsletters from 2 health insurance companies (n=37, 37.0%), followed by social media (mainly Facebook advertisements) and recruitment via family, friends, or work contacts (both n=21, 21.0%). More labor-intensive strategies, such as in-person flyer distribution or a local newspaper interview, yielded only a few participants (both n=4; see Table S4 in Multimedia Appendix 2).

Since recruiting participants from lower SEP backgrounds is often reported as challenging [14], we also examined recruitment effectiveness across SEP groups. No significant differences in recruitment strategy were found between SEP groups (χ^2_{18}=20.06; *P*=.33; Cramér *V*=0.33).

Study Adherence

A significant association was found between PF usage (low, moderate, and high) and study adherence at T1 (χ^2_2=9.32; *P*=.009; Cramér *V*=0.31). T1 completers (n=77) were more likely to be in the high-usage group (n=26, 33.8%) than noncompleters (n=7, 9.1%). Conversely, most noncompleters (n=15, 63.6%) were in the low-usage group, compared to 29.9% (n=23) of completers.

Discussion

Principal Findings

This single-arm, pre-post, convergent mixed methods study primarily aimed to investigate the feasibility and acceptability of PF, an mHealth intervention with a virtual coach offering real-time, personalized feedback to support both smoking cessation and PA promotion. Additional aims included exploring associations between baseline characteristics and feasibility and acceptability outcomes, as well as examining the feasibility of conducting the research study. Overall, PF showed adequate feasibility and acceptability, and study procedures proved feasible. Furthermore, the mixed methods design yielded valuable insights into participants' experiences and areas for improving PF and virtual coach interventions in general.

The main findings showed that PF usage varied considerably across participants. Descriptive analyses indicated moderate-to-good satisfaction and usability of PF, but virtual coach acceptance was somewhat negative. High PF users reported greater satisfaction with PF, usability of PF, and coach acceptance. Frequent connection issues between the smartwatch and coach emerged as a disruptive factor, reported more often in the low- and moderate-usage groups than in the high-usage group. Qualitative results provided in-depth insights into PF's feasibility and acceptability, encompassing both positive and negative experiences. Exploratory analyses suggested that high PF users were older than those with moderate and low usage and that the eHealth literacy scale "Feel safe and in control" was weakly positively correlated with usability. Importantly, PF was perceived as similarly feasible and acceptable across SEP groups, aligning with one of the research project's goals. Finally, there was substantial initial interest in PF, and study retention postintervention was 77%, which is relatively high given typical attrition in eHealth studies [21,71].

Understanding Variation in Usage and Experiences

Variation in PF usage and experiences may first be explained by user engagement, conceptualized as (1) the extent of usage and (2) the subjective experience (ie, attention, interest, and affect) [72]. Regarding usage extent, quantitative data showed that weekly coach interaction and intervention duration were generally high; yet, completion of core intervention components varied widely. This may reflect the intervention's personalization but also differences in satisfaction and usability. Subjective engagement, inferred from qualitative data, also varied: some participants found PF motivating and enjoyable, whereas others did not. Increased attention and interest were reported when content prompted reflection or self-reflection, felt personally relevant, or was delivered empathetically. Participants reported the need for evolving goals and content to maintain relevance, challenge, and a sense of progress. This may support perceived competence, potentially leading to high engagement or "flow," consistent with psychological flow theory [73]. This emphasizes the value of personalization, as shown in prior studies [32,74,75]. A second explanation for variation in usage and experiences concerns technical issues. These issues caused frustration and loss of interest for some participants, mirroring findings from a previous review [25]. A particularly disruptive issue was the smartwatch-coach connection, interfering with step tracking and goal personalization. Findings indicated that these issues were associated with lower PF usage and early dropouts, consistent with a review on mHealth interventions for PA promotion [75]. Although many connection issues were difficult to trace, we shared lessons learned (Multimedia Appendix 3) to help mitigate them in future studies. For example, implementing real-time monitoring systems may help detect missing sensor data early and allow researchers to intervene before data loss undermines engagement or intervention delivery. Together, these findings highlight 2 potential factors underlying variability in usage and experiences, offering guidance for future interventions.

Challenges and Opportunities for the Virtual Coach

The virtual coach was the least appreciated component of PF. Qualitative feedback showed participants generally valued its supportive, empathetic, and nonjudgmental style but were critical of its repetitive, static communication and limited adaptability to user input or personal situations. Similar concerns have also been reported in a previous review of virtual coaches for smoking cessation [25]. One potential strategy to improve the coach's communication is the use of natural language processing, which can increase variation and adaptability by training the system to classify and generate responses based on large textual datasets [1,21,25,76]. While recent advances make this a promising approach that might increase the coach's acceptability, such models can produce potentially misleading, inappropriate, or nonfactual outputs [21,77,78]. It remains crucial to balance flexibility (eg, responsiveness to context) with controllability, consistency in content delivery, and safety. As rule-based systems offer greater control [21,76], a hybrid approach combining rule-based and probabilistic techniques may be most suitable for future virtual coach interventions [79].

XSL•FO
RenderX

Accessibility and Inclusivity

We aimed to make PF accessible to individuals often underserved by health behavior change interventions, including those with a lower SEP, limited eHealth literacy, or digital skills [21,25]. Exploratory analyses revealed no significant differences in feasibility and acceptability across SEP groups. Although the study was underpowered to detect small effects, there was a trend toward higher PF satisfaction among participants with a lower SEP. Moreover, participants scoring higher on the "feel safe and in control" eHealth Literacy subscale reported greater usability of PF. Qualitative feedback, especially from older participants with self-reported lower digital skills, indicated that they expected PF to be easier for younger participants or would have appreciated additional support (eg, from a human assistant). However, experiences varied, as some older participants with limited digital skills found PF manageable. Interestingly, participants in the high-usage group were significantly older than those in the low- and moderate-usage groups. These findings suggest that older age does not necessarily limit the ability to use PF, possibly because PF was designed to be suitable for individuals with lower digital skills. Furthermore, older users may have had fewer prior experiences with digital interventions, which can lead to lower expectations and, consequently, a more positive appraisal of the intervention [75]. Finally, dropouts did not differ significantly from completers in SEP, age, or eHealth literacy. These findings are noteworthy given concerns that virtual coaches (for smoking cessation) may exacerbate health disparities [21,25]. Developing accessible and inclusive interventions is therefore essential, and these findings indicate that PF may be a promising step in that direction.

Strengths and Limitations

This study has several strengths. First, PF integrates behavior change techniques commonly used in smoking cessation [37] and PA promotion [38], as well as behavioral theories. Additionally, we incorporated features previously reported as lacking in virtual coaches, such as relational strategies [25], and addressed literature gaps, including transparent intervention reporting [25], which we provided in our published protocol [43]. Moreover, PF was developed by an interdisciplinary team in collaboration with end users [36], likely enhancing feasibility and acceptability. Second, we conducted an in-depth evaluation of PF's feasibility and acceptability. These aspects remain underresearched in virtual coaches [1], despite their importance for long-term adoption [24-26,33,34]. Our convergent mixed methods approach captured both general patterns and nuanced user experiences, helping to identify key strengths and areas for improvement. Third, initial interest in PF was high, and the retention rate was approximately 77% at postintervention (4 months after baseline). This is a positive outcome given the typically high dropout rates in digital interventions [21,71]. Furthermore, recruiting participants from lower socioeconomic backgrounds is often challenging [14] but essential to promote digital inclusion. Therefore, we applied recruitment strategies to reach a broad, diverse audience, resulting in a SEP distribution comparable to the general Dutch population [68].

This study also has some limitations. First, we did not include a dedicated user engagement questionnaire (eg, the short form of the User Engagement Scale [80]), which might have provided a more explicit assessment of engagement. However, there is ongoing debate about the best way to measure engagement in virtual coach interventions, and many studies rely solely on objective usage data, overlooking users' subjective experiences [1,25]. By applying a mixed methods approach, we inferred subjective engagement from qualitative data and identified potential barriers and facilitators. Second, an error in the SEP measurement prevented distinguishing between lower-level (1-2) and middle-level (3-4) vocational education. As a result, all participants with vocational education were classified as middle SEP, likely underestimating the lower SEP group. Third, postintervention completers generally showed higher PF usage, which may have led to an overestimation of feasibility and acceptability. Furthermore, although we aimed to recruit a diverse group for the interviews, participants who were more engaged with PF may have been more likely to participate. To address this, we explicitly invited participants with different levels of PF usage, leading to interviews with 3 low PF users. This may have helped mitigate potential overestimation.

Research and Practical Implications

Several implications emerged for further development of PF and virtual coaches in general. Key areas for improvement included greater content variation, enhanced adaptability of the coach to user input and personal situations, integration of intervention into a single app, and a more reliable smartwatch-coach connection. Although PF incorporated personalization strategies (eg, adaptive step goals), qualitative feedback highlighted the need for additional personalization. For example, adapting the coach to users' preferred coaching style (eg, more directive vs autonomy-supportive) may enhance engagement [81,82]. As engagement can change over time [28], personalization based on dynamic user states (eg, motivation) may also be beneficial [25,83]. Besides enhancing personalization, identifying which populations these interventions best serve is important. Our findings revealed varying preferences; for instance, some participants valued a virtual coach for anonymity, low-threshold access, and a sense of control, while others favored a human coach for greater accountability. These insights can guide future research on aligning population characteristics with different delivery modes of virtual coach interventions. At the same time, broad accessibility should be promoted to avoid exacerbating health disparities, for instance, by exploring different implementation formats. PF was evaluated as a standalone intervention, which may be suitable for those preferring self-guided support or facing barriers to human care. Adding low-level human involvement could further support individuals with lower digital skills (as indicated by our qualitative findings) or lower SEP [14], while considering limited health care resources and ethical allocation. Strategies such as human feedback messages [83] or adaptive designs escalating from low-intensity digital support to human involvement at certain moments [84] may help achieve this balance. Future research should explore how to balance effectiveness, resource use, and equity in virtual coach interventions.

XSL•FO
RenderX

Conclusions

PF shows adequate feasibility and acceptability as a virtual coach-based intervention for smoking cessation and PA promotion. It is accessible to groups often underserved by behavior change interventions, such as those with lower SEP or older age. Usage varied considerably, potentially due to differences in engagement and technical issues (eg, smartwatch-coach connection). Participants valued the coach's empathetic style but noted limitations in its adaptability. This underscores the need for additional personalization, for example, through hybrid approaches combining rule-based systems with natural language processing. Variation in participant experiences highlights the importance of identifying which populations are best served by virtual coaches while promoting broad accessibility to avoid exacerbating health disparities. One strategy is to add low-level human involvement while accounting for limited health care resources.

Overall, PF shows potential as an accessible, inclusive multibehavior change intervention that could benefit public health. This study lays the groundwork for follow-up research evaluating PF's effectiveness and suitable implementation strategies, and provides recommendations for the development of future virtual coach interventions.

Acknowledgments

The authors would like to thank the Perfect Fit advisory board members, members of the Perfect Fit consortium, including collaborating public and private parties, experts who shared their knowledge during expert sessions, and several students for their contributions to the development of the Perfect Fit intervention and this research study.

The members of the Perfect Fit consortium are as follows:

Willem-Paul Brinkman (Department of Intelligent Systems, Delft University of Technology, Delft, The Netherlands), Sven van der Burg (Netherlands eScience Center, Amsterdam, The Netherlands), Winifred A. Gebhardt (Department of Psychology, Unit Health, Medical and Neuropsychology, Leiden University, Leiden, The Netherlands), Joost N. Kok (Faculty of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands), Mark Neerincx (Department of Intelligent Systems, Delft University of Technology, Delft, The Netherlands), Jasper Reenalda (Biomedical Signals and Systems, University of Twente, Enschede, The Netherlands)

Declaration of generative AI and AI-assisted technologies in the writing process:

During the preparation of this work, the authors used ChatGPT (OpenAI, 2025) to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

This work is part of the multidisciplinary research project Perfect Fit. This research received funding from the Netherlands Organisation for Scientific Research (NWO) program Commit2Data - Big Data & Health (project number 628.011.211). The program was funded by the following parties: NWO, the Netherlands Organisation for Health Research and Development (ZonMw), Hartstichting, Ministry of Health, Welfare and Sport (VWS), Health Holland, and the Netherlands eScience Center. The publication reflects only the authors' views and the funders are not liable for any use that may be made of the information contained herein.

Data Availability

Data generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: MHMvV, EM, NHC, AV
Data curation: MHMvV, NA, KMP, WB, BLS, RAvE
Formal analysis: MHMvV
Funding acquisition: EM
Investigation: MHMvV, RAvE
Methodology: MHMvV, EM, AV
Project administration: MHMvV, EM, RAvE, AV
Resources: MHMvV, EM, NA, KMP, WB, BLS, AV
Software: NA, WB, BLS
Supervision: EM, NHC, AV
Visualization: MHMvV
Writing – original draft: MHMvV
Writing – review & editing: MHMvV, EM, NA, KMP, WB, BLS, RAvE, NHC, AV

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed reporting of patient and public involvement (PPI) using the Guidance for Reporting Involvement of Patients and the Public (GRIPP) 2 short form.
[DOCX File, 25 KB - humanfactors_v13i1e83456_app1.docx]

Multimedia Appendix 2
Data preparation and additional results.
[DOCX File, 36 KB - humanfactors_v13i1e83456_app2.docx]

Multimedia Appendix 3
Sensor connection issues: technical lessons learned and recommendations.
[DOCX File, 22 KB - humanfactors_v13i1e83456_app3.docx]

Checklist 1
Good Reporting of a Mixed Methods Study (GRAMMS) checklist.
[DOCX File, 20 KB - humanfactors_v13i1e83456_app4.docx]

References

1. Oh YJ, Zhang J, Fang ML, Fukuoka Y. A systematic review of artificial intelligence chatbots for promoting physical activity, healthy diet, and weight loss. Int J Behav Nutr Phys Act 2021 Dec 11;18(1):160. [doi: 10.1186/s12966-021-01224-6] [Medline: 34895247]
2. McCool J, Dobson R, Whittaker R, Paton C. Mobile health (mHealth) in low- and middle-income countries. Annu Rev Public Health 2022 Apr 5;43(1):525-539. [doi: 10.1146/annurev-publhealth-052620-093850] [Medline: 34648368]
3. Santarossa S, Kane D, Senn CY, Woodruff SJ. Exploring the role of in-person components for online health behavior change interventions: can a digital person-to-person component suffice? J Med Internet Res 2018 Apr 11;20(4):e144. [doi: 10.2196/jmir.8480] [Medline: 29643048]
4. Domin A, Spruijt-Metz D, Theisen D, Ouzzahra Y, Vögele C. Smartphone-based interventions for physical activity promotion: scoping review of the evidence over the last 10 years. JMIR Mhealth Uhealth 2021 Jul 21;9(7):e24308. [doi: 10.2196/24308] [Medline: 34287209]
5. Gold N, Yau A, Rigby B, Dyke C, Remfry EA, Chadborn T. Effectiveness of digital interventions for reducing behavioral risks of cardiovascular disease in nonclinical adult populations: systematic review of reviews. J Med Internet Res 2021 May 14;23(5):e19688. [doi: 10.2196/19688] [Medline: 33988126]
6. Noncommunicable diseases. World Health Organization. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2026-06-02]
7. Physical activity. World Health Organization. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/physical-activity [accessed 2026-06-02]
8. Bryazka D, Reitsma MB, Abate YH, et al. Forecasting the effects of smoking prevalence scenarios on years of life lost and life expectancy from 2022 to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet Public Health 2024 Oct;9(10):e729-e744. [doi: 10.1016/S2468-2667(24)00166-X]
9. Duan YP, Shang BR, Liang W, Du GH, Yang M, Rhodes RE. Effects of eHealth-based multiple health behavior change interventions on physical activity, healthy diet, and weight in people with noncommunicable diseases: systematic review and meta-analysis. J Med Internet Res 2021;23(2):e23786. [doi: 10.2196/23786]
10. Prochaska JJ, Prochaska JO. A review of multiple health behavior change interventions for primary prevention. Am J Lifestyle Med 2011 May;5(3). [doi: 10.1177/1559827610391883] [Medline: 24358034]
11. Cijfers roken. Trimbos-Instituut. URL: https://www.trimbos.nl/kennis/cijfers/roken/ [accessed 2026-06-02]
12. Inzicht in mensen die niet en weinig bewegen in nederland. Rijksinstituut voor Volksgezondheiden Milieu Ministerie van Volksgezondheid,Welzijn en Sport. URL: https://www.rivm.nl/publicaties/inzicht-in-mensen-die-niet-en-weinig-bewegen-in-nederland#:~:text=Een%20deel%20van%20de%20bevolking,beweegt%20(ruim%2053%20procent) [accessed 2026-06-02]
13. Kloosterman R, Akkermans M, Reep C, Aa M. (On)gezonde leefstijl 2022: opvattingen, motieven en gedragingen. Centraal Bureau voor de Statistiek. URL: https://www.cbs.nl/nl-nl/longread/rapportages/2023/on--gezonde-leefstijl-2022-opvattingen-motieven-en-gedragingen/8-roken [accessed 2026-06-02]

XSL•FO
RenderX

14. Al-Dhahir I, Reijnders T, Faber JS, et al. The barriers and facilitators of eHealth-based lifestyle intervention programs for people with a low socioeconomic status: scoping review. J Med Internet Res 2022 Aug 24;24(8):e34229. [doi: 10.2196/34229] [Medline: 36001380]
15. Kunicki ZJ, Hallgren M, Uebelacker LA, Brown RA, Price LH, Abrantes AM. Examining the effect of exercise on the relationship between affect and cravings among smokers engaged in cessation treatment. Addict Behav 2022 Feb;125:107156. [doi: 10.1016/j.addbeh.2021.107156] [Medline: 34710842]
16. Zhou YH, Feng WX, Guo YG, Wu JH. Effect of exercise intervention on smoking cessation: a meta-analysis. Front Physiol 2023 Aug 8;14:37614760. [doi: 10.3389/fphys.2023.1221898]
17. Berkovitch A, Kivity S, Klempfner R, et al. Time-dependent relation between smoking cessation and improved exercise tolerance in apparently healthy middle-age men and women. Eur J Prev Cardiol 2015 Jun;22(6):807-814. [doi: 10.1177/2047487314535116] [Medline: 24817697]
18. West R, Papadakis S. Stop smoking services: increased chances of quitting. National Centre for Smoking Cessation and Training. URL: https://www.ncsct.co.uk/publications/Stop_smoking_services_impact_on_quitting [accessed 2026-06-02]
19. Howlett N, Trivedi D, Troop NA, Chater AM. Are physical activity interventions for healthy inactive adults effective in promoting behavior change and maintenance, and which behavior change techniques are effective? A systematic review and meta-analysis. Transl Behav Med 2019 Jan 1;9(1):147-157. [doi: 10.1093/tbm/iby010] [Medline: 29506209]
20. van der Kleij R, Kasteleyn MJ, Meijer E, et al. SERIES: eHealth in primary care. Part 1: concepts, conditions and challenges. Eur J Gen Pract 2019 Oct;25(4):179-189. [doi: 10.1080/13814788.2019.1658190] [Medline: 31597502]
21. Bendotti H, Lawler S, Chan GCK, Gartner C, Ireland D, Marshall HM. Conversational artificial intelligence interventions to support smoking cessation: a systematic review and meta-analysis. Digit Health 2023;9(20552076231211634):20552076231211634. [doi: 10.1177/20552076231211634] [Medline: 37928336]
22. Keij B, Versluis A, Alblas EE, Keuper JJ, Tuyl LHD, Vaart R. E-healthmonitor 2023: stand van zaken digitale zorg. Nivel. 2024. URL: https://www.nivel.nl/nl/publicatie/e-healthmonitor-2023-stand-van-zaken-digitale-zorg [accessed 2026-06-02]
23. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
24. Versluis A, van Luenen S, Meijer E, et al. SERIES: eHealth in primary care. Part 4: addressing the challenges of implementation. Eur J Gen Pract 2020 Dec;26(1):140-145. [doi: 10.1080/13814788.2020.1826431] [Medline: 33025820]
25. He L, Balaji D, Wiers RW, Antheunis ML, Krahmer E. Effectiveness and acceptability of conversational agents for smoking cessation: a systematic review and meta-analysis. Nicotine Tob Res 2023 Jun 9;25(7):1241-1250. [doi: 10.1093/ntr/ntac281] [Medline: 36507916]
26. Milne-Ives M, de Cock C, Lim E, et al. The effectiveness of artificial intelligence conversational agents in health care: systematic review. J Med Internet Res 2020 Oct 22;22(10):e20346. [doi: 10.2196/20346] [Medline: 33090118]
27. Perski O, Crane D, Beard E, Brown J. Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. Digit Health 2019;5:2055207619880676. [doi: 10.1177/2055207619880676] [Medline: 31620306]
28. He L, Braggaar A, Basar E, Krahmer E, Antheunis M, Wiers R. Exploring user engagement through an interaction lens: what textual cues can tell us about human-chatbot interactions. 2024 Jul 8 Presented at: CUI '24; Jul 8-10, 2024 URL: https://dl.acm.org/doi/proceedings/10.1145/3640794 [doi: 10.1145/3640794.3665536]
29. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
30. Brickwood KJ, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819. [doi: 10.2196/11819] [Medline: 30977740]
31. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23. [doi: 10.2196/mhealth.8873] [Medline: 29343463]
32. Bickmore T, Schulman D, Yin LX. Maintaining engagement in long-term interventions with relational agents. Appl Artif Intell 2010 Jul 1;24(6):648-666. [doi: 10.1080/08839514.2010.492259] [Medline: 21318052]
33. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061. [doi: 10.1136/bmj.n2061] [Medline: 34593508]
34. Suijkerbuijk A, Bruijn A, Stok M, Zorginnovaties JM. Zorginnovaties. een checklist voor succesvolle implementatie. Rijksinstituut voor Volksgezondheiden Milieu Ministerie van Volksgezondheid, Welzijn en Sport. 2022. URL: https://www.rivm.nl/documenten/zorginnovaties-checklist-voor-succesvolle-implementatie [accessed 2026-06-02]
35. van Vliet MHM, van Eersel RA, Poot CC, et al. A practical step-by-step approach for patient and public involvement in eHealth intervention research: lessons learned from three case projects. Internet Interv 2026 Mar;43:100896. [doi: 10.1016/j.invent.2025.100896] [Medline: 41551032]

XSL•FO
RenderX

36. Versluis A, Penfornis KM, van der Burg SA, et al. Targeting key risk factors for cardiovascular disease in at-risk individuals: developing a digital, personalized, and real-time intervention to facilitate smoking cessation and physical activity. JMIR Cardio 2024 Dec 20;8(1):e47730. [doi: 10.2196/47730] [Medline: 39705698]
37. Michie S, Hyder N, Walia A, West R. Development of a taxonomy of behaviour change techniques used in individual behavioural support for smoking cessation. Addict Behav 2011 Apr;36(4):315-319. [doi: 10.1016/j.addbeh.2010.11.016] [Medline: 21215528]
38. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
39. Burke PJ, Stets JE. Identity Theory: Oxford University Press; 2009.
40. West R, Brown J. Theory of Addiction, 2nd edition: Wiley-Blackwell; 2013. [doi: 10.1002/9781118484890]
41. Oyserman D, Elmore K, Smith G. Self, self-concept, and identity. In: Handbook of Self and Identity, 2nd edition: The Guilford Press; 2012:69-104.
42. Marlatt GA, Donovan DM, editors. Relapse Prevention: Maintenance Strategies in the Treatment of Addictive Behaviors: The Guilford press; 2007.
43. van Vliet MHM, Versluis A, Chavannes NH, et al. Protocol of a mixed-methods evaluation of Perfect Fit: a personalized mHealth intervention with a virtual coach to promote smoking cessation and physical activity in adults. Digit Health 2024;10:20552076241300020. [doi: 10.1177/20552076241300020] [Medline: 39640962]
44. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008 Apr;13(2):92-98. [doi: 10.1258/jhsrp.2007.007074] [Medline: 18416914]
45. Poot CC, Meijer E, Fokkema M, Chavannes NH, Osborne RH, Kayser L. Translation, cultural adaptation and validity assessment of the Dutch version of the eHealth Literacy Questionnaire: a mixed-method approach. BMC Public Health 2023 May 30;23(1):1006. [doi: 10.1186/s12889-023-15869-4] [Medline: 37254148]
46. Kayser L, Karnoe A, Furstrand D, et al. A multidimensional tool based on the eHealth Literacy Framework: development and initial validity testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36. [doi: 10.2196/jmir.8371] [Medline: 29434011]
47. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991 Sep;86(9):1119-1127. [doi: 10.1111/j.1360-0443.1991.tb01879.x] [Medline: 1932883]
48. Godin G. The Godin-Shephard Leisure-Time Physical Activity Questionnaire. Health Fit J Can 2011 Jan;4(1):18-22. [doi: 10.14288/hfjc.v4i1.82]
49. Amireault S, Godin G. The Godin-Shephard Leisure-Time Physical Activity Questionnaire: validity evidence supporting its use for classifying healthy adults into active and insufficiently active categories. Percept Mot Skills 2015 Apr;120(2):604-622. [doi: 10.2466/03.27.PMS.120v19x7] [Medline: 25799030]
50. Brooke J. SUS-a quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry: Taylor & Francis; 1996:189-194.
51. Staniszewska S, Brett J, Simera I, et al. GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research. Res Involv Engagem 2017;3(13):13. [doi: 10.1186/s40900-017-0062-2] [Medline: 29062538]
52. Nederlands Huisartsen Genootschap. Cardiovasculair risicomanagement. NHG-Richtlijnen. URL: https://richtlijnen.nhg.org/standaarden/cardiovasculair-risicomanagement [accessed 2026-06-02]
53. Hennink MM, Kaiser BN, Marconi VC. Code saturation versus meaning saturation: how many interviews are enough? Qual Health Res 2017 Mar;27(4):591-608. [doi: 10.1177/1049732316665344] [Medline: 27670770]
54. Forerunner® 55. Garmin Ltd. URL: https://www.garmin.com/en-US/p/741137/ [accessed 2026-06-02]
55. The treatment platform in mental healthcare. NiceDay Nederland. 2024. URL: https://nicedaynederland.nl/en/home-en/ [accessed 2026-02-06]
56. Baccinelli W, van der Burg S, Richardson R, et al. Reusable virtual coach for smoking cessation and physical activity coaching. 2022 Sep 6 Presented at: IVA '22; Sep 6-9, 2022 URL: https://dl.acm.org/doi/proceedings/10.1145/3514197 [accessed 2026-05-06] [doi: 10.1145/3514197.3551252]
57. Burg S, Baccinelli W, Richardson R, et al. PerfectFit. Zenodo. URL: https://doi.org/10.5281/zenodo.10837093 [accessed 2026-06-02]
58. Meijer E, Gebhardt WA, Van Laar C, Kawous R, Beijk S. Socio-economic status in relation to smoking: the role of (expected and desired) social support and quitter identity. Soc Sci Med 2016 Aug;162:41-49. [doi: 10.1016/j.socscimed.2016.06.022] [Medline: 27328056]
59. Hummel K, Candel M, Nagelhout GE, et al. Construct and predictive validity of three measures of intention to quit smoking: findings from the International Tobacco Control (ITC) Netherlands Survey. Nicotine Tob Res 2018 Aug 14;20(9):1101-1108. [doi: 10.1093/ntr/ntx092] [Medline: 28472427]
60. Penfornis KM, Gebhardt WA, Rippe RCA, Van Laar C, van den Putte B, Meijer E. My future-self has (not) quit smoking: an experimental study into the effect of a future-self intervention on smoking-related self-identity constructs. Soc Sci Med 2023 Mar;320:115667. [doi: 10.1016/j.socscimed.2023.115667] [Medline: 36641885]

XSL•FO
RenderX

61. Provoost S, Kleiboer A, Ornelas J, et al. Improving adherence to an online intervention for low mood with a virtual coach: study protocol of a pilot randomized controlled trial. Trials 2020 Oct 16;21(1):860. [doi: 10.1186/s13063-020-04777-2] [Medline: 33066805]
62. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a relational agent by hospital patients with depressive symptoms. Interact Comput 2010 Jul 1;22(4):289-298. [doi: 10.1016/j.intcom.2009.12.001] [Medline: 20628581]
63. Tukey JW. The future of data analysis. Ann Math Statist 1962 Mar;33(1):1-67. [doi: 10.1214/aoms/1177704711]
64. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13(1):117. [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
65. Ritchie J, Lewis J, Nicholls CM, Ormston R. Qualitative Research Practice: A Guide for Social Science Students and Researchers: Sage Publications Ltd; 2013.
66. Electronic data capture (EDC). Castor. 2025. URL: https://www.castoredc.com/electronic-data-capture-system/ [accessed 2026-06-02]
67. Hopewell S, Chan AW, Collins GS, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. Lancet 2025 Apr 14;405(10489):S0140-6736(25)00672-5. [doi: 10.1016/S0140-6736(25)00672-5] [Medline: 40245901]
68. Wat is het onderwijsniveau van nederland? Nederlandin cijfers.: Centraal Bureau voor de Statistiek URL: https://longreads.cbs.nl/nederland-in-cijfers-2024/wat-is-het-onderwijsniveau-van-nederland/ [accessed 2026-06-02]
69. Meneses-Gaya ID, Zuardi AW, Loureiro SR, Crippa JDS. Psychometric properties of the Fagerström Test for nicotine dependence. J Bras Pneumol 2009 Jan;35(1):73-82. [doi: 10.1590/S1806-37132009000100011]
70. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
71. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng XQ, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283. [doi: 10.2196/20283] [Medline: 32990635]
72. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
73. Csikszentmihalyi M. Beyond Boredom and Anxiety: Experiencing Flow in Work and Play: Jossey-Bass; 2000.
74. He LW, Basar E, Wiers RW, Antheunis ML, Krahmer E. Chatting your way to quitting: a longitudinal exploration of smokers' interaction with a cessation chatbot. Internet Interv 2025 Mar;39:100806. [doi: 10.1016/j.invent.2025.100806]
75. Carter DD, Robinson K, Forbes J, Hayes S. Experiences of mobile health in promoting physical activity: a qualitative systematic review and meta-ethnography. PLoS ONE 2018;13(12):e0208759. [doi: 10.1371/journal.pone.0208759] [Medline: 30557396]
76. Almusharraf F, Rose J, Selby P. Engaging unmotivated smokers to move toward quitting: design of motivational interviewing-based chatbot through iterative interactions. J Med Internet Res 2020 Nov 3;22(11):e20251. [doi: 10.2196/20251] [Medline: 33141095]
77. Abroms LC, Yousefi A, Wysota CN, Wu TC, Broniatowski DA. Assessing the adherence of ChatGPT chatbots to public health guidelines for smoking cessation: content analysis. J Med Internet Res 2025 Jan 30;27:e66896. [doi: 10.2196/66896] [Medline: 39883917]
78. Meyer S, Elsweiler D. LLM-based conversational agents for behaviour change support: a randomised controlled trial examining efficacy, safety, and the role of user behaviour. Int J Hum Comput Stud 2025 May;200:103514. [doi: 10.1016/j.ijhcs.2025.103514]
79. Basar E, Balaji D, He L, et al. HyLECA: a framework for developing hybrid long-term engaging controlled conversational agents. 2023 Jul 19 Presented at: CUI '23; Jul 19-21, 2023 URL: https://dl.acm.org/doi/proceedings/10.1145/3571884 [accessed 2026-06-05] [doi: 10.1145/3571884.3604404]
80. O'Brien HL, Cairns P, Hall M. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. Int J Hum Comput Stud 2018 Apr;112:28-39. [doi: 10.1016/j.ijhcs.2018.01.004]
81. Gabriele JM, Carpenter BD, Tate DF, Fisher EB. Directive and nondirective e-coach support for weight loss in overweight adults. Ann Behav Med 2011 Apr;41(2):252-263. [doi: 10.1007/s12160-010-9240-2] [Medline: 21108032]
82. Smit ES, Bol N. From self-reliers to expert-dependents: identifying classes based on health-related need for autonomy and need for external control among mobile users. Media Psychol 2020 May 3;23(3):391-414. [doi: 10.1080/15213269.2019.1604235]
83. Albers N, Melo FS, Neerincx MA, Kudina O, Brinkman WP. Psychological, economic, and ethical factors in human feedback for a chatbot-based smoking cessation intervention. NPJ Digit Med 2025 May 31;8(1):326. [doi: 10.1038/s41746-025-01701-3] [Medline: 40450111]
84. Castro O, Mair JL, Zheng S, et al. The LvL UP trial: protocol for a sequential, multiple assignment, randomised controlled trial to assess the effectiveness of a blended mobile lifestyle intervention. Contemp Clin Trials 2025 Mar;150:107833. [doi: 10.1016/j.cct.2025.107833] [Medline: 39900289]

XSL•FO
RenderX

Abbreviations

eHLQ: eHealth Literacy Questionnaire
FTND: Fagerström Test for Nicotine Dependence
GRAMMS: Good Reporting of a Mixed Methods Study
GRIPP2: revised version of Guidance for Reporting Involvement of Patients and the Public
GSLTPAQ: Godin-Shephard Leisure-Time Physical Activity Questionnaire
HSD: honestly significant difference
mHealth: mobile health
PA: physical activity
PF: Perfect Fit
PPI: patient and public involvement
SEP: socioeconomic position
SMART: specific, measurable, achievable, relevant, and time-bound
SUS: System Usability Scale
WHO: World Health Organization

Edited by E Borycki, A Kushniruk; submitted 03.Sep.2025; peer-reviewed by M Mondellini, X Su; revised version received 14.Apr.2026; accepted 21.Apr.2026; published 14.Jul.2026.

Please cite as:
van Vliet MHM, Meijer E, Albers N, Penfornis KM, Baccinelli W, Scheltinga BL, van Eersel RA, Chavannes NH, Versluis A, The Perfect Fit Consortium
Evaluation of the Feasibility and Acceptability of Perfect Fit, a Virtual Coach–Based mHealth Intervention for Smoking Cessation and Physical Activity in Adults: Mixed Methods Study
JMIR Hum Factors 2026;13:e83456
URL: https://humanfactors.jmir.org/2026/1/e83456
doi:10.2196/83456

© Milon H M van Vliet, Eline Meijer, Nele Albers, Kristell M Penfornis, Walter Baccinelli, Bouke L Scheltinga, Roxy A van Eersel, Niels H Chavannes, Anke Versluis, The Perfect Fit Consortium. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 14.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring Use, Usability, and Functionality in the Iterative Testing of a Co-Designed Preventative Mental Health Smartphone App for University Students: Comparison of Monash Thrive App Development Trials

Melinda McCabe[1], BA Hons, PhD; Michelle Byrne[1], PhD; Judith Gullifer[2], PhD; Isabelle Smart[1], BA Hons; Chunlei Chang[1], PhD; Sanket Singh[1], MA; Gagan Khara[1], MSc; Kim Cornish[1], PhD

[1]Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Level 5, 18 Innovation Walk, Clayton, Victoria, Australia

[2]School of Psychological Sciences, Monash University, Clayton, Victoria, Australia

Corresponding Author:
Melinda McCabe, BA Hons, PhD
Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Level 5, 18 Innovation Walk, Clayton, Victoria, Australia

Abstract

Background: The emergence of digital tools in physical and mental health spaces is a necessary development to meet the growing need for preventative recognition and early care. Such tools must be appropriately tested throughout the development process to ensure that they are fit for purpose and both empirically and clinically valid. However, it is also critical that these tools are developed in an iterative and user-focused manner to ensure that they meet user needs for usability and functionality.

Objective: This study explores the iterative testing of use, usability, and functionality key performance indicators of the Monash Thrive app, a preventative mental health app focusing on mental health literacy and mood check-in features. It was hypothesized that there would be an improvement in all variables in the second iteration testing of the app compared to the initial iteration.

Methods: Two iterative trials of the app were compared, including an initial trial of the minimal viable product (MVP) of the app across 4 weeks of use and a secondary trial of the beta version of the app, redeveloped post-MVP testing, across 12 weeks of use. All recruitment and testing were delivered online through email surveys.

Results: Retention rates for both trial 1 and 2 were low (59% and 46%, respectively). Despite redevelopment and resulting hypotheses, there was no significant improvement observed across any variable when comparing the 2 trials. No difference was identified across any Mobile App Rating Scale subscales, and a significant difference was detected (P=.04; 95% CI −0.59 to 9.35), indicating higher usability scale scores in the MVP app trial compared to the beta.

Conclusions: Results did not support the study hypotheses. Such findings could be due to changing expectations of smartphone apps, relatively low retention throughout the trials, or the lack of a control or comparison group for which to compare ratings. However, these lack of findings demonstrate the core importance of iterative testing to ensure that changes or upgrades are delivering the desired improvement in use, usability, and functionality.

Trial Registration: ANZCTR ACTRN12622001054707; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=383757 and ANZCTR ACTRN12623000131651; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=384450

(JMIR Hum Factors 2026;13:e67818) doi:10.2196/67818

KEYWORDS

mental health; prevention; smartphone app; digital tool; university student; usability; iteration; design; co-design; systems; development

Introduction

Digital Tools in the Mental Health Space

Digital health tools including websites, smartphone apps, telehealth services, and other online or digitally facilitated services have become critical in the evolution of mental health care globally. Recent systematic reviews have highlighted the emergence and testing of digital health tools used in areas such as substance use disorders, mood and trauma-related disorders, and a wide range of other mental health conditions [1]. Although still a new and emerging technology in the mental health space,

XSL•FO
RenderX

websites, apps, and other digital tools are the clear future steps forward to assist in the availability, accessibility, and scalability of mental health support. These benefits are clear in preventative mental health care, which describes the digital tools, specifically websites and apps, targeting and tailored to individuals who may be struggling with their mental health and well-being without having or necessarily meeting the criteria for a mental health diagnosis. Preventative tools and care can be designed for and deployed among the broader population with the overarching goal of promoting awareness, recognition, and early help-seeking well before clinical symptoms emerge. Many such apps have been developed and tested in a clinical and research setting, focusing on preventative care or support care [2-9]. Although there are differing levels of efficacy and effectiveness across these tools depending on core outcomes and use rates, almost all of these apps are either not available for widespread use or have been removed from relevant app stores. In the mental health space, efficacy is critical and requires a strict testing protocol to ensure that users are capable of benefiting from the tool. However, this is not enough to ensure that these tools are available for ongoing use and can evolve to meet the needs of the users over time.

Efficacy and Usability

Efficacy of preventative mental health tools, particularly those tools developed from a clinical or research-based setting, is often measured during development through structured research testing. In the mental health space, such testing often uses clinically validated measures of stress, well-being, and other measures to assess whether use of the app has caused a measurable change in mental health symptoms after a period of use [4-14]. Although such scales and the focus on mental health outcomes are necessary to ensure that the digital tool being developed improves the health and mental health outcomes being targeted, efficacy measured through mental health outcomes is not able to account for elements critical to uptake and use. Usability must also be a core focus of tool testing and design, specifically usability in terms of ease of use and how intuitive the tool is. Usability also includes the burden or time cost of the user to effectively engage with the tool (ie, how long does it take to learn how to use) and the ability of the tool to effectively deliver on its intended functionality. Such elements of design and delivery are critical to ensure the uptake and ongoing use of the app required to make sure that the intended user base interacts with the tool long enough to see mental health improvements outside the confines of clinical trial or research settings.

Including Critical Design Stages in Research

The focus on usability is a well-established and accepted foundation of product design in industry. The importance of usability, to ensure a product can be used effectively and efficiently by the target user, can be the difference between the success or failure of any new product or smartphone app, regardless of how effective the tool may be. In health and mental health spaces, the focus on usability can be overlooked or seen as secondary to the objective efficacy or evidence base of a product; however, if the use of the tool is dependent on a user-driven uptake and ongoing use, it is critical that the usability, engagement, functionality, and efficacy are all considered key performance indicators. Such indicators can be measured and identified by specific tools including the System Usability Scale (SUS) [15-18] and the Mobile App Rating Scale (MARS) [19-25]. These scales demonstrate how distinct and interconnected each of these usability measures is. The MARS contains 5 subscales, exploring a wide range of elements contributing to app quality. These subscales include engagement (how entertaining and interesting the app is), functionality (performance and ease of use), aesthetics (quality of layout and graphics), information (quality and quantity of information), and subjective quality (user use, recommendations, or desire to pay for use). The SUS is a validated scale that can be used to assess a wide variety of tools, not just smartphone apps. Originally developed in 1996 [15], this tool is still relevant, exploring whether participants would use the tool regularly, if they would need support or training to use the tool, and if they felt confident using and navigating the system or tool. Improving usability is an ongoing process that takes place in both initial and ongoing development phases. Such development processes require iterative stages of product development that rely upon user testing and feedback throughout the development process, including ongoing testing with final release products to ensure a product evolution that can adapt to meet user needs. Many theories and processes exist to outline these necessary steps; however, one of the most recognizable in literature is the user-centered systems design (UCSD) [26]. The UCSD process highlights the importance of an agile, user-focused, and iterative design process. Using an iterative and circular structure of (1) vision and plan, (2) analyze user needs, (3) design for usability with prototypes, (4) construct and deploy in iterations, (5) evaluate with users in context, and (6) feedback and integration into the next iteration, cycling back into stage 1 to repeat the process. Two of the greatest challenges throughout this process are the ongoing iterative testing and development and ensuring the focus on user-focused testing and the importance of user feedback. Iterative testing and subsequent redevelopment require substantial time and financial commitments, often better used in environments that have ongoing funding and continuous development goals. This may be why iteration is often given in commercial settings, but can often be underused in academic or health settings where funding may be limited by short-term grants. Iterative testing and development is not a new concept in literature and has been used in app testing [27,28]; however, the use and reporting of iterative testing in academia are inconsistent, where some studies refer to iterative testing as stages of testing of the same tool from different perspectives [3,29]. A focus on the end user during development is more common and is continuing to grow as a new standard for intervention testing and can be referred to as co-design or consultation. Some inconsistency remains when defining how end users are included in the design process.

Potential to Develop Engaging Tools Using Evidence

The term “co-design” can mean many things in the fields of research, development, and health services [30,31]. For many research-based and academic-focused teams working on a co-design process, this means the consultation and feedback gathering from experts in the field, such as qualified health care

XSL•FO
RenderX

providers or implementation experts or those with lived experience such as caregivers or advocates. Although consultation with these expert bodies is important, particularly in a health setting, it does not consider the needs of the individual user from the experience of the individual user. This means the co-design and design consultation happening in these academic or clinical settings can bias the opinions of a small number of experts rather than potential users who are themselves experts in lived experience and are better equipped to inform developers about their own desired functionality or barriers to use. It is this information that can be critical to user uptake, endorsement, and ongoing use. In a broader industry setting, outside the world of academia, co-design is the focus and importance of end-user consultation at iterative stages of development to inform and guide planning and final decision-making for the end product. The landscape of how these digital tools are being developed in medical and academic settings is evolving with more studies emerging, demonstrating both user-focused co-design and iterative design process [32,33]. A core foundation of these steps toward the future of health and mental health digital tool development in academic and medical settings is the incorporation of scales and key performance indicators, such as those measured by MARS or SUS, that focus solely on the user experience, engagement, and likelihood of further use and success of the tool, in addition to the health or clinical scales used to measure changes in participants over time. The inclusion of scales such as the MARS and SUS will ensure that researchers alongside developers are focusing on both the efficacy of the tool in terms of their intervention of interest (mental health outcomes) and also whether the tool is a usable, engaging, and effective way of delivering their interventions. It is critical to ensure that not only is a health or mental health tool founded in strong evidence, but that it is also usable to the degree that an individual user would willingly use the tool long enough to see benefit outside the extrinsic motivators of a research trial, clinical oversight, or other kind of incentive.

This Research

The iterative testing of product usability and functionality is a fundamental element of all product development and testing. In the digital mental health space, app testing often focuses on mental health and health-related outcomes, focusing on clinical or research efficacy. The inclusion of design and usability-focused measures is slowly being incorporated into such development and testing; however, the importance of iterative testing is still underreported in the literature. The Monash Thrive app is a preventative mental health app designed and developed by a team of both mental health researchers and educators and experts in user experience and systems design. The design process of the Monash Thrive app followed the iterative and user-focused steps highlighted in the UCSD process. The focus of the app is to promote mental health awareness and literacy, the early recognition of mental health concerns, and early help-seeking behaviors of university students. The Monash Thrive app was designed to deliver a range of mental health literacy tools and promote preventative mental health behaviors such as journaling [34] and mindfulness [35]. At the time of development (2020 - 2022), no single app or tool was freely available to university students, which allowed users to test and incorporate a range of positive mental health tools. The aim of this study is to explore the user-focused testing, design, and usability key performance indicators of a preventative mental health app designed with and for a university student population across 2 iterative app testing trials using both the initial minimum viable product (MVP) and the second iteration (beta) app prototypes. It is hypothesized that the trial of the second iteration of the Monash Thrive app (beta) will show improvements in all use, usability, and functionality outcome variables when compared to the initial iteration and trial (MVP) of the Monash Thrive app.

Methods

Participants

Trial 1

Participants were recruited using an online recruitment and eligibility survey that was open for a period of 13 weeks. This survey included an explanatory statement and consent form. All eligible participants who had responded to this survey by the close date were recruited and provided with details to download and install the Monash Thrive app. All respondents who met the eligibility criteria and responded before the start deadline were then recruited. The inclusion criteria for trial 1 required participants to be currently enrolled at Monash University at the time of participation, to be located in Australia due to app accessibility, and to have a working smartphone. Surveys including items not included in this study, app use, feedback, and bug reporting were sent to participants weekly. The items of interest in this study were included in the week 4 (of 4) surveys. Initially, 135 participants were recruited throughout the 4 weeks of active app use and 6 weeks of survey testing of trial 1. From the initial 135 participants, only 80 completed the survey containing the core items of interest; 1 participant did not complete all items of interest on the MARS and did not complete the SUS. A total of 79 (59%) participants completed all items of interest in this study and are therefore included in the following analysis. All 79 participants were located in Australia, Victoria, at the time of testing. Approximately 70% (55/79) of these participants were international students, 79% (62/79) identified as female, and 58% (46/79) of the sample were in their first year of study. The average age of the participants was 21 (SD 4.96) years, and the age range of the final sample was 18 to 50 years. Across users, 63% (50/79) of participants used an iPhone, and the remaining participants used a range of Android devices including Samsung, OnePlus, and Redmi.

Trial 2

Similar to trial 1, an online recruitment and eligibility survey (including explanatory statement and consent form) was open for 11 weeks, and all eligible participants were contacted after survey close with details to download and install the Monash Thrive app. Participants from trial 1 who indicated that they would be interested in taking part in another trial were also contacted with an invitation to complete the eligibility survey. Similar to trial 1, all respondents to the eligibility survey who

XSL•FO
RenderX

met the inclusion criteria were recruited and onboarded. Trial 2 included 5 waves of recruitment, where new responses to the onboarding survey were included in the next wave of recruitment until the deadline. The same inclusion criteria were used as in trial 1; however, for the final analysis, those who completed trial 1 were removed from the trial 2 data in this study. New surveys were delivered to participants fortnightly, measuring nonreported items of interest and app use, feedback, and bug reporting. The items of interest in this study were included in the week 10 (of 12) surveys. Initial recruitment for trial 2 included 100 participants; by the end of the trial, 46 participants completed all items of interest across the 12 weeks of testing. From the initial 100 participants, only 49 participants completed the survey including the 2 core items of interest; of these 49, 3 did not complete all the items of MARS. Of the 46 (46%) remaining participants included in the final analysis for trial 2, 74% identified as female, 24% as male, and the remaining 2% did not identify with the listed options. Of the final sample, 38 were international students (83%), and just under half (46%) were in their first year of study at the time of the trial, and all but 1 participant was located in Australia at the time of recruitment. Of the final sample for trial 2, 59% used an Apple smartphone, and the remaining used a variety of Android devices. The average age of the sample was 23 (SD 5.86) years.

Measures

Monash Thrive App

The Monash Thrive app is a co-designed preventative mental health app that has been developed, tested, and deployed by a team of both mental health researchers and educators, alongside expert app developers and designers to ensure that it met both the expectations of the market, users, and industry, and was able to meet the efficacy standards of empirical testing. The app focuses on increasing the users' personal awareness of not only their own mental health, but also their understanding of mental health and psychology as a field. The overarching goal of the app is to improve awareness and increase rates of early help-seeking for mental health concerns well before the onset of clinically significant symptoms.

The first functional iteration of the Monash Thrive app, the MVP, was tested among participants in trial 1. The core functionality of the MVP was based on a multistage mood check-in including a diary entry. Based on users' "check-ins," they were then recommended mental health literacy resources as a short-form "blog" that can be read by either flipping through a series of cards or in long form. The entire series of resources was also available for users to explore. Alongside these interactive tools, the MVP also included a character "buddy," Moody, who sent reminders and responded to mood check-ins. The Monash Thrive app also included reminders to engage with the app; these were set at 9 AM and 6 PM daily, and these reminders could be blocked using phone settings. The MVP iteration of the Monash Thrive app was designed and built based on a combination of market research focusing on the core functionality of the most popular mental health and well-being focus smartphone apps and informal community-based interviewing where the lead designers and developers engaged with students in community settings to informally discuss app use and desired characteristics and features. From this, the core elements of a mood check-in and accompanying student-written mental health resources (peer-developed) formed the basis of the Monash Thrive app and therefore the minimal viable necessary features. The second iteration of the app, tested in the second trial, included all the same functionality as the MVP with the addition of a personalized analytics page and calendar function that allowed users to reflect on their mood check-ins over time. The second iteration of the app also allowed users to customize or change reminder notifications in the app. New developments were based on user feedback from both the qualitative and quantitative survey data and informal user focus groups via Zoom (Zoom Video Communications) data collected in trial 1. Example screenshots of the described functionality can be found in Multimedia Appendix 1. User feedback and market research formed the basis for feature planning and development. Throughout both trials, the Monash Thrive app was only available for download for participants of each trial through software testing platforms (eg, TestFlight), and the app was not available publicly through app stores.

App Quality

The MARS [19] was used to measure the quality and functionality of the first and second iterations of the Monash Thrive app. The MARS is an extensive 5-point Likert scale with 29 items, exploring engagement (5 items), functionality (4 items), aesthetics (3 items), information (7 items), subjective quality (4 items), and an app-specific score (6 items), as well as a total score. All items have individually coded responses ranging from 1=negative to 5=positive. The MARS also includes 6 items focusing on app-specific perceived impact of the digital tool, and these items are tailored to the app in question and are answered using a 5-point Likert scale of 1=strongly disagree to 5=strongly agree. The MARS is scored by calculating the mean score across all items within each subsection of the scale. Initial development of the MARS app demonstrated strong reliability (α=.90) [19]. For this study, 4 items were removed from the MARS total scale (accuracy, goals, visual information, and evidence base), as they were not relevant to the initial version of the app being tested. Therefore, the information subscore and total score will not include these 4 items and will be calculated using the average of the remaining items. The reliability of the MARS items (total) used in this study was α=.93 for trial 1 and α=.92 for trial 2.

Usability

The SUS [15] was used to measure the usability of the Monash Thrive app in trial 1 and trial 2. The SUS comprises 10 items scored on a 5-point Likert scale of 0=strongly disagree to 4=strongly agree. All even-numbered items are negatively worded and are therefore reverse-scored before all items are summated and multiplied by 2.5 to allow for a final possible score range of 0 - 100. Previous analysis of digital health apps demonstrated a benchmark SUS score of 68 (SD 12.5), which could be considered a "good" score [36]. A 2018 study found that the SUS was still being used by up to 43% of industry usability studies [17] and has consistently been cited as a scale easily understood by a variety of users (researchers, developers, users, and project managers) [37].

XSL•FO
RenderX

App Use

The Monash Thrive app was designed to not collect and store user data, including automatically collected use and engagement analytics. This was to ensure complete privacy and safety of user mental health–related information such as mood and emotion check-ins and diary entries. As such, to measure app use, a short series of self-report items was created by researchers to assess average use in minutes over the course of the previous 7 days to measure and track the average time (minutes per day) spent using the app over the course of both trials. These items listed the day and date of the previous 7 days and provided participants with a sliding scale from 0 to 60 minutes each day in which to respond. Participants were instructed that if the app was used for more than 60 minutes to just indicate 60 minutes of use. Given the anticipated time of use per day was between 5 and 10 minutes, researchers did not anticipate any issues caused by a reported use ceiling. Participant average scores were calculated as the average of at least 3 inputs from participants. If participants did not input at least 3 responses, they did not have an average use score and were therefore not included in the relevant analysis.

Procedure

Trial 1 included 6 weeks of testing across 4 weeks of app use (4-week surveys followed by 2 weeks of informal focus group testing). Expression of interest, mental health variables, and demographics were collected prior to app use. Based on these initial responses, participants were then invited to participate and guided through the download and installation process by using step-by-step manuals specially designed for iOS and Android devices. Once onboarding was complete, participants received a follow-up email each week containing surveys on use and app feedback (each week), mental health outcomes (weeks 2 and 4), and app quality and usability (week 4). This was then followed by informal focus groups to further explore user feedback through open discussion via Zoom. The design of these focus groups followed a market research model, and due to the unstructured nature of the discussion, it was not included in this analysis. Trial 2 included 12 weeks of testing and app use, with demographic, onboarding, and baseline mental health variables collected prior to testing. During the testing period, participants were surveyed every 2 weeks with use and app feedback (weeks 2, 4, 6, 8, 10, and 12), mental health battery (weeks 6 and 12), and app quality and usability (week 10). All surveys included a reminder for any participant who had not completed the survey after 1 week. All surveys and reminders were self-reported measures and sent to participants via email and completed online in their own time. Convenience sampling was used across both trials 1 and 2, and advertisements were used to invite participants to participate, both online and in targeted student locations. Participants in trial 1 were also invited to take part in trial 2 if they opted in for recontact. No participants who took part in trial 1 were included in the findings for trial 2.

Ethical Considerations

Across both trials, participants were reimbursed with an online gift card sent at the end of the trial. For each survey the participants completed, they added Aus $10 to the final amount they would be reimbursed upon the conclusion of the trial. Participants in trial 1 were also reimbursed an additional Aus $20 for participating in the informal focus groups. Participant payments for trial 1 were completed in May 2022, where the average exchange rate from the Australian dollar to the US dollar was US $0.705. Payments for trial 2 were completed in December 2022, where the average exchange rate from the Australian dollar to the US dollar was US $0.6745. Participation in both trials was voluntary and provided a full explanation of the study via explanatory statements and informed consent as part of the recruitment survey. Any and all potentially identifiable data were removed from this data analysis or displayed in an aggregate form to ensure the privacy of the participants. Both trial 1 (ID# 30475) and trial 2 (ID# 32761) were approved by the Monash University Human Ethics Committee. Both trial 1 (ACTRN12622001054707) and trial 2 (ACTRN12623000131651) were registered with the Australian and New Zealand Clinical Trials Registry prior to testing.

Results

All data analysis was completed using SPSS Statistics (IBM Corp). A series of 1-tailed *t* test analyses (α=.05) was used to assess the differences between the app design and quality across trial 1 and trial 2. Normality was assumed based on kurtosis and skewness variables (<1.96) aside from the average time of use (kurtosis=2.28). No participants who were included in the sample of trial 1 were also included in trial 2. Homogeneity of variances was assumed for all variables aside from MARS functionality and MARS aesthetics based on the Levene test for equality of variances (*P*>.05). As such, average time of use, MARS functionality, and MARS aesthetics were analyzed using the Welch *t* test [38].

Descriptive statistics for all analyses can be found in Table 1. All inferential statistics can be found in Table 2.

For all variables, significance measured via *P* values was only reported for app usability when comparing both trials. However, this difference was in a direction opposite to what was hypothesized, and as such, the null hypothesis was retained. This finding is also supported by the reported CI.

XSL•FO
RenderX

Table . Descriptive statistics.

	Trial 1 (n=79), mean (SD)	Trial 2 (n=49), mean (SD)
App use (minutes)	5.59 (2.91)	4.94 (3.82)
Engagement	3.44 (0.62)	3.44 (0.62)
Functionality	4.23 (0.61)	4.17 (0.53)
Aesthetics	4.17 (0.65)	4.14 (0.53)
Information	3.88 (0.72)	3.93 (0.64)
Subjective quality	2.91 (0.67)	3.03 (0.74)
App utility	3.88 (0.73)	3.97 (0.60)
MARS[a] total score	3.73 (0.50)	3.76 (0.47)
App usability (SUS[b])	79.23 (12.85)	74.89 (15.45)

[a]MARS: Mobile App Rating Scale.

[b]SUS: System Usability Scale.

Table . One-tailed *t* test outcome statistics across trials.

	95% CI	*t* test (*df*)	*P* value
App use (minutes)	−0.94 to 1.05	0.10 (192)	.46
Engagement	−0.27 to 0.17	−0.45 (127)	.33
Functionality	−0.16 to 0.24	0.40 (114.93)	.35
Aesthetics	−0.18 to 0.22	0.19 (115.77)	.43
Information	−0.29 to 0.21	−0.34 (126)	.37
Subjective quality	−0.37 to 0.12	−1.0 (126)	.16
App utility	−0.34 to 0.13	1.74 (126)	.18
MARS[a] total score	−0.22 to 0.14	−0.46 (123)	.32
App usability (SUS[b])	−0.59 to 9.35	1.74 (126)	.04

[a]MARS: Mobile App Rating Scale.

[b]SUS: System Usability Scale.

Discussion

This study explored design and functionality key performance indicators of a preventative mental health app (Monash Thrive app) across 2 stages of iterative testing, an MVP trial and a beta trial after a redesign and development process based on the MVP testing. Key variables of interest across this study included average app use, system usability, and various functionality and perceived app quality markers outlined by the MARS including engagement, functionality, aesthetics, quality of information, subjective quality, and app utility. When comparing app system usability and functionality markers across trials, a significant difference was only identified for app usability across trials; however, this significant difference demonstrated a decline in system usability across trials, the opposite of what was predicted in the study hypotheses.

Although there was no improvement across trials, the average scores of the MARS and the MARS individual domains for both trials indicate a positive response for each of the items within each domain. The strongest domains for both trials were app functionality, exploring how accurate, quick, easy to learn, and intuitive the app is, and app aesthetics, exploring size and arrangement of visual elements, resolution and quality of visual elements, and how subjectively "good" the app looks. This was followed by the quality of information provided within the app (information), app utility and efficacy based on target outcomes (app utility), how entertaining, interesting, customizable, and engaging the app is (engagement), and finally, subjective rating of the app and recommendation to others (subjective app quality). These scores are lower than similar ratings of health apps completed by researchers or health experts [22,39-43]; however, there is much fluctuation within these published average scores (ranging from 3.08 to 4.2 for research reporting mean total scores). Understanding how these scores relate to other studies that used the scale with a user-based sample is limited.

The lack of change in average scores across time for all MARS variables does not support the previously outlined hypotheses. These findings do highlight the critical importance of ongoing testing throughout the iterative development stages, where initial findings in the first app trial indicated strong scores; however, the implementation of suggested changes and retesting of the beta iteration of the app demonstrated similar average scores

XSL•FO
RenderX

on the MARS variables. Without iterative testing, developers may infer that changes based on user feedback must improve all design outcomes; however, without ongoing testing, this cannot be assumed.

Similar to the findings with MARS outcomes, the usability score from the SUS demonstrates strong results both in trial 1 and trial 2 when compared to similar health app testing [36]. When initially comparing the 2 trials directly, there was a significant difference between the 2; however, this was not supported by the accompanying CI outcomes of the same test, and the difference was in the opposite direction to what was hypothesized. Much like the findings for the MARS, these unanticipated differences are evidence that the ongoing testing of iterations throughout development is critical to understand how changes, even if recommended by users in previous testing, may not improve system usability. In this study, the changes across app iterations increased functionality, with the introduction of personalized mood analytics and additional steps available to review mood check-ins over time through the new calendar view function. These additional features, though based on feedback from users, may have increased the complexity of the app without the inclusion of necessary changes to ensure that engaging with these new features is intuitive to users. These findings may have also been impacted by ever-changing expectations and rising standards of digital tools, a greater expectation of a tool known to be the "second iteration" of a tool, or high expectations from a tool associated with their university. Although no users in trial 1 also responded in trial 2, it is possible that expectations of mental health digital tools evolved between the 2 trials.

Across both trials, the average daily use of the app, reported by participants, was approximately 5 minutes (trial 1: mean 6, SD 3 minutes; trial 2: mean 5, SD 3 minutes). This reported use matches the intended daily use of the app, where daily interactions should consist of a mood check and engaging with a mental health resource based on their mood check-in. The beta app also provides users the ability to review their own mood check-ins over time; however, this was unlikely to add much time to the interactions.

This study, as a comparison of 2 iterative trials, is an example of the importance of iterative testing across key design and performance indicators. Although iteration is not uncommon in an app design process, it is not often reported or explored through literature in the same manner as we might see single trial or mental health outcomes tested. By systematically comparing all variables of interest across app iterations, developers can ensure that changes and new developments are improving the app from the user's perspective and experience. Although these findings represent one of the first empirical comparisons of outcomes across iterations of the same app or tool, there are limitations in the design and interpretation of these findings. The varying length of each trial and the difference in time frame are worth noting when interpreting the findings. Trial 1 extended over a 4-week period of active app use, where the items of interest were collected after 4 weeks of use, and trial 2 extended over 12 weeks of use, where the items of interest were collected at 10 weeks of use. Although not previously highlighted in any of the scales used in this study, there may have been some bias and changes in user response due to the differences in time spent using the app. Both trials could have been improved by the use of a more structured trial protocol such as the inclusion of a control group, blinding across groups, the inclusion of another intervention for comparison (eg, a nondigital intervention or another popular app), and the collection of additional relevant demographic information such as digital literacy, typical phone use, and socioeconomic status. A high number of international students when comparing trial 1 and trial 2 should also be noted (70% and 83%, respectively); this may be due to the advertisement of the Monash Thrive app testing trials, where promotion on site and online was more heavily promoted in international student spaces. This skew toward international students may impact findings if participants experienced any language or cultural barriers to engaging in the mental health content of the app that was primarily built from Western research and principles. Future research may consider the inclusion of more culture-focused survey items such as qualitative open-ended questions to better understand how their digital tool may be interpreted and endorsed across a range of cultural and linguistic backgrounds. This study also demonstrates a significant and likely influential attrition rate of participants over time, where 42% and 46% of participants did not continue responding to surveys from trial 1 by week 4 and trial 2 by week 10, respectively. Although attrition is not uncommon in app testing trials [2-9], the drop in participation over time is critical to consider when discussing the usability of an app, where the discontinued completion of surveys likely reflects a discontinued use of the app despite the monetary incentive provided by each trial. Future studies may consider including attrition as a measure itself to reflect usability and continued use. Future studies may also benefit from the incorporation of objective measures of use and engagement. The app tested in this study was designed to not collect and send user data to protect user privacy; as such, all outcomes were collected through separate surveys including daily use, which required self-reporting and was collected once a week (for each day in the previous week) in trial 1 and once every 2 weeks (for each day in the previous 2 weeks) in trial 2. This does leave considerable margin for error in participant recollection and self-reporting of their use. Overall, any replication or similar research should undergo considerable improvement in trial structure, similar to that of randomized controlled trials.

In conclusion, this study is an example of the importance of incorporating design-, usability-, and functionality-based key performance indicators to app testing in a research and clinical setting. These outcomes provide the insight necessary to ensure that research-based and clinical digital tools in both the health and mental health space are not only found in clinical evidence but also usable and engaging enough for users to ensure the uptake and ongoing use necessary to see benefits. This study also highlights the importance of iterative testing throughout the design process, where developers may assume that the next iteration of design would demonstrate better usability and functionality scores; however, current findings demonstrate that even when changes are based on user feedback and are designed to increase app usability, such improvements may not be reflected in user testing. Iterative testing is important to ensure

XSL•FO
RenderX

that app development is moving in the right direction and developers are not just assuming improvement in app usability and reception alongside the addition or changes in the future. This process is the foundation of a user-centered design process [26], which outlines the importance of continued testing and feedback incorporation as the tool evolves. The process of developing and testing these mental health and health digital tools is time-consuming and costly; however, none so more costly than the final development and deployment of a tool just to see low uptake and poor retention rates, resulting often in the removal of the app from circulation; as such, it is critically important that all relevant key performance indicators are included in testing of digital tools, iteratively, as they are being developed.

Acknowledgments

The Monash Thrive app development and testing were part of a collaboration between Monash University, the Monash Thrive team, and Allianz Care Australia. The authors are very grateful to their collaborators and were proud to develop and deliver this tool to all Monash staff and students.

Funding

Funding was received from Allianz Care Australia as part of a Research Agreement, focused on developing preventative mental health support tools for Monash University international students.

Authors' Contributions

MM contributed to trial conceptualization, implementation and completion, analysis, and reporting. MM, IS, CC, SS, and GK were all developers of both the content for the Monash Thrive app (MM and IS) and the development of the backend (CC and GK) and frontend (SS and GK) of the Monash Thrive app. MM and KC contributed to the funding acquisition. KC, JG, and MB provided leadership and review of this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshot images of the core functionality of the Monash Thrive app including (A) the home page with app character “Moody” and recommended mental health resources, (B) mood check known as “wheel of feels,” (C) second stage mood check-in including journal entry, (D) mental health resources library, (E) example of a mental health resource including short form (card swipe) and long form below, and (F) personalized mood analytics with calendar diary icon to review mood check-in timeline (available in app beta, trial 2).
[DOCX File, 3819 KB - humanfactors_v13i1e67818_app1.docx]

References

1. Philippe TJ, Sikder N, Jackson A, et al. Digital health interventions for delivery of mental health care: systematic and comprehensive meta-review. JMIR Ment Health 2022 May 12;9(5):e35159. [doi: 10.2196/35159] [Medline: 35551058]
2. González-García M, Álvarez JC, Pérez EZ, Fernandez-Carriba S, López JG. Feasibility of a brief online mindfulness and compassion-based intervention to promote mental health among university students during the COVID-19 pandemic. Mindfulness (N Y) 2021;12(7):1685-1695. [doi: 10.1007/s12671-021-01632-6] [Medline: 34025814]
3. Øksnebjerg L, Woods B, Waldemar G. Designing the ReACT app to support self-management of people with dementia: an iterative user-involving process. Gerontology 2019;65(6):673-685. [doi: 10.1159/000500445] [Medline: 31266031]
4. Bakker D, Kazantzis N, Rickwood D, Rickard N. A randomized controlled trial of three smartphone apps for enhancing public mental health. Behav Res Ther 2018 Oct;109:75-83. [doi: 10.1016/j.brat.2018.08.003] [Medline: 30125790]
5. Deady M, Glozier N, Calvo R, et al. Preventing depression using a smartphone app: a randomized controlled trial. Psychol Med 2022 Feb;52(3):457-466. [doi: 10.1017/S0033291720002081] [Medline: 32624013]
6. Lee RA, Jung ME. Evaluation of an mHealth app (DeStressify) on university students’ mental health: pilot trial. JMIR Ment Health 2018 Jan 23;5(1):e2. [doi: 10.2196/mental.8324] [Medline: 29362209]
7. Shand F, Mackinnon A, O’Moore K, et al. The iBobbly Aboriginal and Torres Strait Islander app project: study protocol for a randomised controlled trial. Trials 2019 Apr 5;20(1):198. [doi: 10.1186/s13063-019-3262-2] [Medline: 30953556]
8. Tighe J, Shand F, Ridani R, Mackinnon A, De La Mata N, Christensen H. Ibobbly mobile health intervention for suicide prevention in Australian Indigenous youth: a pilot randomised controlled trial. BMJ Open 2017 Jan 27;7(1):e013518. [doi: 10.1136/bmjopen-2016-013518] [Medline: 28132007]

XSL•FO
RenderX

9. Wiljer D, Shi J, Lo B, et al. Effects of a mobile and web app (Thought Spot) on mental health help-seeking among college and university students: randomized controlled trial. J Med Internet Res 2020 Oct 30;22(10):e20790. [doi: 10.2196/20790] [Medline: 33124984]
10. Catuara-Solarz S, Skorulski B, Estella-Aguerri I, et al. The efficacy of "Foundations," a digital mental health app to improve mental well-being during COVID-19: proof-of-principle randomized controlled trial. JMIR Mhealth Uhealth 2022 Jul 1;10(7):e30976. [doi: 10.2196/30976] [Medline: 34978535]
11. Deady M, Collins DAJ, Glozier N, et al. Naturalistic evaluation of HeadGear: a smartphone app to reduce depressive symptoms in workers. Behav Ther 2025 May;56(3):529-542. [doi: 10.1016/j.beth.2024.01.001] [Medline: 40287181]
12. Kenny R, Fitzgerald A, Segurado R, Dooley B. Is there an app for that? A cluster randomised controlled trial of a mobile app-based mental health intervention. Health Informatics J 2020 Sep;26(3):1538-1559. [doi: 10.1177/1460458219884195] [Medline: 31702409]
13. Litvin S, Saunders R, Jefferies P, Seely H, Pössel P, Lüttke S. The impact of a gamified mobile mental health app (eQuoo) on resilience and mental health in a student population: large-scale randomized controlled trial. JMIR Ment Health 2023 Jul 21;10(1):e47285. [doi: 10.2196/47285] [Medline: 37477955]
14. Six SG, Byrne KA, Aly H, Harris MW. The effect of mental health app customization on depressive symptoms in college students: randomized controlled trial. JMIR Ment Health 2022 Aug 9;9(8):e39516. [doi: 10.2196/39516] [Medline: 35943788]
15. Brooke J. SUS: a "quick and dirty" usability scale. In: Usability Evaluation In Industry: CRC Press; 1996.
16. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum–Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
17. Lewis JR. The System Usability Scale: past, present, and future. Int J Hum–Comput Interact 2018 Jul 3;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
18. Kaya A, Ozturk R, Altin Gumussoy C. Usability measurement of mobile applications with System Usability Scale (SUS). In: Calisir F, Cevikcan E, Camgoz Akdag H, editors. Industrial Engineering in the Big Data Era: Springer International Publishing; 2019:389-400. [doi: 10.1007/978-3-030-03317-0_32]
19. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile App Rating Scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27. [doi: 10.2196/mhealth.3422] [Medline: 25760773]
20. Adam A, Hellig JC, Perera M, Bolton D, Lawrentschuk N. "Prostate Cancer Risk Calculator" mobile applications (Apps): a systematic review and scoring using the validated user version of the Mobile Application Rating Scale (uMARS). World J Urol 2018 Apr;36(4):565-573. [doi: 10.1007/s00345-017-2150-1] [Medline: 29222595]
21. Amor-García M, Collado-Borrell R, Escudero-Vilaplana V, et al. Assessing apps for patients with genitourinary tumors using the Mobile Application Rating Scale (MARS): systematic search in app stores and content analysis. JMIR Mhealth Uhealth 2020 Jul 23;8(7):e17609. [doi: 10.2196/17609] [Medline: 32706737]
22. Cheah KJ, Abdul Manaf Z, Fitri Mat Ludin A, Razalli NH, Mohd Mokhtar N, Md Ali SH. Mobile apps for common noncommunicable disease management: systematic search in app stores and evaluation using the Mobile App Rating Scale. JMIR Mhealth Uhealth 2024 Mar 12;12(1):e49055. [doi: 10.2196/49055] [Medline: 38532298]
23. Hatem S, Long JC, Best S, Fehlberg Z, Nic Giolla Easpaig B, Braithwaite J. Mobile apps for people with rare diseases: review and quality assessment using Mobile App Rating Scale. J Med Internet Res 2022 Jul 26;24(7):e36691. [doi: 10.2196/36691] [Medline: 35881435]
24. Kim BY, Sharafoddini A, Tran N, Wen EY, Lee J. Consumer mobile apps for potential drug-drug interaction check: systematic review and content analysis using the Mobile App Rating Scale (MARS). JMIR Mhealth Uhealth 2018 Mar 28;6(3):e74. [doi: 10.2196/mhealth.8613] [Medline: 29592848]
25. LeBeau K, Huey LG, Hart M. Assessing the quality of mobile apps used by occupational therapists: evaluation using the user version of the Mobile Application Rating Scale. JMIR Mhealth Uhealth 2019 May 1;7(5):e13019. [doi: 10.2196/13019] [Medline: 31066712]
26. Gulliksen J, Göransson B, Boivie I, Blomkvist S, Persson J, Cajander Å. Key principles for user-centred systems design. Behav Inf Technol 2003 Nov;22(6):397-409. [doi: 10.1080/01449290310001624329]
27. Oliveira ER, Branco AC, Carvalho D, et al. An iterative process for the evaluation of a mobile application prototype. SN Comput Sci 2022 Jul;3(4):262. [doi: 10.1007/s42979-022-01153-6]
28. Jibb LA, Cafazzo JA, Nathan PC, et al. Development of a mHealth real-time pain self-management app for adolescents with cancer: an iterative usability testing study [Formula: see text]. J Pediatr Oncol Nurs 2017;34(4):283-294. [doi: 10.1177/1043454217697022] [Medline: 28376666]
29. Fairman AD, Yih ET, McCoy DF, et al. Iterative design and usability testing of the Imhere system for managing chronic conditions and disability. Int J Telerehabil 2016;8(1):11-20. [doi: 10.5195/ijt.2016.6194] [Medline: 27563387]
30. Slattery P, Saeri AK, Bragge P. Research co-design in health: a rapid overview of reviews. Health Res Policy Syst 2020 Feb 11;18(1):17. [doi: 10.1186/s12961-020-0528-9] [Medline: 32046728]
31. Bevan Jones R, Stallard P, Agha SS, et al. Practitioner review: co‐design of digital mental health technologies with children and young people. Child Psychol Psychiatry 2020 Aug;61(8):928-940. [doi: 10.1111/jcpp.13258]

XSL•FO
RenderX

32. Adams Z, Grant M, Hupp S, et al. Acceptability of an mHealth app for youth with substance use and mental health needs: iterative, mixed methods design. JMIR Form Res 2021 Dec 24;5(12):e30268. [doi: 10.2196/30268] [Medline: 34951593]
33. Fox S, Brown LJE, Antrobus S, et al. Co-design of a smartphone app for people living with dementia by applying agile, iterative co-design principles: development and usability study. JMIR Mhealth Uhealth 2022 Jan 14;10(1):e24483. [doi: 10.2196/24483] [Medline: 35029539]
34. Sohal M, Singh P, Dhillon BS, Gill HS. Efficacy of journaling in the management of mental illness: a systematic review and meta-analysis. Fam Med Community Health 2022 Mar;10(1):e001154. [doi: 10.1136/fmch-2021-001154] [Medline: 35304431]
35. Galante J, Friedrich C, Dawson AF, et al. Mindfulness-based programmes for mental health promotion in adults in nonclinical settings: a systematic review and meta-analysis of randomised controlled trials. PLoS Med 2021 Jan;18(1):e1003481. [doi: 10.1371/journal.pmed.1003481] [Medline: 33428616]
36. Hyzy M, Bond R, Mulvenna M, et al. System Usability Scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022 Aug 18;10(8):e37290. [doi: 10.2196/37290] [Medline: 35980732]
37. Zwakman DS, Pal D, Arpnikanondt C. Usability evaluation of artificial intelligence-based voice assistants: the case of Amazon Alexa. SN Comput Sci 2021;2(1):28. [doi: 10.1007/s42979-020-00424-4] [Medline: 33458698]
38. Allen P, Bennett K, Heritage B. SPSS Statistics Version 22: A Practical Guide, 3rd edition: Cengage Learning Australia Pty Limited; 2014.
39. Escriche-Escuder A, De-Torres I, Roldán-Jiménez C, et al. Assessment of the quality of mobile applications (apps) for management of low back pain using the Mobile App Rating Scale (MARS). Int J Environ Res Public Health 2020 Dec 9;17(24):9209. [doi: 10.3390/ijerph17249209] [Medline: 33317134]
40. Mandracchia F, Llauradó E, Tarro L, Valls RM, Solà R. Mobile phone apps for food allergies or intolerances in app stores: systematic search and quality assessment using the Mobile App Rating Scale (MARS). JMIR Mhealth Uhealth 2020 Sep 16;8(9):e18339. [doi: 10.2196/18339] [Medline: 32936078]
41. Roberts AE, Davenport TA, Wong T, Moon HW, Hickie IB, LaMonica HM. Evaluating the quality and safety of health-related apps and e-tools: adapting the Mobile App Rating Scale and developing a quality assurance protocol. Internet Interv 2021 Apr;24:100379. [doi: 10.1016/j.invent.2021.100379] [Medline: 33777705]
42. Salehinejad S, Niakan Kalhori SR, Hajesmaeel Gohari S, Bahaadinbeigy K, Fatehi F. A review and content analysis of national apps for COVID-19 management using Mobile Application Rating Scale (MARS). Inform Health Soc Care 2021 Mar 2;46(1):42-55. [doi: 10.1080/17538157.2020.1837838] [Medline: 33164594]
43. Werner NE, Brown JC, Loganathar P, Holden RJ. Quality of mobile apps for care partners of people with Alzheimer disease and related dementias: Mobile App Rating Scale evaluation. JMIR Mhealth Uhealth 2022 Mar 29;10(3):e33863. [doi: 10.2196/33863] [Medline: 35348467]

Abbreviations

MARS: Mobile App Rating Scale
MVP: minimum viable product
SUS: System Usability Scale
UCSD: user-centered systems design

Edited by M Waite; submitted 24.Oct.2024; peer-reviewed by G Jones, R Chigevenga, Z Davey; revised version received 31.Aug.2025; accepted 18.Feb.2026; published 12.Nov.2026.

Please cite as:
McCabe M, Byrne M, Gullifer J, Smart I, Chang C, Singh S, Khara G, Cornish K
Exploring Use, Usability, and Functionality in the Iterative Testing of a Co-Designed Preventative Mental Health Smartphone App for University Students: Comparison of Monash Thrive App Development Trials
JMIR Hum Factors 2026;13:e67818
URL: https://humanfactors.jmir.org/2026/1/e67818
doi:10.2196/67818

© Melinda McCabe, Michelle Byrne, Judith Gullifer, Isabelle Smart, Chunlei Chang, Sanket Singh, Gagan Khara, Kim Cornish. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 12.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Role of Rating Valence in AI Skin Cancer App Acceptance: Eye-Tracking and Questionnaire Study

Inga Jagemann[1], MSc; Sabrina Hegner[2], Prof Dr; Gerrit Hirschfeld[1], Prof Dr

[1]School of Business, University of Applied Sciences and Arts Bielefeld, Interaktion 1, Bielefeld, Germany

[2]School of International Business, City University of Applied Sciences Bremen, Bremen, Germany

Corresponding Author:
Inga Jagemann, MSc
School of Business, University of Applied Sciences and Arts Bielefeld, Interaktion 1, Bielefeld, Germany

Abstract

Background: Artificial intelligence–based skin cancer screening apps (AISCSAs) offer diagnostic potential but face limited adoption. App store cues, such as ratings, may influence acceptance; yet, little is known about how users cognitively process app store information in high-stakes health contexts. To address this gap, eye-tracking was used to measure visual attention while participants evaluated a mock AISCSA app store listing.

Objective: This study aimed to test whether a single negative rating captures visual attention and whether an extended technology acceptance model (TAM) can predict behavioral intention to use (BI) AISCSAs.

Methods: Participants (N=76) evaluated a mock app store listing for an AISCSA under positive (n=42) or negative (n=34) rating conditions while their eye movements were recorded. Analyses combined fixation durations in defined areas of interest (AOIs) with self-reported measures of perceived usefulness (PU), perceived ease of use (PEOU), trust, BI, willingness to pay, and the self-rated importance of app attributes.

Results: Normalized fixation durations (seconds per square pixel) revealed the highest attention to the description (0.166 s/px^2), followed by the reviews (0.11 s/px^2) and the ratings (0.04 s/px^2), while the price and the data protection received the least attention. Of the 5 self-rated app attributes, only reviews correlated positively with fixation durations on the reviews-AOI (r=0.28; P=.01). Rating valence had no significant effect on gaze patterns, PU, PEOU, trust, BI, or willingness to pay (all Ps>.05). However, PEOU (P=.001), PU (P<.001), and trust (P<.001) were significantly correlated with BI.

Conclusions: Although the expected attentional capture effect of the negative rating was not observed, the weak or nonexistent associations between fixation durations on the AOIs and the self-rated importance of app attributes suggest that eye-tracking captures aspects of information processing that are not directly reflected in self-reported evaluations. These findings indicate that eye-tracking provides a more direct approximation of actual user behavior by revealing implicit attentional processes beyond what is captured by questionnaires. While the technology acceptance model constructs and trust predicted BI, rating valence alone did not affect acceptance or gaze behavior. In high-stakes health contexts, textual information may outweigh rating valence in driving adoption. Future research should explore conditions under which rating valence matters, including more extreme rating contrasts, variations in accompanying review texts, and the influence of individual differences such as preexisting attitudes toward artificial intelligence and levels of artificial intelligence literacy.

(JMIR Hum Factors 2026;13:e93489) doi:10.2196/93489

KEYWORDS

eye-tracking; mobile health; mHealth; artificial intelligence; AI; technology acceptance model; TAM; skin cancer screening

Introduction

Background

Skin cancer is a highly prevalent global health concern, with melanoma, the deadliest subtype, posing a significant challenge. Early detection is crucial for improving survival rates, and clinical guidelines recommend full-body skin examinations by trained dermatologists and the use of dermoscopy for assessing skin lesions [1,2]. However, a growing shortage of dermatologists is exacerbating diagnostic delays [3,4]. Even today, more than half of all skin cancers are first noticed by patients themselves rather than detected during professional screenings [5], which has driven the exploration of artificial intelligence (AI) to support skin cancer detection. Empirical studies show that deep learning models can achieve diagnostic performance comparable to human experts in analyzing dermoscopic images for melanoma [6,7]. When integrated into mobile apps, these systems provide immediate risk assessments through automated, deep learning–based image analysis. Despite their potential to support early detection, adoption rates remain

XSL•FO
RenderX

low, as many users continue to prefer conventional dermatological consultations [8-10].

This reluctance reflects a broader adoption gap identified in research on AI in health care—a discrepancy between the high diagnostic performance of AI systems and their limited acceptance in real-world practice [11,12]. While this gap is well-documented, less is known about the specific digital cues that drive such skepticism. Given that users increasingly rely on app store information to evaluate AI-based health apps, understanding how user-generated ratings shape acceptance becomes critical—particularly as these ratings represent a powerful yet understudied influence on decision-making in high-stakes health care scenarios.

Despite extensive research on rating valence in e-commerce, 3 key gaps remain. First, the role of rating valence in high-stakes health contexts, such as AI–based skin cancer screening apps (AISCSAs), remains largely unexplored. Second, prior research has not systematically integrated rating valence into the technology acceptance model (TAM) [13], particularly with regard to its causal effects on perceived usefulness (PU), perceived ease of use (PEOU), trust, and outcomes such as behavioral intention to use (BI) and willingness to pay (WTP). Existing studies predominantly rely on observational, nonexperimental designs, limiting causal inference. In contrast, this study uses an experimental manipulation of rating valence (positive vs negative) to directly examine its causal impact on these constructs. Third, prior work relies almost exclusively on self-reported measures and thus provides limited insight into the underlying cognitive processes, neglecting objective behavioral data such as eye-tracking.

To address these gaps, this study develops an integrative framework that combines the TAM, trust, and rating valence within a dual-process perspective of technology acceptance. By linking experimentally manipulated rating valence with both self-reported measures and eye-tracking data, we provide novel insights into how rating valence causally shapes acceptance beliefs and intentions, how users visually process app store information, and how closely self-reported evaluations align with actual attentional behavior.

Theoretical Foundation and Research Hypotheses

Effects of Rating Valence on Technology Acceptance

Traditional technology acceptance research, particularly the TAM, assumes that technology evaluation is primarily a deliberate and analytical process. The TAM posits that PU and PEOU are the key determinants of BI [13]. However, dual-process theory suggests that human cognition is governed by 2 interacting systems: type 1 processing, which is fast, intuitive, and heuristic-based, and type 2 processing, which is slower, reflective, and cognitively effortful [14]. Because individuals tend to act as “cognitive misers,” they typically default to type 1 processing and rely on heuristic cues, particularly in uncertain or information-rich environments [14].

In digital marketplaces, such as app stores, these heuristic processes are highly relevant. User-generated star ratings serve as salient cues that shape users’ initial impressions before more deliberate evaluations take place. In this context, ratings and reviews function as key signals of product reliability by conveying aggregated user experiences and recommendations [15-17]. Meta-analytic evidence confirms that review valence exerts a powerful effect on purchase intentions [18]. While positive reviews enhance perceived quality and download likelihood [19-21], negative reviews often exert even stronger effects because they are perceived as more diagnostic and trustworthy [22,23]. This asymmetry is attributed to the negativity bias, where negative information attracts greater attention and has a stronger impact on decisions than positive ones [24-26].

Taken together, this suggests that rating valence may act as an early heuristic signal that shapes subsequent evaluations before more analytical processing occurs. Acceptance beliefs, such as PU, PEOU, trust, BI, and WTP, may already be biased by initial exposure to rating information. Importantly, WTP appears less stable than BI and may be influenced in a more complex and indirect manner, as it reflects distinct evaluative judgments and is shaped by both functional and psychological factors [27-29]. Building on this reasoning, this study introduces an experimental manipulation of rating valence, enabling a direct test of how external heuristic cues causally influence the formation of PU, PEOU, trust, and outcomes such as BI and WTP. Accordingly, we address the following research question: What is the effect of positive versus negative ratings on behavioral intentions? To empirically examine this proposed preprocessing role of rating valence across different outcome variables, the following hypothesis is formulated:

- H1: Rating valence significantly influences user perceptions and intentions, such that a negative rating leads to lower (a) trust, (b) PU, (c) PEOU, (d) BI, and (e) WTP.

Extending the TAM

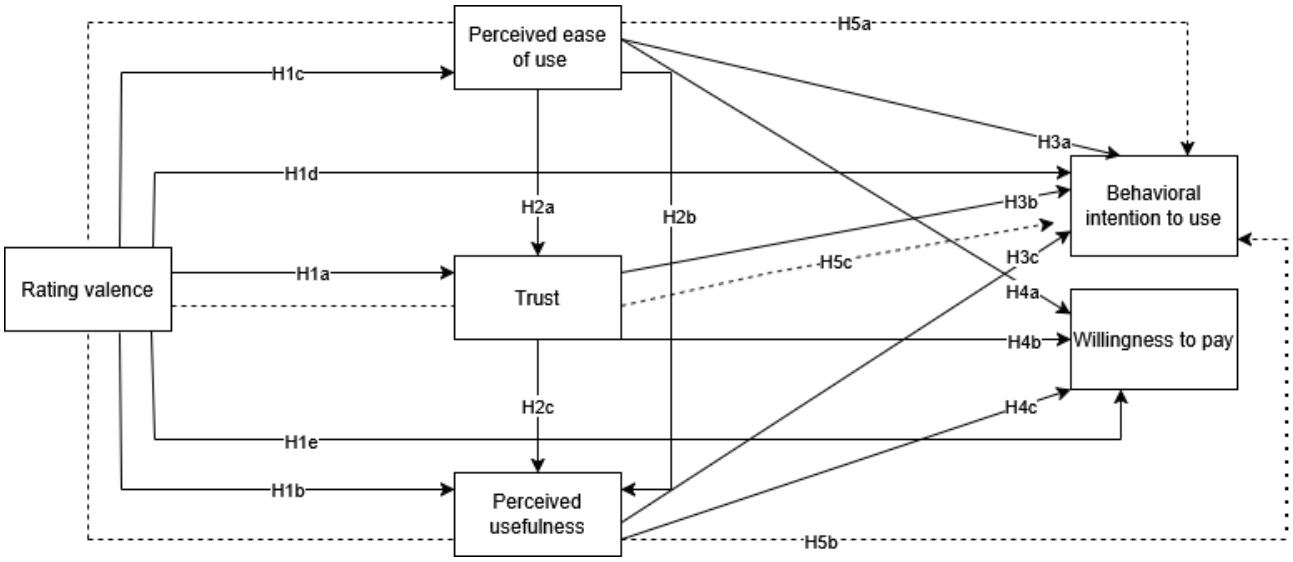
While the previous hypothesis proposed that rating valence influences initial evaluations, it remains unclear through which mechanisms these effects translate into acceptance decisions. The TAM provides a well-established framework to explain how user beliefs shape behavioral intentions. However, in AI-based health care contexts, the TAM alone may not fully capture user acceptance. A recent review by Lee et al [30] shows that the core TAM constructs are insufficient in complex health care environments, where additional factors, such as trust, organizational support, and education, significantly influence adoption. Due to the complexity, opacity, and perceived unpredictability of intelligent systems, trust emerges as an essential additional construct. Trust reflects the belief that a system performs reliably and safely, thereby reducing perceived risk and supporting decision-making under uncertainty [31,32]. Empirical evidence indicates that trust not only directly affects BI [22,33,34] but also positively influences PU [35-38]. Based on this framework, the following hypotheses are proposed (Figure 1):

- H2: Within the AISCSA context, PEOU is positively associated with (a) trust and (b) PU, and (c) trust is positively associated with PU.
- H3: BI is positively associated with (a) PEOU, (b) trust, and (c) PU.

XSL•FO
RenderX

- H4: WTP is positively associated with (a) PEOU, (b) trust, and (c) PU.

Figure 1. Integrative research model incorporating rating valence and trust into the technology acceptance model to predict behavioral intention to use and willingness to pay. H: hypothesis.



We want to test whether the TAM is able to explain the effects of the experimental manipulation. Prior research extending the TAM in AI-based health care contexts predominantly relies on cross-sectional survey designs to examine relationships between the TAM constructs and additional factors (eg, trust and privacy concerns) without experimentally manipulating external inputs [33,34,39,40]. While these approaches allow for the estimation of theoretically grounded causal pathways, they are based on observational (nonexperimental) data and therefore remain limited in establishing causal effects. Against this background, a key question is whether the TAM can adequately explain how externally manipulated rating information translates into behavioral intentions. Accordingly, we address the following research question: To what extent can the TAM explain the effect of rating valence on behavioral intentions? Building on this reasoning, we propose the following hypotheses:

- H5: The effect of rating valence on BI is mediated by (a) PEOU, (b) PU, and (c) trust.

Visual Attention and the Processing of Heuristic Cues

While heuristic cues such as rating valence are assumed to influence evaluations, this assumption does not account for how users cognitively process such information at the moment of exposure. Traditional technology acceptance research has largely relied on self-reported measures to assess acceptance beliefs, thereby conceptualizing decision-making as a consciously accessible and predominantly rational process [13]. However, such measures are limited, as individuals are often unable to accurately report the processes underlying their decisions due to recall bias and introspective inaccessibility [41-43]. Eye-tracking offers an important window into these processes by capturing objective and continuous measures of visual attention and cognitive engagement. According to the eye-mind hypothesis, individuals process what they fixate on, making visual fixation a reliable indicator of ongoing cognitive processing [44,45]. In this context, fixation duration reflects cognitive effort and information processing depth, with longer fixations indicating more intensive processing [46]. From a dual-process perspective, fixation patterns can be interpreted as indicators of different processing modes: shorter fixations are associated with intuitive, heuristic-based (type 1) processing, whereas longer fixations suggest more deliberate and analytical (type 2) processing [47,48]. Importantly, eye-tracking captures nonconscious attentional processes and thus provides an unobtrusive and behaviorally grounded measure of information processing that complements self-reported evaluations. Here, we use eye-tracking to improve our understanding in 3 ways.

First, we investigate where people look for information in the app store when browsing for health-related apps. In digital environments, attention is inherently selective and shaped by both stimulus characteristics and user goals. Prior eye-tracking research suggests that users tend to focus on visually salient elements such as images [49-53]. However, app store-specific evidence indicates that decision-making often relies more heavily on textual information, particularly descriptions and user reviews [54]. At the same time, existing findings point to substantial variability in attention patterns. Eye-tracking studies show that while some users engage extensively with review content, others allocate more attention to product selection, with differences observed across user groups such as different age cohorts [55]. Conversely, privacy- and security-related information typically receives limited attention, especially when presented in less salient formats [56]. Users may overlook privacy policies despite expressing concerns and rely on heuristic cues when apps are perceived as necessary [57,58]. Taken together, these findings suggest that attention allocation in app store environments is neither uniform nor fully explained by existing theory. In particular, it remains unclear which types of information users prioritize when evaluating high-risk health apps. To address this gap, we descriptively analyze gaze patterns.

Second, we test whether users' consciously reported importance of app attributes corresponds to their actual allocation of visual

XSL•FO
RenderX

attention during decision-making. While self-reports and eye-tracking capture different levels of cognitive processing [59], they are expected to converge on attributes that users deem personally relevant. If an attribute is integrated into an individual's conscious value system, it should actively guide attentional resources toward related information. Accordingly, we address the following research question: To what extent are the self-rated importance of app attributes reflected in users' gaze patterns? Based on this reasoning, the following hypothesis is proposed:

- H6: There is a positive relationship between the self-rated importance of app attributes and actual visual attention (fixation durations) to the corresponding areas of interest (AOIs).

Third, we test the effect of our experimental manipulation on visual attention. Consistent with the negativity bias [24-26], prior eye-tracking research shows that negative information attracts greater visual attention and elicits more intensive cognitive processing [60,61]. Such heightened attention may reflect a shift from intuitive, automatic processing (type 1) to more deliberate and analytical processing (type 2), particularly when users encounter risky or conflicting information [47]. Supporting this view, evidence suggests that negative or misleading stimuli can activate type 2 processing and prompt deeper evaluation compared to consensus-based stimuli [62]. Based on this reasoning, the following hypothesis is proposed:

- H7: A single negative rating increases visual attention to user reviews, as reflected in longer fixation durations on the reviews-AOI.

Moderating Role of Attention and Individual Differences

While rating valence shapes perceptions through the TAM-related mechanisms, its impact is unlikely to be uniform across users and situations. Dual-process theory suggests that the influence of heuristic cues depends on processing depth: information that is processed more extensively exerts a stronger effect on subsequent judgments [63]. Visual attention provides an observable indicator of this processing depth, as information receiving more attention is more likely to be incorporated into decision-making [64]. Beyond mere information acquisition, research indicates that visual attention actively biases decisions through a causal amplification effect. According to this mechanism, fixations are not neutral; rather, they amplify the relative value signal of the attended information. Consequently, longer fixations increase the weighting of a specific cue, such as rating valence, within the value integration process, thereby exerting a disproportionate influence on the final evaluation [65,66]. This implies that rating valence should have a stronger impact on user beliefs and intentions when it receives greater visual attention. In other words, the effect of rating valence is not fixed but depends on how deeply the cue is processed and integrated into the decision-making process. To capture this conditional effect, we propose that visual attention moderates the relationship between rating valence and acceptance-related outcomes:

- H8: The effect of rating valence on (a) PU, (b) PEOU, (c) trust, and (d) BI is moderated by visual attention, such that higher fixation durations on the ratings-AOI strengthen this relationship.

Prior research suggests that individual differences influence how users process and respond to app store information. Factors such as age, sex, education, prior experience with health apps, and self-reported AI knowledge may shape reliance on heuristic cues as well as attention allocation and acceptance judgments [67-69]. Less experienced users, for instance, may rely more strongly on ratings, whereas more knowledgeable users may engage in more analytic processing. However, existing evidence does not support a comprehensive set of directional hypotheses for each characteristic. Therefore, these variables are included as control and exploratory factors to provide additional insights into users' attention, perceptions, and behavioral intentions.

Methods

Participants and Recruitment

Participants were recruited via university mailing lists, flyers, and student WhatsApp groups. Eligibility criteria included a minimum age of 18 years and fluency in German. Data collection took place between April 28 and May 31, 2025. A total of 95 participants were initially recruited; 19 were excluded due to insufficient eye-tracking accuracy or failed attention checks, resulting in a final sample of 76 participants (positive condition: n=42; negative condition: n=34). An a priori power analysis (G*Power; α=.05; power=0.95; effect size=0.80) indicated a required sample size of at least 35 participants per group, confirming sufficient statistical power. In response to the reviewers' and editor's comments, we have streamlined and refined our hypotheses. Importantly, this revision involved a structural clarification rather than changes to the underlying analyses or hypothesis testing. The updated hypotheses and their correspondence to the original versions are presented in Table S1 in Multimedia Appendix 1.

Eye-Tracking Setup

Gaze data (fixation duration) were recorded using a Tobii Pro Spark eye tracker (60 Hz), mounted below a 22-inch laptop screen (1920×1080 pixels, 60 Hz). Sessions were conducted individually in a controlled laboratory environment, with participants seated approximately 45 - 55 cm from the screen [70]. A standard 9-point calibration was performed prior to the experiment, and only calibrations with a validation accuracy below 0.80° were accepted [71].

Materials

The stimulus for the eye-tracking experiment was a mock-up of an Apple app store listing for a fictitious skin cancer screening app ("SkinScan"), modeled after existing apps. To reduce familiarity bias, app name and content were adapted. Two versions of SkinScan were created, identical in all aspects except for the middle of the 3 user star ratings: the positive rating condition showed two 5-star ratings and one 4-star rating, while the negative rating condition replaced the middle 5-star with a 1-star rating to introduce the experimental manipulation (Multimedia Appendix 2)Multimedia Appendix 2. All other elements, including the overall star rating and review text, were

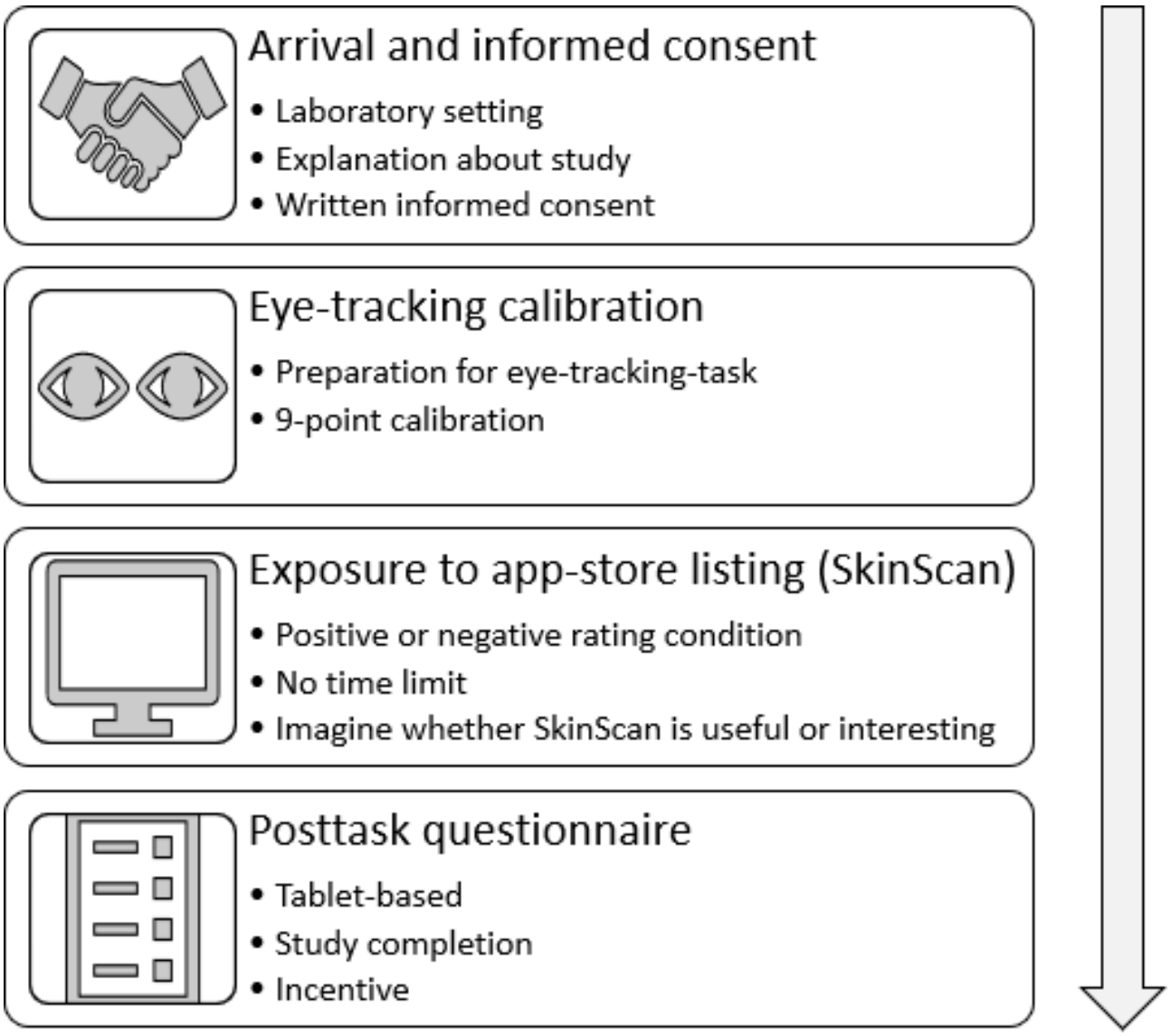
XSL•FO
RenderX

identical across conditions. Both app versions were integrated into and presented through Tobii Pro Lab.

Procedure

After providing informed consent, participants were seated in front of the eye-tracking system. They were instructed to view the app page as they normally would and to imagine evaluating its usefulness. Subsequently, a 9-point calibration procedure was conducted. Following calibration, the experiment began. No time limit was imposed, and gaze behavior was recorded continuously throughout the task. After stimulus exposure, participants completed a questionnaire on a tablet device and received compensation (see Figure 2 for the procedure of the experiment). A small pretest (n=3) using a think-aloud approach was conducted to ensure clarity and usability of the study design.

Figure 2. Schematic representation of the study workflow.



Posttask Questionnaire

The questionnaire consisted of 29 items. The first section measured PEOU, PU, trust, and BI using validated multi-item scales, with each construct measured by 4 items (see Table S2 in Multimedia Appendix 1 for a complete list of items and scale specifications). The second section assessed the self-rated importance of app attributes (eg, price, reviews, and data protection) on 5-point Likert scales (1=not at all important and 5=very important). The third section included self-rated AI knowledge, WTP, and demographic variables (age, sex, previous experience with health apps, and education). All scales were adapted from prior research. Following established scale development procedures [72,73], content modifications were made to reflect the study's objectives. To ensure linguistic accuracy, all items were translated using a back-translation procedure [74]. An attention-check item was included to ensure data quality. A full list of questionnaire items is provided in Table S2 in Multimedia Appendix 1.

Data Analysis

To assess potential common method bias, the Harman single-factor test was conducted [75]. An exploratory factor analysis including all questionnaire items was performed. Parallel analysis indicated a multifactor structure, and the single-factor solution accounted for only 30% of the total variance. Moreover, the 1-factor model showed poor fit

XSL•FO
RenderX

(χ^2_{135}=4364.2; P<.001). These results suggest that common method bias is unlikely to be a significant concern in this study.

Fixation durations (in seconds) were recorded within predefined AOIs (Multimedia Appendix 3)Multimedia Appendix 3 and processed using Tobii Pro Lab. Fixation duration data were aggregated into composite categories (eg, reviews, ratings, and description) and normalized by AOI size (seconds per square pixel) to account for differences in area. Mean fixation durations per AOI were then calculated at the participant level. Participants with insufficient tracking accuracy or missing data were excluded (n=19), yielding a final sample of 76.

Prior to analysis, scatterplots were examined to assess linearity and identify influential outliers. Pearson correlations were used to test relationships between variables, and independent-samples 2-tailed *t* tests (including Cohen *d*) were conducted to compare experimental conditions. Multiple regression analyses were performed to assess predictors of BI and WTP. Mediation and moderation effects were tested using regression-based approaches with bootstrapped CIs. All analyses were conducted at a significance level of P<.05, with 95% CIs reported.

Ethical Considerations

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. An ethics approval was granted by the University of Applied Sciences and Arts Bielefeld (2025_002), and the study was preregistered. Informed consent was required from all participants in this study. Participants received approximately US $5.83 compensation.

Results

Sample Characteristics

The final sample consisted of 76 participants. Of these, 42 of 76 (55%) participants were randomly assigned to view the positive rating condition, while 34 of 76 (45%) participants saw the negative rating condition. As shown in Table 1, participants had a mean age of 22.88 (SD 3.13; range 18-33) years and a median age of 22 (IQR 21-24) years. The majority of participants were female (53/76, 70%). Most reported having a general school leaving certificate (49/76, 64%), while 13 of 76 (17%) held a university degree. The mean level of self-rated AI knowledge was 2.65 (SD 0.91). None of the participants had prior experience with skin cancer screening apps.

Table . Characteristics of the study sample.

Characteristics	Values
Age (years)	
Mean (SD)	22.88 (3.13)
Range	18-33
Median (IQR)	22 (21-24)
Sex, n (%)	
Female	53 (70)
Male	23 (30)
Education, n (%)	
General school leaving certificate (eg, Abitur)	49 (64)
Vocational school qualification (eg, apprenticeship)	12 (16)
University degree (eg, bachelor)	13 (17)
PhD	—[a]
Other	1 (1)
No answer	1 (1)
Previous experience with skin cancer screening apps, n (%)	
Yes	—[a]
No	76 (100)

[a]Not available.

Effects of Rating Valence on Technology Acceptance

To address the research question of how a single negative rating affects behavioral intentions, independent-samples comparisons were conducted to examine the effect of rating valence on trust, PU, PEOU, BI, and WTP (Table 2). The results indicated no significant differences between positive and negative rating conditions for any of the examined variables (all Ps>.05). Thus, H1a-e are not supported.

XSL•FO
RenderX

Table . Effects of rating valence on the technology acceptance model constructs.

Hypothesis (H)	Variable	Positive rating (n=42), mean (SD)	Negative rating (n=34), mean (SD)	*P* value, overall	Cohen *d*
H1a	Trust	2.48 (0.61)	2.62 (0.70)	.35	0.22
H1b	PU[a]	3.78 (0.64)	3.71 (0.71)	.64	0.11
H1c	PEOU[b]	4.23 (0.48)	4.12 (0.62)	.41	0.20
H1d	BI[c]	2.96 (0.84)	3.14 (1.01)	.42	−0.19
H1e	WTP[d]	19.5 (76.4)	63.1 (229)	.30	−0.27

[a]PU: perceived usefulness.

[b]PEOU: perceived ease of use.

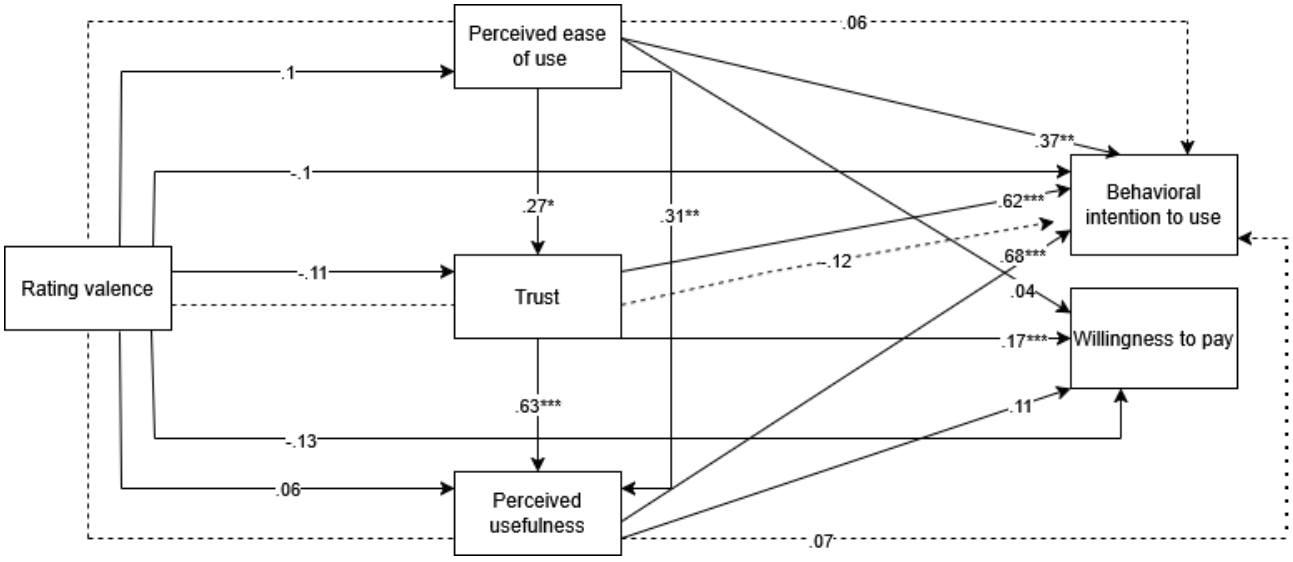
[c]BI: behavioral intention to use.

[d]WTP: willingness to pay.

Extending the TAM

To investigate the research question concerning the applicability of the TAM framework in connecting rating valence to behavioral intentions, Pearson correlation coefficients were calculated to examine the strength of the relationships between the hypothesized constructs (Figure 3). PEOU was positively correlated with PU (H2b; r=0.31; P=.007) and trust (H2a; r=0.27; P=.02). Trust also showed a strong positive correlation with PU (H2c; r=0.63; P<.001). Regarding BI, all proposed constructs were significantly positively correlated: PEOU (H3a; r=0.37; P=.001), trust (H3b; r=0.62; P<.001), and PU (H3c; r=0.68; P<.001). In addition, Pearson correlation analyses revealed a small but significant positive association between trust and WTP (H4b; r=0.17; P<.001). In contrast, WTP was not significantly correlated to PU (H4c; r=0.11; P=.35) or PEOU (H4a; r=0.04; P=.76).

Figure 3. Summary of the results of the extended technology acceptance model. Dashed lines illustrate mediation analysis. H: hypothesis. *P=.05, **P<.01, ***P<.001.



To further examine the predictors of BI, a multiple linear regression analysis was conducted including rating valence (condition), PU, PEOU, and trust. The overall model was significant ($F_{4,71}$=22.36; P<.001), explaining a substantial proportion of variance in behavioral intention (R^2=0.56; adjusted R^2=0.53). PU emerged as the strongest predictor of BI (β=.65; P<.001). Trust also showed a significant positive effect (β=.37; P=.01), while PEOU had a marginal positive influence (β=.27; P=.06). In contrast, rating valence (condition) did not significantly predict BI (β=−.20; P=.18). Overall, the findings closely resemble the individual correlations, with the exception that the effect of trust weakens once PU is included, pointing toward a potential mediation effect.

To examine whether rating valence and the TAM-related constructs predict WTP, a multiple linear regression analysis was conducted including condition, PU, PEOU, and trust. The overall model was not significant ($F_{4,71}$=0.77; P=.55) and explained only a negligible proportion of variance in WTP (R^2=0.04; adjusted R^2=−0.01). None of the predictors showed a significant effect on WTP, including rating valence (β=−.20; P=.32), PU (β=.02; P=.85), PEOU (β=.00; P=.99), and trust (β=.10; P=.39). Overall, the regression results largely mirror the correlation findings. However, the initially observed association between trust and WTP did not remain significant in the regression model, suggesting that this relationship is weak and not robust when controlling for other variables.

XSL•FO
RenderX

To examine whether the effect of rating valence on BI is mediated by PEOU, PU, and trust (H5a-c), mediation analyses were conducted (see Table 3 and dashed lines in Figure 3). Results revealed no significant indirect effects for any of the mediators, as all CIs included 0. Additionally, no significant total or direct effects of rating valence were observed. Thus, H5 is not supported.

Table . Mediation effects of rating valence on behavioral intention to use via PEOU[a], PU[b], and trust.

Hypothesis (H)	Mediator	Indirect effect (ACME[c]), estimate (95% CI)	Direct effect (ADE[d]), estimate (95% CI)	Total effect, estimate (95% CI)
H5a	PEOU	0.06 (−0.10 to 0.25)	−0.24 (−0.63 to 0.18)	−0.17 (−0.57 to 0.25)
H5b	PU	0.07 (−0.22 to 0.35)	−0.24 (−0.56 to 0.05)	−0.17 (−0.60 to 0.23)
H5c	Trust	−0.12 (−0.40 to 0.14)	−0.05 (−0.38 to 0.30)	−0.17 (−0.55 to 0.28)

[a]PEOU: perceived ease of use.

[b]PU: perceived usefulness.

[c]ACME: average causal mediation effect.

[d]ADE: average direct effect.

Visual Attention and the Processing of Heuristic Cues

To examine the research question of where users look for information when browsing for a skin cancer screening app, we analyzed how participants allocate visual attention across the app store interface. During data preprocessing, individual AOIs were aggregated into composite AOIs. Specifically, all review-related AOIs were combined into a reviews-AOI, all rating-related AOIs into a ratings-AOI, icon and screenshots into a design-AOI, and information and description into a description-AOI. To control for differences in AOI size, fixation durations were normalized and expressed as seconds per pixel² (s/px^2). As shown in Table 4, the description-AOI attracted the longest fixation durations (mean 0.16, SD 0.10 s/px^2), followed by the reviews-AOI (mean 0.11, SD 0.07 s/px^2) and the ratings-AOI (mean 0.04, SD 0.04 s/px^2). In contrast, the price-AOI (mean 0.02, SD 0.03 s/px^2) and the data protection-AOI (mean 0.02, SD 0.02 s/px^2) received the least attention.

Table . Normalized fixation durations by area of interest (AOI).

AOI	Values, n	Mean (SD) (s/px^2)	95% CI
Design	76	0.04 (0.02)	0.04 - 0.05
Description	76	0.16 (0.10)	0.14 - 0.19
Reviews	76	0.11 (0.07)	0.09 - 0.13
Ratings	68	0.04 (0.04)	0.03 - 0.05
Data protection	76	0.02 (0.02)	0.02 - 0.03
Price	76	0.02 (0.03)	0.01 - 0.03

To examine the research question of whether the self-rated importance of app attributes is reflected in users' gaze patterns, Pearson correlations were computed between participants' ratings and normalized fixation durations on the corresponding AOIs (Table 5). Results revealed a significant positive correlation for reviews (r=0.28; P=.01), while all other relationships were nonsignificant (Ps>.05). Thus, only H6b is supported.

XSL•FO
RenderX

Table . Pearson correlations between participants' self-rated importance of app attributes and fixation durations on corresponding areas of interest (AOIs).

Hypothesis (H)	Normalized fixation durations on AOI (s/px^2)	Self-rated importance of app attributes	App attribute, mean (SD)	*r* (95% CI)
H6a	Design	Importance of design	4.03 (0.83)	−0.11 (–0.32 to 0.11)
H6b	Reviews	Importance of reviews	4.21 (0.80)	0.28[a] (0.05 to 0.47)
H6c	Ratings	Importance of reviews	4.21 (0.80)	0.17 (–0.07 to 0.39)
H6d	Price	Importance of price	4.57 (0.63)	≈0 (–0.26 to 0.22)
H6e	Description	Importance of description	4.72 (0.53)	−0.21 (–0.41 to 0.01)
H6f	Data protection	Importance of data protection	3.82 (1.05)	0.08 (–0.14 to 0.30)

[a]*P*<.05.

To address the research question of how rating valence affects gaze patterns, fixation duration on the manipulated review-AOI was compared across rating conditions. Absolute fixation durations were used, as the AOI size was identical across conditions. Results showed no significant difference between the positive (mean 6.17, SD 4.15 seconds) and negative conditions (mean 6.21, SD 3.98 seconds; *P*=.97; *d*=−0.01). Although participants in the negative condition showed a slight tendency toward longer viewing times, this effect was not statistically significant. Thus, H7 is not supported.

Moderating Role of Attention and Individual Differences

To test whether visual attention moderates the relationship between rating valence and key constructs (H8a-d), moderation analyses were conducted using fixation durations on the ratings-AOI (Table S3 in Multimedia Appendix 1). Results indicated no significant interaction effects between rating valence and fixation durations on PU, PEOU, trust, or BI (all *P*s>.05). Thus, H8 is not supported.

Analyses of individual differences revealed no significant effects of demographic or experiential variables on the main outcome measures (trust, PU, PEOU, BI, and WTP; all *P*s>.05). Notably, both age and prior experience showed limited variance within the sample, which restricts the interpretability of their relationships with the outcome variables. In particular, none of the participants reported prior experience with skin cancer screening apps. Furthermore, self-reported AI knowledge was not significantly associated with any of the outcome variables (all *P*s>.05). Similarly, no significant differences were observed across sex or education levels (all *P*s>.05)

Exploratory Analysis

Furthermore, we explored potential sex differences in self-rated app attribute importance and gaze patterns. Overall, female and male participants evaluated most app attributes similarly. No significant sex differences were observed for the importance of price, data protection, or evidence (all *P*s>.05). However, female participants rated the importance of reviews (*d*=0.63; *P*=.01) and design (*d*=0.69; *P*=.02) significantly higher than male participants (Table 6).

XSL•FO
RenderX

Table . Sex differences in self-rated importance of app attributes and normalized area of interest (AOI) fixation durations.

Measure and attribute or AOI	Female (n=53), mean (SD)	Male (n=23), mean (SD)	*P* value, overall	Cohen *d*
Self-rated importance				
Price	4.66 (0.52)	4.39 (0.84)	.17	0.43
Data protection	3.83 (1.03)	3.83 (1.11)	.99	<.001
Evidence	4.70 (0.57)	4.78 (0.42)	.48	−0.16
Reviews	2.36 (0.79)	3.87 (0.76)	.01[a]	0.63
Design	4.21 (0.72)	3.65 (0.98)	.02[a]	0.69
Normalized fixation durations on AOI (s/px^2)				
Price	0.02 (0.03)	0.02 (0.03)	.81	−0.05
Data protection	0.03 (0.03)	0.02 (0.02)	.26	0.24
Ratings	0.05 (0.05)	0.03 (0.02)	.03[a]	0.47
Reviews	0.12 (0.07)	0.09 (0.05)	.12	0.34
Design	0.05 (0.03)	0.04 (0.01)	.08	0.36
Description	0.17 (0.11)	0.16 (0.07)	.46	0.15

[a]*P*<.05.

With regard to gaze patterns, female participants demonstrated longer fixation durations than male participants across all AOIs; however, this difference reached statistical significance only for ratings (d=0.47; P=.03). No significant sex differences were observed for fixation durations on data protection, description, reviews, design, or price (all Ps>.05; Table 6).

Discussion

Principal Findings

This study investigated the acceptance of AISCSAs by integrating the TAM with eye-tracking and the experimental manipulation of rating valence. Importantly, by systematically varying rating valence as an external heuristic cue, the study enabled a direct test of how such cues causally influence the formation of PU, PEOU, and trust, as well as BI and WTP. Our findings reveal a hierarchy of visual attention dominated by textual descriptions and social proof while also identifying a significant gap between self-reported attribute importance and actual gaze behavior. Although the core TAM constructs (PU, PEOU, and trust) emerged as robust predictors of BI, the experimental manipulation of rating valence did not significantly alter acceptance beliefs or attention patterns in this specific laboratory context.

Effects of Rating Valence on Acceptance and Attention

Contrary to our expectations, rating valence did not significantly influence trust, PU, PEOU, BI, or WTP. Likewise, the negative rating condition did not increase visual attention to the review section. Thus, neither H1 nor H7 was supported. At first glance, these findings appear inconsistent with research on negativity bias and online review effects, which often show that negative information attracts more attention and exerts a disproportionately strong influence on judgments [24,60]. However, our findings are more plausibly interpreted as evidence of a boundary condition than as evidence against the relevance of rating valence per se.

A first explanation concerns the subtlety of the manipulation. In this study, rating valence was altered only minimally, replacing 1 of 3 visible ratings with a single 1-star evaluation, while all other cues remained favorable. Such a minor discrepancy may have been insufficient to shift appraisals of a high-stakes health app. This is consistent with meta-analytic evidence by Purnawirawan et al [76], which demonstrates that valence effects are not uniform; rather, they are highly dependent on the degree of extremity and the evaluative context, such as product type and brand familiarity. In high-stakes environments, a single negative cue may lack the weight necessary to override a predominantly positive information environment. In addition, in high-stakes contexts, users may engage in more motivated and systematic processing, thereby reducing the relative impact of simple heuristic cues such as rating valence.

A second explanation is that app store ratings may have limited informational value in digital health contexts. Several studies show that consumer ratings are poor proxies for clinically relevant attributes such as quality, safety, and privacy. For example, Hyzy et al [77] found that ratings and downloads do not reliably reflect professional assessments, while de Chantal et al [78] reported a clear gap between user ratings and expert evaluations. Similarly, Levine et al [79] demonstrated that star ratings correlate only weakly with critical dimensions such as privacy, security, and interoperability, with even highly rated apps often performing poorly in these areas. This suggests that rating information may not provide users with meaningful signals for evaluating clinically relevant aspects of health apps. As a result, ratings may play a limited role in shaping acceptance beliefs.

XSL•FO
RenderX

From a dual-process perspective, the results indicate that the manipulated rating did not cross the threshold required to redirect processing from a default heuristic impression (type 1) to more effortful elaboration (type 2). While negative information can trigger deeper scrutiny [62], this is most likely when it is salient, coherent, and clearly diagnostic. Here, the cue may have been noticeable but not sufficiently consequential to reorganize attention or alter technology beliefs. This interpretation also fits the absence of any moderation by fixation durations on the ratings-AOI (H8): without a sufficiently strong evaluative signal, there was little for visual attention to amplify. Importantly, the null findings for H1 and H7 should not be interpreted as evidence that rating valence is irrelevant in digital health. Rather, they indicate that isolated, weakly negative rating cues are insufficient to meaningfully influence acceptance beliefs or increase visual attention in a high-risk medical AI context.

TAM Validity and the Role of Trust in AI Health Apps

Despite the absence of direct effects of rating valence, the internal structure of the extended TAM was strongly supported. The positive relationships between PEOU, PU, and trust (H2a-c), as well as their associations with BI (H3a-c), are consistent with prior research, demonstrating the robustness of the TAM in digital health contexts [37,40,80]. The regression results further refine this picture by highlighting a differentiated role of the constructs: PU emerged as the strongest predictor of BI, followed by trust, while PEOU showed only a marginal effect. This pattern suggests that usefulness constitutes the primary decision criterion, whereas ease of use operates mainly indirectly via PU.

The clear hierarchy of predictors reveals a specific decision-making logic in high-stakes AI health contexts. The dominance of PU suggests that for potentially life-impacting technologies such as AISCSAs, users follow a strongly instrumental evaluation logic, prioritizing diagnostic utility over interface convenience. In such contexts, perceived functional value appears capable of compensating for usability barriers, indicating that effectiveness outweighs ease of use. This finding is consistent with prior research in health care contexts, showing that PU is typically the strongest predictor of behavioral intention [34,40].

At the same time, our findings highlight the important role of trust in AI-based health care. While PU emerged as the strongest predictor of BI, trust remains as an independent predictor and is particularly relevant in contexts characterized by uncertainty and limited user expertise. In such settings, users rely on systems whose diagnostic processes are not easily interpretable, making trust a necessary condition for acting on PU. In this sense, trust may function as an enabling mechanism or boundary condition that determines whether PU translates into behavioral intention. This interpretation aligns with trust theory, which emphasizes its importance under conditions of uncertainty and risk [81,82], and with prior extensions of the TAM that position trust as a core belief construct [34,83-86].

Interestingly, the marginal effect of PEOU challenges the traditional emphasis on usability as a key driver of technology acceptance [13] and aligns with prior research, showing that PU is typically the dominant predictor, particularly in high-stakes contexts [34,37]. In this study, PEOU appears to play a secondary role once a system is perceived as useful and trustworthy. Users may be willing to tolerate complexity if it is associated with higher perceived competence or clinical rigor. This suggests that simplicity is not the primary objective in AI-based health care apps, with reliability and effectiveness outweighing usability considerations.

In contrast, WTP was not explained by the included variables. As WTP was measured using an open-ended response format, the variable was not normally distributed. To account for this, additional nonparametric correlation analyses (Spearman ρ) were conducted, yielding the same pattern of results. WTP was not significantly associated with PU or PEOU, and its relationship with trust remained weak (H4a and H6c not supported). This indicates that WTP represents a distinct and less stable evaluative judgment that cannot be fully explained by the traditional TAM constructs [23,28]. Instead, monetary evaluations appear to be driven more strongly by psychological and normative factors, such as perceived risk, fairness, and emotional or social considerations [27]. Accordingly, while the TAM provides a robust framework for explaining acceptance intentions, it appears less suitable for capturing economic decision-making without incorporating additional constructs.

The lack of significant mediation (H5) and moderation (H8) effects is directly tied to the nonsignificant main effect of rating valence. Without an initial “push” from the experimental condition, the “amplification effect” of visual attention, where fixations serve to increase the weighting of attended information [65,66], had no baseline signal to amplify. These findings underscore the robustness of the TAM framework while highlighting that single-cue manipulations may be insufficient to shift these deeply interrelated constructs.

Visual Attention Allocation and Information Processing

The eye-tracking results reveal a clear attentional hierarchy: users primarily focused on the app description, followed by reviews and ratings, while price and data protection received little attention. This pattern suggests that, in high-stakes contexts such as AI-based health apps, users prioritize information that helps them understand the app’s functionality and potential clinical value, indicating more deliberate and analytical processing (type 2). This finding is consistent with prior research showing that textual information plays a central role in app evaluation, despite developers often emphasizing visual design [54,87].

At the same time, the secondary attention given to reviews and ratings highlights that social proof remains relevant, particularly under uncertainty. These cues are easily interpretable and visually salient, making them effective heuristics for initial evaluation and risk reduction [19,21,60]. However, the fact that they received comparatively less attention than the description suggests that social proof complements rather than dominates decision-making in this context. Overall, the findings indicate that users integrate both analytical and heuristic processing, with greater weight placed on substantive, information-rich content.

XSL•FO
RenderX

In contrast, the data protection-AOI and the price-AOI received minimal attention. The limited focus on these elements is consistent with prior eye-tracking research, showing that users often overlook pricing information and security indicators in digital environments [58,88,89]. Several factors may explain this pattern. First, many health apps follow freemium models, where pricing information is either absent or only becomes relevant after installation. Second, although users frequently report privacy as a major concern [90,91], the complexity and legalistic nature of privacy policies may discourage active engagement. In addition, users may rely on implicit trust in platform-level regulation or assume a baseline level of data protection for health-related apps [92]. Finally, the low visual salience of these sections likely contributes to their neglect, as prior research shows that subtle privacy indicators are easily overlooked, whereas more prominent cues attract attention [56]. Taken together, these findings provide behavioral evidence for the well-documented "privacy paradox": despite expressing concerns about data protection, users allocate little attention to related information during decision-making.

The Discrepancy Between Self-Reports and Gaze Patterns

A central finding of this study is the limited alignment between self-reported importance and actual visual attention, with the exception of reviews (H6b). While participants who rated reviews as important also allocated more attention to them, no such relationship was observed for design, price, evidence, or data protection (H6a and H6c-f). This discrepancy suggests that self-reports and gaze data capture different aspects of the decision-making process. In line with prior research, eye-tracking reflects implicit attentional processes that are not accessible through retrospective self-reports [41].

From a dual-process perspective, this divergence can be interpreted as a misalignment between type 1 and type 2 processing [14]. Self-reported importance ratings primarily reflect deliberative, reflective evaluations (type 2), often shaped by normative expectations or socially desirable responses, particularly in sensitive contexts such as health care. In contrast, eye-tracking captures spontaneous, heuristic-driven allocation of attention during actual interaction (type 1). This gap is particularly pronounced for data protection. Although participants rated privacy as highly important, it received little visual attention during the task. This pattern is consistent with prior findings on the "privacy paradox," where expressed concerns do not translate into corresponding behavior [57,58]. Taken together, these findings highlight that technology acceptance cannot be understood as a purely rational process, as suggested by traditional TAM frameworks. Instead, it reflects an interplay between reflective evaluations and implicit attentional processes. By capturing these behaviorally grounded mechanisms, eye-tracking provides a valuable complement to self-report measures and offers a more nuanced understanding of how users evaluate health-related AI systems.

Individual Differences in AI App Evaluation

Our exploratory data suggest sex-specific patterns in app evaluation, with female participants assigning greater importance to reviews and design and showing longer fixation durations on ratings. This aligns with prior research, indicating that female participants tend to engage in more comprehensive information processing and pay closer attention to user-generated content when evaluating digital products [60,67,93]. Beyond sex, recent studies highlight the broader relevance of sociodemographic factors in shaping AI acceptance in health. For instance, Aras et al [68] show that younger individuals, female participants, and those with higher income levels are more likely to adopt AI in health care. Similarly, Méndez-Suárez et al [69] find that positive attitudes toward AI-based products are more prevalent among younger individuals and those with higher socioeconomic status. Complementing these findings, Kauttonen et al [94] demonstrate that individuals with moderate knowledge exhibit the highest acceptance of AI apps in health, whereas both low and high knowledge levels are associated with more cautious evaluations.

The absence of significant effects for AI knowledge and prior experience in this study likely reflects the relative homogeneity of the student sample, which limits variability in key characteristics. This highlights an important boundary condition of the findings. Future research should therefore examine more diverse populations, particularly older users and individuals with varying socioeconomic backgrounds and levels of digital literacy, to better capture how individual differences shape trust and acceptance in AI-based health care contexts.

Limitations

This study, while offering valuable insights, is subject to several limitations that should be acknowledged. First, although the stimulus used in the eye-tracking experiment ensured a high degree of ecological validity by closely resembling an actual app store page, the experimental environment and the task itself remained artificial and hypothetical; for instance, participants did not autonomously decide to search for an AISCSA. This setting may not fully capture the complexity and dynamics of real-world app evaluation and decision-making. Moreover, the absence of real consequences, such as actually downloading or using the app, could have influenced participants' engagement, attention patterns, and responses. Second, all participants were young university students with a narrow age range and no prior experience with skin cancer screening apps. This homogeneity limits the generalizability of the findings to other populations, especially older users or individuals with specific health concerns, who may differ in their levels of trust, digital literacy, and motivation. Therefore, the findings should be interpreted with caution, particularly when generalizing to more diverse or clinically relevant populations. Third, the internal consistency of some constructs, such as PU, PEOU, and trust, was relatively low, with Cronbach α values falling slightly below the conventional 0.70 threshold. While these levels are often considered acceptable in exploratory research within emerging fields, they suggest a moderate reliability, which should be taken into account when interpreting the strength of the observed relationships. Future research could benefit from further refining these scales to enhance their psychometric robustness in the context of AI-based health apps.

XSL•FO
RenderX

Conclusions

Addressing a key gap in prior research, this study extends the TAM by experimentally manipulating rating valence as an external variable and examining its effects on trust and core acceptance constructs. While prior research has largely relied on observational data, this approach enables a direct test of how external information influences the TAM-related belief structures.

The findings confirm that PU, PEOU, and trust remain significant predictors of behavioral intention in the context of AISCSAs. However, rating valence showed no direct effect on user acceptance or visual attention. This suggests that in high-stakes health contexts, users are less influenced by simple social-proof cues and instead prioritize substantive, information-rich content.

The primary implication for developers and marketers of health-related AI tools is that relying solely on positive ratings to drive adoption may be an insufficient strategy. Instead, efforts should be directed toward providing transparent, evidence-based information, as the description-AOI received the longest fixation duration. Building acceptance through verifiable information appears to be more critical than leveraging simple social proof in this domain. Moreover, designing a more appealing data protection section may increase user attention, rendering them better-informed consumers. Importantly, the weak or nonexistent associations between fixation durations on specific AOIs and the self-rated importance of app attributes indicate that eye-tracking captures aspects of information processing that are not directly reflected in self-reported questionnaires. These findings demonstrate that combining self-report measures with behavior-based methods, such as eye-tracking, provides a more comprehensive understanding of user acceptance processes in health-related AI systems. This dual approach captures both reflective, self-reported evaluations (type 2) and implicit, attention-driven processes (type 1), which are not accessible through questionnaires alone.

Funding

This research was conducted within the framework of the project “SAIL: SustAInable Lifecycle of Intelligent Socio-Technical Systems,” funded by the Ministry of Culture and Science of the State of North Rhine-Westphalia (grant NW21-059B).

Data Availability

The anonymized dataset, analysis scripts (R code), and codebook generated and analyzed during this study are publicly available in the Zenodo repository [95].

Authors' Contributions

Conceptualization: IJ, SH, GH
Data curation: IJ
Formal analysis: IJ, GH
Supervision: GH
Writing—original draft: IJ
Writing—review and editing: IJ, SH, GH

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview of preregistered hypotheses from the original submission and revised hypotheses reported in this paper. The table summarizes modifications made in response to reviewer and editor feedback during the peer-review process.
[DOCX File, 30 KB - humanfactors_v13i1e93489_app1.docx]

Multimedia Appendix 2
Visual representation of the experimental stimulus: positive (top) and negative (bottom) rating conditions. The manipulation focuses on the central review, where the star rating was varied while keeping the review text identical to isolate the effect of rating valence on user perception.
[PNG File, 233 KB - humanfactors_v13i1e93489_app2.png]

Multimedia Appendix 3
Predefined areas of interest for the simulated app-store interface. The colored regions represent the specific areas used to fixation duration.
[PNG File, 38 KB - humanfactors_v13i1e93489_app3.png]

References

XSL•FO
RenderX

1. S3-leitlinie prävention von hautkrebs. AWMF. 2021. URL: https://register.awmf.org/assets/guidelines/032-052OLk_S3_Praevention-Hautkrebs_2021-09.pdf [accessed 2025-06-12]
2. Melanoma: assessment and management. NICE. 2022. URL: https://www.nice.org.uk/guidance/ng14/resources/melanoma-assessment-and-management-pdf-1837271430853?utm_source=chatgpt.com [accessed 2026-01-08]
3. Ersser SJ, Effah A, Dyson J, et al. Effectiveness of interventions to support the early detection of skin cancer through skin self-examination: a systematic review and meta-analysis. Br J Dermatol 2019 Jun;180(6):1339-1347. [doi: 10.1111/bjd.17529] [Medline: 30561006]
4. Kis A, Augustin M, Augustin J. Regionale fachärztliche versorgung und demographischer wandel in Deutschland—szenarien zur dermatologischen versorgung im jahr 2035 [Regional dermatology workforce supply and demographic change in Germany: scenarios for dermatologic care in 2035]. J Dtsch Dermatol Ges 2017 Dec;15(12):1199-1210. [doi: 10.1111/ddg.13379_g] [Medline: 29228477]
5. Avilés-Izquierdo JA, Molina-López I, Rodríguez-Lomba E, Marquez-Rodas I, Suarez-Fernandez R, Lazaro-Ochaita P. Who detects melanoma? Impact of detection patterns on characteristics and prognosis of patients with melanoma. J Am Acad Dermatol 2016 Nov;75(5):967-974. [doi: 10.1016/j.jaad.2016.07.009] [Medline: 27645105]
6. Brinker TJ, Hekler A, Enk AH, et al. Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task. Eur J Cancer 2019 May;113:47-54. [doi: 10.1016/j.ejca.2019.04.001] [Medline: 30981091]
7. Codella NCF, Nguyen QB, Pankanti S, et al. Deep learning ensembles for melanoma recognition in dermoscopy images. IBM J Res Dev 2017;61(4/5):5. [doi: 10.1147/JRD.2017.2708299]
8. Gaube S, Biebl I, Engelmann MKM, Kleine AK, Lermer E. Comparing preferences for skin cancer screening: AI-enabled app vs dermatologist. Soc Sci Med 2024 May;349:116871. [doi: 10.1016/j.socscimed.2024.116871] [Medline: 38640741]
9. Jagemann I, Wensing O, Stegemann M, Hirschfeld G. Acceptance of medical artificial intelligence in skin cancer screening: choice-based conjoint survey. JMIR Form Res 2024 Jan 12;8:e46402. [doi: 10.2196/46402] [Medline: 38214959]
10. Baldauf M, Fröehlich P, Endl R. Trust me, I'm a doctor—user perceptions of AI-driven apps for mobile health diagnosis. In: Cauchard J, Löchtefeld M, editors. Presented at: 19th International Conference on Mobile and Ubiquitous Multimedia; Nov 22-25, 2020. [doi: 10.1145/3428361.3428362]
11. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. Lancet Digit Health 2021 Sep;3(9):e599-e611. [doi: 10.1016/S2589-7500(21)00132-1] [Medline: 34446266]
12. Longoni C, Bonezzi A, Morewedge CK. Resistance to medical artificial intelligence. J Consum Res 2019 Dec 1;46(4):629-650. [doi: 10.1093/jcr/ucz013]
13. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep 1;13(3):319-340. [doi: 10.2307/249008]
14. Evans J, Stanovich KE. Dual-process theories of higher cognition: advancing the debate. Perspect Psychol Sci 2013 May;8(3):223-241. [doi: 10.1177/1745691612460685] [Medline: 26172965]
15. Bickart B, Schindler RM. Internet forums as influential sources of consumer information. J Interact Mark 2001 Aug;15(3):31-40. [doi: 10.1002/dir.1014]
16. Park S, McCallister J. The effects of social proof marketing tactics on nudging consumer purchase. J Stud Res 2023;12(3). [doi: 10.47611/jsrhs.v12i3.4887]
17. Abdul Talib YY, Mat Saat R. Social proof in social media shopping: an experimental design research. SHS Web Conf 2017;34:02005. [doi: 10.1051/shsconf/20173402005]
18. Qiu K, Zhang L. How online reviews affect purchase intention: a meta-analysis across contextual and cultural factors. Data Inform Manag 2024 Jun;8(2):100058. [doi: 10.1016/j.dim.2023.100058]
19. Luan J, Shan W, Wang Y, Xiao J. How easy-to-process information influences consumers over time: online review vs. brand popularity. Comput Human Behav 2019 Aug;97:193-201. [doi: 10.1016/j.chb.2019.03.028]
20. Sällberg H, Wang S, Numminen E. The combinatory role of online ratings and reviews in mobile app downloads: an empirical investigation of gaming and productivity apps from their initial app store launch. J Market Anal 2023 Sep;11(3):426-442. [doi: 10.1057/s41270-022-00171-w]
21. Sung E, Chung WY, Lee D. Factors that affect consumer trust in product quality: a focus on online reviews and shopping platforms. Humanit Soc Sci Commun 2023;10(1). [doi: 10.1057/s41599-023-02277-7]
22. Kim Y, Peterson RA. A meta-analysis of online trust relationships in e-commerce. J Interact Mark 2017 May;38(1):44-54. [doi: 10.1016/j.intmar.2017.01.001]
23. Lee J, Oh Y, Kim M, Cho B, Shin J. Willingness to use and pay for digital health care services according to 4 scenarios: results from a national survey. JMIR Mhealth Uhealth 2023 Mar 29;11:e40834. [doi: 10.2196/40834] [Medline: 36989025]
24. Cui G, Lui HK, Guo X. The effect of online consumer reviews on new product sales. Int J Electron Commer 2012 Oct;17(1):39-58. [doi: 10.2753/JEC1086-4415170102]
25. Zhuang M, Cui G, Peng L. Manufactured opinions: the effect of manipulating online product reviews. J Bus Res 2018 Jun;87:24-35. [doi: 10.1016/j.jbusres.2018.02.016]

XSL•FO
RenderX

26. Park C, Lee TM. Information direction, website reputation and eWOM effect: a moderating role of product type. J Bus Res 2009 Jan;62(1):61-67. [doi: 10.1016/j.jbusres.2007.11.017]
27. Fu H, He W, Feng K, Guo X, Hou C. Understanding consumers' willingness to pay for circular products: a multiple model-comparison approach. Sustain Product Consum 2024 Mar;45:67-78. [doi: 10.1016/j.spc.2023.12.005]
28. Xie Z, Chen J, Or CK. Consumers' willingness to pay for eHealth and its influencing factors: systematic review and meta-analysis. J Med Internet Res 2022 Sep 14;24(9):e25959. [doi: 10.2196/25959] [Medline: 36103227]
29. McDougall JA, Furnback WE, Wang BCM, Mahlich J. Understanding the global measurement of willingness to pay in health. J Mark Access Health Policy 2020;8(1):1717030. [doi: 10.1080/20016689.2020.1717030] [Medline: 32158523]
30. Lee AT, Ramasamy RK, Subbarao A. Understanding psychosocial barriers to healthcare technology adoption: a review of TAM technology acceptance model and unified theory of acceptance and use of technology and UTAUT frameworks. Healthcare (Basel) 2025 Jan 27;13(3):250. [doi: 10.3390/healthcare13030250] [Medline: 39942440]
31. Gulati S, Sousa S, Lamas D. Modelling trust in human-like technologies. Presented at: Proceedings of the 9th Indian Conference on Human-Computer Interaction; Dec 16-18, 2018. [doi: 10.1145/3297121.3297124]
32. Sousa S, Lamas D, Dias P. A model for human-computer trust. In: Hutchison D, Kanade T, Kittler J, editors. Learning and Collaboration Technologies Designing and Developing Novel Learning Experiences: Springer International Publishing; 2014:128-137.
33. Kamal SA, Shafiq M, Kakria P. Investigating acceptance of telemedicine services through an extended technology acceptance model (TAM). Technol Soc 2020 Feb;60:101212. [doi: 10.1016/j.techsoc.2019.101212]
34. Liu K, Tao D. The roles of trust, personalization, loss of privacy, and anthropomorphism in public acceptance of smart healthcare services. Comput Human Behav 2022 Feb;127:107026. [doi: 10.1016/j.chb.2021.107026]
35. Matar A, Aloqaily AN. The mediating influence of perceived usefulness on consumer behaviour towards driving e-wallet adoption in Jordan. J Open Innov 2025 Dec;11(4):100651. [doi: 10.1016/j.joitmc.2025.100651]
36. Pitts G, Motamedi S. What drives students' use of AI chatbots? Technology acceptance in conversational AI. arXiv. Preprint posted online on Feb 24, 2026. [doi: 10.48550/arXiv.2602.20547]
37. Adnan A, Irvine RE, Williams A, Harris M, Antonacci G. Improving acceptability of mHealth apps-the use of the technology acceptance model to assess the acceptability of mHealth apps: systematic review. J Med Internet Res 2025 May 7;27:e66432. [doi: 10.2196/66432] [Medline: 40334265]
38. Beldad AD, Hegner SM. Running frequently with an app to be fantastic! Determinants of runtastic usage continuation intention among German users. Int J Mob Commun 2022;20(2):174. [doi: 10.1504/IJMC.2022.121436]
39. Dhagarra D, Goswami M, Kumar G. Impact of trust and privacy concerns on technology acceptance in healthcare: an Indian perspective. Int J Med Inform 2020 Sep;141:104164. [doi: 10.1016/j.ijmedinf.2020.104164] [Medline: 32593847]
40. Park JH, Lee CW, Do C. Examining users' acceptance intention of health applications based on the technology acceptance model. Healthcare (Basel) 2025 Mar 9;13(6):596. [doi: 10.3390/healthcare13060596] [Medline: 40150446]
41. Nisbett RE, Wilson TD. Telling more than we can know: verbal reports on mental processes. Psychol Rev 1977;84(3):231-259. [doi: 10.1037/0033-295X.84.3.231]
42. Matukin M, Ohme R, Boshoff C. Toward a better understanding of advertising stimuli processing. J Advert Res 2016 Jun;56(2):205-216. [doi: 10.2501/JAR-2016-017]
43. Segijn CM, Voorveld HAM, Vandeberg L, Smit EG. The battle of the screens: unraveling attention allocation and memory effects when multiscreening. Hum Commun Res 2017 Apr;43(2):295-314. [doi: 10.1111/hcre.12106]
44. Just MA, Carpenter PA. A theory of reading: from eye fixations to comprehension. Psychol Rev 1980 Jul;87(4):329-354. [doi: 10.1037/0033-295X.87.4.329] [Medline: 7413885]
45. Rayner K. Eye movements in reading and information processing: 20 years of research. Psychol Bull 1998 Nov;124(3):372-422. [doi: 10.1037/0033-2909.124.3.372] [Medline: 9849112]
46. Fu H, Manogaran G, Wu K, Cao M, Jiang S, Yang A. Intelligent decision-making of online shopping behavior based on internet of things. Int J Inf Manage 2020 Feb;50:515-525. [doi: 10.1016/j.ijinfomgt.2019.03.010]
47. Moravec PL, Minas RK, Dennis AR. Fake news on social media: people believe what they want to believe when it makes no sense at all1. MIS Q 2019 Dec 1;43(4):1343-1360. [doi: 10.25300/MISQ/2019/15505]
48. van der Gijp A, Ravesloot CJ, Jarodzka H, et al. How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology. Adv Health Sci Educ Theory Pract 2017 Aug;22(3):765-787. [doi: 10.1007/s10459-016-9698-1] [Medline: 27436353]
49. Leiva LA, Xue Y, Bansal A, et al. Understanding visual saliency in mobile user interfaces. Presented at: 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '20); Oct 5-8, 2020. [doi: 10.1145/3379503.3403557]
50. Li T, Pavlou PA. How does social influence really affect consumer decisions? Insights from an eye tracking study. SSRN J 2014. [doi: 10.2139/ssrn.2510098]
51. Maslowska E, Segijn CM, Vakeel KA, Viswanathan V. How consumers attend to online reviews: an eye-tracking and network analysis approach. Int J Advert 2020 Feb 17;39(2):282-306. [doi: 10.1080/02650487.2019.1617651]
52. Shan Y, Ji M, Xie W, Lam KY, Chow CY. Public trust in artificial intelligence applications in mental health care: topic modeling analysis. JMIR Hum Factors 2022 Dec 2;9(4):e38799. [doi: 10.2196/38799] [Medline: 36459412]

XSL•FO
RenderX

53. Shan W, Wang Y, Luan J, Tang P. The influence of physician information on patients' choice of physician in mHealth services using China's Chunyu Doctor app: eye-tracking and questionnaire study. JMIR Mhealth Uhealth 2019 Oct 23;7(10):e15544. [doi: 10.2196/15544] [Medline: 31647466]
54. Montazami A, Ann Pearson H, Kenneth Dubé A, Kacmaz G, Wen R, Shajeen Alam S. Why this app? How educators choose a good educational app. Comput Educ 2022 Jul;184:104513. [doi: 10.1016/j.compedu.2022.104513]
55. Ponnusamy M, Divya Venkatesh J, Das S, Nanda G. Examining the role of customer reviews in online purchasing using eye tracking. In: HCI International 2025 Posters: Springer; 2025. [doi: 10.1007/978-3-031-94153-5_26]
56. Guerra M, Milanese R, Deodato M, Perozzi V, Fasano F. Visual attention and privacy indicators in android: insights from eye tracking. Presented at: 10th International Conference on Information Systems Security and Privacy; Feb 26-28, 2024. [doi: 10.5220/0012437600003648]
57. Furman S, Theofanos M. Preserving privacy—more than reading a message. In: Hutchison D, Kanade T, Kittler J, editors. Universal Access in Human-Computer Interaction: Design for All and Accessibility Practice: Springer International Publishing; 2014:14-25. [doi: 10.1007/978-3-319-07509-9_2]
58. Pan Y, Ruan Y, Chang M, Lyu D, Li Y. Read or skip privacy policies when installing apps on wearable devices: the roles of perceived necessity and threat clues. Humanit Soc Sci Commun 2024;11(1). [doi: 10.1057/s41599-024-02989-4]
59. Duchowski A. Eye Tracking Methodology: Springer; 2007.
60. Chen T, Samaranayake P, Cen X, Qi M, Lan YC. The impact of online reviews on consumers' purchasing decisions: evidence from an eye-tracking study. Front Psychol 2022;13(865702):865702. [doi: 10.3389/fpsyg.2022.865702] [Medline: 35756238]
61. Skaramagkas V, Giannakakis G, Ktistakis E, et al. Review of eye tracking metrics involved in emotional and cognitive processes. IEEE Rev Biomed Eng 2023;16:260-277. [doi: 10.1109/RBME.2021.3066072] [Medline: 33729950]
62. Fu H, Li Y, Wu Z, Tan Y, He W, Zuo J. Immunization effect of inoculation messages: interventions against consensus and misinformation to improve public acceptance of recycled water. Humanit Soc Sci Commun 2025;12(1). [doi: 10.1057/s41599-025-05675-1]
63. Chaiken S. Heuristic versus systematic information processing and the use of source versus message cues in persuasion. J Pers Soc Psychol 1980;39(5):752-766. [doi: 10.1037/0022-3514.39.5.752]
64. Wedel M, Pieters R. Eye tracking for visual marketing. Found Trends Mark 2008 Aug 19;1(4):231-320. [doi: 10.1561/1700000011]
65. Armel KC, Beaumel A, Rangel A. Biasing simple choices by manipulating relative visual attention. Judgm Decis Mak 2008 Jun;3(5):396-403. [doi: 10.1017/S1930297500000413]
66. Krajbich I, Armel C, Rangel A. Visual fixations and the computation and comparison of value in simple choice. Nat Neurosci 2010 Oct;13(10):1292-1298. [doi: 10.1038/nn.2635] [Medline: 20835253]
67. Jagemann I, Stegemann M, von Brachel R, Hirschfeld G. Gender differences in preferences for mental health apps in the general population—a choice-based conjoint analysis from Germany. BMC Psychiatry 2024 Oct 14;24(1):682. [doi: 10.1186/s12888-024-06134-y] [Medline: 39402505]
68. Aras S, Drakos C, Manimangalam V, et al. Influencing public acceptance of artificial intelligence (AI) in healthcare delivery. Front Digit Health 2025;7:1664345. [doi: 10.3389/fdgth.2025.1664345] [Medline: 41623889]
69. Méndez-Suárez M, Monfort A, Hervas-Oliver JL. Are you adopting artificial intelligence products? Social-demographic factors to explain customer acceptance. Eur Res Manag Bus Econ 2023 Sep;29(3):100223. [doi: 10.1016/j.iedeen.2023.100223]
70. How to position participants and the eye tracker. Tobii Pro. 2023. URL: https://connect.tobii.com/s/article/how-to-position-participants-and-the-eye-tracker?language=en_US [accessed 2025-05-05]
71. Tobii Pro Spark: product description. Tobii. 2024. URL: https://go.tobii.com/tobii-pro-spark-product-description [accessed 2025-05-16]
72. Gerbing DW, Anderson JC. An updated paradigm for scale development incorporating unidimensionality and its assessment. J Mark Res 1988 May;25(2):186-192. [doi: 10.1177/002224378802500207]
73. Rossiter JR. The C-OAR-SE procedure for scale development in marketing. Int J Res Mark 2002 Dec;19(4):305-335. [doi: 10.1016/S0167-8116(02)00097-6]
74. Brislin RW. Back-translation for cross-cultural research. J Cross Cult Psychol 1970 Sep;1(3):185-216. [doi: 10.1177/135910457000100301]
75. Fuller CM, Simmering MJ, Atinc G, Atinc Y, Babin BJ. Common methods variance detection in business research. J Bus Res 2016 Aug;69(8):3192-3198. [doi: 10.1016/j.jbusres.2015.12.008]
76. Purnawirawan N, Eisend M, De Pelsmacker P, Dens N. A meta-analytic investigation of the role of valence in online reviews. J Interact Mark 2015 Aug;31(1):17-27. [doi: 10.1016/j.intmar.2015.05.001]
77. Hyzy M, Bond R, Mulvenna M, et al. Don't judge a book or health app by its cover: user ratings and downloads are not linked to quality. PLoS One 2024;19(3):e0298977. [doi: 10.1371/journal.pone.0298977] [Medline: 38437233]
78. de Chantal PL, Chagnon A, Cardinal M, Faieta J, Guertin A. Evidence of user-expert gaps in health app ratings and implications for practice. Front Digit Health 2022;4:765993. [doi: 10.3389/fdgth.2022.765993] [Medline: 35252957]

XSL•FO
RenderX

79. Levine DM, Co Z, Newmark LP, et al. Design and testing of a mobile health application rating tool. NPJ Digit Med 2020;3(74):74. [doi: 10.1038/s41746-020-0268-9] [Medline: 32509971]
80. Zin K, Kim S, Kim HS, Feyissa IF. A study on technology acceptance of digital healthcare among older Korean adults using extended TAM (extended technology acceptance model). Adm Sci 2023;13(2):42. [doi: 10.3390/admsci13020042]
81. Rousseau DM, Sitkin SB, Burt RS, Camerer C. Not so different after all: a cross-discipline view of trust. Acad Manage Rev 1998 Jul;23(3):393-404. [doi: 10.5465/amr.1998.926617]
82. Mayer RC, Davis JH, Schoorman FD. An integrative model of organizational trust. Acad Manage Rev 1995 Jul;20(3):709. [doi: 10.2307/258792]
83. Beldad AD, Hegner SM. Expanding the technology acceptance model with the inclusion of trust, social influence, and health valuation to determine the predictors of German users' willingness to continue using a fitness app: a structural equation modeling approach. Int J Hum Comput Interact 2018 Sep 2;34(9):882-893. [doi: 10.1080/10447318.2017.1403220]
84. Benbasat I, Wang W. Trust in and adoption of online recommendation agents. J Assoc Inform Syst 2005;6(3):72-101. [doi: 10.17705/1jais.00065]
85. Hegner SM, Beldad AD, Brunswick GJ. In automatic we trust: investigating the impact of trust, control, personality characteristics, and extrinsic and intrinsic motivations on the acceptance of autonomous vehicles. Int J Hum Comput Interact 2019 Nov 26;35(19):1769-1780. [doi: 10.1080/10447318.2019.1572353]
86. Zhang T, Tao D, Qu X, et al. Automated vehicle acceptance in China: social influence and initial trust are key determinants. Transp Res Part C Emerg Technol 2020 Mar;112:220-233. [doi: 10.1016/j.trc.2020.01.027]
87. Kim E, Lin JS, Sung Y. To app or not to app: engaging consumers via branded mobile apps. J Interact Advert 2013 Jan;13(1):53-65. [doi: 10.1080/15252019.2013.782780]
88. Arianezhad M, Camp LJ, Kelley T, Stebila D. Comparative eye tracking of experts and novices in web single sign-on. In: CODASPY '13: Proceedings of the Third ACM Conference on Data and Application Security and Privacy: Association for Computing Machinery; 2013:105-116. [doi: 10.1145/2435349.2435362]
89. Whalen K, Inkpen M. Gathering evidence: use of visual security cues in web browsers. In: GI '05: Proceedings of Graphics Interface 2005: Canadian Human-Computer Communications Society, University of Waterloo; 2005:137-144. [doi: 10.5555/1089508.1089532]
90. Alhammad N, Alajlani M, Abd-Alrazaq A, Epiphaniou G, Arvanitis T. Patients' perspectives on the data confidentiality, privacy, and security of mHealth apps: systematic review. J Med Internet Res 2024 May 31;26:e50715. [doi: 10.2196/50715] [Medline: 38820572]
91. Auxier B, Rainie L, Anderson M, Perrin A, Kumar M, Turner E. Americans' attitudes and experiences with privacy policies and laws. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/2019/11/15/americans-attitudes-and-experiences-with-privacy-policies-and-laws/ [accessed 2026-01-09]
92. DiGA und DiPA datenschutzkriterien. Bundesinstitut für Arzneimittel und Medizinprodukte. 2024. URL: https://www.bfarm.de/DE/Medizinprodukte/Aufgaben/DiGA-und-DiPA/Datenschutzkriterien/_node.html [accessed 2025-11-25]
93. Hwang YM, Lee KC. Using an eye-tracking approach to explore gender differences in visual attention and shopping attitudes in an online shopping environment. Int J Hum Comput Interact 2018 Jan 2;34(1):15-24. [doi: 10.1080/10447318.2017.1314611]
94. Kauttonen J, Rousi R, Alamäki A. Trust and acceptance challenges in the adoption of AI applications in health care: quantitative survey analysis. J Med Internet Res 2025 Mar 21;27:e65567. [doi: 10.2196/65567] [Medline: 40116853]
95. Jagemann I, Hegner S, Hirschfeld G. Dataset for: augmenting self-reports: using eye-tracking and questionnaire data to elucidate the role of rating valence in AI skin cancer app acceptance. Zenodo. 2026. URL: https://zenodo.org/records/19553460 [accessed 2026-06-08]

Abbreviations

AI: artificial intelligence
AISCSA: artificial intelligence–based skin cancer screening app
AOI: area of interest
BI: behavioral intention to use
PEOU: perceived ease of use
PU: perceived usefulness
TAM: technology acceptance model
WTP: willingness to pay

XSL•FO
RenderX

Edited by NV Kalckreuth; submitted 13.Feb.2026; peer-reviewed by H Fu, Ö SEBETCi; revised version received 19.Apr.2026; accepted 19.Apr.2026; published 11.Jun.2026.

Please cite as:
Jagemann I, Hegner S, Hirschfeld G
The Role of Rating Valence in AI Skin Cancer App Acceptance: Eye-Tracking and Questionnaire Study
JMIR Hum Factors 2026;13:e93489
URL: https://humanfactors.jmir.org/2026/1/e93489
doi:10.2196/93489

© Inga Jagemann, Sabrina Hegner, Gerrit Hirschfeld. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 11.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Quality and Characteristics of Digital Mental Health Apps: Mixed Methods Study

Maciej Marek Zych[1], PhD; Raymond Bond[1], PhD; Maurice Mulvenna[1], MPhil, PhD; Jorge Martinez Carracedo[1], MSc, PhD; Lu Bai[2], PhD; Tim Andrews[3], BA

[1]School of Computing, Ulster University, 2-24 York Street, Belfast, United Kingdom

[2]School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast, United Kingdom

[3]ORCHA, Daresbury, United Kingdom

Corresponding Author:
Maciej Marek Zych, PhD
School of Computing, Ulster University, 2-24 York Street, Belfast, United Kingdom

Abstract

Background: There are around 20,000 mental health apps available in app stores. The Organisation for the Review of Care and Health Apps (ORCHA), a United Kingdom digital health compliance company, has assessed a number of digital mental health apps with regard to their quality, professional and clinical assurance, data privacy, and user experience. This study analyzes the data that were collected by ORCHA when they assessed mental health apps.

Objective: This study aimed to examine the characteristics of mental health apps regarding their quality, target users, features, underpinning evidence, and data privacy.

Methods: A dataset comprising ORCHA Baseline Review assessments of over 2000 digital health apps, including 436 mental health apps, was used. This study uses exploratory data analysis to gain insight into the quality and characteristics of mental health apps. Methods such as descriptive and inferential statistics, k-modes clustering, and association rule mining were used to explore the quality of mental health apps as well as reveal insights into the different cost types, target users, app features, data types, and evidence of app content.

Results: Information provision, data capture, and data sharing were the most common features within the 436 mental health apps. The examined apps primarily targeted the following groups: adults (n=229, 52.5%), everyone (n=184, 42.2%), and teens (n=135, 31%). The cost of apps has not been linked to the quality of mental health apps, although paid apps or apps with in-app purchases may include additional services. Indicated user acceptance or benefit is the most common type of evidence provided by these mental health apps. A total of 241 (55.3%) apps included a qualified professional in app development, and 251 (57.6%) apps provided evidence within the app that the developer validated any guidance with relevant reliable information sources or references. Usage data and email were the most commonly collected data types. Association rule mining showed that email, IP address, name, and usage data are often co-collected by the same apps. K-modes cluster analysis showed that mental health apps can be categorized into 2 clusters, where one cluster of apps (n=182, 41.7%) collected more data than apps in the other cluster.

Conclusions: Mental health apps are commonly targeted for everyone to use, but many apps are targeted toward teens or adults. Our study suggests that many publicly available mental health apps did not take the precautions (such as the involvement of appropriate health professionals, literature references, or conducting tests) to ensure that their content is valid and research based. Greater effort on behalf of mental health app developers is needed to ensure that the public is provided with high-quality apps. Moreover, our study indicates that the mental health apps that collect more data tend to score better on the ORCHA Baseline Review assessment.

(JMIR Hum Factors 2026;13:e67944) doi:10.2196/67944

KEYWORDS

digital mental health; mental health apps; machine learning; health informatics; data privacy; mobile health

Introduction

According to a World Health Organization report [1], there are nearly 1 billion people living with a diagnosed mental health condition, and most do not have access to effective care. Moreover, during the COVID-19 pandemic, mental health issues became more predominant and prevalent [2-4] and have remained prevalent as of 2024 [1]. Mental health conditions affect many people's lives; for example, a study from 2020 [5] showed that social stigma is a contributing factor to unemployment in people living with mental health issues. With

XSL•FO
RenderX

the decline in mental health, it is important to mitigate a global mental health crisis.

While apps might be useful, it is important to understand their quality. For example, a study from 2020 has shown that college counseling centers in the United States often recommend unsafe and not up-to-date mental health apps to students [6]. Another paper from 2016 [7] has shown that suicide prevention apps may also provide potentially harmful content. Moreover, Sander et al [8] indicated that the public can have difficulty identifying high-quality apps for posttraumatic stress disorder. Beyond apps, other digital tools also require further investigation; for example, social media has been linked to an increase in mental health issues [9,10].

Arguably, the most accessible tools for mental health care in the public are mental health apps or websites [11]. Estimates show that there are around 20,000 mental health apps available in app stores [11,12]. However, it is important to identify mental health apps based on their quality before health care professionals recommend them to patients or clients.

The Organisation for the Review of Care and Health Apps (ORCHA) [13] is a United Kingdom–based digital health compliance company that specializes in the quality assessment of digital health apps. ORCHA is currently working with 70% of National Health Service organizations within England and provides digital health app libraries, hosted by various health care organizations, that contain information about digital health apps and their quality. The ORCHA Baseline Review (OBR) is a proxy for digital health apps' compliance with best practice standards. Specifically, the OBR assessment results provide information (including an assessment score between 0 and 100) regarding a digital health app's compliance with best practices in the domains of professional and clinical assurance (PCA), data privacy (DP), and user experience (UX), allowing end users and clinical professionals to make informed decisions on whether to use or recommend these digital health apps. National Institute for Health and Care Excellence (NICE) [14] provides guidance on digital health app research. ORCHA uses Evidence Standard Framework (ESF) [15], developed by NICE, to categorize digital health apps, and the ESF has been influential in the development of the OBR assessment tool.

It has been suggested that mental health apps can enhance but not replace existing mental health services [16]. Similarly to a 2023 study [16], a meta-analysis from 2021 [17] studied the treatment of depression using apps and found that there was no difference in the outcomes between human-guided digital interventions and face-to-face psychotherapy. For apps to do more good than harm, it is important to recommend only mental health apps of sufficient quality for public use. A meta-review of meta-analyses from 2020 stated that more studies related to the quality of mental health apps are needed [18]. Previous studies have assessed the UX, engagement, and popularity of mental health apps [19,20]. This paper also extends the work conducted in 2023 [21], which examined the quality of digital health apps across different health care domains or categories.

Expanding on these studies, the objective of this study was to examine the characteristics of mental health apps, specifically regarding their quality, the level of underpinning evidence, and aspects related to DP. ORCHA has provided a dataset of OBR results and documentation from their assessment of 436 digital mental health apps. While there are studies regarding the use of conversational agents for mental health [22,23], these were not included in this study. To achieve our objective, the following research questions were examined:

1. How do mental health apps score on the OBR assessment tool (as indicated by ORCHA, PCA, UX, and DP scores)?
2. Is there a difference in ORCHA, PCA, UX, or DP scores among mental health apps that are free and those that are paid or contain in-app purchases?
3. Who are the target users of mental health apps?
4. What features are included in mental health apps, and which are linked to the ORCHA score?
5. What evidence is presented to validate the content of mental health apps?
6. What are mental health apps' use and development characteristics?
7. What data are collected by mental health apps?

By answering the abovementioned questions, we believe we will be able to broadly report on the characteristics of digital mental health apps. This will allow us to identify areas of concern where improvements can and should be made regarding digital health app quality.

Methods

The Dataset and Tools

ORCHA provided an assessment dataset of 2127 digital health apps. All the apps were assessed using the OBR tool, Version 6 (OBR V6). The assessment and app categorization of all the digital health apps were conducted by 2 ORCHA reviewers. A third reviewer was included only if there was any dispute between the original 2 reviewers during an assessment. It takes around 6 months for an ORCHA reviewer to be trained on how to use the OBR tool and to be considered ready to carry out live reviews using the tool. The training involves teaching the new reviewer about each area (UX, PCA, and DP) of the OBR. Training is carried out either in person or via online meetings. The dataset used included digital health app assessments that were published between January 18, 2021, and January 6, 2022.

From the 2127 digital health apps, a subset of 436 mental health apps was created and used for this study. The only selection criterion was that an app had been categorized as a "mental health app." Multimedia Appendix 1 states more specific categories (mutually inclusive) of mental health apps that were included.

The OBR V6 tool is the latest version of the ORCHA assessment tool, which consists of approximately 300 objective (mostly dichotomous yes/no) assessment questions in 3 domain areas: PCA, DP, and UX. Each of the areas is scored individually on a scale from 0 to 100 and combined into an overall ORCHA score. Moreover, the collected data of OBR assessments of mental health apps included information such as target users, features, and associated cost types (free, paid, and in-app purchases). Information relating to app development was also collected, such as whether there is a suitably qualified

XSL•FO
RenderX

professional involved in the development team of the app. All the collected data allowed us to examine what sort of evidence is being provided regarding the content of mental health apps.

NICE, a UK-based organization, provides guidance relating to health technologies, including health apps. NICE has produced an ESF that can be used to classify digital health apps using A, B, and C tiers based on their functionality, risk, and regulatory status. All the assessed mental health apps were assigned a NICE ESF tier. Tier A indicates that the digital health apps provide health and social care services with no measurable user outcome. Tier B denotes that the digital health apps can provide a 2-way communication between users and health care professionals and provide health care information or a health diary. Tier C indicates that digital health apps provide preventative behavioral change aimed at health issues; they may allow users to self-manage a specific condition, indicate that the digital health apps provide or guide treatment for a condition, record and transmit data about this condition to a professional, caregiver, or third party without a user's input, contain a calculator that impacts treatment, provide diagnostics for a specific condition, or guide a diagnosis. Since the NICE tiers are dependent upon functionality, risk, and regulatory status, it would be inappropriate to, for example, use OBR on a tier C digital health app using OBR tier B score weighting. See Checklist 1 for included topics and items in this paper.

Data Analysis

Descriptive statistics were used to study the data. In this study, box plots, line plots, and heatmaps were used for the exploratory visualization of data. Descriptive statistics were used when conducting analysis of OBR scores, cost types, target users, features, presented evidence, use and development characteristics, and collected data types. Quantiles were used to represent the spread of domain (OBR) scores from 0% to 100% in increments of 25%. Median and IQR were calculated for some of the data.

The Wilcoxon rank sum test was used to compare 2 samples of data to determine any statistically significant differences between them. A *P* value less than or equal to .05 was considered statistically significant. A Bonferroni-corrected α value was calculated when multiple hypothesis testing was conducted. This was used when comparing ORCHA scores between apps that had a specific feature and those that did not.

A random forest algorithm was used to determine feature importance, with mean square error (increase in mean squared error) and node purity (increase in node purity) calculations, when predicting the overall ORCHA score when using app features as predictors (eg, goal setting, gamification).

Association rule mining (a priori algorithm) was used to identify commonly co-occurring types of data collected by mental health apps. K-modes cluster analysis was performed on the data types. K-modes was chosen as a clustering method because the data were recorded in binary format. An optimal number of clusters was determined using the NbClust package in R [24]. The package uses 30 indices to determine the best number of clusters for a dataset.

Ethical Considerations

This secondary data analysis study gained ethical approval from Ulster University (Ethics Filter Committee, Faculty of Computing, Engineering and the Built Environment, project number: CEBE_RE-22 - 002). The process undertaken by ORCHA ensures that digital health app developers are aware of their score and are given time to contest findings of the assessment, which may be amended if developers provide additional relevant information. All reviews, unless explicitly requested to be removed by the developer, are covered as suitable for research in ORCHA's privacy policy [25]. Informed consent was obtained by ORCHA to use the data from the assessments for the purposes of research, provided that the assessed apps are anonymized in publications.

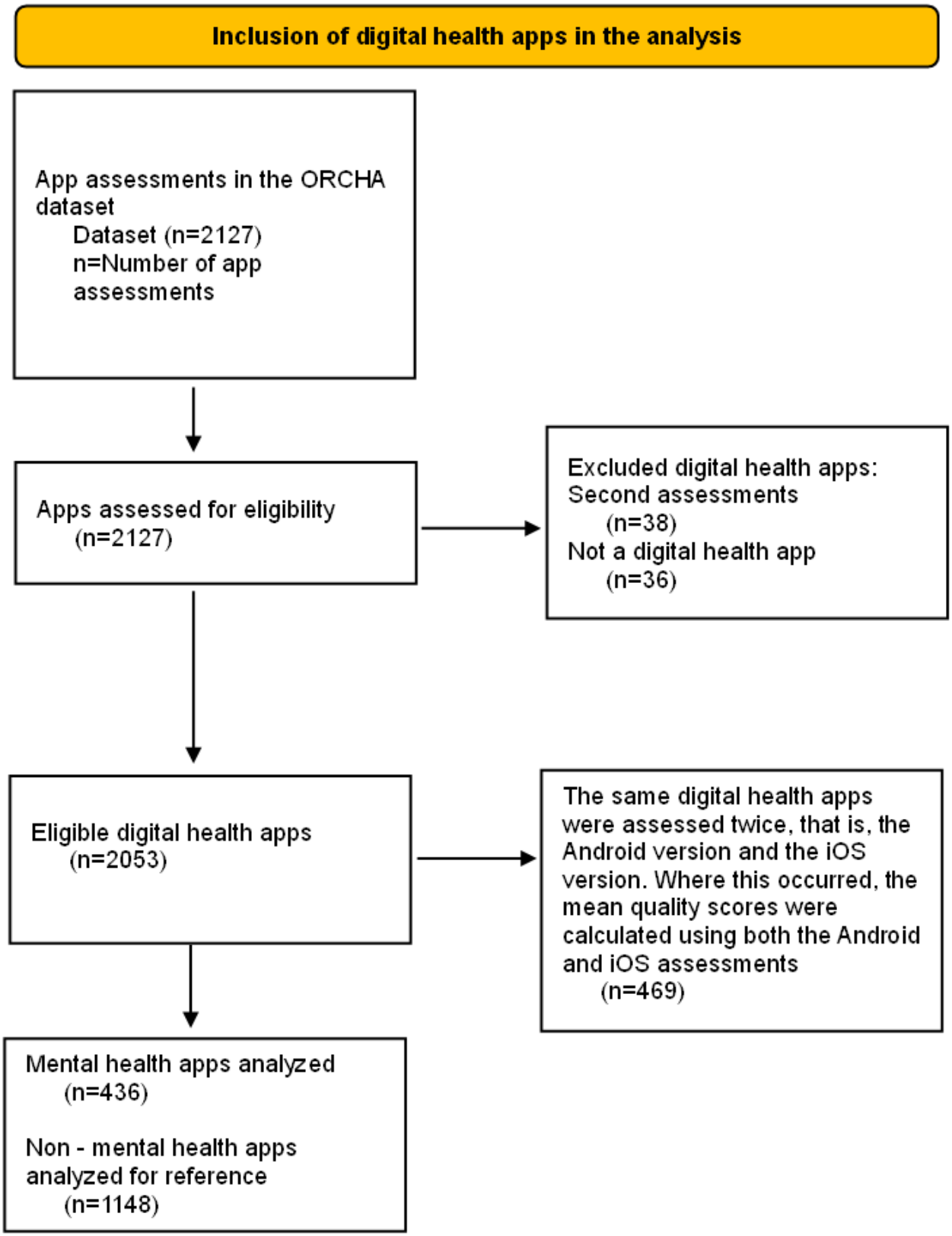
Results

App Selection

Figure 1 shows the process of selecting mental health apps from the ORCHA dataset.

XSL•FO
RenderX

Figure 1. Mental health app selection diagram. ORCHA: Organisation for the Review of Care and Health Apps



XSL•FO
RenderX

How Do Mental Health Apps Score on the OBR Assessment Tool (as Indicated by ORCHA, PCA, UX, and DP Scores)?

Table 1 shows domain scores per NICE ESF tier. Color coding was applied as follows: red for values greater than 75, green for values from 65 to 75, and blue for values less than 65. Since there were only 2 mental health apps assigned to tier A, they were not included in Table 1.

As can be seen in Table 1, the median ORCHA score for 436 mental health apps was 60 (Tier B: 62 and Tier C: 55). ORCHA, together with the National Health Service, determined that an ORCHA threshold score of 65 indicates adequate compliance with the assessment criteria to be recommended for use by the public. Most mental health apps do not meet that threshold. This finding demonstrates that greater effort is needed on behalf of developers to produce higher-quality mental health apps.

Figure 2 visualizes NICE ESF tiers per domain score with box plots. When separated into NICE ESF tiers, the sample sizes were as follows: tier A, 2; tier B, 332; and tier C, 102.

Table . Mental health apps' ORCHA scores per tier.

Tier and scores	Values, median (IQR; range)
All apps (n=436)	
ORCHA[a]	60 (51-73; 25.0-96.0)
UX[b]	75.2 (69.8-79.6; 27.4-92.7)
PCA[c]	46.1 (30.3-75.3; 7.57-98.5)
DP[d]	65.6 (52.5-73.1; 4.28-99.3)
Tier B (n=332)	
ORCHA	62 (52-73; 28.0-90.5)
UX	75.2 (70.2-79.6; 32.6-88.3)
PCA	47.7 (30.3-75.3; 11.4-95.3)
DP	65 (52.5-73.1; 16.5-95.3)
Tier C (n=102)	
ORCHA	55 (46.3-71; 25.0-96.0)
UX	72.5 (68.6-77.8; 27.4-92.7)
PCA	39.5 (31.1-68; 7.57-98.5)
DP	67.3 (51.7-75.3; 4.28-94.9)

[a]ORCHA: Organisation for the Review of Care and Health Apps.

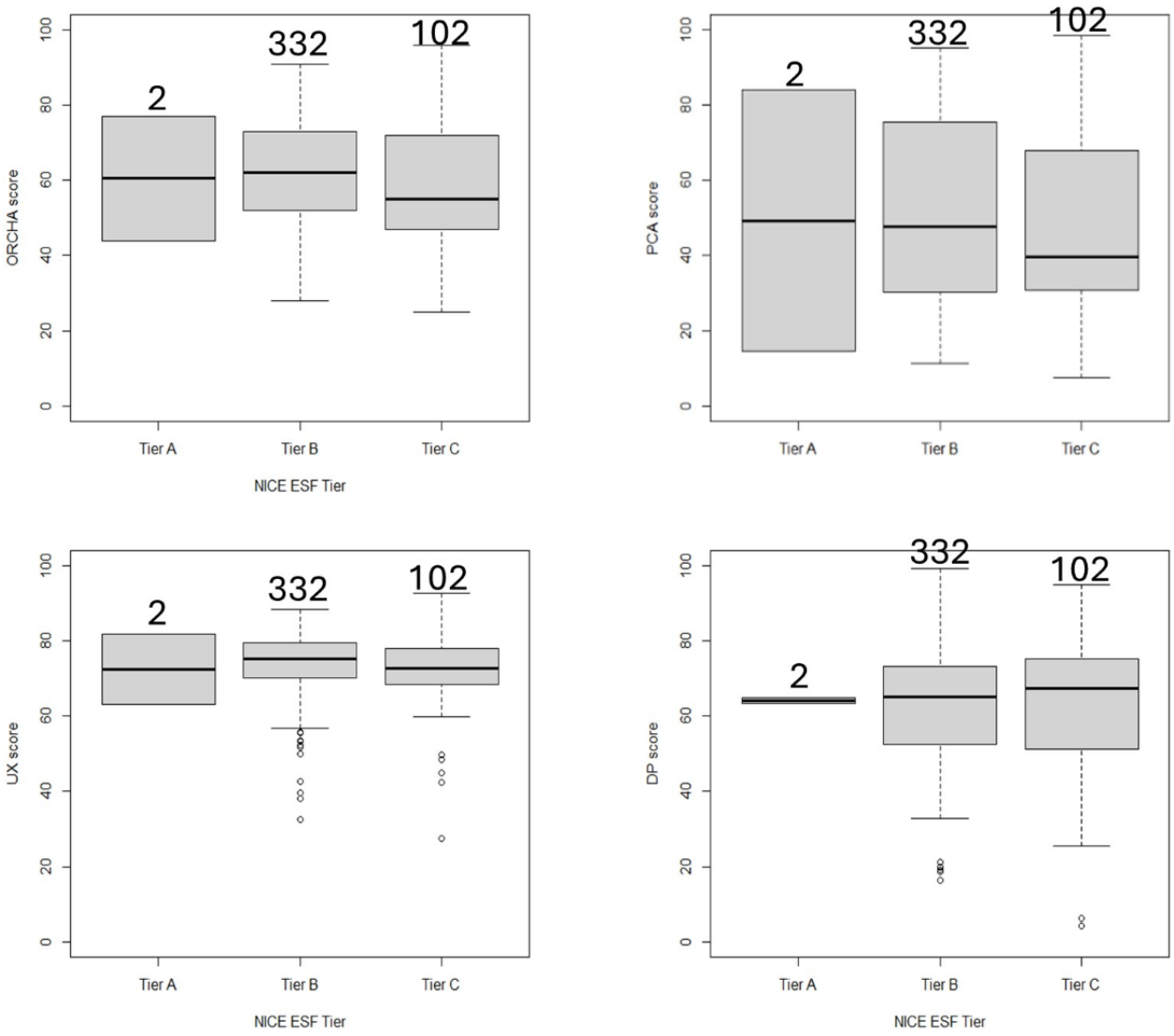
[b]UX: user experience.

[c]PCA: professional and clinical assurance.

[d]DP: data privacy.

XSL•FO
RenderX

Figure 2. Quality scores per assessment domain (user experience [UX], data privacy, and professional clinical assurance) and per National Institute for Health and Care Excellence (NICE) Evidence Standard Framework (ESF) Tier (Tier A, B, and C). DP: data privacy; ORCHA: Organisation for the Review of Care and Health Apps; PCA: professional and clinical assurance.



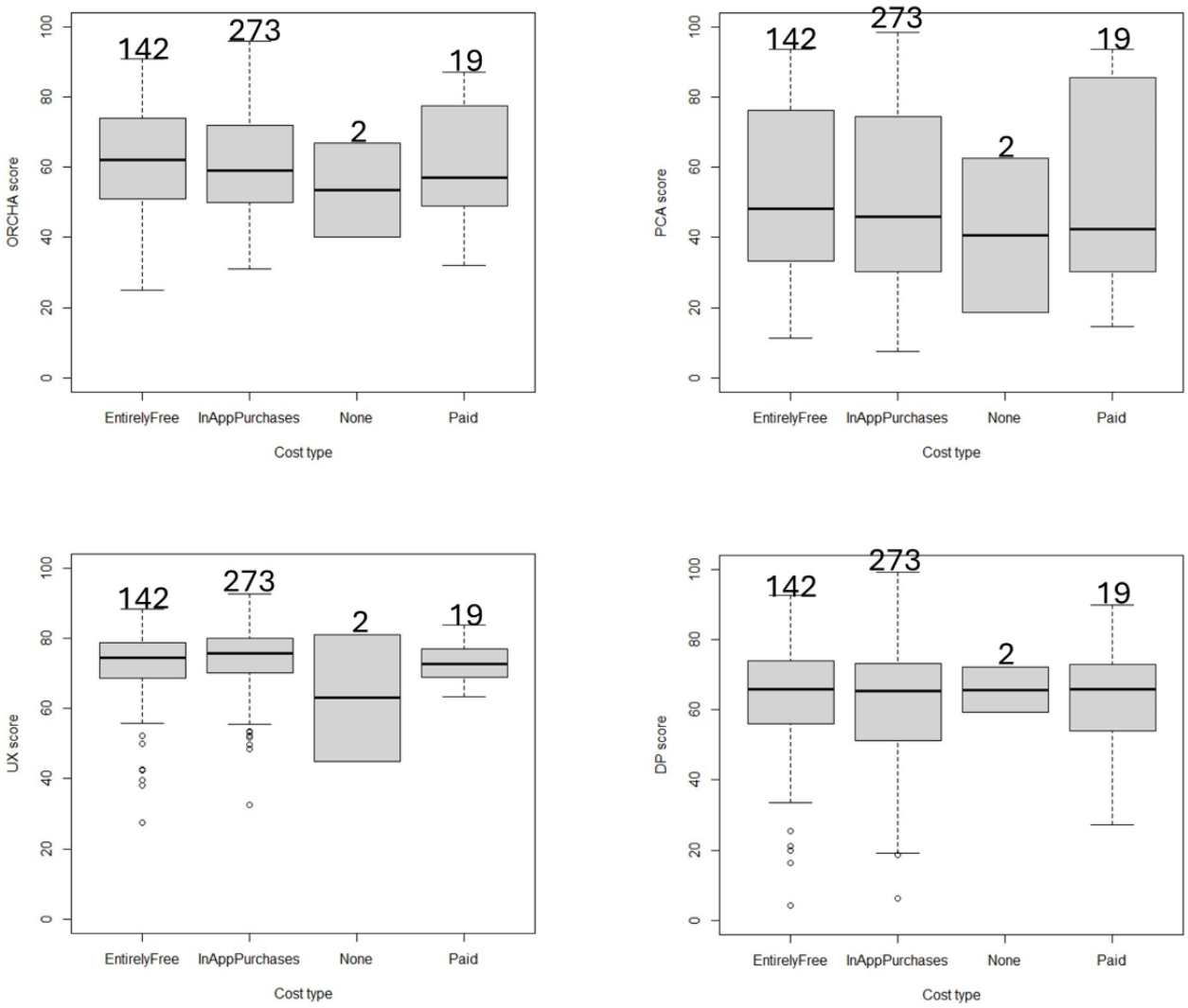
Is There a Difference in ORCHA Score, PCA, UX, or DP Among Mental Health Apps That Are Free and Those That Are Paid or Contain In-App Purchases?

Figure 3 shows domain scores against cost type. As can be seen from the box plots, the quality of a mental health app does not appear to be influenced by cost. This is further confirmed by the Wilcoxon rank sum test, where none of the cost types achieved a statistically significant ($P \leq .05$) difference when compared for each of the domain scores. When separated by cost types, the sample sizes were as follows: none: 2, paid: 19, entirely free: 142, and in-app purchases: 273.

This study has shown that there appears to be no difference in OBR scores in relation to cost. Figure 3 shows that the distributions of OBR scores (ORCHA, PCA, UX, and DP) for mental health apps that are free, paid, and those with in-app purchases are similar, with a nonstatistically significant ($P \leq .05$) difference when using the Wilcoxon 2-sample rank sum test. Nevertheless, it is possible that paid apps or those with in-app purchases may include features not offered by those apps that are available for free. Further study is needed to determine whether paid apps or those with in-app purchases offer features or services that are unavailable in free digital mental health apps.

XSL•FO
RenderX

Figure 3. Cost analysis box plots per domain score. DP: data privacy; ORCHA: Organisation for the Review of Care and Health Apps; PCA: professional and clinical assurance; UX: user experience.



Who Are the Target Users of Mental Health Apps?

Table 2 shows the target users of the 436 mental health apps that were included in this study. Target users are presented in descending order of frequency. The table shows that most mental health apps appear to be targeted at adults, followed by everyone and teens.

Table . Mental health apps' target users (N=436).

Target users	Frequency, n (%)
Adults	229 (52.5)
Everyone	184 (42.2)
Teenagers	135 (31)
Older adults	51 (11.7)
Children	27 (6.19)
Preteens	20 (4.59)
Clinicians	5 (1.15)
Patients	2 (0.459)
Infants	2 (0.459)
Health care administrators	1 (0.229)
Caregivers	1 (0.229)

XSL•FO
RenderX

What Features Are Included in Mental Health Apps and Which Are Linked to ORCHA Score?

In this section, we aim to identify features that may improve the use and quality of digital mental health apps. Table 3 shows the features that were recorded for the 436 mental health apps included in this study and their frequency. Features are presented in descending order of frequency. The table indicates that the most common features are information provision, followed by data capture and data sharing.

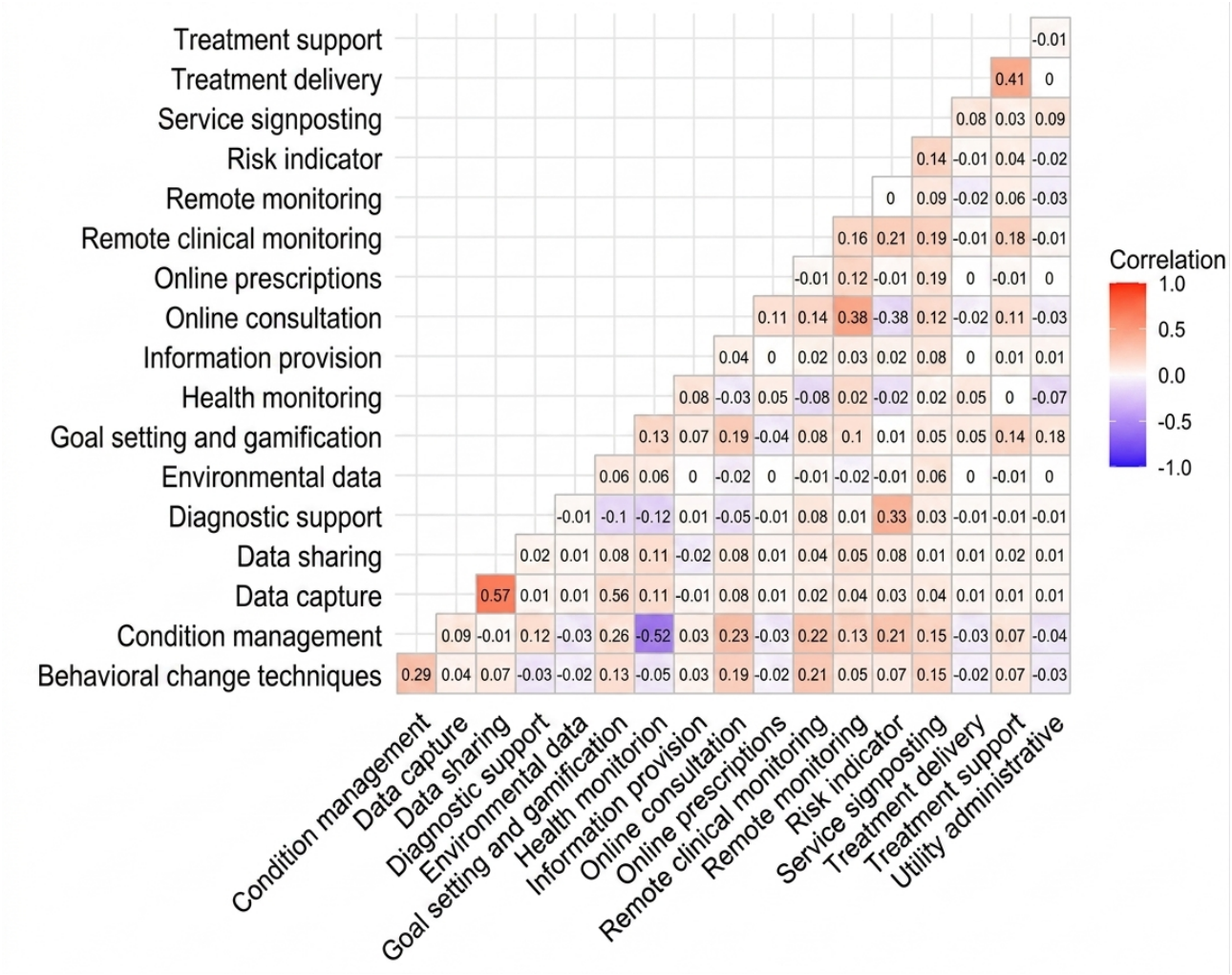
Figure 4 depicts pairwise Spearman correlations among the features. The highest being between data capture and data sharing, with a Spearman correlation of 0.57. Figure 5 depicts the importance of features (increase in mean squared error and increase in node purity), with ORCHA scores as the independent variable and features as the dependent variables. Multimedia Appendix 2 depicts feature importance and correlation for each tier.

Table . Feature frequency (N=436).

Features	Frequency, n (%)
Information provision	433 (99.3)
Data capture	431 (98.9)
Data sharing	421 (96.6)
Health monitoring	215 (49.3)
Goal setting and gamification	184 (42.2)
Service signposting	159 (36.5)
Condition management	107 (24.5)
Online consultation	69 (15.8)
Remote monitoring	62 (14.2)
Behavioral change techniques	56 (12.8)
Risk indicator	34 (7.80)
Remote clinical monitoring	17 (3.90)
Treatment support	6 (1.38)
Diagnostic support	6 (1.38)
Utility/administrative	2 (0.459)
Treatment delivery	1 (0.229)
Online prescriptions	1 (0.229)
Environmental data	1 (0.229)

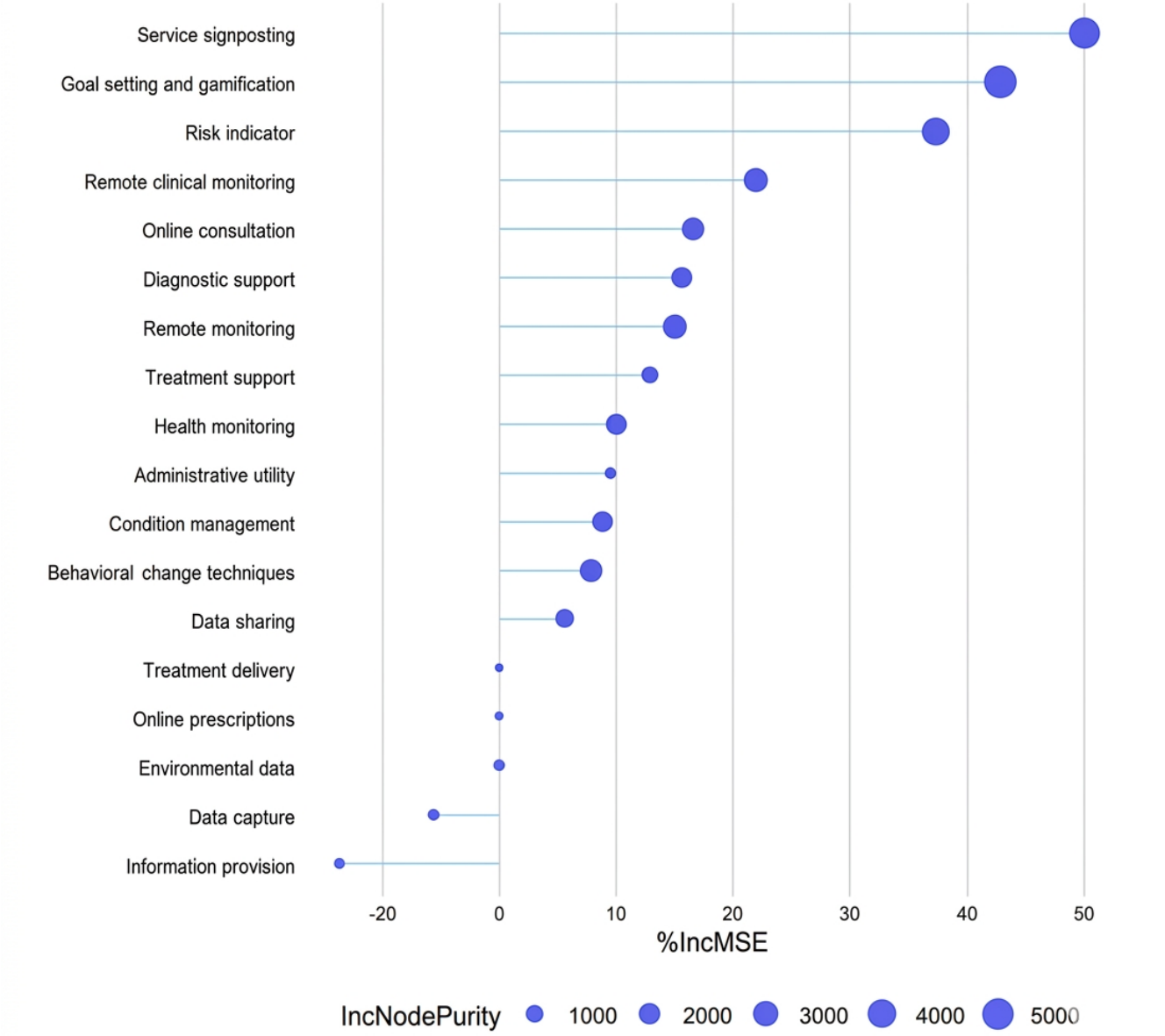
XSL•FO
RenderX

Figure 4. Spearman correlations among features.



XSL•FO
RenderX

Figure 5. Feature importance with random forest for the Organisation for the Review of Care and Health Apps (ORCHA) score. incMSE: increase in mean squared error; incNodePurity: increase in node purity.



Random forest feature importance was used to identify which features are linked to the ORCHA score (Figure 5). Service signposting and goal setting and gamification were identified as the 2 most important features that are positively associated with the ORCHA score, whereas the risk indicator was negatively associated with the ORCHA score.

Table 4 depicts the top 3 features for mental health apps altogether and when partitioned by NICE ESF tier (tier B and tier C). It also states the number of apps with and without the feature, with their ORCHA scores' median and IQR. The Wilcoxon rank sum test with a Bonferroni-corrected α value was calculated to check for statistically significant differences between the 2 distributions of apps regarding the ORCHA score among apps with and without the feature. See Multimedia Appendices 2 and 3 for more information regarding features for each tier and data visualizations.

XSL•FO
RenderX

Table . Top 3 features for each tier. Bonferroni corrected α value is .05/9 ≈ .006.

Tier and feature	Apps with feature, n	Apps with feature ORCHA[a] score, median (IQR)	Apps without feature, n	Apps without feature ORCHA score, median (IQR)	Wilcoxon test (*P* value)	Effect size (Cohen W)
All apps						
Service signposting	159	66 (55.0-76.5)	277	56 (48.0-71.0)	<.001	0.271
Goal setting and gamification	184	67.5 (54.0-78.0)	252	57 (48.0-68.0)	<.001	0.156
Risk indicator	34	49.5 (45.3 -57.0)	402	61 (51.0-74.0)	<.001	0.844
Tier B						
Service signposting	105	70 (59.0-77.0)	227	57 (49.0-70.0)	<.001	0.368
Goal setting and gamification	127	68 (54.5-78.0)	205	60 (49.0-70.0)	<.001	0.235
Online consultation	38	67 (60.5-74.0)	294	61 (51.0-73.0)	.01	0.771
Tier C						
Goal setting and gamification	57	62 (52.0-76.0)	45	51 (46.0-57.0)	<.001	0.118
Risk indicator	33	48 (45.0-57.0)	69	57 (51.0-76.0)	<.001	0.353
Behavioral change techniques	32	72 (55.0-82.0)	70	53 (46.0-59.5)	<.001	0.373

[a]ORCHA: Organisation for the Review of Care and Health Apps.

What Evidence Is Presented to Validate the Content of Mental Health Apps?

Table 5 depicts the types of evidence that were presented in the OBR assessment of mental health apps, separated by NICE ESF tier, and by their medical device status in Table 6. Some mental health apps can be classified as medical devices if they conduct diagnoses or provide treatments. The data indicate that indicated user acceptance (295/436, 67.7%) and observational studies (59/436, 13.5%) are the most commonly presented evidence types for the content of digital mental health apps. However, 118 of 436 (27.1%) digital mental health apps provided no evidence.

Tables 5 and 6 show that the most common evidence presented by mental health apps during the assessment is indicative of user acceptance or benefit (295/436, 67.7%). The rarest type of evidence presented was meta-analysis or systematic review (10/436, 2.29%). More evidence from academic and clinical sources may improve the quality of mental health apps. Relying mainly on user acceptance or benefit does not verify that the content of those apps is of clinical standard and value. Moreover, 118 out of 436 (27.1%) apps did not present any evidence. In the time of a mental health crisis and the accessibility of mental health apps, it should be a requirement for all mental health apps to include a qualified professional in the app development lifecycle, as well as validate any guidance provided by an app with relevant information sources and references to mitigate any risk associated with its use.

XSL•FO
RenderX

Table . Evidence presented in descending order of frequency per National Institute for Health and Care Excellence Evidence Standard Framework tier.

Tier and evidence type	Frequency, n (%)
Tier A (n=2)	
None	1 (50)
Survey	1 (50)
Observational	1 (50)
Tier B (n=332)	
Indicated user acceptance/benefit	229 (69)
None	96 (28.9)
Observational	35 (10.5)
Pilot study	23 (6.93)
Randomized controlled trial	18 (5.42)
Survey	13 (3.92)
Meta-analysis/systematic review	4 (1.20)
Tier C (n=102)	
Indicated user acceptance/benefit	66 (64.7)
Observational	23 (22.5)
None	21 (20.6)
Pilot study	16 (15.7)
Randomized controlled trial	14 (13.7)
Meta-analysis/systematic review	6 (5.88)
Survey	4 (3.92)

Table . Evidence presented in descending order of frequency per medical device.

Device and evidence type	Frequency, n (%)
Not a medical device (n=414)	
Indicated user acceptance/benefit	282 (68.1)
None	110 (26.6)
Observational	56 (13.5)
Pilot study	38 (9.18)
Randomized controlled trial	27 (6.52)
Survey	16 (3.86)
Meta-analysis/systematic review	10 (2.42)
Medical device (n=22)	
Indicated user acceptance/benefit	13 (59.1)
None	8 (36.4)
Randomized controlled trial	5 (22.7)
Observational	3 (13.6)
Survey	2 (9.09)
Pilot study	1 (4.55)
Meta-analysis/systematic review	0 (0)

Table 7 depicts the characteristics of evidence that were provided. The table presents which methods were used to validate mental health apps' evidence, separated by NICE ESF tiers. The key takeaway from the table appears to be that most mental health apps did not involve a qualified professional, and most of the presented evidence did not involve any calculation

XSL•FO
RenderX

of the *P* value, as it was deemed not applicable (NA). Table 7 shows that many of the mental health apps in this study did not provide information on the characteristics of the evidence, such as details on the sample size used, use of a comparator, or *P* values, indicating that even when evidence is presented for mental health apps, important related information is missing.

Table . Evidence characteristics across tiers.

Characteristics	Tier A (apps: n=2), n (%)	Tier B (apps: n=332), n (%)	Tier C (apps: n=102), n (%)
Is there a suitably qualified professional involved in the development team of the app?	Yes: 1 (50); No: 1 (50)	Yes: 169 (50.9); No: 163 (49.1)	Yes: 71 (69.6); No: 31 (30.4)
Is the sample size appropriate?	Yes: 1 (50); No: 0 (0%); N/A[a]: 1 (50)	Yes: 35 (10.5); No: 11 (3.31); N/A: 286 (86.1)	Yes: 23 (22.5); No: 7 (6.86); N/A: 72 (70.6)
Does the evidence found provide a *P* value?	Yes: 1 (50); No: 0 (0); N/A: 1 (50)	Yes: 36 (10.8); No: 10 (3.01); N/A: 286 (86.1)	Yes: 28 (27.5); No: 2 (1.96); N/A: 72 (70.6)
Does the *P* value demonstrate significance ($P<.05$)?	Yes: 1 (50); No: 0 (0); N/A: 1 (50)	Yes: 35 (10.5); No: 1 (0.301); N/A: 296 (89.2)	Yes: 27 (26.5); No: 1 (0.980); N/A: 74 (72.5)
Does the *P* value demonstrate near significance ($P<.2$)?	Yes: 0 (0); No: 0 (0); N/A: 2 (100)	Yes: 1 (0.301); No: 3 (0.904); N/A: 328 (98.8)	Yes: 1 (0.980); No: 1 (0.980); N/A: 100 (98)
Is there a comparator?	Yes: 1 (50); No: 0 (0); N/A: 1 (50)	Yes: 28 (8.43); No: 18 (5.42); N/A: 286 (86.1)	Yes: 26 (25.5); No: 4 (3.92); N/A: 72 (70.6)
Is the comparator validated?	Yes: 0 (0); No: 1 (50); N/A: 1 (50)	Yes: 9 (2.71); No: 21 (6.33); N/A: 302 (91.0)	Yes: 9 (8.82); No: 17 (16.7); N/A: 76 (74.5)

[a]Not applicable.

What Are Mental Health Apps' Use and Development Characteristics

Table 8 depicts the characteristics of mental health apps regarding their use and development. This table states what key information is available to users of digital mental health apps and whether key professionals were involved. This study shows that 241 out of 436 (55.3%) apps included a qualified professional in app development. A total of 251 out of 436 (57.6%) apps provided evidence within the app that the developer validated any guidance with relevant reliable information sources or references.

XSL•FO
RenderX

Table . Mental health apps characteristics regarding their use and development.

Question	MH[a] apps that answered the question, n	Yes, n (%)	No, n (%)
Is there a suitably qualified professional involved in the development team of the app?	436	241 (55.3)	195 (44.7)
Does the organization behind the app have relevant credentials?	436	26 (5.96)	410 (94)
Is there evidence of an endorsement by a relevant body?	436	78 (17.9)	358 (82.1)
Are organizations using the app?	436	110 (25.2)	326 (74.8)
Is there a statement that it has been positively evaluated or validated by a relevant health care professional?	436	76 (17.4)	360 (82.6)
Is there evidence within the app that the developer has validated any guidance with relevant reliable information sources or references?	436	251 (57.6)	185 (42.4)
Is there a statement or any evidence showing that appropriate safeguarding measures are in place around peer-support and other communication functions within the platform?	171	141 (82.5)	30 (17.5)
Does the developer make clear risks associated with using the app?	436	27 (6.19)	409 (93.8)
Is there a privacy policy clearly available via the app/web app/website?	374	364 (97.3)	10 (2.67)
Does the developer fully inform the user of how they will collect data about them?	407	376 (92.4)	31 (7.62)
Is there a statement that confirms the app's compliance to GDPR[b] 2016 (or Data Protection Act 2018)?	371	160 (43.1)	211 (56.9)
Are users informed of their rights with regard to their data? Are users clearly informed of the individual privacy rights they are entitled to expect under GDPR?	365	270 (74)	95 (26)

[a]MH: mental health.

[b]GDPR: General Data Protection Regulation.

What Data Is Typically Collected by Mental Health Apps?

Table 9 depicts the types of data that mental health apps collect. Non–mental health apps have been included as a comparator. The table is sorted in descending order of digital mental health app frequency. The most common data type is usage data (357/436, 81.9%), followed by email (322/436, 73.9%). Percentage delta indicates the difference between mental health apps and non–mental health apps. The Fisher exact test (with Bonferroni-corrected α) was used to check for statistical significance between mental health apps and non–mental health apps. Many mental health apps collected data that may not be necessary to provide a mental health service. These were cookies or web beacons (233/436, 53.4%), location data (212/436, 48.6%), and full address or postcode (125/436, 28.7%). However, non–mental health apps also collected these data with a similar proportion; see Table 9 for a list of what types of data were collected. Figure 6 depicts occurrences of data types that were collected by the same apps. Multimedia Appendix 4 depicts the top 10 rules resulting from an association rule mining analysis, showing that email, IP address, name, and usage data are often co-collected by the same apps.

Figure 6 depicts occurrences of data types that were collected together by digital mental health apps. The most common occurrences appear in the top left.

XSL•FO
RenderX

Table . Mental health apps data collection. Bonferroni corrected α=0.002.

Data type	Frequency of MH[a] apps (N=436), n (%)	Frequency of all other apps (N=1148), n (%)	Delta (difference), %	*P* value (Fisher exact test)
Usage data	357 (81.9)	860 (74.9)	7	.003
Email	322 (73.9)	792 (69)	4.9	.07
IP address	295 (67.7)	680 (59.2)	8.5	.002
Name (full name, nickname, or first name only)	265 (60.8)	727 (63.3)	2.5	.35
Cookies/web beacons, etc (used for tracking an individual's online browsing behaviors/movements)	233 (53.4)	576 (50.2)	3.2	.26
Other	228 (52.3)	573 (49.9)	2.4	.40
Location	212 (48.6)	487 (42.4)	6.2	.03
Other unique device identifiers	191 (43.8)	432 (37.6)	6.2	.03
General wellness data	183 (42)	419 (36.5)	5.5	.049
Age/DOB[b]	176 (40.4)	515 (44.9)	4.5	.11
Number (mobile number/device number/home phone number)	155 (35.6)	446 (38.9)	3.3	.25
Gender (self-declared or observed)	149 (34.2)	430 (37.5)	3.3	.24
Full address/postcode	125 (28.7)	355 (30.9)	2.2	.39
Physical and/or mental health data	122 (28)	413 (36)	8	.003
Card/payment/financial information	102 (23.4)	229 (19.9)	3.5	.15
Username	98 (22.5)	216 (18.8)	3.7	.11
Physical description	91 (20.9)	279 (24.3)	3.4	.16
Employment/career history	44 (10.1)	105 (9.15)	0.95	.56
Race/ethnic origin	31 (7.11)	68 (5.92)	1.19	.42
Lifestyle (marital status/family/social circumstance)	29 (6.65)	75 (6.53)	0.12	.91
No information on data collection provided	28 (6.42)	102 (8.89)	2.47	.12
Device IMEI[c] number	27 (6.19)	38 (3.31)	2.88	.02
Education (qualifications/professional training/awards)	19 (4.36)	57 (4.97)	0.61	.69
Sexual orientation or sex life	16 (3.67)	46 (4.01)	0.34	.89
General identifier eg, NHS[d] No	15 (3.44)	54 (4.70)	1.26	.34
Beliefs/opinions (political, religious, or philosophical)	15 (3.44)	15 (1.31)	2.13	.01
None (data are stored on the device)	9 (2.06)	24 (2.09)	0.03	>.99
Offenses committed/alleged (criminal proceedings/outcomes/sentence)	2 (0.459)	1 (0.087)	0.372	.19

XSL•FO
RenderX

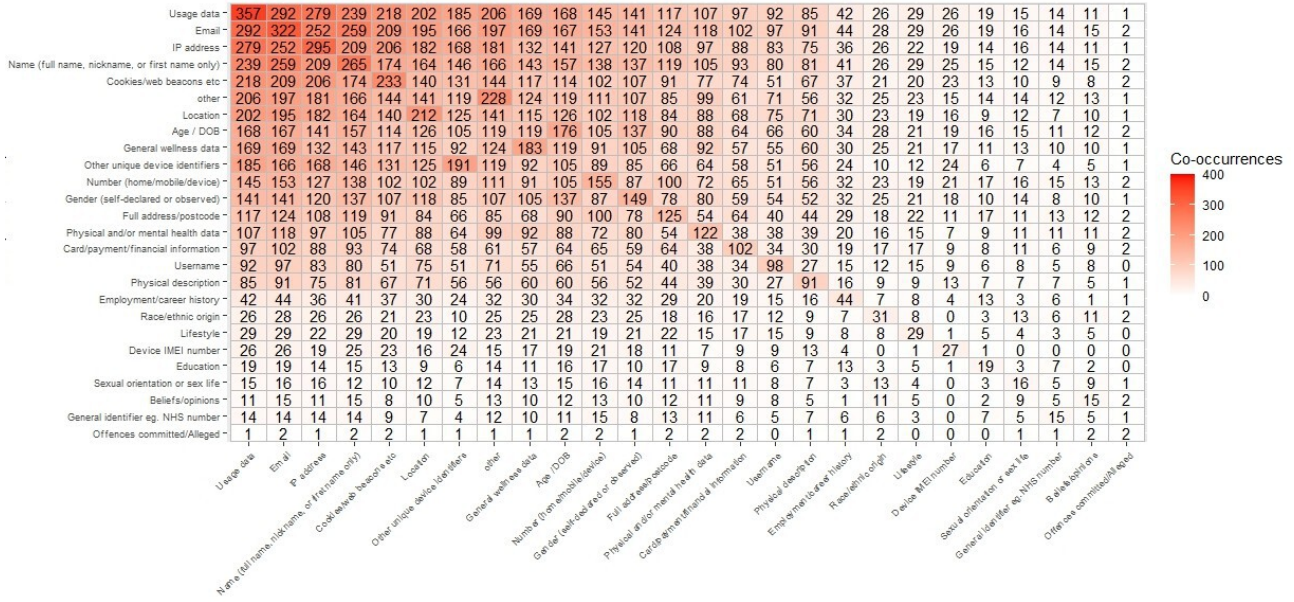
[a]MH: mental health.

[b]DOB: date of birth.

[c]IMEI: International Mobile Equipment Identity.

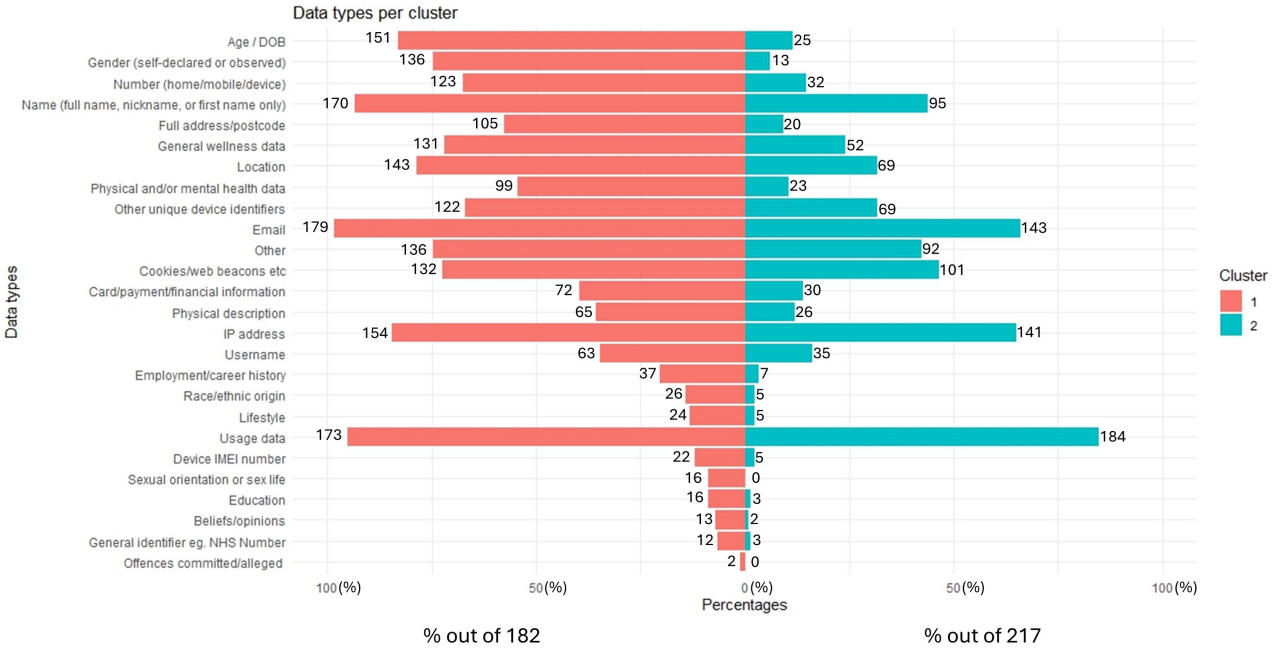
[d]NHS: National Health Service.

Figure 6. Heatmap of occurrences of collected data. Apps with "No information on data collection provided" and "None (data is stored on the device)." A total of 399 mental health apps are represented in the figure. DOB: date of birth; IMEI: International Mobile Equipment Identity.



The results show that mental health apps can be split into 2 clusters. This is visualized in Figure 7, where mental health apps in cluster 1 collected more data than cluster 2. Moreover, on average (median), mental health apps in cluster 1 present a higher ORCHA score than apps in cluster 2. This indicates that mental health apps that collected more data, on average (median), outperformed those that collected less. Previous work analyzed the top 27 ranked Google Play Store mental health apps regarding their DP [26]. It found that many of the apps had DP-related issues such as unnecessary permissions, insecure cryptography implementations, leaks of personal data, credentials in logs, web requests, as well as risk of user profiling. Further analysis is needed to confirm whether it is data collection that contributed to the apps achieving higher ORCHA scores in this study, as indicated by cluster analysis.

Figure 7. Bar plots showing frequency of data types per cluster, ordered in descending order of percentage delta (% difference) between clusters. DOB: date of birth; IMEI: International Mobile Equipment Identity; NHS: National Health Service.



XSL•FO
RenderX

K-modes cluster analysis was conducted on 399 of 436 mental health apps (apps with data types "No information on data collection provided" and "None" were excluded). Figure 7 depicts bar plots for each of the data types per cluster. The x-axis represents percentages (corresponding to the size of each cluster) of frequency for each of the data types per cluster. Multimedia Appendix 5 depicts the results of K-modes clustering conducted on the 399 mental health apps.

Multimedia Appendices 4 and 5 depict the results of association rule mining and K-modes clustering conducted on data types, respectively.

Discussion

Principal Findings

The objective of this study was to examine the characteristics of mental health apps. A total of 436 digital mental health apps were assessed in this study. This study found that most mental health apps do not meet the necessary threshold for quality of ORCHA score of 65. Many apps do not follow precautions such as involvement of professionals in the app development process, and organizations behind the apps often do not have the necessary credentials. This study analyzed the distribution of OBR scores across domains (PCA, UX, and DP) and across NICE ESF tiers. The study also analyzed the relationship between quality and cost and the importance of app features in relation to their quality, as indicated by the ORCHA score. Moreover, this study states the types of evidence that underpin mental health apps and the characteristics of app development, use, and DP.

The findings of this study demonstrate that many publicly available mental health apps may not take precautions (such as involvement of appropriate health professionals, literature references, or conducting tests) to ensure that their content is valid and research based. Moreover, as shown by previous research [6-8,18,19], there is a place for improvement regarding the quality of mental health apps. The inclusion of an appropriate health professional in the development of a mental health app could potentially improve the quality of an app, as they would be able to advise on the content of an app and any necessary features.

A study analyzing the mental health app marketplace found that, for mental health apps, user ratings and downloads are not linked with app features, function, or evidence [27]. Similarly, a study that analyzed digital health apps across 26 categories found that user ratings and downloads are not linked to quality [28]. Since user ratings and downloads are not reliable indicators, this study of mental health app characteristics and quality is necessary to further their improvement. Moreover, this indicates that users may be unable to judge whether a mental health app is of appropriate quality and may require expert recommendations. Previous work has been done trying to help in the selection of mental health apps by proposing a digital health app database [29] of apps that were assessed using the American Psychiatric Association app evaluation model [30]. Similarly, the ORCHA database could be used [31], allowing for better selection of digital health apps, which may improve their adoption. A study focused on opportunities and tensions surrounding the use of technology-enabled mental health services found that clinicians saw potential in using technology-enabled mental health care services [32]. In the context of mental health apps, this study could be helpful in making technology-enabled mental health care services improve in quality.

A study on implementation strategies for mental health apps in health care settings [33] explored strategies such as "Exploration Phase Strategies." This strategy considers the use of evidence-based services based on the needs of those in a setting. This study on the characteristics of mental health apps may influence future strategies, as it provides insight into the quality of mental health apps, the types of evidence that are collected, and the data types that are collected. This study may be useful in determining what is needed for a mental health app to be considered useful and reliable.

Limitations

The OBR evolved from earlier versions during the height of the COVID-19 pandemic. Originally, it was created as a more stringent version so that ORCHA could recommend the most compliant digital health apps to members of the UK population with confidence. ORCHA tested a new version of the OBR on a selection of highly compliant digital health apps (as determined by prior versions of the OBR). This set of 30 digital health apps served as the pilot group, with the subsequent 2097 (original dataset used in this study, Figure 1) digital health apps being assessed with ORCHA's typical assessment approach of categorizing digital health apps into categories ordered by number of downloads and assessing the most downloaded digital health app in each category, followed by the second, and so forth.

This study is limited to the characteristics that were included in the OBR assessment, meaning that features not included in the OBR may have a strong influence on quality but were not considered in this study. For example, target users such as gender or mental health symptoms were not included. Future studies could focus on additional characteristics. This study was exploratory; a more in-depth study in terms of target users and features could be conducted, for example, by analyzing feature importance for different categories of mental health apps.

This study was conducted using 436 mental health apps; a study with a larger sample size could lead to better results. Using different assessment tools (other than the OBR) could provide different results, as different characteristics and questions may be included, and a study comparing these findings with findings from other tools could lead to more insights.

Conclusion

The objective of this study was to examine the characteristics of mental health apps regarding their quality, target users, features, underpinning evidence, and DP. To achieve this objective, 7 exploratory research questions were proposed and answered. In this study, common features and data types were examined regarding their link to quality, as indicated by the ORCHA score. The key findings of the study indicate that the quality of mental health apps could be improved, as the median

XSL•FO
RenderX

ORCHA score for mental health apps is 60, which is below the ORCHA threshold score of 65. The findings of this study demonstrate that many publicly available mental health apps did not take precautions (such as involving appropriate health professionals, literature references, or conducting tests) to ensure that their content is valid and research based. Greater effort on behalf of mental health app developers is needed to ensure that the public is provided with high-quality apps.

Acknowledgments

No artificial intelligence has been used to write this paper.

Funding

The authors declared that no financial support was received for this work.

Data Availability

The authors had special access to the Organisation for the Review of Care and Health Apps (ORCHA) dataset, which is not publicly available. However, the dataset may be shared upon request by ORCHA.

Authors' Contributions

Conceptualization: MMZ, RB, MM, JMC, LB, TA
Data curation: MMZ
Formal analysis: MMZ
Investigation: MMZ, RB, MM, JMC, LB, TA
Methodology: MMZ, RB, MM, JMC, LB, TA
Resources: TA
Supervision: RB, MM, JMC, LB
Validation: RB, MM, JMC, LB
Visualization: MMZ
Writing – original draft: MMZ
Writing – review & editing: MMZ, RB

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mental health apps' categories and frequency.
[DOCX File, 22 KB - humanfactors_v13i1e67944_app1.docx]

Multimedia Appendix 2
Features' frequency separated by tier.
[DOCX File, 209 KB - humanfactors_v13i1e67944_app2.docx]

Multimedia Appendix 3
Box plots of the three most important features for each tier against Organisation for the Review of Care and Health Apps (ORCHA) scores.
[DOCX File, 75 KB - humanfactors_v13i1e67944_app3.docx]

Multimedia Appendix 4
Association rule mining analysis for mental health apps' data types.
[DOCX File, 17 KB - humanfactors_v13i1e67944_app4.docx]

Multimedia Appendix 5
K-modes cluster analysis for mental health apps' data types.
[DOCX File, 20 KB - humanfactors_v13i1e67944_app5.docx]

Checklist 1
PRISMA checklist.

XSL•FO
RenderX

[DOCX File, 275 KB - humanfactors_v13i1e67944_app6.docx]

References

1. World Mental Health Report: transforming mental health for all. : World Health Organization; 2022 URL: https://iris.who.int/server/api/core/bitstreams/40e5a13a-fe50-4efa-b56d-6e8cf00d5bfa/content [accessed 2026-04-11]
2. Nochaiwong S, Ruengorn C, Thavorn K, et al. Global prevalence of mental health issues among the general population during the coronavirus disease-2019 pandemic: a systematic review and meta-analysis. Sci Rep 2021 May 13;11(1):10173. [doi: 10.1038/s41598-021-89700-8] [Medline: 33986414]
3. Xiong J, Lipsitz O, Nasri F, et al. Impact of COVID-19 pandemic on mental health in the general population: a systematic review. J Affect Disord 2020 Dec 1;277:55-64. [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]
4. Moitra M, Owens S, Hailemariam M, et al. Global mental health: where we are and where we are going. Curr Psychiatry Rep 2023 Jul;25(7):301-311. [doi: 10.1007/s11920-023-01426-8] [Medline: 37256471]
5. Brouwers EPM. Social stigma is an underestimated contributing factor to unemployment in people with mental illness or mental health issues: position paper and future directions. BMC Psychol 2020 Apr 21;8(1):36. [doi: 10.1186/s40359-020-00399-0] [Medline: 32317023]
6. Melcher J, Torous J. Smartphone apps for college mental health: a concern for privacy and quality of current offerings. Psychiatr Serv 2020 Nov 1;71(11):1114-1119. [doi: 10.1176/appi.ps.202000098] [Medline: 32664822]
7. Larsen ME, Nicholas J, Christensen H. A systematic assessment of smartphone tools for suicide prevention. PLoS ONE 2016;11(4):e0152285. [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]
8. Sander LB, Schorndanner J, Terhorst Y, et al. "Help for trauma from the app stores?" A systematic review and standardised rating of apps for post-traumatic stress disorder (PTSD). Eur J Psychotraumatol 2020;11(1):1701788. [doi: 10.1080/20008198.2019.1701788] [Medline: 32002136]
9. Beyari H. The relationship between social media and the increase in mental health problems. Int J Environ Res Public Health 2023 Jan 29;20(3):2383. [doi: 10.3390/ijerph20032383] [Medline: 36767749]
10. Braghieri L, Levy R, Makarin A. Social media and mental health. Am Econ Rev 2022 Nov 1;112(11):3660-3693. [doi: 10.1257/aer.20211218]
11. Torous J, Roberts LW. Needed innovation in digital health and smartphone applications for mental health: transparency and trust. JAMA Psychiatry 2017 May 1;74(5):437-438. [doi: 10.1001/jamapsychiatry.2017.0262] [Medline: 28384700]
12. Schueller SM, Neary M, O'Loughlin K, Adkins EC. Discovery of and interest in health apps among those with mental health needs: survey and focus group study. J Med Internet Res 2018 Jun 11;20(6):e10141. [doi: 10.2196/10141] [Medline: 29891468]
13. ORCHA. URL: https://orchahealth.com/ [accessed 2022-06-02]
14. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/ [accessed 2022-05-03]
15. Evidence standards framework for digital health technologies. National Institute of Health and Care Excellence. 2018 Dec 10. URL: https://www.nice.org.uk/corporate/ecd7 [accessed 2021-11-06]
16. Bond RR, Mulvenna MD, Potts C, O'Neill S, Ennis E, Torous J. Digital transformation of mental health services. Npj Ment Health Res 2023 Aug 22;2(1):13. [doi: 10.1038/s44184-023-00033-y] [Medline: 38609479]
17. Moshe I, Terhorst Y, Philippi P, et al. Digital interventions for the treatment of depression: a meta-analytic review. Psychol Bull 2021 Aug;147(8):749-786. [doi: 10.1037/bul0000334] [Medline: 34898233]
18. Lecomte T, Potvin S, Corbière M, et al. Mobile apps for mental health issues: meta-review of meta-analyses. JMIR Mhealth Uhealth 2020 May 29;8(5):e17458. [doi: 10.2196/17458] [Medline: 32348289]
19. Kaveladze BT, Wasil AR, Bunyi JB, Ramirez V, Schueller SM. User experience, engagement, and popularity in mental health apps: secondary analysis of app analytics and expert app reviews. JMIR Hum Factors 2022 Jan 31;9(1):e30766. [doi: 10.2196/30766] [Medline: 35099398]
20. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567. [doi: 10.2196/14567] [Medline: 31573916]
21. Hyzy M, Bond R, Mulvenna M, et al. Quality of digital health interventions across different health care domains: secondary data analysis study. JMIR Mhealth Uhealth 2023 Nov 23;11:e47043. [doi: 10.2196/47043] [Medline: 37995121]
22. Booth F, Potts C, Bond R, et al. A mental health and well-being chatbot: user event log analysis. JMIR Mhealth Uhealth 2023 Jul 6;11:e43052. [doi: 10.2196/43052] [Medline: 37410539]
23. Dosovitsky G, Pineda BS, Jacobson NC, Chang C, Escoredo M, Bunge EL. Artificial intelligence chatbot for depression: descriptive study of usage. JMIR Form Res 2020 Nov 13;4(11):e17065. [doi: 10.2196/17065] [Medline: 33185563]
24. Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust: an R package for determining the relevant number of clusters in a data set. J Stat Softw 2014;61:1-36. [doi: 10.18637/jss.v061.i06]
25. Privacy policy. ORCHA. URL: https://www.orchahealth.com/privacy-policy [accessed 2026-04-27]
26. Iwaya LH, Babar MA, Rashid A, Wijayarathna C. On the privacy of mental health apps: an empirical investigation and its implications for app development. Empir Softw Eng 2023;28(1):2. [doi: 10.1007/s10664-022-10236-0] [Medline: 36407814]

XSL•FO
RenderX

27. Lagan S, D'Mello R, Vaidyam A, Bilden R, Torous J. Assessing mental health apps marketplaces with objective metrics from 29,190 data points from 278 apps. Acta Psychiatr Scand 2021 Aug;144(2):201-210. [doi: 10.1111/acps.13306] [Medline: 33835483]
28. Hyzy M, Bond R, Mulvenna M, et al. Don't judge a book or health app by its cover: user ratings and downloads are not linked to quality. PLoS ONE 2024;19(3):e0298977. [doi: 10.1371/journal.pone.0298977] [Medline: 38437233]
29. Lagan S, Aquino P, Emerson MR, Fortuna K, Walker R, Torous J. Actionable health app evaluation: translating expert frameworks into objective metrics. NPJ Digit Med 2020;3:100. [doi: 10.1038/s41746-020-00312-4] [Medline: 32821855]
30. The app evaluation model. American Psychiatric Association (APA). URL: https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/the-app-evaluation-model [accessed 2024-08-26]
31. Health app library. ORCHA. URL: https://us.orchahealth.com/digital-health-products/health-app-library/ [accessed 2023-01-19]
32. Lattie EG, Nicholas J, Knapp AA, Skerl JJ, Kaiser SM, Mohr DC. Opportunities for and tensions surrounding the use of technology-enabled mental health services in community mental health care. Adm Policy Ment Health 2020 Jan;47(1):138-149. [doi: 10.1007/s10488-019-00979-2] [Medline: 31535235]
33. Graham AK, Lattie EG, Powell BJ, et al. Implementation strategies for digital mental health interventions in health care settings. Am Psychol 2020 Nov;75(8):1080-1092. [doi: 10.1037/amp0000686] [Medline: 33252946]

Abbreviations

DP: data privacy
ESF: Evidence Standard Framework
NICE: National Institute for Health and Care Excellence
OBR: ORCHA Baseline Review
ORCHA: Organisation for the Review of Care and Health Apps
PCA: professional and clinical assurance
UX: user experience

Edited by A Kushniruk; submitted 25.Oct.2024; peer-reviewed by A Mushquash, N Shen; revised version received 31.Jan.2026; accepted 19.Mar.2026; published 11.May.2026.

Please cite as:
Zych MM, Bond R, Mulvenna M, Martinez Carracedo J, Bai L, Andrews T
The Quality and Characteristics of Digital Mental Health Apps: Mixed Methods Study
JMIR Hum Factors 2026;13:e67944
URL: https://humanfactors.jmir.org/2026/1/e67944
doi:10.2196/67944

© Maciej Marek Zych, Raymond Bond, Maurice Mulvenna, Jorge Martinez Carracedo, Lu Bai, Tim Andrews. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 11.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Participant Engagement With a Digital Behavioral Health App for Chronic Pain: Descriptive Secondary Analysis of a Feasibility Randomized Controlled Trial

Susan M Zbikowski[1,2], PhD; Jo Masterson[1], RN; Yohali Burrola-Mendez[3], PhD; Chialing Hsu[3], MS; Kris Pui Kwan Ma[3], PhD; Ying Zhang[3], MD; Deanna Waters[1], LMHC; Kari A Stephens[3], PhD

[1]2Morrow, Inc, 12020 113th Ave NE, Ste 295, Kirkland, WA, United States

[2]inZights Consulting, LLC, Seattle, WA, United States

[3]Department of Family Medicine, University of Washington, Seattle, WA, United States

Corresponding Author:
Susan M Zbikowski, PhD
2Morrow, Inc, 12020 113th Ave NE, Ste 295, Kirkland, WA, United States

Abstract

Background: Chronic pain is a widespread condition that impairs quality of life and is often managed primarily with medications. National guidelines now recommend nonpharmacologic, mind-and-body, and behavioral approaches as first-line or complementary treatments. However, access to these evidence-based options remains limited. Digital health technologies offer a scalable way to deliver integrative, self-care interventions that empower patients to live well with pain.

Objective: This study examined engagement with and perceived usefulness of a patient- and health care professional–informed mobile app designed to deliver behavioral and educational content to support pain self-management.

Methods: Adult primary care patients with chronic pain were enrolled in a 12-week feasibility trial. The app included lessons addressing the physical, emotional, and social aspects of pain; tracking and personalized insights; self-screenings; and optional in-app coaching. Participants completed baseline and 3-month surveys assessing usability and satisfaction. Engagement was evaluated through app analytics and milestone completion.

Results: Of 49 patients assigned to the app, 40 (81.6%) activated it. Participants used the app for an average of 27.3 (SD 25.2) unique days and completed an average of 25.5 (SD 22.5) core lessons. Engagement highlights included 42.5% (17/40) completion of the valued living module, 25.0% (10/40) completion of all lessons, and 50.0% (20/40) use of daily check-ins. Usability ratings were high, with 86.7% (26/30) reporting that the app helped them better understand or manage their pain and 90.0% (27/30) recommending it to others.

Conclusions: Adults with chronic pain engaged with the program and reported high satisfaction with this evidence-informed digital mind-and-body intervention. Findings from this feasibility study suggest the potential for digital tools to support access to nonpharmacologic, integrative pain self-care and complement traditional clinical approaches.

Trial Registration: ClinicalTrials.gov NCT06000007; https://clinicaltrials.gov/study/NCT06000007

(JMIR Hum Factors 2026;13:e88122) doi:10.2196/88122

KEYWORDS

chronic pain; self-management; digital health; behavioral health; acceptance and commitment therapy

Introduction

Chronic pain is one of the most common reasons why adults seek medical care [1]. It affects over 50 million adults in the United States—more than 20% of the population—with approximately 8.5% experiencing pain that limits daily functioning, including participation in work and social activities [2]. The burden is not equally distributed; rates are higher among individuals with lower income and educational levels and those with public health insurance [3]. Chronic pain is associated with long-term opioid use, emotional distress, and disability [4,5] and is estimated to cost the US economy between US $560 and US $635 billion annually in direct health care expenditures and lost productivity [6].

In response to the opioid epidemic, national guidelines recommend behavioral interventions as first-line or complementary treatments for chronic pain [7,8]. Evidence supports cognitive behavioral therapy, acceptance and commitment therapy, and mindfulness-based interventions [9-11], yet these treatments are underused in routine care. Primary care providers, often the first point of contact for patients [12], face barriers to implementing these strategies,

XSL•FO
RenderX

including limited training in behavioral approaches for pain; time constraints; and system-level barriers such as insufficient insurance coverage, access challenges, and limited availability of trained professionals [12-14]. Because most behavioral pain treatments are delivered in specialty settings, scalability and accessibility are limited [13-15]. The National Academy of Medicine has called for the development of innovative strategies to reduce these barriers and support more equitable and effective pain care [12].

Digital health technologies, including mobile apps, can extend the reach of evidence-based behavioral interventions by delivering support asynchronously outside traditional clinical settings and being integrated into broader care models. They align with needs to diversify service delivery and equip medical professionals, including those in primary care, with accessible tools for pain management [13,16-18]. However, reviews show that many consumer apps lack credible content or health care professional input, have limited functionality, and fail to engage users [19,20]. This highlights the need for engaging, evidence-based, patient- and health care professional–informed apps that address the full biopsychosocial experience of pain and can be used in clinical practice.

We conducted a feasibility study of a self-directed mobile app designed to deliver evidence-based educational content and behavioral strategies to adults with chronic pain. This study evaluated participant engagement, including patterns of use, frequency of use, and feature use. Insights into user interaction with digital behavioral tools are essential for refining intervention design, informing clinical integration, and guiding future research and scale-up.

Methods

Study Design

We analyzed engagement data collected during a feasibility randomized control trial of a mobile behavioral health app designed to help patients treated in primary care for chronic pain management. The feasibility trial was a 2-arm randomized controlled trial designed to assess the feasibility of trial implementation (recruitment, randomization, treatment, and retention) and explore preliminary impacts on pain, pain interference, and mental health. This paper presents a descriptive analysis of user experience and app use data among participants in the intervention arm, who had access to the mobile behavioral health app during the study period. Participants were randomized using a stratified approach based on health care system and age to ensure balanced allocation across health systems and approximate the national age distribution of adults living with chronic pain.

Ethical Considerations

This study was conducted as part of a National Institutes of Health Small Business Innovation Research–funded collaboration, with 2Morrow Inc serving as the grant recipient and developer of the intervention and the University of Washington serving as the academic research partner.

All study procedures, including recruitment, consent, data collection, and follow-up, were reviewed and approved by the University of Washington Institutional Review Board (STUDY00016340) and were conducted in accordance with the approved study protocols and applicable federal regulations. All patients provided informed consent prior to participation. Patient privacy and confidentiality were protected through multiple safeguards. Data were stored on secure, access-controlled systems, and all analytic datasets were deidentified prior to analysis. No directly identifying information was included in analytic files. App use data were captured automatically through the intervention platform and linked to study data using unique participant identifiers. Data handling procedures complied with applicable privacy and confidentiality requirements. Patients assigned to the Salty program were provided with 2Morrow's privacy policy at the time of program activation, which outlines how user data are collected, used, and protected.

Patients received a US $50 electronic gift card for each completed survey (up to US $200 total).

Recruitment and Eligibility

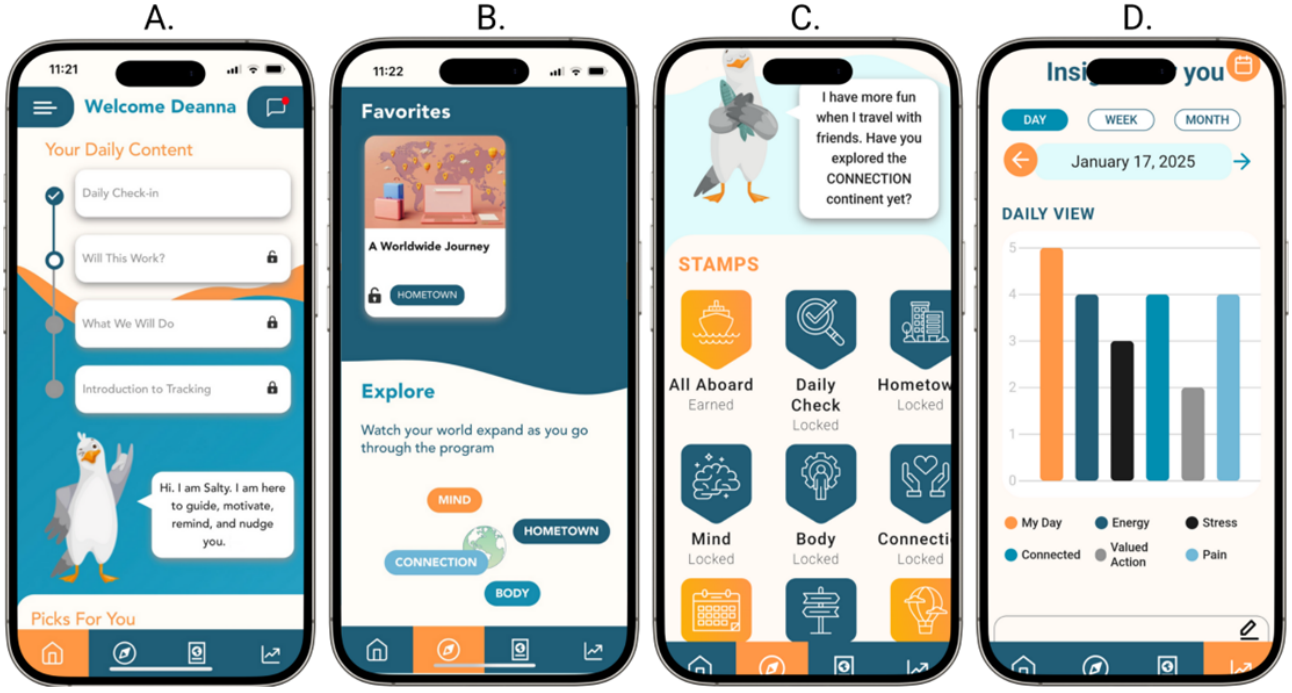
Primary care patients with chronic pain were identified through electronic health records at 4 health systems (60 clinics) in Washington and Idaho, sent an email invitation to participate in the study, and directed to complete a screening survey to determine eligibility. Eligible participants had at least one documented chronic pain condition identified through electronic health records. Types of pain represented in the study population included back pain, neck pain, noninflammatory joint pain, abdominal or bowel pain, fibromyalgia, and headache disorders. Patients were selected to participate if they were actively receiving care for their chronic pain (≥2 primary care visits in the previous year), currently endorsed struggling with chronic pain for 3 months or longer, endorsed having pain interference in their lives (≥55 on the Patient-Reported Outcomes Measurement Information System Pain Interference scale [21,22]), could use an app in English, and had a smartphone compatible with the behavioral health app.

Patients randomized to the intervention group received access to the Salty for Chronic Pain app, a novel digital behavioral health intervention developed by 2Morrow Inc (Figure 1). To be included in this study and analysis, patients had to successfully activate the app, which included downloading it and completing their in-app personal profile.

Patients were invited to participate via email and directed to an online interest survey. The survey provided 3 options: proceed to eligibility screening, request a call from a study coordinator for additional information, or opt out of the study. Patients who completed the eligibility screening were informed of their eligibility status at the end of the survey. Those who were eligible were directed to an electronic consent form, which allowed them to confirm participation, request further discussion with a study coordinator, or decline participation. Patients who confirmed participation provided informed consent electronically prior to enrollment.

XSL•FO
RenderX

Figure 1. Screenshots of the pain app: (A) daily task recommendations, (B) the section where patients can explore other core lesson categories, (C) the passport stamp page for viewing completed milestones, and (D) the tool that provides insights from the daily check-in data.



Program Design and Features

Salty for Chronic Pain (pain app) is a 12-week self-directed program designed to educate and support patients managing the emotional, social, and behavioral aspects of living with chronic pain with the goal of reducing pain interference and improving function and quality of life. The program was created with input from patients, health care professionals, and pain behavioral health specialists during phase 1 of the Small Business Innovation Research trial; health care professionals indicated that they wanted additional options for their patients with pain that were evidence based but did not require much of their time [23].

The program used the metaphor of guiding patients on a personal journey to explore the physical, emotional, and social aspects of pain to learn how to live life despite having pain. The program included a total of 58 core lessons divided into 4 sections and additional bonus content. The first core section, known as "Hometown," included 22 lessons related to the concept of valued living and living with pain based on principles from acceptance and commitment therapy. The other sections covered the physical (16 lessons), emotional (13 lessons), and social (7 lessons) aspects of pain.

The program included tools for gaining personal insights into pain and daily functioning. The "Daily Check-in" allowed patients to rate their overall day ("awesome" to "awful"), as well as energy, activity, stress, social connection, valued activities, and pain (all on 5-point scales), with results displayed graphically across daily, weekly, or monthly views. A journal feature was provided for recording additional reflections, and patients could optionally connect a Fitbit (provided at no cost) to track activity, sleep, and heart rate. Self-screenings were available in the "Self-Discovery" section, including the Depression, Anxiety, and Stress Scale-21 (DASS-21 [24]); Valued Living Questionnaire (VLQ; [25]); Drug Use Disorders Identification Test (DUDIT [26]); and Alcohol Use Disorders Identification Test (AUDIT [27]). Although the program was intended to be self-guided, participants could message with a coach on the app about lessons, personal discoveries, and progress. These features and others on the app are described in Textbox 1.

XSL•FO
RenderX

Textbox 1. Summary of core features and functionality on the app.

Profile

- Description: pain-related questions that patients complete to set up their account
- Purpose: benchmark data and personalization

Core lessons

- Description: 58 short lessons related to 4 domains (valued living—acceptance and commitment therapy concept, physical aspects of pain, emotional aspects of pain, and social aspects of pain)
- Purpose: pain information, education, and coping skills

Bonus lessons

- Description: additional content that patients could access
- Purpose: additional education, tips, and patient stories

Daily check-in

- Description: a brief daily questionnaire for patients to rate various aspects of living with pain
- Purpose: a personal pain tracking tool

Journal

- Description: patients could record their pain experience in their own words
- Purpose: a flexible and personal pain tracking and insight tool

Personalized messaging

- Description: brief in-app messages
- Purpose: provided to direct, motivate, or reinforce concepts

Sensor integration

- Description: participants could link a Fitbit device to the app to track additional insights
- Purpose: objective tracking of sleep, activity, and heart rate

Self-screening questionnaires

- Description: 4 standard questionnaires for patients to explore alcohol and drug use, mental health, and valued living
- Purpose: to explore and better understand their substance use, mental health, and what is most important to them

Insight dashboard

- Description: a visualization tool to explore pain insights
- Purpose: ability to view the interconnection among self-reported pain, energy, activity, stress, social connection, and other metrics

Milestone achievement

- Description: participants could receive up to 12 stamps on their passport page
- Purpose: summarize and reward progress on the app

Monthly check-in

- Description: a brief monthly questionnaire
- Purpose: in-app survey to measure pain, interference, mental health, quality of life, and satisfaction each month after a patient enrolls in and uses the program

Message a coach

- Description: asynchronous chat tool with a coach
- Purpose: meant to provide additional support and encouragement for patients who wanted to connect with a coach; patients could share insights and achievements as well

XSL•FO
RenderX

Navigation and pacing features

- Patients were instructed to complete a certain number of lessons along with a daily check-in each day that they visited the app
- The app required completion of all Hometown lessons before moving on to another section
- Patients needed to complete each core section before selecting another one to complete
- Patients could complete the self-screening tools only once per month
- Save a favorite: a feature on the app to save content that participants found helpful

How the Intervention Worked

Each time patients opened the app, they were prompted to complete a "Daily Check-in" and up to 3 scheduled lessons and could view optional bonus content. They were encouraged to engage with the app 2 to 3 times per week over 12 weeks, with lesson access paced to discourage rapid completion and promote sustained use. All participants began with valued living content in the "Hometown" section before advancing to the "Social," "Physical," or "Emotional" modules. Completion of all lessons within a category was required before moving on as content was designed to build sequentially. Participants could bookmark favorite lessons and retained access to app content and features for 100 days.

App Milestone Completion

Patients earned up to 12 milestone stamps (Table 1) as a gamified feature to promote engagement and track progress. Stamps were displayed on the "Passport" tab and awarded for completing the profile, core content (values, body, mind, connection, and bonus), at least 10 "Daily Check-ins," and a monthly check-in, as well as for connecting a Fitbit, messaging a coach, completing self-screenings, and visiting all recommended sections.

Table . Passport stamps for the 12 milestones completed on the app (N=40).

Milestone	Stamp name	Participants receiving the stamp, n (%)
Completion of profile	All Aboard	40 (100.0)
Completion of daily check-in 10 times	Daily Check	20 (50.0)
User visited the "Self-Discovery," "Passport" (stamp location), "Explore" navigator tool, and "Insights" pages	Traveler	19 (47.5)
One self-screening (alcohol or substance use, depression and anxiety screening, or valued living questionnaire) completed	Discover	17 (42.5)
Hometown core content completed (22 lessons)	Hometown	17 (42.5)
Mind core content completed (13 lessons)	Mind	11 (27.5)
Body core content completed (16 lessons)	Body	13 (32.5)
Connection core content completed (7 lessons)	Connection	11 (27.5)
15 bonus lessons completed	Extras	14 (35.0)
First monthly check-in completed	Monthly Check	14 (35.0)
One engagement with a coach	Coach	13 (32.5)
Synchronizing the Fitbit	Sync Up	5 (12.5)

Data Collection

Demographics (age, gender, race and ethnicity, educational level, and employment status) and pain history (type and duration) were collected at screening or baseline. Use metrics were captured from the app data system. App metrics included app use milestones achieved (eg, earned passport stamps) and use metrics, including unique days of use, duration (days between first and last use), core lessons completed, feature use (tracking and self-screenings), and in-app actions.

User experience was measured at 3 months via the mHealth App Usability Questionnaire (MAUQ [28]) and a custom perceived usefulness survey developed by 2Morrow, Inc. Acceptability and usefulness were derived from the MAUQ Ease of Use and Satisfaction Domain (Q1-8). The custom survey asked patients to rate the helpfulness of the app overall and of specific features, value, and the likelihood that they would recommend the app to others.

Recruitment, screening, and survey administration were managed by the University of Washington. Patients had 2 weeks to complete each survey.

Analyses

Descriptive statistics (frequencies, means, ranges, and percentages) were calculated for demographics and engagement, describing the types of app content, features, and tools accessed.

XSL•FO
RenderX

Results

Study Population

Of the 49 patients randomized to the pain app intervention arm, 40 (81.6%) met the inclusion criteria for this study (ie, successfully activated the program by downloading the app and completing their profile). Participants reported living with chronic pain for 3 or more years; were aged 47.3 (SD 17.3) years on average (the most common age group was 30-44 years); and identified as women for the most part followed by men, with a smaller proportion (3/40, 7.5%) identifying as nonbinary. Nearly half (19/40, 47.5%) of the participants identified as White individuals, with the remainder (21/40, 52.5%) identifying as Asian, Pacific Islander, or Hawaiian; Black; multiracial; or other. Educational attainment ranged from high school to postgraduate education, with most participants (32/39, 82.1%) reporting at least some college education. At baseline, a substantial proportion of the participants reported a history of mental health conditions, reflecting the clinical complexity of the study population. Employment status varied, including part-time work, full-time employment, and disability status. Counts and percentages are reported in Table 2.

XSL•FO
RenderX

Table . Participant characteristics (N=40).

	Participants, n (%)
Gender	
Women	25 (62.5)
Men	12 (30.0)
Nonbinary	3 (7.5)
Race or ethnicity	
Asian American	5 (12.5)
Black or African American	5 (12.5)
Native Hawaiian or other Pacific Islander	2 (5.0)
Two or more races	6 (15.0)
White	19 (47.5)
Other	3 (7.5)
Age group (years)	
<30	8 (20.0)
30 - 44	12 (30.0)
45-64	12 (30.0)
≥65	8 (20.0)
Educational level (n=39)	
Lower than high school	3 (7.7)
High school graduate or GED[a]	4 (10.3)
Some college but did not complete degree	5 (12.8)
2- or 4-year degree	10 (25.6)
More than 4-year college degree	17 (43.6)
Employment status (n=38)	
Employed full time	15 (39.5)
Employed part time	5 (13.2)
Contract or temporary	1 (2.6)
Voluntarily unemployed	3 (7.9)
Involuntarily unemployed	1 (2.6)
Unable to work or disabled	8 (21.1)
Retired	5 (13.2)
Pain duration	
3 months to 1 year	3 (7.5)
1 year to 3 years	5 (12.5)
≥3 years	32 (80.0)

[a]GED: General Educational Development.

Outcome Measures

App Use

Passport Stamps Earned

All participants earned at least one passport stamp, with a mean of 4.9 (SD 3.8; range 1 - 12) stamps per user, indicating broad engagement with the app. The most frequently earned stamps reflected onboarding and routine use, including completing the profile (40/40, 100.0%), completing at least 10 daily check-ins (20/40, 50.0%), and exploring multiple app sections (19/40, 47.5%). A total of 42.5% (17/40) completed at least one core content module, with all of these participants completing the valued living ("Hometown") module; in this subgroup, 58.8% (10/17) completed all 4 content modules, and 82.4% (14/17) accessed bonus content. Overall, 25.0% (10/40) of the

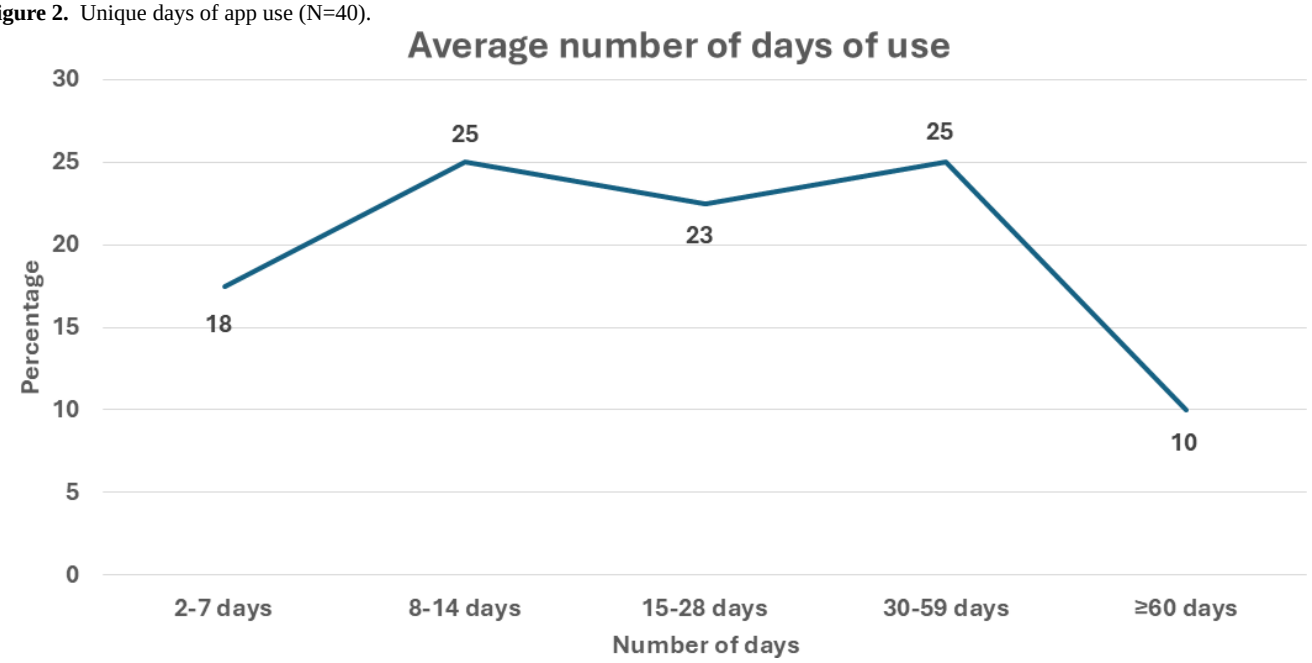
XSL•FO
RenderX

participants earned 9 or more stamps, suggesting a subset engaged deeply with multiple features, whereas others demonstrated more selective use.

Lesson Completion and Days of Use

Participants completed a mean of 25.5 (SD 22.5, range 2 - 58) core lessons. A total of 42.5% (17/40) completed the valued living module, and 25.0% (10/40) completed all 58 core lessons across the 4 content areas. Participants engaged with the app on an average of 27.3 (SD 25.2, range 2 - 118) unique days over the 12-week period; 57.5% (23/40) used the app on 15 or more days, and 75.0% (30/40) interacted with the app across a span of 30 to 100 calendar days (Figure 2).

Figure 2. Unique days of app use (N=40).



Additional Engagement Indicators

Use of tracking and optional features varied. In total, 82.5% (33/40) completed at least one milestone related to pain insight activities (daily check-ins, self-screenings, or monthly check-ins), with over 40% (17/40, 42.5%) completing at least one mental health or substance use self-screening. Completion rates for each self-screening ranged from 22% to 30%: DUDIT (9/40, 22.5%), VLQ (10/40, 25%); AUDIT (10/40, 25.0%), and DASS-21 (12/40, 30.0%). A smaller proportion engaged with coaching features (13/40, 32.5%) or synchronized a Fitbit device (5/40, 12.5%). Across all users, mean in-app activity totaled 2123 (SD 2579.8, range 48 - 10,332) actions, reflecting substantial variability in depth of interaction.

User Survey Results

Overview

The response rate to the follow-up survey at 3 months was 85.0% (34/40) for the MAUQ ease of use and satisfaction domain questions (Table 3). Due to an error in the survey build, the custom perceived usefulness survey response rate was somewhat lower (30/40, 75.0%), and missing responses were mostly due to random missingness (Table 4).

XSL•FO
RenderX

Table . Frequency distribution of participant responses to the mHealth App Usability Questionnaire ease of use and satisfaction domain items (N=34).

Item	Strongly disagree, n (%)	Disagree, n (%)	Somewhat disagree, n (%)	Neither agree nor disagree, n (%)	Somewhat agree, n (%)	Agree, n (%)	Strongly agree, n (%)	Somewhat agreed, agreed, or strongly agreed, n (%)
The app was easy to use	0 (0.0)	0 (0.0)	2 (5.9)	5 (14.7)	2 (5.9)	15 (44.1)	11 (32.4)	28 (82.4)
It was easy for me to learn to use the app	0 (0.0)	0 (0.0)	2 (5.9)	3 (8.8)	2 (5.9)	14 (41.2)	13 (38.2)	29 (85.3)
I like the interface of the app[a]	0 (0.0)	1 (3.0)	3 (9.1)	6 (18.2)	2 (6.1)	11 (33.3)	10 (30.3)	23 (69.7)
The information in the app was well organized, so I could easily find the information I needed	0 (0.0)	0 (0.0)	3 (8.8)	6 (17.6)	10 (29.4)	6 (17.6)	9 (26.5)	25 (73.5)
I feel comfortable using this app in social settings	2 (5.9)	2 (5.9)	1 (2.9)	6 (17.6)	7 (20.6)	6 (17.6)	10 (29.4)	23 (67.6)
The amount of time involved in using this app has been fitting for me	1 (2.9)	2 (5.9)	3 (8.8)	4 (11.8)	4 (11.8)	10 (29.4)	10 (29.4)	24 (70.6)
I would use this app again[a]	1 (3.0)	0 (0.0)	2 (6.1)	5 (15.2)	4 (12.1)	10 (30.3)	11 (33.3)	25 (75.8)
Overall, I am satisfied with this app	1 (2.9)	0 (0.0)	1 (2.9)	8 (23.5)	4 (11.8)	9 (26.5)	11 (32.4)	24 (70.6)

[a]Results are reported based on 33 valid responses.

XSL•FO
RenderX

Table . Frequency distribution of participant responses to the pain app perceived usefulness survey (N=30).

Item	Not at all, n (%)	Slightly, n (%)	Moderately, n (%)	Very, n (%)	Extremely, n (%)	Slightly, moderately, very, or extremely, n (%)
The Salty Program helped me with my pain, for example, to better understand or manage my pain	4 (13.3)	7 (23.3)	11 (36.7)	6 (20.0)	2 (6.7)	26 (86.7)
Was using the Salty program worth your time?	4 (13.3)	6 (20.0)	8 (26.7)	9 (30.0)	3 (10.0)	26 (86.7)
Please rate how helpful the following features in the Salty Program were to you:						
The lessons and content in the program	3 (10.0)	4 (13.3)	9 (30.0)	11 (36.7)	3 (10.0)	27 (90.0)
Having the program divided into different topics ranging from physical, emotional, social aspects of pain	4 (13.3)	3 (10.0)	3 (10.0)	14 (46.7)	6 (20.0)	26 (86.7)
Having content related to how pain may affect you physically	4 (13.3)	4 (13.3)	4 (13.3)	15 (50.0)	3 (10.0)	26 (86.7)
Having content related to how pain may affect you emotionally	4 (13.3)	3 (10.0)	3 (10.0)	13 (43.3)	7 (23.3)	26 (86.7)
Having content related to how pain may affect you socially	4 (13.3)	4 (13.3)	4 (13.3)	13 (43.3)	5 (16.7)	26 (86.7)
How much you got to do each day	5 (16.7)	6 (20.0)	8 (26.7)	10 (33.3)	1 (3.3)	25 (83.3)
Having to complete an entire section of lessons before moving to other topics or themes	7 (23.3)	6 (20.0)	6 (20.0)	9 (30.0)	2 (6.7)	23 (76.7)
Daily check-ins to track how things are going for you	3 (10.0)	6 (20.0)	5 (16.7)	9 (30.0)	7 (23.3)	27 (90.0)
Viewing insights from daily tracking	5 (16.7)	4 (13.3)	7 (23.3)	10 (33.3)	4 (13.3)	25 (83.3)
The option to connect with a coach	8 (26.7)	4 (13.3)	9 (30.0)	6 (20.0)	3 (10.0)	22 (73.3)
Earning passport stamps	9 (30.0)	8 (26.7)	7 (23.3)	3 (10.0)	3 (10.0)	21 (70.0)
Opportunity to receive a Fitbit	9 (30.0)	3 (10.0)	9 (30.0)	4 (13.3)	5 (16.7)	21 (70.0)
Seeing Fitbit data in the app	10 (33.3)	3 (10.0)	9 (30.0)	2 (6.7)	6 (20.0)	20 (66.7)
Length of the program (100 d)	3 (10.0)	4 (13.3)	10 (33.3)	12 (40.0)	1 (3.3)	27 (90.0)

XSL•FO
RenderX

Item	Not at all, n (%)	Slightly, n (%)	Moderately, n (%)	Very, n (%)	Extremely, n (%)	Slightly, moderately, very, or extremely, n (%)
Self-discovery surveys in the app to learn more about your mood, drinking, etc	5 (16.7)	6 (20.0)	6 (20.0)	5 (16.7)	8 (26.7)	25 (83.3)
Videos that you watched	4 (13.3)	9 (30.0)	8 (26.7)	5 (16.7)	4 (13.3)	26 (86.7)

MAUQ Ease of Use and Satisfaction Domains

Overall usability and satisfaction ratings were favorable (Table 3). Most participants reported that the app was easy to learn (29/34, 85.3%) and use (28/34, 82.4%). More than 70% (25/34, 73.5%) agreed that the app was well organized and the time commitment was appropriate, were satisfied overall, and would use it again.

Custom Perceived Usefulness Survey

Core features—including daily check-ins; structured lessons; and content addressing the physical, emotional, and social aspects of pain—were rated as helpful by 76.7% (23/30) to 90.0% (27/30) of the participants. While features such as receiving passport stamps, Fitbit integration, and coach messaging received somewhat lower ratings, more than two-thirds (20/30, 66.7%) still found them useful. Overall, 90% (27/30) of respondents reported that they would recommend the app to others. Additionally, 23.3% (7/30) rated the daily check-ins and emotional content as extremely helpful, with comparable ratings across other content areas (Table 4).

Discussion

Principal Findings

This study evaluated the Salty for Chronic Pain app, a behavioral digital health program designed to support adults living with chronic pain. Unlike most studies on digital health programs for pain that report only frequency or duration of use, our evaluation incorporated milestone completion, lesson progression, self-tracking, and perceptions of usefulness, offering a more comprehensive view of engagement and potential value as a complement to clinical care. Patients demonstrated sustained engagement for nearly a month, advanced through structured lessons, and reported high usability and satisfaction, suggesting that the program appeared feasible and was generally well received. These findings add to a growing body of research showing that digital platforms can deliver behavioral pain self-management strategies in ways that are usable, acceptable, and aligned with evidence-based approaches [10,16,23,29].

Although direct comparisons to other studies are limited due to differences in design, content, and metrics, our findings are broadly consistent with those of prior digital pain intervention research [30,31]. Activation rates were high, suggesting that patients could begin using the app with minimal support, which may help reduce burden on clinical staff. Most participants met recommended frequency targets, indicating that the app design supported regular use without intensive reminders. These results are comparable to those of Thomson et al [32], who reported that 77% of participants achieved minimum engagement during a 6-week intervention.

On average, patients completed more than 25 core lessons, and engagement extended beyond the introductory module, with a subset completing all lessons and others using bonus content. This level of engagement is noteworthy given that lessons were sequenced, requiring completion of one section before progressing, and the valued living module (“Hometown”) was mandatory before others could be accessed. In addition, no automated reminders were delivered due to a technical issue. Hence, engagement occurred without automated reminders, suggesting that participants were self-motivated. It is possible that greater use would have been achieved with automated reminders or more flexible sequencing. At the same time, these findings raise important questions about how much content exposure is necessary to achieve meaningful improvements in pain, acceptance, and functioning. Some patients may benefit from a targeted subset of lessons, whereas others may require more extensive exposure. Clinical outcomes will be reported separately. Future refinements, such as adding flexible sequencing or automated prompts, may increase completion rates and help identify the “dose” of content most strongly associated with clinical benefit or perceived value.

These findings highlight the potential of digital behavioral health tools to support access to evidence-based pain self-management strategies. Participants engaged with structured lessons and interactive features that promoted reflection, self-monitoring, and awareness of valued living—processes associated with acceptance, coping, and resilience in chronic pain populations [33,34]. The results suggest that such strategies can be delivered in digital form with high usability and satisfaction; however, engagement metrics reflect exposure to content rather than therapeutic dose or clinical effect.

Several limitations should be considered when interpreting these findings. The relatively small sample size reflects the feasibility nature of the trial and limits generalizability to all adults with chronic pain. The eligibility criteria and study context further shape interpretation: participants were required to have smartphone access, English-language proficiency, and clinically significant pain interference, and most reported long-standing pain. Accordingly, the sample may overrepresent individuals who were digitally ready or motivated to engage with a self-directed program. The engagement patterns observed in this study may not generalize to all adults with chronic pain,

XSL•FO
RenderX

particularly those with limited digital access or differing readiness to engage in behavioral interventions.

Taken together, these findings suggest that this type of digital behavioral health program may be particularly well suited for individuals who are motivated to engage in self-directed care and interested in nonpharmacologic or complementary approaches to chronic pain management. The observed engagement without automated reminders further suggests potential fit for users who prefer flexible, low-burden support delivered outside traditional clinical settings.

Engagement and satisfaction ratings may have also been influenced by study participation, incentives associated with survey completion, novelty effects, or app design constraints such as fixed lesson sequencing.

Despite these limitations, this study provides early evidence that an evidence-informed, user-centered mobile program can be feasibly delivered and meaningfully engaged with by adults receiving care for chronic pain. These findings can inform intervention refinement and support the design of larger studies to evaluate engagement patterns; predictors of sustained use; and effects on pain interference, functioning, and quality of life [17,18].

Conclusions

Patients successfully activated the program without support, most met recommended use targets, and many engaged deeply with the available content and insight tools. These findings are promising for the use of self-directed evidence-based digital behavioral health apps in clinical pain management. While patients rated the features favorably, additional research is needed to determine how much exposure is necessary to achieve clinical benefit. Future studies should evaluate the impact of these tools on pain interference and functioning, identify engagement patterns linked to improved outcomes, and test refinements such as flexible sequencing or automated reminders. Addressing these questions can help digital behavioral health programs expand access; supplement usual care; and support more equitable, scalable approaches to chronic pain management.

Acknowledgments

The authors wish to thank MG Prado, B Keiser, and T Anastas for assistance with developing measures, screening and recruitment, and study management; the Data Query Extraction Standardization Translation Coordinating Center for assistance with recruiting; and Distributed Ambulatory Research in Therapeutics Network (DARTNet) for programming and collecting baseline and follow-up measures. Some of the results reported in this paper were presented as a poster at PAINWeek 2025, Las Vegas, Nevada. ChatGPT, an artificial intelligence language model developed by OpenAI, was used to assist with editing and improving the flow of the manuscript after it was written by the authors. All artificial intelligence–generated suggestions were critically reviewed and approved by the authors prior to inclusion in the final manuscript.

Funding

The research reported in this publication was supported by the National Center for Complementary and Integrative Health and the National Institute of Neurological Disorders and Stroke of the National Institutes of Health under award R44AT011593 and leveraged resources from the University of Washington Institute of Translational Health Sciences funded by the National Center for Advancing Translational Sciences of the National Institutes of Health under award UL1TR002319. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

The data that support the findings of this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: SMZ, JM, KAS
Data curation: YB-M, KPKM, YZ, KAS
Formal analysis: SMZ, YB-M, KPKM, YZ, CH, KAS
Funding acquisition: JM, KAS
Investigation: SMZ, DW, KAS
Methodology: SMZ, JM, YB-M, KPKM, YZ, SW, KAS
Project administration: DW
Resources: DW
Writing—original draft: SMZ
Writing—review and editing: SMZ, JM, YB-M, KPKM, YZ, CH, DW, KAS

Conflicts of Interest

JM is the chief executive officer of 2Morrow Inc, which is a digital health company that creates evidence-based digital therapeutic programs for behavior change and chronic disease management. JM is also the National Institutes of Health grant principal investigator for this Small Business Innovation Research grant–based study. DW is vice president of Participant Success at

XSL•FO
RenderX

2Morrow Inc and was involved in the development and oversight of the digital behavioral health app for chronic pain evaluated in this study. 2Morrow Inc may benefit from positive findings related to this product. SMZ was a paid consultant working for 2Morrow Inc. She currently serves as a noncompensated scientific advisor to 2Morrow Inc and has been granted stock options in the company. She does not currently hold equity shares. All other authors declare no other conflicts of interest.

References

1. Chronic pain and complementary health approaches. National Center for Complementary and Integrative Health. URL: https://www.nccih.nih.gov/health/chronic-pain-and-complementary-health-approaches-usefulness-and-safety [accessed 2026-03-20]
2. Lucas JW, Sohi I. Chronic pain and high-impact chronic pain in U.S. adults, 2023. NCHS Data Brief 2024 Oct(518):CS355235. [doi: 10.15620/cdc/169630] [Medline: 39751180]
3. Dahlhamer J, Lucas J, Zelaya C, et al. Prevalence of chronic pain and high-impact chronic pain among adults - United States, 2016. MMWR Morb Mortal Wkly Rep 2018 Sep 14;67(36):1001-1006. [doi: 10.15585/mmwr.mm6736a2] [Medline: 30212442]
4. Pitcher MH, Von Korff M, Bushnell MC, Porter L. Prevalence and profile of high-impact chronic pain in the United States. J Pain 2019 Feb;20(2):146-160. [doi: 10.1016/j.jpain.2018.07.006] [Medline: 30096445]
5. Kathiresan P, Rao R, Joshi T, et al. Chronic noncancer pain and opioid addiction: diagnostic and management challenges. Indian J Palliat Care 2020;26(4):544-547. [doi: 10.4103/IJPC.IJPC_232_19] [Medline: 33623322]
6. Gaskin DJ, Richard P. The economic costs of pain in the United States. J Pain 2012 Aug;13(8):715-724. [doi: 10.1016/j.jpain.2012.03.009] [Medline: 22607834]
7. Dowell D, Haegerich TM, Chou R. CDC guideline for prescribing opioids for chronic pain--United States, 2016. JAMA 2016 Apr 19;315(15):1624-1645. [doi: 10.1001/jama.2016.1464] [Medline: 26977696]
8. Pain Management Best Practices Inter-Agency Task Force Report. : US Department of Health and Human Services; 2019 URL: https://www.hhs.gov/sites/default/files/pain-mgmt-best-practices-draft-final-report-05062019.pdf [accessed 2026-03-20]
9. Ehde DM, Dillworth TM, Turner JA. Cognitive-behavioral therapy for individuals with chronic pain: efficacy, innovations, and directions for research. Am Psychol 2014;69(2):153-166. [doi: 10.1037/a0035747] [Medline: 24547801]
10. Veehof MM, Oskam MJ, Schreurs KM, Bohlmeijer ET. Acceptance-based interventions for the treatment of chronic pain: a systematic review and meta-analysis. Pain 2011 Mar;152(3):533-542. [doi: 10.1016/j.pain.2010.11.002] [Medline: 21251756]
11. Garland EL, Manusov EG, Froeliger B, Kelly A, Williams JM, Howard MO. Mindfulness-oriented recovery enhancement for chronic pain and prescription opioid misuse: results from an early-stage randomized controlled trial. J Consult Clin Psychol 2014 Jun;82(3):448-459. [doi: 10.1037/a0035798] [Medline: 24491075]
12. Institute of Medicine (US) Committee on Advancing Pain Research, Care, and Education. Relieving Pain in America: A Blueprint for Transforming Prevention, Care, Education, and Research: National Academies Press; 2011. [doi: 10.17226/13172]
13. Keefe FJ, Abernethy AP, Campbell LC. Psychological approaches to understanding and treating disease-related pain. Annu Rev Psychol 2005;56:601-630. [doi: 10.1146/annurev.psych.56.091103.070302] [Medline: 15709948]
14. Mularski RA, White-Chu F, Overbay D, Miller L, Asch SM, Ganzini L. Measuring pain as the 5th vital sign does not improve quality of pain management. J Gen Intern Med 2006 Jun;21(6):607-612. [doi: 10.1111/j.1525-1497.2006.00415.x] [Medline: 16808744]
15. Kazdin AE, Blase SL. Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspect Psychol Sci 2011 Jan;6(1):21-37. [doi: 10.1177/1745691610393527] [Medline: 26162113]
16. Keogh E, Rosser BA, Eccleston C. e-Health and chronic pain management: current status and developments. Pain 2010 Oct;151(1):18-21. [doi: 10.1016/j.pain.2010.07.014] [Medline: 20674174]
17. Ruehlman LS, Karoly P, Enders C. A randomized controlled evaluation of an online chronic pain self management program. Pain 2012 Feb;153(2):319-330. [doi: 10.1016/j.pain.2011.10.025] [Medline: 22133450]
18. McBeth J, Prescott G, Scotland G, et al. Cognitive behavior therapy, exercise, or both for treating chronic widespread pain. Arch Intern Med 2012 Jan 9;172(1):48-57. [doi: 10.1001/archinternmed.2011.555] [Medline: 22082706]
19. Lalloo C, Jibb LA, Rivera J, Agarwal A, Stinson JN. "There's a pain app for that": review of patient-targeted smartphone applications for pain management. Clin J Pain 2015 Jun;31(6):557-563. [doi: 10.1097/AJP.0000000000000171] [Medline: 25370138]
20. Reynoldson C, Stones C, Allsop M, et al. Assessing the quality and usability of smartphone apps for pain self-management. Pain Med 2014 Jun;15(6):898-909. [doi: 10.1111/pme.12327] [Medline: 24422990]
21. Amtmann D, Cook KF, Jensen MP, et al. Development of a PROMIS item bank to measure pain interference. Pain 2010 Jul;150(1):173-182. [doi: 10.1016/j.pain.2010.04.025] [Medline: 20554116]
22. Pain interference scoring manual. : HealthMeasures; 2022 URL: https://www.healthmeasures.net/images/PROMIS/manuals/Scoring_Manual_Only/PROMIS_Pain_Interference_Scoring_Manual_03June2022.pdf [accessed 2026-03-21]

XSL•FO
RenderX

23. Ma KP, Stephens KA, Geyer RE, et al. Developing digital therapeutics for chronic pain in primary care: a qualitative human-centered design study of providers' motivations and challenges. JMIR Form Res 2023 Feb 3;7:e41788. [doi: 10.2196/41788] [Medline: 36735284]
24. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales, 2nd edition: Psychology Foundation of Australia; 1996.
25. Wilson KG, Sandoz EK, Kitchens J, Roberts M. The Valued Living Questionnaire: defining and measuring valued action within a behavioral framework. Psychol Rec 2010 Apr;60(2):249-272. [doi: 10.1007/BF03395706]
26. Berman AH, Bergman H, Palmstierna T, Schlyter F. Drug Use Disorders Identification Test (DUDIT) manual. Karolinska Institutet. 2003. URL: https://media.cdn-a-klinikkasaatio.fi/wp-content/uploads/sites/3/2023/12/dudit_manual.pdf [accessed 2026-03-21]
27. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption--II. Addiction 1993 Jun;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
28. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500. [doi: 10.2196/11500] [Medline: 30973342]
29. Zhang CQ, Leeming E, Smith P, Chung PK, Hagger MS, Hayes SC. Acceptance and commitment therapy for health behavior change: a contextually-driven approach. Front Psychol 2017;8:2350. [doi: 10.3389/fpsyg.2017.02350] [Medline: 29375451]
30. Kravitz RL, Schmid CH, Marois M, et al. Effect of mobile device-supported single-patient multi-crossover trials on treatment of chronic musculoskeletal pain: a randomized clinical trial. JAMA Intern Med 2018 Oct 1;178(10):1368-1377. [doi: 10.1001/jamainternmed.2018.3981] [Medline: 30193253]
31. Jamison RN, Jurcik DC, Edwards RR, Huang CC, Ross EL. A pilot comparison of a smartphone app with or without 2-way messaging among chronic pain patients: who benefits from a pain app? Clin J Pain 2017 Aug;33(8):676-686. [doi: 10.1097/AJP.0000000000000455] [Medline: 27898460]
32. Thomson CJ, Pahl H, Giles LV. Randomized controlled trial investigating the effectiveness of a multimodal mobile application for the treatment of chronic pain. Can J Pain 2024 Aug 19;8(1):2352399. [doi: 10.1080/24740527.2024.2352399] [Medline: 39175941]
33. Vowles KE, Fink BC, Cohen LL. Acceptance and commitment therapy for chronic pain: a diary study of treatment process in relation to reliable change in disability. J Contextual Behav Sci 2014 Apr;3(2):74-80. [doi: 10.1016/j.jcbs.2014.04.003] [Medline: 27818931]
34. Goesling J, Lin LA, Clauw DJ. Psychiatry and pain management: at the intersection of chronic pain and mental health. Curr Psychiatry Rep 2018 Mar 5;20(2):12. [doi: 10.1007/s11920-018-0872-4] [Medline: 29504088]

Abbreviations

AUDIT: Alcohol Use Disorders Identification Test
DASS-21: Depression, Anxiety, and Stress Scale-21
DUDIT: Drug Use Disorders Identification Test
MAUQ: mHealth App Usability Questionnaire
VLQ: Valued Living Questionnaire

Edited by A Kushniruk; submitted 19.Nov.2025; peer-reviewed by E Ross, S P; revised version received 10.Feb.2026; accepted 21.Feb.2026; published 09.Apr.2026.

Please cite as:
Zbikowski SM, Masterson J, Burrola-Mendez Y, Hsu C, Ma KPK, Zhang Y, Waters D, Stephens KA
Participant Engagement With a Digital Behavioral Health App for Chronic Pain: Descriptive Secondary Analysis of a Feasibility Randomized Controlled Trial
JMIR Hum Factors 2026;13:e88122
URL: https://humanfactors.jmir.org/2026/1/e88122
doi:10.2196/88122

© Susan M Zbikowski, Jo Masterson, Yohali Burrola-Mendez, Chialing Hsu, Kris Pui Kwan Ma, Ying Zhang, Deanna Waters, Kari A Stephens. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 9.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR

XSL•FO
RenderX

Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

User Acceptance of Smart Home Emergency Response Systems: Mixed Methods Study

Michael Pantförder[1,2], MSc; Oliver Krüger[3], MSc; Annika Rietz[2], MSc; Colin Dzieia[2], MSc; Sven Meister[1,2], Prof Dr

[1]Health Informatics, Faculty of Health, Witten/Herdecke University, Alfred-Herrhausen-Straße 50, Witten, Germany

[2]Healthcare Department, Fraunhofer Institute for Software and Systems Engineering, Dortmund, Germany

[3]Office of Fire, Rescue and Civil Protection, Dortmund, Germany

Corresponding Author:
Michael Pantförder, MSc
Health Informatics, Faculty of Health, Witten/Herdecke University, Alfred-Herrhausen-Straße 50, Witten, Germany

Abstract

Background: Smart home emergency response systems (SHERS) leverage existing smart home infrastructure to detect critical events and alert emergency services without manual activation. Unlike personal emergency response systems, which require users to trigger alarms, SHERS initiate alerts autonomously. Although technically feasible, user acceptance has received limited empirical attention.

Objective: This study examined factors influencing the intention to adopt SHERS in private households, identifying key facilitators and barriers to acceptance.

Methods: A mixed methods study followed the Double Diamond framework. In the discover/define phases, expert interviews (n=3) and secondary data analysis informed persona and scenario development. In the "develop" phase, brainwriting workshops (6-3-5 method, n=12) generated design requirements translated into a low-fidelity prototype. In the "deliver" phase, an online survey (n=85) assessed acceptance using the Technology Usage Inventory. Structural equation modeling tested hypothesized relationships, and methodological triangulation integrated qualitative and quantitative findings.

Results: Perceived accessibility was the strongest positive predictor of intention to use (β=0.33, *P*=.02), while skepticism showed a marginally negative effect (β=−0.34, *P*=.06). The model explained 66% of variance in behavioral intention (R^2=0.66). Triangulation confirmed that concerns about complexity, false alarms, and data privacy underlie these effects. Experts emphasized that technology should support rather than replace human decision-making; workshop participants stressed intuitive setup and user control over alarm cancellation.

Conclusions: SHERS acceptance is primarily associated with perceived accessibility, while skepticism may act as a barrier. Developers should prioritize seamless integration with existing ecosystems, clear feedback mechanisms to prevent false alarms, and strong data protection.

(JMIR Hum Factors 2026;13:e93003) doi:10.2196/93003

KEYWORDS

smart home emergency response systems; SHERS; user acceptance; emergency response; mixed methods research; Double Diamond; structural equation modeling

Introduction

Background

Emergency response systems usually assume that affected individuals can seek help by calling emergency services, activating alarm buttons, or evacuating dangerous areas [1]. However, physical limitations, cognitive impairments, or environmental factors often stop timely self-rescue. This is especially true for older adults living alone and people facing sudden medical issues [2]. In Germany, almost 6 million people aged 65 years and older live in single-person households [3], which puts a significant number at risk of delays in emergency response.

Smart home technologies provide a technical basis for addressing these gaps [4]. Recent developments in ambient assisted living and health monitoring show that these systems can spot anomalies, track daily activities, and detect potential emergencies through patterns of inactivity or changes in the environment [5,6]. Building on these abilities, smart home emergency response systems (SHERS) represent a new type of safety-critical application that can automatically identify emergency situations and send alerts to emergency services, caregivers, or nearby responders without requiring users to take action [7].

Such systems may draw on heterogeneous sensor configurations already present in smart home environments, including motion

XSL•FO
RenderX

and presence sensors, environmental detectors for smoke or carbon monoxide, pressure-based bed and chair sensors that track prolonged inactivity, and wearable devices capable of detecting falls or abnormal vital signs [5,6]. By combining data from multiple sources, SHERS can distinguish routine patterns from potential emergencies and initiate alerts accordingly.

Unlike traditional personal emergency response systems that rely on manual activation, such as wearable alarm buttons [1], SHERS change the approach from user-initiated to system-initiated emergency management. When individuals cannot act, automated detection can close the delay between event and response and may help reduce mortality [8].

Related Work

Research on smart home applications for health and safety has grown considerably, with ambient assisted living and telecare studies showing that sensor-based monitoring can support older adults [1,9-11]. Intelligent fire-detection systems have shown improved accuracy over standalone smoke detectors [12,13]. However, most current implementations operate as closed-loop systems that alert residents locally but still rely on human action to contact emergency services.

Behavioral research shows that people may fail to hear alarms, underestimate risks, or attempt to resolve emergencies independently [14-18]. These behavioral bottlenecks show that systems must be able to act autonomously when human response is delayed or absent.

At the infrastructure level, frameworks such as the Internet of Emergency Services [19] and smart emergency response systems [20] focus primarily on professional responders rather than end-user integration.

Technology acceptance research identifies usefulness, ease of use, trust, and privacy as key determinants of smart home adoption [21]. Similar patterns have been observed specifically for health-related smart home applications, where concerns about data security and loss of autonomy can inhibit adoption even when functional benefits are recognized [22]. However, safety-critical automation that acts on behalf of users in emergencies remains underexplored.

Research Gap

Despite the technical feasibility of SHERS, empirical evidence on how users perceive *automated* emergency notification remains limited [23]. Three specific gaps emerge from the literature:

First, acceptance research rarely addresses safety-critical automation, where concerns about false alarms and privacy may differ from convenience-oriented applications [1,9,24].

Second, most studies examine acceptance factors in isolation rather than modeling their joint influence on behavioral intention. Understanding how psychological dispositions such as technology anxiety and curiosity, functional factors such as physical independence, and social factors such as prosocial orientation interact to shape adoption intentions requires multivariate approaches.

Third, the perspective of emergency professionals is rarely integrated into the design process.

Objectives

This study addresses these gaps by investigating the adoption and intention to use SHERS in private households. Guided by a user-centered design approach, it pursues three objectives:

1. *Discover and define* the operational context, user needs, and design requirements for SHERS through expert interviews with emergency professionals and scenario-based design workshops.
2. *Develop and evaluate* a system concept that integrates qualitative insights into concrete design artifacts (personas, scenarios, and low-fidelity prototypes).
3. *Assess the determinants* of user acceptance through a quantitative survey examining how psychological, functional, and social factors influence intention to use.

By triangulating qualitative and quantitative findings within a structured design framework, this study seeks to provide both conceptual guidance for SHERS development and empirical evidence on the factors shaping public acceptance of automated emergency technologies.

Methods

Research Approach

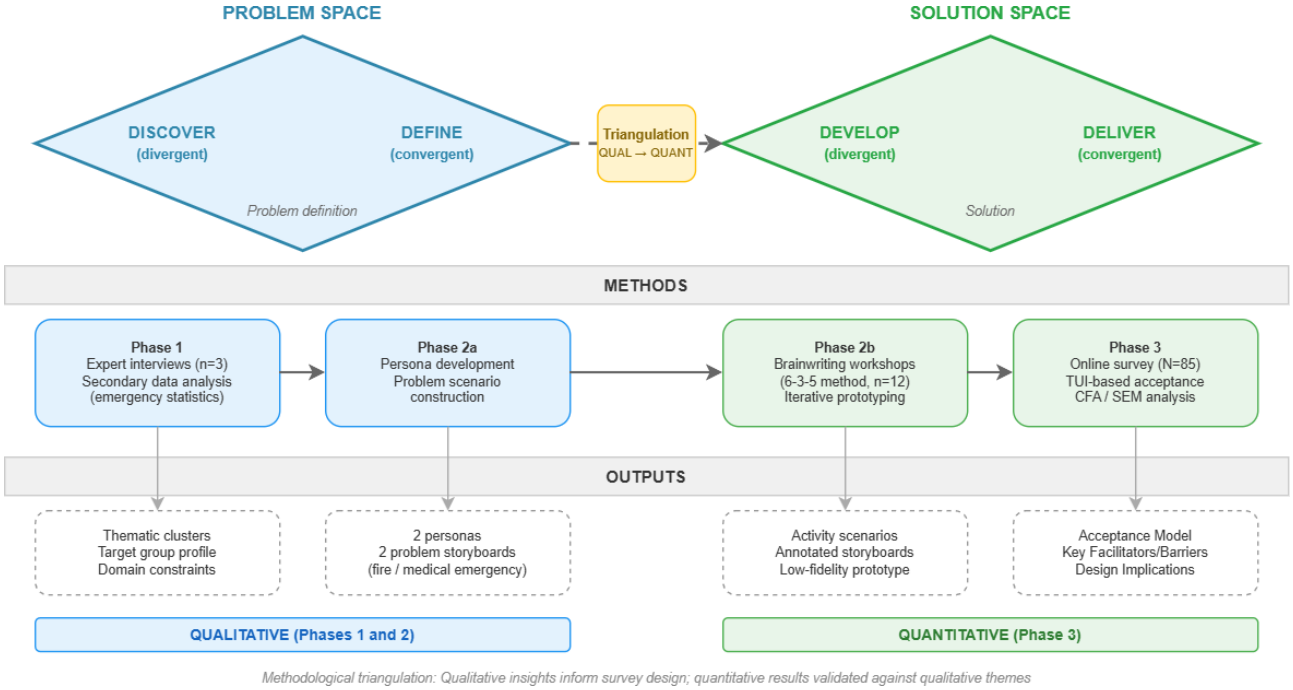
This study adopted a user-centered design approach [25] operationalized through the Double Diamond framework [26], which structures innovation processes into 4 iterative phases: discover, define, develop, and deliver. This framework involves different user groups to ensure solutions reflect actual needs [27,28].

The study combined qualitative and quantitative methods across the 4 phases in a sequential exploratory design [29]. The reporting of this mixed methods approach is informed by the Mixed Methods Article Reporting Standards (JARS-MMARS) [30], which provide guidance on transparently documenting how qualitative and quantitative components are integrated. By synthesizing insights from expert interviews, participatory design workshops, and a structured acceptance survey, findings from each phase informed and validated subsequent stages.

Figure 1 illustrates the research design, mapping each phase to its corresponding methods, outputs, and how findings were iteratively integrated.

XSL•FO
RenderX

Figure 1. Overview of the Double Diamond research design with methods and outputs per phase. CFA: confirmatory factor analysis; SEM: structural equation modeling; TUI: Technology Usage Inventory.



Ethical Considerations

The ethics committee of Witten/Herdecke University approved the research project on September 7, 2022 (protocol code: S-152/2022). All participants provided informed consent prior to participation. For the online survey, consent was obtained electronically on the first page of the questionnaire. All data were collected anonymously; no personally identifiable information was stored. No financial or material incentives were offered for participation in any phase of this study.

Phase 1: Discover—Domain Analysis and Problem Identification

The "discover" phase aimed to understand the operational context of emergency response and identify user groups who would benefit most from smart home-based emergency notification. This phase combined qualitative contextual inquiry with secondary data analysis.

Expert Interviews

We conducted 3 semistructured interviews with key stakeholders across the emergency response chain: a paramedic, a fire department unit leader, and a control center dispatcher. All 3 experts were affiliated with the Dortmund Fire Department, which served as the operational partner in the ADLeR (Automatisiertes Detektions-, Melde- und Leitsystem für Rettungskräfte [Automated Detection, Notification, and Guidance System for Emergency Responders]) project. Participants were purposively selected to represent the 3 core operational perspectives along the emergency response process: emergency dispatch and coordination, tactical incident management, and frontline medical care. This maximum-variation approach prioritized structural coverage over the number of interviews. As the interviews served primarily exploratory purposes within the "discover" phase, the resulting themes were subsequently validated and expanded through 2 participatory workshops with 12 additional participants (phase 2).

Following a problem-centered approach [31], interviews explored challenges in alerting, communication, and decision-making under uncertainty. Sessions were conducted using a virtual whiteboard, enabling participants to cocreate digital mind maps that captured workflows, pain points, and preliminary solution ideas. This participatory method aligns with principles of contextual design [32].

Interview data were thematically clustered by 2 researchers (OK and MP) to identify cross-cutting needs and constraints. Each researcher independently reviewed the mind map outputs and assigned thematic labels. Discrepancies were resolved through discussion until consensus was reached. As this step served primarily exploratory purposes, detailed COREQ (Consolidated Criteria for Reporting Qualitative Research) [33] reporting was not applied; however, the interview guide and mind map outputs are provided in Multimedia Appendix 1.

Secondary Data Analysis

To complement the qualitative findings, descriptive analysis was conducted using publicly available demographic data and anonymized emergency call statistics. The emergency data were provided by the Fire and Rescue Service of the City of Dortmund (Germany) and comprised all medically induced emergency calls recorded in 2021 (n=19,200). The dataset included age group classifications but contained no personally identifiable information. We analyzed age-stratified distributions to identify high-incidence population segments and inform target group selection for subsequent design activities.

Phase 1 Output

The combined findings provided a domain-grounded foundation for phase 2, including identified user needs, operational constraints, and a validated target group profile.

XSL•FO
RenderX

Phase 2: Define and Develop—Scenario-Based Design and Prototyping

Building on insights from phase 1, phase 2 used scenario-based design techniques [34] to translate empirical findings into concrete design artifacts. This phase corresponds to both the "define" stage (problem framing through personas and scenarios) and the "develop" stage (solution ideation through participatory workshops) of the Double Diamond model.

Persona and Scenario Construction

Based on the thematic clusters from expert interviews and demographic indicators from the statistical analysis, we developed 2 representative user personas, capturing attributes such as age, living situation, health status, and technology familiarity. These personas guided the creation of 2 problem scenarios, a residential fire and a fall-related medical emergency, representing prototypical high-risk situations. Each scenario was visualized through low-fidelity storyboards depicting the current emergency response process without technological intervention, highlighting behavioral bottlenecks and communication gaps.

Participatory Design Workshops

We conducted 2 brainwriting workshops to validate and expand the initial scenarios. Brainwriting is a structured written ideation technique distinct from verbal brainstorming. Participants independently generate ideas in writing before passing them to others for elaboration, which reduces dominance effects and encourages balanced contributions [35]. We applied the 6-3-5 variant, in which 6 participants each write 3 ideas within 5 minutes, then rotate sheets for iterative refinement.

Workshop 1 involved 6 emergency professionals (first responders and dispatch personnel) who reviewed the problem scenarios and contributed design ideas using the 6-3-5 method. The session focused on identifying pain points, desired system behaviors, and integration requirements.

Workshop 2 involved 6 domain experts from academic and applied research institutions with expertise in aging societies, human-machine interaction, and sociotechnical systems. This session addressed system-level considerations, including usability, information interpretability, and edge-case behavior.

Based on workshop outputs, the problem scenarios were iteratively transformed into activity scenarios incorporating envisioned system functionalities. These were synthesized into a low-fidelity prototype combining key system elements into a coherent, user-facing concept.

Data Analysis

We conducted workshop sessions online using digital whiteboards. Exported materials served as the primary qualitative data source. Two researchers (AR and MP) independently analyzed the content following qualitative content analysis principles, inductively coding idea clusters to identify recurring patterns and design requirements. Discrepancies were resolved through iterative discussion until consensus was reached. Reporting follows the COREQ guideline [33]; details are provided in Multimedia Appendix 2.

Phase 2 Output

Validated activity scenarios, annotated storyboards, and a low-fidelity prototype serving as stimulus material for the acceptance evaluation.

Phase 3: Deliver—Acceptance Evaluation

The "deliver" phase evaluated user acceptance of the proposed system concept through a quantitative online survey, completing the Double Diamond cycle by testing the developed solution with end users.

Survey Design and Constructs

Survey constructs were based on the Technology Usage Inventory (TUI) [36], which extends traditional acceptance models by incorporating individual traits including curiosity (NEU), anxiety (ANG), interest (INT), perceived usefulness (NÜT), user-friendliness (BEN), skepticism (SKE), and accessibility (ZUG). Abbreviations reflect the original German instrument. In the TUI framework, accessibility refers to the perceived availability of and ease of access to the technology, including how readily it can be obtained, set up, and integrated into daily routines. This construct does not refer to disability-related accessibility in the sense of barrier-free design. Two additional validated measures were included: the social compatibility (SOC) scale [37] for prosocial attitudes and the activities of daily living (ADL) scale [38] for functional independence.

The survey used a pre-post design: participants first completed demographic questions, smart home experience items, and baseline constructs (NEU, ANG, ADL, and SOC). They were then exposed to visual storyboards illustrating the emergency scenarios with integrated smart home functions (developed in phase 2). Post-exposure, participants rated the remaining TUI constructs and intention to use (ITU). This design enabled measurement of both general dispositions and concept-specific evaluations.

A pilot test (n=5) ensured item clarity and technical robustness. Survey implementation followed CHERRIES (Checklist for Reporting Results of Internet E-Surveys) reporting standards [39]; details are provided in Multimedia Appendix 3.

Participants

The survey was implemented as an open online questionnaire using LimeSurvey, hosted on a secure university server. It was distributed between October 2022 and June 2023 through multiple channels: housing cooperatives and senior advisory councils served as the primary recruitment partners, and the Dortmund Fire Department shared the survey link via social media (X, Facebook, and LinkedIn). Participants were also able to share the link independently. No financial or material incentives were offered. The only inclusion criterion was the ability to complete the German-language questionnaire; no further restrictions regarding age, health status, or technology experience were applied. A total of 128 individuals accessed the survey, of whom 85 completed all 4 sections (completion rate: 66%). Cookies were used to prevent repeated participation from the same device. The final sample included adults aged 20 to 89 years.

XSL•FO
RenderX

Statistical Analysis

We performed analyses in R (version 4.3.3) using lavaan (version 0.6 - 20). We evaluated internal consistency using Cronbach α, with values ≥0.70 considered acceptable [40]. A 2-stage modeling strategy combined confirmatory factor analysis (CFA) and structural equation modeling (SEM). Given ordered-categorical measurement and the number of indicators per construct, a diagonally weighted least squares (DWLS) estimator with robust corrections was used. We evaluated model fit using scaled indices (chi-square, comparative fit index [CFI], Tucker-Lewis index [TLI], root mean square error of approximation [RMSEA], and standardized root mean square residual [SRMR]), with CFI/TLI >0.90 and RMSEA <0.08 interpreted as acceptable.

While no universally accepted minimum sample size exists for DWLS estimation, simulation research suggests that DWLS with robust corrections can yield adequate parameter estimates and fit indices at sample sizes below 200, particularly when indicators are ordinal and factor loadings are moderate to strong [41,42]. Nevertheless, with 85 participants and 9 latent predictors, the ratio of sample size to model complexity is limited, which may reduce statistical power for detecting moderate effects and increase the uncertainty of individual path estimates. The results should therefore be interpreted with appropriate caution regarding the precision of nonsignificant coefficients.

Phase 3 Output

Validated acceptance model identifying key facilitators and barriers to adoption intention.

Methodological Integration and Triangulation

Findings from each phase directly informed subsequent stages. Specifically, the thematic clusters from expert interviews identified key concerns (such as false alarms, data privacy, and complexity of setup) that shaped the selection of acceptance constructs for the survey. The design requirements and activity scenarios generated in the workshops were translated into annotated storyboards that served as the stimulus material for the quantitative evaluation, ensuring that participants rated a concept grounded in domain-specific insights rather than an abstract technology description.

Methodological triangulation combined qualitative methods (interviews and workshops) with quantitative methods (survey and SEM), enabling cross-validation of findings. Table 1 summarizes the mapping of methods to phases and outputs.

Table . Mapping of Double Diamond phases to methods and outputs.

Phase	Double Diamond stage	Methods	Key outputs
1	Discover	Expert interviews (n=3) and secondary data analysis	Thematic clusters and target group identification
2a	Define	Persona development and problem scenario construction	2 personas and 2 problem storyboards
2b	Develop	Brainwriting workshops (12 participants) and iterative prototyping	Activity scenarios and low-fidelity prototype
3	Deliver	Online survey (n=85) and CFA[a]/SEM[b] analysis	Acceptance model and design implications

[a]CFA: confirmatory factor analysis.

[b]SEM: structural equation modeling.

Following JARS-MMARS recommendations, the qualitative components (phases 1 and 2) adhere to COREQ reporting standards [33], while the quantitative component (phase 3) follows CHERRIES guidelines [39]. The integration of both strands occurred at multiple points: qualitative findings shaped the survey stimulus materials, and quantitative results were interpreted in the context of qualitative themes identified earlier. This sequential design meant that the acceptance evaluation was built on domain-specific insights instead of abstract constructs.

Results

Phase 1: Domain Analysis and Target Group Identification

Thematic Clustering From Expert Interviews

Across the 3 interviews conducted, a total of 82 individual idea nodes were cocreated using digital whiteboarding tools, resulting in 3 structured mind maps (Multimedia Appendix 1). Thematic clustering revealed recurring barriers and requirements along 3 operational layers: dispatch coordination, responder navigation, and lay responder integration.

A central theme concerned the limited situational awareness available to dispatchers at the time of call intake. Interviewees described frequent uncertainty due to missing contextual data, such as the caller's precise location, environmental hazards, or patient-specific health risks.

First responders, in turn, reported considerable time losses caused by restricted access to buildings and unclear apartment layouts. Navigating unfamiliar environments, particularly under time pressure, was cited as a major operational burden. Participants noted that prior knowledge about access points, apartment configurations, or mobility limitations could significantly streamline response efforts.

Finally, the role of lay responders proved both promising and problematic. While early arrival can be beneficial, the lack of standardized guidance and limited integration into formal rescue

XSL•FO
RenderX

coordination restricts their effectiveness. Suggestions included structured onboarding, risk-adapted task delegation, and access to context-specific information prior to arrival.

Demographic Analysis and Emergency Statistics

Descriptive analysis of age-stratified emergency statistics confirmed the relevance of older adults as a disproportionately affected population segment. Although individuals aged 60 years and above represent less than one-third of the general population, they accounted for 64% (12,232/19,200) of all medically induced emergency calls in 2021 (Table 2). This disproportionality indicates that age is not merely a proxy for population size but a significant factor in emergency system usage. Thus, older adults were selected as the primary target group for subsequent design phases.

Table . Age distribution of medical emergencies in domestic settings reported in Dortmund (Germany), a subset of total emergency calls in 2021 (N=19,200).

Age group (years)	Emergencies, n (%)
0 - 9	666 (3.47)
10 - 19	438 (2.28)
20 - 29	1291 (6.72)
30 - 39	1310 (6.82)
40 - 49	1258 (6.55)
50 - 59	2005 (10.44)
60 - 69	2611 (13.60)
70 - 79	3479 (18.12)
80 - 89	4644 (24.19)
90 - 99	1457 (7.59)
100 - 110	41 (0.21)

These findings emphasize the need for improved information availability, environmental transparency, and support mechanisms for nonprofessional actors, which directly informed the scenario development in phase 2.

Phase 2: Scenario Development and Requirements Engineering

Findings From Scenario-Based Brainwriting

The structured idea generation yielded a diverse set of system-relevant concepts, which were subsequently clustered into key design implications. Participants emphasized the importance of real-time contextual awareness enabled by ambient and body-worn sensor data, precise geolocation, and adaptive alarm routing tailored to user roles and incident types. Suggestions also included infrastructure-level support for building access and orientation, such as dynamic lighting, illuminated door signage, and smart locking mechanisms.

Several contributions addressed systemic integration challenges, highlighting the need for interoperability across heterogeneous devices and platforms. The envisioned system was expected to balance automation with user override options and to provide differentiated access to sensitive data, depending on responder qualification and proximity. Participants also raised concerns regarding privacy, false-positive prevention, and reliability under edge-case conditions.

Scenario Iteration and Consolidation

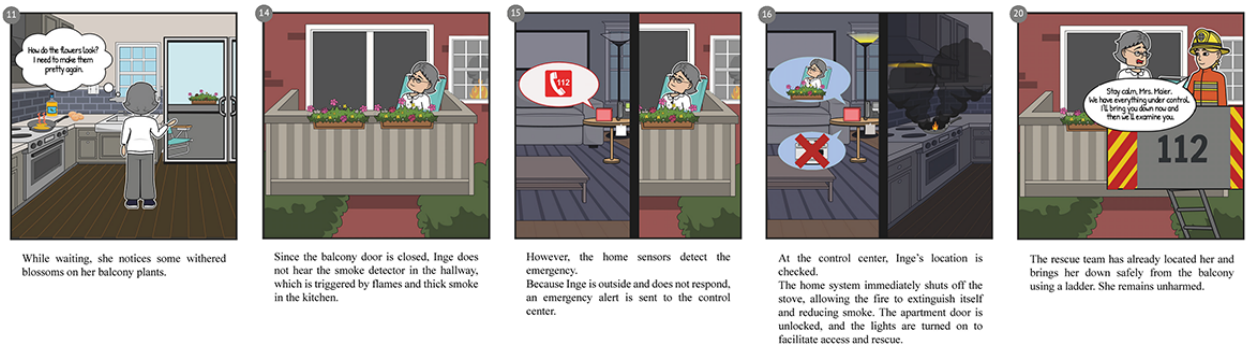
Based on these inputs, we revised the initial scenarios into detailed activity scenarios incorporating technical functionalities and user interactions, visualized through annotated low-fidelity storyboards.

A subsequent workshop evaluated these scenarios, focusing on usability, privacy, and error tolerance. Feedback was synthesized into a low-fidelity prototype illustrating key components and user touchpoints.

Figure 2 illustrates selected panels from the fire emergency scenario, depicting the sequence from unnoticed hazard through autonomous detection, system-initiated intervention, and successful responder arrival (see Multimedia Appendix 2 for complete storyboards).

XSL•FO
RenderX

Figure 2. Selected panels from the fire emergency scenario illustrating autonomous system intervention.



The resulting artifacts, including 29 design requirements and annotated storyboards (Multimedia Appendix 2), served as the conceptual foundation for the acceptance study conducted in phase 3.

Phase 3: Evaluation and Statistical Analysis

Descriptive Analysis

A total of 85 participants completed all 4 sections of the survey. The sample included individuals aged 20 to 89 years, with the most represented groups being those aged 30 - 39 years (n=23, 27%) and 20 - 29 years (n=16, 19%). A majority of participants identified as male (n=55, 65%), and most lived either in one-person (n=28, 33%) or two-person households (n=36, 42%). Regarding education, 44% (n=37) reported a college degree, while 29% (n=25) held a secondary school qualification.

Current smart home usage was reported by 61% (52/85) of participants, with 33% (28/85) indicating an intention to adopt such technologies within the next 12 months. Adoption intentions varied slightly by age, gender, and living arrangement (see Table 3).

Table . Demographic distribution of participants and their smart home technology usage patterns, including current usage and future intentions.

	Count (n=85), n (%)	Current smart home usage, n (%)	Planned smart home adoption, n (%)
Total		52 (61)	28 (33)
Sex			
Female	30 (35)	15 (50)	5 (17)
Male	55 (65)	37 (67)	23 (42)
Age (years)			
20 - 29	16 (19)	9 (56)	6 (38)
30 - 39	23 (27)	14 (61)	9 (39)
40 - 49	13 (15)	9 (69)	4 (31)
50 - 59	12 (14)	8 (67)	5 (42)
60 - 69	9 (11)	7 (78)	3 (33)
70 - 79	8 (9)	4 (50)	1 (12)
80 - 89	4 (5)	1 (25)	0 (0)
Education			
Intermediate school	25 (29)	12 (48)	4 (16)
High school	15 (18)	11 (73)	7 (47)
College	37 (44)	24 (65)	13 (35)
Doctorate/PhD	8 (9)	5 (62)	4 (50)
Living situation			
Single-person household	28 (33)	18 (64)	6 (21)
Two persons	36 (42)	23 (64)	13 (36)
Three or more persons	21 (25)	11 (52)	9 (43)

Descriptive statistics for all psychometric constructs and individual items are summarized in Table 4. Participants reported high levels of functional independence, with mean scores above 4.6 across all ADL items (eg, ADL1: mean 4.81, SD 0.72). Social compatibility showed moderate to high values (SOC1-SOC5: mean>4.2), though lower agreement was

XSL•FO
RenderX

observed for items reflecting burdensome social responsibility (SOC6-SOC9: mean≈3‐4).

XSL•FO
RenderX

Table . Descriptive statistics of item responses.

Item	Mean (SD)
ADL (activities of daily living)	
ADL1	4.81 (0.72)
ADL2	4.68 (0.88)
ADL3	4.86 (0.54)
ADL4	4.66 (0.92)
SOC (social compatibility)	
SOC1	5.00 (1.02)
SOC2	4.49 (1.14)
SOC3	5.12 (1.07)
SOC4	4.95 (1.06)
SOC5	4.24 (1.26)
SOC6	3.01 (1.56)
SOC7	3.91 (1.57)
SOC8	4.18 (1.37)
SOC9	4.34 (1.34)
SOC10	4.34 (1.32)
NEU (curiosity)	
NEU1	6.04 (1.19)
NEU2	4.26 (1.96)
NEU3	5.72 (1.1)
NEU4	4.82 (1.72)
ANG (anxiety)	
ANG1	2.38 (1.58)
ANG2	2.31 (1.47)
ANG3	2.07 (1.32)
ANG4	2.01 (1.34)
INT (interest)	
INT1	5.58 (1.49)
INT2	4.86 (1.54)
INT3	4.88 (1.58)
INT4	4.96 (1.58)
NÜT (usefulness)	
NÜT1	5.99 (1.11)
NÜT2	4.31 (1.63)
NÜT3	4.84 (1.57)
NÜT4	4.31 (1.59)
SKE (skepticism)	
SKE1	4.20 (1.7)
SKE2	2.54 (1.45)
SKE3	2.34 (1.31)
SKE4	2.36 (1.28)
BEN (user-friendliness)	

XSL•FO
RenderX

Item	Mean (SD)
BEN1	5.20 (1.35)
BEN2	5.21 (1.22)
BEN3	5.09 (1.44)
ZUG (accessibility)	
ZUG1	3.60 (1.49)
ZUG2	4.41 (1.48)
ZUG3	4.38 (1.46)
ITU (intention to use)	
ITU1	78.15 (20.53)
ITU2	70.34 (23.98)
ITU3	81.59 (17.95)

Pre-exposure assessments revealed high curiosity (NEU1: mean 6.04, SD 1.19) and low technology-related anxiety (ANG1: mean 2.38, SD 1.58). Following the storyboard intervention, participants rated the proposed system as generally useful (NÜT1: mean 5.99, SD 1.11) and user-friendly (BEN1: mean 5.20, SD 1.35), with only moderate skepticism. Accessibility received more varied evaluations (ZUG1: mean 3.60, SD 1.49).

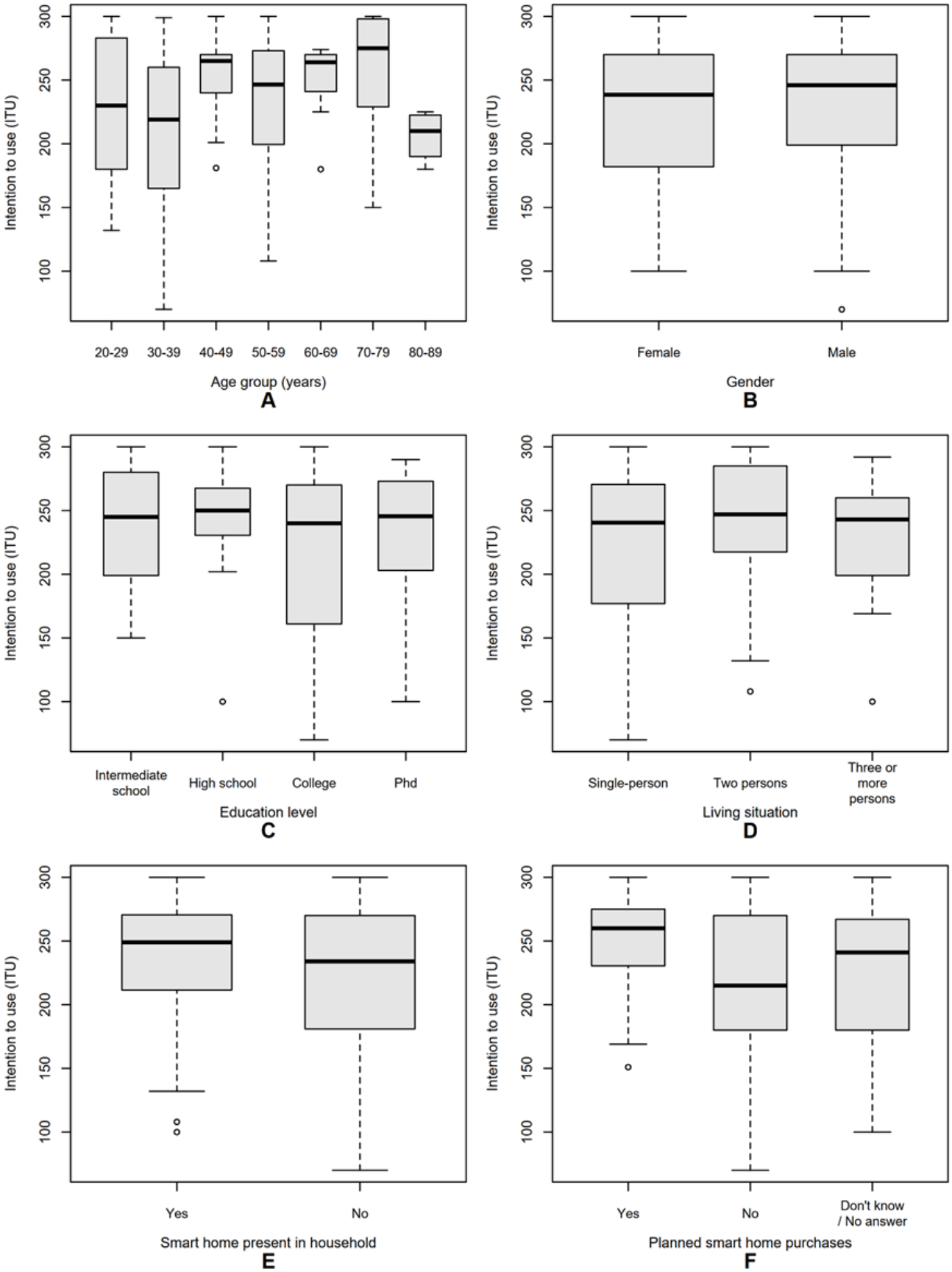
We measured the ITU via a 0‐100 visual analog scale. Participants reported moderately high usage intentions, with item means ranging from 70.34 (SD 23.98) to 81.59 (SD 17.95) across the three ITU items.

Sociodemographic Variation in Intention to Use

Before conducting the latent-variable analyses, we performed exploratory comparisons to examine whether ITU differed across sociodemographic and contextual subgroups. Figure 3 displays the distribution of ITU scores by age group, education level, gender, living situation, and smart-home–related characteristics (ie, presence of smart-home devices in the household and planned future purchases).

XSL•FO
RenderX

Figure 3. Distribution of intention to use (ITU) scores across sociodemographic and contextual subgroups. Panels show median, IQR, and outliers for (A) age group, (B) education level, (C) gender, (D) living situation, (E) planned smart-home purchases, and (F) smart-home presence in the household.



Across age categories (Figure 3A), median ITU scores appeared largely comparable, with slightly higher intentions in the youngest (20 - 29 years) and older (70 - 79 years) participants, but no consistent age-related trend. Regarding smart-home–related factors, participants who already had smart-home devices in their household (Figure 3E) tended to

XSL•FO
RenderX

report slightly higher ITU scores than those without such devices, although the spread of values was wide and overlapping. Similarly, those who planned to purchase smart-home technologies in the near future (Figure 3F) reported somewhat elevated intentions relative to those without purchase plans, but again without clear statistical separation. Given the absence of significant mean-level differences in ITU across sociodemographic subgroups, these variables were not included as exogenous predictors in the SEM.

Overall, these analyses indicated no meaningful sociodemographic bias in ITU across the examined groups. This supports the focus on psychological and attitudinal constructs as the main determinants in the subsequent latent-variable modeling.

Preliminary Reliability and Correlational Analyses

Before estimating the latent variable model, we examined reliability and zero-order associations among the observed constructs. All multi-item scales demonstrated good internal consistency, with Cronbach α values ranging from 0.72 (BEN) to 0.93 (ADL) (see Table 5).

Table . Internal consistency and bivariate associations of study constructs with intention to use.

Construct	Cronbach α	Correlation with ITU[a]	β coefficient	*P* value
ADL (activities of daily living)	0.930	−0.064	−1.265	.56
SOC (social compatibility)	0.838	0.369	2.512	.001
NEU (curiosity)	0.813	0.366	4.158	.001
ANG (anxiety)	0.828	−0.112	−1.341	.31
INT (interest)	0.869	0.236	2.515	.03
NÜT (usefulness)	0.784	0.600	7.230	<.001
SKE (skepticism)	0.762	−0.539	−6.824	<.001
BEN (user-friendliness)	0.720	0.238	6.204	.03
ZUG (accessibility)	0.780	0.441	6.661	<.001

[a]ITU: intention to use.

Bivariate correlations showed that NÜT (r=0.60, P<.001), ZUG (r=0.44, P<.001), SOC (r=0.37, P=.001), NEU (r=0.37, P=.001), INT (r=0.24, P=.03), and BEN (r=0.24, P=.03) each correlated positively with ITU. In contrast, SKE correlated negatively (r=–0.54, P<.001), while ANG (r=–0.11, P=.31) and ADL (r=–0.06, P=.56) showed small, nonsignificant associations.

Simple linear regressions mirrored these bivariate patterns: positive effects on ITU emerged for NÜT, ZUG, SOC, NEU, INT, and BEN (all P≤.03), whereas SKE exerted a significant negative effect (β=–6.82, P<.001). Neither ANG nor ADL reached statistical significance (P>.30).

Together, these findings supported the proposed directionality of the theoretical model and justified proceeding to latent-variable analyses.

Confirmatory Factor and Structural Equation Modeling

We applied a comprehensive confirmatory factor and SEM approach using a DWLS estimator with robust corrections, appropriate for ordered-categorical indicators and the comparatively high number of observed variables per construct. The 9-factor measurement model, plus a 3-indicator factor for ITU, demonstrated an acceptable overall model fit on scaled indices (χ^2_{815}=1052.21, P<.001; CFI=0.928; TLI=0.920; RMSEA=0.059, 90% CI 0.048 - 0.069; SRMR=0.107; Bentler-SRMR=0.0923) (Table 6).

XSL•FO
RenderX

Table . Scaled global model fit (DWLS[a], robust) for the confirmatory factor and structural model (n=85).

Index	Value	Interpretation
Chi-square (*df*=815)	1052.21 (*P*<.001)	N/A[b]
CFI[c]	0.928	Acceptable
TLI[d]	0.920	Acceptable
RMSEA[e]	0.059 (90% CI 0.048 - 0.069)	Good
SRMR[f]	0.107	Moderate
Bentler-SRMR	0.092	Acceptable given categorical indicators

[a]DWLS: diagonally weighted least squares.

[b]N/A: not applicable.

[c]CFI: comparative fit index.

[d]TLI: Tucker-Lewis index.

[e]RMSEA: root mean square error of approximation.

[f]SRMR: standardized root mean square residual.

Although both SRMR indices were slightly above conventional thresholds [43], simulation research indicates that SRMR tends to be inflated when using categorical estimators such as DWLS, particularly in complex models with many indicators [44]. The combination of CFI and TLI values close to 0.93 and an RMSEA below 0.06, given the categorical measurement level and the model's complexity, supports an adequate and stable representation of the empirical covariance structure. All standardized factor loadings were positive, substantial, and statistically significant (*P*<.001), ranging from 0.46 to 0.99, with most exceeding 0.70. No cross-loadings or Heywood cases were detected, and the factor solution remained stable across estimation runs, confirming the stability of the measurement specification.

The SEM maintained an acceptable overall fit that closely mirrored the CFA solution and accounted for 66.1% of the variance in ITU (R^2=0.661). Among the latent predictors, ZUG was the strongest positive determinant (β=0.325, *P*=.02). In contrast, SKE showed a negative association with ITU that approached but did not reach conventional significance levels (β=−0.336, *P*=.06). NÜT (β=0.357, *P*=.15) and SOC (β=0.129, *P*=.25) showed positive but nonsignificant trends, while BEN, INT, NEU, ANG, and ADL displayed no meaningful associations (|β|≤0.26, all *P*>.10; Table 7). A comparison between the full model (with all 9 predictors) and a reduced model (retaining only significant predictors) revealed minimal differences in fit indices (ΔCFI=−0.003; ΔRMSEA=0.004; ΔSRMR=0.004). Although a nested model test indicated a statistically significant decrement in fit (*P*=.001), the negligible changes in practical fit indices suggest that retaining theoretically motivated but statistically nonsignificant constructs did not materially distort the overall model.

Table . Standardized path coefficients predicting intention to use (scaled SEM[a] and DWLS[b]).

Predictor → ITU (intention to use)	β (std)	SE	*z* score	*P* value	Significance
ZUG (accessibility)	0.325	0.239	2.33	.02	Significant
SKE (skepticism)	−0.336	0.312	−1.85	.06	Marginal
NÜT (usefulness)	0.357	0.431	1.42	.15	—[c]
SOC (social compatibility)	0.129	0.191	1.16	.25	—
BEN (user-friendliness)	−0.164	0.321	−0.87	.38	—
INT (interest)	0.070	0.323	0.37	.71	—
NEU (curiosity)	0.002	0.373	0.01	.99	—
ANG (anxiety)	−0.005	0.416	−0.02	.98	—
ADL (activities of daily living)	−0.255	0.324	−1.35	.18	—

[a]SEM: structural equation modeling.

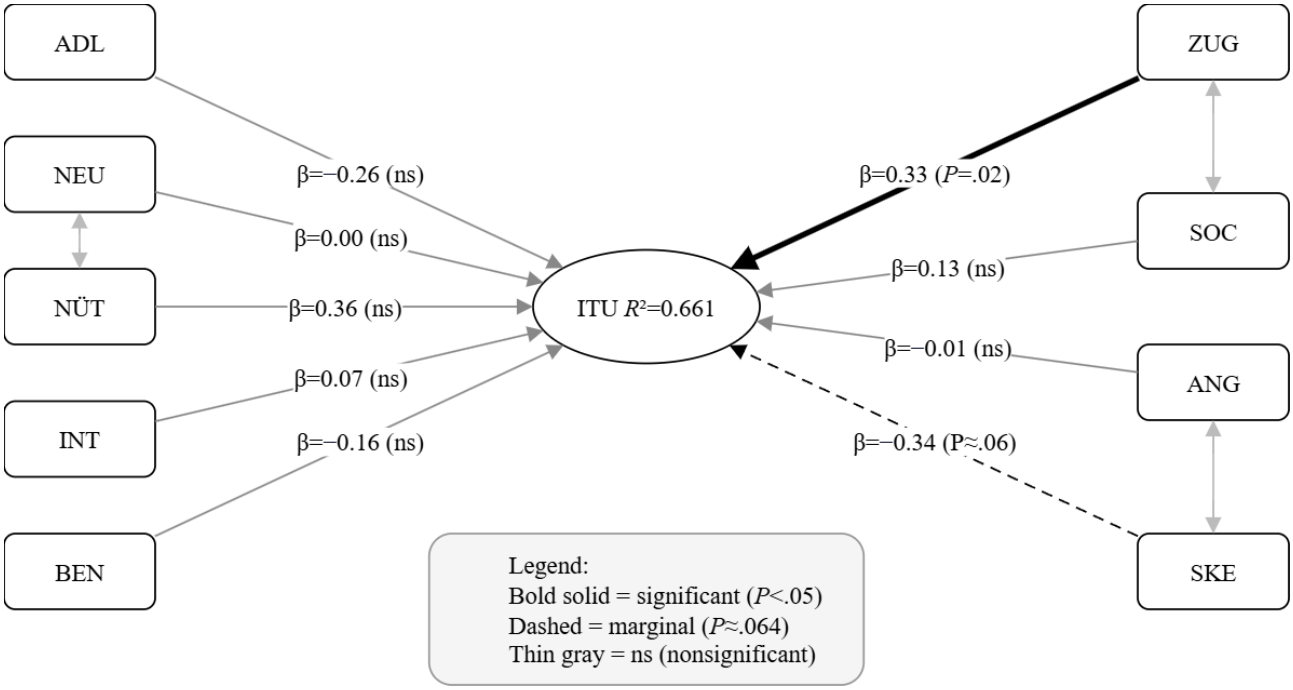
[b]DWLS: diagonally weighted least squares.

[c]Not applicable.

XSL•FO
RenderX

The final structural equation model is depicted in Figure 4, illustrating the pattern of relationships among the 9 latent predictors and ITU. All exogenous latent constructs were allowed to covary, controlling for shared variance among psychological, functional, and social factors. These correlations are depicted as curved double-headed arrows in the structural model but were not interpreted further, as they were not part of the primary hypotheses.

Figure 4. Structural equation model predicting intention to use (ITU) from psychological, functional, and social determinants. ADL: activities of daily living; ANG: anxiety; BEN: user-friendliness; INT: interest; ITU: intention to use; NEU: curiosity; NÜT: usefulness; SKE: skepticism; SOC: social compatibility; ZUG: accessibility.



To assess the robustness of the structural findings, we conducted several complementary diagnostics. To ensure that the constructs included in the latent-variable model were empirically distinct and not affected by multicollinearity, we performed additional post hoc diagnostics. The latent factor correlations extracted from the CFA model indicated satisfactory discriminant validity, as none of the interfactor correlations approached unity and all remained well below critical thresholds ($|r|<0.85$). This supports the conceptual distinctness of the modeled constructs.

Variance inflation factors (VIF) confirmed the absence of problematic multicollinearity, with values ranging from 1.24 to 1.81 (all VIF<5), indicating that the predictors were sufficiently independent. Alternative model specifications (eg, method factors, item parceling) provided no meaningful improvement and were not retained.

The findings provide a coherent picture of the psychological and contextual factors underlying intention to use. The next section discusses their theoretical and practical implications.

Discussion

Principal Findings

This mixed methods study identified key drivers and barriers of SHERS adoption by combining domain analysis, scenario-based design, and SEM.

The structural model explained two-thirds of the variance in behavioral intention (R^2=0.66), underscoring the strong explanatory power of psychological and usability-related factors. Among all predictors, ZUG had the strongest positive effect on intention to use, emphasizing that participants' perception of how easily the technology can be set up and integrated matters most for adoption. Conversely, SKE showed a negative association with intention that approached statistical significance (P=.06), suggesting that residual doubt about reliability, privacy, or false alarms may reduce willingness to adopt. Other theoretically relevant constructs such as NÜT and SOC displayed positive but nonsignificant trends, indicating that their effects may be more indirect or context dependent.

These findings suggest that beyond functional utility, the perceived ease of obtaining and setting up the technology may be more relevant to adoption intentions than its perceived benefits alone.

It is important to note that only ZUG reached conventional statistical significance as a predictor of intention to use. The negative trend for SKE (P=.06) is suggestive but does not meet the conventional threshold. All remaining constructs were nonsignificant in the multivariate model, despite several showing meaningful bivariate associations with intention to use. This pattern is consistent with the limited statistical power of the sample (n=85) to detect moderate effects among multiple correlated predictors and may also partly reflect the storyboard-based evaluation format, which limits experiential assessment of constructs such as BEN. Accordingly, interpretations involving nonsignificant predictors should be considered exploratory.

XSL•FO
RenderX

Integration of Qualitative and Quantitative Findings

Methodological triangulation allowed cross-validation of qualitative and quantitative findings.

The quantitative analysis identified ZUG as the strongest predictor of intention to use. This finding directly mirrors qualitative themes: Expert 3 (unit leader) noted that "technical setup is complex and costly," stressing that incentives or mandates would be necessary. Experts also stressed integration with existing infrastructure, proposing connections between "smart-watches, home emergency systems, and fire alarm systems" [Expert 3]. These convergent insights suggest accessibility represents a reliable determinant of adoption.

SKE was a negative predictor (P=.06), reflecting concerns about reliability, false alarms, and data privacy. Expert 1 (paramedic) proposed a "filter function to distinguish false alarms from actual emergencies" and mechanisms to "allow user intervention to cancel the alarm." Expert 2 (dispatcher) emphasized that cameras should only be activated "after explicit request and consent." These findings suggest skepticism is rooted in concrete operational and privacy concerns.

NÜT showed a positive but nonsignificant effect despite strong bivariate correlations. Experts recognized value for "vulnerable groups" [Expert 3], yet emphasized that determining whether "it is truly an emergency" should remain a "decision to be made by the incident commander" [Expert 3], and that "situational assessment by people on site is necessary" [Expert 2]. This tension may explain the attenuated effect. However, this interpretation remains exploratory, as the nonsignificant path coefficient may also reflect insufficient statistical power to detect a moderate effect at n=85, or measurement limitations arising from the storyboard-based evaluation format.

An overarching theme concerns the balance between automation and human control. Experts consistently emphasized that "human decision-making competence cannot be replaced" [Expert 1] and that "technology should initially only propose a classification" [Expert 1]. Expert 3 cautioned that "weighing up 'deploy or not' under the influence of data or a second opinion is highly critical." These insights explain why both accessibility and low skepticism are central to acceptance.

Triangulation strengthens confidence in results while revealing contextual nuances.

Comparison With Prior Work

The identified role of accessibility is consistent with findings from previous acceptance studies in digital health and smart-home contexts, where usability barriers frequently outweighed perceived benefits in determining behavioral intention [45,46].

The negative association between skepticism and intention mirrors earlier observations in related domains such as telemonitoring and artificial intelligence–assisted care [1,9]. Concerns about data privacy, false alarms, and perceived loss of control are well-documented inhibitors of trust in automated systems. The present results suggest that skepticism may reduce adoption intentions, although the association did not reach conventional significance (P=.06). This pattern is consistent with prior findings and points to trust calibration as a relevant design consideration.

Unlike some prior studies that found strong demographic effects, such as age-related declines in technology acceptance or gender-specific differences in curiosity and anxiety [47-49], the current analysis revealed no significant sociodemographic bias in intention to use. This suggests that, when users are adequately informed and exposed to a concrete system concept, psychological and design-related factors may play a larger role than demographic predispositions in shaping adoption intentions. This aligns with recent evidence indicating that participatory exposure and scenario-based demonstrations can reduce pre-existing reservations toward unfamiliar technologies [50].

Practical Implications

From a design perspective, the findings highlight the importance of making accessibility tangible through intuitive setup, clear user guidance, and compatibility with existing smart home ecosystems. Developers should focus on reducing friction points such as complex configuration, opaque data handling, or unclear error feedback.

Addressing skepticism requires trust-building measures that communicate reliability and data protection standards. Clear alert protocols, privacy assurances, and demonstrated reliability under real conditions can increase confidence in automated emergency technologies. For policymakers and emergency service providers, these findings support the case for public communication and certification frameworks focused on safety and interoperability.

The consistency between expert and user perspectives observed in this study supports co-design approaches that bring emergency professionals and end users together early in the development process. Such collaboration can connect technical, psychological, and organizational perspectives and support the wider adoption of SHERS.

Limitations

Several limitations should be considered. First, the quantitative evaluation relied on a conceptual rather than interactive prototype presented through annotated storyboards. Although this format is suitable for early-phase acceptance testing [34], it may have limited participants' ability to evaluate usability and reliability. Constructs that depend on direct interaction experience, particularly BEN, may be underestimated or imprecisely captured when assessed on the basis of visual scenarios alone. This could partly explain the nonsignificant effect of BEN in the structural model, as participants lacked the hands-on experience needed to form differentiated usability judgments. Conversely, constructs reflecting general attitudes, such as curiosity or skepticism, are less likely to be affected by prototype fidelity. Future studies should use functional or simulated prototypes to validate these findings under more realistic conditions.

Second, although DWLS estimation is robust to smaller samples with ordinal data, the sample size (n=85) limits generalizability and the ability to test subgroup interactions. The recruitment strategy relied on housing cooperatives and senior advisory

XSL•FO
RenderX

councils, which may have introduced selective bias by overrepresenting individuals with existing institutional ties and underrepresenting isolated, low-income, or digitally disengaged populations. These groups may stand to benefit most from SHERS but are typically harder to reach through organizational channels. The sample was also skewed toward higher educated (44% college degree) and male (65%) participants, which may affect the transferability of findings. Individuals with limited digital literacy or those living in underresourced housing conditions may hold different attitudes toward automated emergency systems. Larger and more diverse samples, including targeted recruitment of digitally excluded and lower-income populations, are needed to confirm the stability and generalizability of the observed effects.

The limited sample size also increases the risk of type II errors, particularly for predictors with moderate effect sizes. Several constructs with meaningful bivariate associations with intention to use, such as NÜT (r=0.60), did not reach significance in the multivariate model. While this pattern may partly reflect shared variance among correlated predictors, insufficient power to detect moderate effects in the presence of 9 simultaneous paths cannot be ruled out.

Third, self-reported data may have introduced social desirability bias. Behavioral or longitudinal observations could strengthen future analyses.

Finally, contextual variables (cost, connectivity, and interoperability) were not included and should be addressed in future research.

Future Research

Future research should examine long-term acceptance trajectories and how trust develops during actual system use.

Cross-stakeholder studies should examine how acceptance differs between end users, emergency professionals, and service operators.

Evaluating artificial intelligence–driven emergency systems under field conditions will be necessary for safe deployment.

Conclusions

This study examined the psychological and functional determinants of SHERS acceptance. Accessibility was the strongest statistically significant predictor of intention to use, while skepticism showed a negative trend that suggests it may function as a barrier to adoption. These results indicate that technological development alone is insufficient without user-centered design, open communication, and strategies that address adoption concerns early.

The mixed methods approach allowed us to connect design-level insights with statistical evidence on acceptance factors. Taken together, the findings offer practical orientation for developing emergency technologies that are trusted, usable, and compatible with the living environments of vulnerable populations.

Acknowledgments

The authors thank all participating experts and survey respondents for their valuable contributions to this research.

Funding

This research is part of the project ADLeR, funded by the German Federal Ministry of Research, Technology and Space (BMFTR) under grant 13N16805.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

MP conceptualized the study, designed the mixed methods framework, conducted data analysis, and prepared the manuscript. OK conducted the expert interviews in phase 1 and assisted in the preparation of the demographic and contextual data. AR cofacilitated the scenario-based design workshops and contributed to qualitative data analysis and manuscript revision. CD contributed to data analysis and proofreading of the manuscript. SM supervised the study and provided critical feedback on methodological design and interpretation of results. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Expert interview materials, including the semistructured interview guide and thematic mind maps from phase 1 domain analysis.
[DOCX File, 844 KB - humanfactors_v13i1e93003_app1.docx]

Multimedia Appendix 2

XSL•FO
RenderX

Qualitative reporting and design artifacts, including the COREQ checklist for phase 2 workshops and design requirements derived from scenario-based brainwriting sessions.
[DOCX File, 3388 KB - humanfactors_v13i1e93003_app2.docx]

Multimedia Appendix 3
Survey instrument and CHERRIES checklist for phase 3 acceptance evaluation, including the adapted Technology Usage Inventory questionnaire.
[DOCX File, 36 KB - humanfactors_v13i1e93003_app3.docx]

References

1. Stokke R. The personal emergency response system as a technology innovation in primary health care services: an integrative review. J Med Internet Res 2016 Jul 14;18(7):e187. [doi: 10.2196/jmir.5727] [Medline: 27417422]
2. Li KF. Smart home technology for telemedicine and emergency management. J Ambient Intell Humaniz Comput 2013;4(5):535-546. [doi: 10.1007/s12652-012-0129-8] [Medline: 32218875]
3. Fast 6 millionen ältere menschen leben allein. Statistisches Bundesamt. URL: https://www.destatis.de/DE/Presse/Pressemitteilungen/2021/09/PD21_N057_12411.html [accessed 2026-4-1]
4. Marikyan D, Papagiannidis S, Alamanos E. A systematic review of the smart home literature: a user perspective. Technol Forecast Soc Change 2019 Jan;138:139-154. [doi: 10.1016/j.techfore.2018.08.015]
5. Wilhelm S, Wahl F. Emergency detection in smart homes using inactivity score for handling uncertain sensor data. Sensors (Basel) 2024 Oct 12;24(20):6583. [doi: 10.3390/s24206583] [Medline: 39460065]
6. Morita PP, Sahu KS, Oetomo A. Health monitoring using smart home technologies: scoping review. JMIR Mhealth Uhealth 2023 Apr 13;11(1):e37347. [doi: 10.2196/37347] [Medline: 37052984]
7. Rashidi P, Mihailidis A. A survey on ambient-assisted living tools for older adults. IEEE J Biomed Health Inform 2013 May;17(3):579-590. [doi: 10.1109/jbhi.2012.2234129] [Medline: 24592460]
8. Mowbray F, Mills F, Symons C, Amlôt R, Rubin GJ. A systematic review of the use of mobile alerting to inform the public about emergencies and the factors that influence the public response. J Contingencies Crisis Manag 2024 Mar;32(1):e12499. [doi: 10.1111/1468-5973.12499]
9. Steele R, Lo A, Secombe C, Wong YK. Elderly persons' perception and acceptance of using wireless sensor networks to assist healthcare. Int J Med Inform 2009 Dec;78(12):788-801. [doi: 10.1016/j.ijmedinf.2009.08.001] [Medline: 19717335]
10. Aggar C, Sorwar G, Seton C, Penman O, Ward A. Smart home technology to support older people's quality of life: a longitudinal pilot study. Int J Older People Nursing 2023 Jan;18(1). [doi: 10.1111/opn.12489]
11. Kim KI, Gollamudi SS, Steinhubl S. Digital technology to enable aging in place. Exp Gerontol 2017 Feb;88:25-31. [doi: 10.1016/j.exger.2016.11.013] [Medline: 28025126]
12. Kweon SJ, Park JH, Park CO, Yoo HJ, Ha S. Wireless kitchen fire prevention system using electrochemical carbon dioxide gas sensor for smart home. Sensors (Basel) 2022 May 24;22(11):3965. [doi: 10.3390/s22113965] [Medline: 35684586]
13. Saeed F, Paul A, Rehman A, Hong WH, Seo H. IoT-based intelligent modeling of smart home environment for fire prevention and safety. J Sens Actuator Netw 2018 Mar;7(1):11. [doi: 10.3390/jsan7010011]
14. Proulx G, Laroche C, Latour JC. Audibility problems with fire alarms in apartment buildings. Proc Hum Factors Ergon Soc Annu Meet 1995 Oct;39(15):989-993. [doi: 10.1177/154193129503901505]
15. Bruck D, Thomas I. Towards a better smoke alarm signal – an evidence based approach. Fire Saf Sci 2008;9:403-414. [doi: 10.3801/IAFSS.FSS.9-403]
16. Taylor M, Appleton D, Fielding J, Oakford G. An exploration of householder injuries sustained fighting dwelling fires. Fire Saf J 2022 Jan;127:103519. [doi: 10.1016/j.firesaf.2021.103519]
17. Proulx G, Fahy RF. The time delay to start evacuation: review of five case studies. Fire Saf Sci 1997;5:783-794. [doi: 10.3801/IAFSS.FSS.5-783]
18. Buffington T, Ezekoye OA. Statistical analysis of fire department response times and effects on fire outcomes in the United States. Fire Technol 2019 Nov;55(6):2369-2393. [doi: 10.1007/s10694-019-00870-4]
19. Damaševičius R, Bacanin N, Misra S. From sensors to safety: Internet of Emergency Services (IoES) for emergency response and disaster management. J Sens Actuat Netw 2023;12(3):41. [doi: 10.3390/jsan12030041]
20. Zander J, Mosterman PJ, Padir T, Wan Y, Fu S. Cyber-physical systems can make emergency response smart. Procedia Eng 2015;107:312-318. [doi: 10.1016/j.proeng.2015.06.086]
21. Rhee JH, Ma JH, Seo J, Cha SH. Review of applications and user perceptions of smart home technology for health and environmental monitoring. J Comput Des Eng 2022 May 11;9(3):857-889. [doi: 10.1093/jcde/qwac030]
22. Dermody G, Wadsworth D, Dunham M, Glass C, Fritz R. Factors affecting clinician readiness to adopt smart home technology for remote health monitoring: systematic review. JMIR Aging 2024 Dec 5;7:e64367. [doi: 10.2196/64367] [Medline: 39012852]
23. Mashal I, Shuhaiber A, AL-Khatib AW. User acceptance and adoption of smart homes: a decade long systematic literature review. Int J Data Netw Sci 2023;7(2):533-552. [doi: 10.5267/j.ijdns.2023.3.017]

XSL•FO
RenderX

24. Vrančić A, Zadravec H, Orehovački T. The role of smart homes in providing care for older adults: a systematic literature review from 2010 to 2023. Smart Cities 2024 Aug;7(4):1502-1550. [doi: 10.3390/smartcities7040062]
25. ISO 9241-210:2019: Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems. International Organization for Standardization. 2019 Jul. URL: https://www.iso.org/standard/77520.html [accessed 2026-04-01]
26. The Double Diamond. Design Council. 2019. URL: https://www.designcouncil.org.uk/our-resources/the-double-diamond/ [accessed 2026-04-01]
27. Soehnchen C, Rietz A, Weirauch V, Meister S, Henningsen M. Creating an intercultural user-centric design for a digital sexual health education app for young women in resource-poor regions of Kenya: qualitative self-extended Double Diamond model for requirements engineering analysis. JMIR Form Res 2023 Nov 3;7:e50304. [doi: 10.2196/50304] [Medline: 37921860]
28. Wolf E, Morisse K, Meister S. Identifying design requirements for an interactive physiotherapy dashboard with decision support for clinical movement analysis of musicians with musculoskeletal problems: qualitative user research study. JMIR Hum Factors 2025 Jul 16;12:e65029. [doi: 10.2196/65029] [Medline: 40669045]
29. Creswell JW, Clark P, Vicki L. Designing and Conducting Mixed Methods Research, 3rd edition: SAGE Publications; 2018.
30. Levitt HM, Bamberg M, Creswell JW, Frost DM, Josselson R, Suárez-Orozco C. Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: the APA Publications and Communications Board task force report. Am Psychol 2018 Jan;73(1):26-46. [doi: 10.1037/amp0000151] [Medline: 29345485]
31. Witzel A, Reiter H. The Problem-Centred Interview: Principles and Practice: SAGE Publications; 2012. [doi: 10.4135/9781446288030]
32. Beyer H, Holtzblatt K. Contextual Design: Defining Customer-Centered Systems: Morgan Kaufmann Publishers; 1997.
33. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
34. Carroll JM. Scenario-based design. In: Making Use: Scenario-Based Design of Human-Computer Interactions: MIT Press; 2003:45-70.
35. VanGundy AB. Brain writing for new product ideas: an alternative to brainstorming. J Consum Mark 1984 Feb 1;1(2):67-74. [doi: 10.1108/eb008097]
36. Kothgassner OD, Felnhofer A, Hauk N, Kastenhofer E, Gomm J, Kryspin-Exner I. TUI: Technology Usage Inventory Manual. : FFG Forschungsförderungsgesellschaft; 2012 URL: https://www.ffg.at/sites/default/files/allgemeine_downloads/thematische%20programme/programmdokumente/tui_manual.pdf [accessed 2026-04-01]
37. Scherhorn G, Haas H, Hellenthal F, Seibold S. Sozialverträglichkeit. In: Zusammenstellung Sozialwissenschaftlicher Items Und Skalen (ZIS): ZIS - GESIS Leibniz Institute for the Social Sciences; 1999. [doi: 10.6102/ZIS196]
38. Bös K, Abel T, Woll A, Niemann S, Tittlbach S, Schott N. Der Fragebogen zur Erfassung des motorischen Funktionsstatus (FFB-Mot). Diagnostica 2002 Apr;48(2):101-111. [doi: 10.1026//0012-1924.48.2.101]
39. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34. [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
40. Nunnally JC. An overview of psychological measurement. In: Wolman BB, editor. Clinical Diagnosis of Mental Disorders: A Handbook: Springer; 1979:97-146. [doi: 10.1007/978-1-4684-2490-4_4]
41. Rhemtulla M, Brosseau-Liard P, Savalei V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. Psychol Methods 2012;17(3):354-373. [doi: 10.1037/a0029315]
42. Sass DA, Schmitt TA, Marsh HW. Evaluating model fit with ordered categorical data within a measurement invariance framework: a comparison of estimators. Struct Equ Modeling 2014 Apr 3;21(2):167-180. [doi: 10.1080/10705511.2014.882658]
43. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
44. Shi D, Lee T, Maydeu-Olivares A. Understanding the model size effect on SEM fit indices. Educ Psychol Meas 2019 Apr;79(2):310-334. [doi: 10.1177/0013164418783530]
45. Perotti L, Strutz N. Evaluation and intention to use the interactive robotic kitchen system AuRorA in older adults. Z Gerontol Geriat 2023 Nov;56(7):580-586. [doi: 10.1007/s00391-022-02105-8]
46. Fink F, Kalter I, Steindorff JV, Helmbold HK, Paulicke D, Jahn P. Identifying factors of user acceptance of a drone-based medication delivery: user-centered design approach. JMIR Hum Factors 2024;11:e51587. [doi: 10.2196/51587]
47. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021 Aug 17;21(1):1556. [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]

XSL•FO
RenderX

48. Kotval-K Z, Pithwa I. Affinity and usage of technology in general, and telehealth in particular amongst Michigan older adults during the COVID-19 pandemic. Healthcare (Basel) 2024 Jan 4;12(1):117. [doi: 10.3390/healthcare12010117] [Medline: 38201022]
49. Gell NM, Rosenberg DE, Demiris G, LaCroix AZ, Patel KV. Patterns of technology use among older adults with and without disabilities. Gerontologist 2015 Jun;55(3):412-421. [doi: 10.1093/geront/gnt166] [Medline: 24379019]
50. Granić A. Educational technology adoption: a systematic review. Educ Inf Technol 2022 Aug;27(7):9725-9744. [doi: 10.1007/s10639-022-10951-7]

Abbreviations

ADL: activities of daily living
ANG: anxiety
BEN: user-friendliness
CFA: confirmatory factor analysis
CFI: comparative fit index
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
COREQ: Consolidated Criteria for Reporting Qualitative Research
DWLS: diagonally weighted least squares
INT: interest
ITU: intention to use
JARS-MMARS: Mixed Methods Article Reporting Standards
NEU: curiosity (neugier)
NÜT: perceived usefulness
RMSEA: root mean square error of approximation
SEM: structural equation modeling
SHERS: smart home emergency response systems
SKE: skepticism
SOC: social compatibility
SRMR: standardized root mean square residual
TLI: Tucker-Lewis index
TUI: Technology Usage Inventory
VIF: variance inflation factor
ZUG: accessibility

Edited by A Kushniruk; submitted 06.Feb.2026; peer-reviewed by K Sames, S Boonlua; revised version received 10.Mar.2026; accepted 10.Mar.2026; published 20.Apr.2026.

Please cite as:
Pantförder M, Krüger O, Rietz A, Dzieia C, Meister S
User Acceptance of Smart Home Emergency Response Systems: Mixed Methods Study
JMIR Hum Factors 2026;13:e93003
URL: https://humanfactors.jmir.org/2026/1/e93003
doi:10.2196/93003

© Michael Pantförder, Oliver Krüger, Annika Rietz, Colin Dzieia, Sven Meister. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 20.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

AI-Enhanced Conversational Agents for Personalized Asthma Support in People With Asthma: Factors for Engagement, Value, and Efficacy in a Cross-Sectional Survey Study

Laura Moradbakhti[1], Dr Phil; Dorian Peters[2], PhD; Jennifer K Quint[1], PhD; Björn Schuller[3], Dr-Ing; Darren Cook[1], PhD; Rafael A Calvo[1], PhD

[1]Dyson School of Design Engineering, Faculty of Engineering, Imperial College London, South Kensington Campus, 25 Exhibition Road, London, United Kingdom

[2]Institute for Technology and Humanity, University of Cambridge, Cambridge, United Kingdom

[3]Department of Computing, Faculty of Engineering, Imperial College London, London, United Kingdom

Corresponding Author:
Laura Moradbakhti, Dr Phil
Dyson School of Design Engineering, Faculty of Engineering, Imperial College London, South Kensington Campus, 25 Exhibition Road, London, United Kingdom

Abstract

Background: Asthma-related deaths in the United Kingdom are the highest in Europe, and only 30% of patients access basic care. There is a need for alternative approaches to reaching people with asthma to provide health education, self-management support, and better bridges to care.

Objective: This study aimed to examine patients' interest in using a chatbot for asthma and to identify factors that influence engagement. Automated conversational agents (specifically, mobile chatbots) present opportunities for providing alternative and individually tailored access to health education, self-management support, and risk self-assessment. But would patients engage with a chatbot, and what factors influence engagement?

Methods: We present results from a patient survey (N=1257) developed by a team of asthma clinicians, patients, and technology developers, conducted to identify optimal factors for efficacy, value, and engagement with an asthma chatbot.

Results: Results indicate that most adults with asthma (53%) are interested in using a chatbot. The patients most likely to do so are those who believe their asthma is more serious and are less confident in their self-management. Results also indicate enthusiasm for 24/7 access, personalization, and for WhatsApp (Meta) as the preferred access method (compared to app, voice assistant, SMS text messaging, or website).

Conclusions: Obstacles to uptake include security and privacy concerns and skepticism of technological capabilities. We present detailed findings and consolidate these into 7 recommendations for developers to optimize the efficacy of chatbot-based health support.

(JMIR Hum Factors 2026;13:e80979) doi:10.2196/80979

KEYWORDS

asthma; chatbot; conversational agent; digital health; WhatsApp; artificial intelligence; AI; psychological needs; self-determination theory

Introduction

Background

Eight million people (12% of the population) in the United Kingdom are estimated to have a diagnosis of asthma [1]. Despite effective treatments and extensive efforts, the UK death rate from asthma is the highest in Europe. Moreover, 65% of people with asthma in the United Kingdom do not receive the professional care they are entitled to (eg, a yearly review, feedback on how to use their inhaler correctly, or an asthma action plan) [2]. In the latest Asthma + Lung UK survey [3], it was reported that only 30% of patients with asthma received basic care in 2021, the lowest proportion since 2015. Moreover, according to the same report, almost half (44%) of the patients who are admitted to the hospital did not receive follow-up care, with younger adults 18-29 years receiving the lowest level of basic asthma care [4].

One of the reasons for this is believed to be poor self-assessment of risk by those with asthma. For instance, according to Asthma + Lung UK, the illness is often not taken seriously enough, as reports show that 1 in 6 people in the United Kingdom do not know, or are not sure, if asthma can be fatal [5]. Moreover,

XSL•FO
RenderX

health education and self-management behaviors play a critical role.

In the United Kingdom, asthma-associated care costs are at least £1.1 billion per year [6], and the bulk of the burden centers on differences in asthma control rather than severity. People with severe uncontrolled asthma are prone to more symptoms, night-time awakenings, rescue medication use, and worse self-reported outcomes [7] and are estimated to represent 3 times the cost to the health care system compared with patients with severe but controlled disease [7]. Studies have shown that informed self-management that is accompanied by professional reviews improves asthma control, reduces exacerbations, and improves overall quality of life [8].

These statistics demonstrate an urgent need for alternative approaches to reaching people with asthma to (1) improve risk self-assessment, (2) improve health literacy and self-management, and (3) encourage them to seek traditional care as needed. One way to support people with asthma could be through a specialized conversational agent (CA). A CA can be any type of dialogue system that has natural language processing capabilities and can respond to input using human language. Moreover, the input and output are not restricted to text but could also happen via voice. Although CAs can be embodied avatars or physical robots, for this study, we refer only to CAs that are nonembodied chatbots used via apps such as WhatsApp (Meta).

New generative chatbot technologies, such as ChatGPT (OpenAI), are rapidly becoming pervasive, and people are likely to continue to consult them for health care advice despite safety risks [9-11]. Conversational interfaces based on large language models can be intuitive, engaging, and provide a level of personalization in their responses based on user input. However, these technologies rely on probabilistic models trained on information from across the internet to generate answers that have never been vetted. These systems are known to fabricate information and present it confidently, sometimes adding artificial citations to sources. Such “hallucinations” could be dangerous or even life-threatening when the system is responding to medical questions. As such, misinformation remains a significant problem for publicly available generative artificial intelligence (AI) services, rendering them unsafe for health applications. The development of a trustworthy conversational system must involve health experts, including patients and clinicians, and must exclusively provide health information that has been verified as safe and accurate for the specific context.

There is existing evidence that mobile technology can be used effectively for health interventions to help people learn about their illness, complete self-assessments, track symptoms, and be directed to appropriate care options. Several bespoke apps for asthma [12-15] and a smaller number of studies on CAs designed for children and adolescents have been reported in the literature [16,17]. However, research is missing that might guide an effective mobile CA for adults with asthma. While the technology itself holds promise, successful outcomes are entirely contingent on human acceptance and use, and research is lacking on what adult patients’ concerns, expectations, preferences, and motivations would make such a conversation-based health intervention valuable to them. Furthermore, how these needs might vary for the groups that most stand to benefit from alternative interventions (eg, patients with poor asthma control, with lower health literacy, or who do not access traditional health services) is not clear.

There are many open questions related to the potential for meaningful patient engagement with a conversation-based asthma intervention. For example, what would patients be looking for from conversational health support? What would be necessary for them to value and trust a technology enough to use it? How does existing technology use play a part? While CAs have a lot of potential to help people with asthma, they are a relatively new type of technology in the public sphere and, in addition to common concerns about privacy, people may mistrust a CA’s ability to help. Also, there is not enough understanding of what motivates or hinders patients’ use of CAs in general, or how socioeconomic variables may relate to lack of trust in, or engagement with, CAs.

The study aimed to provide answers to these questions and contribute to a better understanding of the features required to make a successful CA for asthma based on the expectations, needs, and motivations of adult patients with asthma. Additionally, attention is given to different groups of patients and how developers might most effectively meet the needs of those who stand to benefit the most from such digital services.

Related Work

This section describes previous work in the 3 intersecting areas relevant to this project: mobile technologies for health broadly and asthma specifically; conversational mobile technologies for health broadly and asthma specifically; and the use of WhatsApp as a platform for conversational mobile digital health.

Mobile Technologies for Asthma

Mobile technologies focusing on health care support have been proven to be cost-effective across health domains [2]. Specifically, there is evidence that mobile technology can be used effectively to help patients learn about their illness, complete self-assessments, track symptoms, and be directed to appropriate care options. According to a systematic review, 87% of studies reported improved adherence and 53% reported improved health outcomes for users [14]. Moreover, they are a cost-effective way to reach patients in a personalized and time-efficient manner, for example, by reminding patients to engage in self-management tasks (eg, medication adherence, electronically accessible action plans, and so on) [2]. With respect to asthma specifically, bespoke apps are the most common approach to mobile interventions for asthma. For example, several bespoke apps for young people [13,14] have demonstrated positive outcomes for improving self-management. For adults, the ASTHMAXcel mobile app was designed to improve asthma knowledge and outcomes [18]. The app was tested in multiple studies and revealed an increase in asthma knowledge [18,19] as well as a decrease in clinical outcomes such as emergency department visits, hospitalizations, and prednisone usage [20] among patients.

XSL•FO
RenderX

CAs for Health

Previous studies have shown that CAs can help patients with a range of health care tasks, such as delivering clinical and health risk information, providing elements of basic care, and supporting behavior change [21] through personalized dialogue. The use of CAs for health has been extensively reviewed [22-24].

With respect to asthma specifically, 3 studies explored the use of CAs for young people. Rhee et al [25] described an automated mobile phone–based asthma self-management aid for adolescents that is able to interpret conversational English SMS text messaging describing asthma symptoms. The system was tested over 2 weeks by 15 adolescent-parent dyads, and response rates for daily messages were 81% - 97% for the adolescents. Following the intervention, participants' awareness of both symptoms and triggers, their sense of control, and treatment adherence were higher, with benefits to the partnership between parents and adolescents.

Kadariya et al [16] tested kBot, a chatbot capable of interacting via both text and voice in the form of an Android app, designed to support pediatric patients with asthma (aged 8 - 15 years). Overall, kBot received good technology acceptance and system usability scores from clinicians (n=8) and researchers (n=8), but no patient testing was reported.

A study by Kowatsch et al [26] tested MAX, a CA that was designed to increase knowledge of asthma and behavioral skills, such as inhalation technique, among 10- to 15-year-olds with asthma. MAX incorporated different modes of communicating with the CA; for example, care professionals could communicate with it via email, patients via a mobile app, and family members via SMS text messaging. The results showed not only high acceptability of the CA by all stakeholders but also improved cognitive and behavioral skills. These studies with children and adolescents suggest promise for conversation-based digital health care. However, they are limited to tests with small groups of pediatric patients. Since late-onset asthma often remains undiagnosed [27] and these patients regularly experience more severe forms of asthma compared with people with early-onset asthma [27-29], it is important to assess the needs of adults and understand how they differ from those of pediatric patients, particularly with regard to health literacy and the self-management of symptoms.

Messaging platforms are widely used across age groups in the United Kingdom and elsewhere. According to Statista [30], 80% of people aged 16 - 64 years in the United Kingdom use WhatsApp, and 59.5% use Facebook Messenger (Meta). As such, the potential for conversation-based support for adults could be equally promising, but a better understanding of adult users and what might drive them to engage with conversational tools in real-world settings is needed, including how this may vary between different groups with respect to asthma severity, control, and other factors.

WhatsApp and Digital Health

Although CAs are often delivered as part of custom apps, they can also leverage existing and familiar chat-based tools such as SMS text messaging, WhatsApp, or web-based messaging systems. Indeed, we began our project with the intention of creating a bespoke app, but early engagement with patients and patient advocates led us to change this plan and shift to the widespread platform, WhatsApp. As mentioned in the "CAs for Health" section, most of the UK adult population already uses WhatsApp, which is probably why it was favored by our patient participants. Using a familiar and widespread platform also meant that we could remove the additional friction and technical barriers associated with bandwidth, storage capacity, usability, learnability, and concerns about trustworthiness that can come with downloading and using a new app for the first time.

Previous research has demonstrated positive results for the use of WhatsApp in health care settings. For example, an online questionnaire assessing patient satisfaction with WhatsApp for sharing health information, collected from 14 WhatsApp asthma and allergy groups, concluded that WhatsApp is an effective social media program for improving patient knowledge and behavior [31]. Furthermore, a study by Asthma UK offered young adults access to an asthma nurse who would assist them in managing their asthma via WhatsApp [32]. The service was well received, with 67.1% of participants reporting improved confidence in asthma management post usage.

Similarly, a study conducted across 5 Latin American countries confirmed an interest in WhatsApp as a tool to receive information about asthma. Irrespective of age (mean age of 43 years), 61.5% of patients in the study reported the highest interest in WhatsApp (as compared to SMS text messaging, Facebook, Twitter [X Corp], and email) as a tool for communication [33]. It is noteworthy that patients reported SMS text messaging as the most used service (69.9%) while still indicating that WhatsApp would be the most preferred tool for asthma communication. WhatsApp has several advantages over SMS text messaging, which may help explain its penetration. It is often more cost-effective than SMS text messaging because SMS text messaging is charged differently by mobile network providers than web-based messaging platforms such as WhatsApp; unlike SMS text messaging, the geographic reach of WhatsApp is global (ie, not limited to the mobile service region); and it provides better support for group communication, audio, video, and the option to retain an account even when one's phone number changes.

Despite the above evidence for the potential benefits of using WhatsApp, it is not as widely used as websites or custom apps. One reason (which applies to all messaging platforms) is that text- and voice-based conversations do not provide as many affordances (ie, visual displays and menu interfaces) as other media. Another reason, more specific to WhatsApp, involves privacy concerns. For example, to receive research ethics approval for a follow-up study, the research team carried out thorough discussions with both the ethics committee and the university's data protection office over the use of WhatsApp. Some of the concerns raised and how we recommend addressing them are listed below:

- Sensitive data: a health app could include sensitive data. Given these concerns, we recommend not asking users for

XSL•FO
RenderX

any sensitive information and asking only for the minimum data required to provide functionality.
- Data use by Meta: because WhatsApp is owned by Meta (formerly Facebook), there is no way of knowing how they will use data from chatbot conversations. To address these concerns, we recommend including an explanatory section specific to WhatsApp in the participant information sheet. It is also important to note that, as of September 2023, WhatsApp uses end-to-end encryption, meaning the conversation is readable only by the 2 communicating parties, in this case, the patient and our servers. This means that not even WhatsApp can read the messages. As such, the onus is on the entity hosting the chatbot servers to manage data securely and ethically rather than on Meta.
- Local storage vulnerabilities: local backups of conversations stored on a user's phone could be compromised if the phone is lost. To address these concerns, we recommend instructing users on how to turn off the backup if they wish.
- Management of identifiable data: to further address privacy concerns, we propose that it is not necessary to request any identifiable information and that only the WhatsApp phone number be used to identify a user within the system. Conversely, the user can be invited to give a nickname rather than their real name. For data analysis, data can be anonymized further (eg, phone numbers and nicknames can be removed).

Summary

The studies described in the "Introduction" section, including the asthma chatbot studies with children and adolescents, the survey studies exploring the potential use of WhatsApp in health care, and significant evidence for the efficacy of mobile interventions more broadly within health care, all provide promising evidence for a new conversation-based digital health care approach to complement primary and secondary care. These are summarized in Table 1 to clarify visually how existing work is distributed with respect to the technology used and whether it refers to health broadly or asthma specifically.

Table . Overview of categories under which the related work studies fall.

Category	For health (broadly)	For asthma (specifically)
Mobile technology (broadly; eg, apps, bespoke, and so on)	Ramsey et al [14] and Whittamore [2]	MAX (Kowatsch et al [26]) and ASTHMAXcel (Hsia et al [18,20])
WhatsApp (Meta; specific chat technology)	Saeed Tayeb [31]	Calderón et al [33] and Cumella et al [32]
Conversational agents (automated chatbots that can be used through chat technologies)	Singh et al [21]	kBot (Kadariya et al [16]) and Rhee et al [25]

What remains to be explored is a nuanced understanding of what adult users themselves would use and value, what might drive them to engage with conversational health tools in the real world, and how this might differ for diverse groups (eg, based on differences in disease severity, symptom control, confidence with technology, or trust in the health care system). For this study, we specifically sought this improved understanding for the population of adults with asthma within the United Kingdom, and we did so with a view to improving asthma risk self-assessment, asthma health literacy, and access to care. As such, our study aimed to answer the following research questions: (1) To what extent do different adults with asthma in the UK trust the health care system, and does this level of trust correlate with interest in using a chatbot for asthma? (2) Are those people who are more interested in using an asthma care chatbot more or less likely to already be accessing basic health care services (ie, general practitioner [GP])? (3) In what ways would adults with asthma prefer to access an asthma care chatbot? (4) What features, characteristics, and style preferences for a chatbot would maximize the likelihood of adults with asthma engaging meaningfully with it? (5) To what extent does a person's confidence using technology affect their willingness to use a chatbot for asthma? (6) Is there a relationship between an individual's interest in using an asthma care chatbot and their self-assessment of the seriousness of their asthma?

Methods

Overview

The survey study described herein is part of a larger overarching project funded by the UK Engineering and Physical Sciences Research Council (EPSRC) and conducted in partnership with Asthma + Lung UK, an asthma charity organization. A description of the overarching project is available in the protocol paper by Calvo et al [34].

Recruitment

Participants were recruited through YouGov [35]. To fulfill our recruitment criteria, we recruited only participants with self-reported asthma from the YouGov database. Although no medical proof of their asthma was requested, 1117 participants stated that they had received a diagnosis for their asthma. Participants were paid based on YouGov rates.

Survey

A survey comprising 38 closed- and 12 open-ended questions was conducted with adults living with asthma to gain a better understanding of how design and demographic features might influence interest in, and preferences for, a chatbot to support asthma management. A full list of the survey questions is provided in the Multimedia Appendix 1.

Analysis

The quantitative results were analyzed using SPSS (version 28; IBM Corp) [36] to run all statistical analyses of the quantitative

XSL•FO
RenderX

survey results. Firstly, descriptive statistics were analyzed to gain a better understanding of the sample and participants' general preferences for an asthma chatbot. Following the descriptive statistics, further analyses were conducted with a focus on users' trust in the health care system and to highlight differences between participants who were and were not interested in the chatbot. For all analyses, the significance level was set at α=.05.

Qualitative analysis of the open-ended free-text responses in the survey was conducted via a process of inductive thematic analysis following the steps outlined in Braun and Clarke [37] and using NVIVO (version 11 for Mac; QSR International) [38]. The aim was two-fold: (1) to identify more detailed insights underlying quantitative answers to the closed-ended questions and (2) to highlight patterns of responses or themes salient across participants' free-text contributions. In other words, text responses were coded iteratively, and themes generated were based on repeated patterns of ideas shared across participants. Resulting themes highlight some of the more common motivations and rationales for the quantitative results.

Ethical Considerations

The survey questions were approved by the Imperial College London ethics committee (reference no #21|C7403).

Results

Quantitative Results Overview

A total of 1257 people with self-reported asthma completed the survey (43% male, 57% female, and 4% left the gender question blank). In total, 1117 (88.9%) participants reported that they had received an asthma diagnosis. Of these, 47.8% of participants reported that their asthma was diagnosed younger than the age of 16 years, while 51.7% of participants reported that their asthma was diagnosed older than the age of 16 years, and 0.5% of participants preferred not to provide this information. Overall, 98.9% of participants received their diagnosis more than 6 months ago, 0.8% of participants preferred not to say when their diagnosis happened, and 0.3% of participants were diagnosed with asthma within the last 6 months at the time of participation in the study.

Most participants were older than 40 years; specifically, 44% were older than 55 years, 30% between 41 and 55 years, 19% were between 31 and 40 years, and 7% were between 18 and 30 years. Most participants had either completed a university degree (48%) or trade/technical/vocational training (19%), while 33% indicated that secondary school was their highest education level. For 0.4% of participants, the highest education level was completion of primary school. Fifty-six percent of participants spent most of their lives in a medium to small town, 23% in a rural area, and 21% in a large city.

Sixty-seven (5.3%) participants from our sample identified as being part of a minority ethnic group, while 1178 (93.7%) participants did not, and 12 (1%) participants preferred not to disclose. The number of participants who identified as part of a minority ethnic group was too small to allow for comparisons between participants who do and do not identify as such; however, some correlations with minority group identification and other variables are reported.

A summary of all key results per section is provided in Table 2. Table 3 summarizes the results that distinguish between the preferences of participants interested in an asthma chatbot and the preferences of participants who are not interested in an asthma chatbot for their asthma management.

XSL•FO
RenderX

Table . Summary of key results from each section.

Section	Results
Trust in health care system	• The majority of participants trust the health care system (83%). • People who are interested in the asthma chatbot have more trust in the health care system than those who are not interested. • Around 79.5% of participants seek support from a GP[a]. • Participants with an interest in the asthma chatbot seek more support from a GP than those who were not interested. • The more education participants have, the more likely they are to trust the health care system. • Participants who are more confident in their ability to use technology have a higher trust in the health care system. • Older participants trust the health care system more than younger participants. • Participants who are already confident in dealing with their own asthma trust the health care system more.
Communication preferences	• Around 59.7% of participants were interested in the voice feature. • Participants with a general interest in the asthma chatbot were also more interested in the voice feature of the asthma chatbot. • Around 85.5% of participants would prefer to talk to a human about their asthma. • Participants who were not interested in the asthma chatbot have a stronger preference for in-person conversations about their asthma.
Conversational style	• Most respondents preferred a “reassuring” and “friendly” conversational style and an asthma chatbot that felt “like talking to a nurse.” • However, a significant number of participants selected other preferences, for example, a “direct” approach or an experience that was more “like talking to a doctor.” • Overall, it can be concluded that different styles suit different people.
Technology experience	• The large majority of participants were confident in their use of technology (95%). • Participants who were less confident in their use of technology were less interested in trying a chatbot for asthma. • Participants who were confident in their use of technology, as well as participants who were interested in using an asthma chatbot, used WhatsApp (Meta) and other messenger apps significantly more often compared with participants who were less confident in their use of technology and those who were not interested in an asthma chatbot. • The majority of participants had used a virtual assistant in the past (89.3%). • Participants who had used a virtual assistant in the past were more likely to be interested in using a chatbot for their asthma.
Confidence in self-management	• Most participants felt either very confident (51.5%) or somewhat confident (45.6%) about managing their asthma, with only 2.9% of participants who did not feel confident. • Participants who were not interested in the asthma chatbot reported being more confident in managing their asthma. • The seriousness of their asthma was rated by participants, with 6.2% reporting it asvery serious, 31.1% reporting it as somewhat serious, and 62.7% reporting it as not serious. • Participants who were not interested in the asthma chatbot rated their asthma as less serious.

[a]GP: general practitioner.

XSL•FO
RenderX

Table . Summary of results that distinguish participants interested vs not Interested in an asthma chatbot.

Interested in an asthma chatbot	Not interested in an asthma chatbot
• More trust in the health care system • More likely to ask for support from a general practitioner (GP) than those • More likely to also be interested in a feature that detects asthma from the sound of the voice • Tend to rate their asthma as more serious	• Stronger preference for in-person conversations about their asthma • More confident in managing their asthma

Descriptive Results

YouGov shares some data they collect from all registered users. Existing YouGov data on WhatsApp use was available for 1011 participants and showed that 85% used WhatsApp. According to our own questions, 74% of participants used smartphone messaging apps daily, and 60% specifically used WhatsApp daily.

In contrast, 92.5% of participants indicated that they had never used a mobile app to help with their asthma (eg, to track their symptoms). Instead, participants were more likely to turn to existing familiar technologies (ie, internet search) or to trusted people (eg, nurses, doctors, or family members) for asthma information. The graph in Multimedia Appendix 2 shows which platform and/or source of information participants currently use to obtain information about asthma (multiple answer options; N=1257).

Participants were asked about which platform they would prefer to use to chat with an asthma chatbot, and they were allowed to select more than one. The majority indicated a preference for WhatsApp (refer to Multimedia Appendix 3), closely followed by a website.

Overall, the majority of participants (53%) indicated that they would indeed be interested in using an asthma chatbot via a messaging service such as WhatsApp (17% very interested and 36% somewhat interested).

An even higher number of participants (59%; 22% very interested and 37% somewhat interested) indicated interest in an asthma chatbot that could detect asthma severity from the sound of their voice, a technological feature currently in development by our research team. For this question, we asked participants: “Imagine AsthmaBot could detect your asthma severity by listening to the sound of your voice. You would speak into your phone to provide a short voice recording, and AsthmaBot would analyze this recording for signals (that only a computer can detect) to determine your asthma severity. How interested would you be in trying this voice feature?”

Trust in the Health Care System

To analyze participants’ trust toward the health care system, Spearman rank-order correlations were conducted. There was a significant negative correlation between participants’ education level and their trust in the health care system (coded as 1: participant trusts the system; and 2: participant does not trust the system; $r_s(1255)=-0.07$ and $P=.014$). That is, the higher a participant’s educational attainment, the more likely they were to trust the health care system.

Participants’ confidence in technology significantly positively correlated with their trust in the health care system ($r_s(1255)=0.190$ and $P<.001$). In other words, participants who were more confident in their ability to use technology also had higher trust in the health care system.

Participants’ trust in the health care system just fell short of a significant negative correlation with their identification as a minority ethnic group (or not; $r_s(1255)=-0.054$ and $P=.056$). Therefore, we can only report that there is a tendency for participants who do not identify as part of a minority ethnic group to trust the health care system more than people who do identify as an ethnic minority. This would be an important question to explore in a future study with a larger sample and would support previous research on experiences of discrimination and distrust in the health care system [39].

Age correlated significantly and negatively with trust in the health care system ($r_s(1255)=-0.07$ and $P=.009$), indicating that older participants were more likely to trust the health care system than younger ones.

Participants’ confidence in dealing with their asthma significantly correlated with participants’ trust in the health care system, $r_s(1255)=0.101$ and $P<.001$, showing that participants who were confident in managing their asthma trusted the health care system more. This may suggest that positive experiences with the health care system led to better management and, therefore, confidence in the ability to manage their asthma. However, there are many possible explanations for such a result: for example, people who trust the system less may not access health care and therefore may not get what they need to manage their asthma well; people who have had negative experiences with the system may not have gotten what they need to manage their asthma well despite accessing it; or people with asthma that is more difficult to treat and manage may more clearly see the limitations of the health care system.

Comparing Participants Interested vs Not Interested in an Asthma Chatbot

To allow a straightforward interpretation of the data, some of the variable answers were grouped. The example below (for the variable “How interested would you be in using AsthmaBot?”) shows how the Likert-scale answers were grouped into a binary format (Textbox 1).

XSL•FO
RenderX

Textbox 1. Categorical variable split between participants interested in the asthma chatbot and not interested in the asthma chatbot.

Interested

- Very interested
- Somewhat interested

Not interested

- Neutral
- Not particularly interested
- Not at all interested

There was no significant correlation between participants' interest in using an asthma chatbot and either their age, education level, or identification as a minority. This suggests that neither age, education level, nor identification with a minority ethnic group is likely to be a barrier to engagement.

To account for multiple comparisons, we applied the Bonferroni correction across all tests, which focus on analyzing factors that influence participants' interest in the asthma chatbot. This includes the 8 main chi-square comparisons as well as 1 Mann-Whitney *U* test. The adjusted significance level threshold for these tests is $P<.0056$. All tests answering different research questions were not included in this correction. The Spearman correlations were treated as exploratory analyses and are not included in this correction.

Health Care Usage Predicts Interest

We conducted a chi-square test for association to measure the relationship between participants' trust in the health care system and their interest in an asthma chatbot. The relationship is significant ($\chi^2_1=9.75$; $P=.002$; $\phi=.088$). Overall, the majority of participants trusted the health care system (83%), while 17% did not. Out of the participants who were not interested in an asthma chatbot, 20.1% indicated they did not trust the health care system, while 79.9% did. Of the participants who were interested in using an asthma chatbot, a higher percentage trusted the health care system (86.5%), and a lower percentage of participants did not trust the health care system (13.5%). This indicates that people who are interested in an asthma chatbot were more likely to trust the health care system than those who are not interested.

To check for a difference between participants' interest in an asthma chatbot and their current support from a GP, we conducted another chi-square test for association. The relationship is significant ($\chi^2_1=12.72$; $P<.001$; $\phi=.101$). Overall, 20.5% of participants do not seek support from a GP, while 79.5% do. Of the participants who were interested in the chatbot, 16.7% did not already receive support from a GP, while 83.3% did. For participants who were not interested in the chatbot, 24.8% did not get support from a GP, while 75.2% did. This suggests that overall, participants with an interest in the asthma chatbot were also more likely to be seeking support from a GP. This aligns with our results showing that people who see their asthma as more serious are more likely to be interested in a chatbot. Moreover, participants' perception of the seriousness of their asthma positively and significantly correlated with the Asthma Control Questionnaire items: difficulty sleeping due to asthma ($r_s(1255)=0.313$, $P<.001$); asthma symptoms during the day ($r_s(1255)=0.279$, $P<.001$); asthma affecting usual activities ($r_s(1255)=0.361$, $P<.001$); and asthma caused an accident and emergency (A&E) visit in the last 2 years ($r_s(1255)=0.275$, $P<.001$). Taken together, these findings suggest that at least some participants who were not interested in a chatbot lack interest because they did not feel their asthma was serious enough to motivate use. Qualitative results provide additional evidence for this conclusion (refer to section Reasons for Lack of Interest and Reservations).

Communication Preferences: Participants Would Value Asthma Voice Recognition but Still Favor Human Support

Regarding participants' interest in a feature that detects asthma severity through the sound of one's voice, voice recording could be more appealing (eg, novelty and perceived accuracy) or less appealing (eg, privacy concerns) than the standard text-based chatbot. Therefore, we analyzed the relationship between interest in such a feature and general interest in an asthma chatbot with a chi-square test for association. Again, both variables were coded as interested vs not interested. The relationship is significant ($\chi^2_1=347.100$; $P<.001$; $\phi=0.525$). Overall, 59.7% of participants were interested in the voice feature, while 40.3% were not. Out of the interested participants, 84.1% were interested in the voice feature, while 15.9% were not. Out of participants who were not interested in an asthma chatbot, only 32.4% were interested in the voice feature, and 67.6% were not. These differences show that most participants with a general interest in an asthma chatbot were also interested in a voice feature to detect severity, and that a subgroup of participants not interested in an asthma chatbot was, however, interested in voice-based severity detection for asthma.

A chi-square test for association was conducted to test the relationship between participants' preference for talking to a human about their asthma and their interest in an asthma chatbot. The relationship is significant ($\chi^2_1=17.05$; $P<.001$; $\phi=.116$). Overall, 85.5% of participants reported that they would prefer to talk to a human about their asthma. Of the participants interested in an asthma chatbot, 81.7% would prefer to talk to a human, while 18.3% would not. Of the participants who were not interested in an asthma chatbot, an even higher percentage preferred talking to a human (89.9%), while only 10.1% did not. The findings demonstrate that participants who were not

XSL•FO
RenderX

interested in using a chatbot for asthma, unsurprisingly, have a stronger preference for in-person conversations about their asthma. It is also worth noting that while the large majority would prefer to speak to a human about their asthma, a similar majority had sought asthma information via the internet. This suggests that although human interaction is preferred, when it is not available or easily accessible, most people turn to nonhuman options for support.

Conversational Style: Most Prefer a Chatbot That Is Friendly, Reassuring, and Like a Nurse

Research suggests that defining personality dimensions and a clear conversational style to guide the dialogue of a chatbot provides a more consistent and effective user experience [40]. Chatbot conversational style can vary across different personality dimensions; for example, it might be more or less formal or informal, authoritative, submissive, or convey warmth through word choice. To determine the kind of conversational style that most people with asthma would feel comfortable interacting with in a chatbot, the survey provided a list of personality descriptors and asked participants to select those they believe to be most appropriate for an asthma chatbot. Among participants interested in an asthma chatbot, the most frequently selected personality descriptors were “reassuring” (n=343), “like talking to a nurse” (n=335), “friendly” (n=316), “caring” (n=261), “good listener” (n=198), “like talking to a doctor” (n=196), “direct” (n=164), and “informal” (n=147). It is notable, however, that a significant number of participants selected other preferences, for example, favoring a “direct” approach or an experience more “like talking to a doctor,” suggesting different styles suit different people.

Technology Experience Predicts Interest

To assess the relationship between participants’ confidence in using technology and their interest in an asthma chatbot, we conducted a chi-square test for association. The relationship is significant (χ^2_1=13.73; P<.001; ϕ=0.105). Overall, the large majority of participants were confident in their use of technology (95%). Of the participants who were not interested in an asthma chatbot, 92.6% were confident in using technology, while 7.4% were not. Of the interested participants, a higher percentage were confident in technology (97.1%), while only 2.9% were not. In other words, and perhaps unsurprisingly, people who reported less confidence with technology use in general were also less interested in trying a chatbot for asthma.

Moreover, 2 Mann-Whitney U tests showed that participants who reported more confidence in their ability to use technology used WhatsApp and other messenger apps significantly more frequently in comparison to people who felt less confident with technology use (U=23652.50; Z=−5.636; P<.001) and that participants who were interested in using an asthma chatbot also used both WhatsApp and other messenger apps significantly more frequently in comparison to participants who were not interested in using a chatbot (U=175969.50; Z=−3.688; P<.001).

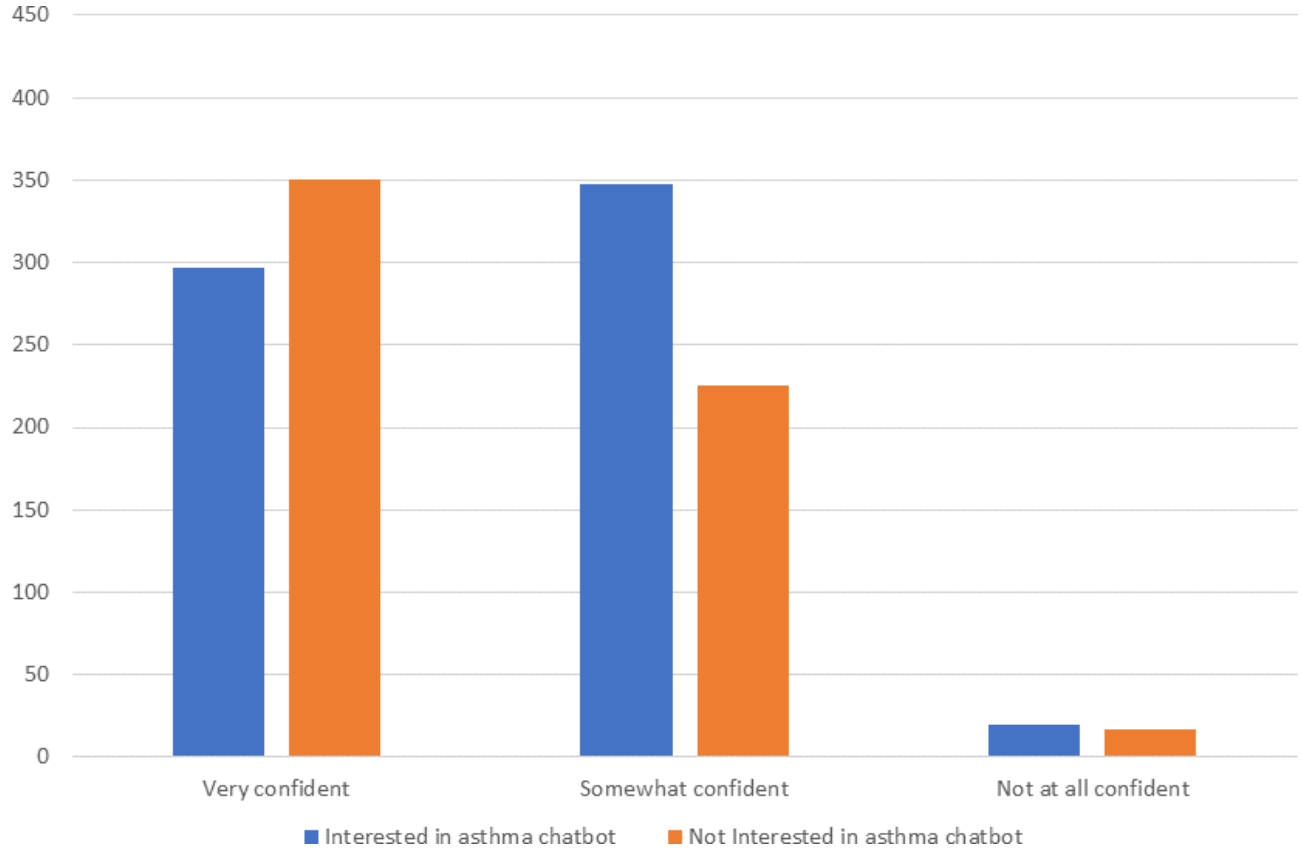
We also conducted a chi-square test for association to measure the relationship between participants’ previous use of a virtual assistant and participants’ interest in an asthma chatbot. The relationship is significant (χ^2_1=10.11; P=.001; ϕ=0.090). Overall, the majority of participants reported having used a virtual assistant in the past (89.3%). Of the participants who were not interested in an asthma chatbot, 86.3% had used a virtual assistant before, while 13.7% had not. Of participants who were interested, a higher percentage had used a virtual assistant previously (91.9%), while only 8.1% had not. As such, people who had used a virtual assistant in the past were more likely to be interested in using a chatbot for their asthma.

Lower Confidence in Disease Self-Management Predicts Interest

A chi-square test for association was conducted to assess the relationship between participants’ confidence in the self-management of their asthma and their interest in an asthma chatbot. The relationship is significant (χ^2_2=26.84; P<.001; ϕ=0.146). The majority of participants felt either very confident (51.5%) or somewhat confident (45.6%) about managing their asthma, with only 2.9% of participants reporting that they did not feel confident. Of those interested in the asthma chatbot, 44.7% were very confident, 52.3% were somewhat confident, and 3% were not confident at all. In comparison, out of the participants who were not interested in an asthma chatbot, a higher percentage indicated being very confident in managing their asthma (59.1%), 38% were somewhat confident, and only 2.9% were not confident. Therefore, participants who are less confident about their own ability to manage their asthma were more likely to be interested in an asthma chatbot. For an overview, refer to Figure 1. This provides support for the utility of a chatbot in improving management skills and confidence among those who feel they lack them.

XSL•FO
RenderX

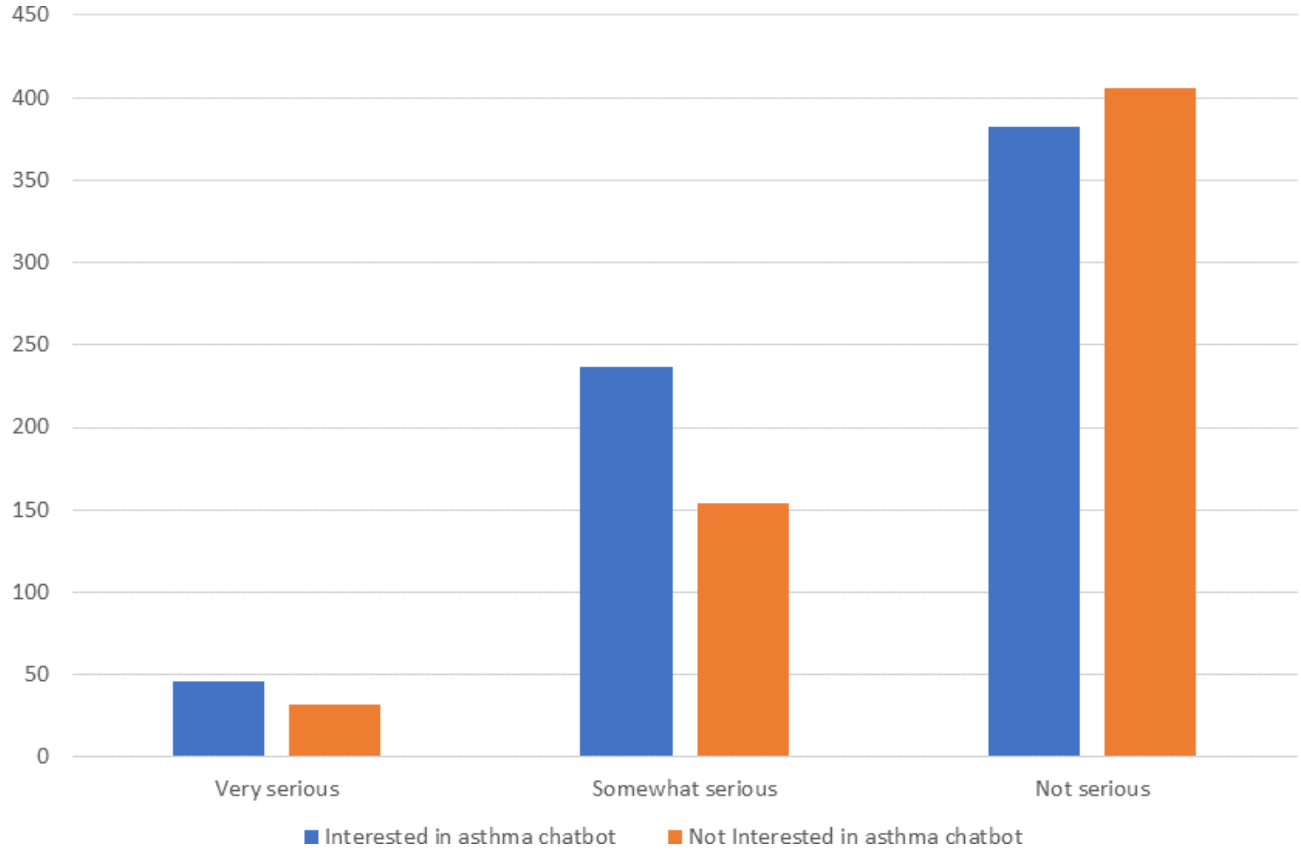
Figure 1. Bar chart showing the confidence level of participants dealing with their asthma, who are either Interested or not Interested in the asthma chatbot.



To assess the relationship between participants' own perception of the severity of their asthma and their interest in using an asthma chatbot, a chi-square test for association was conducted. The relationship is significant (χ^2_2=16.68; P<.001; ϕ=0.115). Asthma seriousness was rated by participants, with 6.2% rating their asthma as very serious, 31.1% as somewhat serious, and 62.7% as not serious. Of the participants who were interested in the asthma chatbot, 6.9% rated their asthma as very serious, 35.6% as somewhat serious, and 57.4% as not serious. Participants who were not interested rated their asthma as less serious overall, with 5.4% indicating that their asthma is very serious, 26% stating that their asthma is somewhat serious, and 68.6% rating their asthma as not serious. Overall, participants who were not interested in an asthma chatbot rated their asthma as less serious. For an overview, refer to Figure 2.

XSL•FO
RenderX

Figure 2. Bar chart showing the level of severity at which participants rate their asthma, who are either interested or not interested in the asthma chatbot.



Qualitative Results Overview

All of the 1257 survey respondents provided at least 1 free-text response to an open question. While quantitative analysis of closed survey questions reveals characteristics correlated with those most interested in an asthma chatbot, open-text results provide explicit insights into the reasons some participants are interested while others are not. In alignment with this "interest vs no interest" focus applied to the quantitative analysis, the qualitative analysis of open-question text was also primarily organized around these 2 categories, including reasons given for these and associated characteristics.

Open-text responses also provided additional information on specific technologies and information sources in relation to survey questions about where participants seek asthma information and what technologies they already use to support self-management.

Reasons for Lack of Interest and Reservations

Data were organized into a set of 4 recurring themes based on the types of reasons and reservations expressed in relation to lower interest in an asthma chatbot. These are provided in Table 4, along with example quotations from respondent comments.

XSL•FO
RenderX

Table . Summary of reasons for not being interested in an asthma chatbot.

Theme	Example quotations from the open comments
Theme 1: technology skepticism	• "I'm unsure it would get details correct and misdiagnose." [P21] • "They're not real, just programs. So they make mistakes and are not to be taken seriously." [P351] • "I dislike AI[a]." [P9] • "I would rather talk to my doctor or asthma nurse." [P109]
Theme 2: personalization skepticism	• "I'm not sure I would use it; these things tend to give general advice and if you need advice specifically for you then I'm not sure it could do that for everyone." [P231] • "I worry that it would tell me basic things that I already know or give generic advice. If it was very tailored to me that might help." [P163] • "My concern is these virtual assistants are programmed to answer specific questions whilst I may want to ask something that they are not equipped to answer!" [P455]
Theme 3: low asthma severity	• "I don't feel that my asthma is severe enough to need anything more than my annual check at the moment." [P367] • "If my asthma was worse than it is, I would find it more useful." [P855]
Theme 4: security and privacy concerns	• "The system gets hacked, and my data is used nefariously." [P35] • "I'd be uncomfortable with the idea of my medical details being stolen either by a hack or 'man in the middle' style attack." [P633]

[a]AI: artificial intelligence.

The first theme, "technology skepticism," captures expressions of doubt in the accuracy and trustworthiness of chatbot technology. For example, a number of participants expressed skepticism about the capacity of current technology to provide information that is accurate and safe (referring to the possibility of misdiagnoses and mistakes). Some participants expressed a general dislike for technological solutions ("I dislike AI" [P9]) or referenced the technology's nonhuman nature as an inherent problem ("They're not real…and are not to be taken seriously" [P352]), or relatedly, the simple preference for human solutions ("I would rather talk to my doctor or asthma nurse" [P109]).

The second recurring theme, "personalization skepticism," refers to responses expressing doubt in a technology's ability to respond in a personally relevant and specific way. For example, participant 163 stated, "I worry that it would tell me basic things that I already know or give generic advice. If it was very tailored to me that might help."

The third recurring reason for lack of interest in a chatbot was perceived low asthma severity, resulting in a perceived lack of need for technological support (eg, "I don't feel that my asthma is severe enough to need anything more than my annual check at the moment" [P367]).

Finally, a fourth reason for hesitancy in using an asthma chatbot was concern about system security and data privacy, including misuse by a malicious third party (eg, "The system gets hacked, and my data is used nefariously" [P35]). For more examples, refer to Table 4.

In addition to expressions of doubt, concern, or reservation with respect to using an asthma chatbot, many participants expressed optimism and interest in the potential for a chatbot to provide benefits. The recurring themes around these responses are summarized in Table 5. They include the capacity for a chatbot to provide education and self-management support, for example, by providing users with health information, warning of future asthma attacks, and helping manage symptoms; and in doing these things, providing an alternative to traditional health services (eg, "Support for my symptoms without risking a trip to the doctors" [P125] and "I have mild asthma and am concerned it may get worse, an asthma bot may help me control it because the National Health Service (NHS) website is next to useless" [P480]).

XSL•FO
RenderX

Table . Summary of reasons for interest in an asthma chatbot.

Theme	Example quotations from the open comments
Theme 1: health education and self-management support	• "Help control my asthma." [P120] • "Support for my symptoms without risking a trip to the doctors." [P125] • "To try and predict potential future asthma attacks in case I'm more at risk." [P240] • "It sounds easy and convenient and help manage my asthma worries easily." [P259] • "Learn new things about controlling asthma." [P337] • "It seems like a fun way to learn about the condition." [P453] • "I have mild asthma and am concerned it may get worse, an asthma chatbot may help me control it because the NHS[a] website is next to useless." [P480]
Theme 2: ease of access	• "It already knows the range of symptoms and can come up with something concrete very quickly, instead of me wasting someone's actual time." [P371] • "I need help now but cannot see my GP[b] for seven weeks." [P428] • "It is easier than getting appointment with nurse." [P449] • "I would use it rather than visit A&E[c] out of hours." [P485] • "An easy to access service." [P53] • "If it was free." [P107] • "Free to use and could be helpful before going to see a doctor." [P152] • "It seems like a good tool. It being free would be best." [P226]
Theme 3: trusted endorsement	• "That it is run by the NHS in conjunction with asthma UK not a commercial computer program." [P356] • "Endorsement by my GP practice." [P370]

[a]NHS: National Health Service.

[b]GP: general practitioner.

[c]A&E: accident and emergency.

Another set of responses captured various ways a chatbot provides ease of access to support, as compared to traditional health services ("I need help now but cannot see my GP for seven weeks" [P428]), with an emphasis by some participants on the importance of such a service being available for free ("If it was free" [P107] and "It being free would be best" [P226]).

Finally, some participants highlighted the importance of trusted endorsement by a trustworthy organization, such as their GP or the NHS, rather than a commercial company, as being important to their interest in using the technology (eg, "That it is run by the NHS in conjunction with Asthma UK, not a commercial computer program" [P356]).

Sources of Asthma Support: Sources and Apps

Finally, participants used open-text responses to provide the names of information sources and asthma technologies used. Specifically, 74 participants listed specific technologies they already use, or have used in the past, to get support for their asthma. Those best represented in the text responses included the NHS app (listed by 17 participants), Asthma UK (listed by 9), and Peak Flow (listed by 6).

Discussion

Key Findings and Recommendations

This study is the first to assess and contribute detailed insights on the motivational factors, features, and characteristics necessary for ensuring an asthma chatbot is of value to adults with asthma and that would maximize the likelihood of meaningful engagement. These findings are based on the closed and free-text responses of adults with asthma sharing views about their expectations, needs, and preferences, which were then correlated with differences in background, technology confidence, asthma history, and asthma self-management confidence. The goal is to contribute toward meaningful engagement with an easy-access asthma support technology to help address the care gap. Key findings that can most readily guide the design and development of future asthma chatbots are summarized herein.

Recommendation 1: Prioritize Ease of Access and Consider Existing Technologies Rather Than a Custom App

Overall, our results suggest that a significant percentage of UK adults with asthma (53% of our survey participants) would be interested in using a chatbot to support their asthma, and that an even higher number would be interested in using a voice-based risk detection feature (59% of our survey participants). Moreover, our findings provide a strong rationale for delivering the chatbot via WhatsApp instead of through a custom asthma app, since the large majority of our participants had already installed WhatsApp on their phone (85% of 1011 participants) and reported it to be their most common method of sending messages. Finally, WhatsApp was the number one

XSL•FO
RenderX

platform of preference for interacting with an asthma chatbot. Alternatively, a website was rated the second most popular channel for accessing an asthma chatbot. Although apps have distinct advantages that will sometimes warrant the extra friction, few participants had used an app for asthma, and their 2 top preferences for access leveraged familiar and easily accessible technology.

This aligns with qualitative findings in which participants highlighted ease of access as a major motivator for using an asthma chatbot. Ease of access had three main facets, including (1) that it is available 24/7, including when appointments are not available, (2) that it can (and should) be made available for free to ensure access, and (3) that it is easier to access than health care professionals. How we assessed ease of access aligns with 2 out of 4 key factors of the Unified Theory of Acceptance and Use of Technology model (UTAUT) [41], namely effort expectancy (the level of convenience and usability when using an information system) and facilitating conditions (the belief that the technical infrastructure to support the use of the system exists), which have been shown to positively influence behavioral intention and use behavior of information systems.

Recommendation 2: Center the Needs of People With More Serious Asthma Who Feel They Need Additional Support

Our findings indicate that an asthma chatbot is of particular interest to those who believe their asthma to be more serious. As such, a CA could be most beneficial for those patients who need additional support, either because their asthma is indeed more serious or because they experience more anxiety around it and less confidence in self-management. This is supported by the further results showing that participants who feel less confident in managing their asthma are also more likely to be interested in a chatbot than those who are already confident. Qualitative data confirms this and suggests that at least some people may also feel they do not always get adequate or timely access to traditional care (refer to Table 5). For example, participant open responses included “could be helpful before going to see a doctor” and “I need help now but cannot see my GP for seven weeks.” Thus, collectively, our results suggest there is a strong opportunity to integrate CAs as a supplementary, easily accessible, 24/7 service for patients who are in need of additional support.

Relatedly, participants who would be interested in using an asthma chatbot were also more likely to seek support from a GP in comparison to those who were not. This is consistent with the finding that participants interested in a chatbot have a greater need for support, as they rate their asthma as more serious.

In summary, our results show that an asthma chatbot is likely to be of greatest benefit to (1) patients who believe their asthma to be serious, (2) patients who feel less confident about managing their asthma, and (3) patients who are looking for immediate support while they wait for an appointment with a health care professional.

Recommendation 3: Include Conversational Content for Asthma Education, Risk Assessment, and Customized Self-Management Advice

Across both quantitative and qualitative responses, results demonstrate that participants would value easy access to educational information about asthma, support in assessing and preventing asthma attacks, and support for self-management—particularly tailored support that goes beyond the generic advice broadly available. The clinicians involved in this study add that there is great potential for a chatbot to have conversations about topics that are important to positive health outcomes but that there often is not enough time to cover during medical appointments. These topics include full explanations of what asthma is, what the medicine is for, how and why it works, proper technique, and conversations to support identifying and avoiding triggers. Previous studies have demonstrated that patient self-management not only allows patients to better manage long-term health conditions but also decreases visits to both GP offices and emergency departments [42,43]. A CA can improve health literacy, and this would not only fill in gaps but also provide repeated access to information over time, thus augmenting clinical practice.

Recommendation 4: Craft a Conversational Design That Is Reassuring and Like Talking to a Nurse

Our survey results with respect to people’s preferences for the conversational style (personality) of a chatbot revealed a set of majority preferences, with additional groups of alternate preferences. Most respondents preferred an asthma chatbot that was “reassuring,” “friendly,” and “like talking to a nurse.” It is notable, however, that a significant alternative group favored a “direct” approach or an experience that was more “like talking to a doctor.” Designers and developers could approach this diversity in a number of ways, including by (1) opting for the majority preference, (2) targeting a specific audience, or (3) allowing users to select personalities from a set of options. From a Self-Determination Theory perspective [44], providing options for personalization would support users’ need for autonomy. Fulfillment of all 3 basic psychological needs during technology interaction (autonomy, competence, and relatedness) predicts positive user experience and well-being [45-47], as well as higher intention to use a system [48,49].

Recommendation 5: Address Security and Privacy Concerns

The group of participants who reported more interest in an asthma chatbot also showed more overall trust in the health care system compared with participants who were not interested. This finding suggests that participants who are not interested in the chatbot might be less inclined overall to seek external support for their asthma. Future studies could investigate whether this is simply due to less serious asthma and effective self-management, or a more general lack of trust in external support (eg, negative experience, language barriers). The qualitative analysis provided some evidence for both explanations, indicating a combination of reasons is likely. Specifically, on the one hand, we identified a group of participants who were not interested in the chatbot and believed their asthma to be less severe (refer to theme 3 in Table 4). On

XSL•FO
RenderX

the other hand, we also identified a group who were skeptical of AI solutions and their accuracy in general.

Qualitative results revealed a number of themes that provide richer explanations for the quantitative findings. Specifically, that there were at least four categories of reasons given for lack of interest in an asthma chatbot: (1) technology skepticism (doubt in the technology's ability to be accurate or useful), (2) personalization skepticism (doubt in the technology's ability to provide anything other than very basic and generic advice, or to provide specific advice tailored to specific needs), (3) low asthma severity (perception that their asthma is not severe enough to warrant the use of technological support), and (4) security and privacy concerns (concerns over data misuse or hacking).

A successful chatbot would have to address these concerns to reach ambivalent users. These patient concerns demonstrate the importance of providing clear information and transparency around data security and the technology owner. Our results align with previous studies demonstrating that technology skepticism [50,51], as well as privacy concerns [52-54], can hinder the successful integration of new technologies in health care settings. Moreover, as the results show that previous exposure to a virtual assistant had a positive impact on patients' interest in the CA, these issues may improve over time as exposure increases [55-58].

Addressing security and privacy concerns would, at minimum, require very transparent disclosure of how data would be used and kept private (ie, not shared beyond the system), communicated in easy-to-understand ways. Granular consent options could be provided if providers wanted to use data for other purposes, such as research, and these would need to be opt-in. Sharing data with health care providers or systems (eg, NHS) could be made voluntary and fully under user control. Granular controls could also allow users to share only certain types of information and to withdraw consent [59]. To mitigate risks around data security, newer methods such as federated learning (decentralized model training without the distribution of raw patient data) or blockchain-based federated learning (an additional security layer for data sharing) present alternatives to traditional machine learning approaches and combine technological advances with adherence to legislative requirements (eg, HIPAA [Health Insurance Portability and Accountability Act] or GDPR [General Data Protection Regulation]) [60-62].

Recommendation 6: Address Technology Skepticism

While security and privacy concerns have to do with threats to personal autonomy, technology skepticism was evidenced by participant responses to nonbelief in the technology's capabilities, accuracy, and fitness to purpose. Some participants expressed aversion to using technology in favor of human interaction or general aversion to "AI." Others were doubtful that a chatbot system could provide advice that was tailored enough to the individual to be useful. Aiming to address these doubts, both technologically (making the doubts unfounded) and/or through messaging (communicating the genuine benefits and limitations) is likely to increase the audience for an asthma chatbot.

However, trust in these systems is likely to be a moving target. Given the rapid growth of generative AI chatbots such as ChatGPT (OpenAI), future studies could explore whether increased exposure to these types of chatbots impacts either positively or negatively on trust in related technologies. Although an asthma chatbot should only provide advice vetted by clinicians, generic generative AI chatbots such as ChatGPT and Claude (Anthropic) generate advice that is not safety-checked by a human and may therefore respond erroneously with so-called "hallucinations" or fabrications, which, in a health context, could be highly dangerous. As such, educating patients about the limitations and distinctions among various conversational technologies may become critical going forward. It is crucial to establish AI literacy among patients and health care providers to ensure the effective, ethical, and safe integration of emerging technologies [63,64]. In other words, communicating that there is a clear distinction between (1) general-purpose systems, (2) unproven apps that claim to be for health, and (3) specialized health systems with reliable third-party credentials will be part of the effort. However, there are no easy answers here, as trust must be deserved. To tackle these challenges, a multistakeholder approach should be implemented [53,65]. This involves a difficult, ongoing, and system-wide effort being addressed by academics, health care professionals, and technologists simultaneously as we attempt to find ways to balance the potential of new technologies with the clear risks of harm and challenges for regulation in the face of relentless technological change.

Recommendation 7: Consider a Trusted Endorsement

A final point on trust came from participants who highlighted the importance that a chatbot be endorsed by trusted authorities, such as their GP, an asthma charity, and not a commercial enterprise. For health care systems, trust is being repeatedly named as one of the key factors driving technology adoption even in comparison to the typical Unified Theory of Acceptance and Use of Technology (UTAUT) predictors: effort expectancy, performance expectancy, social influence, and ease of use [66-68]. This may be attributable to the fact that health care is a more sensitive and private topic. Consequently, it is imperative to engage trusted authorities in the development of digital health care solutions. Transparency about the inclusion of trusted authorities will provide users with the necessary trust to engage with novel CAs and other digital health care solutions.

By understanding people's doubts and the reservations that form obstacles to technology adoption, developers can target these concerns through technological improvements, interaction design, and education.

Limitations and Future Work

A limitation to the generalizability of the findings within this study is that it was conducted within the United Kingdom, and with the UK health care system in mind. Results with respect to access to primary care, trust in the health care system, or access to technologies, for example, are likely to vary across other countries. Another important consideration is that the survey was conducted online through an online survey service. Confidence in technology use was reported as very high within

XSL•FO
RenderX

our sample, but results could be different, for example, if a paper-based survey were conducted via community centers.

It is also relevant to highlight that our data rely on participants' self-reports of having asthma (rather than medical proof). To mitigate the risk of inaccurate self-reporting, we asked participants to specify when they were diagnosed by a doctor and to answer questions on symptoms and asthma control. We believe the added detail in questions, combined with the high number of participants who took part in the study, mitigates the potential effects of inaccurate self-reporting.

Another limitation involves our small sample of participants who identify with a minority ethnic group. We did not recruit enough participants from ethnic minority groups through the YouGov database to be able to explore research questions that could contribute to helping these groups specifically get their needs met. Our results are therefore not generalizable across ethnic groups, and additional research is needed to address the specific needs of ethnic minorities. We believe that this is an important research area for future studies, since there are health disparities demonstrated for minority ethnic communities, and easily accessible technologies, such as CAs, might contribute to addressing some of these. We would recommend using alternative methods of recruitment to reach participants from minority ethnic groups.

Relatedly, the majority of participants in our sample get support from a GP for their asthma, and trust in the health care system was high. Follow-up studies could focus specifically on patient groups who do not fall within either of these categories to gain a better understanding of how these groups (who are not accessing traditional care) might be better supported and how a chatbot might play a role.

Finally, in this study, participants did not have the chance to test an actual asthma chatbot. Instead, they were asked to make hypothetical judgments about their interests and preferences, so their responses were speculative and based on their previous experiences. Ideally, a longitudinal study should be conducted to test the long-term effects of, and adherence to, a fully built CA for asthma support. In a next follow-up study, we will test the feasibility of a prototype designed in alignment with the findings of this study. The follow-up study will provide insights into real-world patterns of use, perception of benefit, and final evaluations.

Conclusion

Overall, this study provides evidence for demographically broad interest in an asthma chatbot among adults with asthma in the United Kingdom. It is the first study to explore the potential of a CA to support adults with asthma, since previous research in this area focused on pediatric asthma instead. We were able to identify key factors that correlate with greater interest (ie, easy access, a fast alternative to GP or nurse, lower confidence in asthma self-management, and perception of asthma as serious) and less interest (lack of trust in AI, security concerns, preference for in-person contact, and higher confidence in self-management). These findings offer important insights for the design and development of a CA for asthma management and demonstrate the demand and opportunity in the United Kingdom for providing alternative support for asthma management. The results indicate that CAs can help to improve risk self-assessment, asthma control, and provide easy access to basic care outside of appointments with a health care professional.

Acknowledgments

We are grateful to Asthma + Lung UK and our patient advisory board participants for their constant support.

Funding

This project was funded by the UK Engineering and Physical Sciences Research Council (grant no EP/W002477/1). This research was supported by the National Institute for Health and Care Research (NIHR) Imperial Biomedical Research Centre (BRC).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full list of survey questions used in the study.
[PDF File, 111 KB - humanfactors_v13i1e80979_app1.pdf]

Multimedia Appendix 2
Overview of channels participants currently use to receive information about asthma.
[PNG File, 43 KB - humanfactors_v13i1e80979_app2.png]

Multimedia Appendix 3
Overview of participants' preferred platform to use the asthma chatbot.
[PNG File, 50 KB - humanfactors_v13i1e80979_app3.png]

XSL•FO
RenderX

References

1. What is the prevalence of asthma? NICE. 2022. URL: https://www.nice.org.uk/cks-uk-only#:~:text=In%20the%20UK%2C%20over%208,people%20are%20receiving%20asthma%20treatment [accessed 2023-06-22]
2. Whittamore A. Technology can revolutionise supported self-management of asthma. : Primary Care Respiratory Society UK; 2017 URL: https://www.pcrs-uk.org/sites/default/files/SupportedSelfCare_SupportGroupsPerspective.pdf [accessed 2026-03-03]
3. Levelling up lung health how improving lung health will achieve levelling up by delivering a healthier, fairer society. : Asthma + Lung UK; 2022 URL: https://www.asthmaandlung.org.uk/sites/default/files/Levelling%20up%20lung%20health_what%20needs%20to%20be%20done.pdf
4. On the edge: how inequality affects people with asthma. : Asthma UK; 2018 URL: https://www.asthmaandlung.org.uk/sites/default/files/2023-03/auk-health-inequalities-final.pdf
5. Lung conditions kill more people in the UK than anywhere in Western Europe. Asthma + Lung UK. URL: https://www.asthmaandlung.org.uk/media/press-releases/lung-conditions-kill-more-people-uk-anywhere-western-europe#:~:text=ue%20your%20breath-,Lung%20conditions%20kill%20more%20people%20in%20the%20UK%20than%20anywhere,3 [accessed 2023-01-19]
6. Mukherjee M, Stoddart A, Gupta RP, et al. The epidemiology, healthcare and societal burden and costs of asthma in the UK and its member nations: analyses of standalone and linked national databases. BMC Med 2016 Aug 29;14(1):113. [doi: 10.1186/s12916-016-0657-8] [Medline: 27568881]
7. Chen S, Golam S, Myers J, Bly C, Smolen H, Xu X. Systematic literature review of the clinical, humanistic, and economic burden associated with asthma uncontrolled by GINA Steps 4 or 5 treatment. Curr Med Res Opin 2018 Dec;34(12):2075-2088. [doi: 10.1080/03007995.2018.1505352] [Medline: 30047292]
8. Pinnock H. Supported self-management for asthma. Breathe (Sheff) 2015 Jun;11(2):98-109. [doi: 10.1183/20734735.015614] [Medline: 26306110]
9. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell 2023;6:1169595. [doi: 10.3389/frai.2023.1169595] [Medline: 37215063]
10. Li J, Dada A, Kleesiek J, Egger J. ChatGPT in healthcare: a taxonomy and systematic review. MedRXiv. Preprint posted online on Mar, 2023. [doi: 10.1101/2023.03.30.23287899]
11. Vaishya R, Misra A, Vaish A. ChatGPT: Is this version good for healthcare and research? Diabetes & Metabolic Syndrome: Clinical Research & Reviews 2023 Apr;17(4):102744. [doi: 10.1016/j.dsx.2023.102744]
12. Jaimini U, Thirunarayan K, Kalra M, Venkataraman R, Kadariya D, Sheth A. "How is my child's asthma?" Digital phenotype and actionable insights for pediatric asthma. JMIR Pediatr Parent 2018;1(2):e11988. [doi: 10.2196/11988] [Medline: 31008446]
13. Kagen S, Garland A. Asthma and allergy mobile apps in 2018. Curr Allergy Asthma Rep 2019 Feb 2;19(1):6. [doi: 10.1007/s11882-019-0840-z] [Medline: 30712150]
14. Ramsey RR, Caromody JK, Voorhees SE, et al. A systematic evaluation of asthma management apps examining behavior change techniques. J Allergy Clin Immunol Pract 2019;7(8):2583-2591. [doi: 10.1016/j.jaip.2019.03.041] [Medline: 30954644]
15. Venkataramanan R, Thirunarayan K, Jaimini U, et al. Determination of personalized asthma triggers from multimodal sensing and a mobile app: observational study. JMIR Pediatr Parent 2019 Jun 27;2(1):e14300. [doi: 10.2196/14300] [Medline: 31518318]
16. Kadariya D, Venkataramanan R, Yip HY, Kalra M, Thirunarayanan K, Sheth A. KBot: knowledge-enabled personalized chatbot for asthma self-management. : IEEE; 2019 Presented at: 2019 IEEE International Conference on Smart Computing (SMARTCOMP) p. 138-143. [doi: 10.1109/SMARTCOMP.2019.00043]
17. Suehs CM, Vachier I, Galeazzi D, et al. Standard patient training versus Vik-Asthme chatbot-guided training: "AsthmaTrain" - a protocol for a randomised controlled trial for patients with asthma. BMJ Open 2023 Feb 21;13(2):e067039. [doi: 10.1136/bmjopen-2022-067039] [Medline: 36810168]
18. Hsia B, Mowrey W, Keskin T, et al. Developing and pilot testing ASTHMAXcel, a mobile app for adults with asthma. J Asthma 2021 Jun;58(6):834-847. [doi: 10.1080/02770903.2020.1728770] [Medline: 32046564]
19. Chan A, Kodali S, Lee GY, et al. Evaluating the effect and user satisfaction of an adapted and translated mobile health application ASTHMAXcel© among adults with asthma in Pune, India. J Asthma 2023 Aug;60(8):1513-1523. [doi: 10.1080/02770903.2022.2155188] [Medline: 36511602]
20. Hsia BC, Wu S, Mowrey WB, Jariwala SP. Evaluating the ASTHMAXcel mobile application regarding asthma knowledge and clinical outcomes. Respir Care 2020 Aug;65(8):1112-1119. [doi: 10.4187/respcare.07550] [Medline: 32487751]
21. Singh B, Olds T, Brinsley J, et al. Systematic review and meta-analysis of the effectiveness of chatbots on lifestyle behaviours. NPJ Digit Med 2023 Jun 23;6(1):118. [doi: 10.1038/s41746-023-00856-1] [Medline: 37353578]
22. Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 1;25(9):1248-1258. [doi: 10.1093/jamia/ocy072] [Medline: 30010941]

XSL•FO
RenderX

23. Parmar P, Ryu J, Pandya S, Sedoc J, Agarwal S. Health-focused conversational agents in person-centered care: a review of apps. NPJ Digit Med 2022 Feb 17;5(1):21. [doi: 10.1038/s41746-022-00560-6] [Medline: 35177772]
24. Sundareswaran V, Sarkar A. Chatbots RESET: a framework for governing responsible use of conversational AI in healthcare. World Economic Forum. 2020. URL: https://www3.weforum.org/docs/WEF_Governance_of_Chatbots_in_Healthcare_2020.pdf [accessed 2023-01-19]
25. Rhee H, Allen J, Mammen J, Swift M. Mobile phone-based asthma self-management aid for adolescents (mASMAA): a feasibility study. Patient Prefer Adherence 2014;8:63-72. [doi: 10.2147/PPA.S53504] [Medline: 24470755]
26. Kowatsch T, Schachner T, Harperink S, et al. Conversational agents as mediating social actors in chronic disease management involving health care professionals, patients, and family members: multisite single-arm feasibility study. J Med Internet Res 2021 Feb 17;23(2):e25060. [doi: 10.2196/25060] [Medline: 33484114]
27. Li J, Ye L, She J, Song Y. Clinical differences between early- and late-onset asthma: a population-based cross-sectional study. Can Respir J 2021;2021:8886520. [doi: 10.1155/2021/8886520] [Medline: 33574971]
28. de Nijs SB, Venekamp LN, Bel EH. Adult-onset asthma: is it really different? Eur Respir Rev 2013 Mar 1;22(127):44-52. [doi: 10.1183/09059180.00007112] [Medline: 23457164]
29. Wu TJ, Chen BY, Lee YL, Hsiue TR, Wu CF, Guo YL. Different severity and severity predictors in early-onset and late-onset asthma: a Taiwanese population-based study. Respiration 2015;90(5):384-392. [doi: 10.1159/000439310] [Medline: 26418355]
30. Alexander K. Most used messenger by brand in the UK as of september 2025. Statista. 2023. URL: https://www.statista.com/forecasts/997945/most-used-messenger-by-brand-in-the-uk [accessed 2026-03-03]
31. Saeed Tayeb MM. Whatsapp education for asthma and allergy patients: pros and cons. GJO 2019;18(5) [FREE Full text] [doi: 10.19080/GJO.2019.18.555999]
32. Cumella A, King J, Sinton H, Walker S. Engaging younger people in their asthma management – findings from a pilot telemedicine asthma nurse service. 2020 Sep 7 Presented at: ERS International Congress 2020 abstracts p. 4805. [doi: 10.1183/13993003.congress-2020.4805]
33. Calderón J, Cherrez A, Ramón GD, et al. Information and communication technology use in asthmatic patients: a cross-sectional study in Latin America. ERJ Open Res 2017 Jul;3(3):00005-02017. [doi: 10.1183/23120541.00005-2017] [Medline: 28717641]
34. Calvo RA, Peters D, Moradbakhti L, et al. Assessing the feasibility of a text-based conversational agent for asthma support: protocol for a mixed methods observational study. JMIR Res Protoc 2023 Feb 2;12:e42965. [doi: 10.2196/42965] [Medline: 36729586]
35. YouGov PLC. URL: https://yougov.com/ [accessed 2026-03-10]
36. IBM SPSS Statistics for Windows, version 25.0. IMB Corp. 2017. URL: https://www.ibm.com/products/spss?utm_content=SRCWW&p1=Search&p4=346823877772&p5=p&p9=88618557002&gclsrc=aw.ds&gad_source=1&gad_campaignid=6537010&gbraid=0AAAAD-QsTPB7ZjuFQkBFkgmVJT8gdHFAhQhChMIdLo2YKWwMVHBx217SDCEAAYASAAEgLnD_BwE [accessed 2026-03-10]
37. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
38. Ask more from your data with the depth and power of NVivo. Lumivero. 2020. URL: https://lumivero.com/products/nvivo/
39. Paul E, Fancourt D, Razai M. Racial discrimination, low trust in the health system and COVID-19 vaccine uptake: a longitudinal observational study of 633 UK adults from ethnic minority groups. J R Soc Med 2022 Nov;115(11):439-447. [doi: 10.1177/01410768221095241] [Medline: 35512716]
40. Callejas Z, López-Cózar R, Ábalos N, Griol D. Affective conversational agents: the role of personality and emotion in spoken interactions. In: Conversational Agents and Natural Language Interaction: Techniques and Effective Practices: IGI Global; 2011:203-222. [doi: 10.4018/978-1-60960-617-6.ch009]
41. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. Manag Inf Syst Q 2003 Sep 1;27(3):425-478. [doi: 10.2307/30036540]
42. Barker I, Steventon A, Deeny S. Patient activation is associated with fewer visits to both general practice and emergency departments: a cross-sectional study of patients with long-term conditions. Clin Med 2017;17(Suppl 3):s15. [doi: 10.7861/clinmedicine.17-3s-s15]
43. Deeny S, Thorlby R, Steventon A. Briefing: reducing emergency admissions: unlocking the potential of people to better manage their long-term conditions. : The Health Foundation; 2018 URL: https://www.health.org.uk/sites/default/files/Reducing-Emergency-Admissions-long-term-conditions-briefing.pdf [accessed 2026-03-03]
44. Deci EL, Ryan RM. The “what” and “why” of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 2000 Oct;11(4):227-268. [doi: 10.1207/S15327965PLI1104_01]
45. Peters D, Calvo RA, Ryan RM. Designing for motivation, engagement and wellbeing in digital experience. Front Psychol 2018;9:797. [doi: 10.3389/fpsyg.2018.00797] [Medline: 29892246]
46. Peters D. Wellbeing supportive design – research-based guidelines for supporting psychological wellbeing in user experience. Int J Hum Comput Interact 2023 Aug 27;39(14):2965-2977. [doi: 10.1080/10447318.2022.2089812]

XSL•FO
RenderX

47. Ryan RM, Deci EL. Self-determination theory: basic psychological needs in motivation, development, and wellness. : Guilford Publications; 2017 URL: https://www.google.com/books/edition/Self_Determination_Theory/th5rDwAAQBAJ?hl=en&gbpv=1&dq=inauthor:%22Richard+M.+Ryan%22 [accessed 2026-03-04] [doi: 10.1521/978.14625/28806]
48. Moradbakhti L, Leichtmann B, Mara M. Development and validation of a basic psychological needs scale for technology use. Psychol Test Adapt Dev 2024 Jan 1;5(1):26-45. [doi: 10.1027/2698-1866/a000062]
49. Moradbakhti L, Schreibelmayr S, Mara M. Do men have no need for "feminist" artificial intelligence? Agentic and gendered voice assistants in the light of basic psychological needs. Front Psychol 2022;13:855091. [doi: 10.3389/fpsyg.2022.855091] [Medline: 35774945]
50. Offermann J, Rohowsky A, Ziefle M. Emotions of scepticism, trust, and security within the acceptance of telemedical applications. Int J Med Inform 2023 Sep;177:105116. [doi: 10.1016/j.ijmedinf.2023.105116] [Medline: 37307720]
51. Sobaih AEE, Chaibi A, Brini R, Abdelghani Ibrahim TM. Unlocking patient resistance to AI in healthcare: a psychological exploration. Eur J Investig Health Psychol Educ 2025 Jan 8;15(1):6. [doi: 10.3390/ejihpe15010006] [Medline: 39852189]
52. Alhammad N, Alajlani M, Abd-Alrazaq A, Epiphaniou G, Arvanitis T. Patients' perspectives on the data confidentiality, privacy, and security of mHealth apps: systematic review. J Med Internet Res 2024 May 31;26:e50715. [doi: 10.2196/50715] [Medline: 38820572]
53. Gondi DS, Bandaru VKR, Sathish K, Kanthi Kumar K, Ramakrishnaiah N, Bhutani M. Balancing innovation with patient privacy amid ethical challenges of AI in health care. In: International Conference on Innovative Computing and Communication: Springer Nature; 2025:555-576.
54. Richardson JP, Smith C, Curtis S, et al. Patient apprehensions about the use of artificial intelligence in healthcare. NPJ Digit Med 2021 Sep 21;4(1):140. [doi: 10.1038/s41746-021-00509-1] [Medline: 34548621]
55. Bartneck C, Suzuki T, Kanda T, Nomura T. The influence of people's culture and prior experiences with Aibo on their attitude towards robots. AI & Soc 2006 Nov 3;21(1-2):217-230. [doi: 10.1007/s00146-006-0052-7]
56. Glikson E, Woolley AW. Human trust in artificial intelligence: review of empirical research. Ann R Coll Physicians Surg Can 2020 Jul;14(2):627-660. [doi: 10.5465/annals.2018.0057]
57. Rossi A, Holthaus P, Dautenhahn K, Koay KL, Walters ML. Getting to know pepper: effects of people's awareness of a robot's capabilities on their trust in the robot. 2018 Presented at: Proceedings of the 6th international conference on human-agent interaction p. 246-252. [doi: 10.1145/3284432.3284464]
58. Waytz A, Heafner J, Epley N. The mind in the machine: anthropomorphism increases trust in an autonomous vehicle. J Exp Soc Psychol 2014 May;52:113-117. [doi: 10.1016/j.jesp.2014.01.005]
59. Narkhede MR, Wankhede NI, Kamble AM. Enhancing patient autonomy in data ownership: privacy models and consent frameworks for healthcare. JDH 2025:1-23. [doi: 10.55976/jdh.4202513361-23]
60. Abbas SR, Abbas Z, Zahir A, Lee SW. Federated learning in smart healthcare: a comprehensive review on privacy, security, and predictive analytics with IoT integration. Healthcare (Basel) 2024 Dec 22;12(24):2587. [doi: 10.3390/healthcare12242587] [Medline: 39766014]
61. Ullah Tariq U, Sabrina F, Mamunur Rashid M, et al. Blockchain-based secured data sharing in healthcare: a systematic literature review. IEEE Access 2025;13:45415-45435. [doi: 10.1109/ACCESS.2025.3547953]
62. Xu G, Zhou Z, Dong J, Zhang L, Song X. A blockchain-based federated learning scheme for data sharing in industrial internet of things. IEEE Internet Things J 2023;10(24):21467-21478. [doi: 10.1109/JIOT.2023.3298196]
63. Ang CS. Developing AI literacy in healthcare education: bridging the gap in competency assessment. Discov Educ 2025;4(1):372. [doi: 10.1007/s44217-025-00812-z]
64. Liu Z, Zou W, Lin C. Exploring the influence of privacy concerns, AI literacy, and perceived health stigma on AI chatbot use in healthcare: an uncertainty reduction approach. Patient Educ Couns 2025 Nov;140:109271. [doi: 10.1016/j.pec.2025.109271] [Medline: 40712403]
65. Williamson SM, Prybutok V. Balancing privacy and progress: a review of privacy challenges, systemic oversight, and patient perceptions in AI-driven healthcare. Appl Sci (Basel) 2024;14(2):675. [doi: 10.3390/app14020675]
66. AlQudah AA, Al-Emran M, Shaalan K. Technology acceptance in healthcare: a systematic review. Appl Sci (Basel) 2021;11(22):10537. [doi: 10.3390/app112210537]
67. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172. [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
68. Tung FC, Chang SC, Chou CM. An extension of trust and TAM model with IDT in the adoption of the electronic logistics information system in HIS in the medical industry. Int J Med Inform 2008 May;77(5):324-335. [doi: 10.1016/j.ijmedinf.2007.06.006] [Medline: 17644029]

Abbreviations

AI: artificial intelligence
CA: conversational agent
EPSRC: Engineering and Physical Sciences Research Council

XSL•FO
RenderX

GDPR: General Data Protection Regulation
GP: general practitioner
HIPAA: Health Insurance Portability and Accountability Act
NHS: National Health Service
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by A Bamgboje-Ayodele, A Kushniruk; submitted 26.Jul.2025; peer-reviewed by SW Lee, SM Shaffi; revised version received 22.Dec.2025; accepted 22.Jan.2026; published 11.Mar.2026.

Please cite as:
Moradbakhti L, Peters D, Quint JK, Schuller B, Cook D, Calvo RA
AI-Enhanced Conversational Agents for Personalized Asthma Support in People With Asthma: Factors for Engagement, Value, and Efficacy in a Cross-Sectional Survey Study
JMIR Hum Factors 2026;13:e80979
URL: https://humanfactors.jmir.org/2026/1/e80979
doi:10.2196/80979

© Laura Moradbakhti, Dorian Peters, Jennifer K Quint, Björn Schuller, Darren Cook, Rafael A Calvo. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 11.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

A User-Centered Interface Design Framework for the DELONELINESS System in Older Adults: Design Indicator Development and Prioritization

Yi Zhou[1], PhD; Jessica Rees[2], PhD; Faith Matcham[3], PhD; Michela Antonelli[4], PhD; Sebastien Ourselin[4], Prof Dr; Wei Liu[1], Prof Dr

[1]Department of Engineering, King's College London, Strand Campus, Strand, London, United Kingdom

[2]Department of Global Health and Social Medicine, King's College London, London, United Kingdom

[3]School of Psychology, University of Sussex, Brighton, United Kingdom

[4]School of Biomedical Engineering & Imaging Sciences, King's College London, London, United Kingdom

Corresponding Author:
Wei Liu, Prof Dr
Department of Engineering, King's College London, Strand Campus, Strand, London, United Kingdom

Abstract

Background: Loneliness among older adults has become a major public health concern associated with cognitive decline, depression, and increased health care use. The advancement of digital health technologies such as wearable devices, smart home systems, and mobile health apps provides new opportunities to monitor and mitigate loneliness through continuous physiological and behavioral assessment. However, the effectiveness of such technologies largely depends on user interface design, ensuring that older adults can understand, trust, and comfortably engage with the technology. Existing research on interactive platforms and user interfaces for psychological and emotional monitoring has mainly focused on usability testing and technology feasibility, with limited attention to structured design frameworks that integrate psychological, emotional, and accessibility dimensions for older adults. Furthermore, most studies rely on qualitative assessments and lack quantitative prioritization of design indicators.

Objective: This study aimed to build on the DELONELINESS (Design for Healthy Ageing: a Smart System to Decrease Loneliness for Older People) system to develop a user-centered hierarchical framework of interface design indicators for older adults.

Methods: A mixed methods design was applied, integrating literature search, qualitative focus group analysis, and expert consultation to build an initial indicator pool. A hierarchical indicator structure with 7 first-level and 26 second-level indicators was developed. The analytic hierarchy process was used to assign indicator weights through surveys of 20 experts with academic or professional experience in human-computer interaction, digital health, gerontology, and health informatics. Based on the weighted results, 3 interface design solutions were developed and comparatively evaluated using the Technique for Order Preference by Similarity to Ideal Solution.

Results: All expert judgment matrices satisfied the analytic hierarchy process consistency requirement (consistency ratio <0.1). The Kendall coefficient of concordance indicated good agreement among experts for both first-level indicators (W=0.313; *P*<.001) and second-level indicators (W=0.156; *P*<.001). Among the 7 first-level indicators, trust and safety (weight=0.206), ease of use (weight=0.187), and accessibility (weight=0.167) received the highest weights, indicating their importance in enhancing user confidence and engagement. Technique for Order Preference by Similarity to Ideal Solution evaluation results showed that design solution 2 achieved the highest overall performance score (relative closeness coefficient C=0.877), emphasizing clear interaction pathways, visual clarity, and guided feedback as key factors for optimal usability.

Conclusions: This study developed a user-centered framework for interface design in loneliness monitoring among older adults by integrating user insights, literature-derived indicators, and expert consensus, and providing a structured data-driven approach to prioritizing design requirements. The proposed framework bridges subjective user experience with objective evaluation, offering practical guidance for developing empathetic, inclusive, and trustworthy digital mental health technologies for older adults.

(JMIR Hum Factors 2026;13:e88263) doi:10.2196/88263

KEYWORDS

older adults; loneliness; digital health; user-centered design; human-computer interaction; interface design

XSL•FO
RenderX

Introduction

With the rapid growth of the global aging population, loneliness has become a major public health concern in aging societies. It is estimated that up to 1 in 4 older adults worldwide currently experience social isolation [1-3]. Research has demonstrated that older adults are particularly vulnerable to loneliness and social isolation. As people age, they may experience loneliness and social isolation due to reduced social interactions caused by various factors, including physical and cognitive changes and losses, living alone, or the loss of family or friends [4-6]. Prolonged loneliness or social isolation has been linked to a higher risk of depression, dementia, physical decline, chronic illness, and even premature mortality [7,8]. Furthermore, loneliness increases the demand for health care and long-term support services, placing substantial strain on health care systems [5]. Therefore, there is an urgent need for cost-effective and accessible interventions to alleviate loneliness among older adults.

In recent years, the development of digital health technologies such as wearable devices, smart home systems, and mobile health apps has created new opportunities to monitor and mitigate loneliness through continuous physiological and behavioral assessment. For example, smartwatches and sensor-based wearables can monitor posture as well as physiological parameters linked to loneliness, such as heart rate, respiration, and blood oxygen levels [9-12]. Microphone or camera-based smart furniture and environmental monitoring devices can detect activity patterns, and artificial intelligence (AI)-driven interactive platforms are increasingly being used to provide emotional support and facilitate social interaction among older adults [13-15]. Advances in flexible electronics and smart textiles have further enabled the seamless integration of monitoring technologies into garments or furniture, creating unobtrusive and continuous passive monitoring solutions [10,16,17]. These systems can capture multimodal signals such as heart rate, respiration, and movement to infer emotional and social states and deliver personalized interventions. However, despite significant progress in data collection and AI-based analytical technologies, the effectiveness of these systems largely depends on the design of their user interfaces [18,19]. The design of user interfaces, particularly those for older adults, should ensure that users can understand, trust, and comfortably interact with the system, while also promoting inclusivity and long-term engagement to support psychological well-being [20].

Older adults may face greater challenges when interacting with digital technologies and user interfaces due to limited digital literacy, cognitive and sensory changes, and age-related physical decline [21]. For instance, the loss of fine motor skills can make it difficult to interact with small virtual buttons [22]. In addition, visual impairments can hinder the perception of detail, color, and contrast, and hearing loss can reduce responsiveness to auditory feedback cues [23]. Therefore, the design of interactive platforms and user interfaces should consider these age-related changes to ensure inclusivity and usability. Beyond these physiological and cognitive considerations, emotional comfort and psychological acceptance are essential in determining older adults' engagement with digital systems. Many older adults may experience stigma associated with mental health or hold generational beliefs that discourage emotional disclosure [24,25]. As a result, they may be reluctant to use mental health-related technologies, perceiving them as intrusive or impersonal. Increasing evidence suggests that systems lacking empathy, warmth, and reassurance may unintentionally reinforce feelings of loneliness rather than alleviate them [26,27].

Although digital mental health interventions have gained increasing attention, research specifically focusing on the design of interactive platforms and user interfaces for psychological and emotional monitoring among older adults remains limited [20]. Most current studies have mainly focused on usability testing or technical feasibility, such as investigating conversational agents, social platforms, and mobile apps for emotional well-being in older adults [28-31]. And these studies mainly adopted in-depth qualitative evaluation or user testing, providing valuable insights into user experiences and contextual needs. However, they generally do not extend to systematic or quantitatively prioritized design frameworks that integrate technological, emotional, and psychological factors. To address this gap, this study developed a user-centered framework for interface design in loneliness monitoring technologies for older adults. By integrating qualitative focus group analysis, systematic literature search, and expert consultation, this study established a comprehensive design indicator framework based on the analytic hierarchy process (AHP) [32]. Interface design solutions were then developed according to these indicators and evaluated using the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) [33]. Building on the DELONELINESS (Design for Healthy Ageing: a Smart System to Decrease Loneliness for Older People) smart textile loneliness monitoring system we previously developed [34], we aimed to construct a hierarchical design indicator model reflecting the cognitive, emotional, and functional needs of older adults. Expert consensus was used to assign weights to these indicators, followed by the development and evaluation of interface design solutions. Through this process, the study generates a quantified prioritization of interface design requirements for loneliness monitoring in older adults and provides an evidence-based decision model that can be adapted to other age-friendly digital mental health interfaces. This integrated approach bridges user-centered design principles with data-driven decision analysis, providing a solid foundation for the future development of digital mental health technologies and interactive platforms for older adults.

Methods

Study Design

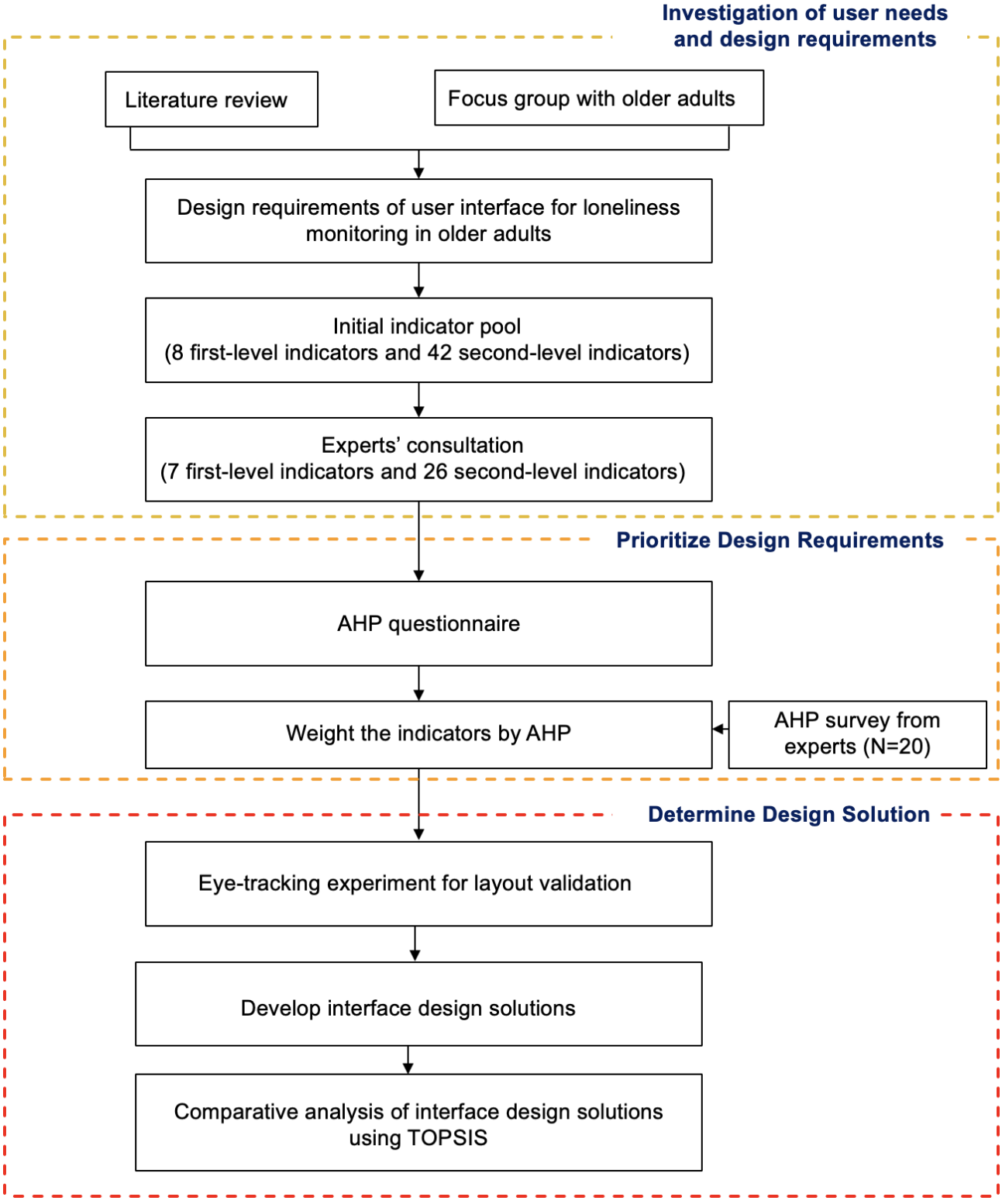
This study applied a mixed method to prioritize interface design indicators and to develop interface solutions for the DELONELINESS system for older adults. The DELONELINESS system is a smart garment and furniture-based platform designed to monitor loneliness in older adults [34]. Through its associated user interfaces, the system provides older users with real-time feedback on various physiological and behavioral parameters such as heart rate, respiration rate, and daily activity levels, while also visualizing the current loneliness

XSL•FO
RenderX

status and offering recommendations for health-promoting activities.

An initial pool of design indicators was constructed based on a literature review and qualitative analysis of previous focus groups with older adults [34]. Following screening by experts of human-computer interaction (HCI), an AHP survey was conducted to determine the relative weights and priorities of the indicators. To further verify the applicability of the design indicators derived from the AHP analysis in actual interface structures, an eye-tracking experiment was conducted to compare different interface layouts in terms of information accessibility and cognitive clarity. Then, the findings of the eye-tracking experiment and user requirements with higher weight values derived from AHP were integrated to produce three interface design solutions for comparative analysis. Finally, the design solutions were evaluated using the TOPSIS to validate their overall feasibility. The research process is presented in Figure 1.

Figure 1. The research process. AHP: analytic hierarchy process; TOPSIS: Technique for Order Preference by Similarity to Ideal Solution.



XSL•FO
RenderX

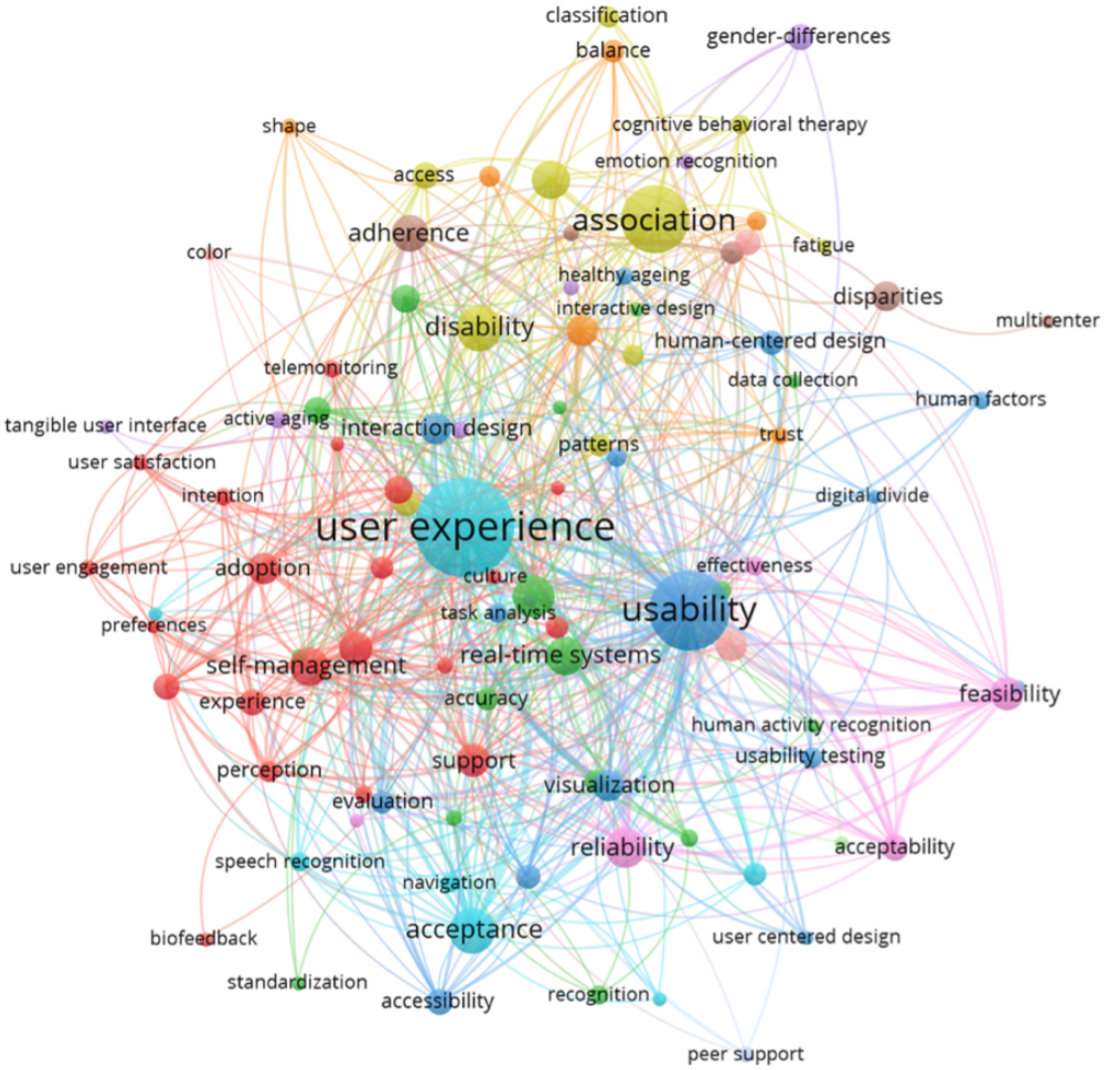
Construction of Initial Indicator Pool

The initial indicator pool was constructed based on a comprehensive literature search and qualitative analysis of focus groups with older adults. The literature search was conducted across 4 major databases including IEEE Xplore, ACM Digital Library, Scopus, and Web of Science to identify studies related to older adults, health technology platforms, and interface design. The search string was defined as:

("older adults" OR "aging people" OR "aging" OR "senior citizens" OR "elderly") AND ("interface design" OR "user interface" OR "interaction design" OR "human computer interaction" OR "user experience") AND ("loneliness monitoring" OR "mental health" OR "psychological wellbeing" OR "digital health")

From the search results, 3522 highly relevant papers with more than 10 citations were selected for further analysis. First, nondescriptive and generic terms (such as "health" and "technology"), as well as duplicates were removed. Additionally, high-frequency terms were extracted using term frequency–inverse document frequency analysis and visualized [35]. As shown in Figure 2, the node size and color represent word frequency, with larger nodes indicating higher frequency, while node color represents keyword clustering. From this analysis, high-frequency terms were synthesized into preliminary interface-related indicators such as usability, feasibility, accessibility, and reliability.

Figure 2. Co-occurrence analysis of high-frequency terms in research of digital health and interface design for older adults.



Additionally, qualitative thematic analysis was conducted on the results of our previous focus groups with older adults [34]. These focus groups explored older users' lived experiences of loneliness, expectations, and concerns regarding digital health and monitoring technologies for loneliness across diverse living contexts. The transcripts were coded using open coding. Two researchers independently coded data and resolved discrepancies through discussion to ensure consistency. Statements and discussion related to comprehensibility, trust and safety, emotional comfort, and personalization of user interfaces were coded (Multimedia Appendix 1). These findings provided a user-centered perspective and complemented the design requirements identified through the literature search.

By integrating results from both the literature and focus group analysis, an initial indicator pool comprising 8 first-level and 42 second-level indicators was established to demonstrate different aspects of user interface design for loneliness

XSL•FO
RenderX

monitoring in older adults. Then, consultation was undertaken with 8 experts in HCI design. These experts were not involved in the AHP survey. They reviewed the combined results of the literature review and focus group analysis, deleted redundant or overlapping items, and suggested additional indicators where appropriate. Finally, a refined hierarchical indicator structure containing 7 first-level indicators and 26 second-level indicators was determined.

Inclusion Criteria of Experts

To evaluate and prioritize the interface design indicators, experts were recruited to participate in the AHP survey. To ensure the validity and reliability of the weighting results, the inclusion criteria for experts were as follows: (1) professional background: experts were required to have academic or professional expertise in HCI, digital health, gerontology, or health informatics; (2) experience: at least 5 years of professional experience in the relevant fields; (3) academic or professional rank: an intermediate level or higher professional title; and (4) voluntary participation: willingness to participate in the survey and agreement of informed consent for the use of their responses in the research.

A total of 23 experts meeting the above criteria were invited to participate. 20 out of 23 (86.96%) experts completed the questionnaire. All these experts held a master's degree or higher and were mainly composed of university lecturers, postdoctoral researchers, clinical experts, and industry professionals with work experience in HCI design and digital health. The demographic characteristics of the experts are summarized in Table 1.

Table . Characteristics of the analytic hierarchy process (AHP) participants (N=20).

Characteristics	Values, n (%)
Professional background	
Human-computer interaction	7 (35)
Digital health	6 (30)
Gerontology	4 (20)
Health informatics	3 (15)
Occupation	
University lecturers	5 (25)
Postdoctoral researchers	3 (15)
Clinical experts	4 (20)
Industry professionals	8 (40)
Seniority (y)	
6 - 10	14 (70)
11 - 15	3 (15)
16 - 20	2 (10)
20	1 (5)
Professional title	
Middle	5 (25)
Associate senior	8 (40)
Senior	7 (35)
Education	
Master's degree	8 (40)
Doctoral degree	12 (60)

Weight Assignment of Indicators

Based on the refined indicator pool, we developed an AHP questionnaire (Multimedia Appendix 2) comprising 7 first-level indicators and 26 second-level indicators. The questionnaire was structured as a pairwise comparison matrix, requiring experts to judge the relative importance of each indicator in relation to the others. To guide the evaluation, we adopted the Saaty 1 to 9 scale, which is widely applied in multicriteria decision analysis [32]. On this scale, a score of "1" represents equal importance, while larger odd numbers indicate greater importance, such as "3" (moderate importance), "5" (strong importance), "7" (very strong importance), and "9" (absolute importance). Even numbers (2, 4, 6, and 8) were used to allow for more nuanced judgments between categories. This structured approach ensured that expert evaluations could be translated into quantitative weights for each indicator [36,37]. The survey was distributed via email to 23 eligible experts, and 20 experts completed the questionnaire.

XSL•FO
RenderX

The collected data were first entered into Microsoft Excel and then imported into SPSS software (version 29.0.2.0; IBM Corp) to calculate the weights of both first-level and second-level indicators. For each judgment matrix, the largest eigenvalue and its corresponding eigenvector were calculated to determine the relative weights of indicators. The consistency ratio (CR) was used to verify the logical consistency of expert judgments, with CR <0.1 indicating acceptable consistency. To further assess the degree of agreement among experts, the Kendall coefficient of concordance (Kendall W) was applied, with values ranging from 0 to 1, where higher values represent greater agreement [38]. Results were considered statistically significant when P<.05. Finally, the resulting pairwise comparison matrix from all experts was integrated to obtain the final weight of each indicator. The weighted values were normalized so that the sum of all indicator weights equaled 1. These results provided a prioritized structure of interface design requirements, informing the design of following user interface solutions.

Development of Design Solutions

After the weight assignment of design indicators through AHP, we examined the applicability of the prioritized indicators to interface layouts. At this early prototyping stage, an eye-tracking experiment was conducted to inform layout selection and structural design decisions rather than to evaluate end-user experience. The eye-tracking study was designed to assess how different layout types (list-detail, grid, and feed layouts) influenced information localization and visual-cognitive efficiency within the DELONELINESS application. A Tobii Fusion eye tracker (60 Hz) was used to record gaze behavior. The participants were the same 20 experts who had taken part in the previous AHP evaluation. Each expert completed information search tasks in various layout types through locating specific interface elements and triggering screen transitions by clicking on the target area. The system automatically recorded each participant's gaze trajectory and task completion time.

We analyzed 2 primary performance metrics, including gaze path distribution and task completion time (TCT). Based on the results, 3 user interface design solutions were developed. These design solutions varied in layout and user experience but were all aligned with the prioritized requirements identified in the weight assignment stage.

Evaluation of Design Solutions

To evaluate the feasibility of these 3 design solutions, the same group of experts from the AHP phase was invited to participate in a structured assessment. Experts rated each design solution against the indicators obtained from the previous AHP analysis. Ratings were collected using a 9-point Likert scale, ranging from 1 ("Strongly Disagree") to 9 ("Strongly Agree") to quantify the performance of each design solution across the 7 first-level and 26 second-level indicators. After collecting the ratings, the highest and lowest values were excluded, and the mean of the remaining scores was calculated as the composite score for each solution.

To evaluate the reliability of expert ratings, the intraclass correlation coefficient (ICC) was computed. A 2-way random mixed model with absolute agreement was applied. ICC values greater than 0.60 were considered indicative of good consistency, while values above 0.75 indicated excellent consistency. Additionally, we applied the TOPSIS to calculate the evaluation score for each design solution. This method constructs an ideal solution (the best performance across all indicators) and a negative ideal solution (the worst performance across all indicators) and measures the relative distance of each design alternative to these 2 reference points [33]. By calculating the closeness coefficient, these 3 solutions were ranked and the most suitable interface design for the DELONELINESS system was identified.

Ethical Considerations

This research received ethical approval from the Research Ethics Committee of King's College London (reference: LRS/DP-24/25 - 34602). Prior to participation, all participants received a participant information sheet and a consent form via email, which outlined the study objectives, participants' rights, and the voluntary nature of participation. Electronic informed consent was obtained before data collection. Files containing participants' names and contact details were stored separately from research data in password-protected documents, accessible only to the core research team. No compensation was provided for participation.

Results

Weight Assignment of Indicators

To evaluate the reliability of expert judgments, a consistency test was performed on each expert's pairwise comparison matrix. The CR of all matrices was less than 0.1, indicating that all expert judgments met the consistency requirement. In addition, to assess the overall agreement among experts in judging the importance of indicators at each hierarchical level, the Kendall W was calculated for both first and second-level indicators. The Kendall's W value was 0.313 for the first-level indicators (χ^2 =37.606, P<.001) and 0.156 for the second-level indicators (χ^2_1 =59.286, P<.001). These results demonstrate a satisfactory level of logical consistency among experts, supporting the validity of further weight calculation and coordination analysis.

After the consistency test, the AHP was applied to calculate the normalized weights and combination weights of each indicator. The detailed weighting results are presented in Multimedia Appendix 3. Among the first-level indicators, trust and safety (0.206) had the highest overall weight, followed by ease of use (0.187) and accessibility (0.167), indicating that these dimensions were considered the most critical factors in the interface design of the DELONELINESS system. In contrast, emotional comfort (0.080) and personalization (0.092) had relatively lower weights. At the second level, the highest-weighted indicators included user ability to modify permissions (0.084), clear interaction pathways (0.667), and transparent data collection (0.566), showing that user autonomy and system transparency were regarded as key determinants of older user trust and engagement. Other highly ranked indicators, such as offline usability under poor network conditions (0.550) and clear interaction areas (0.521), highlight the importance of system robustness and interface clarity for older adult users.

XSL•FO
RenderX

Overall, these findings indicate that experts prioritized factors that enhance trust, safety, and operational simplicity over purely aesthetic or emotional aspects. The resulting hierarchical weighting structure provided a quantitative foundation for the further design and evaluation of user interface solutions.

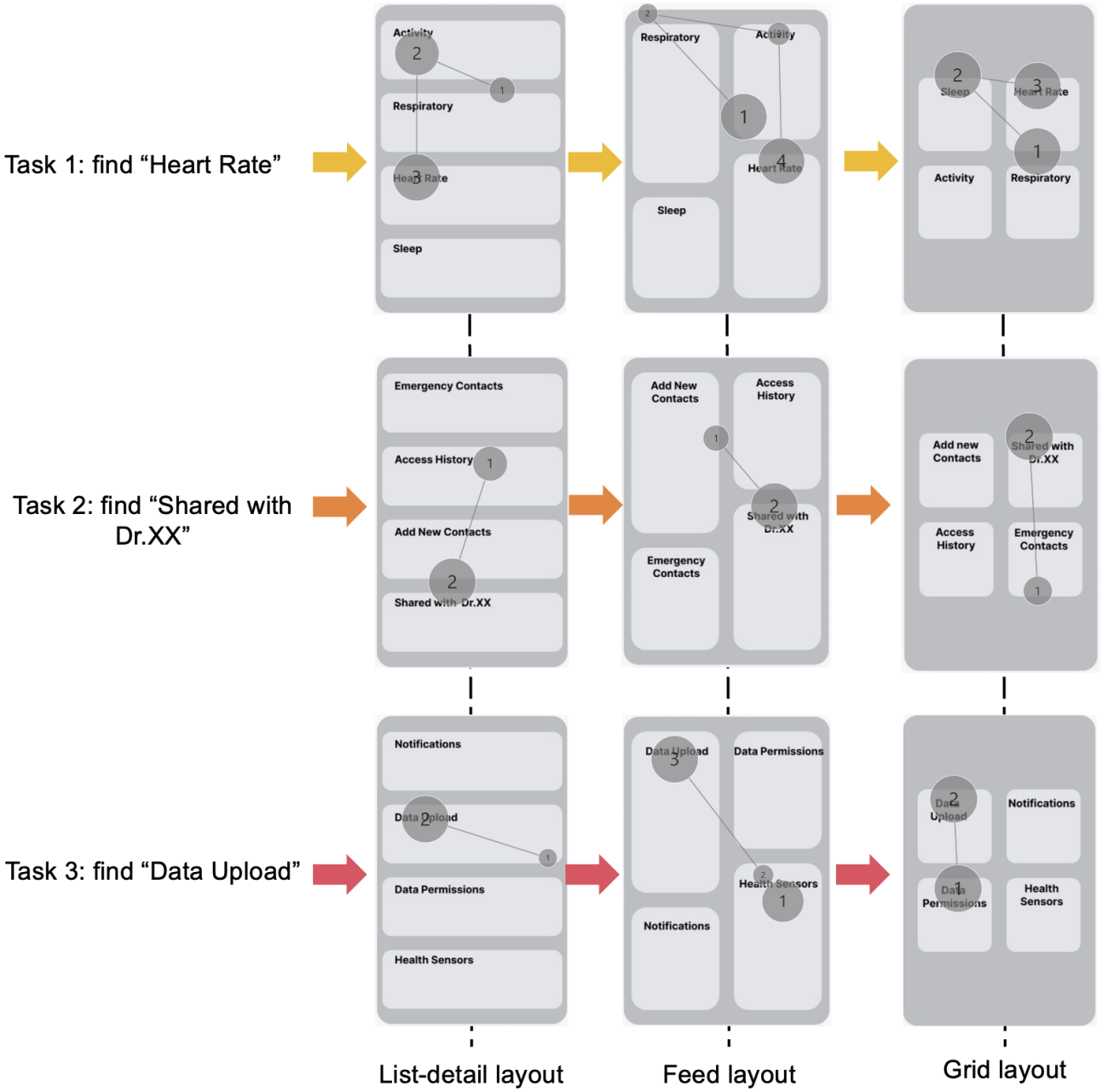
Development of Design Solutions

In the interface design of the DELONELINESS system, we first examined how different layout structures influenced visual-cognitive efficiency and interaction performance. A one-way ANOVA was conducted to analyze the eye-tracking data across different interface layout types. The results revealed significant differences in TCT (P<.05), indicating that interface structure had a significant impact on visual localization efficiency and interaction performance.

Among the 3 layouts, the list–detail layout performed best on TCT (mean 1.33, SD 0.33), lower than the grid layout (mean 1.37, SD 0.33) and the feed layout (mean 1.76, SD 0.65). These findings suggest that the list-detail layout more effectively guided visual attention, reduced information retrieval load, and improved task execution efficiency.

As shown in Figure 3, the list-detail layout showed a more focused and linear visual scanning path, with fixation points concentrated in core information areas. In contrast, the grid layout produced multiple scattered fixation hotspots, whereas the feed layout presented more diffuse but continuous scanning patterns. Overall, the list-detail layout demonstrated outstanding visual and operational efficiency, providing evidence for the following interface prototype design.

Figure 3. Gaze paths across 3 interface layout types in different tasks.



Based on these findings, the hierarchical weighting structure derived from the AHP was integrated into the interface design process. Among all design indicators, trust and safety, ease of use, and accessibility were identified as the most important

XSL•FO
RenderX

dimensions. Therefore, 3 interface design solutions were developed (Figure 4). Each design followed the same hierarchical weighting framework and maintained a consistent functional architecture, including the home, data, activities, sharing, and privacy modules. However, they differed in the expression and emphasis of high-priority dimensions.

Figure 4. User interface design solutions for DELONELINESS (Design for Healthy Ageing: a Smart System to Decrease Loneliness for Older People) system: (A) design solution 1, (B) design solution 2, and (C) design solution 3.



As shown in Figure 4, design solution 1 prioritizes trust and safety and comprehensibility by providing users with transparent control over their data and privacy settings. The homepage displays a simplified loneliness status section, using soft colors and positively framed language to convey feedback. Below, the section titled "What's Happening With Your Data?" allows users to quickly access information about ongoing data collection and its purposes. The "Share and Contacts" page visualizes access permissions and emergency contacts and enables users to view their data-sharing history, which enhances both their sense of security and control. Within the "Privacy Center," users can manage notifications and data permissions and pause data uploads at any time.

Design solution 2 (Figure 4) focuses on ease of use and comprehensibility, addressing users' needs for clarity and intuitive interaction. The homepage features a large central status indicator, accompanied by a simplified health score and encouraging feedback. Navigation follows a linear guided structure to reduce users' cognitive load. Users can access "Quick Controls" for notifications and data uploads, while the "Privacy" section integrates all permission settings in an intuitive toggle-based layout. Additional cues such as tooltips and contextual guidance can help older adults navigate unfamiliar interactions. This design emphasizes clear interaction pathways and immediate guidance, enhancing operational ease and user confidence during daily use.

Design solution 3 (Figure 4) emphasizes accessibility and ensures usability under unstable network conditions. The interface adapts seamlessly across smartphones and tablets and supports a low-bandwidth mode to enable offline access. Key health and loneliness indicators can remain available locally and synchronize automatically once connectivity is restored. The "Accessibility" section offers high-contrast modes, enlarged fonts, and voice-assist features, improving inclusivity and reliability. This design ensures that the system remains usable for older adults with sensory limitations or those living in poor network environments.

Evaluation of Design Solutions

To evaluate the feasibility of these three design solutions, the TOPSIS was applied to compare their performance and identify the optimal interface design for the DELONELINESS system. First, experts rated each design solution across 7 first-level and 26 second-level indicators using a 9-point Likert scale. To assess the reliability of expert ratings, the ICC was calculated. The ICC values for design solution 1, design solution 2, and design solution 3 were 0.761, 0.653, and 0.747, respectively. All ICC values exceeded 0.6, indicating moderate to good consistency among experts. The mean scores of each design solution are shown in Table 2. At the first indicator level, design solution 2 achieved the highest mean scores across most dimensions, particularly in comprehensibility (6.39), ease of use (6.50), and feedback and support (5.67), demonstrating its guided and

XSL•FO
RenderX

user-friendly interaction design. Design solution 1 emphasized trust and data transparency and achieved the highest score in trust and safety (6.17). Design solution 3 focused on accessibility and obtained the relatively higher scores in accessibility (5.23) and comprehensibility (6.22). At the second indicator level, design solution 2 achieved higher average scores in clarity of visual elements (6.61), consistency of interface layout (6.33), and clear interaction areas (6.28), suggesting experts' preference for its clear visual hierarchy and intuitive structure. Design solution 3 performed well in cross-device access (5.43), multisystem compatibility (5.33), and accessibility support functions (4.61), indicating good robustness in supporting diverse usage contexts. Although design solution 1 received slightly lower overall scores, it performed best in transparent data collection (6.17) and user authorization of functions (6.06), demonstrating perceptions of trust and control.

XSL•FO
RenderX

Table . Mean expert ratings of the 3 interface design solutions.

	Design solution 1	Design solution 2	Design solution 3
First-level indicators, mean (SD)			
Comprehensibility	5.28 (0.82)	6.39 (1.14)	6.22 (1.07)
Ease of use	5.26 (1.05)	6.50 (0.64)	6.06 (1.02)
Trust and safety	6.17 (0.98)	5.61 (1.32)	6.11 (1.00)
Feedback and support	5.06 (1.00)	5.67 (1.27)	5.39 (0.82)
Emotional comfort	4.22 (0.72)	5.33 (1.17)	5.22 (1.05)
Personalization	4.22 (1.02)	5.11 (1.15)	5.06 (0.92)
Accessibility	4.44 (0.70)	4.78 (1.23)	5.23 (1.09)
Second-level indicators, mean (SD)			
1.1 Consistency of interface layout	5.17 (1.19)	6.33 (1.45)	6.22 (0.91)
1.2 Clarity of visual elements	5.11 (1.06)	6.61 (1.67)	6.28 (1.08)
1.3 Simplicity of information content	4.94 (1.14)	6.22 (1.38)	6.39 (0.97)
2.1 Clear interaction areas	5.06 (1.08)	6.28 (1.76)	6.06 (1.15)
2.2 Support for multimodal interaction	5.28 (1.16)	6.33 (1.62)	5.83 (0.87)
2.3 Easy correction of errors	5.28 (0.98)	6.22 (1.72)	6.28 (1.02)
2.4 Clear interaction pathways	5.17 (1.14)	6.00 (1.34)	6.06 (1.19)
3.1 Transparent data collection	6.17 (0.94)	6.11 (0.91)	5.72 (1.29)
3.2 User authorization of functions	6.06 (1.15)	6.00 (1.02)	5.96 (1.22)
3.3 User ability to modify permissions	5.39 (0.92)	5.72 (1.21)	5.67 (1.03)
4.1 Real-time health status feedback	4.78 (1.16)	5.50 (1.28)	5.56 (1.19)
4.2 Immediate guidance	4.56 (1.14)	5.22 (1.54)	5.61 (1.33)
4.3 Online assistance	4.50 (1.32)	5.33 (1.44)	5.39 (0.94)
5.1 Emotionally supportive language	4.11 (0.97)	5.06 (1.46)	5.11 (1.04)
5.2 Emotionally supportive images	4.39 (0.94)	5.28 (1.52)	5.28 (1.12)
5.3 Integration with daily life scenarios	4.50 (1.06)	5.28 (1.27)	5.00 (1.21)
6.1 Customizable fonts	4.06 (1.23)	4.89 (1.41)	4.94 (1.10)
6.2 Customizable language	4.39 (1.05)	4.72 (1.59)	4.89 (1.02)
6.3 Customizable notifications	4.00 (1.05)	4.78 (1.35)	4.83 (1.27)
6.4 Customizable mode	4.39 (0.98)	5.00 (1.52)	5.06 (1.34)
6.5 Customizable mood input	4.28 (1.17)	4.89 (1.63)	4.83 (1.06)
7.1 Cross-device access	4.33 (0.91)	4.61 (1.38)	5.43 (1.04)
7.2 Multisystem compatibility	4.28 (0.89)	4.72 (1.59)	5.33 (1.15)
7.3 Accessibility support functions	4.17 (0.76)	4.44 (1.41)	4.61 (1.29)
7.4 Offline usability under poor network conditions	4.22 (0.78)	4.78 (1.61)	4.98 (1.38)

XSL•FO
RenderX

Based on the expert ratings and the AHP indicator weights, the TOPSIS method was used to evaluate the performance of the 3 design solutions. The distances to the positive and negative ideal solutions and the overall closeness coefficients are presented in Table 3. The results showed that design solution 2 had the highest overall score, followed by design solution 3 and design solution 1. These findings indicate that design solution 2 was closest to the ideal interface configuration, which can offer the best balance of trust, usability, and accessibility.

Table . Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) evaluation of design solutions.

	Distance to positive ideal solution (D^+)	Distance to negative ideal solution (D^-)	Relative closeness (C)	Ranking
Design solution 1	0.999	0.002	0.002	3
Design solution 2	0.132	0.942	0.877	1
Design solution 3	0.318	0.855	0.728	2

Discussion

Principal Findings

This study aimed to construct a user-interface design indicator model for the DELONELINESS system, a smart textile system for monitoring loneliness among older adults. To comprehensively integrate design requirements for health technology platforms targeting older users, we applied a mixed methods approach, combining a systematic literature search, qualitative focus group analysis, and expert consultation. These processes led to the construction of a design indicator framework, consisting of 7 first-level indicators and 26 second-level indicators. The AHP was then applied to assign relative weights to each indicator. Based on the derived weights, 3 interface design solutions were developed and evaluated by experts using the TOPSIS to identify the optimal configuration. The findings indicate that trust and safety, ease of use, and accessibility were prioritized over other design dimensions, highlighting the central role of transparency, cognitive simplicity, and inclusive access in interface design for monitoring loneliness among older adults.

The AHP results constructed a hierarchical weighting structure. Among the 7 first-level indicators, trust and safety (0.206), ease of use (0.187), and accessibility (0.167) were identified as the most important dimensions, indicating that data transparency, simplified user operations, and accessibility under different physical and technological conditions are essential for enhancing older adults' confidence and engagement with digital systems. At the second level, user ability to modify permissions (0.084), clear interaction pathways (0.067), and transparent data collection (0.057) received the highest weights, demonstrating the key role of user autonomy and system transparency in fostering trust and usability.

Based on these priorities, three interface design solutions were developed. Each solution followed the same hierarchical framework but emphasized different high-priority dimensions. Specifically, design solution 1 focused on trust and safety, design solution 2 prioritized ease of use and comprehensibility, and design solution 3 emphasized accessibility. Expert evaluation using the TOPSIS method revealed that design solution 2 achieved the highest overall score (C=0.877), followed by design solution 3 (C=0.728) and design solution 1 (C=0.002). These findings suggest that interfaces emphasizing clear interaction pathways, visual clarity, and immediate feedback achieve the most effective balance among trust, usability, and accessibility for older users.

Implications for Practice and Research

In recent years, digital health technologies have played an increasingly important role in addressing loneliness and improving psychological well-being among older adults. Loneliness monitoring platforms, telehealth systems, and mobile mental health apps mainly applied wearable sensors, mobile devices, and AI-driven analytics to provide continuous monitoring and personalized interventions. These systems have shown great potential in detecting early psychosocial risks and promoting engagement in remote health care, particularly for older adults who have limited social support. However, the effectiveness of these technologies largely depends on the design of user interfaces. Therefore, ensuring older users can understand, trust, and comfortably interact with these systems is essential. Furthermore, empathic interface design that resonates with the emotional and relational dimensions of loneliness is important for fostering long-term engagement and psychological support [39]. Such designs not only convey a sense of care but also help bridge the digital divide between older adults and technology, mitigating digital loneliness through inclusive design.

Current research on loneliness and emotional well-being monitoring systems for older adults has primarily focused on usability and acceptability. For example, Choi et al [40] developed a mobile loneliness intervention app, highlighting usability, feedback clarity, and emotional support as key determinants of user acceptance. Rodríguez-Martínez et al [41] proposed a conversational chatbot designed to alleviate loneliness in older adults, emphasizing the importance of privacy and autonomy in interaction design. Xing et al [42] investigated a community engagement platform that supported older adults during the COVID-19 pandemic and underscored the role of data transparency and user control in data sharing. Additionally, Bhattacharya et al [43] demonstrated that multimodal interface design integrating voice, tactile, and visual feedback enhanced the older adults' engagement and comprehension in telehealth applications. While these studies have identified recurring design dimensions such as usability, confidentiality, and accessibility, most relied on qualitative methods or usability testing. To address this gap, this study introduces a user-centered quantifiable hierarchical framework, shifting from single-factor

XSL•FO
RenderX

evaluations to a systematic and multidimensional method for assessing interactive factors.

Theoretically, we introduced a knowledge-driven design pathway grounded in both literature insights and user requirements. Through visualized analysis and keyword co-occurrence mapping of relevant publications, we constructed a research landscape of interface design for mental health technologies in older adults. The qualitative findings were further integrated from focus group discussions, leading to the construction of an initial indicator pool. This process ensured that the theoretical foundation of the framework was both systematic and progressive, enabling the model to reflect not only the expert experience and user needs but also the academic field's evolving focus and research trends. Methodologically, the study integrated the AHP and the TOPSIS to construct a quantifiable hierarchical framework. The AHP was applied to determine experts' judgments on the relative importance of design dimensions, while TOPSIS was used to comprehensively evaluate and prioritize different design solutions based on these weightings, achieving a 3-level validation process spanning theory, structure, and empirical evaluation. This approach transforms interface evaluation from subjective experience into a quantifiable decision-making process, offering a novel means of representing complex psychological and emotional factors in user interface design.

To translate the quantitative results into actionable principles for digital mental health design, we highlighted some major dimensions that received high weighting in the AHP hierarchy. First, trust and safety are the most highly weighted dimensions identified through the AHP, showing that the credibility and transparency of digital platforms are essential for sustained engagement among older adults. Furthermore, significant progress has been made in strategies for health data protection and security, such as encryption technologies, anonymization protocols, and blockchain-based solutions which can improve data privacy and integrity [44-46]. These studies have shown that clear privacy principles and controllability can reduce technology-related anxiety and increase adherence to digital health systems. In the context of loneliness monitoring, fostering user autonomy not only mitigates distrust but also transforms the interaction from passive monitoring to active self-management, thereby promoting psychological empowerment among older adults [47].

Additionally, both experts and older users expressed a strong need for clarity and cognitive simplicity when interacting with digital platforms. The design solution 2 incorporated structured task flows, step-by-step prompts, and consistent layout design, which improved expert ratings for comprehensibility, feedback and support, and ease of use. These findings suggest that guided interaction models, which provide immediate feedback, linear navigation, and supportive cues, can effectively reduce cognitive load and enhance older adults' confidence in digital self-management.

Accessibility was also identified as a key factor in age-friendly interface design. The design indicators such as offline usability under poor network conditions and accessibility support functions received high expert ratings in evaluations. However, the wireless communication technologies commonly used in smart garments and furniture such as Bluetooth and Wi-Fi still face significant challenges, particularly in terms of data quality, sampling rate, and transmission range. Data loss or latency can lead to inaccurate or incomplete monitoring, potentially resulting in misinterpretation of users' health conditions [48]. The restricted range of Bluetooth and Wi-Fi may inadvertently reduce older adults' mobility, especially in spacious living environments or outside the home, when precise data collection is required.

Finally, although personalization and emotional comfort received relatively lower weights in the AHP hierarchy, empathetic and adaptive design remains essential in the context of loneliness monitoring for older adults. This result can be interpreted considering the expert decision-making process. Emotional comfort may be implicitly contingent on more foundational interaction requirements, such as trust, safety, usability, and accessibility. This interpretation aligns with motivational models such as the Maslow hierarchy of needs, in which affective and self-related needs tend to be addressed only after the basic functional and safety needs are satisfied [49]. In the context of digital health interfaces, experts may therefore prioritize dimensions that reduce risk, uncertainty, and cognitive burden before considering emotional engagement. Loneliness is a subjective and context-dependent experience, influenced by emotional states, daily routines, and social relationships [50]. Therefore, interface personalization such as allowing users to adjust the tone of feedback, the frequency of reminders, or the interface mode plays a vital role in enabling the system to respond to diverse psychological and emotional needs. Moreover, emotionally supportive and adaptive design can increase older adults' openness to using digital mental health platforms and foster a sense of emotional stability [51]. For example, emotionally supportive language can be translated into the use of positively framed feedback messages such as "You are doing well today" rather than deficit-oriented alerts [52]. In addition, concerns about psychological acceptance and perceived intrusiveness informed design decisions that minimized the salience of continuous monitoring, such as presenting sensing functions in the background and allowing users to pause data collection or customize notifications [53]. Therefore, emotional support and adaptability should not be regarded as aesthetic additions but as ethical foundations of inclusive digital health design. The goal of digital mental health technologies is not just functional usability, but to foster psychological stability and emotional connectivity through empathic interaction mechanisms tailored to the lived experiences of older adults.

The widespread adoption of digital technologies has developed new opportunities for promoting mental health among older adults while simultaneously revealing persistent digital inequalities in contemporary society. Older populations remain at a disadvantage in terms of accessing, understanding, and trusting digital systems [54,55]. This "digital divide" stems not only from differences in individual capability but also from a broader lack of inclusivity and emotional awareness in technology design. Building on 7 key dimensions, including trust and safety, comprehensibility, accessibility, and emotional

XSL•FO
RenderX

comfort, this study proposes a multifactor hierarchical framework that reconceptualizes "usability" from an individual challenge to a societal design responsibility. This shift suggests that digital health systems should move beyond the narrow focus on "user experience" toward the broader pursuit of citizen well-being. As the interface serves as a mediator between people, society, and technology, it carries the social function of emotional connection and psychological support. Accordingly, the quantitative framework established in this study not only enhances the precision of digital product evaluation but also provides a theoretical pathway for achieving ethical inclusivity and social equity in digital health. Future research should further explore the application of this framework in policy evaluation, geriatric care service systems, and cross-cultural digital health contexts, promoting its integration into real-world practices that advance both technological innovation and social justice.

Strengths and Limitations

The innovation aspect of this study lies in the development of a user-centered quantifiable prioritization framework for interface design in loneliness monitoring among older adults. We integrated qualitative focus group analysis, literature search, and expert consultation to ensure that the final indicator system was both evidence-based and grounded in user needs. Furthermore, we applied the AHP and TOPSIS to achieve the prioritization and empirical validation of interface design solutions, providing a systematic approach for future research in digital health and HCI. In addition, within the context of loneliness monitoring for older adults, psychological and emotional dimensions were integrated into the study. These dimensions are critical for engagement, empathy, and sustained interaction of older adults, but often underrepresented in existing digital health design frameworks. By combining user-centered design principles with data-driven decision analysis, this research contributes to the emerging field of digital mental health for older populations.

However, several limitations should be acknowledged. First, although the number of experts was sufficient for AHP analysis, the limited representation from gerontology and health informatics may have constrained the diversity of perspectives from psychology, data security, and clinical care. This imbalance may have influenced the weighting outcomes by placing greater emphasis on interaction efficiency and technical feasibility, while potentially underrepresenting clinically oriented considerations such as long-term care workflows, symptom interpretation, and psychosocial support needs in later life. Moreover, emotional comfort is inherently subjective and more difficult to operationalize into discrete interface features, which may lead experts to deprioritize it relative to more tangible and measurable design dimensions within a hierarchical decision-making process. Future research should expand the expert pool to enhance representativeness and better capture affective and experiential priorities. Second, the evaluation of design solutions in this study mainly relied on expert judgment and expert-based eye-tracking assessment rather than real-world usability testing with older adults. While the use of experts was appropriate for early-stage evaluation of information architecture, expert participants may not fully represent the emotional characteristics of older adults. Future studies can validate these findings through user trials and incorporate user feedback into the iterative design process. As the user interface developed in this study was built upon our smart textile loneliness monitoring system (the DELONELINESS system), further research is required to examine the usability, accessibility, and user acceptability of the integrated system through follow-up usability testing and in-situ evaluation with older adults to verify whether expert-informed design priorities align with real-world user experiences. Finally, future work may explore the integration of this framework within broader health policy contexts, such as the United Kingdom's National Health Service digital care strategies or other international systems, to adapt the model to evolving user needs and technological ecosystems.

Conclusions

This study developed a user-centered quantifiable prioritization framework for interface design in loneliness monitoring among older adults. By integrating qualitative insights from focus groups, literature-derived indicators, and expert consultation, the study established a comprehensive design indicator system based on the AHP and evaluated design solutions using the TOPSIS. Using user-centered design principles with data-driven decision analysis, this research not only provides a systematic design framework for future research in HCI, but also bridges subjective user experience with objective design evaluation, offering practical insights for developing empathic and inclusive digital mental health technologies for older adults. Beyond the DELONELINESS system, the proposed framework can be adapted to inform the design and evaluation of other digital mental health platforms and age-friendly interactive systems by supporting structured prioritization of usability, trust, accessibility, and emotional considerations across diverse contexts.

Funding

This work is supported by the UK Engineering and Physical Sciences Research Council and the National Institute for Health and Care Research (grant: EP/W031434/1).

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Qualitative thematic analysis of focus groups.
[DOCX File, 18 KB - humanfactors_v13i1e88263_app1.docx]

Multimedia Appendix 2
Analytic hierarchy process questionnaire.
[DOCX File, 20 KB - humanfactors_v13i1e88263_app2.docx]

Multimedia Appendix 3
Weights of indicators for the DELONELINESS (Design for Healthy Ageing: a Smart System to Decrease Loneliness for Older People) system interface design.
[DOCX File, 16 KB - humanfactors_v13i1e88263_app3.docx]

References

1. Guglielmi V, Colangeli L, Parrotta ME, et al. Social isolation and loneliness in non-communicable chronic diseases: impact of COVID-19 pandemic, population aging and technological progress. Nutr Metab Cardiovasc Dis 2025 Jun;35(6):104015. [doi: 10.1016/j.numecd.2025.104015] [Medline: 40189996]
2. Walsh BE, Rottenberg J, Schlauch RC. Why loneliness requires a multidimensional approach: a critical narrative review. Nat Mental Health 2025;3(2):175-184. [doi: 10.1038/s44220-024-00382-3]
3. Teo RH, Cheng WH, Cheng LJ, Lau Y, Lau ST. Global prevalence of social isolation among community-dwelling older adults: a systematic review and meta-analysis. Arch Gerontol Geriatr 2023 Apr;107:104904. [doi: 10.1016/j.archger.2022.104904] [Medline: 36563614]
4. Dahlberg L, McKee KJ, Frank A, Naseer M. A systematic review of longitudinal risk factors for loneliness in older adults. Aging Ment Health 2022 Feb;26(2):225-249. [doi: 10.1080/13607863.2021.1876638] [Medline: 33563024]
5. Aalto UL, Bonin-Guillaume S. Editorial: loneliness among older people exacerbated by the COVID-19 pandemic. J Nutr Health Aging 2023;27(8):617-618. [doi: 10.1007/s12603-023-1968-z] [Medline: 37702333]
6. Benson JA, McSorley VE, Hawkley LC, Lauderdale DS. Associations of loneliness and social isolation with actigraph and self-reported sleep quality in a national sample of older adults. Sleep 2021 Jan 21;44(1):zsaa140. [doi: 10.1093/sleep/zsaa140] [Medline: 32691067]
7. Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspect Psychol Sci 2015 Mar;10(2):227-237. [doi: 10.1177/1745691614568352] [Medline: 25910392]
8. Hawthorne G. Measuring social isolation in older adults: development and initial validation of the friendship scale. Soc Indic Res 2006 Jul;77(3):521-548. [doi: 10.1007/s11205-005-7746-y]
9. Schrempft S, Jackowska M, Hamer M, Steptoe A. Associations between social isolation, loneliness, and objective physical activity in older men and women. BMC Public Health 2019 Jan 16;19(1):74. [doi: 10.1186/s12889-019-6424-y] [Medline: 30651092]
10. Ejupi A, Menon C. Detection of talking in respiratory signals: a feasibility study using machine learning and wearable textile-based sensors. Sensors (Basel) 2018 Jul 31;18(8):2474. [doi: 10.3390/s18082474] [Medline: 30065177]
11. McCrory C, Finucane C, O'Hare C, et al. Social disadvantage and social isolation are associated with a higher resting heart rate: evidence from the Irish Longitudinal Study on Ageing. J Gerontol B Psychol Sci Soc Sci 2016 May;71(3):463-473. [doi: 10.1093/geronb/gbu163]
12. Nath RK, Thapliyal H. Machine learning-based anxiety detection in older adults using wristband sensors and context feature. SN Comput Sci 2021 Sep;2(5):359. [doi: 10.1007/s42979-021-00744-z]
13. Kim JE. Commentary on electronic health record usability and postsurgical outcomes among older adults with dementia. Am J Geriatr Psychiatry 2023 Jul;31(7):501-502. [doi: 10.1016/j.jagp.2023.03.001] [Medline: 36997404]
14. Vuegen L, Van Den Broeck B, Karsmakers P, Van hamme H, Vanrumste B. Monitoring activities of daily living using wireless acoustic sensor networks in clean and noisy conditions. Annu Int Conf IEEE Eng Med Biol Soc 2015;2015:4966-4969. [doi: 10.1109/EMBC.2015.7319506] [Medline: 26737406]
15. Petersen J, Thielke S, Austin D, Kaye J. Phone behaviour and its relationship to loneliness in older adults. Aging Ment Health 2016 Oct;20(10):1084-1091. [doi: 10.1080/13607863.2015.1060947] [Medline: 26133745]
16. Jiang L, Siriaraya P, Choi D, Zeng F, Kuwahara N. Electroencephalogram signals emotion recognition based on convolutional neural network-recurrent neural network framework with channel-temporal attention mechanism for older adults. Front Aging Neurosci 2022;14:945024. [doi: 10.3389/fnagi.2022.945024] [Medline: 36212045]
17. Azeem M, Shahid M, Masin I, Petru M. Design and development of textile-based wearable sensors for real-time biomedical monitoring; a review. J Text Inst 2025 Jan 2;116(1):80-95. [doi: 10.1080/00405000.2024.2318500]
18. Eisma R, Dickinson A, Goodman J, Syme A, Tiwari L, Newell AF. Early user involvement in the development of information technology-related products for older people. Universal Access Inf Soc 2004 Jun 1;3(2):131-140. [doi: 10.1007/s10209-004-0092-z]

XSL•FO
RenderX

19. McMurtrey ME, McGaughey RE, Downey JR. Seniors and information technology: Are we shrinking the digital divide? J Int Technol Inf Manag 2008;17(2). [doi: 10.58729/1941-6679.1113]
20. Thangavel G, Memedi M, Hedström K. Customized information and communication technology for reducing social isolation and loneliness among older adults: scoping review. JMIR Ment Health 2022 Mar 7;9(3):e34221. [doi: 10.2196/34221] [Medline: 35254273]
21. Lee C, Coughlin JF. PERSPECTIVE: older adults' adoption of technology: an integrated approach to identifying determinants and barriers. J of Product Innov Manag 2015 Sep;32(5):747-759. [doi: 10.1111/jpim.12176]
22. Johansson-Pajala RM, Gusdal A, Eklund C, Florin U, Wågert PVH. A codesigned web platform for reducing social isolation and loneliness in older people: a feasibility study. Inform Health Soc Care 2023 Apr 3;48(2):109-124. [doi: 10.1080/17538157.2022.2070068] [Medline: 35579407]
23. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: a literature based framework (MOLD-US). Int J Med Inform 2018 Jun;114:66-75. [doi: 10.1016/j.ijmedinf.2018.03.012] [Medline: 29673606]
24. Ratcliffe J, Kanaan M, Galdas P. Reconceptualising men's loneliness: an interpretivist interview study of UK-based men. Soc Sci Med 2023 Sep;332:116129. [doi: 10.1016/j.socscimed.2023.116129] [Medline: 37531906]
25. Rees J, Liu W, Canson J, et al. Qualitative exploration of the lived experiences of loneliness in later life to inform technology development. Int J Qual Stud Health Well-being 2024 Dec;19(1):2398259. [doi: 10.1080/17482631.2024.2398259] [Medline: 39305060]
26. Rees J, Matcham F, Probst F, et al. Wearables, sensors and the future of technology to detect and infer loneliness in older adults. Gerontechnology 2023 Aug 31;22(2). [doi: 10.4017/gt.2023.22.2.ree.08]
27. Rees J, Liu W, Ourselin S, et al. Understanding the psychological experiences of loneliness in later life: qualitative protocol to inform technology development. BMJ Open 2023 Jun 19;13(6):e072420. [doi: 10.1136/bmjopen-2023-072420] [Medline: 37336536]
28. Al Mahmud A, Long KM, Harrington KD, et al. Developing a digital psychoeducational tool to reduce loneliness in older adults: a design case study. Int J Hum Comput Interact 2022 Apr 3;38(6):499-528. [doi: 10.1080/10447318.2021.1949854]
29. Muuraiskangas S, Leist AK, Braun A, et al. V2me: evaluating the first steps in mobile friendship coaching. J Ambient Intell Smart Environ 2012;4(6):517-534. [doi: 10.3233/AIS-2012-0177]
30. Koceski S, Koceska N. Evaluation of an assistive telepresence robot for elderly healthcare. J Med Syst 2016 May;40(5):121. [doi: 10.1007/s10916-016-0481-x] [Medline: 27037685]
31. Brandenburgh A, van Breda W, van der Ham W, Klein M, Moeskops L, Roelofsma P. Increasing physical and social activity through virtual coaching in an ambient environment. Presented at: Active Media Technology AMT 2014; Aug 11-14, 2014. [doi: 10.1007/978-3-319-09912-5_41]
32. Saaty RW. The analytic hierarchy process—what it is and how it is used. Math Model 1987;9(3-5):161-176. [doi: 10.1016/0270-0255(87)90473-8]
33. Márquez-Barja J, Calafate CT, Cano JC, Manzoni P. An overview of vertical handover techniques: algorithms, protocols and tools. Comput Commun 2011 Jun;34(8):985-997. [doi: 10.1016/j.comcom.2010.11.010]
34. Probst F, Rees J, Aslam Z, et al. Evaluating a smart textile loneliness monitoring system for older people: co-design and qualitative focus group study. JMIR Aging 2024 Dec 17;7(1):e57622. [doi: 10.2196/57622] [Medline: 39688889]
35. Wu HC, Luk RWP, Wong KF, Kwok KL. Interpreting TF-IDF term weights as making relevance decisions. ACM Trans Inf Syst 2008 Jun;26(3):1-37. [doi: 10.1145/1361684.1361686]
36. Chakraborty S, Raut RD, Rofin TM, Chakraborty S. A comprehensive and systematic review of multi-criteria decision-making methods and applications in healthcare. Healthc Anal 2023 Dec;4:100232. [doi: 10.1016/j.health.2023.100232]
37. Schmidt K, Aumann I, Hollander I, Damm K, von der Schulenburg JMG. Applying the Analytic Hierarchy Process in healthcare research: a systematic literature review and evaluation of reporting. BMC Med Inform Decis Mak 2015 Dec 24;15(1):112. [doi: 10.1186/s12911-015-0234-7] [Medline: 26703458]
38. Field AP. Kendall's coefficient of concordance. In: Wiley StatsRef: Statistics Reference Online: John Wiley & Sons, Ltd; 2014. [doi: 10.1002/9781118445112.stat06537]
39. Dwyer L. Loneliness by design: the structural logic of isolation in engagement-driven systems. Int J Environ Res Public Health 2025 Sep 6;22(9):1394. [doi: 10.3390/ijerph22091394] [Medline: 41007539]
40. Choi HK, Lee K, Lee SH. Developmental study on "Smart Silver Care": a mobile application to alleviate loneliness in older adults within the community. Healthcare (Basel) 2023 Aug 23;11(17):2376. [doi: 10.3390/healthcare11172376] [Medline: 37685410]
41. Rodríguez-Martínez A, Amezcua-Aguilar T, Cortés-Moreno J, Jiménez-Delgado JJ. Qualitative analysis of conversational chatbots to alleviate loneliness in older adults as a strategy for emotional health. Healthcare 2024;12(1):62. [doi: 10.3390/healthcare12010062]
42. Xing Y, Kelly RM, Rogerson MJ, Waycott J. "Keeping the Program Going": Technology use by community organizations to support the social connectedness of older adults during the COVID-19 pandemic. Proc ACM Hum-Comput Interact 2023;7:1-26. [doi: 10.1145/3610032]

XSL•FO
RenderX

43. Bhattacharya R, Bose D, Rodriguez RV, Siddique KR. Personalized multi-modal interfaces for cognitive aging: a narrative review of design and technological innovations. Arch Gerontol Geriatr Plus 2025 Dec;2(4):100206. [doi: 10.1016/j.aggp.2025.100206]
44. Aouedi O, Sacco A, Piamrat K, Marchetto G. Handling privacy-sensitive medical data with federated learning: challenges and future directions. IEEE J Biomed Health Inform 2023 Feb;27(2):790-803. [doi: 10.1109/JBHI.2022.3185673] [Medline: 35737624]
45. Isravel DP, Arulkumar D, Raimond K, Issac B. A novel framework for quality care in assisting chronically impaired patients with ubiquitous computing and ambient intelligence technologies. In: Dinesh Peter J, Fernandes SL, editors. Systems Simulation and Modeling for Cloud Computing and Big Data Applications 2020:61-79.
46. Alam T, Gupta R. Federated learning and its role in the privacy preservation of IoT devices. Future Internet 2022;14(9):246. [doi: 10.3390/fi14090246]
47. Cho E, Cho H, Demiris G, et al. Older adults' perceptions and attitudes toward passive sensors to measure loneliness: a qualitative study. Sage Open Aging 2025;11:30495334251325607. [doi: 10.1177/30495334251325607] [Medline: 40611868]
48. Zhou Y, Ratcliffe J, Molteni E, et al. Smart textile systems for loneliness monitoring in older people care: a review of sensing and design innovations. Adv Elect Materials 2025 Oct;11(16):e00300. [doi: 10.1002/aelm.202500300]
49. Maslow AH. A theory of human motivation. Psychol Rev 1943;50(4):370-396. [doi: 10.1037/h0054346]
50. Hawkley LC, Cacioppo JT. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. Ann Behav Med 2010 Oct;40(2):218-227. [doi: 10.1007/s12160-010-9210-8] [Medline: 20652462]
51. Torous J, Linardon J, Goldberg SB, et al. The evolving field of digital mental health: current evidence and implementation issues for smartphone apps, generative artificial intelligence, and virtual reality. World Psychiatry 2025 Jun;24(2):156-174. [doi: 10.1002/wps.21299] [Medline: 40371757]
52. Liu X, Mikels JA, Stine-Morrow EAL. The psycholinguistic and affective processing of framed health messages among younger and older adults. J Exp Psychol Appl 2021 Jun;27(2):201-212. [doi: 10.1037/xap0000285] [Medline: 33749299]
53. Malik AS, Acharya S, Humane S. Exploring the impact of security technologies on mental health: a comprehensive review. Cureus 2024 Feb;16(2):e53664. [doi: 10.7759/cureus.53664] [Medline: 38455776]
54. Győrffy Z, Boros J, Döbrössy B, Girasek E. Older adults in the digital health era: insights on the digital health related knowledge, habits and attitudes of the 65 year and older population. BMC Geriatr 2023 Nov 27;23(1):779. [doi: 10.1186/s12877-023-04437-5] [Medline: 38012565]
55. Knotnerus HR, Ngo HTN, Maarsingh OR, van Vugt VA. Understanding older adults' experiences with a digital health platform in general practice: qualitative interview study. JMIR Aging 2024 Aug 30;7(1):e59168. [doi: 10.2196/59168] [Medline: 39212599]

Abbreviations

AHP: analytic hierarchy process
AI: artificial intelligence
CR: consistency ratio
DELONELINESS: Design for Healthy Ageing: a Smart System to Decrease Loneliness for Older People
HCI: human-computer interaction
ICC: intraclass correlation coefficient
TCT: task completion time
TOPSIS: Technique for Order Preference by Similarity to Ideal Solution

Edited by A Kushniruk; submitted 21.Nov.2025; peer-reviewed by HC Chiu, R Singh; revised version received 11.Jan.2026; accepted 09.Feb.2026; published 06.Mar.2026.

Please cite as:
Zhou Y, Rees J, Matcham F, Antonelli M, Ourselin S, Liu W
A User-Centered Interface Design Framework for the DELONELINESS System in Older Adults: Design Indicator Development and Prioritization
JMIR Hum Factors 2026;13:e88263
URL: https://humanfactors.jmir.org/2026/1/e88263
doi:10.2196/88263

© Yi Zhou, Jessica Rees, Faith Matcham, Michela Antonelli, Sebastien Ourselin, Wei Liu. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 6.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and

XSL•FO
RenderX

reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Usability Evaluation of a Macular Quantitative Square Grid Self-Examination Application in Patients With Macular Disease: Mixed Methods Study

Shu Li[1,2*], MD; Enming Zhang[2*], MD; Jiani Pan[3], BMed; Yan Xu[3], BMed; Kairong Zheng[3], BMed; Xun Xu[1], MD; Qiong Fang[2], PhD

[1]National Clinical Research Center for Eye Diseases, Shanghai, China

[2]School of Nursing, Shanghai Jiao Tong University, No. 227 Chongqing South Road, Shanghai, China, Shanghai, China

[3]Department of Nursing, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Qiong Fang, PhD
School of Nursing, Shanghai Jiao Tong University, No. 227 Chongqing South Road, Shanghai, China, Shanghai, China

Abstract

Background: Digital self-monitoring applications could provide individuals with macular disease with a convenient, quantitative method for tracking metamorphopsia at home; yet the usability of such tools remains to be fully established.

Objective: This study evaluated the usability of a macular quantitative square grid self-examination application, a semiquantitative, touch-based self-monitoring application for macular function.

Methods: This study used a convergent mixed methods design. The application quantifies (1) distortion severity, (2) distortion area, and (3) temporal trends through a 3-step touch interface. A total of 24 adults with neovascular age-related macular degeneration or diabetic macular edema, accompanied by self-reported metamorphopsia, participated in a single supervised test session. A 10-item System Usability Scale (SUS) was used to assess usability, and semistructured interviews were conducted to gather further insights. Quantitative data were summarized descriptively, and qualitative feedback underwent inductive thematic analysis.

Results: A total of 24 participants completed the GridMacuScan application self-assessment, the SUS questionnaire, and 11 participants completed the interview when data saturation was achieved. All eyes that showed distortion on the Amsler grid also produced positive distortion maps on the GridMacuScan application, yielding 100% diagnostic concordance. The mean SUS score was 82.1 (SD 8.7), indicating “good-excellent” usability. The inductive thematic analysis yielded four overarching themes: (1) high usability and positive overall experience, (2) perceived functional advantages, (3) shortcomings and optimization suggestions, and (4) strong willingness for continued use.

Conclusions: The GridMacuScan application demonstrated diagnostic sensitivity comparable to that of the traditional Amsler grid and received high user ratings for usability. Furthermore, it provided quantitative distortion metrics that could facilitate longitudinal disease surveillance. Future research must be conducted to validate performance in unsupervised home environments and investigate whether sustained use improves time-to-disease-progression detection and treatment outcomes.

(JMIR Hum Factors 2026;13:e79699) doi:10.2196/79699

KEYWORDS

metamorphopsia; age-related macular degeneration; diabetic macular edema; mobile health; self-monitoring; usability evaluation; tele-ophthalmology

Introduction

Macular diseases such as age-related macular degeneration (AMD) [1] and diabetic macular edema (DME) [2] are among the leading causes of irreversible central vision loss worldwide. A hallmark symptom of these conditions is metamorphopsia, a visual distortion resulting from structural changes in the macula [3]. Metamorphopsia has been identified as a sensitive functional biomarker of disease onset, progression, and therapeutic response [4,5]. In clinical settings, the timely detection of visual changes is particularly critical during intravitreal anti–vascular endothelial growth factor (anti-VEGF) therapy, where dynamic treatment decisions are contingent on subtle shifts in visual function [6].

However, suboptimal adherence remains a widespread challenge. A recent systematic review revealed that approximately 30% of patients discontinue anti-VEGF therapy before the prescribed duration, with dissatisfaction with

XSL•FO
RenderX

treatment outcomes being the most commonly cited reason (29.9%) [7], which highlights an urgent need for tools that enhance patient engagement by making treatment responses more tangible and visible. Currently, changes in visual function are primarily determined by infrequent outpatient examinations, which exhibit limited capacity for continuous patient monitoring between visits [8].

The Amsler grid, a rudimentary paper-based tool, remains the most prevalent approach for domestic monitoring of central vision due to its affordability and accessibility [9]. However, its use in daily life remains limited, potentially due to the heterogeneity of the generated data, the inability to perform quantitative assessments, the complexity of recording processes, and the difficulty in tracking trend changes [10,11], which complicates the accurate comparison of pre- and posttreatment outcomes [12]. A meta-analysis revealed substantial variability in the sensitivity and specificity of the Amsler grid, particularly in the context of AMD screening [10]. Furthermore, traditional grids may present certain challenges in terms of usability for elderly individuals, such as maintaining central fixation and accurately reporting distortions. These limitations underscore the necessity for simplified, user-friendly tools that prioritize accessibility and consistency in real-world domestic environments [13,14].

In China, structured self-monitoring of visual function among patients with retinal disease has not been widely prioritized. Although Amsler grids are occasionally distributed in paper format, especially in community and primary care settings, they are rarely integrated into standardized follow-up workflows or supported with user training [15]. This finding indicates a systemic deficiency in awareness, patient empowerment, and digital tool adoption in the management of chronic retinal disease [15].

In recent years, digital health innovations have introduced app-based tools to support remote self-monitoring of visual function [16]. Although applications such as MacuFix [17], OdySight [18], and Home Vision Monitor [19] have demonstrated efficacy in Europe and the United States, their implementation in China remains constrained by numerous barriers. For instance, the MacuFix application requires users to discriminate among 4 simultaneously presented visual fields to identify subtle distortions. This interface design presents potential cognitive and operational challenges for older adults with visual impairments [17]. In addition, the majority of these applications are deficient in localization of language, user interface design, and integration with domestic health care systems and platforms [20].

To address these challenges, we developed GridMacuScan, an application compatible with tablet devices, designed for structured, self-administered metamorphopsia assessment using a digital square grid. The name “GridMacuScan” originates from the traditional Amsler grid pattern, while incorporating the terms “macula” and “scan” to emphasize its focus on systematically monitoring and quantifying macular degeneration. The GridMacuScan application enables patients to visualize and quantify visual distortion before and after anti-VEGF treatment, providing real-time data to inform clinician decision-making and support patient self-management.

The aim of this study was to evaluate the usability and acceptability of the GridMacuScan application among individuals with macular disease using a mixed methods approach and to characterize user experience, identify potential barriers to engagement, and generate insights to inform future refinement and clinical integration.

Methods

Overview

This study followed the IDEAS (Integrate, Design, Assess, and Share) framework for developing digital health behavior change interventions [21]. According to the IDEAS framework, rigorous evaluations of digital products through large-scale randomized controlled trials should be preceded by small-scale evaluations to test potential efficacy and usability. Specifically, a questionnaire can be used to evaluate usability and satisfaction, and interviews can be conducted to gain insight into the user experiences.

Study Design

This study employed a convergent mixed methods design to evaluate the usability of the GridMacuScan application. Quantitative data were collected through a standardized survey, and qualitative insights were gathered through semistructured interviews, enabling a comprehensive assessment of the user experience. The convergent design involves the parallel collection and analysis of both types of data, subsequently merged to yield a more thorough understanding than that afforded by either method alone. This approach was selected over other mixed methods designs (eg, explanatory, exploratory, or embedded) because it allows for simultaneous triangulation of findings, balancing the strengths and limitations of each data type [22]. The study was conducted in two phases: (1) the development of the GridMacuScan application tailored for patients with macular disease and (2) a usability evaluation combining quantitative assessment and qualitative feedback.

Participation Recruitment and Eligibility Assessment

Participants were recruited from the ophthalmology outpatient department of Shanghai General Hospital between March 4 and March 22, 2025. The inclusion criteria were as follows: (1) aged 18 years or older; (2) diagnosed during outpatient consultation with AMD or DME in at least 1 eye; (3) presence of metamorphopsia, confirmed using the Amsler grid test; (4) best-corrected visual acuity ≥20/200 in the affected eye; and (5) ability to understand study instructions and complete the assessment with minimal assistance. Exclusion criteria included (1) diagnosis of advanced glaucoma, retinal detachment, or other ocular conditions known to cause metamorphopsia or interfere with central visual field assessment; (2) presence of severe cognitive impairment, motor disability, or other systemic diseases that may hinder independent operation; and (3) presence of bilateral macular disease requiring immediate intervention, which could confound the usability evaluation. For the sample size, Bastien [23] cited studies showing that the majority of usability problems can be identified in a sample of 5 to 15

XSL•FO
RenderX

participants. Assuming a 20% dropout rate for the clinical sample, it was calculated that a minimum of 18 participants would need to be invited to participate in the study.

Development of the GridMacuScan Application

The GridMacuScan application was developed as a web-based self-monitoring tool tailored for individuals with macular disease. The name "GridMacuScan" reflects the application's 3 core components: grid-based visualization, macular disease monitoring, and scanning for distortion. The tool was designed to enable structured, semiquantitative self-assessment of metamorphopsia via a square digital grid, providing both patients and clinicians with meaningful visual feedback over time.

This application is composed of 2 primary modules: basic information collection and dynamic disease monitoring. The basic information module is designed to collect user data, including systemic comorbidities, educational background, and other relevant demographic characteristics, which may inform the feasibility of continued follow-up and digital tool use. The dynamic monitoring module is designed to record patient-reported visual distortions, enabling the quantification of metamorphopsia changes before and after intravitreal anti-VEGF treatment.

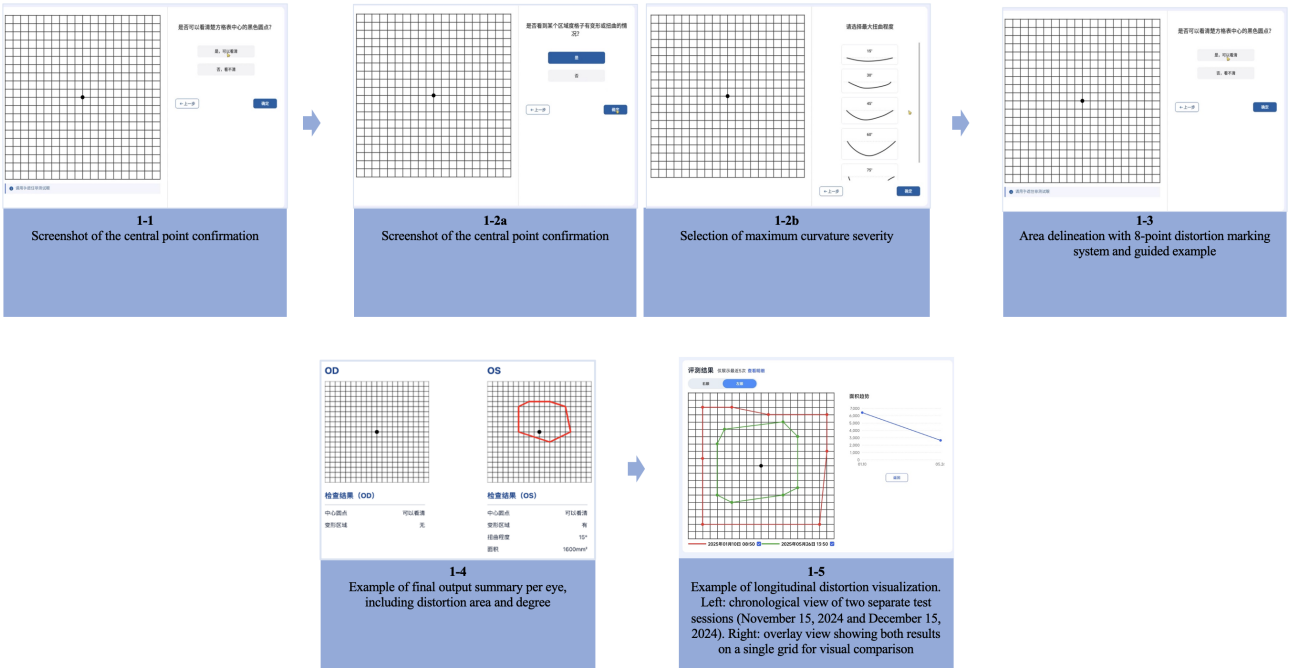
This application also incorporates a backend management platform encompassing 2 core functionalities: data management and disease monitoring. The data module aggregates longitudinal self-assessment results to generate metamorphopsia trend profiles automatically. The monitoring module enables authorized clinical personnel (eg, retinal specialists) or system administrators to review patient-level progression patterns and intervene as deemed necessary.

Self-Assessment Workflow and User Interface Design

The GridMacuScan application's self-assessment workflow comprises the following 3 sequential submodules:

- Step 1: Central point confirmation—Users are prompted to focus on a central black dot displayed within a digital grid and indicate whether it appears clear, blurry, or distorted. This step assesses central fixation stability and initiates the evaluation process (Figure 1).

Figure 1. Self-assessment workflow of the GridMacuScan application.



- Step 2: Severity selection—Users are required to confirm the presence of metamorphopsia (Step 2a in Figure 1). Following this confirmation, they are required to select the severity by choosing from schematic illustrations representing varying degrees of curvature (15° to 75°; Figure 1).
- Step 3: Area delineation—Users must mark the approximate spatial extent of the distortion by selecting grid cells corresponding to the affected area. A rule-based selection strategy instructs users to identify and mark the outermost boundaries of distortion along 4 cardinal directions (superior, inferior, nasal, and temporal), resulting in 8 structured anchor points. A dynamic progress bar (eg, "3/8 points completed") and illustrative prompts were provided to enhance clarity and task completion, based on the standardized 8-point rule, which requires users to mark the outermost distorted grid lines in 4 directions (Step 3 in Figure 1).

Regarding the interface design, the GridMacuScan application prioritizes visual accessibility and ease of use. It is acknowledged that a large number of individuals with vision loss may encounter difficulties in accessing digital technologies. To address this issue, the user interface has been meticulously designed to incorporate large icons, high-contrast color schemes, and low-cognitive-load layouts. These features have been adapted for the use of older adults and individuals with low vision.

XSL•FO
RenderX

Technical Implementation and Data Visualization

From a technical standpoint, the front end was developed using Taro, a cross-platform framework based on HTML and JavaScript, ensuring compatibility across devices. The back end was implemented in Java to enhance data stability and processing efficiency. The tool was developed to be compatible with commonly available Android tablets. It was tested on an 11-inch Redmi Pad SE (Xiaomi Corporation) device designed to prioritize a basic user experience and exceptional value for money, which was selected based on its affordability, screen size, and suitability for domestic use.

In the result visualization module, the GridMacuScan application supports a multilevel display of distortion changes. After each self-assessment, the results are stored in the patient record and visualized as a quantifiable grid. Users can access their historical records and review prior results in either a chronological or layered overlay mode (Step 4 in Figure 1). The system provides date-stamped output with area and curvature comparisons, enabling users and clinicians to observe trends in metamorphopsia severity over time. In the overlay view, changes in the distortion area are visualized through color-coded outlines and anchor points, with the area growth or reduction ratios being calculated automatically to support clinical interpretation and patient engagement (Step 5 in Figure 1).

To preserve the diagnostic logic of conventional tools, the GridMacuScan application adopted a one-to-one digital replication of the traditional Amsler grid layout, as prior studies demonstrated no significant performance differences across grid densities used in electronic Amsler tests [24]. An illustration of the GridMacuScan application output interface is shown in Steps 4 and 5 in Figure 1, which display monocular detection results, including fixation clarity, distortion presence, angular distortion, and area measurement.

Procedures

To ensure the stable operation and implementation of the GridMacuScan application, a backend management team was established. This team comprised 2 research assistants, 2 ophthalmologists, and 1 software engineer. The research assistants were responsible for participant recruitment, daily user management, and backend data organization. Their duties included monitoring the usage status of all registered users, managing login activity, organizing and backing up usage data, and serving as the primary point of contact for participants encountering any operational issues while using the GridMacuScan application. The ophthalmologists conducted a clinical evaluation, with a particular focus on interpreting optical coherence tomography (OCT) images and comparing structural findings with functional outcomes captured by the GridMacuScan application. The software engineer was responsible for maintaining the technical functionality of the application, including backend monitoring of coding or logic errors, resolving system faults promptly, and performing routine updates to ensure long-term operability. The usability evaluation was conducted by the research assistants after the testing period, using both a standardized questionnaire and semistructured interviews to assess user experience and gather feedback for future refinement.

Data Collection

Testing Preparation and Participant Training

Before the initiation of the usability testing phase, the participants' baseline demographic and clinical information was collected, including age, sex, education level, ocular diagnosis, and current use of intravitreal medication. These data were then used to describe the sample characteristics and assess the generalizability of the study findings. Each participant was provided with a standardized user guide and underwent a concise, in-person training session conducted by the research staff. The training program incorporated step-by-step demonstrations using illustrative examples to familiarize users with the functions and operations of the GridMacuScan application. Participants were then invited to complete the full self-assessment procedure independently with minimal supervision.

Quantitative Evaluation of the Usability of the GridMacuScan Application

After completing the test, the System Usability Scale (SUS) is used to quantitatively assess the usability of the GridMacuScan application. All responses were collected anonymously in paper format. SUS is a widely adopted 10-item instrument for assessing the perceived usability of digital health tools [25]. The SUS comprises alternating positive and negative statements, each of which is rated on a 5-point Likert scale ranging from "strongly disagree" to "strongly agree." The individual item scores were transformed and summed by standard procedures and then multiplied by 2.5 to yield a total score ranging from 0 to 100, with higher scores indicating better usability.

Qualitative Assessment of the User Experience With the GridMacuScan Application

A semistructured interview was conducted to complement the quantitative assessment and to gain deeper insights into users' experiences with the GridMacuScan application. Each interview was conducted by a researcher who had undergone training in qualitative methods and lasted approximately 30 minutes. The interview process continued until data saturation was achieved, defined as the point at which no new themes emerged from 2 consecutive interviews. It is important to note that all interviews were audio-recorded and transcribed verbatim for thematic analysis. The interview guide was composed of three key questions: (1) What do you think are the main advantages of this self-monitoring tool? How does it help you manage your condition? (2) What limitations or weaknesses did you find? Do you have any suggestions on how to improve it? (3) Would you be willing to keep using this application to monitor your condition? Why or why not?

Data Analysis

Quantitative Evaluation Data of the Usability of the GridMacuScan Application

Descriptive statistics were used to summarize the participants' sociodemographic and clinical characteristics as well as their SUS scores. Categorical variables, including sex and education level, were reported as frequencies and percentages. Continuous variables, including age and SUS scores, were presented as

XSL•FO
RenderX

means and SDs. The analysis of the quantitative data was conducted using the R Studio software (version 2024.4.2.764; R Foundation for Statistical Computing).

Qualitative Assessment Results of the User Experience With the GridMacuScan Application

Following each interview, the interviewer completed the review of the audio recordings within 24 hours, completed verbatim transcription, and imported the transcripts into NVivo 14.0 (QSR International) for analysis. The transcripts were analyzed systematically using the inductive thematic analysis as described by Braun and Clarke [26]. Two researchers independently reviewed the transcripts multiple times to become thoroughly acquainted with the content. Subsequently, they engaged in open coding to identify significant statements and formulate meanings. The resulting codes were meticulously cross-checked, and any discrepancies were discussed and resolved through a consensus-based approach. The analogous codes were then grouped into subthemes, which were subsequently organized under broader themes that reflected the core dimensions of user experience. A series of illustrative quotations was selected to provide a supporting framework for the themes and subthemes that were identified.

Mixed Methods Integration

A convergent mixed methods approach was used to integrate and interpret the quantitative and qualitative findings [22]. The integration process comprised four distinct steps: (1) analyzing the quantitative data, (2) analyzing the qualitative data, (3) identifying areas of similarity or divergence, and (4) interpreting whether the findings confirmed, expanded, or contradicted each other. Confirmation was established when both data types exhibited analogous usability outcomes. The expansion resulted from the 2 data sources providing divergent yet complementary insights into the user experience. The discordance was identified when the survey and interview findings were inconsistent or contradictory, prompting further interpretive analysis.

Ethical Considerations

The ethical approval for the study protocol was obtained from the Ethics Committees of Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine (2022-KY-021). Before the study began, all participants provided written consent by signing an informed consent form. The participants were told that their participation in the study was voluntary, that they could withdraw at any time, and that all data would be kept strictly confidential and would only be accessible to the researchers. No monetary or material compensation was provided to the participants.

Results

Participants' Characteristics

A total of 24 participants were recruited from the ophthalmology outpatient department of Shanghai General Hospital between March 4 and March 22, 2025. All 24 participants completed the GridMacuScan application self-assessment and the SUS questionnaire. All 24 participants were also required to complete semistructured interviews.

The mean age of the participants was 60.2 (SD 13.9) years. The total number of participants in the study was 24, comprising 13 (54%) men and 11 (46%) women. Regarding education, 13 (54%) participants had attained college or university-level education. The employment status of the participants was evenly distributed, with 12 (50%) participants being employed and 12 (50%) retired. The majority of the participants were diagnosed with AMD (n=14, 58%), while 10 (42%) had DME. The mean number of intravitreal anti-VEGF injections received by participants was 3.5 (SD 1.8), with a median of 3 (IQR 2 - 5). A comprehensive overview of the demographic and clinical characteristics is provided in Table 1.

XSL•FO
RenderX

Table . Demographic characteristics of participants (N=24).

Characteristic	Value
Age (y), mean (SD)	60.2 (13.9)
Gender, n (%)	
Male	13 (54)
Female	11 (46)
Education, n (%)	
Primary school and below	2 (8)
Junior high school	2 (8)
Senior high school	7 (29)
College or university and above	13 (54)
Occupation status, n (%)	
Employed	12 (50)
Retired	12 (50)
Type of disease, n (%)	
AMD[a]	14 (58)
DME[b]	10 (42)
Injection eye, n (%)	
Left eye only	7 (29)
Right eye only	11 (46)
Both eyes	6 (25)
Medication, n (%)	
Aflibercept	3 (12)
Conbercept	12 (50)
Ranibizumab	9 (38)
Number of injections, median (IQR)	3 (2-5)

[a]AMD: age-related macular degeneration.

[b]DME: diabetic macular edema.

Quantitative Evaluation of the Usability of the GridMacuScan Application

Diagnostic Agreement With the Amsler Grid

Before the usability testing, all participants had been confirmed to have metamorphopsia based on traditional Amsler grid screening as per the inclusion criteria. The GridMacuScan application detected visual distortion in all 24 cases, resulting in 100% concordance with the traditional Amsler grid findings for the presence of self-reported metamorphopsia.

SUS Scores

The overall mean SUS score for GridMacuScan application was 82.08 (SD 8.68), indicating “excellent” usability according to established benchmarks (28). Item-level analysis revealed generally positive responses. For illustration, 96% (23/24) of the participants expressed an agreement or strong agreement with the statement “I found the test easy to use” (item 3). Conversely, for negatively framed items, participants indicated high usability. For instance, for the item “I found the test unnecessarily complicated” (item 2), the mean score was 3.50 (SD 0.83) on a scale where lower scores (disagreement) indicate better usability after reverse scoring. A parallel outcome was observed for the statement “I had to learn a lot before using the test” (item 10), for which the mean score was 3.71 (SD 0.46). None of the participants reported that the system was found to be overly complicated (item 8) or that extensive prior learning was required before its usage (item 10). Table 2 presents the mean scores and SDs for all 10 SUS items.

XSL•FO
RenderX

Table . Mean scores and SDs of System Usability Scale (SUS) items.

Item number	SUS item statement	Mean (SD)
1	I can well imagine using the test regularly (P)	3.16 (0.48)
2	I find the test unnecessarily complicated (N)	3.50 (0.83)
3	I find the test easy to use (P)	3.13 (0.45)
4	I think I would need technical support to use the test (N)	2.21 (0.78)
5	I find that the various functions are easy to use (P)	3.46 (0.59)
6	I think there are too many inconsistencies (N)	3.71 (0.46)
7	Most people can learn to use it quickly (P)	3.33 (0.56)
8	I find the operation very complicated (N)	3.38 (0.50)
9	I felt very confident in using the test (P)	3.25 (0.53)
10	I had to learn a lot before using the test (N)	3.71 (0.46)

Task Completion Time

The majority of sessions lasting less than 15 minutes entailed single-eye assessments, whereas those exceeding 20 minutes frequently involved bilateral testing.

Qualitative Assessment of the User Experience With the GridMacuScan Application

Data saturation was achieved after conducting 11 consecutive interviews with participants. The detailed characteristics of the participants are shown in Multimedia Appendix 1.

The thematic analysis of the interview data yielded 4 main themes and 7 subthemes regarding user experience with GridMacuScan. The primary themes that emerged from this analysis were as follows: (1) overall positive experience and ease of use, (2) perceived advantages of GridMacuScan, and (3) shortcomings and optimization suggestions. A detailed summary of themes, subthemes, and illustrative participant quotes is presented in Table S2 in Multimedia Appendix 2.

The participants' reports indicated a predominantly positive experience, with frequent emphasis on highlighting the tool's ease of use. Several key advantages were identified, including the ability to dynamically evaluate and continuously track visual distortion, the convenience of the tool, and an increased sense of confidence in their treatment. The suggestions for improvement primarily focused on developing a WeChat mini-program version to enhance accessibility and optimizing certain module functions, such as providing more detailed feedback on the degree of distortion. The majority of participants also expressed a strong willingness to continue using the GridMacuScan application for self-monitoring.

Mixed Methods Integration

Furthermore, the qualitative findings provided a more in-depth perspective on the specific aspects that contributed to the positive usability scores. For instance, while SUS item 3 ("I find the test easy to use") demonstrated a high score, interview data revealed that features such as a "clear and concise interface" contributed to this perception (Patient 1 quote). Similarly, the mean task completion time was 12.3 (SD 4.96) minutes, indicating the practicality of the proposed application, which was corroborated by qualitative comments on the convenience of the tool.

The quantitative finding that 23 (95.8%) participants would "imagine using the test regularly" (SUS item 1, if this is the interpretation of that item's score from Table 2) was strongly echoed in the qualitative interviews, where a dedicated theme "Hope to continue using" emerged, with participants expressing a clear willingness to integrate the GridMacuScan application into their routine monitoring.

Qualitative feedback also identified areas for future improvement that were not explicitly captured by the SUS, such as the desire for a WeChat mini-program and enhanced analytical features. These suggestions offer valuable insights for future improvements to GridMacuScan, even with its current high usability scores. In general, there was no significant discordance observed between the quantitative and qualitative results. Indeed, the qualitative data effectively served to confirm and expand upon the quantitative usability assessment.

Discussion

Principal Results

The main findings of this study were that the GridMacuScan application demonstrated excellent usability, as evidenced by a high mean SUS score of 82.08 (SD 8.68), and the quantitative result was strongly corroborated by our qualitative findings, where "High usability and positive overall experience" emerged as a dominant theme. The participants frequently described the tool as "simple and intuitive'" and "easy to use." In addition, the tool demonstrated 100% concordance with the traditional Amsler grid in detecting the presence of metamorphopsia and was completed within a feasible timeframe, with a mean completion time of 12.3 (SD 4.96) minutes.

Although traditional self-monitoring tools, such as the Amsler grid, have been employed in macular disease care, digital health innovations offer novel opportunities to overcome the inherent limitations of static, paper-based methods [27]. With the increasing use of anti-VEGF therapies for conditions such as

XSL•FO
RenderX

neovascular AMD and DME, the clinical emphasis has shifted toward long-term monitoring and adherence [28,29]. In this context, the GridMacuScan application signifies a substantial advancement in patient-centered monitoring. Through its intuitive digital interface and standardized 3-step process, it enables the structured, quantifiable, and systematic capture of distortion severity and spatial extent, making it suitable for remote use and serial tracking [6]. The integration of a patient-centered interface with quantitative output effectively addresses the key limitations of paper-based grids, which are often underused in clinical practice due to their noninteractive nature and limited ability to support disease monitoring [17,30].

The design of the GridMacuScan application was guided by the principle of maximizing usability for home-based, semiquantitative monitoring, particularly among older adults. Consequently, we deliberately avoided incorporating features such as eye-tracking or fixation correction, as the added complexity could compromise adoption and ease of use. During design discussions, the tool's visual grid structure and rule-based distortion marking were deemed sufficient for detecting meaningful changes in perceived distortion. This focus on accessible self-monitoring underscores the tool's intended role: it is not meant to replace clinical imaging modalities, such as OCT, but to complement them by supporting regular patient engagement and symptom tracking between visits.

The GridMacuScan application shows significant promise as a tool to augment standard care for patients with macular disease. Its high usability suggests that it could be feasibly integrated into clinical workflows to empower patients with structured self-monitoring, bridge the monitoring gap between clinic appointments, and improve communication and shared decision-making. The backend platform, which aggregates data for clinician review, provides a direct pathway for integrating patient-generated health data into treatment plans.

A primary strength of this study is its rigorous mixed methods convergent design. By integrating quantitative SUS data with rich qualitative insights from interviews, we achieved triangulation of our findings, providing a more robust and comprehensive understanding of the user experience than either method could alone. Furthermore, the study was conducted with the target clinical population—patients with confirmed AMD and DME recruited from an ophthalmology outpatient department—enhancing the real-world relevance of our findings.

However, this study still has several limitations that need to be acknowledged. First, the sample size, although appropriate for identifying key usability issues (N=24), is relatively small and may limit the generalizability of the findings. Second, the participants were recruited from a single, large academic hospital in Shanghai, and a high proportion (n=13, 54%) had a college-level education or higher. This population may not be representative of patients in community or rural settings, who may exhibit different levels of digital literacy. Finally, the study focused on short-term usability rather than long-term utility, which is a key limitation and a crucial aspect that should be considered in future research. Although it was determined that users could operate the tool effectively during an initial encounter, the investigation did not extend to examining whether they would maintain engagement and adhere to regular use over time. Consequently, the ultimate clinical efficacy of the GridMacuScan application, whether its sustained usage results in discernible benefits, such as the earlier detection of disease progression or improved treatment outcomes, remains unmeasured.

To address the study's limitations and build upon its potential, future research should focus on the following key fronts. First, the tool's real-world performance and quantitative outputs must undergo rigorous validation. This entails conducting both unsupervised, home-based studies and longitudinal research to correlate GridMacuScan metrics with objective OCT findings. Second, establishing clinical efficacy is of the utmost importance. A large, multicenter randomized controlled trial is required to rigorously measure long-term adherence and its ultimate impact on clinical outcomes, such as earlier detection of disease progression and changes in treatment frequency. Finally, to ensure long-term values and scalability, implementation research should investigate ways to integrate the GridMacuScan application into existing telemedicine platforms and clinical workflows, thereby ensuring its utility across larger and more diverse patient populations.

Conclusions

This mixed methods study found that the GridMacuScan application is a highly usable and acceptable tool for self-monitoring of metamorphopsia among patients with macular disease in China. It effectively addresses the known limitations of the traditional Amsler grid by providing a quantitative, trackable, and engaging user experience. However, further validation is required to ascertain its long-term adherence and clinical efficacy in a larger trial. The GridMacuScan application is a promising digital health solution with the potential to enhance patient self-management, improve engagement with therapy, and ultimately, support better clinical outcomes in the management of chronic retinal conditions.

Funding

The authors declared no financial support was received for this work.

Data Availability

Due to the protection of personal privacy, the dataset analyzed during the study is not publicly available; however, it can be obtained upon reasonable request from the corresponding author.

XSL•FO
RenderX

Authors' Contributions

All the authors contributed to the successful completion of this research. This study was designed and managed by QF, SL, and XX. SL performed material preparation. JP, KZ, and YX performed data collection. EZ performed data analysis. EZ and SL wrote the first draft of the manuscript. All authors made comments and suggestions on previous editions of the manuscript. The final manuscript was approved after all the authors had read it to their satisfaction.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Demographic characteristics of participants interviewed (N=11).
[DOCX File, 16 KB - humanfactors_v13i1e79699_app1.docx]

Multimedia Appendix 2
Thematic analysis of participant interviews regarding Gridmacuscan application usability (N=11).
[DOCX File, 17 KB - humanfactors_v13i1e79699_app2.docx]

References

1. Thomas CJ, Mirza RG, Gill MK. Age-related macular degeneration. Med Clin North Am 2021 May;105(3):473-491. [doi: 10.1016/j.mcna.2021.01.003] [Medline: 33926642]
2. Zhang J, Zhang J, Zhang C, et al. Diabetic macular edema: current understanding, molecular mechanisms and therapeutic implications. Cells 2022;11(21):3362. [doi: 10.3390/cells11213362]
3. Wiecek E, Lashkari K, Dakin SC, Bex P. A statistical analysis of metamorphopsia in 7106 amsler grids. Ophthalmology 2015 Feb;122(2):431-433. [doi: 10.1016/j.ophtha.2014.09.006] [Medline: 25439598]
4. Wiecek E, Lashkari K, Dakin SC, Bex P. Novel quantitative assessment of metamorphopsia in maculopathy. Invest Ophthalmol Vis Sci 2014 Nov 18;56(1):494-504. [doi: 10.1167/iovs.14-15394] [Medline: 25406293]
5. Hanumunthadu D, Lescrauwaet B, Jaffe M, et al. Clinical update on metamorphopsia: epidemiology, diagnosis and imaging. Curr Eye Res 2021 Dec;46(12):1777-1791. [doi: 10.1080/02713683.2021.1912779] [Medline: 33825600]
6. AREDS2-HOME Study Research Group, Chew EY, Clemons TE, Bressler SB, et al. Randomized trial of a HOME monitoring system for early detection of choroidal neovascularization HOME monitoring of the Eye (HOME) study. Ophthalmology 2014 Feb;121(2):535-544. [doi: 10.1016/j.ophtha.2013.10.027] [Medline: 24211172]
7. Shahzad H, Mahmood S, McGee S, et al. Non-adherence and non-persistence to intravitreal anti-vascular endothelial growth factor (anti-VEGF) therapy: a systematic review and meta-analysis. Syst Rev 2023 Jun 2;12(1):92. [doi: 10.1186/s13643-023-02261-x] [Medline: 37269003]
8. Faes L, Bachmann LM, Sim DA. Home monitoring as a useful extension of modern tele-ophthalmology. Eye (Lond) 2020 Nov;34(11):1950-1953. [doi: 10.1038/s41433-020-0964-3] [Medline: 32405047]
9. Amsler M. L'Examen qualitatif de la fonction maculaire. Ophthalmologica 1947;114(4-5):248-261. [doi: 10.1159/000300476]
10. Faes L, Bodmer NS, Bachmann LM, Thiel MA, Schmid MK. Diagnostic accuracy of the Amsler grid and the preferential hyperacuity perimetry in the screening of patients with age-related macular degeneration: systematic review and meta-analysis. Eye (Lond) 2014 Jul;28(7):788-796. [doi: 10.1038/eye.2014.104] [Medline: 24788016]
11. Bjerager J, Schneider M, Potapenko I, et al. Diagnostic accuracy of the Amsler grid test for detecting neovascular age-related macular degeneration: a systematic review and meta-analysis. JAMA Ophthalmol 2023 Apr 1;141(4):315-323. [doi: 10.1001/jamaophthalmol.2022.6396] [Medline: 36795396]
12. Gross N, Bachmann LM, Islam M, et al. Visual outcomes and treatment adherence of patients with macular pathology using a mobile hyperacuity home-monitoring app: a matched-pair analysis. BMJ Open 2021 Dec 23;11(12):e056940. [doi: 10.1136/bmjopen-2021-056940] [Medline: 34949632]
13. Schuchard RA. Validity and interpretation of Amsler grid reports. Arch Ophthalmol 1993 Jun;111(6):776-780. [doi: 10.1001/archopht.1993.01090060064024] [Medline: 8512478]
14. Liu L, Wang YZ, Bedell HE. Visual-function tests for self-monitoring of age-related macular degeneration. Optom Vis Sci 2014 Aug;91(8):956-965. [doi: 10.1097/OPX.0000000000000324] [Medline: 24987816]
15. Balaskas K, Drawnel F, Khanani AM, Knox PC, Mavromaras G, Wang YZ. Home vision monitoring in patients with maculopathy: current and future options for digital technologies. Eye (Lond) 2023 Oct;37(15):3108-3120. [doi: 10.1038/s41433-023-02479-y] [Medline: 36973405]
16. Daich Varela M, Sanders Villa A, Pontikos N, Crossland MD, Michaelides M. Digital health and wearable devices for retinal disease monitoring. Graefes Arch Clin Exp Ophthalmol 2025 Feb;263(2):279-289. [doi: 10.1007/s00417-024-06634-3] [Medline: 39297890]
17. Claessens D, Ichhpujani P, Singh RB. MacuFix® versus Amsler grid for metamorphopsia categorization for macular diseases. Int Ophthalmol 2022 Jan;42(1):229-238. [doi: 10.1007/s10792-021-02017-3] [Medline: 34420124]

XSL•FO
RenderX

18. Brucker J, Bhatia V, Sahel JA, Girmens JF, Mohand-Saïd S. Odysight: a mobile medical application designed for remote monitoring—a prospective study comparison with standard clinical eye tests. Ophthalmol Ther 2019 Sep;8(3):461-476. [doi: 10.1007/s40123-019-0203-9] [Medline: 31346977]
19. Korot E, Pontikos N, Drawnel FM, et al. Enablers and barriers to deployment of smartphone-based home vision monitoring in clinical practice settings. JAMA Ophthalmol 2022 Feb 1;140(2):153-160. [doi: 10.1001/jamaophthalmol.2021.5269] [Medline: 34913967]
20. Kreienbrinck A, Hanft-Robert S, Forray AI, Nozewu A, Mösko M. Usability of technological tools to overcome language barriers in healthcare- a scoping review. Arch Public Health 2025 Feb 25;83(1):52. [doi: 10.1186/s13690-025-01543-1] [Medline: 40001263]
21. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (integrate, design, assess, and share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016 Dec 16;18(12):e317. [doi: 10.2196/jmir.5927] [Medline: 27986647]
22. Robinson P. Designing and conducting mixed methods research. Aust N Z J Public Health 2007 Aug;31(4):388. [doi: 10.1111/j.1753-6405.2007.00096.x]
23. Bastien JMC. Usability testing: a review of some methodological and technical aspects of the method. Int J Med Inform 2010 Apr;79(4):e18-e23. [doi: 10.1016/j.ijmedinf.2008.12.004] [Medline: 19345139]
24. Augustin AJ, Offermann I, Lutz J, Schmidt-Erfurth U, Tornambe P. Comparison of the original Amsler grid with the modified Amsler grid: result for patients with age-related macular degeneration. Retina 2005 Jun;25(4):443-445. [doi: 10.1097/00006982-200506000-00008] [Medline: 15933590]
25. Lewis JR. The system usability scale: past, present, and future. Int J Hum Comput Interact 2018 Jul 3;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
26. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
27. Marmor MF. A brief history of macular grids: from Thomas Reid to Edvard Munch and Marc Amsler. Surv Ophthalmol 2000;44(4):343-353. [doi: 10.1016/s0039-6257(99)00113-7] [Medline: 10667441]
28. The Diabetic Retinopathy Clinical Research Network. Aflibercept, bevacizumab, or ranibizumab for diabetic macular edema. N Engl J Med 2015 Mar 26;372(13):1193-1203. [doi: 10.1056/NEJMoa1414264]
29. Angermann R, Hofer M, Huber AL, et al. The impact of compliance among patients with diabetic macular oedema treated with intravitreal aflibercept: a 48-month follow-up study. Acta Ophthalmol 2022 Mar;100(2):e546-e552. [doi: 10.1111/aos.14946] [Medline: 34145756]
30. Crossland M, Rubin G. The Amsler chart: absence of evidence is not evidence of absence. Br J Ophthalmol 2007 Mar;91(3):391-393. [doi: 10.1136/bjo.2006.095315] [Medline: 17322466]

Abbreviations

AMD: age-related macular degeneration
anti-VEGF: anti–vascular endothelial growth factor
DME: diabetic macular edema
IDEAS: Integrate, Design, Assess, and Share
OCT: optical coherence tomography
SUS: System Usability Scale

Edited by A Kushniruk; submitted 01.Jul.2025; peer-reviewed by JD Souza; revised version received 15.Feb.2026; accepted 21.Feb.2026; published 18.Mar.2026.

Please cite as:
Li S, Zhang E, Pan J, Xu Y, Zheng K, Xu X, Fang Q
Usability Evaluation of a Macular Quantitative Square Grid Self-Examination Application in Patients With Macular Disease: Mixed Methods Study
JMIR Hum Factors 2026;13:e79699
URL: https://humanfactors.jmir.org/2026/1/e79699
doi:10.2196/79699

© Shu Li, Enming Zhang, Jiani Pan, Yan Xu, Kairong Zheng, Xun Xu, Qiong Fang. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 18.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic

XSL•FO
RenderX

information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluating a Smartphone App to Monitor Blood Pressure in Normotensive Pregnancies, High-Risk Pregnancies, and Women With Preeclampsia: Prospective Longitudinal Feasibility Study

Maria E Andersson[1], PhD; Christine Rubertsson[2], PhD; Elia Psouni[3], PhD; Lena Erlandsson[4], PhD; Camilla Edvinsson[4], MD, PhD; Stefan R Hansson[1], MD, PhD

[1]Department of Obstetrics and Gynecology, Skane University Hospital, Institution of Clinical Sciences L, Lund University, BMC l:12, Lund, Sweden

[2]Department of Health Sciences, Skane University Hospital, Medical Faculty, Lund, Lund University, Lund, Sweden

[3]Department of Psychology, Lund, Lund University, Lund, Sweden

[4]Department of Obstetrics and Gynecology, Institution of Clinical Sciences Lund, Lund University, Lund, Skåne, Sweden

Corresponding Author:
Stefan R Hansson, MD, PhD
Department of Obstetrics and Gynecology, Skane University Hospital, Institution of Clinical Sciences L, Lund University, BMC l:12, Lund, Sweden

Abstract

Background: Antenatal care has been crucial in reducing maternal mortality. Currently, screening programs of pregnant women include blood pressure (BP) measurements, urine protein tests, and the identification of risk factors. Home monitoring can enhance the early detection and management of pregnancy-related hypertension, while also empowering women to take an active role in their own health care.

Objective: This study aimed to evaluate the reliability and accuracy of contactless BP monitoring using the Anura smartphone app and to compare it to conventional manual cuff measurements. This was done in normotensive and high-risk pregnancies, as well as in women diagnosed with preeclampsia. A secondary objective was to assess women's experience using the Anura app.

Methods: Pregnant women with normotensive or high-risk pregnancies were enrolled from pregnancy weeks 8 - 14, and women with preeclampsia were enrolled at the time of diagnosis. The 3 study groups consisted of 132 women with normotensive pregnancies, 40 women with high-risk pregnancies, and 87 women with preeclampsia. They were instructed to use the Anura smartphone app and perform a 30-second facial scan, alongside manual BP measurements, throughout pregnancy. Differences between the 2 methods were analyzed with linear mixed models accounting for repeated measures, reporting beta coefficients with 95% CIs, stratified by patient group and trimester. Outliers were detected visually in the Bland-Altman plots. A digital survey was answered in the Anura app at gestational weeks 37 - 39, about their experiences using the Anura app.

Results: A total of 4932 BP measurements were recorded with Anura, of which 539 had corresponding manual measurements. In normotensive pregnancies, Anura consistently showed slightly higher diastolic values (approximately 5 - 7 mm Hg) and lower systolic values, with significant differences in the second and third trimesters. In high-risk pregnancies, both the systolic and diastolic BP were generally lower with Anura, especially in the second and third trimesters, while women with preeclampsia showed the largest differences, with Anura clearly showing lower systolic and diastolic values. Bland-Altman analyses confirmed these patterns and showed increasing variability and wider limits of agreement in the high-risk pregnancies with preeclampsia. Of 172 women with normotensive and high-risk pregnancies, 56 (32.5%) evaluated their experiences that were predominantly positive, with high perceived safety, better control, and a feeling of increased responsibility for their own health. Some experienced the measurement as somewhat uncomfortable.

Conclusions: The Anura app is well accepted by pregnant women and supported them to take an active role in their own health care. Agreement with manual BP measurements was acceptable in normotensive pregnancies but lower in high-risk and preeclamptic pregnancies. These findings indicate potential for Anura as a complementary self-monitoring tool. Further development is needed to improve the app's accuracy in high-risk groups before broader implementation can be recommended.

(JMIR Hum Factors 2026;13:e70370) doi:10.2196/70370

KEYWORDS

blood pressure; experiences; mobile telephone; preeclampsia; self blood pressure monitoring; pregnancy

XSL•FO
RenderX

Introduction

Antenatal care has been the most important health care intervention when it comes to reducing maternal mortality in modern times [1]. The current national antenatal care program in Sweden includes screening of pregnant women, performing blood pressure (BP) measurements, urine dipstick tests (proteinuria), and obtaining medical history to identify risk factors [2]. Home BP measurements by pregnant women using a validated normal standard BP cuff have shown lower BP levels than measurements at the clinic and a reduced need for antenatal care visits [3,4]. However, many pregnant women find this method to be uncomfortable and cumbersome, which might increase the risk of infrequent use and, consequently, failure to discover significant BP changes associated with such pregnancy risks as preeclampsia. Noncontact and cuffless BP measurement techniques have been studied in different populations with varying results. Studies show both acceptable accuracy and significant margins of error, and reviews point to methodological limitations, particularly in studies of pregnant women [5-8].

Preeclampsia is one of the most common causes of maternal and fetal mortality worldwide and affects 3 - 8% of all pregnant women, corresponding to a total of 8.5 million women annually [9-11]. It is defined as hypertension at 140/90 mmHg after 20 weeks of gestation in combination with maternal organ dysfunction and/or intrauterine growth restriction [2]. Risk factors for preeclampsia, such as diabetes, chronic hypertension, kidney diseases, BMI >30, or African origin, are used as part of risk assessment in the first trimester of pregnancy [12,13] to provide prophylactic treatment with acetylsalicylic acid and reduce the risk of developing preeclampsia. Abnormal calcium and potassium levels have also been shown to be risk factors for pregnancy-related BP [14]. More complex screening algorithms, including biomarkers and Doppler ultrasound, such as the fetal medicine foundation model, have still not been evaluated in Sweden but recently in Denmark [15]. The fetal medicine foundation model, which combines maternal history factors with measurements of mean arterial pressure, uterine artery pulsatility index, and serum placental growth factor levels [16], is well validated but complex, requiring specially trained personnel to perform the Doppler ultrasound. Thus, it is essential to develop a more user-friendly method, available for all pregnant women for early identification of hypertension. A more straightforward and accessible BP monitoring solution would allow the women to be more involved in self-care and their health, both during pregnancy and postpartum.

Apps using smartphones may be a solution for monitoring BP in a user-friendly way. Anura is a smartphone app based on a novel imaging methodology called transdermal optical imaging (TOI) and capitalizes on the fact that light can travel beneath the skin and be reflected due to the skin's translucent properties [17]. The reflected light can then be captured by an optical sensor [18]. TOI uses video-captured images and machine-learning algorithms to extract blood flow information from the facial epidermis. The Anura app can make these measurements in 30 seconds by simply asking the patient to take a 30-second video selfie (Figure 1). Thus, the app may remove the need for cuffs or other technical equipment [19]. Based on a large video dataset from normotensive and hypertensive patients, along with their physiological measurements based on gold-standard medical devices, TOI uses machine-learning algorithms to build computational models to predict a variety of vital signs [20,21].

Figure 1. The user interface of the Anura app.



TOI technology has been shown to accurately determine BP in nonpregnant normotensive participants with a precision comparable to clinical standards using manual measurements [19]. The systolic and diastolic BP predicted from TOI resulted in measurements within 5±8 mm Hg of the reference measurements [22]. However, its performance in pregnant women remains unexplored, and the generalizability to diverse populations has not yet been established. To address this gap, the objective of this study was to evaluate the feasibility of the

XSL•FO
RenderX

Anura app as a method for antenatal BP surveillance of pregnant women.

The primary aim of this study was to evaluate the reliability and accuracy of the Anura app for BP measurement during pregnancy in normotensive pregnancies, high-risk pregnancies, and in women diagnosed with preeclampsia, compared to conventional manual cuff measurements. In addition to the BP measurements, we also evaluated women's experience of using the Anura app by a digital questionnaire.

Methods

Ethical Considerations

The study was conducted with ethical approval from the Swedish Ethical Review Board in Lund, Sweden (DNR 2021 - 03216), and performed in line with the principles of the Declaration of Helsinki. Informed consent was obtained from all participants, who signed consent forms prior to inclusion, and no compensation was provided for participation. Participant confidentiality was strictly maintained. All data were deidentified prior to analysis, and no information that could enable identification of individual participants is presented. Data were stored securely in accordance with institutional and national data protection regulations, with access restricted to the research team.

Design, Setting, and Participants

A total of 3 groups of pregnant women were recruited: those with normotensive pregnancies, those with high-risk pregnancies, and women diagnosed with preeclampsia. High-risk pregnancies were defined according to the Swedish Society of Obstetrics and Gynecology guidelines for hypertensive disorders in pregnancy [2]. Participants were classified as high-risk if they had at least one high-risk factor (eg, chronic hypertension, type 2 diabetes) or 3 or more moderate-risk factors (eg, advanced maternal age, nulliparity, high BMI). Preeclampsia was defined as hypertension (140/90 mmHg) after 20 weeks of gestation in combination with maternal organ dysfunction and/or intrauterine growth restriction [23,24]. The hemolysis, elevated liver enzymes, low platelets syndrome is a severe form of preeclampsia with general organ involvement, especially of the liver. Women diagnosed with preeclampsia were recruited when they were admitted to the hospital. The study was part of a prospective longitudinal feasibility study aiming to follow pregnant women with and without high-risk factors from early pregnancy until 6 - 8 weeks postpartum (Multimedia Appendix 1).

The study was conducted in southern Sweden at 2 large hospital maternity health care units and 9 different antenatal health clinics (AHCs) in the region, from March 2022 to December 2023. The inclusion criteria were: (1) pregnant normotensive women, high-risk pregnancies, and women diagnosed with preeclampsia, (2) having a smartphone that could download the Anura app, (3) 18 years or older, and (4) understanding Swedish. All pregnant women enrolled in AHC during the study period were screened for eligibility and invited to participate. In total, 331 women were approached, of whom 288 agreed to participate. A total of 43 women declined participation, most commonly citing reasons such as mental illness, anxiety, concerns about having high BP, or difficulties using the Anura app. The eligible women received both oral and written information and had the opportunity to consider whether they wished to participate. They were also informed to contact their midwife or the maternal health unit if they experienced symptoms such as headache, general malaise, swelling, chest pain, or if their BP was 140/90 mm Hg or higher.

The included women were asked to use the Anura app to perform facial scans for BP measurements at each AHC or hospital visits. At the visits, the BP was also measured by midwives or health care professionals using a routinely applied manual sphygmomanometer cuff method.

The pregnancies were divided into trimesters, defined as: first trimester (0 - 13+6 gestational week [GW])=0 - 93 days, second trimester (14+0 - 27+6 GW)=94 - 187 days, and the third trimester (28+1 - 42+0 GW)=188 - 283 days.

Data Collection

Demographic data for the included women (Tables 1 and 2) was collected from the medical record system Obstetrics (version 2.18.0.100, Copyright Cerner Sverige AB) [25]. In addition to the standard antenatal program [1], the women were asked to follow a scheme for BP measurements using the Anura app. The midwives or the first author of the study provided information on downloading the Anura app to their private smartphones. Basic personal information (weight, height, age, high BP, or diabetes) was entered directly into the Anura app by the women.

XSL•FO
RenderX

Table . Demographic characteristics of included women.

Characteristic and risk factors for preeclampsia	Normotensive pregnancy, n=132	High-risk pregnancy, n=40	Preeclampsia, n=98
ASA[a] prophylaxis, n (%)	0 (0)	35 (88)	38 (39)
High-risk factors, n (%)			
Autoimmune disease	0 (0)	1 (3)	0 (0)
IUGR[b], IUFD[c], placental abruption	1 (0.8)	4 (10)	6 (6)
Diabetes	0 (0)	0 (0)	6 (6)
Kidney disease	0 (0)	0 (0)	1 (1)
Chronic hypertension	0 (0)	5 (13)	13 (13)
IVF[d] with egg donation	0 (0)	5 (13)	3 (3)
Previous GH[e] delivery before wk 34	0 (0)	1 (3)	0
Moderate risk factors			
Maternal age (y), min-max; mean (SD)	19 - 43; 31.8 (3.9)	19 - 43; 35 (5.2)[f]	18 - 45; 32.2 (4.8)[f]
BMI at booking visit, min-max; mean (SD)	19 - 37; 24.4 (3.7)	21 - 42; 29.3 (5.3)[f]	17 - 61; 27.4 (6.9)[f]
African origin, n (%)	0 (0)	2 (5)	5 (5)
Nulliparity, n (%)	56 (42.7)	8 (20)	56 (57)
Family history of high BP[g], n (%)	38 (29)	19 (48)	32 (33)
Pregnancy interval >10 y, n (%)	2 (1.5)	0 (0)	1 (1)
White coat of hypertension, n (%)	3 (2.3)	0 (0)	1 (1)
Family history of PE, n (%)	6 (4.6)	9 (23)	12 (12)
Previous PE, n (%)	0 (0)	13 (33)	18 (18)
Multiple birth, n (%)	0 (0)	0 (0)	6 (6)
Other risk factors, n (%)			
Mental illness	19 (14.5)	7 (18)	20 (20)

[a]ASA: acetylsalicylic acid.

[b]IUGR: intrauterine growth restriction.

[c]IUFD: intrauterine fetal death.

[d] IVF: in vitro fertilization.

[e]GH: gestational hypertension.

[f]P<.001.

[g]BP: blood pressure.

XSL•FO
RenderX

Table . Characteristics of included women and their newborn children

Characteristic	Normotensive pregnancy, n=132	High-risk pregnancy, n=40	Preeclampsia, n=98
Regulated BP[a] postpartum, n (%)	1 (0.8)	11 (28)	72 (73)
Medication for hypertension at discharge, n (%)	1 (0.8)	9 (23)	65 (66)
SBP[b], min and max; mean (SD)	100 - 150; 122 (9)	120 - 180; 135 (12)[c]	130-190; 156 (13)[c]
DBP[d], min and max; mean (SD)	60 - 95; 76 (8)	70 - 110; 87 (9)[c]	80 - 120; 98 (7)[c]
Preterm birth <wk 37, n (%)	1 (0.8)	1 (3)	36 (37)
Birth <wk 34, n (%)	0 (0)	0 (0)	21 (21)
Gestational weeks of delivery, min-max; mean (SD)	28 - 42; 39.4 (1.6)	35 - 41; 38.9 (1.3)[c]	26 - 40; 38.81 (2.9)[c]
Vaginal birth, n (%)	101 (77.1)	26 (65)	46 (47)
Induction of labor, n (%)	36 (27,5)	18 (45)	56 (56)
Vacuum extraction, n (%)	4 (3.1)	3 (8)	4 (4)
Cesarian section, n (%)	24 (18.3)	9 (23)	51 (52)
Oxytocin augmentation, n (%)	35 (26.7)	9 (23)	16 (16)
Hemorrhage (ml), min-max; mean (SD)	100 - 2100; 510 (344)	100 - 3000; 568 (559)	100 - 2200; 548 (429)
Twins, n (%)	0 (0)	0 (0)	9 (9)
SGA[e], n (%)	1 (0.8)	0 (0)	27 (28)
Days in PP hospital care, min-max; mean (SD)	1 - 7; 2.8 (1.3)	1 - 22; 3.9 (3.6)[c]	2 - 44; 9.8 (6.7)[c]
NICU[f] admission, n (%)	11 (8.4)	2 (5)	36 (37)
Newborn weight (g), min and max; mean (SD)	1900 - 4655; 3616 (477)	2550 - 4438; 3480 (467)	630 - 3830; 2666 (792)[c]
Newborn sex, n (%)			
Boy	60 (46)	18 (45)	49 (50)
Girl	71 (54)	18 (45)	49 (50)

[a]BP: blood pressure.

[b]SBP: systolic blood pressure.

[c]*P*<.001.

[d]DBP: diastolic blood pressure.

[e]SGA: small for gestational age.

[f]NICU: neonatal intensive care unit.

Once the app was downloaded and registered, each woman was assigned a participant code. The code key was stored under lock and key. The BPs were measured in 2 ways: by the midwife using a standard BP cuff or a validated automatic BP monitor, and by the women using the Anura app. The results were automatically stored in the Anura app. The participants manually entered the diastolic blood pressure (DBP) and systolic blood pressure (SBP) measured by the midwife into the Anura app.

In addition, the women were encouraged to measure their BP at home at least once a week, after 15 minutes of rest. For each Anura measurement, the women scanned their faces for 30 seconds at 30 frames/second imaging. The app processed the image frames and used its local algorithms to extract blood flow information. Only the extracted facial blood flow information was packaged, encrypted, and sent to the cloud server in Europe owned by NuraLogix Corporation (Toronto, Canada). The data was processed by the DeepAffex (DFX) Artificial Intelligence Engine [20]. The Anura is in full compliance with the General Data Protection Regulation. In addition, a written agreement between Lund University and NuraLogix was signed to ensure data safety, privacy, and the freedom of data use by the research team.

The women also answered a survey at 37 - 39 GW. The experience of using the Anura app was investigated with a survey consisting of 8 questions. The questions concerned their perceived privacy, sense of control, security, and user-friendliness. The answer options were categorical and reported as percentages for each question.

XSL•FO
RenderX

Data Analysis

In order to determine the number of women to be included in each group, a power calculation was performed using G*power [26]. To achieve adequate power of 80% with a significance level of 0.05 and a medium effect size (f=0.25), a total sample of 269 was required with 90 for each of the 3 groups (normotensive, high-risk, and preeclampsia). Data from Anura was downloaded directly from a dashboard, minimizing the risk of missing data. The clinical data were manually extracted from medical records and double-checked to ensure completeness. All analyses were performed using International Business Machines Corporation SPSS Statistics for Windows, Version 28.0. A P value <.05 was considered statistically significant. Demographic data were analyzed with one-way ANOVA. Bland–Altman analysis, including limits of agreement (LoA), was conducted in Microsoft Excel. All data points from the Anura app were downloaded from the NuraLogix DFX dashboard. For continuous variables, descriptive statistics were presented as minimum, maximum, and mean (SD). Categorical variables were presented in numbers (n) and proportions (%). All calculations for BP were performed separately for SBP and DBP values. The analyses were stratified by patient group and trimester. The data were normally distributed. To account for repeated measures, the 2 BP measuring methods, Anura versus manual BP, were compared using linear mixed models, fitted in R 4.3.3 using package linear and nonlinear mixed effects model version 3.1 - 167. Measuring methods were defined as fixed effect and pair of measurements within person as random effect using an AR1 correlation structure. The difference in measurements is presented as the beta coefficient with 95% CI. Mixed model analyses were stratified by patient group and trimester, and assumptions were assessed graphically.

Bland-Altman plots were used to show outliers and visual comparisons of the Anura measurements against the manual cuff measurements. The Bland-Altman plot, which showed systematic differences and dispersion, did not contain any CIs but instead the 95% LoA, which were calculated assuming normal distribution. All available paired Anura BP measurements were included in the analyses. For participant questionnaires, only complete responses were included. No imputation was applied for missing data.

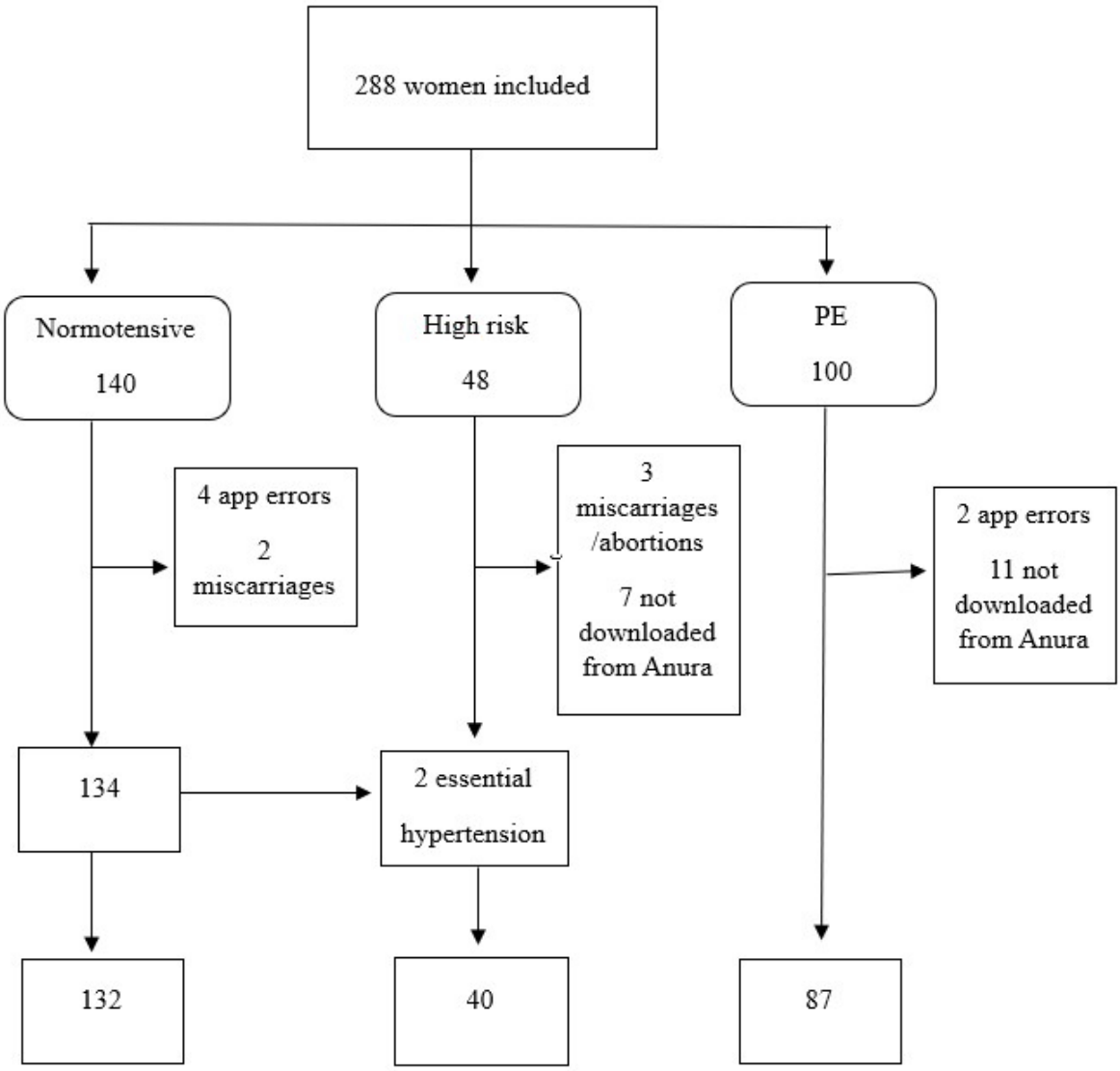
Results

Demographics

In total, 288 women were recruited for participation, but 10.1% (29/288) were excluded due to miscarriage, abortion, or technical problems with the Anura app (Figure 2). The remaining 259 women were divided into 3 groups as follows: 132 were classified as women with normotensive pregnancy, 40 were classified as those with high-risk pregnancies, and 87 were diagnosed with preeclampsia (Figure 2). Data were successfully downloaded from the NuraLogix DFX dashboard. Data could not be downloaded for 11 women with preeclampsia, but their demographic background data are still included in the preeclampsia group shown in Tables 1 and 2, resulting in 98 for this group (Figure 2).

XSL•FO
RenderX

Figure 2. Flowchart for included women in the 3 study groups. PE: preeclampsia.



Demographic data showed that the normotensive pregnancy group was significantly different ($P<.001$) from both the high-risk group and the preeclampsia group regarding pregnancy length, age, height, weight, BMI, DBP, and SBP (Tables 1 and 2). The normotensive pregnancy group had significantly lower ($P<.001$) maternal age and BMI compared to the high-risk and preeclampsia groups. Nearly half (19/40, 47.5%) of the women in the high-risk group had a family history of high BP. Additionally, most women in the high-risk pregnancy group received prophylactic acetylsalicylic acid treatment (35/40, 87.5%).

In the normotensive group, 2/140 women developed gestational hypertensions and were therefore transferred to the high-risk group (Figure 2). Of the 140 women in the normotensive group, 3 women developed preeclampsia some weeks before childbirth and were therefore transferred to the preeclampsia group. Diseases that were identified in the normotensive group were 1 case of asthma, 1 case of Crohn disease, 1 case of epilepsy, 1 case of gestational diabetes, 6 cases of hyperthyroidism, and 1 case of polycystic ovary syndrome. The fetal sex of 2 newborns remains unknown since they were born outside of the study region. Characteristics for the included women and their newborns are shown in Table 2.

In the high-risk group (n=40), 7 women developed preeclampsia, and 6 developed gestational hypertension. Eight had essential hypertension, 1 had borderline high BP, 1 had systemic lupus erythematosus or rheumatoid arthritis, and 2 had a previous history of postpartum depression. In addition, 27.5% had a regulated BP with medication postpartum, and 22.5% were still on antihypertensive medication when discharged. Women in the high-risk group also had significantly ($P<.001$) longer hospital stays.

In the preeclampsia group (n=98), 12 women had a known family history of preeclampsia and 6 had essential hypertension. Additionally, 1 woman developed hemolysis, elevated liver

XSL•FO
RenderX

enzymes, and low platelets syndrome. In the preeclampsia group, 73.5% (72/98) had a regulated BP using medication postpartum and 66.3% (65/98) were still on antihypertensive medication when discharged from the maternity unit. Women in the preeclampsia group also had significantly (P<.001) longer hospital stays, the birthweights were significantly lower, and 36.7% (36/98) of their newborns were admitted to the neonatal intensive care unit.

Blood Pressure Measured With the Anura App Versus Manual Measurements

In total, there were 4932 BP measurements registered in the Anura app in the 3 study groups: 2993 normotensive pregnancies, 853 high-risk pregnancies, and 1072 pregnancies with preeclampsia. Of these measurements, 539 had a corresponding manual BP cuff measurement taken at an antenatal care visit (normotensive pregnancies, n=194; high-risk pregnancies, n=108; and pregnancies with preeclampsia, n=237). Manual BP of the 3 groups and during the trimesters is shown in Table 3. For the preeclampsia group, there were few measurements in the second trimester, since most preeclampsia did not develop until the third trimester.

Table . SBP[a] and DBP[b] for unpaired manual measurements for each trimester.

	Manual		
Blood pressure characteristic	Normotensive pregnancy	High-risk pregnancy	Preeclampsia
Trimester 1	n=53	n=14	—[c]
SBP, mean (SD; min-max)	113 (8.9; 97 - 132)	120 (11.3; 88 - 135)	—
DBP, mean (SD; min-max)	69 (6.1; 54 - 80)	78 (12.1; 64 - 114)	—
Trimester 2	n=57	n=37	n=4
SBP, mean (SD; min-max)	115 (9.6; 97 - 136)	121 (6.4; 109 - 135)	156 (13.8; 140 - 170)
DBP, mean (SD; min-max)	70 (5.6; 60 - 83)	80 (4.8; 70 - 92)	106 (9.7; 99 - 120)
Trimester 3	n=74	n=50	n=185
SBP, mean (SD; min-max)	115 (9.1; 100 - 134)	121 (10.2; 100 - 149)	138 (10.7; 110 - 180)
DBP, mean (SD; min-max)	70 (6.8; 60 - 88)	77 (10.4; 60 - 97)	87 (84; 63 - 113)

[a]SBP: systolic blood pressure.

[b]DBP: diastolic blood pressure.

[c]Not applicable.

Comparisons between Anura and manual BP measurements are presented in Table 4. In the normotensive group, Anura consistently measured higher DBP across all trimesters (β ranging from 4.9 to 6.9 mm Hg; P<.001), while SBP tended to be lower, reaching statistical significance in the second and third trimesters. In the high-risk group, DBP was lower in the second trimester (β –3.5 mm Hg; P=.004), whereas SBP was significantly lower in both the second and third trimesters (β –8.4 and –9.1 mm Hg; P<.001). Among women with preeclampsia, Anura showed markedly lower values, with a mean difference of –6.8 mm Hg for DBP and –16.9 mm Hg for SBP (both P<.001).

XSL•FO
RenderX

Table . Comparison Anura—manual blood pressure measurement linear mixed effects model.

Patient type and trimester	Measurements	Individuals	DBP[a] β (95% CI)	*P* value	SBP[b] β (95% CI)	*P* value
Normotensive						
First	53	46	4.9 (2.8 to 7)	<.001	−2.6 (−5.6 to 0.4)	.09
Second	57	39	5.9 (4.3 to 7.5)	<.001	−3.7 (−6.6 to −0.8)	.01
Third	75	39	6.9 (4.9 to 9)	<.001	−3.2 (−5.7 to −0.7)	.01
High risk						
First	14	13	3 (−2.7 to 8.8)	.28	−2.7 (−7.3 to 1.9)	.23
Second	37	11	−3.5 (−5.9 to −1.2)	.004	−8.4 (−11.7 to −5.1)	<.001
Third	50	9	0.4 (−3.1 to 3.9)	.82	−9.1 (−12.4 to −5.9)	<.001
Preeclampsia						
Second	4	1	__[c]	—	—	—
Third	184	38	-6.8 (−8.2 to −5.4)	<.001	−16.9 (−18.7 to 15.1)	<.001

[a]DBP: diastolic blood pressure.

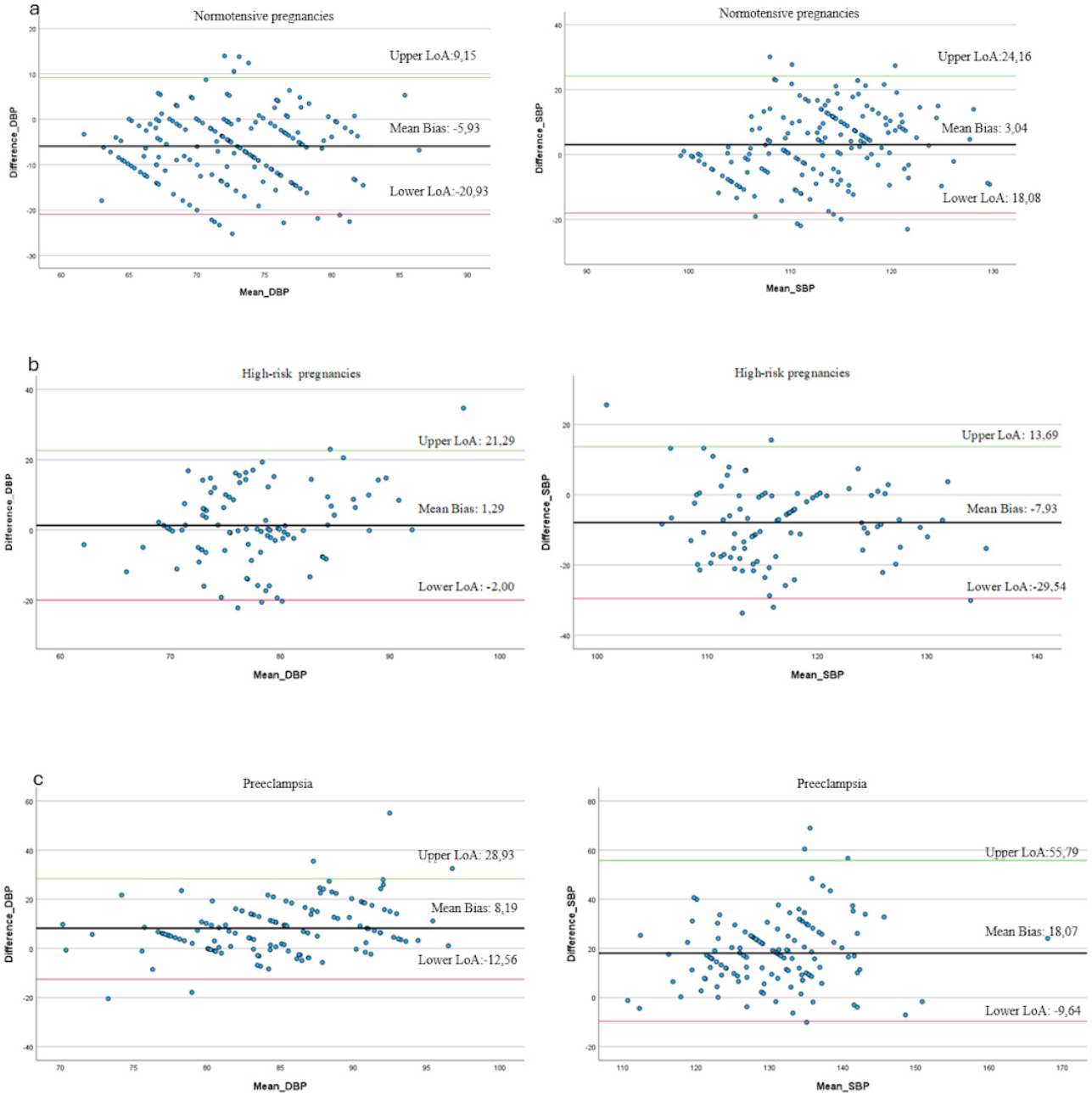
[b]SBP: systolic blood pressure.

[c]Not available.

Bland-Altman plots (Figure 3) were created to illustrate differences between the paired BP measurements. The Bland-Altman analysis for SBP in the normotensive group (Figure 3a) shows a mean difference of 3.04 (SD 7.7) and for DBP a mean difference of −5.93 (SD 7.7). In the high-risk group (Figure 3b), the SBP shows a mean difference of −7.93 (SD 11.0) and for DBP a mean difference of 1.29 (SD 10.8). For the preeclampsia group (Figure 3c), the SBP shows a mean difference of 18.07 (SD 18.1) and for DBP a mean difference of 8.19 (SD 10.6). There was a variation in the Bland-Altman plots of the upper and lower LoA in both the SBP and the DBP in the 3 groups (Figure 3).

XSL•FO
RenderX

Figure 3. Bland-Altman plot differences in DBP and SBP for manually and Anura measurements in (a) normotensive pregnancies, (b) high-risk pregnancies, and (c) preeclampsia. DBP: diastolic blood pressure; LoA: limits of agreement; SBP: systolic blood pressure.



The Experiences of Using Anura

The results from the survey are based on 56/172 (32.3%) women who evaluated their experience of using Anura (Multimedia Appendix 2). Most of the women (48/56, 85.7%) expressed no concern regarding their privacy while using the app, whereas 8/56 (14.3%) women reported being a little worried. Almost all women (55/56, 98.2%) reported no concern about seeing their BP results, with only 1/56 (1.8%) woman indicating slight apprehension. Regarding responsibility for their own health, 31/56 (55.4%) women felt a higher degree of responsibility when using the app. Of these, 2/56 (3.6%) women reported feeling much more responsible, 7/56 (13%) women more responsible, and 22/56 (39.3%) women a little more responsible. In contrast, 25 (44.6%) women reported no increase in responsibility. Perceived safety when using the app was also high: 51/56 (91.1%) women considered it safe, of whom 30/56 (56.6%) rated it as very safe, 21/56 (37.5%) as safe enough, and 5/56 (8.9%) women as only a little safe. Most of the women (48/56, 85.7%) experienced better control of their health with the app. Specifically, 8/56 (14.3%) women reported much better control, 18/56 (32.1%) women better control, and 22/56 (39.3%) women slightly better control. A total of 8/56 (14.2%) women reported no improvement. Similarly, 32/56 (57.1%) women reported an increased understanding of their health status, with 5/56 (8.9%) women indicating much better understanding, 17/56 (30.4%) women better understanding, and 10/56 (17.9%) women a little better understanding. A total of 2/56 (3.6%) women reported no better understanding, while 5/56 (8.9%) women remained neutral. Most women (42/56, 75%) reported that sitting still and looking into the camera during the measurement went well. However, 12/56 (21.4%) women found it unpleasant or

XSL•FO
RenderX

uncomfortable, 1/56 (1.8%) woman reported it as quite unpleasant, and 11/56 (19.6%) women were neutral. The duration of the measurement was generally well accepted: 32/56 (57.1%) women considered the length just right, 22/56 (39.3%) women found it okay, and only 2/56 (3.8%) women felt it was too long.

Discussion

Principal Findings

The primary aim of our study was to evaluate how reliable and accurate the Anura app is for BP measurement during pregnancy, compared to manual cuffs for BP measurement. In addition, we asked women about their experiences of using the app. To the best of our knowledge, this study is the first to validate the Anura technology in pregnant women, including high-risk pregnancies and women diagnosed with preeclampsia, with analyses stratified across 3 trimesters. The use of smartphones for medical monitoring has previously indicated a need for a user-friendly, independent BP monitoring method during pregnancy to enable identification of hypertension. Home monitoring can enhance early detection [3,4] and management of pregnancy-related hypertension, while also empowering women to take an active role in their own health care [27]. This fact was also confirmed in this study, as most women reported a positive experience of using the Anura app and feelings of being in control of their own health.

The main finding in our study suggests that Anura agrees well with manual BP in normotensive women; however, discrepancies between the 2 methods increase in the high-risk and preeclampsia groups, especially for SBP. Among normotensive women, Anura generally showed a higher DBP (4.9 - 6.9 mm Hg) compared to manual measurements, while SBP tended to be slightly lower, with statistical significance in the second and third trimesters. This suggests that the use of Anura BP measurement may be more reliable in women with normotensive pregnancies, while caution is required in high-risk pregnancies and once preeclampsia is diagnosed. The Bland-Altman analysis confirmed this finding, showing that the normotensive group exhibited small differences between the 2 methods, indicating that Anura has good agreement with manual measurements in normotensive women. Our findings in normotensive pregnant women are in line with previous research with male patients and nonpregnant female patients [22]. Regarding the high-risk pregnancies, the analysis showed that DBP was lower in the second trimester (–3.5 mm Hg) and SBP was lower in both the second and third trimesters (–8.4 and –9.1 mm Hg), respectively. These results show that in high-risk pregnancies, the Anura app may be less reliable in the last part of pregnancy compared to earlier stages of pregnancy. The preeclampsia group demonstrated the most pronounced differences, –6.8 mm Hg for DBP and –16.9 mm Hg for SBP. Thus, while the use of Anura BP measurement may be reliable in women with normotensive pregnancies, our results do not support using the Anura app for BP screening in women with high-risk pregnancies or women diagnosed with preeclampsia, suggesting caution in high-risk pregnancies and once preeclampsia is diagnosed. These findings are further corroborated by another study showing that BP measurements were not accurate in a perioperative setting using a similar technique, video plethysmography for contactless measurement [7]. However, Anura has been reported to be highly accurate in predicting the respiratory rate and heart rate of surgical patients in a clinical setting [7,8].

The differences shown in the Bland-Altman analysis suggest that in some cases, the Anura technology either overestimated or underestimated the BP compared to the manual measurements, especially in the preeclampsia group where the deviations were the largest. These deviations are important to consider during further processing of the variables to improve the measurements in the Anura app in the future. The high variation of the upper and lower LoA in the Bland-Altman plots may indicate that the Anura app might be less reliable at extreme BP values, or that its performance varies depending on the context or conditions of the measurements [28]. One possible explanation is that women with high-risk pregnancies and preeclampsia may have more edema in their faces, which could interfere with the TOI technique that fails to make accurate BP measurements. A physiological explanation for this inconsistency in BP may lie in the cardiovascular adaptations that occur during pregnancy. Normally, the blood volume increases during the first 2 trimesters, and the peripheral vascular resistance decreases [29], leading to a 10 - 20 mm Hg reduction in BP. It reaches its lowest point at 18 - 20 weeks of gestation and then returns to prepregnancy levels by the third trimester [30,31]. In preeclampsia, however, there is a maintained vascular resistance, as well as endothelial damage, that results in general edema. Increased edema and hemo-concentrated blood might impact the measurement quality when BP is measured using TOI. Our results indicate that the Anura measurements tended to show lower values compared to manual BP measurements for both DBP and SBP in the preeclampsia group, and to some extent also in the high-risk group. In preeclampsia, the vascular endothelial cells undergo changes that result in inflammation, impaired blood flow regulation, and increased vascular permeability [10,32]. While most pregnant women experience swelling during their pregnancy, this predominantly does not occur until the third trimester in normotensive pregnancies [33].

In previous studies, BP was determined in nonpregnant individuals by Anura with high accuracy [19,22]. In this study, the technology was evaluated under controlled conditions with optimal lighting, a tripod-mounted mobile phone, and using the same device for all measurements. In contrast, this study allowed the women to use their own mobile phones, holding them manually, and having varying lighting conditions at home. Movement and lighting often affect the accuracy of contactless measurement technologies using a mobile phone [21,34]; yet, while knowing their role for accurate BP measurement, we could neither entirely control nor account for them in this study. The women were instructed to rest in a sitting position and hold their mobile phones still at arm's length in front of their faces during the 30 seconds needed for measurements. They were to measure their BP only when at least 3 stars appeared in the Anura app. One star indicates poor lighting conditions, while 5 stars indicate perfect lighting conditions [35]. Any movements of the arm and varying light qualities in the rooms at AHC and

XSL•FO
RenderX

at home could, of course, affect the measurements. This may apply particularly to women admitted to the hospital with preeclampsia, who often had to dim the lights in their rooms due to symptoms and illness related to preeclampsia. Such suboptimal conditions may have impacted the performance of the Anura measurements. However, our results contribute to the advancement of contactless BP monitoring by highlighting the challenges of implementing this technology. These insights can be used to further improve the handling of the smartphone, lighting conditions, and the algorithm in the future.

The use of smartphones for medical monitoring has increased in recent years. As information technology continues to develop, so do the opportunities to find new ways to measure BP with smartphone-based apps [36]. Compared to traditional monitoring, home BP monitoring during hypertensive pregnancies appears to be cost-saving without compromising the safety of the pregnant mother [37]. It increases accessibility and convenience for pregnant women, has the potential to decrease the number of hospital visits [3], enables the early detection of abnormalities, and promotes better control and self-management of their health. It has also been shown that BP measured at home often is lower than in a clinical setting [38]. In addition, home measurements have shown reduced stress and anxiety and enable more individualized care, which in turn can lead to better health outcomes for the women and their children. In a systematic review and meta-analysis [39], it was demonstrated that contactless monitoring technology using consumer-friendly cameras, such as smartphones, is accurate for measuring heart rate compared to other medical devices. However, more studies are needed to assess the accuracy of contactless BP measurements, particularly among pregnant women. In another review, it was also highlighted that no apps for BP had undergone sufficient testing to be recommended in a clinical practice [40]. In a recently published study, the accuracy of contactless monitoring technology for heart rate suggested limitations in accurately measuring BP in a hospital setting. However, the study showed that contactless monitoring technology was both accurate and feasible for measuring respiratory rate in a hospital setting [7]. In an earlier study using Anura, the SBP and DBP predicted from TOI fell within 5±8 mm Hg of reference measurements [22], but in our study, these criteria were not met in the paired data for the 3 groups. In yet another study using Anura, it was also shown that TOI can determine heart rate, heart rate variability, and infer stress of an individual with high accuracy [21]. Given the Anura precise measurement of heart rate and stress, the app remains an interesting tool for assessing BP. By leveraging advanced machine learning techniques and enhancing the app's sensitivity, those responsible for the app may expand its use and reliability.

In this study, most women expressed a positive experience of using the Anura app. They also reported a higher degree of responsibility and control over their health. This agrees with another study that also found the participants to be satisfied with the contactless technology and would recommend it for future clinical settings [7]. The women generally perceived the app as adequate and safe for use, with nearly all women reporting no elevated anxiety when seeing their BP results. Concerns were raised by some women regarding the accuracy of the measurements and the handling of their data, particularly in relation to privacy. These findings suggest that the app has the potential to be well-received by women during pregnancy, though addressing these concerns will be important for its future use.

Limitations and Future Solutions

A key strength of the study is the use of linear mixed models to account for repeated BP measurements within individuals, which provides a robust estimation of differences between the Anura app and manual measurements compared to simpler paired analyses. Stratifying analyses by patient group and trimester allowed us to explore potential variations across different risk profiles and gestational stages. An additional strength is that the analyses have been quality assured through independent review by a statistician.

However, some limitations should be acknowledged. The models did not adjust for potential confounding factors such as maternal age, BMI, or comorbidities, which may influence BP variability and limit the generalizability of the findings. We also acknowledge a lack of insight into the validation of the underlying technology. Our study sought to explore the use of Anura in a clinical setting with pregnant women, with the primary aim to compare measurement methods. With this aim, and the paired design, we see no benefit to adjust the analysis for maternal age, gestational age, or comorbidities. Additionally, the study focused on comparing measurement methods rather than evaluating clinical outcomes, so the diagnostic accuracy and predictive value of the Anura app for conditions such as preeclampsia remain unassessed. The sample sizes in some subgroups, particularly patients with high-risk pregnancies and preeclampsia in early trimesters, were relatively small, which may reduce statistical power and precision, particularly regarding high-risk pregnancies. Due to these limitations, future studies with larger sample sizes and better compliance are needed to further validate these observations in a clinical setting. Finally, while mixed models handle within-subject correlation, physiological variability and measurement conditions may still contribute to residual variability that is not fully captured by the models. Furthermore, while the Anura app has previously shown a measurement error of approximately 5±8 mm Hg [22], we did not conduct a formal error assessment or apply corrections in the present study. We acknowledge this as a limitation and emphasize that our findings should be interpreted as preliminary feasibility results rather than accuracy validation.

Another limitation was that only 32.6% of the women answered the questionnaire evaluating their experiences of using the Anura app, so caution is advised when interpreting these results. Completing surveys through an app shortly before giving birth may not have been an ideal time point for the women. Additionally, we were unable to send reminders within the app. To improve compliance, future studies might consider alternative methods, such as providing the questionnaires directly by the midwife at the AHC or calling the women by phone to gather their perspective on the experience.

Our results are partly in line with previous studies that reported both acceptable accuracy and significant margins of error for

XSL•FO
RenderX

noncontact BP measurements [6]. Discrepancies can be partly explained by differences in study design, population, and measurement conditions, and previous reviews point to methodological limitations, especially in studies of pregnant women [5]. We also acknowledge that Anura was not compared to 24-hour ambulatory BP monitoring in this study, and we note this as a limitation. Future studies should include such comparisons to better evaluate performance across all risk groups. Home BP measurements offer numbers of benefits [3,4,27] and the ability to measure cuffless BP in a home setting appears to be a promising solution for pregnant women in the future.

Overall, the analyses provide useful insights into the feasibility and agreement of the Anura app with standard BP measurements, but future studies incorporating larger samples, clinical outcomes, and adjustments for relevant confounders are needed to more definitively assess its clinical use. Future research should evaluate Anura in larger and more diverse populations, including women with chronic hypertension and multiple comorbidities. Longitudinal studies are needed to assess performance across pregnancy. Beyond BP monitoring, the technology may enable early detection of other pregnancy-related complications. Usability, integration with clinical systems, and data security should also be explored to support safe implementation in AHC.

Conclusion

The Anura app showed a good acceptance among women with normotensive and high-risk pregnancies. They experienced increased security and control over their health. In normotensive pregnancies, the Anura BP showed a good accuracy; however, the reliability was limited in high-risk and preeclampsia groups, especially for the SBP. The app can therefore currently only be considered as a supportive complement in normal pregnancy, but not a substitute for validated clinical equipment. Continued development and adaptation of algorithms, with data from high-risk pregnancies, is necessary to enable broader clinical use. Such development could strengthen preventive work and contribute to better care for women at risk of hypertensive pregnancy complications.

Acknowledgments

The authors would like to warmly thank all the women who participated in the study and to all midwife colleagues who helped us recruit the participants. This study is a collaboration between the Department of Obstetrics and Gynecology, Lund University, Sweden and NuraLogix, Toronto, Canada [41]. They also thank NuraLogix for their support and collaboration. A language review was assisted by AI.

Funding

This work was supported by funding from the Swedish Research Council, ALF Lund University, SUS, and Region Skane foundations (grant 314373.35.135949).

Data Availability

All data and associated protocols are available on request.

Authors' Contributions

Conceptualization: MEA (lead), CR, SRH (equal), EP (supporting)
Data curation: MEA
Formal analysis: MEA, SRH (lead), EP (supporting)
Funding acquisition: LE, CE, SRH
Investigation: MEA, CR, SRH
Methodology: MEA
Project administration: MEA (lead), CR, SRH (equal), EP (supporting)
Resources: LE, SRH
Supervision: CR, EP, LE, CE, SRH
Validation: MEA, SRH
Visualization: MEA (lead), SRH (supporting)
Writing – original draft: MEA (lead), LE, SRH (supporting)
Writing – review & editing: MEA (lead), CR, EP, LE, CE, SRH (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study protocol for the Anura study.
[DOCX File, 27 KB - humanfactors_v13i1e70370_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Responses to the questionnaire on women's experiences of using the Anura app.
[DOCX File, 15 KB - humanfactors_v13i1e70370_app2.docx]

References

1. Berglund A, Bäckström A, Glantz MA, et al. Mödrahälsovård, sexuell och reproduktiv hälsa [Report in Swedish]. : Svensk förening för Obstetrik och Gynekologi; 2016 URL: https://www.sfog.se/wp-content/uploads/2017/03/modrahalsovard-sexuell-och-reproduktiv-halsa.pdf [accessed 2026-01-22]
2. Bergman LB, Hansson S, Hein A, et al. Guidelines for pregnancy with hypertension [Report in Swedish]. : Swedish Association for Obstetrics and Gynecology; 2019 URL: https://www.sfog.se/wp-content/uploads/2025/03/preeklampsi-riktlinje-uppdaterad-240923.pdf [accessed 2026-01-22]
3. Perry H, Sheehan E, Thilaganathan B, Khalil A. Home blood-pressure monitoring in a hypertensive pregnant population. Ultrasound Obstet Gynecol 2018 Apr;51(4):524-530. [doi: 10.1002/uog.19023] [Medline: 29468771]
4. Weinfeld JM, Hart KM, Vargas JD. Home blood pressure monitoring. Am Fam Physician 2021 Sep 1;104(3):237-243. [Medline: 34523884]
5. Bello NA, Woolley JJ, Cleary KL, et al. Accuracy of blood pressure measurement devices in pregnancy: a systematic review of validation studies. Hypertension 2018 Feb;71(2):326-335. [doi: 10.1161/HYPERTENSIONAHA.117.10295] [Medline: 29229741]
6. Nour M, Polat K, Şentürk Ü, Arıcan M. A novel cuffless blood pressure prediction: uncovering new features and new hybrid ML models. Diagnostics (Basel) 2023 Mar 28;13(7):1278. [doi: 10.3390/diagnostics13071278] [Medline: 37046499]
7. Pham C, Poorzargar K, Panesar D, et al. Video plethysmography for contactless blood pressure and heart rate measurement in perioperative care. J Clin Monit Comput 2024 Feb;38(1):121-130. [doi: 10.1007/s10877-023-01074-6] [Medline: 37715858]
8. Poorzargar K, Pham C, Panesar D, et al. Video plethysmography for contactless measurement of respiratory rate in surgical patients. J Clin Monit Comput 2024 Feb;38(1):47-55. [doi: 10.1007/s10877-023-01064-8] [Medline: 37698697]
9. Chappell LC, Cluver CA, Kingdom J, Tong S. Pre-eclampsia. Lancet 2021 Jul 24;398(10297):341-354. [doi: 10.1016/S0140-6736(20)32335-7] [Medline: 34051884]
10. Steegers EAP, von Dadelszen P, Duvekot JJ, Pijnenborg R. Pre-eclampsia. Lancet 2010 Aug 21;376(9741):631-644. [doi: 10.1016/S0140-6736(10)60279-6] [Medline: 20598363]
11. WHO recommendations for prevention and treatment of pre-eclampsia and eclampsia. : World Health Organization; 2011 URL: https://iris.who.int/server/api/core/bitstreams/02f76d6d-4f9f-4662-84fe-e23c186e1594/content [accessed 2026-01-22]
12. Andolf E, Bremme K, Bruss C, et al. Preeklampsi [Report in Swedish]. : svensk förening för obstetrik och gynekologi Arbets- och Referensgrupp för Perinatologi; 2014 URL: https://www.sfog.se/wp-content/uploads/2025/05/arg72_komplett_lrbf4f5598-2309-4013-8dea-3cbb534708ee.pdf?utm_source=chatgpt.com [accessed 2026-01-22]
13. Hypertens pregnancy. National Institute for Health and Care Exallence. 2013. URL: https://www.nice.org.uk/guidance/qs35/chapter/Quality-statement-2-Antenatal-assessment-of-pre-eclampsia-risk [accessed 2025-06-10]
14. Ajong AB, Yakum MN, Aljerf L, et al. Association of hypertension in pregnancy with serum electrolyte disorders in late pregnancy among Cameroonian women. Sci Rep 2023 Nov 28;13(1):20940. [doi: 10.1038/s41598-023-47623-6] [Medline: 38017060]
15. Riishede I, Rode L, Sperling L, et al. Pre-eclampsia screening in Denmark (PRESIDE): National Validation Study. Ultrasound Obstet Gynecol 2023 Jun;61(6):682-690. [doi: 10.1002/uog.26183] [Medline: 36840981]
16. Chaemsaithong P, Sahota DS, Poon LC. First trimester preeclampsia screening and prediction. Am J Obstet Gynecol 2022 Feb;226(2S):S1071-S1097. [doi: 10.1016/j.ajog.2020.07.020] [Medline: 32682859]
17. Dawson JB, Barker DJ, Ellis DJ, et al. A theoretical and experimental study of light absorption and scattering by in vivo skin. Phys Med Biol 1980 Jul;25(4):695-709. [doi: 10.1088/0031-9155/25/4/008] [Medline: 7454759]
18. Stamatas GN, Zmudzka BZ, Kollias N, Beer JZ. Non-invasive measurements of skin pigmentation in situ. Pigment Cell Res 2004 Dec;17(6):618-626. [doi: 10.1111/j.1600-0749.2004.00204.x] [Medline: 15541019]
19. Liu J, Luo H, Zheng PP, Wu SJ, Lee K. Transdermal optical imaging revealed different spatiotemporal patterns of facial cardiovascular activities. Sci Rep 2018;8(1):10588. [doi: 10.1038/s41598-018-28804-0]
20. Anura. Nuralogix Corporation. URL: https://www.anura.ai [accessed 2026-01-22]
21. Wei J, Luo H, Wu SJ, Zheng PP, Fu G, Lee K. Transdermal optical imaging reveal basal stress via heart rate variability analysis: a novel methodology comparable to electrocardiography. Front Psychol 2018;9:98. [doi: 10.3389/fpsyg.2018.00098] [Medline: 29472879]
22. Luo H, Yang D, Barszczyk A, et al. Smartphone-based blood pressure measurement using transdermal optical imaging technology. Circ Cardiovasc Imaging 2019 Aug;12(8):e008857. [doi: 10.1161/CIRCIMAGING.119.008857] [Medline: 31382766]
23. Bergman LB, Hansson S, Hein A, et al. Riktlinjer för hypertonisjukdomar under graviditet [Report in Swedish]. : Svensk förening för obstetrik och gynekologi; 2019 URL: https://www.sfog.se/wp-content/uploads/2025/03/preeklampsi-riktlinje-uppdaterad-240923.pdf [accessed 2025-06-10]

XSL•FO
RenderX

24. Prevention och handläggning av preeklampsi och graviditetsinducerad hypertoni [Report in Swedish]. : Region Skåne; 2024 URL: https://vardgivare.skane.se/siteassets/1.-vardriktlinjer/regionala-riktlinjer---fillistning/preeklampsi-och-graviditetsinducerad-hypertoni.pdf [accessed 2026-01-22]
25. Användarmanual mödrahälsovård obstetrix 2.18.0.100 [Report in swedish]. : Cerner Sverige AB; 2022 URL: https://vardgivare.skane.se/siteassets/6.-it/it-stod-och-tjanster/obstetrix/manualer---fillistning/anvandarmanual-mhv-obstetrix-2.18.0.100.pdf [accessed 2026-01-22]
26. How to determine sample size from G*power. Statistics Solutions. 2025. URL: https://www.statisticssolutions.com/how-to-determine-sample-size-from-gpower [accessed 2026-01-22]
27. Dougall G, Franssen M, Tucker KL, et al. Blood pressure monitoring in high-risk pregnancy to improve the detection and monitoring of hypertension (the BUMP 1 and 2 trials): protocol for two linked randomised controlled trials. BMJ Open 2020 Jan 23;10(1):e034593. [doi: 10.1136/bmjopen-2019-034593] [Medline: 31980512]
28. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [Medline: 2868172]
29. Mahendru AA, Everett TR, Wilkinson IB, Lees CC, McEniery CM. A longitudinal study of maternal cardiovascular function from preconception to the postpartum period. J Hypertens (Los Angel) 2014;32(4):849-856. [doi: 10.1097/HJH.0000000000000090]
30. American College of Obstetricians and Gynecologists. ACOG practice bulletin no. 125: chronic hypertension in pregnancy. Obstet Gynecol 2012 Feb;119(2 Pt 1):396-407. [doi: 10.1097/AOG.0b013e318249ff06] [Medline: 22270315]
31. Soma-Pillay P, Nelson-Piercy C, Tolppanen H, Mebazaa A. Physiological changes in pregnancy. Cardiovasc J Afr 2016;27(2):89-94. [doi: 10.5830/CVJA-2016-021] [Medline: 27213856]
32. Erlandsson L, Ohlsson L, Masoumi Z, et al. Preliminary evidence that blocking the uptake of placenta-derived preeclamptic extracellular vesicles protects the vascular endothelium and prevents vasoconstriction. Sci Rep 2023 Oct 27;13(1):18425. [doi: 10.1038/s41598-023-45830-9] [Medline: 37891193]
33. Melchiorre K, Sharma R, Khalil A, Thilaganathan B. Maternal cardiovascular function in normal pregnancy: evidence of maladaptation to chronic volume overload. Hypertension 2016 Apr;67(4):754-762. [doi: 10.1161/HYPERTENSIONAHA.115.06667] [Medline: 26962206]
34. Yang D, Xiao G, Wei J, Luo H. Preliminary assessment of video-based blood pressure measurement according to ANSI/AAMI/ISO81060-2: 2013 guideline accuracy criteria: Anura smartphone app with transdermal optimal imaging technology. Blood Press Monit 2020 Oct;25(5):295-298. [doi: 10.1097/MBP.0000000000000467] [Medline: 32842022]
35. Barszczyk A, Zhou W, Lee K. Artificial intelligence in medicine. In: AIM and Transdermal Optical Imaging: Springer Nature; 2022:1143-1156. [doi: 10.1007/978-3-030-58080-3_250-1]
36. Barnes KL, Dunivan G, Jaramillo-Huff A, Krantz T, Thompson J, Jeppson P. Evaluation of smartphone pelvic floor exercise applications using standardized scoring system. Female Pelvic Med Reconstr Surg 2019;25(4):328-335. [doi: 10.1097/SPV.0000000000000563] [Medline: 29489554]
37. Xydopoulos G, Perry H, Sheehan E, Thilaganathan B, Fordham R, Khalil A. Home blood-pressure monitoring in a hypertensive pregnant population: cost-minimization study. Ultrasound Obstet Gynecol 2019 Apr;53(4):496-502. [doi: 10.1002/uog.19041] [Medline: 29516615]
38. Kalafat E, Mir I, Perry H, Thilaganathan B, Khalil A. Is home blood-pressure monitoring in hypertensive disorders of pregnancy consistent with clinic recordings? Ultrasound Obstet Gynecol 2018 Oct;52(4):515-521. [doi: 10.1002/uog.19094] [Medline: 29786155]
39. Pham C, Poorzargar K, Nagappa M, et al. Effectiveness of consumer-grade contactless vital signs monitors: a systematic review and meta-analysis. J Clin Monit Comput 2022 Feb;36(1):41-54. [doi: 10.1007/s10877-021-00734-9] [Medline: 34240262]
40. Cho JS, Park JH. Application of artificial intelligence in hypertension. Clin Hypertens 2024 May 1;30(1):11. [doi: 10.1186/s40885-024-00266-9] [Medline: 38689376]
41. Revolutionizing digital health—one face at a time. Nuralogix. TM URL: https://www.nuralogix.ai/ [accessed 2026-02-06]

Abbreviations

AHC: antenatal health clinic
BP: blood pressure
DBP: diastolic blood pressure
DFX: DeepAffex
GW : gestational week
LoA: limits of agreement
NICU: neonatal intensive care unit
SBP: systolic blood pressure
TOI: transdermal optical imaging

XSL•FO
RenderX

Edited by A Kushniruk; submitted 20.Dec.2024; peer-reviewed by F Pettit, L Aljerf; revised version received 04.Dec.2025; accepted 02.Jan.2026; published 18.Feb.2026.

Please cite as:
Andersson ME, Rubertsson C, Psouni E, Erlandsson L, Edvinsson C, Hansson SR
Evaluating a Smartphone App to Monitor Blood Pressure in Normotensive Pregnancies, High-Risk Pregnancies, and Women With Preeclampsia: Prospective Longitudinal Feasibility Study
JMIR Hum Factors 2026;13:e70370
URL: https://humanfactors.jmir.org/2026/1/e70370
doi:10.2196/70370

© Maria E Andersson, Christine Rubertsson, Elia Psouni, Lena Erlandsson, Camilla Edvinsson, Stefan R Hansson. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 18.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Acceptability of a Digital Care App in Patients Undergoing Hip and Knee Arthroplasty: Prospective Cohort Study

Yacine Louni[1], MDCM; Matthew Laroche[1], BCom; Abdulrhman Alnasser[2], MD; Mohammad Abuhaneya[2], MD; Eric Belzile[3], MSc; Sandhya Baskaran[3], MEd, MBA; Jennifer Mutch[2], MDCM, FRCSC; Anthony Albers[2], MDCM, FRCSC

[1]Faculty of Medicine, McGill University, Montreal, QC, Canada

[2]Division of Orthopaedic Surgery, McGill University, St Mary's Hospital, 3830 Avenue Lacombe, Montreal, QC, Canada

[3]St Mary's Hospital Research Center, St Mary's Hospital, Montreal, QC, Canada

Corresponding Author:
Anthony Albers, MDCM, FRCSC
Division of Orthopaedic Surgery, McGill University, St Mary's Hospital, 3830 Avenue Lacombe, Montreal, QC, Canada

Related Article:

This is a corrected version. See correction statement: https://humanfactors.jmir.org/2026/1/e101097

Abstract

Background: Mobile health (mHealth) apps have become more commonly used in orthopedics. However, for these apps to be efficient, patients should be willing to use them, making it essential to understand patients' perspectives of mHealth interventions.

Objective: The aim of this single-center, intent-to-treat, preoperative single-cohort study of 100 patients was to evaluate the acceptability of mymobility (Zimmer-Biomet), an mHealth app designed for the postoperative care of total hip arthroplasty (THA) and total knee arthroplasty (TKA).

Methods: We measured acceptability using the theoretical framework for acceptability (TFA) preoperatively and at 3 months post operation. We also measured satisfaction with app use postoperatively using the Usefulness, Satisfaction, and Ease of Use questionnaire as well as patient-reported outcome measures preoperatively and postoperatively using the Oxford hip and knee scores and the visual analog scale for pain. Patients included were 18 years or older; underwent unilateral primary total hip, total knee, or partial knee arthroplasty; spoke and read French or English; and had a smartphone with internet access. Participants used mymobility in addition to standard government-funded physiotherapy.

Results: The preoperative overall TFA result was 4.2 out of 5, but results decreased significantly postoperatively. There was higher self-efficacy in preoperative TFAs with higher education, and lower acceptability in postoperative TFAs with TKA. The Usefulness, Satisfaction, and Ease of Use questionnaire revealed a good level of satisfaction with the use of the app. Patient-reported outcome measures showed earlier improvement in THA (31.2 d) than in TKA (89.4 d), whereas the visual analog scale showed a rapid decrease in pain with both procedures. Only 1 patient expressed privacy concerns with the use of the app.

Conclusions: There was a good level of acceptability with the use of mymobility for the postoperative management of THA and TKA, although acceptability decreased postoperatively. This decrease could signify high expectations toward the app preoperatively or higher than expected difficulty and pain in the early postoperative period. Acceptability tended to increase with higher education and decrease with TKA. These trends are consistent with prior literature and constitute a potential gap to address for app developers. The influence of the natural recovery process on acceptability remains unclear. Future studies could explore this gap by comparing results in cohorts using the app to cohorts with standard care.

(JMIR Hum Factors 2026;13:e79682) doi:10.2196/79682

KEYWORDS

mobile health; orthopedics; patient satisfaction; telemedicine; telerehabilitation; total joint arthroplasty

Introduction

In Canada, there were 60 705 hip and 70 379 knee replacements performed in the inpatient setting in 2023 - 2024 [1]. These procedures were the third and second most common inpatient surgeries, and these numbers are expected to at least double in the United States (and likely Canada) over the next 15 years due to an aging population [2,3]. Another growing sector in health care is the use of smartphones for patient care. An estimated 71% of the world population uses a smartphone in 2024, and there are more than 350 000 mobile health (mHealth) apps currently available [4,5]. This increasing use of mHealth

XSL•FO
RenderX

apps is also present in various areas of orthopedics, where numerous studies have shown that they are efficient and convenient, that they promote patient adherence and engagement with their treatment, and that they facilitate rehabilitation while promoting positive outcomes [6-13].

Although mHealth has the potential to improve outcomes, patients' attitudes toward these apps must be assessed, as they can only be effective if patients are willing to use them. User acceptability is defined as "a multifaceted construct that reflects the extent to which people delivering or receiving health care intervention consider it to be appropriate based on anticipated or experienced cognitive and emotional responses to the intervention" [14]. Acceptability is an important measure, as it not only provides information on patients' experience with mHealth apps, but also gives an insight into their cognitive and emotional response [14]. This helps determine whether patients genuinely perceive them as appropriate and valuable.

Recent studies investigated patients' perspectives of mHealth interventions in the context of postoperative hip and knee arthroplasty care. These studies found a high level of engagement and increased compliance with treatment exercises when using mHealth interventions [13,15-18]. Other findings include participants being satisfied with these interventions while also finding them easy to use and engaging, and providing better connection to their treating team [13,16,19,20]. These results are promising and could signify that patients are willing to use mHealth apps for their rehabilitation. However, these studies either did not assess acceptability or did so through surveys that were researcher-created or not primarily designed to measure acceptability. According to Perski et al [21], this reduces clarity regarding which aspects of acceptability are satisfied, and which are not. Additionally, high engagement does not necessarily reflect patients' perceptions of the intervention. Engagement might be driven by external factors, such as study participation, financial incentives, or free access to the devices and apps, while patients might still find the app to be time-consuming, repetitive, and burdensome.

Measuring acceptability, especially preoperatively, is therefore relevant as it provides a better perspective of patients' opinions prior to and beyond use, and because it "may usefully be considered an emergent property […] which in turn influences (and is influenced by) user engagement and intervention effectiveness" [21].

We started using the mymobility app (Zimmer Biomet, Warsaw, and Indiana) in 2022 in our community hospital to provide postoperative care to patients receiving outpatient total hip arthroplasty (THA) and total knee arthroplasty (TKA) [22]. This app, available on iOS and Android, currently stands at more than 50 000 downloads on the Google Play Store and is amongst the most downloaded when compared to other apps with the same objective [23]. In this context, our team sought to investigate the acceptability of mymobility, an mHealth app specifically designed for the postoperative care of THA and TKA, using a standardized framework.

Methods

Study Design

The primary aim of this single-center, intent-to-treat (ITT), single-cohort study was to evaluate pre- and postoperative patient acceptability with the use of the mymobility app for the postoperative care of THA and TKA. The secondary outcomes were postoperative patient satisfaction with the app and pre- and postoperative patient-reported outcome measures (PROMs).

Data Collection

Data collected for this study are presented in Multimedia Appendix 1. Demographic data were obtained from patient charts. PROMs were collected either directly through the mymobility app or by one of the authors contacting participants by phone or email. Details of the specific PROMs are provided later in this section. All data were stored securely and deidentified prior to analysis.

Ethical Considerations

This study received ethics approval from the Research Ethics Board of the Montreal West Island's Integrated University Health and Social Services Centers (Biomedical Subcommittee), affiliated with St. Mary's Hospital, Montreal (IRB# 2024-954). All participants provided informed consent prior to participation. Study data, including personal information, remained confidential and were deidentified. No secondary analyses were conducted. Study participants received no financial compensation, but access to the mymobility app was given for free. No identifiable images of participants were included in this manuscript. Under the Personal Information Protection and Electronic Documents ACT, the data collected by mymobility could not be used or shared with third parties, nor could cookies be collected, unless participants gave their explicit consent within the app [24,25]. If consent was given, data could be used or shared by the app following anonymization, with consent revocable at any time. Participants were not required to give consent to data use or sharing, or to cookie collection to use the app and be part of the study. This was made clear to participants as part of their informed consent.

App Description

Mymobility is an app that is accessible either through a smartphone or an Apple Watch. The app features educational content on pre- and postoperative care, exercise routines with metrics to track progress, AI-assisted gait speed analysis to determine the level of recovery, in-app PROMs, and telemedicine with the opportunity to send SMS text messages, videos, or pictures to health care providers (see Multimedia Appendix 2). The app engages patients in their recovery by sending notifications and reminders if users accept them while also keeping track of progress made. It also provides patients with basic and important smartphone functions to use (eg, Wi-Fi) for those with low digital literacy.

XSL•FO
RenderX

Questionnaires

Acceptability

Acceptability was measured using the theoretical framework for acceptability (TFA) questionnaire developed by Sekhon et al [26], which was adapted to our research question (see Multimedia Appendix 3). The questionnaire comprises 8 items answered by participants preoperatively and at 3 months post operation by phone or email. Each statement was evaluated using a 5-point Likert scale: scores 1 to 2 were classified as negative, 3 as neutral, and 4 - 5 as positive. The mean score was calculated for each of the 8 items, as well as an overall score.

Satisfaction

Patient satisfaction was measured using the Usefulness, Satisfaction, and Ease of Use (USE) questionnaire at 3 months post operation (see Multimedia Appendix 4) [27]. Satisfaction was defined as "the net feeling of pleasure or displeasure that results from aggregating all the benefits that a person hopes to receive from interaction with the information system" [27]. The questionnaire was answered by phone or email. Statements were evaluated using a 7-point Likert scale: scores 1 to 3 were classified as negative, 4 as neutral, and 5 to 7 as positive. The mean score was then calculated for each of the 4 sections of the questionnaire. Two additional questions assessing satisfaction were asked and scored on a 5-point Likert scale (see Multimedia Appendix 5).

Patient-Reported Outcome Measures

PROMs were measured using the Oxford Hip Score (OHS) and the Oxford Knee Score (OKS) as well as the visual analog scale (VAS) for pain (scored from 0 to 10) preoperatively and at 1, 3, and 6 months post operation (see Multimedia Appendices 6 and 7) [28-30]. OHS and OKS were answered directly within the app, whereas the VAS was answered by phone or email.

Recruitment

Patients were included in the study if they were booked for unilateral primary THA or TKA, were clinically suited for telehealth care (characterized as age ≥18 y, speak and read French or English, and have a smartphone or a tablet), had internet access, and provided informed consent to participate in the study. Patients were excluded if they did not meet the inclusion criteria, if they were unable or unwilling to provide informed consent, if they were undergoing emergency THA or TKA, or if they were undergoing revision surgeries. Participants had access to the app in addition to normal postoperative care, which included physiotherapy and postoperative follow-up visits with an orthopedic surgeon.

Patients were approached by a member of the research team during their preoperative assessment visit in the clinic. Eligible patients were provided with an informed consent form detailing the study. Once enrolled, a 1-year free access to the mymobility app was provided to the participants by the treating physician or a member of the research team, after which an email with instructions to download the app was sent. Participants were then required to download the app and agree to the terms and conditions to use it [24].

In the end, 100 participants were included in this study and had a profile created in the app. Participants who did not download the app preoperatively were sent a reminder email and given a few days following surgery, after which they were considered as nondownloaders. In the end, 75 participants downloaded the app and 25 did not. Only participants who downloaded the app were provided with the study questionnaires, as the goal was to assess acceptability, satisfaction, and PROMs with use of the app. Patients who did not download the app remained in the study in the context of the ITT format. However, since deciding not to download the app could signify low acceptability, we contacted participants who did not download the app by phone or email to ask the reason behind their decision.

Data Analysis

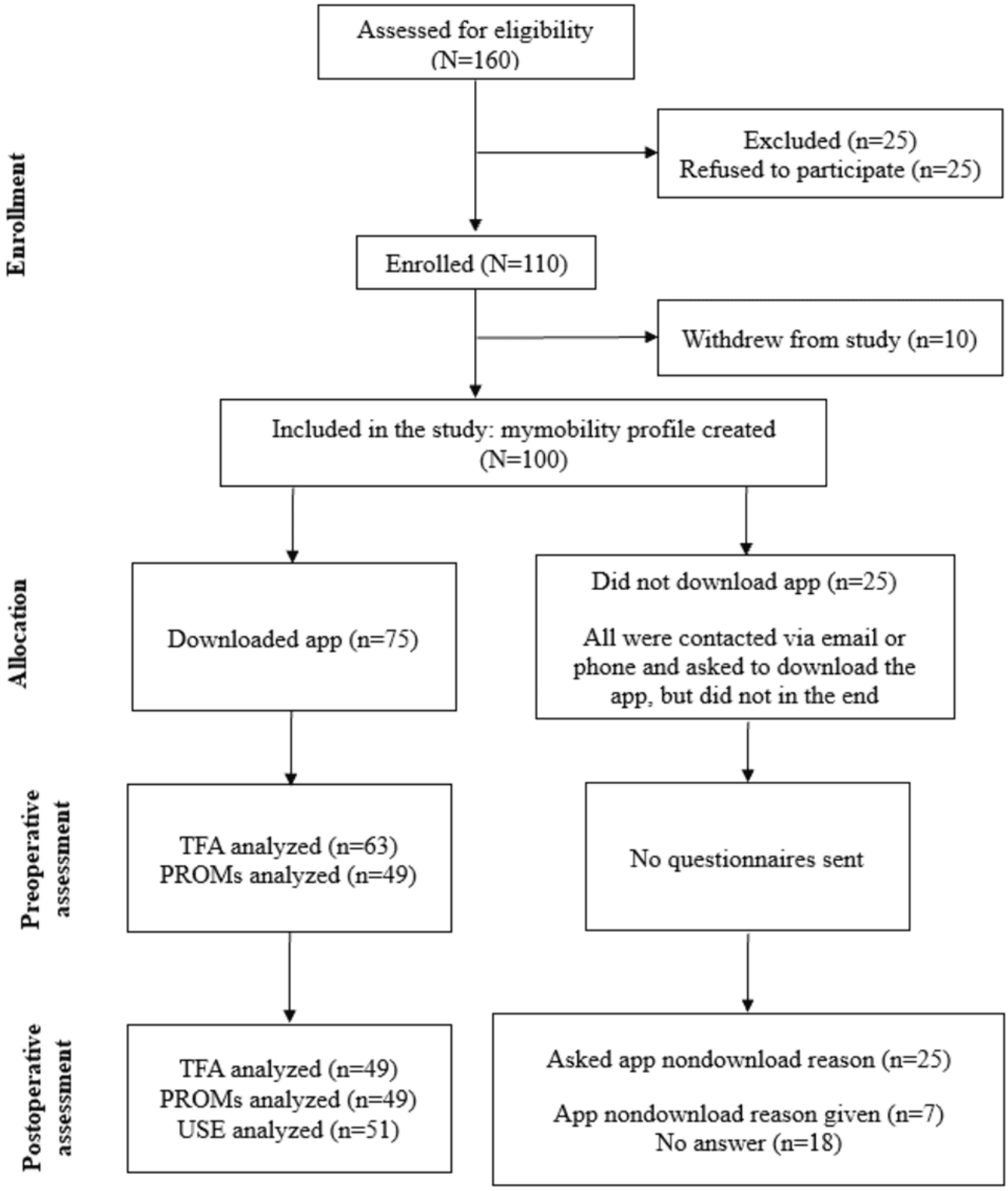
Study data was analyzed using the ITT principle. Data analysis was conducted and reported using STATA 17.0 and R software. Statistical significance was defined as a P value <.05.

Results

The recruitment process is summarized in Figure 1. Baseline characteristics of our population can be found in Table 1.

XSL•FO
RenderX

Figure 1. Study enrollment, participation, and questionnaire distribution flow diagram. PROM: patient-reported outcome measure; TFA: theoretical framework for acceptability; USE: Usefulness, Satisfaction, and Ease of Use.



XSL•FO
RenderX

Table . Baseline demographic and clinical characteristics of the study population.

Baseline variables	Participants (N=100)
Female, n (%)	51 (52)
Missing	1 (1)
Age (years), mean (SD)	67.0 (10.7)
Ethnicity, n (%)	
Caucasian	75 (82)
Asian	9 (10)
Arab	3 (3)
Other	5 (5)
Missing	8 (8)
Education, n (%)	
Graduate	20 (21)
Undergraduate	27 (29)
Postsecondary (nonuniversity)	25 (27)
Primary/secondary	22 (23)
Missing	6 (6)
Employment, n (%)	
Full/part time	31 (34)
Retired	56 (61)
Unemployed	5 (5)
Missing	8 (8)
Language used, n (%)	
English	64 (69)
French	23 (25)
Both	6 (6)
Missing	7 (7)
Comorbidities, n (%)	
Yes[a]	39 (39)
None	61 (61)
Hypertension	18 (31)
Diabetes	10 (17)
Obesity	6 (10)
Thyroid	4 (7)
Cholesterol	4 (7)
Other	17 (29)
Procedure, n (%)	
TKA[b]	55 (55)
THA[c]	42 (42)
PKA[d]/UKA[e]	3 (3)
Income (CAD $[f]), n (%)	
<25,000	8 (9)
25,000 - 50,000	15 (17)

XSL•FO
RenderX

Baseline variables	Participants (N=100)
50,000 - 75,000	10 (11)
>100,000	24 (27)
Prefer not to say/unknown	13 (15)
Missing	11 (11)
Area of residency, n (%)	
Urban	72 (77)
Rural	22 (23)
Missing	6 (6)

[a]Participants who answered “Yes” can have multiple comorbidities.

[b]TKA: total knee arthroplasty.

[c]THA: total hip arthroplasty.

[d]PKA: partial knee arthroplasty.

[e]UKA: unicompartmental knee arthroplasty.

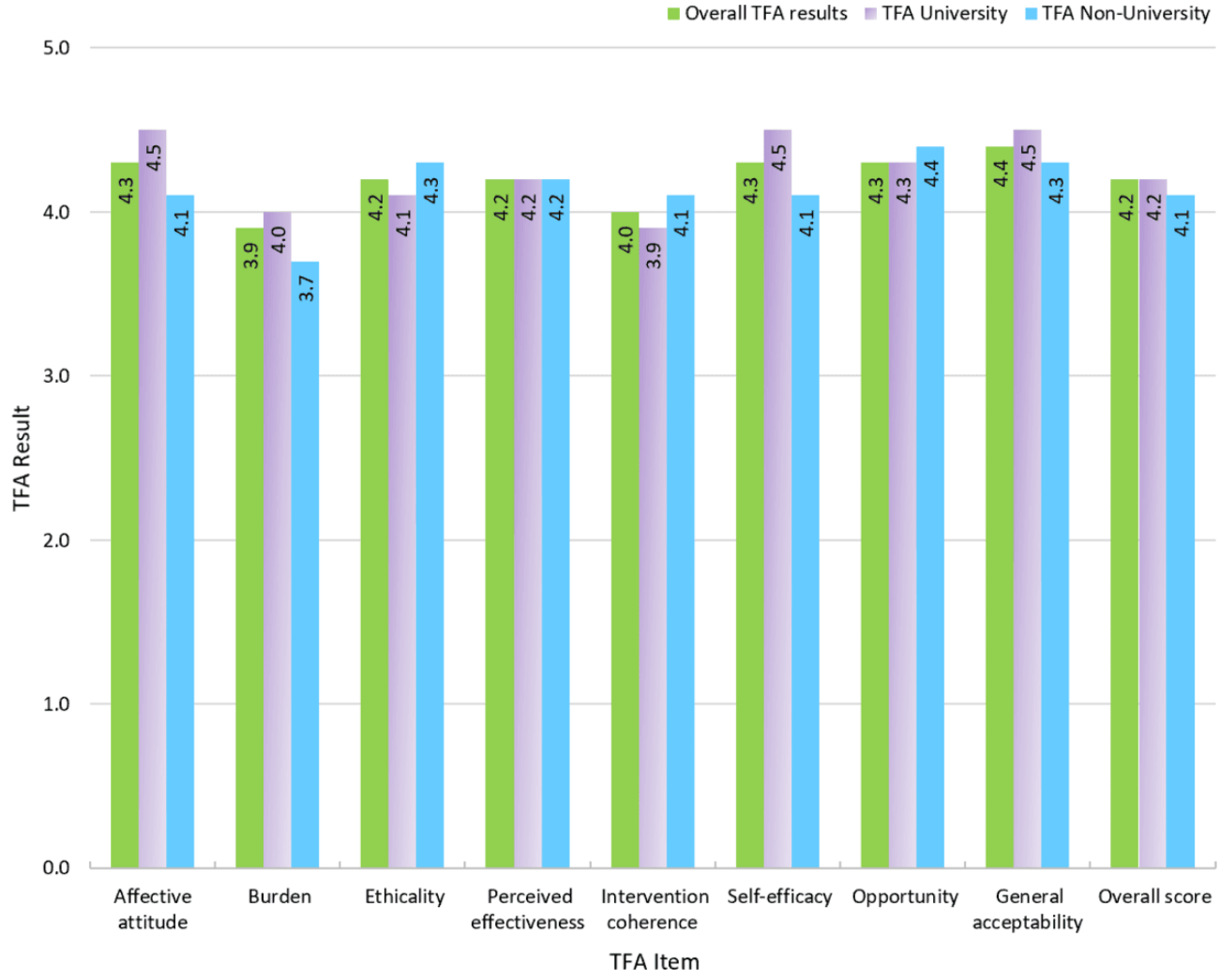
[f]A currency exchange rate of CAD $1=US $0.75 is applicable.

Acceptability

The overall score for the preoperative TFA questionnaires was 4.2 (SD 0.6; Table 2). We found a statistically significant increase in self-efficacy with a university level of education compared to non-university (*P*=.04), but no difference related to age, sex, or employment (Figure 2). There was a statistically significant decrease between pre- and postoperative overall TFA (*P*=.007), as well as in pre- and postoperative perceived effectiveness (*P*=.01) and self-efficacy (*P*=.008; Table 3). Furthermore, when looking at TFA results by procedure, there was a statistically significant decrease with TKA in multiple TFA items, including overall acceptability (*P*=.008; Figure 3). There were no statistically significant differences between pre- and postoperative overall TFA nor in any of the TFA items with THA.

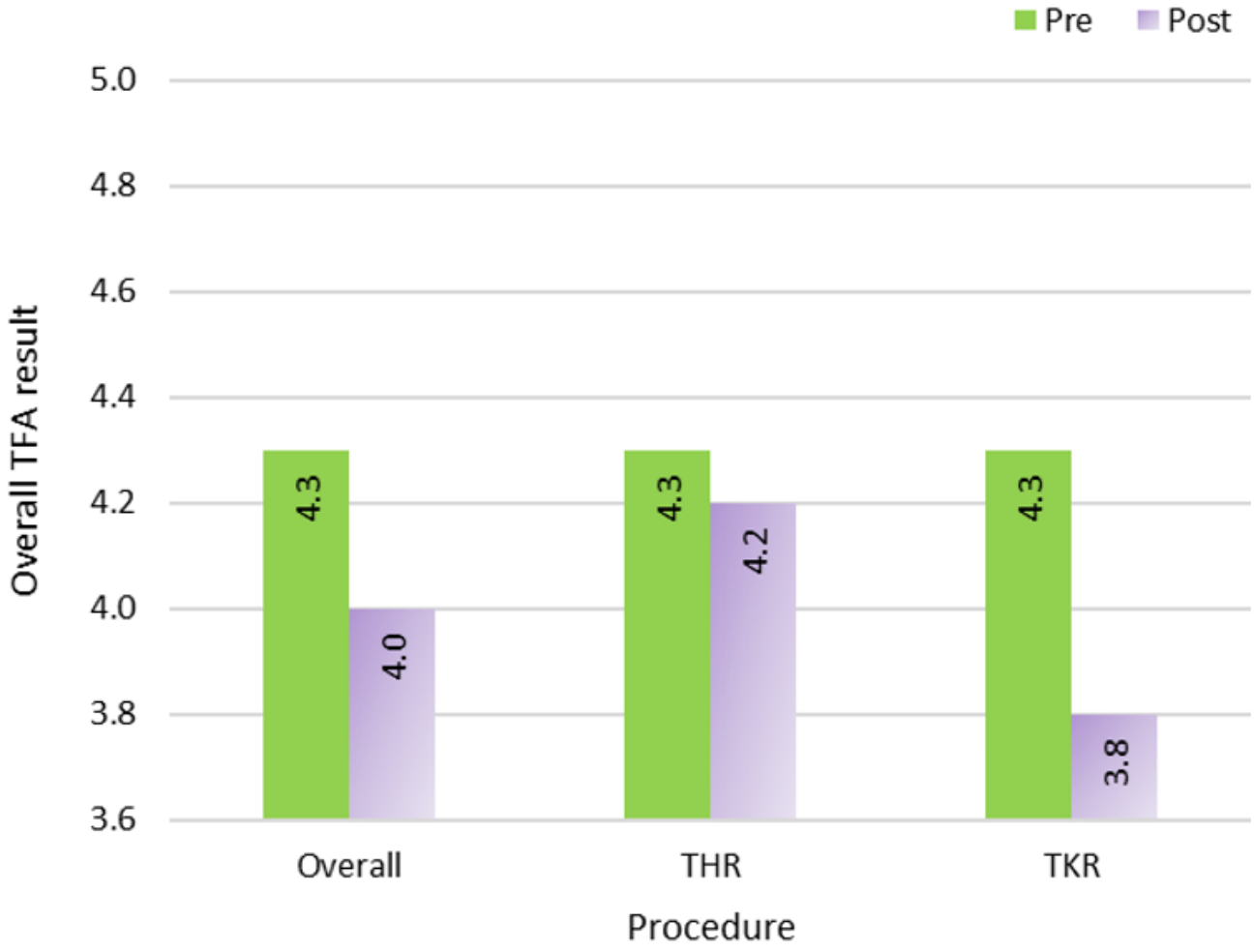
XSL•FO
RenderX

Figure 2. Preoperative TFA results (all items) for the overall population and by education level. TFA: theoretical framework for acceptability.



XSL•FO
RenderX

Figure 3. Comparison between pre- and postoperative overall TFA results for the overall population and by procedure. TFA: theoretical framework for acceptability; THR: total hip replacement; TKR: total knee replacement.



XSL•FO
RenderX

Table . Preoperative TFA[a] results and subgroup analysis by demographic category[b].

TFA items	Overall (n=63), mean (SD)	Age			Sex			Education			Employment		
		<70 (n=34), mean (SD)	≥70 (n=29), mean (SD)	*P* value (*t* test)	Male (n=33), mean (SD)	Female (n=30), mean (SD)	*P* value (*t* test)	Nonuniversity (n=28), mean (SD)	University (n=32), mean (SD)	*P* value (*t* test)	Not working (n=36), mean (SD)	Working (n=23), mean (SD)	*P* value (*t* test)
Affective attitude	4.3 (0.9)	4.3 (0.8)	4.4 (0.8)	.84	4.4 (0.8)	4.1 (1.0)	.21	4.1 (1.0)	4.5 (0.8)	.08	4.4 (0.8)	4.3 (1.1)	.67
Burden	3.9 (0.9)	3.9 (1.0)	3.8 (0.9)	.63	4.0 (0.7)	3.9 (1.1)	.47	3.7 (1.2)	4.0 (0.7)	.20	3.8 (1.0)	4.0 (0.9)	.63
Ethicality	4.2 (1.0)	4.1 (1.1)	4.3 (1.1)	.32	4.1 (1.1)	4.3 (0.8)	.34	4.3 (0.9)	4.1 (1.1)	.54	4.1 (1.1)	4.4 (0.8)	.22
Perceived effectiveness	4.2 (0.8)	4.1 (1.0)	4.3 (0.6)	.23	4.2 (0.9)	4.2 (0.9)	.96	4.2 (0.8)	4.2 (0.9)	.98	4.2 (0.7)	4.3 (1.2)	.55
Intervention coherence	4.0 (0.9)	3.8 (1.1)	4.3 (0.6)	.05	4.0 (0.9)	4.0 (1.0)	.99	4.1 (0.7)	3.9 (1.1)	.49	4.1 (0.8)	3.8 (1.2)	.21
Self-efficacy	4.3 (0.7)	4.3 (0.9)	4.3 (0.5)	.93	4.3 (0.8)	4.3 (0.7)	.99	*4.1 (0.9)*	*4.5 (0.6)*	*.04*	4.3 (0.5)	4.3 (1.1)	.89
Opportunity	4.3 (0.9)	4.2 (1.0)	4.5 (0.7)	.16	4.3 (0.9)	4.3 (0.8)	.88	4.4 (0.7)	4.3 (0.9)	.40	4.4 (0.7)	4.2 (1.0)	.27
General acceptability	4.4 (0.5)	4.3 (0.6)	4.4 (0.5)	.74	4.3 (0.5)	4.4 (0.6)	.87	4.3 (0.5)	4.5 (0.5)	.09	4.3 (0.5)	4.5 (0.6)	.40
Overall score[c]	4.2 (0.6)	4.1 (0.6)	4.3 (0.5)	.22	4.2 (0.5)	4.2 (0.7)	.78	4.1 (0.6)	4.2 (0.5)	.55	4.2 (0.5)	4.2 (0.6)	.94

[a]TFA: theoretical framework for acceptability.

[b]Statistically significant results are given in italic.

[c]Burden, ethicality, and opportunity costs were reversed to compute the overall TFA score.

XSL•FO
RenderX

Table . Change in TFA[a] from preoperative to 3 months post operation[b].

TFA items	Overall (n=49)			TKA[c] (n=27)			THA[d] (n=22)		
	Pre, mean (SD)	Post, mean (SD)	*P* value (*t* test)	Pre, mean (SD)	Post, mean (SD)	*P* value (*t* test)	Pre, mean (SD)	Post, mean (SD)	*P* value (*t* test)
Affective attitude	4.5 (0.6)	4.2 (1.3)	.09	4.5 (0.6)	4.1 (1.3)	.22	4.6 (0.7)	4.3 (1.3)	.26
Burden	4.0 (0.8)	4.0 (1.1)	.91	3.9 (1.0)	3.8 (1.3)	.70	4.1 (0.7)	4.3 (0.8)	.36
Ethicality	4.3 (0.9)	3.9 (1.3)	.11	*4.4 (0.9)*	*3.6 (1.4)*	*.04*	4.2 (0.9)	4.3 (1.1)	.89
Perceived effectiveness	*4.3 (0.8)*	*3.8 (1.3)*	*.01*	*4.3 (0.7)*	*3.7 (1.4)*	*.02*	4.2 (0.9)	4.0 (1.2)	.28
Intervention coherence	4.1 (0.9)	3.8 (1.2)	.09	*4.2 (0.8)*	*3.6 (1.3)*	*.01*	3.9 (1.1)	4.0 (1.0)	.69
Self-efficacy	*4.4 (0.5)*	*4.0 (1.2)*	*.008*	*4.4 (0.5)*	*3.8 (1.4)*	*.02*	4.5 (0.6)	4.3 (0.9)	.21
Opportunity	4.4 (0.7)	4.1 (1.1)	.08	4.4 (0.7)	4.0 (1.3)	.10	4.4 (0.8)	4.3 (0.7)	.54
General acceptability	4.4 (0.5)	4.3 (1.0)	.56	4.4 (0.5)	4.1 (1.1)	.28	4.4 (0.5)	4.5 (0.8)	.63
Overall score[e]	*4.3 (0.5)*	*4.0 (0.8)*	*.007*	*4.3 (0.5)*	*3.8 (0.9)*	*.008*	4.3 (0.4)	4.2 (0.6)	.41

[a]TFA: theoretical framework for acceptability.

[b]Statistically significant results are given in italic.

[c]TKA: total knee arthroplasty.

[d]THA: total hip arthroplasty.

[e]Burden, ethicality, and opportunity costs were reversed to compute the overall TFA score.

Satisfaction

The USE questionnaire and the additional satisfaction questions revealed good levels of satisfaction with the use of the app (Table 4 and Figure 4). Patients consistently rated the app as useful and easy to integrate into their recovery process.

XSL•FO
RenderX

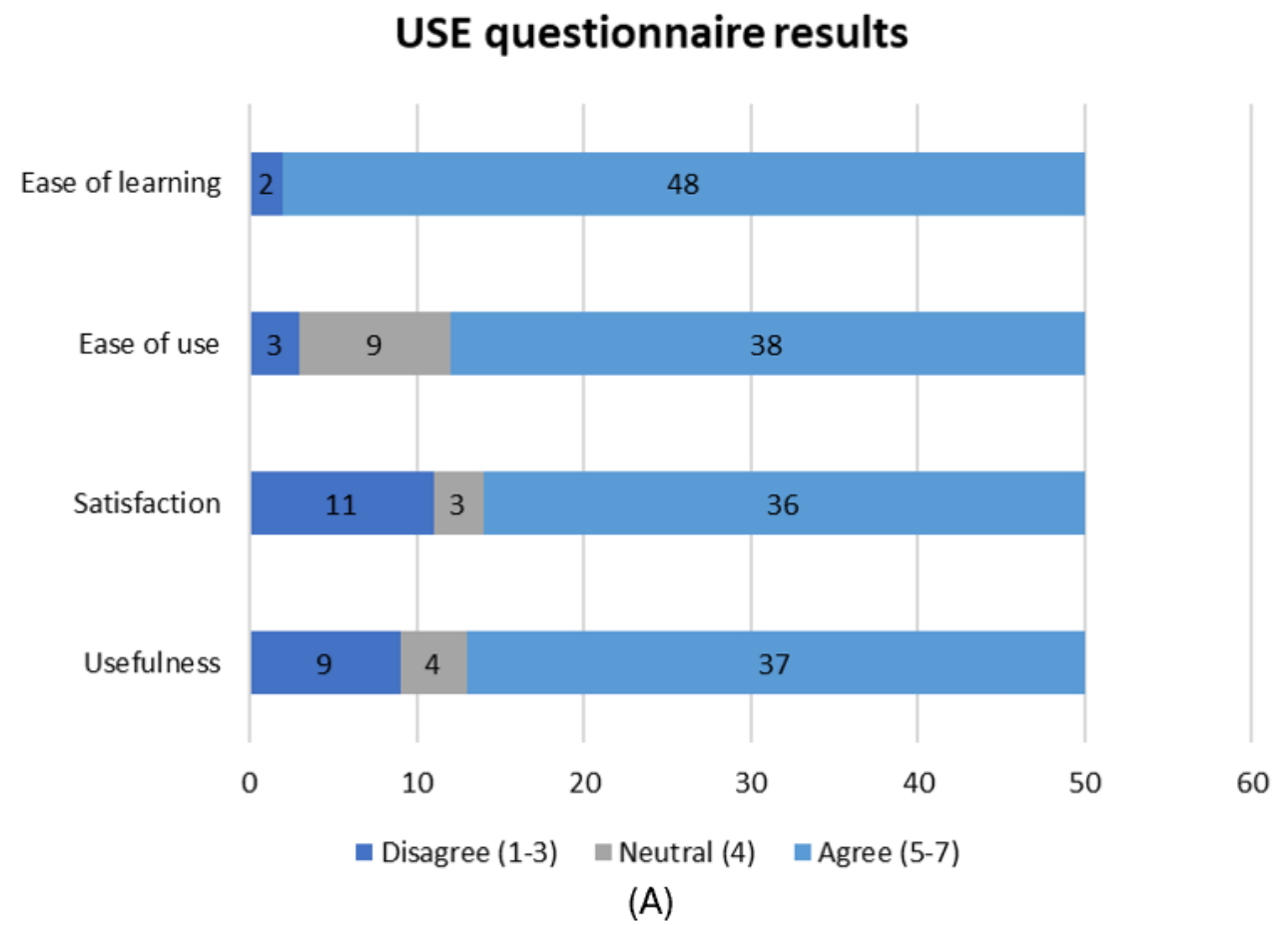
Table . Usefulness, Satisfaction, and Ease of Use questionnaire results 3 months post operation.

Variable	Participants (n=51)
Usefulness score (1-7)	
Disagree (1-3), n (%)	9 (18)
Neutral (4), n (%)	4 (8)
Agree (5-7), n (%)	37 (74)
Missing, n (%)	1 (1.9)
Mean (SD)	5.0 (1.5)
Satisfaction (1-7)	
Disagree (1-3), n (%)	11 (22)
Neutral (4), n (%)	3 (6)
Agree (5-7), n (%)	36 (72)
Missing, n (%)	1 (1.9)
Mean (SD)	4.9 (1.5)
Ease of use (1-7)	
Disagree (1-3), n (%)	3 (6)
Neutral (4), n (%)	9 (18)
Agree (5-7), n (%)	38 (76)
Missing, n (%)	1 (1.9)
Mean (SD)	5.4 (1.1)
Ease of learning (1-7)	
Disagree (1-3), n (%)	2 (4)
Neutral (4), n (%)	0 (0)
Agree (5-7), n (%)	48 (96)
Missing, n (%)	1 (1.9)
Mean (SD)	5.9 (0.9)
Additional questions	
Level of satisfaction with postoperative care (1-5)	
Disagree (1-2), n (%)	6 (12)
Neutral (3), n (%)	4 (8)
Agree (4-5), n (%)	41 (80)
Mean (SD)	4.1 (1.2)
Likelihood to recommend care to friend or family member (1-5)	
Disagree (1-2), n (%)	8 (16)
Neutral (3), n (%)	3 (6)
Agree (4-5), n (%)	40 (78)
Mean (SD)	4.2 (1.3)

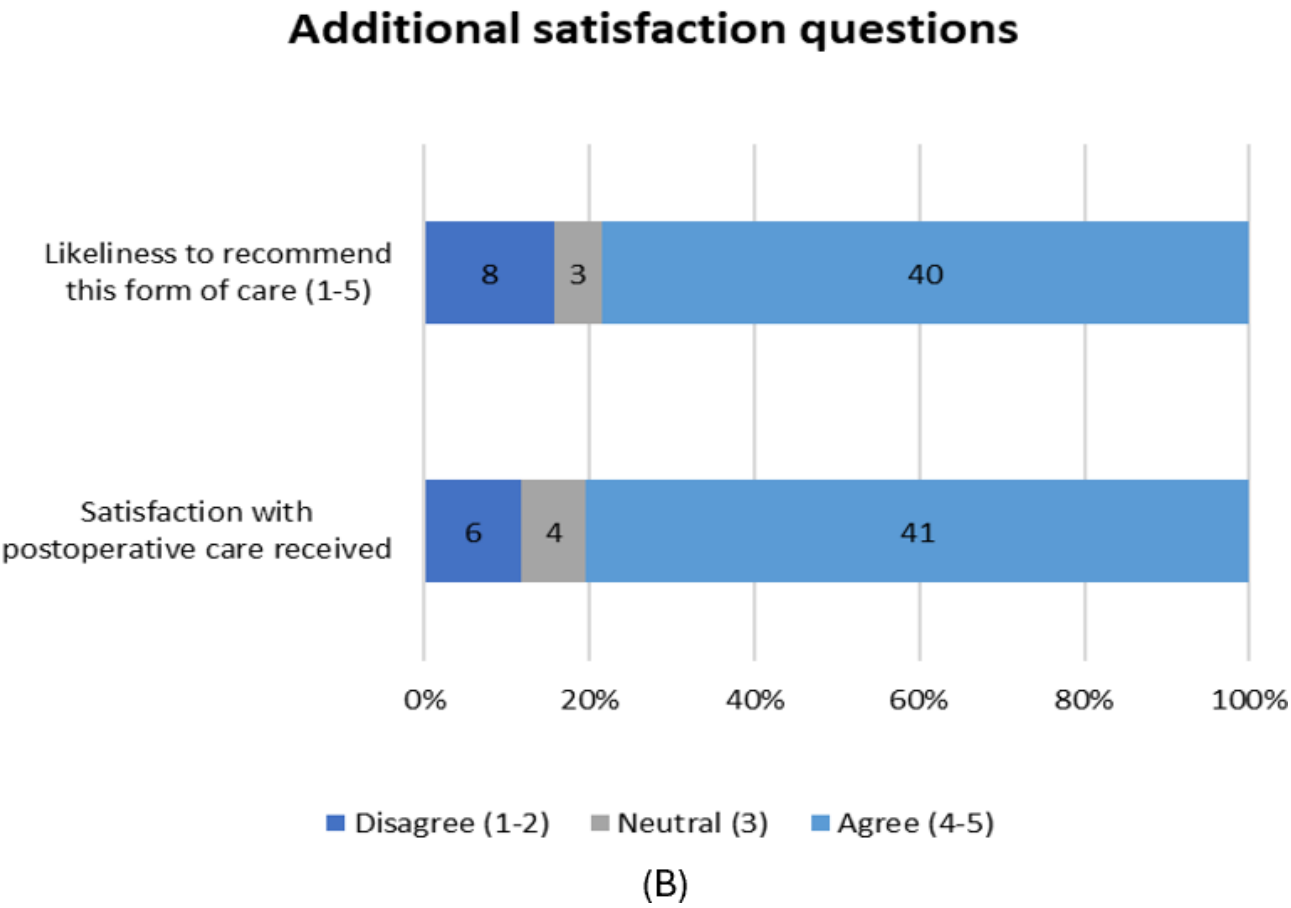
XSL•FO
RenderX

Figure 4. Satisfaction questionnaire results: (A) USE questionnaire results on a stacked bar chart for Likert data; (B) additional satisfaction question results on a 100% stacked bar chart. USE: Usefulness, Satisfaction, and Ease of Use.

XSL•FO
RenderX



(A)



(B)

XSL•FO
RenderX

Patient-Reported Outcome Measures

There was a statistically significant increase in overall Oxford score and OHS at 31.2 days post operation, but only at 89.4 days post operation in OKS (Table 5 and Figure 5). As for the VAS, there was a significant decrease in overall, hip, and knee VAS scores at 31.2 days post operation.

Table . Oxford scores and VAS[a] results at 1, 3, and 6 months post operation[b].

Sample time point	Preop[c]		First postop (with preop data)[d]				Second postop (with preop data)[e]				Third postop (with preop data)[f]			
	Participants, n	Mean (SD)	Participants, n	Mean (SD)	Mean difference vs pre	*P* value (*t* test)	Participants, n	Mean (SD)	Mean difference vs pre	*P* value (*t* test)	Participants, n	Mean (SD)	Mean difference vs pre	*P* value (*t* test)
Oxford score overall (0 - 48)	49	24.1 (10.1)	49	*30.4 (8.6)*	*6.3*	*<.001*	21	*37.8 (5.7)*	*10.4*	*<.001*	9	*39.1 (3.8)*	*12.2*	*.002*
Oxford Hip Score (0 - 48)	20	22.8 (11.8)	20	*35.0 (6.4)*	*12.3*	*<.001*	8	*40.5 (4.4)*	*13.1*	*.01*	1	—[g]	—	—
Oxford Knee Score (0 - 48)	29	25.0 (8.9)	29	27.2 (8.6)	2.3	.17	13	*36.1 (5.9)*	*8.7*	*.001*	8	*39.9 (3.3)*	*12.5*	*.004*
VAS overall (0 - 10)	49	4.6 (2.6)	49	*2.3 (2.1)*	*−2.3*	*<.001*	22	*0.9 (1.1)*	*−3.1*	*<.001*	9	*1.1 (0.4)*	*−2.4*	*.007*
VAS hip (0 - 10)	20	4.5 (2.6)	20	*1.5 (1.4)*	*−3.0*	*<.001*	9	*0.7 (0.3)*	*−2.7*	*.009*	1	—	—	—
VAS knee (0 - 10)	29	4.7 (2.6)	29	*2.9 (2.2)*	*−1.8*	*<.001*	13	*1.0 (0.4)*	*−3.4*	*<.001*	8	*0.8 (0.3)*	*−2.7*	*.004*

[a]VAS: visual analog scale.

[b]Statistically significant results are given in italic.

[c]Gap with surgery (days): mean 18.8 (SD 10.1, range 3-36).

[d]Gap with surgery (days): mean 31.2 (SD 2.8, range 30-44).

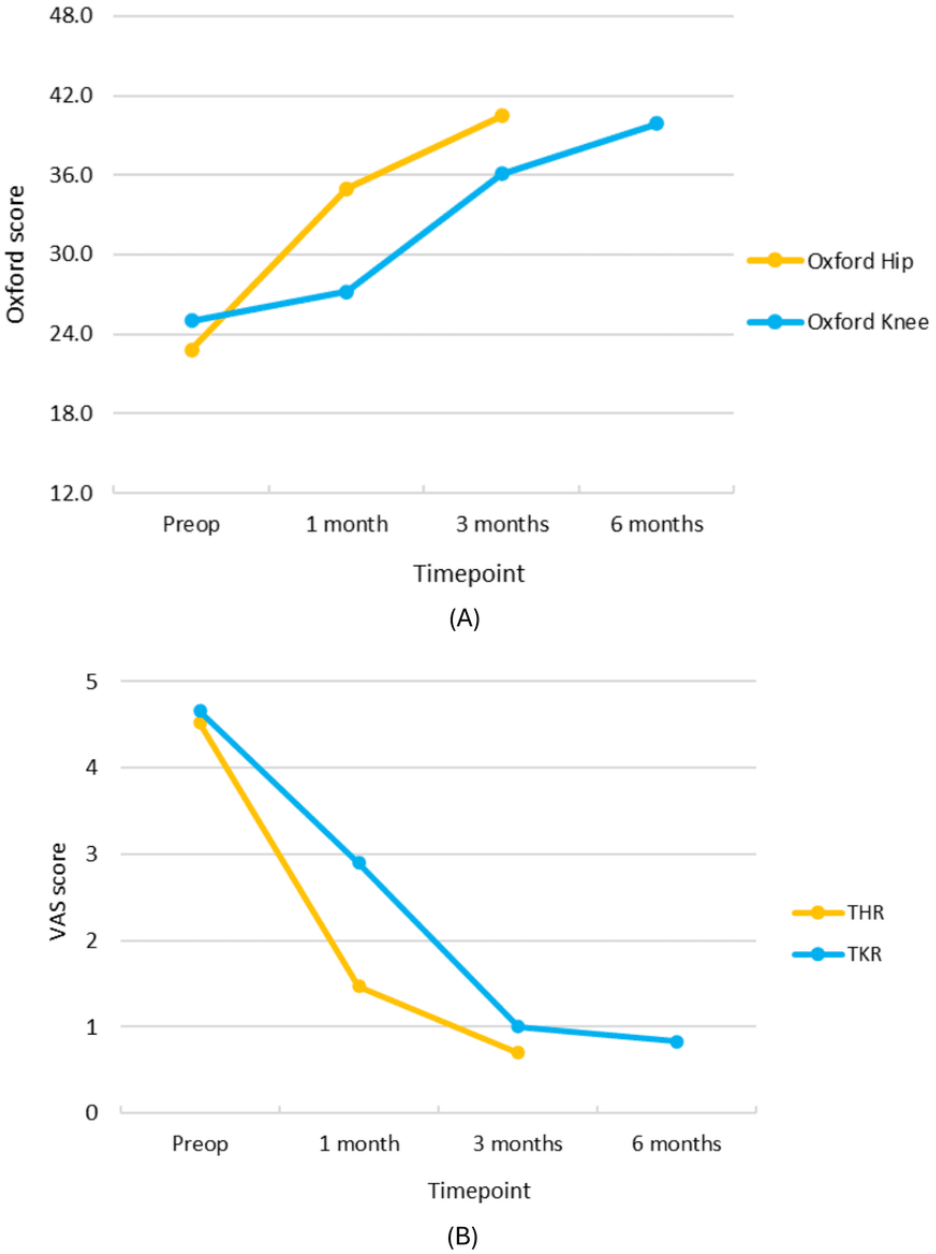
[e]Gap with surgery (days): mean 89.4 (SD 5.1, range 67-93).

[f]Gap with surgery (days): mean 180.1 (SD 0.4).

[g]Not available.

XSL•FO
RenderX

Figure 5. Pre- and postoperative (1, 3, and 6 mo) patient-reported outcome measure results by procedure: (A) Oxford scores and (B) VAS scores. THR: total hip replacement; TKR: total knee replacement; VAS: visual analog scale.



App Nondownload Rate

Of the 25 participants who did not download the app, 7 provided a reason. Their responses included technological concerns (n=2), feeling preoccupied and anxious preparing for the procedure (n=1), family issues (n=2), missing download instructions (n=1), and privacy concerns (n=1). The remaining 18 participants were also contacted on multiple occasions but never responded.

Discussion

Principal Findings

In this study, we investigated the acceptability of the use of the mymobility app for the postoperative care of 100 participants undergoing THA and TKA. We sought to investigate acceptability using the TFA to ensure all aspects of acceptability were measured.

Preoperatively, overall acceptability was satisfactory, with all but one TFA item scoring above 4 out of 5. The sole exception was burden (3.9/5), which may reflect lower familiarity with smartphone apps among older participants in our sample. We also found that higher education was associated with a higher level of self-efficacy. While general and overall acceptability were not significantly higher with higher education, this trend aligns with Wang et al's [15], Nuveo et al's [31], and Lee et al's [32] results, who described trends of higher engagement and adherence with higher education and lower receptivity to digital health technologies with lower education. Thus, these findings reveal the importance for app developers to consider this gap and tailor their content to diverse populations, notably

XSL•FO
RenderX

through the inclusion of images or videos, which are more easily accessible than heavy-text content.

When comparing pre- and postoperative TFAs, although overall scores were satisfactory at both time points, we found a significant decrease in overall acceptability postoperatively. We suspect this decrease is attributed to elevated expectations toward the app, which were adjusted following its use. However, we believe acceptability could also have been influenced by the natural process of recovery. Indeed, the decline in TFA was significant in TKA, but not in THA. PROM results showed the same trend, with no improvement until 3 months post operation in TKA as opposed to 1 month post operation in THA. This is consistent with Booth et al's [19] and Bourne et al's [33] results showing that TKA patients felt a higher need for in-person rehabilitation as opposed to THA and that short-term outcomes are superior in THA than in TKA. When considering the TFA items that decreased in TKA, which are linked to the perceived helpfulness of the app, this pushes us to believe that the decrease in postoperative acceptability was likely also driven by the surgical outcomes. This illustrates the possibility for app developers to better address the challenges in TKA rehabilitation, notably through adjustment of patient expectations following this procedure.

When looking at participants who did not download the app, reasons included concerns regarding technology in 2 cases, which might be due to the age of the participants (77-79 y). Additionally, 1 participant expressed privacy concerns. Although we explained how data collection and privacy functioned within the app, we still expected this number to be higher. This result is a positive sign for patient trust toward platforms used to provide care, although the need to discuss patient confidentiality within the app remains important.

Comparison With Previous Work

Multiple studies have investigated patient perspectives on the use of mHealth apps for joint arthroplasty. These studies all reported a high level of engagement, adherence, and satisfaction [13,15-19,31]. Factors that influenced the results positively included higher education and supportive environments, whereas those that did negatively included higher age, lower economic status, and worsened physical or psychological condition [15,17,31]. Patients also described better health care accessibility and reduced isolation with these interventions [20]. Our results showed that higher education was associated with greater acceptability. However, unlike prior work, we did not observe other demographic characteristics to significantly influence acceptability.

Strengths and Limitations

The main strength of this study is the measure of acceptability using a validated tool (TFA), which allowed us to assess all domains of this concept. Using this framework, we measured acceptability preoperatively, prior to exposure to the app, and postoperatively, after using the app. This allowed us to interpret how preexposure expectations were met following use of the app. Other strengths include the use of a single app, which reduced heterogeneity; the ITT design, which reflects real-world adherence; and the diverse population.

This study also has limitations. The main limitation is the absence of a control group, which limits the ability to attribute the results specifically to the app. This limitation is a result of ethical concerns brought by our institutional review board, which prohibited us from having a control group that could be denied a beneficial intervention. We also observed a high rate of unfilled questionnaires, which could have created a response bias. We believe this is due to most questionnaires being distributed by email or phone, creating a limiting step in communication. Other important limitations include the use of a single app, limiting the generalizability of the results; a potential selection bias toward patients willing to use the app; and statistical analysis being only performed with *t* tests, which might have limited the power of the conclusions. In addition, there was no assessment of baseline technological literacy, which could have affected results despite the app's technological guidance features.

Conclusion

There was a good level of acceptability with the use of mymobility for the postoperative management of THA and TKA, although acceptability decreased postoperatively. This decrease could signify high expectations toward the app preoperatively or higher than expected difficulty and pain in the early postoperative period. Higher education was associated with higher preoperative acceptability. TKA was associated with lower postoperative acceptability, which could be related to the delayed recovery in TKA when compared to THA. These 2 trends have previously been described in the literature, indicating potential gaps to address for app developers. We found a good level of satisfaction with the app, and Oxford scores showed better outcomes in THA than TKA, whereas VAS showed equally significant decreases in pain with both procedures. Privacy concerns were rare, with only 1 patient reporting them.

The impact of the natural recovery process in total joint arthroplasty on the acceptability of mHealth remains unclear. This gap could be addressed through a comparison of acceptability pre- and postoperatively in cohorts using the app and cohorts with standard care.

Acknowledgments

The authors sincerely thank St. Mary's Hospital for its support. The authors declare that no generative artificial intelligence was used in the preparation and writing of this article. All authors reviewed the results, approved the final version of the manuscript, and agreed to be accountable for all aspects of the work.

XSL•FO
RenderX

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

All data generated or analyzed during this study are included in this published article (and the multimedia appendices).

Conflicts of Interest

A Albers is a paid consultant for Zimmer-Biomet. The other authors declare no conflicts of interest.

Multimedia Appendix 1
Overview of data collected by the research team and by the mymobility app.
[DOCX File, 16 KB - humanfactors_v13i1e79682_app1.docx]

Multimedia Appendix 2
Overview of the mymobility functions and interface. (A) Features of the app with depictions of the app interface and (B) Walk-AI, the mymobility artificial intelligence–assisted gait speed analysis predicting and determining level of recovery based on gait speed.
[DOCX File, 640 KB - humanfactors_v13i1e79682_app2.docx]

Multimedia Appendix 3
Modified theoretical framework for acceptability questionnaire.
[DOC File, 51 KB - humanfactors_v13i1e79682_app3.doc]

Multimedia Appendix 4
Usefulness, Satisfaction, and Ease of Use questionnaire.
[DOC File, 37 KB - humanfactors_v13i1e79682_app4.doc]

Multimedia Appendix 5
Additional satisfaction questions.
[DOC File, 33 KB - humanfactors_v13i1e79682_app5.doc]

Multimedia Appendix 6
Oxford Hip and Oxford Knee Scores.
[DOC File, 82 KB - humanfactors_v13i1e79682_app6.doc]

Multimedia Appendix 7
Visual analog scale for pain.
[DOC File, 268 KB - humanfactors_v13i1e79682_app7.doc]

References

1. Hospital stays in Canada, 2021–2022. : Canadian Institute for Health Information; 2023 Feb 23 URL: https://bit.ly/3N6xCNM [accessed 2025-12-09]
2. Shichman I, Roof M, Askew N, et al. Projections and epidemiology of primary hip and knee arthroplasty in Medicare patients to 2040-2060. JB JS Open Access 2023;8(1):e22.00112. [doi: 10.2106/JBJS.OA.22.00112] [Medline: 36864906]
3. Aging and chronic diseases: a profile of Canadian seniors. : Public Health Agency of Canada; 2020 URL: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/aging-chronic-diseases-profile-canadian-seniors-report.html [accessed 2025-12-09]
4. Sherif A. Global smartphone penetration rate as share of population from 2016 to 2024. Statista. 2025. URL: https://www.statista.com/statistics/203734/global-smartphone-penetration-per-capita-since-2005/ [accessed 2025-12-09]
5. Wells C, Spry C. An overview of smartphone apps: CADTH horizon scan. : Ottawa (ON): Canadian Agency for Drugs and Technologies in Health; 2022 URL: https://www.ncbi.nlm.nih.gov/books/NBK595384 [accessed 2025-12-09]
6. Andrawis JP, Muzykewicz DA, Franko OI. Mobile device trends in orthopedic surgery: rapid change and future implications. Orthopedics 2016;39(1):e51-e56. [doi: 10.3928/01477447-20151228-01] [Medline: 26730684]
7. Dionisi S, Giannetta N, Di Simone E, et al. The use of mHealth in orthopedic surgery: a scoping review. Int J Environ Res Public Health 2021 Nov 28;18(23):12549. [doi: 10.3390/ijerph182312549] [Medline: 34886274]

XSL•FO
RenderX

8. Hayat J, Ramadhan M, Gonnah AR, Alfadhli A, Al-Naseem AO. The role of mobile health technology in perioperative spinal care: a systematic scoping review and narrative synthesis. Cureus 2024 Feb;16(2):e54254. [doi: 10.7759/cureus.54254] [Medline: 38496189]
9. Mapinduzi J, Ndacayisaba G, Verbrugghe J, Timmermans A, Kossi O, Bonnechère B. Effectiveness of mHealth interventions to improve pain intensity and functional disability in individuals with hip or knee osteoarthritis: a systematic review and meta-analysis. Arch Phys Med Rehabil 2025 Feb;106(2):280-291. [doi: 10.1016/j.apmr.2024.06.008] [Medline: 38945508]
10. Pritwani S, Shrivastava P, Pandey S, et al. Mobile and computer-based applications for rehabilitation monitoring and self-management after knee arthroplasty: scoping review. JMIR Mhealth Uhealth 2024 Jan 26;12:e47843. [doi: 10.2196/47843] [Medline: 38277195]
11. Dieter V, Janssen P, Krauss I. Efficacy of the mHealth-based exercise intervention re.flex for patients with knee osteoarthritis: pilot randomized controlled trial. JMIR Mhealth Uhealth 2024 Sep 9;12:e54356. [doi: 10.2196/54356] [Medline: 39250181]
12. Özden F, Sarı Z. The effect of mobile application-based rehabilitation in patients with total knee arthroplasty: a systematic review and meta-analysis. Arch Gerontol Geriatr 2023 Oct;113:105058. [doi: 10.1016/j.archger.2023.105058] [Medline: 37172329]
13. Monárrez R, Mohamadi A, Drew JM, Abdeen A. Mobile application's effect on patient satisfaction and compliance in total joint arthroplasty: a systematic review and meta-analysis. J Am Acad Orthop Surg Glob Res Rev 2023 Sep 1;7(9):e22.00200. [doi: 10.5435/JAAOSGlobal-D-22-00200] [Medline: 37678829]
14. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88. [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
15. Wang Q, Lee RLT, Hunter S, Zhu A, Chan SWC. Patient engagement in a mobile app-based rehabilitation program for total hip or knee arthroplasty: secondary data analysis of a randomized controlled trial. JMIR mHealth uHealth 2024 Oct 1;12:e57635. [doi: 10.2196/57635] [Medline: 39353187]
16. Hong M, Loeb J, Yang M, Bailey JF. Postoperative outcomes of a digital rehabilitation program after total knee arthroplasty: retrospective, observational feasibility study. JMIR Form Res 2022 Sep 19;6(9):e40703. [doi: 10.2196/40703] [Medline: 36121690]
17. Wang Q, Gao J, Zhu A, Cheng C, Chan SWC. Patient engagement in mobile technology-based rehabilitation for arthroplasty: a scoping review. J Clin Nurs 2025 Sep;34(9):3489-3503. [doi: 10.1111/jocn.17802] [Medline: 40346770]
18. Ramkumar PN, Haeberle HS, Ramanathan D, et al. Remote patient monitoring using mobile health for total knee arthroplasty: validation of a wearable and machine learning-based surveillance platform. J Arthroplasty 2019 Oct;34(10):2253-2259. [doi: 10.1016/j.arth.2019.05.021] [Medline: 31128890]
19. Booth MW, Riegler V, King JS, Barrack RL, Hannon CP. Patients' perceptions of remote monitoring and app-based rehabilitation programs: a comparison of total hip and knee arthroplasty. J Arthroplasty 2023 Jul;38(7S):S39-S43. [doi: 10.1016/j.arth.2023.04.032] [Medline: 37088222]
20. Wang Q, Lee RLT, Hunter S, Chan SWC. Patients' experiences of using a mobile application-based rehabilitation programme after total hip or knee arthroplasty: a qualitative descriptive study. BMC Nurs 2023 Jul 27;22(1):246. [doi: 10.1186/s12912-023-01409-3] [Medline: 37496003]
21. Perski O, Short CE. Acceptability of digital health interventions: embracing the complexity. Transl Behav Med 2021 Jul 29;11(7):1473-1480. [doi: 10.1093/tbm/ibab048] [Medline: 33963864]
22. Mymobility® with Apple Watch. Zimmer Biomet. 2023. URL: https://www.zimmerbiomet.com/en/products-and-solutions/zb-edge/mymobility.html#01-overview [accessed 2025-12-09]
23. Mymobility® Patient App. Google Play. URL: https://play.google.com/store/apps/details?id=com.mymobility [accessed 2025-12-09]
24. Privacy notice. Zimmer Biomet. 2023. URL: https://www.zimmerbiomet.com/en/corporate/privacy-notice/privacy-notice-english.html [accessed 2025-12-09]
25. PIPEDA requirements in brief. Office of the Privacy Commissioner of Canada. URL: https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/pipeda_brief/ [accessed 2025-12-09]
26. Sekhon M, Cartwright M, Francis JJ. Development of a theory-informed questionnaire to assess the acceptability of healthcare interventions. BMC Health Serv Res 2022 Mar 1;22(1):279. [doi: 10.1186/s12913-022-07577-3] [Medline: 35232455]
27. Lund AM. Measuring usability with the USE questionnaire. Usability Interface 2001;8(2):3-6 [FREE Full text]
28. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. J Bone Joint Surg Br 1996 Mar;78(2):185-190. [doi: 10.1302/0301-620X.78B2.0780185] [Medline: 8666621]
29. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. J Bone Joint Surg Br 1998 Jan;80-B(1):63-69. [doi: 10.1302/0301-620X.80B1.0800063]
30. Kersten P, White PJ, Tennant A. Is the pain visual analogue scale linear and responsive to change? An exploration using Rasch analysis. PLoS ONE 2014;9(6):e99485. [doi: 10.1371/journal.pone.0099485] [Medline: 24921952]

XSL•FO
RenderX

31. Nuevo M, Modrego A, Rodríguez-Rodríguez D, et al. The influence of sociodemographic and health factors on adherence to home-based rehabilitation after fast-track total knee arthroplasty: secondary analysis of a randomized controlled trial. Disabil Rehabil 2025 Sep;47(18):4728-4737. [doi: 10.1080/09638288.2025.2458191] [Medline: 39879639]
32. Lee CMY, Thomas E, Norman R, et al. Educational attainment and willingness to use technology for health and to share health information - the reimagining healthcare survey. Int J Med Inform 2022 Aug;164:104803. [doi: 10.1016/j.ijmedinf.2022.104803] [Medline: 35644052]
33. Bourne RB, Chesworth B, Davis A, Mahomed N, Charron K. Comparing patient outcomes after THA and TKA: is there a difference? Clin Orthop Relat Res 2010 Feb;468(2):542-546. [doi: 10.1007/s11999-009-1046-9] [Medline: 19760472]

Abbreviations

mHealth: mobile health
OHS: Oxford Hip Score
OKS: Oxford Knee Score
PROM: patient-reported outcome measure
TFA: theoretical framework for acceptability
THA: total hip arthroplasty
TKA: total knee arthroplasty
USE: Usefulness, Satisfaction, and Ease of Use
VAS: visual analog scale

Edited by A Kushniruk; submitted 29.Jun.2025; peer-reviewed by W Glinkowski, X Zhang; revised version received 23.Oct.2025; accepted 20.Nov.2025; published 27.Jan.2026.

Please cite as:
Louni Y, Laroche M, Alnasser A, Abuhaneya M, Belzile E, Baskaran S, Mutch J, Albers A
Acceptability of a Digital Care App in Patients Undergoing Hip and Knee Arthroplasty: Prospective Cohort Study
JMIR Hum Factors 2026;13:e79682
URL: https://humanfactors.jmir.org/2026/1/e79682
doi:10.2196/79682

© Yacine Louni, Matthew Laroche, Abdulrhman Alnasser, Mohammad Abuhaneya, Eric Belzile, Sandhya Baskaran, Jennifer Mutch, Anthony Albers. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 27.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Providing Psychological Therapy Support and Improving Postdischarge Data Collection: Preliminary Evaluation of the "Paddle" App

Graham R Thew[1,2,3], DPhil; Michelle Lee[4], PhD; Ineke Wolsey[5], MA; Fiona Giles[2], PGDip; Emma Hayden[6], BSc; Grace Jell[6], PhD; Jamie Peacock[3], MSc; John Pimm[2], PhD; Joanne Ryder[3], MSc; Alison Salvadori[6], DClinPsy; David M Clark[1], DPhil

[1]Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom
[2]NHS Buckinghamshire Talking Therapies, Oxford Health NHS Foundation Trust, High Wycombe, United Kingdom
[3]NHS Oxfordshire Talking Therapies, Oxford Health NHS Foundation Trust, Oxford, United Kingdom
[4]Charlie Waller Institute, School of Psychology and CLS, University of Reading, Reading, United Kingdom
[5]Brighter Futures Together CIC, Reading, United Kingdom
[6]NHS Berkshire Talking Therapies, Berkshire Healthcare NHS Foundation Trust, Reading, United Kingdom

Corresponding Author:
Graham R Thew, DPhil
Department of Experimental Psychology
University of Oxford
Life and Mind Building
South Parks Road
Oxford, OX1 3EL
United Kingdom
Phone: 44 01865 281607
Email: graham.thew@psy.ox.ac.uk

Abstract

Background: Clinical trials of psychological therapies such as cognitive behavioral therapy typically show sustained posttreatment effects. However, less is known about individuals' outcomes following treatment in routine practice, and additionally, patient representatives have highlighted a need for better postdischarge support.

Objective: The "Paddle" app aims to address these issues. Paddle allows patients to store their therapy materials, reflect on sessions, and monitor their well-being during and after treatment, completing outcome questionnaires and monthly reviews.

Methods: Study 1 evaluated patients' use of Paddle during treatment in routine practice. Patient and therapist surveys explored feasibility, acceptability, and helpfulness. Study 2 examined the feasibility of using the app post discharge, inviting users to submit monthly questionnaires for 6 months, and seeking feedback via survey.

Results: Our findings indicate that Paddle was feasible to implement with few technical problems. Although not all patients wanted to use an app in Study 1, uptake was 66% (111/168) and users found it acceptable and helpful for organizing and remembering therapy information. Most reported using the app on a weekly or fortnightly basis. In Study 2, a total of 321 patients downloaded and used the app at least once, of whom 49% (156/321) submitted follow-up data. Of those who reliably improved during treatment, 73% (86/118) remained so throughout the 6-month follow-up period. Among all users, 20% (31/156) showed further reliable improvement at least once compared to their end-of-treatment score. We introduce the concept of "reliable relapse," which occurred for 36% (32/89) of users who had reliably recovered during treatment, highlighting that some experience fluctuations or deterioration. Feedback highlighted Paddle's value in helping people self-monitor and prioritize their well-being after treatment, with 81% (50/62) suggesting a postdischarge follow-up period longer than 6 months would be helpful.

Conclusions: These preliminary findings suggest that Paddle shows promise in supporting patients to collate therapy resources and monitor their well-being during and after treatment. It may help to improve rates of follow-up data collection, which warrants further investigation.

(JMIR Hum Factors 2026;13:e68671) doi:10.2196/68671

XSL•FO
RenderX

KEYWORDS

anxiety; app; cognitive behavioral therapy; depression; NHS Talking Therapies; outcome measurement

Introduction

Psychological therapies, such as cognitive behavioral therapy (CBT), have a strong evidence base demonstrating their efficacy in the treatment of anxiety and depression [1-3]. They are recommended as first-line treatment options by health guidance organizations such as the National Institute for Health and Social Care Excellence [4,5], due to there being stronger evidence for sustained gains compared to other treatments such as medication [6]. Evidence from randomized controlled trials suggests that, on average, the clinical gains made during treatment are well-maintained after treatment ends, with studies typically showing sustained effects in follow-up assessments 12-24 months posttreatment [7-9]. However, although improvements are maintained on average, some people show further gains, and some people show deterioration during the follow-up period.

Services providing such therapies within routine clinical practice are often limited in their ability to provide posttreatment support to patients, and to monitor the outcomes of those after they have been discharged. In England, data from the National Health Service Talking Therapies programme for anxiety and depression (NHS TT) in 2022-2023 showed that only 1.3% of patients received a follow-up appointment after they had been discharged [10]. This is perhaps not surprising because these services are not commissioned to collect follow-up data, but it does highlight that we know little about how such patients progress after leaving treatment, and that efforts to increase postdischarge data collection may be valuable for our understanding of the maintenance of treatment gains in routine practice.

A small number of studies have examined posttreatment relapse rates in cohorts of patients who had recovered following low-intensity CBT. Observed relapse rates included 29% in the first 9 months [11] and 53% within 12 months [12]. Factors increasing the likelihood of relapse have included residual depression symptoms at the end of treatment, younger age, unemployment, and nonlinear treatment responses [12,13]. Such studies provide helpful indications about which patients may be at greater risk of relapse. They also highlight areas where further research is needed. For example, studies to date have focused only on low-intensity treatment, so our understanding of posttreatment outcomes across whole services (which include high-intensity treatments and sequential low- and high-intensity interventions) is limited. Studies to date have focused only on patients who recovered during treatment, meaning less is known about those who showed reliable improvement in symptoms but did not fully recover. Finally, the time-to-relapse analyses used in most studies to date place less focus on what happens after a relapse, suggesting there is more to understand about the duration of relapses and the frequency of fluctuating symptom trajectories in this period. Furthering our knowledge of patient outcomes in the posttreatment period is an important step toward better tailoring of treatments or developing targeted posttreatment support to those who may particularly benefit. This might reduce the need for patients to return to services [14], enhancing clinical effectiveness and the efficient use of services' resources.

Importantly, the need to improve posttreatment support and prevent relapse has also been identified as a priority by patients. Patient and public involvement (PPI) on this topic in the Thames Valley region led to the development of the "Staying Well" guidance and support package that has been recommended for use as a component of low-intensity interventions in NHS TT services (materials available at Oxford Centre for Anxiety Disorders and Trauma, 2024) [15]. PPI contributors also highlighted that therapies such as CBT often result in a large number of materials, such as worksheets, diaries, flashcards, pictures, and diagrams, and that it can be challenging to organize and store these in ways that are accessible and useful both during therapy and afterward. Many had kept a folder or notebook but found that if they were out or away from home, they were not able to remind themselves of the key points or reminders they had generated from treatment. As one contributor put it, "a folder is no good if you are having a panic attack on the bus."

These discussions led to the development of an app ("Paddle"; Global Initiative Ltd) as a collaboration between PPI contributors, regional services, clinicians, and software developers [16]. Digital tools have been suggested as a helpful method to enhance the delivery of psychological therapies [17] and their role in preventing relapse has been identified as a priority research area [18]. Paddle is a secure, personal storage app in which a patient can add and organize their materials from therapy, as well as keep notes or reflections about their sessions. The app also includes easily accessible emergency and crisis information. Finally, Paddle allows patients to complete routine outcome measures within the app, sending these results directly to services, and presenting graphs and information about these questionnaires to the patient, to support them in monitoring their own well-being both during and after treatment. It is hoped that Paddle represents a useful and effective adjunct to standard psychological therapy interventions, which may support patients in organizing, accessing, and remembering their therapy materials during and beyond treatment.

This paper reports the findings of 2 studies evaluating the use of Paddle as part of routine service provision in 3 NHS TT services. The research questions were as follows:

1. Is Paddle feasible to use alongside standard psychological therapy interventions?
2. Do patients and therapists find the app acceptable?
3. Is it feasible to use Paddle to collect outcome data following the end of treatment?
4. What are patients' views on using Paddle to monitor their own well-being after discharge?

XSL•FO
RenderX

Study 1

Methods

Design

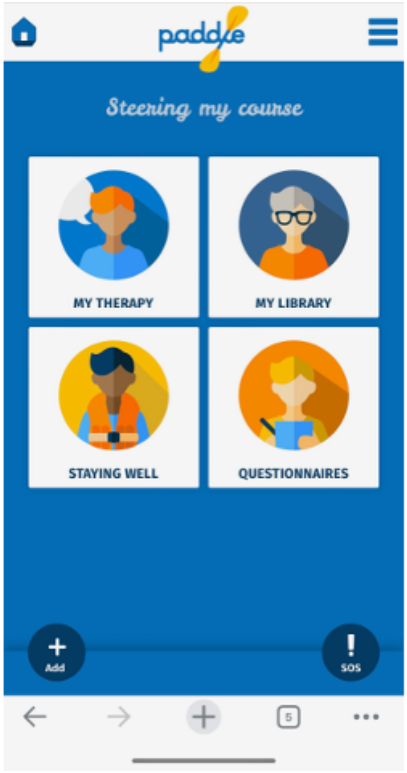
This was a survey-based study of participants starting treatment in the participating NHS TT services (Berkshire, Buckinghamshire, and Oxfordshire) whose therapist introduced and discussed Paddle with them. Both low- and high-intensity interventions were included.

App Description

Paddle is a prescribed (invite-only) app available to patients receiving treatment for anxiety and/or depression in NHS TT services. It is designed to be used solely by the patient as an adjunct to their psychological therapy intervention at low or high intensity, allowing them to add, organize, and store their therapy-related information in one secure location. These materials are then easily accessible for the patient both during treatment and afterward. It does not therefore deliver therapy content but is an adjunctive tool designed to support psychological interventions such as CBT. Paddle meets all relevant technical and data security standards (DCB0129 of the NHS Data Coordination Board, ISO [International Organization for Standardization] 9001, and ISO 27001). Since the completion of the studies described here, Paddle has been approved and registered as a Class I medical device.

During Study 1, the app allowed patients to upload information to 3 areas: "My Therapy," where patients could enter notes, actions, reflections, and materials from their sessions; "My Library," where patients could develop a bespoke set of resources, including images, audio files, documents, websites, and telephone numbers; and "Staying Well," where patients could add information and resources for emergencies, setbacks, or difficult moments (eg, a safety plan, relapse prevention plan, or other key points to remind themselves of during and after treatment). In addition, an "SOS" section of the app allows patients to easily access their most important resources when needed. Figure 1 shows a screenshot of the app homepage [16].

Figure 1. Screenshot of a phone showing the Paddle app home page.



Participants

A group of 18 therapists (12 psychological well-being practitioners and 6 CBT therapists) across the 3 participating services attended a half-day in-person training session to learn about the aims, features, and functionality of the app, how to introduce Paddle to their patients, and how to generate an app invite for patients.

These therapists then discussed Paddle with eligible patients on their caseload. The only exclusion criteria were patients using an internet-based platform to undertake treatment, for example, SilverCloud (SilverCloud Health) [19], iCT-SAD (internet-delivered cognitive therapy for social anxiety disorder) [20] or iCT-PTSD (internet-delivered cognitive therapy for posttraumatic stress disorder) [21], or those attending a brief psychoeducation-only group.

Paddle was discussed with 203 patients, and 168 (83%) agreed to be sent an invite to download the app. Of these, 111 (66%) downloaded and logged into the app at least once.

Procedure

Therapists from the participating NHS TT services who attended the training were asked to introduce the Paddle app to each new patient during the first or second treatment session between

XSL•FO
RenderX

October 2019 and February 2020. Patients were signposted to the Paddle website for further information and sent a sheet of frequently asked questions about the app. Therapists then asked patients if they wished to be sent an invite to download and use the app. Those who agreed and downloaded the app were then able to use it as they wished for the duration of their treatment. Therapists were not required to prompt discussion or use of Paddle but could do so if clinically relevant.

At the point of discharge from treatment, the therapist invited the patient to complete an online survey (using the Qualtrics survey platform; Qualtrics LLC). All 203 patients were invited. A branching survey design was used to present questions relevant to respondents. For those who had requested an invite to download Paddle, questions were asked about their use of and experience with the app. For those who declined to be sent an invitation, they were invited to share feedback about their reasons for doing so. Response types on the survey therefore included Likert and frequency response scales, multiple choice, and free-text questions.

At the end of the study period, the participating therapists were also invited to complete an online survey to gather their impressions of Paddle based on their interactions with patients, and record any technical issues they had experienced when sending invites.

Data Analysis

Responses to both the patient survey and therapist survey were analyzed descriptively.

Ethical Considerations

The study was registered and approved as a pilot service evaluation project with the research and development departments of the participating NHS trusts (Oxford Health NHS Foundation Trust and Berkshire Healthcare NHS Foundation Trust). Ethics approval was sought and granted for this process within the University of Reading (Reference UREC 19/02). All participants provided consent to participate and for their data to be used to support service evaluation and improvement. Any identifiable data was removed prior to analysis. Compensation was not provided.

Results

Patient Survey

A total of 68 responses were received, giving an overall response rate of 33% (68/203). Respondents comprised 51 people who had logged in at least once and 17 who declined to be invited to Paddle. The response rate of those who had logged in was therefore 46% (51/111), and of those who declined was 49% (17/45). There were no responses from patients who had received an invite to download the app but had not logged in.

The age range of Paddle users, as self-reported on the survey, was 16-59 years, with the most common reported age range being 20-29 years. The age profile of nonusers was older, ranging from 20 to 69 years, with the most common reported age range being 60-69 years. Overall, 70% of this sample were 50 years or older, compared with 10% of the sample who had chosen to download the app.

Paddle users all reported using other apps in everyday life, with 69% (35/51) using other health-related apps. On average, they rated their confidence in using health-related apps as 78/100 (SD 11.61). In contrast, only 12% (2/17) of the nonusers group reported using other apps, with only one person using a health-related app. Mean confidence was 10/100 (SD 2.89).

The majority (45/51, 88%) of Paddle users indicated that they had downloaded the app because their clinician had recommended it, and 63% (32/51) because they wanted to have their therapy materials together and to hand. Paddle nonusers gave 2 main reasons for declining to use the app: a lack of confidence with this type of technology and concerns about information security.

Paddle users reported using all sections of the app, including “My Therapy” (46/51, 90%), “My Library” (36/51, 71%), and “Staying Well” (40/51, 78%). Most reported using the app on a weekly (29/51, 57%) or fortnightly (12/51, 24%) basis during treatment, which likely reflected use in relation to each therapy session. Some users used the app more than once per week (8/51, 16%) and some less than fortnightly (2/51, 4%). Users’ agreement ratings with statements about Paddle are shown in Table 1. These reflected positive feedback regarding the feasibility, acceptability, and helpfulness of the app. Free-text comments indicated that users liked that this app had been developed by patients for patients, and that everything was in one place to refer back to. Various comments indicated the app had also been helpful in organizing users’ thinking: for example, “Summarising the sessions in the app helped me to focus and remember the techniques,” “It was really helpful writing a summary from each session to get my thoughts together,” “It’s helped me to organise my thoughts, I can look back on this when I need to,” “It helped me to think about what I needed to focus on and what I’d learnt.” Comments also indicated that users felt the app will be a helpful resource to use following the end of treatment.

XSL•FO
RenderX

Table 1. Number of Paddle users agreeing with statements about the app.

Statement	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Paddle has helped me make the most of treatment	0	1	13	20	17
Paddle has been useful to me	0	1	12	19	19
Paddle has been straightforward to use	0	0	13	22	16
I've liked using Paddle	0	1	14	19	18
I plan on using Paddle now treatment has ended	0	5	12	23	11
I would recommend Paddle to others having treatment for anxiety or depression	0	1	14	21	15

Therapist Survey

Eleven responses to the therapist survey were received, a response rate of 61%. Responses came from 7 psychological well-being practitioners and 4 CBT therapists. Therapists' comments highlighted patients' interest in the app and willingness to receive an invite to download it. They noted the value of discussing it together with the patient ("Explaining about Paddle in the session is helpful – especially if there are questions"). Consistent with findings from the patient survey, age and technological confidence were reported as factors that might be associated with not wanting to use the app ("They might not be used to using apps"; "It's been my older clients who are more hesitant"). No major technical problems or barriers to use were reported. Therapists' impressions of their patients' experiences using the app alongside treatment were positive ("I had one client who said how useful it was to have a place to store everything together," "Clients seem to like it and are finding it useful," "I've had very positive feedback about it," "Clients were definitely using it between sessions to record info and help them prepare for subsequent sessions").

Discussion

Principal Findings

The findings of this study provided an initial indication that it was feasible for therapists to introduce and discuss Paddle with their patients, and for patients to download and use the app during treatment. No major technical issues were observed. Acceptability was indicated by the high rate (168/203, 83%) of those willing to receive an invite to the app, and of those who downloaded and used the app at least once (66% of those invited, 111/168). Most patients reported using the app regularly during treatment. The positive feedback provided an initial indication that the app was valued and perceived as helpful by the majority of respondents, though we cannot say whether nonrespondents had similar views. The findings also suggest that not everyone found the idea of using an app alongside therapy to be helpful, which may be associated with technological confidence and concerns about information security.

These initial findings that the app could be implemented during treatment and that it could be a helpful tool for patients indicated that progression to the next phase of development was warranted. This involved adding the functionality to complete outcome questionnaires within the app.

Study 2

Methods

Design

This was a single-cohort observational study, examining (1) whether the app could be used to collect outcome data following discharge and (2) whether users found using the app helpful to self-monitor their well-being in this phase. Unlike in Study 1, we invited patients at the point of discharge to download and use Paddle to support their posttreatment well-being. All features of the app were available, including the functionality to complete routine outcome measures within the app. Patients were invited to complete these questionnaires on a monthly basis for the first 6 months following discharge.

App Description

The app was identical to that used in Study 1, with the addition of a "Questionnaires" area, which allowed users to complete the same routine outcome measures they had used during treatment. It also provided users with a brief explanation of each questionnaire and what it measures. After completing a set of questionnaires, users could view graphs and tables of their scores over time, allowing them to compare these with their scores at the start and end of treatment. Based on the trajectory of the 3 most recent scores on each questionnaire, and the current score in relation to the standard clinical caseness threshold for that measure (described in National Collaborating Centre for Mental Health, 2018) [22], Paddle provided brief feedback and using short question prompts, encouraged users to reflect on what had been going well or less well and to continue building on strategies they were finding helpful, referring to their "Staying Well" materials as required.

Users' submitted questionnaire data were automatically integrated with the electronic care record used by the relevant NHS TT service. It was explained to users that because they were not currently under the care of the service (having been discharged), the scores they submit to the service were for information only, and if they wanted to access additional support, they should contact the service or other health care providers as normal.

One of the routine outcome measures, the Patient Health Questionnaire 9-item version (PHQ-9) [23], contains an item relating to risk of suicide or self-harm. It was agreed with services that in the course of looking to improve follow-up data collection after discharge, there is a responsibility to respond

XSL•FO
RenderX

appropriately if significant risk is indicated. Based on discussion with clinicians, service managers, and patient representatives, we established the following procedure for this study. If any risk was indicated on the PHQ-9 (ie, a response >0 on the self-harm and suicidality item), the app presented 4 brief additional questions to the user, exploring plans, preparations, behavioral intent, and protective factors. All PHQ-9 data, plus any responses to the additional questions, were sent directly to the relevant service. A nominated clinician in each service reviewed these on a regular basis in conjunction with the clinical notes. Based on their clinical judgment and local service procedures, the clinician would contact the patient if necessary to conduct further risk assessment and management. Users were informed that services are not able to respond to mental health crises and provided with a list of emergency and crisis service details for them to contact if needed.

Participants

Patients were eligible if they reported a main problem of depression or an anxiety disorder and had received at least 2 therapy sessions, at low- or high-intensity. NHS TT services deliver National Institute for Health and Care Excellence (NICE)-recommended therapies, most commonly CBT but also approaches such as counselling or interpersonal therapy. They were not eligible if their last course of treatment prior to discharge used an internet-based treatment platform, because these treatments typically allow users to continue to access the treatment material for a fixed period after discharge. Patients were also not eligible if they were being referred on for further treatment in another service.

Across the 3 participating NHS TT services, 7714 patients were discharged within the 13-week study recruitment period (October 2021 to January 2022), of whom 4063 were eligible and invited to participate. Of these, 321 patients (8%) downloaded and used the app at least once.

Procedure

All clinical staff received a brief training webinar explaining about Paddle and how to introduce it to eligible patients. Where possible, the treating therapist introduced and explained Paddle during the final treatment session and sent the invitation to download the app. However, in most cases, this was not possible, so a designated clinician in each service reviewed the list of recently discharged patients and sent the app invite by email.

Paddle users were free to use all areas of the app as they wished. They received a monthly email invitation to submit their questionnaires, with a reminder 1 week later. At the end of their 6-month period post discharge, anyone who had logged into Paddle at least once was invited to complete a feedback survey via the Qualtrics survey platform. Questions focused on the use of the questionnaire area of the app, and the experience of using Paddle post discharge.

Data Analysis

App usage, clinical outcomes, and user feedback were analyzed descriptively. The demographic and clinical representativeness of app users was analyzed using 2-tailed *t* tests and chi-squared tests.

Measures

As is standard in NHS TT services, all participants were given standardized and well-validated measures of depression (PHQ-9; Kroenke et al [23]), generalized anxiety (Generalized Anxiety Disorder 7-item version; GAD-7; Spitzer et al [24]), and general functioning (Work and Social Adjustment Scale; WSAS; Mundt et al [25]). Where a participant's treatment focused on a specific anxiety disorder, Paddle administered a relevant additional measure, matching that which was used during treatment. Anxiety disorder specific measures (ADSM) used in this study were Health Anxiety Inventory [26], Obsessive Compulsive Inventory [27], Posttraumatic Stress Disorder Checklist for *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders* [Fifth Edition]) [28], Panic Disorder Stress Scale [29], Patient Health Questionnaire 15 item [30], and Social Phobia Inventory [31]. These measures are also well validated and used as standard within NHS TT services [22]. Participants were also asked to provide their employment status and indicate their use of health care services in the past month.

Classification of Clinical Outcomes

Clinical outcomes were categorized following standard NHS TT definitions, which are calculated based on changes in the PHQ-9 and GAD-7 scores. Where ADSM scores are available, these are used in place of the GAD-7. Reliable improvement is indicated where scores on the PHQ-9 and/or GAD-7/ADSM decrease by an amount greater than the reliable change index for the measure, and neither measure shows a reliable increase. Reliable deterioration is indicated by an equivalent increase in scores, with neither measure showing a reliable decrease. Reliable recovery is indicated where a patient starts treatment above the clinical caseness threshold on the PHQ-9 and/or the GAD-7/ADSM, shows reliable improvement, and has an end score below caseness on both measures. For further details, see the NHS TT manual [22].

For this study, we developed the concept of "reliable relapse," which is the conceptual opposite of reliable recovery. Reliable relapse is indicated where a patient is below caseness on both PHQ-9 and GAD-7/ADSM, shows reliable deterioration, and has an end score that is above caseness on either or both measures. By requiring that scores must have shown a deterioration that is clinically meaningful and more than might be expected by measurement error alone (ie, reliable deterioration), this definition avoids classifying a small change from just below to just above the clinical cutoff as a relapse.

Ethical Considerations

The study registration and ethical considerations were identical to Study 1. The study was registered and approved as a pilot service evaluation project with the research and development departments of the participating NHS trusts (Oxford Health NHS Foundation Trust and Berkshire Healthcare NHS Foundation Trust). Ethics approval was sought and granted for this process within the University of Reading (Reference UREC 19/02). All participants provided consent to participate and for their data to be used to support service evaluation and

XSL•FO
RenderX

improvement. Any identifiable data were removed prior to analysis. Compensation was not provided.

Results

Use of the App to Complete Questionnaires

Of the 321 patients who downloaded and used the app at least once, 156 (49%) submitted at least one set of questionnaires and thus provided follow-up data. No major technical problems occurred. Patterns of data completion across the 6-month follow-up period were variable: 30 (19%) users provided data in all 6 months, 17 (11%) provided data on 5 occasions, 19 (12%) on 4 occasions, 17 (11%) on 3 occasions, 20 (13%) on 2 occasions, and 53 (34%) on 1 occasion. ADSMs were used by 17% of those who submitted questionnaires.

A total of 487 PHQ-9 questionnaires were completed during the study, of which 78 (16%) included a score of 1 or more on the risk item, and 23 (5%) showed an increase in score compared to the last available datapoint. Upon review of the PHQ-9 and additional risk questions by the designated clinician in each service, seven users were contacted for further risk assessment and management as appropriate.

Representativeness of App Users

We compared the demographics of Paddle users with those who were invited but did not log in. No significant differences were observed with respect to age (t_{4060}=–0.031; P=.98, or gender: χ^2_2=1.5; P=.47). The proportion of those from non-White ethnic backgrounds (ie, selecting options other than “White British,” “White Irish,” and “White Other”) was lower among Paddle users (10.26%) compared to nonusers (14.36%; χ^2_1=3.9; P=.049). The same pattern of results was shown when comparing Paddle users with all other people who were discharged within the study recruitment period: age (t_{7711}=–0.084; P=.93; gender: χ^2_2=1.8; P=.41; ethnicity: χ^2_1=6.5; P=.01).

To compare treatment characteristics between Paddle users who submitted follow-up data and those who did not, we compared the number of treatment sessions received. Users who provided follow-up data had received a greater number of sessions (mean 15.05, SD 10.89) compared to those who did not (mean 11.64, SD 7.85; Mann-Whitney U=9858.50; P=.004).

Clinical Outcomes

Clinical outcomes were explored for the 156 Paddle users who submitted follow-up questionnaires. We started by examining their clinical outcomes during treatment. The rate of reliable improvement during treatment was 76% (118/156), and reliable deterioration was 4% (7/156). Reliable recovery was shown by 61% of users during treatment (89/146 users who started treatment above caseness). The mean PHQ-9 score was 14.45 (SD 5.66) at the start of treatment, and 7.24 (SD 5.18) at the end of treatment. The mean GAD-7 score was 12.47 (SD 4.97) at the start of treatment, and 6.21 (SD 4.57) at the end of treatment.

An average of the PHQ-9 and GAD-7 scores provided by each user during the 6-month follow-up period was calculated as an indicator of symptom severity across this phase. The mean PHQ-9 score during follow-up was 8.40 (SD 5.21), and the mean GAD-7 score was 7.31 (SD 4.99). Paired t tests indicated that the follow-up scores were an average of 1.2 points higher on the PHQ-9 compared to the end of treatment (t_{155}=–3.81; P<.001, Cohen d=0.31) and 1.1 points higher on the GAD-7 (t_{154}=–3.85; P<.001; Cohen d=0.31). Although significant, the observed effect sizes suggested these increases were considerably smaller than the decreases seen when comparing the follow-up scores to the start of treatment scores (PHQ-9 t_{155}=13.62; P<.001; Cohen d=1.09; GAD-7 t_{154}=12.01; P<.001; Cohen d=0.96).

We then looked at individual change trajectories over the follow-up period, comparing each monthly follow-up score to the start of treatment score. First, we examined the 118 users who had reliably improved during treatment. This showed that 86 (73%) users stayed reliably improved across the 6-month period, 26 (22%) had at least one score that no longer reflected reliable improvement, and 4 (3%) showed reliable deterioration at at least one point. Two users (2%) had insufficient data for this analysis. The results also showed that of the 29 people in this group who had reliably improved but not reliably recovered during treatment, 5 (17%) went on to demonstrate reliable recovery during the follow-up.

Second, we examined those who had not reliably improved during treatment (n=38). We found that 7 users (18%) went on to reliably improve or reliably recover and remained so throughout follow-up. Three (8%) users showed reliable improvement or reliable recovery that was not sustained. Nine (24%) users showed no reliable changes in score throughout follow-up, and the remaining 19 users (50%) showed reliable deterioration at any point.

Third, we examined the group of users who had reliably recovered during treatment (n=89). There were 46 (52%) users who remained reliably recovered throughout the 6-month period, and a further 20 (22%) who remained either reliably recovered or reliably improved at each follow-up timepoint. Conversely, 18 (20%) users had at least one score that no longer reflected reliable improvement, and 4 (4%) showed reliable deterioration at at least one follow-up occasion. One (1%) user had insufficient data for this analysis. For this group, we also examined rates of reliable relapse by comparing each monthly follow-up score to the end-of-treatment score. This showed that 32 users (36%) showed a reliable relapse at any point during follow-up. In 9 of these instances, subsequent available data showed that the relapse was not sustained.

Finally, we examined changes within the follow-up period by comparing each monthly follow-up score to the end-of-treatment score for all cases (n=156). This showed that compared to when users ended treatment, 11 (7%) showed reliable improvement at every available timepoint during follow-up. Nine users (6%) showed a series of follow-up scores that reflected either reliable improvement or stability (ie, no reliable change compared to the end of treatment). Moreover, 72 (46%) showed stability at each available timepoint. In addition, 29 (19%) showed a series of follow-up scores that reflected either stability or reliable deterioration, and 23 (15%) showed reliable deterioration at

XSL•FO
RenderX

every available timepoint during follow-up. There were 11 (7%) users with a fluctuating series of scores that included both reliable improvement and reliable deterioration, and 1 (1%) user with insufficient data for this analysis.

Employment and Health Care Use

There were 109/156 (70%) Paddle users who were in employment at the start of their treatment. Of these, 102 (94%) remained employed across the treatment and follow-up period (1 moved to sickness or disability benefits, 2 experienced a temporary period of unemployment, 2 became students, 1 retired, and 1 was unemployed at the end of follow-up). There were 11/156 (7%) users who were unemployed at the start of their treatment, 7 (64%) of whom moved into employment by the end of treatment, and remained in employment or study throughout the follow-up period.

Of the 153 users who provided information on their health care use at any point during follow-up, 45 (29%) reported accessing further health care services in the month prior to completing the questionnaire. Examples included general practitioner support, voluntary helplines, private psychological therapy, or other postdischarge support from NHS TT, such as online peer support (Browne et al [32]) or an additional booster session.

Patient Survey

Of the 321 people who had used Paddle at least once, 62 (19%) responded to the feedback survey. Most reported they had downloaded Paddle because they thought it might be a helpful tool to store their information, reflect on progress, and stay well following treatment. Nearly all respondents had used Paddle to check their latest scores on the questionnaires (95%), with the majority also checking how their scores had changed over time (89%), and reviewing their progress using the monthly prompt questions (61%). Three respondents reported minor technical problems experienced when logging into the app, but noted these were eventually resolved. No other technical problems were reported. When asked about the duration of the follow-up questionnaire period, 19% of respondents felt that 6 months was appropriate, 47% said they would prefer 12 months, and 34% felt the monthly questionnaire prompts should continue indefinitely.

Users' agreement ratings with statements about Paddle are shown in Table 2. These reflected positive feedback regarding the feasibility, acceptability, and helpfulness of the app. Free-text comments highlighted the app's role in supporting users to monitor their well-being, for example: "It helped me to keep an eye on myself following the end of therapy by completing the same questions I'd done beforehand," "It's helped me to check in regularly when life's been busy and make sure the new habits I'd built in therapy had stuck," "It's been a great resource to keep my mental health a top priority," "The app has been really helpful – and I am of the older generation who didn't grow up with the internet and this technology, so that's really a tribute to the Paddle app."

Table 2. Number of Paddle users agreeing with statements about the app.

Statement	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Paddle has been straightforward to use	0	1	16	25	20
Paddle helped me to stay well	0	1	19	26	16
I would recommend Paddle to other people who've had treatment for anxiety or depression	0	1	16	27	18

General Discussion

Principal Findings

Taken together, the 2 studies reported here demonstrated preliminary evidence that the Paddle app is feasible to implement and acceptable to patients, with good potential to support patients during and after treatment and collect postdischarge outcome data. Feasibility was indicated by the successful use of Paddle within NHS TT services, with minimal technical problems evident across the stages of inviting participants, downloading and using the app, submitting data and integrating this with services. The app was also feasible from a risk management perspective, with services only needing to respond to indications of risk on 7 occasions.

Acceptability was indicated in Study 1 by the high uptake rate (66%) and use of the app as evidenced by patient and therapist feedback. Study 2 focused specifically on the postdischarge period, and thus invited participants at the point of discharge. Although appropriate for the study aims, this design meant participants had not been using the Paddle app to develop a set of useful therapy resources during treatment. They were therefore not familiar with the benefits of Paddle before being asked to use it to submit outcome data post discharge. Perhaps as a result, uptake was low (8%). However, despite the nonoptimal conditions for starting to use Paddle, 49% (156/321) of users submitted outcome data on the app and feedback suggested they found it acceptable to help monitor their well-being after treatment.

These studies provide an initial indication of the ways in which the app may have a positive impact on patients, though further studies are needed to examine this in more depth. Feedback highlighted the potential role of the app in supporting patients to organize and remember their thoughts from each session, keep their therapy materials together, and help patients to self-monitor their well-being after treatment had ended. Using the observed uptake rate of 66% (111/168) at the start of treatment, and the follow-up data collection rate of 49% (156/321), a tentative estimate would be that if used in routine practice, the app would be collecting follow-up data on around 32% of patients in NHS TT services. This is considerably higher than typically collected in the participating services (10%) and in these services nationally (1.3%). Although the findings demonstrate that not all patients will choose to use an app or to

XSL•FO
RenderX

complete questionnaires after discharge, this suggests Paddle may offer a helpful route to improve follow-up data collection and understand more about how patients do after treatment.

Analyses indicated that the cohort of app users was representative of patients using the participating services regarding gender and age, though users from White ethnic backgrounds were overrepresented. The group of app users had an average of 13 treatment sessions and a good, reliable recovery rate (89/146, 61%), indicating the sample may have overrepresented those who did well in treatment. Although this provided a greater opportunity to demonstrate reliable relapse in the present study, future research could explore ways to broaden the use of Paddle to reach those who received fewer sessions and those from a wider range of ethnicities.

Although the present follow-up period of 6 months in Study 2 was slightly shorter than in other studies, the observed 36% relapse rate among those who reliably recovered during treatment is consistent with previous literature, falling between existing estimates of 29% [11] and 53% [12]. When analyzing all Paddle users together, regardless of their outcomes during treatment, 40% (63/156) showed reliable deterioration at any point during the follow-up phase. The data suggested that some of these increases in symptoms were not sustained, resolving later during follow-up, but this was only possible to explore where a subsequent timepoint was available. It does suggest, however, that some observed relapses or symptom deteriorations are short-lived, and may reflect natural fluctuations that are qualitatively different from more sustained relapses where additional support may be more beneficial.

We also found considerable evidence of posttreatment stability in Study 2, with a total of 74% (66/89) of those who reliably recovered during treatment remaining reliably recovered or reliably improved across the 6-month follow-up period. Similarly, among the cohort of patients who had reliably improved during treatment, 73% (86/118) remained so across the follow-up period. These findings are in line with the broader CBT literature, suggesting clinical gains are well-sustained following treatment [7-9]. It is also notable that 5 cases who reliably improved during treatment went on to reliably recover during follow-up, and in total, 20% (31/156) of all cases showed reliable improvement at one or more follow-up timepoints compared to their end-of-treatment scores. This demonstrates that further clinical gains can occur after treatment ends.

Employment data indicated that most Paddle users were in employment at the start of treatment and remained so throughout follow-up. Only 11 were unemployed at the start of treatment, but the majority of these moved into employment by the end of treatment and remained so across follow-up. These findings are therefore promising given the association between employment and well-being [33], though the extent to which they directly relate to NHS TT support is unclear and larger studies are needed.

Limitations

These initial studies have some limitations to consider when interpreting the findings. First, although we made good efforts to elicit feedback from those who chose not to use Paddle, the extent of data here was limited, despite providing informative findings. Second, the uptake of the app in Study 2 was low when considered as a proportion of the total number of people invited. However, low uptake rates can be common among health apps with many contributing factors [34], and as discussed earlier, it is possible that, having completed treatment, patients were reluctant to start using the app to collect all their therapy materials together. A helpful next study would recruit participants at the start of treatment as per Study 1, allow users to build their therapy resources through treatment, then explore the rates of continued use and questionnaire completion post discharge. Third, although the sample of app users included people from a range of ethnic backgrounds, those from White backgrounds were overrepresented. Further work to support greater diversity among app users is needed. Fourth, the classification of clinical outcomes was only possible based on the available data. Although data availability was good, with 53% of users providing at least 3 datapoints, it is possible that certain evidence of stability or fluctuations in scores could not be demonstrated. Finally, as it was not possible to include a control group in Study 2, we cannot quantify the extent to which patients' scores in the follow-up period were influenced by using Paddle to self-monitor their well-being, though participant feedback did suggest that Paddle was helpful in this regard. More broadly, it would be beneficial for studies in this area to examine postdischarge employment outcomes in greater detail alongside clinical outcomes. It would also be valuable to use methods that more clearly distinguish temporary lapses from more sustained episodes of relapse. The metric of reliable relapse developed for this study may offer a useful starting point for this.

Conclusions

Overall, the results suggest that Paddle could offer a promising route toward improving support for patients both during and after treatment. Refinements such as further streamlining the risk management procedure are now possible due to the app having medical device status, and based on patient feedback, extending the follow-up period to 12 months may be appropriate. Future studies could helpfully explore users' experience of the app in greater detail and examine factors associated with reliable deterioration or reliable relapse that may guide the provision of further support for these patients.

Acknowledgments

The authors would like to thank the Thames Valley Patient Forum, the Oxford Academic Health Sciences Network, Global Initiative Ltd, the Paddle champions, staff, and patients of the participating services.

XSL•FO
RenderX

Funding

This project is supported by the National Health Service England and NHS Talking Therapies national teams, the Wellcome Trust (200796), and by the National Institute for Health and Care Research (NIHR) Oxford Health Biomedical Research Centre (NIHR203316). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to privacy and ethical restrictions but are available from the corresponding author on reasonable request.

Conflicts of Interest

IW is the appointed Paddle Lead within Brighter Futures Together, a not-for-profit Community Interest Company which hosts the Paddle app. The company had no role in data analysis, interpretation, or the decision to submit this work for publication.

References

1. Cuijpers P, Cristea IA, Karyotaki E, Reijnders M, Huibers MJH. How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence. World Psychiatry 2016;15(3):245-258 [FREE Full text] [doi: 10.1002/wps.20346] [Medline: 27717254]
2. Hofmann SG, Smits JAJ. Cognitive-behavioral therapy for adult anxiety disorders: a meta-analysis of randomized placebo-controlled trials. J Clin Psychiatry 2008;69(4):621-632 [FREE Full text] [doi: 10.4088/jcp.v69n0415] [Medline: 18363421]
3. Watts SE, Turnell A, Kladnitski N, Newby JM, Andrews G. Treatment-as-usual (TAU) is anything but usual: a meta-analysis of CBT versus TAU for anxiety and depression. J Affect Disord 2015;175:152-167. [doi: 10.1016/j.jad.2014.12.025] [Medline: 25618002]
4. Common mental health problems: identification and pathways to care. NICE. 2011. URL: https://www.nice.org.uk/guidance/cg123 [accessed 2023-07-11]
5. Depression in adults: treatment and management. NICE. 2022. URL: https://www.nice.org.uk/guidance/ng222 [accessed 2026-01-20]
6. Hollon SD, Stewart MO, Strunk D. Enduring effects for cognitive behavior therapy in the treatment of depression and anxiety. Annu Rev Psychol 2006;57:285-315 [FREE Full text] [doi: 10.1146/annurev.psych.57.102904.190044] [Medline: 16318597]
7. Bandelow B, Sagebiel A, Belz M, Görlich Y, Michaelis S, Wedekind D. Enduring effects of psychological treatments for anxiety disorders: meta-analysis of follow-up studies. Br J Psychiatry 2018;212(6):333-338. [doi: 10.1192/bjp.2018.49] [Medline: 29706139]
8. Cuijpers P, Hollon S, van Straten A, Bockting C, Berking M, Andersson G. Does cognitive behaviour therapy have an enduring effect that is superior to keeping patients on continuation pharmacotherapy? A meta-analysis. BMJ Open 2013;3(4):e002542 [FREE Full text] [doi: 10.1136/bmjopen-2012-002542] [Medline: 23624992]
9. van Dis EAM, van Veen SC, Hagenaars MA, Batelaan NM, Bockting CLH, van den Heuvel RM, et al. Long-term outcomes of cognitive behavioral therapy for anxiety-related disorders: a systematic review and meta-analysis. JAMA Psychiatry 2020;77(3):265-273 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.3986] [Medline: 31758858]
10. NHS Talking Therapies for anxiety and depression, Annual reports, 2023-24. NHS England. URL: https://digital.nhs.uk/data-and-information/publications/statistical/nhs-talking-therapies-for-anxiety-and-depression-annual-reports/2023-24 [accessed 2026-01-20]
11. Palacios JE, Enrique A, Mooney O, Farrell S, Earley C, Duffy D, et al. Durability of treatment effects following internet-delivered cognitive behavioural therapy for depression and anxiety delivered within a routine care setting. Clin Psychol Psychother 2022;29(5):1768-1777 [FREE Full text] [doi: 10.1002/cpp.2743] [Medline: 35466486]
12. Ali S, Rhodes L, Moreea O, McMillan D, Gilbody S, Leach C, et al. How durable is the effect of low intensity CBT for depression and anxiety? Remission and relapse in a longitudinal cohort study. Behav Res Ther 2017;94:1-8 [FREE Full text] [doi: 10.1016/j.brat.2017.04.006] [Medline: 28437680]
13. Lorimer B, Delgadillo J, Kellett S, Lawrence J. Dynamic prediction and identification of cases at risk of relapse following completion of low-intensity cognitive behavioural therapy. Psychother Res 2021;31(1):19-32 [FREE Full text] [doi: 10.1080/10503307.2020.1733127]
14. Lorimer B, Kellett S, Giesemann J, Lutz W, Delgadillo J. An investigation of treatment return after psychological therapy for depression and anxiety. Behav Cogn Psychother 2023;52(2):149-162. [doi: 10.1017/s1352465823000322]
15. OXCADAT resources: resources for cognitive therapy for PTSD, social anxiety disorder and panic disorder. Oxford Centre for Anxiety Disorders and Trauma. URL: https://oxcadatresources.com/ [accessed 2026-01-20] [WebCite Cache ID 76nRqtXMT]

XSL•FO
RenderX

16. Paddle AW. Welcome to Paddle: Providing support for your therapeutic journey. Paddle. URL: https://paddleapp.org/ [accessed 2026-01-20]
17. Using digital tools across the NHS Talking Therapies for Anxiety and Depression pathway. NHS England. URL: https://future.nhs.uk/NHSTalkingTherapies/view?objectId=167261413 [accessed 2023-07-12]
18. Hollis C, Sampson S, Simons L, Davies EB, Churchill R, Betton V, et al. Identifying research priorities for digital technology in mental health care: results of the James Lind Alliance Priority Setting Partnership. Lancet psychiatry 2018;5(10):845-854 [FREE Full text] [doi: 10.1016/S2215-0366(18)30296-7] [Medline: 30170964]
19. Richards D, Enrique A, Eilert N, Franklin M, Palacios J, Duffy D, et al. Erratum: author correction: a pragmatic randomized waitlist-controlled effectiveness and cost-effectiveness trial of digital interventions for depression and anxiety. NPJ Digit Med 2020;3(1):91 [FREE Full text] [doi: 10.1038/s41746-020-0298-3] [Medline: 32637656]
20. Clark DM, Wild J, Warnock-Parkes E, Stott R, Grey N, Thew G, et al. More than doubling the clinical benefit of each hour of therapist time: a randomised controlled trial of internet cognitive therapy for social anxiety disorder. Psychol Med 2023;53(11):5022-5032 [FREE Full text] [doi: 10.1017/S0033291722002008] [Medline: 35835726]
21. Ehlers A, Wild J, Warnock-Parkes E, Grey N, Murray H, Kerr A, et al. Therapist-assisted online psychological therapies differing in trauma focus for post-traumatic stress disorder (STOP-PTSD): a UK-based, single-blind, randomised controlled trial. Lancet Psychiatry 2023;10(8):608-622. [doi: 10.1016/s2215-0366(23)00181-5]
22. NHS Talking Therapies for anxiety and depression manual. NHS England. URL: https://www.england.nhs.uk/publication/the-improving-access-to-psychological-therapies-manual/ [accessed 2026-01-20]
23. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
24. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
25. Mundt J, Marks I, Shear M, Greist J. The work and social adjustment scale: a simple measure of impairment in functioning. Br J Psychiatry 2002;180:461-464. [doi: 10.1192/bjp.180.5.461] [Medline: 11983645]
26. Salkovskis PM, Rimes KA, Warwick HMC, Clark DM. The health anxiety inventory: development and validation of scales for the measurement of health anxiety and hypochondriasis. Psychol Med 2002;32(5):843-853. [doi: 10.1017/s0033291702005822] [Medline: 12171378]
27. Foa EB, Kozak MJ, Salkovskis PM, Coles ME, Amir N. The validation of a new obsessive-compulsive disorder scale: the Obsessive-Compulsive Inventory. Psychol Assess 1998;10(3):206-214. [doi: 10.1037//1040-3590.10.3.206]
28. Weathers FW, Litz BT, Keane TM, Palmieri PA, Marx BP, Schnurr PP. PTSD checklist for DSM-5 (PCL-5). U.S. Department of Veterans Affairs. 2013. URL: https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp [accessed 2026-01-20]
29. Shear MK, Brown TA, Barlow DH, Money R, Sholomskas DE, Woods SW, et al. Multicenter collaborative panic disorder severity scale. Am J Psychiatry 1997;154(11):1571-1575. [doi: 10.1176/ajp.154.11.1571] [Medline: 9356566]
30. Kroenke K, Spitzer RL, Williams JBW. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. Psychosom Med 2002;64(2):258-266. [doi: 10.1097/00006842-200203000-00008] [Medline: 11914441]
31. Connor KM, Davidson JR, Churchill LE, Sherwood RH, Weisler R, Foa E. Psychometric properties of the social phobia inventory (SPIN) new self-rating scale. Br J Psychiatry 2000;176(4):379-386 [FREE Full text] [doi: 10.1192/bjp.176.4.379]
32. Browne NL, Carragher NO, O'Toole A, Pimm J, Ryder J, Thew GR. Evaluating user experiences of SHaRON: an online CBT-based peer support platform. Cogn Behav Ther 2022;15:e18. [doi: 10.1017/s1754470x22000150]
33. Modini M, Joyce S, Mykletun A, Christensen H, Bryant RA, Mitchell PB, et al. The mental health benefits of employment: results of a systematic meta-review. Australas Psychiatry 2016;24(4):331-336. [doi: 10.1177/1039856215618523] [Medline: 26773063]
34. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]

Abbreviations

ADSM: anxiety disorder specific measures
CBT: cognitive behavioral therapy
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
GAD-7: Generalized Anxiety Disorder 7-item version
iCT-PTSD: internet-delivered cognitive therapy for posttraumatic stress disorder
iCT-SAD: internet-delivered cognitive therapy for social anxiety disorder
ISO: International Organization for Standardization
NHS TT: National Health Service Talking Therapies
NICE: National Institute for Health and Care Excellence
PHQ-9: Patient Health Questionnaire 9-item version

XSL•FO
RenderX

PPI: patient and public involvement
WSAS: Work and Social Adjustment Scale

Edited by A Kushniruk, Y Senathirajah; submitted 11.Nov.2024; peer-reviewed by L Yang, A Eisingerich; comments to author 12.Nov.2025; revised version received 07.Jan.2026; accepted 13.Jan.2026; published 24.Feb.2026.

Please cite as:
Thew GR, Lee M, Wolsey I, Giles F, Hayden E, Jell G, Peacock J, Pimm J, Ryder J, Salvadori A, Clark DM
Providing Psychological Therapy Support and Improving Postdischarge Data Collection: Preliminary Evaluation of the "Paddle" App
JMIR Hum Factors 2026;13:e68671
URL: https://humanfactors.jmir.org/2026/1/e68671
doi:10.2196/68671
PMID:

©Graham R Thew, Michelle Lee, Ineke Wolsey, Fiona Giles, Emma Hayden, Grace Jell, Jamie Peacock, John Pimm, Joanne Ryder, Alison Salvadori, David M Clark. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 24.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Deployment of an Activity Monitoring Program to Complement a Clinical Intervention for Veterans With Gulf War Illness: Qualitative Study

Selene S Mak[1], MPH, PhD; Pauline McManus[2], BSN, RN; Steven Greer[2], MS; Kari Haws[2], PhD; Carrie J Carlson[2], DO; Helena K Chandler[2], PhD; Omowunmi Osinubi[2], MSc, MBA, MD; Garrett I Ash[3], PhD

[1]Center for the Study of Healthcare Innovation, Implementation & Policy (CSHIIP), Veterans Affairs Greater Los Angeles Healthcare System, Los Angeles, CA, United States

[2]War Related Illness and Injury Study Center - VA New Jersey Health Care System (WRIISC-VANJHCS), East Orange, NJ, United States

[3]Pain Research, Informatics, Multi-morbidities, and Education (PRIME) Center, Veterans Affairs Connecticut Healthcare System, West Haven, CT, United States

Corresponding Author:
Selene S Mak, MPH, PhD
Center for the Study of Healthcare Innovation, Implementation & Policy (CSHIIP)
Veterans Affairs Greater Los Angeles Healthcare System
11301 Wilshire Blvd
Los Angeles, CA, 90073
United States
Phone: 1 310 478 3711
Fax: 1 310 268 4320
Email: selene.mak@va.gov

Abstract

Background: Many veterans who served in the Gulf experience Gulf War Illness (GWI), a chronic multisymptom condition associated with fatigue, pain, gastrointestinal problems and respiratory issues, mood/cognitive issues, and sleep difficulties. These symptoms contribute to decreased function, increased mental health needs, and poor quality of life. The Veterans Affairs War Related Illness and Injury Study Center in New Jersey developed a 26-week virtual health coaching intervention to support symptom management for veterans with GWI. In 2023, a consumer-grade smartwatch was added as part of an activity monitoring program to complement this program.

Objective: The purpose of this project was to assess the feasibility and acceptability of including a smartwatch-based activity monitoring component to complement a virtual health coaching program for veterans with GWI.

Methods: Twenty-four veterans enrolled in the health coaching program were invited to participate in the activity monitoring component. Participants attended a virtual orientation to set up the smartwatch, and verbal consent to share data through a Health Insurance Portability and Accountability Act (HIPAA)-compliant platform was obtained. Program feasibility was assessed by evaluating wear-time percentage and duration of use. Acceptability was assessed using two items from a monthly survey and through a midprogram semistructured interview. Quantitative data were summarized descriptively, and qualitative data were analyzed using a coding scheme adapted from Sekhon et al's Theoretical Framework for Acceptability (TFA).

Results: Twenty veterans agreed to participate in the program (mean age 49 years; 7/20, 35% female; 19/20, 94% non-Hispanic White; 11/20, 55% first-time smartwatch users). Twelve participants (60%) wore the watch for the full 26 weeks. Among participants who completed 26 weeks, median daily wear-time completeness exceeded 80% for 25 weeks. Most participants (12/20, 60%) reported that wearing the smartwatch helped them achieve their wellness goals, and the majority (16/20, 80%) said they would recommend using the smartwatch for activity monitoring to other veterans. Qualitative findings supported acceptability across TFA domains. One adverse event was reported (minor skin irritation that resolved after changing the smartwatch band to a hypoallergenic watch band).

Conclusions: Within this clinical program, pairing a smartwatch with virtual health coaching for veterans with GWI was feasible and acceptable. Activity monitoring integrated into an existing intervention may support symptom self-management and augment patient education and engagement. As no prior activity monitoring programs specific to veterans with GWI have been described, these findings could inform future program development and implementation within this population.

XSL•FO
RenderX

(JMIR Hum Factors 2026;13:e82177) doi:10.2196/82177

KEYWORDS

smartwatch; digital health; feasibility; Gulf war illness; HIPAA

Introduction

An estimated one-third of veterans who served in the 1990-1991 Persian Gulf War experience chronic multisymptom illnesses, commonly referred to as Gulf War Illness (GWI) [1,2]. Veterans with GWI often report a range of symptoms, including fatigue, pain, gastrointestinal problems, respiratory issues, mood/cognitive issues, and sleep difficulties that may be related to exposures during their deployment [1,3]. These chronic health concerns and symptoms contribute to decreased function, increased mental health needs, and poor quality of life [1,3]. Because the etiology of GWI remains poorly understood, treatment for GWI has largely focused on managing and alleviating symptoms [3]. Prior interventions targeting physical activity, pain, cognition, gut health, and mindfulness have shown mixed results [4].

In 2021, the Veterans Affairs (VA) War Related Illness and Injury Study Center in New Jersey (WRIISC-NJ) developed a virtual 26-week health coaching intervention to support individualized symptom management for veterans with GWI [5]. The program combines group-based sessions for peer support with individual health coaching tailored to participants' symptom profile. In 2023, a smartwatch-based activity monitoring component was added to support physical activity and sleep tracking as part of this program.

Consumer-grade smartwatches are now widely available and are used by a substantial proportion of US adults [6]. Veterans' interest in and use of smartwatches has increased in the past few years [7]. These devices use accelerometers and optical sensors to capture physical activity and heart rate. While smartwatches are not used for diagnosis, the information provided can serve as "baseline" for the individual, which can lead to greater awareness of health behavior patterns [8,9]. When paired with behavioral interventions, smartwatches may offer participants added support toward achieving personal health goals such as weight loss, reduced sedentary behavior, and increased physical activity [9,10].

Assessing feasibility and acceptability is a critical early step in the development of digital health interventions. Using definitions from Proctor et al's [11] implementation outcomes framework, feasibility refers to the extent to which an intervention can be successfully performed within a given setting, and acceptability reflects the degree to which the intervention is perceived as satisfactory or agreeable [11,12]. Acceptability and feasibility are closely linked, as an intervention may be feasible to deliver but not acceptable to participants or acceptable but difficult to implement [13,14]. In a resource-limited environment, assessing feasibility and acceptability helps determine whether an intervention is a reasonable use of time and effort before broader implementation or scale-up. These evaluations identify key facilitators and barriers based on participant feedback and willingness to engage in intervention activities, and findings can be used to refine program design and delivery [15]. Consistent with behavior change theories such as the Theory of Planned Behavior [16], sustained engagement with digital health tools is essential for the intervention to have any practical impact. Understanding feasibility and acceptability is particularly important for veterans with GWI, given their high symptom burden and complex care needs.

The purpose of this project was to assess the feasibility and acceptability of including a smartwatch-based activity monitoring component to complement a virtual health coaching program for veterans with GWI. From a human factors perspective, feasibility and acceptability reflected veterans' perceptions of how the technology fit into their daily routines and use capabilities in real-world settings. Since there is no known activity monitoring program for veterans with GWI, evaluating feasibility and acceptability is critical to future implementation of any activity monitoring program for this population.

Methods

Ethical Considerations

This quality improvement study aimed to inform the development of an activity monitoring program to complement a clinical intervention for veterans with GWI. The VA New Jersey Healthcare System Institutional Review Board determined that this quality improvement project met criteria for quality improvement and was considered non-research; therefore, formal institutional review board review was not required. The CONSORT (Consolidated Standards of Reporting Trials) extension for pilot and feasibility studies was used for reporting (Multimedia Appendix 1) [17,18].

Study Participants

Veterans with GWI who completed a comprehensive evaluation at the WRIISC-NJ were invited to participate in a 26-week virtual health coaching intervention program. Two cohorts were enrolled: April 2023 to September 2023 and October 2023 to March 2024. GWI diagnosis was determined clinically through WRIISC-NJ specialty evaluation.

Procedures

Participants received a smartwatch (Fitbit Sense 2) as part of a VA device distribution initiative [19] as well as a user guide. A videoconferencing orientation provided an overview of device setup, smartwatch capabilities (eg, tracks physical activity and heart rate), program expectations, risk mitigation (ie, taking breaks from wearing the smartwatch to mitigate skin irritation), and data privacy and sharing through a third-party Health Insurance Portability and Accountability Act (HIPAA)-compliant platform (Fitabase). Verbal consent for data sharing was obtained. Participants did not receive monetary compensation but could keep the smartwatch after program completion. Feasibility and acceptability were assessed using

XSL•FO
RenderX

smartwatch wear-time data, monthly electronic surveys, and a semistructured interview.

Activity Monitoring Program

Participants were instructed to wear the smartwatch, keep it charged, sync data using the manufacturer's mobile app at least once every 7 days, and agree to share data with the project team. These procedures were intended to reflect typical real-world use rather than impose artificial workflow constraints on participants. No additional instructions were provided regarding when or how long to wear the device beyond wearing it as tolerated. The program was designed to capture participant behavior in a free-living environment. A nurse educator reviewed battery status and "last synced" data twice per week and contacted participants if more than 6 days had elapsed without data syncing. During the postprogram session with the clinical director, participants were presented with a report summarizing their average weekly step count and sleep duration.

Quantitative Data Collection

Program feasibility was assessed using the number of participants recruited and enrolled, average daily wear-time percentage, and duration of smartwatch use (weeks) prior to discontinuation. Adverse events related to smartwatch use were recorded. Program acceptability was assessed using two items from the monthly Qualtrics survey: (1) the extent to which the smartwatch has helped them achieve their wellness goals (0=not at all; 10=definitely) and (2) the likelihood of recommending the program to another veteran by using a modified 1-item net promoter score (0=not at all likely; 10=extremely likely) [20].

Qualitative Data Collection

Questions focused on smartwatch use, perceived benefit, and anticipated continued use. These questions were loosely adapted from interview guides from other interventions conducted by the senior researchers (GA, SM) that had included questions about acceptability of technology use. Questions included, "How do you use your Fitbit?" and "Do you think you'll keep wearing the watch after the program has ended? Why or why not?" A nurse educator (author PM) trained in motivational interviewing contacted the participants by phone after the program midpoint to conduct a semistructured interview regarding smartwatch use. Two attempts were made to reach each participant. Interviews were scheduled to coincide with other program evaluation activities when possible. Interviews lasted approximately 10 minutes, and responses were recorded using a Microsoft Teams form.

Data Analysis

Survey Data

The number of completed surveys was recorded. For each survey item, the range and frequency of the final response were summarized. Group net promoter score was calculated by subtracting the percentage of detractors (scores 0-6) from promoters (scores 9-10); scores range from –100 to 100, with scores above 0 considered favorable [21].

Smartwatch Data

Step count and heart rate data were downloaded as Microsoft Excel files and stored on a VA secure server accessible only to the project team. Data were accessed by authorized project staff through the Fitabase platform and exported manually for analysis. Twenty-four-hour wear time, defined as the presence of a heart rate signal, was calculated as a weekly percentage for each participant. Weekly group distributions were visualized using box plots. Anomalous days (eg, steps recorded without heart rate signal) were excluded (~1%). All analyses were conducted using Microsoft Excel.

Semistructured Interviews

Interview data were analyzed using the Theoretical Framework for Acceptability (TFA) [22]. Sekhon et al [22] developed TFA as a comprehensive guide to assess this multifaceted measure. Six constructs were applied: affective attitude, burden, perceived effectiveness, intervention coherence, opportunity costs, and self-efficacy (Table 1).

Ethicality was excluded because no ethical or moral issues were identified in this context. Data were transferred from a Microsoft Teams form to Microsoft Word and then uploaded into Atlas.ti (version 23.2.1) for further analysis [23]. Two team members (SM, GA) piloted the coding scheme on 4 documents together. After minor revisions to the coding scheme, the same researchers coded subsequent documents independently and reconciled conflicts with team discussion.

XSL•FO
RenderX

Table 1. Coding scheme adapted from Sekhon et al's [22] Theoretical Framework of Acceptability (TFA).

Construct of acceptability	Definition of construct	Notes
Affective attitude	• How a new smartwatch user feels about using the smartwatch • How an individual with previous experience using a smartwatch feels about using the program-issued smartwatch compared to another brand's smartwatch	Whether a participant has had previous experience with another smartwatch will likely affect his/her perception of using a smartwatch
Burden	• The perceived amount of effort that is required to use the smartwatch	Comments may provide insight into barriers that should be addressed before a large-scale trial/future implementation of program
Intervention coherence	• The extent to which the participant understands how to use the smartwatch and how it can continuously collect various forms of data passively	Whether participants understood that the smartwatch should be worn "continuously" to allow for collection of data continuously will affect smartwatch wear time. Device wear time is often a proxy for an acceptability.
Opportunity costs	• The extent to which benefits, profits, or values must be given up by using the smartwatch	Comments may provide insight to barriers that should be addressed before a large-scale trial/future implementation of program
Perceived effectiveness	• The extent to which using the smartwatch is perceived as likely to achieve its purpose	Our interview questions proposed two purposes of the smartwatch: (1) whether the smartwatch has helped the participant achieve his/her wellness goals and (2) whether participants reviewed the data collected by the smartwatch.
Self-efficacy	• The participant's confidence that they can perform the behavior(s) required to use the smartwatch • The participant's confidence that they can perform the behavior(s) required to use the smartwatch led to actual use of the smartwatch's tracking function	Self-efficacy is central to technology acceptance because it reflects whether a user believes they can successfully use the device. The degree of confidence will affect actual use.

Results

Participant Characteristics

Twelve individuals enrolled in each of the two cohorts (n=24). Three later withdrew from the health coaching program, and 1 declined to participate in the activity monitoring program, resulting in a final sample of 20 participants (mean age 49 years; 7/20, 35% female; 19/20, 94% non-Hispanic White; 11/20, 55% first-time smartwatch users; Table 2). Nineteen participants completed at least one of the 6 monthly surveys, including 11 who completed all 6 surveys, 4 who completed 5 surveys, and 4 who completed 2-4 surveys. All participants completed a semistructured interview.

Table 2. Participant characteristics (N=20).

Demographic characteristics	Values, n (%)
Sex	
Female	7 (35)
Male	13 (65)
Age group (y)	
30-40	3 (15)
41-50	6 (30)
51-60	11(55)
Race/ethnicity	
Non-Hispanic White	19 (94)
Declined to answer	1 (6)
First-time smartwatch user	11 (55)

Survey Responses

Eighteen participants responded at least once to the question, "To what extent did the Fitbit program help you achieve your wellness goals?" Final responses ranged from 0 (not helpful at all; n=1), 1-5 (n=5), 6-8 (n=9), to 9-10 (very helpful; n=3). Among the 17 participants who completed at least two surveys,

XSL•FO
RenderX

the largest observed change was from 3 to 10. All but one participant showed an increase from baseline to final response.

Nineteen participants responded at least once to the recommendation question. Final ratings ranged from 0 to 6 (n=4), 7 to 8 (n=8), and 9 to 10 (n=7). The final net promoter score was 15.8, indicating a favorable score. All but 4 participants reported increased recommendation scores over time.

Twelve participants who rated the device as helpful (rating greater than 5) also indicated they would recommend the smartwatch to other veterans (rating greater than or equal to 7). Four of 6 participants who did not rate the device helpful would still recommend the device to other veterans.

Wear Time

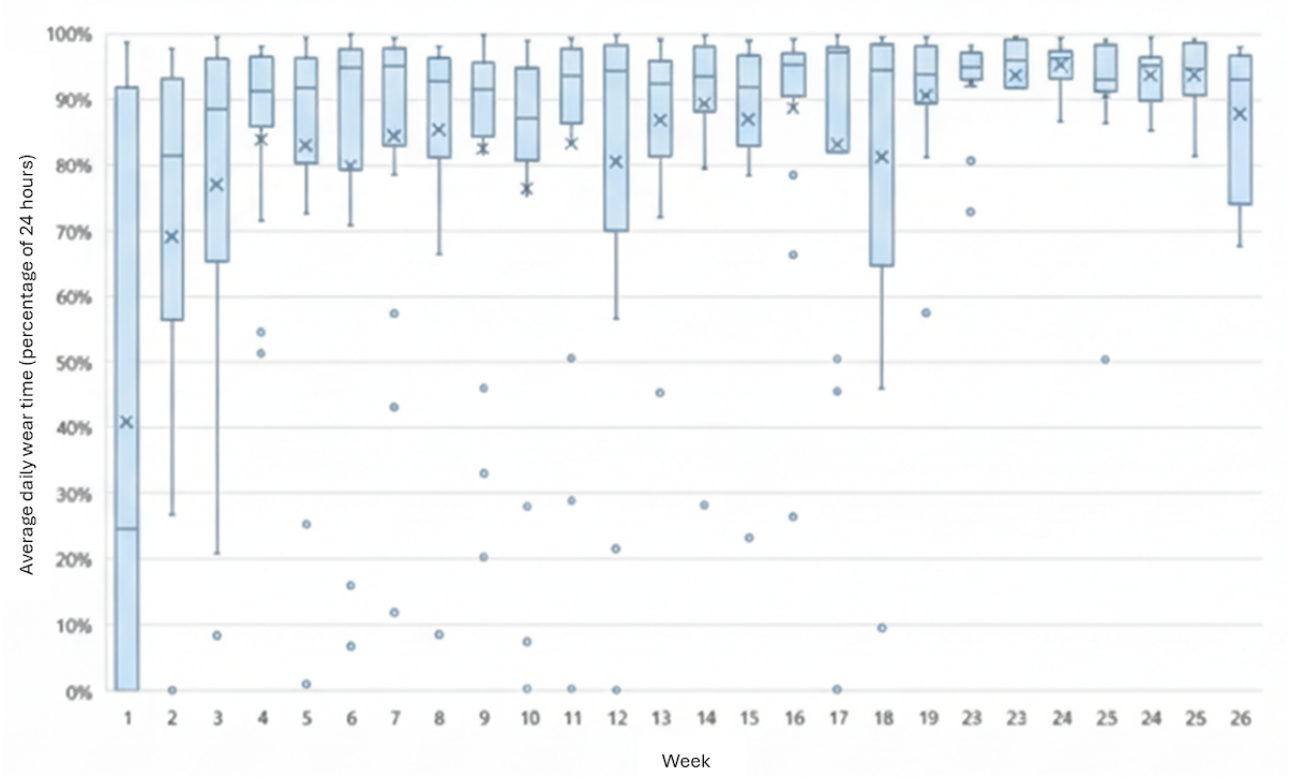
Total wear time ranged from 8 weeks to 26 weeks (median 26, IQR 18-26). Twelve participants (60%) wore the smartwatch for the full 26 weeks. The remaining 8 participants discontinued use between 8 and 12 weeks (n=3), 16 and 18 weeks (n=4), and 23 weeks (n=1).

Average daily wear-time completeness ranged from 10% (2.4/24 hours) to 96% (23.0/24 hours) across participants. Among those who completed 26 weeks, daily completeness ranged from 76% to 96% (18.2-23.0/24 hours; mean 87% [20.9 hours], SD 6% [1.0 hours]). Prior to device discontinuation, 6 participants demonstrated similarly high completeness (mean 85% [20.0 hours], SD 8% [2.0 hours]).

First-Time Smartwatch Users

Of the 11 first-time smartwatch users, they were more likely to be younger than 50 years (8/11, 73%) and female (5/11, 45%). First-time users showed higher rates of early discontinuation (6/11, 55%) and low daily completeness (2/11, 18%). Figure 1 shows the weekly distributions of daily wear-time completeness across participants. In most weeks, wear-time distributions were negatively skewed, with a subset of participants demonstrating substantially lower wear time than the group median. By approximately week 18, this pattern diminished as participants who discontinued use had stopped contributing data. The remaining 12 participants who wore the smartwatch for the full 26 weeks generally maintained consistently high weekly wear time.

Figure 1. Wear time by week of program. The blue box is the IQR and represents the middle 50% of wear time for the group. The bars indicate typical range (1.5 times IQR). Dots indicate outlier points (outside 3 times IQR), horizontal dash indicates median, and X indicates mean. Participants who stopped wearing smartwatch altogether (between 8 and 12 weeks [n=3], between 16 and 18 weeks [n=4], at 23 weeks [n=1]) are excluded at the point of discontinuation.



Adverse Events

There was one adverse event reported. One participant reported minor skin irritation related to the watch band, which resolved after switching to a hypoallergenic watch band.

Semistructured Interviews

Findings from the semistructured interviews are organized below by domain from TFA. Representative quotes are provided in Table 3.

XSL•FO
RenderX

Table 3. Illustrative quotes.

Construct of acceptability	Quotes
Affective attitude	• …In general, I do not like the Fitbit anywhere near as much as my Apple iWatch v3 Bluetooth. The Fitbit's Bluetooth disconnects constantly throughout the day, won't allow me to change the temperature measurement from Celsius to Fahrenheit, rings for an entire minute after I pick-up my phone, has less apps capability, but [it] does have better sleep monitoring and battery life. [Participant 8] • …I like it more and more, especially for the sleep monitoring because it has helped me attend to my sleep. [Participant 12] • …I love it! I would recommend it to others. [Participant 14] • …I stopped wearing it. It doesn't help me with my needs. I'm just not comfortable wearing it. I can't explain. Intrusive comes to mind but I don't know if that is the right word? Inconvenient? [Participant 3]
Burden	• …I thought it would be difficult to wear but it's not. [Participant 15] • …I don't even notice I'm wearing it anymore. [Participant 6] • …I do wear it except for when I'm showering and charging. I [sometimes forget] to put it on after charging, but I do wear it almost all of the time. [Participant 20]
Intervention coherence	• …I wear it all the time and look at my activity and my steps. [Participant 9] • …I wear it during the day and at night. I take it off to charge or take a shower. [Participant 11] • …I wear it in the day and when I sleep. I take it off to charge or when I shower. [Participant 15] • …I'm wearing it because I'm in the [health coaching] program. [Participant 19]
Opportunity costs	• …I had to adjust to wearing it all the time in the beginning but I am used to it now. [Participant 20] • …I was using a Garmin to affect the way I do things so I'm wearing the Fitbit to do that now. [Participant 11] • …I'm wearing [the Fitbit] every day, but I prefer my Garmin. [Participant 10]
Perceived effectiveness	• …I like it more and more, especially for the sleep monitoring because it has helped me attend to my sleep. [Participant 12] • …I use it for everything…I've lost weight! I watch my food intake. [Participant 14] • …I use the step counter, it saves the weekly reports so I look at those, heart rate and exercise, calorie counter and running heart rate checker if I'm feeling stressed…I also don't use the watch as a phone. I use it to track my health. [Participant 14] • …I use it to track my sleep but it doesn't always show up. The sleep isn't accurate with tracking…According to the watch I'm not sleeping. [Participant 15]
Self-efficacy	• …Yes. I look at my steps and my sleep. Sometimes when I look at the data I'm surprised by how many I have moved around. I recently got a penguin badge for all the walking I did when I visited the zoo with my niece a few weeks back - almost 6 miles! Even this morning I was surprised at how far I walked with the dogs. I didn't think it was that far. [Participant 7] • …I think it causes me to move more and makes me want to hit more steps. I set a goal of 10K steps but sometimes I do more. I'll do more as it goes on. I joined a gym since I started the program and I walk on the treadmill there. [Participant 15] • …I look at it more now to track my steps, sleep, and HR. I started looking at the other apps available now that I'm more familiar with the watch. [Participant 6] • …I like to look at the data and track my activity and progress. I'll look and see I haven't been as active as I thought, and so I will take the dogs for another walk so I can reach my goal of 8k steps. [Participant 1]

Affective Attitude

Most participants reported a positive overall attitude toward using the smartwatch, with many noting good usability and low interaction burden during daily activities. Some expressed immediate enthusiasm for the device, while others, particularly first-time smartwatch users, became more comfortable over time as they learned its features. Participants with prior smartwatch experience often compared devices and expressed brand preferences; however, these preferences did not appear to limit the continued use of the program-issued smartwatch, suggesting that acceptability of activity monitoring was not driven by brand alone. One participant discontinued use, citing discomfort and inconvenience with wearing the device.

Burden

The perceived effort required to use the smartwatch was generally described as low. Participants most commonly cited routine charging and remembering to put the watch back on after charging as minor inconveniences. These issues did not appear to substantially interfere with ongoing use. Only one participant reported skin irritation, which resolved with a band change.

Intervention Coherence

Participants demonstrated a clear understanding of how the smartwatch fit within the larger health coaching program. Participants consistently articulated that wearing the device, syncing data, and sharing information with the program team were expected components of participation. There was little

XSL•FO
RenderX

evidence of confusion regarding device purpose or data collection.

Opportunity Costs

There were only a handful of comments related to the idea that wearing the smartwatch caused the participant to give up something else. No individual was overly concerned about loss of time related to managing the device. One participant shared that while he preferred another brand of smartwatch because it tracked his runs better, he gave up wearing the other watch and wore the program-issued smartwatch instead while participating in the health coaching program.

Perceived Effectiveness

Most participants provided specific examples about functionalities of the smartwatch. They mentioned “steps,” “sleep,” “heart rate,” and one participant mentioned “water intake” and “calories.” They reported that the act of monitoring increased their awareness of physical activity and encouraged them to be more active or pay closer attention to sleep habits. While participants generally described the smartwatch data as useful for self-monitoring, several expressed mixed views about the perceived accuracy of certain metrics, particularly sleep data.

Self-Efficacy

Participants uniformly expressed confidence in their ability to use the smartwatch. Particularly for those with previous experience using another smartwatch, they leveraged their previous experience to use the program-issued smartwatch, and some reported tinkering with the applications. The majority of the participants recognized that the smartwatch can be used to track some health behaviors and reported independently reviewing their data to inform daily activity choices.

Discussion

Principal Findings

In this program, smartwatch-based activity monitoring was incorporated into a virtual health coaching program for veterans with GWI with sustained use and participation. High rates of sustained participation and daily wear-time completeness indicate that most participants were willing and able to engage with activity monitoring over an extended period. Favorable survey scores and qualitative feedback further support the acceptability of this approach. From a human factors standpoint, the combination of sustained wear, low reported burden, and high self-efficacy suggests that the device-user interaction was well aligned with participants’ capabilities and daily workflows. Program observations support the feasibility and acceptability of incorporating activity monitoring into health coaching programs for veterans with GWI.

This paper presents findings from the first known activity monitoring program specifically for veterans with GWI. Prior studies have shown that wearable activity trackers can be implemented in populations with chronic disease and are commonly used to support physical activity and self-monitoring, which is consistent with the patterns observed in this program [24-28]. Using a smartwatch for activity monitoring may help inform the design of personalized symptom-management strategies and may provide signals regarding treatment response. This may be applicable for veterans with GWI with decreased cognitive function, since the smartwatch may serve as a cognitive support tool to reduce tracking burden, create accountability, increase awareness of how behavior can affect health, and prompt discussions with care team [27].

Assessing feasibility and acceptability are conventional practices for researchers when introducing new program components, particularly in unique clinical populations such as veterans with GWI who have complex and variable health needs. A feasibility pilot would also explore barriers and pain points related to technology use prior to broader implementation. Technology acceptance frameworks such as the Technology Acceptance Model describe how perceived usefulness and perceived ease of use shape attitudes toward technology and influence actual use [29]. Although our qualitative analysis was guided by TFA, these constructs align conceptually with the Technology Acceptance Model and helped clarify why participants did or did not accept the activity monitoring program.

Prior studies have identified factors such as social features and engagement strategies that influence the acceptability of mobile health components, including social connectivity [30], gamification [31], and ability to choose among available devices [32]. Device-agnostic platforms may increase flexibility and user choice; however, they can also introduce challenges related to data integration and standardization across devices [32,33]. Issuing a single device to all participants can simplify data management and support consistent clinical feedback, but it may also introduce tension for participants who prefer another brand of smartwatch. Our findings reflected this complexity. Several participants expressed loyalty to other smartwatch brands; however, most continued to wear the program-issued device, suggesting that brand preference alone did not determine acceptability of the activity monitoring program. Similar to prior digital health programs, sustained engagement tends to decline over time in app-based and wearable-based interventions [34,35].

Acceptability in this program appeared to be multifaceted and strongly shaped by clinical context. Survey results suggested that perceived helpfulness was closely aligned with willingness to recommend the program, supporting overall acceptability. At the same time, qualitative data indicated that some participants viewed the smartwatch as less useful, preferred a different brand, or questioned specific device features. Despite these reservations, many continued to wear the program-issued device because it was embedded within the larger health coaching program. For these participants, participation in the core health coaching intervention and their commitment to the program appeared to outweigh dissatisfaction with the specific device. This suggests that acceptability of the activity monitoring component was influenced not only by the smartwatch itself but also by veterans’ motivation to fully engage in the broader health coaching program.

Limitations

This quality improvement project was designed to assess feasibility and acceptability rather than effectiveness; therefore,

XSL•FO
RenderX

causal inferences cannot be made. Smartwatch data were collected in an uncontrolled free-living environment and could not be independently validated. Although the sample size was modest, it was sufficient to detect common feasibility challenges in pilot work [36]. Participants were recruited from a single VA center, which may limit generalizability to the broader population of veterans with GWI. Because activity monitoring was implemented as part of a health coaching program, our findings may not generalize to a standalone activity monitoring program or to other populations.

Conclusions

The growing availability of smartwatches creates an opportunity to support symptom self-management for veterans with GWI. In this program, activity monitoring was implemented to complement a 26-week virtual health coaching program. The program demonstrated high feasibility and acceptability within this clinical setting, supporting activity monitoring as a feasible way to augment patient education and ongoing support for veterans with GWI. Program acceptability appeared to be influenced by both the usability of the device and its integration within the existing health coaching program.

When conducting feasibility work, programs should be designed to reflect the specific setting and population of interest. Collecting both quantitative and qualitative data provides important context for understanding use and engagement. Assessing feasibility and acceptability can then be used to inform implementation strategies tailored to supporting sustained participation in the program.

Acknowledgments

The authors would like to thank Elijah Sacra, Clarissa Kussin, and Sandhya Bandi for their support of the program. Carolyn Alish of Sterling Medical Communications LLC provided editorial support. The views expressed in this article are those of the authors and do not necessarily represent the views of the US Department of Veterans Affairs or the United States government. Generative artificial intelligence (AI) tools were used solely to assist with grammar and style editing. No AI tools were used to generate substantive content. The authors reviewed all edits and take full responsibility for the final text.

Funding

No funding was received for this study. However, GIA's time was supported by the National Institute of Diabetes, Digestive, and Kidney Diseases of the National Institutes of Health under a mentored research scientist development award (K01DK129441).

Data Availability

The data that support the findings of this study are available on request from the corresponding author.

Authors' Contributions

Conceptualization, formal analysis, methodology, visualization, writing – original draft: SSM, GIA
Data collection: PM, SG, SSM, GIA
Writing – review & editing: KH, CJC, HKC, OO, PM, SG, SSM, GIA
All authors read and approved the final manuscript.

Conflicts of Interest

GIA is a scientific advisor for WellModeled LLC, has a provisional patent filed for a digital system for lifestyle medicine (047162-5346-P1US), and receives professional services from Calm.com (nominal fee), Labfront (full fee), and GlucoseZone (full fee). GIA attests that none of these entities supported the work and none of them influenced the work, its design, or outcomes. GIA conducted the work outside of his responsibilities and affiliations with these entities. The other authors declare no conflicts of interest.

Multimedia Appendix 1
CONSORT-EHEALTH Checklist (version 1.6.1).
[PDF File (Adobe PDF File), 988 KB - humanfactors_v13i1e82177_app1.pdf]

References

1. Institute of Medicine. Gulf War and Health: Treatment for Chronic Multisymptom Illness. Washington, DC: The National Academies Press; 2013.
2. Timeline of US period of service eras (1900-present). US Department of Veterans Affairs, Office of Health Equity. URL: https://www.va.gov/HEALTHEQUITY/docs/Period_of_Service_Timeline_OHE02102017.pdf [accessed 2025-11-29]
3. National Academies of Sciences, Engineering, and Medicine. Gulf War and Health: Volume 10: Update of Health Effects of Serving in the Gulf War. Washington, DC: The National Academies Press; 2016.

XSL•FO
RenderX

4. Nugent SM, Freeman M, Ayers CK, Winchell KA, Press AM, O'Neil ME, et al. A systematic review of therapeutic interventions and management strategies for Gulf War illness. Mil Med 2021 Jan 30;186(1-2):e169-e178. [doi: 10.1093/milmed/usaa260] [Medline: 33128563]
5. Haws K, Mak S, Greer S, Kussin CA, Sacra E, Carlson CJ, et al. A virtual functional medicine-based interdisciplinary and integrative intervention for Gulf war illness. Mil Med 2024 Aug 19;189(Suppl 3):99-105. [doi: 10.1093/milmed/usae054] [Medline: 39160805]
6. National tracking poll 2301120. Morning Consult. 2023. URL: https://pro-assets.morningconsult.com/wp-uploads/2023/02/16130638/2301120_crosstabs_MC_HEALTH_DIGITAL_HEALTH_DEVICES_Adults.pdf [accessed 2025-12-20]
7. Saleem JJ, Wilck NR, Murphy JJ, Herout J. Veteran and staff experience from a pilot program of health care system-distributed wearable devices and data sharing. Appl Clin Inform 2022 May;13(3):532-540 [FREE Full text] [doi: 10.1055/s-0042-1748857] [Medline: 35613912]
8. Swan M. The quantified self: fundamental disruption in big data science and biological discovery. Big Data 2013 Jun;1(2):85-99. [doi: 10.1089/big.2012.0002] [Medline: 27442063]
9. Wieczorek M, O'Brolchain F, Saghai Y, Gordijn B. The ethics of self-tracking. A comprehensive review of the literature. Ethics & Behavior 2022 Jun 06;33(4):239-271. [doi: 10.1080/10508422.2022.2082969]
10. Hackshaw KV, Plans-Pujolras M, Rodriguez-Saona LE, Moore MA, Jackson EK, Sforzo GA, et al. A pilot study of health and wellness coaching for fibromyalgia. BMC Musculoskelet Disord 2016 Nov 08;17(1):457 [FREE Full text] [doi: 10.1186/s12891-016-1316-0] [Medline: 27821160]
11. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
12. Karsh B. Beyond usability: designing effective technology implementation systems to promote patient safety. Qual Saf Health Care 2004 Oct;13(5):388-394 [FREE Full text] [doi: 10.1136/qhc.13.5.388] [Medline: 15465944]
13. Weiner BJ, Lewis CC, Stanick C, Powell BJ, Dorsey CN, Clary AS, et al. Psychometric assessment of three newly developed implementation outcome measures. Implement Sci 2017 Aug 29;12(1):108 [FREE Full text] [doi: 10.1186/s13012-017-0635-3] [Medline: 28851459]
14. Klaic M, Kapp S, Hudson P, Chapman W, Denehy L, Story D, et al. Implementability of healthcare interventions: an overview of reviews and development of a conceptual framework. Implement Sci 2022 Jan 27;17(1):10 [FREE Full text] [doi: 10.1186/s13012-021-01171-7] [Medline: 35086538]
15. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
16. Ajzen I. The theory of planned behavior. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
17. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS consensus group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot Feasibility Stud 2016;2:64 [FREE Full text] [doi: 10.1186/s40814-016-0105-8] [Medline: 27965879]
18. Lancaster GA, Thabane L. Guidelines for reporting non-randomised pilot and feasibility studies. Pilot Feasibility Stud 2019;5:114 [FREE Full text] [doi: 10.1186/s40814-019-0499-1] [Medline: 31608150]
19. Amrhein A. VA and Fitbit help support veterans' health and wellness during COVID-19 pandemic. VA News.: US Department of Veterans Affairs; 2021. URL: https://news.va.gov/83470/va-fitbit-veterans-covid-19-pandemic/ [accessed 2024-12-20]
20. Reichheld FF. The one number you need to grow. Harv Bus Rev 2003 Dec;81(12):46-54. [Medline: 14712543]
21. Carpenter A. What is a good Net Promoter Score? Qualtrics. 2021. URL: https://www.qualtrics.com/articles/customer-experience/good-net-promoter-score/ [accessed 2024-12-20]
22. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
23. ATLAS.ti. URL: https://atlasti.com [accessed 2024-11-15]
24. Gell NM, Grover KW, Savard L, Dittus K. Outcomes of a text message, Fitbit, and coaching intervention on physical activity maintenance among cancer survivors: a randomized control pilot trial. J Cancer Surviv 2020 Feb;14(1):80-88. [doi: 10.1007/s11764-019-00831-4] [Medline: 31776849]
25. Malhotra R, Rahimi S, Agarwal U, Katz R, Kumar U, Garimella PS, et al. The impact of a wearable activity tracker and structured feedback program on physical activity in hemodialysis patients: the Step4Life pilot randomized controlled trial. Am J Kidney Dis 2023 Jul;82(1):75-83 [FREE Full text] [doi: 10.1053/j.ajkd.2022.12.011] [Medline: 36801430]
26. Mattison G, Canfell O, Forrester D, Dobbins C, Smith D, Töyräs J, et al. The influence of wearables on health care outcomes in chronic disease: systematic review. J Med Internet Res 2022 Jul 01;24(7):e36690 [FREE Full text] [doi: 10.2196/36690] [Medline: 35776492]
27. Ferguson T, Olds T, Curtis R, Blake H, Crozier AJ, Dankiw K, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022 Aug;4(8):e615-e626 [FREE Full text] [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]

XSL•FO
RenderX

28. Feng S, Mäntymäki M, Dhir A, Salmela H. How self-tracking and the quantified self promote health and well-being: systematic review. J Med Internet Res 2021 Sep 21;23(9):e25171 [FREE Full text] [doi: 10.2196/25171] [Medline: 34546176]
29. Davis F. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 1989;13(3):319-340. [doi: 10.2307/249008]
30. Triantafyllidis A, Kondylakis H, Katehakis D, Kouroubali A, Alexiadis A, Segkouli S, et al. Smartwatch interventions in healthcare: a systematic review of the literature. Int J Med Inform 2024 Oct;190:105560. [doi: 10.1016/j.ijmedinf.2024.105560] [Medline: 39033723]
31. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022 Apr 26;24(4):e35120 [FREE Full text] [doi: 10.2196/35120] [Medline: 35471414]
32. Mustafa AS, Ali N, Dhillon JS, Alkawsi G, Baashar Y. User engagement and abandonment of mHealth: a cross-sectional survey. Healthcare (Basel) 2022 Jan 24;10(2):221 [FREE Full text] [doi: 10.3390/healthcare10020221] [Medline: 35206837]
33. Robertson E, Naghavi N, Wipperman MF, Tuckwell K, Effendi M, Alaj R, et al. Digital measurement of mobility in pulmonary arterial hypertension: a structured review of an emerging area. Digit Health 2024;10:20552076241277174 [FREE Full text] [doi: 10.1177/20552076241277174] [Medline: 39291158]
34. Lattie EG, Cohen KA, Hersch E, Williams KDA, Kruzan KP, MacIver C, et al. Uptake and effectiveness of a self-guided mobile app platform for college student mental health. Internet Interv 2022 Mar;27:100493 [FREE Full text] [doi: 10.1016/j.invent.2021.100493] [Medline: 35036331]
35. Ash GI, Mak SS, Haughton AD, Augustine M, Bodurtha PO, Axtell RS, et al. College campus-based physical activity support: retrospective longitudinal analysis of intra- versus inter-personal components for uptake and outcome prediction. Journal of Medical Internet Research 2025 Jun 16;13:e51707 [FREE Full text] [doi: 10.2196/51707] [Medline: 40523272]
36. Viechtbauer W, Smits L, Kotz D, Budé L, Spigt M, Serroyen J, et al. A simple formula for the calculation of sample size in pilot studies. J Clin Epidemiol 2015 Nov;68(11):1375-1379. [doi: 10.1016/j.jclinepi.2015.04.014] [Medline: 26146089]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
GWI: Gulf war illness
HIPAA: Health Insurance Portability and Accountability Act
TFA: Theoretical Framework for Acceptability
VA: Veterans Affairs
WRIISC-NJ: War Related Illness and Injury Study Center in New Jersey

Edited by A Kushniruk; submitted 10.Aug.2025; peer-reviewed by S Huang, C Gissel, BL Niles; comments to author 03.Sep.2025; revised version received 04.Dec.2025; accepted 28.Dec.2025; published 20.Mar.2026.

Please cite as:
Mak SS, McManus P, Greer S, Haws K, Carlson CJ, Chandler HK, Osinubi O, Ash GI
Deployment of an Activity Monitoring Program to Complement a Clinical Intervention for Veterans With Gulf War Illness: Qualitative Study
JMIR Hum Factors 2026;13:e82177
URL: https://humanfactors.jmir.org/2026/1/e82177
doi:10.2196/82177
PMID:

©Selene S Mak, Pauline McManus, Steven Greer, Kari Haws, Carrie J Carlson, Helena K Chandler, Omowunmi Osinubi, Garrett I Ash. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 20.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Continuous Glucose Monitoring for Personalized Nutrition in Real-World Vively App Users: Retrospective Observational Study

Michelle R Jospe[1], PhD; Kelli M Richardson[1], PhD; Susan M Schembre[1], PhD

Georgetown Lombardi Comprehensive Cancer Center, Washington, DC, United States

Corresponding Author:
Susan M Schembre, PhD
Georgetown Lombardi Comprehensive Cancer Center
2115 Wisconsin Avenue NW Suite 300
Washington, DC, 20007
United States
Phone: 1 202 444 2223
Email: ss4731@georgetown.edu

Abstract

Background: The rising popularity of apps that sync with continuous glucose monitors (CGMs) reflects growing interest in on-demand, personalized care. These platforms combine real-time glucose biofeedback with self-monitored behaviors to optimize metabolic health among individuals with and without diabetes. However, little is known about user characteristics, engagement patterns, or factors associated with sustained use of CGM-integrated digital health apps in real-world settings.

Objective: This study aimed to describe user demographics, CGM usage patterns, and food logging behaviors among Vively app users and to identify characteristics of sustained engagement with CGM wear and food tracking.

Methods: We conducted a retrospective observational study of Vively app users between August 2021 and February 2025. Vively is a commercial digital health app that integrates with Abbott FreeStyle Libre CGMs to deliver personalized nutrition guidance. Users with at least 1 day of CGM wear were included. Primary outcomes were CGM wear duration (total days) and food logging engagement. Factors associated with engagement were identified using negative binomial regression for CGM wear and hurdle negative binomial models for food logging, adjusting for age, sex, BMI, baseline glucose, and device connectivity; the food logging model additionally adjusted for CGM wear category.

Results: The analytical sample included 7647 users (4782/6905, 69.3% female, mean age 44.4, SD 10.9 years, mean BMI 27.8, SD 6.1 kg/m²). Users wore CGMs for a median of 15 (IQR 14-30) days, with 42.7% (3263/7647) completing one full wear period (13-15 days) and 30.3% (2315/7647) completing 2 or more wear periods (≥28 days). Most users (7013/7647, 91.7%) logged food at least once, with a median of 47 (IQR 18-91) food entries over 12 days. Food logging declined progressively during CGM wear (mean 63.2%, SD 8) and dropped sharply after sensor removal (mean 2.4%, SD 1.1). In multivariate models, higher baseline glucose was associated with longer CGM wear (incidence rate ratio [IRR] 1.15, 95% CI 1.13-1.17) but fewer food logging days (IRR 0.96, 95% CI 0.94-0.98). Connected device syncing showed the strongest association for both CGM wear (IRR 1.32, 95% CI 1.28-1.37) and food logging (IRR 1.45, 95% CI 1.39-1.51). Older age and female sex were associated with higher engagement in both behaviors.

Conclusions: This large-scale analysis reveals distinct engagement patterns with CGM-integrated digital health applications. Food logging was largely concurrent with active CGM wear, dropping dramatically in CGM-free periods. The divergent associations of baseline glucose levels, with longer CGM wear but reduced food logging, may reflect different motivational drivers for passive monitoring versus active behavior tracking. These findings have important implications for designing sustainable digital health interventions that maintain user engagement beyond periods of biological feedback, though replication in more diverse samples and studies accounting for diabetes status and socioeconomic factors is needed.

(JMIR Hum Factors 2026;13:e80734) doi:10.2196/80734

KEYWORDS

precision health; digital health; metabolic health; personalized nutrition; blood glucose self-monitoring; biological feedback

XSL•FO
RenderX

Introduction

In recent years, a growing number of companies have begun offering direct-to-consumer web- or app-based programs that are paired with continuous glucose monitors (CGMs) to optimize well-being and improve metabolic health. These programs represent an emerging application of precision nutrition: a technology-enabled approach that leverages individual variability in genetic, metabolic, phenotypic, and behavioral factors to optimize dietary recommendations and health outcomes [1,2]. By providing real-time, CGM-based biofeedback on glucose responses to meals and lifestyle behaviors, these platforms are increasingly marketed to individuals without diabetes as tools for metabolic optimization. Users are encouraged to adjust their diet, meal timing, and physical activity based on their personalized guidance to promote weight loss and glycemic stability, an indicator of metabolic health [3]. Early clinical trials have begun to demonstrate the effectiveness of combining CGM with personalized nutrition therapy on health outcomes among people without diabetes [4-6].

Advances in CGM technology and artificial intelligence, the precision health movement, and increasing device accessibility [5] have driven the expansion of commercial CGM-integrated apps. While CGMs were originally developed for insulin-dependent diabetes management, manufacturers have expanded access by releasing over-the-counter versions in several countries, enabling broader use among health-conscious consumers. Paired with mobile apps that translate sensor data into behavioral insights, CGMs are being positioned not only as medical devices but also as tools for health behavior change [6-9]. Yet most existing research on CGM-integrated apps has focused on diabetes management populations [10-12], with less attention to predominantly nondiabetic consumers using intermittent wear patterns for metabolic optimization. Despite the increasing interest in CGM-based mobile health apps, there is little empirical evidence describing who uses these tools, how they are used in everyday, nonclinical settings, and which factors are associated with sustained engagement. Most commercial platforms do not publicly report user characteristics or behavioral engagement metrics, leaving a critical gap in real-world evidence that limits both scientific understanding and the design of effective, scalable interventions.

This study addresses this gap by leveraging large-scale observational data from real-world users of the Vively app (Vively Health Pty Ltd), a commercial, CGM-integrated platform focused on personalized nutrition and lifestyle modification. We characterize user demographics, CGM usage patterns, and food and activity tracking, and identify associations with sustained engagement. In doing so, this work provides rare, real-world insight into how individuals interact with CGM-based feedback tools in consumer settings, with implications for the design and evaluation of precision digital health interventions.

Methods

Study Design

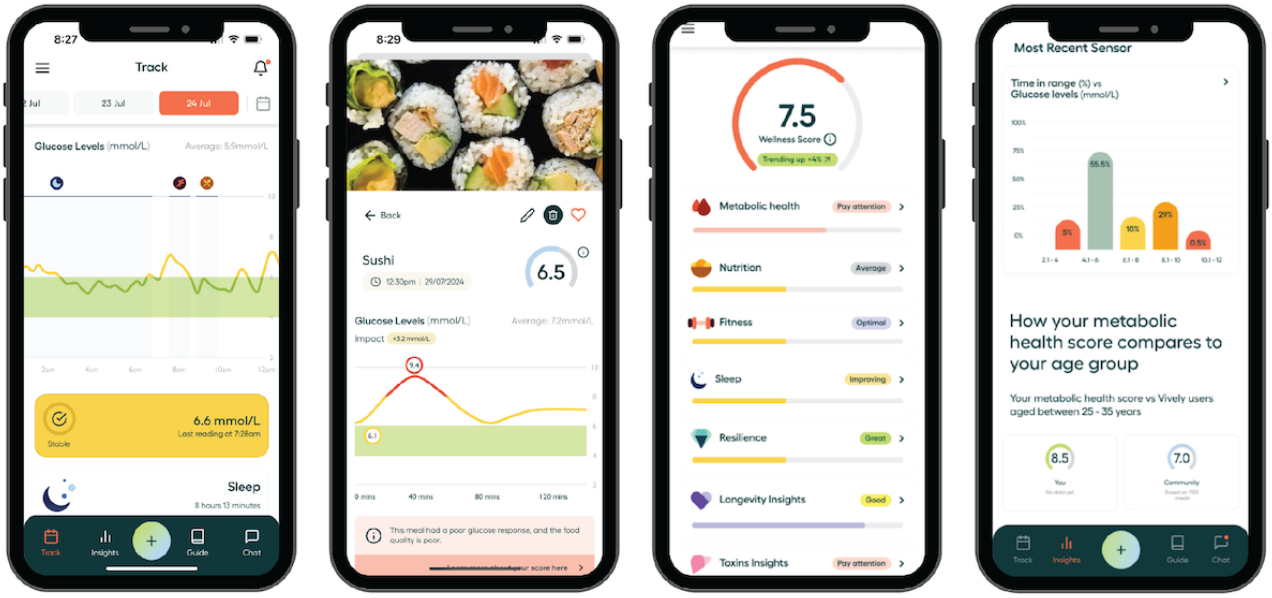
This was a retrospective observational study of Vively app users. The cohort entry window included users who initiated app use between August 2021 and November 2024. The follow-up window extended through February 2025, ensuring all users had at least 3 months of potential engagement data for analysis. Users were eligible for inclusion if they had at least one day of CGM sensor wear during the study period. The final analytical sample included 7647 users. This study followed STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for reporting observational research [13] (Multimedia Appendix 1).

Vively CGM Program

The Vively app is a commercially available digital health application that integrates with the 14-day Abbott FreeStyle Libre 2 and 3 CGM. It delivers real-time glucose feedback, meal scoring, and behavioral guidance to support metabolic health, weight management, and disease prevention. Designed for users with and without diabetes, the app combines CGM data with self-reported food and activity logs, allowing users to observe how lifestyle factors affect glucose levels (Figure 1).

XSL•FO
RenderX

Figure 1. Screenshots of the Vively app.



Vively includes meal-level scoring based on both nutrient quality and postprandial glucose responses. Users are encouraged to log meals during and between CGM wear periods to receive personalized feedback and track changes over time. The app can also sync with wearable devices to capture data on physical activity, sleep, and stress. Optional support from a registered dietitian is available as a paid add-on, offering one-on-one feedback for users seeking additional guidance.

Users typically begin their Vively experience by wearing a CGM sensor for 14 days to establish a baseline of glucose patterns and responses to food, activity, and lifestyle factors. During this period, users are encouraged to log meals and workouts to help the app generate personalized insights. After this initial wear period, users may continue to engage with the app without wearing a sensor. During these times, users can still log meals and receive guidance based on their previous CGM data and overall dietary patterns. Users are advised to repeat CGM wear approximately every 3 months as a refresher and to receive updated feedback. This intermittent approach is intended to balance the benefits of biological feedback with long-term usability and cost.

Study Outcomes

User engagement was assessed using 2 outcomes: daily food logging and CGM wear duration. Daily food logging was treated as a continuous variable (total food logging days) and as a binary outcome indicating whether a user ever logged food. CGM wear duration reflects the total number of days with CGM data. In addition to modelling total CGM wear duration, we categorized wear duration to reflect the intended, intermittent, CGM use pattern (2 weeks, every 3 months): less than 1 full wear duration (<13 days), 1 full wear duration (13-15 days), 1 to 2 wear durations (16-27 days), and 2 or more wear durations (≥28 days). A CGM wear duration of 13-15 days will herein be referred to as 1 CGM wear, and a CGM wear duration of 15 or more days will be referred to as 2 (or more) CGM wears. Food logging served as our primary measure of app engagement, since users received feedback mainly on this behavior, while CGM sensor wear was used as a secondary measure of engagement.

Diet quality and diet glucose scores, as derived by the Vively app, were extracted as indicators of the average nutritional and glycemic quality of meals logged over the cumulative CGM wear and food logging duration. The diet quality score (range 0-10) reflects the nutritional composition of meals, including macronutrient distribution, degree of food processing, and whether the food contained alcohol. The diet glucose score (range 0-10) is derived from CGM data and captures postprandial glucose responses over a 2-hour window following each reported meal, and includes area under the curve, time-in-target range, and peak glucose levels. For each user, average scores were calculated across all eligible meals to summarize how meals aligned with the app's metabolic goals. Higher scores for each diet-related outcome are more favorable.

Physical activity data were extracted from the Vively app as steps or workouts. Users with physical activity data were those who synced a connected device, such as a smartwatch or smartphone, to the Vively app. Step counts were inferred from any non-null step data linked to a third-party source (eg, Apple Health, Garmin). Workout data were collected from both third-party sources and user-logged events and were analyzed descriptively as a proxy for physical activity engagement.

Statistical Analysis

We used descriptive statistics to summarize user characteristics, CGM wear, and app usage, reporting mean (SDs) or median (IQRs), as appropriate for the data distribution. Diet scores, step counts, and workout frequency were summarized descriptively and not included in modelling. Step counts were restricted to a plausible daily range of 500-100,000 steps. The lower bound was chosen to exclude days likely reflecting nonwear rather than genuine sedentary behavior, as very low counts are implausible for a full wear day [14]. Days with step counts more than 3 SDs below an individual's mean were additionally

XSL•FO
RenderX

removed; only below-mean outliers were excluded, as high step counts remain physiologically plausible.

CGM wear and food logging were the outcomes of interest, modeled separately as correlates of app engagement. Daily food logging rates were calculated across the first 14 days of the first CGM wear period and the first sensor-free period following it. The analytical sample on each day was restricted to users who had reached that wear day or break day. For wear days, users with shorter initial wear periods were excluded from subsequent days' calculations. For break days, users who initiated a second sensor wear were excluded from subsequent days' calculations, as their first sensor-free period had ended. To identify factors associated with food logging, we used a hurdle, negative binomial model. The hurdle model accounts for zero inflation and overdispersion in the count of logging days by modelling two components: (1) a logistic regression estimating the odds of ever logging food (≥1 day vs none), and (2) a zero-truncated negative binomial regression estimating the number of logging days among users who logged at least once. To reduce the influence of extreme values, food logging days were winsorized at the 99th percentile (111 days) prior to modelling, retaining all users in the analysis while reducing the influence of the 76 users (1%) with extreme logging durations [15]. Variables in both components included age, sex (reference: female), BMI, baseline mean glucose, and connected device use. The CGM wear category was included in the food logging model as an adjustment variable to account for the association between wear duration and food logging opportunity. As the CGM wear category represents a study exposure rather than a user characteristic, its coefficients are not presented in the regression forest plot.

Total CGM wear days were modeled using a standard negative binomial regression, using the same set of variables. Food logging days were excluded from this model to avoid adjusting for a potential downstream behavior. All continuous variables were modeled linearly and checked for collinearity. Exponentiated coefficients are reported as odds ratios or incidence rate ratios (IRRs), with corresponding 95% CIs. All regression models estimate associations between variables. As this is an observational study without randomization or experimental manipulation, findings should be interpreted as descriptive associations rather than causal effects.

User characteristics and engagement outcomes of interest were compared across the predefined categories of CGM use using pairwise tests. For continuous variables, pairwise independent 2-tailed *t* tests were conducted; for categorical variables, pairwise Fisher exact tests were used. All *P* values were adjusted for multiple comparisons using the Bonferroni correction. Effect size was calculated using Eta-squared (η^2) for continuous variables and Cramér V for categorical variables. Users with missing values for any of the variables were excluded from regression models; no imputation was performed. Missing data were observed for age (n=274, 3.6%), BMI (n=293, 3.8%), and baseline mean glucose (n=218, 2.9%), resulting in 502 users (6.6%) excluded from both regression models, leaving an analytical sample of 7145. All analyses were conducted in R (version 4.2.2; R Foundation for Statistical Computing), using the *pscl* and *MASS* packages for count models.

Ethical Considerations

The Georgetown University Institutional Review Board reviewed the study and determined this secondary analysis of deidentified data to be exempt (STUDY00008558). On account creation, Vively users agree to the Health Privacy Policy, which explicitly permits the use of deidentified operational data for research purposes. Users may withdraw consent at any time by opting out within the app or by contacting Vively support. Data extraction was performed by Vively prior to transfer in accordance with their privacy policy and the terms of the data-sharing agreement, which governed how opt-out requests were handled. Data were deidentified before transfer. Data were transferred via a secure file transfer protocol under a formal data-sharing agreement that specified permissible uses, storage requirements, and publication rights, and prohibited reidentification. No direct identifiers were available to investigators. No compensation was provided. Vively provided deidentified data exports under a data-sharing agreement and had no role in study design, analysis, interpretation, or the decision to submit. Investigators retained full control of the analytic code and results.

Results

Vively User Characteristics

The analysis included a total of 7647 individuals who used Vively with at least 1 day of CGM sensor wear (Table 1). The cohort was predominantly female, with roughly equal proportions in the normal (2555/7354, 34.7%), overweight (2518/7354, 34.2%), and obesity (2164/7354, 29.4%) BMI categories. Age ranged from 18 to 88 years. Most users were based in Australia (7091/7571, 93.7%), where Vively Health is headquartered.

XSL•FO
RenderX

Table 1. User, engagement, and behavioral characteristics across continuous glucose monitor (CGM) wear categories.

Characteristics	All	<1	1	1-2	≥2	*P* value	Effect size[a,b,c]
Number of CGM wears, n (%)	7647 (100)	1298 (17.0)	3263 (42.7)	771 (10.1)	2315 (30.3)	—[d]	—
CGM wear duration (days), mean (SD)	27.8 (37.1)	8.5[e] (3.1)	14.6[f] (0.7)	22.0[g] (3.7)	59.1[h] (55.6)	<.001	0.32[b]
Sex, n (%)						.08	0.03[c]
Male	2120 (30.7)	379 (33.3)	876 (29.2)	208 (31.2)	657 (31.4)	—	—
Female	4782 (69.3)	760 (66.7)	2126 (70.7)	459 (68.8)	1437 (68.6)	—	—
Age (y), mean (SD)	44.4 (10.9)	41.2[e] (10.4)	43.8[f] (11.0)	44.2[f] (10.7)	47.0[g] (10.6)	<.001	0.03[b]
BMI (kg/m²), mean (SD)	27.8 (6.1)	28.1[e](5.8)	27.4[f] (5.7)	28.1[e] (6.1)	28.0[e] (6.6)	<.001	0.00[b]
Mean baseline CGM glucose (mmol/L)[i], mean (SD)	5.8 (1.1)	5.7[e] (1.0)	5.7[e](0.8)	5.7[e] (1.1)	6.0[f] (1.4)	<.001	0.01[b]
CGM days with food logging[i], mean (SD)	55.7 (35.1)	50.7[e] (36)	62.9[f] (34.6)	50.4[e] (33.5)	50.1[e] (33.9)	<.001	0.03[b]
Meals per day, mean (SD)	3.5 (1.4)	3.2[e] (1.5)	3.7[f] (1.5)	3.4[g] (1.3)	3.4[g] (1.3)	<.001	0.02[b]
Vively diet quality score[j], mean (SD)	8.1 (0.7)	8[e] (0.9)	8.1[f] (0.7)	8.1[f] (0.7)	8.2[g] (0.7)	<.001	0.01[b]
Vively diet glucose score[k], mean (SD)	8.7 (0.8)	8.8[e] (0.7)	8.7[f] (0.7)	8.8[e] (0.7)	8.6[f] (0.9)	<.001	0.01[b]
Users with connected devices, n (%)	4668 (61)	685[e] (52.8)	1886[f] (57.8)	483[g] (62.6)	1614[h] (69.7)	<.001	0.13[c]
Steps per day, mean (SD)	3831 (4024)	3769 (3999)	3803 (4084)	3954 (4216)	3854 (3910)	.90	0.00[b]

[a]Effect size calculated using Eta-squared (η^2) for continuous variables and Cramér V for categorical variables. These reflect different types of association and are not directly comparable.

[b]For η^2: small ≈ 0.01, medium ≈ 0.06, large ≈ 0.14.

[c]For Cramér V (df=3): small ≈ 0.06, medium ≈ 0.17, large ≈ 0.29.

[d]Not applicable.

[e,f,g,h]Values with e, f, g, h superscript letters indicate statistically significant differences in pairwise comparisons between CGM wear categories (<1, 1, 1-2, ≥2 wears) based on 2-sided independent *t* tests or Fisher exact tests with Bonferroni correction.

[i]Calculated from hours 24 to 96 of continuous glucose monitor wear.

[j]A high diet quality score indicates better nutrient composition and less food processing.

[k]A high diet glucose score indicates a more stable and controlled glucose response, while a low score suggests potentially more rapid and significant spikes in blood sugar.

CGM Wear Patterns

Users wore a CGM for a median of 15 (IQR 14-30) days. There were 2 peaks in total wear durations: 42.7% (3263/7647) of users had 1 CGM wear (13-15 days), and 30.3% (2315/7647) of users had 2 or more CGM wears (≥28 days; Figure 2). For those with 2 CGM wears (either partial or full wears, 16-30 days; 1583/7647, 20.7%), the median number of days between CGM wears was 31 (IQR 6-85) days, and patterns varied widely (Figure 3). There were 141 users (1.8%) with 10 or more CGM wears (≥140 days). Reasons for partial CGM wear (<14 days; 2069/7647, 27.1%,) were not recorded.

XSL•FO
RenderX

Figure 2. Distribution of total continuous glucose monitor wear days per user (N=7647). CGM: continuous glucose monitor.

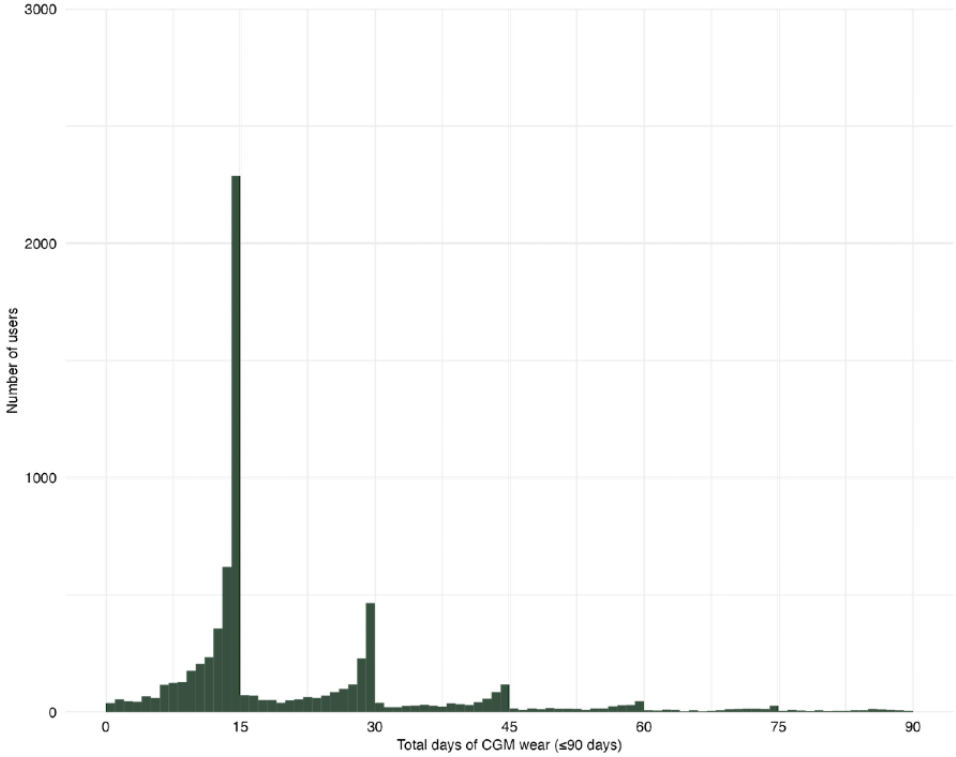
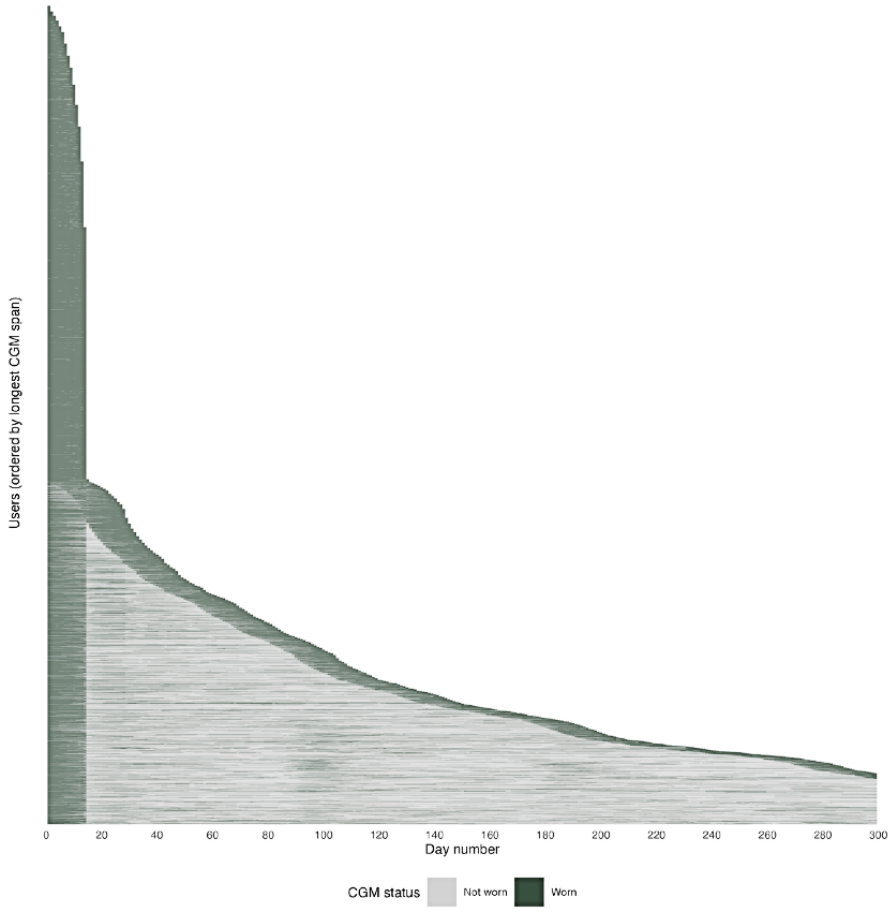


Figure 3. Continuous glucose monitor wear patterns among users over 300 days (N=7647). Each row represents an individual user, ordered by the longest span of CGM wear. Green tiles indicate days with continuous glucose monitor wear, and grey tiles indicate days without continuous glucose monitor wear. CGM: continuous glucose monitor.



XSL•FO
RenderX

Food Logging

Most users (7013/7647, 91.7%) logged at least 1 food item. The median number of entries was 47 (IQR 18-91) items over a median of 12 (IQR 6-19) days. One hundred users logged on for at least 100 days, and 8 users logged food on more than 300 days. As with CGM wear, the most common cumulative food logging duration was 13-15 days (960/7647, 12.6%). On days with food entries, users reported an average of 3.5 (SD 1.4) meals per day.

Food logging declined progressively during CGM wear and dropped sharply after sensor removal. During the first 14 days of the initial CGM wear period, daily food logging rates averaged 63.2% (SD 8), declining from 72.3% (5519/7638) on the first wear day to 50.7% (2500/4934) on the 14th wear day. After sensor removal, daily food logging rates averaged 2.4% (SD 1.1), declining from 5.4% (415/7636) on the first day after sensor removal to 1.6% (95/6091) on the 14th day after sensor removal.

Diet scores were presented to users for each logged meal and were scored from 0 to 10, with higher scores suggesting better food choices. Diet quality scores, which were based on nutrient composition and degree of food processing, were available for 77.5% (421,418/543,476 meals) of logged meals across 6876 users. To receive a score, a meal needed to contain at least one ingredient with nutrient information. The cumulative, per-user mean diet quality score was 8.1 (SD 0.7). Diet glucose scores, based on postprandial glucose response, were available for 86.3% of logged meals (469,279/543,476 meals) across 6917 users. The cumulative, per-user mean diet glucose score was 8.7 (SD 0.8).

Vively User, Engagement, and Behavioral Characteristics Across CGM Wear Categories

When comparing Vively users, engagement, and behavioral characteristics across CGM wear categories, most between-group differences were small in magnitude ($\eta^2 \leq 0.03$; Cramér V ≤0.13) and should be interpreted cautiously given that statistical significance is expected across nearly all comparisons at this sample size (Table 1). Compared to users with 1 total CGM wear, users with 2 or more CGM wears were older (47.0 (SD 10.6) years vs 43.8 (SD 11) years, $P<.05$), they had the highest mean baseline CGM glucose, they reported fewer meals per day (on average), and had slightly higher Vively diet quality scores. Users with 1 total CGM wear also had the highest proportion of CGM days with food logging (62.9%, SD 34.6), with users in all other wear categories logging on a smaller proportion of their CGM days (50.1%-50.7%). Users with less than 1 CGM wear were the youngest users; they were least likely to have a connected device synced with Vively, they reported the fewest number of meals per day, and had the lowest Vively diet quality score. Users with 1-2 CGM wears were not markedly different from users with 1 CGM wear or 2 or more CGM wears.

Steps and Workouts

Most users (4668/7647, 61%) connected Vively to a third-party device. The most common devices were the Apple Watch (3026/7647, 39.6%) and the Garmin watch (625/7647, 8.2%). Among those with connected devices, step counts were recorded on average for 80.6 (SD 79.2) days, with a mean daily step count of 3831 (SD 4024). Exercise logging was high, with 78.8% (6029/7647) users logging at least one workout. The number of workouts per user varied widely (median 28, IQR 7-93).

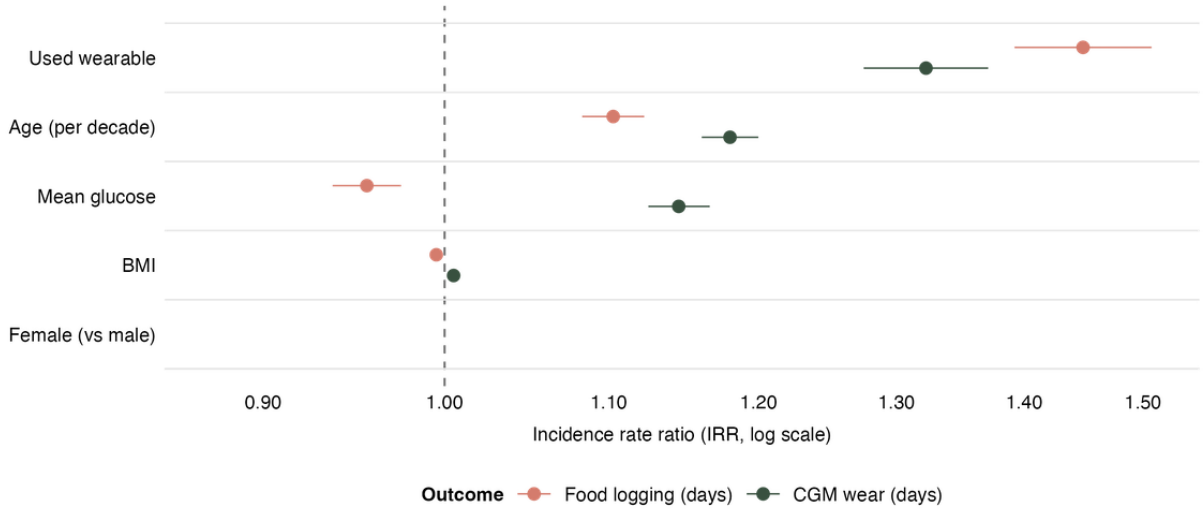
Factors Associated With Food Logging and CGM Wears

Regression analyses examined associations between measured user characteristics and engagement, but could not control for unmeasured factors such as diabetes status, medications, or socioeconomic circumstances.

Figure 4 depicts the factors associated with app engagement, measured as food logging (total days) and CGM wear duration (total days). Baseline mean glucose showed divergent associations across outcomes, with higher glucose associated with greater CGM wear but reduced food logging. Device syncing, older age, and being female were associated with higher engagement for both outcomes. Associations with BMI were small and inconsistent across models.

XSL•FO
RenderX

Figure 4. Factors associated with engagement behaviors. Incidence rate ratios with 95% CIs from negative binomial models of food logging days and continuous glucose monitor wear days. Incidence rate ratios are shown on a log scale; values >1 indicate positive associations. CGM: continuous glucose monitor; IRR: incidence rate ratio.



Among the 7013 Vively users (91.7% of 7647) who logged food at least once, females logged 14% more days than males (IRR 1.14, 95% CI 1.09-1.19). Each additional decade of age was associated with an 10% increase in logging days (IRR 1.10, 95% CI 1.08-1.12), and each 1 mmol/L higher mean glucose was associated with 4% fewer logging days (IRR 0.96, 95% CI 0.94-0.98). Users with a connected device logged 45% more days than those without (IRR 1.45, 95% CI 1.39-1.51). Logging days were marginally lower with higher BMI (IRR 0.995, 95% CI 0.992-0.998).

CGM wear duration was positively associated with baseline glucose (IRR 1.15, 95% CI 1.13-1.17). Similar to food logging, use of a connected device was associated with more CGM days (IRR 1.32, 95% CI 1.28-1.37). CGM wear days were 8% higher among women than men (IRR 1.08, 95% CI 1.04-1.12) and were slightly higher among users with higher BMI (IRR 1.01, 95% CI 1.00-1.01) and older age (IRR 1.18 per decade, 95% CI 1.16-1.20).

Discussion

Principal Results

This large-scale observational analysis of Vively users offers a rare window into real-world use of a digital health app that integrates CGM with self-tracked behaviors. The cohort of 7647 users was predominantly female and middle-aged, with a broad range of BMIs and mostly normoglycemic glucose levels. Most users wore a CGM for 15 days, with distinct peaks in wear duration at 13 to 15 and 28 to 30 days, reflecting 1 or 2 sensor periods, respectively. Users with repeat CGM use (2 or more wears) tended to be older and had higher baseline glucose levels, though differences in user characteristics across CGM wear categories were generally small in magnitude. Food logging, used as an indicator of app engagement, was largely confined to periods of active CGM wear, declining progressively during the wear period and dropping sharply after sensor removal. Several user characteristics were associated with engagement, with some variation by outcome. Higher glucose levels were linked to longer CGM wear but fewer food logging days. Consistent associations of greater engagement across both behaviors included device syncing, older age, and identifying as female. BMI showed small and inconsistent associations with engagement.

The level of engagement among this cohort of Vively app users, specifically in food logging, was notable, even when accounting for potential self-selection bias. The observed level of self-directed adherence suggests a motivational factor, likely driven by the immediacy and personal relevance of glucose feedback. This aligns with prior research showing that combining food logging with self-monitoring of behavioral outcomes, such as self-weighing during weight loss efforts, is associated with higher engagement and greater goal attainment [16]. Other studies have found that personalized feedback independently improves adherence to food logging [17-19]. CGM-based feedback may provide particularly salient motivation, as it offers immediate, personalized insights into the physiological consequences of food choices. While our observational design cannot determine whether CGM feedback directly causes increased engagement or whether highly motivated individuals are more likely to both wear CGMs and log food consistently, previous research has shown that self-monitoring diet is itself a key behavior in achieving health goals and has been identified as one of the strongest predictors of weight loss [16,20-22]. Nevertheless, food logging dropped sharply after CGM removal, even though the Vively app continues to provide personalized dietary feedback in the absence of sensor wear, suggesting that engagement with active self-monitoring behaviors is closely tied to the availability of real-time biological feedback.

Several factors were consistently associated with overall engagement with the app, including both food logging and CGM wear. Device syncing emerged as one of the strongest relationships across both outcomes, suggesting that integration of multiple data sources may be associated with higher engagement by providing more complete and rewarding feedback. Previous research has found that integrating

XSL•FO
RenderX

third-party devices, such as wearables, into interventions has been associated with improved engagement [23] and enhanced physical activity outcomes [24,25]. As wearable technology becomes more sophisticated and apps increasingly integrate multiple devices and data streams, the role of device connectivity in sustaining user engagement warrants further investigation. Understanding both how users interact with digital health technologies and whether that interaction supports behavior change is essential for designing effective interventions [26]. Older age was also associated with higher engagement in our sample, though this relationship is inconsistently observed across studies of nutrition app use [27]. Women were more likely to engage in both food logging and CGM wear. However, as is common in behavioral research, our sample included a greater proportion of women, which may have influenced these findings.

We found that baseline glucose showed divergent effects on engagement. While users with higher glucose were more likely to persist with CGM wear, they were paradoxically less likely to log food, potentially reflecting lower motivation for active diet tracking or greater burden of behavior change. Users with higher glucose levels may perceive greater value in the sensor data itself, focusing on monitoring rather than on modifiable behaviors such as dietary tracking. While these patterns suggest potential mechanisms, such as passive monitoring requiring less effort than active logging, or users with higher glucose prioritizing observation over intervention, our cross-sectional, observational design prevents us from determining the directionality or causality of these relationships. Longitudinal experimental studies would be needed to test these hypothesized mechanisms.

Comparison With Prior Work

Our findings are consistent with emerging evidence on engagement patterns in CGM-integrated digital health applications. The divergent associations between baseline glucose and engagement behaviors, with higher glucose associated with greater CGM wear but reduced food logging, echo patterns observed in other studies. Kumbara et al [11] reported that diabetes management app users with baseline mean glucose >10 mmol/L (180 mg/dL) were substantially less likely to log food intake, exercise, sleep, and medication use. This suggests that elevated glucose may signal lower capacity or motivation for active self-monitoring despite greater clinical need. Dehghani Zahedani et al [10] reported similar engagement patterns, with nondiabetic participants showing better adherence to food and activity logging during a 10-day CGM intervention. While their superior metabolic outcomes (91.7% of nondiabetic participants improving time in range compared to 58.3% with type 2 diabetes) could reflect better engagement, metabolic differences and medication use likely contribute substantially to these outcome differences independent of engagement behavior. The consistent pattern of reduced engagement among those with elevated glucose nevertheless suggests that metabolic dysregulation may create barriers to effortful self-monitoring, whether through competing priorities, psychological burden, or lower self-efficacy.

Our results also support the expected finding that passive monitoring requires less sustained effort than active tracking. Böhm et al [12] documented this pattern in a diabetes management app, where users engaged more consistently with automated data entry, such as CGM, compared to manual input like food and medication logs. More recent real-world evidence confirms that CGM adherence patterns predict glycemic outcomes, with higher adherence associated with greater hemoglobin A_{1c} improvements [28]. We observed similar dynamics, with CGM wear showing more sustained engagement than food logging, particularly after the initial sensor period. While the lower effort required for passive monitoring compared to active tracking is self-evident, passive data collection methods also eliminate recall bias and reduce administrative burden compared to manual self-monitoring [29], which may explain their superior adherence patterns. Quantifying this difference in real-world commercial settings provides empirical support for prioritizing low-burden technologies in intervention design, particularly for individuals managing chronic conditions or elevated health risks who may have limited capacity for intensive self-monitoring.

Our study contributes to this evidence base by examining engagement in a predominantly nondiabetic, commercial app context. While prior research has focused primarily on clinical diabetes management platforms, we demonstrate that these engagement dynamics also occur in consumer-oriented precision nutrition applications targeting metabolic optimization. The progressive decline in food logging from an average of 63.2% during wear to 2.4% after sensor removal underscores how strongly engagement depends on the presence of immediate biological feedback. This pattern emerged even among self-selected, paying users presumably motivated to improve their health, which has implications for designing sustainable digital health interventions that must maintain user engagement beyond periods of wearable sensor use.

Strengths and Limitations

As an observational study without a control group, randomization, or experimental manipulation, our findings represent associations rather than causal relationships. We cannot determine whether CGM feedback directly drives engagement behaviors or whether unmeasured factors influence both CGM adoption and self-monitoring behaviors. Additionally, we cannot establish temporal ordering, for example, whether baseline glucose influences subsequent engagement, or whether pre-existing engagement patterns predict glucose levels. To assess whether differential follow-up time could explain engagement differences across CGM wear categories, we compared available follow-up time across groups. Median follow-up time was broadly comparable across categories (45-70 weeks), and the correlation between follow-up duration and total wear days was negligible (Spearman ρ=0.06), suggesting that observed differences in wear reflect true engagement rather than observation time.

The dataset lacks information on diabetes status, medications, comorbidities, and socioeconomic status, factors that may influence both glucose levels and motivation to engage with health tracking technologies. These unmeasured factors could

XSL•FO
RenderX

confound observed associations. As an observational dataset of paying users of a commercial wellness app, the sample is self-selected and skewed toward women and Australian residents, likely representing health-motivated early adopters of CGM technology rather than typical patients or the general population. While we characterize this as "real-world" data because it reflects naturalistic, voluntary app use in commercial settings (as opposed to controlled research conditions with prescribed protocols), findings may not generalize to nonpaying users, clinical populations, men, or other geographic regions. Furthermore, step count was based on data from users' smartwatches or phones, without confirmation that devices were worn continuously, leading to potential underestimation when devices were removed or worn inconsistently.

However, the study's strengths include its large sample size and ecological validity, providing rare insight into real-world use of a CGM-integrated digital health app outside of research or clinical settings. The availability of granular engagement data across both passive and active behaviors also enables a nuanced understanding of how users interact with biological feedback tools. While future studies with comprehensive health and socioeconomic data and more diverse samples are needed to enhance generalizability and disentangle these relationships, the current findings offer important baseline information about engagement patterns among early adopters of CGM-integrated wellness technology for hypothesis generation and intervention design.

Conclusions

The rising popularity of apps that sync with CGMs reflects growing interest in on-demand, personalized care. This large-scale analysis of a commercial CGM-integrated app in naturalistic use underscores how user characteristics and engagement patterns among self-selected, paying users can inform the design of mobile-enabled tools that move healthcare beyond traditional models. While these findings reflect engagement among health-motivated early adopters, they provide important insights into the behavioral dynamics of CGM-integrated digital health interventions. As sensors become more affordable and longer-lasting [30], future research should explore not only how to sustain engagement beyond CGM use but also how long sensor wear continues to influence behavior. Identifying the behavioral drivers and specific features that support sustained engagement will be essential for developing interventions that are both scalable and effective [31-34]. This work should also expand to other apps and digital health platforms that use biofeedback to support personalized nutrition [35], to understand which tools and contexts most effectively promote lasting dietary behavior change. While our findings are associational and cannot establish causation, they provide valuable descriptive evidence of real-world usage patterns that can inform hypothesis generation for future experimental research. Gaps in our data, particularly the absence of diabetes diagnoses, medication information, comorbidity data, and socioeconomic indicators, highlight the need for future studies to comprehensively assess these factors to better understand how health status and social determinants influence CGM engagement patterns. Translating lessons from early adopters to broader, more diverse populations will require attention to access, usability, and behavioral support [33,36], especially for those who may benefit most from metabolic feedback but are less intrinsically motivated to actively self-monitor. Frameworks for understanding technology implementation, adoption, and sustainability can help guide efforts to scale these interventions across diverse settings and populations [37,38].

Acknowledgments

This manuscript was prepared with language editing assistance from ChatGPT-4o (OpenAI).

Funding

The authors declared no financial support was received for this work.

Data Availability

Aggregate summary statistics and analysis code are available from the corresponding author on reasonable request. The raw data are not publicly available due to commercialization restrictions.

Authors' Contributions

MRJ, SMS, and KMR contributed to the concept, design, and execution of the study. MRJ curated the data, conducted all analyses, and prepared the figures. MRJ and SMS co-wrote the manuscript. All authors contributed to manuscript review and editing.

Conflicts of Interest

MRJ declares past consultations with ZOE Limited and WW International, Inc. KMR declares current employment at WW International, Inc. Vively provided deidentified data exports under a data-sharing agreement and had no role in study design, analysis, interpretation, or the decision to submit. Investigators retained full control of the analytic code and results.

Multimedia Appendix 1
STROBE checklist.
[PDF File (Adobe PDF File), 125 KB - humanfactors_v13i1e80734_app1.pdf]

XSL•FO
RenderX

References

1. Berciano S, Figueiredo J, Brisbois TD, Alford S, Koecher K, Eckhouse S, et al. Precision nutrition: Maintaining scientific integrity while realizing market potential. Front Nutr 2022;9:979665 [FREE Full text] [doi: 10.3389/fnut.2022.979665] [Medline: 36118748]
2. Livingstone KM, Ramos-Lopez O, Pérusse L, Kato H, Ordovas JM, Martínez JA. Precision nutrition: a review of current approaches and future endeavors. Trends in Food Science & Technology 2022;128:253-264. [doi: 10.1016/j.tifs.2022.08.017]
3. Suh S, Kim JH. Glycemic variability: how do we measure it and why is it important? Diabetes Metab J 2015;39(4):273-282 [FREE Full text] [doi: 10.4093/dmj.2015.39.4.273] [Medline: 26301188]
4. Basiri R, Cheskin LJ. Personalized nutrition therapy without weight loss counseling produces weight loss in individuals with prediabetes who are overweight/obese: a randomized controlled trial. Nutrients 2024;16(14):2218 [FREE Full text] [doi: 10.3390/nu16142218] [Medline: 39064661]
5. Dixon W, Kim S, Levonian D, Gusz D, Fouladgar-Mercer S, Skyler JS. Novel glucose metric "Latest Spike Time" correlated with weight loss at six months in people with obesity using the signos system. Diabetes Technol Ther 2025;27(1):19-26. [doi: 10.1089/dia.2024.0222] [Medline: 39078656]
6. Richardson KM, Jospe MR, Bohlen LC, Crawshaw J, Saleh AA, Schembre SM. The efficacy of using continuous glucose monitoring as a behaviour change tool in populations with and without diabetes: a systematic review and meta-analysis of randomised controlled trials. Int J Behav Nutr Phys Act 2024;21(1):145 [FREE Full text] [doi: 10.1186/s12966-024-01692-6] [Medline: 39716288]
7. Jospe MR, Richardson KM, Saleh AA, Bohlen LC, Crawshaw J, Liao Y, et al. Leveraging continuous glucose monitoring as a catalyst for behaviour change: a scoping review. Int J Behav Nutr Phys Act 2024 Jul 10;21(1):74 [FREE Full text] [doi: 10.1186/s12966-024-01622-6] [Medline: 38987796]
8. Jospe MR, Kendall M, Schembre SM, Roy M. Real-world effectiveness of glucose-guided eating using the data-driven fasting app among adults interested in weight and glucose management: observational study. JMIR Form Res 2025;9:e65368 [FREE Full text] [doi: 10.2196/65368] [Medline: 40338170]
9. Veluvali A, Dehghani Zahedani A, Hosseinian A, Aghaeepour N, McLaughlin T, Woodward M, et al. Impact of digital health interventions on glycemic control and weight management. NPJ Digit Med 2025 Jan 09;8(1):20 [FREE Full text] [doi: 10.1038/s41746-025-01430-7] [Medline: 39789102]
10. Dehghani Zahedani A, Shariat Torbaghan S, Rahili S, Karlin K, Scilley D, Thakkar R, et al. Improvement in Glucose Regulation Using a Digital Tracker and Continuous Glucose Monitoring in Healthy Adults and Those with Type 2 Diabetes. Diabetes Ther 2021 Jul;12(7):1871-1886 [FREE Full text] [doi: 10.1007/s13300-021-01081-3] [Medline: 34047962]
11. Kumbara AB, Iyer AK, Green CR, Jepson LH, Leone K, Layne JE, et al. Impact of a Combined Continuous Glucose Monitoring-Digital Health Solution on Glucose Metrics and Self-Management Behavior for Adults With Type 2 Diabetes: Real-World, Observational Study. JMIR Diabetes 2023 Sep 11;8:e47638 [FREE Full text] [doi: 10.2196/47638] [Medline: 37590491]
12. Böhm AK, Jensen ML, Sørensen MR, Stargardt T. Real-world evidence of user engagement with mobile health for diabetes management: longitudinal observational study. JMIR Mhealth Uhealth 2020;8(11):e22212 [FREE Full text] [doi: 10.2196/22212] [Medline: 32975198]
13. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007 Oct 20;335(7624):806-808 [FREE Full text] [doi: 10.1136/bmj.39335.541782.AD] [Medline: 17947786]
14. Tudor-Locke C, Craig CL, Aoyagi Y, Bell RC, Croteau KA, De Bourdeaudhuij I, et al. How many steps/day are enough? For older adults and special populations. Int J Behav Nutr Phys Act 2011 Jul 28;8:80 [FREE Full text] [doi: 10.1186/1479-5868-8-80] [Medline: 21798044]
15. Flatt RE, Thornton LM, Tregarthen J, Argue S, Bulik CM. How engagement changes over time in a digital eating disorder app: observational study. JMIR Mhealth Uhealth 2025;13:e68824 [FREE Full text] [doi: 10.2196/68824] [Medline: 41027004]
16. Patel ML, Hopkins CM, Brooks TL, Bennett GG. Comparing self-monitoring strategies for weight loss in a smartphone app: randomized controlled trial. JMIR Mhealth Uhealth 2019;7(2):e12209. [doi: 10.2196/12209] [Medline: 30816851]
17. Turk MW, Elci OU, Wang J, Sereika SM, Ewing LJ, Acharya SD, et al. Self-monitoring as a mediator of weight loss in the SMART randomized clinical trial. Int J Behav Med 2013 Dec;20(4):556-561 [FREE Full text] [doi: 10.1007/s12529-012-9259-9] [Medline: 22936524]
18. Patel ML, Wakayama LN, Bennett GG. Self-monitoring via digital health in weight loss interventions: a systematic review among adults with overweight or obesity. Obesity (Silver Spring) 2021;29(3):478-499. [doi: 10.1002/oby.23088] [Medline: 33624440]
19. Rivera-Romero O, Gabarron E, Ropero J, Denecke K. Designing personalised mHealth solutions: an overview. J Biomed Inform 2023;146:104500 [FREE Full text] [doi: 10.1016/j.jbi.2023.104500] [Medline: 37722446]
20. Turner-McGrievy GM, Dunn CG, Wilcox S, Boutté AK, Hutto B, Hoover A, et al. Defining Adherence to Mobile Dietary Self-Monitoring and Assessing Tracking Over Time: Tracking at Least Two Eating Occasions per Day Is Best Marker of

XSL•FO
RenderX

Adherence within Two Different Mobile Health Randomized Weight Loss Interventions. J Acad Nutr Diet 2019 Sep;119(9):1516-1524 [FREE Full text] [doi: 10.1016/j.jand.2019.03.012] [Medline: 31155473]
21. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
22. Raber M, Liao Y, Rara A, Schembre SM, Krause KJ, Strong L, et al. A systematic review of the use of dietary self-monitoring in behavioural weight loss interventions: delivery, intensity and effectiveness. Public Health Nutr 2021 Dec;24(17):5885-5913 [FREE Full text] [doi: 10.1017/S136898002100358X] [Medline: 34412727]
23. Rayward AT, Vandelanotte C, Van Itallie A, Duncan MJ. The association between logging steps using a website, app, or fitbit and engaging with the 10,000 steps physical activity program: observational study. J Med Internet Res 2021;23(6):e22151 [FREE Full text] [doi: 10.2196/22151] [Medline: 34142966]
24. Tang MSS, Moore K, McGavigan A, Clark RA, Ganesan AN. Effectiveness of wearable trackers on physical activity in healthy adults: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2020;8(7):e15576. [doi: 10.2196/15576] [Medline: 32706685]
25. Ferguson T, Olds T, Curtis R, Blake H, Crozier AJ, Dankiw K, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022 Aug;4(8):e615-e626 [FREE Full text] [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
26. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
27. Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, et al. Factors Influencing Adherence to mHealth Apps for Prevention or Management of Noncommunicable Diseases: Systematic Review. J Med Internet Res 2022 May 25;24(5):e35371 [FREE Full text] [doi: 10.2196/35371] [Medline: 35612886]
28. Nemlekar PM, Hannah KL, Green CR, Norman GJ. Association between adherence, A1C improvement, and type of continuous glucose monitoring system in people with type 1 diabetes or type 2 diabetes treated with intensive insulin therapy. Diabetes Ther 2024;15(3):639-648 [FREE Full text] [doi: 10.1007/s13300-023-01529-8] [Medline: 38289464]
29. Germini F, Noronha N, Borg Debono V, Abraham Philip B, Pete D, Navarro T, et al. Accuracy and Acceptability of Wrist-Wearable Activity-Tracking Devices: Systematic Review of the Literature. J Med Internet Res 2022 Jan 21;24(1):e30791 [FREE Full text] [doi: 10.2196/30791] [Medline: 35060915]
30. Huang X, Yao C, Huang S, Zheng S, Liu Z, Liu J, et al. Technological Advances of Wearable Device for Continuous Monitoring of Glucose. ACS Sens 2024 Mar 22;9(3):1065-1088. [doi: 10.1021/acssensors.3c01947] [Medline: 38427378]
31. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
32. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
33. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
34. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
35. Richardson KM, Jospe MR, Saleh AA, Clarke TN, Bedoya AR, Behrens N, et al. Use of Biological Feedback as a Health Behavior Change Technique in Adults: Scoping Review. J Med Internet Res 2023 Sep 25;25:e44359 [FREE Full text] [doi: 10.2196/44359] [Medline: 37747766]
36. Guideline: Recommendations on Digital Interventions for Health System Strengthening. Geneva, Switzerland: World Health Organization; 2019. URL: https://www.ncbi.nlm.nih.gov/books/NBK541902/ [accessed 15-12-2025]
37. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
38. Greenhalgh T, Maylor H, Shaw S, Wherton J, Papoutsi C, Betton V, et al. The NASSS-CAT Tools for Understanding, Guiding, Monitoring, and Researching Technology Implementation Projects in Health and Social Care: Protocol for an Evaluation Study in Real-World Settings. JMIR Res Protoc 2020 May 13;9(5):e16861 [FREE Full text] [doi: 10.2196/16861] [Medline: 32401224]

Abbreviations

CGM: continuous glucose monitor

XSL•FO
RenderX

IRR: incidence rate ratio
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by J Sarvestan, LR Buis; submitted 15.Jul.2025; peer-reviewed by C Wang, Z Zandesh; comments to author 04.Nov.2025; accepted 10.Mar.2026; published 29.Apr.2026.

Please cite as:
Jospe MR, Richardson KM, Schembre SM
Continuous Glucose Monitoring for Personalized Nutrition in Real-World Vively App Users: Retrospective Observational Study
JMIR Hum Factors 2026;13:e80734
URL: https://humanfactors.jmir.org/2026/1/e80734
doi:10.2196/80734
PMID:

©Michelle R Jospe, Kelli M Richardson, Susan M Schembre. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 29.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Support, Monitoring, and Reminder Technology for Mild Dementia (SMART4MD) for People With Mild Cognitive Impairment and Their Informal Caregivers: Cost-Effectiveness Analysis

Zartashia Ghani[1], PhD; Johan Jarl[2], PhD; Peter Anderberg[1], PhD; Johan Sanmartin Berglund[1], PhD; Maria Quintana Aparicio[3], PhD; Pilar Barnestein-Fonseca[4], PhD; Selim Cellek[5], MD, PhD; Fermín Mayoral-Cleries[6], PhD; Maite Garolera[7], PhD; Gloria Guerrero-Pertiñez[6], MPGS; Karen Hayden[5], MSc; Carmel Moore[5], PhD; Jufen Zhang[5], PhD; Dominic Trepel[8], PhD; Sanjib Saha[2], PhD

[1]Blekinge Institute of Technology, Karlskrona, Sweden

[2]Department of Clinical Science (Malmö), Health Economics Unit, Lund, Skane, Sweden

[3]Consorci Sanitari de Terrassa, Torrebonica, Spain

[4]Instituto CUDECA de Estudios e Investigación en Cuidados Paliativos, Fundación CUDECA, Málaga, Spain

[5]Anglia Ruskin University, Chelmsford, United Kingdom

[6]Mental Health Clinical Management Unit, Regional University Hospital of Malaga, Malaga, Spain

[7]Consorci Sanitari de Terrassa, Barcelona, Spain

[8]Trinity College Dublin, Dublin, Ireland

Corresponding Author:
Sanjib Saha, PhD
Department of Clinical Science (Malmö)
Health Economics Unit
Forum Medicum, Sölvegatan 19
Lund, Skane, 223 62
Sweden
Phone: 46 736947976
Email: sanjib.saha@med.lu.se

Abstract

Background: Support, Monitoring, and Reminder Technology for Mild Dementia (SMART4MD), a customized tablet app, was developed to improve or maintain the quality of life of people with mild cognitive impairment (MCI) and their informal caregivers.

Objective: This study conducts an 18-month economic evaluation of the SMART4MD app, in addition to standard care, compared with standard care alone in Sweden and Spain, from a health care provider perspective.

Methods: In a pragmatic randomized controlled trial, people with MCI and their informal caregivers were randomized to the intervention and control groups. Health care costs, quality-adjusted life-years (QALYs), and incremental cost-effectiveness ratios (ICERs) were measured in 345 Swedish people with MCI and their informal caregivers, and in 347 Spanish people with MCI.

Results: The analysis showed higher incremental costs and lower QALYs for Swedish people with MCI than for controls, whereas higher incremental costs and higher QALYs were observed for Spanish people with MCI. The intervention was not found to be cost-effective for Swedish informal caregivers, with an ICER of €78,000/QALY (€1=US $1.16).

Conclusions: The differing findings regarding cost-effectiveness for people with MCI in Sweden and Spain highlight the need for further research with extended follow-up, ideally involving a larger sample size and conducted across different national contexts.

Trial Registration: ClinicalTrials.gov NCT03325699; https://clinicaltrials.gov/ct2/show/NCT03325699

(JMIR Hum Factors 2026;13:e77808) doi:10.2196/77808

KEYWORDS

cost-effectiveness; mHealth; dementia; mild cognitive impairment; Informal Caregivers

XSL•FO
RenderX

Introduction

Memory loss, or a decline in cognitive abilities such as language and spatial skills, is an early sign of diminishing mental abilities, referred to as mild cognitive impairment (MCI). MCI is a condition characterized by cognitive decline that exceeds what is typical for a person's age and education level, yet does not significantly impact daily activities [1]. This condition has the potential to progress into dementia over time. Approximately half of the people with MCI develop dementia within 3 years [2]. Dementia imposes a substantial financial burden on the health care sector, as well as emotional and financial stress on close relatives. The annual incidence of dementia in Sweden is estimated at 25,000 cases, with a corresponding societal cost of €45,000 (€1=US $1.16) per person [3]. The age-adjusted incidence rate of dementia among individuals aged 65 years or older in Spain is reported to be 6.66 per 1000 persons [4]. The total societal cost of dementia in Spain was reported to be €8268.5 million in 2018 [5]. According to the Dementia in Europe Yearbook 2019 [6], the prevalence of dementia in Sweden and Spain by 2050 is projected to be approximately 2.63% and 3.99% of the total population, respectively. Research further demonstrates that people with dementia incur higher health care costs up to 10 years before diagnosis compared with those without dementia [7].

MCI poses challenges to activities of daily living. People with MCI need active assistance from informal caregivers, for example, with taking medication, keeping medical appointments, and managing other daily activities. Caregiving responsibilities increase as the cognitive abilities of people with MCI decline over time [8]. This is expected to lead to emotional and mental stress among informal caregivers and affect the quality of life (QoL) of both people with MCI and their informal caregivers [9].

Smartphone/tablet apps have been suggested as useful tools to assist people with MCI in performing activities of daily living [10]. Regular use of smartphone/tablet apps by people with MCI may help them develop routines that become even more important, and thus beneficial, with increasing cognitive decline. Maintaining autonomy is expected to ease the burden on informal caregivers and improve the well-being of both people with MCI and their informal caregivers. Based on this notion, the tablet-based app Support, Monitoring, and Reminder Technology for Mild Dementia (SMART4MD) was developed to improve or maintain the QoL of people with MCI and their informal caregivers [11]. A pragmatic randomized controlled trial (RCT) was conducted in 3 European countries (Sweden, Spain, and Belgium) to study the effectiveness of the app in comparison with standard care over an 18-month period, from December 2017 to September 2020 (ClinicalTrials.gov registration number NCT03325699).

Our findings regarding the efficacy of the SMART4MD app over the 18-month trial period indicated a significant improvement in the quality of life at 6-month (measured as quality-adjusted life-years [QALYs]) of informal caregivers and dyads in the intervention group (authors' unpublished observations). Economic evaluations of such interventions can assist health policy makers and reimbursement agencies in making informed and transparent decisions. A short-term economic evaluation of the SMART4MD intervention, based on 6-month trial data, has previously been published. The results indicate that the intervention might yield benefits for informal caregivers in terms of reducing caregiving burden and improving their QoL [12]. However, the short follow-up period is not expected to fully capture the effects on health and QoL. Therefore, the purpose of this study is to assess the 18-month cost-effectiveness of the SMART4MD tablet-based intervention, in addition to standard care, compared with standard care, from the perspective of the health care provider in the Swedish and Spanish settings.

Methods

Study Design and Reporting Standards

This economic evaluation was conducted following the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) guidelines (see Multimedia Appendix 1), using data from the SMART4MD trial [13].

The SMART4MD Trial

SMART4MD aims to improve or maintain the quality of life of people with MCI and their informal caregivers. SMART4MD is a customized mobile health (mHealth) app for tablets, adapted for people with MCI. The app was developed by Healthbit Ltd [14] in consultation with people with MCI, their informal caregivers, and health care professionals. The app's primary features include medication and health care appointment reminders, cognitive support activities such as games and photos, and a health information feature that allows users to share their daily health status or health-related issues with others. The idea was to support people with MCI while also helping establish a routine in their daily lives that could be maintained as the disease progressed.

The SMART4MD trial was conducted at 4 sites in Sweden, Spain, and Belgium: (1) Blekinge Institute of Technology (BTH), Karlskrona, Sweden; (2) Consorci Sanitaria de Terrassa (CST), Barcelona, Spain; (3) Servicio Andaluz de Salud, Málaga, Spain; and (4) University College Leuven-Limburg, Belgium. One of the Spanish sites (Servicio Andaluz de Salud) was excluded from the analysis due to the unavailability of cost data, whereas the Belgian site was excluded due to insufficient participants (5 dyads) and the unavailability of both cost and effect data. However, we performed a scenario analysis for people with MCI from 1 Spanish site (CST), where inpatient and outpatient cost data were available only for people with MCI. Thus, this study conducted the primary analysis using complete 18-month trial data for both people with MCI and informal caregivers from the Swedish site (BTH), and a scenario analysis only for people with MCI from the Spanish site (CST).

The recruitment of participants was carried out through primary care and secondary care services (memory clinics), as well as outpatient clinics, day hospitals, specialist mental health care units, geriatric medicine units, and neurology service units. Older adults (≥55 years) who could administer their own medication and had a Mini-Mental State Examination (MMSE)

XSL•FO
RenderX

score ranging from 20 to 28 were eligible for inclusion. Further inclusion criteria were that participants were physically fit to use the app, had an informal caregiver, and had experienced impaired cognition for more than 6 months. Older adults were excluded if they scored above 11 on the 15-item Geriatric Depression Scale, indicating severe depression [15], or if they had a life expectancy of less than 3 years due to a life-limiting illness. A research nurse used information collected at the baseline visit and from participants' medical records to assess life expectancy.

A total of 345 dyads (people with MCI and their main informal caregiver) were randomized to either the intervention or control group, resulting in 173 dyads in the intervention group and 172 dyads in the control group. In addition, 347 people with MCI (intervention group, n=174; control group, n=173) participated from the Spanish site (CST). Participants randomly assigned to the intervention group received the SMART4MD app on data-enabled tablets, along with standard care. The intervention group also received a 1.5-hour training session, led by a research nurse, to demonstrate the functions of the app and was instructed to use the app daily, with assistance from their informal caregiver if needed. As data-enabled tablets and training were provided to trial participants, no technical prerequisites regarding familiarity with, or ownership of, smartphones or tablets were required for participation.

Participants in the control group received only standard care. Standard care in Sweden includes an annual in-person or telephonic visit to a physician for routine checkups and the renewal of drug prescriptions, as well as necessary laboratory tests, if required. Standard care in Spain includes a visit to a health center, where patients receive diagnostic, therapeutic, and preventive services and are potentially referred to specialists if necessary [16]. As all dyads continued to seek health care as needed, the extent of health care contacts varied between participants due to differences in general health, comorbidities, and health-seeking behavior. Further information on the SMART4MD app and the trial is described in detail elsewhere [11].

Ethics Considerations

The study was conducted in accordance with the ethical principles of the Declaration of Helsinki (World Medical Association, 2013). Ethical approval for the SMART4MD trial was obtained from the Regional Ethical Review Board in Sweden (LU numbers 650-00 and 744-00) and the local research ethics committee in Spain (approval number 02-16-107-029). Before participation, all older adults received comprehensive oral and written information about the study, including its objectives, data protection measures, and the voluntary nature of participation. Written informed consent was obtained from each participant before enrollment and reaffirmed at each data collection point.

Resource Use Measurement and Estimating Costs

A health care provider's (regional council's) perspective was adopted; thus, only costs incurred by the health care provider were included. Costs related to community, social, and home support services were not included, as these costs fall on municipalities rather than the health care provider. All costs were estimated in Swedish kronor (SEK) and reported in 2020 euros (€), using an exchange rate of 10.4867 SEK/€ [17].

All types of health care utilization for both inpatient and outpatient care were included in the study. Outpatient care includes both primary care and specialized outpatient clinics. Costs arising from the intervention include the purchasing cost of tablets, mobile data, and the time of the research nurse who provided training on how to use the app. All managerial costs related to the development and execution of the RCT were classified as research costs and were therefore excluded from the analysis.

Health care utilization and costs were collected from the Blekinge regional council health care registers [18], where costs were assigned to each health care episode using diagnosis-related groups [19]. In cases of missing outpatient health care costs, the average costs of visits in primary care and in each specialized outpatient clinic, stratified by type of health care personnel, were used (Table S1 in Multimedia Appendix 2). In a few cases, both cost and type of health care personnel were missing; in these instances, the average cost for each type of visit in specialized clinics was used. In cases where inpatient care costs were missing, the average daily cost was used.

Measurement of Effectiveness

EQ-5D-3L Index Score

The EQ-5D-3L questionnaire is a preference-based instrument that includes 5 attributes: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each attribute can be rated at 3 levels: no problem, some problem, and extreme problem, creating a total of 243 possible health states [20]. These health states are translated into a score ranging from 0 (equivalent to dead) to 1 (perfect health), based on population preferences [21]. The EQ-5D-3L questionnaire was collected for both people with MCI and informal caregivers at baseline and at 3 follow-up points (6, 12, and 18 months). EQ-5D-3L utilities [20] were estimated using the Swedish experience-based tariff [22].

Quality-Adjusted Life-Years

QALYs were included as the primary outcome measure for both people with MCI and informal caregivers, based on the EQ-5D-3L index score. The change in QALYs was estimated using the area-under-the-curve approach [23]. QALYs were calculated for each follow-up assessment at 6, 12, and 18 months. A 180-day window, with a tolerance of +18 or –18 days, was designated for each visit in the SMART4MD trial. However, as is common in trials, some participants' visits occurred outside this 180-day period. Consequently, QALYs were calculated using the actual number of days between visits instead of the fixed 180-day intervals. To determine the total QALYs for the entire 18-month trial period, the QALYs from each follow-up assessment were summed, and participants who missed any follow-up assessments were excluded.

Quality of Life in Alzheimer's Disease Scale

Quality of Life in Alzheimer's Disease Scale (QoL-AD) measures disease-specific quality of life and comprises 13 items

XSL•FO
RenderX

covering interpersonal, environmental, functional, physical, and psychological aspects of people with dementia [24]. Scores of 1 (poor), 2 (fair), 3 (good), and 4 (excellent) were combined to yield a total score, with higher totals indicating better quality of life [25]. We employed both self-rated and proxy-rated versions of QoL-AD. People with MCI were interviewed to obtain QoL-AD scores, while informal caregivers completed a self-administered questionnaire. Following Logsdon et al [26], we combined the 2 ratings into a weighted composite QoL-AD score, in which the rating of people with MCI was given 2 times the weight of the informal caregiver's proxy rating.

Mini-Mental State Examination

The MMSE score measures cognitive functioning [27]. It is a standardized assessment tool used to evaluate the cognitive abilities of people with MCI. This test focuses on memory, attention, orientation, and language skills, with a maximum score of 30 [27]. A higher MMSE score suggests a more positive outcome, while a score below 24 indicates the presence of dementia. However, there is no agreed-upon cutoff score for the MMSE to define MCI in the scientific literature [28].

Short-Form 12-Item Zarit Burden Interview

Caregiving burden was included as a secondary outcome measure for informal caregivers and was estimated using the short-form 12-item Zarit Burden Interview (ZBI-12). This self-reported instrument contains 12 items measuring the effect of caregiving activities on the informal caregiver's leisure time and overall health. A 5-point Likert scale (0=never to 5=nearly always) was used to score each question. The maximum score of the ZBI-12 is 48 points, where a score of less than 17 represents no burden [29]. To align with other outcome measures, we inverted the ZBI-12 score so that a higher score indicates a more favorable outcome. Thus, a score equal to or above 31 represents no burden.

Analysis of Cost-Effectiveness

The results are presented as an incremental cost-effectiveness ratio (ICER), which is the ratio of between-group differences in costs and effects, measured from baseline to the end of follow-up (18 months). According to the Swedish National Board of Health and Welfare, an ICER above 500,000 SEK (€48,876) is considered a high cost per QALY gained and is therefore used as the willingness-to-pay (WTP) threshold in this study [30]. Thus, if the estimated cost per QALY was below this threshold, the use of the SMART4MD app was considered cost-effective from a health care provider's perspective, compared with standard care. We also expressed the results in terms of net monetary benefit (NMB), which is calculated as follows: (incremental benefit × WTP) − incremental cost. The NMB provides results in monetary terms with a known threshold value of benefit. A positive NMB indicates that the tablet-based app is cost-effective compared with standard care. Some advantages of calculating NMB are that it is linear, has a simple sampling distribution, and is more stable than the ICER in cases of marginal differences in effectiveness between the intervention and comparator [31]. As both Swedish and Spanish guidelines recommend using a 3% discount rate in economic evaluations, we applied this rate to discount both QALYs and costs for the last 6 months of the trial [32].

The uncertainty around the cost-effectiveness results was explored using 5000 bootstrap resamples of the ICER, presented graphically on a 4-quadrant cost-effectiveness (CE) plane [33]. A cost-effectiveness acceptability curve (CEAC) was also derived to present the probability of the intervention being cost-effective at different WTP thresholds.

Sensitivity and Subgroup Analyses

We conducted several sensitivity and subgroup analyses to assess the uncertainty around the base-case results. Sensitivity analyses were reported for (1) people with MCI, (2) caregivers, and (3) dyads, where applicable (Textbox 1).

XSL•FO
RenderX

Textbox 1. Additional analyses.

1.Multiple imputation for missing quality-adjusted life-years

Missing information on the EQ-5D-3L index score was imputed using multiple imputation (20 datasets), based on baseline age, sex, education, living arrangements, EQ-5D-3L index score, Quality of Life in Alzheimer's Disease Scale score of people with mild cognitive impairment, Mini-Mental State Examination score, 12-item Zarit Burden Interview score, and intervention/control group [34]. After generating the multiple datasets, pooled estimates were computed [35].

2. Intervention cost

The intervention costs include the tablet purchasing cost, mobile data, and the time of the research nurse conducting training sessions on app usage. As these costs are not expected to be covered by the health care provider in the case of widespread implementation, they were included only in a sensitivity analysis.

3. Removing zero health care cost

During the trial, people with mild cognitive impairment were expected to have health care contacts due to their age and health condition. Thus, dyads with zero health care costs were removed in a sensitivity analysis, as these could be considered outliers.

4. Subgroup analysis—sex

Results were stratified by sex to explore whether they were influenced by differences in health-seeking behavior between men and women.

5. Subgroup analysis—age

Results were stratified by age (up to 70 vs ≥70 years) to explore whether they were influenced by differences in familiarity with and use of tablets, which is assumed to be correlated with age [36,37].

6. Subgroup analysis—Mini-Mental State Examination

Results were stratified based on baseline Mini-Mental State Examination score (≤26 vs >26) [38] to explore whether they were influenced by the level of cognitive decline at the time of introduction to the app. Furthermore, prior research indicates that this cutoff is advantageous due to its enhanced specificity and sensitivity [38,39]. Moreover, this threshold aligns with the median of our sample.

7. Subgroup analysis—Zarit burden

Results were stratified based on baseline Zarit burden score (no burden [12-item Zarit Burden Interview >31] vs some burden [12-item Zarit Burden Interview ≤31]) to explore whether they were influenced by the level of caregiving burden.

8. User behavior

Usage of the Support, Monitoring, and Reminder Technology for Mild Dementia (SMART4MD) app was assessed by the number of times it was launched during the first 12-month period (information was unavailable for the last 6 months of the trial). Results were stratified based on high versus low usage (median cutoff of 125 launches in our sample during the first year) to explore whether more frequent use of the app influenced the results.

Scenario Analyses

Two scenario analyses were performed. The first scenario analysis was conducted for one of the Spanish sites (CST), where cost data for people with MCI were available. A Spanish tariff was used to calculate QALYs [40]. Although Spain does not use an explicit threshold to establish the cost-effectiveness of health care interventions, a threshold of €30,000 per QALY is the most frequently cited in Spanish economic evaluation studies [41]. Therefore, we also used €30,000 per QALY as the threshold to establish the cost-effectiveness of the SMART4MD app for people with MCI participating at the Spanish site in this study. Health care utilization data were collected from Hospital de Terrassa, Terrassa Sanitary Consortium, Terrassa (Barcelona). All costs were estimated in euros (€) and reported in 2019 prices [42].

The second scenario analysis was conducted to evaluate the cost-effectiveness of people with MCI who participated from both BTH and CST sites, utilizing costs, QALYs, QoL-AD, and MMSE scores. The European EQ-5D VAS tariff was used to calculate QALYs in the second scenario [43], and a threshold value of €30,000 per QALY was applied [41].

Statistical Analysis

Intergroup (between intervention and control groups) statistical differences were assessed using an independent *t* test (unpaired and 2-tailed) for continuous variables and a chi-square test for categorical variables. Intragroup (within intervention and control groups) statistical differences were assessed using a paired *t* test. The scores of both people with MCI and their informal caregivers were combined to calculate dyad scores, and the results were presented as mean (SD) or mean differences. All analyses were conducted using Stata/SE 15.1 (StataCorp LP).

Results

Baseline Characteristics and Follow-Up Outcomes

Table 1 presents baseline and 18-month follow-up demographic characteristics of people with MCI and informal caregivers, where no considerable differences were observed between the intervention and control groups at baseline. At the 18-month follow-up, both people with MCI and dyads had significantly lower QALYs (P=.047 and .03, respectively) in the intervention group compared with the control group at the 12-month follow-up, whereas informal caregivers (P=.02) and dyads (P=.03) had significantly higher EQ-5D-3L index scores in the intervention group compared with the control group at the 18-month follow-up (Table S2 in Multimedia Appendix 2).

XSL•FO
RenderX

After 18 months, of the 345 dyads, 73 (21.2%) had left the trial, with the most prominent reason being "simply do not want to participate" (45/73, 62%), followed by physical (9/73, 12%) and cognitive reasons (6/73, 8%; see Figure S1 in Multimedia Appendix 2). Further, we excluded 5 informal caregivers who had not reported any health outcomes. Three people with MCI had changed their primary caregivers during the trial. As it is not reasonable to compare different caregivers in terms of health outcomes, these cases were excluded from the analyses (see Figure S1 in Multimedia Appendix 2).

The dropout versus nondropout analysis of baseline characteristics for people with MCI showed that those who dropped out had significantly worse EQ-5D-3L index score at baseline in both the intervention (*P*=.03) and control (*P*=.0097) groups. People with MCI who dropped out of the control group were more likely to live with a spouse or partner and had worse MMSE scores. The dropout analysis for informal caregivers in the control group showed that, at baseline, those who dropped out had higher Zarit burden scores and reported worse proxy QoL-AD scores on behalf of people with MCI (see Tables S3 and S4 in Multimedia Appendix 2).

Table 1. Demographic characteristics of people with MCI[a] and informal caregivers at baseline and 18-month follow-up.

People with MCI							Informal caregiver					
	Baseline			18-month follow-up			Baseline			18-month follow-up		
	Intervention (n=173)	Control (n=172)	*P* value	Intervention (n=131)	Control (n=141)	*P* value	Intervention (n=173)	Control (n=172)	*P* value	Intervention (n=127)	Control (n=137)	*P* value
Age (years), mean (SD)	76 (5.06)	76 (5.18)	.74	77 (5.10)	78 (5.07)	.79	70 (10.45)	69 (11.35)	.51	71 (10.39)	71 (11.22)	.93
Gender, n (%)			.47			.54			.67			.69
Male	97 (56.1)	103 (59.9)		76 (58.0)	87 (61.7)		57 (32.9)	53 (30.8)		39 (30.7)	39 (28.5)	
Female	76 (43.9)	69 (40.1)		55 (42.0)	54 (38.3)		116 (67.1)	119 (69.2)		88 (69.3)	98 (71.5)	
Education[b], n (%)			.07			.01			.37			.30
Elementary education	57 (32.9)	64 (37.2)		47 (35.9)	52 (36.9)		44 (25.4)	36 (20.9)		38 (29.9)	31 (22.6)	
Secondary education	57 (32.9)	38 (22.1)		44 (33.6)	27 (19.1)		64 (37.0)	60 (34.9)		45 (35.4)	50 (36.5)	
Higher education	58 (33.5)	70 (40.7)		39 (29.8)	62 (44.0)		63 (36.4)	75 (43.6)		42 (33.1)	56 (40.9)	
Civil status, n (%)			.64			.52			.64			.67
Single	46 (26.6)	42 (24.4)		38 (29.0)	36 (25.5)		23 (13.3)	20 (11.6)		19 (15.0)	18 (13.1)	
Married/living together	127 (73.41)	130 (75.6)		93 (71.0)	105 (74.5)		150 (86.7)	152 (88.4)		108 (85.0)	119 (86.9)	
Living arrangements, n (%)			.57			.01			.50			.29
Single	41 (23.7)	42 (24.4)		35 (26.7)	25 (17.7)		18 (10.4)	18 (10.5)		13 (10.2)	7 (5.1)	
Spouse/common law partner	130 (75.1)	130 (75.6)		93 (71.0)	109 (77.3)		151 (87.3)	147 (85.5)		109 (85.8)	124 (90.5)	
Children	1 (0.6)	0 (0)		1 (0.8)	0 (0)		2 (1.2)	6 (3.5)		1 (0.8)	0 (0)	
Other	1 (0.6)	0 (0)		2 (1.5)	7 (5.0)		2 (1.2)	1 (0.6)		4 (3.1)	6 (4.4)	

[a]MCI: mild cognitive impairment.

[b]n=130 people with MCI in the intervention group at follow-up; n=125 carers in the intervention group at follow-up.

Cost Measures

The average per-person cost for people with MCI was €7032 in the intervention group and €7024 in the control group over the 18-month period (see Table S5 in Multimedia Appendix 2). Outpatient care accounted for the majority of health care costs in both groups, representing total costs of €5466 (77.73%) in the intervention group and €5639 (80.28%) in the control group. The average cost per informal caregiver in the intervention group was €5371, while in the control group it was €4903. In contrast to the intervention group, informal caregivers in the control group had higher outpatient and lower inpatient costs.

XSL•FO
RenderX

However, no statistically significant differences were observed between the 2 groups (*P*=.63).

Effect Measures

Over the 18-month period, people with MCI in both groups experienced a statistically significant decline (*P*<.001 in both groups) in their quality of life (composite QoL-AD score). However, there was a significant improvement in MMSE scores for people with MCI in both groups (*P*<.001 in the intervention group and *P*=.02 in the control group; see Table S6 in Multimedia Appendix 2). Informal caregivers in both groups also experienced a decline in their quality of life (EQ-5D-3L index score) over the study period; however, this decline was statistically significant only in the control group (*P*<.001). The inverse ZBI-12 score indicated an insignificant decrease (*P*=.52) in caregivers' burden in the intervention group after 18 months. It further illustrates that dyads in the control group experienced a statistically significant decline in quality of life (EQ-5D-3L index score; *P*=.001) over the 18-month period. Intergroup differences for all effect measures were insignificant (for people with MCI: EQ-5D-3L, *P*=.72; composite QoL-AD, *P*=.73; for informal caregiver: EQ-5D-3L, *P*=.07; ZBI-12, *P*=.25; and for dyad: EQ-5D-3L, *P*=.25), except for MMSE scores, where the intervention group showed a statistically significant improvement (*P*=.03) compared with the control group.

Cost-Effectiveness Analysis

The intervention was dominated by standard care for people with MCI (Table 2), as the intervention group had higher costs (€9) and lower QALYs (−0.015) compared with the control group. However, the differences were not statistically significant for either costs (*P*=.99) or QALYs (*P*=.46). Additionally, the negative NMB (–€742) suggests that the intervention was not cost-effective. Of the 5000 incremental CE pairs for QALYs, 1908 (38.16%) were in the northwest quadrant (more costly and less effective) of the CE plane (Figure 1). At a WTP of €48,876 per QALY, the CEAC indicated an approximately 30% likelihood of the intervention being cost-effective (Figure 1). The ICERs for the composite QoL-AD and MMSE scores were €56 and €15 per unit gain, respectively.

For informal caregivers, the intervention group demonstrated an improvement in QALYs (0.006) at higher costs (€468) compared with the control group. The resulting ICER of €78,000 per QALY and a negative NMB (−€175) indicate that the intervention is not cost-effective at the WTP threshold of €48,876 per QALY. On the CE plane, 2157 out of 5000 (43.14%) CE pairs were in the northeast quadrant, indicating higher costs and higher effectiveness. The CEAC indicated that the likelihood of the intervention being cost-effective for caregivers was approximately 48% at a WTP of €48,876 per QALY (Figure 1). Moreover, due to a reduction in caregiving burden in the intervention group compared with the control group, the ICER for ZBI-12 was €613 per unit reduction in caregiving burden.

Table 2 shows higher costs (€630) and lower QALYs (−0.014) for dyads in the intervention group compared with the control group. These results, along with the negative NMB (−€1314), indicate that the intervention is dominated by standard care. The probability of the intervention being cost-effective at a WTP of €48,876 was estimated to be around 25% (Figure 1).

XSL•FO
RenderX

Table 2. Differences in pooled mean cost and health effects with 95% CI, ICERs[a], and NMB[b] (€[c])—BTH[d].[e]

Participants	Costs				Effects				ICERs	NMB
	Sample size[f] (intervention/control), n	Intervention, mean (SD)	Control, mean (SD)	ΔC (Bootstrap 95% CI)	Sample size[f] (intervention/control), n	Intervention, mean (SD)	Control, mean (SD)	ΔE (Bootstrap 95% CI)		
People with mild cognitive impairment										
Aggregated quality-adjusted life-years	128/143	7032 (9272)	7024 (7976)	9 (–2082 to 2099)	127/143	1.282 (0.16)	1.297 (0.16)	–0.015 (–0.05 to 0.02)	Dominated	–742
Composite quality of life in Alzheimer disease	128/143	7032 (9272)	7024 (7976)	9 (–2082 to 2099)	120/133	–1.525 (3.56)	–1.687 (3.82)	0.162 (–0.75 to 1.07)	56	N/A[g]
Mini-Mental State Examination	128/143	7032 (9272)	7024 (7976)	9 (–2082 to 2099)	130/141	1.046 (2.37)	0.454	0.592 (0.06 to 1.12)	15	N/A
Carer										
Quality-adjusted life years	123/135	5371 (8509)	4903 (7014)	468 (–1432 to 2368)	122/132	1.299 (0.13)	1.292 (0.15)	0.006 (–0.03 to 0.04)	78,000	–175
12-item Zarit Burden Interview	123/135	5371 (8509)	4903 (7014)	468 (–1432 to 2368)	126/136	0.286 (4.94)	–0.478 (5.63)	0.764 (–0.51 to 2.04)	613	N/A
Dyad										
Quality-adjusted life years	123/135	11,877 (12,566)	11,247 (8810)	630 (–2038 to 3299)	121/132	2.590 (0.19)	2.604 (0.19)	–0.014 (–0.06 to 0.03)	Dominated	–1314

[a]ICER: incremental cost-effectiveness ratio.

[b]NMB: net monetary benefit.

[c]€1=US $1.16.

[d]BTH: Blekinge Institute of Technology.

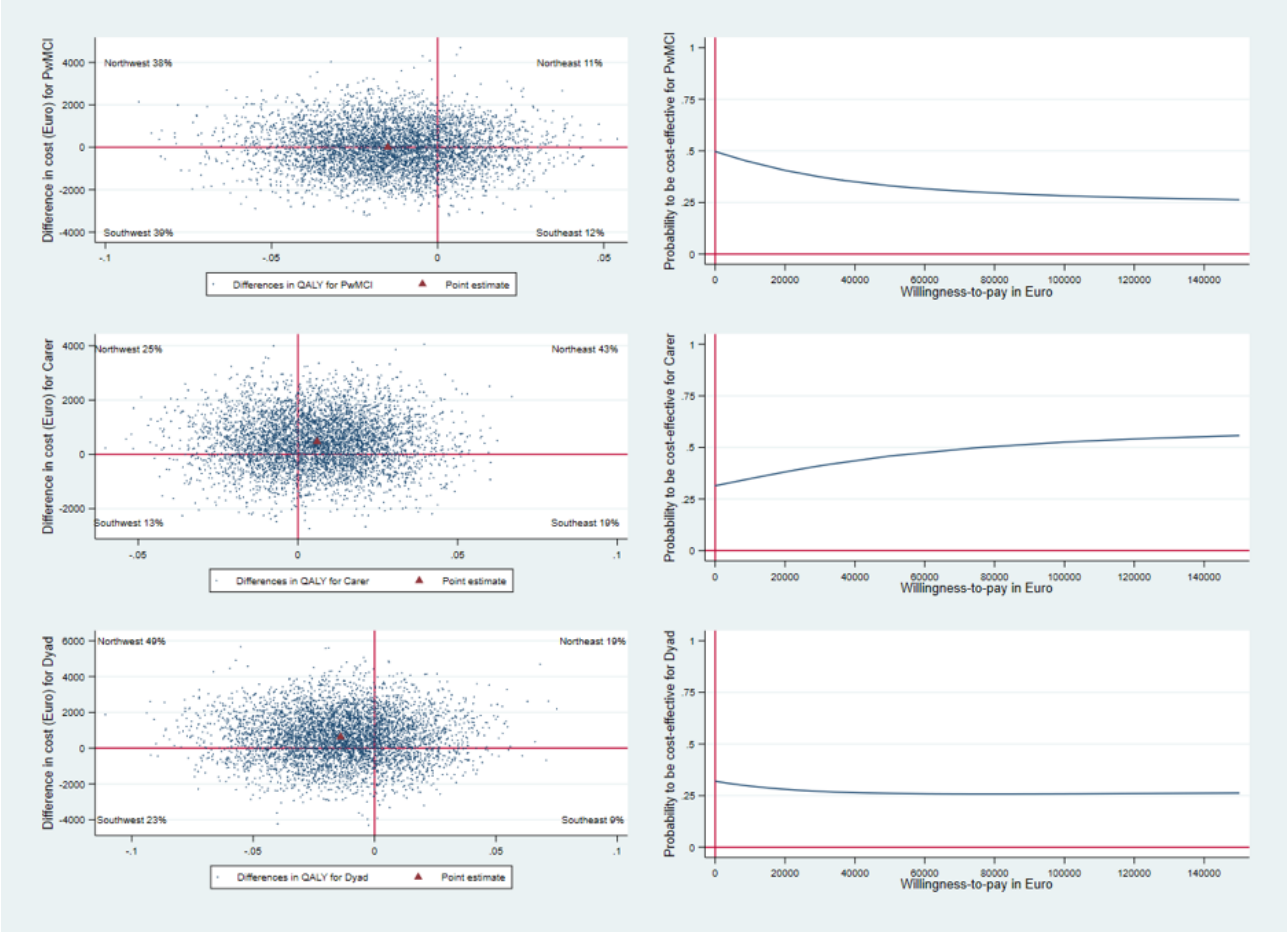
[e]Incremental effect with a positive value represents improved outcomes. We reversed the 12-item Zarit Burden Interview scores to obtain this. Deceased people with mild cognitive impairment are included in quality-adjusted life years and cost calculations. The incremental costs for dyads do not add up to the individual with mild cognitive impairment and informal caregivers cost differences due to different sample sizes.

[f]The number of participants available in the intervention group is followed by the number of participants available in the control group.

[g]N/A: not applicable.

XSL•FO
RenderX

Figure 1. Cost-effectiveness (CE) plane from the health care provider perspective and cost-effectiveness acceptability curve (CEAC), showing the probability that Support, Monitoring, and Reminder Technology for Mild Dementia (SMART4MD) is cost-effective across different willingness-to-pay (€) thresholds per quality-adjusted life-year (QALY) gained.



Sensitivity and Subgroup Analyses

Table 3 presents the results of the sensitivity and subgroup analyses. While numerical changes in costs and effects were observed across different analyses, the base-case findings remained generally consistent. Some exceptions were observed. Sensitivity analyses for people with MCI older than 70 years and with MMSE ≤26 showed that the intervention was less costly and less effective. For informal caregivers, the intervention was dominant among low app users. Further, the intervention was cost-effective for informal caregivers older than 70 years, female informal caregivers, those with health care costs greater than zero, and when QALYs were imputed.

XSL•FO
RenderX

Table 3. Sensitivity analyses from the health care provider perspective in ICERs[a]—Blekinge Institute of Technology.[b]

Number	Scenarios	Sample size[c], n	Changes in cost (bootstrap 95% CI)	Changes in effect (quality-adjusted life-years; bootstrap 95% CI)	ICER (€[d])
Intervention	Control				
People with mild cognitive impairment					
Base case	128/127	143/143	9 (–2082 to 2099)	–0.015 (–0.05 to 0.02)	Dominated
1. Imputed quality-adjusted life-years change	128/173	143/172	9 (–2082 to 2099)	–0.003 (–0.04 to 0.03)	Dominated
2. With intervention cost	128/127	143/143	182 (–1909 to 2272)	–0.015 (–0.05 to 0.02)	Dominated
3. Removing zero health care cost	122/121	138/138	100 (–2013 to 2213)	–0.014 (–0.05 to 0.03)	Dominated
4a. Men only	74/74	89/88	–445 (–3162 to 2272)	–0.010 (–0.06 to 0.04)	44,500
4b. Women only	54/53	54/54	824 (–2310 to 3958)	–0.017 (–0.07 to 0.03)	Dominated
5a. Age ≤70 years	18/18	19/19	1484 (–2057 to 5025)	0.028 (–0.03 to 0.09)	53,000
5b. Age >70 years	110/109	124/124	–204 (–2473 to 2065)	–0.022 (–0.06 to 0.02)	9273
6a. Mini-Mental State Examination score ≤26	48/47	39/39	–1783 (–5028 to 1461)	–0.016 (–0.09 to 0.06)	111,438
6b. Mini-Mental State Examination score >26	80/80	104/104	853 (–1863 to 3569)	–0.008 (–0.05 to 0.03)	Dominated
7a. High app use >125	66/65	143/143	–159 (–2751 to 2434)	–0.002 (–0.04 to 0.04)	79,500
7b. Low app use ≤125	61/61	143/143	158 (–2504 to 2819)	–0.007 (–0.04 to 0.03)	Dominated
Informal caregiver					
Base case	123/122	135/132	468 (–1432 to 2368)	0.006 (–0.03 to 0.04)	78,000
1. Imputed quality-adjusted life-years change	123/173	135/170	468 (–1432 to 2368)	0.011 (–0.02 to 0.04)	42,545
2. Removing zero health care cost	110/109	116/114	300 (–1807 to 2406)	0.012 (–0.03 to 0.05)	25,000
3a. Men only	39/39	39/38	960 (–2969 to 4888)	–0.005 (–0.04 to 0.03)	Dominated
3b. Women only	84/83	96/94	254 (–1932 to 2440)	0.008 (–0.04 to 0.05)	31,750
4a. Age ≤70 years	55/55	57/56	663 (–1231 to 2558)	–0.024 (–0.08 to 0.03)	Dominated
4b. Age >70 years	68/67	78/76	435 (–2579 to 3449)	0.028 (–0.01 to 0.07)	15,536
5a. 12-Item Zarit Burden Interview score >31 (no burden)	117/116	125/122	631 (–1369 to 2632)	0.009 (–0.03 to 0.04)	70,111
5b. 12-Item Zarit Burden Interview score ≤31 (some burden)	6/6	10/10	–2378 (–5722 to 966)	–0.069 (–0.19 to 0.05)	34,464
6a. High app use >125	64/63	135/132	1528 (–1215 to 4271)	–0.003 (–0.04 to 0.04)	Dominated
6b. Low app use ≤125	59/59	135/132	–682 (–2603 to 1240)	0.016 (–0.03 to 0.06)	Dominant
Dyads (people with mild cognitive impairment plus an informal caregiver)					
Base case	123/121	135/132	630 (–2038 to 3299)	–0.014 (–0.06 to 0.03)	Dominated
1. Imputed quality-adjusted life-years change	123/173	135/170	630 (–2038 to 3299)	0.008 (–0.04 to 0.05)	78,750
2. With intervention cost	123/121	135/132	804 (–1865 to 3472)	–0.014 (–0.06 to 0.03)	Dominated
3. Removing zero health care cost	123/121	134/131	547 (–2149 to 3243)	–0.013 (–0.06 to 0.03)	Dominated

[a]ICER: incremental cost-effectiveness ratio.

[b]Incremental effect with positive value represents improved outcomes. We reversed 12-Item Zarit Burden Interview scores to obtain this.

[c]Number of participants available for cost estimation first, followed by the number of participants available for health effects.

[d]€1=US $1.16.

XSL•FO
RenderX

Scenario Analyses

Results for the scenario analysis of people with MCI who participated in the CST site are presented in Table 4. The intervention was more costly and more effective in terms of QALYs, with an ICER of €3653/QALY and an NMB of €1897, indicating that the intervention is cost-effective at the threshold of €30,000/QALY. The results for the composite QoL-AD score showed that the intervention was more costly and more effective, with an ICER of €17,533 per unit gain. However, a negative effect of the intervention in terms of MMSE indicates that the intervention was dominated by standard care. All differences in costs and outcomes between the intervention and control groups were statistically insignificant (difference in cost, *P*=.66; QALY Spain, *P*=.29; QoL-AD, .98; and MMSE, .55). Further, intergroup differences in demographic characteristics and health outcomes for the Spanish site (CST) are presented in Table S7 in Multimedia Appendix 2. The MMSE score was statistically significant at baseline (*P*=.04), indicating that people with MCI in the intervention group had lower cognitive abilities than those in the control group.

The results of the scenario analyses for people with MCI from both the BTH and CST sites indicated that the intervention was less costly and more effective compared with the control group, as measured by QALYs, QoL-AD, and MMSE scores.

Table 4. Scenario analyses of differences in pooled mean cost and health effects for people with mild cognitive impairment with 95% CI, ICERs[a], and NMB[b] (€[c]).[d]

Scenario analysis	Costs				Health effects					
	Sample size[e] (intervention/control), n	Intervention, mean (SD)	Control, mean (SD)	ΔC (bootstrap 95% CI)	Sample size[e] (intervention/control), n	Intervention, mean (SD)	Control, mean (SD)	ΔE (bootstrap 95% CI)	ICERs	NMB
Consorci Sanitaria de Terrassa										
Mean quality-adjusted life-years	70/69	1192 (4023)	929 (3051)	263 (–918 to 1444)	69/67	1.104 (0.41)	1.033 (0.41)	0.072 (–0.06 to 0.21)	3653	1897
Quality of life in Alzheimer disease	70/69	1192 (4023)	929 (3051)	263 (–918 to 1444)	68/75	0.211 (3.84)	0.196 (3.34)	0.015 (–1.17 to 1.20)	17533	N/A[f]
Mini-Mental State Examination	70/69	1192 (4023)	929 (3051)	263 (–918 to 1444)	73/80	–1.274 (4.05)	–0.913 (3.44)	–0.361 (–1.54 to 0.82)	Dominated	N/A
Consorci Sanitaria de Terrassa plus Blekinge Institute of Technology										
Mean quality-adjusted life-years	198/212	4968 (8302)	5040 (7349)	–72 (–1585 to 1440)	196/210	1.122 (0.26)	1.121 (0.26)	0.00021 (–0.05 to 0.05)	Dominant	2232
Quality of life in Alzheimer disease	198/212	4968 (8302)	5040 (7349)	–72 (–1585 to 1440)	188/208	–0.897 (3.75)	–1.008 (3.76)	0.111 (–0.63 to 0.85)	Dominant	N/A
Mini-Mental State Examination	198/212	4968 (8302)	5040 (7349)	–72 (–1585 to 1440)	203/221	0.21 (3.27)	–0.04 (2.79)	0.25 (–0.33 to 0.83)	Dominant	N/A

[a]ICER: incremental cost-effectiveness ratio.

[b]NMB: net monetary benefit.

[c]€1=US $1.16.

[d]No significant differences in incremental costs and effects were found (independent *t* test).

[e]Number of participants available for cost estimation first, followed by number of participants available for health effects.

[f]N/A: not applicable.

Discussion

Principal Findings

We performed an economic evaluation of the SMART4MD trial from a health care provider's perspective over the trial duration. Our findings indicate that, except for MMSE scores, there were no significant differences in effects and costs between the intervention and control groups. The estimated cost-effectiveness ratios show that the intervention is dominated by standard care for both people with MCI and dyads, and it was not found to be cost-effective for informal caregivers when QALYs were the outcome. However, for the Spanish site, the intervention was cost-effective, with an ICER of €3337/QALY.

In the base-case analysis, the intervention was not cost-effective in Sweden at the country-specific WTP threshold, whereas it was cost-effective in Spain. This divergence is likely driven primarily by structural and contextual differences between settings. First, country-specific cost inputs played an important

XSL•FO
RenderX

role: higher unit costs of health care utilization and intervention delivery in Sweden increased incremental costs relative to usual care, while lower unit costs in Spain resulted in smaller incremental costs. Second, baseline risk and disease severity differed between countries, with the Spanish population exhibiting a less favorable risk profile; consequently, the same relative treatment effect translated into larger absolute health gains and more favorable ICERs in Spain. Third, variation in usual care meant that more intensive standard care in Sweden reduced the marginal benefit of the intervention, whereas in Spain, a less intensive usual care comparator allowed greater incremental benefit. Taken together, these factors suggest that the observed cross-country differences in cost-effectiveness are driven more by contextual variation in costs, baseline risk, and usual care than by differences in the intrinsic effectiveness of the intervention itself, underscoring the need for caution when extrapolating our findings to health systems with substantially different cost structures and treatment pathways.

The findings of this study for people with MCI concur with our previously published short-term (6-month) economic evaluation of the SMART4MD trial [12]. Both studies indicate that the intervention is dominated by standard care for people with MCI and dyads in terms of QALYs and is not cost-effective for informal caregivers at a WTP threshold of €48,876 per QALY [12]. While the ICER for people with MCI, when aggregating data from the Swedish and Spanish sites, indicates a dominant point estimate, the differences in costs and QALYs were not statistically significant.

It is well established in the literature that using the EQ-5D-3L instrument might not be the most suitable approach for assessing the quality of life of people with dementia. This is because the EQ-5D-3L does not encompass attributes that adequately capture cognitive function [44], relationships with caregivers, and social support, which are recognized as important factors in measuring quality of life for people with dementia [45]. However, we also included other health outcome measures, such as MMSE scores, QoL-AD, and caregiving burden (ZBI-12 score), to cover these aspects to some extent and complement the results of the primary outcome measure.

For people with MCI, the ICERs for the remaining 2 outcome measures were €55 per unit gain in composite QoL-AD and €15 per unit gain in MMSE score, respectively. It is uncertain whether these values can be deemed cost-effective due to the absence of established cost-effectiveness thresholds for QoL-AD and MMSE. However, low ICER values may indicate that the societal value associated with a 1-unit change in these outcome measures is likely to be higher.

Similar to previous effectiveness and cost-effectiveness studies on the SMART4MD trial [12,46], this study also indicates an improvement in the quality of life of informal caregivers in the intervention group compared with the control group. This may suggest that informal caregivers gained the most from the intervention, even though it was intended to be used primarily by people with MCI themselves. It is possible that the app served as a reminder for informal caregivers regarding medication schedules and doctors' appointments for people with MCI. Furthermore, it is important to acknowledge that the involvement of multiple informal caregivers for an individual with dementia may enhance the benefits of SMART4MD, potentially leading to an underestimation of its impact.

The available research on the cost-effectiveness of mHealth interventions aimed at older adults with chronic conditions in a home-care setting is inconclusive [47]. For example, some studies have shown that mHealth interventions for patients with chronic obstructive pulmonary disease in the United Kingdom [48] and in Denmark [49] were not cost-effective. However, mHealth interventions were deemed cost-effective for older adults diagnosed with Parkinson disease [50] and for cardiac rehabilitation [42]. Therefore, comprehensive evidence regarding the cost-effectiveness of mHealth for older adults living with chronic conditions in general, and with MCI or dementia in particular, appears to be lacking. The generalizability of a specific mHealth intervention is also limited due to several factors, such as the learning curve of health care professionals and participants [48,49], the level of engagement of health care professionals [49], and participants' adherence rates [50].

In our previous study [12], we reasoned that the cost-effectiveness of the SMART4MD intervention might be underestimated due to the short (6-month) follow-up period. This study, conducted over a longer follow-up period of 18 months, still did not find significant differences between the intervention and control groups, especially with respect to QoL outcome measures. One reason could be that the 18-month period may still not be long enough to capture the full health effects. A model-based economic evaluation could be used to extrapolate the health effects of such an intervention over a lifetime horizon, which was beyond the scope of this study. By contrast, short follow-up periods are common in dementia research due to its rapid progression [51]. Thus, it is possible that the SMART4MD app is not a cost-effective intervention, as suggested by the results of our study. However, the results also show a statistically significant improvement in MMSE scores for people with MCI in the intervention group compared with the control group. The insignificant results for QoL outcome measures might be due to an underpowered trial with respect to cost-effectiveness analysis. The sample size of this trial was based on clinical effectiveness rather than cost-effectiveness. However, it is important to note that no significant differences were observed for QoL-AD, which was the primary outcome measure of the trial and was used to calculate the sample size [11].

A high dropout rate is a common issue when mHealth interventions are introduced [52], and the SMART4MD trial is no exception. At the 18-month follow-up, 42 of the 173 (24.2%) participants in the intervention group and 31 of the 172 (18.0%) in the control group had left the trial. As previously discussed [12,46], a potential explanation for leaving the trial could be the perceived burden of participation. However, a previously published qualitative study, using individual semistructured interviews, was conducted with 16 people with MCI who participated in the SMART4MD trial. The findings showed that these individuals did not perceive the app's reminder feature as beneficial [53], potentially affecting their level of engagement with the app. This is also evident from the data, which show that only 195 of 312 (62.5%) individuals with MCI accessed

XSL•FO
RenderX

the app between 6 and 12 months, compared with 299 out of 312 (95.8%) during the initial 6 months. Given that the difference in dropout rates between the groups was not statistically significant (P=.15), we assume that it does not have a meaningful impact on the cost-effectiveness results of this study. However, we conducted our primary analysis on an intention-to-treat basis, which preserves the benefits of randomization and reflects the effectiveness of offering the intervention under real-world patterns of uptake and adherence, including the observed decline in app use over time. In this context, a per-protocol analysis restricted to highly adherent users would likely yield more favorable ICERs but would be less representative of routine practice and more vulnerable to selection bias. We therefore did not perform such an analysis and interpret our results as estimates of cost-effectiveness under realistic, rather than optimal, adherence.

Missing cost data due to attrition were handled within the same multiple imputation framework used for missing outcomes, assuming data were missing at random, conditional on treatment group, baseline covariates, and observed intermediate costs and effects. All relevant cost components and utility measures at each time point were included in the imputation model, and imputed individual-level costs were then used to construct total costs over the trial horizon and estimate ICERs, with results pooled across imputations using rules proposed by Rubin [35]. While this approach is recommended over complete-case analysis, it still relies on untestable assumptions about the missingness mechanism; we therefore interpret the cost-effectiveness results with appropriate caution.

Study Strengths

The study has several strengths. Given the limited evidence available on the cost-effectiveness of mHealth interventions designed for people with MCI and their informal caregivers [12], this study extends the knowledge base for this type of intervention. Moreover, this economic evaluation is based on a pragmatic RCT [54], which improves the generalizability of the findings to older adults who visit primary care facilities seeking assistance with memory-related concerns. The absence of data on medication is one limitation of this study, as it is an important component of the health care provider's perspective. Furthermore, it is recommended in the literature to collect health care utilization data before randomization for a specific period, such as 1, 3, or 12 months. This prerandomization cost data could serve as a predictor of postrandomization costs and potentially capture variations in health care use before and after the trial [55]. Unfortunately, in our study, we did not have access to prerandomization cost data. Also, due to the absence of cost data from Spanish sites, the findings of this study relied exclusively on Swedish data; consequently, the results may not represent all participating sites. Finally, we were unable to analyze from a societal perspective due to the absence of crucial cost categories. These include the costs of social support, informal care, and participants' out-of-pocket expenses. Considering the positive outcomes for caregivers in this study, and that informal care constitutes around 50% of the total cost of dementia [56], an analysis from a societal perspective would have been informative.

Due to a lack of data regarding the app's actual utilization, our analysis is restricted to examining its presence, which offers people with MCI the opportunity to use it. Consequently, we also lack data on any behavioral changes associated with app usage. Therefore, the focus of this study was on enhancement in health-related quality of life, as such an improvement in QoL would suggest a behavioral change. Given the results, further investigations are necessary to determine whether the intervention failed to induce behavioral changes in participants or whether any behavioral changes observed had no impact on the primary outcome measure of health-related quality of life.

Neither the researchers nor the participants were blinded to the intervention and control groups, which could have introduced bias. However, blinding may be less critical in pragmatic trials that aim to closely replicate real-world practice [57]. Future studies could address this issue by providing an app with limited features to the control group. Another limitation was the lack of detailed characterization of informal caregivers. Key details, such as the primary caregiver's relationship with people with MCI, caregiving hours per day or week, and employment status, were not recorded. Additionally, the impact of the SMART4MD app was not analyzed based on caregiver type (spouse, adult child, or other), which could have offered valuable insights into which group benefited most from the intervention.

Conclusions

In conclusion, total costs and QALYs over the 18-month trial period were not significantly different between the intervention and control groups, indicating that the SMART4MD intervention was neither better nor worse than standard care. The results suggest that this type of tablet-based app for people with MCI and dyads is not cost-effective.

Acknowledgments

The authors are grateful to everyone who volunteered to participate in the SMART4MD trial, as well as to the staff for their collaboration in data collection and management. The authors also thank Region Blekinge for providing health care utilization data.

Funding

This study is part of SMART4MD (Support, Monitoring, and Reminder Technology for Mild Dementia), a European Commission Horizon 2020 project (grant 643399; SMART4MD Project Factsheet, H2020 CORDIS, European Commission). SS is an Atlantic Fellow for Equity in Brain Health at the Global Brain Health Institute (GBHI) and acknowledges support from GBHI, the Alzheimer's Association, and the Alzheimer's Society (grant GBHI ALZ UK-22-869016). Neither the financial sponsors nor

XSL•FO
RenderX

Healthbit Ltd played any role in the design, execution, analysis, or interpretation of the data; the write-up of the study; or the decision to publish.

Authors' Contributions

PA, JSB, MQA, PBF, SC, FMC, MG, GGP, KH, CM, and JZ designed and conducted the SMART4MD trial. ZG, SS, JJ, JSB, and PA conceived and designed the research. ZG performed the analyses. ZG, SS, and JJ interpreted the results. ZG wrote the first draft. JJ, PA, JSB, MQA, PBF, SC, FMC, MG, GGP, KH, CM, JZ, DT, and SS revised the manuscript for important intellectual content, and all authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Consolidated Health Economic Evaluation Reporting Standards (CHEERS) checklist.
[DOCX File , 23 KB - humanfactors_v13i1e77808_app1.docx]

Multimedia Appendix 2
Additional analysis.
[DOCX File , 830 KB - humanfactors_v13i1e77808_app2.docx]

References

1. Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, International Psychogeriatric Association Expert Conference on mild cognitive impairment. Mild cognitive impairment. Lancet 2006 Apr 15;367(9518):1262-1270. [doi: 10.1016/S0140-6736(06)68542-5] [Medline: 16631882]
2. van Maurik IS, Vos SJ, Bos I, Bouwman FH, Teunissen CE, Scheltens P, et al. Biomarker-based prognosis for people with mild cognitive impairment (ABIDE): a modelling study. The Lancet Neurology 2019 Nov;18(11):1034-1044. [doi: 10.1016/s1474-4422(19)30283-2]
3. Wimo A, Jönsson L, Fratiglioni L, Sandman PO, Gustavsson A, Sköldunger A, et al. The societal costs of dementia in Sweden 2012 - relevance and methodological challenges in valuing informal care. Alzheimers Res Ther 2016 Nov 18;8(1):59-11 [FREE Full text] [doi: 10.1186/s13195-016-0215-9] [Medline: 27986093]
4. Andreu-Reinón ME, Huerta JM, Gavrila D, Amiano P, Mar J, Tainta M, et al. Incidence of dementia and associated factors in the EPIC-Spain Dementia Cohort. JAD 2020 Nov 10;78(2):543-555. [doi: 10.3233/jad-200774]
5. Meijer E, Casanova M, Kim H, Llena-Nozal A, Lee J. Economic costs of dementia in 11 countries in Europe: estimates from nationally representative cohorts of a panel study. Lancet Reg Health Eur 2022 Sep;20:100445 [FREE Full text] [doi: 10.1016/j.lanepe.2022.100445] [Medline: 35781926]
6. Alzheimer Europe. Dementia in Europe Yearbook 2019 - Estimating the Prevalence of Dementia in Europe. Luxembourg City, Luxembourg: Alzheimer Europe; 2019.
7. Persson S, Saha S, Gerdtham U, Toresson H, Trépel D, Jarl J. Healthcare costs of dementia diseases before, during and after diagnosis: longitudinal analysis of 17 years of Swedish register data. Alzheimers Dement 2022 Dec 21;18(12):2560-2569 [FREE Full text] [doi: 10.1002/alz.12619] [Medline: 35189039]
8. Ahn IS, Kim J, Kim S, Chung JW, Kim H, Kang HS, et al. Impairment of instrumental activities of daily living in patients with mild cognitive impairment. Psychiatry Investig 2009 Sep;6(3):180-184 [FREE Full text] [doi: 10.4306/pi.2009.6.3.180] [Medline: 20046393]
9. Fisher GG, Franks MM, Plassman BL, Brown SL, Potter GG, Llewellyn D, et al. Caring for individuals with dementia and cognitive impairment, not dementia: findings from the aging, demographics, and memory study. J Am Geriatr Soc 2011 Mar;59(3):488-494 [FREE Full text] [doi: 10.1111/j.1532-5415.2010.03304.x] [Medline: 21391939]
10. Martínez-Alcalá CI, Pliego-Pastrana P, Rosales-Lagarde A, Lopez-Noguerola JS, Molina-Trinidad EM. Information and communication technologies in the care of the elderly: systematic review of applications aimed at patients with dementia and caregivers. JMIR Rehabil Assist Technol 2016 May 02;3(1):e6 [FREE Full text] [doi: 10.2196/rehab.5226] [Medline: 28582258]
11. Anderberg P, Barnestein-Fonseca P, Guzman-Parra J, Garolera M, Quintana M, Mayoral-Cleries F, et al. The effects of the Digital Platform Support Monitoring and Reminder Technology for Mild Dementia (SMART4MD) for people with mild cognitive impairment and their informal carers: protocol for a pilot randomized controlled trial. JMIR Res Protoc 2019 Jun 21;8(6):e13711 [FREE Full text] [doi: 10.2196/13711] [Medline: 31228177]
12. Ghani Z, Saha S, Jarl J, Andersson M, Berglund JS, Anderberg P. Short term economic evaluation of the digital platform “Support, Monitoring and Reminder Technology for Mild Dementia” (SMART4MD) for people with mild cognitive impairment and their informal caregivers. JAD 2022 Apr 19;86(4):1629-1641. [doi: 10.3233/jad-215013]

XSL•FO
RenderX

13. Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs AH, Carswell C, CHEERS 2022 ISPOR Good Research Practices Task Force. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. BMJ 2022 Jan 11;376:e067975 [FREE Full text] [doi: 10.1136/bmj-2021-067975] [Medline: 35017145]
14. healthbit. 2020 Nov 18. URL: https://www.healthbit.com/ [accessed 2026-04-15]
15. Greenberg S. The Geriatric Depression Scale (GDS). HIGN. 2012. URL: https://hign.org/sites/default/files/2020-06/Try_This_General_Assessment_4.pdf [accessed 2026-04-15]
16. Serrano JP, Latorre JM, Gatz M. Spain: promoting the welfare of older adults in the context of population aging. Gerontologist 2014 Oct;54(5):733-740 [FREE Full text] [doi: 10.1093/geront/gnu010] [Medline: 24632624]
17. Annual average exchange rates. The Central Bank of Sweden [Sveriges Riskbank]. 2020. URL: https://www.riksbank.se/en-gb/statistics/search-interest--exchange-rates/annual-average-exchange-rates/?y=2020&m=12&s=Comma&f=y [accessed 2022-05-17]
18. egionala priser och ersättningar för södra sjukvårdsregionen [Regional prices and reimbursements for the southern health care region]. Södra regionvårdsnämnden [Southern Regional Care Board]. Stockholm, Sweden: Södra regionvårdsnämnden [Southern Regional Care Board]; 2017. URL: https://sodrasjukvardsregionen.se/ [accessed 2026-04-15]
19. Diagnosis related group (DRG). Hawaii Medical Service Association (HMSA). URL: https://hmsa.com/portal/PROVIDER/zav_pel.fh.DIA.650.htm [accessed 2024-09-05]
20. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol group. Ann Med 2001 Jul;33(5):337-343. [doi: 10.3109/07853890109002087] [Medline: 11491192]
21. Dolan P. Modeling valuations for EuroQol health states. Med Care 1997 Nov;35(11):1095-1108. [Medline: 9366889]
22. Burström K, Sun S, Gerdtham U, Henriksson M, Johannesson M, Levin L, et al. Swedish experience-based value sets for EQ-5D health states. Qual Life Res 2014 Mar;23(2):431-442 [FREE Full text] [doi: 10.1007/s11136-013-0496-4] [Medline: 23975375]
23. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. Health Econ 2005 May;14(5):487-496. [doi: 10.1002/hec.944] [Medline: 15497198]
24. Hoe J, Katona C, Roch B, Livingston G. Use of the QOL-AD for measuring quality of life in people with severe dementia--the LASER-AD study. Age Ageing 2005 Mar;34(2):130-135. [doi: 10.1093/ageing/afi030] [Medline: 15713856]
25. Torisson G, Stavenow L, Minthon L, Londos E. Reliability, validity and clinical correlates of the Quality of Life in Alzheimer's disease (QoL-AD) scale in medical inpatients. Health Qual Life Outcomes 2016 Jun 14;14(1):90-98 [FREE Full text] [doi: 10.1186/s12955-016-0493-8] [Medline: 27301257]
26. Logsdon R, Gibbons L, McCurry S, Teri L. Quality of life in Alzheimer's disease: patient and caregiver reports. Journal of Mental health and Aging 1999;5:21-32 [FREE Full text]
27. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
28. Gluhm S, Goldstein J, Loc K, Colt A, Liew CV, Corey-Bloom J. Cognitive performance on the Mini-Mental State Examination and the Montreal Cognitive Assessment across the healthy adult lifespan. Cogn Behav Neurol 2013 Mar;26(1):1-5 [FREE Full text] [doi: 10.1097/WNN.0b013e31828b7d26] [Medline: 23538566]
29. Bédard M, Molloy DW, Squire L, Dubois S, Lever JA, O'Donnell M. The Zarit Burden Interview: a new short version and screening version. Gerontologist 2001 Oct;41(5):652-657. [Medline: 11574710]
30. Utvärdering av metoder i hälsa- och sjukvården och insatser i socialtjänsten. SBU. 2017. URL: https://www.sbu.se/globalassets/ebm/metodbok/sbushandbok.pdf [accessed 2024-09-05]
31. Paulden M. Calculating and interpreting ICERs and net benefit. Pharmacoeconomics 2020 Aug 11;38(8):785-807. [doi: 10.1007/s40273-020-00914-6] [Medline: 32390067]
32. Attema AE, Brouwer WBF, Claxton K. Discounting in economic evaluations. Pharmacoeconomics 2018 Jul 19;36(7):745-758 [FREE Full text] [doi: 10.1007/s40273-018-0672-z] [Medline: 29779120]
33. Drummond M, Sculpher M, Claxton K, Stoddart G, Torrance G. Methods for the Economic Evaluation of Health Care Programmes. Oxford, UK: Oxford University Press; 2015.
34. Lavori PW, Dawson R, Shera D. A multiple imputation strategy for clinical trials with truncation of patient data. Stat Med 1995 Sep 15;14(17):1913-1925. [doi: 10.1002/sim.4780141707] [Medline: 8532984]
35. Rubin D. Multiple Imputation for Nonresponse in Surveys. New York: Wiley & Sons; 1987.
36. Hunsaker A, Hargittai E. A review of Internet use among older adults. New Media & Society 2018 Jul 16;20(10):3937-3954. [doi: 10.1177/1461444818787348]
37. Zhang H, Wang H, Yan H, Wang X. Impact of internet use on mental health among elderly individuals: a difference-in-differences study based on 2016-2018 CFPS data. Int J Environ Res Public Health 2021 Dec 23;19(1):1-13 [FREE Full text] [doi: 10.3390/ijerph19010101] [Medline: 35010361]
38. Kukull W, Larson E, Teri L, Bowen J, McCormick W, Pfanschmidt M. The Mini-Mental State Examination score and the clinical diagnosis of dementia. J Clin Epidemiol 1994 Sep;47(9):1061-1067. [doi: 10.1016/0895-4356(94)90122-8] [Medline: 7730909]

XSL•FO
RenderX

39. Salis F, Costaggiu D, Mandas A. Mini-Mental State Examination: optimal cut-off levels for mild and severe cognitive impairment. Geriatrics 2023 Jan 12;8(1):12. [doi: 10.3390/geriatrics8010012]
40. Badia X, Roset M, Herdman M, Kind P. A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. Med Decis Making 2001;21(1):7-16. [doi: 10.1177/0272989X0102100102] [Medline: 11206949]
41. Vallejo-Torres L, García-Lorenzo B, Serrano-Aguilar P. Estimating a cost-effectiveness threshold for the Spanish NHS. Health Econ 2018 Apr;27(4):746-761. [doi: 10.1002/hec.3633] [Medline: 29282798]
42. National Statistics Institute/Instituto Nacional de Estadistica (INE). Spain in figures 2018. National Statistics Institute/Instituto Nacional de Estadistica (INE). 2018. URL: https://tinyurl.com/3swnt9se [accessed 2024-09-05]
43. Greiner W, Weijnen T, Nieuwenhuizen M, Oppe S, Badia X, Busschbach J, et al. A single European currency for EQ-5D health states. Results from a six-country study. Eur J Health Econ 2003 Sep 1;4(3):222-231. [doi: 10.1007/s10198-003-0182-5] [Medline: 15609189]
44. Hounsome N, Orrell M, Edwards RT. EQ-5D as a quality of life measure in people with dementia and their carers: evidence and key issues. Value Health 2011 Mar;14(2):390-399 [FREE Full text] [doi: 10.1016/j.jval.2010.08.002] [Medline: 21402307]
45. Neumann PJ. Health utilities in Alzheimer's disease and implications for cost-effectiveness analysis. Pharmacoeconomics 2005;23(6):537-541. [doi: 10.2165/00019053-200523060-00002] [Medline: 15960551]
46. Ghani Z, Eindrathant P, Saha S, Jarl J, Cleries F, Garolera M. Effectiveness of the digital platform, Support, Monitoring and Reminder Technology for Mild Dementia (SMART4MD) for people with mild cognitive impairment and their informal caregivers. Master's Programme in Public Health thesis. Lund University. 2025 Aug 9. URL: https://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=9098146&fileOId=9098148 [accessed 2026-04-14]
47. Ghani Z, Jarl J, Sanmartin Berglund J, Andersson M, Anderberg P. The cost-effectiveness of mobile health (mHealth) interventions for older adults: systematic review. Int J Environ Res Public Health 2020 Jul 22;17(15):1-11 [FREE Full text] [doi: 10.3390/ijerph17155290] [Medline: 32708016]
48. Stoddart A, van DPM, Pinnock H, Hanley J, McCloughan L, Todd A, et al. Telemonitoring for chronic obstructive pulmonary disease: a cost and cost-utility analysis of a randomised controlled trial. J Telemed Telecare 2015 Mar;21(2):108-118. [doi: 10.1177/1357633X14566574] [Medline: 25586810]
49. Witt Udsen F, Lilholt PH, Hejlesen O, Ehlers L. Cost-effectiveness of telehealthcare to patients with chronic obstructive pulmonary disease: results from the Danish 'TeleCare North' cluster-randomised trial. BMJ Open 2017 May 17;7(5):e014616. [doi: 10.1136/bmjopen-2016-014616]
50. Cubo E, Mariscal N, Solano B, Becerra V, Armesto D, Calvo S, et al. Prospective study on cost-effectiveness of home-based motor assessment in Parkinson's disease. J Telemed Telecare 2016 Jul 08;23(2):328-338. [doi: 10.1177/1357633x16638971]
51. Green C, Shearer J, Ritchie CW, Zajicek JP. Model-based economic evaluation in Alzheimer's disease: a review of the methods available to model Alzheimer's disease progression. Value in Health 2011 Jul;14(5):621-630. [doi: 10.1016/j.jval.2010.12.008]
52. Bidargaddi N, Musiat P, Winsall M, Vogl G, Blake V, Quinn S, et al. Efficacy of a web-based guided recommendation service for a curated list of readily available mental health and well-being mobile apps for young people: randomized controlled trial. J Med Internet Res 2017 May 12;19(5):e141. [doi: 10.2196/jmir.6775] [Medline: 28500020]
53. Piculell E, Skär L, Sanmartin Berglund J, Anderberg P, Bohman D. Using a mobile application for health communication to facilitate a sense of coherence: experiences of older persons with cognitive impairment. Int J Environ Res Public Health 2021 Oct 28;18(21):11332 [FREE Full text] [doi: 10.3390/ijerph182111332] [Medline: 34769845]
54. Ramsey SD, Willke RJ, Glick H, Reed SD, Augustovski F, Jonsson B, et al. Cost-effectiveness analysis alongside clinical trials II-An ISPOR Good Research Practices Task Force report. Value Health 2015 Mar;18(2):161-172 [FREE Full text] [doi: 10.1016/j.jval.2015.02.001] [Medline: 25773551]
55. Glick H, Doshi J, Sonnad S, Polsky D. Economic Evaluation in Clinical Trials. Oxford, UK: Oxford University Press; 2014.
56. Dementia statistics. Alzheimer's Disease International (ADI). 2023. URL: https://www.alzint.org/about/dementia-facts-figures/dementia-statistics/#:~:text=There%20are%20over%2010%20million,new%20case%20every%203.2%20seconds [accessed 2023-01-31]
57. Monaghan TF, Agudelo CW, Rahman SN, Wein AJ, Lazar JM, Everaert K, et al. Blinding in clinical trials: seeing the big picture. Medicina (Kaunas) 2021 Jun 24;57(7):647 [FREE Full text] [doi: 10.3390/medicina57070647] [Medline: 34202486]

Abbreviations

BTH: Blekinge Institute of Technology
CEAC: cost-effectiveness acceptability curve
CHEERS: Consolidated Health Economic Evaluation Reporting Standards
CST: Consorci Sanitaria de Terrassa
ICER: incremental cost-effectiveness ratio
MCI: mild cognitive impairment

XSL•FO
RenderX

mHealth: mobile health
MMSE: Mini-Mental State Examination
NMB: net monetary benefit
QALY: quality-adjusted life-year
QoL-AD: Quality of Life in Alzheimer's Disease Scale
RCT: randomized controlled trial
SEK: Swedish kronor
SMART4MD: Support, Monitoring, and Reminder Technology for Mild Dementia
WTP: willingness to pay
ZBI-12: 12-item Zarit Burden Interview

Edited by A Kushniruk; submitted 21.May.2025; peer-reviewed by Y Bobrova; comments to author 19.Mar.2026; revised version received 26.Mar.2026; accepted 27.Mar.2026; published 22.May.2026.

Please cite as:
Ghani Z, Jarl J, Anderberg P, Sanmartin Berglund J, Aparicio MQ, Barnestein-Fonseca P, Cellek S, Mayoral-Cleries F, Garolera M, Guerrero-Pertiñez G, Hayden K, Moore C, Zhang J, Trepel D, Saha S
Support, Monitoring, and Reminder Technology for Mild Dementia (SMART4MD) for People With Mild Cognitive Impairment and Their Informal Caregivers: Cost-Effectiveness Analysis
JMIR Hum Factors 2026;13:e77808
URL: https://humanfactors.jmir.org/2026/1/e77808
doi:10.2196/77808
PMID:

©Zartashia Ghani, Johan Jarl, Peter Anderberg, Johan Sanmartin Berglund, Maria Quintana Aparicio, Pilar Barnestein-Fonseca, Selim Cellek, Fermín Mayoral-Cleries, Maite Garolera, Gloria Guerrero-Pertiñez, Karen Hayden, Carmel Moore, Jufen Zhang, Dominic Trepel, Sanjib Saha. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 22.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Early Deployment of an Integrated Digital Platform (shamiriOS) for Scalable Youth Mental Health Service Delivery in Kenya: Development and Usability Study

Shadrack Lilan[1], BSc; Tom L Osborn[1], BA; Wendy Mmbone[1], BSc; Jean Kasudi[1], BPharm; Benny Otieno[1], BA; Ichami Etyang[1], BA; Jianing Tu[2], MBA, MSc; Edmund Korley[3], BSc; Christine Wasanga[1,4], PhD

[1]Shamiri Institute, Nairobi, Kenya

[2]Columbia University, New York, NY, United States

[3]The Agency Fund, San Francisco, CA, United States

[4]Kenyatta University, Nairobi, Kenya

Corresponding Author:
Shadrack Lilan, BSc
Shamiri Institute
13th Floor, CMS Africa, Chania Avenue
Nairobi, 00506
Kenya
Phone: 254 112540760
Email: shadrack.lilan@shamiri.institute

Abstract

Background: Scaling youth mental health services in low-resource settings requires digital infrastructure that supports not just clinical delivery but the full operational, supervisory, and engagement demands of community-based, task-shifted models. Existing platforms—whether commercial health systems, open-source medical records, or consumer-facing wellness apps—address fragments of this need, but none provide the integrated, offline-capable, and affordable architecture required for lay-provider delivery at scale.

Objective: We introduce shamiriOS, an open-source, modular digital platform comprising 3 interlinked suites—the Shamiri Digital Hub (SDH) for operational management, Rafi for youth engagement, and the Shamiri Provider Platform (SPP) for clinical workflows—designed for scalable, stepped-care youth mental health delivery in Kenya. Our objectives were (1) to conduct an environmental scan of existing platforms and characterize their limitations, (2) to describe the user-centered design and development of shamiriOS, and (3) to report early deployment outcomes across centralized and decentralized settings between 2023 and 2024.

Methods: We conducted a structured environmental scan of 6 case management platforms and 28 youth-facing mental health apps, assessing cost, usability, open-source availability, customizability, offline capability, and suitability for task-shifted delivery. Based on identified gaps, SDH was built as a browser-based operation platform, and Rafi was developed as a native mobile app (Android or iOS) with an offline-first architecture. The SPP was adapted from an existing electronic medical record system. Development followed a user-centered design process with community consultation, including cocreation workshops with 77 university-aged youths. Deployment was evaluated using use analytics, usability ratings, and Net Promoter Scores.

Results: No reviewed platform met the combined requirements for stepped-care delivery. SDH was deployed across 11 sites serving 76,344 youths via 1195 lay providers and 111 clinical supervisors by Q1 2024. Staff reported high satisfaction (usability: mean 8.36/10, SD 1.49; Net Promoter Score: mean 8.63/10, SD 1.46). Rafi achieved 74.7% (n=3737) registration at Mount Kenya University, with 50.4% (508/1008) booking therapy sessions, but sustained engagement with self-guided features declined to near 0 by 9 months.

Conclusions: shamiriOS demonstrates the feasibility of building modular, open-source digital infrastructure for scalable, task-shifted youth mental health delivery. Its component-based architecture is designed for adaptation to other contexts, though extension would require participatory redesign. The most significant obstacles to impact lie not in platform design but in implementation readiness, incentive alignment, and institutional integration. Future priorities include SPP deployment, artificial intelligence–assisted supervision features (shamiriAI), and strengthening sustained engagement.

(JMIR Hum Factors 2026;13:e79107) doi:10.2196/79107

XSL•FO
RenderX

KEYWORDS

digital health platforms; digital mental health; task-shifting; stepped-care; youth mental health; open-source platforms; implementation science; AI in mental health; artificial intelligence; low- and middle-income countries; Kenya

Introduction

Background

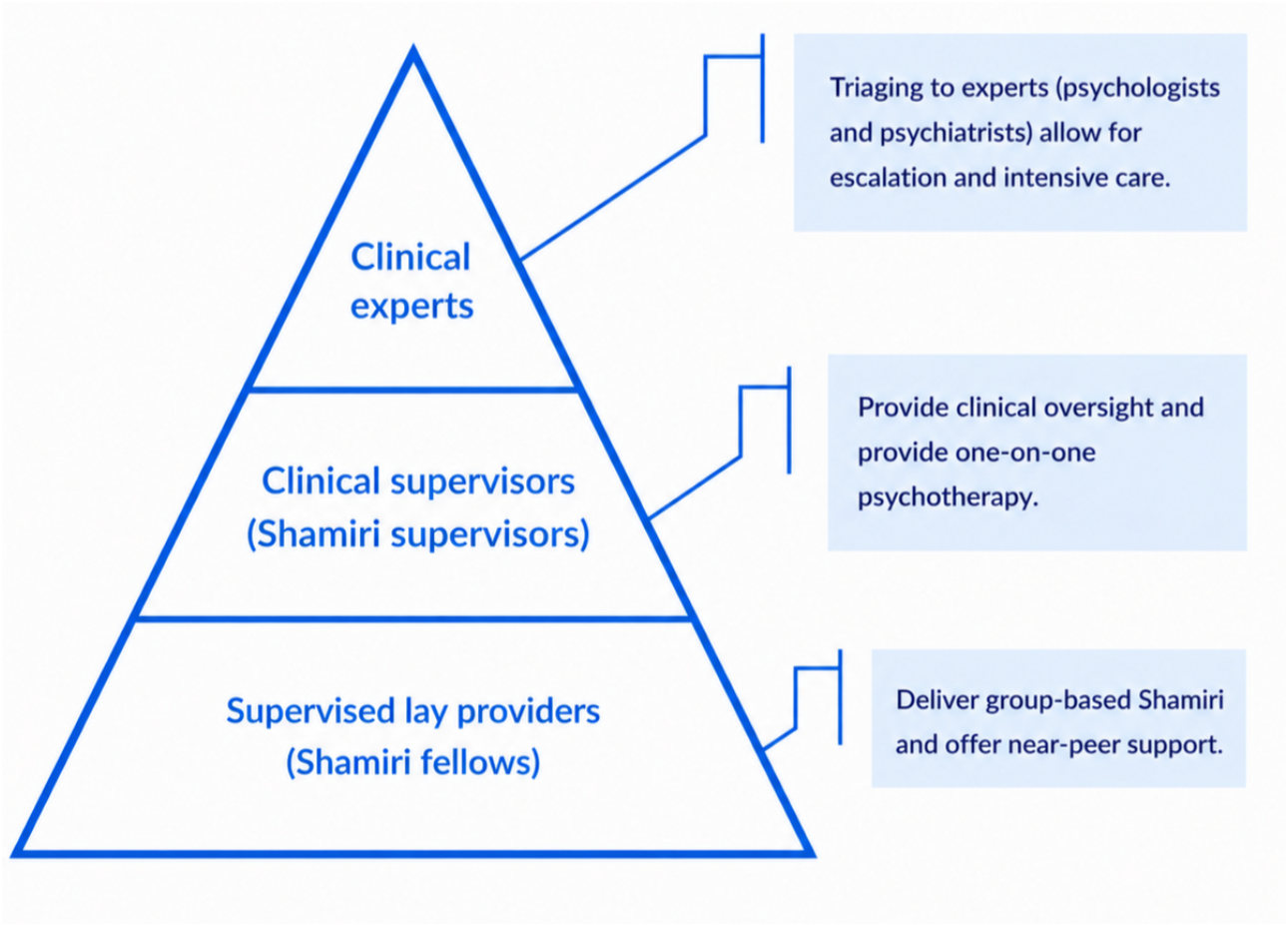
Youth mental health challenges are among the most pressing global health issues today [1,2]. Half of all lifetime mental disorders begin by age 14 [3,4]. Most go untreated—particularly in low- and middle-income countries (LMICs), where shortages of trained clinicians and fragile health systems produce a significant treatment gap [1,2,5,6]. Closing this gap requires care models that are cost-effective at scale, contextually grounded, and delivered beyond traditional clinical settings into community spaces [1,7,8].

Stepped Care Models Can Close the Treatment Gap

Task-shifting—training nonprofessionals to deliver structured, evidence-based treatments under supervision—has emerged as a promising strategy for closing the treatment gap [9,10]. When embedded within stepped-care systems, where the intensity of care is matched to individual need [9-12], this approach maximizes scarce resources, reduces costs, and leverages community infrastructure for delivery [12,13].

The Shamiri model is an example of task-shifting within a stepped-care system: a 3-tier school-based mental health program delivered by lay providers and clinical staff (Figure 1) [14,15]. In the first tier, lay providers aged 18-22 years are trained and supervised to deliver brief, group-based psychological interventions—typically weekly sessions of 4 weeks with groups of 6 to 15 students [16,17]. Clinical supervisors with backgrounds in psychology or clinical social work form the second tier, providing oversight of lay providers and direct one-on-one psychotherapy where needed [16,18,19]. Severe or complex cases are escalated to clinical psychologists and psychiatrists in the third tier [16,18,19]. Randomized controlled trials demonstrate that the Shamiri model significantly reduces depression and anxiety symptoms in secondary school students, with effects sustained at least 7 months after the intervention [14,15,20,21]. Cost-effectiveness analyses place delivery at approximately US $7.48 per student in 2024—among the most affordable youth mental health interventions globally [22,23].

Figure 1. The Shamiri model: a 3-tier stepped care model for youth mental health.



Despite these outcomes, scaling innovations like Shamiri while maintaining fidelity and efficacy across diverse contexts remains a challenge. Two main dissemination strategies have been used [24,25]. The centralized approach, where the originating

XSL•FO
RenderX

organization retains direct control, allows for tighter quality control but limits reach [23]. The decentralized—or "train-the-trainer"—approach, where other organizations implement independently, enables rapid scale-up but carries risks to fidelity and consistency [24,25]. Recent research comparing both approaches in Kenya found both effective at improving youth outcomes, though decentralized models require more robust support structures and greater funding to maintain effectiveness and fidelity [23].

As implementation moves across contexts and implementers, additional operational challenges arise—particularly for small or resource-constrained community-based organizations [23,26,27]. Coordinating multitiered teams, managing logistics and financial workflows, tracking real-time outcomes, and maintaining consistent training and supervision all pose barriers to broader uptake of otherwise effective interventions [23,26,27].

Technology as a Catalyst for Scale

Digital platforms can address these bottlenecks. Purpose-built technology can reduce administrative burden, standardize supervision workflows, enable real-time monitoring, and facilitate client transitions across care tiers [13,28,29]. For task-shifted stepped-care models specifically, digital tools can provide lay providers with structured session prompts, automate operational workflows, support triage, and coordinate across roles [13,28-31].

Most existing digital mental health tools fall short of what community-based delivery in Kenya—and similar contexts—requires [13,30-32]. Available platforms are typically designed for high-resource clinical environments, carry prohibitive licensing costs, lack customizability for task-shifted workflows, and offer little or no offline functionality [29,33,34]. Multiple recent reviews confirm this infrastructure gap: digital mental health interventions developed for LMICs remain limited in number, and those that exist rarely include the operational architecture needed for stepped-care delivery, provider supervision, or scalable implementation [29,35-37]. Platforms built to coordinate lay-provider delivery, integrate operational and clinical workflows, and function in bandwidth-constrained settings are largely absent from the literature [29,37]. This gap is not unique to Kenya: across sub-Saharan Africa, South Asia, and Southeast Asia—settings characterized by severe mental health workforce shortages and limited digital infrastructure—the distance between the promise of digital health tools and the availability of platforms suited for community-based delivery remains wide [1,38].

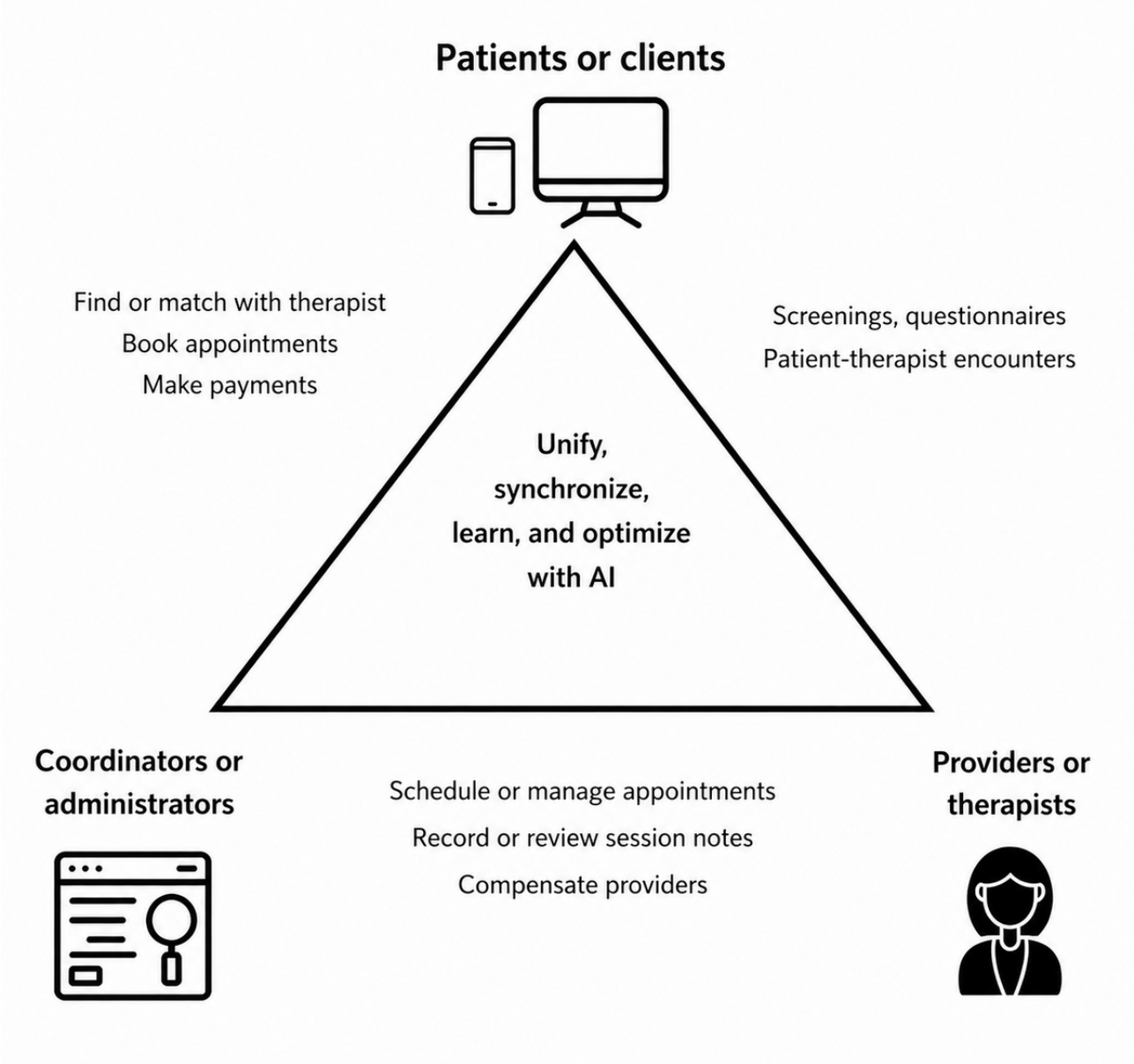
Development of shamiriOS: A 3-Product Platform for Delivery at Scale

To address this gap, we developed shamiriOS—an open-source digital platform built to support the scalable delivery of task-shifted youth mental health services in Kenya and similar settings. The name reflects its purpose: an integrated operating system for service delivery, not a conventional computer operating system. The platform comprises 3 interoperable product suites—the Shamiri Digital Hub (SDH), the Shamiri Provider Platform (SPP), and Rafi—each designed for a distinct user group within a task-shifted care approach like the Shamiri model (Figure 2):

- SDH is an operational command center for program management, coordination, supervision oversight, and financial workflows.
- SPP is a clinical workflow tool supporting fellows, supervisors, and therapists in delivering and documenting care.
- Rafi is a client-facing mobile app through which young people access care, navigate support pathways, and engage with mental health resources. Available on Android and iOS, Rafi is built with an offline-first architecture to remain functional in the low-bandwidth environments common across Kenyan schools and universities.

XSL•FO
RenderX

Figure 2. shamiriOS overview. AI: artificial intelligence.



These 3 products function as a unified delivery ecosystem: SDH runs the program, SPP supports the providers, and Rafi connects young people to care. Each component can be deployed independently or in combination, depending on an implementing organization's context and capacity. Additionally, each component is "modular" in design: its features can be configured and adapted to meet the specific needs of different implementing partners, contexts, and end users.

Although the broader vision for shamiriOS encompasses all 3 components, this paper focuses primarily on the development and deployment of SDH and Rafi—built from the ground up—and on the initial integration of SPP, adapted from an existing clinical platform. shamiriOS builds on the capabilities and limitations of existing platforms, including case management systems such as CommCare and OpenMRS (Open Medical Record System), and youth-facing tools reviewed in our environmental scan [29,35-37]. Its distinctive contribution is an integrated, 3-product architecture combining operational management, clinical workflows, and client-facing engagement within a single open-source, context-specific system—a combination not found in existing platforms in Kenya or comparable settings. Successful adaptation for other countries is feasible with appropriate participatory redesign; the findings reported here are specific to the Kenyan context in which the platform was developed and validated.

Study Aims

This paper has 3 aims. First, we present an environmental scan of existing digital platforms to characterize their limitations for stepped-care youth mental health delivery in Kenya. Second, we describe the iterative, user-centered design and development process of shamiriOS. Third, we evaluate early outcomes from its deployment across centralized and decentralized implementation settings in Kenya between 2023 and 2024, reporting on adoption, user satisfaction, and feature refinement.

Together, these 3 elements—environmental scan, platform development, and early deployment—establish a foundational evidence layer for the ongoing development, evaluation, and

XSL•FO
RenderX

contextual adaptation of integrated digital mental health platforms in settings where such infrastructure is critically needed.

Methods

Overview

This section describes the review, development, and deployment processes underlying shamiriOS. We first describe the environmental scan of existing platforms that informed our decision to build a custom solution. We then outline the iterative development of each component—the SDH and Rafi, built from the ground up, and the SPP, adapted from an existing clinical system. We describe technical access specifications, the engagement measurement framework used to evaluate deployment outcomes, and the phased deployment of the platform across centralized and decentralized settings in Kenya between 2023 and 2024.

Environmental Scan of Existing Platforms

Overview and Evaluation Criteria

Prior to development, we conducted an environmental scan to assess whether existing digital platforms could support the operational, clinical, and engagement requirements of the Shamiri stepped-care model in Kenya. The scan assessed market availability, feature coverage, and contextual fit for community-based, task-shifted delivery—it was not a systematic review of clinical efficacy. We evaluated potential tools across 3 envisioned system functions: a backend operations platform (SDH), a client-facing engagement tool (Rafi), and a clinical provider platform (SPP).

Platforms were evaluated against 6 criteria tailored to the Kenyan implementation context: cost (upfront and ongoing), ease of use for administrators and end users, customizability for task-shifted workflows, open-source availability, suitability for mental health interventions, and appropriateness for low-resource settings including offline capability. These criteria were determined through prior field experience and consultation with Shamiri's operations and clinical teams (Table 1).

Table 1. Platform evaluation criteria and justification.

Criterion	Rationale for LMIC[a] or Kenyan context
Cost (upfront and ongoing)	Must be affordable for grassroots and mid-sized organizations; includes licensing, hosting, and support costs.
Ease of use	Must be usable by lay providers, supervisors, and adolescent clients with varying levels of digital literacy.
Customizability	Must accommodate task-shifting workflows, school-based implementation, and stepped-care triage.
Open-source availability	Ensures accessibility, adaptability, and long-term sustainability without vendor lock-in.
Suitability for mental health	Requires support for psychosocial content, triage workflows, and group or individual care formats.
Low-resource appropriateness	Must support offline use, low bandwidth environments, and options for local hosting.

[a]LMIC: low- and middle-income country.

Backend Operations Platforms

Between May and July 2023, we conducted a desktop scan of global case management platforms through public documentation, product demonstrations, and technical repositories. Six platforms were reviewed: Salesforce, Dimagi CommCare, KoboToolbox, Ona, ODK Collect, and Google Workspace. Platforms were assessed for their ability to support real-time multisite coordination, stipend disbursement tracking, session attendance logging, triage escalation, and role-differentiated access across lay providers, supervisors, and coordinators. Search methods and assessment criteria are described in Multimedia Appendix 1.

Client-Facing Youth Mental Health Apps

Searches were conducted on the Google Play Store and Apple App Store using terms including mental health, therapy, depression, and youth wellbeing. Consistent with prior research methodology, we examined the top 10 results for each term [39]. Eligible apps had a minimum of 100,000 downloads (Android) or a 4.0+ star rating (iOS), were available in English and accessible in East Africa, and had been updated within the prior 12 months. Apps focused solely on physical wellness or designed exclusively for clinicians were excluded. A total of 28 apps met eligibility criteria and were assessed across cost structure, provider matching and support features, validated assessment tools, delivery model, self-help resources, and user engagement mechanisms including gamification. Reviewed apps were disaggregated by whether they claim an explicit evidence base (eg, cognitive behavioral therapy and interpersonal therapy) versus general wellness positioning to enable a more precise assessment of their suitability for stepped-care delivery. Full details are in Multimedia Appendix 1.

Clinical Provider Platforms

Existing electronic medical record (EMR) and clinical workflow systems were reviewed for suitability as a provider-facing platform, including District Health Information System 2 (DHIS2), OpenMRS, Bahmni, and EasyClinic. Systems were assessed for clinical documentation capability, teletherapy support, appointment scheduling, analytics dashboards, data security, and ease of customization for a task-shifted delivery context.

XSL•FO
RenderX

Development of the shamiriOS Platform

Overview and Design Approach

The environmental scan confirmed that no existing solution fully met the operational, clinical, and youth engagement needs of the Shamiri model in Kenya. We initiated development of shamiriOS in mid-2023, conceptualizing it as a 3-product digital ecosystem: the SDH, Rafi, and the SPP. Initial development focused on building SDH and Rafi from the ground up while adapting SPP from an existing solution, EasyClinic (Table 2).

Table 2. Development of shamiriOS components.

Component and version	Period	Features	Feedback
Shamiri Digital Hub			
Version 1	Late 2022 to early 2023	Session attendance tracking, fellow performance monitoring, supervisory reviews	Improved mobile responsiveness, integration of payment and stipend tracking tools, support for multisite coordination
Version 2	Early to late 2023	Supervisor and coordinator dashboards, stipend reconciliation workflows, referral tracking, automated task reminders	Iterative feedback via field interviews and monthly user experience reviews
Version 3	Late 2023 to present	Multirole platform, multisite implementation management, Metabase-powered analytics dashboards, offline functionality	Development continues, phased rollouts planned for 2025 and beyond
Rafi			
Version 1.0	May 2022	Daily motivational content, well-being check-ins, therapy and group session booking, monthly score tracking, gamified system	Need for more frequent interaction, personalization, engaging visuals
Version 2.0	Early 2023	Daily mood check-ins, interactive reward system, artificial intelligence–generated mood art, triage pathway, community feed	Strong acceptability and usability, need for more modular content, deeper personalization
Version 3.0	Early 2024	Expanded self-help toolkit, upgraded triage flows, redesigned interface, anonymous messaging, session scheduling, offline functionality	Feasibility of integrating technology-driven triage and engagement tools into institutional frameworks
Shamiri Provider Platform			
Version 1.0	2024 to present	Client management, teletherapy support, session documentation, analytics dashboards, syncing user records, real-time therapist availability, appointment scheduling	Emphasis on data security, user-centered design, interoperability

The development of shamiriOS was guided by a user-centered design process with community consultation, adapted to the operational demands of building integrated health technology infrastructure in Kenya [40,41]. There is no single template for co-design and user-centered design; rather, the methods must be tailored to the nature of what is being designed, the stakeholders involved, and the constraints of the implementation environment [42-45]. The shared philosophy—centering the needs, experiences, and contexts of end users—remained constant across components; the specific methods varied by component type.

For shamiriOS, we distinguish 2 types of design work. The operational infrastructure components—SDH and SPP—were designed primarily as administrative and clinical workflow tools, and their development was guided by consultations with Shamiri's operations, clinical, and research teams. Rafi, the youth-facing component, was developed through a more participatory co-design process: youth end users shaped interface design, content tone, and feature prioritization through structured feedback sessions, card-sorting exercises, and prototype walkthroughs. Cocreation sessions for the initial version involved 77 students across 3 Kenyan universities (Multimedia Appendix 2).

Across all components, end users—lay providers, clinical supervisors, hub coordinators, and youth participants—acted as consultants and testers, informing design through structured feedback, usability tests, and pain-point mapping. Strategic decisions regarding platform architecture, feature prioritization, and technical implementation remained with Shamiri's technical and clinical leadership. This approach sits at the consultation end of the community engagement spectrum rather than representing fully community-led development; its implications and our future commitments are discussed in the Discussion section.

Development of the SDH

Overview

The SDH was designed to replace fragmented operational tools—spreadsheets, WhatsApp messaging, and paper logs—that were prone to error and coordination breakdowns as the number of supervisors, fellows, and implementation sites grew. The goal was a centralized platform enabling real-time monitoring, stipend disbursement, and data-driven supervision across multisite teams.

SDH development occurred across 3 cycles between 2022 and 2024, each guided by user consultation with Shamiri's

XSL•FO
RenderX

operations, clinical, and research teams. Each development cycle followed a hybrid Agile-Waterfall methodology. Agile development is an iterative approach in which work is broken into short, time-boxed cycles called sprints—typically 2 to 4 weeks—with working software produced at the end of each cycle based on user feedback [46]. Waterfall is a sequential approach in which requirements are fully scoped before development begins [47]. The hybrid model we used scoped core feature requirements in advance (Waterfall) but delivered them through sprint-based cycles with regular stakeholder check-ins and usability testing (Agile), allowing responsiveness to evolving user needs while maintaining predictable development milestones [46,47]. This approach maintained predictable milestones while remaining responsive to evolving user needs.

Version 1: Prototype Development (Late 2022 to Early 2023)

Version 1 was built on Bubble, a no-code development platform, and served as a minimum viable product to address the most critical operational needs: session attendance tracking, fellow oversight, and performance review. It was deployed in a small usability pilot across 3 implementation hubs and 20 field staff to assess basic usability, workflow alignment, and feasibility of replacing paper-based tracking. User feedback identified 3 priorities for subsequent development: improved mobile responsiveness, integration of payment and stipend tracking, and support for multisite coordination.

Version 2: Supervisor-Centered Platform Expansion (Early to Late 2023)

Version 2 rebuilt SDH on a scalable web app architecture: a Node.js and Express backend, a Next.js frontend, and a PostgreSQL database, with components communicating via RESTful (Representational State Transfer) application programming interfaces (APIs), JSON Web Token–based authentication, and role-based access control (RBAC) at the API layer. Full technical specifications, including the data architecture and security model, are in Multimedia Appendix 3.

Version 2 introduced supervisor and coordinator dashboards, stipend reconciliation workflows, referral tracking, and automated task reminders. It was deployed across 5 hubs, with iterative feedback collected via field interviews and monthly user experience reviews.

Version 3: Multirole Platform and Offline Functionality (Late 2023 to Present)

Version 3 began development in late 2023 to accommodate scaling demands and a more diverse user base. The platform was expanded to include fellows, supervisors, coordinators, and clinical leads; introduced multisite implementation management; and integrated Metabase-powered analytics dashboards. Metabase is an open-source business intelligence tool that enables nontechnical users to query databases and generate visual reports without writing code [48,49]; its integration allows hub coordinators and supervisors to access real-time program data without developer support. Version 3 also introduced offline-first data capture with automatic synchronization upon network reconnection, supporting use in the low-connectivity environments common across Kenyan schools and community centers. Local server deployment options are under active development.

Development of Rafi

Overview

Rafi (derived from rafiki, Kiswahili for “friend”) is a client-facing native mobile app for young people—primarily university students aged 17 to 22 years—supporting mental health screenings, patient-therapist encounters, and community engagement within Shamiri’s stepped-care model. Rafi is available on Android (Google Play Store) and iOS (Apple App Store) and was designed with an offline-first architecture: core features—including self-guided content (affirmations, journaling, and goal tracking), mood check-ins, and well-being score tracking—remain functional without internet access. Features requiring connectivity—therapy booking, artificial intelligence (AI)–generated mood artwork, community feed posts, and data synchronization—queue locally and sync automatically on reconnection. This design decision was informed by the environmental scan finding that only 5 of 28 reviewed youth-facing apps offered any offline capability, despite the intermittent-connectivity environments common across Kenyan universities and schools.

Conception and Design

Rafi’s development followed the user-centered design process described earlier. Youth end users were engaged as testers and consultants across each version through structured feedback sessions, card-sorting exercises, prototype walkthroughs, and focus group discussions. Cocreation sessions for the initial version involved 77 students across 3 universities (Multimedia Appendix 2). Architectural and design decisions were made by the Shamiri technical team.

Rafi 1.0: Initial Prototype

Version 1.0 was developed as a mobile app available on Android, iOS, and a supplementary web portal to maximize accessibility across Kenya’s variable connectivity landscape. Core features included daily and weekly mental health check-ins, access to the Shamiri-Digital self-guided intervention, low-cost caregiver connections, clinical referral pathways, and gamification elements (points and rewards). Version 1.0 was piloted at 7 Kenyan universities through partnerships with peer counseling groups (Multimedia Appendix 2).

Rafi 2.0: Stable Platform

Following pilot feedback, version 2.0 introduced mood-based AI-generated artwork (the “wellness gallery”), enhanced triage tools, and expanded the gamification system. It was cocreated with 77 youth users through structured feedback sessions and deployed across select student clubs at additional universities using QR code–based registration and peer ambassador–led outreach.

Rafi 3.0: Integrated Platform

Version 3.0 was designed to address persistent retention challenges from earlier deployments. Key design priorities included gamified onboarding sequences, mood-responsive motivational content, and more direct pathways connecting

XSL•FO
RenderX

users to therapists. The interface was restructured to reduce feature overload, and the backend was refactored for tighter SPP integration, supporting seamless transitions from self-guided to provider-supported care. Version 3.0 was the platform used in the Mount Kenya University (MKU) institutional deployment described below.

Development of the SPP

The SPP was adapted from EasyClinic, an existing EMR solution selected for its modular design, customizable clinical workflows, and prior use in resource-constrained health care settings. Rather than building a clinical platform from the ground up, we adapted an existing, maintained solution to reduce development cost and time to deployment. Development followed an iterative sprint-based approach incorporating therapist feedback, usability testing, and emphasis on data security and interoperability within the shamiriOS architecture.

Adaptations include (1) migration of legacy patient records from Rafi and SDH to EasyClinic to establish unified clinical histories; (2) integration of real-time therapist availability and appointment scheduling into Rafi's booking interface, synchronized with the EasyClinic backend; and (3) subscription and plan management enabling users to initiate and manage therapy plans through Rafi with real-time synchronization to EasyClinic. Full technical specifications are in Multimedia Appendix 4.

Technical Access and Compatibility

All shamiriOS components are web-based or app-based and require no operating system modification, installation privileges, or proprietary software. This was a deliberate design choice to maximize accessibility for nontechnical users and reduce dependency on institutional IT infrastructure across diverse Kenyan implementation settings.

SDH and SPP are web apps accessible via any modern browser on desktop, laptop, or mobile devices. Both incorporate offline-first data capture with automatic synchronization on reconnection. Rafi is a native mobile app available on Android and iOS; core self-help content and check-in features are accessible offline, while therapy booking, triage escalation, and data synchronization require an active internet connection. A supplementary web portal provides browser-based access for users without smartphone access, synced with app data. Table 3 summarizes the minimum technical requirements and offline capabilities for each component.

Table 3. Technical access requirements for shamiriOS components[a].

Component	Access method	Supported browsers or platforms	Device or OS[b] compatibility	Offline capability
Shamiri Digital Hub	Web app	Chrome 90+, Firefox 88+, Safari 14+, Edge 90+	Desktop, laptop, or mobile—Windows, macOS, Linux, Android, iOS	Partial: offline data capture with automatic sync on reconnection
Rafi	Native mobile app+supplementary web portal	Google Play (Android), App Store (iOS), modern browser (web portal)	Android 8.0+, iOS 12+, web portal via modern browser	Limited: self-help content and check-ins accessible offline; therapy booking, triage, and sync require connectivity
Shamiri Provider Platform	Web app (adapted from EasyClinic)	Chrome 90+, Firefox 88+, Safari 14+, Edge 90+	Desktop, laptop, or mobile—Windows, macOS, Linux, Android, iOS	Not available in current version

[a]All data transmission is encrypted via HTTPS. Authentication uses Open Authorization 2.0 protocols across all components. Full technical specifications are in Multimedia Appendices 2-4.

[b]OS: operating system.

Engagement Measurement Framework

Engagement with digital health platforms is multifaceted and context-dependent. We adapted a framework defining engagement as the extent to which users interact with each shamiriOS component in ways intended by the design and required for operational or clinical benefit [50]. We distinguish four levels:

1. Initial engagement: account creation, onboarding completion, and first active use.
2. Ongoing engagement: frequency, duration, and consistency of continued use over time.
3. Deep engagement: use of advanced or therapeutically substantive features requiring sustained investment (eg, journaling, clinical triage documentation, and supervisor analytics review).
4. Clinical or operational engagement: actions directly tied to service delivery outcomes (eg, therapy session booking, weekly session attendance logging, and triage escalation).

Engagement is reported separately for each component, reflecting their distinct purposes and user populations:

- SDH (operational or staff-facing): Measured through platform login frequency, data entry compliance rates, feature adoption across user roles, and staff-reported usability and acceptability (single-item satisfaction rating and Net Promoter Score [NPS]). Weekly engagement aligned with supervision and reporting cycles constitutes the target pattern for SDH users.
- Rafi (youth-facing): Measured through registration and onboarding completion (initial engagement), daily and monthly active user (MAU) rates (ongoing engagement), self-guided feature use such as journaling and mood

XSL•FO
RenderX

check-ins (deep engagement), and therapy booking and triage escalation (clinical engagement). Episodic or needs-driven engagement may be more clinically appropriate than sustained daily use for this population; we report both frequency-based and action-based metrics.

What constitutes meaningful engagement differs by user type and platform purpose. For youths using Rafi, therapy booking and triage pathway use may be more clinically significant than daily login frequency. For supervisors using SDH, weekly engagement aligned with supervision cycles was operationally sufficient. For clinicians using SPP, session-by-session engagement was anticipated. Engagement data were drawn from backend use logs, session analytics via Metabase, and NPS surveys administered after deployment [51].

Implementation Evaluation

Our evaluation used pragmatic metrics derived from implementation science instruments [50-53]: measures used were (1) usability—a single-item satisfaction rating on a 1-10 scale ("How easy is the platform to use?"), administered to supervisors and hub coordinators after deployment; (2) acceptability—a NPS item on a 1-10 scale ("How likely are you to recommend this platform to a colleague?") [51]; (3) engagement—backend use analytics tracking login frequency, feature adoption rates, data entry completeness, and retention over time, reported separately by component; and (4) qualitative feedback—open-ended survey questions and semistructured interviews conducted during and after training sessions.

The choice of pragmatic over standardized measures reflected several constraints. Sprint-based development cycles required rapid, actionable feedback that standard implementation science instruments—which typically require extensive administration time and psychometric analysis—could not reliably provide. The simultaneous development and evaluation of the platform also blurred the boundary between formative user research and summative implementation evaluation, making it difficult to administer standardized instruments at appropriate intervals. Field teams requested brief feedback mechanisms to reduce survey fatigue during intensive deployment.

Deployment of the shamiriOS Platform

Overview

Deployment of SDH and Rafi occurred in phased waves between 2023 and 2024, expanding the user base, geographic coverage, and implementation complexity across waves. SPP deployment is ongoing and is not reported here; a follow-on publication will report SPP deployment and clinical workflow integration separately. Table 4 provides a deployment overview.

Table 4. Development of shamiriOS components (Rafi and Shamiri Digital Hub).

Deployment and wave	Details	Time frame	Sites	Size
Shamiri Digital Hub				
Wave 1	Initial deployment in centralized school-based hubs	Q4 2023	5 school-based hubs in Nairobi and Kiambu	358 fellows
Wave 2	Expansion to decentralized partner sites and diverse implementation contexts	Early 2024	6 additional sites	1312 implementation team members
Rafi				
Rafi 1.0	Limited university pilot with peer counseling groups	2022	7 universities in Kenya	N/A[a]
Rafi 2.0	Soft-launched across select student clubs	2023	Other universities and colleges	N/A
Rafi 3.0	Formal deployment at scale through partnership with Mount Kenya University	Early 2024	Mount Kenya University	5000 students

[a]N/A: not applicable.

Deployment of the SDH

Wave 1: Centralized School-Based Hubs (Q4 2023)

The first SDH deployment targeted 5 school-based hubs in Nairobi and Kiambu counties, selected for high program activity, digital readiness, and established supervisory infrastructure. The goal was to transition day-to-day operations—previously managed through spreadsheets, WhatsApp, and manual logs—to the SDH platform. Training was delivered to 40 supervisors, 6 hub coordinators, and approximately 358 fellows through remote onboarding modules, in-person workshops, and helpdesk support. Each user received a role-based account scoped to their operational responsibilities. Adoption was monitored through backend use metrics and periodic supervisor check-ins, with real-time feedback informing feature refinements.

Wave 2: Decentralized Partner Sites (Early 2024)

In early 2024, deployment expanded to 6 additional sites including periurban schools, community centers, and regional nongovernmental organizations operating as local implementation partners. The platform was configured to support site-level customization including localized workflows. Asynchronous training modules, role-specific onboarding videos, and quick-start guides were developed for the more dispersed user base. Site-level implementation champions facilitated peer-to-peer learning beyond formal training periods. Weekly virtual support check-ins were offered during each new site's first month of onboarding.

XSL•FO
RenderX

Deployment of Rafi

Rafi 1.0 (2022)

Rafi 1.0 was piloted through partnerships with peer counseling groups at 7 Kenyan universities: the University of Nairobi, Kenyatta University, Jomo Kenyatta University of Agriculture and Technology, Technical University of Kenya, Cooperative University, MKU, and Maasai Mara University. Students were recruited through peer counseling clubs and campus outreach events, then onboarded and asked to provide structured feedback on usability, content tone, and visual design.

Rafi 2.0 (2023)

Rafi 2.0 was soft-launched across select student clubs at additional universities and colleges using QR code–based registration and peer ambassador–driven outreach. Students received promotional content and brief onboarding messages introducing Rafi's core features: wellness check-ins, journaling, triage alerts, and therapy booking.

Rafi 3.0—MKU Deployment (Early 2024)

Rafi 3.0 was deployed at scale through a formal partnership with MKU, serving a student population of approximately 25,000, with a targeted deployment cohort of 5000 enrolled students. Deployment included structured digital training for therapists and clinical supervisors covering the provider-facing dashboard and triage flagging protocols; a launch campaign conducted with the university's student services office and student associations, including peer-led classroom outreach, mental health fairs, and "Rafi Club" engagement events; and a dedicated peer ambassador network tasked with driving registrations. In-app analytics tracked early use trends and informed weekly engagement strategies. Backend dashboards allowed provider teams real-time oversight of flagged users and support queue management.

Ethical Considerations

The activities described in this paper comprised platform development, an environmental scan of existing digital tools, operational evaluation of the SDH using routine program monitoring data and satisfaction surveys from adult implementation staff, participatory design consultations with university students during Rafi's development, and aggregated platform use analytics. Per the National Guidelines for Ethical Conduct of Biomedical Research Involving Human Participants in Kenya [54], mandatory ethics review applies to research involving human participants who are "exposed to manipulation, intervention, observation, or other interaction with investigators" or who "become individually identifiable through investigators' collection, preparation or use of biological material, or medical or other records." The activities reported here did not meet these criteria: data collection was operational or design-oriented, no sensitive personal or health data were collected, and the intent was product development rather than systematic generation of new knowledge about individuals. Formal ethics committee review was therefore not sought. All data handling complied with Kenya's Data Protection Act [55] and Shamiri's institutional data governance framework.

Results

Findings From the Environmental Scan

Backend Operations Platforms

None of the 6 platforms reviewed met the combined requirements for cost, offline capability, and customizability for task-shifted delivery in Kenya. Salesforce and comparable enterprise systems offered extensive functionality but at licensing costs prohibitive for grassroots organizations. Dimagi CommCare and KoboToolbox offered greater affordability and open-source availability but lacked role-differentiated access, stipend management workflows, and real-time multisite coordination. No platform met all 6 evaluation criteria simultaneously; the highest-scoring alternatives met a maximum of 4 of 6. Full results are in Table 5.

Table 5. Review of existing case management and health systems.

Platform	Cost	Ease of use	Customizability	Open source	Ideal use case
Salesforce Health Cloud	High	Moderate	High	No	Care coordination, shared records, team collaboration
Dimagi CommCare	Medium	High	Moderate	No	Mobile-first field data collection
OpenMRS[a]	Low	Moderate	High	Yes	Customizable EMRs[b] for low-resource clinical settings
Odoo	Medium	High	High	Yes	Operations and resource management
Ona OpenSRP	Low	Moderate	High	Yes	Last-mile community health, offline syncing
Microsoft Dynamics	High	Moderate	High	No	Enterprise health care and operations

[a]OpenMRS: Open Medical Record System.

[b]EMR: electronic medical record.

Youth-Facing Digital Mental Health Apps

Of the 28 apps reviewed, 12 were explicitly evidence-based (citing cognitive behavioral therapy, interpersonal therapy, or other clinical frameworks), and 16 were general wellness apps. No app met all evaluation criteria. Common limitations included subscription-based pricing inaccessible to Kenyan university students, the absence of validated clinical content in local languages, no triage or escalation pathway to human providers, and limited or no offline functionality. Only 3 of the 28 apps

XSL•FO
RenderX

offered provider-matching features, and none were designed as part of an integrated stepped-care model with backend operational management. Full results are in Table 6, disaggregated by evidence-base status. Further details are in Multimedia Appendix 1.

Table 6. Review of existing mental health apps.

Category or feature	Apps with feature (n=28), n (%)	Evidence-based apps only (n=12), n (%)	General wellness apps only (n=16), n (%)
Evidence-base classification	—[a]	12 (43)	16 (57)
Freemium cost model	23 (82.1)	9 (75)	14 (87.5)
Subscription only	5 (17.9)	3 (25)	2 (12.5)
Provider access			
Therapist access	8 (28.6)	7 (58.3)	1 (6.3)
Peer support forums	11 (39.3)	4 (33.3)	7 (43.8)
AI[b] chatbot	3 (10.7)	1 (8.3)	2 (12.5)
Triage or clinical features			
Symptom checker or triage	7 (25)	6 (50)	1 (6.3)
Provider matching	5 (17.9)	4 (33.3)	1 (6.3)
Self-help resources			
Journaling	9 (32.1)	5 (41.7)	4 (25)
Meditation	13 (46.4)	6 (50)	7 (43.8)
Guided CBT[c] content	10 (35.7)	10 (83.3)	0 (0)
Operational or offline features			
Offline functionality	4 (14.3)	1 (8.3)	3 (18.8)
Lay-provider or tiered supervision workflows	0 (0)	0 (0)	0 (0)
Integrated escalation to in-person care	0 (0)	0 (0)	0 (0)

[a]Not available.

[b]AI: artificial intelligence.

[c]CBT: cognitive behavioral therapy.

Clinical Provider Platforms

Open-source EMR systems including OpenMRS and DHIS2 offered strong clinical documentation features but required substantial technical capacity for implementation and maintenance beyond the reach of most Shamiri implementation partners. EasyClinic was identified as the strongest candidate: it offered client management, teletherapy support, session documentation, appointment scheduling, and analytics dashboards at an accessible cost, with a modular architecture amenable to customization. These findings supported the adaptation of EasyClinic rather than the de novo development of the SPP.

Current State of shamiriOS

Overview

As of early 2025, shamiriOS comprises three interlinked product suites: (1) the SDH, the operational backbone for program management and real-time monitoring; (2) Rafi, a native mobile app (Android and iOS) through which young people access care and mental health resources; and (3) the SPP, an EMR and clinical workflow tool adapted from EasyClinic. SDH and Rafi are fully deployed across multiple centralized and decentralized sites; the SPP is currently in integration testing. Each suite is modular: features can be configured to meet the specific needs of different implementing partners and contexts.

Current State of the SDH

Roles and Permissions

SDH operates through a role-based permission system aligned with Shamiri's delivery hierarchy. Five user roles are defined, each scoped to their operational function (Table 7):

- Program administrators hold platform-wide permissions covering multisite oversight, intervention fidelity reporting, stipend management, and global configuration.
- Hub coordinators manage specific hubs, with permissions to assign fellows and supervisors, validate session attendance records, approve stipend submissions, and escalate triage cases.
- Supervisors verify weekly fellow activities, document supervision sessions, and initiate triage referrals for at-risk youths.

XSL•FO
RenderX

- Fellows (lay providers) log session data, track student attendance, and report implementation activities within their assigned hubs. Permissions are currently limited to data entry and personal activity tracking; future enhancements will expand access to include performance dashboards.
- Clinical leads access escalated clinical cases and review triage documentation across all sites.

Table 7. Roles and permissions in Shamiri Digital Hub.

Users	Roles and permissions
Program administrators	Possess platform-wide permissions, enabling oversight of multisite activities, intervention fidelity reporting, stipend management, and global settings configuration.
Hub coordinators	Manage specific hubs, with permissions to assign fellows and supervisors, validate session attendance records, approve stipend submissions, and escalate triage cases.
Supervisors	Responsible for the verification of weekly fellow activities, documentation of supervision sessions, and initiation of triage referrals for at-risk youths.
Fellows (lay providers)	Access the platform directly to log session data, track student attendance, and report key implementation activities. Their permissions are limited to data entry and personal activity tracking within assigned hubs. Future system enhancements may expand their access to include feedback tools and performance dashboards to support self-monitoring and reporting efficiency.
Clinical leads	Access escalated clinical cases and review triage documentation across all implementation sites.

Each user interaction is timestamped and logged, supporting full auditability and quality assurance monitoring across the implementation pipeline.

Software Architecture

SDH is built on a scalable, microservice-inspired architecture. The frontend is developed in ReactJS and deployed via Vercel's content delivery network, optimized for performance across variable connectivity environments on desktop and mobile browsers. The backend uses a Node.js runtime interfacing with a PostgreSQL database hosted on Amazon Web Services Relational Database Service; system interactions are managed through a RESTful API framework using JSON Web Tokens for session authentication and encrypted communications. A Metabase analytics layer provides real-time dashboards on session attendance, intervention frequency, supervision compliance, and flagged clinical concerns. Offline-first functionality enables data capture in low-connectivity settings with automatic synchronization on reconnection.

Features

The SDH system provides the following features in support of the Shamiri operational model:

- Staff and school management: Dynamic assignment of fellows to school-based groups, tracking intervention session coverage, and historical site performance monitoring.
- Session attendance and tracking: Supervisors enter weekly attendance, attach school visit summaries, and receive automated reminders to ensure data completeness.
- Performance monitoring: Visual analytics dashboards display fellow engagement metrics, session quality ratings, and supervisor oversight reports.
- Financial management: Integrated workflows for stipend submission, validation, and approval through multitiered administrative review pathways.
- Triage and clinical escalation: Supervisors flag students requiring advanced support, triggering structured triage documentation and clinical review workflows.
- Audit trail and compliance monitoring: Comprehensive backend logs maintain records of user interactions, system edits, and data submissions for quality assurance and troubleshooting.

Current State of Rafi

Roles and Permissions

Rafi's role and permission system balances user privacy and personalization with clinical oversight. Three roles are defined: (1) users (youths) create and manage personal accounts, complete check-ins, access self-help content, and manage therapy bookings and sharing preferences—all personally identifiable data are encrypted and access-controlled; (2) therapists and clinical staff access a provider dashboard displaying triaged users and appointment requests and may view anonymized user content when the user grants permission; and (3) administrators manage content deployment, update activity flows, and monitor platform use metrics.

Software Architecture

Rafi is delivered as a native mobile app (Android and iOS) with a supplementary web portal for users without smartphone access. The frontend is built with the Remix framework, hosted on Vercel, with a responsive design for mobile browsers. The backend uses a hybrid Node.js and Python architecture supporting modular feature development—including AI-generated affirmations and journaling workflows. Structured data (user check-ins and activity logs) are stored in PostgreSQL; media files (eg, journal artwork) are stored in Digital Ocean object storage. Integrations include Google Calendar API for therapist scheduling, Firebase Cloud Messaging for push notifications, and AI services for mood-based artwork generation and personalized reflection prompts. All backend services use HTTPS encryption and Open Authorization 2.0 protocols.

Features

Rafi's feature set is designed to engage youths in a nonclinical but supportive way, supporting well-being monitoring, self-help access, and connection to care:

XSL•FO
RenderX

- Wellness check-ins: Daily and weekly self-ratings across 5 domains (social connection, motivation, academic stress, and others) generating a personalized wellness score and surfacing tailored content.
- AI mood art (Wellness Gallery): Weekly check-in responses are transformed into AI-generated artwork that users can share on the community feed.
- Therapy booking: Users directly schedule individual or group sessions through integrated booking modules connected to therapist availability in the SPP.
- Triage pathways: Algorithmic triggers escalate users to therapist connection workflows based on risk indicators detected during check-ins.
- Community feed: A moderated, anonymous social space where users can share mood art and connect with peers, monitored by peer counselors and the administrator team.
- Reward hub: Gamification elements including badges, streak tracking, and unlockable achievements to promote continued engagement.
- Web portal: A lightweight browser-accessible version providing basic check-in and booking functionality, synced with app data.

Core features—self-guided content (affirmations, journaling, and goal tracking), daily mood check-ins, and well-being score tracking—remain fully functional without internet connectivity. Features requiring server connectivity—therapy booking, AI-generated mood artwork, community feed posts, and data synchronization—queue locally and sync automatically on reconnection.

Current State of the SPP

The SPP was adapted from EasyClinic and is currently in integration testing. Formal deployment results are not reported here, as the SPP reached production readiness after the observation period of this study. Adaptations completed to date include migration of legacy patient records from Rafi and SDH to EasyClinic to establish unified clinical histories; integration of real-time therapist availability into Rafi's booking interface; and subscription plan management enabling users to manage therapy plans through Rafi with synchronization to EasyClinic. The SPP is actively used in Shamiri's clinic-based services. A dedicated follow-on publication will report SPP deployment, clinical workflow integration, and cross-component data flows.

Engagement With shamiriOS Components

Engagement With the SDH

Wave 1 Outcomes (Q4 2023)

By the end of 2023, SDH supported program management across all 5 target hubs, coordinating 358 fellows serving over 20,697 students. Platform acceptability was assessed via postdeployment NPS survey administered to supervisors and hub coordinators (n=22). Respondents reported a mean likelihood of recommending SDH of 8.63 (SD 1.46) and a mean user experience satisfaction score of 8.36 (SD 1.49) on 10-point scales. Qualitative feedback emphasized ease of navigation, the value of real-time data capture, and substantial reductions in administrative workload relative to prior spreadsheet and WhatsApp-based systems. Stipend workflow automation and centralized fellow attendance tracking were the most frequently cited valued features (Multimedia Appendix 3).

Wave 2 Outcomes (Q1 2024)

By the end of Q1 2024, SDH supported 1195 fellows, 111 supervisors, and over 76,344 students across all active hubs—representing a 3.3-fold increase in fellows, a 2.8-fold increase in supervisors, and a 3.7-fold increase in student reach relative to wave 1. Ongoing monitoring included automated alerts for low engagement, biweekly platform review meetings with operations staff, and structured feedback channels informing subsequent sprint planning cycles (Multimedia Appendix 3).

Engagement With Rafi

Overview

Rafi engagement is reported across 4 dimensions: initial engagement (registration and onboarding), ongoing engagement (MAU rates), deep engagement (self-guided feature use), and clinical or operational engagement (therapy booking and triage escalation). As noted in the Methods section, episodic or needs-driven engagement may be more clinically meaningful than sustained daily use for a youth mental health platform; we report both frequency-based and action-based metrics.

Rafi 1.0—Initial Pilot (2022)

Of the 1349 students who registered interest, 857 (63.5%) enrolled as active users by the end of 2022. Among enrolled users, 74% (634/857) engaged with the self-guided digital wellness courses—the platform's highest-engagement feature. In total, 37.1% (318/857) booked at least 1 therapy session, and 33% (105/318) of those who booked attended.

Rafi 2.0—Expanded Deployment (2023)

During the 2.0 deployment, 1008 students registered accounts. Therapy booking improved markedly: 50.4% (508/1008) of users booked at least 1 session (compared to 318/857, 37.1% in version 1.0), and 44.9% (228/508) of those attended (compared to 105/318, 33%). Self-guided wellness module engagement remained high at 77.3% (779/1008). Monthly active engagement declined significantly after the first 4 weeks, however—a persistent attrition pattern that directly informed the gamified re-engagement architecture of Rafi 3.0.

Rafi 3.0—Institutional Deployment, MKU (2024)

Initial Engagement

Between January and September 2024, a total of 3737 of the 5000 targeted MKU students registered on Rafi—a 74.7% initial registration rate and the largest single-cohort registration in the platform's deployment history.

Ongoing Engagement and Attrition

Despite strong initial registration, MAU rates declined significantly across the 9-month observation window—from 7.4% in January 2024 to 0% by September—indicating near-complete disengagement among the registered cohort by the end of the observation period. Daily active user rates showed a consistent downward trend from the weeks following launch.

XSL•FO
RenderX

Feature-Specific Engagement

Across the observation period, 3.8% (n=142) of registered users used the daily check-in tool, 1.5% (n=56) used journaling features, 0.5% (n=20) booked therapy sessions, and 0.4% (n=15) used the Ask a Therapist chat feature. Affirmations and other low-touch tools demonstrated limited traction (<0.2%). An NPS survey administered in June 2024 (n=275) found that 15% (n=41) of respondents had attended at least 1 therapy session; among users who engaged with self-guided tools, 99.4% (n=141) subsequently booked therapy appointments.

Implementation Barriers and Facilitators: Preliminary Observations

The following observations were documented through operational feedback channels—training session notes, helpdesk tickets, monthly check-in meetings, NPS survey comments, and sprint planning discussions (Multimedia Appendices 2 and 3).

Barriers

Internet Connectivity

Inconsistent connectivity in periurban schools and community centers caused data synchronization failures and incomplete session logging in SDH. Supervisors at some sites reported traveling to areas with stronger connectivity to submit weekly reports, reducing data timeliness. For Rafi, the connectivity requirement for therapy booking and triage constrained clinical pathway use among students at lower-resourced sites.

Device Constraints

Not all fellows had personal smartphones capable of running modern browser-based apps; device sharing at some hubs created bottlenecks for timely data entry. For Rafi, storage constraints on student smartphones contributed to postregistration uninstalls—limited phone storage was cited in helpdesk and peer ambassador feedback as a contributing factor.

Engagement Infrastructure Readiness

The MKU deployment of Rafi 3.0 launched before its engagement infrastructure was fully operational. Push notification workflows, in-app communication features, and the gamification reward system were not functioning as intended at launch. This prevented timely re-engagement of users experiencing early drop-off and left the team without the real-time use tracking needed to respond to emerging attrition patterns.

Ambassador Incentivization Misalignment

The peer ambassador model at MKU incentivized students financially based on registration counts, producing high initial sign-ups but not the values-based peer promotion needed to drive sustained engagement. Postlaunch operational review found that peer counselors—who combined mental health training with authentic relational credibility—were substantially more effective at converting registrations into active users than financially incentivized ambassadors.

Absence of Institutional Integration

At MKU, the absence of a formal Memorandum of Understanding with university leadership limited institutional buy-in and delayed clinical pathway activation. Caregivers and clinical staff were not fully onboarded at the time of the student launch, meaning therapy booking and triage escalation features were not immediately functional during the highest-engagement window following sign-up.

Facilitators

Alignment With Existing Workflows

SDH adoption was strongest at sites, where the platform mirrored and enhanced existing workflows rather than replacing them entirely. Users at these sites reported reduced learning curves and described the platform as intuitive to adopt.

Role-Based Access Design

The role-differentiated account structure in SDH was consistently cited as a facilitator. Users received dashboards and input fields scoped to their specific responsibilities, reducing cognitive overload and increasing the perceived relevance of the platform to day-to-day tasks.

Peer Counselor-Led Outreach

Peer counselors with mental health training and established peer relationships demonstrated significantly higher engagement conversion than incentivized ambassadors. This has directly informed the revised outreach strategy for subsequent Rafi deployments, shifting from large launch-event acquisition toward gradual rollout through trusted peer networks.

Free Access Model

The absence of licensing fees for both SDH and Rafi was consistently identified by implementation partners as a critical enabler, removing the cost barrier that had prevented uptake of commercially licensed platforms reviewed in the environmental scan.

Offline-First Functionality in SDH

The ability to capture session data without active internet connectivity was consistently highlighted by supervisors at periurban sites as a key facilitator, reducing dependency on stable connectivity for core operational data entry.

Discussion

Principal Findings

This study documents the development and early deployment of shamiriOS, an open-source, modular digital platform for scalable, task-shifted youth mental health service delivery in Kenya [15,20,23]. Three principal findings emerge.

First, an environmental scan confirmed that no existing platform adequately met the combined operational, clinical, and youth engagement requirements of the Shamiri stepped-care model. Reviewed platforms failed on at least 2 of 6 evaluation criteria—most commonly cost, offline capability, and customizability for task-shifted workflows. This reflects not a critique of those platforms, but confirmation that the infrastructure gap in community-based mental health delivery in Kenya was real and unaddressed by the global product landscape [34].

Second, the SDH demonstrated strong early feasibility and acceptability. Deployed across 5 hubs by the end of 2023 and

XSL•FO
RenderX

expanding to 11 sites supporting over 76,344 students by Q1 2024, SDH replaced fragmented operational systems with real-time, role-differentiated digital infrastructure. Implementation staff reported high satisfaction (usability: mean 8.36, SD 1.49; NPS: mean 8.63, SD 1.46), and qualitative feedback consistently identified reductions in administrative burden and improvements in data timeliness and completeness. These results are consistent with evidence that well-designed operational platforms improve efficiency and oversight capacity in task-shifted delivery systems in LMICs [29,31,35-37,56].

Third, Rafi demonstrated strong initial reach—74.7% (n=3737) registration among the targeted MKU cohort—but significant engagement attrition, with MAU rates declining from 7.4% in January 2024 to near 0 by September. This pattern is common in digital mental health globally [37,39,57-60], but operational analysis identified specific, modifiable causes beyond structural or population-level explanations [39,58-60]. Among users who engaged with self-guided features, 99.4% subsequently booked therapy appointments—a cross-sectional finding requiring cautious interpretation, but one suggesting that self-help engagement may function as a gateway to clinical help-seeking for those who remain on the platform.

shamiriOS in the Context of Existing Digital Mental Health Platforms

shamiriOS contributes to a growing landscape of digital mental health platforms in LMICs [29-31,35-37]. What distinguishes it is not any single feature, similar apps exist, but 3 architectural characteristics that together address gaps in the current platform landscape [61,62].

First, shamiriOS integrates operational management (SDH), youth engagement (Rafi), and clinical workflows (SPP) within a single interoperable ecosystem. Most existing platforms address one of these functions; few attempt integration across all 3. This reflects our view that sustainable digital mental health infrastructure must support not just the end-user experience, but the full organizational and clinical system within which that experience is embedded [63,64].

Second, the platform uses a modular, component-based design in which each suite operates as an independent service with its own codebase, database, and deployment pipeline, connected through defined API contracts. An implementing organization could adopt SDH without deploying Rafi or adapt Rafi for a different clinical context without requiring the full SDH infrastructure. RBAC is configurable per deployment—the same instance can serve a secondary school program with 3 user roles or a university-based stepped-care service with 10—without code changes. This modularity is deliberate: no 2 implementation contexts share identical operational requirements, and platform rigidity is a common reason digital health tools fail when transferred across settings [63,64].

Third, the open-source architecture—built on Node.js, PostgreSQL, and React, with RESTful APIs and a phased Fast Healthcare Interoperability Resources compliance roadmap—lowers the barrier to adaptation by other technical teams [63-65]. Offline-first data capture, customizable reporting, and flexible intervention workflows are configurable features rather than hardcoded assumptions [63-65].

To the best of our knowledge, no comparable integrated open-source platform has been developed and deployed for community-based stepped-care mental health delivery in Kenya or comparable East African settings. We make no claim to novelty for the global landscape, where numerous digital mental health platforms exist [30,31,35]. The contribution is contextual: a platform built from the ground up for the operational, cultural, and infrastructural realities of implementation in Kenya and similar low-resource settings.

Extending shamiriOS to nonmental-health domains is a plausible future direction, but currently hypothetical. The modular architecture could, in principle, support other task-shifted, community-based delivery programs, but such extensions would require domain-specific user research and evaluation rather than direct platform transfer.

We believe publishing shamiriOS as open source serves an important purpose: making the architecture and code available for other teams to adapt, critique, and build upon. The precedent set by OpenMRS and DHIS2—adapted across dozens of countries while maintaining core architectural integrity—suggests that platform-level adaptability is achievable when accompanied by community-led localization [65-68]. Implementation toolkits, training modules, and documentation are being developed to lower the barrier for organizations in Kenya and comparable settings, with the explicit expectation that substantial localization will be required.

shamiriOS Can Support Backend Operations Within Task-Shifted Youth Mental Health Systems

The SDH was feasible and acceptable across both centralized and decentralized implementation contexts. The transition from fragmented workflows—spreadsheets, manual stipend logs, and WhatsApp-based supervision coordination—to a unified real-time platform was achieved across 5 hubs in Q4 2023, expanding to 11 sites supporting 1195 fellows, 111 supervisors, and over 76,344 students by Q1 2024. A 3.7-fold increase in student reach within a single year, maintained alongside high satisfaction scores, suggests that the platform can scale without deterioration in usability or staff experience. An NPS of 8.63 and a usability rating of 8.36 are consistent with evidence that well-designed operational platforms improve efficiency and oversight capacity in task-shifted delivery systems [30,31,35,56]. Qualitative feedback highlighted 3 valued features: centralized real-time visibility of fellow attendance across sites; automated stipend workflows, which reduced manual verification errors; and Metabase analytics dashboards, which allowed nontechnical supervisors to interrogate program data without developer support. The last is particularly relevant for scale: building analytical self-sufficiency into the platform reduces long-term dependency on the central technical team [30,31,35,56].

Operational feedback from the wave 2 expansion revealed several implementation-relevant patterns. Adoption was strongest at sites where SDH mirrored and enhanced existing workflows rather than replacing them wholesale—consistent with implementation science evidence that technology adoption

XSL•FO
RenderX

is smoother when tools augment rather than disrupt established practice [31,64]. Device sharing at some hubs created bottlenecks in timely data entry, temporarily reducing completeness in the first weeks of deployment. Offline-first data capture—introduced in version 3—addressed connectivity constraints but not device access. Future deployments to settings with low device-to-user ratios may need to treat hardware access as a precondition for adoption or consider lighter-weight interfaces optimized for brief interactions on shared devices.

Engagement Is a Challenge for Youth-Facing Apps Like Rafi

The gap between registration and sustained engagement was striking. High initial registration confirms that broad uptake is achievable through institutional partnerships and peer-led outreach. The subsequent decline—with sustained active use falling well below the 4% at 6 months common in reviews of consumer mental health apps [39,58]—is consistent with the literature but requires careful disaggregation.

Qualitative feedback identified several contributing factors: low perceived ongoing need among students who registered out of curiosity during launch, feature overload from multiple simultaneous engagement pathways, device storage constraints, and notification fatigue. These represent different problems requiring different solutions—notification calibration addresses fatigue but does nothing for users with genuinely low perceived need or stigma concerns [69-71].

More structurally, Rafi was designed around continuous self-monitoring—daily check-ins, journaling, and affirmations—which requires users to perceive ongoing value in routine self-assessment. For a predominantly healthy student population, this model may not fit typical help-seeking patterns, which tend to be episodic and crisis-driven rather than habitual [69-71]. Future iterations should consider whether a shift toward episodic engagement—designed for crisis-prompted access to care rather than daily wellness tracking—would better match actual user needs [39].

The finding that 99.4% of Rafi users who engaged with self-guided tools subsequently booked a therapy appointment—though cross-sectional and not causal—suggests that for students who do engage, the platform functions effectively as a gateway to higher-intensity care [69,71]. This points to a segmented engagement strategy: prioritizing acquisition and early retention for students with active mental health needs, rather than sustained mass engagement across the full student population.

shamiriOS Features in Development

Overview

Ongoing development is guided by field feedback, platform analytics, and emerging implementation needs [18,20,23,72,73].

SDH is being upgraded for more robust multisite operations, with planned enhancements including automated triage escalation for flagged students, expanded school and hub-level reporting, credential management, and intervention quality dashboards [18,23]. Standardized onboarding and role-based interfaces are being embedded to support consistency, as the platform scales to new partner organizations [18,23].

The next version of Rafi is designed to address the engagement challenges described earlier. Design priorities include gamified onboarding sequences, mood-responsive motivational content, and more direct pathways connecting students to therapists. Personalized content journeys adapted to user behavior are intended to reduce early drop-off [32,74,75]. The interface is being restructured to reduce feature overload, beginning with a simpler onboarding pathway before exposing users to the full feature set. Rafi's architecture is also being refactored for tighter SPP integration, supporting seamless transitions from self-guided to provider-supported care [30,32].

SPP is not reported as a deployed component in this study by design: it reached production readiness after the observation period, and its deployment and clinical integration merit are ongoing. Still, SPP integration continues with development priorities including real-time therapist scheduling, digital session documentation, and bidirectional data flow between clinical and operational layers. Historical client data migration is in progress to ensure continuity of care across platforms [33,34].

Future Development: AI Integration

A key area of planned development is the integration of AI tools—referred to internally as shamiriAI—across 3 priority domains: AI-assisted provider training and supervision, including automated fidelity scoring of intervention sessions; precision-matching of users to content and care levels based on symptom severity and engagement behavior; and symptom network analysis to inform modular intervention design. The potential benefits—particularly for supervision at scale, where human oversight capacity is a binding constraint in task-shifted delivery—are significant [31,76].

AI integration in mental health carries well-documented risks that are amplified in the Kenyan context [77-80]. These include dangerous clinical advice from language models, inappropriate substitution of AI for human therapeutic relationships, exploitation of sensitive personal data, and perpetuation of biases embedded in training datasets [77,81,82]. Dominant large language models are trained overwhelmingly on English-language [83], Western psychological literature [84,85], creating fundamental misalignment with Kenyan contexts where the experience of mental health may differ from Western models [86-90]. Regulatory frameworks in most African countries have not kept pace with the speed of AI deployment [77,78,91].

For these reasons, shamiriOS approaches AI integration as a learning-first rather than deployment-first activity. Guiding principles include an "AI-in-the-loop" design retaining human agency over care; training data grounded in African languages and contexts; community oversight involving youths, fellows, and clinicians as reviewers rather than only end users; strict data governance; and prospective safety monitoring. shamiriAI is presented as a set of research questions to be answered through rigorous evaluation. Future phases will assess feasibility, acceptability, cultural appropriateness, and safety before any AI feature enters service delivery [77].

XSL•FO
RenderX

Data Protection and Engagement Considerations

shamiriOS was designed to improve access to youth mental health services, but digital data collection introduces risks requiring explicit acknowledgment. Kenya's Data Protection Act (2019) provides a legal framework [91], but enforcement capacity is limited, and the gap between legal protection and practical data security in community-based implementations remains significant [92-94].

Current data protection measures include end-to-end HTTPS encryption, data deidentification, RBACs, separation of operational data (SDH) from clinical and personal data (Rafi and SPP), secure cloud hosting with encrypted databases, and Open Authorization 2.0 authentication. Rafi does not collect passive sensor data. Anonymous community feed features allow engagement without linking posts to clinical data.

These measures reduce but do not eliminate risk [92-94]. Future work will strengthen protections through user-controlled data deletion and full data export; transparent disclosure about what data are collected, how they are stored, and scenarios under which they could be accessed by third parties; and, where technically feasible, local device storage options that reduce cloud dependency [92-94]. shamiriOS's data governance policies will be published openly as part of the open-source repository to enable external scrutiny.

shamiriOS was developed through a user-centered design process [40,41]—not community-led development. End users informed design decisions as consultants and testers; strategic and architectural decisions were made by Shamiri's technical and clinical leadership. This reflects pragmatic constraints of real-world development and believe that the level of inclusion on feature development, content priorities, and data governance is appropriate for platforms at this stage. Future governance structures should consider greater community ownership [42,45].

Limitations

Several limitations should inform the interpretation of these findings. The environmental scan was structured but not exhaustive. Proprietary, region-specific, or non-English–language platforms may not have been captured, and the scan assessed market availability and feature coverage rather than clinical efficacy; it should not be read as a systematic review.

We did not use validated implementation science measures for feasibility, acceptability, or usability. Pragmatic industry metrics—NPS, MAU rates, and feature adoption rates—provided actionable feedback for development sprints but lack the reliability and cross-study comparability of instruments such as the System Usability Scale [95], Acceptability of Intervention Measure [53], or Feasibility of Intervention Measure [53]. Future evaluations of the stable platform should use these standardized instruments.

Barriers and facilitators were documented through operational feedback channels—training sessions, helpdesk tickets, monthly check-ins, and NPS qualitative responses—rather than formal implementation research using validated frameworks such as Consolidated Framework for Implementation Research or Exploration, Preparation, Implementation, Sustainment [96,97]. Our understanding of why engagement varied across sites and user groups is therefore incomplete. Future studies must prioritize rigorous implementation science methods.

Deployment results—particularly Rafi's MKU engagement data—reflect a 9-month observation window. Longitudinal engagement patterns, clinical outcomes, and cost-effectiveness at scale have not been assessed. The 99.4% therapy-booking rate is a cross-sectional observation that cannot establish causality.

Finally, shamiriOS used user-centered design with consultation, not fully participatory or community-led design. Future studies should consider a full participatory approach and investigate the implications of different approaches on cultural grounding, local ownership, and long-term sustainability.

Conclusions

The development and early deployment of shamiriOS demonstrate the feasibility of modular, open-source digital infrastructure for scalable, task-shifted youth mental health service delivery in Kenya. SDH showed strong feasibility and acceptability at scale, with satisfaction scores comparing favorably with digital health benchmarks and a 3.7-fold expansion in reach over a single year. Rafi demonstrated broad initial reach but surfaced important, actionable lessons about the conditions required for sustained engagement—lessons specific enough to have directly informed deployment strategy for future cohorts.

shamiriOS is not a finished product. Its full value will only be realized once all 3 components are deployed, integrated, and rigorously evaluated. The evidence here establishes a foundation: the core infrastructure works, it is valued by its users, and the most significant obstacles to impact lie not in platform design but in implementation readiness, incentive structures, and institutional integration—all of which are addressable.

We envision shamiriOS as an open-source, contextually grounded infrastructure for decentralized mental health service delivery in Kenya, available for adaptation by teams in comparable settings—with adaptation requiring necessary contextual redesign. By publishing both the platform and an account of its early deployment, we aim to contribute not only a technical resource but a set of implementation lessons to the broader community working to close the youth mental health treatment gap in Kenya and beyond.

Acknowledgments

The authors would like to acknowledge the contributions of the Agency Fund team, including Linus Wong, Robert On, Paul Wang, and James Walsh, who provided critical support during the conceptualization, design, and early development phases of

XSL•FO
RenderX

the platform. The authors also thank Rosine Baseke, Veronica Ngatia, and other members of the Shamiri Institute staff whose input and collaboration were instrumental in shaping the design and functionality of shamiriOS.

Funding

The development of shamiriOS was supported by funding from the Agency Fund, the Fund for Innovation in Development, and the Inner Foundation.

Conflicts of Interest

SL, TLO, WM, JK, BO, IE, and CW are current or former employees of Shamiri Institute, the nonprofit organization that developed and maintains shamiriOS. Shamiri Institute received funding from the Agency Fund, the Fund for Innovation in Development, and the Inner Foundation to support platform development. JT and EK are affiliated with external organizations that provided technical or funding support. No author received personal financial benefit from shamiriOS, and the platform is released as open-source software. The authors report no other conflicts of interest.

Multimedia Appendix 1
Environmental scan: methodology and reviewed platforms.
[DOCX File , 97 KB - humanfactors_v13i1e79107_app1.docx]

Multimedia Appendix 2
Rafi: co-design methodology and version history.
[DOCX File , 96 KB - humanfactors_v13i1e79107_app2.docx]

Multimedia Appendix 3
Shamiri Digital Hub: technical architecture, development, and deployment outcomes.
[DOCX File , 103 KB - humanfactors_v13i1e79107_app3.docx]

Multimedia Appendix 4
Shamiri Provider Platform: EasyClinic adaptation and Fast Healthcare Interoperability Resources compliance roadmap.
[DOCX File , 98 KB - humanfactors_v13i1e79107_app4.docx]

References

1. World mental health report: transforming mental health for all. World Health Organization. 2022. URL: https://www.who.int/publications/i/item/9789240049338 [accessed 2026-04-04]
2. The Lancet. Better understanding of youth mental health. Lancet 2017;389(10080):1670. [doi: 10.1016/S0140-6736(17)31140-6] [Medline: 28463125]
3. Fuhrmann D, Knoll LJ, Blakemore S. Adolescence as a sensitive period of brain development. Trends Cogn Sci 2015;19(10):558-566. [doi: 10.1016/j.tics.2015.07.008] [Medline: 26419496]
4. Axelson DA, Birmaher B. Relation between anxiety and depressive disorders in childhood and adolescence. Depress Anxiety 2001;14(2):67-78. [doi: 10.1002/da.1048] [Medline: 11668659]
5. Mental health atlas 2017. World Health Organization. 2018. URL: https://www.who.int/publications/i/item/mental-health-atlas-2017 [accessed 2026-04-14]
6. The global health observatory. World Health Organization. 2019. URL: https://www.who.int/data/gho [accessed 2026-04-10]
7. Osborn TL, Wasanga CM, Ndetei DM. Transforming mental health for all. BMJ 2022;377:o1593. [doi: 10.1136/bmj.o1593] [Medline: 35772771]
8. Ndetei DM, Mutiso V, Osborn T. Moving away from the scarcity fallacy: three strategies to reduce the mental health treatment gap in LMICs. World Psychiatry 2023;22(1):163-164 [FREE Full text] [doi: 10.1002/wps.21054] [Medline: 36640407]
9. Bolton P. Global Mental Health and Psychotherapy: Importance of Task-Shifting and a Systematic Approach to Adaptation. Amsterdam, Netherlands: Elsevier; 2019:11-24.
10. Joshi R, Alim M, Kengne AP, Jan S, Maulik PK, Peiris D, et al. Task shifting for non-communicable disease management in low and middle income countries—a systematic review. PLoS One 2014;9(8):e103754 [FREE Full text] [doi: 10.1371/journal.pone.0103754] [Medline: 25121789]
11. Kern L, Rusnak LK. Tiered mental health support: meeting the needs of all students. In: Weist MD, McQuillin SD, Kern L, editors. Scaling Effective School Mental Health Interventions and Practices. Cham: Springer Nature Switzerland; 2024:19-35.
12. Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency. Narrative literature review. Br J Psychiatry 2005;186:11-17. [doi: 10.1192/bjp.186.1.11] [Medline: 15630118]

XSL•FO
RenderX

13. Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. Behav Res Ther 2017;88:19-25 [FREE Full text] [doi: 10.1016/j.brat.2016.08.012] [Medline: 28110672]
14. Osborn TL, Wasil AR, Venturo-Conerly KE, Schleider JL, Weisz JR. Group intervention for adolescent anxiety and depression: outcomes of a randomized trial with adolescents in Kenya. Behav Ther 2020;51(4):601-615. [doi: 10.1016/j.beth.2019.09.005] [Medline: 32586433]
15. Osborn TL, Venturo-Conerly KE, Arango GS, Roe E, Rodriguez M, Alemu RG, et al. Effect of Shamiri layperson-provided intervention vs study skills control intervention for depression and anxiety symptoms in adolescents in Kenya: a randomized clinical trial. JAMA Psychiatry 2021;78(8):829-837 [FREE Full text] [doi: 10.1001/jamapsychiatry.2021.1129] [Medline: 34106239]
16. Venturo-Conerly K, Roe E, Wasil A, Osborn T, Ndetei D, Musyimi C, et al. Training and supervising lay providers in Kenya: strategies and mixed-methods outcomes☆☆☆. Cogn Behav Pract 2022;29(3):666-681. [doi: 10.1016/j.cbpra.2021.03.004]
17. Osborn TL, Venturo-Conerly KE, Wasil AR, Rodriguez M, Roe E, Alemu R, et al. The Shamiri group intervention for adolescent anxiety and depression: study protocol for a randomized controlled trial of a lay-provider-delivered, school-based intervention in Kenya. Trials 2020;21(1):938 [FREE Full text] [doi: 10.1186/s13063-020-04732-1] [Medline: 33225978]
18. Kahi S, Memba L, Syan A, Ngatia V, Venturo-Conerly K, Wasanga C, et al. Implementation of a school-based risk management protocol within a task-shifted mental healthcare model. Glob Ment Health (Camb) 2025;12:e127. [doi: 10.1017/gmh.2025.10073] [Medline: 41292604]
19. Venturo-Conerly KE, Wasil AR, Osborn TL, Puffer ES, Weisz JR, Wasanga CM. Designing culturally and contextually sensitive protocols for suicide risk in global mental health: lessons from research with adolescents in Kenya. J Am Acad Child Adolesc Psychiatry 2022;61(9):1074-1077. [doi: 10.1016/j.jaac.2022.02.005] [Medline: 35217169]
20. Venturo-Conerly KE, Osborn TL, Rusch T, Ochuku BK, Johnson NE, van der Markt A, et al. Testing the Shamiri intervention and its components with Kenyan adolescents during the COVID-19 pandemic: outcomes of a universal, 5-arm randomized controlled trial. J Am Acad Child Adolesc Psychiatry 2025;64(7):786-798. [doi: 10.1016/j.jaac.2024.04.015] [Medline: 38851382]
21. Venturo-Conerly KE, Osborn TL, Alemu R, Roe E, Rodriguez M, Gan J, et al. Single-session interventions for adolescent anxiety and depression symptoms in Kenya: a cluster-randomized controlled trial. Behav Res Ther 2022;151:104040. [doi: 10.1016/j.brat.2022.104040] [Medline: 35168011]
22. Kacmarek CN, Johnson NE, Osborn TL, Wasanga C, Weisz JR, Yates BT. Costs and cost-effectiveness of Shamiri, a brief, layperson-delivered intervention for Kenyan adolescents: a randomized controlled trial. BMC Health Serv Res 2023;23(1):827 [FREE Full text] [doi: 10.1186/s12913-023-09856-z] [Medline: 37542304]
23. Ochuku B, Osborn TL, Nerima D, van der Markt A, Rusch T, Omune H, et al. Testing pathways to scale: study protocol for a three-arm randomized controlled trial of a centralized and a decentralized ("Train the Trainers") dissemination of a mental health program for Kenyan adolescents. Trials 2023 Aug 13;24(1):526 [FREE Full text] [doi: 10.1186/s13063-023-07539-y] [Medline: 37574545]
24. Yarber L, Brownson CA, Jacob RR, Baker EA, Jones E, Baumann C, et al. Evaluating a train-the-trainer approach for improving capacity for evidence-based decision making in public health. BMC Health Serv Res 2015;15:547 [FREE Full text] [doi: 10.1186/s12913-015-1224-2] [Medline: 26652172]
25. Starr K, Hattendorf L. The doer and the payer: a simple approach to scale. Standford Soc Innov Rev 2015. [doi: 10.48558/ET3G-NY28]
26. Le PD, Eschliman EL, Grivel MM, Tang J, Cho YG, Yang X, et al. Barriers and facilitators to implementation of evidence-based task-sharing mental health interventions in low- and middle-income countries: a systematic review using implementation science frameworks. Implement Sci 2022;17(1):4 [FREE Full text] [doi: 10.1186/s13012-021-01179-z] [Medline: 35022081]
27. Wasil AR, Osborn TL, Venturo-Conerly KE, Wasanga C, Weisz JR. Conducting global mental health research: lessons learned from Kenya. Glob Ment Health (Camb) 2021;8:e8 [FREE Full text] [doi: 10.1017/gmh.2021.7] [Medline: 34026238]
28. Mudiyanselage KWW, De Santis KK, Jörg F, Saleem M, Stewart R, Zeeb H, et al. The effectiveness of mental health interventions involving non-specialists and digital technology in low-and middle-income countries—a systematic review. BMC Public Health 2024;24(1):77 [FREE Full text] [doi: 10.1186/s12889-023-17417-6] [Medline: 38172713]
29. Wani C, McCann L, Lennon M, Radu C. Digital mental health interventions for adolescents in low- and middle-income countries: scoping review. J Med Internet Res 2024;26:e51376 [FREE Full text] [doi: 10.2196/51376] [Medline: 39471371]
30. Naslund JA, Aschbrenner KA, Araya R, Marsch LA, Unützer J, Patel V, et al. Digital technology for treating and preventing mental disorders in low-income and middle-income countries: a narrative review of the literature. Lancet Psychiatry 2017;4(6):486-500 [FREE Full text] [doi: 10.1016/S2215-0366(17)30096-2] [Medline: 28433615]
31. Naslund JA, Shidhaye R, Patel V. Digital technology for building capacity of nonspecialist health workers for task sharing and scaling up mental health care globally. Harv Rev Psychiatry 2019;27(3):181-192 [FREE Full text] [doi: 10.1097/HRP.0000000000000217] [Medline: 30958400]

XSL•FO
RenderX

32. Carter H, Araya R, Anjur K, Deng D, Naslund JA. The emergence of digital mental health in low-income and middle-income countries: a review of recent advances and implications for the treatment and prevention of mental disorders. J Psychiatr Res 2021;133:223-246 [FREE Full text] [doi: 10.1016/j.jpsychires.2020.12.016] [Medline: 33360867]
33. Philippe TJ, Sikder N, Jackson A, Koblanski ME, Liow E, Pilarinos A, et al. Digital health interventions for delivery of mental health care: systematic and comprehensive meta-review. JMIR Ment Health 2022;9(5):e35159 [FREE Full text] [doi: 10.2196/35159] [Medline: 35551058]
34. Merchant R, Torous J, Rodriguez-Villa E, Naslund JA. Digital technology for management of severe mental disorders in low-income and middle-income countries. Curr Opin Psychiatry 2020;33(5):501-507 [FREE Full text] [doi: 10.1097/YCO.0000000000000626] [Medline: 32520747]
35. Chau LW, Lam RW, Minas H, Hayashi K, Nguyen VC, O'Neil J. Digital health interventions for depression and anxiety in low- and middle-income countries: rapid scoping review. JMIR Ment Health 2025 Aug 22;12(5):e68296 [FREE Full text] [doi: 10.2196/68296] [Medline: 40846319]
36. Lehtimaki S, Martic J, Wahl B, Foster KT, Schwalbe N. Evidence on digital mental health interventions for adolescents and young people: systematic overview. JMIR Ment Health 2021;8(4):e25847 [FREE Full text] [doi: 10.2196/25847] [Medline: 33913817]
37. Alagarajah J, Ceccolini D, Butler S. Digital mental health interventions for treating mental disorders in young people based in low-and middle-income countries: a systematic review of the literature. Glob Ment Health (Camb) 2024;11:e74. [doi: 10.1017/gmh.2024.71] [Medline: 39314998]
38. Mental health atlas 2024. World Health Organization. 2024. URL: https://www.who.int/publications/i/item/9789240114487 [accessed 2026-04-04]
39. Wasil AR, Venturo-Conerly KE, Shingleton RM, Weisz JR. A review of popular smartphone apps for depression and anxiety: assessing the inclusion of evidence-based content. Behav Res Ther 2019;123:103498. [doi: 10.1016/j.brat.2019.103498] [Medline: 31707224]
40. Hagen P, Collin P, Metcalf A. Design of Evidence-Based Online Youth Mental Health Promotion, Prevention, Early Intervention and Treatment. Australia: Young and Well CRC; 2011.
41. Orlowski SK, Lawn S, Venning A, Winsall M, Jones GM, Wyld K, et al. Participatory research as one piece of the puzzle: a systematic review of consumer involvement in design of technology-based youth mental health and well-being interventions. JMIR Hum Factors 2015;2(2):e12 [FREE Full text] [doi: 10.2196/humanfactors.4361] [Medline: 27025279]
42. Holeman I, Kane D. Human-centered design for global health equity. Inf Technol Dev 2019;26(3):477-505 [FREE Full text] [doi: 10.1080/02681102.2019.1667289] [Medline: 32982007]
43. Belindar. Human-centered design is not a one-size-fits all solution. HCDExchange. 2023. URL: https://hcdexchange.org/human-centered-design-is-not-a-one-size-fits-all/ [accessed 2026-03-01]
44. Singh DR, Sah RK, Simkhada B, Darwin Z. Potentials and challenges of using co-design in health services research in low- and middle-income countries. Glob Health Res Policy 2023;8(1):5 [FREE Full text] [doi: 10.1186/s41256-023-00290-6] [Medline: 36915174]
45. Göttgens I, Oertelt-Prigione S. The application of human-centered design approaches in health research and innovation: a narrative review of current practices. JMIR Mhealth Uhealth 2021;9(12):e28102 [FREE Full text] [doi: 10.2196/28102] [Medline: 34874893]
46. Beck K, Beedle M, Van BA. Manifesto for Agile Software Development. 2001. URL: https://agilemanifesto.org/ [accessed 2026-04-10]
47. Royce RW. Managing the development of large software systems: concepts and techniques. 1987 Presented at: ICSE '87: Proceedings of the 9th international conference on Software Engineering; March 1, 1987; Monterey, CA, United States p. 328-388.
48. Abraham T. Metabase Up and Running: Introduce Business Intelligence and Analytics to Your Company and Make Better Business Decisions. Birmingham: Packt Publishing Ltd; 2020.
49. Santos B, Sério F, Abrantes S. Open source business intelligence tools: Metabase and Redash. 2019 Presented at: Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2019); September 17-19, 2019; Vienna, Austria p. 467-474. [doi: 10.5220/0008351704670474]
50. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
51. Reichheld FF. The one number you need to grow. Harv Bus Rev 2003;81(12):46-54, 124. [Medline: 14712543]
52. Glasgow RE, Harden SM, Gaglio B, Rabin B, Smith ML, Porter GC, et al. RE-AIM planning and evaluation framework: adapting to new science and practice with a 20-year review. Front Public Health 2019;7:64 [FREE Full text] [doi: 10.3389/fpubh.2019.00064] [Medline: 30984733]
53. Weiner BJ, Lewis CC, Stanick C, Powell BJ, Dorsey CN, Clary AS, et al. Psychometric assessment of three newly developed implementation outcome measures. Implement Sci 2017;12(1):108 [FREE Full text] [doi: 10.1186/s13012-017-0635-3] [Medline: 28851459]

XSL•FO
RenderX

54. National guidelines for ethical conduct of biomedical research involving human participants in Kenya. National Commission for Science, Technology and Technology. 2020. URL: https://tinyurl.com/y7wednc3 [accessed 2026-05-19]
55. The data protection act. Kenya Subsidiary Legislation. URL: https://www.odpc.go.ke/wp-content/uploads/2024/03/THE-DATA-PROTECTION-GENERAL-REGULATIONS-2021-1.pdf [accessed 2026-05-19]
56. Berardi C, Antonini M, Jordan Z, Wechtler H, Paolucci F, Hinwood M. Barriers and facilitators to the implementation of digital technologies in mental health systems: a qualitative systematic review to inform a policy framework. BMC Health Serv Res 2024;24(1):243 [FREE Full text] [doi: 10.1186/s12913-023-10536-1] [Medline: 38408938]
57. Baka E, Tan Y, Wong B, Xing Z, Yap P. A scoping review of digital interventions for the promotion of mental health and prevention of mental health conditions for young people. Oxf Open Digit Health 2025;3:oqaf005. [doi: 10.1093/oodh/oqaf005] [Medline: 40230873]
58. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
59. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
60. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018;21(3):116-119 [FREE Full text] [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
61. Arora PG, Collins TA, Dart EH, Hernández S, Fetterman H, Doll B. Multi-tiered systems of support for school-based mental health: a systematic review of depression interventions. School Ment Health 2019;11(2):240-264. [doi: 10.1007/s12310-019-09314-4]
62. Ladegard K, Alleyne S, Close J, Hwang MD. The role of school-based interventions and communities for mental health prevention, tiered levels of care, and access to care. Child Adolesc Psychiatr Clin N Am 2024;33(3):381-395. [doi: 10.1016/j.chc.2024.03.004] [Medline: 38823811]
63. Bathelt F, Lorenz S, Weidner J, Sedlmayr M, Reinecke I. Application of modular architectures in the medical domain—a scoping review. J Med Syst 2025;49(1):27. [doi: 10.1007/s10916-025-02158-3] [Medline: 39964566]
64. Casanova R, Villa-Garzon FA, Branch-Bedoya JW. Architectural patterns for health information systems: a systematic review. Front Digit Health 2025;7:1694839 [FREE Full text] [doi: 10.3389/fdgth.2025.1694839] [Medline: 41333107]
65. Mamlin BW, Biondich P, Wolfe B, Fraser H, Jazayeri D, Allen C, et al. Cooking up an open source EMR for developing countries: OpenMRS—a recipe for successful collaboration. AMIA Annu Symp Proc 2006;2006:529-533 [FREE Full text] [Medline: 17238397]
66. Braa J, Sahay S. The DHIS2 Open Source Software Platform: Evolution Over Time and Space. Cambridge, MA: Massachusetts Institute of Technology; 2017:1-13.
67. Dehnavieh R, Haghdoost A, Khosravi A, Hoseinabadi F, Rahimi H, Poursheikhali A, et al. The District Health Information System (DHIS2): a literature review and meta-synthesis of its strengths and operational challenges based on the experiences of 11 countries. Health Inf Manag 2019;48(2):62-75. [doi: 10.1177/1833358318777713] [Medline: 29898604]
68. Byrne E, Sæbø JI. Routine use of DHIS2 data: a scoping review. BMC Health Serv Res 2022;22(1):1234 [FREE Full text] [doi: 10.1186/s12913-022-08598-8] [Medline: 36203141]
69. Jin S, Kim B, Han K. "I Don't Know Why I Should Use This App": holistic analysis on user engagement challenges in mobile mental health. 2025 Presented at: CHI '25: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems; April 26-May 1, 2025; Yokohama, Japan p. 1-23. [doi: 10.1145/3706598.3713732]
70. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
71. Rickwood DJ, Deane FP, Wilson CJ. When and how do young people seek professional help for mental health problems? Med J Aust 2007;187(S7):S35-S39. [doi: 10.5694/j.1326-5377.2007.tb01334.x] [Medline: 17908023]
72. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013;1(2):160-171 [FREE Full text] [doi: 10.9745/GHSP-D-13-00031] [Medline: 25276529]
73. Mehl G, Labrique A. Prioritizing integrated mHealth strategies for universal health coverage. Science 2014;345(6202):1284-1287. [doi: 10.1126/science.1258926] [Medline: 25214614]
74. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
75. Klein A, Clucas J, Krishnakumar A, Ghosh SS, Van Auken W, Thonet B, et al. Remote digital psychiatry for mobile mental health assessment and therapy: MindLogger platform development study. J Med Internet Res 2021;23(11):e22369 [FREE Full text] [doi: 10.2196/22369] [Medline: 34762054]

XSL•FO
RenderX

76. Espino Carrasco DK, Palomino Alcántara MDR, Arbulú Pérez Vargas CG, Santa Cruz Espino BM, Dávila Valdera LJ, Vargas Cabrera C, et al. Sustainability of AI-assisted mental health intervention: a review of the literature from 2020-2025. Int J Environ Res Public Health 2025;22(9):1382 [FREE Full text] [doi: 10.3390/ijerph22091382] [Medline: 41007526]
77. Stade EC, Stirman SW, Ungar LH, Boland CL, Schwartz HA, Yaden DB, et al. Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. Npj Ment Health Res 2024;3(1):12 [FREE Full text] [doi: 10.1038/s44184-024-00056-z] [Medline: 38609507]
78. Hua Y, Na H, Li Z, Liu F, Fang X, Clifton D, et al. A scoping review of large language models for generative tasks in mental health care. NPJ Digit Med 2025;8(1):230 [FREE Full text] [doi: 10.1038/s41746-025-01611-4] [Medline: 40307331]
79. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth 2018;6(11):e12106 [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]
80. Varghese MA, Sharma P, Patwardhan M. Public perception on artificial intelligence-driven mental health interventions: survey research. JMIR Form Res 2024;8:e64380 [FREE Full text] [doi: 10.2196/64380] [Medline: 39607994]
81. van Heerden AC, Pozuelo JR, Kohrt BA. Global mental health services and the impact of artificial intelligence-powered large language models. JAMA Psychiatry 2023;80(7):662-664. [doi: 10.1001/jamapsychiatry.2023.1253] [Medline: 37195694]
82. Straw I, Callison-Burch C. Artificial intelligence in mental health and the biases of language based models. PLoS One 2020;15(12):e0240376 [FREE Full text] [doi: 10.1371/journal.pone.0240376] [Medline: 33332380]
83. Na H, Hua Y, Wang Z. A survey of large language models in psychotherapy: current landscape and future directions. 2025 Presented at: Findings of the Association for Computational Linguistics: ACL 2025; April 8, 2026; Vienna, Austria p. 7362-7376. [doi: 10.18653/v1/2025.findings-acl.385]
84. Lawrence HR, Schneider RA, Rubin SB, Matarić MJ, McDuff DJ, Jones Bell M. The opportunities and risks of large language models in mental health. JMIR Ment Health 2024;11:e59479 [FREE Full text] [doi: 10.2196/59479] [Medline: 39105570]
85. Hendriks T, Warren MA, Schotanus-Dijkstra M, Hassankhan A, Graafsma T, Bohlmeijer E, et al. How WEIRD are positive psychology interventions? A bibliometric analysis of randomized controlled trials on the science of well-being. J Posit Psychol 2018;14(4):489-501. [doi: 10.1080/17439760.2018.1484941]
86. Kleinman AM. Depression, somatization and the "new cross-cultural psychiatry". Soc Sci Med (1967) 1977;11(1):3-10. [doi: 10.1016/0037-7856(77)90138-x] [Medline: 887955]
87. Osborn TL, Kleinman A, Weisz J. Complementing standard western measures of depression with locally co-developed instruments: a cross-cultural study on the experience of depression among the Luo in Kenya. Transcult Psychiatry 2021;58(4):499-515. [doi: 10.1177/13634615211000555] [Medline: 33818199]
88. Kirmayer LJ, Ryder AG. Culture and psychopathology. Curr Opin Psychol 2016;8:143-148. [doi: 10.1016/j.copsyc.2015.10.020] [Medline: 29506790]
89. Haroz EE, Ritchey M, Bass JK, Kohrt BA, Augustinavicius J, Michalopoulos L, et al. How is depression experienced around the world? A systematic review of qualitative literature. Soc Sci Med 2017;183:151-162 [FREE Full text] [doi: 10.1016/j.socscimed.2016.12.030] [Medline: 28069271]
90. Owoyemi A, Owoyemi J, Osiyemi A, Boyd A. Artificial intelligence for healthcare in Africa. Front Digit Health 2020;2:6 [FREE Full text] [doi: 10.3389/fdgth.2020.00006] [Medline: 34713019]
91. Data protection act. Kenya Law. 2019. URL: https://new.kenyalaw.org/akn/ke/act/2019/24/eng@2022-12-31 [accessed 2026-04-10]
92. Ireri S, Waiganjo P, Ochieng D, Kagiri M, Anindo M, Adoyo M, et al. An exploration of the successful scale-up of the electronic community health information system in Kenya. Oxf Open Digit Health 2025;3:oqaf020. [doi: 10.1093/oodh/oqaf020] [Medline: 40977915]
93. Waiyaiya E, Nyangena J, Nyotu R, Wamicwe J, Nyaga N, Wanyee S, et al. Digital health convergence workshop: the Kenyan experience. Oxf Open Digit Health 2024;2:oqae014. [doi: 10.1093/oodh/oqae014] [Medline: 40230980]
94. Wambua RN, Oduor C, Macharia J. Digital health model for equitable access to universal health coverage in Kenya. Discov Health Systems 2025;4(1):146. [doi: 10.1007/s44250-025-00314-7]
95. Brooke J. SUS—a quick and dirty usability scale. Usability Eval Ind 1996;189:4-7 [FREE Full text]
96. Breimaier HE, Heckemann B, Halfens RJG, Lohrmann C. The Consolidated Framework for Implementation Research (CFIR): a useful theoretical framework for guiding and evaluating a guideline implementation process in a hospital-based nursing practice. BMC Nurs 2015;14:43 [FREE Full text] [doi: 10.1186/s12912-015-0088-4] [Medline: 26269693]
97. Moullin JC, Dickson KS, Stadnick NA, Rabin B, Aarons GA. Systematic review of the Exploration, Preparation, Implementation, Sustainment (EPIS) framework. Implement Sci 2019;14(1):1 [FREE Full text] [doi: 10.1186/s13012-018-0842-6] [Medline: 30611302]

Abbreviations

AI: artificial intelligence

XSL•FO
RenderX

API: application programming interface
DHIS2: District Health Information System 2
EMR: electronic medical record
LMIC: low- and middle-income country
MAU: monthly active user
MKU: Mount Kenya University
NPS: Net Promoter Score
OpenMRS: Open Medical Record System
RBAC: role-based access control
REST: Representational State Transfer
SDH: Shamiri Digital Hub
SPP: Shamiri Provider Platform

Edited by A Kushniruk; submitted 16.Jun.2025; peer-reviewed by M Pulier, T Aung; comments to author 18.Nov.2025; revised version received 01.Mar.2026; accepted 19.Mar.2026; published 03.Jun.2026.

Please cite as:
Lilan S, Osborn TL, Mmbone W, Kasudi J, Otieno B, Etyang I, Tu J, Korley E, Wasanga C
Early Deployment of an Integrated Digital Platform (shamiriOS) for Scalable Youth Mental Health Service Delivery in Kenya: Development and Usability Study
JMIR Hum Factors 2026;13:e79107
URL: https://humanfactors.jmir.org/2026/1/e79107
doi:10.2196/79107
PMID:

©Shadrack Lilan, Tom L Osborn, Wendy Mmbone, Jean Kasudi, Benny Otieno, Ichami Etyang, Jianing Tu, Edmund Korley, Christine Wasanga. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 03.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Effect of a Personalized Mobile App on Glucose Control in Adults With Prediabetes and Type 2 Diabetes: Exploratory Pilot Randomized Controlled Trial

Elena Lalama[1], MSc; Marta Csanalosi[1], MSc; Stefan Kabisch[1,2], Dr med; Saskia Wilson-Barnes[3], PhD; Lazaros P Gymnopoulos[4], MSc; Kosmas Dimitropoulos[4], PhD; Konstantinos Rouskas[5], PhD; Argiriou Anagnostis[5], Prof Dr; Ioannis Oikonomidis[6], MS; Leontios Hadjileontiadis[7], PhD; Veronique Cornelissen[8], PhD; Maria Hassapidou[9], PhD; Ioannis Pagkalos[9], PhD; Sofia Balula Dias[10], PhD; Kathryn Hart[3], PhD; Andreas F H Pfeiffer[1], Dr med

[1]Department of Endocrinology and Metabolic Diseases, Charité - Universitätsmedizin Berlin, Berlin, Berlin, Germany

[2]German Center for Diabetes Research, Neuherberg, Bavaria, Germany

[3]School of Biosciences, Discipline: Nutrition, Food and Exercise Science, University of Surrey Faculty of Health and Medical Sciences, Surrey, United Kingdom

[4]Information Technologies Institute, Centre for Research and Technology Hellas, Thessaloniki, Greece

[5]Institute of Applied Biosciences, Centre for Research & Technology Hellas, Thessaloniki, Greece

[6]INTRASOFT International (Greece), Athens, Attica, Greece

[7]Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

[8]Department of Rehabilitation Sciences, KU Leuven, Leuven, Flanders, Belgium

[9]Department of Nutritional Sciences and Dietetics, International Hellenic University, Thessaloniki, Greece

[10]Faculdade de Motricidade Humana, University of Lisboa, Lisbon, Lisbon, Portugal

Corresponding Author:
Elena Lalama, MSc
Department of Endocrinology and Metabolic Diseases
Charité - Universitätsmedizin Berlin
Hindenburgdamm 30
Berlin, Berlin, 12203
Germany
Phone: 49 30450514422
Email: elena.lalama@charite.de

Abstract

Background: The incidence of type 2 diabetes (T2D) continues to increase, and the lack of individualized therapy strategies hinders patient engagement with and commitment to a healthy lifestyle. The PROTEIN project aimed to facilitate users in choosing healthy living, thereby improving their metabolism and T2D management.

Objective: This study aims to assess the efficacy of a personalized mobile app to achieve a 5% time in range (TIR) improvement over a 12-week intervention in adults with prediabetes or T2D.

Methods: We conducted an exploratory pilot randomized controlled trial with 21 individuals with T2D or prediabetes who used a continuous glucose monitoring system and the PROTEIN mobile app for personalized meals and exercise recommendations based on their glucose levels and physical activity.

Results: The TIR of the participants increased (P<.05; from 71.8%, SD 27.3% to 76%, SD 28.1%) with individual use of the PROTEIN app but did not achieve a 5% improvement overall; however, given the exploratory design and small sample size, this finding should be interpreted with caution. Glycated hemoglobin, fasting blood glucose, and body weight did not fluctuate throughout the 12-week intervention. The dropout rate was high, and the average duration of use of the PROTEIN app was 42 (range 5-84) days.

Conclusions: Our results showed a modest increase in TIR with the use of the PROTEIN app; however, considering the exploratory design and small sample size, this finding should be interpreted as preliminary. Integrating wearables and automated personalization for well-being is an innovative approach that must keep pace with the accelerated development of ever-evolving technologies. The COVID-19 pandemic was a major obstacle to recruitment in our clinical trial.

XSL•FO
RenderX

Trial Registration: ClinicalTrials.gov NCT05951140; https://clinicaltrials.gov/study/NCT05951140

(JMIR Hum Factors 2026;13:e87692) doi:10.2196/87692

KEYWORDS

type 2 diabetes; personalized nutrition; glucose management; time in range; CGM; continuous glucose monitoring

Introduction

Identifying optimal diets for individuals with type 2 diabetes (T2D) and prediabetes remains challenging [1]. The World Health Organization defines mobile health (mHealth) as medical and public health practices supported by mobile devices, which have become increasingly prevalent [2]. Together, these developments highlight the growing potential of digital tools to support personalized diabetes management.

Advances in artificial intelligence (AI) now enable the processing of large datasets to deliver personalized nutrition advice that aligns with individual values and preferences, potentially supporting sustained patient engagement in healthy dietary choices [3,4]. However, despite the constant and rapid evolution of diabetes-related mobile apps, few offer adaptive, evidence-based nutrition guidance for patients with T2D in Germany [5-7].

Given the significant interindividual variability in postprandial glucose responses, generic recommendations are insufficient. Integrating data from continuous glucose monitoring (CGM) and user preferences allows for more tailored interventions, using algorithms powered by machine learning and AI to enhance diabetes self-management [6,8-11].

Within the European H2020 "PROTEIN: PeRsOnalized nutriTion for hEalthy living" project (2019-2022), we aimed to develop and evaluate a personalized nutrition and exercise mobile app to support daily T2D management [12]. The PROTEIN project involved 20 partners from 11 European countries, focusing on health promotion through multidisciplinary collaboration. Our exploratory pilot study specifically assessed whether the PROTEIN app could improve dietary compliance and glucose management in T2D, using time in range (TIR) as the primary outcome - a metric not commonly used as a primary endpoint in similar randomized controlled trials (RCTs) [7]. TIR is the proportion of time that a person spends within a target range of glucose levels, usually between 70 and 180 mg/dL, as predetermined by a health practitioner or the person with diabetes [13].

Following Germany's 2019 digital health reform, only approved apps can be prescribed for diabetes management, ensuring patient safety and reliability [14]. Only 6 of 59 mHealth apps have been approved for people with diabetes [15]. The PROTEIN trial provides prospective, controlled, long-term evidence on AI-driven mobile apps that deliver personalized recommendations using CGM data, user interaction, and preferences. This adaptive AI approach addressing engagement and patient engagement challenges is novel in digital therapeutics.

To our knowledge, this adaptive approach has not been reported in previous studies [16-19]. Therefore, our exploratory pilot study aimed to automatically customize meal plans and physical activity (PA) according to user phenotype and food preferences by integrating CGM data of patients with T2D. The primary objective was to increase the TIR of the study participants by 5% while improving eating behaviors.

Expanding upon existing research findings, we aimed to address the following questions:

1. Does TIR increase when using the PROTEIN app?
2. Are there associated changes in weight?

Methods

Study Design

We conducted a 12-week RCT in Germany (June-December 2022) to evaluate the PROTEIN app's impact on glucose management in T2D.

The PROTEIN System

Developed and tested [20-24] by a European consortium, the PROTEIN system comprised:

1. Nutrium dashboard: used by nutritionists to manage participants and create meal plans using country-specific food composition datasets [25].
2. PROTEIN app: enabled participants to view, plan, and log meals and activities, and interact with personalized recommendations.
3. User dashboard: provided participants with an alternative visualization tool through a web browser.

The AI-driven advisor included a food and activity recommender system, a reasoning-based decision support system [8] built on the Nutrition & Activity Ontology for Healthy Living, which encoded the rules and relations used by the AI advisor [26], and generated individualized meal and activity plans every two weeks. These were based on the user profile, preferences, CGM data, and aligned with the "knowledge tracker," a framework for the meal and PA suggestions [27] that was integrated into the application. Nutrition requirements followed European guidelines [28].

Intervention

We conducted a 12-week intervention. The participants used the PROTEIN app and the CGM Freestyle Libre (Abbott Diabetes Care, Wiesbaden, Germany) to collect data to enable further personalization. We randomized and allocated the participants into two groups after successful screening:

1. Start-group: participants used the activity tracker, CGM, and PROTEIN app for 12 weeks, followed by a 6-week

XSL•FO
RenderX

period without the PROTEIN app, but where the wearables continued to be used.

2. Wait-group: participants only used the activity tracker and CGM for 6 weeks, followed by a 12-week period of using the PROTEIN app as well.

This study design, as shown in Figure 1, allowed us to clearly understand the influence of the PROTEIN app. The randomization sequence was computer-generated (Microsoft Excel 2019, Microsoft Corporation) and was performed on-site by the study team. Stratification considered sex, age (cut-off 60 years), and BMI (cut-off 30kg/m^2) at screening. Given the nature of the trial, blinding was not possible for either participants or study personnel. After generating the randomization sequence, participants were assigned on-site by the study team, without formal allocation concealment.

Figure 1. Study design illustrating the start-group (blue, top panel) and the wait-group (green, bottom panel).



The participants allocated to the start-group attended their first clinical visit, where baseline clinical and laboratory assessments, including blood count, lipid panel, liver profile, uric acid, C-reactive protein, fasting glucose, hemoglobin A_{1c} (HbA_{1c}), ankle-brachial index, and body composition, were collected after an overnight fast.

On a typical day, the participants would open the app to review the recommended meal and activity plans, confirm or delete items, rate meals on a 1- to 5-star scale, and add alternative foods or activities.

Participants and Recruitment

Participants were recruited via hospital and newspaper advertisements, outreach to physicians, and online postings. Eligibility criteria are given in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Prediabetes or T2D (HbA_{1c} ≥5.7%, fasting glucose ≥100 mg/dL, or 2-hour glucose ≥140 mg/dL) [29]
- Age >18 years
- BMI 20-45 kg/m^2

Exclusion criteria

- Type 1 diabetes
- Severe metabolic conditions
- Insulin, sulfonylureas, and glucagon-like peptide-1 (GLP-1) agonists
- No Android smartphone
- Use of continuous glucose monitoring (CGM) or activity tracking apps in the last 3 months

Some eligibility criteria were further specified during study implementation to ensure participant safety. In particular, individuals using sulfonylureas and glucagon-like peptide-1 receptor agonists were excluded, although these criteria were not explicitly detailed in the trial registry. The BMI inclusion range (20-45 kg/m^2) was applied consistently during recruitment. After telephone screening, 71 candidates were invited for on-site screening.

The PROTEIN trial conducted in Germany was retrospectively registered after completion of data collection at ClinicalTrials.gov: registration number NCT05951140 [30];

XSL•FO
RenderX

that said, the protocol and outcomes had been defined prior to data analysis.

Meal Plan Personalization

Meal plans were developed using the PROTEIN knowledge tracker, tailored for prediabetes or T2D as shown in Table 1, and based on participant profile (age, sex, diagnosis, weight, dietary history, goals, allergies, intolerances, and religious preferences). Plans were updated biweekly, with participant feedback and CGM data guiding further personalization.

Table 1. Dietary framework of the knowledge tracker developed for prediabetes and type 2 diabetes.

Nutrient recommendations	Male	Female
Energy intake (kcal)	2100 (300)	2100 (300)
Salt (g)	<5	<5
CHO[a] (%EI[b])	45 (10)	45 (10)
Sugar (%EI)	<10	<10
Protein (g/kg/BW[c])	0.8-1.4	0.8-1.4
Fat (%EI)	30 (10)	30 (10)
Fiber (g/day)	30-45	30-45
Fruit (portions)	2	2
Vegetables (portions)	≥3	≥3

[a]CHO: carbohydrates.

[b]EI: energy intake.

[c]BW: body weight.

Participants interacted with the app by confirming, deleting, or rating meals (1-5 stars) and activities. Customization increased as participants engaged with the app, especially during the initial 2 weeks, enabling the AI to learn user preferences. Engagement metrics were monitored to assess correlations with health behavior changes.

Glucose Data

Each participant used a Freestyle Libre CGM system, with data uploaded weekly to inform recommendations. A TIR ≥70% (equivalent to HbA_{1c} ≤6.7%) was set as the target, following the international consensus on CGM [31].

If TIR dropped below 70%, the app automatically reduced daily carbohydrate recommendations by 10% for the following week. This allowed a stepwise personalization within the ranges of evidence-based dietary recommendations for the dietary management of diabetes [32,33]. Foods, apart from vegetables and pulses, causing high glucose excursions (≥140 mg/dL for 2-4 hours postprandial, on ≥2 occasions), activated a push notification directly to the user and were excluded from the meal plans the following week.

The data extracted from the CGM was only used if at least 70% of the glucose data was available per day. We based this modified approach on the recommendations for interpretations of TIR [31]. We compared the average TIR of the user during the time they were actively using the app to the average TIR during their planned waiting time in the study protocol. This approach was used to account for the fact that most participants did not use the application for the planned 12 weeks.

Sample Size and Power Calculation

The study aimed for 300 participants to detect a 5% TIR difference with 80% power (β=0.20) and α=0.05, enabling detection of medium to small effects (Cohen d=0.348) [34]. Due to recruitment challenges, only 27 participants were enrolled, and 21 were included in the final analysis. The study is substantially underpowered for our primary aim, and the analysis that focused on periods of active app use should be interpreted as exploratory.

Statistical Analysis

Initially, an as-treated analysis was planned for all completers. Due to low adherence and high dropout, the originally planned RCT analysis was not feasible. Therefore, the analyses presented are exploratory and were limited to periods of active app use compared with nonuse days. This approach introduces selection bias and does not preserve the randomized design. No intention-to-treat analysis was possible due to the limited sample size and missing data. Subgroup analyses (start-group vs wait-group) and analyses restricted to periods of active app use were not prespecified and should be interpreted as exploratory. Active app use days were defined as days with at least one logged interaction with the PROTEIN app and valid CGM data. Noncompliance was defined as <30% app use (<25 days) [35]. Statistical analyses were performed using SPSS 26.0 and Prism 8. Data are presented as mean (SD) in tables or mean (SEM) in figures. Paired 2-tailed t tests or nonparametric equivalents were used as appropriate, with significance set at P<.05. Outliers (>3× IQR) were excluded.

Due to early termination of recruitment, high attrition, and deviation from the planned randomized analysis, this study should be considered an exploratory pilot study. All analyses

XSL•FO
RenderX

are hypothesis-generating and do not provide confirmatory evidence of effectiveness.

Registered Outcomes and Reporting

To ensure transparency between the prespecified trial registration and the analyses reported in the results section, all registered outcomes were systematically compared with the outcomes analyzed. A summary of registered primary and secondary outcomes, their reporting status in this manuscript, and reasons for nonreporting where applicable, is provided in Table 2.

Table 2. Comparison of registered outcomes and outcomes reported in this manuscript.

Outcome (as registered at ClinicalTrials.gov)	Type	Reported in this manuscript	Reason for nonreporting (if applicable)
Time in range (TIR)	Primary	Yes	—[a]
Improve eating behavior toward healthier choices	Primary	No	Insufficient and inconsistent dietary data and absence of a predefined validated combined measure
Adherence to the dietary recommendations of the app	Secondary	Exploratory and descriptive only	Exploratory indicators (eg, app engagement, meal ratings) reported
Glycemic metabolism (HbA_{1c}[b], fasting glucose)	Secondary	Yes	—
Sedentary time	Secondary	No	Data are incomplete and not consistently recorded across participants
Energy intake	Secondary	Partially	Reported descriptively (Table 7); not analyzed as a formal outcome due to limited sample size
Microbiome	Secondary	No	Insufficient samples available for robust analysis
Polygenic risk score	Secondary	No	Not available for the analyzed cohort

[a]Not applicable.

[b]HbA_{1c}: glycated hemoglobin A_{1c}

In particular, the primary outcome “healthier eating behavior” could not be formally analyzed due to insufficient and inconsistent dietary data.

Ethical Considerations

The pilot study PROTEIN was conducted at the University Hospital Charité. The ethics committee of the Charité University Hospital Berlin approved all trial procedures and amendments (approval number EA4/110/20 on September 11, 2020). We retrospectively registered the trial at ClinicalTrials.gov (registration number NCT05951140) after completion of data collection due to administrative delays; we defined the protocol and outcomes prior to data analysis. Written informed consent was obtained from all participants.

CONSORT Guidelines

This randomized controlled trial was conducted in accordance with the CONSORT (Consolidated Standards of Reporting Trials) 2025 guidelines for reporting RCTs (Multimedia Appendix 1) [36].

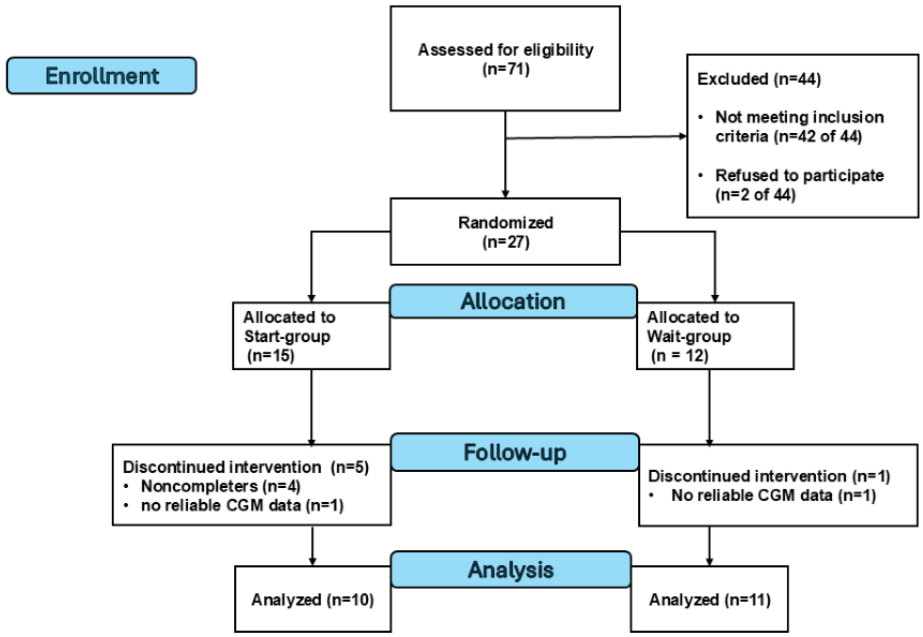
Results

Demographic Characteristics

Overall, 27 participants (n=13 male, n=14 female) were enrolled in the study. Twelve participants (44%) stopped using the PROTEIN app after 12 days, 11 of whom continued to use the wearables. In total, 21 participants had reliable CGM data and were included in the analysis (12 female and 9 male) (Figure 2). The baseline characteristics of the participants are shown in Table 3.

XSL•FO
RenderX

Figure 2. Flowchart depicting the progression of study participants throughout the pilot study. CGM: continuous glucose monitoring.



XSL•FO
RenderX

Table 3. Baseline characteristics of the analyzed study participants.

Characteristics	Baseline (n=21)
Age (y), mean (SD)	57 (10.7)
Sex (female/male), n	12/9
Body composition, mean (SD)	
Body weight (kg)	90.5 (19.2)
BMI (kg/m^2)	30.5 (6.6)
Waist circumference (cm)	103.5 (12.5)
Hip circumference (cm)	111.5 (14.8)
WHR[a]	0.93 (0.08)
Fat mass (kg)	26.5 (13.4)
Fat mass (%)	28.4 (10.2)
Blood pressure (mmHg), mean (SD)	
Systolic	137 (12)
Diastolic	88 (8)
HbA_{1c}[b] (%)	6.4 (0.9)
Fasting blood glucose (mg/dl)	120.4 (34.3)
Liver transaminases (U/l), mean (SD)	
ALT[c] (GPT[d])	28 (13)
γ-GT	50 (65)
Lipid concentrations (mg/dl)	
Total cholesterol	195 (54)
Triglyceride	143 (87)
LDL[e]-cholesterol	121 (47)
Prescribed oral antidiabetic medication	
0	15
1	2
2	4

[a]WHR: waist to hip ratio.

[b]HbA_{1c}: glycated hemoglobin A_{1c}.

[c]ALT: Alanine aminotransferase.

[d]GPT: glutamate pyruvate alanine aminotransferase.

[e]LDL: low-density lipoprotein.

CGM Data

We compared the TIR of the participants during the waiting phase (without the use of the PROTEIN app) to the intervention (with the use of the PROTEIN app). TIR increased significantly (*P*=.04), rising from 71.8% (SD 27.3%) to 76% (SD 28.1%). This represented a 4.2% (9%). Improvement, not meeting the 5% target set.

We then examined potential differences between the wait- and the start-group. In the start-group (n=10), the TIR decreased (not statistically significant) from 68.9 (SD 32.7%) to 64.9 (SD 31.1%) (*P*=.14) once they stopped using the application, whereas the TIR of the wait-group (n=11) increased from 78.0 (SD 23%) to 82.4 (SD 22.9%) (*P*=.10) once they started using the application. Having considered the positive effect on the TIR, we investigated whether starting to use the PROTEIN app first (start: 4, SD 9.4 vs wait: 4.3, SD 8) yielded a stronger change in the TIR. No significant difference in the change of the TIR between the 2 groups was seen (*P*=.94).

Routine Glucose Parameters

HbA_{1c} and fasting blood glucose did not change significantly during the study, as shown in Tables 4 and 5. Values remained comparable across study visits within both groups. No statistically significant differences were observed between the intervention and waiting phases.

XSL•FO
RenderX

Table 4. Glycated hemoglobin A1c separated by group and study phase.

Group phase[a]	Visit	Participants, n	HbA_{1c}[b] (%), mean (SD)	*P* value
Start group				
Intervention				.11
	V1	8	6.8 (1)	
	V2	8	6.5 (0.6)	
Waiting				.47
	V2	5	6.3 (0.7)	
	V3	5	6.4 (0.9)	
Both				.22
	V1	6	6.8 (1.1)	
	V3	6	6.6 (0.9)	
Wait group				
Waiting				.26
	V1	10	6.2 (0.9)	
	V2	10	6.1 (0.8)	
Intervention				.36
	V2	4	6 (0.7)	
	V3	4	6.1 (0.6)	
Both				.41
	V1	4	6 (0.6)	
	V3	4	6.1 (0.6)	

[a]HbA_{1c} changes during the different phases (intervention and waiting) of the pilot. “Intervention” represents the period of time where the participants use the app; “Waiting” represents the waiting phase without the use of the app; and “both” represents both phases together. P represents *P* values from *t* test for analyzing differences within each group.

[b]HbA^{1c}: glycated hemoglobin A_{1c}.

XSL•FO
RenderX

Table 5. Fasting blood glucose separated by group and study phase.

Group phase[a]	Visit	Participants, n	Fasting blood glucose (mg/dL), Mean (SD)	*P* value
Start group				
Intervention				.31
	V1	7	134.7 (41.8)	
	V2	7	122.9 (24.7)	
Waiting				.50
	V2	5	115.4 (26.1)	
	V3	5	125.2 (48.2)	
Both				.72
	V1	5	128.8 (39.2)	
	V3	5	131.6 (48.8)	
Wait group				
Waiting				.72
	V1	10	109.5 (29.8)	
	V2	10	104.5 (19.7)	
Intervention				.07
	V2	4	101.3 (17.3)	
	V3	4	107.5 (17.7)	
Both				.07
	V1	4	101.3 (17.5)	
	V3	4	107.5 (17.7)	

Body Weight Change

We analyzed the wait-group and start-group separately. Body weight remained stable within and between the wait-group and start-group as shown in Table 6. Mean values were similar across study visits in both groups. No statistically significant changes were observed over time.

XSL•FO
RenderX

Table 6. Body weight separated by group and study phase.

Group phase[a]	Visit	Participants, n	Weight (kg), mean (SD)	*P* value
Start group				
Intervention				.92
	V1	8	96.5 (18)	
	V2	8	96.6 (17.3)	
Waiting				.14
	V2	5	95.7 (18.9)	
	V3	5	94.9 (19.3)	
Both				.11
	V1	6	93.5 (21.7)	
	V3	6	91 (19.7)	
Wait group				
Waiting				.65
	V1	11	87.3 (20.3)	
	V2	11	86.8 (21.3)	
Intervention				.24
	V2	5	91.2 (26.7)	
	V3	5	92 (26.7)	
Both				.87
	V1	5	92.1 (25.9)	
	V3	5	92 (26.7)	

Use of the PROTEIN App

The intended duration of the PROTEIN app usage was 84 days (12 weeks), but only 2 participants completed the full period. To account for this, we assessed the time period between first and last use of the PROTEIN app, irrespective of intermittent pauses. Participants' average app usage was 42 days (range 5-84 days). Contrary to expectations, there was no significant correlation between days of app usage and change in TIR (r= –0.224; *P*=.33). No adverse events or study-related harms were observed during the study period.

Personalization

The process of personalization is described in detail in the PROTEIN AI-Advisor [8].

To test the accuracy of the recommended meal plans, the data from the nutrition and activity plans were extracted and compared to the knowledge tracker, as seen in Table 7. We did not detect any high glucose excursions caused by pulses or vegetables, nor exclusions of specific foods.

Table 7. Nutrient recommendations via the PROTEIN app compared to the knowledge tracker.

Nutrient recommendations	Values, mean (SD; n=21)	Knowledge tracker (male/female)
Energy intake (kcal)	2048 (376)	2100 (300)/1750 (250)
Protein (g/kg/BW)	1.2 (0.2)	0.8-1.4
Carbohydrates (%EI)	49 (5)	45 (10)
Fat (%EI)	31 (4)	30 (10)
Fiber (g/d)	36 (6)	30-45
Fruit (portions/day)	3 (1)	2
Vegetables (portions/day)	5 (2)	≥3

Overall, the average star ratings ranged from 3 out of 5 stars for lunch and dinner to 3,5 out of 5 stars for breakfast, afternoon snack, and supper, and 4 out of 5 stars for morning snack.

In addition, we investigated the total variety of meal composition that was yielded for 17 participants; 159 unique options for breakfast were recommended, 114 for morning snacks, 145 for lunches, 112 for afternoon snacks, 123 for dinner, and 27 for

XSL•FO
RenderX

suppers. Furthermore, 6 (29%) participants reported trying new meals that they had not thought of before as a result of using the PROTEIN app.

Discussion

Our study addressed two research questions: whether the PROTEIN app increases TIR and whether it affects body weight. The results indicate a modest improvement in TIR (4.2%) during app use, with no change in weight. In summary, our data indicated that the TIR improved with the use of the PROTEIN app for only an average of 12 days, but this fell short of the expected 5% improvement and of the expected 12-week usage period. Unexpectedly, we did not find a significant correlation between the duration of app use and the improvement of the TIR, which could be due to the small sample size and narrow range of usage durations. Despite the high rates of noncompliance in the intervention phase, the TIR improved in parallel with the use of the mobile nutrition app, suggesting a potential signal that a tailored recommendation approach delivered with wearables through a mobile app may be associated with small increases in TIR and may encourage greater concordance with dietary recommendations. However, this observation is exploratory and cannot be distinguished from selection bias or confounding, and any effect on dietary concordance remains uncertain. Our study, therefore, also illustrates the need to balance technological capabilities with user experience since using an app with many features and wearables is time-consuming, contributing to participant fatigue and dropout.

Overall, participants who took part in the intervention with valid CGM data (n=21) had an improvement of 4.2% in their TIR while they were using the PROTEIN app, suggesting that even though they did not follow the recommendations every day, there seems to be a degree of agreement and engagement while using the PROTEIN system [35,37]. It is noteworthy that the outcome relies on the adequate resources of the participants and may be influenced by different factors, including family network, profession, and socio-economic status.

Although the 4.2% increase in TIR reached statistical significance, its clinical relevance remains uncertain, because of the high baseline TIR (mean >70%) and the lack of change in HbA_{1c}, fasting glucose, and weight. International consensus recommends achieving and maintaining TIR ≥70% and states that a 5% change in TIR is strongly associated with a 0.5% decrease in HbA_{1c}, which is a clinically significant reduction. We therefore interpret the observed TIR change as a modest signal of potential benefit that requires confirmation in larger, adequately powered studies.

We anticipated that participants would use the PROTEIN app for a full 12 weeks; however, this was not achieved by the majority. This was largely due to the app being perceived as not user-friendly and occasionally too time-consuming. Defining dropouts was a difficult task, since several studies defined this differently. We based our definition of adherence on the days the app was used, namely, at least 30% of the planned time. According to this definition, the dropout rate in the study (43%) is consistent with the work reported by Meyerowitz-Katz et al [38].

Accordingly, after analyzing the differences between the two groups, we observed that the participants' TIRs improved regardless of whether they started with the intervention or the waiting phase. This may suggest that wearables alone are not sufficient to improve glucose management, but that extra feedback, in this case in the shape of an app, led participants toward better glucose management, although this cannot be determined definitively from the current data. The TIR of the start-group did not decrease significantly after they stopped using the PROTEIN app, suggesting there is a legacy effect for the users. The participants kept using the sensors as planned (12 weeks during intervention and 6 weeks in the waiting phase), which allowed us to compare the two periods of time. From these data, we observed patterns consistent with improved glucose management during the PROTEIN app use; however, causality cannot be inferred due to the exploratory design and potential selection bias. A dietary protocol before intervention could be added in further research to provide a better understanding of dietary change.

Furthermore, prior studies have shown that CGM is an important tool for the management of T2D [13,39,40]. When we analyzed our participants according to the assigned group, we can see that using wearables without guidance on how to effectively use them did not cause a change in the glucose levels. Interestingly, the participants who used the PROTEIN app right away (start-group) had a decrease in the TIR once they stopped using it, but this decrease was not statistically significant. This suggests that 12 weeks may not be enough to maintain a health-supporting behavior for a longer period. As no significant changes in weight were seen over the study period, improvements in TIR appear to be driven by the consumption of more adequate meals, independent of the group they were randomized into. Therefore, we cannot determine whether the PROTEIN app influences their caloric intake, but the findings are consistent with changes in food selection.

Overall, the PROTEIN application generally provided suitable meals for individuals with T2D and prediabetes. However, on a few occasions, the recommended protein amounts fell short when compared to the knowledge tracker. The first 2 weeks during which the participants were expected to train the app proved to be excessively time-consuming, leading to a decline in the compliance of our participants. Additional factors, including managing multiple devices and integrating the protocol into daily routines, were also identified as reasons for disengagement. The mobile app was programmed to change the meal plan for the whole week once the subject skipped a meal and added a new one instead. Moreover, the PROTEIN app sometimes failed to "learn" or adapt after two weeks, despite participants having the same breakfast daily.

The PROTEIN AI-advisor adjusted carbohydrate targets in response to low TIR and was programmed to exclude foods associated with repeated high excursions. We could not detect the frequency and magnitude of these adaptations in practice. Moreover, participants reported cases in which the system did not sufficiently adapt to repeated meal patterns, indicating that

XSL•FO
RenderX

the current personalization was only partially realized and requires further optimization.

It is important to acknowledge that 21 participants had reliable CGM data, making the trial substantially underpowered and exploratory relative to the original sample size calculation with 300 participants. All suggestions of effect should be interpreted as exploratory signals rather than solid evidence.

Furthermore, limiting analyses to periods of active app use introduces substantial selection bias toward more engaged users, likely underestimates variability in real-world use, and may overestimate potential effects.

Several limitations apply to our exploratory pilot study. Recruitment and retention were challenging, exacerbated by the COVID-19 pandemic and technical barriers (eg, Android-only compatibility). These factors limited the sample size and statistical power, highlighting the need for larger, more inclusive studies. Several prespecified outcomes were registered in the clinical trial registry but are not reported in this manuscript because of the limited sample size and incomplete data collection. A detailed comparison of registered and reported outcomes is provided in Table 2. Even though we saw a moderate difference in the TIRs while using the PROTEIN app, we were unable to assess participants' TIR prior to using the wearables and/ or using the application, which could provide further robust evidence as to the health behavior changes.

Therefore, we cannot fully exclude the effects influencing intra-participant TIR patterns, as preintervention blind CGM data periods were not available.

Despite these limitations, the study provides preliminary indications of the feasibility and potential of combining mHealth, AI, and CGM data to support individualized diabetes care; nevertheless, it must be recognized that our own evidence indicates the intervention does not feasibly support increased TIR, except for a short period and for a small proportion of app users. This emphasizes important limitations regarding the broader clinical use and sustainability of the approach.

Our exploratory pilot study adds novel data on the feasibility, adherence patterns, and preliminary glycemic impact of such an AI-driven system under real-world conditions.

Future research should focus on improving user engagement, expanding compatibility, and evaluating long-term clinical outcomes in larger, more diverse populations.

Conclusions

The level of personalization offered by the PROTEIN app needs to be enhanced [9], which could help reduce the time required for the app to learn and adapt. For example, allowing users to select in advance their favorite ingredients or meals could streamline the learning process. Ideally, greater integration of image recognition features, allowing participants to take and upload pictures of their meals rather than manually entering ingredients, and the ability to automatically upload and correlate CGM data with food intake would significantly improve the user and researcher experience. These conclusions are based on this combination of published evidence and participant input.

Considering the strong limitation in sample size and engagement relative to the original protocol, our study results should be seen as an exploratory trial that primarily provides information regarding feasibility, engagement, and potential effects.

Finally, the PROTEIN app was associated with a modest increase in TIR in adults with prediabetes or T2D; however, this exploratory finding does not establish effectiveness and requires confirmation in adequately powered trials, suggesting a possible role for AI-driven, personalized mHealth solutions for diabetes management. While the primary endpoint was not fully achieved, the study supports further development and evaluation of adaptive digital interventions designed to enhance personalized engagement with dietary recommendations and improve glycemic control.

Acknowledgments

Generative artificial intelligence tools were not used in the design, conduct, analysis, or drafting of this manuscript. Copilot was used for proofreading.

Data Availability

The data that support the findings of this study, trial protocol, and statistical analysis are not publicly available to preserve individuals' privacy under the European General Data Protection Regulation but are available from the corresponding author (elena.lalama@charite.de) upon reasonable request.

Funding

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 817732 (PROTEIN: PeRsOnalized nutriTion for hEalthy living).

Authors' Contributions

Conceptualization, data curation, formal analysis, investigation, methodology, validation, writing – original draft, and writing – review and editing: EL, MC, SK, AFHP
Conceptualization, data curation, review, and editing: SWB, KH, SBD, VC
Methodology, validation, and conceptualization: LPG, KD, KR, AA, YO, LH, JMB, RL, MH, IP

XSL•FO
RenderX

Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Multimedia Appendix 1
CONSORT checklist.
[DOCX File , 33 KB - humanfactors_v13i1e87692_app1.docx]

References

1. Davies MJ, Aroda VR, Collins BS, Gabbay RA, Green J, Maruthur NM, et al. Management of hyperglycaemia in type 2 diabetes, 2022. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetologia 2022 Dec;65(12):1925-1966 [FREE Full text] [doi: 10.1007/s00125-022-05787-2] [Medline: 36151309]
2. Kay M, Santos J, Takane M. mHealth: New horizons for health through mobile technologies. In: Global Observatory for eHealth series. Geneva: World Health Organisation; 2011.
3. Hyun MK, Lee JW, Ko S, Hwang JS. Improving glycemic control in type 2 diabetes using mobile applications and e-coaching: a mixed treatment comparison network meta-analysis. J Diabetes Sci Technol 2022;16(5):1239-1252 [FREE Full text] [doi: 10.1177/19322968211010153] [Medline: 33980055]
4. Eberle C, Löhnert M, Stichling S. Effectiveness of disease-specific mhealth apps in patients with diabetes mellitus: scoping review. JMIR Mhealth Uhealth 2021;9(2):e23477 [FREE Full text] [doi: 10.2196/23477] [Medline: 33587045]
5. Essén A, Stern AD, Haase CB, Car J, Greaves F, Paparova D, et al. Health app policy: international comparison of nine countries' approaches. NPJ Digit Med 2022 Mar 18;5(1):31. [doi: 10.1038/s41746-022-00573-1] [Medline: 35304561]
6. Fleming GA, Petrie JR, Bergenstal RM, Holl RW, Peters AL, Heinemann L. Diabetes digital app technology: benefits, challenges, and recommendations. A consensus report by the European Association for the Study of Diabetes (EASD) and the American Diabetes Association (ADA) diabetes technology working group. Diabetologia 2020;63(2):229-241. [doi: 10.1007/s00125-019-05034-1] [Medline: 31802144]
7. Corsica JA, Kelly MC, Bradley LE, Konsor MM, Wilson EJ, Quinones IC, et al. Mobile apps for diabetes self-management: an updated review of app features and effectiveness. J Behav Med 2025 Feb;48(1):137-148. [doi: 10.1007/s10865-024-00525-y] [Medline: 39489882]
8. Stefanidis K, Tsatsou D, Konstantinidis D, Gymnopoulos L, Daras P, Wilson-Barnes S, et al. PROTEIN AI Advisor: a Knowledge-Based Recommendation Framework Using Expert-Validated Meals for Healthy Diets. Nutrients 2022 Oct 21;14(20):4435 [FREE Full text] [doi: 10.3390/nu14204435] [Medline: 36297118]
9. Dias SB, Oikonomidis Y, Diniz JA, Baptista F, Carnide F, Bensenousi A, et al. Users' perspective on the AI-based smartphone PROTEIN app for personalized nutrition and healthy living: a modified technology acceptance model (mTAM) approach. Front Nutr 2022;9:898031 [FREE Full text] [doi: 10.3389/fnut.2022.898031] [Medline: 35879982]
10. Ji C, Jiang T, Liu L, Zhang J, You L. Continuous glucose monitoring combined with artificial intelligence: redefining the pathway for prediabetes management. Front Endocrinol (Lausanne) 2025;16:1571362. [doi: 10.3389/fendo.2025.1571362] [Medline: 40491592]
11. Liu X, Zhang J. Continuous glucose monitoring: a transformative approach to the detection of prediabetes. J Multidiscip Healthc 2024;17:5513-5519 [FREE Full text] [doi: 10.2147/JMDH.S493128] [Medline: 39600717]
12. Personalized nutrition for healthy living. European Commission. URL: https://cordis.europa.eu/project/id/817732 [accessed 2026-05-23]
13. Ehrhardt N, Al Zaghal E. Continuous glucose monitoring as a behavior modification tool. Clin Diabetes 2020;38(2):126-131 [FREE Full text] [doi: 10.2337/cd19-0037] [Medline: 32327884]
14. Digitale gesundheitsanwendungen (DiGA) [Web page in German]. Bundesministerium für Gesundheit. URL: https://www.bundesgesundheitsministerium.de/themen/krankenversicherung/online-ratgeber-krankenversicherung/arznei-heil-und-hilfsmittel/digitale-gesundheitsanwendungen [accessed 2026-05-23]
15. DiGA-verzeichnis [Web page in German]. Bundesinstitut für Arzneimittel und Medizinprodukte. URL: https://diga.bfarm.de/de [accessed 2026-05-23]
16. Akulova M, Nesterova A, Isaacs D, Molodkin E, Ushakov M. Effect of mobile-based application usage on time in range and time above range in patients with diabetes mellitus: a pilot cohort retrospective study. MRAJ 2023;11(7.1). [doi: 10.18103/mra.v11i7.1.4047]
17. Dehghani Zahedani A, Shariat Torbaghan S, Rahili S, Karlin K, Scilley D, Thakkar R, et al. Improvement in glucose regulation using a digital tracker and continuous glucose monitoring in healthy adults and those with type 2 diabetes. Diabetes Ther 2021 Jul;12(7):1871-1886 [FREE Full text] [doi: 10.1007/s13300-021-01081-3] [Medline: 34047962]
18. Verma R, Bhardwaj S, Lathia T, Kalra S, Ranadive R, Tanna S, et al. Personalized glycemic response led digital therapeutics program improves time in range in a period of 14 days. Int J Diabetes Dev Ctries 2023 Jun;43(3):425-432 [FREE Full text] [doi: 10.1007/s13410-022-01111-1] [Medline: 35892065]

XSL•FO
RenderX

19. Kitazawa M, Takeda Y, Hatta M, Horikawa C, Sato T, Osawa T, et al. Lifestyle intervention with smartphone app and isCGM for people at high risk of type 2 diabetes: randomized trial. J Clin Endocrinol Metab 2024 Mar 15;109(4):1060-1070 [FREE Full text] [doi: 10.1210/clinem/dgad639] [Medline: 37931069]
20. Patra E, Kokkinopoulou A, Wilson-Barnes S, Hart K, Gymnopoulos LP, Tsatsou D, et al. Personal goals, user engagement, and meal adherence within a personalised AI-based mobile application for nutrition and physical activity. Life (Basel) 2024 Sep 27;14(10):1238 [FREE Full text] [doi: 10.3390/life14101238] [Medline: 39459538]
21. Hart K, Wilson-Barnes S, Lanham-New S, Gymnopoulos L, Dimitropoulos K, Stefanidis K, et al. Can a mobile application deliver appropriate, acceptable and usable personalised nutrition advice to UK adults. Proc Nutr Soc 2024 Jan 08;82(OCE5). [doi: 10.1017/s0029665123004135]
22. Hart KH, Wilson-Barnes S, Stefanidis K, Tsatsou D, Gymnopoulos L, Dimitropoulos K, et al. The suitability of dietary recommendations suggested By artificial intelligence technology via a novel personalised nutrition mobile application. Proc Nutr Soc 2022 Feb 08;81(OCE1):489. [doi: 10.1017/S0029665122000374]
23. Decorte E. A pilot trial guiding development and investigating feasibility and usability as well as preliminary effectiveness of the PROTEIN application: Personalized nutrition for healthy living. Eur J Prev Cardiol 2023;30(Supplement_1):zwad125.220. [doi: 10.1093/eurjpc/zwad125.220]
24. Wilson-Barnes SL. The evaluation of a personalised nutrition and physical activity tool to facilitate lifestyle changes for adults with poor-quality diets. Proc Nutr Soc 2023;82(OCE1):E42. [doi: 10.26226/m.63a07d2a410a2800136d1809]
25. Nutrium. 2020. URL: https://nutrium.com/en [accessed 2026-05-23]
26. Tsatsou D, Lalama E, Wilson-Barnes SL, Hart K, Cornelissen V, Buys R, et al. NAct: the nutrition and activity ontology for healthy living. Frontiers Artif Intelligence Applications 2021:129-143. [doi: 10.3233/faia210377]
27. Wilson‐Barnes S, Gymnopoulos LP, Dimitropoulos K, Solachidis V, Rouskas K, Russell D, et al. Personalised nutrition for healthy kiving: the PROTEIN project. Nutrition Bulletin 2021;46(1):77-87. [doi: 10.1111/nbu.12482]
28. European Food Safety Authority (EFSA). Dietary reference values for nutrients summary report. EFSA 2017 Dec;14(12):e15121. [doi: 10.2903/sp.efsa.2017.e15121]
29. ElSayed NA, Aleppo G, Aroda VR, Bannuru RR, Brown FM, Bruemmer D, et al. 2. Classification and diagnosis of diabetes: standards of care in diabetes-2023. Diabetes Care 2023 Jan 01;46(Suppl 1):S19-S40 [FREE Full text] [doi: 10.2337/dc23-S002] [Medline: 36507649]
30. Personalised nutrition for healthy living (PROTEIN). ClinicalTrials.gov. URL: https://clinicaltrials.gov/study/NCT05951140 [accessed 2026-06-11]
31. Battelino T, Danne T, Bergenstal RM, Amiel SA, Beck R, Biester T, et al. Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range. Diabetes Care 2019 Aug;42(8):1593-1603 [FREE Full text] [doi: 10.2337/dci19-0028] [Medline: 31177185]
32. Evert AB, Dennison M, Gardner CD, Garvey WT, Lau KHK, MacLeod J, et al. Nutrition therapy for adults with diabetes or prediabetes: a consensus report. Diabetes Care 2019 May;42(5):731-754 [FREE Full text] [doi: 10.2337/dci19-0014] [Medline: 31000505]
33. DiabetesNutrition Study Group (DNSG) of the European Association for the Study of Diabetes (EASD). Evidence-based European recommendations for the dietary management of diabetes. Diabetologia 2023;66(6):965-985. [doi: 10.1007/s00125-023-05894-8] [Medline: 37069434]
34. Desveaux L, Agarwal P, Shaw J, Hensel JM, Mukerji G, Onabajo N, et al. A randomized wait-list control trial to evaluate the impact of a mobile application to improve self-management of individuals with type 2 diabetes: a study protocol. BMC Med Inform Decis Mak 2016 Nov 15;16(1):144 [FREE Full text] [doi: 10.1186/s12911-016-0381-5] [Medline: 27842539]
35. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
36. Hopewell S, Chan A, Collins GS, Hróbjartsson A, Moher D, Schulz KF, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. BMJ 2025 Apr 14;389:e081123. [doi: 10.1136/bmj-2024-081123] [Medline: 40228833]
37. Oftedal B, Bru E, Karlsen B. Motivation for diet and exercise management among adults with type 2 diabetes. Scand J Caring Sci 2011;25(4):735-744. [doi: 10.1111/j.1471-6712.2011.00884.x] [Medline: 21443543]
38. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
39. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
40. Danne T, Nimri R, Battelino T, Bergenstal RM, Close KL, DeVries JH, et al. International consensus on use of continuous glucose monitoring. Diabetes Care 2017 Dec;40(12):1631-1640 [FREE Full text] [doi: 10.2337/dc17-1600] [Medline: 29162583]

XSL•FO
RenderX

Abbreviations

AI: artificial intelligence
CGM: continuous glucose monitoring
HbA1c: glycated hemoglobin A1c
mHealth: mobile health
PA: physical activity
PDA: personal digital assistants
PROTEIN: personalized nutrition for healthy living
RCT: randomized controlled trial
T2D: type 2 diabetes
TIR: time in range
WHO: World Health Organization

Edited by J Sarvestan; submitted 14.Nov.2025; peer-reviewed by J Zhang, O Owolo; comments to author 24.Mar.2026; revised version received 04.May.2026; accepted 06.May.2026; published 24.Jun.2026.

Please cite as:
Lalama E, Csanalosi M, Kabisch S, Wilson-Barnes S, Gymnopoulos LP, Dimitropoulos K, Rouskas K, Anagnostis A, Oikonomidis I, Hadjileontiadis L, Cornelissen V, Hassapidou M, Pagkalos I, Balula Dias S, Hart K, Pfeiffer AFH
Effect of a Personalized Mobile App on Glucose Control in Adults With Prediabetes and Type 2 Diabetes: Exploratory Pilot Randomized Controlled Trial
JMIR Hum Factors 2026;13:e87692
URL: https://humanfactors.jmir.org/2026/1/e87692
doi:10.2196/87692
PMID:

©Elena Lalama, Marta Csanalosi, Stefan Kabisch, Saskia Wilson-Barnes, Lazaros P Gymnopoulos, Kosmas Dimitropoulos, Konstantinos Rouskas, Argiriou Anagnostis, Ioannis Oikonomidis, Leontios Hadjileontiadis, Veronique Cornelissen, Maria Hassapidou, Ioannis Pagkalos, Sofia Balula Dias, Kathryn Hart, Andreas F H Pfeiffer. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 24.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Authors' Reply: Empathy Cuts Both Ways in Clinical AI

Sang Won Bae[1], PhD; Tongze Zhang[1], PhD; Tammy Chung[2], PhD; Anind Dey[3], PhD

[1]Department of Systems Engineering, Human-Computer Interaction and Human-Centered AI Systems Lab, AI for Healthcare Lab, Charles V. Schaefer, Jr. School of Engineering and Science, Stevens Institute of Technology, 1 Castle Point Terrace, Hoboken, NJ, United States

[2]Institute for Health, Healthcare Policy and Aging Research, Rutgers Biomedical and Health Sciences, Rutgers University, Newark, NJ, United States

[3]Information School, University of Washington, Seattle, WA, United States

Corresponding Author:
Sang Won Bae, PhD
Department of Systems Engineering, Human-Computer Interaction and Human-Centered AI Systems Lab, AI for Healthcare Lab, Charles V. Schaefer, Jr. School of Engineering and Science, Stevens Institute of Technology, 1 Castle Point Terrace, Hoboken, NJ, United States

Related Articles:

https://humanfactors.jmir.org/2026/1/e89005

https://humanfactors.jmir.org/2026/1/e108778

(JMIR Hum Factors 2026;13:e110739) doi:10.2196/110739

KEYWORDS

artificial intelligence; AI; clinical decision support systems; large language models; empathy; explainable AI; trust; human factors; reliance; human-AI interaction

Sorin and Klang's Letter to the Editor [1] regarding our evaluation [2] of an empathic clinical decision support system raises an important distinction between perceived trust and objectively warranted reliance. Emotion-adaptive communication may make accurate advice easier to understand and accept, but it could also make inaccurate advice harder to question or reject [1]. We agree that such communication should not indiscriminately increase acceptance of AI output.

Our study compared explanation delivery; it did not evaluate clinical effectiveness or appropriate reliance. The study's empathy-enhanced and basic conditions used the same cannabis-intoxication prediction and explainable AI pipeline; the empathic condition added affect sensing and emotion-adaptive delivery. We measured perceived usability, clarity, trustworthiness, and reliability—not decision accuracy, error detection, reliance, or patient outcomes. Separately reported predictive-model performance therefore does not establish that the increase in perceived trust was warranted. Consistent with this measurement concern, Sorin et al's systematic review [3] found that all but 1 of 12 studies assessed large language model empathy subjectively, underscoring the need for objective behavioral outcomes.

The relationships among accuracy, trust, and appropriate reliance require nuance to understand [4]. Users may consult decision support because they cannot independently determine whether advice is correct. Their ability to evaluate recommendations varies with clinical expertise, AI literacy, case difficulty, prior beliefs, and confidence—important considerations in our heterogeneous sample [2]. Objective accuracy is necessary but not sufficient for trustworthy use. Accurate information may be underused when presented confusingly, too technically, or without sufficient context [2]. Conversely, empathic delivery must not conceal uncertainty, weaken explanation fidelity, or make unsupported advice appear authoritative. Our design cannot determine whether higher trust reflected improved comprehension, persuasive delivery, or both.

Our study addressed the first communication barrier. We did not manipulate advice correctness; therefore, our data neither demonstrate nor rule out empathy-induced overreliance on AI-generated output. Participants reported that adaptation sometimes changed response focus or specificity, and we acknowledged the need for guardrails preserving clinical precision and explanation fidelity [2]. The same concern applies when an underlying prediction or recommendation is inaccurate, because inaccurate AI advice can impair clinical judgment [5].

Future studies should examine trust mechanisms beyond ratings. A useful design might, for example, vary advice correctness and communication style, hold the clinical scenario and evidentiary content constant across communication-style conditions, and assess effects across levels of clinician expertise. Outcomes could include objective comprehension, pre- and postadvice decisions, confidence calibration, verification behavior, acceptance of correct advice, rejection of incorrect advice, and both overreliance and underreliance on AI-generated output. Uncertainty communication and explanation fidelity also could be assessed.

Accordingly, our findings provide preliminary evidence that emotion-adaptive delivery can improve perceived clarity, usability, and trust; they do not establish improved decision

XSL•FO
RenderX

accuracy or appropriate reliance. The goal should not be maximal perceived trust but warranted trust and appropriate reliance grounded in system performance, transparent uncertainty, faithful explanation, and users' ability to interpret the evidence.

Acknowledgments

GPT-5.6 Sol (OpenAI) was used for language editing. The authors take full responsibility for the content.

Funding

This study was supported by the National Institute on Drug Abuse of the National Institutes of Health (U01DA056472 and R21 DA043181). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Sorin V, Klang E. Empathy cuts both ways in clinical Al. JMIR Hum Factors 2026;13:e108778. [doi: 10.2196/108778]
2. Zhang T, Bae SW, Chung T, Dey AK. Emotion-adaptive large language model-driven clinical decision support: user evaluation of the empathic clinical decision support system framework for trust and explainability. JMIR Hum Factors 2026 May 22;13:e89005. [doi: 10.2196/89005] [Medline: 42172616]
3. Sorin V, Brin D, Barash Y, et al. Large language models and empathy: systematic review. J Med Internet Res 2024 Dec 11;26:e52597. [doi: 10.2196/52597] [Medline: 39661968]
4. Schemmer M, Kuehl N, Benz C, Bartos A, Satzger G. Appropriate reliance on AI advice: conceptualization and the effect of explanations. In: IUI '23: Proceedings of the 28th International Conference on Intelligent User Interfaces: Association for Computing Machinery; 2023:410-422. [doi: 10.1145/3581641.3584066]
5. Gaube S, Suresh H, Raue M, et al. Do as AI say: susceptibility in deployment of clinical decision-aids. NPJ Digit Med 2021 Feb 19;4(1):31. [doi: 10.1038/s41746-021-00385-9] [Medline: 33608629]

Edited by A Schaffeler; submitted 29.Aug.2026; this is a non–peer-reviewed article;accepted 17.Sep.2026; published 02.Oct.2026.

Please cite as:
Bae SW, Zhang T, Chung T, Dey A
Authors' Reply: Empathy Cuts Both Ways in Clinical AI
JMIR Hum Factors 2026;13:e110739
URL: https://humanfactors.jmir.org/2026/1/e110739
doi:10.2196/110739

© Sang Won Bae, Tongze Zhang, Tammy Chung, Anind Dey. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 2.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Empathy Cuts Both Ways in Clinical AI

Vera Sorin[1], MD; Eyal Klang[2], MD

[1]Mayo Clinic, Rochester, MN, United States

[2]Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Boston, MA, United States

Corresponding Author:
Eyal Klang, MD
Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Boston, MA, United States

Related Articles:

Comment in: https://humanfactors.jmir.org/2026/1/e110739

https://humanfactors.jmir.org/2026/1/e89005

(JMIR Hum Factors 2026;13:e108778) doi:10.2196/108778

KEYWORDS

artificial intelligence; AI; clinical decision support systems; large language models; empathy; explainable AI; trust; human factors; reliance; human-AI interaction

Zhang et al [1] found that participants rated an emotion-adaptive clinical decision support system as clearer, easier to use, more reliable, and more trustworthy than a basic system. Both systems used the same core prediction model and explainability outputs. The empathic system added emotion sensing and a dialogue model that changed how the explanation was delivered.

A clinician who is stressed or confused may understand a difficult explanation better when the system changes how it speaks. But empathy cuts both ways. It can make correct advice easier to accept. It may also make wrong advice harder to reject.

The study measured users' ratings of trust, reliability, and clarity. It did not test whether the empathic system helped them make more accurate decisions or spot errors. The prediction models were tested separately, but that does not show whether the extra trust was deserved. A system can feel more reliable without being more accurate. A systematic review of large language model empathy found that 11 of 12 studies assessed empathy subjectively [2]. That gap matters because the goal is not the highest possible trust. Trust should match what the system can actually do. Users should accept correct AI advice and reject wrong advice [3].

The authors already identify one part of the problem. Participants reported that emotion adaptation sometimes changed the focus or specificity of the answer. Some replies became overly supportive and did not address the clinical question directly. The authors therefore call for controls that protect clinical precision and decision clarity [1]. But even a direct and precise answer can be wrong. When that answer is also empathic, it may be harder to reject.

This is not only a theoretical concern. In a clinical study, inaccurate advice reduced physicians' diagnostic accuracy [4]. Another study of health chatbots found that empathic responses increased perceived warmth, trust, and intentions to use the system [5]. These findings do not prove that empathy increases acceptance of wrong clinical advice. They show why the question needs to be tested. One objection is that empathy may raise trust because it improves understanding, not because it persuades. That may be true. The test should separate these effects. Better understanding should help users accept correct advice and reject wrong advice. Persuasion alone may increase acceptance of both.

Future studies should compare correct and wrong AI advice in empathic and neutral language. The clinical content should stay the same. Users could make a judgment before seeing the advice and again afterward. The main outcomes should be final decision accuracy, acceptance of correct advice, rejection of wrong advice, and a direct test of what users understood. Trust ratings can remain, but they should not stand alone.

Emotion-adaptive clinical AI should not make every answer more trusted. It should help trust follow correctness.

Acknowledgments

ChatGPT (GPT-5.6 Pro; OpenAI), accessed through the ChatGPT web interface, was used to improve grammar, spelling, and readability. The authors take full responsibility for the final content.

XSL•FO
RenderX

Funding

The authors declared no financial support was received for this work.

Authors' Contributions

VS and EK conceived the argument, drafted and revised the manuscript, and approved the final version.

Conflicts of Interest

None declared.

References

1. Zhang T, Bae SW, Chung T, Dey AK. Emotion-adaptive large language model-driven clinical decision support: user evaluation of the empathic clinical decision support system framework for trust and explainability. JMIR Hum Factors 2026 May 22;13:e89005. [doi: 10.2196/89005] [Medline: 42172616]
2. Sorin V, Brin D, Barash Y, et al. Large language models and empathy: systematic review. J Med Internet Res 2024 Dec 11;26:e52597. [doi: 10.2196/52597] [Medline: 39661968]
3. Schemmer M, Kuehl N, Benz C, Bartos A, Satzger G. Appropriate reliance on AI advice: conceptualization and the effect of explanations. In: IUI '23: Proceedings of the 28th International Conference on Intelligent User Interfaces: Association for Computing Machinery; 2023:410-422. [doi: 10.1145/3581641.3584066]
4. Gaube S, Suresh H, Raue M, et al. Do as AI say: susceptibility in deployment of clinical decision-aids. NPJ Digit Med 2021 Feb 19;4(1):31. [doi: 10.1038/s41746-021-00385-9] [Medline: 33608629]
5. Seitz L. Artificial empathy in healthcare chatbots: does it feel authentic? Comput Hum Behav Artif Hum 2024 Jan;2(1):100067. [doi: 10.1016/j.chbah.2024.100067]

Edited by A Schaffeler; submitted 04.Aug.2026; this is a non–peer-reviewed article;accepted 17.Sep.2026; published 02.Oct.2026.

Please cite as:
Sorin V, Klang E
Empathy Cuts Both Ways in Clinical AI
JMIR Hum Factors 2026;13:e108778
URL: https://humanfactors.jmir.org/2026/1/e108778
doi:10.2196/108778

© Vera Sorin, Eyal Klang. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 2.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mobile Health for Postpartum Contraception Knowledge in India (mMitra): Realist Evaluation Using Moderation Analysis and Structural Equation Models

Abinaya Chandrasekar[1], MPH, PhD; Emma Slaymaker[2], MSc, PhD; John Ford[3], MBChB, MSc, PhD; Neha Madhiwalla[4], PhD; Emily Warren[5], PhD; Caroline Free[2], MBChB, MSc, PhD; Ona L McCarthy[5], MSc, PhD

[1]Cicely Saunders Institute, Florence Nightingale Faculty of Nursing, Midwifery & Palliative Care, King's College London, Bessemer Road, London, England, United Kingdom

[2]Department of Population Health, Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, England, United Kingdom

[3]Wolfson Institute of Population Health, Faculty of Medicine and Dentistry, Queen Mary University of London, London, England, United Kingdom

[4]Research Division, ARMMAN, Mumbai, Maharashtra, India

[5]Department of Health Services Research and Policy, Faculty of Public Health and Policy, London School of Hygiene & Tropical Medicine, London, England, United Kingdom

Corresponding Author:
Abinaya Chandrasekar, MPH, PhD
Cicely Saunders Institute, Florence Nightingale Faculty of Nursing, Midwifery & Palliative Care, King's College London, Bessemer Road, London, England, United Kingdom

Abstract

Background: Indian National Family Health Survey data from 2021 suggest that 9.4% of women of reproductive age had an unmet need for spacing or limiting future pregnancies. Randomized controlled trials evaluating mobile health interventions (programs delivered via mobile phones) provide evidence of their potential benefit in addressing the determinants of misconceptions and lack of knowledge regarding postpartum family planning (PPFP) methods. The mMitra program has been implemented since 2014 and uses automated voice-messaging to deliver pregnancy and postpartum care information to socioeconomically vulnerable women in the Mumbai metropolitan area. However, the mMitra program has not been adequately evaluated to understand its impact on PPFP knowledge and outcomes.

Objective: This study aims to assess whether, how, why, for whom, and in what contexts the mMitra program affects PPFP awareness by analyzing cohort follow-up survey data, call duration, and attendance to measure the impact of listenership and by refining context-mechanism-outcome configurations (CMOCs) by testing for effect modification in logistic regression and structural equation models (SEMs).

Methods: This study presents the quantitative findings from the realist evaluation of the mMitra program. Moderator analysis (through testing for effect modification in a logistic regression model) and SEMs were used to test CMOCs from an accompanying realist review.

Results: After adjusting for religion as a potential confounder and including phone access as an effect modifier in the logistic regression model, mMitra call duration, a proxy measure for program engagement, had a modest effect on knowledge of one or more highly effective contraceptive methods (odds ratio 1.012, 95% CI 1.002 - 1.021; *P*=.02). In addition, 77% of women measured at baseline did not complete the endline survey at 6 months, potentially due to cultural expectations surrounding women returning to natal homes during their final months of pregnancy. SEMs show some evidence of an association between shared phone ownership and mMitra program engagement, potentially via the mechanism of women's ability to negotiate phone access.

Conclusions: This study adds to the growing body of evidence on the usefulness of SEM for testing CMOCs in realist evaluations. Issues with data collection and management, and the absence of survey data from locations outside Mumbai, limit the generalizability of the conclusions from this study.

(JMIR Hum Factors 2026;13:e97650) doi:10.2196/97650

KEYWORDS

mHealth; postpartum; family planning; realist evaluation; structural equation modeling; multivariable regression

XSL•FO
RenderX

Introduction

Background

Approximately 218 million women of reproductive age (15 - 49 years) in low- and middle-income countries (LMICs) have an unmet need for family planning (FP) [1-4]. Indian National Family Health Survey Data from 2021 suggest that 9.4% of women of reproductive age had an unmet need for spacing or limiting future pregnancies [1]. The total unmet need for FP in India also significantly varies according to demographic determinants, such as educational attainment, the number of antenatal care visits, and place of residence [5,6]. Poor knowledge on the availability of FP services and the diversity of FP methods influence contraceptive nonuse [7]. Misconceptions regarding the effectiveness and potential side-effects associated with modern contraceptives also contribute to contraceptive nonuse among rural Indian women [8]. A descriptive study of Indian District Level Household Survey Data from 2007 and 2008 confirms this finding; women continue to use traditional, less effective methods of FP due to lack of information dissemination on modern methods by local health services [9].

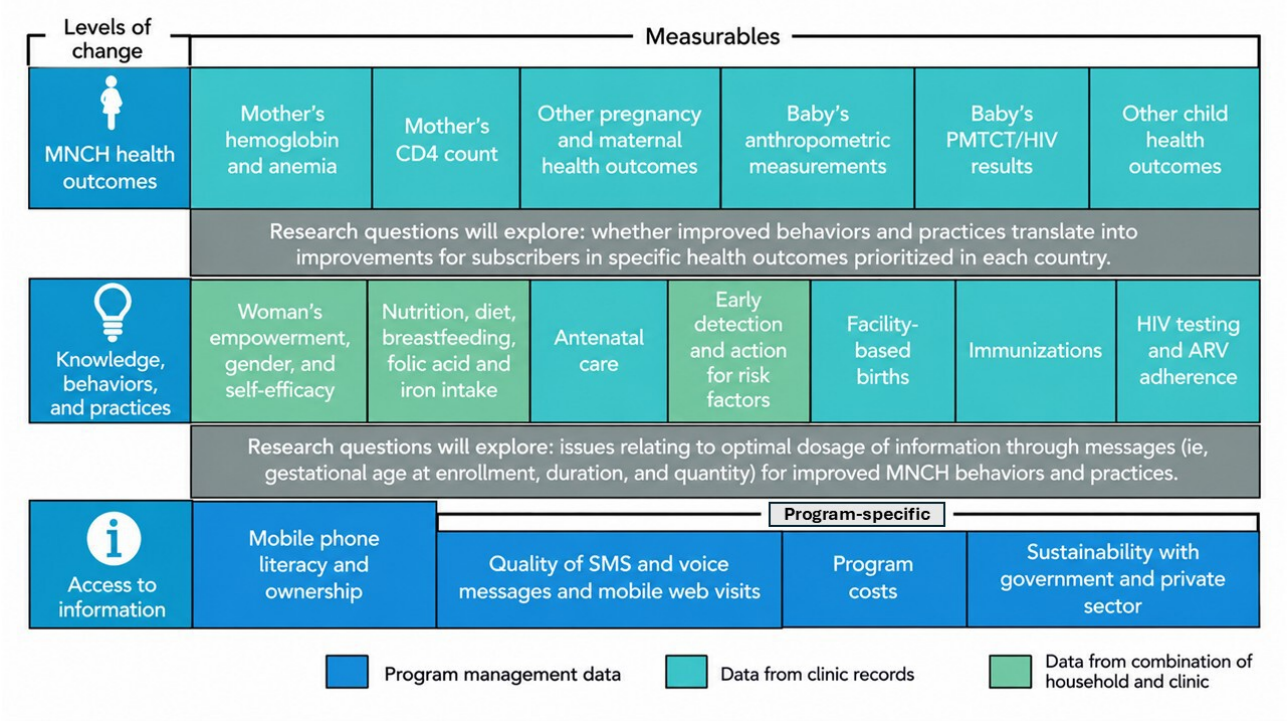
Globally, mobile phone ownership has grown exponentially over the last decade [10], becoming an avenue for many individuals from LMICs to access crucial health information and services [10]. Mobile health (mHealth) includes any health program supported by or delivered through mobile technologies such as cellular phones, personal monitoring devices, and other wireless devices [11]. mHealth interventions have been used to bridge the knowledge gap on FP methods across population subsets [12]. One example is the mobile for reproductive health program piloted in Kenya and Tanzania, which was successful in improving participants' knowledge of FP services [13]. Findings from the randomized controlled trial of the Kilkari intervention in Madhya Pradesh, India, suggest that women in the intervention group were more likely to ask health care providers about FP methods compared to women in the control groups (risk ratio 1.04, 95% CI 1.01 - 1.07; P<.001) [14].

The "mMitra" program, initially piloted in Mumbai and currently being implemented in 9 Indian states, provides pregnancy and postpartum care information to participants through 141 automated voice messages [2,15]. mMitra has reached approximately 3.6 million women across India and is underpinned by the assumption that access to easily understandable and engaging information aligning with women's antenatal and postnatal stages will encourage them to perform necessary self-care measures and "seek needed health services" [16]. mMitra is modeled after the Mobile Alliance for Maternal Action intervention from South Africa, which uses the same theory of change, and proposes that sending messages tailored to women's gestational stage would motivate women to change pregnancy and postpartum self-care behaviors, which would subsequently improve maternal and child health outcomes [17]. The mMitra program messages were developed by BabyCenter, using guidelines on pregnancy self-care and infant development produced by the World Health Organization (WHO) and the Indian Health Mission [18]. For mMitra subscribers in the Mumbai Metropolitan Region, messages are delivered in Hindi or Marathi and range between 1 and 2.5 minutes in duration. mMitra messages are delivered in the chosen language and timeslot of every enrolled woman. The automated voice-message delivery system makes 9 attempts to reach every woman for each of the 141 messages. An incoming missed call system and mMitra call center service are also available to offer technical support to enrolled women.

Women are enrolled either at government health facilities by trained health workers or in communities by trained outreach workers, enabling coverage of socioeconomically vulnerable populations. By using widely available mobile phones, the program aims to address gaps in counseling and health communication within routine maternal and child health services. The program theory of change for the mMitra program proposes that delivering timely and comprehensive health information increases the knowledge of recommended practices, including adequate birth spacing between consecutive pregnancies. Increased knowledge can enhance women's motivations and perceived feelings of self-efficacy to adopt recommended practices such as postpartum family planning (PPFP). The theory of change underpinning the Mobile Alliance for Maternal Action series of interventions emphasizes the causal pathway between obtaining knowledge, subsequent motivation, and the adoption of self-care practices to improve maternal, reproductive, and child health behaviors (Figure 1).

XSL•FO
RenderX

Figure 1. Mobile Alliance for Maternal Action (MAMA) theory of change and priority outcomes applied to the development of the mMitra program (reproduced with permission from Mechael et al [16]). ARV: antiretroviral; CD4: cluster of differentiation 4; MNCH: maternal, newborn, and child health outcomes; PMTCT: prevention of mother-to-child transmission.



There is a gap in the existing evidence on the effectiveness of the mMitra program. Murthy et al [2,3] evaluated the program's impact on subscribers' postpartum health behaviors. Subscribers were no more likely (odds ratio [OR] 0.82, 95% CI 0.64 - 1.04) to discuss adopting an FP method or report having used an FP method compared to controls [2]. Although mMitra has scaled up across Indian geographies and contexts since the pseudorandomized trial reported above, there is a need for updated evidence on the reach and effectiveness of the program [15]. In order to understand how complex social interventions change outcomes, evaluations should explore both how various contexts can alter the intervention's functions and also how these contextual features alter the ways individuals interact with mHealth interventions [19]. To do this, mMitra listenership, engagement, and knowledge were evaluated through a cohort study (Appendix A in Multimedia Appendix 1).

We conducted a realist evaluation of mMitra [20,21] to assess the program's influence on PPFP awareness, given the substantial impact of contextual factors, such as age, income, ability to access adequate health services, individual empowerment, and interpersonal relationships [22,23]. The realist approach employs the generative model of causality that emphasizes the role of contexts and conditions in which mechanisms are activated to produce program outcomes, thus aligning with the evidence gap on the mMitra program [3,21,24]. This realist evaluation builds upon the findings of an accompanying realist review by testing and further refining the context-mechanism-outcome configurations (CMOCs) concerning whether, how, for whom, and in what contexts the mMitra program influences PPFP knowledge. The realist review identified contextual factors influencing participant engagement with mHealth interventions, such as women's access to mobile phones, household power dynamics, and social support. This realist evaluation involved a secondary analysis of quantitative data from the aforementioned cohort study. While there is ongoing debate about the usefulness of quantitative methods to test and refine CMOCs in realist evaluations [24-26], we propose that quantitative methods are useful for testing theorized causal pathways regarding how and for whom program resources change outcomes, and for assessing the strength of the association between theoretical components leading to program outcomes [21,27-29]. We propose that moderator analysis, through testing for effect modification in logistic regression models, and structural equation models (SEMs) allow for testing and refining CMOCs from the accompanying realist review and contextualize these findings within the mMitra program.

Aims and Objectives

Aim

The aim of this study is to quantitatively explore how, for whom, and in what contexts the mMitra program influences PPFP knowledge.

Objectives

The objectives of this study are

1. to analyze mMitra cohort study follow-up survey data, call durations, and attendance to evaluate the effect of mMitra listenership on PPFP awareness.
2. to refine CMOCs identified in the accompanying realist review of how, whether, and for whom the mMitra program works through testing for effect modification in logistic regression models and structural equation modeling.

XSL•FO
RenderX

Methods

Study Design

We propose that moderator analysis, through testing for effect modification in logistic regression models, and structural equation modeling allow for testing and refining CMOCs from the accompanying realist review and contextualize these findings within the mMitra program. Findings have been reported in accordance with RAMESES II (Realist and Meta-Narrative Evidence Syntheses: Evolving Standards II) reporting standards for realist evaluations [30].

Data Collection

We performed a secondary analysis of cohort study follow-up survey data and system data on call attendance and duration collected from mMitra subscribers. The original cohort study comprised data collection at baseline and 4 follow-up points administered at 3-month intervals. Routinely collected system data on call attendance and duration were extracted only for the time period corresponding to enrollment and participation in the cohort study. The second follow-up survey, approximately 6 months after enrollment, was the end line of this study because it included questions on PPFP knowledge.

Ethical Considerations

Ethical approval for the mMitra cohort study was provided by ARMMAN's Institutional Review Board (Appendix B in Multimedia Appendix 1). Ethical approval for the secondary analysis of cohort study survey data was provided by the Observational/Interventions Research Ethics Committee at The London School of Hygiene & Tropical Medicine and ARMMAN's Institutional Review Board (Appendix C in Multimedia Appendix 1).

Exposure Variables

Engagement with the mMitra program was measured in 2 ways, resulting in the following two exposure variables:

1. Total mMitra call duration—total duration in minutes of mMitra calls listened to by each subscriber.
2. Call duration of mMitra calls only on FP—duration in minutes of mMitra calls only on the topic of FP listened to by each subscriber.

Outcome Variable

PPFP awareness, the outcome variable, was defined as subscribers' awareness of one or more highly effective contraceptive (HEC) methods and was measured as a binary variable.

Contextual Variables and Mechanisms

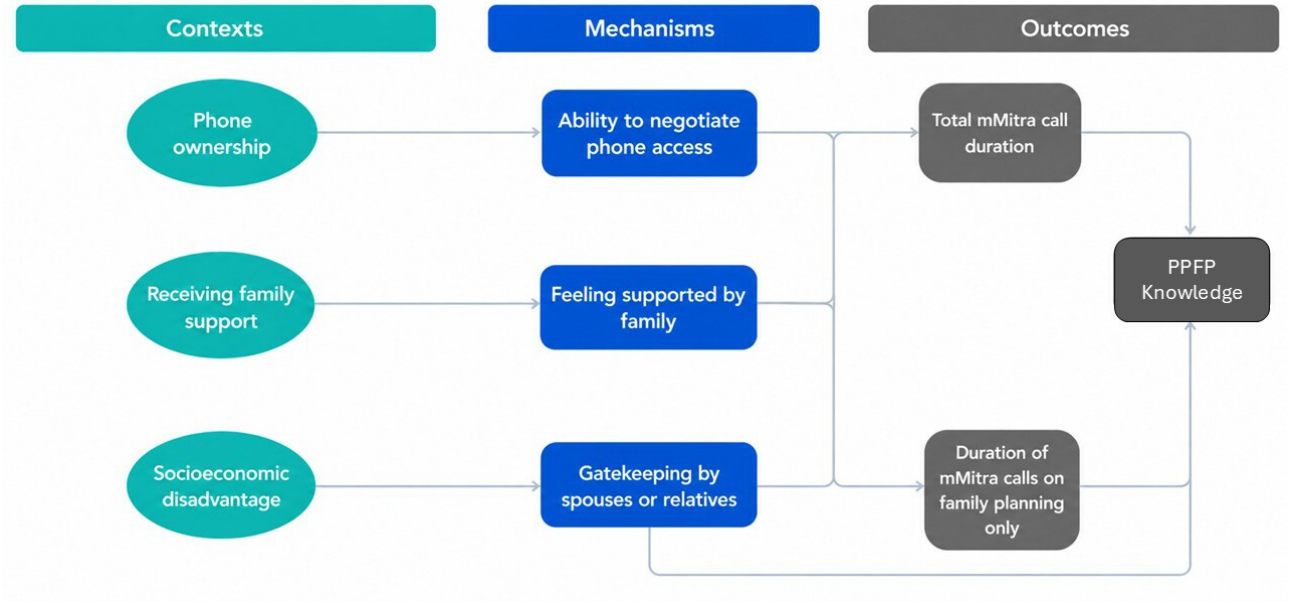
Variables representing potential realist contexts included demographic characteristics, education, religion, and indicators of phone access and household dynamics measured in the baseline survey. Age (continuous, 18 - 43 years), caste (open or general/other backward caste/scheduled caste, scheduled tribe, or nomadic tribe), religion (Hindu/Muslim/Buddhist/Other), and educational attainment (less than secondary education/secondary education or higher) were considered potential confounders of the association between mMitra call duration and PPFP awareness and were measured in the baseline survey. Age was considered a potential confounder because older women are more likely to have greater knowledge of PPFP needs [31-33], as well as a greater understanding of adverse pregnancy outcomes [31-33]. This means that they may be more willing to learn about pregnancy-related self-care measures, leading to greater engagement with the mMitra program. Women who have greater educational attainment may be more inclined to learn about pregnancy and postpartum care and thus engage more with the mMitra program. Women who have greater educational attainment may also have greater preexisting knowledge of PPFP methods [34-37]. Phone access was constructed as a composite binary variable by combining binary indicator variables that answered the questions outlined in Appendix D in Multimedia Appendix 1.

Women who responded with all the high-access responses for the binary indicator variables (outlined in Appendix D in Multimedia Appendix 1) were categorized as having "High access" to mobile phones. For the purposes of constructing this access variable, women who did not provide "high-access responses" for all the indicator variables were considered to have "low access" to mobile phones. Women with missing data for any of the variables were coded as missing data in this composite "Access" variable.

Latent variables, representing mechanisms identified in the realist review, were constructed using combinations of observed variables and used to test mechanisms using SEMs. These mechanisms and their potential to generate PPFP awareness improvement through greater engagement with mHealth programs are rooted in evidence from the accompanying realist review. We started by identifying variables in the dataset that could theoretically reflect the latent mechanisms to be tested based on evidence from the realist review (Appendix G in Multimedia Appendix 1). We tested Cronbach α values for each combination of observed variables that could make up the latent mechanism variables to test for construct validity and internal consistency. The combination of observed variables with the highest Cronbach α value was selected. The following CMOCs were tested using structural equation modeling (Figure 2). Different combinations of the measured variables were used to construct the mobile phone access variable used in the logistic regression model and the latent variables used in the SEMs. This is to reflect the overarching realist logic underpinning this study. These variables (Appendix D in Multimedia Appendix 1) reflect women's access to mobile phones and consider household dynamics that could influence these concepts, which can be explored through logistic regression models. However, certain elements of phone access (eg, whose phone is registered with the mMitra program) may influence other elements of phone access such as a women's confidence to ask to use shared mobile phones from other family members, which is better explored through constructing SEMs.

XSL•FO
RenderX

Figure 2. Context-mechanism-outcome configurations (CMOCs) from accompanying realist review tested through structural equation models (SEMs). PPFP: postpartum family planning.



Data Analysis

We conducted univariate analysis of the exposure, outcome, potential confounders, and potential effect modifiers to describe the demographic characteristics of the study sample. The results are presented as frequencies and percentages for categorical or binary variables and as means and SDs for continuous variables.

We summarized call durations by sociodemographic characteristics and compared them using *z*-tests. Bivariate analysis was conducted between the exposure and outcome variables to calculate a crude OR. Mantel-Haenszel OR estimates were used to assess potential confounding or effect modification of the association between mMitra call duration and PPFP awareness. Potential confounders not on the causal pathway were considered for inclusion in the logistic regression model if there was evidence of an association ($P \leq .05$) between the confounder and both the exposure and outcome variables.

In the multivariable analysis, 2 logistic regression models were built: one for each of the 2 exposure variables. Confounders and effect modifiers (Table S1 in Multimedia Appendix 2) were added sequentially based on bivariate analysis findings and rationalized by evidence from the literature. A threshold of $P \leq .05$ was used to determine the statistical significance of estimates. Crude ORs, adjusted ORs, and their associated 95% CIs are reported to 2 or 3 decimal places to ensure that the small differences in effect estimates are reflected in the results.

We built 3 SEMs to represent causal pathways (mechanisms) identified in CMOCs from the accompanying realist review (Figure 2). These included (1) receiving and feeling supported by family, (2) ability to negotiate phone access, and (3) gatekeeping of mMitra phone access by spouses and relatives. Principal component analysis and confirmatory factor analysis were used to assess the structure and reliability of covariance among observed variables forming each latent construct. We used a conservative Cronbach α threshold of ~.60 to account for inconsistencies and error in the reporting and measurement of observed variables [38-40]. SEMs were built using Stata 18 (StataCorp LLC) [41] and statistical significance was defined as $P \leq .05$. As alpha values informed latent variable construction, goodness-of-fit estimates were not used to assess model fit. The decision to treat "Receiving family support" (ie, do family members encourage you to listen to mMitra calls?) as a contextual factor distinct from the mechanism of "feeling supported by family" was based on realist review findings: emotional support and confidence to engage with mHealth were enabled by practical access to a phone [42]. PPFP awareness and mMitra engagement were modeled as outcomes in the CMOCs, reflecting the realist review's identification of mHealth engagement as an outcome produced by mechanisms operating within specific contexts. Given the exploratory and theory-testing nature of this analysis, we interpreted findings cautiously and emphasized patterns consistent with hypothesized CMOCs.

Results

Study Population Characteristics

Baseline data were collected from 884 mMitra subscribers; a large percentage of these subscribers were lost to follow-up and were not captured at end line (the second follow-up survey data were collected approximately 6 months after baseline). There were negligible differences in demographic characteristics of mMitra subscribers between those who completed the second follow-up survey and those who were lost to follow-up.

Out of 884 individuals who were captured at baseline, 683 (77.3%) women were lost to follow-up and 201 women completed follow-up survey 2. There were no significant differences in demographics or access among women lost to follow-up and those captured at end line (Table S1 in Multimedia Appendix 2). Table 1 describes the number of minutes the 201 women in the end line survey had spent listening to mMitra calls and to the call segments containing FP messages. Table S2 in Multimedia Appendix 3 presents the

XSL•FO
RenderX

proportion of the end line respondents who were aware of one or more highly effective methods of contraception. There were small variations by sociodemographic characteristics in both the total duration of calls listened to and the duration of FP calls, but *z*-test results (not shown) indicated that these were all likely to be chance variations (Table 2). Awareness of at least one highly effective method of contraception did not differ by age, education, or gestational age at enrollment. It was highest among women of open or general caste, Muslim women, and those with low access to a mobile phone at baseline (Table S2 in Multimedia Appendix 3).

Table . Awareness of modern methods of contraception among respondents to follow-up survey, by sociodemographic characteristics.

Characteristic	Number of respondents, n	Traditional methods only, n (%)	Modern HEC[a] methods, n (%)
Total	171	84 (49.12)	87 (50.88)
Age (y)			
<20	2	1 (50.00)	1 (50.00)
20 - 30	142	70 (49.30)	72 (50.70)
≥31	27	13 (48.15)	14 (51.85)
Highest completed education			
Less than secondary education	32	19 (59.38)	13 (40.62)
Secondary education or higher	139	65 (46.76)	74 (53.24)
Gestational age at enrollment (mo)			
<6	126	61 (48.41)	65 (51.59)
≥6	45	23 (51.11)	22 (48.89)
Caste			
Open or general	101	40 (39.60)	61 (60.40)
OBC[b]	16	13 (81.25)	3 (18.75)
SC[c]/ST[d]/NT[e]	31	18 (58.06)	13 (41.94)
Religion			
Hindu	77	45 (58.44)	32 (41.56)
Muslim	78	25 (32.05)	53 (67.95)
Buddhist	12	10 (83.33)	2 (16.67)
Other	1	1 (100)	0 (0.00)
Access			
Low access	63	28 (44.44)	35 (55.56)
High access	58	32 (55.17)	26 (44.83)

[a]HEC: highly effective contraceptive.

[b]OBC: other backward caste.

[c]SC: scheduled caste.

[d]ST: scheduled tribe.

[e]NT: nomadic tribe.

XSL•FO
RenderX

Table . Demographic characteristics of mMitra cohort study population.

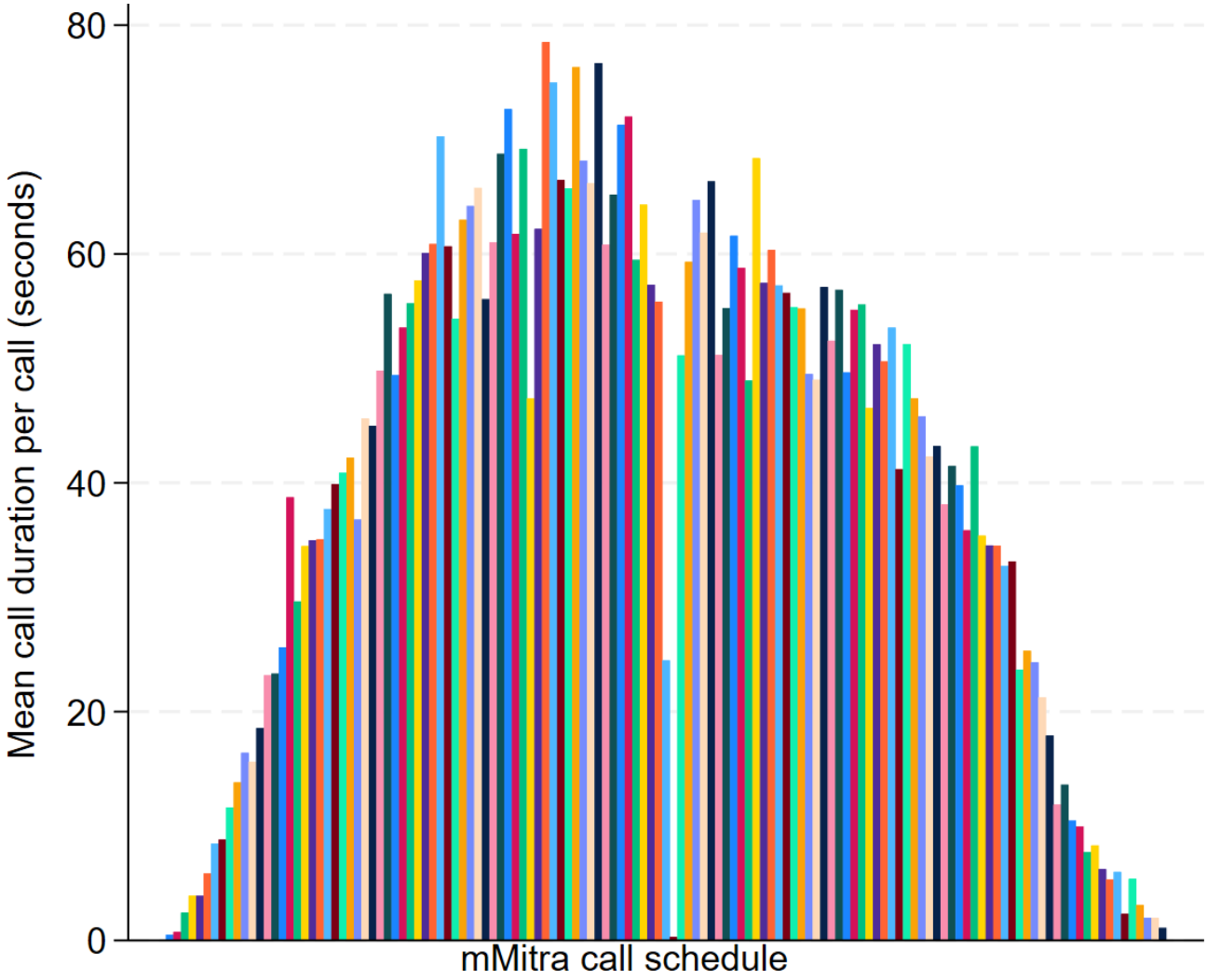
Characteristic	Lost to follow-up (n=683)	Followed up in survey 2 (n=201)	Total (N=884)
Age at enrollment (y), mean (SD)	25.39 (4.00)	26.28 (4.00)	25.61 (4.01)
Total call duration (min), mean (SD)	73.93 (67.01)	118.44 (68.58)	84.05 (69.87)
FP[a]-only call duration (min), mean (SD)	5.75 (4.62)	8.24 (4.60)	6.39 (4.74)
Highest completed education, n (%)			
Less than secondary education	125 (19.8)	37 (18.4)	162 (19.4)
Secondary education or higher	507 (80.2)	164 (81.6)	671 (80.6)
Gestational age, n (%)			
Less than 6 months	438 (64.1)	144 (71.6)	582 (65.8)
6 months or more	245 (35.9)	57 (28.4)	302 (34.2)
Caste, n (%)			
Open or general	346 (73.0)	120 (69.4)	466 (72.00)
Other backward caste (OBC)	45 (9.5)	17 (9.8)	62 (9.6)
Scheduled caste/scheduled tribe/nomadic tribe	83 (17.5)	36 (20.8)	119 (18.4)
Religion, n (%)			
Hindu	308 (49.6)	95 (48.0)	403 (49.2)
Muslim	281 (45.2)	90 (45.5)	371 (45.3)
Buddhist	28 (4.5)	12 (6.1)	40 (4.9)
Other	4 (0.6)	1 (0.5)	5 (0.6)
Access, n (%)			
Low access	218 (52.5)	72 (52.2)	290 (52.4)
High access	197 (47.5)	66 (47.8)	263 (47.6)
Total, n (%)	683 (77.3)	201 (22.7)	884 (100)

[a]FP: family planning.

Mean call duration per call was highest during the mMitra calls scheduled for later in pregnancy (Figure 3). Mean call duration per call also tapers off during the final calls of the mMitra schedule. The second message delivered during week 39 had the lowest call duration (mean 0.31, SD 6.12 seconds) and most women in the dataset did not attend this call (n=750).

XSL•FO
RenderX

Figure 3. Mean call duration per call of the mMitra call schedule.



On average, women who were more aware (knew one or more HEC methods in the follow-up survey) had listened to 123.38 minutes of the mMitra program, whereas women who were less aware (only knew traditional FP methods) listened to a mean total mMitra call duration of 120.37 (SD 71.37) minutes. Women who were more aware had a mean duration of FP-only calls of 8.05 (SD 4.47) minutes, whereas women who were less aware had a mean duration of FP-only calls of 8.83 (SD 4.71) minutes.

Crude Association Between Call Durations and PPFP Knowledge

There was no evidence of an association between total mMitra call duration per participant and PPFP awareness (OR 1.00, 95% CI 0.997 - 1.004; *P*=.83). There was some evidence of an association between religion and total mMitra call duration (OR 0.61, 95% CI 0.40 - 0.92; *P*=.02) as well as between religion and PPFP awareness (OR 2.15, 95% CI 1.16 - 3.99; *P*=.02; Multimedia Appendix 4). Therefore, religion was adjusted for as a confounder in the logistic regression model of the effect of total mMitra call duration on PPFP awareness. No variables fulfilled the criteria for confounding in the logistic regression model of the association between the duration of FP-only calls and PPFP awareness. There was no evidence of an association between the duration of FP-only calls and PPFP awareness (OR 0.97, 95% CI 0.91 - 1.03; *P*=.32).

Multivariable Analysis—Logistic Regression

Main Findings

There was some evidence of an association between total mMitra call duration and PPFP awareness after including an interaction term reflecting phone access as an effect modifier and religion as a confounder (OR 1.012, *P*=.02). Of the 3 CMOCs tested via SEM (feeling supported by family, ability to negotiate phone access, and gatekeeping by spouses or family), there was some evidence of an indirect association between context (phone ownership) and outcomes (mMitra program engagement) via the mediating mechanism of women's ability to negotiate phone access.

Association Between Total mMitra Call Duration and PPFP Awareness

After including an interaction term reflecting phone access as an effect modifier and religion as a confounder, there was some evidence of an association between total mMitra call duration and PPFP awareness. Women with low access to mobile phones (effect modifier) were the reference group for this regression model. For every 1-minute increase in total mMitra call duration, mMitra subscribers with low access to mobile phones had 1.012

XSL•FO
RenderX

times greater odds of identifying one or more HEC methods (OR 1.012, 95% CI 1.002 - 1.021; *P*=.02). There was no evidence of an association between total mMitra call duration and PPFP awareness among women with high access to mobile phones (OR 0.997, 95% CI 0.989 - 1.005; *P*=.48). Likelihood-ratio test estimates (*P*=.02) provide evidence that the logistic regression model is a better fit with the addition of the interaction term between phone access and total mMitra call duration. Religion was associated with the outcome as a confounder.

Association Between the Duration of mMitra Calls on FP Only and PPFP Awareness

We did not conduct a multivariable regression analysis with the duration of FP-only calls as the exposure variable because no significant confounders or effect modifiers were identified in the bivariate analysis.

SEMs

SEMs were built to test 3 CMOCs. These CMOCs were chosen because latent variables could plausibly be constructed using the measured variables in the mMitra cohort study dataset. We measured the direct association between context and outcome and the indirect association between context and outcome via a mediating mechanism (indirect association).

Feeling Supported by Family

Cronbach α for this SEM indicated good reliability (0.65). There was no evidence of a direct context-outcome or indirect context-mechanism-outcome association in this SEM (Table 3).

Table . Direct (context to outcome) and indirect (context to outcome via mechanism) effects of family support context-mechanism-outcome configurations (CMOCs) detailing the mechanism of feeling supported by family tested via structural equation modeling (context: receives family support).

Outcome	Direct effect (context to outcome)		Indirect effect (context to outcome via mechanism)	
	β or OR[a] (95% CI)	*P* value	β or OR (95% CI)	*P* value
Total mMitra call duration	1.89[b] (−48.62 to 52.41)	.94	10.78[b] (−37.82 to 59.38)	.66
Duration of calls on FP[c] only	.70[b] (−2.99 to 4.38)	.71	−.04[b] (−3.59 to 3.50)	.98
PPFP[d] awareness	0.77[e] (0.38 to 1.46)	.47	1.32[e] (0.68 to 2.58)	.41

[a]OR: odds ratio.

[b]Denotes β values.

[c]FP: family planning.

[d]PPFP: postpartum family planning.

[e]Denotes OR values.

Ability to Negotiate Phone Access

Cronbach α for this SEM indicated moderate reliability (0.59). There was some evidence of an indirect association between phone ownership (context) and total mMitra call duration (outcome) via the mechanism of women's ability to negotiate phone access (β=−16.54, 95% CI −31.61 to −1.47; *P*=.03). This means women who shared mobile phones with husbands or relatives listened to 16.54 fewer minutes of the mMitra program because they were unable to negotiate phone access compared to women who used their own mobile phones to engage with the mMitra program. There was also some evidence of an association between phone ownership (context) and the duration of mMitra calls for FP only (outcome) via the mechanism of women's ability to negotiate phone access (β=−1.41, 95% CI −2.52 to −0.29; *P*=.01; Table 4).

Table . Direct (context to outcome) and indirect (context to outcome via mechanism) effects of context-mechanism-outcome configurations (CMOCs) detailing the mechanism of women's ability to negotiate phone access tested via structural equation modeling (context: phone registered with mMitra [own phone or shared phone with spouse/relatives]).

Outcome	Direct effect (context to outcome)		Indirect effect (context to outcome via mechanism)	
	β or OR[a] (95% CI)	*P* value	β or OR (95% CI)	*P* value
Total mMitra call duration	.93[b] (−23.08 to 24.94)	.94	−16.54[b] (−31.61 to −1.47)	.03[c]
Duration of FP[d]-only calls	−.32[b] (−2.09 to −1.41)	.72	−1.41[b] (−2.52 to −0.29)	.01[c]
PPFP[e] awareness	1.33[f] (0.83 to 2.14)	.24	0.87[f] (0.63 to 1.20)	.40

[a]OR: odds ratio.

[b]Denotes β values.

[c]*P*≤.05 was considered statistically significant.

[d]FP: family planning.

[e]PPFP: postpartum family planning.

[f]Denotes OR values.

XSL•FO
RenderX

Gatekeeping by Spouses or Relatives

Cronbach α for this SEM indicated moderate reliability (0.61). There was some evidence of a direct association between socioeconomic disadvantage (context) and PPFP awareness (outcome; OR 0.57, 95% CI 0.33 - 1.01; *P*=.054; Table 5).

Table . Direct (context to outcome) and indirect (context to outcome via mechanism) effects of context-mechanism-outcome configurations (CMOCs) detailing gatekeeping by spouses and relatives tested via structural equation modeling (context: socioeconomic disadvantage).

Outcome	Direct effect (context to outcome)		Indirect effect (context to outcome via mechanism)	
	β or OR[a] (95% CI)	*P* value	β or OR (95% CI)	*P* value
Total mMitra call duration	–14.57[b] (–40.21 to 11.07)	.27	1.27[b] (–5.57 to 8.11)	.72
Duration of FP[c]-only calls	–1.76[b] (–3.64 to 0.11)	.07	.09[b] (–0.38 to 0.56)	.72
PPFP[d] awareness	0.57[e] (0.33 to 1.01)	.054[f]	1.00[e] (0.99 to 1.01)	.95

[a]OR: odds ratio.

[b]Denotes β values.

[c]FP: family planning.

[d]PPFP: postpartum family planning.

[e]Denotes OR values.

[f]*P*≤.05 was considered statistically significant.

Discussion

Main Findings

This analysis found some evidence for a modest association between total mMitra call duration and PPFP awareness. Although the effect size per additional minute of listening to the mMitra program was small, women, on average, listened to approximately 118 (SD 68.58) minutes of the mMitra program. Thus, the effect of this program could compound over time to lead to a substantial increase in awareness. Shared ownership of mobile phones (context) lowered women's ability to access the program (mechanism), which led to slightly poorer engagement (outcome). Building on the logistic regression models, an examination of specific domains of phone access through the SEM suggests that access can influence engagement with the program contingent on context. Although women with low-access to mobile phones were most likely to benefit from the mMitra program, the program was not designed intentionally to increase phone access, and therefore engagement, in this group. There was also some evidence that women who experienced gatekeeping by spouses and relatives of a shared mobile phone (context) were less likely to identify HEC methods (outcome). There was a 77% loss to follow-up by 6 months. It is important to note that this study aimed to unpick the associations between many co-occurring complex concepts using all available data, and the findings of this study were interpreted by acknowledging these limitations.

Interpretation

There was evidence of a significant difference in the effect of listening to the mMitra program between women who had low access to a mobile phone and those who had high access to a mobile phone. It is possible that women with high access to mobile phones may have had less knowledge to gain from listening to the mMitra program. At end line, the proportion of women aware of a highly effective method of contraception was at 55%, higher in the low access group than in the high access group (44%); there may have been a larger increase in awareness among the low access group if they started from a lower baseline level of knowledge. Awareness of PPFP was conducted during the second follow-up survey, administered 6 months after enrollment in the mMitra program, which means we are not able to test this. This limits our ability to conclusively determine the salience of the identified mechanisms and their ability to generate program engagement and PPFP awareness outcomes.

The difference in the association of mMitra listenership with PPFP awareness between women with low access to a mobile phone and those with high access may be explained by differences in the relevance of the other program content. FP messages make up less than 10% of the overall mMitra program content, and the messages are spread over the course of several months. If the other program content was less useful for women with high access to a mobile phone, perhaps due to existing high levels of knowledge, they may have "switched off" and been less receptive to the FP messages, even though they listened to the same amount as women with low access.

It is possible that the mMitra program potentially activates women's interest in self-care during pregnancy and the postpartum period, and that the full dose of content is not required to elicit changes in PPFP awareness. mMitra subscribers may then choose to access contraceptive information through alternative sources such as health care providers. mMitra subscribers do not exist in isolation, and there are several information sources they can access to improve their health literacy during pregnancy and the postpartum period that may lead to a drop in their active engagement with the mMitra program. We do not know what caused women to disengage. A greater understanding of how to tailor the mode of delivery, and how much of the mMitra program women are willing to receive may help build on the theories of tailored and personalized messaging to activate motivation for health behavior change, which reflects a key CMOC developed and refined through the findings of the accompanying realist review.

XSL•FO
RenderX

We assessed exposure to the intervention as a continuous variable; however, there may not be a linear relationship between listening to the program and increase in knowledge. There may be a threshold at which increased exposure to the mMitra program does not affect PPFP awareness, or, conversely, a minimum level of exposure below which no effect is observed. A dose-response relationship may still exist on either side of a threshold in mMitra call duration. We had no reason to suspect a threshold effect, as each message delivered content on different topics. Messages were "drip-fed" to participants, and listening to complete messages was necessary to receive all planned information. Future research on patterns of engagement with the mMitra program could provide better insight into the dose of program content required to fully engage women and yield improvements in PPFP awareness and outcomes.

The analysis used two proxy variables for mMitra program engagement: (1) total mMitra call duration and (2) duration of mMitra calls on FP. However, neither of these variables captured participants' active listening to the call content. The measurement of participants' completion of each message and whether "dropped" calls were reinitiated by participants could have added to the depth of analysis on participants' true engagement with the program. The mMitra program's delivery system, using one-way automated voice-messaging, may contribute to our inability to assess the accuracy of the theorized CMOCs when applied to real-time data. For instance, incorporating more advanced conversational or bidirectional messaging channels in the mMitra program may aid in assessing women's interactions with the program content, as seen in recent interventions [43]. For instance, the PROMPTS SMS platform designed and delivered by Jacaranda Health in 2023 used AI capabilities to enable a conversational experience between participants and the interventions. This allowed for triaging of concerns and offering personalized solutions in real time. While the mMitra program offers a similar telephone help desk and reminds subscribers of the availability of this service at the end of each message, this service can only be accessed by subscribers who feel motivated to call the number and converse with someone regarding their concerns. Measuring these elements and assessing their associations with program engagement outcomes could strengthen the methodological approach of SEM within realist evaluation.

A large percentage of women (~71%) who were enrolled at baseline did not complete the second follow-up survey, which reduced the statistical power for the analysis. Call duration information is available for the nonrespondents and shows that this loss to follow-up occurred across the 6-month period. There were no significant differences between those who responded to the second survey and nonrespondents with respect to age, educational attainment, caste, or religion. Women who did not complete the second survey had listened to fewer minutes of calls and may have lost interest in the mMitra program (only 128 out of 683 women who did not respond to the end line survey listened to the last scheduled mMitra call). Women who enrolled in the program at a later gestational age were less likely to have responded to the follow-up survey, perhaps because they were occupied with newborn care duties. Among those who responded to the follow-up survey, there was no difference in engagement by gestational age at enrollment. Those lost to follow-up may have differed from those who were retained in factors associated with accessing mobile phones (such as household income, cellular network infrastructure, and digital literacy) and FP knowledge, if these were not correlated with the demographic characteristics outlined in Table S1 in Multimedia Appendix 2. This might have biased the results if women who felt they had gained from the program were the most likely to take part in the follow-up survey. However, there are other, more neutral, explanations for the poor response in the follow-up survey. It is common for women to temporarily move to their natal home during late pregnancy leading up to delivery, and this may have led to nonresponse.

SEM provided evidence supporting the CMOC pertaining to gatekeeping of phone access and PPFP awareness as well as associations between phone ownership (context) and mMitra program engagement (outcome). Since the data used to build the SEMs were observational, we cannot conclusively determine a causal link between the tested intervention mechanisms and their potential to generate outcomes related to improvements in PPFP awareness or mMitra program engagement.

The loss to follow-up within 6 months of enrollment may be indicative of systemic issues with program engagement. Since there was no difference in demographics between the groups lost to follow-up and those who were retained, it is difficult to predict why women were less likely to continue engaging with the program even while residing in their natal homes. The loss to follow-up may also reflect women's lack of interest in engaging with the program. A gradual decrease in call duration per call during the final messages of the mMitra program may indicate waning interest or motivation to continue listening to the calls as women may perceive their or their child's risk of adverse health outcomes to be lower than during the pregnancy or initial postpartum period.

Comparison to Existing Literature and Implications for Future Research

Evidence from the literature suggests that repetitive information delivered through automated calls may lead to subscriber disinterest and disengagement from the program; call dropout rates steadily decreased in a sexual and reproductive health mHealth intervention for adolescents in Sierra Leone [44]. Alshurafa et al [45] also suggest that there is a disconnect between metrics used to measure doses received by participants of mHealth programs and the true "engagement" required to activate health behavior changes. The commentary by Alshurafa et al [45] on the connection between intervention engagement and health behavior change further suggests that there is still no consensus in the literature regarding the optimal dose of mHealth programs required to elicit health behavior change. Findings from the accompanying realist review theorized the importance of gestational stage-based messaging and tailored content to activate motivation to engage with an mHealth program and potentially improve PPFP awareness or awareness [42]. Future research could build on this study's findings by aiming to appropriately measure and identify a dose-response relationship between mMitra engagement and PPFP behaviors.

XSL•FO
RenderX

Lower mean call durations at the end of the mMitra call schedule potentially align with the results of the accompanying realist review, which suggests that the motivation of women who were lost to follow-up may have been activated if their enrollment in the program occurred earlier in their pregnancy to capitalize on their inherent high intrinsic motivation to seek health information [46-54]. Women may have enrolled in the program due to social pressure and expectations from trusted health professionals rather than intrinsic motivation or interest. This may have impacted the low mean call durations in the initial calls of the mMitra program. Longer call durations picking up later in pregnancy may correlate with increased anxieties in women regarding the health of their babies. Previous studies of nonuse attrition in an mHealth program targeting adolescent health behaviors suggest that participants' perceptions of the benefits of the program were a key determinant of continued engagement [55]. An evaluation conducted by Mauriello et al [56] of an mHealth program providing prenatal and postpartum care to women with low socioeconomic status in Rhode Island suggested that continued efforts to engage with participants through community outreach at health care centers could be an effective strategy to reduce dropout rates [56]. The mMitra program theoretically employs a similar "tech + touch" approach to encourage behavior change [15]. Perhaps, more frequent opportunities for interaction between mMitra subscribers and stakeholders could yield greater engagement with the program and potential behavior change. It would be helpful for future research on the mMitra program to also address the contexts in which the mechanism of subscriber disinterest is activated to yield the outcome of poor program engagement.

While this study tests CMOCs that are rooted in empirical evidence from the accompanying realist review, the findings could be strengthened if the realist evaluation approach were embedded within the cohort study design and data collection procedures. This type of data collection and analysis would also align with current debates on the use of quantitative indicators to estimate associations within CMOCs.

These findings contribute to the growing body of literature exploring the usefulness of structural equation modeling as a method to test and refine CMOCs in realist evaluations. Ford et al [57] used SEM to test CMOCs representing access to primary care facilities by socioeconomically disadvantaged older people living in rural communities. Although they concluded that SEM is a feasible and appropriate method for quantitative analysis in realist evaluations, they state that this analysis should form part of a mixed methods approach to realist evaluations of complex social phenomena. Abildgaard et al [58] used structural equation modeling to analyze data from a cluster randomized trial aimed at understanding the interactions between organizational contexts and intervention mechanisms in yielding improvements to the psychosocial work environment. They used qualitative interview data to complement their SEMs, thus concurring with the Ford et al [57] approach to using SEM in realist evaluations. As far as we know, there have not been any studies using structural equation modeling to investigate the mediating effects of mechanisms on mHealth interventions and their influence on PPFP outcomes. Based on the findings from this study, it is evident that SEM can be used to test predefined CMOCs that are built through primarily qualitative data or qualitatively analyzed empirical evidence from the literature. Collecting primary quantitative data that explicitly represent CMOC components would positively build on the strengths of SEM used in this study.

Strengths and Limitations

This analysis of cohort study survey data contributes to the growing body of evidence, supporting the inclusion of quantitative data to test and refine emerging program theories. A key strength of this evaluation is that it was informed by and empirically tested the validity of CMOCs from an accompanying realist review through evidence from an mHealth program implemented in an LMIC setting [42]. However, this study is not without its limitations. Overreliance on categorical data also limited model fit in many instances because these variables could not capture the variability in the data. Furthermore, insufficient availability of data from mMitra subscribers from locations other than Mumbai limited the ability to generalize conclusions on the effectiveness of the program. The duration of the study follow-up period, and the duration of the mMitra program by extension, may have contributed to the high loss to follow-up. The assessment of knowledge outcomes at a single time point is a notable limitation. In addition, it is important to note that engagement in the program was a proxy measurement, and it is unclear whether the call-duration measure captured participants' attention to, and comprehension of, the delivered messages.

Conclusions

To our knowledge, this is the first realist evaluation to be directly informed by a realist review of mHealth program implementation in LMICs, addressing FP knowledge and awareness. Findings from the quantitative study suggest a modest impact but highlight the importance of phone access as a contextual factor influencing mMitra's effectiveness. Future research on the mMitra program should focus on testing additional CMOCs from the literature to assess their relevance and impact, especially when scaling across diverse cultural and geographic contexts.

Generative AI Use

The authors declare that generative AI (GenAI) tools (Microsoft CoPilot and GPT-6) were used in the research and writing process. According to the GAIDeT taxonomy (2026), the following tasks were delegated to GenAI tools under full human supervision: proofreading and editing. Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes.

XSL•FO
RenderX

Funding

The authors declared no financial support was received for this work.

Data Availability

The datasets generated or analyzed during this study are not publicly available currently as they are being processed for storage in a data repository. Data are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: AC
Data curation: AC
Formal analysis: AC
Funding acquisition: AC
Investigation: AC
Methodology: AC
Project administration: AC
Resources: NM
Software: AC
Supervision: ES, JF, NM, EW, CF
Validation: ES, JF, NM, EW, CF
Visualization: AC
Writing – original draft: AC
Writing – review & editing: AC, ES, JF, NM, EW, CF

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data and variables used to create a composite “mobile phone access” variable.
[DOCX File, 582 KB - humanfactors_v13i1e97650_app1.docx]

Multimedia Appendix 2
Construction of composite mobile phone access variable.
[DOCX File, 16 KB - humanfactors_v13i1e97650_app2.docx]

Multimedia Appendix 3
Time follow-up survey respondents spent listening to mMitra calls and sections of calls containing family planning messages, by sociodemographic characteristics.
[DOCX File, 19 KB - humanfactors_v13i1e97650_app3.docx]

Multimedia Appendix 4
Adjusted odds ratios for the association between total mMitra call duration and postpartum family planning awareness.
[DOCX File, 15 KB - humanfactors_v13i1e97650_app4.docx]

References

1. Devaraj K, Gausman J, Mishra R, Kumar A, Kim R, Subramanian SV. Trends in prevalence of unmet need for family planning in India: patterns of change across 36 states and union territories, 1993-2021. Reprod Health 2024 Apr 9;21(1):48. [doi: 10.1186/s12978-024-01781-6] [Medline: 38594726]
2. Murthy N, Chandrasekharan S, Prakash MP, et al. Effects of an mHealth voice message service (mMitra) on maternal health knowledge and practices of low-income women in India: findings from a pseudo-randomized controlled trial. BMC Public Health 2020 Jun 1;20(1):820. [doi: 10.1186/s12889-020-08965-2] [Medline: 32487065]
3. Murthy N, Chandrasekharan S, Prakash MP, et al. The impact of an mHealth voice message service (mMitra) on infant care knowledge, and practices among low-income women in India: findings from a pseudo-randomized controlled trial. Matern Child Health J 2019 Dec;23(12):1658-1669. [doi: 10.1007/s10995-019-02805-5] [Medline: 31584144]
4. Sully EA, Biddlecom A, Darroch JE, et al. Adding it up: investing in sexual and reproductive health 2019. : Guttmacher Institute; 2020 URL: https://www.guttmacher.org/sites/default/files/report_pdf/adding-it-up-investing-in-sexual-reproductive-health-2019.pdf [accessed 2026-10-29]

XSL•FO
RenderX

5. World population dashboard. United Nations Population Fund. 2023. URL: https://www.unfpa.org/data/world-population-dashboard [accessed 2026-10-29]
6. Singh SK, Kashyap GC, Sharma H, Mondal S, Legare CH. Changes in discourse on unmet need for family planning among married women in India: evidence from NFHS-5 (2019-2021). Sci Rep 2023 Nov 22;13(1):20464. [doi: 10.1038/s41598-023-47191-9] [Medline: 37993605]
7. Yadav K, Agarwal M, Shukla M, Singh JV, Singh VK. Unmet need for family planning services among young married women (15-24 years) living in urban slums of India. BMC Womens Health 2020 Sep 3;20(1):187. [doi: 10.1186/s12905-020-01010-9] [Medline: 32883262]
8. Hall MAK, Stephenson RB, Juvekar S. Social and logistical barriers to the use of reversible contraception among women in a rural Indian village. J Health Popul Nutr 2008 Jun;26(2):241-250. [Medline: 18686557]
9. Ram F, Shekhar C, Chowdhury B. Use of traditional contraceptive methods in India & its socio-demographic determinants. Indian J Med Res 2014 Nov;140 Suppl(Suppl 1):S17-S28. [Medline: 25673538]
10. Connected women: the mobile gender gap report 2021. : GSM Association (GSMA); 2021 URL: https://www.gsma.com/wp-content/uploads/2025/12/The-Mobile-Gender-Gap-Report-2021.pdf [accessed 2026-10-29]
11. mHealth: new horizons for health through mobile technologies: second global survey on ehealth. : World Health Organization; 2011 URL: https://iris.who.int/server/api/core/bitstreams/ad1b13c0-7c82-47b4-8dd5-f0a26c3a3cc3/content [accessed 2026-10-29]
12. Jennings L, Gagliardi L. Influence of mHealth interventions on gender relations in developing countries: a systematic literature review. Int J Equity Health 2013 Oct 16;12:85. [doi: 10.1186/1475-9276-12-85] [Medline: 24131553]
13. L'Engle K, Vadhat H. Mobile phone interventions for reproductive health (m4rh): testing the feasibility of text messaging to improve family planning. : Family Health International (FHI); 2009 URL: https://www.fhi360.org/wp-content/uploads/drupal/documents/formative_results_m4rh.pdf [accessed 2026-10-29]
14. LeFevre AE, Shah N, Scott K, et al. The impact of a direct to beneficiary mobile communication program on reproductive and child health outcomes: a randomised controlled trial in India. BMJ Glob Health 2022 Jul;6(Suppl 5):e008838. [doi: 10.1136/bmjgh-2022-008838] [Medline: 35835477]
15. mMITRA. ARMMAN. URL: https://www.armman.org/mmitra [accessed 2026-10-29]
16. Mechael P, Kaonga NN, Chandrasekharan S, et al. The elusive path toward measuring health outcomes: lessons learned from a pseudo-randomized controlled trial of a large-scale mobile health initiative. JMIR mHealth uHealth 2019 Aug 21;7(8):e14668. [doi: 10.2196/14668] [Medline: 31436165]
17. Lessons from country programs implementing the Mobile Alliance for maternal action programs in Bangladesh, South Africa, India and Nigeria, 2010–2016. : USAID; 2017 URL: https://www.ictworks.org/wp-content/uploads/2018/09/MAMA-Full-Report.pdf [accessed 2026-10-29]
18. A strategic approach to reproductive, maternal, newborn, child and adolescent health (RMNCH+A) in India. : Ministry of Health and Family Welfare, Government of India; 2013 URL: https://nhm.gov.in/images/pdf/RMNCH%2BA/RMNCH%2BA_Strategy.pdf [accessed 2026-10-29]
19. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061. [doi: 10.1136/bmj.n2061] [Medline: 34593508]
20. Kabongo EM, Mukumbang FC, Delobelle P, Nicol E. Understanding the influence of the MomConnect programme on antenatal and postnatal care service utilisation in two South African provinces: a realist evaluation protocol. BMJ Open 2019 Jul 1;9(7):e029745. [doi: 10.1136/bmjopen-2019-029745] [Medline: 31266842]
21. Pawson R, Tilley N. Realistic Evaluation: SAGE Publications; 1997.
22. Ewerling F, Wehrmeister FC, Victora CG, Raj A, McDougal L, Barros AJ. Is women's empowerment associated with coverage of RMNCH interventions in low- and middle-income countries? An analysis using a survey-based empowerment indicator, the SWPER. J Glob Health 2021 Mar 1;11:04015. [doi: 10.7189/jogh.11.04015] [Medline: 33791094]
23. Kunjumen T, Okech M, D, Asamani JA, Mohamed N, Nuruzzaman M. Multi-country case studies on planning RMNCH services using WISN methodology: Bangladesh, Ghana, Kenya, Sultanate of Oman and Papua New Guinea. Hum Resour Health 2022 Jan 28;19(Suppl 1):155. [doi: 10.1186/s12960-021-00671-3] [Medline: 35090493]
24. Bonell C, Fletcher A, Morton M, Lorenc T, Moore L. Realist randomised controlled trials: a new approach to evaluating complex public health interventions. Soc Sci Med 2012 Dec;75(12):2299-2306. [doi: 10.1016/j.socscimed.2012.08.032] [Medline: 22989491]
25. Bonell C, Warren E, Fletcher A, Viner R. Realist trials and the testing of context-mechanism-outcome configurations: a response to Van Belle et al. Trials 2016 Oct 1;17(1):478. [doi: 10.1186/s13063-016-1613-9] [Medline: 27716303]
26. Van Belle S, Wong G, Westhorp G, et al. Can "realist" randomised controlled trials be genuinely realist? Trials 2016 Jul 7;17(1):313. [doi: 10.1186/s13063-016-1407-0] [Medline: 27387202]
27. Bhaskar R. A Realist Theory of Science: Harvester Press; 1978.
28. What is a mechanism? What is a programme mechanism? : RAMESES Project; 2017 URL: https://ramesesproject.org/media/RAMESES_II_What_is_a_mechanism.pdf [accessed 2026-10-29]
29. Jagosh J. Retroductive theorizing in Pawson and Tilley's applied scientific realism. J Crit Realism 2020 Mar 14;19(2):121-130. [doi: 10.1080/14767430.2020.1723301]

XSL•FO
RenderX

30. Wong G, Westhorp G, Manzano A, Greenhalgh J, Jagosh J, Greenhalgh T. RAMESES II reporting standards for realist evaluations. BMC Med 2016 Jun 24;14(1):96. [doi: 10.1186/s12916-016-0643-1] [Medline: 27342217]
31. Bhatta DN, Aryal UR. Paternal factors and inequity associated with access to maternal health care service utilization in Nepal: a community-based cross-sectional study. PLoS One 2015;10(6):e0130380. [doi: 10.1371/journal.pone.0130380] [Medline: 26107621]
32. Hassen SS, Lelisho ME. Determining factors associated with the prevalence of knowledge, attitude, and practice in seeking skilled maternal healthcare services among women in a remote area of Gesha district. BMC Health Serv Res 2022 Nov 4;22(1):1318. [doi: 10.1186/s12913-022-08710-y] [Medline: 36329430]
33. Mekonnen BD, Gelagay AA, Lakew AM. Knowledge and associated factors of postpartum contraceptive use among women in the extended postpartum period in Gondar City, Northwest Ethiopia. Open Access J Contracept 2021;12:7-15. [doi: 10.2147/OAJC.S290337] [Medline: 33531846]
34. Mehare T, Mekuriaw B, Belayneh Z, Sharew Y. Postpartum contraceptive use and its determinants in Ethiopia: a systematic review and meta-analysis. Int J Reprod Med 2020;2020:5174656. [doi: 10.1155/2020/5174656] [Medline: 31970195]
35. Okedo-Alex IN, Akamike IC, Ezeanosike OB, Uneke CJ. Determinants of antenatal care utilisation in sub-Saharan Africa: a systematic review. BMJ Open 2019 Oct 7;9(10):e031890. [doi: 10.1136/bmjopen-2019-031890] [Medline: 31594900]
36. Sharma V, Mohan U, Das V, Awasthi S. Socio demographic determinants and knowledge, attitude, practice: survey of family planning. J Family Med Prim Care 2012 Jan;1(1):43-47. [doi: 10.4103/2249-4863.94451] [Medline: 24479000]
37. Wakuma B, Mosisa G, Etafa W, et al. Postpartum modern contraception utilization and its determinants in Ethiopia: a systematic review and meta-analysis. PLoS One 2020;15(12):e0243776. [doi: 10.1371/journal.pone.0243776] [Medline: 33315904]
38. Johnson S, Johnson R. Conceptualising and interpreting reliability. : Office of Qualifications and Examinations Regulation (Ofqual); 2009 URL: https://dera.ioe.ac.uk/id/eprint/11077/1/2010-02-05-conceptualising-and-interpreting-reliability.pdf [accessed 2026-10-29]
39. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. Res Sci Educ 2018 Dec;48(6):1273-1296. [doi: 10.1007/s11165-016-9602-2]
40. Barbera J, Naibert N, Komperda R, Pentecost TC. Clarity on Cronbach's alpha use. J Chem Educ 2021 Feb 9;98(2):257-258. [doi: 10.1021/acs.jchemed.0c00183]
41. Stata. URL: https://www.stata.com/ [accessed 2026-10-29]
42. Chandrasekar A, Warren E, Free C, et al. mHealth interventions for postpartum family planning in LMICs: a realist review. PLOS Glob Public Health 2024;4(7):e0003432. [doi: 10.1371/journal.pgph.0003432] [Medline: 39024319]
43. Vatsa R, Chang W, Akinyi S, et al. Impact evaluation of a digital health platform empowering Kenyan women across the pregnancy-postpartum care continuum: a cluster randomized controlled trial. PLoS Med 2025 Feb;22(2):e1004527. [doi: 10.1371/journal.pmed.1004527] [Medline: 39899612]
44. Chukwu E, Gilroy S, Dickson KE. Project design and technology trade-offs for implementing a large-scale sexual and reproductive health mHealth intervention: lessons from Sierra Leone. Front Digit Health 2023;5:1060376. [doi: 10.3389/fdgth.2023.1060376] [Medline: 36994145]
45. Alshurafa N, Jain J, Alharbi R, Iakovlev G, Spring B, Pfammatter A. Is more always better?: discovering incentivized mHealth intervention engagement related to health behavior trends. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Dec;2(4). [doi: 10.1145/3287031] [Medline: 32318650]
46. Dyer JD. Maternal health behavior change: women's experiences as participants of an mHealth program in Timor-Leste [Master's thesis]. : University of Washington, School of Public Health; 2015 URL: https://digital.lib.washington.edu/server/api/core/bitstreams/00de64d7-9676-4c5b-aa8f-23b1cb029b62/content [accessed 2026-10-29]
47. Frenn M, Malin S, Bansal NK. Stage-based interventions for low-fat diet with middle school students. J Pediatr Nurs 2003 Feb;18(1):36-45. [doi: 10.1053/jpdn.2003.6] [Medline: 12610786]
48. Jones RM, Kimenju G, Subbiah S, Styles A, Pearson N, Rajasekharan S. A short message service (SMS) increases postpartum care-seeking behavior and uptake of family planning of mothers in peri-urban public facilities in Kenya. PLoS One 2020;15(9):e0239213. [doi: 10.1371/journal.pone.0239213] [Medline: 32997684]
49. Unger JA, Ronen K, Perrier T, et al. Short message service communication improves exclusive breastfeeding and early postpartum contraception in a low- to middle-income country setting: a randomised trial. BJOG 2018 Nov;125(12):1620-1629. [doi: 10.1111/1471-0528.15337] [Medline: 29924912]
50. Lawrence WT, Haslam C. Smoking during pregnancy: where next for stage-based interventions? J Health Psychol 2007 Jan;12(1):159-169. [doi: 10.1177/1359105307071750] [Medline: 17158849]
51. Maslowsky J, Frost S, Hendrick CE, Trujillo Cruz FO, Merajver SD. Effects of postpartum mobile phone-based education on maternal and infant health in Ecuador. Int J Gynaecol Obstet 2016 Jul;134(1):93-98. [doi: 10.1016/j.ijgo.2015.12.008] [Medline: 27126905]
52. McConnell M, Rothschild CW, Ettenger A, Muigai F, Cohen J. Free contraception and behavioural nudges in the postpartum period: evidence from a randomised control trial in Nairobi, Kenya. BMJ Glob Health 2018;3(5):e000888. [doi: 10.1136/bmjgh-2018-000888] [Medline: 30364345]

XSL•FO
RenderX

53. O'Brien OA, Lindsay KL, McCarthy M, et al. Influences on the food choices and physical activity behaviours of overweight and obese pregnant women: a qualitative study. Midwifery 2017 Apr;47:28-35. [doi: 10.1016/j.midw.2017.02.003] [Medline: 28236733]
54. PROMPTS: a personalized digital health navigator for new and expecting mothers. : Jacaranda Health; 2024 URL: https://jacarandahealth.org/ypoagriw/2024/08/PROMPTS_Jacaranda-Health.pdf [accessed 2026-10-29]
55. Maenhout L, Peuters C, Cardon G, Crombez G, DeSmet A, Compernolle S. Nonusage attrition of adolescents in an mHealth promotion intervention and the role of socioeconomic status: secondary analysis of a 2-arm cluster-controlled trial. JMIR mHealth uHealth 2022 May 10;10(5):e36404. [doi: 10.2196/36404] [Medline: 35536640]
56. Mauriello LM, Van Marter DF, Umanzor CD, Castle PH, de Aguiar EL. Using mHealth to deliver behavior change interventions within prenatal care at community health centers. Am J Health Promot 2016 Sep;30(7):554-562. [doi: 10.4278/ajhp.140530-QUAN-248] [Medline: 26305603]
57. Ford JA, Jones A, Wong G, Clark A, Porter T, Steel N. Access to primary care for socio-economically disadvantaged older people in rural areas: exploring realist theory using structural equation modelling in a linked dataset. BMC Med Res Methodol 2018 Jun 19;18(1):57. [doi: 10.1186/s12874-018-0514-x] [Medline: 29914411]
58. Abildgaard JS, Nielsen K, Wåhlin-Jacobsen CD, Maltesen T, Christensen KB, Holtermann A. 'Same, but different': a mixed-methods realist evaluation of a cluster-randomized controlled participatory organizational intervention. Hum Relat 2020 Oct;73(10):1339-1365. [doi: 10.1177/0018726719866896]

Abbreviations

CMOC: context-mechanism-outcome configuration
FP: family planning
HEC: highly effective contraceptive
LMIC: low- and middle-income country
mHealth: mobile health
OR: odds ratio
PPFP: postpartum family planning
RAMESES II: Realist and Meta-Narrative Evidence Syntheses: Evolving Standards II
SEM: structural equation model
WHO: World Health Organization

Edited by A Kushniruk; submitted 09.Apr.2026; peer-reviewed by A Lakdawala, P Katyal; revised version received 31.Aug.2026; accepted 02.Oct.2026; published 18.Nov.2026.

Please cite as:
Chandrasekar A, Slaymaker E, Ford J, Madhiwalla N, Warren E, Free C, McCarthy OL
Mobile Health for Postpartum Contraception Knowledge in India (mMitra): Realist Evaluation Using Moderation Analysis and Structural Equation Models
JMIR Hum Factors 2026;13:e97650
URL: https://humanfactors.jmir.org/2026/1/e97650
doi:10.2196/97650

© Abinaya Chandrasekar, Emma Slaymaker, John Ford, Neha Madhiwalla, Emily Warren, Caroline Free, Ona L McCarthy. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 18.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Long-Term Effects of Multimedia Education and "Foot Alerts" Through the MyU App on Diabetic Foot Care in Individuals With Moderate-High Risk: Pilot Randomized Controlled Trial

Ameenah Alawadhi[1], BSc, MSc, PhD; Kay Scarsbrook Khan[2,3], BSc; Grace Messenger[2,4], BSc; Alice Cusworth[2], BSc; Mohammad Assi[5], BSc; Stuart R Gray[6], BSc, MSc, PhD; Ebaa Al-Ozairi[1], BMedSci, MBChB, MRCPE, CCT, FACN, FACE

[1]Clinical Care Research and Trials Unit, Dasman Diabetes Institute, P.O. Box 1180, Dasman 15462, Kuwait City, Kuwait

[2]Podiatry Unit, Dasman Diabetes Institute, Kuwait City, Kuwait

[3]Podiatry Unit, Harrogate and District NHS Foundation Trust, North Yorkshire, United Kingdom

[4]Podiatry Unit, North Cumbria Integrated Care NHS Foundation Trust, Cumbria, United Kingdom

[5]Technology Department, Dasman Diabetes Institute, Kuwait City, Kuwait

[6]School of Cardiovascular and Metabolic Health, University of Glasgow, Glasgow, United Kingdom

Corresponding Author:
Ebaa Al-Ozairi, BMedSci, MBChB, MRCPE, CCT, FACN, FACE
Clinical Care Research and Trials Unit, Dasman Diabetes Institute, P.O. Box 1180, Dasman 15462, Kuwait City, Kuwait

Abstract

Background: Diabetic foot ulceration (DFU) is the leading cause of nontraumatic amputations in people with diabetes. Research shows that improving patient awareness can result in short-term improvements, but Cochrane reviews report insufficient high-quality evidence.

Objective: This study aims to investigate the effects of multimedia presentation and smartphone alerts to enhance long-term knowledge and foot care behaviors in individuals at moderate-to-high risk of DFU.

Methods: Participants were randomized to a control group, receiving usual diabetic foot care advice (n=40), or an intervention group, receiving a multimedia diabetic foot care presentation and regular "foot alerts" through the MyU smartphone app on top of usual care (n=37). Patient's knowledge and behaviors related to diabetic foot care were assessed at baseline and after 12 months. Repeated measures ANOVA was conducted in both intention-to-treat and per-protocol analyses to evaluate the intervention's effectiveness.

Results: The findings were consistent across intention-to-treat and per-protocol analyses. In the intervention group, the number of podiatry visits was positively correlated with improved foot care behavior (r=0.408; P=.02), while the control group showed a negative correlation (r=−0.402; P=.02). No significant correlations were observed with knowledge scores. Although no significant time×group interactions were seen, the main effects of time were found for both knowledge (η^2=0.12; P=.004) and behavior scores (η^2=0.31; P<.001). Post hoc analysis showed a decline in knowledge scores in the control group (Cohen d=−0.24; P=.007) and improvements in behavior scores in both groups (Cohen d: intervention=0.61, control=0.63; all P<.001).

Conclusions: The MyU app−based multimedia intervention was associated with improved foot care behaviors over 12 months, indicating potential benefits as an adjunct to usual diabetic foot care. However, no significant changes in diabetic foot care knowledge were observed. These findings suggest that while the applied digital multimedia tool may support behavior change, further research is needed to enhance knowledge retention and clinical impact. The study revealed that multimedia education alone may not be effective for long-term improvement in foot self-care knowledge and behavior among individuals at moderate-high DFU risk, but the reinforcement of educational material during follow-up podiatry visits could be effective.

Trial Registration: ClinicalTrials.gov NCT03934944; https://www.clinicaltrials.gov/study/NCT03934944

(JMIR Hum Factors 2026;13:e78261) doi:10.2196/78261

KEYWORDS

behavior; DFU; diabetic foot ulceration; foot care; multimedia education; smartphone

XSL•FO
RenderX

Introduction

The escalating incidence of diabetes worldwide has turned it into a global pandemic impacting over 500 million individuals [1]. Health costs associated with diabetes are expected to reach US $1054 billion by 2045, up from US $966 billion in 2021 [2], with 80% of this spent on managing avoidable complications [3]. One of the most common long-term complications of diabetes is diabetic foot ulceration (DFU), with a lifetime risk of developing a DFU of between 12% and 25% [4-6]. DFU continues to be the leading cause of nontraumatic lower limb amputation [7], accounting for 84% of lower limb amputations [8]. Studies have shown that people with diabetes are 10 - 30 times more likely to have a lower limb amputation than someone without diabetes [9,10]. Furthermore, Armstrong et al [11] reported a 5-year mortality rate in people with DFU of as much as 43% at 1 year, worse than some of the most common cancers, and postamputation rates have shown that only 50% of patients survive the first year following amputation [12].

The early detection and treatment of minor foot injuries have been shown to reduce amputation rates by between 49% and 85% [13], yet amputation rates continue to rise [14]. Improved patient awareness and correct foot self-care practices can interrupt the amputation pathway [6]. Although systematic reviews have demonstrated the efficacy of foot care education in enhancing self-care behaviors and knowledge [15-17], there are insufficient data to support its clinical use in reducing ulceration and amputation incidence [18,19]. Furthermore, the delivery of education sessions is particularly difficult for those with limited literacy, and debates are ongoing about whether video-based or printed materials are most effective [20,21]. Comprehensive foot care education, which encompasses assessment, discussion, counseling, home visits, and telephone calls, has been proposed to enhance diabetic foot care and educational programs [18]. However, a systematic review of 6 randomized controlled trials (RCTs) revealed insufficient evidence to support its efficacy [22]. More recently, it has been suggested that mobile apps can effectively prevent DFU recurrence [23] and improve diabetic foot care outcomes by incorporating at least 1 information communication tool [24]. However, many of the RCTs focused on patients at low risk, limiting their utility to support improved clinical outcome in the measure of reducing the rate of ulceration [18,19,22].

Given these findings and the estimated financial savings in the prevention of DFU, there is a need to focus on its prevention [25]. This study, therefore, aims to (1) establish if the use of a multimedia presentation and weekly alerts from a smartphone app as an education tool improves long-term knowledge and foot care behaviors of patients with diabetes when compared to usual care, (2) assess the association between the change in participants' knowledge and behavior and the number of podiatry appointments over the study period, and (3) compare the number of new DFUs and hemoglobin A_{1c} (HbA_{1c}) levels between the 2 study groups.

Methods

Study Design

This randomized, investigator-blinded, 2-armed, pilot study was conducted on patients with diabetes who visited Dasman Diabetes Institute (Kuwait City, Kuwait) clinics from January 2019 to May 2024. Each participant was fully informed about the study prior to giving their written informed consent.

Eligibility Criteria

Participants had to meet the following inclusion criteria: (1) adult patients (≥18 years of age) with type 2 diabetes or adults with type 1 diabetes for ≥5 years, (2) be of medium or high DFU risk, defined as having at least 1 diabetic foot risk factor (loss of pain perception, peripheral vascular disease, foot deformity, history of DFU, or amputation), (3) able to understand study procedures and comply with them for the entire length of the study, and (4) must own a smartphone with internet access and agree to have the phone app uploaded for the duration of the study. People with chronic kidney disease, cognitive impairment, active psychiatric illness, inability to give written informed consent, hearing or visual impairment, or phone app inaccessibility for more than 4 weeks were excluded from the study.

Procedures

Overview

The study consisted of 2 main visits (at baseline and 12 mo) and up to 6 interim follow-up visits during the 1-year study period. A total of 98 participants were randomized by an independent researcher to either the control (“Usual Care” section) or intervention (“Multimedia Educational Intervention” section) arm in a 1:1 ratio. Participants were instructed not to disclose any information about their assigned group to the research team to maintain confidentiality. All written information for this study went through a process of back translation from English to Arabic, focusing on conceptual rather than literal translations and using natural language for a broad audience [26]. Prior to the intervention period, participants underwent diabetic foot assessment and were classified based on their diabetic foot risk as per NICE (National Institute for Health and Care Excellence) guidelines [27], with participants with moderate risk attending 2 additional follow-up visits at 6-month intervals and participants with high risk attending 6 additional follow-up visits at 2-month intervals. At the baseline visit, HbA_{1c} was measured, and participants received their assigned diabetic foot care education route. Then, they were asked to complete a diabetic foot care knowledge questionnaire and a foot care and diabetes self-care behaviors questionnaire. Interim follow-up visits included a visual inspection of feet, verbal advice in the usual manner (for both groups), and an educational audio-visual prerecorded presentation (for intervention group only). At the 12-month visit, all participants completed a foot care knowledge questionnaire, a foot care and diabetes self-care behaviors questionnaire, and underwent an HbA_{1c} test.

XSL•FO
RenderX

Usual Care

The control group received usual care and the usual route of education with educational leaflets to take home. Participants received verbal advice during their podiatry appointment about daily foot inspections, wearing well-fitting shoes, what to do in an emergency, and the use of emollient. More specific advice was given depending on the participant's risk classification. This usual route involved information by the podiatrist translated by a bilingual but not native Arabic-speaking nurse.

Multimedia Educational Intervention

The intervention group received an educational presentation and a smartphone app providing weekly foot alerts for the duration of the study, in addition to the usual care, education route, and leaflet. The details of the MyU app interface, including the layout and key features used to deliver educational content, are illustrated in Figure 1. The educational presentation was a 10-minute audio-visual prerecorded presentation based on the internationally recognized advice on preventing diabetic foot complications including good diabetes control, daily foot inspections, and footwear advice [27-29]. The research podiatry clinic allowed only 1 participant to attend at a time, ensuring privacy for viewing the educational video. In addition, a teaching- or education-based smartphone app called MyU was installed onto their smartphones, allowing educational information to be uploaded and delivered remotely. The app also remotely recorded the number of times the content was accessed by each participant. Adherence to the intervention was monitored based on these access logs, where each "access" was defined as opening the app and viewing the educational content. However, more detailed engagement metrics such as video completion or time spent on content were not captured. Adherence (%) was calculated as the percentage of scheduled videos viewed, based on the total video views recorded by the MyU app for each participant. Both the absolute number of videos viewed (count) and the corresponding percentage adherence were reported. This app also delivered a weekly "foot-alert" (notification) to remind the participant to view the uploaded educational video. There were 12 unique educational videos "alerts" repeated at 3-month cycles. Repeating the educational method has been shown to improve health literacy [30]. All foot alerts were uploaded in both Arabic and English.

XSL•FO
RenderX

Figure 1. MyU app interface showing the layout and key features used to deliver educational content. The home screen provides access to 12 educational videos, which were assigned to participants according to their recruitment date. Weekly push notifications (shown at the bottom of the screen) prompted users to engage with the app and view the assigned videos displayed on the Newsfeed page.



Foot Care Knowledge Questionnaire

The foot care knowledge questionnaire used in this study has been adapted from Pollock et al [31] and Rheeder et al [32]. The questionnaire covers 2 sections. The first section consisted of 7 questions with 3 options each: correct, incorrect, and don't know. A correct answer was awarded 1 point, while incorrect or "don't know" responses received 0 points. The second section consisted of 5 questions with multiple-choice options. A correct answer was awarded 1 point, while an incorrect answer received 0 points. The maximal total score is 12 points, with higher scores indicating a better understanding of foot care knowledge.

Foot Care and Diabetes Self-Care Behaviors Questionnaire

The foot care and diabetes self-care behaviors questionnaire has been adopted from the Summary of Diabetes Self-Care Activities (SDSCA). The SDSCA has been found to have both reliability and validity as a standard measure of diabetes self-management [33]. Furthermore, this test has been shown to be reliable and valid, giving consistent results when translated into Arabic in a sample size of 243 participants [34]. The SDSCA scale measures the frequency of each self-care activity in the last 7 days for 4 aspects related to diabetes routine: foot care (8 scaled questions), blood-glucose testing (2 scaled questions), medications adherence (2 scaled questions), and smoking habits (2 questions). For this analysis, the score of the scaled questions (from 0 to 7 d) was calculated as the average

XSL•FO
RenderX

of responses within each section [33]. A higher total score indicates better foot self-care behavior.

Sample Size

Although this is a pilot study, a sample size of 98 with a 1:1 ratio between intervention and control group (49:49) was calculated using G*power software (version 3.9.1.2; Heinrich-Heine-Universität Düsseldorf). An a priori difference between 2 independent means tests was performed to find a significance interaction with a power of 80% and probability of type 1 error of 0.05. The effect from a study by Baba et al [35] was a significant mean change of 1.8 (SD 2.6) reduction in foot score (based on the presence and severity of 15 podiatry disorders; n=78) compared to a 0.1 (SD 2.6) reduction in the foot score after receiving interactive education (n=76) over 3 months (effect size=0.58).

Statistical Analysis

Data analysis was performed using SPSS Statistics (version 29.0; IBM Corp). Descriptive statistics were represented as mean (SD) for continuous variables and as frequencies for categorical variables. The Shapiro-Wilk test was used to assess normality for each continuous variable to determine the appropriate use of parametric or nonparametric tests. Chi-square tests and independent 2-tailed *t* tests were conducted to test to examine baseline differences between the groups and to compare the frequency of new DFU cases between the study groups. The Wilcoxon signed-rank test was used when the assumption of normality was violated.

Both intention-to-treat (ITT) and per-protocol (PP) analyses were conducted. The ITT analysis included all randomized participants with the available outcome data, regardless of adherence or dropout status. A linear mixed model was used to evaluate group, time, and group×time interaction effects on the outcome measures: foot care knowledge scores, foot care, blood-glucose testing, and medication adherence behaviors.

Post hoc pairwise comparisons were conducted using the matched pair 2-tailed Student *t* test or Wilcoxon signed-rank test, with Bonferroni adjustment applied for multiple comparisons. Bivariate Pearson or Spearman correlation analyses were performed to evaluate the strength and direction of relationships between the change in foot care knowledge scores and foot self-care behavior: scores and the number of follow-up visits. Changes in scores were calculated as 12 month values − baseline values. *P* values <.05 were considered statistically significant.

Ethical Considerations

Ethical approval was granted by the Dasman Diabetes Institute Ethical Review Committee (RA HM-2018-044). The study was conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines and was registered at ClinicalTrials.gov (NCT03934944). The trial was conducted and reported in accordance with the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) 2011 guidelines (Checklist 1).

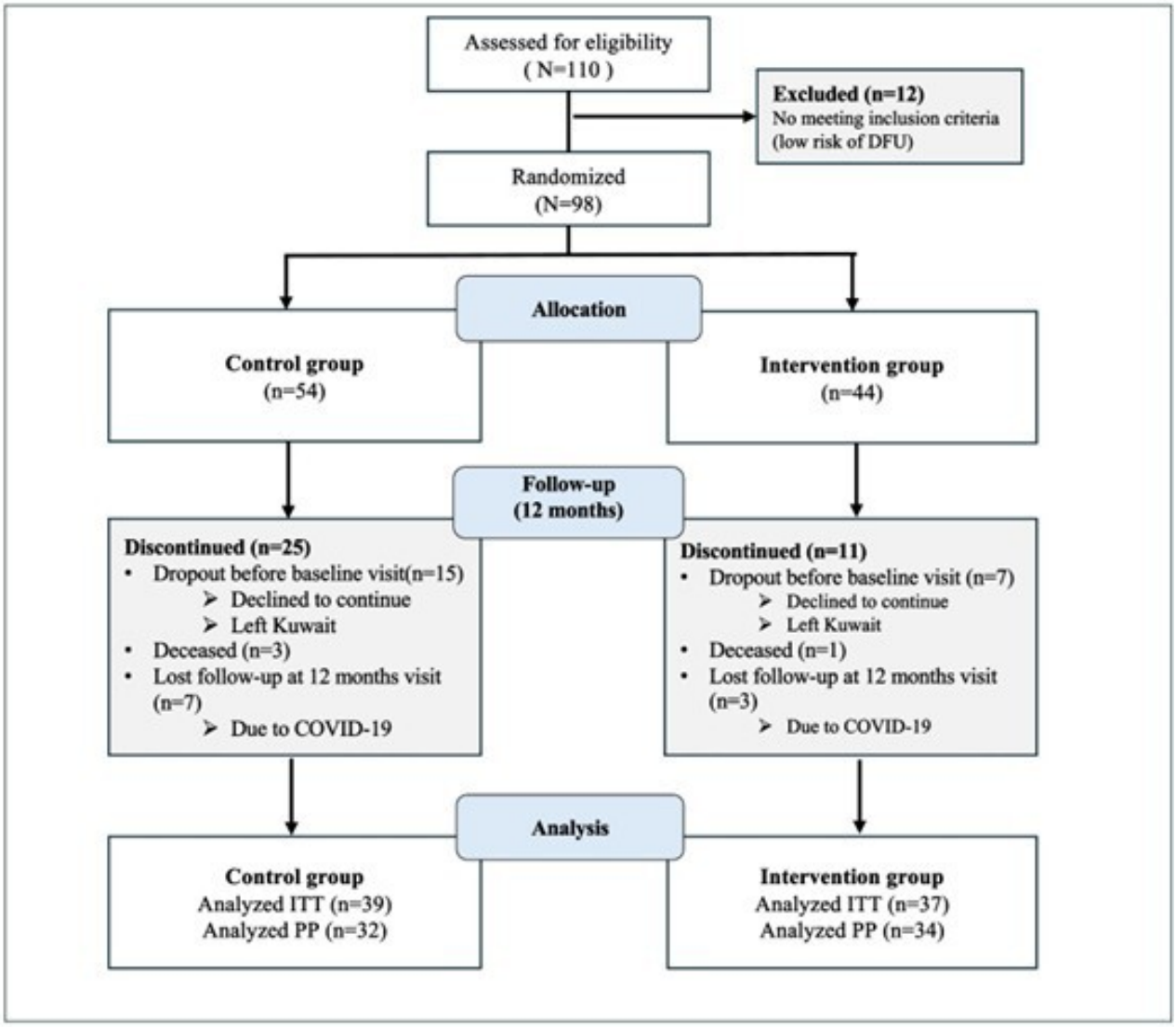
Written informed consent was obtained from all participants prior to enrollment. Participant privacy and confidentiality were strictly maintained throughout the study: all data were deidentified and stored on secure, password-protected servers accessible only to authorized research personnel, and no personally identifiable information was included in the analysis or reported in the manuscript. Participants did not receive financial compensation for their participation.

Results

A total of 66 participants completed the study visits and were included in the final PP analysis (Figure 2). Additionally, the ITT analysis included 77 participants who were randomized and had at least 1 outcome measure available. Table 1 shows the baseline characteristics of study participants in both ITT and PP populations. In the ITT population, the mean age of the participants was 60 (SD 7.85) years in the intervention group and 62 (SD 9.31) years in the control group. In both groups, the majority of the study participants were male participants, had attained an undergraduate degree, were nonsmokers, and had type 2 diabetes for over 20 years. In the intervention group, 75.7% (n=28) were classified as having a high DFU risk compared to a 62.5% (n=20) in the control group. All baseline characteristics were statistically comparable between the groups in both ITT and PP analyses, except for BMI, which showed a statistically significant difference between the groups in the ITT sample (*P*=.01) but not in the PP sample. Dropout analysis revealed that 8 participants from the control group (6 with high DFU risk and 2 with moderate risk) and 3 participants from the intervention group (all high risk) discontinued the study.

XSL•FO
RenderX

Figure 2. Participants flow chart. DFU: diabetic foot ulceration; ITT: intention-to-treat; PP: per-protocol.



XSL•FO
RenderX

Table . General characteristics according to study group and analysis carried out.

Variable	ITT[a]			PP[b]		
	Intervention (n=37)	Control (n=40)	*P* value	Intervention (n=34)	Control (n=32)	*P* value
Age (y), mean (SD)	60.43 (7.85)	62.39 (9.81)	.34	61.06 (7.87)	62.70 (9.41)	.44
BMI (kg/m^2), mean (SD)	34.05 (5.69)	36.94 (3.2)	.01[c]	34.05 (5.69)	32.09 (6.04)	.18
Sex, n (%)			.83			.65
Male	24 (65)	25 (63)		21 (62)	20 (63)	
Female	13 (35)	15 (38)		13 (38)	12 (38)	
Ethnicity, n (%)			.57			.49
Arabic	2 (5)	3 (8)		2 (6)	3 (9)	
Asian	0 (0)	1 (3)		0 (0)	1 (3)	
Kuwaiti	35 (95)	36 (90)		32 (94)	28 (88)	
Education level, n (%)			.23			.21
Junior school	8 (22)	2 (5)		6 (18)	1 (3)	
Senior school	9 (24)	11 (28)		9 (27)	7 (23)	
Undergraduate	14 (38)	22 (55)		14 (41)	19 (61)	
Postgraduate	7 (19)	4 (10)		5 (15)	4 (13)	
Smoking status, n (%)			.51			.21
Smoker	6 (16)	5 (13)		6 (18)	3 (9)	
Nonsmoker	26 (65)	23 (62)		26 (81)	21 (62)	
Ex-smoker	7 (21)	5 (13)		7 (21)	3 (9)	
Type of diabetes, n (%)			.58			.33
Type 1	3 (8)	2 (5)		3 (9)	1 (3)	
Type 2	34 (92)	38 (95)		31 (97)	31 (91)	
Duration of diabetes (y), n (%)			.82			.83
0 - 9	4 (11)	5 (13)		4 (13)	3 (9)	
10 - 19	11 (30)	14 (35)		9 (28)	11 (34)	
>20	22 (60)	21 (53)		19 (59)	18 (56)	
HbA_{1c}[d], mean (SD)	8.32 (1.87)	8.14 (1.61)	.65	8.29 (2.04)	8.17 (1.68)	.61
Overall ABPI[e], mean (SD)	1.17 (0.13)	1.18 (0.07)	.95	1.186 (0.21)	1.13 (0.14)	.56
Average VPT[f], mean (SD)	52.60 (118.3)	50.86 (113.7)	.94	53.43 (123.55)	55.84 (129.16)	.94
Risk factors, n (%)						
PAD[g]	7 (19)	6 (15)	.57	6 (18)	5 (16)	.82
LOPS[h]	33 (89)	37 (93)	.41	30 (88)	30 (92)	.43
Foot deformity	22 (60)	19 (48)	.13	21 (62)	16 (50)	.33
History of DFU[i]	18 (49)	16 (40)	.5	16 (47)	13 (41)	.59
History of amputation	7 (19)	8 (20)	.61	7 (21)	7 (22)	.89
History of Charcot	5 (14)	4 (10)	.56	4 (12)	3 (9)	.75
DFU risk classification, n (%)			.12			.13

XSL•FO
RenderX

Variable	ITT[a]		PP[b]	
High	28 (76)	25 (63)	25 (76)	18 (58)
Moderate	9 (24)	15 (38)	9 (24)	14 (42)

[a]ITT: intention-to-treat.

[b]PP: per protocol.

[c]*P value <.05

[d]HbA_{1c}: hemoglobin A_{1c}.

[e]ABPI: ankle-brachial pressure index.

[f]VPT: vibration perception threshold.

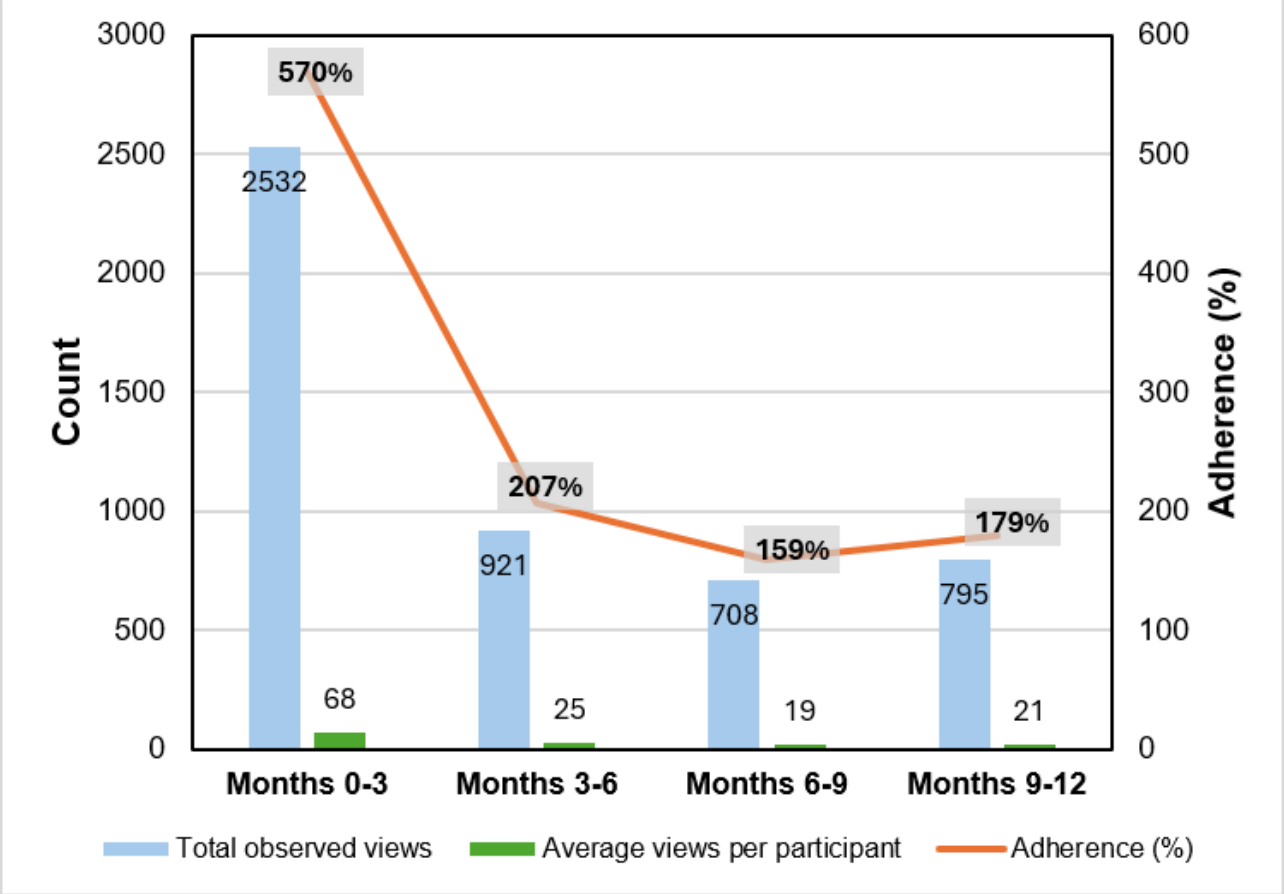
[g]PAD: peripheral artery disease.

[h]LOPS: diabetic peripheral neuropathy with loss of protective sensation.

[i]DFU: diabetic foot ulceration.

Adherence was estimated using the total video views recorded by the MyU app. The actual views exceeded the maximum expected views per interval (444 views), especially at baseline (mo 0 - 3: 2532 views, 570%), indicating repeated viewing. Engagement declined over time (mo 9 - 12: 795 views, 179%). These counts serve as a group-level proxy for adherence (Figure 3).

Figure 3. Total video views (blue bars), average views per participant (green bars), and adherence percentage (line) across study time points, based on MyU app data. Values represent group-level engagement with 12 videos per interval. Adherence (%) represents the percentage of the scheduled videos viewed, calculated from the total video views recorded by the MyU app. Both the absolute number of videos viewed (count) and the corresponding percentage are shown.



Both ITT and PP analyses revealed no significant time×group interactions for diabetic foot care knowledge and behavior scores. However, there were significant main effects of time for both outcomes (ITT: $F_{1,70}$=6.98, P=.01, partial η^2=0.09 for knowledge; $F_{1,70}$=28.36, P<.001, partial η^2=0.29 for behavior; PP: $F_{1,64}$=8.760; P=.004, partial η^2=0.12 for knowledge scores; $F_{1,64}$=28.139; P<.001, partial η^2=0.31 for behavior scores). Post hoc within-group comparisons showed a significant reduction in diabetic foot care knowledge scores (ITT: mean difference −0.878, SE 0.366 points, P value=.02, Cohen d=−0.21; PP: mean difference −1, SE 0.357 points, P value=.007, Cohen d=−0.22) in the control group, with no change in the intervention group (ITT: mean difference −0.45, SE 0.344

XSL•FO
RenderX

points, P=.2, Cohen d=−0.38; PP: mean difference −0.47, SE 0.346 points, $P_{t\,test}$=.18, Cohen d=−0.48). For diabetic foot care behavior, post hoc within-group comparisons showed a significant increase in both control (ITT: mean difference 1.129, SE 0.29 points, P<.001, Cohen d=0.62; PP: mean difference 1.109, SE 0.3 points, P<.001, Cohen d=0.61) and intervention (ITT: mean difference 1.139, SE 0.3 points, P<.001, Cohen d=0.60; PP: mean difference 1.121, SE 0.29 points, P<.001, Cohen d=0.63) groups. No interaction or main effects were seen in other self-care behaviors scores such as blood-glucose testing and medication adherence (Tables 2 and 3).

Table . Intention-to-treat (ITT) analysis: comparison of knowledge and behavior of diabetic foot self-care between the groups before and after the intervention (n=77; intervention group n=37 and control group n=40).

Group	Baseline, estimated mean (SE)	12 months, estimated mean (SE)	Within-group comparison, P value (Cohen d)	Time effect, P value (η^2)	Group effect, P value (η^2)	Interaction, P value (η^2)
Knowledge of diabetic foot care score						
Intervention	6.51 (0.31)	6.10 (0.29)	.2 (–0.38)	.01[a] (0.09)	.42 (0.01)	.39 (0.01)
Control	6.42 (0.27)	5.54 (0.36)	.02[a] (–0.21)	—[b]	—	—
Foot care and diabetes self-care behaviors						
Foot-care score						
Intervention	3.49 (0.19)	4.64 (0.36)	.001[a] (0.60)	<.001[a] (0.29)	.12 (0.04)	.98 (<0.001)
Control	3.04 (0.17)	4.16 (0.36)	.001[a] (0.62)	—	—	—
Blood-glucose testing score						
Intervention	4.77 (0.43)	5.16 (0.48)	.45 (–0.07)	.06 (0.05)	.07 (0.001)	.41 (0.009)
Control	4.30 (0.42)	5.34 (0.42)	.08 (0.03)	—	—	—
Medication adherence score						
Intervention	5.50 (0.35)	5.63 (0.42)	.83 (0.29)	.92 (0.0001)	.83 (0.0006)	.67 (0.002)
Control	5.60 (0.38)	5.38 (0.38)	.68 (0.12)	—	—	—

[a]P<.05.

[b]Em dashes indicate values that are identical for both groups because these results represent overall ANOVA model effects rather than group-specific estimates.

XSL•FO
RenderX

Table . Per-protocol (PP) analysis: comparison of knowledge and behavior of diabetic foot self-care between the groups before and after the intervention (n=66; intervention group n=34 and control group n=32).

Group	Baseline, estimated mean (SE)	12 months, estimated mean (SE)	Within-group comparison, *P* value (Cohen *d*)	Time effect, *P* value (η^2)	Group effect, *P* value (η^2)	Interaction, *P* value (η^2)
Knowledge of diabetic foot care score						
Intervention	6.62 (1.87)	6.15 (1.74)	.18 (–0.48)	.004[a] (0.12)	.49 (0.007)	.29 (0.017)
Control	6.63 (1.75)	5.63 (1.87)	.007[a] (–0.24)	—[b]	—	—
Foot care and diabetes self-care behaviors						
Foot-care score						
Intervention	3.51 (1.21)	4.63 (2.2)	<.001[a] (0.63)	<.001[a] (0.31)	.13 (0.034)	.97 (<0.001)
Control	3.03 (0.98)	4.14 (1.55)	<.001[a] (0.61)	—	—	—
Blood-glucose testing score						
Intervention	4.66 (2.71)	5.07 (2.91)	.41 (–0.07)	.09 (0.043)	.91 (0.0002)	.49 (0.01)
Control	4.3 (2.75)	5.29 (2.46)	.13 (–0.07)	—	—	—
Medication adherence score						
Intervention	5.35 (2.32)	5.61 (2.51)	.66 (–0.26)	.98 (<0.001)	.68 (0.003)	.79 (0.001)
Control	5.46 (2.54)	5.34 (2.26)	.71 (–0.15)	—	—	—

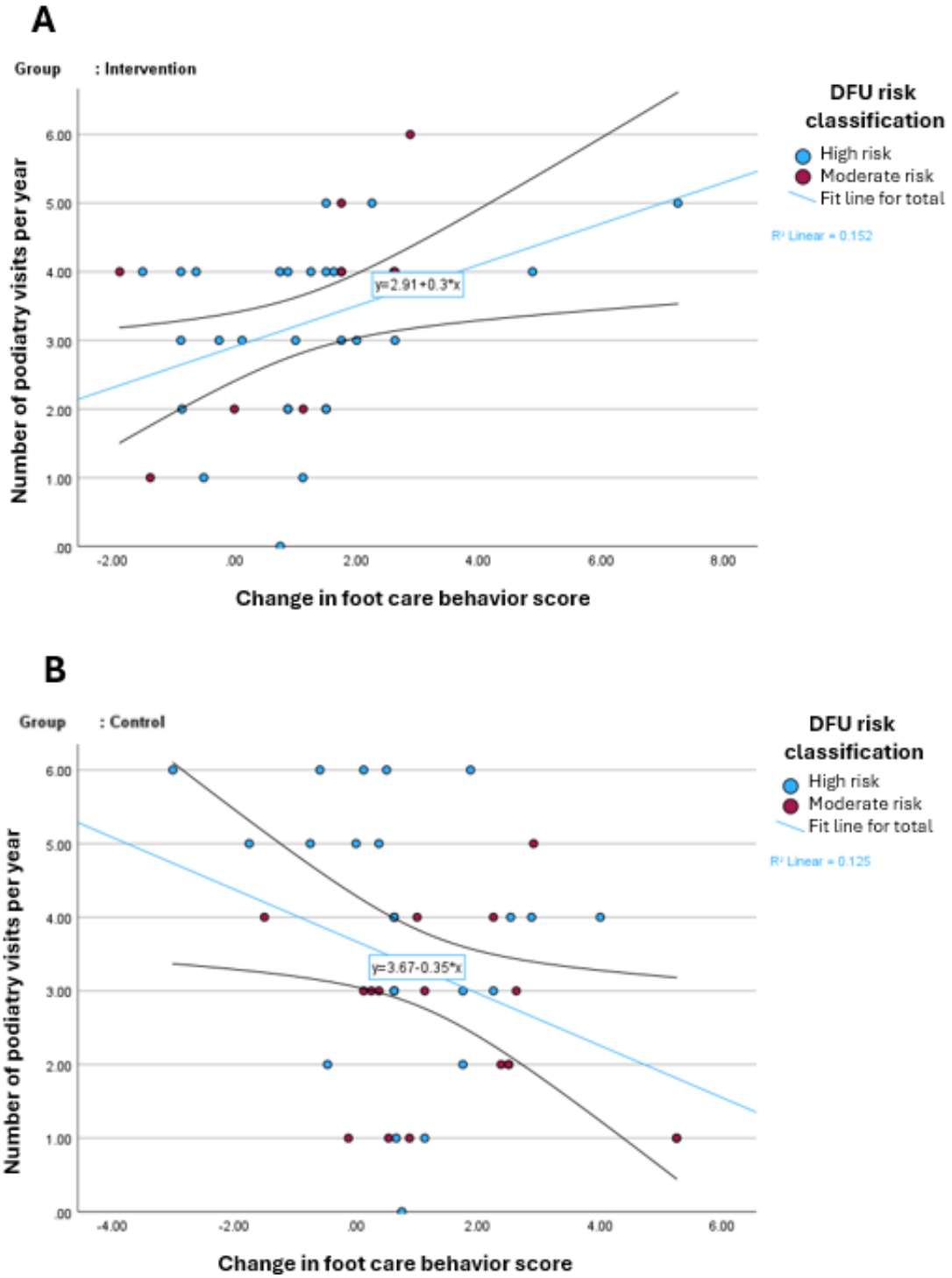
[a]*P*<.05.

[b]Em dashes indicate values that are identical for both groups because these results represent overall ANOVA model effects rather than group-specific estimates.

The mean of the number of podiatry follow-up visits over the 12 months was comparable between the ITT group (control: mean 3.27, SD 1.7 vs intervention: mean 3.24, SD 1.3 visits; *P*=.93) and PP group (control: mean 3.3, SD 1.6 vs intervention: mean 3.26, SD 1.3 visits; *P*=.83). Similarly, among participants at a high DFU risk, those in the intervention attended mean 3.52 (SD 1.2; ITT) and mean 3.64 (SD 1.1; PP) visits, while those in the control group attended mean 3.72 (SD 1.6; ITT) and mean 3.77 (SD 1.6; PP) visits. Among participants at a moderate DFU risk, the intervention group attended a mean of 2.25 (SD 1.48; ITT) visits, in both ITT and PP analyses, while the control group attended a mean of 2.53 (SD 1.6; ITT) visits and a mean of 2.76 (SD 1.58; PP) visits. The number of podiatry visits, however, was positively associated with the change in foot care behavior in the intervention group (ITT: *r*=0.390, *P*=.02; PP: *r*=0.408, *P*=.02), while the control group showed a negative correlation (ITT: *r*=−0.354, *P*=.03; PP: *r*=−0.402, *P*=.02; Figure 4). In ITT analysis, a significant negative association was observed in the control group between the number of podiatry visits and the change in the knowledge scores (*r*=−0.357; *P*=.02), but no such association was found in the intervention group. In the PP analysis, no significant associations were observed in either group. The frequency of new DFU during the course of the study was comparable between the 2 groups (ITT: control=17.5% vs intervention=21.6%, *P*=.64; PP: control=18.8% vs intervention=17.6%, *P*=.91). Both ITT and PP analyses revealed no statistically significant difference in the HbA_{1c} at 12 months between the intervention (ITT: mean 7.86%, SD 0.38%; PP: mean 7.93%, SD 1.08%) and the control (ITT: mean 8.28%, SD 0.36%; PP: mean 8.45%, SD 1.4%) groups, with a mean difference of 0.58% (95% CI −0.93 to 2.09; *P*=.44) in ITT, and 0.52% (95% CI −0.14 to 1.19; *P*=.12) in the PP analyses.

XSL•FO
RenderX

Figure 4. Association between the number of podiatry visits and the change in diabetic foot care behavior in the intervention group (**A**) and control group (**B**). DFU: diabetic foot ulceration.



Discussion

Principal Findings

This pilot study investigated the effectiveness of a multimedia educational approach in improving foot care knowledge and behavior among individuals with diabetes over a 12-month period. Overall, this study found that there was no long-term benefit to using multimedia educational tools in addition to standard usual podiatry care to improve participants' knowledge and behavior. However, post hoc comparisons showed that while both usual care and multimedia educational approaches enhanced the foot care behaviors of participants, the usual care settings resulted in a decline in their foot care knowledge, whereas the multimedia educational settings maintained it. The effect on glycemic control, however, was marginal and not clinically relevant (nonsignificant reduction in HbA_{1c} of only 0.52%). The consistency of results between the ITT and PP analyses supports the robustness of our findings and indicates that participant dropout did not substantially bias the outcomes.

XSL•FO
RenderX

This consistency increases confidence that the observed effects are likely to be representative of real-world settings.

Delivering the appropriate foot self-care education is a crucial strategy for reducing the risk of DFUs in people with diabetes [36]. Though there is currently no consensus regarding the optimal education method on diabetic foot care among health care professionals [37]. In light of technological advancements, several interventional studies using smartphone apps and media aids were conducted to educate patients about foot self-care [38]. But their effectiveness remains unclear. To the best of our knowledge, this is the first study that examined the long-term (12 mo) effects of multimedia educational tools on diabetic foot care including both audio-visual presentation and smartphone app. In support of our findings, a short-term (3 mo) RCT on 120 individuals with diabetes found that an audio-visual educational presentation, supplementary to standard diabetic care and regular reinforcement of education in outpatient clinic, significantly enhanced knowledge scores by +1.17 (95% CI 0.7 - 1.64, *P* value<.001) compared to those receiving standard care alone (+0.1, 95% CI −0.3 to 0.5, *P*=.62), whereas the practice scores were significantly improved in both groups (intervention: +1.6, 95% CI 1.09 - 2.11, *P*<.001; control: +0.48, 95% CI 0.16 - 0.8, *P*=.004) [39]. A 3-week educational program involving presentation slides, booklets, and group discussions significantly improved knowledge (intervention: mean 8.08, SD 0.88 vs control: mean 7.17, SD 1.91; *P*=.02) and behavior scores (intervention: mean 19.32, SD 6.48 vs control: mean 18.96, SD 7.36; *P*=.0001) compared to usual care [40].

Contrary to our findings regarding the use of a smartphone app, a recent RCT study involving 88 participants found that m-DAKBAS, an educational mobile diabetic foot care app that involved information, prevention, and management interfaces, significantly improved knowledge scores over a 6-month period (intervention: mean 16.83, SD 1.56 vs control: mean 15.05, SD 2.17; *P*=.0001), with no significant differences in behavior scores (intervention: mean 62.59, SD 7.76 vs control: mean 59.45, SD 10.53*; P*=.23) [41]. Furthermore, a 2-month use of the MobileDiabetes self-care app, which allowed patients to improve their self-care practices with flexibility in timing, location, and choices, has been shown to increase both patients' self-care knowledge and behavior by 17% and 22%, respectively [42]. In a 1-month RCT, an animation-supported app (Mobile Diabetic Foot Care Education; M-DFCE) that delivered 2 push alerts per week and included a cartoon animation providing basic education on daily foot care to prevent foot wounds was evaluated [43]. The experimental group outperformed the control group in terms of foot care behavior (mean difference: 11.28, SD 10.47 for intervention vs 0.6, SD 24.85 for control; *P* =.01) and knowledge (mean difference: 0.87, SD 1.21 for intervention vs 0.01, SD 1.25 for control; *P*=.002) when compared to baseline [43]. According to a recent systematic review, for patients with diabetes, using a mobile health app enhanced their awareness of the disease and their capacity for self-care in studies of durations of less than 6 months [44]. A study involving 58 patients with uncontrolled diabetes found that integrating self-management, through peer-supporting video, a quiz game, and a feedback system, with the Diabetes Care App for 5 weeks, improved foot care behavior significantly (*P*<.01) when compared to usual care [45]. Similar improvements in foot care practices were observed on patients with type 2 diabetes after 1 month [46] and 3 months [47] of using self-management integration with smartphone apps. Furthermore, the use of smartphone apps and alerts interventions has been associated with improved glycemic control [48], which is contrary to our findings. We found a minor and comparable decrease in HbA_{1c} among the 2 groups, which might be reflected by their scores of the blood glucose-testing and medication adherence. While multimedia education on foot self-care is essential for improving the knowledge and behavior of individuals at a moderate-to-high risk of DFUs, it did not have a significant impact on glycemic control or long-term behavior changes.

In this study, there was no discernible difference in the overall incidence of DFU across the groups, although the overall incidence and thus the study power were low. This may suggest that the use of multimedia aids in foot self-care does not offer additional benefits above the usual care education. In agreement with our findings, a systematic review of 6 RCTs found insufficient evidence for the benefit of an integrated care approach, which involved combining multiple DFU prevention strategies at different levels of care (including the patient, health care provider, and health care structure), indicating the need for more high-quality studies [22]. Another systematic review of 11 RCTs, on the other hand, showed that educational technologies were protective against the incidence of lower limb amputations (relative risk=0.53, 95% CI 0.31 - 0.90; *P*=.02) and DFU (relative risk=0.40, 95% CI 0.18 - 0.90; *P*=.03), despite the lack of evidence of certainty assessment [49]. Other studies suggested that a more focused and intensive educational approach should be adopted to reduce the incidence of DFU. An RCT on people with type 2 diabetes found that 2-hour focused education sessions, including practical exercises on foot care behaviors, were effective in preventing the incidence of DFU during 6-month follow-ups compared to the control group (incidence of DFU=0% vs 10%; *P*=.01) [50]. Similarly, a quasi-experimental study found that combining educational sessions with foot assessment and care reduced the recurrence rate of DFU to 13.3% compared to 33.3% in the control group [51]. Intensive education approaches that included training and customized footwear have been shown to reduce the incidence rate of new DFU compared to usual care by 18% and 31%, respectively [52]. The results from a systematic review and meta-analysis of RCTs demonstrated that, in comparison to the control group, an intensive educational approach—a 45- to 1-hour education session reinforced by written instructions—significantly decreased the risk of DFU incidence (relative risk 0.37, 95% CI 0.14-1.01; *P*=.05) [53]. Nonetheless, the study showed that there was a substantial degree of variability in the studies, with 91% heterogeneity [53]. Thus, multimedia aids may not offer additional benefits beyond usual care in reducing the incidence of DFU among individuals at a moderate-high risk of DFUs, suggesting the need for more targeted educational approaches. Although our study did not find a significant effect on self-care behavior, this should not be interpreted as evidence against the long-term effectiveness of all media-based tools, which may require extended follow-ups, greater interactivity, or more tailored content to

XSL•FO
RenderX

achieve meaningful and sustained behavioral change. In addition, future studies should consider evaluating the effectiveness of media-based educational tools in lower-risk populations, where such interventions may have a greater impact on preventive behaviors and outcomes.

Strengths and Limitations

The strength of this pilot study is the long-term duration of the intervention at 12 months. However, a major limitation of this study is the unexpected COVID-19 global pandemic and associated government restrictions, which significantly impacted study conduct and led to lower enrollment and higher dropout rates than expected. This was potentially because, as a high-risk group for COVID-19, participants were reluctant to increase their contact with others unnecessarily. Although there was a higher dropout among participants with a high risk, particularly in the control group, our ITT analysis yielded results consistent with the PP analysis, suggesting that differential attrition did not substantially bias the study findings. The small sample size limits the generalizability of our findings, and thus, future studies with larger and more diverse populations are needed to validate these results and better evaluate the effectiveness of app-based interventions. Another limitation is that the MyU app lacked interactive features, which may have reduced its appeal—particularly among older participants [54,55]—although this was not formally assessed through usability testing or structured feedback during the study.

Conclusions and Future Directions

In conclusion, the study findings revealed that a multimedia education approach alone to improve foot self-care knowledge and behavior was unfeasible and not engaging for long-term use in people at a moderate-high DFU risk. However, the reinforcement of educational material in the follow-up podiatry visit might be effective in achieving persistent changes in foot care behavior of this patient group. Given the scope of this study, further confirmative RCT studies, with a larger sample size, are needed.

Acknowledgments

The authors would like to thank the participants for volunteering to take part in the study. The authors would also like to acknowledge the support of the nurses, clinical laboratory, and podiatry department at the Dasman Diabetes Institute.

Funding

The study has been funded by the Dasman Diabetes Institute, Kuwait.

Data Availability

The data supporting the results of this study are available from the corresponding author upon reasonable request, with a data transfer agreement if required.

Authors' Contributions

Conceptualization: EAO, GM.
Data curation: AC, GM, KSK, MA.
Data validation: AA, SRG.
Formal analysis: AA.
Funding acquisition: EAO, GM.
Project administration: AC, EAO, KSK, MA, SRG.
Resources: AC, GM, KSK, MA.
Study design: EAO, GM, MA.
Supervision: AC, EAO, KSK, MA, SRG.
Visualization: AA, SRG.
Writing – original draft: AA.
Writing – review & editing: AA, AC, EAO, GM, KSK, MA, SRG.

Conflicts of Interest

None declared.

Checklist 1
CONSORT-EHEALTH checklist.
[PDF File, 2620 KB - humanfactors_v13i1e78261_app1.pdf]

References

1. Ong KL, Stafford LK, McLaughlin SA. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2023 Jul 15;402(10397):203-234. [doi: 10.1016/S0140-6736(23)01301-6] [Medline: 37356446]

XSL•FO
RenderX

2. Magliano DJ, Boyko EJ. IDF DIABETES ATLAS, 10th edition: International Diabetes Federation; 2021. URL: http://www.ncbi.nlm.nih.gov/books/NBK581934 [accessed 2025-12-13]
3. State of the nation (England 2016): time to take control of diabetes. Diabetes UK. 2016. URL: https://www.diabetes.org.uk/node/1753 [accessed 2025-12-13]
4. Reiber GE, Vileikyte L, Boyko EJ, et al. Causal pathways for incident lower-extremity ulcers in patients with diabetes from two settings. Diabetes Care 1999 Jan;22(1):157-162. [doi: 10.2337/diacare.22.1.157] [Medline: 10333919]
5. Abbott CA, Garrow AP, Carrington AL, et al. Foot ulcer risk is lower in South-Asian and African-Caribbean compared with European diabetic patients in the U.K.: the North-West diabetes foot care study. Diabetes Care 2005 Aug;28(8):1869-1875. [doi: 10.2337/diacare.28.8.1869] [Medline: 16043725]
6. Singh N, Armstrong DG, Lipsky BA. Preventing foot ulcers in patients with diabetes. JAMA 2005 Jan 12;293(2):217-228. [doi: 10.1001/jama.293.2.217] [Medline: 15644549]
7. Boulton AJM. The diabetic foot: grand overview, epidemiology and pathogenesis. Diabetes Metab Res Rev 2008;24 Suppl 1:S3-S6. [doi: 10.1002/dmrr.833] [Medline: 18442166]
8. Pecoraro RE, Reiber GE, Burgess EM. Pathways to diabetic limb amputation. Basis for prevention. Diabetes Care 1990 May;13(5):513-521. [doi: 10.2337/diacare.13.5.513] [Medline: 2351029]
9. Siitonen OI, Niskanen LK, Laakso M, Siitonen JT, Pyörälä K. Lower-extremity amputations in diabetic and nondiabetic patients. A population-based study in eastern Finland. Diabetes Care 1993 Jan;16(1):16-20. [doi: 10.2337/diacare.16.1.16] [Medline: 8422771]
10. Trautner C, Haastert B, Giani G, Berger M. Incidence of lower limb amputations and diabetes. Diabetes Care 1996 Sep;19(9):1006-1009. [doi: 10.2337/diacare.19.9.1006] [Medline: 8875099]
11. Armstrong DG, Swerdlow MA, Armstrong AA, Conte MS, Padula WV, Bus SA. Five year mortality and direct costs of care for people with diabetic foot complications are comparable to cancer. J Foot Ankle Res 2020 Mar 24;13(1):16. [doi: 10.1186/s13047-020-00383-2] [Medline: 32209136]
12. Lim TS, Finlayson A, Thorpe JM, et al. Outcomes of a contemporary amputation series. ANZ J Surg 2006 May;76(5):300-305. [doi: 10.1111/j.1445-2197.2006.03715.x] [Medline: 16768686]
13. Schaper NC, van Netten JJ, Apelqvist J, et al. Practical guidelines on the prevention and management of diabetic foot disease (IWGDF 2019 update). Diabetes Metab Res Rev 2020 Mar;36 Suppl 1(S1):e3266. [doi: 10.1002/dmrr.3266] [Medline: 32176447]
14. Ezzatvar Y, García-Hermoso A. Global estimates of diabetes-related amputations incidence in 2010-2020: a systematic review and meta-analysis. Diabetes Res Clin Pract 2023 Jan;195:110194. [doi: 10.1016/j.diabres.2022.110194] [Medline: 36464091]
15. Raju BN, Mateti UV, Mohan R, et al. Educational interventions and its impact on the treatment outcomes of diabetic foot ulcer patients. J Diabetol 2022;13(1):1-7. [doi: 10.4103/jod.jod_99_21]
16. Yıldırım Ayaz E, Dincer B, Oğuz A. The effect of foot care education for patients with diabetes on knowledge, self-efficacy and behavior: systematic review and meta-analysis. Int J Low Extrem Wounds 2022 Sep;21(3):234-253. [doi: 10.1177/15347346221109047] [Medline: 35711163]
17. Saltar L, Sahar J. The intervention of foot care education in the prevention of diabetic foot ulcers: a literature review. 2020 Presented at: International Summit on Science, Technology, and Humanity (ISETH); Dec 15-16, 2020 URL: https://proceedings.ums.ac.id/iseth/article/view/1276 [accessed 2021-08-25]
18. Sampson EO, Abdul Manaf R, Ismail S, Shahar HK, Udeani TK. Improving foot self-care practices through health education intervention programs among diabetic patients: a systematic review. Malays J Med Health Sci 2023;19(4):315-325. [doi: 10.47836/mjmhs.19.4.44]
19. Dorresteijn JAN, Kriegsman DMW, Assendelft WJJ, Valk GD. Patient education for preventing diabetic foot ulceration. Cochrane Database Syst Rev 2014 Dec 16;2014(12):CD001488. [doi: 10.1002/14651858.CD001488.pub5] [Medline: 25514250]
20. Campbell FA, Goldman BD, Boccia ML, Skinner M. The effect of format modifications and reading comprehension on recall of informed consent information by low-income parents: a comparison of print, video, and computer-based presentations. Patient Educ Couns 2004 May;53(2):205-216. [doi: 10.1016/S0738-3991(03)00162-9] [Medline: 15140461]
21. Murphy PW, Chesson AL, Walker L, Arnold CL, Chesson LM. Comparing the effectiveness of video and written material for improving knowledge among sleep disorders clinic patients with limited literacy skills. South Med J 2000 Mar;93(3):297-304. [Medline: 10728518]
22. Hoogeveen RC, Dorresteijn JAN, Kriegsman DMW, Valk GD. Complex interventions for preventing diabetic foot ulceration. Cochrane Database Syst Rev 2015 Aug 24;2015(8):CD007610. [doi: 10.1002/14651858.CD007610.pub3] [Medline: 26299991]
23. Pratiwi LD, Haryanto J, Wahyudi AS. The effect of foot self care and diabetes self-management mobile application in preventing foot ulcer recurrence: a systematic review study. Malays J Med Health Sci 2023;19:109-116 [FREE Full text]
24. Obilor HN, Achore M, Woo K. Use of information communication technology tools in diabetic foot ulcer prevention programs: a scoping review. Can J Diabetes 2022 Jul;46(5):535-548. [doi: 10.1016/j.jcjd.2021.11.009] [Medline: 35739048]

XSL•FO
RenderX

25. Bus SA, van Netten JJ. A shift in priority in diabetic foot care and research: 75% of foot ulcers are preventable. Diabetes Metab Res Rev 2016 Jan;32 Suppl 1(S1):195-200. [doi: 10.1002/dmrr.2738] [Medline: 26452160]
26. Process of translation and adaptation of instruments. World Health Organization. 2016. URL: https://www.scirp.org/reference/referencespapers?referenceid=3514337 [accessed 2025-07-06]
27. Diabetic foot problems: prevention and management. National Institute for Health and Care Excellence. 2015 Aug 26. URL: https://www.nice.org.uk/guidance/ng19 [accessed 2025-07-03]
28. Management of diabetes: a national clinical guideline. Scottish Intercollegiate Guidelines Network. URL: https://www.sign.ac.uk/assets/sign116.pdf [accessed 2025-12-09]
29. American Diabetes Association. Microvascular complications and foot care: standards of medical care in diabetes—2018. Diabetes Care 2018 Jan;41(Suppl 1):S105-S118. [doi: 10.2337/dc18-S010] [Medline: 29222381]
30. Sheridan SL, Viera AJ, Krantz MJ, et al. The effect of giving global coronary risk information to adults: a systematic review. Arch Intern Med 2010 Feb 8;170(3):230-239. [doi: 10.1001/archinternmed.2009.516] [Medline: 20142567]
31. Pollock RD, Unwin NC, Connolly V. Knowledge and practice of foot care in people with diabetes. Diabetes Res Clin Pract 2004 May;64(2):117-122. [doi: 10.1016/j.diabres.2003.10.014] [Medline: 15063604]
32. Rheeder P, Venn M, de Korte E, van Zyl D. Knowledge of foot care in people with diabetes in a tertiary care setting. J Endocrinol Metab Diabetes S Afr 2008 Dec;13(3):105-108. [doi: 10.1080/22201009.2008.10872180]
33. Toobert DJ, Hampson SE, Glasgow RE. The summary of diabetes self-care activities measure: results from 7 studies and a revised scale. Diabetes Care 2000 Jul;23(7):943-950. [doi: 10.2337/diacare.23.7.943] [Medline: 10895844]
34. AlJohani KA, Kendall GE, Snider PD. Psychometric evaluation of the Summary of Diabetes Self-Care Activities-Arabic (SDSCA-Arabic): translation and analysis process. J Transcult Nurs 2016 Jan;27(1):65-72. [doi: 10.1177/1043659614526255] [Medline: 24841470]
35. Baba M, Duff J, Foley L, Davis WA, Davis TME. A comparison of two methods of foot health education: the Fremantle Diabetes Study Phase II. Prim Care Diabetes 2015 Apr;9(2):155-162. [doi: 10.1016/j.pcd.2014.05.004] [Medline: 24929632]
36. Schaper NC, van Netten JJ, Apelqvist J, et al. Practical guidelines on the prevention and management of diabetic foot disease (IWGDF 2019 update). Diabetes Metab Res Rev 2020 Mar;36(S1):e3266. [doi: 10.1002/dmrr.3266] [Medline: 32176447]
37. Jeffcoate WJ, Vileikyte L, Boyko EJ, Armstrong DG, Boulton AJM. Current challenges and opportunities in the prevention and management of diabetic foot ulcers. Diabetes Care 2018 Apr;41(4):645-652. [doi: 10.2337/dc17-1836] [Medline: 29559450]
38. Kustini K, Sari Y. Use of mobile applications in increasing knowledge of diabetes mellitus foot care. J Med Health Stud 2023;4(6):21-24. [doi: 10.32996/jmhs.2023.4.6.3]
39. Rahaman HS, Jyotsna VP, Sreenivas V, Krishnan A, Tandon N. Effectiveness of a patient education module on diabetic foot care in outpatient setting: an open-label randomized controlled study. Indian J Endocrinol Metab 2018;22(1):74-78. [doi: 10.4103/ijem.IJEM_148_17] [Medline: 29535941]
40. Beiranvand S, Fayazi S, Asadizaker M. Effect of educational programs on the knowledge, attitude, and practice of foot care in patients with diabetes. Jundishapur J Chronic Dis Care 2015;4(2). [doi: 10.5812/jjcdc.26540]
41. Kilic M, Karadağ A. Developing and evaluating a mobile foot care application for persons with diabetes mellitus: a randomized pilot study. Wound Manag Prev 2020 Oct;66(10):29-40. [doi: 10.25270/wmp.2020.10.2940] [Medline: 33048829]
42. Guo SHM, Chang HK, Lin CY. Impact of mobile diabetes self-care system on patients' knowledge, behavior and efficacy. Comput Ind 2015 May;69:22-29. [doi: 10.1016/j.compind.2014.11.001]
43. Dincer B, Bahçecik N. The effect of a mobile application on the foot care of individuals with type 2 diabetes: a randomised controlled study. Health Educ J 2021 Jun;80(4):425-437. [doi: 10.1177/0017896920981617]
44. Sadler S, Gerrard J, Searle A, et al. The use of mHealth apps for the assessment and management of diabetes-related foot health outcomes: systematic review. J Med Internet Res 2023 Oct 4;25:e47608. [doi: 10.2196/47608] [Medline: 37792467]
45. Firdaus M, Jittanoon P, Boonyasopun U, Che Hasan MK. The effect of mHealth program on behavior modification and health outcomes among patients with diabetes: a randomized controlled trial study. Belitung Nurs J 2023;9(5):437-447. [doi: 10.33546/bnj.2664] [Medline: 37901368]
46. Chen SM, Jo Wu CJ. Investigating the effects of digital foot self-management program on enhancing self-efficacy and self-care behavior among community-dwelling older adults with type 2 diabetes: a randomized controlled trial. Digit Health 2023;9:20552076231220791. [doi: 10.1177/20552076231220791] [Medline: 38107978]
47. Suniyadewi NW, Nursalam N, Sufyanti Arief Y, et al. Effect of android-based mobile diabetic foot early self-assessment on diabetic foot prevention behaviors of Indonesian patients with type 2 diabetes. J Holist Nurs Midwifery 2024;34(4):309-315. [doi: 10.32598/jhnm.34.4.2633]
48. Moschonis G, Siopis G, Jung J, et al. Effectiveness, reach, uptake, and feasibility of digital health interventions for adults with type 2 diabetes: a systematic review and meta-analysis of randomised controlled trials. Lancet Digit Health 2023 Mar;5(3):e125-e143. [doi: 10.1016/S2589-7500(22)00233-3] [Medline: 36828606]

XSL•FO
RenderX

49. Lira JAC, Rocha Á, Bezerra SMG, Nogueira PC, Santos A, Nogueira LT. Effects of educational technologies on the prevention and treatment of diabetic ulcers: a systematic review and meta-analysis. Rev Lat Am Enfermagem 2023;31:e3944. [doi: 10.1590/1518-8345.6628.3944] [Medline: 37341258]
50. Monami M, Zannoni S, Gaias M, Nreu B, Marchionni N, Mannucci E. Effects of a short educational program for the prevention of foot ulcers in high-risk patients: a randomized controlled trial. Int J Endocrinol 2015;2015(1):615680. [doi: 10.1155/2015/615680] [Medline: 26448748]
51. Pratama K, Fahrain J, et al. Prevention strategy for ulcer recurrence in patients with type II diabetes mellitus: a quasi-experimental study. Iran J Nurs Midwifery Res 2023;28(2):139-143. [doi: 10.4103/ijnmr.ijnmr_432_21] [Medline: 37332375]
52. Bharat Kotru SK. Intervention of diabetes foot care practices on the prevention of new diabetic foot ulcers in patients with type 2 diabetes mellitus. J Diabetes Metab 2025;06(2):1-6. [doi: 10.4172/2155-6156.1000494]
53. Adiewere P, Gillis RB, Imran Jiwani S, Meal A, Shaw I, Adams GG. A systematic review and meta-analysis of patient education in preventing and reducing the incidence or recurrence of adult diabetes foot ulcers (DFU). Heliyon 2018 May;4(5):e00614. [doi: 10.1016/j.heliyon.2018.e00614] [Medline: 29872752]
54. Steinert A, Eicher C, Haesner M, Steinhagen-Thiessen E. Effects of a long-term smartphone-based self-monitoring intervention in patients with lipid metabolism disorders. Assist Technol 2020;32(2):109-116. [doi: 10.1080/10400435.2018.1493710] [Medline: 29944463]
55. Su J, Dugas M, Guo X, Gao GG. Influence of personality on mHealth use in patients with diabetes: prospective pilot study. JMIR mHealth mHealth 2020 Aug 10;8(8):e17709. [doi: 10.2196/17709] [Medline: 32773382]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
DFU: diabetic foot ulceration
HbA_{1c}: hemoglobin A_{1c}
ITT: intention-to-treat
NICE: National Institute for Health and Care Excellence
PP: per-protocol
RCT: randomized controlled trial
SDSCA: Summary of Diabetes Self-Care Activities

Edited by A Kushniruk; submitted 29.May.2025; peer-reviewed by C Law, DV Coppini, Y Gao; revised version received 15.Sep.2025; accepted 14.Oct.2025; published 13.Jan.2026.

Please cite as:
Alawadhi A, Khan KS, Messenger G, Cusworth A, Assi M, Gray SR, Al-Ozairi E
Long-Term Effects of Multimedia Education and "Foot Alerts" Through the MyU App on Diabetic Foot Care in Individuals With Moderate-High Risk: Pilot Randomized Controlled Trial
JMIR Hum Factors 2026;13:e78261
URL: https://humanfactors.jmir.org/2026/1/e78261
doi:10.2196/78261

© Ameenah Alawadhi, Kay Scarsbrook Khan, Grace Messenger, Alice Cusworth, Mohammad Assi, Stuart R Gray, Ebaa Al-Ozairi. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 13.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Co-Created Digital Pretherapy Psychoeducation for Outpatients in Specialized Mental Health Care: Usability Evaluation and Patient Satisfaction Study

Henrik Pedersen[1,2]; Tatiana Skliarova[1]; Liv S Engvik[1]; Arthur Mandahl[3]; Carlos De las Cuevas[4]; Audun Havnen[1,5]; Mariela Loreto Lara-Cabrera[2,6]

[1]Nidaros DPS, St Olav's University Hospital, Trondheim, Norway

[2]Department of Mental Health, Norwegian University of Science and Technology, Trondhem, Norway

[3]Vårres Regional User-Led Center, Trondheim, Norway

[4]Department of Internal Medicine, Dermatology and Psychiatry, Universidad de La Laguna, La Laguna, Spain

[5]Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway

[6]Nidelv DPS, St Olav's University Hospital, Trondheim, Norway

Corresponding Author:
Henrik Pedersen
Nidaros DPS
St Olav's University Hospital
Østmarkveien 49
Trondheim, 7050
Norway
Phone: 47 91697343
Email: henrik.pedersen@ntnu.no

Abstract

Background: Specialized mental health services are facing high demand, potentially leading to lower-quality care. One solution may be to prepare patients for attending treatment. Digital pretherapy psychoeducation may be particularly relevant. However, the effectiveness of such an intervention depends on user engagement and satisfaction, and usability is therefore one of the most important factors.

Objective: This article has 2 objectives. Study 1 describes the development of StartHelp, a co-created digital pretherapy psychoeducation program for patients on waiting lists before their first consultation in outpatient specialized mental health services. Study 2 explores the usability of StartHelp, aiming to identify potential issues and assess whether the intervention is suitable for further evaluation in a randomized controlled trial.

Methods: Guided by co-creation principles, we developed StartHelp in accordance with the Guidance for Reporting Involvement of Patients and the Public (GRIPP2) checklist. To assess the app's usability, we recruited 10 patients from specialized mental health care to complete tasks during individual think-aloud interviews. Afterward, they completed questionnaires, including open-ended questions, an item assessing perceived video quality, 2 versions of the System Usability Scale (SUS), the 4-item Client Satisfaction Questionnaire (CSQ-4), and a modified version of the CSQ-4 (CSQ-Video). The StartHelp project group discussed the results, and solutions to the identified issues were proposed and implemented.

Results: Study 1 involved the development of StartHelp over 24 months. The app comprises 27 tasks, including 24 videos and links to 14 websites containing in-depth information. Study 2 involved usability testing with 5 men and 5 women. SUS scores for StartHelp's videos indicated good usability, with a mean of 83.7. By contrast, SUS scores related to navigating StartHelp's overarching architecture were barely acceptable, with a mean of 63.6. For the CSQ-4, the sample mean was 12.3, indicating moderate satisfaction. Mean scores on the CSQ-Video (10.9) indicated satisfaction in the lower moderate range. However, patients perceived the videos as high quality and rated them as nonoffensive. The qualitative findings supported the quantitative results. The usability tests revealed 1 major issue and several minor issues. The primary issue concerned navigation of the overarching technological infrastructure on which StartHelp was developed, rather than StartHelp itself. To address these issues and impracticalities related to interacting with the overarching technological infrastructure, we made minor changes to the StartHelp app.

XSL•FO
RenderX

Conclusions: Through a collaborative co-creation process, we developed StartHelp, a digital pretherapy psychoeducation program. Usability testing indicated that the content itself was highly usable, with video-related SUS scores suggesting good usability, whereas navigation of the overarching technological infrastructure received lower scores. Although patients rated the clinical material as high quality and nonoffensive, satisfaction scores were lower than anticipated. Nevertheless, StartHelp was deemed ready for further testing in a clinical trial.

(JMIR Hum Factors 2026;13:e80130) doi:10.2196/80130

KEYWORDS

usability testing; mental health; mental disorder; patient-centered care; internet-delivered; patient satisfaction; usability; digital health; patient and public involvement

Introduction

Background

Globally, the high prevalence of mental disorders represents a growing public health concern [1,2]. This prevalence leads to heightened barriers to care in specialized mental health services [2] and results in a decline in service quality due to the measures implemented to address the high demand [3]. Such measures may include more restricted services or longer waiting times before the initial consultation, which could result in unmet needs [2], lower attendance [3], worsening of mental health symptoms during the waiting period [3], or changes in patient-reported variables that predict positive outcomes in future therapy [4]. To address this issue, a pluralistic approach is likely necessary, with a simultaneous focus on traditional psychotherapy, adjunctive interventions, and innovative technology-assisted approaches [2,3,5,6].

One type of intervention that may help meet this demand is the preparation of patients for specialized mental health services through psychoeducation [7-11]. Such pretherapy interventions have demonstrated promising results by increasing patients' knowledge of treatment options [9]. Furthermore, improvements in patient satisfaction [12] and patient-reported outcomes [7,12-14], reported in both the short term [8] and long term, suggest an association with more favorable treatment outcomes [7,8].

One dilemma faced by pretherapy interventions is the allocation of therapist resources: specifically, whether therapists should dedicate their time to preparing or treating patients. One way to address this issue is to involve user representatives as close collaborators, delivering pretherapy interventions in a group format [7-9], thereby reducing the number of health personnel required. Another solution is to provide pretherapy psychoeducation digitally [10], particularly given its potential for scalability [6,15,16] and lower barriers to entry [10]. However, researchers have highlighted the lack of rigorous testing of pretherapy digital interventions [10], both during their development and in clinical trials [15,16]. Some studies have also identified adverse effects of digital pretherapy [10], indicating that preparation is not without risk and that thorough evaluation is crucial. Moreover, the outcomes of digital interventions are highly dependent on user engagement and satisfaction, with usability being the primary barrier to engagement [17]. In other words, usefulness, ease of use, and satisfaction are core components of usability [18].

Recent developments have underscored the importance of patient and public involvement in research [19,20]. This involvement can begin during the development of health interventions [20-22], a practice commonly referred to as "co-creation" [22-24]. The process of co-creation goes beyond gathering user feedback; it involves active collaboration with all relevant nonacademic stakeholders throughout the research and development phases, grounded in a patient-centered perspective [25]. This approach results in more relevant content, increased engagement, and ultimately more effective implementation [20,22-24,26]. Despite these potential benefits, however, patients and user representatives are rarely engaged in the co-creation of mental health interventions [22], including both digital and nondigital pretherapy interventions [10,26].

StartHelp

To address these gaps in the literature, we developed StartHelp (Norwegian: "StartHjelp"), a co-created digital pretherapy psychoeducation program for patients awaiting their first consultation at outpatient specialized mental health services. StartHelp is designed to address the gap between referral and a patient's initial appointment with a therapist. The primary goal in developing StartHelp was to provide patients with essential information before starting treatment in a digital, automated way. This information includes legally mandated information, such as patient rights, as well as details regarding the psychological assessment process, practical considerations, and coping strategies during the waiting period. Moreover, different treatment approaches are presented, including video consultations, group therapy, therapist-guided internet-based interventions, medical treatment, and physical therapy. The aim of providing this information to patients before their first consultation is to enhance patient-centered care by preparing them for their upcoming therapy. Within this framework, StartHelp aims to improve patient satisfaction by making the waiting period more meaningful, clarifying the patient's role in therapy, establishing realistic expectations, and alleviating apprehension about the initial sessions.

Objectives and Hypotheses

This study aimed to evaluate whether StartHelp was ready for real-world testing in a clinical trial and had 2 objectives. The first was to describe the development process of StartHelp (study 1). The second objective was to investigate its usability by assessing program navigation, potential adverse effects, and suggestions for improving 4 of StartHelp's videos (study 2). We hypothesized that StartHelp would achieve a System

XSL•FO
RenderX

Usability Scale (SUS) score of >70, based on benchmark values for digital health apps [27] and Scandinavian patient portals [28]. Additionally, we hypothesized that StartHelp would achieve medium levels of satisfaction.

Methods

Study 1: Presentation of StartHelp and Its Development

Overview

We used the Guidance for Reporting Involvement of Patients and the Public (GRIPP2) checklist [19] as a reporting framework to ensure transparent reporting of user involvement (see Multimedia Appendix 1).

Ethics Approval

All necessary ethical approvals were obtained before the study began. The study protocol and consent form were evaluated and approved by the Regional Committee for Medical Research Ethics in Mid Norway (*Regional etisk komite* [regional ethics committee]: 2024/729793) and the Norwegian Agency for Shared Services in Education and Research (Norwegian: Kunnskapssektorens tjenesteleverandør [SIKT]: 2024/987092). When initially contacted, patients were told that participation was independent of their treatment process and that their decision whether to participate would not affect their treatment in any way. They were also informed that they could send an SMS text message at any time to cancel the appointment without having to explain why. Patients formally provided informed consent by signing a consent form on-site before starting the testing. The form consisted of information about the study, how the data were handled to secure anonymity, and patients' right to opt out of the study at any time without any consequences for treatment.

Description of StartHelp

StartHelp is a mobile app accessed through the official regional patient portal [29]. In StartHelp, patients are prompted with tasks at predetermined intervals, consisting of psychoeducation delivered through text and video. This stepwise approach serves 2 purposes. First, we believe that delivering information in this manner ensures that patients are not overwhelmed, allowing them to review the information carefully. Second, we reasoned that this approach may promote ongoing engagement and foster a sense of continuity during patients' time on the waiting list.

StartHelp's dashboard consists of a list of tasks that automatically updates. Clicking on a task reveals a brief text consisting of 1 or 2 paragraphs, and most tasks provide hyperlinks to websites containing more in-depth information [30]. The destinations of these links include websites of official Norwegian health authorities, patient organizations, official legal authorities, and "closed-off" websites hosted by the university hospital and created by the project group. These websites are accessible only through the app and not via search engines. On these websites, short videos accompanied by text provide more detailed information. The videos range in length from 1 to 3 minutes, with 1 video exceeding 4 minutes. In total, StartHelp consists of 27 tasks, 14 websites, and 24 videos. Figure 1 presents an overview of the task content and the timing of task administration. Figures 2 and 3 show screenshots of StartHelp and a website with a video and accompanying text. Figure 4 depicts the StartHelp dashboard in its right frame.

The intention was for every patient to read the initial information and watch most of the videos before selecting the in-depth information relevant to them. As patients must actively press a button to mark a task as complete, the app can track their activity, and checkmarks indicate which tasks have been completed.

Figure 1. StartHelp content and timeline.



XSL•FO
RenderX

Figure 2. Visual presentation of tasks in StartHelp.

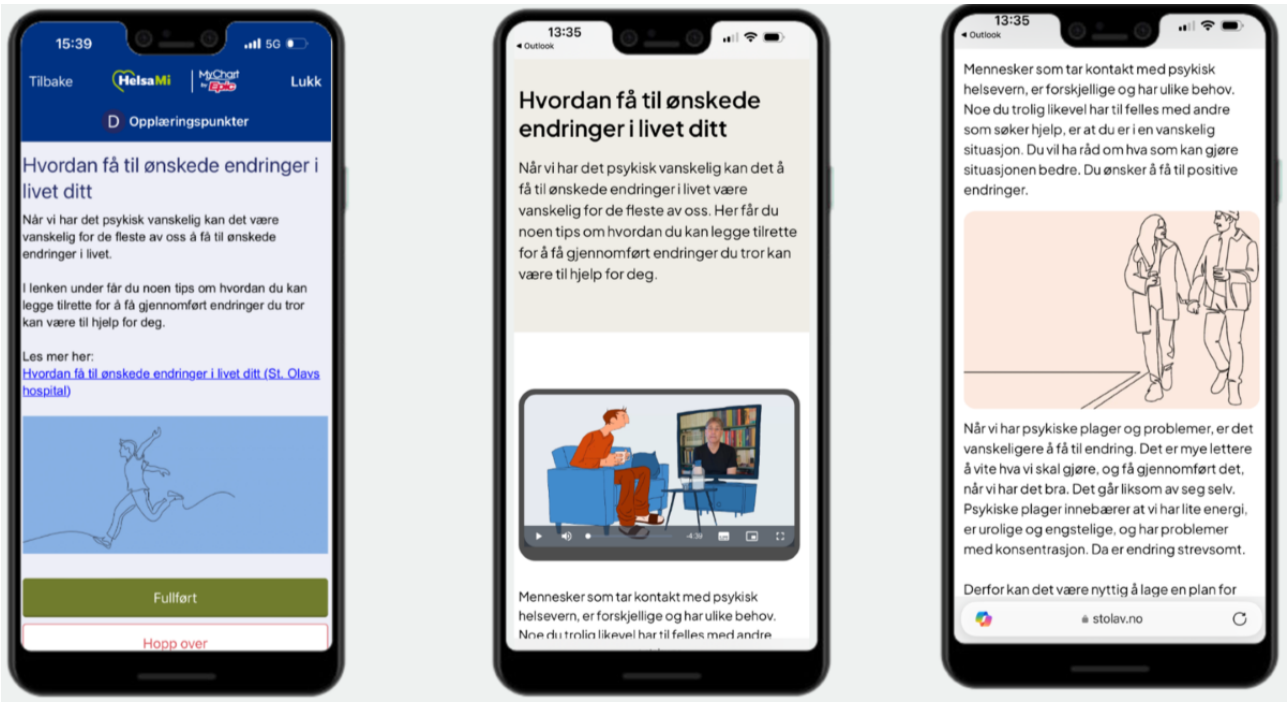
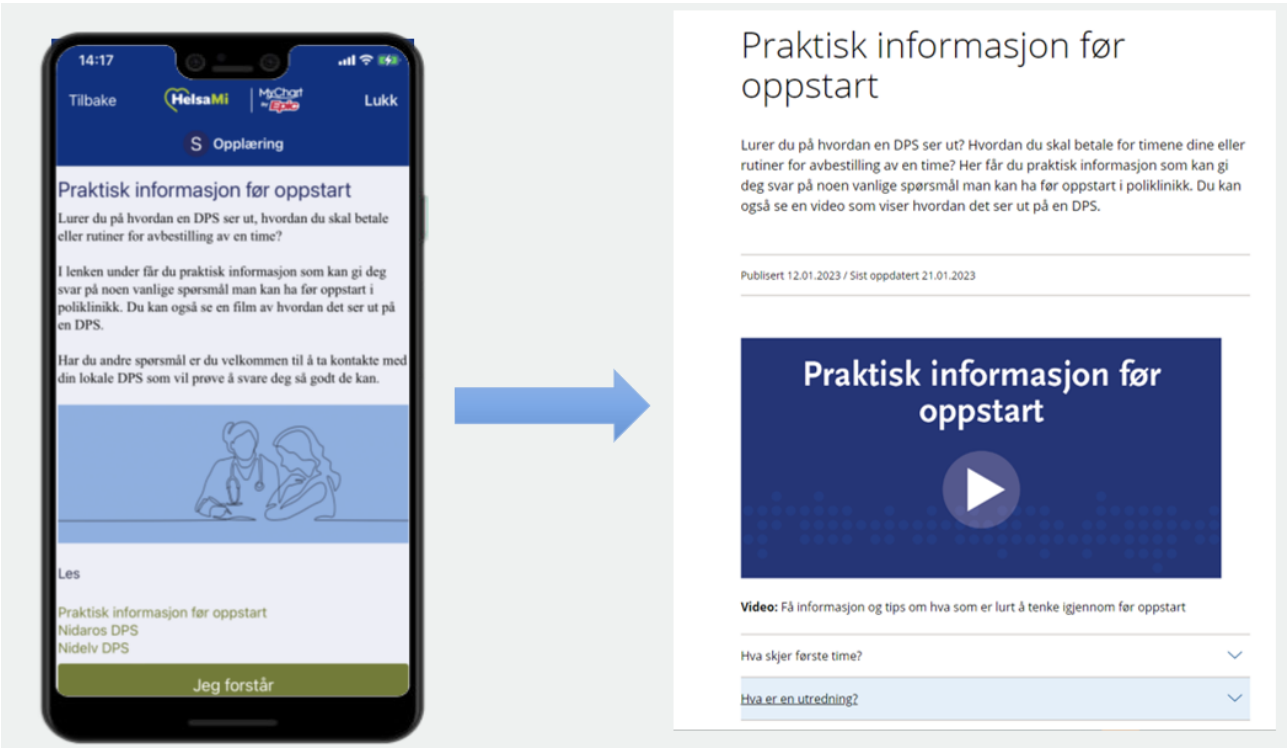
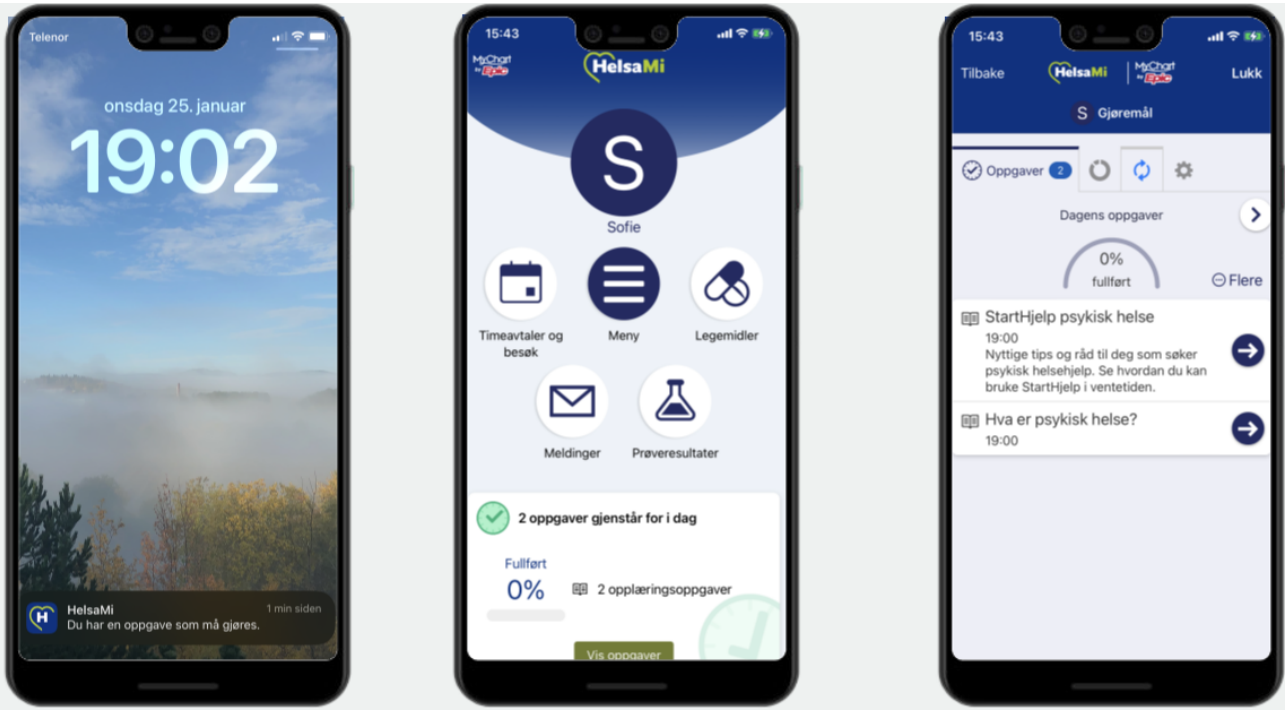


Figure 3. Visual presentation of a task and website with in-depth information.



XSL•FO
RenderX

Figure 4. Visual presentation of phone notifications, the MyHealth front page, and StartHelp dashboard.



Description of the Overarching Technological Infrastructure

StartHelp is not a standalone app; it is built upon an overarching technological infrastructure known as MyHealth (Norwegian: HelsaMi). MyHealth serves as a patient portal connected to the regional patient administration and journal system in the Central Norway health region. Patients need MyHealth to access StartHelp. StartHelp is free to use and does not require user registration.

MyHealth was designed to provide patients and, occasionally, their relatives with easy access to patient information and health care services. It offers a range of functions, including completing questionnaires relevant to the patient's care; providing secure video consultations; notifying the health care provider of the patient's arrival at the clinic; posing questions to the health care provider; receiving information; and providing an overview of planned consultations, personal information, medical prescriptions, and journal entries. Figure 4 presents screenshots of MyHealth and StartHelp, with the MyHealth front page displayed in the middle. MyHealth is comparable to other patient portals in other health regions across Norway and other Nordic countries due to collaborations with the same health system providers [28,31,32].

In addition to serving as a patient portal, one of MyHealth's goals is to provide a secure infrastructure integrated into the regional journal system, enabling the development of additional innovative programs tailored to specific patient groups without involving private third-party actors.

Development of StartHelp

User Involvement and Co-Creation

User involvement was deemed essential from the beginning of the development process, both in principle and to ensure the quality, relevance, and implementation of the final product [20,22]. As the objective was to digitize prior pretherapy educational interventions [7-9], user representatives who had collaborated in their design and delivery were contacted. We strove to ensure that these user representatives held a partner role, meaning they would have equal influence with other members of the project group in making important decisions throughout the process [33]. In addition to user representatives, the project group included experienced nonacademic clinicians, psychologists, psychiatrists, psychiatric nurses, and clinic middle management to ensure that all relevant perspectives were represented [34].

First, the group reviewed previous interventions and current service needs. User representatives highlighted frustrations regarding long wait times and limited communication. Consequently, the team identified a critical need for information regarding treatment options, self-help, patient rights, and user organizations, with a preference for video content over written text.

Development of Video Manuscripts and Informational Websites

The project group decided to develop 2 general types of videos for StartHelp: one featuring a clinician speaking directly to the camera, accompanied by visual animations illustrating key points (experts talking), and animated videos with a voice-over.

Most of the work on the manuscripts for these videos was conducted collaboratively with the project group via videoconference. We collectively outlined the objectives for each video, aiming for concision. Thus, patients who find the content relevant may feel encouraged to explore the in-depth information provided through the links. The intention was that this proactive method of seeking relevant information would also engage patients at an early stage. We then allocated responsibility for the manuscripts among the project group

XSL•FO
RenderX

members and revised them together in subsequent meetings. In this process, the user representatives exerted significant influence on the phrasing.

After completing the video manuscripts, the project group contacted an advertising agency with an existing contractual relationship with the university hospital. The agency filmed all the videos, produced the animations, and recorded the voice-overs, which underwent iterative testing in 2 workshops with 6 participants each. Participants, recruited by user representatives, included individuals with attention-deficit/hyperactivity disorder, autism spectrum disorder, or both. This approach helped ensure that the content was accessible and comprehensible to a broad audience.

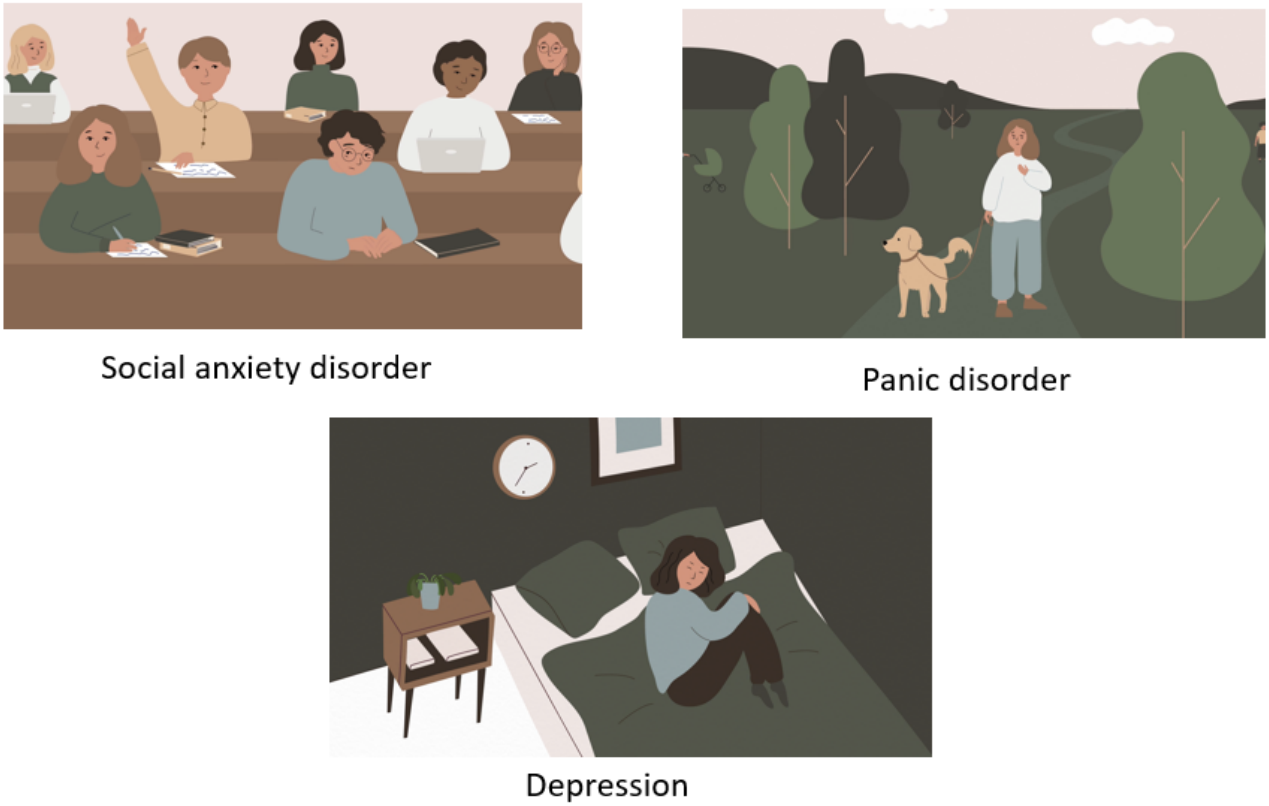
The feedback that emerged from the workshops prompted significant design changes. Participants rejected an initial "cartoon character" style as too similar to commercials and too distracting. We adopted a "1-line drawing" style that gradually appears on screen and decided to maintain this graphic style throughout StartHelp. Based on user requests, the project group also created an additional introductory video.

Website Design

After finalizing the video design, we created websites with comprehensive information. These websites use the hospital's web domain, which imposed template constraints, but we maintained the videos' style throughout. The content was designed for easy scanning, using headlines that function as drop-down menus.

During the development of the websites, the project group realized the need for 3 additional videos addressing panic disorder, social anxiety, and depression. The videos aimed to lower barriers to entry for specific treatment options, such as therapist-led internet treatment and group therapy. The videos used composite narratives derived from interviews with 3 patients per diagnosis to ensure anonymity and relatability. Unlike the earlier content, these videos were not tested in workshops and were therefore included in the usability testing. The graphic profile of these videos is illustrated in Figure 5.

Figure 5. The graphic profile of the videos depicting specific mental issues.



Initial Paper Testing

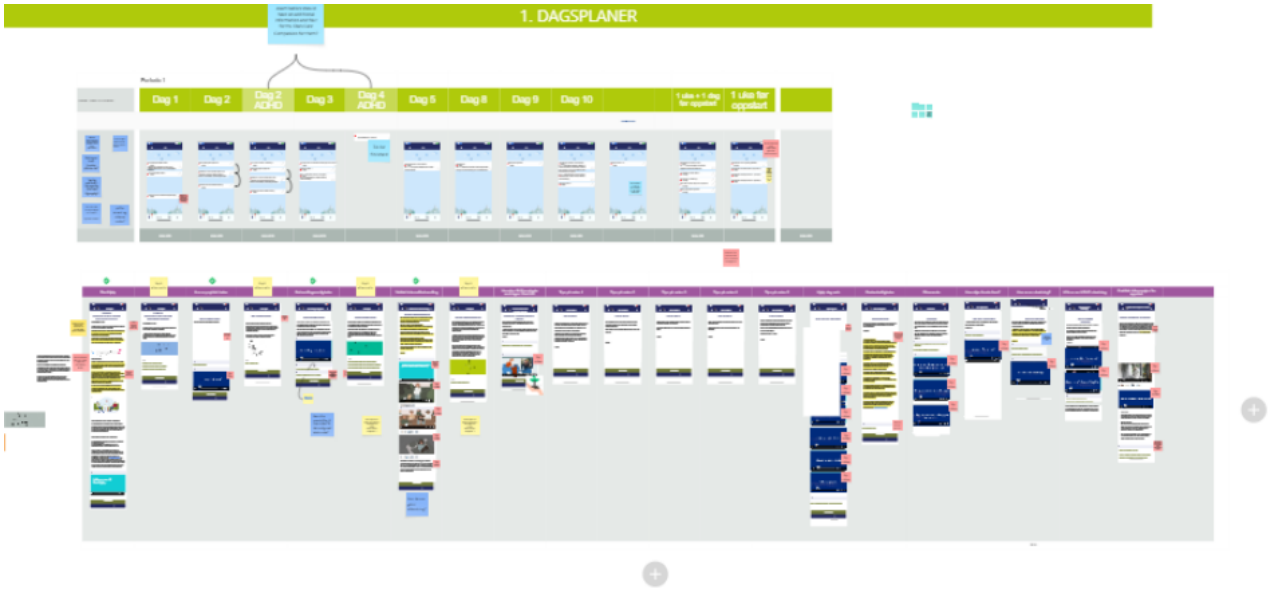
In January 2023, the project group conducted a "paper test" of a StartHelp prototype before integrating it into MyHealth. The group outlined the prototype in Miro Board, a cloud-based interactive whiteboard. The visual drafts were printed on paper and tested on 4 individuals aged 25-64 who were recruited from the associated outpatient mental health clinic.

In this test, patients navigated back and forth within the paper-based prototype app. They described which buttons they would press, and a flowchart guided them to different pages in the paper prototype. Two members of the project group participated in each user test: 1 conducted the test, and the other took notes based on the users' feedback. The test produced 10 key insights. These insights were then categorized according to recurring themes. Figure 6 depicts this process.

Finally, the project group conducted preliminary checks for major technical bugs before planning and conducting the usability testing outlined in study 2.

XSL•FO
RenderX

Figure 6. Visualization of paper prototype testing.



Study 2: Usability Testing

Methods and Procedures

Study 2 explored the usability of StartHelp, aiming to clarify usability issues that may hinder a clinical trial. Usability testing was conducted between May and July 2024 using a convenience sample of patients receiving treatment at a Norwegian outpatient mental health clinic. These patients were contacted because StartHelp's user base will primarily consist of individuals from the general clinical mental health population. Although StartHelp is intended for patients who have not yet started treatment, we recruited patients who had attended at least one consultation.

This approach ensured that participants had prior experience with the MyHealth patient portal, facilitating differentiation between StartHelp itself and the overarching portal infrastructure, MyHealth. We considered that this distinction would have been more difficult if both systems had been introduced simultaneously for the first time. During usability testing, the facilitator explicitly presented StartHelp as a distinct program accessed within MyHealth, and participants were instructed to evaluate StartHelp separately from the patient portal.

Research on usability testing has found that samples of 5 participants typically identify approximately 80% of usability issues [35]. To increase the likelihood of identifying not only common but also less frequent and potentially critical usability issues, we chose to include 10 participants. A sample of this size is considered sufficient to approach thematic saturation in qualitative usability data and to capture a broad range of user experiences relevant to refining the intervention before evaluation in a clinical trial. As recommended by Maramba et al [36], both qualitative and quantitative data were collected and interpreted together. This approach enables a comprehensive assessment of usability while highlighting specific issues that require attention.

To recruit patients for testing, the first author discussed the project with clinicians at a community mental health center in central Norway. There were no formal exclusion criteria other than requiring acute care due to suicide risk, current psychosis, or an inability to provide informed consent. Clinicians then asked patients whether they would be interested in learning more about a research project concerning a digital pretherapy intervention. If patients expressed interest and consented to be contacted, the first author (HP) called them, explained the project details (eg, the testing goals and expected duration), and scheduled a date and time for testing.

Data Collection and Analysis

The individual testing comprised 3 phases and lasted between 45 minutes and 1 hour in total. First, patients completed a series of questionnaires; afterward, they were assigned a set of tasks that included watching 4 selected videos before completing another set of questionnaires. At the end, they were asked the open-ended question, "Do you think all relevant issues regarding the app you tested today have been addressed in our conversation and the open questions in the questionnaires?" The think-aloud interviews were conducted by a facilitator (HP) who received training in the think-aloud methodology, including guidance on using neutral prompts and minimizing leading questions to reduce potential interactional bias. A member of the research team was present throughout all phases to observe and provide assistance. If participants had any questions, the facilitator responded with standard phrases such as "What do you think?" "What would you do?" "Are there any ambiguities we should become aware of?" or "If you find it difficult to circle an answer, please circle the response that you first thought of."

Tasks

After answering the first set of questions, participants were instructed to complete a series of tasks. The instructions provided to participants before starting are presented in Textbox 1. The tasks themselves are presented in Textbox 2. During task completion, the concurrent think-aloud method was used [37]. The facilitator observed all participants during the tasks but sat

XSL•FO
RenderX

at a desk elsewhere in the room while they completed the questionnaires.

Textbox 1. Instructions given orally to participants before starting on tasks.

- "Now, you will receive a series of tasks from me. These tasks will be quite straightforward and will involve navigating the program called StartHelp. The purpose of this exercise is to determine whether the program is easy to use."
- "The tasks may be easy for you to complete, but if they prove challenging, this information is valuable to us. Only by identifying what is difficult or illogical and trying to understand why can we determine what improvements are necessary. Therefore, we encourage you to think out loud while you solve the tasks. If, for example, I ask you to find a video titled 'social phobia', try your best while talking about your thought process."
- "If at any point you find yourself stuck or uncertain about what to do, let me know. I will take note of what was difficult and assist you before moving on to the next task."
- "I will be recording what you say so I can later write it down. This will help me to better remember everything that is said. None of what we do today will be stored with names, and the audio file will be deleted as soon as I have transcribed it. I will let you know when I have started the recording, and the first thing I would like you to do at that point is to confirm that you are aware I am recording, with a simple yes. After that, we will proceed with the tasks."
- "While we carry out the tasks, you will also see some of the videos we have created for the final program. We have selected these videos specifically because we want your feedback. After the tasks, you will be asked what you think about them."
- "Let me know after you have completed each of the tasks."

Textbox 2. Tasks given to participants during testing.

- "I have activated the StartHelp program for you. The first thing I would like you to check is whether you have received any form of SMS text message or email regarding this."
- "Then, view this notification and open StartHelp."
- "When you are inside the patient portal, the first thing I want you to visit is a page called 'StartHelp mental health'. When you are on this page, please read the information provided. After that, you may press 'Complete task'."
- "Next, I would like you to locate a page that provides information about various treatment options. Watch the video on this page."
- "After this, navigate around the same page and see if you can: (1) Find a video discussing various treatment options and watch it. (2) Find a video about social anxiety disorder and watch it. (3) Find a video about panic disorder and watch it. (4) Find a video about depression and watch it."
- "After this, you can return to the app and 'complete the task'."
- "Now you can completely exit the app, returning to your phone's home screen. After that, attempt to access StartHelp again."
- "Now, I would like you to locate a page that says something about patient rights and click on the link provided. Afterward, I want you to try to get back to the app."
- "Then, I would like you to navigate back to the social phobia video so that you can watch it again."
- "If this is difficult, please return to the 'Welcome to StartHelp' page, where you can find instructions on how to locate it."
- "Now, I would like you to return and locate a page that contains information about 'e-treatment', read the content provided there, and click the button to contact 'The centre for eTreatment'."

Videos Tested

Participants were shown 4 specific videos during task completion that the project group had determined required a more comprehensive evaluation. One video presented various treatment options available at the outpatient clinic, including face-to-face therapy (both individual and group therapy), video teletherapy, therapist-guided internet cognitive behavioral therapy, pharmacotherapy, and psychomotor physiotherapy. This video aimed to raise patients' awareness of different treatment options and to encourage discussions with their therapist regarding treatment preferences while assessing treatment needs. The graphic profile of this video is presented in Figure 3. The other 3 videos addressed challenges associated with specific common diagnoses, such as social phobia, panic disorder, and depression. The graphic profile of these videos is illustrated in Figure 5.

These 4 videos, in particular, were tested for 2 primary reasons. First, as mentioned previously, the 3 animated videos depicting specific diagnoses had not yet been evaluated by patients or user representatives. Second, the "treatment options" video was tested because user representatives emphasized its particular importance.

Questionnaires

Demographic Variables

In addition to the aforementioned questionnaires, participants were asked to provide demographic information, including age, gender, marital status, living situation, education level, and employment status.

XSL•FO
RenderX

Open-Ended Questions

To further explore potential usability issues, the questionnaire included the open-ended questions presented in Textbox 3. Patients provided free-text responses.

Textbox 3. Open-ended questions.

- Before completing tasks and watching videos.
  - Do you know about the various treatment options available at the community mental health center? If so, which ones?
- After completing tasks and watching videos.
  - Do you know about the various treatment options available at the community mental health center? If so, which ones?
  - Does the MyHealth app have any specific areas for improvement?
  - Do the videos you watched have any specific areas for improvement?
  - Do you believe there are aspects of these videos that some patients may find condescending, offensive, or uncomfortable, either in the short or long term?

The System Usability Scale

The SUS [38] is one of the most widely used questionnaires for measuring the usability of digital systems and technology. It consists of 10 items, half positively phrased (eg, "I thought the system was easy to use.") and half negatively phrased (eg, "I found the system unnecessarily complex."). Each question is scored on a 5-point scale, ranging from 0 (strongly disagree) to 4 (strongly agree). Scores are typically multiplied by 2.5 to create an intuitive score range of 0-100 [39]. Benchmarks exist for educational tools [40], digital health apps [27], and patient portals [28] based on Scandinavian norms.

Given StartHelp's integration with the MyHealth portal, the SUS was adapted into 2 different versions. Participants completed the SUS twice: once with MyHealth navigation and functionality in mind (SUS-MyHealth; eg, login, navigation between services; "I thought that MyHealth was easy to use") and once with StartHelp content in mind (SUS-Videos; eg, clarity, structure, and presentation of tasks and videos; "I found the videos unnecessarily complex"). This approach was used to reinforce the distinction between the overarching patient portal and the StartHelp program during evaluation.

Perceived Quality of the Videos and Adverse Effects

To assess perceived quality, we used a single-item scale ranging from 0 to 100, where 0 indicated *very low quality* and 100 represented the *highest possible quality.* Moreover, to assess how uncomfortable or offensive patients found the videos, we used a scale from 0 to 5, ranging from *not at all* to a *very large degree.*

Four-Item Client Satisfaction Questionnaire

In this study, we used 4-item Client Satisfaction Questionnaire (CSQ-4) scores to assess participants' baseline satisfaction, not to evaluate the effectiveness of treatment or the StartHelp intervention. The CSQ-4 [41,42], a shortened version of the 8-item Client Satisfaction Questionnaire [43,44], is a 4-item questionnaire that measures overall treatment satisfaction. Each item is answered on a 4-point scale, ranging from *no, definitely not/quite dissatisfied* to *yes, definitely/very satisfied*, yielding a possible total score range of 4-16. The Norwegian translation of the CSQ-4 [42,45] has shown psychometric properties comparable to those of the English version [41]. Cutoff values are based on Larsen et al [43], where a score of 4-10.5 indicates low satisfaction, 10.6-13.5 indicates medium satisfaction, and 13.6-16 indicates high satisfaction.

Client Satisfaction Questionnaire-Video

In addition to the standard CSQ-4, we used a modified version to assess satisfaction with the video content in StartHelp (hereafter referred to as CSQ-Video). Items were adapted to refer specifically to the videos (eg, "In an overall, general sense, how satisfied are you with the videos you have received?"). This adapted version has not been psychometrically validated and was therefore used for exploratory purposes only, to inform further refinement of the intervention rather than to evaluate acceptability or effectiveness.

Quantitative and Qualitative Data Analyses

Given the small number of patients in our sample, we analyzed all quantitative measures descriptively. We interpreted quantitative data alongside qualitative data to inform refinement and assess readiness for further evaluation in a clinical trial, rather than to establish effectiveness. Qualitative data consisted of think-aloud interview transcripts collected during task completion and participants' written responses to open-ended questions. These data were analyzed iteratively, with all comments and suggestions reviewed. Potential usability barriers were coded into categories, and comments were organized into themes [46]. Afterward, the working group responsible for the development of StartHelp discussed both the themes and categories.

Answers to questions (considering both before and after; see Textbox 3) "Do you know about the various treatment options available at the community mental health center? If so, which ones?" were first categorized and then counted. The project group organized the responses into meaningful categories: different treatment modalities and treatments that differed in ways that make a practical difference. For example, "psychoeducation" and "psychoeducation for family members" were counted as 2 separate suggestions, as the inclusion of family members was considered significant. Conversely, "talking to a psychologist" and "talking therapy" were not regarded as distinct treatment options. However, "individual

XSL•FO
RenderX

psychotherapy" and "psychotherapy over video" were classified as distinct, because awareness of video consultations may lead a patient to make meaningful choices during treatment. The listings of treatment options from all patients have been translated and are included in Multimedia Appendix 2.

Integration of Results Into the Development Process Through Discussion

After the first author (HP) synthesized the results, they were presented to the project group for discussion. The project group examined each patient's suggestion and issue in relation to StartHelp's overall goals. Potential solutions were then proposed. During this process, several key questions were considered: "Is a change in StartHelp warranted?" "What is the ideal change?" and "Is a change in StartHelp possible?"

Results

Development Process

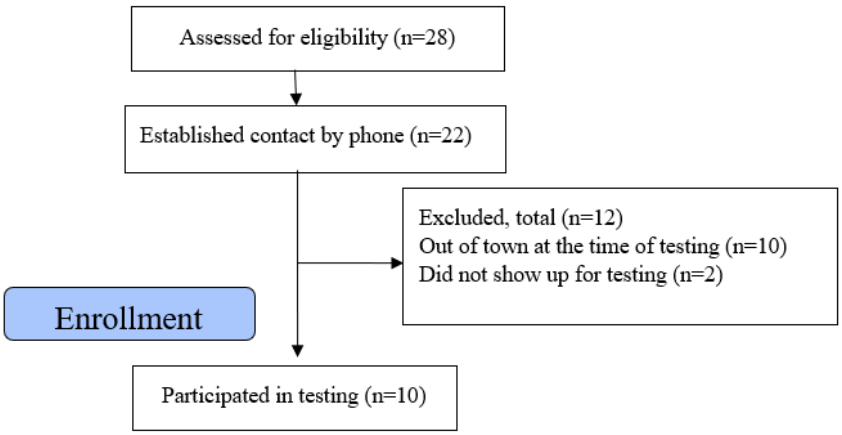
The development process spanned approximately 2 years, from February 12, 2021, to January 26, 2023. During this period, we created StartHelp, consisting of a total of 27 tasks. These tasks included 24 videos and links to 14 websites containing in-depth information. A discussion of the usability testing results follows the presentation of these findings.

The insights from the initial prototype testing resulted in the following themes: the MyHealth app, the information letter regarding StartHelp, the questionnaire, general concerns, tasks within StartHelp, information in MyHealth, information on hospital websites, and videos. This feedback influenced the initial integration of StartHelp into MyHealth. As StartHelp was one of the first programs integrated into MyHealth, feedback from the user tests also led to change requests concerning MyHealth's core infrastructure, which may benefit future programs. A digital representation of the qualitative analysis process of the Post-it notes created during prototype testing and the abstracted themes is provided in Multimedia Appendix 3.

Usability Study Results: Participants

To assess usability, we invited 28 patients to participate. Of these, 6 did not answer their phones after 2 attempts, 10 were out of town at the time of testing because it took place during the summer months, and 2 withdrew after scheduling an appointment. Recruitment concluded when 10 participants had been included. Figure 7 presents a flowchart of this process.

Figure 7. Flowchart of the usability study recruitment process.



The mean age of the sample was 31.8 (range 21-44) years. Half of the participants (n=5) were women, and 6 of the 10 were married or cohabiting. Most (n=8) had completed 3 or more years of higher education, and 7 were on sick leave. Table 1 presents the collected demographic information for each patient.

XSL•FO
RenderX

Table 1. Demographic information about participants.

Patient ID	Gender; age (years)	Marital status	Living situation	Education[a]	Employment status	Work domain
Patient 1	Female; 44	Not married	Living with someone	Higher secondary education	Currently on sick leave	Management/administration
Patient 2	Female; 33	Married/cohabiting	Living with someone	Higher education—3 years or less	Currently on sick leave	Cultural work/design/music/sport
Patient 3	Male; 31	Married/cohabiting	Living with someone	Higher education—3 years or less	Currently on sick leave	Health care
Patient 4	Female; 40	Married/cohabiting	Living with someone	Higher education—5 years or more	Employed	Education
Patient 5	Female; 33	Not married	Living with someone	Higher education—3 years or less	Partial sick leave	Education
Patient 6	Male; 31	Married/cohabiting	Living with someone	Higher education—3 years or less	Currently on sick leave	Management/administration
Patient 7	Male; 26	Married/cohabiting	Living with someone	Higher education—5 years or more	Employed	Management/administration
Patient 8	Male; 23	None of the answers fit my situation	Living with someone	Higher education—3 years or less	None of the answers fit my situation	Management/administration
Patient 9	Female; 36	Married/cohabiting	Living with someone	Higher education—3 years or less	Partial sick leave	Health care
Patient 10	Male; 21	Not married	Living with someone	Higher secondary education	Currently on sick leave	Construction

[a]Education refers to the highest completed education level.

Usability Issues Emerging From Task Completion

Overall, patients experienced few usability issues during task completion and made few suggestions for improvement, although one type of issue was mentioned by all participants. These issues primarily involved navigating the overarching app infrastructure, MyHealth, rather than StartHelp itself or the videos. Participants generally felt that too many features were presented that were not relevant to them at this time. This made the MyHealth home page difficult to navigate, especially when trying to find videos they had previously watched (n=10), which had to be accessed elsewhere than through the StartHelp dashboard after completing the related task. In the original design, tasks were automatically removed from the StartHelp dashboard shortly after completion. Other comments included that the text in MyHealth was too small (n=1) and that the "sheer number of options presented at once was overwhelming" (n=3). Finally, 3 patients had their language settings in MyHealth set to English instead of Norwegian, which made the StartHelp instructions for navigating the MyHealth menu redundant, as they referred only to the Norwegian menu.

Finally, the fourth video, which addressed depression, had not been properly uploaded. This issue was discovered during testing, and 4 patients were therefore unable to view or comment on the video. After that day of testing, the project team contacted the information technology personnel responsible, who reuploaded the video, checked all other videos, and reviewed the upload procedures to determine whether the problem was caused by a systematic error.

Quantitative Data

Treatment Options

Table 2 provides an overview of the treatment options reported by patients. Several general trends emerged from the questionnaires; 4 of the 10 patients reported a greater number of available treatment options after viewing the videos, with the total number of treatment options mentioned increasing from 19 to 31. Notably, 4 patients reported knowing no treatment options before task completion. All treatment options mentioned by patients are listed in Multimedia Appendix 2.

XSL•FO
RenderX

Table 2. Overview of treatment options mentioned by patients.

Treatment options[a]	Frequency before videos, n	Frequency after videos, n
Individual psychotherapy	6	9
Group psychotherapy	4	7
Video consultations	1	3
Therapist-guided internet treatment	2	1
Group exposure therapy	1	1
Psychoeducation	2	2
Family-oriented psychoeducation	1	1
Physical therapy	1	2
Counseling by a vocational counselor	1	1
Pharmacotherapy	0	3
Total suggestions	19	30

[a]A complete overview of the treatment options mentioned by patients is provided in Multimedia Appendix 2.

System Usability Scale and the Perceived Quality of the Videos and Adverse Effects

All patients had either equal scores on the SUS-MyHealth (mean 63.6) and the SUS-Videos (mean 85) or higher scores on the SUS-Videos.

The depression video received the highest rating; however, this score may be artificially inflated, as only 6 participants viewed the video. Moreover, the 4 participants who did not view the depression video tended to rate the videos somewhat lower than the other 6. The video on treatment options received the second-highest rating.

Most patients rated the videos as not at all or only minimally offensive (mean 1.2; range 0-5). However, 1 patient rated the videos as uncomfortable or offensive "to a very large degree," while another rated them "to some degree." Table 3 presents an overview of all patient scores.

Table 3. Descriptive presentation of survey data.

Patient ID			Perceived quality of the videos tested				
	SUS[a] MyHealth	SUS[a] videos	Treatment options	Panic disorder	Social anxiety disorder	Depression	Feeling of discomfort when watching videos
Patient 1	47.5	87.5	100	100	100	__[b]	0
Patient 2	57.5	75	60	70	80	—	3
Patient 3	35	72	80	50	10	—	5
Patient 4	85	87.5	90	30	30	—	1
Patient 5	95	95	100	90	90	90	1
Patient 6	65	90	95	95	95	95	0
Patient 7	47.5	70	80	90	90	90	1
Patient 8	—	87.5	70	70	70	70	0
Patient 9	77.5	87.5	90	100	100	100	1
Patient 10	62.5	85	60	60	70	70	0
Sample mean	63.6[c]	83.7	82.5	75.5	73.5	85.8	1.2

[a]SUS: System Usability Scale.

[b]Not available (patients 1-4 are missing perceived quality scores on the "depression" video due to technical problems during testing. The mean for this video was therefore calculated only for patients 5-10; patient 8 did not answer the question).

[c]This number was divided by 9 due to 1 missing datapoint.

Satisfaction

Regarding the CSQ-4, the sample mean was 12.3, indicating medium satisfaction with treatment [43]. The score range was 9-15. The mean score on the CSQ-Video was somewhat lower (10.9), indicating satisfaction in the lower medium range [43]. The score range for the CSQ-Video was 8-14. Table 4 presents

XSL•FO
RenderX

an overview of the means and SDs for all items, in addition to the total scores.

Table 4. Satisfaction scores.

Item	CSQ-4[a], mean (SD)	CSQ-Video[b], mean (SD)
Met needs	3.2 (0.63)	2.8 (0.42)
Improvement in self-efficacy	2.9 (0.88)	2.3 (0.48)
Overall satisfaction	3.3 (0.82)	3.1 (0.74)
Come back	3.3 (0.48)	2.7 (0.67)
Total score[c]	12.3 (2.26)	10.9 (1.91)

[a]CSQ-4: 4-item Client Satisfaction Questionnaire; score range 4-16.

[b]CSQ-Video: a modified version of the CSQ-4 for evaluating informational videos; score range 4-16.

[c]A score of 4-10.5 is regarded as low satisfaction, 10.6-13.5 medium satisfaction, and 13.6-16 high satisfaction.

Qualitative Data: Usability Issues

In response to the open-ended questions regarding potential improvements, 4 patients suggested improving the MyHealth app by tidying up the menu and reducing the number of options displayed at once. Another suggestion was to keep tasks visible on the StartHelp dashboard after completion. This would eliminate the need to exit the StartHelp dashboard and navigate through the MyHealth app to locate previously completed tasks elsewhere in the menu. In the current design, tasks were automatically removed from the StartHelp dashboard upon completion and stored under “Completed tasks.”

Patients made almost no suggestions for improvements regarding the videos. However, they raised 3 main concerns. First, 6 patients noted that not all individuals would identify with the content, stating, for example, “Social anxiety is so different for each person” or “If I do not experience my anxiety the way the video presents it, will I begin to think I do not have the disorder?” By contrast, 1 patient reflected on the risk of “self-diagnosis” that viewers might engage in while watching the videos. Second, 3 participants expressed frustration that the videos “only presented problems” and did not offer ways to cope with the symptoms. These 3 patients emphasized the need for further clarification of the types of treatment “we could offer them” to manage the issues presented in the videos.

Third, 5 patients expressed concerns that the videos might be distressing for some viewers, especially those dealing with the issues presented. For example, 1 patient reported becoming aware of their heartbeat while watching. As the videos were intended to help individuals on the waiting list identify with the scenarios presented, we considered this feedback potentially concerning and followed up with additional questions. We asked how distressed they found the videos and whether they had personal experience with the mental disorders portrayed. Interestingly, none of the patients reported significant distress in the form of anxiety, although they speculated that someone else might find the videos more distressing. The quantitative data corroborate these findings. Moreover, patients who spontaneously reported previous experiences with social anxiety or panic attacks rated these videos more highly than the other videos.

One patient reported difficulty concentrating during the 2 videos that presented clinical vignettes (only 2 were viewed due to a malfunction with the depression video) and described the sound design as chaotic. The same patient rated the videos as uncomfortable or offensive “to a very large degree.” In a follow-up question—“This is very important to us. We want to know what about the videos we made caused you to feel uncomfortable, and in what way.”—the patient explained that their discomfort stemmed from frustration at being unable to concentrate during the videos, rather than from anxiety or indignation. When asked further, the patient stated that they had no issues with the video on treatment options.

Addressing the Usability Test Results in the Project Group

Issues Identified As Important

The project group discussed all major themes and the usability issues identified. The most pressing concern was the navigation issue, which made it difficult for patients to locate completed tasks in MyHealth outside the StartHelp dashboard. Most of the discussion focused on how to resolve these issues without modifying the MyHealth infrastructure, which would require changes deeper in the technological architecture. After reviewing the information presented to patients, the project group concluded that the guidance on how to schedule the first session as a video consultation required improvement. Other issues discussed included potential discomfort when watching the videos and whether they effectively served their intended purpose. Regarding the MyHealth app, an issue raised by user representatives during preliminary testing of the software following the paper prototype tests was that they experienced the login procedures as unnecessarily cumbersome.

We communicated all findings to the owners of MyHealth, which is part of the regional hospital administration, through 2 complementary approaches. First, we held verbal discussions to explore potential solutions that could be implemented without significantly altering MyHealth’s core features. Second, any new programs or modifications to existing programs require final approval from the owners—a safety mechanism that ensures appropriate regulation of MyHealth as a platform for innovation. Therefore, all requested changes were formally registered and archived.

XSL•FO
RenderX

Changes Made to MyHealth and StartHelp

Few changes were made to MyHealth during the development and usability testing of StartHelp. However, new login alternatives were added, which now include login via fingerprint or facial recognition, as well as faster login using a personalized numeric PIN (personal identification number). Following the usability testing of StartHelp, 1 improvement implemented was the provision of clearer information on how patients can request a video consultation for their first appointment. This information is now prominently displayed on the web page containing more in-depth details about the first session. We did not modify the videos themselves, as the treatment options video already presents video consultations as an option and encourages viewers to consult the additional information on the website. Notably, it was this video that prompted 1 patient to request more concrete information on the topic. Additionally, information on how to locate previously completed tasks is now clearly presented in bold at the bottom of every task: "Remember that all completed tasks can be found in the MyHealth menu under 'Education' and then 'StartHelp.' Go there if you wish to view your previous tasks." Moreover, we updated the app to allow tasks to remain on the dashboard for a full week after completion.

Regarding the videos, we concluded that they were unlikely to cause unnecessary discomfort. One patient rated the 3 videos containing patient stories as uncomfortable or offensive "to a very large degree," primarily due to chaotic sound and graphic design that made it difficult to concentrate. This patient did not report any issues with the treatment options video, which had been revised based on workshop feedback.

The project group reasoned that individuals who have difficulty tolerating the current video design are unlikely to be part of the target population for the associated treatments, which focus specifically on social anxiety disorder, panic disorder, or depression without concurrent disorders. Additionally, the videos are optional, presented alongside written information, and accessed primarily by users seeking in-depth information about these treatment modalities. Although some degree of emotional discomfort may be unavoidable when videos encourage patients to identify with patient stories, the overall reported discomfort during testing was relatively low.

Keeping StartHelp's goal in mind—namely, to inform patients and encourage active involvement in treatment decisions so that they can better tailor their treatment—we chose not to modify the videos themselves. Instead, we revised the surrounding text to strengthen the connection between the videos and the clinic's specific treatment options. For instance, using therapist-guided internet therapy as an example:

> *Panic disorder, social anxiety, and depression are three disorders currently being treated through therapist-guided internet therapy at the Centre for eTherapy. If you believe that one of these disorders may be your primary concern and are interested in trying out therapist-guided internet therapy, here is how to contact us for an assessment to determine whether this approach may be beneficial for you. Therapist-guided internet therapy typically has much shorter waiting lists, and participating in this assessment will not remove you from the current waiting list for other treatments.*

We concluded that the aforementioned changes were sufficient. As the project group did not anticipate any new critical usability issues arising from the proposed changes, no additional isolated usability testing was deemed necessary. The remaining potential issues were considered best evaluated within the context of a clinical trial.

Discussion

Principal Findings

This study had 2 objectives: to describe the development of StartHelp and to explore its usability and perceived user satisfaction. The purpose was to evaluate whether StartHelp was ready for further evaluation in a clinical trial. The study, therefore, represents an exploratory, early-stage evaluation rather than definitive evidence of StartHelp's usability or user satisfaction, as such evidence would require further testing in its intended real-world context. In the following sections, we discuss the development process in accordance with the GRIPP2 guidelines [19], the usability findings, user satisfaction with the video content, and implications for future research.

Development Process

When developing StartHelp, we aimed to ensure that user representatives were involved as "partners" [33]. To achieve this, we engaged user representatives who had prior experience working with members of the project team and were familiar with the interventions on which StartHelp is based [7-9]. The project was designed with a long-term collaborative relationship in mind, reinforcing the academic team's commitment to the perspectives and goals of user representatives. Our approach was informed by previous research on user involvement, which indicates that even in nonhierarchical collaborations, an indirect hierarchy can develop, especially when partnerships are short term [20]. However, partnering with the same user representatives over an extended period may also have drawbacks, as it could limit the diversity of viewpoints represented.

To mitigate this, we included patients with no prior experience working with the project group in the initial focus groups and subsequent usability testing. Another way to mitigate this issue could have been to involve at least one patient with no prior experience working with the project group. During the development phase, before usability testing, we primarily involved patients with attention-deficit/hyperactivity disorder or autism spectrum disorder, which may have introduced bias. As previously stated, however, we reasoned that if StartHelp is accessible to these populations, it should also be accessible to most other populations.

Usability

In line with the recommendations of Maramba et al [36], the usability evaluation included both quantitative and qualitative data. The quantitative data revealed some discrepancies. The average SUS scores for the sample were 63.6 for the MyHealth

XSL•FO
RenderX

app and 83.7 for the videos. Notably, none of the patients rated the MyHealth app higher than the videos, indicating that navigating the overarching MyHealth app was more complicated than engaging with the informational video content and using the StartHelp dashboard. According to Bangor et al [47], a score of 83.7 is considered good to excellent, whereas a score of 63.7 is deemed acceptable. Moreover, a mean score of 63.6 is below the benchmark for digital health apps (68) [39] and for other Norwegian patient portals (72.1) [28]. This suggests that MyHealth's usability may be lower than that of the average health app or patient portal. Intuitively, there should not be such a large difference in SUS scores between the Norwegian benchmark values and those observed for MyHealth. This discrepancy may be explained either by the small sample size or by interactions between StartHelp's features and MyHealth, which may result in usability challenges not present when using MyHealth alone. This finding is explored further below in the discussion of the qualitative data. However, the marked difference between usability ratings for StartHelp and the MyHealth portal further suggests that participants were able to distinguish between the 2 systems during testing.

Satisfaction and Video Quality

Mean scores and SDs on the CSQ-4 are comparable to those observed in Norwegian specialist mental health care [42,48,49], suggesting that the current sample was representative in this regard. Moreover, none of the patients who participated in the usability testing reported extreme CSQ-4 scores before testing, making it less likely that overall treatment satisfaction biased the results. By contrast, scores on the CSQ-Video indicated satisfaction in the lower medium range [43]. Few studies have used the CSQ in the context of pretherapy interventions [8,50], and only 1 used the CSQ to evaluate patients' satisfaction with the intervention itself [8]. By contrast, Zwick and Attkisson [50] used it to assess how their intervention affected patients' perceptions of services provided after treatment ended. Lara-Cabrera et al [8] reported an item-level mean of 3.28 on the CSQ-8 [8], which is higher than the item-level mean in this study (2.73). This discrepancy may be attributed to the use of different CSQ versions; however, this explanation seems unlikely because the CSQ-8 and CSQ-4 have demonstrated similar psychometric properties and mean values in outpatients receiving specialized mental health care [42,51]. The discrepancy is more likely related to CSQ's association with respondents' personal needs at the time of assessment [44,52] or to the expected clinical utility within the study context. We would not expect the brief, nonclinical context of this study's usability testing to be sufficient to address participants' current problems in a general sense (ie, to alleviate mental health symptoms). This interpretation is further supported by the reported perceived quality of the videos, which was considerably higher than the CSQ-Video scores. For example, patient 5 reported low satisfaction (10; cutoff 10.5 [49]) yet rated the video quality as high (mean score 92.5 out of 100). Nevertheless, further research on the full StartHelp intervention—tested in its intended context, in which patients awaiting treatment receive the complete program—is needed to more comprehensively evaluate user satisfaction with the program.

Qualitative Findings

In general, the qualitative data derived from patients' task completion and responses to open-ended questions corroborated the quantitative findings. Most reported usability problems stemmed from attributes of the MyHealth app (eg, default language settings and text size) and from its navigation. Patients reported no usability issues with the StartHelp dashboard. Other researchers have noted that a major barrier to the widespread integration and implementation of digital interventions is poor or outdated information technology infrastructure [16,53,54]. Our results may reflect a similar tension between the infrastructure and the specific interventions designed within it. Building apps within an overarching app and patient portal, such as MyHealth, offers several advantages, including secure handling of patient data and integration with the patient administration system to provide easy access to other health information collected through routine clinical practice. One downside, however, is that app development within such infrastructures may be constrained in terms of design and available features. Moreover, adding new features to an app or intervention may prove difficult if doing so requires fundamental changes to the overarching technological infrastructure. It is also challenging to predict whether such modifications will result in unintended interactions with other functions or integrated apps. We attempted to mitigate these challenges by conducting initial prototype testing and limiting changes to MyHealth, instead modifying only aspects of StartHelp, as discussed above in the "Changes Made to MyHealth and StartHelp" section. These changes were intended to minimize the need for patients to switch between the MyHealth menu and the StartHelp dashboard.

Patients also offered reflections on the videos, which primarily fell into 2 categories: their function and their potential to cause discomfort. Ultimately, we modified the accompanying information rather than the videos themselves. Interestingly, although several patients expressed concern that the videos might disturb others and provoke anxiety symptoms, they reported experiencing few symptoms themselves. Although we did not alter the videos, we ensured that none were mandatory and that all content was also readily available in text format.

Strengths and Limitations

Five users are often sufficient to identify approximately 80% of usability problems [35]. However, this depends largely on the nature and number of issues present in the program; common and critical problems are generally easier to detect. Although we cannot be certain that all usability issues were identified, most of the tasks assigned to patients were completed without difficulty. Therefore, we consider our sample size of 10 adequate for identifying critical usability issues before further testing in a clinical trial. A strength of the study is that actual patients were involved in the testing and that both quantitative and qualitative data were collected [36]. However, a sample of 10 is small for quantitative estimates of usability and satisfaction, and further evaluation with larger samples is required. The qualitative data consisted primarily of written responses to open-ended questions; in-depth interviews could provide additional insights.

XSL•FO
RenderX

Another strength is the involvement of user representatives, as well as the study's novelty. To our knowledge, no other digital pretherapy interventions have been developed for a general outpatient population in specialized mental health care in close collaboration with user representatives.

Although sex was evenly distributed in our sample, most participants were highly educated and relatively young. We consider the sample representative of the clinical population treated at the outpatient clinic and at other similar Norwegian outpatient specialized mental health services [55]. Nevertheless, the absence of older adults or individuals with little formal education is a limitation. Although these groups represent minorities in similar outpatient settings [53], further real-world testing should examine whether they encounter difficulties using StartHelp—for example, by assessing lower app usage or higher dropout rates. As we developed StartHelp for a Norwegian context, the results may not be generalizable to international settings. However, some issues that emerged during testing may be transferable, as developing programs within existing infrastructures may be necessary for full implementation within current health care information technology systems. Moreover, MyHealth is comparable to other patient portals across Norway and other Nordic countries due to collaborations with the same health system providers [35,37,38].

Another limitation is the staged setting in which the testing took place. We assessed the usability of the program itself; however, we do not know whether issues specific to more naturalistic use may arise during a potential feasibility trial involving a waiting list population, in which most users are unlikely to explore all aspects of the app simultaneously. The current study design is well-suited to identifying concrete usability and technical issues related to navigating and using StartHelp. However, further research is needed to evaluate how patients use StartHelp before their first consultation and how this influences patient satisfaction.

Additionally, further research is needed to establish StartHelp's usability and to evaluate its preliminary effects on patient satisfaction with treatment. Conducting a randomized controlled feasibility trial that recruits patients awaiting their first consultation in specialized mental health care would be particularly valuable. Moreover, StartHelp's intended downstream effects on treatment outcomes—by enhancing treatment satisfaction and patient-reported outcomes associated with improved clinical outcomes—should be investigated further.

Conclusions

Through a collaborative co-creation process, we developed StartHelp, a pretherapy psychoeducation program for outpatients awaiting specialized mental health treatment. During usability testing, a few issues related to StartHelp itself emerged; however, participants reported difficulties navigating the overarching MyHealth portal, which received usability scores lower than hypothesized. The project group discussed and addressed the problems identified during usability testing, thereby minimizing the need to navigate the overarching patient portal. Although the perceived quality and usability of the clinical material were high, patients reported lower satisfaction than anticipated. Taken together, these findings suggest that StartHelp is suitable for further evaluation in a feasibility trial conducted within its intended context, involving outpatients awaiting specialized mental health treatment.

Authors' Contributions

Conceptualization: HP, TS, LSE, AM, CDC, AH, MLLC
Data collection: HP
Design: TS, AM, MLLC
Formal analysis: HP
Funding acquisition: HP, MLLC
Methodology: HP, CDC, AH, MLLC
Supervision: AH, MLLC
Writing – original draft: HP
Writing – review & editing: TS, LSE, AM, AH, CDC, MLLC

Conflicts of Interest

None declared.

Multimedia Appendix 1
The Guidance for Reporting Involvement of Patients and the Public (GRIPP2) checklist.
[PDF File (Adobe PDF File), 421 KB - humanfactors_v13i1e80130_app1.pdf]

Multimedia Appendix 2
Treatment options mentioned by patients during usability testing.
[PDF File (Adobe PDF File), 246 KB - humanfactors_v13i1e80130_app2.pdf]

Multimedia Appendix 3

XSL•FO
RenderX

Post-it notes created during prototype testing and the abstracted themes.
[PDF File (Adobe PDF File), 294 KB - humanfactors_v13i1e80130_app3.pdf]

References

1. World Health Organization. World Mental Health Report: Transforming Mental Health for All. Geneva, Switzerland: World Health Organization; 2022.
2. Kazdin AE. Perceived need for treatment for mental disorders: a review and critical evaluation. Clin Psychol Rev 2025 Jul;119:102591. [doi: 10.1016/j.cpr.2025.102591] [Medline: 40381402]
3. Thomas KA, Schroder AM, Rickwood DJ. A systematic review of current approaches to managing demand and waitlists for mental health services. MHRJ 2020 Nov 05;26(1):1-17. [doi: 10.1108/mhrj-05-2020-0025]
4. Biringer E, Sundfør B, Davidson L, Hartveit M, Borg M. Life on a waiting list: how do people experience and cope with delayed access to a community mental health center? Scand Psychol 2015 Apr 25;2:1-16. [doi: 10.15714/scandpsychol.2.e6]
5. Wampold BE. Brave new world: mental health services 25 years since Dodo and 25 years in the future. Psychother Res 2023 Apr 10;33(4):533-534. [doi: 10.1080/10503307.2022.2154177] [Medline: 36628467]
6. Cuijpers P. The Dodo Bird and the need for scalable interventions in global mental health-a commentary on the 25th anniversary of Wampold et al. (1997). Psychother Res 2023 Apr 28;33(4):524-526. [doi: 10.1080/10503307.2022.2138792] [Medline: 36305328]
7. Koksvik JM, Linaker OM, Gråwe RW, Bjørngaard JH, Lara-Cabrera ML. The effects of a pretreatment educational group programme on mental health treatment outcomes: a randomized controlled trial. BMC Health Serv Res 2018 Aug 29;18(1):665-610 [FREE Full text] [doi: 10.1186/s12913-018-3466-2] [Medline: 30157839]
8. Lara-Cabrera M, Gjerden M, Gråwe RW, Linaker O, Steinsbekk A. Short-term effects of a peer co-led educational programme delivered before mental health treatment: a randomised controlled trial. Patient Educ Couns 2016 Jul;99(7):1257-1261 [FREE Full text] [doi: 10.1016/j.pec.2016.02.006] [Medline: 26905956]
9. Lara-Cabrera ML, Salvesen, Nesset MB, De las Cuevas C, Iversen VC, Gråwe RW. The effect of a brief educational programme added to mental health treatment to improve patient activation: a randomized controlled trial in community mental health centres. Patient Educ Couns 2016 May;99(5):760-768 [FREE Full text] [doi: 10.1016/j.pec.2015.11.028] [Medline: 26682971]
10. Jardine J, Bowman R, Doherty G. Digital interventions to enhance readiness for psychological therapy: scoping review. J Med Internet Res 2022 Aug 30;24(8):e37851 [FREE Full text] [doi: 10.2196/37851] [Medline: 36040782]
11. Walitzer KS, Dermen KH, Connors GJ. Strategies for preparing clients for treatment. A review. Behav Modif 1999 Jan 01;23(1):129-151. [doi: 10.1177/0145445599231006] [Medline: 9926524]
12. Miglietta E, Belessiotis-Richards C, Ruggeri M, Priebe S. Scales for assessing patient satisfaction with mental health care: a systematic review. J Psychiatr Res 2018 May;100:33-46. [doi: 10.1016/j.jpsychires.2018.02.014] [Medline: 29482063]
13. Hibbard JH. Patient activation and the use of information to support informed health decisions. Patient Educ Couns 2017 Jan;100(1):5-7. [doi: 10.1016/j.pec.2016.07.006] [Medline: 27432014]
14. Hibbard JH, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. Health Aff (Millwood) 2013 Feb;32(2):207-214. [doi: 10.1377/hlthaff.2012.1061] [Medline: 23381511]
15. Bakker N, Knoop J, Hutting N, Heerkens Y, Engels J, Staal J, et al. OP0201-PARE Development through co-creation of a personalized, multimodal, physiotherapist-led, work-oriented intervention to increase work ability in working people with rheumatoid arthritis or axial spondyloarthritis. Annals of the Rheumatic Diseases 2022 Jun;81:132-133. [doi: 10.1136/annrheumdis-2022-eular.477]
16. Balcombe L, De Leo D. Digital mental health challenges and the horizon ahead for solutions. JMIR Ment Health 2021 Mar 29;8(3):e26811 [FREE Full text] [doi: 10.2196/26811] [Medline: 33779570]
17. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug 05;21(3):116-119 [FREE Full text] [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
18. British Standards Institution. Ergonomic Requirements for Office Work With Visual Display Terminals (VDTs): Guidance on Usability. Geneva, Switzerland: International Organization for Standardization; 1998.
19. Staniszewska S, Brett J, Simera I, Seers K, Mockford C, Goodlad S, et al. GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research. BMJ 2017 Aug 02;358:j3453 [FREE Full text] [doi: 10.1136/bmj.j3453] [Medline: 28768629]
20. Vaughn LM, Whetstone C, Boards A, Busch MD, Magnusson M, Määttä S. Partnering with insiders: a review of peer models across community-engaged research, education and social care. Health Soc Care Community 2018 Nov 07;26(6):769-786. [doi: 10.1111/hsc.12562] [Medline: 29512217]
21. Greenhalgh T, Jackson C, Shaw S, Janamian T. Achieving research impact through co-creation in community-based health services: literature review and case study. Milbank Q 2016 Jun 06;94(2):392-429 [FREE Full text] [doi: 10.1111/1468-0009.12197] [Medline: 27265562]

XSL•FO
RenderX

22. Åkerblom KB, Ness O. Peer workers in co-production and co-creation in mental health and substance use services: a scoping review. Adm Policy Ment Health 2023 Mar 17;50(2):296-316 [FREE Full text] [doi: 10.1007/s10488-022-01242-x] [Medline: 36396756]
23. Voorberg WH, Bekkers VJJM, Tummers LG. A systematic review of co-creation and co-production: embarking on the social innovation journey. Public Management Review 2014 Jun 30;17(9):1333-1357. [doi: 10.1080/14719037.2014.930505]
24. McMullin C, Needham C. Co-production and co-creation. In: Co-Production in Healthcare. New York, NY: Routledge; 2018:151-160.
25. Schofield J, Galea-Singer S, Baldacchino A. Co-production in community-based substance use disorder treatment services: a scoping review. Current Opinion in Psychiatry 2025;38(4):265-273. [doi: 10.1097/yco.0000000000001011]
26. Valentine AZ, Hall SS, Sayal K, Hall CL. Waiting-list interventions for children and young people using child and adolescent mental health services: a systematic review. BMJ Ment Health 2024 Feb 01;27(1):e300844 [FREE Full text] [doi: 10.1136/bmjment-2023-300844] [Medline: 38302408]
27. Hyzy M, Bond R, Mulvenna M, Bai L, Dix A, Leigh S, et al. System Usability Scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022 Aug 18;10(8):e37290 [FREE Full text] [doi: 10.2196/37290] [Medline: 35980732]
28. Kujala S, Simola S, Wang B, Soone H, Hagström J, Bärkås A, Cajander, et al. Benchmarking usability of patient portals in Estonia, Finland, Norway, and Sweden. Int J Med Inform 2024 Jan;181:105302 [FREE Full text] [doi: 10.1016/j.ijmedinf.2023.105302] [Medline: 38011806]
29. HelsaMi. URL: https://tjenester.helsami.no/HelsaMi/Authentication/Login [accessed 2026-02-20]
30. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 01;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
31. Hertzum M, Ellingsen G. The implementation of an electronic health record: comparing preparations for Epic in Norway with experiences from the UK and Denmark. Int J Med Inform 2019 Sep;129:312-317. [doi: 10.1016/j.ijmedinf.2019.06.026] [Medline: 31445272]
32. Jormanainen V, Vehko T, Lindgren M, Keskimäki I, Kaila M. Implementation, adoption and use of the Kanta services in Finland 2010-2022. Stud Health Technol Inform 2023 May 18;302:227-231. [doi: 10.3233/SHTI230108] [Medline: 37203652]
33. Roche B, Flicker S, Guta A. Peer research in action I: models of practice. Wellesley Institute. Toronto, ON, Canada: Wellesley Institute; 2011. URL: https://www.wellesleyinstitute.com/wp-content/uploads/2011/02/Models_of_Practice_WEB.pdf [accessed 2026-02-22]
34. Doherty G, Coyle D, Matthews M. Design and evaluation guidelines for mental health technologies. Interacting with Computers 2010 Jul;22(4):243-252. [doi: 10.1016/j.intcom.2010.02.006]
35. Nielsen J, Landauer T. A mathematical model of the finding of usability problems. In: CHI '93: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 1993 May 1 Presented at: INTERCHI93: Conference on Human Factors in Computing; April 24-29, 1993; Amsterdam, The Netherlands p. 206-213. [doi: 10.1145/169059.169166]
36. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
37. Van DHM, De JM, Schellens P. Evaluation of an informational web site: three variants of the think-aloud method compared. University of Twente. 2007 Feb. URL: https://ris.utwente.nl/ws/portalfiles/portal/6459986/Haak07evalutation.pdf [accessed 2026-02-22]
38. Brooke J. SUS-a quick and dirty usability scale. In: Usability Evaluation in Industry. Boca Raton, FL: CRC Press; 1996:4-7.
39. Lewis JR. The System Usability Scale: past, present, and future. International Journal of Human–Computer Interaction 2018 Mar 30;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
40. Vlachogianni P, Tselios N. Perceived usability evaluation of educational technology using the System Usability Scale (SUS): a systematic review. Journal of Research on Technology in Education 2021 Feb 23;54(3):392-409. [doi: 10.1080/15391523.2020.1867938]
41. Greenfield TK. The role of client satisfaction in evaluating university counseling services. Eval Program Plann 1983 Jan;6(3-4):315-327. [doi: 10.1016/0149-7189(83)90011-3] [Medline: 10267259]
42. Pedersen H, Skliarova T, Attkisson CC, Lara-Cabrera ML, Havnen A. Measuring patient satisfaction with four items: validity of the client satisfaction questionnaire 4 in an outpatient population. BMC Psychiatry 2023 Nov 07;23(1):808-808 [FREE Full text] [doi: 10.1186/s12888-023-05310-w] [Medline: 37936112]
43. Larsen DL, Attkisson C, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. Eval Program Plann 1979 Jan;2(3):197-207. [doi: 10.1016/0149-7189(79)90094-6] [Medline: 10245370]
44. Attkisson C, Zwick R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. Eval Program Plann 1982 Jan;5(3):233-237. [doi: 10.1016/0149-7189(82)90074-x] [Medline: 10259963]

XSL•FO
RenderX

45. Pedersen H, Havnen A, Lara-Cabrera ML. Measuring client satisfaction with digital services: validity and reliability of a short-form digital tool. iProc 2022 Aug 23;8(1):e41472. [doi: 10.2196/41472]
46. Kenter RMF, Schønning A, Inal Y. Internet-delivered self-help for adults with ADHD (MyADHD): usability study. JMIR Form Res 2022 Oct 21;6(10):e37137 [FREE Full text] [doi: 10.2196/37137] [Medline: 36269662]
47. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. International Journal of Human–Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
48. Skliarova T, Gråwe RW, Vaag J, Mandahl A, Kolltveit I, Løvås I, et al. Pilot randomised controlled trial on the feasibility and intervention satisfaction with an educational group programme for adults with attention deficit hyperactivity disorder and their caregivers. BMC Psychiatry 2025 Feb 20;25(1):161-112 [FREE Full text] [doi: 10.1186/s12888-025-06570-4] [Medline: 39979888]
49. Skliarova T, Lara-Cabrera ML, Hafstad H, Havnen A, Saether SG, Salvesen, et al. Feasibility, acceptability and preliminary evaluation of a user co-facilitated psychoeducational programme: a feasibility proof-of-concept randomised control trial. BMC Psychiatry 2024 Sep 16;24(1):615-613 [FREE Full text] [doi: 10.1186/s12888-024-06015-4] [Medline: 39285365]
50. Zwick R, Attkisson CC. Effectiveness of a client pretherapy orientation videotape. Journal of Counseling Psychology 1985;32(4):514-524. [doi: 10.1037//0022-0167.32.4.514]
51. Pedersen H, Havnen A, Brattmyr M, Attkisson CC, Lara-Cabrera ML. A digital Norwegian version of the client satisfaction questionnaire 8: factor validity and internal reliability in outpatient mental health care. BMC Psychiatry 2022 Oct 31;22(1):671-678 [FREE Full text] [doi: 10.1186/s12888-022-04281-8] [Medline: 36316661]
52. Nguyen TD, Attkisson C, Stegner BL. Assessment of patient satisfaction: development and refinement of a service evaluation questionnaire. Eval Program Plann 1983 Jan;6(3-4):299-313. [doi: 10.1016/0149-7189(83)90010-1] [Medline: 10267258]
53. Jones K, Daniels H, Heys S, Lacey A, Ford DV. Toward a risk-utility data governance framework for research using genomic and phenotypic data in safe havens: multifaceted review. J Med Internet Res 2020 May 15;22(5):e16346 [FREE Full text] [doi: 10.2196/16346] [Medline: 32412420]
54. Sittig DF, Singh H. COVID-19 and the need for a national health information technology infrastructure. JAMA 2020 Jun 16;323(23):2373-2374. [doi: 10.1001/jama.2020.7239] [Medline: 32421178]
55. Pedersen PB, Lilleeng SE. Distriktspsykiatriske tjenester (rapport IS-2825). Helsedirektoratet. 2019. URL: https://tinyurl.com/5y9hzzvz [accessed 2026-02-22]

Abbreviations

CSQ-4: 4-item Client Satisfaction Questionnaire
CSQ-Video: a modified version of the Client Satisfaction Questionnaire 4-item version
GRIPP2: Guidance for Reporting Involvement of Patients and the Public
PIN: personal identification number
SIKT: Kunnskapssektorens tjenesteleverandør (Norwegian Agency for Shared Services in Education and Research)
SUS: System Usability Scale

Edited by P Santana-Mancilla; submitted 04.Jul.2025; peer-reviewed by V P?rez Jover, R Marshall; comments to author 05.Dec.2025; revised version received 26.Jan.2026; accepted 28.Jan.2026; published 26.Feb.2026.

Please cite as:
Pedersen H, Skliarova T, Engvik LS, Mandahl A, De las Cuevas C, Havnen A, Lara-Cabrera ML
Co-Created Digital Pretherapy Psychoeducation for Outpatients in Specialized Mental Health Care: Usability Evaluation and Patient Satisfaction Study
JMIR Hum Factors 2026;13:e80130
URL: https://humanfactors.jmir.org/2026/1/e80130
doi:10.2196/80130
PMID:

©Henrik Pedersen, Tatiana Skliarova, Liv S Engvik, Arthur Mandahl, Carlos De las Cuevas, Audun Havnen, Mariela Loreto Lara-Cabrera. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 26.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Utilization of AI Among Medical Students and Development of AI Education Platforms in Medical Institutions: Cross-Sectional Study

Xiaokang Shi[1,2], BS; Zewu Jiang[1], MM; Li Xiong[1], BS; Ka-Chun Siu[3], PhD; Zhen Chen[1,2], MD, PhD

[1]Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Building 9, 3rd Floor, No. 160, Pujian Road, Pudong New Area, Shanghai, China

[2]School of Medicine, Tongji University, Shanghai, China

[3]Department of Health and Rehabilitation Sciences, College of Allied Health Professions, University of Nebraska Medical Center, Omaha, NE, United States

Corresponding Author:
Zhen Chen, MD, PhD
Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Building 9, 3rd Floor, No. 160, Pujian Road, Pudong New Area, Shanghai, China

Abstract

Background: The emergence of artificial intelligence (AI) is driving digital transformation and reshaping medical education in China. Numerous medical schools and institutions are actively implementing AI tools for case-based learning, literature analysis, and lecture support. This expanding application is accelerating the adoption of localized AI platforms, which are poised to become integral components in the coming years.

Objective: The primary aim of this study was to investigate the current use of AI tools among medical students, including usage frequency, commonly used platforms, and purposes of use. The second aim was to explore students' needs and expectations toward AI-powered medical education platforms by collecting and assessing student feedback, and to identify practical requirements across disciplines and academic stages to inform more effective platform design.

Methods: Based on the task-technology fit model and 5 hypotheses, an anonymous online questionnaire was conducted to assess AI usage in learning, gather student feedback on AI-powered medical education platforms, and evaluate expected functionalities. The survey was conducted from March 1 to May 31, 2025, using a convenience sampling method to recruit medical students from various disciplines across Shanghai, China. The sample size was determined at 422, accounting for a 10% rate of invalid responses. The questionnaire was developed and distributed online via Wenjuanxing and promoted through WeChat groups and in-person interviews. Data analysis was conducted employing IBM SPSS Statistics (v 27.0).

Results: A total of 428 valid questionnaires were collected. The average frequency of AI-assisted learning among medical students was 5.06 (SD 2.05) times per week. Over 90% (388/428) of the students used more than 2 AI tools in their daily tasks. Students from different disciplines, educational stages, and academic systems demonstrated different usage patterns and expectations for AI-powered medical education platforms.

Conclusions: AI technology is widely accepted by medical students and is extensively applied across various aspects of medical education. Significant differences are observed in usage patterns across disciplines, educational stages, and academic systems. Understanding the actual needs of students is crucial for the construction of AI-powered medical education platforms.

(JMIR Hum Factors 2026;13:e81652) doi:10.2196/81652

KEYWORDS

AI chatbots; artificial intelligence; cross-sectional study; medical education; medical schools; medical students; technology acceptance

Introduction

The rapid development of artificial intelligence (AI) has profoundly accelerated the digital transformation of medical education worldwide. AI demonstrates significant potential across multiple domains of medical education, including case-based teaching, literature analysis, and lecture support [1,2].

Globally, previous studies have documented the successful integration of large language models (LLMs) and conversational agents in medical education, significantly enhancing teaching effectiveness [3,4]. Gilson et al [5] reported that ChatGPT and other LLMs have been deployed to simulate clinical reasoning sessions, automate feedback on student essays, and generate customized practice questions, demonstrating measurable gains in diagnostic accuracy and pedagogical efficiency. Simultaneously, students' mastery of both theoretical knowledge and practical skills has markedly improved, resulting in better

XSL•FO
RenderX

learning outcomes and clinical performance. These advances address existing challenges in medical education and open up promising pathways for its future development [6,7].

In China, supportive national policies and rapid technological advances have jointly facilitated the localized application of AI in medical education [8]. The national "AI Plus" Implementation Guidelines outline the strategic direction for the deep integration of AI technologies with public welfare services [9]. Meanwhile, Shanghai's pioneering Medical Artificial Intelligence Work Plan specifically proposes establishing an "intelligent medical education and training" platform, emphasizing the development of smart training environments using generative AI technologies [10]. This provides clear policy support and practical guidance for higher education institutions to develop localized AI−powered medical education platforms [11,12].

With the widespread adoption of AI, students' learning methods, habits, and institutional teaching models have been rapidly reshaped. However, systematic data on medical students' current AI usage patterns and their practical needs across different educational stages and disciplines remain limited. A deeper understanding of these aspects is crucial for effectively guiding the development and optimization of future AI−powered medical education platforms. Therefore, this study aimed to collect medical students' current use of AI and their practical needs across different educational stages and academic disciplines. We also explored effective strategies for developing AI-powered medical education platforms, with the goal of providing recommendations to guide the development and optimization of future AI−powered medical education platforms.

Methods

Participants and Procedures

This study employed a convenience sampling method to conduct a cross-sectional survey among medical students from various medical universities in Shanghai, China, with data collected from March 1 to May 31, 2025.

The inclusion criteria were (1) current enrollment in a medical program and (2) provision of informed consent to participate. The exclusion criteria were (1) nonmedical students, (2) students not currently enrolled in any program, (3) students attending medical schools outside Shanghai, and (4) students who declined participation.

Theoretical Framework and Hypotheses

This study is based on the task-technology fit (TTF) model. Its core proposition is that technology's effectiveness is determined not by its attributes or user attitudes alone but by the fit between technological functionalities and user task requirements.

Based on the TTF framework and a review of relevant literature [13,14], this study proposes 5 hypotheses. These hypotheses aim to examine differences in usage status, functional needs, and expectations regarding AI-powered medical education platforms among medical students across different disciplines, educational stages, and program types while identifying factors with a significant influence.

- H1: Perceived task-technology fit has a significant positive impact on platform satisfaction.
- H2: Disciplinary background moderates the relationship between TTF and platform satisfaction.
- H3: Program type (full-time vs part-time) moderates the relationship between TTF and platform satisfaction.
- H4: Educational stage moderates the relationship between TTF and platform satisfaction.
- H5: The frequency of AI usage has a significant positive impact on platform satisfaction.

Instrument Pretesting and Validation

Guided by these 5 hypotheses, we designed a concise online questionnaire. Prior to the formal survey, a pilot test was conducted through in-person interviews with 23 postgraduate clinical medicine students from the same institution to evaluate content validity, item clarity, and internal consistency reliability. The expected platform functions scale (Cronbach α=0.825, items with zero variance were excluded) and the most frequently used AI platforms scale (Cronbach α=0.858, items with zero variance were excluded) showed high internal consistency.

Questionnaire Design

Based on the feedback from the pilot study, we made appropriate revisions to certain items and their phrasing in the questionnaire. Our questionnaire covered 3 sections with 13 items (see the questionnaire in Multimedia Appendix 1): the first section is general information, including age, gender, major, school, educational stage, and academic program type; the second section is the current use of AI tools, including the frequency of use (average per week), preferred platforms, and usage purposes (eg, theoretical learning, literature assistance, among others); and the third section is current status and expectations for AI-powered medical education platforms (whether the institution has developed an AI-powered medical education platform, satisfaction with the platform, and expected functions of the future platform) [13].

To ensure data completeness, all questionnaire items were set as mandatory, and participants were required to complete all the questions before submission. The survey platform would automatically record device type and completion time.

Ethical Considerations

This study received an ethics exemption (EX-2025 - 017) from the Medical Ethics Committee of Renji Hospital, Shanghai Jiao Tong University School of Medicine, as it utilizes anonymized data, operates under standard informed consent protocols, and involves no sensitive biological materials or procedures.

We obtained informed consent from all student participants before the survey, providing full details about the study's purpose, procedures, and privacy protections. No compensation was provided, as the study involved minimal burden and no anticipated harm. All the data were strictly protected to ensure confidentiality and prevent any risk of information leakage. To this end, access was restricted to authorized research team members, and the data were used solely for analysis and reporting within this study.

XSL•FO
RenderX

Sample Size

To ensure adequate statistical power and precision for the intended analyses, we used a standard sample size calculation formula [15]. Assuming a 95% CI, a margin of error of 0.05, and an expected population proportion of 0.5, the minimum required sample size was calculated to be 384. Drawing on previous studies [16,17] and to improve the generalizability of the results, we further accounted for a 10% invalid questionnaire rate, resulting in a final target sample size of 422.

$n=Z^2 \cdot P \cdot (1-P) E^2$ (95% CI, Z=1.96, E=0.05, P=.50)

Finite population correction was not applied due to the use of convenience sampling and structural heterogeneity across institutions, which precluded the definition of a single unified sampling frame.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics (version 27.0). Descriptive statistics were computed using appropriate measures for each variable type: continuous variables were summarized with means and SDs, while categorical variables were presented as frequencies and percentages.

For group comparisons involving categorical variables, chi-square tests were employed. Multiple response analyses were conducted using multiple response sets combined with chi-square tests, with Bonferroni correction applied to account for multiple comparisons. For ordinal data or continuous variables violating normality assumptions, non-parametric tests (Mann-Whitney *U* for 2-group comparisons and Kruskal-Wallis *H* for multigroup comparisons) were utilized, with post hoc analyses performed where appropriate.

ANOVA was used for comparing continuous variables across multiple groups, while MANOVA was employed for analyses involving multiple continuous dependent variables. Multivariable analyses included linear regression for continuous outcomes and logistic regression for binary outcomes.

All statistical tests used a 2-tailed significance threshold of $P<.05$, with appropriate corrections for multiple testing implemented where necessary [18].

Results

Participants' Characteristics

A total of 440 questionnaires were collected. After excluding 12 responses from nonmedical students, students at institutions outside Shanghai, and nonenrolled individuals, 428 valid questionnaires were retained, yielding an effective response rate of 97.3%. The questionnaires were returned with complete and valid responses, thus containing no missing data.

Participants were drawn from 7 medical schools in Shanghai, with 188 (43.92%) male participants and 240 (56.07%) female participants, and a median age of 22 (IQR 20.07-24.62) years. Among them, 223 (52.10%) were undergraduate students, 174 (40.65%) were master's students, and 31 (7.24%) were doctoral students. The sample covered 8 major disciplines: clinical medicine, basic medicine, rehabilitation therapy, nursing, public health and epidemiology, pharmacy, traditional Chinese medicine, and medical engineering.

Frequency of AI Use Among Medical Students

Most respondents reported regular use of AI tools in academic work and daily tasks, with a mean usage frequency of 5.06 (SD 2.05) times per week. The mean (SD) frequency among undergraduates, master's students, and doctoral students was 5.09 (1.97), 4.99 (2.11), and 5.19 (2.27) times per week, respectively, with no significant differences based on educational stage or gender (Table 1).

XSL•FO
RenderX

Table . Current usage of artificial intelligence (AI) in medical students' learning

Variables	Educational stage			Chi-square (*df*)	Cramer *V*	*P* value[a]	Gender		Chi-square (*df*)	Cramer *V*	*P* value[a]	Total (n=428), n (%)
	Undergraduate students (n=223), n (%)	Master's students (n=174), n (%)	Doctoral students (n=31), n (%)				Male (n=188), n (%)	Female (n=240), n (%)				
Average weekly use of AI for learning				7.97 (6)	0.10	.24			6.29 (3)	0.12	.10	
≤1	15 (6.73)	21 (12.07)	4 (12.90)				16 (8.51)	24 (10)				40 (9.35)
2 - 3	58 (26.01)	34 (19.54)	6 (19.35)				36 (19.15)	62 (25.83)				98 (22.90)
4 - 6	52 (23.32)	44 (25.29)	4 (12.90)				40 (21.28)	60 (25.00)				100 (23.36)
≥7	98 (43.95)	75 (43.10)	17 (54.84)				96 (51.06)	94 (39.17)				190 (44.39)
AI platform used by medical students												
DeepSeek	213 (95.52)	163 (93.68)	28 (90.32)	1.67 (2)	0.06	.43	176 (93.62)	228 (95.00)	0.38 (1)	0.03	.54	404 (94.39)
Doubao	138 (61.88)	99 (56.90)	13 (41.94)	4.74 (2)	0.11	.09	97 (51.60)	153 (63.75)	6.41 (1)	0.12	.01	250 (58.41)
Kimi	127 (56.95)	108 (62.07)	13 (41.94)	4.57 (2)	0.10	.10	98 (52.13)	150 (62.50)	4.66 (1)	0.10	.03	248 (57.94)
ChatGPT	156 (69.96)	103 (59.20)	21 (67.74)	5.08 (2)	0.11	.08	130 (69.15)	150 (62.50)	2.06 (1)	0.07	.15	280 (65.42)
ChatGLM	28 (12.56)	14 (8.05)	0 (0)	5.88 (2)	0.12	.05	17 (9.04)	25 (10.42)	0.23 (1)	0.02	.64	42 (9.81)
Claude	24 (10.76)	9 (5.17)	3 (9.68)	4.03 (2)	0.10	.13	22 (11.70)	14 (5.83)	4.71 (1)	0.11	.03	36 (8.41)
Gemini	29 (13.00)	13 (7.47)	0 (0)	7.02 (2)	0.13	.03	27 (14.36)	15 (6.25)	7.84 (1)	0.14	.005	42 (9.81)
Other[b]	22 (9.87)	11 (6.32)	1 (3.23)	2.70 (2)	0.08	.26	16 (8.51)	18 (7.50)	0.15 (1)	0.02	.70	34 (7.94)
Number of AI platforms used by medical students				17.36 (14)	0.14	.24			3.86 (7)	0.10	.80	
2	53 (23.77)	45 (25.86)	8 (25.81)				46 (24.47)	60 (25.00)				106 (24.77)
≥2	210 (94.17)	153 (87.93)	25 (80.65)				166 (88.30)	222 (92.50)				388 (90.65)
≥3	157 (70.40)	108 (62.07)	17 (54.83)				120 (63.83)	162 (67.50)				282 (65.89)

[a]Chi-square test and Bonferroni correction were applied for multiple comparisons (α=.00625).

[b]AI platforms, including Qwen Chat, ERNIE Bot, Tencent Yuanbao, Poe, and Grok, were grouped together for analysis due to their relatively small sample sizes.

AI Platforms Used by Medical Students

In terms of AI platform selection, most medical students favor mainstream tools such as DeepSeek (n=404, 94.39%), Doubao (n=250, 58.41%), and ChatGPT (n=280, 65.42%) for their daily study and work. Survey results further indicated that the current use of multiple AI platforms has become a common practice among medical students, with over 90% (388/428) of the students reporting the use of 2 or more platforms, and over 60% (282/428) of the students reported using 3 or more platforms (Table 1).

Gender-based differences were observed in the adoption of specific AI platforms, with selected variations reaching statistical significance (χ^2_1=7.84; *P*=.005). Nonetheless, the overall variety of the platforms utilized remained consistent

XSL•FO
RenderX

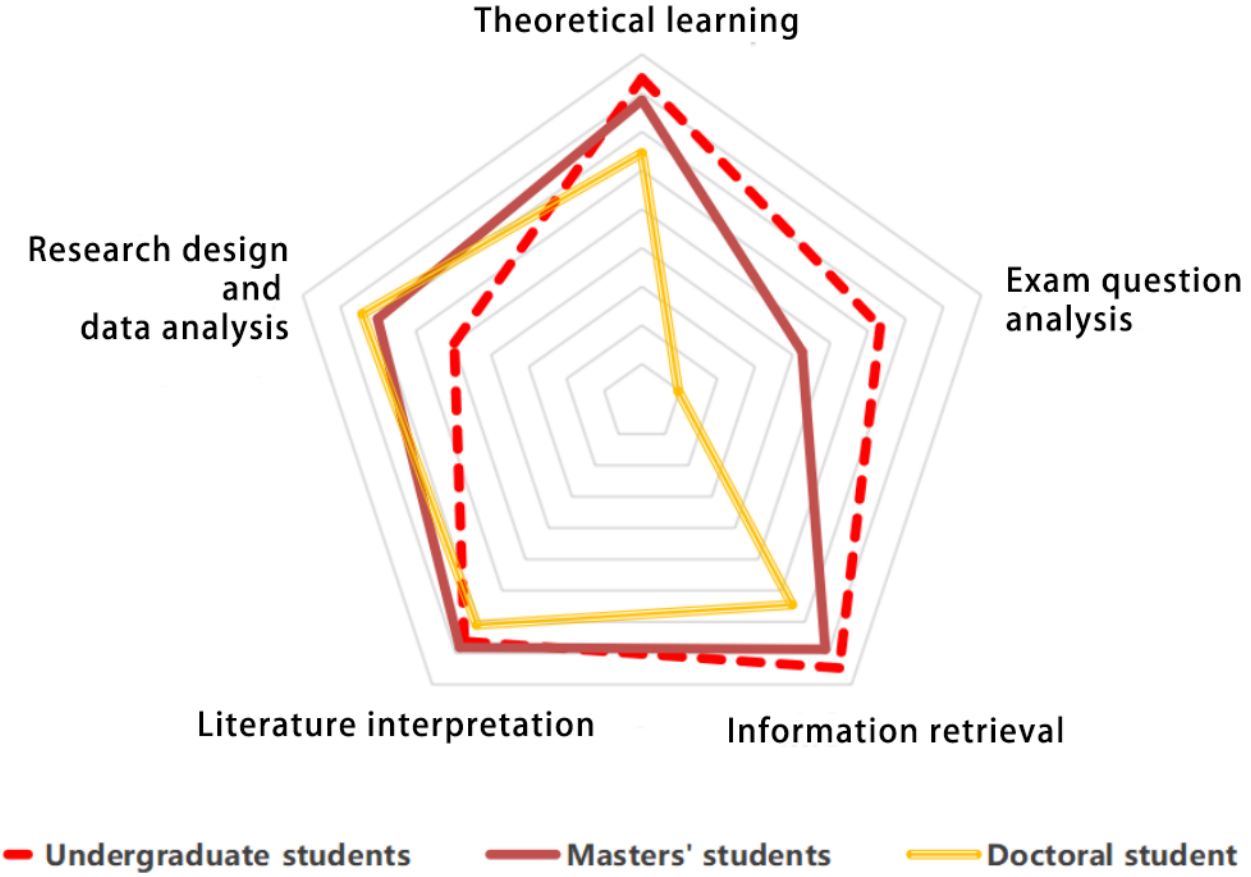
across genders, as evidenced by a comparable number of the tools used (χ^2_7=3.86; *P*=.80).

AI Usage Patterns by Academic Program Type and Educational Stage

The application of AI in medical education spans multiple areas, including theoretical learning, question analysis, literature translation, and scientific research. Based on the literature review and preliminary survey feedback, our questionnaire categorized the application areas into 5 domains: "theoretical learning," "exam question analysis," "information retrieval," "literature interpretation," and "research design and data analysis."

Regarding educational stage, undergraduate students primarily used AI for exam preparation, while graduate students focused more on research tasks such as study design and data analysis (Figure 1).

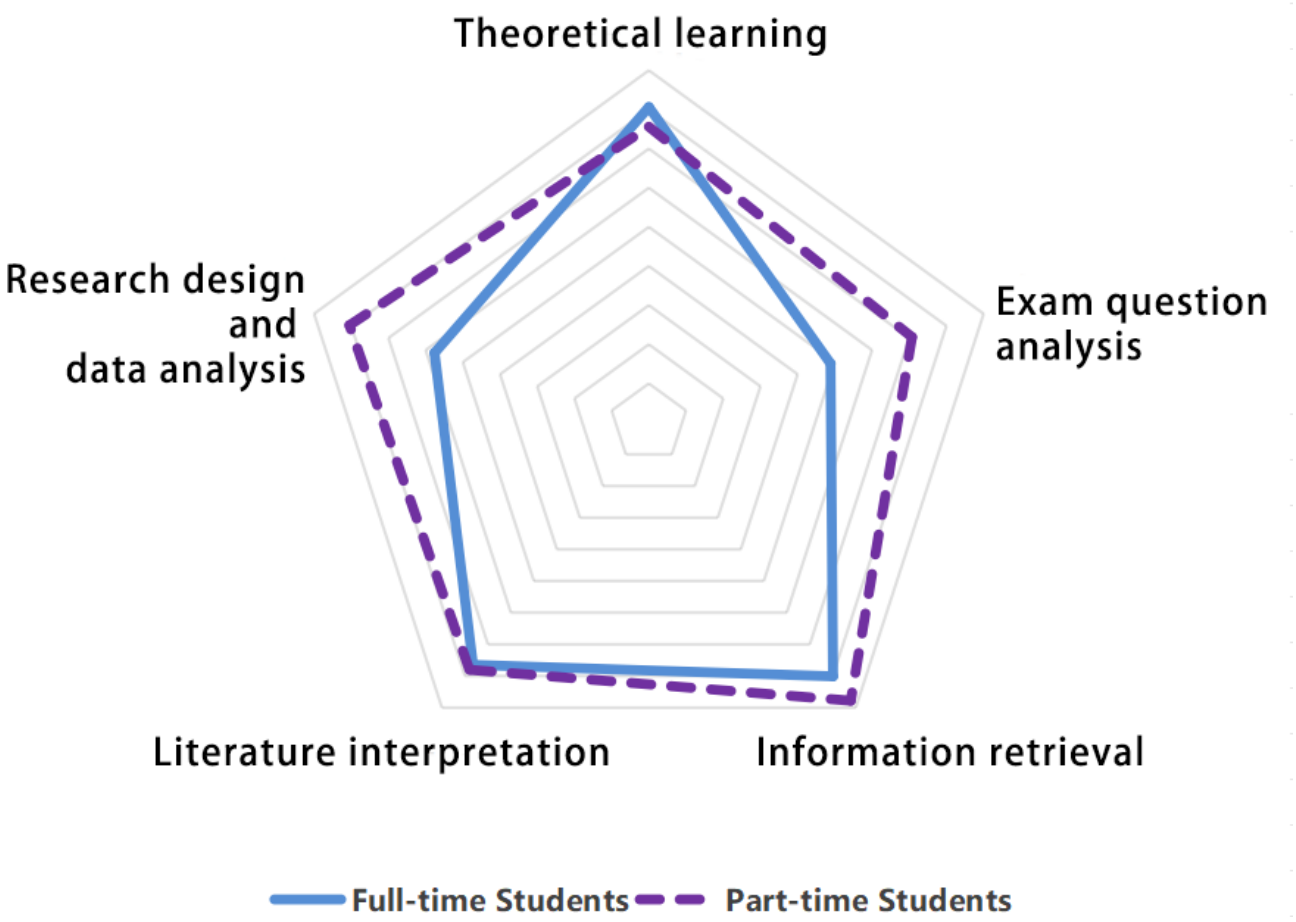
Figure 1. Usage of artificial intelligence (AI) tools in learning among medical students from different educational stages.



When grouped by academic program type, part-time students demonstrated stronger needs for AI support in practical research and exam question analysis. This preference pattern appears related to their need to balance studies with work commitments, requiring efficient learning solutions that yield immediate academic and professional applicability (Figure 2).

XSL•FO
RenderX

Figure 2. Usage of artificial intelligence (AI) tools in learning among medical students from different academic program types (full-time vs part-time).



AI Usage Patterns by Disciplines

Medical students from different academic disciplines exhibited distinct priorities in their application of AI-powered learning (Table 2). Initial analysis using a multiple response test revealed a significant overall difference in AI usage patterns across disciplines (χ^2_{30}=53.62; *P*=.005), prompting subsequent pairwise comparisons between disciplines with a Bonferroni-adjusted alpha of 0.01.

XSL•FO
RenderX

Table . Comparative analysis of artificial intelligence (AI) function usage across different medical disciplines.

Domain	Clinical medicine (n=245), n (%)	Nursing (n=65), n (%)	Rehabilitation therapy (n=31), n (%)	Basic medicine (n=31), n (%)	Public health and epidemiology and pharmacy[a] (n=11), n (%)	Traditional Chinese medicine (n=22), n (%)	Medical engineering (n=23), n (%)	Chi-square (*df*)	Cramer *V*	*P* value[b]
Theoretical learning	201 (82.04)	50 (76.92)	24 (77.42)	26 (83.87)	8 (72.73)	16 (72.73)	18 (78.26)	2.58 (6)	0.08	.86
Exam question analysis	139 (56.73)	37 (56.92)	12 (38.71)	14 (45.16)	4 (36.36)	7 (31.82)	5 (21.74)	18.49 (6)	0.21	.005
Information retrieval	202 (82.45)	56 (86.15)	19 (61.29)	28 (90.32)	8 (72.73)	15 (68.18)	18 (78.26)	13.89 (6)	0.18	.03
Literature interpretation	184 (75.10)	55 (84.62)	22 (70.97)	26 (83.87)	10 (90.91)	12 (54.55)	19 (82.61)	11.80 (6)	0.17	.07
Research design and data analysis	144 (58.78)	35 (53.85)	17 (54.84)	23 (74.19)	9 (81.82)	15 (68.18)	13 (56.52)	7.02 (6)	0.13	.32

[a]Disciplines including public health and epidemiology and pharmacy were grouped together for analysis due to their relatively small sample sizes.

[b]Chi-squared test and Bonferroni correction were applied for multiple comparisons (α=.01).

Across all disciplines, the usage demand for theoretical learning was similarly high, with no statistically significant differences identified.

Regarding exam question analysis, students in clinical medicine demonstrated significantly higher usage demand than those in medical engineering (χ^2_1=10.36; *P*=.001).

In terms of information retrieval, students from rehabilitation therapy demonstrated significantly lower usage demand compared to students in clinical medicine (χ^2_1=7.72; *P*=.005), nursing (χ^2_1=7.59; *P*=.006), and basic medicine (χ^2_1=7.12; *P*=.008).

For literature interpretation, traditional Chinese medicine students demonstrated a significantly lower level of usage demand relative to students in nursing (χ^2_1=8.40; *P*=.004).

Across disciplines, usage demand for research design and data analysis was moderate, peaking non-significantly among public health and epidemiology and pharmacy students.

Correlates of Satisfaction With the Institutional AI Platform

With the rapid development of AI, many universities have launched localized AI platforms. Our survey investigated the availability of institution-specific AI−powered medical education platforms among medical students. Furthermore, we assessed student satisfaction with these platforms.

Approximately one-fifth (86/428) of the respondents reported that their institutions had developed such platforms. Satisfaction scores among these users exhibited significant variation. The average satisfaction score among the 86 users was 72.23 (SD 21.84), distributed as 40 (46.51%) satisfied, 28 (32.56%) neutral, and 18 (20.93%) dissatisfied.

Nonparametric tests revealed a significant difference in satisfaction levels by gender (*U*=686.50; *z*=−2.06; *P*=.04). No significant associations were found with academic program type, educational stage, or discipline. A subsequent multivariable regression that included these variables and usage frequency identified no significant predictors. To assess the model's reliability, collinearity diagnostics were performed, and they revealed no substantial multicollinearity (Table 3).

XSL•FO
RenderX

Table . Medical students' satisfaction with the artificial intelligence (AI)−powered medical education platforms at their institutions.

Predictor	Unstandardized coefficient		Standardized coefficient	*P* value	VIF[a]	95% CI for unstandardized coefficients (*B*)
	B	SE	β			
Constant	71.97	13.93	—[b]	<.001	—	44.22 to 99.72
Academic program type (reference: part-time)						
Full-time	−3.18	9.72	−0.04	.74	1.20	−22.55 to 16.19
Gender (reference: female)						
Male	8.52	5.12	0.20	.10	1.11	−1.68 to 18.72
Frequency	−0.39	1.34	−0.03	.77	1.08	−3.06 to 2.28
Discipline (reference: clinical medicine)						
Nursing	5.59	6.85	0.10	.42	1.14	−8.05 to 19.24
Rehabilitation therapy	−4.14	13.76	−0.04	.76	1.08	−31.55 to 23.26
Basic medicine	−0.36	7.38	−0.01	.96	1.11	−15.07 to 14.35
Public health and epidemiology and pharmacy	10.54	12.64	0.10	.41	1.20	−14.64 to 35.71
Medical engineering	−6.18	24.99	−0.03	.81	1.22	−55.96 to 43.61
Educational stage (reference: undergraduate student)						
Master's student	−2.43	5.97	−0.05	.69	1.33	−14.32 to 9.46
Doctoral student	8.59	10.83	0.10	.43	1.29	−12.99 to 30.18

[a]VIF: variance inflation factor.

[b]Not applicable.

The lack of significant predictors for satisfaction should be interpreted with caution. This result may reflect the inherently subjective and multifaceted nature of satisfaction, which can be influenced by factors beyond the scope of this study. Future research employing longitudinal or mixed methods designs is needed to unravel the complex drivers of user satisfaction.

Correlates of Expected Functions for the Institutional AI Platform

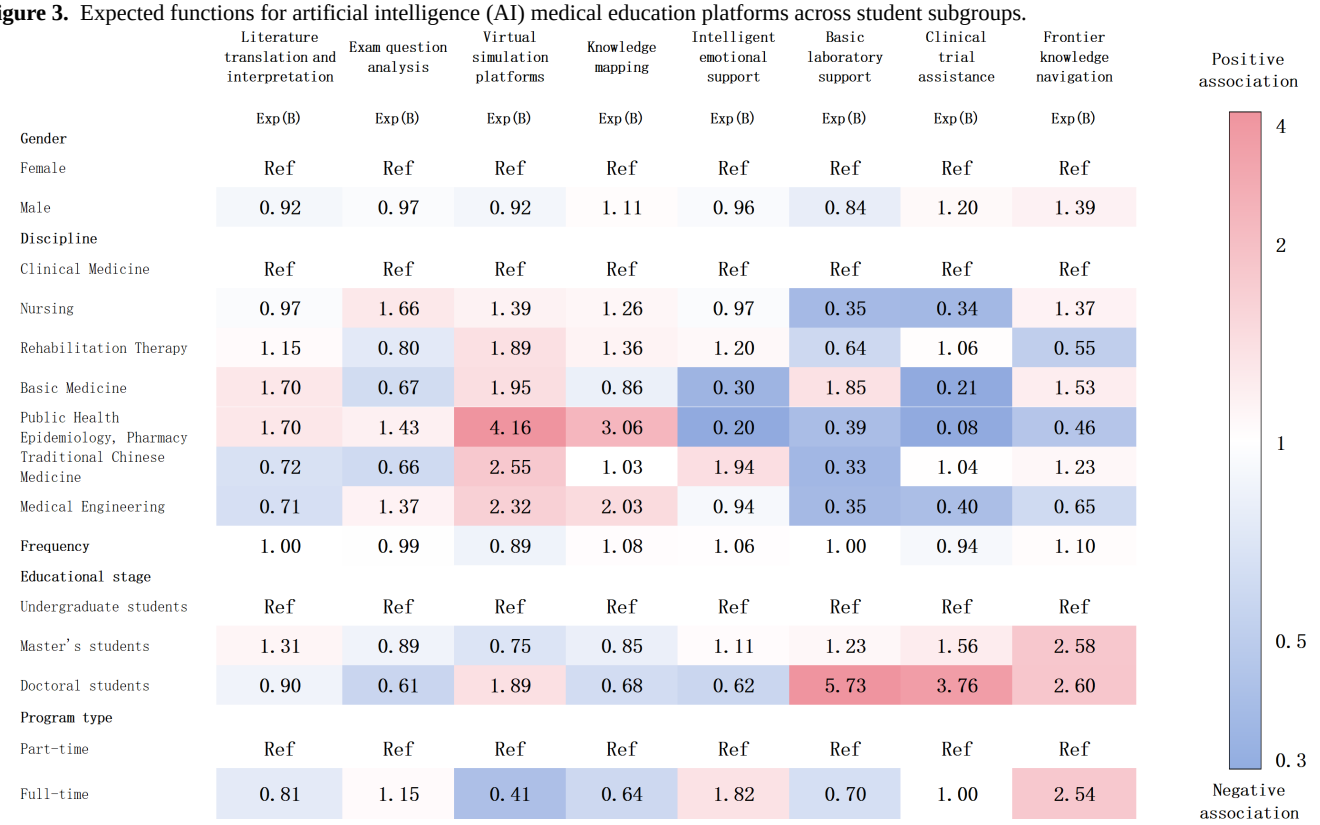
The majority (342/428) of the students reported that their institutions had not yet launched an AI platform specifically centered on medical education. To investigate the needs of this group, this survey further explored their practical needs and expectations for upcoming platforms.

Drawing on the current practical applications of AI technology and feedback from medical students, along with a review of relevant literature, our questionnaire categorized the expected functions of AI-powered medical education platforms into 8 aspects: literature translation and interpretation, exam question analysis, clinical trial assistance, basic laboratory support, knowledge mapping, virtual simulation platforms, frontier knowledge navigation, and intelligent emotional support.

Nonparametric testing revealed significant subgroup differences in the demand for specific platform functions. Guided by these initial findings, we advanced the analysis using binary logistic regression within a more rigorous multivariable framework. For comparing the relative effects of multiple predictors, the results are presented as odds ratios (Exp(B)) and visualized in the accompanying heatmap (Figure 3).

XSL•FO
RenderX

Figure 3. Expected functions for artificial intelligence (AI) medical education platforms across student subgroups.



Regarding gender, the regression analysis revealed a consistent pattern: no statistically significant associations were found between gender and any of the anticipated platform functions.

Analysis of academic program type (full-time vs part-time) revealed no significant differences in functional expectations. However, full-time students demonstrated a numerically stronger preference for frontier knowledge navigation (B=0.93; Exp(B)=2.54; P=.14), while they were less inclined toward Virtual Simulation Platforms (B=−0.90; Exp(B)=0.41; P=.08).

Regarding the educational stage, doctoral students demonstrated a significantly stronger preference for AI-assisted clinical trial support (B=1.32; Exp(B)=3.76; P=.01) and basic laboratory assistance (B=1.75; Exp(B)=5.73; P=.001) compared to undergraduates, while master's students showed significantly higher demand for frontier knowledge navigation (B=0.95; Exp(B)=2.58; P=.02).

Across different academic disciplines, medical students showed varying expectations for the functions of upcoming AI-powered medical education platforms. Students in clinical medicine demonstrated a significantly stronger demand for clinical trial assistance (B=−1.09; Exp(B)=0.34; P=.008) and Basic Laboratory Support (B=−1.05; Exp(B)=0.35; P=.005) compared to those in nursing. Conversely, nursing students were inclined toward personalized guidance on theoretical exam preparation, though this tendency did not reach statistical significance (B=0.51; Exp(B)=1.66; P=.08). Notably, students in clinical medicine demonstrated a significantly higher demand for clinical trial assistance compared to students in other disciplines. This difference reached statistical significance when compared specifically with students in basic medicine (B=−1.58; Exp(B)=0.21; P=.01) and those in public health and epidemiology and pharmacy (B=−2.54; Exp(B)=0.08; P=.02).

Test of the TTF Hypotheses

Based on the results of our prior data analysis and guided by the 5 research hypotheses derived from the TTF theory, we employed appropriate statistical methods to test these hypotheses (Table 4).

XSL•FO
RenderX

Table . Hypothesis testing results of the task-technology fit (TTF) model.

Hypothesis	Path relationship	Statistic[a]	Effect size[b]	*P* value	Supported
H1	TTF (current use) → satisfaction	$F(5, 80)=1.11$	Adjusted $R^2=0.006$	.36	No
H2	Discipline → TTF (current use)	$F(36, 2526)=1.60$	Partial $\eta^2=0.02$	.01	Yes
	Discipline → TTF (expected function)	$F(42, 840)=1.25$	Partial $\eta^2=0.06$	.13	No
H3	Academic program type → TTF (current use)	$F(6, 421)=3.32$	Partial $\eta^2=0.05$	.003	Yes
	Academic program type → TTF (expected function)	$F(7, 140)=1.22$	Partial $\eta^2=0.06$	.29	No
H4	Educational stage → TTF (current use)	$F(12, 842)=6.51$	Partial $\eta^2=0.09$	<.001	Yes
	Educational stage → TTF (expected function)	$F(14, 280)=1.53$	Partial $\eta^2=0.07$	.10	No
H5	Frequency → satisfaction	$B=-0.39$	95% CI for $B=-3.06$ to 2.28	.77	No

[a]Analytical methods were hypothesis-specific: ANOVA for a single outcome variable (H1); multivariate ANOVA for multiple outcome variables (H2-H4); linear regression with unstandardized coefficients for predictive modeling (H5).

[b]Effect size measures were selected and reported in accordance with the conventions for the specific statistical procedures applied.

The analytical approach was tailored to the characteristics of the variables under examination. For H1, ANOVA was used, which indicated no significant association between current AI usage patterns and satisfaction. For H2, H3, and H4, MANOVA was applied: the results showed that discipline, academic program type, and educational stage each had a significant influence on TTF (current use); however, none of these factors demonstrated a significant effect on TTF (expected function). For H5, linear regression analysis found no significant relationship between frequency and satisfaction.

Discussion

While previous studies have described broad trends in AI adoption among medical students, this study identifies the specific drivers of heterogeneity within a defined cohort from Shanghai. Through a granular subgroup analysis, we demonstrate how disciplinary background, educational stage, and program type significantly shape distinct patterns in usage frequency, functional preferences, and perceived value.

Popularity of AI Technology Among Medical Students

The rapid advancement of AI technology is demonstrating the potential to reshape traditional paradigms in medical education. For medical students, the integration of digital and intelligent technologies has significantly enhanced instructional quality and learning outcomes. With ongoing technological advancement, AI is now widely adopted and has become an essential part of students' academic work [19].

The survey revealed that medical students now engage with AI tools on a frequent basis, reflecting a notable shift in usage patterns compared to earlier adoption phases [20,21]. This suggests that, with ongoing technological advancement and the growing accessibility of AI, students are increasingly adopting this new technology and using it more in their learning.

In terms of LLM selection, most students preferred mainstream models, encompassing both domestic platforms such as DeepSeek and Doubao and international ones like ChatGPT. The choice appears to have been influenced by factors such as accessibility and performance. While mainstream LLMs meet most students' academic requirements, many still select specific models based on personal preferences and practical needs. This reflects a growing trend toward individualized AI tool usage.

Regarding mainstream LLM selection, international findings show that ChatGPT is widely popular globally [22,23]; however, its adoption among the surveyed medical students in Shanghai remains lower compared to DeepSeek. This observed usage pattern aligns with DeepSeek's established localization strategy and regulatory compliance within the Chinese environment. China's regulatory policies require GenAI services to complete local filing and security assessments [24]. ChatGPT, lacking such compliance, is inaccessible through conventional channels. Furthermore, China's exclusion from OpenAI's supported countries creates additional access barriers [25]. In contrast, DeepSeek operates in full compliance with these requirements, ensuring seamless accessibility for Chinese users and thus gaining a competitive edge in the local AI market.

The analysis of AI tool preferences revealed distinct gender−based patterns. Despite using a similar number of AI platforms, with mainstream tools being central for both groups, male students demonstrated greater enthusiasm for emerging options, whereas female students adopted a more cautious approach. These observed differences align with previous

XSL•FO
RenderX

findings on gendered perceptions of AI technology [26-28]. Notably, this variation further underscores the importance of developing flexible, multiplatform strategies to accommodate diverse user preferences in medical education.

The integration of AI technology has become a defining feature of contemporary medical education. Our findings reveal a clear consensus among students across all academic backgrounds on adopting a multiplatform approach. Regardless of educational stage or discipline, medical students are actively leveraging diverse AI tools, flexibly selecting platforms according to specific learning scenarios and practical needs [29].

Group Differences in AI Tool Usage for Academic Purposes

Medical knowledge covers a broad spectrum and involves extensive interdisciplinary integration. Although various medical specialties are inherently connected, they differ significantly in teaching method, clinical practice, and scientific research. These distinctions are also reflected in the application of AI.

From an educational stage perspective, undergraduate students tended to emphasize AI-assisted analysis of theoretical exam questions more than master's and doctoral students, who placed greater focus on AI's role in supporting practical research. This difference reflects the distinct teaching priorities at each educational stage: undergraduate students face a heavy burden of theoretical courses and professional qualification exams, so their learning focus tends to be on theoretical exam question analysis and knowledge mapping. In contrast, graduate students, facing less exam pressure, focus more on research-related tasks and thus place greater emphasis on AI as a tool for practical research support [30-32].

From the perspective of academic program type (full-time vs part-time), part-time students had broader and more comprehensive needs for AI-assisted learning compared to full-time students. Their demand for AI in scientific research and test analysis was more pronounced. This stems directly from the dual pressures faced by part-time medical students, who balance both academic studies and daily work. In dealing with research and exams, they tend to be more reliant on AI for assistance and guidance to alleviate the burden of learning [33].

From a disciplinary perspective, different majors have a significant impact on the direction of AI-assisted learning [34]. The data showed that students majoring in clinical medicine and nursing were more focused on AI-assisted exam question analysis [35], while traditional Chinese medicine students had a notably lower demand for literature translation and interpretation. Students in basic medicine, public health and epidemiology, and pharmacy were more concerned with AI's role in supporting research work. These differences directly reflect the distinct curricular focus, professional requirements, and learning objectives characterizing each discipline [7,20,36].

Student-Centered Design of Platform Functions

At present, the application of AI in medical education is still in an exploratory stage. Major computer and internet companies, both domestically and internationally, have launched AI models with diverse functionalities. In parallel, universities and research institutions have adopted various strategies to expand the potential of this emerging field. For example, the AI application platform at Tongji University integrates multiple large models such as DeepSeek, OpenAI, and Tongyi Qianwen, offering a wide range of AI tools. Faculty and students can select these tools based on their professional and academic requirements [37]. Similarly, Shanghai Jiao Tong University has launched the "Jiao Xiao Zhi" agent management platform, integrating LLMs including DeepSeek for localized AI deployment. The platform enables faculty and students to create customized AI assistants, streamlining teaching and research tasks through its secure campus−based infrastructure [38].

This survey focused on medical students' satisfaction with and expectations for AI-powered medical education platforms. While overall satisfaction was relatively high, a significant minority (approximately one-quarter) expressed dissatisfaction, with notable variations across academic stages and disciplines.

To address these divergent needs, we propose a tiered design framework for future platforms:

For undergraduate students: Platforms should prioritize structured learning support. This includes AI-powered tools for adaptive test preparation aligned with standardized exams, interactive virtual patient cases for foundational clinical exposure, and personalized review systems that target individual knowledge gaps.

For postgraduate students (master's or PhD): The focus should shift to research and specialized skill development. Key features should encompass advanced literature interpretation aids, data analysis modules for processing experimental or clinical data, and AI assistants for research design and grant writing, catering to their deep engagement in academic research.

Discipline-specific customization: Further refinement should distinguish between clinical and basic medicine disciplines. Clinically oriented tracks would benefit from advanced diagnostic simulators and patient management tools, while basic medicine tracks require robust support for experimental design, genomic data analysis, and scientific visualization.

By moving beyond a one-size-fits-all model to adopt such a stratified and discipline-aware approach, AI platforms can achieve deeper integration into medical education, ultimately enhancing both student satisfaction and educational outcomes [9,39].

Limitations

Due to constraints in personnel and resources, this study employed a convenience sampling method among medical students in Shanghai. This nonprobability sampling approach may have introduced selection bias, such as an overrepresentation of students already enthusiastic about AI, and limits the generalizability of the findings beyond similar urban, well-resourced contexts.

Moreover, this study relied on self-reported measures for key metrics—including AI usage frequency and satisfaction—for which formal psychometric validation was not conducted. This approach increased susceptibility to recall and social desirability biases across all these measures. Furthermore, as subjective

XSL•FO
RenderX

constructs, they were vulnerable to varying personal interpretations and benchmarks. This was particularly relevant for abstract constructs like satisfaction. Consequently, these subjective data may not fully capture the nuanced realities of actual user behavior.

Future research could therefore prioritize multicenter, large-sample designs that utilize standardized instruments. Such approaches would provide more robust and generalizable evidence to effectively guide the development of AI-powered medical education platforms.

Conclusions

This study conducted a cross-sectional survey across major medical institutions in Shanghai, China, collecting and analyzing the current use and practical needs of AI among medical students from different educational stages, academic program types, and disciplines. Our findings clearly indicate that AI is widely applied in medical education and has become a common tool for student learning. Students from different disciplines, educational stages, genders, and academic program types show significant differences in their functional demands for AI-assisted learning.

Furthermore, our study investigated the current status of AI-powered medical education platforms and explored students' expectations for such platforms. It is evident that the rapid integration of AI in medical education holds great promise, and our findings provide evidence-based support to guide the future development of AI-powered medical education platforms.

Acknowledgments

We thank all the medical students who participated in this survey for taking time to complete the questionnaires and providing valuable data for this study. We also appreciate those who assisted in questionnaire distribution and collection, which helped improve the response rate and data quality of the research.

We would like to extend special thanks to the Clinical Research Center of Renji Hospital for their guidance in study design, methodology, and ethics, as well as their ongoing technical and logistical support throughout the study.

GenAI was used for language polishing and reference formatting during manuscript preparation. The authors reviewed and edited all content and take the full responsibility for the work.

Funding

This work was funded by National Key Research and Development Program of China (2024YFF0618400), Pudong New Area Health Commission (PWR12024-08), and Yuanshen Rehabilitation Institute Afflicted to Shanghai Jiao Tong University School of Medicine (yskf3-23-1107-2; yskf2-23-1107-2).

Data Availability

The datasets generated and analyzed during this study are not publicly available due to privacy restrictions in the informed consent agreements. However, anonymized data and related documentation are available from the corresponding author upon reasonable request for academic use, subject to a data use agreement and required citation. Requests should include a brief research proposal and will be reviewed within 4 - 6 weeks.

Authors' Contributions

Conceptualization: XS (lead), ZJ (equal)
Formal analysis: XS (lead)
Investigation: XS (lead), LX (supporting)
Methodology: ZJ (lead), LX (supporting), KCS (supporting)
Project administration: ZC (lead)
Supervision: ZC (lead)
Writing – original draft: XS (lead), KCS (supporting)
Writing – review & editing: ZC (lead), XS (supporting), ZJ (supporting), LX (supporting), KCS (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed items of our questionnaire.
[DOCX File, 17 KB - humanfactors_v13i1e81652_app1.docx]

Checklist 1
CHERRIES checklist.

XSL•FO
RenderX

[PDF File, 44 KB - humanfactors_v13i1e81652_app2.pdf]

References

1. Moro C, Štromberga Z, Raikos A, Stirling A. The effectiveness of virtual and augmented reality in health sciences and medical anatomy. Anat Sci Educ 2017 Nov;10(6):549-559. [doi: 10.1002/ase.1696] [Medline: 28419750]
2. Boillat T, Prithishkumar IJ, Suresh D, Naidoo N. Integrating virtual reality to enhance remote teaching of anatomy during unprecedented times. Anat Cell Biol 2025 Mar 31;58(1):112-121. [doi: 10.5115/acb.24.197] [Medline: 39623255]
3. Lee H. The rise of ChatGPT: exploring its potential in medical education. Anat Sci Educ 2024 Dec;17(9):1779. [doi: 10.1002/ase.2496] [Medline: 39087851]
4. Wu Z, Liu Z, Zhang Z. Research on the current situation and future governance of artificial intelligence in the field of medical education. Chin Med Ethics 2024;37(9):1093-1100. [doi: 10.12026/j.issn.1001-8565.2024.09.12]
5. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2024 Feb 27;10:e57594. [doi: 10.2196/57594] [Medline: 38412478]
6. Wang J, Luo L, Zhao C, Zeng X, Tang R, He S. Practice research of VR-based immersion programs in the clinical emergency. Smart 2022;8(10):176-178. [doi: 10.19335/j.cnki.2096-1219.2022.10.053]
7. Cherrez-Ojeda I, Gallardo-Bastidas JC, Robles-Velasco K, et al. Understanding health care students' perceptions, beliefs, and attitudes toward AI-powered language models: cross-sectional study. JMIR Med Educ 2024 Aug 13;10:e51757. [doi: 10.2196/51757] [Medline: 39137029]
8. Zou L, Lin H, Sun L. Policy text research on AI-enabled medical education. Chin Med Education Technol 2025;39(1):23-28. [doi: 10.13566/j.cnki.cmet.cn61-1317/g4.202501005]
9. Opinions of the state council on deepening the implementation of the "artificial intelligence+" initiative. State Council of the People's Republic of China. 2025. URL: https://cset.georgetown.edu/publication/china-ai-plus-opinions-2025/ [accessed 2025-12-17]
10. Shanghai medical AI development work plan (2025-2027). : General Office of the Shanghai Municipal People's Government; 2024 URL: https://www.shanghai.gov.cn/nw12344/20241224/0f60c01551784720899dbb911e9d5f08.html [accessed 2025-12-12]
11. Zhang J, Li C, Lyu W. The application prospects and challenges of artificial intelligence in medical education, medical research, and clinical practice. Chin J Gen Pract 2024;22(7):1085-1089. [doi: 10.16766/j.cnki.issn.1674-4152.003572]
12. Xiu J, Li Y, Wang Z, et al. Exploration on smart medical education based on 5G technology: taking the construction and application of "5G+Smart Education" in Guizhou Medical University as an example. J Guizhou Med Univ 2024;49(12):1864-1872. [doi: 10.19367/j.cnki.2096-8388.2024.12.021]
13. Tortella F, Palese A, Turolla A, et al. Knowledge and use, perceptions of benefits and limitations of artificial intelligence chatbots among Italian physiotherapy students: a cross-sectional national study. BMC Med Educ 2025 Apr 18;25(1):572. [doi: 10.1186/s12909-025-07176-w] [Medline: 40251635]
14. Tao W, Yang J, Qu X. Utilization of, perceptions on, and intention to use AI chatbots among medical students in China: national cross-sectional study. JMIR Med Educ 2024 Oct 28;10:e57132. [doi: 10.2196/57132] [Medline: 39466038]
15. Lwanga SK, Lemeshow S. Sample size determination in health studies: a practical manual. : World Health Organization; 1991 URL: https://iris.who.int/items/9c2e5da4-3785-4fec-9dbc-841e4ae0d98c [accessed 2025-12-12]
16. Ziapour A, Darabi F, Janjani P, Amani MA, Yıldırım M, Motevaseli S. Factors affecting medical artificial intelligence (AI) readiness among medical students: taking stock and looking forward. BMC Med Educ 2025 Feb 18;25(1):264. [doi: 10.1186/s12909-025-06852-1] [Medline: 39966878]
17. Sami A, Tanveer F, Sajwani K, et al. Medical students' attitudes toward AI in education: perception, effectiveness, and its credibility. BMC Med Educ 2025 Jan 17;25(1):82. [doi: 10.1186/s12909-025-06704-y] [Medline: 39833834]
18. Weidener L, Fischer M. Artificial intelligence in medicine: cross-sectional study among medical students on application, education, and ethical aspects. JMIR Med Educ 2024 Jan 5;10:e51247. [doi: 10.2196/51247] [Medline: 38180787]
19. Ganjavi C, Eppler M, O'Brien D, et al. ChatGPT and large language models (LLMs) awareness and use. A prospective cross-sectional survey of U.S. medical students. PLOS Digit Health 2024 Sep;3(9):e0000596. [doi: 10.1371/journal.pdig.0000596] [Medline: 39236008]
20. Alharbi MK, Syed W, Innab A, Basil A Al-Rawi M, Alsadoun A, Bashatah A. Healthcare students attitudes opinions perceptions and perceived obstacles regarding ChatGPT in Saudi Arabia: a survey-based cross-sectional study. Sci Rep 2024 Oct 1;14(1):22800. [doi: 10.1038/s41598-024-73359-y] [Medline: 39354005]
21. Pan G, Ni J. A cross sectional investigation of ChatGPT-like large language models application among medical students in China. BMC Med Educ 2024 Aug 23;24(1):908. [doi: 10.1186/s12909-024-05871-8] [Medline: 39180023]
22. Ajalo E, Mukunya D, Nantale R, et al. Widespread use of ChatGPT and other artificial intelligence tools among medical students in Uganda: a cross-sectional study. PLoS ONE 2025;20(1):e0313776. [doi: 10.1371/journal.pone.0313776] [Medline: 39787055]
23. von Garrel J, Mayer J. Artificial intelligence in studies—use of ChatGPT and AI-based tools among students in Germany. Humanit Soc Sci Commun 2023;10(1):799. [doi: 10.1057/s41599-023-02304-7]

XSL•FO
RenderX

24. Interim measures for the management of generative artificial intelligence services. : China Cyberspace Administration; 2023 URL: https://openresearch-repository.anu.edu.au/server/api/core/bitstreams/7577e6a0-db26-4dab-9273-8d13a6883a40/content [accessed 2025-12-12]
25. ChatGPT supported countries. OpenAI. URL: https://help.openai.com/en/articles/7947663-chatgpt-supported-countries-and-territories [accessed 2025-12-12]
26. Duan S, Liu C, Rong T, Zhao Y, Liu B. Integrating AI in medical education: a comprehensive study of medical students' attitudes, concerns, and behavioral intentions. BMC Med Educ 2025 Apr 23;25(1):599. [doi: 10.1186/s12909-025-07177-9] [Medline: 40269824]
27. Zhai J, Zhang H. Critical thinking disposition of medical students in Anhui Province, China: a cross-sectional investigation. BMC Med Educ 2023 Sep 8;23(1):652. [doi: 10.1186/s12909-023-04646-x] [Medline: 37684599]
28. Abdelhafiz AS, Farghly MI, Sultan EA, Abouelmagd ME, Ashmawy Y, Elsebaie EH. Medical students and ChatGPT: analyzing attitudes, practices, and academic perceptions. BMC Med Educ 2025 Feb 5;25(1):187. [doi: 10.1186/s12909-025-06731-9] [Medline: 39910593]
29. George Pallivathukal R, Kyaw Soe HH, Donald PM, Samson RS, Hj Ismail AR. ChatGPT for academic purposes: survey among undergraduate healthcare students in Malaysia. Cureus 2024 Jan;16(1):e53032. [doi: 10.7759/cureus.53032] [Medline: 38410331]
30. Xu S, Su T, Wang H, Xu J, Liu P, Tang Y. On the differences and reasons of online courses demand for military medical students at different educational levels: taking an example of a military medical university. Chin J Med Educ Res 2021;20(11):1334-1338. [doi: 10.3760/cma.j.cn116021-20200630-00681]
31. Zhao W, Zhu B. Learn to learn, learn to study, learn to create: different focuses for different levels of students. Acad Degree Graduate Education 2011(2):52-56. [doi: 10.3969/j.issn.1001-960X.2011.02.013]
32. You S, Zhou J, Li J, Kong M, Tan W. Practice and reflection on cultivating top-notch innovative talents in life sciences and medicine. Bull Chin Acad Sci 2025;40(2):260-272. [doi: 10.16418/j.issn.1000-3045.20241106001]
33. Alsobhi M, Khan F, Chevidikunnan MF, Basuodan R, Shawli L, Neamatallah Z. Physical therapists' knowledge and attitudes regarding artificial intelligence applications in health care and rehabilitation: cross-sectional study. J Med Internet Res 2022 Oct 20;24(10):e39565. [doi: 10.2196/39565] [Medline: 36264614]
34. Li Y, Xu J, Jia C, Zhai X. Investigation of college students' generative artificial intelligence (GAI) usage status and its implication: taking Zhejiang University as an example. Open Educ Res (1):89-98. [doi: 10.13966/j.cnki.kfjyyj.2024.01.010]
35. Wu N, Chen S, Yu B, Chen X. The current situation and reflection on the application of generative artificial intelligence assisted learning for vocational nursing students. Chin J Nurs Educ (4):403-409. [doi: 10.3761/j.issn.1672-9234.2025.04.004] [Medline: 24063287]
36. Liu S, Wright AP, Patterson BL, et al. Using AI-generated suggestions from ChatGPT to optimize clinical decision support. J Am Med Inform Assoc 2023 Jun 20;30(7):1237-1245. [doi: 10.1093/jamia/ocad072] [Medline: 37087108]
37. DeepSeek Tongji localized model user guide. Sohu. URL: https://www.sohu.com/a/862666794_712343 [accessed 2025-10-12]
38. Jiao Xiao Zhi Agent Management Platform. Xi'an Jiaotong University. URL: https://agent.xjtu.edu.cn/index.htm [accessed 2025-10-12]
39. Elhassan SE, Sajid MR, Syed AM, Fathima SA, Khan BS, Tamim H. Assessing familiarity, usage patterns, and attitudes of medical students toward ChatGPT and other chat-based AI apps in medical education: cross-sectional questionnaire study. JMIR Med Educ 2025 Jan 30;11:e63065. [doi: 10.2196/63065] [Medline: 39883912]

Abbreviations

AI: artificial intelligence
LLM: large language model
TTF: task-technology fit model

Edited by A Kushniruk; submitted 31.Jul.2025; peer-reviewed by AR Hoque, B Mohan, F Anupama, P Usuemerai, R Mamidipally; revised version received 02.Dec.2025; accepted 04.Dec.2025; published 08.Jan.2026.

Please cite as:
Shi X, Jiang Z, Xiong L, Siu KC, Chen Z
Utilization of AI Among Medical Students and Development of AI Education Platforms in Medical Institutions: Cross-Sectional Study
JMIR Hum Factors 2026;13:e81652
URL: https://humanfactors.jmir.org/2026/1/e81652
doi:10.2196/81652

XSL•FO
RenderX

© Xiaokang Shi, Zewu Jiang, Li Xiong, Ka-Chun Siu, Zhen Chen. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 8.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Evaluation of Combined Educational Methods on Motivational Interviewing for Final-Year Medical Students: Mixed Methods Study

Isaraporn Thepwongsa[1], MFM, MD, PhD; Radhakrishnan Muthukumar[2], MClinEmbryol, MBBS, PhD; Pat Nonjui[1], MD

[1]Department of Community, Family and Occupational Medicine, Faculty of Medicine, Khon Kaen University, Khon Kaen, Khon Kaen, Thailand
[2]Academic Affairs, Faculty of Medicine, Khon Kaen University, 123 Mittraparb Road, Muang Khon Kaen, Khon Kaen, Thailand

Corresponding Author:
Radhakrishnan Muthukumar, MClinEmbryol, MBBS, PhD
Academic Affairs, Faculty of Medicine, Khon Kaen University, 123 Mittraparb Road, Muang Khon Kaen, Khon Kaen, Thailand

Abstract

Background: Motivational interviewing (MI) is a patient-centered communication approach that supports health behavior change; yet, its integration into undergraduate medical curricula remains inconsistent. Combined learning models that comprise face-to-face instruction with structured web-based components may strengthen MI training, but evidence supporting their effectiveness among medical students, particularly in Asian contexts, is limited.

Objective: This study evaluated the impact of a combined MI educational model on final-year medical students' MI knowledge, confidence, and application in real patient encounters during clinical rotations.

Methods: This study used a sequential explanatory mixed methods design. The quantitative component used a before-and-after study to evaluate changes in MI knowledge and confidence among final-year medical students enrolled in an Ambulatory Care course in 2024. All 130 students participated in a 2-hour interactive MI workshop, and 120 completed pre- and postintervention questionnaires assessing MI knowledge and self-reported confidence. Students were also provided access to a 3-hour web-based MI learning module, and learning-management system analytics were used to track engagement. The qualitative component consisted of semistructured interviews with 12 purposively selected students, conducted to explore their experiences applying MI during clinical encounters. Quantitative data were analyzed using paired-samples *t* tests, and qualitative data were analyzed using inductive conventional content analysis. Findings from both components were integrated during interpretation to provide a comprehensive understanding of the educational intervention.

Results: Students demonstrated a significant improvement in MI knowledge following the educational intervention (pretest mean 8.87, SD 2.69; posttest mean 15.04, SD 2.99; $t_{1\ \ 1}$ =–18.45; P<.001; η^2=0.74). After the workshop, 96.9% (126/130) of students reported applying MI with patients, and 92.3% (n=120) agreed that the combined learning approach was adequate for supporting clinical use. Learning analytics data showed that 76.9% (n=100) of students enrolled in the web-based MI module, and 51% (n=51) completed all lessons. Students most frequently applied MI when counseling patients with diabetes, hypertension, and dyslipidemia, especially related to diet, physical activity, and medication adherence. Interview findings indicated that students mainly used brief MI, were most comfortable with engaging and focusing, and developed greater empathy, confidence, and patient-centered communication skills. Challenges included limited time during consultations, clinical workload, and difficulty applying all MI processes to complex cases.

Conclusions: A combined MI learning approach integrating a short workshop with a web-based course was associated with higher MI knowledge scores and greater self-reported confidence among students, as well as reported use of MI-informed communication strategies during clinical encounters. Students perceived MI as a practical and ethically grounded communication approach that can enhance patient engagement, particularly in the management of chronic diseases. Introducing MI training longitudinally through a spiral curriculum, with opportunities for repeated practice and reinforcement, may help strengthen behavior-change communication competencies in undergraduate medical education.

(JMIR Hum Factors 2026;13:e89126) doi:10.2196/89126

KEYWORDS

motivational interviewing; medical student; medical education; behavioral changes; combined educational methods; web-based learning; communication skills

XSL•FO
RenderX

Introduction

Background

Patient-centered communication is a fundamental component of effective clinical practice and plays a critical role in promoting health behavior change and improving clinical outcomes [1-5]. Physicians frequently engage in patient education during routine consultations, particularly when managing chronic diseases such as diabetes, hypertension, and cardiovascular conditions that require sustained lifestyle modification and self-management [1]. Evidence suggests that structured patient education interventions can improve clinical outcomes across a wide range of conditions, including cardiovascular disease, diabetes, respiratory disorders, and mental health conditions [1,2]. Effective patient education is typically supported by communication approaches that emphasize patient engagement, individualized counseling, and shared decision-making [6,7].

Among the various communication strategies used to support behavioral change, motivational interviewing (MI) has emerged as one of the most widely studied and evidence-based approaches [8]. MI is a collaborative, patient-centered counseling method developed by Miller and Rollnick [8] that aims to strengthen individuals' motivation and commitment to change by exploring and resolving ambivalence [9]. The approach emphasizes empathy, collaboration, and respect for patient autonomy while guiding patients toward identifying their own motivations for change. Over the past 3 decades, MI has been applied across a wide range of health care settings and clinical conditions [9,10].

A growing body of evidence demonstrates that MI can effectively support improvements in health behaviors and clinical outcomes [9-15]. Systematic reviews and meta-analyses have shown that MI interventions are associated with increased physical activity, smoking cessation, weight reduction, improved glycemic control, and reductions in alcohol consumption and other health-risk behaviors [11-14]. MI has also been successfully implemented in primary care settings, where physicians frequently counsel patients on lifestyle modification and chronic disease management [15-17]. However, despite its demonstrated effectiveness, studies indicate that motivational and patient-centered communication techniques remain underused in routine clinical practice, partly due to insufficient training among health care professionals [16,17].

Recognizing the importance of behavior-change communication, medical education programs have increasingly explored ways to incorporate MI training into undergraduate curricula [1]. Previous studies have shown that MI can be successfully taught to medical students and other health professional trainees using a variety of instructional strategies, including workshops, role-play exercises, coaching, and supervised practice [18,19]. These educational interventions have been associated with improvements in students' communication skills, confidence, and readiness to support patient behavioral change [20-22]. In addition, training in MI may help students develop transferable clinical communication skills such as reflective listening, open-ended questioning, and affirmation, which are valuable across many areas of clinical practice [6,23].

Despite these promising findings, several challenges remain in integrating MI training into medical curricula. Existing educational programs vary widely in duration, teaching methods, and assessment approaches, and there is no widely accepted standard model for MI education in undergraduate medical training [1,18,19,24-26]. In many programs, MI instruction is limited to short workshops without sufficient opportunities for reinforcement, practice, or longitudinal integration. As a result, the sustained application of MI skills in clinical practice remains inconsistent [10,24,25].

Multifaceted or combined educational methods comprising face-to-face instruction with web-based learning modules have been proposed as a promising strategy for improving MI training [14,27]. Such approaches allow learners to acquire foundational knowledge through flexible online learning while using in-person sessions for interactive practice and feedback. Emerging evidence suggests that combined educational methods can enhance learner engagement, knowledge retention, and skill development compared with traditional teaching methods alone [18,19,28]. However, relatively few studies have examined the effectiveness of combined MI training models among undergraduate medical students, particularly in Asian and low- and middle-income educational contexts [18,19,26].

Although MI training has been studied in medical education, relatively little research has examined how combined educational approaches comprising in-person workshops with web-based modules can be implemented within undergraduate medical curricula in resource-constrained educational settings. This study therefore evaluated the implementation of a combined learning approach for introducing MI to final-year medical students in Thailand and explored students' learning experiences and reported application of MI during clinical training.

Our Medical School Curriculum

Our medical school, located in northeastern Thailand, enrolls approximately 310 students annually. The 6-year undergraduate medical curriculum consists of a premedical phase, a preclinical phase, and a clinical phase, totaling 259 academic credits [29,30]. The first year focuses on premedical education (38.5 credits), while the second and third years emphasize preclinical sciences (76.5 credits). Clinical education occurs during years 4 to 6 and comprises 96 credits during years 4 and 5, followed by 48 credits during the final-year clinical rotations [30].

The learning outcomes of the curriculum are categorized into 3 domains: fundamental knowledge, professional competency, and professionalism. Communication skills are included as a core competency under the professionalism domain. Communication training is integrated throughout both preclinical and clinical courses; however, structured training specifically focused on behavior-change communication techniques has not been mandatory in the curriculum.

Before the 2024 academic year, some students may have been exposed to MI informally during clinical rotations in psychiatry. Beginning in the 2024 academic year, MI training was formally introduced as a compulsory component of the final-year

XSL•FO
RenderX

curriculum within the Ambulatory Care in Primary Care Setting course, where this study was conducted.

Final-year medical students participating in this course have already completed core clinical rotations and are actively involved in outpatient care under faculty supervision. During these rotations, students conduct patient interviews, participate in clinical decision-making, and provide counseling for chronic disease management. This clinical environment provides opportunities for students to apply behavior-change communication techniques such as MI during real patient encounters.

Methods

Study Design

Overview

This study used a sequential explanatory mixed methods design consisting of 2 complementary phases. In the first phase, a quantitative before-and-after evaluation was conducted to assess changes in final-year medical students' knowledge in MI following participation in the educational intervention.

In the second phase, semistructured qualitative interviews were conducted after completion of the quantitative data collection to further explore students' experiences in learning and applying MI during clinical encounters. The qualitative component was designed to help explain and contextualize the quantitative findings, particularly regarding how students applied MI in real clinical settings and the challenges they encountered when using MI with patients. The integration of quantitative and qualitative findings provided contextual insights into students' learning experiences and complemented the quantitative results.

As the training was implemented as part of the required curriculum for all students, a comparison group was not feasible; therefore, this study was designed as an educational program evaluation rather than an experimental effectiveness study.

Participants

This study's population consisted of final-year medical students enrolled in the "Ambulatory Care in Primary Care Setting" course during the 2024 academic year (May 2024 to April 2025). This course is a required component of the clinical curriculum and includes outpatient training in primary care settings.

Students who participated in the pilot testing of the questionnaire instruments (n=28) were excluded from this study to avoid potential bias related to prior exposure to the survey tools.

Sample size estimation was conducted using G*Power (version 3.1.9.7; Heinrich-Heine-Universität Düsseldorf) for paired comparisons of pre- and posttest scores [31]. Assuming a moderate effect size of 0.5, a statistical power of 80%, and a significance level of 0.05, the minimum required sample size was estimated to be approximately 50 participants.

The assumed effect size of 0.5 was selected based on conventions for moderate effect sizes commonly used in educational intervention studies and prior research evaluating communication skills and MI training among health care trainees.

Although all eligible final-year medical students enrolled in the course were invited to participate in this study, the power analysis was conducted before data collection to verify that the available cohort size would be sufficient to detect meaningful differences in knowledge scores between pre- and posttests. As the available cohort included 130 students, the sample size exceeded the minimum requirement estimated by the power analysis.

Educational Interventions

The educational intervention was designed to introduce foundational principles of MI and to provide opportunities for students to apply MI concepts in clinical contexts. The intervention combined a face-to-face interactive workshop with a structured web-based learning module and brief MI guides, allowing students to acquire theoretical knowledge and practice communication strategies relevant to behavior-change counseling in primary care.

Workshop

A 2-hour interactive workshop on MI was provided. The students attended the workshop in small rotation groups (25 - 28 students per group; 6 rotations throughout the academic year). The workshop facilitators (IT and PN) are faculty members in family medicine and medical education with experience in behavioral counseling in primary care. They have completed formal MI training programs and are continuing professional development activities related to MI.

The workshop content included an overview of health-behavior change theory, the principles and spirit of MI, and the 4 core MI processes (engaging, focusing, evoking, and planning). Students were introduced to the use of key MI communication strategies, including open-ended questions, affirmations, reflective listening, and summaries.

Interactive learning activities were incorporated to support skill development. These included case-based discussions, role-play exercises, and guided demonstrations of brief MI techniques. Students practiced applying MI strategies to simulated clinical scenarios commonly encountered in primary care, particularly in counseling patients with chronic conditions such as diabetes and hypertension.

Web-Based Course

In addition to the workshop, students were provided access to a structured web-based MI learning module designed to reinforce key concepts and support self-directed learning. The module contained approximately 3 hours of learning material and included short lessons covering MI principles, stages of change, and practical applications in patient consultations.

The online course incorporated interactive elements such as quizzes, self-reflection exercises, and case-based learning activities to encourage active engagement. The web-based curriculum was developed following established principles for effective online education [32]. The web-based module was provided as supplementary learning material to reinforce the

XSL•FO
RenderX

workshop content. Students could voluntarily access and complete the course at their own pace through the medical school's learning management system.

Completion of the web-based module was defined according to the platform's predefined criteria, which required students to complete all lessons and associated activities in the online course.

Brief MI Guide

A brief MI guide was provided as a downloadable resource within the web-based course and distributed during the workshop. The guide summarized key MI communication principles and included practical examples of open-ended questions, reflective statements, and strategies for exploring patients' motivation for behavior change. It also provided practical prompts to help students incorporate MI-informed communication into routine outpatient consultations. The guide served as a quick reference tool during clinical training rather than a comprehensive MI training manual. The resource was adapted from the "A Taste of MI" exercise included in the MI foundational training developed by Miller, Rollnick, and Moyers [33] and has been described in our previous publication [27].

The learning objectives of the intervention were to (1) introduce students to the principles and spirit of MI; (2) familiarize students with the four MI processes—engaging, focusing, evoking, and planning—and with fundamental MI communication skills, including open-ended questions, affirmations, reflective listening, and summaries; and (3) support the initial application of MI-informed communication strategies, including brief MI techniques, during clinical encounters.

Given the limited duration of the training, the program was designed to provide introductory exposure to MI concepts and basic communication strategies rather than to develop full MI proficiency.

The interactive workshop was a required component of the Ambulatory Care course, and therefore, all enrolled students attended the workshop session. In contrast, the web-based MI learning module was provided as a supplementary self-directed learning resource, allowing students to access and complete the lessons voluntarily at their own pace through the learning management system.

Quantitative Component

A before-and-after design was applied to assess students' knowledge before and after participation in the MI training. Students completed the pretest questionnaire before participating in the MI workshop. The baseline questionnaire included demographic characteristics (age and sex) as well as information about students' prior exposure to MI and other communication strategies used in patient counseling. Additional items assessed whether students had previously received training related to MI and their experiences with behavior-change communication during clinical rotations. These baseline variables were collected to characterize this study's population and to provide contextual information regarding students' prior familiarity with MI and communication skills before participating in the educational intervention. The posttest questionnaire was administered immediately after completion of the workshop to assess immediate changes in knowledge and confidence following the instructional session. The web-based module was provided as supplementary learning material that students could access voluntarily after the workshop to reinforce MI concepts during their clinical training. The instruments were developed by the research team (IT and RM) based on the relevant literature [14] and reviewed by experts in medical education and general practice to ensure content validity. Three independent reviewers assessed each item for clarity, relevance, and necessity. The questionnaires included multiple-choice, yes or no, open-ended, and Likert-scale questions evaluating knowledge, attitudes, and self-reported application toward MI.

Knowledge scores were calculated by summing the number of correct responses to the multiple-choice questions, generating a total knowledge score for each participant in both the pre- and posttests.

For attitudinal measures, the postintervention questionnaire included items assessing students' self-reported confidence in applying MI during patient consultations. Confidence was assessed across the 4 MI processes—engaging, focusing, evoking, and planning. Participants rated their confidence using a 5-point Likert scale ranging from 1 (not at all confident) to 5 (very confident). Students also reported the extent to which they applied MI during clinical encounters using a 5-point Likert scale ranging from 1 (used the least) to 5 (used the most). The questionnaire items were developed based on the MI framework described by Miller and Rollnick and were reviewed by experts in medical education and general practice to ensure content validity. After piloting with 28 students, minor revisions were made. Internal consistency reliability was assessed for the Likert-scale items measuring confidence related to MI, yielding a Cronbach α coefficient of 0.93. The survey instrument used in this study is available from the corresponding author upon reasonable request.

Qualitative Component

Semistructured interviews were conducted with students who had completed both the pre- and posttests. The semistructured interview guide was developed by the research team based on this study's objectives, findings from the quantitative phase, and relevant literature on MI education and communication skills training in medical education. The interview questions explored several key domains, including students' experiences applying MI in clinical encounters, which MI processes or techniques they used most frequently, perceived changes in their communication style following the training, challenges encountered when attempting to apply MI during patient consultations, perceived benefits of MI for patient communication, and students' perspectives on the most appropriate timing and methods for teaching MI in the medical curriculum. All interview participants were drawn from the same cohort of final-year medical students who participated in the quantitative component of this study.

Students were informed about the qualitative component after completing the MI training and posttest questionnaire and were invited to participate in voluntary interviews. Students who

XSL•FO
RenderX

expressed interest were contacted by the research team and scheduled for an interview.

A total of 12 students were selected using purposive sampling to ensure representation from different periods of course participation during the academic year. Specifically, students were selected from early, middle, and late course rotations in order to capture a range of experiences with learning and applying MI during clinical training.

The sample size was guided by the principle of thematic sufficiency commonly used in qualitative research exploring participant experiences. During analysis, the research team observed that key themes began to recur across interviews, suggesting that sufficient depth of information had been obtained for the purposes of this exploratory qualitative component.

Interviews were conducted in May 2025 via Google Meet and lasted approximately 15 - 20 minutes. The relatively brief duration reflected scheduling constraints during students' clinical rotations; however, the semistructured format enabled focused discussion of key topics related to students' experiences with learning and applying MI. All interviews were audio-recorded with participants' consent, transcribed verbatim, and returned to participants for verification to ensure transcription accuracy.

The semistructured interview guide used in this study is available from the corresponding author upon reasonable request.

Data Analysis

Quantitative Analysis

Statistical analyses were performed using IBM SPSS Statistics for Windows (version 20.0; IBM Corp). Paired-samples *t* tests were used to compare the mean knowledge scores before and after the intervention. Analyses were conducted using pairwise deletion, meaning that only participants who had both pre- and posttest responses for the relevant variable were included in the paired comparisons. Participants with missing responses in either the pretest or posttest were excluded from the paired analysis for that variable.

These measures were designed to assess students' perceived knowledge, confidence, and self-reported use of MI-informed communication strategies, rather than to provide an objective assessment of MI competency.

Qualitative Analysis

The interview data were analyzed using conventional content analysis, following the methodological approach described by Hsieh and Shannon [34]. An inductive approach was adopted, involving repeated reading of transcripts to identify initial codes, which were subsequently grouped into broader categories and themes. Initial coding was conducted by 1 researcher (IT), after which emerging codes, categories, and interpretations were reviewed and discussed with the research team to enhance analytical rigor and credibility.

Integration of Quantitative and Qualitative Data

The 2 data strands were integrated during interpretation to provide a comprehensive understanding of the intervention's impact. Quantitative findings demonstrated measurable improvements in knowledge and confidence, while qualitative insights illustrated how students experienced, applied, and reflected on MI in clinical contexts. This integration allowed qualitative findings to complement and provide contextual insights into the quantitative results, strengthening the overall interpretation of this study's findings.

Ethical Considerations

This study was approved by the Khon Kaen University Human Research Ethics Committee (project number HE631031). Participation was voluntary, and all students were informed of their rights to withdraw at any stage without academic consequences. Informed consent was obtained from all participants before data collection. Confidentiality and anonymity were maintained throughout the research process. Participants were encouraged to share honest feedback, as their insights were considered valuable for improving educational approaches in behavioral change communication training. Participation was voluntary, and no incentives were provided.

Results

Overview

Table 1 shows demographic data of the participating final-year medical students. None of the participants had prior MI training or course completion within the past 3 years. Table 2 shows student participation in components of the MI training program. All 130 final-year medical students attended the required 2-hour MI workshop as part of the Ambulatory Care course. The web-based MI learning module was offered as a voluntary supplementary learning resource. Learning-management system analytics showed that 100 (76.9%) students accessed or enrolled in the online module. Among those who accessed the module, 51 (51%) students completed all required lessons and activities according to the platform's completion criteria. Across students who accessed the module, the mean number of lessons completed was 9 of 19 lessons, indicating varying levels of engagement with the online content. Participants' enrollment time in the course ranged from 0 to 481 minutes, with a mean of 71.98 (SD 92.49). The number of completed lessons by participants ranged from 0 to 19 lessons, with a mean of 9.19 (SD 8.87). These findings indicate variable engagement with the web-based module, reflecting the self-directed nature of the online component.

XSL•FO
RenderX

Table . Demographic data of the participating final-year medical students.

Demographic data	Values, n (%)
Sex (n=128)	
Males	69 (53.9)
Females	59 (46.1)
Age (year; n=127)	
Mean (SD)	23.58 (0.61)
Range	22-25
Prior exposure to MI[a] within the past 3 years (n=127)	
Yes	18 (14.2)
No	109 (85.8)
Prior MI training or course completion within the past 3 years (n=127)	
Yes	0 (0.0)
No	127 (100.0)
Based on your experiences, which techniques did you use to help patients change their behaviors (n=127)	
Giving information	120 (94.5)
Advising	112 (88.2)
Order or instruct to do	12 (9.4)
Motivating	65 (51.2)
Never	3 (2.4)

[a]MI: motivational interviewing.

Table . Student participation in components of the MI[a] training program.

Training component	Values
Attended MI workshop (required), n (%)	130 (100)
Accessed or enrolled in web-based module, n (%)	100 (76.9)
Completed all web-based lessons, n/N	51/130
Lessons completed (among enrolled students; N=19), mean (SD)	9 (8.87)

[a]MI: motivational interviewing.

Most final-year medical students expressed strong needs and expectations regarding MI. Before completing the MI workshops, the majority wanted to strengthen their skills in choosing words carefully, applying stages of change, motivating patients, and enhancing communication for health-behavioral change (Table 3). Their learning expectations were consistently high, with mean scores near 4 on a 5-point scale, particularly for MI skills, knowledge, and helping patients control their diseases through MI (Table 4). Regarding specific behavioral areas, students prioritized skills related to oral medication adherence, diet and nutrition, doctor appointment adherence, and injection drug administration (Table 5). Overall, the data highlight the strong motivation among students to develop practical MI communication and counseling competencies that directly support patient self-management and lifestyle modification.

XSL•FO
RenderX

Table . Skills that participants wanted to develop to help patients with diabetes, other chronic diseases, or the general population in modifying health behaviors (before completion of the MI[a] workshops).

Skills that participants wanted to improve (n=127)	Values, n (%)
Choosing words carefully to enhance motivation	96 (75.6)
Applying the stage of change	95 (74.8)
Motivating	90 (70.9)
Communication skills to enhance motivation	90 (70.9)
Counseling	84 (66.1)
Facilitating successful health-behavioral change in patients	84 (66.1)

[a]MI: motivational interviewing.

Table . Participating final-year medical students' expectations for learning the MI[a] course.

Expectation in learning on MI (n=127)	Values, mean (SD)[b]
MI skills	4.06 (0.84)
Control patients' illnesses or diseases after applying MI with them	3.99 (0.72)
MI knowledge	3.97 (0.80)
Patients can change their behaviors after applying MI with them	3.94 (0.77)
Prompt use for MI	3.91 (0.87)
Teaching colleagues about MI	3.62 (0.91)

[a]MI: motivational interviewing.

[b]Mean was calculated using a 5-point Likert scale ranging from 1 (not at all expected), 2 (slightly expected), 3 (moderately expected), 4 (very much expected), and 5 (fully expected).

Table . Specific behaviors that the participating final-year medical students needed skills and knowledge in helping patients change.

Specific behaviors	Values, mean (SD)[a]
Oral medication administration	4.24 (0.79)
Diet and nutrition	4.20 (0.76)
Doctor appointment adherence	4.20 (0.83)
Injection drug administration	4.15 (0.84)
Exercise	4.13 (0.80)
Counseling skills for behavioral change	4.09 (0.75)
Behavioral changes in all aspects	4.06 (0.83)
Updating knowledge and skills in patient care	4.05 (0.83)
Working with the patient's family	3.94 (0.82)

[a]Mean was calculated using a 5-point Likert scale ranging from 1 (least needed), 2 (slightly needed), 3 (moderately needed), 4 (very needed), and 5 (most needed).

After attending the MI workshop, 126 of 130 (96.9%) final-year medical students reported attempting to apply elements of MI during patient encounters within the preceding 2 weeks. Most participants (n=120, 92.3%) indicated that the combination of a 2-hour workshop and web-based learning was sufficient for applying MI in clinical practice. Students perceived MI as a useful communication approach for supporting health-behavioral change (mean 4.47 of 5, SD 0.67).

After completing the MI workshop, the final-year medical students who applied MI techniques in real clinical settings predominantly used MI to support patients in improving their diet, physical activity, and medication adherence (Table 6), particularly among those with diabetes, hypertension, and dyslipidemia (Table 7). Students reported applying elements of MI during patient encounters and demonstrated moderate to high confidence in applying MI across all 4 MI processes (Table 8). The highest confidence score was observed for engaging (mean 3.73, SD 0.78), followed by focusing (mean 3.65, SD 0.74), planning (mean 3.52, SD 0.87), and evoking (mean 3.38, SD 0.90). These findings suggest that students felt relatively comfortable initiating and structuring MI conversations but reported slightly lower confidence in more advanced MI processes such as evoking change talk.

XSL•FO
RenderX

Table . Issues that the participating final-year medical student applied MI[a] to the patients.

Issues that the students applied MI with the patients (n=126)	Values, n (%)
Diet change	117 (92.9)
Exercise	100 (79.4)
Medication intake	67 (53.2)
Quitting alcohol consumption	43 (34.1)
Decision-making about treatment	42 (33.3)
Smoking cessation	39 (31)
Doctor appointment adherence	37 (29.4)

[a]MI: motivational interviewing.

Table . Types of patients with whom the final-year medical student applied MI[a].

Type of patients (n=126)	Values, n (%)
Diabetes	119 (94.4)
Hypertension	112 (88.9)
Dyslipidemia	111 (88.1)
Obesity	47 (37.3)
Stroke	12 (9.5)
Cardiovascular disease	8 (6.3)

[a]MI: motivational interviewing.

Table . MI[a] processes that were used by the final-year medical student to help patients change their behaviors and their level of confidence to use each MI process.

MI process (n=126)	Level of use, mean (SD)[b]	Level of confidence, mean (SD)[c]
Engaging	3.95 (0.93)	3.73 (0.78)
Focusing	3.59 (0.93)	3.65 (0.74)
Evoking	3.47 (0.97)	3.38 (0.90)
Planning	3.48 (1.10)	3.52 (0.87)

[a]MI: motivational interviewing.

[b]Mean was calculated using a 5-point Likert scale ranging from 1 (used the least), 2 (used a little), 3 (used moderately), 4 (used a lot), and 5 (used the most).

[c]Mean was calculated using a 5-point Likert scale ranging from 1 (not at all confident), 2 (somewhat not confident), 3 (neutral), 4 (confident), and 5 (very confident).

Comparing Participating Medical Students' Knowledge Scores Before and After Workshop Completion

A total of 130 final-year medical students participated in the MI workshop. Of these, 126 (96.9%) students completed the pretest questionnaire, and 120 (87.5%) students completed the posttest questionnaire. Students who did not complete the posttest were primarily absent during the postintervention data collection session or did not submit the questionnaire. Paired analyses were conducted using data from the 120 students who completed both the pre- and posttest. The proportion of missing data for the paired analysis was therefore small and unlikely to substantially affect the results. A paired-sample *t* test was conducted to examine differences in knowledge scores before and after participation in the MI training. There was a statistically significant difference in the knowledge test scores from time 1 (mean 8.87, SD 2.69) to time 2 (mean 15.04, SD 2.99), t_{119}=−18.45, P<.001 (2-tailed). The mean increase in knowledge test score was −6.18 (SD 3.66), 95% CI −6.84 to −5.51. The eta-squared statistic indicated a large effect size (eta-squared=0.74).

Insights From Participant Interviews

A total of 12 final-year medical students voluntarily participated in the semistructured interviews, representing 3 groups categorized by their enrollment period in the MI course: students who completed the course at the beginning of the 2024 academic year (May to August 2024), coded as group A; those who completed it in the middle of the academic year (September to December 2024), coded as group B; and those who completed it at the end of the academic year (January to April 2025), coded as group C. To maintain participant confidentiality while

XSL•FO
RenderX

distinguishing individual responses, each quotation is identified using a code consisting of a letter and a number. The letter (A, B, or C) represents the interview group based on the period of course participation, and the number represents the individual participant within that group.

The qualitative analysis identified several themes related to students' experiences learning and applying MI during clinical training, including experiences using MI in clinical encounters, changes in communication style, barriers to implementation, perceived advantages and benefits of MI, development of confidence, and students' perspectives on MI education.

According to the semistructured interviews (Table 9), the final-year medical students reported applying MI primarily in brief forms during patient encounters, particularly with patients managing chronic conditions such as diabetes, hypertension, dyslipidemia, and gout. They identified engaging as the most frequently used and comfortable skill, whereas evoking and planning were used less often due to time limitations and lower confidence. After completing the MI course, the students observed noticeable improvements in their communication, becoming more empathetic, reflective, and patient-centered. They perceived MI as a valuable approach that enhanced patient motivation, strengthened doctor-patient relationships, and promoted ethical and collaborative care. Nonetheless, challenges such as heavy workloads, time constraints, and communication barriers with some patients were reported. Most students expressed increased confidence in using MI and preferred learning it during their fifth year, when clinical exposure allowed sufficient opportunity for practice before the externship year. They also emphasized that effective MI instruction should integrate lectures with practical components, such as role-play, video demonstrations, and feedback sessions, ideally delivered through blended or flipped classroom models, to reinforce both theoretical and applied communication skills.

XSL•FO
RenderX

Table . The participating final-year medical students' experiences in and perceptions toward the use of MI[a] with patients.

Final-year medical students' experiences in and perceptions toward the use of brief MI with patients	Comments
Theme 1: students' experiences applying MI in clinical encounters • Final-year medical students reported applying MI during patient encounters, most often in the form of brief MI rather than the full 4-process model (A1-A4, B1-B5, and C1-C3). They most commonly used MI with patients managing NCDs[b] such as diabetes, hypertension, dyslipidemia, or gout (B1, A1, B3, C1, and C3), and in some cases to support smoking cessation (C3) or alcohol reduction (B2). • Across interviews, students identified engaging as the MI skill they used most frequently and felt most comfortable with, while evoking and planning were used less often because of limited time, clinical workload, or lower confidence in applying the full sequence (A1, B2, and B4). Many students explained that although they could not always perform all MI processes, they consistently incorporated key elements such as open-ended questioning, reflection, and affirmation into routine consultations (B5 and C3).	• "I used MI with patients with diabetes, hypertension, dyslipidemia and gout." (B1) • "In general, I try to explore the root cause of the patient's problem. I haven't had the opportunity to use the full MI process yet, but I try to engage them in talking about what happened and what consequences have occurred and tried brief MI with all patients with diabetes." (B2) • "I use every MI process and apply brief MI with all patients, mainly focusing on engaging. Sometimes patients ask, 'Will I get to see your supervisor?' or say things to make me feel like I'm not trustworthy enough, and insist that they would rather talk with my supervisor." (B4) • "Engaging is the basic process and the easiest to use. Affirmation is important for maintaining positive communication. Open-ended questions have some limitations because patients may not understand us, making them harder to use. I had already used some of these techniques before knowing MI through this course, and during the training, I felt good realizing that I had been using them already." (B5) • "Engaging is used very often, especially when adjusting medication for diabetic patients. I ask questions like, 'Why is the control not achieved?' or 'Is there anything that can be changed?'—but I haven't gone as far as the evoking process yet for the full MI but I tried to used brief MI. I also use the concept of the stages of change to assess which stage the patient is in and whether they are ready to change their behavior." (A1) • "I use Engaging in the outpatient department because we usually focus only on the patient's presenting problem. In reality, we can inspire or spark awareness in the patient. For example, a patient came with diabetes and also smoked. At first, I didn't focus on the smoking issue, but I tried to use reflection, which helped the patient realize that the number of cigarettes he smoked was quite high. When someone reflected this back to him, the patient said, 'I know it's a lot, right, doctor?' In that case, I went through all four MI processes. It felt like small talk that helps initiate the conversation." (C3)

XSL•FO
RenderX

Final-year medical students' experiences in and perceptions toward the use of brief MI with patients	Comments
Theme 2: changes in students' communication style after MI training • Final-year medical students reported clear changes in their communication style after completing the MI course, most commonly integrating engaging, open-ended questioning, and affirmation into their routine consultations (B1, B2, B4, and C1). Engaging was the MI process used most frequently, often combined with focusing to help patients better understand their condition and explore their own reasons for behavioral change (B4 and C1). Several students also described increased use of reflection and empathic listening, which helped patients express their motivations more freely and led to deeper conversations (C3). • Although evoking and planning were used less consistently, students reported feeling more confident overall and more capable of supporting patient-led change, moving away from directive advice-giving toward a more collaborative style of consultation (A1 and C3).	• "Used Engaging and tried to ask open-ended questions. Changed my way of working significantly after learning MI. After the course, I felt more confident using it and even tried applying it with my family members." (B1) • "After learning about Motivational Interviewing, what I've continued to use is open-ended questions. I ask open-ended questions to let them share, and then I practice empathic listening. Before the course, I had used affirmations occasionally, but after the course, I began using them more often. Giving affirmations helps patients feel proud of themselves and more motivated." (B2) • "The skill I continue to use regularly is engaging. I even applied it with my grandfather at home regarding his smoking. We had conversations using engaging and focusing, and as a result, he reduced his smoking. Affirmation is the skill I use the most." (B4) • "Engaging is probably the skill that has stayed with me after the course because I don't usually see patients long term. Most of the time, I see them for just one visit. So, I use engaging and focusing—patients don't always understand what they have, but we help them understand themselves better and improve their compliance." (C1) • "I use reflection more often now. At first, I would just keep asking questions and taking notes based on what the patient said. But when I started using reflection, it made me pause and think, and it also helped the patient reflect on themselves at the same time. As for affirmation, it is very important. Positive reinforcement has always worked well. What the patient gains is the feeling that someone cares and is interested in them, which helps them open up more. This makes it easier when it comes time to offer advice." (C3) • "Before and after the course, I think there's a big difference. I had never really helped patients change their behavior, but in real life, it's very important. If we can help them reach a stage where they're ready to change, it has a significant impact on the condition they are dealing with." (A1)
Theme 3: challenges and barriers to using MI in clinical practice • Final-year medical students reported several challenges when attempting to use MI in clinical settings. A major barrier was time pressure and high patient load, which made it difficult to perform all MI processes and often resulted in reverting to traditional directive advice-giving (B3, A2, and B5). Many students also noted that their limited experience and developing confidence in MI techniques led to hesitation, disrupted communication flow, or incomplete use of the approach (B1). • Students described difficulties in choosing the right wording, keeping conversations focused, and managing open-ended questions—which sometimes led patients off-topic and required more time than was available (A3 and C2). Additional obstacles included challenges communicating with older adult patients who had hearing problems or preferred to speak only with a doctor rather than a student (B4). Some students also encountered patients with fixed beliefs or preconceived online information, which made behavioral change difficult unless MI skills were used fluently and confidently (C1).	

XSL•FO
RenderX

Final-year medical students' experiences in and perceptions toward the use of brief MI with patients	Comments
	• "While using it, I couldn't think of how to proceed, so the interaction wasn't smooth. I've only just started using it and I'm not yet proficient. My skill and experience are still limited." (B1) • "How to use words so they understand us. Choosing the right words/change talks is challenging. If I could follow up with the same patient, I think I would be able to use this effectively—it would be doable if it's the same person. In real life, it takes a lot of time to listen. I believe it can be applied to everyone, but for now, I haven't used it with every patient." (C2) • "With open-ended questions, patients often talk about many things off-topic, which takes a lot of time." (A3) • "Because of the time constraints and the high number of patients, there's very little time left for each consultation. If I try to use MI, it becomes impractical—I won't be able to finish on time, and it makes other patients wait longer." (B3) • "Elderly patients often have trouble hearing, so I have to speak loudly, and they are usually not very cooperative. It takes a long time to talk with them. Patients prefer to make plans with the doctor (not with the student)." (B4) • "Because of limited time in real-life practice, I may not be able to use every step—only the basic ones. Patients may not come for counseling but rather for treatment. Some principles may not be fully understood at first, until I gain more experience using them with patients." (B5) • "If we are unable or lack the necessary skills, we may not be able to apply the principles in practice. Many patients nowadays have access to technology. They search for information on their own and come with fixed ideas or a fixed mindset. If we don't have the skills, we won't be able to help them." (C1) • "Conversations take time, and because of the large number of patients, we are unable to complete the full process. The time we have is limited, so we end up giving direct advice in the usual way—such as about alcohol or diet—which fits within the time limit, but the patients still don't change." (A2)

XSL•FO
RenderX

Final-year medical students' experiences in and perceptions toward the use of brief MI with patients	Comments
Theme 4: perceived benefits of MI for patient communication and clinical practice • Final-year medical students described multiple benefits of applying MI in routine consultations, particularly in improving communication and strengthening the doctor-patient relationship. Students perceived MI as a valuable approach for establishing shared goals and involving patients in decision-making, rather than relying on 1-way directive advice (B3 and A2). This collaborative communication style helped patients feel listened to and respected, which in turn increased trust and engagement during consultations (B2 and A3). • Many students reported that using MI enhanced patients' motivation and willingness to participate in their own care. They observed that when patients were encouraged to express their concerns and motivations, they became more confident and proactive in managing their health conditions (B4 and C3). Students noted that this approach was particularly useful in managing chronic diseases such as diabetes, where long-term behavior change is essential for effective care (B5). • Students also highlighted that MI helped them better understand patients' perspectives, emotions, and barriers to behavioral change, thereby fostering greater empathy in clinical interactions (C3 and A1). Several participants emphasized that MI aligns with patient-centered and ethical communication, helping to reduce misunderstandings and potentially decrease the likelihood of patient dissatisfaction or complaints (C1). • In addition to benefits for patient care, students reported that using MI positively influenced their own professional development. They described feeling more confident and satisfied when consultations became more collaborative and when patients responded positively to their communication approach (A1 and C3). Some students also noted that MI improved their broader communication skills, including active listening, reflective responses, and awareness of personal biases during clinical encounters (B5 and C2). • Finally, several participants suggested that MI may contribute to more efficient long-term care. They perceived that patients who felt motivated and supported were more likely to engage in self-management, which could reduce repeated consultations for the same behavioral issues. For many students, practicing MI also reinforced their sense of professional identity as a physician who communicates with empathy, respect, and patient-centered care (A1, A2, and B2).	

XSL•FO
RenderX

Final-year medical students' experiences in and perceptions toward the use of brief MI with patients	Comments
	• "This was the first thing that interested me—how to speak. It helps guide me a lot when talking with patients. I think it's very important." (B1) • "This is a skill every doctor should have. Before, I would just tell patients what to do without exploring how difficult behavioral change can be or how much time it takes to change. Now I use a patient-centered approach, finding a shared middle ground together." (B2) • "MI is important. Giving advice is necessary, but advice alone is a one-way communication and then it ends. MI, on the other hand, allows patients to make their own decisions. If we do it well, good outcomes will follow—like quitting smoking and preventing disease. With better clinical practice, the results we aim for will also improve." (A2) • "We don't force the patient to do anything — we plan together and see things from their perspective. It's better than just telling the patient what to do. It helps us understand the ways that lead to change. Even though we don't follow up with the patient afterward, during the interaction we can feel that the patient understands and has goals for change. It creates a good relationship with the patient." (A3) • "MI is important because it helps us find a shared goal. If the doctor doesn't do this, the patient may not want to make any changes—it feels as if someone is just ordering them, so they don't want to follow. MI helps patients want to improve themselves and want to collaborate with our guidance. The benefit for us is that it helps us understand patients better. It's not just about applying a theory and expecting results to happen automatically." (B3) • "MI is important. It's clear that patients become more cooperative. It's evident that MI truly helps motivate patients, and it helps us understand them better — we gain more insight into their background and situation. Patients are more willing to cooperate as a result." (B4) • "Using MI in the care of patients with diabetes makes management easier, if it is used correctly and fluently." (B5) • "MI is very important. Behavior is something that stays with patients for long time. What happen is the patient is often not understood by the doctor because we've never asked about their thoughts and feelings—we still don't really know each other. MI provides a natural way for us to do this. It gives us a structure to understand the patient. Behavior is hard to change. MI helps us reach the patient's feelings and talk about their attitudes, which is extremely helpful." (C3) • "It benefits us by helping us empathize with patients and listen to them more attentively. Sometimes we don't feel like listening to patients, even when what they say affects their treatment. It reminds us not to let our own bias get in the way, and to act professionally. It also helps us observe the patient's gestures and behavior." (B5) • "MI is important, yet it's something no one has ever brought up. MI is a part of medical ethics—listening without judgment and applying the principles of patient-centered care. We didn't recognize its importance before or realize what proper ethical principles really entail—we just practiced without knowing their significance." (C1) • "The benefit for myself is that I feel, in my heart, like a good doctor. The patient is happy, and they leave the consultation with a sense of satisfaction." (A1) • "It helps patients open up to us more, as if we were talking with a family member. When we encourage them and give positive support, that positivity comes back to us. The patient is happy, and we are happy — we receive positive energy. It makes us appreciate the value of caring for patients even more." (C3) • "The benefit for ourselves is improved communication skills, which help not only with patients but in general as well." (C2) • "The outpatient setting is very important. If there is proper guidance on how to speak with patients, it helps them feel comfortable and satisfied when they listen, and they become more cooperative. That's exactly what MI does." (A1) • "Doctor–patient relationship becomes more open, allowing patients to trust the doctor and feel understood. It gives them the opportunity to think for themselves, and they feel more confident and trusting.

XSL•FO
RenderX

Final-year medical students' experiences in and perceptions toward the use of brief MI with patients	Comments
	As a result, our work improves and becomes lighter—our workload is reduced because patients understand how to take care of themselves." (A2)
Theme 5: development of confidence in applying MI • After completing the MI course and practicing its techniques during clinical rotations, most final-year medical students reported increased confidence in communicating with patients and supporting behavioral change (B1, B3, and B4). Several students noted a shift from feeling uncertain or hesitant in offering guidance before the course to becoming more comfortable and proactive in conversations involving lifestyle changes or complex emotions (B4 and C3). • Students described how MI gave them a clearer framework for patient interaction, helping them move beyond simply giving advice to learning how to guide patients through reflective discussions and motivate change (A3 and B5). A few emphasized that increased confidence was also tied to stronger medical knowledge, which enabled more fluent and effective use of MI in practice (A1). • Although confidence was not immediate for everyone, many students shared that through repeated use—both in clinical encounters and informal settings—they became more adept at integrating MI naturally into their consultations (A2 and B5).	• "More confident in giving advice to patients, more willing to speak up, and becoming more familiar with it." (B1) • "More confident, and during OPD rotations I use MI more often." (B3) • "More confident now. Before the course, I didn't dare to give advice to patients." (B4) • "I am more confident in helping patients change their behavior. At first, I didn't know there was a principle behind it, I would just give advice or tell them what to do. In the future, I will use MI in situations where it is appropriate, but there are still limitations due to the high number of patients coming to the OPD." (B5) • "To be confident in speaking and using MI, we must first be confident in our medical knowledge. Once we are confident in our knowledge, we will feel more confident during patient consultations, and the more we speak, the more fluent and confident we become in using MI." (A1) • "More confident immediately after the course, but not at the highest level yet. After practicing for a while following the course, I became much more confident. I tried using it on my own and no longer needed additional training." (A2) • "Before learning MI, everything was scattered and without direction. I didn't know what to do and had no framework. • Now, I feel more confident." (C3) • "There is now a clearer starting framework. Before, I would give advice based on the knowledge I had — 1, 2, 3, 4 — without giving patients the chance to talk about why they couldn't follow it. After learning MI, I understand patients more, can help motivate them, and can apply a more effective approach to giving advice." (A3)
Theme 6: students' perspectives on when MI should be taught in the curriculum • Most final-year medical students agreed that year 5 is the most appropriate time to introduce MI in the curriculum, as it aligns with outpatient clinical exposure and provides enough time to practice before entering the externship year (B1, B2, and B4). Students noted that by year 5, they have sufficient medical knowledge and experience to make MI training meaningful and applicable, especially during rotations involving real patient care and behavioral counseling (A3 and B3). • A few students recommended introducing MI theory earlier, in years 3 or 4, followed by more practical reinforcement in years 5 and 6 to help develop fluency over time (C1 and A2). Some felt that teaching MI too late—such as near the end of year 6—would limit the opportunity to apply and refine skills in varied clinical settings (B2 and A2). • However, a small number suggested that year 6 is ideal, viewing it as the stage when clinical confidence matures enough to fully integrate MI into patient care (B5 and C3). Overall, the majority favored a staggered approach, with early theoretical exposure followed by increasing depth and practice in years 5 and 6.	

XSL•FO
RenderX

Final-year medical students' experiences in and perceptions toward the use of brief MI with patients	Comments
	• "I would like to learn it before entering year 6, so that I can practice more and apply it across different block rotations. It could be early in year 5 or during the family medicine rotation in year 5. Since in year 5 we have OPD sessions, while in year 4 we mostly focus on history taking and physical examination, learning it in year 5 would be better. Then, by the time we reach year 6, when we are more involved in the patient management, we would be able to use it more fluently." (B1) • "In years 5 and 6, there are more opportunities to explore patient care. In year 5, there is the OPD Ambulatory Medicine course. If MI is taught in year 5 and then applied in Ambulatory Care in Primary Care during year 6, that would be ideal. However, in year 6, if a student takes the block toward the end of the year, they would learn the techniques late and have less time to practice. Therefore, learning MI in year 5 would be better." (B2) • "Years 4 - 5 are okay because that's when we start working in the outpatient department. In year 4, we have a rotation in community hospitals, and I heard that the nurses there use MI with diabetic patients. So if we could learn it in year 4, that would be good. Then in years 5 - 6, we would be able to apply it in real practice." (B3) • "Learning it in year 5 would be good because we have outpatient rotations and can apply it. By year 6, we'd have a solid enough understanding to use it directly in clinical practice." (B4) • "I think teaching it during the extern year (year 6) is suitable because students already have confidence in patient care. But if it's taught before the extern year, students may not be ready yet." (B5) • "MI should start being taught theoretically in years 3 - 4. After learning the theory, students can then observe patients, and later receive reinforcement teaching again. In year 4, the focus should be on engaging and focusing, because that year is mainly about history taking and physical examination. In years 5 - 6, the teaching should shift to evoking and planning, because students begin learning about treatment and management. This would help students understand MI more deeply. Those who already enjoy talking with patients benefit a lot when they learn MI, but those who are more academically focused, not like to listen or talk may not use it as much." (C1) • "In year 4, it can be practiced during the family medicine rotation. If students start practicing in year 4 and continue in years 5 - 6, their skills will become more complete. If it is taught near the end of the extern year, they may not have many opportunities to apply it in different wards." (A2) • "Year 6, because that's when we get to use it in real practice. In year 4, during OPD rotations, it's difficult to apply because most of the learning is through lectures and observing instructors. If we had to do it ourselves, it would be hard to apply, since the focus in that year is mainly on history taking and physical examination." (C3) • "Years 5 - 6 are suitable because by the time we reach the stage of giving advice, we already have the necessary knowledge and understand the importance of advising patients. If it is taught too early, we won't have enough information or knowledge to really help patients. If it is taught too early, we will forget it, won't see its importance, and won't be focused. In years 5 - 6, students are more likely to want to use it." (A3)

XSL•FO
RenderX

Final-year medical students' experiences in and perceptions toward the use of brief MI with patients	Comments
Theme 7: preferred educational approaches for learning MI • Final-year medical students emphasized the importance of combining lectures with hands-on practice when teaching MI. While onsite lectures and web-based modules were considered helpful for providing foundational knowledge (B1, B3, and B4), most students stressed that practice-oriented learning—especially through role-play, video demonstrations, and feedback or coaching sessions—is essential for building confidence and applying MI effectively in real clinical settings (B2, C1, and C2). • Several students recommended a flipped classroom model, where theoretical content is delivered online before in-person sessions focused on skill application and reflection (B3 and B4). They also suggested that extended or repeated practice sessions, including opportunities to rehearse full MI processes and observe expert demonstrations (eg, videos or live modeling), would help address initial confusion—particularly around more complex components such as evoking and planning (A3 and B5). • Overall, students viewed a combined educational approach—with early theoretical input, case-based role-play, instructor feedback, and opportunities for ongoing practice—as key to mastering MI during undergraduate medical training.	• "Onsite lectures are good, but there is still not enough practice. When applying it with patients, there is still a lack of feedback. I would like there to be a practice session with feedback or coaching. Video demonstrations are useful and could be added to the web-based course, or a workshop could be organized for students or physicians who are interested, perhaps as an elective." (B1) • "I was impressed by the session where we practiced brief MI. I was able to try applying it with patients myself. The lecture is important, but I would like to have hands-on practice right away. The web-based format is already detailed and complete. We could watch the online clips before, during, or after trying it with patients. For the theory part, it should be summarized briefly and concisely, and then we should move straight to practice." (B2) • "Lecture+ web-based learning is already good. But if there were a wrap-up session, it would be even better, because in the classroom I couldn't keep up. If we could study the web-based content first and have some time to digest it, we would understand the in-class session better, since this is a new topic for us. Web-based learning is helpful because I can't always follow everything during the live class. • Role-play and videos would also be useful because they make it easier to apply MI and help build confidence in using it." (B3) • "I prefer learning onsite and then reviewing it again online. • Another option is to have students learn the theory on their own through the web-based course, and then the instructor can give a brief summary in class and let students do role-play." (B4) • "The teaching format is appropriate as it is. However, the session where we practiced brief MI was too short. The session should be extended, and we should also try practicing the full MI process. I would like to add a video demonstration because the current case studies are only on paper. After this, reviewing the principles doesn't require additional courses—I think we can continue practicing on our own." (B5) • "What I remember most from the course is how to speak well with patients. Even though it's been a long time, I still remember it clearly because I saw the instructor demonstrate it during the OPD rotation, especially how to approach difficult patients. The instructor was a good role model, and seeing the correct practice has stayed with me to this day." (A1) • "Lectures alone may not be enough—there should be more practice included. Perhaps add role-play demonstrations for difficult patients, with role-play sessions that classmates can observe. Coaching is challenging, but it would be very helpful if available. The content is sufficient, but what's missing is how to apply it—how to practice correctly and in what way. Skills should be refreshed regularly through small sessions, such as annual courses or workshops, so that when we apply them in practice, we have a solid foundation to rely on." (C1) • "Organize group role-plays in front of the class and have the instructor provide feedback and advice. • For physicians who want to use it, the training is sufficient, but continuous practice is necessary." (C2) • "I could follow about 50% of the lecture. When practicing, I didn't really understand what the evoking process meant or how to use it. I was still confused when trying it, but once I understood the concept, I felt that if we practiced with roleplay first and then had a follow-up lecture, it would be better. Learning from video demonstrations is difficult if we don't immediately encounter real cases. Maybe we could divide into groups for roleplay, then go back and watch the videos, and the next day do hands-on practice—that might work better. More time should be given for learning and self-practice, with more opportunities to practice." (A3)

[a]MI: motivational interviewing.

[b]NCD: noncommunicable diseases.

XSL•FO
RenderX

Discussion

Principal Findings

This mixed methods evaluation examined a combined MI training program for final-year medical students. Quantitative findings showed higher MI knowledge scores and greater self-reported confidence after participation in the training program. In addition, most students reported applying elements of MI during clinical encounters, primarily in brief interactions related to counseling patients with chronic diseases.

Qualitative findings provided further insight into students' learning experiences. Participants described increased awareness of patient-centered communication strategies, including reflective listening, open-ended questioning, and affirmations. However, students also reported challenges in applying more advanced MI processes, such as evoking and planning, due to time constraints, clinical workload, and limited opportunities for repeated practice. Together, these findings suggest that the combined learning approach was associated with improvements in knowledge and perceived communication confidence, while also highlighting practical challenges in implementing MI during routine clinical encounters.

It is important to note that these findings reflect self-reported learning outcomes and perceptions of MI use, rather than objectively measured MI competency or verified patient behavior change.

Comparison With Previous Studies

Our findings are broadly consistent with previous studies suggesting that brief MI training can increase learners' knowledge and confidence in behavior-change communication [18,19,35,36]. Similar to prior research, students in this study reported feeling most comfortable using the early MI processes of engaging and focusing, while evoking and planning were more difficult to apply without extended practice or supervision. Previous research has also suggested that blended learning approaches combining online instruction with interactive teaching may enhance learner engagement and flexibility in medical education [28].

However, it is important to note that many studies evaluating MI education rely primarily on self-reported outcomes, and evidence regarding objective skill acquisition remains limited. Therefore, while this study contributes additional evidence from a Thai undergraduate medical education context, further research using objective assessment tools and controlled study designs is needed to better evaluate the effectiveness of MI training programs for undergraduate medical students.

Consistent with qualitative findings, students perceived limited consultation time as a barrier to applying MI. These perceptions are consistent with previous reports from trainees who are newly learning MI techniques and may require additional time to formulate reflective responses and guide behavior-change conversations [27]. However, evidence from the broader MI literature suggests that when clinicians become more proficient in MI, the approach can be incorporated efficiently into routine clinical consultations without substantially increasing encounter time [9,37]. Some studies have even reported that effective MI communication may improve the efficiency of patient interactions by helping clinicians focus more quickly on patients' motivations and priorities for behavior change [9,38]. Therefore, the time-related challenges reported by students in the present study may reflect their early stage of learning rather than an inherent limitation of MI itself. As students gain more experience and practice integrating MI skills into clinical encounters, the perceived time burden may decrease.

Strengths and Weaknesses

A major strength of this study lies in its mixed methods design, which allows integration of quantitative findings with qualitative insights to provide a broader understanding of students' learning experiences. The inclusion of an entire cohort of final-year medical students reduced the selection bias and enhanced representativeness. The use of validated instruments and expert review increased internal validity.

Several limitations should be considered when interpreting the findings of this study. First, this study used a single-group before-and-after design without a control group, which prevents causal attribution of observed changes to the educational intervention. Improvements in knowledge or confidence may also reflect concurrent clinical learning experiences, maturation effects, or testing effects during the academic year.

Second, several outcomes, including students' reported application of MI in clinical encounters, were based on self-reported measures, which may be subject to recall bias or social desirability bias. These measures, therefore, reflect students' perceptions rather than objectively verified MI competence. In addition, the qualitative interviews were relatively brief (approximately 15 - 20 minutes) due to scheduling constraints during students' clinical rotations, which may have limited the depth of exploration of participants' experiences.

Third, the duration of the intervention was relatively brief (approximately 5 hours of combined learning), which may be sufficient for introducing foundational MI concepts but is unlikely to support mastery of more advanced MI skills without continued practice, feedback, and supervision [10,39].

The relatively low completion rate of the web-based module may also reflect challenges related to implementation fidelity, as not all students were fully engaged with the online component of the intervention. However, the web-based course was designed as a supplementary resource to support students who wished to review or reinforce knowledge delivered during the onsite sessions. Participation in the web-based module was optional, and therefore variations in uptake among students were expected. As this study was not originally designed to compare outcomes between students who completed the web-based module and those who did not, subgroup analyses were not conducted. Future studies may explore whether different levels of engagement with online learning components influence educational outcomes.

Finally, this study was conducted at a single medical school, and therefore, the findings may not be generalizable to other educational settings or health care systems.

XSL•FO
RenderX

Implications and Future Study

The findings of this study offer practical implications for undergraduate medical education. A combined learning structure that integrates a short workshop with a self-paced web-based module may provide a flexible approach for introducing MI concepts within a busy curriculum, although engagement with the online component varied among students. The web-based component provides flexibility, while the in-person workshop facilitates experiential learning and peer interaction.

Students' preference for learning MI during the fifth year supports the alignment of MI training with clinical exposure to maximize relevance and skill transfer. A spiral curriculum, introducing MI principles during preclinical years and reinforcing practical skills during clinical rotations, could further strengthen integration.

Addressing barriers such as workload and limited practice time will require institutional support, including formative feedback, supervision, and longitudinal reinforcement. Embedding MI within broader communication skills training and assessment frameworks may enhance sustainability.

Future research should incorporate objective measures of MI competency, such as validated behavioral coding systems or structured communication assessments, to more accurately evaluate skill acquisition. Longitudinal studies are also needed to examine the retention of MI skills and their potential impact on patient outcomes over time. In addition, multicenter and cross-cultural research in low- and middle-income countries would help determine how combined MI educational approaches can be adapted to different health care systems and educational environments.

Overall, the findings suggest that the intervention was feasible within an undergraduate medical curriculum and was associated with higher knowledge scores and increased perceived confidence among participating students, while qualitative findings illustrated how students perceived and attempted to apply MI principles during clinical training.

Conclusions

In summary, this study evaluated a combined educational approach comprising a brief workshop and a web-based learning module to introduce MI to final-year medical students. The findings suggest that the training was associated with improvements in students' MI knowledge and self-reported confidence in behavior-change communication. Students also reported applying elements of MI during clinical encounters, particularly when counseling patients with chronic diseases.

While these findings indicate that combined MI education is a feasible approach for introducing communication skills within undergraduate medical curricula, the results should be interpreted cautiously due to the absence of a control group and the reliance on self-reported outcomes. Further research using controlled study designs, objective assessment tools, and longitudinal follow-up is needed to better evaluate the impact of MI training on clinical competency and patient outcomes.

Acknowledgments

We sincerely thank the production house teams of the Faculty of Medicine, Khon Kaen University, for their invaluable support in developing and producing a structured web-based course on motivational interviewing and the stage of change. Generative artificial intelligence tools were used solely for language editing and grammatical refinement. All scientific content, analysis, interpretations, and references were developed and verified by the authors, who take full responsibility for the accuracy, originality, and integrity of this paper.

Funding

This study was supported by the Faculty of Medicine, Khon Kaen University, Thailand (IN63237).

Data Availability

The data from this research project are available upon reasonable request.

Conflicts of Interest

None declared.

References

1. Simonsmeier BA, Flaig M, Simacek T, Schneider M. What sixty years of research says about the effectiveness of patient education on health: a second order meta-analysis. Health Psychol Rev 2022 Sep;16(3):450-474. [doi: 10.1080/17437199.2021.1967184] [Medline: 34384337]
2. Hinchageri SS. To evaluate the effectiveness of patient education on clinical outcome in hospitalized diabetic patients by clinical pharmacist. J Drug Delivery Ther 2012 Jul;2(4). [doi: 10.22270/jddt.v2i4.191]
3. Hines LA. Educating patients on value-based health care to improve clinical outcomes. MIH 2023;8(1):32. [doi: 10.69554/CYCH3634]

XSL•FO
RenderX

4. Balea LB, Gulestø RJA, Xu H, Glasdam S. Physicians', pharmacists', and nurses' education of patients about antibiotic use and antimicrobial resistance in primary care settings: a qualitative systematic literature review. Front Antibiot 2024;3:1507868. [doi: 10.3389/frabi.2024.1507868] [Medline: 39850331]
5. Garshasbi S, Khazaeipour Z, Fakhraei N, Naghdi M. Evaluating knowledge, attitude and practice of health-care workers regarding patient education in Iran. Acta Med Iran 2016 Jan;54(1):58-66. [Medline: 26853292]
6. Howard LM, Williams BA. A focused ethnography of baccalaureate nursing students who are using motivational interviewing. J Nurs Scholarsh 2016 Sep;48(5):472-481. [doi: 10.1111/jnu.12224] [Medline: 27314559]
7. Wofford JL, Wells MD, Singh S. Best strategies for patient education about anticoagulation with warfarin: a systematic review. BMC Health Serv Res 2008 Feb 14;8:40. [doi: 10.1186/1472-6963-8-40] [Medline: 18275605]
8. Miller W, Rollnick S. The method of motivational interviewing in: motivational interviewing: preparing people for change. : Guilford Press; 2013 URL: https://bluepeteraustralia.wordpress.com/wp-content/uploads/2012/12/motivational-interviewing.pdf [accessed 2026-04-20]
9. Rubak S, Sandbaek A, Lauritzen T, Christensen B. Motivational interviewing: a systematic review and meta-analysis. Br J Gen Pract 2005 Apr;55(513):305-312. [Medline: 15826439]
10. Söderlund LL, Madson MB, Rubak S, Nilsen P. A systematic review of motivational interviewing training for general health care practitioners. Patient Educ Couns 2011 Jul;84(1):16-26. [doi: 10.1016/j.pec.2010.06.025] [Medline: 20667432]
11. Zhu S, Sinha D, Kirk M, et al. Effectiveness of behavioural interventions with motivational interviewing on physical activity outcomes in adults: systematic review and meta-analysis. BMJ 2024;386:e078713. [doi: 10.1136/bmj-2023-078713]
12. West DS, DiLillo V, Bursac Z, Gore SA, Greene PG. Motivational interviewing improves weight loss in women with type 2 diabetes. Diabetes Care 2007 May;30(5):1081-1087. [doi: 10.2337/dc06-1966] [Medline: 17337504]
13. Frost H, Campbell P, Maxwell M, et al. Effectiveness of motivational interviewing on adult behaviour change in health and social care settings: a systematic review of reviews. PLoS ONE 2018;13(10):e0204890. [doi: 10.1371/journal.pone.0204890] [Medline: 30335780]
14. Thepwongsa I, Muthukumar R, Kessomboon P. Motivational interviewing by general practitioners for type 2 diabetes patients: a systematic review. Fam Pract 2017 Aug 1;34(4):376-383. [doi: 10.1093/fampra/cmx045] [Medline: 28486622]
15. Christie D, Channon S. The potential for motivational interviewing to improve outcomes in the management of diabetes and obesity in paediatric and adult populations: a clinical review. Diabetes Obes Metab 2014 May;16(5):381-387. [doi: 10.1111/dom.12195] [Medline: 23927612]
16. Kaczmarek T, Kavanagh DJ, Lazzarini PA, Warnock J, Van Netten JJ. Training diabetes healthcare practitioners in motivational interviewing: a systematic review. Health Psychol Rev 2022 Sep;16(3):430-449. [doi: 10.1080/17437199.2021.1926308] [Medline: 33970799]
17. Moran J, Bekker H, Latchford G. Everyday use of patient-centred, motivational techniques in routine consultations between doctors and patients with diabetes. Patient Educ Couns 2008 Nov;73(2):224-231. [doi: 10.1016/j.pec.2008.07.006] [Medline: 18701234]
18. Lei LYC, Chew KS, Chai CS, Chen YY. Evidence for motivational interviewing in educational settings among medical schools: a scoping review. BMC Med Educ 2024 Aug 8;24(1):856. [doi: 10.1186/s12909-024-05845-w] [Medline: 39118104]
19. Kaltman S, Tankersley A. Teaching motivational interviewing to medical students: a systematic review. Acad Med 2020 Mar;95(3):458-469. [doi: 10.1097/ACM.0000000000003011] [Medline: 31577585]
20. Darnton R, Thomson R, McKimm J. Video consultations: a potential gamechanger in clinical education. Br J Gen Pract 2022 Sep;72(722):416-417. [doi: 10.3399/bjgp22X720509]
21. Kim GJ, Gahlot A, Magsombol C, et al. Feasibility of a self-directed upper extremity training program to promote actual arm use for individuals living in the community with chronic stroke. Arch Rehabil Res Clin Transl 2024 Mar;6(1):100316. [doi: 10.1016/j.arrct.2023.100316] [Medline: 38482107]
22. Rimayanti MU, Taylor NF, Shields N, Prendergast LA, O'Halloran PD. Health professionals can sustain proficiency in motivational interviewing with a moderate amount of training: an intervention fidelity study. J Contin Educ Health Prof 2024;44(2):105-111. [doi: 10.1097/CEH.0000000000000522] [Medline: 37366578]
23. King A, Hoppe RB. "Best practice" for patient-centered communication: a narrative review. J Grad Med Educ 2013 Sep;5(3):385-393. [doi: 10.4300/JGME-D-13-00072.1] [Medline: 24404300]
24. Bell DL, Roomaney R. Exploring the barriers that prevent practitioners from implementing motivational interviewing in their work with clients. SWORK 2020;56(4):416. [doi: 10.15270/56-4-884]
25. Adams RJ. Improving health outcomes with better patient understanding and education. Risk Manag Healthc Policy 2010;3:61-72. [doi: 10.2147/RMHP.S7500] [Medline: 22312219]
26. Ng WCE, Koura R, Khaider KB, Chew CSE, Davis C. Effectiveness of a hybrid, obesity-specific counselling programme in improving medical students' self-efficacy and motivational interviewing skills for paediatric obesity counselling. BMC Med Educ 2025 Feb 11;25(1):224. [doi: 10.1186/s12909-024-06589-3] [Medline: 39934793]
27. Thepwongsa I, Nonjui P, Muthukumar R, Sripa P. Impact of motivational interviewing education on general practitioners' and trainees' learning and diabetes outcomes in primary care: mixed methods study. JMIR Med Educ 2025 Sep 16;11:e75916. [doi: 10.2196/75916] [Medline: 40957062]

XSL•FO
RenderX

28. Erschens R, Fahse B, Festl-Wietek T, et al. Training medical students in motivational interviewing using a blended learning approach: a proof-of-concept study. Front Psychol 2023;14:1204810. [doi: 10.3389/fpsyg.2023.1204810] [Medline: 37546454]
29. Sripa P, Thepwongsa I, Muthukumar R. Factors associated with the entry to general practice training: a multicentre study. Med Teach 2020 Dec;42(12):1394-1400. [doi: 10.1080/0142159X.2020.1811846] [Medline: 32878524]
30. Mei A, Gao D, Jiang J, Qiao T, Wang F, Li D. The medical education systems in China and Thailand: a comparative study. Health Sci Rep 2022 Nov;5(6):e826. [doi: 10.1002/hsr2.826] [Medline: 36210882]
31. Bączek M, Zagańczyk-Bączek M, Szpringer M, Jaroszyński A, Wożakowska-Kapłon B. Students' perception of online learning during the COVID-19 pandemic: a survey study of polish medical students. Medicine (Abingdon) 2021 Feb 19;100(7):e24821. [doi: 10.1097/MD.0000000000024821]
32. Cook DA, Dupras DM. A practical guide to developing effective web-based learning. J Gen Intern Med 2004 Jun;19(6):698-707. [doi: 10.1111/j.1525-1497.2004.30029.x] [Medline: 15209610]
33. Motivational interviewing - foundational. Psychwire. URL: https://psychwire.com/motivational-interviewing/mi-foundational [accessed 2026-04-23]
34. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
35. Poirier MK, Clark MM, Cerhan JH, Pruthi S, Geda YE, Dale LC. Teaching motivational interviewing to first-year medical students to improve counseling skills in health behavior change. Mayo Clin Proc 2004 Mar;79(3):327-331. [doi: 10.4065/79.3.327] [Medline: 15008606]
36. Arnett MC, Evans MD, Stull CL. Brief motivational interviewing: evaluation of a skills-based education program. J Dent Hyg 2022 Aug;96(4):46-56. [Medline: 35906078]
37. Rollnick S, Miller WR, Butler CC. Motivational Interviewing in Health Care: Helping Patients Change Behavior, 2nd edition: The Guilford Press; 2022.
38. Lundahl B, Moleni T, Burke BL, et al. Motivational interviewing in medical care settings: a systematic review and meta-analysis of randomized controlled trials. Patient Educ Couns 2013 Nov;93(2):157-168. [doi: 10.1016/j.pec.2013.07.012] [Medline: 24001658]
39. Schwalbe CS, Oh HY, Zweben A. Sustaining motivational interviewing: a meta‐analysis of training studies. Addiction 2014 Aug;109(8):1287-1294. [doi: 10.1111/add.12558]

Abbreviations

MI: motivational interviewing

Edited by A Kushniruk; submitted 07.Dec.2025; peer-reviewed by A Carcone, S Zhao; revised version received 13.Mar.2026; accepted 01.Apr.2026; published 29.Apr.2026.

Please cite as:
Thepwongsa I, Muthukumar R, Nonjui P
Evaluation of Combined Educational Methods on Motivational Interviewing for Final-Year Medical Students: Mixed Methods Study
JMIR Hum Factors 2026;13:e89126
URL: https://humanfactors.jmir.org/2026/1/e89126
doi:10.2196/89126

© Isaraporn Thepwongsa, Radhakrishnan Muthukumar, Pat Nonjui. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 29.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Understanding Barriers to Effective Injury Care by Medical Trainees and Traffic Law Enforcement First Responders in Low-Income Contexts in Uganda (Motor Registry Project Part 2): Convergent Mixed Methods Analysis

Herman Lule[1,2,3], MBChB, MMED; Benson Oguttu[4], MBChB; Micheal Mugerwa[5], MSc; Michael Lowery Wilson[1,6], PhD; Jussi P Posti[7,8], MD, PhD

[1]Injury Epidemiology and Prevention (IEP) Research Group, Turku Brain Injury Centre, Division of Clinical Neurosciences, Turku University Hospital and University of Turku, Vähä Hämeenkatu 1B, Turun Yliopisto, Turku, Finland

[2]Center for Health Equity in Surgery and Anesthesia (CHESA), Institute of Global Health Sciences, University of California San Francisco, San Francisco, CA, United States

[3]Department of Surgery, Kiryandongo Regional Referral Hospital, Kigumba, Uganda

[4]Department of Public Health, Botswana-Havard Health Partnership, Gaborone, Botswana

[5]Department of Public Health Sciences, California Baptist University, Riverside, CA, United States

[6]Section for Oral Health, Heidelberg Institute of Global Health (HIGH), Heidelberg University, Heidelberg, Germany

[7]Neurocenter, Department of Neurosurgery and Turku Brain Injury Center, Turku University Hospital, Turku, Finland

[8]Division of Clinical Neurosciences, Neurosurgery, University of Turku, Turku, Finland

Corresponding Author:
Herman Lule, MBChB, MMED
Injury Epidemiology and Prevention (IEP) Research Group, Turku Brain Injury Centre, Division of Clinical Neurosciences, Turku University Hospital and University of Turku, Vähä Hämeenkatu 1B, Turun Yliopisto, Turku, Finland

Abstract

Background: Trauma-related mortality exhibits a marked social gradient, driven by access barriers and health inequities worldwide. These barriers jeopardize progress toward Sustainable Development Goals 3 and 4.

Objective: This study aimed to investigate the prehospital and in-hospital barriers to timely injury care as perceived by frontline trainee physicians and traffic law enforcement professionals during real-time treatment execution in Uganda. Additionally, we aimed to highlight the potential impact of these barriers on trauma outcomes.

Methods: This study used a convergent mixed methods approach. Qualitative data were collected through structured interviews and face-to-face focus groups with diverse teams of 500 frontline trainee physicians, including surgical residents, interns, medical students, and traffic law enforcement professionals. Directed content analyses for structured interviews and focus groups were conducted in NVivo (version 14, release 2023; QSR International). Quantitative data were concurrently collected using a survey questionnaire from the Motorcycle Trauma Outcome Registry, which included 1003 patients with trauma. We categorized barriers as prehospital or in-hospital barriers and as trauma team-related, patient-related, or health system–related barriers. Multilevel restricted maximum likelihood logistic regression analyses of factors associated with barriers to injury care were analyzed using Stata (version 15.0; StataCorp LLC). Odds ratios (ORs) and 95% CIs were reported; statistical significance was defined as $P<.05$.

Results: Qualitative analyses identified key prehospital barriers, including delays in emergency medical services activation, ambulance arrival, and transportation. In-hospital barriers were primarily shortage of supplies, delays in identifying life-threatening injuries, and insufficient critical care services. Quantitatively, among the 1003 audited patients with trauma, 42% (416/1003) faced barriers during treatment. The most common obstacles were delays in treatment decisions (232/1003, 23%) and securing necessary supplies (180/1003, 18%). The presence of barriers was independently associated with a 3-fold increased likelihood of unfavorable Glasgow Outcome Scale scores (OR 3.15, 95% CI 2.23 - 4.66; $P<.001$) for neurological injuries and was linked to a 4-fold increase in odds of 90-day mortality (OR 4.20, 95% CI 2.25 - 6.94; $P<.001$). After adjusting for injury severity and clustering effects by hospital teams and resources, the presence of barriers was associated with arrival by public means (adjusted OR [aOR] 1.62, 95% CI 1.09 - 2.41; $P=.02$), increasing age (aOR 1.01, 95% CI 1.00 - 1.03; $P=.01$), sustaining 1 or more injuries requiring admission (aOR 1.92, 95% CI 1.18 - 3.14; $P=.01$ vs aOR 3.69, 95% CI 1.95 - 6.98; $P<.001$), and a severe Kampala Trauma Score of ≤6 (aOR 2.71, 95% CI 1.37 - 5.37; $P=.004$).

Conclusions: Multiple barriers to trauma care are more frequent for severe injuries and are associated with poorer neurological outcomes and higher mortality. These findings indicate the need for targeted, multifaceted interventions that incorporate frontline

XSL•FO
RenderX

health workers' perspectives to improve trauma care delivery in low-resource settings facing both prehospital and in-hospital barriers.

International Registered Report Identifier (IRRID): RR2-10.2196/55297

(JMIR Hum Factors 2026;13:e84774) doi:10.2196/84774

KEYWORDS

emergency medical services; health care disparities; medical training; wounds and injuries; mixed methods

Introduction

Globally, injury-related mortality is characterized by significant socioeconomic gradients [1]. These social inequalities manifest through suboptimal prehospital care access, fragmented health care systems, and suboptimal postcrash care, including the unmet need for physical and mental rehabilitation after injuries, in both low- and high-income countries (HICs) [2]. For instance, research highlights substantial regional disparities in trauma-related mortality rates, particularly in Europe, where Eastern Europe experiences rates nearly 3 times higher than those in Western Europe [1,3]. In the United States, compelling evidence demonstrates desperate trauma outcomes that are linked to disparities in ethnicity and insurance coverage, where uninsured persons of low socioeconomic status are the most vulnerable to adverse injury outcomes, whereas in Australia, rural geographic disparities to injury care are well documented [4]. Research from Canada mandates that issues of health care inequities be addressed by medical trainees and well-established physicians through their civic duties as patients' advocates [5].

In Asia and Africa, the burden of injury is disproportionately higher in nations with poor economies, contributing to approximately 4.5 million premature deaths annually in low- and middle-income countries (LMICs) [3,6,7]. For instance, it is estimated that 36% of injury-related deaths in LMICs are linked to delays or barriers to injury care [8]. These barriers are highlighted in the United Nations General Assembly resolution 74/299, with a rapidly approaching target deadline: to halve road traffic crash–related global mortality and ensure well-being for all individuals by 2030, congruent to Sustainable Development Goals (SDG3) [9]. However, whereas recent trends show declining inequalities in injury care access among HICs, in LMICs, socioeconomic inequities, inadequate infrastructure, and shortage of human resources for health have continued to exacerbate these challenges, resulting in disparities in access to injury care both between and within regions, for instance, in urban versus rural areas [3,6,10].

In Africa, the disparities in injury care are manifested by suboptimal emergency medical services coverage in most countries in the region [11,12]. For instance, despite the high volume of trauma cases in Uganda and Rwanda, significant challenges persist, including inadequate insurance coverage and limited resources for trauma care. These challenges also encompass a weak prehospital care system characterized by shortage of ground ambulances and evacuation of injured patients predominantly by traffic police, a scarcity of trauma specialists, and the constraints of rigid, nondecentralized referral systems [13,14]. These issues not only impact injury outcomes but also have consequences on the learning and clinical experiences of medical students and resident doctors who train in such environments, including burnouts and poor retention of the health care workforce, after graduation from medical schools [15]. Medical training and capacity building have been identified as top priorities for addressing African health system challenges [15]; however, research that positions trainee physicians as policy advocates for improving the social determinants of health at individual and practice levels, particularly through identifying barriers that impact the care of patients in their regular work environment, is limited.

Positioning medical trainees as patient advocates is essential for building a competent health workforce, particularly in LMICs where resource constraints worsen health disparities and create imbalances in geographic and specialty distributions. Current academic programs in Uganda focus primarily on biomedical education, failing to connect medical training with community-based interprofessional practice, social determinants of health, and population needs. These connections are crucial for preparing future physicians to assume progressive responsibilities aligned with community priorities. Systematic reviews indicate that integrating experiential learning, reflective practice, and feedback in education fosters long-term professional identity formation, empowering trainees to influence local and national health policies, which is vital for addressing the dual burden of disease and workforce deficits in LMICs [16,17].

We previously reported on a Motorcycle Trauma Outcome Registry (MOTOR) project that considered medical trainees and traffic law enforcement first responders as sustainable human resources for injury care in rural settings with a positive impact on injury outcomes through trauma team training [18,19]. This ancillary study formed an integral component of tertiary outcomes to supplement the MOTOR project by detailing the perspectives of these cadres regarding barriers to injury care in low-resourced contexts.

Coping with patient care demands, sufficient health system resources, and an enabling work environment are essential for the success of medical trainees and frontline health workers in their learning and patient care responsibilities. These demands intensify when caring for patients with trauma due to the emergency and life-threatening nature of their injuries. Therefore, it is crucial to understand the barriers that may hinder effective treatment of such patients from the perspectives of trainee physicians, who serve as frontline health workers, and from traffic law enforcement professionals who work in the capacity of paramedics in low-resource settings. However, existing studies often lack a comprehensive analysis of the barriers to injury care faced at the individual patient level, which

XSL•FO
RenderX

are influenced by geographical and cultural contexts [13,14,20]. In Uganda, studies that have attempted to investigate some of these barriers have limited their scope to prehospital factors mainly in urban settings and have excluded insights from frontline health care workers such as trainee physicians who predominate the health system workforce [21,22]. This phenomenon can be attributed, in part, to the challenges associated with acquiring information from key frontline stakeholders. These challenges arise from their demanding work schedules, remote geographical locations, privacy concerns, and the aspiration to manage the heterogeneity of research participants, which often proves to be realistically unattainable.

Therefore, the aim of this study was to identify the prehospital and in-hospital barriers to timely injury care as perceived by trainee physicians and traffic law enforcement professionals who are the first point of contact and are actively engaged in the care of injured patients in a low-resource context, as well as to elucidate how these barriers impact injury outcomes. We structured the identified barriers into 2 frameworks, focusing on both prehospital and in-hospital challenges, while examining both infrastructural and human resource factors. Additionally, we aimed to delineate the specific barriers encountered at both the individual-patient and trauma-team levels during the real-time execution of treatment.

Methods

Study Design

This was a convergent mixed methods study nested within a broader MOTOR project, which investigated the impact of rural trauma teams and training on outcomes of motorcycle-related injuries [18]. To effectively address the complex interconnected geopolitical, infrastructural, and human resources issues that delay the execution of lifesaving interventions, mixed methods research was essential to fully understand the specific barriers to timely injury care in low-resource settings, particularly from the perspectives of health care personnel involved in the initial patient contact. Prior research had identified mixed methods as particularly suitable for investigating complex health care systems [23]. In their recent systematic review, Whitaker and colleagues [20] have recommended that future research assessing barriers to injury care should use innovative mixed methods to investigate both prehospital and in-hospital delays, as well as the associated care processes and outcomes. In this study, we fitted a 2-framework approach into the concept of 3 temporal delays in accessing treatment within the context of LMICs, including seeking, reaching, and receiving injury care, as described by Vaca et al [24]. This theoretical framework permitted the convergence of qualitative and quantitative data to gain a deeper understanding of the obstacles to injury care and the characteristics of patients who experience these barriers. The research was documented in accordance with good reporting of mixed methods studies guidelines [25], whereas the qualitative element was reported in line with the COREQ (Consolidated Criteria for Reporting Qualitative Research) [26].

Study Settings

The study was conducted from September 2, 2019, to August 27, 2023, at 6 regional tertiary hospitals in Uganda: Jinja, Hoima, Fort Portal, Mubende, Kiryandongo, and Kampala International University Teaching Hospital, which also serve as surgical residency and internship training centers. Multidisciplinary teams of consultant surgeons, residents, interns, and medical students operate the trauma and casualty units. Patients with trauma typically arrive via ground ambulances, traffic police patrols, or public transport, where they are assessed and treated by the multidisciplinary teams. These facilities operate as level III trauma centers and collectively serve a population of 18 million across 50 districts and 3 major cities in Uganda.

Study Population and Source of Data

Qualitative data were collected from surgical residents, interns (doctors or nurses), and third- and fifth-year medical students associated with the surgery departments of the participating hospitals. Additionally, we engaged law enforcement professionals specializing in road traffic regulations, who serve as first responders to road traffic–related injuries within the municipalities that constituted the catchment areas of these hospitals. The qualitative data were collected through investigator-administered structured interviews and audio-recorded face-to-face focus groups. The structured interviews had coded predetermined themes regarding barriers to injury care identified from previous literature and featured open-ended options to cater for emerging new themes [8,20]. These interviews were administered as part of a pretest conducted prior to the development training of rural trauma teams previously described for the frontline health workers [27,28]. On the other hand, quantitative data were gathered from investigator-administered survey questionnaires for patients who sustained motorcycle-related injuries, received care at any of the participating hospitals during the study period, and participated in the MOTOR project [18,29].

Sample Size Estimation and Sampling Methods

Given the complexity of the research topic and the cultural, geographical, and professional diversity of our study population, we aimed to recruit 50-80 participants for structured interviews from each of the 6 geographically distinct hospitals. We consecutively recruited samples of surgery residents, intern physicians, nurses, medical students, and purposively selected traffic rescue police involved in immediate contact with injured individuals. Participants were contacted from those who had opted into the rural trauma team development training project to complete the 15-minute structured interviews [18]. Participants who completed the structured interviews were subsequently recruited for 40-minute focus groups, organized into 100 teams. Each focus group comprised 5 participants and was managed by 3 research team members: HL administered the interviews, MM recorded field notes, and BO moderated the group sessions, verbalizing new responses verbatim. The groups consisted of varying compositions: rescue police alone (N=20), medical trainees alone (N=20), or a mix of both (N=60). Each participant was recruited only once. This data triangulation approach aimed to further explore the context of emerging themes through follow-up questions and to stimulate new themes based on team dynamics, aligning with an initial strategy to incorporate these groups into rural trauma teams [28].

XSL•FO
RenderX

To minimize institutional and selection bias, a transparent eligibility criterion was established. Trainee participants were stratified by hospital resources capacity and by study year as a rough estimate of expertise, followed by random assignment into 5-member groups using a blinded lottery method. Additionally, bias was further mitigated by triangulating data from focus groups and interviews and by continuously assessing data saturation to ensure a diversity of experiences.

We assumed data saturation and ceased recruitment when no new information, themes, or insights emerged in the monitoring log after 3 consecutive interviews and across different groups at various study sites, in accordance with previous studies [30,31]. Theme saturation was achieved after conducting 15-17 focus groups at each of the 6 hospitals, totaling 100 groups with 434 medical professionals and 66 traffic law enforcement officers (total sample=500). Data collection was capped upon reaching saturation to optimize time and resources, thereby avoiding redundancy while ensuring the comprehensiveness of emerging themes from diverse professional contexts [32]. Systematic reviews indicate that depending on the complexity of the research question and diversity of the research population, 9-17 interviews are sufficient for qualitative studies involving homogeneous populations, while a minimum of 50 interviews may be appropriate for culturally diverse populations [30,31].

The quantitative component assessed barriers to injury care for 1003 patients with trauma treated at the 6 study sites. The quantitative sample size was derived from an open-source R Shiny application for the MOTOR trial, which was a cluster randomized, parallel controlled trial with discrete time decay correlation structures for multiple periods as described by Lule et al [18].

Eligibility Criteria

The qualitative component included trauma care frontliners, such as chief surgical residents, medical interns, traffic police officers stationed on major highways for immediate road traffic accident responses, and medical students in surgical and traumatology clinical rotations. Medical participants without prior exposure to surgery or emergency department rotations were excluded, as they were presumed to lack experiences with patients with trauma and the associated barriers to care. Additionally, participants who anticipated being unable to complete both the structured interviews and focus groups due to commitments, such as police officers on duty, were also excluded to ensure comprehensive capture of each participant's experiences regarding the studied phenomenon.

The quantitative component included patients aged 2-80 years who had sustained neurological or musculoskeletal injuries resulting from a motorcycle crash and attended the emergency departments of the 6 study sites within 24 hours during the study period. To minimize confounding from injury mechanisms and comorbidities affecting neurocognitive and musculoskeletal outcomes, we excluded participants with a documented history of stroke and neuropsychiatric disorders, as outlined in the study protocols [18,27].

Data Collection

Data were collected and precoded by global surgery and doctoral researchers HL and MM according to themes identified in the existing literature [20]. The data collection process was conducted simultaneously using pretested tools [18,28,29]. The qualitative component was part of a broader face-to-face patient-public consultative engagement for the MOTOR project [18]. After explaining the research objective and the procedure for completing the structured interviews, we asked participants to identify, assess, and rank the barriers. Subsequently, we used focus groups to gain an in-depth understanding through follow-up questions, to probe novel barrier themes, and to fully capture the experiences of participants (Multimedia Appendix 1). This information was intended to inform the establishment of emergency medical services, trauma teams, and trauma registries focused on motorcycle crash–related neurological and musculoskeletal injuries, which represent Uganda's most significant trauma burden [33].

In the quantitative component, a prospective MOTOR trial project was established at 6 regional referral hospitals in Uganda [18,19], which captured quantitative data as detailed in Multimedia Appendix 2. Consultants, surgical residents, and interns involved in trauma care documented barriers encountered during the execution of definitive injury care for 1003 patients with trauma (MOTOR participants). Data on demographics, injury severity based on the Glasgow Coma Scale (GCS), and the Kampala Trauma Score (KTS) [34] were collected at baseline. On the other hand, trauma outcomes, including all-cause mortality, the Glasgow Outcome Scale (GOS) for traumatic brain injuries [35], and the Trauma Outcome Measure Scale (TOMS) score [36] for musculoskeletal injuries, were recorded 90 days postinjury.

Quantitative data were captured using precoded paper-based questionnaires and later uploaded to the secure REDCap (Research Electronic Data Capture; Vanderbilt University) electronic system hosted by Turku University in Finland [37]. Information was captured if a barrier was encountered during the execution of treatment, and each documented barrier triggered an alert, prompting an audit of the patient's case file during weekly virtual meetings via Zoom (Zoom Communications, Inc). These 15- to 30-minute meetings, summarized in field notes, clarified the nature of the barriers and included verbatim anonymous documentation of responses from research assistants (designated 2-year experienced medical officers) and attending clinicians. Corrections were made for any inaccurate entries identified during the audits. This comprehensive approach aimed to capture a nuanced understanding of barriers affecting injury care access in real time to enhance future interventions.

Data Analysis

We analyzed data from each component of this mixed methods study independently. The basis of this approach was to enrich the interpretation of the main outcomes of the MOTOR trial project in a broader context.

XSL•FO
RenderX

Qualitative Analyses

Directed content analysis for both structured interviews and focus groups was conducted in NVivo (version 14, release 2023; QSR International). We used an iterative approach to track and document the prevalence of emerging themes using a log of nodes in NVivo. The results from primary qualitative data for both structured interviews and focus groups were consistently noncontradictory; thus, the overall triangulated results from these 2 qualitative approaches were presented. Where applicable, verbatim responses were presented as quotes to provide further context from open-ended questions.

Quantitative Analyses

Concerning prehospital barriers (delay 1), participants selected from 6 responses. Provided with scores ranging from a minimum of 1 to a maximum of 6, participants ranked the barriers using a 6-point Likert scale, where 1 indicated "not a barrier," 2 "very weak barrier," 3 "weak barrier," 4 "averagely weak barrier," 5 "strong barrier," and 6 "extraordinarily strong barrier." A barrier category was considered significant (major theme) if it attained an overall mean score exceeding the prespecified hypothesized average of 3.5; otherwise, it was classified as a minor theme.

For in-hospital team barriers (delay 2) and system barriers (delay 3), there were 4 potential responses presented to medical participants, with Likert scores ranging from 1 ("least important barrier") to 4 ("very important barrier"). While initial qualitative data collection targeted emergent themes, Likert items were incorporated with open-ended questions in a mixed methods design to enrich qualitative inquiry and facilitate triangulation with quantitative data. Predetermined Likert response categories (eg, "not a barrier" to "very important barrier") quantified complex qualitative constructs such as attitudes, perceptions, and preferences. These response patterns informed purposive sampling for follow-up qualitative interviews across demographic groups and clinical experience levels. This approach bridged quantitative precision and qualitative depth by elucidating reasons behind each of the numeric Likert choices. The barrier category was considered important (major theme) if it surpassed the prespecified hypothesized average of 2.5.

The overall ranking of barriers was determined by calculating weighted mean scores, derived from the sum of individual Likert component scores multiplied by the number of responses for each component, subsequently divided by the valid sample size. This methodology has been previously validated for assessing barriers to trauma care in LMICs [20]. For example, in relation to prehospital delays, responses were scored on a Likert scale as follows: 1 (not a barrier), 2 (very weak barrier), 3 (weak barrier), 4 (averagely weak barrier), 5 (strong barrier), and 6 (extraordinarily strong barrier).

$$\text{Weighted mean barrier score} = [1(n1)+2(n2)+3(n3)+4(n4)+5(n5)+6(n6)] \div N$$

where n_1=number of people responding per component, and N=sample size.

To analyze data from the MOTOR, first, we dichotomized the responses into 2 categories, that is, whether a barrier was present (yes or no). Second, whenever a barrier was identified, responses were further categorized into three categories: (1) team-related barriers that lead to delays in the identification, prioritization, or timely referral of life-threatening injuries by the emergency teams; (2) individual barriers, either from patients or from clinicians, that contributed to delays in decisions regarding surgery, initiation of treatment, or referral of injuries exceeding local capacity; and (3) in-hospital system barriers that caused delays in securing essential supplies, including oxygen, anesthetics, sutures, and blood products, or delayed access to a functional laboratory, imaging diagnostics, theater space, or intensive critical care services. These categories were subsequently represented as proportions based on their relative contribution to the total ranks. The mean rank scores for each barrier across various cadres (third- and fifth-year medical students, intern, and police) were compared with the hypothesized mean using the nonparametric Kruskal-Wallis equality of proportions rank test.

Finally, to determine factors associated with treatment barriers and their impact on injury outcomes, we assessed normality with the Levene test and performed bivariate and multilevel restricted maximum likelihood mixed-effects logistic regression analyses using Stata (version 15.0; StataCorp LLC) after adjusting for clustering effects. The mvaghermite integration method was used to explore potential correlations between barriers and injury outcomes, including all-cause mortality and morbidity at 90 days postinjury, as measured by the GOS [35] and TOMS [36]. Confounding and effect modification were evaluated using Cochran-Mantel-Haenszel statistics. Variables demonstrating a P value of <.2 in the Mantel-Haenszel strata and those with biological relevance to injury severity were included in the final model.

Recalling that the severity of injury is an important determinant of its outcome, the effects of barriers on injury outcomes were stratified by injury severity at baseline based on the GCS and KTS II for neurological and musculoskeletal injuries, respectively [34]. The selected outcome tools were chosen due to the primary focus of the MOTOR trial project on neurological and musculoskeletal trauma [18], which represent the most significant injuries resulting from motorcycle crashes [33]. For neurological injuries, outcomes were dichotomized as favorable (if GOS score 4 - 5) or unfavorable (if GOS score 1 - 3), whereas for musculoskeletal injuries, outcomes were dichotomized as favorable (if TOMS equaled or exceeded initial Trauma Expectation Factor Score) or unfavorable (if TOMS was less than Trauma Expectation Factor Score), congruent to previous studies [35,36]. A detailed description of how these tools were used to evaluate the morbidity associated with injuries, including assessments for pain, physical function, disability, satisfaction with treatment, and overall quality of life after injury, is reported in the MOTOR trial protocol [18].

Convergence

Qualitative responses were analyzed in conjunction with quantitative metrics for each construct by 2 independent researchers (HL and MM). Any contradictions identified during this analysis were addressed by a third researcher (BO). The convergence status was categorized as "convergent" if the

XSL•FO
RenderX

qualitative findings aligned with the quantitative results, "complementary" if the combined outcomes were assessed as synergistic, and "divergent" if the 2 results were inconsistent. Divergence indicated contradiction, as reflected in the notation (Qualitative≠Quantitative). In contrast, complementarity examined the different aspects of a phenomenon from multiple perspectives, leveraging the strengths of each method to cover more content than either method alone, thereby providing fuller, richer, multidimensional, yet nonoverlapping insight (Quantitative+Qualitative). Conversely, the convergent approach required consistency of outcomes across methods (Qualitative=Quantitative).

Ethical Considerations

For this research involving human participants, all methods were performed in accordance with the ethical standards stipulated in the declaration of Helsinki and its later amendments, and in accordance with the relevant local guidelines and regulations. Ethical approval was obtained from the Research and Ethics Committees of Mbarara University of Science and Technology (reference: 1/7; 05/5 - 9) and the Uganda National Council for Science and Technology (reference: SS 5082). All participants or their legally authorized representatives provided written informed consent prior to enrollment into the study. We ensured confidentiality and anonymity of participants through the use of codes and generic pseudonyms, such as "doctor," "nurse," or "police" in the reports. Each participant was compensated with transport reimbursement of US $7.50 as approved by the research and ethics committee.

Results

Qualitative Findings

Of the 500 trauma care frontliners who participated in structured interviews and focus groups for qualitative analysis (100% response rate), 27% (135/500) were females, 73% (365/500) were males, 13% (66/500) were traffic law enforcement officers, and 87% (434/500) were medical trainees.

Prehospital Barriers (Delay 1)

The 3 highest ranked barriers to accessing injury care identified in the assessment of prehospital factors were as follows: delays in the activation and mobilization of emergency medical services (105/500, 21%); delays attributable to prehospital response times, including the arrival of ambulances at the scene (95/500, 19%); and prolonged transportation times to the hospital (75/500, 15%) (Figure 1).

We found that while the police evacuated the patients with trauma, there were no designated rural networks of emergency rescue police to intervene for traffic crash victims.

> *The role of evacuation of traffic crash victims is not specifically designated to particular officers within our jurisdiction which means any road traffic officer on duty can oversee such cases.* [Participant no. 2, traffic law enforcement officials]

In addition, limited human resource capacity was reported as a key obstacle for the law enforcement professionals to participate in prehospital evacuation of injured patients.

> *Often we are understaffed and not able to undertake transfer responsibilities timely as this would mean leaving one's station vacant which would not be acceptable in our line of duty.* [Participant no. 36, law enforcement official]

Furthermore, a lack of key focal people concerned with coordinating emergencies at recipient hospitals and suboptimal communication technologies was cited as a key barrier to timely injury care.

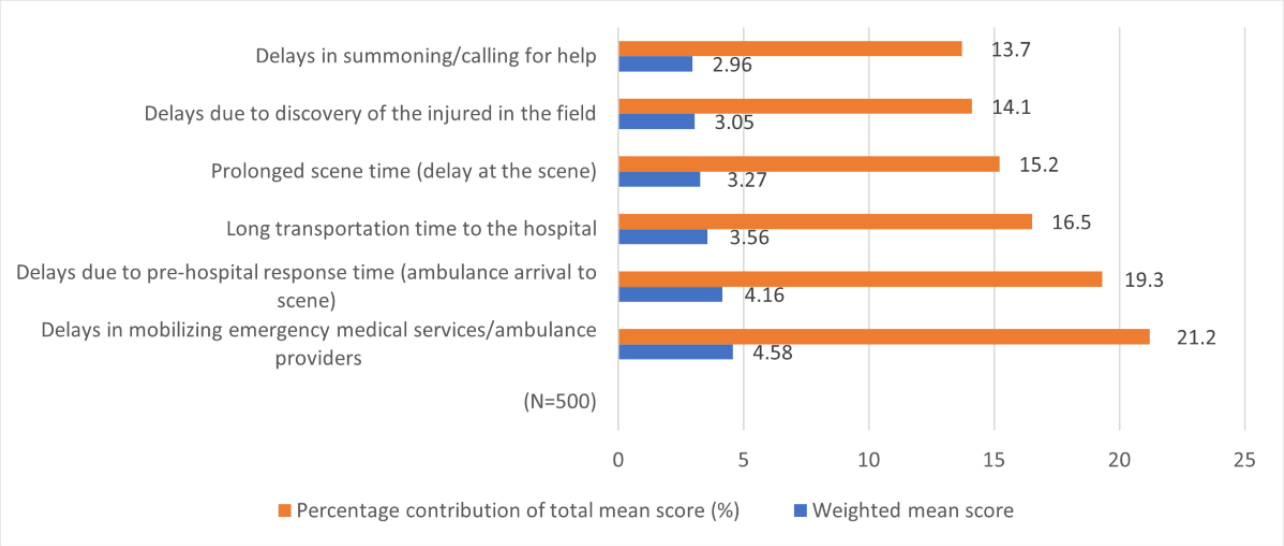
> *Of course, you want to communicate timely but the poor communication technologies and network breakdown will let you down. Our intercom is not linked to that of the hospital emergency departments unless one is chanced to have physical contact there. Nevertheless, we are required to drop them (patients) by the government emergency hospital departments.* [Participant no. 11, traffic police commander]

We established that there were also concerns of bias and mistrust that led to nondisclosure of injury-related information, which could delay the emergency evacuation process and insurance compensation claims.

> *We would be delighted to have an integration of emergency evacuation rescue services that is well coordinated along with the health ministry, I guess this would increase our public trust. You want to help evacuate these accident victims, but they think you are interrogating, arresting, and prosecuting them for their traffic crimes.* [Participant no. 47, police officer]

XSL•FO
RenderX

Figure 1. Overall weighted mean scores for prehospital barriers (delay 1) ranked from lowest to highest.



In-Hospital Trauma Team and Individual-Patient Barriers (Delay 2)

The highest ranked in-hospital team barriers were delays in identification of life-threatening injuries (117/434, 27%) and in executing referrals to higher trauma centers (113/434, 26%) (Figure 2).

We found that significant financial limitations prevented patients from obtaining emergency laboratory or imaging investigations that required out-of-pocket payments at private facilities, contributing to unnecessary delays. Additionally, there were no clear referral protocols linking the public to private facilities for specialized investigations or interventions that were not available in public facilities.

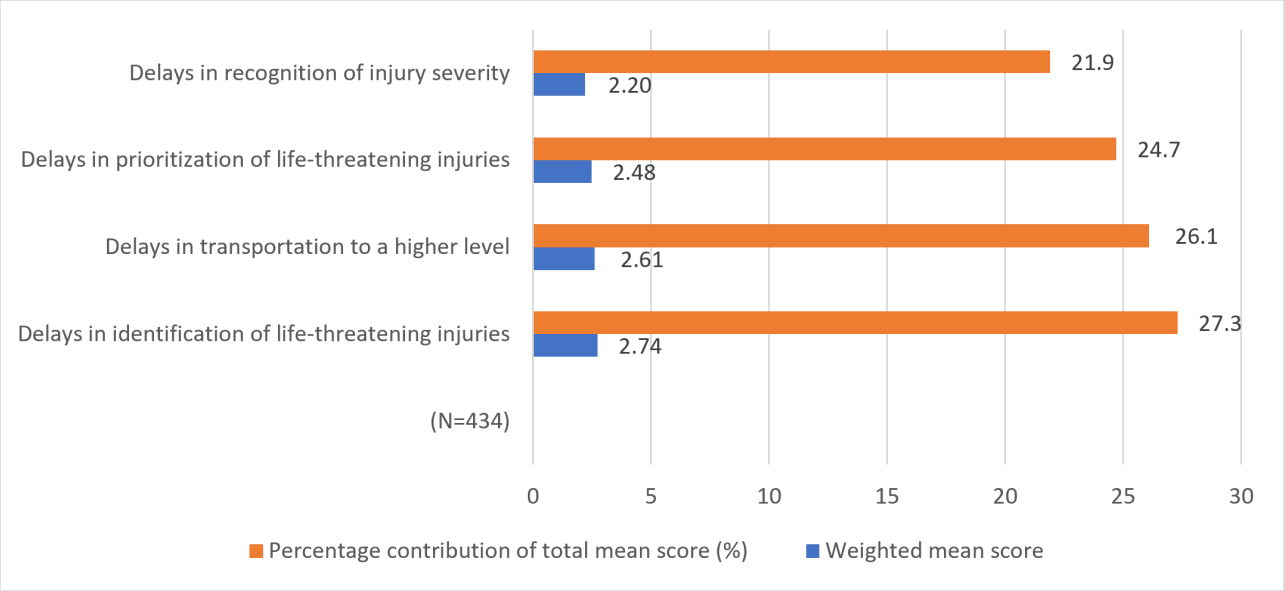
> *We wanted to obtain head and brain CT scans, but when the service in our government national referral hospital was down, the patient's next of kin delayed mobilizing financial resources to outsource one from private. We do not have direct referral pathways to private hospitals so then we had to wait until patients' relatives availed us with the results of the investigations before operating.* [Expert participant no. 1, attending neurosurgeon]

The trauma teams from qualitative clinical audits raised concerns about delays in hospital arrivals due to reliance on public transportation and dysfunctional ambulance systems. Additionally, there was limited capacity among emergency medical services staff and critical care supplies for life-threatening injuries, compounded by suboptimal health insurance coverage and communication barriers with the families of the injured individuals.

> *He arrived late in a taxi as there was no fuel in the ambulance from the referring hospital. The staff there say were understaffed and were unable to accompany the patient as no one would cover their duty. They could have placed a chest tube though, but they were out of stock. Our ICU is full though, so we plan to refer him to a private hospital after a chest tube, but we are facing difficulties with direct communication, he has no attendants available yet as his family is still mobilizing resources for ICU costs. He does not seem to be insured.* [Expert participant no. 3, attending anesthesiologist]

XSL•FO
RenderX

Figure 2. Overall mean weighted scores for in-hospital trauma team barriers (delay 2).



In-Hospital System Barriers (Delay 3)

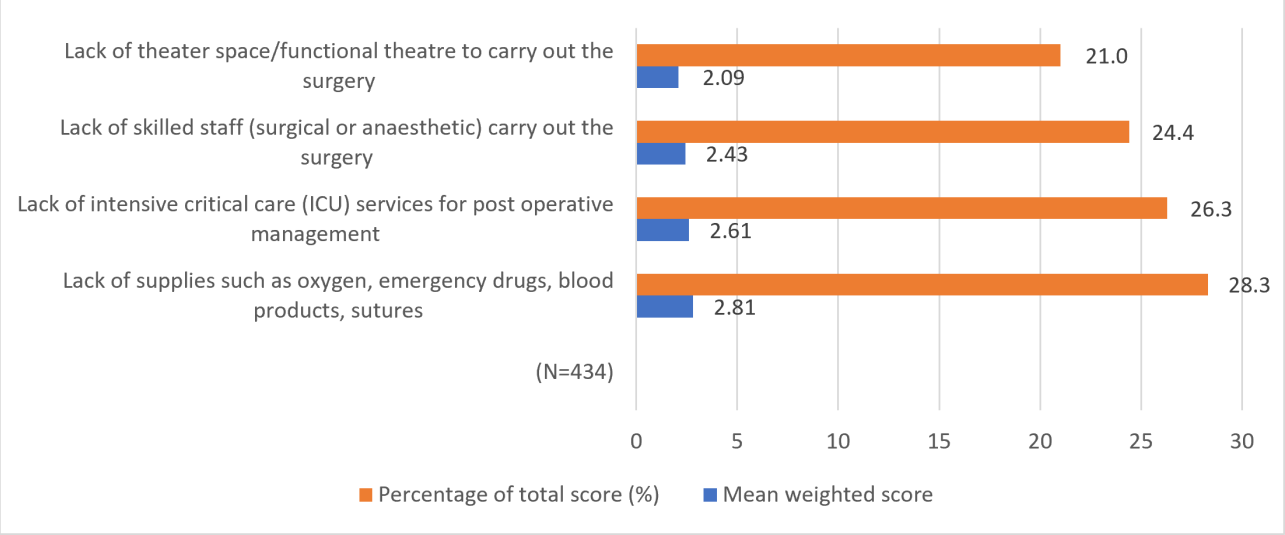
The highest ranked in-hospital system barriers to injury care were a shortage of emergency medical supplies (123/434, 28%) and suboptimal critical care services (114/434, 26%) (Figure 3).

Qualitative clinical audits revealed a limited range of laboratory diagnostic services available in public facilities, contributing to delays in treatment decisions. Additionally, a general shortage of essential medical and surgical supplies for both pre- and postoperative care, including pharmaceuticals and blood products, was noted.

> *My role as an intern was to stabilize the patient before theatre once the decision to operate was made by the consultant. It frustrated me when I could not get the lab results and blood products on time because of having no reagents, yet my boss was on my neck. I was waiting for the relatives to bring the results from private.* [Participant no. 133, intern doctor]

> *He was supposed to get labetalol for his intracranial pressure before brain surgery; I am not sure how far his relatives have gone with lobbying for that. The orthopedic team also says he has not yet bought the implants for his fractures. There were social issues there, but I think the team will gather when the resources are available.* [Participant no. 154, surgery resident]

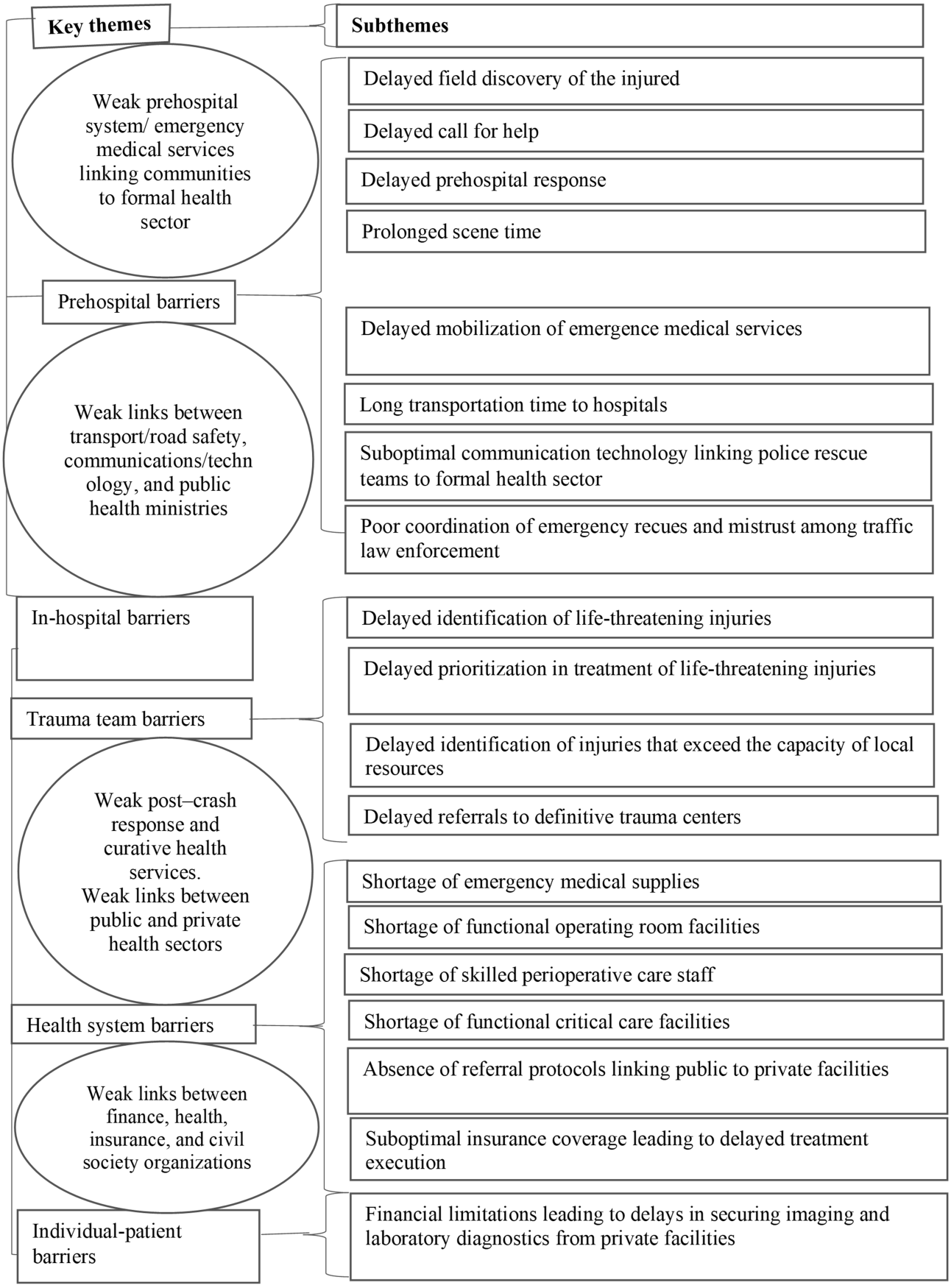
Figure 3. Mean weighted scores for in-hospital barriers after decision to operate (delay 3).



A summary of key themes and subthemes derived from the qualitative analyses regarding barriers to injury care is presented in Figure 4.

XSL•FO
RenderX

Figure 4. Thematic map summarizing themes and subthemes emerging from qualitative analyses of barriers to injury care.



XSL•FO
RenderX

Quantitative Findings

Overview

Kruskal-Wallis test showed statistically significant differences in how the various cadres ranked the various barriers (Multimedia Appendix 3). Of the 1003 patients with trauma in the MOTOR for quantitative analysis, there were 82% (817/1003) males and 19% (186/1003) females. The overall median age (IQR) was 28 (22-37) years. The median (IQR) commute distance was 5 (2-10) km, and median (IQR) prehospital time was 2 (1-3) hours from accident scene to the emergency departments. Neurotrauma was prevalent in 95% (949/1003), whereas musculoskeletal injuries occurred in 80% (799/1003) of participants. A majority (745/1003, 74%) of participants sustained polytrauma. The median (IQR) baseline injury severity scores–based GCS and KTS were 14 (11-15) and 8 (7-9), respectively.

Characteristics of Patients Who Encounter Barriers to Injury Care

We analyzed 90-day complete follow-up data for 88% (887/1003) of individuals with traumatic brain injuries, and for 80% (799/1003) of individuals with musculoskeletal injuries. The characteristics of those who experienced barriers are summarized in Table 1. This group included patients who were critically ill, with 64% (272/426) presenting with a moderate to severe KTS and 45% (142/426) exhibiting moderate to severe GCS scores. Furthermore, 31% (133/426) of these individuals sustained multiple injuries that necessitated admission. For individuals with traumatic brain injuries, the proportion of unfavorable outcomes was significantly higher among those who reported barriers (84/366, 23% compared with 45/521, 9%), with χ^2_1 (N=887)=35.44 and P<.001. In contrast, for those with musculoskeletal injuries, there was no statistically significant difference in the proportions of unfavorable trauma outcomes between individuals who encountered barriers and those who did not (74/311, 24% vs 121/494, 25%), as indicated by χ^2_1 (N=805)=0.05 and P=.82. However, the mortality rate was higher among individuals who reported barriers to injury care than among those with no such barriers (55/366, 15% vs 27/521, 5%), with χ^2_1 (N=887)=24.83 and P<.001, as shown in Table 1.

XSL•FO
RenderX

Table . Characteristics of patients with trauma who experienced barriers to injury care.[a]

Variable	Barrier to receiving injury care reported		Total	Chi-square (*df*)	*P* value
	No	Yes			
Age (years), median (IQR)	22 (27 - 35)	22 (30 - 40)	28 (22 - 37)	N/A[b]	<.001
Sex, n (%)				0.19 (1)	.62
Male	767 (81.6)	50 (79.4)	817 (81.5)		
Female	173 (18.4)	13 (20.6)	186 (18.5)		
Commute distance from accident scene to emergency department (km), median (IQR)	5 (2-10)	5 (2-10)	5 (2-10)	N/A	.83
Musculoskeletal injuries present?, n (%)				6.98 (1)	.008
No	134 (23.2)	70 (16.4)	204 (20.3)		
Yes	443 (76.8)	356 (83.6)	799 (79.7)		
Neurological injuries present?, n (%)				0.30 (1)	.58
No	33 (5.7)	21 (4.9)	54 (5.4)		
Yes	544 (94.3)	405 (95.1)	949 (94.6)		
Respiratory rate at admission (breaths per minute), n (%)				54.76 (2)	<.001
≤9	3 (0.7)	5 (0.9)	8 (0.8)		
10 - 29	292 (70.2)	519 (88.4)	811 (80.9)		
≥30	121 (29.1)	63 (10.7)	184 (18.3)		
Systolic BP[c] at admission (mm Hg), n (%)				21.73 (2)	<.001
≤49	2 (0.5)	0 (0)	2 (0.2)		
50 - 89	100 (24.0)	79 (13.5)	179 (17.8)		
≥90	314 (75.5)	508 (86.5)	822 (82.0)		
Accident scene to emergency department admission interval (hours), median (IQR)	1.8 (1.0 - 3.0)	2.0 (1.0 - 3.0)	2.0 (1.0 - 3.0)	N/A	.88
Referral decision to dispatch interval (hours), median (IQR)	2.5 (1.7 - 3.8)	3.0 (2.0 - 4.0)	2.8 (1.8 - 3.9)	N/A	.78
Number of serious injuries requiring an admission, n (%)				70.84 (2)	<.001
>1	78 (13.3)	133 (34.4)	221 (22.0)		
1	390 (66.4)	233 (56.0)	623 (62.1)		
None	119 (20.3)	40 (9.6)	159 (15.9)		
KTS[d], median (IQR)	8 (6-9)	9 (8-9)	8 (7-9)	N/A	<.001
KTS categories, n (%)				42.05 (1)	<.001
Mild (scores 9 - 10)	328 (56.9)	154 (36.2)	482 (48.1)		
Moderate (scores 7 - 8) or severe (score ≤6)	249 (43.1)	272 (63.9)	521 (51.91)		
GCS[e], median (IQR)	13 (10 - 15)	15 (12 - 15)	14 (11 - 15)	N/A	<.001
GCS categories, n (%)				24.10 (1)	<.001
Mild (scores 13 - 15)	404 (70.0)	234 (54.9)	638 (63.6)		

XSL•FO
RenderX

Variable	Barrier to receiving injury care reported		Total	Chi-square (*df*)	*P* value
	No	Yes			
Moderate (scores 9 - 12) or severe (score ≤8)	173 (30.0)	192 (45.1)	365 (36.4)		
GOS[f], median (IQR)	4 (4-5)	5 (4-5)	5 (4-5)	N/A	<.001
GOS categories, n (%)				35.44 (1)	<.001
Favorable outcome (scores 4 - 5)	476 (91.4)	282 (77.1)	758 (85.5)		
Unfavorable outcome (scores 1 - 3)	45 (8.6)	84 (23.0)	129 (14.5)		
TOMS[g] categories, n (%)				0.05 (1)	.82
Favorable	373 (75.5)	237 (76.2)	610 (75.8)		
Unfavorable	121 (24.5)	74 (23.8)	195 (24.2)		
Injury outcome at 90 days, n (%)				24.83 (1)	<.001
Survived	494 (94.8)	311 (85.0)	805 (90.8)		
Died	27 (5.2)	55 (15.0)	82 (9.2)		

[a]Level of statistical significance (*P*=.05). Data are n (%) unless specified otherwise. The characteristics of all those included in analyses are demonstrated.

[b]Not applicable.

[c]BP: blood pressure.

[d]KTS: Kampala Trauma Score.

[e]GCS: Glasgow Coma Scale.

[f]GOS: Glasgow Outcome Scale.

[g]TOMS: Trauma Outcome Measure Scale score.

In-Hospital Audit of Nature of Common Barriers to Injury Care for the Patient Participants (Delay 3)

From 1003 audited trauma cases, only 59% (587/1003) received treatment for their injuries without their clinical teams encountering any barriers in the provision of definitive care. Hospitals whose trauma teams received rural trauma team development training identified more barriers than their counterparts (256/501, 51% vs 170/502, 34%; *P*<.001). The proportion of patients who faced 1 or more barriers was 23.1% (232/1003) at the individual (patient) level, 6.3% (63/1003) at trauma-team level, and 17.9% (180/1003) at hospital-system level. Team-level barriers referred to obstacles that resulted in delays in the activation of emergency teams, as well as in the identification, prioritization, recognition, or timely referral of life-threatening injuries. Individual-level barriers were those that impeded timely decisions regarding the initiation of treatment or the referral of injuries that exceed local capacity. In contrast, in-hospital system–level barriers were systemic obstacles that caused delays in obtaining essential supplies, such as oxygen, anesthetics, sutures, blood products, and diagnostics, as well as in securing access to theater space and intensive or critical care services.

Factors Associated With Health Care Access Barriers and Their Impact on Injury Outcomes

The results of factors associated with barriers to injury care are presented in Table 2. After adjusting for clustering effects by hospital teams and resources, mixed-effects regression analyses indicated that the factors independently associated with barriers to injury care included increasing age, involvement in motorcycle-car crashes, the multiplicity of injuries requiring admission, the presence of moderate or severe injuries as assessed by both the KTS and the GCS, the occurrence of axial head lesions, and the necessity for craniotomy.

Further analysis demonstrated that the presence of barriers was independently associated with unfavorable GOS scores (odds ratio [OR] 3.15, 95% CI 2.23 - 4.66; *P*<.001) for neurological injuries and was linked to a 4-fold increase in odds of 90-day mortality (OR 4.20, 95% CI 2.25 - 6.94; *P*<.001). However, barriers were not significantly associated with unfavorable TOMS scores for musculoskeletal injuries.

Multilevel analyses identified key factors associated with barriers, including arrival by public means (adjusted OR [aOR] 1.62, 95% CI 1.09 - 2.41; *P*=.02), increasing age (aOR 1.01, 95% CI 1.00 - 1.03; *P*=.01), and sustaining 1 or more injuries requiring admission (aOR 1.92, 95% CI 1.18 - 3.14; *P*=.01 vs aOR 3.69, 95% CI 1.95 - 6.98; *P*<.001), as well as a severe KTS score of ≤6 (aOR 2.71, 95% CI 1.37 - 5.37; *P*=.004). Conversely, a negative head CT result and skull fractures were associated with lower odds of encountering barriers in both bivariate and multilevel analyses as shown in (Table 2).

XSL•FO
RenderX

Table . Multilevel mixed-effects logistic regression analyses of factors associated with barriers to injury care.[a]

Variable	Crude odds ratio	95% CI	*P* value	Adjusted odds ratio	95% CI	*P* value
Age (years)	1.018	1.008-1.027	<.001	1.014	1.003-1.025	.01
Sex						
Female	Reference	N/A[b]	N/A	N/A	N/A	N/A
Male	0.900	0.650-1.247	.53	N/A	N/A	N/A
Injury mechanism						
Motorcycle-motorcycle crash	Reference	N/A	N/A	N/A	N/A	N/A
Motorcycle-pedestrian crash	1.051	0.761-1.450	.76	N/A	N/A	N/A
Motorcycle-car crash	1.747	1.238-2.464	.001	N/A	N/A	N/A
Motorcycle-static object	0.988	0.648-1.505	.95	N/A	N/A	N/A
Commute distance from injury to emergency department (km)						
<5	Reference	N/A	N/A	N/A	N/A	N/A
>5	0.803	0.621-1.037	.09	N/A	N/A	N/A
Mode of arrival at emergency department						
By ambulance	Reference	N/A	N/A	N/A	N/A	N/A
By public taxi/motorcycle	1.120	0.813-1.545	.49	1.621	1.092-2.405	.02
By other means (walk-in)	0.367	0.133-1.015	.05	0.635	0.192-2.101	.46
Prehospital care received prior to arrival at emergency department?						
Yes	Reference	N/A	N/A	N/A	N/A	N/A
No	1.214	0.940-1.569	.14	N/A	N/A	N/A
Number of serious injuries requiring admission[c]						
None	Reference	N/A	N/A	N/A	N/A	N/A
1	2.132	1.423-3.194	<.001	1.923	1.179-3.137	.01
>1	6.285	3.948-10.007	<.001	3.693	1.954-6.980	<.001
Injury severity based on Kampala Trauma Score						
Mild (scores 9‐10)	Reference	N/A	N/A	N/A	N/A	N/A
Moderate (scores 7‐8)	1.567	1.168-2.103	.003	0.987	0.638-1.526	.95
Severe (score ≤6)	4.913	3.390-7.120	<.001	2.713	1.372-5.368	.004
Injury severity based on Glasgow Coma Scale score						
Mild (scores 13‐15)	Reference	N/A	N/A	N/A	N/A	N/A
Moderate (9-12)	1.98	1.441-2.721	<.001	1.169	0.642-2.129	.61
Severe (score ≤8)	3.166	1.99-5.027	<.001	1.330	0.607-2.915	.48
Head CT[d] findings						
CT not indicated	Reference	N/A	N/A	N/A	N/A	N/A
Negative CT results	0.448	0.310-0.646	<.001	0.543	0.353-0.834	.01

XSL•FO
RenderX

Variable	Crude odds ratio	95% CI	*P* value	Adjusted odds ratio	95% CI	*P* value
Extra-axial lesions	1.230	0.845-1.791	.28	0.855	0.465-1.572	.62
Axial lesions	1.983	1.180-3.333	.01	0.753	0.377-1.503	.42
Skull fractures	0.421	0.249-0.711	.001	0.279	0.139-0.558	<.001
Prehospital interval (hours)						
<1	Reference	N/A	N/A	N/A	N/A	N/A
>1	1.106	0.839-1.459	.48	N/A	N/A	N/A
Referral-dispatch interval (hours)						
<1	Reference	N/A	N/A	N/A	N/A	N/A
>1	0.621	0.461-0.838	.002	N/A	N/A	N/A
Neurosurgical treatment						
Watchful waiting	Reference	N/A	N/A	N/A	N/A	N/A
Craniotomy	1.530	1.134-2.065	.01	0.718	0.358-1.437	.35
Decompressive craniectomy	1.734	0.767-3.922	.19	0.383	0.125-1.180	.10
Glasgow Outcome Scale category						
Favorable (scores 4‐5)	Reference	N/A	N/A	N/A	N/A	N/A
Unfavorable (scores 1-3)	3.151	2.231-4.658	<.001	0.543	0.267-1.107	.09
Trauma Outcome Measure Scale score						
Favorable	Reference	N/A	N/A	N/A	N/A	N/A
Unfavorable	1.071	0.737-1.555	.74	N/A	N/A	N/A
90-day injury outcomes						
Alive	Reference	N/A	N/A	N/A	N/A	N/A
Died	4.202	2.245-6.937	<.001	1.228	0.0512-2.945	.65

[a]Mixed-effects logistic regression integration method: mvaghermite, number of observations=861, number of treatment arms=2, average size per treatment arm=501, residual intraclass correlation for intervention vs control arm for adjusted model=0.05, 95% CI (0.01-0.31), Akaike information criterion=1037.4, and Bayesian information criterion=1127.8.

[b]Not applicable.

[c]Serious injury refers to one that warrants admission on its own merit.

[d]CT: computed tomography.

Convergence

The aggregated data indicated that the components of delay 1 exhibited divergence, while convergence was observed for delays 2 and 3. A summary of the integration between qualitative and quantitative findings is presented in Table 3.

XSL•FO
RenderX

Table . A joint display summary showing the intersection of qualitative and quantitative results.

Theme/construct	Quantitative metrics	Quantitative findings	Qualitative quotes	Convergence status	Implications and overall interpretation
Delay 1: delays in decision to seek injury care	Prehospital interval in hours	Prehospital intervals of more than 1 hour were not independently associated with reported barriers to injury care (OR[a] 1.11, 95% CI 0.84 - 1.46; *P*=.48).	• "You want to help evacuate these accident victims from scenes, but they think you are interrogating, arresting, and prosecuting them for their traffic crimes." • "He arrived late in a taxi as there was no fuel in the ambulance from the referring hospital."	Divergent	Prehospital time was confounded by several factors including public mistrust, scene time, commute distance, and ambulance availability, which impacted decisions to seek injury care.
Delay 2: delays in reaching injury care	Referral decision to dispatch interval in hours	Referral to dispatch intervals of more than 1 hour were associated with lower odds of reported barriers to injury care (OR 0.62, 95% CI 0.46-0.84; *P*=.002).	• "Due to understaffing we are often unable to undertake transfer responsibilities timely." • "You want to communicate the transfer timely but the poor communication technologies and network breakdown will let you down." • "Our intercom is not directly linked to hospital emergency departments."	Convergence	Prolonged referral intervals were linked to barriers in accessing timely definitive injury care. Hospitals would delay referrals when the demands of patient care did not exceed the available local resources or when the necessary resources for executing the referrals were unavailable.
Delay 3: delays in receiving adequate injury care.	Emergency department arrival to head CT[b] imaging as a measure of in-hospital delays	Negative CT results were associated with lower odds of reporting barriers (OR 0.54, 95% CI 0.35-0.83; *P*=.01). Conversely, severe Kampala Trauma Scores ≤6 were associated with increased odds for reporting barriers to injury care (OR 2.71, 95% CI 1.37 - 5.37; *P*=.004).	• "We do not have direct referral pathways to private hospitals for CT." • "The patient's next of kin is still mobilizing financial resources to outsource CT from private." • "Our intensive care unit is full, so we plan to refer him to a private hospital, but his family is still mobilizing resources for ICU costs."	Convergence	Critically ill patients with severe injuries face obstacles in accessing injury care such as dysfunctional referral protocols and suboptimal insurance coverage compared with less severely injured individuals.

XSL•FO
RenderX

Theme/construct	Quantitative metrics	Quantitative findings	Qualitative quotes	Convergence status	Implications and overall interpretation
Injury outcomes at 90 days	Injury morbidity based on proportions of unfavorable outcomes and proportions/odds of mortality	Multiplicity of injuries was associated with increased odds of encountering barriers in the continuum of injury care (OR 3.69, 95% CI 1.95 - 6.98; *P*<.001). Encountering barriers was independently associated with 3-fold increase in unfavorable Glasgow Outcome Scores (OR 3.15, 95% CI 2.23 - 4.66; *P*<.001) and 4-fold increase in 90-day mortality (OR 4.202, 95% CI 2.245 - 6.937; *P*<.001).	• "He was supposed to get labetalol for his intracranial pressure from head injury before brain surgery; the relatives need to lobby for that. The orthopedic team also says he has not yet bought the implants for his fractures. The multidisciplinary team will gather when everything is in place." • "We had to wait until patients' relatives availed us with the results of the investigations from private before operating."	Complementary	Multiple injuries were linked to delayed access to critical surgical care interventions. These delays could contribute to undesired injury outcomes and increased mortality.

[a]OR: odds ratio.

[b]CT: computed tomography.

Discussion

Principal Findings

The objective of this study was to identify critical obstacles within the continuum of trauma care, both before and during hospital admission, as perceived by first responders. We found that there was a convergence of key themes between those reported by frontline health workers during the qualitative interviews and those elicited by the attending clinicians during the real-time execution of patient treatment in the quantitative trauma registry audit. Our results indicated notable systemic inefficiencies in the activation of emergency medical services and the arrival of ambulances, which may exacerbate the severity of injuries and negatively impact patient outcomes. Moreover, a lack of designated rural networks for emergency responses, as highlighted by the law enforcement officials, exacerbates the situation in remote areas. Furthermore, the absence of clear protocols for traffic crash victim evacuation may negatively impact the quality of care provided. Similar impediments have been earlier identified in remote settings of other LMICs, highlighting the need to strengthen rural trauma systems [20,38].

Qualitative analyses uncovered organizational and system-level factors encountered by law enforcement and referring clinicians. For instance, understaffing and the absence of clearly defined roles for emergency road traffic response personnel underscore the necessity for strategic workforce management and explicit assignment of responsibilities. The complex interplay among understaffing, role ambiguity, and task delays driven by burnouts due to heavy workload degrades emergency care processes. These effects reflect a mismatch between people, tasks, and tools, which collectively undermines the health system's ability to deliver safe, high-quality care. Within the Systems Engineering Initiative for Patient Safety, these workforce-related conditions operate as work system determinants that shape how care tasks are allocated, coordinated, and executed.

Moreover, there were reported inefficient communication technologies that impeded effective trauma coordination between emergency responders and recipient hospitals, delaying patient transfers. Research shows that such inefficiencies could result in the loss of critical information and compromise patient care [13]. In a recent scoping review of handover practices between paramedics, emergency department staff, and trauma teams, lost information due to communication gaps was associated with diagnostic errors, missed injuries, and increased mortality [39].

Furthermore, we identified psychosocial constraints to an effective emergency response framework manifested by public bias and mistrust of law enforcement trauma evacuation teams. This mistrust led to withholding valuable information regarding the circumstances of injury, which would be essential for assessing the liability and settlement of insurance claims, thereby reducing out-of-pocket expenditure. These observations reinforce the need for comprehensive community engagement and awareness campaigns to build public trust. Statutorily integrating police services with health care systems may bridge these gaps [40]. However, this unified approach requires clearly defined dispatch protocols with explicit role specifications, structured assessment checklists, and quality safeguards with clinical oversight of scope of practice to ensure patient safety and compliance with data-protection policies.

XSL•FO
RenderX

We found in-hospital barriers characterized by a shortage of surgical supplies, which delayed the identification and treatment of life-threatening injuries. Access to essential medical supplies and a competent health workforce are an integral component of World Health Organization (WHO) building blocks for resilient health systems [41]. A sustainable robust logistical framework is mandatory to ensure that clinicians can quickly assess, diagnose, and treat life-threatening conditions. In quantitative analyses, the plausible correlation between the presence of barriers, injury severity, and unfavorable outcomes further emphasizes that these systemic issues not only hinder immediate care but also have lasting implications for the recovery and survival of very critically ill patients whose care demands exceed local resources. The increased odds of mortality among those who encounter obstacles highlight the urgent need for interventions aimed at streamlining care processes and resource allocation within trauma care systems. Comparable research conducted in Rwanda, Ghana, and South Africa assessed the application of multiple data collection methods to identify context-specific blockages to injury care that contributed to mortality and disability, and our findings are consistent with these studies [38].

Finally, the obstacles identified by the frontline workers during the real-time execution of treatment, such as inadequate diagnostic resources, transportation failure, and financial difficulties, spotlight systemic challenges within the Ugandan health care sector. Sustainable health care financing is an additional core WHO building block because it protects patients from catastrophic expenditures [41]. In a recent Delphi study conducted in Malawi [42], the absence of readily available physical resources emerged as the most significant factor contributing to in-hospital delays. Moreover, the high financial cost of care and the unaffordability of emergency transport were identified as the primary factors leading to prehospital delays [42]. These challenges reflect inadequate resource allocation to trauma care and to the broader health sector, which has negatively impacted the quality of health systems in both LMICs and HICs [43,44]. For instance, a systematic review of barriers to injury care in 21 American countries revealed that a lack of resources and equipment constituted 58% (33/57) of barriers, followed by the absence of protocols, understaffing, transport and logistics challenges, and financial limitations [45]. Further research from Australia and Europe elaborates more on these challenges, highlighting rural-urban and socioeconomic disparities amid declining social support structures [46,47].

Poor-quality health care and inadequate financial risk protection are now considered the major drivers of mortality, surpassing inadequate health access, which has delayed the achievement of SDG3 targets 3.6 and 3.8 in vulnerable populations, irrespective of country-specific development index [43,48]. Furthermore, the reported inefficient communication technologies that impede effective trauma coordination, coupled with inadequate infrastructural capabilities influenced by social factors in health systems, could impede the attainment of SDG9 [48]. Thus, attaining universal health coverage should not be regarded as a stand-alone but rather a multisectoral initiative that incorporates elements of SDG9. This would foster technological innovations for medical equipment, build inclusive infrastructure such as safe road networks and accessible hospitals with functional emergency departments, and establish resilient industries that foster research development and equitable medical education to address low-quality and health access barriers [49]. Our results indicate weak connections between the formal and informal sectors, public and private sectors, and health and technology sectors. Addressing these gaps through strong advocacy by civil society organizations could meet the care demands of severely injured patients and enhance the overall learning and work environment for trainee physicians in low-resource settings, thereby improving the quality of their medical education (SDG 4).

Policy Implications

Our study has identified multiple significant obstacles present at the various stages of the trauma care continuum. Most of these identified challenges are applicable to other LMICs [20,38,42]. Consequently, the hindrances to injury care from the perspectives of trainee physicians and traffic law enforcement first responders must be addressed to accurately conceptualize their expected roles in the trauma care continuum and patients' advocacy. Addressing these barriers requires multidisciplinary approaches focusing on quality prehospital response, swift safe transportation, and robust postcrash care systems. Stakeholders, including the National Road Safety Council, Ugandan Ministry of Health, Ministry of Education, and Regional Police Commanders, should engage in planning meetings to address these critical issues.

All the financial, logistical, and system-level impediments to trauma care elicited in this study are direct obstacles to the WHO health systems building blocks [41]. This study provides context for the implementation strategies needed to address gaps in injury care. In Uganda, policymakers should strengthen the existing Integrated Disease Surveillance and Response system by adapting monitoring tools including electronic logistics management and health information systems to track stockouts of essential medical supplies, human resources observatories to monitor health workforce distribution and service delivery inequities, and national health accounts to monitor health financing, public expenditure, and financial risk protection. These measures, aligned with WHO guidance and the Global Health Observatory, would form a coherent framework for assessing and improving health sector performance regarding trauma care [41].

Regarding individual-patient and trauma-team delays, the interplay between limited insurance coverage and insufficient disposable incomes leads to substantial delays as caretakers mobilize funds for critical investigations and emergency surgical procedures. For instance, a brain computed tomographic scan can cost a substantial amount of up to US $130, commensurate to the average monthly salary of green-collar workers in Uganda [24]. Consequently, trauma teams delay surgical decisions, which may preclude high-quality surgical outcomes amidst potential threat for medical-legal litigation, as patients and families increasingly assert their autonomy in treatment decisions. The likelihood that individuals with severe injuries have high care demands, which may exceed local resources, necessitates advocacy for improvement in the health system to

XSL•FO
RenderX

adequately address these needs. To achieve this goal, stewardship that promotes good governance and team leadership, together with routine community health insurance engagement, should accompany health system performance assessments in line with the WHO building blocks for sustainable health systems.

In-hospital system delays in injury care were primarily attributed to a lack of supplies and critical care services, with these constraints ranking second in an audit of patient handling. Key supplies such as oxygen, anesthetics, and blood products, as well as access to operating theaters and intensive care units, are all mandatory requisites for the survival of patients with major trauma. Access to good-quality service delivery such as critical care underpins one of the WHO building blocks. Previous studies indicate that dual deficiencies in both equipment and human resources imply a severe hindrance to lifesaving interventions, contributing to a substantial proportion of preventable trauma deaths in LMICs [13,20]. To mitigate these adversities, investment in reliable supply chain management through private-public partnerships, improving ambulance coordination through sufficient staffing, and establishing financial support systems for uninsured patients could streamline referral pathways to ensure continuity of care in emergency situations. The use of noncontributory funding sources generated through public taxation, pooling funds across multiple public sectors, and strong accountability to track performance of the designated funds could mitigate the shortage of supplies and safeguard patients against catastrophic expenditures [48].

Strengths and Limitations

The strengths of this study dwell in its mixed methods approach, which enabled a nuanced analysis of barriers encountered at various stages of real-world injury care. Additionally, by focusing on the perspectives of frontline workers directly involved in trauma care, the study provides valuable insights into context-specific challenges encountered in LMIC settings. Moreover, the study contributes valuable evidence on the intersection of trauma systems, medical education, and health equity in low-resource settings, which is relevant to the attainment of SDGs 3 and 4.

However, there were limitations that should be recognized. First, the use of Likert scales may pose challenges for interpretation although they were applied to mixed methods to enrich qualitative inquiry and facilitate triangulation with quantitative data. While the study used parametric tests to enhance statistical power, this approach might not align with best practices for exploratory research that lacks prior hypotheses [50].

Second, the study assumes that accumulated Likert items can serve as a reasonable surrogate for continuous interval measurements. Although this assumption facilitates the computation of overall mean scores to enable objective ranking for the barriers, it may potentially compromise the precision of central tendency measures [51]. Moreover, while we endeavored to interpret the data objectively through triangulation and use of parametric tests, we acknowledge the possibility of bias arising from the integration of diverse data paradigms which used varying collection and sampling methods. This concern reflects a significant and ongoing debate within the field of mixed methods research [52].

Third, the reliance on data saturation during interviews may have unintentionally constrained the identified barriers to dominant themes, thereby silencing divergent minority opinions. This represents an inherent methodological issue in qualitative research, which often prioritizes the depth and richness of data over numerical targets [31]. Moreover, we acknowledge that recall in qualitative responses and the missing 90-day outcome data may have introduced systematic bias.

Finally, we acknowledge that aggregating injury severity into subcategories of mild, moderate, and severe, and outcomes as favorable versus unfavorable, may lead to a loss of information, contributing to residual confounding in assessing the true impact of barriers on injury outcomes. Moreover, in view of other host and environmental factors that impact injury outcomes but were not assessed in this study, patients with less severe injuries may have had minimal care needs, which could prevent attending clinicians from identifying significant in-hospital barriers during their treatment, compared with the severely injured patients, an issue that future studies should address. Nevertheless, the study assumed that stratifying trauma outcomes by injury severity categories would minimize this confounding.

Conclusions

This mixed methods study evaluated the key barriers to injury care within the context of frontline health workers in low-resource settings. We identified significant barriers in the prehospital and in-hospital trauma care continuum, which could adversely impact patient outcomes. These systemic challenges require multifaceted targeted interventions that integrate operational improvements to address system-level factors and logistical frameworks through community engagement. The results of this study warrant an urgent need to secure sustainable hospital supplies and establish clear referral protocols that link the public to private facilities. These interventions would enable timely treatment decisions to optimize trauma care delivery and treatment outcomes, especially for those who sustain severe injuries. Future research should be robust multicenter prospective cohort studies that integrate mixed methods to investigate how these barriers impact long-term injury outcomes in diverse socioeconomic settings.

Acknowledgments

The authors wish to express their gratitude to the staff at various medical referral institutions in Uganda, including Jinja, Mbarara, Hoima, Mubende, Kiryandongo, and Kampala International University Teaching Hospitals for providing space for data collection. Furthermore, the authors acknowledge all study participants for their willingness to dedicate their time to participate in this investigation and the MOTOR trial registry project investigators for permitting triangulation of the datasets used in this study.

XSL•FO
RenderX

The authors declare the use of generative artificial intelligence in the research writing process. Under full human supervision, Microsoft CoPilot and Poe Assistant were used during proofreading, editing, and refining grammatical errors. The responsibility for the final paper lies entirely with the authors. The views expressed in this paper are solely those of the authors and do not necessarily represent the official positions of their respective institutions.

Funding

HL received funding support through participation in fellowships from the University of Turku Graduate School, Turku University Hospital (TYKS) Foundation, TYKS Neurocenter, and Center for Health Equity in Surgery and Anesthesia (CHESA) at the University of California San Francisco. Additionally, JPP was supported by the Research Council of Finland and Sigrid Jusélius Foundation. The funding agencies had no role in the design, execution, or reporting of the research findings.

Data Availability

The anonymized datasets generated and/or analyzed during this study are available in the Mendeley data repository [29].

Authors' Contributions

HL contributed to conceptualization, data curation, investigation, methodology, project administration, resources, formal analysis, software, visualization, writing (original draft, review, and editing), and funding acquisition. BO participated in data curation, investigation, project administration, validation, and writing (review and editing). MM contributed to formal analysis, software, visualization, validation, and writing (review and editing). JPP and MLW contributed to resources, investigation, funding acquisition, supervision, validation, and writing (review and editing). All authors read and approved the last version of the paper.

Conflicts of Interest

JPP has received speaker’s fees from Sanofi SA, the Finnish Medical Association, Wellbeing services county of North Karelia, and Finnish Association of Otorhinolaryngology—Head and Neck Surgery and travel expenses reimbursement and expert fees from the National Institute of Neurological Disorders and Stroke. All other authors declare that they have no conflicts of interest.

Multimedia Appendix 1
Barriers to injury care identified through qualitative structured interviews and focus groups.
[DOCX File, 36 KB - humanfactors_v13i1e84774_app1.docx]

Multimedia Appendix 2
Motorcycle trauma and outcome registry data sheet.
[DOCX File, 89 KB - humanfactors_v13i1e84774_app2.docx]

Multimedia Appendix 3
Mean rank scores for different barriers to injury care by the various cadres.
[DOCX File, 17 KB - humanfactors_v13i1e84774_app3.docx]

Checklist 1
COREQ checklist.
[PDF File, 163 KB - humanfactors_v13i1e84774_app4.pdf]

Checklist 2
GRAMMS checklist.
[PDF File, 108 KB - humanfactors_v13i1e84774_app5.pdf]

References

1. Laflamme L, Burrows S, Hasselberg M. Socioeconomic differences in injury risks: a review of findings and a discussion of potential countermeasures. World Health Organization. URL: https://apps.who.int/iris/handle/10665/350275 [accessed 2026-02-15]
2. Barthélemy EJ, Affana CK, Asfaw ZK, et al. Racial and socioeconomic disparities in neurotrauma: research priorities in the New York metropolitan area through a global neurosurgery paradigm. World Neurosurg 2022 Sep;165:51-57. [doi: 10.1016/j.wneu.2022.06.031] [Medline: 35700861]
3. Haagsma JA, Charalampous P, Ariani F, et al. The burden of injury in Central, Eastern, and Western European sub-region: a systematic analysis from the Global Burden of Disease 2019 Study. Arch Public Health 2022 May 20;80(1):142. [doi: 10.1186/s13690-022-00891-6] [Medline: 35590340]

XSL•FO
RenderX

4. Morgan JM, Calleja P. Emergency trauma care in rural and remote settings: challenges and patient outcomes. Int Emerg Nurs 2020 Jul;51:100880. [doi: 10.1016/j.ienj.2020.100880] [Medline: 32622226]
5. Luft LM. The essential role of physician as advocate: how and why we pass it on. Can Med Educ J 2017 Jun;8(3):e109-e116. [Medline: 29098052]
6. Alayande B, Chu KM, Jumbam DT, et al. Disparities in access to trauma care in Sub-Saharan Africa: a narrative review. Curr Trauma Rep 2022;8(3):66-94. [doi: 10.1007/s40719-022-00229-1] [Medline: 35692507]
7. James SL, Castle CD, Dingels ZV, et al. Global injury morbidity and mortality from 1990 to 2017: results from the Global Burden of Disease Study 2017. Inj Prev 2020 Oct;26(Supp 1):i96-i114. [doi: 10.1136/injuryprev-2019-043494] [Medline: 32332142]
8. Whitaker J, Nepogodiev D, Leather A, Davies J. Assessing barriers to quality trauma care in low and middle-income countries: a Delphi study. Injury 2020 Feb;51(2):278-285. [doi: 10.1016/j.injury.2019.12.035] [Medline: 31883865]
9. Stewart B, Hollis S, Amato SS, Bulger E, Mock C, Reynolds T. Trauma care and development assistance: opportunities to reduce the burden of injury and strengthen health systems. Bull World Health Organ 2019 May 1;97(5):371-373. [doi: 10.2471/BLT.18.213074] [Medline: 31551634]
10. Haider AH, Weygandt PL, Bentley JM, et al. Disparities in trauma care and outcomes in the United States: a systematic review and meta-analysis. J Trauma Acute Care Surg 2013 May;74(5):1195-1205. [doi: 10.1097/TA.0b013e31828c331d] [Medline: 23609267]
11. Plummer V, Boyle M. EMS systems in lower-middle income countries: a literature review. Prehosp Disaster Med 2017 Feb;32(1):64-70. [doi: 10.1017/S1049023X1600114X] [Medline: 27938449]
12. Mould-Millman NK, Dixon JM, Sefa N, et al. The state of emergency medical services (EMS) systems in Africa. Prehosp Disaster Med 2017 Jun;32(3):273-283. [doi: 10.1017/S1049023X17000061] [Medline: 28228178]
13. Ningwa A, Muni K, Oporia F, et al. The state of emergency medical services and acute health facility care in Uganda: findings from a National Cross-Sectional Survey. BMC Health Serv Res 2020 Jul 9;20(1):634. [doi: 10.1186/s12913-020-05508-8] [Medline: 32646519]
14. Nzasabimana P, Ignatowicz A, Alayande BT, et al. Barriers to equitable access to quality trauma care in Rwanda: a qualitative study. BMJ Open 2023 Sep 28;13(9):e075117. [doi: 10.1136/bmjopen-2023-075117] [Medline: 37770259]
15. Oleribe OO, Momoh J, Uzochukwu BS, et al. Identifying key challenges facing healthcare systems in Africa and potential solutions. Int J Gen Med 2019;12:395-403. [doi: 10.2147/IJGM.S223882] [Medline: 31819592]
16. Sadeq A, Guraya SS, Fahey B, et al. Medical professionalism education: a systematic review of interventions, outcomes, and sustainability. Front Med 2025;12:1522411. [doi: 10.3389/fmed.2025.1522411]
17. Al-Worafi YM. Medical and health sciences education training in developing countries: overview. In: Al-Worafi YM, editor. Handbook of Medical and Health Sciences in Developing Countries: Springer International Publishing; 2023:1-16. [doi: 10.1007/978-3-030-74786-2_148-1]
18. Lule H, Mugerwa M, Ssebuufu R, et al. Effect of rural trauma team development on the outcomes of motorcycle accident-related injuries (motor registry project): protocol for a multicenter cluster randomized controlled trial. JMIR Res Protoc 2024 May 7;13:e55297. [doi: 10.2196/55297] [Medline: 38713507]
19. Lule H, Mugerwa M, Abio A, et al. Impact of rural trauma team development education on prehospital time, referral-to-dispatch interval, and neurological and musculoskeletal injury outcomes: cluster randomized controlled trial. JMIR Hum Factors 2026 Apr 20;13:e82591. [doi: 10.2196/82591] [Medline: 42008599]
20. Whitaker J, O'Donohoe N, Denning M, et al. Assessing trauma care systems in low-income and middle-income countries: a systematic review and evidence synthesis mapping the Three Delays Framework to injury health system assessments. BMJ Glob Health 2021 May;6(5):e004324. [doi: 10.1136/bmjgh-2020-004324] [Medline: 33975885]
21. Bayiga Zziwa E, Muhumuza C, Muni KM, Atuyambe L, Bachani AM, Kobusingye OC. Road traffic injuries in Uganda: pre-hospital care time intervals from crash scene to hospital and related factors by the Uganda Police. Int J Inj Contr Saf Promot 2019 Jun;26(2):170-175. [doi: 10.1080/17457300.2018.1535511] [Medline: 30541384]
22. Balikuddembe JK, Ardalan A, Khorasani-Zavareh D, Nejati A, Raza O. Weaknesses and capacities affecting the prehospital emergency care for victims of road traffic incidents in the greater Kampala metropolitan area: a cross-sectional study. BMC Emerg Med 2017 Oct 3;17(1):29. [doi: 10.1186/s12873-017-0137-2] [Medline: 28974202]
23. Wasti SP, Simkhada P, van Teijlingen ER, Sathian B, Banerjee I. The growing importance of mixed-methods research in health. Nepal J Epidemiol 2022 Mar;12(1):1175-1178. [doi: 10.3126/nje.v12i1.43633] [Medline: 35528457]
24. Vaca SD, Kuo BJ, Nickenig Vissoci JR, et al. Temporal delays along the neurosurgical care continuum for traumatic brain injury patients at a tertiary care hospital in Kampala, Uganda. Neurosurgery 2019 Jan 1;84(1):95-103. [doi: 10.1093/neuros/nyy004] [Medline: 29490070]
25. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008 Apr;13(2):92-98. [doi: 10.1258/jhsrp.2007.007074] [Medline: 18416914]
26. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042]

XSL•FO
RenderX

27. Lule H, Mugerwa M, SSebuufu R, Kyamanywa P, Posti JP, Wilson ML. Rural trauma team development training amongst medical trainees and traffic law enforcement professionals in a low-income country: a protocol for a prospective multicenter interrupted time series. Int J Surg Protoc 2024 Mar;28(1):12-19. [doi: 10.1097/SP9.0000000000000013] [Medline: 38433864]
28. Lule H, Mugerwa M, Ssebuufu R, Kyamanywa P, Posti JP, Wilson ML. Prospective multi-centre analysis of rural trauma team development training for medical trainees and traffic law enforcement professionals in Uganda: an interrupted time series study. BMC Med Educ 2025 Jan 29;25(1):148. [doi: 10.1186/s12909-025-06755-1] [Medline: 39881413]
29. Lule H, Mugerwa M, Abio A, et al. Motorcycle Trauma Outcome Registry (MOTOR) trial dataset. : Mendeley Data; 2025. [doi: 10.17632/BGPMKPCWDT.1]
30. Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: a systematic review of empirical tests. Soc Sci Med 2022 Jan;292:114523. [doi: 10.1016/j.socscimed.2021.114523] [Medline: 34785096]
31. Ahmed SK. Sample size for saturation in qualitative research: debates, definitions, and strategies. J Med Surg Public Health 2025 Apr;5:100171. [doi: 10.1016/j.glmedi.2024.100171]
32. Rahimi S, Khatooni M. Saturation in qualitative research: an evolutionary concept analysis. Int J Nurs Stud Adv 2024 Jun;6:100174. [doi: 10.1016/j.ijnsa.2024.100174] [Medline: 38746797]
33. Lule H, Mugerwa M, Abio A, et al. Factors associated with all-cause mortality and morbidity of motorcycle crash-related neurological and musculoskeletal injuries in Uganda: the MOTOR cluster randomised trial ancillary study. Inj Prev 2026 Jan 2:ip-2025-045737. [doi: 10.1136/ip-2025-045737] [Medline: 41482470]
34. Damulira J, Muhumuza J, Kabuye U, et al. New trauma score versus Kampala Trauma Score II in predicting mortality following road traffic crash: a prospective multi-center cohort study. BMC Emerg Med 2024 Jul 29;24(1):130. [doi: 10.1186/s12873-024-01048-0] [Medline: 39075406]
35. Fitzgerald E, Donaldson L, Hammond N, et al. Methods of functional outcome assessment in neurocritical care trials: a systematic review and meta-analysis. Aust Crit Care 2025 Mar;38(2):101124. [doi: 10.1016/j.aucc.2024.09.009] [Medline: 39489650]
36. Fang C, Platz A, Müller L, et al. Evaluation of an expectation and outcome measurement questionnaire in ankle fracture patients: The Trauma Expectation Factor Trauma Outcomes Measure (TEFTOM) Eurasia study. J Orthop Surg (Hong Kong) 2020;28(1):2309499019890140. [doi: 10.1177/2309499019890140] [Medline: 31916492]
37. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
38. Equi-Trauma Collaborative, Odland ML, Abdul-Latif AM, et al. Equitable access to quality trauma systems in low-income and middle-income countries: assessing gaps and developing priorities in Ghana, Rwanda and South Africa. BMJ Glob Health 2022 Apr;7(4):e008256. [doi: 10.1136/bmjgh-2021-008256] [Medline: 35410954]
39. Harthi N, Almuwallad A, Albargi H, et al. Trauma handover practices between ambulance clinicians, emergency department staff and trauma teams: scoping review. BMC Emerg Med 2026 Jan 27;26(1):57. [doi: 10.1186/s12873-026-01480-4] [Medline: 41588499]
40. Salhi RA, Iyengar S, da Silva Bhatia B, Smith GC, Heisler M. How do current police practices impact trauma care in the prehospital setting? a scoping review. J Am Coll Emerg Physicians Open 2023 Jun;4(3):e12974. [doi: 10.1002/emp2.12974] [Medline: 37229183]
41. Jauhar SM, Singh A, Rishipathak P, Kulkarni M. Navigating health systems: a scoping review of WHO Building Blocks and Control Knobs frameworks. Hosp Top 2025 Aug 13:1-27. [doi: 10.1080/00185868.2025.2542447] [Medline: 40802244]
42. Whitaker J, Njawala T, Nyirenda V, et al. Identifying and prioritising barriers to injury care in northern Malawi, results of a multifacility multidisciplinary health facility staff survey. PLoS One 2024;19(9):e0308525. [doi: 10.1371/journal.pone.0308525] [Medline: 39264901]
43. Kruk ME, Gage AD, Arsenault C, et al. High-quality health systems in the Sustainable Development Goals era: time for a revolution. Lancet Glob Health 2018 Nov;6(11):e1196-e1252. [doi: 10.1016/S2214-109X(18)30386-3] [Medline: 30196093]
44. Vărzaru AA. Assessing the relationships of expenditure and health outcomes in healthcare systems: a system design approach. Healthcare (Basel) 2025 Feb 7;13(4):352. [doi: 10.3390/healthcare13040352] [Medline: 39997227]
45. Kinder F, Mehmood S, Hodgson H, Giannoudis P, Howard A. Barriers to trauma care in south and central America: a systematic review. Eur J Orthop Surg Traumatol 2022 Aug;32(6):1163-1177. [doi: 10.1007/s00590-021-03080-3] [Medline: 34392445]
46. Hafeez-Baig S, Buckley L, Midwinter M. Trauma outcomes based on remoteness of injury in Australia: a systemic review. Aust J Rural Health 2025 Feb;33(1):e13216. [doi: 10.1111/ajr.13216] [Medline: 39803766]
47. Sethi D, Racioppi F, Baumgarten I, Bertollini R. Reducing inequalities from injuries in Europe. Lancet 2006 Dec 23;368(9554):2243-2250. [doi: 10.1016/S0140-6736(06)68895-8] [Medline: 17189036]
48. Collins TE, Akselrod S, Atun R, et al. Converging global health agendas and universal health coverage: financing whole-of-government action through UHC. Lancet Glob Health 2023 Dec;11(12):e1978-e1985. [doi: 10.1016/S2214-109X(23)00489-8] [Medline: 37973345]
49. Ugwu OPC, Ugwu OPC, Alum EU, et al. Sustainable Development Goals (SDGs) and resilient healthcare systems: addressing medicine and public health challenges in conflict zones. Medicine (Baltimore) 2025 Feb 14;104(7):e41535. [doi: 10.1097/MD.0000000000041535] [Medline: 39960902]

XSL•FO
RenderX

50. Fagerland MW, Sandvik L, Mowinckel P. Parametric methods outperformed non-parametric methods in comparisons of discrete numerical variables. BMC Med Res Methodol 2011 Apr 13;11:44. [doi: 10.1186/1471-2288-11-44] [Medline: 21489231]
51. South L, Saffo D, Vitek O, Dunne C, Borkin MA. Effective use of Likert scales in visualization evaluations: a systematic review. Comput Graph Forum 2022 Jun;41(3):43-55. [doi: 10.1111/cgf.14521]
52. Liu Y. Paradigmatic compatibility matters: a critical review of qualitative-quantitative debate in mixed methods research. Sage Open 2022 Jan;12(1):21582440221079920. [doi: 10.1177/21582440221079922]

Abbreviations

aOR: adjusted odds ratio
COREQ : Consolidated Criteria for Reporting Qualitative Research
GCS: Glasgow Coma Scale
GOS: Glasgow Outcome Scale
HIC: high-income country
KTS: Kampala Trauma Score
LMIC: low- and middle-income country
MOTOR: Motorcycle Trauma Outcome Registry
OR: odds ratio
REDCap: Research Electronic Data Capture
SDG: Sustainable Development Goal
TOMS: Trauma Outcome Measure Scale
WHO: World Health Organization

Edited by A Bamgboje-Ayodele; submitted 27.Sep.2025; peer-reviewed by N Tuan, X Zhang; revised version received 29.Apr.2026; accepted 04.May.2026; published 10.Jun.2026.

Please cite as:
Lule H, Oguttu B, Mugerwa M, Wilson ML, Posti JP
Understanding Barriers to Effective Injury Care by Medical Trainees and Traffic Law Enforcement First Responders in Low-Income Contexts in Uganda (Motor Registry Project Part 2): Convergent Mixed Methods Analysis
JMIR Hum Factors 2026;13:e84774
URL: https://humanfactors.jmir.org/2026/1/e84774
doi:10.2196/84774

© Herman Lule, Benson Oguttu, Micheal Mugerwa, Michael Lowery Wilson, Jussi P Posti. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 10.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

E-Learning Modules Based on Bloom Taxonomy and the Miller Pyramid for First-Year Indian Medical Students: Randomized Controlled Study in Medical Education

Archana Prabu Kumar[1], MBBS, MD, PhD; Padmavathi R[2], MBBS, MD, PhD; Maheshkumar K[3], BNYS, MSc, PhD; Abirami Omprakash[2], MBBS, MD, PhD; Prabu Kumar Chokkalingam Mani[4], MBBS, MD; B W C Sathiyasekaran[5], MBBS, MD; Maruthy K N[6], MBBS, MD

[1]Medical Education Department, College of Medicine and Health Sciences, Arabian Gulf University, Bahrain, Bahrain

[2]Department of Physiology, Sri Ramachandra Medical College and Research Institute Sri Ramachandra Institute of Higher Education and Research, No 1 Ramachandra Nagar, Chennai, Tamil Nadu, India

[3]Department of Physiology and Biochemistry, Government Yoga and Naturopathy Medical College and Hospital, Chennai, Tamil Nadu, India

[4]Department of Pathology, Government Hospitals – Salmaniya Medical Complex, Manama, Bahrain

[5]Sree Balaji Medical College, Chennai, Tamil Nadu, India

[6]Department of Physiology, PES Institute of Medical Sciences and Research, Kuppam, Tamil Nadu, India

Corresponding Author:
Padmavathi R, MBBS, MD, PhD
Department of Physiology, Sri Ramachandra Medical College and Research Institute Sri Ramachandra Institute of Higher Education and Research, No 1 Ramachandra Nagar, Chennai, Tamil Nadu, India

Abstract

Background: Competency-based medical education (CBME) in India emphasizes early competency formation, higher-order cognitive processing, and self-directed learning. Although e-learning is widely adopted, there is limited evidence on structured e-modules explicitly designed using Bloom taxonomy and the Miller pyramid for Indian MBBS students.

Objective: This study aims to design, validate, implement, and evaluate CBME-aligned e-modules for first-year MBBS foundational subjects (anatomy, physiology, and biochemistry) and to compare their effectiveness with traditional teaching on cognitive, psychomotor, and affective learning outcomes using Kirkpatrick levels 1 and 2.

Methods: A randomized controlled study was conducted among 690 first-year medical undergraduates (control: n=370; intervention: n=320). The intervention group received Sharable Content Object Reference Model–based interactive e-modules through Moodle (Modular Object-Oriented Dynamic Learning Environment) in addition to standard lectures, while the control group received lectures only. e-Modules were designed using Bloom taxonomy and the Miller pyramid, validated by internal and external experts, and implemented following Kern 6-step approach. Learning outcomes were assessed using structured feedback (Kirkpatrick level 1) and end-of-block internal assessments comprising multiple choice questions, short notes, and objective structured clinical examinations (Kirkpatrick level 2).

Results: Students in the intervention group performed significantly better across all cognitive levels compared with the control group: remember (mean 6.52, SD 2.24 vs mean 5.26, SD 2.89; P<.001), understand (mean 2.92, SD 1.45 vs mean 2.55, SD 1.13; P<.001), apply (mean 3.43, SD 1.06 vs mean 2.56, SD 1.08; P<.001), and analyze (mean 3.01, SD 0.83 vs mean 2.53, SD 1.11; P<.001). Psychomotor scores (objective structured clinical examination manipulation: mean 4.55, SD 1.12 vs mean 4.10, SD 1.42; P<.001) and affective domain scores (mean 3.02, SD 0.81 vs mean 2.03, SD 0.81; P<.001) were also significantly higher in the intervention group. Subgroup analysis showed the largest gains among medium achievers across domains.

Conclusions: CBME-aligned e-modules significantly enhanced student performance in cognitive, psychomotor, and affective domains compared with traditional teaching alone, with particularly pronounced benefits for medium achievers. Well-designed e-modules represent a scalable, adaptable strategy to support CBME implementation across diverse medical education settings in India.

(JMIR Hum Factors 2026;13:e84339) doi:10.2196/84339

KEYWORDS

competency-based medical education; CBME; e-module; Bloom taxonomy; Miller pyramid; medical undergraduates; digital learning

XSL•FO
RenderX

Introduction

Background

The National Medical Commission in India introduced a new approach, known as competency-based medical education (CBME), in 2015 and officially implemented it in 2019 [1]. This updated curriculum encourages new ways of learning, including self-directed learning; elective modules; and training in attitude, ethics, and communication [2,3]. It is aimed at directing medical education to focus more on helping students build essential skills, such as critical thinking and decision-making.

While traditional lectures often struggle to maintain student interest and foster deeper understanding and critical reasoning skills [4,5], e-learning (computer-assisted learning), which uses electronic devices to support teaching and learning, is becoming increasingly popular [6]. A common way e-learning helps is through blended learning, which mixes classroom teaching with online materials, such as tutorials [7]. One such tool is e-modules. An e-module is defined as an e-learning strategy, which focuses on 1 or 2 learning concepts, usually incorporating a blend of teaching and assessment tools that may include multimedia, video clips, direct lecture instruction, game-based activities, and even links to social media [8,9] to help students stay engaged and learn better [7]. Such tools have been shown to improve short-term memory and keep learners engaged [10]. They are found useful in early medical training, where complex theory can be better understood through interactive content [11]. However, for e-learning to be successful, it must consider cognitive limitations and learning preferences [12].

Rationale

The COVID-19 pandemic led to a rapid increase in digital learning tools [13], and research over time has shown that these tools can help students gain both knowledge and skills [14,15]. CBME is based on well-known learning frameworks, such as Bloom taxonomy [16] and the Miller pyramid [17], which help structure learning in a step-by-step way. Yet, the use of digital modules built on these frameworks is limited [18]. There is a clear gap, and few studies have developed or tested e-modules created specifically for Indian medical students, especially those based on Bloom and Miller models [19].

Moreover, to evaluate training programs, studies in the past have not applied comprehensive evaluations of the Kirkpatrick model to assess e-learning interventions in medical education [20]. A study evaluating a teaching workshop for health care staff has used the first 3 levels, demonstrating improvements in satisfaction, knowledge, and behavior [8]. Such evaluations specifically targeting e-learning modules for Indian medical undergraduates are limited.

Study Objectives

Thus, this study aimed to develop, validate, implement, and evaluate e-modules tailored for Indian medical undergraduates. The objectives of the study are as follows:

1. To design e-modules for first-year MBBS foundational disciplines (anatomy, physiology, and biochemistry), using Bloom taxonomy and the Miller pyramid to target cognitive, affective, and psychomotor competencies outlined in the Indian CBME curriculum.
2. To evaluate the effectiveness of these e-modules, implemented through Moodle-SCORM (Modular Object-Oriented Dynamic Learning Environment–Sharable Content Object Reference Model), in comparison with traditional classroom teaching by examining learning outcomes across cognitive, psychomotor, and affective domains using Kirkpatrick levels 1 and 2.

This study provides one of the first large-scale evaluations of theory-driven e-modules designed specifically for Indian medical undergraduates, addressing a critical gap in CBME-aligned digital learning research. By addressing these objectives, this study seeks to bridge the gap between pedagogical strategies and technological advancements in medical education, ultimately enhancing the competency and preparedness of future health care professionals.

Methods

Study Design and Setting

This randomized controlled study was conducted between 2015 and 2018 at a tertiary medical university in south India. The study followed the Kern 6-step approach [7,21] for curriculum development and aimed to evaluate the effectiveness of e-learning modules, designed using Bloom taxonomy and the Miller pyramid, in enhancing learning outcomes among first-year MBBS students.

Ethical Considerations

This study received ethical approval from the Institutional Ethics Committee of Sri Ramachandra University, Chennai, India (Approval No: IEC-NI/12/OCT/30/53, approved on March 24, 2014). All procedures were conducted in accordance with the ethical standards of the institutional review board and the principles of the Declaration of Helsinki.

Participation in the study was voluntary, and all participants were informed about the purpose and procedures of the study prior to participation. Written informed consent was obtained from all participants. Participant privacy and confidentiality were strictly maintained. Data were anonymized prior to analysis, and no personally identifiable information was included in the dataset used for analysis or in the reporting of results. Participants did not receive any financial compensation or incentives for their participation in this study.

Participants and Randomization

A total of 690 first-year MBBS students enrolled during the study period were invited to participate. Students who completed all baseline assessments and attended the scheduled end-of-block internal assessments (EBIAs) were included in the final analysis.

Students were randomized into control and intervention groups using the Research Randomizer software (version 4.0; Geoffrey C. Urbaniak and Scott Plous; Multimedia Appendix 1).

- Control group: This group received standard PowerPoint (Microsoft Corp) lectures

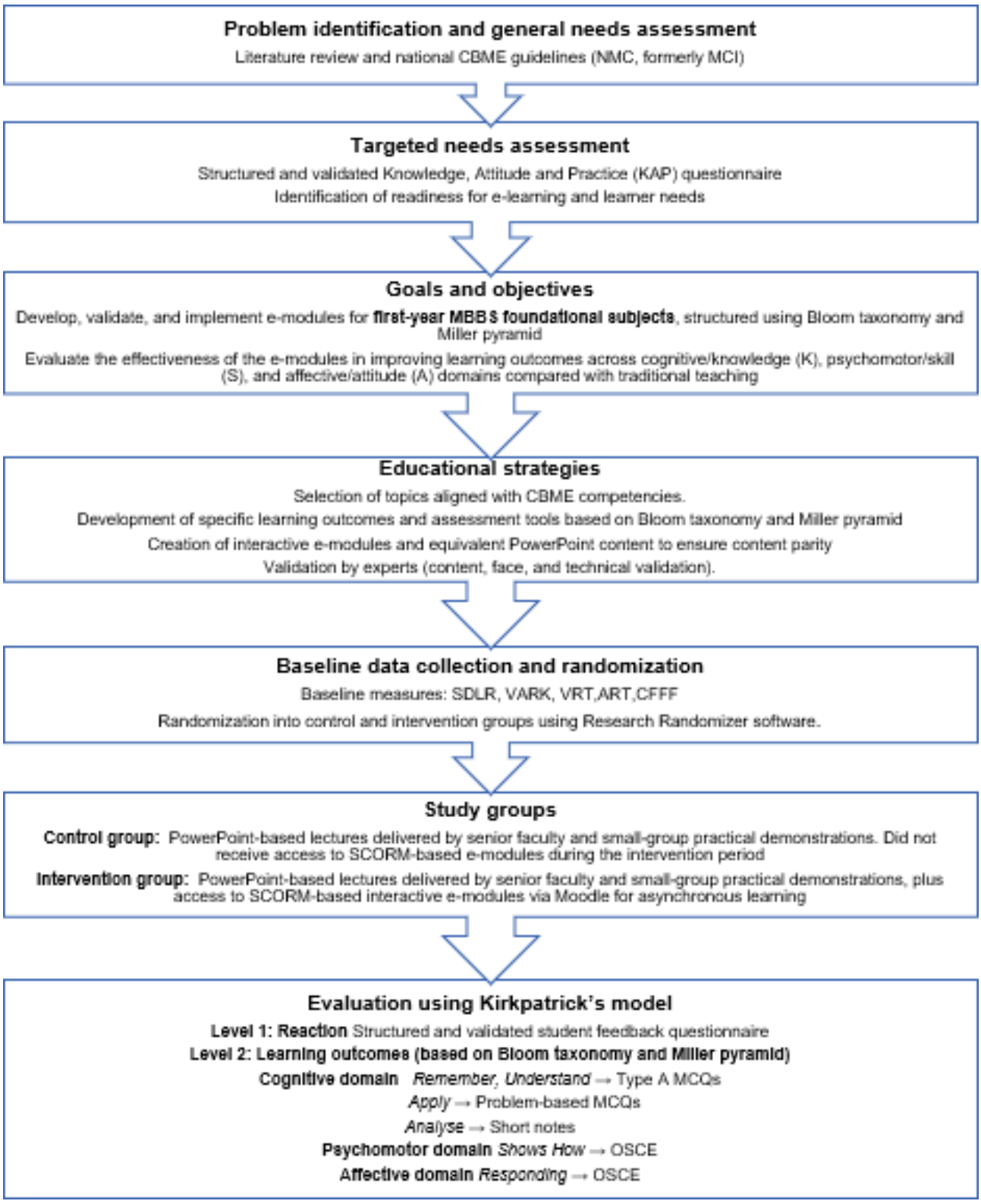
XSL•FO
RenderX

- Intervention group: This group received both PowerPoint lectures and e-modules via Moodle with unique login credentials (Multimedia Appendix 2)

All students in the control group gained access to e-modules after the end-of-block assessment to ensure equitable learning opportunities.

The steps followed for data collection are shown in Figure 1.

Figure 1. Flowchart of curriculum development, implementation, and evaluation steps based on Kern 6-step approach and the Kirkpatrick model. ART: auditory reaction time; CBME: competency-based medical education; CFFF: critical flicker fusion frequency; MCI: Medical Council of India; NMC: National Medical Commission; OSCE: objective structured clinical examination; SDLR: self-directed learning readiness; SLO: specific learning outcomes; VARK: visual auditory reading kinesthetic; VRT: visual reaction time.



XSL•FO
RenderX

Targeted Needs Assessment

The data were collected using a structured and validated knowledge, attitude, and practice questionnaire (Multimedia Appendix 3) [6,10]. Knowledge, attitude, and practice analysis revealed high digital readiness, with all students owning internet-enabled devices, most confident in basic software use and strongly favoring interactive e-modules in education.

The results emphasized the need to develop self-directed e-learning tools tailored to the diverse learning styles of Indian medical undergraduates. E-learning strategies were identified as essential for the rapid and effective implementation of new curricula.

Educational Intervention

Selection of Topics

Four topics were selected based on global or national significance, core learning relevance, feasibility, faculty expertise, and alignment with the curriculum. A list of e-modules addressing the specific CBME competencies and learning objectives aligned with respective assessment tools was categorized under Bloom taxonomy and the Miller pyramid (Multimedia Appendix 4). The selected e-modules covered core foundational topics from anatomy, physiology, and biochemistry that map directly to first-year CBME competencies.

Development of e-Modules

e-Modules were developed as structured, interactive learning tools aligned with CBME competencies. Content (text, images, and animations) was created collaboratively by subject experts and educational technologists using standard multimedia authoring software. Visual elements were sourced from textbooks and open-access repositories in accordance with fair-use guidelines, and additional illustrations were created where needed (Multimedia Appendix 5).

Platform (Moodle-SCORM) and Delivery Approach

Multimedia components, including narration, animations, and interactive elements, were combined to produce SCORM-compliant packages suitable for delivery via Moodle. Each module included embedded self-assessment exercises (multiple choice questions [MCQs], case scenarios, and concept-based questions) with automated scoring and immediate feedback (Multimedia Appendices 6 and 7).

To ensure content equivalence, identical learning material (text and images) was used to create accompanying PowerPoint presentations for the control group, thereby avoiding content bias between instructional formats.

Validation of Learning Materials

The e-modules and PowerPoint were proofread and face-validated by students (second-year undergraduate students who were not part of this study). Content validation and technical validation were done by 2 internal and 2 external experts, following a structured checklist (Multimedia Appendix 8). Suggestions by the students and experts were incorporated wherever possible and feasible. There were certain similarities and differences between PowerPoint (2016) and the e-modules used in this study. Both platforms made use of text, images, audio, video, and animations. However, e-modules offered several interactive features over PowerPoint, such as interactive images, easy navigation with a toolbar or index, self-assessment with specific feedback, competitive scoring, and access restrictions based on performance. Additionally, e-modules enabled the tracking of student performance and the generation of reports, making them a more dynamic and data-driven learning tool compared to PowerPoint. Modules were converted into Sharable Content Object Reference Model (SCORM, 2016) format for integration into a Moodle (2016) [22].

Copyright and Intellectual Property

The developed e-modules were submitted for copyright registration and granted under L-83129/2019.

Baseline Measures

Before implementation, the data on self-directed learning readiness [23] and learning styles (Visual, Aural, Read/Write, and Kinesthetic questionnaire version 7.8) [24], visual reaction time, auditory reaction time, and critical flicker fusion frequency were collected. These measures were used to describe the learner profile and explore potential associations with learning outcomes, without influencing group allocation or intervention intensity. Prior authorization was obtained for administering these questionnaires (Multimedia Appendix 9). Although the self-directed learning readiness scale was administered during the study period (2015 - 2018), we have cited the 2021 validation study as the most current reference establishing construct validity in Indian MBBS learners. The scale itself predates this reference, and no post-2018 findings were incorporated into data collection or analysis.

Assessment Framework

The effectiveness of e-modules was based on the Kirkpatrick learning evaluation model [25]. Table 1 shows the different levels of the Kirkpatrick learning evaluation model, learning domains, and corresponding tools of evaluation used in this study.

XSL•FO
RenderX

Table . Evaluation of the effectiveness of e-modules based on the Kirkpatrick learning evaluation model.

Kirkpatrick model and domain	Taxonomy Bloom level	Taxonomy Miller pyramid	Tools for evaluation
Level 1: Reaction			
—[a]	—	—	• Structured and validated feedback questionnaire containing open- and close-ended questions
Level 2: Learning			
Cognitive	• Remember • Understand • Apply • Analyze	• Knows • Knows how • Knows how • Knows how	• MCQs[b,c] • Problem-based MCQs[b] • Problem-based MCQs[b] • Short notes[b]
Attitude	• Responding	• Shows how	• OSCE[d]
Skill	• Manipulation[e]	• Shows how	• OSCE

[a]Not applicable.

[b]End of block internal assessment.

[c]MCQs: multiple choice questions.

[d]OSCE: objective structured clinical examination.

[e]Classification by Dave, 1970 [26].

Level 1: Evaluation of Reaction

The students' reactions were evaluated using a structured and validated feedback questionnaire containing open- and closed-ended questions (Multimedia Appendix 10). The quantitative results of the evaluation of reaction and qualitative data were transcribed word by word verbatim in a Word document and analyzed thematically using a deductive coding approach (Multimedia Appendix 11).

Level 2: Evaluation of Learning

Evaluation of learning was assessed through MCQs (designed appropriately to assess "Remember," "Understand," and "Apply" levels of Bloom taxonomy) and short notes (designed to assess "Analysis" level) in the EBIA.

Standardization of EBIA

Fifteen MCQs (combination of levels of remember, understand, and application), 4 short notes (analysis level), and 1 structured essay totaled 50 marks. Question papers followed standardized blueprints and were administered within 90 minutes under invigilated conditions. The difficulty index was comparable with other EBIAs held during the academic year.

Outcome Measures and Scoring Procedures

Each internal assessment was graded by a single designated faculty member, as per the examination policy of the host university. All faculty involved in this research were exempted from any duty related to assessment and evaluation as per the examination or research ethics policy of the host university. This practice was intentionally adopted to maintain scoring consistency and minimize variability or rater-related bias (eg, halo effect). To further ensure fairness and uniformity, all answer sheets were anonymized using dummy roll numbers prior to evaluation. Additionally, the evaluating faculty member was provided with a standardized marking guide (answer key or scoring rubric) that detailed the expected key points and the corresponding marks to be awarded. This ensured transparent and objective grading across all responses.

Statistical Analysis

The data were analyzed using standard descriptive and inferential statistics. Continuous variables were summarized as means and SDs. Group comparisons were performed using independent 2-tailed *t* tests.

Effect sizes (Cohen *d*) were calculated to quantify the magnitude of differences between the groups across cognitive, psychomotor, and affective domains. Subgroup analyses were conducted to examine differential effects among high, medium, and low academic achievers. Students were categorized into low achievers (<40% in previous assessments) and high achievers (>70% in previous assessments) to analyze differential learning impacts within the control and intervention groups. A significance level of $P<.05$ was adopted for all analyses.

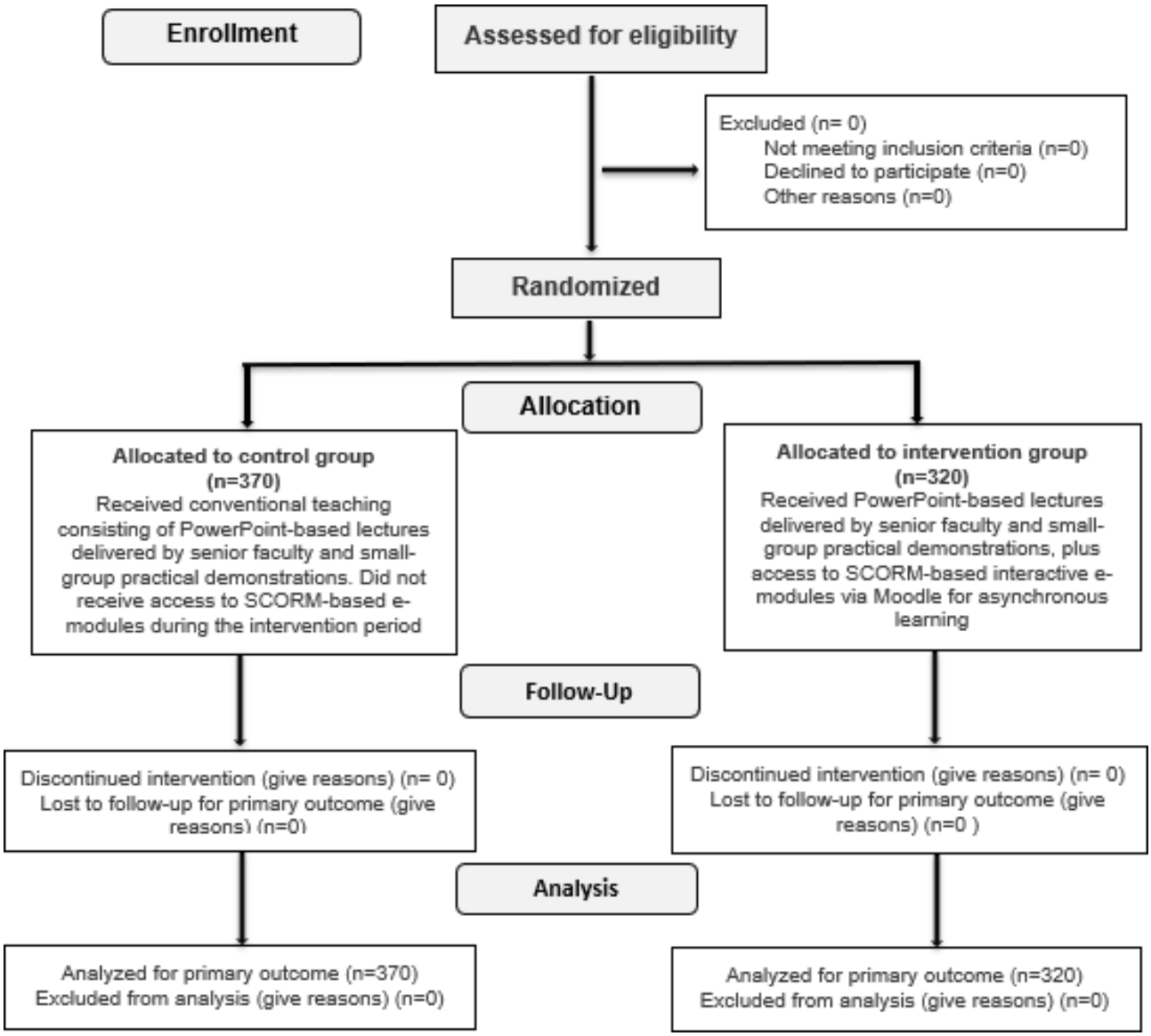
Results

Participant Characteristics

A total of 690 first-year undergraduate students participated in this study. The steps followed in participant enrollment, allocation, follow-up, and analysis are shown in Figure 2. The participants randomly assigned to the control groups were comparable in terms of baseline characteristics (Table 2). All students completed baseline assessments and the EBIAs.

XSL•FO
RenderX

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) 2025 flow diagram. Flow diagram of the progress through the phases of a randomized trial of 2 groups (ie, enrollment, intervention allocation, follow-up, and data analysis). Moodle: Modular Object-Oriented Dynamic Learning Environment; SCORM: Sharable Content Object Reference Model [27].



XSL•FO
RenderX

Table . Demographic details of study participants.

Characteristic	Overall (N=690)	Control (n=370)	Intervention (n=320)	*P* value
Gender, n (%)				.06
Male	272 (39.4)	149 (40.3)	123 (38.4)	
Female	418 (60.6)	221 (59.7)	197 (61.6)	
Attendance, mean (SD)	84.70 (8.06)	84.82 (7.93)	84.55 (8.22)	.03
Academic performance, mean (SD)	54.94 (12.20)	54.36 (12.43)	55.61 (11.92)	.02
Level of achievers, n (%)				.06
High	89 (12.9)	44 (11.9)	45 (14.1)	
Low	114 (16.5)	65 (17.6)	49 (15.3)	
Medium	487 (70.6)	261 (70.5)	226 (70.6)	
Visual reaction time (ms), mean (SD)	—[a]	284 (92)	240 (71)	.07
Auditory reaction time (ms), mean (SD)	—	236 (105)	212 (61)	.07
Critical flicker fusion frequency (Hz), mean (SD)	—	27 (3)	28 (5)	.06
Self-directed learning readiness, mean (SD)	—	156 (12)	153 (14)	.09

[a]Not applicable.

Kirkpatrick Level 1: Student Reactions

During level 2, the intervention group received e-modules in addition to traditional lectures and outperformed the control group across all cognitive domains. The intervention group showed superior performance in motor and affective domains, indicating that e-modules significantly improved student performance not only in knowledge acquisition but also in practical skills and attitudinal engagement. Table 3 shows that students in the intervention group consistently outperformed the control group across all cognitive levels. The largest differences were observed in the apply and analyze domains, indicating that the e-modules were particularly effective in promoting higher-order thinking. All comparisons were statistically significant ($P<.001$), demonstrating a meaningful advantage of the blended e-learning approach.

XSL•FO
RenderX

Table . Evaluation of the effectiveness of e-modules on Kirkpatrick model level 2: learning.

Domain[a] and level	Total (N=690)	Control (n=370)	Intervention (n=320)	Effect size (Cohen *d*)	*P* value
Knowledge, mean (SD)					<.001
Remember (direct MCQs[b], 10 marks)	5.84 (2.68)	5.26 (2.89)	6.52 (2.24)	0.48	
Understand (concept-based MCQs, 5 marks)	2.73 (1.30)	2.55 (1.13)	2.92 (1.45)	0.29	
Application (problem-based MCQs, 5 marks)	2.96 (1.15)	2.56 (1.08)	3.43 (1.06)	0.81	
Analysis (SAQ[c], 5 marks)	2.75 (1.02)	2.53 (1.11)	3.01 (0.83)	0.48	
Motor, mean (SD)					
Manipulation (OSCE[d], 6 marks)	4.31 (1.31)	4.10 (1.42)	4.55 (1.12)	0.35	<.001
Affective, mean (SD)					
Responding (OSCE, 4 marks)	2.49 (0.96)	2.03 (0.82)	3.02 (0.81)	1.21	<.001

[a]Comparison of scores across cognitive (Bloom's 4 levels), motor (manipulation, Dave level), and affective (responding, Bloom domain) between the control and intervention groups.

[b]MCQ: multiple choice question.

[c]SAQ: short answer question.

[d]OSCE: objective structured clinical examination.

Kirkpatrick Level 2: Learning Outcomes

Cognitive Outcomes

Students who received the e-modules demonstrated significantly higher performance across all cognitive levels assessed (Table 3). Effect sizes ranged from small to large: remember (*d*=0.48; 6.52 vs 5.26; *P*<.001), understand (*d*=0.29; 2.92 vs 2.55; *P*<.001), application (*d*=0.81; 3.43 vs 2.56; *P*<.001), and analysis (*d*=0.48; 3.01 vs 2.53; *P*<.001). The large effect size for application skills indicates that e-modules were particularly effective in enhancing students' problem-solving abilities.

Psychomotor Outcomes

The intervention group demonstrated superior psychomotor performance (*d*=0.35; 4.55 vs 4.10; *P*<.001), suggesting that visual demonstrations and procedural guidance in e-modules effectively enhanced motor skill acquisition.

Affective Outcomes

Affective skills showed the most substantial improvement (*d*=1.21; 3.02 vs 2.03; *P*<.001), with the largest effect size across all domains. This highlights the exceptional value of structured e-modules with visual explanations, role-modeling, and repeated exposure to professional behaviors for developing interpersonal competencies and professional attitudes—outcomes traditionally challenging to achieve through conventional teaching.

Subgroup Analysis

The impact of e-modules was found to vary across different achievement groups (Table 4). Subgroup analysis revealed that the intervention was particularly effective among medium achievers, showing significant improvements across all domains (*P*<.001 for all comparisons). High achievers in the intervention group showed significant improvements in application (*P*<.001) and affective domains (*P*<.001). Low achievers demonstrated significant improvements in understanding (*P*<.001), application (*P*<.001), analysis (*P*=.01), and affective domains (*P*<.001).

XSL•FO
RenderX

Table . Evaluation of the effectiveness of e-modules on Kirkpatrick model level 2: learning.

Domain[a] and level	High achievers			Low achievers			Medium achievers		
	Control	Intervention	*P* value	Control	Intervention	*P* value	Control	Intervention	*P* value
Knowledge, mean (SD)									
Remember (Direct MCQ[b]s, 10 marks)	5.14 (2.72)	6.18 (2.10)	.04	5.49 (2.87)	5.88 (1.96)	.39	5.22 (2.93)	6.73 (2.29)	<.001
Understand (Concept-Based MCQs, 5 marks)	2.45 (1.11)	2.58 (1.25)	.62	2.49 (1.16)	3.65 (1.60)	<.001	2.59 (1.13)	2.84 (1.41)	=.03
Application (Problem-Based MCQs, 5 marks)	2.55 (1.04)	4.38 (0.49)	<.001	2.63 (1.13)	3.63 (0.49)	<.001	2.54 (1.07)	3.20 (1.12)	<.001
Analysis (SAQ[c], 5 marks)	2.16 (1.12)	2.80 (0.84)	.003	2.62 (1.11)	3.08 (0.79)	.01	2.57 (1.11)	3.04 (0.83)	<.001
Motor, mean (SD)									
Manipulation (OSCE[d], 6 marks)	4.02 (1.47)	4.56 (1.03)	.05	4.00 (1.46)	4.47 (1.10)	.05	4.14 (1.40)	4.57 (1.15)	<.001
Affective, mean (SD)									
Responding (OSCE, 4 marks)	2.48 (0.76)	3.16 (0.74)	<.001	1.77 (0.77)	3.08 (0.76)	<.001	2.02 (0.82)	2.98 (0.83)	<.001

[a]Comparison of scores of high, medium, and low achievers between the control and intervention groups.

[b]MCQ: multiple choice question.

[c]SAQ: short answer question.

[d]OSCE: objective structured clinical examination.

Discussion

Summary of Principal Findings

In this randomized controlled study of 690 first-year MBBS students, we found that e-modules, when added to traditional teaching, significantly improved cognitive, psychomotor, and affective domain scores compared with lectures alone, with the greatest gains seen among medium achievers. The study contributes to medical education literature by demonstrating how structured, theory-driven e-modules, explicitly designed for the Indian CBME environment, can support competency development at scale. While prior studies have examined e-learning in general, few have contextualized module design within India's national curriculum reforms, digital readiness variations, and large cohort teaching realities. By aligning the e-modules with Bloom and Miller frameworks and evaluating them using the Kirkpatrick model, this study provides a comprehensive pedagogical and contextual analysis that is highly relevant for Indian medical colleges transitioning to CBME.

Interpretation of Findings Using Educational Frameworks: Integrating Bloom Taxonomy and the Miller Pyramid in e-Module Design

The decision to implement the e-modules within first-year MBBS foundational subjects (anatomy, physiology, and biochemistry) was intentional, as these disciplines carry a high cognitive load and require extensive visualization and conceptual integration. Early exposure to structured, interactive digital learning in these subjects can strengthen foundational understanding, support the development of higher-order competencies, and better prepare students for subsequent clinical learning, where the CBME framework expects early competency formation.

The development and implementation of the e-module were carefully structured around Bloom taxonomy and the Miller pyramid to ensure a progressive, competency-based learning experience. These frameworks provided a solid foundation for designing digital learning tools that promote cognitive development and practical skill acquisition, aligning well with CBME.

To ensure a structured learning progression, Bloom taxonomy guided the organization of learning objectives within the module. This structured approach ensured that students moved beyond rote memorization [28] and developed deeper conceptual understanding and analytical skills. In addition to cognitive development, the Miller pyramid was integrated to ensure that learning translated into practical competence. By blending theoretical content with multimedia demonstrations and self-assessments, the module ensured that students were not only gaining theoretical knowledge but also demonstrating their ability to apply concepts in practical scenarios.

XSL•FO
RenderX

This study further evaluated the effectiveness of e-modules in medical education using the Kirkpatrick learning evaluation model, which assesses learning outcomes across 4 levels. The study specifically examined level 1 (reaction) and level 2 (learning), focusing on cognitive, affective (attitude), and psychomotor (skill) domains. e-Modules were found to significantly enhance student engagement, knowledge acquisition, skill performance, and professional attitudes in students, reinforcing their role as an effective supplement to traditional teaching methods.

Level 1: Reaction—Student Perception of e-Modules

e-Modules effectively captured students' interest and enhanced engagement, both of which are critical for promoting active learning. The use of interactive modules allows students to learn at their own pace, integrating clinical knowledge with foundational sciences, crucial for effective learning in medical education [29]. It has also been seen in previous works that integrated modular systems are more effective than traditional methods, promoting critical thinking and innovation among students [30]. These systems allow flexibility and adaptability to individual learning needs, which can lead to improved academic performance and satisfaction [30]. Students also expressed a strong preference for a blended learning approach, advocating for e-modules as a supplement rather than a replacement for traditional teaching.

Level 2: Learning—Cognitive, Affective, and Psychomotor Outcomes

In terms of knowledge recall (remember), students in the intervention group were significantly better than those in the control group in recalling factual knowledge. However, low-achieving students did not show statistically significant improvements, likely due to difficulties in retaining information without contextual application. To enhance their learning, strategies, such as spaced repetition and adaptive quizzes, could be incorporated in e-modules to reinforce knowledge retention and provide targeted support [31].

For conceptual comprehension (understand), the e-module significantly improved comprehension for low and medium achievers, while high achievers showed minimal gains. This could be attributed to the fact that high-achieving students already have a strong conceptual foundation, making additional learning less impactful for them. To cater to this subgroup, future iterations of the module should integrate advanced case–based learning and real-world problem-solving tasks to challenge their cognitive abilities further [32]. When assessing problem-solving skills (apply), students in the intervention group showed significantly higher improvements. This finding underscores the module's effectiveness in developing clinical reasoning and problem-solving abilities. Medium and low achievers benefited the most, suggesting that structured digital learning can help bridge performance gaps and create a more equitable learning experience. The effectiveness of e-modules in promoting critical thinking and higher-order learning (analyze) was evident through significant improvements in analysis-level assessments.

The impact of e-modules varied across different student groups. Medium achievers showed the most substantial gains across all cognitive levels. Their foundational knowledge gaps were likely addressed through the structured nature of the e-modules, allowing them to engage more effectively with new learning strategies. Low achievers demonstrated notable improvements in comprehension and application, indicating that e-modules serve as effective platforms for conceptual learning. In contrast, high achievers exhibited smaller relative gains, possibly due to their prior familiarity with the concepts covered. This finding aligns with past works, which also found that digital learning tools have the most impact on students with moderate baseline proficiency, while high-achieving students experience diminishing returns [33,34]. Students with moderate proficiency are often seen to benefit from digital learning tools as they enhance essential digital skills and cater to their specific learning needs, which are crucial for academic success [33,34].

Conversely, high-achieving students may already possess the skills and motivation to excel without the additional support of digital tools, potentially diminishing their relative impact [35,36]. Also, self-directed and problem-based learning approaches are more effective for high-achieving medical students, as they provide greater cognitive flexibility and autonomy compared to structured digital modules [37]. Studies have emphasized that digital learning tools designed for high achievers should incorporate more advanced problem-solving tasks, adaptive difficulty levels, and research-oriented challenges to maintain engagement and maximize learning outcomes [38-40].

Effectiveness of e-Modules Across Progressive Levels (the Miller Pyramid)

This study's findings align with the Miller framework, demonstrating that e-modules were particularly beneficial in progressing students from basic knowledge acquisition to higher-order thinking and practical application.

Knowledge Acquisition (Knows)—Remembering and Understanding

Students in the intervention group outperformed those in the control group in both of these domains. e-Modules enhanced students' ability to recall and comprehend medical concepts more effectively than traditional lectures alone. This aligns with previous studies, where digital learning tools significantly improved retention and understanding in medical students by integrating multimedia, interactive elements, and adaptive learning [40,41].

Additionally, the effectiveness of e-learning in knowledge acquisition has been reported in studies on flipped classrooms [42]. Students who engaged with online modules before in-person sessions demonstrated better long-term retention compared to those relying solely on lectures.

Application of Knowledge (Knows How)—Application and Analysis

Low achievers and high achievers have performed significantly well in the intervention group, suggesting that e-modules facilitated deeper learning and cognitive processing, allowing

XSL•FO
RenderX

students to synthesize and apply knowledge to real-world scenarios. It was similarly found in past studies that interactive e-learning modules significantly improved students' problem-solving skills in clinical case studies [43-46]. Digital resources, when designed with active learning strategies, such as those integrated into these e-modules, enhance analytical reasoning, reduce cognitive overload [47], and improve knowledge retention and conceptual understanding [48,49].

Demonstration of Skills (Shows How)—Psychomotor Domain

The motor skills of students assessed using OSCEs were also significantly higher in the intervention group. The results suggest that interactive, visual, and simulation-based e-modules are effective in improving hands-on procedural skills, which is an essential aspect of medical training. The finding concurs with other studies that the use of computer technology enables learners to be active in the learning process, construct knowledge, develop problem-solving skills, and analyze the given situation to discover alternative solutions [50]. Studies have also found that virtual simulations and video-based learning improve students' psychomotor skills, particularly in clinical scenarios where real-life practice opportunities are limited [51,52]. Simulation-based e-learning has been found as effective as traditional hands-on training in improving procedural skills [53]. Moreover, the integration of e-modules with interactive video demonstrations likely enhanced motor learning [54].

Professionalism and Attitudes (Does)—Affective Domain

The affective domain is often neglected in traditional didactic education. Studies suggest that interactive, self-paced learning methods enhance students' motivation and self-directed learning behaviors [55], developing greater autonomy in medical education [56]. The incorporation of self-assessment, feedback, and real-life clinical scenarios in e-modules likely contributed to this improvement.

Implications for CBME Implementation

The successful implementation of the e-module demonstrates that technology can enhance medical education by making complex concepts more accessible and promoting independent learning. One of the key takeaways is the role of technology as a catalyst for learning. The e-module supports self-directed learning by allowing students to engage with content at their own pace. This aligns with the growing emphasis on student autonomy and flexible learning within medical education [56]. Another critical aspect is the scalability and adaptability of the e-module. Its modular design allows for easy adaptation across various medical disciplines.

Recommendations for Future Research

Future e-module designs should include differentiated instructions to address varied learning needs. CBME can benefit from hybrid models that combine e-modules with case-based learning and simulations, developing deeper engagement and critical thinking, particularly among high achievers. Adding gamified elements and competitive assessments may help sustain motivation across all student levels. Future research should extend evaluation to Kirkpatrick levels 3 and 4, examining behavior change during clinical postings and long-term impact on competency attainment in later MBBS phases. The integration of advanced adaptive learning, analytics dashboards, and gamification may further strengthen national-level CBME implementation.

Limitations of the Study

This study has several limitations. The study was conducted between 2015 and 2018, and some technologies used (such as Adobe Flash) are now deprecated. While this does not affect the internal validity of the study, it limits immediate scalability of the exact modules in their original form. However, as all modules were exported into Moodle-compatible SCORM packages, the instructional design and learning architecture remain fully transferable to modern authoring tools. Additionally, the literature cited in this study spans both the study period and recent years, consistent with standard academic practice for studies reporting older datasets. Contemporary sources were intentionally included to ensure that the interpretation and implications of the findings align with current developments in CBME, e-learning research, and medical education frameworks. These later references did not inform the original study design or data collection but were used solely for contextualization. It involved only first-year medical students, limiting the generalizability of the findings. The focus was on short-term outcomes (Kirkpatrick levels 1 and 2), without examining long-term behavior changes or clinical impact. While most students had access to devices and the internet, differences in internet speed or device quality were not fully considered, which may have affected learning. Instructor-related bias is also possible, as differences in teaching style could have influenced results despite standardized content. Finally, while student feedback was gathered, detailed engagement data, such as time spent, interaction frequency, or assessment attempts, were not thoroughly analyzed.

Conclusion

Well-designed e-modules, when combined with traditional teaching, can significantly enhance learning by using clear goals aligned with Bloom taxonomy and the Miller pyramid for undergraduate medical students. They helped students not only remember and understand content but also apply it and develop stronger attitudes and skills. Students across performance levels benefited, with average performers improving the most. Positive feedback and better test scores indicate that e-modules are a valuable addition to medical education, though they work best as supplements rather than replacements. These tools can be expanded to other subjects and learner levels to support more flexible and engaging learning. With CBME now uniformly implemented across India, scalable e-modules aligned with national competencies could promote more equitable learning, especially in resource-limited institutions. The findings support national strategies that use structured e-learning to strengthen and standardize instructional quality.

XSL•FO
RenderX

Funding

The authors declared no financial support was received for this work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The screenshots of implementation of e-modules.
[PDF File, 492 KB - humanfactors_v13i1e84339_app1.pdf]

Multimedia Appendix 2
Sample of research randomizer results.
[PDF File, 62 KB - humanfactors_v13i1e84339_app2.pdf]

Multimedia Appendix 3
Knowledge, Attitude, and Practice towards e-learning.
[PDF File, 163 KB - humanfactors_v13i1e84339_app3.pdf]

Multimedia Appendix 4
List of e-modules addressing specific competency-based medical education (CBME) competencies and corresponding domains.
[PDF File, 208 KB - humanfactors_v13i1e84339_app4.pdf]

Multimedia Appendix 5
Overview of the development of e-modules.
[PDF File, 84 KB - humanfactors_v13i1e84339_app5.pdf]

Multimedia Appendix 6
Screenshot of steps of creating an e-module.
[PDF File, 476 KB - humanfactors_v13i1e84339_app6.pdf]

Multimedia Appendix 7
E module assessment questions.
[PDF File, 1194 KB - humanfactors_v13i1e84339_app7.pdf]

Multimedia Appendix 8
Tool used for validation exercise.
[PDF File, 667 KB - humanfactors_v13i1e84339_app8.pdf]

Multimedia Appendix 9
Request for permission to use self-directed learning readiness (SDLR) scale and Visual, Aural, Read/Write, and Kinesthetic (VARK) questionnaire.
[PDF File, 164 KB - humanfactors_v13i1e84339_app9.pdf]

Multimedia Appendix 10
Feedback form for e-module.
[PDF File, 449 KB - humanfactors_v13i1e84339_app10.pdf]

Multimedia Appendix 11
Results of level 1 evaluation of the Kirkpatrick model.
[PDF File, 155 KB - humanfactors_v13i1e84339_app11.pdf]

Checklist 1
CONSORT checklist.
[DOCX File, 37 KB - humanfactors_v13i1e84339_app12.docx]

XSL•FO
RenderX

References

1. Nema N, Srivastava R, Bose S. An insight into competency-based undergraduate curriculum and its application in ophthalmology. J clin ophthalmol res 2022;10(2):91-97. [doi: 10.4103/jcor.jcor_7_22]
2. Inamdar P, Inamdar DP, Anuradha B, Undi M. Elective module in undergraduate medical curriculum: implementation and its perception. Curricula 2025;4(1):287-302. [doi: 10.17509/curricula.v4i1.80717]
3. Elyorbek R, Sojida U, Gulnora S, et al. Recommendations to develop teaching methods in medical sciences [Article in Russian]. Int J Med Sci Public Health Res 2024;5(10):24-37. [doi: 10.37547/ijmsphr/Volume05Issue10-04]
4. Jay R, Davenport C, Patel R. Clinical reasoning-the essentials for teaching medical students, trainees and non-medical healthcare professionals. Br J Hosp Med (Lond) 2024 Jul 30;85(7):1-8. [doi: 10.12968/hmed.2024.0052] [Medline: 39078902]
5. Mariani AW, Terra RM, Pêgo-Fernandes PM. E-learning: from useful to indispensable tool. Sao Paulo Med J 2012;130(6):357-359. [doi: 10.1590/s1516-31802012000600001] [Medline: 23338731]
6. Omprakash K, Kumar AP, Sagar V, Sathiyasekaran BWC, Ramaswamy P. Medical Education/Original Article needs assessment among students of health professions education for the introduction of E-learning in a South Indian Tertiary Care University. Indian J Physiol Pharmacol 2019;63(1):42-48 [FREE Full text]
7. Kumar Basak S, Wotto M, Bélanger P. E-learning, M-learning and D-learning: conceptual definition and comparative analysis. E-Learn Digit Media 2018 Jul;15(4):191-216. [doi: 10.1177/2042753018785180]
8. Omprakash A, Kumar AP, Kuppusamy M, Sathiyasekaran BWC, Ravinder T, Ramaswamy P. Validation of metacognitive awareness inventory from a private medical university in India. J Educ Health Promot 2021;10:324. [doi: 10.4103/jehp.jehp_39_21] [Medline: 34761010]
9. Zhang Z, Cao T, Shu J, Liu H. Identifying key factors affecting college students' adoption of the e-learning system in mandatory blended learning environments. Interact Learn Environ 2022 Jul 4;30(8):1388-1401. [doi: 10.1080/10494820.2020.1723113]
10. Prabu Kumar A, Omprakash A, Chokkalingam Mani PK, et al. E-learning and E-modules in medical education—A SOAR analysis using perception of undergraduate students. PLoS ONE 2023;18(5):e0284882. [doi: 10.1371/journal.pone.0284882] [Medline: 37205679]
11. Abdulrahaman MD, Faruk N, Oloyede AA, et al. Multimedia tools in the teaching and learning processes: a systematic review. Heliyon 2020 Nov;6(11):e05312. [doi: 10.1016/j.heliyon.2020.e05312] [Medline: 33195834]
12. Vaona A, Banzi R, Kwag KH, et al. E-learning for health professionals. Cochrane Database Syst Rev 2018 Jan 21;1(1):CD011736. [doi: 10.1002/14651858.CD011736.pub2] [Medline: 29355907]
13. Darejeh A, Mashayekh S, Marcus N. Cognitive-based methods to facilitate learning of software applications via E-learning systems. Cogent Education 2022 Dec 31;9(1). [doi: 10.1080/2331186X.2022.2082085]
14. Levidze M. Mapping the research landscape: a bibliometric analysis of e-learning during the COVID-19 pandemic. Heliyon 2024 Jul 15;10(13):e33875. [doi: 10.1016/j.heliyon.2024.e33875] [Medline: 39027537]
15. Mahdavi Ardestani SF, Adibi S, Golshan A, Sadeghian P. Factors influencing the effectiveness of e-learning in healthcare: a fuzzy ANP study. Healthcare (Basel) 2023 Jul 16;11(14):2035. [doi: 10.3390/healthcare11142035] [Medline: 37510477]
16. Adams NE. Bloom's taxonomy of cognitive learning objectives. J Med Libr Assoc 2015 Jul;103(3):152-153. [doi: 10.3163/1536-5050.103.3.010] [Medline: 26213509]
17. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990 Sep;65(9 Suppl):S63-S67. [doi: 10.1097/00001888-199009000-00045] [Medline: 2400509]
18. Mirbahai L, Adie JW. Applying the utility index to review single best answer questions in medical education assessment. Arch Epid Pub Health 2020;2(1). [doi: 10.15761/AEPH.1000113]
19. Bates R. A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. Eval Program Plann 2004 Aug;27(3):341-347. [doi: 10.1016/j.evalprogplan.2004.04.011]
20. Heydari MR, Taghva F, Amini M, Delavari S. Using Kirkpatrick's model to measure the effect of a new teaching and learning methods workshop for health care staff. BMC Res Notes 2019 Jul 10;12(1):388. [doi: 10.1186/s13104-019-4421-y] [Medline: 31292006]
21. Thomas PA, Kern DE, Hughes MT, Tackett SA, Chen BY. Curriculum Development for Medical Education: A Six-Step Approach: Johns Hopkins University Press; 2022. [doi: 10.56021/9781421444116]
22. Ruiz JG, Mintzer MJ, Issenberg SB. Learning objects in medical education. Med Teach 2006 Nov;28(7):599-605. [doi: 10.1080/01421590601039893] [Medline: 17594550]
23. Kumar AP, Omprakash A, Mani PKC, et al. Validation of internal structure of Self-Directed Learning Readiness Scale among Indian medical students using factor analysis and the structural equation modelling approach. BMC Med Educ 2021 Dec 11;21(1):614. [doi: 10.1186/s12909-021-03035-6] [Medline: 34895214]
24. Fleming ND, Mills C. Not another inventory, rather a catalyst for reflection. J Educ Dev 1992 Jun;11(1):137-155. [doi: 10.1002/j.2334-4822.1992.tb00213.x]
25. Yardley S, Dornan T. Kirkpatrick's levels and education "evidence". Med Educ 2012 Jan;46(1):97-106. [doi: 10.1111/j.1365-2923.2011.04076.x] [Medline: 22150201]
26. Dave RH. Psychomotor levels. In: Armstrong RJ, editor. Developing and Writing Behavioral Objectives: Educational Innovators Press; 1970:20-21.

XSL•FO
RenderX

27. Hopewell S, Chan AW, Collins GS, et al. CONSORT 2025 statement: Updated guideline for reporting randomised trials. BMJ 2025 Apr 14;389:e081123. [doi: 10.1136/bmj-2024-081123] [Medline: 40228833]
28. Bankar MN, Bankar NJ, Singh BR, Bandre GR, Shelke YP. The role of E-content development in medical teaching: how far have we come? Cureus 2023 Aug;15(8):e43208. [doi: 10.7759/cureus.43208] [Medline: 37692742]
29. Kausar U, Noor MSA, Sohail MU, Rana MH. Evaluation of integrated modular & traditional curriculum learning environment in medical education. J Univ Med Dent Coll 2024;15(4). [doi: 10.37723/jumdc.v15i4.1061]
30. Price DW, Wang T, O'Neill TR, et al. The effect of spaced repetition on learning and knowledge transfer in a large cohort of practicing physicians. Acad Med 2025 Jan 1;100(1):94-102. [doi: 10.1097/ACM.0000000000005856] [Medline: 39250798]
31. Choe AI, Woodard S, Thompson BM, Walter V, Fotos JS, Kasales CJ. Spaced education: randomized trial comparing learning efficiency of the adaptive versus nonadaptive spaced education systems among radiology residents. J Am Coll Radiol 2022 Jun;19(6):706-710. [doi: 10.1016/j.jacr.2022.03.010] [Medline: 35472369]
32. Jiang D, Huang D, Wan H, et al. Effect of integrated case-based and problem-based learning on clinical thinking skills of assistant general practitioner trainees: a randomized controlled trial. BMC Med Educ 2025 Jan 14;25(1):62. [doi: 10.1186/s12909-025-06634-9] [Medline: 39806361]
33. Getenet S, Cantle R, Redmond P, Albion P. Students' digital technology attitude, literacy and self-efficacy and their effect on online learning engagement. Int J Educ Technol High Educ 2024;21(1). [doi: 10.1186/s41239-023-00437-y]
34. Romi IM. Digital skills impact on university students' academic performance: an empirical investigation. Edelweiss Appl Sci Technol 2024;8(5):2126-2141. [doi: 10.55214/25768484.v8i5.1965]
35. Petratos S, Botha R, Mtshabe M. Digital skills' impact on the use of digital learning tools. Presented at: Proceedings of the 23rd European Conference on e-Learning - ECEL 2024; Oct 24-25, 2024 URL: https://papers.academic-conferences.org/index.php/ecel/article/view/2995 [accessed 2026-02-27] [doi: 10.34190/ecel.23.1.2995]
36. Aulakh J, Wahab H, Richards C, Bidaisee S, Ramdass PVAK. Self-directed learning versus traditional didactic learning in undergraduate medical education: a systemic review and meta-analysis. BMC Med Educ 2025 Jan 16;25(1):70. [doi: 10.1186/s12909-024-06449-0] [Medline: 39815233]
37. Drigas A, Karyotaki M. Online and other ICT-based assessment tools for problem-solving skills. Int J Emerg Technol Learn 2016;11(4):56-60. [doi: 10.3991/ijet.v11i04.5339]
38. Ifenthaler D, Gibson DC, Zheng L. The dynamics of learning engagement in challenge-based online learning. Presented at: 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT); Jul 9-13, 2018. [doi: 10.1109/ICALT.2018.00049]
39. McGee RG, Wark S, Mwangi F, et al. Digital learning of clinical skills and its impact on medical students' academic performance: a systematic review. BMC Med Educ 2024 Dec 18;24(1):1477. [doi: 10.1186/s12909-024-06471-2] [Medline: 39696150]
40. Subramanian A, Timberlake M, Mittakanti H, Lara M, Brandt ML. Novel educational approach for medical students: improved retention rates using interactive medical software compared with traditional lecture-based format. J Surg Educ 2012;69(2):253-256. [doi: 10.1016/j.jsurg.2011.12.007] [Medline: 22365876]
41. Delungahawatta T, Dunne SS, Hyde S, et al. Advances in e-learning in undergraduate clinical medicine: a systematic review. BMC Med Educ 2022 Oct 7;22(1):711. [doi: 10.1186/s12909-022-03773-1] [Medline: 36207721]
42. Förster M, Maur A, Weiser C, Winkel K. Pre-class video watching fosters achievement and knowledge retention in a flipped classroom. Comput Educ 2022 Apr;179:104399. [doi: 10.1016/j.compedu.2021.104399]
43. Ertl S, Wadowski PP, Löffler-Stastka H. Improving students' performance via case-based e-learning. Front Med (Lausanne) 2024;11:1401124. [doi: 10.3389/fmed.2024.1401124] [Medline: 39835111]
44. Villatoro T, Lackritz K, Chan JSY. Case-based asynchronous interactive modules in undergraduate medical education. Acad Pathol 2019;6:2374289519884715. [doi: 10.1177/2374289519884715] [Medline: 31700991]
45. Stollar F, Cerutti B, Aujesky S, Scherly D, Nendaz M, Galetto-Lacour A. E-learning modules to improve clinical reasoning and practice: a prospective comparative study. MedEdPublish (2016) 2023;13:39. [doi: 10.12688/mep.19449.2] [Medline: 38813067]
46. Skulmowski A, Xu KM. Understanding cognitive load in digital and online learning: a new perspective on extraneous cognitive load. Educ Psychol Rev 2022 Mar;34(1):171-196. [doi: 10.1007/s10648-021-09624-7]
47. Deslauriers L, McCarty LS, Miller K, Callaghan K, Kestin G. Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. Proc Natl Acad Sci U S A 2019 Sep 24;116(39):19251-19257. [doi: 10.1073/pnas.1821936116] [Medline: 31484770]
48. Do HN, Do BN, Nguyen MH. 3How do constructivism learning environments generate better motivation and learning strategies? The Design Science Approach. Heliyon 2023 Dec;9(12):e22862. [doi: 10.1016/j.heliyon.2023.e22862] [Medline: 38125439]
49. Haleem A, Javaid M, Qadri MA, Suman R. Understanding the role of digital technologies in education: a review. Sustainable Oper Comput 2022;3(1):275-285. [doi: 10.1016/j.susoc.2022.05.004]
50. Clerkin R, Patton D, Moore Z, Nugent L, Avsar P, O'Connor T. What is the impact of video as a teaching method on achieving psychomotor skills in nursing? A systematic review and meta-analysis. Nurse Educ Today 2022 Apr;111:105280. [doi: 10.1016/j.nedt.2022.105280] [Medline: 35139443]

XSL•FO
RenderX

51. İsmailoğlu EG, Orkun N, Eşer İ, Zaybak A. Comparison of the effectiveness of the virtual simulator and video-assisted teaching on intravenous catheter insertion skills and self-confidence: a quasi-experimental study. Nurse Educ Today 2020 Dec;95:104596. [doi: 10.1016/j.nedt.2020.104596] [Medline: 33002745]
52. Elendu C, Amaechi DC, Okatta AU, et al. The impact of simulation-based training in medical education: a review. Medicine (Baltimore) 2024 Jul 5;103(27):e38813. [doi: 10.1097/MD.0000000000038813] [Medline: 38968472]
53. Shahrezaei A, Sohani M, Taherkhani S, Zarghami SY. The impact of surgical simulation and training technologies on general surgery education. BMC Med Educ 2024 Nov 13;24(1):1297. [doi: 10.1186/s12909-024-06299-w] [Medline: 39538209]
54. Altinpulluk H, Kilinc H, Alptekin G, Yildirim Y, Yumurtaci O. Self-directed learning and intrinsic motivation levels in MOOCs. Open Praxis 2023 Jul 20;15(2):149-161. [doi: 10.55982/openpraxis.15.2.556]
55. Jin X, Jiang Q, Xiong W, Pan X, Zhao W. Using the online self-directed learning environment to promote creativity performance for university students. Educ Technol Soc 2022;25(2):130-147 [FREE Full text]
56. Liu F, Niu P, Yu S, et al. Investigating students' perceptions of the medical education environment and learning autonomy in blended learning in pediatrics: a cross-sectional study. BMC Med Educ 2024 Dec 18;24(1):1464. [doi: 10.1186/s12909-024-06490-z] [Medline: 39696173]

Abbreviations

CBME: competency-based medical education
EBIA: end-of-block internal assessment
MCQ: multiple choice question
Moodle-SCORM: Modular Object-Oriented Dynamic Learning Environment

Edited by J Waterson; submitted 18.Sep.2025; peer-reviewed by A Karthikeyan, Y Asada; revised version received 11.Dec.2025; accepted 12.Jan.2026; published 07.Apr.2026.

Please cite as:
Prabu Kumar A, R P, K M, Omprakash A, Chokkalingam Mani PK, Sathiyasekaran BWC, K N M
E-Learning Modules Based on Bloom Taxonomy and the Miller Pyramid for First-Year Indian Medical Students: Randomized Controlled Study in Medical Education
JMIR Hum Factors 2026;13:e84339
URL: https://humanfactors.jmir.org/2026/1/e84339
doi:10.2196/84339

© Archana Prabu Kumar, Padmavathi R, Maheshkumar K, Abirami Omprakash, Prabu Kumar Chokkalingam Mani, B W C Sathiyasekaran, Maruthy K N. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 7.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Gamified Formative Assessment in a First-Year Medical Morphology Course: Qualitative Study

Álvaro Herrera Alcaíno[1], MSc, MD; Sofía Paz Araya Figueroa[2], BSc; Valentina Francisca Riquelme Vilches[1], BSc; Eduardo Vargas Zurita[2], BSc; Renata Valentina Cozzi Ahumada[2], BSc; Camila Riquelme Bahamondes[3], MSc; Carlos Alberto Rosas Chuñil[2], PhD

[1]Faculty of Medicine, San Sebastián University, Lota 2465, Providencia, Santiago, Metropolitana, Chile

[2]Faculty of Medicine, San Sebastián University, Valdivia, Los Ríos, Chile

[3]Faculty of Medicine, San Sebastián University, Concepción, Biobío, Chile

Corresponding Author:
Álvaro Herrera Alcaíno, MSc, MD
Faculty of Medicine, San Sebastián University, Lota 2465, Providencia, Santiago, Metropolitana, Chile

Abstract

Background: Formative assessment plays a central role in medical education by offering low-risk practice and timely feedback, particularly in visually and practically demanding courses, such as morphology. Gamified digital tools can foster active learning; however, their educational value relies on pedagogical integration rather than technology alone. The Technological Pedagogical Content Knowledge (TPACK) model provides a framework for understanding this relationship.

Objective: This study explored students' and a teacher's perceptions of Kahoot! as a gamified formative assessment tool in a first-year medical morphology course, identifying facilitators, tensions, and perceived educational value through the TPACK framework.

Methods: This qualitative, exploratory study adopted an interpretive phenomenological orientation. It was conducted within the framework of the RAÍCES 2 (Red de ayudantes de investigación comprometidos con la educación en salud; Network of Student Researchers Committed to Health Professions Education) educational research project at Universidad San Sebastián (Valdivia, Chile). Kahoot! was implemented at the end of class sessions to simulate the visual and time demands of authentic practical examinations. Data collection included 2 online focus groups with first-year medical students (n=14) and 1 semistructured interview with the course teacher (n=1). Transcripts were analyzed using deductive coding informed by the TPACK framework and the project's analytical axes, combined with reflexive thematic analysis following Braun and Clarke. ATLAS.ti was used to support the analysis.

Results: Six interrelated themes emerged: (1) techno-disciplinary alignment, reflecting the value of authentic, high-resolution anatomical images and time constraints mirroring practical examinations; (2) Kahoot! as a formative closure, serving as an exit check that guided self-assessment and the identification of learning gaps; (3) from score to learning, highlighting the importance of immediate postquestion instructor feedback explaining both correct and incorrect answers; (4) motivation and socioemotional climate, whereby gamification enhanced engagement but triggered evaluative anxiety and leaderboard frustration, which were buffered by participant anonymity; (5) implementation conditions, mediated by connectivity, device availability, and time allocation; and (6) projection and improvement, highlighting the need for institutional question repositories and systematic curriculum integration.

Conclusions: Kahoot! effectively functions as a formative assessment strategy when intentionally aligned with disciplinary and practical evaluative demands. Its educational value is not inherent to the software but is determined by explicit pedagogical mediation and careful attention to students' socioemotional dynamics.

(JMIR Hum Factors 2026;13:e90651) doi:10.2196/90651

KEYWORDS

medical education; formative assessment; gamification; Kahoot!; qualitative research; Technological Pedagogical Content Knowledge; TPACK; morphology; anatomy

XSL•FO
RenderX

Introduction

Formative Assessment and Gamification in Medical Education

In medical education, formative assessment plays a key role by allowing students to practice in academically low-risk environments and to receive timely feedback. These opportunities are particularly relevant for preparing students for high-stakes assessments, such as anatomical structure recognition exercises, which are common in courses with a strong practical and visual component.

In this context, the use of digital strategies for formative purposes has steadily increased in medical education. A recent review published in JMIR Medical Education highlights the expansion of digital teaching [1], noting that gamification tools, such as web-based quizzes, can help synthesize essential content [2] and support engaging synchronous learning activities for students [3]. Available evidence suggests that, when appropriately implemented, these tools are associated with higher levels of participation and satisfaction, thereby supporting learning processes.

Anatomical Training Demands and the Technological Pedagogical Content Knowledge Framework

In September 2025, an educational activity was implemented at Universidad San Sebastián (Valdivia, Chile) for first-year medical students enrolled in the morphology I course. The intervention consisted of an end-of-class formative assessment activity using the gamification tool Kahoot! It was intentionally designed to simulate the specific conditions, time constraints, and pressure encountered during authentic practical assessments in the course, such as gymkhanas (station-based examinations focused on timed anatomical structure identification).

Although Kahoot! is widely used for reviewing cross-cutting content [2], its effectiveness largely depends on context and pedagogical design. Previous studies have described this platform as a motivating tool in health professions education [3,4]; however, its impact does not reside solely in the technology itself, but rather in how it is pedagogically integrated. From the theoretical perspective of Technological Pedagogical Content Knowledge (TPACK), the effective use of educational technologies requires a balanced articulation among technology, pedagogy, and disciplinary content [5,6].

Despite its widespread use, evidence about how this type of tool is experienced in highly visual disciplines, such as morphology, remains limited. In particular, there is limited understanding of students' and teachers' perceptions regarding its contribution to learning, as well as the facilitators and barriers to its implementation. This study explored students' and a teacher's experiences with Kahoot! as a formative assessment tool in morphology I. Using qualitative methods and the TPACK framework, we identified implementation barriers, facilitators, and perceived learning outcomes.

In addition, this study extends beyond describing participants' perceptions by identifying pedagogical conditions and transferable design principles for gamified formative assessment in visually intensive medical courses. Interpreting these findings through the TPACK framework provides guidance for designing technology-enhanced formative assessment activities that align disciplinary content, pedagogy, and digital tools.

The aim of this study was to explore students' and the teacher's perceptions regarding the use of Kahoot! as a gamified formative assessment tool in a first-year medical morphology course. Specifically, this study sought to identify the facilitators, tensions, and perceived educational value of this tool when interpreted through the TPACK framework. Ultimately, we intend to provide transferable pedagogical conditions and design principles to guide educators in integrating technology-enhanced assessments within visually intensive medical disciplines.

Methods

Study Design and Context

This study followed a qualitative, exploratory design informed by an interpretive phenomenological orientation, which guided the study's focus on understanding participants' lived experiences and the meanings they attributed to the use of Kahoot! within the morphology course. Rather than adopting a specific phenomenological method, this orientation served as an epistemological framework to support an interpretive analysis of participants' accounts. The study was conducted within the framework of the RAÍCES 2 (Red de ayudantes de investigación comprometidos con la educación en salud; Network of Student Researchers Committed to Health Professions Education) project (2025) and reports findings derived from an educational intervention implemented in the morphology I course in the Medicine program at Universidad San Sebastián, Valdivia campus, Chile. Data collection took place in September 2025.

Participants and Recruitment

Participants were first-year medical students enrolled in the morphology I course, which had a total enrollment of 76 students during the intervention period. A voluntary sampling strategy was used. Inclusion criteria were as follows: (1) enrollment in the Medicine program at Universidad San Sebastián, Valdivia campus; (2) registration in the morphology I course in September 2025; and (3) participation in the Kahoot!-based educational activity. No exclusion criteria were applied.

After the intervention, students who met the inclusion criteria were invited in person during class sessions to participate in the study. In addition, the invitation was disseminated through social media, along with a link to a participation form. In total, 14 students agreed to participate, representing approximately 11% of the eligible student cohort. As participation was voluntary, this proportion reflects a self-selected sample of students willing to share their experiences.

Given the qualitative and exploratory nature of the study, the sample was considered adequate for the study's aims, providing sufficient depth and diversity of experiences to support an interpretive understanding of participants' perspectives. Consistent with qualitative research, emphasis was placed on the richness and relevance of the data rather than on statistical representativeness.

XSL•FO
RenderX

Prior to data collection, participants were invited to attend an in-person meeting at Universidad San Sebastián, Valdivia campus, during which the objectives of the study were explained and written informed consent was obtained. The volunteers were then assigned to 2 focus groups, each composed of 7 students. In addition, an individual semistructured interview was conducted with the course teacher, who was responsible for implementing the intervention, to complement the students' perspectives.

Data Collection

Data collection was conducted remotely in September 2025 through synchronous sessions held via Microsoft Teams. Two focus groups with students and 1 individual interview with the course teacher were held.

All sessions were semistructured and guided by interview protocols developed in advance by the research team. Separate guides were designed for students and for the teacher in order to capture their distinct perspectives.

The student focus group guide explored topics such as perceptions of Kahoot! as a formative assessment tool, its influence on learning and motivation, facilitators and barriers to its use, and its alignment with the visual and evaluative demands of the course.

The teacher interview guide focused on pedagogical intentions, implementation decisions, perceived educational value, challenges encountered, and reflections on the integration of technology, pedagogy, and content. The development of both guides was informed by the TPACK framework, ensuring alignment with the study's theoretical perspective. The full interview and focus group guides are provided in Multimedia Appendix 1.

Each focus group was moderated by a member of the research team. The first focus group was facilitated by EVZ, the second by VFRV, and the teacher interview was conducted by SPAF. None of the moderators had a direct teaching relationship with the participants in order to minimize potential biases related to academic hierarchies. The sessions lasted approximately 60 minutes. No field notes were taken. With participants' informed consent, all sessions were audio recorded for subsequent analysis. The recordings were transcribed verbatim; only the transcription of the first focus group was supported by AI-assisted tools. Subsequently, all transcripts were manually reviewed and corrected by members of the research team to ensure accuracy and fidelity to the original recordings.

Data Analysis

The analysis was conducted within an interpretive qualitative framework consistent with the study's phenomenological orientation, acknowledging that meaning is coconstructed through participants' accounts and researchers' interpretive engagement.

Data from the focus groups and the teacher interview were analyzed using a hybrid abductive analytic approach that combined sensitizing concepts derived from the TPACK framework (technology, pedagogy, and content) and the analytical axes of the RAÍCES project (motivation, facilitators, barriers, perceived learning, and transfer), together with an inductive, reflexive thematic analysis following Braun and Clarke [7]. Rather than functioning as a closed coding template, these theoretical dimensions served as an initial analytical scaffold to support familiarization with the data and the preliminary organization of participants' accounts. This method was selected for its theoretical flexibility and its compatibility with interpretive qualitative approaches, allowing for the identification of patterns of shared meaning while preserving sensitivity to participants' lived experiences. This process enabled deeper analysis and the integration of both theoretical categories and meanings constructed from participants' lived experiences.

Throughout the analytic process, the analysis followed a hybrid abductive approach in which theoretical concepts and data-driven insights informed one another iteratively. The TPACK dimensions and the RAÍCES analytical axes were used as an initial organizing structure, a sensitizing framework that provided a preliminary scaffold for indexing participants' accounts, rather than as a closed coding scheme. Building on this structure, inductive open coding was conducted to capture meanings not anticipated by the framework, such as the salience of the visual component and the emotional ambivalence of competition. Theme construction was therefore reflexive and primarily inductive: deductively derived categories were retained only when supported by the data, and emergent codes that exceeded the framework were preserved and developed into themes in their own right. Throughout this process, the research team moved iteratively between the empirical material and the theoretical framework, so that the final themes reflect participants' lived experiences while remaining interpretable through the TPACK lens. The full analytical workflow is summarized in Multimedia Appendix 2, which illustrates the iterative integration of deductive template coding and inductive reflexive thematic analysis, showing how theoretical categories served as an initial analytic scaffold while allowing themes to emerge from the data.

This hybrid analytic process generated the 6 thematic outcomes described in the *Results* section: "Techno-Disciplinary Alignment and Visuality," "Kahoot! as a Formative Closure: Rapid Check, Review, and Gap Diagnosis," "From Score to Learning: Feedback as a Condition for Being Truly Formative," "Motivation and socioemotional climate: between play, competition, and psychological safety," "Conditions of implementation: facilitators and barriers situated in the course context," and "Projection and improvement: from a single activity to a sustainable and transferable strategy."

The analysis was conducted collaboratively by members of the research team. The ATLAS.ti (Lumivero, LLC) software was used as a support tool for the management, organization, and coding of qualitative data.

Ethical Considerations

The study was approved by the institutional scientific ethics committee (CEC; protocol number 40 - 25, approved on June 25, 2025). Participation was voluntary, and all participants signed an informed consent form prior to data collection. Participants were also informed that they could withdraw from

XSL•FO
RenderX

the study at any time, without any academic or other consequences.

The collected data were accessible exclusively to the research team and were stored in private folders within the institution's secure cloud storage system (OneDrive). To ensure confidentiality, participants' names were replaced with pseudonyms in transcripts and in the presentation of results, with identifying information kept under restricted access.

Researcher Positionality and Reflexivity

To ensure qualitative rigor, the research team engaged in continuous reflexivity throughout the study design, data collection, and analysis. The primary analysis was conducted by 4 fourth-year medical students (SPAF, EVZ, VFRV, and RVCA) and supervised by 2 faculty physicians (AHA and CRB), all of whom were actively involved in the RAÍCES 2 educational project. The student researchers shared a common institutional and academic background with the participants, having previously completed the same morphology I course at Universidad San Sebastián. This insider positionality facilitated a deep, empathetic understanding of the students' lived experiences and the specific pressures of practical assessments. However, it also introduced potential biases about preconceived notions of student anxiety and course difficulty.

To manage these subjectivities and maintain an interpretive balance, reflexivity was operationalized through a collaborative, iterative coding process. The student researchers held weekly peer-debriefing meetings to compare preliminary codes generated in ATLAS.ti and discuss their interpretive lenses.

When disagreements emerged regarding the boundary or meaning of specific codes or subthemes, they were not resolved by a majority vote; instead, they were critically debated and cross-examined alongside the raw transcripts and under the guidance of the faculty members until an analytical consensus was reached. Alternative interpretations of participants' accounts were deliberately explored during these discussions, and emerging themes were repeatedly challenged against the full dataset to ensure that interpretations remained grounded in the data rather than reflecting researchers' initial assumptions or preferred explanations. Themes were iteratively refined through multiple rounds of dialogue, moving back and forth between the empirical data and the TPACK theoretical framework to ensure that the final interpretations truly reflected the participants' accounts rather than the researchers' preexisting assumptions.

Results

Theme 1: Techno-Disciplinary Visual Alignment

Kahoot!'s use was sustained by its intuitive interface and its direct alignment with the visual demands of the course. Participants valued the tool when it effectively simulated the formal objective structured practical examinations, such as gymkhanas (a local term for timed, station-based practical assessments analogous to objective structured practical examinations) by prioritizing the visual recognition of anatomical structures under time pressure. Quick entry via QR codes facilitated a seamless start without disrupting the laboratory flow.

Furthermore, although the image-based pin function required a brief initial learning curve regarding pointer manipulation, it demonstrated high formative potential for anatomical tasks. Ultimately, both students and the teacher emphasized that the success of the activity depended critically on using authentic, high-resolution images identical to physical laboratory resources (eg, physical models and anatomical tables) to effectively train the clinical eye.

Theme 2: Kahoot! as a Formative Closing Tool

Implemented as an end-of-class activity, Kahoot! effectively structured the laboratory session's closure. Acting as an integrative "exit ticket," the platform provided a crucial opportunity for immediate self-assessment. It allowed students to immediately apply the reviewed concepts and contrast their subjective perceived understanding with their objective performance.

Rather than introducing new material, both students and the teacher valued the tool primarily as a diagnostic resource to identify learning gaps and estimate the achievement of expected outcomes. Consequently, this rapid formative check helped guide subsequent independent study. However, its impact on long-term study habits varied; while some students integrated the platform into their personal study routines to simplify content, others viewed it strictly as a classroom complement that did not alter their established study methods.

Theme 3: The Role of Immediate Explanatory Feedback

While Kahoot!'s playful elements such as speed and ranking initially engaged students, these gamified features alone were insufficient for meaningful learning. The tool's true formative value depended entirely on the teacher's pedagogical mediation. Participants stressed that merely identifying correct or incorrect answers was inadequate; explicit explanatory feedback was required to understand the rationale behind the errors. Crucially, both students and the teacher noted that this feedback needed to be immediate, leveraging the brief interval between questions while the visual stimulus remained active in students' working memory.

However, this created a persistent pedagogical tension: teachers had to constantly balance the game's dynamic rhythm with the necessity of pausing for deeper conceptual discussions, a task that proved particularly challenging with complex or extensive content.

Furthermore, the effectiveness of this feedback loop was heavily reliant on precise question design. Learning was notably hindered when images lacked clarity or when items used terminology that deviated from the course's standard anatomical vocabulary.

Theme 4: Socioemotional Dynamics of Gamification

Gamification effectively disrupted the standard classroom routine, adding dynamism and recapturing sustained students' attention. While competition served as a powerful motivational

XSL•FO
RenderX

driver for personal improvement, it also generated emotional ambivalence. Students experienced initial evaluative nervousness, and participants recognized a risk of demotivation or frustration for those consistently achieving lower rankings. In this context, the platform's anonymity was identified as a crucial safeguard. It fostered psychological safety by allowing students to participate and take academic risks without the fear of public failure.

Theme 5: Contextual and Logistical Implementation Factors

The successful implementation of Kahoot! was heavily mediated by contextual and logistical variables. Functioning effectively as an end-of-class closure activity (10 - 15 min), it helped recapture waning attention without disrupting laboratory dynamics. However, students faced minor technological barriers, such as device-battery anxiety and reliance on institutional Wi-Fi, raising concerns about digital equity and exclusion. While the teacher found the platform intuitive to facilitate, students recognized the substantial workload required to design high-quality visual quizzes. Isolated technical and terminological incidents occurred but were quickly resolved without detracting from the overall positive evaluation.

Theme 6: Projections for Sustainable Pedagogical Use

Looking forward, participants projected that systematizing the use of Kahoot!, rather than using it as an isolated intervention, would yield a deeper impact on consistent study habits. To ensure sustainability and alleviate individual faculty workloads, students recommended developing a standardized, institutionally supported question bank. This repository should ensure disciplinary balance, specifically by integrating more histology items alongside macroscopic anatomy. Ultimately, participants affirmed that the core competencies practiced through this gamified tool, such as rapid image recognition and decision-making under time constraints, were directly transferable to formal practical assessments.

Reflexive thematic analysis identified 6 main themes that synthesize students' and the teacher's perceptions of the use of Kahoot! as a formative assessment strategy in morphology I. These themes reflect how the educational value of the tool emerged from the interaction between the visual and temporal design of the activities, the teacher's pedagogical mediation, the socioemotional climate generated by gamification, and contextual implementation conditions. Table 1 presents an interpretive synthesis of the identified themes. To preserve the inductive depth of the participants' accounts, these findings are presented chronologically based on their lived experiences, using the TPACK framework strictly as a subsequent interpretive lens in the discussion section.

Table . Main themes identified from the reflexive thematic analysis of the use of Kahoot! in morphology I.

Theme	Core meaning
Alignment with disciplinary and visual demands	Kahoot! was perceived as valuable when its visual design, task structure, and available time reproduced conditions similar to practical assessments and laboratory work.
Kahoot! as a formative closing activity	The activity functioned as a rapid formative check at the end of class, supporting self-assessment, identification of learning gaps, and guidance for subsequent study.
From score to learning: feedback as a critical condition	The formative value of the tool depended on immediate explanatory feedback from the teacher, which transformed game results into opportunities for meaningful learning.
Motivation and socioemotional climate	Gamification increased engagement and attention but also generated ambivalent emotional responses related to competition, time pressure, and performance visibility.
Contextual conditions of implementation	Factors such as connectivity, device availability, class time management, and teacher preparation conditioned the quality of the experience.
Projection toward sustainable use	Participants envisioned greater educational impact through more systematic use, improved item design, disciplinary balance, and institutional support.

Discussion

Principal Findings and Interpretation

In summary, the main findings of this study indicate that Kahoot! can effectively function as a formative assessment tool when its visual and temporal design aligns with authentic practical examinations. Furthermore, as experienced by the participants, its educational value is not inherent in the technology itself but relies crucially on the participating instructor's immediate explanatory feedback to mitigate competitive anxiety and transform gamified scores into meaningful learning.

The primary perception of Kahoot! as a low-stakes formative environment, rather than a summative measurement, highlights a successful pedagogical integration of technology to alleviate evaluative anxiety. From the perspective of the TPACK framework, this shift illustrates the pedagogical use of technology to support formative assessment, demonstrating how a digital tool technological knowledge (TK) can be intentionally steered away from evaluative pressure to serve a purely

XSL•FO
RenderX

diagnostic purpose pedagogical knowledge (PK) within a demanding discipline content knowledge (CK). By leveraging technological affordances specifically for formative closure, the classroom dynamic transitions from performance-oriented competition to a psychologically safe learning space where errors are embraced as pedagogical milestones for gap diagnosis.

Crucially, Kahoot!'s pedagogical value depended entirely on immediate explanatory feedback. Students found scores and rankings insufficient without clear justifications for correct and incorrect answers. In this sense, learning did not occur during the act of responding to the question but rather during the guided reflection that followed it, positioning the teacher's mediation as a central component perceived to support the transformation of errors into meaningful learning opportunities.

This friction between engagement and cognitive processing illuminates the critical pedagogical and socioemotional dimensions of technology use in high-stakes environments. Viewed through the lens of the TPACK framework, this finding reveals that when the inherent characteristics of a technology (TK), such as Kahoot!'s speed and countdown timer, override the needs of the content (CK), a direct tension arises with the pedagogical objectives (PK). In disciplines with a high cognitive and visual load, like morphology, technological fluency can conflict with the time required for conceptual consolidation.

This correlation between item design flaws and student errors underscores the critical need for a tight integration of disciplinary content, pedagogical decisions, and technological affordances. Viewed through the lens of the TPACK framework, this issue illustrates what happens when the technological execution (TK) fails to precisely mirror the unique visual demands of the content (CK). In a highly visual discipline like morphology, where anatomical pattern recognition is a core competency, the graphical fidelity of a digital tool is not a mere esthetic preference; it is a pedagogical prerequisite. When images lack clarity or distractors introduce terminological variance, the technological tool (TK) creates a barrier that interferes with the pedagogical goal (PK) of accurate self-assessment.

The coexistence of enthusiasm and evaluative anxiety further illuminates the pedagogical and socioemotional dimensions of technology use, revealing that gamified environments activate learners at both cognitive and affective levels. Viewed through the TPACK framework, the emotional impact of a digital tool (TK) is not neutral; rather, it is heavily modulated by how game mechanics (such as leaderboards and speed points) intersect with the pedagogical climate (PK) designed by the instructor. While competition serves as a powerful driver for engagement, persistent low rankings introduce socioemotional friction in which continuous peer comparison risks shifting student anxiety from a state of productive alertness to frustration and demotivation. In the face of these adverse situations, strategies that foster student security become paramount (such as real-time reassurance provided by the instructor and student anonymity), thereby ensuring that the competitive drive does not compromise psychological safety. Specifically, anonymity acts as a structural prerequisite embedded within the technological interface (TK) that mitigates the fear of public exposure, promoting broader participation by allowing students to take cognitive risks even when unsure of their answers. Without these supportive, feedback-oriented instances (PK) to buffer the pressures of the game, the technological tool (TK) loses its formative value and defaults to an evaluative stressor. Indeed, this anonymity facilitates active participation without fear of public exposure, reducing the dread of making mistakes, whereas competition remains an ambivalent strategy that requires careful pedagogical scaffolding to prevent demotivation when peer comparison becomes persistent.

The tight interdependence among these emergent findings demonstrates that a digital gamified environment alone remains pedagogically inert without deliberate technological, pedagogical, and content alignment. This systemic interaction offers a clear projection toward sustainable use and provides transferable design principles for educators in other visually intensive medical disciplines, such as histology, pathology, or radiology. To successfully transfer these insights, instructors must move away from using gamified tools as isolated recreational games. Instead, sustainable implementation requires a balanced articulation within the TPACK framework: first, ensuring that the technological execution (TK) achieves authentic visual alignment with disciplinary content demands (CK); second, positioning immediate, explicit teacher feedback (PK) as a strict prerequisite to automated scoring; and third, actively managing the socioemotional climate to preserve psychological safety. Rather than a mere classroom intervention, these conditions establish a reproducible pedagogical model that transforms gamified formative assessment into a sustainable tool for diagnostic gap recognition across medical curricula.

Comparison With Prior Work

The findings of this study are consistent with the literature indicating that the educational value of digital technologies in medical education depends less on the tool itself and more on its integration with the pedagogical and disciplinary demands of the learning context. In their review of the digital teaching landscape in medicine, Yeung et al [1] describe that technological interventions tend to be better valued when they are coherently aligned with learning objectives and with the practices inherent to each course. In line with this, our findings show that Kahoot! was perceived as useful mainly when it aligned with the practical performance required in morphology I, particularly the recognition of anatomical structures through authentic images under time constraints comparable to the course's practical assessments.

Previous studies have documented that the use of quiz-based and game-based platforms can foster engagement and content review among medical students. Both Yuenyongviwat and Bvonpanttarananon [2] and Cordero-Cordero et al [4] report positive perceptions of Kahoot! as a learning support strategy. However, the results of the present study add an important nuance by showing that a playful and competitive component alone does not guarantee learning perceived as meaningful. In this case, explanatory feedback provided by the teacher emerged as the central condition that transformed the game dynamics into a genuinely formative experience, in line with what Rodríguez Salvador and Arboniés Ortiz [8] propose regarding

XSL•FO
RenderX

the role of feedback in formative assessment in medical education.

From a motivational standpoint, the results partially align with studies describing gamification as a strategy capable of increasing student participation and interest. Rivera et al [3] note that technology-mediated interactive activities can foster engagement in medical teaching contexts. However, the present study offers a more complex view by evidencing that motivation associated with Kahoot! is highly ambivalent. While competition was described as stimulating for many students, accounts of evaluative anxiety and potential demotivation also emerged when peer comparison became persistent. This observation is consistent with findings by Kirsch and Spreckelsen [9], who warn that competitive gamification may generate initial enjoyment but not necessarily sustained benefits for intrinsic motivation or learning.

Crucially, this socioemotional vulnerability underscores broader digital health risks documented in recent literature. As demonstrated by Song et al [10] in their evaluation of classroom technology, inappropriate or excessive use of unintegrated digital tools can be positively correlated with academic burnout among university students. To mitigate these psychological risks, Song et al [10] emphasize the importance of enhancing teacher leadership in the digital classroom and strengthening student resilience. This directly supports our findings, suggesting that when a tool like Kahoot! is used under high time and visual constraints, the emotional toll can easily tip toward burnout or frustration if the tool functions in isolation. Therefore, robust pedagogical design and active instructor mediation are required to safeguard the socioemotional climate.

Taken together, these integrated findings strongly reinforce the propositions of the TPACK framework by Mishra and Koehler [5,6], showing that the effectiveness of educational technologies in medical training does not lie, in the technological component in isolation but in the deliberate interaction between disciplinary knowledge (content), pedagogical decisions, and the intentional use of technology. From this perspective, the present study contributes situated qualitative evidence that deepens the understanding of how gamified tools such as Kahoot! are experienced by students and teachers when integrated into practical medical contexts characterized by high cognitive, evaluative, and socioemotional demands.

Conceptual Contribution of the Study

Rather than demonstrating the effectiveness of a specific digital platform, this study identifies a set of pedagogical conditions that appear to be essential for designing effective gamified formative assessment in visually intensive medical courses. These principles were inductively derived from 6 themes identified in the reflexive thematic analysis and subsequently interpreted using the TPACK framework.

First, gamified activities should reproduce the authentic visual and temporal demands of disciplinary assessments to ensure alignment with learning objectives. Second, game mechanics should be embedded within intentional pedagogical design rather than functioning as isolated motivational elements. Third, immediate explanatory feedback should accompany each response to transform game performance into meaningful learning. Finally, implementation should explicitly consider students' socioemotional experiences by balancing engagement, competition, and psychological safety.

Together, these findings extend the application of the TPACK framework by proposing transferable design principles that can inform the development of gamified formative assessment beyond the specific context of morphology I and the use of Kahoot!

Strengths and Limitations

This study is distinguished by a comprehensive qualitative approach that enabled a rich and detailed understanding of students' and the teacher's experiences regarding the use of Kahoot! in morphology I. Methodological triangulation, integrating 2 student focus groups and 1 semistructured interview with the teacher, strengthened the credibility of the findings by contrasting pedagogical intentions with students' perceptions and needs.

The combination of a deductive theoretical framework based on the TPACK model and the RAÍCES project axes with an inductive reflexive thematic analysis (Braun and Clarke [7]) allowed for a systematic and nuanced interpretation of the data, identifying both expected patterns and relevant emergent elements, such as the importance of the visual component and the ambivalence of competition.

Among the limitations, the study was conducted in a single institutional context, the Universidad San Sebastián Valdivia campus, and with a specific cohort of medical students, which limits the transferability of the findings to other settings. Additionally, it is necessary to acknowledge that the faculty perspective is derived from a single instructor, meaning these specific pedagogical reflections may not represent the broader teaching body of the institution. Crucially, because this research was based on a single Kahoot! intervention, a potential "novelty effect" must be critically considered. The high levels of student motivation, focus, and even the acute evaluative anxiety reported may have been heavily influenced by the unique, disruptive nature of the activity within the standard laboratory routine. If the tool were integrated into a repeated or mandatory framework, these dynamics would likely shift; repeated exposure might normalize the format, thereby mitigating acute time-related anxiety, yet it concurrently risks inducing technological fatigue and diminishing the situational interest observed in this isolated trial.

Furthermore, our recruitment relied on a voluntary sampling strategy, which introduces a distinct self-selection bias. It is highly probable that the 14 students who chose to participate represented more academically engaged or more positive voices within the cohort, potentially overrepresenting favorable perceptions of Kahoot! Consequently, the perspectives of students who experienced severe demotivation, lower academic performance, or technological alienation may have been underrepresented, meaning the full spectrum of socioemotional friction was not entirely captured. Finally, the qualitative approach did not incorporate objective measures of academic performance, and the implementation under a bring your own

XSL•FO
RenderX

device model introduced variability in devices and connectivity, which heavily conditioned the individual user experience.

Future Directions

Future studies could evaluate the impact of a systematic and longitudinal implementation of Kahoot! as a formative assessment strategy in morphology I on objective academic performance indicators, using quasi-experimental designs or comparisons between course sections. It would also be valuable to explore, in a controlled manner, the effects of design variables identified as critical in this study—such as the quality of visual materials, the use of timed responses, and the presence of explanatory feedback—while incorporating complementary measures such as motivation, self-efficacy, cognitive load, and assessment-related anxiety.

Conclusions

Principal Insights and Disciplinary Transferability

This study explored the experiences of students and a teacher regarding the use of Kahoot! as a formative assessment strategy in the morphology I laboratory, showing that its pedagogical value does not lie in the technological tool itself, but rather in its alignment with the disciplinary, evaluative, and pedagogical demands of the course. Kahoot! was perceived as useful when it allowed students to practice the recognition of structures using authentic images under time constraints and in formats comparable to real practical assessments.

In this context, Kahoot! can constitute a relevant and transferable strategy in health sciences education, provided that it is implemented intentionally, with well-designed items, explicit pedagogical feedback, and a conscious consideration of its cognitive and emotional implications. Conceptually, our findings regarding the interplay between rapid image recognition and real-time faculty mediation possess significant transferability to other visually intensive medical disciplines, such as histology, pathology, and radiology. In these fields, undergraduate students face a steep learning curve characterized by a high visual cognitive load, in which they must transition from merely viewing complex structures to instantly recognizing clinical patterns under diagnostic time constraints. By demonstrating that gamified environments can effectively simulate these authentic evaluative pressures, our study offers a broader pedagogical model: digital tools serve as a valuable sandbox to test the "clinical eye," but their educational efficacy remains largely dependent on immediate instructor deconstruction of visual errors to prevent cognitive overload and safely foster diagnostic competence.

The findings indicate that within the scope of the analyzed experiences, the tool functioned primarily as a formative closing activity, supporting self-assessment, diagnosis of learning gaps, and consolidation of learning, rather than as a summative measurement mechanism. However, its formative impact was critically conditioned by the participating instructor's mediation, particularly through the provision of explanatory feedback following each question.

From a socioemotional perspective, gamification fostered motivation, participation, and attention but also activated ambivalent emotional responses associated with competition and time management. These results underscore the need for careful pedagogical design that considers not only students' engagement but also learners' psychological safety.

TPACK-Informed Design Principles and Practical Recommendations

Taken together, the results support the TPACK framework by suggesting that the perceived use and educational value of gamified tools in medical education emerge from the interaction between technology, pedagogy, and disciplinary content. More broadly, these findings suggest that the effectiveness of gamified formative assessment depends less on the digital platform itself than on the pedagogical conditions under which it is implemented. Authentic alignment with disciplinary tasks, immediate explanatory feedback, intentional instructor mediation, and attention to students' socioemotional experiences emerged as potentially transferable design principles that may guide the implementation of gamified formative assessment across visually intensive health professions education.

These findings suggest that educators implementing gamified formative assessment should approach these tools not as stand-alone games but as highly intentional pedagogical interventions. Instructors in visually demanding disciplines should specifically design digital quizzes to mirror the authentic time constraints and visual fidelity of real-world practical assessments. Furthermore, rather than relying on the platform's automated scoring, educators must actively deliver immediate, explanatory feedback between questions. This real-time mediation is essential to transforming the inherent competitive anxiety of gamification into a psychologically safe and meaningful learning opportunity.

Acknowledgments

The authors thank the medical students and the course teacher who participated in this study and shared their perspectives, as well as Universidad San Sebastián for the institutional support provided for the development of the RAÍCES 2 (Red de ayudantes de investigación comprometidos con la educación en salud; Network of Student Researchers Committed to Health Professions Education) project.

Generative AI Use

Generative AI tools (ChatGPT; OpenAI) were used exclusively to support the improvement of the manuscript's academic writing (language, clarity, and readability of the text) and to assist in the translation of the manuscript from Spanish into English, with the aim of enhancing clarity, coherence, and linguistic style. In addition, AI-assisted transcription was used only for the first

XSL•FO
RenderX

student focus group as an initial transcription aid. The resulting transcript was subsequently reviewed and manually corrected by members of the research team to ensure accuracy and fidelity to the original audio recording. AI tools were not used for the transcription of the remaining sessions, nor were they used for data generation, coding, data analysis, interpretation, or decision-making related to the study. All analytic decisions, theme development, and interpretation of the findings remained the sole responsibility of the research team. All sections of the manuscript were originally written by the authors, who reviewed, edited, and approved the final English version and retained full responsibility for the content, data interpretation, and conclusions presented, as well as for the accuracy, originality, and integrity of the manuscript, including all references and citations.

Funding

This study was supported by the Vicerrectoría Académica and the Centro de Investigación para la Educación Superior (Fondo USS-FIN-2026_APC_VRA-012) and by institutional resources from Universidad San Sebastián.

Data Availability

Due to the qualitative nature of the study and the risk of participant reidentification, the full transcripts are not publicly available. Deidentified excerpts and the analytical framework may be shared upon reasonable request, subject to ethical approval and data protection requirements.

Authors' Contributions

Conceptualization: SPAF, EVZ, VFRV, RVCA
Data curation: SPAF, EVZ, VFRV, RVCA
Formal analysis: SPAF, EVZ, VFRV, RVCA, AHA, CARC
Investigation: SPAF, EVZ, VFRV, RVCA
Methodology: SPAF, EVZ, VFRV, RVCA
Supervision: AHA, CRB, SPAF, CARC
Writing – original draft: SPAF, EVZ, VFRV, RVCA
Writing – Review and editing: SPAF, EVZ, VFRV, RVCA, AHA, CRB, CARC

Conflicts of Interest

None declared.

Multimedia Appendix 1
Focus group and semistructured interview guides.
[DOCX File, 41 KB - humanfactors_v13i1e90651_app1.docx]

Multimedia Appendix 2
Methodological flow diagram of the qualitative data analysis process.
[PDF File, 652 KB - humanfactors_v13i1e90651_app2.pdf]

References

1. Yeung AWK, Parvanov ED, Hribersek M, et al. Digital teaching in medical education: scientific literature landscape review. JMIR Med Educ 2022 Feb 9;8(1):e32747. [doi: 10.2196/32747] [Medline: 35138260]
2. Yuenyongviwat V, Bvonpanttarananon J. Using a web-based quiz game as a tool to summarize essential content in medical school classes: retrospective comparative study. JMIR Med Educ 2021 Apr 29;7(2):e22992. [doi: 10.2196/22992] [Medline: 33913813]
3. Rivera R, Smart J, Sakaria S, et al. Planning engaging, remote, synchronous didactics in the COVID-19 pandemic era. JMIR Med Educ 2021 May 11;7(2):e25213. [doi: 10.2196/25213] [Medline: 33872191]
4. Cordero-Cordero G, Guevara-Vizcaíno C, Erazo-Álvarez C. Kahoot! como herramienta de gamificación del aprendizaje: una experiencia con estudiantes de Medicina [Article in Spanish]. 593 Digit Publ CEIT 2022;7(4-2):328-341. [doi: 10.33386/593dp.2022.4-2.1426]
5. Mishra P, Koehler MJ. Technological Pedagogical Content Knowledge: a framework for teacher knowledge. Teach Coll Rec 2006 Jun;108(6):1017-1054. [doi: 10.1111/j.1467-9620.2006.00684.x]
6. Koehler MJ, Mishra P. What is technological pedagogical content knowledge? Contemp Issues Technol Teach Educ 2009;9(1):60-70 [FREE Full text]
7. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
8. Rodríguez Salvador JJ, Arboniés Ortiz JC. El feedback como herramienta docente en la formación clínica. 1.ª parte [Article in Spanish]. Educ Med 2018 Mar;19(2):120-124. [doi: 10.1016/j.edumed.2018.02.001]

XSL•FO
RenderX

9. Kirsch J, Spreckelsen C. Caution with competitive gamification in medical education: unexpected results of a randomised cross-over study. BMC Med Educ 2023 Apr 19;23(1):259. [doi: 10.1186/s12909-023-04258-5] [Medline: 37072842]
10. Song W, Wang Z, Zhang R. Classroom digital teaching and college students' academic burnout in the post COVID-19 era: a cross-sectional study. Int J Environ Res Public Health 2022 Oct 17;19(20):13403. [doi: 10.3390/ijerph192013403] [Medline: 36293983]

Abbreviations

CK: content knowledge
PK: pedagogical knowledge
RAÍCES 2: Red de ayudantes de investigación comprometidos con la educación en salud (Network of Student Researchers Committed to Health Professions Education)
TK: technological knowledge
TPACK: Technological Pedagogical Content Knowledge

Edited by C LeRouge; submitted 01.Jan.2026; peer-reviewed by D Son, WAW Yahaya; revised version received 13.Jul.2026; accepted 10.Aug.2026; published 22.Oct.2026.

Please cite as:
Herrera Alcaíno Á, Araya Figueroa SP, Riquelme Vilches VF, Vargas Zurita E, Cozzi Ahumada RV, Riquelme Bahamondes C, Rosas Chuñil CA
Gamified Formative Assessment in a First-Year Medical Morphology Course: Qualitative Study
JMIR Hum Factors 2026;13:e90651
URL: https://humanfactors.jmir.org/2026/1/e90651
doi:10.2196/90651

© Álvaro Herrera Alcaíno, Sofía Paz Araya Figueroa, Valentina Francisca Riquelme Vilches, Eduardo Vargas Zurita, Renata Valentina Cozzi Ahumada, Camila Riquelme Bahamondes, Carlos Alberto Rosas Chuñil. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 22.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Quality of Informed Consent and Interface Usability in Primary Care e-Consultation: Cross-Sectional Study

Caitlin Parfitt-Ford[1], BSc; Lisa Ballard[2,3], DHealthPsych; Adriane Chapman[1,3], PhD

[1]School of Electronics and Computer Science, Faculty of Engineering and Physical Sciences, University of Southampton, Southampton, United Kingdom
[2]Primary Care, Population Science, and Medical Education, Faculty of Medicine, University of Southampton, Southampton, England, United Kingdom
[3]Southampton Biomedical Research Centre, National Institute for Health and Care Research, Southampton, England, United Kingdom

Corresponding Author:
Caitlin Parfitt-Ford, BSc
School of Electronics and Computer Science
Faculty of Engineering and Physical Sciences
University of Southampton
University Road
Southampton, SO17 1BJ
United Kingdom
Phone: 44 2380595000
Email: cpf1g20@soton.ac.uk

Related Article:

Correction of: https://humanfactors.jmir.org/2026/1/e78483

(JMIR Hum Factors 2026;13:e103250) doi:10.2196/103250

In "Quality of Informed Consent and Interface Usability in Primary Care e-Consultation: Cross-Sectional Study" [1], the authors made one clarification.

The name of the first author was changed to Caitlin.

The correction will appear in the online version of the paper on the JMIR Publications website, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Reference

1. Parfitt-Ford C, Ballard L, Chapman A. Quality of informed consent and interface usability in primary care e-consultation: cross-sectional study. JMIR Hum Factors 2026 Feb 09;13:e78483 [FREE Full text] [doi: 10.2196/78483] [Medline: 41662699]

Submitted 02.Jun.2026; this is a non–peer-reviewed article;accepted 24.Jun.2026; published 12.Aug.2026.

Please cite as:
Parfitt-Ford C, Ballard L, Chapman A
Correction: Quality of Informed Consent and Interface Usability in Primary Care e-Consultation: Cross-Sectional Study
JMIR Hum Factors 2026;13:e103250
URL: https://humanfactors.jmir.org/2026/1/e103250
doi:10.2196/103250
PMID:

©Caitlin Parfitt-Ford, Lisa Ballard, Adriane Chapman. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 12.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information,

XSL•FO
RenderX

a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Human-AI Interaction With AI-Assisted Tumor Overlays in Pediatric Whole-Body Magnetic Resonance Imaging: Exploratory Reader Study

Abhishek Moturu[1,2,3,4*], HBSc, MSc; Olurotimi Komolafe[5,6*], BSc, MBBS; Sayali Joshi[5,6], MBBS, DMRE, MSc; Andrea S Doria[5,6], MSc, MBA, MD, PhD; Anna Goldenberg[1,2,3,4,7,8], PhD

[1]Department of Computer Science, University of Toronto, 40 St George Street, Toronto, ON, Canada
[2]Vector Institute, Toronto, ON, Canada
[3]Genetics and Genome Biology, Hospital for Sick Children, Toronto, ON, Canada
[4]Temerty Centre for Artificial Intelligence Research and Education in Medicine, University of Toronto, Toronto, ON, Canada
[5]Department of Diagnostic & Interventional Radiology, Hospital for Sick Children, Toronto, ON, Canada
[6]Department of Medical Imaging, University of Toronto, Toronto, ON, Canada
[7]CIFAR, Toronto, ON, Canada
[8]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada
[*]these authors contributed equally

Corresponding Author:
Abhishek Moturu, HBSc, MSc
Department of Computer Science, University of Toronto, 40 St George Street, Toronto, ON, Canada

Abstract

Background: AI tools have the potential to enhance personalized clinical care, particularly in radiology. However, their integration into clinical workflows remains complex, especially in pediatric oncology, where early cancer detection is critical. Children with Li-Fraumeni syndrome (LFS), a rare cancer predisposition disorder, undergo regular surveillance whole-body magnetic resonance imaging (wbMRI), which presents an opportunity for AI-assisted tumor detection.

Objective: We evaluated the feasibility of an AI-assisted overlay for highlighting tumor-like regions in pediatric surveillance wbMRI and explored how access to the overlay influenced radiologist workflow, candidate-lesion marking behavior, follow-up recommendations, and perceived workload.

Methods: We developed a patch-based AI segmentation model trained on augmented 2D slices from 675 surveillance wbMRI volumes of pediatric patients with LFS. The model was designed to highlight regions with high tumor probability. A reader study was conducted with 2 radiologists who independently reviewed wbMRI cases both with and without AI assistance. We measured evaluation time, number and location of reader-marked candidate lesions, type of follow-up recommendation, and subjective feedback using structured questionnaires.

Results: AI assistance altered interpretation workflows for both radiologists, with mixed effects. On average, the time required to evaluate each case increased when using the AI tool for both radiologists. However, one radiologist had an increase in the number of candidate lesion locations selected with the tool, and one had a decrease in the number of candidate lesion locations selected with the tool. Subjective feedback indicated that one of the radiologists reported lower mental demand with the AI tool, while both radiologists reported lower stress with the AI tool. Interrater variability was evident, underscoring the need for personalized calibration of AI tools.

Conclusions: AI-assisted wbMRI interpretation can improve tumor detection in pediatric cancer surveillance by reducing false negatives. However, its influence on workflow efficiency and interradiologist variability highlights the importance of careful implementation. Successful integration requires addressing challenges such as improving the predictive precision of AI models, offering intuitive end-user designs and instructions, and building trust in AI outputs. AI outputs can influence workflow and behavior in reader-specific ways. Clinical translation will require larger, randomized, multireader studies and model refinement to reduce false positives and quantify lesion-level reader performance. This can help ensure better patient outcomes in addition to reduced clinician burnout.

(JMIR Hum Factors 2026;13:e81066) doi:10.2196/81066

XSL•FO
RenderX

KEYWORDS

machine learning; radiology; segmentation; pediatric oncology; whole-body magnetic resonance imaging; wbMRI; Li-Fraumeni syndrome; tumor detection; human-AI interaction; interpretability in AI

Introduction

Integrating AI tools into clinical practice marks a paradigm shift in health care, with the potential for enhanced diagnostic accuracy, improved workflows, and personalized patient care. In fact, a 2019 survey of health professionals found that up to 71% of surveyed physicians opined that AI would improve and impact the field of medicine within the next decade [1]. The potential of AI tools in health care extends across various domains, including radiology [2], pathology [3], dermatology [4], ophthalmology [5], general practice [6], surgery [7], neurology [8], cardiology [9], and more. By leveraging vast amounts of medical data, AI algorithms can help identify patterns and correlations that may be more difficult for human clinicians to recognize due to the size, rarity, or complexity of certain clinical conditions. As a result, this can enable earlier and more accurate diagnoses and subsequent interventions. Machine learning models have shown promising task-specific performance in selected medical-imaging applications, although performance and clinical benefit vary substantially across tasks, datasets, and implementation settings [10].

In addition to aiding in clinical diagnostics, AI tools can enhance clinical decision-making [11]. AI-based decision support systems can synthesize plans from multimodal data and can continuously learn from new data, improving recommendations over time and adapting to the latest clinical guidelines and research findings [12]. The use of AI in clinical settings also addresses the growing concern of clinician burnout caused by an increase in workload in interpreting imaging examinations without a corresponding increase in manpower [13]. Yu et al [14] further support the complexity of human-AI collaboration in diagnostic settings and indicate significant variability in how radiologists respond to AI assistance. They found that radiologists' performance was easily influenced by AI errors in their large-scale study on AI-aided chest X-ray interpretation, but conventional factors such as experience or familiarity with AI did not predict the impact of AI support. There is also inherent variability in how radiologists use and interact with AI tools, which is influenced by personal biases, cognitive strategies, and the opacity of AI algorithms [15-18].

Integrating AI into clinical practice also raises concerns about data privacy [19], algorithmic bias [20], and the need for rigorous validation and regulation [21]. Ensuring that AI tools are reliable, explainable, and complement rather than replace clinical judgment is critical to ensuring their acceptance, since replacement is a significant fear in the field, for clinicians and patients receiving care alike [22,23]. Thus, a collaborative approach, where AI augments the clinician's expertise instead of acting as a substitution, is essential for the successful adoption of these technologies.

In this paper, we studied the interaction between radiologists and a specific AI tool that was developed for the interpretation of tumors in whole-body magnetic resonance imaging (wbMRI) performed as part of an imaging protocol for cancer surveillance [24]. Early detection of tumors in patients with cancer predisposition disorders such as Li-Fraumeni syndrome (LFS) through clinical-imaging protocols has been shown to improve the prognosis of patients [25]. However, the detection and segmentation of tumors in pediatric wbMRIs is challenging for many reasons. In particular, variations in the age, sex, body shape, body size, and bone density of a developing child may influence the anatomical and physiological properties that contribute to image formation. In addition, variations in tumor size, location, magnetic resonance imaging (MRI) signal brightness, type and settings of the MRI scanner, and MRI artifacts are a few of the factors that radiologists must consider when evaluating wbMRIs in search of tumors in patients with LFS [26]. Although the malignancy of a tumor cannot always be identified from an MRI examination alone, developing a tool to assist radiologists in finding tumors by highlighting areas of high tumor probability may save time, reduce the chance of missing tumors, and minimize the amount of invasive tests. However, there is a lack of tools for finding tumors in wbMRIs for the reasons outlined above. To our knowledge, AI-assisted tumor detection has been less studied in pediatric surveillance wbMRI than in adult or single-organ imaging settings. This setting is distinct because lesions are often very small relative to the full image volume, tumor prevalence is low, and the whole-body field of view introduces substantial anatomical heterogeneity. Accordingly, our contribution is both technical and human-factors-oriented: we describe a patch-based tumor-likelihood overlay for pediatric LFS surveillance wbMRI and an exploratory reader study, evaluating how the AI overlay influences workflow, candidate-lesion marking behavior, follow-up recommendations, and perceived workload.

In order to find areas of high tumor probability within wbMRIs, we use segmentation. The use of deep learning methods to perform biomedical segmentation tasks effectively, especially in low-data availability domains, has seen great strides in recent years [27,28]. In spite of this, the comparative sizes of wbMRI volumes and tumors are so vastly different that the signal-to-noise ratios are often too low for proper tumor detection and segmentation when using full volumes or generative approaches [29,30]. Patch-based approaches show promise in this regard for segmenting very small objects in large spaces [31-33].

After developing our AI tool to highlight areas of high tumor probability in wbMRIs, we performed a validation study with 2 radiologists, each reviewing several wbMRIs with and without the assistance of the AI tool, to analyze the effect of the tool on the overall evaluation time, candidate-lesion marking behavior, follow-up recommendations, and questionnaire responses. Based on our findings from this study and relevant literature, we provide insights on how to effectively incorporate AI tools into health care to navigate existing challenges and uncover unexplored opportunities. Based on these findings and relevant

XSL•FO
RenderX

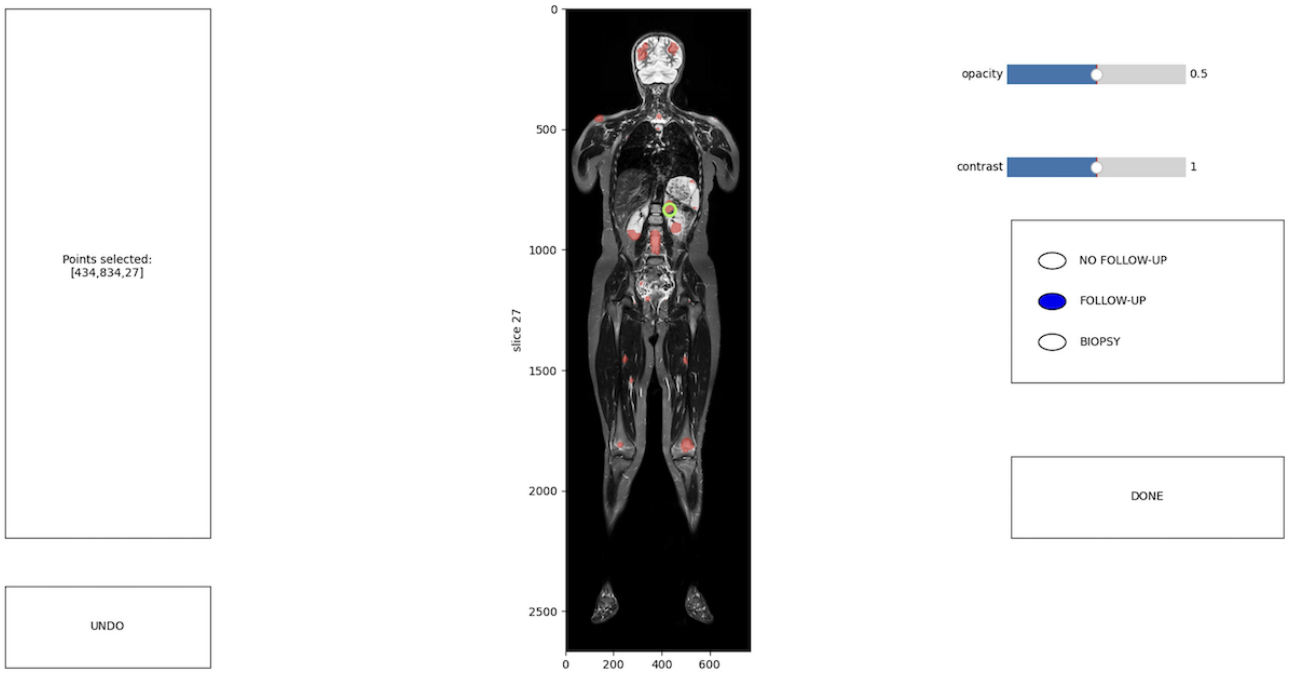
literature, we discuss practical considerations for integrating such overlays into health care workflows.

Methods

AI Tool Development

We developed our AI tool, shown in Figure 1, starting from its initial conception and progressing through multiple iterations to refine it based on best practices and feedback. We carefully integrated feedback from our clinical collaborators at every stage, ensuring that both the underlying algorithm and the user-facing components aligned with real-world diagnostic needs. The model-development and reader-study components are reported separately below.

Figure 1. The app that allows scrolling through a volume with the AI tool overlaying masks onto the regions of high tumor probability (in red), selected points (candidate lesion locations) on the image (in a green circle), a list of selected points on the left, an opacity slider and contrast slider on the top right, and the a reader-selectable follow-up recommendation on the right.



We trained the AI tool to recognize areas of high tumor probability using datasets generated by obtaining 2D patches from the slices of 675 coronal Short Tau Inversion Recovery (STIR) wbMRI examinations in which volumes were assessed from patients with LFS aged 0 to 18 years, followed in the Oncology Department of a large tertiary pediatric hospital after the study received Research Ethics Board approval from our institution. Only 27 out of 675 (4.0%) volumes were tumor-positive, containing 60 tumors in total, and only 120 out of 27,437 (0.44%) slices contained tumor signal. We have approximately 40 slices per volume, with each slice being approximately 3000×800 pixels. This created an extreme class-imbalance setting, with tumor tissue representing a minute fraction of the total image area. Because the natural prevalence of tumor-containing tissue was too low for stable patch-based training, patch sampling was intentionally enriched for tumor-containing patches, as described below. The diameter of each tumor section ranged from about 10 to 100 pixels, representing only a small fraction of each wbMRI volume. We used 4-fold cross-validation and a test set while ensuring that each of the train and validation folds, as well as the test set, was separated by patient volumes. A total of 5707 patches of size 128×128 pixels were extracted after split assignment, with 3995 (approximately 70%) used for training/validation and 1712 (approximately 30%) for testing. Within each split, approximately 25% (999/3995 and 428/1712) of the sampled patches contained tumor tissue. This proportion was enforced intentionally for model development and did not reflect the natural prevalence of tumor-containing tissue in surveillance wbMRI. Positive patches were drawn primarily from tumor-containing slices and centered on annotated lesions, with additional jittered samples to vary lesion position; negative patches were sampled from tumor-free regions, tumor-free volumes, and physiologically bright STIR regions that commonly generate false positives. Data splitting was performed at the patient level before patch extraction so that all patches from a given patient remained within a single training, validation, or test partition. The results on the volume-level cross-validation set and test set were obtained by running the trained patch-based model in a sliding-window manner on each slice, with 25% pixel overlap between adjacent patches to reduce edge misses.

Ethical Considerations

The study received approval from the Research Ethics Board at The Hospital for Sick Children (SickKids) under the protocol “Machine Learning (ML) Approach to Improve Early Cancer Detection” (number 1000063239; principal investigator: AG).

Patches, Augmentations, and Synthetic Tumors

Rather than using all possible patches, we found it more efficient and effective to primarily train the model using patches that

XSL•FO
RenderX

contained tumors centrally (Figure 2). The masks for the corresponding patches were taken from the same location within the mask volumes. We also included a few patches with their centers moved around randomly to account for variations in the location of a tumor within any given patch. To improve the specificity of the tool, we also included several patches from elsewhere in the body to train the model to recognize false positives. For instance, we included patches from parts of the body that are inherently bright on wbMRI STIR sequences, to ensure that the model can differentiate between inherently bright body regions and tumors, which also appear bright. Finally, we included a few patches from volumes without tumors, which were obtained from the same general region as volumes containing tumors, in order to help the model differentiate between patches with and without tumors obtained from the same part of the body.

Figure 2. (A) A patch and mask with tumors. (B) A patch and mask without tumors.



To improve robustness, we applied augmentations intended to approximate routine manipulation during MRI review, including zoom, brightness variation, and noise modulation. We also included synthetically inserted tumors of varying size, shape, brightness, and location to increase exposure to rare small-lesion patterns [34,35].

Model and Losses

The U-Net model [28] was used along with a compound loss function containing asymmetric unified focal loss [36], Dice loss [37], contour perimeter loss [38], and contour difference loss to address extreme foreground-background imbalance and to improve lesion boundary localization. The respective coefficients α, β, γ, and δ of each of these losses were tunable hyperparameters.

$$(1) \text{Loss}_{\text{patch}} = \alpha \times \text{Loss}_{\text{AUF}} + \beta \times \text{Loss}_{D} + \gamma \times \text{Loss}_{\text{CP}} + \delta \times \text{Loss}_{\text{CD}}$$

Each of the 4 parts of the loss function (Equation 1) served a specific purpose. Loss_{AUF} and Loss_{D} dealt with the imbalance between the hard class (tumor pixels) and easy class (nontumor pixels), and Loss_{CP} and Loss_{CD} ensured that the tumor was properly localized and sized with limited overguessing or underguessing.

Radiologist Evaluation

Setup

To evaluate our AI tool designed to highlight areas of high tumor probability on pediatric wbMRIs from children with a known cancer predisposition syndrome, based on the patch-based model run across wbMRIs in a sliding window fashion as described above, we gave 2 radiologists (Olurotimi Komolafe and Sayali Joshi, each with 3 years of radiology experience after training, respectively, who were supervised by a senior radiologist, Andrea S Doria, with over 20 years of radiology experience after training) detailed instructions on how to label 25 wbMRI volumes with the aid of the AI tool and 25 wbMRIs without the aid of the AI tool. The tool provided radiologists with the option to have AI overlays of predicted masks of high tumor probability regions to potentially help them identify tumors more efficiently and accurately. This was an exploratory, nonrandomized, noncounterbalanced reader study. Each radiologist reviewed 25 cases with AI overlays, followed by 25 different cases without AI overlays.

A timer was incorporated into each wbMRI, which automatically started as soon as each radiologist began the evaluation. The first 25 wbMRIs assigned to each radiologist contained the AI-assisted tumor prediction overlays. The app, as pictured in Figure 1 (note that the red prediction mask overlay is only shown for volumes 1 - 25 and not shown for volumes 26 - 50), shows the user interface we designed for evaluation. When reading the AI-assisted MRI volumes, the radiologists had the option to scroll up and down through each volume using the up and down keys on the keyboard (or scroll using the mouse or trackpad). They also had the option to change the degree of opacity of the red prediction mask overlays using an opacity scroller, which enabled them to transition between fully transparent and fully opaque. The default opacity was set to 50%. The app also incorporated a contrast tool to alter the visual contrast of the original wbMRI from 0 to 2, where 0=low contrast and 2=high contrast. The default contrast was set to 1. To improve detail, the app also incorporated zoom in, zoom out, and pan features, whose buttons were placed on the bottom left of the screen (Figure 1). These augmentations, some of which were suggested by the participating radiologists, were incorporated to simulate the native MRI reporting tools, which the radiologists were accustomed to in reporting wbMRIs.

The primary task was to place a click at the center of each suspected lesion or other candidate abnormality within a wbMRI volume. We refer to these clicks as candidate lesion locations rather than tumors because each click represented a reader's suspicion rather than a verified lesion. Clicking on a region placed a green circle over that region, with a corresponding coordinate from the region added to a "Points selected" box (Figure 1). This box contained all the points (candidate lesion locations) selected for that specific wbMRI volume and served

XSL•FO
RenderX

the purpose of providing visual feedback of all candidate lesion locations to the radiologists. This way, they could see if any points had been selected in error. To undo any errors, an "Undo" button was provided, which deleted the previous selection circle and the corresponding error coordinates from the selection box.

After selecting all the candidate lesion locations that they suspected to contain tumors/lesions, they were then asked to choose whether the lesion required "No Follow-Up," "Follow-Up," or "Biopsy." The options to do this and finish the evaluation were provided on the right side of the screen (Figure 1). Invariably, the radiologists' choice as to whether or not a lesion required follow-up was reflective of their consideration of the lesion to be malignant, suspicious, indeterminate (biopsy or follow-up), or benign/physiological (no follow-up). After clicking on a choice of follow-up, the incorporated timer automatically stopped, and the radiologist was redirected to a short Likert survey to evaluate their reporting experience for that particular volume. The survey evaluations were not timed.

Survey

The survey consisted of the following questions, ranked on a scale from 1 ("not at all") to 5 ("a great deal"):

Q1. How mentally demanding was the task?

Q2. How hurried or rushed was the pace of the task?

Q3. How successful were you in accomplishing what you were asked to do?

Q4. How hard did you have to work to accomplish your level of performance?

Q5. How insecure, discouraged, irritated, stressed, and annoyed were you?

After concluding their reading of the first 25 wbMRI volumes, which contained tumor prediction overlays, the radiologists then read another set of 25 wbMRI volumes, following the exact same procedures, but without tumor prediction overlays. Data obtained from each MRI volume—with and without the AI overlays—included the amount of time it took to finish the labeling, the candidate lesion locations selected, the follow-up recommendation, and the survey responses.

Results

To better study the discrepancy between the performance at the patch level and the volume level, the following metrics were considered at each level. Although it is desirable for diagnostic tools to have both low false-positive and false-negative rates, there is often a trade-off in trying to minimize both.

Given the priority of minimizing false negatives, we used a threshold to check whether at least 10% of the tumor pixels have been predicted by the segmentation model. The dice score was easily affected by the amount of tumor pixels correctly predicted and therefore did not need to be rigorously optimized, as predicting the general tumor location is sufficient. Similarly, the pixel-based false-positive rate needs to be higher for the segmentation model to properly identify parts of tumors that would otherwise be missed. We achieved a 1.54% increase on average across the validation folds in the binary tumor detection score at the patch level, as shown in Table 1, with 0.9550 using U-Net without any modifications and 0.9704 using U-Net with all of the modifications discussed in this paper.

Table . Train/validation results for each additional modification to the U-Net model with the average dice score, true-positive rate (TPR), false-positive rate (FPR), and binary tumor detection score (true for a patch if TPR>10%) across the folds[a].

U-Net modifications	Dice score	TPR	FPR	TPR>10%
$Loss_D$	0.9097/0.8200[b]	0.9874/0.9391	0.0080/0.0096[b]	0.9926/0.9550
+ $Loss_{AUF}$	0.8105/0.7463	0.9796/0.9478	0.0170/0.0181	0.9911/0.9616
+ $Loss_{CP}$	0.8182/0.7439	0.9788/0.9486	0.0167/0.0440	0.9883/0.9666
+ $Loss_{CD}$	0.8328/0.7644	0.9834/0.9457	0.0179/0.0139	0.9908/0.9651
+ Augmentations	0.7703/0.7732	0.9767/0.9476	0.0229/0.0137	0.9865/0.9660
+ Synthetic tumors	0.8014/0.7577	0.9754/0.9527[b]	0.0167/0.0155	0.9858/0.9704[b]

[a]We treat true-negative patches as perfect matches, which count toward the patches where TPR >10%. Each row consists of all the additions in the previous rows.

[b]We denote the best validation results for each metric. Note that the best TPR and TPR>10% come from including all of the U-Net modifications.

At the patch level, in addition to recording each of the losses and the total weighted loss, the true-positive rate (nonthresholded and thresholded at intervals of every 10% of tumor detection), the true-negative rate, false-positive rate, false-negative rate, positive predictive value, negative predictive value, accuracy, balanced accuracy, Jaccard index, false discovery rate, and false omission rate were recorded for both the training and validation sets. These epoch-based metrics were considered when selecting the best model. Figures 3 and 4 illustrate representative test-set predictions.

XSL•FO
RenderX

Figure 3. Examples of contours of the true and predicted masks (red outline) for (A) true positive, (B) true negative, (C) false positive, and (D) false negative when identifying a tumor.

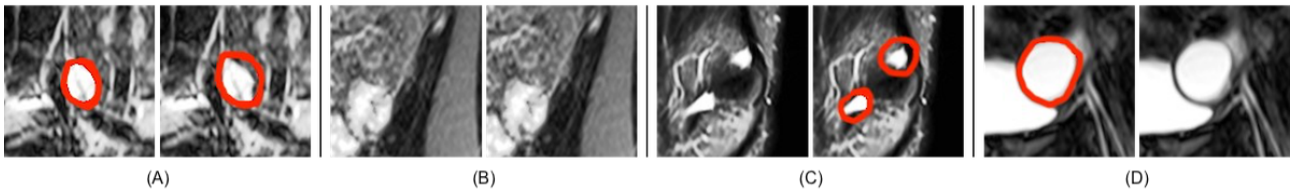
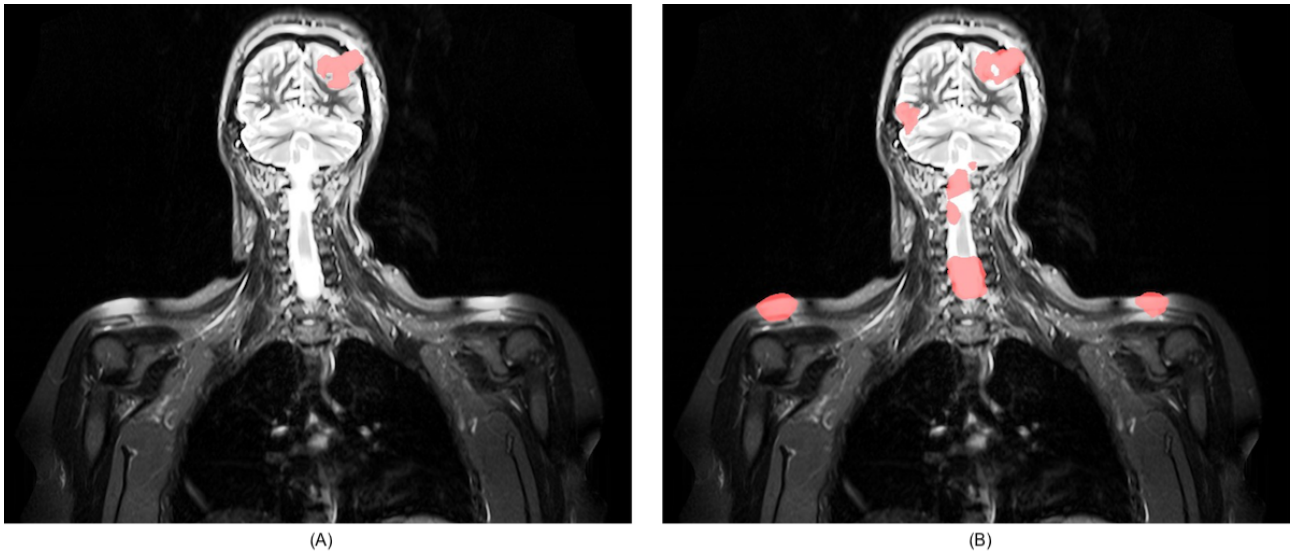


Figure 4. Representative test-set example. (A) Ground-truth tumor annotation. (B) Model output overlaid in red. In addition to the true lesion, nontumor highlighted areas are visible, illustrating the false-positive burden that radiologists would need to dismiss in clinical use.



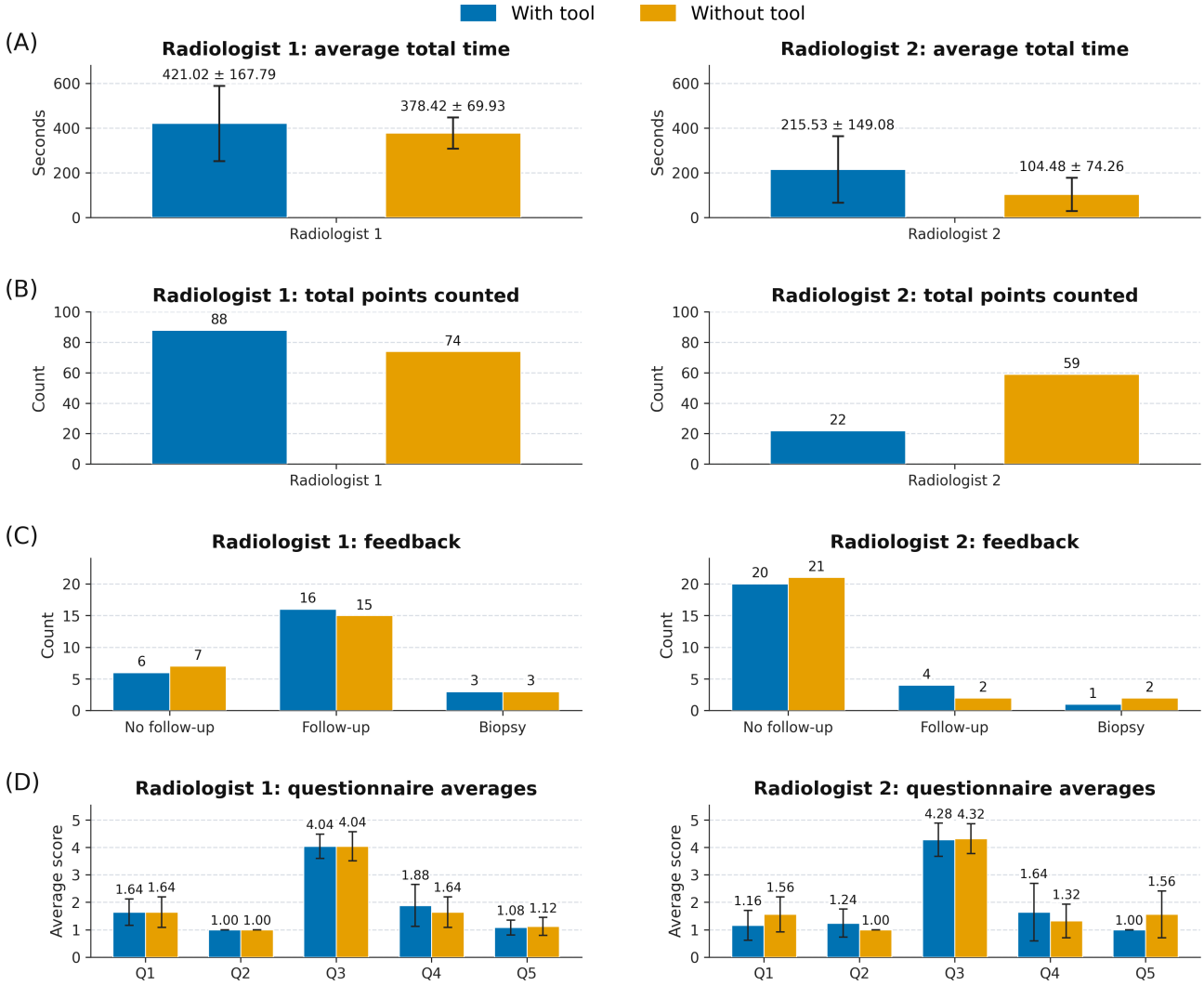
The detection rates reported below are lesion-level and tumor-positive volume-level metrics from sliding-window inference on full wbMRI volumes; a tumor was counted as detected when the model overlapped at least 10% of the annotated tumor pixels. On the cross-validation set, 27 out of 53 (50.9%; 95% Wilson CI 37.9% - 63.9%) tumors were detected in 16 out of 21 (76.2%; 95% Wilson CI 54.9% - 89.4%) tumor-positive volumes. On the held-out test set, 6 out of 7 (85.7%; 95% Wilson CI 48.7% - 97.4%) tumors were detected in 5 out of 6 (83.3%; 95% Wilson CI 43.6% - 97.0%) tumor-positive volumes. Given the very small number of positive cases, these intervals are wide and should be interpreted cautiously. The false positive rates in each volume ranged from approximately 0.5% to 2%. However, it is important to note that tumors only comprised around 0.01% to 0.0001% of each given volume. Therefore, there were 50 to 20,000 times more false positive pixels (usually brighter spots not representing tumors) than true positive pixels (representing tumors). This discrepancy between the false positive vs true positive pixels may not be an issue if radiologists are experienced and can readily dismiss clear cases of groupings of false positive pixels (such as variants of normality in bone marrow signal intensity) as false positives based on their specialized knowledge and complementary clinical and laboratory assessments. This false positive burden is an important limitation because even visually dismissible false positive regions may increase reader verification workload and influence how radiologists respond to the AI overlay.

Two radiologists, designated as radiologist 1 and radiologist 2, independently assessed pediatric wbMRIs under 2 conditions: with and without the AI-assisted diagnostic tool. We systematically analyzed their differences in labeling strategies, assessment duration, recommendations for clinical follow-up, and subjective evaluations of their workflow experience.

As shown in Figure 5, the average total time to evaluate each wbMRI for both radiologists was longer with the AI tool than without the AI tool. The use of the AI tool impacted the radiologists differently. While radiologist 1 spent more total time labeling the studies than radiologist 2, there was only a marginal time difference spent on labeling the studies with the AI tool than without the AI tool (Figure 5)—indicating that the AI tool only slightly increased the reporting times. Across both conditions, radiologist 1 marked twice as many candidate lesion locations as radiologist 2 (162 vs 81) and in the AI-assisted cases, radiologist 1 marked four times as many candidate lesion locations as radiologist 2 (22 vs 88) (Figure 5). In addition, radiologist 1 recommended follow-ups or biopsies more often than radiologist 2. The consideration of several more lesions as potentially worrisome is a rational explanation as to why radiologist 1 spent more time on labeling the studies than radiologist 2.

XSL•FO
RenderX

Figure 5. Reader-study results by radiologist and condition. (A) Mean evaluation time per volume; error bars indicate SD. (B) Total number of reader-marked candidate lesion locations across cases. (C) Counts of follow-up recommendations (no follow-up, follow-up, and biopsy). (D) Mean postcase Likert scores (1-5) for mental demand, hurriedness, perceived success, effort, and negative affect/stress; error bars indicate SD. "With tool" refers to cases read with the assistance of the AI overlay tool, and "Without tool" refers to cases read without it.



In contrast, although the AI-assisted samples had a higher mean reporting time for radiologist 2, it reduced the number of candidate lesion locations as potentially containing a tumor/lesion, in comparison to labeling without the AI tool. Given the low density of tumor slices in the volumes, the AI tool may have been helpful in reducing the number of false positives for radiologist 2 (Figure 5). Radiologist 2 also selected a smaller number of lesions for biopsy/follow-up than radiologist 1.

With regard to feedback on follow-up suggestions, the 2 radiologists overlapped 11 out of 25 (44.0%) times in their follow-up suggestions with the tool and 10 out of 25 (40.0%) times without the tool. With the tool, the 11 overlaps came from 6 overlaps in "No follow-up," 4 overlaps in "Follow-up," and 1 overlap in "Biopsy." Without the tool, the 10 overlaps came from 7 overlaps in "No follow-up," 2 overlaps in "Follow-up," and 1 overlap in "Biopsy."

Finally, the questionnaire was used to evaluate the subjective experience after each case. Reader 1 reported slightly higher effort with AI assistance (Q4) and slightly lower negative affect/stress with AI assistance (Q5). Reader 2 reported lower mental demand with AI assistance (Q1), higher hurriedness (Q2), similar perceived success (Q3), higher effort (Q4), and lower negative affect/stress (Q5). Because only 2 readers participated, these questionnaire findings should be interpreted descriptively rather than as generalized user-preference outcomes.

Discussion

This study contributes both a pediatric surveillance wbMRI tumor-likelihood overlay and an exploratory reader-study evaluation of how such an overlay may affect radiologist workflow. In addition to machine learning metrics, other considerations such as deployment scenario, human-computer interaction, and individual differences must be taken into account.

The low signal-to-noise ratio of pediatric wbMRI scans and the limited number of positive cases contributed to labeling difficulties in this study. While the AI tool aimed to mitigate the challenges of tumor detection, radiologists still seemed to double-check the AI's positive and negative suggestions, possibly subconsciously, in addition to accomplishing the

XSL•FO
RenderX

required task. This increased the time and impacted the efficiency of their assessments in different ways based on the radiologist.

Model development occurred in an extremely low-prevalence setting. The small number of tumor-positive volumes limited the diversity of tumor appearances, sizes, and anatomic locations represented in training and likely contributed to unstable performance estimates and persistent false positives. In addition, the intentionally enriched patch-sampling strategy improved learnability but does not reproduce real-world prevalence. A limitation of this study is that the model showed modest lesion-level performance with a substantial false positive burden. Radiologist behavior may reflect responses to an occasionally imperfect or noisy overlay in addition to the effect of a clinically reliable tumor-detection tool. These factors limit robustness and generalizability, and external validation in larger, more diverse, preferably multicenter cohorts is needed.

In the United States, radiologists have faced a high number of lawsuits due to missed diagnoses [39]. Therefore, radiologists generally prefer to "over-call" a finding rather than "under-call" or miss out on potentially lethal findings. For a tool designed to detect tumors to be adopted, it is crucial to have a low false negative rate. However, this does not justify the tool having a very high false positive rate, which then introduces a new challenge of differentiating between the true positives and the false positives, which in itself may also have adverse legal ramifications.

The main purpose of surveillance wbMRI in patients with cancer predisposition disorders is to avoid missed tumors. However, false positives are not a trivial trade-off in this setting. If the tool highlights many nontumor regions, radiologists must spend additional time verifying them, clinician trust in the AI overlay may decline, and patients and families may be exposed to additional follow-up imaging, biopsy discussions, or uncertainty. In Li-Fraumeni surveillance, where families already face substantial psychological burden, false positive prompts may also increase anxiety if they trigger further workup for ultimately benign findings. Reducing false positives while preserving sensitivity is therefore central to clinical adoption.

Our results reflect the nuanced impact of AI tools on clinical practice, with both radiologists demonstrating markedly different responses to the tool's guidance. The AI tool appeared to prolong the diagnostic process for one of the participating radiologists, possibly due to second-guessing the AI's recommendations, overrelying on the recommendations, or trying to verify its findings. In contrast, the other radiologist seemed to use the tool to streamline their decision-making in the sense that although more time was spent with the AI tool, the number of candidate lesion locations selected was much lower with the tool than without the tool, suggesting that the AI tool was able to root out a lot of false-positive cases. The subjective survey responses also suggest that the same AI overlay can be experienced as either helpful or effortful, depending on the individual reader, reinforcing that user experience should be considered reader-specific rather than uniform.

The divergent behavior of the 2 readers suggests that identical AI overlays may not be interpreted consistently across radiologists. One reader marked substantially more candidate lesions and recommended follow-up more often overall, whereas the other was more conservative. This variability indicates that overlay interpretation is itself a human-factors problem: the same AI output may be integrated differently depending on reader thresholds, risk tolerance, and trust calibration. Future deployment studies should therefore include reader onboarding, calibration, and interface testing rather than assuming a one-size-fits-all interaction model.

While many AI tools often achieve diagnostic results that are comparable to radiologists, their narrow task focus and need for large, annotated datasets limit their effectiveness. Furthermore, the "black box" nature of neural networks means that radiologists often second-guess AI outputs. The differences between radiologists' performance in this study suggest that some radiologists might spend extra time verifying AI outputs or modifying their usual workflow to incorporate AI feedback.

Our exploratory reader study reveals how individual radiologists' diagnostic workflows change with the introduction of AI tools in varying ways and illustrates how individual biases and preferences affect human-AI collaboration. Another study suggests that self-confidence, not confidence in AI, affects the decision to accept or reject AI suggestions the most [40]. Given that only 2 readers were studied, these differences should be interpreted as preliminary evidence of reader-specific heterogeneity rather than as a comprehensive characterization of interreader variability in clinical practice.

Our analysis revealed that the AI tool impacted radiologists differently. Radiologist 1 displayed little difference in average evaluation time with or without the AI tool, while radiologist 2 had a lower mean evaluation time labeling without the tool than with it. The AI tool appeared to streamline radiologist 2's decision-making process but also led to fewer candidates being selected, suggesting better exclusion of false positives. Despite the variance in time spent to conduct the requested tasks, the overlap in follow-up suggestions was similar for both radiologists with and without the AI tool in less than half the cases. More caution seems to be taken with the AI tool than without the AI tool with regard to the follow-up suggestions. Additionally, questionnaire results demonstrated varying subjective impacts on mental demand, hurriedness, irritation, workload, and perceived success with and without the tool.

Following the evaluation of our AI tool by 2 radiologists, we identified several challenges and opportunities in improving the usability of AI tools in health care. Based on the lessons learned from this study, the key insights are as follows:

- Different health care professionals with similar levels of training may interact with AI tools in various ways, influenced by personal biases and cognitive strategies, so calibrating AI tools to each individual represents a potentially promising future direction to explore.
- The opaque or black-box nature of many AI algorithms can lead to distrust and second-guessing of AI outputs by health care professionals; therefore, improved explainability and

XSL•FO
RenderX

interpretability may mitigate clinician skepticism and reduce redundant verification behaviors.
- Having clear and comprehensive training for health care professionals on how to effectively use AI tools is important in ensuring the proper use of the tools and improving satisfaction with their utilization.
- Poor integration of AI tools into existing workflows can disrupt routine practices and reduce overall efficiency, so employing a continuous evaluation and improvement process to refine AI tools based on real-world performance and user feedback may prove useful.
- The introduction of AI tools can increase cognitive load and mental demand if the interactive component is not up to par. Therefore, making health care professionals a part of the discussion on how to develop user-friendly AI tools that can help reduce friction with existing workflows is crucial.
- The technical complexity of AI tools can pose a barrier to their effective use by health care professionals who are not technically inclined. Therefore, it is important to create user-centric designs that serve the needs and preferences of health care professionals from a wide spectrum of experiences.
- Fostering collaboration between AI developers, health care professionals, and researchers in the development process to ensure AI tools are clinically relevant and serve precise clinical needs.
- Encouraging the use of AI tools as supportive aids, rather than replacements for human imaging interpretation, can help avoid overconfidence or underconfidence in the AI tools.

In summary, this study addresses a clinically specific and technically challenging setting, pediatric surveillance wbMRI in LFS, and shows that AI overlays can influence radiologist workflow in reader-specific ways. We did not formally disentangle appearance-driven signals from implicit contextual/location priors, so future work should test whether the model remains sensitive to tumors in less commonly represented locations. The reader study should be interpreted as preliminary. Only 2 radiologists participated, and the design was not randomized or counterbalanced. Therefore, the observed differences in timing, candidate-lesion marking, and subjective responses should be viewed as descriptive and hypothesis-generating rather than definitive estimates of clinical workflow impact or trust or as robust estimates of population-level reader effects. The nuanced relationship between radiologists and AI tools requires further research to optimize these systems for diverse diagnostic workflows. This study emphasizes the interplay between trust, validation, and interpretation in human-AI collaboration, highlighting that a one-size-fits-all strategy may be inadequate for optimizing AI tools in health care.

Acknowledgments

The authors used the generative AI tool ChatGPT by OpenAI to assist with editing the manuscript for grammar, spelling, word choice, and flow. All suggested revisions were carefully reviewed and further revised by the authors.

Funding

This work was supported by the Hospital for Sick Children, the Vector Institute for Artificial Intelligence, the University of Toronto, Natural Sciences and Engineering Research Council of Canada, Canadian Institutes of Health Research, CIFAR, and the Mark Foundation for Cancer Research.

Data Availability

The data analyzed during this study are not publicly available because they contain sensitive pediatric health information and are subject to institutional privacy, data governance, and research ethics restrictions. The authors are not authorized to share the underlying data, including upon individual request. Inquiries regarding these restrictions may be directed to the corresponding author.

Conflicts of Interest

ASD reports research grants from Terry Fox Foundation, grants from PSI Foundation, grants from Society of Pediatric Radiology, grants from Radiological Society of North America, grants from CARRA Foundation, and grants from Novo Nordisk Access to Insight Grant outside the submitted work.

References

1. Scheetz J, Rothschild P, McGuinness M, et al. A survey of clinicians on the use of artificial intelligence in ophthalmology, dermatology, radiology and radiation oncology. Sci Rep 2021 Mar 4;11(1):5193. [doi: 10.1038/s41598-021-84698-5] [Medline: 33664367]
2. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H. Artificial intelligence in radiology. Nat Rev Cancer 2018 Aug;18(8):500-510. [doi: 10.1038/s41568-018-0016-5] [Medline: 29777175]
3. Chang HY, Jung CK, Woo JI, et al. Artificial intelligence in pathology. J Pathol Transl Med 2019 Jan;53(1):1-12. [doi: 10.4132/jptm.2018.12.16] [Medline: 30599506]

XSL•FO
RenderX

4. Du-Harpur X, Watt FM, Luscombe NM, Lynch MD. What is AI? Applications of artificial intelligence to dermatology. Br J Dermatol 2020 Sep;183(3):423-430. [doi: 10.1111/bjd.18880] [Medline: 31960407]
5. Kapoor R, Walters SP, Al-Aswad LA. The current state of artificial intelligence in ophthalmology. Surv Ophthalmol 2019;64(2):233-240. [doi: 10.1016/j.survophthal.2018.09.002] [Medline: 30248307]
6. Blease C, Kaptchuk TJ, Bernstein MH, Mandl KD, Halamka JD, DesRoches CM. Artificial intelligence and the future of primary care: exploratory qualitative study of UK general practitioners' views. J Med Internet Res 2019 Mar 20;21(3):e12802. [doi: 10.2196/12802] [Medline: 30892270]
7. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. Ann Surg 2018 Jul;268(1):70-76. [doi: 10.1097/SLA.0000000000002693] [Medline: 29389679]
8. Patel UK, Anwar A, Saleem S, et al. Artificial intelligence as an emerging technology in the current care of neurological disorders. J Neurol 2021 May;268(5):1623-1642. [doi: 10.1007/s00415-019-09518-3] [Medline: 31451912]
9. Lopez-Jimenez F, Attia Z, Arruda-Olson AM, et al. Artificial intelligence in cardiology: present and future. Mayo Clin Proc 2020 May;95(5):1015-1039. [doi: 10.1016/j.mayocp.2020.01.038] [Medline: 32370835]
10. Rajpurkar P, Irvin J, Zhu K, et al. Chexnet: radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv. Preprint posted online on Nov 14, 2017. [doi: 10.48550/arXiv.1711.05225]
11. Harish V, Morgado F, Stern AD, Das S. Artificial intelligence and clinical decision making: the new nature of medical uncertainty. Acad Med 2021 Jan 1;96(1):31-36. [doi: 10.1097/ACM.0000000000003707] [Medline: 32852320]
12. Soenksen LR, Ma Y, Zeng C, et al. Integrated multimodal artificial intelligence framework for healthcare applications. NPJ Digit Med 2022 Sep 20;5(1):149. [doi: 10.1038/s41746-022-00689-4] [Medline: 36127417]
13. Harris E. AI-drafted responses to patients reduced clinician burnout. JAMA 2024 May 7;331(17):1440. [doi: 10.1001/jama.2024.5157] [Medline: 38607634]
14. Yu F, Moehring A, Banerjee O, Salz T, Agarwal N, Rajpurkar P. Heterogeneity and predictors of the effects of AI assistance on radiologists. Nat Med 2024 Mar;30(3):837-849. [doi: 10.1038/s41591-024-02850-w] [Medline: 38504016]
15. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res 2020 Jun 19;22(6):e15154. [doi: 10.2196/15154] [Medline: 32558657]
16. Chong L, Zhang G, Goucher-Lambert K, Kotovsky K, Cagan J. Human confidence in artificial intelligence and in themselves: the evolution and impact of confidence on adoption of AI advice. Comput Human Behav 2022 Feb;127:107018. [doi: 10.1016/j.chb.2021.107018]
17. Choudhury A. Factors influencing clinicians' willingness to use an AI-based clinical decision support system. Front Digit Health 2022;4:920662. [doi: 10.3389/fdgth.2022.920662] [Medline: 36339516]
18. Kumar P, Chauhan S, Awasthi LK. Artificial intelligence in healthcare: review, ethics, trust challenges & future research directions. Eng Appl Artif Intell 2023 Apr;120:105894. [doi: 10.1016/j.engappai.2023.105894]
19. Khalid N, Qayyum A, Bilal M, Al-Fuqaha A, Qadir J. Privacy-preserving artificial intelligence in healthcare: techniques and applications. Comput Biol Med 2023 May;158:106848. [doi: 10.1016/j.compbiomed.2023.106848] [Medline: 37044052]
20. Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in health care. JAMA 2019 Dec 24;322(24):2377-2378. [doi: 10.1001/jama.2019.18058] [Medline: 31755905]
21. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. J Am Med Inform Assoc 2020 Mar 1;27(3):491-497. [doi: 10.1093/jamia/ocz192] [Medline: 31682262]
22. Huisman M, Ranschaert E, Parker W, et al. An international survey on AI in radiology in 1,041 radiologists and radiology residents part 1: fear of replacement, knowledge, and attitude. Eur Radiol 2021 Sep;31(9):7058-7066. [doi: 10.1007/s00330-021-07781-5] [Medline: 33744991]
23. Schlicker N, Langer M. Towards warranted trust: a model on the relation between actual and perceived system trustworthiness. Presented at: MuC '21: Proceedings of Mensch und Computer 2021; Sep 5-8, 2021; Ingolstadt, Germany p. 325-329. [doi: 10.1145/3473856.3474018]
24. Moturu A, Joshi S, Doria AS, Goldenberg A. Volume-based performance not guaranteed by promising patch-based results in medical imaging. Presented at: I Can't Believe It's Not Better! - Understanding Deep Learning Through Empirical Falsification; Dec 3, 2022 URL: https://proceedings.mlr.press/v187/moturu23a/moturu23a.pdf [accessed 2026-07-04]
25. Villani A, Shore A, Wasserman JD, et al. Biochemical and imaging surveillance in germline TP53 mutation carriers with Li-Fraumeni syndrome: 11 year follow-up of a prospective observational study. Lancet Oncol 2016 Sep;17(9):1295-1305. [doi: 10.1016/S1470-2045(16)30249-2] [Medline: 27501770]
26. Darge K, Jaramillo D, Siegel MJ. Whole-body MRI in children: current status and future applications. Eur J Radiol 2008 Nov;68(2):289-298. [doi: 10.1016/j.ejrad.2008.05.018] [Medline: 18799279]
27. Liu L, Cheng J, Quan Q, Wu FX, Wang YP, Wang J. A survey on U-shaped networks in medical image segmentations. Neurocomputing 2020 Oct;409:244-258. [doi: 10.1016/j.neucom.2020.05.070]
28. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. Presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015); Oct 5-9, 2015. [doi: 10.1007/978-3-319-24574-4_28]
29. Chang A, Suriyakumar V, Moturu A, et al. 3D reasoning for unsupervised anomaly detection in pediatric WbMRI. arXiv. Preprint posted online on Mar 24, 2021. [doi: 10.48550/arXiv.2103.13497]

XSL•FO
RenderX

30. Chang A, Suriyakumar VM, Moturu A, Tewattanarat N, Doria A, Goldenberg A. Using generative models for pediatric wbMRI. arXiv. Preprint posted online on Jun 1, 2020. [doi: 10.48550/arXiv.2006.00727]
31. Bernal J, Kushibar K, Cabezas M, Valverde S, Oliver A, Llado X. Quantitative analysis of patch-based fully convolutional neural networks for tissue segmentation on brain magnetic resonance imaging. IEEE Access 2019;7:89986-90002. [doi: 10.1109/ACCESS.2019.2926697]
32. Cordier N, Delingette H, Ayache N. A patch-based approach for the segmentation of pathologies: application to glioma labelling. IEEE Trans Med Imaging 2016 Apr;35(4):1066-1076. [doi: 10.1109/TMI.2015.2508150] [Medline: 26685225]
33. Cui Z, Yang J, Qiao Y. Brain MRI segmentation with patch-based CNN approach. Presented at: 2016 35th Chinese Control Conference (CCC); Jul 27-29, 2016; Chengdu, China p. 7026-7031. [doi: 10.1109/ChiCC.2016.7554465]
34. Chang A, Moturu A. Detecting early stage lung cancer using a neural network trained with patches from synthetically generated x-rays. : Department of Computer Science, University of Toronto; 2019 URL: https://www.cs.toronto.edu/pub/reports/na/Project_Report_Moturu_Chang_2.pdf [accessed 2026-07-04]
35. Moturu A, Chang A. Creation of synthetic x-rays to train a neural network to detect lung cancer. : Department of Computer Science, University of Toronto; 2018 URL: https://www.cs.toronto.edu/pub/reports/na/Project_Report_Moturu_Chang_1.pdf [accessed 2026-07-04]
36. Yeung M, Sala E, Schönlieb CB, Rundo L. Unified focal loss: generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. Comput Med Imaging Graph 2022 Jan;95:102026. [doi: 10.1016/j.compmedimag.2021.102026] [Medline: 34953431]
37. Milletari F, Navab N, Ahmadi SA. V-Net: fully convolutional neural networks for volumetric medical image segmentation. Presented at: 2016 Fourth International Conference on 3D Vision (3DV); Oct 25-28, 2016. [doi: 10.1109/3DV.2016.79]
38. Jurdi RE, Petitjean C, Honeine P, Cheplygina V. A surprisingly effective perimeter-based loss for medical image segmentation. Presented at: Medical Imaging with Deep Learning (MIDL); Jul 7-9, 2021 p. 158-167 URL: https://proceedings.mlr.press/v143/el-jurdi21a/el-jurdi21a.pdf [accessed 2026-07-04]
39. Whang JS, Baker SR, Patel R, Luk L, Castro A III. The causes of medical malpractice suits against radiologists in the United States. Radiology 2013 Feb;266(2):548-554. [doi: 10.1148/radiol.12111119] [Medline: 23204547]
40. Chong L, Raina A, Goucher-Lambert K, Kotovsky K, Cagan J. The evolution and impact of human confidence in artificial intelligence and in themselves on AI-assisted decision-making in design. J Mech Des 2023;145(3):031401. [doi: 10.1115/1.4055123]

Abbreviations

LFS: Li-Fraumeni syndrome
MRI: magnetic resonance imaging
STIR: Short Tau Inversion Recovery
wbMRI: whole-body magnetic resonance imaging

Edited by A Kushniruk; submitted 22.Jul.2025; peer-reviewed by SM Shaffi, S Ye, TM Alsady; revised version received 01.Jun.2026; accepted 23.Jun.2026; published 07.Aug.2026.

Please cite as:
Moturu A, Komolafe O, Joshi S, Doria AS, Goldenberg A
Human-AI Interaction With AI-Assisted Tumor Overlays in Pediatric Whole-Body Magnetic Resonance Imaging: Exploratory Reader Study
JMIR Hum Factors 2026;13:e81066
URL: https://humanfactors.jmir.org/2026/1/e81066
doi:10.2196/81066

© Abhishek Moturu, Olurotimi Komolafe, Sayali Joshi, Andrea S Doria, Anna Goldenberg. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 7.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effects of AI Assistance Timing on Pharmacists' Trust in Automated Pill Recognition Technology: Within-Participants Experimental Study

Jin Yong Kim[1], PhD; Brigid Rowell[2], MA; Megan Whitaker[2], MHI; Qiyuan Chen[1], MSE; Raed Al Kontar[1], PhD; Corey Lester[2], PharmD, PhD; Xi Jessie Yang[1], PhD

[1]Industrial and Operations Engineering, University of Michigan, 1640 IOE, 1205 Beal Ave, Ann Arbor, MI, United States

[2]College of Pharmacy, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Xi Jessie Yang, PhD
Industrial and Operations Engineering, University of Michigan, 1640 IOE, 1205 Beal Ave, Ann Arbor, MI, United States

Abstract

Background: Image-based pill verification systems demonstrate high model accuracy. However, their effectiveness in pharmacy practice depends on how pharmacists interact with the AI output. The timing of AI advice is one design factor that influences these interactions, yet its impact on pharmacists' moment-to-moment trust dynamics during medication verification requires further investigation.

Objective: This study aims to investigate how the timing and conditionality of AI assistance shape pharmacists' trust dynamics during medication verification.

Methods: Between April and December 2024, 50 licensed pharmacists completed a browser-based simulated medication dispensing task with 2 AI types: ex-ante advice (AI advice given concurrently with clinical information) and ex-post advice (AI advice given after an initial diagnosis). Ex-post advice was further divided into the involved ex-post and the not-involved ex-post conditions. The experiment used a within-participants design with varying AI types and AI recognition patterns (right fill-correct recognition, right fill-incorrect recognition, wrong fill-correct recognition, and wrong fill-incorrect recognition). The primary outcomes were trust adjustment magnitude and trust adjustment, which were analyzed using mixed-effects linear regression models.

Results: Trust adjustment magnitude differed significantly across AI assistance conditions. The involved ex-post condition led to the highest magnitude of trust adjustment, followed by ex-ante advice (mean difference 9.65, 95% CI 8.55-10.75; $P<.001$), with the not-involved ex-post condition showing the lowest magnitude (mean difference 3.33, 95% CI 2.63-4.03; $P<.001$). Analysis by each recognition pattern revealed significant differences in trust adjustment when the right drugs were incorrectly rejected. In this pattern, the involved ex-post condition led to larger trust decrements than both ex-ante advice (mean difference −2.13, 95% CI −3.83 to −0.44; $P=.008$) and not-involved ex-post conditions (mean difference −7.32, 95% CI −13.69 to −.95; $P=.009$). A marginal difference was observed between ex-ante advice and not-involved ex-post conditions (mean difference −5.19, 95% CI −11.55 to 1.17; $P=.051$). No significant differences were observed for other recognition patterns.

Conclusions: Both the timing and conditionality of AI assistance influenced pharmacists' trust dynamics. Disagreement-based AI interventions (involved ex-post) that incorrectly challenged pharmacists led to a substantial trust decrement, whereas the not-involved AI intervention resulted in more stable trust fluctuations. These findings highlight the importance of designing AI systems that align intervention strategies with user expertise and task demands to foster appropriate trust in safety-critical workflows.

Trial Registration: ClinicalTrials.gov NCT06245044; https://clinicaltrials.gov/study/NCT06245044

(JMIR Hum Factors 2026;13:e92822) doi:10.2196/92822

KEYWORDS

artificial intelligence; human-computer interaction; medication verification; medication errors; trust in automation

XSL•FO
RenderX

Introduction

Background

Pharmacists play a critical role in ensuring that patients receive the correct medication, particularly during medication verification, in which the medication filled inside the bottle must match the prescription label [1]. Errors at this stage, such as dispensing the wrong drug, strength, or formulation, are a major source of preventable adverse drug events, with error rates that can exceed 1% in both community and hospital settings [2-4]. These errors adversely impact patient safety and trigger emergency room visits and hospitalizations [5]. Contributing factors include high workload, fatigue, technology limitations, lack of trust in automation, and the high cognitive demands placed on pharmacists who must manage multiple tasks simultaneously [6-9]. Because medication verification is both labor-intensive and cognitively demanding, tools that reduce dispensing errors have the potential to substantially improve patient outcomes and decrease health care costs [10].

To mitigate these risks, image-based pill verification systems have been introduced [11-15]. Advances in deep learning have further improved the performance of these systems, with recent models achieving accuracies greater than 98% [16]. However, the successful integration of AI systems into pharmacy workflows depends not only on model accuracy but also on how humans perceive, trust, interpret, and act on AI output [8,17,18]. Prior human factors and human-computer interaction (HCI) research highlights two design dimensions that strongly influence human-AI interaction: (1) the amount and type of model information presented and (2) the timing of AI assistance [19,20]. These dimensions shape cognitive load, situational awareness, and the development of trust in autonomous and AI systems, defined as the belief that an autonomous or AI agent will support an individual's goals under conditions of uncertainty [21].

The amount and type of model information can play a key role in promoting clinician understanding of clinical decision support systems (CDSSs). To overcome the barrier of the "black-box" nature of many of these CDSSs, in which model outputs provide little insight into how predictions are generated, research has emphasized the importance of enhancing system transparency through explanatory feedback and uncertainty visualizations [9,22-26]. For instance, Kompa et al [26] investigated how different levels of explanation in a CDSS supporting balance disorder diagnosis influenced health care professionals' trust and reliance. Participants used either a selective CDSS that presented examination results alone or a comprehensive CDSS that additionally provided explanatory context. Their findings showed that explanations positively affected trust, whereas limited explanations led users to question the system's reliability [26]. Kim et al [9] examined how presenting model uncertainty information in a medication verification AI affected pharmacists' trust. Their pill characteristic plots conveyed match-mismatch information based on visual attributes such as imprint, color, shape, and score, and the uncertainty plot illustrated the predicted probability of each set of predictions summarized in a histogram [11]. By manipulating the presence of uncertainty visualization as a between-participants factor, they found that communicating uncertainty improved pharmacists' end trust, suggesting that transparency fosters more calibrated trust and active engagement with AI systems [9].

Another critical design factor influencing human-AI interaction is the timing of AI advice [20]. In clinical decision-making, AI systems may present recommendations before the human makes the initial decision (ex-ante AI advice) or after the human forms an initial judgment (ex-post AI advice) [19,27]. Ex-ante systems provide continuous feedback throughout the task [28,29], supporting real-time integration of AI insights, but increasing the risk of cognitive overload or anchoring when AI outputs are incorrect [30-32]. In contrast, ex-post systems require the human decision maker to commit to an initial decision before receiving AI advice [33-36]. While this approach can preserve human decision autonomy and reduce the anchoring effect, it may exacerbate confirmation bias [37-39], contribute to algorithm aversion [35,40,41], and introduce workflow interruptions when AI intervenes unnecessarily [42].

Empirical findings on the optimal timing of AI assistance remain mixed. Yin et al [27] investigated how the timing of AI advice affects doctors' diagnostic decision-making. In their experiment, they recruited and asked physicians to diagnose clinical cases. The authors compared ex-post advice with ex-ante advice. They found that physicians engaged more deeply with ex-post advice, resulting in higher diagnostic accuracy (task performance) compared with ex-ante advice. Conversely, Fogliato et al [43] explored a human-AI workflow sequence with veterinary radiologists, who reviewed X-ray images and made diagnostic judgments under two conditions: (1) a 1-step workflow (ex-ante advice), where AI inferences were shown before radiologists' diagnosis and (2) a 2-step workflow (ex-post advice), where radiologists made an initial diagnosis first, then saw the AI's recommendations. The authors reported that radiologists' diagnostic performance was better in the ex-ante advice system compared to the ex-post advice system. With the ex-post advice help, radiologists rarely revised their initial judgments and rated the AI as less useful [43].

Properties of Trust Dynamics

Trust in autonomous and AI systems has been extensively studied across domains, including health care [9,17,21]. Trust is a critical determinant of how people rely on autonomous systems and is strongly influenced by system performance [44-46].

Most existing literature has adopted a snapshot approach to trust measurement, typically assessing trust through posttask or end-of-experiment questionnaires. While informative, this approach overlooks that trust is a dynamic variable and evolves over the course of repeated, moment-to-moment interactions with autonomous systems [45,47,48].

Yang et al [45] identified and summarized 3 key trust dynamics properties. First, continuity suggests that trust at a given moment (i) is strongly associated with trust at the immediately preceding moment ($i - 1$). Trust tends to increase following automation successes and decrease following failures [9,46]. Second, negativity bias highlights the asymmetric impact of autonomy

XSL•FO
RenderX

performance, wherein trust is difficult to build but can be lost quickly. Prior studies have shown that negative experiences with automation failures exert a stronger influence on trust than positive experiences [8,49-51]. Third, stabilization refers to the tendency for trust to stabilize over repeated interactions with the same system, as users gain familiarity and form more stable expectations [52].

Present Study

Building on these findings, the present study investigates how the timing of AI assistance shapes pharmacists' trust dynamics during medication verification. We examine 2 modes of AI engagement: ex-post advice (AI advice given after an initial diagnosis) and ex-ante advice (AI advice given concurrently with clinical information). The ex-ante advice continuously displays predictive information throughout the task, and the ex-post advice intervenes only when its prediction conflicts with the pharmacist's initial decision. Ex-ante advice may support continuous decision-making but risks cognitive overload and anchoring, whereas ex-post advice may reduce unnecessary attentional demands but risks disruptive false alarms [30,32-34,36].

By comparing these timing strategies, we aim to identify how different modes of AI engagement influence trust adjustment in a safety-critical and expert-driven medication verification task.

In addition, prior empirical validations of trust dynamics properties (continuity, negativity bias, and stabilization) have relied primarily on simplified tasks and nonexpert participant populations [8,45,48]. By examining licensed pharmacists engaged in realistic medication verification tasks, the present study extends prior work and strengthens the ecological validity of trust dynamics properties in a safety-critical health care context.

Methods

Ethical Considerations

This study was exempt from institutional review board oversight by the University of Michigan (HUM#00241223). All participants provided electronic informed consent prior to the experiment, and data were collected anonymously. Participants received US $100 upon completing the study.

Participants

Pharmacists were recruited via email through the Minnesota Pharmacy Practice-Based Research Network, Pharmacy Practice Enhancement and Active Research Link in Wisconsin, and the University of Michigan College of Pharmacy Pharmacist Preceptor Network. Participants were eligible for the study if they were licensed in the United States, were at least 18 years old, and had access to a laptop or desktop computer with a webcam. Pharmacists were excluded if they needed assistive technology to operate a computer, wore eyeglasses with more than 1 power, had uncorrected cataracts, intraocular implants, glaucoma, or permanently dilated pupils, and/or had eye movement or alignment abnormalities (eg, lazy eye, strabismus, nystagmus). Fifty licensed US pharmacists participated in the study between April and December 2024.

AI Model

The AI model used in this study is a Bayesian neural network (BNN) designed to predict the National Drug Code (NDC) for dispensed pills while estimating the prediction uncertainty [11,16]. This was achieved by applying random dropout to a ResNet-34 convolutional neural network architecture [53,54]. Instead of outputting a single probability distribution for each class, the model generated 50 probability vectors for each image. The final NDC prediction was determined by averaging those vectors and selecting the class with the highest mean probability.

The model was trained on a dataset of 432,974 images representing 352 distinct NDC classes from a mail-order pharmacy in the United States, which uses a robotic system to count pills, fill and label vials, capture images of the contents, and seal the vials [55]. The dataset contains 1 year of robot-captured images of solid oral medication (ie, tablets and capsules) inside prescription vials, with each image associated with its corresponding NDC and annotated with attributes such as color, shape, size, manufacturer, tablet scoring, and imprint. The number of available images for each NDC varied from 3 to 12105, with a median of 540 (IQR 257-1291).

The medications in the dataset were categorized into 12 distinct colors: white (42.1%), yellow (12.3%), pink (9.1%), orange (7.1%), multicolor (5.9%), green (5.2%), red (5.2%), blue (4.8%), brown (3.8%), purple (3.1%), turquoise (0.7%), and gray (0.7%). Additionally, seven unique shapes were categorized: round (49.6%), oval (33.4%), capsule (16.2%), hexagon-6-sided (0.4%), triangle (0.3%), trapezoid (0.1%), and pentagon-5-sided (0.04%).

Experimental Apparatus

Overview

The study utilized Labvanced, a browser-based experimental platform, to conduct simulated medication dispensing tasks. Participants accessed the testbed through a web browser on their personal computers, which enabled the software to collect eye-tracking data through the webcam while also recording response decisions and reaction times for each task.

Experimental Task

In this experiment, participants performed a simulated medication verification task supported by varying types of imperfect automated pill recognition systems [9,12,13]. The user interface displayed a reference image of prescription medication, the prescription details, and an image of the filled medication. In the 2 AI-assistance conditions, the AI-generated decision support was also presented. The objective was to determine whether the filled medication matched the prescription (ie, reference image). If the 2 images matched, the correct response was to click "accept"; if they differed, the appropriate response was to click "reject."

The experimental stimuli, including the reference NDC and corresponding filled medication images, were carefully selected from the dataset of 432,974 images. The selection prioritized a

XSL•FO
RenderX

broad representation of medication colors and shapes while excluding blurry images to ensure clarity. To reduce potential learning effects, each reference NDC was presented no more than twice during the experiment. To ensure sufficient exposure to both correct and incorrect AI predictions, we constructed a 300-image experimental dataset (representing 154 distinct NDCs) by selectively sampling from the larger dataset to increase the proportion of naturally occurring incorrect predictions, resulting in an overall AI accuracy of approximately 79%. NDCs could occur no more than twice to prevent bias. AI outputs were not artificially modified, and all predictions reflect the model's original performance. Because this sampling approach preserved the model's inherent error patterns, prediction errors were not uniformly distributed across medication characteristics (see Table 1).

Table . Distribution of medication characteristics for the experimental dataset (n=300).

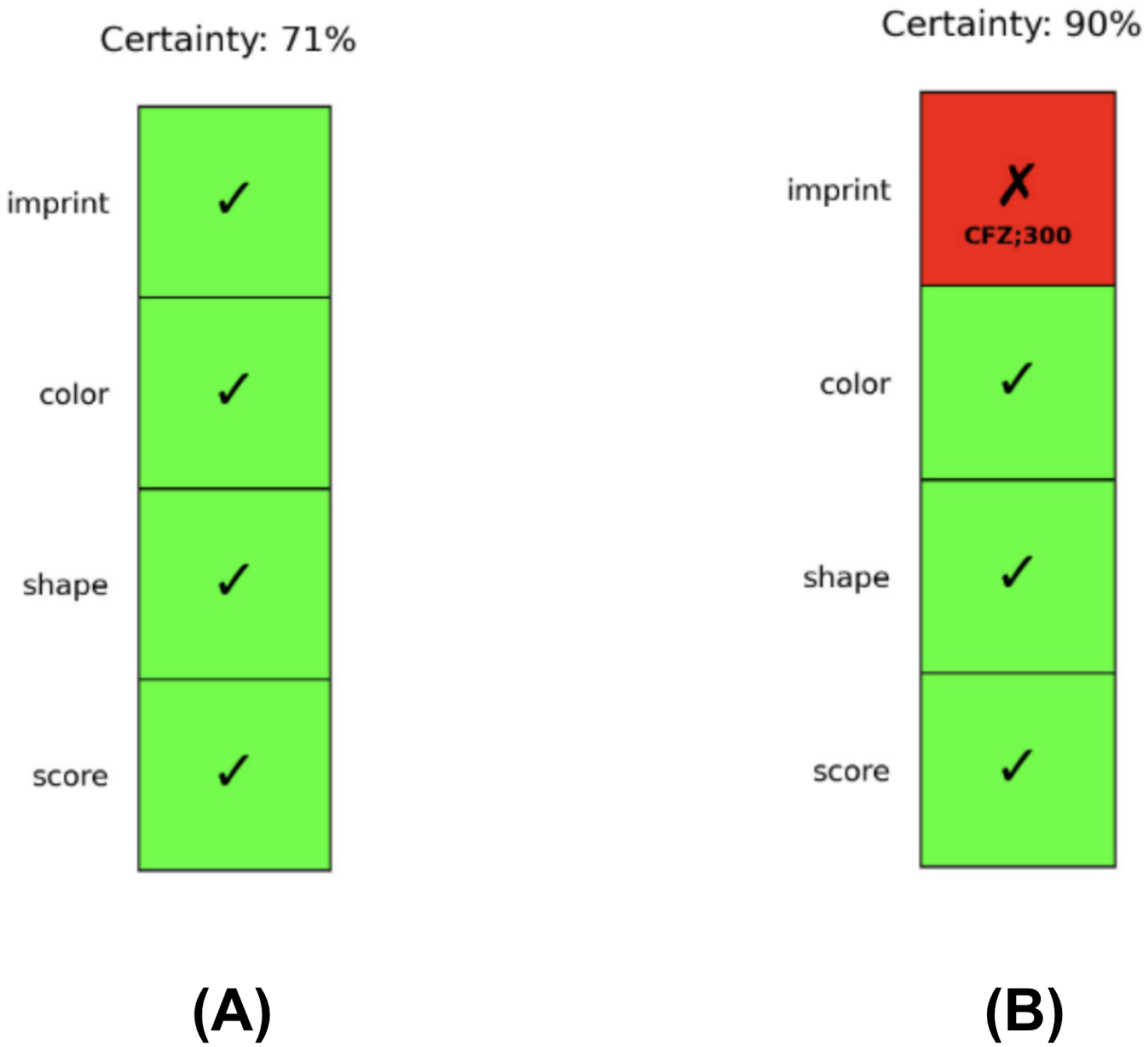
Characteristics	Filled, n (%)	Reference, n (%)
Color		
White	154 (51)	160 (53)
Yellow	42 (14)	34 (11)
Pink	24 (8)	22 (7)
Orange	21 (7)	25 (8)
Multicolor	4 (1)	4 (1)
Green	14 (5)	14 (5)
Red	7 (2)	7 (2)
Blue	12 (4)	12 (4)
Brown	14 (5)	14 (5)
Purple	4 (1)	4 (1)
Turquoise	4 (1)	4 (1)
Shape		
Round	181 (60)	187 (62)
Oval	80 (27)	72 (24)
Capsule	35 (12)	37 (12)
Hexagon (6-sided)	2 (1)	2 (1)
Triangle	2 (1)	2 (1)

To characterize the nature of AI errors, the visual similarity between the reference medication and the AI-predicted medication was examined. Most incorrect predictions (93%) occurred between medications sharing the same color and shape (ie, visually similar medications with different NDCs). A small proportion involved a shape mismatch (5%) or a color mismatch (2%), and none involved mismatches in both color and shape.

Two types of AI assistance, both powered by the same underlying BNN model, were integrated into the interface. The AI model communicated its confidence (or uncertainty) and feature matches through a checkbox plot indicating 4 key pill characteristics: imprint, color, shape, and score (Figure 1A). Matching characteristics were represented using green checkboxes with checkmarks. Mismatches were represented using red checkboxes marked with an "X," along with a textual description of the AI-predicted characteristic (eg, the predicted imprint shown in Figure 1B). The AI's confidence was represented by the predicted probability of the highest-ranked NDC class.

XSL•FO
RenderX

Figure 1. Checkbox plots. (A) Checkbox plot indicating a match and (B) checkbox plot indicating a mismatch. Matching characteristics were represented using green checkboxes with checkmarks. Mismatches were represented using red checkboxes marked with an "X," along with a textual description of the AI-predicted characteristic.



The ex-ante advice condition continuously displayed the checkbox plot before the participant made a decision (Figure 2A). Feedback was displayed after the participant made the decision, indicating the accuracy of the participant's selection and the AI's prediction (Figure 2B). On the other hand, the ex-post advice condition served as a "safeguard," appearing only when the participant's response contradicted the AI's prediction. For example, if the participant initially selected "accept" (Figure 3A) when the AI predicted a mismatch, the AI intervened and prompted a reverification by displaying a message, such as: "You selected Accept. The AI disagrees. Confirm or change your response" (Figure 3B). The AI assistance was intentionally imperfect, occasionally providing incorrect recommendations. Although the AI model's accuracy was 98.5%, it was deliberately lowered to 79% to ensure sufficient exposure to both AI's errors and AI's correct guidance outputs. This manipulation enabled the systematic observation of pharmacists' responses to AI successes and failures. The selected reliability level was informed by prior human factors research suggesting that when automation reliability falls below approximately 70%, its costs may outweigh its benefits, highlighting the importance of studying user responses across a range of imperfect, yet still functional, automation performance levels [56].

XSL•FO
RenderX

Figure 2. The ex-ante advice condition displayed the checkbox plot before the participant made a decision. (A) Ex-ante advice selection and (B) ex-ante advice feedback. Feedback was displayed after the initial decision.

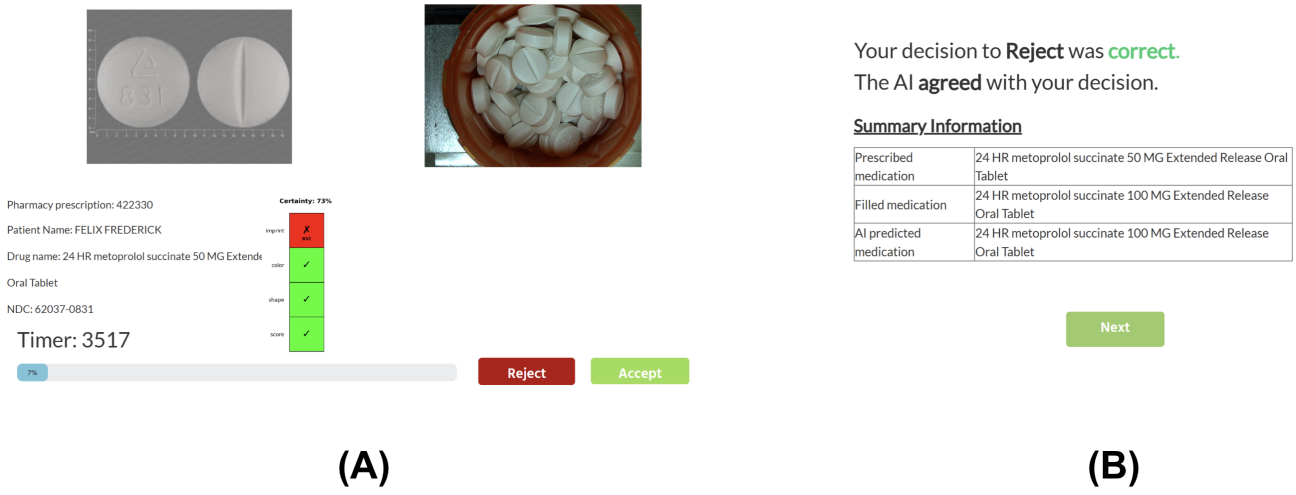
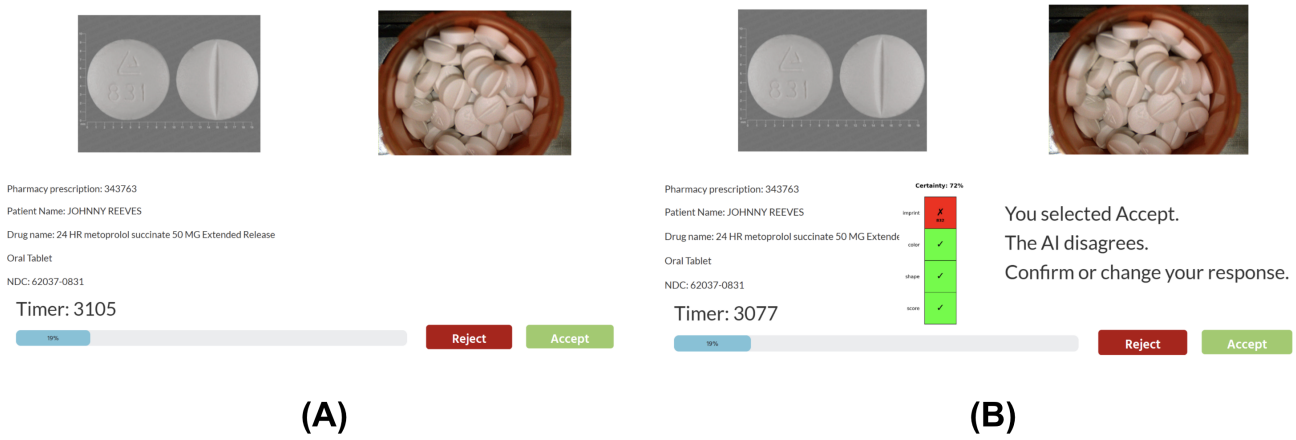


Figure 3. The ex-post advice condition served as a "safeguard," appearing only when the participant's response contradicted the AI's prediction. (A) Ex-post advice initial selection and (B) ex-post advice intervention.



Participants completed 3 blocks of 100 medication verification trials, with each block corresponding to each AI assistance condition, including a no-assistance control condition. In the AI-assisted blocks, participants received trial-by-trial feedback on indicating whether their decision to accept or reject was correct and whether the AI's recommendation was accurate (ie, "Your decision to accept was correct. The AI agreed with your decision"). Following each AI-assisted trial, participants rated their trust in the AI using a visual analog scale ranging from 0 to 100, anchored by "Not at all trust" (0) on the far left to "Completely trust" (100) on the far right. Because the no-assistance control condition did not involve AI support, no trust ratings were obtained.

Experimental Design

The experiment used a within-participants design with varying AI types (ex-ante advice and ex-post advice) and AI recognition patterns (right fill-correct recognition, right fill-incorrect recognition, wrong fill-correct recognition, and wrong fill-incorrect recognition). Within the ex-post advice condition, AI assistance was further classified as the involved ex-post condition (ex-post-AI disagreed with the human decision) and the not-involved ex-post condition (ex-post-AI agreed with the human decision). The agreement conditions depended on participants' responses and therefore could not be predetermined. Block order was counterbalanced across participants using a Latin square design. Each AI type block consisted of 100 medication verification trials randomly selected by the experimental platform from the larger trial pool, including 60 right fill-correct recognition, 16 right fill-incorrect recognition, 19 wrong fill-correct recognition, and 5 wrong fill-incorrect recognition trials.

Measures

After each trial i, participants report their $trust(i)$ in the AI. We calculate a trust adjustment as:

$$Trust\ adjustment(i) = Trust(i) - Trust(i-1), \text{ where } i = 2, 3, ..., 100$$

Since the moment-to-moment trust is reported after each trial, 99 trust adjustments are obtained from each participant.

Participant demographics, trust propensity, and eye-tracking data were also collected. Trust propensity was evaluated as a potential covariate during model development, but was not a significant predictor and did not alter the results; therefore, it was excluded from the final models. Because the present study focuses on within-participant trust adjustment, demographic and eye-tracking analyses are beyond the scope of this paper and are reported in a companion manuscript.

XSL•FO
RenderX

Experimental Procedure

Before the experiment, participants met with a member of the research team via Zoom (Zoom Communications Inc.) to verify that technical and environmental requirements were met, including a functioning webcam, adequate lighting, and a quiet workspace. Once confirmed, participants received a secure link and password to access the Labvanced software. An instructional video then introduced the mock verification process using the study interface, explaining that the objective was to determine whether an image of a filled medication vial matched the prescribed reference image.

The video also outlined the 2 AI assistance features.

Prior to starting the verification tasks, participants completed a demographic questionnaire, the Occupational Fatigue Exhaustion/Recovery Scale (OFER), and the trust propensity survey, followed by calibration procedures for the eye-tracking software [57,58].

During each trial, a "reference image" (the prescribed medication) appeared on the left, and a "filled vial image" (the dispensed medication) appeared on the right, along with relevant prescription details (Figure 2A). Participants decided whether to accept or reject the filled medication based on this comparison. Each participant completed 100 mock verification trials under the ex-ante advice condition (Figure 2A), 100 trials under the ex-post advice condition (Figure 3B), and 100 trials without AI assistance, in a counterbalanced order. The 6 block orders were assigned approximately evenly across 50 participants (4 sequences: n=8 each; 2 sequences: n=9 each). After each block of 100 trials, participants completed postcondition surveys, including the trust survey on a 5-point Likert scale, the system usability scale, the NASA Task Load Index (NASA-TLX) survey, and nonmandatory free-response feedback questions [59-61].

Analysis

We analyzed trust adjustment magnitude and trust adjustment using mixed-effects linear regression models. The models included categorical predictors and random intercepts to account for repeated measures within participants. Predictors included AI type and AI recognition pattern, forming a 3 × 4 design. We first examined differences between AI types (ex-ante advice and ex-post advice) and then further decomposed the ex-post advice into the involved ex-post condition and the not-involved ex-post condition for a 3-level comparison between interaction conditions. This decomposition was prespecified to examine the trust-dynamics associated with disagreement-triggered versus agreement-based AI interactions. Because assignment to the involved and not-involved conditions depends on pharmacists' initial decisions, these comparisons were response-contingent and are treated as exploratory analyses rather than confirmatory tests.

AI recognition patterns consisted of 4 levels: right fill-correct recognition, right fill-incorrect recognition, wrong fill-correct recognition, and wrong fill-incorrect recognition. Because the study focused on pharmacists' trust in AI-assisted decision-making, we restricted the analysis to AI-assisted conditions and excluded data from the no-AI-assistance control block.

We conducted regression 2-tailed *t* tests to compare mean trust outcomes across AI types within each AI recognition pattern. To account for multiple pairwise comparisons, *P* values for post hoc comparisons were adjusted using the Benjamini-Hochberg false discovery rate procedure. We conducted all analyses using mixed-effects linear regression models implemented in the lme4 package, applying the Kenward-Roger method to estimate degrees of freedom [62]. All analyses were conducted in the R statistical software (version 4.2.2) [63].

Results

Overview

Observations with trust-adjustment values exceeding ±3 SDs from the mean were excluded prior to analysis to reduce the influence of extreme outliers. To evaluate order effects, block order was included as a fixed-effect covariate. The likelihood ratio test showed that adding block order did not improve model fit for either outcome (trust adjustment magnitude: χ^2_9=8.88; *P*=.11; trust adjustment: χ^2_9=6.25; *P*=.28). Therefore, the results from the model excluding block order are presented.

Comparison Between AI Types

Trust Adjustment Magnitude Between Ex-Ante and Ex-Post Advice

AI type ($F_{1,9562.1}$=3.45, η^2=0.0004; *P*=.06) marginally affected the magnitude of trust adjustment. Ex-ante advice (estimated marginal mean [EMM] 6.13, 95% CI 4.28-7.97) led to marginally higher trust adjustment magnitude compared to ex-post advice (EMM 5.60, 95% CI 3.76-7.45; Table 2).

XSL•FO
RenderX

Table . Mean and SD of trust adjustment magnitude and trust change for different AI recognition patterns. AI recognition pattern to confusion matrix representation[a].

AI type and AI recognition pattern (fill accuracy - (human) AI)	Trust adjustment magnitude	Trust adjustment	Observations, n
Ex-ante advice	5.94 (15.23)		
Right-correct		2.14 (9.03)	2868
Wrong-correct		1.60 (9.90)	1088
Right-incorrect		−15.98 (26.98)	76
Wrong-incorrect		−12.78 (19.85)	800
Ex-post advice	5.37 (15.23)		
Right-correct		1.83 (8.47)	2841
Wrong-correct		1.68 (7.64)	1045
Right-incorrect		−17.84 (29.24)	90
Wrong-incorrect		−9.57 (16.71)	798
Involved Ex-Post	15.44 (26.36)		
Right-(incorrect)-correct		0.07 (13.74)	99
Wrong-(incorrect)-correct		2.75 (7.57)	95
Right-(correct)-incorrect		−18.09 (29.33)	68
Wrong-(correct)-incorrect		−8.96 (16.70)	769
Not-involved Ex-Post	2.60 (8.31)		
Right-(correct)-correct		1.90 (8.21)	2742
Wrong-(correct)-correct		1.57 (7.64)	950
Right-(incorrect)-incorrect		−11.14 (26.06)	22
Wrong-(incorrect)-incorrect		−11.45 (16.97)	29

[a]AI recognition pattern to confusion matrix representation. 472 right fill and correct recognition: correct rejection; wrong fill and correct recognition: correct detection; 473 right fill and incorrect recognition: false positive; wrong fill and incorrect recognition: false negative.

Trust Adjustment Magnitude Between Ex-Ante, Involved Ex-Post, and Not-Involved Ex-Post

When ex-post advice was further separated into involved and not-involved ex-post, the interaction condition ($F_{2,9565.1}$=380, η^2=0.07; P<.001) significantly affected the magnitude of trust adjustment.

A post hoc comparison between interaction conditions revealed that involved ex-post condition (EMM 15.78, 95% CI 13.62-17.94) led to significantly higher trust adjustment magnitude compared to ex-ante advice (EMM 6.13, 95% CI 4.15-8.11; $t_{9555.51}$=21.00, mean difference 9.65, 95% CI 8.55-10.75, Cohen d=0.72; P<.001) and not-involved ex-post condition (EMM 2.80, 95% CI 0.80-4.79; $t_{9556.27}$=27.55, mean difference 12.98, 95% CI 11.85-14.11, Cohen d=0.97; P<.001). Ex-ante advice led to a higher trust adjustment magnitude compared to not-involved ex-post condition ($t_{9554.52}$=11.46, mean difference 3.33, 95% CI 2.63-4.03, Cohen d=.25; P<.001; Table 2).

Comparison Between AI Types Within Each Recognition Pattern

Trust Adjustment Between Ex-Ante and Ex-Post Advice for Each Recognition Pattern

Recognition pattern ($F_{3,9555.9}$=811.37, η^2=0.2; P<.001) significantly affected trust adjustment; however, AI type did not influence trust adjustment ($F_{1,9548.6}$=2.47, η^2=0.0003; P=.12). However, a marginal interaction effect was observed ($F_{3,9550.7}$=2.52, η^2=0.0008; P=.06).

Because of the marginal interaction effect, we compare Ex-Ante and ex-post advice within each recognition pattern. Significant differences were observed when the right drugs were incorrectly rejected. Ex-post advice (EMM −17.84, 95% CI −19.06 to −16.62) showed larger trust decrement compared to ex-ante advice (EMM −15.97, 95% CI −17.19 to −14.76; $t_{9556.76}$=−2.67, mean difference −1.87, 95% CI −3.24 to −0.49, Cohen d=−0.13; P=.008). For other recognition patterns, no significant differences were observed.

XSL•FO
RenderX

Trust Adjustment Between Ex-Ante, Involved Ex-Post, and Not-Involved Ex-Post for Each Recognition Pattern

Both interaction condition ($F_{2,9558.7}$=3.89, η^2=0.0008; P=.03) and recognition pattern ($F_{3,9553.7}$=484.11, η^2=0.13; P<.001) significantly affected trust adjustment. In addition, a significant interaction effect was observed ($F_{6,9570.8}$=2.12, η^2=.0013; P=.048).

Because of the significant interaction effect, we compared interaction conditions within each recognition pattern. Significant differences were observed when the right drugs were incorrectly rejected. Involved ex-post condition (EMM −18.11, 95% CI −19.34 to −16.87) showed larger trust decrement compared to ex-ante advice (EMM −15.97, 95% CI −17.19 to −14.76; $t_{9553.27}$=−3.02, mean difference −2.13, 95% CI −3.83 to −0.44, Cohen d=−0.15; P=.008) and not-involved ex-post condition (EMM −10.79, 95% CI −15.96 to −5.61; $t_{9577.26}$=−2.75, mean difference −7.32, 95% CI −13.69 to −0.95, Cohen d=−0.52; P=.009). A marginal difference was observed between ex-ante advice and not-involved ex-post condition ($t_{9569.61}$=−1.94, mean difference −5.19, 95% CI −11.55 to 1.17, Cohen d=−0.37; P=.052). For other recognition patterns, no significant differences were observed.

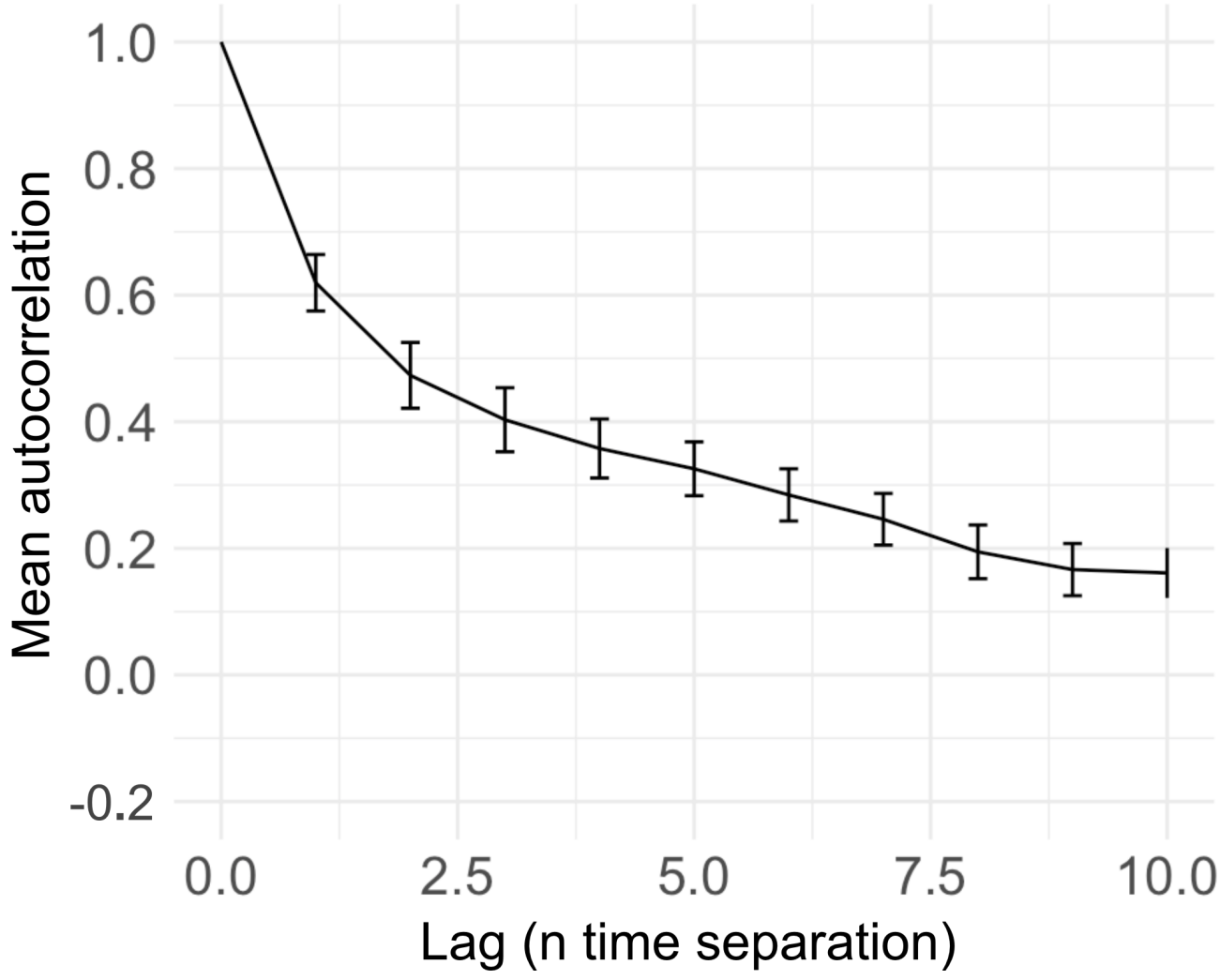
Evaluating Trust Properties: Continuity, Negativity Bias, and Stabilization

To explore the key characteristics of trust dynamics, the recognition patterns were grouped into 2 categories: correct AI predictions (right drug correctly approved and *wrong* drug correctly rejected) and incorrect AI predictions (right drug *incorrectly* rejected and *wrong* drug *incorrectly* accepted).

Continuity

Comparing instances of correct versus incorrect AI predictions revealed that successful AI predictions generally increased user trust, while failures led to decreases. In addition, trust levels at a given moment (t) were strongly correlated with trust from the preceding moment (t − 1). To assess how current trust ratings relate to past ones, we calculated the autocorrelation of trust ratings. As shown in Figure 4, trust at time t closely aligns with trust at time t − 1. However, as the time interval between ratings grew longer, the average autocorrelation declined, indicating that trust correlation weakens over time (Figure 4).

Figure 4. Autocorrelation of trust as a function of time separation. The error bars represent ±2 SEs.



XSL•FO
RenderX

Negativity Bias

To test for negativity bias, we compared the magnitude of trust increases following correct predictions with the magnitude of trust decreases following incorrect ones. The results indicated a significant difference between the two: correct AI predictions (mean 2.98, SD 7.33) produced smaller changes than incorrect ones (mean 15.80, SD 23.60) (t_{4726}=30.59, mean difference −12.7, 95% CI −13.6 to −11.9, standardized effect size=−1.15; P<.001).

Stabilization

To examine the stabilization property, linear mixed models were used to analyze trust adjustments over time (sequential success [or failure] occurrence).

The findings revealed that the magnitude of trust adjustments decreased significantly across occurrences for both correct predictions (β=−.037, 95% CI −0.055 *to* −0.019, $t_{4021.10}$=−4.10; P<.001) and incorrect predictions (β=−.402, 95% CI −0.659 to −0.155, $t_{827.46}$=−3.20*;* P=.001).

Discussion

Principal Findings

This study investigated how the timing of AI assistance shaped pharmacists' trust during AI-assisted medication dispensing. We implemented two AI advice timing conditions. In the ex-ante advice condition, the AI presented recognition predictions before pharmacists' initial decision. In the ex-post advice condition, the AI provided selective, postdecision intervention. The ex-post advice was further separated into the involved ex-post condition, which intervened when disagreed with the pharmacist, and the not-involved ex-post condition, which withheld intervention when in agreement.

Overall, differences in trust adjustment magnitude across AI types were modest, with ex-ante advice eliciting slightly higher trust adjustment magnitude than ex-post advice. However, when AI generated incorrect recommendations on initially right fills, ex-ante advice resulted in significantly smaller trust decrement than ex-post advice. Further differentiation of ex-post advice revealed an ordered pattern, where involved ex-post that actively disagreed in the decision process prompted the greatest trust adjustment magnitude, followed by ex-ante advice, with not-involved ex-post yielding the smallest magnitude. This ordering was consistent across the right-incorrect recognition pattern.

In addition, pharmacists' trust dynamics demonstrated ecological validity of trust dynamics properties within a realistic verification context. The findings indicate that pharmacists' trust in AI is dynamic and influenced by the timing, level of involvement, and situational performance of AI assistance during medication verification.

Interpretation of Findings

Comparison Between AI Types

Comparison between the ex-ante and ex-post advice revealed a marginally higher magnitude for ex-ante advice compared to the ex-post advice, which aligns with the findings of Yin et al [27] that the ex-post advice resulted in superior diagnostic quality. When we further decomposed the ex-post advice paradigm into involved ex-post and not-involved ex-post to examine more nuanced trust dynamics associated with AI agreement and disagreement, the involved ex-post condition yielded the highest magnitude of trust adjustment, followed by the ex-ante advice, and then the not-involved ex-post condition. This pattern suggests that interruptive, disagreement-based interventions may exert a strong influence on trust. The ex-ante advice also required pharmacists to evaluate the AI's accuracy during verification, leading to greater trust fluctuations than in the not-involved ex-post condition, which preserved workflow continuity and imposed no extra attentional demand.

Within Recognition Pattern Comparisons

We further analyzed trust adjustment behavior across AI recognition patterns, which considered both the correctness of the filled medication (*right fill* or *wrong fill*) and the accuracy of the AI's prediction (*correct* or *incorrect*). Significant differences in trust adjustment across ex-ante and ex-post advice occurred when *the right fill was incorrectly rejected*. In this pattern, the ex-post advice led to a larger trust decrement compared to the ex-ante advice. For *the right fill was incorrectly rejected* pattern, significant differences in trust adjustment also occurred across interaction conditions. The involved ex-post condition led to the largest trust decrement compared to the other interaction conditions. The ex-ante advice showed a marginally greater trust decrement compared to the not-involved ex-post condition.

Understanding this pattern requires examining how each interaction condition responded in scenarios where AI incorrectly rejected the right drug. In the ex-ante advice condition, the AI's incorrect prediction was visible from the start. Mismatch indicators (ie, red checkboxes) may have falsely anchored pharmacists' attention toward an assumed mismatch and led to a moderate trust decrement [30,32]. In the involved ex-post condition, after pharmacists made the correct initial decision to accept the right drug, AI erroneously classified the mismatch and intervened with a false alert. This prompted the pharmacist to reconsider a correct initial decision, and the largest trust decrement was observed. This mirrors findings that incorrect ex-post alerts elicit strong trust decrements [42]. In contrast, the not-involved ex-post condition functioned as a passive agreement system. Under this condition, the AI silently agreed with the pharmacists' incorrect initial rejection of the right drug. Here, the error was shared between pharmacists and the AI, and the smallest trust decrement was observed.

Operationally, an incorrect rejection of the right drug represents an unnecessary interruption in the dispensing process. Following the AI's incorrect suggestion would require the pharmacist to reverify or refill a medication that was ready to be dispensed, introducing inefficiency. The larger trust decrement observed in this scenario likely does not stem from task consequence. When pharmacists' accurate initial decisions were falsely contradicted by the involved ex-post condition, the error was salient, yielding strong trust decrements. Conversely, when both

XSL•FO
RenderX

pharmacists and AI were wrong (not-involved ex-post), error attribution was shared, producing the smallest trust adjustment.

Validating Trust Dynamics Properties

Beyond evaluating AI interaction types, this study validated 3 trust dynamics properties—continuity, negativity bias, and stabilization [8,45,64]—with an ecologically realistic pharmacy verification environment. Prior validations have often relied on simplified tasks and nonexpert populations. By involving licensed pharmacists performing authentic pill verification, this study strengthens the ecological validity of trust-dynamics in human-AI interaction.

With licensed pharmacists, we demonstrated that trust evolves in a temporally dependent manner, such that trust at a given moment is shaped by prior interactions, strengthening following correct AI performance and declining after errors. We also observed negativity bias, reflecting the asymmetric influence of system performance, whereby trust is more sensitive to AI failures than to successes, making it difficult to build but easy to lose. Finally, pharmacists' trust exhibited stabilization, showing a tendency to converge over repeated interactions as users developed more consistent expectations of the system's behavior.

Implications

These findings underscore the need for careful design of conditional intervention systems in safety-critical workflows. Ex-post advice assistance, implemented as a disagreement-triggered intervention, carries risks of confirmation bias, low revision rates, and negative reactions to incorrect alerts [40,43]. For pharmacists, who are expert decision makers and typically accurate, false AI interventions impose unnecessary cognitive costs. Frequent exposure to erroneous interventions may lead to over-skepticism toward the AI, undermining the intended safety benefit. Conversely, ex-ante advice avoids disruptive alerts but introduces anchoring and information-overload risks of Ex-Ante paradigms [30,31]. Both modalities present distinct trade-offs between autonomy support, cognitive load, and error-correction potential.

Designers of medication verification AI must consider these dynamics when determining when and how AI assistance should be delivered. For example, hybrid approaches, such as confidence-gated alerts, adaptive interventions, or selective suppression of low-confidence interruptions, may mitigate the negative impacts observed with the involved ex-post condition.

Limitations and Future Directions

We acknowledge several limitations of this study and propose directions for future research. First, the AI system's accuracy was experimentally reduced from its operational performance (98.5%) to 79% to ensure sufficient exposure to both correct and incorrect recommendations within a feasible number of trials. Although this approach enabled systematic examination of trust adjustments, it may limit generalizability to real-world settings where AI errors may be less frequent. In high-accuracy environments, errors may be more salient and unexpected, potentially eliciting different trust dynamics. Future research should investigate how trust evolves under near-realistic levels of AI reliability.

Second, in the involved ex-post condition, AI disagreement and workflow interruption were inherently coupled. Thus, the observed aggregated trust adjustment magnitude and trust decrement may reflect not only disagreement with the AI recommendation itself but also the cognitive and attentional disruption introduced by mid-task interventions. Future work should investigate the independent effects of disagreement content and workflow interruption on trust adjustment.

Third, the present analyses focused on self-reported trust adjustment, captured through trial-by-trial trust ratings, rather than subsequent behavioral adaptation to the AI support. As a result, the study does not explicitly highlight whether the trust pattern predicts behavioral responses, such as overriding AI recommendations, changes in decision strategies, or adjustments in reaction time. The asymmetric trust pattern observed in the study (negativity bias) may also relate to the broader literature on algorithm aversion, which suggests humans might respond differently to algorithmic mistakes, to an extent that they may disregard using systems even if they outperform humans [40,65]. Future work should investigate how moment-to-moment trust dynamics influence the subsequent decision-making behavior during repeated interactions with AI systems.

Conclusions

This study examined pharmacists' trust dynamics in response to variations in AI assistance timing and conditionality during the medication verification task. By comparing ex-ante advice and ex-post advice paradigms, and further decomposing ex-post advice into involved and not-involved conditions, the findings demonstrate that both timing and conditionality influence trust. In particular, ex-ante advice resulted in greater overall trust adjustment magnitude than in ex-post advice. However, when further decomposed, disagreement-based interventions that incorrectly challenge right expert decisions produced the largest trust decrements and greatest trust fluctuations.

These results extend prior work on AI advice timing by showing that the effectiveness of ex-post advice depends on the nature of AI involvement. While ex-post advice has been shown to enhance overall decision quality, the present findings indicate that the degree to which the AI agrees or disagrees with users' initial decision plays a critical role in trust adjustments beyond the AI advice timing alone. For expert users performing safety-critical tasks, such as pharmacists performing medication verification, erroneous AI interruptions may influence trust more negatively than continuous AI support. This highlights the need to align AI intervention strategies with user expertise and task demands.

Finally, this work also validates the view that trust in AI is a dynamic, interaction-driven construct that evolves through repeated encounters with system behavior. By validating trust dynamics properties in an ecologically realistic pharmacy verification task, this study contributes to human-AI interaction research that emphasizes the importance of capturing moment-to-moment trust adjustments.

XSL•FO
RenderX

Acknowledgments

During the preparation of this work, the authors used ChatGPT (OpenAI, 2026) to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

This study is supported by the National Institutes of Health's National Library of Medicine grant R01LM013624.

Data Availability

The datasets generated during this study are available on the ClinicalTrials.gov website under the identifier NCT06245044.

Authors' Contributions

JYK handled data curation, investigation, formal analysis, validation, visualization, writing – original draft, and writing – review and editing. BR handled investigation, project administration, and writing – review and editing. MW conducted investigation and writing – review and editing. QC supervised software, validation, and writing – review and editing. RAK contributed to conceptualization, funding acquisition, methodology, and supervision. CL contributed to conceptualization, funding acquisition, methodology, project administration, supervision, validation, and writing – review and editing. XJY managed conceptualization, funding acquisition, methodology, supervision, validation, writing – original draft, and writing – review and editing.

Conflicts of Interest

None declared.

References

1. Karia A, Norman R, Robinson S, et al. Pharmacist's time spent: Space for Pharmacy-based Interventions and Consultation TimE (SPICE)—an observational time and motion study. BMJ Open 2022 Mar 2;12(3):e055597. [doi: 10.1136/bmjopen-2021-055597] [Medline: 35236731]
2. Szeinbach S, Seoane-Vazquez E, Parekh A, Herderick M. Dispensing errors in community pharmacy: perceived influence of sociotechnical factors. Int J Qual Health Care 2007 Aug;19(4):203-209. [doi: 10.1093/intqhc/mzm018] [Medline: 17567597]
3. Um IS, Clough A, Tan ECK. Dispensing error rates in pharmacy: a systematic review and meta-analysis. Res Social Adm Pharm 2024 Jan;20(1):1-9. [doi: 10.1016/j.sapharm.2023.10.003] [Medline: 37848350]
4. Campbell PJ, Patel M, Martin JR, et al. Systematic review and meta-analysis of community pharmacy error rates in the USA: 1993–2015. BMJ Open Qual 2018;7(4):e000193. [doi: 10.1136/bmjoq-2017-000193] [Medline: 30306141]
5. Reiner G, Pierce SL, Flynn J. Wrong drug and wrong dose dispensing errors identified in pharmacist professional liability claims. J Am Pharm Assoc (2003) 2020;60(5):e50-e56. [doi: 10.1016/j.japh.2020.02.027] [Medline: 32217085]
6. Chui MA, Look KA, Mott DA. The association of subjective workload dimensions on quality of care and pharmacist quality of work life. Res Social Adm Pharm 2014;10(2):328-340. [doi: 10.1016/j.sapharm.2013.05.007] [Medline: 23791360]
7. Chui MA, Mott DA. Community pharmacists' subjective workload and perceived task performance: a human factors approach. J Am Pharm Assoc (2003) 2012;52(6):e153-e160. [doi: 10.1331/JAPhA.2012.11135] [Medline: 23229977]
8. Kim JY, Lester C, Yang XJ. Beyond binary decisions: evaluating the effects of AI error type on trust and performance in AI-assisted tasks. Hum Factors 2025 Oct;67(10):1062-1083. [doi: 10.1177/00187208251326795] [Medline: 40104968]
9. Kim JY, Marshall VD, Rowell B, et al. The effects of presenting AI uncertainty information on pharmacists' trust in automated pill recognition technology: exploratory mixed subjects study. JMIR Hum Factors 2025 Feb 11;12:e60273. [doi: 10.2196/60273] [Medline: 39932773]
10. Larios Delgado N, Usuyama N, Hall AK, et al. Fast and accurate medication identification. NPJ Digit Med 2019;2(1):10. [doi: 10.1038/s41746-019-0086-0] [Medline: 31304359]
11. Zheng Y, Rowell B, Chen Q, et al. Designing human-centered AI to prevent medication dispensing errors: focus group study with pharmacists. JMIR Form Res 2023 Dec 25;7(1):e51921. [doi: 10.2196/51921] [Medline: 38145475]
12. Tsai CC, Kim JY, Chen Q, et al. Effect of artificial intelligence helpfulness and uncertainty on cognitive interactions with pharmacists: randomized controlled trial. J Med Internet Res 2025 Jan 31;27:e59946. [doi: 10.2196/59946] [Medline: 39888668]
13. Lester C, Rowell B, Zheng Y, et al. Effect of uncertainty-aware AI models on pharmacists' reaction time and decision-making in a web-based mock medication verification task: randomized controlled trial. JMIR Med Inform 2025 Apr 18;13(1):e64902. [doi: 10.2196/64902] [Medline: 40249341]
14. Heo J, Kang Y, Lee S, Jeong DH, Kim KM. An accurate deep learning-based system for automatic pill identification: model development and validation. J Med Internet Res 2023 Jan 13;25:e41043. [doi: 10.2196/41043] [Medline: 36637893]
15. Palenychka R, Lakhssassi A, Palenychka M. Verification of medication dispensing using the attentive computer vision approach. Proc IEEE Int Symp Circuits Syst 2018:1-4. [doi: 10.1109/ISCAS.2018.8351202]

XSL•FO
RenderX

16. Chen Q, Al Kontar R, Nouiehed M, Yang XJ, Lester C. Rethinking cost-sensitive classification in deep learning via adversarial data augmentation. INFORMS J Data Sci 2025;4(1):1-19. [doi: 10.1287/ijds.2022.0033] [Medline: 42610103]
17. Gombolay M, Yang XJ, Hayes B, et al. Robotic assistance in the coordination of patient care. Int J Rob Res 2018;37(10):1300-1316. [doi: 10.1177/0278364918778344]
18. Wang EH, Gross CP, Tilburt JC, et al. Shared decision making and use of decision AIDS for localized prostate cancer: perceptions from radiation oncologists and urologists. JAMA Intern Med 2015 May;175(5):792-799. [doi: 10.1001/jamainternmed.2015.63] [Medline: 25751604]
19. Steyvers M, Kumar A. Three challenges for AI-assisted decision-making. Perspect Psychol Sci 2024;19(5):722-734. [doi: 10.1177/17456916231181102] [Medline: 37439761]
20. Zhang Q, Yang XJ, Robert LP. What and when to explain? A survey of the impact of explanation on attitudes toward adopting automated vehicles. IEEE Access 2021;9:159533-159540. [doi: 10.1109/ACCESS.2021.3130489]
21. Lee JD, See KA. Trust in automation: designing for appropriate reliance. Hum Factors 2004;46(1):50-80. [doi: 10.1518/hfes.46.1.50_30392] [Medline: 15151155]
22. von Eschenbach WJ. Transparency and the black box problem: why we do not trust AI. Philos Technol 2021;34(4):1607-1622. [doi: 10.1007/s13347-021-00477-0]
23. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019 Oct 29;17(1):195. [doi: 10.1186/s12916-019-1426-2] [Medline: 31665002]
24. Cai CJ, Winter S, Steiner D, Wilcox L, Terry M. "Hello AI": uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. Proc ACM Hum-Comput Interact 2019 Nov 7;3(CSCW):1-24. [doi: 10.1145/3359206]
25. Begoli E, Bhattacharya T, Kusnezov D. The need for uncertainty quantification in machine-assisted medical decision making. Nat Mach Intell 2019;1:20-23. [doi: 10.1038/s42256-018-0004-1]
26. Kompa B, Snoek J, Beam AL. Second opinion needed: communicating uncertainty in medical machine learning. NPJ Digit Med 2021 Jan 5;4(1):4. [doi: 10.1038/s41746-020-00367-3] [Medline: 33402680]
27. Yin J, Ngiam KY, Tan SSL, Teo HH. Designing AI-based work processes: how the timing of AI advice affects diagnostic decision making. Manage Sci 2025 Nov;71(11):9361-9383. [doi: 10.1287/mnsc.2022.01454]
28. Brennan M, Puri S, Ozrazgat-Baslanti T, et al. Comparing clinical judgment with the MySurgeryRisk algorithm for preoperative risk assessment: a pilot usability study. Surgery 2019 May;165(5):1035-1045. [doi: 10.1016/j.surg.2019.01.002] [Medline: 30792011]
29. Tschandl P, Rinner C, Apalla Z, et al. Human-computer collaboration for skin cancer recognition. Nat Med 2020;26(8):1229-1234. [doi: 10.1038/s41591-020-0942-0] [Medline: 32572267]
30. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. Science 1974 Sep 27;185(4157):1124-1131. [doi: 10.1126/science.185.4157.1124] [Medline: 17835457]
31. Buçinca Z, Malaya MB, Gajos KZ. To trust or to think. Proc ACM Hum-Comput Interact 2021 Apr 13;5(CSCW1):1-21. [doi: 10.1145/3449287]
32. Rastogi C, Zhang Y, Wei D, Varshney KR, Dhurandhar A, Tomsett R. Deciding fast and slow: the role of cognitive biases in AI-assisted decision-making. Proc ACM Hum-Comput Interact 2022 Mar 30;6(CSCW1):1-22. [doi: 10.1145/3512930]
33. Aoki T, Yamada A, Aoyama K, et al. Clinical usefulness of a deep learning-based system as the first screening on small-bowel capsule endoscopy reading. Dig Endosc 2020 May;32(4):585-591. [doi: 10.1111/den.13517] [Medline: 31441972]
34. Dratsch T, Chen X, Rezazade Mehrizi M, et al. Automation bias in mammography: the impact of artificial intelligence BI-RADS suggestions on reader performance. Radiology 2023 May;307(4):e222176. [doi: 10.1148/radiol.222176] [Medline: 37129490]
35. Gaube S, Suresh H, Raue M, et al. Do as AI say: susceptibility in deployment of clinical decision-aids. NPJ Digit Med 2021 Feb 19;4(1):31. [doi: 10.1038/s41746-021-00385-9] [Medline: 33608629]
36. Green B, Chen Y. The principles and limits of algorithm-in-the-loop decision making. Proc ACM Hum-Comput Interact 2019 Nov 7;3(CSCW):1-24. [doi: 10.1145/3359152]
37. Yaniv I, Choshen‐Hillel S. Exploiting the wisdom of others to make better decisions: suspending judgment reduces egocentrism and increases accuracy. Behav Decis Mak 2012 Dec;25(5):427-434. [doi: 10.1002/bdm.740]
38. Klayman J. Varieties of confirmation bias. Psychol Learn Motiv 1995;32:385-418. [doi: 10.1016/S0079-7421(08)60315-1]
39. Nickerson RS. Confirmation bias: a ubiquitous phenomenon in many guises. Rev Gen Psychol 1998 Jun;2(2):175-220. [doi: 10.1037/1089-2680.2.2.175]
40. Dietvorst BJ, Simmons JP, Massey C. Algorithm aversion: people erroneously avoid algorithms after seeing them err. J Exp Psychol Gen 2015 Feb;144(1):114-126. [doi: 10.1037/xge0000033] [Medline: 25401381]
41. Wang D, Yang Q, Abdul A, Lim BY. Designing theory-driven user-centric explainable AI. Proc 2019 CHI Conf Hum Factors Comput Syst 2019:1-15. [doi: 10.1145/3290605.3300831]
42. Lehman CD, Wellman RD, Buist DSM, et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. JAMA Intern Med 2015 Nov;175(11):1828-1837. [doi: 10.1001/jamainternmed.2015.5231] [Medline: 26414882]

XSL•FO
RenderX

43. Fogliato R, Chappidi S, Lungren M, et al. Who goes first? Influences of human-AI workflow on decision making in clinical imaging. Proc ACM Conf Fairness Accountab Transpar 2022 Jun 21:1362-1374. [doi: 10.1145/3531146.3533193]
44. de Visser EJ, Peeters MMM, Jung MF, et al. Towards a theory of longitudinal trust calibration in human–robot teams. Int J of Soc Robotics 2020;12(2):459-478. [doi: 10.1007/s12369-019-00596-x]
45. Yang XJ, Guo Y, Schemanske C. From trust to trust dynamics: combining empirical and computational approaches to model and predict trust dynamics. In: Human-Automation Interaction: Transportation: Springer; 2022:253-265. [doi: 10.1007/978-3-031-10784-9_15]
46. Lee J, Moray N. Trust, control strategies and allocation of function in human-machine systems. Ergonomics 1992 Oct;35(10):1243-1270. [doi: 10.1080/00140139208967392] [Medline: 1516577]
47. Wischnewski M, Krämer N, Müller E. Measuring and understanding trust calibrations for automated systems: a survey of the state-of-the-art and future directions. Proc 2023 CHI Conf Hum Factors Comput Syst 2023:1-16. [doi: 10.1145/3544548.358119]
48. Yang XJ, Schemanske C, Searle C. Toward quantifying trust dynamics: how people adjust their trust after moment-to-moment interaction with automation. Hum Factors 2023 Aug;65(5):862-878. [doi: 10.1177/00187208211034716] [Medline: 34459266]
49. Manzey D, Reichenbach J, Onnasch L. Human performance consequences of automated decision aids: the impact of degree of automation and system experience. J Cogn Eng Decis Mak 2012;6(1):57-87. [doi: 10.1177/1555343411433844]
50. Lee JD, Moray N. Trust, self-confidence, and operators' adaptation to automation. Int J Hum Comput Stud 1994 Jan;40(1):153-184. [doi: 10.1006/ijhc.1994.1007]
51. Yang XJ, Wickens CD, Hölttä-Otto K. How users adjust trust in automation: contrast effect and hindsight bias. Proc Hum Factors Ergon Soc Annu Meet 2016;60:196-200. [doi: 10.1177/1541931213601044]
52. Yang XJ, Unhelkar VV, Li K, Shah JA. Evaluating effects of user experience and system transparency on trust in automation. Proc ACM/IEEE Int Conf Hum Robot Interact 2017 Mar 6:408-416. [doi: 10.1145/2909824.3020230]
53. Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. Presented at: Proceedings of the 33rd International Conference on Machine Learning; Jun 19-24, 2016; New York City, NY p. 1050-1059 URL: https://proceedings.mlr.press/v48/gal16.pdf [accessed 2026-08-27]
54. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. IEEE Conf Comput Vis Pattern Recognit 2016:770-778. [doi: 10.1109/CVPR.2016.90]
55. Lester CA, Li J, Ding Y, Rowell B, Kontar RA. Performance evaluation of a prescription medication image classification model: an observational cohort. NPJ Digit Med 2021 Jul 27;4(1):118. [doi: 10.1038/s41746-021-00483-8] [Medline: 34315995]
56. Wickens CD, Dixon SR. The benefits of imperfect diagnostic automation: a synthesis of the literature. Theor Issues Ergon Sci 2007 May;8(3):201-212. [doi: 10.1080/14639220500370105]
57. Merritt SM, Heimbaugh H, LaChapell J, Lee D. I trust it, but I don't know why: effects of implicit attitudes toward automation on trust in an automated system. Hum Factors 2013 Jun;55(3):520-534. [doi: 10.1177/0018720812465081] [Medline: 23829027]
58. Winwood PC, Winefield AH, Dawson D, Lushington K. Development and validation of a scale to measure work-related fatigue and recovery: the Occupational Fatigue Exhaustion/Recovery Scale (OFER). J Occup Environ Med 2005 Jun;47(6):594-606. [doi: 10.1097/01.jom.0000161740.71049.c4] [Medline: 15951720]
59. Muir BM, Moray N. Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. Ergonomics 1996 Mar;39(3):429-460. [doi: 10.1080/00140139608964474] [Medline: 8849495]
60. Hart SG. NASA-Task Load Index (NASA-TLX); 20 Years Later. Proc Hum Factors Ergon Soc Annu Meet 2006 Oct;50(9):904-908. [doi: 10.1177/154193120605000909]
61. Brooke J. SUS—a quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation in Industry: CRC Press; 1996:189-194. [doi: 10.1201/9781498710411]
62. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67:1-48. [doi: 10.18637/jss.v067.i01]
63. The R Project for Statistical Computing. 2023. URL: https://www.R-project.org/ [accessed 2026-08-27]
64. Guo Y, Yang XJ. Modeling and predicting trust dynamics in human–robot teaming: a Bayesian inference approach. Int J Soc Robotics 2021 Dec;13(8):1899-1909. [doi: 10.1007/s12369-020-00703-3]
65. Bogert E, Schecter A, Watson RT. Humans rely more on algorithms than social influence as a task becomes more difficult. Sci Rep 2021 Apr 13;11(1):8028. [doi: 10.1038/s41598-021-87480-9] [Medline: 33850211]

Abbreviations

BNN: Bayesian neural network
CDSS: clinical decision support system
EMM: estimated marginal mean
HCI: human-computer interaction

XSL•FO
RenderX

NASA-TLX: NASA Task Load Index
NDC: National Drug Code
OFER: Occupational Fatigue Exhaustion/Recovery Scale

Edited by J Waterson, S Law; submitted 03.Feb.2026; peer-reviewed by AR Ashraf, N Rudolf; revised version received 21.Jul.2026; accepted 06.Aug.2026; published 11.Sep.2026.

Please cite as:
Kim JY, Rowell B, Whitaker M, Chen Q, Al Kontar R, Lester C, Yang XJ
Effects of AI Assistance Timing on Pharmacists' Trust in Automated Pill Recognition Technology: Within-Participants Experimental Study
JMIR Hum Factors 2026;13:e92822
URL: https://humanfactors.jmir.org/2026/1/e92822
doi:10.2196/92822

© Jin Yong Kim, Brigid Rowell, Megan Whitaker, Qiyuan Chen, Raed Al Kontar, Corey Lester, Xi Jessie Yang. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 11.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Clinician Trust and Human Factors in AI-Enabled Clinical Decision Support in Acute Care: Mixed Methods Study

Meghana Darla[1,2], BDS, MS; Danielle Miltz[3], PA-C; Khushboo Chandnani[2], MS; Saptarshi Purkayastha[2], PhD; John W Diehl[4], MD; Sivasubramanium V Bhavani[5], MS, MD

[1]Department of Otolaryngology, School of Medicine, Emory University, Emory University Hospital Midtown, 550 Peachtree St NE, Suite 1135, Atlanta, GA, United States

[2]Department of Biomedical Engineering and Informatics, Luddy School of Informatics, Computing, and Engineering, Indiana University, Indianapolis, IN, United States

[3]Surgical Transplant Intensive Care Unit, Emory Critical Care Center, Emory Healthcare, Atlanta, GA, United States

[4]Department of Emergency Medicine, School of Medicine, Emory University, Atlanta, GA, United States

[5]Department of Medicine, School of Medicine, Emory University, Atlanta, GA, United States

Corresponding Author:
Meghana Darla, BDS, MS
Department of Otolaryngology, School of Medicine, Emory University, Emory University Hospital Midtown, 550 Peachtree St NE, Suite 1135, Atlanta, GA, United States

Abstract

Background: AI has the potential to enhance clinical decision-making in high-acuity settings such as intensive care units (ICUs) and emergency departments (EDs). However, despite promising performance, many AI-driven clinical decision support systems (AI-CDSSs) face poor adoption due to issues of trust, workflow disruption, and alert fatigue. Understanding the human factors that shape clinician acceptance is critical to guide safe and effective implementation of AI-CDSS in acute care. Theoretical frameworks, including the Systems Engineering Initiative for Patient Safety (SEIPS) 2.0 model and the technology acceptance model (TAM), suggest that successful adoption requires addressing sociotechnical interactions among clinician trust, system design, organizational readiness, and task complexity, yet few empirical studies have applied these frameworks to AI-CDSSs in acute care settings.

Objective: This study aimed to evaluate emergency medicine and critical care clinicians' perceptions of AI-CDSSs and to identify key factors influencing adoption, including trust, design preferences, and workflow integration.

Methods: A SEIPS 2.0–informed mixed methods study evaluated ICU and ED clinicians from Emory Healthcare on perceptions of AI in clinical practice. An expert-reviewed survey (N=57) assessed clinician perceptions, trust, and implementation preferences. Semistructured interviews (n=11) included A/B testing of AI-CDSSs and clinical sepsis scenarios to explore decision-making in context. Transcripts were thematically analyzed using the Braun and Clarke framework in ATLAS.ti (version 26, ATLAS.ti Scientific Software Development). Quantitative data were analyzed descriptively. This study assessed clinician perceptions using mock alerts and hypothetical scenarios rather than real-world AI-CDSS deployment.

Results: Trust in AI varied significantly by patient acuity (Cochran Q=30.40, *P*<.001): stable patients (43/57, 75%, 95% CI 63% - 85%), deteriorating patients (27/57, 47%, 95% CI 35% - 60%), and critically ill patients in the ICU and undifferentiated patients in the ED (25/57, 44%, 95% CI 32% - 57% for each scenario). Internal consistency was acceptable-to-good across three scales (Cronbach α: AI Perception=.891, Trust=.743, Implementation=.740; McDonald ω: AI Perception=0.895, Trust=0.782, Implementation=0.746). Barriers included overreliance, insufficient training, and data quality concerns. For the exploratory AI-CDSS design, clinicians preferred opt-in alerts (10/11, 91%), evidence-linked recommendations (7/11, 64%), and avoiding overt mention of increased AI acceptance (8/11, 73%). Thematic analysis yielded 36 themes across six domains: trust and transparency, alert usability, workflow fit, data concerns, training needs, and perceived clinical impact. Clinicians favored AI-CDSSs that preserved autonomy, minimized disruption, and provided transparent rationales.

Conclusions: Adoption of AI-CDSSs in critical care is not solely a technical issue but a human-factors challenge centered on trust, transparency, and workflow compatibility. These findings support future testing of a phased implementation approach—beginning with lower-acuity applications where clinician trust is highest, then gradually extending to higher-acuity scenarios with enhanced transparency and override mechanisms. This graduated strategy addresses the critical interdependencies among people (trust), tools (design), organizations (training), and tasks (clinical complexity) identified in this study.

(JMIR Hum Factors 2026;13:e95472) doi:10.2196/95472

XSL•FO
RenderX

KEYWORDS

clinical decision support systems; artificial intelligence; critical care; human factors; clinician surveys; trust in automation; technology acceptance model; SEIPS model; Systems Engineering Initiative for Patient Safety

Introduction

The integration of AI tools into clinical practice offers an opportunity to enhance health care delivery, particularly in high-acuity settings like the intensive care unit (ICU) and emergency department (ED), where rapid, data-driven decisions are crucial [1]. Despite the technological promise, the successful implementation of AI tools into frontline clinical practice is filled with challenges [2]. The history of health information technology is replete with examples of powerful tools that failed to gain traction due to a mismatch with clinical workflow. This mismatch is often attributed to a lack of attention to "human factors," which refer to the complex interplay between technology, users, and the clinical environment [3]. The adoption of AI is similarly challenged by significant barriers, including a lack of trust, concerns about workflow disruption, and the potential for alert fatigue [3-5]. Overcoming these barriers requires a deep understanding of clinician perceptions, needs, and priorities.

A clinical decision support system (CDSS) is a practical application that can integrate AI into clinical decision-making [4]. A targeted intervention can be suggested to the clinician through a CDSS from a validated machine learning algorithm. For instance, in the Precision Resuscitation With Crystalloids in Sepsis (PRECISE) trial, a validated machine learning algorithm identifies a subgroup of infected patients with a potential mortality benefit from balanced crystalloid instead of normal saline, and a CDSS prompts the clinician to change the order [6]. Since the integration of CDSSs into electronic health records (EHRs), there have been multiple studies on the "five rights" of CDSSs, which help guide the creation of effective alerts [7]; however, less is understood on how to effectively incorporate a CDSS that utilizes an AI recommendation. Because successful AI-based CDSS (AI-CDSS) implementation depends on more than the technical performance of the algorithm, a human factors framework is needed to examine how clinician trust, workflow, alert design, and organizational context shape adoption [8,9].

Using multiple complementary frameworks to evaluate implementation of an AI tool can provide a more holistic understanding. We primarily draw on the Systems Engineering Initiative for Patient Safety (SEIPS) 2.0 model [10], which conceptualizes health care systems as comprising five interacting components: people, tasks, tools and technologies, organization, and environment. SEIPS 2.0 emphasizes that patient safety and quality outcomes emerge from complex, bidirectional interactions among these components, making it particularly suitable for studying the adoption of AI-CDSSs, which simultaneously affects clinician cognition (people), decision-making workflows (tasks), interface design requirements (tools), implementation strategy (organization), and care setting constraints (environment). We also incorporate constructs from the technology acceptance model (TAM) [11], specifically Perceived Usefulness and Perceived Ease of Use, and Jian et al's [12] Trust in Automated Systems framework, which characterizes trust across belief, intention, and behavior dimensions. In addition, we draw on the three-layer trust model of Hoff and Bashir [13], which distinguishes dispositional, situational, and learned trust and offers an integrative lens for trust in automation. This multiframework approach responds to calls for theoretically grounded, human-centered evaluations of health care AI [8,9].

In this study, we address the gap in knowledge on applying AI-CDSSs to high acuity settings through a mixed methods evaluation to investigate ED and ICU clinicians' perceptions of AI use in clinical practice and to identify and analyze the facilitators and barriers to AI adoption.

Methods

Study Design

A mixed methods human factors evaluation was conducted at Emory Healthcare in Atlanta, Georgia. The study employed a cross-sectional, exploratory design and combined a quantitative survey with qualitative, semistructured interviews. We used a convergent (concurrent triangulation) design in which the quantitative and qualitative strands were collected in parallel and given equal priority (QUAN+QUAL). Integration occurred at the methods level, through shared constructs across the survey and interview guide, and at the interpretation level, through a joint display. The mixed methods reporting follows the Good Reporting of a Mixed Methods Study (GRAMMS) criteria [14] (Checklist 1). SEIPS 2.0 also guided the study operationally across data collection, analysis, and interpretation. Survey domains and the semistructured interview guide were developed to reflect SEIPS 2.0 work-system components, including people, tasks, tools and technologies, organization, and environment. These domains were later used as sensitizing concepts during codebook refinement, theme development, and interpretation. The objective was to explore clinician perceptions of AI-CDSSs.

Recruitment and Participants

Eligible participants included clinicians in the ICU and ED, who were actively involved in direct patient care. Recruitment was conducted from February to April 2025 through multiple channels: invitation emails distributed via ICU and ED provider listservs at all Emory Healthcare hospitals, institutional review board (IRB)–approved flyers posted in clinical units, presentations during departmental meetings, and word-of-mouth referrals among clinical staff. Flyers included a QR code directing interested individuals to the online REDCap survey platform. Participants could opt to complete the survey, participate in an interview, or both. Participation was voluntary and could be discontinued at any point.

XSL•FO
RenderX

Data Collection

Survey Procedures

The survey instrument was developed through an iterative consensus process with subject matter experts in critical care and health informatics. The instrument assessed four primary domains: (1) clinician demographics and prior experience with AI, (2) general perceptions of AI's role and impact, (3) trust in AI recommendations across different clinical scenarios, and (4) the importance of various AI usability features. The survey was administered online via REDCap (Vanderbilt University) and remained open from February 9 to April 1, 2025. Items used Likert-scale (5-point), multiple-choice, and open-ended response formats.

The survey was voluntary and distributed through ICU and ED provider listservs, IRB-approved flyers with QR codes, departmental presentations, and word-of-mouth referral. The survey was administered electronically through REDCap, and participants reviewed study information and provided informed consent before participation. No survey weighting was applied. Since recruitment occurred through both listserv-based and open-ended channels, the exact number of clinicians who received or viewed the invitation could not be determined, and a precise response rate could not be calculated. A Checklist for Reporting Results of Internet E-Surveys (CHERRIES) checklist was completed [15] (Checklist 2).

The survey items were organized into four construct groups for analysis: (1) AI perception (6 items mapping to TAM perceived usefulness: accuracy, efficiency, variability, the two complementarity items, and reliability; the negatively worded overreliance item was analyzed as a standalone item), (2) trust scenarios (4 binary items across patient acuity levels, informed by Jian et al's [12] trust-in-automation framework), (3) trust beliefs (3 Likert items: reliability, understanding-trust link, complementarity), and (4) implementation factors (6 importance-rated items mapping to unified theory of acceptance and use of technology [UTAUT]; facilitating conditions and TAM perceived ease of use [11,16]).

Interview Procedures

Clinicians who completed the survey were subsequently invited to participate in semistructured interviews conducted via Zoom (Zoom Communications, Inc) from March 20 to April 7, 2025. Interviews lasted approximately 30 to 45 minutes and followed a structured guide developed by the research team based on existing literature regarding AI-CDSS adoption. All interviews were conducted by the lead analyst (MD, who has clinical training in dentistry and informatics). The interview component is reported in accordance with the Consolidated Criteria for Reporting Qualitative Research (COREQ) [17] (Checklist 3). Interview questions explored the same domains as the survey while allowing for deeper examination of participants' reasoning and preferences. The interviews incorporated A/B testing of different AI-CDSS interface designs. Alert prototypes were derived from the AI-CDSS utilized in the PRECISE trial, which encourages clinicians to consider changing their order from normal saline to balanced crystalloid in a population of infected patients that have been shown to have a mortality benefit with balanced crystalloid using an AI algorithm [6]. Participants evaluated three distinct pairs of mock-up alerts (version A vs version B; Multimedia Appendices 1-4), each designed to assess specific aspects of usability and information framing. For each pair, participants selected their preferred design and articulated detailed reasoning for their choice. The presentation order of A/B test pairs was consistent across participants; randomization or counterbalancing was not employed, which may introduce order effects. Additionally, each A/B pair tested multiple design variables simultaneously (eg, opt-in vs opt-out combined with different action language), limiting the ability to isolate individual design effects.

Participants were also presented with a clinical case scenario designed to elicit nuanced perspectives on the practical application of an AI-CDSS. Nurses were excluded from this section since the decisions made may be outside their scope of practice. The scenario depicted a 54-year-old male with multiple comorbidities presenting with septic shock. This case was used to anchor the discussion and probe opinions across five decision points in a critical care resuscitation (Multimedia Appendix 5).

All interviews were audio-recorded using Zoom's built-in recording function, transcribed verbatim using automated transcription with subsequent manual review for accuracy, and deidentified prior to analysis. Complete survey and interview instruments are provided in Multimedia Appendices 6 and 7 respectively.

Data Analysis

Survey responses were analyzed using descriptive statistics (frequencies, proportions, means, and SDs) computed in Python (SciPy, NumPy). Proportions are reported with 95% Wilson score CI, preferred over Wald intervals for small samples and extreme proportions. Likert-type items (5-point scales) are reported with means, SDs, and 95% CIs. Internal consistency of multi-item scales was assessed using Cronbach α. The McDonald ω (total, from a one-factor model) is analyzed alongside α because ω provides a less restrictive estimate of internal consistency and is recommended as a practical complement to α [18]. Trust across four patient acuity levels was compared using Cochran Q test for related binary samples, with pairwise McNemar exact tests and Bonferroni correction (6 comparisons, adjusted α=.0083). Holm step-down and Benjamini-Hochberg false discovery rate adjustments are presented alongside Bonferroni, and a Friedman test on the full 5-point trust ratings served as a nonparametric sensitivity analysis. Medians and IQRs are reported with means and SDs for Likert-type items in a supplementary table (Multimedia Appendix 8). Associations between categorical variables (department×perception, role×AI usage, AI experience×trust) were assessed using chi-square tests or Fisher exact tests when expected cell counts fell below 5. Since more than 20% of expected cell counts were below 5 in these cross-tabulations, they were treated as exploratory sensitivity analyses, and Monte Carlo permutation *P* values (fixed margins, 100,000 simulated tables) were computed alongside the asymptotic tests. Given the exploratory sample size (N=57), we report exact counts alongside percentages and note cell-count limitations for cross-tabulations. For the multiselect item on desired AI

XSL•FO
RenderX

applications, respondents could choose any number of options, and percentages are reported as shares of all selections. Exploratory subgroup analyses compared ICU and ED respondents on each main outcome using 95% Wilson CIs, Fisher exact tests for binary outcomes, and Mann-Whitney *U* tests for ordinal items.

Interview transcripts and open-ended survey responses were analyzed using thematic analysis following the six-phase framework described by Braun and Clarke [19]. Two reviewers (MD and KC) independently coded the data using ATLAS.ti (version 26; ATLAS.ti Scientific Software Development) and Microsoft Excel. Coding was conducted both deductively, based on the structure of the interview guide, and inductively to capture emergent patterns. The reviewers first reviewed transcripts to identify recurring ideas, phrases, and concepts, which were then labeled as codes and organized into a shared codebook, through consensus discussions with a third reviewer (SP), refining the codebook across multiple coding cycles. During codebook refinement and theme development, codes and excerpts were reviewed in relation to SEIPS 2.0 work-system domains while allowing inductive themes to emerge from the data. Related codes were then collated into preliminary themes and subthemes based on recurring patterns in the interview transcripts. These preliminary themes were reviewed against the underlying excerpts, refined through consensus discussions among the analytic team, and revised to reduce overlap and improve conceptual clarity. Themes addressing similar concepts were then merged into six overarching thematic clusters based on recurrence across responses, relevance to the study aims, clinical interpretability, and alignment with SEIPS 2.0 work-system domains. The final cluster structure was validated by comparing each cluster against representative excerpts, resolving discrepancies through coder discussion, and integrating the qualitative findings with survey results. Integration types in the joint display (convergent, complementary, expansion, or emergent) were assigned through consensus discussion between the two coders. Open-ended survey responses were coded using the same codebook to ensure analytical coherence across data sources. An audit trail was maintained throughout the analytic process, and regular debriefing sessions among coders ensured analytic rigor. Intercoder agreement was assessed at early coding stages using percentage agreement, which exceeded 85% prior to consensus discussion. A third reviewer (SP) not directly involved in the clinical trial adjudicated discrepancies.

Ethical Considerations

The study received approval from the Emory University and Indiana University IRBs (IRB #STUDY00008904 and Protocol #26246 respectively), and all participants provided informed consent prior to participation and data collection. All study procedures adhered to ethical guidelines for human subjects' research. All data were deidentified, and transcripts and survey data were stored on secure, encrypted servers. Participants' identities were protected throughout data analysis and reporting.

We acknowledge that the positionality of the research team, several members of which are embedded in the PRECISE trial [6], may introduce bias toward favorable framing of AI-CDSSs. The lead analyst (MD) conducted interviews with prior knowledge of AI-CDSS design principles; however, the dual-coder approach, inclusion of clinician coinvestigators (DM, JWD, SVB) with direct patient care responsibilities, analytic debriefing, and team-based consensus review during analysis helped mitigate interpretive bias. We have attempted to present both supportive and critical clinician perspectives throughout.

Results

Survey Findings

Respondent Characteristics

Fifty-seven clinicians completed the survey (98.2% completion rate, with 1 partially incomplete). The sample comprised ICU providers (39/57, 68%, 95% CI 56% - 79%), ED providers (13/57, 23%, 95% CI 14% - 35%), and other settings or the general ward (5/57, 9%, 95% CI 4% - 19%). Respondents represented diverse clinical roles: ICU advanced practice providers (APP; 15/57, 26%), nurses (12/57, 21%), residents and fellows (9/57, 16%), intensivists (8/57, 14%), emergency physicians (7/57, 12%), and ED APPs, pharmacists, and other roles (6/57, 11%). Experience ranged from <1 year (2/57, 4%) to >15 years (9/57, 16%), with the largest group at 1 - 5 years (18/57, 32%).

Of 57 respondents, 22 (39%, 95% CI 27% - 51%) reported prior AI use in clinical practice, while 35 (61%) had not. Among AI users (n=22), usage frequency was daily (n=5, 23%, 95% CI 10% - 43%), weekly (n=11, 50%, 95% CI 31% - 69%), monthly (n=3, 14%), and rarely (n=3, 14%). Confidence in AI recommendations among users was generally positive: confident (n=8, 36%, 95% CI 20% - 57%), moderately confident (n=9, 41%, 95% CI 23% - 61%), slightly confident (n=3, 14%), and not at all confident (n=2, 9%). Combined positive confidence (confident+moderately confident) was 77% (95% CI 57% - 90%).

Perception of AI Role and Impact

Fifty-four percent (31/57, 95% CI 42% - 67%) of survey respondents perceived AI as "helpful but supplementary," 23% (13/57, 95% CI 14% - 35%) considered it "transformative and essential," 19% (n=11) were "neutral/uncertain," and only 4% (2/57, 95% CI: 1% - 12%) viewed AI as "potentially disruptive." Trust in AI varied significantly with patient acuity (Cochran Q_3=30.40; P<.001): stable ward patients elicited the highest combined trust (somewhat trust+strongly trust) at 75% (95% CI 63% - 84%), followed by deteriorating ward patients at 47% (95% CI 35% - 60%), undifferentiated ED patients at 44% (95% CI 32% - 57%), and critically ill ICU patients at 44% (95% CI 32% - 57%). Notably, strong distrust increased markedly with acuity: from 2% for stable patients to 7% for deteriorating patients to 12% for critically ill ICU patients. Overall, 68 % (95% CI 56% - 79%) of respondents expressed ethical concerns about AI in clinical decision-making. Regarding overreliance, 79% (95% CI 67% - 88%) agreed or strongly agreed that overreliance on AI poses risks to patient outcomes.

Likert-scale responses further characterized clinician attitudes: most clinicians agreed that AI enhances care efficiency (36/57,

XSL•FO
RenderX

63%) and complements clinical expertise (39/57, 68% to 43/57, 75%), while over three-quarters reported that understanding AI outputs increases trust (45/57, 79%). At the same time, concerns about overreliance on AI were common (45/57, 79%). Confidence in AI's reliability was more moderate, with 47% (27/57) agreeing that AI provides reliable insights. Internal consistency was considered good for the 6-item AI Perception scale (Cronbach α=0.891; McDonald ω=.895) and acceptable for the 3-item Trust subscale (α=0.743; ω=.782) and the 6-item Implementation scale (α=0.740; ω=.746), all exceeding the 0.70 acceptability threshold. Detailed Likert-scale statistics are presented in Table 1.

Table . Clinician attitudes toward AI-based clinical decision support system (Likert-scale responses).

Statement	Mean (SD)	95% CI (mean)	Median (IQR)	Agree and strongly agree combined % (95% CI)
AI improves accuracy of decisions	3.54 (0.78)	3.34 - 3.75	4 (3-4)	56 (43 - 68)
AI enhances care efficiency	3.79 (0.80)	3.58 - 4.00	4 (3-4)	63 (50 - 75)
AI reduces decision variability	3.51 (0.78)	3.31 - 3.71	4 (3-4)	54 (42 - 67)
AI complements clinical expertise	3.74 (0.77)	3.54 - 3.94	4 (3-4)	68 (56 - 79)
AI insights complement expertise	3.86 (0.74)	3.67 - 4.05	4 (4-4)	75 (63 - 85)
AI provides reliable insights	3.35 (0.88)	3.12 - 3.58	3 (3-4)	47 (35 - 60)
Overreliance on AI poses risks	4.05 (0.91)	3.82 - 4.29	4 (4-5)	79 (67 - 88)
Understanding AI outputs increases trust	4.14 (0.97)	3.89 - 4.39	4 (4-5)	79 (67 - 88)

Inferential analysis revealed that trust varied significantly across patient acuity levels (Cochran Q_3=30.40; *P*<.001). Pairwise McNemar tests with Bonferroni correction (6 comparisons, adjusted α=.0083) confirmed that trust in stable patients was significantly higher than in deteriorating (*P*<.001), ICU (*P*<.001), and ED (all *P*<.001) scenarios, while the three nonstable scenarios did not differ significantly from each other. Conclusions were unchanged when Holm step-down and Benjamini-Hochberg false discovery rate adjustments were used instead of Bonferroni. A Friedman test on the full 5-point trust ratings confirmed the acuity effect (χ^2_3=33.5; *P*<.001; Kendall W=0.20). In exploratory sensitivity analyses, the cross-tabulations had more than 20% of expected cell counts below 5 and are not confirmatory. The association between department and AI perception was significant under the asymptotic test (χ^2_9=17.4*;* *P*=.04) but was not robust under Monte Carlo permutation testing (*P*=.08). The association between role and AI usage was significant (χ^2_7=16.5*;* *P*=.02; permutation *P*=.01). Both findings require confirmation in larger samples. Fisher exact test for AI experience and trust in stable patients was not significant (OR 0.79; 95% CI 0.23-2.69; *P*=.76).

Barriers to AI Adoption

Among the 22 respondents with AI experience, insufficient training (n=12, 55%, 95% CI 35% - 73%) and lack of trust in AI recommendations (n=12, 55%, 95% CI 35% - 73%) were the most frequently cited barriers, followed by workflow disruption (n=7, 32%, 95% CI 16% - 53%) and other factors (n=4, 18%, 95% CI 7% - 39%). Only 2 respondents (9%, 95% CI 3% - 28%) reported no barriers. The values here are reported among AI users only, as barriers are most meaningful for those with direct experience. Given the small denominator (n=22), these proportions have wide CIs and should be interpreted cautiously.

Desired Features and Implementation Priorities

Respondents universally emphasized the importance of explainability: 75% (43/57) rated transparency as "very important" and 25% (14/57) as "important," achieving 100% combined endorsement (95% CI 94% - 100%). Furthermore, 79% (45/57, 95% CI 67% - 88%) agreed or strongly agreed that understanding how AI generates its outputs increases their trust. Implementation factor importance ratings (on a 5-point scale from "Not important" to "Essential") were: nondisruptive to workflow (mean 4.47, SD 0.66; 91%, 95% CI 81% - 96%), positive patient outcomes (mean 4.35, SD 0.81; 88%, 95% CI 77% - 94%), clear recommendations (mean 4.14, SD 0.81; 81%, 95% CI 69% - 89%), alignment with clinical judgment (mean 4.11, SD 0.75; 84%, 95% CI 73% - 92%), training provided (mean 3.91, SD 0.97; 70%, 95% CI 57% - 81%), and intuitive interface design (mean 3.89, SD 0.88; 75%, 95% CI 63% - 85%). Internal consistency for the 6-item implementation scale was acceptable (Cronbach α=0.740).

Respondents could select multiple desired AI applications, so percentages are shares of the 213 total selections. The most desired applications were antibiotic selection (41/213, 19.2%), image interpretation (37/213, 17.4%), ventilator management (30/213, 14.1%), sepsis prediction (28/213, 13.1%), deterioration prediction (26/213, 12.2%), fluid management (24/213, 11.3%), and vasopressor management (18/213, 8.5%). Preferences varied by clinical role: nurses prioritized image interpretation, deterioration prediction, and sepsis prediction, while intensivists preferred antibiotic selection and image interpretation. By department, ICU clinicians most valued antibiotic selection

XSL•FO
RenderX

(28/136 ICU selections, 20.6%), while ED clinicians showed more evenly distributed preferences across use cases.

Exploratory ICU vs ED Subgroup Comparisons

In exploratory subgroup analysis, prior AI use was more common among ED respondents (10/13, 77%; 95% CI 50% - 92%) than among ICU respondents (12/39, 31%; 95% CI 19% - 46%; *P*=.008). No other main outcome differed significantly between the two departments. For example, combined trust for a stable ward patient was 28/39 (72%; 95% CI 56% - 83%) among ICU respondents and 11/13 (85%; 95% CI 58% - 96%) among ED respondents. Full subgroup statistics with 95% Wilson CIs are provided in Multimedia Appendix 9. The ED subgroup was small, so these comparisons are hypothesis generating only.

Interview Findings

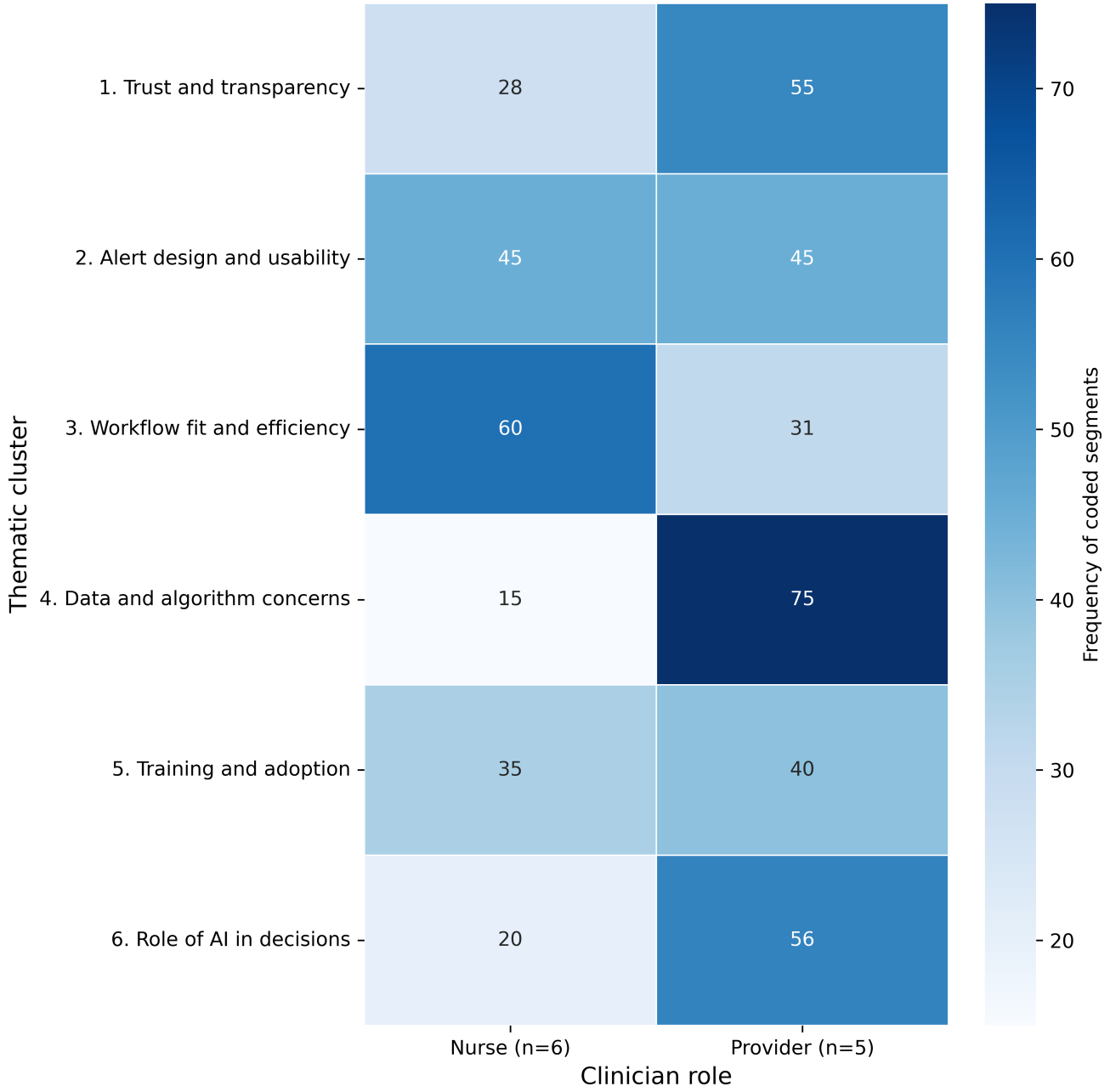
Overview

Semistructured interviews were conducted with 11 participants, including physicians (3), APPs (2), and nurses (6). The thematic analysis for the interviews identified 36 detailed themes from approximately 690 coded segments. To support synthesis and clinical interpretability, these themes were organized into six overarching thematic clusters: (1) trust, transparency, and human control (2) alert design, interface, and usability (3) workflow fit and efficiency (4) data accuracy and algorithm concerns (5) training, rollout, and adoption barriers (6) role of AI in clinical decision making. The six clusters serve as the primary analytic structure. The 36 detailed themes, related subthemes, and representative quotations are provided in Multimedia Appendix 10.

Thematic saturation was considered achieved after the tenth interview, when no new codes or subthemes emerged. Following Malterud et al's [20] information power framework, our interview sample (n=11) was deemed sufficient given the narrow study aim, specific participant group, strong interview dialogue quality, and application of established theoretical frameworks. A frequency-based analysis of thematic codes confirmed progressive code saturation across the final three interviews. This pattern is consistent with the code saturation thresholds described by Hennink et al [21]. A frequency of thematic discussion is illustrated in Figure 1.

XSL•FO
RenderX

Figure 1. Frequency of thematic discussion by clinician role. This heatmap shows the frequency of coded segments for each thematic cluster, compared between nurses and providers. The "provider" category includes physician, resident, and advanced practice provider participants.



Cluster 1: Trust, Transparency, and Human Control

Trust emerged as a central but conditional concept in clinician attitudes toward AI. Participants described trust as dependent on transparency, logic, and the ability to maintain control over the final clinical decision. The need for explainability was repeatedly emphasized, particularly in relation to opaque or "black-box" algorithms: This qualitative finding aligns with survey data showing that 79% (95% CI 67% - 88%) of respondents agreed or strongly agreed that understanding how AI generates its outputs increases trust, while only 47% (95% CI 35% - 60%) agreed that AI systems provide reliable and accurate insights, a 32 percentage-point gap suggesting that trust is contingent on transparency mechanisms rather than assumed reliability.

> *If it's going to be a black box in like a neural network, it needs to at least make logical sense.* [P8, Emergency Physician]

Clinicians consistently emphasized the importance of human oversight and override authority:

> *If you're thinking of a supportive tool, you should be able to override the tool's recommendation.* [P9, Nurse Practitioner]
>
> *I only trust AI 50%. I still go through the reasoning myself and apply it to the situation at hand.* [P6, Nurse]

Even among those optimistic about AI's potential, trust was framed as "trust, but verify," rooted in the belief that decisions must always align with clinical reasoning and patient context.

XSL•FO
RenderX

> *It depends on the AI system…it's a trust, but verify. It needs to make logical sense. If it doesn't, or it doesn't have robust data behind it, then I'm not gonna trust it.* [P8, Emergency Physician]

> *AI can pop up and say "Hey, this number looks weird," but it's still up to the human to say...we are going to continue with our plan of care...* [P8, Emergency Physician]

Clinicians viewed AI as an "extra set of eyes" that enhances safety but must coexist with human interpretation:

> *I think it's an extra set of eyes…an extra layer of safety. But you still need humans putting all the pieces together.* [P5, Nurse]

> *You never know who put the original information in....If there was some transparency with that, then I think the trust level would be higher.* [P6, Nurse]

Cluster 2: Alert Design and Interface Usability

Participants emphasized user control, simplicity, and relevance in the design of alerts and interfaces. The majority preferred opt-in alerts that allow deliberate review rather than automatic notifications. Concise recommendations coupled with links to supporting evidence were considered ideal:

> *Links to the study rationale as to why it's doing it are always going to be helpful.* [P8, Emergency Physician]

> *If I was in clinical settings...I'd love to have a short version, and then click to see the studies when I actually have time.* [P2, Nurse]

Clinicians highlighted the importance of balancing best-practice guidance with individualized patient care:

> *Version A [opt-in] allows the clinician to take a moment and review which one is better for the patient.* [P2, Nurse]

A subset of participants expressed fatigue toward AI branding, noting that overuse of the term "AI" could reduce acceptance or trigger trust issues:

> *I'm also feeling this myself, and I do AI research. People are just getting burnt out by the term.* [P11, ICU Physician]

> *If we see "recommended by AI" on every single thing...then that could trigger trust issues...* [P5, Nurse]

Some envisioned adaptive electronic medical record (EMR) systems that would learn user patterns to reduce redundant tasks and improve efficiency:

> *The idea of a learning EMR that adapts to my patterns and shows the information I need…without all the extra clicking, would be a real time saver.* [P11, ICU Physician]

Cluster 3: Workflow Integration and Efficiency

Several participants viewed AI as a tool to enhance administrative efficiency and relieve documentation burden:

> *AI impact on nursing workflow…most importantly clinical documentation and note-taking, because many times when you are the ICU clinician, you are busy.* [P6, Nurse]

> *Ambient listening can save time for providers in terms of summarizing notes.* [P9, Nurse Practitioner]

Clinicians repeatedly noted that workflow integration would determine whether AI-CDSS tools succeed or fail in practice. Systems that disrupt time-sensitive tasks or introduce additional clicks were viewed negatively, even if algorithmically advanced. Reliable data inputs and minimal disruption were essential for trust and adoption:

> *It's a great tool, but if you make it harder to use or make me click more, people will avoid it.* [P5, Nurse]

Participants favored AI implementations that fit seamlessly into existing workflows, augmenting rather than complicating clinical processes.

> *You have to find the right use case. If it's interrupting the flow or making my job harder, it won't be used.* [P6, Nurse]

Cluster 4: Data Accuracy and Algorithm Concerns

Across interviews, data integrity surfaced as a recurring theme. Participants questioned whether AI models could operate effectively given the variability and inaccuracies of EHR data. Concerns about algorithmic reliability often traced back to inconsistent documentation or sensor readings rather than mistrust of AI itself.

> *AI is only as good as the data it's trained on.* [P7, Physician]

> *There's a lot of garbage when it comes to vital signs, like a ridiculous amount.* [P8, Emergency Physician]

> *If you're using a model that's based on garbage…it will harm people.* [P1, Nurse]

Clinicians worried that even a well-designed algorithm could amplify inequities or generate unsafe recommendations if trained on biased or incomplete datasets.

> *It would be harmful in patient populations where we don't have a lot of data…broadly applying things that may be true for the majority may not apply to the minority.* [P11, ICU Physician]

A related barrier noted by several participants was automation bias, reflecting apprehension that clinicians might overtrust AI outputs without understanding their context or training data:

> *I worry about automation bias…people treating what AI says as golden, without knowing how it was trained or the context it's appropriate for.* [P11, ICU Physician]

Cluster 5: Training, Rollout, and Adoption Barriers

The theme of training and implementation was strongly represented across participants. Many clinicians cited past experiences with EHR rollouts that lacked context or interactivity:

XSL•FO
RenderX

> *I've had experiences where things in the clinical setting get rolled out without much explanation or any explanation at all.* [P10, Nurse Practitioner]
>
> *There was no opportunity to interact or ask questions…I would want to know the reason it was being implemented, the benefit of it.* [P2, Nurse]

Participants consistently favored peer-led "super user" programs that provided accessible, hands-on training:

> *Emory is very good at doing that, any time a new system is introduced, we have super users.* [P6, Nurse]
>
> *I think if there was a dedicated person…that would be helpful; everyone is already overburdened with work.* [P3, Resident]

Most participants rejected static, online formats:

> *I don't want to take a class, and I don't want to do an online course or PowerPoint.* [P3, Resident]

Effective rollout was described as transparent, contextualized, and supported by leadership and time allocation.

Cluster 6: Role of AI in Clinical Decision-Making

Clinicians characterized AI as a complementary aid that can enhance clinical awareness and decision-making but not replace human expertise. They underscored that human touch and situational awareness remain irreplaceable. AI was described as a diagnostic "extension," comparable to tools like ultrasound:

> *I think it should be an extension of our physical exam…just like ultrasound…but with more caution.* [P11, ICU Physician]
>
> *My ideal use is: reinforce what I already thought, or help me think in a new way.* [P8, Emergency Physician]
>
> *You can't replace the hands and body of a nurse.* [P5, Nurse]

Several participants recognized AI's emerging role in detecting subtle findings and supporting clinical intuition, expressing optimism.

> *I think it's going to be a pillar of medical care…for predicting decompensation based on vitals and labs.* [P4]
>
> *AI program caught a small subdural hematoma the radiologist missed...it was really cool to see that.* [P9]
>
> *We're still in the infancy for it….But I think this is going to be the way in which we get a better sense of personalized medicine and healthcare.* [P11, ICU Physician]

Despite optimism, clinicians cautioned that overreliance could erode core clinical skills, noting that excessive reliance on automated systems could diminish traditional clinical capabilities:

> *My clinical exam skills are not on par with my predecessors…the people coming after me, their skill sets have eroded even more.* [P11, ICU Physician]

A comprehensive summary integrating survey and interview findings is presented in Table 2.

XSL•FO
RenderX

Table . Joint display: integration of survey and interview findings.

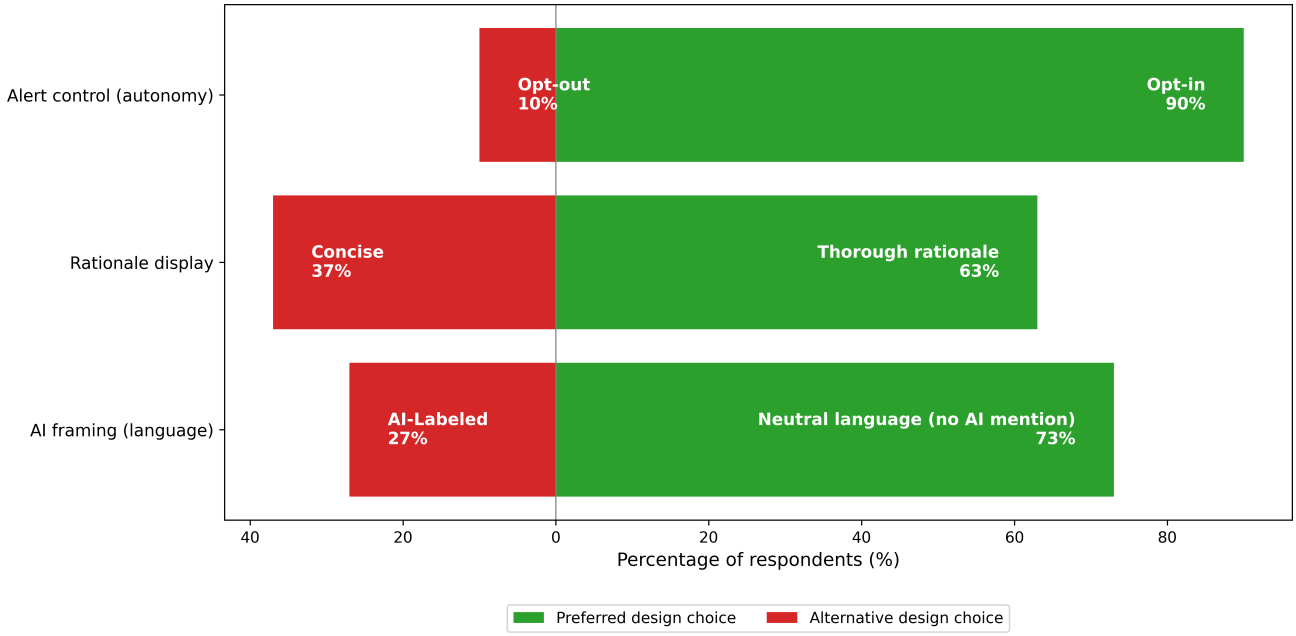
Survey finding (quantitative)	Interview theme (qualitative)	Integration and interpretation
Trust decreases with acuity: 75% stable → 47% deteriorating → 44% each for ICU and ED (Cochran Q=30.40, *P*<.001)	Cluster 1: "I only trust AI 50%...I still go through the reasoning myself" (P6). Trust described as conditional and context-dependent.	Convergent: Both data sources confirm acuity-dependent trust. Qualitative data reveals this reflects clinicians' belief that higher-acuity decisions require greater human judgment, not blanket AI distrust.
55% of AI users cited insufficient training as barrier; 55% cited lack of trust	Cluster 5: "Things get rolled out without much explanation" (P4). Participants favored peer-led "super user" programs over online training.	Expansion: Survey identifies training as top barrier; interviews explain why past negative rollout experiences create lasting skepticism. Training format (hands-on vs online) matters more than training availability.
Exploratory A/B preference observation: 10/11 participants preferred opt-in alerts; 8/11 preferred non-AI-branded alerts	Cluster 2: "People are just getting burnt out by the term [AI]" (P11). Preference for clinician-controlled, evidence-framed alerts.	Convergent: Strong alignment between survey preferences and interview explanations. Debranded, opt-in design reflects desire for autonomy and reduced alert fatigue. These exploratory mock-alert preferences require replication and should not be interpreted as causal design conclusions.
47% agreed AI provides reliable insights; 79% concerned about overreliance	Cluster 4: "AI is only as good as the data it's trained on" (P7). Data quality concerns central to distrust.	Complementary: Survey captures magnitude of distrust; interviews reveal root cause, concerns about training data quality and EHR data integrity rather than AI capability per se.
100% rated transparency as important or very important; 79% said understanding outputs increases trust	Cluster 1: "If it's a black box...it needs to at least make logical sense" (P8). Transparency enables "trust, but verify" approach.	Convergent: Universal agreement on transparency importance. Interviews reveal clinicians want rationale not for blind acceptance but for active verification, consistent with maintaining clinical expertise.
54% view AI as "helpful but supplementary"; 68% agree AI complements expertise	Cluster 6: "It should be an extension of our physical exam" (P11). AI as "extra set of eyes," not replacement.	Convergent: Majority frame AI as augmentation tool. Both sources align on AI-as-supplement, with interviews providing the metaphor of AI as safety net.
No direct survey item on AI branding (emergent finding from interviews only)	Cluster 2: "If we see recommended by AI on every single thing...that could trigger trust issues" (P5).	Emergent: AI-branding fatigue identified exclusively through qualitative data, a novel finding that could not have been captured by survey alone, demonstrating the value of the mixed-methods design.

A/B Testing

In exploratory A/B testing of mock CDS alerts, 91% (10/11) of participants preferred an opt-in version over an opt-out version. When comparing a CDSS alert with a short, concise recommendation vs a more thorough explanation of the recommendation including a hyperlink to the study evidence, most participants (7/11, 64%) preferred the thorough explanation with the study rationale hyperlink included, while 18% (2/11) preferred having a balance of brief, concise alert messages, with a clickable hyperlink to the study evidence for reference at point of care. Seventy-three percent of participants (8/11) preferred the CDSS alert that did not mention AI in its recommendation. All respondents emphasized the importance of AI explainability (Figure 2). Further discussion of interface design considerations derived from the A/B testing results is presented in Multimedia Appendix 11. Each A/B pair varied several design features at once and the presentation order was fixed, so these percentages are exploratory preference observations. They should not be attributed to any single design element, and they require replication before generalization.

XSL•FO
RenderX

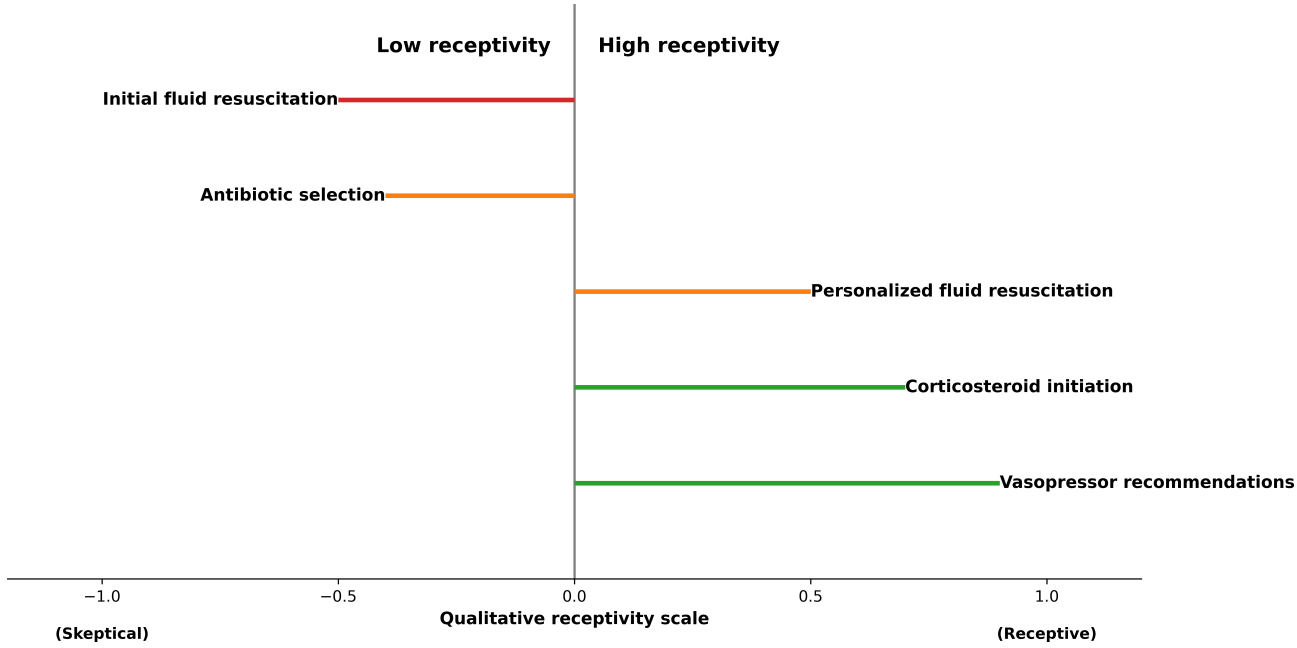
Figure 2. A/B testing preferences.



Case Scenario

A clinical sepsis scenario was presented to a subgroup of five providers (three physicians and two APPs). Their perceptions of five potential AI-driven interventions revealed a varied pattern (Figure 3). Clinicians were skeptical of AI for straightforward, protocolized tasks like initial fluid resuscitation and antibiotic selection in an uncomplicated case and were more receptive when decisions involved higher complexity like ongoing fluid management, vasopressor selection, and corticosteroid initiation, though this was often conditional on trusting the underlying data. The detailed responses are summarized in Multimedia Appendix 12. This component included only five providers and excluded nurses, although nurses participated in all other survey and interview components. It is presented as illustrative qualitative evidence rather than a basis for broader conclusions.

Figure 3. Physicians' perceptions on AI-driven clinical decision support systems in sepsis case scenarios. We created a qualitative receptivity scale (–1.0=skeptical to +1.0=receptive) to illustrate clinician responses to the five AI-driven interventions. Placement on the scale was assigned by the coding reviewers through consensus, informed by thematic coding of case scenario responses and not derived from a quantitative measure, with left indicating skepticism, center reflecting conditional acceptance, and right showing consistent openness. The scale is conceptual rather than numerical, serving to highlight relative differences in receptivity.



XSL•FO
RenderX

Discussion

Principal Findings

This study demonstrates that while clinicians in high-acuity settings are optimistic about AI's potential, their willingness to adopt these tools is contingent on a core set of principles. The AI-CDSS should be transparent in its algorithm, integrate seamlessly into complex workflows, and respect clinician autonomy and judgment. To bridge the gap between technological innovation and meaningful clinical use, the focus must shift from the capabilities of the algorithm alone to address the needs and perceptions of the clinicians for successful implementation. Our findings reflect perceived acceptability, elicited with mock alerts and hypothetical scenarios, and not actual use, implementation success, or patient outcomes. The study is therefore best understood as an early-stage formative evaluation, consistent with Phase 0 of the Developmental and Exploratory Clinical Investigations of Decision Support Systems Driven by AI (DECIDE-AI) framework [22].

Viewed through the SEIPS 2.0 framework [10], our findings map onto all five work system components: people (clinicians' trust varies by role, experience, and acuity context), tasks (AI-CDSS must fit complex, time-pressured clinical decision-making), tools and technologies (alert design preferences including opt-in mechanisms and debranded interfaces), organization (training infrastructure, rollout strategy, and peer champion models), and environment (ICU and ED contexts with distinct workflow pressures). Our results demonstrate bidirectional interactions: organizational training deficits reduce trust in the technology, while poor technology design increases task burden, illustrating why isolated interventions targeting single SEIPS components are unlikely to succeed. Interpreted through Hoff and Bashir's [13] three-layer model of trust, the acuity-dependent trust gradient reflects situational trust, and the AI-branding fatigue described in interviews reflects learned trust shaped by prior technology experiences [8]. While several of these findings converge with prior work, such as clinicians valuing transparency and explainability being important for adoption [23-31], this study identifies three distinctive insights for successful CDSS creation: preference for lower acuity scenarios, a preference-effectiveness paradox, and emphasis on training and implementation. We use the term preference-effectiveness paradox to mean that the alert designs clinicians prefer are not necessarily the designs that prove most effective in practice.

Initial AI-CDSSs should be strategic to target lower acuity scenarios while maintaining clinician autonomy to foster clinician confidence in their implementation. The demonstration of acuity-dependent trust in this study (decreasing from 75% [43/57] in stable patients to 44% [25/57] in critically ill patients) suggests a graduated, SEIPS-informed implementation strategy that accounts for task complexity. We frame this strategy as a recommendation for future testing because this study did not evaluate implementation or measure changes in confidence over time. This strategy allows a CDSS to demonstrate its reliability and value in a safer context, creating a foundation of trust before it is applied to higher acuity situations. Study participants most frequently requested AI-CDSSs for radiology and antibiotic selection, which are both lower acuity scenarios. Adaptability of AI tools at lower stakes situations has been shown to be effective in clinical scenarios like reading chest X-rays (CXRs) [32-34]. While implementation of AI assistance in CXR reading has been directed at radiologists, these applications may also have value in integration for critical care or ED clinicians. For instance, AI has the capability to effectively identify a pneumothorax, which could assist in earlier detection in resource-limited hospitals [35]. This algorithm could then, through a CDSS, prompt the provider to review or re-examine the CXR. For antibiotic selection, an AI-CDSS could make recommendations based on the hospital antibiogram and patient's former infections and antibiotic exposure. One study showed promising development of an AI algorithm that effectively detected *Stenotrophomonas maltophilia* with antibiotic resistance, and a CDSS could more rapidly alert a clinician compared to the standard practice [36]. Integration of AI in these lower acuity situations could help establish trust between the clinician and AI-CDSS [37]. As exposure to AI-CDSSs increases, these perceptions may change, and ongoing evaluation would be helpful to capture evolving attitudes and possibly allow for higher acuity applications. Our understanding of acuity with AI-CDSS recommendations remains theoretical, and future studies should compare the deployment of AI-CDSSs in variable acuity settings.

Our study demonstrated the importance of preserving clinician autonomy for acceptance of the AI-CDSS, which can be done by designing an "opt-in" over an "opt-out" system. Preserving clinician control over when and how AI recommendations are provided may support trust and perceived usefulness of the system. However, it is important to acknowledge that clinician preferences do not necessarily predict optimal AI-CDSS effectiveness in practice. Further studies are needed to compare different AI-CDSS design models in actual clinical settings. While opt-in designs may be preferred, their passive nature may result in lower compliance rates compared to more intrusive opt-out configurations in practice, as evidenced from prior clinical decision support research [5]. This demonstrates the preference-effectiveness paradox, in which autonomy-preserving designs may be preferred in fast-paced settings like the ED and ICU workflows, but optional alerts may be missed or deprioritized. Autonomy-preserving designs should therefore be paired and evaluated with monitoring of alert uptake and safety outcomes in future research. From a SEIPS perspective, this highlights an important tension between the person (clinician preferences) and organization (system effectiveness) components and underscores the need for future implementation studies to evaluate different alert design strategies in real clinical settings.

Interestingly, there was a strong preference for alerts that did not mention AI. It may reflect AI fatigue, distrust, framing effects, or concern about accountability. Debranding also creates an ethical tension with transparency because clinicians and patients may reasonably expect to know when a recommendation is AI-derived [23,26]. This exploratory finding requires replication and dedicated study before it informs design practice.

XSL•FO
RenderX

Finally, our study highlights that even a perfectly designed tool will fail if not implemented thoughtfully. "Insufficient training" was identified as a primary barrier, along with causing lack of trust. Participants reflected on prior EHR rollouts that were characterized by one-way, module-based training and minimal opportunity for feedback, which often resulted in disengagement. Conversely, interactive, peer-led "super-user" models were described as more effective and engaging. These "super-users" should have a strong understanding of the AI systems and algorithms, as well as familiarity with organizational change management processes. In addition to technical expertise, trainers should possess strong communication and listening skills to effectively engage with clinical teams and translate technical system capabilities into practical clinical use [38]. Consistent with prior studies, participants reported that training is more meaningful when it explains not only how to use the system but also why it is being implemented and what value it provides [39,40]. Participants also highlighted the importance of transparency about AI model development, including its training data and validation process [41,42]. Broader literature similarly points to the significance of addressing "change fatigue" and providing protected time for clinicians to participate in training to ensure effective engagement [43]. Training that is interactive, context-specific, and psychologically informed may mitigate resistance and improve confidence in AI-CDSS use [39,44].

Limitations and Future Directions

This study should be interpreted within the context of several limitations. First, data collection occurred within a single academic medical center, which may limit generalizability. Perceptions may differ in community hospitals and resource-limited settings with less AI exposure, and a multicenter trial is an explicit next step. Second, both interview and survey participation were voluntary, which may have introduced selection bias toward individuals with particularly strong opinions about AI. Recruitment channels were open-ended, so a response rate could not be computed, and respondents could not be compared with the underlying clinician population. In addition, the positive confidence estimates among respondents with prior AI use should be interpreted cautiously because this subgroup was small (n=22); positive selection of AI-engaged clinicians may therefore bias trust and adoption findings upward. Third, the mock alerts used in A/B testing and clinical scenarios may not fully capture the complexity of real-time clinical decision-making. In addition, these mock alerts were derived from a single AI-CDSS prototype associated with the PRECISE trial, and several members of the research team are also investigators in that trial. Therefore, findings should be interpreted as formative guidance rather than broadly generalizable evidence of AI-CDSS implementation across settings. Fourth, our sample size may have limited the power for some findings. The sample was primarily ICU based; ED findings are exploratory, and several cross-tabulations had sparse cells and are reported as sensitivity analyses. Fifth, trust in AI is dynamic and evolves with exposure [8]; our cross-sectional design captures a single time point. Future work could include a longitudinal extension, such as a resurvey of the same cohort after deployment of the PRECISE trial AI-CDSS, to capture learned trust trajectories [13,22]. Finally, this study focused exclusively on clinician perceptions of AI-CDSSs; however, the growing literature on shared decision-making in AI-augmented care emphasizes that patient perspectives are equally critical for successful AI-CDSS adoption [45]. We also note that SEIPS 3.0 has extended the model to include patient journey considerations [46]. Future directions should include multicenter replication of these findings, longitudinal evaluation of the temporal evolution of clinician trust in AI-CDSSs, and input from patient and public advocates to better understand patient-centered acceptability, transparency, and perceived benefit.

In conclusion, this study demonstrates that adoption of AI-CDSSs in critical care is not solely a technical issue but a human-factors challenge that requires focus on trust, transparency, and workflow compatibility. These conclusions are based on clinician perceptions, mock alerts, and hypothetical scenarios in a single academic medical center; the proposed implementation strategies should be interpreted as exploratory and hypothesis-generating. Their effectiveness and generalizability require validation through prospective multicenter studies and real-world AI-CDSS implementation.

Acknowledgments

We thank the clinicians from Emory Healthcare who participated in the survey and interviews, as well as the research staff who supported recruitment and data collection. We also acknowledge the contributions of the Precision Resuscitation With Crystalloids in Sepsis (PRECISE) trial team. No generative AI tools were used for any portion of the manuscript.

Funding

This study was supported by the Kaiser Permanente AIM-HI award and NHLBI R01 HL175626.

Data Availability

Deidentified survey data underlying this study are available from the corresponding author on reasonable request. Due to institutional requirements and participant confidentiality, interview transcripts cannot be publicly shared.

Authors' Contributions

Conceptualization: DM, SP, SVB

XSL•FO
RenderX

Data curation: MD
Formal analysis: MD, KC, SP
Funding acquisition: SVB
Investigation: MD
Methodology: MD, DM, SP, SVB
Project administration: MD, DM, SVB
Resources: DM, SP, SVB, JWD
Supervision: DM, SP, SVB
Validation: SP, MD, DM, SVB
Visualization: MD
Writing – original draft: MD, KC
Writing – review & editing: MD, DM, KC, SP, JWD, SVB

Conflicts of Interest

Several members of the research team are investigators in the Precision Resuscitation With Crystalloids in Sepsis (PRECISE) trial [6], which is the source of the AI-based clinical decision support system (AI-CDSS) prototype used to derive the mock alerts in this study. The authors declare no other competing interests.

Multimedia Appendix 1
A/B testing mock-up pair 1 – critical alert design. Version A displays two user-action buttons (“ACCEPT”/“REJECT”) allowing clinicians to confirm or decline a recommended fluid change. Version B shows an auto-executed substitution (“Lactated Ringer’s Infusion has been ordered instead”), requiring manual cancellation to reverse.
[PNG File, 236 KB - humanfactors_v13i1e95472_app1.png]

Multimedia Appendix 2
A/B testing mock-up pair 2 – evidence-framing design. Version A presents a brief alert message without supporting citation. Version B adds an evidence-based rationale and a hyperlink to the referenced research.
[PNG File, 229 KB - humanfactors_v13i1e95472_app2.png]

Multimedia Appendix 3
A/B testing mock-up pair 3 – algorithm-attribution design. Version A explicitly attributes the recommendation to an AI algorithm. Version B conveys the same guidance using clinical logic and omits any mention of AI.
[PNG File, 175 KB - humanfactors_v13i1e95472_app3.png]

Multimedia Appendix 4
A/B testing alert designs presented to interview participants.
[DOCX File, 13 KB - humanfactors_v13i1e95472_app4.docx]

Multimedia Appendix 5
Clinical case scenario for physicians and advanced practice providers.
[DOCX File, 12 KB - humanfactors_v13i1e95472_app5.docx]

Multimedia Appendix 6
Survey questionnaire.
[DOCX File, 14 KB - humanfactors_v13i1e95472_app6.docx]

Multimedia Appendix 7
Interview questionnaire.
[DOCX File, 14 KB - humanfactors_v13i1e95472_app7.docx]

Multimedia Appendix 8
Supplementary statistical tables.
[DOCX File, 19 KB - humanfactors_v13i1e95472_app8.docx]

Multimedia Appendix 9
Exploratory intensive care unit (ICU) vs emergency department (ED) subgroup analysis.

XSL•FO
RenderX

[DOCX File, 16 KB - humanfactors_v13i1e95472_app9.docx]

Multimedia Appendix 10
Comprehensive overview of 36 themes with subthemes and representative quotes.
[DOCX File, 29 KB - humanfactors_v13i1e95472_app10.docx]

Multimedia Appendix 11
Supplementary insights.
[DOCX File, 19 KB - humanfactors_v13i1e95472_app11.docx]

Multimedia Appendix 12
Summary of clinician perceptions in the sepsis scenario.
[DOCX File, 13 KB - humanfactors_v13i1e95472_app12.docx]

Checklist 1
GRAMMS checklist.
[DOCX File, 13 KB - humanfactors_v13i1e95472_app13.docx]

Checklist 2
CHERRIES checklist.
[DOCX File, 19 KB - humanfactors_v13i1e95472_app14.docx]

Checklist 3
COREQ checklist.
[DOCX File, 16 KB - humanfactors_v13i1e95472_app15.docx]

References

1. Elhaddad M, Hamam S. AI-driven clinical decision support systems: an ongoing pursuit of potential. Cureus 2024 Apr;16(4):e57728. [doi: 10.7759/cureus.57728] [Medline: 38711724]
2. Scott IA, van der Vegt A, Lane P, McPhail S, Magrabi F. Achieving large-scale clinician adoption of AI-enabled decision support. BMJ Health Care Inform 2024 May 30;31(1):e100971. [doi: 10.1136/bmjhci-2023-100971] [Medline: 38816209]
3. Sittig DF, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Qual Saf Health Care 2010 Oct;19 Suppl 3(Suppl 3):i68-i74. [doi: 10.1136/qshc.2010.042085] [Medline: 20959322]
4. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17. [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
5. Blecker S, Pandya R, Stork S, et al. Interruptive versus noninterruptive clinical decision support: usability study. JMIR Hum Factors 2019 Apr 17;6(2):e12469. [doi: 10.2196/12469] [Medline: 30994460]
6. Bhavani SV, Holder A, Miltz D, et al. The Precision Resuscitation With Crystalloids in Sepsis (PRECISE) trial: a trial protocol. JAMA Netw Open 2024 Sep 3;7(9):e2434197. [doi: 10.1001/jamanetworkopen.2024.34197] [Medline: 39292459]
7. Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A roadmap for national action on clinical decision support. J Am Med Inform Assoc 2007;14(2):141-145. [doi: 10.1197/jamia.M2334] [Medline: 17213487]
8. Tun HM, Rahman HA, Naing L, Malik OA. Trust in artificial intelligence-based clinical decision support systems among health care workers: systematic review. J Med Internet Res 2025 Jul 29;27:e69678. [doi: 10.2196/69678] [Medline: 40772775]
9. Shneiderman B. Human-centered artificial intelligence: reliable, safe & trustworthy. Int J Hum Comput Interact 2020 Apr 2;36(6):495-504. [doi: 10.1080/10447318.2020.1741118]
10. Holden RJ, Carayon P, Gurses AP, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 2013;56(11):1669-1686. [doi: 10.1080/00140139.2013.838643] [Medline: 24088063]
11. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep 1;13(3):319-340. [doi: 10.2307/249008]
12. Jian JY, Bisantz AM, Drury CG. Foundations for an empirically determined scale of trust in automated systems. Int J Cogn Ergon 2000 Mar;4(1):53-71. [doi: 10.1207/S15327566IJCE0401_04]
13. Hoff KA, Bashir M. Trust in automation: integrating empirical evidence on factors that influence trust. Hum Factors 2015 May;57(3):407-434. [doi: 10.1177/0018720814547570] [Medline: 25875432]

XSL•FO
RenderX

14. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008 Apr;13(2):92-98. [doi: 10.1258/jhsrp.2007.007074] [Medline: 18416914]
15. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34. [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
16. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003 Sep 1;27(3):425-478. [doi: 10.2307/30036540]
17. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
18. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. Br J Psychol 2014 Aug;105(3):399-412. [doi: 10.1111/bjop.12046] [Medline: 24844115]
19. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
20. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
21. Hennink MM, Kaiser BN, Marconi VC. Code saturation versus meaning saturation: how many interviews are enough? Qual Health Res 2017 Mar;27(4):591-608. [doi: 10.1177/1049732316665344] [Medline: 27670770]
22. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med 2022 May;28(5):924-933. [doi: 10.1038/s41591-022-01772-9] [Medline: 35585198]
23. Rosenbacke R, Melhus Å, McKee M, Stuckler D. How explainable artificial intelligence can increase or decrease clinicians' trust in AI applications in health care: systematic review. JMIR AI 2024 Oct 30;3:e53207. [doi: 10.2196/53207] [Medline: 39476365]
24. Sibbald M, Abdulla B, Keuhl A, Norman G, Monteiro S, Sherbino J. Electronic diagnostic support in emergency physician triage: qualitative study with thematic analysis of interviews. JMIR Hum Factors 2022 Sep 30;9(3):e39234. [doi: 10.2196/39234] [Medline: 36178728]
25. Bergquist M, Rolandsson B, Gryska E, et al. Trust and stakeholder perspectives on the implementation of AI tools in clinical radiology. Eur Radiol 2024 Jan;34(1):338-347. [doi: 10.1007/s00330-023-09967-5] [Medline: 37505245]
26. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 2020 Nov 30;20(1):310. [doi: 10.1186/s12911-020-01332-6] [Medline: 33256715]
27. Cai CJ, Winter S, Steiner D, Wilcox L, Terry M. "Hello AI": uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. Proc ACM Hum-Comput Interact 2019 Nov 7;3(CSCW):1-24. [doi: 10.1145/3359206]
28. Tanaka M, Matsumura S, Bito S. Roles and competencies of doctors in artificial intelligence implementation: qualitative analysis through physician interviews. JMIR Form Res 2023 May 18;7:e46020. [doi: 10.2196/46020] [Medline: 37200074]
29. Wu Y, Wu M, Wang C, Lin J, Liu J, Liu S. Evaluating the prevalence of burnout among health care professionals related to electronic health record use: systematic review and meta-analysis. JMIR Med Inform 2024 Jun 12;12:e54811. [doi: 10.2196/54811] [Medline: 38865188]
30. Trinkley KE, Kahn MG, Bennett TD, et al. Integrating the practical robust implementation and sustainability model with best practices in clinical decision support design: implementation science approach. J Med Internet Res 2020 Oct 29;22(10):e19676. [doi: 10.2196/19676] [Medline: 33118943]
31. Stewart J, Freeman S, Eroglu E, et al. Attitudes towards artificial intelligence in emergency medicine. Emerg Med Australas 2024 Apr;36(2):252-265. [doi: 10.1111/1742-6723.14345] [Medline: 38044755]
32. Lee S, Shin HJ, Kim S, Kim EK. Successful implementation of an artificial intelligence-based computer-aided detection system for chest radiography in daily clinical practice. Korean J Radiol 2022 Sep;23(9):847-852. [doi: 10.3348/kjr.2022.0193] [Medline: 35762186]
33. Shin HJ, Lee S, Kim S, Son NH, Kim EK. Hospital-wide survey of clinical experience with artificial intelligence applied to daily chest radiographs. PLoS One 2023;18(3):e0282123. [doi: 10.1371/journal.pone.0282123] [Medline: 36862644]
34. Hua D, Petrina N, Sacks AJ, et al. Towards human-AI collaboration in radiology: a multidimensional evaluation of the acceptability of AI for chest radiograph analysis in supporting pulmonary tuberculosis diagnosis. JAMIA Open 2025 Feb;8(1):ooae151. [doi: 10.1093/jamiaopen/ooae151] [Medline: 39911582]
35. Hillis JM, Bizzo BC, Mercaldo S, et al. Evaluation of an artificial intelligence model for detection of pneumothorax and tension pneumothorax in chest radiographs. JAMA Netw Open 2022 Dec 1;5(12):e2247172. [doi: 10.1001/jamanetworkopen.2022.47172] [Medline: 36520432]
36. Lin TH, Chung HY, Jian MJ, et al. Innovative strategies against superbugs: developing an AI-CDSS for precise Stenotrophomonas maltophilia treatment. J Glob Antimicrob Resist 2024 Sep;38:173-180. [doi: 10.1016/j.jgar.2024.06.004] [Medline: 38909685]

XSL•FO
RenderX

37. Shamszare H, Chaudhry Z, Berenji M, Choudhury A. Conceptualizing clinicians' trust in artificial intelligence as a function of their expertise, workload, patient outcome, diagnosis difficulty, and AI accuracy: a systems thinking approach. IEEE Access 2025;13:119601-119618. [doi: 10.1109/ACCESS.2025.3586555]
38. Nair M, Svedberg P, Larsson I, Nygren JM. A comprehensive overview of barriers and strategies for AI implementation in healthcare: mixed-method design. PLoS One 2024;19(8):e0305949. [doi: 10.1371/journal.pone.0305949] [Medline: 39121051]
39. Scipion CEA, Manchester MA, Federman A, Wang Y, Arias JJ. Barriers to and facilitators of clinician acceptance and use of artificial intelligence in healthcare settings: a scoping review. BMJ Open 2025 Apr 15;15(4):e092624. [doi: 10.1136/bmjopen-2024-092624] [Medline: 40233955]
40. Russell RG, Lovett Novak L, Patel M, et al. Competencies for the use of artificial intelligence-based tools by health care professionals. Acad Med 2023 Mar 1;98(3):348-356. [doi: 10.1097/ACM.0000000000004963] [Medline: 36731054]
41. Lewis AE, Weiskopf N, Abrams ZB, et al. Electronic health record data quality assessment and tools: a systematic review. J Am Med Inform Assoc 2023 Sep 25;30(10):1730-1740. [doi: 10.1093/jamia/ocad120] [Medline: 37390812]
42. Ranwala RADLMK, Andrade AQ. Enhancing AI clinical decision support trust: design workshop insights from general practitioners. Stud Health Technol Inform 2025 Aug 7;329:593-597. [doi: 10.3233/SHTI250909] [Medline: 40775927]
43. Sittig DF, Lakhani P, Singh H. Applying requisite imagination to safeguard electronic health record transitions. J Am Med Inform Assoc 2022 Apr 13;29(5):1014-1018. [doi: 10.1093/jamia/ocab291] [Medline: 35022741]
44. Pavuluri S, Sangal R, Sather J, Taylor RA. Balancing act: the complex role of artificial intelligence in addressing burnout and healthcare workforce dynamics. BMJ Health Care Inform 2024 Aug 24;31(1):e101120. [doi: 10.1136/bmjhci-2024-101120] [Medline: 39181545]
45. Moy S, Irannejad M, Manning SJ, et al. Patient perspectives on the use of artificial intelligence in health care: a scoping review. J Patient Cent Res Rev 2024;11(1):51-62. [doi: 10.17294/2330-0698.2029] [Medline: 38596349]
46. Carayon P, Wooldridge A, Hoonakker P, Hundt AS, Kelly MM. SEIPS 3.0: Human-centered design of the patient journey for patient safety. Appl Ergon 2020 Apr;84:103033. [doi: 10.1016/j.apergo.2019.103033] [Medline: 31987516]

Abbreviations

APP: advanced practice provider
CDSS: clinical decision support system
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
COREQ: Consolidated Criteria for Reporting Qualitative Research
CXR: chest X-ray
DECIDE-AI: Developmental and Exploratory Clinical Investigations of Decision Support Systems Driven by AI
ED: emergency department
EHR: electronic health record
EMR: electronic medical record
GRAMMS: Good Reporting of a Mixed Methods Study
ICU: intensive care unit
IRB: institutional review board
PRECISE: Precision Resuscitation With Crystalloids in Sepsis
SEIPS: Systems Engineering Initiative for Patient Safety
TAM: technology acceptance model
UTAUT: unified theory of acceptance and use of technology

Edited by A Kushniruk; submitted 17.Mar.2026; peer-reviewed by D Yoo, G Nzeigwe, M Chakit; revised version received 24.Aug.2026; accepted 08.Sep.2026; published 02.Oct.2026.

Please cite as:
Darla M, Miltz D, Chandnani K, Purkayastha S, Diehl JW, Bhavani SV
Clinician Trust and Human Factors in AI-Enabled Clinical Decision Support in Acute Care: Mixed Methods Study
JMIR Hum Factors 2026;13:e95472
URL: https://humanfactors.jmir.org/2026/1/e95472
doi:10.2196/95472

© Meghana Darla, Danielle Miltz, Khushboo Chandnani, Saptarshi Purkayastha, John W Diehl, Sivasubramanium V Bhavani. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 2.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits

XSL•FO
RenderX

unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Public Perceptions and Determinants of Preference for Robotic-Assisted Surgery in the United Arab Emirates: Cross-Sectional Study

Yasir Ahmed Mohammed Elhadi[1], MPH; Aminu S Abdullahi[1], MPH; Abdelrahman Mohamed Alblooshi[2], MD; Abdulla Alhosani[2], MD; Arwa Al Khalidi[1], MPH; Mohammed Saleem[2], MD; Shamma Rashed Mohammed Alrashdi[3], BSc; Reem Saleh Aljaberi[4], MSc; Azhar T Rahma[1], PhD

[1]Institute of Public Health, College of Medicine and Health Sciences, United Arab Emirates University, Tawam St, Al Ain, United Arab Emirates

[2]College of Medicine and Health Sciences, United Arab Emirates University, Al Ain, United Arab Emirates

[3]Department of Statistics and Data Analytics, College of Business and Economics, United Arab Emirates University, Al Ain, United Arab Emirates

[4]Department of Biological Sciences, College of Medicine and Health Sciences, Khalifa University, Abu Dhabi, United Arab Emirates

Corresponding Author:
Azhar T Rahma, PhD
Institute of Public Health, College of Medicine and Health Sciences, United Arab Emirates University, Tawam St, Al Ain, United Arab Emirates

Abstract

Background: Robotic-assisted surgery (RAS) is increasingly being integrated into surgical services worldwide. Successful implementation of robotic surgical technologies depends not only on clinical effectiveness but also on public trust, perceptions of safety, and willingness to use technology-enabled care. In the United Arab Emirates, the adoption of RAS is expanding rapidly; however, evidence regarding public perceptions and factors influencing acceptance remains limited.

Objective: This study aims to assess awareness, understanding, perceptions, and factors independently associated with preference for RAS among residents in the United Arab Emirates.

Methods: A community-based cross-sectional survey was conducted between July and November 2025 among adults aged 18 years or older residing across multiple emirates in the United Arab Emirates. Participants were recruited using a nonprobability convenience sampling approach through hospitals, shopping malls, community events, university mailing lists, and social media platforms. A bilingual Arabic-English questionnaire adapted from previously validated robotic surgery perception instruments was administered. Descriptive statistics summarized participant responses. Associations between participant characteristics and perceptions of RAS were examined using chi-square and Fisher exact tests. To adjust for potential confounding, a multivariable Firth-penalized logistic regression model was fitted to identify factors independently associated with preference for RAS over traditional surgeon-performed surgery. Statistical analyses were conducted using R (version 4.4.1).

Results: A total of 508 participants were included. Although 85% (n=432) had previously heard of RAS, only 38% (n=195) demonstrated an accurate functional understanding of robotic surgical systems. Overall, 33% (n=169) preferred RAS, while 35% (n=178) perceived robotic-assisted procedures as safe. The most frequently reported concerns were robotic malfunction (n=360, 71%), reduced human involvement (n=239, 47%), and procedural costs (n=183, 36%). Perceived benefits included improved surgical precision (n=331, 65%) and reduced complications (n=183, 36%). In multivariable analysis, participants who perceived RAS as safe had substantially higher odds of preferring a robot-assisted surgeon than those who were uncertain about its safety (adjusted odds ratio [aOR]=3.00, 95% CI 1.56 - 5.87; P<.001). Conversely, participants who perceived RAS as unsafe had substantially lower odds of preferring a robot-assisted surgeon than those who were uncertain about its safety (aOR=0.08, 95% CI 0.02 - 0.26; P<.001). No other factors were independently associated with surgery preference after adjustment.

Conclusions: Within this predominantly highly educated convenience sample of United Arab Emirates residents, awareness of RAS was high, but accurate understanding and acceptance remained limited. Perceived safety emerged as the factor shaping preference for RAS, highlighting the importance of trust and confidence in technology-enabled health care. These findings suggest that public education, transparent communication regarding surgeon oversight, and patient engagement strategies may be important components of the successful implementation of robotic surgical services in the United Arab Emirates.

(JMIR Hum Factors 2026;13:e97975) doi:10.2196/97975

XSL•FO
RenderX

KEYWORDS

robot-assisted surgery; United Arab Emirates; health services research; human-robot interaction; technology adoption; technology acceptance; surgery

Introduction

Robotic-assisted surgery (RAS) has emerged as a major advancement in minimally invasive surgical care, transforming operative practice across multiple specialties through enhanced precision, dexterity, visualization, and ergonomic support for surgeons [1,2]. Over the past 2 decades, the global adoption of robotic platforms has expanded rapidly in general surgery, urology, gynecology, cardiothoracic surgery, and neurosurgery, supported by technological innovations and the increasing integration of AI-enhanced capabilities into surgical systems [3,4]. Contemporary robotic systems increasingly incorporate digital interfaces, image-guided navigation, machine learning–supported decision assistance, and real-time data integration, positioning RAS as an important component of the broader digital transformation of health care [5-7]. Despite growing clinical adoption and evidence supporting several perioperative benefits, including reduced blood loss, shorter hospital stay, and improved surgical precision in selected procedures, public acceptance remains essential for the successful integration of robotic technologies into routine health care delivery [8,9].

Acceptance of emerging health care technologies is influenced not only by clinical effectiveness but also by human factors, including trust, perceived safety, transparency, perceived autonomy, and comfort with human-machine interaction [10-13]. The technology acceptance model and related frameworks emphasize that perceived usefulness, perceived ease of use, and trust strongly influence willingness to adopt novel technologies [14-16]. In the context of robotic surgery, concerns regarding machine autonomy, loss of human control, technical malfunction, depersonalization of care, and ethical implications may shape patient attitudes independently of objective clinical outcomes [17-19]. Public perceptions of AI-integrated technologies are also increasingly influenced by broader societal anxieties surrounding automation and decision-making by intelligent systems, particularly in high-risk settings such as health care [13,20-22]. Understanding these human-factor determinants is therefore critical for ensuring patient-centered implementation and equitable access to advanced surgical technologies.

Existing international evidence demonstrates substantial variability in public awareness, understanding, and acceptance of RAS across different populations and health care systems. Studies from the United States and Europe have reported moderate public awareness but persistent misconceptions regarding robotic autonomy and surgeon involvement during robotic procedures [13,23]. Surveys conducted in Saudi Arabia, Kuwait, and Jordan similarly identified limited functional understanding of RAS, concerns regarding safety and technical failure, and demographic differences in acceptance, particularly across age and gender groups [10,24-27]. A recent scoping review further highlighted that public perceptions of robotic surgery are commonly characterized by uncertainty, limited trust, and confusion regarding the role of the surgeon during robotic procedures [28]. Although several studies reported positive perceptions regarding precision and innovation, many participants remained reluctant to personally undergo robotic-assisted procedures due to concerns about safety, cost, and reduced human involvement [23,24,28].

In the United Arab Emirates, rapid investment in digital health infrastructure and advanced medical technologies has accelerated the implementation of robotic surgical systems within tertiary health care centers [29]. Robotic-assisted procedures are increasingly performed across major public and private hospitals, reflecting the country's broader health care innovation agenda and strategic emphasis on AI integration within health services. Nevertheless, evidence regarding public perceptions of RAS in the United Arab Emirates remains limited. Existing studies have primarily focused on health care professionals and medical students rather than the general public [30,31]. Earlier research examining public attitudes toward robotic surgery in the United Arab Emirates reported skepticism and limited awareness; however, this study was conducted more than a decade ago and may no longer reflect contemporary perceptions within the context of rapidly evolving health care technologies and increased national investment in digital medicine [32]. This study aimed to examine public awareness, understanding, perceptions, and preference for RAS among residents in the United Arab Emirates. Specifically, the study sought to (1) assess public awareness and functional understanding of RAS; (2) examine perceived benefits, risks, and trust-related concerns associated with RAS; and (3) identify demographic and experiential factors associated with preference for RAS and perceptions of safety.

Methods

Study Design and Setting

A community-based cross-sectional survey was conducted between July and November 2025 across multiple emirates in the United Arab Emirates, including Abu Dhabi, Dubai, Sharjah, Ajman, Ras Al-Khaimah, Fujairah, and Umm Al-Quwain. The study was reported in line with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) recommendations [33].

Participants and Sampling

Eligible participants were adults aged 18 years or older residing in the United Arab Emirates who were able to read either Arabic or English. Individuals professionally involved in robotics, AI, biomedical engineering, robotic systems development, or related technical fields were excluded to minimize expert-knowledge bias and ensure that responses reflected perceptions of the general public. A nonprobability convenience sampling approach was used. Participants were recruited through both in-person and online approaches to maximize demographic and geographic diversity. In-person recruitment was conducted in hospitals,

XSL•FO
RenderX

shopping malls, community centers, and public events across multiple emirates. The online survey was disseminated through university mailing lists and social media platforms, including WhatsApp, Instagram, LinkedIn, and X (formerly Twitter).

The optimal required sample size was estimated using the single population proportion formula:

$$n=Z2P(1-P)d2$$

Assuming a 50% expected prevalence, a 95% confidence level, a 5% margin of error, and a z score of 1.96. The calculated minimum sample size was 384 participants. To account for incomplete responses and potential nonresponse, the recruitment target was increased, and a final sample of 508 participants was included in the analysis.

Data Collection Tool

The bilingual (Arabic-English) questionnaire was adapted from previously validated RAS perception surveys conducted in the Gulf region and internationally [24,27,32]. The instrument underwent expert review by specialists in public health, digital health, and RAS to assess content validity, clarity, and cultural appropriateness for the UAE context. Minor modifications in wording and terminology were implemented following expert feedback to improve comprehensibility and contextual relevance.

Translation of the questionnaire followed the World Health Organization forward-backward translation methodology [34]. The English version was initially translated into Arabic by bilingual researchers and was then independently back-translated into English to ensure semantic equivalence and conceptual consistency between versions. Pilot testing was conducted among 20 bilingual UAE residents representing the target population to evaluate readability, comprehension, cultural appropriateness, and survey flow. Feedback obtained during pilot testing was used to refine question wording and improve usability. Pilot participants were excluded from the final analysis. The questionnaire included domains assessing demographic characteristics, technology experience, awareness and understanding of RAS, perceptions and attitudes toward RAS, and prior experiences with robotic surgery. The perception and attitude items demonstrated acceptable internal consistency within the study sample (Cronbach α=0.81).

Measures

The questionnaire consisted primarily of closed-ended items divided into five domains: (1) demographic characteristics, (2) technology experience, (3) awareness and understanding of RAS, (4) perceptions and attitudes toward RAS, and (5) patient-reported experiences among participants with prior exposure to robotic surgery. Demographic variables included age, gender, nationality, educational attainment, occupation, marital status, monthly income, health insurance status, chronic disease status, and medical travel history. Awareness and understanding of RAS were assessed using items evaluating prior exposure to robotic surgery, information sources, and functional understanding of robotic systems. Participants were classified as having "good understanding" if they correctly identified that robotic surgery involves a surgeon operating robotic arms from a control console. Perceptions and attitudes toward RAS included perceived safety, trust in robotic technologies, perceived benefits and risks, willingness to undergo RAS, concerns regarding robotic malfunction or loss of human control, and beliefs regarding the future role of robotic surgery in health care.

Ethical Considerations

This study was conducted in accordance with the research standards outlined in the Declaration of Helsinki. Ethical approval was obtained from the Social Sciences Research Ethics Committee at United Arab Emirates University (reference number: ERSC_2025_5861). All participants received an information sheet explaining the study objectives, the voluntary nature of participation, anonymity, and confidentiality measures. Electronic or written informed consent was obtained before participation. Participation was voluntary, and respondents were assured of their right to withdraw at any time without consequence. No identifiable personal information was collected. All data were collected anonymously and handled in accordance with ethical research standards. Participants were not financially compensated for participation.

Statistical Analysis

Missing data were minimal (<2% across all variables). Participants with missing data for variables included in a specific analysis were excluded only from that analysis. Data were analyzed using the R statistical software (version 4.4.1; R Foundation for Statistical Computing). Categorical variables were summarized using frequencies and percentages. Exact binomial 95% CIs were calculated for key proportions.

The variable assessing participants' understanding of RAS included 5 response options. Participants were classified as having a good understanding of RAS only if they correctly identified that robotic surgery involves a surgeon sitting at a control unit and operating robotic arms. All other responses, including "I don't know," were classified as limited understanding. For the ease of interpretation in inferential analyses, age was recategorized from 6 original categories (18 - 24, 25 - 34, 35 - 44, 45 - 54, 55 - 64, and ≥65 y) into 2 groups (18 - 44 y and ≥45 y).

Associations between participants' characteristics and 2 key outcomes, including preference for and perceived safety of RAS, were assessed using the chi-square test or Fisher exact test, as appropriate. Variables examined included demographic characteristics, health insurance status, chronic disease status, medical travel history, parental status, awareness of RAS, familiarity with individuals who had undergone robotic surgery, perceived availability of RAS in the UAE, sources of information about RAS, understanding of RAS, comfort with modern technology, and frequency of technology use. Statistical significance was defined as a 2-sided *P* value <.05. To adjust for potential confounding, a multivariable Firth-penalized logistic regression model was fitted to identify factors independently associated with preference for RAS. This method was used to address sparse data and potential quasi-complete separation [35]. The primary outcome was preference for RAS versus traditional surgeon, as participants with no preference were excluded from the primary analysis. Variables were selected a priori based on theoretical relevance and included

XSL•FO
RenderX

gender, age, nationality, monthly income, knowing someone who had undergone RAS, understanding of RAS, and perceived safety of RAS. Adjusted odds ratios (aORs) and 95% CIs were reported. To assess the robustness of the findings, a sensitivity analysis was performed by redefining the outcome as preference for a robot-assisted surgeon versus all other responses (traditional surgeon or no preference) using the same multivariable Firth-penalized logistic regression model.

Results

Participant Recruitment and Characteristics

A total of 508 participants were included in the final analysis. The study sample comprised 57% (n=291) female participants and 55% (n=282) Emiratis (Table 1). Most participants (n=328, 65%) were aged 18 - 44 years, while 35% (n=180) were aged 45 years or older. The majority (n=316, 62%) resided in the Abu Dhabi Emirate, held a bachelor's degree or higher qualification (n=402, 79%), were married (n=317, 62%), and had children (n=322, 63%). Approximately one-third of participants worked in health care–related fields, including 20% who identified as health care professionals. Regarding monthly income, 31% (n=157) reported earning more than 20,000 AED (United Arab Emirates dirham; 1 AED=0.27 USD as of August 4, 2026) per month, while 13% (n=64) reported earning less than 5000 AED per month. Chronic medical conditions were reported by 22% (n=113) of the participants, 29% (n=145) reported having health insurance coverage, and 16% (n=79) reported a history of medical travel.

XSL•FO
RenderX

Table . Participants' characteristics (N=508).

Characteristic	Participants, n (%)
Gender	
Female	291 (57.3)
Male	217 (42.7)
Age (y)	
18 - 24	102 (20.1)
25 - 34	103 (20.3)
35 - 44	123 (24.2)
45 - 54	94 (18.5)
55 - 64	72 (14.2)
≥65	14 (2.8)
Nationality	
Emirati	282 (55.5)
Non-Emirati	226 (44.5)
Residence	
Abu Dhabi	316 (62.2)
Ajman	16 (3.2)
Al-Fujairah	11 (2.2)
Dubai	92 (18.1)
Ras Al-Khaimah	12 (2.4)
Sharjah	59 (11.6)
Umm Al-Quwain	2 (0.3)
Education	
High school	68 (13.4)
Diploma	38 (7.5)
Bachelor's degree	270 (53.1)
Master's degree	132 (26)
Work field	
Business or administration	149 (29.3)
Education	58 (11.4)
Engineering	73 (14.4)
Medical or health care	160 (31.5)
Others	68 (13.4)
Marital status	
Divorced	22 (4.3)
Married	317 (62.4)
Single	167 (32.9)
Widow	2 (0.4)
Have children	322 (63.4)
Health care worker	100 (19.7)
Monthly income (AED[a])	
<5000	64 (12.6)
5000 - 9000	51 (10)

XSL•FO
RenderX

Characteristic	Participants, n (%)
10,000 - 20,000	100 (19.7)
>20,000	157 (30.9)
Prefer not to answer	136 (26.8)
Have a chronic condition	113 (22.2)
Medical travel (yes)	79 (15.6)
Have insurance coverage	145 (28.5)

[a]AED: United Arab Emirates dirham.

Perceived Safety of RAS

Overall, 35% of the participants (95% CI 31%-39%) perceived RAS as safe, whereas 22% (95% CI 19% - 26%) considered it unsafe, and 43% (95% CI 38% - 47%) reported uncertainty regarding its safety (Table S1 in Multimedia Appendix 1). Several participant characteristics were significantly associated with perceptions of safety. Male participants were more likely than female participants to perceive RAS as safe (50% vs 24%; *P*<.001). Participants aged 45 years or older were also more likely to perceive robotic surgery as safe compared with younger adults aged 18 - 44 years (42% vs 31%; *P*=.01). Emirati nationality, higher monthly income, and health care worker status were also associated with more favorable safety perceptions. Participants with health insurance coverage were substantially more likely to perceive RAS as safe than uninsured participants (51% vs 29%; *P*<.001). Participants reporting chronic medical conditions also demonstrated more positive safety perceptions compared with those without chronic conditions (41% vs 33%; *P*=.04).

Prior exposure and understanding strongly influenced safety perceptions. Participants who knew someone who had undergone robotic surgery were more likely to perceive RAS as safe than those without such exposure (66% vs 36%; *P*<.001). Similarly, participants who had previously heard of robotic surgery (38% vs 16%; *P*<.001), demonstrated accurate functional understanding of RAS (42% vs 31%; *P*=.02), or preferred robotic-assisted procedures over traditional surgery (58% vs 16%; *P*<.001) were significantly more likely to report positive perceptions regarding safety. Participants exposed to information regarding robotic surgery through television, newspapers, medical websites, medical conferences or seminars, health care professionals, or personal experience demonstrated significantly greater confidence in the safety of RAS than those without exposure through these channels. Educational attainment, having children, medical travel history, comfort with modern technology, frequency of technology use, and learning about robotic surgery through social media or university lectures were not significantly associated with perceived safety.

Perceptions Regarding the Future Role and Precision of RAS

Most participants expressed optimism regarding the future role of RAS within health care systems. Overall, 67% (n=340) believed that robotic surgery would eventually become a standard surgical approach, while 61% (n=310) perceived robotic systems to be more precise than human surgeons. Despite these favorable perceptions regarding technological advancement and precision, the willingness to personally undergo robotic-assisted procedures remained limited. Only 37% (n=188) reported that they would choose RAS if it represented the only available treatment option, whereas 31% (n=158) stated that they would still decline robotic-assisted procedures under such circumstances. Additional perception-related responses are presented in Table S2 in Multimedia Appendix 1.

Perceived Benefits and Concerns Regarding RAS

Participants reported both perceived benefits and concerns regarding RAS. The most frequently reported perceived benefit was improved surgical precision (n=331, 65%), followed by reduced complications and surgical risks (n=183, 36%). Additionally, 20% (n=102) believed RAS could reduce postoperative pain, while 20% (n=102) associated robotic surgery with faster postoperative recovery. Concerns regarding RAS were widespread, with 94% (n=478) of the participants reporting at least one concern related to robotic surgical systems. The most commonly reported concern was the possibility of a robotic malfunction during surgery (n=360, 71%). Additional concerns included reduced involvement of human surgical skills (n=239, 47%), high procedural costs (n=183, 36%), and ethical or technical concerns related to robotic technologies (n=133, 26%).

Perceived Suitability of RAS Across Surgical Specialties

Participants identified several surgical specialties they perceived as particularly suitable for RAS. General surgery was the most frequently identified specialty (n=218, 43%), followed by cardiac surgery (n=188, 37%) and neurosurgery (n=168, 33%). Additional specialties perceived as appropriate for robotic-assisted procedures included orthopedic surgery (n=152, 30%), urological surgery (n=147, 29%), and cardiothoracic surgery (n=97, 19%).

Preferences for RAS and Associated Factors

Among the 508 participants, 33% (95% CI 29% - 38%) preferred RAS, 50% (95% CI 45% - 54%) preferred traditional surgeon-performed procedures, and 17% (95% CI 14% - 21%) reported no preference (Table 2). The distribution of surgery preference differed significantly by gender (*P*<.001), age (*P*<.001), nationality (*P*<.001), monthly income (*P*=.002), having children (*P*=.023), health insurance coverage (*P*<.001), and perceived safety of RAS (P<.001). The proportion preferring RAS was higher among male than female participants (40% vs

XSL•FO
RenderX

29%), participants aged 45 years or older than those aged 18–44 years (46% vs 27%), Emirati than non-Emirati participants (39% vs 26%), participants with a monthly income of 20,000 AED or more than those with lower incomes (41% vs 28%), participants with children than those without children (37% vs 26%), and among participants with health insurance coverage than among those without coverage (42% vs 23%).

Prior awareness and exposure to robotic surgery were also significantly related to the distribution of surgery preference. The distribution differed significantly according to whether participants knew someone who had undergone RAS (*P*<.001), had previously heard of robotic surgery (*P*<.001), their functional understanding of RAS (P=.002), and whether they believed that RAS was available in the United Arab Emirates (*P*<.001). The proportion preferring RAS was higher among participants who knew someone who had undergone RAS than among those who did not (60% vs 32%), those who had previously heard of robotic surgery than among those who had not (36% vs 18%), those with good rather than poor functional understanding of RAS (35% vs 32%), and those who believed that RAS was available rather than unavailable in the United Arab Emirates (46% vs 27%). The distribution of surgery preference also differed significantly according to whether participants had learned about RAS through television (*P*=.027), social media (*P*=.009), or personal experience (*P*=.012). The corresponding proportions preferring RAS among participants with versus without exposure through these channels were 37% versus 32%, 37% versus 29%, and 57% versus 32%, respectively.

XSL•FO
RenderX

Table . Preferences for surgery and participants' characteristics[a].

Characteristic	Surgery preference			
	Robot-assisted surgeon (n=169)	Traditional surgeon (n=253)	No preference (n=86)	*P* value
Overall (95% CI)	33% (29% - 38%)	50% (45% - 54%)	17% (14% - 21%)	—[b]
Gender, n (%)				<.001
Female	83 (29)	171 (59)	37 (13)	
Male	86 (40)	82 (38)	49 (23)	
Age (y), n (%)				<.001
18 - 44	87 (27)	183 (56)	58 (18)	
≥45	82 (46)	70 (39)	28 (16)	
Nationality, n (%)				<.001
Emirati	110 (39)	115 (41)	57 (20)	
Non-Emirati	59 (26)	138 (61)	29 (13)	
Children, n (%)				.02
Yes	120 (37)	155 (48)	47 (15)	
No	49 (26)	98 (53)	39 (21)	
Monthly income (AED[c]), n (%)				.002
<20,000	61 (28)	125 (58)	29 (13)	
≥20,000	64 (41)	62 (39)	31 (20)	
Have insurance coverage, n (%)				<.001
Yes	61 (42)	49 (34)	35 (24)	
No	22 (23)	58 (62)	14 (15)	
Not sure	86 (32)	146 (54)	37 (14)	
Know someone who had RAS[d], n (%)				<.001
Yes	28 (60)	13 (28)	6 (13)	
No	114 (32)	180 (50)	67 (19)	
Heard of RAS, n (%)				<.001
Yes	155 (36)	199 (46)	78 (18)	
No	14 (18)	54 (71)	8 (11)	
Learned about RAS from television, n (%)				.03
Yes	53 (37)	59 (41)	32 (22)	
No	116 (32)	194 (53)	54 (15)	
Learned about RAS from social media, n (%)				.009
Yes	107 (37)	129 (44)	57 (19)	
No	62 (29)	124 (58)	29 (13)	
Learned about RAS from personal experience, n (%)				.01
Yes	13 (57)	5 (22)	5 (22)	
No	156 (32)	248 (51)	81 (17)	
Understanding of RAS, n (%)				.002
Good	68 (35)	81 (42)	46 (24)	
Limited	101 (32)	172 (55)	40 (13)	
Believes RAS is available in the United Arab Emirates, n (%)				<.001

XSL•FO
RenderX

Characteristic	Surgery preference			
	Robot-assisted surgeon (n=169)	Traditional surgeon (n=253)	No preference (n=86)	*P* value
Yes	91 (46)	69 (35)	38 (19)	
No	4 (27)	10 (67)	1 (6.7)	
Not sure	74 (25)	174 (59)	47 (16)	
Believes RAS is safe, n (%)				<.001
Yes	98 (56)	40 (22)	40 (22)	
No	6 (5.5)	103 (91)	4 (3.5)	
Not sure	65 (30)	110 (51)	42 (19)	

[a]Other factors explored but not found to be significantly correlated with surgery preferences include education, health care worker, having a chronic condition, medical travel, comfort level with using modern technology, frequency of technology use, other ways of learning about RAS— newspapers, medical websites, medical conferences or seminars, health care professionals, university or academic lectures, and other sources.

[b]Not applicable.

[c]AED: United Arab Emirates dirham.

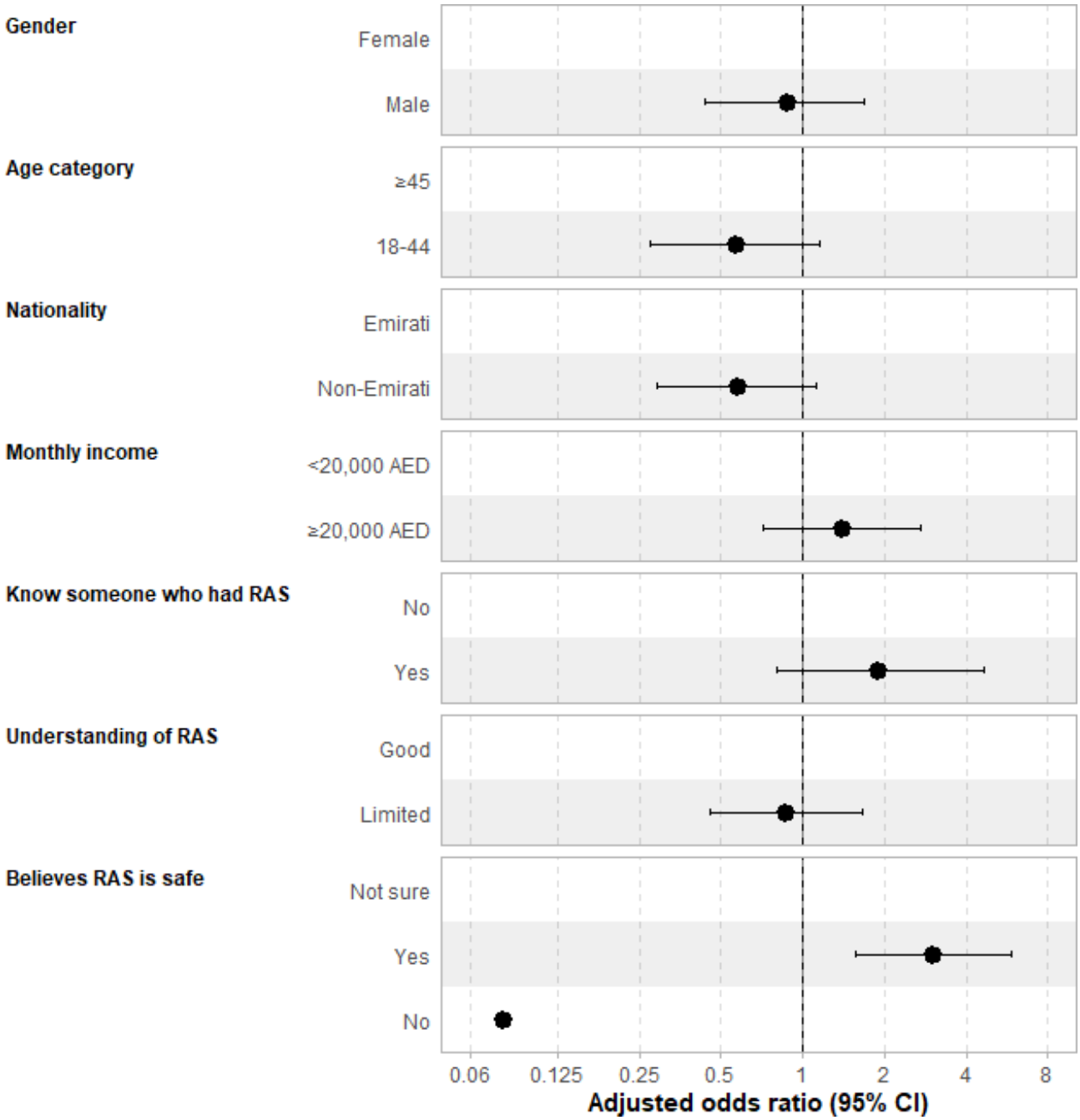
[d]RAS: robotic-assisted surgery.

In the adjusted analysis using Firth penalized logistic regression, perceived safety of RAS was the only factor independently associated with preference for a robot-assisted surgeon over a traditional surgeon (Figure 1). Participants who perceived RAS as safe had substantially higher odds of preferring a robot-assisted surgeon than those who were uncertain about its safety (aOR=3.00, 95% CI 1.56 - 5.87; *P*<.001). Conversely, participants who perceived RAS as unsafe had substantially lower odds of preferring a robot-assisted surgeon than those who were uncertain about its safety (aOR=0.08, 95% CI 0.02 - 0.26; *P*<.001). No other factors were independently associated with surgery preference after adjustment.

A sensitivity analysis redefining the outcome as preference for a robot-assisted surgeon versus all other responses (traditional surgeon or no preference) yielded materially similar findings (Table S3 in Multimedia Appendix 1). Perceived safety remained the only consistent independent factor associated with preference, although younger age (18 - 44 y) also reached statistical significance in the sensitivity analysis.

XSL•FO
RenderX

Figure 1. Multivariable logistic regression analysis of factors associated with preference for a robot-assisted surgeon over a traditional surgeon. Odds ratios are adjusted for all variables shown. Error bars represent 95% CIs. The dashed vertical line denotes the null value (odds ratio=1). RAS: robotic-assisted surgery.



Discussion

Principal Findings

This study provides insights into perceptions of RAS among a convenience sample of predominantly highly educated residents in the United Arab Emirates. Although most participants had previously heard of RAS, only a minority demonstrated an accurate functional understanding of surgeon control and robotic operation. In this sample, acceptance of robotic-assisted procedures was relatively limited: only one-third of the participants preferred RAS to conventional surgery, and approximately one-third perceived robotic surgery as safe. Concerns regarding robotic malfunction, reduced human involvement, and technical reliability were common. Nevertheless, many participants simultaneously recognized potential benefits of robotic surgery, particularly improved surgical precision and reduced complications, highlighting an important tension between perceived innovation and perceived risk. Importantly, perceived safety emerged as the strongest independent factor associated with preference for RAS. Participants who believed RAS was safe were substantially more likely to prefer robotic-assisted procedures, even after adjustment for demographic and experiential factors. These findings suggest that trust and confidence in the safety of robotic surgical technologies may play a more important role in acceptance than most demographic characteristics.

XSL•FO
RenderX

The findings of this study are broadly consistent with international literature demonstrating limited public understanding and cautious acceptance of RAS despite increasing global adoption of robotic technologies in health care [28,36,37]. RAS across multiple countries is frequently characterized by uncertainty, misconceptions regarding robotic autonomy, and limited trust in machine-assisted procedures [37]. Similar findings have been reported in studies from the United States, Europe, and Gulf countries, where awareness of robotic surgery may be moderate, but functional understanding remains poor [28,38-41].

In the present study, many participants expressed uncertainty about the surgeon's role during robotic procedures, with misconceptions about robotic autonomy and machine decision-making remaining common. These findings align with previous surveys reporting that large proportions of the public incorrectly believe that robotic systems operate independently of surgeons or possess autonomous surgical capabilities [23,25,26,42]. Such misconceptions likely reflect broader societal anxieties regarding AI-driven technologies and automation in health care settings [12,13]. These findings are consistent with the concept that acceptance of robotic surgery may be influenced not only by perceived clinical outcomes but also by trust, transparency, and understanding of human oversight within technologically mediated care [15,16].

Despite recognizing potential benefits of robotic surgery, participants in this study demonstrated relatively cautious willingness to personally undergo robotic-assisted procedures. Similar patterns have been observed internationally [17,24,36,40]. Previous studies have shown that patients may perceive robotic surgery as technologically advanced yet remain hesitant to accept it for their own treatment [17,36]. This apparent contradiction likely reflects underlying concerns regarding safety, reliability, and perceived loss of human control during surgery. In the present study, concerns regarding robotic malfunction represented the most frequently reported barrier to acceptance, consistent with previous studies identifying technical failure and loss of surgeon control as major public concerns surrounding RAS [17,18,23].

Participants in this study also perceived robotic surgery as associated with improved surgical precision and better clinical outcomes, including reduced complications, lower postoperative pain, and faster recovery. These findings align with previous international and regional studies reporting that the public commonly associates robotic-assisted procedures with enhanced precision and minimally invasive surgical benefits [25,28,42,43]. However, these perceived advantages were often offset by concerns about safety, ethical implications, procedural costs, and reduced human interaction in care. Similar concerns regarding depersonalization of care and diminished surgeon presence have been reported in qualitative studies examining patient perceptions of robotic-assisted procedures [38,44].

The present findings suggest that human factors play a central role in shaping public acceptance of RAS. Trust, perceived safety, understanding of surgeon oversight, and familiarity with robotic technologies were strongly associated with acceptance of RAS in bivariate analyses. Participants with prior exposure to robotic surgery, either through personal experience or knowing someone who had undergone robotic-assisted procedures, were substantially more likely to perceive RAS as safe and preferable to conventional surgery. This finding is consistent with previous evidence showing that familiarity and direct exposure are associated with lower uncertainty and greater trust in emerging medical technologies [28,36].

The study findings indicate that media exposure may help shape perceptions of RAS. Many participants reported learning about robotic surgery primarily through television, social media, internet sources, or word of mouth rather than directly from health care professionals. Similar patterns have been reported internationally, particularly in Middle Eastern populations [42,43]. Although media exposure may improve awareness of robotic technologies, inaccurate or sensationalized portrayals of robotic surgery may also contribute to fear, misunderstanding, and exaggerated perceptions of risk [45].

Public misconceptions regarding robotic autonomy and technical failure may therefore partly reflect the quality and framing of publicly available information regarding robotic surgery. Previous studies have demonstrated that targeted educational interventions and physician-led explanations can substantially improve patient understanding and confidence regarding robotic-assisted procedures [36,39].

Importantly, the multivariable analysis demonstrated that perceived safety was the only independent factor associated with preference for RAS. Participants who believed RAS was safe were substantially more likely to prefer robotic-assisted procedures than traditional surgeon-performed surgery, even after adjustment for demographic and experiential factors. In contrast, most demographic characteristics lost statistical significance after adjustment. These findings suggest that trust in the safety of robotic systems may be more influential than demographic characteristics in determining acceptance of robotic-assisted medical procedures.

From a health services research perspective, these findings have important implications for the implementation and adoption of technology-enabled surgical services. Successful integration of RAS into routine clinical practice may depend not only on technological capability and clinical effectiveness but also on addressing patient concerns regarding safety, reliability, and human oversight. Implementation strategies that strengthen public trust and improve understanding of surgeon control during robotic procedures may therefore play a critical role in promoting acceptance and uptake of robotic-assisted surgical services.

Limitations

This study has several limitations. First, the use of nonprobability convenience sampling may have introduced selection bias and limited the generalizability of the findings. Second, the sample included a disproportionately high proportion of participants with bachelor's degrees or higher qualifications and a substantial proportion working in health care–related fields, which may not reflect the demographic composition of the broader United Arab Emirates' population. Third, the cross-sectional design precludes causal inference

XSL•FO
RenderX

regarding factors associated with perceptions of RAS. Finally, all data were self-reported and may be subject to recall and social desirability biases. Therefore, the findings should be interpreted as reflecting perceptions within this study sample rather than the UAE population.

Acknowledgments

During manuscript preparation, the authors used ChatGPT (OpenAI) to assist with language editing, to improve readability, and to enhance the clarity and grammar of selected sections of the manuscript. The tool was not used for study conception, study design, data collection, statistical analysis, data interpretation, generation or verification of references, or formulation of the scientific conclusions. All AI-assisted text were critically reviewed, fact-checked, substantially revised where appropriate, and approved by the authors. The authors take full responsibility for the accuracy, integrity, and originality of the manuscript and have verified all references and scientific statements included in the final version.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: ATR and YAME
Data curation: ASA, AAK, MS, SRMA, RSA
Formal analysis: ASA and YAME
Investigation: ASA, AAK, MS, SRMA, RSA
Methodology: ATR, YAME, AA
Project administration: ATR
Validation: ATR and YAME
Visualization: YAME
Writing – original draft: YAME
Writing – review and editing: YAME, AA, ASA, AAK, MS, SRMA, RSA, ATR
All authors approved the manuscript

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File, 30 KB - humanfactors_v13i1e97975_app1.docx]

References

1. Wah JNK. The rise of robotics and AI-assisted surgery in modern healthcare. J Robot Surg 2025 Jun 20;19(1):311. [doi: 10.1007/s11701-025-02485-0] [Medline: 40540146]
2. Morris B. Robotic surgery: applications, limitations, and impact on surgical education. MedGenMed 2005 Sep 27;7(3):72. [Medline: 16369298]
3. Fairag M, Almahdi RH, Siddiqi AA, et al. Robotic revolution in surgery: diverse applications across specialties and future prospects review article. Cureus 2024 Jan;16(1):e52148. [doi: 10.7759/cureus.52148] [Medline: 38344598]
4. Yeung AWK, Torkamani A, Butte AJ, et al. The promise of digital healthcare technologies. Front Public Health 2023;11:1196596. [doi: 10.3389/fpubh.2023.1196596] [Medline: 37822534]
5. Rojas Burbano JC, Ruiz NI, Rojas Burbano GA, Guacho Inca JS, Arragan Lezama CA, González MS. Robot-assisted surgery: current applications and future trends in general surgery. Cureus 2025 Apr;17(4):e82318. [doi: 10.7759/cureus.82318] [Medline: 40376327]
6. Karunanayake N. Next-generation agentic AI for transforming healthcare. Inform Health 2025 Sep;2(2):73-83. [doi: 10.1016/j.infoh.2025.03.001]
7. O P. Robotic surgery powered by AI: precision and automation in the operating room. SunText Rev Med Clin Res 2025;6(2). [doi: 10.51737/2766-4813.2025.125]
8. Handa A, Gaidhane A, Choudhari SG. Role of robotic-assisted surgery in public health: its advantages and challenges. Cureus 2024 Jun;16(6):e62958. [doi: 10.7759/cureus.62958] [Medline: 39050344]

XSL•FO
RenderX

9. Riad A, Hadid M, Elomri A, et al. Advancements and challenges in robotic surgery: a holistic examination of operational dynamics and future directions. Surg Pract Sci 2025 Sep;22:100294. [doi: 10.1016/j.sipas.2025.100294] [Medline: 40697312]
10. McDermott H, Choudhury N, Lewin-Runacres M, Aemn I, Moss E. Gender differences in understanding and acceptance of robot-assisted surgery. J Robot Surg 2020 Feb;14(1):227-232. [doi: 10.1007/s11701-019-00960-z] [Medline: 31049775]
11. Otto L, Schlieter H, Harst L, Whitehouse D, Maeder A. The telemedicine community readiness model-successful telemedicine implementation and scale-up. Front Digit Health 2023;5:1057347. [doi: 10.3389/fdgth.2023.1057347] [Medline: 36910573]
12. Fuller P, Kennedy S, Ball M, et al. Understanding the challenges of robotic-assisted surgery adoption: perspectives from stakeholders and the general population on human-interaction, built environment, and training. Appl Ergon 2025 Jan;122:104403. [doi: 10.1016/j.apergo.2024.104403] [Medline: 39418740]
13. Hölgyesi Á, Zrubka Z, Gulácsi L, et al. Robot-assisted surgery and artificial intelligence-based tumour diagnostics: social preferences with a representative cross-sectional survey. BMC Med Inform Decis Mak 2024 Mar 27;24(1):87. [doi: 10.1186/s12911-024-02470-x] [Medline: 38553703]
14. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172. [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
15. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003 Sep 1;27(3):425-478. [doi: 10.2307/30036540]
16. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep 1;13(3):319-340. [doi: 10.2307/249008]
17. Szabó B, Őrsi B, Csukonyi C. Robots for surgeons? Surgeons for robots? Exploring the acceptance of robotic surgery in the light of attitudes and trust in robots. BMC Psychol 2024 Jan 24;12(1):45. [doi: 10.1186/s40359-024-01529-8] [Medline: 38267998]
18. Brar G, Xu S, Anwar M, Talajia K, Ramesh N, Arshad SR. Robotic surgery: public perceptions and current misconceptions. J Robot Surg 2024 Feb 22;18(1):84. [doi: 10.1007/s11701-024-01837-6] [Medline: 38386115]
19. Wong SW, Wong AL. Ergonomic challenges and barriers of robotic surgery—a review. Ann Laparosc Endosc Surg 2025;10:28-28. [doi: 10.21037/ales-25-10]
20. Kim JJH, Soh J, Kadkol S, et al. AI Anxiety: a comprehensive analysis of psychological factors and interventions. AI Ethics 2025 Aug;5(4):3993-4009. [doi: 10.1007/s43681-025-00686-9]
21. Rahman MM, Babiker A, Yankouskaya A, AlShakhsi S, Ali R. Revisiting the interplay of risks, expected benefits, and ethical implications of AI across varying automation and criticality levels in Arab and UK contexts. J Responsible Technol 2026 Jun;26:100163. [doi: 10.1016/j.jrt.2026.100163]
22. Brauner P, Glawe F, Liehner GL, Vervier L, Ziefle M. Mapping public perception of artificial intelligence: expectations, risk–benefit tradeoffs, and value as determinants for societal acceptance. Technol Forecast Soc Change 2025 Nov;220:124304. [doi: 10.1016/j.techfore.2025.124304]
23. Boys JA, Alicuben ET, DeMeester MJ, et al. Public perceptions on robotic surgery, hospitals with robots, and surgeons that use them. Surg Endosc 2016 Apr;30(4):1310-1316. [doi: 10.1007/s00464-015-4368-6] [Medline: 26173543]
24. Buabbas AJ, Aldousari S, Shehab AA. An exploratory study of public' awareness about robotics-assisted surgery in Kuwait. BMC Med Inform Decis Mak 2020 Jul 1;20(1):140. [doi: 10.1186/s12911-020-01167-1] [Medline: 32611407]
25. Aldousari SA, Buabbas AJ, Yaiesh SM, Alyousef RJ, Alenezi AN. Multiple perceptions of robotic-assisted surgery among surgeons and patients: a cross-sectional study. J Robot Surg 2021 Aug;15(4):529-538. [doi: 10.1007/s11701-020-01136-w] [Medline: 32776285]
26. de Andres-Sanchez J, Almahameed AA, Arias-Oliva M, Pelegrin-Borondo J. Correlational and configurational analysis of factors influencing potential patients' attitudes toward surgical robots: a study in the Jordan University community. Mathematics 2022;10(22):4319. [doi: 10.3390/math10224319]
27. AlNaim AA, AlNaim N, Al Nasser F, Albash L, Almulhim M, Albinsaad L. Public's perception and knowledge of using robotics in general surgery in Eastern Region, Saudi Arabia. Ann Afr Med 2024 Jul 1;23(3):299-306. [doi: 10.4103/aam.aam_182_23] [Medline: 39034550]
28. McDonnell C, Devine M, Kavanagh D. The general public's perception of robotic surgery—a scoping review. Surgeon 2025 Apr;23(2):e49-e62. [doi: 10.1016/j.surge.2024.12.001] [Medline: 39658498]
29. Kumar A. Arab health 2025: why robot-assisted surgeries are rising in UAE. Gulf News. 2025. URL: https://gulfnews.com/uae/science/dubai-20-surgeries-are-robotic-assisted-says-top-experts-1.500024391 [accessed 2025-07-22]
30. Barkati N, Ntefeh N, Okasha A, Takshe AA, ElKhatib R, Chelli S. Robotic assisted surgery in the United Arab Emirates: healthcare experts' perceptions. J Robot Surg 2023 Dec;17(6):2799-2806. [doi: 10.1007/s11701-023-01716-6] [Medline: 37733210]
31. Mousa AH, Nukaly HY, Abujamai JZ, et al. Robotic surgery perception and attitude among medical students in the gulf countries: a multinational cross-sectional study. Research Square. Preprint posted online on Apr 12, 2023. [doi: 10.21203/rs.3.rs-2793219/v1]
32. Zineddine M, Arafa N. Attitude towards robot assisted surgery: UAE context. 9th Int Conf Innov Inf Technol 2013:175-179. [doi: 10.1109/Innovations.2013.6544414]

XSL•FO
RenderX

33. von Elm E, Altman DG, Egger M, et al. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007 Oct 20;335(7624):806-808. [doi: 10.1136/bmj.39335.541782.AD] [Medline: 17947786]
34. WHODAS 2.0 translation package: translation and linguistic evaluation protocol and supporting material. : World Health Organization; 2010 URL: https://terrance.who.int/mediacentre/data/WHODAS/Guidelines/WHODAS%202.0%20Translation%20guidelines.pdf. [accessed 2026-08-22]
35. Firth D. Bias reduction of maximum likelihood estimates. Biometrika 1993;80(1):27-38. [doi: 10.1093/biomet/80.1.27]
36. Chu CM, Agrawal A, Mazloomdoost D, et al. Patients' knowledge of and attitude toward robotic surgery for pelvic organ prolapse. Female Pelvic Med Reconstr Surg 2019;25(4):279-283. [doi: 10.1097/SPV.0000000000000556] [Medline: 29369840]
37. Stai B, Heller N, McSweeney S, et al. Public perceptions of artificial intelligence and robotics in medicine. J Endourol 2020 Oct;34(10):1041-1048. [doi: 10.1089/end.2020.0137] [Medline: 32611217]
38. Herling SF, Palle C, Moeller AM, Thomsen T. The experience of robotic-assisted laparoscopic hysterectomy for women treated for early-stage endometrial cancer: a qualitative study. Cancer Nurs 2016;39(2):125-133. [doi: 10.1097/NCC.0000000000000260] [Medline: 25881811]
39. Wongkietkachorn A, Boonyawong P, Rhunsiri P, Tantiphlachiva K. Hands-on surgical training workshop: an active role-playing patient education for adolescents. J Cancer Educ 2017 Sep;32(3):543-548. [doi: 10.1007/s13187-016-0988-3] [Medline: 26792783]
40. Lavelle M, Haziza S, Huntley S, Conte B, D'Apuzzo M, Hernandez VH. Patients' perception of technology: an update of patients' understanding of robotics and navigation in total joint arthroplasty. Muscles Ligaments Tendons J 2022;12(3). [doi: 10.32098/mltj.03.2022.15]
41. Abdelaal MS, Wiafe BM, Khan IA, et al. Robotic-assisted total knee arthroplasty: what are patients' perspectives, understanding and expectations? J Arthroplasty 2023 Sep;38(9):1726-1733. [doi: 10.1016/j.arth.2023.03.020] [Medline: 36924858]
42. Al Dihan FA, Alghamdi MA, Aldihan FA, Alamer NM, Alshahrani FA, Alqarni A. Knowledge, attitude, awareness, and future expectations of robotic surgery in patients attending surgical specialties clinics. Cureus 2024 Mar;16(3):e56523. [doi: 10.7759/cureus.56523] [Medline: 38646294]
43. Khalil SS, Altinbas SK, Desai K, Sirota I, Silverman HJ, Nezhat FR. Patient understanding, attitudes, and perceptions of laparoscopic and robotic-assisted laparoscopic gynecologic procedures. Obstet Gynecol 2015;125(Supplement 1):91S. [doi: 10.1097/AOG.0000000000000816]
44. Ammer E, Mandt LS, Silbersdorff IC, Kahl F, Hagmayer Y. Robotic anxiety-parents' perception of robot-assisted pediatric surgery. Children (Basel) 2022 Mar 11;9(3):399. [doi: 10.3390/children9030399] [Medline: 35327771]
45. Ficko Z, Koo K, Hyams ES. High tech or high risk? An analysis of media reports about robotic surgery. J Robot Surg 2017 Jun;11(2):211-216. [doi: 10.1007/s11701-016-0647-z] [Medline: 27778227]

Abbreviations

AED: United Arab Emirates dirham
aOR: adjusted odds ratio
RAS: robotic-assisted surgery
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by S Law; submitted 11.Apr.2026; peer-reviewed by B Zmuk, N Kamil; revised version received 30.Jun.2026; accepted 30.Jun.2026; published 11.Sep.2026.

Please cite as:
Elhadi YAM, Abdullahi AS, Alblooshi AM, Alhosani A, Al Khalidi A, Saleem M, Alrashdi SRM, Aljaberi RS, Rahma AT
Public Perceptions and Determinants of Preference for Robotic-Assisted Surgery in the United Arab Emirates: Cross-Sectional Study
JMIR Hum Factors 2026;13:e97975
URL: https://humanfactors.jmir.org/2026/1/e97975
doi:10.2196/97975

© Yasir Ahmed Mohammed Elhadi, Aminu S Abdullahi, Abdelrahman Mohamed Alblooshi, Abdulla Alhosani, Arwa Al Khalidi, Mohammed Saleem, Shamma Rashed Mohammed Alrashdi, Reem Saleh Aljaberi, Azhar T Rahma. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 11.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The

XSL•FO
RenderX

complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Clinicians' Trust in AI-Based Clinical Recommendations Across Controlled Primary Care–Style Clinical Vignettes: Nurse-Dominant Experimental Pilot Study

Avishek Choudhury[1*], MS, PhD; Yeganeh Shahsavar[2*], PhD; Ayse P Gurses[2,3], MS, MPH, PhD

[1]Industrial and Management Systems Engineering, Benjamin M Statler College of Engineering and Mineral Resources, West Virginia University, Morgantown, WV, United States

[2]Armstrong Institute Center for Health Care Human Factors, Johns Hopkins Medicine, Johns Hopkins University, Baltimore, MD, United States

[3]Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

[*]these authors contributed equally

Corresponding Author:
Avishek Choudhury, MS, PhD
Industrial and Management Systems Engineering
Benjamin M Statler College of Engineering and Mineral Resources
West Virginia University
321 Engineering Sciences Bldg
1306 Evansdale Drive
Morgantown, WV, 26506
United States
Phone: 1 3042939431
Email: avishek.choudhury@mail.wvu.edu

Abstract

Background: Safe integration of AI-enabled clinical decision support requires understanding whether users' trust and behavioral reliance are appropriately calibrated to recommendation quality.

Objective: This pilot study examined a 3-item vignette-level trust construct and accept or reject behavior across sequential primary care–style clinical vignettes.

Methods: A total of 68 health care professionals, including 59 (86.8%) registered nurses and 9 (13.2%) physicians, each evaluated 21 clinical vignettes from 1 of 2 series. Each series contained 10 correct and 11 intentionally incorrect AI recommendations. Participants accepted or rejected each recommendation and responded to 3 questionnaire items measuring trust, perceived transparency, and likelihood of acting on the recommendation before receiving correctness feedback and points. The unrounded item mean formed the trust composite. In the behavioral analysis, 50 unrecorded accept or reject responses were classified as rejections. A pooled cross-classified linear mixed-effects model was constructed to assess associations between the trust composite, recommendation correctness, vignette position, and participant and vignette characteristics. It included random intercepts for participant and vignette item, with case series included as a nuisance adjustment.

Results: Participants accepted 374 of 748 (50%) incorrect recommendations and rejected 112 of 680 (16.47%) correct recommendations. Incorrect recommendations received lower prefeedback trust than correct recommendations (β=−0.772, 95% CI −1.036 to −0.507; standardized effect=−0.419). Trust declined modestly across vignette positions (β=−0.027, 95% CI −0.048 to −0.005; standardized effect=−0.088), with a decline in case series A but not case series B. Baseline intention to use AI was positively associated with trust (β=0.391, 95% CI 0.110-0.672; standardized effect=0.243), whereas higher perceived diagnostic difficulty was negatively associated with trust (β=−0.474, 95% CI −0.648 to −0.301; standardized effect=−0.257). In the exploratory lagged model, trust in the preceding recommendation was associated with trust in the current recommendation (β=0.266, 95% CI 0.210-0.322; standardized effect=0.264).

Conclusions: Self-reported trust differentiated correct from incorrect recommendations in aggregate, while acceptance and rejection were not fully aligned with recommendation correctness. These preliminary findings identify a potential evaluation problem: lower trust in incorrect advice does not necessarily imply that users will reject it. The findings do not establish real-world clinical effects.

XSL•FO
RenderX

(JMIR Hum Factors 2026;13:e97649) doi:10.2196/97649

KEYWORDS

artificial intelligence; clinical decision support systems; trust composite; human-AI interaction; trust calibration; behavioral reliance; registered nurses; clinical vignettes; perceived difficulty; repeated interaction

Introduction

AI-enabled clinical decision support systems (CDSS) have the potential to improve diagnostic and treatment workflows across health care settings by augmenting clinicians' productivity, particularly in environments characterized by workload and uncertainty [1-3]. Yet, the adoption of AI at the bedside has been slower than expected [4]. One major reason for this slow uptake is clinicians' trust in AI systems [5]. Clinicians must feel confident that AI recommendations are reliable and safe before they are willing to incorporate them into their clinical workflow. However, integrating AI into health care without fully understanding how clinicians interact with these systems and how their trust develops over time can create risks for patient safety [6,7].

AI performance may fluctuate depending on the clinical scenario, data quality, or system limitations [8]. If clinicians overtrust AI recommendations, they may accept incorrect advice, while undertrust may lead them to ignore useful guidance. Both situations can negatively affect patient care. Therefore, to safely and effectively integrate AI into clinical practice, it is essential to understand how clinicians develop, adjust, and act on trust in AI systems, as well as the factors that influence this trust [6,9]. In many ways, understanding these trust dynamics is just as important as improving the technical performance of AI algorithms themselves [10].

In clinical environments, which are high-paced and high-stakes work systems, maintaining appropriate trust in AI systems is critical. Health care workers (nurses and physicians) often need to make rapid decisions while managing multiple sources of information, and in such settings, clinicians must quickly determine when to accept, question, or override AI recommendations. While they commonly adopt a trust-but-verify approach when working with residents or colleagues, AI systems differ in that they lack shared accountability and interactive clarification. Trust in AI can be confounded with a user's self-confidence [11,12]. If a user lacks self-confidence, they may report lower trust in the AI not because the AI is unreliable, but because they doubt their own capacity to use it safely. Distinguishing self-confidence from trust is necessary to identify the right remedies (improve model explanations vs improve user training). Additionally, positive past performance may gradually increase trust in AI recommendations [13]. At the same time, as trust grows, users may become less likely to critically scrutinize AI outputs, allowing erroneous recommendations to pass through the decision-making process without sufficient evaluation. Conversely, negative past experiences with AI may make users more skeptical of the system, reducing their willingness to rely on AI recommendations even when they are accurate [5,7,14-16]. These dynamics highlight the importance of understanding how nurses' and physicians' trust evolve through repeated interactions with AI systems. Although AI decision support tools often perform with high accuracy, infrequent errors can hinder trust in the technology.

Real-world clinical work is demanding, and trust in automation is known to vary with workload, uncertainty, and perceived task difficulty [17-19]. A growing literature has examined global attitudes toward AI in health care, yet far fewer studies have explored how trust in AI can change in a clinical setting [10,20]. Existing work has largely focused on isolated psychological or demographic predictors such as prior trust, confidence, or experience without simultaneously examining how these individual characteristics interact with contextual pressures such as time constraints and perceived diagnostic complexity [21,22].

Building on trust calibration and human-automation reliance frameworks, we conceptualized trust as a dynamic evaluation that should ideally correspond to the perceived trustworthiness and observed reliability of the AI system. Foundational trust-in-automation theory argues that trust guides reliance when users cannot fully inspect or understand an automated system, and that safe human-automation interaction requires "appropriate reliance" rather than either blind acceptance or blanket rejection of automation [23,24]. Automation-reliance research further distinguishes subjective trust from observable patterns of use, misuse, disuse, and overreliance, emphasizing that users may behave in ways that do not perfectly match their stated trust. This distinction is especially important in clinical decision support, where automation bias can occur when clinicians overrely on CDSS recommendations and reduce independent information seeking or verification [25]. Recent health care AI studies and reviews similarly emphasize that clinicians' trust in AI-CDSS is shaped by clinical reliability, transparency, validation, usability, prior experience, and professional expertise, and that trust should be examined separately from behavioral reliance on AI advice [16,26,27]. Therefore, appropriate reliance occurs when clinicians accept useful or correct AI recommendations and question or reject recommendations that are unsafe, incorrect, or insufficiently supported; overreliance and underreliance represent mismatches between AI reliability and clinician behavior. In repeated AI-assisted decisions, trust may be influenced by dispositional attitudes toward AI, the visible plausibility of each recommendation, feedback from prior clinical vignettes, and task conditions such as difficulty and time constraints [13]. Therefore, this pilot study distinguished reported trust-related evaluation from behavioral acceptance of AI recommendations.

This experimental pilot study had 3 aims. The primary aim was to characterize how the vignette-level trust composite changed across 21 sequential AI-assisted clinical vignettes and how it was associated with recommendation correctness, vignette position, baseline AI attitudes, participant characteristics, and perceived diagnostic difficulty. The second aim was to describe whether behavioral reliance, measured by accepting or rejecting

XSL•FO
RenderX

recommendations, aligned with programmed correctness. The exploratory aim was to examine the association between trust in consecutive recommendations.

Methods

Study Design

This online experimental pilot study used repeated observations across 2 fixed series of 21 clinical vignettes. For each clinical vignette, participants reviewed a fictional patient presentation and a large language model (LLM)–generated diagnosis and recommendation, indicated whether they accepted or rejected the recommendation, and rated their trust. The same response process was repeated sequentially for all 21 clinical vignettes in the assigned series. No real patient encounters or patient outcomes were evaluated.

Clinical Vignettes

Before generating any study materials, the authors provided Current Medical Diagnosis and Treatment 2010 [28] to ChatGPT (GPT-4; OpenAI) as reference material. ChatGPT was then used to draft short, fictional clinical vignettes and their associated diagnoses and recommendations. The clinical vignettes concerned common adult presentations intended to be understandable without subspecialty knowledge. They were not drawn from real patient records. The generation tasks involved creating the patient descriptions and associated reference diagnoses and recommendations, with intentionally incorrect participant-facing alternatives included where required by the experimental design.

The authors reviewed the generated materials before expert physician review. They compared the initial diagnoses with the textbook and checked the clinical vignettes and recommendations for readability, completeness, clinical coherence, and internal consistency. A physician subsequently conducted the expert physician review, assessing the plausibility of each presentation, the compatibility of the proposed diagnosis with the clinical information, and the consistency of each participant-facing recommendation with its intended correct or incorrect status. She finalized the reference diagnoses and approved the materials for use in the experiment. No corrections to the initial reference diagnoses were required during either the authors' textbook-based review or the physician's review.

The absence of diagnosis-level corrections should be distinguished from the deliberately incorrect recommendations used in the experiment. Reference diagnoses established what was considered correct for each clinical vignette; participant-facing recommendations could intentionally differ from those diagnoses to implement the planned correctness sequence. These incorrect recommendations were experimental stimuli, not diagnostic errors left unaddressed during review.

Multimedia Appendix 1 describes the production and review procedure and provides prompt templates for reference-guided vignette generation, diagnosis and recommendation generation, and intentionally incorrect alternatives.

Participant Recruitment

Participants were recruited between August and September 2025 through Centiment, a paid online audience-panel service [29]. The survey was distributed to the provider's health care worker panel. Eligibility was based on self-reported professional designation as a registered nurse (RN) or physician and reported clinical experience. RNs were included because they are important users of digital clinical information and decision support and routinely contribute to assessment, monitoring, triage, escalation, and care coordination [21,22]. The clinical-vignette task examined evaluation of AI recommendations and reliance behavior; it was not intended to test participants' independent diagnostic or prescribing competence.

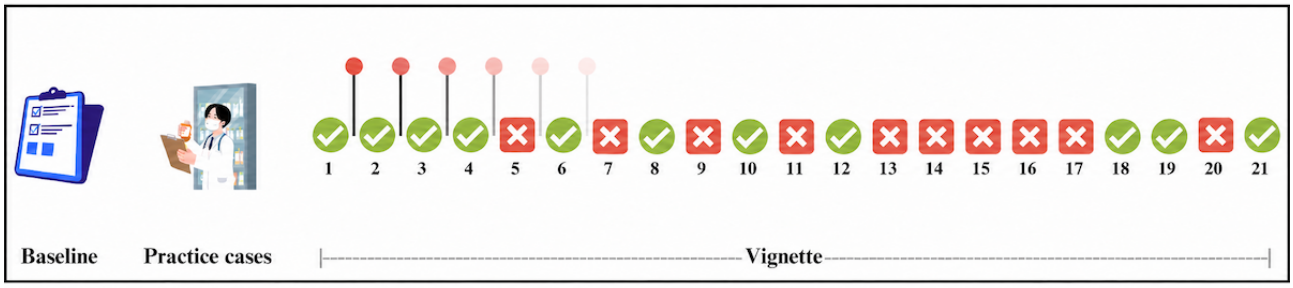
No independent credential, licensure, employer, or National Provider Identifier verification was performed. Centiment applies proprietary quality procedures, including duplicate and automated-response detection, device or browser fingerprinting, ReCAPTCHA, and fraud scoring. These procedures were performed by the panel provider; the research team did not receive IP addresses or device identifiers. The number invited, prescreened, or excluded by the provider and the numerical fraud-score threshold were not available to the research team.

Study Procedures

The 42 clinical vignettes were allocated to two 21-vignette series of broadly comparable scope and expected difficulty, without formal matching on independent prestudy difficulty ratings. Each series included 10 correct and 11 intentionally incorrect recommendations, giving a programmed accuracy rate of 47.62%. Both began with 4 correct recommendations followed by 2 incorrect recommendations; later correctness positions differed. Figure 1 shows the series-specific sequences. Check marks (✓) indicate recommendations classified as correct; cross marks (x) indicate intentionally incorrect recommendations. The two case series shared the same status through Trial 8 but differed at selected later trial positions. Accept or reject decisions and ratings were submitted before correctness feedback, and points were displayed for each trial. Participants rated the perceived difficulty of every clinical vignette so that variation in subjective difficulty could be described and included in the analysis.

XSL•FO
RenderX

Figure 1. Actual programmed sequence of AI recommendation correctness across the 21 analyzed cases.



Participants completed the study in one Qualtrics session and were randomly assigned to case series A or B. The session included electronic consent, task instructions, baseline questions, a demonstration vignette, and the 21 analyzed clinical vignettes. A programmed attention-check screen appeared between the fifth and sixth analyzed clinical vignettes. The demonstration vignette and attention-check screen were excluded from the analyses. Case series A had no fixed response window; case series B used a 45-second response window.

On each clinical-vignette screen, participants reviewed the patient presentation and AI recommendation, indicated whether they accepted or rejected the recommendation, and responded to 3 trust questionnaire items and a perceived-difficulty question. The accept or reject item preceded the trust questions to prioritize eliciting behavioral reliance before explicitly asking participants to reflect on trust. All responses appeared on the same screen and were submitted before correctness feedback or points were displayed.

After submission, participants received feedback about the programmed correctness of the AI recommendation. They received +10 points for accepting a correct recommendation or rejecting an incorrect recommendation and −10 points for the opposite decisions. Feedback and scores from preceding clinical vignettes were intentionally available to inform responses to subsequent recommendations. In case series B, the page advanced when the 45-second response window expired; the analytic handling of unrecorded decisions is specified below under the “Analysis” section.

Outcome Measurement

The primary outcome was vignette-level trust in the AI recommendation. Trust was operationalized through 3 investigator-developed questionnaire items addressing perceived trust, perceived transparency, and likelihood of acting on the recommendation. These items were specified as indicators of the study’s trust construct rather than as 3 separate primary outcomes; their unrounded mean was the trust composite. The conceptual framing drew on trust-in-automation literature [23,24]. The observed accept or reject response was a separate behavioral-reliance outcome. Baseline measures are summarized in Table 1, and the clinical-vignette measures are given in Table 2.

XSL•FO
RenderX

Table 1. Baseline participant measures.

Construct	Survey items	Scale or scoring
Professional designation	What is your professional or academic background?	• Registered nurse • Physician
Clinical experience	How many years of experience do you have in your current field or line of work?	• <1 year • 1-2 years • 3-5 years • 6-10 years • >10 years
Prior AI use	In the last 6 months, have you ever used AI in medical practice?	• Yes • No
Baseline trust in AI (3-item composite)	General trust in AI, perceived transparency of AI recommendations, and likelihood of acting on an AI recommendation.	• Each item: 1=none at all to 7=completely • Composite=unrounded mean of available items (minimum 2 of 3)
Intent to use AI	I would like to use AI in my clinical practice.	• 1=strongly disagree to 5=strongly agree
Self-confidence	Nine retained items: I handle new situations with relative comfort and ease; I feel positive and energized about life; I keep trying even after others have given up; If I work hard to solve a problem, I will find the answer; I achieve the goals I set for myself; people give me positive feedback on my work and achievements; when I overcome an obstacle, I think about the lessons I have learned; I believe that if I work hard, I will achieve my goals; I have contact with people with similar skills and experience whom I consider successful.	• Each item: 1=not at all to 5=very often • Mean of 9 retained items

Table 2. Clinical-vignette measures completed before correctness feedback.

Construct	Survey item	Scale
Perceived diagnostic difficulty	• How difficult do you think it would be to diagnose this patient case based on the information provided without AI assistance?	• 1=very easy to 5=very difficult
Vignette-level trust in the AI recommendation (3-item composite)	• How much do you trust the AI-based recommendation? • How transparent do you find the AI-based recommendation? • How likely are you to act upon the AI-based recommendation?	• Each item: 1=none at all to 7=completely • Composite=unrounded mean of the 3 items

Separate 1-factor confirmatory factor analysis (CFA) summaries based on polychoric correlations were aligned to the 21 analyzed clinical-vignette positions. Average variance extracted (AVE) ranged from 0.76 to 0.92, coefficient omega from 0.90 to 0.97, and Cronbach α from 0.87 to 0.96. The 3 items showed strong internal consistency; these results do not make the self-reported composite equivalent to observed acceptance or rejection. Vignette-specific factor loadings and reliability estimates are reported in Multimedia Appendix 2.

Baseline trust was calculated as the unrounded mean of the corresponding 3 baseline items, using at least 2 available responses, as specified in Table 1. Cronbach α was 0.938 among participants with complete baseline items. Self-confidence was the unrounded mean of the 9 retained 1-5 items in Table 1, informed by self-efficacy theory [30]. The supplied self-confidence CFA reported AVE=0.65, coefficient Ω=0.94, and Cronbach α=0.91 (Multimedia Appendix 2).

Data Collection

Before the clinical-vignette sequence, participants reported their professional designation, clinical experience, prior AI use, baseline trust, intention to use AI, and self-confidence. During each clinical vignette, the accept or reject decision was elicited as a binary choice, the three trust items as 1-7 ratings, and perceived diagnostic difficulty as a 1-5 rating. Responses were linked by anonymous study record, case series, and clinical-vignette position for repeated-measures analysis.

The first 20 eligible participants formed a pilot test to assess survey access, clarity of instructions, screen functioning, and whether participants could understand and answer the questions. They met the same eligibility criteria and completed the same consent process, measures, clinical-vignette sequence, and feedback procedure as the remaining participants. No technical or comprehension problems were identified, and no changes were made; these responses were retained in the final analytic sample.

XSL•FO
RenderX

Analysis

The hypotheses concerned the trust composite: H1a predicted a difference between correct and intentionally incorrect recommendations after accounting for vignette position and participant and vignette variability; H1b predicted a positive association between trust in consecutive recommendations; H2 predicted higher trust among participants with stronger baseline intention to use AI; and H3 predicted lower trust for clinical vignettes perceived as more difficult.

The demonstration vignette and attention-check screen were excluded. All 3 trust items and perceived-difficulty responses were complete for the 68 analyzed participants across 21 clinical vignettes, yielding 1428 participant-vignette observations. The data were organized in long format. The 5-point difficulty rating was categorized as lower difficulty (very easy or easy) or higher difficulty (moderate, difficult, or very difficult); a sensitivity model retained the original 1-5 score. Perceived difficulty was treated as a concurrent subjective task-context predictor rather than as a preexposure causal confounder.

Recommendation correctness was coded separately for each case series using the programmed status and cross-checked against Qualtrics task-score fields. Clinical-vignette position was centered at position 11. Correctness remained partly linked to vignette content, position, and the fixed reliability sequence; coefficients involving correctness were therefore interpreted as adjusted associations rather than causal effects.

The primary analysis pooled all 68 participants. A parsimonious cross-classified linear mixed-effects model assessed associations with the trust composite and included random intercepts for participant and vignette item, with 42 distinct vignette items. Fixed effects were recommendation correctness, centered clinical-vignette position, case series, professional designation, baseline trust, intention to use AI, self-confidence, clinical experience (≤10 vs >10 years), and perceived difficulty. Continuous participant-level predictors were mean-centered. Case series was a nuisance adjustment because vignette content and response window varied together; between-series differences were not attributed to response timing.

The primary model was fitted separately within each case series, and a pooled vignette position × case series model examined differences in sequence slopes. A secondary interaction model examined whether the incorrect-versus-correct trust difference varied with vignette position, baseline trust, intention to use AI, self-confidence, experience, or perceived difficulty. These interactions were excluded from the parsimonious primary model and interpreted as hypothesis-generating. An exploratory lagged model used positions 2-21 to relate trust in the current recommendation to trust in the immediately preceding recommendation while adjusting for current correctness, vignette position, case series, difficulty, and participant variables. Models with and without preceding-vignette trust were compared by maximum likelihood; the lagged coefficient was interpreted as a temporal association rather than evidence of a causal mechanism.

The study and analyses were not preregistered. Models were estimated by restricted maximum likelihood in Python 3.13 (Python Software Foundation) using *statsmodels* (version 0.14.6). Maximum likelihood was used for the lagged-model comparison. All reported models converged. Estimates were summarized using coefficients, SEs, 2-sided Wald tests, and 95% CIs. Standardized effects came from parallel models with z-standardized outcomes and continuous predictors, retaining the original categorical coding. Marginal R^2 described variance explained by fixed effects; conditional R^2 included fixed and random effects. Trust-composite end point frequencies were also examined.

Behavioral reliance was summarized by acceptance and rejection relative to programmed recommendation correctness. A total of 50 case series B observations had no recorded accept or reject response. For the primary descriptive analysis, these observations were assigned to the rejection category under the assumption that participants who trusted a recommendation sufficiently to act on it would record acceptance within the 45-second response window. This rule retained all presented clinical vignettes and treated unrecorded decisions as nonacceptance; it did not establish that participants actively rejected or distrusted the recommendation. A descriptive sensitivity summary excluded unrecorded decisions instead of recoding them; the resulting counts and denominators are reported in Multimedia Appendix 3.

Ethical Considerations

The West Virginia University (WVU) Institutional Review Board (IRB) approved the study under the WVU Flexibility Review Model (protocol number 2502117081). All participants provided electronic informed consent before beginning the survey. The study did not request names, contact information, IP addresses, licensure identifiers, employer information, or other direct identifiers; the analytic dataset contained anonymous responses. Study files were stored in password-protected locations accessible only to authorized research personnel. Participants were recruited and compensated through Centiment according to the provider's panel procedures; the exact participant payment was not available to the research team.

This was an online clinical-vignette experiment without real patient encounters or patient outcomes. Applicable CONSORT-AI (Consolidated Standards of Reporting Trials–Artificial Intelligence) reporting principles informed descriptions of eligibility, allocation, participant flow, recommendation presentation, human-AI interaction, and analysis [31].

Results

Participant Flow and Characteristics

A total of 113 individuals participated in the survey. Sixty-eight met the eligibility and completeness criteria and were included in the analysis. The panel provider did not supply the number invited or prescreened, the number excluded through provider-level fraud or quality procedures, or sufficient information to distinguish ineligible from incomplete records among the remaining 45 participants. The final sample included 59/68 (86.76%) RNs and 9/68 (13.24%) physicians, with 34 participants assigned to each case series. All 68 analyzed

XSL•FO
RenderX

participants provided complete responses to the 3 trust questionnaire items and the perceived-difficulty item across 21 clinical vignettes.

Baseline descriptive statistics, reported as means and SDs, are presented in Table 3. The mean baseline trust composite was 4.37 (SD 1.64); mean intention to use AI was 3.35 (SD 1.16); and mean self-confidence was 4.02 (SD 0.70).

Table 3. Participant characteristics and baseline measures[a].

Characteristic	Case series A (n=34)	Case series B (n=34)	Total (n=68)
Professional designation, n (%)			
Registered nurse	31 (91.18)	28 (82.35)	59 (86.76)
Physician	3 (8.82)	6 (17.65)	9 (13.24)
Experience (years), n (%)			
<1	2 (5.88)	0 (0.00)	2 (2.94)
1-2	1 (2.94)	3 (8.82)	4 (5.88)
3-5	8 (23.53)	2 (5.88)	10 (14.71)
6-10	10 (29.41)	4 (11.76)	14 (20.59)
>10	13 (38.24)	25 (73.53)	38 (55.88)
Prior AI use, n (%)			
Yes	16 (47.06)	16 (47.06)	32 (47.06)
No	18 (52.94)	18 (52.94)	36 (52.94)
Baseline trust composite, mean (SD)	4.28 (1.66)	4.46 (1.64)	4.37 (1.64)
Intent to use AI, mean (SD)	3.41 (1.08)	3.29 (1.24)	3.35 (1.16)
Self-confidence, mean (SD)	3.95 (0.66)	4.09 (0.74)	4.02 (0.70)

[a]The baseline trust composite is the unrounded mean of general trust, perceived transparency, and likelihood-to-act items. Self-confidence is the unrounded mean of 9 retained items.

Perceived Diagnostic Difficulty

Of the 714 difficulty ratings in each series, 342 (47.90%) were categorized as higher difficulty in case series A and 355 (49.72%) in case series B. The position-specific distributions are shown in Multimedia Appendix 3.

Trust Across Clinical Vignettes

The trust composite varied across the 21 clinical vignettes in both series (Figure S2 in Multimedia Appendix 3). Case series A increased from a mean of 5.04 at position 1 to 5.38 at position 4, decreased to 3.56 at the first incorrect recommendation at position 5, and ended at 3.21. Case series B began at 4.80, decreased to 3.54 at position 5, and ended at 4.32. Grouped means by case series, difficulty, and correctness are shown in Multimedia Appendix 3.

Across 1428 trust composites, 123 (8.61%) were at the minimum of 1 and 177 (12.39%) were at the maximum of 7; 300 scores (21.01%) occurred at either end point. Scores of 2 or lower accounted for 17.72%, and scores of 6 or higher accounted for 27.03%.

Behavioral Reliance

Across both series, 374 of 748 (50%) incorrect recommendations were accepted, and 112 of 680 (16.47%) correct recommendations were in the rejection category. Each series contributed 714 decision instances: 374 with incorrect and 340 with correct recommendations. Incorrect recommendations were accepted 185 times in case series A (49.47% of incorrect recommendations; 25.91% of all series A observations) and 189 times in case series B (50.53%; 26.47% of all series B observations). Correct recommendations were in the rejection category 52 times in case series A (15.29% of correct recommendations; 7.28% of all series A observations) and 60 times in case series B (17.65%; 8.40%).

At clinical vignette 5, the first incorrect recommendation in both series, 16 participants in case series A accepted and 18 rejected the recommendation; in case series B, 12 accepted and 22 responses were classified in the rejection category. Acceptance was above zero for every incorrect recommendation. When both series returned to a correct recommendation at position 17, 25 participants in case series A and 28 in case series B accepted it. Position-specific counts are shown in Multimedia Appendix 3.

Of the 50 unrecorded case series B decisions, 26 occurred for correct and 24 for incorrect recommendations. When only recorded responses were counted, incorrect recommendations were accepted in 374 of 724 (51.66%) observations, and correct recommendations were explicitly rejected in 86 of 654 (13.15%) observations. In case series B alone, explicit rejection of correct recommendations was 34 of 314 (10.83%) recorded responses, compared with 60 of 340 (17.65%) under rejection-category coding. Complete denominators for both summaries are provided in Multimedia Appendix 3.

XSL•FO
RenderX

Primary Mixed-Effects Model

Table 4 presents the pooled parsimonious cross-classified model. Incorrect AI recommendations were associated with lower prefeedback trust than correct recommendations (β=−0.772, 95% CI −1.036 to −0.507; standardized effect=−0.419). Trust also declined modestly across clinical vignettes (β=−0.027, 95% CI −0.048 to −0.005; standardized effect=−0.088). Intention to use AI was positively associated with trust (β=0.391, 95% CI 0.110-0.672; standardized effect=0.243), whereas higher perceived difficulty was negatively associated with trust (β=−0.474, 95% CI −0.648 to −0.301; standardized effect=−0.257). CIs for the remaining fixed effects included zero. The sensitivity model retaining the original 1-5 difficulty score yielded the same substantive pattern; each 1-point increase in difficulty was associated with a 0.404-point lower trust composite (95% CI −0.495 to −0.312).

Table 4. Parsimonious cross-classified mixed-effects model predicting the vignette-level trust composite[a].

Predictor	β (SE)	95% CI	*P*	Standardized effect
Intercept	4.795 (0.249)	4.307 to 5.283	<.001	—[b]
Vignette position (centered)	−0.027 (0.011)	−0.048 to −0.005	.02	−0.088
AI recommendation incorrect	−0.772 (0.135)	−1.036 to −0.507	<.001	−0.419
Case series B vs A (nuisance adjustment)	0.174 (0.300)	−0.414 to 0.762	.56	0.094
Professional designation (physician)	0.243 (0.401)	−0.542 to 1.029	.54	0.132
Baseline trust composite	0.072 (0.095)	−0.114 to 0.259	.45	0.064
Intent to use AI	0.391 (0.143)	0.110 to 0.672	.006	0.243
Self-confidence	0.047 (0.198)	−0.341 to 0.435	.81	0.018
Experience >10 years	0.007 (0.294)	−0.569 to 0.583	.98	0.004
Perceived diagnostic difficulty: higher	−0.474 (0.089)	−0.648 to −0.301	<.001	−0.257

[a]N=1428 observations from 68 participants and 42 vignette items. Reference categories: correct recommendation, case series A, registered nurse, ≤10 years of experience, and lower perceived difficulty. The case series was included only as a nuisance adjustment. Marginal R^2=0.176; conditional R^2=0.500. Standardized effects were obtained from a parallel model in which the outcome and continuous predictors were z-standardized; categorical predictors retained their original coding.

[b]Not applicable.

Sensitivity Analyses by Case Series

Sensitivity analyses showed that lower trust for incorrect recommendations was evident in both case series: case series A, β=−0.850 (95% CI −1.197 to −0.502), and case series B, β=−0.676 (95% CI −0.996 to −0.356). The longitudinal pattern differed across vignette sets. Trust declined with vignette position in case series A (β=−0.056, 95% CI −0.085 to −0.027) but not in case series B (β=0.004, 95% CI −0.022 to 0.030). In the pooled vignette position × case series model, the case series B slope was 0.064 points per clinical vignette less negative than the case series A slope (95% CI 0.027-0.102). Higher perceived difficulty was negatively associated with trust in both series, whereas intention to use AI was positively associated with trust only in case series B (Table 5).

XSL•FO
RenderX

Table 5. Parsimonious models fitted separately within each case series[a].

Predictor	Series A			Series B		
	β (SE)	95% CI	*P* value	β (SE)	95% CI	*P* value
Intercept	4.918 (0.269)	4.390 to 5.446	<.001	4.846 (0.458)	3.948 to 5.744	<.001
Vignette position (centered)	−0.056 (0.015)	−0.085 to −0.027	<.001	0.004 (0.013)	−0.022 to 0.030	.77
AI recommendation incorrect	−0.850 (0.177)	−1.197 to −0.502	<.001	−0.676 (0.163)	−0.996 to 0.356	<.001
Professional designation: physician	1.144 (0.642)	−0.113 to 2.402	.07	−0.354 (0.564)	−1.459 to 0.752	.53
Baseline trust composite	−0.060 (0.131)	−0.317 to 0.198	.65	0.134 (0.128)	−0.118 to 0.385	.30
Intent to use AI	0.236 (0.203)	−0.163 to 0.634	.25	0.619 (0.211)	0.205 to 1.033	.003
Self-confidence	−0.211 (0.273)	−0.745 to 0.324	.44	0.045 (0.313)	−0.568 to 0.657	.89
Experience >10 years	−0.305 (0.367)	−1.024 to 0.414	.41	0.173 (0.486)	−0.780 to 1.126	.72
Perceived diagnostic difficulty: higher	−0.604 (0.144)	−0.887 to −0.321	<.001	−0.344 (0.103)	−0.546 to −0.141	<.001

[a]Each model included 34 participants, 714 observations, 21 vignette items, and random intercepts for participant and vignette item. Marginal or conditional R^2=0.178/0.438 for case series A and 0.323/0.601 for case series B. The pooled vignette position × case series model is summarized in the text.

Exploratory Interaction Model

The exploratory model suggested that sensitivity to incorrect recommendations varied with several characteristics. The incorrect-versus-correct difference was more negative among participants with more than 10 years of experience (interaction β=−0.703, 95% CI −0.998 to −0.407) and at higher self-confidence (interaction β=−0.360 per scale point, 95% CI −0.571 to −0.150), but less negative on clinical vignettes categorized as higher difficulty (interaction β=0.327, 95% CI 0.045-0.609). Interactions with vignette position, baseline trust, and intention to use AI had CIs that included zero. Because this interaction-heavy model was secondary and participant-level power was limited, these findings are considered hypothesis-generating. Full model estimates are reported in Multimedia Appendix 3.

Exploratory Lagged Model

In the exploratory lagged model, the preceding clinical vignette's trust composite was positively associated with current-vignette trust (β=0.266, 95% CI 0.210-0.322; standardized effect=0.264). Incorrect current-vignette recommendations remained associated with lower trust (β=−0.753, 95% CI −1.037 to −0.469; standardized effect=−0.405), intention to use AI remained positively associated with trust, and higher perceived difficulty remained negatively associated with trust. Vignette position and the remaining participant-level variables had CIs that included zero. Adding preceding-vignette trust improved model fit relative to the otherwise identical model without the lagged term. Full coefficients and model-fit statistics are reported in Multimedia Appendix 3.

Hypothesis Summary

H1a was supported because incorrect recommendations received lower vignette-level trust than correct recommendations. H1b was supported statistically because the preceding clinical vignette's trust composite was positively associated with current-vignette trust, although this result is interpreted as a temporal association rather than evidence of a causal carryover mechanism. H2 was supported because baseline intention to use AI was positively associated with vignette-level trust. H3 was supported because higher perceived diagnostic difficulty was negatively associated with vignette-level trust.

Discussion

Principal Findings and Contributions

Participants assigned lower trust to incorrect than to correct AI recommendations, yet accepted half of the incorrect recommendations and rejected approximately 1 in 6 correct recommendations under the stated behavioral coding rule. This aggregate pattern identifies a potential gap between self-reported trust and reliance calibrated to recommendation correctness; it does not establish that individual participants recognized errors before accepting them. Trust in consecutive recommendations was positively associated; baseline intention to use AI was associated with higher trust, and greater perceived difficulty with lower trust. The pooled decline across vignette positions was evident in case series A but not case series B.

Together, these results suggest that clinical AI evaluation should focus on 2 related but distinct questions: whether users can judge the quality of individual AI recommendations and whether those judgments lead to appropriate reliance. The findings should nevertheless be interpreted as preliminary because they were obtained from a small, nurse-dominant sample completing simulated primary care–style tasks.

Dynamic Trust and Recommendation Correctness

Participants gave substantially lower trust ratings to incorrect recommendations than to correct recommendations. Because the trust questions were completed before current-vignette feedback was shown, this difference suggests that participants were, on average, sensitive to features of the vignette and the

XSL•FO
RenderX

apparent plausibility of the AI recommendation. The average difference was therefore present before explicit feedback about that recommendation. This result is consistent with recent repeated-interaction studies. Kahr et al [32] found that people reported greater trust and reliance when AI advice came from a more accurate model, while Yang et al [33] showed that users adjust trust from one interaction to the next and react particularly strongly to automation failures.

The pattern of trust across the 21 clinical vignettes was less consistent. Trust declined in case series A but remained comparatively stable in case series B. This qualification is important because it means that the pooled decline should not be interpreted as a universal process in which trust steadily erodes whenever users encounter AI errors. Instead, the course of trust appears to depend on which cases users see, how convincing the recommendations appear, and where correct and incorrect recommendations occur in the sequence. This result is consistent with Rittenberg et al [34], who found that trust did not always rise and fall in direct proportion to automation reliability. In their experiments, trust sometimes declined even when system reliability improved, and trust was difficult to rebuild after users had experienced a poorly performing system. Kahr et al [32], by contrast, observed increasing trust when participants repeatedly interacted with high-accuracy AI advice. Considered together, these studies and the present findings indicate that there may not be one standard trust trajectory. Trust depends on reliability, but it is also shaped by starting impressions, the order and type of errors, task characteristics, and the user's ability to evaluate each recommendation. Because recommendation correctness, case content, and vignette position were not independently randomized in this study, the lower trust assigned to incorrect recommendations cannot be attributed solely to correctness. Some of the difference may also reflect the particular cases used or their position in the sequence. The main conclusion is therefore that participants showed aggregate sensitivity to recommendation quality, while the way trust changed over time remained dependent on the specific case series.

Trust in Consecutive Recommendations

Trust in the immediately preceding clinical vignette was positively associated with trust in the current clinical vignette. In everyday terms, participants appeared to carry part of their recent impression of the AI into the next case rather than evaluating every recommendation in complete isolation. This makes sense in repeated AI use: after seeing several useful recommendations, a person may approach the next recommendation more favorably, whereas recent errors may create greater caution. This finding is consistent with evidence from Kahr et al [32], Yang et al [33], and Rittenberg et al [34] that early reliability experiences can influence later trust and its recovery. This study extends this work to repeated clinical-vignette tasks. However, the lagged finding should not be described as proof of trust inertia. Consecutive ratings from the same participant will often resemble one another potentially because individuals have relatively stable response styles and general attitudes. The most defensible interpretation is that trust showed temporal continuity where the previous rating provided useful information about the next rating, but the study cannot determine whether this occurred because of memory of previous AI performance, accumulated feedback, a stable personal response pattern, or another cognitive process.

The practical implication remains important. Trust in AI should be studied as a developing process rather than as a one-time opinion.

Reported Trust and Behavioral Reliance

One of the most important contributions of this study is the separate examination of reported trust and actual accept or reject behavior. Lower average trust for incorrect recommendations coexisted with frequent acceptance of those recommendations. This finding supports prior concerns about automation bias in clinical decision support. Gaube et al [35] found that inaccurate advice reduced physicians' diagnostic accuracy and that substantial subgroups of both radiologists and physicians with less radiology expertise repeatedly followed incorrect advice. This study identifies a related concern in a different task: lower average trust for incorrect advice did not coincide with uniformly low acceptance of that advice. A recommendation may be accepted because it appears plausible, reduces the effort required to reach a decision, or is difficult to verify independently.

At the same time, the present behavioral pattern differs from findings reported by Küper et al [27]. In their 2025 study of 223 dermatologists, participants relied more strongly on correct than incorrect AI advice, but overall reliance on AI was low and self-reliance was high. In this study, by contrast, half of the incorrect recommendations were accepted. This difference may reflect several features: Küper et al [27] studied specialists performing dermatology image-classification tasks, whereas this study included a nurse-dominant sample evaluating short primary care–style vignettes.

The binary accept or reject decision used in this study provides a clear behavioral indicator, but actual clinical reliance may be more complicated. Sivaraman et al [9] found that intensive care unit (ICU) clinicians did not always simply accept or reject AI treatment advice; they often accepted some parts, rejected others, or delayed action, a process described as negotiation. Thus, the current findings clearly demonstrate misalignment between reference correctness and behavior, but future studies should allow more nuanced responses that better reflect how health care professionals use recommendations in practice.

Together, the aggregate findings support measuring trust and behavioral reliance separately rather than assuming that lower self-reported trust ensures rejection of incorrect advice. The present analyses did not directly test a within-participant causal pathway from trust to the accept or reject decision.

Baseline AI Attitudes and Perceived Difficulty

Participants who entered the study with a stronger intention to use AI generally reported higher trust across the individual cases. This finding is consistent with recent reviews showing that perceived usefulness, openness to technology, previous experience, and expectations about AI influence health care professionals' acceptance of AI systems [5,6,16]. This study extends that literature by showing that a general willingness to

XSL•FO
RenderX

use AI was related not only to overall acceptance, but also to evaluations of specific recommendations during repeated interactions. However, the baseline trust composite itself was not clearly associated with vignette-level trust after the other variables were considered. This suggests that general trust and intention to use AI may not represent the same thing. A participant may express general trust in AI but remain unwilling to use it because of workflow, accountability, or professional concerns. Conversely, someone may be willing to use AI as a helpful tool without assuming that every recommendation is trustworthy. This distinction should be studied more directly in future work.

Participants also reported lower trust in cases they perceived as more difficult. This finding suggests that difficult or ambiguous cases may make participants less certain not only about their own judgment, but also about the quality of the AI recommendation. It complements the finding reported by Gaube et al [35] that the effect of inaccurate clinical advice varied across individual cases and was especially concerning when cases involved difficult-to-recognize diagnostic features. Because perceived difficulty was measured during the same clinical vignette, this study cannot conclude that difficulty caused lower trust. A case may have seemed difficult because the vignette was ambiguous, because the AI recommendation appeared inconsistent, because the participant lacked relevant knowledge, or because several of these factors occurred together. The appropriate takeaway is that participants trusted AI recommendations less on cases they personally experienced as more difficult.

Limitations and Future Research

Several limitations constrain interpretation. First, the study used short primary care–style clinical vignettes rather than real encounters. The cases were drafted with LLM assistance, anchored to a 2010 textbook, and reviewed by one physician, but were not independently adjudicated by multiple practicing primary care clinicians. They may not reflect current guidelines, contemporary diagnostic pathways, or the complexity of clinical practice.

Second, the sample was small and nurse-dominant. Registered nurses represented 86.76% (59/68) of participants, whereas physicians represented 13.24% (9/68). Although nurses are important users of clinical information and decision support, they do not ordinarily make all primary care diagnoses or prescribe treatments independently. The mismatch between the simulated task and usual RN scope may have affected perceived difficulty, trust, and reliance, limiting ecological validity and generalization to either physician diagnosis or routine nursing practice.

Third, professional designation was self-reported and was not independently verified through licensure, employer, or National Provider Identifier records. The third-party panel did not provide complete counts for invitations, prescreening, or provider-level quality exclusions. Fourth, the participant-level sample limited power for professional-group comparisons and the exploratory interaction tests; their estimates should be interpreted as hypothesis-generating.

Fifth, the two case series differed in vignette content, later correctness positions, and response window. These features were not independently varied, so between-series differences cannot be attributed to response timing or case characteristics. The differing clinical vignette slopes demonstrate that the pooled longitudinal pattern was sensitive to vignette set. Sixth, correctness followed fixed series-specific sequences rather than being randomized or counterbalanced, leaving correctness partly linked to vignette content, vignette position, and reliability phase.

The handling of unrecorded decisions is an additional limitation. Classifying the 50 unrecorded case series B responses as rejection represented nonacceptance, not a verified active decision. It may overstate deliberate rejection and understate acceptance among participants who provided an explicit response. This matters particularly for apparent underreliance: 26 unrecorded responses to correct recommendations contributed to the rejection category. The recorded-response sensitivity summary illustrates this dependence on coding but cannot establish what the unrecorded decisions would have been. Future studies should retain timeout or nonresponse as a separate behavioral outcome rather than assume that it is equivalent to rejection.

Conclusions

This nurse-dominant pilot study illustrates why clinical AI evaluation should examine both self-reported trust and observed reliance across clinical vignettes. Incorrect recommendations received lower trust on average, but half were accepted; approximately 1 in 6 correct recommendations fell into the rejection category under the stated coding rule. These findings identify a potential evaluation problem rather than a demonstrated patient-safety effect. Larger studies using independently validated, role-aligned clinical vignettes, counterbalanced experimental factors, and separate recording of active rejection and nonresponse are needed to determine how recommendation quality, feedback, and difficulty shape appropriate reliance.

Acknowledgments

The authors gratefully acknowledge Dr Zaira Chaudhry for her expert physician review and finalization of the clinical vignettes, reference diagnoses, and AI recommendations following the authors' textbook-based checks. Her review assessed clinical coherence, plausibility, and consistency with each recommendation's intended correctness status before deployment. The authors declare the use of generative AI (GenAI) in the research and writing process. According to the GAIDeT (Generative AI Delegation Taxonomy; 2025), the following tasks were delegated to GenAI tools under full human supervision: literature search and systematization, code generation, reproducibility testing, proofreading and editing, and summarizing text. The GenAI tools used

XSL•FO
RenderX

were Apple AI and ChatGPT 5.6. Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes. The declaration was submitted by all authors.

Funding

The authors declared no financial support was received for this work.

Authors' Contributions

AC was responsible for conceptualization, survey development, figure development, data collection, experimental design, and manuscript drafting. YS contributed to data analysis, survey development, figure development, data collection, experimental design, and manuscript drafting. APG contributed to critical manuscript revision. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Clinical vignettes, generation procedure, and prompts.
[DOCX File , 42 KB - humanfactors_v13i1e97649_app1.docx]

Multimedia Appendix 2
Measurement structure and reliability.
[DOCX File , 40 KB - humanfactors_v13i1e97649_app2.docx]

Multimedia Appendix 3
Supplementary Analyses and Results.
[DOCX File , 536 KB - humanfactors_v13i1e97649_app3.docx]

References

1. Chustecki M. Benefits and risks of AI in health care: narrative review. Interact J Med Res 2024;13:e53616 [FREE Full text] [doi: 10.2196/53616] [Medline: 39556817]
2. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]
3. Taylor RA, Sangal RB, Smith ME, Haimovich AD, Rodman A, Iscoe MS, et al. Leveraging artificial intelligence to reduce diagnostic errors in emergency medicine: challenges, opportunities, and future directions. Acad Emerg Med 2025;32(3):327-339. [doi: 10.1111/acem.15066] [Medline: 39676165]
4. Fackler J, Ghobadi K, Gurses A. Algorithms at the bedside: moving past development and validation. Pediatr Crit Care Med 2024;25(3):276-278. [doi: 10.1097/PCC.0000000000003437] [Medline: 38451799]
5. Lambert SI, Madi M, Sopka S, Lenes A, Stange H, Buszello C, et al. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. NPJ Digit Med 2023;6(1):111 [FREE Full text] [doi: 10.1038/s41746-023-00852-5] [Medline: 37301946]
6. Steerling E, Siira E, Nilsen P, Svedberg P, Nygren J. Implementing AI in healthcare-the relevance of trust: a scoping review. Front Health Serv 2023;3:1211150 [FREE Full text] [doi: 10.3389/frhs.2023.1211150] [Medline: 37693234]
7. Scott IA, van der Vegt A, Lane P, McPhail S, Magrabi F. Achieving large-scale clinician adoption of AI-enabled decision support. BMJ Health Care Inform 2024;31(1):e100971 [FREE Full text] [doi: 10.1136/bmjhci-2023-100971] [Medline: 38816209]
8. Choudhury A, Chaudhry Z. Large language models and user trust: consequence of self-referential learning loop and the deskilling of health care professionals. J Med Internet Res 2024;26:e56764 [FREE Full text] [doi: 10.2196/56764] [Medline: 38662419]
9. Sivaraman V, Bukowski LA, Levin J, Kahn JM, Perer A. Ignore, trust, or negotiate: understanding clinician acceptance of AI-based treatment recommendations in health care. 2023 Presented at: CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems; 2023 April 23 - 28; Hamburg, Germany p. 1-18. [doi: 10.1145/3544548.3581075]
10. Choudhury A, Shamszare H. Human factors influencing trust in healthcare artificial intelligence: systematic literature review. IISE Trans Occup Ergon Hum Factors 2026;14(2):116-131. [doi: 10.1080/24725838.2026.2625655] [Medline: 41674135]
11. Wong KKL, Han Y, Cai Y, Ouyang W, Du H, Liu C. From trust in automation to trust in AI in healthcare: a 30-year longitudinal review and an interdisciplinary framework. Bioengineering (Basel) 2025;12(10):1070 [FREE Full text] [doi: 10.3390/bioengineering12101070] [Medline: 41155069]

XSL•FO
RenderX

12. Li X, Zong Q, Cheng M. The impact of medical explainable artificial intelligence on nurses' innovation behaviour: a structural equation modelling approach. J Nurs Manag 2024;2024:8885760. [doi: 10.1155/2024/8885760] [Medline: 40224836]
13. Shamszare H, Chaudhry Z, Berenji M, Choudhury A, CHOUDHURY A, CHOUDHURY A, et al. Conceptualizing clinicians' trust in artificial intelligence as a function of their expertise, workload, patient outcome, diagnosis difficulty, and AI accuracy: a systems thinking approach. IEEE Access 2025;13:119601-119618. [doi: 10.1109/access.2025.3586555]
14. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc 2012;19(1):121-127 [FREE Full text] [doi: 10.1136/amiajnl-2011-000089] [Medline: 21685142]
15. Vereschak O, Alizadeh F, Bailly G, Caramiaux B. Trust in ai-assisted decision making: perspectives from those behind the system and those for whom the decision is made. 2024 Presented at: CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems; 2024 May 11 - 16; Honolulu, HI, USA p. 1-14. [doi: 10.1145/3613904.3642018]
16. Tun HM, Rahman HA, Naing L, Malik OA. Trust in artificial intelligence-based clinical decision support systems among health care workers: systematic review. J Med Internet Res 2025;27:e69678 [FREE Full text] [doi: 10.2196/69678] [Medline: 40772775]
17. Sato T, Inman J, Politowicz MS, Chancey ET, Yamani Y. A meta-analytic approach to investigating the relationship between human-automation trust and attention allocation. Proc Hum Factors Ergon Soc Annu Meet 2023;67(1):959-964. [doi: 10.1177/21695067231194333]
18. Kiani A, Uyumazturk B, Rajpurkar P, Wang A, Gao R, Jones E, et al. Impact of a deep learning assistant on the histopathologic classification of liver cancer. NPJ Digit Med 2020;3:23 [FREE Full text] [doi: 10.1038/s41746-020-0232-8] [Medline: 32140566]
19. Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, et al. Human-computer collaboration for skin cancer recognition. Nat Med 2020;26(8):1229-1234 [FREE Full text] [doi: 10.1038/s41591-020-0942-0] [Medline: 32572267]
20. Chaudhry ZS, Choudhury A. Clinical applications of artificial intelligence in occupational health: a systematic literature review. J Occup Environ Med 2024;66(12):943-955. [doi: 10.1097/JOM.0000000000003212] [Medline: 39190393]
21. Choudhury A. Toward an ecologically valid conceptual framework for the use of artificial intelligence in clinical settings: need for systems thinking, accountability, decision-making, trust, and patient safety considerations in safeguarding the technology and clinicians. JMIR Hum Factors 2022;9(2):e35421 [FREE Full text] [doi: 10.2196/35421] [Medline: 35727615]
22. Browne JT, Bakker S, Yu B, Lloyd P, Ben AS. Trust in clinical AI: expanding the unit of analysis. In: HHAI2022: Augmenting Human Intellect. Amsterdam, the Netherlands: IOS Press; 2022:96-113.
23. Lee JD, See KA. Trust in automation: designing for appropriate reliance. Hum Factors 2004;46(1):50-80. [doi: 10.1518/hfes.46.1.50_30392] [Medline: 15151155]
24. Hoff KA, Bashir M. Trust in automation: integrating empirical evidence on factors that influence trust. Hum Factors 2015;57(3):407-434. [doi: 10.1177/0018720814547570] [Medline: 25875432]
25. Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. J Am Med Inform Assoc 2017;24(2):423-431 [FREE Full text] [doi: 10.1093/jamia/ocw105] [Medline: 27516495]
26. Sakamoto T, Harada Y, Shimizu T. Facilitating trust calibration in artificial intelligence-driven diagnostic decision support systems for determining physicians' diagnostic accuracy: quasi-experimental study. JMIR Form Res 2024;8:e58666 [FREE Full text] [doi: 10.2196/58666] [Medline: 39602469]
27. Küper A, Lodde GC, Livingstone E, Schadendorf D, Krämer N. Psychological factors influencing appropriate reliance on AI-enabled clinical decision support systems: experimental web-based study among dermatologists. J Med Internet Res 2025;27:e58660 [FREE Full text] [doi: 10.2196/58660] [Medline: 40184614]
28. McPhee SJ, Papadakis MA. Current Medical Diagnosis and Treatment 2010. New York, NY: McGraw-Hill Medical; 2010.
29. Audience panel and research services. Centiment. URL: https://www.centiment.co/ [accessed 2026-09-10]
30. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
31. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AICONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med 2020;26(9):1364-1374 [FREE Full text] [doi: 10.1038/s41591-020-1034-x] [Medline: 32908283]
32. Kahr PK, Rooks G, Willemsen MC, Snijders CCP. Understanding trust and reliance development in AI advice: assessing model accuracy, model explanations, and experiences from previous interactions. ACM Trans. Interact. Intell. Syst 2024;14(4):1-30. [doi: 10.1145/3686164]
33. Yang XJ, Schemanske C, Searle C. Toward quantifying trust dynamics: how people adjust their trust after moment-to-moment interaction with automation. Hum Factors 2023;65(5):862-878 [FREE Full text] [doi: 10.1177/00187208211034716] [Medline: 34459266]
34. Rittenberg BSP, Holland CW, Barnhart GE, Gaudreau SM, Neyedli HF. Trust with increasing and decreasing reliability. Hum Factors 2024;66(12):2569-2589 [FREE Full text] [doi: 10.1177/00187208241228636] [Medline: 38445652]
35. Gaube S, Suresh H, Raue M, Merritt A, Berkowitz SJ, Lermer E, et al. Do as AI say: susceptibility in deployment of clinical decision-aids. NPJ Digit Med 2021;4(1):31 [FREE Full text] [doi: 10.1038/s41746-021-00385-9] [Medline: 33608629]

XSL•FO
RenderX

Abbreviations

AVE: average variance extracted
CDSS: clinical decision support systems
CFA: confirmatory factor analysis
CONSORT-AI: Consolidated Standards of Reporting Trials–Artificial Intelligence
ICU: intensive care unit
IRB: Institutional Review Board
LLM: large language model
RN: registered nurse
WVU: West Virginia University

Edited by S Law; submitted 08.Apr.2026; peer-reviewed by M Saremi, Y Harada; comments to author 23.Jun.2026; revised version received 11.Sep.2026; accepted 11.Sep.2026; published 08.Oct.2026.

Please cite as:
Choudhury A, Shahsavar Y, Gurses AP
Clinicians' Trust in AI-Based Clinical Recommendations Across Controlled Primary Care–Style Clinical Vignettes: Nurse-Dominant Experimental Pilot Study
JMIR Hum Factors 2026;13:e97649
URL: https://humanfactors.jmir.org/2026/1/e97649
doi:10.2196/97649
PMID:

©Avishek Choudhury, Yeganeh Shahsavar, Ayse P Gurses. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 08.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Development and Initial Usability Evaluation of SSEndoSim, a Free, Web-Based Endodontic Diagnostic Simulator: Cross-Sectional Study

Sarang Suresh*, BDS; Priya Rani*, BDS; Feroze Kalhoro, BDS

Department of Operative Dentistry, Liaquat University of Medical & Health Sciences, Jamshoro, Sindh, Pakistan

*these authors contributed equally

Corresponding Author:
Sarang Suresh, BDS
Department of Operative Dentistry, Liaquat University of Medical & Health Sciences, Jamshoro, Sindh, Pakistan

Abstract

SSEndoSim, a free, open access, mobile, web-based endodontic diagnostic simulation application for undergraduate dental education, demonstrated favorable usability rated on a System Usability Scale (SUS) witht a composite 100-point score (mean 75.30, SD 14.72) and high perceived educational value (PEV) on a 5-point scale (mean 4.62, SD 0.40) in a cross-sectional usability evaluation of final-year bachelor of dental surgery (BDS) students at Liaquat University of Medical and Health Sciences (LUMHS), Jamshoro, Pakistan, in April 2026, establishing usability and learner acceptability.

(JMIR Hum Factors 2026;13:e100407) doi:10.2196/100407

KEYWORDS

dental education; endodontic diagnosis; gamification; mobile learning; simulation; serious game; System Usability Scale

Introduction

Endodontic diagnosis requires simultaneous pulpal and periapical diagnoses using American Association of Endodontists terminology [1], with treatment following the European Society of Endodontology S3-level guideline [2]. Undergraduate diagnostic accuracy is variable, and lectures and clinical exposure alone are insufficient to develop it consistently [3]. Simulation enables repeated engagement with curated cases as well as immediate feedback [4]. Free, browser-based tools may help where commercial simulation platforms are inaccessible.

We report the development of SSEndoSim, a free, web-based endodontic diagnostic simulator with automated scoring and structured feedback, and an initial evaluation of its usability and perceived educational value (PEV) among final-year bachelor of dental surgery (BDS) students ahead of a planned trial of learning effectiveness.

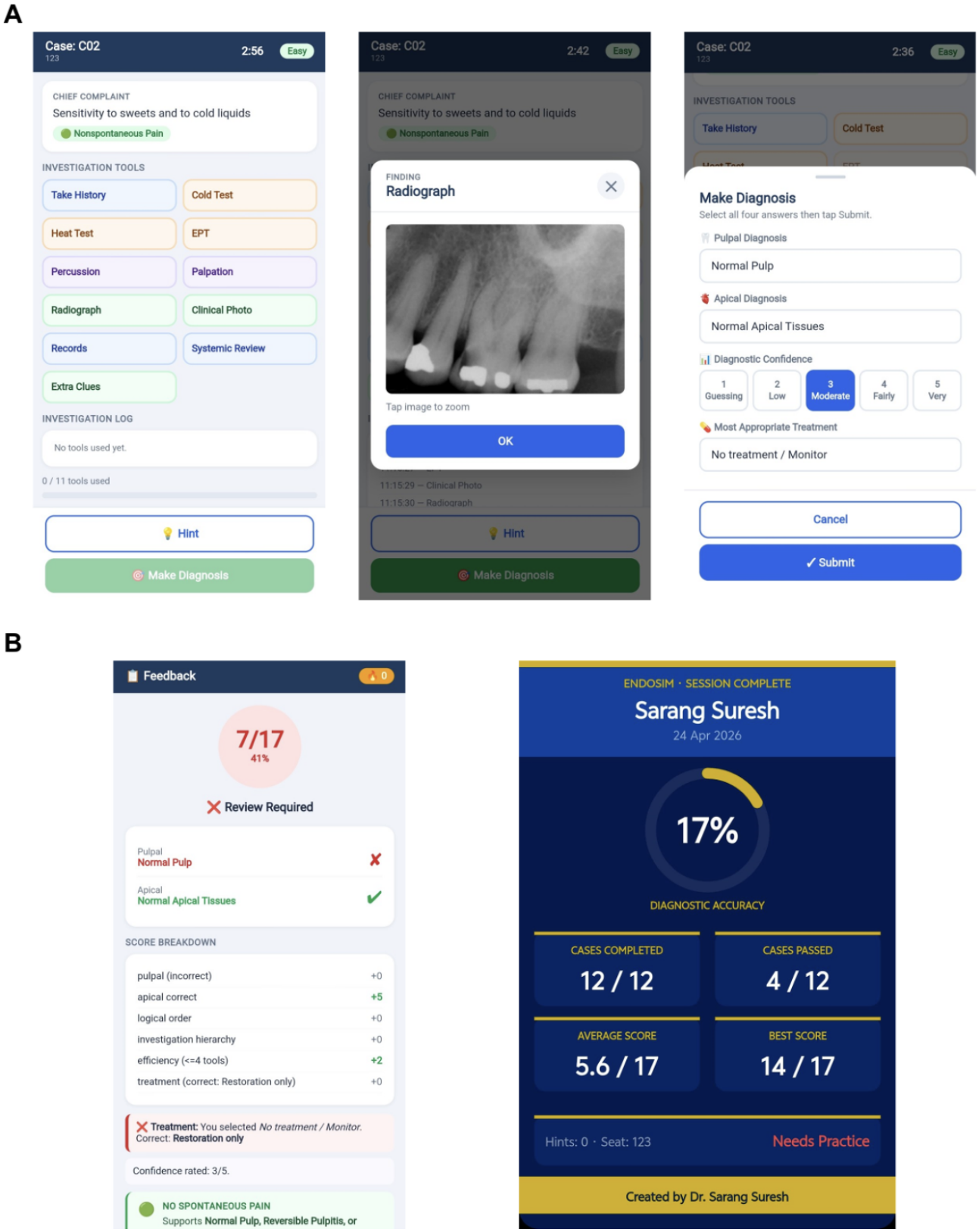
Methods

Application Development

SSEndoSim [5] was designed so that every case demands a simultaneous pulpal and periapical diagnosis per American Association of Endodontists terminology, and every submission receives immediate, objective feedback. It runs in any web browser without installation. The student selects investigations, submits both diagnoses, chooses treatment, and rates their confidence in the diagnoses. A 17-point rubric scores the diagnosis (10 points), treatment (2), and guideline-consistent investigation (5, with deductions), and feedback follows immediately (Figure 1). Rubric weights and difficulty labels were set by investigator judgment, and 2 senior endodontists reviewed the cases and rubric (expert review, not formal content validation). The 12-case bank spanned 6 of 7 pulpal and 4 of 6 periapical categories (4 easy, 6 moderate, and 2 hard). Cases were sampled without replacement within a session; reloading or changing difficulty began a new sequence, so cases could recur across sessions. Twenty postgraduate trainees and faculty piloted SSEndoSim informally beforehand (verbal feedback only), and their prewindow records were excluded from this study by date. SSEndoSim is for education with simulated cases, not diagnosis of real patients. Multimedia Appendix 1 details SSEndoSim's development, rubric, case bank, and log definitions.

XSL•FO
RenderX

Figure 1. Screen captures of the SSEndoSim interface. (A) Case investigation and diagnosis workflow: investigation tool panel, radiographic finding display, and combined pulpal diagnosis, periapical diagnosis, confidence, and treatment selection panel. (B) Automated postcase feedback (diagnosis result and score breakdown) and end-of-session summary card.



Participants and Measures

All 75 students in the eligible final-year BDS cohort at Liaquat University of Medical and Health Sciences (LUMHS), Jamshoro, Pakistan, were invited to participate in this study by institutional email; those who volunteered were included. Participation was voluntary, uncompensated, and without effect on teaching, assessment, or educational opportunities. Data were collected from April 12 to 26, 2026. Usability was measured with a standard, unmodified 10-item System Usability

XSL•FO
RenderX

Scale (SUS) in English [6]. PEV was measured with an exploratory 5-item investigator-developed scale (Table 1); all 15 items were scored on a scale of 1 to 5 (1=strongly disagree, 5=strongly agree). The PEV composite (participant means averaged; Cronbach α=.61) is exploratory; item-level results are primary. Sessions were reconstructed from log time using a 30-minute inactivity threshold, on which the session count depends. Because the questionnaire (roll number) and application (seat identifier) used different identifiers, individual SUS scores could not be linked to usage, so the SUS could not be examined by exposure level. Free-text comments were grouped into recurring topics by 1 investigator (descriptive summary; Multimedia Appendix 1). Analyses are descriptive (Python, version 3.11; Python Software Foundation), and no sample-size calculation was conducted.

Table . System Usability Scale (SUS)[a] and perceived educational value (PEV)[b] item results 33 final-year dental students at Liaquat University of Medical and Health Sciences (LUMHS), Jamshoro, Pakistan.

Item	Statement	Mean (SD), scale of 1 to 5
SUS1	I think that I would like to use SSEndoSim frequently.	4.61 (0.83)
SUS2	I found SSEndoSim unnecessarily complex.	1.85 (0.97)
SUS3	I thought SSEndoSim was easy to use.	4.06 (1.17)
SUS4	I think I would need technical support to use SSEndoSim.	1.91 (1.16)
SUS5	I found the various functions in SSEndoSim were well integrated.	4.00 (0.97)
SUS6	I thought there was too much inconsistency in SSEndoSim.	2.03 (1.13)
SUS7	I would imagine that most people would learn to use SSEndoSim very quickly.	4.21 (1.14)
SUS8	I found SSEndoSim very cumbersome to use.	2.21 (1.24)
SUS9	I felt very confident using SSEndoSim.	4.09 (0.98)
SUS10	I needed to learn a lot of things before I could get started with SSEndoSim.	2.85 (1.48)
PEV1	SSEndoSim helped me understand endodontic diagnostic classification better than reading alone.	4.36 (0.86)
PEV2	The immediate automated feedback after each case was educationally valuable.	4.76 (0.50)
PEV3	The scoring system motivated me to think carefully before submitting a diagnosis.	4.70 (0.64)
PEV4	I would recommend SSEndoSim to fellow dental students for self-directed learning.	4.76 (0.44)
PEV5	SSEndoSim should be included as a formal component of the endodontics curriculum.	4.52 (0.71)

[a]SUS composite score (0 - 100): mean 75.30, SD 14.72.

[b]PEV composite score (1-5, exploratory): mean 4.62, SD 0.40.

Ethical Considerations

The research ethics committee of LUMHS, Jamshoro, approved this study (LUMHS/REC/1575; March 15, 2026). Invitations were sent and subsequent written electronic consent sought through the institutional email and questionnaire system. Roll numbers, used only to screen duplicates, were deleted before analysis. Deidentified data were held by the principal investigator alone, who lectured the cohort, had no assessment role, and did not know who participated.

Results

Of 75 invited students, 33 (44%) attempted at least 1 case and submitted a complete questionnaire; of these, 17 (52%) completed all 12 cases. Logs recorded 43 sessions and 423 scored case attempts, comprising 301 unique user-case combinations and 122 repeats, with 16 users (48%) repeating at least 1 case. The median number of unique cases attempted was 12 (IQR 5.5 - 12), and the median time spent using the app was 22.4 minutes (IQR 10.7 - 29.0).

The mean SUS score was 75.30 (SD 14.72; 95% CI 70.08 - 80.52; median 80.0, IQR 66.2 - 87.5; range 45.0 - 100.0). The distribution showed no strong evidence of

XSL•FO
RenderX

departure from normality (Shapiro-Wilk *P*=.35). Individual scores were as follows: 3 (9%) scored below 51, 7 (21%) score from 51 to 67, 6 (18%) scored from 68 to 79, 10 (30%) scored from 80 to 89, and 7 (21%) scored 90 or above (Multimedia Appendix 1). Item means are shown in Table 1; among the negatively worded items, the onboarding item scored highest (2.85, SD 1.48). PEV was high (mean 4.62, SD 0.40), with automated feedback and peer recommendation scoring highest (both 4.76). Of the 19 free-text comments received, the most common topics were feedback and scoring clarity (n=4; Multimedia Appendix 1).

Discussion

SSEndoSim's mean SUS score (75.30, SD 14.72) is comparable to other endodontic and dental diagnostic education tools evaluated with the SUS: the Endo 10 endodontic diagnosis app (median 77.5) [7], an endodontic e-learning application (81.4) [8], and a virtual scenario–based clinical-reasoning system for dental students (70.14) [9]. Samples and administration differ, so values are broadly consistent, not directly comparable. Participants rated the feedback favorably, and immediate feedback is one component associated with deliberate practice [4]; however, these data do not show that use of the app resulted in corrected diagnostic errors in users. Score-based elements may encourage use, but engagement alone does not establish learning benefit [10,11], and high recommendation and curricular-integration ratings reflect learner perception, not curriculum readiness. The open-ended data and scores for the onboarding item point to onboarding as the clearest usability target, and strengthening introductory guidance is the immediate development priority.

This single-institution study used a voluntary, self-selected sample of participants who were possibly more receptive to mobile learning than nonparticipants. Because the developer taught the cohort, perceived-authority and social-desirability effects cannot be excluded. The PEV measure was not pretested and showed modest internal consistency. Logs could not be linked to questionnaire responses, exposure varied, and only 17 participants completed all 12 cases. There was no comparator group or pre-post learning measure. The findings, therefore, concern usability and learner perception, not educational effectiveness.

A free, web-based, guideline-aligned simulator was usable and well regarded by final-year BDS students in a resource-constrained setting. Its browser-based architecture and teacher-editable case bank may be adaptable to other areas of dental education, although this was not examined here. Whether SSEndoSim improves diagnostic competence remains to be tested.

All 15 items were scored on a 5-point Likert scale (1=strongly disagree, 5=strongly agree). Values are raw item means (SD) and are not transformed for display. SUS items 2, 4, 6, 8, and 10 are negatively worded: lower raw means indicate more favorable responses, and these items are reverse scored only when computing the SUS composite. The PEV composite is the mean of each participant's 5 PEV items, averaged across participants, and is exploratory (Cronbach α=.61). All values are descriptive; no hypothesis tests were performed. Full response distributions can be found in Multimedia Appendix 1.

Acknowledgments

The authors used Claude (Anthropic) to assist with coding the SSEndoSim application and for language editing of the manuscript. The clinical content of the application was drawn from published guidelines and authored by the investigators; it was not generated by AI. The authors independently checked all statistical calculations against the raw data. Generative AI was not used to create study data, results, or references. The authors reviewed and approved the final manuscript and took responsibility for its content.

Funding

The authors declared no financial support was received for this work.

Data Availability

The deidentified item-level questionnaire dataset and the aggregate application-log summaries analyzed in this study are available from the corresponding author on reasonable request.

Authors' Contributions

SS conceived the study, developed the application, analyzed the data, and drafted the manuscript. PR contributed to data collection, data analysis, and technical refinement of the application. FK designed and supervised the study. All authors reviewed and approved the final manuscript.

Conflicts of Interest

SS developed the SSEndoSim application and declares this nonfinancial developer interest. The authors declare no financial conflicts of interest.

Multimedia Appendix 1

XSL•FO
RenderX

Development account, 17-point scoring rubric, case-bank composition, application log definitions, distribution of individual SUS scores by band, response distributions for all items, and open-ended comment categories.
[DOCX File, 24 KB - humanfactors_v13i1e100407_app1.docx]

References

1. Glickman GN. AAE Consensus Conference on Diagnostic Terminology: background and perspectives. J Endod 2009 Dec;35(12):1619-1620. [doi: 10.1016/j.joen.2009.09.029] [Medline: 19932336]
2. Duncan HF, Kirkevang LL, Peters OA, et al. Treatment of pulpal and apical disease: The European Society of Endodontology (ESE) S3-level clinical practice guideline. Int Endod J 2023 Oct;56 Suppl 3(S3):238-295. [doi: 10.1111/iej.13974] [Medline: 37772327]
3. Alobaoid MA, Aldowah O, Karobari MI. Endodontic clinical diagnostic skills amongst undergraduate dental students: cross-sectional study. Healthcare (Basel) 2022 Aug 30;10(9):1655. [doi: 10.3390/healthcare10091655] [Medline: 36141267]
4. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. Acad Emerg Med 2008 Nov;15(11):988-994. [doi: 10.1111/j.1553-2712.2008.00227.x] [Medline: 18778378]
5. Suresh S. EndoSim. URL: https://ss-endosim.netlify.app/ [accessed 2026-09-02]
6. Jordan PW, Thomas B, McClelland IL. 'SUS: A "Quick and Dirty" Usability Scale', in Usability Evaluation In Industry: CRC Press; 1996:207-212. [doi: 10.1201/9781498710411-35]
7. Abuabara A, de Castro JP, Locks MEN, et al. Evaluation of Endo 10 mobile application as diagnostic tool in endodontics. J Clin Exp Dent 2023 Aug;15(8):e612-e620. [doi: 10.4317/jced.60342] [Medline: 37674608]
8. Färber CM, Lemos M, Said Yekta-Michael S. Effect of an endodontic e-learning application on students' performance during their first root canal treatment on real patients: a pilot study. BMC Med Educ 2022 May 23;22(1):394. [doi: 10.1186/s12909-022-03463-y] [Medline: 35606729]
9. Cao Y, Yuan S, Pan Y, et al. Assessing a virtual scenario-based training system for enhancing clinical reasoning and communication in dental education. BMC Oral Health 2026 Jan 21;26(1):367. [doi: 10.1186/s12903-026-07723-x] [Medline: 41566298]
10. Alkhattab OR. Gamification in endodontic education: a pilot study on student engagement and perceived learning outcomes. BMC Med Educ 2025 Aug 28;25(1):1214. [doi: 10.1186/s12909-025-07753-z] [Medline: 40877855]
11. Papadakis S, Karakose T. Gamification and student achievement: potential benefits, limitations, and effective use in educational environments. Educ Process Int J 2025;19(1):1. [doi: 10.22521/edupij.2025.19.529]

Abbreviations

BDS: bachelor of dental surgery
LUMHS: Liaquat University of Medical and Health Sciences
PEV: perceived educational value
SUS: System Usability Scale

Edited by S Law; submitted 05.May.2026; peer-reviewed by PCA Watanabe, S Papadakis; revised version received 28.Jul.2026; accepted 13.Aug.2026; published 10.Sep.2026.

Please cite as:
Suresh S, Rani P, Kalhoro F
Development and Initial Usability Evaluation of SSEndoSim, a Free, Web-Based Endodontic Diagnostic Simulator: Cross-Sectional Study
JMIR Hum Factors 2026;13:e100407
URL: https://humanfactors.jmir.org/2026/1/e100407
doi:10.2196/100407

© Sarang Suresh, Priya Rani, Feroze Kalhoro. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 10.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Digital App-Based Intervention (Platoniq) to Reduce Loneliness: Single-Arm Pilot Study

Silvan Hornstein[1,2], MSc, PhD; Jasmina Schmidt[1], BSc; Lara Sachs[1], MSc; Tobias Krieger[3], Prof Dr; Daniel Stachnik[1], MSc; Ulrike Lueken[1,4], Prof Dr

[1]Department of Psychology, Humboldt-Universität zu Berlin, Rudower Chausee 18, Berlin, Germany

[2]SHS Social Health Solutions UG (haftungsbeschränkt), Berlin, Germany

[3]Department of Clinical Psychology and Psychotherapy, Institute of Psychology, University of Bern, Bern, Switzerland

[4]German Center for Mental Health (DZPG), Partner site Berlin-Potsdam, Berlin, Germany

Corresponding Author:
Silvan Hornstein, MSc, PhD
Department of Psychology, Humboldt-Universität zu Berlin, Rudower Chausee 18, Berlin, Germany

Abstract

This single-arm pilot study of a psychological, app-based intervention (platoniq) found significantly reduced loneliness after 4 and 8 weeks.

Trial Registration: German Clinical Trials Register DRKS00038467; https://drks.de/search/en/trial/DRKS00038467/details

(JMIR Hum Factors 2026;13:e99374) doi:10.2196/99374

KEYWORDS

loneliness; social health; digital health; pilot study; digital mental health

Introduction

Loneliness, the subjectively perceived lack of social relationships, is affecting a significant share of people across countries and age groups [1]. As this is associated with numerous adverse health consequences [2], effective interventions to tackle loneliness are highly needed. Although a substantial body of evidence has examined such interventions and found that psychological approaches in particular can effectively reduce loneliness [3], considerably more research is still required. Loneliness is linked to hypervigilance and negative social cognition, reinforcing withdrawal in a cycle [4], which cognitive-behavioral techniques such as social skills training and cognitive restructuring aim to break. Particularly, low-threshold digital interventions are an interesting yet underinvestigated approach with few available studies [5]. Consequently, the objective of this preregistered [6] pilot study was the evaluation of a digital app-based intervention against loneliness.

Methods

Ethical Considerations

Ethical approval for this study was granted by the ethical committee of the Institute for Psychology of the Humboldt University Berlin (identifier: 2025 - 58).

Study Design

This pilot study was designed as a one-arm, noncontrolled trial. Loneliness, the primary outcome, was assessed using the 9-item version of the UCLA Loneliness Scale (UCLA-9) [7] at baseline, 4 weeks, and 8 weeks. Additionally, perceived social support [8], perceived stress [9], and social competencies [10] were measured at baseline and at week 8 as secondary outcomes.

Sample

Study participants were recruited through convenience sampling via several channels, such as social media and university newsletters. A total of 72 participants matched our inclusion criteria (UCLA-9 score ≥18, reflecting preexisting low-to-moderate loneliness; ≥18 years old; German-speaking; no acute mental health crisis; and not currently seeking treatment), completed at least 1 of the follow-up questionnaires, and accessed the intervention with their phone. The median age was 35 (IQR 27.0-48.75; mean 39.3, SD 14.23) years and 60 (83.3%) participants were female. Of the 72 participants, 10 (13.9%) reported a migration background; 40 (55.6%) were employed full-time; and the sample was highly educated, with the majority (n=50, 69.4%) holding a university degree.

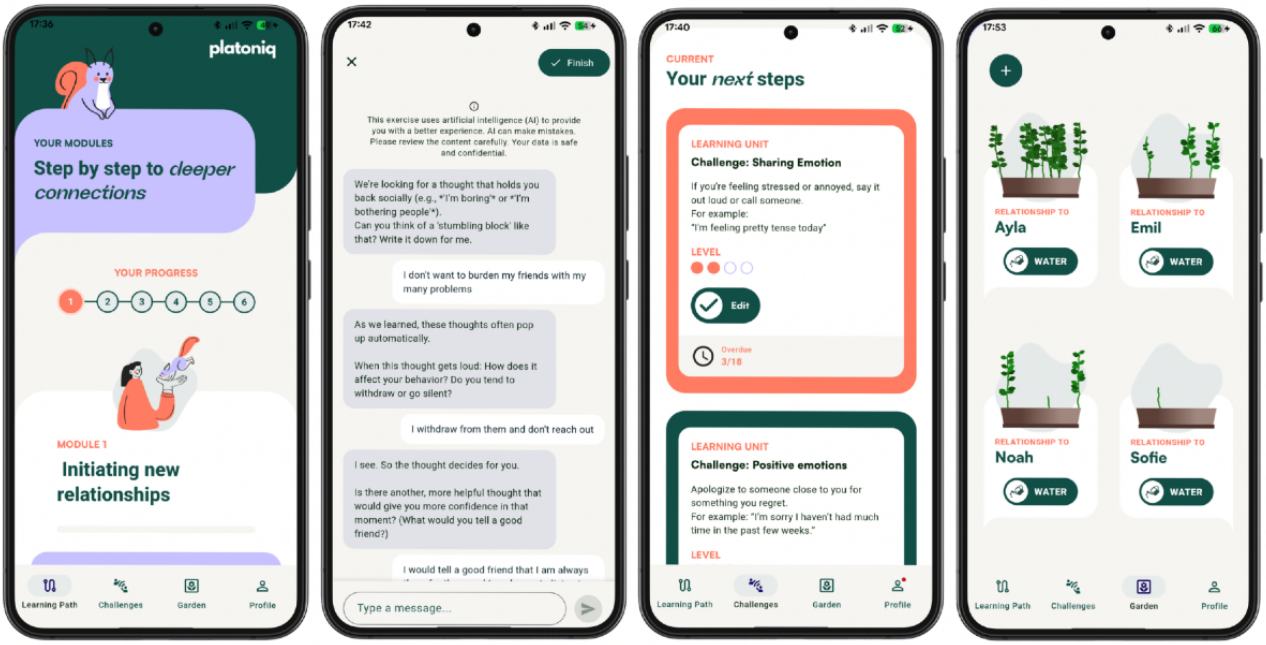
Intervention

The app-based intervention platoniq (SHS Social Health Solutions UG) was used in this study. The digital intervention is structured into 6 modules, addressing diverse aspects of social relationships and loneliness. Exercises are adapted from social skills training and cognitive behavioral therapy, following a

XSL•FO
RenderX

cognitive model of loneliness [4]. One new module is unlocked each week, covering the following topics in sequence: initiating relationships, deepening relationships, emotional support, conflict management, setting boundaries, and social relationships and own needs. Within each module, exercises have to be completed in sequential order (see Figure 1). In addition, the app includes a tracking mechanism for social relationships, allowing participants to monitor their social interactions over the course of the program. Notifications reminded participants to use the app, and chat support was available for technical questions.

Figure 1. Screenshots of the intervention app, platoniq, which includes exercises (eg, social skills training and cognitive restructuring), behavioral experiments, and a tracking mechanism.



Results

Out of the 72 included participants, 67 (93%) provided answers to the UCLA-9 after 4 weeks from the intervention start. UCLA-9 scores decreased from baseline (mean 25.6, SD 3.7) to week 4 (mean 23.5, SD 4.2). A paired *t* test showed a significant medium-to-large effect for loneliness (t_{66}=5.8; *P*<.001; *d*=0.71), with an average reported decrease on the UCLA-9 of 2.2 (SD 3.1). Approximately 75% (50/67) of participants reported a decrease, and 45% (30/67) of all participants reported a decrease of at least 3 points. This has been considered clinically meaningful in past work, following a reliable change index derived for the UCLA-9 in a comparable sample [4]. Higher baseline UCLA-9 scores were associated with greater reduction (*r*=0.25; *P*=.04). The effect remained stable for the 44 participants who responded to the questionnaires after 8 weeks, with scores decreasing from baseline (mean 26.2, SD 4.2) to week 8 (mean 24.0, SD 5.2; t_{43}=3.8; *P*<.001; *d*=0.57), reporting an average reduction of 2.21 (SD 3.9). Additionally, effects were found for interpersonal competencies (t_{43}=−3.1; *P*=.004; *d*=−0.46) as well as perceived social support (t_{43}=−2.4; *P*=.02; *d*=−0.37). However, while there was an average decrease of 0.77 (SD 6.8) in perceived stress, this effect was not statistically significant (t_{43}=0.75; *P*=0.45; *d*=0.13). No significant baseline differences were found between completers and noncompleters (ie, no 8-week questionnaire) at 8 weeks across demographic or clinical variables, suggesting limited attrition bias. Noncompleters showed markedly lower overall engagement (mean 14.8, SD 13.3 vs 30.9, SD 20.1 events; $t_{69.87}$=4.10; *P*<.001).

On average, the included participants used the app for approximately 128 (SD 119) minutes, corresponding to a weekly usage of around 21 (SD 19.8) minutes. On average, 25 (SD 19.3) exercises were completed, as well as 6 (SD 4.6) behavioral experiments, and the tracking mechanism was used 18 (SD 19.3) times. The small (*r*=0.14) correlation between usage time and reduction was nonsignificant in the given sample (*P*=.24).

Discussion

In this uncontrolled pilot study, an app-based digital health intervention targeting loneliness in individuals with at least low-to-moderate levels of loneliness showed preliminary signals of reduced loneliness. The observed within-group effect size (*d*=0.71) falls within the range reported across existing internet-based interventions (*d*=0.33 to *d*=1.47) [5], corroborating the general promise of this intervention format. However, given the single-arm design, this interpretation is limited, and randomized controlled trials are needed to confirm the intervention's effectiveness.

The sample was predominantly female and highly educated, with considerable dropout, limiting generalizability. Additionally, user satisfaction and barriers to use were not systematically assessed. Finally, the association between

XSL•FO
RenderX

baseline severity and reduction may partly reflect regression to the mean. Interestingly, contrary to our initial hypothesis, the effect on perceived stress in this study was not significant. This may be due to limited statistical power to detect small effects. Taken together, these preliminary results suggest app-based cognitive behavior therapy interventions may offer a promising approach to addressing loneliness as a pressing public health challenge.

Generative AI Use

The authors used Claude (Anthropic) to assist with language editing and wording of the manuscript. The tool was not used to generate scientific content, analyze data, or interpret results. All authors reviewed and edited the output and take full responsibility for the content of this publication.

Funding

The article processing charge was funded by the Open Access Publication Fund of Humboldt-Universität zu Berlin.

Data Availability

Due to legal and privacy restrictions, the user data used in this study cannot be shared. Access to the data is limited to authorized researchers under strict confidentiality agreements.

Authors' Contributions

SH conceptualized the study, developed the methodology, conducted the data collection, curated and analyzed the data, administered the project, and wrote the original draft of the manuscript. JS contributed to the conceptualization, data collection, and project administration. LS and DS contributed to the development of the intervention and provided resources. TK advised on the study design and methodology. UL supervised the project. All authors critically reviewed and approved the final version of the manuscript.

Conflicts of Interest

SH, LS, and DS are cofounders and stakeholders of SHS Social Health Solutions UG, which provided the intervention that was evaluated in this study. JS, UL, and TK report no conflicts of interest.

References

1. Surkalim DL, Luo M, Eres R, et al. The prevalence of loneliness across 113 countries: systematic review and meta-analysis. BMJ 2022 Feb 9;376:e067068. [doi: 10.1136/bmj-2021-067068] [Medline: 35140066]
2. Leigh-Hunt N, Bagguley D, Bash K, et al. An overview of systematic reviews on the public health consequences of social isolation and loneliness. Public Health 2017 Nov;152:157-171. [doi: 10.1016/j.puhe.2017.07.035] [Medline: 28915435]
3. Lasgaard M, Qualter P, Løvschall C, et al. Are loneliness interventions effective for reducing loneliness? A meta-analytic review of 280 studies. Am Psychol 2026 Jan;81(1):36-52. [doi: 10.1037/amp0001578] [Medline: 41129341]
4. Skoko A, Seewer N, Mund M, Krieger T. Revisiting the cognitive and behavioral aspects of loneliness: insights from different measurement approaches. PLoS ONE 2025 Apr 28;20(4):e0321931. [doi: 10.1371/journal.pone.0321931] [Medline: 40294013]
5. Imwinkelried M, Seewer N, Berger T, Krieger T. Internet-based self-help for loneliness-a systematic review [Article in German]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2024 Oct;67(10):1161-1171. [doi: 10.1007/s00103-024-03934-1] [Medline: 39117979]
6. Trial DRKS00038467: evaluation study social health app platoniq. German Clinical Trials Register. URL: https://drks.de/search/en/trial/DRKS00038467 [accessed 2026-04-24]
7. Luhmann M, Bohn J, Holtmann J, Koch T, Eid M. I'm lonely, can't you tell? Convergent validity of self- and informant ratings of loneliness. J Res Pers 2016 Apr;61:50-60. [doi: 10.1016/j.jrp.2016.02.002]
8. Kliem S, Mößle T, Rehbein F, Hellmann DF, Zenger M, Brähler E. A brief form of the Perceived Social Support Questionnaire (F-SozU) was developed, validated, and standardized. J Clin Epidemiol 2015 May;68(5):551-562. [doi: 10.1016/j.jclinepi.2014.11.003] [Medline: 25499982]
9. Schneider EE, Schönfelder S, Domke-Wolf M, Wessa M. Measuring stress in clinical and nonclinical subjects using a German adaptation of the Perceived Stress Scale. Int J Clin Health Psychol 2020 May;20(2):173-181. [doi: 10.1016/j.ijchp.2020.03.004] [Medline: 32550857]
10. Coroiu A, Meyer A, Gomez-Garibello CA, Brähler E, Hessel A, Körner A. Brief form of the Interpersonal Competence Questionnaire (ICQ-15): development and preliminary validation with a German population sample. Eur J Psychol Assess 2015 Oct;31(4):272-279. [doi: 10.1027/1015-5759/a000234]

Abbreviations

UCLA-9: 9-item version of the UCLA Loneliness Scale

XSL•FO
RenderX

Edited by A Kushniruk; submitted 24.Apr.2026; peer-reviewed by AKC Chau, C Baxter, S Hassan; revised version received 10.Aug.2026; accepted 08.Sep.2026; published 22.Oct.2026.

Please cite as:
Hornstein S, Schmidt J, Sachs L, Krieger T, Stachnik D, Lueken U
A Digital App-Based Intervention (Platoniq) to Reduce Loneliness: Single-Arm Pilot Study
JMIR Hum Factors 2026;13:e99374
URL: https://humanfactors.jmir.org/2026/1/e99374
doi:10.2196/99374

© Silvan Hornstein, Jasmina Schmidt, Lara Sachs, Tobias Krieger, Daniel Stachnik, Ulrike Lueken. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 22.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Age-Related Differences in Mental Health Status at Psychiatric Discharge Among Patients Across 10 Acute Care Facilities in Alberta, Canada: Cross-Sectional Study

Wanying Mao[1], PhD; Ernest Owusu[2], PhD; Reham Shalaby[2], PhD, MD; Hossam Eldin Elgendy[2], PhD; Belinda Agyapong[2], PhD; Ejemai Eboreime[3], MD, PhD; Peter H Silverstone[2], MD, PhD, Prof Dr; Pierre Chue[2], PhD, Prof Dr; Wesley Vuong[4], MPH; Arto Ohinmaa[5], PhD; Andrew J Greenshaw[2], PhD, Prof Dr; Vincent I O Agyapong[2,3], MD, PhD, Prof Dr

[1]AI for Mental Health Program, Department of Psychiatry, St. Michael's Hospital, Unity Health Toronto, Toronto, ON, Canada

[2]Department of Psychiatry, University of Alberta, Edmonton, AB, Canada

[3]Department of Psychiatry, Dalhousie University, 5909 Veterans Memorial Lane, Abbie J. Lane Memorial Building QEII Health Sciences Centre, Halifax, NS, Canada

[4]Clinical Operations Informatics Office, Health Shared Services, Edmonton, AB, Canada

[5]School of Public Health, University of Alberta, Edmonyon, AB, Canada

Corresponding Author:
Vincent I O Agyapong, MD, PhD, Prof Dr
Department of Psychiatry, University of Alberta, Edmonton, AB, Canada

Abstract

Background: The transition from psychiatric inpatient care to community settings is associated with heightened vulnerability and increased risk of adverse mental health outcomes. Understanding age-related differences in mental health status at discharge may inform the targeting of postdischarge interventions and potentially improve long-term outcomes.

Objective: This study examined the associations between age and baseline demographic, clinical, and mental health characteristics among participants enrolled in the Text4Support program following discharge from psychiatric inpatient care.

Methods: Baseline data collected between March 8, 2022, and February 29, 2024, from 1106 participants discharged from inpatient psychiatric units across Alberta, Canada, were analyzed. Participants were grouped by age (18 - 25, 26 - 40, and >40 years). Mental health outcomes were assessed using validated self-report measures. Chi-square tests and 1-way ANOVA or Welch ANOVA were used to examine age-related differences in categorized and continuous outcomes, respectively.

Results: Significant age-related differences were observed across several mental health outcomes. Younger participants (18 - 25 years) reported higher levels of depression, anxiety, and suicidal ideation, along with lower resilience and recovery scores compared with participants aged older than 40 years. Differences in quality of life were also observed, with younger adults reporting greater anxiety/depression and participants aged older than 40 years reporting more mobility-related problems.

Conclusions: Marked age-related disparities in mental health status were evident at the point of psychiatric discharge, with younger adults demonstrating greater vulnerability. These findings underscore the importance of age-informed risk stratification and tailored transitional care strategies following inpatient psychiatric treatment and suggest that different strategies may be required depending upon age.

Trial Registration: ClinicalTrials.gov NCT05133726; https://clinicaltrials.gov/study/NCT05133726

(JMIR Hum Factors 2026;13:e92360) doi:10.2196/92360

KEYWORDS

mental health; digital mental health; Text4Support; psychiatric care; aging

Introduction

The transition from psychiatric inpatient care to community-based mental health services represents a particularly vulnerable period for individuals experiencing mental health conditions [1-3]. This phase is frequently associated with adverse clinical outcomes, including symptom exacerbation, relapse, hospital readmission, and an increased mortality risk [1-4]. Substantial evidence indicates that patients discharged from psychiatric facilities encounter significant challenges in the early postdischarge period, reflecting unmet clinical needs and limited continuity of care [1-4]. Understanding the clinical characteristics of patients at the point

XSL•FO
RenderX

of discharge is therefore essential for informing targeted interventions and improving postdischarge outcomes.

Age is a key demographic factor that has been shown to influence the presentation, severity, and course of mental health conditions across the lifespan [5,6]. Younger and older adults often differ markedly in symptom profiles, coping strategies, service use patterns, and treatment engagement [6]. For example, younger adults may present with higher levels of affective symptoms and psychosocial stressors, whereas older adults may experience greater functional impairment, comorbidity, and barriers to accessing mental health care [6-8]. Despite these differences, age-related variations in mental health symptom burden at the point of psychiatric discharge remain underexplored, particularly within real-world inpatient populations.

Age-sensitive mental health care is currently most clearly articulated for particular stages of the adult life course. Canadian guidance for emerging adults recommends a developmentally responsive continuum of care based on individuals' changing needs rather than rigid age-based service thresholds [9]. Guidance for later-life mental health care similarly emphasizes integrated, recovery-oriented coordination among community services, primary care, general mental health, and specialized services [10]. More broadly, a life-course approach recognizes that developmental, functional, and social needs change throughout adulthood and should be considered in service planning [11]. Nevertheless, specific guidance for age-responsive psychiatric discharge planning across the full adult age range remains limited, supporting the need to examine age-related differences at this important transition in care.

In the Canadian context, psychiatric inpatient populations represent a clinically complex and heterogeneous group. According to the Canadian Institute for Health Information (CIHI), individuals hospitalized for mental health and substance use disorders experience some of the highest rates of early readmission following discharge [12]. These patterns highlight the importance of identifying baseline factors that may influence postdischarge trajectories. However, most existing studies focus on postdischarge outcomes or intervention effects, with relatively limited attention to how demographic characteristics such as age are associated with mental health status at the time of inpatient recruitment.

In Alberta, Canada, the Text4Support program was developed to address gaps in community-based mental health support through the provision of daily supportive text messages, alongside complementary peer support initiatives [13]. As part of the broader Text4Support research program, baseline data were collected from psychiatric inpatients at the time of recruitment, prior to the initiation of any intervention. This dataset provides a unique opportunity to examine age-related differences in mental health symptom severity within a real-world inpatient cohort at a critical transition point in care. Although the current analysis does not examine intervention effects, characterizing age-related differences at discharge provides important context for interpreting subsequent findings from the parent trial and may identify age-related needs to consider when developing and evaluating postdischarge supports.

The present study uses baseline data from the Text4Support project to examine the relationship between age and multiple mental health outcomes among psychiatric inpatients at the time of recruitment. Using validated instruments, including the Generalized Anxiety Disorder 7-item scale (GAD-7), the Patient Health Questionnaire-9 (PHQ-9), the World Health Organization-Five Well-Being Index (WHO-5), the Brief Resilience Scale (BRS), and the EQ-5D-5L, this study aims to assess age-related differences in anxiety, depression, overall well-being, resilience, and quality of life. By characterizing baseline mental health symptom profiles across age groups, this research seeks to inform the development of age-sensitive interventions and contribute to a more nuanced understanding of patient needs during the transition from inpatient to community-based mental health care. The analysis was intended to provide a descriptive, age-stratified profile of baseline mental health status rather than to estimate an independent causal effect of age or develop a multivariable prediction model.

Methods

Study Setting and Design

This study draws on data collected during the initial phase of an innovative program designed to enhance support and care for individuals discharged from inpatient mental health using e-mental health tools and peer worker services. The broader program adopted a pragmatic stepped-wedge cluster-randomized design and was implemented across 10 major health care sites in Alberta, located in Edmonton, Calgary, and Grande Prairie, between March 8, 2022, and February 29, 2024 [13]. Its purpose was to evaluate the effectiveness of postdischarge interventions, specifically supportive text messaging (Text4Support) alone or in combination with peer support, compared with treatment as usual [13]. This study involved secondary analysis of baseline data from the stepped-wedge trial and examines associations between age and multiple mental health measures among program participants. The present secondary analysis is restricted to baseline assessments completed before the delivery of Text4Support or peer support and therefore does not evaluate the effects of either intervention.

Data Collection and Inclusion Criteria

A face-to-face recruitment process was implemented with nurse managers and clinical staff assisting the research team by identifying potential participants who were nearing discharge from inpatient psychiatric units. Eligible individuals were provided with detailed study information, including an information leaflet. Those who agreed to participate signed a paper-based consent form and completed a self-administered questionnaire on a tablet device. The survey was developed in REDCap (Vanderbilt University), a secure, browser-based platform for creating and managing online surveys and research databases, with support from members of the research team [14]. Data collected included sociodemographic characteristics (eg, age, gender, ethnicity, relationship status, employment status, education level, and housing status) and clinical information (eg, discharge diagnosis, anxiety, depression, and

XSL•FO
RenderX

well-being levels). All enrolled participants were included in the analysis.

Participants were required to have a diagnosed mental health condition, be preparing for discharge from an inpatient psychiatric unit, be 18 years of age or older, and own a mobile device with an active phone number. They also needed the ability to receive and read English-language text messages and to provide written informed consent. Individuals planning to travel outside the city during the 6-month follow-up period were excluded due to the in-person nature of the peer support intervention. Phone numbers and health care numbers were collected as primary identifiers.

Ethical Considerations

Ethical approval was obtained from the University of Alberta Health Research Ethics Board (reference number: Pro00111459), with additional operational approval granted by the regional health authority. Written informed consent was obtained from all participants.

Sample Size Calculation

With 3 age clusters, a 2-tailed α level of 0.05% and 80% power (corresponding to a Type II error probability, β=0.20), and assuming a moderate effect size corresponding to Cohen *f*=0.25 (approximately equivalent to a pairwise Cohen *d* of 0.5 with a common SD of 1.0), we estimated that a total sample size of 159 participants (53 per age cluster) would be required to detect statistically significant differences between groups using a 1-way ANOVA [15].

Outcome Measures

GAD-7

The GAD-7 was used to assess participants' anxiety symptoms. This self-report measure consists of 7 items evaluating the frequency and severity of anxiety symptoms experienced over the preceding 2 weeks. Each item is rated on a 4-point Likert scale from 0 ("not at all") to 3 ("nearly every day"), producing a total score ranging from 0 to 21, with higher scores indicating more severe anxiety [16]. The GAD-7 is widely applied in both clinical and research settings due to its strong psychometric properties. Using the recommended cutoff score of 10, the scale demonstrates a sensitivity of 89% and a specificity of 82% for detecting generalized anxiety disorder. It also shows strong test-retest reliability (intraclass correlation coefficient=0.83) and excellent internal consistency (Cronbach α=0.92) [16,17].

PHQ-9

The PHQ-9 was administered to assess depressive symptoms. Derived from the 9 *DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition)* diagnostic criteria for major depressive disorder, this self-administered tool measures the frequency of depressive symptoms over the past 2 weeks. Items are rated from 0 ("Never") to 3 ("Nearly every day"), yielding total scores between 0 and 27 [18,19]. The PHQ-9 functions both as a diagnostic aid for major depressive disorder and as a measure of symptom severity. A score of 10 or above is commonly recommended to indicate clinically significant depression. Severity categories include mild (5), moderate (10), moderately severe (15), and severe (20). The scale has demonstrated good internal consistency (Cronbach α=0.85) and strong convergent validity, showing moderate-to-high correlations with measures such as the WHO-5, Hospital Anxiety and Depression Scale (HADS)–depression, HADS-anxiety, and GAD-7. At a cutoff of ≥10, sensitivity and specificity are both approximately 88% [18-20].

Item 3 of the PHQ-9 ("PHQ sleep") [21,22] and item 9 (suicidal ideation) [23] provide specific indicators for sleep disturbance and suicide risk, respectively. The sleep item assesses difficulties falling asleep, staying asleep, or oversleeping in the prior 2 weeks. Research in primary care settings supports its effectiveness as a brief yet reliable screening tool for sleep problems without adding substantial response burden [21,22,24]. For instance, MacGregor et al [24] reported strong correlations between PHQ sleep scores and the Insomnia Severity Index, with a score of 1 offering an optimal balance between sensitivity and specificity.

PHQ-9 item 9 assesses suicidal ideation ("Thoughts that you would be better off dead or hurting yourself in some way") and serves as a significant predictor of suicide risk independent of overall depression severity [23]. Higher scores on this item have been associated with markedly increased risk of suicide attempts and deaths for up to 2 years following assessment [25]. The item's sensitivity (0.69) and specificity (0.84) further support its value as a routine screening indicator for suicide risk in primary care, facilitating earlier intervention for individuals who might otherwise go unrecognized [26].

WHO-5

The WHO-5 was used to evaluate participants' overall mental well-being. This brief, positively worded self-report measure comprises 5 items reflecting subjective well-being, such as "I have felt cheerful and in good spirits," "I have felt calm and relaxed," and "My daily life has been filled with things that interest me" [27]. Each item is rated on a 6-point Likert scale from 0 ("not at all") to 5 ("all the time"), and raw scores are converted to a standardized scale ranging from 0 (lowest well-being) to 100 (highest well-being). Scores below 50 indicate reduced well-being and suggest the need for further clinical assessment. The WHO-5 demonstrates strong internal consistency (Cronbach α=0.90) and good convergent validity, including a significant negative correlation with the PHQ-9 ($r=-0.73$; $P<.001$).

BRS

The BRS was used to measure participants' resilience [28]. The scale contains 6 items, 3 positively worded and 3 negatively worded, rated on a 5-point Likert scale from 1 to 5, resulting in total scores ranging from 6 to 30 [29]. The BRS has acceptable internal consistency (α=0.66) and test-retest reliability ($r=0.67$) [29]. A respondent's final score is calculated by dividing the total score by the number of completed items, with higher values indicating greater resilience.

Previous research classifies individual BRS mean item scores from 1.00 to 2.99 as low resilience, 3.00 to 4.30 as normal resilience, and 4.31 to 5.00 as high resilience [28,29]. These values represent individual-level classification thresholds rather than group means; therefore, SDs are not applicable. In the

XSL•FO
RenderX

present study, the continuous BRS total score was used in analyses comparing resilience levels across age groups. For categorical analyses, individual mean item scores in the normal and high ranges (3.00 - 5.00) were combined into a single category and compared with scores in the low-resilience range (1.00 - 2.99).

EQ-5D-5L

The EQ-5D-5L, a widely used preference-based measure of health-related quality of life, was also administered. The instrument includes a descriptive system comprising 5 health dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression [30,31]. Each dimension has 5 response levels ranging from 1 (no problems) to 5 (extreme problems). For this study, responses were dichotomized into 2 categories: “No Problems” (level 1) and “Problems” (levels 2 - 5). The EQ-5D-5L is recognized for its reliability and validity in describing health status across diverse populations, with test-retest reliability reported between 0.65 and 0.91 [32,33].

Statistical Analysis

All statistical analyses were conducted using SPSS software (version 25; IBM Corp) [34]. Descriptive statistics were used to summarize participants’ sociodemographic and clinical characteristics by age group, with categorical variables presented as frequencies and percentages and continuous variables reported as means and SDs.

Chi-square tests were performed to examine associations between age groups and categorized mental health measures, assessing whether the distributions of these indicators differed significantly across age groups.

Age was categorized as 18 to 25, 26 to 40, and >40 years. The 18- to 25-year category was selected to represent emerging adulthood, a developmental period characterized by major transitions in education, employment, relationships, and independent living [35,36]. The remaining cut points were selected pragmatically to permit comparisons across subsequent adult age ranges while maintaining adequate sample sizes in each group. These categories are study-defined and do not represent established clinical thresholds. In particular, the older than 40 years category includes both middle-aged and older participants and was therefore not interpreted as a uniform older-adult group. For continuous mental health outcomes, including anxiety, depression, sleep disturbance, suicidal ideation, resilience, well-being, and health-related quality of life, 1-way ANOVA was conducted to compare mean scores across age groups. When the assumption of homogeneity of variance was violated, Welch ANOVA was applied. For outcomes demonstrating significant overall group differences, post hoc pairwise comparisons were conducted using Tukey HSD and Games-Howell tests to identify specific age-group differences. The analyses were designed as unadjusted, descriptive comparisons across age groups. Education level, employment status, and discharge diagnosis were summarized by age group to characterize the sample but were not entered as covariates because independent predictor modeling was outside the scope of the present analysis.

Results

Baseline Data Analysis

Participants ranged in age from 18 to 81 years, with a mean age of 34.06 years and an SD of 14.05. Of the 1106 participants, 403 (36.4%) were aged 18 to 25 years, 378 (34.2%) were aged 26 to 40 years, and 325 (29.4%) were over 40 years of age. Among respondents, 468 (42.3%) identified as male, 606 (54.8%) as female, and 32 (2.9%) as another gender. The majority (n=681, 61.6%) identified as Caucasian. More than half (n=568, 51.4%) reported completing a high school education, 652 (59.0%) were single, and 585 (52.9%) were unemployed. In terms of living arrangements, 459 (41.5%) participants lived with family or friends. A total of 294 (26.6%) respondents reported having a diagnosis of depression. A detailed summary of participant characteristics is presented in Table 1.

XSL•FO
RenderX

Table . Distribution of sociodemographic and clinical characteristics among the study participants.

Variables	Age groups (years), n (%)			
	18-25 (n=403)	26-40 (n=378)	>40 (n=325)	Total (N=1106)
Gender				
Male	157 (39.0)	168 (44.4)	143 (44.0)	468 (42.3)
Female	226 (56.1)	199 (52.6)	181 (55.7)	606 (54.8)
Other	20 (5.0)	11 (2.9)	1 (0.3)	32 (2.9)
Ethnicity				
Caucasian	210 (52.1)	235 (62.2)	236 (72.6)	681 (61.6)
Indigenous	43 (10.7)	31 (8.2)	28 (8.6)	102 (9.2)
Black	63 (15.6)	41 (10.8)	16 (4.9)	120 (10.8)
Asian	60 (14.9)	40 (10.6)	25 (7.7)	125 (11.3)
Mixed/other	27 (6.7)	31 (8.2)	20 (6.2)	78 (7.1)
Education level				
Less than high school	15 (3.7)	16 (4.2)	10 (3.1)	41 (3.7)
High school diploma	297 (73.7)	161 (42.6)	110 (33.8)	568 (51.4)
Postsecondary education	77 (19.1)	196 (51.9)	190 (58.5)	463 (41.9)
Prefer not to say	14 (3.5)	5 (1.3)	15 (4.6)	34 (3.1)
Relationship status				
Married/partnered/common-law	75 (18.6)	110 (29.1)	138 (42.5)	323 (29.2)
Single	306 (75.9)	237 (62.7)	109 (33.5)	652 (59.0)
Separated or divorced	0 (0)	24 (6.3)	63 (19.4)	87 (7.9)
Widowed	0 (0)	3 (0.8)	9 (2.8)	12 (1.1)
Prefer not to say	22 (5.5)	4 (1.1)	6 (1.8)	32 (2.9)
Employment status				
Employed	91 (22.6)	134 (35.4)	103 (31.7)	328 (29.7)
Unemployed	215 (53.3)	222 (58.7)	148 (45.5)	585 (52.9)
Student	78 (19.4)	10 (2.6)	1 (0.3)	89 (8.0)
Retired	0 (0)	4 (1.1)	64 (19.7)	68 (6.1)
Other	19 (4.7)	8 (2.1)	9 (2.8)	36 (3.3)
Housing status				
Own home	16 (4.0)	64 (16.9)	141 (43.1)	221 (20.0)
Rented accommodation	82 (20.3)	149 (39.4)	127 (39.1)	358 (32.4)
Live with family or friend	285 (70.7)	139 (36.8)	35 (10.8)	459 (41.5)
Couch surfing/shelter/street/other	20 (5.0)	26 (6.9)	22 (6.8)	68 (6.1)
Primary mental health diagnosis				
Depression	99 (24.6)	84 (22.2)	111 (34.2)	294 (26.6)
Bipolar disorder	68 (16.9)	89 (23.5)	66 (20.3)	223 (20.2)
Anxiety	56 (13.9)	40 (10.6)	47 (14.5)	143 (12.9)
Schizophrenia	63 (15.6)	80 (21.2)	35 (10.8)	178 (16.1)
Personality disorder	70 (17.4)	34 (9.0)	3 (0.9)	107 (9.7)
Substance use disorder	4 (1.0)	4 (1.1)	7 (2.2)	15 (1.4)
Other	43 (10.7)	47 (12.4)	56 (17.2)	146 (13.2)

XSL•FO
RenderX

Chi-Squared Analysis

Tables 2 and 3 present the chi-square analyses examining baseline clinical measures across 3 age groups (18 - 25, 26 - 40, and >40 years). Several mental health indicators showed significant variation by age. Resilience, as measured using the BRS, differed significantly across groups (P=.04). Several categorical mental health outcomes differed significantly across age groups. Low resilience was reported by 60.8% (n=245) of participants aged 18 to 25 years, compared with 52.6% (n=199) of those aged 26 to 40 years and 53.4% (n=173) of those aged older than 40 years (χ^2_2=6.36; P=.04). Poor well-being was reported by 52.5% (n=211), 42.9% (n=162), and 50.9% (n=165) of the 3 groups, respectively (χ^2_2=8.12; P=.02).

Table . Clinical baseline measures by age group (BRS[a], WHO-5[b], PHQ-9[c], and suicidal ideation).

Variables	Age groups (years), n (%)			χ^2 (*df*)	*P* value
	18-25[d]	26-40	>40		
BRS				6.36 (2)	.04[e]
Normal-high resilience	158 (39.2)	179 (47.4)	151 (46.6)		
Low resilience	245 (60.8)	199 (52.6)	173 (53.4)		
WHO-5				8.12 (2)	.02[e]
Good well-being	191 (47.5)	216 (57.1)	159 (49.1)		
Poor well-being	211 (52.5)	162 (42.9)	165 (50.9)		
PHQ-9				6.81 (2)	.03[e]
Unlikely depression	153 (38.1)	169 (44.7)	153 (47.2)		
Likely depression	249 (61.9)	209 (55.3)	171 (52.8)		
Sleep issues				4.49 (2)	.11
No sleep issues	70 (17.4)	78 (20.6)	77 (23.8)		
With sleep issues	332 (82.6)	300 (79.4)	247 (76.2)		
Suicidal ideation				19.09 (2)	<.001[e]
No	169 (42.0)	210 (55.6)	181 (55.9)		
Yes	233 (58.0)	168 (44.4)	143 (44.1)		
GAD-7[f]				3.57 (2)	.17
Unlikely anxiety	246 (61.3)	234 (62.1)	216 (67.7)		
Likely anxiety	155 (38.7)	143 (37.9)	103 (32.3)		

[a]BRS: Brief Resilience Scale.

[b]WHO-5: World Health Organization-Five Well-Being Index.

[c]PHQ-9: Patient Health Questionnaire-9.

[d]Percentages are based on available responses. Measure-specific denominators for the 18- to 25-, 26- to 40-, and >40-year groups, respectively, were as follows: BRS: n=403, 378, and 324; WHO-5, PHQ-9, sleep issues, and suicidal ideation: n=402, 378, and 324; and GAD-7: n=401, 377, and 319.

[e]Significant *P* value.

[f]GAD-7: Generalized Anxiety Disorder 7-item scale.

XSL•FO
RenderX

Table . Clinical baseline measures by age group (EQ-5D-5L).

Variable	Age groups (years), n (%)			χ^2 (*df*)	*P* value
	18-25[a]	26-40	>40		
Mobility				22.65 (2)	<.001[b]
No problems	318 (79.3)	298 (79.0)	209 (65.5)		
Problems present	83 (20.7)	79 (21.0)	110 (34.5)		
Self-care				4.47 (2)	.11
No problems	310 (77.3)	307 (81.4)	239 (74.9)		
Problems present	91 (22.7)	70 (18.6)	80 (25.1)		
Usual activity				5.30 (2)	.07
No problems	213 (53.1)	208 (55.2)	149 (46.7)		
Problems present	188 (46.9)	169 (44.8)	170 (53.3)		
Anxiety/depression				12.63 (2)	<.001[b]
No problems	75 (18.7)	108 (28.6)	89 (27.9)		
Problems present	326 (81.3)	269 (71.4)	230 (72.1)		
Pain/discomfort				4.14 (2)	.13
No	169 (42.1)	173 (45.9)	122 (38.2)		
Yes	232 (57.9)	204 (54.1)	197 (61.8)		

[a]Percentages are based on available responses. The measure-specific denominators for the 18- to 25-, 26- to 40-, and >40-year groups were n=401, n=377, and n=319, respectively.

[b]Significant *P* value.

The proportion screening positive for probable depression was 61.9% (n=249) among participants aged 18 to 25 years, 55.3% (n=209) among those aged 26 to 40 years, and 52.8% (n=171) among those aged older than 40 years (χ^2_2=6.81; *P*=.03). Suicidal ideation, assessed using PHQ-9 item 9, was reported by 58.0%, 44.4%, and 44.1% of the 3 groups, respectively (χ^2_2=19.09; *P*<.001). No statistically significant age-group differences were observed in categorical sleep difficulties or anxiety.

For the EQ-5D-5L dimensions, mobility problems were reported by 20.7% (n=83) of participants aged 18 to 25 years, 21.0% (n=79) of those aged 26 to 40 years, and 34.5% (n=110) of those aged older than 40 years (χ^2_2=22.65, *P*<.001). Problems in the anxiety/depression dimension were reported by 81.3% (n=326), 71.4% (n=269), and 72.1% (n=230) of the 3 groups, respectively (χ^2_2=12.63, *P*<.001). No statistically significant age-group differences were found for self-care, usual activities, or pain/discomfort.

ANOVA Analysis

One-way ANOVA was conducted to examine age-group differences in continuous baseline mental health measures across 3 age categories (18 - 25 y, 26 - 40 y, and >40 y; Table 4). When the assumption of homogeneity of variance was violated, Welch ANOVA was applied. For outcomes showing significant omnibus effects, post hoc pairwise comparisons were conducted using Games-Howell tests to identify specific age-group differences (Table 5). Consistent with the chi-square analyses of categorized outcomes, significant age-related differences were observed for resilience, depressive symptoms, anxiety symptoms, suicidal ideation, and recovery, with younger participants generally reporting poorer mental health outcomes.

XSL•FO
RenderX

Table . ANOVA test parameters for different age groups.

Variable	Age groups (years), mean (SD)			*F* test (*df*)	*P* value
	18-25	26-40	>40		
BRS[a]	15.96 (4.67)	17.34 (5.14)	17.09 (4.90)	8.73 (2,1102)	<.001[b]
WHO-5[c]	49.63 (22.85)	53.31 (22.38)	51.11 (24.39)	2.48 (2,1101)	.08
PHQ-9[d]	12.12 (6.33)	11.03 (6.70)	10.46 (6.56)	6.17 (2,1101)	<.001[b]
Suicidal ideation	0.94 (1.01)	0.69 (0.92)	0.68 (0.92)	9.53 (2,1101)	<.001[b]
Sleep issues	1.50 (0.99)	1.46 (1.03)	1.43 (1.07)	.41 (2,1101)	.67
GAD-7[e]	9.02 (5.38)	8.71 (6.01)	7.44 (5.35)	7.74 (2,1094)	<.001[b]
EQ-5D-5L	87.69 (14.87)	92.08 (15.16)	90.11 (16.01)	8.07 (2,1100)	<.001[b]

[a]BRS: Brief Resilience Scale.

[b]Significant *P* value.

[c]WHO-5: World Health Organization-Five Well-Being Index.

[d]PHQ-9: Patient Health Questionnaire-9.

[e]GAD-7: Generalized Anxiety Disorder 7-item scale.

XSL•FO
RenderX

Table . Pairwise comparison of the baseline mean scores on the scales for participants of different age groups.

Variable, study group, and paired groups	Mean difference (SE)	*P* value	95% CI
BRS[a] (y)			
Study group: 18-25			
Paired group: 26-40	−1.38 (0.35)	<.001[b]	−2.20 to −0.55
Paired group: >40	−1.13 (0.36)	.01[b]	−1.97 to 0.29
Study group: 26-40			
Paired group: 18-25	1.38 (0.35)	<.001[b]	0.55 to 2.20
Paired group: >40	0.25 (0.38)	.79	−0.64 to 1.14
Study group: >40			
Paired group: 18-25	1.13 (0.36)	.01[b]	0.29 to 1.97
Paired group: 26-40	−0.25 (0.38)	.79	−1.14 to 0.64
PHQ-9[c] (y)			
Study group: 18-25			
Paired group: 26-40	1.08 (0.47)	.05	−0.01 to 2.18
Paired group: >40	1.66 (0.48)	<.001[b]	0.53 to 2.79
Study group: 26-40			
Paired group: 18-25	−1.08 (0.47)	.05	−2.18 to 0.01
Paired group: >40	0.58 (0.50)	.48	−0.60 to 1.76
Study group: >40			
Paired group: 18-25	−1.66 (0.48)	<.001[b]	−2.79 to −.053
Paired group: 26-40	−0.58 (0.50)	.48	−1.76 to 0.60
Suicidal Ideation (y)			
Study group: 18-25			
Paired group: 26-40	0.26 (0.07)	<.001[b]	0.10 to 0.42
Paired group: >40	0.26 (0.07)	<.001[b]	0.10 to 0.43
Study group: 26-40			
Paired group: 18-25	−0.26 (0.07)	<.001[b]	−0.42 to −0.10
Paired group: >40	0.01 (0.07)	>.99	−0.16 to 0.17
Study group: >40			
Paired group: 18-25	−0.26 (0.07)	<.001[b]	−0.43 to −0.10
Paired group: 26-40	−0.01 (0.07)	>.99	−0.17 to 0.16
GAD-7[d] (y)			
Study group: 18-25			
Paired group: 26-40	0.31 (0.41)	.72	−0.64 to 1.28
Paired group: >40	1.59 (0.41)	<.001[b]	0.64 to 2.53
Study group: 26-40			
Paired group: 18-25	−0.31 (0.41)	.72	−1.28 to 0.65
Paired group: >40	1.27 (0.43)	.01	0.26 to 2.29
Study group: >40			
Paired group: 18-25	−1.59 (0.40)	<.001[b]	−2.53 to −0.64

XSL•FO
RenderX

Variable, study group, and paired groups	Mean difference (SE)	*P* value	95% CI
Paired group: 26-40	1.27 (0.43)	.01	−2.28 to −0.26
EQ-5D-5L (y)			
Study group: 18-25			
Paired group: 26-40	−4.40 (1.08)	<.001[b]	−6.92 to −1.87
Paired group: >40	−2.42 (1.16)	.09	−5.15 to 0.30
Study group: 26-40			
Paired group: 18-25	4.40 (1.08)	<.001[b]	1.87 to 6.92
Paired group: >40	1.97 (1.18)	.22	−0.91 to 4.75
Study group: >40			
Paired group: 18-25	2.42 (1.16)	.09	−0.30 to 5.15
Paired group: 26-40	−1.97 (1.18)	.22	−4.76 to 0.81

[a]BRS: Brief Resilience Scale.

[b]Significant *P* value.

[c]PHQ-9: Patient Health Questionnaire-9.

[d]GAD-7: Generalized Anxiety Disorder 7-item scale.

Resilience, assessed using the BRS, differed significantly across age groups (Welch $F_{2, 717.35}$=8.98, P<.001), with a small effect size (η^2≈0.02). Post hoc Games-Howell comparisons indicated that participants aged 18 to 25 years reported significantly lower resilience than those aged 26 to 40 years (P<.001) and those over 40 years of age (P<.05); in contrast, no significant difference was observed between the 2 older age groups.

Depressive symptom severity, measured by the PHQ-9, also varied significantly by age (Welch $F_{2, 718.19}$=6.34, P=.002), with a small effect size (η^2≈0.01). Games-Howell post hoc tests revealed that participants aged 18 to 25 years reported significantly higher depression scores than those over 40 years of age (P<.001), aligning with the higher prevalence of likely depression observed in the youngest age group in the categorical analyses.

Anxiety severity, assessed using the GAD-7, demonstrated significant age-group differences (Welch $F_{2, 715.12}$=8.33, P<.001), with a small-to-moderate effect size (η^2≈0.02). Post hoc analyses showed that both the 18 to 25 year and 26 to 40 year groups reported significantly higher anxiety scores compared with participants over 40 years of age (P<.001 and P=.01, respectively), consistent with the chi-square findings, indicating higher rates of anxiety among younger participants.

Suicidal ideation, measured using item 9 of the PHQ-9, also differed significantly across age groups (Welch $F_{2, 725.35}$=9.05; P<.001), with a small-to-moderate effect size (η^2≈0.02). Games-Howell post hoc comparisons indicated that participants aged 18 to 25 years reported significantly higher levels of suicidal ideation than both the 26 to 40 years and older than 40 years groups (both P<.001), with no significant difference between the 2 older groups.

Recovery scores showed significant age-related variation (Welch $F_{2, 713.12}$=8.38, P<.001), with a small effect size (η^2≈0.01). Post hoc analyses revealed that participants aged 18 to 25 years reported significantly lower recovery scores than those aged 26 to 40 years (P<.001), mirroring the lower resilience and poorer mental health profiles observed among younger participants.

Quality of life, measured using the EQ-5D-5L index score, also differed significantly across age groups ($F_{2, 1100}$=8.07, P<.001), with a small effect size. Participants aged 18 to 25 years reported significantly lower EQ-5D-5L scores compared with those aged 26 to 40 years, indicating poorer overall perceived health status among younger participants, while scores among participants aged over 40 years fell between the 2 younger age groups.

No significant age-group differences were observed for overall well-being as measured by the WHO-5 (Welch $F_{2, 714.09}$=2.60, P=.08), or for sleep disturbance assessed using the PHQ-9 sleep item (Welch $F_{2, 714.35}$=0.41, P=.67).

Discussion

Principal Findings

This study examined age-related differences in baseline demographic, clinical, and mental health characteristics among individuals enrolled in the Text4Support program following discharge from inpatient psychiatric care. By analyzing both categorized and continuous mental health measures, the findings provide a nuanced understanding of how age is associated with mental health vulnerability at the point of discharge, a critical transitional period. Overall, the results indicate that younger adults, particularly those aged 18 to 25 years, experience disproportionately higher levels of psychological distress, lower resilience, and poorer recovery-related outcomes compared with older age groups.

Baseline analyses revealed a high prevalence of mental health symptoms across the entire sample, underscoring the substantial burden of psychiatric distress at hospital discharge. Rates of

XSL•FO
RenderX

depression, anxiety, poor well-being, and suicidal ideation were considerable, reinforcing previous evidence that the postdischarge period represents a time of heightened vulnerability [37-39]. These findings align with broader literature highlighting increased risks of relapse, emergency service use, and mortality following psychiatric discharge, particularly when continuity of care is limited. The substantial symptom burden observed among participants preparing for discharge warrants careful clinical interpretation. Screening positive for depression or endorsing PHQ-9 item 9 does not, by itself, establish that discharge was inappropriate or that inpatient treatment was inadequate. Discharge decisions are informed by comprehensive clinical assessment, including symptom change, functional status, immediate risk, treatment response, available supports, and the feasibility of community follow-up. Moreover, PHQ-9 item 9 is a screening indicator of thoughts of death or self-harm and should not be interpreted as a stand-alone assessment of imminent suicide risk [40-42]. Because this study did not examine inpatient treatment processes, changes during hospitalization, or the clinical basis for discharge decisions, conclusions regarding the quality of acute inpatient care cannot be drawn.

Age-stratified analyses demonstrated consistent and meaningful differences across several mental health indicators. Younger participants reported significantly higher levels of depressive and anxiety symptoms, greater suicidal ideation, and lower resilience and recovery scores compared with participants aged over 40 years. These patterns were evident in both chi-square analyses of categorized outcomes and Welch ANOVA analyses of continuous measures, lending robustness to the findings. The convergence of findings across analytical approaches indicates that age-group differences were present across several mental health measures. However, because the analyses were unadjusted, these differences may also reflect variation in education, employment, diagnosis, or other sociodemographic and clinical characteristics.

The heightened vulnerability observed among individuals aged 18 to 25 years may be explained by a combination of developmental, social, and structural factors [35,36]. Young adulthood is a period marked by ongoing identity formation, emotional regulation development, and life transitions related to education, employment, and relationships. Psychiatric hospitalization during this stage may disrupt these processes and compound existing stressors [35]. In addition, younger adults in this sample were more likely to report low resilience and poorer recovery, which may reflect fewer established coping strategies or limited access to stable social and economic support. These findings are consistent with previous research demonstrating that younger psychiatric patients often experience more severe symptomatology and poorer psychosocial functioning at discharge compared with older cohorts [35,36].

In contrast, participants aged over 40 years generally demonstrated lower levels of depression and anxiety, fewer reports of suicidal ideation, and higher resilience and recovery scores. While older adults may face their own set of challenges, including chronic illness or social isolation, they may also benefit from accumulated coping skills, greater emotional regulation, or more stable life circumstances [43]. The observed age gradient in mental health outcomes highlights the importance of considering age as a key factor in postdischarge risk stratification and intervention planning.

Quality-of-life analyses further supported these age-related differences. Participants aged older than 40 years reported a higher prevalence of mobility problems, reflecting age-related physical health concerns, whereas younger participants reported substantially higher levels of anxiety and depression within the EQ-5D-5L framework. This distinction underscores the need for age-sensitive approaches to postdischarge care that address not only psychiatric symptoms but also broader functional and quality-of-life domains relevant to different age groups [44]. Viewed in relation to existing service frameworks, the greater psychological distress observed among participants aged 18 to 25 years supports calls for developmentally responsive follow-up and stronger continuity of care for emerging adults [9]. The higher prevalence of mobility problems among participants aged older than 40 years also indicates that discharge planning should consider functional and physical health needs alongside psychiatric symptoms. However, because the older than 40 years category included both middle-aged and older participants, this finding should not be interpreted as evidence specific to geriatric mental health services.

Notably, no significant age-group differences were observed for overall well-being measured by the WHO-5 or for sleep disturbance when analyzed as continuous outcomes. These findings may suggest that certain aspects of subjective well-being and sleep difficulties are pervasive across age groups at discharge, potentially reflecting the shared impact of acute psychiatric hospitalization. Alternatively, these measures may be less sensitive to age-related variation at baseline, warranting further investigation in longitudinal analyses.

Although this baseline analysis did not evaluate Text4Support, the observed age-related patterns may inform hypotheses for subsequent intervention analyses, including whether engagement or outcomes vary across age groups. In particular, the greater psychological distress and lower resilience observed among participants aged 18 to 25 years may be relevant when considering the content, intensity, and delivery of postdischarge support. However, the present findings do not establish that Text4Support is effective or that any age-specific adaptation would improve outcomes.

From a service delivery perspective, these findings support incorporating both symptom severity and age-specific considerations into discharge planning and transitional care. Patients with persistent symptoms may benefit from comprehensive clinical and suicide risk assessment before discharge, collaborative safety planning, medication reconciliation, a written crisis plan, and confirmed follow-up arrangements [38,42,45]. A warm handover to community providers and proactive postdischarge contact may reduce gaps in care, particularly when suicide risk has been identified [42,46,47]. Younger patients may benefit from more structured follow-up, proactive engagement, and interventions targeting resilience and recovery skills, while participants aged older than 40 years may require additional attention to functional limitations and physical health concerns. Continued symptom

XSL•FO
RenderX

monitoring can support early identification of deterioration; however, supportive text messaging should be considered an adjunct to, rather than a replacement for, timely clinical follow-up and continuing care [46-50].

Limitations

Several limitations should be acknowledged. First, this cross-sectional analysis was based on baseline data, precluding causal inference and conclusions about symptom trajectories over time. It also cannot determine the effectiveness, accessibility, or acceptability of Text4Support, whether engagement with the intervention differs by age, or whether age modifies intervention response. Second, the mental health outcomes were assessed using self-report instruments and may therefore be affected by reporting bias. Third, the older than 40 years category encompassed a broad age range and included participants at different stages of adulthood. Combining middle-aged and older participants may have obscured age-specific patterns within this group; consequently, the findings should not be interpreted as representing a distinct older-adult population. Finally, education level, employment status, and discharge diagnosis were reported descriptively in Table 1 but were not examined jointly in relation to the mental health outcomes. The reported age-group comparisons are therefore unadjusted and cannot be interpreted as effects of age independent of these sociodemographic and clinical characteristics. Other factors not included in the present analysis, including broader socioeconomic circumstances and social support, may also influence mental health status at discharge. Future hypothesis-driven analyses should examine these characteristics and their potential interactions with age.

Despite these limitations, this study provides valuable insight into age-related disparities in mental health status at psychiatric discharge within a large, real-world sample. The findings underscore the need for age-informed transitional care strategies and support the integration of digital mental health interventions within broader, flexible care pathways. Future research should examine how age interacts with intervention engagement and outcomes over time and whether tailored approaches can mitigate the elevated risks observed among younger adults during this critical period.

Conclusions

This study demonstrates significant age-related differences in baseline mental health and quality-of-life outcomes among individuals discharged from psychiatric inpatient care and enrolled in the Text4Support program. Younger adults, particularly those aged 18 to 25 years, exhibited higher levels of psychological distress, lower resilience, and poorer recovery-related outcomes compared with participants aged older than 40 years, underscoring their heightened vulnerability during the transition to community care.

The findings emphasize the importance of incorporating age as a key consideration in postdischarge assessment and planning. Participants aged 18 to 25 years with elevated baseline risk profiles may require more tailored or intensified transitional care, whereas participants aged older than 40 years may benefit from supports that address functional limitations alongside mental health needs.

These baseline findings provide context for future evaluation of Text4Support and other postdischarge supports, including whether intervention engagement and outcomes vary across age groups. Integrating baseline risk stratification into transitional care planning may enhance the effectiveness of postdischarge supports and improve outcomes for vulnerable populations.

Generative AI Use

The authors stated that ChatGPT (OpenAI) was used to assist with language editing and improving text clarity but not for data collection, statistical analysis, or interpretation of the study findings. They attest that there was no use of generative AI technology in the generation of text, figures, or other informational content of this manuscript. They also confirmed that all AI-assisted content was reviewed and revised by the authors, who take full responsibility for the accuracy, originality, and integrity of the final manuscript.

Funding

This study was funded through a grant by Alberta Innovates Health Solutions (grant application number: 202010086). The content reported herein is that of the authors. The funders had no role in the study design, the decision to publish, or the writing of this protocol.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the sensitive nature of participant-level mental health data and the need to protect participant confidentiality. Deidentified data may be available from the corresponding author upon reasonable request, subject to applicable institutional and ethical requirements.

Authors' Contributions

Conceptualization: VIOA
Data curation: WM, EO, RS, HEE, EE, VIOA
Formal analysis: RS, WM, EO, HEE
Funding acquisition: VIOA
Investigation: EO, RS, HEE, BA, EE, PHS, PC, WV, AO, AJG, VIOA

XSL•FO
RenderX

Methodology: WM, EE, RS, EO, VIOA
Project administration: VIOA
Resources: PHS, PC, WV, AO, EE, AJG, VIOA
Supervision: AJG, VIOA
Visualization: WM
Writing – original draft: WM
Writing – review & editing: WM, EO, RS, HEE, BA, EE, PHS, PC, WV, AO, AJG, VIOA.

Conflicts of Interest

None declared.

References

1. Mao W, Shalaby R, Owusu E, et al. Status after hospital discharge: an observational study of the progression of patients' mental health symptoms six weeks after hospital discharge. J Clin Med 2023 Dec 7;12(24):7559. [doi: 10.3390/jcm12247559] [Medline: 38137628]
2. Smith TE, Haselden M, Corbeil T, et al. Effect of scheduling a post-discharge outpatient mental health appointment on the likelihood of successful transition from hospital to community-based care. J Clin Psychiatry 2020 Sep 15;81(5):20m13344. [doi: 10.4088/JCP.20m13344] [Medline: 32936543]
3. Mao W, Shalaby R, Owusu E, et al. Supportive text messaging and peer support for patients in the 6 months following discharge from a psychiatric admission: mental health outcomes from a cluster-randomized controlled trial. J Clin Med 2025 Nov 21;14(23):8262. [doi: 10.3390/jcm14238262] [Medline: 41375565]
4. Gillard S, White R, Miller S, Turner K. Open access support groups for people experiencing personality disorders: do group members' experiences reflect the theoretical foundations of the Sun Project? Psychol Psychother 2015 Mar;88(1):87-104. [doi: 10.1111/papt.12031] [Medline: 24872313]
5. Mirowsky J, Ross CE. Age and depression. J Health Soc Behav 1992 Sep;33(3):187-205. [Medline: 1401846]
6. Gupta U. Age, mental health and well-being. J Psychosoc Res 2016;11(1):147-156 [FREE Full text]
7. Zisook S, Lesser I, Stewart JW, et al. Effect of age at onset on the course of major depressive disorder. Am J Psychiatry 2007 Oct;164(10):1539-1546. [doi: 10.1176/appi.ajp.2007.06101757] [Medline: 17898345]
8. Lavingia R, Jones K, Asghar-Ali AA. A systematic review of barriers faced by older adults in seeking and accessing mental health care. J Psychiatr Pract 2020 Sep;26(5):367-382. [doi: 10.1097/PRA.0000000000000491] [Medline: 32936584]
9. Taking the next step forward: building a responsive mental health and addictions system for emerging adults. : Mental Health Commission of Canada (MHCC); 2015 Jul 17 URL: https://mentalhealthcommission.ca/resource/taking-the-next-step-forward-building-a-responsive-mental-health-and-addictions-system-for-emerging-adults/ [accessed 2026-10-10]
10. MacCourt P, Wilson K, Tourigny-Rivard MF. Guidelines for comprehensive mental health services for older adults in Canada. : Mental Health Commission of Canada (MHCC); 2011 Jun 1 URL: https://mentalhealthcommission.ca/wp-content/uploads/2021/08/Senior_Care_Guideline_2018_eng.pdf [accessed 2026-10-10]
11. Framework to implement a life course approach in practice. : World Health Organization; 2025 URL: https://www.who.int/publications/i/item/9789240112575 [accessed 2026-10-10]
12. Health indicators 2008. : Canadian Institute for Health Information; 2008 URL: https://publications.gc.ca/collections/collection_2008/cihi-icis/H115-16-2008E.pdf [accessed 2026-10-22]
13. Eboreime E, Shalaby R, Mao W, et al. Reducing readmission rates for individuals discharged from acute psychiatric care in Alberta using peer and text message support: protocol for an innovative supportive program. BMC Health Serv Res 2022 Mar 12;22(1):332. [doi: 10.1186/s12913-022-07510-8] [Medline: 35279142]
14. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
15. Kohn MA, Senyak J. Sample Size Calculators. 2021 Dec 20. URL: https://sample-size.net/about-us/ [accessed 2026-11-06]
16. Williams N. The GAD-7 questionnaire. Occup Med 2014 Apr 1;64(3):224. [doi: 10.1093/occmed/kqt161]
17. Kroenke K, Spitzer RL, Williams JBW, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007 Mar 6;146(5):317-325. [doi: 10.7326/0003-4819-146-5-200703060-00004] [Medline: 17339617]
18. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
19. Mao W, Adu M, Eboreime E, et al. Post-traumatic stress disorder, major depressive disorder, and wildfires: a fifth-year postdisaster evaluation among residents of Fort McMurray. Int J Environ Res Public Health 2022 Aug 8;19(15):9759. [doi: 10.3390/ijerph19159759] [Medline: 35955114]
20. Beard C, Hsu KJ, Rifkin LS, Busch AB, Björgvinsson T. Validation of the PHQ-9 in a psychiatric sample. J Affect Disord 2016 Mar 15;193:267-273. [doi: 10.1016/j.jad.2015.12.075] [Medline: 26774513]

XSL•FO
RenderX

21. Lequerica AH, Watson E, Dijkers MP, et al. The utility of the Patient Health Questionnaire (PHQ-9) sleep disturbance item as a screener for insomnia in individuals with moderate to severe traumatic brain injury. J Head Trauma Rehabil 2022;37(5):E383-E389. [doi: 10.1097/HTR.0000000000000746] [Medline: 35125428]
22. Morin CM. Insomnia: Psychological Assessment and Management: Guilford press; 1993. URL: https://psycnet.apa.org/record/1993-98362-000 [accessed 2026-10-10]
23. Simon GE, Rutter CM, Peterson D, et al. Does response on the PHQ-9 Depression Questionnaire predict subsequent suicide attempt or suicide death? Psychiatr Serv 2013 Dec 1;64(12):1195-1202. [doi: 10.1176/appi.ps.201200587] [Medline: 24036589]
24. MacGregor KL, Funderburk JS, Pigeon W, Maisto SA. Evaluation of the PHQ-9 Item 3 as a screen for sleep disturbance in primary care. J Gen Intern Med 2012 Mar;27(3):339-344. [doi: 10.1007/s11606-011-1884-5] [Medline: 21948205]
25. Louzon SA, Bossarte R, McCarthy JF, Katz IR. Does suicidal ideation as measured by the PHQ-9 predict suicide among VA patients? Psychiatr Serv 2016 May 1;67(5):517-522. [doi: 10.1176/appi.ps.201500149] [Medline: 26766757]
26. Bauer AM, Chan YF, Huang H, Vannoy S, Unützer J. Characteristics, management, and depression outcomes of primary care patients who endorse thoughts of death or suicide on the PHQ-9. J Gen Intern Med 2013 Mar;28(3):363-369. [doi: 10.1007/s11606-012-2194-2] [Medline: 22936288]
27. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176. [doi: 10.1159/000376585] [Medline: 25831962]
28. Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The Brief Resilience Scale: assessing the ability to bounce back. Int J Behav Med 2008;15(3):194-200. [doi: 10.1080/10705500802222972] [Medline: 18696313]
29. Minh Nguyen T, Tran MAQ, Hoang Le GN. Psychometric validation of the Brief Resilience Scale in the sample of Vietnamese university student. J Coll Stud Ment Health 2023 Jul 2;38(3):648-663. [doi: 10.1080/87568225.2023.2231348]
30. Devlin N, Roudijk B, Ludwig K. Value Sets for EQ-5D-5L: A Compendium, Comparative Review & User Guide: Springer; 2022.
31. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec;20(10):1727-1736. [doi: 10.1007/s11136-011-9903-x] [Medline: 21479777]
32. Feng YS, Kohlmann T, Janssen MF, Buchholz I. Psychometric properties of the EQ-5D-5L: a systematic review of the literature. Qual Life Res 2021 Mar;30(3):647-673. [doi: 10.1007/s11136-020-02688-y] [Medline: 33284428]
33. Dams J, Rimane E, Steil R, Renneberg B, Rosner R, König HH. Reliability, validity and responsiveness of the EQ-5D-5L in assessing and valuing health status in adolescents and young adults with posttraumatic stress disorder: a randomized controlled trail. Psychiatr Q 2021 Jun;92(2):459-471. [doi: 10.1007/s11126-020-09814-6] [Medline: 32803473]
34. Downloading IBM SPSS statistics 25. IBM Support. 2017. URL: https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-25 [accessed 2026-10-10]
35. Arnett JJ. Emerging adulthood. A theory of development from the late teens through the twenties. Am Psychol 2000 May;55(5):469-480. [Medline: 10842426]
36. Arnett JJ. Emerging adulthood: what is it, and what is it good for? Child Dev Perspect 2007 Dec 1;1(2):68-73. [doi: 10.1111/j.1750-8606.2007.00016.x]
37. Chung DT, Ryan CJ, Hadzi-Pavlovic D, Singh SP, Stanton C, Large MM. Suicide rates after discharge from psychiatric facilities: a systematic review and meta-analysis. JAMA Psychiatry 2017 Jul 1;74(7):694-702. [doi: 10.1001/jamapsychiatry.2017.1044] [Medline: 28564699]
38. Vigod SN, Kurdyak PA, Dennis CL, et al. Transitional interventions to reduce early psychiatric readmissions in adults: systematic review. Br J Psychiatry 2013 Mar;202(3):187-194. [doi: 10.1192/bjp.bp.112.115030] [Medline: 23457182]
39. Walter F, Carr MJ, Mok PLH, et al. Premature mortality among patients recently discharged from their first inpatient psychiatric treatment. JAMA Psychiatry 2017 May 1;74(5):485-492. [doi: 10.1001/jamapsychiatry.2017.0071] [Medline: 28296989]
40. Na PJ, Yaramala SR, Kim JA, et al. The PHQ-9 Item 9 based screening for suicide risk: a validation study of the Patient Health Questionnaire (PHQ)-9 Item 9 with the Columbia Suicide Severity Rating Scale (C-SSRS). J Affect Disord 2018 May;232:34-40. [doi: 10.1016/j.jad.2018.02.045] [Medline: 29477096]
41. Suicide risk assessment toolkit: a resource for healthcare workers and organizations. : Mental Health Commission of Canada (MHCC) and Canadian Patient Safety Institute (CPSI); 2021 URL: https://mentalhealthcommission.ca/wp-content/uploads/2021/02/Suicide-risk-assessment-toolkit.pdf [accessed 2026-10-10]
42. Care transitions procedure (addiction & mental health). Alberta Health Services. 2023 Jan 9. URL: https://www.albertahealthservices.ca/info/Page7451.aspx [accessed 2026-10-10]
43. Thomas ML, Kaufmann CN, Palmer BW, et al. Paradoxical trend for improvement in mental health with aging: a community-based study of 1,546 adults aged 21–100 years. J Clin Psychiatry 2016 Aug;77(8):e1019-e1025. [doi: 10.4088/JCP.16m10671] [Medline: 27561149]
44. Prince M, Patel V, Saxena S, et al. No health without mental health. The Lancet 2007 Sep;370(9590):859-877. [doi: 10.1016/S0140-6736(07)61238-0]

XSL•FO
RenderX

45. Stanley B, Brown GK, Brenner LA, et al. Comparison of the safety planning intervention with follow-up vs usual care of suicidal patients treated in the emergency department. JAMA Psychiatry 2018 Sep 1;75(9):894-900. [doi: 10.1001/jamapsychiatry.2018.1776] [Medline: 29998307]
46. Che SE, Gwon YG, Kim KH. Follow-up timing after discharge and suicide risk among patients hospitalized with psychiatric illness. JAMA Netw Open 2023 Oct 2;6(10):e2336767. [doi: 10.1001/jamanetworkopen.2023.36767] [Medline: 37812420]
47. Madsen T, Egilsdottir E, Damgaard C, Erlangsen A, Nordentoft M. Assessment of suicide risks during the first week immediately after discharge from psychiatric inpatient facility. Front Psychiatry 2021;12:643303. [doi: 10.3389/fpsyt.2021.643303] [Medline: 33959051]
48. Agyapong VIO, Juhás M, Ohinmaa A, et al. Randomized controlled pilot trial of supportive text messages for patients with depression. BMC Psychiatry 2017 Aug 2;17(1):286. [doi: 10.1186/s12888-017-1448-2] [Medline: 28768493]
49. Agyapong VIO, Milnes J, McLoughlin DM, Farren CK. Perception of patients with alcohol use disorder and comorbid depression about the usefulness of supportive text messages. Technol Health Care 2013;21(1):31-39. [doi: 10.3233/THC-120707] [Medline: 23358057]
50. Agyapong VIO, Shalaby R, Hrabok M, et al. Mental health outreach via supportive text messages during the COVID-19 pandemic: improved mental health and reduced suicidal ideation after six weeks in subscribers of Text4Hope compared to a control population. Int J Environ Res Public Health 2021 Feb 23;18(4):2157. [doi: 10.3390/ijerph18042157] [Medline: 33672120]

Abbreviations

BRS: Brief Resilience Scale
CIHI: Canadian Institute for Health Information
***DSM-IV*:** *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*
GAD-7: Generalized Anxiety Disorder 7-item scale
HADS: Hospital Anxiety and Depression Scale
PHQ-9: Patient Health Questionnaire-9
WHO-5: World Health Organization-Five Well-Being Index

Edited by C Jacob; submitted 28.Jan.2026; peer-reviewed by A Seeherunwong, A Vuolukka; revised version received 20.Aug.2026; accepted 28.Sep.2026; published 09.Nov.2026.

Please cite as:
Mao W, Owusu E, Shalaby R, Elgendy HE, Agyapong B, Eboreime E, Silverstone PH, Chue P, Vuong W, Ohinmaa A, Greenshaw AJ, Agyapong VIO
Age-Related Differences in Mental Health Status at Psychiatric Discharge Among Patients Across 10 Acute Care Facilities in Alberta, Canada: Cross-Sectional Study
JMIR Hum Factors 2026;13:e92360
URL: https://humanfactors.jmir.org/2026/1/e92360
doi:10.2196/92360

© Wanying Mao, Ernest Owusu, Reham Shalaby, Hossam Eldin Elgendy, Belinda Agyapong, Ejemai Eboreime, Peter H Silverstone, Pierre Chue, Wesley Vuong, Arto Ohinmaa, Andrew J Greenshaw, Vincent I O Agyapong. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 9.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Factors Influencing Patient Satisfaction Within the Multistage Online Consultation Process in Internet Hospitals in Beijing: Empirical Analysis

Yi Wang[1*], MD; Yujia Han[2*], BA; Mengyu Duan[3*], MD; Yani Wang[4], PhD

[1]School of Public Health, Capital Medical University, Beijing, China

[2]Yanjing Medical College, Capital Medical University, Beijing, China

[3]Graduate School, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

[4]Medical Technology College, Capital Medical University, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Yani Wang, PhD
Medical Technology College
Capital Medical University
No 4, Dadong Road, Shunyi District
Beijing, 101300
China
Phone: 86 15910272366
Email: wangyani@ccmu.edu.cn

Abstract

Background: Internet hospitals are playing a significant role in medical care with their potential to provide widely accessible outpatient service delivery via information technologies. Current research on patients' satisfaction with internet hospitals mainly focus on physician-patient relationship and patient demand, and less is considered about the whole process of online consultation.

Objective: This study aims to identity the factors influencing patient satisfaction considering the entire process of online consultation to help physicians deliver better online medical services and to help physical hospitals operate internet hospitals more effectively.

Methods: Based on the service quality (SERVQUAL) theoretical model, questionnaire items were designed for the 5 dimensions of reliability, assurance, responsiveness, empathy, and tangibility. The research samples for data collection were 355 patients on the internet hospital platform operated by a tertiary general hospital in Beijing. Confirmatory factor analysis was performed for dimensions of the measurement model, and path analysis was performed for the structural model.

Results: The current consultation process of internet hospitals, as perceived by patients, did not yet meet their expectations, and the overall satisfaction rate of patients was only 7.15. The 5 dimensions of reliability, responsiveness, empathy, tangibility, and assurance had different positive predictive effects on patient satisfaction in internet hospitals to a certain extent. Among them, tangibility exerted the greatest impact on patient satisfaction in internet hospitals (β=0.25) but obtained the lowest score among the sampled participants (mean=6.78).

Conclusions: Internet hospitals and physicians should focus on the factors of the above dimensions, especially the tangibility of medical services in the multistage online consultation, thus better serving patients and promoting the sustainable development of internet health care.

(JMIR Hum Factors 2026;13:e79213) doi:10.2196/79213

KEYWORDS

internet hospital; multistage of online consultation process; patient satisfaction; SERVQUAL model; service quality; structural equation modeling

XSL•FO
RenderX

Introduction

The internet hospital is a new medical model that realizes online disease diagnosis and treatment services by using information technology to provide services to patients with common diseases or chronic diseases or living in remote and rural areas [1,2]. On such platforms, patients can consult physicians about their disease diagnosis, treatment plan, and prognosis without the limits of time and space [3,4]. In June 2024, the number of internet medical users in China had reached 418 million, accounting for 37.7% of the total internet users. There are 3340 internet hospitals in China, and the annual volume of internet-based medical services provided exceeds 100 million medical visits. The convenience, accessibility, and continuity of medical services in internet hospitals have injected great momentum into the development of China's public health services [5,6].

Internet hospitals have developed rapidly and become an important part of China's medical service system [7]. Many entity hospitals have set up their own internet hospital to provide web-based medical services. Compared with physical hospitals, internet hospitals can reduce patients' waiting time, increase doctor-patient interaction efficiency, and improve the accessibility of medical service quality [8]. Patient satisfaction is an important basis for evaluating health care quality, and patients will have high satisfaction when they are treated in an expected way and receive high service quality [9]. However, despite the convenience and accessibility of online health care services, there is still room for improvement in patient satisfaction with internet hospitals [3].

Existing research on patients' satisfaction primarily focus on online physician-patient interaction [4,10] and physician provision [7,11] and tends to examine offline physical hospitals [12,13]. For example, Li et al [4] evaluated the impact of internet hospitals on physician-patient relationship during outpatient visits and suggested improving physicians' online communication skills and strengthening the level of trust between physicians and patients from a patient perspective. Chang et al [10] investigated how physician-patient online interaction affects the integration of online and offline health services. Wu et al [7] studied patient satisfaction from the perspective of physicians and proposed that physicians' response, detailed style, and emotional comfort positively affects patient satisfaction. Yang et al [11] explored the effect of physicians' appearance, smile, and positive emotions on patients' selection and evaluation behavior and reported interesting and valuable conclusions. However, these studies have only focused on patient satisfaction at a specific stage, without exploring the whole process of patient diagnosis and treatment in internet hospitals.

Online consultation in internet hospitals is a complex and dynamic process, which includes preconsultation, consultation, postconsultation, and online and offline integration [14]. When patients have diagnosis and treatment needs, they can apply to internet hospitals for online consultation services. In the preconsultation stage, the patients need to decide which physician to choose by comparing information displayed in the internet hospitals [15]. Then, patients will communicate with the selected physician to conduct diagnosis and treatment in the consultation stage. In the postconsultation stage, medicines purchased online will be delivered on time without damage. Online and offline integration needs to be performed when the patient's condition is unusual or serious [10].

Patient satisfaction is often linked with health care service quality. Parasuraman et al [16] initially developed a service quality model to explain the potential causes of service quality issues and help managers improve service quality. The model was further improved as a measurement tool, the service quality (SERVQUAL) questionnaire, which consists of 5 dimensions: reliability, assurance, responsiveness, empathy, and tangibility, to measure users' expectation and perception of the service quality provided by the entity and reflects their attitude toward the service provider [17]. Previous studies have evaluated the quality of medical services in hospitals based on the SERVQUAL questionnaire [18-20]. Han et al [3] evaluated patients' expected and perceived service qualities of internet hospitals by using the service quality questionnaire with a national representative sample. In this study, we designed measure items of the questionnaire based on the 5 dimensions of the SERVQUAL theoretical model to evaluate the service quality from a multistage online consultation process such as preconsultation, consultation, postconsultation, and online-offline integration. The measurement dimensions in each phase of the diagnosis and treatment process can correspond to the SERVQUAL model.

This study aims to identify and clarify the factors influencing patient satisfaction in internet hospitals in Beijing across the multistage online consultation process. Exploring this problem can help such platforms optimize service design, thereby guiding internet hospitals to better serve patients and attract more users. We expect that problems existing in the online diagnosis and treatment process of internet hospitals can be uncovered from the perspective of patients, and suggestions on upgrading and improving internet hospitals can be put forward. Ultimately, internet diagnosis and treatment can better meet the needs of patients in the process of use and management, improve patients' satisfaction with internet hospitals, and thus promote the development of internet diagnosis and treatment.

Methods

We conducted a web-based cross-sectional survey of patients from an internet hospital platform of Beijing Chaoyang hospital, a tertiary general hospital which provides an integration of online and offline medical services for patients.

Ethical Considerations

According to the "Ethical Review Measures for Human Life Sciences and Medical Research" (approved by the National Science and Technology Ethics Committee) issued by the National Health Commission and the Ministry of Science and Technology of the People's Republic of China on February 27, 2023 (Document No. GWNJ-KE-2023-4), and according to Article 32 of Chapter 3 of the Ethical Review Regulations [21], this study was exempted from ethic review.

XSL•FO
RenderX

Before the questionnaire survey, the professional staff from the information center of the Beijing Chaoyang hospital carefully verified the questionnaire dimensions to ensure compliance with ethical standards. The survey data were anonymized and kept confidential for the respondents. During the data collection stage, our data was collected through the form of online questionnaires. All participants signed the "Informed Consent Form", clearly informing them of the research purpose, confidentiality terms, and the right to withdraw voluntarily. In the questionnaire we clearly stated that the collected data would only be used for academic research. We strictly abide by the Personal Information Protection Law. The questionnaires were anonymous and did not involve patients' names. Consequently, this study did not require ethical approval. Participants who completed a questionnaire could receive a reward of $0.80 USD. During the data processing process, we strictly abided by data protection regulations and took measures to ensure data security.

Instrument Development

Combining the existing scale of patient satisfaction in offline hospitals and the operating status of internet hospitals in Beijing, questionnaire items were designed based on the 5 dimensions of the SERVQUAL model. Our questionnaire consisted of 2 parts: basic information and satisfaction evaluation. The basic information included characteristics of the patients such as gender, age, education, monthly income, occupation, and the frequency of using internet hospitals. Satisfaction evaluation covered all stages of the patient's internet online diagnosis and treatment process, including prediagnosis, diagnosis, postdiagnosis, payment, purchase of medication, and online and offline referrals, with 4 items on reliability, 5 items on guarantees and responsiveness, 4 items on empathy, 3 items on tangibility, and 4 items on overall patient satisfaction, totaling 25 questions. Multimedia Appendix 1 lists the 25 items used in this study. The items were measured using a 10-point Likert scale (1=strongly disagree, 2=disagree, 3=slightly disagree, 4=somewhat disagree, 5=indifferent, 6=general, 7=somewhat agree, 8=slightly agree, 9=agree, 10=strongly agree).

Data Collection

To ensure the scientific aspect and rationality of the questionnaire scale, we adopted snowball sampling to conduct a small-scale pretest for patients who have used the internet hospital of a tertiary hospital in Beijing; a total of 44 questionnaires were issued, and 30 valid questionnaires were obtained after eliminating invalid questionnaires, with an effective rate of 68%. The sample data were tested for reliability and validity using SPSS software (version 24.0; IBM Corp). The Cronbach α coefficients of each dimension were greater than 0.7, and the overall Cronbach α coefficient was greater than 0.9, indicating that the internal consistency of the items was better and the overall reliability of the sample was higher. The Kaiser-Meyer-Olkin value was 0.75 and the *P* value was less than .001, meaning that the validity of the sample was acceptable. Therefore, this questionnaire can be used in subsequent studies.

In the formal survey, the questionnaire was released to the internet hospital platform of a tertiary hospital in Beijing, and convenient random sampling was adopted to distribute the questionnaires. To ensure data quality, we implemented a rigorous data screening procedure. Given that it takes 86 seconds at least to read the questionnaire from the beginning to the end, 20 questionnaires with the answer time of less than 86 seconds were excluded. Additionally, 23 questionnaires that selected one option for all questions and 6 questionnaires that failed an attention check were excluded. Finally, 355 valid questionnaires were obtained, with an effective rate of 87.9% (355/404).

Data Analysis

SPSS (version 27.0) and AMOS (version 23.0; IBM Corp) software were used for data analysis. We first employed descriptive statistics to analyze the demographic information of the patients and descriptive statistics of 25 observed variables. Structural equation modeling was then employed to examine the relationships between measurement items and structural variables to identify the causal relationships between structural variables [22,23]. Among them, confirmatory factor analysis was performed to ensure the reliability and validity for the measurement model, and path analysis was performed to estimate the influencing relationships for the structural model [24].

Results

Descriptive Statistics

Table 1 presents the demographic characteristics of the respondents. Among the respondents, males and females were evenly divided, with slightly more females than males. The largest number of people were in the 26-45 years age group, accounting for 31.5% (112/355) of the total sample. Respondents below the high school/secondary school level accounted for nearly half (174/355) of the total sample. Respondents with a monthly income of less than US $409.80 accounted for 38% (135/355) of the total sample. The frequency of using internet hospitals is measured as often (more than 3 times a month), frequently (2-3 times a month), occasionally (once a month), rarely (almost never). More than half (295/355, 83.1%) of the population employed internet hospitals at an average frequency.

XSL•FO
RenderX

Table 1. Demographic statistics of the sample (N=355).

Demographic characteristics, categories	Values, n (%)
Gender	
Female	167 (47.1)
Male	188 (52.9)
Age (y)	
<18	27 (7.6)
18-25	51 (14.4)
26-45	112 (31.5)
46-60	108 (30.4)
>60	57 (16.1)
Education	
High school or below	174 (49)
Junior college	70 (19.7)
Undergraduate	76 (21.4)
Postgraduate	35 (10)
Monthly income	
Less than US $409.80	135 (38)
US $409.90-US $819.70	130 (36.6)
US $819.80-US $1366.10	58 (16.3)
US $1366.20-US $2049.20	20 (5.6)
More than US $2049.20	12 (3.4)
Frequency of using internet hospital (frequently)	295 (83.1)

Table 2 provides a descriptive statistics for these items, including item mean and standard deviation, skewness, kurtosis, minimum, and maximum values. The skewness of each item ranged from –0.10 to –1.03, and the kurtosis ranged from 0.30 to 1.75. Since all values fell within the absolute value of 2, the data of each construct were normally distributed [25]. The mean value of each item was greater than the median 5.5, demonstrating that most patients using internet hospitals had a strong sense of identity with the potential variables set by the research. The 5 dimensions of reliability, assurance, responsiveness, empathy, and tangibility proposed in this paper could be regarded as core variables and the following detailed analysis was performed according to the data. However, the average score values of the 5 dimensions were not very high (only around 7), indicating that patients are not very satisfied with the whole process of diagnosis and treatment in the internet hospitals.

XSL•FO
RenderX

Table 2. Descriptive statistics for the items.

Constructs (mean), items	Mean (SD)	Skewness	Kurtosis	Minimum	Maximum
Reliability (mean=6.98)					
a1	6.98 (2.12)	–0.87	0.62	1	10
a2	6.94 (2.10)	–0.72	0.47	1	10
a3	7.00 (2.20)	–0.97	0.83	1	10
a4	7.01 (2.12)	–0.95	0.92	1	10
Assurance (mean=6.95)					
b1	6.95 (2.10)	–0.79	0.53	1	10
b2	6.90 (2.15)	–0.83	0.49	1	10
b3	6.92 (2.25)	–1.04	0.88	1	10
b4	6.97 (2.19)	–0.97	0.73	1	10
b5	6.99 (2.22)	–0.95	0.73	1	10
Responsiveness (mean=7.05)					
c1	6.97 (1.98)	–0.83	0.80	1	10
c2	7.06 (1.87)	–0.77	0.92	1	10
c3	7.08 (2.01)	–0.98	1.20	1	10
c4	7.12 (1.99)	–0.99	1.28	1	10
c5	7.01 (2.09)	–0.10	1.28	1	10
Empathy (mean=7.02)					
d1	7.09 (2.07)	–0.97	1.11	1	10
d2	6.88 (2.07)	–0.94	0.98	1	10
d3	7.01 (2.07)	–0.88	0.88	1	10
d4	7.11 (2.14)	–0.91	0.78	1	10
Tangibility (mean=6.78)					
e1	6.77 (2.26)	–0.78	0.30	1	10
e2	6.77 (2.16)	–0.91	0.61	1	10
e3	6.79 (2.25)	–0.84	0.41	1	10
Overall satisfaction (mean=7.15)					
f1	7.19 (1.96)	–0.89	1.28	1	10
f2	7.07 (1.94)	–0.87	1.38	1	10
f3	7.19 (1.95)	–0.93	1.17	1	10
f4	7.14 (1.90)	–1.03	1.75	1	10

Measurement Model

We tested the internal reliability, convergent validity, and discriminant validity for confirmatory factor analysis to evaluate the fitness between the 5 constructs and their indicators. The values of composite reliability, average variance extracted, Cronbach α, and factor loadings were calculated to assess the internal reliability and convergent validity. As shown in Table 3, the factor loading values were all greater than 0.84, reaching the desired level of estimation. The *P* value of the unstandardized coefficients was less than .01, meaning that the observed variables were significant for the latent variables. The reliability coefficients, composite reliability, and Cronbach α were all above the threshold value of 0.7, and average variance extracted were greater than 0.71, indicating that all reflective constructs exhibited good reliability and convergent validity [26,27].

XSL•FO
RenderX

Table 3. Construct reliability and validity analysis.

Items	Factor loading	*P* value	Reliability coefficient	Composite reliability	Cronbach α	Average variance extracted
Reliability				0.93	0.93	0.78
a1	0.89	<.001	0.80			
a2	0.87	.01	0.75			
a3	0.88	.01	0.78			
a4	0.89	—[a]	0.79			
Assurance				0.94	0.94	0.77
b1	0.87	<.001	0.76			
b2	0.87	<.001	0.76			
b3	0.87	<.001	0.76			
b4	0.89	<.001	0.80			
b5	0.87	—	0.75			
Responsiveness				0.92	0.93	0.71
c1	0.85	.01	0.71			
c2	0.84	<.001	0.71			
c3	0.84	<.001	0.70			
c4	0.84	<.001	0.70			
c5	0.86	—	0.74			
Empathy				0.91	0.92	0.73
d1	0.85	<.001	0.73			
d2	0.86	.01	0.73			
d3	0.84	.01	0.71			
d4	0.87	—	0.75			
Tangibility				0.90	0.90	0.75
e1	0.89	<.001	0.80			
e2	0.86	<.001	0.74			
e3	0.86	—	0.74			
Overall satisfaction				0.91	0.91	0.71
f1	0.85	—	0.72			
f2	0.84	<.001	0.71			
f3	0.84	.01	0.70			
f4	0.85	<.001	0.72			

[a]Not applicable.

For discriminant validity, the square root of average variance extracted for each construct on the diagonal was much larger than the correlation coefficients of the other constructs, indicating good discriminant validity of the measurement model, as shown in Table 4. The correlation coefficients of assurance, reliability, responsiveness, empathy, tangibility, and overall patient satisfaction were 0.88, 0.88, 0.84, 0.85, 0.87, and 0.51, respectively, with *P* values of less than .001, indicating that there was a significant positive correlation between the constructs and patient satisfaction.

XSL•FO
RenderX

Table 4. Discriminant validity analysis.

	Reliability	Assurance	Responsiveness	Empathy	Tangibility	Overall satisfaction
Reliability	0.88	__[a]	—	—	—	—
Assurance	0.36	0.88	—	—	—	—
Responsiveness	0.34	0.48	0.84	—	—	—
Empathy	0.36	0.42	0.39	0.85	—	—
Tangibility	0.35	0.40	0.33	0.35	0.87	—
Overall satisfaction	0.36	0.42	0.39	0.39	0.43	0.51

[a]Not applicable.

Structural Model

We performed absolute fit, added fit, and reduced fit to verify the fit degree of the structural model. The fitting effect of the structural model is shown in Table 5, with a chi-square of 321.6, a df value of 260, and chi-square minimum/df of 1.24, which was between 1 and 3. The goodness-of-fit index, measuring how much of the variance and covariance of the observed data can be explained by the model, was 0.93 greater than the threshold of 0.90 and the root mean square error of approximation value, which measures the gap between the theoretical model and the perfect fitting model, was 0.03 less than the threshold of 0.05, indicating that the model had a good absolute fit. The normed fit index value, incremental fit index value, Tucker-Lewis Index, and comparative fit index explain the degree of fit of the hypothetical model and the independent model from various aspects, whose values were all greater than the threshold of 0.90, meaning that the model had a good added fit with the actual data. The Parsimony-adjusted normed fit index value, Parsimony-adjusted comparative fit index value, and Parsimony-adjusted goodness-of-fit index explain the simplicity of the model based on the assessment of fit. The values were all greater than the threshold of 0.5, indicating that the model had a reduced fit [28]. Overall, the fitness of the structural model was good.

Table 5. Goodness-of-fit test.

Indicators	Model indicator values	Reference standard	Whether or not fit
Absolute fit			
Chi-square minimum/df	1.24	<3	Fit
Root mean square error of approximation	0.03	<0.05	Fit
Goodness-of-fit index	0.93	>0.9	Fit
Added fit			
Normed fit index	0.96	>0.9	Fit
Incremental fit index	0.99	>0.9	Fit
Tucker-Lewis index	0.99	>0.9	Fit
Comparative fit index	0.99	>0.9	Fit
Reduced fit			
Parsimony-adjusted normed fit index	0.83	>0.5	Fit
Parsimony-adjusted comparative fit index	0.86	>0.5	Fit
Parsimony-adjusted goodness-of-fit index	0.75	>0.5	Fit

The results of structural model are reported in Table 6, which presents the estimated coefficients, the standardized and nonstandardized coefficients, standard error, critical ratio, and the associated *P* values. The coefficients represented the factor loading from patient satisfaction to the 5 dimensions of reliability, assurance, responsiveness, empathy, and tangibility. Standard error was used to judge the relationship between different variables in the model, as well as its strength and direction, the smaller and better the standard error, indicating that the parameter estimation was more reliable, and vice versa, indicating that the parameter estimation of the uncertainty was greater. Critical ratio was used to judge whether the variables in the model had a good continuity. If it was greater than or equal to 1, it showed that the model had good continuity and could be considered existing important and stable relationships between the variables. On the contrary, it meant that the model had a large uncertainty and need to be further modified. A *P* value of less than .05 indicated that the dimension had a positive effect on patient satisfaction.

XSL•FO
RenderX

Table 6. Path coefficients and significance level analysis.

	Standardized coefficient	Nonstandardized coefficient	Standard error	Critical ratio	*P* value
Reliability	0.15	0.13	0.05	2.39	.02
Assurance	0.12	0.11	0.05	2.10	.04
Responsiveness	0.15	0.14	0.06	2.56	.01
Empathy	0.15	0.13	0.05	2.50	.01
Tangibility	0.25	0.22	0.05	4.28	.01

As shown in Table 6, the standardized coefficients of the 5 dimensions of assurance, reliability, responsiveness, empathy, and tangibility on patient satisfaction were all greater than 0, and the *P* value was less than .05, which proved that there was a significantly positive correlation between all 5 dimensions on the overall satisfaction of patients in the internet hospitals, that is, the improvement of the 5 dimensions could correspondingly affect the improvement of overall patient satisfaction. The standard errors were all positive and small in value, indicating that there was a positive effect of the 5 dimensions on overall satisfaction, and the parameter estimation of the path was reliable. The critical values were all greater than 1, indicating that the model had good continuity and there was an important and significant relationship between the 5 dimensions and overall patient satisfaction. In summary, it could be seen that the 5 dimensions played a positive role in predicting overall patient satisfaction.

Discussion

Main Findings and Suggestions

This study aims to investigate the factors influencing patient satisfaction in the multistage online consultation process based on 5 dimensions of the SERVQUAL theoretical model. To summarize, the current consultation process of internet hospitals as perceived by patients did not yet meet their expectations, and the overall satisfaction rate of patients was only 7.15. We will discuss this in detail from the following aspects.

First, our study reveals a critical paradox: among the various dimensions measuring the service quality of internet hospitals, tangibility received the lowest patient ratings, but it was identified as the most influential factor affecting overall satisfaction. The results are similar to other studies using the SERVQUAL model that emphasized the importance of tangibility [19,20]. The intangible nature of the service remains a primary bottleneck for the development of internet health care, while successfully making the virtual service tangible can yield significant returns in terms of patient trust and satisfaction. However, improving tangibility necessitates a profound experience redesign. The online user interface should evolve beyond basic functionality to become a communication medium itself, which can be achieved through visual service flow maps and multidimensional presentations of the medical teams, thereby making the intangible diagnostic and treatment process clear, predictable, and structured [10]. Besides, hospitals should integrate online and offline support systems to create a seamless chain of trust by establishing end-to-end tracking systems for prescription approval, medication delivery, and offline examination appointments. In summary, the future development of internet hospitals should not solely focus on stacking technological features but must strive to transform cold online interactions into a warm and perceptible service experience. By implementing the proposed 2-pronged strategy for enhancing tangibility, internet hospitals can effectively mitigate their inherent disadvantage of service intangibility, which will pave the way for building robust patient satisfaction and achieving sustainable growth.

Second, the findings clearly show that reliability is a core factor driving patient satisfaction in internet hospitals. Reliability assesses the ability of internet hospitals to deliver services accurately and dependably [4,18]. To strengthen reliability, internet hospitals must strive to consistently fulfill their promises to patients throughout the entire process from preconsultation, during consultation, to postconsultation. First, preconsultation focuses on establishing clear expectations, requiring physicians to adhere strictly to appointment times and ensuring complete fee transparency through detailed cost breakdowns. Second, during and postconsultation, the emphasis shifts to delivering a high-quality service loop. Previous studies have shown that online physicians are a key factor in improving patient satisfaction [11]; internet hospitals should be aware of the importance of physicians' appearance and smile on patients' selection and evaluation behavior. Physicians should communicate with professionalism and empathy (eg, a smile) to build trust and compensate for the lack of face-to-face interaction. Concurrently, drug delivery must be punctual, accurate, and with intact packaging, representing the final fulfillment of the service promise. Finally, underpinning all services is technical assurance. Internet hospitals must prioritize network security and patient privacy protection, maintained by dedicated IT teams to ensure system stability and data security [8]. In essence, by standardizing processes, ensuring transparency, and rigorously honoring commitments within this framework, internet hospitals can systematically build reliability, leading to a significant increase in patient satisfaction.

Responsiveness also demonstrated a significant positive impact on patient satisfaction in internet hospitals. Notably, among the 5 SERVQUAL dimensions, responsiveness received the highest score from the participants. This dimension pertains to the service efficiency of internet hospitals, reflected specifically in the promptness of staff replies to patient inquiries and whether the exact time for each service is communicated in advance [29]. To continuously improve responsiveness, internet hospitals need to optimize multiple aspects. First, IT support should be strengthened to ensure that medical staff can promptly address patient needs during consultations and provide quick feedback

XSL•FO
RenderX

and solutions to postconsultation issues. Second, in response to patient-reported problems such as difficulty contacting doctors and delayed responses, root causes such as network latency, system lag, or unreasonable interface design should be thoroughly analyzed and addressed through technological upgrades and process optimization. Additionally, internet hospitals should focus on enhancing service synergy, reducing communication delays, and avoiding consultation interruptions or efficiency losses due to poor communication. On the other hand, internet hospitals should further expand the application scenarios of responsiveness. For example, timely push of hospital resource information (eg, doctor schedules, latest services) and the implementation of rapid response mechanisms for payment, refunds, and other processes can comprehensively improve service efficiency and patient experience. In the future, internet hospitals can leverage intelligent technologies (eg, AI pre-replies, smart triage) to enhance responsiveness, reducing the workload of medical staff while further improving service efficiency and patient satisfaction.

Fourth, a significant positive correlation exists between empathy and patient satisfaction in internet hospitals. Improving the empathy of internet hospitals needs to focus on whether patients feel cared for in the whole process of diagnosis and treatment [30]. When internet hospitals provide personalized services, demonstrate genuine concern, and proactively address patient needs, patients feel valued, which fosters positive emotions and, consequently, increases their satisfaction [3]. For instance, internet hospitals should cater to the needs of special patient groups such as older adults and persons with disabilities. Implementing dedicated application modes such as an “Elderly Mode” with optimized font styles and sizes can directly address their specific requirements and reflect a commitment to inclusive, humanistic care. Furthermore, as internet hospitals primarily serve patients with chronic conditions or those facing mobility challenges, the scheduling system must be patient-centric. The allocation of appointment slots and the number of available time slots should not only align with patient demand but also be balanced reasonably with physicians' available consultation hours. Additionally, offline hospitals can enhance empathy by establishing dedicated service counters for their internet hospital services. These counters could assist patients with practical tasks such as printing electronic prescriptions or reports and serve as an introduction point for first-time users, guiding them through the platform's features. This integration of online and offline support embodies a holistic and empathetic approach to patient care.

Finally, among the 5 service quality dimensions, assurance was found to have the least impact on patient satisfaction, a finding that aligns with practical realities. In the context of this study, assurance primarily pertains to the technical competence of online physicians and the effectiveness of the treatment received. Patients who generally lack professional medical knowledge and face information asymmetry are often unable to directly evaluate this dimension, making the sense of assurance less tangible in an internet hospital setting. Nonetheless, it is indisputable that physicians' professional knowledge remains the cornerstone of a credible internet hospital service [14]. Therefore, entities operating these platforms, including traditional hospitals, must prioritize the rigorous selection and training of online physicians. Deploying high-caliber medical talent is essential to enhance the perceived assurance and, consequently, patient satisfaction. To this end, offline hospitals should assign their most qualified medical teams to online duties, ensuring that the quality of consultation achieves parity with in-person visits. Furthermore, strengthening the integration between online and offline services is crucial. By creating seamless referral channels and strategically shifting appropriate follow-up care burdens from offline to online platforms, health care providers can build a cohesive and trustworthy system. This integration not only enhances assurance but also encourages patient adoption and sustained use of internet hospitals throughout their entire health care journey.

Conclusion and Limitations

This study analyzes patient satisfaction in the multistage online consultation process in internet hospitals. Based on the SERVQUAL model, we designed questionnaire questions for the 5 dimensions of reliability, assurance, responsiveness, empathy, and tangibility, and collected data from patients of the internet hospital of a tertiary hospital in Beijing. We found that the 5 factors of reliability, responsiveness, empathy, tangibility, and assurance all had different positive predictive effects on patient satisfaction in internet hospitals to a certain extent. Our results provide patients, physicians, and online health communities with guidelines that support them in widely adopting internet hospitals and effectively improving their decision-making.

This paper has some shortcomings while achieving certain results. First, this paper determined 25 observed variables of the 5 dimensions according to previous literature and the actual situation of a tertiary hospital internet hospital in Beijing. However, the operation mode of each internet hospital is different, and the applicability of the variable selection is worth considering. Second, the cross-sectional design limits the ability to establish causal relationships between the 5 dimensions of SERVQUAL theoretical model and patient satisfaction. Future research could use longitudinal studies to examine the pathways of causality and the possible changes in patient satisfaction over time. In addition, the survey respondents were patients using online health services in Beijing. As an emerging thing, patients in different regions may present different satisfaction levels due to ethnic and cultural differences, and future research could explore patients in different regions.

Acknowledgments

We are thankful for the support of the General Project of Social Sciences Program of Beijing Municipal Education Commission (SM202310025001).

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Measurement items of constructs.
[DOCX File , 18 KB - humanfactors_v13i1e79213_app1.docx]

References

1. Han Y, Lie RK, Guo R. The internet hospital as a telehealth model in China: systematic search and content analysis. J Med Internet Res 2020 Jul 29;22(7):e17995 [FREE Full text] [doi: 10.2196/17995] [Medline: 32723721]
2. Lai Y, Chen S, Li M, Ung COL, Hu H. Policy interventions, development trends, and service innovations of internet hospitals in China: documentary analysis and qualitative interview study. J Med Internet Res 2021 Jul 20;23(7):e22330 [FREE Full text] [doi: 10.2196/22330] [Medline: 34283025]
3. Han T, Wei Q, Wang R, Cai Y, Zhu H, Chen J, et al. Service quality and patient satisfaction of internet hospitals in China: cross-sectional evaluation with the service quality questionnaire. J Med Internet Res 2024 Nov 08;26:e55140 [FREE Full text] [doi: 10.2196/55140] [Medline: 39514849]
4. Li D, Hu Y, Liu S, Li G, Lu C, Yuan S, et al. The effect of using internet hospitals on the physician-patient relationship: patient perspective. Int J Med Inform 2023 Jun;174:105058. [doi: 10.1016/j.ijmedinf.2023.105058] [Medline: 37002986]
5. Zhi L, Yin P, Ren J, Wei G, Zhou J, Wu J, et al. Running an internet hospital in China: perspective based on a case study. J Med Internet Res 2021 Sep 16;23(9):e18307 [FREE Full text] [doi: 10.2196/18307] [Medline: 34342267]
6. Xu X, Cai Y, Wu S, Guo J, Yang L, Lan J, et al. Assessment of internet hospitals in China during the COVID-19 pandemic: national cross-sectional data analysis study. J Med Internet Res 2021 Jan 20;23(1):e21825 [FREE Full text] [doi: 10.2196/21825] [Medline: 33417586]
7. Wu DC, Zhao X, Wu J. Online physician-patient interaction and patient satisfaction: empirical study of the internet hospital service. J Med Internet Res 2023 Aug 24;25:e39089 [FREE Full text] [doi: 10.2196/39089] [Medline: 37616031]
8. Chen S, Lai K, Guo X, Zhang X. The influence of digital health technology on the allocation of regional medical resources in China. Health Policy and Technology 2025 Jun;14(3):101013. [doi: 10.1016/j.hlpt.2025.101013]
9. Jannati N, Nakhaee N, Yazdi-Feyzabadi V, Tjondronegoro D. A cross-sectional online survey on patients' satisfaction using store-and-forward voice and text messaging teleconsultation service during the COVID-19 pandemic. Int J Med Inform 2021 Jul;151:104474 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104474] [Medline: 33965682]
10. Chang Y, Hsu P, Wang Y, Chang P. Integration of online and offline health services: the role of doctor-patient online interaction. Patient Educ Couns 2019 Oct;102(10):1905-1910. [doi: 10.1016/j.pec.2019.04.018] [Medline: 31279612]
11. Yang Y, Ma C, Liu H. The beauty premium: physicians' facial expressions and patients' selection and evaluation behavior in online health communities. Health Policy and Technology 2024 Sep;13(4):100895. [doi: 10.1016/j.hlpt.2024.100895]
12. Aiken LH, Sermeus W, Van den Heede K, Sloane DM, Busse R, McKee M, et al. Patient safety, satisfaction, and quality of hospital care: cross sectional surveys of nurses and patients in 12 countries in Europe and the United States. BMJ 2012 Mar 20;344:e1717 [FREE Full text] [doi: 10.1136/bmj.e1717] [Medline: 22434089]
13. Aiken LH, Lasater KB, Sloane DM, Pogue CA, Fitzpatrick Rosenbaum KE, Muir KJ, US Clinician Wellbeing Study Consortium. Physician and nurse well-being and preferred interventions to address burnout in hospital practice: factors associated with turnover, outcomes, and patient safety. JAMA Health Forum 2023 Jul 07;4(7):e231809 [FREE Full text] [doi: 10.1001/jamahealthforum.2023.1809] [Medline: 37418269]
14. Song H, Guo X, Wu T. Exploring the impact of free live-streamed medical consultation on patient engagement and patient satisfaction in the multistage online consultation process: a quasi-experimental design. Decision Support Systems 2025 Apr;191:114422. [doi: 10.1016/j.dss.2025.114422]
15. Guo S, Guo X, Fang Y, Vogel D. How doctors gain social and economic returns in online health-care communities: a professional capital perspective. Journal of Management Information Systems 2017 Aug 17;34(2):487-519. [doi: 10.1080/07421222.2017.1334480]
16. Parasuraman A, Zeithaml VA, Berry LL. A conceptual model of service quality and its implications for future research. Journal of Marketing 1985 Sep 01;49(4):41-50. [doi: 10.1177/002224298504900403]
17. Parasuraman A, Zeithaml V, Berry L. SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. J Retail 1988;64(1):A [FREE Full text]
18. Alizadeh T, Noveiry MJS, Karkhah S, Leyli EK, Kohan KS, Vajargah PG. Service quality in Iranian hospitals from patients' perspectives based on the SERVQUAL model: a systematic review and meta-analysis. Health Sci Rep 2025 Jan;8(1):e70362 [FREE Full text] [doi: 10.1002/hsr2.70362] [Medline: 39846046]
19. Jonkisz A, Karniej P, Krasowska D. SERVQUAL method as an "Old New" tool for improving the quality of medical services: a literature review. Int J Environ Res Public Health 2021 Oct 13;18(20):10758 [FREE Full text] [doi: 10.3390/ijerph182010758] [Medline: 34682499]

XSL•FO
RenderX

20. Jonkisz A, Karniej P, Krasowska D. The SERVQUAL method as an assessment tool of the quality of medical services in selected Asian countries. Int J Environ Res Public Health 2022 Jun 26;19(13):7831 [FREE Full text] [doi: 10.3390/ijerph19137831] [Medline: 35805492]
21. Notice on Issuing the Measures for Ethical Review of Life Science and Medical Research Involving Human Subjects. Chapter Three: Ethical Review (Article 32). National Health Commission of the People's Republic of China. 2023. URL: https://www.nhc.gov.cn/qjjys/c100016/202302/6b6e447b3edc4338856c9a652a85f44b.shtml [accessed 2026-03-02]
22. Anderson JC, Gerbing DW. Structural equation modeling in practice: a review and recommended two-step approach. Psychological Bulletin 1988 May;103(3):411-423. [doi: 10.1037/0033-2909.103.3.411]
23. Hair JF, Ringle CM, Sarstedt M. PLS-SEM: indeed a silver bullet. J Mark Theory Pract 2014 Dec 08;19(2):139-152. [doi: 10.2753/mtp1069-6679190202]
24. Fornell C, Bookstein FL. Two structural equation models: LISREL and PLS applied to consumer exit-voice theory. J Mark Res 1982 Nov 01;19(4):440-452. [doi: 10.1177/002224378201900406]
25. Park EJ, Kim HY, Nho J, Ko E, Boyes AW. Validity and reliability of the Korean version of supportive care needs survey-short form 34 for patients with cancer: a methodological study. Asian Nurs Res (Korean Soc Nurs Sci) 2025 Feb;19(1):60-68 [FREE Full text] [doi: 10.1016/j.anr.2024.12.003] [Medline: 39701517]
26. Hair JF, Risher JJ, Sarstedt M, Ringle CM. When to use and how to report the results of PLS-SEM. EBR 2019 Jan 14;31(1):2-24. [doi: 10.1108/ebr-11-2018-0203]
27. George D, Mallery. SPSS for Windows Step by Step: A Simple Guide and Reference18.0 Update (11th ed). Boston, MA: Allyn & Bacon; Nov 12, 2010.
28. MacKenzie S, Podsakoff P, Podsakoff N. Construct measurement and validation procedures in MISbehavioral research: integrating newexisting techniques. MIS. Q 2011;35(2):293-334. [doi: 10.2307/23044045]
29. Goula A, Stamouli M, Alexandridou M, Vorreakou L, Galanakis A, Theodorou G, et al. Public hospital quality assessment. Evidence from greek health setting using SERVQUAL model. Int J Environ Res Public Health 2021 Mar 25;18(7):3418 [FREE Full text] [doi: 10.3390/ijerph18073418] [Medline: 33806126]
30. Wu H, Li M, Zhang L. Comparing physician and large language model responses to influenza patient questions in the online health community. Int J Med Inform 2025 May;197:105836. [doi: 10.1016/j.ijmedinf.2025.105836] [Medline: 39986122]

Abbreviations

SERVQUAL: service quality

Edited by A Kushniruk; submitted 17.Jun.2025; peer-reviewed by Y Chow, M Al Zoubi; comments to author 20.Sep.2025; revised version received 23.Dec.2025; accepted 02.Jan.2026; published 16.Mar.2026.

Please cite as:
Wang Y, Han Y, Duan M, Wang Y
Factors Influencing Patient Satisfaction Within the Multistage Online Consultation Process in Internet Hospitals in Beijing: Empirical Analysis
JMIR Hum Factors 2026;13:e79213
URL: https://humanfactors.jmir.org/2026/1/e79213
doi:10.2196/79213
PMID:

©Yi Wang, Yujia Han, Mengyu Duan, Yani Wang. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 16.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Exploring English and Swedish General Practitioners' Behavioral Intentions to Use Telemedicine: Comparative Study

Michael Harris[1,2*], PhD; Gordon Taylor[1], PhD; Miriam Pikkemaat[3,4*], PhD; Hans Thulesius[3*], PhD; Veronica Milos Nymberg[3,4*], PhD

[1]College of Medicine and Health, University of Exeter, Exeter, United Kingdom

[2]Bern Institute of Primary Health Care (BIHAM), University of Bern, Bern, Switzerland

[3]Center for Primary Health Care Research, Department of Clinical Sciences Malmö, Lund University, Jan Waldenströms gata 35, Malmö, Sweden

[4]University Clinic Primary Care, Skåne University Hospital, Region Skåne, Sweden

[*]these authors contributed equally

Corresponding Author:
Veronica Milos Nymberg, PhD
Center for Primary Health Care Research, Department of Clinical Sciences Malmö, Lund University, Jan Waldenströms gata 35, Malmö, Sweden

Abstract

Background: Although telemedicine grew rapidly during the COVID-19 pandemic, instruments to assess general practitioners' (GPs) attitudes and behavioral intentions to use it are scarce. In Sweden, the Physicians' Attitudes and Intentions to use Telemedicine (PAIT) questionnaire was developed from the "theory of planned behavior" in 2019 and translated into English in 2022.

Objective: The aim of this study was to explore similarities and differences between behavioral intentions and predictors of intentions to use telemedicine among GPs in England and Sweden.

Methods: This study compared attitudes, behavioral intentions, and self-reported use of telemedicine after the COVID-19 pandemic among 52 GPs in England and 101 GPs in Sweden. The PAIT questionnaire has 33 items with 7-point Likert scale options ranging from "strongly disagree" to "strongly agree," examining 3 predictors of intentions: attitudes (12 items), subjective norms (6 items), perceived behavioral control (9 items), and "intentions" (6 items) to use telemedicine; 22 items assess use of telemedicine tools, general questions about telemedicine, training experience, free-text comments, and demographic and background questions.

Results: Both English and Swedish GPs reported little training and low use of telemedicine after the COVID-19 pandemic. Swedish GPs had significantly higher mean scores for intentions to use telemedicine in daily practice compared with English GPs. More positive attitudes and higher perceived behavioral control were significantly associated with higher behavioral intention scores in both English and Swedish GPs.

Conclusions: While our results are exploratory due to sample size constraints, these findings provide insights into the similarities and differences between English and Swedish GPs regarding telemedicine adoption—attitudes, behavioral intentions, and self-reported use of telemedicine assessed by the PAIT questionnaire—which proved useful for cross-country comparisons and could be used for further international studies.

(JMIR Hum Factors 2026;13:e73609) doi:10.2196/73609

KEYWORDS

telemedicine; general practitioners; GP; theory of planned behavior; behavioral intentions; attitudes; perceived behavioral control; subjective norms

Introduction

Background

Telemedicine has developed at a rapid pace and increased in use worldwide in recent years [1,2], partly due to social distancing policies introduced during the COVID-19 pandemic [3-6]. Current telemedicine technologies are available at a low cost, accessible to those living in rural areas, and they have the potential to decrease the workload for general practitioners (GPs) [7,8].

Telemedicine is defined by the United Kingdom's National Health Service as "The use of telecommunication and information technology providing remote health assessments and therapeutic interventions" [9]. Another definition is "health care service delivered by health care providers in a patient-centered manner, from a geographical distance using ICT (information and communication technology)" [10].

XSL•FO
RenderX

Telehealth is a broader term encompassing the integration of telecommunication in the practice, promotion, and management of health care. Methods of communication in telemedicine include video consultations, telephone calls, emails, as well as mobile apps to monitor chronic diseases (mobile health), digital diagnostic algorithms, triage, and web guideline support for treatments [7,8,11,12].

While studies in the United Kingdom have shown that GPs have a positive attitude toward telemedicine, some have criticized the lack of IT integration with existing medical systems and reported a lack of telemedicine experience, especially before the COVID-19 pandemic [3,7]. Surveys in the United States and Sweden have shown similar conclusions: positive attitudes toward telemedicine but little use of it [11,13-16]. The main obstacles for use have been reported to be a lack of experience and understanding of the use of these communication methods and technologies [13,16]. The increased use of telemedicine in health care [2,6] has raised concerns about both GP workload increase and decrease [13], with Salisbury et al [17] warning of an increase in GPs' workload when integrating telemedicine as a means of digital consultation before face-to-face consultation. Workload is only decreased if at least half of digital consultations are nonurgent matters, and it depends on the demand and requirements of the consultations [17]. There has been a call for research on telehealth and telemedicine to evaluate how they could improve medical consultations and disease monitoring [1,14]. In this study, telemedicine refers to remote clinical consultations between patients and clinicians, such as by telephone, video, chat, or secure email. It does not include remote monitoring devices, digital diagnostics, or triage algorithms.

Theoretical Framework

Over the past few decades, several theories have emerged to explain end-user acceptance and adoption of new technologies [10,18-22]. One of the most commonly applied frameworks for understanding telemedicine adoption among health care staff is the theory of planned behavior (TPB).

The TPB explains how behavioral intentions influence the likelihood of performing an action. According to the theory, these intentions are shaped by an individual's attitudes toward the behavior, subjective norms—defined as the individual's perception of social expectations regarding behavior change—and perceived behavioral control, which reflects the individual's belief in their ability to perform the behavior [23].

The TPB is widely recognized for its practical applicability and predictive validity. A validated manual is available for implementing the TPB, enabling the development of questionnaires designed to assess behavioral intentions [23].

This study aims to build upon the Swedish study by Pikkemaat et al [11], conducted in 2019 before the COVID-19 pandemic. That study examined physicians' attitudes and intentions toward using telemedicine, using a newly designed survey instrument based on the TPB. The Swedish-language instrument, known as the Physicians' Attitudes and Intentions to use Telemedicine (PAIT) survey, was distributed to 820 potential respondents across 160 health care centers in 2 southern Swedish regions, yielding 198 (24.1%) responses. The study found high levels of intention and positive attitudes toward using telemedicine, with perceived behavioral control emerging as the strongest predictor of telemedicine adoption. Given that the COVID-19 pandemic accelerated the use of telemedicine, the study suggested that intentions to use digital tools in primary care may have increased since 2019, warranting further exploration across different contexts.

To date, no studies have examined the intentions of English GPs to adopt telemedicine. Building on insights from the Swedish study conducted before the COVID-19 pandemic [11] and the assumption that telemedicine adoption may have shifted in primary care following the pandemic, we decided to collect new survey data from Sweden and compare it with data from England, both within the primary care setting.

The objective of this study was to explore and compare the intentions of English and Swedish GPs to use telemedicine using the TPB-validated questionnaire. Additionally, the study aimed to assess how attitudes, subjective norms, and perceived behavioral control influence the intentions of GPs in both countries to adopt telemedicine.

Methods

Survey Design

The PAIT survey instrument was originally designed and written in Swedish by Pikkemaat et al [11], based on the manual from the University of London Institutional Repository for creating questionnaires using the TPB, constructed by Francis et al [24]. The PAIT survey includes questions about experiences with digital contact methods, chronic disease monitoring with digital tools, and participant demographics. Digital contact methods encompass the various telemedicine approaches GPs might use to communicate with patients, such as telephone, online chat, video consultation, and email.

The original PAIT-Swedish survey instrument also included questions regarding experiences with artificial intelligence (AI) in telemedicine, focusing on diagnostic algorithms, triage, and support for treatment guidelines. However, as responses from the Swedish survey indicated very low use and experience of AI technologies among GPs [11], these questions were excluded from the English survey.

Translation and Validation Process

The PAIT survey was translated into English by a Swedish-English bilingual master's student, with minor cultural adaptations to suit the UK health care context. To ensure linguistic equivalence and conceptual accuracy, the English version underwent back-translation into Swedish, followed by a detailed comparison with the original Swedish version. This process was overseen by the research team, ensuring that no essential meaning was lost during translation.

The newly adapted English-language version, named "Behavioural Intentions towards Telemedicine" (BIT), was designed to yield scores comparable to both the pre–COVID-19 pandemic PAIT-Swedish survey [11] and the results from a

XSL•FO
RenderX

repeat PAIT-Swedish survey conducted following the COVID-19 pandemic (Multimedia Appendix 1).

Validation and Reliability Testing

The translated BIT survey was validated following established guidelines, including the Checklist for Reporting Results of Internet E-Surveys (CHERRIES), as recommended by the *Journal of Medical Internet Research* [25] (Checklist 1). The PAIT-Swedish version had previously undergone rigorous testing for construct validity and reliability using the Churchill Paradigm and COSMIN (Consensus-based Standards for the Selection of Health Measurement Instruments) guidelines. This involved a pilot test with 24 participants and evaluation by 5 proficient physicians in telemedicine to ensure the questionnaire accurately captured the research questions.

Internal consistency was assessed using Cronbach α, with items scoring below an acceptable threshold removed to improve the instrument's reliability. The final version demonstrated Cronbach α scores ranging between 0.6 and 0.95. Temporal stability was further tested by administering the questionnaire to the same 24 respondents after 2 weeks, with 12 respondents completing the retest. The Pearson correlation coefficient for test-retest reliability was 0.63, prompting minor adjustments to question structures for the final version.

Survey Distribution

The anonymous web-based BIT survey was distributed to GPs located in southwest England between March and May 2023. We contacted GPs through email and social media (Twitter and Facebook), as well as through the Royal College of General Practitioners, the British Medical Journal, GP magazine, Pulse magazine, Local Medical Committees, "GP online," "eLearning," "GP survival," and Somerset GP Education. An introductory email, including the survey or a request to share it, was sent to the managers of these groups, explaining the purpose of the research. The survey was only distributed if the group managers agreed to facilitate its circulation.

The survey for Swedish GPs was distributed between March and May 2022 to managers at the same 160 primary health care centers in southern Sweden as in the 2019 survey.

Statistics

Given the small English sample size (n=52) that limited the power of the study, analyses were exploratory. Likert scale responses were converted to numerical scores, and demographic responses were coded accordingly. Likert scale scores ranged from 1 to 7, where 1 represented "not at all," "harmful," "bad," "unpleasant," or "useless," and 7 represented "to a great extent," "beneficial," "good," "pleasant," or "useful."

We converted respondents' years of practice into 3 categories (<15 years, 15 - 29 years, and ≥30 years). For each respondent and topic, we calculated the mean intention scores from questions 17 and 18, covering email and/or online chat, video consultation, and digital tools for chronic disease monitoring.

We also calculated scores for the core concepts of the TPB that influence behavioral intentions. This included mean scores for subjective norms (questions 15 and 16), perceived behavioral control (questions 12 to 14), and attitudes (questions 6, 8, and 11). Self-reported current use of telemedicine included using different digital tools (email, chat, and video). Data were analyzed using SPSS (version 28.0; IBM Corp). Descriptive statistics were generated separately for English and Swedish GPs across the various outcomes. Group differences between English and Swedish GPs were tested using the 2-tailed Student *t* test.

To explore the association between predictor scores and behavioral intentions to use telemedicine among English and Swedish GPs, we performed a split-sample multiple regression analysis.

Data analysis followed a consistent approach for both the English and Swedish surveys, with one exception: in the Swedish survey, respondents could select "I don't know" for Likert scale questions. These responses were categorized as "Not Applicable" during the analysis.

Ethical Considerations

The English ethical review was conducted by the University of Exeter Medical School's Research Ethics Committee. Ethics approval was granted by the review board in March 2023 (845606). The study was carried out in accordance with the Declaration of Helsinki (Fortaleza, Brazil, October 2013), the UK Policy Framework for Health and Social Care Research (2020), and the general principles of Good Clinical Practice E6 (R2). Data storage complied with UK General Data Protection Regulation requirements and was securely maintained within the University of Exeter database. The study was deemed to pose no risk and did not involve the collection of any personal, health, or sensitive data.

The Swedish Ethical Review Authority provided an advisory statement confirming that, according to the Swedish Ethical Review Act (SFS 2003:460), ethical vetting was not required for this type of study because it involved anonymous data and did not inquire about participants' own health or other sensitive topics. The statement provided by the Swedish Ethical Review Authority in November 2018 states the following: "In accordance with current ethical and data-protection standards, the project does not require formal ethical review, as no personal data are processed and the information utilized cannot be linked to identifiable living individuals. Should it later become evident that personal data are involved, only the processing of sensitive personal data—such as information pertaining to racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, or data relating to health or sexual life—would necessitate formal ethical approval."

All participants who returned questionnaires provided informed consent to participate in the study. No monetary compensation was offered to the participants.

Results

We received 52 responses from English GPs and 101 responses from Swedish GPs. The demographics of the respondents varied significantly between the two groups (Table 1). The mean age

XSL•FO
RenderX

of English GPs was 51.2 years, compared with 47.8 years for Swedish GPs.

Table . Demographic characteristics of respondents.

Characteristics	English GPs[a] (n=52), n (%)	Swedish GPs (n=101), n (%)
Gender		
Women	37 (71.2)	51 (50.5)
Men	15 (28.8)	48 (47.5)
Nonbinary	0 (0.0)	1 (1.0)
Training in telemedicine		
Yes	17 (32.7)	37 (36.6)
No	35 (67.3)	64 (63.4)
Practice (years)		
<15	8 (15.4)	57 (56.4)
15 - 29	23 (44.2)	36 (35.6)
≥30	21 (40.4)	8 (7.9)

[a]GP: general practitioner.

Gender distribution also differed, among 52 English GPs, 15 (28.8%) identified as men and 37 (71.2%) as women, while among 101 Swedish GPs, 48 (47.5%) identified as men and 51 (50.5%) as women.

Regarding telemedicine training, 17 (32.7%) English GPs reported having received such training, compared with 37 (36.6%) Swedish GPs. Additionally, English GPs had more years of practice on average than their Swedish counterparts.

Swedish GPs had significantly (P<.001) higher intentions to use email and/or online chat, video consultations, and digital tools to monitor chronic diseases compared with their English counterparts (Table 2).

The theoretical predictors of behavior derived from the TPB revealed significant differences between English and Swedish GPs, with Swedish GPs scoring higher for both intentions and predictors of behavioral intentions to use telemedicine (Table 3).

Table . Comparison of intentions to use different telemedicine tools between English and Swedish general practitioners (GPs).

Topics	English GPs (n=52), mean (SD)	Swedish GPs (n=101), mean (SD)	*P* value
Email and/or online chat	3.6 (2.3)	4.9 (1.8)[a]	<.001
Video consultation	2.8 (1.9)[b]	4.7 (2.0)[c]	<.001
Digital tools to monitor chronic diseases	4.3 (1.9)	5.6 (1.8)[d]	<.001

[a]n=99.
[b]n=50.
[c]n=100.
[d]n=97.

Table . Distribution of mean scores for theoretical constructs derived from the theory of planned behavior predictors (subjective norms, attitudes, and perceived behavioral control) on intentions among English and Swedish general practitioners (GPs).

Constructs	English GPs (n=52), mean (SD)	Swedish GPs (n=101), mean (SD)	*P* value
Subjective norms	2.8 (1.3)	3.4 (1.4)	.03
Attitudes	4.1 (1.1)[a]	4.1 (1.1)[b]	<.001
Perceived behavioral control	3.6 (1.0)	4.4 (1.1)	<.001
Intentions	3.7 (1.6)	4.8 (1.7)[c]	<.001

[a]n=51.
[b]n=96.
[c]n=99.

XSL•FO
RenderX

Overall, the self-reported use of telemedicine was low among both English and Swedish GPs (Table 4). However, the mean scores for self-reported current use indicated that English GPs were significantly more likely to use email and/or online chat (P=.006). There were no significant differences between the two groups regarding the use of video consultations (P=.76) or digital tools for monitoring chronic diseases (P=.73).

For both English and Swedish GPs, significant associations were observed between attitudes, perceived behavioral control, and intentions to use telemedicine (Table 5). The model had a higher explanatory power for the English GPs (R^2=0.664) than for the Swedish GP (R^2=0.258). However, given the modest sample sizes, particularly in the English cohort, the multivariable regression analyses should be interpreted cautiously as they may be underpowered to detect smaller effect sizes.

Table . Self-reported current use of email and/or online chat, video consultations, and digital tools for chronic disease monitoring among English and Swedish general practitioners (GPs).

Tools	English GPs (n=52), mean (SD)	Swedish GPs (n=101), mean (SD)	*P* value
Email and/or online chat	2.2 (1.2)	1.7 (1.0)	.006
Video consultation	2.0 (1.5)	1.9 (1.3)[a]	.76
Digital tools to monitor chronic diseases	2.2 (1.7)[b]	2.1 (1.6)[c]	.73

[a]n=96.

[b]n=50.

[c]n=98.

Table . Split sample multiple regression analysis with intention scores from English and Swedish general practitioners (GPs) as dependent variables.

Predictors[a]	Intentions of English GPs[b]		Intentions of Swedish GPs[c]	
	U-B[d] (95% CI[e])	*P* value	U-B (95% CI)	*P* value
Subjective norms	0.082 (−0.139 to 0.303)	.46	0.042 (−0.174 to 0.258)	.70
Attitudes	0.783 (0.453 to 1.113)	<.001	0.4 (0.124 to 0.676)	.005
Perceived behavioral control	0.497 (0.141 to 0.853)	.007	0.52 (0.244 to 0.796)	<.001
Constant	−1.58 (−2.765 to −0.387)	.01	0.451 (−1.303 to 2.204)	.61

[a]R^2 indicates the proportion of variance in the dependent variable (intention to use telemedicine) explained by the independent variables (attitudes, subjective norms, and perceived behavioral control).

[b]R^2=0.664.

[c]R^2=0.258.

[d]U-B represents the unstandardized B value, testing whether the coefficient is equal to zero.

[e]The 95% CI reflects the range within which the true B value is likely to fall.

Discussion

Principal Results

A major finding in this study is the low self-reported use of telemedicine among both English and Swedish GPs, even after the COVID-19 pandemic. Practitioners from both countries also reported limited training in telemedicine. Despite this, Swedish GPs demonstrated significantly higher mean scores for intentions to use telemedicine in daily practice compared with their English counterparts. More positive attitudes and greater perceived behavioral control were significantly associated with higher behavioral intention scores among both English and Swedish GPs. In addition, the consistently low levels of reported training in both countries suggest that insufficient training may itself be an important contributor to lower intentions to use telemedicine, underscoring the need for more structured educational efforts.

Strengths and Limitations

A major strength of this study is the use of a validated, theory-based questionnaire (PAIT), adapted and applied in both Swedish and English primary care contexts. The questionnaire was developed based on the TPB framework and had previously been validated among Swedish GPs, demonstrating strong internal consistency and construct validity. The English version, adapted for the UK health care setting, underwent a translation and back-translation process, ensuring linguistic equivalence and conceptual accuracy. This adaptation enhances the study's ability to compare findings between English and Swedish GPs while maintaining methodological robustness.

Another strength lies in the use of theory-based survey items that have been previously used in other primary care studies, ensuring that the measured constructs are relevant and reliable. The exclusion of questions related to AI, which had shown limited relevance in the original Swedish survey, further streamlined the questionnaire for the UK primary care context.

XSL•FO
RenderX

However, the study has several limitations. The small English sample limits generalizability, so the findings should be considered exploratory. Future studies should recruit larger samples. One key limitation is the difficulty in estimating the response rate due to the recruitment methods used. As the survey was distributed through various GP networks and professional organizations, the total number of GPs who received the invitation remains unknown. Consequently, it is not possible to draw firm conclusions regarding the generalizability of the results.

Additionally, the sample of English respondents was not fully representative of the national GP population. Of the English participants, 71.2% (37/52) were female, compared with 57% of the GP workforce in England as of May 2023. This overrepresentation of female GPs may have introduced bias in the findings.

Another limitation was the challenge of identifying the specific GP groups that responded to the questionnaire. The distribution method involved contacting GP organizations and networks across southwestern England, with responses collected anonymously. While anonymity is crucial for encouraging honest responses, it also means that duplicate responses could not be detected. Nevertheless, given the time constraints faced by GPs and the voluntary nature of participation, the risk of duplicate responses is believed to be very low.

Comparison With Prior Work

The low self-reported use of telemedicine in this study is an interesting finding, despite expectations of increased use after the COVID-19 pandemic. One explanation might be that, during the COVID-19 pandemic, practices had a vital role in providing physical contact to organize and manage accurate diagnostics [18] with the purpose of avoiding hospitalization.

Our study shows that Swedish GPs had significantly higher mean scores for predictors and behavioral intentions to use telemedicine compared with English GPs, which might have several explanations. Since 2014, patients in all geographic areas in Sweden have been broadly introduced to video consultations mainly through private health care providers of digital care, leading to substantially increased costs for the regions, raising concerns regarding accessibility, inequity of care, continuity of care, and effective resource use [2]. Today, video consultations, chat-based digital tools, and a national digital contact platform are available across most Swedish regions. Many private digital health care providers have also acquired previously established primary care centers, offering integrated “digi-physical” care.

While the United Kingdom has made significant progress in health care IT, it faces ongoing challenges related to interoperability, adoption, and funding. Although access to telemedicine and digital services is improving, it remains inconsistent across regions, potentially influencing English GPs’ lower intentions to adopt telemedicine. In contrast, Sweden has developed a more integrated health care IT system with higher levels of telemedicine use, prioritizing patient convenience and accessibility.

Other Swedish studies have shown positive attitudes toward telemedicine in primary care, reporting that medical staff seem to appreciate the flexibility of working in a digital context [19], but few consider this a full-time career option [4]. While psychologists report higher intentions to use video consultations compared with physicians and nurses in a similar theory-based survey [15], physicians stress the need for evaluation of telemedicine tools regarding clinical usefulness, patient safety, and involvement of end users in the implementation process [13]. There are few studies discussing English GPs’ intentions to use telemedicine, with similar findings [3,7]. In this study, English GPs responded with low intentions to use telemedicine and little self-reported experience and training on the topic. Telemedicine and its use have been widely debated recently, regarding the practical benefits and setbacks and the optimal methods to achieve adoption [1,20]. Greenhalgh et al [26] discuss the importance of the correct adoption method of telemedicine to remain considerate of both GPs and patients, to maximize the benefits of telemedicine and minimize the setbacks.

Telemedicine appears to have a significant positive effect on many challenges that primary care faces today, such as accessibility or increased patient-centeredness [27]. There is therefore a need for a validated and efficient method of testing GPs’ intentions of using telemedicine to understand predictors of behavioral intentions and allow successful implementation.

Implications for Research and Practice

Before a more positive attitude toward the use of telemedicine can develop, it is essential to have seamless integration of telemedicine platforms with existing health care systems; comprehensive training for GPs and staff to use telemedicine tools effectively; access to remote diagnostic devices (eg, digital stethoscopes and blood pressure monitors) for patients to use at home; and increased funding for telemedicine infrastructure and services. GPs need to be involved in the development and refinement of telemedicine systems to ensure that the tools meet their needs. A cultural shift is unlikely to occur until skepticism among GPs has been addressed by promoting telemedicine’s ability to enhance accessibility and reduce unnecessary in-person visits.

The method of this study could be used for further studies either in England among English GPs or in other English-speaking countries, or it could be translated and used in other countries to strengthen the knowledge of GPs’ intentions to use telemedicine. This could help better understand the demand and benefits of certain telemedicine adoption methods in either local or global implementation structures and provide a better understanding of simple intentions. Future studies with larger and more representative samples should include training as a formal predictor of behavioral intentions, given its potential relevance for shaping GPs’ intentions to adopt telemedicine.

Conclusions

This study showed that English GPs reported lower intentions to use telemedicine compared with Swedish GPs. Practitioners from both countries also had little training in telemedicine. The behavioral intentions to use telemedicine among both English

XSL•FO
RenderX

and Swedish GPs were mainly influenced by attitudes and perceived behavioral control. The PAIT questionnaire provided valuable insights into similarities and differences between English and Swedish GPs regarding attitudes, behavioral intentions, and self-reported telemedicine use. The PAIT could serve as a useful tool for cross-country comparisons of telemedicine adoption, enabling assessments of attitudes, intentions, and perceived behavioral control across health care systems and jurisdictions. This potential is further enhanced by the recent translation of the PAIT survey into Latvian and Turkish languages by collaborating researchers, broadening its applicability for international studies.

Acknowledgments

The authors would like to thank the master's student Oliver Ras attending the medical school at the University of Exeter who translated the Swedish survey into the English context and distributed the survey to the English general practitioners. The authors are also thankful to the Swedish and English general practitioners who answered the surveys.

Funding

The authors received no financial support for the research or authorship of this article.

Authors' Contributions

Conceptualization: MH, GT, MP, HT, VMN
Data curation: MH, VMN
Formal analysis: MH, VMN
Methodology: MH, GT, MP, HT
Writing—original draft: VMN
Writing—review and editing: MH, GT, MP, HT, VMN

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey about physicians' attitudes and experiences of telemedicine in primary care.
[DOCX File, 38 KB - humanfactors_v13i1e73609_app1.docx]

Checklist 1
CHERRIES checklist.
[DOCX File, 18 KB - humanfactors_v13i1e73609_app2.docx]

References

1. Snoswell CL, Chelberg G, De Guzman KR, et al. The clinical effectiveness of telehealth: a systematic review of meta-analyses from 2010 to 2019. J Telemed Telecare 2023 Oct;29(9):669-684. [doi: 10.1177/1357633X211022907] [Medline: 34184580]
2. Ekman B, Thulesius H, Wilkens J, Lindgren A, Cronberg O, Arvidsson E. Utilization of digital primary care in Sweden: descriptive analysis of claims data on demographics, socioeconomics, and diagnoses. Int J Med Inform 2019 Jul;127:134-140. [doi: 10.1016/j.ijmedinf.2019.04.016] [Medline: 31128825]
3. Elawady A, Khalil A, Assaf O, Toure S, Cassidy C. Telemedicine during COVID-19: a survey of health care professionals' perceptions. Monaldi Arch Chest Dis 2020 Sep 22;90(4). [doi: 10.4081/monaldi.2020.1528] [Medline: 32959627]
4. Hägglund M, Ekwall AK, Davoody N, Farrokhnia N. Escape to the future - a qualitative study of physicians' views on the work environment, education, and support in a digital context. BMC Med Inform Decis Mak 2023 Oct 19;23(1):231. [doi: 10.1186/s12911-023-02337-7] [Medline: 37858147]
5. Tu K, Sarkadi Kristiansson R, Gronsbell J, et al. Changes in primary care visits arising from the COVID-19 pandemic: an international comparative study by the International Consortium of Primary Care Big Data Researchers (INTRePID). BMJ Open 2022 May 9;12(5):e059130. [doi: 10.1136/bmjopen-2021-059130] [Medline: 35534063]
6. Milos Nymberg V, Ellegård LM, Kjellsson G, et al. Trends in remote health care consumption in Sweden: comparison before and during the first wave of the COVID-19 pandemic. JMIR Hum Factors 2022 Feb 2;9(1):e33034. [doi: 10.2196/33034] [Medline: 34846304]
7. Carter M, Fletcher E, Sansom A, Warren FC, Campbell JL. Feasibility, acceptability and effectiveness of an online alternative to face-to-face consultation in general practice: a mixed-methods study of WebGP in six Devon practices. BMJ Open 2018 Feb 15;8(2):e018688. [doi: 10.1136/bmjopen-2017-018688] [Medline: 29449293]

XSL•FO
RenderX

8. Anthony B. Use of telemedicine and virtual care for remote treatment in response to COVID-19 pandemic. J Med Syst 2020 Jun 15;44(7):132. [doi: 10.1007/s10916-020-01596-5] [Medline: 32542571]
9. NHS data model and dictionary. National Health Service. URL: https://www.datadictionary.nhs.uk [accessed 2026-03-05]
10. Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: systematic review. J Med Internet Res 2019 May 21;21(5):e13117. [doi: 10.2196/13117] [Medline: 31115340]
11. Pikkemaat M, Thulesius H, Milos Nymberg V. Swedish primary care physicians' intentions to use telemedicine: a survey using a new questionnaire - Physician Attitudes and Intentions to Use Telemedicine (PAIT). Int J Gen Med 2021 Jul 15;14:3445-3455. [doi: 10.2147/IJGM.S319497] [Medline: 34295177]
12. Brant H, Atherton H, Ziebland S, McKinstry B, Campbell JL, Salisbury C. Using alternatives to face-to-face consultations: a survey of prevalence and attitudes in general practice. Br J Gen Pract 2016 Jul;66(648):e460-e466. [doi: 10.3399/bjgp16X685597] [Medline: 27215571]
13. Glock H, Milos Nymberg V, Borgström Bolmsjö B, et al. Attitudes, barriers, and concerns regarding telemedicine among Swedish primary care physicians: a qualitative study. Int J Gen Med 2021;14:9237-9246. [doi: 10.2147/IJGM.S334782] [Medline: 34880663]
14. Olofsson S, Karlsson F, Pikkemaat M, et al. Physicians' intentions to use digital tools - a comparative survey, before and after the COVID-19 pandemic, in Southern Sweden. Scand J Prim Health Care 2024 Dec;42(4):497-514. [doi: 10.1080/02813432.2024.2346133] [Medline: 38696226]
15. Maack K, Karlsson F, Gillberg N, Wikström E, Pikkemaat M, Milos Nymberg V. Behavioural intentions towards use of digital video consultations in primary care: a survey study on physicians', nurses' and psychologists' perceptions in Swedish primary care. BMJ Open 2024 Nov 7;14(11):e082724. [doi: 10.1136/bmjopen-2023-082724] [Medline: 39515864]
16. Tipre M, Scarinci IC, Pandya VN, et al. Attitudes toward telemedicine among urban and rural residents. J Telemed Telecare 2024 May;30(4):722-730. [doi: 10.1177/1357633X221094215] [Medline: 35578537]
17. Salisbury C, Murphy M, Duncan P. The impact of digital-first consultations on workload in general practice: modeling study. J Med Internet Res 2020 Jun 16;22(6):e18203. [doi: 10.2196/18203] [Medline: 32543441]
18. Rouidi M, Elouadi A, Hamdoune A. Acceptance and use of telemedicine technology by health professionals: development of a conceptual model. Digit Health 2022 Feb;8:20552076221081693. [doi: 10.1177/20552076221081693] [Medline: 35223077]
19. Frennert S, Erlingsdóttir G, Muhic M, Rydenfält C, Milos Nymberg V, Ekman B. "It increases my ability to influence my ways of working": a qualitative study on digitally mediated patient management in primary healthcare. Scand J Caring Sci 2023 Mar;37(1):88-105. [doi: 10.1111/scs.13099] [Medline: 35833314]
20. Rahimi B, Nadri H, Lotfnezhad Afshar H, Timpka T. A systematic review of the technology acceptance model in health informatics. Appl Clin Inform 2018 Jul;9(3):604-634. [doi: 10.1055/s-0038-1668091] [Medline: 30112741]
21. Venkatesh V, Thong JY, Xu X. Unified theory of acceptance and use of technology: a synthesis and the road ahead. J Assoc Inf Syst 2016;17(5):328-376. [doi: 10.17705/1jais.00428]
22. May C, Harrison R, Finch T, et al. Understanding the normalization of telemedicine services through qualitative evaluation. J Am Med Inform Assoc 2003;10(6):596-604. [doi: 10.1197/jamia.M1145] [Medline: 12925553]
23. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
24. Francis J, Eccles MP, Johnston M, et al. Constructing Questionnaires Based on the Theory of Planned Behaviour: A Manual for Health Services Researchers: University of Newcastle; 2004.
25. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34. [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
26. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629. [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
27. Van Poel E, Vanden Bussche P, Collins C, et al. Patient safety in general practice during COVID-19: a descriptive analysis in 38 countries (PRICOV-19). Fam Pract 2025 Feb 7;42(2):cmae059. [doi: 10.1093/fampra/cmae059] [Medline: 39475357]

Abbreviations

AI: artificial intelligence
BIT: Behavioural Intentions towards Telemedicine
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
COSMIN: Consensus-based Standards for the Selection of Health Measurement Instruments
GP: general practitioner
PAIT: Physicians' Attitudes and Intentions to use Telemedicine
TPB: theory of planned behavior

XSL•FO
RenderX

Edited by A Kushniruk; submitted 07.Mar.2025; peer-reviewed by O Allela; revised version received 13.Feb.2026; accepted 19.Feb.2026; published 20.Mar.2026.

Please cite as:
Harris M, Taylor G, Pikkemaat M, Thulesius H, Milos Nymberg V
Exploring English and Swedish General Practitioners' Behavioral Intentions to Use Telemedicine: Comparative Study
JMIR Hum Factors 2026;13:e73609
URL: https://humanfactors.jmir.org/2026/1/e73609
doi:10.2196/73609

© Michael Harris, Gordon Taylor, Miriam Pikkemaat, Hans Thulesius, Veronica Milos Nymberg. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 20.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Comparison of Mask-Wearing Behavior on Social Media and Its Relationship With Demographic Characteristics During the COVID-19 Pandemic: Social Media Data Analysis Between the United States and Japan

Kiki Ferawati[1,2], PhD; Wan Jou She[1,3], PhD; Shoko Wakamiya[1], PhD; Eiji Aramaki[1], PhD

[1]Graduate School of Science and Technology, Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma, Japan

[2]Faculty of Mathematics and Natural Sciences, Sebelas Maret University, Surakarta, Indonesia

[3]Faculty of Information and Human Sciences, Kyoto Institute of Technology, Kyoto, Japan

Corresponding Author:
Eiji Aramaki, PhD
Graduate School of Science and Technology, Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma, Japan

Abstract

Background: Social media is one of the most accessible and extensive sources of data for tracking and understanding public reactions to COVID-19 policies. Cultural differences between the United States and Japan have resulted in highly distinctive policies and public reactions in each country.

Objective: This study aims to analyze the public opinions surrounding COVID-19 mask mandate through 1,102,876 and 560,873 geo-tagged tweets from the United States and Japan during the period from 2020 to 2022. We conducted 3 stages of analysis—relevance to COVID-19 masks, stance for or against masking, and whether the tweets indicate users wearing masks—to understand individuals' stance towards the mask mandate and their actual mask-wearing behavior.

Methods: We adopted a semisupervised approach to enhance BERT (Bidirectional Encoder Representations from Transformers) classification results due to data imbalance, which were then visualized through time series and map representations.

Results: In the United States, our data showed that individuals with a bachelor's degree or higher, as well as those living in states with higher household incomes, are positively correlated with positive attitudes toward mask-wearing. In contrast, in Japan, those with higher education levels or individuals aged 65 years and older were positively correlated with tweets categorized as having a stance against the mask mandate. Key events in Japan, such as the announcement of the state of emergency and the Olympics, served as major triggers for the number boost in public opinion.

Conclusions: Our analysis of over 1.6 million tweets from the United States and Japan revealed that public opinion shifted notably in response to major events and policy changes during the COVID-19 pandemic. While some trends align with previous research, correlations with education, age, and income suggest that social media data may reflect underlying societal divisions and algorithm-driven biases.

(JMIR Hum Factors 2026;13:e78236) doi:10.2196/78236

KEYWORDS

COVID-19; mask-wearing; cultural differences; classification; map; correlation

Introduction

Countries around the world have adopted a diverse strategy for managing the COVID-19 pandemic, as summarized in the government responses index in 2020 when COVID-19 first impacted the globe [1]. The implementation of different types of policies by countries is influenced by various factors, including demographics and geographical considerations. Among the measures taken to curb the spread of the disease, the mask mandate is recognized as one of the most widely adopted policies in several countries. Early in the pandemic, the Centers for Disease Control and Prevention (CDC) advised people in the United States to wear masks as a preventive measure [2], which was also adopted in a few other countries [3]. While the Japanese government did not mandate mask-wearing, its population had already been accustomed to wearing masks even before the COVID-19 outbreak, during the pandemic, and after [4].

The comparison between the United States and Japan is a frequently studied topic among researchers due to cultural contrasts and variations in Western and Eastern societal conventions [5,6]. Cultural views also contributed to how people in the 2 countries reacted differently as a response to health

XSL•FO
RenderX

policies during the pandemic [7]. During the COVID-19 outbreak, both countries displayed contrasting attitudes regarding policy creation to control the spread of the disease and adhere to mask mandates. Undoubtedly, the vast number of social media posts is often viewed as a valuable resource for comprehending, analyzing, and even informing policymakers about societal perceptions or attitudes toward an occurrence. However, a significant obstacle for cross-cultural comparison studies is handling data from multilingual social media data. From the process of data labeling, classification, and processing, the differences between the characteristics of the 2 languages might provide interesting results to be analyzed.

Social media serves as a cost-effective and accessible alternative for assessing public responses in addition to surveys [8]. Changes in people's behavior during pandemics can be observed on social media platforms like X (formerly Twitter), as demonstrated during the 2009 H1N1 pandemic [8,9] and the COVID-19 pandemic [10]. X reflects real-time insights into individuals' reactions [11,12], making it a convenient yet highly sensitive tool for assessing the risk associated with events or interventions that may have downstream effects on policy-making and implementation [13,14]. Social media might also show declining sentiment on topics related to COVID-19 restrictions throughout the pandemic, as described in the results of a study using tweets in the 3 cities in the United States [15]. Other than previously mentioned studies, the content from social media posts, such as tweets, can also be used to assess possible outbreaks by using tools such as large language models [16].

The debate surrounding masking has generated one of the most widely discussed topics throughout the pandemic, which could be clearly traced and documented on X posts [17]. Previous studies using location information in X have been conducted, such as to predict the prevalence of ZIKV cases in the United States by filtering tweets [18]. The studies used a bounding box as the boundary of the observed area, excluding tweets from Canada and Mexico. Spatial mapping shows similar patterns observed across the United States. Another study also uses geotagged tweet data, covering discussions about the #MeToo movement in the United States [19]. The results show further discourse about the movement, including the descriptions of abuse often related to the case. These existing studies suggested that geotagged tweets might provide information about ongoing social media discussions in a certain area.

This study aims to compare the discussion about masking in COVID-19 between the United States and Japan on social media and its changes throughout the pandemic stages. In the COVID-19 period, the 2 countries have different approaches to the response to the pandemic, resulting in different policies and rules. This difference was also reflected in how people responded, which can also be seen on social media. Mask-wearing, as one of the policies during COVID-19, generates a lot of discussion among the public. Some are supporting, and some are against mask-wearing. Despite their small number, people who did not want to wear masks are a vocal minority in sharing their opinion [20]. Due to all those differing opinions, while mask-wearing is indeed an important effort to mitigate the spread of COVID-19, there are still many debates about the necessity of masks [14,21]. A geographical focus in the analysis was also interesting to explore, as a survey showed that mask-wearing as a public health behavior might exhibit spatial heterogeneity, suggesting a different approach in the area based on country-level analysis in the United States [22].

In this study, we discussed the public opinions surrounding the COVID-19 mask mandate through geo-tagged tweets from the United States and Japan during the period from 2020 to 2022 in 3 stages of analysis: starting from its relevance to COVID-19 masks, stance for or against masking, and whether the tweets indicate users wearing masks. By discussing the factors and correlation of the observed tweets, we aim to understand individuals' stance towards the mask mandate and their actual mask-wearing behavior.

Methods

Ethical Considerations

This study did not require participants to be involved in any physical or mental intervention. This research also did not use personally identifiable information, thus exempted from institutional review board approval in accordance with the Ethical Guidelines for Medical and Health Research Involving Human Subjects stipulated by the Japanese national government.

Data

The data in this study consists of geo-tagged tweets collected between January 1, 2020, and December 31, 2022, using Academic Research Access (discontinued in mid-2023). This dataset comprised 1,102,876 English tweets and 560,873 Japanese tweets. Note that we used the same dataset as our previous study on multilanguage data annotation [23]. Examples of paraphrased tweets showing their opinion about mask-wearing are shown in Textbox 1.

XSL•FO
RenderX

Textbox 1. Example of tweets in English and Japanese.

English:

- Against: Wearing this mask at work is exhausting. I'm sweating like crazy even from the smallest tasks.
- Not against: Unexpected perk of wearing a mask: it keeps your face nice and warm in freezing weather! \#MaskUp \#StayWarm \#WinterHack

Japanese:

- Against: 普段あまりこういうことは言わないようにしてるけど, どうしてまだみんな外でマスクをしてるの？室内でも特に必要とは思えないし, もう外してもいいんじゃないかな, 本当に。 (I usually try not to say things like this, but why is everyone still wearing masks outside? I don't really see the need for them indoors either. Isn't it okay to stop wearing them already? Honestly.)
- Not against: それに, マスクをしていない人の割合もかなり高め。ちょっと気が緩みすぎてない？ (Besides, the number of people not wearing masks is pretty high. Aren't people getting a little too careless?)

Demographic data were accessed from the Portal Site of Official Statistics of Japan [24] and Census Bureau data of the United States [25], obtained from the 2020 census. Previous studies have indicated that demographic characteristics such as sex, ethnicity, age, marital status, and employment status affect participation in preventive behaviors in the United States [26]. In this study, we focused on several demographics for population characteristics at the state/prefecture level: total population, population by age, education level, information on foreign members/race, and income (Table 1). For Japan, the education level was obtained from the number of employed persons aged 15 years and older.

Table . Demographics variable used in correlation calculation.

Demographics	Variable
Population[a]	• Total population • Population of working age[b] • Population of elderly (older than 65 years)
Education level[c]	• Population completed bachelor or higher
Income	• Household income
Race information	• Race index (the United States), Number of foreign residents (Japan)

[a]Total population from 2020 census.

[b]Percentage of people aged 18 - 64 years (the United States) and 15 - 64 years (Japan).

[c]Percentage of people completed bachelor or higher from population aged older than 25 years (the United States) from American Community Survey (2021) and population of employed person aged older than 15 years (Japan) from 2020 census.

Preprocessing of Tweets

The first preprocessing step involved validating each tweet's location tag. Given the various types of location information in geo-tagged tweets, we focused on city-level information found in the "full-name" entity of the tweets. Instances where the geo-tagged location did not match the city lists in the United States and Japan were removed [27,28]. Next, we applied basic preprocessing steps, including changing usernames to common handles, addressing links, converting emojis to text, and removing duplicates and NAs.

Finally, we ensured that the tweets contained the keyword "mask" for English tweets and "マスク" (mask in Japanese) for Japanese tweets. In some cases, these keywords appeared as usernames or links, providing insufficient information or relevance to the topics and were thus eliminated in the process.

Data Annotation

The annotation for our 1100 pairs of tweets was done by 4 annotators, 2 for each English and Japanese. We used Cohen kappa as the measure to evaluate interannotator agreement between annotators [29] and interpret the agreement results [30]. The complete detailed approach and evaluation are explained in Ferawati et al [23]. To combine the annotated label between the annotators, we assigned the agreed label as the final label and assigned the disagreed label as not relevant, not against, and unknown for each stage.

Classifier Training

X studies often use human annotators with the aid of machine learning classifiers [31,32]. We designed a multilingual guideline to obtain a suitable data annotation for this study [23]. The results of the annotation show a case of imbalanced data, with one category having a much smaller number of tweets compared with another. However, due to the time and resources required for human annotation, it is not practical to annotate a large number of tweets. To address this issue, we propose a semisupervised approach by using the BERT (Bidirectional Encoder Representations from Transformers) score from BERT-based classification method as additional data for training. By including additional training sets from predicted tweets, we hope to improve the performance of the classifier to

XSL•FO
RenderX

solve the cases, especially in public perception and attitudes for COVID-19 masking in the 2 countries.

An existing study explained that BERT surpassed traditional machine learning algorithms, such as Logistic Regression, Support Vector Machine, Random Forest, and Naive Bayes, where BERT was supported by its ability to understand contextual information appearing in social media [33]. BERT was also widely used to analyze and derive conclusions. Based on the results of these studies, we decided to use BERT-based models for our study. We trained the annotated datasets using BERT-based classifiers, RoBERTa base model [34] for English tweets and BERT base Japanese [35] for Japanese tweets. To evaluate the models, we set aside 15% of the annotated data as the test set and applied 5-fold cross-validation to the remaining data. The final model was selected from the fold with the best results on the test set, evaluated using F_1-macro to address data imbalance.

Due to the limited number of annotated data, we included all 1100 annotated tweets. For each language, tweets annotated with the same categories by both annotators were assigned to those categories. In cases of conflicting annotations—for example, where annotator 1 marked a tweet as not relevant while annotator 2 marked it as relevant—we categorized the tweet as not relevant. The same approach was applied for the stance stage (categorized as unclear, then assigned to not against) and mask-wearing (categorized as unknown).

Semisupervised Approach

We observed an imbalance in the categories of our annotated data, which likely contributed to the low performance of our models. Due to constraints in obtaining additional annotated data, we implemented a semisupervised approach for the stance and mask-wearing stages of the tweets. The semisupervised approach was a common measure taken to obtain more annotated data without a significant increase in annotation costs [36]. In this study, we stored the BERT outputs from the predictions and filtered them based on determined thresholds for each category. We considered 2 methods of adding pseudo-labeled data: adding an equal number of tweets to both categories and adding tweets only to the minority category. Using softmax-transformed values of the BERT output for each class, we selected a set of data to be added to the training. The objective for implementing the approach is to obtain labels with a higher confidence from the obtained softmax-transformer values as an additional sample for training. We maintained the chosen split of training and validation from the previous step in classifier training.

The selected tweets from this step were then included as additional training data using the same parameters and model as in the previous subsection. The data for validation and testing remained the same as the baseline. We experimented with several combinations of thresholds for each class in each stage (against and not for stance; no, yes, and unknown for mask-wearing), while endeavoring to keep the number of selected additional tweets balanced. When determining the threshold value for the selection, we selected a certain value to balance our objectives: retaining sufficient confident predictions for the reliability of our pseudo-label and having enough additional samples to meaningfully expand our dataset. The threshold was set by observing the transformed BERTScore from the prediction results, plotting them into a histogram with the aid of descriptive statistics of the score to determine the appropriate threshold for filtering. We then compared the results of each training.

Correlation

We calculated the correlation between the predicted percentages and demographic data from the 2 countries, the United States and Japan, using Pearson correlation. In this study, we considered 3 levels of significance for the P values with a significance level of .05, .01, and .001.

Results

Interannotator Agreement

For the annotation phase, we used only a subset of the entire dataset. Considering the reasonable workload of our annotators, we selected 1100 tweets from each language, as reported in another publication focusing on the annotation of a multicultural dataset [23]. Although the agreement levels for stance and mask-wearing stage are still less than <0.60—a score that was categorized as moderate by Cohen (1960) [29] but inadequate by McHugh (2012) [30]—they are deemed acceptable for our study (Table 2).

Table . Cohen kappa for the annotation results. The percentage in each round shows the number of samples annotated. The agreement is calculated for each round and each part of the samples.

Stage	Language	
	English	Japanese
1: Relevancy	0.79	0.92
2a: Stance	0.46	0.59
2b: Mask-wearing	0.55	0.47

The number of tweets varies substantially across the country, with certain states or prefectures having far more tweets than others. The difference in number is primarily observed in densely populated areas, such as the capital or popular destinations. Among tweets identified as relevant to COVID-19, California leads the United States in mask-related discussions, while Tokyo tops the list in Japan. A list showing the top and bottom tweet counts for states (in the United States) and prefectures (in Japan) is provided (Table 3).

XSL•FO
RenderX

Table . Top and bottom 3 of tweet counts in the United States and Japan.

Ranking	United States		Japan	
	State	Tweet count	Prefecture	Tweet count
Top 3				
	California	113,148	Tokyo	85,836
	Texas	75,517	Kanagawa	27,737
	New York	60,317	Osaka	26,725
Bottom 3				
	Vermont	1314	Kochi	1197
	North Dakota	1079	Tottori	1149
	Wyoming	776	Shimane	1049

Classifier Training

We fine-tuned roberta-base and bert-base-japanese-v3 with our annotated tweets to predict the rest of the data, a method commonly used for social media data [37,38]. We trained the classifier using this annotated data and then used it to predict the remaining tweets. The parameters for our final model used the AdamW optimizer, with a batch size of 32, 10 epochs with early stopping callback, learning rate of 0.00002, and weight decay of 0.01.

We used the entire sample of 1100 tweets for relevancy classification (Table 4). In the first stage of relevancy classification, we achieved the highest F_1-macro scores for both English and Japanese texts compared with the majority baseline. However, the stance and mask-wearing stage did not yield similar results, showing only a marginal increase from the majority baseline. This indicates that the models are still unable to correctly predict the minority categories. These results also mirror the annotation results, where annotators struggled to achieve higher interannotator agreement.

Table . Results of BERT (Bidirectional Encoder Representations from Transformers)-based classifier. Only tweets annotated as relevant by both annotators are included as a sample for Stage 2a: Stance and Stage 2b: Mask-wearing. As the category is heavily imbalanced for all the stages, we calculated the majority baseline by assigning majority classes as the predicted category in each stage: relevant for relevancy, not against for stance, and unknown for mask-wearing.

Stage	Language	Number of samples, n	Majority baseline	F_1-macro
1: Relevancy				
	English	1100	0.44	0.78
	Japanese	1100	0.39	0.75
2a: Stance				
	English	880	0.47	0.61
	Japanese	706	0.45	0.69
2b: Mask-wearing				
	English	880	0.31	0.31
	Japanese	706	0.29	0.40

Semisupervised Approach and the Final Classifier

By adopting a semisupervised approach to train our classifier, we were able to improve second-stage F_1-scores that were nearly as high as those from the first stage—0.68 for English tweets and 0.72 for Japanese tweets.

Based on the results for the stance stage, the chosen threshold with a more balanced number of tweets in each category (eg, an additional 650 in the "against" category, bringing the total to 755 "against" and 775 "not against" for English data) achieved a better F_1-macro score than other alternatives.

When it comes to classifying mask-wearing behavior, maintaining a balanced sample for re-training is challenging due to the imbalance of labels in the annotated data. In English tweets, only 46 tweets were predicted as "not wearing mask" out of the entire dataset, so we included all the predicted tweets in the model. The best result was an F_1-score of 0.487, a substantial increase from the previous score of 0.301 with the original annotated data. For Japanese tweets, our best model only yielded an F_1-score of 0.48.

For the mask-wearing stage, the distribution of BERT score outputs was much more skewed than for stance in both languages. Because it was challenging to maintain balanced numbers for each category, we experimented with various threshold combinations to assess performance. Due to the difficulty in balancing the numbers, the improvement in the

XSL•FO
RenderX

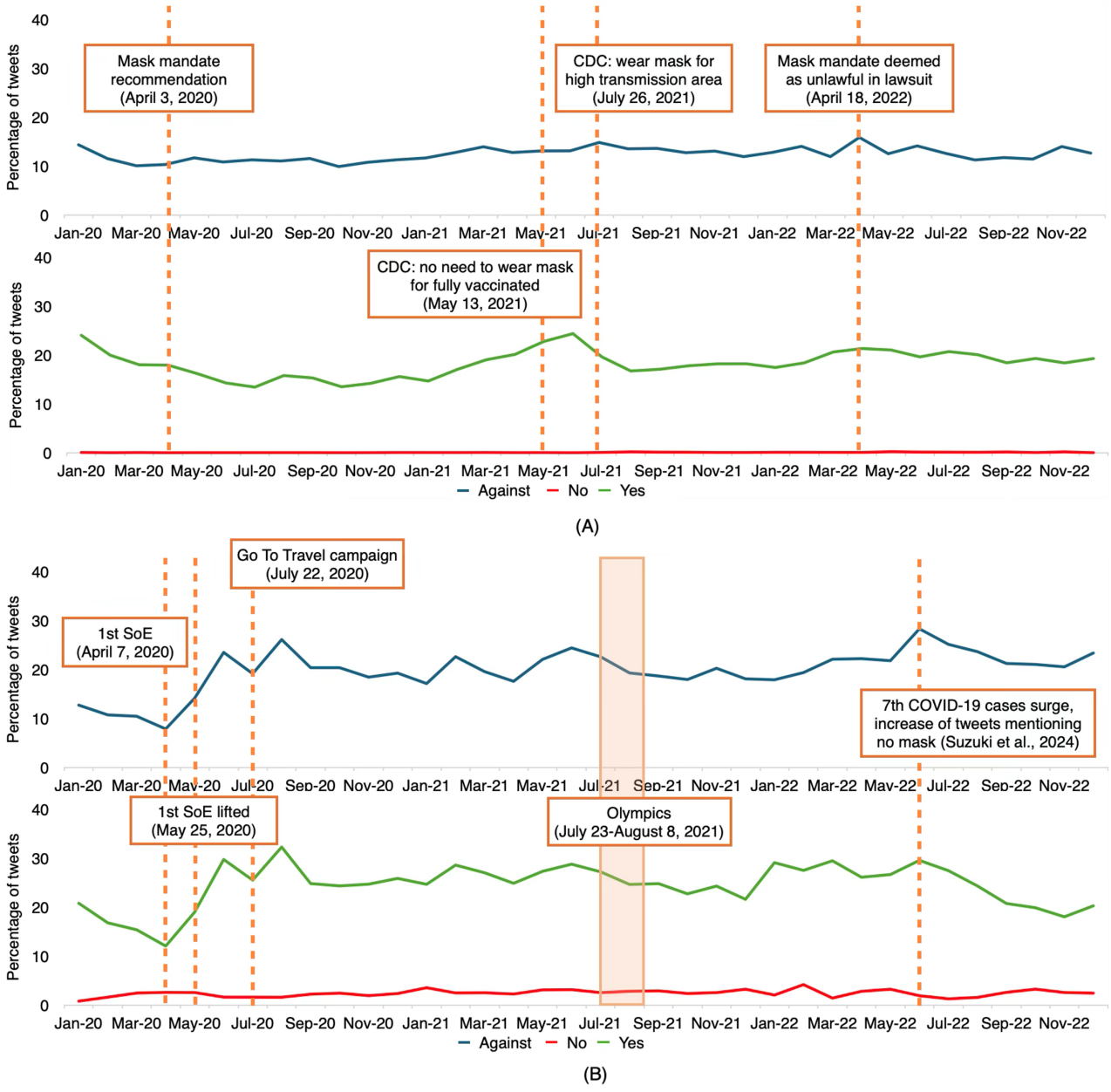
model performance seems rather limited. Therefore, it should be considered the best our classifiers could achieve and is acceptable to be used in our following tasks.

The final classifier model was trained on a combined dataset of the original annotated data and the semisupervised data. The remaining tweets were classified using the final model. We deemed the F_1 -score from the previous semisupervised results acceptable for our study and used it as a basis for generating a map of stance toward the mask mandate and actual mask-wearing behavior from related tweets in the United States and Japan, using both English and Japanese tweets.

Time Series Observation of the Classification Results

The tweets were collected over a 3-year period, capturing the trend from early, peak, to post-peak periods of the COVID-19 pandemic. Throughout this time, various events and policies occurred, potentially influencing tweet volume and content (Figure 1). In the figure, no observable trend in the stance against COVID-19 masking is evident in the United States, suggesting little change in people's opinion about mask-wearing from early 2020 to the end of 2022. Both countries show a decreasing trend in tweets against masking from January 2020 to April, followed by an upward trend after relevant departments released recommendations about mask-wearing.

Figure 1. Time series plot of tweets against masking in (**A**) the United States and (**B**) Japan. Some events and policies happening throughout the period are highlighted in the figure. The upper part of the graph shows the percentage of tweets against mask-wearing, while the lower part shows tweets mentioning users wearing a mask (green) and not wearing a mask (red) [39]. CDC: Centers for Disease Control and Prevention; SoE: state of emergency.



Throughout the observed period, the CDC released several guidelines on the utilization of masks as a preventive measure. They initially mandated the use of masks in April 2020 and later revised the recommendation after the development of vaccines, such as lifting the recommendation to wear masks for fully vaccinated individuals in May 2021. In the month

XSL•FO
RenderX

following CDC's revision of their mask-wearing guidance to no need to wear a mask if vaccinated, tweets classified as "not wearing mask" reached their lowest point, while tweets classified as "wearing mask" peaked (Figure 1). Subsequently, the CDC released another recommendation to wear masks in high-transmission areas in July 2021, which is reflected in interesting changes observed in the time series plot of mask-wearing stages. The number of tweets indicating "not wearing mask" substantially increased from 0.03% in June 2021 to 0.21% in August 2021, while tweets about "wearing mask" saw a decrease from 24.41% in June 2021 to 16.76% in August 2021. In April 2022, as a result of a lawsuit regarding mask mandates deemed unlawful in Florida, nationwide mask mandates for airplanes and public transportation were no longer enforced [40]. As shown in Figure 1A, in the same month, a peak was observed in the stance against mask mandates in the United States, and tweets classified as "not wearing mask" peaked a month after the ruling, observed in May 2022.

In Japan, there are 3 noticeable fluctuations in the figure, closely related to COVID-19 policies. We observed some significant changes when the government announced a state of emergency (SoE) in April 2020 and subsequently lifted it in May 2020. After a dip in the percentage of tweets expressing a stance against masking after the first SoE in April 2020, the first peak appeared a month after the first SoE was lifted. These fluctuations are also consistent with findings from another study on responses to COVID-19 waves in Japan [41]. Another study discussed that there are indeed signs of disruptions due to COVID-19 in X, which shows top concerns during the SoE period and some impacts of the pandemic on societies in Japan [42]. The second peak appeared at the beginning of the Olympics. The Tokyo Olympics, held from July 23 to August 8, 2021, also influenced the observed changes in the figure. Before the Olympics began, in June 2021, there was a slight increase in tweets expressing a stance against mask-wearing, as shown in Figure 1B. The third peak appeared in June 2022, coinciding with the 7th surge in COVID-19 cases and an increase in "no mask" tweets, as observed by Suzuki et al [39].

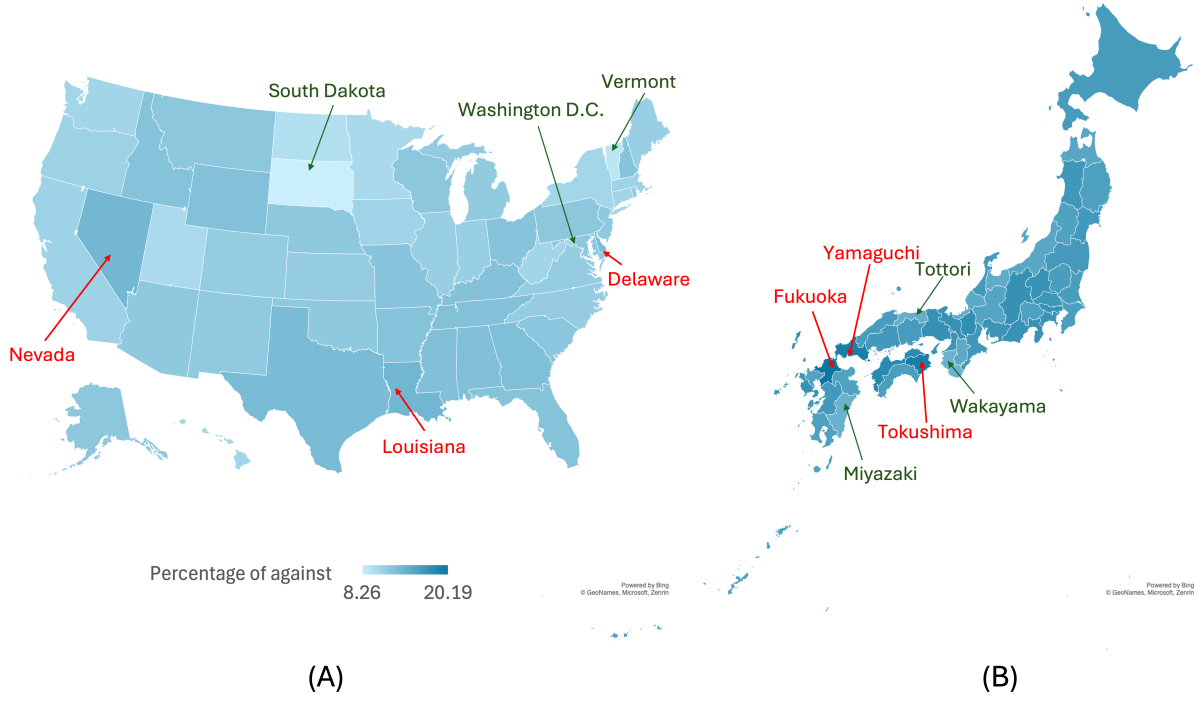
Map Comparison

Tweets Against Masking

The final classifications from the selected model were presented in the form of maps, depicting the percentage of tweets classified as against mask-wearing in the United States (at the state level) and Japan (at the prefecture level). Table 3 reveals a significant discrepancy in tweet frequencies between states and prefectures. For instance, California has over 100,000 relevant tweets, while other states like Wyoming have fewer than a thousand tweets. Similarly, in Japan, Tokyo, the capital, has the highest number of relevant tweets, with over 85,000, while Tottori has the fewest tweets, with approximately 1000.

In the United States, Louisiana, Nevada, and Delaware were the 3 states with the highest percentage of tweets against masking, ranging from 13.28% to 13.75% (Figure 2). The 3 states with the lowest percentage are Washington, DC; Vermont; and South Dakota, with about 8% - 9%. On the other hand, in Japan, the percentage of tweets against masking is observed to be higher, ranging from 19.28% to 20.19% in Fukuoka, Yamaguchi, and Tokushima prefectures, with the lowest percentage observed at about 14% in Tottori, Miyazaki, and Wakayama prefectures.

Figure 2. Percentage of tweets against masking in (**A**) the United States and (**B**) Japan. The same color scale is used for both the United States and Japan. The top 3 prefectures/states with the highest percentage are marked in red text and arrow, while the bottom 3 are marked in green text and arrow.



XSL•FO
RenderX

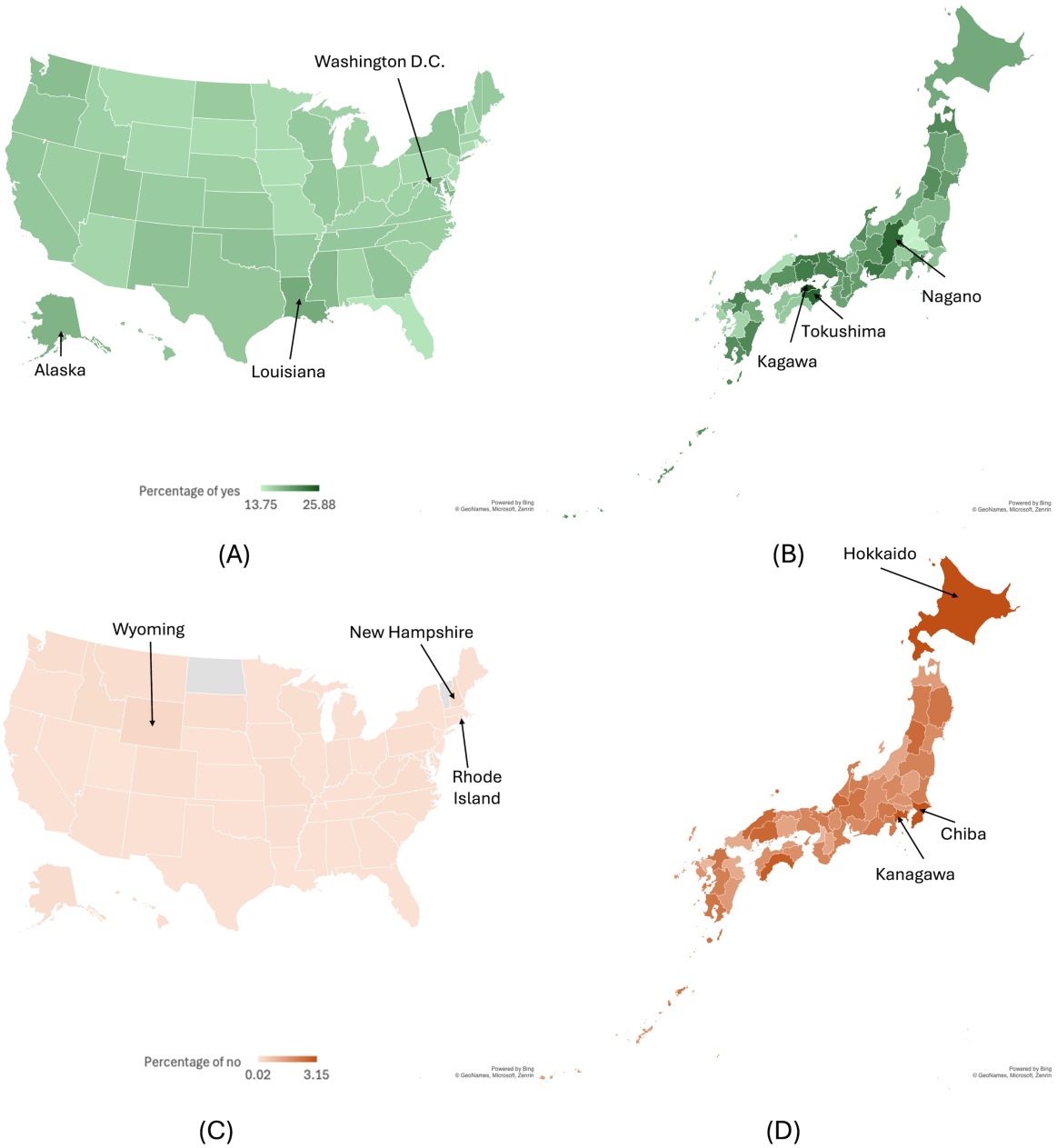
Mask-Wearing Classification Results

The analysis and maps for mask-wearing behavior tweets were divided into 2 categories: those classified as “not wearing” and “wearing” masks (Figure 3). Since some states or prefectures have a much smaller number of tweets, the count of tweets classified as “not wearing” is reduced accordingly. For English tweets, the total count of “not wearing” tweets was very small, at 574, resulting in no tweets classified in the category in several states and causing the percentage to appear as 0. Wyoming has the highest percentage of tweets classified as “not wearing,” at 0.26% (2 out of 776). In contrast, Louisiana leads in the percentage of “wearing” mask tweets, at 19.12% (2122 out of 11,100); followed by Washington, DC (1638 out of 8845, 18.52%); and Alaska with about 18.33% (333 out of 1817). However, a careful interpretation is necessary for these figures, as the number of tweets was relatively small, especially for the tweets mentioning not wearing masks. The tweets were much smaller in the behavioral aspects of mask-wearing, so the resulting maps might not have enough data for a state-level conclusion.

On the Japan side, there were 8462 tweets classified as “not wearing” masks, a higher number compared with the United States results. However, a big portion of these tweets was still concentrated in Tokyo, with over 2000 tweets, while the other 31 prefectures had fewer than 100 tweets each. In terms of the percentage of tweets classified as “not wearing” masks, Hokkaido has the highest percentage at 3.15%. For tweets classified as “wearing” masks, Kagawa has the highest percentage at 25.87%.

XSL•FO
RenderX

Figure 3. Percentage of tweets classified as wearing a mask in (**A**) the United States and (**B**) Japan, and tweets classified as not wearing a mask in (**C**) the United States and (**D**) Japan. The state/prefecture name shown denotes the top 3 areas with the highest percentage. The same color scale is used for both the United States and Japan. Gray shows no predicted data for the area.



Correlation With Demographics Data

We calculated the correlation between stance percentages and the demographic data from the country census for both the United States and Japan (Table 5). In the United States, we observed a negative correlation of −0.384 between household income and stance against masking percentages. Although this correlation is relatively low, it is still significant within the 1% level of significance threshold. This suggests that states with higher incomes tend to have lower percentages of tweets against masking. Additionally, education level also shows a significant negative correlation, showing that with an increase in the population completing higher studies, there was an observed decrease in tweets with a negative stance towards the mask mandate. In Japan, we found that the number of people older than 65 years showed a positive correlation with the percentage of tweets against the mask mandate at the prefecture level. Similar results were observed in the number of people who completed a bachelor or higher degree.

XSL•FO
RenderX

Table . Correlation calculation of tweets percentage with the demographics data from the United States and Japan. The table only displayed a significant correlation.

Variable	Correlation	95% CI	*P* values
The United States			
Population completed bachelor and higher	−0.456	−0.653 to −0.200	.001
Household income	−0.353	−0.577 to −0.080	.013
Japan			
Population of elderly (older than 65 years)	0.295	0.009 to 0.537	.044
Population completed bachelor and higher	0.310	0.024 to 0.548	.034

Discussion

Principal Results

Key Events That Drove the Surge in COVID-19–Related Tweets

The time series plot illustrates the changes in stance against masking and mask-wearing tweets in the United States and Japan. The major fluctuations in the data points often coincide with governments announcing mask-related policies or major public events such as lockdown or the Olympics during the COVID-19 period. Such a phenomenon suggested that individuals tend to be prompted to express themselves on social media by major movements from government or public events, as previous studies also indicated [43-46]. In addition to the temporal trend, the geographical visualizations from the final results provide a general landscape of public responses to the mask-related policies divided by the municipal regions on social media (referred to as “X” in this study) in the United States and Japan. Information regarding the percentage of tweets against masking is displayed for each state or prefecture (Figure 3). Interestingly, Washington, DC, emerges as the state with the lowest proportion of tweets against masking and the highest percentage of tweets mentioning mask-wearing. This suggests that the state exhibits less resistance to mask-wearing and actively follows mask guidance.

Based on the correlation results between tweets and demographic data, education level and household income were found to be significantly correlated in the United States. Specifically, the population of people who have completed a bachelor’s degree and higher is negatively correlated with a stance against masking. This result indicates that with the increase in the higher education levels of the population in the states, we observed a decrease in the number of tweets against masking. This observation aligns with findings from a previous study based on a survey in the United States, which reported that individuals with at least a college degree are less likely to refrain from wearing face masks [47]. Additionally, the same study found that households with lower incomes were more likely to exhibit the same behavior of not wearing masks, an insight consistent with our findings.

Previous studies indicate that the behavior of older and younger people changes throughout the pandemic, with older individuals becoming more cautious about preventive measures as the pandemic progresses [48]. Although the majority of X users fall within the young adult category (aged 18 - 29 years), there are still a considerable number of older X users in the United States, as revealed by a survey conducted in 2023 [49]. Since this study calculated correlations for the entire period of data collection, these changes in opinion might not have been captured well. Throughout the study period, there have been several changes in mask-wearing rules, compliance, and when masks are mandated in some areas, which also affects how people responded and their opinion towards mask-wearing [50]. New information about mask-wearing released by the authority can also result in changes in public attitude toward mask-wearing [51], such as when people initially support mask-wearing and then change their opinion, or if they are against masking in the beginning and change to accepting masks in the later stage of the pandemic. This was also supported by results of the study, where a change in the sentiment is observed in the 3 major cities in the United States [15].

In Japan, the number of people aged older than 65 years shows a positive correlation with the number of tweets expressing their stance against masking, in line with a previous study in Hiroshima, Japan [52]. Based on existing studies summarizing responses in 27 countries, including Japan, people over 60 did not really comply with wearing masks outside of their home [53], which is also reflected in the results of this study [54]. The overall results of the correlation calculated are, however, quite small. Such a result indicated that there could be more confounding factors influencing the results. Below, we discuss some potential factors suggested in previous literature.

Health policy differences in the states within the United States might be one of the possible factors affecting the mask-wearing behavior reflected in social media. There had been observable differences, such as physical distancing rules across states and the relation of the measure to racial and socioeconomic differences in the area [55]. There was also an existing study documenting mask-wearing policy across the states and its enforcement [56]. The result indicates that overall, the mask is not strongly enforced in the states. The resistance to mask-wearing in the United States could also be viewed through cultural aspects of people living in the country [57]. A study focusing on the cultural and political aspects of mask-wearing

XSL•FO
RenderX

suggested that future attempts at mask-wearing policy need to consider the cultural meaning of masks to the public [58].

A difference in response to policies was observed in Japan, with results from a study showing the high trust and compliance in government policies during COVID-19 [59]. An internet survey showed that some habits of the Japanese, such as wearing masks prior to the pandemic, were also related to their responses to the preventive measures for COVID-19 [60]. These policies and behavioral insights might serve as additional variables to further explain the observed web-based behavior in X. The insights can be used in handling information dissemination in social media, for example, by the government in countries to design the most effective way to communicate with the public on social media.

Interpreting Conflicting Correlations: Mobility, Social Media, and Educational Shifts

Most of our findings aligned with previous research. However, unexpectedly, individuals with a bachelor's degree or higher also showed a positive correlation with opposition to the mask mandate, which contradicts earlier studies [48]. One possible cause for the inconsistency was the difference in the demographics between people using X and the census data used for the correlation calculation. In the United States, the percentage of X users with at least a bachelor's degree is higher than that of the general public, which is also observed in the income levels [61]. In Japan, based on a survey conducted by the Ministry of Internal Affairs and Communications, about 80% of people in their 20s were X users. The penetration rate is the highest in the younger population, but there were also users in their 60s (18.9%) [62]. According to a study conducted using X data, even within the social media users' demographics, there was a bias due to the choice to participate in a certain topic for social media discussion [63]. This disparity might affect the correlation results with tweet percentages, as the demographics being calculated are not X-specific demographics but public demographics based on census and surveys.

Another possible cause of the inconsistency in results may also have been caused by the mobility of people throughout the COVID-19 period in Japan. The tweet data were collected over a 3-year period, while the education level was obtained from the 2020 census. A previous study mentioned that before the pandemic, people were centralized in the Tokyo metropolitan area, a pattern disrupted during COVID-19 [64]. The changes in migration rates were also influenced by the emergence of remote work during the pandemic, prompting younger people to move to rural areas.

The increase in tweets mentioning "no-mask," as shown in the study by Suzuki et al [39], might also contribute to the conflicting correlation results, particularly in the latter half of 2022. Another reason might be caused by changes in the education style during the early period of COVID-19 to adapt to the changing policy. The changes to web-based education proved to be overwhelming for students, with their self-reported mental health status being worsened [65]. The anxiety and depression experienced during COVID-19 also increased students' social media usage [66]. This might be reflected in opinions about COVID-19 and its measures, such as mask-wearing, where similar cases were observed in studies about disruptions caused by COVID-19 in social media [42].

Our findings indicated that mask-wearing behavior was observed in social media discourse in both the United States and Japan. The map showing the area with a high percentage of yes and no to mask-wearing can be considered a guideline in how to communicate or release guidelines in the event of a pandemic, with more attention to the area with a high percentage of "no" in mask-wearing. Based on the number of tweets, we may also launch more information on social media to reach more people in the area. The correlation results show some significant correlations observed with population data of the states and prefectures, with some of them consistent with existing findings. These results suggest that the social media data has the potential to aid in research, with some caution due to the limited nature of the geo-tagged data in social media. Focusing on other aspects of analysis, such as other potential factors in correlation analysis or additional data, can be a future research direction to improve the results.

The Health Belief Model was believed to be useful for understanding and investigating the participation in COVID-19 preventive measures in the United States [67]. A study using the Understanding America Survey showed that the Health Belief Model results were consistent with existing studies, with mask-wearing behavior positively endorsed as one of the preventive measures. A study comparing public perceptions in different countries during the pandemic, Japan and the United States included, also showed that higher perceived effectiveness was one of the factors affecting how the public responded to preventive behavior, such as mask-wearing or handwashing [68]. To make the public more accepting of preventive measures or policies, the public needs to trust the effectiveness of the act. The role of publishing and spreading the effectiveness can be entrusted to social media, with emphasis in areas with higher percentages in the stance of not wearing masks.

Limitations

This study focuses solely on the English and Japanese languages for analyzing short and informal text, specifically tweet posts. In the second stage, we encountered numerous ambiguous and unclear tweets regarding attitude. Analyzing such stages proved challenging due to the short text and ambiguity of tweet content, which was often insufficient for accurately determining attitudes. Therefore, we chose to exclude a neutral option, limiting choices to either against or not.

Although X data might be subject to sampling bias due to its limited user base, the data were proven useful for infodemiology research, especially in identifying and monitoring public discussion topics during the pandemic [69,70]. Population-wise, our sample is restricted to individuals who voluntarily provide location information in their tweets. Despite the limitation of such data, it reflects the opinion of the users to a degree, as reported in an existing study that showed that geotagged tweets were able to display public opinion of people in the United States toward COVID-19 vaccinations [71]. Disparities in tweet counts across states/prefectures may influence the final percentage results displayed in the figures, which are based on the entire 3 years of data. We calculated the percentage based

XSL•FO
RenderX

on the tweets observed in each state as one of the measures to address the potential problems due to the location data. The number of tweets in each state varies, so we avoid general bias caused by comparing the tweets based on their count, but based on the percentage of the classified tweets. Due to the data collection period and collective analysis after, the resulting figure, especially in the form of map visualization, may not sufficiently capture early COVID-19 period changes and post-pandemic trends shown in the time series plot.

A previous study identified around 8.4% of tweets as bots from around 200,000 accounts collected in the early period of the pandemic, March to June 2020 [72]. Another study mentioned that bots focused more on political content in the COVID-19 pandemic discourse compared with health content, such as actions taken by the public during the pandemic [73]. The bots mostly discussed information related to COVID-19 or pandemic statistics. In this study, we focused on mitigating bots' influence, such as artificial amplification by removing duplicate tweets, one of which often appeared in posts by bot accounts. This action ensured that no single automated message could statistically dominate our social media dataset or skew our demographic correlations. To verify this approach, we conducted a manual audit in the annotated dataset, which was sampled randomly from the entire tweets data. We did not find apparent bot accounts during the check in our sample. Given this minimal presence, we conclude that the presence of bots does not significantly skew the overall stance analysis or correlations in our study.

The limited number of "not wearing" mask annotations constrained the number of predicted instances in the rest of the tweets. Consequently, the final predicted percentages for "not wearing" masks were also very small, only 0.07% for English tweets and 2.33% for Japanese tweets. This problem, caused by a small number of tweets annotated as "not wearing mask" was also further added by a surge in tweets mentioning "no-mask" after June 2022 [39]. This surge, as documented by Suzuki et al [39], may contribute to the disparity from previous studies and introduce bias into the correlation calculations, requiring people to interpret the results cautiously. To address these challenges, future improvements in classification and additional annotations are necessary for enhanced accuracy.

The initial score of the classification results obtained was not satisfactory. The F_1-score for stance against masking is 0.61 for English and 0.69 for Japanese. It was even smaller for the mask-wearing stage, with 0.31 for English and 0.40 for Japanese. While it is higher than the majority baseline, it is still not performing as well for the classification. There is a slight increase in the F_1-score after adding pseudo-labeled data, as observed in the stance and mask-wearing stage. However, the scores were still not as high, indicating that the topic to be classified is indeed challenging, as both human annotators and classifiers struggle to achieve a high score in classifying stance against masking and mask-wearing information of the users.

Regarding the mask-wearing stage, the number of tweets annotated as "not-wearing" masks is very small in the annotation phase (10 for English and 11 for Japanese). This small number affects the classification because when we split the data, we were working with an even smaller number of tweets for training, validation, and testing. This makes it even harder to classify the "not-wearing-mask" tweets, resulting in a low F_1-score in the testing phase. This issue is further highlighted in the classification of the remaining tweets, with only 0.07% predicted as "not-wearing" masks in English and 2.33% in Japanese out of all the predicted tweets. The low number of tweets predicted as "not-wearing" masks is also apparent in Figure 3, especially in the United States case. This limitation warrants caution in interpreting our results, especially since the annotation target was particularly challenging.

Due to the policy of X, we are unable to obtain the demographics of its users; thus, we considered using the demographics of the general population in the country. However, there are differences between the demographics of X users and the general population [61,62] which might affect the results of the correlation calculated in this study.

Conclusions

Our findings revealed a clear shift in public opinions at several key moments, often coinciding with major policy changes or public events, such as mask mandates and the Olympics, observed from 1,102,876 English tweets and 560,873 Japanese tweets from the United States and Japan between early 2020 and late 2022, covering the entire duration of the COVID-19 pandemic. Based on the correlation observed in the social media data, some variables are significantly correlated with the number of tweets, such as education levels, the population of the elderly, and household income. Some of our findings align with previous research across major social media platforms, suggesting that public opinions reflected in social media data are often interconnected. However, divisions based on education level, household income, and generational differences also indicate that social media interactions may be inherently biased, partly due to recommendation algorithms that encourage clicks and foster echo chambers.

Funding

This work was supported by JST SICORP (grant number JPMJSC2107), JSPS KAKENHI (grant number JP22K12041), and the Cross-ministerial Strategic Innovation Promotion Program (SIP) on "Integrated Health Care System" (grant number JPJ012425), Japan.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to X's data sharing policy, but the IDs are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

KF collected the data. KF, WJS, and SW conducted the experiments and analyzed the results. SW and EA provided critical comments. All authors contributed to the design and execution of the study and reviewed the manuscript. EA supervised the project.

Conflicts of Interest

None declared.

References

1. Hale T, Angrist N, Kira B, Petherick A, Phillips T, Webster S. Variation in government responses to COVID-19. Blavatnik School of Government. 2020. URL: https://www.bsg.ox.ac.uk/research/covid-19-government-response-tracker [accessed 2024-01-12]
2. Netburn D. A timeline of the CDC's advice on face masks. Los Angeles Times. 2021. URL: https://www.latimes.com/science/story/2021-07-27/timeline-cdc-mask-guidance-during-covid-19-pandemic [accessed 2024-02-14]
3. Näher AF, Schulte-Althoff M, Kopka M, Balzer F, Pozo-Martin F. Effects of face mask mandates on COVID-19 transmission in 51 countries: retrospective event study. JMIR Public Health Surveill 2024 Mar 8;10:e49307. [doi: 10.2196/49307] [Medline: 38457225]
4. Miura A, Murayama A, Kitamura H. Behind the mask: analyzing the dual influence of social norms on pandemic behavior in Japan 1. Jpn Psychol Res 2024. [doi: 10.1111/jpr.12520]
5. Noguchi K. Examination of the content of individualism/collectivism scales in cultural comparisons of the USA and Japan. Asian J Social Psychol 2007 Sep;10(3):131-144. [doi: 10.1111/j.1467-839X.2007.00220.x]
6. Ishii K, Eisen C. Socioeconomic status and cultural difference. Oxf Res Encycl Psychol 2020. [doi: 10.1093/acrefore/9780190236557.013.584]
7. Uricher R, Nakayama M, Uchida Y. Cultural differences in behavioral regulation under the collective threat of COVID‐19: more adjustment in Japan and more influence in the United States. J Appl Social Pyschol 2025 Jul;55(7):483-496. [doi: 10.1111/jasp.13103]
8. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of tweets during the 2009 H1N1 outbreak. PLoS ONE 2010 Nov 29;5(11):e14118. [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
9. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS ONE 2011 May 4;6(5):e19467. [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
10. Dyer J, Kolic B. Public risk perception and emotion on Twitter during the Covid-19 pandemic. Appl Netw Sci 2020;5(1):99. [doi: 10.1007/s41109-020-00334-7] [Medline: 33344760]
11. Su Y, Venkat A, Yadav Y, Puglisi LB, Fodeh SJ. Twitter-based analysis reveals differential COVID-19 concerns across areas with socioeconomic disparities. Comput Biol Med 2021 May;132:104336. [doi: 10.1016/j.compbiomed.2021.104336] [Medline: 33761419]
12. Imran M, Castillo C, Diaz F, Vieweg S. Processing social media messages in mass emergency. ACM Comput Surv 2015 Jul 21;47(4):1-38. [doi: 10.1145/2771588]
13. Thorpe Huerta D, Hawkins JB, Brownstein JS, Hswen Y. Exploring discussions of health and risk and public sentiment in Massachusetts during COVID-19 pandemic mandate implementation: a Twitter analysis. SSM Popul Health 2021 Sep;15:100851. [doi: 10.1016/j.ssmph.2021.100851] [Medline: 34355055]
14. Wang Y, Hao H, Platt LS. Examining risk and crisis communications of government agencies and stakeholders during early-stages of COVID-19 on Twitter. Comput Human Behav 2021 Jan;114:106568. [doi: 10.1016/j.chb.2020.106568]
15. Saito R, Tsugawa S. Understanding citizens' response to social activities on Twitter in US metropolises during the COVID-19 recovery phase using a fine-tuned large language model: application of AI. J Med Internet Res 2025 Feb 11;27:e63824. [doi: 10.2196/63824] [Medline: 39932775]
16. Deiner MS, Deiner NA, Hristidis V, et al. Use of large language models to assess the likelihood of epidemics from the content of tweets: infodemiology study. J Med Internet Res 2024 Mar 1;26:e49139. [doi: 10.2196/49139] [Medline: 38427404]
17. Al-Ramahi M, Elnoshokaty A, El-Gayar O, Nasralah T, Wahbeh A. Public discourse against masks in the COVID-19 era: infodemiology study of Twitter data. JMIR Public Health Surveill 2021 Apr 5;7(4):e26780. [doi: 10.2196/26780] [Medline: 33720841]
18. Masri S, Jia J, Li C, et al. Use of Twitter data to improve Zika virus surveillance in the United States during the 2016 epidemic. BMC Public Health 2019 Jun 14;19(1):761. [doi: 10.1186/s12889-019-7103-8] [Medline: 31200692]
19. Modrek S, Chakalov B. The #MeToo movement in the United States: text analysis of early Twitter conversations. J Med Internet Res 2019 Sep 3;21(9):e13837. [doi: 10.2196/13837] [Medline: 31482849]

XSL•FO
RenderX

20. Taylor S, Asmundson GJG. Negative attitudes about facemasks during the COVID-19 pandemic: the dual importance of perceived ineffectiveness and psychological reactance. PLoS ONE 2021;16(2):e0246317. [doi: 10.1371/journal.pone.0246317] [Medline: 33596207]
21. Eikenberry SE, Mancuso M, Iboi E, et al. To mask or not to mask: modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. Infect Dis Model 2020;5:293-308. [doi: 10.1016/j.idm.2020.04.001]
22. Taube JC, Susswein Z, Bansal S. Spatiotemporal trends in self-reported mask-wearing behavior in the United States: analysis of a large cross-sectional survey. JMIR Public Health Surveill 2023 Mar 6;9(1):e42128. [doi: 10.2196/42128] [Medline: 36877548]
23. Ferawati K, She WJ, Wakamiya S, Aramaki E. Synchronizing approach in designing annotation guidelines for multilingual datasets: a COVID-19 case study using english and japanese tweets. In: Prabhakaran V, Dev S, Benotti L, Hershcovich D, Cabello L, Cao Y, et al, editors. Presented at: Proceedings of the 2nd Workshop on Cross-Cultural Considerations in NLP; Aug 16, 2024; Bangkok, Thailand p. 32-41 URL: https://aclanthology.org/2024.c3nlp-1 [accessed 2026-04-01] [doi: 10.18653/v1/2024.c3nlp-1.3]
24. Portal site of official statistics of Japan. e-Stat. 2024. URL: https://www.e-stat.go.jp/en [accessed 2024-03-29]
25. Census bureau data. United States Census Bureau. 2024. URL: https://data.census.gov [accessed 2024-03-29]
26. Li S, Feng B, Liao W, Pan W. Internet use, risk awareness, and demographic characteristics associated with engagement in preventive behaviors and testing: cross-sectional survey on COVID-19 in the United States. J Med Internet Res 2020 Jun 16;22(6):e19782. [doi: 10.2196/19782] [Medline: 32501801]
27. United States cities database. Simplemaps. 2022. URL: https://simplemaps.com/data/us-cities [accessed 2024-01-02]
28. MIT. List of cities in Japan. GitHub. 2019. URL: https://github.com/nobuf/list-of-cities-in-japan [accessed 2024-01-02]
29. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960 Apr;20(1):37-46. [doi: 10.1177/001316446002000104]
30. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012;22(3):276-282. [doi: 10.11613/BM.2012.031]
31. O'Dea B, Wan S, Batterham PJ, Calear AL, Paris C, Christensen H. Detecting suicidality on Twitter. Internet Interv 2015 May;2(2):183-188. [doi: 10.1016/j.invent.2015.03.005]
32. Mozetič I, Grčar M, Smailović J. Multilingual Twitter sentiment classification: the role of human annotators. PLoS ONE 2016;11(5):e0155036. [doi: 10.1371/journal.pone.0155036] [Medline: 27149621]
33. Bikku T, Jarugula J, Kongala L, Tummala ND, Vardhani Donthiboina N. Exploring the effectiveness of BERT for sentiment analysis on large-scale social media data. Presented at: 2023 3rd International Conference on Intelligent Technologies (CONIT); Jun 23-25, 2023; Hubli, India p. 1-4. [doi: 10.1109/CONIT59222.2023.10205600]
34. Liu Y, Ott M, Goyal N, et al. RoBERTa: a robustly optimized BERT pretraining approach. corr. . Preprint posted online on Jul 26, 2019 URL: http://arxiv.org/abs/1907.11692
35. Tohoku NLP. Pretrained Japanese BERT models. Hugging Face. 2023. URL: https://huggingface.co/tohoku-nlp/bert-base-japanese-v3 [accessed 2024-05-22]
36. Duarte JM, Berton L. A review of semi-supervised learning for text classification. Artif Intell Rev 2023 Sep;56(9):9401-9469. [doi: 10.1007/s10462-023-10393-8]
37. Klein AZ, Kunatharaju S, O'Connor K, Gonzalez-Hernandez G. Automatically identifying self-reports of COVID-19 diagnosis on Twitter: an annotated data set, deep neural network classifiers, and a large-scale cohort. J Med Internet Res 2023 Jul 3;25:e46484. [doi: 10.2196/46484] [Medline: 37399062]
38. Klein AZ, Gutiérrez Gómez JA, Levine LD, Gonzalez-Hernandez G. Using longitudinal Twitter data for digital epidemiology of childhood health outcomes: an annotated data set and deep neural network classifiers. J Med Internet Res 2024 Mar 25;26:e50652. [doi: 10.2196/50652] [Medline: 38526542]
39. Suzuki R, Iizuka Y, Sugawara H, Lefor AK. Wearing masks is easy but taking them off is difficult - A situation in Japan during COVID-19 pandemic and after. Dialogues Health 2024 Jun;4:100172. [doi: 10.1016/j.dialog.2024.100172] [Medline: 38516216]
40. Shepardson D, Singh RK, Mason J. US will no longer enforce mask mandate on airplanes, trains after court ruling. Reuters. 2022. URL: https://www.reuters.com/legal/government/us-judge-rules-mask-mandate-transport-unlawful-overturning-biden-effort-2022-04-18/ [accessed 2024-05-29]
41. Karako K, Song P, Chen Y, Tang W, Kokudo N. Overview of the characteristics of and responses to the three waves of COVID-19 in Japan during 2020-2021. BST 2021;15(1):1-8. [doi: 10.5582/bst.2021.01019]
42. Kamba M, She WJ, Ferawati K, Wakamiya S, Aramaki E. Exploring the impact of the COVID-19 pandemic on Twitter in Japan: qualitative analysis of disrupted plans and consequences. JMIR Infodemiology 2024 Apr 1;4:e49699. [doi: 10.2196/49699] [Medline: 38557446]
43. Ramos PJ, Ferawati K, Liew K, Aramaki E, Wakamiya S. Emotion analysis of writers and readers of Japanese tweets on vaccinations. Presented at: Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis; May 26, 2022; Dublin, Ireland p. 95-103. [doi: 10.18653/v1/2022.wassa-1.10]
44. Wakamiya S, Kawai Y, Aramaki E. Twitter-based influenza detection after flu peak via tweets with indirect information: text mining study. JMIR Public Health Surveill 2018 Sep 25;4(3):e65. [doi: 10.2196/publichealth.8627] [Medline: 30274968]

XSL•FO
RenderX

45. Liang G, Zhao J, Lau HYP, Leung CK. Using social media to analyze public concerns and policy responses to COVID-19 in Hong Kong. ACM Trans Manage Inf Syst 2021 Dec 31;12(4):1-20. [doi: 10.1145/3460124]
46. Liao Q, Yuan J, Dong M, Yang L, Fielding R, Lam WWT. Public engagement and government responsiveness in the communications about COVID-19 during the early epidemic stage in China: infodemiology study on social media data. J Med Internet Res 2020;22(5):e18796. [doi: 10.2196/18796]
47. Liu E, Arledge S. Individual characteristics and demographics associated with mask wearing during the COVID-19 pandemic in the United States. Asian J Soc Health Behav 2022;5(1):3-9. [doi: 10.4103/shb.shb_148_21]
48. Kim JK, Crimmins EM. How does age affect personal and social reactions to COVID-19: results from the national understanding America study. PLoS ONE 2020;15(11):e0241950. [doi: 10.1371/journal.pone.0241950] [Medline: 33170903]
49. Social media fact sheet. Pew Research Center. 2024. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2024-05-29]
50. Howard J, Huang A, Li Z, et al. An evidence review of face masks against COVID-19. Proc Natl Acad Sci USA 2021 Jan 26;118(4):e2014564118. [doi: 10.1073/pnas.2014564118]
51. Esmaeilzadeh P. Public concerns and burdens associated with face mask-wearing: Lessons learned from the COVID-19 pandemic. Prog Disaster Sci 2022 Jan;13:100215. [doi: 10.1016/j.pdisas.2022.100215] [Medline: 35036901]
52. Sugimura M, Chimed-Ochir O, Yumiya Y, et al. The association between wearing a mask and COVID-19. Int J Environ Res Public Health 2021 Aug 30;18(17):9131. [doi: 10.3390/ijerph18179131] [Medline: 34501719]
53. Daoust JF. Elderly people and responses to COVID-19 in 27 countries. PLoS ONE 2020;15(7):e0235590. [doi: 10.1371/journal.pone.0235590] [Medline: 32614889]
54. Takahashi Y, Kubota H, Shigeto S, Yoshida T, Yamagata Y. Diverse values of urban-to-rural migration: a case study of Hokuto City, Japan. J Rural Stud 2021 Oct;87:292-299. [doi: 10.1016/j.jrurstud.2021.09.013]
55. Zang E, West J, Kim N, Pao C. U.S. regional differences in physical distancing: evaluating racial and socioeconomic divides during the COVID-19 pandemic. PLoS One 2021;16(11):e0259665. [doi: 10.1371/journal.pone.0259665] [Medline: 34847174]
56. Jacobs P, Ohinmaa AP. The enforcement of statewide mask wearing mandates to prevent COVID-19 in the US: an overview. F1000Res 2020;9:1100. [doi: 10.12688/f1000research.25907.1] [Medline: 33520192]
57. Timpka T, Nyce JM. Face mask use during the COVID-19 pandemic-the significance of culture and the symbolic meaning of behavior. Ann Epidemiol 2021 Jul;59:1-4. [doi: 10.1016/j.annepidem.2021.03.012] [Medline: 33836290]
58. Kemmelmeier M, Jami WA. Mask wearing as cultural behavior: an investigation across 45 U.S. states during the COVID-19 pandemic. Front Psychol 2021;12:648692. [doi: 10.3389/fpsyg.2021.648692] [Medline: 34366970]
59. Gotanda H, Miyawaki A, Tabuchi T, Tsugawa Y. Association between trust in government and practice of preventive measures during the COVID-19 pandemic in Japan. J Gen Intern Med 2021 Nov;36(11):3471-3477. [doi: 10.1007/s11606-021-06959-3] [Medline: 34159544]
60. Suppasri A, Kitamura M, Tsukuda H, et al. Perceptions of the COVID-19 pandemic in Japan with respect to cultural, information, disaster and social issues. Prog Disaster Sci 2021 Apr;10:100158. [doi: 10.1016/j.pdisas.2021.100158] [Medline: 34095808]
61. Sizing up Twitter users. Pew Research Center. 2019. URL: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/ [accessed 2024-07-17]
62. FY2023 survey report on usage time of information and communications media and information behavior. : Ministry of Internal Affairs and Communications Japan; 2024 URL: https://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/pressrelease/2024/pdf/000382186_20240621_4.pdf [accessed 2026-04-01]
63. Pokhriyal N, Valentino BA, Vosoughi S. Quantifying participation biases on social media. EPJ Data Sci 2023;12(1):26. [doi: 10.1140/epjds/s13688-023-00405-6]
64. Fielding T, Ishikawa Y. COVID-19 and migration: a research note on the effects of COVID-19 on internal migration rates and patterns in Japan. Popul Space Place 2021 Aug;27(6):e2499. [doi: 10.1002/psp.2499] [Medline: 34512211]
65. Nishimura Y, Ochi K, Tokumasu K, et al. Impact of the COVID-19 pandemic on the psychological distress of medical students in Japan: cross-sectional survey study. J Med Internet Res 2021;23(2):e25232. [doi: 10.2196/25232]
66. Noda T, Nagaura H, Tsutsumi T, et al. A cross-sectional study of the psychological impact of the COVID-19 pandemic on undergraduate and graduate students in Japan. J Affect Disord Rep 2021 Dec;6:100282. [doi: 10.1016/j.jadr.2021.100282]
67. Smail E, Schneider KE, DeLong SM, et al. Health beliefs and preventive behaviors among adults during the early COVID-19 pandemic in the United States: a latent class analysis. Prev Sci 2021 Nov;22(8):1013-1022. [doi: 10.1007/s11121-021-01273-0] [Medline: 34275054]
68. Fujii R, Suzuki K, Niimi J. Public perceptions, individual characteristics, and preventive behaviors for COVID-19 in six countries: a cross-sectional study. Environ Health Prev Med 2021 Mar 3;26(1):29. [doi: 10.1186/s12199-021-00952-2] [Medline: 33657995]
69. Xue J, Chen J, Hu R, et al. Twitter discussions and emotions about the COVID-19 pandemic: machine learning approach. J Med Internet Res 2020 Nov 25;22(11):e20550. [doi: 10.2196/20550] [Medline: 33119535]
70. Boon-Itt S, Skunkan Y. Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study. JMIR Public Health Surveill 2020 Nov 11;6(4):e21978. [doi: 10.2196/21978] [Medline: 33108310]

XSL•FO
RenderX

71. Hu T, Wang S, Luo W, et al. Revealing public opinion towards COVID-19 vaccines with Twitter data in the United States: spatiotemporal perspective. J Med Internet Res 2021 Sep 10;23(9):e30854. [doi: 10.2196/30854] [Medline: 34346888]
72. Suarez-Lledo V, Alvarez-Galvez J. Assessing the role of social bots during the COVID-19 pandemic: infodemic, disagreement, and criticism. J Med Internet Res 2022 Aug 25;24(8):e36085. [doi: 10.2196/36085] [Medline: 35839385]
73. Chang HCH, Ferrara E. Comparative analysis of social bots and humans during the COVID-19 pandemic. J Comput Soc Sci 2022;5(2):1409-1425. [doi: 10.1007/s42001-022-00173-9] [Medline: 35789937]

Abbreviations

BERT: Bidirectional Encoder Representations from Transformers
CDC: Centers for Disease Control and Prevention
SoE: state of emergency

Edited by A Kushniruk; submitted 29.May.2025; peer-reviewed by K Jermsittiparsert; revised version received 14.Feb.2026; accepted 20.Feb.2026; published 08.Apr.2026.

Please cite as:
Ferawati K, She WJ, Wakamiya S, Aramaki E
Comparison of Mask-Wearing Behavior on Social Media and Its Relationship With Demographic Characteristics During the COVID-19 Pandemic: Social Media Data Analysis Between the United States and Japan
JMIR Hum Factors 2026;13:e78236
URL: https://humanfactors.jmir.org/2026/1/e78236
doi:10.2196/78236

© Kiki Ferawati, Wan Jou She, Shoko Wakamiya, Eiji Aramaki. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 8.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Using Digital Mental Health and Measurement-Based Care With Youth Experiencing Suicidal Thoughts and Behaviors: Qualitative Study of Clinician Barriers and Facilitators

Lia Norman[1,2,3], BAID, BA; Katherine Bright[2,3,4], RN, MN, PhD; Karina Pintson[2,3]; Julia Hews-Girard[2,3,5], RN, PhD; Adrijana D'Silva[1], PhD; Courtney Habina[1,2,6], BSc; Diana F Pricop[1,2,6], BA, MA; Geneca Henry[1,2,6], BA, MA; Sarah Daniel[2,3]; Lauren Volcko[2,3], BMSc; Katelyn Greer[1,2,3]; Jason Gondziola[7], BA, MBA; Leanne Stamp[7], BEd; Melanie Fersovitch[7], BN; Karen Moskovic[7], BN; Jessica Bradley[7], RN, MN; Haley M LaMonica[8], BSc, MA, PhD; Ian Hickie[8], AM, MD; Frank Iorfino[8], BSc, MBMSc, PhD; David Johnson[7,9], MD; Paul Arnold[1,2,6,10], MD, PhD; Gina Dimitropoulos[2,3,6,10], MSW, RSW, RMFT, PhD

[1]Department of Psychiatry, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[10]Alberta Children's Hospital Research Institute, University of Calgary, Calgary, AB, Canada

[2]Mathison Centre for Mental Health Research and Education, University of Calgary, Calgary, AB, Canada

[3]Faculty of Social Work, University of Calgary, MacKimmie Tower 413, 420 Campus Lane NW, Calgary, AB, Canada

[4]School of Nursing and Midwifery, Faculty of Health, Community, and Education, Mount Royal University, Calgary, Canada

[5]Faculty of Nursing, University of Calgary, Calgary, Canada

[6]Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada

[7]Department of Pediatrics, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[8]Brain and Mind Centre, The University of Sydney, Camperdown, Australia

[9]Maternal Newborn Child and Youth Strategic Clinical Network, Alberta Health Services, Calgary, Canada

Corresponding Author:
Gina Dimitropoulos, MSW, RSW, RMFT, PhD
Mathison Centre for Mental Health Research and Education, University of Calgary, Calgary, AB, Canada

Abstract

Background: Youth suicide is a public health crisis. In addition to suicide mortality, many youth experience and live with suicidal thoughts and behaviors (STBs). STBs have serious consequences for youth mental health and are associated with suicide. Despite recognition of the incidence and severity of STBs, barriers to accessing support are prevalent. Digital mental health (DMH) and digitally delivered measurement-based care (MBC) may improve access to treatment among youth and enhance clinical response to suicide risk and crises.

Objective: This study aimed to explore clinician perceptions of the barriers and facilitators to using DMH and MBC with youth experiencing STBs.

Methods: As part of a larger implementation science project, a DMH and MBC platform was implemented in youth-serving mental health service settings in communities across Alberta, Canada. The platform included a multidimensional assessment package and embedded an automated suicide escalation protocol that notified clinicians when youth reported STBs. In participating service settings, 32 interviews were conducted with clinicians using DMH and MBC. We used the Consolidated Framework for Implementation Research and theoretical thematic analysis to identify barriers, facilitators, and relevant themes.

Results: Four overarching themes were identified that described barriers and facilitators to using DMH and MBC with youth experiencing STBs: (1) service setting and professional practice barriers, (2) DMH platform barriers, (3) service setting and professional practice facilitators, and (4) DMH platform facilitators. Clinicians' recommendations were presented in alignment with the identified barrier themes.

Conclusions: Given the association between STBs and suicide, attending to STBs is a critical component of suicide prevention. Our findings suggest that the assessment of STBs using DMH may address the youth suicide crisis by facilitating disclosure and rapid clinical action. Moreover, multidimensional assessment may reveal important information about STBs and risk factors, with advantages for personalized care and system-level improvement. Clinicians delivering DMH and MBC must be supported by applying their recommendations and continuously strategizing to mitigate barriers and leverage facilitators for sustained implementation.

XSL•FO
RenderX

(JMIR Hum Factors 2026;13:e81505) doi:10.2196/81505

KEYWORDS

digital mental health; digital health; measurement-based care; mental health care; early intervention; mental health; suicide; suicidal thoughts and behaviors; youth; young adults; qualitative research; implementation science

Introduction

Background

Youth suicide is a public health crisis [1-4]. In Canada, suicide is the second leading cause of death among youth and young adults aged 15-34 years [5] and accounts for approximately 25% of deaths among youth aged 15-24 years [6]. A study of the epidemiological impact of youth suicide in countries with the highest human development index found that Canada has the third highest crude suicide mortality rate among youth in the 15-24 age range [7]. Each suicide signifies a young person who matters, years of potential life lost prematurely, and exerts a heartrending impact on families, friends, and communities [8]. Beyond suicide mortality, many youth experience and live with suicidal thoughts and behaviors (STBs) [5,9]. STBs comprise all facets of suicidality: active suicidal ideation (with intent to act), passive suicidal ideation (without intent to act), deliberate self-harm, preparatory acts (eg, planning suicide), and suicide attempts [10]. Youth aged 15-19 years exhibit the highest rates of urgent care visits and hospitalization related to self-harm in Canada [5]. Notably, STBs have serious and possibly debilitating consequences for youth's physical and mental health and well-being [8].

Risk factors for STBs and suicide are heterogeneous, fluctuating, and interactive [4,5,11]. At the systemic and community level, residing in a remote or rural area [5,12], social and economic deprivation [3,5], natural disasters, and war [8,13,14] may elevate youth's vulnerability. Other powerful systemic and community influences include experiences of marginalization [4], minoritization, discrimination, prejudice, loss of cultural identity [3], and racism [5]. For youth, interpersonal factors such as negative family and peer relationships, the death of a family member or friend [3], and peer victimization [3,15] are instrumental. Experiences of child maltreatment [16-18], witnessing intimate partner violence [3], gender-based violence [5,18], and other trauma also contribute to psychosocial stress and potential self-harm. Finally, individual factors interact with and are affected by broader systemic, community, and interpersonal dynamics [4,8]. Psychological distress [19] and feelings of worthlessness, self-loathing, isolation, and other emotions [3] are associated with youth STBs. Although not all youth experiencing STBs have a psychiatric disorder (eg, posttraumatic stress disorder), mental ill-health and comorbidities increase youth's risk [10]. A personal history of STBs is a salient contributing factor to youth suicide [2,20]. Biologically, adolescence represents a period of vulnerability to developing STBs due to rapid neurodevelopmental and physical changes [4], limited cognitive control, and lower inhibition [1]. Isolation and disrupted routines during the COVID-19 pandemic interrupted typical adolescent development [4] and increased youth's risk of suicidality [1,21,22] and suicide [1,4].

Due to pervasive social inequities, risk factors and related prevalence of STBs and suicide are not equally distributed across population groups [4,5,8]. Indigenous peoples (First Nations, Métis, and Inuit) in Canada experience risk factors including intergenerational trauma, marginalization, discrimination, racism, loss of land and culture, and differential access to resources [5,23-25] (eg, mental health services, housing, and clean water) as a consequence of historical and present colonialism and oppression [23,24]. Social and health disparities resulting from ongoing injustices contribute to elevated rates of suicide among Indigenous youth [24,26]; among Indigenous youth aged 15 to 24, the suicide rate is 5 to 6 times higher than that of non-Indigenous populations [23,27]. The suicide rate among Inuit youth in this age range is 11 times higher [23,27] and may be increasing among Inuit children aged 10-14 years [24]. However, Indigenous populations are diverse [25], and suicide is not a universal phenomenon in all Indigenous communities [23,25,28]. Community contextual factors (eg, environmental location and Indigenous governance in the community) and aspects of intersectional identity (eg, gender, sexual orientation, and social location) greatly affect risk factors for suicide [28].

Among other racialized populations, collective and individual experiences of marginalization, discrimination, racism [29,30], racial trauma [31], barriers to health resources [32], and other factors also contribute to suicide risk. There is a notable lack of race-based and ethnicity-based data [29,32] documenting suicide rates, and a lack of research to better understand unique pathways to suicidality among heterogeneous racialized communities and individuals in Canada [29]. Similarly, as gender identity and sexual orientation are not typically collected in death records, suicide rates among Canadian Two-Spirit, lesbian, gay, bisexual, transgender, queer, questioning, intersex, and asexual (2SLGBTQIA+) youth are unknown [5] or underreported. Existing data demonstrate that transgender and sexual minority youth are at higher risk of suicidal ideation and attempted suicide than their cisgender, heterosexual peers [33]. As with other historically and presently marginalized populations, heightened risk of STBs among 2SLGBTQIA+ youth occurs due to external inequities and influences, such as minority stress in various forms across societal and interpersonal levels [34-36] and potential victimization [37]. Experiences of STBs are again impacted by intersectionality and within- and between-group differences [33,35,37].

Despite recognition of the prevalence of STBs and suicide in youth, and the disproportionate burden in some communities, many youth struggling with STBs are unable to access appropriate support and treatment [10,38]. Barriers to accessing youth mental health care in Canada include limited resources in remote and rural areas, long waitlists, the cost of specialized services, a lack of support for transition-aged youth, stigma toward mental health concerns, and minimal or absent culturally

XSL•FO
RenderX

competent care [39]. Among youth with access to health care systems, barriers to disclosing STBs are well documented. Many youth experience concerns about the confidentiality of their disclosure [40-42], fears of reexperiencing past harmful reactions to disclosure [42], and anxiety about hospitalization [41]. Feelings of hopelessness [40], isolation [3], and shame [41], along with fears of stigma [3], may also prevent youth from talking about STBs. Notably, 2 studies highlight that youth may not disclose STBs to clinicians if clinicians do not directly ask questions about suicidality [40,42]. A lack of clarity around roles and responsibilities for responding to suicide risk, worries about community follow-up (eg, when youth are sent to hospital but not admitted), limited time, competing priorities, and the need for training [43] may deter clinicians from initiating conversations about STBs with youth.

Digital Mental Health and Measurement-Based Care for Youth Experiencing STBs: Overview of Existing Evidence

Considering the barriers to youth mental health care and the sensitivity around discussing STBs, tools that improve youth's access to treatment and address STBs are urgently needed. Digital mental health (DMH) includes various web- and mobile-based apps that may be used to self-manage mental health concerns [38] or used in conjunction with clinician care [19,38,44,45]. Specific to STBs, DMH modalities such as telemedicine [44], videos, games, apps, supportive text messaging [38], and wearable devices [20] leverage technology to support youth experiencing ongoing suicidality and youth in acute crises. DMH may reduce barriers to treatment access by improving accessibility irrespective of youth's geographic location [19,46], eliminating waitlists, providing low- or no-cost care options, expediting access to information [45], and increasing anonymity [46]. Some research indicates that DMH may be accessible for communities experiencing systemic marginalization, provided that barriers to access are addressed, and DMH is co-designed with the community it intends to support [47-49]. Co-design enables DMH to recognize the unique experiences of diverse groups of youth, meet their self-identified needs [48,49], and improve cultural relevance and relatability [47]. Importantly, youth perceive technologies as acceptable [44,50] and even preferable [51,52] means of crisis management and intervention. Youth who have not previously shared their STBs [53] and youth with difficulties communicating verbally [52,54] may find that DMH facilitates communication and disclosure about STBs with clinicians. Furthermore, DMH can play a role in safety planning by providing psychoeducation, teaching coping strategies [38], connecting youth to crisis resources and emergency contacts (eg, clinicians and caregivers) [38,55], and enabling rapid clinical response [53].

A significant advantage of DMH is its ability to deliver assessments of STBs and subsequently facilitate measurement-based care (MBC) [56]. In mental health care, MBC involves integrating continuous measurement of a person's outcomes across one or multiple health domains into their treatment. The digital delivery of MBC through DMH apps and websites, alongside traditional services, offers a promising way to identify youth experiencing STBs and detect changes in their suicidality over time [19,53]. Clinicians monitoring youth's MBC assessments in DMH who are alerted of suicide risk can respond by initiating contact with youth, developing a safety plan, and coordinating and escalating youth's level of care. A recent mixed methods study reported that DMH and MBC improved clinicians' response to suicide risk [57]. Given that the direct assessment of STBs is associated with a higher likelihood of disclosure [41], and that youth express wanting clinicians to initiate difficult conversations about STBs [40], DMH and MBC hold potential to improve awareness of and communication about suicidality for both youth and clinicians.

A small number of studies have documented clinician concerns about using DMH and MBC with youth. For instance, responding to crises may be outside clinicians' scope of practice [58]—particularly if crisis alerts are received outside standard working hours [58-61]. This contributes to apprehension about liability due to potential delays in responding to youth at risk of suicide [58,59]. Additionally, clinicians report having limited time to implement digital tools due to low staffing availability [45], heavy workloads [58,60], and a lack of dedicated time in their schedules to review online assessments [58]. The remote nature of DMH raises concerns about clinicians' lack of control to manage crises in person (eg, being unable to direct the patient to hospital) [46]. For these reasons, implementing DMH and MBC to support youth may be challenging without changes to staffing and roles [60].

Literature exploring the perspectives of clinicians responsible for using DMH and MBC with youth experiencing STBs appears limited. While review papers summarize various types of DMH for use with this population [20,38,44,51,52], and studies with youth [19,50,53,55] and caregiver participants [55] provide valuable insights, there is a lack of implementation research specific to clinicians and digitally delivered MBC. We identified 9 relevant studies that included clinician perceptions: 2 investigated other types of DMH [45,46], 3 examined monitoring tools similar to MBC [54,61,62], and 2 considered DMH and/or MBC more broadly [57,60]. Two studies were previously conducted by our research and implementation team in preimplementation project phases [58,59].

Study Purpose

Suicide is preventable [8], and suicide prevention is a global imperative recognized by the World Health Organization (WHO) [8,63], the United Nations Sustainable Development Goals [63], and the Government of Canada [5]. Suicide prevention involves multisectoral collaboration [8], empowering communities, youth, clinicians [5], and families with mental health resources, as well as early identification and intervention [8]. Implementing DMH and MBC in youth-serving settings may support suicide prevention efforts; however, the inclusion of clinician perspectives remains a gap in the literature. In-depth qualitative research exploring postimplementation clinician perspectives is necessary to determine what clinicians need to use DMH and MBC effectively and relationally with youth. This study therefore aimed to explore clinicians' perceptions and experiences through the following research question: "What are

XSL•FO
RenderX

the barriers and facilitators to using DMH and MBC with youth experiencing STBs?"

This study is part of a larger mixed methods implementation science project evaluating the implementation of a DMH and MBC platform in youth mental health service settings in rural and urban communities across Alberta. The larger project, implementation process, and platform are described in previous publications [58,59,64-66]. Among several outcome objectives, the project is assessing the impact of implementation on emergency department use, youth mental health, and suicide attempts. This study extends our prior work by presenting data collected postimplementation, focusing explicitly on youth experiencing STBs, and using a well-established implementation science framework to identify barriers, facilitators, and clinician recommendations specific to this topic.

In this paper, unless specifically referring to MBC, we hereinafter refer to DMH and MBC collectively as "DMH."

Methods

The Innowell DMH Platform: Description and Functions for Youth Experiencing STBs

The Innowell platform (Innowell PTY Ltd) was developed in Australia through co-design with individuals with lived experience of mental illness and health care professionals [56,57,67,68]. It is web-based and contains a multidimensional assessment package of clinical questionnaires for 20 biopsychosocial domains. Youth complete the initial baseline package and can repeat all or selected questionnaires as needed. Once complete, results are presented in a dashboard with color-coded "health cards" that indicate the severity (low-high) of youth's needs. Visual graphs that reflect improvement or decline over time are generated as questionnaires are retaken. In addition, relevant apps, e-tools, and community-specific resources are provided. The platform enables MBC and encourages collaboration by allowing shared access to youth data for youth and their clinicians. Clinicians can personalize care by monitoring results according to service setting procedures, exploring questionnaire responses with youth, and engaging youth in shared decision-making about their treatment.

Innowell includes a Suicidal Thoughts and Behaviors domain and embeds an automated suicide escalation protocol [56,67,68]. In the STBs domain, questions come from 2 evidence-based measures covering current and historical STBs and their intent (eg, suicidal ideation with or without a plan or attempt): the Suicidal Ideation Attributes Scale [69] and the Colombia-Suicide Severity Rating Scale [70]. Any expression of STBs in the questionnaire immediately triggers a pop-up notification with crisis helplines and community-specific crisis resources. If the severity of risk meets a threshold determined by the service setting, a notification is also sent to the youth's clinician and any other support person monitoring their DMH profile. The notification remains "flagged" for clinicians until they have contacted the client to confirm their safety, develop a safety plan, book an appointment, or any other action deemed appropriate by the youth and clinicians. In this way, the protocol enables identification, monitoring, and rapid clinical response to suicide risk.

Service settings implementing Innowell developed procedures for responding to crises and advised clinicians not to check the platform or respond to notifications outside working hours. However, many clinicians continued to monitor the platform at other times.

Service Settings and Participant Recruitment

Participating primary care networks (PCNs), specialized mental health services, schools, and community-based organizations in Alberta implemented the platform between 2021 and 2024. Implementation was facilitated by implementation leads located in participating communities who provided DMH education, training, and support to clinicians. Clinicians invited youth aged 15-24 years to use the platform as part of their care. To recruit clinicians with differing experiences with DMH, the research and implementation team invited clinicians who used the platform extensively, moderately, and minimally to participate.

Data Collection

A semistructured interview guide was developed based on the research and implementation team's expertise in implementation science and integrating DMH in clinical practice. Three topic areas were included in the guide: (1) implementation experience, which explored participants' roles, their use of DMH, and barriers and facilitators to implementation; (2) impact, which considered the impact of DMH on workflows, supporting youth with acute care needs (including youth experiencing STBs), and youth engagement in treatment; and (3) sustainability, which addressed participants' willingness to continue using DMH and recommended changes to facilitate sustained implementation. The semistructured approach allowed participants to respond to planned topics and questions, while offering opportunities for reflections related to STBs and other pertinent issues to occur at different times. This study's topic was inductively derived from the content of the interviews.

Individual interviews were conducted online via Zoom (Zoom Communications, Inc) to accommodate participants' various locations in Alberta, and facilitated by a senior member of the research team and one research assistant (RA). Both facilitators had no prior relationship with participants. The duration of interviews varied from 60 to 90 minutes. All interviews were recorded, transcribed, anonymized, and reviewed by the research team before data analysis to ensure their anonymity and accuracy.

Data Analysis

Qualitative Descriptive Approach

This study followed a qualitative descriptive approach [71]. In contrast to qualitative methods involving recontextualization of the data [72], qualitative description seeks to explore the data using low-inference description that closely reflects participants' own words about their perceptions and experiences [71]. Descriptive methods are well-suited to health research and intervention studies, as their straightforward documentation of why the intervention did or did not work, and what could improve, can be directly applied to create clinically meaningful

XSL•FO
RenderX

change [73]. With consideration of our study's purpose, the qualitative descriptive approach was chosen to provide a rich description of participant-identified barriers and facilitators to using DMH with youth experiencing STBs.

Qualitative description is adaptable and allows for several possible means of analyzing the data [73]. We conducted our analysis in three phases: (1) deductive coding using a recognized implementation science framework [74] to systematically organize the entirety of the data; (2) rating constructs (codes) to identify what constructs served as barriers and facilitators; and (3) conducting a theoretical thematic analysis [75] to identify patterns within the constructs, group constructs into themes, and provide an in-depth description of themes in response to the research question.

Analysis Phase 1: Deductive Coding Using the Consolidated Framework for Implementation Research

The Consolidated Framework for Implementation Research (CFIR) is a highly cited implementation science resource designed to assess determinants of innovation implementation from the perspectives of involved individuals and teams [74]. Deductive coding using the CFIR can identify factors influencing implementation related to the innovation itself, larger systems, the organization, individuals delivering and receiving the innovation, and implementation processes, and inform recommendations for sustained integration. Using the CFIR enhanced our adherence to low-inference qualitative description across all phases of analysis by providing a framework in which to situate and understand our data, thereby reducing bias and overinterpretation and improving consistency between coders.

The CFIR includes 5 domains to structure research data [74] (refer to Table 1; definitions are paraphrased for brevity and adapted for relevance to this study). Each domain contains constructs and subconstructs (39 in total) with definitions for use in coding that may be understood within the implementation context and adapted as necessary.

Table . Consolidated Framework for Implementation Research (CFIR) overarching domains (adapted from Damschroder et al [74]).

CFIR[a] domain	Paraphrased description
Innovation	This domain includes factors related to the innovation being implemented (DMH[b]). CFIR constructs consider the innovation's evidence-base, design, adaptability, complexity, cost, and other innovation features that may influence implementation.
Outer Setting	This domain includes factors related to the larger setting and systems in which the inner setting is situated (eg, health care system and legal system). CFIR constructs consider local conditions, partnerships, laws, policies, financing, and other external conditions that may influence implementation.
Inner Setting	This domain includes factors related to the setting where the innovation is directly implemented (youth-serving mental health service settings). CFIR constructs consider the inner setting's structure, staffing, culture, and other internal factors that may influence implementation. Constructs also consider whether the inner setting is accepting of change, if implementation is prioritized, and the compatibility of the innovation with inner setting processes.
Individuals	This domain includes factors related to individuals involved in implementation (clinicians and youth). CFIR constructs consider the capability, opportunity, and motivation of innovation deliverers and recipients to implement and receive the innovation, and whether the innovation meets deliverer and recipient needs.
Implementation Process	This domain includes factors related to the process of implementing the innovation (DMH). CFIR constructs consider initial assessments of deliverer and recipient needs, planning, strategizing, and collaborating during implementation, engaging recipients, and other processes required to implement the innovation.

[a]CFIR: Consolidated Framework for Implementation Research.

[b]DMH: digital mental health.

Before coding, reviewed transcripts were uploaded to a qualitative data analysis platform (Taguette [Rampin]) [76]. The research team familiarized themselves with the data by reading each transcript and memoing their observations for team discussion. Transcripts were then divided between 2 RAs and independently coded using the CFIR. RAs and the research team met weekly to share their progress, discuss questions, and resolve uncoded quotes that RAs sought feedback on. If disagreements about coding arose, the team referred to the CFIR construct definitions for clarification.

The final step involved creating a copy of the larger codebook and removing all constructs that did not contain quotes related to STBs. Within the remaining constructs, any quotes with unrelated content were removed to ensure that irrelevant topics were excluded from the second phase of analysis. Domains and constructs deemed applicable to this study are defined in the "Results" section.

XSL•FO
RenderX

Analysis Phase 2: Construct Rating—Assessment of Barriers and Facilitators

In the second phase, constructs that contained quotes related to STBs were rated using the CFIR Index Manual for Administration and Scoring [77]. The CFIR Index is a practical tool for research and quality improvement designed to classify CFIR constructs as barriers and facilitators to implementation and to assess the influence, valence, and strength of the constructs. In our study, we followed the step-by-step process outlined in the manual to identify constructs as barriers or facilitators to using DMH with youth experiencing STBs.

RAs first rated individual quotes as a negative (-), positive (+), or mixed (X) factor (containing both negative and positive sentiments). Using a 5-point Likert scale, RAs then applied a number between −2 and +2 to determine the strength of the factor's influence on implementation. Most quotes related to STBs were clearly strong barriers or facilitators and rated accordingly as −2 or +2. To establish the construct's overall rating as a barrier or facilitator, RAs observed the distribution of rated quotes and rated each construct according to the majority.

Analysis Phase 3: Theoretical Thematic Analysis

After rating relevant constructs, we conducted a theoretical thematic analysis as described by Braun and Clarke [75]. Theoretical thematic analysis is a deductive method of identifying and describing patterns in the data by applying preexisting theories or frameworks to answer a research question in detail. Guided by the CFIR, completed phases of deductive coding and construct rating produced barrier and facilitator constructs that acted as the framework for interpreting the data and clustering constructs into themes [78]. We followed Braun and Clarke's [75] six-stage iterative analytic process [78] to explore data related to clinicians' perceptions and experiences of using DMH with youth experiencing STBs, contained within relevant constructs. As with previous phases, the theoretical thematic analysis remained focused on observable details in the data in alignment with the qualitative descriptive approach [71].

In the first stage, RAs refamiliarized themselves with the data by rereading coded transcripts with the research question in mind and writing memos about relevant patterns. The second stage, coding, was previously completed. In the third stage, RAs examined the constructs and their associated quotes and began to combine them into larger patterns, themes, and subthemes. This involved considering what "central organizing concept" [78] was shared by each cluster of constructs. Developing a clear understanding of what constructs and ideas fit within a particular theme led to identifying more distinct groupings. Due to the CFIR's broad definitions of constructs, some constructs contained multiple ideas and were therefore included in more than one theme. RAs ensured that ideas did not overlap between themes—despite the interconnectedness of the data overall. Subthemes captured distinct concepts but shared the same underlying concept as the theme.

In the fourth stage, themes were reviewed to confirm their central organizing concepts, ensure their relevance to the research question, and check that no key ideas in the dataset were missed. This step involved examining whether the themes told an accurate and meaningful story about the data. Finally, the fifth and sixth stages entailed defining and naming the themes and producing the paper. Exemplar participant quotes were thoughtfully selected to illustrate specific impressions in theme titles and descriptions.

Positionality

The research and implementation team includes researchers, clinicians, and clinician-researchers at various stages in their careers and with expertise in diverse disciplines (social work, psychology, psychiatry, nursing, medicine, education, and software design and development). It is vital to acknowledge that the topic of STBs and suicide is deeply significant to the authors. Some authors carry personal lived experience with STBs, have lost friends to suicide, have worked as clinicians with youth who completed suicide, and have worked in mental health advocacy alongside parents and caregivers who lost their children to suicide. Others have worked as researchers and facilitators with youth experiencing suicidality. These experiences have shaped our perspectives on life—and inevitably impacted the way we perceived and interacted with the data. We engaged in reflexivity by recognizing our unique perspectives, partaking in group discussions so that multiple perspectives were reflected in the research, and continuously referring to the data to ensure that participants' voices remained central to the analysis.

Trustworthiness and Authenticity

This study adheres to the 5 standards used to assess trustworthiness and authenticity in qualitative research: objectivity, dependability, credibility, transferability, and application [71]. Objectivity was addressed by describing our methods in detail, recognizing potential biases, and reporting our positionality. To achieve dependability, we standardized data collection by using the same interviewer and interview guide across interviews, determining data analysis procedures based on the research question, and using a credible theoretical framework (the CFIR) to guide the analysis and remain close to the data. Researcher triangulation was achieved by 2 RAs independently coding the transcripts and subsequently discussing their findings with the larger team. To assess credibility, findings were shared with implementation leads to confirm that the analysis reflected their experiences with implementing DMH and engaging with clinicians using the platform. A discussion of transferability and generalizability of the findings is included in the "Strengths and Limitations" subheading. Finally, we considered the application of the study by presenting clinician and researcher recommendations for addressing the barriers to using DMH with youth experiencing STBs (refer to "Results" section and "Practical Implications and Researcher Recommendations" subheading).

Ethical Considerations

This study was approved by the University of Calgary Research Ethics Board (REB20-1137). Potential participants were informed about the study purpose and procedures, that participation was voluntary, and that they could withdraw at any time. Before participation, all participants provided written informed consent and completed a demographic survey via

XSL•FO
RenderX

REDCap (Research Electronic Data Capture; Vanderbilt University) secure software hosted at the University of Calgary. All identifying information was omitted from the paper to protect participants' privacy and confidentiality. Monetary compensation was not provided because participation in the project entailed implementing DMH within the scope of participants' regular duties.

Results

Participant Characteristics

Clinicians (N=32) from youth-serving mental health service settings participated in interviews: 17 from specialized mental health services, 9 from schools and community-based organizations, and 6 from PCNs. A majority were from rural communities (n=21, 66%). Participants occupied varying professions, including psychology and counseling (n=14, 44%), social work (n=11, 34%), and other roles (n=7, 22%). Most participants had worked for their service setting for ≤5 years (n=19, 59%). A majority worked full-time (n=25, 78%) and had graduate-level education (n=20, 63%). Participants' ages ranged from 20 - 29 to 50 - 69 years. Most participants reported their assigned sex as female (n=25, 78%) and most identified as women (n=24, 75%). A majority were heterosexual (n=23, 72%). Just over half of participants had European ethnic origins (n=18, 56%), and others reported various ethnic origins (n=13, 41%). Several variables were collapsed into "Other" categories due to low numbers to protect participants' identities. Full participant demographics are provided in Table 2.

XSL•FO
RenderX

Table . Participant demographics (N=32).

Variable	Value, n (%)
Community setting	
Rural community	21 (66)
Urban community	11 (34)
Service setting	
Specialized mental health services	17 (53)
Schools and community-based organizations	9 (28)
Primary care networks	6 (19)
Profession	
Psychology or counseling	14 (44)
Social work	11 (34)
Other	6 (19)
Missing	1 (3)
Length of time with organization	
<1-5 years	19 (59)
≥6 years	12 (38)
Missing	1 (3)
Full- or part-time work	
Full time	25 (78)
Part time	6 (19)
Missing	1 (3)
Education	
Graduate school	20 (63)
Bachelor’s degree	8 (25)
Other	3 (9)
Missing	1 (3)
Age (years)[a]	
20 - 29	8 (25)
40 - 49	8 (25)
30 - 39	7 (22)
50 - 69	6 (19)
Not reported	2 (6)
Missing	1 (3)
Assigned sex	
Female	25 (78)
Male	5 (16)
Missing	2 (6)
Gender identity	
Woman	24 (75)
Man	5 (16)
Missing	3 (9)
Sexual orientation	
Heterosexual	23 (72)

XSL•FO
RenderX

Variable	Value, n (%)
Other	8 (25)
Missing	1 (3)
Ethnicity	
European origins	18 (56)
Other	13 (41)
Missing	1 (3)

[a]Some variables were collapsed into an "Other" category due to low numbers. Contents of the "Other" categories are not reported to protect participants' identities. Not reported age range: during data cleaning, we identified inconsistencies between some participants' reported age (eg, ≤19 years) and education level (eg, graduate school). As such combinations were not feasible, these cases (n=2) were not reported to maintain data integrity. The extent of clinicians' platform use is not reported to protect participants' identities.

Theoretical Thematic Analysis

Overview

The CFIR delineated barriers and facilitators to using DMH with youth experiencing STBs across all 5 CFIR domains: Innovation, Outer Setting, Inner Setting, Individuals, and Implementation Process. This reflects the salience of the matter in connection with participants' service settings, professional practice, and DMH. Rated constructs that informed the theoretical thematic analysis are defined in Table 3.

XSL•FO
RenderX

Table . Synthesis of results.

Theme and subtheme[a]	CFIR[b] domain and construct included in theme, with CFIR definitions
1: Service Setting and Professional Practice Barriers	
1.1: Service setting structure does not support immediate crisis response	• Inner Setting: Structural Characteristics—Work Infrastructure (includes factors related to staffing and staff roles) • Inner Setting: Compatibility (includes factors related to the compatibility of the innovation with existing inner setting processes) • Individuals: Innovation Deliverers—Opportunity (includes factors related to the capacity, availability, and power of deliverers to effectively implement the innovation) • Individuals: Innovation Deliverers—Motivation (includes factors related to deliverers' motivation to implement and sustain use of the innovation)
1.2: DMH[c] incites tension between clinicians' responsiveness and burden of responsibility	• Outer Setting: Policies and Laws (includes factors related to laws, regulations, and professional guidelines that may affect implementation; eg, mandatory reporting laws and code of ethics) • Inner Setting: Compatibility (includes factors related to the compatibility of the innovation with existing inner setting processes) • Individuals: Innovation Deliverers—Opportunity (includes factors related to the capacity, availability, and power of deliverers to effectively implement the innovation) • Individuals: Innovation Deliverers—Motivation (includes factors related to deliverers' motivation to implement and sustain use of the innovation)
1.3: Clinicians express nuanced concerns about youth who report STBs[d] in DMH	• Implementation Process: Assessing Needs—Innovation Recipients (includes factors related to assessing the needs of recipients to inform implementation) • Individuals: Innovation Deliverers—Motivation (includes factors related to deliverers' motivation to implement and sustain use of the innovation) • Individuals: Innovation Deliverers—Opportunity (includes factors related to the capacity, availability, and power of deliverers to effectively implement the innovation)
2: DMH Platform Barriers	
2.1: Youth struggle to interpret the DMH STBs assessment and cope with results	• Innovation: Evidence-base (includes factors related to existing evidence to support use of the innovation) • Innovation: Complexity (includes factors related to the complexity of the innovation) • Innovation: Design (includes factors related to the design of the innovation) • Individuals: Innovation Recipients—Capability (includes factors related to recipients' ability and knowledge to effectively use the innovation)
2.2: Continuous DMH crisis notifications for youth with persistent or past STBs are unhelpful	• Innovation: Design (includes factors related to the design of the innovation) • Implementation Process: Assessing Needs—Innovation Recipients (includes factors related to assessing the needs of recipients to inform implementation)
3: Service Setting and Professional Practice Facilitators	
3.1: DMH supports the therapeutic relationship between clinicians and youth	• Innovation: Relative Advantage (includes factors related to whether the innovation enhances existing practice and processes) • Inner Setting: Culture—Recipient-centeredness (includes factors related to the inner setting's commitment to supporting innovation recipients) • Inner Setting: Mission Alignment (includes factors related to whether the innovation aligns with the inner setting's mission and purpose) • Implementation Process: Teaming (includes factors related to team collaboration and coordination during implementation)
4: DMH Platform Facilitators	

XSL•FO
RenderX

Theme and subtheme[a]	CFIR[b] domain and construct included in theme, with CFIR definitions
4.1: DMH enables disclosure, identification, and early intervention	• Innovation: Relative Advantage (includes factors related to whether the innovation enhances existing practice and processes) • Individuals: Innovation Recipients—Need (includes factors related to whether recipients have needs that the innovation could address)

[a]All domains, constructs, and subconstructs are from Damschroder et al [74]. Definitions are paraphrased for brevity and adapted for relevance to this study.

[b]CFIR: Consolidated Framework for Implementation Research.

[c]DMH: digital mental health.

[d]STB: suicidal thoughts and behaviors.

During theming, we noted that service setting and professional practice factors were greatly interconnected. Four overarching themes were therefore identified that describe (1) service setting and professional practice barriers, (2) DMH platform barriers, (3) service setting and professional practice facilitators, and (4) DMH platform facilitators (refer to Table 3). Relevant CFIR constructs are noted in *italics* in each theme description.

Theme 1: Service Setting and Professional Practice Barriers

Overview

Theme 1 explores barriers to using DMH with youth experiencing STBs that occur in the larger organization and within clinicians' daily practice. This includes subtheme 1.1 (service setting structure does not support immediate crisis response), subtheme 1.2 (DMH incites tension between clinicians' responsiveness and burden of responsibility), and subtheme 1.3 (clinicians express nuanced concerns about youth who report STBs in DMH).

Subtheme 1.1: "Kids Don't Just Go Into Crisis Monday to Friday, 8:30 AM to 4:30 PM": Service Setting Structure Does Not Support Immediate Crisis Response

In most interviews, the structure and purpose of service settings as "not a crisis clinic" (P29) was a principal factor hindering clinicians' ability and *motivation* to use DMH with youth experiencing STBs. This service setting barrier affected participants' overall experience with DMH implementation and contributed to individual-level hurdles to integrating MBC specifically. Although DMH is accessible 24/7 and youth can complete the STBs domain at any time, participants questioned the *compatibility* of DMH that incorporates crisis notifications with their service settings and practice due to restricted clinic hours. "Time limitations" (P12) led to delayed responses to youth who scored high for STBs after clinic hours, on weekends, and during participants' holidays. One participant captured the stress shared by many clinicians: "It's terrible when you come to work and you look [at DMH] and see that they've been suicidal since Friday night, and now it's Monday or Tuesday…" (P09). Despite knowing that DMH provides immediate access to crisis resources, including crisis helplines and prompts to visit emergency services, participants felt uneasy with the situation being outside their control and being unable to guarantee that youth would receive timely and appropriate support.

Service setting structure and purpose also related to clinicians' *opportunity* to integrate DMH and crisis response into their daily workflows and processes. For many participants, DMH was considered "an unfair ask of clinicians, who don't have flexibility in their schedule" (P29) to check and respond to crisis notifications during working hours. Concerns about clinicians' limited availability were frequently tied to concerns that they may be unable to fulfill their existing roles and responsibilities if using DMH. One participant stated:

> *What happens when we get that suicidal flag? Then what? I have a full day booked with 20 patients and I don't have time to cold call this person and make sure they're doing well…* [P22]

In the context of their *work infrastructure*, such as task distribution and staffing levels, several participants consequently struggled to see DMH as a viable means to support youth in crisis.

Subtheme 1.2: "You Should Be Able to Leave Work at Work": DMH Incites Tension Between Clinicians' Responsiveness and Burden of Responsibility

Clinicians' internal conflict about receiving crisis notifications outside working hours was a significant consideration related to the *compatibility* of DMH with their practice. This involved a tension between their compassion and desire to meet youth's urgent needs, and feeling emotionally burdened by crises that occurred in their personal time. One participant articulated consequences to clinicians' mental health:

> *I can't ignore it. I don't care if it's a day off, I'm going to call that person. I don't mind doing that, because I care about my people… But for my own mental health, there were some nights where you get an email […], and I'm sitting there stressing about this individual who has just flagged as high-risk for suicide.* [P04]

Several clinicians also noted ongoing prospective anxiety about crises arising outside work, regardless of whether this transpired in reality. One participant wondered, "…in the middle of the night, are we supposed to respond? […] I would probably not sleep until I got to talk to them" (P14), and another expressed, "If something did happen where they completed [suicide], it would break your heart, that we weren't there for them" (P09). In some cases, feelings of burden negatively impacted clinicians' general perceptions of DMH and their *motivation* to continue using it. Some participants had removed or wanted to eliminate

XSL•FO
RenderX

their work email from their phone to set explicit boundaries between their practice and private lives.

Concerning clinicians' inner tension around DMH, participants shared additional worries about *policies and laws* dictating the need to attend to youth experiencing STBs within a set timeframe. Clinicians' lack of *opportunity* for crisis response meant that some participants were wary to "take on liability that I know I can't attend to" (P29). For one participant, the potential implications of not responding arose in practice and created great discomfort. They described a situation in which a client was assessed as stable during their appointment, but later scored high for STBs in DMH:

> *…it was a question of liability. I believe it was a Friday, I had very little time to follow up, and the question was 'how' and 'what if' […]. So, it was a real moral dilemma for me of, how do I follow up now that I have this information, now that it's in my hands? […] I didn't appreciate that part of my role is somehow to follow up and always be monitoring if they're scoring high on risk […]. Ultimately, we're still somewhat responsible if we do have information in between sessions if the risk changes, and I wasn't comfortable with that.* [P20]

Due to their apprehension about liability, some participants desired "…some version of this program [DMH] where it feels less like my responsibility to check in on, to be a part of […] …something that alleviates my level, or flattens and equalizes my level of responsibility with the client" (P31).

Subtheme 1.3: "How to Respond in an Effective, Timely, Equitable Way": Clinicians Express Nuanced Concerns About Youth Who Report STBs in DMH

As clinicians engaged in *assessing the needs* of youth to guide DMH implementation, they encountered unique concerns about youth who experience STBs self-reporting their symptoms. Some participants had worked with youth that they characterized as "underreporters" (P01) who revised their responses in the DMH STBs domain or stopped using DMH entirely after being flagged as high-risk. Despite prefacing that the purpose of crisis notifications is to confirm youth's immediate safety and engage in safety planning, some youth were "not very keen" (P16) about the follow-up process and "hesitant to go back" (P19) to using DMH. Participants observed that youth may minimize the severity of their STBs due to discomfort with seeing their results or fears that their caregivers would be notified. This was understood as limiting the purpose and practicality of using DMH with certain youth.

Beyond underreporting, a more prevalent issue affecting clinicians' *motivation* to continue offering DMH was their concern that some youth "overreported" their STBs to receive a quicker response from clinicians and a higher level of care to validate their experiences. Some participants felt that youth used DMH "like a bat signal" (P24) to communicate in-the-moment feelings of distress, which they did not perceive as genuine crises. One participant explained:

> *If they plot it [STBs] in the moment in which they have acuity or crisis, then it becomes real. […] Something about the fact that it now exists on that platform […], it almost enhances or amplifies what it means. Because it's available in that moment and there's that protective element of anonymity, people just shoot for the stars, so to speak.* [P31]

"Overreporting" was noted as common among youth with complex diagnoses and acute care needs, who experience "a lower level of tolerance for suicidality […] because of the structure of their thoughts" (P01). Importantly, one participant eloquently stated the "nuances" (P16) of why youth may recurrently score high for STBs in DMH:

> *…if [DMH] is the only thing that they have, that might be the only place that they're comfortable informing that 'I am in crisis'. […] If it was just left up to the platform itself… It's hard for youth to understand, and it's hard for them to not show that they're in crisis as well, if that makes sense.* [P16]

Their sentiment illustrates that DMH may sincerely reflect the perspectives of some youth, and that youth may use DMH as a cry for help.

Considering their therapeutic practice and aims, participants saw that youth reporting STBs in DMH—and the follow-up process—may inadvertently oppose distress tolerance practices learned in treatment and reinforce ineffective coping. Most participants were unsure how to manage frequent crisis notifications that they perceived did not represent youth's true level of distress, and had limited *opportunity* to do so. One participant voiced the uncertainty shared by other clinicians:

> *I felt really torn… […]. I really struggled with how to respond in an effective, timely, equitable way, while also establishing and modelling healthy attachment and healthy boundaries. I still don't know the answer. […] I don't want to ignore the button, and then one day it is a crisis and they do need immediate support […]. But, I have a full caseload now and I don't necessarily have time to take away from another client in order to respond to what I'm thinking is a crisis, but it isn't a crisis.* [P07]

Collectively, clinicians faced multilayered concerns when integrating DMH into their practice with youth with complex needs who have varying ways of communicating their suicidality and need for support.

Theme 2: DMH Platform Barriers

Overview

Theme 2 explores barriers to using DMH with youth experiencing STBs that relate to qualities of DMH itself. This includes subtheme 2.1 (youth struggle to interpret the DMH STBs domain and cope with results) and subtheme 2.2 (continuous DMH crisis notifications for youth with persistent or past STBs are unhelpful).

XSL•FO
RenderX

Subtheme 2.1: "This Doesn't Feel Safe": Youth Struggle to Interpret the DMH STBs Domain and Cope With Results

A DMH barrier for clinicians was the *design* and *complexity* of the STBs domain and its associated assessment questions. Based on their experience and feedback from youth, participants remarked that questions about STBs were poorly worded and contained "gray areas" (P30) that could be interpreted in multiple ways. One participant shared,

> *…a good example would be […], have you ever had a suicidal thought? 'Yes'. And then they're like, but it was 20 years ago, or 10 years […], it was a long time ago. When it's not relevant anymore, it's still relevant on that survey.* [P30]

Several participants commented that youth's *capability* to comprehend the questions and accurately reflect their current STBs in their responses was subsequently limited by the lack of context associated with some of the wording. Moreover, concerns emerged related to youth's capacity to understand the overall purpose and context of the STBs domain—and the necessary follow-up that may occur as a result of their responses. One participant spoke to their experience working with young clients:

> *…a lot of children […], they do make comments about wanting to end their life or feelings of wanting to die. That can be for a multitude of reasons. It's not always suicidal ideation… It can come from wishing to be with a family member that's in heaven. It could be a way of expressing, 'I just don't want to feel angry anymore', or 'I can't deal with this really big feeling' […]. So, I worry about children being on the platform and not understanding the context…* [P21]

Given the lack of clarity, some participants questioned the STBs domain's effectiveness, accuracy, accessibility, and other *evidence-base* factors. One participant expressed, "I'm going to be very blunt; I don't trust the suicide scale on the platform" (P20), and another said, "I don't think that the suicide screen available on the platform is efficient" (P21).

Another concern among clinicians was youth's *capability* to cope with reading and seeing the visual of their STBs assessment results. If youth scored high for STBs, participants worried about how youth might "interpret or sit with this information, especially if the client is presenting with challenges to stress tolerance with new information" (P25). One participant captured clinicians' fears about youth completing the STBs domain by themselves:

> *…expecting them to do all of that on their own, without any co-regulation, without any coaching… […]. This doesn't feel safe, especially if they are suicidal youth, that doesn't cope well.* [P17]

Accordingly, some participants needed to work through DMH with youth to ensure their safety during and after completing the questionnaire. This was perceived negatively due to taking additional time from youth's sessions and limiting clients' self-determination to discuss issues that matter to them.

Subtheme 2.2: "Pigeonholed and Stuck There": Continuous DMH Crisis Notifications for Youth With Persistent or Past STBs Are Unhelpful

In discussing the DMH automated suicide escalation protocol, many clinicians mentioned that the crisis notifications were "very sensitive for our population" (P24). Participants referred to the prevalence of persistent suicidality among their clients to contextualize their concerns with the *design* of the system. If youth experiencing ongoing STBs completed the STBs domain on a regular basis, clinicians were frequently notified about youth whom they were already aware of and closely monitoring. Some participants found this unnecessary and felt it did not add value to their practice. With regard to *assessing the needs* of youth with persistent STBs, participants observed that youth themselves were discouraged by being consistently flagged as "permanent high-risk" (P06). One participant explained:

> *I had some youth who got very frustrated that it always flagged, even if they weren't actively suicidal […]. The patients who have chronic suicidal ideation, I feel like they were pigeonholed and stuck there, whereas they were making incredible gains in a lot of ways, but that portion of the platform didn't depict that.* [P04]

Another participant shared their understanding of youth's thought process: "It was like, 'this is something I go through. I'm not actually gonna do it. I actually don't want to die, right now'" (P24). Regardless of how much time had passed, youth with personal histories of STBs (eg, a past suicidal episode) who disclosed this in DMH also triggered repeated crisis notifications for clinicians.

A related barrier concerning the *design* of the protocol was clinicians' perception that continuous crisis notifications impeded their delivery of strengths-based therapy. Some participants found the automated protocol to be stigmatizing, and considered this "…a barrier for the client, but as well as for me as a practitioner—I don't necessarily want to put my client in that position" (P06). Crisis notifications further distracted from other therapeutic goals that were part of youth's treatment plan. One participant described their position:

> *…when we're flagging, then all of a sudden we're re-diverting to that […]. That becomes the focus, rather than, how do we actually go backwards and look at resiliency, and how to help people grow and develop and work through things. We're always talking about the end, fearful thing…* [P17]

Although participants recognized the necessity of addressing suicidality, they highlighted that working from a strengths-based approach can underlie and support this broader intent.

Theme 3: Service Setting and Professional Practice Facilitators

Overview

Theme 3 explores facilitators to using DMH with youth experiencing STBs that occur in the larger organization and within clinicians' daily practice. This includes subtheme 3.1

XSL•FO
RenderX

(DMH supports the therapeutic relationship between clinicians and youth).

Subtheme 3.1: "You're Not Alone": DMH Supports the Therapeutic Relationship Between Clinicians and Youth

The ability for DMH to facilitate the therapeutic alliance was seen as a significant *relative advantage* of DMH for clinicians' practice, compared with traditional services alone. Across the interviews, participants consistently said that youth using DMH and clinicians' access to their results opened the door to important conversations about suicidality that may not have occurred otherwise. Participants felt that their reaching out to discuss youth's STBs was "validating for the clients" (P11) and enabled youth to "feel like they get to tell their own story—I think it made them feel like somebody gave a darn about them" (P04). When clinicians received a crisis notification, most participants found that the follow-up process also had a profoundly positive effect on their connection with youth. One participant shared their impactful experience responding to youth in crisis:

> *As a practitioner who is very relationship-based, to be able to reach out to somebody when they have identified they're in need, and to do it quickly, I just think it's such a gift […]. To be able to reach out to that person and say, 'you're not alone, these are the resources we have, we're here for you'… I can only imagine being on the receiving end […]. When I think of a youth cohort, you're often feeling unseen and unheard, and not valued or validated. For someone to be able to reach out in a time like that, I think is a huge benefit of the application. […] I feel like this tool, for me… It's been like the magic wand that has allowed me to have a different kind of relationship.* [P18]

Notably, some participants noticed that following up about youth's STBs was particularly powerful for youth who struggle to trust adults and clinicians. Responding to crisis notifications demonstrated clinicians' reliability and responsiveness to youth. One participant stated:

> *I think there was something pretty powerful of, I tell you that I'm going to get this email, and I do actually follow through with following up with you. I think that did a lot for the relationship… […] They appreciated that an adult actually did do what they say they're going to do.* [P04]

Together with one-to-one relationships, DMH aided in cultivating a positive relationship among youth and the service setting at large. Participants provided numerous examples of *teaming* in which their team was able to "reach and wrap services around" (P18) youth experiencing STBs. For instance, one participant's colleagues rapidly and effectively responded to their client's crisis notification while they were out of office. They voiced, "It was nice to know that the days you are away or not feeling well… if someone falls down, they're still going to get caught. They're still going to have a lifeline, even if you're not it" (P15). Another participant described how the alerts emphasized the strength and cohesion of their team: "…it was a success because we learned how effective our clinician team was […] …it was a good experiment in how a tool like this could unite clinicians on our team to take quick action" (P07).

Taking into account the advantages to their practice and service setting, some participants perceived a *mission alignment* between the relationship-building capabilities of DMH and the relational values of their organizations. Participants also felt that DMH and associated setting-wide collaboration to meet youth's needs contributed to their service settings' culture of *recipient-centeredness*.

Theme 4: DMH Platform Facilitators

Overview

Theme 4 explores facilitators to using DMH with youth experiencing STBs that relate to qualities of DMH itself. This includes subtheme 4.1 (DMH enables disclosure, identification, and early intervention).

Subtheme 4.1: "To Be Able to Intervene and Make Sure Someone's Life Is Safe": DMH Enables Disclosure, Identification, and Early Intervention

In a majority of interviews, clinicians noted that DMH expanded their existing awareness about youth on their caseloads who experience STBs—and the extent and urgency of their struggles. Information about youth's STBs gleaned from DMH was considered a strong *relative advantage* even among participants who faced substantial barriers to implementing DMH with youth in crisis. For youth who "barely speak" (P13), "are not forthcoming with that information" (P12), or "would never ever approach suicidality head-on" (P10), DMH created a safe space for disclosure and facilitated identification of STBs by their clinician. Participants also recognized that the STBs domain "discovered or identified" youth whom they "had no knowledge that they were suicidal" (P09), in addition to aiding identification in cases where they were already conscious of possible risk. Some participants were uncomfortable using DMH as their primary means of identification, but appreciated it as a supplementary tool alongside in-person assessment and clinical judgment.

Following disclosure and identification, clinicians stressed another *relative advantage* of DMH: "to be able to intervene and make sure someone's life is safe" (P29). Participants frequently spoke to the benefits of DMH with regard to early intervention, the provision of timely support, the ability to share internal and external resources, and how this met the *needs* of youth. Sometimes, this included expediting youth on waitlists to receive earlier and more intensive treatment to address their STBs. One participant shared that overcoming clinician discomfort to discuss STBs identified via DMH may prevent later escalation and the need for emergency services:

> *I know clinicians are generally scared about suicide, and it's always uncomfortable. But actually, when you look at it from the perspective of, let's get in there early and have that conversation. I haven't needed to send anybody to the hospital, or get any services involved immediately […]. We've gone through a suicide [safety] plan […]. It's allowed for youth to*

XSL•FO
RenderX

> *talk through what they could do, what they might do.* [P05]

Participants also felt that DMH itself had a "big role to play" (P19) in early intervention by providing resources, such as apps and e-tools, that youth can use to self-manage their STBs and reinforce coping skills between appointments.

Finally, some clinicians were optimistic that documentation about STBs in DMH might improve the quality of the youth mental health care system overall. One participant contemplated:

> *I think there's value to measurement-based care […]. The more that we're able to research, document, the better. It's actually scary, how often something happens in the medical field that we don't have proper documentation of, we don't have proper measurement of. We don't know that 70 people were suicidal. We have absolutely no idea when, why, circumstances, how suicidal they were… The more information that we have, and can collectively share, is better.* [P13]

They and other clinicians saw the potential for documentation and resultant system-level improvement as an overarching *relative advantage* of DMH—and ultimately, a gain for young people themselves who require support with their experiences of suicidality.

Clinician Recommendations

Clinicians often strategized potential solutions to the identified service setting, professional practice, and DMH platform barriers. Their recommendations are provided in Table 4 in accordance with the overarching barrier themes.

Table . Clinician recommendations for addressing the barriers to using digital mental health (DMH) with youth experiencing suicidal thoughts and behaviors (STBs).

Theme and barrier	Clinician recommendations
Service Setting and Professional Practice Barriers	
Service settings do not support immediate crisis response due to restricted clinic hours.	• Service settings implementing DMH[a] should integrate a designated crisis response professional or crisis response team from within or outside the service setting who are available to respond to youth flagged in DMH. • This professional or team should have flexible availability within and outside service setting hours.
Clinicians have limited opportunity to integrate crisis response within their schedules.	• Refer to above recommendations.
Clinicians face barriers related to responding to youth in crisis outside working hours, including feeling burdened in their personal time and concerns about liability.	• Refer to above recommendations. • Outside service setting hours, DMH should notify local and available crisis service settings instead of notifying clinicians.
Clinicians have limited opportunity to respond and are unsure how to respond to youth in crisis that they perceive are "overreporting" their symptoms.	• The DMH STBs[b] assessment should include questions that assess "bigger picture" experiences with STBs, in addition to in-the-moment feelings of distress and suicidality.
DMH Platform Barriers	
Some wording in the DMH STBs assessment can be interpreted multiple ways and lacks context.	• Revise wording to make assessment questions clearer and easier for youth to interpret. • Clinicians note that youth in crisis are not thinking clearly; ensure that the assessment and crisis resources are easy for youth to navigate.
Some youth may struggle to interpret the DMH STBs assessment and cope with reading and seeing the visual of their results.	• Complete assessment with youth in-session; however, clinicians note that this may take additional time.
Youth with persistent or past STBs are continuously flagged as high-risk in the platform.	• The assessment should clearly distinguish between past and present STBs. • DMH should include an open text box for youth to provide comments and context about STBs (eg, when they occurred—past vs present). • DMH should include an open text box for youth to provide follow-up comments and context about STBs (eg, did they access crisis resources and are they still suicidal). • If youth are no longer in active crisis and report this, DMH should "un-flag" youth. • DMH should show improvement for youth with persistent STBs who are improving yet still require close monitoring.

[a]DMH: digital mental health.

[b]STB: suicidal thoughts and behaviors.

XSL•FO
RenderX

Discussion

Principal Findings

This study explored clinicians' perceptions of using DMH with youth experiencing STBs. A theoretical thematic analysis identified barriers and facilitators related to clinicians' service settings and professional practice, and DMH itself. Our findings highlight important topics for discussion, practical implications and recommendations, and directions for future research.

In the first theme, participants shared that their service settings were not intended to provide immediate and all-hours crisis response to youth who score high for STBs in DMH. Concerns about the compatibility of DMH due to clinic hours and school closures were expressed in preimplementation focus groups conducted by our research and implementation team [58,59]. Similarly, other studies document clinician apprehension about youth reporting their STBs outside working hours [60-62], while a clinician is on leave [61,62], or if a clinician works part-time [61]. Our findings reveal additional fears that clinicians are unable to ascertain that the crisis resources provided in DMH (eg, crisis helplines or community-specific crisis resources) will adequately support youth in their moment of need. If crisis notifications occur during working hours, clinicians' ability to respond while fulfilling their existing roles and responsibilities is still limited by hectic schedules, competing demands [58], and high-acuity client caseloads [60].

The first theme also describes that clinicians' deep care for their clients' well-being extends beyond clinic hours to impact their personal time and thoughts. Although previous studies acknowledge that DMH may contribute to clinician burden [61,62], our findings more clearly define the emotional impact of crisis notifications as a unique barrier to clinicians' acceptance of DMH. This study restates another preimplementation worry that clinicians may be held legally responsible if crisis notifications are not immediately attended to due to working hours and capacity, and an adverse event occurs [58,59].

A final and noteworthy finding in the first theme is that youth experiencing STBs may under- or overreport their symptoms in DMH. Underreporting or nondisclosure of STBs may relate to fears about confidentiality [40], potential harmful reactions from clinicians [42], or other factors, which may contribute to some youth's preference for DMH that is self-guided [38]—vs DMH that provides shared access to youth's data (eg, Innowell). Our study also identifies a marked incongruity between clinicians' assessment of the criticality of youth's STBs and youth's expressed vulnerability and urgency in the DMH STBs domain. As said by one participant, frequent crisis notifications may occur if youth are only comfortable and able to alert clinicians about their STBs through online mediums [52]. Participants' perception that responding to continuous alerts may reinforce ineffective coping is consistent with Hetrick et al's [62] finding that clinicians perceive that DMH may unintentionally create learned helplessness.

Of relevance to all concerns in the first theme, 2 studies exploring both clinician and youth perspectives about DMH emphasize youth's insight into clinician barriers [61,62]. Hetrick et al [62] found that youth were understanding of clinicians' working hours, the difficulty of responding to crisis notifications after hours, and were aware that crisis resources (eg, crisis helplines) may not guarantee immediate support. Youth participants in a co-design study shared worries that monitoring risk via DMH may be burdensome for clinicians and thus may negatively impact the therapeutic alliance [61]. While mindful of clinicians' time and capacity, youth did want their clinician to know if they had used crisis resources and received support [62]. Some youth felt that automated protocols should have sufficiently high thresholds for determining the severity of risk to reduce clinician burden [61].

In the second theme, participants perceived the DMH STBs domain's complex wording and lack of context as barriers to youth using DMH and DMH accurately reflecting youth's distress. Preimplementation data similarly captured concerns that DMH may be unsuitable for youth experiencing STBs, difficult to navigate, and difficult to interpret [58,59]. This study evidences such concerns manifesting in practice and adds that, in addition to the assessment itself, results may be challenging for youth to understand and cope with without clinician support. Accessibility, ease of use [52,54], and "clear and precise questions" [54] are noted as facilitators to youth with STBs engaging with DMH.

The second theme also describes the sensitivity of the DMH automated suicide escalation protocol and frequency of crisis notifications, contextualized by participants' acute client populations. It is worth noting that a majority of our participants were rurally located. Rural communities in Canada face a higher burden of mental health crises [12], due in part to inequitable access to mental health resources, stigma, and economic pressures [79,80]. This aligns with our finding that many youth in rural areas experience ongoing STBs or have past STBs—and thereby trigger frequent alerts. Prior work explains how youth seeing their decline in DMH may create feelings of disappointment, shame, guilt, or even the urge to engage in self-harm [50]. Further, a study about best practices for working with youth experiencing STBs [40] supports our participants' concern that continuous crisis notifications may oppose strengths-based therapeutic approaches. Youth participants were uncomfortable being labeled as at-risk, felt that risk assessment may be reductive, and desired more encouraging assessment language to facilitate help-seeking [40]. Despite this, youth accentuated the importance of talking about STBs with clinicians and wanted clinicians to ask about STBs.

Findings in the third theme demonstrate that DMH can facilitate conversations about STBs between clinicians and youth. Other researchers have addressed the potential for DMH to instigate a collaborative dialogue for youth who have not yet shared their STBs [53], youth who feel ashamed and are therefore reticent to disclose STBs in-person, and youth who are more comfortable expressing themselves online [52]. In turn, conversations about STBs and safety planning may build therapeutic rapport [44,53]. Our study adds that DMH may promote a stronger relationship between youth and the service setting by uniting clinician teams to support youth in distress. In contrast to positive findings, Sundram et al [61] found that some clinicians felt that DMH

XSL•FO
RenderX

had no impact on their relationship with youth. Systematic reviews note that DMH may only promote deeper relationships for youth experiencing STBs if it is leveraged as a complement to traditional care, rather than replacing it fully [44,52].

Finally, the fourth theme illustrates the life-saving capacity of DMH by facilitating disclosure, identification, and early intervention for STBs. Disclosure of STBs may be more likely in DMH vs face-to-face assessment given youth's comfort with technology [45,52,54,60]. As a result, clinicians are better able to detect youth's urgent needs and understand the breadth of their symptoms [19,53,54,68]. Our finding that DMH supports early intervention is congruent with several studies reporting that DMH improves service settings' and clinicians' response to suicide risk by assisting clinical decision-making about youth's treatment [19,54,57,68] and escalating youth on waitlists to receive immediate care [19,53]. Additionally, previous work indicates that apps and e-tools for STBs (such as those provided in Innowell) may bridge the gap in-between sessions [52]—particularly if clinicians practice using the resources with youth to increase their comfort level [40]. This study also brings to light that information about STBs in DMH may be documented and acted upon to improve youth mental health services and targeted supports for suicidal youth.

Practical Implications and Researcher Recommendations

The fact that many of clinicians' preimplementation concerns were realized when using DMH in practice underscores the validity of their apprehension, and the need to hear and respond to their concerns. In relation to service setting and professional practice barriers, we found that adaptations are required to ensure that youth with the courage to share their STBs receive the support that they need. We echo our participants' recommendation that service settings implementing DMH should designate staff with scheduling availability and flexibility to monitor and respond to crisis notifications in DMH. Service settings might consider adding multiple clinicians or assigning a clinical team as support persons to youth's DMH profile, as well as including caregivers or other trusted adults as emergency contacts. Automatically alerting all-hours crisis service settings in the community, rather than clinicians, may also alleviate burden outside working hours—while still documenting STBs in DMH for later viewing. Concerns about liability should be addressed by providing clinicians using DMH with sufficient training and understanding of local laws, policies, and standardized clinic procedures for responding to youth in crisis. Along with procedural changes, we call for increased organizational support for clinicians whose own mental health is affected by working with youth experiencing STBs. Access to mental health resources and regular debriefing with clinician peers and leadership may support clinicians in managing stress and vicarious trauma. Regarding youth who underreport their STBs, clinicians should consider DMH as similar to paper-and-pencil assessments and rely upon their clinical judgment and relationship with youth to determine if youth are minimizing their distress. Clinicians should recognize that youth that they perceive are "overreporting" their STBs may not feel that they are exaggerating and may require more intensive treatment. Before using DMH, creating clear expectations of clinicians' availability and response times to crisis notifications may be helpful so that youth are not distressed by not receiving an immediate response.

To resolve DMH barriers, DMH developers should differentiate crisis notifications for youth who (1) are continuously suicidal or have past STBs, vs (2) youth experiencing active crises requiring urgent support. Reporting ongoing or past STBs could trigger a "secondary" alert that suggests checking in with youth in-session rather than rapid clinical action. Our participants' recommendation for DMH to include open text boxes may allow youth to comment on assessment questions they are unsure of and add context to their responses. In Sundram et al's [61] study, youth proposed a higher threshold for alerts and suggested that some youth may be capable of identifying and communicating their level of urgency in DMH.

This study emphasizes the value of implementation frameworks to assess implementation processes, identify DMH-specific barriers, and guide the design and adaptation of DMH itself. For instance, applying frameworks such as the CFIR [74] iteratively throughout DMH development and preimplementation project phases may uncover key contextual factors to address early in design (eg, outer and inner setting context, clinician capacity, and considerations for youth with ongoing or past STBs), improve alignment between DMH features and clinical processes, and increase responsiveness to user needs.

Our findings also offer practical implications for augmenting personalized care using DMH and youth's data. Considering individual risk factors for STBs and youth's unique perceptions and experiences, personalization is a facilitator to engaging with DMH [52]. Previous work highlights that multidimensional DMH assessment may detect personal underlying issues impacting STBs [50,53] and track patterns in youth's thoughts and behaviors [50]. By recognizing contributing factors beyond the symptom itself, DMH may enhance youth's self-knowledge [50] and facilitate the delivery of personalized care by clinicians [53]. We further suggest that DMH include features that enable youth and clinicians to codevelop a personalized safety plan with coping strategies, apps, e-tools, and emergency contacts.

Aligned with our larger project's outcome objectives, evidence shows that early intervention for STBs may mitigate hospitalization [81] and suicide attempts [82]. We advocate for the continued implementation of DMH in health care systems and other youth-serving institutions and organizations (eg, schools) to meet youth where they are, identify their areas of need, and intervene for STBs at early stages.

Future Directions

As digital interventions for suicidality evolve, DMH tools that assess multiple health domains remain fundamental to capturing holistic data and recognizing the multifactorial nature of risk factors for STBs. Research on DMH specific to STBs is another important area for attention and development. At present, novel approaches include (but are not limited to) natural language processing systems that detect STBs [44], real-time monitoring tools that assess STBs at the time they occur [20], machine learning [83] and continuous-time trajectory models [84] that

XSL•FO
RenderX

predict STBs, and just-in-time adaptive interventions that provide personalized in-the-moment support [11]. There is a lack of research on DMH for STBs that acknowledges and incorporates aspects of culture [52] and intersectionality. Given the compounded risk factors experienced by some populations (eg, youth who are racialized and/or 2SLGBTQIA+), community co-design of DMH for STBs is a significant and urgent area for growth, and an opportunity to support youth who are underserved and excluded by current health care systems. Trauma-informed DMH apps for youth exposed to trauma and experiencing STBs is another direction for future research. For all types of DMH, the inclusion of youth perspectives and the lived experience lens is imperative to shape accessible, inclusive, and effective resources.

All existing and future DMH developments necessitate rigorous testing and evaluation of implementation in a variety of settings. Our research and implementation team is continuing to implement DMH and share our findings of the perspectives of clinicians [58,59,64], leadership, and young people receiving care [65] to identify obstacles and practical solutions.

Strengths and Limitations

To our knowledge, this qualitative analysis is one of few studies investigating clinician perceptions of using DMH with youth experiencing STBs. The inclusion of clinicians working in PCNs, specialized mental health services, schools, and community-based organizations in rural and urban locations across Alberta, Canada, provides insights from diverse settings. Clinicians' ethnic diversity is another strength within our demographics—although we were unable to report specific ethnic origins to protect participants' identities. During data analysis, the CFIR offered a comprehensive theoretical framework for producing a rich description of themes. The involvement of researchers with lived experience of STBs and who have been affected by youth suicide aided in understanding the complexity of experiences of STBs, receiving treatment, and working with youth as clinicians.

A limitation within our demographics is that most participants identified as women. While the gender distribution is illustrative of mental health clinicians in Canada, the addition of perspectives of men and gender diversity may have enriched our findings. Our relatively small sample size of 32 clinicians in Alberta may not be representative of other communities and contexts. Finally, this paper outlines the perceptions of clinicians only; including youth along with clinicians may have resulted in different findings and recommendations.

Conclusions

This study sought to identify barriers and facilitators to using DMH with youth experiencing STBs in youth-serving mental health service settings across Alberta, Canada. Clinicians perceived service setting and professional practice barriers, including that the structure and purpose of service settings do not support immediate response to crises flagged in DMH, DMH crisis notifications create tension between clinicians' responsiveness and the emotional burden of monitoring DMH outside working hours, and clinicians are uncertain how to manage and respond to youth with complex needs who communicate their distress in DMH in differing ways. The complexity of the DMH STBs domain and continuous crisis notifications for youth with persistent or past suicidality were identified as barriers in DMH. Despite their concerns, clinicians saw that using DMH collaboratively with youth supported the therapeutic relationship and aligned with their service settings' mission and values. Disclosure, identification, and early intervention for STBs enabled by DMH significantly benefited youth and inspired clinicians to continue implementing digital tools.

Given the association between STBs and suicide, attending to STBs is a critical component of suicide prevention. Our findings demonstrate that direct assessment of STBs using DMH may address the youth suicide crisis by facilitating disclosure and rapid clinical action to support early intervention. Moreover, multidimensional assessment may reveal important information about STBs and risk factors, with advantages for personalized care and system-level improvement. Clinicians delivering DMH must be supported by applying their recommendations and continuously strategizing to mitigate barriers and leverage facilitators for sustained implementation.

Acknowledgments

We express our thanks to Marianne Barker for facilitating interviews with clinicians, and Emma Thomsen for working with youth as a Digital Navigator. We are grateful for the clinicians who invested their time in learning and implementing digital mental health (DMH) and shared their perspectives on using DMH in practice. Finally, we extend our deepest gratitude to the youth and their parents and caregivers who have profoundly impacted our lives, and our work in the mental health field. We are honored to have known you and carry your stories with us every day.

Artificial intelligence tools were not used during the preparation or writing of this manuscript.

Funding

The authors disclose receipt of the following financial support for the research, authorship, and publication of this paper. This work was supported by the Partnership for Research and Innovation in the Health System (PRIHS) Program, Alberta Health Services, the Alberta Children's Hospital Foundation, and the Canadian Institutes of Health Research (CIHR; operating grant: SPOR iCT Rewarding Success). IH is supported by an NHMRC L3 Investigator Grant (GNT2016346). FI is supported by an NHMRC EL1 Investigator Grant (GNT2018157).

XSL•FO
RenderX

Data Availability

The datasets generated and analyzed during this study are not publicly available, in compliance with ethics requirements to protect participant confidentiality and privacy.

Authors' Contributions

Conceptualization was carried out by GD, LN, DJ, PA, KB, JHG, CH, HML, IH, FI, JG, LS, MF, KM, and JB. Data curation was performed by LN and KP. Formal analysis was conducted by LN, KP, GD, KB, JHG, AD, CH, DFP, GH, SD, LV, and KG. Funding acquisition was led by GD. Investigation was conducted by LN and KP. Methodology was developed by GD, KB, and LN. Project administration was carried out by GD, KB, JHG, and CH. Resources were provided by GD and KB. Supervision was provided by GD, KB, JHG, CH, and GH. Writing—original draft was prepared by LN, GD, and KB. Writing—review and editing was contributed by LN, GD, KB, KP, JHG, AD, CH, DFP, GH, SD, LV, KG, JG, LS, MF, KM, JB, HML, IH, FI, DJ, and PA.

Conflicts of Interest

IH is the Co-Director of Health and Policy at the Brain and Mind Centre (BMC), University of Sydney. The BMC operates early-intervention youth services at Camperdown under contract to headspace. He is the Chief Scientific Advisor to, and a 3.2% equity shareholder in, InnoWell Pty Ltd, which aims to transform mental health services through the use of innovative technologies.

References

1. Bersia M, Koumantakis E, Berchialla P, et al. Suicide spectrum among young people during the COVID-19 pandemic: a systematic review and meta-analysis. EClinicalMedicine 2022 Dec;54:101705. [doi: 10.1016/j.eclinm.2022.101705] [Medline: 36338787]
2. Castellví P, Lucas-Romero E, Miranda-Mendizábal A, et al. Longitudinal association between self-injurious thoughts and behaviors and suicidal behavior in adolescents and young adults: a systematic review with meta-analysis. J Affect Disord 2017 Jun;215:37-48. [doi: 10.1016/j.jad.2017.03.035] [Medline: 28315579]
3. Grimmond J, Kornhaber R, Visentin D, Cleary M. A qualitative systematic review of experiences and perceptions of youth suicide. PLoS One 2019;14(6):e0217568. [doi: 10.1371/journal.pone.0217568] [Medline: 31188855]
4. Kim S, Park J, Lee H, et al. Global public concern of childhood and adolescence suicide: a new perspective and new strategies for suicide prevention in the post-pandemic era. World J Pediatr 2024 Sep;20(9):872-900. [doi: 10.1007/s12519-024-00828-9] [Medline: 39008157]
5. The National Suicide Prevention Action Plan (2024 to 2027). Government of Canada. 2024. URL: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/national-suicide-prevention-action-plan-2024-2027.html [accessed 2026-04-27]
6. Findlay L. Depression and suicidal ideation among Canadians aged 15 to 24. : Statistics Canada; 2017 Jan 18 URL: https://www150.statcan.gc.ca/n1/en/pub/82-003-x/2017001/article/14697-eng.pdf?st=k5XIZjZ4 [accessed 2026-04-27]
7. Doran CM, Kinchin I. Economic and epidemiological impact of youth suicide in countries with the highest human development index. PLoS ONE 2020;15(5):e0232940. [doi: 10.1371/journal.pone.0232940] [Medline: 32428034]
8. Preventing suicide: a global imperative. World Health Organization. 2014. URL: https://www.who.int/publications/i/item/9789241564779 [accessed 2026-04-27]
9. Suicide, self-harm, and suicide-related behaviours in Canada. Government of Canada. 2025. URL: https://health-infobase.canada.ca/mental-health/suicide-self-harm/ [accessed 2026-04-27]
10. Turecki G, Brent DA. Suicide and suicidal behaviour. Lancet 2016 Mar 19;387(10024):1227-1239. [doi: 10.1016/S0140-6736(15)00234-2] [Medline: 26385066]
11. Coppersmith DDL, Dempsey W, Kleiman EM, Bentley KH, Murphy SA, Nock MK. Just-in-time adaptive interventions for suicide prevention: promise, challenges, and future directions. Psychiatry (Abingdon) 2022;85(4):317-333. [doi: 10.1080/00332747.2022.2092828] [Medline: 35848800]
12. Mahinpey N, Pollock NJ, Liu L, Contreras G, Thompson W. Self-harm and rurality in Canada: an analysis of hospitalization data from 2015 to 2019. Soc Psychiatry Psychiatr Epidemiol 2023 Aug;58(8):1161-1170. [doi: 10.1007/s00127-023-02463-7] [Medline: 37029322]
13. Kar N. Depression in youth exposed to disasters, terrorism and political violence. Curr Psychiatry Rep 2019 Jul 4;21(8):73-11. [doi: 10.1007/s11920-019-1061-9] [Medline: 31270638]
14. Rahmani M, Silverman AL, Thompson A, Pumariega A. Youth suicidality in the context of disasters. Curr Psychiatry Rep 2023 Nov;25(11):587-602. [doi: 10.1007/s11920-023-01454-4] [Medline: 37768444]
15. Biswas T, Scott JG, Munir K, et al. Global variation in the prevalence of suicidal ideation, anxiety and their correlates among adolescents: a population based study of 82 countries. EClinicalMedicine 2020 Jul;24:100395. [doi: 10.1016/j.eclinm.2020.100395] [Medline: 33015595]
16. Angelakis I, Gillespie EL, Panagioti M. Childhood maltreatment and adult suicidality: a comprehensive systematic review with meta-analysis. Psychol Med 2019 May;49(7):1057-1078. [doi: 10.1017/S0033291718003823] [Medline: 30608046]

XSL•FO
RenderX

17. Lippard ETC, Nemeroff CB. The devastating clinical consequences of child abuse and neglect: increased disease vulnerability and poor treatment response in mood disorders. Am J Psychiatry 2020 Jan 1;177(1):20-36. [doi: 10.1176/appi.ajp.2019.19010020] [Medline: 31537091]
18. Ng QX, Yong BZJ, Ho CYX, Lim DY, Yeo WS. Early life sexual abuse is associated with increased suicide attempts: an update meta-analysis. J Psychiatr Res 2018 Apr;99:129-141. [doi: 10.1016/j.jpsychires.2018.02.001] [Medline: 29454220]
19. Iorfino F, Davenport TA, Ospina-Pinillos L, et al. Using new and emerging technologies to identify and respond to suicidality among help-seeking young people: a cross-sectional study. J Med Internet Res 2017 Jul 12;19(7):e247. [doi: 10.2196/jmir.7897] [Medline: 28701290]
20. Kleiman EM, Glenn CR, Liu RT. Real-time monitoring of suicide risk among adolescents: potential barriers, possible solutions, and future directions. J Clin Child Adolesc Psychol 2019;48(6):934-946. [doi: 10.1080/15374416.2019.1666400] [Medline: 31560584]
21. Bell IH, Nicholas J, Broomhall A, et al. The impact of COVID-19 on youth mental health: a mixed methods survey. Psychiatry Res 2023 Mar;321:115082. [doi: 10.1016/j.psychres.2023.115082] [Medline: 36738592]
22. Madigan S, Korczak DJ, Vaillancourt T, et al. Comparison of paediatric emergency department visits for attempted suicide, self-harm, and suicidal ideation before and during the COVID-19 pandemic: a systematic review and meta-analysis. Lancet Psychiatry 2023 May;10(5):342-351. [doi: 10.1016/S2215-0366(23)00036-6] [Medline: 36907199]
23. Giroux R, Homer K, Kassam S, et al. Mental health and suicide in Indigenous communities in Canada. : Canadian Federation of Medical Students; 2018 URL: https://www.suicideinfo.ca/wp-content/uploads/gravity_forms/6-191a85f36ce9e20de2e2fa3869197735/2018/07/Mental-Health-and-Suicide-in-Indigenous-Communities-in-Canada_oa.pdf [accessed 2026-04-27]
24. Kumar MB, Tjepkema M. Suicide among First Nations people, Métis and Inuit (2011-2016): Findings from the 2011 Canadian Census Health and Environment Cohort (CanCHEC). : Statistics Canada; 2019 URL: https://publications.gc.ca/collections/collection_2019/statcan/99-011-x2019001-eng.pdf [accessed 2026-04-27]
25. White J, Mushquash C. We belong: life promotion to address Indigenous suicide. : Thunderbird Partnership Foundation; 2016 URL: https://wisepractices.ca/wp-content/uploads/2017/12/White-Mushquash-2016-FINAL.pdf [accessed 2026-04-27]
26. Barker B, Goodman A, DeBeck K. Reclaiming Indigenous identities: culture as strength against suicide among Indigenous youth in Canada. Can J Public Health 2017 Jun 16;108(2):e208-e210. [doi: 10.17269/cjph.108.5754] [Medline: 28621659]
27. Hus Y, Segal O. Unravelling suicide and related behaviours in Indigenous youth and young adults in the Canadian context. Neuropsychiatr Dis Treat 2024;20:2073-2094. [doi: 10.2147/NDT.S479491] [Medline: 39512427]
28. Ansloos J. Rethinking Indigenous suicide. IJIH 2018;13(2):8-28. [doi: 10.32799/ijih.v13i2.32061]
29. Experiences with suicide: African, Caribbean, and Black communities in Canada. : Mental Health Commission of Canada; 2024 URL: https://mentalhealthcommission.ca/resource/experiences-with-suicide-african-caribbean-and-black-communities-in-canada/ [accessed 2026-05-07]
30. Molock SD, Boyd RC, Alvarez K, et al. Culturally responsive assessment of suicidal thoughts and behaviors in youth of color. Am Psychol 2023 Oct;78(7):842-855. [doi: 10.1037/amp0001140] [Medline: 36913280]
31. Williams MT, Khanna Roy A, MacIntyre MP, Faber S. The traumatizing impact of racism in Canadians of colour. Curr Trauma Rep 2022;8(2):17-34. [doi: 10.1007/s40719-022-00225-5] [Medline: 35345606]
32. Fante-Coleman T, Jackson-Best F. Barriers and facilitators to accessing mental healthcare in Canada for Black youth: a scoping review. Adolescent Res Rev 2020 Jun;5(2):115-136. [doi: 10.1007/s40894-020-00133-2]
33. Kingsbury M, Hammond NG, Johnstone F, Colman I. Suicidality among sexual minority and transgender adolescents: a nationally representative population-based study of youth in Canada. CMAJ 2022 Jun 6;194(22):E767-E774. [doi: 10.1503/cmaj.212054] [Medline: 35667666]
34. Gorse M. Risk and protective factors to LGBTQ+ youth suicide: a review of the literature. Child Adolesc Soc Work J 2022 Feb;39(1):17-28. [doi: 10.1007/s10560-020-00710-3]
35. Green AE, Price MN, Dorison SH. Cumulative minority stress and suicide risk among LGBTQ youth. Am J Community Psychol 2022 Mar;69(1-2):157-168. [doi: 10.1002/ajcp.12553] [Medline: 34534356]
36. Wallace ER, O'Neill S, Lagdon S. Risk and protective factors for suicidality among lesbian, gay, bisexual, transgender, and queer (LGBTQ+) young people, from countries with a high global acceptance index (GAI), within the context of the socio-ecological model: a scoping review. J Adolesc 2024 Jul;96(5):897-924. [doi: 10.1002/jad.12308] [Medline: 38372179]
37. Williams AJ, Jones C, Arcelus J, Townsend E, Lazaridou A, Michail M. A systematic review and meta-analysis of victimisation and mental health prevalence among LGBTQ+ young people with experiences of self-harm and suicide. PLoS ONE 2021;16(1):e0245268. [doi: 10.1371/journal.pone.0245268] [Medline: 33481862]
38. Kruzan KP, Biernesser C, Hoffmann JA, Meyerhoff J. Digital interventions for adolescents and young adults experiencing self-injurious thoughts and behaviors. Curr Treat Options Psychiatry 2024 Jun;11(2):76-89. [doi: 10.1007/s40501-024-00318-9] [Medline: 39525358]
39. Murray JK, Knudson S. Mental health treatment and access for emerging adults in Canada: a systematic review. Front Public Health 2023;11:1088999. [doi: 10.3389/fpubh.2023.1088999] [Medline: 37501945]

XSL•FO
RenderX

40. Bellairs-Walsh I, Perry Y, Krysinska K, et al. Best practice when working with suicidal behaviour and self-harm in primary care: a qualitative exploration of young people's perspectives. BMJ Open 2020 Oct 28;10(10):e038855. [doi: 10.1136/bmjopen-2020-038855] [Medline: 33115897]
41. Fox KR, Bettis AH, Burke TA, Hart EA, Wang SB. Exploring adolescent experiences with disclosing self-injurious thoughts and behaviors across settings. Res Child Adolesc Psychopathol 2022 May;50(5):669-681. [doi: 10.1007/s10802-021-00878-x] [Medline: 34705197]
42. McGillivray L, Rheinberger D, Wang J, Burnett A, Torok M. Non-disclosing youth: a cross sectional study to understand why young people do not disclose suicidal thoughts to their mental health professional. BMC Psychiatry 2022 Jan 4;22(1):3. [doi: 10.1186/s12888-021-03636-x] [Medline: 34983460]
43. Davis M, Siegel J, Becker-Haimes EM, et al. Identifying common and unique barriers and facilitators to implementing evidence-based practices for suicide prevention across primary care and specialty mental health settings. Arch Suicide Res 2023;27(2):192-214. [doi: 10.1080/13811118.2021.1982094] [Medline: 34651544]
44. Forte A, Sarli G, Polidori L, Lester D, Pompili M. The role of new technologies to prevent suicide in adolescence: a systematic review of the literature. Medicina (Kaunas) 2021 Jan 26;57(2):109. [doi: 10.3390/medicina57020109] [Medline: 33530342]
45. Rheinberger D, Baffsky R, McGillivray L, et al. Examining the feasibility of implementing digital mental health innovations into hospitals to support youth in suicide crisis: interview study with young people and health professionals. JMIR Form Res 2023 Nov 16;7:e51398. [doi: 10.2196/51398] [Medline: 37971790]
46. Gilmore AK, Ward-Ciesielski EF. Perceived risks and use of psychotherapy via telemedicine for patients at risk for suicide. J Telemed Telecare 2019 Jan;25(1):59-63. [doi: 10.1177/1357633X17735559] [Medline: 28990454]
47. Bakhti R, Daler H, Ogunro H, Hope S, Hargreaves D, Nicholls D. Exploring engagement with and effectiveness of digital mental health interventions in young people of different ethnicities: systematic review. J Med Internet Res 2025 Apr 7;27:e68544. [doi: 10.2196/68544] [Medline: 40194267]
48. Hicks LJ, Toombs E, Lund J, Kowatch KR, Hopkins C, Mushquash CJ. Expanding our understanding of digital mental health interventions for Indigenous youth: an updated systematic review. J Telemed Telecare 2025 Aug;31(7):919-960. [doi: 10.1177/1357633X241239715] [Medline: 38584397]
49. Lucassen M, Samra R, Iacovides I, et al. How LGBT+ young people use the internet in relation to their mental health and envisage the use of e-therapy: exploratory study. JMIR Serious Games 2018 Dec 21;6(4):e11249. [doi: 10.2196/11249] [Medline: 30578194]
50. Kruzan KP, Mohr DC, Reddy M. How technologies can support self-injury self-management: perspectives of young adults with lived experience of nonsuicidal self-injury. Front Digit Health 2022;4:913599. [doi: 10.3389/fdgth.2022.913599] [Medline: 35847416]
51. Holland M, Hawks J, Morelli LC, Khan Z. Risk assessment and crisis intervention for youth in a time of telehealth. Contemp Sch Psychol 2021;25(1):12-26. [doi: 10.1007/s40688-020-00341-6] [Medline: 33425481]
52. Luo M, Yue Y, Du N, Xiao Y, Chen C, Huan Z. Needs for mobile and internet-based psychological intervention in patients with self-injury and suicide-related behaviors: a qualitative systematic review. BMC Psychiatry 2024 Jan 4;24(1):26. [doi: 10.1186/s12888-023-05477-2] [Medline: 38178028]
53. Chong MK, Hickie IB, Ottavio A, et al. A digital approach for addressing suicidal ideation and behaviors in youth mental health services: observational study. J Med Internet Res 2024 Dec 18;26:e60879. [doi: 10.2196/60879] [Medline: 39693140]
54. Hetrick SE, Goodall J, Yuen HP, et al. Comprehensive online self-monitoring to support clinicians manage risk of suicide in youth depression. Crisis 2017 May;38(3):147-157. [doi: 10.1027/0227-5910/a000422] [Medline: 27659516]
55. McManama O'Brien KH, LeCloux M, Ross A, Gironda C, Wharff EA. A pilot study of the acceptability and usability of a smartphone application intervention for suicidal adolescents and their parents. Arch Suicide Res 2017;21(2):254-264. [doi: 10.1080/13811118.2016.1182094] [Medline: 27136115]
56. Iorfino F, Cross SP, Davenport T, et al. A digital platform designed for youth mental health services to deliver personalized and measurement-based care. Front Psychiatry 2019;10:595. [doi: 10.3389/fpsyt.2019.00595] [Medline: 31507465]
57. McKenna S, Piper S, Capon W, et al. The polarization of clinician and service staff perspectives after the use of health information technology in youth mental health services: implementation and evaluation study. JMIR Hum Factors 2023 Jul 25;10:e42993. [doi: 10.2196/42993] [Medline: 37490321]
58. Bassi EM, Bright KS, Norman LG, et al. Perceptions of mental health providers of the barriers and facilitators of using and engaging youth in digital mental-health-enabled measurement based care. Digit Health 2024;10:20552076241253093. [doi: 10.1177/20552076241253093] [Medline: 38726214]
59. Dimitropoulos G, Bassi EM, Bright KS, et al. Implementation of an electronic mental health platform for youth and young adults in a school context across Alberta, Canada: thematic analysis of the perspectives of stakeholders. JMIR Ment Health 2024 Jan 17;11:e49099. [doi: 10.2196/49099] [Medline: 38231558]
60. Lattie EG, Nicholas J, Knapp AA, Skerl JJ, Kaiser SM, Mohr DC. Opportunities for and tensions surrounding the use of technology-enabled mental health services in community mental health care. Adm Policy Ment Health 2020 Jan;47(1):138-149. [doi: 10.1007/s10488-019-00979-2] [Medline: 31535235]

XSL•FO
RenderX

61. Sundram F, Hawken SJ, Stasiak K, et al. Tips and traps: lessons from codesigning a clinician e-monitoring tool for computerized cognitive behavioral therapy. JMIR Ment Health 2017 Jan 11;4(1):e3. [doi: 10.2196/mental.5878] [Medline: 28077345]
62. Hetrick SE, Robinson J, Burge E, et al. Youth codesign of a mobile phone app to facilitate self-monitoring and management of mood symptoms in young people with major depression, suicidal ideation, and self-harm. JMIR Ment Health 2018 Jan 23;5(1):e9. [doi: 10.2196/mental.9041] [Medline: 29362208]
63. Suicide worldwide in 2019: global health estimates. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/9789240026643 [accessed 2026-05-07]
64. Barker M, Hews-Girard J, Pinston K, et al. Organizational factors impacting the implementation of a digital mental health tool in Alberta's mental health care of youth and young adults. Digit Health 2025;11:20552076241310341. [doi: 10.1177/20552076241310341] [Medline: 39801578]
65. Daniel S, Volcko L, Bassi E, et al. Exploring youth perspectives on digital mental health platforms: qualitative descriptive study. JMIR Hum Factors 2025 May 29;12:e69907. [doi: 10.2196/69907] [Medline: 40440597]
66. Hews-Girard J, Bright K, Barker M, et al. Mental health provider and youth service users' perspectives regarding implementation of a digital mental health platform for youth: a survey study. Digit Health 2024;10:20552076241289179. [doi: 10.1177/20552076241289179] [Medline: 39421310]
67. Davenport TA, LaMonica HM, Whittle L, et al. Validation of the InnoWell platform: protocol for a clinical trial. JMIR Res Protoc 2019 May 31;8(5):e13955. [doi: 10.2196/13955] [Medline: 31152524]
68. LaMonica HM, Iorfino F, Lee GY, et al. Informing the future of integrated digital and clinical mental health care: synthesis of the outcomes from project synergy. JMIR Ment Health 2022 Mar 9;9(3):e33060. [doi: 10.2196/33060] [Medline: 34974414]
69. van Spijker BAJ, Batterham PJ, Calear AL, et al. The suicidal ideation attributes scale (SIDAS): community-based validation study of a new scale for the measurement of suicidal ideation. Suicide Life Threat Behav 2014 Aug;44(4):408-419. [doi: 10.1111/sltb.12084] [Medline: 24612048]
70. Posner K, Brown GK, Stanley B, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry 2011 Dec;168(12):1266-1277. [doi: 10.1176/appi.ajp.2011.10111704] [Medline: 22193671]
71. Colorafi KJ, Evans B. Qualitative descriptive methods in health science research. HERD 2016 Jul;9(4):16-25. [doi: 10.1177/1937586715614171] [Medline: 26791375]
72. Hall S, Liebenberg L. Qualitative description as an introductory method to qualitative research for master's-level students and research trainees. Int J Qual Methods 2024 Jan;23:23. [doi: 10.1177/16094069241242264]
73. Doyle L, McCabe C, Keogh B, Brady A, McCann M. An overview of the qualitative descriptive design within nursing research. J Res Nurs 2020 Aug;25(5):443-455. [doi: 10.1177/1744987119880234] [Medline: 34394658]
74. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated Consolidated Framework for Implementation Research based on user feedback. Implement Sci 2022 Oct 29;17(1):75. [doi: 10.1186/s13012-022-01245-0] [Medline: 36309746]
75. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
76. About Taguette. Taguette. URL: https://www.taguette.org/about.html [accessed 2026-04-27]
77. Assefa M, McGovern M. The Consolidated Framework for Implementation Research (CFIR) index manual for administration and scoring. : Stanford University School of Medicine; 2019 URL: https://med.stanford.edu/content/dam/sm/cdi/documents/CFIR-Index-Manual_5.06.19.pdf [accessed 2026-04-27]
78. Terry G, Hayfield N, Clarke V, Braun V. Thematic analysis. In: Willig C, Rogers WS, editors. The SAGE Handbook of Qualitative Research in Psychology, 2nd edition: SAGE Publications, Inc; 2017. [doi: 10.4135/9781526405555.n2]
79. Paquet J, Hibbard K, Brett-MacLean P. The landscape of non-psychotic psychiatric illness in rural Canada: a narrative review. Rural Remote Health 2024 Feb;24(1):1-11. [doi: 10.22605/RRH8341] [Medline: 38331714]
80. Rauch K, Ryan K, Ramsey D, et al. Suicidality in rural communities: a scoping review of research in Canada. Can J Community Ment Health 2023 May 1;42(1):41-57. [doi: 10.7870/cjcmh-2023-005]
81. Arango A, Gipson PY, Votta JG, King CA. Saving lives: recognizing and intervening with youth at risk for suicide. Annu Rev Clin Psychol 2021 May 7;17(1):259-284. [doi: 10.1146/annurev-clinpsy-081219-103740] [Medline: 33544628]
82. Brent DA, McMakin DL, Kennard BD, Goldstein TR, Mayes TL, Douaihy AB. Protecting adolescents from self-harm: a critical review of intervention studies. J Am Acad Child Adolesc Psychiatry 2013 Dec;52(12):1260-1271. [doi: 10.1016/j.jaac.2013.09.009] [Medline: 24290459]
83. McHugh CM, Ho N, Iorfino F, et al. Predictive modelling of deliberate self-harm and suicide attempts in young people accessing primary care: a machine learning analysis of a longitudinal study. Soc Psychiatry Psychiatr Epidemiol 2023 Jun;58(6):893-905. [doi: 10.1007/s00127-022-02415-7] [Medline: 36854811]
84. Varidel M, Hickie IB, Prodan A, et al. Dynamic learning of individual-level suicidal ideation trajectories to enhance mental health care. NPJ Ment Health Res 2024 Jun 7;3(1):26. [doi: 10.1038/s44184-024-00071-0] [Medline: 38849429]

XSL•FO
RenderX

Abbreviations

2SLGBTQIA+: Two-Spirit, Lesbian, Gay, Bisexual, Transgender, Queer or Questioning, Intersex, Asexual, + other sexual orientations and gender identities
CFIR: Consolidated Framework for Implementation Research
DMH: digital mental health
MBC: measurement-based care
PCN: primary care network
RA: research assistant
REDCap: Research Electronic Data Capture
STB: suicidal thoughts and behaviors
WHO: World Health Organization

Edited by A Kushniruk; submitted 29.Jul.2025; peer-reviewed by A AL-Asadi, DG Goldberg; revised version received 25.Mar.2026; accepted 27.Mar.2026; published 18.May.2026.

Please cite as:
Norman L, Bright K, Pintson K, Hews-Girard J, D'Silva A, Habina C, Pricop DF, Henry G, Daniel S, Volcko L, Greer K, Gondziola J, Stamp L, Fersovitch M, Moskovic K, Bradley J, LaMonica HM, Hickie I, Iorfino F, Johnson D, Arnold P, Dimitropoulos G
Using Digital Mental Health and Measurement-Based Care With Youth Experiencing Suicidal Thoughts and Behaviors: Qualitative Study of Clinician Barriers and Facilitators
JMIR Hum Factors 2026;13:e81505
URL: https://humanfactors.jmir.org/2026/1/e81505
doi:10.2196/81505

© Lia Norman, Katherine Bright, Karina Pintson, Julia Hews-Girard, Adrijana D'Silva, Courtney Habina, Diana F Pricop, Geneca Henry, Sarah Daniel, Lauren Volcko, Katelyn Greer, Jason Gondziola, Leanne Stamp, Melanie Fersovitch, Karen Moskovic, Jessica Bradley, Haley M. LaMonica, Ian Hickie, Frank Iorfino, David Johnson, Paul Arnold, Gina Dimitropoulos. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 18.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Service Users' Views on Digital Remote Monitoring for Psychosis: Survey Study

Xiaolong Zhang[1], PhD; Emily Eisner[1,2], PhD; Daniela Di Basilio[1,3], PhD; Cara Richardson[1], PhD; Joseph Firth[1,2], PhD; Sandra Bucci[1,2], DClinPsy

[1]Division of Psychology and Mental Health, School of Health Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, The University of Manchester, 1st Floor, Jean McFarlane Building, 176 Oxford Rd, Manchester, United Kingdom

[2]Greater Manchester Mental Health NHS Foundation Trust, Manchester, United Kingdom

[3]Division of Health Research, Faculty of Health and Medicine, Lancaster University, Lancaster, United Kingdom

Corresponding Author:
Sandra Bucci, DClinPsy
Division of Psychology and Mental Health, School of Health Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, The University of Manchester, 1st Floor, Jean McFarlane Building, 176 Oxford Rd, Manchester, United Kingdom

Abstract

Background: Digital remote monitoring using smartphones and wearable devices is a promising solution for psychosis management, where precise, time-sensitive intervention is crucial. Combining active symptom monitoring (ASM) and passive sensing (PS) can support self-management by allowing remote, low-burden mental health monitoring.

Objective: This study aimed to explore (1) views on collecting data using ASM and PS methods and comfort levels with different types of data gathered via these methods, (2) views on using smartphones and wearable devices in the context of mental health care, and (3) the ownership and usage of smartphones and wearable devices.

Methods: We conducted a cross-sectional survey study with service users with psychosis in the United Kingdom between March 2023 and March 2024.

Results: A total of 309 participants completed the survey. They reported mixed views on using ASM and PS technologies for monitoring mental health, with more participants endorsing the concept than opposing it (ASM: n=145, 46.9% and PS: n=132, 42.7%). However, the type of data gathered using these methods was an important factor. Collecting personal information was deemed less acceptable (P<.001) than other data types (physical health, mental health, environment, and nonpersonal device information).

Conclusions: We found that participants were comfortable with using apps and wearables for digital remote monitoring, though personal information was less acceptable than other data types due to privacy and surveillance concerns. This highlights the importance of further exploring trust issues related to digital monitoring and ensuring that end users have choices regarding the types of data that digital systems gather and share with mental health services.

(JMIR Hum Factors 2026;13:e86152) doi:10.2196/86152

KEYWORDS

digital remote monitoring; psychosis; service user; passive sensing; smartphone; wearable

Introduction

Background

Conventional methods of monitoring and treating psychosis rely on service users recalling symptoms over the preceding period, often weeks or months, at a scheduled appointment time with a member of their care team [1,2]. However, the accuracy of symptom recall is often compromised due to infrequent clinic appointments in overstretched mental health services [3,4]. This poses challenges for psychosis management, where precise, time-sensitive treatment is crucial. In this context, mental health services are increasingly advocating for the adoption of digital tools to overcome the limitations of conventional assessment and intervention approaches [5].

Digital remote monitoring (DRM) in the context of mental health care refers to the use of digital technology to track and assess mental health symptoms remotely. It includes active symptom monitoring ([ASM]; self-reported data) and passive sensing ([PS]; sensor-derived data), enabling continuous, low-burden monitoring and the opportunity for timely, personalized interventions [3,6,7]. PS using smartphones and wearables can collect extensive sensor data, from which features can be extracted to infer clinical states [8]. In psychosis specifically, a recent systematic review found that the most used

XSL•FO
RenderX

devices for PS were accelerometers and smartphones, and the most commonly derived features were sleep duration and time spent sedentary [9]. Despite being in its early stages, recent advancements in DRM for psychosis show its potential for assessment and treatment [10]. Specifically, it has been used to monitor psychotic symptoms [11,12], functional outcomes [12], and sleep [13]. Beyond symptom monitoring, digital tools have also been used for relapse prediction. A recent systematic review found that, as of 2022, 52 studies had used ASM or PS to track early warning signs and other antecedent behaviors of relapse or symptom worsening, reflecting the rapid growth of these technologies in mental health care [14].

Numerous surveys have demonstrated a continuous increase in digital technology ownership among individuals with psychosis [15-17]. A recent survey study reported a 90% smartphone ownership rate [18], suggesting that access to devices may no longer be a significant barrier to using digital health technologies in psychosis care. Moreover, the survey indicated that most participants used instant messaging and social media multiple times per day, while health-related apps were rarely used [18]. However, despite this accessibility, service user acceptance and engagement with digital health systems have remained low, partly due to the insufficient understanding of stakeholders' preferences and needs [19]. Indeed, a recent review exploring the barriers and facilitators of engagement with digital mental health interventions (DMHIs) highlighted that successful implementation depends on the functionality of digital health systems, the relevance of their content to users' specific needs, and users perceiving value and benefit from using these interventions [20]. Additionally, demographic and other factors may affect engagement, but their effects are inconsistent across studies [20]; for example, some studies found higher engagement among older adults [21,22], while others found higher engagement among younger people [23,24] or mixed effects [25]. While previous research has explored general barriers and facilitators to engagement, there is a gap in understanding perceptions of PS technologies in this population.

Aims

This study aimed to understand the views of service users with psychosis on using personal digital devices for remote symptom monitoring. We aimed to explore (1) views on collecting data using ASM and PS methods specifically in the context of an individual's mental health care, and comfort levels with different types of data gathered via these methods; (2) broad views on using smartphones and wearable devices in the context of mental health care; and (3) how individuals with psychosis own and use smartphones and wearable devices, as well as the barriers they perceive to their use. In this study, digital health tools refer to apps, sensing features on a smartphone, wearable devices, or smartwatches that are used for monitoring and managing mental and physical health (eg, monitoring symptoms, counting steps, helping with sleep routines, and setting medication reminders).

Methods

Participants

We conducted a one-off survey study with service users with psychosis between March 2023 and March 2024. To enhance inclusivity, we deliberately applied a broad criterion for psychosis, incorporating both the use of antipsychotic medication and the presence of psychotic symptoms (ie, hallucinations, delusions, confused, and disturbed thoughts) in addition to a formal diagnosis where available. More specifically, eligibility criteria were (1) self-reported diagnosis of a schizophrenia-spectrum disorder, taking antipsychotic medication, or experiencing psychotic symptoms; (2) aged 18 years or older; and (3) provision of informed consent. Participants not sufficiently fluent in English to complete the survey were excluded from participation. To reach a broad sample, we used a wide range of recruitment methods: social media platforms (eg, X (X Corp) and Reddit (Reddit, Inc), displaying posters in mental health services, asking clinicians to identify potential participants from their caseload, recruitment via the Clinical Research Network across England, existing research cohorts of the research team, and third sector organizations.

Data Collection

The survey (Table S1 in Multimedia Appendix 1) took approximately 20 - 30 minutes to complete and gathered demographic information and explored participants' experiences with and opinions of using smartphone apps and wearable devices for mental health management. It also explored views on using digital health tools in clinical practice. The survey was developed based on a review of existing literature on DMHIs in mental health care [26-37]. To enhance participants' understanding of DRM, the survey included brief explanations of ASM and PS, accompanied by images of various digital devices. Participants could choose to complete the questionnaire in either an online or paper format based on their preference. The Qualtrics (Qualtrics, LLC) platform [38] was used to deliver the online version, accessed via a URL link or a QR code. The participant information sheet was embedded in the survey for potential participants to read, followed by a series of screening questions to assess eligibility. After participants completed the screening questions, those who were not eligible were shown a thank-you message explaining their ineligibility, and the survey ended. Eligible participants were directed to the survey questions. For participants completing a paper questionnaire, a study pack containing the paper survey (identical in content to the online version) and a return postage-paid envelope was sent. Paper questionnaire responses were then entered manually into the Qualtrics database. To monitor data integrity and minimize the risk of survey fraud, study data were reviewed monthly by the research team. Summary completion numbers were shared with participating NHS trusts, and local contacts were asked to confirm that the number of surveys received aligned with their recruitment logs. Patient and public contributors in our digital mental health group reviewed survey items for acceptability and relevance.

Data Analysis

Quantitative analysis was performed using R (R Core Team and R Foundation for Statistical Computing) [39]. Descriptive statistics, including frequencies and percentages, were calculated. To understand the differences in participants' comfort level with different features collected via DRM, we

XSL•FO
RenderX

used principal component analysis (PCA) to cluster features into high-level data-type groups (eg, personal data and health data). The clustering process was informed by a previous study on individuals' perspectives on sharing different types of digital remote monitoring data [40]. Using this as a conceptual framework, an initial grouping of features was proposed and subsequently refined through iterative discussions among authors XZ, EE, and SB. PCA biplots were examined to assess how features grouped across components, and the pair of components that provided the clearest visual separation between feature groups was used to guide clustering. Following that, we conducted the Kruskal-Wallis test to compare participants' comfort levels between data-type groups. Dunn post hoc test with Bonferroni correction was applied to identify the specific significant differences. $P<.05$ was considered statistically significant. Data visualization was created using R and Microsoft Excel [41]. We used content analysis to summarize the free-text data [42]. Initial coding was conducted by author XZ, and the results were subsequently reviewed and refined by authors EE and SB. When divergence occurred, a meeting was scheduled to discuss the discrepancies until agreement was reached. NVivo (version 12; QSR International Pty Ltd, now known as Lumivero) was used to analyze free-text data [43].

Ethical Considerations

The study was approved by the North West–Greater Manchester West Research Ethics Committee (reference: 22/NW/0246). All participants had the option to enter a £50 (US $67.40) prize draw at the end of the survey as compensation for their time. A participant information sheet was embedded in the survey for potential participants to read.

Results

Demographics

In total, 309 participants completed the survey. The median age was 38.5 (IQR 28 - 50) years. There was an approximately even gender split (male: n=145, 47% and female: n=153, 49%). Almost all participants reported their gender was the same as their sex assigned at birth (n=298, 96.4%). Most participants were White (n=244, 79%), single (n=173, 56%), and had no caring responsibilities (n=213, 68.9%). A third of participants were employed (n=99, 32%), over half were living with family (n=159, 51.5%), and nearly three-quarters had received further education or higher (n=223, 72.2%). Full sample demographic characteristics are shown in Table 1.

XSL•FO
RenderX

Table . Baseline characteristics and service use (N=309).

Characteristic	Value
Age (years), median (IQR)	38.5 (28-50)
Sex, n (%)	
Female	153 (49.51)
Male	145 (46.93)
Nonbinary or third gender	4 (1.29)
Other	3 (0.97)
Prefer not to say or unsure	1 (0.32)
Missing	3 (0.97)
Ethnicity, n (%)	
Asian or Asian British	22 (7.11)
Black, Black British, Caribbean, or African	18 (5.82)
White	244 (78.96)
Other ethnic group	3 (0.97)
Mixed or multiple ethnic groups	15 (4.85)
Prefer not to say	1 (0.32)
Missing	6 (1.94)
Relationship, n (%)	
Single	173 (55.99)
Cohabiting	14 (4.53)
In a relationship	41 (13.27)
Married	48 (15.53)
Divorced	11 (3.56)
Widowed	3 (0.97)
Other	1 (0.32)
Prefer not to say	10 (3.24)
Missing	8 (2.59)
Job, n (%)	
Employed	99 (32.04)
Full-time parent or carer	3 (0.97)
Homemaker	2 (0.65)
Out of work and looking for work	16 (5.18)
Out of work and not looking for work	16 (5.18)
Retired	17 (5.50)
Student	22 (7.12)
Unable to work	43 (13.92)
Unemployed	61 (19.74)
Voluntary work	15 (4.85)
Prefer not to say	3 (0.97)
Other	4 (1.29)
Missing	8 (2.59)
Education, n (%)	
Primary school	6 (1.94)

XSL•FO
RenderX

Characteristic	Value
Secondary school (up to GCSEs[a])	59 (19.09)
Further education (sixth form, college, or equivalent)	78 (25.24)
Trade, technical, or vocational training	11 (3.56)
Diploma or equivalent	38 (12.30)
University bachelor's degree	52 (16.83)
University master's degree	31 (10.03)
PhD or higher	13 (4.21)
Other	3 (0.97)
Prefer not to say	10 (3.24)
Missing	8 (2.59)
Living arrangements, n (%)	
Living alone	106 (34.30)
Living with family	159 (51.46)
Living with others (eg, friends)	11 (3.56)
Living in supported housing	13 (4.21)
Other	10 (3.24)
Prefer not to say	3 (0.97)
Missing	7 (2.27)
Care duties, n (%)	
No	213 (68.93)
Yes	77 (24.92)
Missing	19 (6.15)

[a]GCSE: General Certificate of Secondary Education.

Views on Collecting Data Using Active Monitoring and PS Methods

As shown in Table 2, a greater number of participants endorsed using ASM and PS for remote monitoring (ASM: n=145, 46.9% and PS: n=132, 42.7%) and sharing this information with their care team (ASM: n=144, 46.6% and PS: n=134, 43.4%) compared to those who opposed it. Nevertheless, many participants felt uncomfortable using these technologies for data collection (ASM: n=56, 18.1% and PS: n=73, 23.6%) and sharing the data with their care team (ASM: n=61, 19.7% and PS: n=60, 19.4%), and a notable number of participants found these technologies unhelpful (ASM: n=45, 14.6% and PS: n=47, 15.2%). For the questions reported in Table 2, a substantial proportion of participants reported "unsure" (range 18.5% - 21.7%) or did not answer the questions (range 14.6% - 15.5%).

XSL•FO
RenderX

Table . Comfort with and perceived helpfulness of digital monitoring via active symptom monitoring and passive sensing.

Items	Active symptom monitoring, n (%)	Passive sensing, n (%)
How comfortable do you feel with university researchers using it to collect information about your feelings, behaviors, general health, or whereabouts? (please tick to indicate)		
Comfortable	145 (46.93)	132 (42.72)
Uncomfortable	56 (18.12)	73 (23.62)
Unsure	63 (20.39)	57 (18.45)
Missing	45 (14.56)	47 (15.21)
How helpful would it be in helping you manage your mental health or well-being?		
Helpful	128 (41.42)	133 (43.04)
Unhelpful	45 (14.56)	47 (15.21)
Unsure	91 (29.45)	82 (26.54)
Missing	45 (14.56)	47 (15.21)
Would you be comfortable with information about whether you have been sleeping, exercising, traveling, or phoning people being collected via…		
Smartphone	37 (11.97)	37 (11.97)
Wearable device	34 (11)	25 (8.09)
Both	126 (40.78)	115 (37.22)
Neither	67 (21.68)	85 (27.51)
Missing	45 (14.56)	47 (15.21)
How comfortable do you feel with information about whether you have been sleeping, exercising, traveling, or phoning people being collected and shared with a member of your clinical or mental health team?		
Comfortable	144 (46.6)	134 (43.37)
Uncomfortable	61 (19.74)	60 (19.42)
Unsure	59 (19.09)	67 (21.68)
Missing	45 (14.56)	48 (15.53)

Figure 1 shows participants' responses regarding the acceptability of using a smartphone or wearable to collect 20 specific types of information. Only 4 out of the 20 data types or features received over 50% endorsement by participants. Specifically, heart rate (55%), fears of deteriorating (55%), sleep patterns (53%), and mood (53%) were endorsed ("agree" or "strongly agree") by participants. PCA was performed to further explore participants' comfort levels with different categories of data. Although principal components 1 and 2 explained most of the variability (66.6% and 11.1%, respectively), the clusters on the biplot of these 2 components were not clear (refer to Figure S1 in Multimedia Appendix 1). Instead, we used the biplot of principal component 2 and principal component 3 (which explained 4.3% of the variability) to guide the clustering (refer to Figure S2 in Multimedia Appendix 1). Features were clustered into 5 data types: physical health information (eg, heart rate), mental health information (eg, mood), personal information (eg, location), environmental information (eg, weather), and nonpersonal device information (eg, battery level). The Kruskal-Wallis test showed that participants' comfort levels differed significantly between data-type groups (P<.001). Dunn post hoc test for pairwise comparisons showed that participants were significantly less comfortable sharing personal information compared to all other data types (P<.001). Participants felt significantly more comfortable sharing mental health and physical health data than for all other data types (P<.05), except for environmental data (no significant difference between mental health data and environmental data; P≥.99). There was no significant difference in comfort levels between sharing mental health data and physical health data (P=.24), nor between environmental data and nonpersonal device data (P=.57; refer to Table 3).

XSL•FO
RenderX

Figure 1. Acceptability of collecting specific types of information using a smartphone or wearable device.

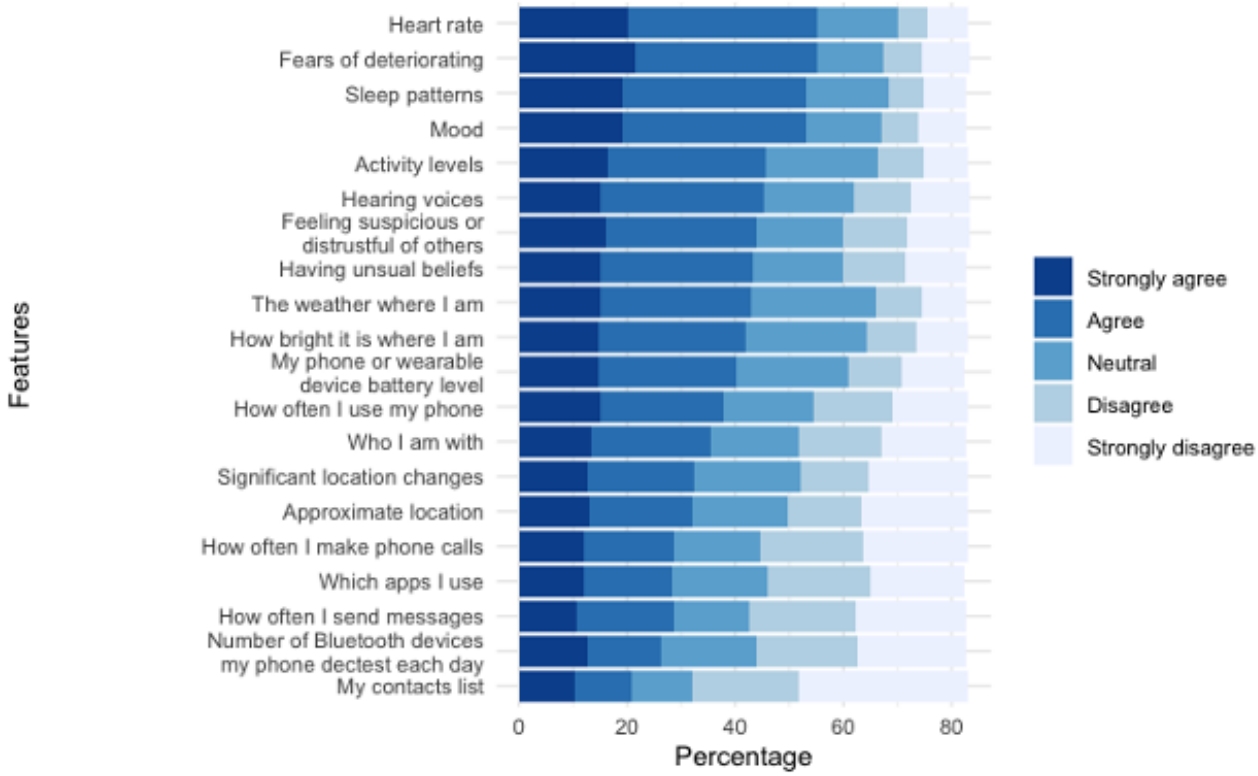


Table . Pairwise comparison of participants' comfort level with sharing different types of data with researchers.

Comparison	*z* value	Unadjusted *P* value	Adjusted *P* value[a]
Environmental information–nonpersonal device information	1.91	.06	.57
Environmental information–mental health information	−1.17	.24	≥.99
Nonpersonal device information–mental health information	−3.45	.001	.006
Environmental information–personal information	8.08	<.001	<.001
Nonpersonal device information–personal information	5.67	<.001	<.001
Mental health information–personal information	12.94	<.001	<.001
Environmental information–physical health information	−2.87	.004	.04
Nonpersonal device information–physical health information	−4.96	<.001	<.001
Mental health information–physical health information	−2.25	.02	.24
Personal information–physical health information	−13.31	<.001	<.001

[a]*P* values were adjusted for multiple comparisons using the Bonferroni correction.

Participants were also asked to share any other concerns not listed in the survey in a free-text question (refer to Tables S2 and S3 in Multimedia Appendix 1). The top concerns about ASM were "risk of data misuse" (29/91 free-text responses, 32%), "feeling uncomfortable with being monitored" (19/91, 21%), and "tracking symptoms making people feel worse" (14/91, 15%). For PS, participants expressed concerns such as "feeling uncomfortable with passive sensing" (19/57 free-text responses, 33%), "it is invasive and unhelpful" (13/57, 23%),

XSL•FO
RenderX

and "preferring to use it collaboratively rather than being monitored by a clinician" (9/57, 16%).

Views on Using Smartphones and Wearable Devices in the Context of Mental Health Care

The concept of using an app or a wearable device to support mental health between appointments was highly acceptable. A notable 59.9% (n=185) were willing to use an app, while an even higher 65% (n=201) expressed interest in using a wearable device. When it comes to integrating apps or wearables into mental health management, the most favored option was a blended approach of either an app (n=124, 51.7%) or a wearable (n=102, 41.1%) combined with face-to-face care. Moreover, the preferred method for receiving prompts to complete symptom monitoring tasks was a combination of prescheduled prompts and self-initiated responses, rather than relying solely on either method. Participants preferred this approach for both smartphone apps (n=86, 35.8%) and wearable devices (n=74, 29.8%).

Regarding the perceived impact of using apps and wearables, more than half of the participants reported that using such technologies would have no negative impact on their mood (apps: n=135, 56.3% and wearables: n=140, 56.5%) or psychotic symptoms (apps: n=117, 48.8% and wearables: n=144, 58.1%). However, when asked about the positive impact, half of the participants expressed a neutral stance on the statement that using a smartphone app (n=121, 50.4%) or a wearable device (n=128, 51.6%) makes them feel happy. Over half (n=130, 54.2%) reported that using smartphone apps could help them feel connected with others, whereas slightly fewer (n=122, 49.2%) participants reported this for wearable devices. In terms of impacts on overall mental health and well-being, 40.8% (n=98) and 37.9% (n=94) either "agreed" or "strongly agreed" that using smartphone apps and wearable devices, respectively, could help their overall mental health and well-being. A small proportion of participants considered using smartphone apps (n=48, 20%) or wearable devices (n=50, 20.2%) to be unhelpful for overall mental health and well-being.

As shown in Table S4 in Multimedia Appendix 1, the ranking of advantages, based on the total percentage of participants who reported "agree" or "strongly agree," was consistent across both smartphones and wearables. Top advantages were the ability to be used at any time, in any location (app: 189/240, 78.8% participants and wearables: 193/248, 77.8% participants), sharing information in real time with the care team (app: 180/240, 75% participants and wearables: 180/248, 72.6% participants), the ability to identify triggers and patterns (app: 171/240, 71.3% participants and wearables: 172/248, 69.4% participants), and being able to record and reflect on symptoms and experiences over time (app: 176/240, 73.3% participants and wearables: 168/248, 67.7% participants). Additional advantages specific to smartphone apps are shown in Table S5 in Multimedia Appendix 1. When participants were asked the extent to which barriers might impact their likelihood of using smartphone apps or wearable devices for mental health care (refer to Table S6 in Multimedia Appendix 1), the primary concern for apps was that they might be used as an excuse to replace face-to-face care (141/240, 58.8% participants). In contrast, cost was the main barrier for wearables (144/248, 58.1% participants).

Ownership and Usage of Smartphones and Wearable Devices

Of the 309 participants, 276 (89.3%) owned or had access to a smartphone; in contrast, 95 (30.7%) participants owned or had access to a wearable device. The percentages of iPhone (n=138, 44.7%) and Android (Google LLC; n=141, 45.6%) smartphone users were comparable. For participants using a wearable device, Apple Watch was the most used device (n=32, 10.4%), followed by Fitbit (n=28, 9.1%) and Samsung (n=16, 5.2%). Most participants reported that they faced no barriers to owning or using a smartphone (n=191, 61.8%). For those who did encounter barriers, paranoia or suspiciousness related to technology (n=37, 12%), affordability (n=32, 10.4%), and digital literacy (n=27, 8.7%) were the top reasons. Most participants (n=215, 69.6%) identified barriers to owning or using a wearable device. The primary barrier was a lack of interest (n=94, 30.4%), followed by affordability (n=82, 26.5%) and a perceived lack of need to use it (n=45, 14.6%). In terms of frequency of use, smartphones and smartphone apps were the most frequently used digital tools, with 77% (n=238) and 60.2% (n=186) of participants, respectively, reporting using these multiple times a day. In contrast, in line with the ownership rates, 60.5% (n=187) stated they did not use wearable devices. Nearly half of the participants (n=147, 47.6%) had used health-related apps. Among those, physical health apps were the most used (n=99, 32%), while just over a quarter (n=80, 25.9%) reported using mental health apps.

Regarding participants' preferences for the appearance of wearable devices, we selected Fitbit as an example of a consumer-grade device and the Empatica E4 as an example of a research-grade device. These 2 devices were chosen because they are frequently used in digital mental health research [9]. More participants chose the Fitbit (n=168, 54.4% participants) with a black strap (n=106, 50.7% participants) due to its more stylish design compared to the Empatica E4 (n=41, 13.3% participants), Fitbit with steel blue strap (n=52, 24.9% participants), and Fitbit with lunar white strap (n=49, 23.4% participants). Regardless of which type of device they selected, most participants were willing to use it daily (E4: 25/41, 61% participants and Fitbit: 121/168, 72% participants) and for longer than 1 year (E4: 21/41, 51.2% participants and Fitbit: 87/168, 51.8% participants). Participants identified the most important features of wearable devices as privacy (n=186, 89% participants), physical comfort (n=185, 88.5% participants), reasonable price (n=181, 86.6% participants), long battery life (n=181, 86.6% participants), and ease of use (n=174, 83.3% participants). A full list of features, ranked by importance, is shown in Table S7 in Multimedia Appendix 1.

Discussion

Principal Findings

We conducted a one-off survey with a large sample of individuals in the United Kingdom who self-reported a diagnosis of schizophrenia-spectrum disorder, were taking antipsychotic

XSL•FO
RenderX

medication, or were experiencing psychotic symptoms, to gather their perspectives on DRM of symptoms, including ASM and PS, using smartphone apps and wearable devices. Both smartphone apps (59.9%) and wearable devices (65%) were considered acceptable tools for digital monitoring in mental health care. Most participants preferred a blended approach that combined digital tools with in-person support and were willing to use wearable devices daily and for longer than a year to manage their mental health. Nevertheless, although more participants accepted than rejected the use of ASM and PS technologies for monitoring than those who opposed it, the acceptance rate was still below 50%. This suggests that user-level barriers must be addressed to facilitate the adoption of these methods in mental health care. Consistent with previous studies [17,18,44], smartphone ownership rates were high, and smartphones were frequently used, with participants reporting few barriers to owning or using one. In contrast, wearables were less commonly owned, less frequently used, and participants faced more barriers (eg, cost) to owning or using them. To further investigate the persistent barriers to accessibility and acceptability of digital interventions, qualitative studies, or the incorporation of qualitative interviews alongside quantitative surveys in mixed methods research would be valuable for elucidating nuanced contextual factors that may influence uptake.

Our sample was predominantly White (79%), with fewer participants from ethnic minority groups (eg, Black, 5.8%), which is less representative of the ethnic diversity typically observed in populations experiencing psychosis [45]. Additionally, 32% of participants were employed, which is higher than the employment rate reported in the broader psychosis population, which ranges from 1% to 18.5% [46]. Although the gender distribution in our sample was relatively balanced, we recruited slightly more females than males (49.5% vs 46.9%), diverging from the general trend in psychosis incidence, which is typically higher among males [47]. These differences suggest the need for further research to better understand the views of underrepresented groups on the use of personal digital devices for remote symptom monitoring.

Participants expressed varying levels of comfort with sharing different types of data with a research team. The most acceptable data types to be collected from a DRM tool were “physical health data” and “mental health data,” such as heart rate, fears of deteriorating, sleep patterns, and mood. Conversely, the least acceptable types of data were “personal information,” including contact lists, frequency of sending messages, and app usage. This is in line with a previous survey study on the general population’s comfort levels with sharing PS data for assessing depression and anxiety [40]. The study, which examined preferences for data sharing with clinicians, electronic health records (EHRs), and family members, found that participants were significantly more comfortable sharing health-related information (physical activity, mood, and sleep) than personal information (communication logs, location, and social activity) across all 3 recipient groups [40]. One explanation is that people might be more comfortable sharing health-related data because they see it as clinically relevant, whereas passive monitoring of personal behaviors (eg, communication and location) may feel more intrusive and less clinically relevant. Additionally, free-text responses in our survey study revealed that some participants were concerned about data misuse, disliked being monitored, and felt that monitoring was invasive and unhelpful. Taken together, the findings also indicate that privacy concerns remain a significant issue for people and pose a barrier to implementing digital monitoring in clinical practice. Some researchers have argued that training and explanation of the purpose of tracking to end users may reduce concerns [48]. However, our findings suggest that building trust may be more difficult than simply explaining the nonharmful nature of data tracking. Additionally, the complexity of methods such as machine learning, which is used to analyze large volumes of data collected from digital tools, makes it challenging even for clinicians to fully understand how the data will be used [49]. As the field moves toward collecting large-scale data from diverse populations, building trust between service users and digital systems is crucial.

Regarding common concerns about the potential negative effects of digital technologies on mental health, over half of the participants reported that using apps and wearables did not worsen their mood or psychotic symptoms. More participants found these technologies helpful for mental health and well-being rather than unhelpful and found that apps and wearables fostered a sense of connection, though a substantial proportion did not endorse such statements. These findings are consistent with a previous survey study [18], highlighting the complex relationship between technology use and mental health. Furthermore, serious negative effects, such as adverse events, remain an area of uncertainty. A recent review noted that while adverse events are rarely recorded and reported in DMHI trials, it is unclear whether this is due to a lack of occurrence or inadequate reporting [50]. Therefore, continued research is needed to better understand and monitor potential risks, ensuring the safety of digital health technologies in different populations.

Many participants had not used any health-related apps (42.4%). This finding echoes a previous US survey, which found that service users with psychosis rarely used their digital devices for health-related purposes, such as mental health symptom monitoring (14%) and sleep tracking (14%) [51]. Additionally, we found that for participants who had used a health-related app, physical health apps were more commonly used than mental health apps. Similarly, previous studies have identified high levels of interest in using physical health apps among young people with psychosis [52,53]. Given the increased risk of physical health conditions in individuals with severe mental health problems [54], and the frequent neglect of these issues in current mental health services [54,55], integrating both mental and physical health features in the same app may be preferred by service users. Moreover, most participants preferred DRM tools to be blended with face-to-face care. This is consistent with recent research highlighting the growing emphasis on blended care models, where digital tools complement traditional in-person services to enhance accessibility, engagement, and continuity of mental health support [56-58].

Despite the high rates of device ownership, a substantial proportion of participants reported that they faced barriers to owning or using smartphones and wearables. Specifically, for

XSL•FO
RenderX

smartphones, paranoia or suspiciousness related to technology, affordability, and digital literacy were the top reasons, whereas for wearables, top barriers were a lack of interest, affordability, and a perceived lack of need to use them. These findings suggest that wearables may be perceived as less essential for daily life compared to smartphones. However, affordability was reported as a barrier for both technologies, suggesting that cost limits access to digital technology for individuals with psychosis. A recent systematic review indicated that providing supportive infrastructure to users, including devices and free internet access, is a potential solution for advancing digital health equity [59]. For example, in the United Kingdom, the NHS could directly provide devices through its funding, as using personal health budgets for device provision is feasible in NHS City and Hackney [60]. Therefore, device provision in standard practice should be considered if digital mental health becomes an essential component of mental health care.

Strengths and Limitations

We reached a large, geographically diverse sample, enhancing the robustness and generalizability of our findings. We explicitly asked individuals who experience psychosis about their perspectives on PS, a rapidly developing subfield of digital mental health that has been largely unexplored in previous research. Regarding limitations, we relied on participants to self-report their diagnosis or experience of psychosis without independent validation. Also, this self-selected sample may have been more familiar, interested, and comfortable with using digital tools than the wider clinical population, as individuals who choose to participate in research are often more engaged with services. Moreover, the missing data rate of some questions was substantial, ranging from 1.3% to 15.5%, which can affect the reliability and validity of the findings. Additionally, although the survey included some free-text items, most data were quantitative, collected through structured response formats. As such, the study may not capture the depth or nuance that qualitative methods typically provide when exploring service user views in detail. Finally, our sample lacked ethnic diversity, with most participants being White, and almost all participants were in contact with mental health services. The findings may not be generalizable to individuals from diverse backgrounds, who may have different experiences, preferences, and barriers related to digital health technologies. Future research should aim to include more diverse samples to ensure that DMHIs are equitable and accessible to all service users.

Conclusions and Clinical Implications

Given that certain types of personal information (eg, contact lists, message frequency, and app usage) are perceived as less acceptable for collection, DRM should prioritize transparent data governance, user consent, and customizable privacy settings to build trust and alleviate concerns about surveillance and data misuse. Developers need to adhere to regulatory standards, such as the Digital Technology Assessment Criteria [61], throughout the process of designing and developing DRM tools. Additionally, service users should have control over the data they share, with whom, and for what purpose, for instance, through implementing digital dashboards to present key health metrics to service users and allow for their active participation in care management and decisions over data sharing with clinical teams. Alongside this, clinicians should also play a role in explaining the purpose of DRM (what data is being collected, why, and what it will be used for), emphasizing its clinical benefits and safety profile while addressing potential privacy-related apprehensions. Although most participants did not perceive DRM as harming mental health, a substantial proportion expressed concerns. Therefore, clinicians should proactively assess potential unintended effects, such as increased anxiety, paranoia, or distress related to feeling monitored, and implement strategies to ensure the safe and supportive use of these technologies. The higher reported usage of physical health apps compared to mental health apps highlights the potential for developing integrated platforms that address both mental and physical health needs. Given the high prevalence of physical health problems in psychosis, combining mental health tracking with physical health features (eg, activity tracking, sleep monitoring, and medication reminders) could enhance user engagement and support holistic care. To ensure the practical integration of such platforms into existing clinical workflows, it is essential to leverage established digital infrastructures, particularly EHRs. Integrating digital health apps with EHRs would allow clinicians to seamlessly access and interpret integrated data on both physical and mental health, supporting comprehensive and person-centered care. Adopting interoperability standards would further support data exchange across health care settings, enabling the holistic management of physical and mental health within routine practice. By incorporating these considerations, integrated platforms have the potential to significantly enhance service user outcomes by enabling the early detection of both psychiatric symptom deterioration and cardiometabolic risks, thereby facilitating more personalized and timely interventions. Blended approaches, using DRM alongside face-to-face care, should be considered as a priority for clinical implementation, as it appears to be participants' preference. Finally, should digital technologies become important in improving clinical care, they should be provided at no cost to service users, especially wearables, as cost appears to be a main barrier to owning or using this technology.

Acknowledgments

The authors would like to thank the participants who took part in the survey study, as well as the health care professionals and researchers who supported the study. We also thank individuals within the respective National Health Service (NHS) services for their assistance. No AI tools were used in the preparation of this manuscript.

XSL•FO
RenderX

Funding

This project was funded by a research professorship (NIHR300794) from the National Institute for Health and Care Research (NIHR), awarded to author SB.

Data Availability

The datasets used and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

XZ contributed to writing the original draft, reviewing and editing, as well as visualization, formal analysis, investigation, data curation, and project administration. EE contributed to writing review, editing, conceptualization, and formal analysis. DDB contributed to writing review, editing, investigation, and project administration. CR contributed to writing review, editing, conceptualization, investigation, and project administration. JF contributed to writing the review, editing, and conceptualization. SB contributed to writing review, editing, validation, supervision, resources, project administration, conceptualization, investigation, funding acquisition, and data curation. SB is the guarantor.

Conflicts of Interest

SB is a co-founder of CareLoop Health Ltd, a spin-out company from The University of Manchester that develops and markets digital therapeutics for schizophrenia and a digital screening app for postnatal depression. JF is supported by a UK Research and Innovation Future Leaders Fellowship (MR/Y033876/1) and has provided consultancy and/or advisory services to Atheneum, Bayer, ParachuteBH Ltd, LLMental, HedoniaUSA, and Arthur D Little, independent of this work.

Multimedia Appendix 1
Service user survey and additional analyses.
[DOCX File, 1397 KB - humanfactors_v13i1e86152_app1.docx]

References

1. Andrade LH, Alonso J, Mneimneh Z, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol Med 2014 Apr;44(6):1303-1317. [doi: 10.1017/S0033291713001943] [Medline: 23931656]
2. Two-fifths of patients waiting for mental health treatment forced to resort to emergency or crisis services. RCPsych. 2020. URL: https://www.rcpsych.ac.uk/news-and-features/latest-news/detail/2020/10/06/two-fifths-of-patients-waiting-for-mental-health-treatment-forced-to-resort-to-emergency-or-crisis-services [accessed 2022-10-21]
3. Ben-Zeev D, Brian R, Wang R, et al. CrossCheck: integrating self-report, behavioral sensing, and smartphone use to identify digital indicators of psychotic relapse. Psychiatr Rehabil J 2017 Sep;40(3):266-275. [doi: 10.1037/prj0000243] [Medline: 28368138]
4. Insel TR. Digital phenotyping: a global tool for psychiatry. World Psychiatry 2018 Oct;17(3):276-277. [doi: 10.1002/wps.20550] [Medline: 30192103]
5. Smith KA, Blease C, Faurholt-Jepsen M, et al. Digital mental health: challenges and next steps. BMJ Ment Health 2023 Feb;26(1):e300670. [doi: 10.1136/bmjment-2023-300670]
6. Torous J, Wisniewski H, Bird B, et al. Creating a digital health smartphone app and digital phenotyping platform for mental health and diverse healthcare needs: an interdisciplinary and collaborative approach. J Technol Behav Sci 2019 Jun;4(2):73-85. [doi: 10.1007/s41347-019-00095-w]
7. Torous J, Bucci S, Bell IH, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct;20(3):318-335. [doi: 10.1002/wps.20883] [Medline: 34505369]
8. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 May 8;13(1):23-47. [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
9. Bladon S, Eisner E, Bucci S, et al. A systematic review of passive data for remote monitoring in psychosis and schizophrenia. npj Digit Med 2025 Jan 27;8(1). [doi: 10.1038/s41746-025-01451-2]
10. Lane E, D'Arcey J, Kidd S, et al. Digital phenotyping in adults with schizophrenia: a narrative review. Curr Psychiatry Rep 2023 Nov;25(11):699-706. [doi: 10.1007/s11920-023-01467-z] [Medline: 37861979]
11. Henson P, Barnett I, Keshavan M, Torous J. Towards clinically actionable digital phenotyping targets in schizophrenia. npj Schizophr 2020 May 5;6(1). [doi: 10.1038/s41537-020-0100-1]
12. Depp CA, Bashem J, Moore RC, et al. GPS mobility as a digital biomarker of negative symptoms in schizophrenia: a case control study. npj Digit Med ;2(1). [doi: 10.1038/s41746-019-0182-1]

XSL•FO
RenderX

13. Meyer N, Joyce DW, Karr C, et al. The temporal dynamics of sleep disturbance and psychopathology in psychosis: a digital sampling study. Psychol Med 2022 Oct;52(13):2741-2750. [doi: 10.1017/S0033291720004857] [Medline: 33431090]
14. Gleeson JF, McGuckian TB, Fernandez DK, et al. Systematic review of early warning signs of relapse and behavioural antecedents of symptom worsening in people living with schizophrenia spectrum disorders. Clin Psychol Rev 2024 Feb;107:102357. [doi: 10.1016/j.cpr.2023.102357] [Medline: 38065010]
15. Watson A, Mellotte H, Hardy A, Peters E, Keen N, Kane F. The digital divide: factors impacting on uptake of remote therapy in a South London psychological therapy service for people with psychosis. J Ment Health 2022 Nov 2;31(6):825-832. [doi: 10.1080/09638237.2021.1952955]
16. Luther L, Buck BE, Fischer MA, Johnson-Kwochka AV, Coffin G, Salyers MP. Examining potential barriers to mHealth implementation and engagement in schizophrenia: phone ownership and symptom severity. J Technol Behav Sci 2022 Mar;7(1):13-22. [doi: 10.1007/s41347-020-00162-7] [Medline: 35434267]
17. Zhang X, Lewis S, Chen X, Berry N, Bucci S. Technology use and attitudes towards digital mental health in people with severe mental health problems: a survey study in China. Front Psychiatry 2023;14. [doi: 10.3389/fpsyt.2023.1261795]
18. Eisner E, Berry N, Bucci S. Digital tools to support mental health: a survey study in psychosis. BMC Psychiatry 2023 Dec;23(1). [doi: 10.1186/s12888-023-05114-y]
19. Bucci S, Schwannauer M, Berry N. The digital revolution and its impact on mental health care. Psychol Psychother 2019 Jun;92(2):277-297. [doi: 10.1111/papt.12222] [Medline: 30924316]
20. Eisner E, Faulkner S, Allan S, et al. Barriers and facilitators of user engagement with digital mental health interventions for people with psychosis or bipolar disorder: systematic review and best-fit framework synthesis. JMIR Ment Health 2025 Jan 20;12:e65246. [doi: 10.2196/65246] [Medline: 39832352]
21. Arnold C, Villagonzalo KA, Meyer D, et al. Predicting engagement with an online psychosocial intervention for psychosis: exploring individual- and intervention-level predictors. Internet Interv 2019 Dec;18:100266. [doi: 10.1016/j.invent.2019.100266] [Medline: 31890619]
22. Achtyes ED, Ben-Zeev D, Luo Z, et al. Off-hours use of a smartphone intervention to extend support for individuals with schizophrenia spectrum disorders recently discharged from a psychiatric hospital. Schizophr Res 2019 Apr;206:200-208. [doi: 10.1016/j.schres.2018.11.026] [Medline: 30551981]
23. Almeida R, Sousa TJ, Couto AS, Marques AJ, Queirós CM, Martins CL. Development of weCope, a mobile app for illness self-management in schizophrenia. Arch Clin Psychiatry (São Paulo) 2019;46(1):1-4. [doi: 10.1590/0101-60830000000182]
24. Lopez-Morinigo JD, Barrigón ML, Porras-Segovia A, et al. Use of ecological momentary assessment through a passive smartphone-based app (eB2) by patients with schizophrenia: acceptability study. J Med Internet Res 2021 Jul 26;23(7):e26548. [doi: 10.2196/26548] [Medline: 34309576]
25. Ben-Zeev D, Scherer EA, Gottlieb JD, et al. mHealth for schizophrenia: patient engagement with a mobile phone intervention following hospital discharge. JMIR Ment Health 2016 Jul 27;3(3):e34. [doi: 10.2196/mental.6348] [Medline: 27465803]
26. Torous J, Friedman R, Keshavan M. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. JMIR mHealth uHealth 2014 Jan 21;2(1):e2. [doi: 10.2196/mhealth.2994] [Medline: 25098314]
27. Torous J, Chan SR, Yee-Marie Tan S, et al. Patient smartphone ownership and interest in mobile apps to monitor symptoms of mental health conditions: a survey in four geographically distinct psychiatric clinics. JMIR Ment Health ;1(1):e5. [doi: 10.2196/mental.4004]
28. Miller BJ, Stewart A, Schrimsher J, Peeples D, Buckley PF. How connected are people with schizophrenia? Cell phone, computer, email, and social media use. Psychiatry Res 2015 Feb;225(3):458-463. [doi: 10.1016/j.psychres.2014.11.067]
29. Naslund JA, Aschbrenner KA, Bartels SJ. How people with serious mental illness use smartphones, mobile apps, and social media. Psychiatr Rehabil J 2016;39(4):364-367. [doi: 10.1037/prj0000207]
30. Robotham D, Satkunanathan S, Doughty L, Wykes T. Do we still have a digital divide in mental health? A five-year survey follow-up. J Med Internet Res 2016 Nov 22;18(11):e309. [doi: 10.2196/jmir.6511] [Medline: 27876684]
31. Lal S, Dell'Elce J, Tucci N, Fuhrer R, Tamblyn R, Malla A. Preferences of young adults with first-episode psychosis for receiving specialized mental health services using technology: a survey study. JMIR Ment Health 2015;2(2):e18. [doi: 10.2196/mental.4400] [Medline: 26543922]
32. Gay K, Torous J, Joseph A, Pandya A, Duckworth K. Digital technology use among individuals with schizophrenia: results of an online survey. JMIR Ment Health 2016 May 4;3(2):e15. [doi: 10.2196/mental.5379] [Medline: 27146094]
33. Nicholas J, Boydell K, Christensen H. Beyond symptom monitoring: consumer needs for bipolar disorder self-management using smartphones. Eur Psychiatry 2017 Jul;44:210-216. [doi: 10.1016/j.eurpsy.2017.05.023] [Medline: 28692910]
34. Matthews M, Murnane E, Snyder J, et al. The double-edged sword: a mixed methods study of the interplay between bipolar disorder and technology use. Comput Human Behav 2017 Oct;75:288-300. [doi: 10.1016/j.chb.2017.05.009]
35. Bogart K, Wong SK, Lewis C, et al. Mobile phone text message reminders of antipsychotic medication: is it time and who should receive them? A cross-sectional trust-wide survey of psychiatric inpatients. BMC Psychiatry 2014 Jan 22;14(1):15. [doi: 10.1186/1471-244X-14-15] [Medline: 24447428]
36. Jain N, Singh H, Koolwal GD, Kumar S, Gupta A. Opportunities and barriers in service delivery through mobile phones (mHealth) for severe mental illnesses in Rajasthan, India: a multi-site study. Asian J Psychiatr 2015 Apr;14:31-35. [doi: 10.1016/j.ajp.2015.01.008] [Medline: 25701069]

XSL•FO
RenderX

37. Ben-Zeev D, Davis KE, Kaiser S, Krzsos I, Drake RE. Mobile technologies among people with serious mental illness: opportunities for future services. Adm Policy Ment Health 2013 Jul;40(4):340-343. [doi: 10.1007/s10488-012-0424-x] [Medline: 22648635]
38. Homepage. Qualtrics. URL: https://www.qualtrics.com [accessed 2026-04-13]
39. The R project for statistical computing. The R Foundation. URL: https://www.R-project.org/ [accessed 2026-04-13]
40. Nicholas J, Shilton K, Schueller SM, Gray EL, Kwasny MJ, Mohr DC. The role of data type and recipient in individuals' perspectives on sharing passively collected smartphone data for mental health: cross-sectional questionnaire study. JMIR Mhealth Uhealth 2019 Apr 5;7(4):e12578. [doi: 10.2196/12578] [Medline: 30950799]
41. Microsoft Excel. Microsoft. URL: https://office.microsoft.com/excel [accessed 2026-04-13]
42. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
43. Ask more from your data with the depth and power of NVivo. Lumivero. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2026-04-13]
44. Bell IH, Thompson A, Valentine L, Adams S, Alvarez-Jimenez M, Nicholas J. Ownership, use of, and interest in digital mental health technologies among clinicians and young people across a spectrum of clinical care needs: cross-sectional survey. JMIR Ment Health 2022 May 11;9(5):e30716. [doi: 10.2196/30716] [Medline: 35544295]
45. Fearon P, Kirkbride JB, Morgan C, et al. Incidence of schizophrenia and other psychoses in ethnic minority groups: results from the MRC AESOP Study. Psychol Med 2006 Nov;36(11):1541-1550. [doi: 10.1017/S0033291706008774] [Medline: 16938150]
46. Psychosis data report. : Public Health England; 2016 Sep URL: https://www.gov.uk/government/publications/psychosis-data-report [accessed 2026-05-01]
47. Kirkbride JB, Errazuriz A, Croudace TJ, et al. Incidence of schizophrenia and other psychoses in England, 1950-2009: a systematic review and meta-analyses. PLOS ONE 2012;7(3):e31660. [doi: 10.1371/journal.pone.0031660] [Medline: 22457710]
48. Rodriguez-Villa E, Rozatkar AR, Kumar M, et al. Cross cultural and global uses of a digital mental health app: results of focus groups with clinicians, patients and family members in India and the United States. Glob Ment Health (Camb) 2021;8:e30. [doi: 10.1017/gmh.2021.28] [Medline: 34512999]
49. Zhang X, Eisner E, Di Basilio D, Richardson C, Firth J, Bucci S. Healthcare professionals' views on implementing digital health tools in psychosis: A national survey in the UK. BMJ Digit Health 2025 Jun;1(1):e000005. [doi: 10.1136/bmjdhai-2025-000005]
50. Allan S, Ward T, Eisner E, et al. Adverse events reporting in digital interventions evaluations for psychosis: a systematic literature search and individual level content analysis of adverse event reports. Schizophr Bull 2024 Nov 8;50(6):1436-1455. [doi: 10.1093/schbul/sbae031] [Medline: 38581410]
51. Noel VA, Acquilano SC, Carpenter-Song E, Drake RE. Use of mobile and computer devices to support recovery in people with serious mental illness: survey study. JMIR Ment Health 2019 Feb 20;6(2):e12255. [doi: 10.2196/12255] [Medline: 30785401]
52. Naslund JA, Aschbrenner KA. Technology use and interest in digital apps for mental health promotion and lifestyle intervention among young adults with serious mental illness. J Affect Disord Rep 2021 Dec;6:100227. [doi: 10.1016/j.jadr.2021.100227] [Medline: 40027518]
53. Firth J, Sawyer C, Sainsbury J, et al. Using physical health apps to promote healthy lifestyles in youth mental healthcare: a nationwide perspective-gathering exercise of over 400 service users. Psychiatry Res 2024 Dec;342:116187. [doi: 10.1016/j.psychres.2024.116187] [Medline: 39305826]
54. Firth J, Siddiqi N, Koyanagi A, et al. The Lancet Psychiatry Commission: a blueprint for protecting physical health in people with mental illness. Lancet Psychiatry 2019 Aug;6(8):675-712. [doi: 10.1016/S2215-0366(19)30132-4]
55. Bailey JM, Bartlem KM, Wiggers JH, et al. Systematic review and meta-analysis of the provision of preventive care for modifiable chronic disease risk behaviours by mental health services. Prev Med Rep 2019 Dec;16:100969. [doi: 10.1016/j.pmedr.2019.100969]
56. Garety P, Ward T, Emsley R, et al. Effects of SlowMo, a blended digital therapy targeting reasoning, on paranoia among people with psychosis: a randomized clinical trial. JAMA Psychiatry 2021 Jul 1;78(7):714-725. [doi: 10.1001/jamapsychiatry.2021.0326] [Medline: 33825827]
57. Valentine L, McEnery C, Bell I, et al. Blended digital and face-to-face care for first-episode psychosis treatment in young people: qualitative study. JMIR Ment Health 2020 Jul 28;7(7):e18990. [doi: 10.2196/18990] [Medline: 32720904]
58. Gumley AI, Bradstreet S, Ainsworth J, et al. The EMPOWER blended digital intervention for relapse prevention in schizophrenia: a feasibility cluster randomised controlled trial in Scotland and Australia. Lancet Psychiatry 2022 Jun;9(6):477-486. [doi: 10.1016/S2215-0366(22)00103-1] [Medline: 35569503]
59. Wilson S, Tolley C, Mc Ardle R, et al. Recommendations to advance digital health equity: a systematic review of qualitative studies. npj Digit Med ;7(1):173. [doi: 10.1038/s41746-024-01177-7]
60. Blueprinting. NHS England. 2021. URL: https://transform.england.nhs.uk/key-tools-and-info/digital-playbooks/mental-health-digital-playbook/a-digital-recovery-platform-for-severe-mental-illness/ [accessed 2026-04-13]

XSL•FO
RenderX

61. Digital Technology Assessment Criteria (DTAC): guidance for buyers and suppliers. NHS England. 2021. URL: https://transform.england.nhs.uk/key-tools-and-info/digital-technology-assessment-criteria-dtac/ [accessed 2025-06-17]

Abbreviations

ASM: active symptom monitoring
DMHI: digital mental health intervention
DRM: digital remote monitoring
EHR: electronic health record
PCA: principal component analysis
PS: passive sensing

Edited by H Witteman; submitted 20.Oct.2025; peer-reviewed by A Priyam, J D'Arcey; revised version received 16.Mar.2026; accepted 20.Mar.2026; published 05.May.2026.

Please cite as:
Zhang X, Eisner E, Di Basilio D, Richardson C, Firth J, Bucci S
Service Users' Views on Digital Remote Monitoring for Psychosis: Survey Study
JMIR Hum Factors 2026;13:e86152
URL: https://humanfactors.jmir.org/2026/1/e86152
doi:10.2196/86152

© Xiaolong Zhang, Emily Eisner, Daniela Di Basilio, Cara Richardson, Joseph Firth, Sandra Bucci. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 5.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

The Impact of Tobacco Smoking on Treatment Response Among Patients With Psoriasis Undergoing Biologic Treatment: Prospective Observational Study

Fanlingzi Shen[1*], MD; Yuning Ding[1*], MD; Xiuqi Zhang[2*], MD; Quanruo Xu[3*], MD; Zhen Duan[3], MD; Ruiqi Cai[3], MD; Rui Zhang[1], MD; Xiangjin Gao[1], MD; Ruiping Wang[1,2,3], MD, PhD

[1]Clinical Research Center, Shanghai Skin Disease Hospital, 1278 Baode Road, Shanghai, Shanghai, China

[2]School of Medicine, Tongji University, Shanghai, Shanghai, China

[3]School of Public Health, Shanghai University of Traditional Chinese Medicine, Shanghai, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Ruiping Wang, MD, PhD
Clinical Research Center, Shanghai Skin Disease Hospital, 1278 Baode Road, Shanghai, Shanghai, China

Abstract

Background: Tobacco smoking is viewed as a behavioral risk factor for psoriasis initiation and progress, even among those undergoing biologic treatment. However, evidence regarding the association between tobacco smoking and treatment response to biologics among patients with psoriasis is limited.

Objective: This study aimed to explore the impact of tobacco smoking on the efficacy of biologic treatment in patients with psoriasis.

Methods: Patients with psoriasis undergoing biologic treatment were recruited from 2022 to 2024 at the Shanghai Skin Disease Hospital. Demographic characteristics and smoking habits were collected using a structured questionnaire. Clinical features and treatment efficacy were assessed and recorded by dermatologists at baseline and weeks 4, 8, 12, 24, and 48 after treatment, and the Psoriasis Area and Severity Index (PASI) 75 and PASI 90 measures were calculated for treatment efficacy evaluation.

Results: A total of 192 patients with psoriasis were included, of whom 78 (40.6%) were tobacco smokers, with a higher smoking prevalence observed in male patients (74/154, 48.1%). The PASI 75 response rates at weeks 4, 8, 12, 24, and 48 were 29.2% (56/192), 54.2% (104/192), 78.6% (151/192), 84.5% (153/181), and 82.7% (134/162), respectively. The PASI 90 response rates increased from 13.0% (25/192) at week 4 to 62.4% (113/181) at week 24 and 59.9% (97/162) at week 48. Logistic regression analysis indicated that nonsmoking patients with psoriasis had a high PASI 75 response rate. The adjusted odds ratios were 2.57 (95% CI 1.19 - 5.53), 2.61 (95% CI 1.34 - 5.08), 2.62 (95% CI 1.13 - 6.04), 2.27 (95% CI 0.89 - 5.75), and 2.75 (95% CI 1.01 - 7.49) at weeks 4, 8, 12, 24, and 48, respectively. Moreover, nonsmoking patients with psoriasis also had a higher PASI 90 response rate than those who smoked. The odds ratios ranged from 1.32 (95% CI 0.49 - 3.54) to 2.59 (95% CI 1.21 - 5.55). Correlation analysis showed that both tobacco smoking duration and daily cigarette consumption were negatively correlated with the reduction in PASI score at weeks 4 to 48 after treatment (P<.05).

Conclusions: Tobacco smoking was negatively associated with treatment response among patients with psoriasis undergoing biologic treatment, especially among patients with longer tobacco smoking duration and higher daily cigarette consumption.

(JMIR Hum Factors 2026;13:e90963) doi:10.2196/90963

KEYWORDS

psoriasis; tobacco smoking; biologics; treatment response; Psoriasis Area and Severity Index; PASI

Introduction

Psoriasis is a chronic and inflammatory skin disease, and most patients with psoriasis have some detriment to their quality of life [1]. Most patients with mild to moderate psoriasis can control their condition with topical medications or phototherapy, but these may not be sufficient for moderate to severe cases [2]. The emergence of biologics has brought about a major breakthrough in the treatment of psoriasis, particularly playing a positive role in patients with moderate to severe psoriasis. Biologics precisely target certain cytokines in inflammatory pathways during the pathological process of psoriasis, such as tumor necrosis factor, interleukin (IL)-17, and IL-23, to prevent inflammatory reactions and achieve efficient clearance of psoriasis lesions [3,4].

XSL•FO
RenderX

In recent years, guidelines on the treatment of psoriasis in Europe and the United States have shown growing emphasis and stronger recommendations for biologics [2,5]. The guidelines for the diagnosis and treatment of psoriasis in China also increasingly recommend biologics in clinical treatment, and the "Guidelines for the treatment of psoriasis with biologics and small-molecule drugs in China" were published in 2024, providing recommendations for clinicians on the use of biologic therapies [6,7]. With the increasing accessibility of biologics and support from medical insurance policies, more and more patients with psoriasis benefit from biologics. Previous studies have shown that biologics are efficacious in the treatment of psoriasis and are associated with greater improvements in quality of life compared with other treatment modalities [3,8,9]. However, some patients still fail to achieve satisfactory therapeutic responses in clinical practice, and factors such as genetics, obesity, tobacco smoking, alcohol consumption, and disease-related features (eg, disease duration and age at diagnosis) may contribute to the heterogeneity of responses to biologics [10,11].

Tobacco smoking is a widespread unhealthy lifestyle habit, with approximately 1.14 billion current smokers worldwide (approximately 30% in China), resulting in enormous health and economic consequences [12]. The impact of tobacco smoking on health has been extensively studied, and smoking is recognized as being closely associated with the development of various cardiovascular diseases, pulmonary diseases, and cancers [13-15]. The association between tobacco smoking and psoriasis is also receiving increasing attention. Previous studies have demonstrated that tobacco smoking elevates the risk of psoriasis and exacerbates disease severity, making the condition more difficult to control [16-18]. However, the impact of smoking on the efficacy of biologic treatment in patients with psoriasis remains inconclusive. Some studies have reported that smoking has a negative impact on the efficacy of biologic treatment for psoriasis [19-25], whereas other studies have failed to detect a significant effect of smoking on treatment outcomes [26-28]. Therefore, this study aimed to explore the impact of tobacco smoking on the efficacy of biologic treatment among patients with psoriasis in Shanghai, with the expectation of providing valuable insights for clinical practice regarding psoriasis treatment.

Methods

Study Population

A hospital-based prospective observational study was conducted in Shanghai Skin Disease Hospital from 2022 to 2024. Patients with psoriasis vulgaris aged ≥18 years, both male and female, and without a migration plan within a year were included and followed up on for 48 weeks, and patients who were unable to provide informed consent or who had neurological or psychiatric disorders were excluded. In this study, we extracted data from 192 patients with psoriasis who received biologic treatment with secukinumab.

Data Collection

Data were collected through a self-designed questionnaire administered by dermatologists after receiving uniform training. The main contents of the questionnaire were as follows: (1) demographic characteristics, including age, gender, educational level, monthly income, residence, marital status, and BMI; (2) information on tobacco smoking status, including smoking initiation age, smoking duration, and number of daily consumed cigarettes; and (3) information on psoriasis duration and severity. The severity of psoriasis was assessed by dermatologists using the Psoriasis Area and Severity Index (PASI), body surface area (BSA), and Physician Global Assessment (PGA) at baseline, week 4, week 8, week 12, week 24, and week 48.

Definition and Classification

In this study, the age of patients with psoriasis was classified as <35, 35 to 45, 46 to 60, and >60 years. Educational level was categorized into 4 groups (primary school or lower, junior high school, senior high school, and college or higher) based on the highest level of education completed. Monthly income (in CNY) was categorized into <¥3000, ¥3000 to ¥5000, ¥5001 to ¥10,000, and >¥10,000 (¥1=US $0.14). Residence was classified as urban and rural, and marital status was classified as married; unmarried; and divorced, widow or widower, or other. The BMI of patients with psoriasis was categorized as <23.9 kg/m^2 (lower or normal weight), 24.0 to 28.0 kg/m^2 (overweight), and >28.0 kg/m^2 (obesity).

A smoker was defined as a person who had smoked at least 100 cigarettes in their lifetime. Smoking initiation age was classified as <25 years and ≥25 years. Smoking duration was calculated as the age at the time of the study minus the smoking initiation age and then stratified into <20 years and ≥20 years. The number of daily consumed cigarettes was categorized as <20 and ≥20.

The PASI 75, PASI 90, and PGA (1/0) rate were calculated and used to evaluate the treatment response. The PASI 75 and PASI 90 were defined as patients with a reduction in PASI score of ≥75% and ≥90%, respectively, calculated using the formula [(PASI at baseline – PASI at week *t*)/PASI at baseline] × 100%. The PGA (1/0) rate referred to the proportion of patients with a PGA score of 1 or 0 among all patients.

Statistical Analysis

Data analysis was performed using the SAS software (version 9.4; SAS Institute). Descriptive statistics were presented as means and SDs or medians and IQRs as appropriate for quantitative variables and frequency counts and proportions for qualitative variables. The 2-tailed Student *t* test, Mann-Whitney *U* test, or Pearson chi-square test was performed to examine the statistical difference between groups. The temporal trend of treatment response (PASI 75, PASI 90, and PGA [1/0]) at different weeks among patients with psoriasis was evaluated using the Cochran - Armitage trend test. Logistic regression was applied to calculate the odds ratios (ORs) and 95% CIs to explore the association between tobacco smoking and treatment response among patients with psoriasis at weeks 4, 8, 12, 24, and 48. Scatterplots and quadratic polynomial regression were used to explore the association among smoking duration, daily cigarette consumption, and PASI score reduction. A *P* value of less than .05 (2 tailed) was considered statistically significant.

XSL•FO
RenderX

Ethical Considerations

This study was approved by the institutional review board of Shanghai Skin Disease Hospital (2022-25). Informed consent was obtained before starting the study, and the study was strictly performed in accordance with the Declaration of Helsinki. Patient participation in this study was voluntary, and patients could withdraw at any time. Participants' privacy and confidentiality were strictly protected. All data were deidentified before analysis and were accessible only to the researchers of this study. No compensation was provided to participants.

Results

Overview

A total of 192 patients with psoriasis undergoing biologic treatment were included in the final data analysis, of whom 154 (80.2%) were male. The average age of the patients was 41.2 (SD 18.2) years, with 41.1% (n=79) of patients aged >45 years. Among the patients, 49.0% (n=94) had a college or higher educational level, whereas 56.8% (n=109) reported monthly incomes exceeding ¥5000. Approximately 66.7% (n=128) of patients were urban residents, and 58.9% (n=113) of them were married. The average BMI of patients was 25.8 (SD 4.3) kg/m^2, and 65.1% (n=125) were overweight or obese. The median disease duration was 13.5 (IQR 7.0-21.0) years, and the median values of the PASI, BSA, and PGA scores at baseline (week 0) were 12.2 (IQR 9.8-18.6), 18.5 (IQR 12.0-33.3), and 3.0 (IQR 2.0-3.0), respectively. Data in Table 1 indicate that female patients were older (t_{190}=–2.64; P=.009) and had higher educational levels (χ^2_1=5.4; P=.02), higher incomes (χ^2_3=15.2; P=.002), and longer disease duration (z=–2.18; P=.03), and there was a higher proportion of married individuals in this group (χ^2_2=11.8; P=.003) but a lower proportion of overweight or obesity (χ^2_2=11.1; P=.004) and lower PGA scores than among male patients (z=–3.73; P<.001). The differences were all statistically significant (P<.05; Table 1).

XSL•FO
RenderX

Table . The baseline characteristics of patients with psoriasis undergoing biologic treatment (N=192).

Characteristic	Total	Male patients (n=154)	Female patients (n=38)	*P* value
Age (y), mean (SD)	41.2 (18.2)	39.5 (18.5)	48.1 (15.4)	.009
Age group (y), n (%)				.04
<35	76 (39.6)	67 (43.5)	9 (23.7)	
35-45	37 (19.3)	28 (18.2)	9 (23.7)	
46-60	42 (21.9)	32 (20.8)	10 (26.3)	
>60	37 (19.3)	27 (17.5)	10 (26.3)	
Educational level, n (%)				.02
Primary school or lower	48 (25.0)	45 (29.2)	3 (7.9)	
Junior high school	25 (13.0)	19 (12.3)	6 (15.8)	
Senior high school	25 (13.0)	19 (12.3)	6 (15.8)	
College or higher	94 (49.0)	71 (46.1)	23 (60.5)	
Monthly income (¥; ¥1=US $0.14), n (%)				.002
<3000	47 (24.5)	45 (29.2)	2 (5.3)	
3000-5000	36 (18.8)	32 (20.8)	4 (10.5)	
5001-10,000	101 (52.6)	71 (46.1)	30 (78.9)	
>10,000	8 (4.2)	6 (3.9)	2 (5.3)	
Residence, n (%)				.16
Urban	128 (66.7)	99 (64.3)	29 (76.3)	
Rural	64 (33.3)	55 (35.7)	9 (23.7)	
Marital status, n (%)				.003
Married	113 (58.9)	82 (53.2)	31 (81.6)	
Unmarried	48 (25.0)	46 (29.9)	2 (5.3)	
Divorced, widow or widower, or other	31 (16.1)	26 (16.9)	5 (13.2)	
BMI (kg/m^2), mean (SD)	25.8 (4.3)	26.4 (4.1)	23.5 (4.4)	<.001
BMI (kg/m^2), n (%)				.004
<23.9 (low or normal weight)	67 (34.9)	45 (29.2)	22 (57.9)	
24.0-28.0 (overweight)	73 (38.0)	63 (40.9)	10 (26.3)	
>28.0 (obesity)	52 (27.1)	46 (29.9)	6 (15.8)	
Disease duration (y), median (IQR)	13.5 (7.0-21.0)	12.5 (7.0-20.0)	19.5 (11.0-25.0)	.03
PASI[a] at week 0, median (IQR)	12.2 (9.8-18.6)	12.3 (10.0-19.2)	11.3 (8.0-16.8)	.30
BSA[b] at week 0 (%), median (IQR)	18.5 (12.0-33.3)	18.5 (12.0-32.0)	16.8 (11.0-33.5)	.62
PGA[c] at week 0, median (IQR)	3.0 (2.0-3.0)	3.0 (2.0-3.0)	2.2 (2.0-2.7)	<.001

[a]PASI: Psoriasis Area and Severity Index. The PASI indicates the severity of psoriasis, with a higher value representing a more severe condition (0-72).

[b]BSA: body surface area. BSA indicates the affected body surface, with a higher value representing a more severe condition (0%-100%).

[c]PGA: Physician Global Assessment. The PGA indicates the physician's overall evaluation of psoriasis, with a higher score representing a more severe condition (0-5).

XSL•FO
RenderX

Tobacco Smoking Among Patients With Psoriasis

As shown in Table 2, of the 192 patients with psoriasis undergoing biologic treatment, 78 (40.6%) were tobacco smokers, including 74 (94.9%) male and 4 (5.1%) female patients. The prevalence of tobacco smoking was higher in male patients (74/154, 48.1%) than in female patients (4/38, 10.5%), and the difference was statistically significant (χ^2_1=17.8; *P*<.001).

Table . Tobacco smoking characteristics among patients with psoriasis (N=192).

Variable	Total	Male patients (n=154)	Female patients (n=38)	*P* value
Tobacco smoking, n/N (%)				<.001
Yes	78/192 (40.6)	74/154 (48.1)	4/38 (10.5)	
No	114/192 (59.4)	80/154 (51.9)	34/38 (89.5)	
Smoking initiation age (y), mean (SD)	24.0 (4.9)	24.0 (4.8)	24.0 (6.7)	.99
Smoking initiation age (y), median (IQR)	24.0 (20.0-26.0)	24.5 (20.0-26.0)	21.0 (20.0-28.0)	.67
Smoking initiation age (y), n/N (%)				.64
<25	40/78 (51.3)	37/74 (50.0)	3/4 (75.0)	
≥25	38/78 (48.7)	37/74 (50.0)	1/4 (25.0)	
Smoking duration (y), mean (SD)	21.2 (16.1)	20.5 (15.7)	32.8 (21.6)	.14
Smoking duration (y), median (IQR)	20.0 (9.0-35.0)	19.0 (9.0-35.0)	35.0 (14.5-51.0)	.21
Smoking duration (y), n/N (%)				.64
<20	38/78 (48.7)	37/74 (50.0)	1/4 (25.0)	
≥20	40/78 (51.3)	37/74 (50.0)	3/4 (75.0)	
Cigarettes consumed per d, mean (SD)	16.9 (7.9)	17.1 (8.1)	13.8 (4.8)	.42
Cigarettes consumed per d, median (IQR)	20.0 (10.0-20.0)	20.0 (10.0-20.0)	12.5 (10.0-17.5)	.30
Cigarettes consumed per d, n/N (%)				.34
<20	31/78 (39.7)	28/74 (37.8)	3/4 (75.0)	
≥20	47/78 (60.3)	46/74 (62.2)	1/4 (25.0)	

The average and median age of smoking initiation was 24.0 (SD 4.9; IQR 20.0-26.0) years, and 51.3% (40/78) of patients began tobacco use at <25 years. The mean smoking duration was 21.2 (SD 16.1) years, and the median duration was 20.0 (IQR 9.0-35.0) years. A total of 51.3% (40/78) of patients had smoked for more than 20 years. The mean daily cigarette consumption among patients was 16.9 (SD 7.9) cigarettes, with a median of 20 (IQR 10.0-20.0) cigarettes. In total, 60.3% (47/78) of patients smoked more than 20 cigarettes per day. The differences between male and female patients who smoked tobacco in smoking initiation age (t_{76}=0.01 and *P*=.99; *z*=–0.42 and *P*=.67; χ^2_1=0.2 and *P*=.64), smoking duration (t_{76}=–1.48 and *P*=.14; *z*=–1.26 and *P*=.21; χ^2_1=0.2 and *P*=.64), and daily cigarette consumption (t_{76}=0.80 and *P*=.42; *z*=–1.04 and *P*=.30; χ^2_1=0.9 and *P*=.34) were not statistically significant (Table 2).

Treatment Outcomes Among Patients With Psoriasis

Table 3 illustrates the longitudinal outcomes of PASI, BSA, and PGA scores, along with the achievement of PASI 75, PASI 90, and PGA (1/0) at weeks 4, 8, 12, 24, and 48 after treatment. During the treatment period, the achievement of PASI 75, PASI 90, and PGA (1/0) showed an increasing trend (*Z*=9.42 and *P*<.001; *Z*=8.53 and *P*<.001; and *Z*=3.14 and *P*=.002, respectively). The response rate increased from 29.2% (56/192) at week 4 to 82.7% (134/162) at week 48 for PASI 75, from 13.0% (25/192) at week 4 to 59.9% (97/162) at week 48 for PASI 90, and from 47.4% (91/192) at week 4 to 69.8% (113/162) at week 48 for PGA (1/0; Table 3).

XSL•FO
RenderX

Table . Treatment outcomes among patients with psoriasis at weeks 4, 8, 12, 24, and 48.

Variable	Weeks after treatment				
	Week 4 (n=192)	Week 8 (n=192)	Week 12 (n=192)	Week 24 (n=181)	Week 48 (n=162)
PASI[a]					
Mean (SD)	6.7 (6.3)	4.3 (5.1)	2.2 (3.1)	1.6 (2.9)	1.9 (3.0)
Median (IQR)	4.7 (2.1-10.2)	2.5 (0.8-5.7)	1.2 (0.0-3.0)	0.4 (0.0-1.8)	0.6 (0.0-2.4)
PASI 75, n (%)	56 (29.2)	104 (54.2)	151 (78.6)	153 (84.5)	134 (82.7)
PASI 90, n (%)	25 (13.0)	57 (29.7)	104 (54.2)	113 (62.4)	97 (59.9)
BSA[b] (%)					
Mean (SD)	14.9 (15.9)	8.9 (12.1)	3.8 (7.9)	1.6 (2.8)	2.5 (6.0)
Median (IQR)	10.0 (4.0-19.8)	5.0 (1.0-12.5)	1.1 (0.0-4.0)	0.4 (0.0-1.8)	0.2 (0.0-2.5)
PGA[c]					
Mean (SD)	1.6 (0.9)	1.2 (1.0)	0.9 (1.0)	0.9 (1.0)	1.1 (1.1)
Median (IQR)	2.0 (1.0-2.0)	1.0 (0.0-2.0)	1.1 (0.0-1.0)	1.0 (0.0-1.0)	1.0 (0.0-2.0)
PGA (1/0), n (%)	91 (47.4)	124 (64.6)	145 (75.5)	136 (75.1)	113 (69.8)

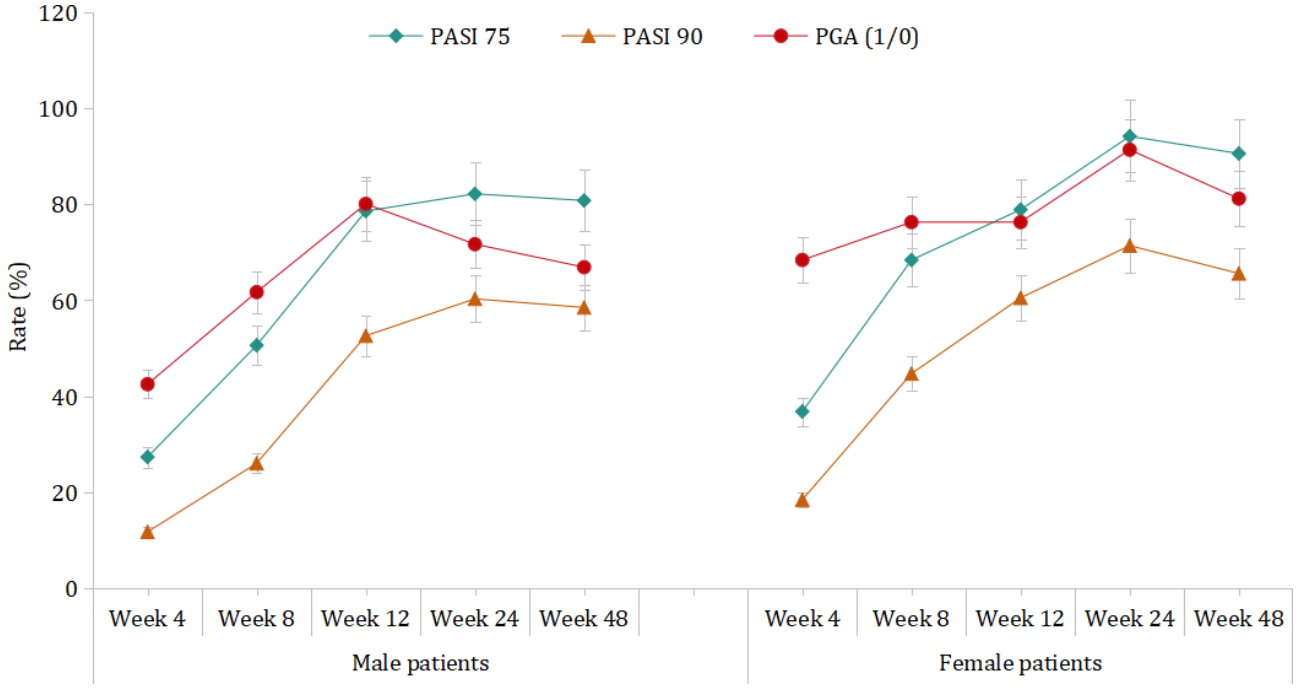
[a]PASI: Psoriasis Area and Severity Index. Indicates the severity of psoriasis, with a higher value representing a more severe condition (0-72).

[b]BSA: body surface area. Indicates the affected body surface, with a higher value representing a more severe condition (0%-100%).

[c]PGA: Physician Global Assessment. Indicates the physician's overall evaluation of psoriasis, with a higher score representing a more severe condition (0-5).

Stratification by gender revealed that, in male patients, the PASI 75, PASI 90, and PGA (1/0) response rates all showed an increasing trend (Z=8.35 and P<.001; Z=7.84 and P<.001; and Z=2.85 and P=.004, respectively). In female patients, only the PASI 75 and PASI 90 response rates increased with biologic treatment at different weeks (Z=4.47 with P<.001 and Z=3.47 with P<.001, respectively) but not the PGA (1/0) response rate (Z=1.42; P=.16; Figure 1). Moreover, data in Figure 1 also indicate that female patients typically achieved better response rates (PASI 75, PASI 90, and PGA [1/0]) than male patients at the same evaluation time points.

Figure 1. Treatment responses (Psoriasis Area and Severity Index [PASI] 75, PASI 90, and Physician Global Assessment [PGA] [1/0]) of male and female patients at weeks 4, 8, 12, 24, and 48.



XSL•FO
RenderX

Association Between Tobacco Smoking and Treatment Response

Logistic regression analysis indicated that patients with psoriasis who did not smoke tobacco had higher PASI 75 response rates than those who smoked tobacco, and the ORs were 2.57 (95% CI 1.19-5.53), 2.61 (95% CI 1.34-5.08), 2.62 (95% CI 1.13-6.04), 2.27 (95% CI 0.89-5.75), and 2.75 (95% CI 1.01-7.49) at weeks 4, 8, 12, 24, and 48, respectively, even with the adjustment of potential confounders. Patients who did not smoke tobacco also had higher PASI 90 response rates than those who smoked. The ORs were 1.32 (95% CI 0.49-3.54), 1.78 (95% CI 0.85-3.71), 1.71 (95% CI 0.88-3.33), 1.38 (95% CI 0.70-2.71), and 2.59 (95% CI 1.22-5.55) at weeks 4, 8, 12, 24, and 48, respectively. For PGA (1/0), patients who did not smoke tobacco tended to have higher response rates than those who smoked tobacco at weeks 4, 8, 12, 24, and 48. The ORs were 1.08 (95% CI 0.57-2.05), 1.14 (95% CI 0.58-2.24), 1.55 (95% CI 0.70-3.39), 1.52 (95% CI 0.69-3.31), and 1.18 (95% CI 0.54-2.59), respectively (Table 4).

Table . The association between tobacco smoking and treatment response (Psoriasis Area and Severity Index [PASI] 75, PASI 90, and Physician's Global Assessment [PGA] [1/0]) among patients with psoriasis at weeks 4 to 48.

Variable	Response rate (nonsmokers vs smokers; %)	OR[a] (95% CI)		
		Model A[b]	Model B[c]	Model C[d]
PASI 75				
Week 4	36.8 vs 17.9	2.67 (1.34-5.33)	2.56 (1.23-5.31)	2.57 (1.19-5.53)
Week 8	64.9 vs 38.5	2.96 (1.63-5.38)	2.54 (1.35-4.78)	2.61 (1.34-5.08)
Week 12	84.2 vs 70.5	2.23 (1.11-4.49)	2.23 (1.04-4.79)	2.62 (1.13-6.04)
Week 24	90.4 vs 76.6	2.87 (1.24-6.64)	2.33 (0.96-5.67)	2.27 (0.89-5.75)
Week 48	89.8 vs 74.3	3.03 (1.28-7.20)	2.72 (1.07-6.91)	2.75 (1.01-7.49)
PASI 90				
Week 4	14.9 vs 10.3	1.53 (0.63-3.75)	1.33 (0.50-3.51)	1.32 (0.49-3.54)
Week 8	36.0 vs 20.5	2.18 (1.11-4.25)	1.80 (0.88-3.69)	1.78 (0.85-3.71)
Week 12	59.6 vs 46.2	1.72 (1.00-3.09)	1.67 (0.90-3.11)	1.71 (0.88-3.33)
Week 24	66.3 vs 57.1	1.48 (0.81-2.72)	1.31 (0.68-2.51)	1.38 (0.70-2.71)
Week 48	68.2 vs 50.0	2.14 (1.13-4.06)	2.30 (1.14-4.66)	2.59 (1.21-5.55)
PGA (1/0)				
Week 4	51.3 vs 42.3	1.44 (0.80-2.57)	1.10 (0.59-2.07)	1.08 (0.57-2.05)
Week 8	67.5 vs 60.3	1.37 (0.75-2.50)	1.13 (0.60-2.14)	1.14 (0.58-2.24)
Week 12	78.2 vs 74.3	1.23 (0.63-2.45)	1.43 (0.69-2.99)	1.55 (0.70-3.39)
Week 24	80.8 vs 68.4	1.94 (0.98-3.85)	1.55 (0.74-3.22)	1.52 (0.69-3.31)
Week 48	72.7 vs 66.2	1.36 (0.70-2.67)	1.21 (0.58-2.50)	1.18 (0.54-2.59)

[a]OR: odds ratio.

[b]Model A: univariate logistic regression.

[c]Model B: multivariate logistic regression with adjustment for age and sex.

[d]Model C: multivariate logistic regression with adjustment for age, educational level, sex, marital status, monthly income, and BMI.

Among the 40.6% (78/192) of patients with psoriasis who smoked tobacco, the PASI score at baseline was positively correlated with tobacco smoking years (r=0.39; P=.03) and daily cigarette consumption (r=0.31; P=.04). Moreover, the PASI score reduction at weeks 4, 8, 12, 24, and 48 from baseline (week 0) was negatively correlated with tobacco smoking years (Figure 2) and daily cigarette consumption (Figure 3).

XSL•FO
RenderX

Figure 2. Correlation analysis between the Psoriasis Area and Severity Index (PASI) score at week 0 or PASI score reduction at weeks 4 to 48 and the years of tobacco smoking among smokers.

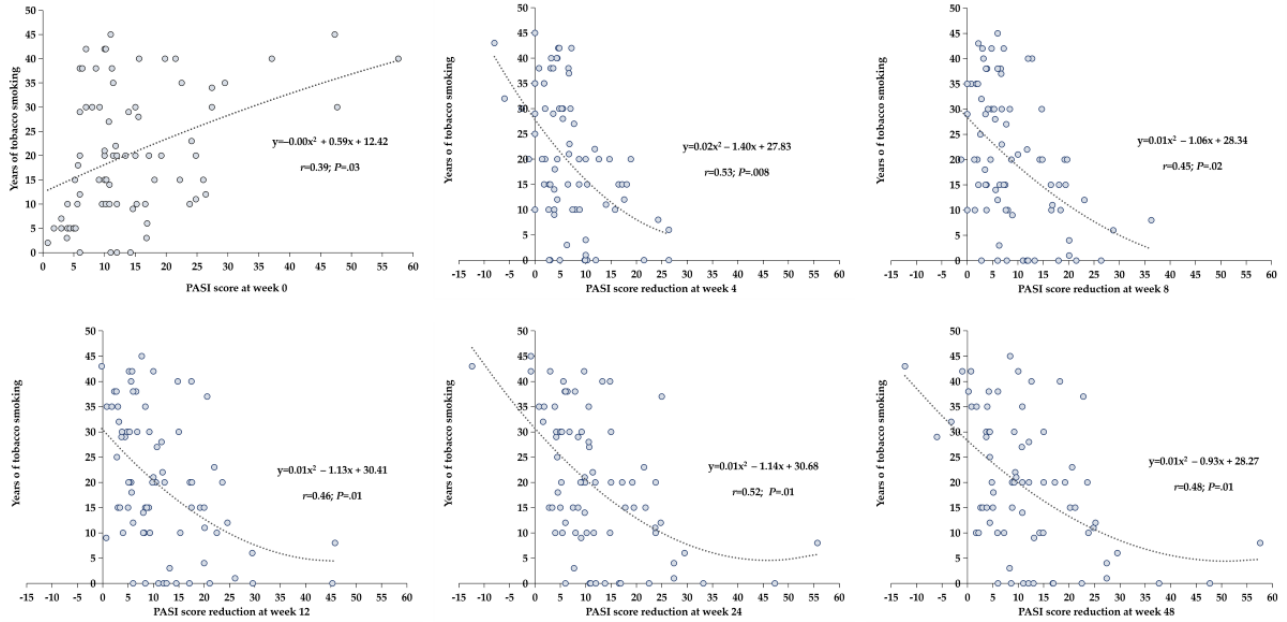
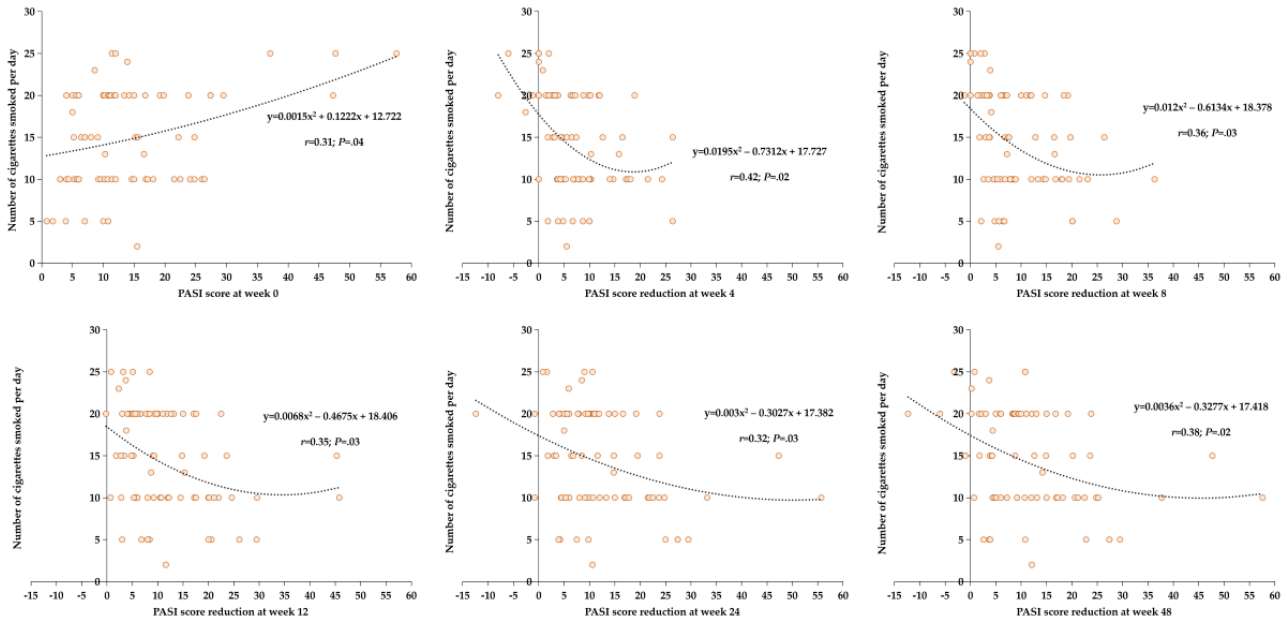


Figure 3. Correlation analysis between the Psoriasis Area and Severity Index (PASI) score at week 0 or PASI score reduction at weeks 4 to 48 and the number of cigarettes smoked per day.



Discussion

Principal Findings

This study explored the impact of tobacco smoking on the treatment efficacy of biologics in 192 patients with psoriasis based on real-world clinical data. In this study, the smoking prevalence among patients with psoriasis was 40.6% (78/192), which was consistent with that in previous studies [29-31]. The findings of this study indicated that patients with psoriasis who did not smoke had better treatment response than those who smoked, especially in the PASI 75 evaluation. Moreover, the reduction in PASI score at weeks 4, 8, 12, 24, and 48 from baseline was negatively correlated with both tobacco smoking years and number of daily cigarettes consumed.

Rapid clearance of skin lesions is the greatest treatment need among patients with psoriasis in China [32]. The emergence and application of biologics have brought about a qualitative leap in the treatment goals for psoriasis, with PASI 90 and PASI 100 becoming the new therapeutic targets for the biologic treatment of psoriasis [2,5,6]. In a previous study, based on data from the Psoriasis Study of Health Outcomes, Lynde et al [33] reported that the response rates for the primary end point (static PGA [0/1] and/or PASI 90) at week 12 were 71.4% and 58.6% for patients receiving anti–IL-17A biologics and other biologics, respectively. Pinter et al [34] compared the long-term treatment effectiveness of different biologics, including ixekizumab, risankizumab, and ustekinumab, and indicated that the response rates for PASI 90 were 40.5% to 64.3% at month 6 and 41.7% to 61.4% at month 12. In this study, we observed similar results,

XSL•FO
RenderX

with overall PASI 75 response rates of 78.6% (151/192), 84.5% (153/181), and 82.7% (134/162); PASI 90 response rates of 54.2% (104/192), 62.4% (113/181), and 59.9% (97/162); and PGA (1/0) rates of 75.5% (145/192), 75.1% (136/181), and 69.8% (113/162) at weeks 12, 24, and 48, respectively. The response rates for PASI 75 and PASI 90 at week 24 were slightly higher than those at week 48, which might be attributed to an increased rate of treatment discontinuation after receiving 24 weeks of treatment among patients.

Tobacco smoking is recognized as an important risk factor for psoriasis [35,36]. In this study, we noticed that tobacco smoking was negatively associated with treatment response at weeks 4, 8, 12, 24, and 48 among patients with psoriasis undergoing treatment with biologics. Compared to patients who smoked tobacco, patients who did not smoke tobacco had 2.75 and 2.59 times higher odds of achieving PASI 75 and PASI 90 at week 48, respectively, which is consistent with most previous studies. Two meta-analyses have reported that patients with psoriasis who smoke exhibit poorer response to biologics [20,23]. In a multicenter longitudinal cohort study, Warren et al [24] found that smokers had reduced odds of achieving PASI 90 at 6 and 12 months after initiating biologic therapies. These studies all suggest a negative impact of tobacco smoking on treatment response to biologics for psoriasis.

Previous studies have indicated that smoking intensity and duration are positively correlated with the severity of psoriasis in a dose-dependent fashion [17,18,37]. In this study, we observed consistent results demonstrating that PASI score was positively correlated with tobacco smoking years and number of daily consumed cigarettes. Moreover, we focused on the relationship between smoking and treatment efficacy among patients with psoriasis undergoing treatment with biologics during the 48-week follow-up period. The results showed that the reduction in PASI score was negatively correlated with tobacco smoking duration and tobacco smoking intensity. However, the specific mechanism of how tobacco smoking affects the efficacy of biologics for psoriasis still requires further research and exploration.

In this study, we also noticed that female patients had better treatment responses than male patients in the PASI 75, PASI 90, and PGA (1/0). There are several possible explanations for this gender disparity. First, female patients in this study had lower disease severity than male patients at baseline, which may have made it easier for female patients to achieve disease improvement. Second, BMI may also play a role in this difference. In this study, the average BMI of female patients was significantly lower than that of male patients, and higher BMI is considered to be associated with increased disease severity and reduced treatment response in psoriasis [38]. Pharmacokinetic studies have shown that excessive body weight can have a negative impact on drug clearance and volume of distribution [38]. Third, the proportion of female patients with a college education or higher was higher than that of male patients, and a higher educational level often leads to a better treatment adherence and a subsequently better treatment response [39]. Fourth, the lower tobacco smoking proportion might have also contributed to the better treatment response among female patients with psoriasis.

A key strength of this study was that the clinical data of patients with psoriasis were extracted directly from the health information system, eliminating recall bias, which ensured high data quality. The collection of data for over 48 weeks, allowing for the assessment of the impact of tobacco smoking on biologic efficacy at multiple time points, was another strength. Additionally, we used a variety of evaluation tools to assess patient outcomes, including the PASI, BSA, and PGA, thereby providing more and multidimensional data support for the study results.

Limitations

This study has several limitations. First, the 192 patients were recruited from only 1 hospital, which ensured high internal authenticity; however, the generalization of the findings to other patients with psoriasis is limited. Second, the smoking status of patients was self-reported, which may be subject to social desirability bias and recall bias. The lack of information on secondhand smoke and e-cigarette exposure may have underestimated the impact of smoking on treatment efficacy to some extent. Third, the number of female patients included in this study was relatively low, especially among smokers, which may have resulted in the findings more strongly reflecting the situation of male patients. Future studies should consider conducting investigations with larger sample sizes to comprehensively assess the impact of smoking on treatment response in patients with psoriasis with different demographic characteristics.

Conclusions

The prevalence of tobacco smoking among patients with psoriasis undergoing biologic treatment was high, especially among male patients. Tobacco smoking was negatively associated with treatment response among patients with psoriasis undergoing biologic treatment, especially among those with longer tobacco smoking duration and higher daily cigarette consumption. Therefore, we recommend that dermatologists actively assess tobacco smoking status among patients and encourage them to quit smoking to improve their treatment response.

Acknowledgments

The authors would like to thank the dermatologists at Shanghai Skin Disease Hospital for their data collection support in this study.

XSL•FO
RenderX

Funding

This study was supported by the Clinical Research Three-Year Plan of Shanghai Hospital Development Center (SHDC2022CRS053 and SHDC2024CRX032), Clinical Research Program of Shanghai Municipal Health Commission (202240371), and Intelligence Fund of Shanghai Skin Disease Hospital (2021KYQD01). The funders had no role in the study design, data collection and analysis, decision to publish, or manuscript preparation.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: RW
Data curation: ZD, RC
Formal analysis: YD, XZ, QX
Investigation: FS, XZ, QX, ZD, XG
Supervision: RZ, XG
Writing—original draft: FS, YD
Writing—review and editing: RW
All authors contributed to the paper and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Griffiths CEM, Armstrong AW, Gudjonsson JE, Barker J. Psoriasis. Lancet 2021 Apr 3;397(10281):1301-1315. [doi: 10.1016/S0140-6736(20)32549-6] [Medline: 33812489]
2. Menter A, Strober BE, Kaplan DH, et al. Joint AAD-NPF guidelines of care for the management and treatment of psoriasis with biologics. J Am Acad Dermatol 2019 Apr;80(4):1029-1072. [doi: 10.1016/j.jaad.2018.11.057] [Medline: 30772098]
3. Singh R, Koppu S, Perche PO, Feldman SR. The cytokine mediated molecular pathophysiology of psoriasis and its clinical implications. Int J Mol Sci 2021 Nov 26;22(23):12793. [doi: 10.3390/ijms222312793] [Medline: 34884596]
4. Ferrara F, Verduci C, Laconi E, et al. Current therapeutic overview and future perspectives regarding the treatment of psoriasis. Int Immunopharmacol 2024 Dec 25;143(Pt 1):113388. [doi: 10.1016/j.intimp.2024.113388] [Medline: 39405929]
5. Nast A, Smith C, Spuls PI, et al. EuroGuiDerm Guideline on the systemic treatment of psoriasis vulgaris – part 1: treatment and monitoring recommendations. J Eur Acad Dermatol Venereol 2020 Nov;34(11):2461-2498. [doi: 10.1111/jdv.16915] [Medline: 33349983]
6. Chinese Society of Dermatology, China Dermatologist Association, Dermatology and Venereology Specialized Committee of Chinese Association of Integrative Medicine. Guidelines for the treatment of psoriasis with biologics and small-molecule drugs in China (2024). Chin J Dermatol 2024;57(11):976-997. [doi: 10.35541/cjd.20240294]
7. Committee on Psoriasis, Chinese Society of Dermatology. Guideline for the diagnosis and treatment of psoriasis in China. Chin J Dermatol 2023;56(7):573-625. [doi: 10.35541/cjd.20220839]
8. Min S, Wang D, Xia J, Lin X, Jiang G. The economic burden and quality of life of patients with psoriasis treated with biologics in China. J Dermatolog Treat 2023 Dec;34(1):2247106. [doi: 10.1080/09546634.2023.2247106] [Medline: 37621166]
9. Takahashi H, Iinuma S, Tsuji H, Honma M, Iizuka H. Biologics are more potent than other treatment modalities for improvement of quality of life in psoriasis patients. J Dermatol 2014 Aug;41(8):686-689. [doi: 10.1111/1346-8138.12544] [Medline: 25040258]
10. Lebwohl MG, Armstrong AW, Alexis AF, Lain EL, Jacobson AA. Efficacy of brodalumab in patients with psoriasis and risk factors for treatment failure: a review of post hoc analyses. Dermatol Ther (Heidelb) 2024 Oct;14(10):2709-2726. [doi: 10.1007/s13555-024-01264-3] [Medline: 39264399]
11. Karczewski J, Poniedziałek B, Rzymski P, Adamski Z. Factors affecting response to biologic treatment in psoriasis. Dermatol Ther 2014;27(6):323-330. [doi: 10.1111/dth.12160] [Medline: 25053228]
12. GBD 2019 Tobacco Collaborators. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990–2019: a systematic analysis from the Global Burden of Disease Study 2019. Lancet 2021 Jun 19;397(10292):2337-2360. [doi: 10.1016/S0140-6736(21)01169-7] [Medline: 34051883]
13. Alhindal M, Janahi J, D'Angelo EC, et al. Impact of smoking on cardiovascular health: mechanisms, epidemiology and specific concerns regarding congenital heart disease. Int J Cardiol Congenit Heart Dis 2025 Apr 2;20:100581. [doi: 10.1016/j.ijcchd.2025.100581] [Medline: 40242035]

XSL•FO
RenderX

14. Safiri S, Carson-Chahhoud K, Noori M, et al. Burden of chronic obstructive pulmonary disease and its attributable risk factors in 204 countries and territories, 1990-2019: results from the Global Burden of Disease Study 2019. BMJ 2022 Jul 27;378:e069679. [doi: 10.1136/bmj-2021-069679] [Medline: 35896191]
15. Secretan B, Straif K, Baan R, et al. A review of human carcinogens--part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. Lancet Oncol 2009 Nov;10(11):1033-1034. [doi: 10.1016/s1470-2045(09)70326-2] [Medline: 19891056]
16. Armstrong AW, Bohannan B, Mburu S, et al. Patient perspectives on psoriatic disease burden: results from the global psoriasis and beyond survey. Dermatology 2023;239(4):621-634. [doi: 10.1159/000528945] [Medline: 37075723]
17. Richer V, Roubille C, Fleming P, et al. Psoriasis and smoking: a systematic literature review and meta-analysis with qualitative analysis of effect of smoking on psoriasis severity. J Cutan Med Surg 2016 May;20(3):221-227. [doi: 10.1177/1203475415616073] [Medline: 26553732]
18. Wei L, Chen S, Zhang Z, et al. Prevalence of tobacco smoking and its association with disease severity among patients with psoriasis in China: a cross-sectional study. Front Med (Lausanne) 2022 May 12;9:883458. [doi: 10.3389/fmed.2022.883458] [Medline: 35646971]
19. Zorlu O, Bülbül Başkan E, Yazici S, et al. Predictors of drug survival of biologic therapies in psoriasis patients. J Dermatolog Treat 2022 Feb;33(1):437-442. [doi: 10.1080/09546634.2020.1763240] [Medline: 32351141]
20. Zhou H, Wu R, Kong Y, Zhao M, Su Y. Impact of smoking on psoriasis risk and treatment efficacy: a meta-analysis. J Int Med Res 2020 Oct;48(10):300060520964024. [doi: 10.1177/0300060520964024] [Medline: 33121308]
21. Hong EJ, Yoon HJ, Kwon MJ, et al. Factors associated with achieving complete skin clearance compared to almost complete skin clearance in patients with moderate to severe psoriasis treated with biologics: a retrospective chart review. Ann Dermatol 2024 Apr;36(2):91-98. [doi: 10.5021/ad.23.074] [Medline: 38576247]
22. Schwarz CW, Loft N, Rasmussen MK, et al. Predictors of response to biologics in patients with moderate-to-severe psoriasis: a Danish nationwide cohort study. Acta Derm Venereol 2021 Oct 26;101(10):adv00579. [doi: 10.2340/actadv.v101.351] [Medline: 34642768]
23. Hjort G, Schwarz CW, Skov L, Loft N. Clinical characteristics associated with response to biologics in the treatment of psoriasis: a meta-analysis. JAMA Dermatol 2024 Aug 1;160(8):830-837. [doi: 10.1001/jamadermatol.2024.1677] [Medline: 38888917]
24. Warren RB, Marsden A, Tomenson B, et al. Identifying demographic, social and clinical predictors of biologic therapy effectiveness in psoriasis: a multicentre longitudinal cohort study. Br J Dermatol 2019 May;180(5):1069-1076. [doi: 10.1111/bjd.16776] [Medline: 30155885]
25. Egeberg A, Conrad C, Gorecki P, et al. Response types and factors associated with response types to biologic therapies in patients with moderate-to-severe plaque psoriasis from two randomized clinical trials. Dermatol Ther (Heidelb) 2024 Mar;14(3):745-758. [doi: 10.1007/s13555-024-01123-1] [Medline: 38485863]
26. Kinahan CE, Mazloom S, Fernandez AP. Impact of smoking on response to systemic treatment in patients with psoriasis: a retrospective case-control study. Br J Dermatol 2015 Feb;172(2):428-436. [doi: 10.1111/bjd.13359] [Medline: 25142556]
27. Constantin MM, Bucur S, Mutu CC, et al. The impact of smoking on psoriasis patients with biological therapies in a Bucharest hospital. J Pers Med 2021 Jul 30;11(8):752. [doi: 10.3390/jpm11080752] [Medline: 34442396]
28. Anzengruber F, Augustin M, Radtke MA, et al. Smoking does not alter the therapy response to systemic anti-psoriatic therapies: a two-country, multi-centre, prospective, non-interventional study. Acta Derm Venereol 2019 Sep 1;99(10):871-877. [doi: 10.2340/00015555-3221] [Medline: 31099404]
29. May F, Armstrong J, Vogrin S, Baker C, Foley P. Drug survival of biological therapies in smokers and non-smokers with psoriasis: a retrospective cohort study using data from the Australasian Psoriasis Registry. Australas J Dermatol 2025 May;66(3):e140-e147. [doi: 10.1111/ajd.14450] [Medline: 40028779]
30. Salihbegovic EM, Kurtalic N, Omerkic E. Smoking cigarettes and consuming alcohol in patients with psoriasis. Mater Sociomed 2021 Mar;33(1):30-33. [doi: 10.5455/msm.2021.33.30-33] [Medline: 34012347]
31. Adışen E, Uzun S, Erduran F, Gürer MA. Prevalence of smoking, alcohol consumption and metabolic syndrome in patients with psoriasis. An Bras Dermatol 2018 Mar;93(2):205-211. [doi: 10.1590/abd1806-4841.20186168] [Medline: 29723384]
32. Ding W, Yao M, Wang Y, et al. Patient needs in psoriasis treatment and their influencing factors: a nationwide multicentre cross-sectional study in China. Indian J Dermatol 2023;68(5):587. [doi: 10.4103/ijd.ijd_611_23] [Medline: 38099128]
33. Lynde C, Riedl E, Maul JT, et al. Comparative effectiveness of biologics across subgroups of patients with moderate-to-severe plaque psoriasis: results at week 12 from the PSoHO study in a real-world setting. Adv Ther 2023 Mar;40(3):869-886. [doi: 10.1007/s12325-022-02379-9] [Medline: 36515803]
34. Pinter A, Costanzo A, Khattri S, et al. Comparative effectiveness and durability of biologics in clinical practice: month 12 outcomes from the international, observational psoriasis study of health outcomes (PSoHO). Dermatol Ther (Heidelb) 2024 Jun;14(6):1479-1493. [doi: 10.1007/s13555-023-01086-9] [Medline: 38113010]
35. Laguardia F, Owczarczyk-Saczonek A, Maggi P. The role of smoking in psoriasis. Postepy Dermatol Alergol 2024 Apr;41(2):143-148. [doi: 10.5114/ada.2023.133582] [Medline: 38784922]
36. Roszkiewicz M, Dopytalska K, Szymańska E, Jakimiuk A, Walecka I. Environmental risk factors and epigenetic alternations in psoriasis. Ann Agric Environ Med 2020 Sep 11;27(3):335-342. [doi: 10.26444/aaem/112107] [Medline: 32955211]

XSL•FO
RenderX

37. Yang X, Lu L, Zhou H, Xu X, Lin B. Prevalence of tobacco smoking and its association with disease severity and psoriatic arthritis among patients with nail psoriasis in China: a cross-sectional study. Front Med (Lausanne) 2025 Dec 19;12:1660724. [doi: 10.3389/fmed.2025.1660724] [Medline: 41446844]
38. Pirro F, Caldarola G, Chiricozzi A, et al. Impact of body mass index on the efficacy of biological therapies in patients with psoriasis: a real-world study. Clin Drug Investig 2021 Oct;41(10):917-925. [doi: 10.1007/s40261-021-01080-z] [Medline: 34537921]
39. Teixeira A, Teixeira M, Gaio R, et al. Influence of clinical and psychosocial factors on the adherence to topical treatment in psoriasis. Healthcare (Basel) 2024 Apr 12;12(8):822. [doi: 10.3390/healthcare12080822] [Medline: 38667584]

Abbreviations

BSA: body surface area
IL: interleukin
OR: odds ratio
PASI: Psoriasis Area and Severity Index
PGA: Physician Global Assessment

Edited by A Kushniruk; submitted 07.Jan.2026; peer-reviewed by B Li, D Zhu; revised version received 21.Jan.2026; accepted 09.Feb.2026; published 04.Mar.2026.

Please cite as:
Shen F, Ding Y, Zhang X, Xu Q, Duan Z, Cai R, Zhang R, Gao X, Wang R
The Impact of Tobacco Smoking on Treatment Response Among Patients With Psoriasis Undergoing Biologic Treatment: Prospective Observational Study
JMIR Hum Factors 2026;13:e90963
URL: https://humanfactors.jmir.org/2026/1/e90963
doi:10.2196/90963

© Fanlingzi Shen, Yuning Ding, Xiuqi Zhang, Quanruo Xu, Zhen Duan, Ruiqi Cai, Rui Zhang, Xiangjin Gao, Ruiping Wang. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 4.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Sex Representation and User Preferences in Pain Drawing Body Charts in Back Pain Research: Multimethod Study

Sonja Schläpfer[1], PhD; Casey Murphy[1], MBA; Yanick X Lukic[2,3], PhD; Concepción Campos-Asensio[4], BPharm, MLIS; Giovanni Colacicco[5], MD; Beatrice A Lauber[5], DVM; Sophie Masneuf[6], PhD; Luana Nyirö[7], DCM; Georg W Kajdi[8], MD; Reto Sutter[8], Prof Dr Med; Christoph J Laux[9], MD; Philipp Ackermann[2], PhD; Jesús López-Alcalde[1,10,11], MPH, MD; Claudia M Witt[1], Prof Dr Med

[1]Complementary and Integrative Digital Health, Institute of Primary Care, University of Zurich and University Hospital Zurich, Sonneggstrasse 6, Zurich, Switzerland

[10]Faculty of Medicine, Universidad Francisco de Vitoria (UFV), Madrid, Spain

[11]Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), Unidad de bioestadística clínica, Hospital Universitario Ramón y Cajal, (CIBERESP), Cochrane Madrid, Madrid, Spain

[2]Institute of Computer Science, School of Engineering, Zurich University of Applied Sciences, Winterthur, Switzerland

[3]School of Medicine, University of St.Gallen, St.Gallen, Switzerland

[4]Medical Library, Hospital Universitario de Getafe, Madrid, Spain

[5]Institute of Anatomy, University of Zurich, Zurich, Switzerland

[6]Neuroscience Center Zurich, ETH and University of Zurich, Zurich, Switzerland

[7]Department of Chiropractic Medicine, Balgrist University Hospital, University of Zurich, Zurich, Switzerland

[8]Radiology Department, Balgrist University Hospital, University of Zurich, Zurich, Switzerland

[9]University Spine Center Zurich, Balgrist University Hospital, University of Zurich, Zurich, Switzerland

Corresponding Author:
Sonja Schläpfer, PhD
Complementary and Integrative Digital Health, Institute of Primary Care, University of Zurich and University Hospital Zurich, Sonneggstrasse 6, Zurich, Switzerland

Abstract

Background: Pain drawing (PD) body charts are widely used in back pain research, but the representation of sex in these charts has not been systematically evaluated.

Objective: This study aims to evaluate sex representation in PD body charts used in back pain research, assess the perception of a newly designed sex-neutral body chart, and explore user preferences for sex representation in PD body charts.

Methods: We conducted a multimethod study comprising: (1) a scoping review to assess sex representation and the reporting of sex in PD body charts in back pain literature, (2) an expert opinion study where anatomy experts evaluated the perceived sex of extracted body charts, and (3) a survey among a representative sample of UK adults with and without back pain to assess the perception of a newly designed sex-neutral body chart and explore preferences for sex representation in PD body charts.

Results: From 349 full-text papers, 108 articles met the inclusion criteria. Most (103/108, 95.4%) did not report the sex of the body charts used, and only 5.6% (6/108) included both male and female charts. Experts showed fair to moderate agreement (Fleiss κ=0.306; Gwet AC1=0.456) in assessing the sex of charts, with most charts assessed as male based on majority ratings (59/108, 54.6%) and classified as male-biased relative to the sex distribution of study participants (76/108, 67.7%). The newly designed sex-neutral body chart was perceived as sex-neutral by 68.5% (204/298) of survey participants across diverse groups. However, perceptions varied by racial group: 73% (181/248) of White participants viewed it as sex-neutral, compared to 42.5% (17/40) of participants from smaller racial groups (χ^2_2=15.9; *P*=.001). Female participants slightly preferred female charts (85/154, 55.2%); male participants preferred sex-neutral ones (88/144, 61.1%). Nonetheless, most female participants (82/154, 53.3%) and male participants (107/144, 74.3%) considered the option to choose between male, female, and sex-neutral chart versions unimportant.

Conclusions: Our study reveals reporting gaps and a predominant male bias in the representation of sex in PD body charts used in back pain research. The newly developed sex-neutral body chart was widely perceived as sex-neutral, offering a promising step toward more inclusive pain assessment. However, variations in perception across racial groups highlight the need for cultural considerations in design. These findings underscore the potential of sex-neutral and culturally sensitive body charts to enhance the inclusivity and equity of back pain research and clinical practice.

(JMIR Hum Factors 2026;13:e76175) doi:10.2196/76175

XSL•FO
RenderX

KEYWORDS

pain drawing; pain map; body chart; sex; gender; sex/gender bias; back pain

Introduction

A pain drawing (PD) is an illustration of the human body that allows individuals to draw the location and extent of their pain. These visual tools support the assessment of subjective pain, may help determine appropriate treatment strategies, and facilitate communication between patients and health care professionals [1]. Since their introduction in 1949, PDs have evolved from traditional paper-and-pencil formats into digital and, in some cases, 3D representations [2]. The body charts used in these PDs vary considerably [2], with some representing male or female figures, while others may be perceived as sex-neutral.

The first sex-specific body charts, featuring both male and female versions, were introduced in 1988 by Uden et al [3], marking an important advancement in acknowledging the need for diversity in pain reporting tools [2]. However, despite this progress, subsequent research has largely neglected the role of diversity in body chart design. Most studies have examined how a participant's sex influences pain depiction by comparing PDs from female and male participants [4,5], while overlooking whether the sex of the body chart itself influences pain reporting. In addition to this gap, it remains unclear to what extent male, female, and sex-neutral body charts are currently used in pain research.

The sex of body charts used in pain research is often poorly documented, and the prevalence of different sex representations remains largely unknown. Preliminary studies suggest that male or sex-neutral charts with male features are more commonly used than female or truly sex-neutral alternatives. Most body charts are described using terms such as "human body outline," without explicitly identifying them as male, female, or sex-neutral [6]. This ambiguity raises important questions about how a "universal" human body is visually represented across sexes and genders, and how individuals perceive and interpret these depictions. A prevailing assumption among researchers is that most body charts are sex-neutral with distinctly male features, while truly sex-neutral representations are rare [1,7]. However, this perspective is not grounded in a comprehensive review, highlighting the need for further investigation into potential sex and gender biases in body chart design.

Another important consideration is whether individuals prefer male, female, or sex-neutral body charts. While some studies indicate a preference for sex-specific body charts [1,8,9], their findings may be limited by the fact that body charts intended to be sex-neutral are not always perceived or interpreted as such. Understanding individuals' preferences is critical, as these preferences may influence how users engage with the charts, their comfort in using them, and the accuracy of their pain reporting. When individuals feel represented by the body chart, they may be more likely to provide accurate and detailed pain descriptions, potentially leading to improved diagnostic outcomes [1]. Exploring these preferences also supports the development of inclusive, user-centered tools that reflect the diversity of patient populations—an essential step toward more equitable health care.

This study examines how sex is represented in PD body charts within the back pain literature, analyzing both the descriptions provided in published articles and anatomy experts' assessments of the perceived sex of the charts. Furthermore, we evaluated a newly developed sex-neutral body chart, designed by an interdisciplinary team of illustrators, computer scientists, physicians (specializing in anatomy, orthopedics, radiology, and integrative medicine), and psychologists [10]. Our aim was to determine whether this chart is perceived by individuals as truly sex-neutral, in alignment with its intended design. We also explored a range of individual factors that may influence these perceptions, alongside participants' preferences for sex-specific versus sex-neutral body charts. Taken together, this study sought to address the following research questions:

1. What is the prevalence of male, female, and sex-neutral PD body charts in the back pain literature—from its inception to the present—as reported in the articles and as assessed by experts? Additionally, what is the level of agreement among experts in their sex-based classifications of the charts?
2. How do individuals perceive a body chart explicitly designed to be sex-neutral, and what individual factors influence whether they interpret the chart as male, female, or sex-neutral?
3. In the context of back pain assessment, do individuals prefer male, female, or sex-neutral body charts, and how important is having a choice to them?

Throughout this paper, we use the terms *sex* and *gender* in accordance with the SAGER (Sex and Gender Equity in Research) guidelines [11]. We define *sex* in body charts based on anatomical characteristics, including primary and secondary sexual organs typically associated with male and female bodies, and in humans as sex assigned at birth. *Gender* refers to socially constructed roles, behaviors, and identities in humans.

Methods

Study Design

This research adopted a multimethod design comprising a scoping review, an expert opinion study, and a survey—each focusing on different aspects of sex representation in PD body charts within the context of back pain. The expert opinion study incorporated insights from anatomy experts across various medical and health sciences disciplines, while the survey gathered responses from individuals both with and without back pain. This inclusive approach ensured the integration of diverse perspectives throughout the study. We submitted the study synopsis to the Ethics Committee of Zurich, Switzerland, and after review, they stated that the study does not fall under the regulation of the Human Research Act of Switzerland (ethics ID: 2022 - 01475). The scoping review protocol was registered

XSL•FO
RenderX

in the Open Science Framework on July 1, 2024, prior to starting the process of selecting studies for the review [12].

Ethical Considerations

The study was conducted in accordance with ethical standards. Privacy and confidentiality were ensured by pseudonymizing all data in the expert opinion study and by collecting anonymized data in the survey, with no personal identifiers recorded. All data were securely stored. Participation was voluntary, and participants were informed of their right to withdraw at any time. Participants provided electronic informed consent. Each participant received £0.75 (US $1) as compensation for their time.

Inclusion Criteria

Review

The scoping review considered studies that have used, designed, or evaluated PDs with a full-body or torso template in paper or digital format. Studies were included if they assessed pain location and involved adults (aged ≥18 y) with back pain, regardless of participants' demographic characteristics (such as age, sex, or pain type and severity). The review considered all primary research studies published in English or German, conducted in any country or clinical setting. Secondary studies—including systematic reviews, narrative reviews, clinical practice guidelines, and consensus documents—were excluded.

Expert Opinion Study

The expert opinion study included experts with academic qualifications and professional expertise in medicine or related health sciences, with an equal distribution of male and female experts.

Survey

The survey included English-speaking adults aged 18 years or older who resided in the United Kingdom, had provided informed consent, and either had or had not experienced back pain, regardless of its type or severity. To ensure that the sample was approximately representative of the UK population, participants also had to meet demographic criteria for sex, age, and race based on simplified UK Census data.

Recruitment Procedure

Review

The search strategy for the scoping review aimed to identify relevant published studies. An initial, limited search of PubMed was conducted to locate key articles, followed by an analysis of the text words in titles and abstracts, as well as the index terms used to describe those articles. Insights from this analysis informed the development of a comprehensive search strategy incorporating the identified keywords and index terms, as outlined in Multimedia Appendix 1. The final search was conducted by a health librarian (CC-A) in the MEDLINE database via Ovid, covering all records from database inception through May 3, 2024.

Expert Opinion Study

Experts in anatomy and related health sciences were recruited through personal invitations extended via the professional networks of the research team. We aimed to engage a total of 6 experts—3 male and 3 female individuals—to ensure balanced representation and to capture a diverse range of perspectives across relevant areas of expertise.

Survey

The anonymous survey was conducted on Prolific [13], with a recruitment goal of 300 participants—an approximate sample size considered adequate for robust descriptive analyses in survey research [14]. Eligible individuals from the Prolific participant pool were able to self-select into the study if they met the inclusion criteria. Prior to beginning the 5-minute survey, participants provided electronic informed consent through the platform. Each participant received £0.75 (US $1) as compensation for their time.

Assessments

Review

All identified citations were imported into EndNote (version 21.2; Clarivate Analytics), where duplicates were identified and removed. The unique records were then transferred to Rayyan [15] for study selection. The selection process involved full-text screening of all articles, including appendices, to detect images of PD body charts. Each record was independently assessed by one reviewer (CM or SS), with 10% of excluded articles cross-checked by a second reviewer (SS or CM) according to predefined eligibility criteria. Discrepancies were resolved through discussion between the two reviewers or, if needed, by consultation with a third reviewer (JL-A or CMW). Articles that mentioned the use of PD body charts but did not include images were only retained if the type of body chart used was explicitly referenced and the images could be located through the cited source.

Data were extracted from papers included in the scoping review using the data extraction tool listed in Multimedia Appendix 2. This tool was piloted with 8 articles to ensure its reliability and comprehensiveness. Two reviewers (SS and CM) independently extracted data from the first 10 articles to establish consistency. For the remaining articles, one reviewer (either CM or SS) extracted the data, while a second reviewer cross-verified 10% of the extracted data to maintain quality control. The extracted data included specific details about the populations, concepts, contexts, and study methods relevant to the scoping review's questions and objectives. Any disagreements between the reviewers were resolved through discussion or, when necessary, by consulting a third reviewer (JL-A or CMW).

Expert Opinion Study

Following data extraction, our experts (GC, BAL, SM, LN, GWK, and CJL) independently assessed the sex representation of the PD body charts identified in the scoping review. Images of the body charts were compiled in a Microsoft Excel file for review. Experts were asked to classify each body chart by selecting one of the following predefined categories from a drop-down menu: "Male," "Female," "Sex-neutral," "Male and

XSL•FO
RenderX

female (>1 template)," or "Unclear." As the original articles rarely provided detailed descriptions of the body charts, there were no predefined correct or incorrect answers; rather, the goal was to capture the experts' subjective assessments informed by their anatomical expertise. In cases where the same body chart appeared multiple times—occasionally in slightly different versions—experts were instructed to assess each version individually. When a body chart included multiple figures with different sex representations, the category "Male and female (>1 template)" was used, where applicable. If none of the predefined categories were suitable, experts had the option to elaborate in a designated comments field.

Survey

The survey aimed to examine participants' perceptions of body charts, focusing specifically on how they classified the chart designed by Lukic et al [10] (as female, male, or sex-neutral), their preferences for selecting a body chart to represent their pain (female, male, or sex-neutral), and the importance they placed on having the option to select from these versions. Additionally, we collected self-reported demographic information, including age, current and past experiences with back pain, country of residence, sex assigned at birth, gender identity, educational attainment, and race. The survey also included additional questions related to the acceptability and usability of the body chart; however, these were not related to the research questions addressed in this paper and were analyzed separately [10].

Analyses

Review

We conducted a descriptive analysis of the extracted data, focusing on the geographical location of the studies, whether the sex of the body chart used was reported, and whether single or multiple body charts were used. We also assessed whether the studies reported the sex of the participants and, based on this information, whether they included both male and female participants, or exclusively one sex.

Expert Opinion Study

Due to the imbalanced distribution of ratings—51.2% of which fell into the "Male" category—both Fleiss κ and Gwet AC1 were calculated to assess inter-rater agreement. Fleiss κ provides a conservative estimate that accounts for prevalence effects, representing a "worst-case" scenario, while the Gwet AC1 offers a more robust measure in the presence of imbalanced data, reflecting a "best-case" scenario [16]. While Fleiss κ is a widely used standard metric, the Gwet AC1 is increasingly recommended for its reliability under uneven category distributions. Agreement levels for both statistics were interpreted using the guidelines established by Landis and Koch [17]. For additional context, raw percentage agreement was also calculated as a straightforward, unadjusted measure of observed consensus among raters. Analyses were performed using the *irr* (version 0.84.1), *irrCAC* (version 1.0), and *DescTools* (version 0.99.59) packages in R (R Foundation for Statistical Computing). Agreement was assessed across 5 categories reflecting experts' judgments of the body chart's sex: "Male," "Female," "Sex-neutral," "Male and female (>1 template)," and "Unclear." The analyses were performed for all raters collectively, as well as separately for female raters, male raters, and for all raters excluding one identified as a potential outlier based on preliminary rating patterns. Additionally, we analyzed majority ratings, defined as instances where at least 4 out of 6 raters assigned the same rating. We also evaluated the extent to which expert ratings aligned with explicit body chart descriptions reported in the literature. Finally, we investigated whether the sex classification assigned by experts corresponded with the sex of the study participants as reported in the articles, considering body charts rated as "sex-neutral" to be a match for participants of any sex.

Survey

Descriptive statistics were computed for the entire sample, followed by subgroup analyses to compare responses across sex, age groups, back pain experience (current or past), education level, and race. Chi-square tests were conducted to assess whether the distribution of responses differed significantly between subgroups. For the subgroup analyses, dichotomous groups were formed where appropriate: For age, participants were categorized into 2 groups based on the mean age: those aged older than 46 years (above the mean) and those aged 46 years or younger. Educational levels were categorized into "Non-tertiary education" (including "No formal qualification," "Lower secondary qualification," "Upper secondary qualification without university access," and "University entry qualification") and "Tertiary education" (including "Post-secondary below bachelor's level," "Bachelor's level," and "Master's level and above"). Race was grouped into "Largest racial group" (participants identifying as White) and "Other racial groups" (including those identifying as Asian, Black, Latino or Hispanic, Middle Eastern, Indigenous, Aboriginal or First Nations, or "Other," including multiracial participants). Participants who selected "I prefer not to answer" were excluded from race-based subgroup analyses. Due to the small number of participants with diverse gender identities, subgroup analyses by gender were not feasible. For descriptive purposes, free-text responses indicating "female" or "woman" were categorized as "woman," and those indicating "male" or "man" as "man." Responses unrelated to gender identity (eg, referring to sexual orientation) were classified as "other."

Results

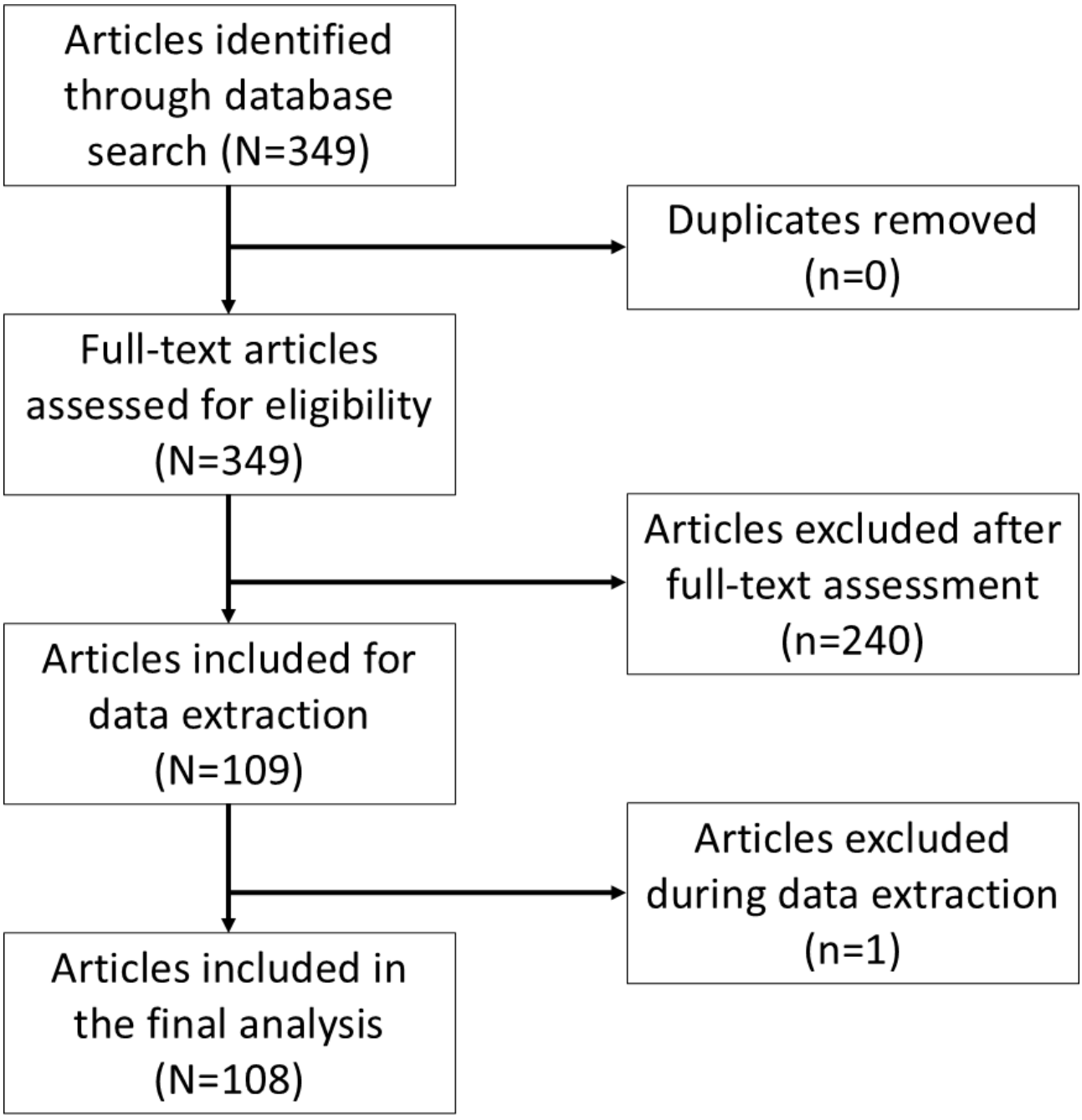
Scoping Review

Our search yielded 349 articles, of which 109 were selected for further analysis after full-text screening. During data extraction, one article was excluded because it used actual participant photos instead of an illustrated body chart for PD, resulting in a final sample of 108 articles (see Figure 1 and full citations in Multimedia Appendix 3). The articles referred to studies predominantly conducted in Europe (n=60, 55.6%) and North America (n=31, 28.7%) and were published between 1988 and 2022. Notably, none of the articles explicitly stated an intention to use a sex-neutral body chart, and only 5 (4.6%) articles reported the sex of the body chart used. Among these, 4 articles [8,18-20] used 2 different body charts described as male and female, while 1 article [21] used a single body chart described

XSL•FO
RenderX

as male. Out of the 108 articles, 6 (5.6%) used 2 different body charts [3,8,18-20,22], while the remaining 102 (94.4%) used a single body chart. One of the studies using two charts included only female participants [22]. Among studies using a single body chart, 70.6% (72/102) included both male and female participants [4,21,23-92], 8.8% (9/102) included only female participants [93-101], and 2.0% (2/102) included only male participants [102,103]. In 18.6% (19/102) of studies, the sex or gender of participants was not clearly reported; these studies used general terms such as "patients," "subjects," or "employees," without further specification [104-122].

Figure 1. Flowchart of literature search and study selection



Expert Opinion Study

The expert panel consisted of 6 raters—3 female and 3 male raters—including 4 physicians (an orthopedic spine surgeon, a radiologist, a chiropractor, and an anatomy lecturer), an anatomical dissection specialist, and a neuroscientist. All experts self-identified as White. Each expert independently assessed the sex representation of body charts from 108 articles, resulting in a total of 648 individual ratings. The distribution of ratings was as follows: 332 (51.2%) ratings were classified as "male," 158 (24.4%) as "sex-neutral," 74 (11.4%) as "female," 27 (4.2%) as "male and female" (for charts depicting two different bodies), 57 (8.8%) as "unclear," and 0 (0%) as "other." Female experts classified 1.42 times more body charts as "male" (195 ratings, 60.2% of their total) compared to male experts (137 ratings, 42.3% of their total). Conversely, male experts assigned more than twice as many body charts as "sex-neutral" (108 ratings, 33.3% of their total) than the female experts (50 ratings, 15.4%

XSL•FO
RenderX

of their total). See Figure 2 for a visual representation of these findings.

Figure 2. Body chart ratings by category and expert sex. The categories represent experts' ratings of the perceived sex of the body charts. "Mixed" refers to body charts that included multiple figures perceived as both male and female. Each group of experts—female and male—contributed 324 ratings, for a total of 648 ratings.

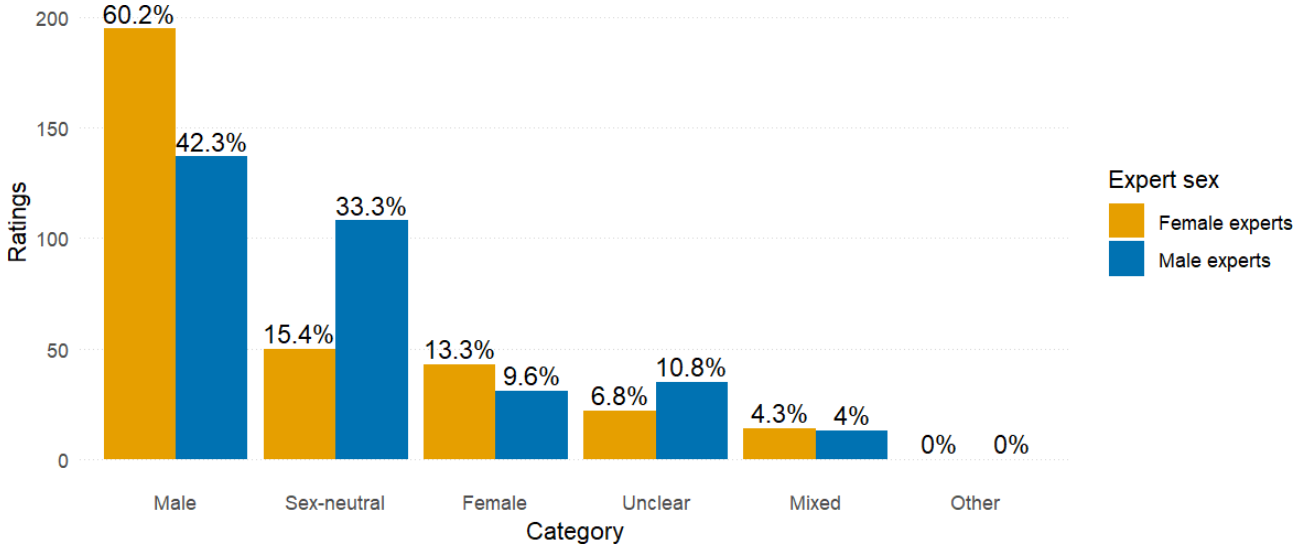


Table 1 presents the inter-rater agreement metrics for the overall sample and subgroups. Across all raters, the Fleiss κ indicated fair agreement, while the Gwet AC1 suggested moderate agreement. The simple percentage agreement was 10.2% (see Multimedia Appendix 4 for images that were rated identically). Subgroup analyses revealed the following patterns: for the 3 female raters, there was moderate agreement for both Fleiss κ and Gwet AC1, and the simple percentage agreement was 52.8%. The 3 male raters revealed slight agreement for Fleiss κ and fair agreement for Gwet AC1, with one suspected outlier; the simple percentage agreement was 13.9%. After excluding the suspected outlier, agreement among the remaining 5 raters increased to moderate agreement (Fleiss κ) and substantial (Gwet AC1), with a percentage agreement of 49.1%. All interpretations should be considered alongside the confidence intervals in Table 1, and percentage agreement values should be interpreted with caution due to the limited nuance.

Table . Inter-rater agreement metrics.

Group and metric	Metric value	95% CI	Interpretation
All raters (n=6)			
Fleiss κ	0.306	0.276 - 0.336	Fair agreement
Gwet AC1	0.456	0.401 - 0.511	Moderate agreement
Percentage agreement	10.2	N/A[a]	—[b]
Female raters (n=3)			
Fleiss κ	0.404	0.337 - 0.47	Moderate agreement
Gwet AC1	0.587	0.496 - 0.679	Moderate agreement
Percentage agreement	52.8	N/A	—
Male raters (n=3)			
Fleiss κ	0.098	0.031 - 0.165	Slight agreement
Gwet AC1	0.251	0.188 - 0.314	Fair agreement
Percentage agreement	13.9	N/A	—
Without male outlier (n=5)			
Fleiss κ	0.488	0.452 - 0.524	Moderate agreement
Gwet AC1	0.647	0.571 - 0.724	Substantial agreement
Percentage agreement	49.1	N/A	—

[a]N/A: not applicable.

[b]—: not available.

XSL•FO
RenderX

We also examined majority ratings, defined as at least 4 out of 6 raters assigning the same classification. Body charts were categorized as "male" in more than half of the articles (59/108, 54.6%), "female" in 10 (9.3%) articles, "sex-neutral" in 8 (7.4%) articles, "male and female" in 4 (3.7%) articles, and "unclear" in 1 (0.9%) article. In 26 articles (24.1%), ratings were too inconsistent to achieve a majority.

Of the 108 articles, only 5 explicitly described the body charts as male, female, or sex-neutral. Four of these articles included 2 body charts each, described as "male" and "female." In three of these articles, all raters classified the sex of the charts in line with the article descriptions. In the fourth, however, rater assessments diverged: four classified both charts as "female," and two as "sex-neutral." The fifth article included a single body chart described as male, which 5 experts identified as "male" and one as "unclear."

Finally, we assessed alignment between the expert-assigned sex of the body charts and the sex of study participants as reported in the articles. Of the 108 articles, 40 (37.0%) provided unclear descriptions of participants' sex or lacked a majority rating from the experts. In the remaining 68 articles, 22.1% (15/68) showed a match between the chart and participant sex. The majority, 67.7% (46/68), were male-biased (ie, male body charts used in studies not exclusively involving male participants), while 10.3% (7/68) were female-biased.

Survey

A total of 300 individuals participated in the survey; however, two were excluded due to technical issues, resulting in a final sample of 298 participants. Descriptive statistics for the sample are presented in Table 2, and subgroup comparisons, including chi-square test results, are shown in Table 3. The majority of respondents (204/298, 68.5%) perceived Lukic et al's [10] body chart as sex-neutral. In contrast, 20.1% (60/298) identified it as representing a male body, while 11.4% (34/298) perceived it as female. Chi-square tests revealed no significant differences in response distributions based on sex, age group, current or past back pain experience, or educational background. However, a significant association was observed for race: participants who identified as White were more likely to perceive the body chart as sex-neutral compared to those who identified as Asian; Black; Latino or Hispanic; Middle Eastern; Indigenous, Aboriginal, or First Nations; or "Other" (Table 3).

XSL•FO
RenderX

Table . Survey sample characteristics (n=298).

Characteristics	Value
Age (y)	
Mean (SD)	46.16 (15.35)
Min	18
Max	79
Current back pain, n (%)	
Yes	154 (51.7)
Back pain in past, n (%)	
Yes	239 (80.2)
Participants' sex, n (%)	
Female	154 (51.7)
Male	144 (48.3)
Participants' gender, n (%)	
Woman	148 (49.7)
Man	133 (44.6)
Other	2 (0.7)
Missing	15 (5)
Education, n (%)	
Nontertiary	134 (45)
No formal qualification	2
Lower secondary qualification	21
Upper secondary qualification without university access	80
University entry qualification	31
Tertiary	164 (55)
Post-secondary below bachelor's level	10
Bachelor's level	96
Master's level and higher	58
Race, n (%)	
Asian	24 (8.1)
Black	12 (4)
Indigenous, Aboriginal, or First Nations	0 (0)
Latino or Hispanic	1 (0.3)
Middle Eastern	6 (2)
Other	6 (2)
White	248 (83.2)
I prefer not to answer	5 (1.7)

XSL•FO
RenderX

Table . Responses to the question: "Does this body chart represent a female, male, or sex-neutral body?" by participants' sex, age, back pain experience, education, and race.

Variable	Body chart, n (%)			Chi-square (*df*)	*P* value
	Sex-neutral	Male	Female		
Participants' sex				0.3 (2)	.92
Female (n=154)	104 (67.5)	31 (20.1)	19 (12.3)		
Male (n=144)	100 (69.4)	29 (20.1)	15 (10.4)		
Age groups (y)				4.3 (2)	.13
≤46 (n=139)	87 (62.6)	34 (24.5)	18 (13)		
>46 (n=159)	117 (73.6)	26 (16.4)	16 (10.1)		
Current back pain				0.9 (2)	.66
No (n=144)	97 (67.4)	32 (22.2)	15 (10.4)		
Yes (n=154)	107 (69.5)	28 (18.2)	19 (12.3)		
Past back pain				1.1 (2)	.57
No (n=59)	38 (64.4)	12 (20.3)	9 (15.3)		
Yes (n=239)	166 (69.5)	48 (20.1)	25 (10.5)		
Educational attainment				1.0 (2)	.61
Nontertiary (n=134)	93 (69.4)	24 (17.9)	17 (12.7)		
Tertiary (n=164)	111 (67.7)	36 (22)	17 (10.4)		
Race				15.9 (2)	.001
Other racial groups[a] (n=40)	17 (42.5)	13 (32.5)	10 (25)		
Largest racial group[b] (n=248)	181 (73)	44 (17.7)	23 (9.3)		

[a] "Other racial groups" includes participants who identified as Asian, Black, Middle Eastern, Latino or Hispanic, Indigenous, Aboriginal, First Nations, or "Other" (including multiracial).

[b] "Largest racial group" includes participants who identified as White.

Most participants (189/298, 63.4%) indicated that it was not important to them to be able to choose between female, male, or sex-neutral versions of the body chart. This included 53.3% (82/154) of female participants and 74.3% (107/144) of male participants. When given the option, over half of the female participants (85/154; 55.2%) said they would choose a female body chart, while 46.1% (71/154) preferred a sex-neutral body chart, and 2 (1.3%) selected a male body chart. In contrast, nearly two-thirds of male participants (88/144, 61.1%) preferred a sex-neutral body chart, 36.1% (52/144) chose a male body chart, 2.8% (4/144) opted for a female body chart. Two participants did not identify with the sex assigned to them at birth; both were assigned male and identified as women. One rated the body chart as male, preferred a female body chart, and expressed that having the option to choose is important. The other rated the chart as sex-neutral, preferred a sex-neutral chart, and reported that the option to choose is not important. Due to the small number of participants with diverse gender identities, no conclusions can be drawn regarding their perceptions and preferences.

Discussion

Principal Results

Our review revealed that the vast majority of articles in the back pain literature do not explicitly report the sex of the body chart used, nor do they indicate an intention of sex neutrality. Very few articles included both male and female versions of body charts. In our expert opinion study, over half of the body charts were classified as "male" and did not accurately reflect the sex of the study participants. Specifically, two-thirds were identified as male-biased—meaning male body charts were used in studies that did not exclusively involve male participants. Agreement among expert raters ranged from fair to moderate, with stronger agreement observed among female raters than male raters. This suggests that assessing the sex of body charts is inherently challenging. While many charts may have been designed to be sex-neutral, they are not consistently perceived as such. In contrast, Lukic et al's [10] body chart—explicitly designed as sex-neutral by an interdisciplinary team—was perceived as sex-neutral by two-thirds of male and female survey participants, regardless of age, back pain experience (current or past), or education level. However, a significant difference emerged by race: participants who identified as White were more likely to

XSL•FO
RenderX

perceive the chart as sex-neutral compared to those who did not identify as White. In terms of preference, female participants showed a slight inclination toward female body charts over sex-neutral ones, while male participants clearly preferred sex-neutral charts over male ones. Notably, both male and female participants generally considered the ability to choose among body chart versions to be of little importance. These findings suggest that a well-designed sex-neutral body chart is broadly accepted by both male and female users and holds the potential for promoting greater sex and gender inclusivity in PDs.

Comparison With Prior Work

Despite the introduction of sex-specific PD body charts in the late 1980 s [3], our findings show that most subsequent studies have continued to use a single body chart—typically perceived as male—to represent both male and female participants. This supports previous observations noting that most PD body charts used in the literature are male [1,7]. To more accurately reflect the diversity of study participants, future studies could either provide different versions of body charts that align with the sex or gender of participants [8,19,20] or offer a sex-neutral version that is broadly perceived as truly neutral [10].

The variability in expert ratings underscores the subjectivity involved in perceiving sex representations. What one expert identifies as male, female, or sex-neutral may differ from another's interpretation, even when both have similar anatomical training and expertise. This echoes earlier findings showing variability in how study participants perceive sex in body charts [7]. Such discrepancies may stem from individual experiences and implicit biases, as well as from the design of body charts that omit explicit sex-specific cues. While the absence of these cues is likely intended to enhance neutrality, it may also make interpretation difficult, leading to the fair to moderate agreement levels observed in our study. This suggests that many of the body charts found in the literature were likely intended to be sex-neutral but were not consistently interpreted as such—instead, they are most often interpreted as "male."

A male bias has been well-documented in human sex perception and anatomical illustrations. Research indicates that when faced with ambiguity or uncertainty, humans are more likely to perceive a figure as male [123,124]. This perceptual bias may have evolutionary roots: misidentifying a potentially dangerous individual (typically male) as female could have higher survival costs, aligning with instinctual self-preservation mechanisms [123]. Therefore, "male" often becomes the default in sex perception, while "female" is only assumed when explicit female sex cues are present [124]. This phenomenon aligns with our findings and may help explain why body charts with ambiguous or unclear sex cues were predominantly perceived as male. Additionally, the design of the body charts themselves—even when designed to be sex-neutral—may reflect a bias toward male figures. Such bias is not unique to PDs but mirrors a broader trend in anatomical illustrations, where male figures are overrepresented, reinforcing the male body as the normative standard [125-129].

Given this observed male bias and the imbalance in sex representation, it is essential to consider the needs and preferences of individuals currently underrepresented in PD tools. Our survey results suggest that a truly sex-neutral body chart is well-accepted by both male and female individuals. While female individuals showed a slight preference for a female-specific body chart, the option to choose between different body chart versions was generally not considered crucial. This nuance builds on the findings of Egsgaard et al [1], who reported a preference among female participants for female body charts. It is important to note, however, that participants in Egsgaard et al's study [1] were only presented with male and female options—without the inclusion of a sex-neutral alternative. Interestingly, while female participants in that study expressed a clear preference for female body charts over male ones, more than half indicated that the ability to choose was not particularly important. This aligns with our results, suggesting that although preferences exist, they may not be a crucial factor in pain reporting. The methodological difference in body chart options presented likely accounts for the partial discrepancy in findings between our study and the study by Egsgaard et al [1]. Together, these insights highlight the potential value of including truly sex-neutral body charts in PD tools, as they may offer a more inclusive and representative option for diverse patient populations while still accommodating those with a preference for sex-specific representations.

Limitations

Our scoping review aimed for comprehensiveness but was limited by language restrictions (English and German) and the exclusive use of the MEDLINE database. These constraints may have introduced selection bias, potentially excluding relevant literature and culturally specific body charts. Future research should expand language criteria and database inclusion to ensure a more comprehensive review. Another potential limitation is that some articles may have displayed only a single body chart, despite using multiple charts in their study. This could have led to misinterpretations and inaccurate conclusions in our review.

Regarding the expert assessment, an important limitation was the lack of cultural diversity among the experts, as all experts self-identified as White. Future studies should aim to include a more diverse panel to assess how cultural backgrounds might influence perceptions of the body charts.

Similarly, of the 298 survey participants, only 40 did not identify as White, which limits the generalizability of our findings to racially diverse populations. Despite this, the survey sample was broad in terms of age and educational background and approximately balanced in terms of sex. Future research should explore how the body chart designed by Lukic et al [10] is perceived by individuals from a broader range of racial and cultural backgrounds to better understand its cross-cultural applicability and inclusivity.

Clinical Implications

The observed acceptance of our sex-neutral body chart across sexes highlights the potential of sex-neutral PD tools to offer a more universally applicable approach to pain assessment. These tools may be particularly valuable in clinical settings where traditional sex-specific charts are impractical or where sex

XSL•FO
RenderX

information is unavailable or deemed nonessential—for example, in emergency care settings, large-scale epidemiological studies, or in the care of nonbinary and transgender individuals. This approach aligns with a growing shift toward inclusivity in medical tools, such as the development of non–sex-specific growth charts for transgender youth [125]. Moreover, sex-neutral PD tools could be particularly beneficial in digital health applications, where standardized, sex-neutral inputs would facilitate automated analysis and improve consistency across diverse populations.

Conclusions

Our scoping review and subsequent studies reveal important gaps and biases in the reporting and representation of sex in PD body charts used in back pain research. The majority of charts are perceived as male, often misrepresenting the study population they are intended to depict. Expert assessments showed fair to moderate agreement, suggesting the complexity of sex perception and indicating that many charts—despite possibly being intended as sex-neutral—are not consistently perceived as such. Our study offers a promising solution: a newly designed sex-neutral body chart, developed with expert input, was widely perceived as sex-neutral across diverse participant groups. However, variations in perception across racial groups highlight the importance of incorporating cultural considerations into design. These findings underscore the potential of a truly sex-neutral body chart to support greater sex and gender balance in pain assessment tools. Future research should prioritize the development and evaluation of culturally sensitive body chart designs to further promote inclusivity and equity in both back pain research and clinical practice.

Funding

This research was supported by the Digitalization Initiative of the Zurich Higher Education Institutions. The views expressed in this publication are those of the authors and do not necessarily reflect those of the funding body. The funder had no role in the study's design, data collection, analysis, interpretation, or manuscript writing.

Authors' Contributions

Conceptualization: CMW (lead), SS (supporting), CM (equal), YXL (equal), CC-A (equal), GC (equal), RS (equal), CJL (equal), PA (equal), JL-A (supporting)
Data curation: SS
Formal analysis: SS
Funding acquisition: CMW (equal), PA (equal)
Investigation: SS (lead), CM (equal), YXL (equal), CC-A (equal), GC (equal), BAL (equal), SM (equal), LN (equal), GWK (equal), CJL (equal), JL-A (supporting), CMW (supporting)
Methodology: SS (lead), YXL (equal), JL-A (equal), CMW (equal)
Project administration: SS (lead), JL-A (supporting)
Resources: CMW
Supervision: CMW (lead), JL-A (supporting)
Visualization: SS
Writing – original draft: SS
Writing – review & editing: CMW (lead), CM (equal), YXL (equal), CC-A (equal), GC (equal), BAL (equal), SM (equal), LN (equal), GWK (equal), RS (equal), CJL (equal), PA (equal), JL-A (equal)

Conflicts of Interest

CMW has active research grants to the University for digital health projects from the Digitalization Initiative of the Zurich Higher Education Institutions, the Swiss Cancer Research Foundation, the Swiss National Science Foundation, the German Health Care Innovation Fund, and Newsense Lab GmbH and has received honoraria from Swiss hospitals for scientific presentations on digitalization and artificial intelligence in medicine. The radiology department of the Balgrist University Hospital has an academic research collaboration with Siemens Healthineers and Bayer. The remaining authors declare no conflicts of interest.

Multimedia Appendix 1
Search strategy.
[DOCX File, 14 KB - humanfactors_v13i1e76175_app1.docx]

Multimedia Appendix 2
Data extraction tool.
[XLSX File, 13 KB - humanfactors_v13i1e76175_app2.xlsx]

Multimedia Appendix 3
References to included studies.

XSL•FO
RenderX

[DOCX File, 130 KB - humanfactors_v13i1e76175_app3.docx]

Multimedia Appendix 4
Body charts with 100% agreement among all 6 experts.
[DOCX File, 1268 KB - humanfactors_v13i1e76175_app4.docx]

References

1. Egsgaard LL, Christensen TS, Petersen IM, Brønnum DS, Boudreau SA. Do gender-specific and high-resolution three dimensional body charts facilitate the communication of pain for women? A quantitative and qualitative study. JMIR Hum Factors 2016 Jul 20;3(2):e19. [doi: 10.2196/humanfactors.5693] [Medline: 27440737]
2. Shaballout N, Neubert TA, Boudreau S, Beissner F. From paper to digital applications of the pain drawing: systematic review of methodological milestones. JMIR mHealth mHealth 2019 Sep 5;7(9):e14569. [doi: 10.2196/14569] [Medline: 31489841]
3. Udén A, Aström M, Bergenudd H. Pain drawings in chronic back pain. Spine (Phila Pa 1976) 1988 Apr;13(4):389-392. [doi: 10.1097/00007632-198804000-00002] [Medline: 2970123]
4. Toomingas A. Characteristics of pain drawings in the neck-shoulder region among the working population. Int Arch Occup Environ Health 1999 Mar;72(2):98-106. [doi: 10.1007/s004200050344] [Medline: 10197481]
5. George SZ, Bialosky JE, Wittmer VT, Robinson ME. Sex differences in pain drawing area for individuals with chronic musculoskeletal pain. J Orthop Sports Phys Ther 2007 Mar;37(3):115-121. [doi: 10.2519/jospt.2007.2399] [Medline: 17416126]
6. Jamison RN, Fanciullo GJ, Baird JC. Usefulness of pain drawings in identifying real or imagined pain: accuracy of pain professionals, nonprofessionals, and a decision model. J Pain 2004 Nov;5(9):476-482. [doi: 10.1016/j.jpain.2004.08.004] [Medline: 15556825]
7. Szczypien N, Ruchay Z, Ruchay Z, Müller SV, Kaiser C, Klawonn F. Sex-specific differences in pain localization in female patients with endometriosis: a comparison of sexless and female human body outlines. Brain Behav 2023 Dec;13(12):e3285. [doi: 10.1002/brb3.3285] [Medline: 37853673]
8. Cruder C, Falla D, Mangili F, et al. Profiling the location and extent of musicians' pain using digital pain drawings. Pain Pract 2018 Jan;18(1):53-66. [doi: 10.1111/papr.12581] [Medline: 28466572]
9. Neubert TA, Dusch M, Karst M, Beissner F. Designing a tablet-based software app for mapping bodily symptoms: usability evaluation and reproducibility analysis. JMIR mHealth uHealth 2018 May 30;6(5):e127. [doi: 10.2196/mhealth.8409] [Medline: 29848470]
10. Lukic YX, Schläpfer S, Holler A, et al. Open-source anatomical illustration for digital back pain documentation: development and evaluation. JMIR Preprints. . [doi: 10.2196/preprints.75737]
11. Heidari S, Babor TF, De Castro P, Tort S, Curno M. Sex and gender equity in research: rationale for the SAGER guidelines and recommended use. Res Integr Peer Rev 2016;1(2):2. [doi: 10.1186/s41073-016-0007-6] [Medline: 29451543]
12. Schläpfer S, Lopez-Alcalde J, Murphy C, et al. Scoping review. Open Science Framework. 2025 Mar 20. URL: https://osf.io/pbuch/ [accessed 2024-07-01]
13. Prolific. 2024. URL: https://www.prolific.com [accessed 2026-01-17]
14. Lavrakas P. Encyclopedia of Survey Research Methods: Sage Publications, Inc; 2008. [doi: 10.4135/9781412963947]
15. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(1):210. [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
16. Gwet K. Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters: Advanced Analytics, LLC; 2014.
17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [doi: 10.2307/2529310] [Medline: 843571]
18. Alter BJ, Anderson NP, Gillman AG, Yin Q, Jeong JH, Wasan AD. Hierarchical clustering by patient-reported pain distribution alone identifies distinct chronic pain subgroups differing by pain intensity, quality, and clinical outcomes. PLoS ONE 2021;16(8):e0254862. [doi: 10.1371/journal.pone.0254862] [Medline: 34347793]
19. Barbero M, Moresi F, Leoni D, Gatti R, Egloff M, Falla D. Test-retest reliability of pain extent and pain location using a novel method for pain drawing analysis. Eur J Pain 2015 Sep;19(8):1129-1138. [doi: 10.1002/ejp.636] [Medline: 25565607]
20. Barbero M, Piff M, Evans D, Falla D. Do metrics derived from self-reported and clinician-reported pain drawings agree for individuals with chronic low back pain? Musculoskelet Sci Pract 2023 Nov;68:102871. [doi: 10.1016/j.msksp.2023.102871] [Medline: 37832508]
21. Hildebrandt J, Franz CE, Choroba-mehnen B, Temme M. The use of pain drawings in screening for psychological involvement in complaints of low-back pain. Spine (Phila Pa 1986) 1988 Jun;13(6):681-685. [doi: 10.1097/00007632-198813060-00016] [Medline: 2972073]
22. Starzec M, Truszczyńska-Baszak A, Tarnowski A, Rongies W. Pregnancy-related pelvic girdle pain in Polish and Norwegian women. J Manipulative Physiol Ther 2019 Feb;42(2):117-124. [doi: 10.1016/j.jmpt.2019.03.002] [Medline: 31126521]

XSL•FO
RenderX

23. Abbott JH, Foster M, Hamilton L, Ravenwood M, Tan N. Validity of pain drawings for predicting psychological status outcome in patients with recurrent or chronic low back pain. J Man Manip Ther 2015 Feb;23(1):12-19. [doi: 10.1179/2042618613Y.0000000046] [Medline: 26309377]
24. Albert HB, Hansen JK, Søgaard H, Kent P. Where do patients with MRI-confirmed single-level radiculopathy experience pain, and what is the clinical interpretability of these pain patterns? A cross-sectional diagnostic accuracy study. Chiropr Man Therap 2019;27(1):50. [doi: 10.1186/s12998-019-0273-8] [Medline: 31608144]
25. Alev L, Fujikoshi S, Yoshikawa A, et al. Duloxetine 60 mg for chronic low back pain: post hoc responder analysis of double-blind, placebo-controlled trials. J Pain Res 2017;10:1723-1731. [doi: 10.2147/JPR.S138297] [Medline: 28769588]
26. Andersen T, Christensen FB, Hansen ES, Bünger C. Pain 5 years after instrumented and non-instrumented posterolateral lumbar spinal fusion. Eur Spine J 2003 Aug;12(4):393-399. [doi: 10.1007/s00586-003-0547-6] [Medline: 12756629]
27. Andersen T, Christensen FB, Høy KW, et al. The predictive value of pain drawings in lumbar spinal fusion surgery. Spine J 2010 May;10(5):372-379. [doi: 10.1016/j.spinee.2010.02.002] [Medline: 20421073]
28. Apeldoorn AT, Bosselaar H, Ostelo RW, et al. Identification of patients with chronic low back pain who might benefit from additional psychological assessment. Clin J Pain 2012 Jan;28(1):23-31. [doi: 10.1097/AJP.0b013e31822019d0] [Medline: 21677570]
29. Bertilson BC, Brosjö E, Billing H, Strender LE. Assessment of nerve involvement in the lumbar spine: agreement between magnetic resonance imaging, physical examination and pain drawing findings. BMC Musculoskelet Disord 2010 Sep 10;11:202. [doi: 10.1186/1471-2474-11-202] [Medline: 20831785]
30. Bishop MD, Horn ME, George SZ, Robinson ME. Self-reported pain and disability outcomes from an endogenous model of muscular back pain. BMC Musculoskelet Disord 2011 Feb 2;12:35. [doi: 10.1186/1471-2474-12-35] [Medline: 21288349]
31. Bishop MD, Horn ME, Lott DJ, Arpan I, George SZ. Magnitude of spinal muscle damage is not statistically associated with exercise-induced low back pain intensity. Spine J 2011 Dec;11(12):1135-1142. [doi: 10.1016/j.spinee.2011.11.005] [Medline: 22208857]
32. Boudreau SA, Badsberg S, Christensen SW, Egsgaard LL. Digital pain drawings: assessing touch-screen technology and 3D body schemas. Clin J Pain 2016 Feb;32(2):139-145. [doi: 10.1097/AJP.0000000000000230] [Medline: 25756558]
33. Burkey AR, Kanetsky PA. Development of a novel location-based assessment of sensory symptoms in cancer patients: preliminary reliability and validity assessment. J Pain Symptom Manage 2009 May;37(5):848-862. [doi: 10.1016/j.jpainsymman.2008.05.013] [Medline: 19059751]
34. Falla D, Peolsson A, Peterson G, et al. Perceived pain extent is associated with disability, depression and self-efficacy in individuals with whiplash-associated disorders. Eur J Pain 2016 Oct;20(9):1490-1501. [doi: 10.1002/ejp.873] [Medline: 27149920]
35. Friedrich M, Gittler G, Pieler-Bruha E. Misleading history of pain location in 51 patients with osteoporotic vertebral fractures. Eur Spine J 2006 Dec;15(12):1797-1800. [doi: 10.1007/s00586-006-0065-4] [Medline: 16463199]
36. Hägg O, Fritzell P, Hedlund R, et al. Pain-drawing does not predict the outcome of fusion surgery for chronic low-back pain: a report from the Swedish Lumbar Spine Study. Eur Spine J 2003 Feb;12(1):2-11. [doi: 10.1007/s00586-002-0427-5] [Medline: 12592541]
37. Chan CW, Goldman S, Ilstrup DM, Kunselman AR, O'Neill PI. The pain drawing and Waddell's nonorganic physical signs in chronic low-back pain. Spine (Phila Pa 1976) 1993 Oct 1;18(13):1717-1722. [doi: 10.1097/00007632-199310000-00001] [Medline: 8235853]
38. Cleland JA, Childs JD, Palmer JA, Eberhart S. Slump stretching in the management of non-radicular low back pain: a pilot clinical trial. Man Ther 2006 Nov;11(4):279-286. [doi: 10.1016/j.math.2005.07.002] [Medline: 16380286]
39. Corrêa LA, Bittencourt JV, Ferreira ADS, Reis F, de Almeida RS, Nogueira LAC. The reliability and concurrent validity of PainMAP software for automated quantification of pain drawings on body charts of patients with low back pain. Pain Pract 2020 Jun;20(5):462-470. [doi: 10.1111/papr.12872] [Medline: 31961038]
40. De Oliveira Paes Leme M, Yuan SLK, Oliveira Magalhães M, Ferreira de Meneses SR, Marques AP. Pain and quality of life in knee osteoarthritis, chronic low back pain and fibromyalgia: a comparative cross-sectional study. Reumatismo 2019 Jul 9;71(2):68-74. [doi: 10.4081/reumatismo.2019.1104] [Medline: 31309776]
41. Dean LE, Arnold L, Crofford L, et al. Impact of moving from a widespread to multisite pain definition on other fibromyalgia symptoms. Arthritis Care Res (Hoboken) 2017 Dec;69(12):1878-1886. [doi: 10.1002/acr.23214] [Medline: 28182834]
42. Dyrehag LE, Widerström-Noga EG, Carlsson SG, et al. Relations between self-rated musculoskeletal symptoms and signs and psychological distress in chronic neck and shoulder pain. Scand J Rehabil Med 1998 Dec;30(4):235-242. [doi: 10.1080/003655098443986] [Medline: 9825388]
43. Ekman P, Möller H, Hedlund R. Predictive factors for the outcome of fusion in adult isthmic spondylolisthesis. Spine (Phila Pa 1976) 2009 May 15;34(11):1204-1210. [doi: 10.1097/BRS.0b013e3181a19e66] [Medline: 19444068]
44. Galve Villa M, Palsson TS, Boudreau SA. Spatiotemporal patterns of pain distribution and recall accuracy: a dose-response study. Scand J Pain 2022 Jan 27;22(1):154-166. [doi: 10.1515/sjpain-2021-0032] [Medline: 34343420]
45. Georgopoulos V, Akin-Akinyosoye K, Smith S, McWilliams DF, Hendrick P, Walsh DA. An observational study of centrally facilitated pain in individuals with chronic low back pain. Pain Rep 2022;7(3):e1003. [doi: 10.1097/PR9.0000000000001003] [Medline: 35441119]

XSL•FO
RenderX

46. Gerhardt A, Hartmann M, Blumenstiel K, Tesarz J, Eich W. The prevalence rate and the role of the spatial extent of pain in nonspecific chronic back pain--a population-based study in the south-west of Germany. Pain Med 2014 Jul;15(7):1200-1210. [doi: 10.1111/pme.12286] [Medline: 24341931]
47. Gerhardt A, Eich W, Janke S, Leisner S, Treede RD, Tesarz J. Chronic widespread back pain is distinct from chronic local back pain: evidence from quantitative sensory testing, pain drawings, and psychometrics. Clin J Pain 2016 Jul;32(7):568-579. [doi: 10.1097/AJP.0000000000000300] [Medline: 26379077]
48. Gerhardt A, Eich W, Treede RD, Tesarz J. Conditioned pain modulation in patients with nonspecific chronic back pain with chronic local pain, chronic widespread pain, and fibromyalgia. Pain 2017 Mar;158(3):430-439. [doi: 10.1097/j.pain.0000000000000777] [Medline: 27902566]
49. Goldstein P, Ashar Y, Tesarz J, Kazgan M, Cetin B, Wager TD. Emerging clinical technology: application of machine learning to chronic pain assessments based on emotional body maps. Neurotherapeutics 2020 Jul;17(3):774-783. [doi: 10.1007/s13311-020-00886-7] [Medline: 32767227]
50. Grunnesjö M, Bogefeldt J, Blomberg S, Delaney H, Svärdsudd K. The course of pain drawings during a 10-week treatment period in patients with acute and sub-acute low back pain. BMC Musculoskelet Disord 2006 Aug 11;7:65. [doi: 10.1186/1471-2474-7-65] [Medline: 16901354]
51. Hartvigsen J, Davidsen M, Hestbaek L, Sogaard K, Roos EM. Patterns of musculoskeletal pain in the population: a latent class analysis using a nationally representative interviewer-based survey of 4817 Danes. Eur J Pain 2013 Mar;17(3):452-460. [doi: 10.1002/j.1532-2149.2012.00225.x] [Medline: 23042697]
52. Hayashi K, Arai YCP, Morimoto A, et al. Associations between pain drawing and psychological characteristics of different body region pains. Pain Pract 2015 Apr;15(4):300-307. [doi: 10.1111/papr.12173] [Medline: 24571521]
53. Kapitza C, Lüdtke K, Tampin B, Ballenberger N. Application and utility of a clinical framework for spinally referred neck-arm pain: a cross-sectional and longitudinal study protocol. PLoS One 2020;15(12):e0244137. [doi: 10.1371/journal.pone.0244137] [Medline: 33370389]
54. Kennedy DJ, Mattie R, Nguyen Q, Hamilton S, Conrad B. Glenohumeral joint pain referral patterns: a descriptive study. Pain Med 2015 Aug;16(8):1603-1609. [doi: 10.1111/pme.12797] [Medline: 26184392]
55. Hoshino H, Sasaki N, Ide K, Yamato Y, Watanabe Y, Matsuyama Y. Effect of central sensitization inventory on the number of painful sites and pain severity in a Japanese regional population cohort. J Orthop Sci 2022 Jul;27(4):929-934. [doi: 10.1016/j.jos.2021.05.003] [Medline: 34120827]
56. Hsieh PH, Chang Y, Chen DW, Lee MS, Shih HN, Ueng SWN. Pain distribution and response to total hip arthroplasty: a prospective observational study in 113 patients with end-stage hip disease. J Orthop Sci 2012 May;17(3):213-218. [doi: 10.1007/s00776-012-0204-1] [Medline: 22427017]
57. Huang P, Sengupta DK. How fast pain, numbness, and paresthesia resolves after lumbar nerve root decompression: a retrospective study of patient's self-reported computerized pain drawing. Spine (Phila Pa 1976) 2014 Apr 15;39(8):E529-E536. [doi: 10.1097/BRS.0000000000000240] [Medline: 24480941]
58. Knox PJ, Simon CB, Pohlig RT, et al. Movement-evoked pain versus widespread pain: a longitudinal comparison in older adults with chronic low back pain from the Delaware Spine Studies. J Pain 2023 Jun;24(6):980-990. [doi: 10.1016/j.jpain.2023.01.012] [Medline: 36706887]
59. Lam K, Peolsson A, Soldini E, et al. Larger pain extent is associated with greater pain intensity and disability but not with general health status or psychosocial features in patients with cervical radiculopathy. Medicine (Baltimore) 2021 Feb 26;100(8):e23718. [doi: 10.1097/MD.0000000000023718] [Medline: 33663038]
60. Larsen LH, Hirata RP, Graven-Nielsen T. Pain-evoked trunk muscle activity changes during fatigue and DOMS. Eur J Pain 2017 May;21(5):907-917. [doi: 10.1002/ejp.993] [Medline: 28106312]
61. Larsen LH, Hirata RP, Graven-Nielsen T. Reorganized trunk muscle activity during multidirectional floor perturbations after experimental low back pain: a comparison of bilateral versus unilateral pain. J Pain 2016 Feb;17(2):223-235. [doi: 10.1016/j.jpain.2015.10.012] [Medline: 26548972]
62. MacDowall A, Robinson Y, Skeppholm M, Olerud C. Anxiety and depression affect pain drawings in cervical degenerative disc disease. Ups J Med Sci 2017 Jun;122(2):99-107. [doi: 10.1080/03009734.2017.1319441] [Medline: 28503982]
63. Masferrer R, Prendergast V, Hagell P. Colored pain drawings: preliminary observations in a neurosurgical practice. Eur J Pain 2003;7(3):213-217. [doi: 10.1016/S1090-3801(02)00113-1] [Medline: 12725843]
64. McCormick ZL, Sperry BP, Boody BS, et al. Pain location and exacerbating activities associated with treatment success following basivertebral nerve ablation: an aggregated cohort study of multicenter prospective clinical trial data. Pain Med 2022 Jul 20;23(Suppl 2):S14-S33. [doi: 10.1093/pm/pnac069] [Medline: 35856332]
65. O'Neill S, Jensen TS, Kent P. Computerized quantification of pain drawings. Scand J Pain 2019 Dec 18;20(1):175-189. [doi: 10.1515/sjpain-2019-0082] [Medline: 31605599]
66. Ohnmeiss DD. Repeatability of pain drawings in a low back pain population. Spine (Phila Pa 1976) 2000 Apr 15;25(8):980-988. [doi: 10.1097/00007632-200004150-00014] [Medline: 10767812]
67. Ohnmeiss DD, Vanharanta H, Ekholm J. Degree of disc disruption and lower extremity pain. Spine (Phila Pa 1976) 1997 Jul 15;22(14):1600-1605. [doi: 10.1097/00007632-199707150-00015] [Medline: 9253095]

XSL•FO
RenderX

68. Ohrbach R, Sharma S, Fillingim RB, Greenspan JD, Rosen JD, Slade GD. Clinical characteristics of pain among five chronic overlapping pain conditions. J Oral Facial Pain Headache 2020;34(Suppl):s29-s42. [doi: 10.11607/ofph.2573] [Medline: 32975539]
69. Pfingsten M, Baller M, Liebeck H, Strube J, Hildebrandt J, Schöps P. Psychometric properties of the pain drawing and the Ransford technique in patients with chronic low back pain. Schmerz 2003 Oct;17(5):332-340. [doi: 10.1007/s00482-003-0223-0] [Medline: 14513340]
70. Prins MR, van der Wurff P, Groen GJ. Chronic low back pain patients with accompanying leg pain: the relationship between pain extent and pain intensity, disability and health status. J Back Musculoskelet Rehabil 2013;26(1):55-61. [doi: 10.3233/BMR-2012-00350] [Medline: 23411649]
71. Provenzano DA, Fanciullo GJ, Jamison RN, McHugo GJ, Baird JC. Computer assessment and diagnostic classification of chronic pain patients. Pain Med 2007 Oct;8 Suppl 3:S167-S175. [doi: 10.1111/j.1526-4637.2007.00379.x] [Medline: 17877526]
72. Ris I, Barbero M, Falla D, et al. Pain extent is more strongly associated with disability, psychological factors, and neck muscle function in people with non-traumatic versus traumatic chronic neck pain: a cross sectional study. Eur J Phys Rehabil Med 2019 Feb;55(1):71-78. [doi: 10.23736/S1973-9087.18.04977-8] [Medline: 30156084]
73. Rising DW, Bennett BC, Hursh K, Plesh O. Reports of body pain in a dental student population. J Am Dent Assoc 2005 Jan;136(1):81-86. [doi: 10.14219/jada.archive.2005.0032] [Medline: 15693502]
74. Sorensen CJ, Johnson MB, Callaghan JP, George SZ, Van Dillen LR. Validity of a paradigm for low back pain symptom development during prolonged standing. Clin J Pain 2015 Jul;31(7):652-659. [doi: 10.1097/AJP.0000000000000148] [Medline: 25171636]
75. Southerst D, Stupar M, Côté P, Mior S, Stern P. The reliability of measuring pain distribution and location using body pain diagrams in patients with acute whiplash-associated disorders. J Manipulative Physiol Ther 2013 Sep;36(7):395-402. [doi: 10.1016/j.jmpt.2013.05.023] [Medline: 23891480]
76. Spyridonis F, Ghinea G. A pilot study to examine the relationship of 3D pain drawings with objective measures in mobility impaired people suffering from low back-pain. Annu Int Conf IEEE Eng Med Biol Soc 2010;2010:3895-3898. [doi: 10.1109/IEMBS.2010.5627668] [Medline: 21097077]
77. Sator-Katzenschlager S, Fiala K, Kress HG, et al. Subcutaneous target stimulation (STS) in chronic noncancer pain: a nationwide retrospective study. Pain Pract 2010;10(4):279-286. [doi: 10.1111/j.1533-2500.2009.00351.x] [Medline: 20230450]
78. Tesarz J, Gerhardt A, Leisner S, Janke S, Treede RD, Eich W. Distinct quantitative sensory testing profiles in nonspecific chronic back pain subjects with and without psychological trauma. Pain 2015 Apr;156(4):577-586. [doi: 10.1097/01.j.pain.0000460350.30707.8d] [Medline: 25790450]
79. Vingård E, Mortimer M, Wiktorin C, et al. Seeking care for low back pain in the general population: a two-year follow-up study: results from the MUSIC-Norrtälje Study. Spine (Phila Pa 1976) 2002 Oct 1;27(19):2159-2165. [doi: 10.1097/00007632-200210010-00016] [Medline: 12394932]
80. Vogel S, Magerl W, Treede RD, Schilder A. Dose-dependent pain and pain radiation after chemical stimulation of the thoracolumbar fascia and multifidus muscle: a single-blinded, cross-over study revealing a higher impact of fascia stimulation. Life (Basel) 2022 Feb 25;12(3):340. [doi: 10.3390/life12030340] [Medline: 35330091]
81. Wand BM, Keeves J, Bourgoin C, et al. Mislocalization of sensory information in people with chronic low back pain: a preliminary investigation. Clin J Pain 2013 Aug;29(8):737-743. [doi: 10.1097/AJP.0b013e318274b320] [Medline: 23835768]
82. Tucker KJ, Fels M, Walker SR, Hodges PW. Comparison of location, depth, quality, and intensity of experimentally induced pain in 6 low back muscles. Clin J Pain 2014 Sep;30(9):800-808. [doi: 10.1097/AJP.0000000000000026] [Medline: 25098553]
83. van den Hoven LHJ, Gorter KJ, Picavet HSJ. Measuring musculoskeletal pain by questionnaires: the manikin versus written questions. Eur J Pain 2010 Mar;14(3):335-338. [doi: 10.1016/j.ejpain.2009.06.002] [Medline: 19699125]
84. Weßollek K, Kowark A, Czaplik M, Rossaint R, Kowark P. Pain drawing as a screening tool for anxiety, depression and reduced health-related quality of life in back pain patients: a cohort study. PLoS One 2021;16(10):e0258329. [doi: 10.1371/journal.pone.0258329] [Medline: 34634060]
85. Faucett J, Meyers J, Tejeda D, Janowitz I, Miles J, Kabashima J. An instrument to measure musculoskeletal symptoms among immigrant Hispanic farmworkers: validation in the nursery industry. J Agric Saf Health 2001 Aug;7(3):185-198. [doi: 10.13031/2013.5442] [Medline: 11587195]
86. Albarran J, Durham B, Gowers J, Dwight J, Chappell G. Is the radiation of chest pain a useful indicator of myocardial infarction? A prospective study of 541 patients. Accid Emerg Nurs 2002 Jan;10(1):2-9. [doi: 10.1054/aaen.2001.0304] [Medline: 11998580]
87. Cooper G, Bailey B, Bogduk N. Cervical zygapophysial joint pain maps. Pain Med 2007;8(4):344-353. [doi: 10.1111/j.1526-4637.2006.00201.x] [Medline: 17610457]
88. Ohnmeiss DD, Vanharanta H, Ekholm J. Relation between pain location and disc pathology: a study of pain drawings and CT/discography. Clin J Pain 1999 Sep;15(3):210-217. [doi: 10.1097/00002508-199909000-00008] [Medline: 10524474]

XSL•FO
RenderX

89. Ohnmeiss DD, Vanharanta H, Ekholm J. Relationship of pain drawings to invasive tests assessing intervertebral disc pathology. Eur Spine J 1999;8(2):126-131. [doi: 10.1007/s005860050141] [Medline: 10333151]
90. Ohnmeiss DD, Vanharanta H, Guyer RD. The association between pain drawings and computed tomographic/discographic pain responses. Spine (Phila Pa 1976) 1995 Mar 15;20(6):729-733. [doi: 10.1097/00007632-199503150-00015] [Medline: 7604350]
91. Spyridonis F, Ghinea G. 3-D pain drawings and seating pressure maps: relationships and challenges. IEEE Trans Inf Technol Biomed 2011 May;15(3):409-415. [doi: 10.1109/TITB.2011.2107578] [Medline: 21257385]
92. Visscher CM, Lobbezoo F, de Boer W, van der Meulen M, Naeije M. Psychological distress in chronic craniomandibular and cervical spinal pain patients. Eur J Oral Sci 2001 Jun;109(3):165-171. [doi: 10.1034/j.1600-0722.2001.00008.x] [Medline: 11456346]
93. Abichandani D, Barbero M, Cescon C, et al. Can people with chronic neck pain recognize their own digital pain drawing? Pain Physician 2020 Mar;23(2):E231-E240. [doi: 10.36076/ppj.2020/23/E231] [Medline: 32214308]
94. Clark EM, Hutchinson AP, McCloskey EV, et al. Lateral back pain identifies prevalent vertebral fractures in post-menopausal women: cross-sectional analysis of a primary care-based cohort. Rheumatology (Oxford) 2010 Mar;49(3):505-512. [doi: 10.1093/rheumatology/kep414] [Medline: 20015975]
95. Killinger KA, Boura JA, Peters KM. Pain in interstitial cystitis/bladder pain syndrome: do characteristics differ in ulcerative and non-ulcerative subtypes? Int Urogynecol J 2013 Aug;24(8):1295-1301. [doi: 10.1007/s00192-012-2003-9] [Medline: 23208005]
96. Nilsson-Wikmar L, Pilo C, Pahlbäck M, Harms-Ringdahl K. Perceived pain and self-estimated activity limitations in women with back pain post-partum. Physiother Res Int 2003;8(1):23-35. [doi: 10.1002/pri.269] [Medline: 12701463]
97. Sturesson B, Udén G, Udén A. Pain pattern in pregnancy and “catching” of the leg in pregnant women with posterior pelvic pain. Spine (Phila Pa 1976) 1997 Aug 15;22(16):1880-1883. [doi: 10.1097/00007632-199708150-00013] [Medline: 9280024]
98. Türp JC, Kowalski CJ, O’Leary N, Stohler CS. Pain maps from facial pain patients indicate a broad pain geography. J Dent Res 1998 Jun;77(6):1465-1472. [doi: 10.1177/00220345980770061101] [Medline: 9649175]
99. Louw A, Schmidt SG, Louw C, Puentedura EJ. Moving without moving: immediate management following lumbar spine surgery using a graded motor imagery approach: a case report. Physiother Theory Pract 2015;31(7):509-517. [doi: 10.3109/09593985.2015.1060656] [Medline: 26395828]
100. Waling K, Sundelin G, Ahlgren C, Järvholm B. Perceived pain before and after three exercise programs—a controlled clinical trial of women with work-related trapezius myalgia. Pain 2000 Mar;85(1-2):201-207. [doi: 10.1016/s0304-3959(99)00265-1] [Medline: 10692619]
101. Petersen EJ, Thurmond SM. Differential diagnosis in a patient presenting with both systemic and neuromusculoskeletal pathology: resident’s case problem. J Orthop Sports Phys Ther 2018 Jun;48(6):496-503. [doi: 10.2519/jospt.2018.7652] [Medline: 29406836]
102. Sandmark H, Nisell R. Measurement of pain among electricians with neck dysfunction. Scand J Rehabil Med 1994 Dec;26(4):203-209. [Medline: 7878395]
103. Türp JC, Kowalski CJ, Stohler CS. Temporomandibular disorders--pain outside the head and face is rarely acknowledged in the chief complaint. J Prosthet Dent 1997 Dec;78(6):592-595. [doi: 10.1016/s0022-3913(97)70010-6] [Medline: 9421788]
104. Fukui S, Ohseto K, Shiotani M. Patterns of pain induced by distending the thoracic zygapophyseal joints. Reg Anesth 1997;22(4):332-336. [doi: 10.1016/s1098-7339(97)80007-7] [Medline: 9223198]
105. Jefferson JR, McGrath PJ. Back pain and peripheral joint pain in an industrial setting. Arch Phys Med Rehabil 1996 Apr;77(4):385-390. [doi: 10.1016/s0003-9993(96)90089-1] [Medline: 8607764]
106. Hüllemann P, Keller T, Kabelitz M, Freynhagen R, Tölle T, Baron R. Pain drawings improve subgrouping of low back pain patients. Pain Pract 2017 Mar;17(3):293-304. [doi: 10.1111/papr.12470] [Medline: 27334429]
107. Kloimstein H, Likar R, Kern M, et al. Peripheral nerve field stimulation (PNFS) in chronic low back pain: a prospective multicenter study. Neuromodulation 2014 Feb;17(2):180-187. [doi: 10.1111/ner.12139] [Medline: 24320718]
108. Lorenc T, Gołębiowski M, Syganiec D, Glinkowski WM. Associations between patient report of pain and intervertebral foramina changes visible on axial-loaded lumbar magnetic resonance imaging. Diagnostics (Basel) 2022 Feb 23;12(3):563. [doi: 10.3390/diagnostics12030563] [Medline: 35328116]
109. Lee H, Nicholoson LL, Adams RD, Bae SS. Body chart pain location and side-specific physical impairment in subclinical neck pain. J Manipulative Physiol Ther 2005 Sep;28(7):479-486. [doi: 10.1016/j.jmpt.2005.07.004] [Medline: 16182021]
110. Mann NH III, Brown MD, Enger I. Expert performance in low-back disorder recognition using patient pain drawings. J Spinal Disord 1992 Sep;5(3):254-259. [doi: 10.1097/00002517-199209000-00002] [Medline: 1387821]
111. Ohlund C, Eek C, Palmbald S, Areskoug B, Nachemson A. Quantified pain drawing in subacute low back pain. Validation in a nonselected outpatient industrial sample. Spine (Phila Pa 1976) 1996 May 1;21(9):1021-1030. [doi: 10.1097/00007632-199605010-00005] [Medline: 8724085]
112. Pande KC, Khurjekar K, Kanikdaley V. Correlation of low back pain to a high-intensity zone of the lumbar disc in Indian patients. J Orthop Surg (Hong Kong) 2009 Aug;17(2):190-193. [doi: 10.1177/230949900901700214] [Medline: 19721150]

XSL•FO
RenderX

113. Parker H, Wood PL, Main CJ. The use of the pain drawing as a screening measure to predict psychological distress in chronic low back pain. Spine (Phila Pa 1976) 1995 Jan 15;20(2):236-243. [doi: 10.1097/00007632-199501150-00022] [Medline: 7716631]
114. Potier T, Tims E, Kilbride C, Rantell K. Evaluation of an evidence based quality improvement innovation for patients with musculoskeletal low back pain in an accident and emergency setting. BMJ Qual Improv Rep 2015;4(1):u205903.w2411. [doi: 10.1136/bmjquality.u205903.w2411] [Medline: 26734383]
115. Sanders NW, Mann NH III, Spengler DM. Pain drawing scoring is not improved by inclusion of patient-reported pain sensation. Spine (Phila Pa 1976) 2006 Nov 1;31(23):2735-2741. [doi: 10.1097/01.brs.0000244674.99258.f9] [Medline: 17077744]
116. Serif T, Ghinea G. Recording of time-varying back-pain data: a wireless solution. IEEE Trans Inf Technol Biomed 2005 Sep;9(3):447-458. [doi: 10.1109/titb.2005.847514] [Medline: 16167699]
117. Vucetic N, Määttänen H, Svensson O. Pain and pathology in lumbar disc hernia. Clin Orthop Relat Res 1995 Nov(320):65-72. [Medline: 7586844]
118. Bryner P. Extent measurement in localised low-back pain: a comparison of four methods. Pain 1994 Nov;59(2):281-285. [doi: 10.1016/0304-3959(94)90081-7] [Medline: 7892026]
119. Mann NH III, Brown MD, Hertz DB, Enger I, Tompkins J. Initial-impression diagnosis using low-back pain patient pain drawings. Spine (Phila Pa 1976) 1993 Jan;18(1):41-53. [doi: 10.1097/00007632-199301000-00008] [Medline: 8434324]
120. Sanders NW, Mann NH. Automated scoring of patient pain drawings using artificial neural networks: efforts toward a low back pain triage application. Comput Biol Med 2000 Sep;30(5):287-298. [doi: 10.1016/s0010-4825(00)00013-5] [Medline: 10913774]
121. Takata K, Hirotani H. Pain drawing in the evaluation of low back pain. Int Orthop 1995;19(6):361-366. [doi: 10.1007/BF00178350] [Medline: 8567153]
122. Mann NH III, Brown MD, Enger I. Statistical diagnosis of lumbar spine disorders using computerized patient pain drawings. Comput Biol Med 1991;21(6):383-397. [doi: 10.1016/0010-4825(91)90040-g] [Medline: 1838724]
123. Gaetano J, van der Zwan R, Oxner M, et al. Converging evidence of ubiquitous male bias in human sex perception. PLoS One 2016;11(2):e0148623. [doi: 10.1371/journal.pone.0148623] [Medline: 26859570]
124. Gandolfo M, Downing PE. Asymmetric visual representation of sex from human body shape. Cognition 2020 Dec;205:104436. [doi: 10.1016/j.cognition.2020.104436] [Medline: 32919115]
125. Mendelsohn KD, Nieman LZ, Isaacs K, Lee S, Levison SP. Sex and gender bias in anatomy and physical diagnosis text illustrations. JAMA 1994 Oct 26;272(16):1267-1270. [doi: 10.1001/jama.1994.03520160051042] [Medline: 7933371]
126. Morgan S, Plaisant O, Lignier B, Moxham BJ. Medical students and professional anatomists do not perceive gender bias within imagery featuring anatomy. Clin Anat 2017 Sep;30(6):711-732. [doi: 10.1002/ca.22922] [Medline: 28547745]
127. Morgan S, Plaisant O, Lignier B, Moxham BJ. Sexism and anatomy, as discerned in textbooks and as perceived by medical students at Cardiff University and University of Paris Descartes. J Anat 2014 Mar;224(3):352-365. [doi: 10.1111/joa.12070] [Medline: 23781866]
128. Parker R, Larkin T, Cockburn J. A visual analysis of gender bias in contemporary anatomy textbooks. Soc Sci Med 2017 May;180:106-113. [doi: 10.1016/j.socscimed.2017.03.032] [Medline: 28343109]
129. Beresheim A, Zepeda D, Pharel M, Soy T, Wilson AB, Ferrigno C. Anatomy's missing faces: an assessment of representation gaps in atlas and textbook imagery. Anat Sci Educ 2024;17(5):1055-1070. [doi: 10.1002/ase.2432] [Medline: 38695348]

Abbreviations

PD: pain drawing
SAGER: Sex and Gender Equity in Research

Edited by A Kushniruk; submitted 17.Apr.2025; peer-reviewed by B Pervis; accepted 28.Dec.2025; published 09.Feb.2026.

Please cite as:
Schläpfer S, Murphy C, Lukic YX, Campos-Asensio C, Colacicco G, Lauber BA, Masneuf S, Nyirö L, Kajdi GW, Sutter R, Laux CJ, Ackermann P, López-Alcalde J, Witt CM
Sex Representation and User Preferences in Pain Drawing Body Charts in Back Pain Research: Multimethod Study
JMIR Hum Factors 2026;13:e76175
URL: https://humanfactors.jmir.org/2026/1/e76175
doi:10.2196/76175

© Sonja Schläpfer, Casey Murphy, Yanick X Lukic, Concepción Campos-Asensio, Giovanni Colacicco, Beatrice A Lauber, Sophie Masneuf, Luana Nyirö, Georg W Kajdi, Reto Sutter, Christoph J Laux, Philipp Ackermann, Jesús López-Alcalde, Claudia

XSL•FO
RenderX

M Witt. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 9.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Parental Influence on Children's Media Use in South Korea: National Population-Based Study

Ji Young Kim[1], PhD; Ah Jung Yang[2], PhD; Hye Eun Lee[3], PhD

[1]Department of Psychology, School of Behavioral Sciences and Education, Pennsylvania State University, Middletown, PA, United States

[2]Communication & Media Research Center, College of Social Science, Ewha Womans University, Seoul, Republic of Korea

[3]Department of Communication & Media, College of Social Science, Ewha Womans University, #403 Ewha-POSCO, 52 Ewhayeodae-gil, Seodaemun-gu, Seoul, Republic of Korea

Corresponding Author:
Hye Eun Lee, PhD
Department of Communication & Media, College of Social Science, Ewha Womans University, #403 Ewha-POSCO, 52 Ewhayeodae-gil, Seodaemun-gu, Seoul, Republic of Korea

Abstract

Background: To better understand the effects of media use on children, it is essential to examine the various factors influencing the media use of digital native children. In the situational context, parental media usage, parents' attitudes toward media, and parenting styles have all been identified as significant factors influencing children's media use. This study focuses on the key factors and examines these relationships in greater depth, drawing on existing research to understand their impact on the media usage patterns of digital native children.

Objective: This study examines parental influences related to young children's media use in Korea over a 3-year period (2022 - 2024) using independent, nationally representative cohorts.

Methods: Using multigroup structural equation modeling, we analyzed data from 3 independent parent–reported cohorts (for 2022, n=1058; for 2023, n=1020; for 2024, n=1020) to investigate how parental media habits, attitudes, and distinct parenting styles predict children's daytime and nighttime media consumption.

Results: The online survey results revealed that parental media time, particularly for mothers, consistently correlated with higher levels of children's daytime media use (β=.002-.003). Positive parental attitudes toward media increased children's daytime media use (β=.028-.102), whereas negative attitudes had a limited effect (β=−.069-.140). Among the 7 parenting styles, positive parenting consistently reduced children's daytime media use in 2022 and 2023 (β=−.228 for 2022, β=−.215 for 2023), but harsh punishment emerged as the strongest factor in daytime media use in 2024 (β=−.078 for 2022, β=−.090 for 2023, and β=−.072 for 2024). Notably, parenting styles showed no significant effect on children's nighttime media use throughout the study, suggesting that parental influence may be more effective during daytime hours.

Conclusions: This analysis extends existing research by differentiating media use patterns across time periods and highlights the evolving influence of parenting styles. These findings have implications for the development of targeted parental guidelines for managing young children's media exposure, especially as digital media continues to become a pervasive part of daily life.

(JMIR Hum Factors 2026;13:e75292) doi:10.2196/75292

KEYWORDS

children media use; parent attitude; parent media use; parental mediation; parenting style

Introduction

Parental Influences on Children's Media Use in South Korea

An increasing number of children access media at younger ages, with many actively consuming it. In the United States, 70% of children under the age of 1 year and 91% of children aged 2-3 years access media several times per week [1]. Similarly, in South Korea, approximately 60% of children aged 3-9 years are exposed to television before their first birthday, and 30% begin using smartphones before the age of 1 year [2]. This growing trend is in contrast with the recommendation by the American Academy of Pediatrics [3], which advises against media use for children under the age of 2 years [4]. Despite the recommendation, children's media consumption remains high. The typical 8- to 10-year-old children spend almost 8 hours daily across different media, and this number increases for older children and teenagers to more than 11 hours per day [1].

While media may not be a leading cause of major health concerns, previous research suggests that media exposure contributes to various risks and health problems [3]. For example, younger children's media use, as early as preschool

XSL•FO
RenderX

years, has been associated with developmental difficulties such as reduced physical fitness and psychosocial health [5,6]. Specifically, engagement in risky online social behaviors [7], poor academic performance [8], obesity [9,10], reduced quality of life [11], and impaired executive function [12] have been associated with increased children's media use. Additionally, previous studies have shown that excessive television viewing can lead to cognitive, language, and social-emotional delays [13-16]. Research has further shown that infants and toddlers gain more developmentally beneficial skills through play with physical objects rather than screen-based devices [17,18].

Numerous longitudinal [19] and cross-sectional studies [5,13,20] showed that children with greater screen time had a higher level of behavioral difficulties. A longitudinal study investigating children between the ages of 2 and 6 years showed an association between increased media use and poorer well-being outcomes, even though the results varied for boys and girls [13]. Additionally, each extra hour of television viewing per week between the ages of 2 and 4 years was associated with a 5% increase in body mass index at the age of 10 years, a 9% - 10% rise in the consumption of soda and snacks, and a 13% reduction in outdoor weekend activities [21]. A study from South Korea with survey data from mothers and teachers of 5- and 6-year-old children showed that increased use of smart devices was associated with increased levels of aggression [22]. Altogether, the findings suggest that children's excessive consumption of media may have adverse developmental and behavioral effects.

To better understand the effects of media use on children, it is essential to examine the various factors influencing the media use of digital native children (ie, children who have access to and utilize a wide range of media and platforms from birth) [23]. One key factor is the physical environment, or in other words, the advancements in digital media technology that enable children to use media freely without restrictions of time and space through traditional means such as television as well as portable devices like smartphones and tablet PCs. Another key factor is the situational context, such as the influence of parents, given that parents shape their children's behaviors and attitudes as primary role models [24-26]. Parental media usage [27-29], attitudes toward media [30,31], and parenting styles [26,32] have all been identified as significant factors influencing children's media use. In addition, the impact of these factors varies according to the developmental stage or age of the children [25,26,32]. This study seeks to focus on the second key factor and examine these relationships in greater depth, drawing on existing research to understand their impact on the media usage patterns of digital native children.

Although extensive research has examined children's overall media or screen time, much less is known about how parental factors influence media use across different times of day. This is a critical gap because daytime and nighttime media use may have distinct developmental implications (eg, school readiness, sleep quality, and behavioral regulation). Moreover, little evidence exists on how these influences shift across post-COVID years when children's reliance on screen-based media has been both normalized and expanded to include educational contexts [33,34]. This study addresses this problem by examining parental media use, attitudes, and parenting styles as predictors of children's daytime and nighttime media use in a large, national Korean sample drawn from three independent cohorts across 3 years (2022 - 2024). In this study, "media use" is defined operationally as children's screen-based engagement, including television, smartphones, tablets, and computers, as reported by parents in a 24-hour matrix of daily activities. Conceptually, we focus on electronic and online media rather than print or offline formats, as these account for the majority of children's daily exposure in contemporary contexts.

Parents' Media Use and Social Learning Theory

Children's screen time can be explained by Bandura's social learning theory [35], which posits that learning and behavior are shaped by observing others in one's environment. Young children learn by observing their parents, siblings, and others around them, picking up on daily habits, social interactions, and responses to various situations, including media use in the home. When parents model media behaviors, their children are likely to replicate those patterns. For instance, a national survey found that children often imitate their fathers or older siblings by playing with game controllers, even when they lack the skills to use them effectively [36].

Parents' prolonged or habitual consumption of media at home may serve as a role model for children's media use. Research indicates that parents' media use and habitual patterns of consumption are associated with increased media use time and habitual use among children [37]. Notably, children's media use was more closely related to their parents' habitual consumption of media than to parental restrictions on media use. Further, studies have shown that children's screen time tends to be higher when their mothers report substantial media use [38,39]. These findings suggest that parents' media use may significantly influence children's ability to regulate their own media consumption. Accordingly, we propose:

- H1: A parent's screen time will be positively related to a child's media use.

Parent Attitudes and Media

Parent attitudes are another important factor in children's screen time. Parents play a critical role in establishing rules within the household, with their influence on media consumption being particularly pronounced in younger children. Parent attitudes toward media, whether positive or negative, shape the rules governing media use, thereby directly influencing the amount of time children spend engaging with media [26]. Given their substantial influence on their children's behaviors [40], parents often modify their behaviors as a strategy to affect changes in their children's behaviors [41].

In the United States, a survey focused on electronic media use among children aged 6 months to 6 years indicated that parents had varied attitudes toward media use [36]. Some parents encouraged media use, finding it beneficial for children and useful for household management, such as keeping children occupied while doing chores. While similar proportions of parents felt that television had both positive (38%) and negative (31%) effects on learning, the majority supported computers as beneficial (70%) but video games as detrimental (49%) [29]. It is possible that parents aim to help their children maximize the

XSL•FO
RenderX

benefits of the rich resources offered by video media while simultaneously shielding them from harmful content [42].

In general, the children of parents who view media positively tend to have higher screen time. Among preschoolers, parent attitudes and beliefs about media were strong indicators of children's screen time [30]. For younger (0 - 2 y) and older (5 - 6 y) children, positive parent attitudes have been shown to significantly contribute to exceeding the recommended screen time limits set by the American Academy of Pediatrics [3,43], particularly in relation to television viewing [44]. Although parent attitudes significantly influenced the television and computer use of children, their influence on mobile devices was less pronounced [31].

Additionally, parental rules based on positive attitudes tend to support media use in the home, while negative attitudes encourage restrictions. The more critically parents evaluate media, the more likely they are to engage positively in parental mediation (active or restrictive) to mitigate the negative effects of their children's media exposure [41]. Parental mediation has been shown to effectively influence children's responses to media exposure [45], and experts have even advocated for increasing the level of parental mediation provided by parents [46]. Therefore, we hypothesize:

- H2: Parents with a positive attitude toward media will have children with higher levels of media use.
- H3: Parents with a negative attitude toward media will have children with lower levels of media use.

Parenting Style and Media Use

Finally, parents considerably contribute to children's media use, indicating that parents may mediate children's media use and, in turn, the overall developmental outcome. For example, while children with increased media use showed a decreased level of prosocial behavior, the level of parent-child interactions mediated the association between the children's media use and level of prosocial behavior [19]. Specifically, parenting style—or specific sets of parental mediation strategies and behaviors—may explain children's varying levels of media use. Previous studies have shown that parental control over children's access to smart media influences the amount of time younger children spend using such media [47]. Research in Korea has further indicated that among parental mediation strategies, supervisory control tends to be more pronounced before the age of 6 years but diminishes after the age of 7 years. In particular, for children aged 5-6 years, multiple factors—including media use patterns, self-regulatory abilities, and parental supervision—exert significant influence. This suggests that age-specific approaches to parents' mediation are necessary [48]. Research also indicates that mediation practices vary not only in form (eg, restrictive, active, or co-use) but also in effectiveness, depending on the child's age. For younger children, restrictive mediation (eg, setting time limits) has been shown to reduce overall screen exposure, whereas older children and adolescents benefit more from active mediation approaches that involve discussion, explanation, and joint engagement [25,32,49-51].

Baumrind [52] and Maccoby and Martin [53] identified four different parenting styles: authoritative, authoritarian, indulgent, and neglectful. Each parenting style varied in their levels of demandingness and responsiveness. Generally, studies showed that children with authoritative parents (ie, highly demanding and responsive) had the most positive developmental outcome, while children with authoritarian (ie, highly demanding but not responsive) and permissive (ie, not demanding but highly responsive) parents had more negative outcomes.

The Ghent Parental Behavior Scale, developed by Leeuwen and Vermulst [54], categorizes parenting strategies into 9 distinct domains: positive parenting, autonomy, rules, monitoring, discipline, harsh punishment, material reward, inconsistent discipline, and ignoring. A material reward is a tangible incentive, such as a toy, snack, small gift, or even cash, provided to a child in exchange for a specific behavior or compliance. Positive parenting styles, such as permissive and neglectful parenting, are characterized by strategies including autonomy, positive parenting, and rules. In contrast, negative parenting styles, such as authoritative or authoritarian approaches, are associated with parenting strategies including monitoring, harsh punishment, material rewards, and ignoring [54].

A recent study by Lee et al [26] showed that a child's daytime media use between the ages of 4 and 6 years significantly increased for parents who gave more autonomy to their children (ie, permissive and neglectful parenting style), and a child's nighttime media use decreased with the parenting style of discipline (ie, authoritative and authoritarian parenting style). Lee et al also found that parents' use of material rewards was the strongest predictor of the child's nighttime media use, as the child's nighttime media use significantly increased when parents practiced material reward. Similarly, other studies showed that children between the ages of 10 and 11 years exhibited higher levels of screen exposure when they had parents with lower levels of control over their children [27,32]. The parental behavior in this example aligns with a permissive and neglectful parenting style, where parents provide more autonomy to their children and incorporate positive parenting. On the contrary, studies showed that parents who exercise higher levels of control and support would engage in behaviors or strategies employing active and restrictive mediation [32,55]. Specific strategies would include monitoring and rules, discipline, harsh punishment, and material reward. Collectively, the literature suggests that children with permissive and neglectful parents would be more likely to have higher levels of media use, while children with authoritarian and authoritative parents would be more likely to have lower levels of media use [26].

Despite the clear implications of the parent's role in shaping children's media use, previous studies have focused on descriptive and correlational analyses. As Lee et al [26] suggested, further investigation on the relationship between parental determinants, such as parenting style and children's media use, is needed. The primary goal of the study is to determine the relationship between parenting style and children's media use across various age groups. The researcher distributed surveys to parents because young children cannot reliably self-report their daily media use, and parent reports are generally considered appropriate for estimating children's media use, such

XSL•FO
RenderX

as video and app games, computer use, and television viewing [56]. Based on previous findings that parenting style may potentially have important effects on children's media use, we hypothesize the following:

- H4: Specific parenting styles will be associated with children's level of media use.

The secondary goal of the study is to extend the findings of Lee et al [26] by replicating the previous findings. Therefore, we incorporated the 2 hypotheses related to parenting style used in Lee et al [26]:

- H5: Children who have parents with permissive and neglectful parenting styles (ie, exercise positive parenting and give autonomy to children) will have higher levels of media use.
- H6: Children who have parents with authoritative and authoritarian parenting styles (ie, monitoring, rules, discipline, harsh punishment, and material reward) will have lower levels of media use.

Previous findings support a clear relationship between parenting style and children's media use. To extend the existing literature, it is imperative to identify the parental determinant that explains children's media use, and the researcher of this study aims to answer the following research question:

- RQ1: Among parent's media use, media attitudes, and parenting styles, what is the most influential factor on children's media use?

Methods

Participants

A total of 1058 parents of children aged 5-7 years, 1020 parents of children aged 6-8 years, and 1020 parents of children aged 7-9 years completed questionnaires between April 26 and May 9, 2022; October 26 and November 9, 2023; and July 31 and August 21, 2024, respectively. All participants completed the questionnaire through an online survey conducted by a Korean survey company, Macromill Embrain, which recruited participants from its national panel pool. Each year's sample primarily consisted of an independent cross-sectional group, although approximately 5% - 10% of the participants may have overlapped across years due to the nature of the panel system. However, the study did not track individuals over time, and analyses were conducted separately for each dataset. The age and sex distribution of the children was balanced across each dataset. Detailed demographic characteristics for all three groups are presented in Table 1.

XSL•FO
RenderX

Table . Demographic characteristics of participants and their child (age, gender, education level, income, employment, and family members).

Year and demographic characteristics	Participants, n (%)
2022	
Age (child) (y)	
5	359 (33.9)
6	352 (33.3)
7	347 (32.8)
Gender (child)	
Male	535 (50.6)
Female	523 (49.4)
Age (parent) (y)	
20	15 (1.4)
30	578 (54.6)
40	453 (42.8)
50	12 (1.1)
Gender (parent)	
Male	129 (12.2)
Female	929 (87.8)
Family members	
2	11 (1.0)
3	301 (28.4)
4	576 (54.4)
5	109 (10.3)
>6	61 (5.8)
Education level (father)	
High school or below	144 (13.6)
College or university	781 (73.8)
Master’s degree or above	133 (12.6)
Education level (mother)	
High school or below	108 (10.2)
College or university	834 (58.8)
Master’s degree or above	116 (11.0)
Monthly income (family) (USD)	
<4000	341 (32.2)
4000 - 5000	235 (22.2)
5000 - 6000	161 (15.2)
>6000	321 (30.3)
Employment (father)	
Full-time job	993 (93.9)
Part-time job	50 (4.7)
No job	15 (1.4)
Employment (mother)	
Full-time job	519 (49.1)
Part-time job	171 (16.2)

XSL•FO
RenderX

Year and demographic characteristics	Participants, n (%)
No job	368 (34.8)
Total	1058 (100)
2023	
Age (child) (y)	
6	340 (33.3)
7	340 (33.3)
8	340 (33.3)
Gender (child)	
Male	510 (50.0)
Female	510 (50.0)
Age (parent) (y)	
20	5 (0.5)
30	468 (45.9)
40	541 (53.0)
50	6 (0.6)
Gender (parent)	
Male	144 (14.1)
Female	876 (85.9)
Family members	
2	317 (31.1)
3	547 (53.6)
4	105 (10.3)
5	31 (3.0)
>6	20 (2.0)
Education level (father)	
High school or below	128 (12.6)
College or university	759 (74.4)
Master’s degree or above	133 (13.0)
Education level (mother)	
High school or below	90 (8.8)
College or university	803 (78.7)
Master’s degree or above	127 (12.5)
Monthly income (family) (USD)	
<4000	205 (20.1)
4000 - 5000	222 (21.8)
5000 - 6000	173 (16.9)
>6000	420 (41.2)
Employment (father)	
Full-time job	968 (94.9)
Part-time job	41 (4.0)
No job	11 (1.1)
Employment (mother)	
Full-time job	576 (56.5)

XSL•FO
RenderX

Year and demographic characteristics	Participants, n (%)
Part-time job	160 (15.7)
No job	284 (27.8)
Total	1020 (100)
2024	
Age (child) (y)	
7	340 (33.3)
8	340 (33.3)
9	340 (33.3)
Gender (child)	
Male	510 (50.0)
Female	510 (50.0)
Age (parent) (y)	
20	7 (0.7)
30	369 (36.2)
40	628 (61.6)
50	16 (1.6)
Gender (parent)	
Male	133 (13.0)
Female	887 (87.0)
Family members	
2	307 (31.1)
3	549 (53.8)
4	119 (11.7)
5	33 (3.2)
>6	12 (1.2)
Education level (father)	
High school or below	123 (12.1)
College or university	741 (72.6)
Master’s degree or above	156 (15.3)
Education level (mother)	
High school or below	82 (8.0)
College or university	810 (79.5)
Master’s degree or above	128 (12.5)
Monthly income (family) (USD)	
<4000	172 (16.9)
4000 - 5000	197 (19.3)
5000 - 6000	179 (17.5)
>6000	472 (46.3)
Employment (father)	
Full-time job	976 (95.7)
Part-time job	40 (3.9)
No job	4 (0.4)
Employment (mother)	

XSL•FO
RenderX

Year and demographic characteristics	Participants, n (%)
Full-time job	602 (59.0)
Part-time job	153 (15.0)
No job	265 (26.0)
Total	1020 (100)

Ethical Considerations

The study received ethical approval from the Institutional Review Board of Ewha Womans University (Institutional Review Board approval number: EWHA-202103-0028-01). All data were anonymized prior to analysis, and no personally identifiable information was stored or retained. Informed consent was obtained from all participants involved in the study. As the participants were members of a registered survey company panel, an honorarium was provided based on the survey response time at a rate of 100 KRW (approximately US $0.67) per minute. Participants who discontinued the survey prior to completion were also compensated with a prorated stipend corresponding to their participation time.

Instrument and Measures

Overview

The questionnaire was initially developed in English and then translated into Korean, with translation equivalence verified by bilingual researchers. In addition to the main variables, media use (ie, time spent on media via television, personal device, smartphone) and demographic information of both children and their parents were collected. Tables 2, 3, and 4 present the reliabilities and descriptive statistics of the variables, along with their correlations. Composite variables were calculated once unidimensionality and acceptable reliability were confirmed. All variables were measured with 5-point Likert scales (1=“strongly disagree” to 5=“strongly agree”) unless otherwise noted.

XSL•FO
RenderX

Table . Reliabilities, correlations, means, and standard deviations of the main variables (2022)[a].

2022	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Child's age at first use	—[b]														
Child's locus of control	–0.05	(0.75)													
Mother's media time	0.01	0.17[c]	—												
Father's media time	0.04	0.13[c]	0.62[c]	—											
Positive attitude toward media use	0.10[c]	–0.02	0.10[c]	0.05	(0.90)										
Negative attitude toward media use	0.01	0.44[c]	0.03	0.07[d]	–0.26[c]	(0.89)									
Positive parenting	–0.05	–0.19[c]	–0.06	–0.03	–0.09[c]	–0.12[c]	(0.89)								
Monitoring	0.01	–0.06[d]	–0.07[d]	–0.02	0.10[c]	0.02	0.48[c]	(0.72)							
Rules	–0.07[d]	–0.11[c]	0.01	0.02	0.01	–0.01	0.52[c]	0.44[c]	(0.84)						
Discipline	0.00	0.13[c]	0.02	0.04	0.05	0.21[c]	–0.05	0.10[c]	0.21[c]	(0.76)					
Harsh punishment	0.05	0.26[c]	0.11[c]	0.07[d]	0.10[c]	0.22[c]	–0.36[c]	–0.16[c]	–0.20[c]	0.34[c]	(0.92)				
Material reward	0.05	0.25[c]	0.11[c]	0.10[c]	0.21[c]	0.12[c]	–0.04	0.07[d]	0.00	0.21[c]	0.25[c]	(0.76)			
Autonomy	0.05	–0.10[c]	0.01	0.06	0.13[c]	–0.03	0.45[c]	0.33[c]	0.37[c]	0.03	–0.13	0.07[d]	(0.80)		
Child's daytime media use	–0.08[d]	0.34[c]	0.26[c]	0.17[c]	0.04	0.10[c]	–0.12[c]	–0.07[d]	–0.06	0.05	0.05	0.08[c]	–0.04		—
Child's nighttime media use	–0.01	0.12[d]	0.12[c]	0.11[c]	–0.05	0.02	–0.05	–0.06	–0.04	0.05	0.08[c]	0.01	–0.03	0.09**[c]	

XSL•FO
RenderX

2022	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Mean (SD; range)	3.33 (1.36; 1 - 8)	2.88 (0.82; 1 - 5)	120.67 (83.81; 0 - 600)	127.10 (86.03; 0 - 600)	2.92 (0.70; 1 - 5)	3.09 (0.68; 1 - 5)	4.01 (0.47; 1.91 - 5)	3.79 (0.60; 1 - 5)	4.12 (0.46; 1.86 - 5)	3.31 (0.66; 1 - 5)	1.71 (0.86; 1 - 4.75)	2.83 (0.84; 1 - 5)	3.83 (0.53; 2 - 5)	2.37 (1.49; 0 - 13)	0.19 (0.48; 0 - 4.29)

[a]Reliabilities, calculated using Cronbach α, are reported in parentheses on the diagonal.

[b]Not applicable.

[c]*P*<.001.

[d]*P*<.05.

XSL•FO
RenderX

Table . Reliabilities, correlations, means, and standard deviations of the main variables (2023)[a].

2023	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Child's age at first use	—[b]														
Child's locus of control	–0.04	(0.76)													
Mother's media time	0.01	0.16[c]	—												
Father's media time	–0.03	0.15[c]	0.59[c]	—											
Positive attitude toward media use	0.11[c]	–0.05	–0.02	–0.08[c]	(0.92)										
Negative attitude toward media use	–0.01	0.42	0.02	0.05	–0.22[c]	(0.90)									
Positive parenting	–0.03	–0.22[c]	–0.13[c]	–0.07[d]	0.15[c]	–0.11[c]	(0.9)								
Monitoring	–0.03	–0.05	–0.05	0.01	0.08[c]	0.05	0.50[c]	(0.69)							
Rules	–0.07[d]	–0.11[c]	–0.02	0.02	0.01	–0.01	0.54[c]	0.49[c]	(0.85)						
Discipline	–0.04	0.13[d]	0.02	0.04	0.07[d]	0.13[c]	0.04	0.23[c]	0.26[c]	(0.74)					
Harsh punishment	0.06	0.25[c]	0.12[c]	0.09[c]	0.16[c]	0.13[c]	–0.35[c]	–0.16[c]	–0.25[c]	0.25[c]	(0.92)				
Material reward	0.06	0.26[c]	0.07[d]	0.03	0.13[c]	0.16[c]	–0.08[d]	–0.00	–0.06[d]	0.17[c]	0.25[c]	(0.77)			
Autonomy	0.01	–0.13[c]	–0.04	–0.03	0.07[d]	–0.05	0.47[c]	0.31[c]	0.38[c]	0.11[c]	–0.15[c]	0.02	(0.81)		
Child's daytime media use	–0.10[c]	0.32[c]	0.25[c]	0.22[c]	–0.05	0.10[c]	–0.10[c]	–0.05	–0.01	0.04	0.02	0.05	–0.03		—
Child's nighttime media use	–0.01	0.15[c]	0.11[c]	0.10[c]	–0.03	0.04	–0.08	–0.04	–0.04	0.06	0.05	0.07[d]	–0.01	0.16[c]	

XSL•FO
RenderX

2023	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Mean (SD; range)	3.46 (1.50; 1 - 9)	2.83 (0.84; 1 - 5)	117.53 (81.51; 0 - 610)	125.65 (83.92; 0 - 600)	2.85 (0.72; 1 - 4.89)	3.16 (0.67; 1 - 5)	4.04 (0.48; 2.09 - 5)	3.90 (0.57; 1.75 - 5)	4.32 (0.50; 2.33 - 5)	3.50 (0.72; 1 - 5)	1.70 (0.83; 1 - 4.75)	2.79 (0.83; 1 - 5)	3.86 (0.52; 2 - 5)	2.35 (1.48; 0 - 10.71)	0.22 (0.46; 0 - 4)

[a]Reliabilities, calculated using Cronbach α, are reported in parentheses on the diagonal.

[b]Not applicable.

[c]*P*<.001.

[d]*P*<.05.

XSL•FO
RenderX

Table . Reliabilities, correlations, means, and standard deviations of the main variables (2024)[a].

2024	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Child's age at first use	—[b]														
Child's locus of control	–0.06	(0.76)													
Mother's media time	0.02	0.13[c]	—												
Father's media time	–0.06[d]	0.08[c]	0.54[c]	—											
Positive attitude toward media use	0.04	0.02	0.04	–0.03	(0.93)										
Negative attitude toward media use	–0.01	0.23[c]	0.01	0.01	–0.32[c]	(0.90)									
Positive parenting	–0.01	–0.11[c]	–0.11[c]	–0.07[d]	0.08[d]	–0.08[d]	(0.87)								
Monitoring	–0.02	–0.04	–0.02	0.02	0.03	0.05	0.47	(0.72)							
Rules	–0.05	–0.06	–0.02	0.04	–0.08[d]	0.03	0.54	0.45[c]	(0.92)						
Discipline	–0.02	0.10[c]	0.07[d]	0.06[d]	0.07[d]	0.18[c]	0.06	0.15[c]	0.14[c]	(0.72)					
Harsh punishment	0.02	0.11[c]	0.12[c]	0.09[c]	0.14[c]	0.12[c]	–0.29[c]	–0.14[c]	–0.24[c]	0.37[c]	(0.92)				
Material reward	0.04	0.15[c]	0.03	0.02	0.17[c]	0.14[c]	–0.17[c]	0.01	–0.14[c]	0.23[c]	0.37[c]	(0.77)			
Autonomy	0.04	–0.04	–0.06	–0.04	0.06	0.02	0.40[c]	0.22[c]	0.32[c]	0.01	–0.11	0.01	(0.80)		
Child's daytime media use	–0.13[c]	–0.23[c]	0.18[c]	0.14[c]	–0.05	0.05	–0.05	–0.05	–0.01	–0.02	0.02	0.02	0.03	–0.02	—
Child's nighttime media use	–0.08[d]	0.13[c]	0.15[c]	0.13[c]	0.04	0.01	–0.01	–0.01	0.00	0.04	0.02	0.02	0.06[d]	0.02	0.23[c]

XSL•FO
RenderX

2024	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Mean (SD; range)	3.73 (1.69; 1 - 10)	3.01 (0.61; 1 - 5)	124.52 (85.88; 0 - 600)	140.77 (87.96; 0 - 600)	2.54 (0.79; 1 - 4.89)	3.32 (0.67; 1 - 5)	4.03 (0.46; 2.27 - 5)	3.94 (0.57; 1.75 - 5)	4.43 (0.52; 2.33 - 5)	3.21 (0.60; 1 - 5)	1.73 (0.87; 1 - 5)	2.60 (0.85; 1 - 5)	3.87 (0.57; 1 - 5)	3.48 (2.50; 0 - 14)	0.34 (0.74; 0 - 10)

[a]Reliabilities, calculated using Cronbach α, are reported in parentheses on the diagonal.

[b]Not applicable.

[c]$P<.001$.

[d]$P<.05$.

Children's Age at First Media Use

Parents were asked about their child's age at first media exposure with the question, "How old was your child when he/she first started watching media content?" The response options included "less than 12 months," "1 year," "2 years," "3 years," "4 years," "5 years," "6 years," and "7 years."

Parents' Media Time

Each participant reported both their own and their spouse's media watching time. Two questions were used: "How many hours and minutes do you spend watching media on a typical weekday?" and "How many hours and minutes do you spend to watch media on a typical weekend?" Responses were averaged after weighing (ie, (weekday time × 5 days + weekend × 2 days)/7)) and converted to minutes.

Parents' Positive and Negative Attitudes Toward Media Use

Based on scales by Elias and Sulkin [57] and Nikken and Jansz [58], 9 items measured positive attitudes (eg, "I think watching media will positively influence my child's behavioral development."). Negative attitudes were measured with 2 dimensions: intellectual and social. Each dimension included 4 items, such as "I think watching media will hurt my child's creativity." for the intellectual dimension and "I believe watching media will negatively affect my child's play with friends." for the social dimension.

Parenting Styles

The Ghent Parental Behavior Scale [54] was used to measure the parenting styles. This scale has demonstrated valid structure in diverse samples. Seven dimensions were included due to reliability concerns with 2 original dimensions (inconsistent discipline and ignoring). The included dimensions are positive parenting (11 items: eg, "I make time to listen to my child when he/she wants to tell me something."), monitoring (4 items: eg, "I keep track of the friends my child is seeing."), rules (6 items: eg, "I teach my child to obey rules."), discipline (4 items: eg, "When my child does something I don't approve of, I punish him/her."), harsh punishment (4 items: eg, "I spank my child when he/she is disobedient."), material reward (3 items: eg, "I reward my child with money or a small gift for good behavior."), and autonomy (3 items: eg, "I teach my child to solve problems independently."). Items excluded for reliabilities included 1 ("I teach my child respect for the authorities.") from the rules dimension and 2 items ("When my child has been misbehaving, I give him/her a chore for punishment." and "It happens that I don't punish my child after he/she has done something that is not allowed.") from harsh punishment dimension. All items and dimensions can be referenced in Leeuwen and Vermulst's measurement study [54].

Child's Media Time

Participants selected a cell in a 24-hour matrix to indicate when a child watched media during both weekdays and weekends. Daytime use was defined as viewing between 7:00 AM and 9:00 PM, while viewing before 7:00 AM or after 9:00 PM was classified as nighttime use. This classification was based on developmental sleep guidelines: given that elementary schools in Korea typically begin at 9:00 AM, and the American Academy of Sleep Medicine [59] recommends 10 - 13 hours of sleep for children aged 3 - 5 and 9 - 12 hours for children aged 6 - 12, we estimated that 10 - 11 hours of sleep starting from 9:00 PM would allow children to function healthily on a daily basis. In practical terms, a 9:00 PM bedtime enables children to sleep until approximately 7:00-8:00 AM, providing sufficient time to prepare for and arrive at school. In addition, "screen time" was measured during 2022 - 2024, when academic activities had resumed their normal patterns following the COVID-19 pandemic, thereby minimizing the potential impact of external variables. Media time was averaged and reported in hours. Due to the varied and skewed nature of the data, standardized *z* scores were used for analysis.

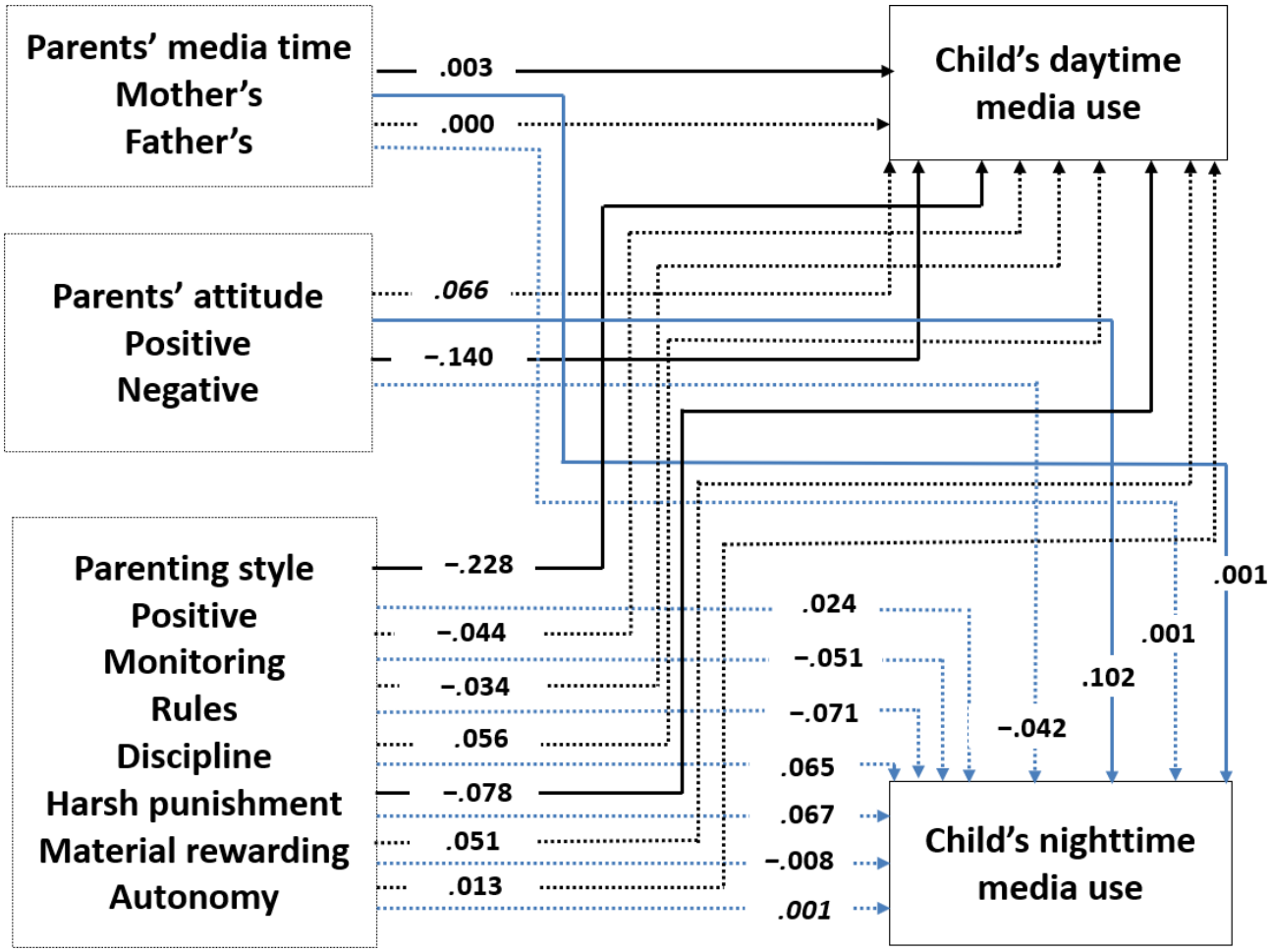
Results

Multigroup structural equation modeling was conducted to test the hypotheses and research question using Mplus 8.0 [60], which uses the maximum likelihood estimation method. To evaluate the model fit, confirmatory fit index (CFI), goodness-of-fit index (GFI), and normed fit index (NFI) were used.

Acceptable goodness-of-fit indices were obtained for the overall model [61]: $\chi^2_1=1.2$, $P=.27$, CFI=0.99, GFI=1.00, NFI=0.99 for 2022, $\chi^2_1=8.9$, $P<.01$, CFI=0.99, GFI=1.00, NFI=0.99 for 2023, and $\chi^2_1=44.7$, $P<.001$, CFI=0.98, GFI=0.99, NFI=0.98 for 2024. The estimated coefficients are presented in Figures 1-3.

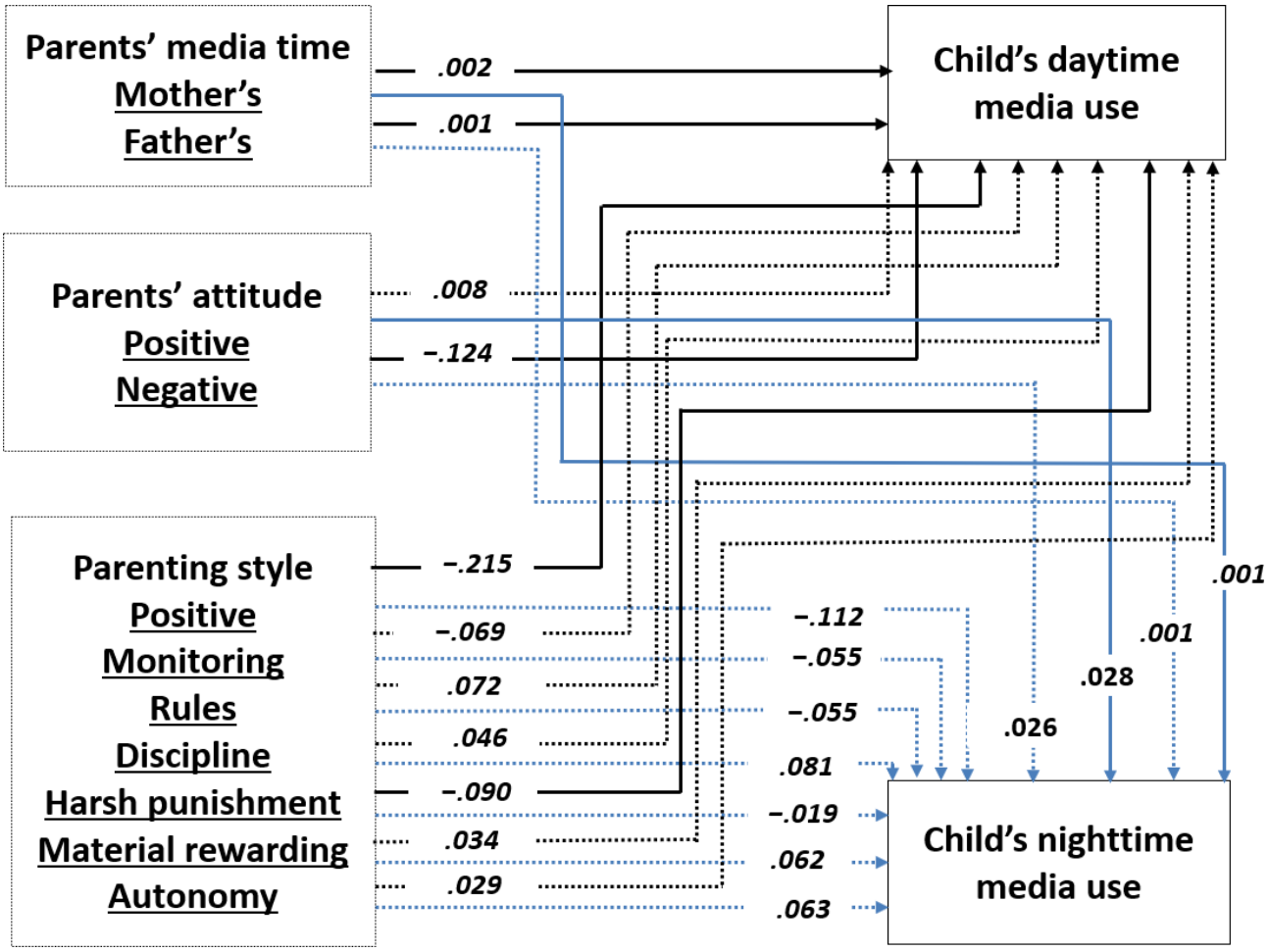
XSL•FO
RenderX

Figure 1. The final model of the relationships among a child's daytime and nighttime media use, parents' media time, parents' attitude toward media use, and parenting style in 2022. Note that the values are the observed standardized path coefficients. Solid lines indicate significant coefficients at $P<.05$, whereas dotted lines represent nonsignificant coefficients. In addition, paths leading to nighttime media use are shown in light blue, while paths leading to daytime media use are shown in black.



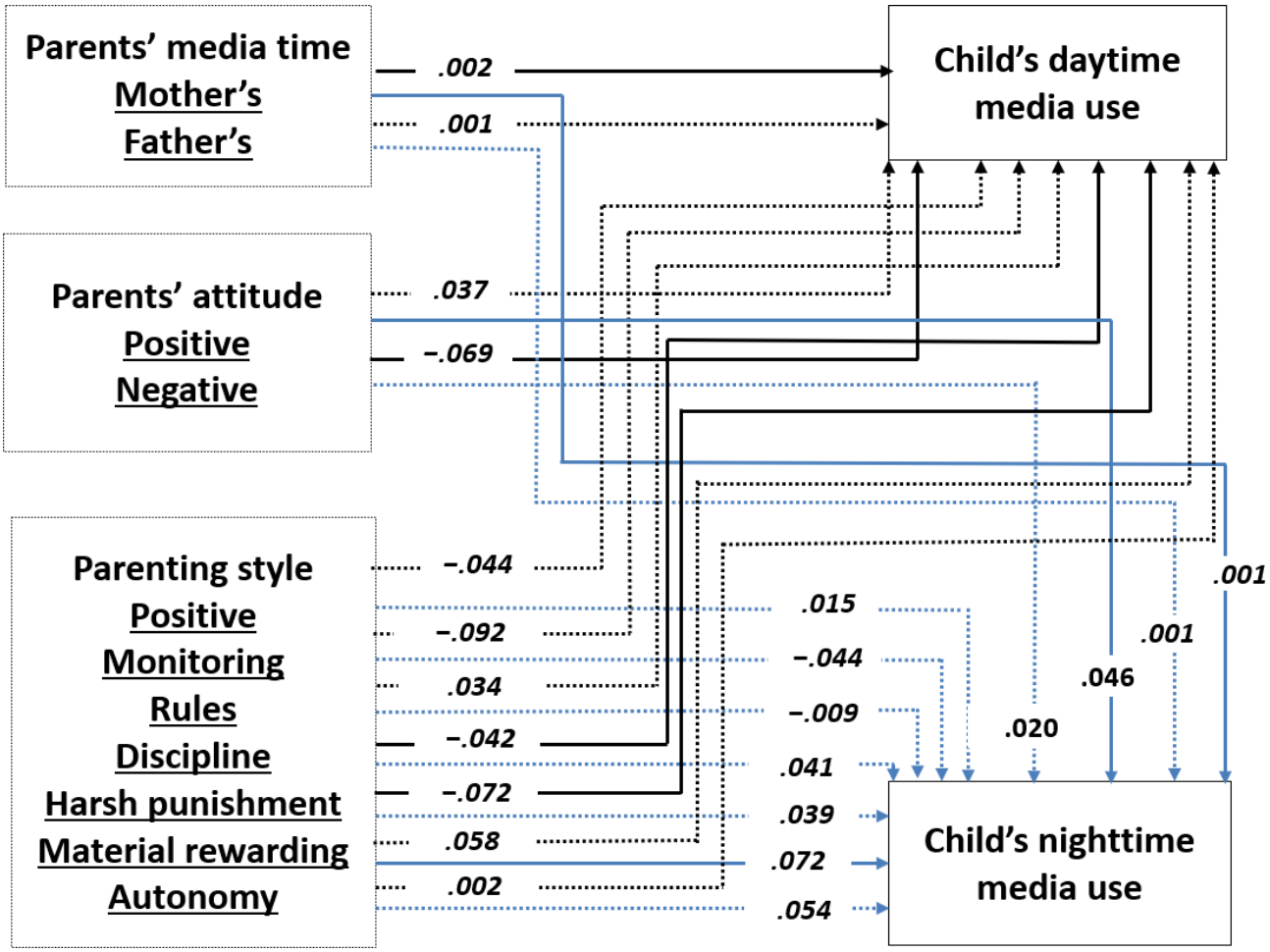
XSL•FO
RenderX

Figure 2. The final model of the relationships among a child's daytime and nighttime media use, parents' media time, parents' attitude towards media use, and parenting style in 2023. Note that the values are the observed standardized path coefficients. Solid lines indicate significant coefficients at $P<.05$, whereas dotted lines represent non-significant coefficients. In addition, paths leading to nighttime media use are shown in light blue, while paths leading to daytime media use are shown in black.



XSL•FO
RenderX

Figure 3. The final model of the relationships among a child's daytime and nighttime media use, parents' media time, parents' attitude toward media use, and parenting style in 2024. Note that the values are the observed standardized path coefficients. Solid lines indicate significant coefficients at $P<.05$, whereas dotted lines represent nonsignificant coefficients. In addition, paths leading to nighttime media use are shown in light blue, while paths leading to daytime media use are shown in black.



H1 predicted that parents with a higher level of screen time would have children who have a higher level of media use. The results partially supported this hypothesis. Regarding a child's daytime media use, mother's media time had a positive effect consistently in 2022, 2023, and 2024 (β=.002-.003), but father's media time showed a significant positive effect on child's daytime in 2023 only (β=.001). Regarding a child's nighttime media use, mother's media time showed a positive effect in 2022, 2023, and 2024 (β=.001).

H2 and H3 hypothesized that parents' positive and/or negative attitudes toward media use would affect child's media use. Parents' positive attitude toward media use increased a child's nighttime media use (β=.028-.102) and parents' negative attitude decreased a child's daytime media use (β=−.069-.140). H2 and H3 were supported.

Finally, H4 predicted that parents with permissive and neglectful parenting styles would be positively associated with children's media use, while parents with authoritative and authoritarian parenting styles would be negatively related to child's media use. Positive parenting style decreased a child's daytime media use (β=−.228 for 2022, β=−.215 for 2023), and harsh punishment (β=−.078 for 2022, β=−.090 for 2023, and β=−.072 for 2024) among the 7 parenting styles significantly decreased a child's daytime media use. Material reward was positively associated with child's nighttime media use (β=.072) in 2024.

RQ1 examined which factors were the most influential among parents' media use, attitude on media, and parenting styles. The results showed that parents' positive parenting style was the strongest predictor of a child's daytime media use in 2022 and 2023, but parents' harsh punishment was the strongest predictor in 2024. Parents' positive and negative attitudes on media use were the strongest predictor of a child's nighttime media use in 2022 only.

Discussion

This study investigated the connection between young children's media use and various parental influences, including media time, attitudes toward media, and parenting styles. The results across 2022, 2023, and 2024 consistently demonstrate the critical role parents play in shaping their children's media use habits. Similar to previous studies, it was observed that when parents—especially mothers—engage in higher media use, their children's media consumption during both daytime and nighttime increases, supporting the findings from past research [26,38,39]. This aligns with the idea that parental behavior directly influences children's media use habits.

XSL•FO
RenderX

Prior research has typically collapsed media consumption into a single metric, overlooking potential temporal differences. Our findings show that parental influences such as parent attitudes and parenting styles are differentially affected by the temporal contexts (ie, daytime vs nighttime). Specifically looking at parent attitudes, across the 3 years, a negative parent attitude toward media consistently correlated with decreased daytime media use, but a positive parent attitude toward media consistently correlated with increased nighttime media use. These consistent relationships across 3 years reflect the findings by Cingel and Krcmar [30] that emphasize the predictive power of parental attitudes on children's screen time.

Parenting style showed distinct influences on children's media use patterns over the 3-year period. In 2022 and 2023, positive parenting emerged as a significant factor in reducing children's daytime media use, highlighting that nurturing and supportive approaches were effective in moderating screen time during the day. By 2024, however, harsh punishment had become the strongest predictor of reduced daytime media use, suggesting a shift where more stringent disciplinary actions had a notable impact on limiting screen time. This change may reflect an evolving role of disciplinary measures in influencing children's media habits over time.

Contrary to expectations, none of the parenting styles showed a significant association with children's nighttime media use across the 3 years except 1. In 2024, material reward was positively associated with a child's nighttime media use. The findings indicated that parenting styles, while impactful on daytime media consumption, may have a limited role in controlling screen time during nighttime hours. Overall, the findings of daytime and nighttime media use suggest that while positive parenting consistently influences daytime media use, the emergence of harsh punishment and material reward as a stronger factor in 2024 points to potential shifts in the effectiveness of various parenting strategies over time.

The findings from a study focused on a younger age group (aged 3-6 y) than this study [26] demonstrated that children's daytime media use increased when parents provided more autonomy to their children. Conversely, nighttime media use decreased among children of parents who emphasized discipline. Children's nighttime media use was significantly increased when parents employed material rewards as a motivational tool. This study extended the findings of Lee et al [26] by investigating an older group of children (ie, aged 5-7 y). The findings from the 2 studies collectively suggest that as children grow older, the effectiveness of these parental interventions, particularly in reducing nighttime media use, diminishes significantly. These findings contribute to a broader understanding of the role of parental determinants in shaping children's media use and underscore the necessity for age-specific parental guidelines.

Previous results consistently highlighted the value of parental mediation in amplifying the positive effects and mitigating the adverse effects of children's media exposure [45,62]. These findings have encouraged scholars to highlight the critical role of parental engagement in managing children's media use [41]. However, a significant gap persists in the implementation of parental controls. For instance, most American adolescents report having no rules or restrictions from their parents regarding either the type of television content or the amount of time spent viewing [1].

Parents play a crucial role in ensuring safe and beneficial use of digital devices, particularly given that children's cognitive and functional abilities are still in developmental stages. Studies have shown that parents can actively promote children's growth and development through well-informed media management practices. Rather than relying on restrictive approaches, Wu et al [63] proposed integrating restrictive, instructive, and joint approach strategies to create a balanced and effective framework. This combined approach encourages collaboration and communication between parents and children, facilitating the establishment of developmentally appropriate media use habits that support overall growth.

Furthermore, these results extend previous research on parental determinants of children's media use and underscore the importance of establishing specific parental guidelines tailored to different times of day. The differentiation between daytime and nighttime media use is particularly relevant, as studies have shown that media use affects multiple developmental domains, including sleep quality, cognitive development, language, and socioemotional and physical health [13-16]. The variations observed over the years suggest that a longitudinal approach can provide a more comprehensive understanding of parental influences on media habits.

From a policy perspective, these findings support the creation of targeted recommendations that help parents foster healthy media use patterns at home. With children increasingly exposed to media, guidelines that encourage informed parental choices could mitigate the potential risks associated with excessive screen time. This evidence-based approach can empower parents to optimize on- and off-screen activities for their children's development.

However, this study is not without limitations. First, it is important to acknowledge that the data were collected during COVID-19 and post-COVID years (2022 - 2024), when screen time became both more prevalent and more socially normalized. In this period, children's daytime media use may not solely reflect leisure activities but may also encompass school-related or educational screen exposure. Second, parental self-reports remain susceptible to social desirability bias, especially given the awareness of recommended screen time limits [64]. Future research should incorporate observational methods to provide a more objective measure of media use. Additionally, distinguishing between types of media would allow for a more nuanced understanding of the effects of specific content on children's behavior and development, as different media types (eg, gaming vs educational content) may have distinct impacts. Similarly, a limitation of this study is the reliance on parent-reported 24-hour matrix data, which, while useful for distinguishing daytime and nighttime use, does not capture the content, context, or quality of children's media exposure. Prior work highlights that screen time measures often overlook these dimensions and that more comprehensive approaches are needed to assess family media environments [65,66]. Future studies

XSL•FO
RenderX

should integrate both time-based metrics and contextual measures to provide a fuller understanding of children's media use. Although the survey company's online panel is designed to mirror the Korean population in terms of key demographic characteristics (eg, sex, age, and family structure), the sampling method is not entirely random. Therefore, the findings should not be interpreted as being representative of the whole population. Rather, the results should be considered in the context of the study's sample, and caution should be exercised when generalizing these patterns to all Korean parents and children. Finally, the correlational nature of this study does not imply causation; thus, further research should explore potential genetic and environmental moderators that could influence the relationship between parental behavior and children's media habits.

In conclusion, the study across 3 years underscores those parental influences, including media time, attitudes, and specific parenting strategies, continue to shape children's media behaviors. By accounting for temporal patterns, researchers and policymakers can provide parents with tailored guidance to support balanced media use, enhancing children's overall well-being in the digital age.

Funding

This work was supported by the Ministry of Education of the Republic of the Korea and National Research Foundation of Korea (NRF-2020S1A3A2A02095619).

Conflicts of Interest

None declared.

References

1. Rideout VJ, Foehr UG, Roberts DF. Generation M2: media in the lives of 8- to 18-year-olds. : Kaiser Family Foundation; 2010 URL: https://www.kff.org/wp-content/uploads/2013/04/8010.pdf [accessed 2026-01-09]
2. Children media use survey [Report in Korean]. : Korean Press Foundation; 2020 URL: https://www.kpf.or.kr/front/research/consumerDetail.do?seq=591511 [accessed 2026-01-09]
3. American Academy of Pediatrics, Council on Communications and Media. Children, adolescents, and the media. Pediatrics 2013 Nov;132(5):958-961. [doi: 10.1542/peds.2013-2656] [Medline: 28448255]
4. Lammers SM, Woods RJ, Brotherson SE, Deal JE, Platt CA. Explaining adherence to American Academy of Pediatrics screen time recommendations with caregiver awareness and parental motivation factors: mixed methods study. JMIR Pediatr Parent 2022 Apr 5;5(2):e29102. [doi: 10.2196/29102] [Medline: 35380541]
5. Poulain T, Vogel M, Neef M, et al. Reciprocal associations between electronic media use and behavioral difficulties in preschoolers. Int J Environ Res Public Health 2018 Apr 21;15(4):814. [doi: 10.3390/ijerph15040814] [Medline: 29690498]
6. LeBlanc AG, Spence JC, Carson V, et al. Systematic review of sedentary behaviour and health indicators in the early years (aged 0-4 years). Appl Physiol Nutr Metab 2012 Aug;37(4):753-772. [doi: 10.1139/h2012-063] [Medline: 22765839]
7. Pillay S, Vijaindren A. Keep the smartphones away from your kids. New Straits Times. 2016. URL: http://www.nst.com.my/news/2016/05/147114/keep-smartphonesaway-your-kids [accessed 2026-01-09]
8. Burns MK, Pulles SM, Maki KE, et al. Accuracy of student performance while reading leveled books rated at their instructional level by a reading inventory. J Sch Psychol 2015 Dec;53(6):437-445. [doi: 10.1016/j.jsp.2015.09.003] [Medline: 26563597]
9. Hagen EW, Starke SJ, Peppard PE. The association between sleep duration and leptin, ghrelin, and adiponectin among children and adolescents. Curr Sleep Medicine Rep 2015 Dec;1(4):185-194. [doi: 10.1007/s40675-015-0025-9]
10. Miller DP. Maternal work and child overweight and obesity: the importance of timing. J Fam Econ Iss 2011 Jun;32(2):204-218. [doi: 10.1007/s10834-010-9244-x]
11. Palmer S. Toxic Childhood: How the Modern World Is Damaging Our Children and What We Can Do about It: Orion; 2005. URL: https://openlibrary.org/books/OL40293153M/Toxic_Childhood [accessed 2026-01-09]
12. Huber B, Yeates M, Meyer D, Fleckhammer L, Kaufman J. The effects of screen media content on young children's executive functioning. J Exp Child Psychol 2018 Jun;170:72-85. [doi: 10.1016/j.jecp.2018.01.006] [Medline: 29448235]
13. Hinkley T, Verbestel V, Ahrens W, et al. Early childhood electronic media use as a predictor of poorer well-being: a prospective cohort study. JAMA Pediatr 2014 May;168(5):485-492. [doi: 10.1001/jamapediatrics.2014.94] [Medline: 24639016]
14. Reid Chassiakos YL, Radesky J, Christakis D, Moreno MA, Cross C, Council on Communications and Media. Children and adolescents and digital media. Pediatrics 2016 Nov;138(5):e20162593. [doi: 10.1542/peds.2016-2593] [Medline: 27940795]
15. Tomopoulos S, Dreyer BP, Berkule S, Fierman AH, Brockmeyer C, Mendelsohn AL. Infant media exposure and toddler development. Arch Pediatr Adolesc Med 2010 Dec;164(12):1105-1111. [doi: 10.1001/archpediatrics.2010.235] [Medline: 21135338]

XSL•FO
RenderX

16. Zimmerman FJ, Christakis DA, Meltzoff AN. Associations between media viewing and language development in children under age 2 years. J Pediatr 2007 Oct;151(4):364-368. [doi: 10.1016/j.jpeds.2007.04.071] [Medline: 17889070]
17. Brown A, Smolenaers E. Parents' interpretations of screen time recommendations for children younger than 2 years. J Fam Issues 2018 Jan;39(2):406-429. [doi: 10.1177/0192513X16646595]
18. Ginsburg KR, American Academy of Pediatrics Committee on Communications, American Academy of Pediatrics Committee on Psychosocial Aspects of Child and Family Health, American Academy of Pediatrics Committee on Psychosocial Aspects of Child and Family Health. The importance of play in promoting healthy child development and maintaining strong parent-child bonds. Pediatrics 2007 Jan;119(1):182-191. [doi: 10.1542/peds.2006-2697] [Medline: 17200287]
19. Zhao J, Zhang Y, Jiang F, et al. Excessive screen time and psychosocial well-being: the mediating role of body mass index, sleep duration, and parent-child interaction. J Pediatr 2018 Nov;202:157-162.e1. [doi: 10.1016/j.jpeds.2018.06.029] [Medline: 30100232]
20. Byun YH, Ha M, Kwon HJ, et al. Mobile phone use, blood lead levels, and attention deficit hyperactivity symptoms in children: a longitudinal study. PLoS One 2013;8(3):e59742. [doi: 10.1371/journal.pone.0059742] [Medline: 23555766]
21. Pagani LS, Fitzpatrick C, Barnett TA, Dubow E. Prospective associations between early childhood television exposure and academic, psychosocial, and physical well-being by middle childhood. Arch Pediatr Adolesc Med 2010 May;164(5):425-431. [doi: 10.1001/archpediatrics.2010.50] [Medline: 20439793]
22. Kim B, Shin H. The relationship between young children's media use time, self-regulation, and problem behaviors. Korean J Hum Ecol 2019 Dec 31;28(6):567-581. [doi: 10.5934/kjhe.2019.28.6.567]
23. Margaryan A, Littlejohn A, Vojt G. Are digital natives a myth or reality? University students' use of digital technologies. Comput Educ 2011 Feb;56(2):429-440. [doi: 10.1016/j.compedu.2010.09.004]
24. An SM, Kang MJ. The effect of mothers' inappropriate smart device usage habits on preschoolers' cognitive ability: sequential mediating effect of preschoolers' smart device immersion tendency and self-regulation. Fam Environ Res 2021 May;59(2):157-168. [doi: 10.6115/fer.2021.012]
25. Collier KM, Coyne SM, Rasmussen EE, et al. Does parental mediation of media influence child outcomes? A meta-analysis on media time, aggression, substance use, and sexual behavior. Dev Psychol 2016 May;52(5):798-812. [doi: 10.1037/dev0000108] [Medline: 26914217]
26. Lee HE, Kim JY, Kim C. The influence of parent media use, parent attitude on media, and parenting style on children's media use. Children (Basel) 2022 Jan 1;9(1):37. [doi: 10.3390/children9010037] [Medline: 35053662]
27. Jago R, Davison KK, Thompson JL, Page AS, Brockman R, Fox KR. Parental sedentary restriction, maternal parenting style, and television viewing among 10- to 11-year-olds. Pediatrics 2011 Sep;128(3):e572-e578. [doi: 10.1542/peds.2010-3664] [Medline: 21859910]
28. Park YM, Kang MJ. Pathway from maternal parenting efficacy, inappropriate motives for allowing smart devices, and smart device dependency to preschoolers' ability to understand minds. Korean J Child Stud 2020 Dec;41(6):9-24. [doi: 10.5723/kjcs.2020.41.6.9]
29. Xu H, Wen LM, Rissel C. Associations of maternal influences with outdoor play and screen time of two-year-olds: findings from the Healthy Beginnings Trial. J Paediatr Child Health 2014 Sep;50(9):680-686. [doi: 10.1111/jpc.12604] [Medline: 24893682]
30. Cingel DP, Krcmar M. Predicting media use in very young children: the role of demographics and parent attitudes. Commun Stud 2013 Sep;64(4):374-394. [doi: 10.1080/10510974.2013.770408]
31. Lauricella AR, Wartella E, Rideout VJ. Young children's screen time: the complex role of parent and child factors. J Appl Dev Psychol 2015 Jan;36:11-17. [doi: 10.1016/j.appdev.2014.12.001]
32. Coyne SM, Radesky J, Collier KM, et al. Parenting and digital media. Pediatrics 2017 Nov;140(Suppl 2):S112-S116. [doi: 10.1542/peds.2016-1758N] [Medline: 29093044]
33. The state of school education: one year into the COVID pandemic. : Organisation for Economic Co-operation and Development; 2021 URL: https://www.oecd.org/content/dam/oecd/en/publications/reports/2021/04/the-state-of-school-education_b929a614/201dde84-en.pdf [accessed 2026-01-09]
34. Bergmann C, Dimitrova N, Alaslani K, et al. Young children's screen time during the first COVID-19 lockdown in 12 countries. Sci Rep 2022 Feb 7;12(1):2015. [doi: 10.1038/s41598-022-05840-5] [Medline: 35132065]
35. Bandura A. Social Learning Theory: Prentice Hall; 1977. URL: https://catalog.nlm.nih.gov/discovery/fulldisplay/alma9915626113406676/01NLM_INST:01NLM_INST [accessed 2026-01-09]
36. Rideout V, Hamel E. The media family: electronic media in the lives of infants, toddlers, preschoolers and their parents. : The Kaiser Family Foundation; 2006 URL: https://www.kff.org/wp-content/uploads/2013/01/7500.pdf [accessed 2026-01-09]
37. Kim HN, Lee SA, Lee EK, Chon JH, Kim SH. Effects of the smart media exposure environment on smart media addiction of young children [Article in Korean]. J Korean Soc Child Care 2014;14(4):127-153 [FREE Full text]
38. Oh J, Park Y. A study on preschoolers' smart media use and parents' perception [Article in Korean]. Korean J Child Care Educ Policy 2019;13(3):3-26 [FREE Full text]
39. Poulain T, Ludwig J, Hiemisch A, Hilbert A, Kiess W. Media use of mothers, media use of children, and parent-child interaction are related to behavioral difficulties and strengths of children. Int J Environ Res Public Health 2019 Nov 22;16(23):4651. [doi: 10.3390/ijerph16234651] [Medline: 31766650]

XSL•FO
RenderX

40. Pequegnat W, Szapocznik J. The role of families in preventing and adapting to HIV/AIDS: issues and answers. In: Working with Families in the Era of HIV/AIDS: Sage; 2000:3-26 URL: https://books.google.co.in/books?id=HvQJAZIL12UC&source=gbs_book_other_versions [accessed 2026-01-09]
41. Rasmussen EC, White SR, King AJ, Holiday S, Densley RL. Predicting parental mediation behaviors: the direct and indirect influence of parents' critical thinking about media and attitudes about parent-child interactions. J Media Lit Educ 2016;8(2):1-21. [doi: 10.23860/JMLE-2016-08-02-01]
42. Arnold LB. Family Communication: Theory and Research: Pearson Education Inc; 2007. URL: https://www.biblio.com/book/family-communication-theory-research-arnold-lorin/d/1449710603?srsltid=AfmBOoqtDQCOnxEkRrRVlGHDcUKRBpbZMqYpkg333S6H3dAxSOQFfnJY [accessed 2026-01-09]
43. American Academy of Pediatrics, Committee on Public Education. American Academy of Pediatrics: children, adolescents, and television. Pediatrics 2001 Feb;107(2):423-426. [doi: 10.1542/peds.107.2.423] [Medline: 11158483]
44. Vandewater EA, Rideout VJ, Wartella EA, Huang X, Lee JH, Shim MS. Digital childhood: electronic media and technology use among infants, toddlers, and preschoolers. Pediatrics 2007 May;119(5):e1006-e1015. [doi: 10.1542/peds.2006-1804] [Medline: 17473074]
45. Chakroff JL, Nathanson AI. Parent and school interventions: mediation and media literacy. In: Calvert SL, Wilson BJ, editors. The Handbook of Children, Media, and Development: Blackwell Publishing; 2008:552-576. [doi: 10.1002/9781444302752]
46. Council on Communications and Media, Brown A. Media use by children younger than 2 years. Pediatrics 2011 Nov;128(5):1040-1045. [doi: 10.1542/peds.2011-1753] [Medline: 22007002]
47. Sanders W, Parent J, Forehand R, Sullivan ADW, Jones DJ. Parental perceptions of technology and technology-focused parenting: associations with youth screen time. J Appl Dev Psychol 2016;44:28-38. [doi: 10.1016/j.appdev.2016.02.005] [Medline: 27795603]
48. Joe SS, Kim BH. A study of factors affecting the amount of children's YouTube use. J Korean Entertain Ind Assoc 2021 Oct 31;15(7):45-57. [doi: 10.21184/jkeia.2021.10.15.7.45]
49. Coyne SM, Rogers A, Holmgren HG, et al. Masters of Media: a longitudinal study of parental media efficacy, media monitoring, and child problematic media use across early childhood in the United States. J Child Media 2023;17(3):318-335. [doi: 10.1080/17482798.2023.2200958] [Medline: 37841526]
50. Huang S, Lai X, Li Y, et al. Does parental media mediation make a difference for adolescents? Evidence from an empirical cohort study of parent-adolescent dyads. Heliyon 2023 Apr;9(4):e14897. [doi: 10.1016/j.heliyon.2023.e14897] [Medline: 37095946]
51. Fam JY, Männikkö N, Juhari R, Kääriäinen M. Is parental mediation negatively associated with problematic media use among children and adolescents? A systematic review and meta-analysis. Can J Behav Sci 2023;55(2):89-99. [doi: 10.1037/cbs0000320]
52. Baumrind D. Current patterns of parental authority. Dev Psychol 1971;4(1, Pt.2):1-103. [doi: 10.1037/h0030372]
53. Maccoby EE, Martin JA. Socialization in the context of the family: parent-child interaction. In: Mussen PH, Hetherington EM, editors. Handbook of Child Psychology: Wiley; 1983, Vol. IV:1-101 URL: https://agris.fao.org/search/en/providers/122535/records/65dde6b74c5aef494fd6d65f [accessed 2026-01-09]
54. Van Leeuwen KG, Vermulst AA. Some psychometric properties of the Ghent Parental Behavior Scale. Eur J Psychol Assess 2004 Jan;20(4):283-298. [doi: 10.1027/1015-5759.20.4.283]
55. Padilla-Walker LM, Coyne SM. "Turn that thing off!" parent and adolescent predictors of proactive media monitoring. J Adolesc 2011 Aug;34(4):705-715. [doi: 10.1016/j.adolescence.2010.09.002] [Medline: 20943262]
56. Wood CT, Skinner AC, Brown JD, et al. Concordance of child and parent reports of children's screen media use. Acad Pediatr 2019 Jul;19(5):529-533. [doi: 10.1016/j.acap.2019.04.002] [Medline: 30981024]
57. Elias N, Sulkin I. YouTube viewers in diapers: an exploration of factors associated with amount of toddlers' online viewing. Cyberpsychol J Psychosoc Res Cyberspace 2017;11(3). [doi: 10.5817/CP2017-3-2]
58. Nikken P, Jansz J. Developing scales to measure parental mediation of young children's internet use. Learn Media Technol 2014 Apr 3;39(2):250-266. [doi: 10.1080/17439884.2013.782038]
59. Paruthi S, Brooks LJ, D'Ambrosio C, et al. Recommended amount of sleep for pediatric populations: a consensus statement of the American Academy of Sleep Medicine. J Clin Sleep Med 2016 Jun 15;12(6):785-786. [doi: 10.5664/jcsm.5866] [Medline: 27250809]
60. Arbuckle JL. 7.0 user's guide. : Amos Development Corporation/SPSS Inc; 2006 URL: http://www.hsc.edu.kw/student/app_manuals/SPSS/Amos%207.0%20User's%20Guide.pdf [accessed 2026-01-09]
61. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
62. Buijzen M, Valkenburg PM. Parental mediation of undesired advertising effects. J Broadcast Electron Media 2005 Jun;49(2):153-165. [doi: 10.1207/s15506878jobem4902_1]
63. Wu CST, Fowler C, Lam WYY, Wong HT, Wong CHM, Yuen Loke A. Parenting approaches and digital technology use of preschool age children in a Chinese community. Ital J Pediatr 2014 May 7;40:44. [doi: 10.1186/1824-7288-40-44] [Medline: 24887105]

XSL•FO
RenderX

64. Rosenman R, Tennekoon V, Hill LG. Measuring bias in self-reported data. Int J Behav Healthc Res 2011 Oct;2(4):320-332. [doi: 10.1504/IJBHR.2011.043414] [Medline: 25383095]
65. Byrne R, Terranova CO, Trost SG. Measurement of screen time among young children aged 0-6 years: a systematic review. Obes Rev 2021 Aug;22(8):e13260. [doi: 10.1111/obr.13260] [Medline: 33960616]
66. Barr R, Kirkorian H, Radesky J, et al. Beyond screen time: a synergistic approach to a more comprehensive assessment of family media exposure during early childhood. Front Psychol 2020;11:1283. [doi: 10.3389/fpsyg.2020.01283] [Medline: 32754078]

Abbreviations

CFI: confirmatory fit index
GFI: goodness-of-fit index
NFI: normed fit index

Edited by H Witteman; submitted 06.Apr.2025; peer-reviewed by G Taylor, R Juhari; revised version received 18.Dec.2025; accepted 22.Dec.2025; published 28.Jan.2026.

Please cite as:
Kim JY, Yang AJ, Lee HE
Parental Influence on Children's Media Use in South Korea: National Population-Based Study
JMIR Hum Factors 2026;13:e75292
URL: https://humanfactors.jmir.org/2026/1/e75292
doi:10.2196/75292

© Ji Young Kim, Ah Jung Yang, Hye Eun Lee. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 28.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

An Internet-Delivered Intervention to Reduce LGBTQ+ Prejudice Among Romanian Teachers: Randomized Controlled Trial

Nastasia Sălăgean[1,2*], PhD; Ioana Maria Latu[2,3*], PhD; Torill Marie Bogsnes Larsen[4], PhD; Andreea Bogdana Isbă oiu[2,5], PhD; Florin Alin Sava[2], PhD

[1]Department of Scientific Research in Economy, Law and Human-Environment Interaction, Institute for Advanced Environmental Research, West University of Timi oara, Boulevard V Pârvan 4, Timisoara, Romania

[2]Department of Psychology, Faculty of Sociology and Psychology, West University of Timisoara, Timisoara, Romania

[3]School of Psychology, Queen's University Beflast, Beflast, United Kingdom

[4]Department of Health Promotion and Development, Faculty of Psychology, University of Bergen, Bergen, Norway

[5]Department of Psychology and Education Sciences, Transilvania University of Brasov, Brasov, Romania

[*]these authors contributed equally

Corresponding Author:
Nastasia Sălăgean, PhD
Department of Scientific Research in Economy, Law and Human-Environment Interaction, Institute for Advanced Environmental Research, West University of Timi oara, Boulevard V Pârvan 4, Timisoara, Romania

Abstract

Background: Discrimination can greatly impact both physical and mental health due to frequent stressors. Younger individuals, particularly those under the age of 17 years, are more adversely affected by victimization. Within the European Union, Romania exhibits poor rankings concerning LGBTQ+ (lesbian, gay, bisexual, transgender/transsexual, queer, and other minority sexual orientations and gender identities) inclusion, with large numbers of LGBTQ+ teenagers experiencing bullying due to their sexual orientation. Given that much of this discrimination and harassment occurs within schools, teachers and counselors are vital in affecting institutional change.

Objective: This study aims to investigate the impact of an intervention on reducing prejudice against the LGBTQ+ community among Romanian teachers and counselors. Most prior interventions of this nature target Western, educated, industrialized, rich, and democratic populations.

Methods: In this randomized controlled trial, we recruited 175 Romanian teachers via a national closed online user group and assigned them to either the experimental or control condition. Participants in the experimental condition received the intervention first and then completed the web-based outcome measures, while those in the control condition completed the measures first and then received the intervention. The intervention, designed for internet-based delivery, consisted of a 1-hour video session led by a pair of researchers. It blended educational information with testimonials of LGBTQ+ people, perspective-taking tasks, and a self-efficacy exercise. We measured LGBTQ+ prejudice (using Attitudes Toward Lesbians and Gay Men Scale, Homophobia Scale, and Attitudes Toward Homosexuals Scale), behavioral intentions, self-efficacy, perspective taking, intergroup disgust sensitivity, intergroup anxiety, empathy, factual knowledge about LGBTQ+ issues, as well as participants' feelings toward lesbian, gay, and bisexual individuals.

Results: Participants in the experimental group (n=89) showed significant reductions in prejudice when using the Attitudes Toward Lesbians and Gay Men Scale ($F_{1,173}$=7.22; P=.008) when compared to the control group (n=86), but not when using the other 2 attitudinal scales. We also found that the experimental group had warmer feelings ($F_{1,173}$=4.40; P=.04; d=0.32), were more likely to engage in supportive behaviors ($F_{1,173}$=13.96; P<.001; d=0.56), displayed more self-efficacy ($F_{1,173}$=9.14; P=.001; d=0.33), had more factual knowledge ($F_{1,173}$=11.98; P=.001; d=0.52), and had a higher ability to take the LGBTQ+ perspective after controlling for contact ($F_{1,172}$=4.77; P=.03; d=0.28). We did not observe significant differences in terms of intergroup disgust sensitivity ($F_{1,173}$=0.816; P=.37), intergroup anxiety for either positive ($F_{1,173}$=.383; P=.54) or negative emotions ($F_{1,173}$=0.51; P=.48), or empathy ($F_{1,173}$=0.02; P=.89).

Conclusions: The intervention offers initial evidence for the effectiveness of a cost-effective and portable online resource for educators and high school counselors, particularly in regions where negative attitudes toward the LGBTQ+ community are prominent. The results show that integrating blended cognitive (information), affective (indirect contact and perspective taking), and behavioral (self-efficacy and empowerment) approaches is a promising avenue for intervention in producing positive outcomes related to LGBTQ+ issues within the school environment.

XSL•FO
RenderX

Trial Registration: ISRCTN Registry ISRCTN 84290049; https://www.isrctn.com/ISRCTN84290049
International Registered Report Identifier (IRRID): RR2-10.2196/54254

(JMIR Hum Factors 2026;13:e63787) doi:10.2196/63787

KEYWORDS

LGBT; discrimination; minority stress; internet; intervention; randomized controlled trial; Romanian teachers; Romania; school teacher; web-based; internet-based delivery; anxiety; empathy; self-efficacy; educators; online session; school; LGBTQ+

Introduction

Background

The LGBTQ+ (lesbian, gay, bisexual, transgender/transsexual, queer, and other minority sexual orientations and gender identities) community has historically faced great adversity, as centuries of prosecution have only recently begun to be overturned. The collapse of communism in Eastern Europe brought several changes in civil societies after they began shifting their focus to prioritizing human rights values and norms [1]. In Romania, homosexuality was officially decriminalized in 2001 because of external and international pressures [1]. Despite these changes, the consequences of the decades of oppression are seen even today, as attitudes toward LGBTQ+ individuals remain overall negative [2]. In fact, when it comes to LGBTQ+ inclusion, Romania tends to perform poorly compared to other EU countries. According to a recent survey [3,4], 45% of LGBTQ+ individuals in Romania reported experiencing discrimination in at least one aspect of their lives, but only 8% reported their experiences to an organization dealing with such issues. Additionally, 43% of the respondents said they faced harassment due to their sexual orientation or gender identity, the second-highest rate among all the EU countries.

The rates mentioned above are particularly concerning when viewed from the perspective of the minority stress model [5]. This model proposes that minorities face unique stressors because of stigmatization and discrimination, and that these can lead to detrimental effects on health. The stress caused by repeated experiences of discrimination can accumulate over time and negatively impact both physical and mental health in the long run.

Several studies have been conducted to test this model. In terms of physical health, a systematic review of 26 studies [6] found that minority stress impacted several physical health outcomes in sexual minorities, including cancer incidence, changes in cardiovascular function, and immune response. Additionally, a different systematic review [7] found that sexual minorities face significantly elevated risks for cardiovascular diseases. Similarly, mental health issues in LGBTQ+ individuals have been widely studied. A systematic review [8] revealed that sexual minorities had a higher risk for mental health issues, substance abuse, and suicide risk. In line with the minority stress model, evidence suggests that individuals who identify as LGBTQ+ also face several obstacles that negatively impact their mental health, such as discrimination, emotional distress, victimization, and barriers to accessing mental health services [9].

Moreover, although data suggest that minority stress affects all LGBTQ+ individuals, it seems that age impacts the stress felt by sexual minorities. Several studies have shown that younger individuals generally report more stress than older ones [10,11]. A meta-analysis revealed that victimization had more negative effects on LGBTQ+ individuals under the age of 17 years as compared to those aged more than 17 years [12]. These findings are corroborated by the impact on suicide risk. A meta-analysis of 35 studies [13] showed that suicide risk was higher for young LGBTQ+ individuals aged 12-20 years compared to cisgender straight young people. Furthermore, sexual minority youth are more likely to experience mental health issues such as anxiety, depression, and impaired academic performance [14]. This effect is especially prevalent within school environments, as supported by a systematic review showing significant negative effects of bias-based bullying for minorities, including LGBTQ+ students [15].

In fact, 44% of LGBTQ+ pupils aged 15 to 17 years in Romania have admitted to hiding their sexual orientation or gender identity in school, the third highest rate in the European Union after Croatia (51%) and Cyprus (47%), and significantly higher than in countries like the Netherlands (16%), France (20%), or Denmark (20%). Disturbingly, half of the respondents aged 15 to 17 years (50%) said that they were bullied at school because of their sexual orientation [3,4]. This finding is further supported by the United Nations Educational, Scientific and Cultural Organization report on school violence [16]. As such, given that discrimination and harassment against LGBTQ+ individuals are quite prevalent in Romanian schools and among students [3,4], it is therefore imperative to create programs that help reduce the stigma and prejudice that Romanian LGBTQ+ students face.

We decided to focus such efforts on teachers as they are perhaps the most significant group that can affect institutional change. Teachers can influence their students' experiences by creating inclusive classroom norms, making these interventions most effective. In fact, survey evidence suggests that supportive communities and schools buffer the negative effects of bias-based bullying for LGBTQ+ students in schools [17] . This finding is backed up by qualitative data, with interviews in schools showing the importance of teacher attitudes and education in creating safe and inclusive environments for LGBTQ+ students [18]. Overall, a recent rapid realist review of interventions to promote LGBTQ+ inclusivity in schools [19] showed that interventions work best when the staff in the school are trained, including education on sexuality, gender issues, as well as how to be an effective ally. Teacher training was related to LGBTQ+ students experiencing less victimization, increased safety, greater self-esteem, improved mental health, and

XSL•FO
RenderX

improved academic performance. Moreover, teacher training improved the effectiveness of teacher involvement in gay-straight alliances, which directly impacts LGBTQ+ youth's outcomes in schools [20].

In this paper, we aim to present the findings of a teacher-oriented intervention whose protocol has already been published [21]. The main objective of the intervention is to use a 1-hour online session that focuses on education and contact as primary training elements. The session is meant to help high school teachers cope with cases of witnessing bullying acts against LGBTQ+ students and to improve their knowledge and attitudes toward LGBTQ+ students.

To create a safe and inclusive learning environment for LGBTQ+ students, it is crucial to not only address teacher biases toward them, but also to equip teachers with the skills and knowledge needed to intervene in LGBTQ+-related victimization. This includes providing teachers with training on how to recognize and respond to instances of LGBTQ+ bullying or discrimination, as well as model appropriate behaviors and attitudes toward LGBTQ+ students. By taking a proactive and comprehensive approach to support LGBTQ+ students, we can help ensure that they feel valued, respected, and included in the classroom, thus reducing the risk of significant mental and physical health costs.

According to a meta-analysis [22], interventions can be effective in reducing sexual prejudice. The effect size of these interventions can range from one-third to one-half of an SD. Furthermore, the analysis highlights the most effective strategies for reducing different outcomes. These strategies include educational interventions, contact with LGBTQ+ individuals, and a combination of education and contact. It is worth noting that the majority of interventions aimed at reducing homophobia have been conducted among undergraduate students in Western countries, with none conducted in Eastern Europe or other regions that still harbor clear animosity and prejudice toward the LGBTQ+ community. While some recent research has tested interventions in other countries such as Jamaica [23] or Brazil [24], in Romania (or Eastern Europe), no such interventions have been tested, particularly on teachers.

However, equally important is how the effect of these interventions is measured. The term "homophobia" has garnered a lot of attention since its inception in the 1970s [25]. Several researchers [26,27] have argued the term should not be used because of its inaccuracy, as the inclusion of the word "phobia" would suggest an anxiety-related measure rather than an attitudinal one. Additionally, the misnomer would mainly focus on attitudes toward gay men and leave out lesbians and bisexuals. This conceptual confusion has seeped into the measurements used when measuring the impact of the interventions. Currently, several measures exist that claim to test sexual prejudice, but few interventions have used more than 1 instrument. It would be, therefore, beneficial to include several measures to see which, if any, would better capture the effects of the intervention. Besides prejudice, behavioral, cognitive, and emotional outcomes are also vital in predicting and ensuring significant changes in the classroom.

In order to ensure the success of interventions, it is crucial to consider the cultural and institutional context in which they are implemented. A recent qualitative analysis [28] has shown that participants frequently criticize interventions for their lack of alignment with the context in which they are conducted. This can be used as a justification for resisting change. To achieve better outcomes, interventions should be tailored to the community's unique characteristics, including their beliefs, values, traditions, and social norms.

Aims and Hypotheses

The main objective of this research is to test an intervention plan that aims to improve Romanian teachers' LGBTQ+ outcomes. Our intention is to consider particular cultural and institutional characteristics, rather than simply implementing previous intervention strategies. Our approach involves using education and contact as primary training elements, consistent with meta-analytic findings [22] on training effectiveness. However, we design the educational components according to the specific needs of our target group. For example, we incorporate information on the biological (rather than social) causes of homosexuality, as this is a common misbelief among Romanians. The protocol for this study is peer-reviewed and published in Research Protocols [21].

Additionally, we addressed potential feelings or perceptions of threat (such as the belief that exposure to LGBTQ+ individuals will "cause" children to become gay) as, according to intergroup threat theory [29], these are important predictors of anti-gay bias. We included several elements that had been found useful in other interventions, such as perspective-taking [30] and self-efficacy [31,32]. We also ensured that these elements were culturally appropriate. The trial protocol details were reported elsewhere [21], including a complete list of intervention components, subcomponents, and contents. Further details are given in Table 1 in a previous research paper [21].

Through a randomized controlled trial, we aim to measure the impact of the intervention on the attitudes, behavior, cognition, and emotions of teachers toward the LGBTQ+ individuals compared to those randomly assigned to the control condition. Additionally, we also measure the participants' factual knowledge about LGBTQ+ issues in the classroom, as well as their attitudes toward LGBTQ+ individuals.

Methods

Ethical Considerations

The ethical review committee of the West University of Timisoara, Romania, approved this study (notice 74505/10.11.2022) after reviewing the procedure, measures, and materials used. Each participant was informed of the data collected and the details of the intervention on 3 separate occasions: upon signing up for the intervention session, upon beginning the intervention, and when responding to the questionnaires. Participants provided informed consent by clicking a radio button on an online sign-up page containing written information about the study.

Upon finishing the online multimedia intervention, respondents received an automatic unique code that they had to submit in a

XSL•FO
RenderX

separate form to confirm their full participation in the intervention. Once participants confirmed their full participation, they were sent a certificate of participation that could be used to earn continuing education credits. They were also given the opportunity to enter a raffle to win a gift certificate worth 500 RON (equivalent to US $110 or €100).

All research data (outcome measures) were collected anonymously without recording any personal or identifiable information. However, to ensure that participants were indeed teachers, and to generate the participation certificates, the researchers collected personal information (email and name) that was stored in a separate database and never associated with the actual study results.

Participant Recruitment and Eligibility Criteria

By working with the local center for educational resources and assistance to distribute an online message to all their national members, we aimed to recruit teachers or counselors employed in Romanian schools. The interested teachers and school counselors had the opportunity to sign up for 1 of the 17 advertised sessions across 3 months (from December 2022 to February 2023), scheduled at different times during the day, depending on their availability. To ensure more participation, the teachers who signed up for a session were sent a reminder email 24 hours before the commencement of the session.

All participants had to be either teachers or school counselors and be fluent in Romanian, as the intervention was conducted in this language. Additionally, because the trial was advertised and conducted online, respondents had to have basic computer or internet literacy to access the recruitment form and the Zoom meeting links (Zoom Communications, Inc).

Study Design

The experimental design used in this intervention was a randomized controlled trial. We randomly assigned participants to either the experimental or control condition using a 2-group design. Those in the experimental condition received the intervention first and then completed the outcome measures, while those in the control condition completed the outcome measures first and then received the intervention. All participants had to fill out the outcome measures through the use of online questionnaires. This design was chosen for ethical reasons so that all participants would benefit from the intervention and associated resources by the end of their participation. Given the rigorous random assignment to the training and control conditions, we did not expect any baseline differences to influence the outcome measures. We did, however, conduct analyses with and without controlling for habitual contact with LGBT individuals. We chose this approach because discussions in the contact literature suggested that baseline levels of contact can potentially lead to a selection bias of participants to these studies [33,34] and as such, controlling those levels can help us isolate the intervention's specific effect beyond habitual exposure.

Participants were grouped into sessions of up to 30 individuals, depending on their availability. Cluster randomization was done at the session level by the lead researcher, with each session being randomly assigned to either the experimental or control condition. The randomization was done with the use of an online number generator. Individual randomization within the session was not possible given that the intervention was presented to all participants at the same time.

Sessions were scheduled outside of typical working hours on different days and times to ensure there were no systematic biases due to participants' session choice.

This study did not involve any risk to the participants' physical or mental integrity, and no unintended harms were observed during implementation.

In addition to being published in an academic journal [21], the study protocol was also registered in an international clinical study registry (ISRCTN84290049) [35]. The study is reported in accordance with the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines as per the completed checklist (Checklist 1).

Intervention

To ensure a standardized experience for all participants, regardless of the session they attended, each session was led by a pair of researchers who were tasked with delivering scripted instructions, answering participant queries, and ensuring that the study was completed at a consistent pace with minimal distractions.

The stand-alone intervention was designed for internet-based delivery and was multimedia in nature. It contained a recorded animated presentation that contained information on terms, threat reduction, and effects of stigma, behavioral tools, testimonials of LGBTQ+ people, a perspective-taking task, and a self-efficacy exercise. In total, the intervention lasted 50 minutes. The full details on all the components of the intervention are presented in Table 1 in a previous research paper describing the protocol in detail [21].

Sample Size Estimation

In order to determine the appropriate sample size for our study, we undertook a power analysis. Drawing on an average effect size of d=0.66 from a meta-analysis by Bartos et al [22], we aimed to achieve a statistical power of 0.80 as per the Cohen [36] recommendation that type II error should be limited to a probability of 0.20. Our analysis indicated a required sample size of 122 participants, although we elected to over-recruit in order to account for multiple outcomes. There was no pre-established stopping rule, and recruitment continued until all interested participants had been given the opportunity to participate in 1 of the 17 scheduled sessions.

All participants who completed their participation in the study were included in the data analysis, with no exclusions.

Outcomes

Overview

In order to ensure the reliability of the final scores for each outcome, we verified that the Cronbach α coefficient passed the 0.7 threshold. Additionally, to ensure the accuracy of the

XSL•FO
RenderX

results in our research, we used 3 different scales in measuring sexual prejudice and antigay bias.

Attitudes Toward Lesbians and Gay Men Scale: Measure of Sexual Orientation Prejudice

This 10-item scale measured beliefs and attitudes toward gay men and lesbians ("I think male homosexuals are disgusting" and "Sex between two women is just plain wrong.") [37]. Items were rated on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). After reverse coding 4 items, scores were averaged into a final score with higher values denoting more negative attitudes. Reliability for this measure was very good (Cronbach α=0.88).

Homophobia Scale (HS): Attitudes Toward Gay Individuals

This scale was composed of 25 items that assessed attitudes toward gay individuals, as well as social avoidance and aggression toward them [38]. Examples of items from the scale included, "Gay people make me nervous" and "I make derogatory remarks about gay people." Participants rated each item on a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). After reverse coding 9 items, the scores were averaged to obtain a final score. Higher scores indicate more negative attitudes toward gay people. The scale has shown excellent reliability with a Cronbach α of 0.93.

Attitudes Toward Homosexuals Scale: Assessment of Negative Beliefs

We used an additional scale to measure the attitudes toward and avoidance of gay people [39]. The 12 items included statements such as "Homosexuality is disgusting in the eyes of God," and "If I can, I prefer to not be in the company of homosexuals." Participants rated the statements on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). Out of all, 5 items were reverse scored, and the statements were averaged into a final score with higher values denoting more negative attitudes. Reliability for this measure was excellent (Cronbach α=0.91).

Self-Efficacy

The original 10-item scale [40] was adapted for working with LGBTQ+ students in a school setting. The scale measured self-efficacy in dealing with issues related to LGBTQ+ students. Sample items included "If I try hard, I can solve difficult issues related to LGBTQ+ students" and "I can deal with unexpected situations that arise with LGBTQ+ students." Ratings ranged from 1 ("Not at all true for me") to 4 ("Perfectly true for me"), and higher scores indicating more self-efficacy in dealing with LGBTQ+–related behaviors in the classroom. The scale showed an excellent reliability with a Cronbach α of 0.92.

Behavioral Intentions

We used a 16-item scale to evaluate the propensity of teachers to exhibit supportive professional conduct in the classroom pertaining to LGBTQ+ issues [39,40], including discussing queries regarding sexual orientation with students or having books about gay and lesbian issues in the classroom. Each item was rated on a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). A final score was obtained by averaging the responses, with higher scores indicating a greater willingness to engage in LGBTQ+ supportive behaviors in the classroom. The measure showed excellent reliability (Cronbach α=0.95).

Factual Knowledge About LGBTQ+ Issues

We designed a questionnaire consisting of 7 questions to evaluate the level of understanding of the participants on LGBTQ+ topics. The questions were carefully crafted based on the training content and covered a range of topics such as the biological basis of gender and the elevated risk of suicide among LGBTQ+ youth. The items were rated as either true or false, and the final score was calculated by adding up the number of correct responses. Participants with higher scores demonstrated a greater knowledge of and familiarity with LGBTQ+ issues.

Feeling Thermometer

We assessed the teachers' attitudes toward gay, lesbian, and bisexual individuals using a feeling thermometer [41]. We asked them to rate their feelings toward each group using a sliding thermometer scale that ranged from 0 (indicating very negative feelings) to 100 (indicating very positive feelings). We then averaged the ratings for the 3 groups into an overall lesbian, gay, and bisexual (LGB) feelings thermometer. The scale showed excellent reliability (Cronbach α=0.98).

Intergroup Disgust Sensitivity

We used a 7-item scale to evaluate the repulsion toward LGBTQ+ groups [42]. The scale included statements such as "After shaking hands with someone who has a different sexual orientation, even if their hands were clean, I would want to wash my hands." Participants rated items on a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). One of the items was reverse-scored. The final score was calculated by averaging all the scores, with higher values indicating more disgust toward the LGBTQ+ community. Reliability for this measure was very good (Cronbach α=0.83).

Intergroup Anxiety

This scale consisted of 10 items that were designed to measure anxiety when interacting with people of another sexual orientation [43]. Participants were asked to rate the likelihood of feeling several emotions such as embarrassed, unsure, happy, or accepted, on a scale of 1 (not at all) to 10 (extremely). We calculated 2 scores, one for positive emotions and another for negative emotions. The reliability for both positive (Cronbach α=0.82) and negative emotions (Cronbach α=0.88) the reliability of the scale was very good.

Perspective Taking

This measure consisted of 5 items that assessed how well individuals can empathize with LGBTQ+ people [44]. Participants were asked to rate their ability to understand the issues that are important to LGBTQ+ people. The overall reliability of the scale was very good, with a Cronbach α score of 0.88.

Toronto Empathy Questionnaire

This scale consisted of 16 items that evaluated participants' level of empathy toward others [45]. Examples of statements

XSL•FO
RenderX

in the scale included "I feel upset when someone is treated disrespectfully." Participants were asked to rate the statements on a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Among all, 7 items were reverse coded, and the final score was computed by averaging all items. Higher scores indicated higher levels of empathy. Reliability for this measure was very good (Cronbach α=0.80).

Control Variables

Demographics

We asked respondents to indicate their age, gender, and sexual orientation. As all respondents were either teachers or counselors, they are required to have at least a bachelor's degree in their respective field. Therefore, respondents were not asked to provide their education level.

Contact With LGBT People

We inquired about the frequency of contact (eg, speaking) with individuals who identify as gay, lesbian, bisexual, and transgender on a scale of 1 (almost daily) to 6 (never).

The Duke University Religion Index was used to measure religiosity [46]. The scale included 5 items that assessed religious involvement, organizational and nonorganizational religious activity, as well as subjective religiosity or intrinsic religiosity. The responses to each item were averaged to obtain a single score, which indicated the level of religiosity. Higher scores suggested greater religiosity. Cronbach α for this scale was 0.81, suggesting a very good reliability.

We also asked participants to indicate their political ideology on a 100-point scale from very conservative to very liberal or progressive.

Statistical Analysis

Quantitative data from the surveys were analyzed using SPSS (version 24; IBM Corp) [47]. To examine our hypotheses, a series of one-way between-participants ANOVAs were performed. The objective of this analysis was to compare outcomes between the participants in the experimental and control groups. We also conducted the analyses while controlling for contact with LGBTQ+ individuals (ANCOVA), given that people's different experiences may influence their response to training, particularly when it comes to affective outcomes. The results were presented as F test and P values, along with descriptive statistics such as n, mean, and SD. In order to estimate the effect size using means and SDs, we calculated Cohen d.

There was no deviation in the analysis from the registered protocol [21]. However, we decided to include a post hoc analysis in Multimedia Appendix 1, to explain some inconsistent results in the case of one outcome—the HS. We also controlled the analyses using the Benjamini-Hochberg false discovery rate (FDR) method to correct for any type I errors that could occur when testing for multiple hypotheses [48]. These analyses were conducted using the R programming language (version 4.5.1; R Foundation for Statistical Computing) [49]. The results are presented with both the corrected (noted as P_{FDR}) and uncorrected P values.

Results

Descriptive Statistics

A total of 382 individuals signed up to attend one of the advertised sessions. Further examination revealed that 6 of the submissions were duplicated, rendering 376 unique individuals that signed up. Of these, 175 out of 382 (45.81%) participants actually participated and finished the intervention, and 201 individuals did not attend despite repeated reminders. A CONSORT (Consolidated Standards of Reporting Trials) flowchart of the phases of the intervention is provided in Figure 1. A full summary of the descriptive statistics, for the full sample (N=175), as well as for the experimental group (n=89) and control group (n=86), is presented in Table 1. Also included in the table are the statistical analyses made to check whether there were any significant differences between the 2 groups. As can be seen, no significant differences were found between the experimental group and the control group in terms of age, gender, sexual orientation, religiosity, ideology, or LGBT contact. This confirms successful randomization across the 2 experimental conditions.

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flowchart.

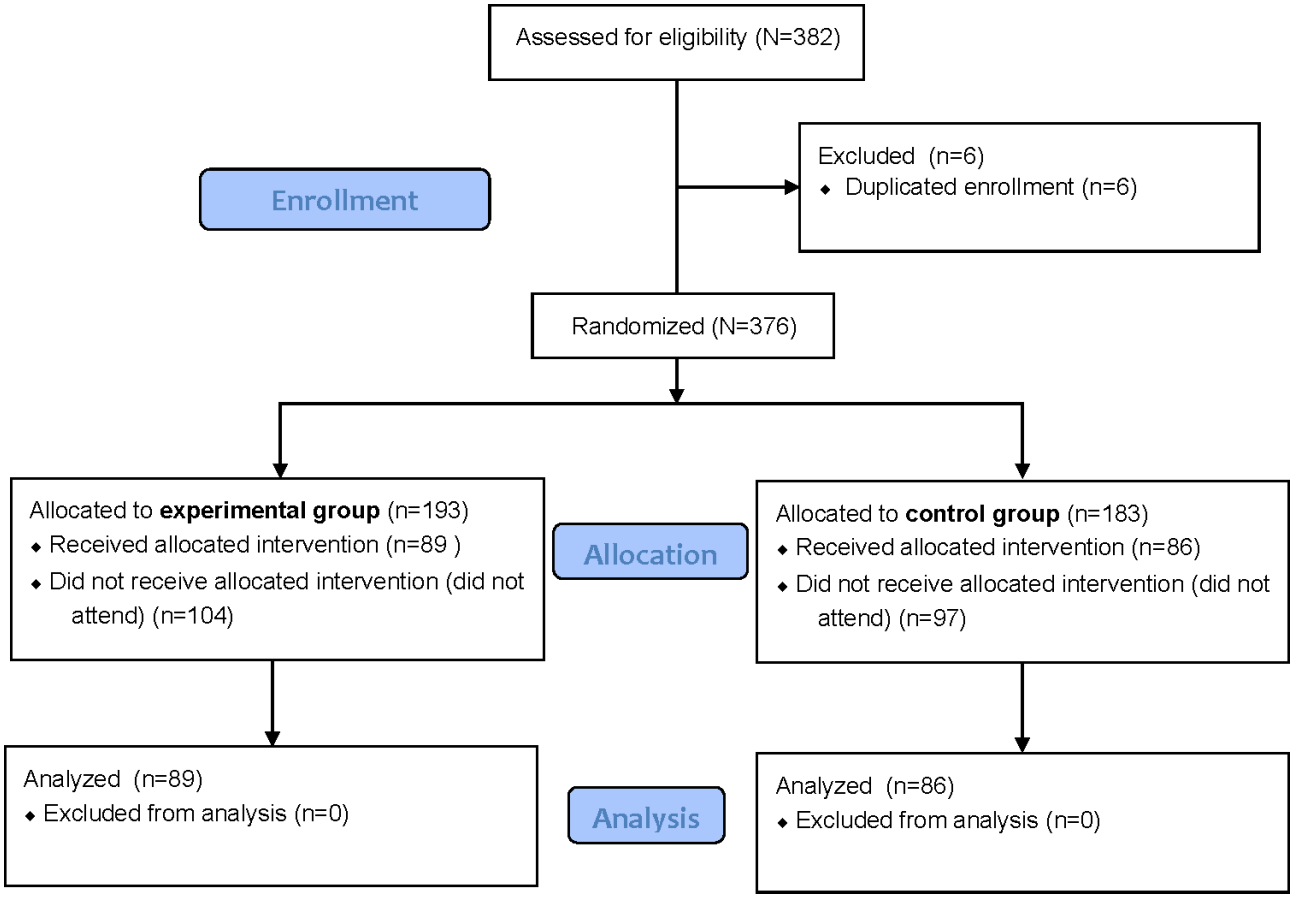


Table . Summary of descriptive statistics (N=175).

Variables	Full sample	Experimental group (n=89)	Control group (n=86)	Chi-square or *t* test (df)	*P* value
Age (y), mean (SD)	43.24 (7.77)	43.34 (7.97)	43.14 (7.62)	0.168 (173)[a]	.87
Gender, n (%)				1.65 (2)[b]	.44
Man	13 (7.4)	8 (9)	5 (5.8)		
Woman	161 (92)	81 (91)	80 (93)		
Nonbinary	1 (0.6)	0 (0)	1 (1.2)		
Sexual orientation, n (%)				4.06 (4)[b]	.40
Heterosexual	146 (83.4)	70 (78.7)	76 (88.4)		
Bisexual	5 (2.9)	3 (3.4)	2 (2.3)		
Other	6 (3.4)	4 (4.5)	2 (2.3)		
Do not know	2 (1.1)	2 (2.2)	0 (0)		
PNTS[c]	16 (9.1)	10 (11.2)	6 (7)		
Religiosity, mean (SD)	3.43 (0.98)	3.46 (0.94)	3.40 (1.03)	0.41 (173)[a]	.69
Ideology, mean (SD)	62.94 (24.86)	65.18 (23.9)	60.62 (25.75)	1.22 (173)[a]	.23
LGBT[d] contact, mean (SD)	5.36 (1.03)	5.39 (0.99)	5.33 (1.08)	0.360 (173)[a]	.72

[a]2-tailed *t* test (*df*).

[b]Chi-square (*df*).

[c]PNTS: prefer not to say.

[d]LGBT: lesbian, gay, bisexual, and transgender/transsexual.

XSL•FO
RenderX

Outcomes

Significant differences were observed for the Attitudes Toward Lesbians and Gay Men Scale ($F_{1,173}$=7.22; P=.008; P_{FDR}=.02), where teachers in the control group had more negative attitudes (mean 2.70, SD 0.93) than those in the experimental group (mean 2.36, SD 0.77). The effect size, as indicated by Cohen *d*, was small to medium (d=0.41). Teachers who completed the intervention reported less negative attitudes toward lesbians and gays compared to teachers who did not complete the intervention. When controlling for the effect of contact, the experimental condition remained significant ($F_{1,172}$=9.16; P=.003; P_{FDR}=.009; d=0.46).

Interestingly, for the HS, while we did see lower scores for the experimental group (mean 1.95, SD 0.64) than the control group (mean 2.08, SD 0.76), these differences were not statistically significant ($F_{1,173}$=1.58; P=.21; P_{FDR}=.32). This effect remained nonsignificant when controlling for contact ($F_{1,172}$=2.36; P=.13; P_{FDR}=.20). The same can be observed for the Attitudes Toward Homosexuals Scale, where no statistically significant differences ($F_{1,173}$=1.67; P=.20; P_{FDR}=.32) were observed between the experimental group (mean 2.14, SD 0.65) and the control group (mean 2.31, SD 0.88). This effect also remained nonsignificant when controlling for contact ($F_{1,172}$=2.40; P=.12; P_{FDR}=.20).

Initially, significant differences were seen for the LGB feelings thermometer, in that respondents in the experimental group had higher (therefore warmer) feelings toward LGB community members (mean 71.80, SD 27.35) when compared to respondents in the control group (mean 62.64, SD 30.35; $F_{1,173}$=4.40; P=.04; P_{FDR}=.09; d=0.32), though the effect size was small, but after applying the LGB FDR correction, the main result failed to reach significance. When controlling for the effect of contact, the experimental condition remained significant even after applying the FDR correction ($F_{1,172}$=5.82; P=.02; P_{FDR}=.048).

We did not find significant differences between the 2 groups in terms of intergroup disgust sensitivity ($F_{1,173}$=0.82; P=.37; P_{FDR}=.49), intergroup anxiety for either positive ($F_{1,173}$=0.38; P=.54; P_{FDR}=.59) or negative emotions ($F_{1,173}$=0.42; P=.52; P_{FDR}=.59), or empathy ($F_{1,173}$=0.02; P=.89; P_{FDR}=.89). Even when controlling for the effects of contact with LGBTQ+ individuals, we did not see any significant results for intergroup disgust sensitivity ($F_{1,172}$=1.11; P=.29; P_{FDR}=.39), intergroup anxiety for either positive ($F_{1,172}$=0.500; P=.48; P_{FDR}=.52) or negative emotions ($F_{1,172}$=0.51; P=.48; P_{FDR}=.52), or empathy ($F_{1,172}$=0.04; P=.84; P_{FDR}=.84).

However, we did see significant differences in terms of behavioral intentions, in that respondents in the experimental group were more willing to engage in helping behaviors (mean 4.19, SD 0.68) than those in the control group (mean 3.75, SD 0.89; $F_{1,173}$=13.96; P<.001; P_{FDR}=.006; d=0.56), even after controlling for contact ($F_{1,172}$=19.54; P<.001; P_{FDR}=.004).

We also saw that factual knowledge was significantly higher in respondents from the experimental group (mean 5.04, SD 1.24) than those in the control group (mean 4.37, SD 1.32; $F_{1,173}$=11.98; P=.001; P_{FDR}=.006; d=0.52), even when controlling for contact ($F_{1,172}$=12.03; P=.001; P_{FDR}=.004). Furthermore, respondents in the experimental group (mean 3.18, SD 0.42) had significantly higher levels of self-efficacy than those in the control group (mean 2.94; SD 0.93; $F_{1,173}$=9.14; P=.003; P_{FDR}=.01; d=0.33), even after looking at the effects of contact ($F_{1,172}$=11.82; P=.001; P_{FDR}=.004).

Finally, the difference between the experimental and control conditions in terms of perspective taking was just shy of reaching significance ($F_{1,173}$=3.27; P=.07; P_{FDR}=.14; d=0.28), such that respondents in the experimental group had a slightly higher ability to take the LGBTQ+ perspective (mean 3.87, SD 0.69) than those in the control group (mean 3.66, SD 0.80). This analysis only reached significance after controlling for contact ($F_{1,172}$=4.77; P=.03; P_{FDR}=.06), although we note that the effect was small, and after applying the FDR correction, the results were again insignificant.

We conducted the above analyses in line with the preregistered research protocol published in a study by Latu et al [21].

Discussion

Principal Results

Our intervention suggests that even a 1-hour online session that combines informative materials (education, factual knowledge, and behavioral tools to address bullying and support LGBTQ+ students) and testimonials (vicarious contact) produces small to average positive effects on various LGBTQ+ outcomes. The intervention decreased teachers' negative attitude toward LGBTQ+ topics, increased their self-efficacy in dealing with LGBTQ+ issues in the classroom, and led to a higher level of behavioral intentions to engage in supportive behaviors for LGBTQ+ students facing bullying in school contexts. It should be noted that the cognitive and behavioral outcomes were most stable, with the attitudinal outcome being relatively inconsistent across measures and types of analysis. Our inconsistent results across several scales of homophobia or antigay attitudes, together with our supplementary analyses, suggest that latent variables may be at play and that future research should systematically investigate the attitudinal components of homophobia or antigay bias for appropriate measurement.

The online intervention, however, did not lead to significant differences between the experimental and the control groups in terms of intergroup disgust or intergroup anxiety, nor on more general and nonspecific measures such as their empathy level toward other people. This may be because such emotional responses, compared to behavioral and cognitive components, are less malleable especially in the context of LGBTQ+ biases which are deeply ingrained in the Romanian mainstream culture. Our findings may also suggest that it was the educational and behavioral components of our intervention that were most effective. Improved affective outcomes, on the other hand, may result from the contact component of interventions [50]. Although vicarious contact was used in our intervention via recorded video testimonials, more direct and prolonged forms

XSL•FO
RenderX

of contact with LGBTQ+ individuals may have the power to change affective responses.

The small to average effect sizes obtained for various outcomes align with an existing meta-analysis [22], in which interventions meant to reduce sexual prejudices obtained similar effect sizes, namely from one-third to one-half of an SD. As already explained elsewhere [21], most of these interventions have been implemented on undergraduate students from Western countries, and none has been implemented in Eastern Europe.

Likewise, our results suggest the opportunity to implement a multipurpose intervention, as suggested in prior studies [24]. Despite its shortness, the intervention showed multiple benefits at the attitudinal, cognitive, affective, and behavioral intention levels.

Taking together, our results are encouraging and a promising starting point for addressing LGBTQ+ biases in teachers. The findings show that an online intervention that is relatively low-cost in terms of personnel and resources could be implemented at a larger scale. Likewise, the session could represent a valuable resource to be added to a more complex intervention. The resource we developed could be an excellent tool for teachers in their professional development training when it comes to topics such as accepting sexual and gender diversity or when required to tackle LGBTQ+–related bullying in school contexts.

Another strength of this work is that the intervention was implemented in a country (Romania) which maintains strong negative attitudes toward LGBTQ+ individuals in the public sphere, with roots in conservative and religious beliefs. Such an educational resource turns away the attention from the moral or religious to psychological aspects such as perspective-taking and diversity acceptance, including vicarious contact with LGBTQ+ people and the problems they face in school settings. The emphasis on the humanistic view in teachers, backed up by education, seems to work even in less tolerant environments concerning LGBTQ+ issues. Therefore, the program behind the current intervention can be seen as a useful and more portable resource for guidance counselors in high schools as well as for other specialists who are interested in upscaling the positive results obtained in various high schools. When evaluating our findings, we believe that the intervention is especially useful given that it led teachers to perceive themselves as more equipped to intervene when LGBTQ+ students experience bullying acts.

Limitations

A significant limitation of this study is the absence of long-term follow-up measurement of outcomes, with measurement being conducted in the same session as the intervention. Although these findings do not allow for inferences about long-term effects, we know that the outcomes of such interventions generally tend to be short-term. For example, a contact intervention [50] showed reduced LGBTQ+ negativity in the short term, but levels returned to baseline 7 days after. In contrast, those in the control condition who were not exposed to the intervention showed even more increased LGBTQ+ negativity 7 days later. Such findings suggest that our intervention, although short-term, can potentially lead to longer-term prevention of worsening of biases. We also suggest that, to secure long-term effects and change inclusivity norms in schools, our intervention could be part of a long-term curriculum of development for teachers.

Another limitation is that there may have occurred biases related to self-selection. Participants were included in the trial based on their voluntary consent and without any certain financial benefit (except for their inclusion in a raffle with a small chance of winning some financial benefits). Although their effort was minimal (attending an online 1-h session excluding the time required for the completion of the outcome questionnaires), it still could have led to the selection of a biased sample from the teachers' population, as most likely teachers with strong negative attitudes toward LGBTQ+ people were more reluctant to attend our study. However, due to random assignment to conditions and the presence of significant effects, we believe there is sufficient evidence about the efficacy of the training.

Another limitation of the study is related to the lack of blinding to the experimental conditions. Because of the nature of the intervention and how it was organized, we could not effectively blind respondents and experimenters to the nature of how the trial worked, thus increasing the risk of a type I error. Future intervention could use artificial intelligence or other automated methods to deliver the intervention in a standardized way, to avoid any demand characteristics or experimenter effects.

A final issue is around the gender composition of our sample, which predominantly consists of women. This gender composition is in line with worldwide trends [51] but also national trends in the Romanian education system [52]. However, theoretically, future studies could investigate the effects of gender on the efficiency of such training tools and adapt them accordingly.

Conclusions

This intervention is a promising resource that can serve as a valuable and relatively low-cost resource (no specialized human intervention is needed) for teachers and high school counselors, particularly in countries where negative attitudes toward LGBTQ+ people prevail. Our study also offers suggestions as to how such interventions could be improved and embedded in a longer-term program to increase classroom inclusiveness for LGBTQ+ students.

Acknowledgments

The authors thank Sebastian E Barto , Lavinia Chirică, and Andre Rădulescu for their advice and assistance in developing the intervention.

XSL•FO
RenderX

Funding

The research leading to these results has recieved funding from the NO grants 2014-2021, under project contract 17/2020.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional analyses.
[DOCX File, 54 KB - humanfactors_v13i1e63787_app1.docx]

Checklist 1
CONSORT-EHEALTH (V 1.6.1) checklist.
[PDF File, 1041 KB - humanfactors_v13i1e63787_app2.pdf]

References

1. Margarit D. LGBTQ rights, conservative backlash and the constitutional definition of marriage in Romania. Gender, Place & Culture 2019 Nov 2;26(11):1570-1587. [doi: 10.1080/0966369X.2019.1567472]
2. Wilson K. Attitudes toward LGBT people and their rights in Europe. In: Oxford Research Encyclopedia: Politics: Oxford University Press; 2020. [doi: 10.1093/acrefore/9780190228637.013.1335]
3. A long way to go for LGBTI equality. European Union Agency for Fundamental Rights. 2020. URL: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-lgbti-equality-1_en.pdf [accessed 2025-07-30]
4. A long way to go for LGBTI equality EU-LGBTI II: country data—Romania. : European Union Agency for Fundamental Rights; 2020 URL: https://fra.europa.eu/sites/default/files/fra_uploads/lgbti-survey-country-data_romania.pdf [accessed 2025-07-30]
5. Meyer IH. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. Psychol Bull 2003 Sep;129(5):674-697. [doi: 10.1037/0033-2909.129.5.674] [Medline: 12956539]
6. Flentje A, Heck NC, Brennan JM, Meyer IH. The relationship between minority stress and biological outcomes: a systematic review. J Behav Med 2020 Oct;43(5):673-694. [doi: 10.1007/s10865-019-00120-6] [Medline: 31863268]
7. Caceres BA, Brody A, Luscombe RE, et al. A systematic review of cardiovascular disease in sexual minorities. Am J Public Health 2017 Apr;107(4):e13-e21. [doi: 10.2105/AJPH.2016.303630] [Medline: 28207331]
8. Plöderl M, Tremblay P. Mental health of sexual minorities. a systematic review. Int Rev Psychiatry 2015;27(5):367-385. [doi: 10.3109/09540261.2015.1083949] [Medline: 26552495]
9. Moagi MM, van Der Wath AE, Jiyane PM, Rikhotso RS. Mental health challenges of lesbian, gay, bisexual and transgender people: an integrated literature review. Health SA 2021;26(1):1487. [doi: 10.4102/hsag.v26i0.1487] [Medline: 33604059]
10. Vale M, Bisconti T. Risk and resilience: highlighting minority stress in a lifespan sample of sexual minorities. Innov Aging 2022 Dec 20;6(Supplement_1):386-387. [doi: 10.1093/geroni/igac059.1523]
11. Vale MT, Bisconti TL. Age differences in sexual minority stress and the importance of friendship in later life. Clin Gerontol 2021;44(3):235-248. [doi: 10.1080/07317115.2020.1836107] [Medline: 33143546]
12. Fedewa AL, Ahn S. The effects of bullying and peer victimization on sexual-minority and heterosexual youths: a quantitative meta-analysis of the literature. J GLBT Fam Stud 2011 Jul;7(4):398-418. [doi: 10.1080/1550428X.2011.592968]
13. di Giacomo E, Krausz M, Colmegna F, Aspesi F, Clerici M. Estimating the risk of attempted suicide among sexual minority youths: a systematic review and meta-analysis. JAMA Pediatr 2018 Dec 1;172(12):1145-1152. [doi: 10.1001/jamapediatrics.2018.2731] [Medline: 30304350]
14. Oswalt SB, Wyatt TJ. Sexual orientation and differences in mental health, stress, and academic performance in a national sample of U.S. college students. J Homosex 2011;58(9):1255-1280. [doi: 10.1080/00918369.2011.605738] [Medline: 21957858]
15. Sapouna M, de Amicis L, Vezzali L. Bullying victimization due to racial, ethnic, citizenship and/or religious status: a systematic review. Adolescent Res Rev 2023 Sep;8(3):261-296. [doi: 10.1007/s40894-022-00197-2]
16. Behind the numbers: ending school violence and bullying. : United Nations Educational, Scientific and Cultural Organization; 2019 URL: https://www.unicef.org/media/66496/file/behind-the-numbers.pdf [accessed 2025-07-30]
17. Gower AL, Watson RJ, Erickson DJ, Saewyc EM, Eisenberg ME. LGBQ youth's experiences of general and bias-based bullying victimization: the buffering role of supportive school and community environments. Int Journal of Bullying Prevention 2021 Jun;3(2):91-101. [doi: 10.1007/s42380-020-00065-4] [Medline: 34926991]
18. Steck AK, Perry D. Challenging heteronormativity: creating a safe and inclusive environment for LGBTQ students. J Sch Violence 2018 Apr 3;17(2):227-243. [doi: 10.1080/15388220.2017.1308255]

XSL•FO
RenderX

19. Schlief M, Stefanidou T, Wright T, Levy G, Pitman A, Lewis G. A rapid realist review of universal interventions to promote inclusivity and acceptance of diverse sexual and gender identities in schools. Nat Hum Behav 2023 Apr;7(4):556-567. [doi: 10.1038/s41562-023-01521-z] [Medline: 36781989]
20. Day JK, Fish JN, Grossman AH, Russell ST. Gay-straight alliances, inclusive policy, and school climate: LGBTQ youths' experiences of social support and bullying. J Res Adolesc 2020 Feb;30 Suppl 2(Suppl 2):418-430. [doi: 10.1111/jora.12487] [Medline: 30861243]
21. Latu I, Sălăgean N, Larsen TMB, Isbasoiu AB, Sava FA. Testing the effectiveness of an intervention to improve Romanian teachers' LGBT+-related attitudes, cognitions, behaviors, and affect: protocol for a randomized controlled trial. JMIR Res Protoc 2024 Apr 23;13:e54254. [doi: 10.2196/54254] [Medline: 38652533]
22. Bartoş SE, Berger I, Hegarty P. Interventions to reduce sexual prejudice: a study-space analysis and meta-analytic review. J Sex Res 2014;51(4):363-382. [doi: 10.1080/00224499.2013.871625] [Medline: 24754358]
23. West K. Understanding and reducing sexual prejudice in Jamaica: theoretical and practical insights from a severely anti-gay society. J Sex Res 2018;55(4-5):472-485. [doi: 10.1080/00224499.2017.1416055] [Medline: 29411998]
24. Costa AB, Pase PF, de Camargo ES, et al. Effectiveness of a multidimensional web-based intervention program to change Brazilian health practitioners' attitudes toward the lesbian, gay, bisexual and transgender population. J Health Psychol 2016 Mar;21(3):356-368. [doi: 10.1177/1359105316628748] [Medline: 26987830]
25. Smith KT. Homophobia: a tentative personality profile. Psychol Rep 1971 Dec;29(3):1091-1094. [doi: 10.2466/pr0.1971.29.3f.1091] [Medline: 5139344]
26. Herek GM. Beyond "Homophobia": Thinking about sexual prejudice and stigma in the twenty-first century. Sex Res Soc Policy 2004 Apr;1(2):6-24. [doi: 10.1525/srsp.2004.1.2.6]
27. Schiffman JB, Delucia-Waack JL, Gerrity DA. An examination of the construct of homophobia. J LGBT Issues Couns 2006 May 10;1(1):75-93. [doi: 10.1300/J462v01n01_06]
28. Barto SE, Hegarty P. Negotiating theory when doing practice: a systematic review of qualitative research on interventions to reduce homophobia. J Homosex 2019;66(9):1262-1286. [doi: 10.1080/00918369.2018.1500780] [Medline: 30095380]
29. Stephan WG, Renfro CL, Davis M. The role of threat in intergroup relations. In: Wagner U, Tropp LR, Finchilescu G, Tredoux C, editors. Improving Intergroup Relations: Building on the Legacy of Thomas F Pettigrew: Blackwell Publishing; 2008:55-72. [doi: 10.1002/9781444303117]
30. Tompkins TL, Shields CN, Hillman KM, White K. Reducing stigma toward the transgender community: an evaluation of a humanizing and perspective-taking intervention. Psychol Sex Orientat Gend Divers 2015;2(1):34-42. [doi: 10.1037/sgd0000088]
31. van Zomeren M, Postmes T, Spears R. Toward an integrative social identity model of collective action: a quantitative research synthesis of three socio-psychological perspectives. Psychol Bull 2008 Jul;134(4):504-535. [doi: 10.1037/0033-2909.134.4.504] [Medline: 18605818]
32. Stewart TL, Latu IM, Branscombe NR, Denney HT. Yes we can!: prejudice reduction through seeing (inequality) and believing (in social change). Psychol Sci 2010 Nov;21(11):1557-1562. [doi: 10.1177/0956797610385354] [Medline: 20889931]
33. Pettigrew TF, Tropp LR. A meta-analytic test of intergroup contact theory. J Pers Soc Psychol 2006 May;90(5):751-783. [doi: 10.1037/0022-3514.90.5.751] [Medline: 16737372]
34. Al Ramiah A, Hewstone M. Intergroup contact as a tool for reducing, resolving, and preventing intergroup conflict: evidence, limitations, and potential. Am Psychol 2013 Oct;68(7):527-542. [doi: 10.1037/a0032603] [Medline: 24128316]
35. Testing the effectiveness of an intervention to improve Romanian teachers' outcomes related to LGBT+ inclusion. ISRCTN. 2023. URL: https://www.isrctn.com/ISRCTN84290049 [accessed 2025-12-12]
36. Cohen J. The earth is round (p <.05). Am Psychol 1994;49(12):997-1003. [doi: 10.1037/0003-066X.49.12.997] [Medline: 8122813]
37. Herek GM, McLemore KA. Attitudes Toward Lesbians and Gay Men Scale. In: Davis CM, Yarber WL, Bauserman R, Schreer G, Davis SL, editors. Handbook of Sexuality-Related Measures, 2nd edition: Sage Publications; 1998:392-394.
38. Wright LW, Adams HE, Bernat J. Development and validation of the homophobia scale. J Psychopathol Behav Assess 1999 Dec;21(4):337-347. [doi: 10.1023/A:1022172816258]
39. Altemeyer B. Enemies of Freedom: Understanding Right-Wing Authoritarianism: Jossey-Bass; 1988. URL: https://cmc.marmot.org/Record/.b12238223 [accessed 2025-12-12]
40. Schwarzer R, Jerusalem M. Self-efficacy as a resource factor in stress appraisal processes. In: Schwarzer R, editor. Self-Efficacy: Thought Control of Action: Hemisphere Publishing Corp; 1992:195-213 URL: https://psycnet.apa.org/record/1992-97719-009 [accessed 2025-12-12]
41. Lavrakas PJ. Feeling thermometer. In: Encyclopedia of Survey Research Methods: Sage; 2008:275-277. [doi: 10.4135/9781412963947]
42. Hodson G, Choma BL, Boisvert J, Hafer CL, MacInnis CC, Costello K. The role of intergroup disgust in predicting negative outgroup evaluations. J Exp Soc Psychol 2013 Mar;49(2):195-205. [doi: 10.1016/j.jesp.2012.11.002]
43. Stephan WG, Stephan CW. Intergroup anxiety. J Soc Issues 1985 Oct;41(3):157-175. [doi: 10.1111/j.1540-4560.1985.tb01134.x]

XSL•FO
RenderX

44. Dessel AB. Measuring the effects of intergroup dialogue on teachers' attitudes, feelings, and behaviors regarding lesbian, gay, and bisexual students and parents [Dissertation]. : University of Tennessee; 2008 URL: https://trace.tennessee.edu/utk_graddiss/426/ [accessed 2025-12-12]
45. Spreng RN, McKinnon MC, Mar RA, Levine B. The Toronto Empathy Questionnaire: scale development and initial validation of a factor-analytic solution to multiple empathy measures. J Pers Assess 2009 Jan;91(1):62-71. [doi: 10.1080/00223890802484381] [Medline: 19085285]
46. Koenig HG, Büssing A. The Duke University Religion Index (DUREL): a five-item measure for use in epidemological studies. Religions (Basel) 2010;1(1):78-85. [doi: 10.3390/rel1010078]
47. IBM SPSS Statistics for Windows, Version 240. IBM Corp. 2016. URL: https://www.ibm.com/support/pages/release-notes-ibm-spss-statistics-240 [accessed 2025-12-12]
48. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Ann Statist 2001;29(4):1165-1188. [doi: 10.1214/aos/1013699998]
49. R Core Team. R: a language and environment for statistical computing. : R Foundation for Statistical Computing; 2025 URL: https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf [accessed 2025-12-12]
50. Cramwinckel FM, Scheepers DT, Wilderjans TF, de Rooij RJB. Assessing the effects of a real-life contact intervention on prejudice toward LGBT people. Arch Sex Behav 2021 Oct;50(7):3035-3051. [doi: 10.1007/s10508-021-02046-0] [Medline: 34505215]
51. Global education monitoring report 2020: gender report, a new generation: 25 years of efforts for gender equality in education. : United Nations Educational, Scientific and Cultural Organization; 2020 URL: https://www.unesco.org/gem-report/en/publication/new-generation-25-years-efforts-gender-equality-education [accessed 2025-10-09]
52. Report on the state of pre-university education in Romania 2019 [Report in Romanian]. : Ministerul Educatiei; 2020 URL: https://www.edu.ro/sites/default/files/_fi%C8%99iere/Minister/2020/Transparenta/Stare%20invatamant/Stare%20preuniversitar_rev_5.07.2021.pdf [accessed 2025-10-09]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
FDR: false discovery rate
HS: Homophobia Scale
LGB: lesbian, gay, and bisexual
LGBTQ+: lesbian, gay, bisexual, transgender/transsexual, queer, and other minority sexual orientations and gender identities

Edited by A Kushniruk; submitted 29.Jun.2024; peer-reviewed by R Willems; revised version received 03.Oct.2025; accepted 20.Nov.2025; published 16.Jan.2026.

Please cite as:
Sălăgean N, Latu IM, Larsen TMB, Isbă oiu AB, Sava FA
An Internet-Delivered Intervention to Reduce LGBTQ+ Prejudice Among Romanian Teachers: Randomized Controlled Trial
JMIR Hum Factors 2026;13:e63787
URL: https://humanfactors.jmir.org/2026/1/e63787
doi:10.2196/63787

© Nastasia Sălăgean, Ioana Maria Latu, Torill Marie Bogsnes Larsen, Andreea Bogdana Isbă oiu, Florin Alin Sava. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 16.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Understanding Psychologists' Usage, Knowledge, and Attitudes Toward Digital Mental Health Solutions for Refugees and Migrants: Exploratory Cross-Sectional Survey in Sweden

Kristofer Vernmark[1], PhD; Anahita Geranmayeh[2], MSc; Naira Topooco[3], PhD; Gerhard Andersson[1,4,5], PhD; Shervin Shahnavaz[2], PhD

[1]Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden

[2]Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, & Stockholm Health Care Services, Stockholm, Region Stockholm, Sweden

[3]Department of Psychology, Uppsala University, Uppsala, Sweden

[4]Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden

[5]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

Corresponding Author:
Kristofer Vernmark, PhD
Department of Behavioural Sciences and Learning
Linköping University
Campus Valla
Linköping, 58183
Sweden
Phone: 46 733211074
Email: kristofer.vernmark@liu.se

Abstract

Background: The rising number of refugees and migrants has created growing mental health needs that health care systems struggle to address. Providing assessment and treatment for mental health problems in a digital format could help increase access to care and facilitate the provision of adapted interventions. Psychologists are key stakeholders in the delivery and influence of clinical services within routine care settings, but there are limited data on their perspectives regarding the use of digital solutions to assess and treat common mental health problems in refugees and migrants.

Objective: This study aimed to examine psychologists' usage, knowledge, and attitudes toward digital mental health solutions for assessing and treating common mental health problems in refugees and migrants within the Swedish health care system.

Methods: A cross-sectional online survey was conducted among psychologists in Sweden between December 2023 and February 2024. Responses included Likert-scale items and categorical variables, which were analyzed using descriptive statistics, independent samples *t* tests, and Fisher exact test to explore differences between subgroups.

Results: A total of 81 psychologists responded to the survey. Among them, 58 (72%) were women, and nearly half (40/81, 49%) worked in a public health care region. Respondents showed the highest acceptance for guided internet-based cognitive behavioral therapy (ICBT), blended treatment, and videoconferencing therapy. Only 20% (16/81) reported using digital solutions for refugees or migrants with mental health problems. Most respondents had low or very low knowledge of digital assessment and screening (61/81, 75%) and digital treatment (58/81, 72%) for these groups. Those using digital formats for refugees and migrants, or working in a setting that did so, had significantly higher ratings on all 5 knowledge items compared to those that did not ($P<.001$ to $P=.01$). Respondents emphasized the importance of digital solutions being provided in refugees' and migrants' native languages (70/81, 86%) and being culturally adapted (56/81, 69%). Those using digital formats for refugees and migrants considered cultural adaptation less necessary ($P=.05$). The preferred implementation approach was through specialized or decentralized units in primary care (66/81, 81%).

Conclusions: While psychologists recognize the potential of digital mental health solutions, significant barriers remain, including limited knowledge and experience with using digital formats for refugees and migrants. Psychologists prefer digital solutions in the native language of refugees and migrants that are implemented at the primary care level. The need for cultural adaptation should be further explored. Addressing psychologists' preferences could facilitate the future integration and implementation of digital formats for refugees and migrants in routine care settings.

XSL•FO
RenderX

(JMIR Hum Factors 2026;13:e75263) doi:10.2196/75263

KEYWORDS

digital mental health; internet-based; ICBT; psychologist; stakeholder; refugee; migrant; implementation; health care; knowledge; attitudes

Introduction

The world is experiencing an unprecedented rise in the number of forcibly displaced people. By the end of 2024, over 123 million people across the globe were displaced or stateless, having left their homes due to war and conflict, natural disasters, consequences of climate change, or political changes, including 31 million refugees [1]. Refugees often endure pre-, peri-, and postmigration challenges that increase the risk of developing mental health problems [2] and complicate integration in the new host country [3]. Research in Swedish and international contexts shows that refugees and certain migrant groups experience elevated symptoms of depression, anxiety, and posttraumatic stress [4-6] and that these problems are also prevalent among children and adolescents [7]. Although effective treatment options exist for these conditions [8], the use of health care services among refugees and migrants differs from that of the native-born population. Refugees and migrants tend to underuse available resources and experience barriers to care due to cultural differences such as language difficulties, lack of awareness, and mental health stigma [9-12], leading to untreated mental health problems and further societal marginalization and exclusion from health care systems.

One way of increasing the reach of mental health interventions and overcoming the lack of access and availability of mental health professionals [13] is by providing digital psychological treatment. This format of delivery has expanded considerably during the last 2 decades [14,15] and was further accelerated during the COVID-19 pandemic [16,17]. Internet-based treatment and videoconferencing therapy have been found effective for common mental health problems, including social phobia, generalized anxiety, panic disorder, posttraumatic stress disorder, and depression [18,19], yielding similar results as face-to-face treatment [20,21]. Other digital formats, such as chat, email therapy, and smartphone apps, are being provided by mental health care professionals and have shown promising results [22-25]. Also, the use of digital material together with face-to-face sessions in a blended format, a combination often preferred by providers [26], has been shown to be effective for common mental health problems [27]. Modern information technology is also used for the assessment of mental health status. Providing questionnaires and screening measures online has been shown to be equivalent to paper and pencil provision [28] and can be used to screen for refugee mental health problems [29]. Digital mental health interventions have been shown to be less effective for culturally and linguistically diverse groups such as ethnic minorities, indigenous communities, migrants, and refugees [30,31]. Few studies so far have examined digital solutions tailored to the specific needs of refugees and migrants [32-34], even though there exist specific frameworks for the cultural adaptation process [35], studies describing the adaptation process [36], and research suggesting that culturally adapting psychological treatment can increase both compliance and effect [37]. There is also a need to learn more on how digital interventions can be scaled and integrated into health care systems in different settings, including factors related to acceptability, practice, and structure [38].

When introducing and implementing a new method, the influence of health care professionals as stakeholders is crucial and considered an important determinant in implementation frameworks [39]. As implementation is essentially a social process, the views of health care professionals also exert a strong influence on dissemination efforts [40,41] and they have been shown to be gatekeepers that strongly influence the implementation of internet-delivered treatment in routine mental health care settings [42]. Concerning mental health services, psychologists constitute an important health care profession often engaged in implementing and providing new services [43]. In Sweden, the term “Psychologist” is a protected title, and psychologists are key stakeholders, as they play an important role in the health care system, assessing and treating common mental health problems in primary care and providing interventions for more severe problems in specialized settings specifically targeting refugee health problems. The psychology profession has undergone a digital transition within the Swedish health care system in private and public settings, both before [44] and during the COVID-19 pandemic [45], with many already using digital modalities for assessment and treatment in their everyday work [15,46]. The opinions of therapists, including psychologists, in this transition have generally been positive [44,47]. However, far less is known about psychologists’ knowledge and attitudes toward the use of digital mental health solutions for refugees and migrants. Sweden experienced a significant increase in the number of refugees and asylum seekers during the 2010s, peaking in 2015, primarily from Arabic-, Farsi-, and Dari-speaking countries such as Syria, Iraq, Iran, and Afghanistan [48], Today, Arabic is the second most commonly spoken language in Sweden [49], and migrants and refugees with insufficient Swedish proficiency often cannot access mental health treatment without interpreter support or the use of translated material. Therefore, there is a specific need to learn more about psychologists’ perspectives of providing digital interventions for refugees and migrants not fluent in the Swedish language.

The objective of the survey was to explore psychologists’ use, knowledge, and attitudes toward digital solutions for the assessment and treatment of common mental health problems in refugees and migrants with limited proficiency in Swedish.

Methods

Study Context

The survey was conducted as part of the “SAHA” (“health” in Arabic) project, an academic collaboration aimed at developing

XSL•FO
RenderX

and evaluating digital assessments and interventions for refugees and migrants speaking Arabic, Dari, or Farsi [50].

Study Procedure and Participants

The Checklist for Reporting Results of Internet E-Surveys (CHERRIES) was used to guide the reporting of the survey [51]. A filled-out version of the checklist can be found in Multimedia Appendix 1. An online self-report survey was distributed between December 2023 and February 2024. A convenience sampling method was used to collect the data. Survey respondents were licensed psychologists working mainly in private, public, or nongovernmental settings in Sweden. The survey intentionally targeted both psychologists working with refugees and migrants and those without such experience but working with interventions for common mental health problems to explore potential differences in responses based on workplace and experience with the target groups. Study information and an invitation to participate in the survey were distributed using a variety of digital channels of communication available for the research group. This included emails to selected individual representatives from a private psychology consultant company, a private health care provider, 7 public health care settings, 2 nongovernmental organizations (NGOs), and 2 professional interest groups for psychologists. Selected representatives received information about the study to distribute within their organizations. Recruitment also included distributing information about the study through mailing lists for employees at 2 public research and clinical centers, one mailing list for cognitive behavioral therapy therapists, and social media posts in 6 closed professional Facebook (Meta Platforms) groups for psychologists (≈200-10,000 members). Data on actual reach through these different distribution channels are not available. Posts and emails described the study as a survey for psychologists about their perspectives on using digital solutions for mental health problems in refugees and migrants and included a link to the online survey. Reminders were sent once or twice through the above-mentioned distribution channels. Invitations included a brief description of the study, the intended respondents, and a link to the survey (see Multimedia Appendix 2).

Survey Development

The authors KV and AG developed the initial version of the survey, drawing on a previous study by the research group at Linköping University that examined knowledge, attitudes, acceptability, and near-future expectations regarding digital treatment formats for depression [26]. Questions considering the provision of care were added to further address implementation perspectives. The new survey was reviewed by all researchers within the SAHA project who had expertise in digital assessment and treatment, refugee and migrant groups, and cultural adaptations. Based on the team review, changes were made concerning terminology, comprehensibility, the inclusion of topics related to the research project, themes regarding refugees and digital mental health, and alignment with overarching analytic entities such as knowledge and attitudes. A beta version of the survey was distributed again to the SAHA research group, as well as 3 psychologists representing the public, private, and nongovernmental (NGO) sectors, to test for contextual fit. Based on this feedback, the survey was again modified, including adjustments of important terminology, grammatical errors, and clarification of how questions were related to the purpose of the study, until a final version of the survey named SAHA-Stakeholder was decided upon. The terms “migrant” and “refugee” were defined at the beginning of the survey for respondents and were used consistently throughout. As phrased, these terms included individuals with no or limited proficiency in Swedish and did not specify current residence status (eg, individuals with no residence permit, asylum seekers, or holders of temporary or permanent permits).

The final version of the survey included 26 closed questions and 5 open-ended free text questions. It was distributed through the Iterapi platform (Linköping University, Sweden), a secure and widely used solution for providing online surveys, questionnaires, and internet-based interventions [52], and took approximately 15-20 minutes to complete. Two of the open-ended questions, addressing specific barriers and facilitators to implementation, were analyzed separately using conventional content analysis and will be reported in a separate manuscript. Closed questions were conveyed as binary questions (yes or no), nominal scale alternatives (eg, children, youth and adolescents, adults, and older adults), or multiple-choice selections. All knowledge questions used 5-point Likert scale answers, ranging from 1 (very low) to 5 (very high). Three of the closed questions included an option for the respondent to provide a free-text answer if the available answers did not apply to the respondent. Data on ethnicity were not collected due to restrictions on this type of data collection in Sweden. The survey was developed and distributed in Swedish. For a translated English version of the survey and a detailed description of the Iterapi platform, see Multimedia Appendices 3 and 4.

Data Analysis

All quantitative data were processed and analyzed using descriptive and inferential statistical procedures in IBM SPSS Statistics (version 29.0; IBM Corp). All inferential statistics were considered exploratory due to the limited sample size and the absence of an a priori power analysis. Responses to the survey questions on respondent characteristics from the survey questions (items 1-8, 12, and 13) were recoded into binary variables to ensure adequate cell sizes in correlational analyses and *t* tests, thereby supporting valid statistical comparisons and improving the interpretability of results. Details of the recategorization are provided in Multimedia Appendix 5.

Due to the low expected frequencies in some cells, Fisher exact test was used as an alternative to the chi-square test for analyzing nominal and ordinal data. Moreover, adjusted standardized residuals were examined to assess the impact of each cell in analyses exceeding the 2x2 table format, with a value of 1.96 indicating significant deviations from expected frequencies. Since this analysis provides descriptive indicators of cell contributions rather than effect size measures, and most tables exceeded the 2x2 format, odds ratios and confidence intervals were not calculated.

Two-tailed *t* tests were used to analyze Likert-scale responses to the 5 knowledge questions and compare groups using the

XSL•FO
RenderX

binary categories. The data were treated as continuous variables, in line with the common use of parametric tests for Likert-type scales in social science research [53-55]. The Shapiro-Wilk test results indicated that the data were not normally distributed. However, *t* tests are generally considered robust to such violations, based on the central limit theorem, when sample sizes are well above 30, such as in this study. To assess the robustness of the findings given these deviations, a nonparametric alternative—the Mann-Whitney *U* test—was conducted as a sensitivity analysis. The results were consistent with those of the independent samples *t* test, with one exception. Free-text responses from the 3 open-ended questions on currently used digital tools, preferred digital tools, and gatekeepers, as well as from the 3 "Other" response options, were grouped into categories but not formally analyzed. They are included in the discussion section only to provide contextual understanding.

Ethical Considerations

Ethical approval was obtained from the Swedish Ethical Review Authority (2022-04274-02). All participants provided informed consent digitally before accessing the survey. Participation was voluntary, and respondents could not proceed to the questionnaire without providing consent (see Multimedia Appendix 6). Participants were informed about the purpose of the study, the voluntary nature of participation, their right to withdraw at any time prior to submission, and how their data would be used. Data were collected anonymously through the online survey platform and stored securely in accordance with applicable data protection regulations. Only the research team had access to the data. The results are reported in aggregate form to prevent identification of individual participants. No financial or other incentives were provided for participation.

Results

Sample Characteristics

The study website received 501 unique visitors, of whom 122 provided informed consent. Among these, 111 respondents started the survey, and 81 (73%) completed it. Only completers were included in the analysis. Sample characteristics of the respondents can be found in Table 1. The most common free-text answer in the "Other" category for "Level of care" was working in a private practice (n=3), while 10 respondents did not provide an explanation.

XSL•FO
RenderX

Table 1. Sample characteristics of the respondents (N=81). Percentages may not total 100% due to rounding.

Characteristics	Values, n (%)
Age (years)	
25-34	23 (28)
35-44	30 (37)
45-54	15 (18)
55-64	10 (12)
≥65	3 (4)
Gender	
Women	58 (72)
Men	19 (23)
Nonbinary	1 (1)
Unsure	1 (1)
Prefer not to answer	2 (2)
Organization	
Health care region	40 (49)
Municipality	3 (4)
Government	7 (9)
Private	19 (23)
Sole proprietorship	1 (1)
Nonprofit or NGO[a]	8 (10)
Other	3 (4)
Level of care	
Primary care	21 (26)
Specialized care	32 (40)
Other	19 (23)
Not applicable	9 (11)
Targeted age group (years)	
Children (0-15)	14 (17)
Adolescents or young adults (16-25)	10 (12)
Adults (26-64)	56 (69)
Older adults (≥65)	1 (1)
Manager position	
Yes	7 (9)
No	74 (91)

[a]NGO: nongovernmental organization.

Experience of Refugee and Migrant Mental Health Interventions

Of respondents, 28 out of 81 (35%) reported that they worked in a setting that provided interventions for mental health problems that were specifically targeting refugees or migrants. A large proportion of respondents, 60 out of 81 (74%), reported having experience working with Arabic-, Dari-, or Farsi-speaking refugees or migrants. Most respondents did not speak Arabic, Dari, or Farsi, apart from 3 who reported being fluent in Dari or Farsi and 2 who had a limited understanding of Arabic. In terms of first-hand personal experience working with mental health problems in refugees, respondents stated that they had experience of psychotherapeutic or psychological treatments (48/81, 59%), assessment or diagnosis (41/81, 51%), psychosocial support (30/81, 37%), referrals (20/81, 25%), prevention (8/81, 10%), and medical or pharmacological treatment (2/81, 2%). A total of 15 respondents (18%) had no firsthand experience working with refugees or migrants.

XSL•FO
RenderX

Use of Digital Solutions for Migrants and Refugees

Respondents reported limited availability of digital solutions at their workplace for addressing mental health problems among refugees and migrants (15/81, 19%). A large proportion of the sample responded that they were not using digital solutions for these groups, or they were using them only for other target groups (see Table 2).

Table 2. Psychologists reported use of digital assessment and treatment for refugees and migrants with mental health problems (N=81).

Reported use of solutions for refugees and migrants	Values, n (%)
Assessment or screening, and treatment	11 (14)
Assessment or screening	3 (4)
Treatment	2 (2)
No, but for other target groups	27 (33)
No use of digital tools	38 (47)

Knowledge

The knowledge of digital interventions for refugees was limited in the sample, with 75% (61/81) rating their knowledge of digital screening and assessment as low or very low, and 72% (58/81) rating their knowledge of digital treatment formats similarly (see Table 3). More than half of the respondents (42/81, 52%) stated a lack of knowledge about the organization of care for refugees and migrants.

Explorative *t* tests indicated that those working in settings that provided interventions for refugees or migrants had significantly more knowledge of mental health assessment and treatment for refugees and migrants (t_{79}=3.920; P<.001), guidelines and organization of care (t_{79}=4.371; P<.001), and cultural considerations and adaptations (t_{79}=2.269; P=.03). Similar results were found for those with personal experience of working with refugees or migrants (t_{79}=3.446; P<.001; t_{79}=3.079; P=.003; t_{79}=2.925; P=.004) and those with specific experience of working with Dari-, Farsi-, and Arabic-speaking refugees or migrants (t_{79}=3.892; P<.001; t_{79}=2.189; P=.03; t_{79}=2.722; P=.008). Respondents in workplace settings that provided interventions for refugees or migrants also had significantly higher knowledge of digital treatment formats (t_{79}=2.978; P=.005). Those working in settings that provided digital formats for mental health problems for refugees or migrants and were themselves using digital formats for refugees or migrants had significantly higher scores on all 5 knowledge questions (t_{79}=2.699 to 3.717; P<.001 to P=.01). Additionally, respondents employed at a health care region rated their knowledge of guidelines and organization of care for refugees significantly lower (t_{79}=–2.419; P=.02), and the sensitivity analysis indicated that psychologists working in specialist care settings reported greater knowledge of digital screening and assessment tools (U=228; Z=–2.113; P=.04).

Table 3. Knowledge about the provision, organization, and cultural adaptation of mental health interventions for refugees and migrants (N=81).

Knowledge about	Frequency[a]	Mean (SD)
Assessment and/or treatment	1=6; 2=20; 3=30; 4=17; 5=8	3.08 (1.08)
Guidelines and organization of care for assessment and/or treatment	1=24; 2=18; 3=24; 4=8; 5=7	2.46 (1.26)
Digital assessment or screening tools	1=41; 2=20; 3=13; 4=4; 5=3	1.86 (1.09)
Digital treatment	1=33; 2=25; 3=13; 4=7; 5=3	2.04 (1.12)
Cultural adaptations of assessment and/or treatment	1=8; 2=21; 3=29; 4=14; 5=9	2.94 (1.13)

[a]Responses were measured on a 5-point Likert scale ranging from 1 (very low) to 5 (very high).

Acceptability and Cultural Adaptation

Most respondents indicated a need to provide assessment and treatment in a digital format for refugees and migrants within the Swedish health care system (56/81, 69%). One in four selected the “Do not know” alternative (21/81, 26%). A significant proportion of respondents stated that digital solutions should be provided in the mother tongue of refugees and migrants (70/81, 86%) and that digital assessment and treatment should be culturally adapted (56/81, 69%). Those currently using, or having personal experience of, digital formats for refugees or migrants had significantly more “No” answers on the question if digital interventions should be culturally adapted than those without experience (P=.48; residual 2.3).

The respondents considered it most acceptable to provide digital treatment for refugees and migrants in the form of guided internet-based cognitive behavioral therapy (ICBT), by video, or in a blended treatment format (see Table 4). Respondents in specialist settings were significantly more positive about the use of video for refugees and migrants (P=.007) than those working in primary care, and women were more positive than men about self-help apps (P=.004).

XSL•FO
RenderX

Table 4. In what format can digital treatment be provided for refugees and/or migrants (multiple choice; N=81).

Digital format	Yes, n (%)	No, n (%)
Guided ICBT[a]	66 (82)	15 (18)
Unguided ICBT	42 (52)	39 (48)
Blended treatment	62 (77)	19 (23)
Internet-based or blended (other than CBT[b])	47 (58)	34 (42)
Video	62 (77)	19 (23)
Self-help Apps	45 (56)	36 (44)
Chat	36 (44)	45 (56)
No digital treatment is suitable	2 (3)	79 (97)

[a]ICBT: internet-based cognitive behavioral therapy.

[b]CBT: cognitive behavioral therapy.

For most diagnoses and common mental health problems in refugees and migrants, a moderate level of severity was considered the highest level that could be treated using digital interventions (see Table 5). Fisher exact test showed that participants employed outside of a public health care region were more likely to report that moderate and severe depression could be treated with digital formats (*P*=.02). In the residual analysis, the "Moderate" response contributed 1.9, and the "Severe" response contributed 2.5. Respondents from public health care regions were more likely to report that posttraumatic stress disorder (PTSD) treatment for refugees and migrants was not suitable for digital formats (*P*=.04), with the "Not at all" response contributing 2.8 in the residual analysis. Those working in settings that had no digital formats for refugee or migrant mental health problems considered mild PTSD to be a more suitable option for digital treatment (*P*=.02).

Table 5. Highest level of severity that can be treated with digital interventions (N=81). Percentages may not total 100% due to rounding.

Diagnose	Not at all, n (%)	Mild, n (%)	Moderate, n (%)	Severe, n (%)	Do not know, n (%)
Depression	2 (2)	16 (20)	44 (54)	6 (7)	13 (16)
Social anxiety disorder	2 (2)	12 (15)	36 (44)	17 (21)	14 (17)
Panic disorder	1 (1)	17 (21)	33 (41)	15 (18)	15 (18)
Generalized anxiety disorder	7 (9)	21 (26)	28 (35)	11 (14)	14 (17)
PTSD[a]	7 (9)	21 (26)	28 (35)	11 (14)	14 (17)
Insomnia	2 (2)	10 (12)	31 (38)	24 (30)	14 (17)
Complicated grief	3 (4)	17 (21)	25 (31)	15 (18)	21 (26)

[a]PTSD: posttraumatic stress disorder.

Preferred Provision of Care and Access to Digital Tools

Most respondents answered "No" to the question of whether refugees and migrants receive the care they need through Sweden's health care system (58/81, 72%). Only 4 said "Yes" (5%), and 19 (23%) chose to answer, "Do not know." Respondents in workplace settings that provided interventions for refugees or migrants had a higher proportion indicating that refugees and migrants do not receive the mental health care they need in the current health care system (*P*=.04; residual=2.6). Those in settings that did not provide such interventions had a higher proportion of "Do not know" responses (*P*=.04; residual=2.0).

Eighteen (22%) respondents worked in settings where digital treatment was already provided, and the other respondents were divided on whether such interventions should be implemented at their workplace (31/81, 38%) or not (32/81, 40%). Those working in settings that did not provide digital formats for refugee or migrant mental health had a significantly higher number of "Does not, but should" responses than those working in settings that did (*P*<.001; residual=2.2). Respondents preferred interventions for refugees and migrants to be provided through specialized units in primary care (38/81, 47%), followed by decentralized dissemination in primary care (28/81, 35%) and secondary or specialized care (6/81, 7%). Significantly more respondents working at the primary care level at the time of this writing believed that digital formats should be provided through specialist settings than those already working in such settings (*P*=.03; residual 2.6). In contrast, respondents working in specialist care settings were more likely to believe that dissemination should occur in nonspecialized settings in primary care compared to psychologists working there (*P*=.03; residual 1.9).

XSL•FO
RenderX

Discussion

Principal Findings

In this survey, 81 psychologists working primarily in public, private, and NGO health care settings provided their perspectives on the use of digital formats for the assessment and treatment of common mental health problems in refugees and migrants. Results from the survey showed that guided ICBT, blended treatment, and videoconferencing therapy were the most accepted digital formats for treating refugee and migrant mental health problems, although respondents had limited experience and low knowledge of using these or other digital solutions for refugees and migrants. There was a strong consensus in favor of providing digital solutions in refugees' native language. Although a majority believed that interventions should also be culturally adapted for better fit, explorative analyses showed that those currently working with digital formats for refugees or migrants considered cultural adaptations to be less necessary. Psychologists preferred digital solutions to be delivered at the primary care level, either through centralized units or a decentralized approach.

Role of Knowledge and Personal Experience

The lack of knowledge and use of digital formats for refugees and migrants among respondents highlights the current landscape in digital mental health, where interventions and assessment procedures are mainly targeting (and being created for) Western, native-speaking populations [56]. This lack of personal experience and knowledge among psychologists could act as a barrier to the use of digital formats, as seen in other studies [57].

Since increasing knowledge about digital solutions has been shown to increase perceived advantages and reduce perceived disadvantages of digital mental health interventions [58], awareness- and knowledge-raising activities for psychologists could serve as an important first step in health care settings where implementation of digital formats is being discussed or planned. The use of digital formats among psychologists in Sweden may facilitate the transition to similar alternatives for refugees and migrants, as therapists' experience in delivering online mental health interventions during COVID-19 was associated with higher acceptance of these formats [59]. For those not using digital formats at the time of this writing, there is a need to examine how such a transition is perceived [44], recognizing that those transitioning may perceive both negative and mixed views on using new technologies for the provision of care [60].

Positive Attitudes and Preferred Digital Solutions

The preference for ICBT, videoconferencing, and blended treatment among respondents may not be surprising, as these are the most commonly used digital delivery formats in the Swedish health care system by psychologists and have been widely studied by researchers [19,27,61]. An interesting finding was that guided ICBT received more positive ratings than blended treatment, which contrasts with earlier surveys showing the opposite [26,57,62]. This could be due to the fact that ICBT programs have been available for 8 years through the public national online platform for internet-delivered treatment and support [63] and are frequently delivered by psychologists. Additionally, it may be preferred because it can facilitate the provision of evidence-based treatment in Arabic, Dari, or Farsi, without requiring the psychologist to be fluent in these languages, thus overcoming the barrier of using an interpreter [64]. The positive attitudes toward the use of video in specialist settings may be explained by the fact that it was the preferred digital solution in the transition to digital formats in psychiatric care during the COVID-19 pandemic [45]. Another perspective is that videoconferencing therapy has been used for diagnoses often treated in specialist settings (eg, PTSD) and can be perceived as more easily adapted for more severe conditions than other digital formats [21]. This and the generally positive attitudes toward digital interventions expressed by psychologists, including the use of smartphone apps among women, could facilitate the transition to digital alternatives, as acceptability has an impact on implementation and scalability [65].

Need for Translated and Culturally Adapted Solutions

Even though most respondents had at least some experience working with refugees or migrants, including working with Arabic-, Dari-, or Farsi-speaking patients, no psychologists in this survey could speak Arabic fluently. Since Arabic is the second most widely spoken mother tongue in Sweden [49], and a significant proportion of refugees originate in countries where Arabic, Dari, or Farsi is spoken [48], there is an obvious gap between the need for linguistically adapted solutions and the language skills of psychologists providing interventions. Increasing the workforce by employing psychologists who speak Arabic, Dari, or Farsi as their mother tongue or educating culture brokers to facilitate understanding between patients and health care providers [10] are viable options but will take time and not be sufficient. Instead, using translated assessment and treatment materials or facilitating the use of interpreters in digital formats, such as video, could swiftly increase access by adapting existing technology and content. The preference for culturally adapted digital solutions seen in this study has been confirmed by other studies, for example, in a survey of New Zealand mental health clinicians, where three-quarters expressed a need to tailor solutions for specific populations, including immigrants and refugees [62]. As existing frameworks simplify the process of cultural adaptation [35,66,67], and culturally adapted interventions have been found to lead to better outcomes [37], adapting digital interventions could increase the possibility of dissemination in routine care settings. An interesting finding in our exploratory analysis was that those already working with digital interventions for refugees or migrants considered cultural adaptations less necessary. It could be that these psychologists had prior experience with nonadapted versions being sufficient, which led them to answer "No" to the question of whether cultural adaptations were necessary. This finding can be seen in the light of a recent meta-analysis that found no added effects for culturally adapted digital interventions in low-income settings [68], making the question of cultural adaptations a complex issue that requires further examination [69].

XSL•FO
RenderX

Appropriate Level of Severity for Digital Interventions

Given that there are few studies on digital interventions specifically targeting refugees [70] and migrants [71,72] at the time of writing, the acceptability among respondents of digital interventions for conditions such as depression and panic disorder could be influenced by research on internet-delivered interventions in general, which supports the efficacy for mild to moderate common mental health problems [73]. An increase in digital mental health research targeting refugees and migrants could help inform key stakeholders, making their opinions more based on evidence and less dependent on personal preferences and attitudes toward digital formats. The acceptability of using digital solutions for more severe mental health problems, such as for severe levels of insomnia, social anxiety, and depression in this study, confirms findings from other surveys [26,74]. This suggests that using digital interventions for more difficult psychiatric conditions should be examined further, both from an effectiveness and provider perspective. On the other hand, respondents working in a public health care region and those working in settings where digital solutions for refugees and migrants were not available responded more cautiously about using digital formats for PTSD. Since PTSD is more common among refugees [4] and often treated in specialized settings in Sweden, the lower degree of acceptability could be attributed to it being considered a more difficult diagnosis to treat, contributing to it being viewed as unsuitable for digital formats or only preferable for milder forms of mental health problems. New innovative digital interventions for PTSD could also have the potential to increase acceptability, being more adapted to refugee needs and degree of complexity [75].

Provision of Care and the Perceived Treatment Gap

The already recognized gap in both high- and low-income settings between the mental health burden of refugees and migrants and their access to care [76,77] was confirmed by respondents in this survey. If digital interventions are to help close this gap, it is essential to understand how they should be provided in routine care settings [78], including through broad dissemination that extends beyond trauma-focused treatment in specialized settings [64]. An interesting finding was that respondents in specialized settings preferred implementation in primary care settings, and vice versa, which highlights the difficulty of organizing care for refugees in the existing health care systems. ICBT has been organized in both a centralized (specialized) and decentralized fashion in Sweden [15]. When comparing 2 organizational models in Swedish primary care, a decentralized and a more concentrated model, both models had specific advantages that could make them viable options also from a refugee perspective [79]. A centralized model could increase therapist and organizational competence of refugee mental health and provision of adapted digital interventions, but a decentralized model could make digital treatment more accessible to refugees as they receive their care in the same setting as others enduring mild to moderate mental health problems. When views of outpatient mental health care deliverers (psychotherapists and psychiatrists) were investigated in a large survey study in Switzerland, almost half of the respondents reported having treated refugees and asylum seekers within the last 12 months with a median waiting time of only 3 weeks [64]. This result indicates that a more decentralized model of provision, which could include digital interventions, has the possibility of increasing access to care for refugees compared to more selective and highly specialized services.

As implementation research highlights the importance of decisions and involvement in outer and inner contexts, along with the influence of individual stakeholders, in successful implementation efforts [80], it is crucial to ensure psychologists feel involved in the setup and implementation of digital solutions. Free text responses on potential gatekeepers showed that respondents were unsure of which individuals or organizations were the most important for the facilitation of successful implementation and stated that gatekeepers were politicians and leaders in health care settings, not health care staff. This finding, along with other data from this study showing generally low knowledge about the organization of care for refugees, may also highlight a knowledge and perceived influence gap, not only in the provision of care but also in understanding the organization of care and the role of one's own profession. One way forward could therefore be to increase the interest of professional psychology societies in helping to establish clinical guidelines and influence the organization and provision of digital mental health in health care contexts [81].

Limitations

The sample size was very small for a national survey and restricts the generalizability of the findings and increases the risk of overinterpretation. As such, all findings should be interpreted cautiously. Also, the use of convenience sampling may introduce self-selection bias into the sample and does not permit an accurate estimation of the number of potential respondents who received information about the survey. To contextualize the sample characteristics, we compared the demographic distribution of respondents with national data on licensed psychologists in Sweden. The sample showed a similar distribution with respect to age and gender [82,83]. However, due to the convenience-based recruitment strategy, the findings cannot be considered representative of all psychologists in Sweden. The survey may also have attracted individuals with an existing interest in digital solutions, and results could reflect a more positive view of digital interventions than what is typical among psychologists. Although the survey aimed to include clinical psychologists in health care settings, 9 (11%) respondents reported that the question about level of care (eg, primary or specialized) was not applicable to their work. These respondents may therefore not fully represent a health care context, which should be considered when interpreting the results.

There are several commonly used terms that overlap in their definition and are closely related, such as "migrant," "immigrant," "refugee," "asylum seeker," and "forcibly displaced individuals." The term "asylum seeker," as well as distinctions within refugees and migrants based on residence status (no permit, temporary permit, and permanent permit), was not included in the survey. This omission may have reduced internal validity, as respondents might have answered differently had the survey differentiated between individuals in the asylum process and those with temporary or permanent residence

XSL•FO
RenderX

permits, given the legal differences in access to health care and exposure to mental health stressors across these groups. Our definitions of refugees, migrants, and digital solutions were included in the introduction to the survey (see Multimedia Appendix 7) and were based on established terminology and a consensus process within the research group. We received no feedback from the participants indicating that further clarification was needed regarding the terms used in the survey.

The inferential analyses were exploratory in nature, as no a priori power analysis was conducted, and the sample size was relatively small. To facilitate statistical analysis and increase the number of data points in each cell, survey responses were combined into binary categories. This categorization may limit the conclusions drawn from the inferential statistics used in the study. The survey included 3 open-ended questions on the used digital tools, preferred digital tools, and gatekeepers for dissemination. Responses were not subjected to formal content analysis procedures due to the design of the questions and the nature of the material provided by participants, which primarily consisted of short answers or simple descriptions.

Future Directions

Psychologists can strongly influence the provision of care and the implementation of new methods, including digital mental health interventions. It is therefore important to examine further how they and other mental health professionals perceive how and if digital interventions for refugees and migrants should be implemented in a health care context [42]. Understanding the perspectives of psychologists who have yet to use digital formats is vital to address potential resistance and ensure successful implementation. There is a need for further stakeholder surveys, in-depth qualitative interviews, continuous monitoring of psychologists' knowledge and skills, and increased use of specific implementation frameworks to guide further analysis and dissemination efforts. Strategies to enhance learning and acceptance of digital formats among health care professionals are crucial, as awareness-raising activities and targeted training programs could help reduce barriers and lead to greater adoption and inclusion of marginalized groups such as refugees and migrants.

Conclusions

This survey represents an important first step in understanding psychologists' perspectives on using digital mental health solutions for refugees and migrants in Sweden. While knowledge and experience were limited, attitudes were generally positive, with guided ICBT, blended treatment, and videoconferencing therapy being the most accepted formats. The emphasis on linguistic accessibility underscores the need for adapted solutions in mental health care provision for refugees and migrants. However, the need for cultural adaptation is more complex and requires further exploration. Despite limitations, such as potential self-selection bias and exploratory statistical analyses, the findings provide valuable insights that can inform future research and implementation efforts. Addressing knowledge gaps, fostering acceptance of digital solutions, and developing scalable, adapted interventions are crucial steps toward alleviating the mental health burden for refugees and migrants in Sweden and beyond.

Acknowledgments

All authors made substantial contributions to the study. KV and AG conceptualized and designed the study. The survey was created by KV and AG with significant input from GA, NT, and SS. All authors contributed to the acquisition of data. KV conducted the statistical analyses. KV wrote the first draft of the manuscript, and all other authors contributed with crucial input on all parts of the manuscript. We want to acknowledge all the researchers in the SAHA research group for their input at various stages of the survey development, as well as the webmaster George Vlaescu for his support with the survey platform. We also extend our thanks to all the respondents who participated in the survey.

Funding

The study was supported by a grant from the Swedish Research Council, "Development and evaluation of internet-supported treatment for recently arrived children and adolescents with psychiatric problems."

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author for noncommercial purposes on reasonable request.

Conflicts of Interest

None declared

Multimedia Appendix 1
Filled out version of the Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[DOCX File , 25 KB - humanfactors_v13i1e75263_app1.docx]

Multimedia Appendix 2

XSL•FO
RenderX

English translation of the recruitment material used to invite participants to the online survey.
[DOCX File , 15 KB - humanfactors_v13i1e75263_app2.docx]

Multimedia Appendix 3
English translation of full online survey.
[DOCX File , 26 KB - humanfactors_v13i1e75263_app3.docx]

Multimedia Appendix 4
Description of the Iterapi platform used for survey administration and data collection.
[DOCX File , 25 KB - humanfactors_v13i1e75263_app4.docx]

Multimedia Appendix 5
Full list of recategorized variables used for statistical analyses.
[DOCX File , 20 KB - humanfactors_v13i1e75263_app5.docx]

Multimedia Appendix 6
English translation of digital informed consent form presented to participants prior to accessing the online survey.
[DOCX File , 19 KB - humanfactors_v13i1e75263_app6.docx]

Multimedia Appendix 7
Definitions of specific terms used in the survey.
[DOCX File , 13 KB - humanfactors_v13i1e75263_app7.docx]

References

1. Global trends: forced displacement in 2024. United Nations High Commissioner for Refugees (UNHCR). URL: https://www.unhcr.org/global-trends-report-2024 [accessed 2025-09-30]
2. Raftree L. Designing safe digital mental health and psycho–social support (MHPSS) for displaced and stateless adolescents. United Nations High Commissioner for Refugees (UNHCR). URL: https://www.unhcr.org/innovation/wp-content/uploads/2023/03/Designing-Safe-Digital-Mental-Health-and-Psychosocial-Support-Executive-Summary.pdf [accessed 2025-10-07]
3. Dang HH, Trinh T, Verme P. Do refugees with better mental health better integrate? Evidence from the building a new life in Australia longitudinal survey. Health Econ 2023;32(12):2819-2835. [doi: 10.1002/hec.4750] [Medline: 37671984]
4. Tinghög P, Malm A, Arwidson C, Sigvardsdotter E, Lundin A, Saboonchi F. Prevalence of mental ill health, traumas and postmigration stress among refugees from Syria resettled in Sweden after 2011: a population-based survey. BMJ Open 2017;7(12):e018899 [FREE Full text] [doi: 10.1136/bmjopen-2017-018899] [Medline: 29289940]
5. Alpak G, Unal A, Bulbul F, Sagaltici E, Bez Y, Altindag A, et al. Post-traumatic stress disorder among Syrian refugees in Turkey: a cross-sectional study. Int J Psychiat Clin 2014;19(1):45-50. [doi: 10.3109/13651501.2014.961930]
6. Bogic M, Njoku A, Priebe S. Long-term mental health of war-refugees: a systematic literature review. BMC Int Health Hum Rights 2015;15:29 [FREE Full text] [doi: 10.1186/s12914-015-0064-9] [Medline: 26510473]
7. Gormez V, Kılıç HN, Orengul AC, Demir MN, Demirlikan Ş, Demirbaş S, et al. Psychopathology and associated risk factors among forcibly displaced syrian children and adolescents. J Immigr Minor Health 2018;20(3):529-535. [doi: 10.1007/s10903-017-0680-7] [Medline: 29204726]
8. Turrini G, Purgato M, Acarturk C, Anttila M, Au T, Ballette F, et al. Efficacy and acceptability of psychosocial interventions in asylum seekers and refugees: systematic review and meta-analysis. Epidemiol Psychiatr Sci 2019;28(4):376-388. [doi: 10.1017/S2045796019000027] [Medline: 30739625]
9. Derr AS. Mental health service use among immigrants in the united states: a systematic review. Psychiatr Serv 2016;67(3):265-274. [doi: 10.1176/appi.ps.201500004] [Medline: 26695493]
10. Kirmayer LJ, Narasiah L, Munoz M, Rashid M, Ryder AG, Guzder J, et al. Common mental health problems in immigrants and refugees: general approach in primary care. Can Med Assoc J 2011;183(12):E959-E967. [doi: 10.1503/cmaj.090292] [Medline: 20603342]
11. Satinsky E, Fuhr DC, Woodward A, Sondorp E, Roberts B. Mental health care utilisation and access among refugees and asylum seekers in Europe: a systematic review. Health Policy 2019;123(9):851-863. [doi: 10.1016/j.healthpol.2019.02.007]
12. Kiselev N, Pfaltz M, Haas F, Schick M, Kappen M, Sijbrandij M, et al. Structural and socio-cultural barriers to accessing mental healthcare among syrian refugees and asylum seekers in Switzerland. Eur J Psychotraumatol 2020;11(1):1717825. [doi: 10.1080/20008198.2020.1717825] [Medline: 32128044]
13. Sijbrandij M, Acarturk C, Bird M, Bryant RA, Burchert S, Carswell K, et al. Strengthening mental health care systems for Syrian refugees in Europe and the middle East: integrating scalable psychological interventions in eight countries. Eur J Psychotraumatol 2017;8(sup2):1388102 [FREE Full text] [doi: 10.1080/20008198.2017.1388102] [Medline: 29163867]

XSL•FO
RenderX

14. Zale A, Lasecke M, Baeza-Hernandez K, Testerman A, Aghakhani S, Muñoz RF, et al. Technology and psychotherapeutic interventions: bibliometric analysis of the past four decades. Internet Interv 2021;25:100425 [FREE Full text] [doi: 10.1016/j.invent.2021.100425] [Medline: 34401384]
15. Vernmark K, Buhrman M, Carlbring P, Hedman-Lagerlöf E, Kaldo V, Andersson G. From research to routine care: a historical review of internet-based cognitive behavioral therapy for adult mental health problems in Sweden. Digit Health 2024;10:20552076241287059. [doi: 10.1177/20552076241287059] [Medline: 39381804]
16. Ellis LA, Meulenbroeks I, Churruca K, Pomare C, Hatem S, Harrison R, et al. The application of e-Mental health in response to COVID-19: scoping review and bibliometric analysis. JMIR Ment Health 2021;8(12):e32948. [doi: 10.2196/32948] [Medline: 34666306]
17. De Witte NAJ, Carlbring P, Etzelmueller A, Nordgreen T, Karekla M, Haddouk L, et al. Online consultations in mental healthcare during the COVID-19 outbreak: an international survey study on professionals' motivations and perceived barriers. Internet Interv 2021;25:100405. [doi: 10.1016/j.invent.2021.100405] [Medline: 34401365]
18. Andersson G, Carlbring P, Titov N, Lindefors N. Internet interventions for adults with anxiety and mood disorders: a narrative umbrella review of recent meta-analyses. Can J Psychiatry 2019;64(7):465-470. [doi: 10.1177/0706743719839381] [Medline: 31096757]
19. Thomas N, McDonald C, de Boer K, Brand RM, Nedeljkovic M, Seabrook L. Review of the current empirical literature on using videoconferencing to deliver individual psychotherapies to adults with mental health problems. Psychol Psychother 2021;94(3):854-883 [FREE Full text] [doi: 10.1111/papt.12332] [Medline: 33620133]
20. Hedman-Lagerlöf E, Carlbring P, Svärdman F, Riper H, Cuijpers P, Andersson G. Therapist-supported Internet-based cognitive behaviour therapy yields similar effects as face-to-face therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. World Psychiatry 2023;22(2):305-314. [doi: 10.1002/wps.21088] [Medline: 37159350]
21. Fernandez E, Woldgabreal Y, Day A, Pham T, Gleich B, Aboujaoude E. Live psychotherapy by video versus in-person: a meta-analysis of efficacy and its relationship to types and targets of treatment. Clin Psychol Psychother 2021;28(6):1535-1549. [doi: 10.1002/cpp.2594] [Medline: 33826190]
22. Vernmark K, Lenndin J, Bjärehed J, Carlsson M, Karlsson J, Oberg J, et al. Internet administered guided self-help versus individualized e-mail therapy: a randomized trial of two versions of CBT for major depression. Behav Res Ther 2010;48(5):368-376. [doi: 10.1016/j.brat.2010.01.005] [Medline: 20152960]
23. Nunes-Zlotkowski KF, Shepherd HL, Beatty L, Butow P, Shaw JM. Blended psychological therapy for the treatment of psychological disorders in adult patients: systematic review and meta-analysis. Interact J Med Res 2024;13:e49660. [doi: 10.2196/49660] [Medline: 39470720]
24. Topooco N, Berg M, Johansson S, Liljethörn L, Radvogin E, Vlaescu G, et al. Chat- and internet-based cognitive-behavioural therapy in treatment of adolescent depression: randomised controlled trial. BJPsych Open 2018;4(4):199-207 [FREE Full text] [doi: 10.1192/bjo.2018.18] [Medline: 29988969]
25. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019;2(1):118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
26. Topooco N, Riper H, Araya R, Berking M, Brunn M, Chevreul K, E-COMPARED consortium. Attitudes towards digital treatment for depression: a European stakeholder survey. Internet Interv 2017;8:1-9 [FREE Full text] [doi: 10.1016/j.invent.2017.01.001] [Medline: 30135823]
27. Erbe D, Eichert HC, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
28. Austin DW, Carlbring P, Richards JC, Andersson G. Internet administration of three commonly used questionnaires in panic research: equivalence to paper administration in Australian and Swedish samples of people with panic disorder. Int J Test 2006;6(1):25-39. [doi: 10.1207/s15327574ijt0601_2]
29. Meurling J, Rondung E, Leiler A, Wasteson E, Andersson G, Richards D, et al. An online tiered screening procedure to identify mental health problems among refugees. BMC Psychiatry 2023;23(1):7 [FREE Full text] [doi: 10.1186/s12888-022-04481-2] [Medline: 36597066]
30. Karyotaki E, Ebert DD, Donkin L, Riper H, Twisk J, Burger S, et al. Do guided internet-based interventions result in clinically relevant changes for patients with depression? An individual participant data meta-analysis. Clin Psychol Rev 2018;63:80-92. [doi: 10.1016/j.cpr.2018.06.007] [Medline: 29940401]
31. Whitehead L, Talevski J, Fatehi F, Beauchamp A. Barriers to and facilitators of digital health among culturally and linguistically diverse populations: qualitative systematic review. J Med Internet Res 2023;25:e42719 [FREE Full text] [doi: 10.2196/42719] [Medline: 36853742]
32. Cuijpers P, Heim E, Ramia JA, Burchert S, Carswell K, Cornelisz I, et al. Effects of a WHO-guided digital health intervention for depression in Syrian refugees in Lebanon: a randomized controlled trial. PLoS Med 2022;19(6):e1004025. [doi: 10.1371/journal.pmed.1004025] [Medline: 35737665]

XSL•FO
RenderX

33. Demetry Y, Wasteson E, Lindegaard T, Abuleil A, Geranmayeh A, Andersson G, et al. Individually tailored and culturally adapted internet-based cognitive behavioral therapy for Arabic-speaking youths with mental health problems in Sweden: qualitative feasibility study. JMIR Form Res 2023;7:e46253 [FREE Full text] [doi: 10.2196/46253] [Medline: 37999955]
34. Lindegaard T, Wasteson E, Demetry Y, Andersson G, Richards D, Shahnavaz S. Investigating the potential of a novel internet-based cognitive behavioural intervention for Dari and Farsi speaking refugee youth: a feasibility study. Internet Interv 2022;28:100533 [FREE Full text] [doi: 10.1016/j.invent.2022.100533] [Medline: 35433279]
35. Heim E, Kohrt BA. Cultural adaptation of scalable psychological interventions: a new conceptual framework. Clin Psychol Eur 2019;1(4):1-12. [doi: 10.32872/cpe.v1i4.37679]
36. Nygren T, Berg M, Sarkohi A, Andersson G. Development of an internet-based cognitive behavioral therapy self-help program for Arabic-speaking immigrants: mixed-methods study. JMIR Res Protoc 2018;7(12):e11872 [FREE Full text] [doi: 10.2196/11872] [Medline: 30563819]
37. Hall GCN, Ibaraki AY, Huang ER, Marti CN, Stice E. A meta-analysis of cultural adaptations of psychological interventions. Behav Ther 2016;47(6):993-1014. [doi: 10.1016/j.beth.2016.09.005] [Medline: 27993346]
38. Woodward A, Burchert S, Barry AS, Broerse JE, Sondorp E, Bold A, et al. Scalability of digital psychological innovations for refugees: a comparative analysis in Egypt, Germany, and Sweden. SSM Ment Health 2023;4:100231. [doi: 10.1016/j.ssmmh.2023.100231]
39. Nilsen P. Overview of theories, models and frameworks in implementation science. In: Handbook on Implementation Science. Massachusetts: Edward Elgar Publishing Limited; 2020:8-31.
40. Damschroder LJ, Reardon CM, Lowery JC. The consolidated framework for implementation research (CFIR). In: Handbook on Implementation Science. Massachusetts: Edward Elgar Publishing Limited; 2020:88-113.
41. Rogers EM. Diffusion of preventive innovations. Addict Behav 2002;27(6):989-993. [doi: 10.1016/s0306-4603(02)00300-3] [Medline: 12369480]
42. Davies F, Shepherd HL, Beatty L, Clark B, Butow P, Shaw J. Implementing web-based therapy in routine mental health care: systematic review of health professionals' perspectives. J Med Internet Res 2020;22(7):e17362 [FREE Full text] [doi: 10.2196/17362] [Medline: 32706713]
43. Graham AK, Lattie EG, Powell BJ, Lyon AR, Smith JD, Schueller SM, et al. Implementation strategies for digital mental health interventions in health care settings. Am Psychol 2020;75(8):1080-1092 [FREE Full text] [doi: 10.1037/amp0000686] [Medline: 33252946]
44. Weineland S, Ribbegårdh R, Kivi M, Bygdell A, Larsson A, Vernmark K, et al. Transitioning from face-to-face treatment to iCBT for youths in primary care - therapists' attitudes and experiences. Internet Interv 2020;22:100356. [doi: 10.1016/j.invent.2020.100356] [Medline: 33318951]
45. Erlandsson A, Forsström D, Rozental A, Werbart A. Accessibility at what price? therapists' experiences of remote psychotherapy with children and adolescents during the COVID-19 pandemic. J Infant Child Adolesc Psychotherapy 2022;21(4):293-308. [doi: 10.1080/15289168.2022.2135935]
46. Blix M, Dahlstrand A. Innovationer inom digitala tjänster för psykisk hälsa - utveckling och hinder. VINNOVA. 2021 Jun 03. URL: https://www.vinnova.se/contentassets/24ca94ab27274b94a32471db28da8ec0/bilaga-3_digitala-tjanster.pdf [accessed 2025-10-07]
47. Kivi M, Eriksson MCM, Hange D, Petersson EL, Björkelund C, Johansson B. Experiences and attitudes of primary care therapists in the implementation and use of internet-based treatment in Swedish primary care settings. Internet Interv 2015;2(3):248-256. [doi: 10.1016/j.invent.2015.06.001]
48. Asylum seekers 2002-2023. SCB. URL: https://www.scb.se/hitta-statistik/sverige-i-siffror/manniskorna-i-sverige/asylsokande-i-sverige/#time [accessed 2024-10-09]
49. Aharon A, Löfdahl M. Arabiska i göteborgs språkliga landskap arabic in gothenburg's linguistic landscape. Institutet för språk och folkminnen. URL: http://urn.kb.se/resolve?urn=urn:nbn:se:sprakochfolkminnen:diva-2508 [accessed 2025-10-07]
50. SAHA - digital interventions for mental health among refugees. Linköping University. URL: https://liu.se/en/research/saha [accessed 2025-03-25]
51. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet e-Surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
52. Vlaescu G, Alasjö A, Miloff A, Carlbring P, Andersson G. Features and functionality of the Iterapi platform for internet-based psychological treatment. Internet Interv 2016;6:107-114 [FREE Full text] [doi: 10.1016/j.invent.2016.09.006] [Medline: 30135819]
53. Sullivan GM, Artino AR. Analyzing and interpreting data from likert-type scales. J Grad Med Educ 2013;5(4):541-542. [doi: 10.4300/JGME-5-4-18] [Medline: 24454995]
54. Winter JD, Dodou D. Five-Point likert items: t test versus Mann–Whitney–Wilcoxon (Addendum added October 2012). Pract Assess Res Eval 2010;15(11). [doi: 10.7275/bj1p-ts64]
55. Willits F, Theodori G, Luloff A. Another Look at Likert Scales. J Rural Soc Sci 2016;31(3):Article 6 [FREE Full text] [doi: 10.34068/jrss]
56. Heim E, Mewes R, Ramia JA, Glaesmer H, Hall B, Harper Shehadeh M, et al. Reporting cultural adaptation in psychological trials – The RECAPT criteria. Clin Psychol Eur 2021;3:1-25. [doi: 10.32872/cpe.6351]

XSL•FO
RenderX

57. Mendes-Santos C, Weiderpass E, Santana R, Andersson G. Portuguese psychologists' attitudes toward internet interventions: exploratory cross-sectional study. JMIR Ment Health 2020;7(4):e16817. [doi: 10.2196/16817] [Medline: 32250273]
58. Donovan CL, Poole C, Boyes N, Redgate J, March S. Australian mental health worker attitudes towards cCBT: What is the role of knowledge? Are there differences? Can we change them? Internet Interv 2015;2(4):372-381. [doi: 10.1016/j.invent.2015.09.001]
59. Aafjes-van Doorn K, Békés V, Prout TA. Grappling with our therapeutic relationship and professional self-doubt during COVID-19: will we use video therapy again? Couns Psychol Q 2020 Jun 03;34(3-4):473-484 [FREE Full text] [doi: 10.1080/09515070.2020.1773404]
60. Ibragimov K, Palma M, Keane G, Ousley J, Crowe M, Carreño C, et al. Shifting to tele-mental health in humanitarian and crisis settings: an evaluation of médecins sans frontières experience during the COVID-19 pandemic. Confl Health 2022;16(1):6 [FREE Full text] [doi: 10.1186/s13031-022-00437-1] [Medline: 35164807]
61. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. World Psychiatry 2019;18(1):20-28 [FREE Full text] [doi: 10.1002/wps.20610] [Medline: 30600624]
62. Rawnsley C, Stasiak K. Unlocking the digital toolbox — a mixed methods survey of New Zealand mental health clinicians' knowledge, use and attitudes towards digital mental health interventions. J Technol Behav Sci 2024;9(4):875-885. [doi: 10.1007/s41347-024-00403-z]
63. Program På Stöd- Och Behandlingsplattformen (SoB). SKR. URL: https://extra.skr.se/skr/tjanster/rapporterochskrifter/publikationer/programpastodochbehandlingsplattformensob.76953.html [accessed 2025-10-07]
64. Kiselev N, Morina N, Schick M, Watzke B, Schnyder U, Pfaltz MC. Barriers to access to outpatient mental health care for refugees and asylum seekers in Switzerland: the therapist's view. BMC Psychiatry 2020;20(1):378. [doi: 10.1186/s12888-020-02783-x] [Medline: 32680485]
65. Klaic M, Kapp S, Hudson P, Chapman W, Denehy L, Story D, et al. Implementability of healthcare interventions: an overview of reviews and development of a conceptual framework. Implement Sci 2022;17(1):10 [FREE Full text] [doi: 10.1186/s13012-021-01171-7] [Medline: 35086538]
66. Cheng VWS, Piper SE, Ottavio A, Davenport TA, Hickie IB. Recommendations for designing health information technologies for mental health drawn from self-determination theory and co-design with culturally diverse populations: template analysis. J Med Internet Res 2021;23(2):e23502 [FREE Full text] [doi: 10.2196/23502] [Medline: 33565985]
67. Bernal G, Jiménez-Chafey MI, Domenech Rodríguez MM. Cultural adaptation of treatments: a resource for considering culture in evidence-based practice. Prof Psychol Res Pract 2009;40(4):361-368. [doi: 10.1037/a0016401]
68. Karyotaki E, Miguel C, Panagiotopoulou OM, Harrer M, Seward N, Sijbrandij M, et al. Digital interventions for common mental disorders in low- and middle-income countries: a systematic review and meta-analysis. Glob Ment Health (Camb) 2023;10:e68 [FREE Full text] [doi: 10.1017/gmh.2023.50] [Medline: 38024798]
69. Blomenkamp M, Kiesel A, Baumeister H, Lehr D, Unterrainer J, Sander LB, et al. Assessing the cultural fit of a digital sleep intervention for refugees in Germany: wualitative study. JMIR Form Res 2025;9:e65412 [FREE Full text] [doi: 10.2196/65412] [Medline: 40179371]
70. Mabil-Atem JM, Gumuskaya O, Wilson RL. Digital mental health interventions for the mental health care of refugees and asylum seekers: integrative literature review. Int J Ment Health Nurs 2024;33(4):760-780. [doi: 10.1111/inm.13283] [Medline: 38291740]
71. Reich H, Zürn D, Mewes R. Engaging turkish immigrants in psychotherapy: development and proof-of-concept study of a culture-tailored, web-based intervention. Clin Psychol Eur 2021;3(4):e5583 [FREE Full text] [doi: 10.32872/cpe.5583] [Medline: 36398285]
72. Fabian KE, Foster KT, Chwastiak L, Turner M, Wagenaar BH. Adapting a transdiagnostic digital mental health intervention for use among immigrant and refugee youth in Seattle: a human-centered design approach. Transl Behav Med 2023;13(11):867-875 [FREE Full text] [doi: 10.1093/tbm/ibad041] [Medline: 37418614]
73. Andrews G, Basu A, Cuijpers P, Craske M, McEvoy P, English C, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: an updated meta-analysis. J Anxiety Disord 2018;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
74. Kerst A, Zielasek J, Gaebel W. Smartphone applications for depression: a systematic literature review and a survey of health care professionals' attitudes towards their use in clinical practice. Eur Arch Psychiatry Clin Neurosci 2020;270(2):139-152. [doi: 10.1007/s00406-018-0974-3] [Medline: 30607530]
75. Riisager LG, Larsen JE, Huniche L, Christiansen TB, Moeller SB. Collaborative development of a self-tracking assisted psychotherapy treatment concept for refugees with complex posttraumatic stress disorder: participatory action research. JMIR Form Res 2025;9:e66663 [FREE Full text] [doi: 10.2196/66663] [Medline: 41056543]
76. Alonso J, Liu Z, Evans-Lacko S, Sadikova E, Sampson N, Chatterji S, et al. Treatment gap for anxiety disorders is global: results of the world mental health surveys in 21 countries. Depress Anxiety 2018;35(3):195-208 [FREE Full text] [doi: 10.1002/da.22711] [Medline: 29356216]
77. Evans-Lacko S, Aguilar-Gaxiola S, Al-Hamzawi A, Alonso J, Benjet C, Bruffaerts R, et al. Socio-economic variations in the mental health treatment gap for people with anxiety, mood, and substance use disorders: results from the WHO World Mental Health (WMH) surveys. Psychol Med 2017;48(9):1560-1571. [doi: 10.1017/s0033291717003336]

XSL•FO
RenderX

78. Folker AP, Mathiasen K, Lauridsen SM, Stenderup E, Dozeman E, Folker MP. Implementing internet-delivered cognitive behavior therapy for common mental health disorders: a comparative case study of implementation challenges perceived by therapists and managers in five European internet services. Internet Interv 2018;11:60-70 [FREE Full text] [doi: 10.1016/j.invent.2018.02.001] [Medline: 30135761]
79. Gervind E, Salem MB, Svanborg C, Nyström ME, Lilja JL, Kaldo V, et al. The influence of organizational models on the implementation of internet-based cognitive behavior therapy in primary care: a mixed methods study using the RE-AIM framework. Internet Interv 2024;35:100698 [FREE Full text] [doi: 10.1016/j.invent.2023.100698] [Medline: 38174208]
80. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated consolidated framework for implementation research based on user feedback. Implement Sci 2022;17(1):75 [FREE Full text] [doi: 10.1186/s13012-022-01245-0] [Medline: 36309746]
81. Nogueira-Leite D, Diniz JM, Cruz-Correia R. Mental health professionals' attitudes toward digital mental health apps and implications for adoption in Portugal: mixed methods study. JMIR Hum Factors 2023;10:e45949 [FREE Full text] [doi: 10.2196/45949] [Medline: 37266977]
82. Resultat av tillgång och efterfrågan på de 22 legitimationsyrkena. Socialstyrelsen. 2022 Feb 28. URL: https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/ovrigt/2022-2-7759-bilaga1.pdf [accessed 2025-10-08]
83. Statistik om legitimerad hälso- och sjukvårdspersonal 2018 samt arbetsmarknadsstatus 2017. Socialstyrelsen. 2019. URL: https://www.socialstyrelsen.se/statistik-och-data/statistik/alla-statistikamnen/halso-och-sjukvardspersonal/ [accessed 2025-10-15]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
ICBT: internet-based cognitive behavioral therapy
NGO: nongovernmental organization
PTSD: posttraumatic stress disorder

Edited by C Jacob; submitted 31.Mar.2025; peer-reviewed by N Kiselev, G Yilmaz; comments to author 06.Sep.2025; accepted 30.Jan.2026; published 03.Mar.2026.

Please cite as:
Vernmark K, Geranmayeh A, Topooco N, Andersson G, Shahnavaz S
Understanding Psychologists' Usage, Knowledge, and Attitudes Toward Digital Mental Health Solutions for Refugees and Migrants: Exploratory Cross-Sectional Survey in Sweden
JMIR Hum Factors 2026;13:e75263
URL: https://humanfactors.jmir.org/2026/1/e75263
doi:10.2196/75263
PMID:

©Kristofer Vernmark, Anahita Geranmayeh, Naira Topooco, Gerhard Andersson, Shervin Shahnavaz. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 03.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploring Response Patterns to Motivational Messages Supporting Physical Activity: Latent Class Analysis

Chihiro Moriishi[1*], PhD; Takeyuki Oba[1*], PhD; Keisuke Takano[1], PhD; Kentaro Katahira[1], PhD; Kenta Kimura[1,2], PhD

[1]Human Informatics and Interaction Research Institute, National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, Japan

[2]Faculty of Letters, Hosei University, Chiyoda-ku, Tokyo, Japan

[*]these authors contributed equally

Corresponding Author:
Keisuke Takano, PhD
Human Informatics and Interaction Research Institute
National Institute of Advanced Industrial Science and Technology
1-1-1 Higashi
Tsukuba, Ibaraki, 305-8566
Japan
Phone: 81 80 2202 0737
Email: keisuke.takano@aist.go.jp

Abstract

Background: In mobile health care, text messages play an important role in improving physical activity. Recent studies have developed message banks based on theories, including the behavior change technique (BCT) taxonomy. However, little evidence is available for individual differences (ie, who responds to which BCTs presented in messages), which is crucial for optimizing message delivery.

Objective: We investigated how individuals perceive messages supporting physical activity and what clusters of individuals are identified by their responses.

Methods: Japanese-speaking adults (n=2859; mean age 54.5, SD 17.5 years; 1486 women) were presented with messages conveying different BCTs and rated how motivational each message was. The motivation ratings were subjected to latent class analysis to identify clusters of individuals per motivation rating.

Results: A total of 7 clusters were identified. Two clusters gave consistently high ratings across BCT types; one showed a general receptivity to all messages, while the other showed a clearer preference for information-based BCTs. Two clusters showed moderate ratings, both preferring information about consequences but differing in their additional preferences for goal setting vs rewards. Two clusters gave overall low ratings and typically included less active individuals in the preaction stages, both showing a preference for information-based BCTs. The remaining cluster showed the greatest variability in BCT preferences, with the strongest preference for salience of consequences.

Conclusions: These results highlight individual differences in perceived motivations across BCTs, informing what BCTs should be prioritized in delivery. The practical implications for message tailoring are also discussed.

(JMIR Hum Factors 2026;13:e91917) doi:10.2196/91917

KEYWORDS

behavior change technique; clustering algorithm; digital health; physical activity; text message

Introduction

Background

The rapid growth of mobile technology enables remote, interactive, and timely access to relevant health information, expanding the applicability of health behavior change interventions to promote physical activity (PA) [1]. A popular intervention modality is the use of SMS text messages to deliver educational materials, health care tips, individual goals and achievements, and other motivational information. Text message interventions (TMIs) are considered a promising form of mobile health (mHealth) intervention for PA [2] because smartphone ownership is very high and the implementation cost is relatively low (eg, an estimated average cost per message is US $0.10)

XSL•FO
RenderX

[3]. Furthermore, messages allow for quick and easy access to necessary information and require little or no effort for information search, which helps message recipients maintain their interest in and commitment to health behavior goals [4]. Another important advantage is that messages can be tailored to recipients' demographic characteristics and psychosocial attributes, or even personalized for individual recipients (eg, to include their names in message headers) to enhance message acceptability and the efficacy of an intervention [5]. TMIs can also be scaled up as ecological momentary or just-in-time adaptive interventions [6] if messages are designed to target a particular moment, such as when a recipient is most likely to accept and respond to messages.

A literature review [7] identified 6 TMI trials targeting PA, of which 3 reported statistically significant effects of TMIs. A meta-analysis [8] synthesized the results of 10 trials, most of which assessed self-reported PA as an outcome, and estimated an effect size of Cohen *d*=0.50. Another meta-analysis [2] of 13 trials targeting objectively measured PA concluded that TMIs lead to significantly greater postintervention steps (*d*=0.38), whereas the effect on moderate to vigorous PA was smaller and marginal (*d*=0.31). Notably, both meta-analyses documented moderate to high heterogeneity in efficacy [2,8], which can be at least partly attributed to the variability in TMI delivery and implementation. Therefore, exploring how TMIs can be optimized is important, particularly to clarify to whom and what messages should be delivered.

Researchers have developed message banks to address this challenge. Several different messages are registered in a bank, from which practitioners and app-implemented algorithms can select the appropriate ones to deliver to their client users. Text messages take various forms in terms of content and tone, depending on the purpose of the TMI (eg, providing education, enhancing motivation, delivering feedback related to goals, and sending reminders for participants' PA plans) [2]. The development of messages is also based on various sources, such as feedback from experts and patients [9], clinical guidelines [10], and reviews of previous studies [11].

Messages are typically tailored to a target population based on their demographic and psychosocial characteristics. Pathak et al. [12] developed messages for low-income minority individuals with comorbid diabetes and depression by lowering the reading level (below the sixth-grade level) and ensuring cultural and linguistic appropriateness. Feedback from message recipients is also an important source for tailoring messages, which has been suggested to improve the ease of understanding and acceptability of the delivered messages (eg, for patients with chronic obstructive pulmonary disease [9] and coronary heart disease [13]). Moreover, message content and tone can vary across individuals at different stages of behavior change, particularly when integrated into a treatment or intervention. Prior studies have used tailoring focused on recipients' gender and PA levels [14], lifestyle habits among adults with type 2 diabetes (eg, individuals' dietary behavior, PA, and medication) [15], and goal progress among older adults at risk of cardiovascular disease [11]. However, researchers have noted the lack of theoretical underpinnings for several message banks and argued that the design and validation processes are not always transparent [16]. These circumstances critically limit the opportunities for replicating and reproducing TMIs and prevent fair evaluations of message quality [17,18]. Furthermore, having a link to theories appears to directly affect the efficacy of digital interventions. For example, 1 meta-analysis showed larger effects of theory-based interventions than those not based on a specific theory ([19]; however, see also [5,20]). An exceptional study is MacPherson et al [18], who developed messages to deliver in a diabetes prevention program using the behavior change wheel (BCW) [21]. The BCW is an integrative framework used to guide the development of behavior change interventions [22] and can be a theoretical basis for researchers and practitioners to select the intervention approaches or behavior change techniques (BCTs) that are most suitable for addressing individuals' specific barriers and facilitators [18]. BCTs have been structured and numbered in the form of a taxonomy [21] to which 93 different types of BCTs (classified into 16 categories) are registered, such as goal setting—behavior (1.1) and problem-solving/coping planning (1.2). In their message-development process, MacPherson et al [23] identified 43 BCTs relevant for PA (and dietary) behavior change adherence, encompassing barriers and facilitators to engage in target behaviors. For the identified BCTs, messages were developed and refined based on feedback from program coaches and participants. Therefore, each message is tagged with a specific BCT-taxonomy label while maintaining clarity and acceptability for recipients [18]. Example messages include "Think about what exercise goal you want to make today, if any," tagged with the BCT labels of goal setting—behavior (1.1) and outcome (1.3), and "Sometimes we need the support of others to help us stick to our goals. Think about who in your life can help you stick to your exercise plan," labeled as social support—emotional (3.3) and unspecified (3.1). The message bank comprises BCTs that are most frequently implemented in PA intervention, including goal setting, social support, self-monitoring of behavior, and instruction on how to perform the behavior, which have been commonly identified in systematic reviews and meta-analyses on mHealth interventions for various populations (eg, adults, children, and patients with a specific disease [24-31]).

However, as MacPherson et al [18] acknowledged, which specific messages (or BCTs offered in messages) are more effective than others remains unclear. As noted above, although the BCW serves as a foundation for interventions using BCTs, it does not clearly specify which BCTs work best, for whom, in which contexts, and delivered by what means [32]. Although common BCTs, such as goal setting and self-monitoring of behavior, are frequently used in interventions, they are not necessarily perceived as helpful or motivational by all recipients. Thus, an important evidence gap is that individual differences in terms of who finds which messages helpful remain understudied. These issues are of practical importance when tailoring message delivery because knowing which messages should be delivered to whom is an important step toward achieving precision health [33]. To address these individual differences, the Behavior Change Resource Model (BCRM) may be a useful framework [34]. BCRM assumes that BCTs work by acting on individuals' resources and proposes a behavior change framework based on these resources. It

XSL•FO
RenderX

distinguishes three types of resources: facilitating (providing external resources), boosting (building and strengthening internal reflective resources), and nudging (activating affective resources). These 3 resources partly correspond to the capability, opportunity, motivation-behavior (COM-B) model, which is central to the BCW [35], but the 2 frameworks have different roles. COM-B classifies the conditions necessary for behavior. In contrast, BCRM explains how resources are generated and activated in the brain's motivational and reward systems and describes how BCTs act on these resources, based on neurobiological mechanisms. In this way, because BCRM classifies BCTs according to the resources they target, differences in BCT preferences across clusters may reflect differences in the resources each cluster needs or prioritizes. Michaelsen and Esch [34] also suggest identifying the resources needed to move to the next stage of behavior change and selecting BCTs accordingly. For example, individuals in the early stages of behavior change may lack motivation and may prefer BCTs categorized as nudging, which do not require deliberate effort. In contrast, individuals who are already performing the behavior may prefer BCTs categorized as boosting, which involve intentional effort.

Objectives

This study investigated how individuals perceive BCT-tagged messages for improving PA. We focused on individual differences in participants' responses to different BCTs recommended in the messages. Specifically, we assumed that different individuals would find different BCTs to be motivational. Therefore, we used a clustering approach to identify groups of people who showed similar responses to presented messages. Through this analysis, we aimed to determine how many response patterns (ie, clusters) should be assumed in a sample and which messages can be delivered together to a particular group of recipients. Moreover, we explored how frequently implemented BCTs [24-31] would manifest in these clusters. We expected that while most clusters would perceive goal setting and other common BCTs as motivational, some clusters might not prefer them. The analysis was conducted using cross-sectional survey data, in which participants rated how motivational each BCT message was in enhancing PA. Messages were adapted from the bank of MacPherson et al [18], to which we added (1) messages localized for the target population of this study (Japanese-speaking adults) to indicate the exact information source such as a health guideline of the country, and (2) messages framed with gain and loss focuses ([36]: *gain framing*, focusing on benefits of a behavioral alternative, and *loss framing*, emphasizing the associated costs of a behavior) [36-38]. Although framing is not a dimension well integrated into the BCW or BCT taxonomy, it has been applied to behavior change based on prospect theory, which suggests that individuals respond differently depending on whether information is presented from a cost (loss framing) or benefit (gain framing) perspective [36]. However, in the context of PA, it remains unclear which type of framing is more effective [39]. Therefore, examining which individuals are more likely to prefer each type of framing, in combination with BCT preferences, may help inform more precise message delivery design. We exploratorily examined how it appeared across clusters of message response types (eg, a cluster might find loss framing and goal setting more motivational than other message types).

Methods

Participants

Participants (N=2862) were recruited from a database of a survey firm, in which >1.3 million individuals (all residents in Japan) had been registered as potential participants. This study was part of longitudinal surveys started in 2022, and the sampling for this study was conducted in November 2024. The overarching project had multiple purposes, for example, to investigate motivational and environmental factors that influence PA [40] and the prevalence of mHealth tool use among Japanese-speaking adults [41,42]. For this study, invitations were sent to the participants who completed the initial, main survey of the overarching project (n=20,611 [43]). Quota sampling was used to ensure that the data covered each age group and gender. The sample size was determined arbitrarily—we aimed to sample around 100 participants for each age group of 5-years interval for each gender. No missing values were recorded as the survey platform required participants to respond to each item to complete the survey. Participants were allowed to stop responding in the middle of the survey, but in that case, no record was created in the database. Data from 3 participants were deemed unreliable and excluded from analysis (2 reported a height of <100 cm and 1 reported a weight of <10 kg). The final data included 2859 participants. We adhered to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement [44] in conducting and reporting this study.

Ethical Considerations

The study protocol was approved by the Ethics Committee of the National Institute of Advanced Industrial Science and Technology (approval ID: 2022-1279). Participants provided informed consent prior to responding to the survey. Participants were compensated with a voucher for online shopping after the completion of the survey. All data were collected via an online survey company and provided to the research team in anonymized form, such that individual participants could not be identified. The dataset was stored on a password-protected computer accessible only to the research team.

Measures

Message and Motivation Rating

Each participant rated 83 messages, 72 of which were adapted and translated from the bank of MacPherson et al [18], which was originally developed for a type 2 diabetes prevention program conducted in Canada. Messages that included program-specific references or mentions of specific service names were excluded (eg, "Remember the exercise you did at the YMCA with your coach? Try to reach those same target heart rate zones on your own!"). After translation, the content validity of the messages was confirmed by the second and third authors. For the remaining messages, 3 were gain-framed, 3 were loss-framed, and 5 were localized with explicit mentions of the health guidelines and authorities in Japan (as credible

XSL•FO
RenderX

sources; for the specific messages, see Table S1 in Multimedia Appendix 1). Messages were contextualized in terms of digital health, with the instructions indicating that messages were intended to be delivered via AI or a chatbot implemented through a smartphone app supporting PA and exercise. The instructions are as follows:

Below are messages designed to support and motivate exercise and PA. These messages are intended for a broad audience, including those who do not have exercise habits yet and those who already maintain an exercise routine. Please read these messages and rate how motivated you would feel if you received them. The intended context is that individuals seek to change their lifestyle habits to maintain or improve their health, and they received these messages in their daily lives. Please assume that the sender of the messages is a chatbot or AI implemented in an exercise or training app.

Each message was rated on a 5-point scale (1=not at all; 5=very much). The number of BCT labels that appeared in the message bank was not balanced (Table 1). The most frequent label was information about health consequences (5.1), tagged for 14 messages, followed by credible source (9.1), for 11 messages.

XSL•FO
RenderX

Table 1. Number of messages per behavior change technique (BCT) and framing label and mean motivation ratings (n=2859)[a].

BCT/framing	Messages, n (%)	Ratings, mean (SD)
Goals and planning (1)		
Goal setting—behavior (1.1)	1 (0.7)	2.41 (1.06)
Problem solving (1.2)	3 (2.2)	2.29 (1.05)
Goal setting—outcome (1.3)	3 (2.2)	2.41 (1.11)
Action planning (1.4)	7 (5.1)	2.32 (1.07)
Review outcome goals (1.7)	2 (1.5)	2.39 (1.05)
Feedback and monitoring (2)		
Self-monitoring of behavior (2.3)	4 (2.9)	2.25 (1.06)
Social support (3)		
Social support—unspecified (3.1)	5 (3.6)	2.32 (1.09)
Social support—practical (3.2)	1 (0.7)	2.40 (1.14)
Shaping knowledge (4)		
Instruction on how to perform the behavior (4.1)	4 (2.9)	2.40 (1.07)
Behavioral experiments (4.4)	1 (0.7)	2.25 (1.03)
Natural consequences (5)		
Information about health consequences (5.1)	14 (10.2)	2.42 (1.08)
Salience of consequences (5.2)	1 (0.7)	2.59 (1.06)
Information about social and environmental consequences (5.3)	1 (0.7)	2.64 (1.12)
Monitoring of emotional consequences (5.4)	1 (0.7)	2.29 (1.13)
Information about emotional consequences (5.6)	2 (1.5)	2.48 (1.11)
Comparison of behavior (6)		
Social comparison (6.2)	4 (2.9)	2.38 (1.07)
Associations (7)		
Prompts/cues (7.1)	3 (2.2)	2.24 (1.08)
Repetition and substitution (8)		
Behavioral practice/rehearsal (8.1)	1 (0.7)	2.32 (1.06)
Behavioral substitution (8.2)	4 (2.9)	2.39 (1.05)
Habit formation (8.3)	1 (0.7)	2.32 (1.06)
Habit reversal (8.4)	3 (2.2)	2.40 (1.06)
Graded tasks (8.7)	2 (1.5)	2.55 (1.04)
Comparison of outcomes (9)		
Credible source (9.1)	11 (8)	2.35 (1.07)
Pros and cons (9.2)	1 (0.7)	2.40 (1.04)
Reward and threat (10)		
Nonspecific reward (10.3)	7 (5.1)	2.45 (1.07)
Nonspecific incentive (10.6)	1 (0.7)	2.41 (1.07)
Self-incentive (10.7)	2 (1.5)	2.41 (1.08)
Self-reward (10.9)	7 (5.1)	2.45 (1.07)
Regulation (11)		
Reduce negative emotions (11.2)	2 (1.5)	2.36 (1.09)
Conserving mental resources (11.3)	2 (1.5)	2.36 (1.09)
Antecedents (12)		

XSL•FO
RenderX

BCT/framing	Messages, n (%)	Ratings, mean (SD)
Restructuring the physical environment (12.1)	2 (1.5)	2.35 (1.05)
Restructuring the social environment (12.2)	1 (0.7)	2.35 (1.06)
Distraction (12.4)	3 (2.2)	2.37 (1.09)
Identity (13)		
Identification of self as role model (13.1)	3 (2.2)	2.33 (1.11)
Framing/reframing (13.2)	3 (2.2)	2.38 (1.05)
Scheduled consequences (14)		
Reward approximation (14.4)	5 (3.6)	2.46 (1.08)
Reward completion (14.5)	1 (0.7)	2.41 (1.05)
Self-belief (15)		
Verbal persuasion about capability (15.1)	4 (2.9)	2.43 (1.09)
Focus on past success (15.3)	4 (2.9)	2.44 (1.10)
Self-talk (15.4)	3 (2.2)	2.31 (1.06)
Covert learning (16)		
Imaginary reward (16.2)	1 (0.7)	2.36 (1.10)
Gain framing	3 (2.2)	2.40 (1.07)
Loss framing	3 (2.2)	2.35 (1.06)

[a]Some messages had multiple BCT labels.

PA Measure

Subjective PA was assessed using the International Physical Activity Questionnaire-Short Form (IPAQ-SF) [45,46]. The IPAQ-SF asks respondents to report the number of days and duration spent on each of the following three PA dimensions over an average week: (1) walking, (2) moderate-intensity activity, and (3) vigorous-intensity activity. Weekly minutes of PA were calculated and transformed into metabolic equivalents (METs; METs-hour/week) to represent total PA levels. The IPAQ-SF originally included items on sedentary time, which were not used in this study.

Stage-of-Change for PA

Stages of change (SoC) for PA were assessed using the SoC questionnaire [47-50]. Each participant indicated the most applicable statement from the following five statements: “I currently do not exercise and do not intend to start exercising in the future” (Precontemplation); “I currently do not exercise but I am thinking about starting to exercise in the next six months” (Contemplation); “I currently exercise some, but not regularly” (Preparation); “I currently exercise regularly, but have only begun doing so within the last six months” (action); and “I currently exercise regularly and have done so for longer than six months” (Maintenance). The instructions explicitly defined regular exercise as exercising twice or more per week for 20 minutes or longer on each occasion.

Statistical Analysis

Latent class analysis (LCA) was used to identify groups of individuals (ie, clusters) who gave similar ratings to the presented messages [51]. LCA is a type of finite mixture modeling that represents clusters in the form of categorical latent variables estimated from categorical observed variables. We performed LCA on dichotomized motivation ratings (with values coded based on whether they were above or below the overall mean) using maximum likelihood estimation. LCA computes the probabilities for each observation to belong to each cluster; therefore, we attributed each participant to the cluster to which they were most likely to belong [52]. To determine the optimal number of clusters, models were estimated with 2-14 clusters, and model fit was evaluated using the Bayesian Information Criterion (BIC). The characteristics of each identified cluster were inspected based on the (1) mean motivation ratings for each message in each cluster, (2) mean motivation ratings per BCT label in each cluster, and (3) standardized mean ratings per BCT label within each cluster. Within-cluster standardization helped interpret which BCTs each cluster perceived as more (or less) motivational.

Results

Demographic Characteristics and Descriptive Statistics

Table 2 summarizes the participants’ demographic characteristics. More than half of the participants were married and had at least 1 child. Most had graduated from high school, and household income was distributed around JPY 2-6 million (US $12,600-US $37,900), which is comparable to the population distribution in Japan (average JPY 5.45 million [US $34,400]). On average, participants were more active than the recommended level (23 METs-hours/week) by the authorities, and approximately one-third reported being in the maintenance stage. Table 1 presents the mean motivation ratings for each BCT label. Information about social and environmental consequences (5.3) was rated as the most motivational, followed

XSL•FO
RenderX

by salience of consequences (5.2). The least motivational messages were prompts/cues (7.1), self-monitoring of behavior (2.3), and behavioral experiments (4.4).

Table 2. Demographic characteristics and descriptive statistics (n=2859).

Variable	Value
Age (years), mean (SD)	54.49 (17.52)
Women, n (%)	1486 (51.98)
BMI, mean (SD)	21.77 (3.7)
Married, n (%)	1723 (60.27)
One or more children, n (%)	1636 (57.22)
Job, n (%)	1685 (58.94)
Education level, n (%)	
Middle/high school	926 (32.39)
Two-year college or vocational school	627 (21.93)
University or above	1292 (45.19)
Other	14 (0.49)
Household income (JPY; JPY 1=US $0.0063), n (%)	
<2 million	269 (9.41)
2-4 million	659 (23.05)
4-6 million	537 (18.78)
6-8 million	363 (12.7)
8-10 million	211 (7.38)
10 million or above	252 (8.81)
No response or unknown	568 (19.87)
PA[a] (METs[b]-hour/week), mean (SD)	35.84 (56.21)
Stage of change, n (%)	
Precontemplation	767 (26.83)
Contemplation	498 (17.42)
Preparation	519 (18.15)
Action	95 (3.32)
Maintenance	980 (34.28)

[a]PA: physical activity.

[b]MET: metabolic equivalent.

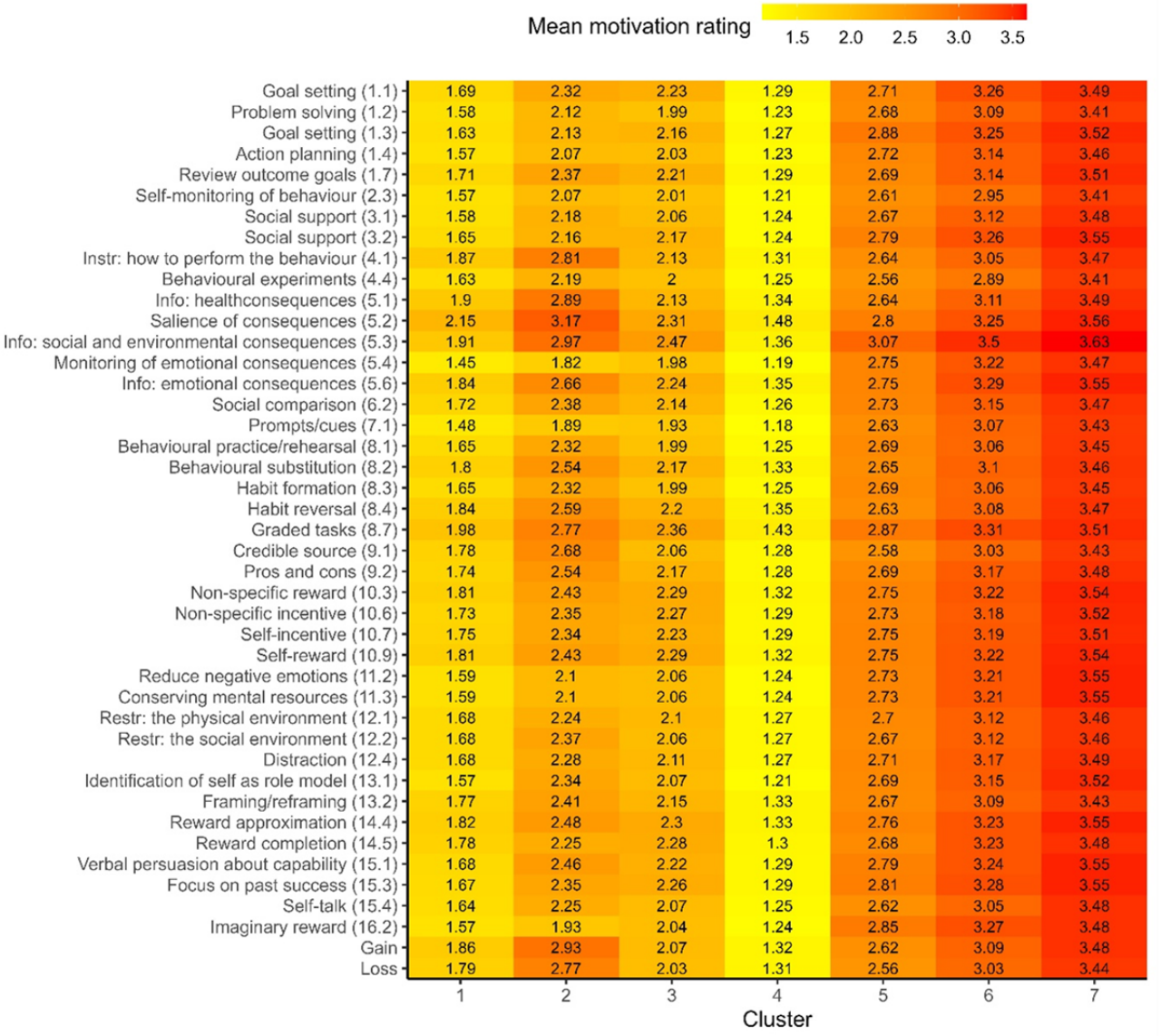
Latent Class Analysis

The LCA identified 7 clusters as the best solution that achieved the minimum BIC (5 clusters=198,459, 6 clusters=197,399, 7 clusters=197,350, 8 clusters=197,413, and higher for the larger number of clusters). The entropy for the 7-cluster solution was 0.943, indicating that the classification accuracy was sufficient.

Among the identified 7 clusters, cluster 7 had the highest motivation ratings for each message, whereas cluster 4 had the lowest ratings (Figure 1). The other clusters fell between clusters 7 and 4, although cluster 2 showed the largest variance across messages, suggesting that it had the most selective preference for different messages.

XSL•FO
RenderX

Figure 1. Mean motivation ratings per behavior change technique and cluster.

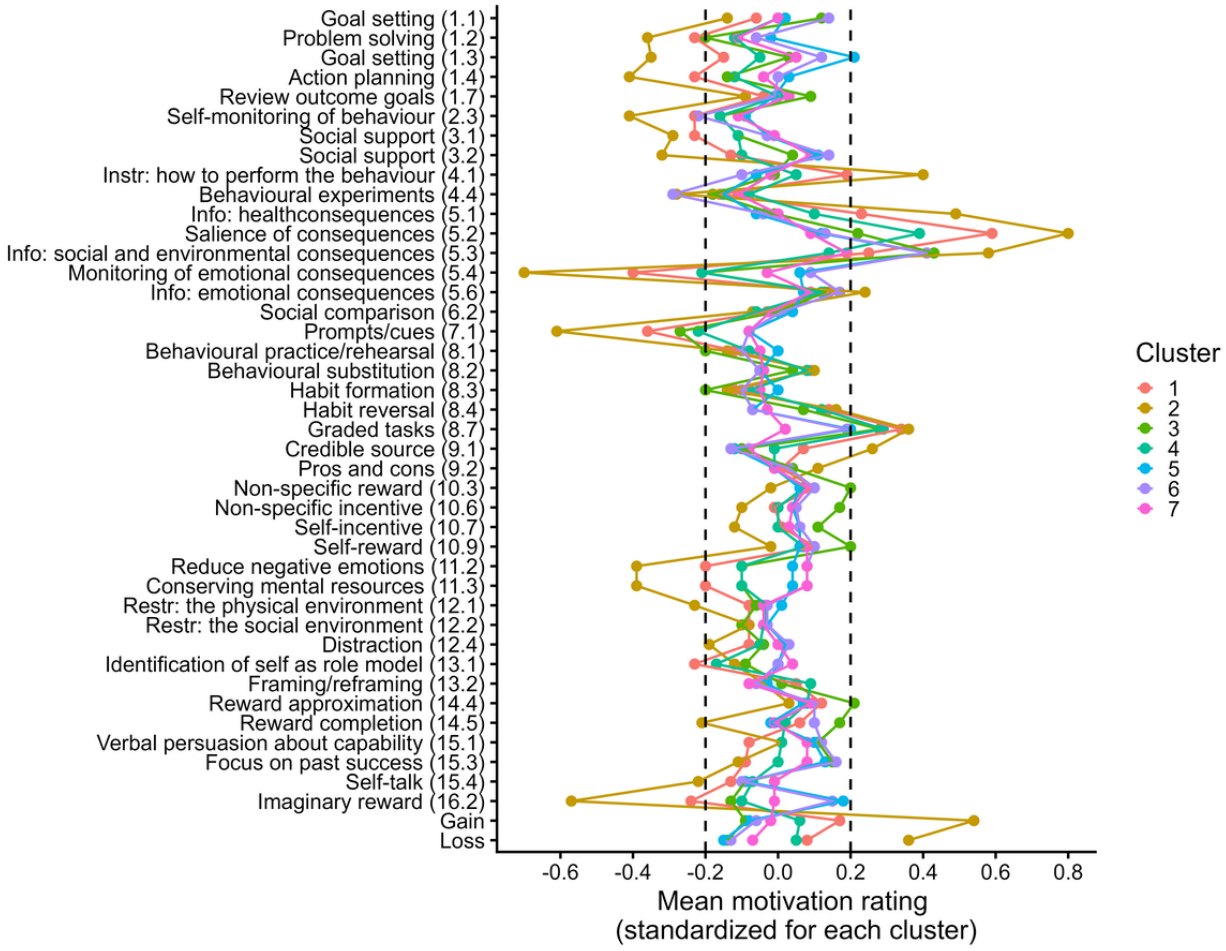


Cluster Characteristics

We calculated the mean motivation ratings per BCT (Figure 1) and standardized them within each cluster to visualize the rating patterns across clusters (Figure 2). Statistical tests did not help interpret between- and within-cluster differences because of the enormous number of possible paired comparisons between BCTs and clusters. Therefore, we focused on (1) the BCTs with the highest and lowest standardized mean motivation ratings within each cluster, (2) the BCTs whose within-cluster standardized mean exceeded the reference threshold in that cluster but not in any other cluster. A within-cluster standardized mean of 0.20 was used as a reference threshold to identify BCTs with meaningful differences in ratings, corresponding to a small effect size based on Cohen's (1988) benchmarks (small: d=0.20; medium: d=0.50; large: d=0.80) [53].

XSL•FO
RenderX

Figure 2. Mean motivation ratings standardized within each cluster.



Cluster 7 gave the highest ratings to any BCTs, with almost no preference for a particular BCT. No BCTs had within-cluster standardized mean values greater than 0.20 or less than –0.20. Cluster 6 also showed high ratings for most BCTs, with most mean motivation ratings above 3.0, but exhibited a clearer preference pattern than cluster 7. Specifically, it showed the highest ratings for information about social and environmental consequences (5.3: within-cluster standardized mean 0.41) and the lowest for behavioral experiments (4.4: within-cluster standardized mean –0.29). Clusters 5 and 3 showed overall moderate ratings for BCTs, with most mean motivation ratings between 2.0 and 3.0. Cluster 5 showed the highest rating for information about social and environmental consequences (5.3: within-cluster standardized mean 0.41), and no BCTs were rated low enough to exceed the reference threshold (no BCTs with a within-cluster standardized mean below –0.20). Goal setting (1.3: within-cluster standardized mean 0.21) exceeded the reference threshold only in this cluster. Similarly, cluster 3 showed the highest ratings for information about social and environmental consequences (5.3: within-cluster standardized mean 0.43) and the lowest for prompts/cues (7.1: within-cluster standardized mean –0.27). Nonspecific rewards (10.3: within-cluster standardized mean 0.20), self-reward (10.9: within-cluster standardized mean 0.20), and reward approximation (14.4: within-cluster standardized mean 0.21) exceeded the reference threshold only in this cluster. Clusters 4 and 1 showed overall low ratings for BCTs, with most mean motivation ratings between 1.0 and 2.0. Cluster 4 showed the highest ratings for salience of consequences (5.2: within-cluster standardized mean 0.39) and the lowest for prompts/cues (7.1: within-cluster standardized mean –0.22). Cluster 1 showed the highest ratings for salience of consequences (5.2: within-cluster standardized mean 0.59) and the lowest for monitoring of emotional consequences (5.4: within-cluster standardized mean –0.40). Identification of self as role model (13.1: within-cluster standardized mean –0.23) exceeded the negative reference threshold only in this cluster. Finally, cluster 2 showed the strongest differentiation in preferences for BCTs. It showed the highest ratings for salience of consequences (5.2: within-cluster standardized mean 0.80) and the lowest for monitoring of emotional consequences (5.4: within-cluster standardized mean –0.70). BCT groups related to goals and planning, feedback and monitoring, and social support did not exceed the positive reference threshold in this cluster, whereas instruction on how to perform the behavior (4.1: within-cluster standardized mean 0.40) exceeded it. High ratings for both gain- and loss-framed messages were also observed (gain: within-cluster standardized mean 0.54; loss: within-cluster standardized mean 0.36).

Among the BCTs commonly used in interventions (ie, goal setting, social support, self-monitoring of behavior, and instruction on how to perform the behavior), goal setting (1.3) exceeded the positive reference threshold only in cluster 5 (within-cluster standardized mean 0.21), whereas it exceeded the negative reference threshold only in cluster 2 (within-cluster

XSL•FO
RenderX

standardized mean –0.35). Instruction on how to perform the behavior (4.1) exceeded the positive reference threshold only in cluster 2 (within-cluster standardized mean 0.40). Self-monitoring of behavior (2.3) exceeded the negative reference threshold in clusters 1, 2, and 6, as did social support (3.1) in clusters 1 and 2, and social support (3.2) in cluster 2.

Table 3 presents the demographic characteristics, self-reported PA levels, and SoC for each cluster. Clusters 7 and 6 were the most male-dominant (191/364, 52% and 288/514, 56%), whereas cluster 2 was the oldest (mean age 63.8 years) among the 7 clusters. BMI showed no significant between-cluster differences. Clusters differed significantly in PA level ($F_{6,\ 2852}$=11.39; P<.001); however, the omnibus effect size was small (ηp^2=.02), indicating that the practical magnitude of between-cluster differences in PA was limited despite statistical significance arising from the large sample. Pairwise comparisons were therefore interpreted with this small overall effect in mind, using the same benchmarks as reference points for effect size interpretation (small: d=0.20; medium: d=0.50; large: d=0.80). Descriptively, cluster 7 showed the highest PA level, followed by cluster 2; this difference was negligible in magnitude (d=0.152). Cluster 6 and cluster 7 differed by a similarly small margin (d=0.200), while cluster 6 and cluster 2 showed almost no difference (d=0.048). Cluster 4 showed the lowest PA level, and the remaining clusters (1, 3, and 5) fell in between, with negligible differences among them (ds<0.05). Approximately half (43%-53%) of clusters 7, 6, and 2 were in the maintenance stage. The least active cluster was cluster 4, in which >80% of the individuals were in the preaction (precontemplation-preparation) stages. In other clusters (1, 3, and 5), the proportions were similar across stages, except for the action stage. Although pairwise comparisons did not detect any significant differences in PA levels between adjacent clusters (eg, cluster 7 vs 2, see Tables S2 and S3 in Multimedia Appendix 1), the high-rating clusters (clusters 7, 6, 5, and 2) had significantly higher PA levels than the low-rating clusters (clusters 1, 3, and 4); t_{2857}=6.062; P<.001; d=0.227). Similarly, the proportion of individuals in the maintenance stage was higher in high-rating clusters than in low-rating ones (odds ratio 2.089, 95% CI 1.78-2.45; P<.001).

Table 3. Demographic characteristics of the latent class analysis–identified clusters.

Variable	Cluster 1 (n=431)	Cluster 2 (n=142)	Cluster 3 (n=338)	Cluster 4 (n=547)	Cluster 5 (n=523)	Cluster 6 (n=514)	Cluster 7 (n=364)	Cluster difference		
								Test statistics	Effect size (ηp^2)	*P* value
Age (years), mean (SD)	59.25 (15.41)	63.75 (13.18)	56.69 (17.24)	53.63 (15.95)	49.86 (17.9)	50.44 (19.15)	56.84 (17.42)	26.08 (6, 2852)[a]	0.05	<.001
Women, n (%)	270 (62.65)	79 (55.63)	189 (55.92)	290 (53.02)	259 (49.52)	226 (43.97)	173 (47.53)	40.11 (6)[b]	N/A[c]	<.001
BMI (kg/m^2), mean (SD)	21.56 (4.00)	21.72 (3.40)	21.96 (3.84)	21.95 (4.05)	21.74 (3.38)	21.63 (3.40)	21.85 (3.56)	0.74 (6, 2852)[a]	0.00	.62
PA[d] (METs[e]-hours/week), mean (SD)	32.22 (48.13)	43.48 (58.13)	34.44 (59.94)	23.06 (53.21)	34.94 (60.30)	40.81 (48.00)	51.93 (64.46)	11.39 (6, 2852)[a]	0.02	<.001
Stage of change, n (%)								380.25 (24)[b]	N/A	<.001
Precontemplation	130 (30.16)	17 (11.97)	79 (23.37)	295 (53.93)	123 (23.52)	76 (14.79)	47 (12.91)			
Contemplation	84 (19.49)	21 (14.79)	68 (20.12)	88 (16.09)	112 (21.41)	78 (15.18)	47 (12.91)			
Preparation	72 (16.71)	30 (21.13)	72 (21.3)	59 (10.79)	106 (20.27)	112 (21.79)	68 (18.68)			
Action	14 (3.25)	5 (3.52)	14 (4.14)	5 (0.91)	23 (4.4)	25 (4.86)	9 (2.47)			
Maintenance	131 (30.39)	69 (48.59)	105 (31.07)	100 (18.28)	159 (30.4)	223 (43.39)	193 (53.02)			

[a]*F* test (*df*).

[b]Chi-square test (*df*).

[c]N/A: not applicable.

[d]PA: physical activity.

[e]MET: metabolic equivalent.

XSL•FO
RenderX

Discussion

Principal Findings

This study investigated how individuals perceive BCTs offered in health messages to support PA. Specifically, we explored which clusters would be identified by individuals' motivation ratings for messages extracted from the extended message bank of MacPherson et al [18]. LCA identified 7 clusters as the optimal solution. Overall, clusters 7 and 6 showed higher ratings across messages, clusters 5 and 3 showed moderate ratings, and clusters 1 and 4 showed lower ratings. Cluster 2 showed the most variable ratings across messages. However, within each broad rating level, clusters differed in their preferences for specific BCTs. The observed heterogeneity in BCT preferences across clusters may be partly explained by differences in the psychological resources individuals require at different stages of behavior change, as proposed by the BCRM framework.

Clusters 7 and 6 showed generally high ratings across different message types. Cluster 7 was labeled "General receptivity" based on its uniformly high ratings across all BCT types, with no particular preference or aversion (no BCTs exceeding a within-cluster standardized mean of ±0.20). Individuals in this cluster had the highest PA levels and approximately half were in the maintenance stage, suggesting the establishment of stable PA habits. Cluster 6 was labeled "Preference for low-effort boosting" based on its highest rating within the cluster for information about social and environmental consequences (5.3) and its lowest rating within the cluster for behavioral experiments (4.4). This cluster showed generally high ratings across BCT types, but with a clearer preference pattern than cluster 7. Both BCTs fall under boosting in the BCRM framework, but they may differ in the degree of active engagement required. In the BCRM, reflective resources refer to internal cognitive capacities that support behavior change, such as beliefs about consequences, intentions, and decision-making processes. While behavioral experiments (4.4) may require active self-engagement, information about consequences (5.3) may allow individuals to build reflective resources, such as beliefs about consequences and decision-making processes, more passively. This suggests that individuals in cluster 6 preferred less effortful forms of boosting. Approximately half of the individuals were in the maintenance stage, where behavior tends to become habitual and automatic, and the need for further change is often reduced [35]. Therefore, individuals in this cluster may have preferred approaches that build reflective resources, such as beliefs about consequences, through passive receipt of information, rather than those requiring active cognitive engagement, such as deliberate self-monitoring or behavioral experimentation.

Clusters 5 and 3 showed overall moderate ratings across BCT types. Cluster 5 was labeled "Preference for outcome information and goal setting" based on its highest rating for information about social and environmental consequences (5.3) within the cluster, and goal setting (1.3), which exceeded the reference threshold only in this cluster, with no BCTs rated notably low (no BCTs below a within-cluster standardized mean of –0.20). Cluster 3 was labeled "Preference for outcome information and rewards over prompts" based on its highest rating within the cluster for information about social and environmental consequences (5.3), rewards such as reward approximation (14.4), which exceeded the positive reference threshold only in this cluster, and its lowest rating within the cluster for prompts/cues (7.1). In both clusters, more than half of individuals were in the preaction stages (precontemplation to preparation; ≥65%), in which knowledge and insight are considered necessary to motivate behavior change [35]. Furthermore, nudging BCTs (reward approximation and prompts/cues), which activate affective resources without requiring conscious deliberation [34], were responded to differently between the 2 clusters. Therefore, even within the same stage, cluster 5 was characterized by a preference for building reflective resources through outcome information, whereas cluster 3 showed a greater emphasis on activating affective resources in addition to such information.

Clusters 4 and 1 showed overall low ratings across BCT types. Cluster 4 was labeled "Low receptivity with preference for outcome information" based on its highest rating within the cluster for salience of consequences (5.2) and its lowest rating within the cluster for prompts/cues (7.1). This cluster had the highest proportion of individuals in the precontemplation stage and the lowest PA levels among all clusters. At a stage where motivation was low and the benefits of PA were not well recognized, prompts/cues that encourage PA may be less acceptable (eg, "Reminders can be super helpful when we are trying to form new habits. Put a reminder in your phone, a note on your fridge, or get creative!"). Instead, BCTs that can enhance intrinsic motivation toward behavior change, such as providing information about outcomes, may be more effective for this group. Cluster 1 was labeled "Preference for external outcome information without self-engagement" based on its highest rating within the cluster for information about social and environmental consequences (5.3), its lowest rating within the cluster for monitoring of emotional consequences (5.4), and identification of self as role model (13.1), which exceeded the negative reference threshold only in this cluster. A large proportion of individuals in this cluster were in the preaction stage. In this preaction stage, where knowledge and insight are considered necessary to motivate behavior change, individuals in this cluster may have preferred approaches that provide external information about outcomes, rather than those requiring active self-reflection or role-based engagement.

Finally, cluster 2 was labeled "Selective preference for concrete guidance with framing sensitivity" based on its greatest variability in BCT preferences among all clusters. Specifically, this cluster showed the highest rating within the cluster for salience of consequences (5.2) and the lowest rating within the cluster for monitoring of emotional consequences (5.4). With the exception of instruction on how to perform the behavior (4.1), commonly used BCT groups (ie, goals and planning, feedback and monitoring, and social support) did not exceed the positive within-cluster standardized mean reference threshold in this cluster. In addition, both gain- and loss-framed messages exceeded the positive reference threshold only in this cluster, with a slightly stronger preference for gain framing. Approximately half of the individuals were in the maintenance

XSL•FO
RenderX

stage, with PA levels similar to cluster 7. From the perspective of prospect theory, gain framing is generally more effective for low-risk preventive behaviors, whereas loss framing is considered more effective for high-risk detection behaviors [54]. Given that many individuals in cluster 2 were already engaging in PA, gain framing may have been slightly more preferred in the context of preventive action. Moreover, cluster 2 had the highest mean age among the clusters, and the messages explicitly referred to conditions more commonly associated with older adults, such as hypertension and diabetes (eg, "Eat more leafy greens and less sugar, and exercise more! These simple actions have been shown to decrease the risk of type 2 diabetes"). Therefore, the self-relevance of the messages may have contributed to the preference for gain-framed messages in cluster 2, although the interaction between age and message framing was not directly tested in this study. Future research should further examine the role of age, including whether the inclusion of age-relevant diseases in messages influences framing effects.

Previous systematic reviews have suggested that behavior change interventions for PA frequently implement the following BCTs: goal setting, social support, self-monitoring of behavior, and instruction on how to perform the behavior [24-31]. Our findings revealed between-cluster differences in the motivation ratings for these common BCTs. For example, goal setting (1.3) exceeded the positive within-cluster standardized mean reference threshold only in cluster 5, whereas it fell below the negative reference threshold only in cluster 2. Instruction on how to perform the behavior (4.1) exceeded the positive reference threshold only in cluster 2. In contrast, self-monitoring of behavior (2.3) fell below the negative reference threshold in clusters 1, 2, and 6, as did social support (3.1) in clusters 1 and 2, and social support (3.2) in cluster 2. These results align with previous findings that frequently used BCTs are not necessarily perceived as helpful or motivational [55,56]. Furthermore, self-monitoring attracts different viewpoints from intervention recipients [57]. Some individuals found diary (manual) recordings of PA repetitive and time-consuming, whereas others found it useful to visualize progress and detect barriers to PA. The effect of self-monitoring on PA is also somewhat mixed, and self-monitoring is suggested to work only when delivered with objective monitoring tools (eg, accelerometers) [58]. Future research must disentangle the discrepancies between the perceived and actual efficacy of self-monitoring considering individual differences because some individuals may perceive it as boring but benefit from it by changing their PA habits.

The findings of this study have several practical implications. When designing a TMI with tailored message delivery, the first decision should be whether the target recipients are open to the TMI. Our results suggest that several clusters rate any message as not motivational. These clusters included a higher proportion of individuals in the preaction stages, suggesting that selecting BCTs that target the resources required at this stage may be effective. The preaction stage is characterized by low motivation for behavior change, with a need for knowledge and insight. Therefore, presenting information about outcomes may be beneficial for building reflective resources for behavior change. In contrast, approaches requiring active self-engagement or providing concrete instructions for action may be less acceptable. Additionally, given the limited reflective resources at this stage, the use of rewards may also be somewhat effective as a supplementary approach. Similarly, practitioners could refer to demographic information when tailoring messages to send. Cluster 2 typically included older people, and this cluster showed relatively higher ratings for both gain- and loss-framed messages, suggesting that age may be a relevant factor in framing preferences, although this was not directly tested in this study. Tailoring variables do not have to be limited to demographic and trait-like characteristics, such as PA and SoC. Fu et al [59] suggested that the key to further facilitating personalization would be for an intervention to assess what BCTs have and have not worked for each individual at an early stage of the intervention. This would clarify the additional support that should be offered to meet individuals' changing needs over the progress of an intervention, considering that different BCTs would be needed at different stages of intervention.

Limitations

This study had several limitations. First, we exclusively recruited Japanese-speaking adults with Internet access. Caution should be exercised when interpreting these results in terms of cultural differences and generalizability. Cluster 2 included several older people who showed unique preferences for framed messages. However, these individuals may not be representative of the older population, which is the generation typically facing the digital divide [60]. Furthermore, the potential influence of selection bias should be considered. Because the participants had previously taken part in surveys conducted by our research group [43], the sample might be more motivated, health-aware, and digitally literate than the general population. This may have generally increased positive responses to mHealth interventions and may also have contributed to the relatively high baseline PA levels, as well as the presence of clusters characterized by high responsiveness to messages (eg, cluster 7). Second, motivation was rated cross-sectionally. Therefore, whether each message triggers actual behavior, and not only motivation or intention, remains unclear. While the association between intention and PA is well established, substantial discrepancies have been observed in experimental studies [61]. Therefore, a high level of motivation in response to messages may not necessarily translate directly into increased PA. Additional constructs that strengthen intention (eg, self-regulation and automaticity) have been proposed. Future research should examine, through experimental and longitudinal designs, whether higher motivation leads to engagement in PA or whether it operates through interactions with other factors. Third, standardizing each message for the tone used and resources referred to (eg, national guidelines) was difficult, mainly because we relied on an established message bank. We focused on the BCTs offered in each message as the key dimension of the analysis; however, motivation ratings could be affected by other factors. Our intention was to analyze an existing message bank to leverage individual differences in motivation ratings for tailoring message delivery. However, a future study could use a factorial design to rigorously control for other factors to clarify what (combinations of) factors determine motivation.

XSL•FO
RenderX

Furthermore, fatigue and reduced concentration associated with evaluating 83 messages may have influenced participants' ratings. Fourth, we specified message delivery in the mHealth context, and the participants were instructed to assume that each message was sent by an AI chatbot implemented in a health care app. This was because this study was intended to be the basis for future mHealth TMIs. However, different senders (eg, human health care professionals) may cause the messages to be perceived differently. Beyond the specific delivery context, the text message format itself has inherent limitations. TMIs remain a valuable component of mHealth due to their cost-effectiveness, ease of personalization, and potential to be scaled up as ecological momentary or just-in-time adaptive interventions. However, not all BCTs may be equally suited to this format. For example, self-monitoring typically assumes a bidirectional process in which individuals record behavior and receive feedback accordingly. The messages used in this study merely prompted recording without reproducing this interactive process, which may have contributed to the relatively low motivation ratings for self-monitoring. Future research should therefore examine how BCTs can be effectively combined with interactive and multimodal features of mHealth tools, such as apps with feedback loops and wearable integration, to clarify which BCTs are most suitable for text message-based delivery and which require more dynamic formats. Fifth, the possibility of cluster solutions other than the 7-cluster solution remains. When conducting a latent profile analysis using continuous variables, models with 9 or more clusters failed to converge, preventing a full comparison across solutions. Although the BIC was lowest among the converged solutions for the 8-cluster solution, we were unable to demonstrate the adequacy of the 8-cluster solution. Therefore, the 7-cluster solution was selected based on the BIC and entropy obtained from the LCA. Nevertheless, the identified clusters should be interpreted as a pragmatic summary of response patterns rather than definitive latent subgroups, as more complex solutions may also exist with continuous variables. Further investigation is needed to determine whether the 7-cluster solution can be replicated in other datasets. Sixth, the number of messages per BCT included in the data bank was unbalanced. As the LCA was conducted using dichotomized motivation ratings for each of the 83 individual messages as input variables, BCTs represented by a greater number of messages may have had a greater influence on the cluster patterns. No weighting or adjustment was applied to account for this imbalance, as it reflects the composition of the original message bank. Its potential impact on the clustering results should therefore be considered when interpreting the findings.

Conclusion

Despite the limitations, we believe that our findings will pave the way for data-driven approaches to tailored health message delivery for enhancing PA. Our analysis identified clusters of individuals who were generally open to any BCTs, and these individuals had typically been active and maintained PA habits. The clusters that did not find the messages to be motivational were generally less active and in the preaction stages. However, they appeared to appreciate information about the consequences of health behaviors, with some also showing a preference for rewards, which may serve as a good starting point should TMIs be offered for inactive individuals. The BCTs that have commonly been used in interventions were not necessarily perceived as motivational. Although we found some individual (or between-cluster) differences, self-monitoring was not consistently rated as motivational across all clusters. This does not mean that self-monitoring and other common BCTs are not effective in promoting PA, but that some care would be appropriate to reduce message recipients' resistance when introducing these BCTs. To further increase message acceptability, it is important to consider the mechanisms of action [34,62] to address individual needs, which can be communicated more explicitly in each message. In this study, BCRM was used as one perspective on mechanisms of action to interpret the psychological resources underlying preferences for different messages. The mechanisms of action serve as another tailoring dimension along with the BCT taxonomy, while identifying what proximate outcomes (eg, knowledge and skills) should be targeted for individual message recipients. Furthermore, drawing on classification systems grounded in self-determination theory, such as that proposed by Ahmadi et al [63], may help clarify the links between these outcomes and underlying motivational processes. Future research should expand the focus to how BCTs could be contextualized in interventions (eg, specifying proximate outcomes and resolving how they could be achieved) to make TMIs more personalized, cover the taxonomy, and achieve dynamic and process-oriented tailoring.

Acknowledgments

We used Claude (Anthropic) to assist in revising the manuscript for clarity and logical flow. All content was reviewed and verified by the authors, who take full responsibility for the accuracy and integrity of the manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Funding

This study was supported by an internal grant funded by the National Institute of Advanced Industrial Science and Technology, and no external funding source was involved.

XSL•FO
RenderX

Authors' Contributions

Conceptualization: KT
Methodology: TO, KT
Formal analysis: TO
Data curation: TO, KT
Project administration: K Kimura
Supervision: KT
Writing—original draft: CM
Writing—review and editing: TO, KT, K Katahira, K Kimura
All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[DOCX File , 37 KB - humanfactors_v13i1e91917_app1.docx]

References

1. Orr JA, King RJ. Mobile phone SMS messages can enhance healthy behaviour: a meta-analysis of randomised controlled trials. Health Psychol Rev 2015;9(4):397-416. [doi: 10.1080/17437199.2015.1022847] [Medline: 25739668]
2. Smith DM, Duque L, Huffman JC, Healy BC, Celano CM. Text message interventions for physical activity: a systematic review and meta-analysis. Am J Prev Med 2020;58(1):142-151 [FREE Full text] [doi: 10.1016/j.amepre.2019.08.014] [Medline: 31759805]
3. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of lifestyle-focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. JAMA 2015;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
4. Douglas N, Free C. 'Someone batting in my corner': experiences of smoking-cessation support via text message. Br J Gen Pract 2013;63(616):e768-e776 [FREE Full text] [doi: 10.3399/bjgp13X674459] [Medline: 24267860]
5. Head KJ, Noar SM, Iannarino NT, Grant Harrington N. Efficacy of text messaging-based interventions for health promotion: a meta-analysis. Soc Sci Med 2013;97:41-48. [doi: 10.1016/j.socscimed.2013.08.003] [Medline: 24161087]
6. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
7. Shaw R, Bosworth H. Short message service (SMS) text messaging as an intervention medium for weight loss: a literature review. Health Informatics J 2012;18(4):235-250 [FREE Full text] [doi: 10.1177/1460458212442422] [Medline: 23257055]
8. Buchholz SW, Wilbur J, Ingram D, Fogg L. Physical activity text messaging interventions in adults: a systematic review. Worldviews Evid Based Nurs 2013;10(3):163-173. [doi: 10.1111/wvn.12002] [Medline: 23746267]
9. Dale MT, Wootton SL, Alison J, McNamara R, Leung R, Spencer L, et al. Development and evaluation of text messages designed for people with COPD on the mobile pulmonary rehabilitation (m-PR™) platform. Physiotherapy 2025;127:101455 [FREE Full text] [doi: 10.1016/j.physio.2024.101455] [Medline: 39951834]
10. Shaw RJ, Bosworth HB, Hess JC, Silva SG, Lipkus IM, Davis LL, et al. Development of a theoretically driven mHealth text messaging application for sustaining recent weight loss. JMIR Mhealth Uhealth 2013;1(1):e5 [FREE Full text] [doi: 10.2196/mhealth.2343] [Medline: 25100678]
11. Ardo J, Lee J, Hildebrand JA, Guijarro D, Ghasemazadeh H, Strömberg A, et al. Codesign of a cardiovascular disease prevention text message bank for older adults. Patient Educ Couns 2021;104(11):2772-2784 [FREE Full text] [doi: 10.1016/j.pec.2021.03.036] [Medline: 33863587]
12. Pathak LE, Aguilera A, Williams JJ, Lyles CR, Hernandez-Ramos R, Miramontes J, et al. Developing messaging content for a physical activity smartphone app tailored to low-income patients: user-centered design and crowdsourcing approach. JMIR Mhealth Uhealth 2021;9(5):e21177 [FREE Full text] [doi: 10.2196/21177] [Medline: 34009130]
13. Thakkar J, Karthikeyan G, Purohit G, Thakkar S, Sharma J, Verma S, et al. Development of macaronic Hindi-English 'Hinglish' text message content for a coronary heart disease secondary prevention programme. Heart Asia 2016;8(2):32-38 [FREE Full text] [doi: 10.1136/heartasia-2016-010789] [Medline: 27752288]
14. Yan AF, Stevens P, Wang Y, Weinhardt L, Holt CL, O'Connor C, et al. mHealth text messaging for physical activity promotion in college students: a formative participatory approach. Am J Health Behav 2015;39(3):395-408. [doi: 10.5993/AJHB.39.3.12] [Medline: 25741684]

XSL•FO
RenderX

15. Zamanillo-Campos R, Fiol-deRoque MA, Serrano-Ripoll MJ, Llobera-Canaves J, Taltavull-Aparicio JM, Leiva-Rus A, et al. Impact of an SMS intervention to support type 2 diabetes self-management: DiabeText clinical trial. Br J Gen Pract 2025;75(756):e457-e465 [FREE Full text] [doi: 10.3399/BJGP.2024.0206] [Medline: 39362693]
16. Sahin C, Courtney KL, Naylor PJ, Rhodes RE. Tailored mobile text messaging interventions targeting type 2 diabetes self-management: a systematic review and a meta-analysis. Digit Health 2019;5:2055207619845279 [FREE Full text] [doi: 10.1177/2055207619845279] [Medline: 31041110]
17. Maar MA, Yeates K, Toth Z, Barron M, Boesch L, Hua-Stewart D, et al. Unpacking the black box: a formative research approach to the development of theory-driven, evidence-based, and culturally safe text messages in mobile health interventions. JMIR Mhealth Uhealth 2016;4(1):e10 [FREE Full text] [doi: 10.2196/mhealth.4994] [Medline: 26800712]
18. MacPherson M, Cranston K, Johnston C, Locke S, Jung ME. Evaluation and refinement of a bank of SMS text messages to promote behavior change adherence following a diabetes prevention program: survey study. JMIR Form Res 2021;5(8):e28163 [FREE Full text] [doi: 10.2196/28163] [Medline: 34448713]
19. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
20. Emeran A, Burrows R, Loyson J, Behardien MR, Wiemers L, Lambert E. The effect of text message-based mHealth interventions on physical activity and weight loss: a systematic review and meta-analysis. Am J Lifestyle Med 2026;20(6):952-974. [doi: 10.1177/15598276241268324] [Medline: 39554952]
21. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
22. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
23. MacPherson MM, Cranston KD, Locke SR, Bourne JE, Jung ME. Using the behavior change wheel to develop text messages to promote diet and physical activity adherence following a diabetes prevention program. Transl Behav Med 2021;11(8):1585-1595 [FREE Full text] [doi: 10.1093/tbm/ibab058] [Medline: 34008852]
24. Howlett N, Trivedi D, Troop NA, Chater AM. Are physical activity interventions for healthy inactive adults effective in promoting behavior change and maintenance, and which behavior change techniques are effective? A systematic review and meta-analysis. Transl Behav Med 2019;9(1):147-157 [FREE Full text] [doi: 10.1093/tbm/iby010] [Medline: 29506209]
25. Eckerstorfer LV, Tanzer NK, Vogrincic-Haselbacher C, Kedia G, Brohmer H, Dinslaken I, et al. Key elements of mHealth interventions to successfully increase physical activity: meta-regression. JMIR Mhealth Uhealth 2018;6(11):e10076 [FREE Full text] [doi: 10.2196/10076] [Medline: 30425028]
26. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
27. Van Rhoon L, Byrne M, Morrissey E, Murphy J, McSharry J. A systematic review of the behaviour change techniques and digital features in technology-driven type 2 diabetes prevention interventions. Digit Health 2020;6:2055207620914427 [FREE Full text] [doi: 10.1177/2055207620914427] [Medline: 32269830]
28. Baumann H, Fiedler J, Wunsch K, Woll A, Wollesen B. mHealth interventions to reduce physical inactivity and sedentary behavior in children and adolescents: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2022;10(5):e35920 [FREE Full text] [doi: 10.2196/35920] [Medline: 35544294]
29. Laranjo L, Ding D, Heleno B, Kocaballi B, Quiroz JC, Tong HL, et al. Do smartphone applications and activity trackers increase physical activity in adults? Systematic review, meta-analysis and metaregression. Br J Sports Med 2021;55(8):422-432. [doi: 10.1136/bjsports-2020-102892] [Medline: 33355160]
30. Qin Y, Wang X, Namkoong K. A meta-analysis of the overall effect of mHealth physical activity interventions for weight loss and the moderating effect of behavioral change theories, techniques, and mobile technologies. Mob Media Commun 2021;10(2):337-359. [doi: 10.1177/20501579211054929]
31. Yerrakalva D, Yerrakalva D, Hajna S, Griffin S. Effects of mobile health app interventions on sedentary time, physical activity, and fitness in older adults: systematic review and meta-analysis. J Med Internet Res 2019;21(11):e14343 [FREE Full text] [doi: 10.2196/14343] [Medline: 31778121]
32. Armitage CJ, Conner M, Prestwich A, de Bruin M, Johnston M, Sniehotta F, et al. Investigating which behaviour change techniques work for whom in which contexts delivered by what means: proposal for an International Collaboratory of Centres for Understanding Behaviour Change (CUBiC). Br J Health Psychol 2021;26(1):1-14 [FREE Full text] [doi: 10.1111/bjhp.12479] [Medline: 33080120]
33. Mauch CE, Edney SM, Viana JNM, Gondalia S, Sellak H, Boud SJ, et al. Precision health in behaviour change interventions: a scoping review. Prev Med 2022;163:107192 [FREE Full text] [doi: 10.1016/j.ypmed.2022.107192] [Medline: 35963310]
34. Michaelsen MM, Esch T. Functional mechanisms of health behavior change techniques: a conceptual review. Front Psychol 2022;13:725644 [FREE Full text] [doi: 10.3389/fpsyg.2022.725644] [Medline: 35369223]

XSL•FO
RenderX

35. Michaelsen MM, Esch T. Understanding health behavior change by motivation and reward mechanisms: a review of the literature. Front Behav Neurosci 2023;17:1151918 [FREE Full text] [doi: 10.3389/fnbeh.2023.1151918] [Medline: 37405131]
36. Rothman AJ, Salovey P. Shaping perceptions to motivate healthy behavior: the role of message framing. Psychol Bull 1997;121(1):3-19. [doi: 10.1037/0033-2909.121.1.3] [Medline: 9000890]
37. Latimer AE, Brawley LR, Bassett RL. A systematic review of three approaches for constructing physical activity messages: what messages work and what improvements are needed? Int J Behav Nutr Phys Act 2010;7:36 [FREE Full text] [doi: 10.1186/1479-5868-7-36] [Medline: 20459779]
38. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. Science 1981;211(4481):453-458. [doi: 10.1126/science.7455683] [Medline: 7455683]
39. Jiang Y, Zhao Y, Xu Q, Tang P, Wang X, Guo Y, et al. The effects of gain-loss framed message on physical activity attitudes, intentions, and behaviors in physically inactive adults: a systematic review and meta-analysis. Front Public Health 2026;14:1782478 [FREE Full text] [doi: 10.3389/fpubh.2026.1782478] [Medline: 41919308]
40. Oba T, Takano K, Katahira K, Kimura K. Exploring individual, social and environmental factors related to physical activity: a network analysis. BMJ Open Sport Exerc Med 2024;10(3):e001983 [FREE Full text] [doi: 10.1136/bmjsem-2024-001983] [Medline: 39104375]
41. Oba T, Takano K, Katahira K, Kimura K. Use patterns of smartphone apps and wearable devices supporting physical activity and exercise: large-scale cross-sectional survey. JMIR Mhealth Uhealth 2023;11:e49148 [FREE Full text] [doi: 10.2196/49148] [Medline: 37997790]
42. Konishi N, Oba T, Takano K, Katahira K, Kimura K. Functions of smartphone apps and wearable devices promoting physical activity: six-month longitudinal study on Japanese-speaking adults. JMIR Mhealth Uhealth 2024;12:e59708 [FREE Full text] [doi: 10.2196/59708] [Medline: 39658011]
43. Oba T, Takano K, Katahira K, Kimura K. Revisiting the transtheoretical model for physical activity: a large-scale cross-sectional study on Japanese-speaking adults. Ann Behav Med 2024;58(3):167-178. [doi: 10.1093/abm/kaad069] [Medline: 38166169]
44. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007;370(9596):1453-1457 [FREE Full text] [doi: 10.1016/S0140-6736(07)61602-X] [Medline: 18064739]
45. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International Physical Activity Questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
46. Murase N, Katsumura T, Ueda C, Inoue S, Shimomitsu T. Validity and reliability of Japanese version of International Physical Activity Questionnaire. J Health Welf Stat 2002;49(11):1-9 [FREE Full text]
47. Oka K. Stages of change for exercise behavior and self-efficacy for exercise among middle-aged adults. Nihon Koshu Eisei Zasshi 2003;50(3):208-215. [doi: 10.11236/jph.50.3_208]
48. Oka K, Hirai K, Tsutsumi T. Psychological factors associated with physical inactivity among middle-aged adults: decisional balance for exercise. Jpn J Behav Med 2003;9(1):23-30. [doi: 10.11331/jjbm.9.23]
49. Marcus BH, Simkin LR. The stages of exercise behavior. J Sports Med Phys Fitness 1993;33(1):83-88. [Medline: 8350613]
50. Reed GR, Velicer WF, Prochaska JO, Rossi JS, Marcus BH. What makes a good staging algorithm: examples from regular exercise. Am J Health Promot 1997;12(1):57-66. [doi: 10.4278/0890-1171-12.1.57] [Medline: 10170436]
51. Sinha P, Calfee CS, Delucchi KL. Practitioner's guide to latent class analysis: methodological considerations and common pitfalls. Crit Care Med 2021;49(1):e63-e79 [FREE Full text] [doi: 10.1097/CCM.0000000000004710] [Medline: 33165028]
52. Karnowski V. Latent class analysis. In: The International Encyclopedia of Communication Research Methods. New Jersey: John Wiley & Sons; 2017:1-10.
53. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
54. Salovey P, Schneider TR, Apanovitch AM. Message framing in the prevention and early detection of illness. In: The Persuasion Handbook: Developments in Theory and Practice. Thousand Oaks: Sage Publications; 2002:391-406.
55. Asbjørnsen RA, Wentzel J, Smedsrød ML, Hjelmesæth J, Clark MM, Solberg Nes L, et al. Identifying persuasive design principles and behavior change techniques supporting end user values and needs in eHealth interventions for long-term weight loss maintenance: qualitative study. J Med Internet Res 2020;22(11):e22598 [FREE Full text] [doi: 10.2196/22598] [Medline: 33252347]
56. Asbjørnsen RA, Smedsrød ML, Solberg Nes L, Wentzel J, Varsi C, Hjelmesæth J, et al. Persuasive system design principles and behavior change techniques to stimulate motivation and adherence in electronic health interventions to support weight loss maintenance: scoping review. J Med Internet Res 2019;21(6):e14265 [FREE Full text] [doi: 10.2196/14265] [Medline: 31228174]
57. Procter S, Mutrie N, Davis A, Audrey S. Views and experiences of behaviour change techniques to encourage walking to work: a qualitative study. BMC Public Health 2014;14:868 [FREE Full text] [doi: 10.1186/1471-2458-14-868] [Medline: 25150004]

XSL•FO
RenderX

58. Compernolle S, DeSmet A, Poppe L, Crombez G, De Bourdeaudhuij I, Cardon G, et al. Effectiveness of interventions using self-monitoring to reduce sedentary behavior in adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2019;16(1):63 [FREE Full text] [doi: 10.1186/s12966-019-0824-3] [Medline: 31409357]
59. Fu E, Neubert SW, Chang A, Smith JD, Graham AK. Characterizing behavior change techniques used in the self-management of binge eating and weight: applying a user-centered design approach. Eat Behav 2022;44:101591 [FREE Full text] [doi: 10.1016/j.eatbeh.2021.101591] [Medline: 34920208]
60. Lu X, Yao Y, Jin Y. Digital exclusion and functional dependence in older people: findings from five longitudinal cohort studies. EClinicalMedicine 2022;54:101708 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101708] [Medline: 36353265]
61. Rhodes RE, de Bruijn GJ. How big is the physical activity intention-behaviour gap? A meta-analysis using the action control framework. Br J Health Psychol 2013;18(2):296-309. [doi: 10.1111/bjhp.12032] [Medline: 23480428]
62. Carey RN, Connell LE, Johnston M, Rothman AJ, de Bruin M, Kelly MP, et al. Behavior change techniques and their mechanisms of action: a synthesis of links described in published intervention literature. Ann Behav Med 2019;53(8):693-707 [FREE Full text] [doi: 10.1093/abm/kay078] [Medline: 30304386]
63. Ahmadi A, Noetel M, Parker P, Ryan RM, Ntoumanis N, Reeve J, et al. A classification system for teachers' motivational behaviors recommended in self-determination theory interventions. J Educ Psychol 2023;115(8):1158-1176. [doi: 10.1037/edu0000783]

Abbreviations

BCRM: Behavior Change Resource Model
BCT: behavior change technique
BCW: behavior change wheel
BIC: Bayesian Information Criterion
COM-B: capability, opportunity, motivation-behavior
IPAQ-SF: International Physical Activity Questionnaire-Short Form
LCA: latent class analysis
MET: metabolic equivalent
mHealth: mobile health
PA: physical activity
SoC: stages of change
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
TMI: text message intervention

Edited by A Choudhury; submitted 22.Jan.2026; peer-reviewed by N Bortoletto, H Tilga; comments to author 29.Apr.2026; accepted 09.Jul.2026; published 19.Aug.2026.

Please cite as:
Moriishi C, Oba T, Takano K, Katahira K, Kimura K
Exploring Response Patterns to Motivational Messages Supporting Physical Activity: Latent Class Analysis
JMIR Hum Factors 2026;13:e91917
URL: https://humanfactors.jmir.org/2026/1/e91917
doi:10.2196/91917
PMID:

©Chihiro Moriishi, Takeyuki Oba, Keisuke Takano, Kentaro Katahira, Kenta Kimura. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 19.Aug.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Exploratory Analysis of the Digital Health Gender Paradox in Consumer-Facing Digital Health Technology Adoption: Cross-Sectional Survey Study

Jessica Kathleen Wallace[1], MDH; Mehak Batra[1], PhD; Tafheem Ahmad Wani[1], PhD; Michael Liem[1], PhD; Kylie Ovenden[1], MBA; James Boyd[1], PhD; Therese Keane[2], DEd; Urooj Raza Khan[1], PhD

[1]School of Psychology and Public Health, La Trobe University, Bundoora, Victoria, Australia

[2]School of Education, La Trobe University, Bundoora, Victoria, Australia

Corresponding Author:
Tafheem Ahmad Wani, PhD
School of Psychology and Public Health
La Trobe University
Kingsbury Drive, Health sciences 2, La Trobe university
Bundoora, 3086
Australia
Phone: 61 451906170
Email: T.Wani@latrobe.edu.au

Abstract

Background: Digital health technologies have become integral to modern health care, yet consumer adoption remains uneven. These technologies span a range of consumer-facing tools, including remote care services, mobile apps, wearable devices, conversational agents, and digital medication services, with uptake varying considerably across tool types and by gender. An earlier analysis of this dataset revealed a gender paradox: women reported higher digital literacy compared to men yet demonstrated lower adoption and curiosity toward digital health tools. The Capability, Opportunity, Motivation – Behavior (COM-B) model proposes that behavior occurs when individuals have the capability (skills and confidence), opportunity (access, affordability, and contextual support), and motivation (trust, perceived usefulness, and willingness) to act.

Objective: This study aimed to examine associations between digital literacy, self-reported COM-B domains, gender, and adoption across digital health technologies and to explore whether these associations differ by gender.

Methods: This cross-sectional quantitative study with an integrated qualitative component enrolled 416 adults and used multivariable logistic regression to examine associations between COM-B domains and adoption across five consumer-facing digital health technologies. Exploratory gender-by-predictor interaction and gender-stratified analyses examined potential gender-specific patterns, while self-reported reasons for nonuse were mapped to COM-B domains using directed content analysis.

Results: Women had lower odds of chatbot/virtual assistant use (adjusted odds ratio [aOR] 0.32, 95% CI 0.18-0.59), marginally lower odds of e-pharmacy use (aOR 0.48, 95% CI 0.22-1.01), and higher odds of mobile health (mHealth) app use (aOR 2.59, 95% CI 1.28-5.25). Of 20 exploratory interaction tests, chatbot digital literacy-by-gender (P=.021) and capability-by-gender (P=.044) interactions were nominally significant. In directed barrier analysis, the gender distribution differed significantly for mHealth apps (P=.009), whereas the wearable result showed a trend (P=.079); motivation-related barriers were most frequent for most technologies.

Conclusions: Behavioral determinants of digital health technology adoption varied by technology type. Women used several technologies less often than men despite reporting higher digital literacy, capability, and motivation, and opportunity emerged as the most consistent factor across the models examined. Exploratory analyses suggested these patterns may differ by gender for selected technologies but did not support a single gender-differentiated pathway. These findings are exploratory and require confirmation in larger studies.

(JMIR Hum Factors 2026;13:e101963) doi:10.2196/101963

XSL•FO
RenderX

KEYWORDS

digital health; gender differences; health equity; digital health literacy; technology adoption; Capability, Opportunity, Motivation – Behavior; COM-B model; mobile health; mHealth; telehealth; artificial intelligence; AI; chatbots

Introduction

Background

Health systems worldwide are under escalating pressure from aging populations, rising chronic disease prevalence, and persistent workforce shortages [1,2]. In response, many countries are accelerating digital transformation to support more sustainable, person-centered models of care. As digital health technologies become embedded in routine health care delivery, attention has increasingly shifted to digital health readiness, including the data, technology, analytic, workforce, and human factors required to realize the benefits of digital health, while minimizing harms [2].

Internationally, a broad range of digital health technologies, such as telehealth, mobile health (mHealth) apps, wearable devices, conversational agents, and e-pharmacies, are being integrated into health care systems. However, adoption remains uneven and is shaped by stakeholder needs, implementation context, trust, infrastructure, and alignment with health care workflows [3,4]. Evidence from Australia reflects a similar pattern: although telehealth use surged during the COVID-19 pandemic, adoption of newer or more autonomous tools has been slower, particularly among older adults, people with lower educational attainment or income, those living in rural or remote areas, and individuals with limited digital or health literacy [5,6]. Accordingly, Australian digital health policy identifies inclusive, person-centered engagement as central to realizing the benefits of digital transformation across the population [7].

Prior Digital Health Technology Adoption/Implementation Research

Research examining digital health technology adoption has expanded rapidly in recent years; however, important limitations remain. Much of the existing literature focuses on single technologies, applies narrow or technology-specific behavioral models, or targets specific subpopulations, such as older adults or individuals with chronic disease [8-10]. In addition, although behavioral determinants are increasingly recognized as important to digital health technology adoption and implementation, they are not always examined using structured behavioral frameworks, and limited attention has been given to how these determinants vary across gender groups or across different types of digital health technologies [8,10,11].

Australian research shows that digital health technology adoption is strongly patterned by socioeconomic status, education, broadband availability, and digital health literacy, which together predict engagement, particularly with national personal health records [12,13]. Consistent with this, implementation reviews continue to identify persistent adoption gaps in rural and underserved populations and note the limited use of behavioral theory in digital health technology adoption research [14]. More broadly, this literature often emphasizes individual digital literacy, while giving less attention to the structural and contextual conditions that shape engagement with digital technologies [4]. The World Health Organization (WHO) Global Strategy on Digital Health similarly notes that literacy alone is insufficient for equitable participation without inclusive design, supportive infrastructure, and equitable access [15].

Together, these gaps mean the evidence base provides an incomplete account of why adoption varies across population groups. In particular, gender is often reported descriptively rather than examined as a factor that may shape how individuals experience, interpret, and engage with digital health technologies. This study analysed behavioral drivers across multiple digital health technologies and examined gender differences in adoption using a structured behavioral framework.

Gendered Patterns of Digital Engagement

A preliminary conference paper using this dataset reported high-level descriptive findings on digital literacy, the current use of digital health technologies, and curiosity toward emerging technologies. The analysis observed that women report higher digital literacy but lower use of some digital health technologies; however, it did not formally examine the behavioral mechanisms underlying these differences using multivariable or Capability, Opportunity, Motivation – Behavior (COM-B)–based analyses [16]. Digital literacy refers to the skills and confidence required to access, understand, and use digital technologies effectively [13,17]. Curiosity, defined as an intrinsic motivation to seek information and explore technologies, is conceptually distinct from digital literacy or confidence [18,19]. This study applied the COM-B framework to examine behavioral determinants of adoption through multivariable regression, gender-by-predictor interaction analyses, gender-stratified models, exploratory nested modeling, and COM-B-guided analysis of reported reasons for nonuse. In this way, the current study moved beyond describing gender differences in literacy and technology use to investigate whether capability, opportunity, and motivation are associated with the observed gender-specific patterns of adoption.

Global analyses highlight persistent gender inequalities in digital participation. Even where women attain high levels of education and digital skills, engagement with digital technologies remains constrained by affordability, safety concerns, time poverty, and structural inequalities in science, technology, engineering, and mathematics (STEM) education [20-23]. Time poverty, limited discretionary time arising from disproportionate domestic and caregiving responsibilities, further restricts opportunities for digital engagement [24].

Consistent with these broader patterns, Australian and international studies have documented a gender gap in digital health engagement. Women, particularly in their roles as patients and carers, may be more likely to experience inequitable or suboptimal digital health interactions [25,26]. Scholars also caution that digital health tools may insufficiently reflect women's needs and that design choices, data governance practices, or algorithms can embed gender bias, while limited integration of gender-responsive and intersectional perspectives

XSL•FO
RenderX

in digital health research may reinforce existing inequalities [25,27].

These dynamics are particularly significant, given women's central roles within health systems. In Australia, women make up a large proportion of the formal health workforce, and primary informal carers are also predominantly female [28,29]. Qualitative evidence from Australia also shows that many women use digital health information, not only as patients, but also to provide advice and support to family members and friends [30]. Consequently, women's engagement with digital health technologies is consequential in its own right and may shape broader patterns of adoption, trust, and diffusion within families and communities [23,30].

Overall, this evidence points to a potential capability-opportunity mismatch in digital health technology adoption: women's digital skills and familiarity with health care settings may coexist with structural, temporal, and contextual barriers that constrain opportunities to use digital technologies [24,25,31]. Addressing these dynamics is essential to ensure that digital transformation does not widen existing inequalities and to support the design and development of technologies that are responsive to the needs, preferences, and contexts of diverse populations [25,32].

Theoretical Framework: COM-B and the Behavior Change Wheel

To examine these dynamics, this study applied the COM-B model, which conceptualizes behavior as the product of interaction between capability, opportunity, and motivation [33]. Capability refers to the psychological and physical capacity to perform a behavior (including relevant knowledge and skills). Opportunity captures external factors, both physical and social, that enable or prompt the behavior. Motivation encompasses reflective and automatic processes, including goals, beliefs, emotions, and habits, that energize and direct behavior [33].

COM-B forms the hub of the Behavior Change Wheel (BCW), a widely used framework for diagnosing behavioral barriers and mapping them to potential intervention strategies and policy levers [33]. Relative to technology adoption models, such as the Technology Acceptance Model and the Unified Theory of Acceptance and Use of Technology, which emphasize constructs such as perceived usefulness, ease of use, performance expectancy, and facilitating conditions [34,35], COM-B provides a broader behavioral lens by incorporating psychological, social, and environmental determinants of behavior. In this study, however, the COM-B constructs were operationalized through self-reported composite indices and should be interpreted as perceived behavioral domains rather than objective measures of physical or psychological capability.

Study Aim

This study aimed to examine gender differences in digital health technology adoption and to explore how self-reported behavioral determinants within the COM-B model are associated with those patterns. The analyses were exploratory and association based; they were not designed to establish causal pathways or mediation.

Using a survey dataset, this study examined technology-specific associations with capability, opportunity, and motivation and evaluated exploratory gender-by-predictor interactions. It addressed gaps in multitechnology behavioral research, while explicitly distinguishing overall pooled predictors from evidence of gender moderation. To translate behavioral insights into practical implications, findings were mapped to the BCW, which linked COM-B determinants to intervention functions and policy strategies that may support more equitable digital health technology adoption. As this was a self-reported study, results reflect perceived rather than objectively measured behaviors; nonetheless, they provide theory-informed insights into gendered determinants of digital health engagement that can inform equitable digital health policy, design, and implementation.

Methods

Overview

A cross-sectional online survey (Research Electronic Data Capture [REDCap]) was conducted to examine Australian adults' digital literacy, engagement with digital health technologies, and preferences for future services. Eligible participants were aged ≥18 years and residing in Australia. As an exploratory pilot examining behavioral pathways and gender differences, no a priori sample size calculation was performed. The achieved sample (N=416) supported multivariable and interaction analyses but may have been underpowered to detect small effects.

The survey was developed collaboratively by academic investigators, Medibank Private Limited, and consumer organizations to ensure relevance and accessibility. It included demographic items and open- and closed-ended questions on digital health experiences and behaviors.

Participants were recruited using convenience sampling through social media, consumer advocacy networks, Medibank communication channels, and research team networks. The sample was not designed to be nationally representative. Data were collected between July 21 and August 24, 2023. Participants could enter a prize draw for an AU $50 (US $36) voucher funded by Medibank Private Limited.

Survey design was informed by the COM-B model and the Australian Digital Capability Framework, with items adapted to the digital health context (Multimedia Appendix 1) [36]. A review by the Consumer Health Forum of Australia and state-based organizations was conducted to optimize accessibility.

Measurement

Predictors

Digital literacy was measured using a 22-item scale covering information/data literacy, communication/collaboration, content creation, safety, and problem solving. Items were averaged, with higher scores indicating greater self-reported digital competence (α=0.81).

Capability, opportunity, and motivation were operationalized as theory-informed composite indices based on

XSL•FO
RenderX

COM-B-informed survey sections. Capability reflected perceived functional readiness, knowledge, skills, and confidence; opportunity reflected perceived access, affordability, availability, infrastructure, and social/professional support; and motivation reflected trust, perceived usefulness and benefits, preferences, interest, intentions, and emotional responses to digital technology. Digital literacy was treated as a related but distinct measure of self-reported digital competence. All measures were self-reported and therefore reflected perceived rather than objectively assessed capability or digital skills.

The Capability index comprised 15 items (α=0.83), the Opportunity index 5 items (α=0.76), and the Motivation index 13 items (α=0.61). These indices were developed for this exploratory study and were not intended as validated psychometric COM-B scales. Items were coded so that higher scores represented greater capability, opportunity, or motivation, with negatively worded items reverse-coded, as appropriate. Item scores were averaged within each domain and standardized (*z*-scores), such that odds ratios represented a 1-SD increase. Exploratory factor analyses indicated that the Opportunity index showed the clearest evidence of unidimensionality, consistent with its internal consistency estimate. The Capability and Motivation indices showed evidence of multidimensionality, suggesting they each capture related but heterogeneous constructs within their respective domains; this is reflected in the lower internal consistency of the Motivation index in particular and should be considered when interpreting findings associated with these composites. Intercorrelations between digital literacy and the three COM-B indices were modest to moderate (r=0.39-0.57), providing preliminary evidence that the constructs were related but distinct.

Complete item-to-domain and subdomain mapping, scoring rules, reverse-coded items, handling of "do not know" responses, item-level missingness, and exploratory measurement analyses are provided in Tables S2-S4 in Multimedia Appendix 2.

Outcomes (Behavior)

Adoption of each technology (telehealth, mHealth apps, wearable devices, chatbots/virtual assistants, and e-pharmacy) was operationalized as self-reported current use and coded as a binary outcome (current user vs nonuser). This measure captured current user status rather than the frequency, duration, or intensity of technology use.

Effect Modifier

Gender (female vs male) was examined as an effect modifier. Interaction terms between gender and each of the four predictors (digital literacy, capability, opportunity, and motivation) were included in separate exploratory models (Table S1 in Multimedia Appendix 2), resulting in 20 interaction tests across five technology outcomes. These analyses were hypothesis generating rather than confirmatory.

Covariates

Participants were grouped by age (≤34, 35-54, and ≥55 years) to reduce sparse cells and maintain adequate group sizes. The exact age in years was not collected; therefore, age could not be modeled as a continuous variable. Models adjusted for age group, education (low/intermediate, bachelor/advanced diploma, postgraduate), occupation (white collar, blue collar/service, other/not in labor force), region (Eastern, Central/Western, Southern/Territory), and residence (metropolitan vs rural/remote). Occupation and region were collapsed from the original survey categories to maintain model parsimony. These groupings were defined during data preparation, and covariates were selected a priori as structural determinants of digital access.

Statistical Analysis

Overview

Quantitative responses, including Likert-scale items, were summarized using descriptive statistics. Qualitative free-text responses were analyzed using directed content analysis guided by the COM-B framework. Cross-tabulations by key demographic variables, including age, gender, education level, and geographic location, were performed using IBM SPSS Statistics, whereas all inferential and multivariable analyses were conducted in Stata version 19.5 [37].

Quantitative Analysis

Participant characteristics were summarized by gender. Categorical variables were compared using χ^2 tests. Continuous or ordinal variables were summarized as median (IQR) and compared using Mann-Whitney *U* tests. Differences in adoption rates by gender were compared using χ^2 tests.

Logistic regression models were used to examine associations between digital health technology adoption and gender, digital literacy, COM-B domains, and sociodemographic covariates. The primary fully adjusted models included gender, digital literacy, capability, opportunity, motivation, age, education, occupation, region, and rural/remote residence. Adjusted odds ratios (aORs) with 95% CIs were reported. Digital literacy, capability, opportunity, and motivation were standardized (*z*-scores), such that their aORs represented a 1-SD increase.

To descriptively examine changes in the gender coefficient after inclusion of the COM-B domains, pre-COM-B-reduced models included gender, digital literacy, and the same sociodemographic covariates as the primary models, but excluded capability, opportunity, and motivation. Covariate definitions and coding were identical between these models and the primary fully adjusted models. These comparisons were descriptive and were not interpreted as evidence of mediation or mechanisms.

Effect moderation by gender was explored by including gender-by-predictor interaction terms for digital literacy, capability, opportunity, and motivation across each of the five technology outcomes, resulting in 20 interaction tests. These analyses were exploratory and hypothesis generating. For nominally significant interactions, adjusted predicted probabilities of technology adoption were estimated at −1 SD, the mean, and +1 SD of the relevant predictor, holding other covariates at their observed values using Stata's *margins* command. Gender-stratified logistic regression models were additionally estimated to characterize within-gender patterns (Table S1 in Multimedia Appendix 2).

Multicollinearity was assessed using variance inflation factors (<2). Model fit was evaluated using likelihood-ratio χ^2 and

XSL•FO
RenderX

pseudo-R^2 statistics. Robust (Huber-White) SEs were used for the primary models. There were no missing data for variables included in the primary multivariable models; consequently, all models included the full analytic sample (N=416).

Given the relatively small number of nonevents for some outcomes, model complexity and robustness were additionally assessed. For each primary model, we examined the number of complete cases, events/nonevents, and estimated parameters. Simplified sensitivity models retained gender, digital literacy, capability, opportunity, motivation, and age, while excluding the remaining sociodemographic covariates. These were fitted for both the primary models and the 20 gender-by-predictor interaction models and compared with the primary fully adjusted models to assess sensitivity to model complexity. A further sensitivity analysis excluding participants in the "other/not in labor force" occupation category was directionally consistent with the primary models (results not shown).

Given the lower internal consistency of the Motivation index (α=0.61), additional sensitivity analyses examined alternative motivation specifications, including exclusion of the e-pharmacy-specific preference item and a reduced composite excluding the three items with the lowest corrected item-rest correlations (Table S5 in Multimedia Appendix 2). Given the exploratory nature of this pilot study and multiple behavioral predictors across technology-specific models, estimates were interpreted cautiously, with emphasis on the magnitude, direction, consistency, and uncertainty of associations.

Qualitative Analysis

For participants who had not adopted specific technologies, reasons for nonuse were captured through open-ended responses and analyzed using directed content analysis guided by the COM-B framework. Responses were imported into Stata version 19.5 and preprocessed prior to coding. Initial categorization was conducted using lexical pattern matching (*regexm* functions) to flag responses corresponding to capability, opportunity, and motivation domains (eg, "no access," "prefer in-person," "not confident"). Lexical matching was used as an initial screening and organizational step rather than as the basis for final classification. The technology-specific coding framework, operational definitions, and illustrative lexical triggers are provided in Table S6 in Multimedia Appendix 2.

Two PhD-qualified researchers independently reviewed all flagged responses against the complete response text and validated or revised the initial classifications based on contextual meaning. Where responses reflected more than one COM-B construct, a primary barrier was assigned according to the predominant reason for nonuse identified from the complete response. Discrepancies between researchers were resolved through iterative discussion and consensus. Because the two researchers' independent judgments were not retained as separate ratings, an interrater agreement statistic was not calculated. Coding rigor was instead supported by the technology-specific framework with documented operational definitions (Table S6 in Multimedia Appendix 2). Coding was blinded to participant gender; gender was linked to the final COM-B classifications only after coding for the gender-stratified descriptive analyses.

Within-gender frequencies of capability, opportunity, and motivation barriers were summarized for each technology and integrated descriptively with the quantitative findings. Because open-ended responses were obtained from a predefined survey sample rather than through iterative qualitative sampling, saturation was not used as a stopping criterion. Given small cell counts in the gender-stratified barrier analyses, Fisher's exact test was used for comparisons.

Ethical Considerations

Ethics approval was obtained from the La Trobe University Low-Risk Human Research Ethics Committee (HREC #23100). Participation was voluntary, with informed consent obtained prior to survey completion. Data were collected anonymously using REDCap, a secure, web-based platform designed to support data collection and management for research. De-identified data were exported and stored securely on a password-protected research drive, with access restricted to the research team.

Results

Participant Profile

Of 416 respondents, 61.1% (n=254) were male and 38.9% (n=162) female. The age distribution differed by gender (P=.019), with women more often aged ≥35 years (n=84, 51.8%) and more likely to hold postgraduate qualifications (n=73, 45.1%) compared to men (P=.001). Occupational distribution differed significantly between genders, with approximately three-quarters (n=123, 75.9%) of women working in white-collar roles compared to 42.9% (n=109) of men (P<0.001), whereas men were more commonly employed in blue-collar or service occupations. Residence area was broadly similar between genders, although regional distribution differed significantly.

Across COM-B domains, women reported higher digital literacy, capability, and motivation, whereas opportunity scores were similar between genders (Table 1). These differences supported examining gender-specific pathways in subsequent analyses.

XSL•FO
RenderX

Table 1. Participant characteristics by gender (N=416).

Characteristics	Male (n=254)	Female (n=162)	*P* value
Age group (years), n (%)			.019[a]
≤34	158 (62.2)	78 (48.2)	—[b]
35-54	68 (26.8)	59 (36.4)	—
≥55	28 (11.0)	25 (15.4)	—
Education level, n (%)			<.001[a]
Low/intermediate	26 (10.2)	19 (11.7)	—
Bachelor/advanced diploma	155 (61.0)	70 (43.2)	—
Postgraduate	73 (28.7)	73 (45.1)	—
Occupation category, n (%)			<.001[a]
White collar	109 (42.9)	123 (75.9)	—
Blue collar/service	143 (56.3)	23 (14.2)	—
Other[c]	2 (0.8)	16 (9.9)	—
Residence area, n (%)			.261[a]
Metropolitan	180 (70.9)	123 (75.9)	—
Rural/remote	74 (29.1)	39 (24.1)	—
Region group, n (%)			<.001[a]
Eastern	202 (79.5)	100 (61.7)	—
Central/western	44 (17.3)	46 (28.4)	—
Southern/territory	8 (3.2)	16 (9.9)	—
Digital literacy, median (IQR); minimum, maximum	3.2 (1.1); 2.1, 4.0	3.6 (0.5); 1.9, 4.0	<.001[d]
Capability index, median (IQR); minimum, maximum	3.0 (0.7); 2.1, 3.8	3.3 (0.5); 2, 3.7	<.001[d]
Opportunity index, median (IQR); minimum, maximum	3.4 (0.6); 1.6, 4.0	3.4 (1.0); 1, 4.0	.372[d]
Motivation index, median (IQR); minimum, maximum	3.4 (0.3); 2.8, 3.8	3.5 (0.2); 2.6, 3.8	<.001[d]

[a]χ^2 test.

[b]Not applicable.

[c]Not in the labor force.

[d]Mann-Whitney *U* test (statistically significant results at $P<.05$).

The Adoption Paradox: Technology-Specific Gender Gaps

Overall, nearly all participants (n=413, 99.3%) reported using at least one digital health technology. Adoption was highest for clinician-mediated modalities, including telehealth and mobile apps.

Gender differences were observed for newer or more autonomous technologies. Women were less likely to use wearable devices, chatbots/virtual assistants, and e-pharmacies (Table 2). After full adjustment, women remained significantly less likely to use chatbots/virtual assistants, whereas the association with e-pharmacy use was marginal and the association with wearable device use was attenuated (Table 3). These findings indicate that gender differences are technology-specific, particularly for tools requiring greater autonomy or trust.

XSL•FO
RenderX

Table 2. Uptake of digital health technologies by gender.

Technology and use	Total participants (N=416), n (%)	Male (n=254), n (%)	Female (n=162), n (%)
Telehealth (*P*=.059[a])[b]			
Yes	385 (92.5)	240 (94.5)	145 (89.5)
No	31 (7.5)	14 (5.5)	17 (10.5)
mHealth[c] app (*P*=.44)[b]			
Yes	323 (77.6)	194 (76.4)	129 (79.6)
No	93 (22.4)	60 (23.6)	33 (20.4)
Wearable device (*P*<.001[d])[b]			
Yes	329 (79.1)	225 (88.6)	104 (64.2)
No	87 (20.9)	29 (11.4)	58 (35.8)
Chatbot/virtual assistant (*P*<.001[d])[b]			
Yes	308 (74.0)	222 (87.4)	86 (53.1)
No	108 (26.0)	32 (12.6)	76 (46.9)
e-Pharmacy (*P*<.001[d])[b]			
Yes	357 (85.8)	235 (92.5)	122 (75.3)
No	59 (14.2)	19 (7.5)	40 (24.7)

[a]*P*=.05-.10 (trend).

[b]χ^2 test.

[c]mHealth: mobile health.

[d]Statistically significant results at *P*<0.05.

Table 3. Multivariable logistic regression models[a] of predictors of digital health technology adoption by gender and COM-B[b] domain.

Predictor[c]	Telehealth, aOR[d] (95% CI), *P* value	mHealth[e] app, aOR (95% CI), *P* value	Wearable device, aOR (95% CI), *P* value	Chatbot/virtual assistant, aOR (95% CI), *P* value	e-Pharmacy, aOR (95% CI), *P* value
Gender (1=female)	1.51 (0.58-3.95), .403	2.59 (1.28-5.25), .008[f]	0.61 (0.33-1.12), .113	0.32 (0.18-0.59), <.001[f]	0.48 (0.22-1.01), .052[g]
Digital literacy (*z*-score)	0.47 (0.24-0.92), .026[f]	0.74 (0.42-1.29), .288	0.70 (0.40-1.21), .199	0.88 (0.54-1.44), .613	0.84 (0.50-1.38), .487
Capability index (*z*-score)	0.88 (0.44-1.73), .703	1.44 (0.81-2.58), .215	0.77 (0.47-1.26), .315	0.56 (0.35-0.89), .016[f]	0.70 (0.43-1.15), .158
Opportunity index (*z*-score)	1.80 (1.15-2.83), .010[f]	3.18 (2.11-4.77), <.001[f]	1.93 (1.33-2.79), <.001[f]	1.16 (0.84-1.59), .380	1.26 (0.89-1.76), .184
Motivation index (*z*-score)	1.50 (0.92-2.45), .101	0.93 (0.70-1.25), .627	1.02 (0.72-1.45), .914	1.26 (0.93-1.70), .135	1.50 (1.05-2.15), .026[f]

[a]Each model was adjusted for age, education, occupation, region, and rural/remote residence.

[b]COM-B: Capability, Opportunity, Motivation – Behavior.

[c]Predictors were standardized (*z*-scores), except gender, so each odds ratio represents the change in odds for a 1-SD increase in the predictor. Gender was entered as a binary categorical variable and was not standardized.

[d]aOR: adjusted odds ratio.

[e]mHealth: mobile health

[f]Statistically significant results at *P*<.05.

[g]*P*=.05-.10 (trend).

Behavioral Determinants of Digital Health Technology Adoption (COM-B Domains and Gender)

Before inclusion of the COM-B domains, pre-COM-B-reduced models, including gender, digital literacy, and sociodemographic covariates, showed significantly lower odds of wearable-device (aOR 0.47, 95% CI 0.26-0.85; *P*=.012), chatbot/virtual assistant (aOR 0.30, 95% CI 0.16-0.54; *P*<.001), and e-pharmacy (aOR 0.41, 95% CI 0.20-0.86; *P*=.018) use among women. Gender was not associated with telehealth (aOR 1.01, 95% CI 0.42-2.44;

XSL•FO
RenderX

P=.981) or mHealth app (aOR 1.36, 95% CI 0.72-2.56; *P*=.342) use. After inclusion of capability, opportunity, and motivation in the primary fully adjusted models (Table 3), the magnitude of these gender associations changed to varying degrees.

Opportunity emerged as the most consistent and influential determinant of digital health technology adoption across technologies. A 1-SD increase in opportunity was associated with higher odds of telehealth, mHealth app, and wearable device use, highlighting the central role of access, affordability, availability, and contextual support in enabling engagement.

Motivation showed a nominally significant positive association with e-pharmacy adoption in the primary model; however, this association should be interpreted cautiously, as it was attenuated and the CI included the null when model complexity was reduced in sensitivity analyses. This pattern may potentially reflect the role of trust and perceived safety in medication-related digital transactions but requires confirmation in larger studies. In contrast, digital literacy showed an inverse association with telehealth use and was not independently associated with other technologies after adjustment, suggesting that technical competence alone is insufficient to drive adoption.

Capability was negatively associated with chatbot/virtual assistant use, indicating that individuals with higher perceived capability were less likely to engage with conversational agents. This may reflect a preference among more capable users for alternative sources of information or care.

After full adjustment, women remained significantly less likely to use chatbots and marginally less likely to use e-pharmacies, while being 2.6 times more likely to use mHealth apps. Collectively, these results indicate that structural and contextual opportunity is the most consistent determinant of engagement across digital health technologies, with patterns of adoption varying by technology type and gender (Table 3).

The primary pooled models examined overall associations between behavioral determinants and technology adoption across the full sample, adjusted for gender and sociodemographic covariates. These overall associations should be distinguished from the exploratory gender moderation analyses presented in the *Gender Moderation and Stratified Behavioral Pathways* section. All primary models included 416 complete cases and 14 predictor parameters. The numbers of events/nonevents were 385/31 for telehealth, 323/93 for mHealth app use, 329/87 for wearable device use, 308/108 for chatbot use, and 357/59 for e-pharmacy use. Simplified-model sensitivity analyses showed that the principal COM-B findings were generally directionally consistent, particularly the opportunity associations with telehealth, mHealth app, and wearable-device use and the capability association with chatbot use. For e-pharmacy use, the motivation association remained positive but was attenuated in the simplified model (aOR 1.35, 95% CI 0.98-1.87; *P*=.070) compared to the primary model (aOR 1.50, 95% CI 1.05-2.15; *P*=.026).

Gender Moderation and Stratified Behavioral Pathways

Exploratory interaction analyses suggested potential gender differences in behavioral associations, most notably for chatbot/virtual assistant use. Nominal interactions were observed for digital literacy (interaction aOR 2.10, 95% CI 1.12-3.96; *P*=.021) and capability (interaction aOR 1.83, 95% CI 1.02-3.29; *P*=.044), as shown in Table S1 in Multimedia Appendix 2. Adjusted predicted probabilities further illustrated these patterns. For capability, the adjusted predicted probability of chatbot use decreased from 92.1% at −1 SD to 70.7% at +1 SD among men and from 67.5% to 55.5% among women, suggesting a steeper decline among men. For digital literacy, the adjusted predicted probability of chatbot use decreased from 89.0% to 76.4% among men but increased from 55.8% to 64.8% among women across the same range.

Interaction estimates were directionally consistent across specifications. In simplified models retaining only the primary predictors and age, chatbot interactions were similar but more precise (digital literacy aOR 2.29, 95% CI 1.28-4.07; capability aOR 1.99, 95% CI 1.17-3.39), and the wearable capability interaction reached nominal significance (aOR 1.82, 95% CI 1.05-3.18) having been nonsignificant under full adjustment (aOR 1.63, 95% CI 0.89-2.99). These models omit education, occupation, region, and rural/remote residence and are reported only as a check on model complexity. The interaction findings are hypothesis generating.

Gender-stratified analyses provided additional descriptive evidence of potentially differing behavioral patterns. Opportunity was positively associated with mHealth app adoption in both genders. For wearable devices, opportunity was positively associated with adoption among both women and men, while higher capability was associated with lower adoption among men. For chatbot use, higher capability was associated with lower adoption among men, whereas no statistically significant association was observed among women. These stratified findings were interpreted descriptively and not as independent evidence of gender differences.

Gendered Barriers to Digital Health Technology Adoption (COM-B Mapping)

Reported barriers were consistent with the COM-B framework (Table 4). A detailed summary of self-reported barriers to digital health technology adoption by COM-B domain is provided in Multimedia Appendix 3, whereas Table 4 provides a brief overview. Motivation-related barriers, such as preference for in-person care or low perceived need, were the most commonly reported category overall, accounting for 41.5% (34/82; mHealth apps) to 69.6% (32/46; e-pharmacy) of nonuse reasons across technologies, and were the leading barriers for telehealth, chatbots/virtual assistants, and e-pharmacy. For wearable devices, opportunity-related barriers were the most frequent (n=36, 46.2%), and for mHealth apps, motivation- and capability-related barriers were equally common (n=34, 41.5% each). Capability-related barriers were otherwise less frequent, ranging from 6.5% (3/46; e-pharmacy) to 21.0% (21/100; chatbots/virtual assistants), whereas opportunity-related constraints, such as limited access, cost, and lack of availability, varied by gender and technology.

XSL•FO
RenderX

Table 4. Self-reported reasons for nonuse of digital health technologies by gender, mapped to COM-B[a] domains.

Technology	Nonusers with a codable reason, n (%)	Opportunity, n/N (%)[b]	Motivation, n/N (%)[b]	Capability, n/N (%)[b]	*P* value[c]
Telehealth	30 (96.8)	M[d]: 4/14 (28.6); F[d]: 6/16 (37.5)	M: 8/14 (57.1); F: 10/16 (62.5)	M: 2/14 (14.3); F: 0	.460
mHealth[e] app	82 (88.2)	M: 5/55 (9.1); F: 9/27 (33.3)	M: 22/55 (40.0); F: 12/27 (44.4)	M: 28/55 (50.9); F: 6/27 (22.2)	.009
Wearable device	78 (89.7)	M: 8/25 (32.0); F: 28/53 (52.8)	M: 12/25 (48.0); F: 22/53 (41.5)	M: 5/25 (20.0); F: 3/53 (5.7)	.079[f]
Chatbot/virtual assistant	100 (92.6)	M: 7/28 (25.0); F: 16/72 (22.2)	M: 14/28 (50.0); F: 42/72 (58.3)	M: 7/28 (25.0); F: 14/72 (19.4)	.755
e-Pharmacy	46 (78.0)	M: 5/16 (31.3); F: 6/30 (20.0)	M: 11/16 (68.8); F: 21/30 (70.0)	M: 0; F: 3/30 (10.0)	.464

[a]COM-B: Capability, Opportunity, Motivation – Behavior.

[b]Percentages are within-gender column percentages.

[c]Fisher's exact test.

[d]M: male; F: female.

[e]mHealth: mobile health.

[f]*P*=.05-.10 (trend). *P*<.05 indicates a statistically significant gender difference. Totals differ from nonuser counts in Table 2 because barrier analyses included only nonusers who provided a codable reason for nonuse.

Women more frequently reported opportunity-related barriers for mHealth apps, including issues of access, affordability ("too expensive"), and contextual support (eg, "never been offered"), whereas men more commonly cited capability-related barriers, including uncertainty and technical skill limitations ("not sure how to use it"). Similar, but nonsignificant, patterns were observed for wearable devices, with women emphasizing access and affordability (n=28, 52.8%), including cost ("too expensive") and lack of device ownership ("don't have a smartwatch"), and men citing technical skill limitations (20%). For telehealth, chatbots, and e-pharmacies, motivational barriers such as low perceived usefulness ("no need") and preference for in person care ("prefer human") predominated for both genders, although opportunity related constraints were also evident, particularly for chatbots ("haven't come across any") and e-pharmacies ("not offered," "haven't had the chance").

These findings suggest that structural barriers, such as cost, limited access, and lack of availability, were more prominent among women for selected technologies, most clearly mHealth apps, where the gender difference in the barrier profile was statistically significant. For mHealth apps, men more frequently reported capability-related barriers, including uncertainty and technical skill limitations. These qualitative findings complemented, but did not fully mirror, the quantitative results. In the quantitative models, higher capability was sometimes associated with a lower likelihood of use among men, particularly for chatbots and wearable devices. This divergence may reflect selective nonadoption among more confident users who perceive limited value in specific technologies. Meanwhile, the pooled quantitative models identified opportunity as an important overall correlate of adoption across several technologies. Taken together, these findings suggest technology-specific opportunity constraints rather than a consistently female-specific opportunity effect across all technologies.

Discussion

Principal Findings

This study applied the COM-B framework to examine gender differences in digital health technology adoption [33]. Although women reported higher digital literacy, capability, and motivation than men, they demonstrated lower unadjusted adoption of several technologies, particularly wearables, chatbots, and e-pharmacies. After adjustment, the most robust gender difference remained for chatbot/virtual assistant use, with a marginal difference observed for e-pharmacy use. These findings contribute to understanding the digital health gender paradox and suggest that lower engagement among women is unlikely to reflect a simple deficit in digital literacy, capability, or willingness to use these technologies. Rather, they are consistent with evidence that gendered digital health inequities may arise through system design, access, governance, and service contexts, rather than individual skill deficits alone [25,38].

The principal contribution of these findings is not simply that opportunity was an important overall predictor of adoption but also that the relationships between behavioral determinants and adoption appeared to vary according to both gender and technology type. Women's lower unadjusted use of several technologies occurred despite their higher reported digital literacy, capability, and motivation, indicating that greater perceived capacity or readiness to use digital technologies does not necessarily translate into greater engagement. Exploratory interaction and gender-stratified analyses further suggested gender-differentiated associations for selected technologies, particularly chatbot/virtual assistant use. These findings extend the digital health gender paradox by suggesting that the behavioral conditions associated with adoption may differ across gender groups, although these patterns were not uniform across

XSL•FO
RenderX

technologies and should be regarded as hypothesis generating rather than evidence of a single gender-specific mechanism.

An important distinction arising from these findings is between the capacity to use a digital health technology and actual engagement with that technology. Digital literacy and perceived capability reflect whether an individual has, or perceives themselves to have, the skills and confidence required to engage, whereas adoption also depends on whether appropriate opportunities for use are available and whether the technology is perceived as relevant or worthwhile. In this study, higher reported digital literacy among women did not consistently translate into greater technology use. This finding aligns with emerging evidence that, although necessary, digital literacy alone is insufficient to drive adoption in the absence of supportive structural conditions [12,13,15,39]. Effective digital health technology use requires not only skills and confidence but also access to technologies, integration into care pathways, and trust in how technologies are designed and delivered. Interventions focused solely on improving digital literacy are therefore likely to have limited impact unless accompanied by strategies that address opportunity-related barriers. It is important to note though that digital literacy and perceived capability are conceptually distinct from actual engagement. Having the capacity or confidence to use a technology does not necessarily mean that it is accessible, affordable, relevant, trusted, integrated into care, or ultimately used. Infodemiological research using Baidu search data similarly illustrates how observable population-level health information-seeking behavior can reveal patterns of digital engagement that are not equivalent to self-reported knowledge, capability, or intentions [40,41]. Although these studies do not directly examine the gender paradox investigated here, they provide a complementary example of the distinction between perceived capacity and observable digital behavior.

Opportunity emerged as one of the most consistent determinants of adoption across several technologies, particularly telehealth, mHealth apps, and wearable devices, reinforcing the importance of structural and contextual factors in enabling digital health engagement. However, the gender-specific pattern was more nuanced. In the self-reported barrier data, women more frequently identified opportunity-related constraints for some technologies, most clearly for mHealth apps, including barriers relating to access, affordability, and whether digital options were available or offered. A similar pattern was observed for wearable devices, although the evidence was less certain. Men, in contrast, more frequently reported capability-related barriers for some technologies. Together with the interaction and stratified analyses, these findings suggest potentially different behavioral patterns across gender groups, while also demonstrating that these patterns depend on the specific technology being considered. The opportunity-telehealth association should be interpreted with particular caution, given the small number of telehealth nonusers (n=31) relative to model parameters, which may contribute to imprecision in this estimate.

Alternative explanations should also be considered. The technologies examined differ substantially in purpose, cost, availability, required device ownership, integration with health care services, and the degree of autonomy required from users. Gender differences in health needs, prior exposure to particular technologies, device ownership, age distribution, occupation, or patterns of health care use may therefore also contribute to the observed adoption differences. In particular, the substantial difference in occupational composition between genders, with approximately three-quarters of women in white-collar roles compared with 43% of men, may have contributed to observed adoption patterns, given that white-collar occupations may involve greater prior exposure to digital technologies; although occupation was included as a covariate in all models, residual confounding from unmeasured aspects of occupational digital exposure cannot be excluded. Although the multivariable analyses adjusted for several sociodemographic characteristics, residual confounding and unmeasured contextual factors cannot be excluded.

Unexpected Findings

After adjusting for COM-B domains, women were more likely than men to adopt mHealth apps. This pattern may indicate that women's engagement with mHealth apps becomes more evident once structural and contextual factors are considered. Given that opportunity was the strongest determinant of mHealth app adoption, the result is consistent with mobile app engagement being associated not only with individual literacy or motivation but also with contextual conditions, such as access, availability, affordability, and service integration. However, because the study was cross-sectional, changes in gender coefficients across adjusted models should not be interpreted as evidence that these factors causally explain the gender difference. This interpretation aligns with prior research showing that digital health engagement is influenced by structural, service, access, and support-related conditions, even when digital literacy is present [13,25,39].

An inverse association between capability and chatbot adoption was observed among men. This may reflect selective nonuse among digitally confident individuals who perceive limited value, reliability, or usefulness in automated health agents. Emerging evidence highlights the role of trust, perceived usefulness, and user concerns in the adoption of artificial intelligence (AI)–enabled health chatbots, including concerns about accuracy, cybersecurity, empathy, and preference for human interaction [42-44].

Motivation showed a positive association with e-pharmacy adoption in the primary model; however, this association was attenuated and the CI included the null in the simplified sensitivity model. This finding should therefore be interpreted with caution and regarded as hypothesis generating. It may potentially reflect the role of trust and perceived safety when technologies involve medication management or financial transactions, consistent with prior research showing that perceived risk, trust, privacy, price, availability, and product-related concerns influence consumers' willingness to purchase medicines online [45,46]; however, confirmation in larger, adequately powered studies is needed.

XSL•FO
RenderX

Addressing Opportunity Barriers in Digital Health Technology Adoption

The distinction between capacity and engagement has practical implications for intervention design. Where opportunity-related barriers were observed in this study, particularly barriers concerning access, affordability, availability, and the offering of digital options, strategies directed solely at increasing digital skills are unlikely to address the full range of conditions associated with nonuse. Instead, implementation approaches may also need to modify the environments in which digital health technologies are made available and offered to consumers. The BCW proposes that opportunity-related barriers can be addressed through intervention functions, such as environmental restructuring and enablement, which modify the context in which behaviors occur [33,47]. Interventions most directly aligned with barriers reported in this study could therefore include improving access to appropriate technologies, reducing financial or device-related barriers, and ensuring that suitable digital options are consistently offered within health care encounters. In the context of digital health, this may include integrating digital services into clinical workflows; proactively offering digital options during health care encounters; and reducing financial, technological, or service access barriers that limit engagement, as suggested by broader implementation literature [7,15,48]. Communication strategies may further support adoption by increasing awareness, improving acceptability, and normalizing digital health services. Patient portal implementation provides a directly relevant example. In a multi-institutional patient portal implementation, uptake was shaped not only by patient interest or capability but also by provider engagement, change management support, leadership endorsement, broad rollout and marketing strategies, technology infrastructure, and policy alignment [49]. Similarly, a systematic review of patient portal adoption found that adoption rates were substantially higher in actively recruited controlled studies than in real-world deployments, suggesting that availability alone is insufficient unless implementation creates practical opportunities for use [50]. These examples illustrate how service design and implementation can influence the practical opportunity to engage with digital services, even among individuals who may already possess the skills required to use them. These findings should also be interpreted within broader gender inequalities in digital participation. Although women increasingly achieve high levels of education and digital skills, structural disparities persist in areas such as access to technology, participation in the digital workforce, and representation in technology development. These factors may influence how digital technologies are designed, implemented, and experienced by users [21,24,51]. Evidence from Australia similarly highlights the importance of gender diversity and inclusion in the digital health workforce, particularly for ensuring that digital health innovation reflects diverse perspectives and needs [38].

Taken together, these findings suggest that digital health technology adoption should not be conceptualized solely as an issue of individual digital literacy. Rather, adoption is shaped by structural opportunity, including access to services, integration into health care pathways, and the broader social context in which digital technologies are developed and deployed. Without interventions that expand opportunity and embed digital services within routine care, digital health initiatives risk reinforcing existing gender inequities despite high levels of individual capability and motivation.

Strengths and Implications

The integration of quantitative modeling with self-reported barrier data strengthened interpretation by allowing behavioral associations identified quantitatively to be considered alongside participants' reported reasons for nonuse. The inclusion of five different digital health technologies was also valuable because it demonstrated that adoption determinants were not uniform across technology types. Digital literacy levels were relatively high within the sample, providing a useful context in which to examine why comparatively high self-reported literacy and capability did not consistently correspond with adoption. Future research should examine longitudinal adoption patterns and evaluate interventions targeting opportunity-related barriers.

Limitations

The cross-sectional study design limits causal inference and does not capture longitudinal changes in digital health technology adoption. Measures of digital literacy and technology use were self-reported and may therefore be subject to recall or social desirability bias. The study used convenience sampling through social media, consumer advocacy networks, Medibank communication channels, and research team networks; therefore, the sample may be subject to selection bias and should not be considered nationally representative of Australian adults. The binary adoption measure did not capture the frequency, duration, or intensity of technology use and may therefore represent different levels of engagement across the five technologies.

The sample was moderately skewed toward men (61.1% male), which may reflect the recruitment channels used and may limit generalizability. The high uptake observed, with 99.3% of respondents reporting use of at least one digital health technology, suggests that the sample may have included individuals with relatively high baseline engagement in digital health. This may limit the extent to which findings apply to populations with lower digital access, lower digital literacy, or less prior exposure to digital health technologies. Although this imbalance may have reduced statistical power to detect smaller gender-by-COM-B interaction effects, analyses were adjusted for key sociodemographic variables and supplemented with exploratory gender-stratified models.

Conclusion

This study provided behavioral evidence on the digital health technology adoption gender paradox, showing that women's lower unadjusted engagement with several digital health technologies occurs despite higher reported digital literacy, capability, and motivation. Using the COM-B framework, the findings suggest that behavioral determinants of adoption may vary by gender and technology type, with opportunity-related barriers, including access, affordability, and inconsistent offering of digital services, appearing particularly relevant to women's nonuse for some technologies. Some findings also suggested that men's nonuse may be more closely linked to capability-related concerns, including confidence and perceived

XSL•FO
RenderX

ease of use. These gendered behavioral patterns suggest that digital transformation initiatives focused primarily on improving digital literacy may not achieve equitable adoption. Addressing structural opportunity barriers through gender-responsive, theory-informed implementation strategies may therefore be important for supporting equitable digital health technology adoption.

Acknowledgments

The authors acknowledge the Consumer Health Forum of Australia and participating consumer organizations for their contributions to survey review and accessibility feedback during survey development. The authors also acknowledge Medibank Private Limited for supporting participant recruitment and funding the participant prize draw voucher. The authors thank all participants who took part in the survey and shared their experiences and perspectives regarding digital health technologies.

Funding Statement

This study received no external competitive funding. Medibank Private Limited provided support for participant recruitment and funded the participant prize draw.

Generative AI Use Disclosure

ChatGPT (OpenAI) was used for language editing and manuscript refinement, including improving grammar, clarity, concision, and wording during manuscript development. All artificial intelligence–assisted content was reviewed and verified by the authors, who take full responsibility for the final manuscript.

Data Availability

The datasets generated and analyzed during the study are not publicly available due to ethics and privacy restrictions associated with participant consent and de-identified survey data but are available from the corresponding author upon reasonable request and subject to ethics approval requirements. For enquiries relating to the research presented in this manuscript, please contact URK.

Authors' Contributions

Conceptualization: URK, ML, JB, TAW, and JKW
Data curation: URK, ML, MB, and TAW
Formal analysis: MB, URK, TAW, and ML
Investigation: JKW, MB, TAW, ML, JB, TK, and KO
Methodology: MB, ML, URK, TAW, and JKW
Project administration: JKW, URK, TAW, and MB
Supervision: URK, JB, and TK
Validation: URK, MB, TAW, ML, JB, TK, and KO
Writing—original draft: JKW
Writing—review and editing: TAW, MB, ML, KO, JB, TK, and URK

Conflicts of Interest

Medibank Private Limited contributed to survey development, supported participant recruitment through its communication channels, and funded the AU $50 (US $36) participant prize-draw voucher, as described in the Funding Statement section. This involvement is disclosed because of the potential for a perceived conflict of interest arising from the participation of a private health insurer in the study. Medibank's involvement was intended to facilitate recruitment and broaden the reach of this consumer-focused digital health survey through its established communication channels. Beyond these activities, Medibank Private Limited played no role in data analysis, interpretation of the findings, manuscript preparation, or the decision to submit the manuscript for publication. The authors declare no other conflicts of interest.

Multimedia Appendix 1
Survey instrument.
[PDF File (Adobe PDF File), 74 KB - humanfactors_v13i1e101963_app1.pdf]

Multimedia Appendix 2
Gender interaction and stratified models.
[DOCX File , 38 KB - humanfactors_v13i1e101963_app2.docx]

XSL•FO
RenderX

Multimedia Appendix 3
Self-reported barriers to digital health technology adoption.
[DOCX File , 21 KB - humanfactors_v13i1e101963_app3.docx]

References

1. Ageing and health. World Health Organization. 2025. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2026-05-19]
2. Organisation for Economic Co-operation and Development. Health at a Glance 2023: OECD Indicators. Paris: OECD Publishing; 2023.
3. Shaw T, McGregor D, Brunner M, Keep M, Janssen A, Barnet S. What is eHealth (6)? Development of a conceptual model for eHealth: qualitative study with key informants. J Med Internet Res 2017 Oct 24;19(10):e324 [FREE Full text] [doi: 10.2196/jmir.8106] [Medline: 29066429]
4. Van Velthoven MH, Cordon C. Sustainable adoption of digital health innovations: perspectives from a stakeholder workshop. J Med Internet Res 2019 Mar 25;21(3):e11922 [FREE Full text] [doi: 10.2196/11922] [Medline: 30907734]
5. Savira F, Orellana L, Hensher M, Gao L, Sanigorski A, Mc Namara K, et al. Use of general practitioner telehealth services during the COVID-19 pandemic in regional Victoria, Australia: retrospective analysis. J Med Internet Res 2023 Feb 07;25:e39384 [FREE Full text] [doi: 10.2196/39384] [Medline: 36649230]
6. Auditor-general report no. 10 2022-23: expansion of telehealth services (performance audit). Australian National Audit Office. 2023. URL: https://www.anao.gov.au/sites/default/files/2023-01/Auditor-General_Report_2022-23_10.pdf [accessed 2026-05-19]
7. National digital health strategy 2023-28. Australian Digital Health Agency (ADHA). 2023. URL: https://www.digitalhealth.gov.au/sites/default/files/documents/national-digital-health-strategy-2023-2028.pdf [accessed 2026-05-19]
8. Hepburn J, Williams L, McCann L. Barriers to and facilitators of digital health technology adoption among older adults with chronic diseases: updated systematic review. JMIR Aging 2025 Sep 11;8:e80000 [FREE Full text] [doi: 10.2196/80000] [Medline: 40934502]
9. Krahe MA, Larkins SL, Adams N. Digital health implementation in Australia: a scientometric review of the research. Digit Health 2024;10:20552076241297729 [FREE Full text] [doi: 10.1177/20552076241297729] [Medline: 39539722]
10. Patel SB, Iqbal FM, Lam K, Acharya A, Ashrafian H, Darzi A. Characterizing behaviors that influence the implementation of digital-based interventions in health care: systematic review. J Med Internet Res 2025 Jun 12;27:e56711 [FREE Full text] [doi: 10.2196/56711] [Medline: 40505130]
11. Voorheis P, Bhuiya AR, Kuluski K, Pham Q, Petch J. Making sense of theories, models, and frameworks in digital health behavior change design: qualitative descriptive study. J Med Internet Res 2023 Mar 15;25:e45095 [FREE Full text] [doi: 10.2196/45095] [Medline: 36920442]
12. Alam K, Mahumud RA, Alam F, Keramat SA, Erdiaw-Kwasie MO, Sarker AR. Determinants of access to eHealth services in regional Australia. Int J Med Inform 2019 Nov;131:103960 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.103960] [Medline: 31518858]
13. Cheng C, Gearon E, Hawkins M, McPhee C, Hanna L, Batterham R, et al. Digital health literacy as a predictor of awareness, engagement, and use of a national web-based personal health record: population-based survey study. J Med Internet Res 2022 Sep 16;24(9):e35772 [FREE Full text] [doi: 10.2196/35772] [Medline: 36112404]
14. Krahe MA, Baker S, Woods L, Larkins SL. Factors that influence digital health implementation in rural, regional, and remote Australia: an overview of reviews and recommended strategies. Aust J Rural Health 2025 Apr;33(2):e70045 [FREE Full text] [doi: 10.1111/ajr.70045] [Medline: 40202368]
15. Global strategy on digital health 2020-2025. World Health Organisation. 2021. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf [accessed 2026-05-19]
16. Thanthiriwattage R, Liem M, Nouman M, Wani T, Marriner T, Ovenden K, et al. Australian healthcare consumers 'curiosity' in digital health technologies. Stud Health Technol Inform 2025 Nov 12;333:14-19 [FREE Full text] [doi: 10.3233/SHTI251568] [Medline: 41235485]
17. McLean P, Oldfield J, Stephens A. Draft Digital Literacy Skills Framework. Canberra, ACT, Australia: Australian Government Department of Education, Skills and Employment; 2020.
18. Cervera RL, Wang MZ, Hayden BY. Systems neuroscience of curiosity. Curr Opin Behav Sci 2020 Oct;35:48-55 [FREE Full text] [doi: 10.1016/j.cobeha.2020.06.011]
19. Kashdan TB, Gallagher MW, Silvia PJ, Winterstein BP, Breen WE, Terhar D, et al. The curiosity and exploration inventory-II: development, factor structure, and psychometrics. J Res Pers 2009 Dec 01;43(6):987-998 [FREE Full text] [doi: 10.1016/j.jrp.2009.04.011] [Medline: 20160913]
20. Cracking the code: girls' and women's education in science, technology, engineering and mathematics. UNESCO. 2017. URL: https://unesdoc.unesco.org/ark:/48223/pf0000253479 [accessed 2026-05-19]
21. Global gender gap report 2025. World Economic Forum. 2025. URL: https://www.weforum.org/publications/global-gender-gap-report-2025/ [accessed 2026-05-19]

XSL•FO
RenderX

22. Women in work 2023: closing the gender pay gap for good: a focus on the motherhood penalty. PricewaterhouseCoopers (PwC). 2023. URL: https://www.pwc.co.uk/economic-services/WIWI/pwc-women-in-work-index-2023.pdf [accessed 2026-05-19]
23. Borges do Nascimento IJ, Abdulazeem HM, Weerasekara I, Marquez J, Vasanthan LT, Deeken G, et al. Transforming women's health, empowerment, and gender equality with digital health: evidence-based policy and practice. Lancet Digit Health 2025 Jun;7(6):100858 [FREE Full text] [doi: 10.1016/j.landig.2025.01.014] [Medline: 40368744]
24. Progress on the sustainable development goals: the gender snapshot 2025. United Nations Women, United Nations Department of Economic and Social Affairs. 2025 Sep. URL: https://tinyurl.com/mpe94vtu [accessed 2026-05-19]
25. Figueroa CA, Luo T, Aguilera A, Lyles CR. The need for feminist intersectionality in digital health. Lancet Digit Health 2021 Aug;3(8):e526-e533 [FREE Full text] [doi: 10.1016/s2589-7500(21)00118-7]
26. Lupton D, Maslen S. How women use digital technologies for health: qualitative interview and focus group study. J Med Internet Res 2019 Jan 25;21(1):e11481 [FREE Full text] [doi: 10.2196/11481] [Medline: 30681963]
27. Joshi A. Big data and AI for gender equality in health: bias is a big challenge. Front Big Data 2024 Oct 16;7:1436019 [FREE Full text] [doi: 10.3389/fdata.2024.1436019] [Medline: 39479339]
28. Informal carers. Australian Institute of Health and Welfare. 2025. URL: https://www.aihw.gov.au/reports/australias-welfare/informal-carers [accessed 2026-05-19]
29. Health workforce. Australian Institute of Health and Welfare. 2024. URL: https://www.aihw.gov.au/reports/workforce/health-workforce [accessed 2026-05-19]
30. Lupton D. The Australian Women and Digital Health Project: comprehensive report of findings. News and Media Research Centre. 2019 Jan. URL: https://apo.org.au/node/220326 [accessed 2026-05-19]
31. Organisation for Economic Co-operation and Development. Development Co-operation Report 2021: Shaping a Just Digital Transformation. Paris, France: OECD Publishing; 2021.
32. Yao R, Zhang W, Evans R, Cao G, Rui T, Shen L. Inequities in health care services caused by the adoption of digital health technologies: scoping review. J Med Internet Res 2022 Mar 21;24(3):e34144 [FREE Full text] [doi: 10.2196/34144] [Medline: 35311682]
33. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
34. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340 [FREE Full text] [doi: 10.2307/249008]
35. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003 Sep 1;27(3):425-478 [FREE Full text] [doi: 10.2307/30036540]
36. Australian digital capability framework: version 1.0. Department of Employment and Workplace Relations, Australian Government. 2022. URL: https://tinyurl.com/yc7wkjpm [accessed 2026-05-19]
37. The complete statistical software for data science. StataCorp LLC. 2025. URL: https://www.stata.com/ [accessed 2026-09-19]
38. James C, Zevgitis L, Arabi S. Gender diversity in Australia's digital health workforce: special report 2024, 2nd ed. 2024 Oct. URL: https://www.telstrahealth.com/wp-content/uploads/2024/10/DiversityReport_Final.pdf [accessed 2026-05-19]
39. Thomas J, McCosker A, Parkinson S, Hegarty K, Featherstone D, Kennedy J, et al. Measuring Australia's Digital Divide: Australian Digital Inclusion Index 2023. Melbourne: ARC Centre of Excellence for Automated Decision-Making and Society, RMIT University, Swinburne University of Technology, and Telstra; 2023.
40. Lin S, Duan L, Xu X, Cao H, Lu X, Wen X, et al. Analyzing online search trends for kidney, prostate, and bladder cancers in China: infodemiology study using Baidu search data (2011-2023). JMIR Cancer 2025 Mar 14;11:e57414 [FREE Full text] [doi: 10.2196/57414] [Medline: 40085845]
41. Wei S, Ma M, Wu C, Yu B, Jiang L, Wen X, et al. Using search trends to analyze web-based interest in lower urinary tract symptoms-related inquiries, diagnoses, and treatments in mainland China: infodemiology study of Baidu index data. J Med Internet Res 2021 Jul 06;23(7):e27029 [FREE Full text] [doi: 10.2196/27029] [Medline: 34255683]
42. Wu PF, Summers C, Panesar A, Kaura A, Zhang L. AI hesitancy and acceptability—perceptions of AI chatbots for chronic health management and long COVID support: survey study. JMIR Hum Factors 2024 Jul 23;11:e51086 [FREE Full text] [doi: 10.2196/51086] [Medline: 39045815]
43. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019;5:2055207619871808 [FREE Full text] [doi: 10.1177/2055207619871808] [Medline: 31467682]
44. Mertz M, Toskovich K, Shields G, Attema G, Dumond J, Cameron E. Exploring trust factors in AI-healthcare integration: a rapid review. Front Artif Intell 2025 Dec 4;8:1658510 [FREE Full text] [doi: 10.3389/frai.2025.1658510] [Medline: 41425055]
45. Limbu YB, Huhmann BA. What influences consumers' online medication purchase intentions and behavior? A scoping review. Front Pharmacol 2024;15:1356059 [FREE Full text] [doi: 10.3389/fphar.2024.1356059] [Medline: 38414739]
46. Ersöz S, Nissen A, Schütte R. Risk, trust, and emotion in online pharmacy medication purchases: multimethod approach incorporating customer self-reports, facial expressions, and neural activation. JMIR Form Res 2023 Dec 25;7:e48850 [FREE Full text] [doi: 10.2196/48850] [Medline: 38145483]

XSL•FO
RenderX

47. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. London, UK: Silverback; 2014.
48. Whitelaw S, Mamas MA, Topol E, Van Spall HGC. Applications of digital technology in COVID-19 pandemic planning and response. Lancet Digit Health 2020 Aug;2(8):e435-e440 [FREE Full text] [doi: 10.1016/S2589-7500(20)30142-4] [Medline: 32835201]
49. Fujioka JK, Bickford J, Gritke J, Stamenova V, Jamieson T, Bhatia RS, et al. Implementation strategies to improve engagement with a multi-institutional patient portal: multimethod study. J Med Internet Res 2021 Oct 28;23(10):e28924 [FREE Full text] [doi: 10.2196/28924] [Medline: 34709195]
50. Fraccaro P, Vigo M, Balatsoukas P, Buchan I, Peek N, van der Veer SN. Patient portal adoption rates: a systematic literature review and meta-analysis. 2017 Presented at: MEDINFO 2017: 16th World Congress on Medical and Health Informatics; August 21-25, 2017; Hangzhou, China p. 79-83. [doi: 10.3233/978-1-61499-830-3-79]
51. I'd blush if I could: closing gender divides in digital skills through education. UNESCO & EQUALS. 2019. URL: https://unesdoc.unesco.org/ark:/48223/pf0000367416 [accessed 2026-05-19]

Abbreviations

AI: artificial intelligence
aOR: adjusted odds ratio
BCW: Behavior Change Wheel
COM-B: Capability, Opportunity, Motivation – Behavior
mHealth: mobile health
REDCap: Research Electronic Data Capture

Edited by P Santana-Mancilla; submitted 21.May.2026; peer-reviewed by S Wei, R Acosta-Diaz; comments to author 11.Aug.2026; revised version received 11.Sep.2026; accepted 11.Sep.2026; published 29.Sep.2026.

Please cite as:
Wallace JK, Batra M, Wani TA, Liem M, Ovenden K, Boyd J, Keane T, Khan UR
Exploratory Analysis of the Digital Health Gender Paradox in Consumer-Facing Digital Health Technology Adoption: Cross-Sectional Survey Study
JMIR Hum Factors 2026;13:e101963
URL: https://humanfactors.jmir.org/2026/1/e101963
doi:10.2196/101963
PMID:

©Jessica Kathleen Wallace, Mehak Batra, Tafheem Ahmad Wani, Michael Liem, Kylie Ovenden, James Boyd, Therese Keane, Urooj Raza Khan. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 29.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Use of Digital Technology to Support the Emotional Well-Being of Older People Who Receive Care at Home From the Perspective of Care Recipients: Qualitative Study

Tanya E Davison[1,2,3], DPsych(Clin); Hilary Davis[4], PhD; Marissa Dickins[1,5], PhD; Sonja Pedell[6], PhD; Tuan Anh Nguyen[7,8,9,10], PhD; Deborah Carveth[1,3], MCoun; Therese Den Dulk[1,3], MCoun; Karen Smith[1,5], PhD; Julie Kelly[1,3], MClinPsych; Sunil Bhar[3], PhD

[1]Research and Innovation, Silver Chain Group Ltd, Melbourne, Victoria, Australia

[2]Department of Design, Faculty of Art, Architecture and Design, Monash University, Melbourne, Victoria, Australia

[3]Department of Psychological Sciences, School of Health Sciences, Swinburne University of Technology, Melbourne, Victoria, Australia

[4]Centre for Social Impact, School of Business, Law and Entrepreneurship, Swinburne University of Technology, Melbourne, Victoria, Australia

[5]School of Public Health and Preventive Medicine, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Victoria, Australia

[6]Centre for Design Innovation, School of Design and Architecture, Swinburne University of Technology, Melbourne, Victoria, Australia

[7]Centre for Mental Health and Brain Sciences, School of Health Sciences, Swinburne University of Technology, Melbourne, Victoria, Australia

[8]Social Gerontology Division, National Ageing Research Institute, Melbourne, Victoria, Australia

[9]College of Health Sciences, VinUniversity, Hanoi, Vietnam

[10]School of Pharmacy and Biomedical Science, Adelaide University, Adelaide, South Australia, Australia

Corresponding Author:
Tanya E Davison, DPsych(Clin)
Research and Innovation
Silver Chain Group Ltd
Level 27, 385 Bourke Street
Melbourne, Victoria, 3000
Australia
Phone: 61 408394218
Email: tanya.davison@silverchain.org.au

Abstract

Background: Older people who receive aged care services at home experience high rates of depression, anxiety, and loneliness, yet have poor access to services to support their well-being. Digital interventions are widely available, but research is required to determine whether they are acceptable to the in-home aged care population.

Objective: This study aimed to explore the perspectives of in-home aged care recipients on the use of digital technologies to support their emotional well-being.

Methods: Eleven older people (aged 70-87 years) who received in-home aged care from a national provider participated in a semistructured interview. The interview aimed to understand care recipients' experiences of using technology, as well as their current access to emotional support and attitudes toward using digital technologies to support their well-being.

Results: Participants reported high access to technology (mean 5, SD 1.6 devices) but varied usage. Most required support to use technology, typically provided by family members, and reported a range of concerns related to technology use. Participants received a variety of informal supports for their emotional well-being, typically from family members and their communities, but only one accessed professional support. Most were open to using technology to support their emotional well-being, and some were willing to receive such support from their aged care provider.

Conclusions: The willingness of older people who receive in-home aged care to use digital technologies to supplement existing informal emotional support is encouraging. Aged care providers are well placed to implement digital well-being programs and can leverage older people's existing access to digital devices but should consider their requirements for technical support. Technologies and implementation strategies should be purpose-designed for this population and address identified barriers.

XSL•FO
RenderX

(JMIR Hum Factors 2026;13:e93823) doi:10.2196/93823

KEYWORDS

aged care; home care; social care; digital technology; older adults; well-being; mental health; digital health; digital literacy; digital usability

Introduction

Globally, population aging has resulted in increased demand for health care and aged care services, with growing numbers of older people requiring assistance to manage increasingly complex needs [1]. Older people typically prefer to remain at home for as long as possible [2], and many countries have observed significant increases in the number accessing in-home aged care (also known as social care). For example, in Australia, the number of older people who received a government-funded Home Care Package to enable them to remain at home nearly quadrupled over the past 10 years [3].

Symptoms of depression and anxiety, and other indicators of poor emotional well-being, such as loneliness, social isolation, grief, and adjustment difficulties, affect more than half of older people who receive aged care at home [4,5] and appear to be increasing alongside other chronic health conditions, frailty, and multimorbidity [6]. Aged care providers commonly provide social support to lonely or socially isolated clients (eg, drive clients to social events or provide company in the home) and may escalate to the older person's medical practitioner for follow-up if they suspect a mental health problem. However, aged care recipients have limited access to specialist services, with recent research indicating that only 2.5% of home care recipients in Australia use government-funded mental health support [7].

With shortages in both aged care and mental health workforces [8,9], alternative approaches are required to address the growing mental health and well-being needs of aged care recipients. Older adults experience significant barriers to accessing outpatient services, especially those with impaired mobility and who live in regions where services are sparse [10]. Digital technologies offer an opportunity to deliver interventions and support at scale to reach large numbers of older people with depression, anxiety, and other aspects of poor emotional well-being who currently have poor access to support and address the current scarcity of mental health in-reach programs in the community [11].

There has been enormous growth in the development of digital mental health and well-being interventions in recent years. These interventions use a variety of technologies and can be grouped into several overlapping categories based on existing taxonomies [12]: (1) synchronous provider-delivered support, such as counseling or psychotherapy delivered by videoconference; (2) guided digital interventions, including web-based or app-based programs supported by a clinician, coach, or facilitator; (3) self-guided tools, such as mobile apps or websites for mood monitoring, psychoeducation, mindfulness, relaxation, or activity planning; (4) social connection technologies, including video calling, online peer-support groups, forums, and social media platforms; and (5) supportive or assistive technologies, such as smart speakers, reminders, and conversational agents. This typology highlights that digital support for emotional well-being extends beyond formal treatment of depression or anxiety to include technologies that foster social connectedness, meaningful activity, and day-to-day self-management in the home environment. The paper uses "emotional well-being" as an umbrella term to encompass mental health symptoms and conditions, as well as related aspects of well-being, including distress, grief, sense of purpose, and social connection. This usage is consistent with recent conceptualizations of emotional well-being as encompassing affective experience, meaning and purpose, and social connection [13].

While most technologies to support emotional well-being have been designed for younger people [14], there is promising evidence that some digital interventions (eg, online cognitive behavior therapy) are effective in reducing symptoms of depression in older people when designed specifically for this population [15], although reviewers have noted that more research is needed [16]. Digital interventions may also be effective in improving loneliness and social connection in older people [17,18]. However, there is little work exploring their use in older people who receive aged care at home.

Older people are generally open to learning to use digital devices [19], and there is an established literature on the use of digital technologies to facilitate health care by older people [20]. Research is also emerging about the views and experiences of older people in using digital technologies to support their emotional well-being, which has been synthesized in recent literature reviews [11,19]. While digital technologies offer convenience and anonymity [19], identified barriers for older people include doubts about the usefulness of digital approaches for emotional well-being, poor access to technology and low digital literacy, as well as physical, cognitive, or sensory impairments, and concerns related to privacy and security [11,19]. Older people with mental health disorders such as depression [19] appear more hesitant to use digital interventions and overall, older people may prioritize the management of other aspects of their health, such as chronic illness, pain, and medications, over mental health and emotional well-being [11]. These findings indicate it is important to consider attitudes toward the use of digital interventions within the broader context of older people's willingness to attend to mental health issues and access emotional support more generally.

There is a particular gap in research examining the use of technology to support emotional well-being among older people who require aged care services to remain at home. Aged care recipients are significantly older than the "older people" recruited to most research studies (eg, in Australia, the median age at commencement of home care services is 81 years [3]) and this population experiences high rates of multimorbidity and frailty [6], as well as other indicators of complexity, such as cognitive decline [21]. While there is growing adoption of

XSL•FO
RenderX

technology by older people generally, the oldest old and those in poor health are more likely to experience a digital divide [22]. Thus, research on use, facilitators, and barriers in relation to digital interventions for emotional well-being derived from general older adult community samples may not generalize to this group. There are indications from interviews with aged care providers that some care recipients are willing to use technology to maintain their social connections, with providers reporting strong uptake of digital communication technologies among their clients during the COVID-19 pandemic [10]. However, there is a lack of published literature on the perspectives of in-home aged care recipients themselves regarding the use of technology to support their emotional well-being.

This study had 3 aims. First, we aimed to gather information from older adults accessing in-home aged care about their experience in using digital technologies in general. Second, we aimed to explore their current access to emotional support and their views on using technology to support their emotional well-being. Finally, we aimed to understand care recipients' willingness to take up such programs when offered by the organization providing their aged care services.

Methods

Study Design and Setting

An interview study was conducted with older people receiving aged care services at home from a single provider. The service provider offers a range of aged care services, including domestic support, personal care, social support, home modifications, nursing, allied health, and transport.

Participants

Study participants were in-home aged care clients who resided in metropolitan Victoria, Australia. Clients were approached by their aged care provider if they were considered to have poor emotional well-being, such as depression or low mood, based on their professional judgment, and were potentially interested in participating in a study on how technology can improve older people's mood and emotional well-being. Participants were not required to have a diagnosis of a mental health disorder, and no mental health or well-being screening was conducted for this study. Eligibility criteria included age of 65 years or older and current receipt of a government-funded aged care service. Participants were not required to have any level of expertise in technology or digital literacy. Exclusion criteria included a diagnosis of dementia or presence of significant cognitive impairment, an acute health-related issue that could interfere with research participation, and difficulty engaging in an interview conducted in English.

Procedure

After an initial conversation with the aged care provider, if the client indicated an interest in taking part in the research, a researcher called them on the telephone, confirmed their willingness to participate, and set up a time to visit their home for the interview. The participant information sheet and consent form were mailed or emailed to all participants prior to the interview. A researcher visited the client's home, obtained written consent, and conducted the audio-recorded interview, using a semistructured interview format. Participants were given the opportunity to have a family carer present during the interview.

Materials

Demographic variables included client age, sex, living arrangement, and geographical location, which were obtained from the participant's care record prior to the home visit. Additional data were collected during the visit: level of education, marital status, and a single item on the presence or absence of a health problem that prevents the participant from doing things people their age can normally do, which was answered "Yes," "No," "Don't Know," or "Prefer Not to Disclose." A full list of demographic questions is available in Multimedia Appendix 1.

A semistructured interview was developed for the study, with interview questions exploring care recipients' existing access to and use of technology, current access to support for their emotional well-being, and their views on the use of digital technologies to improve their mood or emotional well-being, such as web-based platforms, mobile apps, or videoconferences. Questions also explored their views of emotional support being offered by their aged care provider (refer to Multimedia Appendix 1 for interview schedule).

Analysis

Demographic data were analyzed using descriptives (eg, mean, range, and frequency). Qualitative data were transcribed and coded using thematic analysis: a "method for identifying, analyzing and reporting patterns (themes) within data" to find "repeated patterns of meaning" [23]. The 6 phases of thematic analysis described by Braun and Clarke [23], namely, familiarization, generation of initial codes, searching for themes, reviewing themes, and naming themes, followed by report production, were used. No formal codebook was developed. Instead, an inductive strategy was used, which allowed the themes to arise from the data, and the phases of analysis were treated in a reciprocal manner, allowing movement back and forth between the phases. All transcripts were coded by 3 team members (DC, TDD, and HD) using Microsoft Excel. Any discrepancies were resolved through consensus during discussions. Regular discussions were held within the multidisciplinary research team to provide further insight into the themes arising from the analysis. Quotes were selected based on their ability to capture the theme accurately and articulately, ensuring that perspectives were drawn from across the participants.

Ethical Considerations

Ethical approval was granted by the Swinburne University of Technology Human Research Ethics Committee, and data were collected between October and December 2023.

Results

Participant Characteristics

Fifteen clients were nominated for the study by the aged care provider. Two did not meet eligibility criteria (one due to dementia and one due to language issues), 1 could not be

XSL•FO
RenderX

contacted, and 1 declined to participate. The final client sample included 11 aged care recipients (7 female participants and 4 male participants), who were all retired and ranged in age from 70 to 87 (mean 79, SD 5.2) years (Table 1). The sample size of 11 was determined to be sufficient as interviewers identified that saturation had been reached, with no new themes arising in the final interview, and data collection was ceased. Seven participants lived alone, while 4 lived with family members, and 7 were born in Australia. There was a range of formal education: most did not finish secondary school (n=7), but 1 participant had completed secondary schooling, 2 had completed technical or vocational training, and 1 had completed an undergraduate degree. Participants were either married (n=5), widowed (n=4), or divorced or separated (n=2). The majority were living with a health problem that prevented them from doing things people their age can normally do (n=7), but 2 answered “no” to this question, and 2 either were not sure or elected not to answer. Three clients were accompanied by a family carer in the interview, typically their spouse; however, in order to focus on the perspectives of clients themselves, only comments from the clients were included in this analysis.

XSL•FO
RenderX

Table 1. Participant characteristics and technology ownership and usage.

Participant	Sex	Age (years)	Education status	Technology owned	Technology usage
1	Female	78	2 years of secondary school	Smartphone, tablet, computer, smartwatch, and internet access	• Communication (calls, SMS, WhatsApp, and Messenger), entertainment (YouTube), and social media (faith groups on Facebook)
2	Female	79	3 years of secondary school	Smartphone, tablet, computer, smart TV, and internet access	• Communication (calls and SMS), entertainment (puzzles such as Sudoku), information and tasks (plans holidays using Google), and social media (Facebook: reads only, does not post) • Limited use of smartphone
3	Male	83	Technical college	Smartphone, tablet, computer, smart TV, internet access, and electronic book (e-book) reader	• Communication (calls, SMS, email, FaceTime, and Zoom), entertainment (YouTube, Netflix, Amazon Prime, news, and reading books), and information and tasks (internet banking, postal services, maps, monitoring electricity usage, booking appointments with doctors, and managing pension)
4	Male	78	4 years of secondary school	Smartphone, tablet, smart TV, and internet access	• Communication (calls and email [can only read, not respond], video calls, and unable to use SMS), entertainment (Netflix and YouTube), information and tasks (booking appointments with doctors), and social media (Facebook but rarely posts)
5	Female	83	Vocational training	Smartphone, tablet, computer, smart TV, and internet access	• Communication (calls and FaceTime), entertainment (Foxtel, Netflix, Disney+, ABC, and Solitaire, news), and information and tasks (Google)
6	Male	73	University degree	Smartphone, tablet, computers (1 running macOS and 1 running Windows), smart TV, and internet access	• Communication (calls, SMS, WhatsApp, and email, video calls), information and tasks (online courses, Google, Calendar, and Photoshop), entertainment (YouTube, Netflix, news), and social media (Facebook and Twitter: reads only, does not post) • Has difficulty using smartphone given size of text
7	Female	70	2 years of secondary school	Smartphone, computer, smart TV, smartwatch, and internet access	• Communication (calls and SMS), entertainment (games), information and tasks (weather, calendar, Google, monitors steps and heart rate on watch, taking photos, and asks questions using Alexa or Siri), and social media (Facebook) • Unable to use laptop, email, or phone banking
8	Male	86	4 years of secondary school	Smartphone, tablet, smart TV, and internet access	• Communication (calls only), information and tasks (Google), and entertainment (Netflix and YouTube) • Has an email account but managed by daughter • Has access to social media (Facebook) but does not use it • Does not use tablet
9	Female	75	Did not complete primary school	Smartphone, tablet, computer, smart TV, and internet access	• Communication (calls, learning to use SMS, WhatsApp, and email [view only]), information and tasks (takes photos), and social media (Facebook but is limited in what she can do) • Does not use the computer in the house, and limited use of tablet
10	Female	79	2 years of secondary school	Smartphone, computer, no internet at home, and no tablet or smart TV	• Communication (calls and SMS) and information and tasks (taking and sharing photos) • Does not use laptop
11	Female	87	Finished secondary school	Smartphone, tablet, computer, smart TV, smartwatch, internet access, and smart speakers	• Communication (calls, SMS, and email), entertainment (Netflix, Disney+, listening to music on smart speakers, and news), and information and tasks (Google for health information)

XSL•FO
RenderX

Technology Ownership and Use

Results are summarized in Multimedia Appendix 2, including illustrative participant quotes.

Technology Ownership

Participants reported owning 5 different types of technology (range 2-7) on average. Eleven participants owned a smartphone, 9 owned a tablet, 9 owned a laptop or desktop computer, 9 had a smart TV, 3 had a smartwatch, 1 had a smart speaker, and 1 had an electronic book (e-book) reader. Levels of internet access were also high (91%); only 1 participant reported no internet access at home (Table 1).

Technology Use

Collectively, participants reported using technology for a range of reasons, including to access general media, such as news or current affairs programs (n=4); social media (n=7); video entertainment, such as television shows, movies, or YouTube videos (n=7); online games (n=3); and for communication purposes, for example, sending and receiving emails (n=5, though with varying capabilities), SMS text messaging (n=8), and video calls (n=5; Table 1).

Participants commonly streamed television and movies on their smart TV (n=7), but typically only after being set up by family members:

> *My daughter and son bought me a smart TV and they organized for it all to be hooked up. I wouldn't have a clue that they did, but I've watched a lot of Netflix.* [Participant 4, male, 78 years]

The social media platform Facebook (Meta Platforms, Inc) was also commonly viewed (n=7) but participants typically did not post themselves or provide comments:

> *I look at Facebook. I don't know how to get it, to get in to put anything on it...but I do look at it. Yeah, you know, what everybody's saying or looking at, you know.* [Participant 9, female, 75 years]

While 6 care recipients reported using technology to seek information from the internet, only 4 reported using technology to manage physical health issues. One person used a smartwatch to monitor steps taken and heart rate, 1 participant described searching for health advice online, and 2 reported using technologies to manage clinician appointments via app or SMS text messages.

The use of devices varied markedly between participants. Some (n=3) accessed a broad range of digital functionality across multiple devices to support their daily activities, including managing email, calendars and banking, and engaging in online forums. Within this group, 2 participants operated digital devices using a voice-activated digital assistant such as Alexa or Siri, and 1 participant used a smartwatch to take calls and messages:

> *[Smartwatch] has got everything. Yeah, it's got the lot, and you can receive messages, and I can also talk on it too. If I'm out mowing or something, you know?* [Participant 11, female, 87 years]

Remaining participants (n=8) made more limited use of the potential functionality offered by their devices. Many only used their smartphone for calls and/or SMS text messaging (n=6), and in some cases, to play simple games (n=3) or take photographs (n=3):

> *I only ring my family, my husband. He rings me. And well, I send messages to the family. That's all I use it [mobile phone] for. And if I get a photo of it, take a photo of the garden, I'll send it to them and stuff, but that's it.* [Participant 10, female, 79 years]

Several participants (n=5) described difficulty using core functions, such as email and SMS text messaging:

> *If I get an email, I don't know how to respond to it...I don't know how to send a text message, but yeah, I just, I would rather speak to somebody* [Participant 4, male, 78 years]

In addition, a small number of participants (n=4) reported they did not use at least one of the devices (smartphone, tablet, or computer) in their home at all, which in some cases (n=3) had been provided by their family members:

> *I don't even want to know how to use things on my phone. Not interested... I just say I'm just too old to start learning things.* [Participant 2, female, 79 years]

Support in Using Technology

While participants described a wide range of technology use, most participants (n=8) reported receiving at least some support to use technology, typically from a close family member. In some cases (n=3), adult children appeared to routinely take charge of tasks managed through technology. This often involved sharing sensitive or personal information, such as passwords or banking information:

> *The daughter has [my banking information] on her computer because, um, like I got...a message I guess it was saying that the registration was due on the car and so I took it around and she said "there you are, done and paid for" because she can have access to my bank account.* [Participant 4, male, 78 years]

However, one participant indicated that they only turned to family members if they could not resolve the issue themselves:

> *Oh, I'd try to [do it myself], if it was something I absolutely couldn't do...So yes, I did go to [my son-in-law] because I couldn't work out what was going on.* [Participant 11, female, 87 years]

While adult children were often seen as a primary source of information and training in using technology, in some cases (n=2), participants' continued use of technologies required ongoing support that did not eventuate, limiting participants' ability to make full use of the technologies:

> *I need to be shown more than once, that's my problem...I really just want to be able to say, "Yes, I can use the computer," yeah. And you know, years ago, the kids taught me, and I was doing bits and pieces, but now it's just got photos on it.* [Participant 7, female, 70 years]

No participants reported accessing technology support available from technology companies (eg, Apple and Microsoft) or

XSL•FO
RenderX

independent organizations that provide technology support. One participant reported receiving support available in their local community to use technology:

> *When we first moved into the village here, they had a woman come to show us different things. That lasted for about four weeks. Once a week she would come in and show us how to use our laptop, so that's good.* [Participant 5, female, 83 years]

There were, however, some participants (n=4) who were confident in using a range of devices, and described positive attitudes towards learning new technologies, and detailed technological knowledge:

> *Yes, I'm using two computers. One is Macintosh, one is Windows...because Apple doesn't have all the programs.* [Participant 6, male, 73 years]

> *It's just interesting, it's a new challenge. Something nice, new to learn, yeah.* [Participant 11, female, 87 years]

Barriers to Technology Use

Despite technology being common in their homes, participants reported a range of barriers to technology use. Approximately half the participants (n=5) expressed concerns around privacy and scams. For example, this participant discussed low trust in using their smartphone and the high volume of spam calls they were experiencing:

> *Trusting...no! It's no secret anyone can get our data...at least this is the nuisance from our telephone, coming about at least six or seven calls coming or spam, that is a headache.* [Participant 6, male, 73 years]

Cost was identified as a barrier to use of some technologies for 2 participants. One participant on an aged pension described having to choose between having internet access at home and other priorities, leaving them without core technological infrastructure:

> *I can't afford to live and save up a bit of money to go on a holiday and have those sort of things [internet access], as well on an age pension.* [Participant 10, female, 79 years]

Other participants (n=2) raised concerns about breaking the technology or engaging with it in a way that could cause a negative outcome, which limited the technologies they chose to use and how they used them:

> *I will never do phone banking [on an app], okay? Because I'm worried that I might press the wrong button.* [Participant 7, female, 70 years]

> *If I have to learn something new...I just panic if I'm gonna do the wrong thing, you know? And I think and then I'm going to be in big trouble...And that's my fear of it, you know.* [Participant 9, female, 75 years]

Several participants (n=3) reported the need to make minor modifications to their devices, such as increasing font size, to reduce usability barriers:

> *Yes (I use it) but I don't like the phone because it's too small. I can't read and every, every time I have to make [the font] bigger, bigger and read, so it's big.* [Participant 6, male, 73 years]

Support for Emotional Well-Being

Results are summarized in Multimedia Appendix 2, including illustrative participant quotes.

Current Sources of Support

Participants were asked to describe any sources of support they engaged in for their emotional well-being. Only 1 participant reported not receiving emotional support from any source. Seven of the 11 participants reported they received emotional support from family members (in particular, their daughter), which was most commonly provided in person or over the phone:

> *My daughter rings every day. She's just rung about half an hour ago.* [Participant 5, female, 83 years]

> *You just get lonely by yourself. And we just sort of support each other that way, you know...I've got 4 kids. I know I've always got them. But if I've got something on my mind, I'll sort of talk it over with [my daughter], you know?* [Participant 2, female, 79 years]

Six participants reported receiving emotional support from their broader community, although the nature of this support differed, including informal support from friends and neighbors and formal support from a community center. For example, this participant described receiving support from others in their retirement village who had experience with grief:

> *A lot of people in the village here have all that [grief]...all gone through it, and we talk a lot to people there, about [grief and loss]* [Participant 3, male, 83 years]

In contrast, only 1 participant reported receiving current professional support for their emotional well-being, which was organized by her general practitioner (GP). Another participant had also accessed a counselor but several years ago:

> *I've got another 4 visits, but it's up to the doctor, I think to ask for more if I need it. As soon as I mentioned it, he was straight on [organized a referral]. I had [a phone call] the next day.* [Participant 7, female, 70 years]

> *I used to [be depressed], but not anymore...oh, going back how many years? 1995. So, it's been a lot of years and I, I spoke to a counselor and that and yeah, I'm fine.* [Participant 10, female, 79 years]

Barriers to Seeking Support

Remaining participants gave reasons for not seeking professional help. This was commonly expressed in terms of not wanting "to be a burden on anybody" (Participant 4, male, 78 years) or preferring to discuss personal details with their family:

> *I like to talk about myself with my family members, not others.* [Participant 6, male, 73 years]

XSL•FO
RenderX

One participant described difficulty accessing the support he wanted, citing issues with waiting times as a critical reason for discontinuing his search for support:

> *I must admit everything seems to just take so long, and, and you get to the point where you just say "Oh, forget it!" And you don't carry on with it.* [Participant 8, male, 86 years]

Willingness to Engage in Support

Participants were asked if they would be willing to engage in professional help to support their well-being in the future, if available. The majority (n=10) indicated they were open to receiving support for their emotional well-being. However, participants commonly expressed initial reticence before identifying circumstances in which support could be useful:

> *No, I say no...But you never know because a friend of mine committed suicide. So, I say, you know, at the moment, no, I wouldn't use it, but you never know, I mean, probably it's a shame that he didn't speak to somebody...Well, maybe, yeah, maybe in time. Yeah, I would say maybe probably once I give up driving. I think then yes maybe.* [Participant 4, male, 78 years]

> *There are times when you feel that you would probably like to talk to someone else, other than family...It's never struck me as being absolutely necessary, but I think there are times when it'd be nice if it was there.* [Participant 8, male, 86 years]

While the majority (n=10) spoke about support in terms of speaking with a professional, 1 participant noted a preference for peer support for their emotional well-being:

> *I think it'd be the fact that I can talk to somebody else, a stranger who doesn't know the history, but has maybe gone through something like that themselves, or knows of somebody. Umm, but just to get some feedback.* [Participant 7, female, 70 years]

Only 1 participant stated they would not consider engaging in professional emotional support:

> *No, no. Strangers. They'd be strangers. People I didn't know that would be talking, probably from rote. No, no, it wouldn't, wouldn't mean a thing to me.* [Participant 11, female, 87 years]

Use of Digital Technologies to Support Emotional Well-Being

Results are summarized in Multimedia Appendix 2, including illustrative participant quotes.

Current Use of Technology to Support Emotional Well-Being

Participants were asked to share their current experiences using technology to support or enhance their well-being. Seven participants reported diverse experiences, including engaging with music, hobbies such as games, quizzes, or reading, and remote religious support:

> *[I] read a lot, I've got a little Kindle [eReader], and I sit at the computer for half an hour and just see what's happening in the world.* [Participant 3, male, 83 years]

> *On my phone, I have apps that I can go to and bring up a message from [church group].* [Participant 1, female, 78 years]

Barriers to Accessing Digital Emotional Well-Being Support

Participants identified several barriers to use of emotional support delivered online, including barriers to using technology in general, such as poor motivation to engage online and difficulties with hearing, vision, or compromised physical ability:

> *If you ask me if I want online or face to face, I will definitely say face to face ... online, it will be there in my computer, I may not open it...it is easy to put things off if they are online.* [Participant 6, male, 73 years]

> *...with the arthritis, now I don't have the strength in my hand...I use the left hand just to press the buttons.* [Participant 4, male, 78 years]

Two participants shared a perspective that they would require training and assistance to successfully access online emotional support:

> *If there was an app that I could put on my phone, then I probably would use it, as long as I knew how to use it. And if I can take it in, but then I need to be shown more than once. That's my problem. Yeah, it's learning how to use the damn thing.* [Participant 7, female, 70 years]

> *It would need a lot of teaching. I would need a lot of support from my son and my daughter.* [Participant 4, male, 78 years]

Factors such as relevance, trust, and credibility were also identified by participants as being important for uptake in this cohort:

> *It has to have a purpose that can be seen and understood.* [Participant 4, male, age 78]

> *Yeah, has to be trust around it. So I don't. I don't know whether the government can do that ... evaluate those things [digital tools] and give a tick* [Participant 6, male, 73 years]

Support Preferences

Participants were also asked their views on receiving online emotional support. The majority (n=7) of participants expressed a preference for face-to-face support, which was seen as "more personal" (Participant 6, male, 73 years). Despite this, many (n=6) were open to videoconferencing and other digital approaches. Participants expressed that digital approaches could make it easier to access someone to talk to when needed and could also help by providing them with guidance in managing their well-being:

> *Actually, it's very good, very good, because patients like us need some help...I can walk and I can do some things, but there are some people that can't do*

XSL•FO
RenderX

> *anything, so it's better. It's better to have [digital] services...Definitely convenient to people.* [Participant 6, male, 73 years]
>
> *...To get some ideas of what I can do... to use it and talk to someone, I suppose.* [Participant 5, female, 83 years]
>
> *Well, I guess for people, if they're feeling a bit down and that...yeah, that they can, you know, look at or to support them. I mean there must be a lot of lonely people out in the world.* [Participant 4, male, 78 years]

Online emotional support was also seen as potentially appealing to people who were uncomfortable with face-to-face support, as some "don't like face-to-face" [Participant 3, male, 83 years]. Perceived benefits of digital approaches included providing a sense of anonymity and making it easier to establish rapport or access relevant lived or professional experience:

> *Maybe if it was online and I don't see who it is, maybe you could express yourself better. Sometimes I can see people I don't like. I don't feel comfortable with them... you're not going to express yourself if you don't feel comfortable with somebody.* [Participant 2, female, 79 years]

There were mixed views on accessing virtual emotional support provided through their aged care provider, whether using online programs or apps or support via telehealth. Four participants responded they would welcome this, while 5 reported they might be open to receiving this support from their provider:

> *Yeah, definitely. I will use it if I get some service from [aged care provider] or from somebody else I will use it...definitely, I will use it.* [Participant 6, male, 73 years]
>
> *If you can just say "I feel depressed," I can log into [service provider] and say, "I'm a bit down. Can I talk to somebody about something?"* [Participant 3, male, 83 years]

In contrast, 2 care recipients responded negatively to this proposal, with one person perceiving their care provider as providing physical support only, while the other did not welcome emotional support that was provided outside of their family.

Discussion

Principal Findings

This study extends existing literature on the use of digital technologies in later life by focusing on a clinically complex population that has been largely overlooked by researchers but is growing rapidly. Aged in their 70s and 80s, this in-home aged care sample was considerably older than participants in much of the research previously published on this topic, with many researchers defining "older adults" as young as aged 50 or 55 years plus [19,24-26]. Participants in this study generally had limited postschool education, and most reported they were unable to do things others their age could do. Previous research has demonstrated that advanced older age [27-30], poorer health [29,30], and physical disability [28]—all of which are ubiquitous in the home care setting—have a negative impact on use of technology in daily life.

Our findings indicated that while participants generally had good access to digital devices and internet at home, access did not equate to independent or confident use. Technology use varied considerably: some participants used multiple devices for a broad range of activities and functions throughout their daily lives, while others used devices only for basic communication or were unable to use available devices or functions without support. Aged care providers should routinely assess not only device access, but also digital confidence, skills, support needs, and the availability of informal assistance, before introducing digital tools into care.

Most participants used their devices for communication, entertainment, social media, or to access information, while only a minority described using technology to manage their physical health, for example, to monitor their health or to seek health-related information. Participants did not use purpose-designed digital mental health or well-being tools and appeared largely unaware of them, despite their widespread availability in the general population [31]. However, many used technology for hobbies, social connection, music, reading, games, and religious content, which may indirectly support well-being [24]. This suggests that digital well-being tools may be more acceptable if embedded into familiar activities and existing care interactions, rather than presented as stand-alone programs. Aged care providers could address the gap in awareness by recommending apps aligned with their clients' interests that have a strong evidence base, and by incorporating brief digital check-ins via telehealth.

Despite being identified by their aged care provider as likely experiencing poor emotional well-being, only 1 participant was receiving current professional support, although another had previously received counseling, and a further participant had attempted to access support. Most participants were open to receiving professional support, particularly when framed around having someone to talk to, rather than as mental health treatment. This aligns with previous research describing the reticence of in-home aged care recipients to seek mental health care [32] while also suggesting that many accept formal support at key points in their lives if presented in a nonstigmatizing and personally relevant way. Service providers should frame supportive services in terms of helping the person maintain their well-being, engage in valued activities and interests, or implement lifestyle changes.

Most participants expressed a preference for face-to-face or hybrid emotional support rather than fully online, consistent with previous research [19,32]. However, several were open to using telehealth to access mental health specialists, counselors, or people with lived experience, and identified convenience and anonymity, which are important enablers for digital mental health interventions [19]. Participants also recognized the potential for telehealth approaches to address barriers to accessing care, such as limited mobility. Their perceptions may have been shaped by strong current reliance on informal sources of emotional support. Ready access to family or community support may reduce perceived need for digital alternatives and

XSL•FO
RenderX

reinforce preferences for personal, face-to-face interactions. This may help explain participants' preference for in-person support, alongside their openness to digital tools when positioned as complementary to, rather than replacing, existing support networks.

Digital emotional well-being approaches must account for both general barriers to accessing mental health support, such as negative attitudes, low awareness, availability, and accessibility of services [21], as well as digital-specific barriers. In this study, participants identified challenges related to sensory impairments, reduced physical dexterity, low confidence, privacy and scam concerns, and the need for repeated practical instruction. These findings are consistent with previous research showing that older people may require accessible, reliable, and easy-to-use technologies, supported by clear instructions [11,19]. The current findings indicate that training for home care recipients should be ongoing, rather than one-off, with opportunities for repetition and practical demonstration. Approaches such as in-home coaching, simple step-by-step guides, and embedding support within routine care interactions may help to build confidence and sustain engagement over time.

Trusted health care providers may be well placed to introduce suitable digital well-being tools, either by sharing information about available options or embedding selected tools into service delivery [19,26]. Some participants were enthusiastic about the potential to receive emotional support from their aged care provider, while the majority indicated they might be willing to consider this approach, consistent with other recent research [32]. Future research should identify which types of digital supports are most acceptable and feasible in home-based aged care. While many digital mental health offerings focus on managing symptoms of depression and anxiety, in these interviews, care recipients described a broad range of activities that they saw as helpful for their emotional well-being, which offers the possibility of extending the types of interventions for greater engagement and uptake.

Implementation should address older people's concerns about making a mistake or damaging technology, privacy, scams, and the credibility of digital tools, all of which may impact their willingness to engage [10,11,20,26]. Most participants in this study required at least some support in using digital technologies at home, typically provided by their adult children. While some only sought assistance after attempting to resolve technology problems independently, family members routinely completed digital tasks on behalf of other older people. Although family support may often be appropriate during implementation [11,33], reliance on family alone may reinforce dependence and limit autonomy. Encouragingly, even among care recipients who reported poor skills in using technology, many were willing to learn, consistent with previous research [33]. Proactive, no-cost support is likely important for successful implementation, given this sample's poor uptake of existing technological support services available in the community and previous reports of limited use of technology helpdesks [34]. Roles such as digital navigators, technology mediation workers or facilitators within aged care services could support older people's adoption and use of technology [25,26]. Policymakers should ensure that aged care funding enables older people to access not only apps, devices, and internet services where required, but also the practical support needed to use them. Greater coordination between aged care providers, health services, and community organizations (eg, community-based digital literacy programs) may also help ensure older adults can access consistent and trusted sources of assistance.

The findings can be understood through the lens of the Unified Theory of Acceptance and Use of Technology (UTAUT) [35], which proposes that uptake is shaped by performance expectancy (perceived usefulness), effort expectancy (perceived ease of use), social influence, and facilitating conditions. Participants' emphasis on technologies needing a clear and meaningful purpose aligns with performance expectancy, with care recipients being more open to using digital tools when they could see how these would support their daily lives or well-being, or help them engage in valued activities. However, low confidence, fear of making mistakes, and the need for repeated instruction reflect low effort expectancy, highlighting perceived complexity as a key barrier.

Social influence played a central role in shaping technology use, with family members, particularly adult children, playing a central role in introducing and supporting technology use, although in some cases they completed digital tasks on behalf of participants. Technology adoption in this population is not solely an individual process but is strongly embedded within social relationships, which may both facilitate and inadvertently limit independent use. Facilitating conditions were often limited, with most participants relying primarily on informal support networks and reporting limited engagement with structured sources of technological assistance for training and ongoing assistance.

Collectively, applying UTAUT highlights that limited uptake of digital technologies to support emotional well-being in home-based aged care is not primarily driven by lack of interest, but by a combination of modifiable factors related to usability, support, and perceived value, as well as lack of awareness of relevant tools. Technology adoption in this population is strongly shaped by reliance on informal carers and the need for ongoing, hands-on support, rather than access alone. Even when interest is high, low effort expectancy and limited support can constrain engagement. Together, this suggests that improving uptake of digital well-being interventions in this population will require not only demonstrating value but also reducing complexity and embedding sustained support within service models. This study extends existing literature on UTAUT by demonstrating how these constructs operate in a clinically complex, home-based aged care population, where reliance on informal carers and physical limitations may amplify effort expectations and limited access to resources.

Limitations

While this is the first in-depth exploration of the views of in-home aged care recipients in using technology to support their emotional well-being, there are several limitations of this study, most notably the small size of the sample and the potential for sampling bias; those with very low interest in using technology and those uncomfortable discussing emotional well-being may have chosen not to participate in the study. In

XSL•FO
RenderX

addition, the study recruited participants considered by their care workers to have poor emotional well-being, given the broad focus of the research. Some of the participants may have experienced symptoms of depression, anxiety, or other mental health conditions, but their mental health status was not formally assessed, and different results may have been obtained with a sample with confirmed mental health conditions. It is also possible, given the absence of formal criteria provided to referring staff, that some participants did not experience poor emotional well-being at the time of the interview; however, all participants referred to either current or previous periods where they struggled with low mood or aspects of poor well-being during interviews. Aged care recipients from regional or rural areas, those who were non–English-speaking, and those with significant cognitive impairment were not included in this study; additional research is required to understand the perspectives of these populations. Similarly, the research was conducted with recipients of a single aged care provider, and the results may not generalize to the broader population of older people who receive aged care at home. Three participants had a family carer present during their interview, which may have influenced how they responded to interview questions. Finally, the interviews were conducted from October to December 2023 and explored technologies readily available at that time. Novel AI-based tools to support emotional well-being, such as chatbots, were not discussed in interviews. With technology evolving rapidly, research findings should be updated regularly.

Conclusions

Despite the limitations, this study highlights an openness to both formal and informal support for well-being among in-home aged care recipients. Digital technologies could play an important role in supporting a population facing increasing mental health rates [36] and poor access to professional mental health services [7]. As well as increasing access to support, digital technologies have the potential to allow older people to self-manage their well-being. However, our findings suggest that access to everyday digital devices does not necessarily translate into awareness or use of purpose-designed digital well-being tools among frail and clinically complex older people living at home. Aged care providers could play a valuable role in introducing suitable programs to older people, leveraging established trusting relationships and should consider integrating digital tools into existing service delivery to support their clients' well-being [11].

When considering digital initiatives, aged care providers should prioritize those that can be delivered using the devices older people already access in their homes, particularly smartphones, tablets, and smart TVs. Importantly, many care recipients will require significant and ongoing technology training and support in using new technologies, even those accessed via their existing devices, which requires appropriate resourcing [11]. This study identified several barriers that should be addressed throughout implementation and highlighted the need for a highly tailored implementation approach, given the heterogeneity in care recipients' digital literacy and experience using technology—one size will not fit all. The routine use of brief assessments of care recipients' technological readiness has been proposed previously for the aged setting [10]. It may also be beneficial to provide education to aged care workers [10], some of whom may hold negative perceptions about the use of technology in aged care [11,25] and may themselves have low digital readiness [10,11,37], which could impact their willingness to offer these solutions.

Acknowledgments

The authors would like to thank those who participated in this research and the aged care workers who assisted with recruitment.

Funding

This research was funded by Aged Care Research and Industry Innovation Australia (ARIIA). The funder did not take part in planning, conducting, or reporting this research.

Data Availability

The qualitative data generated and analyzed during this study are not publicly available because participants did not consent to the sharing of their interview data. Relevant supporting information is provided within the article and multimedia appendices.

Generative AI Use

The original version of the manuscript was written without the use of AI. During manuscript revision, the authors used Microsoft Copilot to assist with language editing and improve clarity. The authors reviewed and approved all suggested changes and take full responsibility for the accuracy, originality, and integrity of all content in the manuscript.

Authors' Contributions

Conceptualization: TED, MD, SP, TAN, JK, SB
Data curation: MD, HD, DC, TDD
Formal analysis: HD, MD, DC, TDD
Funding acquisition: TED, MD, SP, TAN, JK, SB

XSL•FO
RenderX

Investigation: HD, DC, TDD
Methodology: TED, SB, MD, HD
Project administration: MD
Supervision: TED, MD, SB
Writing – original draft: TED, MD, HD
Writing – review & editing: SP, TAN, JK, DC, TDD, KS, SB

Conflicts of Interest

TED, MD, DC, and KS were employed by Silverchain, the aged care provider through which participants were recruited, and JK received a PhD scholarship funded by Silverchain. Silverchain staff assisted with identifying and initially approaching potential participants; however, participation was voluntary, and Silverchain had no role in determining the study findings or the decision to publish. The authors declare no other conflicts of interest.

Multimedia Appendix 1
Interview schedule.
[DOCX File , 33 KB - humanfactors_v13i1e93823_app1.docx]

Multimedia Appendix 2
Results summary table.
[DOCX File , 42 KB - humanfactors_v13i1e93823_app2.docx]

References

1. Ageing and health. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2026-08-14]
2. Boland L, Légaré F, Perez MMB, Menear M, Garvelink MM, McIsaac DI, et al. Impact of home care versus alternative locations of care on elder health outcomes: an overview of systematic reviews. BMC Geriatr 2017;17(1):20 [FREE Full text] [doi: 10.1186/s12877-016-0395-y] [Medline: 28088166]
3. Admissions into aged care. Australian Institute of Health and Welfare. URL: https://www.gen-agedcaredata.gov.au/topics/admissions-into-aged-care [accessed 2026-10-12]
4. Dickins M, Kelly J, Paxton S, Kwan B, Carveth D, Barker A, et al. The prevalence and correlates of depression and anxiety symptoms in older adults receiving in-home aged care: a cross-sectional survey. Australas J Ageing 2025;44(1):e13396. [doi: 10.1111/ajag.13396] [Medline: 39940120]
5. Wang J, Shang J, Kearney JA. Mental health disorders in home care elders: an integrative review. Geriatr Nurs 2016;37(1):44-60. [doi: 10.1016/j.gerinurse.2015.10.013] [Medline: 26607689]
6. Inacio MC, Lang C, Bray SCE, Visvanathan R, Whitehead C, Griffith EC, et al. Health status and healthcare trends of individuals accessing Australian aged care programmes over a decade: the registry of senior Australians historical cohort. Intern Med J 2021;51(5):712-724 [FREE Full text] [doi: 10.1111/imj.14871] [Medline: 32359019]
7. Thapaliya K, Caughey GE, Crotty M, Williams H, Wesselingh SL, Roder D, et al. Primary, allied health, selected specialists, and mental health service utilisation by home care recipients in Australia before and after accessing the care, 2017-2019. Aging Clin Exp Res 2024;36(1):83 [FREE Full text] [doi: 10.1007/s40520-024-02731-9] [Medline: 38551712]
8. Duty of care: aged care sector in crisis. Committee for Economic Development of Australia. URL: https://cedakenticomedia.blob.core.windows.net/cedamediatest/kentico/media/attachments/duty-of-care-aged-care-sector-in-crisis.pdf [accessed 2025-02-10]
9. National Mental Health Workforce Strategy 2022-2032. Department of Health and Aged Care. 2022. URL: https://www.health.gov.au/sites/default/files/2023-10/national-mental-health-workforce-strategy-2022-2032.pdf [accessed 2025-02-10]
10. Grey E, Baber F, Corbett E, Ellis D, Gillison F, Barnett J. The use of technology to address loneliness and social isolation among older adults: the role of social care providers. BMC Public Health 2024;24(1):108 [FREE Full text] [doi: 10.1186/s12889-023-17386-w] [Medline: 38184519]
11. Peng R, Li X, Guo Y, Ning H, Huang J, Jiang D, et al. Barriers and facilitators to acceptance and implementation of eMental-health intervention among older adults: a qualitative systematic review. Digit Health 2024;10:20552076241234628 [FREE Full text] [doi: 10.1177/20552076241234628] [Medline: 38444518]
12. Pineda BS, Mejia R, Qin Y, Martinez J, Delgadillo LG, Muñoz RF. Updated taxonomy of digital mental health interventions: a conceptual framework. Mhealth 2023;9:28 [FREE Full text] [doi: 10.21037/mhealth-23-6] [Medline: 37492117]
13. Park CL, Kubzansky LD, Chafouleas SM, Davidson RJ, Keltner D, Parsafar P, et al. Emotional well-being: what it is and why it matters. Affect Sci 2023;4(1):10-20 [FREE Full text] [doi: 10.1007/s42761-022-00163-0] [Medline: 37070009]
14. Lattie EG, Stiles-Shields C, Graham AK. An overview of and recommendations for more accessible digital mental health services. Nat Rev Psychol 2022;1(2):87-100 [FREE Full text] [doi: 10.1038/s44159-021-00003-1] [Medline: 38515434]

XSL•FO
RenderX

15. Xiang X, Kayser J, Turner S, Ash S, Himle JA. Layperson-supported, web-delivered cognitive behavioral therapy for depression in older adults: randomized controlled trial. J Med Internet Res 2024;26:e53001 [FREE Full text] [doi: 10.2196/53001] [Medline: 38437013]
16. Goodarzi Z, Holroyd-Leduc J, Seitz D, Ismail Z, Kirkham J, Wu P, et al. Efficacy of virtual interventions for reducing symptoms of depression in community-dwelling older adults: a systematic review. Int Psychogeriatr 2023;35(3):131-141 [FREE Full text] [doi: 10.1017/S1041610222000412] [Medline: 35603891]
17. Chua CMS, Chua JYX, Shorey S. Effectiveness of home-based interventions in improving loneliness and social connectedness among older adults: a systematic review and meta-analysis. Aging Ment Health 2024;28(1):1-10. [doi: 10.1080/13607863.2023.2237919] [Medline: 37466183]
18. Ibarra F, Baez M, Cernuzzi L, Casati F. A systematic review on technology-supported interventions to improve old-age social wellbeing: loneliness, social isolation, and connectedness. J Healthc Eng 2020;2020:2036842 [FREE Full text] [doi: 10.1155/2020/2036842] [Medline: 32765823]
19. Yin R, Martinengo L, Smith HE, Subramaniam M, Griva K, Tudor Car L. The views and experiences of older adults regarding digital mental health interventions: a systematic review of qualitative studies. Lancet Healthy Longev 2024;5(11):100638 [FREE Full text] [doi: 10.1016/j.lanhl.2024.08.007] [Medline: 39491881]
20. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021;21(1):1556 [FREE Full text] [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]
21. McKittrick R, Orellana L, Manias E, Hensher M, Hutchinson AM. Characteristics, care and support needs of older victorians requiring a government-funded home care package: an observational study. Australas J Ageing 2025;44(1):e13400. [doi: 10.1111/ajag.13400] [Medline: 39799959]
22. Yang R, Gao S, Jiang Y. Digital divide as a determinant of health in the U.S. older adults: prevalence, trends, and risk factors. BMC Geriatr 2024;24(1):1027 [FREE Full text] [doi: 10.1186/s12877-024-05612-y] [Medline: 39709341]
23. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2008;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
24. Andrews JA, Brown LJ, Hawley MS, Astell AJ. Older adults' perspectives on using digital technology to maintain good mental health: interactive group study. J Med Internet Res 2019;21(2):e11694 [FREE Full text] [doi: 10.2196/11694] [Medline: 30758292]
25. Elavsky S, Knapova L, Janiš K, Cimler R, Kuhnova J, Cernicky T. Multiple perspectives on the adoption of SMART technologies for improving care of older people: mixed methods study. J Med Internet Res 2024;26:e45492 [FREE Full text] [doi: 10.2196/45492] [Medline: 38324345]
26. LaMonica HM, Davenport TA, Roberts AE, Hickie IB. Understanding technology preferences and requirements for health information technologies designed to improve and maintain the mental health and well-being of older adults: participatory design study. JMIR Aging 2021;4(1):e21461 [FREE Full text] [doi: 10.2196/21461] [Medline: 33404509]
27. Ahlawat H, Darcovich A, Dewhurst M, Feehan E, Hediger V, Maud M. Age is just a number: how older adults view healthy aging. McKinsey Health Institute. URL: https://www.mckinsey.com/mhi/our-insights/age-is-just-a-number-how-older-adults-view-healthy-aging [accessed 2025-12-16]
28. Gell NM, Rosenberg DE, Demiris G, LaCroix AZ, Patel KV. Patterns of technology use among older adults with and without disabilities. Gerontologist 2015;55(3):412-421 [FREE Full text] [doi: 10.1093/geront/gnt166] [Medline: 24379019]
29. Heart T, Kalderon E. Older adults: are they ready to adopt health-related ICT? Int J Med Inform 2013;82(11):e209-e231. [doi: 10.1016/j.ijmedinf.2011.03.002] [Medline: 21481631]
30. Savage RD, Di Nicolo S, Wu W, Li J, Lawson A, Grieve J, et al. The factors associated with nonuse of social media or video communications to connect with friends and family during the COVID-19 pandemic in older adults: web-based survey study. JMIR Aging 2022;5(2):e34793 [FREE Full text] [doi: 10.2196/34793] [Medline: 35344502]
31. Anser MK, Nabi AA, Ahmad I, Abro MMQ, Zaman K. Advancing mental health care: a comprehensive review of digital tools and technologies for enhancing diagnosis, treatment, and wellness. Health Care Sci 2025;4(3):163-178. [doi: 10.1002/hcs2.70018] [Medline: 40568629]
32. Kelly J, O'Callaghan C, Dickins M, Davison T, Schofield P, Bhar S. "A Cup of Tea and a Chat": a qualitative study on the mental healthcare preferences of Australian in‐home aged care recipients and their experiences of accessing mental health services. Health Soc Care Community 2025;2025(1):1-13. [doi: 10.1155/hsc/3221914]
33. Watt J, Fahim C, Straus S, Goodarzi Z. Barriers and facilitators to virtual care in a geriatric medicine clinic: a semi-structured interview study of patient, caregiver and healthcare provider perspectives. Age Ageing 2022;51(1):afab218 [FREE Full text] [doi: 10.1093/ageing/afab218] [Medline: 34794177]
34. Neal DP, Kuiper L, Pistone D, Osinga C, Nijland S, Ettema T, et al. FindMyApps eHealth intervention improves quality, not quantity, of home tablet use by people with dementia. Front Med (Lausanne) 2023;10:1152077 [FREE Full text] [doi: 10.3389/fmed.2023.1152077] [Medline: 37324141]
35. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425-478. [doi: 10.2307/30036540]

XSL•FO
RenderX

36. Dickins M, Davison TE, Smith K. Mental health in aged care: a dramatic increase in absolute numbers of Australians with a mental health condition accessing home care. Aust N Z J Public Health 2025;49(6):100289 [FREE Full text] [doi: 10.1016/j.anzjph.2025.100289] [Medline: 41223634]
37. Kwan BPM, Dickins M, Batchelor F, Williams S, Hwang K, Fulford K, et al. Digital readiness of the in-home aged care workforce: a latent profile analysis. Australas J Ageing 2025;44(1):e70010. [doi: 10.1111/ajag.70010] [Medline: 40044589]

Abbreviations

e-book: electronic book
GP: general practitioner
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by A Kushniruk; submitted 19.Feb.2026; peer-reviewed by L Bovsh, P Haessner; comments to author 29.May.2026; revised version received 16.Sep.2026; accepted 02.Oct.2026; published 28.Oct.2026.

Please cite as:
Davison TE, Davis H, Dickins M, Pedell S, Nguyen TA, Carveth D, Den Dulk T, Smith K, Kelly J, Bhar S
The Use of Digital Technology to Support the Emotional Well-Being of Older People Who Receive Care at Home From the Perspective of Care Recipients: Qualitative Study
JMIR Hum Factors 2026;13:e93823
URL: https://humanfactors.jmir.org/2026/1/e93823
doi:10.2196/93823
PMID:

©Tanya E Davison, Hilary Davis, Marissa Dickins, Sonja Pedell, Tuan Anh Nguyen, Deborah Carveth, Therese Den Dulk, Karen Smith, Julie Kelly, Sunil Bhar. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 28.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Attitudes Toward Digital Health Technology in Neurological Conditions: Qualitative Study of Patients, Caregivers, and Health Care Professionals

Sarah Fearn[1,2], PhD; John Spreadbury[1,2], PhD; Christopher Kipps[1,3], MBBS, PhD

[1]Faculty of Medicine, University of Southampton, Southampton, United Kingdom

[2]NIHR Applied Research Collaboration Wessex, Southampton, United Kingdom

[3]Wessex Neurological Centre, University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom

Corresponding Author:
Sarah Fearn, PhD
Faculty of Medicine
University of Southampton
Institute for Developmental Sciences (IDS) Building, Tremona Road
Southampton General Hospital
Southampton, SO16 6YD
United Kingdom
Phone: 44 (0)23 8120 5253
Email: S.Fearn@soton.ac.uk

Abstract

Background: Digital health technology is increasingly applied across neurological conditions and is now widely accessible from both individual and formal health care perspectives. Nevertheless, comprehensive evidence on attitudes toward digital health technology in this population remains limited.

Objective: This study aimed to investigate attitudes toward digital health technology from the perspectives of people living with neurological conditions, caregivers, and health care professionals.

Methods: An in-depth qualitative design was used, consisting of semistructured interviews with 35 participants (15 people with neurological conditions, 13 caregivers, 7 health care professionals) across multiple neurological conditions. Inductive thematic analysis, guided by a realist approach, was applied separately to each participant group.

Results: Five themes emerged across all 3 groups: (1) openness to use; (2) digital literacy, experience, and ability to use; (3) hesitancy, uncertainty, and perception of benefit; (4) user age and access to technology; and (5) continued need for access to health care professionals. Participants were generally open to digital health technology but reported important caveats, including variable digital literacy, uncertainty about benefit, concerns about age-related access, and a persistent preference for in-person clinical contact. Three findings were particularly significant. First, most participants across all 3 groups were unaware of what digital health tools were available for their neurological condition. Second, neurological symptoms themselves, including fatigue, cognitive impairment, tremor, and photosensitivity, were identified as barriers to technology use that are distinct from, and additional to, the digital literacy barriers documented in the general population. Third, a cross-group comparison revealed that while attitudes were broadly shared, each group evaluated digital health technology from a different perspective: people living with neurological conditions considered it in relation to personal utility and condition-relevance; caregivers in relation to reliability and the ability to escalate concerns; and health care professionals in relation to evidence base and service capacity.

Conclusions: Most participants were open to the uptake and use of digital health technology. The awareness gap and the neurological symptom barrier represent findings with immediate relevance to practice. Active promotion of available, evidence-based digital health tools directed simultaneously at people living with neurological conditions, caregivers, and health care professionals is a prerequisite for meaningful engagement. Digital health tools should be co-designed to account for the specific symptom profiles of people living with neurological conditions and should complement rather than replace in-person care relationships. Implementation strategies should engage all 3 stakeholder groups simultaneously, as meeting the needs of one does not guarantee meeting the needs of another.

(JMIR Hum Factors 2026;13:e107517) doi:10.2196/107517

XSL•FO
RenderX

KEYWORDS

neurological conditions; nervous system diseases; digital health; telemedicine; attitudes; attitude to health; qualitative research; patients; caregivers; health care professionals; mobile phone

Introduction

Neurological conditions are the leading cause of disability-adjusted life years worldwide, affecting over 3 billion people worldwide and representing a significant and increasing health burden for individuals, families, and health systems [1,2]. Digital health technology has emerged as a promising means of extending, personalizing, and supporting neurological care beyond the clinic. Digital health technology is an umbrella term, encompassing labels such as eHealth, mobile health (mHealth), and digital tools, and refers broadly to the application of digital, mobile, and wireless technologies to health and health care [3].

In the United Kingdom, digital transformation has become a strategic priority. The National Health Service (NHS) 10 Year Health Plan for England (2025) outlined a major shift toward digitally enabled, community-based care delivery, and NHS England has invested significantly in digital transformation [4,5]. Potential benefits of this transformation include reduced care inequalities, earlier identification of those at risk of poorer outcomes, and enhanced condition self-management and personalized care. For health services, the anticipated benefits include improved efficiency, better resource allocation, and more integrated care pathways.

Despite significant policy investment and the proliferation of digital tools, variation in uptake and use remains substantial [6]. Understanding the factors that influence whether digital health technologies are adopted and sustained in routine use is therefore a pressing research priority. Attitudes represent a potentially important influence on uptake and use. For the purposes of this study, attitudes are understood as the views, beliefs, and experiences that shape how individuals respond to and engage with digital health technology. A recent systematic review applying the capability, opportunity, and motivation–behavior (COM-B) model to digital health adoption among older adults found that perceived usefulness, social support, and user-centered design were among the most consistent facilitators, while digital literacy gaps, infrastructural barriers, and privacy concerns were among the most prominent obstacles [6]. How these factors interact in the specific context of neurological conditions, however, remains poorly understood.

In the neurological condition context, digital health technology has been applied across a broad range of conditions, including multiple sclerosis, epilepsy, stroke, Parkinson disease, and headache and migraine [7,8]. Existing research indicates that some may exert a positive influence on areas such as mental health, caregiver burden, fatigue, and self-efficacy [9-13]. A rapid review of patient and public involvement (PPI) in the development of digital health technologies for neurological conditions found that the perspectives of people living with neurological conditions were mostly gathered through one-off consultation and that involvement has been inconsistently reported [14]. The range and accessibility of tools continue to expand, yet relatively few are described in the existing literature, and significant uncertainty persists regarding their real-world uptake, effectiveness, and sustainability [7].

Research specifically examining attitudes toward digital health technology in neurological conditions has grown in recent years, though a comprehensive evidence base remains lacking. Existing studies suggest interest in digital health technology and some indication of its use [15-19], as well as a recognition of its potential to improve understanding, self-management, and communication [15,18-20]. More recently, Reid et al [21] found that people with neurodegenerative conditions and their caregivers were broadly supportive of digital technology for health monitoring, while also identifying the complexity of proposed technology and data security as key barriers. However, research also continues to identify negative experiences and concerns, including anxiety, stigma, and stress [18-20]; data privacy and confidentiality [15,18-20]; and mistrust and skepticism [18-20].

Significant gaps remain in this literature. A comprehensive understanding of attitudes toward digital health technology in neurological conditions is lacking, particularly with respect to the nature and sources of those attitudes, the factors perceived to facilitate or hinder engagement with digital tools, and how the perspectives of health care professionals and caregivers compare with those of people living with neurological conditions. The present study was designed to address these gaps by investigating attitudes toward digital health technology in neurological conditions from the perspectives of people living with neurological conditions, caregivers, and health care professionals.

Methods

Design

This study used an in-depth qualitative design consisting of semistructured interviews to investigate attitudes toward digital health technology in the context of neurological conditions. A convenience sampling approach was adopted, whereby participants, after becoming aware of the study, volunteered to participate. Interviews were conducted individually with each participant, with one exception: in 1 caregiver interview, the participant's spouse, who was living with Parkinson disease, was present and contributed to the discussion. This dyadic interview is acknowledged in the relevant quote attribution, and contributions from both speakers were included in the thematic analysis and treated as data from the caregiver context. For the purposes of this study, the term "attitude" was used to encompass the views, beliefs, and experiences of participants in relation to digital health technology. Data were collected between 2022 and 2023, during and shortly after the COVID-19 pandemic. This study is reported in accordance with the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist. The completed checklist is provided in Multimedia Appendix 1.

XSL•FO
RenderX

Participants

A total of 35 participants, consisting of people living with neurological conditions, caregivers, and health care professionals, were recruited from across England.

A broad range of recruitment strategies were used. The majority of people living with neurological conditions and caregivers were recruited through collaboration with national neurological condition support groups and charities, which disseminated study adverts via their websites, member emails, and newsletters. Additional recruitment occurred via the University Hospital Southampton My Medical Record online personal health record, regional neurological condition and caregiver groups, and personal approaches to individuals known to the research team. Health care professionals were recruited via email approaches. Participants varied in neurological condition and in their experience of digital health technology, ranging from those with little or no prior use to those with substantial experience.

Each participant group was analyzed separately. Saturation for each group was considered to have been reached when later interviews yielded few new codes or insights, and a comprehensive account of participant attitudes could already be established. Given the small size of the health care professional group (n=7) and its limited professional and gender diversity, this saturation claim is more tentative for this group than for the patient and caregiver groups.

Materials

Separate semistructured interview schedules were developed for people living with neurological conditions, caregivers, and health care professionals. Questions were grouped into the following sections: use of digital health technology; opinions of digital health technology in general and in the specific neurological condition context; positive and negative aspects of digital health technology; suggestions for improvement; and, where applicable, experience of the My Medical Record online personal health record, used at University Hospital Southampton.

Each schedule also included hypothetical vignettes comprising short scenarios featuring a fictitious third person in a digital health technology and neurological condition context. These were designed to serve as a common stimulus to facilitate discussion of sensitive topics and to reduce social desirability bias. Vignettes depicted scenarios including the introduction of an online care platform, searching for and sharing health-related information, and computer-based medical assessment and advice.

Sociodemographic and neurological condition information was also collected.

A project steering group and PPI contributors reviewed and piloted the interview schedule questions and hypothetical vignettes, providing feedback that informed revisions before their implementation.

Procedure

Interested participants contacted the research team by email or telephone. A researcher then provided access to a participant information sheet and informed consent form, created and stored electronically in REDCap. Interviews were conducted by telephone, lasting between 46 and 125 (mean duration 60, SD 21) minutes. TTelephone interviews were chosen in preference to in-person or video-based interviews to maximize accessibility for participants, because the study was designed during the COVID-19 pandemic when in-person contact was restricted, and to include those without reliable access to online video technology. Interviews followed the same structure: participants first responded to the semistructured interview schedule; those with experience of My Medical Record were then asked supplementary questions specific to that platform (4 of 15 people living with neurological conditions and 2 of 13 caregivers reported experience of the platform); and finally participants responded to the hypothetical vignettes. Participants were given an opportunity at the end of the interview to add any further information. Interviews were recorded using a digital voice recorder and then transcribed verbatim.

Analysis

Thematic analysis was conducted following the framework of Braun and Clarke [22], guided by a realist epistemological position: participant accounts were treated as broadly reflective of their actual attitudes and experiences, rather than as purely discursive constructions of the interview context. Vignette-elicited responses were not analyzed separately from direct interview responses. They were designed to serve as a common stimulus to facilitate discussion of sensitive topics and were therefore integrated fully into the thematic analysis alongside responses to direct questions. Three separate analyses were conducted, one for each participant group. Initial codes were developed inductively by JS through conducting interviews, listening to recordings, and reading transcripts. JS is a male psychologist with formal training and extensive experience in qualitative methods. He approached the data with an interest in understanding the full range of participant attitudes, including those skeptical of or indifferent to digital health technology, and actively sought divergent and negative cases. To establish an initial coding framework, the first 12 transcripts, comprising 6 from people living with neurological conditions, 4 from caregivers, and 2 from health care professionals, were analyzed by hand. At this stage, SF and CK independently reviewed the same 12 transcripts, and the 3 authors met to discuss and compare emerging themes. JS revised the initial coding in light of these discussions. Following these discussions and revisions, JS analyzed the remaining transcripts electronically using NVivo (Lumivero). Each group's transcripts were coded separately, with a distinct set of codes, categories, and subthemes developed for each group. Overarching themes were then derived independently for each group to capture and represent participant attitudes toward digital health technology. JS shared the coding with SF and CK, and themes were refined through further team discussion to explore both the convergence across groups and the distinct perspectives within each. Throughout the analytical process, JS maintained an audit trail of coding decisions and analytical memos. Negative cases and divergent views were actively considered within each theme. The prior professional familiarity of some members of the research team with health care professional participants was

XSL•FO
RenderX

acknowledged and reflected upon during team discussions to mitigate potential influence on interpretation.

Ethical Considerations

Overview

This study received NHS research ethics approval (research ethics committee reference 22/PR/0457; Integrated Research Application System project ID 310073). The study was not exempt from ethics review, and no exemption policy was applied.

Informed Consent

Interested individuals were given a participant information sheet and an electronic informed consent form via REDCap, and all participants gave informed consent before taking part. Participants were told that participation was voluntary and that they could withdraw at any time without giving a reason, but that any data collected up to that point would be retained. Consent to audio recording was confirmed in the consent form and verbally at the start of each interview. In the one dyadic caregiver interview, the caregiver's spouse, who was present and contributed, also gave informed consent.

Privacy and Confidentiality

Interviews were audio-recorded and transcribed verbatim. Each participant was assigned a unique study identifier, and names and other directly identifying details were removed from the transcripts and the quotations reported. Contact details and consent forms were stored separately from the transcripts on secure, access-restricted University of Southampton systems, accessible only to the research team. Quotations are reported with limited descriptors (participant group, condition, and gender), and participant age is given in the text only where it is relevant to the finding, to reduce the risk of identification.

Compensation

Participants received no financial compensation for taking part.

Results

People Living With Neurological Conditions Group

Participants Living With Neurological Conditions

Fifteen people living with neurological conditions were recruited to the study. The mean age of participants was 54.8 (range 19-93, SD 19.3) years; the group included 7 males and 8 females. Neurological conditions represented were epilepsy (n=3), Parkinson disease (n=3), multiple sclerosis (n=3), motor neuron disease (n=3), Huntington disease (n=2), and migraine (n=1).

Five themes represented the attitudes of people living with neurological conditions toward digital health technology: (1) openness to use; (2) digital literacy, experience, and ability to use; (3) hesitancy, uncertainty, and perception of benefit; (4) user age and access to technology; and (5) continued need for access to health care professionals. Each theme below is illustrated with a representative quotation. Further illustrative quotations for each theme and participant group are provided in Multimedia Appendix 2.

Theme 1: Openness to Use

The majority of people living with neurological conditions expressed a broadly positive view of digital health technology, but openness was consistently conditional rather than unconditional, contingent on the technology being genuinely useful, easy to use, and relevant to their specific condition. Underlying motives included a belief that technology could be useful in some way, the perception that health care professionals can take an interest in the patient's condition through digital health technology, and personal interest in using the technology. Reported advantages included speed and efficiency, improved access to information, medication management support, and reduced need to travel to appointments.

> *The main thing is that it's reducing my fatigue levels because I'm not having to travel, I'm not having to walk into the hospital, sit there, and the stress levels of sitting there – thinking, oh, you know, are they going to be running late… yes, it does help the stress levels.* [Patient, multiple sclerosis, male]

Negative aspects reported included technical problems, difficulty remembering to use the technology, perceived excessive complexity, loss of contact with health care professionals, and concerns about misleading information encountered through online health searches and social media.

Theme 2: Digital Literacy, Experience, and Ability to Use

Digital literacy and prior technology experience were identified as key determinants of uptake, but the data revealed an additional and distinctive dimension: for people living with neurological conditions, the condition itself was a potential barrier to technology use, independent of prior digital experience.

> *I've got photosensitivity with my migraine so I cannot look at a screen. When that photosensitivity improves I can, and I can turn the brightness down, so with that feature of my phone…I can just about manage a short interaction.* [Patient, migraine, female]

> *Until I had a disability, I didn't realize how difficult it can be trying to hold a phone with one hand and shaking, whilst trying to stab at some tiny combo box on a webform.* [Patient, Parkinson disease, male]

Theme 3: Hesitancy, Uncertainty, and Perception of Benefit

People living with neurological conditions indicated some hesitancy toward digital health technology. However, the most prominent source of this hesitancy was not resistance to technology but an absence of awareness. Many were uncertain about what tools existed for their specific condition, and some doubted that available technology would be of genuine benefit given the nature of their condition.

> *Just really not aware of what's possible, and why I would need to, I guess… there's nothing that can be done for me, at least at the moment. You think, having got a disease like this there will be all kinds of modern*

XSL•FO
RenderX

tech, but no, there's nothing really at all. [Patient, Parkinson disease, male]

Theme 4: User Age and Access to Technology

Age-related barriers to digital health technology were identified in ways that were both first-person and third-person in character. Some older participants themselves reported limited digital confidence or engagement: one participant aged 70 years with Parkinson disease described using technology minimally and feeling no inclination to develop further digital skills, while a participant aged 81 years described feeling terrified by a video consultation experience and unable to manage without her daughter's help. Other participants, including one aged 70 years with motor neuron disease who reported no personal difficulty, projected barriers onto those older than themselves, describing digital exclusion as a generational issue affecting people beyond their own age group. Concerns were also expressed about equitable access, particularly for those without smartphones or reliable internet access.

I know this is about, you know, neurological [conditions], but the more elderly patients, what do they do? If they don't have a smartphone... You assume that everyone's got the technology, don't we?... There's a lot of poverty here as well, and deprivation… There has to be an alternative. [Patient, multiple sclerosis, female]

Theme 5: Continued Need for Access to Health Care Professionals

Despite openness to digital health technology, the majority of people living with neurological conditions consistently positioned digital health tools as adjuncts to, rather than replacements for, in-person care. Central to this was a desire to be known by their clinician and to have their condition assessed face-to-face, something participants felt could not be replicated digitally. People living with neurological conditions also noted a concern that important clinical information might be missed via remote contact, a preference for face-to-face communication, and a desire for reassurance that they believed could not be adequately provided digitally.

I prefer when I see a consultant or a doctor to see them face to face, in person. I don't like online consultations. And I don't think your GP or anyone can know what's really wrong with you unless you're sitting in front of them. [Patient, epilepsy, female]

Caregiver Group

Caregiver Participants

Thirteen caregivers were recruited to the research. The mean age of caregivers was 65.2 (range 47-77, SD 9.1) years; the group included 6 males and 7 females. Caregiver participants supported people with Huntington disease (n=5), Parkinson disease (n=4), young-onset Parkinson disease (n=1), mixed Parkinson disease and Lewy body dementia (n=1), multiple sclerosis (n=1), and the posterior cortical atrophy variant of Alzheimer disease (n=1).

Five themes represented caregiver attitudes: (1) openness to use (supporting people living with neurological conditions); (2) digital literacy, experience, and ability to use; (3) hesitancy, uncertainty, and perception of benefit; (4) user age and access to technology; and (5) continued need for access to health care professionals.

Theme 1: Openness to Use: Supporting People Living With Neurological Conditions

The majority of caregivers reported openness toward digital health technology, motivated by its perceived potential to be helpful or to offer an easier, quicker, or more efficient means of interacting with health services compared with existing methods. A distinctive finding was that caregivers were also open to supporting the digital health technology use of people living with neurological conditions, including reminding them to use the technology, setting up equipment, explaining information, and mediating communications with health care professionals.

But, um, (wife) has struggled with it just because she's not particularly…technically minded and lacks the confidence, I think, to actually use the thing; to use any of those sort of, err, communication channels on her own. So, I'm always there just to help set them up. Err, for meetings and everything else, but once she's on there it's fine. [Caregiver, Parkinson disease, male]

Theme 2: Digital Literacy, Experience, and Ability to Use

Caregivers identified digital literacy, IT skills, and experience with technology as important determinants of uptake. Those with greater technical ability were seen as more likely to adopt digital health tools with relative ease. Previous negative experiences with technology were highlighted as a deterrent to future engagement.

Previous bad experiences would make them hesitant. Or lack of experience. Anybody that uses it regularly would go straight for it. [Caregiver, Huntington disease, male]

Caregivers also observed how the symptoms of the condition they were supporting created barriers to technology use. One caregiver described watching the person they supported using an iPad:

He has an enormous amount of patience because it takes him ages to get to where he wants to be…with touching the wrong buttons… [Caregiver, Huntington disease, female]

Theme 3: Hesitancy, Uncertainty, and Perception of Benefit

Although most caregivers reported adequate digital skills, many described hesitancy due to experiences of burden or fatigue when using digital health technology. In some instances, this provoked frustration and a preference for direct contact with a health care professional. Frequent software updates, complex forms, and unreliable systems were cited as compounding

XSL•FO
RenderX

sources of burden. Several caregivers reported that existing digital health technology offered minimal practical benefit, and some expressed that investment in medical research to treat the underlying neurological condition was of greater priority.

> *For god's sake stop changing. Because we used to be signed on for the web app... it seems to change every, on average every year. We have been off it for so long now I can't remember. After two or three incidences when you are starting to get it working again, we have just walked away.* [Caregiver, Parkinson disease, male]

Skepticism, partly grounded in prior disappointment with technology that had failed to meet expectations, was also apparent, as was uncertainty about what digital health tools were available in the context of specific neurological conditions.

Theme 4: User Age and Access to Technology

Caregivers, like people living with neurological conditions, perceived that age and access to technology would influence uptake and use. Younger users with greater digital socialization were seen as better equipped to adopt digital health tools. Concerns about the capacity of older adults, including those with progressive neurological decline, to maintain or acquire digital skills were prominent.

> *If they don't have computers at home, if they don't have digital [technology], they just got left out… they are missing out an age range where people basically don't want to know. And people are dropping out of technology, not opting in.* [Caregiver, Parkinson disease, male and patient, Parkinson disease, female]

Theme 5: Continued Need for Access to Health Care Professionals

The majority of caregivers emphasized the continued importance of in-person access to health care professionals, for reasons including the risk of following inaccurate online information, the difficulty of communicating complex concerns digitally, concerns about the technology's capacity to detect serious health changes, and the desire for human reassurance.

> *It's got a lot of potential. One of the things that we've found both of us was that it's nice to meet a person up front and personal first. Then go on to the technology because you feel as if you've built up a rapport with someone before. Otherwise, it just seems a little distant.* [Caregiver, Huntington disease, female]

Health Care Professional Group

Health Care Professional Participants

Seven nurse specialists were recruited to the research as health care professionals. The mean age of this group was 47.0 (range 33-54, SD 6.9) years; all 7 were female. Clinical expertise was varied across multiple sclerosis (n=2), neurological disorders (n=1), epilepsy (n=1), Huntington disease (n=1), motor neuron disease (n=1), and Parkinson disease (n=1). Six participants worked in a hospital-based clinical role and 1 in a community-based advisory role. The sample was modest in size and had limited professional and gender diversity. For this reason, findings from this group should be interpreted as indicative of nurse specialist perspectives rather than as representative of health care professionals in neurology more broadly.

Five themes represented health care professional attitudes: (1) openness to use (promoting self-management and managing caseloads); (2) digital literacy, experience, and ability to use; (3) hesitancy, uncertainty, and perception of benefit; (4) user age and access to technology; and (5) continued need for access to health care professionals.

Theme 1: Openness to Use: Promoting Self-Management and Managing Caseloads

Health care professionals appeared open to the increasing amount of digital health technology being used in their role and for people living with neurological conditions. They identified promotion of self-management among people living with neurological conditions, caseload triage, facilitation of shared decision-making, and symptom monitoring as primary functions of digital health technology in their practice. Online platforms such as My Medical Record were used for messaging, questionnaire administration, sharing treatment information, and enabling people living with neurological conditions to review their own health records. Wearable devices, seizure monitors, and GPS-enabled alarms were also mentioned for supporting the independence of people living with neurological conditions in the community.

> *I absolutely love [the online personal health record]. I think it's the future, and we have to really embrace that. Because our patient numbers are increasing every year, and the only way to keep on top of that is digital. Our patients all have long-term neurological conditions, so they need to be promoted and self-managed.* [Health care professional, multiple sclerosis, female]

Theme 2: Digital Literacy, Experience, and Ability to Use

The majority of health care professionals considered their own digital skills adequate to very good, with several noting that these had improved through use of digital health platforms in their clinical roles.

> *…it's been a hard fast learning curve. Yes, so, I'm, I'm definitely not a tech phob, but I'm equally not an IT expert. So, I'm probably somewhere in the middle, um, but I feel that… I've got to know this system very well and also; I work very closely with the MyMedicalRecord team learning how things work.* [Health care professional, neurological disorders, female]

Health care professional attitude toward digital technology was also identified as likely to influence uptake.

When discussing the digital literacy of their patients, health care professionals identified the progressive nature of some neurological conditions as a particular challenge.

XSL•FO
RenderX

> *There's the perhaps physical or cognitive barriers to using [technology]…people with weak upper limbs, or cognitive effects, the ability to learn new information.* [Health care professional, motor neuron disease, female]

Theme 3: Hesitancy, Uncertainty, and Perception of Benefit

A sense of initial hesitancy or trepidation toward newly introduced digital tools was discernible, particularly when these were introduced alongside existing established working practices. Concerns about workload, time, and the perceived value of new technology were common. Participants indicated that adequate training, ongoing support, clear rationale, and an explanation of anticipated benefits were necessary to facilitate uptake and continued use. Technical difficulties, including system crashes, slow performance, and difficulty locating information, were frequently reported.

> *There's very little research about how useful things really are, so it's very difficult for us to advise patients on 'Yep, you need to get this one', because there's very little evidence base about whether they're useful or effective.* [Health care professional, epilepsy, female]

The majority of health care professionals also acknowledged uncertainty about which digital health tools were available to people living with neurological conditions and expressed reluctance to make specific recommendations without a robust evidence base to draw upon.

Theme 4: User Age and Access to Technology

Health care professionals perceived that user age, access to appropriate technology, and financial resources would meaningfully influence digital health technology uptake and use. People living with neurological conditions who are older, particularly those with progressive conditions affecting cognitive function, were identified as facing significant challenges. The absence of guidance on which technologies to purchase and limited evidence for device efficacy were noted as additional barriers.

> *I think most of us would think of cognition, but actually in reality, access to the actual digital device is the barrier. Not everyone has a smartphone, a tablet, or a laptop; not everyone can afford a computer.* [Health care professional, multiple sclerosis, female]

Theme 5: Continued Need for Access to Health Care Professionals

Health care professionals recognized that digital health technology should function in tandem with, rather than as a replacement for, in-person clinical interaction. Situations in which in-person contact remained necessary included cases where technology use had been unsuccessful, where people living with neurological conditions or caregivers were unable to use digital tools, where immediate clinical decision-making was required, and where the nature of the condition, or individual preference, made in-person consultation more appropriate.

> *I suppose perhaps a concern I would have is that if it were sort of too readily there to replace face-to-face discussions, appointments, that kind of thing... I would see online resource guides as in tandem with conversation. I wouldn't see it as replacing.* [Health care professional, motor neuron disease, female]

Discussion

Principal Findings

This study reports 3 separate qualitative analyses of attitudes toward digital health technology from the perspectives of people living with neurological conditions, caregivers, and health care professionals. Despite separate analyses, 5 themes cut across all 3 groups. These themes of openness to use, digital literacy, hesitancy and perception of benefit, concerns about user age and access, and the desire to preserve in-person contact largely confirm and extend existing evidence and are discussed accordingly. However, 3 findings within and across the themes are particularly noteworthy. First, participants across all 3 groups were largely unaware of what digital health tools were available for their specific neurological condition. This is a finding with direct and immediate practical implications. Second, neurological symptoms themselves, including fatigue, cognitive impairment, tremor, and photosensitivity, were identified as barriers to technology use that are distinct from, and additional to, the digital literacy barriers well documented in the general population. Third, comparison across the 3 groups showed that people living with neurological conditions, caregivers, and health care professionals approached digital health technology from different perspectives, reflecting their respective roles and relationships within care.

Theme 1: Openness to Use

Across all 3 participant groups, the predominant attitude toward digital health technology was one of openness and receptivity. This finding is consistent with the broader literature. A survey of people with neurodegenerative conditions and their caregivers attending a specialist neurology clinic found that participants were supportive of using digital technologies to complement standard clinical care [21]. A qualitative study of patients who had had a stroke similarly identified generally receptive, if mixed, attitudes toward digital health, with participants seeing technology as a convenient means of providing condition-related information and supporting self-management and care access [23]. The present study extends this evidence by capturing the perspectives of people living with neurological conditions, caregivers, and health care professionals simultaneously across a broad range of neurological conditions and by demonstrating that openness extends beyond people living with neurological conditions to include those who support and deliver their care. Openness was consistently conditional rather than unconditional, contingent on the technology being genuinely useful, easy to use, and, notably, relevant to the specific condition. This conditionality is discussed further below in the context of the awareness gap identified across all 3 groups.

XSL•FO
RenderX

Theme 2: Digital Literacy, Experience, and Ability to Use

Personal digital literacy and experience with technology emerged as important determinants of uptake and use across all 3 groups, with greater digital familiarity associated with greater ease of adoption. This is consistent with the literature, including a qualitative study of informal caregivers that identified lack of training as the most common challenge in using digital health technologies [24] and an umbrella review of barriers to health care professional adoption of digital technologies [25]. Previous research has also begun to identify neurological symptoms as barriers to digital health technology use. A literature review of mHealth use in multiple sclerosis identified difficulties with vision, dexterity, and memory as condition-specific barriers, depending on disease progression and severity [26], while a systematic review of digital health technology implementation in neurological rehabilitation identified fatigue, ataxia, pain, limited mobility, and cognitive status among the most frequently reported barriers to patient engagement [27]. Our findings suggest that, in the neurological condition context, digital literacy cannot be understood solely as a function of experience or education. Rather, the ability to use digital health technology was also shaped by the effects of neurological conditions, which could reduce a person's capacity to use digital tools even when they had previously been able to use these with ease. This distinction is important because difficulties using digital health technology may not reflect limited digital literacy in the conventional sense but changes in the physical, sensory, or cognitive capacity required to use the technology. Participants described how photosensitivity and fatigue reduced the time and cognitive resources available for screen use; how tremor and dexterity problems made physical interaction with mobile devices difficult; and how cognitive effects of neurological conditions, including reduced ability to learn and use new platforms, created additional barriers for some. These barriers were identified not only by people living with neurological conditions themselves but also by caregivers managing digital tasks on behalf of the patient and by health care professionals who saw these difficulties as underappreciated in digital health design. Digital health technologies therefore need to accommodate both differences in prior digital experience and the fluctuating, condition-specific effects of neurological symptoms through adaptive and accessible design.

Theme 3: Hesitancy, Uncertainty, and Perception of Benefit

Hesitancy toward digital health technology was evident across all 3 participant groups, underpinned by uncertainty about what was available, skepticism about whether existing tools would be genuinely useful, and, particularly among caregivers, a sense of disappointment with technology that had previously failed to meet expectations. An updated systematic review of digital health adoption among older adults with chronic diseases found that hesitancy was a motivational barrier, driven by mistrust of digital systems, concerns about data privacy, and anxiety about using unfamiliar technology. Conversely, prior positive experiences were identified as a motivator [6]. The difficulty of knowing what is available may be particularly relevant in neurological conditions. A comprehensive review of digital health technology in neurological conditions noted that the landscape remains difficult to navigate, with relatively few tools described in the published literature and uncertainty about which remain in active use [7].

Our findings suggest that this uncertainty extends beyond the complexity of the digital health landscape itself. Participants across all 3 groups (people living with neurological conditions, caregivers, and health care professionals) were largely unaware of the digital health tools available for their specific neurological condition. This awareness gap was evident across the neurological conditions represented in the sample and was reported independently by all 3 participant groups,, suggesting that it is not confined to a particular condition or care setting. Notably, health care professionals, who might be expected to signpost people to available tools, were themselves often uncertain about what was available and reluctant to recommend tools without a stronger evidence base. As a result, people living with neurological conditions and caregivers may be unlikely to identify relevant tools independently, while professionals may also be reluctant to recommend them. Improving awareness may therefore be an important and relatively immediate target for supporting uptake. This needs to involve people living with neurological conditions and caregivers, but also the health care professionals who may be best placed to introduce and signpost relevant tools. In this context, improving access to clear information about available, evidence-based digital health tools may be a more immediate target than attempting to address more established attitudinal and literacy barriers.

Theme 4: User Age and Access to Technology

All 3 participant groups perceived that user age and access to technology would influence uptake, with older users generally seen as less likely to adopt digital health tools. This is consistent with existing literature, but our findings place this issue within the specific context of neurological care. Many neurological conditions disproportionately affect older adults, while progressive disability may further increase the challenges of using technology over time, potentially increasing the risk of digital exclusion. Across all 3 groups, participants emphasized the need to retain nondigital alternatives and identified loaned devices and dedicated support as ways of helping people who might otherwise be excluded.

Theme 5: Continued Need for Access to Health Care Professionals

Despite broad openness to digital health technology, the majority of participants across all 3 groups expressed a persistent desire for in-person access to health care professionals. Digital tools were widely seen as a complement to, rather than a substitute for, direct clinical contact. A mixed methods study of over 1200 people found that 71% preferred in-person visits to telemedicine, with the most common reasons being the perceived ability of clinicians to conduct more accurate examinations and diagnoses in person [28]. A review of digital health implementation from the perspective of health care professionals identified concern about digital health technology impeding the patient-clinician relationship [29]. These issues may be particularly pertinent in neurological condition care, as many conditions involve symptoms that participants felt could only be accurately assessed

XSL•FO
RenderX

face-to-face, such as functional or cognitive decline or fluctuating motor symptoms. It is worth noting that these data were collected in 2022-2023. While the introduction of the NHS 10 Year Health Plan (2025) and subsequent digital investment represent an evolution in the policy landscape since data collection, the findings provide an important perspective on attitudes that may continue to shape the current implementation of digital health.

Cross-Group Comparison

Across the 5 themes, the data revealed both convergence and difference between the 3 participant groups. Convergence was evident in the broad pattern of conditional openness, the identification of digital literacy and access as barriers, uncertainty about what tools were available, concerns about the exclusion of older and less digitally engaged users, and the continued value placed on in-person clinical contact. This consistency across groups suggests that many of the factors shaping views of digital health technology in neurological conditions are shared by those who receive, support, and deliver care.

At the same time, there were also important differences in how the 3 groups approached these issues. People living with neurological conditions considered digital health technology primarily in terms of its personal usefulness: whether it could reduce the burden of managing their condition, support self-management, or reduce the need to travel. Their hesitancy was often related to uncertainty about whether tools existed for their specific condition. Caregivers approached the same issues from the perspective of supporting someone else, considering whether technology would help them provide that support. Their hesitancy more often reflected previous negative experiences with unreliable or frequently changing platforms, as well as concerns that digital-only contact could make it harder to raise concerns directly. Health care professionals tended to consider digital health technology in relation to service capacity and supporting self-management, while their hesitancy centered on a perceived absence of an adequate evidence base, concerns about workload, and uncertainty about how to position digital tools alongside established clinical practice.

The shared desire to preserve in-person contact also reflected these different roles. People living with neurological conditions valued being known by their clinician and having their condition monitored over time; caregivers were concerned that digital tools might reduce their ability to communicate concerns directly and escalate quickly when needed; and health care professionals saw in-person contact as important for clinical assessment and building relationships. These differences are relevant to the design and promotion of digital health tools. Implementation needs to take account of the perspectives of people living with neurological conditions, caregivers, and health care professionals to make sure it meets all 3 groups' needs.

Implications for Practice and Design

Several practical implications follow from these findings. The awareness gap identified across all 3 groups points to a relatively straightforward area for action: improving awareness of available, evidence-based digital health tools among people living with neurological conditions, caregivers, and health care professionals. Health care professionals are particularly important here. They were often as uncertain as patients and caregivers about what tools were available, yet are potentially well placed to introduce and signpost relevant technologies. Clear guidance for health care professionals on available tools may therefore help to improve awareness and support uptake.

The finding that neurological symptoms can act as barriers to technology use, independent of prior digital experience, also has implications for design. Existing accessibility standards do not necessarily address the specific and variable symptoms experienced by people living with neurological conditions. People living with neurological conditions should therefore be involved in the design and testing of digital health technologies, with features such as adjustable screen brightness, voice interaction, simplified navigation, and low-demand interfaces considered as part of the design process rather than added later.

The cross-group comparison also shows that the same tool may be judged against different concerns depending on who is using it. People living with neurological conditions considered personal benefit and relevance to their condition, while caregivers focused more on reliability and whether technology supported rather than replaced direct contact with services. Health care professionals were concerned with the strength of the evidence and the implications for clinical workload. Implementation should therefore take account of all 3 perspectives, and evaluation should consider acceptability to people living with neurological conditions, caregivers, and health care professionals alongside clinical effectiveness.

Finally, the consistent preference for in-person care across all 3 groups suggests that clear communication about how digital engagement will fit alongside, rather than replace, clinical contact may be important in supporting uptake.

Limitations

Several limitations should be acknowledged. The study used a convenience sample, which may limit the representativeness of findings. Recruitment was largely through charities, support groups, and an online personal health record platform and may therefore have underrepresented people who are less engaged with formal support structures or digital technology. Recruitment through neurological condition charities and support groups may also have attracted participants with a stronger interest in their condition, its management, and digital health technology, potentially contributing to the generally positive orientation reported across groups. The recruitment route for participants was not recorded. The choice of telephone interviews may have limited rapport-building and precluded observation of participants' direct interactions with technology and may also have attracted participants more comfortable with telephone as a medium. The relatively small sample size, while consistent with qualitative methodology, precludes quantitative generalization, and the study relied on self-report and may have been influenced by social desirability. Data were collected between 2022 and 2023, during the later stages and immediately following the COVID-19 pandemic. Although subsequent policy developments may have changed the digital health landscape, the findings remain relevant to current implementation,

XSL•FO
RenderX

particularly the structural barriers identified. Findings from health care professionals should be interpreted as indicative of nurse specialist perspectives rather than representative of neurology professionals more broadly, given the modest sample size and limited diversity in gender and professional role. Future research should include more diverse samples, particularly people with limited prior digital engagement and those from lower socioeconomic backgrounds.

Conclusions

The majority of participants were open to the uptake and use of digital health technology, although this openness was consistently conditional. Three findings have particular implications for digital health for people living with neurological conditions. First, many people living with neurological conditions, caregivers, and health care professionals were unaware of what digital health tools were available for their specific condition. Improving awareness may therefore be a relatively immediate and practical target for intervention. Second, the neurological condition itself emerged as a barrier to technology use that is distinct from, and additional to, the digital literacy barriers well documented in the general population but which is not always reflected in current tool design. Third, although the 3 groups shared concerns about digital literacy, uncertainty, perceived benefit, user age and access, and the continuing importance of in-person clinical contact, they approached digital health technology from different positions: personal benefit for people living with neurological conditions, the practical demands of supporting someone else for caregivers, and service capacity for health care professionals. These differences highlight the importance of involving all 3 groups in the design and implementation of digital health tools.

Digital health tools should be co-designed with meaningful input from people living with neurological conditions, caregivers, and health care professionals and should provide evidence-based support while reducing complexity and burden and promoting equitable access, including across neurological symptoms. They should complement rather than replace the in-person care relationships that remain important to this population. Proactive promotion of available tools will also be needed to improve awareness among people living with neurological conditions, caregivers, and health care professionals.

Acknowledgments

The authors gratefully acknowledge the contribution of all study participants, the PPI contributors, the project steering group, and the neurological condition support groups and charities that facilitated participant recruitment, including Parkinson’s UK, the Multiple Sclerosis Society, the Huntington's Disease Association, Epilepsy Action, and the Motor Neurone Disease Association.

Funding

The research was funded by the National Institute for Health and Care Research Applied Research Collaboration Wessex. The funder had no involvement in this study’s design, data collection, analysis, interpretation, or the writing of this paper.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Generative AI Use

The authors declare that generative AI was used in the creation of this manuscript. Generative AI (Claude) was used for proofreading, editing, or language polishing, and word count reduction. The authors verified all outputs of the tool(s) used in the preparation of this manuscript and take full responsibility for the final content of the manuscript.

Authors' Contributions

Conceptualization: SF, JS, CK
Formal analysis: JS, SF, CK
Funding acquisition: CK, SF, JS
Investigation: JS
Methodology: JS, SF, CK
Project administration: JS
Supervision: CK
Writing – original draft: JS, SF
Writing – review and editing: SF, JS, CK

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

COREQ checklist.
[DOCX File , 22 KB - humanfactors_v13i1e107517_app1.docx]

Multimedia Appendix 2
Illustrative participant quotations by theme and participant group.
[DOCX File , 19 KB - humanfactors_v13i1e107517_app2.docx]

References

1. Global status report on neurology. World Health Organization. 2025 Oct 12. URL: https://www.who.int/publications/i/item/9789240116139 [accessed 2026-10-23]
2. GBD 2021 Nervous System Disorders Collaborators. Global, regional, and national burden of disorders affecting the nervous system, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. Lancet Neurol 2024;23(4):344-381 [FREE Full text] [doi: 10.1016/S1474-4422(24)00038-3] [Medline: 38493795]
3. Monitoring and evaluating digital health interventions: a practical guide to conducting research and assessment. World Health Organization. 2016 Dec 12. URL: https://www.who.int/publications/i/item/9789241511766 [accessed 2026-10-23]
4. Fit for the future: 10 year health plan for England. Department of Health and Social Care. 2025 3. URL: https://www.gov.uk/government/publications/10-year-health-plan-for-england-fit-for-the-future [accessed 2026-10-23]
5. Frontline digitisation: creating the conditions for a digital NHS. The NHS Alliance. 2024 Aug 6. URL: https://www.nhsconfed.org/publications/frontline-digitisation [accessed 2026-10-23]
6. Hepburn J, Williams L, McCann L. Barriers to and facilitators of digital health technology adoption among older adults with chronic diseases: updated systematic review. JMIR Aging 2025;8:e80000 [FREE Full text] [doi: 10.2196/80000] [Medline: 40934502]
7. Spreadbury JH, Young A, Kipps CM. A comprehensive literature search of digital health technology use in neurological conditions: review of digital tools to promote self-management and support. J Med Internet Res 2022 Jul 28;24(7):e31929 [FREE Full text] [doi: 10.2196/31929] [Medline: 35900822]
8. Masanneck L, Gieseler P, Gordon WJ, Meuth SG, Stern AD. Evidence from ClinicalTrials.gov on the growth of digital health technologies in neurology trials. NPJ Digit Med 2023;6(1):23 [FREE Full text] [doi: 10.1038/s41746-023-00767-1] [Medline: 36765123]
9. Blom MM, Zarit SH, Zwaaftink RBMG, Cuijpers P, Pot AM. Effectiveness of an internet intervention for family caregivers of people with dementia: results of a randomized controlled trial. PLoS One 2015;10(2):e0116622 [FREE Full text] [doi: 10.1371/journal.pone.0116622] [Medline: 25679228]
10. Chiu T, Marziali E, Colantonio A, Carswell A, Gruneir M, Tang M, et al. Internet-based caregiver support for Chinese Canadians taking care of a family member with Alzheimer disease and related dementia. Can J Aging 2009;28(4):323-336. [doi: 10.1017/S0714980809990158] [Medline: 19925698]
11. D'hooghe M, Van Gassen G, Kos D, Bouquiaux O, Cambron M, Decoo D, et al. Improving fatigue in multiple sclerosis by smartphone-supported energy management: the MS TeleCoach feasibility study. Mult Scler Relat Disord 2018;22:90-96. [doi: 10.1016/j.msard.2018.03.020] [Medline: 29649789]
12. Fischer A, Schröder J, Vettorazzi E, Wolf OT, Pöttgen J, Lau S, et al. An online programme to reduce depression in patients with multiple sclerosis: a randomised controlled trial. Lancet Psychiatry 2015 Mar;2(3):217-223. [doi: 10.1016/S2215-0366(14)00049-2] [Medline: 26359900]
13. Hixson JD, Barnes D, Parko K, Durgin T, Van Bebber S, Graham A, et al. Patients optimizing epilepsy management via an online community: the POEM Study. Neurology 2015;85(2):129-136. [doi: 10.1212/wnl.0000000000001728] [Medline: 26085605]
14. Hanrahan M, Wilson C, Keogh A, Barker S, Rochester L, Brittain K, et al. How can patients shape digital medicine? A rapid review of patient and public involvement and engagement in the development of digital health technologies for neurological conditions. Expert Rev Pharmacoecon Outcomes Res 2025 Feb;25(2):137-154 [FREE Full text] [doi: 10.1080/14737167.2024.2410245] [Medline: 39376020]
15. Bruno E, Simblett S, Lang A, Biondi A, Odoi C, Schulze-Bonhage A, RADAR-CNS Consortium. Wearable technology in epilepsy: the views of patients, caregivers, and healthcare professionals. Epilepsy Behav 2018 Aug;85:141-149. [doi: 10.1016/j.yebeh.2018.05.044] [Medline: 29940377]
16. Gandy M, Karin E, Fogliati V, Meares S, Nielssen O, Titov N, et al. Emotional and cognitive difficulties, help-seeking, and barriers to treatment in neurological disorders. Rehabil Psychol 2018;63(4):563-574. [doi: 10.1037/rep0000241] [Medline: 30247055]
17. Marrie RA, Leung S, Tyry T, Cutter GR, Fox R, Salter A. Use of eHealth and mHealth technology by persons with multiple sclerosis. Mult Scler Relat Disord 2019;27:13-19. [doi: 10.1016/j.msard.2018.09.036] [Medline: 30296732]
18. Simblett SK, Bruno E, Siddi S, Matcham F, Giuliano L, López JH, RADAR-CNS Consortium. Patient perspectives on the acceptability of mHealth technology for remote measurement and management of epilepsy: a qualitative analysis. Epilepsy Behav 2019 Aug;97:123-129. [doi: 10.1016/j.yebeh.2019.05.035] [Medline: 31247523]

XSL•FO
RenderX

19. Simblett SK, Evans J, Greer B, Curtis H, Matcham F, Radaelli M, RADAR-CNS consortium. Engaging across dimensions of diversity: a cross-national perspective on mHealth tools for managing relapsing remitting and progressive multiple sclerosis. Mult Scler Relat Disord 2019 Jul;32:123-132. [doi: 10.1016/j.msard.2019.04.020] [Medline: 31125754]
20. Winberg C, Kylberg M, Pettersson C, Harnett T, Hedvall P, Mattsson T, et al. Feeling controlled or being in control? Apps for self-management among older people with neurological disability. Disabil Rehabil Assist Technol 2021 Aug;16(6):603-608 [FREE Full text] [doi: 10.1080/17483107.2019.1685017] [Medline: 31711351]
21. Reid K, Beswick E, Tam J, Symonds A, Lyle D, Newton J, et al. Perceptions of digital technology use for monitoring health in people living with neurological disorders. J Neurol Sci 2024;462:123072. [doi: 10.1016/j.jns.2024.123072] [Medline: 38885541]
22. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
23. Bally ELS, Cheng D, van Grieken A, Sanz MF, Zanutto O, Carroll A, et al. Patients' perspectives regarding digital health technology to support self-management and improve integrated stroke care: qualitative interview study. J Med Internet Res 2023;25:e42556. [doi: 10.2196/42556] [Medline: 37014677]
24. Zainal H, Xiaohui X, Thumboo J, Seah SJ, Leng LL, Kok Yong FK. Exploring caregiver challenges, digital health technologies, and healthcare support: a qualitative study. Front Digit Health 2025;7:1587162 [FREE Full text] [doi: 10.3389/fdgth.2025.1587162] [Medline: 40547053]
25. do Nascimento IJB, Abdulazeem H, Vasanthan LT, Martinez EZ, Zucoloto ML, Østengaard L, et al. Barriers and facilitators to utilizing digital health technologies by healthcare professionals. NPJ Digit Med 2023;6(1):161. [doi: 10.1038/s41746-023-00899-4] [Medline: 37723240]
26. Gromisch ES, Turner AP, Haselkorn JK, Lo AC, Agresta T. Mobile health (mHealth) usage, barriers, and technological considerations in persons with multiple sclerosis: a literature review. JAMIA Open 2021;4(3):ooaa067 [FREE Full text] [doi: 10.1093/jamiaopen/ooaa067] [Medline: 34514349]
27. Jarvis K, Thetford C, Turck E, Ogley K, Stockley RC. Understanding the barriers and facilitators of digital health technology (DHT) implementation in neurological rehabilitation: an integrative systematic review. Health Serv Insights 2024;17:11786329241229917 [FREE Full text] [doi: 10.1177/11786329241229917] [Medline: 38690403]
28. Moulaei K, Sheikhtaheri A, Fatehi F, Shanbehzadeh M, Bahaadinbeigy K. Patients' perspectives and preferences toward telemedicine versus in-person visits: a mixed-methods study on 1226 patients. BMC Med Inform Decis Mak 2023 Nov 15;23(1):261 [FREE Full text] [doi: 10.1186/s12911-023-02348-4] [Medline: 37968639]
29. Oudbier SJ, Souget-Ruff SP, Chen BSJ, Ziesemer KA, Meij HJ, Smets EMA. Implementation barriers and facilitators of remote monitoring, remote consultation and digital care platforms through the eyes of healthcare professionals: a review of reviews. BMJ Open 2024;14(6):e075833 [FREE Full text] [doi: 10.1136/bmjopen-2023-075833] [Medline: 38858155]

Abbreviations

COM-B: capability, opportunity, and motivation–behavior
COREQ: Consolidated Criteria for Reporting Qualitative Research
mHealth: mobile health
NHS: National Health Service
PPI: patient and public involvement

Edited by Y Gong; submitted 20.Jul.2026; peer-reviewed by MS Ranaiy, P Liu; comments to author 10.Sep.2026; revised version received 01.Oct.2026; accepted 19.Oct.2026; published 19.Nov.2026.

Please cite as:
Fearn S, Spreadbury J, Kipps C
Attitudes Toward Digital Health Technology in Neurological Conditions: Qualitative Study of Patients, Caregivers, and Health Care Professionals
JMIR Hum Factors 2026;13:e107517
URL: https://humanfactors.jmir.org/2026/1/e107517
doi:10.2196/107517
PMID:

©Sarah Fearn, John Spreadbury, Christopher Kipps. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 19.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information,

XSL•FO
RenderX

a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

AI-Assisted Clinical Documentation in Routine Danish General Practice: Quantitative Pre-Post Study

Louise Nørgaard Olsen[1], MD; Philipp Harbig[2,3], MD, PhD; Anna Bay Laurberg[4], BSc (med); Jacob Laurberg[3], MD, PhD; Morten Haaning Charles[3,5,6], MD, PhD

[1]Department of Internal Medicine, Regional Hospital Viborg, Heibergs Allé 5A, Viborg, Denmark

[2]Department of Public Health, Research Unit for General Practice, Aarhus University, Aarhus, Central Jutland, Denmark

[3]General Practice, Aarhus, Denmark

[4]Faculty of Health, Aarhus University, Aarhus, Denmark

[5]Steno Diabetes Center Aarhus, Aarhus University Hospital, Aarhus, Central Jutland, Denmark

[6]Department of Clinical Medicine, Aarhus University, Aarhus, Central Jutland, Denmark

Corresponding Author:
Louise Nørgaard Olsen, MD
Department of Internal Medicine, Regional Hospital Viborg, Heibergs Allé 5A, Viborg, Denmark

Abstract

Background: Administrative workload in general practice limits time for direct patient care. AI-assisted documentation has been proposed as a way to reduce the documentation burden, but evidence from routine primary care settings remains limited.

Objective: This study aimed to evaluate general practitioners' (GPs) acceptance of AI-assisted documentation and its association with documentation time and clinical note quality in routine Danish general practice.

Methods: We conducted a quantitative pragmatic pre-post quality improvement evaluation in Danish general practice. A total of 20 GPs documented 239 consultations before and 236 consultations after implementation of an AI-assisted documentation system. Documentation quality, structure, clinical clarity, and documentation time categories were self-assessed using standardized audit forms completed immediately after each consultation. Technology acceptance and usability were assessed using the technology acceptance model (TAM) and the System Usability Scale (SUS).

Results: Self-assessed documentation structure increased from 3.99 to 4.45, while self-reported documentation time categories decreased from 2.85 to 2.29. Technology acceptance and usability were high (TAM domain means 3.76-4.19; SUS mean 77.5). GP-level paired analyses showed moderate improvements in structure and clarity and a reduction in documentation time. Combined blinded external assessments showed higher postimplementation scores for quality, structure, and clinical clarity, although reviewer-specific ratings diverged, and interrater reliability was low. The association between documentation time and perceived quality was negligible. TAM and SUS indicated high clinician acceptance.

Conclusions: AI-assisted documentation was associated with lower self-reported documentation time categories while maintaining or modestly improving perceived clinical note quality. These findings support the feasibility of AI-assisted documentation in primary care, while highlighting the need for controlled studies with objective time measurement and longer follow-up.

(JMIR Hum Factors 2026;13:e94431) doi:10.2196/94431

KEYWORDS

artificial intelligence; AI; clinical documentation; general practice; primary care; ambient documentation; electronic health records; technology acceptance; clinical workflow

Introduction

Administrative workload in general practice has increasingly constrained the time available for direct patient care and continuity of clinical interaction [1,2]. A substantial proportion of this burden relates to documentation and maintenance of electronic health records (EHRs), which often compete with patient-facing time and contribute to clinician workload and dissatisfaction [3]. In response, digital health innovations aimed at supporting documentation processes have received growing attention.

Recent advances in AI have enabled the development of so-called "ambient" or AI-assisted documentation systems that capture clinician-patient dialogue, generate real-time transcripts, and produce draft clinical notes for clinician review and approval [4-6]. These systems are designed to support clinical documentation by generating draft clinical notes for clinician review and approval rather than functioning as autonomous documentation systems. The clinician retains full responsibility

XSL•FO
RenderX

for reviewing, editing, and approving the final clinical note. International studies suggest that AI-assisted documentation may reduce perceived documentation burden and support workflow efficiency without compromising documentation quality [1,4,5,7]. However, the existing evidence base is heterogeneous, and many studies are conducted in selected settings or under experimental conditions.

Implementation of AI-based documentation in general practice raises specific considerations related to trust, usability, data protection, and professional accountability. In European primary care settings, compliance with the General Data Protection Regulation (GDPR), transparency of data handling, and clinician control over documentation are central prerequisites for adoption [8,9]. Qualitative studies indicate that general practitioners (GPs) perceive both opportunities and reservations regarding AI-supported documentation, emphasizing the importance of usability, reliability, and a clear "human-in-the-loop" workflow [9,10].

Despite growing interest, evidence from routine primary care workflows remains limited, particularly from Scandinavian general practice. Moreover, few studies have examined whether clinicians' self-assessments of AI-assisted documentation quality correspond with independent professional evaluations. External blinded assessment of documentation quality is rarely included in implementation studies, yet such evaluation is relevant to address concerns regarding potential bias in self-reported outcomes.

The aim of the present study was therefore to evaluate the feasibility of implementing an AI-assisted documentation system in routine Danish general practice. Specifically, we aimed to assess (1) GPs' acceptance and perceived usability of the system; (2) perceived changes in documentation time; and (3) changes in clinical note quality, structure, and clinical clarity. To strengthen the evaluation beyond self-assessment alone, we additionally compared GPs' ratings of documentation quality with blinded assessments performed by independent physicians. By combining clinician-reported outcomes with external evaluation under real-world conditions, this study aimed to contribute pragmatic evidence on the role of AI-assisted documentation in everyday primary care practice.

Methods

Study Design and Setting

We conducted a quantitative pragmatic pre-post quality improvement evaluation of the implementation of an AI-assisted clinical documentation system in routine Danish general practice. The primary objective was to evaluate clinician acceptance, perceived usability, and integration of the system into routine clinical workflows rather than patient-level clinical outcomes. The study was carried out within a regional primary care quality cluster ("klynge" in Danish) comprising 7 GP clinics and 20 participating GPs. The cluster had been an established structure for several years before the study and was not formed specifically for this project. It was based on administrative and geographic criteria and was not established based on interest in digital health or information technology, thereby reflecting a routine primary care setting. Participation was voluntary and offered to all GPs within the quality cluster. The participating practices included both solo and partnership practices, but formal measures of practice characteristics, patient population, and digital maturity were not collected. Data collection was conducted between October 2025 and November 2025.

Selection of the AI-Assisted Documentation System

Prior to study initiation, several AI-assisted documentation systems available for Danish general practice were informally explored by the project group with regard to usability and fit with routine clinical workflows. This was a pragmatic local assessment rather than a formal comparative evaluation or procurement process. Noteless (Noteless AS) was selected pragmatically by the project group based on perceived suitability for implementation in routine Danish general practice and its availability for use within the participating quality improvement initiative. The selection was not based on a formal comparative evaluation of available AI-assisted documentation systems. Consequently, the findings should be interpreted as specific to this system, version, and implementation context and may not be directly transferable to other AI-assisted documentation systems.

Participants and Consultations

Participating GPs were instructed to document up to 12 consecutive in-person consultations before implementation of the AI-assisted documentation system. Following system installation, a 2-week familiarization period was allowed to enable integration into routine clinical workflows. After this period, GPs documented up to 12 consecutive in-person consultations using the AI-assisted system.

During both periods, GPs aimed to document 12 consecutive in-person consultations in each period; consultations could be excluded at the GP's discretion if deemed unsuitable, and inclusion continued until 12 eligible consultations were reached when feasible. Documentation time categories and note quality ratings were completed by the participating GP immediately after each documented consultation. No restrictions were placed on patient characteristics or consultation type within these in-person consultations, and participating GPs were encouraged to use the AI-assisted documentation system whenever they considered it appropriate within routine clinical practice. Due to pragmatic clinical conditions, incomplete registrations, and missing ratings, the number of evaluable notes varied between GPs. Only notes with complete outcome ratings were included. GP-level analyses were therefore based on the mean of available preimplementation and postimplementation notes for each GP. This pragmatic approach was chosen to minimize disruption to routine clinical workflows and reduce the additional documentation burden associated with study participation. Consequently, participating GPs were not required to document reasons for excluding individual consultations.

Intervention: AI-Assisted Documentation Workflow

The intervention consisted of using the AI-assisted documentation system Noteless during routine in-person consultations. Immediately before each eligible consultation, the GP initiated a secure personal web browser session and

XSL•FO
RenderX

activated real-time transcription. The system captured clinician-patient dialogue to generate a draft clinical note at the end of the consultation. The GP then reviewed and edited the AI-generated draft before final approval and storage in the EHR. The AI-assisted documentation system operated as a web-based application and was not directly integrated with the participating practices' EHR systems. After reviewing and editing the AI-generated draft, the GP manually transferred the final note into the local EHR by copying and pasting. The specific EHR systems used by the participating practices were not recorded.

Audio capture was used solely to enable real-time transcription and draft note generation. Complete audio recordings were not retained. According to the implemented workflow, audio data were deleted immediately, while transcription outputs and generated draft notes were temporarily retained within a European Union (EU)–based cloud environment for up to 24 hours to allow clinician review and editing, after which they were deleted. These transcription outputs and draft notes were not accessible to the vendor after deletion and were not used for AI model training. Evaluation data were not made accessible to the vendor, and clinical notes used for external assessment were deidentified before sharing.

Data Collection and Outcomes

Data were collected using structured audit forms completed by participating GPs, along with the technology acceptance model (TAM) questionnaire and the System Usability Scale (SUS).

Primary Outcomes (Pre-Post)

The following 4 outcomes were assessed both before and after implementation using structured audit forms: overall satisfaction with the quality of the clinical note, whether the clinical note was well structured, whether another health care provider could understand the patient's clinical course based on the note, and documentation time. The first 3 outcomes were rated on 5-point Likert scales, with higher scores indicating more favorable assessments. Documentation time was recorded using 5 ordinal categories (<1 minute, 1 minute, 2-3 minutes, 4-5 minutes, and >5 minutes) and coded from 1 to 5, with lower values indicating shorter documentation time.

All primary outcomes were self-assessed by participating GPs for each documented consultation.

Postimplementation Additional Outcome

After implementation, GPs additionally rated the perceived impact on time available for direct patient care using a 5-point Likert scale. As no corresponding preimplementation measure was available, this outcome was analyzed descriptively only.

Technology Acceptance and Usability

Technology acceptance was assessed using a modified TAM questionnaire comprising 12 items covering 4 domains: perceived usefulness (4 items), perceived ease of use (4 items), attitude toward use (2 items), and behavioral intention to use (2 items). All TAM items were rated on a 5-point Likert scale ranging from strongly disagree to strongly agree. The TAM questionnaire was translated into Danish by 1 of the investigators for use in this quality improvement project and did not constitute a formally validated Danish version. System usability was assessed using the standard 10-item SUS, also using a 5-point Likert response scale. Participants completed the original English SUS questionnaire, while a Danish translation was provided solely as a reference to support comprehension if needed. The questionnaire explicitly instructed participants to answer the original English items and to consult the Danish translation only if they were uncertain about the meaning of the English wording. SUS responses were converted to a total score ranging from 0 to 100, with higher scores indicating better perceived usability.

Blinded External Assessment of Clinical Notes

To reduce potential self-assessment bias, all preimplementation and postimplementation clinical notes were deidentified and randomly ordered. Patient-identifying information was manually removed before the notes were assigned unique study ID numbers. The deidentified notes were then randomly ordered and compiled into PDF documents for independent review. The reviewers did not access the Noteless system or the EHR. Following completion of the assessments, the study ID numbers were used to restore the original allocation for analysis. Two independent physicians (1 GP and 1 hospital-based rheumatologist), not involved in the project and without prior experience using the AI system, independently evaluated each note while blinded to implementation status. The hospital-based rheumatologist was selected because of extensive clinical experience with referrals from general practice to both hospital and rehabilitation services, providing substantial familiarity with documentation from primary care despite not working in general practice.

External reviewers rated the final clinical notes after GP review, editing, and approval, corresponding to the version stored in the EHR rather than the initial AI-generated draft. Clinical note quality, structure, and clinical clarity were rated using the same 5-point Likert scales as those used by the participating GPs. Each note was evaluated independently by both reviewers. The extent of GP editing of the AI-generated drafts was not quantified.

Statistical Analysis

Analyses were conducted at the GP level to avoid inflation of precision due to multiple notes per clinician. For each GP, mean scores were calculated across all available preimplementation notes and all postimplementation notes for each outcome.

The primary analyses used 2-tailed paired *t* tests on GP-level mean differences. Although the underlying variables were measured on ordinal Likert scales, averaging approximately 12 observations per GP resulted in approximately continuous GP-level outcomes. Inspection of the paired differences showed no evidence of marked skewness or influential outliers. Therefore, 2-tailed paired *t* tests were considered appropriate. Wilcoxon signed-rank tests were additionally performed as sensitivity analyses and yielded consistent conclusions.

Given the exploratory nature of the evaluation, *P* values were interpreted descriptively and presented alongside mean differences, 95% CIs, and paired-samples Cohen d_z, calculated as the mean paired change divided by the SD of the paired changes.

XSL•FO
RenderX

For blinded external assessment, mean preimplementation and postimplementation scores were calculated by averaging ratings across both reviewers prior to GP-level aggregation. Interrater reliability between the 2 blinded reviewers was assessed using intraclass correlation coefficients (ICCs; using a 2-way random-effects model with absolute agreement).

Ethical Considerations

The project was conducted as a pragmatic pre-post quality improvement evaluation within routine clinical practice. According to Danish regulation, the project was classified as a quality improvement initiative and therefore did not require approval from a regional research ethics committee. A data protection impact assessment (DPIA) and a data processing agreement were completed before implementation. All data were handled in accordance with Danish data protection regulations and the GDPR.

Prior to participation, patients received information about the AI-assisted documentation system through information screens in the waiting area, written information available in the consultation room, and verbal information provided by the GP. Patients were given the opportunity to ask questions and decline the use of Noteless before the consultation. Patients verbally consented to participation prior to the consultation. Participation could be declined without any consequences for care. Consultation data were contractually prohibited from being used for AI model training. No identifiable patient information was stored in the study dataset or shared for external assessment.

Results

Study Sample and Data Completeness

A total of 20 GPs from 7 general practice clinics participated in the study. Due to incomplete registrations or inconclusive outcome ratings, not all GPs contributed the intended 12 notes in both periods. In total, 475 clinical notes were included in the analysis (preimplementation notes: n=239; postimplementation notes: n=236). One preimplementation note and 2 postimplementation notes were missing because 2 GPs did not reach 12 documented consultations, and 2 postimplementation notes were excluded due to incomplete or inconclusive ratings. The median number of evaluable notes per GP was 12 (IQR 12-12) both before and after the implementation. Analyses were therefore conducted at the GP level using available data.

All participating GPs contributed to the self-assessment analyses. All included notes were available for blinded external assessment by 2 independent physicians.

Self-Assessed Documentation Quality and Time Use

Descriptive self-assessed documentation scores before and after implementation of AI-assisted documentation are summarized in Table 1.

Table . Mean self-assessed and blinded documentation scores before and after implementation of AI-assisted documentation[a].

Outcomes	Self-assessed documentation, mean (SD)		Blinded documentation, mean (SD)	
	Preintervention score	Postintervention score	Preintervention score	Postintervention score
Quality	4.02 (0.44)	4.05 (0.59)	4.29 (0.40)	4.53 (0.20)
Structure	3.99 (0.49)	4.45 (0.57)	4.30 (0.37)	4.60 (0.18)
Clinical clarity	4.10 (0.46)	4.42 (0.57)	4.42 (0.39)	4.73 (0.19)
Documentation time category	2.85 (0.46)	2.29 (0.68)	—[b]	—

[a]Values are presented as mean (SD) based on general practitioner (GP)–level mean scores from 20 participating GPs. Quality, structure, and clinical clarity were rated on 5-point Likert scales (1="lowest" to 5="highest"). Documentation time was recorded using 5 ordinal categories (1="<1 minute" to 5=">5 minutes"). Self-assessments were completed by participating GPs; blinded assessments were performed by 2 independent physicians.

[b]Em dashes indicate that documentation time category was not assessed in the blinded external assessment.

At a descriptive level, mean self-assessed overall documentation quality showed minimal change after implementation, whereas structural quality and clinical clarity increased. The documentation time category decreased, indicating shorter self-reported documentation time following AI-assisted note generation.

GP-level paired analyses are presented in Table 2. Structural quality was higher after implementation, with a mean increase of 0.45 points (95% CI 0.05-0.84; Cohen d_z=0.50). Clinical clarity also appeared higher after implementation, with a mean increase of 0.31 points (95% CI −0.01 to 0.62; Cohen d_z=0.45). No statistically significant change was observed for overall self-assessed quality.

Self-assessed documentation time category was lower after implementation at the GP level (mean change −0.55; 95% CI −0.94 to −0.16; Cohen d_z=−0.61). A mean change of −0.55 categories corresponded approximately to a shift of half a category toward shorter documentation times at the GP level.

Sensitivity analyses using the Wilcoxon signed-rank test yielded results consistent with the primary 2-tailed paired *t* test analyses. Improvements in documentation structure (*P*=.03) and documentation time category (*P*=.02) remained statistically significant, whereas overall documentation quality (*P*=.60) and clinical clarity (*P*=.08) remained nonsignificant, supporting the robustness of the primary analyses.

XSL•FO
RenderX

Table . General practitioner (GP)–level paired pre-post analyses of self-assessed documentation outcomes[a].

Outcomes	Mean change (SD)	95% CI	Cohen d_z	*P* value[b]
Quality	+0.02 (0.80)	–0.33 to 0.37	0.03	.91
Structure	+0.45 (0.90)	0.05 to 0.84	0.50	.04
Clinical clarity	+0.31 (0.69)	–0.01 to 0.62	0.45	.06
Documentation time category	–0.55 (0.89)	–0.94 to –0.16	–0.61	.01

[a]Analyses were conducted at the GP level (N=20) using 2-tailed paired *t* tests, with each GP's mean score across available consultations. Mean change is reported with 95% CIs. Paired-samples Cohen d_z is reported as the standardized effect size.

[b]*P* values are reported for exploratory purposes only.

Associations Between Documentation Time and Perceived Quality

Exploratory analyses showed a negligible association between documentation time category and self-assessed note quality (GP-level scatter plots; R^2=0.01-0.05), suggesting that longer documentation time was not associated with higher perceived quality.

Blinded External Assessment of Documentation Quality

Blinded external assessments are summarized in Table 1.

For external reviewer 1, mean scores increased across all domains following implementation: overall quality from 4.00 to 4.63, structural quality from 4.00 to 4.66, and clinical clarity from 4.26 to 4.75.

For external reviewer 2, mean overall quality changed from 4.58 before implementation to 4.44 after implementation, while structural quality changed from 4.60 to 4.52. Clinical clarity increased from 4.58 to 4.70. Thus, reviewer-specific ratings were not fully concordant: reviewer 1 rated all 3 domains higher after implementation, whereas reviewer 2 rated clinical clarity higher but overall quality and structure slightly lower after implementation. Accordingly, the blinded external assessment should be regarded as supportive rather than confirmatory evidence.

When combining ratings from both blinded reviewers (Table 3), mean overall quality increased from 4.29 before implementation to 4.53 after implementation (mean change +0.24). Structural quality increased from 4.30 to 4.60 (+0.30), and clinical clarity increased from 4.42 to 4.73 (+0.31).

Table . Blinded external assessment of documentation quality and interrater reliability[a].

Outcomes	Pre-post score change, mean (SD)	95% CI	Cohen d_z	Intraclass correlation coefficient (95% CI)
Quality	0.24 (0.33)	0.08-0.37	0.68	0.19 (0.10-0.28)
Structure	0.30 (0.31)	0.17-0.44	0.98	0.24 (0.15-0.32)
Clinical clarity	0.31 (0.30)	0.18-0.44	1.04	0.34 (0.25-0.42)

[a]Interrater reliability between 2 blinded reviewers was assessed using intraclass correlation coefficients (2-way random effects, absolute agreement). Higher values indicate greater agreement. Analyses were performed at the general practitioner (GP) level (N=20), using each GP's mean score across available consultations.

Interrater Reliability

Overview

Interrater reliability between the 2 blinded external reviewers is presented in Table 3. Interrater reliability was low (ICC 0.19-0.34 across domains), indicating substantial between-rater variability in absolute ratings. Despite this variability, the combined reviewer scores were higher after implementation across all assessed domains. However, reviewer-specific ratings were not fully concordant, and the blinded external assessment should therefore be interpreted cautiously.

Perceived Impact on Time Available for Patient Care

In the postimplementation audit, GPs reported a mean score of 3.86 (SD 1.08) for perceived increased time available for direct patient care. As no corresponding preimplementation measure was available, this outcome is reported descriptively only.

Technology Acceptance and Usability

Technology acceptance was high across all TAM domains. Mean perceived usefulness was 3.76 (SD 0.99), and perceived ease of use was 4.19 (SD 0.78). Attitude toward use was similarly positive, with a mean score of 3.97 (SD 0.97), while behavioral intention to use the system was 3.76 (SD 1.10).

SUS scores indicated good overall usability, with a mean score of 77.5 (SD 14.6), corresponding to high perceived ease of use and learnability.

Discussion

Overview

This study evaluated the implementation of an AI-assisted documentation system in routine Danish general practice, combining self-assessment by participating GPs with blinded

XSL•FO
RenderX

external evaluation. Overall, the findings suggest that AI-assisted documentation appeared feasible to integrate into everyday clinical work and was associated with high clinician acceptance, lower self-reported documentation time categories, and maintained or modestly improved perceived documentation quality. Blinded external assessment provided limited supportive evidence for these findings but should be interpreted cautiously given the low interrater reliability and reviewer-level differences. These findings should be interpreted within the context of an exploratory pragmatic pre-post evaluation and an evidence base that is still evolving for large language model–based ambient documentation systems.

Although the study was designed as a pragmatic quality improvement evaluation, we included GP-level paired pre-post analyses to provide transparent quantitative support for the observed changes. By defining the GP as the primary analytic unit and summarizing each clinician's documentation across multiple consultations, we aimed to balance methodological rigor with the exploratory and practice-oriented nature of the study, in line with recommendations for real-world digital health evaluations.

Principal Findings

Self-assessed documentation quality, structure, and clinical clarity appeared higher following implementation of AI-assisted documentation, while the self-reported documentation time category decreased from 2.85 to 2.29, corresponding to a shift toward shorter documentation time. The largest observed differences were found for structural quality and clinical clarity, whereas overall perceived quality showed minimal change. Importantly, these patterns were also observed by blinded external reviewers, who reported similar directional improvements in overall quality, structure, and clinical clarity.

The overall pattern observed in the blinded assessments was broadly consistent with the GPs' self-assessments, but reviewer-specific results were not fully concordant. One reviewer rated all domains higher after implementation, whereas the second reviewer rated overall quality and structure slightly lower and clinical clarity higher. Together with the low interrater reliability, this indicates that the external assessment should not be interpreted as confirmatory evidence of improved documentation quality. Rather, it provides limited supportive evidence that implementation was not associated with a clear deterioration in clinical clarity or overall documentation quality, while highlighting the subjective nature of evaluating narrative clinical notes.

Documentation Time and Quality

The observed reduction in self-reported documentation time categories without a corresponding decline in perceived quality is clinically relevant. Exploratory analyses demonstrated a negligible association between documentation time and perceived documentation quality, indicating that longer documentation time did not translate into higher-quality notes. This challenges the implicit assumption that perceived documentation quality necessarily depends on time investment and suggests that AI-assisted drafting may support more efficient documentation workflows without sacrificing clinical clarity.

These findings align with emerging international evidence showing that AI-assisted documentation systems, including ambient AI and AI-powered voice-to-text solutions, have been associated with reductions in perceived administrative burden while maintaining or improving documentation quality and clinician workflow [1,4-7,11]. Prior studies have reported reduced perceived workload, improved efficiency, and stable documentation quality in primary care and outpatient settings, although most have relied primarily on self-reported outcomes. By incorporating blinded external evaluation, the present study extends the literature with a more robust assessment of documentation quality under routine clinical conditions.

Our findings are also consistent with a recent longitudinal mixed methods implementation study from Dutch general practice, which similarly reported reduced perceived documentation burden and improved workflow following implementation of an ambient AI scribe. However, the authors also identified unintended consequences, including occasional inaccuracies in AI-generated summaries, challenges during sensitive consultations, and potential interference with clinicians' clinical reasoning [12]. Likewise, the recent *EBioMedicine* systematic review concluded that AI-assisted documentation systems show considerable potential to improve efficiency and patient-centered care while emphasizing persistent concerns regarding transcription errors, safety, and the limited generalizability of the current evidence base [6]. This distinction is also reflected in the recent *The Lancet Primary Care* review, which concluded that clinician satisfaction and perceived reductions in documentation burden have been reported more consistently than objectively measured improvements in efficiency. The present findings are consistent with this pattern, as documentation time was assessed using self-reported documentation time categories rather than objective time measurements [13]. Taken together, these findings suggest that AI-assisted documentation is promising, but that continued evaluation in routine clinical practice remains essential as these technologies evolve.

Acceptance, Usability, and Implementation Considerations

TAM and SUS results indicated high acceptance and perceived usability of the AI-assisted documentation system. High scores for perceived usefulness and ease of use are consistent with established determinants of digital health adoption and suggest that clinicians experienced the system as supportive rather than disruptive to clinical practice [9,10]. Behavioral intention to use also remained high, indicating potential for sustained use beyond the study period.

Implementation of AI-assisted documentation in general practice raises important considerations related to trust, transparency, and data governance. Previous qualitative research has shown that clinicians' willingness to adopt generative AI tools depends on perceived control, reliability, and clarity regarding data handling [9]. Similarly, a recent qualitative study of AI scribes in Norwegian general practice found that clinicians generally perceived ambient documentation systems as supportive of clinical workflows, while also highlighting concerns regarding accuracy, sensitive consultations, and maintaining clinician

XSL•FO
RenderX

oversight. These findings are consistent with the implementation considerations identified in the present study [14]. Similarly, patient perspectives emphasize the importance of transparency and appropriate data governance in the use of AI-based solutions in primary care [15]. The explicit deletion of audio data and short retention period of generated notes in the present implementation may therefore have contributed to both clinician and patient acceptance.

Strengths and Limitations

A key strength of this study is the combination of a pre-post design with blinded external assessment, reducing reliance on self-reported outcomes alone. The use of validated instruments (TAM and SUS) further supports the reliability and interpretability of the findings. Additionally, the study was conducted under routine clinical conditions within a nontechnology-selected group of GPs, enhancing ecological validity. However, formal measures of practice characteristics, including digital maturity, were not collected, limiting the assessment of transferability to other primary care settings.

Several limitations should be considered. The exploratory pre-post design limits causal inference and observed changes may partly reflect learning effects or increased familiarity with documentation practices over time. Because reasons for excluding consultations were not recorded, we were unable to assess whether the case mix differed between the preimplementation and postimplementation periods. The 2-week familiarization period may have been insufficient for all clinicians to fully integrate the system into their workflows, and longer exposure could yield different assessments of usability and efficiency. Documentation time was assessed using self-reported ordinal time categories rather than objectively recorded time measurements. The extent of GP editing of AI-generated drafts was not quantified, and we did not assess lexical characteristics such as note length, number of unique words, or documentation complexity. These measures may provide important complementary information in future evaluations of AI-assisted documentation systems. Consequently, the observed reduction should be interpreted as reflecting clinicians' perceived documentation burden rather than precise time savings in minutes. This distinction is particularly important given that a recent comprehensive review suggests that perceived improvements in documentation burden have been reported more consistently than objectively measured gains in efficiency for AI-assisted documentation systems [13]. Interrater reliability between the 2 blinded external reviewers was low, indicating substantial variability in absolute ratings. In addition, reviewer-specific ratings were not fully concordant, as 1 reviewer rated all domains higher after implementation, whereas the other rated overall quality and structure slightly lower and clinical clarity higher. The blinded external assessment should therefore be interpreted as limited supportive evidence rather than confirmatory evidence of improved documentation quality. We did not formally assess the success of reviewer blinding and therefore cannot exclude the possibility of partial unblinding.

This variability may partly reflect the inherently subjective nature of assessing narrative clinical documentation, even among experienced clinicians. In addition, the 2 reviewers represented different clinical contexts, with 1 working in hospital-based care and the other in general practice. Conceptions of high-quality documentation may therefore differ due to distinct workflows and informational needs rather than differences in documentation quality per se. Hospital-based documentation often emphasizes completeness and detailed clinical information, whereas general practice documentation typically prioritizes concise, structured notes that support rapid overview, continuity of care, and efficient clinical decision-making. Finally, although AI-assisted documentation systems show considerable promise, robust longitudinal real-world evidence for newer large language model–based systems remains limited. Future studies should therefore evaluate not only documentation quality and efficiency but also potential unintended consequences, including effects on clinical reasoning, patient communication, equity, and documentation accuracy as these technologies continue to evolve [13].

Implications and Future Research

Beginning January 1, 2027, Danish patients are expected to have online access to general practice clinical notes in a manner comparable to current access to hospital records. In this context, the quality, structure, and clinical clarity of documentation become increasingly important not only for interprofessional communication but also for patient understanding. The consistently high scores for structure and clinical clarity observed in this study suggest that AI-assisted documentation may support the production of more transparent and patient-readable clinical notes, extending potential benefits beyond efficiency gains alone.

Future research should include controlled study designs, longer follow-up periods, and assessment of patient perspectives, consultation-level characteristics, and clinical outcomes. Future studies should also incorporate objectively measured documentation time alongside clinician-reported outcomes to better distinguish perceived from actual workflow improvements. Further work is needed to evaluate how AI-assisted documentation influences continuity of care, medico-legal documentation quality, interprofessional communication, documentation accuracy, clinical reasoning, and equity across patient populations.

Conclusions

Implementation of AI-assisted clinical documentation in Danish general practice was associated with lower self-reported documentation time categories and no clear deterioration in documentation quality. Clinician acceptance and perceived usability were high. Blinded external assessment provided limited supportive evidence, but low interrater reliability and reviewer-level divergence warrant cautious interpretation. Larger controlled studies with objective time measurement, patient-level and consultation-level data, and longer follow-up are needed to evaluate effectiveness, safety, documentation accuracy, and potential unintended consequences in routine primary care.

XSL•FO
RenderX

Acknowledgments

The authors thank the general practitioners participating in the Aarhus primary care quality cluster for their voluntary participation and engagement in this study. The authors also acknowledge Noteless (Noteless AS) for providing access to the AI-assisted documentation system during the study period and for delivering an introductory online training session at a single cluster meeting. Noteless had no role in study design, selection of outcomes, questionnaire development, data collection, data analysis, interpretation of results, manuscript preparation, or the decision to submit the manuscript for publication. The authors retained full editorial independence. The vendor did not review or approve the manuscript prior to submission and had no access to the study dataset, evaluation data, statistical analyses, or individual-level use data used for the evaluation.

During the preparation of this manuscript, the authors used ChatGPT (OpenAI) for language editing and assistance with manuscript wording. The authors reviewed and revised all AI-assisted content and take full responsibility for the final content of the manuscript.

ChatGPT (OpenAI) was also used to assist in creating the graphical image for the article.

Funding

The authors declare that this study received no external funding.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to privacy and confidentiality considerations related to patient clinical data, but are available from the corresponding author on reasonable request, subject to applicable data protection requirements.

Authors' Contributions

All authors met the criteria for authorship as recommended by the International Committee of Medical Journal Editors (ICMJE) and approved the final manuscript. LNO was responsible for data collection and all communication with participating general practitioners, conducted the data analysis, drafted the manuscript, and approved the final version for publication. PH contributed to the conception and development of the research project, developed the questionnaires administered to the general practitioners, and critically revised the manuscript. ABL contributed to data collection and curation, participated in aspects of the data analysis, and contributed to drafting and critical revision of the manuscript. JL contributed to data collection in clinical practice, interpretation of the results, and drafting and critical revision of the manuscript. MHC conceived and initiated the research project and contributed to study design, interpretation of results, and critical revision of the manuscript.

Conflicts of Interest

None declared.

References

1. Olson KD, Meeker D, Troup M, et al. Use of ambient AI scribes to reduce administrative burden and professional burnout. JAMA Netw Open 2025 Oct 1;8(10):e2534976. [doi: 10.1001/jamanetworkopen.2025.34976] [Medline: 41037268]
2. Bundy H, Gerhart J, Baek S, et al. Can the administrative loads of physicians be alleviated by AI-facilitated clinical documentation? J Gen Intern Med 2024 Nov;39(15):2995-3000. [doi: 10.1007/s11606-024-08870-z] [Medline: 38937369]
3. Rotenstein LS, Apathy N, Holmgren AJ, Bates DW. Physician note composition patterns and time on the EHR across specialty types: a national, cross-sectional study. J Gen Intern Med 2023 Apr;38(5):1119-1126. [doi: 10.1007/s11606-022-07834-5] [Medline: 36418647]
4. Bracken A, Reilly C, Feeley A, Sheehan E, Merghani K, Feeley I. Artificial intelligence (AI) - powered documentation systems in healthcare: a systematic review. J Med Syst 2025 Feb 18;49(1):28. [doi: 10.1007/s10916-025-02157-4] [Medline: 39966286]
5. Guo Y, Wang J, Hu D, et al. Evaluating ambient artificial intelligence documentation: effects on work efficiency, documentation burden, and patient-centered care. J Am Med Inform Assoc 2026 Feb 1;33(2):273-282. [doi: 10.1093/jamia/ocaf180] [Medline: 41100159]
6. Alboksmaty A, Aldakhil R, Hayhoe BW, Ashrafian H, Darzi A, Neves AL. The impact of using AI-powered voice-to-text technology for clinical documentation on quality of care in primary care and outpatient settings: a systematic review. EBioMedicine 2025 Aug;118:105861. [doi: 10.1016/j.ebiom.2025.105861] [Medline: 40694861]
7. Shah SJ, Devon-Sand A, Ma SP, et al. Ambient artificial intelligence scribes: physician burnout and perspectives on usability and documentation burden. J Am Med Inform Assoc 2025 Feb 1;32(2):375-380. [doi: 10.1093/jamia/ocae295] [Medline: 39657021]
8. Falcetta FS, de Almeida FK, Lemos JC, Goldim JR, da Costa CA. Automatic documentation of professional health interactions: a systematic review. Artif Intell Med 2023 Mar;137:102487. [doi: 10.1016/j.artmed.2023.102487] [Medline: 36868684]

XSL•FO
RenderX

9. Blease C, Garcia Sanchez C, Locher C, McMillan B, Gaab J, Torous J. Generative artificial intelligence in primary care: qualitative study of UK general practitioners' views. J Med Internet Res 2025 Aug 6;27:e74428. [doi: 10.2196/74428] [Medline: 40769193]
10. Zhao J, Liu H, Chen Y, Song F. Application of artificial intelligence tools and clinical documentation burden: a systematic review and meta-analysis. BMC Med Inform Decis Mak 2025 Dec 24;26(1):29. [doi: 10.1186/s12911-025-03324-w] [Medline: 41444884]
11. Kuhn T, Basch P, Barr M, Yackel T, Medical Informatics Committee of the American College of Physicians. Clinical documentation in the 21st century: executive summary of a policy position paper from the American College of Physicians. Ann Intern Med 2015 Feb 17;162(4):301-303. [doi: 10.7326/M14-2128] [Medline: 25581028]
12. van Linschoten RC, van Loon CM, Joanknecht L, Bischoff EW. Ambient scribe in general practice: a multi-perspective before-after longitudinal mixed-methods study. NPJ Digit Med 2026 Mar 2;9(1):299. [doi: 10.1038/s41746-026-02454-3] [Medline: 41772212]
13. Laranjo L, Tudor Car L, Payne RE, Neves AL, Kidd M, Jaime Miranda J. Artificial intelligence in primary care: innovation at a crossroads. Lancet Prim Care 2026 Mar;2(3). [doi: 10.1016/j.lanprc.2025.100078] [Medline: 41969643]
14. Nassehi D, Hetlevik Ø, Breivold J. Exploring AI scribes in Norwegian general practice: a qualitative individual interview study. Scand J Prim Health Care 2026 Dec;44(1):2681143. [doi: 10.1080/02813432.2026.2681143] [Medline: 42235527]
15. Mikkelsen JG, Sørensen NL, Merrild CH, Jensen MB, Thomsen JL. Patient perspectives on data sharing regarding implementing and using artificial intelligence in general practice - a qualitative study. BMC Health Serv Res 2023 Apr 4;23(1):335. [doi: 10.1186/s12913-023-09324-8] [Medline: 37016412]

Abbreviations

DPIA: data protection impact assessment
EHR: electronic health record
EU: European Union
GDPR: General Data Protection Regulation
GP: general practitioner
ICC: intraclass correlation coefficient
SUS: System Usability Scale
TAM: technology acceptance model

Edited by A Kushniruk; submitted 01.Mar.2026; peer-reviewed by B Senst, D Wong, U Kirk, V Castillo; revised version received 21.Aug.2026; accepted 08.Sep.2026; published 02.Oct.2026.

Please cite as:
Olsen LN, Harbig P, Laurberg AB, Laurberg J, Charles MH
AI-Assisted Clinical Documentation in Routine Danish General Practice: Quantitative Pre-Post Study
JMIR Hum Factors 2026;13:e94431
URL: https://humanfactors.jmir.org/2026/1/e94431
doi:10.2196/94431

© Louise Nørgaard Olsen, Philipp Harbig, Anna Bay Laurberg, Jacob Laurberg, Morten Haaning Charles. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 2.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Large Language Model–Based Simplification of Digital Therapeutics Explanations for Insomnia and Nicotine Dependence: Two Randomized Online Experiments

JunYoung Seo[1], MS; Moses Yook[1], MS; Dai Jin Kim[2], MD; JunHee Lee[2], MD; Jae Hyun Yoo[2], MD; GiHwan Byeon[2], MD; In Young Choi[1], PhD

[1]Department of Medical Informatics, College of Medicine, The Catholic University of Korea, 222, Banpo-daero, Seocho-gu, Seoul, Republic of Korea
[2]Department of Psychiatry, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

Corresponding Author:
In Young Choi, PhD
Department of Medical Informatics, College of Medicine, The Catholic University of Korea, 222, Banpo-daero, Seocho-gu, Seoul, Republic of Korea

Abstract

Background: Digital therapeutics (DTx) are evidence-based software interventions with the potential to treat health conditions. However, uptake remains limited by low public awareness and overly complex patient education materials that exceed recommended readability levels. Large language models (LLMs) may simplify such content; however, their effects on users' understanding have not been empirically demonstrated.

Objective: This study aimed to examine whether LLM-based simplification of DTx explanatory materials enhances perceived understanding and subjective evaluations of readability, clarity, and comprehensibility compared with manufacturer-provided documents.

Methods: We developed a simplification tool using the GPT-4o application programming interface (API), configured for deterministic outputs and guided by structured readability instructions. Original DTx explanatory materials about insomnia and nicotine dependence were obtained from manufacturers and transformed into simplified versions. Two randomized, between-subject online experiments were conducted (n=1000, with 500 participants in each experiment). Participants were stratified by age and sex and screened for relevance (Insomnia Severity Index ≥8 for the insomnia experiment and smoking ≥5 cigarettes per day for the nicotine dependence experiment). Within each experiment, participants were randomly assigned to review either the original or the LLM-simplified explanatory material. Perceived understanding was assessed before and after exposure. Postexposure evaluations of ease, clarity, and comprehensibility were also collected.

Results: Repeated measures ANOVA revealed significant group×time interaction effects on perceived understanding in both experiments: insomnia ($F_{1,498}$=24.8; P<.001) and nicotine dependence ($F_{1,498}$=14.1; P<.001), with greater improvements observed in the LLM-simplified groups. Mann-Whitney U tests further showed that LLM-simplified explanations were rated as easier, clearer, and more comprehensible than the original versions in both experiments (all P<.05), with small to moderate effect sizes (r=0.11-0.24).

Conclusions: Compared with manufacturer-provided original materials, LLM-simplified DTx explanations led to greater improvements in perceived understanding and subjective evaluations of readability among lay audiences, even after a single exposure. This finding highlights the potential scalability of LLM-based simplification as a strategy to improve the perceived accessibility of health information for lay audiences. Integrating such tools into patient education may enhance how lay audiences perceive and engage with DTx, although further research using objective comprehension measures is needed to confirm these benefits.

Trial Registration: Clinical Research Information Service KCT0011459; https://cris.nih.go.kr/cris/search/detailSearch.do?seq=32221

(JMIR Hum Factors 2026;13:e89451) doi:10.2196/89451

KEYWORDS

digital therapeutics; large language models; text simplification; patient education materials; randomized online experiment

XSL•FO
RenderX

Introduction

Background

Digital therapeutics (DTx) represent a significant advancement in health care and are defined as high-quality software programs delivering evidence-based therapeutic interventions for a wide range of physical, mental, and behavioral conditions [1-3]. Unlike traditional pharmaceuticals and devices, DTx are often cost-effective and designed to encourage sustained user engagement, making them a promising avenue for modern health care [4,5]. DTx have been applied to various conditions, including chronic pain, cardiometabolic diseases (eg, diabetes and hypertension), heart failure, and mental health disorders. They demonstrate potential to complement pharmacotherapy; improve adherence; enable remote monitoring; and expand therapeutic options through emerging technologies, such as virtual reality and wearable sensors [6-10]. In this study, we focused on 2 DTx use cases, insomnia and nicotine dependence, as accessible and representative examples to evaluate the effects of large language model (LLM)–based simplification.

Prior Work

Despite promising applications, the adoption of DTx remains limited [11,12]. Public awareness is low, and many individuals outside the medical community are unfamiliar with DTx or do not fully understand their purpose [13]. For new health technologies to achieve broader societal diffusion, individuals must develop a clear understanding of their value [14-16].

A significant barrier is the inherent complexity of medical information. Research indicates that the average health literacy level of US adults corresponds to that of 13- to 14-year-olds, and medical education materials are recommended to be written at the level of 10- to 11-year-olds [17]. However, over 95% of existing patient education materials require a reading level of at least high school, making them inaccessible to a significant portion of the general population [18]. Producing simplified yet accurate resources remains a resource-intensive and time-consuming process that requires substantial expertise [19-21].

LLMs offer a promising solution. Trained on vast amounts of text data, LLMs can perform natural language tasks, such as summarization, translation, and simplification [22]. Their ability to adapt content to different comprehension levels has been shown to improve the readability of health information, often quantified through automated metrics, such as Flesch-Kincaid grade levels [23-25]. However, these evaluations have rarely examined whether improved readability translates to greater comprehension or more positive user perceptions among lay audiences. This gap highlights the need for empirical studies that directly assess how individuals experience and interpret simplified LLM-generated health information.

Study Objectives

This study aimed to address these gaps by empirically testing whether LLM-based simplification of DTx explanatory materials for insomnia and nicotine dependence enhances perceived understanding and subjective evaluations. Two outcome categories were assessed. The first was the cognitive outcome, which examined whether exposure to LLM-simplified explanations yields greater perceived understanding than exposure to the original explanations. The second was the subjective outcome, which examined whether LLM-based simplification improves evaluations of explanatory materials regarding ease, clarity, and overall comprehensibility.

Specifically, we aimed to determine whether LLM-simplified DTx explanations, compared with the original manufacturer-provided materials, lead to (1) greater improvements in perceived understanding and (2) higher subjective ratings of ease, clarity, and comprehensibility, across 2 independent randomized online experiments.

Methods

Study Design

This study used a randomized, between-subjects online experimental design to evaluate the effectiveness of LLM-simplified DTx explanations. Two independent experiments were conducted, each focusing on a different therapeutic area: insomnia or nicotine dependence.

In both experiments, participants were randomly assigned to one of two groups:

1. Control group—received the original manufacturer-provided DTx explanatory materials
2. Experimental group—received LLM-simplified versions of the same materials

This design allowed direct comparison of the effects of the original versus simplified explanations on participants' perceived understanding and subjective evaluations.

Development of the LLM-based Simplification Tool

To address the challenge of making DTx information more comprehensible to the public while maintaining clinical accuracy, we developed a lightweight text simplification tool. The tool used the GPT-4o (OpenAI) application programming interface (API), a state-of-the-art LLM, for its advanced natural language processing capabilities. The primary goal was to generate simplified text that would be understandable to lay audiences while ensuring fidelity to the original medical content.

In alignment with recommendations for health education materials that suggest a readability level suitable for 10- to 11-year-olds [17], the tool was specifically designed to simplify complex DTx explanations to this level, ensuring accessibility across diverse populations regardless of health literacy.

The simplification process was guided by structured instructions covering content preservation, logical coherence, terminology adaptation, clarity enhancement, tone and style, emphasis, length management, and readability. The full prompt text provided to the LLM is available in Multimedia Appendix 1.

For model configuration, the temperature parameter was set to 0 to minimize randomness and maximize reproducibility, and the top p parameter was fixed at 1 to ensure comprehensive consideration of relevant tokens. This deterministic setup was critical for maintaining precision and consistency in the medical information. These design choices enabled the creation of a tool

XSL•FO
RenderX

that converts complex DTx documents into simplified, patient-friendly explanations without compromising medical accuracy.

DTx Explanatory Materials

To evaluate the generalizability of the proposed text simplification approach, we selected 2 DTx that addressed distinct health conditions: insomnia and nicotine dependence. This design enabled the assessment of the applicability of the LLM-based simplification tool across different therapeutic domains, rather than within a single disease context.

For each DTx, original patient education and explanatory documents were obtained directly from the respective manufacturers. These materials, characterized by high complexity and technical terminology, served as baseline texts for the simplification process. Using the LLM-based tool described above, simplified versions of each document were generated, resulting in two parallel sets of materials for each DTx product: (1) the original manufacturer-provided version and (2) the LLM-simplified version.

The LLM-simplified versions incorporated several modifications, including vocabulary simplification, tone adjustment, and formatting improvements, such as clearer paragraph structuring and emphasis of key points. In contrast, the original manufacturer-provided materials were presented in their original plain-text format without any modification.

To ensure medical accuracy and logical coherence, all LLM-generated materials were independently reviewed by a panel of 3 board-certified psychiatrists who regularly treat conditions such as insomnia and nicotine dependence. This expert validation confirmed that the simplified content faithfully preserved the original medical meaning, avoided factual errors or misinterpretations, and presented information in a logically coherent manner. This review process was essential to maintaining the clinical integrity and trustworthiness of the materials used in this study. The full texts of the original and LLM-simplified explanatory materials are provided in Multimedia Appendix 2.

Participants and Recruitment

In total, 1000 participants were recruited through an online panel service (Macromill Embrain Co, Ltd) [26], with 500 participants allocated to each experiment (n=500, 50% to insomnia DTx and n=500, 50% to nicotine dependence DTx). Stratified assignment was applied during recruitment to ensure comparable sex and age distributions between the control and experimental groups.

The general inclusion criteria were (1) age between 19 and 64 years, (2) access to a smartphone or computer with internet connectivity, and (3) provision of informed consent before participation. Condition-specific screening thresholds were established in consultation with 3 board-certified psychiatrists to enhance contextual relevance while recruiting participants from the general population rather than from patient groups. For the insomnia DTx experiment, only participants with an Insomnia Severity Index score of ≥8 were included, corresponding to at least mild insomnia symptoms [27]. The Insomnia Severity Index is a 7-item self-report scale assessing insomnia symptom severity, with total scores ranging from 0 to 28. For the nicotine dependence DTx experiment, only participants who reported smoking at least 5 cigarettes per day were eligible. These criteria were selected to ensure that participants, although not recruited as patients, shared health-related characteristics that made the DTx materials meaningful and relevant to their personal contexts.

Procedure

This study used an online, randomized, between-subjects experimental design. Participants were recruited through an online panel service and screened according to the general and condition-specific inclusion criteria described above. Only individuals who met eligibility requirements were invited to participate.

After confirming eligibility, participants provided digital informed consent. In each experiment, participants were recruited from an online panel service [26] that collected demographic information (age and sex) during the screening process. To ensure balance across conditions, stratified randomization was applied, assigning participants to either the control group (original materials) or the experimental group (LLM-simplified materials) within each demographic stratum. This approach minimized allocation bias and maintained between-group comparability.

Following the group assignment, participants completed a structured online questionnaire in the following sequence:

- Pre-exposure assessment of perceived understanding of the respective DTx
- Exposure to the assigned explanatory materials (original vs LLM-simplified) with a minimum enforced viewing time of 30 seconds
- Postexposure assessment of perceived understanding of the respective DTx
- Evaluation of explanations, where participants rated the difficulty, clarity, and overall comprehensibility of the materials

All responses were anonymized to ensure the strict protection of participant privacy. After completing the survey, participants received compensation according to the agreement established by the online panel service.

Measures

Perceived Understanding

Participants' perceived understanding of the 2 DTx was assessed both before and after exposure to the explanatory materials. This construct was measured using 5 items adapted from Flynn and Goldsmith subjective knowledge scale [28], modified to capture how much participants believed they understood each DTx. Responses were recorded on a 5-point Likert scale ranging from 0 (not at all) to 4 (very much). The exact wording of the 5 adapted items is provided in the Multimedia Appendix 3. The scale demonstrated good internal consistency, with Cronbach α=0.80 for the insomnia DTx experiment and α=0.82 for the nicotine dependence DTx experiment.

XSL•FO
RenderX

Evaluation of Explanations

Participants' evaluation of the explanatory materials was measured after exposure using single-item ratings on a 0 to 4 scale, where 0 indicated "not at all" and 4 indicated "very much." Three aspects were assessed: (1) ease (how easy or difficult the explanation was to read and follow), (2) clarity (how clear the explanation was perceived overall), and (3) comprehensibility (the extent to which participants felt they could understand the explanation).

All questionnaires were administered in Korean. No objective knowledge tests (eg, recall or factual questions) were included; thus, all outcome measures relied on participants' subjective self-reports.

Ethical Considerations

This study was approved by the Institutional Review Board of the Catholic Medical Center (approval MC25QNSI0039) and was conducted in accordance with the principles of the Declaration of Helsinki [29]. All participants provided informed consent before participation. All participant data were anonymized and kept strictly confidential; no personally identifiable information was collected or retained. Participants received compensation for their participation in accordance with the terms established by the online panel service [26]. This study is reported in accordance with the Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth (CONSORT-EHEALTH) guidelines; the completed checklist is provided in Checklist 1.

Statistical Analysis

All statistical analyses were conducted using Jamovi (version 2.6.26). A significance threshold of α=.05 was applied to all tests. To examine the effect of explanation type on perceived DTx understanding, repeated measures ANOVA was conducted. Each model tested the group (LLM-simplified vs control)×time (before vs after exposure) interaction to determine whether changes in perceived understanding over time significantly differed between groups. The assumptions of normality, homoscedasticity, and sphericity were satisfied.

Participants' postexposure ratings of ease, clarity, and comprehensibility of the DTx explanations were compared between groups using the Mann-Whitney *U* test. This nonparametric approach was selected because of the ordinal nature of the single-item Likert-scale ratings. These analyses evaluated whether LLM-based simplification had a significant influence on participants' subjective evaluations of the materials.

Results

Sample Characteristics

A total of 1000 participants completed the study, with 500 (50%) allocated to the insomnia DTx experiment and 500 (50%) to the nicotine dependence DTx experiment. Stratified assignment ensured comparable sex and age distributions between the control and experimental groups. Demographic information used for stratification was limited to sex and age. Participant demographic characteristics are presented in Table 1.

Table . Participant demographic characteristics by group (insomnia and nicotine dependence experiments).

Characteristic	Insomnia control (n=250)	Insomnia test (n=250)	Nicotine control (n=250)	Nicotine test (n=250)
Sex, n (%)				
Male	125 (50)	125 (50)	125 (50)	125 (50)
Female	125 (50)	125 (50)	125 (50)	125 (50)
Age (years), n (%)				
20	63 (25.2)	63 (25.2)	63 (25.2)	63 (25.2)
30	63 (25.2)	63 (25.2)	63 (25.2)	63 (25.2)
40	62 (24.8)	62 (24.8)	62 (24.8)	62 (24.8)
50-65	62 (24.8)	62 (24.8)	62 (24.8)	62 (24.8)

Effect on Perceived Understanding

No significant main effect of group was observed in either the insomnia experiment ($F_{1,498}$=3.32; P=.07; partial η^2=0.007) or the nicotine dependence experiment ($F_{1,498}$=2.88; P=.09; partial η^2=0.006; Table 2).

XSL•FO
RenderX

Table . Repeated measures ANOVA results for perceived understanding across insomnia and nicotine dependence digital therapeutics (DTx) experiments.

Experiment and effect	*F* test (*df*)	*P* value	Partial η^2
Insomnia DTx			
Group	3.32 (1, 498)	.07	0.007
Time	38.5 (1, 498)	<.001	0.072
Group×time	24.8 (1, 498)	<.001	0.048
Nicotine DTx			
Group	2.88 (1, 498)	.09	0.006
Time	21.5 (1, 498)	<.001	0.041
Group×time	14.1 (1, 498)	<.001	0.028

A significant main effect of time was observed in both the insomnia experiment ($F_{1,498}$=38.5; *P*<.001; partial η^2=0.072) and the nicotine dependence experiment ($F_{1,498}$=21.5, *P*<.001; partial η^2=0.041), with postexposure assessment scores higher than pre-exposure assessment scores across groups.

A significant group×time interaction was observed in both experiments. In the insomnia experiment, the interaction effect was $F_{1, 498}$=24.8; *P*<.001; partial η^2=0.048; and in the nicotine dependence experiment, it was $F_{1, 498}$=14.1; *P*<.001; partial η^2=0.028 (Table 2).

Descriptive statistics for each group and time point are presented in Table 3. As shown in Figure 1, the LLM-simplified group demonstrated greater increases in perceived understanding than the control group in both experiments.

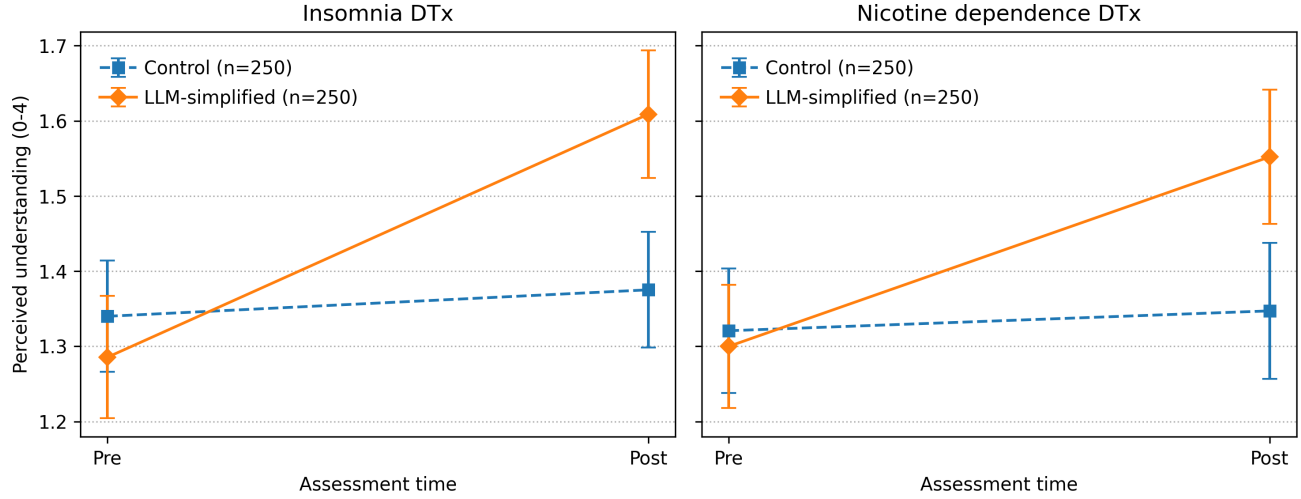
Table . Perceived understanding scores by group and time point (n=250).

Experiment and group	Pretest, mean (SD; 95% CI)	Posttest, mean (SD; 95% CI)
Insomnia DTx[a]		
LLM[b]-simplified	1.29 (0.65; 1.20-1.37)	1.61 (0.68; 1.52-1.69)
Control	1.34 (0.59; 1.27-1.41)	1.38 (0.62; 1.30-1.45)
Nicotine DTx		
LLM-simplified	1.30 (0.66; 1.22-1.38)	1.55 (0.72; 1.46-1.64)
Control	1.32 (0.66; 1.24-1.40)	1.35 (0.73; 1.26-1.44)

[a]DTx: digital therapeutics.

[b]LLM: large language model.

Figure 1. Changes in perceived digital therapeutics (DTx) understanding from pre-exposure assessment to postexposure assessment in both experiments. Points indicate group means, and error bars represent 95% CIs. LLM: large language model.



XSL•FO
RenderX

Evaluation of Explanations

In the insomnia experiment, the LLM-simplified explanations were rated significantly higher for ease (U=27,281; P=.008; r=0.127), clarity (U=27,803; P=.02; r=0.110), and comprehensibility (U=25,125; P<.001; r=0.196). Similarly, in the nicotine dependence experiment, significant group differences were observed for ease (U=23,727; P<.001; r=0.240), clarity (U=27,697; P=.02; r=0.114), and comprehensibility (U=24,738; P<.001; r=0.208). The distribution of ratings across response categories is illustrated in Figure 2.

Figure 2. Group distribution of postexposure ratings (0 - 4) for ease, clarity, and comprehensibility in both experiments. Bars represent 100% stacked proportions. The neutral category (2) is split and plotted on both sides of the center, such that lower ratings (0 - 1) extend to the left of 0 and higher ratings (3-4) extend to the right. This split-neutral presentation allows the direction of responses (negative vs positive) to be visually compared while preserving the full distribution of ratings. Group labels indicate sample sizes (control vs large language model [LLM]–simplified). Between-group differences were tested using 2-sided Mann-Whitney U tests with Holm correction across measures; rank-biserial effect sizes (r) are reported in Table 4. DTx: digital therapeutics.

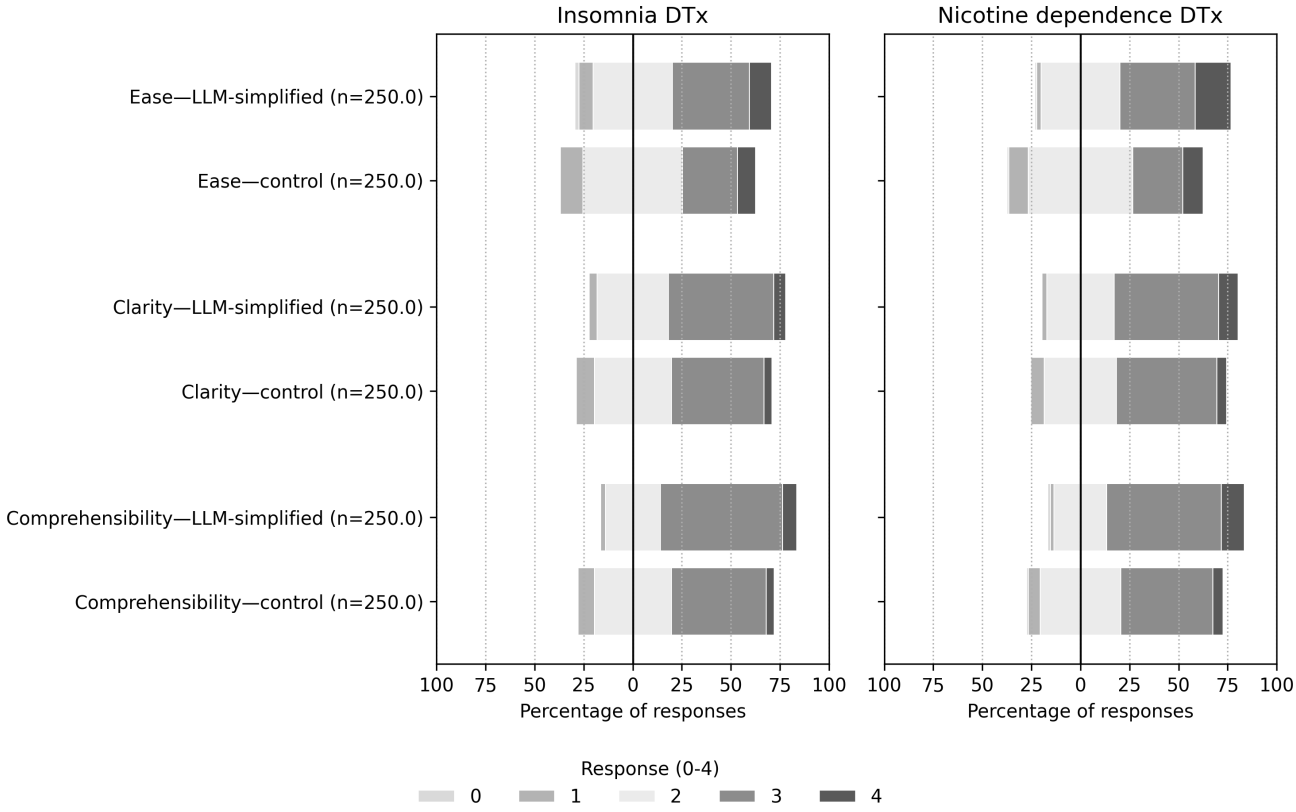


Table . Mann-Whitney U test results comparing subjective evaluation of explanations between groups.

Experiment and measure	U	P value	Adjusted P value[a]	r
Insomnia DTx[b]				
Ease	27,281	.008	.02	0.127
Clarity	27,803	.02	.03	0.11
Comprehensibility	25,125	<.001	.006	0.196
Nicotine DTx				
Ease	23,727	<.001	.006	0.24
Clarity	27,697	.02	.03	0.114
Comprehensibility	24,738	<.001	.006	0.208

[a]Holm correction applied for multiple comparisons.

[b]DTx: digital therapeutics.

Discussion

This study evaluated whether LLM-based simplification of DTx explanatory materials improves perceived understanding and subjective evaluations among lay audiences through 2 independent randomized online experiments targeting insomnia (500/1000, 50%) and nicotine dependence (500/1000, 50%).

XSL•FO
RenderX

Principal Results

Our findings indicated that LLM-based simplification of DTx explanatory materials improves perceived comprehensibility and subjective evaluations of accessibility. In 2 independent experiments on insomnia and nicotine dependence, participants who received LLM-simplified explanations reported significantly greater gains in perceived understanding than those who received the original manufacturer-provided explanations. In addition, the simplified versions were consistently rated as easier, clearer, and more comprehensible, indicating improved subjective evaluations. These findings suggest that LLM-based simplification may improve perceived understanding and subjective evaluations across multiple therapeutic contexts. However, further research using objective comprehension measures and more diverse populations is needed to confirm the robustness and generalizability of these findings.

Although these findings demonstrate improvements in perceived understanding, these improvements do not necessarily indicate accurate or complete comprehension of medical information. Feeling confident about understanding health information may differ from actually understanding the details and appropriate use of a medical intervention. Therefore, future research should incorporate objective knowledge measures to determine whether LLM-based simplification improves actual comprehension in addition to perceived understanding.

The observed differences in subjective evaluations were statistically significant, but the effect sizes were generally small to moderate (r=0.11 - 0.24). These magnitudes suggest modest but consistent improvements in participants' evaluations following a single brief exposure. Given the minimal intervention required, such improvements may still be meaningful in real-world settings where scalable approaches to improving the perceived accessibility of health information are needed.

These findings emphasize the potential for safe and practical applications of LLMs in health care communication. Although concerns about hallucinations highlight the risks of using LLMs to generate new clinical content or make medical decisions [30], their role in transforming clinically validated information into more accessible formats carries considerably lower risk [31]. Prior research has demonstrated that readability can be improved while maintaining factual accuracy [23-25]. This study extends that evidence by incorporating independent reviews from 3 board-certified psychiatrists who confirmed the medical accuracy and logical coherence of the simplified texts. Collectively, these results demonstrate that LLM-based simplification represents a scalable and trustworthy strategy for improving the perceived accessibility and comprehensibility of DTx information among lay audiences.

Comparison With Prior Work

The findings of this study align with a substantial body of research that highlights the challenges of health literacy and the frequent mismatch between the readability of medical information and patients' comprehension skills [17]. Previous studies have consistently shown that most patient education materials are written at a level above the recommended reading level, limiting accessibility for the general population [18]. Recent advances in natural language processing and LLMs suggest that automated text simplification can serve as a promising solution, with evidence demonstrating improvements in readability across a range of health-related documents [22]. However, much of this prior work relied on computational readability metrics (such as Flesch-Kincaid scores) rather than empirical evaluations with end users [23-25].

This study extends this line of research by providing experimental evidence that LLM-simplified explanations not only achieve higher readability but also enhance perceived understanding among lay audiences. Notably, the results demonstrate that accuracy can be preserved, consistent with prior findings that LLM-based simplification improves accessibility without compromising factual correctness [23-25]. Unlike earlier studies, this study combines expert clinical validation by board-certified psychiatrists with large-scale user experiments across 2 therapeutic contexts, thereby offering novel empirical evidence of both accuracy and effectiveness. These findings reinforce the growing recognition that LLM-based simplification may help improve the perceived accessibility of DTx information and support users' perceived understanding of complex health technologies.

Limitations

This study has some limitations. First, participants were recruited from an online panel and allocated through stratified randomization based on sex and age, which may have reduced natural variability and limited the generalizability of the results [32]. Because participation required internet access and familiarity with digital environments, the study sample may have had higher digital literacy than typical patient populations. Therefore, the findings may not fully generalize to individuals with limited digital literacy. In addition, because demographic information was limited to sex and age, other potentially relevant characteristics (such as education, income, health literacy, and prior DTx exposure) were not collected, further limiting the interpretability of the results. In particular, differences in education level and baseline health literacy may influence how individuals respond to simplified health information and should be considered when interpreting the generalizability of the findings. Second, comprehension was assessed only through self-reported perceived understanding rather than objective knowledge tests [33], which limits conclusions about actual comprehension. Third, ease, clarity, and comprehensibility were evaluated using single-item measures, restricting measurement reliability. Fourth, the short-term exposure with a minimum enforced viewing time of 30 seconds leaves the durability and behavioral relevance of the effects uncertain. Fifth, although participants were screened for relevance (eg, minimum insomnia symptoms or daily smoking frequency), they were not clinically diagnosed; thus, caution is warranted when extrapolating these findings to treatment populations. Sixth, the intervention relied on a single LLM (GPT-4o) and specific prompt configurations, underscoring the need for replication using alternative models and implementation conditions to establish robustness. Finally, the intervention involved multiple simultaneous modifications, including vocabulary simplification, tone adjustment, and formatting improvements, such as paragraph structuring and

XSL•FO
RenderX

emphasis of key points. Notably, the original manufacturer-provided materials were presented in plain text without visual enhancements, whereas the LLM-simplified versions incorporated formatting elements, such as bolding, structured paragraph breaks, and emphasis of key terms, as part of the simplification process. Therefore, the comparison was not purely linguistic; differences in visual presentation may have independently contributed to the observed improvements in perceived comprehensibility and subjective evaluations. This constitutes a bundled intervention, and the present design does not allow for disentangling the effects of linguistic simplification from those of visual formatting. Future studies should isolate these components—for example, through factorial designs that independently vary linguistic complexity and visual formatting—to better understand their individual and combined effects.

Conclusions

This study demonstrates that LLM-based simplification of DTx explanatory materials significantly improves perceived understanding and subjective evaluations of readability, clarity, and comprehensibility among lay audiences. By focusing on the transformation of clinically validated content, LLMs can provide a safe and scalable approach to improving the perceived accessibility of digital health information.

Beyond DTx, these findings have broader implications: many types of medical documents, including patient education materials, informed consent forms, and clinical instructions, can be made to feel more accessible through LLM-based simplification. Importantly, by helping patients feel more confident in engaging with health information, LLM-based tools may have the potential to reduce passive compliance and instead promote active engagement, adherence, and shared decision-making.

Future studies should extend these results in several ways. Research involving clinically diagnosed patient populations is necessary to confirm the applicability of this approach in real-world treatment settings. Additionally, objective knowledge assessments and longitudinal designs are needed to determine whether improvements in perceived understanding translate into sustained learning and behavioral change. Finally, evaluations across diverse health conditions, populations, and languages are essential to establish the robustness and generalizability of LLM-based simplification strategies.

Acknowledgments

The authors thank Aimnext Co, Ltd and Innerwave Co, Ltd for their cooperation in providing the original explanatory materials for the digital therapeutics evaluated in this study.

Funding

This work was supported by the Bio-Industrial Technology Development Program (RS-2022-KT217960; Development of an online platform for DTx full-cycle clinical trial support using metaverse technology) funded by the Ministry of Trade, Industry and Resources (MOTIR, Korea). This funding covered both the research process and the article processing charge. The funder had no role in the design of the study, data collection, analysis, interpretation, or manuscript preparation.

Data Availability

The datasets generated and/or analyzed during this study are not publicly available due to participant privacy considerations but are available from the corresponding author on reasonable request, subject to renewed institutional review board approval.

Authors' Contributions

Conceptualization: JYS (lead), IYC (supporting)
Data curation: JYS (lead)
Formal analysis: JYS (lead), MY (supporting)
Funding acquisition: DJK (lead), IYC (supporting)
Investigation: JYS (lead)
Methodology: JYS (lead), MY (supporting)
Supervision: IYC (lead), MY (supporting)
Validation: JHY (equal), JHL (equal), GHB (equal; clinical consultation and review of the large language model–simplified materials)
Writing—original draft: JYS (lead)
Writing—review and editing: JYS (lead), IYC (supporting), MY (supporting), DJK (supporting), JHY (supporting), JHL (supporting), GHB (supporting)
All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Full prompt text used for the large language model.
[DOCX File, 17 KB - humanfactors_v13i1e89451_app1.docx]

Multimedia Appendix 2
Original and large language model–simplified explanatory materials.
[DOCX File, 18 KB - humanfactors_v13i1e89451_app2.docx]

Multimedia Appendix 3
Survey questionnaire.
[DOCX File, 19 KB - humanfactors_v13i1e89451_app3.docx]

Checklist 1
CONSORT-EHEALTH (V 1.6.1) checklist for reporting randomized trials of eHealth interventions.
[PDF File, 1295 KB - humanfactors_v13i1e89451_app4.pdf]

References

1. Hong JS, Wasden C, Han DH. Introduction of digital therapeutics. Comput Methods Programs Biomed 2021 Sep;209:106319. [doi: 10.1016/j.cmpb.2021.106319] [Medline: 34364181]
2. Palanica A, Docktor MJ, Lieberman M, Fossat Y. The need for artificial intelligence in digital therapeutics. Digit Biomark 2020;4(1):21-25. [doi: 10.1159/000506861] [Medline: 32399513]
3. Digital Therapeutics Alliance. URL: https://dtxalliance.org/ [accessed 2025-04-28]
4. Li Y, Gong Y, Zheng B, et al. Effects on adherence to a mobile app-based self-management digital therapeutics among patients with coronary heart disease: pilot randomized controlled trial. JMIR Mhealth Uhealth 2022 Feb 15;10(2):e32251. [doi: 10.2196/32251] [Medline: 34906924]
5. Sapanel Y, Tadeo X, Brenna CT, et al. Economic evaluation associated with clinical-grade mobile app-based digital therapeutic interventions: systematic review. J Med Internet Res 2023 Aug 1;25:e47094. [doi: 10.2196/47094] [Medline: 37526973]
6. Biskupiak Z, Ha VV, Rohaj A, Bulaj G. Digital therapeutics for improving effectiveness of pharmaceutical drugs and biological products: preclinical and clinical studies supporting development of drug + digital combination therapies for chronic diseases. J Clin Med 2024 Jan 11;13(2):403. [doi: 10.3390/jcm13020403] [Medline: 38256537]
7. Rohaj A, Bulaj G. Digital therapeutics (DTx) expand multimodal treatment options for chronic low back pain: the nexus of precision medicine, patient education, and public health. Healthcare (Basel) 2023 May 18;11(10):1469. [doi: 10.3390/healthcare11101469] [Medline: 37239755]
8. Singh B, Hazra P, Roy S, et al. Exploring the need and benefits of digital therapeutics (DTx) for the management of heart failure in India. Cureus 2023 Nov;15(11):e49628. [doi: 10.7759/cureus.49628] [Medline: 38161874]
9. Nordyke RJ, Appelbaum K, Berman MA. Estimating the impact of novel digital therapeutics in type 2 diabetes and hypertension: health economic analysis. J Med Internet Res 2019 Oct 9;21(10):e15814. [doi: 10.2196/15814] [Medline: 31599740]
10. Ferreira-Brito F, Ferreira J, Vieira M, Guerreiro J, Guerreiro T. Digital therapeutics with virtual reality and sensors. In: UbiComp/ISWC '23 Adjunct: Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing: Association for Computing Machinery; 2023.
11. Armeni P, Polat I, De Rossi LM, Diaferia L, Meregalli S, Gatti A. Exploring the potential of digital therapeutics: an assessment of progress and promise. Digit Health 2024;10:20552076241277441. [doi: 10.1177/20552076241277441] [Medline: 39291152]
12. Survey on perception of digital therapeutic devices. Health Insurance Review & Assessment Service. 2024. URL: https://repository.hira.or.kr/handle/2019.oak/3208 [accessed 2026-06-02]
13. Sim B, Ju JH, Kim B, Lee JY. Stakeholders' perceptions regarding digital therapeutics reimbursement in South Korea: qualitative study. JMIR Mhealth Uhealth 2023 Oct 30;11:e47407. [doi: 10.2196/47407] [Medline: 37933414]
14. Feder G, Slade R. The acquisition of information and the adoption of new technology. Am J Agric Econ 1984 Aug;66(3):312-320. [doi: 10.2307/1240798]
15. Straub ET. Understanding technology adoption: theory and future directions for informal learning. Rev Educ Res 2009 Jun;79(2):625-649. [doi: 10.3102/0034654308325896]
16. Tverskoi D, Babu S, Gavrilets S. The spread of technological innovations: effects of psychology, culture and policy interventions. R Soc Open Sci 2022 Jun;9(6):211833. [doi: 10.1098/rsos.211833] [Medline: 35754991]
17. Weiss BD. Health literacy: a manual for clinicians. American Medical Association Foundation and American Medical Association. 2003. URL: http://lib.ncfh.org/pdfs/6617.pdf [accessed 2025-04-28]

XSL•FO
RenderX

18. Morony S, Flynn M, McCaffery KJ, Jansen J, Webster AC. Readability of written materials for CKD patients: a systematic review. Am J Kidney Dis 2015 Jun;65(6):842-850. [doi: 10.1053/j.ajkd.2014.11.025] [Medline: 25661679]
19. Dickinson DK, McCabe A. Bringing it all together: the multiple origins, skills, and environmental supports of early literacy. Learn Disabil Res Pract 2001 Nov;16(4):186-202. [doi: 10.1111/0938-8982.00019]
20. Gal I. Towards probability literacy for all citizens: building blocks and instructional dilemmas. In: Jones GA, editor. Exploring Probability in School: Springer; 2005.
21. Sheppard ED, Hyde Z, Florence MN, McGwin G, Kirchner JS, Ponce BA. Improving the readability of online foot and ankle patient education materials. Foot Ankle Int 2014 Dec;35(12):1282-1286. [doi: 10.1177/1071100714550650] [Medline: 25239196]
22. Abd-Alrazaq A, AlSaad R, Alhuwail D, et al. Large language models in medical education: opportunities, challenges, and future directions. JMIR Med Educ 2023 Jun 1;9:e48291. [doi: 10.2196/48291] [Medline: 37261894]
23. Kirchner GJ, Kim RY, Weddle JB, Bible JE. Can artificial intelligence improve the readability of patient education materials? Clin Orthop Relat Res 2023 Nov 1;481(11):2260-2267. [doi: 10.1097/CORR.0000000000002668] [Medline: 37116006]
24. Rouhi AD, Ghanem YK, Yolchieva L, et al. Can artificial intelligence improve the readability of patient education materials on aortic stenosis? A pilot study. Cardiol Ther 2024 Mar;13(1):137-147. [doi: 10.1007/s40119-023-00347-0] [Medline: 38194058]
25. Patel EA, Fleischer L, Filip P, et al. The use of artificial intelligence to improve readability of otolaryngology patient education materials. Otolaryngol Head Neck Surg 2024 Aug;171(2):603-608. [doi: 10.1002/ohn.816] [Medline: 38751109]
26. Embrain. URL: https://www.embrain.com [accessed 2026-05-02]
27. Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001 Jul;2(4):297-307. [doi: 10.1016/s1389-9457(00)00065-4] [Medline: 11438246]
28. Flynn LR, Goldsmith RE. A short, reliable measure of subjective knowledge. J Bus Res 1999 Sep;46(1):57-66. [doi: 10.1016/S0148-2963(98)00057-5]
29. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
30. Bélisle-Pipon JC. Why we need to be careful with LLMs in medicine. Front Med (Lausanne) 2024;11:1495582. [doi: 10.3389/fmed.2024.1495582] [Medline: 39697212]
31. Bojić L, Zagovora O, Zelenkauskaite A, et al. Comparing large language models and human annotators in latent content analysis of sentiment, political leaning, emotional intensity and sarcasm. Sci Rep 2025 Apr 3;15(1):11477. [doi: 10.1038/s41598-025-96508-3] [Medline: 40181141]
32. Bethlehem J. Selection bias in web surveys. Int Stat Rev 2010 Aug;78(2):161-188. [doi: 10.1111/j.1751-5823.2010.00112.x]
33. Waters EA, Biddle C, Kaphingst KA, et al. Examining the interrelations among objective and subjective health literacy and numeracy and their associations with health knowledge. J Gen Intern Med 2018 Nov;33(11):1945-1953. [doi: 10.1007/s11606-018-4624-2] [Medline: 30120636]

Abbreviations

API: application programming interface
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
DTx: digital therapeutics
LLM: large language model

Edited by Y Quintana; submitted 23.Dec.2025; peer-reviewed by A Rouhi, O Loubani; revised version received 01.May.2026; accepted 02.May.2026; published 10.Jun.2026.

Please cite as:
Seo J, Yook M, Kim DJ, Lee J, Yoo JH, Byeon G, Choi IY
Large Language Model–Based Simplification of Digital Therapeutics Explanations for Insomnia and Nicotine Dependence: Two Randomized Online Experiments
JMIR Hum Factors 2026;13:e89451
URL: https://humanfactors.jmir.org/2026/1/e89451
doi:10.2196/89451

© JunYoung Seo, Moses Yook, Dai Jin Kim, JunHee Lee, Jae Hyun Yoo, GiHwan Byeon, In Young Choi. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 10.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use,

XSL•FO
RenderX

distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Review

Associations of Health Literacy and eHealth Literacy With Patient Portal Access and Use: Systematic Review

Philipp Haessner[1], MBA, MS; Mahitha Abraham[2*]; Aleena Peter[2*]; Abigail Rosario[2*]; Samika Kanaskar[3*]; Crystal L Romero[4], MD, MPH; Megan Gregory[1], PhD

[1]Department of Health Outcomes & Biomedical Informatics, College of Medicine, University of Florida, Gainesville, FL, United States

[2]College of Liberal Arts and Sciences, University of Florida, Gainesville, FL, United States

[3]College of Public Health and Health Professions, University of Florida, Gainesville, FL, United States

[4]Department of Community Health and Family Medicine, College of Medicine, University of Florida, Gainesville, FL, United States

[*]these authors contributed equally

Corresponding Author:
Megan Gregory, PhD
Department of Health Outcomes & Biomedical Informatics
College of Medicine
University of Florida
PO Box 100147
Gainesville, FL, 32610
United States
Phone: 1 352 294 8126
Email: megan.gregory@ufl.edu

Abstract

Background: Effective patient portal access and use require both digital navigation and understanding of clinical information. Health literacy and eHealth literacy reflect these different demands. Although the 2 constructs overlap, prior research often treats them as interchangeable. It remains unclear whether health literacy and eHealth literacy are associated differently across stages of portal engagement and patterns of portal use. A systematic review is therefore needed to identify consistent patterns across studies.

Objective: This study aimed to synthesize associations of health literacy and eHealth literacy with patient portal access, general use, feature-specific use, and frequency of use.

Methods: We searched PubMed, Scopus, CINAHL, PsycINFO, Web of Science, and Embase from inception to March 2026 using keywords related to health literacy, eHealth literacy, patient portals, and portal access or use. We included English-language peer-reviewed articles and gray literature involving adults who used real-world patient portals, measured health literacy or eHealth literacy using validated instruments, and reported quantitative associations with portal access or use. Because studies differed substantially in literacy instruments, portal outcomes, and outcome measures, we used direction-based vote counting, complemented by narrative synthesis. Reporting followed the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) statement, whereas synthesis methods were guided by the Cochrane Handbook for synthesis without meta-analysis. For full-text, peer-reviewed journal articles, we assessed the risk of bias by using the Joanna Briggs Institute (JBI) critical appraisal tools.

Results: Among 1107 screened studies, 30 met the inclusion criteria. Of these, 17 examined health literacy only, 10 examined eHealth literacy only, and 3 examined both constructs in the same sample. Studies focused far more on portal use than on portal access and emphasized general portal use over feature-specific outcomes. Any use was assessed more often than frequency of use. Nine of 10 directionally codable eHealth literacy studies of general portal use showed lower use among individuals with lower eHealth literacy, whereas health literacy findings for general portal use were mixed. Lower health literacy showed a more consistent pattern of lower portal access and several feature-specific outcomes, although comparable eHealth literacy evidence for these outcomes was limited. Most studies met the majority of JBI criteria, although most measured portal use via self-report rather than objective system data.

Conclusions: Nearly every directionally codable eHealth literacy study reported lower portal use among patients with lower eHealth literacy. Lower health literacy showed similar patterns for portal access and several feature-specific outcomes. These consistent associations point to literacy-related disparities at multiple stages of portal engagement, although comparable

XSL•FO
RenderX

within-sample evidence remains limited. Future studies should measure both constructs in the same sample, use comparable modeling approaches, use objective portal log data when possible, and distinguish access, any use, frequency, and feature-specific behaviors to identify where literacy-related disparities emerge.

(JMIR Hum Factors 2026;13:e104485) doi:10.2196/104485

KEYWORDS

patient portals; health literacy; eHealth literacy; digital health; patient participation; systematic review

Introduction

A patient portal is a secure web or mobile interface linked to the electronic health record that provides patients with access to their health information, including laboratory results, visit summaries, immunization records, scheduling, and vital signs [1-3]. In 2024, about 77% of US adults reported being offered online access to their medical records, compared with 52% in 2017 [4]. Despite broader availability, access alone does not ensure meaningful ongoing portal use.

To benefit from a portal, patients must be able to register, log in, navigate relevant features, interpret clinical content, and act on what they read. These tasks involve both digital navigation and comprehension of complex health information [5]. When patients use portals effectively, benefits include better engagement in care, stronger communication with clinicians, and improved management of chronic and preventive care needs [2,6,7]. Yet, the most commonly reported barriers to portal use remain insufficient digital skills or difficulty understanding medical content [8], which indicates that barriers persist even after access is established.

Health literacy and eHealth literacy map directly onto these 2 barriers. Health literacy refers to the ability to obtain, process, and understand health information to make appropriate health-related decisions [9]. Low health literacy affects about one-third of US adults [10] and increases the risk of misunderstanding health information [11]. eHealth literacy was originally defined as the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to a health problem [12]. More recent work has expanded this concept to include the ability to engage with digital technologies in effective, safe, and helpful ways to achieve health goals [13]. eHealth literacy therefore extends beyond information seeking to include the skills needed to use digital health tools, such as patient portals [12].

Although health literacy and eHealth literacy overlap, the 2 are not identical [14]. The Centers for Disease Control and Prevention has cautioned against using them interchangeably [15]. Health literacy focuses primarily on understanding health information and how to act on it [16], whereas eHealth literacy adds the digital skills needed to work with online health information and digital health tools [12,13]. A recent systematic review and meta-analysis found only a weak positive correlation between the 2 constructs (r=0.29), suggesting they capture distinct skill domains rather than variations on a single construct [17]. In patient portal research, this distinction matters: some patients may struggle mainly with digital navigation, others with interpreting the clinical information they find inside the portal.

A second challenge is how portal engagement is measured. Prior research usually defines portal use as whether patients used the portal at all, with less attention to how often they returned [18]. Yet one-time use and sustained use reflect different levels of engagement [3,18]. A patient might register and log in once but never return. Studies focusing only on any use may therefore miss whether literacy-related barriers affect initial uptake, continued use, or both.

Existing reviews have focused mainly on health literacy, giving limited attention to eHealth literacy, and have not clearly separated gaining access to a portal from using it after access is established [3,19]. To our knowledge, no prior review has examined how health literacy and eHealth literacy together relate to portal access, any use, frequency of use, and feature-specific use. Given these gaps, a systematic review is needed to integrate the available evidence across studies and identify consistent patterns. This matters because patients with lower literacy skills may be using portals less, or not at all, while their providers and health systems assume otherwise. As a result, these patients may not recognize critical abnormal test results and clinical changes that require immediate follow-up, delaying diagnosis and treatment. The gap between portal access and meaningful engagement has also been documented among real-world portal users, who frequently encounter difficulty understanding the relevance and clinical significance of test results even after accessing the portal [20]. If health literacy and eHealth literacy relate differently to portal engagement, treating them as interchangeable could misdirect intervention efforts and result in medical errors, such as missed test results. For example, plain-language materials will not reach patients who cannot complete registration, while digital navigation support alone will not address the needs of patients who log in but cannot understand or act on the information they read.

The objective of this systematic review was to synthesize evidence on the associations of health literacy and eHealth literacy with patient portal access, general use, feature-specific use, and frequency of use. The goal of this work is to identify where literacy-related disparities emerge and where targeted support could reduce them.

Methods

Overview

This systematic review was reported in accordance with the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) statement [21]. The synthesis methods were guided by the Cochrane Handbook guidance for synthesis

XSL•FO
RenderX

without meta-analysis [22]. A review protocol was not registered, as this review began as an exploratory assessment of how literacy and portal outcomes were measured.

Eligibility Criteria

We included peer-reviewed articles and gray literature, including conference abstracts and papers, as well as theses, to minimize publication bias. Eligible studies (1) involved adults aged 18 years or older from clinical, community, or online samples; (2) evaluated a real-world patient portal, defined as a portal implemented in routine health care and used by patients rather than a hypothetical or simulated portal; (3) measured health literacy or eHealth literacy using a validated instrument; (4) reported a quantitative association between health literacy or eHealth literacy and patient portal access or use; and (5) were published in English. We excluded studies that evaluated nonportal technologies, such as electronic health record access without a patient portal, or focused only on user perceptions or qualitative outcomes, such as satisfaction, usability, or ease of use.

Information Sources and Search Strategy

We conducted a systematic search in PubMed, Scopus, CINAHL, PsycINFO, Web of Science, and Embase using database-specific free-text search strings related to health literacy, eHealth literacy, patient portals, and portal access or use, combined with Boolean operators. The search strategy was developed by the research team without support from a research librarian. Full search strings for each database are provided in Multimedia Appendix 1. No restrictions were applied to the publication date range. All searches were conducted on March 20, 2026.

Selection Process

All search results were imported into Covidence, where duplicates were identified and removed automatically or manually before title and abstract screening. Five reviewers (PH, MA, AP, AR, and SK) screened titles and abstracts in Covidence using predefined eligibility criteria, with each record assessed independently by 2 of them. Studies that passed the initial title and abstract screening advanced to full-text review, where the same 5 reviewers (PH, MA, AP, AR, and SK) applied the eligibility criteria, again with each article assessed independently by 2 of them. Any disagreements were resolved through discussion and consensus.

Interrater agreement for title and abstract screening, measured with Cohen's kappa (κ), ranged from κ=0.50 to κ=0.57, indicating moderate agreement [23]. The moderate kappa values likely reflect κ's prevalence sensitivity, since most title and abstract screening decisions were exclusions, which can lower kappa despite high observed agreement [24,25]. In contrast, proportion agreement for title and abstract screening ranged from 85% to 97%. Discrepancies arose from unclear abstracts, especially when studies did not clearly distinguish real-world patient portals from hypothetical portals or other digital technologies or did not clearly report quantitative outcomes of portal access or use. For full-text screening, agreement ranged from κ=0.79 to κ=1.00, indicating substantial or almost perfect agreement.

Data Collection Process and Data Items

For each study, we extracted data on key study characteristics, literacy measurement, and patient portal outcomes. Two of the 5 reviewers (PH, MA, AP, AR, and SK) independently extracted and compared data for each study, and any discrepancies were resolved by consensus. Interrater agreement for data extraction was substantial, with κ ranging from 0.76 to 0.80.

Outcome Classification

To support consistent synthesis across heterogeneous studies, we developed an outcome classification scheme during data extraction and applied it across all included studies. The framework and its definitions were derived inductively from the portal outcomes reported in the included studies, refined through reviewer discussion, and then applied consistently across studies. Specifically, we classified outcomes at 2 levels: outcome domain and portal outcome. The outcome domain referred to the broad stage of portal engagement, either portal access or portal use. We defined portal access as the early steps required to gain entry to the portal, and portal use as behaviors occurring after access had been established. The portal outcome referred to the specific action within each domain. This process yielded the following hierarchy:

1. Domain 1: portal access

- Account setup or activation: completion of portal registration, account creation, invitation acceptance, enrollment, or activation
- First portal login: first successful login to the portal after account setup or activation

2. Domain 2: portal use

- General portal use: any overall use of the patient portal after the first login, not tied to a specific feature
- Messaging: viewing, receiving, sending, or replying to portal messages
- Test results viewing: accessing laboratory or other health check results
- Vital sign monitoring: viewing, entering, or monitoring vital signs
- Note viewing: viewing physician or other clinical notes
- Appointments: viewing, scheduling, or managing appointments
- Medications and refills: viewing medications, managing prescriptions, or requesting refills
- Patient-reported outcome (PRO) measures: completing PRO measures, often before appointments
- Health education materials: viewing or accessing educational content in the portal

Following Beal et al [18] suggestions for quantifying portal use, we also classified portal use outcomes by outcome measure. Within the portal use domain, outcome measures were classified as (1) any use, defined as a binary indicator of whether a patient used the portal overall or a specific feature at least once, or (2) frequency of use, reflecting how often the patient used the portal or feature.

XSL•FO
RenderX

Literacy Measurement and Harmonization

We first identified whether each study measured health literacy or eHealth literacy. Most health literacy studies used dichotomous categorical measures, though the labels varied across studies, for example, limited, low, problematic, inadequate, or no limitation. To improve comparability across studies, we harmonized all categorical classifications to a common lower versus higher health literacy framing. A few studies used 3 ordered categories, such as limited, marginal, and adequate health literacy. In those cases, we used the most extreme contrast, for example, limited versus adequate health literacy, to align results with the lower versus higher framing. The original comparison framing and specific thresholds used across studies are shown in Multimedia Appendix 2 for reference. Studies that treated health literacy as a continuous variable were also interpreted directionally, from lower to higher health literacy. We applied the same approach to eHealth literacy. Most eHealth literacy studies used continuous measures, which were interpreted directionally from lower to higher eHealth literacy, whereas categorical classifications were harmonized to the same lower versus higher framing.

Risk-of-Bias Assessment

To assess risk of bias, we used the Joanna Briggs Institute (JBI) critical appraisal tools matched to the study designs, including the analytical cross-sectional [26], the cohort [27], and the quasi-experimental checklists [28]. We restricted the appraisal to full-text, peer-reviewed journal articles. Each criterion was rated yes, no, or unclear. Two of the 5 reviewers (PH, MA, AP, AR, and SK) assessed each study independently and resolved disagreements through consensus discussion. Interrater agreement for risk-of-bias items yielded κ values ranging from 0.63 to 0.75, indicating substantial agreement.

Synthesis Methods

The synthesis first summarized study characteristics descriptively and quantified the number of studies that examined each portal outcome, outcome measure, and literacy type.

Because studies differed substantially in the literacy instruments, portal outcomes, and outcome measures, we judged a pooled quantitative synthesis inappropriate. Instead, we used direction-based vote counting to summarize the direction of associations between health literacy or eHealth literacy and patient portal outcomes, regardless of effect size magnitude and statistical significance [22].

To account for potential confounding, we prioritized fully adjusted estimates when multiple estimates were reported. Effects showing no difference were coded as “no association” and included as a separate category. Some studies reported only statistical significance without effect estimates, whereas others reported descriptive findings only. In these cases, effects were considered noncodable because no clear directional comparison could be derived from the reported results. Studies reporting partially tested findings without a clear direction, or descriptive findings only, were therefore excluded from direction-based coding and summarized using narrative synthesis [29]. Multimedia Appendix 3 presents the study-level effect estimates, directional classifications, and reasons for exclusion from directional coding.

For studies included in direction-based vote counting, we followed the recommended practice of displaying effect directions visually in an effect direction plot [30]. For visual display in the effect direction plot, we grouped studies into smaller (<500 participants) and larger (≥500 participants) samples to provide a simple visual indication of study size across a heterogeneous evidence base.

Data synthesis was conducted at the level of one effect per study for each predefined portal outcome. A single study could contribute multiple effects if it reported different portal outcomes. To maintain a consistent orientation across studies, all effect estimates were standardized to the lower literacy perspective. For example, health literacy estimates originally reported for higher versus lower health literacy were inverted so that all coded effects reflected lower versus higher health literacy. The same approach was applied to eHealth literacy.

We decided not to assess the certainty of the evidence because the synthesis was based on direction-based vote counting rather than pooled effect estimates, which limits the assessment of certainty domains that rely on comparable effect estimates and CIs across studies.

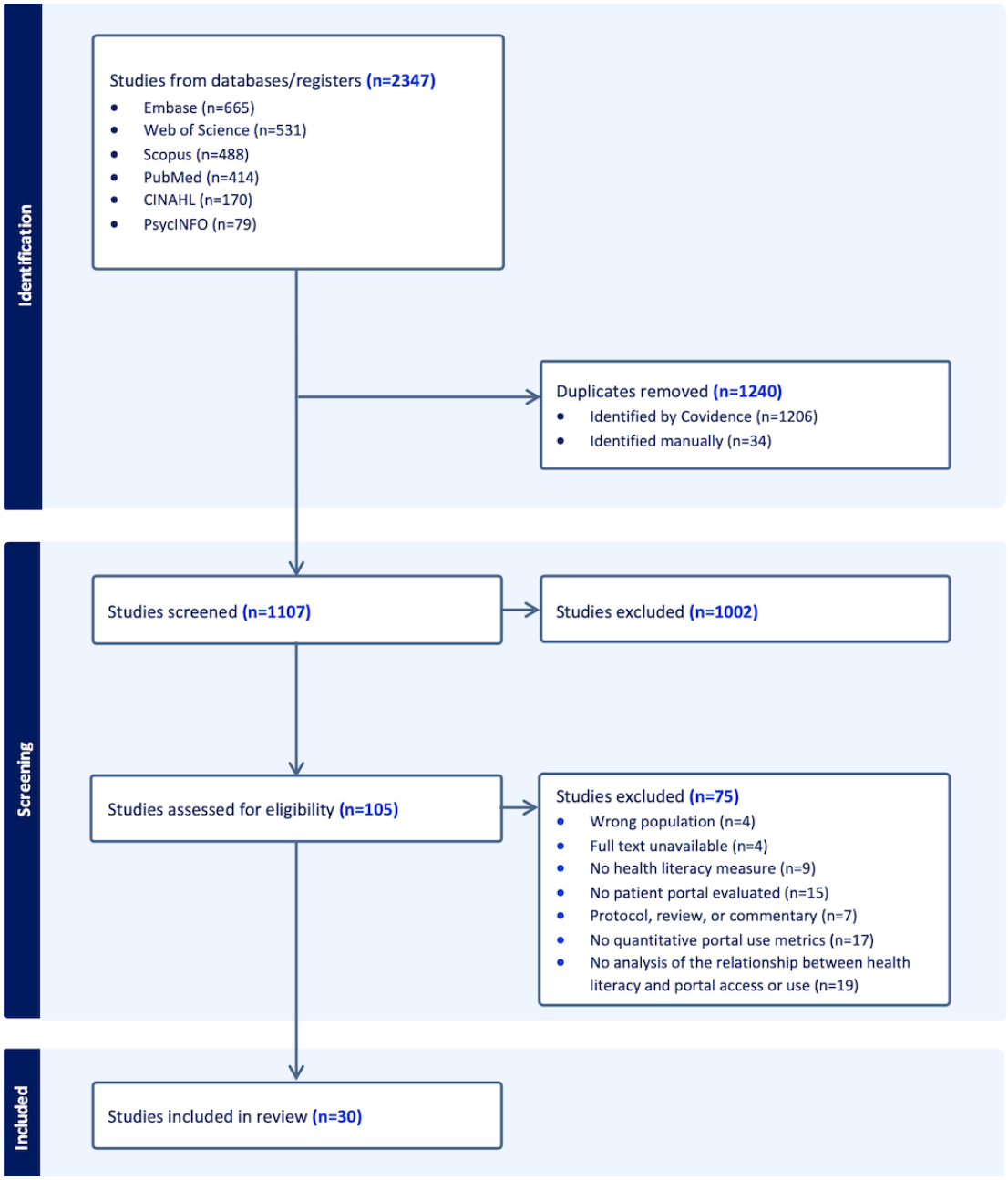
Results

Study Selection

A total of 2347 records were identified through database searches. After removal of 1240 duplicates, 1107 unique records remained for title and abstract screening. Of these, 105 records advanced to full-text review. After full-text assessment, 75 records were excluded, leaving 30 studies [31-60] for data extraction and synthesis. The full study selection process is shown in Figure 1.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.



Study Characteristics

Most included studies were journal articles (26/30, 87%) [33-39,41-45,47-60], while 2 were conference abstracts (2/30, 7%) [31,32], one was a conference paper (1/30, 3%) [40], and one was a thesis (1/30, 3%) [46]. Most studies were conducted in the United States (24/30, 80%) [32,34-43,46-50,52-57,59,60] and used a cross-sectional design (23/30, 77%) [31,34-37,39,41,43-51,53-59]. Sample sizes ranged from 18 to 60159, with most samples drawn from adult clinical populations (24/30, 80%) [31-33,35-47,50,52-57,60]. Survey self-report was the most common portal data source (21/30, 70%) [31,33-37,39,41,43-46,48-51,53-55,58,59], whereas portal log data were used in 8 studies (27%) [32,38,40,42,47,52,57,60] and administrative data in one study (3%) [56]. Health literacy alone was examined in 17 studies (57%) [32-34,36-40,42,47,49,52,55-58,60], eHealth literacy alone in 10 studies (33%) [41,44-46,48,50,51,53,54,59], and both literacy constructs in 3 studies (10%) [31,35,43]. The most common health literacy instruments were the Brief Health Literacy Screener (BHLS) [34,36-40,42,43,52,56] and the Newest Vital Sign (NVS) [35,47,49,55,57,60], whereas eHealth literacy was measured most often with the eHealth Literacy Scale (eHEALS) [35,41,43,45,46,48,50,51,53,54]. Health literacy was usually treated as a categorical variable, most often through dichotomous or ordinal groupings, although thresholds varied across studies and were not always reported. By contrast, most studies using eHEALS treated it as a continuous measure, while 2 studies applied dichotomous eHEALS cutoffs (Multimedia Appendix 2). Portal types included mainly institution-specific portals (25/30, 83%) [32-42,45-54,56,57,59,60], with the remainder using national patient portals (5/30, 17%) [31,43,44,55,58]. Table 1 summarizes the characteristics of all included studies.

XSL•FO
RenderX

Table 1. Characteristics of included studies (N=30).

Author (year)	Publication type	Country	Sample, setting	Clinical condition	Study design, data source	Patient portal	Literacy construct: instrument
Ahmed et al (2023) [34]	Journal article	United States	489 adults, online sample	General population	Cross-sectional, survey	Institutional portal	HL[a]: BHLS[b]
Arcury et al (2017) [35]	Journal article	United States	200 low-income older adults, 3 primary care clinics (urban and rural)	General population	Cross-sectional, survey	Institutional portal	eHL[c]: eHEALS[d] HL: NVS[e]
Bauer et al (2017) [36]	Journal article	United States	186 adults, 5 primary care clinics	General population	Cross-sectional, survey	Institutional portal	HL: BHLS
Cajita et al (2021) [37]	Journal article	United States	100 older adults, academic medical center	General population	Cross-sectional, survey	Institutional portal	HL: BHLS
Cemballi et al (2021) [38]	Journal article	United States	19921 adults, integrated health system	Diabetes	Retrospective longitudinal, portal logs	Kaiser Permanente Online (KP.org)	HL: BHLS
Clarke et al (2021) [39]	Journal article	United States	79 adults, academic medical center	Chronic cardiovascular disease	Cross-sectional, survey	MyChart	HL: BHLS
Davis et al (2015) [40]	Conference paper	United States	60159 adult inpatient admissions, hospital setting	General population	Retrospective, portal logs	My Health at Vanderbilt	HL: BHLS
Deshpande et al (2023) [41]	Journal article	United States	274 general medicine inpatients, academic medical center	General population	Cross-sectional, survey	Institutional portal	eHL: eHEALS
Di Tosto et al (2022) [42]	Journal article	United States	654 hospitalized adults, academic medical center	General population	Secondary analysis, trial portal logs	MyChart Bedside/MyChart	HL: BHLS
Haun et al (2015) [43]	Journal article	United States	819 veterans, 2 VA[f] medical centers	General population	Cross-sectional, survey	My HealtheVet (US VA, national)	eHL: eHEALS HL: BHLS
He et al (2016) [31]	Conference abstract	United Kingdom	175 adults, primary and secondary care clinics	Diabetes	Cross-sectional, survey	My Diabetes My Way (Scotland, national)	eHL: not specified HL: not specified
Holt et al (2019) [44]	Journal article	Denmark	246 adult outpatients, hospital outpatient clinic	General population	Cross-sectional, survey	sundhed.dk (Denmark, national)	eHL: eHLQ[g] & eHLA[h]
Hoogenbosch et al (2018) [45]	Journal article	Netherlands	439 adult outpatients, university hospital	General population	Cross-sectional, survey	Hospital portal	eHL: eHEALS
Kallmerten (2021) [46]	Thesis	United States	20 adults, 3 health care organizations	Heart failure	Cross-sectional, survey	Institutional portal	eHL: eHEALS
Kim et al (2018) [47]	Journal article	United States	18 low-income pregnant women, outpatient prenatal clinic	Pregnancy	Cross-sectional, portal logs and interviews	MyChart	HL: NVS
Lustria et al (2025) [48]	Journal article	United States	276 adults, community sample	At least 1 chronic condition	Cross-sectional, survey	Institutional portal	eHL: eHEALS
Mackert et al (2016) [49]	Journal article	United States	4974 adults, online research panel	General population	Cross-sectional, survey	Institutional portal	HL: NVS
Maroney et al (2021) [50]	Journal article	United States	288 transplant recipients, 2 academic transplant centers	Solid organ transplant	Cross-sectional, survey	Institutional portal	eHL: eHEALS
Neves et al (2021) [51]	Journal article	United Kingdom	1083 adults	General population	Cross-sectional, survey	CIE[i] (United Kingdom)	eHL: eHEALS
Nolla et al (2025) [52]	Journal article	United States	3457 oncology patients, 30 ambulatory oncology clinics	Cancer	Secondary analysis, trial portal logs	MyChart	HL: BHLS

XSL•FO
RenderX

Author (year)	Publication type	Country	Sample, setting	Clinical condition	Study design, data source	Patient portal	Literacy construct: instrument
Philpot et al (2024) [53]	Journal article	United States	1850 adults, multistate multisite academic health system	General population	Cross-sectional, survey	MyChart	eHL: eHEALS
Price-Haywood et al (2017) [54]	Journal article	United States	247 older adults, integrated health system	Hypertension or diabetes	Cross-sectional, survey	MyChart	eHL: eHEALS
Ruiz et al (2016) [55]	Journal article	United States	600 outpatient veterans, VA medical center	General population	Cross-sectional, survey	My HealtheVet (US VA, national)	HL: NVS
Sarkar et al (2011) [56]	Journal article	United States	14102 adults, integrated health system	Diabetes	Cross-sectional, administrative data	Kaiser Permanente Online (KP.org)	HL: BHLS
Smith et al (2015) [57]	Journal article	United States	534 older adults, general internal medicine clinic	General population	Cross-sectional, portal logs	EpicCare portal	HL: NVS
Sundell et al (2022) [58]	Journal article	Sweden	416 adults	General population	Cross-sectional, survey	1177.se (Sweden, national)	HL: S – C & C HL[j] scales
Van der Vaart et al (2014) [33]	Journal article	Netherlands	214 adults, rheumatology center	Rheumatoid arthritis	Pretest-posttest, survey	Hospital portal	HL: FCCHL[k] scales
Wright et al (2022) [59]	Journal article	United States	340 emerging adults, 2 universities	General population	Cross-sectional, survey	Institutional portal	eHL: e-HLS[l]
Yanez et al (2024) [32]	Conference abstract	United States	3464 patients, oncology setting	Cancer	Prospective, portal logs	MyChart	HL: SILS[m]
Yoon et al (2024) [60]	Journal article	United States	536 middle-aged and older primary care patients, academic medical center	Chronic conditions	Retrospective longitudinal, portal logs and survey	MyChart	HL: NVS

[a]HL: health literacy.

[b]BHLS: Brief Health Literacy Screener.

[c]eHL: eHealth literacy.

[d]eHEALS: eHealth Literacy Scale.

[e]NVS: Newest Vital Sign.

[f]VA: Veterans Affairs.

[g]eHLQ: eHealth Literacy Questionnaire.

[h]eHLA: eHealth Literacy Assessment Toolkit.

[i]CIE: Care Information Exchange.

[j]S – C & C HL: Swedish Communicative and Critical Health Literacy.

[k]FCCHL: Functional, Communicative, and Critical Health Literacy.

[l]e-HLS: Electronic Health Literacy Scale.

[m]SILS: Single Item Literacy Screener.

Risk-of-Bias Assessment

Risk of bias was appraised for the 26 journal papers [33-39,41-45,47-60]. Across cross-sectional studies, most JBI criteria were met. The most common risk-of-bias concern was the measurement of patient portal use through self-report rather than objective system data, which affected the reliability of outcome measurement in the majority of studies.

Across cohort studies, most criteria were also met. The most common risk-of-bias concern was inadequate handling of follow-up.

The single quasi-experimental study had risk-of-bias concerns related to the absence of a control group, baseline differences between user groups, a single pre-post time point, incomplete follow-up, and ceiling effects on half of the outcome measures [33]. Risk-of-bias assessments for all criteria and each study are presented in Multimedia Appendix 4.

Results of Synthesis

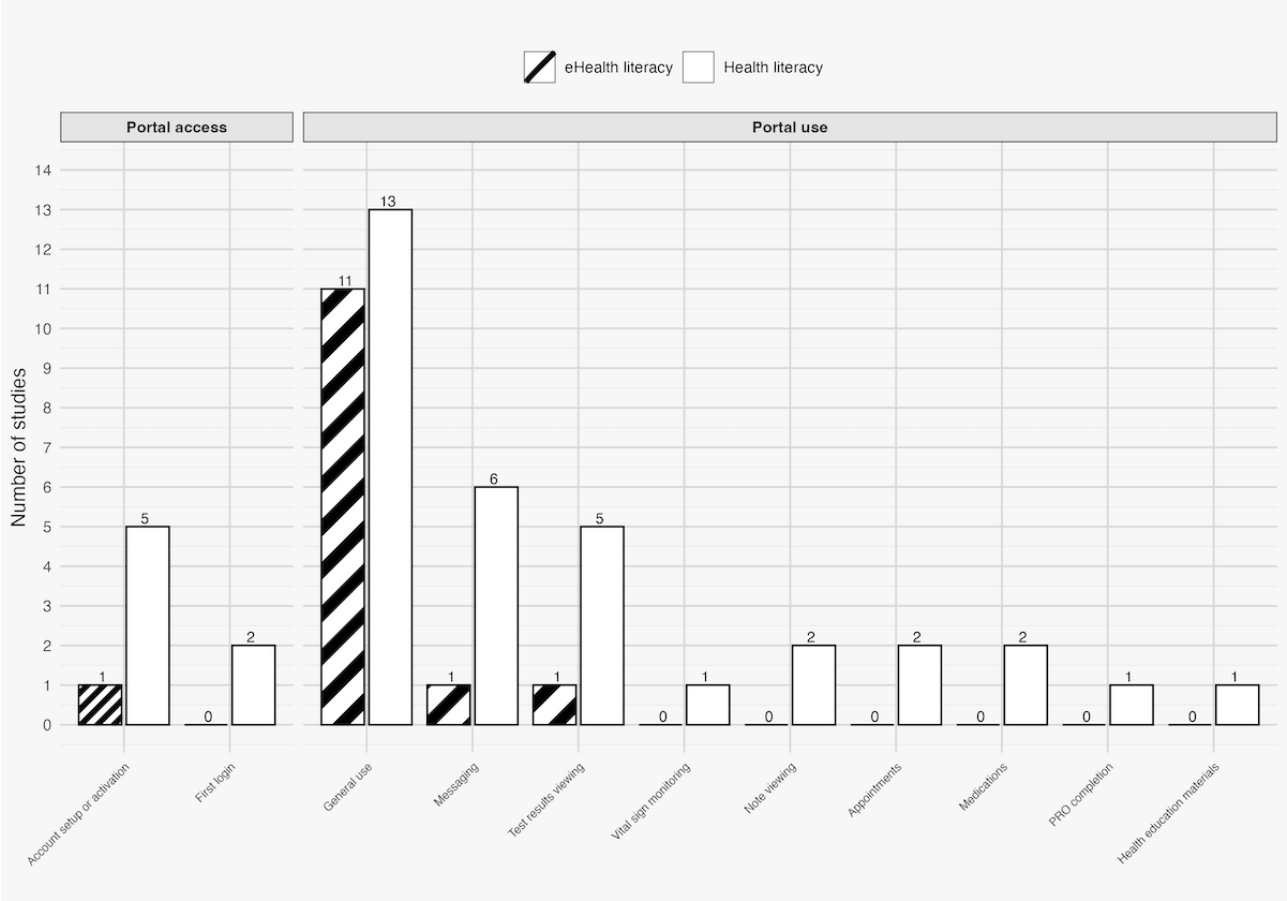
Distribution of Studies Across Portal Outcomes and Literacy Types

Figure 2 shows that the included studies focused far more on portal use than on portal access. Access-related outcomes, such as account setup or registration and first login, were rarely studied. Only 6 studies (20%) [40,47,53,55-57] examined account setup or activation, and 2 studies (7%) [33,56] assessed first login. Most portal features also received limited attention. Across both eHealth literacy and health literacy, general portal

XSL•FO
RenderX

use was the most examined outcome, assessed in 21 unique studies (70%) [31,34-37,39-47,49-51,54,58-60]. Three of these studies assessed both literacy types within the same sample [31,35,43]. Across all outcomes, health literacy was studied much more often than eHealth literacy.

Figure 2. Study counts by portal outcome and literacy type. Studies that assessed both health literacy and eHealth literacy within the same sample were counted once for each literacy type. PRO: patient-reported outcome.



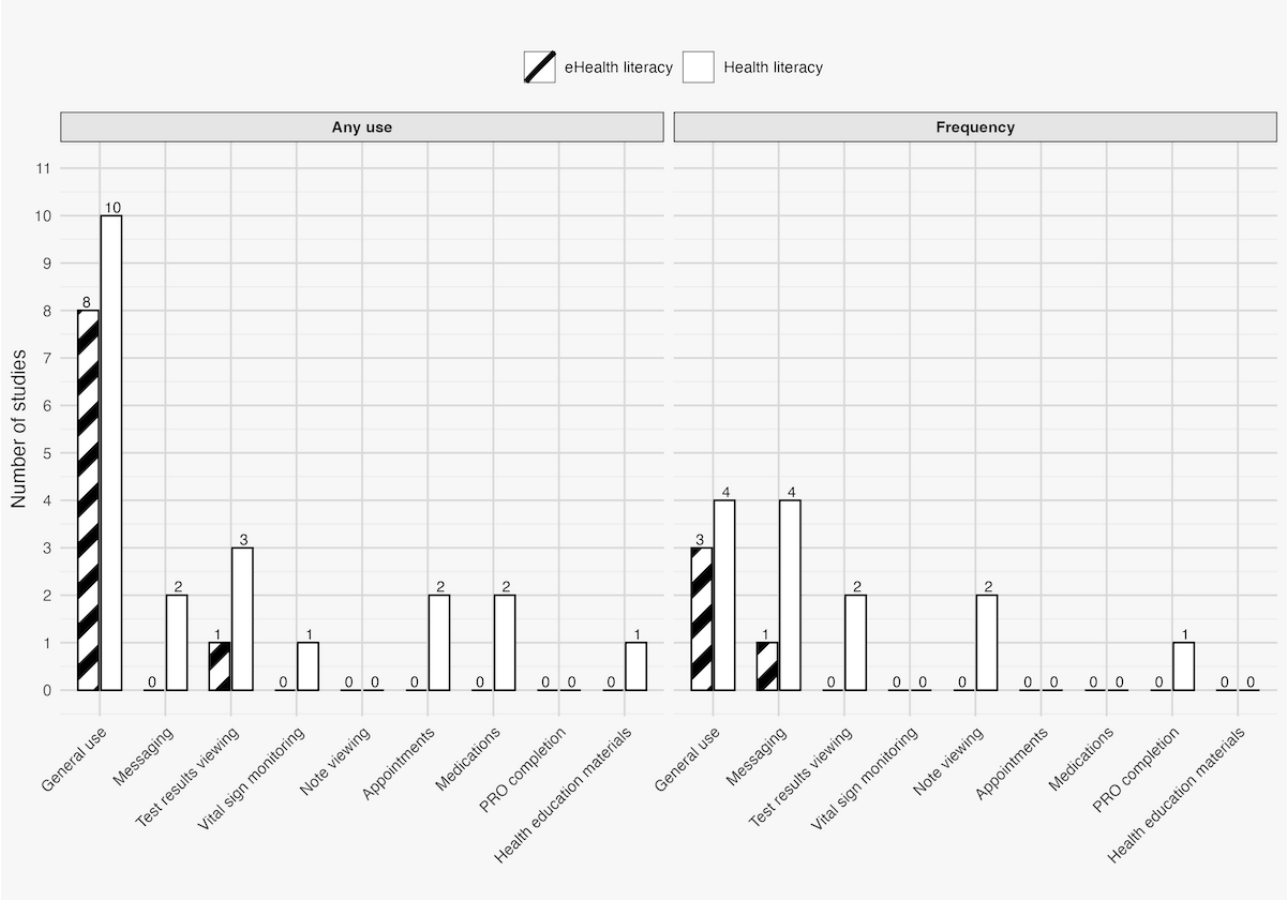
Distribution of Studies Across Outcome Measures and Literacy Types

As shown in Figure 3, studies more often assessed any use than frequency of use. Across both outcome measures, the evidence remained centered on general portal use. Sixteen studies [31,34-37,39-41,44,45,47,49,51,54,58,59] assessed general portal use as any use, and 6 studies [42,43,46,47,50,60] assessed it as frequency of use. Two studies [31,35] in the any-use group and one study [43] in the frequency group contributed both a health literacy and an eHealth literacy analysis.

Among feature-specific outcomes, messaging was the only portal function more often examined using frequency-based measures. Most other specific features were rare or absent in both any-use and frequency analyses. This lack of evidence was even more pronounced among eHealth literacy studies.

XSL•FO
RenderX

Figure 3. Study counts by portal outcome, outcome measure, and literacy type. Studies that assessed both health literacy and eHealth literacy within the same sample were counted once for each literacy type and outcome measure combination. PRO: patient-reported outcome.

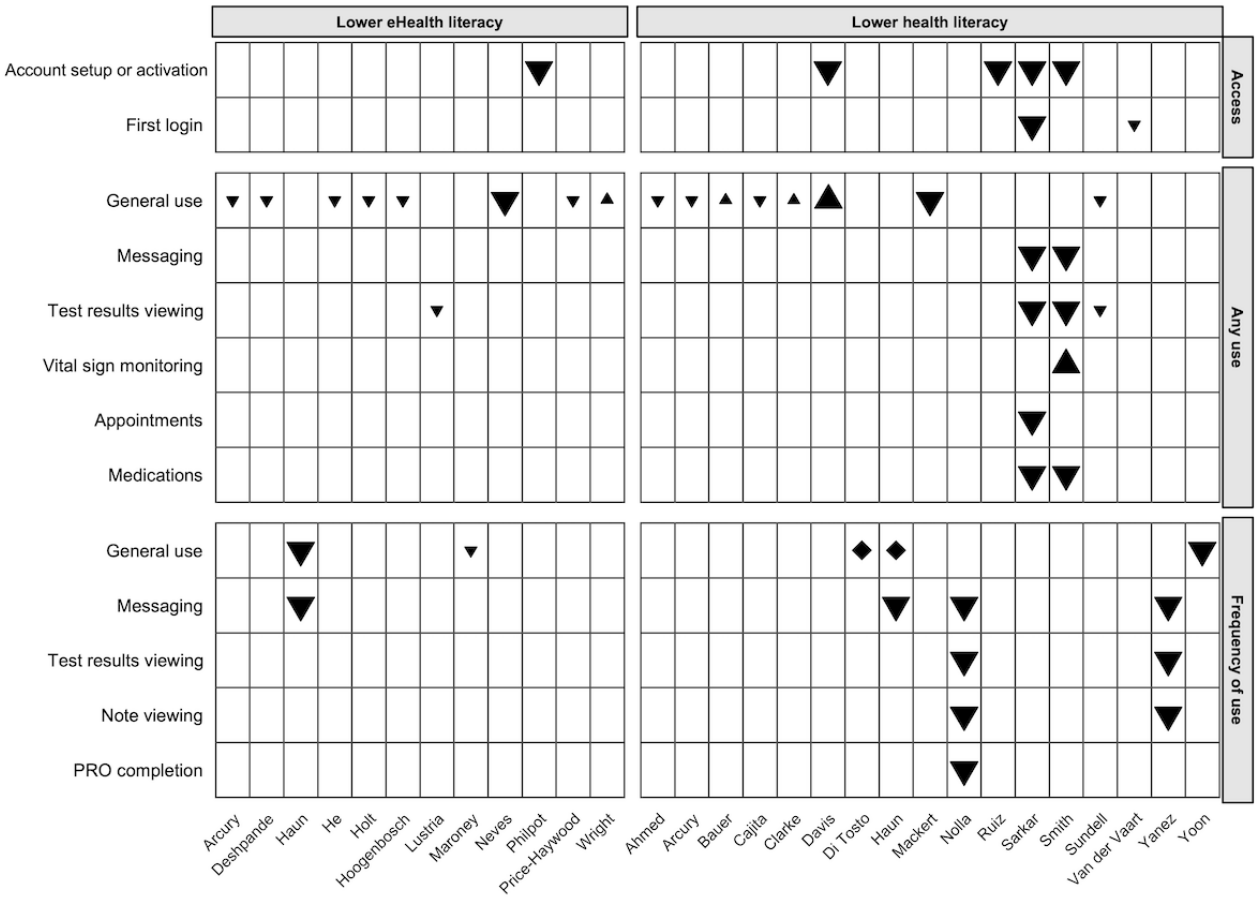


Evidence of Associations Between Lower Health Literacy, Lower eHealth Literacy, and Portal Outcomes and Outcome Measures

Twenty-seven individual studies (90%) [31-37,39-45,48-60] were included in direction-based vote counting. The remaining 3 studies [38,46,47] contributed only to narrative synthesis, and 2 additional studies [31,40] included some findings that could not be directionally coded. Figure 4 [31-37,39-45,48-60] summarizes all identified associations of lower health literacy or eHealth literacy across portal outcomes and outcome measures.

XSL•FO
RenderX

Figure 4. Associations of lower literacy by portal outcome and outcome measure. Direction is coded regardless of statistical significance. Effect direction: upward arrow ▲, higher portal access or use among individuals with lower literacy; downward arrow ▼, lower portal access or use among individuals with lower literacy; diamond ◆, no association. Included sample size: large symbol ≥500 participants; small symbol <500 participants. PRO: patient-reported outcome.



General Portal Use

eHealth Literacy

Nearly all eHealth literacy studies reported lower general portal use, both for any use (7/8) [31,35,41,44,45,51,54] and for frequency of use (2/2) [43,50]. One study [59] found slightly higher any use among those with lower eHealth literacy. Kallmerten [46] was included only in the narrative synthesis and found no significant association between eHealth literacy and portal use frequency, although no effect direction was reported.

Health Literacy

The pattern for health literacy was more mixed. Among studies of any general portal use, 5 of 8 (63%) studies [34,35,37,49,58] reported lower use in groups with lower health literacy, and 3 of 8 (38%) studies [36,39,40] reported the opposite pattern. He et al [31] similarly found no significant association between health literacy and any portal use, although the direction was not reported. For frequency of general portal use, 1 of 3 (33%) studies found lower use among those with lower health literacy [60], whereas the remaining 2 (67%) studies found no association [42,43]. Kim et al [47] was included only in the narrative synthesis and described lower health literacy as a possible barrier to both any use and frequency of use but did not formally test the association.

Portal Access

eHealth Literacy

Only one eHealth literacy study examined access outcomes, finding that lower eHealth literacy was associated with lower account setup or activation [53].

Health Literacy

All studies that examined portal access outcomes, including account setup or activation and first login, found lower access among individuals with lower health literacy [33,40,55-57]. Kim et al [47], included only narratively, similarly suggested that registration may be a barrier for individuals with lower health literacy, although the study did not report a formal test.

Feature-Specific Outcomes

eHealth Literacy

Few eHealth literacy studies examined feature-specific outcomes. For any-use measures, one [48] study reported lower test results viewing among those with lower eHealth literacy. For frequency-based outcomes, another study found lower

XSL•FO
RenderX

messaging frequency among individuals with lower eHealth literacy [43].

Health Literacy

Health literacy studies examined feature-specific outcomes more often and showed consistent negative associations. For any-use measures, lower health literacy was associated with lower messaging use (2/2) [56,57], lower test results viewing (3/3) [56-58], lower appointment use (1/1) [56], and lower medication feature use (2/2) [56,57]. The only exception was vital sign monitoring, for which 1 study reported higher use among those with lower health literacy [57]. Davis et al [40] reported a nonlinear, inverted-U relationship between health literacy and use of appointments and health education materials, so no clear directional interpretation was possible.

The consistency extended to frequency-based outcomes. Lower health literacy was associated with lower note viewing frequency in 2 of 2 studies [32,52], lower PRO measure completion in 1 of 1 study [52], lower messaging frequency in 3 of 3 studies [32,43,52], and lower test results viewing frequency in 2 of 2 studies [32,52]. Cemballi et al [38], included only in the narrative synthesis, also described lower cumulative secure messaging among patients with lower health literacy.

Studies Assessing Both Health Literacy and eHealth Literacy

The 3 studies that assessed both literacy constructs in the same sample allowed a more direct within-study comparison between health literacy and eHealth literacy. In Haun et al [43], lower eHealth literacy was associated with lower frequency of general portal use, whereas health literacy showed no clear directional association [43]. He et al [31] likewise found lower eHealth literacy to be associated with lower any portal use, while health literacy showed no significant association, although the direction of the health literacy effect was not reported. In contrast, Arcury et al [35] found a consistent pattern for both constructs, with portal users showing higher health literacy and higher eHealth literacy than nonusers. Yet, none of these studies simultaneously modeled both constructs in adjusted analyses.

Discussion

Principal Findings

The aim of this review was to synthesize associations of health literacy and eHealth literacy with patient portal access, general use, feature-specific use, and frequency of use. Several main findings were identified. In our review, the evidence base was uneven: health literacy was examined more often than eHealth literacy, included studies focused primarily on general portal use rather than portal access and specific features, and studies relied mainly on any-use measures rather than frequency of use. Despite these imbalances, health literacy and eHealth literacy may not function interchangeably as predictors of portal engagement. Nearly every directionally codable eHealth literacy study reported lower general portal use among patients with lower eHealth literacy. Health literacy showed a more mixed pattern for general use. The 3 studies that assessed both constructs within the same sample mostly supported this contrast, although they remain too few for firm comparative inference. In contrast, lower health literacy was more often associated with lower portal access and feature-specific use, but comparable eHealth literacy evidence for those outcomes remained sparse.

In our review, studies focused predominantly on general portal use, while early engagement stages, such as registration and first login, as well as specific portal features, received less attention. One possible explanation for this pattern is that many included studies recruited participants who were already registered portal users and therefore excluded the access stage from analysis. Another plausible explanation is that overall portal use is easier to measure than feature-specific behaviors, especially in cross-sectional surveys. Yet, treating portal use as a single outcome oversimplifies the staged nature of real-world portal engagement [61]. Access and subsequent portal use represent distinct stages at which disparities may emerge, and access alone does not necessarily translate into continued use [56,57,62]. Studies that measure only overall use may therefore miss whether literacy-related barriers arise during initial access, continued use, or both.

Across the included studies in our review, health literacy was examined more often than eHealth literacy. This imbalance likely reflects the longer history of health literacy as a field. Health literacy became central to health communication research [63] and was originally conceptualized in the context of paper-based health resources [12], well before digital health information became a major focus. Its recognition by the US Department of Health and Human Services further reflects its prominence at the policy level [64]. eHealth literacy, in contrast, was formally introduced only in 2006, as health information seeking increasingly moved into digital environments [12]. Prior digital health research appears to show a similar pattern, with greater emphasis on health literacy than on eHealth literacy [65]. This imbalance matters. Using health literacy as a proxy for eHealth literacy risks misclassifying digital navigation problems as comprehension problems, which may lead health systems to offer the wrong type of support.

A further pattern in our findings was the stronger emphasis on any use than frequency-based outcomes. This pattern is consistent with broader patient portal research, which often reports portal adoption or use without examining how frequently portals are used over time [18]. In our review, most studies relied on cross-sectional surveys to measure portal engagement. Although surveys are appropriate for capturing behavior at a single time point, they are less suited to assessing temporal patterns. Because these designs rely primarily on self-reported rather than objective data, they introduce recall bias [66,67] and provide limited information about when portal use occurred or how often it was repeated. Longitudinal designs using portal audit logs and defined follow-up periods are better suited to capturing repeated use over time [42]. As a result, the literacy evidence base reflects initial uptake more than sustained engagement and may underestimate barriers that emerge only with continued use.

Beyond outcome measurement, the choice of literacy constructs, instruments, and modeling approaches may have contributed to the observed patterns. One possible explanation is conceptual.

XSL•FO
RenderX

Health literacy and eHealth literacy instruments differ substantially in what demands within the portal experience they measure. Across included studies in this review, eHealth literacy was most often assessed with eHEALS, a self-report measure of perceived ability to find, appraise, and use online health information [68]. Health literacy was more often assessed with BHLS, which focuses on perceived difficulty with written clinical materials [69], or NVS, a performance-based test of functional comprehension and numeracy [70]. These differences suggest, but do not establish, that the 2 constructs may align with different aspects of portal engagement. For example, eHealth literacy may be more relevant during digital uptake and navigation, before patients reach portal content, whereas health literacy may become more relevant once patients must interpret and act on test results, clinical notes, or medication information. Our findings offer partial support for this interpretation, as lower health literacy showed more consistent associations with feature-specific portal outcomes than with general portal use. Comparable evidence for eHealth literacy at the feature level, however, remains too limited to allow a firm conclusion. At the same time, the observed difference may not be entirely conceptual and may instead reflect measurement choices. In our review, most studies modeled eHealth literacy as a continuous variable, while health literacy measures were often categorized. Because categorizing a continuous measure reduces information and statistical power [71,72], health literacy associations may have appeared weaker than the underlying relationship. This problem is further compounded by variations in the health literacy thresholds used across studies, with some studies not reporting thresholds at all. The 3 studies that assessed both literacy constructs within the same sample were broadly consistent with the overall pattern; however, they did not simultaneously model both in adjusted analyses. Given this limited within-sample evidence, it is unclear whether health literacy's weaker associations reflect true differences, confounding, measurement choices, or both.

Recommendations for Future Studies and Practical Implications

Given the uncertainty about whether the observed differences between health literacy and eHealth literacy are due to true construct differences, confounding, or both, future portal studies should measure both constructs simultaneously within the same sample using comparable approaches. Modeling both constructs continuously, where appropriate, would preserve statistical power and allow more direct comparison across portal outcomes. Studies should also distinguish portal access from portal use and separate any use from frequency-based outcomes to improve comparability across initial uptake and sustained engagement.

For clinicians and health systems, the choice of literacy measure can help identify the type of portal barrier a patient faces and guide more targeted intervention design. To improve registration, first login, or basic navigation, digital navigation barriers may warrant greater attention. If the goal is to improve understanding and use of portal content after login, health literacy-sensitive communication strategies may be more relevant. For health systems, portal equity assessment should extend beyond account creation or activation. Our findings suggest that literacy-related disparities may persist after access is established, especially for health literacy and feature-specific use, which could go undetected by systems that track only registration or activation. A patient may appear engaged in activation metrics yet never progress to viewing test results, managing medications, or messaging clinicians, the features through which portals affect care. These channels carry abnormal results, medication changes, and clinician messages about findings that change care. The 2 oncology studies in our sample found lower engagement in exactly these channels among patients with lower health literacy [32,52]. As measurement consistency improves, future studies should be better positioned to identify which literacy construct is most informative for specific portal barriers.

Limitations

This review has methodological limitations. First, we used direction-based vote counting. This approach ignores the magnitude of the effect size and gives equal weight to all studies regardless of study quality or risk of bias. Second, most comparisons between health literacy and eHealth literacy were indirect, because only 3 studies assessed both constructs in the same sample. Third, we harmonized heterogeneous literacy measures into a lower-versus-higher framing. This simplification may have obscured differences captured by 3-level categorical distinctions, such as limited, marginal, and adequate health literacy. Fourth, studies often modeled health literacy and eHealth literacy differently, with eHealth literacy typically treated as a continuous variable and health literacy often categorized, further limiting direct comparisons between the 2 constructs. Fifth, the outcome classification framework was developed inductively and was not prespecified, which may have introduced some subjectivity in how outcomes were grouped. Sixth, the review was limited to English-language studies, which may have introduced language bias. Seventh, a review protocol was not registered beforehand, which limits the transparency of the prespecified review methods. Eighth, searches relied on free-text terms and did not include database-specific vocabulary, such as MeSH terms and CINAHL subject headings, which could have decreased search sensitivity. Ninth, manual reference-list or citation searching was not conducted, which may have omitted relevant studies that were not identified through database searches. Finally, because no publication date restriction was applied, some older studies may have evaluated patient portals with different user interfaces and functionalities, which could have affected how directly their findings apply to contemporary portal use.

Limitations identified in the included literature should also be noted. Most studies were conducted in the United States, relied on clinical populations, and evaluated institution-specific portals, which limits generalizability to national portals and community populations. Differences in literacy measures, thresholds, and outcome measures (eg, self-reported use vs portal logs), as discussed above, further limit direct cross-study comparisons.

Conclusions

Current evidence points to a clear equity problem in patient portal engagement. Patient portals are now central to care, but research still relies on broad measures, most often general portal use or binary any-use outcome measures. This approach risks

XSL•FO
RenderX

missing where literacy-related disparities emerge. Across the available evidence, lower eHealth literacy showed a more consistently negative cross-study pattern for general portal use, whereas lower health literacy more often showed negative associations with portal access and several feature-specific outcomes. Because only 3 studies assessed both constructs within the same sample, this contrast should be interpreted as a cross-study pattern rather than a direct comparative finding.

Future research should move beyond broad portal-use measures and examine where disparities occur across the full continuum of portal engagement. Studies should measure both health literacy and eHealth literacy in the same sample, use objective portal log data when possible, and distinguish access, any use, frequency, and feature-specific behaviors. This shift would help health systems identify which patients are left behind after portals become available and design support matched to the barrier: digital navigation, clinical comprehension, or both.

Data Availability

The search strategy, literacy measures, study-level effects, and risk-of-bias assessments used in this review can be found in Multimedia Appendices 1-4. No analytic code was used for this review.

Funding

The authors declared no financial support was received for this work.

Generative AI Use

We used Claude Sonnet 4.6 to check for grammatical mistakes, punctuation errors, and incorrect spelling, and to improve readability after the manuscript was written. We critically reviewed and edited the content as needed and take full responsibility for the content of this work.

Authors' Contributions

Conceptualization: PH
Data curation: PH
Formal analysis: PH, MA, AP, AR, SK, CLR
Methodology: PH
Project administration: PH
Resources: PH
Supervision: MG
Visualization: MA, AP, AR, SK
Writing – original draft: PH
Writing – review & editing: MA, AP, AR, SK, CLR, MG

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strings.
[DOCX File , 21 KB - humanfactors_v13i1e104485_app1.docx]

Multimedia Appendix 2
Literacy measures, comparison framing, and thresholds used.
[DOCX File , 26 KB - humanfactors_v13i1e104485_app2.docx]

Multimedia Appendix 3
Study-level effect estimates and directional coding.
[DOCX File , 34 KB - humanfactors_v13i1e104485_app3.docx]

Multimedia Appendix 4
Risk-of-bias assessment.
[DOCX File , 29 KB - humanfactors_v13i1e104485_app4.docx]

Multimedia Appendix 5
PRISMA 2020 checklist.

XSL•FO
RenderX

[DOCX File , 269 KB - humanfactors_v13i1e104485_app5.docx]

References

1. Roehrs A, da Costa CA, Righi RDR, de Oliveira KSF. Personal health records: a systematic literature review. J Med Internet Res 2017;19(1):e13 [FREE Full text] [doi: 10.2196/jmir.5876] [Medline: 28062391]
2. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013;159(10):677-687 [FREE Full text] [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
3. Irizarry T, DeVito Dabbs A, Curran CR. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
4. What is HINTS? Health Information National Trends Survey (HINTS). URL: https://hints.cancer.gov/ [accessed 2026-04-01]
5. Baldwin JL, Singh H, Sittig DF, Giardina TD. Patient portals and health apps: pitfalls, promises, and what one might learn from the other. Healthc (Amst) 2017;5(3):81-85 [FREE Full text] [doi: 10.1016/j.hjdsi.2016.08.004] [Medline: 27720139]
6. Carini E, Villani L, Pezzullo AM, Gentili A, Barbara A, Ricciardi W, et al. The impact of digital patient portals on health outcomes, system efficiency, and patient attitudes: updated systematic literature review. J Med Internet Res 2021;23(9):e26189 [FREE Full text] [doi: 10.2196/26189] [Medline: 34494966]
7. Dendere R, Slade C, Burton-Jones A, Sullivan C, Staib A, Janda M. Patient portals facilitating engagement with inpatient electronic medical records: a systematic review. J Med Internet Res 2019;21(4):e12779 [FREE Full text] [doi: 10.2196/12779] [Medline: 30973347]
8. Lyles CR, Nelson EC, Frampton S, Dykes PC, Cemballi AG, Sarkar U. Using electronic health record portals to improve patient engagement: research priorities and best practices. Annals of Internal Medicine 2020;172(11_Supplement):S123-S129. [doi: 10.7326/m19-0876]
9. Nielsen-Bohlman L, Kindig DA, Panzer AM, editors. Health Literacy: A Prescription to End Confusion. Washington DC, USA: National Academies Press; 2004.
10. Kutner M, Greenberg E, Jin Y, Paulsen C. The health literacy of America's adults: results from the 2003 national assessment of adult literacy (NCES 2006–483). National Center for Education Statistics. 2006. URL: https://nces.ed.gov/pubs2006/2006483.pdf [accessed 2026-10-17]
11. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011;155(2):97-107 [FREE Full text] [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
12. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
13. Milanti A, Norman C, Chan DNS, So WKW, Skinner H. eHealth literacy 3.0: updating the Norman and Skinner 2006 model. J Med Internet Res 2025;27:e70112 [FREE Full text] [doi: 10.2196/70112] [Medline: 40068166]
14. Monkman H, Kushniruk A, Barnett J, Borycki EM, Greiner LE, Sheets D. Are health literacy and eHealth literacy the same or different? Stud Health Technol Inform 2017;245:178-182. [Medline: 29295077]
15. Health literacy. Centers for Disease Control and Prevention (CDC). URL: https://www.cdc.gov/health-literacy/php/research-summaries/ehealth.html [accessed 2025-10-10]
16. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012;12(1):80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
17. Edwards G, Dorstyn D, Oxlad M. Examining the relationship between health literacy and eHealth literacy in adult populations: a systematic review and meta-analysis. Health Promot Int 2025;40(6):daaf217 [FREE Full text] [doi: 10.1093/heapro/daaf217] [Medline: 41398315]
18. Beal LL, Kolman JM, Jones SL, Khleif A, Menser T. Quantifying patient portal use: systematic review of utilization metrics. J Med Internet Res 2021;23(2):e23493 [FREE Full text] [doi: 10.2196/23493] [Medline: 33629962]
19. Coughlin SS, Stewart JL, Young L, Heboyan V, De Leo G. Health literacy and patient web portals. Int J Med Inform 2018;113:43-48. [doi: 10.1016/j.ijmedinf.2018.02.009] [Medline: 29602432]
20. Giardina TD, Modi V, Parrish DE, Singh H. The patient portal and abnormal test results: an exploratory study of patient experiences. Patient Exp J 2015;2(1):148-154 [FREE Full text] [Medline: 28345018]
21. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
22. McKenzie JE, Brennan SE. Synthesizing and presenting findings using other methods. In: Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, New Jersey, USA: John Wiley & Sons, Inc; 2019:321-347.
23. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012;22(3):276-282. [doi: 10.11613/bm.2012.031]
24. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. J Clin Epidemiol 1993;46(5):423-429. [doi: 10.1016/0895-4356(93)90018-v] [Medline: 8501467]
25. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43(6):543-549. [doi: 10.1016/0895-4356(90)90158-l] [Medline: 2348207]

XSL•FO
RenderX

26. Barker TH, Hasanoff S, Aromataris E, Stone JC, Leonardi-Bee J, Sears K, et al. The revised JBI critical appraisal tool for the assessment of risk of bias for analytical cross-sectional studies. JBI Evid Synth 2026;24(3):401-408. [doi: 10.11124/JBIES-24-00523] [Medline: 40521701]
27. Barker TH, Hasanoff S, Aromataris E, Stone JC, Leonardi-Bee J, Sears K, et al. The revised JBI critical appraisal tool for the assessment of risk of bias for cohort studies. JBI Evid Synth 2025;23(3):441-453. [doi: 10.11124/JBIES-24-00103] [Medline: 39177422]
28. Barker T, Habibi N, Aromataris E, Stone JC, Leonardi-Bee J, Sears K, et al. The revised JBI critical appraisal tool for the assessment of risk of bias for quasi-experimental studies. JBI Evid Synth 2024;22(3):378-388. [doi: 10.11124/JBIES-23-00268] [Medline: 38287725]
29. Popay J, Roberts H, Sowden A. Guidance on the conduct of narrative synthesis in systematic reviews. A Product from the ESRC Methods Programme Version 2006;1(1):b92. [doi: 10.13140/2.1.1018.4643]
30. Boon MH, Thomson H. The effect direction plot revisited: application of the 2019 Cochrane handbook guidance on alternative synthesis methods. Res Synth Methods 2021;12(1):29-33 [FREE Full text] [doi: 10.1002/jrsm.1458] [Medline: 32979023]
31. He J, Cunningham S, Wake D. Patients' barriers to accessing electronic personal health records: experiences with My Diabetes My Way. In: Diabet Med. 2016 Presented at: Proceedings of the Diabetes UK Professional Conference 2016; March 2-4, 2016; Scottish Exhibition and Conference Centre, Glasgow p. 115 URL: https://onlinelibrary.wiley.com/doi/epdf/10.1111/dme.30_13048
32. Yanez B, Lancki N, Garcia SF, Flores AM, Wang Y, Richardson K, et al. Association of symptom severity, anxiety, and health literacy with portal usage among patients with cancer. J Clin Oncol 2024;42(16_suppl):1567-1567. [doi: 10.1200/jco.2024.42.16_suppl.1567]
33. van der Vaart R, Drossaert C, Taal E, Drossaers-Bakker KW, Vonkeman HE, van de Laar MAFJ. Impact of patient-accessible electronic medical records in rheumatology: use, satisfaction and effects on empowerment among patients. BMC Musculoskelet Disord 2014;15(1):102 [FREE Full text] [doi: 10.1186/1471-2474-15-102] [Medline: 24673997]
34. Ahmed N, Brown J, Parau C, McCullers A, Sanghavi K, Littlejohn R, et al. Bridging the digital health divide: characterizing patient portal users and nonusers in the U.S. Med Care 2023;61(7):448-455. [doi: 10.1097/MLR.0000000000001869] [Medline: 37289563]
35. Arcury TA, Quandt SA, Sandberg JC, Miller DP, Latulipe C, Leng X, et al. Patient portal utilization among ethnically diverse low income older adults: observational study. JMIR Med Inform 2017;5(4):e47 [FREE Full text] [doi: 10.2196/medinform.8026] [Medline: 29138129]
36. Bauer AM, Rue T, Munson SA, Ghomi RH, Keppel GA, Cole AM, et al. Patient-oriented health technologies: patients' perspectives and use. J Mob Technol Med 2017;6(2):1-10 [FREE Full text] [doi: 10.7309/jmtm.6.2.1] [Medline: 28936236]
37. Cajita MI, Nilsen ML, Irizarry T, Callan JA, Beach SR, Swartwout E, et al. Predictors of patient portal use among community-dwelling older adults. Res Gerontol Nurs 2021;14(1):33-42 [FREE Full text] [doi: 10.3928/19404921-20200918-03] [Medline: 32966584]
38. Cemballi AG, Karter AJ, Schillinger D, Liu JY, McNamara DS, Brown W, et al. Descriptive examination of secure messaging in a longitudinal cohort of diabetes patients in the ECLIPPSE study. J Am Med Inform Assoc 2021;28(6):1252-1258 [FREE Full text] [doi: 10.1093/jamia/ocaa281] [Medline: 33236117]
39. Clarke MA, Fruhling AL, Lyden EL, Tarrell AE, Bernard TL, Windle JR. The role of computer skills in personal health record adoption among patients with heart disease: multidimensional evaluation of users versus nonusers. JMIR Hum Factors 2021;8(3):e19191 [FREE Full text] [doi: 10.2196/19191] [Medline: 34309574]
40. Davis SE, Osborn CY, Kripalani S, Goggins KM, Jackson GP. Health literacy, education levels, and patient portal usage during hospitalizations. AMIA Annu Symp Proc 2015;2015:1871-1880 [FREE Full text] [Medline: 26958286]
41. Deshpande N, Arora VM, Vollbrecht H, Meltzer DO, Press V. eHealth literacy and patient portal use and attitudes: cross-sectional observational study. JMIR Hum Factors 2023;10:e40105 [FREE Full text] [doi: 10.2196/40105] [Medline: 36705947]
42. Di Tosto G, Walker DM, Sieck CJ, Wallace L, MacEwan SR, Gregory ME, et al. Examining the relationship between health literacy, health numeracy, and patient portal use. Appl Clin Inform 2022;13(3):692-699 [FREE Full text] [doi: 10.1055/s-0042-1751239] [Medline: 35793698]
43. Haun JN, Patel NR, Lind JD, Antinori N. Large-scale survey findings inform patients' experiences in using secure messaging to engage in patient-provider communication and self-care management: a quantitative assessment. J Med Internet Res 2015;17(12):e282 [FREE Full text] [doi: 10.2196/jmir.5152] [Medline: 26690761]
44. Holt KA, Karnoe A, Overgaard D, Nielsen SE, Kayser L, Røder ME, et al. Differences in the level of electronic health literacy between users and nonusers of digital health services: an exploratory survey of a group of medical outpatients. Interact J Med Res 2019;8(2):e8423 [FREE Full text] [doi: 10.2196/ijmr.8423] [Medline: 30950809]
45. Hoogenbosch B, Postma J, de Man-van Ginkel JM, Tiemessen NA, van Delden JJ, van Os-Medendorp H. Use and the users of a patient portal: cross-sectional study. J Med Internet Res 2018;20(9):e262 [FREE Full text] [doi: 10.2196/jmir.9418] [Medline: 30224334]

XSL•FO
RenderX

46. Kallmerten PS. Patient portal use, health information technology self-efficacy, electronic health literacy, and self-management among patients with heart failure: a cross sectional study [PhD thesis]. Duquesne University. 2021. URL: https://dsc.duq.edu/cgi/viewcontent.cgi?article=3042&context=etd [accessed 2026-10-26]
47. Kim J, Mathews H, Cortright LM, Zeng X, Newton E. Factors affecting patient portal use among low-income pregnant women: mixed-methods pilot study. JMIR Form Res 2018;2(1):e6 [FREE Full text] [doi: 10.2196/formative.5322] [Medline: 30684413]
48. Lustria M, Aliche O, Killian M, He Z. Enhancing patient engagement and understanding: is providing direct access to laboratory results through patient portals adequate? JAMIA Open 2025;8(2):ooaf009 [FREE Full text] [doi: 10.1093/jamiaopen/ooaf009] [Medline: 40130170]
49. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
50. Maroney K, Curtis LM, Opsasnick L, Smith KD, Eifler MR, Moore A, et al. eHealth literacy and web-based patient portal usage among kidney and liver transplant recipients. Clin Transplant 2021;35(2):e14184. [doi: 10.1111/ctr.14184] [Medline: 33278846]
51. Neves AL, Smalley KR, Freise L, Harrison P, Darzi A, Mayer EK. Determinants of use of the care information exchange portal: cross-sectional study. J Med Internet Res 2021;23(11):e23481 [FREE Full text] [doi: 10.2196/23481] [Medline: 34762063]
52. Nolla KM, Kuharic M, Lancki N, Walsh-Bailey CL, Flores AM, Garcia SF, et al. Patient portal engagement in oncology: results from the NU IMPACT study in a large health care system. JCO Clin Cancer Inform 2025;9:e2500178. [doi: 10.1200/CCI-25-00178] [Medline: 41370780]
53. Philpot LM, Ramar P, Roellinger DL, Njeru JW, Ebbert JO. Individual-level digital determinants of health and technology acceptance of patient portals: cross-sectional assessment. JMIR Form Res 2024;8:e56493 [FREE Full text] [doi: 10.2196/56493] [Medline: 38695754]
54. Price-Haywood EG, Harden-Barrios J, Ulep R, Luo Q. eHealth literacy: patient engagement in identifying strategies to encourage use of patient portals among older adults. Popul Health Manag 2017;20(6):486-494. [doi: 10.1089/pop.2016.0164] [Medline: 28384076]
55. Ruiz JG, Andrade AD, Hogue C, Karanam C, Akkineni S, Cevallos D, et al. The association of graph literacy with use of and skills using an online personal health record in outpatient veterans. J Health Commun 2016;21(sup2):83-90. [doi: 10.1080/10810730.2016.1193915] [Medline: 27668834]
56. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]
57. Smith S, O'Conor R, Aitken W, Curtis LM, Wolf MS, Goel MS. Disparities in registration and use of an online patient portal among older adults: findings from the LitCog cohort. J Am Med Inform Assoc 2015;22(4):888-895 [FREE Full text] [doi: 10.1093/jamia/ocv025] [Medline: 25914099]
58. Sundell E, Wångdahl J, Grauman Å. Health literacy and digital health information-seeking behavior - a cross-sectional study among highly educated Swedes. BMC Public Health 2022;22(1):2278 [FREE Full text] [doi: 10.1186/s12889-022-14751-z] [Medline: 36471284]
59. Wright JA, Volkman JE, Leveille SG, Amante DJ. Predictors of online patient portal use among a diverse sample of emerging adults: cross-sectional survey. JMIR Form Res 2022;6(2):e33356 [FREE Full text] [doi: 10.2196/33356] [Medline: 35166686]
60. Yoon E, Hur S, Opsasnick L, Huang W, Batio S, Curtis LM, et al. Disparities in patient portal use among adults with chronic conditions. JAMA Netw Open 2024;7(2):e240680 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.0680] [Medline: 38421645]
61. Antonio MG, Petrovskaya O, Lau F. The state of evidence in patient portals: umbrella review. J Med Internet Res 2020;22(11):e23851 [FREE Full text] [doi: 10.2196/23851] [Medline: 33174851]
62. Goel MS, Brown TL, Williams A, Hasnain-Wynia R, Thompson JA, Baker DW. Disparities in enrollment and use of an electronic patient portal. J Gen Intern Med 2011;26(10):1112-1116 [FREE Full text] [doi: 10.1007/s11606-011-1728-3] [Medline: 21538166]
63. Berkman ND, Davis TC, McCormack L. Health literacy: what is it? J Health Commun 2010;15 Suppl 2:9-19. [doi: 10.1080/10810730.2010.499985] [Medline: 20845189]
64. Health literacy in healthy people 2030. Office of Disease Prevention and Health Promotion. URL: https://odphp.health.gov/healthypeople/priority-areas/health-literacy-healthy-people-2030 [accessed 2026-10-07]
65. Kim H, Xie B. Health literacy in the eHealth era: a systematic review of the literature. Patient Educ Couns 2017;100(6):1073-1082. [doi: 10.1016/j.pec.2017.01.015] [Medline: 28174067]
66. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc 2016;9:211-217 [FREE Full text] [doi: 10.2147/JMDH.S104807] [Medline: 27217764]

XSL•FO
RenderX

67. Short M, Goetzel R, Pei X, Tabrizi MJ, Ozminkowski RJ, Gibson TB, et al. How accurate are self-reports? Analysis of self-reported health care utilization and absence when compared with administrative data. J Occup Environ Med 2009;51(7):786-796 [FREE Full text] [doi: 10.1097/JOM.0b013e3181a86671] [Medline: 19528832]
68. Norman C, Skinner H. eHEALS: the eHealth literacy scale. J Med Internet Res 2006;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
69. Chew L, Bradley K, Boyko E. Brief questions to identify patients with inadequate health literacy. Fam Med 2004;36(8):588-594. [Medline: 15343421]
70. Weiss BD, Mays MZ, Martz W, Castro KM, DeWalt DA, Pignone MP, et al. Quick assessment of literacy in primary care: the newest vital sign. Ann Fam Med 2005;3(6):514-522 [FREE Full text] [doi: 10.1370/afm.405] [Medline: 16338915]
71. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ 2006;332(7549):1080. [doi: 10.1136/bmj.332.7549.1080] [Medline: 16675816]
72. MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. Psychol Methods 2002;7(1):19-40 [FREE Full text] [doi: 10.1037/1082-989x.7.1.19] [Medline: 11928888]

Abbreviations

BHLS: Brief Health Literacy Screener
eHEALS: eHealth Literacy Scale
eHL: eHealth literacy
HL: health literacy
JBI: Joanna Briggs Institute
NVS: Newest Vital Sign
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses
PRO: patient-reported outcome

Edited by N von Kalckreuth; submitted 12.Jun.2026; peer-reviewed by L Agrawal, SA Steindal; comments to author 15.Sep.2026; revised version received 01.Oct.2026; accepted 02.Oct.2026; published 02.Nov.2026.

Please cite as:
Haessner P, Abraham M, Peter A, Rosario A, Kanaskar S, Romero CL, Gregory M
Associations of Health Literacy and eHealth Literacy With Patient Portal Access and Use: Systematic Review
JMIR Hum Factors 2026;13:e104485
URL: https://humanfactors.jmir.org/2026/1/e104485
doi:10.2196/104485
PMID:

©Philipp Haessner, Mahitha Abraham, Aleena Peter, Abigail Rosario, Samika Kanaskar, Crystal L Romero, Megan Gregory. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 02.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Demystifying Quality Metrics and Unveiling the True Measure of Quality of Care in Nursing Homes: Mixed Effects Analysis

Suhas Bharadwaj[1,2*], PhD; Haneen Ali[3*], PhD

[1]Department of Industrial and Systems Engineering, College of Engineering, Auburn University, Auburn, AL, United States

[2]Department of Industrial and Systems Engineering, College of Science and Engineering, University of Minnesota, Twin Cities, Minneapolis, MN, United States

[3]Mechanical and Industrial Engineering Department, Faculty of Engineering and Technology, Applied Science Private University, 21 Al Arab Street, Amman, Jordan

[*]all authors contributed equally

Corresponding Author:
Haneen Ali, PhD
Mechanical and Industrial Engineering Department, Faculty of Engineering and Technology, Applied Science Private University, 21 Al Arab Street, Amman, Jordan

Abstract

Background: The 5-Star Quality Rating System for nursing homes plays a central role in evaluating quality of care, although it has both strengths and limitations. This system relies heavily on the Minimum Data Set and derives several quality measures (QMs) from it. In this study, we validated the effectiveness of the 5-Star Quality Rating System for nursing homes and its underlying QMs in estimating quality of care. We constructed a panel dataset of US nursing homes (n=15,416) active from May 2020 to June 2023, retrieving data from three major sources: (1) COVID-19 nursing home data, (2) Payroll-Based Journal data, and (3) nursing home QM snapshots. The outcome variables included (1) resident infection, (2) staff infection, or (3) resident and staff deaths. The predictor variables were the 5-Star Quality Rating System for nursing homes and its underlying QMs classified as structure, process, or outcome (SPO) QMs.

Objective: This study aims to evaluate the effectiveness of nursing home QMs by regressing nursing home COVID-19 outcomes on nursing home QMs, classified using the Donabedian SPO framework. We hypothesized that nursing homes with better structural quality (eg, greater staff availability, better skill mix, and so on), better process quality (eg, lower restraint use and higher vaccination rates), and better outcome quality (eg, lower number of residents with pressure ulcers and a lower number of resident falls) experienced better COVID-19 performance in terms of resident and staff infections and deaths.

Methods: To examine the association between the COVID-19 outcomes and SPO QMs, we imputed missing values in the dataset using random forest. Subsequently, we modeled the imputed dataset using hurdled zero-inflated negative binomial mixed effects models. The zero inflation model included factors influencing initial susceptibility to COVID-19 or factors influencing the possibility of death after COVID-19 had been contracted. The model estimates were conditioned on zero inflation and random effects.

Results: Staffing measures (P<.001 for all variables in all models), health deficiency scores (P<.001 for all variables in at least 1 model), COVID-19 hospitalization rates (P<.001 for all variables in at least 2 models), and vaccinations (P<.001 for all variables in at least 2 models) exhibited meaningful relationships with the COVID-19 outcomes, while the 5-star components, Medicaid dependency, and ownership showed no clear relationships.

Conclusions: Although widely used, the 5-Star Quality Rating System for nursing homes is an unreliable performance measure. Concerted efforts from lawmakers, policy makers, and lobbyists are needed to refine and enhance the measure, thereby ensuring its reliability and effectiveness.

(JMIR Hum Factors 2026;13:e72770) doi:10.2196/72770

KEYWORDS

5-Star Quality Rating; nursing homes; performance measures; outcomes; COVID-19; nursing home industry; nursing home performance; nursing home quality; quality measures; SPO; structure, process, or outcome; Donabedian SPO framework

XSL•FO
RenderX

Introduction

In the United States, nursing homes are essential care providers for more than 1.3 million residents across 15,600 facilities [1,2]. As the third-largest sector in the health care industry, nursing homes employ more than 1.7 million dedicated staff members [2] who cater to the complex needs of the older adults, who are often frail and affected by multimorbidity [3-6]. With projections indicating consistent increases in the older adult population [7] and average life expectancy [8,9], the demand for nursing home care is expected to rise sharply [8,10,11]. This makes it crucial for nursing homes to optimize their resource allocation while providing excellent care. Embracing the concept of quality of care (QoC) is not just a strategic necessity for survival in a competitive market but also a moral obligation to ensure the well-being of older adults.

The United States recorded 649,611 excess deaths more than expected in the first year of the COVID-19 pandemic, a 23% increase over the previous year. The largest share of excess deaths occurred among older adults [12]. In the nursing home setting, older adults were disproportionately affected [2] because of a combination of individual factors, such as advanced age [13,14] and comorbidities [14,15], and organizational factors, such as understaffing, crowding [2], limited supply of personal protective equipment [16], and inadequate infection prevention and control readiness [2,17,18]. While the Centers for Disease Control and Prevention proposed rigorous COVID-19 control measures, including maintaining personal hygiene practices, mask use, self-isolation, mobility restrictions, and physical distancing [19,20], they could not compensate for the chronic noncompliance of infection prevention and control practices. A May 2020 United States Government Accountability Office report found that 82% of Centers for Medicare and Medicaid Services (CMS)–certified nursing homes had at least one infection-related deficiency between 2013 and 2017, with nearly half receiving citations over multiple years [17,21].

Aimed at deterring noncompliance and promoting accountability, CMS launched the Nursing Home Care (NHC) website in 2002 to improve transparency by expanding public access to quality information [22]. However, despite CMS's efforts, this website is not well known and is considered difficult to comprehend. When consulting NHC, consumers are unable to discern meaningful differences between nursing homes to make informed decisions about their well-being and that of their loved ones. To address this problem, the CMS introduced the 5-Star Quality Rating System for nursing homes in 2008, summarizing the detailed information provided on NHC [23].

The 5-Star Nursing Homes Quality Rating System for Nursing Homes is based on data from the Minimum Data Set (MDS) and includes several quality measures (QMs) derived from it. Concerns have been raised about the effectiveness of the 5-Star Quality Rating System for nursing homes because of its reliance on the MDS. The MDS has several limitations, including inconsistent reporting due to ambiguous instructions and subjective items [24]. In addition, comparing across facilities proves difficult when the QMs emphasize rare events. Such events usually lead to large SEs and large CIs, making it challenging to assess true quality differences [24]. The QMs are assumed to be linear; however, surprisingly, they are sometimes nonlinear [24,25]. The MDS QMs are also presumed to use the complete spectrum of possible values even in situations that deviate from medical norms, such as pressure ulcer rates below 2% [24,26]. Furthermore, ascertainment bias, a type of detection bias caused by inadequate recording of relevant QMs, and substantial interrater variability often impact the reliability of MDS data [24].

Additionally, nursing home studies face other challenges beyond those associated with the sector's reliance on the MDS. One problem is the absence of resident-level data, which impedes the establishment of causality. Moreover, endogeneity is inadequately controlled due to the nonindependence of dependent variables and error terms, which leads to biased results [27-29]. Additionally, the omission of risk adjustment for relevant confounders may make it impossible to capture any associations between a confounder and the other variables [27,30]. In this study, we used a panel dataset to evaluate the effectiveness of several nursing home QMs in explaining nursing home QoC in terms of COVID-19 health outcomes. We achieved this by regressing COVID-19 outcomes measured in the US nursing homes during the period May 2020 to June 2023 on nursing home QMs, classified using the Donabedian structure-process-outcome (SPO) framework, which are used in constructing the 5-Star Quality Rating System for nursing homes. Our research contributes to the literature in several ways. First, most scholars who used COVID-19 health outcomes have so far considered only a limited number of nursing home QMs. Moreover, they have relied on small sample sizes or used cross-sectional data over longitudinal data. We addressed these limitations by using a panel dataset of COVID-19 data for the nursing homes in the United States collected during the period May 2020 to July 2023. Second, to tackle the challenge of large SEs associated with rare events, we opted for a larger sample size, thus enhancing statistical power and refining our estimates. Third, to mitigate the concern of inadequate control of endogeneity, we used hierarchical generalized linear mixed effects modeling.

We hypothesized that nursing homes with better structural quality (eg, greater staff availability and better skill mix), better process quality (eg, lower restraint use and higher vaccination rates), and better outcome quality (eg, lower number of residents with pressure ulcers and lower number of resident falls) experienced better COVID-19 performance in terms of resident and staff infections and deaths. An empirical analysis of this issue offers fresh perspectives that were previously unavailable from existing research.

Methods

Data Sources

To test the study's hypotheses, we used a panel dataset created using 3 major publicly available data sources: COVID-19 nursing home data, Payroll-Based Journal data, and nursing home QM snapshots.

XSL•FO
RenderX

COVID-19 health outcomes for the period May 2020 to June 2023 were obtained from COVID-19 nursing home data. The latter were obtained as a single dataset containing weekly summaries of resident and staff infections and deaths. Payroll-Based Journal data are collected quarterly and offer daily summaries of employee weekly hours for different staff, including nursing, non-nursing, employee, and contract staff. Nursing home QM snapshots include monthly summaries of QMs obtained from multiple data sources, such as the MDS, claims, penalties, provider information, survey summaries, and vaccination data.

The data were aggregated to create a unified dataset. Data aggregation was performed using 2 unique identifiers: one identified the facility (ie, the federal provider number), whereas the other identified the period (ie, week, month, or quarter). All statistical analyses were performed using R version 4.0.0 or higher (R Foundation for Statistical Computing) [31], and a 2-sided *P* less than .05 was considered statistically significant.

Dependent Variables

The measures "residents weekly confirmed COVID-19," "residents weekly COVID-19 deaths," "staff weekly confirmed COVID-19," and "staff weekly COVID-19 deaths" monitored the number of residents or staff who tested positive for COVID-19 and those who died due to COVID-19 in a particular week. These figures were reported by providers weekly from May 24, 2020, to June 6, 2023, with the data for the week ending May 25, 2020, potentially including reporting from January 1, 2020, to May 24, 2020.

In the analysis, we combined "residents weekly COVID-19 deaths" and "staff weekly COVID-19 deaths" into a single value representing "total (resident and staff) weekly COVID-19 deaths." We adopted this approach because both variables were characterized by sparsity and a high number of zeros, and combining them in this manner increased statistical power and, subsequently, the variance for modeling of zero-inflated outcome variables [32].

Independent Variables

The set of independent variables was categorized based on the Donabedian SPO framework. According to this framework, the quality of health care, including nursing home care, can be assessed at a facility based on 3 quality components: structure, process, and outcome. *Structure* refers to the settings where health care is provided (eg, buildings and technology infrastructure) and accessibility features. *Process* encompasses the actions taken in giving and receiving care, such as pain management, error prevention, and care follow-ups. *Outcome* indicates the consequences of the provided health care, including mortality rates, readmission rates, and functional status. The quality of health care is contingent upon the interplay of these categories, as Donabedian eloquently stated, "A good structure increases the likelihood of good process, and good process increases the likelihood of good outcomes" [27,33].

Structural QMs

For our analysis, we used a comprehensive array of structural QMs encompassing various nursing home attributes (Textbox 1). To gauge the workforce dynamics, we factored in staffing hours, which were calculated separately for employees and contract staff and for nursing and non-nursing staff; we also examined turnover rates for nursing personnel, registered nurses, and non-nursing staff. Furthermore, we considered characteristics tied to the 5-Star Quality Rating System for nursing homes, including facility fines, penalties, and health and fire safety deficiencies, alongside organizational attributes, such as ownership type, provider category, special focus status, and the presence of resident and family councils, all of which contributed to our assessment of overall performance and stability.

XSL•FO
RenderX

Textbox 1. Structural quality measures (count model predictors). "N" denotes a numeric variable, whereas "C" denotes a categorical variable. The number preceding "C" represents the number of levels in the categorical variable.

Provider information model variables

- Percentage of occupied beds (N)
- Provider type (4C)
- Provider resides in hospital (2C)
- Days since approval to provide Medicare and Medicaid services (N)
- Continuing care retirement community (2C)
- Special focus status (2C)
- Abuse icon (2C)
- Most recent health inspection more than 2 years (2C)
- Provider changed owner in previous 12 months (2C)
- With a resident and family council (4C)

Penalties and staffing model variables

- Employee nursing total weekly hours (N)
- Employee non-nursing total weekly hours (N)
- Contract nursing total weekly hours (N)
- Contract non-nursing total weekly hours (N)

Process QMs

An ample set of process QMs was considered (Textbox 2). Among these were factors pertinent to COVID-19, such as the number of residents hospitalized with confirmed cases and their vaccination status, along with weekly admissions of residents previously treated for the virus. The health care aspects included measures such as the percentage of long-stay residents appropriately receiving pneumococcal and influenza vaccines, the percentage of long-stay residents receiving antianxiety or hypnotic medications, those subjected to physical restraints, and those experiencing adverse effects of in-dwelling catheters. For short-stay residents, the metrics focused on appropriate pneumococcal and influenza vaccination rates and the percentage of residents receiving newly prescribed antipsychotic medications.

Textbox 2. Process quality measures (count model predictors). "N" denotes a numeric variable.

Process model variables

- Residents' weekly COVID-19 admissions (N)
- Residents hospitalized with confirmed COVID-19 (N)
- Residents hospitalized with confirmed COVID-19 and up to date with vaccines (N)
- Percentage of current residents up to date with COVID-19 vaccines (N)
- Percentage of current health care personnel up to date with COVID-19 vaccines (N)
- Percentage of long-stay residents assessed and appropriately given the pneumococcal vaccine (N)
- Percentage of long-stay residents assessed and appropriately given the seasonal influenza vaccine (N)
- Percentage of long-stay residents who have received an antianxiety or hypnotic medication (N)
- Percentage of long-stay residents who have received an antipsychotic medication (N)
- Percentage of long-stay residents who are physically restrained (N)
- Percentage of long-stay residents with catheters inserted and left in their bladders (N)
- Percentage of short-stay residents assessed and appropriately given the pneumococcal vaccine (N)
- Percentage of short-stay residents who have received a newly prescribed antipsychotic medication (N)
- Percentage of short-stay residents assessed and appropriately given the seasonal influenza vaccine (N)

XSL•FO
RenderX

Outcome QMs

The outcome QMs used for the analysis encompassed various measures of resident care and well-being (Textbox 3). Long-stay measures focusing on depressive symptoms, weight loss, deteriorating mobility, increased need for assistance with daily activities, urinary tract infections (UTIs), and loss of bowel or bladder control were included. Short-stay measures evaluated functional improvements in mobility, outpatient emergency department visits, and rehospitalizations within 30 days of admission.

Textbox 3. Outcome quality measures (count model predictors). "N" denotes a numeric variable.

Outcome model variables

- Number of residents with a new positive COVID-19 test result (N)
- Number of staff and personnel with a new positive COVID-19 test result (N)
- Percentage of SNF residents with new or worsened pressure ulcers (N)
- Percentage of high-risk long-stay residents with pressure ulcers (N)
- Percentage of long-stay residents who have experienced one or more falls with major injury (N)
- Percentage of long-stay residents who have depressive symptoms (N)
- Percentage of long-stay residents who have lost too much weight (N)
- Percentage of long-stay residents whose ability to move independently has worsened (N)
- Percentage of long-stay residents whose need for help with daily activities has increased (N)
- Percentage of long-stay residents with urinary tract infections (N)
- Percentage of low-risk long-stay residents who lose control of their bowels or bladders (N)
- Percentage of short-stay residents who have made improvements in function (N)
- Percentage of short-stay residents who have had an outpatient emergency department visit (N)
- Percentage of short-stay residents who have been rehospitalized after a nursing home admission (N)

Penalties and staffing model variables

- Total fines (N)
- Total amount (N)
- Total penalties (N)
- Total days (N)
- Number of facility-reported incidents (N)
- Total nursing staff turnover (N)
- Registered nurse turnover (N)
- Number of administrators who have left the nursing home (N)

Surveys model variables

- Total health deficiencies inspection cycle 1 (N)
- Total health deficiencies inspection cycle 2 (N)
- Total health deficiencies inspection cycle 3 (N)
- Total fire safety deficiencies inspection cycle 1 (N)
- Total fire safety deficiencies inspection cycle 2 (N)
- Total fire safety deficiencies inspection cycle 3 (N)
- Total weighted health survey score (N)

Data Cleaning

Prior to analysis, we removed all the observations in the dataset that contained a flag for either incorrect data or substandard data. This flagging was part of the Centers for Disease Control and Prevention's quality assurance check, conducted by CMS on 8 data fields to ensure they did not contain implausible or erroneous values. Our decision to exclude flagged observations mirrors those made in prior research in the field [34]. Removing flagged observations eliminated 77,684 observations, resulting in a dataset containing 2,429,283 observations. The latter were used in the analysis after imputing with random forest. The

XSL•FO
RenderX

missing value profiles of the variables are shown below (Figures 1-5).

Figure 1. Missing value profile of the structural quality measures: surveys.

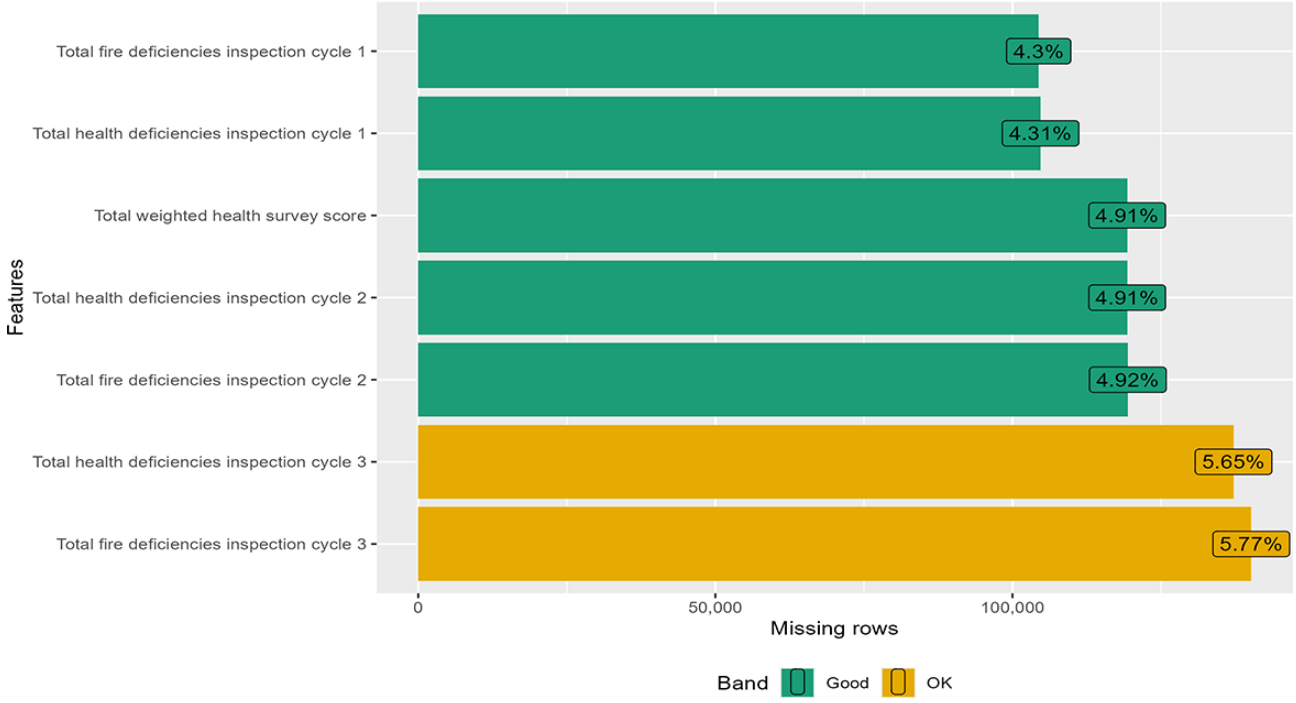
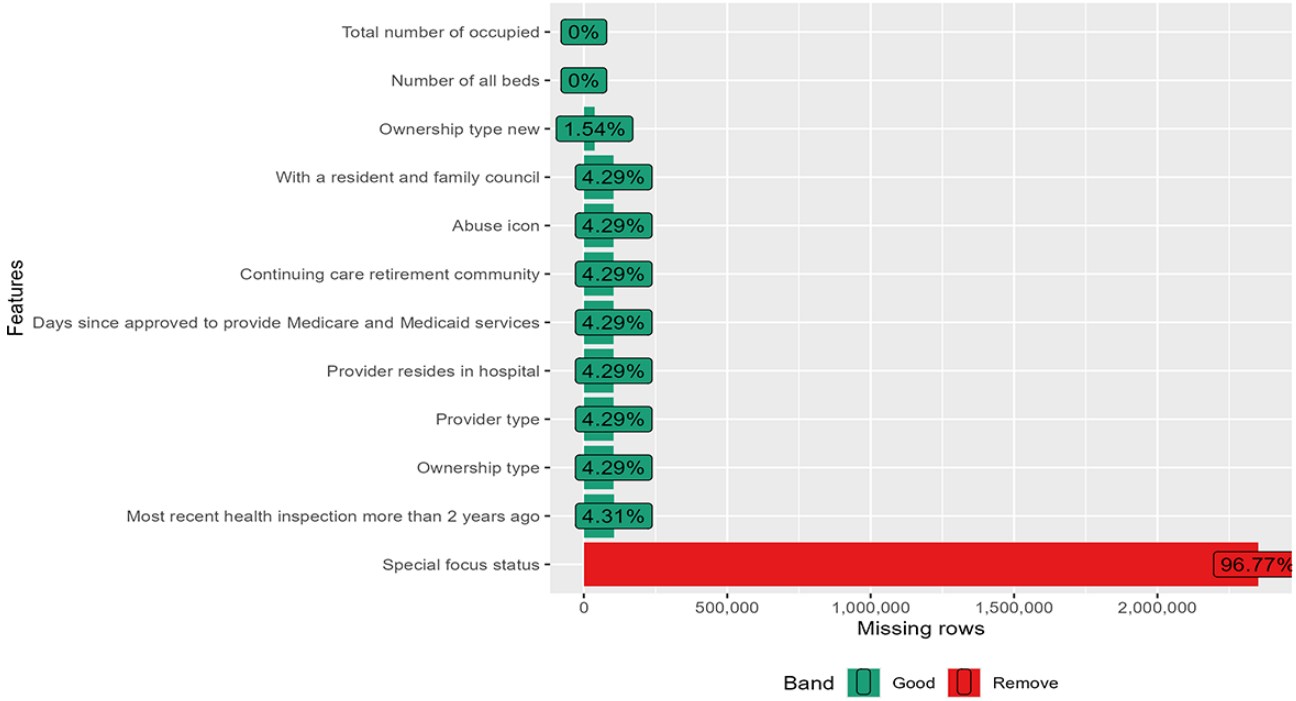


Figure 2. Missing value profile of the structural quality measures: provider information.



XSL•FO
RenderX

Figure 3. Missing value profile of the structural quality measures: penalties and staffing.

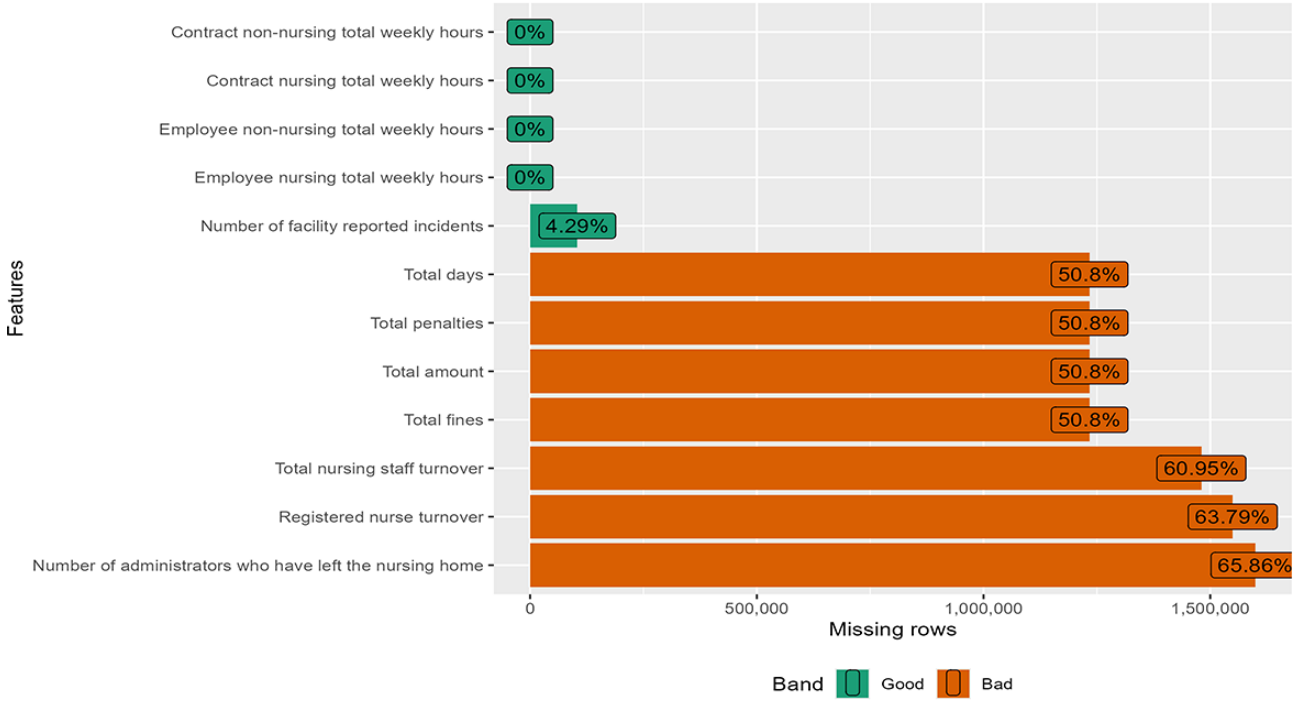
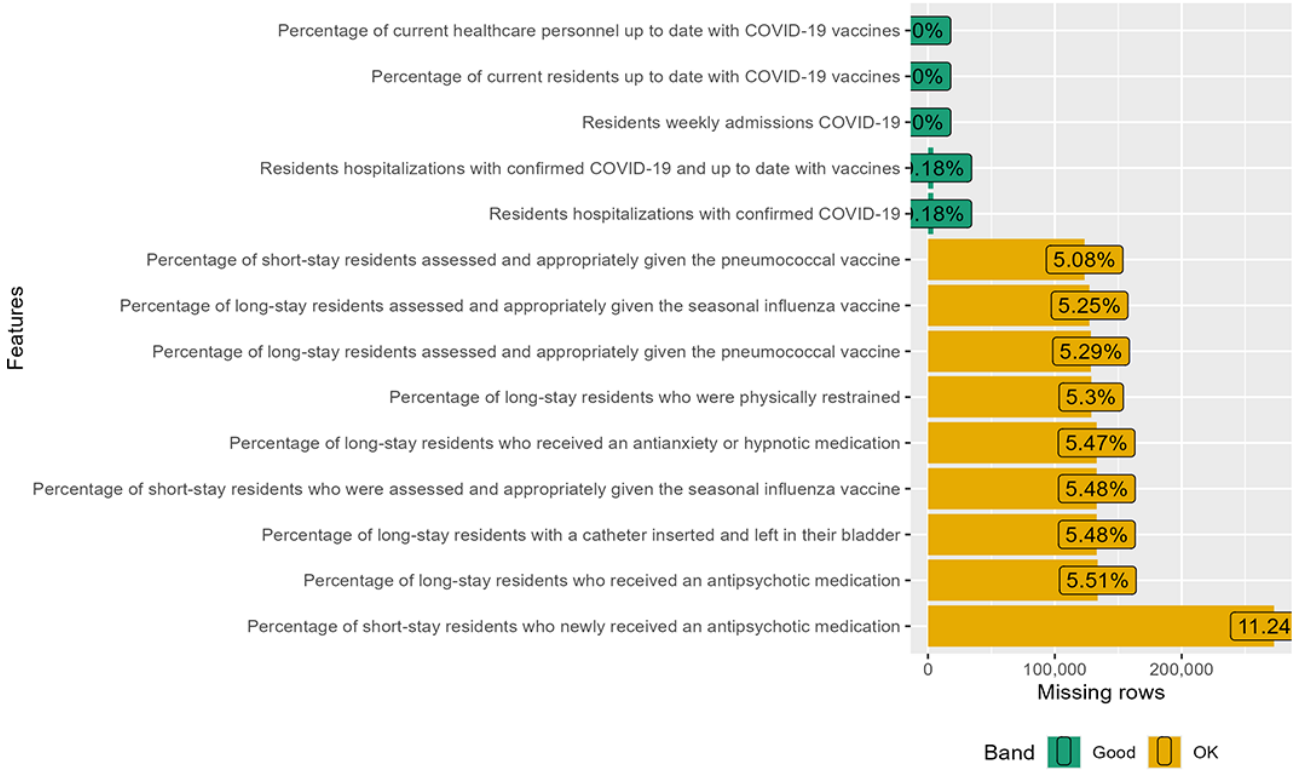
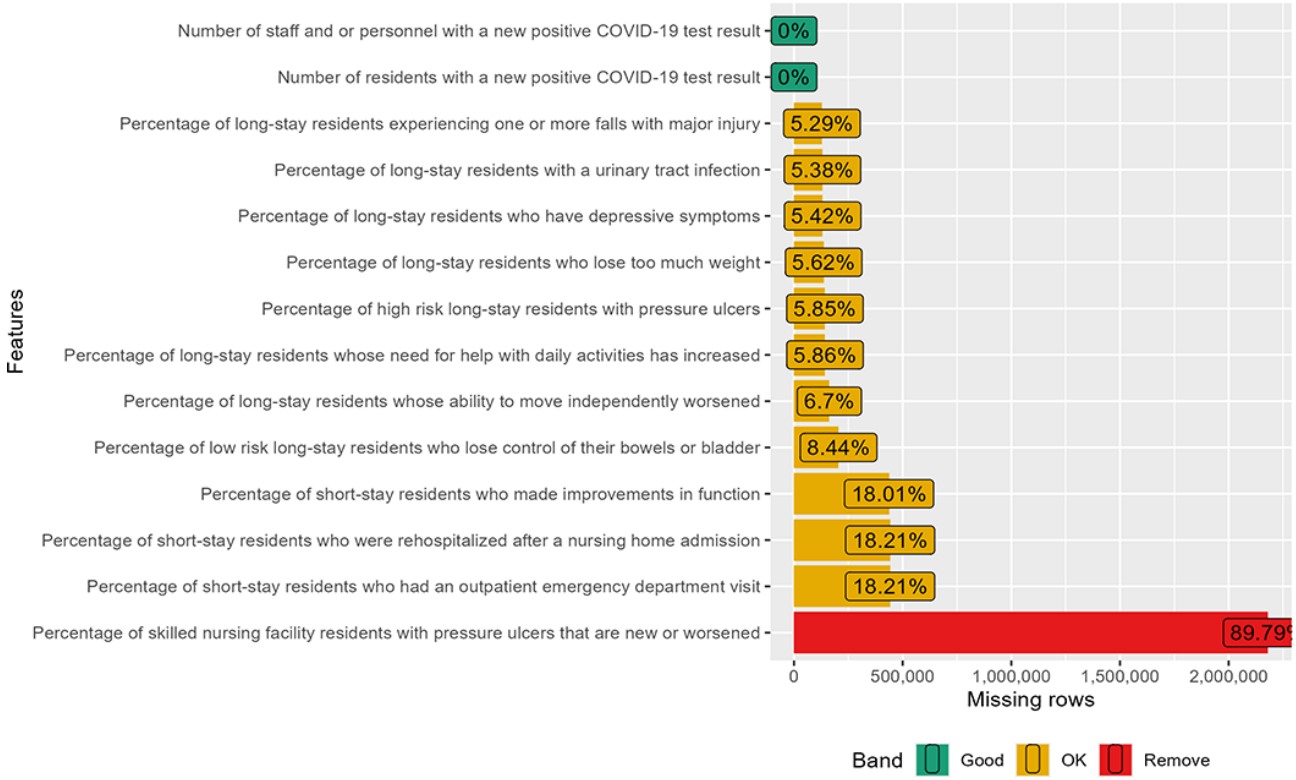


Figure 4. Missing value profile of the process quality measures.



XSL•FO
RenderX

Figure 5. Missing value profile of the outcome quality measures.



Missing Value Imputation

When addressing the presence of missing values in our dataset, we made the deliberate choice to impute rather than exclude them, which aligned with the research goals and the nature of the missing data. In our study, we assumed that the missing data were missing at random (MAR). The MAR mechanism assumes that the probability of missingness depends on the observed data but not on the missing data [35]. As the nursing home data are aggregated at the institution level and include all institutions nationwide, missingness is likely related to differences in administrative or operational factors, such as staffing levels, documentation practices, reporting systems, or governance policies at the state, county, or local levels. We opted for random forest imputation, as it handles various data types seamlessly; it also excels in managing complex interactions and multicollinearity while efficiently scaling to accommodate large datasets. The implementation of random forest in the R package *missranger* [36] resulted in a reduction of up to 50% of imputation errors, and the average proportional deviation in out-of-bag imputation errors remained within the 10% to 15% range. All the variables (except the outcome variables) were imputed using the random forest algorithm. Although the best method to address MAR missingness is to perform multiple imputation [37], we were unable to adopt this approach due to time and resource constraints. The specifications used for random forest imputation are presented in Table 1.

Table . Specifications used for random forest imputation.

Parameter	Value
Predictive mean matching	5
Maximum iterations	20
Number of trees	20
Maximum depth	8

Data Modeling

Data Modeling

Using the imputed dataset, the variable called “percentage of occupied beds” was created using “total number of occupied beds” and “number of all beds.” A total of 10 observations from the calculation resulted in a not-a-number error and were excluded. All the numeric variables were standardized to have a mean of 0 and an SD of 1. All the unordered categorical variables or nominal variables were specified using sum contrasts. For ordered categorical variables or ordinal variables, we used orthogonal polynomial contrasts, and we standardized them onto a uniform scale. Consequently, each contrast column had a mean of 0 and an SD of 1. This approach allowed us to fit the regression models within a standardized framework and easily compare different model coefficients.

Furthermore, all unordered categorical variables containing more than 2 levels were binary encoded before being specified

XSL•FO
RenderX

using sum contrasts. Binary encoding is a machine learning technique that combines hash encoding and one-hot encoding. First, all levels of an unordered categorical variable are expressed using a unique numeric value. Then, the numbers are transformed into a unique combination of zeros and ones, making it efficient to store data that exhibit high cardinality.

The outcome variables were characteristic of count data and exhibited properties such as nonnegativity, integer values, positive skewness, and a greater prevalence of lower count values. Equidispersion assessment revealed significant overdispersion, which rendered the use of a traditional Poisson regression model unsuitable. Therefore, we used a negative binomial regression model. Unlike the Poisson model, the negative binomial model accounts for overdispersion with quadratic parameterization. The presence of excess zeros led to the implementation of a zero-inflated negative binomial (ZINB) model, and the panel structure of the dataset resulted in the specification of a mixed effects model. However, specifying a ZINB model requires the presence of zero inflation at all levels of the grouping variables (ie, states, counties, and facilities). Some of the levels did not have zero inflation; they had zero deflation. On the basis of recommendations in the literature [38], the final model we specified was a truncated ZINB mixed effects model.

We estimated multiple truncated ZINB models, specifying different outcome variables. To model the outcomes "residents weekly confirmed COVID-19" and "staff weekly confirmed COVID-19," the zero inflation (ZI) component of the model included factors influencing initial susceptibility to COVID-19 [39]. These included facility-level characteristics, such as bed size, as well as social determinants of health attributes, including urbanicity, political affiliation of the geographic location, median household income, and local COVID-19 vaccination coverage. To model the outcome "total weekly COVID-19 deaths," the ZI component of the model included factors influencing the possibility of death after COVID-19 had been contracted, such as lagged weekly resident confirmed COVID-19 cases per 1000 residents and lagged county confirmed COVID-19 cases per 1000 people. We assumed that newly infected residents (ie, infected at 1 time point before) had a higher likelihood of death compared to those who were not newly infected (ie, infected more than one time point before). Our assumption was informed by prior research, indicating that individuals who test positive for COVID-19 typically do not begin to develop detectable antibodies until approximately 1 week after the onset of symptoms [40]. Both lagged variables were calculated for a lag of 1. An overview of the outcome variables and their associated ZI model predictors is presented in (Textbox 4).

Textbox 4. Model outcomes and ZI model predictors. Letter "N" denotes a numeric variable, whereas "C" denotes a categorical variable. The number preceding "C" represents the number of levels in the categorical variable.

Residents weekly confirmed COVID-19

- Bed size (3C)
- Urban binary (2C)
- Democrat or Republican (2C)
- Median household income dollars inflation adjusted to data file year ACS (American Community Survey) 2016-2020 (N)
- Partially or fully vaccinated percent (N)

Staff weekly confirmed COVID-19

- Bed size (3C)
- Urban binary (2C)
- Democrat or Republican (2C)
- Median household income dollars inflation adjusted to data file year ACS 2016-2020 (N)
- Partially or fully vaccinated percent (N)

Total weekly COVID-19 deaths

- Lagged weekly resident confirmed COVID-19 cases per 1000 residents (N)
- Lagged county confirmed cases USA Facts new (N)

The model included random slopes and random intercepts for 3 levels of nesting, that is, nursing home facilities nested within counties nested within states. An autoregressive process of order-1 covariance structure was selected to model the correlations among time points for the same individual. The truncated negative binomial distribution with quadratic parameterization implemented in the R package *glmmTMB* [41] under the family "truncated_nbinom2" was used.

Mathematically, the model can be expressed as follows:

$$\mu=E(\text{count} \mid u,NSZ)=\exp(\beta 0+\beta ixi+u) \forall \text{ count}>0$$

$$u\sim N(0,\sigma u2)$$

$$\sigma 2=\text{Var}(\text{count}|u, NSZ)= \mu(1+\mu\theta)\forall \text{ count}>0$$

$$\text{logitp}=\text{logp}1-p=\beta 0(zi)+\beta i(zi)xi$$

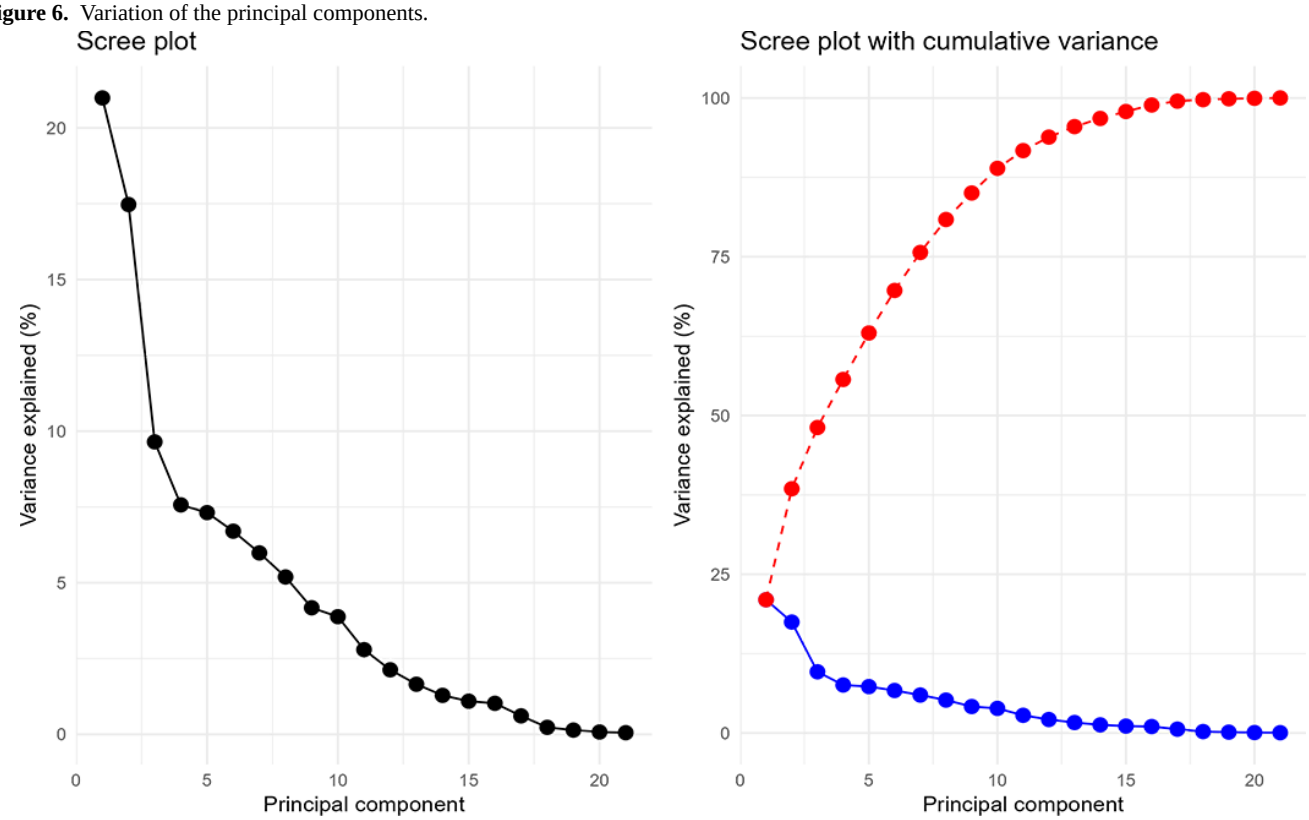
where subscript I denotes the time point, u denotes the random effects terms in the model, NSZ denotes the event "nonstructural

XSL•FO
RenderX

zeros," P=1−Pr(NSZ) is the ZI probability, and θ is the dispersion parameter specific to the family "nbinom2."

For each outcome variable, we created 5 base models. Each model contained a different subset of nursing home QMs, namely, outcome QMs, process QMs, and structural QMs (surveys, provider information, and penalties and staffing). Using a subset of significant predictors (30/58) from the 5 base models, we performed principal component analysis (PCA) for dimensionality reduction on the mean-aggregated dataset, where data were averaged across repeated measures for each entity. Using the loadings (eigenvectors) from this PCA, we calculated principal component scores for the complete dataset, including all the repeated measures. Using the PCA scores as predictors, we created 2 predictive models for the states and the components of the 5-Star Quality Rating System of nursing homes. The first 8 principal components explained 80.86% of the variation present in 30 significant predictors (Figure 6).

Figure 6. Variation of the principal components.



Model Diagnostics

We performed model fit assessment through a combination of graphical diagnostics, goodness-of-fit statistics, and residual analysis. First, observed versus predicted count distributions were compared using histograms to visually assess the model's ability to accurately predict both zero and nonzero values. Second, goodness-of-fit statistics were computed, including the log-likelihood, Akaike Information Criterion, Bayesian Information Criterion, and deviance to evaluate and compare across statistical models. We selected several goodness-of-fit measures to achieve a more robust evaluation of model performance. Third, the Vuong test was applied to compare the ZINB model against standard negative binomial and Poisson models to determine whether the inclusion of a ZI component was statistically justified. Finally, residual diagnostics were performed using the R package *DHARMa* [42], and goodness-of-fit was evaluated through nonparametric, simulation-driven methods. *DHARMa* provided an assessment of outliers, over- or underdispersion, ZI, and the independence of observations by simulating new response data, computing empirical cumulative density functions, and defining residuals based on observed versus simulated values, ensuring a robust model adequacy assessment.

Ethical Considerations

This study is exempt from the institutional review board, as it involved only the use of publicly available data and did not involve human participants, in accordance with 45 CFR 46.102(e). No private, identifiable information was collected or analyzed.

Results

For the models involving outcome QMs as predictors, the number of new COVID-19 cases among residents (β=0.46, SE 0.002; β=0.02, SE 0. 001; and β=0.02, SE 0.003) and staff (β=0.01, SE 0.001; β=0.51, SE 0.002; and β=0.05, SE 0.004) were highly significant, with all the outcomes at the significance levels of P<.001 and P<.001 respectively. The percentage of long-stay residents with UTIs was significant, with staff infections as the outcome (β=−0.004, SE 0.002) at the significance level of P<.10 (P=0.06), resident infections as the outcome (β=−0.01, SE 0.003) at the significance level of P<.001, and total deaths as the outcome (β=−0.03, SE 0.009) at the significance level of P<.01 (P=.001).

XSL•FO
RenderX

For the models involving process QMs as predictors, the indicator concerning residents' weekly COVID-19 admissions (β=0.05, SE 0.003; β=0.04, SE 0.001; and β=0.048, SE 0.002) was significant, with all the outcomes at the significance level of *P*<.001.

Residents' hospitalizations with confirmed COVID-19 (β=0.03, SE 0.002 and β=0.02, SE 0.001), the percentage of current residents up to date with COVID-19 vaccines (β=0.06, SE 0.007 and β=0.02, SE 0.005), and the percentage of long-stay residents assessed and appropriately given the seasonal influenza vaccine (β=0.02, SE 0.005 and β=0.04, SE 0.01) were significant, with at least two of the outcomes at the significance levels of *P*<.001, *P*<.001, and *P*<.001, respectively.

Residents hospitalized with confirmed COVID-19 and up to date with vaccines (β=0.03, SE 0.005 and β=0.05, SE 0.02) and the percentage of long-stay residents with catheters inserted and left in their bladders (β=−0.05, SE 0.005; β=−0.01, SE 0.003; and β=−0.02, SE 0.009) were significant, with at least two of the outcomes at the significance level of *P*<.001 and *P*<.05, respectively.

For the models involving structural QMs (surveys) as predictors, total health deficiencies inspection cycle 2 (β=0.05, SE 0.005 and β=0.01, SE 0.004) was significant, with at least two outcomes at the *P*<.001 significance level.

Total health deficiencies inspection cycle 1 (β=0.04, SE 0.01 and β=0.01, SE 0.004), total health deficiencies inspection cycle 3 (β=0.04, SE 0.01 and β=0.01, SE 0.004), and total fire safety deficiencies inspection cycle 3 (β=0.02, SE 0.01 and β=0.01, SE 0.004) were significant, with at least one of the outcomes at the significance level of *P*<.001, *P*<.05, and *P*<.05, respectively.

For the models involving structural QM (provider information) as predictors, percentage of occupied beds (β=0.04, SE 0.005; β=−0.06, SE 0.004; and β=−0.25, SE 0.01), provider resides in hospital (β=0.19, SE 0.02; β=0.07, SE 0.01; and β=0.13, SE 0.03), and days since approval to provide Medicare and Medicaid services (β=0.08, SE 0.005; β=0.02, SE 0.005; and β=0.04, SE 0.01) were significant, with all outcomes at the *P*<.001, *P*<.001, and *P*<.001 significance levels, respectively.

Continuing care retirement community (β=0.08, SE 0.009 and β=0.06, SE 0.02) and at least one of the 2 binary components of provider type (component 1: β=0.26, SE 0.03; β=0.08, SE 0.03; and β=0.25, SE 0.06), with a resident and family council (component 1: β=0.02, SE 0.005; β=0.05, SE 0.01 and component 2: β=0.08, SE 0.01; β=0.03, SE 0.01; and β=0.16, SE 0.03), and new ownership type (component 2: β=0.05, SE 0.01 and β=−0.07, SE 0.01) were significant, with at least two of the outcomes at the *P*<.001, *P*<.001, *P*<.001, *P*<.001, and *P*<.001 significance level.

For the models involving structural QMs (penalties and staffing) as predictors, all the staffing variables, namely employee nurse total weekly hours (β=0.12, SE 0.006; β=0.09, SE 0.005; and β=0.06, SE 0.01), employee nonnursing total weekly hours (β=−1.06, SE 0.006; β=−0.09, SE 0.005; and β=−0.05, SE 0.01), contract nurse total weekly hours (β=0.1, SE 0.004; β=0.08, SE 0.003; and β=0.10, SE 0.008), and contract nonnursing total weekly hours (β=−0.1, SE 0.01; β=−0.1, SE 0.004; and β=−0.04, SE 0.01) were significant, with all the outcomes at the *P*<.001 significance level, respectively.

Component 3 of provider state (component 3: β=−0.23, SE 0.05; β=−0.23, SE 0.05; and β=−0.20, SE 0.06) was significant at the *P*<.001 significance level. None of the components of the provider county achieved a high level of significance.

The linear (β=0.08, SE 0.008 and β=0.05, SE 0.006), quadratic (β=0.02, SE 0.005 and β=0.02, SE 0.004), and cubic (β=0.03, SE 0.005 and β=0.02, SE 0.003) components of the health inspection rating were significant, with resident infections and staff infections as outcomes at the *P*<.001, *P*<.001, and *P*<.001 significance levels, respectively. However, only the quartic (β=−0.0156, SE 0.0079) component of the health inspection rating was significant, with total deaths as the outcome at the *P*<.05 (*P*=.048) significance level.

The QM rating had significant linear (β=−0.02, SE 0.01) and cubic components (β=−0.01, SE 0.005) in the model, with resident infections as the outcome at the *P*<.05 (*P*=.002) and *P*<.05 (*P*=.045) significance levels, respectively. In addition to the linear (β=0.0164, SE 0.0055) and cubic (β=−0.009, SE 0.003) components, the quartic (β=−0.006, SE 0.003) component was significant at the *P*<.05 (*P*=.003), *P*<.05 (*P*=.01), and *P*<.05 (*P*=.04) significance levels, respectively, in the model with staff infections as the outcome. The quadratic (β=−0.02, SE 0.01) and cubic (β=0.02, SE 0.01) components were significant at the *P*<.05 (*P*=.04) and *P*<.05 (*P*=.045) significance levels, respectively, in the model with total deaths as the outcome.

Staffing rating had a significant linear (β=0.02, SE 0.005) component in the models with staff infections and total deaths as the outcomes at the *P*<.001 significance level. The model with resident infections as the outcome only showed a significant cubic (β=−0.008, SE 0.004) trend at the *P*<.1 (*P*=.06) significance level. In the model with total deaths as the outcome, both the linear (β=0.12, SE 0.01) and quadratic (β=−0.05, SE 0.01) components were significant at the *P*<.001 and *P*<.001 significance levels, respectively, and the quartic (β=−0.01, SE 0.007) component was significant at the *P*<.1 (*P*=.05) significance level.

The discussed results are summarized in Tables 2 and 3. Multimedia Appendix 1, containing the complete model results in the original (log) scale, and Multimedia Appendix 2, containing the visualization of their 95% CIs in the response (exponent) scale, are provided for further reference.

XSL•FO
RenderX

Table . Model results for base models. The table displays variables that were significant when regressed on more than 1 outcome. All estimates are presented on the original (log) scale and are conditioned on the variables included in the conditional model and the model random effects. Complete results are provided in Multimedia Appendix 1 and displayed graphically in Multimedia Appendix 2.

Variable	Model 1: resident infections, estimate (SE)	*P* value	Model 2: staff infections, estimate (SE)	*P* value	Model 3: total deaths, estimate (SE)	*P* value
Intercept	4.60 (0.05)	<.001	3.52 (0.04)	<.001	3.89 (0.06)	<.001
Number of residents with a new positive COVID-19 test result	0.46 (0.002)	<.001	0.02 (0.001)	<.001	0.02 (0.003)	<.001
Number of staff and/or personnel with a new positive COVID-19 test result	0.01 (0.001)	<.001	0.51 (0.002)	<.001	0.05 (0.004)	<.001
Percentage of long-stay residents whose ability to move independently worsened	0.01 (0.003)	.02	0.01 (0.003)	.007	−0.02 (0.01)	.22
Percentage of long-stay residents with a urinary tract infection	−0.02 (0.003)	<.001	−0.004 (0.002)	.06	−0.03 (0.01)	.001
Intercept	4.28 (0.08)	<.001	3.31 (0.06)	<.001	3.37 (0.06)	<.001
Residents weekly admissions COVID-19	0.05 (0.003)	<.001	0.04 (0.001)	<.001	0.05 (0.002)	<.001
Residents hospitalizations with confirmed COVID-19	0.03 (0.002)	<.001	0.02 (0.001)	<.001	0.01 (0.004)	.26
Residents hospitalizations with confirmed COVID-19 and up to date with vaccines	0.0003 (0.002)	.92	0.03 (0.005)	<.001	0.05 (0.02)	.001
Percentage of current residents up to date with COVID-19 vaccines	−0.06 (0.01)	<.001	−0.02 (0.005)	<.001	−0.10 (0.03)	.002
Percentage of current health care personnel up to date with COVID-19 vaccines	0.03 (0.01)	<.001	0.01 (0.004)	.04	−0.001 (0.03)	.99
Percentage of long-stay residents assessed and appropriately given the seasonal influenza vaccine	0.02 (0.01)	<.001	0.0001 (0.004)	.97	0.04 (0.01)	<.001
Percentage of long-stay residents who received an antipsychotic medication	0.04 (0.01)	<.001	0.01 (0.005)	.13	0.2 (0.1)	.06
Percentage of long-stay residents who were physically restrained	−0.02 (0.01)	<.001	−0.002 (0.004)	.70	.02 (0.01)	.09

XSL•FO
RenderX

Variable	Model 1: resident infections, estimate (SE)	*P* value	Model 2: staff infections, estimate (SE)	*P* value	Model 3: total deaths, estimate (SE)	*P* value
Percentage of long-stay residents with a catheter inserted and left in their bladder	−0.05 (0.005)	<.001	−0.01 (0.003)	.04	−0.02 (0.01)	.009
Percentage of short-stay residents who were assessed and appropriately given the seasonal influenza vaccine	−0.04 (0.01)	<.001	0.001 (0.01)	.85	−0.0004 (0.01)	.03
Intercept	4.70 (0.08)	<.001	3.53 (0.06)	<.001	3.86 (0.06)	<.001
Total health deficiencies inspection cycle	0.04 (0.01)	<.001	0.01 (0.005)	.009	0.02 (0.01)	.09
Total health deficiencies inspection cycle 2	0.05 (0.01)	<.001	0.01 (0.004)	<.001	0.01 (0.01)	.30
Total health deficiencies inspection cycle 3	0.04 (0.01)	<.001	0.01 (0.004)	.007	0.02 (0.01)	.10
Total fire deficiencies inspection cycle 3	0.02 (0.005)	<.001	0.01 (0.004)	.006	0.02 (0.01)	.06
Intercept	3.98 (0.08)	<.001	3.35 (0.06)	<.001	3.13 (0.09)	<.001
Percent of occupied beds	0.04 (0.01)	<.001	−0.06 (0.004)	<.001	−0.25 (0.01)	<.001
Provider type b1	0.26 (0.03)	<.001	0.08 (0.03)	.001	0.25 (0.06)	<.001
Provider type b2	−0.02 (0.02)	.07	−0.001 (0.02)	.01	−0.03 (0.05)	.09
Provider resides in hospital	0.18 (0.02)	<.001	0.07 (0.01)	<.001	0.13 (0.03)	<.001
Days since approved to provide Medicare and Medicaid services	0.08 (0.01)	<.001	0.02 (0.005)	<.001	0.04 (0.01)	<.001
Continuing care retirement community	0.08 (0.01)	<.001	−0.01 (0.01)	.43	0.06 (0.02)	<.001
Special focus status	−0.03 (0.004)	<.001	−0.003 (0.004)	.07	−0.01 (0.01)	.29
With a resident and family council b1	0.01 (0.01)	.11	0.02 (0.01)	<.001	0.05 (0.01)	<.001
With a resident and family council b2	0.08 (0.01)	<.001	0.03 (0.01)	.001	0.16 (0.03)	<.001
Ownership type new b1	0.01 (0.01)	<.001	0.02 (0.01)	.06	0.04 (0.01)	.002
Ownership type new b2	0.05 (0.01)	<.001	−0.07 (0.01)	<.001	−0.03 (0.02)	.14
Intercept	4.70 (0.08)	<.001	3.52 (0.06)	<.001	3.85 (0.06)	<.001
Total days	0.01 (0.01)	.08	0.01 (0.004)	.002	0.004 (0.01)	.72
Employee nursing total weekly hours	0.12 (0.01)	<.001	0.09 (0.01)	<.001	0.06 (0.01)	<.001

XSL•FO
RenderX

Variable	Model 1: resident infections, estimate (SE)	*P* value	Model 2: staff infections, estimate (SE)	*P* value	Model 3: total deaths, estimate (SE)	*P* value
Employee nonnursing total weekly hours	−0.11 (0.01)	<.001	−0.09 (0.05)	<.001	−0.05 (0.01)	<.001
Contract nursing total weekly hours	0.10 (0.004)	<.001	0.08 (0.003)	<.001	0.10 (0.01)	<.001
Contract non-nursing total weekly hours	−0.10 (0.01)	<.001	−0.09 (0.004)	<.001	−0.04 (0.01)	<.001
Number of facility reported incidents	0.02 (0.01)	<.001	−0.001 (0.004)	.80	0.02 (0.01)	.06
Registered nurse turnover	0.01 (0.005)	.002	−0.01 (0.003)	.008	0.01 (0.01)	.09

XSL•FO
RenderX

Table . Model results for prediction models. The table displays variables that were significant when regressed on more than one outcome. All estimates are presented on the original (log) scale and are conditioned on the variables included in the conditional model and the model random effects. Complete results are provided in Multimedia Appendix 1 and displayed graphically in Multimedia Appendix 2.

Variable	Model 1: resident infections, estimate (SE)	*P* value	Model 2: staff infections, estimate (SE)	*P* value	Model 3: total deaths, estimate (SE)	*P* value
Intercept rating	4.36 (0.05)	<.001	2.97 (0.05)	<.001	3.45 (0.06)	<.001
PC[a] rating	0.26 (0.01)	<.001	0.17 (0.004)	<.001	0.04 (0.01)	<.001
PC2 rating	0.59 (0.005)	<.001	0.53 (0.003)	<.001	0.02 (0.01)	<.001
PC3 rating	−0.06 (0.01)	<.001	−0.09 (0.004)	<.001	−0.01 (0.009)	.06
PC4 rating	0.12 (0.005)	<.001	0.16 (0.003)	<.001	−0.02 (0.01)	.003
PC5 rating	−0.005 (0.01)	.42	−0.06 (0.005)	<.001	0.03 (0.01)	.005
PC6 - rating	0.17 (0.01)	<.001	0.3 (0.04)	<.001	0.21 (0.01)	<.001
PC7 - rating	0.06 (0.01)	<.001	0.08 (0.004)	<.001	0.25 (0.01)	<.001
PC8 - rating	−0.09 (0.005)	<.001	−0.2 (0.004)	<.001	−0.08 (0.01)	<.001
Health inspection rating.l	0.08 (0.01)	<.001	0.05 (0.01)	<.001	0.02 (0.01)	.16
Health inspection rating.q	0.02 (0.01)	<.001	0.02 (0.004)	<.001	0.004 (0.01)	.67
Health inspection rating.c	0.03 (0.005)	<.001	0.02 (0.003)	<.001	0.01 (0.01)	.05
QM rating.l	−0.02 (0.01)	.002	0.02 (0.01)	.003	0.01 (0.01)	.54
QM rating.c	−0.01 (0.005)	.04	−0.01 (0.003)	.01	0.02 (0.01)	.05
Staffing rating.l	−0.01 (0.01)	.13	0.02 (0.005)	<.001	−0.12 (0.01)	<.001
Intercept - provider state	4.37 (0.06)	<.001	2.98 (0.06)	<.001	3.45 (0.07)	<.001
PC - provider state	−0.24 (0.004)	<.001	−0.15 (0.003)	<.001	−0.05 (0.01)	<.001
PC2 - provider state	0.59 (0.004)	<.001	0.53 (0.003)	<.001	0.03 (0.01)	<.001
PC3 - provider state	−0.08 (0.01)	<.001	−0.1 (0.004)	<.001	−0.01 (0.008)	.05
PC4 - provider state	−0.12 (0.005)	<.001	−0.16 (0.003)	<.001	0.02 (0.01)	<.001
PC5 - provider state	0.005 (0.01)	.74	−0.05 (0.005)	<.001	0.03 (0.01)	<.001
PC6 - provider state	0.16 (0.01)	<.001	0.29 (0.004)	<.001	0.19 (0.01)	<.001
PC7 - provider state	−0.08 (0.01)	<.001	−0.09 (0.004)	<.001	−0.26 (0.01)	<.001
PC8 - provider state	−0.09 (0.005)	<.001	−0.2 (0.004)	<.001	−0.07 (0.01)	<.001
Provider state b3	−0.23 (0.05)	<.001	−0.23 (0.05)	<.001	−0.2 (0.06)	<.001
Provider state b4	−0.01 (0.05)	.01	−0.03 (0.05)	.55	0.11 (0.06)	.07
Intercept - provider county	4.31 (0.07)	<.001	2.96 (0.07)	<.001	3.47 (0.08)	<.001
PC - provider county	−0.24 (0.004)	<.001	−0.15 (0.003)	<.001	−0.05 (0.01)	<.001
PC2 - provider county	0.59 (0.004)	<.001	0.53 (0.003)	<.001	0.03 (0.01)	<.001
PC3 - provider county	−0.08 (0.01)	<.001	−0.1 (0.004)	<.001	−0.01 (0.008)	.09

XSL•FO
RenderX

Variable	Model 1: resident infections, estimate (SE)	*P* value	Model 2: staff infections, estimate (SE)	*P* value	Model 3: total deaths, estimate (SE)	*P* value
PC4 - provider county	−0.12 (0.005)	<.001	−0.16 (0.003)	<.001	0.02 (0.01)	<.001
PC5 - provider county	0.004 (0.01)	.52	−0.05 (0.005)	<.001	0.03 (0.01)	<.001
PC6 - provider county	0.16 (0.01)	<.001	0.29 (0.004)	<.001	0.19 (0.01)	<.001
PC7 - provider county	−0.08 (0.01)	<.001	−0.09 (0.004)	<.001	−0.26 (0.01)	<.001
PC8 - provider county	−0.09 (0.005)	<.001	−0.2 (0.004)	<.001	−0.07 (0.01)	<.001

[a]PC: principal component.

Discussion

Principal Findings

In this study, we used a panel dataset to evaluate the effectiveness of several nursing home QMs in explaining nursing home QoC in terms of COVID-19 outcomes. Staffing measures (*P*<.001 for all variables in all models), health deficiency scores (*P*<.001 for all variables in at least 1 model), COVID-19 hospitalizations (*P*<.001 for all variables in at least 2 models), and vaccinations (*P*<.001 for all variables in at least 2 models) exhibited meaningful relationships with the COVID-19 outcomes. The 5-Star Quality Rating System for nursing homes, Medicaid dependency, and ownership showed no clear relationships with the COVID-19 outcomes.

We found a significant association with the percentage of long-stay residents with UTIs. In a retrospective study, the authors evaluated UTI diagnoses and antibiotic prescriptions in 622 COVID-19 hospital ward patients, and they found that 61% of cases had probably been overdiagnosed [43]. Several researchers studying long-term care facilities have also found evidence of overdiagnosis in the form of inappropriate initiation of antimicrobial treatment in 37% to 61% of patients [43-49].

We also identified significant associations with the percentages of long-stay and short-stay residents assessed and appropriately given the seasonal influenza vaccine. A study compared COVID-19 and seasonal influenza patient groups and suggested that preexisting chronic respiratory conditions more strongly impacted the severity of seasonal influenza than that of COVID-19 [50]. It has also been reported that chronic lower respiratory tract diseases claim the third highest number of lives among adults aged >65 years [51-55]. Therefore, older individuals with comorbidities for both respiratory infections may exhibit stronger symptoms of seasonal influenza than COVID-19, demanding management through the administration of seasonal influenza vaccines.

A significant association was uncovered for the percentage of long-stay residents with catheters inserted and left in their bladders. In a recent systematic review of 67 studies, scholars found that catheter prevalence among nursing home residents varied between 2.2% and 36.4%, with a typical rate of 7.3% [56,57]. Catheter-associated UTIs comprise 32% of all health care–associated infections [57,58]. The fact that catheter use is frequently reported in long-term care residents and is a leading cause of UTI helps explain our results, which show significant associations with COVID-19 outcomes for both the percentage of long-stay residents with UTIs and the percentage of long-stay residents with catheters inserted and left in their bladders.

Our findings indicate a significant association between health deficiency variables and resident and staff COVID-19 cases. During the period 2013 to 2017, more than four-fifths of nursing homes in the United States engaged in deficient infection prevention and control practices, including those that substantially reduced the spread of COVID-19, such as proper hand washing and isolation procedures [21,59]. It is possible that these poor practices continued during the pandemic, making nursing home residents and staff more susceptible to COVID-19 infections.

Several scholars have reported that lower resident density, measured by the proxy variable of the percentage of occupied beds, is associated with a lower prevalence of COVID-19 [59-65]. Our study showed a similar association with resident infections but an inverse association with staff infections and total deaths. We found that hospital-based nursing homes reported fewer COVID-19 cases and deaths than nonhospital-based nursing homes, which is consistent with the results of a study by Tarteret et al [66]. Gorges and Konetzka [67] demonstrated facility-level differences in QoC based on the proportion of residents with Medicaid coverage at the facility. Compared to nursing homes accepting Medicaid only, nursing homes accepting Medicare showed clearly better COVID-19 outcomes in only 2 of 3 models. In 1 model, the difference was less clear due to overlapping 95% CIs.

Evidence from various studies suggests that for-profit nursing homes are more prone to COVID-19 outbreaks compared to nonprofit and government nursing homes [59,61,65,68]. However, some scholars have found no statistically significant link in this regard [34,60,66]. We did not notice a clear difference between for-profit, nonprofit, and government nursing homes. Furthermore, no clear difference in outcomes was observed between nursing homes with or without continuing care retirement communities, resident councils, family councils, or resident and family councils. While not a key characteristic,

XSL•FO
RenderX

nursing homes newly approved to provide Medicare and Medicaid services had better COVID-19 outcomes compared to long-standing nursing homes. This could be because newer nursing homes underwent inspections more recently than previously approved homes as part of their approval process.

Inadequate nursing staff have been linked to a higher likelihood of COVID-19 outbreaks in some studies [59,62,64,65,68-70], but others indicate no significant connection [61,65,69,71]. Our findings show that increased staffing hours for nursing staff resulted in worse COVID-19 outcomes, whereas increased staffing hours for non-nursing staff resulted in better COVID-19 outcomes. We did not explore these associations further based on the types of nursing and nonnursing staff.

Most researchers have found no meaningful connection between nursing home overall quality rating and COVID-19 outcomes [34,60-62,69,72-79]. Similarly, most scholars have found no clear relationships between COVID-19 outcomes and 2 components of the 5-Star Quality Rating System for Nursing Homes, namely staffing rating and health inspection rating [61,64,70,73,78]. Our analysis did not reveal a clear relationship between the components of the 5-Star Quality Rating System for nursing homes and COVID-19 outcomes.

Limitations

This study has some limitations. It only considered temporal autocorrelation, without addressing spatial autocorrelation, thus ignoring potential interdependencies across spatial units. Temporal autocorrelation assumes that changes over time are isolated within units, which is rarely the case in interconnected systems, such as health care networks. We also did not perform multiple dataset imputations, which could have resulted in less robust statistical analyses, as a single imputation set may not adequately reflect the variability inherent in missing data. By generating a range of plausible datasets, MIs allow for more accurate and reliable parameter estimates by accounting for the uncertainty around the imputed values. Future scholars should consider modeling both spatial and temporal autocorrelations, as well as performing MIs, to increase the validity and reliability of their findings.

Conclusions

The evolution of quality in the nursing home industry is truly remarkable. However, sustained progress is essential to meet the needs of the growing older adult population. Although widely used in the nursing home industry, the 5-Star Quality Rating System for nursing homes is an unreliable performance measure. As demonstrated through the results of the study, it is hard to delineate nursing homes' performance based on the star ratings, after accounting for the underlying QMs. Concerted efforts from lawmakers, policy makers, and lobbyists will be required to refine and enhance the measure, thereby ensuring its reliability and effectiveness.

Funding

The authors did not receive support from any funding agency in the public, commercial, or not-for-profit sectors for the submitted work.

Data Availability

The datasets used in this study are available from the author's figshare repositories and are listed in the references.

Authors' Contributions

Conceptualization: HA
Data curation: SB
Formal analysis: SB
Writing – original draft: SB (lead), HA (supporting)
Writing – review & editing: SB (lead), HA (supporting)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Complete set of results tables for all models estimated in the study. Tables are organized by outcome variable and by model specification, distinguishing between base models using the original variables and final models incorporating principal component–derived variables, respectively.
[DOCX File, 102 KB - humanfactors_v13i1e72770_app1.docx]

Multimedia Appendix 2
Complete set of results displayed graphically. Bars denote the point estimates, while whiskers indicate the corresponding 95% CIs. Each figure is divided into 2 panels, representing the conditional model and the zero-inflation model, respectively.
[DOCX File, 1452 KB - humanfactors_v13i1e72770_app2.docx]

XSL•FO
RenderX

References

1. Nursing home care. National Center for Health Statistics. 2020. URL: https://www.cdc.gov/nchs/fastats/nursing-home-care.htm [accessed 2025-09-11]
2. Davidson PM, Szanton SL. Nursing homes and COVID-19: we can and should do better. J Clin Nurs 2020 Aug;29(15-16):2758-2759. [doi: 10.1111/jocn.15297] [Medline: 32281165]
3. McGarry BE, Gandhi AD, Barnett ML. COVID-19 surveillance testing and resident outcomes in nursing homes. N Engl J Med 2023 Mar 23;388(12):1101-1110. [doi: 10.1056/NEJMoa2210063] [Medline: 36947467]
4. Grabowski DC, Chen A, Saliba D. Paying for nursing home quality: an elusive but important goal. J American Geriatrics Society 2023 Feb;71(2):342-348. [doi: 10.1111/jgs.18260]
5. Luo H, Lum TYS, Wong GHY, Kwan JSK, Tang JYM, Chi I. Predicting adverse health outcomes in nursing homes: a 9-year longitudinal study and development of the FRAIL-Minimum Data Set (MDS) quick screening tool. J Am Med Dir Assoc 2015 Dec;16(12):1042-1047. [doi: 10.1016/j.jamda.2015.09.006] [Medline: 26602760]
6. Moore KL, Boscardin WJ, Steinman MA, Schwartz JB. Patterns of chronic co-morbid medical conditions in older residents of U.S. nursing homes: differences between the sexes and across the agespan. J Nutr Health Aging 2014 Apr;18(4):429-436. [doi: 10.1007/s12603-014-0001-y] [Medline: 24676326]
7. Reilev M, Lundby C, Jensen J, Larsen SP, Hoffmann H, Pottegård A. Morbidity and mortality among older people admitted to nursing home. Age Ageing 2019 Dec 1;49(1):67-73. [doi: 10.1093/ageing/afz136] [Medline: 31763682]
8. Vespa J, Armstrong DM, Medina L. Demographic turning points for the United States: population projections for 2020 to 2060. : United States Census Bureau; 2018 URL: https://www.census.gov/content/dam/Census/library/publications/2020/demo/p25-1144.pdf [accessed 2026-01-06]
9. Consumer Financial Protection Bureau Office for Older Americans. Issue spotlight: nursing home debt collection. : Consumer Financial Protection Bureau; 2022 URL: https://www.consumerfinance.gov/data-research/research-reports/issue-spotlight-nursing-home-debt-collection/full-report/ [accessed 2026-01-06]
10. Medina L, Sabo S, Vespa J. Living longer: historical and projected life expectancy in the United States, 1960 to 2060. : United States Census Bureau; 2020 URL: https://www.census.gov/content/dam/Census/library/publications/2020/demo/p25-1145.pdf [accessed 2026-01-06]
11. Mitzner TL, Beer JM, McBride SE, Rogers WA, Fisk AD. Older adults' needs for home health care and the potential for human factors interventions. Proc Hum Factors Ergon Soc Annu Meet 2009 Oct;53(1):718-722. [doi: 10.1177/154193120905301118] [Medline: 25530696]
12. Ruhm CJ. Excess deaths in the United States during the first year of COVID-19. Prev Med 2022 Sep;162:107174. [doi: 10.1016/j.ypmed.2022.107174] [Medline: 35878708]
13. Adams ML, Katz DL, Grandpre J. Population-based estimates of chronic conditions affecting risk for complications from coronavirus disease, United States. Emerg Infect Dis 2020 Aug;26(8):1831-1833. [doi: 10.3201/eid2608.200679] [Medline: 32324118]
14. Williams CS, Zheng Q, White AJ, et al. The association of nursing home quality ratings and spread of COVID-19. J Am Geriatr Soc 2021 Aug;69(8):2070-2078. [doi: 10.1111/jgs.17309] [Medline: 34058015]
15. COVID-19: who's at higher risk of serious symptoms. Mayo Clinic. 2024. URL: https://www.mayoclinic.org/diseases-conditions/coronavirus/in-depth/coronavirus-who-is-at-risk/art-20483301 [accessed 2025-09-11]
16. De Witte J, Noten S, Vinckers F, et al. "You needed to accept the situation": resilience of nursing home residents in times of COVID-19. Can J Aging 2024 Mar;43(1):75-83. [doi: 10.1017/S0714980823000399] [Medline: 37665004]
17. Rubano MD, Kieffer EF, Larson EL. Infection prevention and control in nursing homes during COVID-19: an environmental scan. Geriatr Nurs 2022;43:51-57. [doi: 10.1016/j.gerinurse.2021.10.023] [Medline: 34808421]
18. Dorritie R, Quigley DD, Agarwal M, Tark A, Dick A, Stone PW. Support of nursing homes in infection management varies by US State Departments of Health. J Hosp Infect 2020 Jun;105(2):258-264. [doi: 10.1016/j.jhin.2020.02.007] [Medline: 32068013]
19. Ferdous F. Social distancing vs social interaction for older adults at long-term care facilities in the midst of the COVID-19 pandemic: a rapid review and synthesis of action plans. Inquiry 2021;58:469580211044287. [doi: 10.1177/00469580211044287] [Medline: 34612063]
20. How to protect yourself and others. Centers for Disease Control and Prevention. 2025. URL: https://www.cdc.gov/covid/prevention/index.html [accessed 2025-09-11]
21. Infection control deficiencies were widespread and persistent in nursing homes prior to COVID-19 pandemic. United States Government Accountability Office. 2020. URL: https://www.gao.gov/products/gao-20-576r [accessed 2025-09-11]
22. Nursing home quality initiative. : Centers for Medicare & Medicaid Services; 2003 URL: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/Downloads/NHQIOverView20030731.pdf [accessed 2025-09-11]
23. CMS' five star quality rating system for nursing homes. Centers for Medicare & Medicaid Services. 2014. URL: https://www.cms.gov/newsroom/fact-sheets/cms-five-star-quality-rating-system-nursing-homes [accessed 2025-09-11]
24. Castle NG, Ferguson JC. What is nursing home quality and how is it measured? Gerontologist 2010 Aug;50(4):426-442. [doi: 10.1093/geront/gnq052] [Medline: 20631035]

XSL•FO
RenderX

25. Castle NG, Engberg J. Staff turnover and quality of care in nursing homes. Med Care 2005 Jun;43(6):616-626. [doi: 10.1097/01.mlr.0000163661.67170.b9] [Medline: 15908857]
26. Lyder CH. Pressure ulcer prevention and management. JAMA 2003 Jan 8;289(2):223-226. [doi: 10.1001/jama.289.2.223] [Medline: 12517234]
27. Clemens S, Wodchis W, McGilton K, McGrail K, McMahon M. The relationship between quality and staffing in long-term care: a systematic review of the literature 2008-2020. Int J Nurs Stud 2021 Oct;122:104036. [doi: 10.1016/j.ijnurstu.2021.104036] [Medline: 34419730]
28. Collier E, Harrington C. Staffing characteristics, turnover rates, and quality of resident care in nursing facilities. Res Gerontol Nurs 2008 Jul;1(3):157-170. [doi: 10.3928/19404921-20080701-02] [Medline: 20077960]
29. Spilsbury K, Hewitt C, Stirk L, Bowman C. The relationship between nurse staffing and quality of care in nursing homes: a systematic review. Int J Nurs Stud 2011 Jun;48(6):732-750. [doi: 10.1016/j.ijnurstu.2011.02.014] [Medline: 21397229]
30. Hillmer MP, Wodchis WP, Gill SS, Anderson GM, Rochon PA. Nursing home profit status and quality of care: is there any evidence of an association? Med Care Res Rev 2005 Apr;62(2):139-166. [doi: 10.1177/1077558704273769] [Medline: 15750174]
31. R Foundation for Statistical Computing. 2025. URL: https://www.R-project.org/ [accessed 2026-01-06]
32. McClelland GH. Increasing statistical power without increasing sample size. Am Psychol 2000;55(8):963-964. [doi: 10.1037//0003-066X.55.8.963]
33. Donabedian A. The quality of care: how can it be assessed? JAMA 1988;260(12):1743-1748. [doi: 10.1001/jama.1988.03410120089033]
34. Braun RT, Yun H, Casalino LP, et al. Comparative performance of private equity-owned US nursing homes during the COVID-19 pandemic. JAMA Netw Open 2020 Oct 1;3(10):e2026702. [doi: 10.1001/jamanetworkopen.2020.26702] [Medline: 33112402]
35. Newman DA. Missing data. Organ Res Methods 2014 Oct;17(4):372-411. [doi: 10.1177/1094428114548590]
36. Mayer M. MissRanger: fast imputation of missing values R package version 2.6.0. Comprehensive R Archive Network (CRAN). 2024. URL: https://CRAN.R-project.org/package=missRanger [accessed 2026-01-06]
37. Penny KI, Atkinson I. Approaches for dealing with missing data in health care studies. J Clin Nurs 2012 Oct;21(19-20):2722-2729. [doi: 10.1111/j.1365-2702.2011.03854.x] [Medline: 21895816]
38. Feng CX. A comparison of zero-inflated and hurdle models for modeling zero-inflated count data. J Stat Distrib Appl 2021;8(1):8. [doi: 10.1186/s40488-021-00121-4] [Medline: 34760432]
39. Giri S, Chenn LM, Romero-Ortuno R. Nursing homes during the COVID-19 pandemic: a scoping review of challenges and responses. Eur Geriatr Med 2021 Dec;12(6):1127-1136. [doi: 10.1007/s41999-021-00531-2] [Medline: 34136990]
40. Interim guidelines for COVID-19 antibody testing. Centers for Disease Control and Prevention. URL: https://archive.cdc.gov/#/details?q=https://www.cdc.gov/coronavirus/2019-ncov/hcp/testing/antibody-tests-guidelines.html&start=0&rows=10&url=https://www.cdc.gov/coronavirus/2019-ncov/hcp/testing/antibody-tests-guidelines.html [accessed 2025-09-11]
41. Brooks M, Kristensen K, Benthem K, et al. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R J 2017;9(2):378 [FREE Full text] [doi: 10.32614/RJ-2017-066]
42. Hartig F. DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models (version 046) [R package]. Comprehensive R Archive Network (CRAN). URL: https://CRAN.R-project.org/package=DHARMa [accessed 2025-09-11]
43. Van Laethem J, Wuyts SCM, Pierreux J, et al. Presumed urinary tract infection in patients admitted with COVID-19: are we treating too much? Antibiotics (Basel) 2021 Dec 6;10(12):1493. [doi: 10.3390/antibiotics10121493] [Medline: 34943705]
44. Lim CJ, McLellan SC, Cheng AC, et al. Surveillance of infection burden in residential aged care facilities. Med J Aust 2012 Mar 19;196(5):327-331 [FREE Full text] [doi: 10.5694/mja12.10085] [Medline: 22432671]
45. Loeb M, Simor AE, Landry L, et al. Antibiotic use in Ontario facilities that provide chronic care. J Gen Intern Med 2001 Jun;16(6):376-383. [doi: 10.1046/j.1525-1497.2001.016006376.x] [Medline: 11422634]
46. Montgomery P, Semenchuk M, Nicolle LE. Antimicrobial use in nursing homes in Manitoba. J Geriatr Drug Ther 1995;9(3):55-74. [doi: 10.1300/J089v09n03_05]
47. Peron EP, Hirsch AA, Jury LA, Jump RLP, Donskey CJ. Another setting for stewardship: high rate of unnecessary antimicrobial use in a veterans affairs long-term care facility. J Am Geriatr Soc 2013 Feb;61(2):289-290. [doi: 10.1111/jgs.12099] [Medline: 23405923]
48. Stuart RL, Wilson J, Bellaard-Smith E, et al. Antibiotic use and misuse in residential aged care facilities. Intern Med J 2012 Oct;42(10):1145-1149. [doi: 10.1111/j.1445-5994.2012.02796.x] [Medline: 22472087]
49. Zimmer JG, Bentley DW, Valenti WM, Watson NM. Systemic antibiotic use in nursing homes. A quality assessment. J Am Geriatr Soc 1986 Oct;34(10):703-710. [doi: 10.1111/j.1532-5415.1986.tb04301.x] [Medline: 3760435]
50. Brehm TT, van der Meirschen M, Hennigs A, et al. Comparison of clinical characteristics and disease outcome of COVID-19 and seasonal influenza. Sci Rep 2021 Mar 11;11(1):5803. [doi: 10.1038/s41598-021-85081-0] [Medline: 33707550]
51. Soriano JB, Kendrick PJ, Paulson KR. Prevalence and attributable health burden of chronic respiratory diseases, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Respir Med 2020 Jun;8(6):585-596. [doi: 10.1016/S2213-2600(20)30105-3] [Medline: 32526187]

XSL•FO
RenderX

52. Joshi PR. Pulmonary diseases in older patients: understanding and addressing the challenges. Geriatrics (Basel) 2024 Mar 7;9(2):34. [doi: 10.3390/geriatrics9020034] [Medline: 38525751]
53. Kim V, Criner GJ. Chronic bronchitis and chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2013 Feb 1;187(3):228-237. [doi: 10.1164/rccm.201210-1843CI] [Medline: 23204254]
54. Oelsner EC, Loehr LR, Henderson AG, et al. Classifying chronic lower respiratory disease events in epidemiologic cohort studies. Ann Am Thorac Soc 2016 Jul;13(7):1057-1066. [doi: 10.1513/AnnalsATS.201601-063OC] [Medline: 27088163]
55. Pesce G. Mortality rates for chronic lower respiratory diseases in Italy from 1979 to 2010: an age-period-cohort analysis. ERJ Open Res 2016 Jan;2(1):00093-2015. [doi: 10.1183/23120541.00093-2015] [Medline: 27730182]
56. Czwikla J, Wandscher K, Helbach J, Fassmer AM, Schmiemann G, Hoffmann F. Prevalence of indwelling urinary catheters in nursing home residents: systematic review. Int J Nurs Stud 2023 Sep;145:104555. [doi: 10.1016/j.ijnurstu.2023.104555] [Medline: 37421830]
57. Czwikla J, Schmiemann G, Hoffmann F. Use of indwelling urinary catheters in nursing home residents: results from a cross-sectional study in 21 German nursing homes. BMC Urol 2024 Jun 14;24(1):125. [doi: 10.1186/s12894-024-01512-w] [Medline: 38877475]
58. Nosocomial infections (healthcare-associated infections). Cleveland Clinic. URL: https://my.clevelandclinic.org/health/diseases/16397-avoiding-healthcare-associated-infections-hais [accessed 2025-09-11]
59. Harrington C, Ross L, Chapman S, Halifax E, Spurlock B, Bakerjian D. Nurse staffing and coronavirus infections in California nursing homes. Policy Polit Nurs Pract 2020 Aug;21(3):174-186. [doi: 10.1177/1527154420938707] [Medline: 32635838]
60. Abrams HR, Loomer L, Gandhi A, Grabowski DC. Characteristics of U.S. nursing homes with COVID-19 cases. J Am Geriatr Soc 2020 Aug;68(8):1653-1656. [doi: 10.1111/jgs.16661] [Medline: 32484912]
61. Chatterjee P, Kelly S, Qi M, Werner RM. Characteristics and quality of US nursing homes reporting cases of coronavirus disease 2019 (COVID-19). JAMA Netw Open 2020 Jul 1;3(7):e2016930. [doi: 10.1001/jamanetworkopen.2020.16930] [Medline: 32725243]
62. Gorges RJ, Konetzka RT. Staffing levels and COVID-19 cases and outbreaks in U.S. nursing homes. J Am Geriatr Soc 2020 Nov;68(11):2462-2466. [doi: 10.1111/jgs.16787] [Medline: 32770832]
63. Sun CLF, Zuccarelli E, Zerhouni EGA, et al. Predicting coronavirus disease 2019 infection risk and related risk drivers in nursing homes: a machine learning approach. J Am Med Dir Assoc 2020 Nov;21(11):1533-1538. [doi: 10.1016/j.jamda.2020.08.030] [Medline: 33032935]
64. White EM, Kosar CM, Feifer RA, et al. Variation in SARS-CoV-2 prevalence in U.S. skilled nursing facilities. J Am Geriatr Soc 2020 Oct;68(10):2167-2173. [doi: 10.1111/jgs.16752] [Medline: 32674223]
65. Zhu X, Lee H, Sang H, et al. Nursing home design and COVID-19: implications for guidelines and regulation. J Am Med Dir Assoc 2022 Feb;23(2):272-279. [doi: 10.1016/j.jamda.2021.12.026] [Medline: 34990585]
66. Tarteret P, Strazzulla A, Rouyer M, et al. Clinical features and medical care factors associated with mortality in French nursing homes during the COVID-19 outbreak. Int J Infect Dis 2021 Mar;104:125-131. [doi: 10.1016/j.ijid.2020.12.004] [Medline: 33301993]
67. Gorges RJ, Konetzka RT. Factors associated with racial differences in deaths among nursing home residents with COVID-19 infection in the US. JAMA Netw Open 2021 Feb 1;4(2):e2037431. [doi: 10.1001/jamanetworkopen.2020.37431] [Medline: 33566110]
68. Gopal R, Han X, Yaraghi N. Compress the curve: a cross-sectional study of variations in COVID-19 infections across California nursing homes. BMJ Open 2021 Jan 6;11(1):e042804. [doi: 10.1136/bmjopen-2020-042804] [Medline: 33408208]
69. Li Y, Temkin-Greener H, Shan G, Cai X. COVID-19 infections and deaths among Connecticut nursing home residents: facility correlates. J Am Geriatr Soc 2020 Sep;68(9):1899-1906. [doi: 10.1111/jgs.16689] [Medline: 32557542]
70. Sugg MM, Spaulding TJ, Lane SJ, et al. Mapping community-level determinants of COVID-19 transmission in nursing homes: a multi-scale approach. Sci Total Environ 2021 Jan 15;752:141946. [doi: 10.1016/j.scitotenv.2020.141946] [Medline: 32889290]
71. Bui DP, See I, Hesse EM, et al. Association between CMS quality ratings and COVID-19 outbreaks in nursing homes - West Virginia, March 17-June 11, 2020. MMWR Morb Mortal Wkly Rep 2020 Sep 18;69(37):1300-1304. [doi: 10.15585/mmwr.mm6937a5] [Medline: 32941409]
72. Cai S, Yan D, Intrator O. COVID-19 cases and death in nursing homes: the role of racial and ethnic composition of facilities and their communities. J Am Med Dir Assoc 2021 Jul;22(7):1345-1351. [doi: 10.1016/j.jamda.2021.05.002] [Medline: 34062147]
73. Konetzka RT. Testimony of R. Tamara Konetzka, PhD. in United States Senate Special Committee on Aging. : United States Senate Special Committee on Aging; 2020 URL: https://www.aging.senate.gov/imo/media/doc/SCA_Konetzka_05_21_20.pdf [accessed 2025-09-11]
74. Chen MK, Chevalier JA, Long EF. Nursing home staff networks and COVID-19. Proc Natl Acad Sci U S A 2021 Jan 7;118(1):e2015455118. [doi: 10.1073/pnas.2015455118] [Medline: 33323526]
75. Cronin CJ, Evans WN. Nursing home quality, COVID-19 deaths, and excess mortality. J Health Econ 2022 Mar;82:102592. [doi: 10.1016/j.jhealeco.2022.102592] [Medline: 35104669]

XSL•FO
RenderX

76. Dean A, Venkataramani A, Kimmel S. Mortality rates from COVID-19 are lower in unionized nursing homes. Health Aff (Millwood) 2020 Nov;39(11):1993-2001. [doi: 10.1377/hlthaff.2020.01011] [Medline: 32910688]
77. Figueroa JF, Wadhera RK, Papanicolas I, et al. Association of nursing home ratings on health inspections, quality of care, and nurse staffing with COVID-19 cases. JAMA 2020 Sep 15;324(11):1103-1105. [doi: 10.1001/jama.2020.14709] [Medline: 32790822]
78. Konetzka RT, White EM, Pralea A, Grabowski DC, Mor V. A systematic review of long-term care facility characteristics associated with COVID-19 outcomes. J Am Geriatr Soc 2021 Oct;69(10):2766-2777. [doi: 10.1111/jgs.17434] [Medline: 34549415]
79. Shen K. Relationship between nursing home COVID-19 outbreaks and staff neighborhood characteristics. PLOS ONE 2022;17(4):e0267377. [doi: 10.1371/journal.pone.0267377] [Medline: 35439279]

Abbreviations

CMS: Centers for Medicare and Medicaid Services
MAR: missing at random
MDS: Minimum Data Set
NHC: nursing home care
PCA: principal component analysis
QM: quality measure
QoC: quality of care
SPO: structure, process, or outcome
UTI: urinary tract infection
ZI: zero inflation
ZINB: zero-inflated negative binomial

Edited by A Kushniruk, E Borycki; submitted 17.Feb.2025; peer-reviewed by F Irshaidat, SJ Lane; revised version received 16.Sep.2025; accepted 23.Sep.2025; published 29.Jan.2026.

Please cite as:
Bharadwaj S, Ali H
Demystifying Quality Metrics and Unveiling the True Measure of Quality of Care in Nursing Homes: Mixed Effects Analysis
JMIR Hum Factors 2026;13:e72770
URL: https://humanfactors.jmir.org/2026/1/e72770
doi:10.2196/72770

© Suhas Bharadwaj, Haneen Ali. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 29.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Radiologist Perceptions of an AI Tool for Intracranial Hemorrhage Detection in Teleradiology: Cross-Sectional Survey Study

Andrew J Del Gaizo[1], MD; Jackson K Del Gaizo[2]; Troy A Shahoumian[3], PhD

[1]Moffitt Cancer Center, Tampa, FL, United States

[2]Durham Academy, Durham, NC, United States

[3]Population Health Informatics, Digital Health Office, Veterans Health Administration, Washington DC, DC, United States

Corresponding Author:
Andrew J Del Gaizo, MD
Moffitt Cancer Center
12902 USF Magnolia Drive,
Tampa, FL, 33612
United States
Phone: 1 813 745 4673
Email: andrew.delgaizo@moffitt.org

Abstract

Background: Artificial intelligence (AI) detection tools for intracranial hemorrhage (ICH) are increasingly integrated into radiology workflows. In real-world practice, perceived utility depends not only on diagnostic performance but also on workflow fit, false positive burden, and how clinicians interpret and act on AI outputs.

Objective: This study aimed to characterize radiologists' perceptions of a Food and Drug Administration (FDA)–cleared ICH AI detection tool in a national teleradiology network, including perceived reliability, false positive burden, workflow impact, medicolegal concerns, and self-reported behaviors during routine use.

Methods: We conducted an anonymous cross-sectional survey of radiologists in a national teleradiology practice who had access to an FDA-cleared ICH AI overlay during noncontrast head computed tomography interpretation. Survey items used a 5-point Likert scale. Results are summarized as agreement proportions ("agree" or "strongly agree") with 95% CIs. We compared neuroradiologists with non-neuroradiologists using Fisher exact tests. One primary end point was prespecified: agreement that time spent reviewing examinations with false positive AI alerts outweighed the benefits. Remaining subgroup comparisons were treated as exploratory, with false discovery rate control using the Benjamini-Hochberg procedure.

Results: A total of 65 radiologists responded, including 23 (35.4%) neuroradiologists and 42 (64.6%) non-neuroradiologists. Only 18.5% (12/65; 95% CI 10.9%-29.6%) agreed that false-positive alerts were infrequent enough to be acceptable. Agreement that the AI correctly identified most ICH cases was 32.3% (21/65; 95% CI 22.2%-44.4%), and agreement that the AI rarely missed clinically important hemorrhages was 43.1% (28/65; 95% CI 31.8%-55.2%). Trust in AI output was conditional: 50.8% (33/65; 95% CI 38.9%-62.5%) reported trusting the AI when it agreed with their interpretation, whereas 3.1% (2/65; 95% CI 0.8%-10.5%) reported trusting it when it conflicted with their interpretation. Only 10.8% (7/65; 95% CI 5.3%-20.6%) reported reduced overall interpretation time, whereas 33.8% (22/65; 95% CI 23.5%-46.0%) agreed that time spent reviewing false-positive alerts outweighed the benefits. Self-reported reduced scrutiny after an AI-negative result was uncommon (4/65, 6.2%; 95% CI 2.4%-14.8%). In subgroup analysis, neuroradiologists more often endorsed the primary end point than non-neuroradiologists (12/23, 52.2% vs 10/42, 23.8%; unadjusted P=.03), but no exploratory subgroup differences remained statistically significant after false discovery rate correction. Free-text responses emphasized artifact- and calcification-driven false positives, delayed or inconsistent AI availability, consultation burden, and medicolegal concerns.

Conclusions: In this national teleradiology setting, radiologists reported substantial false positive burden, limited perceived time savings, and strongly conditional trust in an FDA-cleared ICH AI detection tool. Self-reported reduced scrutiny after negative AI outputs was uncommon but present in a minority of cases. These findings support the importance of specificity, interpretability, latency, and workflow-aware implementation when deploying radiology AI tools in practice.

(JMIR Hum Factors 2026;13:e92145) doi:10.2196/92145

XSL•FO
RenderX

KEYWORDS

artificial intelligence; AI; intracranial hemorrhage; computed tomography; teleradiology; workflow; false positives; trust in automation; alert fatigue; human factors; decision support

Introduction

Intracranial hemorrhage (ICH) on noncontrast head computed tomography (CT) is a high-stakes diagnosis in emergency and inpatient care requiring timely identification and communication. Artificial intelligence (AI) systems intended to detect ICH are increasingly incorporated into radiology workflows, including as diagnostic assistive tools during radiologist interpretation and, in some implementations, as tools to support prioritization of urgent cases [1-4]. However, the real-world value of these systems depends not only on discriminative performance but also on how they interact with clinical workflow, how often they generate disruptive false-positive alerts, and how clinicians interpret their outputs in routine practice [2-5].

Post-deployment evaluation is increasingly recognized as essential for clinical AI. DECIDE-AI (Developmental and Exploratory Clinical Investigations of Decision Support Systems Driven by Artificial Intelligence) and related implementation literature emphasize that early live clinical evaluation should address workflow integration, human factors, and unintended burdens rather than focusing on accuracy alone [5]. Real-world studies of radiology AI have similarly shown that local prevalence, operational context, and implementation strategy can substantially shape perceived utility and clinician acceptance [2,3,6]. For example, deployment experience with an AI algorithm in clinical practice has highlighted how false discovery burden can create an “accuracy paradox” in which an apparently high-performing system is experienced as inaccurate because positive alerts are frequently false in a low-prevalence setting [6].

Human factors theory suggests that false alarms, workflow interruptions, and opaque outputs can alter clinician reliance on automation [7-10]. In health care more broadly, high alert burden and repeated alerts can contribute to alert fatigue and reduced responsiveness to clinical decision support alerts [8]. In radiology, incorrect AI outputs may affect reader confidence and performance, particularly when the AI appears authoritative or visually salient [9]. At the same time, survey data alone cannot establish whether AI changes diagnostic behavior; they are better suited to characterizing clinician perceptions, attitudes, and self-reported experiences.

We previously evaluated a Food and Drug Administration (FDA)–cleared ICH AI tool in a national teleradiology network and reported operational observations, including false positives and radiologist-reported effects on interpretation time [1]. The study reported in this paper extends that work by focusing on radiologists’ perceptions during routine exposure to the tool. Specifically, we sought to characterize perceived reliability, false positive burden, workflow impact, medicolegal concerns, and self-reported behaviors during AI-assisted interpretation and explore whether these perceptions differed by neuroradiology subspecialty training.

Methods

Study Design

This was an anonymous cross-sectional survey study conducted within a national teleradiology practice. This study was designed as a real-world implementation evaluation aligned with early-stage clinical AI assessment principles, with emphasis on workflow context and user experience [5].

Ethical Considerations

This evaluation met the US Department of Veterans Affairs criteria for quality assessment and nonresearch and received a determination of nonresearch from the Stanford University Institutional Review Board (protocol 73642). Participation was voluntary and anonymous.

Participants

Eligible participants were practicing radiologists within the teleradiology group who had access to the ICH AI output during routine noncontrast head CT interpretation. Respondents self-reported whether they had neuroradiology fellowship training or primarily practiced neuroradiology, and responses were stratified as neuroradiologists vs non-neuroradiologists.

AI Tool and Workflow Context

The evaluated system was an FDA-cleared and commercially available deep learning algorithm for ICH detection (CINA version 1.0; Avicenna.ai) integrated into the reading workflow for noncontrast head CT. The tool was intended to assist radiologists in identifying suspected ICH during image interpretation and, depending on workflow configuration, could also support prioritization. In practice, radiologists could view AI-generated overlays during interpretation, and the system functioned as an adjunctive diagnostic support tool within routine workflow. The tool did not replace radiologist interpretation, and final clinical interpretation remained the responsibility of the reading radiologist. Because intended use and actual workflow exposure are both relevant to interpretation of user perceptions, we frame these findings as perceptions arising during routine use of the tool in this implementation context.

Survey Instrument

The survey included 15 Likert items spanning the following domains: perceived reliability, false positive burden, workflow integration, time impact, medicolegal concerns, adoption preferences, and self-reported behaviors during AI use. Responses were captured on a 5-point scale (“strongly agree,” “agree,” “neutral,” “disagree,” and “strongly disagree”). Items were developed for face validity based on prior operational experience and relevant human factors concepts, but the instrument was not a formally validated psychometric measure. Accordingly, results were interpreted on the face of each item as descriptive measures of respondent perceptions.

XSL•FO
RenderX

Prespecified End Point and Statistical Analysis

To limit overinterpretation from multiple comparisons, one primary end point was prespecified: agreement with the following statement: "The extra time spent reviewing AI false-positive studies outweighs the benefits." This end point was selected because false positive burden was considered a central determinant of usability and adoption in a high-volume teleradiology environment.

All other item-level subgroup comparisons were considered exploratory. Responses were summarized as agreement proportions, defined as the percentage selecting "agree" or "strongly agree." We report proportions with 2-sided 95% CIs using Wilson intervals. For subgroup comparisons, we used Fisher exact tests on dichotomized responses ("agree" or "strongly agree" vs all other responses). Because responses were condensed from a 5-point scale to a binary end point for interpretability and consistency with implementation-focused survey reporting, some resolution was necessarily lost. Therefore, we interpret subgroup analyses as descriptive and hypothesis generating. False discovery rate control for exploratory subgroup comparisons was applied using the Benjamini-Hochberg procedure. Analyses were performed in Python (version 3.13.5; Python Software Foundation) using exact tests and Wilson interval estimation with *SciPy* (version 1.17.0) and *statsmodels* (version 0.14.6).

Qualitative Analysis

Optional free-text comments were reviewed qualitatively to identify recurring implementation themes. Comments could map to more than one theme. Representative excerpts were selected to illustrate the most common concerns raised by respondents.

Results

Respondent Characteristics

A total of 65 radiologists completed the survey, including 23 (35.4%) neuroradiologists and 42 (64.6%) non-neuroradiologists. Full quantitative results across all survey items are provided in Table 1.

XSL•FO
RenderX

Table 1. Agreement proportions across survey items ("agree" or "strongly agree") overall and stratified by neuroradiology training. The prespecified primary end point is item 7 (N=65)[a].

Item number	Question	Total, n (%; 95% CI)	Neuroradiologists (n=23), n (%)	Non-neuroradiologists (n=42), n (%)	*P* value	*q* value
1	The AI[b] tool correctly identifies most cases of intracranial hemorrhage.	21 (32.3; 22.2-44.4)	8 (34.8)	13 (31.0)	.79	.92
2	False-positive alerts from the AI are infrequent enough to be acceptable.	12 (18.5; 10.9-29.6)	4 (17.4)	8 (19.0)	>.99	>.99
3	The AI rarely misses clinically important hemorrhages.	28 (43.1; 31.8-55.2)	10 (43.5)	18 (42.9)	>.99	>.99
4	I trust the AI output when it agrees with my own interpretation.	33 (50.8; 38.9-62.5)	11 (47.8)	22 (52.4)	.80	.92
5	I trust the AI output when it conflicts with my initial impression.	2 (3.1; 0.8-10.5)	0 (0)	2 (4.8)	.54	.90
6	Using the AI has reduced my overall interpretation time per head CT[c].	7 (10.8; 5.3-20.6)	1 (4.3)	6 (14.3)	.41	.90
7	The extra time spent reviewing AI false-positive studies outweighs the benefits.	22 (33.8; 23.5-46.0)	12 (52.2)	10 (23.8)	.03	—[d]
8	If the AI labels a study as negative, I tend to scrutinize it less thoroughly than usual.	4 (6.2; 2.4-14.8)	0 (0)	4 (9.5)	.29	.90
9	The AI reduces the risk that I will overlook a subtle hemorrhage on a busy shift.	23 (35.4; 24.9-47.5)	7 (30.4)	16 (38.1)	.60	.90
10	Frequent false positive alerts make me second-guess my decisions.	20 (30.8; 20.9-42.8)	4 (17.4)	16 (38.1)	.10	.75
11	Overall, the AI has improved the quality of care our patients receive.	13 (20; 12.1-31.3)	4 (17.4)	9 (21.4)	.76	.92
12	I feel legally vulnerable if I override an AI recommendation.	21 (32.3; 22.2-44.4)	6 (26.1)	15 (35.7)	.58	.90
13	The AI integrates smoothly into our PACS[e]/reporting workflow.	32 (49.2; 37.5-61.1)	13 (56.5)	19 (45.2)	.44	.90
14	I would recommend continued use of this AI tool in our practice.	19 (29.2; 19.6-41.2)	5 (21.7)	14 (33.3)	.40	.90
15	I would like additional feedback or training to use the AI tool optimally.	13 (20; 12.1-31.3)	3 (13.0)	10 (23.8)	.35	.90

[a]Between-group comparisons were conducted using Fisher exact tests. False discovery rate control across exploratory subgroup comparisons was applied using the Benjamini-Hochberg procedure; *q* values are reported for exploratory items.

[b]AI: artificial intelligence.

[c]CT: computed tomography.

[d]Not applicable. Item 7 was the prespecified primary end point; *q* values were not applied to this item.

[e]PACS: picture archiving and communication system.

Perceived Reliability and False Positive Burden

Agreement that the AI correctly identified most cases of ICH was 32.3% (21/65; 95% CI 22.2%-44.4%). Agreement that the AI rarely missed clinically important hemorrhages was somewhat higher at 43.1% (28/65; 95% CI 31.8%-55.2%). In contrast, only 18.5% (12/65; 95% CI 10.9%-29.6%) agreed that false-positive alerts were infrequent enough to be acceptable, indicating that perceived false positive burden was a prominent limitation.

The prespecified primary end point was also consistent with this pattern. Overall, 33.8% (22/65) of respondents (95% CI 23.5%-46.0%) agreed that the extra time spent reviewing examinations with false positive AI alerts outweighed the benefits of the system. These radiologists endorsed the view that false positive workload offset the tool's net value in routine use. Overall agreement across survey items is shown in Table 1.

Trust in AI Output and Perceived Workflow Impact

Responses suggested strongly conditional trust in the AI output. Half (33/65, 50.8%) of respondents (95% CI 38.9%-62.5%) agreed that they trusted the AI when it agreed with their own interpretation. In contrast, only 3.1% (2/65; 95% CI 0.8%-10.5%) agreed that they trusted the AI when it conflicted with their initial impression. This asymmetry suggests that

XSL•FO
RenderX

respondents were much more likely to experience the AI as confirmatory than as persuasive in discrepant cases.

Perceived workflow benefit was limited. Only 10.8% (7/65) of respondents (95% CI 5.3%-20.6%) agreed that AI use reduced their overall interpretation time per head CT. Similarly, 35.4% (23/65; 95% CI 24.9%-47.5%) agreed that the AI reduced the risk that they would overlook a subtle hemorrhage on a busy shift, indicating that perceived detection support was present for some respondents but did not translate into broad perception of time savings.

Self-Reported Behaviors, Quality Perceptions, and Medicolegal Concerns

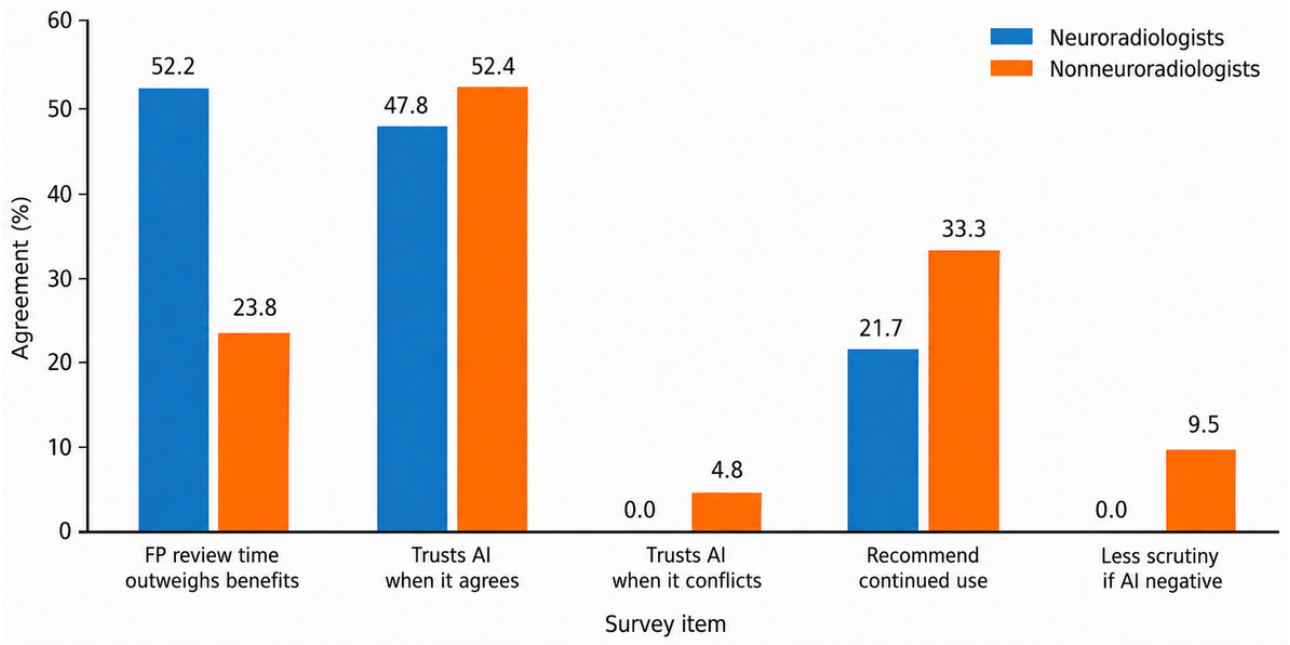
Self-reported reduced scrutiny after a negative AI result was uncommon. Only 6.2% (4/65) of the respondents (95% CI 2.4%-14.8%) agreed with the statement that, if the AI labeled a result as negative, they tended to scrutinize it less thoroughly than usual. Because this was a self-reported perception and not a direct behavioral measurement, we interpreted it narrowly as a descriptive response to the survey item.

Frequent false-positive alerts were also associated with self-reported second-guessing. A total of 30.8% (20/65) of the respondents (95% CI 20.9%-42.8%) agreed that frequent false-positive alerts made them second-guess their decisions. Only 20% (13/65; 95% CI 12.1%-31.3%) agreed that the AI improved the quality of care that patients received. Agreement that the AI integrated smoothly into the picture archiving and communication system or reporting workflow was 49.2% (32/65; 95% CI 37.5%-61.1%), suggesting mixed perceptions of integration. Recommendation for continued use of the tool was endorsed by 29.2% (19/65; 95% CI 19.6%-41.2%). Concern about legal vulnerability when overriding AI output was reported by 32.3% (21/65; 95% CI 22.2%-44.4%). Additional feedback or training to use the tool optimally was desired by 20% (13/65; 95% CI 12.1%-31.3%).

Subgroup Findings by Neuroradiology Training

Neuroradiologists more frequently endorsed the primary end point than non-neuroradiologists (12/23, 52.2% vs 10/42, 23.8%; unadjusted *P*=.03). For exploratory items, patterns were broadly similar between groups. Agreement with trusting the AI when it agreed with the reader's interpretation was 47.8% (11/23) among neuroradiologists and 52.4% (22/42) among non-neuroradiologists. Trust in conflicting AI output remained rare in both groups (0% vs 2/42, 4.8%). Self-reported reduced scrutiny after an AI-negative result was reported by 0% of neuroradiologists and 9.5% (4/42) of non-neuroradiologists. Recommendation for continued use was reported by 21.7% (5/23) of neuroradiologists and 33.3% (14/42) of non-neuroradiologists. After false discovery rate correction, no exploratory subgroup differences remained statistically significant. Descriptive subgroup comparisons for selected survey items are shown in Figure 1.

Figure 1. Descriptive comparison of agreement by subspecialty training for selected survey items. Bars show the percentage of respondents selecting "agree" or "strongly agree" among neuroradiologists and non-neuroradiologists for selected items related to false positive (FP) burden, trust in artificial intelligence (AI) output, recommendation for continued use, and reported less thorough than usual scrutiny of AI-negative results. These subgroup comparisons are descriptive; no exploratory subgroup differences remained statistically significant after false discovery rate correction.



Qualitative Themes From Free-Text Comments

Free-text responses reinforced the quantitative findings and highlighted recurring implementation concerns.

Theme 1: Artifact- and Calcification-Driven False Positives Dominated the User Experience

Many respondents attributed false-positive alerts to parafalcine calcification, skull-adjacent artifacts, motion, or streak artifacts and described these alerts as disruptive or time-consuming.

XSL•FO
RenderX

Theme 2: Perceived Utility Was Limited Unless the Tool Meaningfully Changed Triage

Several respondents indicated that the system primarily identified obvious hemorrhages and offered limited incremental value for experienced readers, although some suggested that workflow-level prioritization or routing features might improve perceived utility.

Theme 3: Latency and Inconsistent Availability Reduced Usefulness

Multiple respondents noted that AI output was not always available at the time of dictation or arrived too late to be helpful.

Theme 4: False-Positive Alerts Increased Second-Guessing and Consultation Burden

Some comments suggested that false-positive AI flags triggered unnecessary consultation or additional cognitive burden.

Theme 5: Medicolegal Concerns Influenced Acceptance

Respondents expressed discomfort with the possibility that delayed or discrepant AI outputs could create liability concerns.

Theme 6: Respondents Wanted Better Localization, Higher Specificity, and More Training

Suggested improvements included more precise localization, lower false positive rates, better timing, and implementation feedback or training.

Aggregated Themes

Aggregated qualitative themes and representative excerpts are summarized in Table 2.

Table 2. Qualitative themes from free-text responses. Comments could map to more than one theme; counts reflect the number of comments mentioning each theme at least once (N=65)[a].

Theme	Description	Count, n (%)	Representative excerpt
False positives or low specificity	Frequent false-positive alerts, commonly attributed to artifacts or calcification; perceived as disruptive and time-consuming	22 (33.8)	“False positives are too frequent...”
Limited utility, discontinue, or low value	Negative overall sentiment, including requests to discontinue the tool or concerns that its value did not justify its burden	13 (20)	“Please remove AI completely.”
Latency or availability	AI[b] output was sometimes delayed or unavailable at the time of dictation or interpretation	7 (10.8)	“Sometimes AI image made available very late...”
Medicolegal concerns	Concerns about legal vulnerability, accountability, or liability if the AI output appeared after sign-off or conflicted with their interpretation	5 (7.7)	“It could potentially ‘frame’ me.”
False negatives or missed hemorrhages	Concern that the tool could miss subtle hemorrhages, particularly smaller or less conspicuous bleeds	4 (6.2)	“I have seen several false negatives.”
Positive or supportive comments	Explicitly positive feedback or support for continued use despite limitations	4 (6.2)	“The AI ICH tool is great!”
Triage, prioritization, or assignment routing	Suggestions that the tool might be more useful if integrated into prioritization or routing workflows	3 (4.6)	“Positive studies should be assigned...and prioritized.”
Consultation burden or second-guessing	Reports that false-positive AI alerts increased consultation burden or second-guessing	3 (4.6)	“I frequently get consults.”
Training or feedback requests	Requests for additional training, implementation guidance, or feedback on use	2 (3.1)	“I would like additional training.”

[a]Aggregated qualitative findings from optional free-text comments are summarized. Because individual comments could mention more than one concern, counts are not mutually exclusive and reflect the number of comments mentioning each theme at least once.

[b]AI: artificial intelligence.

Discussion

Principal Findings

In this survey of radiologists exposed to an FDA-cleared ICH AI detection tool in a national teleradiology environment, 3 findings stood out. First, false positive burden was a central driver of dissatisfaction: fewer than 1 in 5 respondents found false-positive alerts infrequent enough to be acceptable, and one-third (22/65, 33.8%) agreed that time spent reviewing false-positive alerts outweighed the tool’s benefits. Second, perceived workflow benefit was limited: only approximately 1 in 10 respondents reported reduced interpretation time despite a larger minority reporting that the tool might reduce the risk of missing subtle hemorrhages during busy shifts. Third, trust in the AI output was strongly conditional. Respondents were willing to trust the AI much more often when it agreed with their own interpretation than when it conflicted with it.

Taken together, these findings suggest that the practical value of this implementation was constrained less by the abstract idea of AI assistance than by the combined effects of false positive burden, timing, and workflow integration. In other words, the tool appears to have been experienced more as an additional signal to manage than as a consistently efficiency-enhancing aid.

XSL•FO
RenderX

Relationship to Prior Work

These results are consistent with a growing literature showing that post-deployment experience with clinical AI may differ substantially from laboratory or regulatory performance summaries [2-6]. In radiology and other clinical settings, user acceptance is shaped not only by sensitivity and specificity but also by local disease prevalence, the timing of alerts, the interpretability of outputs, and how often the system changes action in a useful direction [5,6]. Scaringi et al [6] described how real-world deployment can produce an "accuracy paradox" in which acceptable technical performance is experienced as poor because positive alerts carry a high false discovery burden in a low-prevalence setting. That framework is highly relevant in this study, where many respondents perceived the false positive rate as a major practical barrier.

These findings also align with human factors literature on alarm burden and trust in automation. High alert burden and repeated false-positive alerts can result in cognitive overhead, desensitization, and reduced perceived usefulness of decision support systems [7,8]. Meanwhile, trust in automation is often selective rather than global; clinicians may accept AI as supportive when it confirms a preexisting judgment but remain reluctant to defer to it when it conflicts with their own assessment [10,11]. Our survey findings fit that pattern closely. However, importantly, because we measured perceptions rather than observed task performance, our data support cautious description of conditional trust patterns, not stronger causal inferences about AI-driven decision change.

Interpreting the Self-Reported Behavior Item Cautiously

A survey item about scrutinizing AI-negative results less thoroughly does not establish a behavioral phenomenon in the same way that a controlled multi-reader study could. Accordingly, we interpret this result narrowly. The 6.2% (4/65) agreement rate should be read only as the proportion of respondents who self-reported less thorough scrutiny after a negative AI result on that item. It should not be construed as direct evidence of automation-induced complacency or measurable diagnostic underreading. This distinction is important, particularly because self-report on professionally sensitive behaviors may be influenced by social desirability bias [9,10].

At the same time, the item remains informative as a perception signal. Even a small minority of respondents endorsing reduced scrutiny suggests that how negative AI outputs are presented, timed, and contextualized may matter. Future work using controlled reader studies, audit-based designs, or prospective workflow observation would be better suited to determining whether and when AI output measurably alters vigilance or diagnostic behavior.

Why False Positive Burden Matters So Much in Teleradiology

The teleradiology environment may amplify the operational importance of false positives. High case volume, rapid turnaround expectations, and distributed communication structures can magnify the cost of reviewing spurious alerts or revisiting already completed interpretations. In such settings, an alert that does not change management may still consume attention, create uncertainty, or generate downstream consultation. The free-text comments in this study point to precisely these burdens: artifact-driven overcalls, delayed output, second-guessing, and extra communication work.

This observation has practical implications for how radiology AI tools should be tuned and evaluated. In some settings, maximizing sensitivity may be appropriate, particularly when AI outputs are used to support prioritization or highlight potentially urgent findings; however, if the resulting false positive burden is too high, this may erode clinician acceptance even when the tool performs as intended at a regulatory or technical level [3,4,6]. Therefore, a workflow-sensitive implementation strategy may matter as much as the operating characteristics of the algorithm itself.

Implications for Implementation

Several implementation lessons emerge from these data. First, specificity matters. The comments suggest that calcification, streak artifacts, motion, and skull-adjacent findings were recurrent sources of false-positive alerts. Improving discrimination in these common failure modes could meaningfully improve user acceptance. Second, interpretability matters. If discrepant alerts are to influence radiologist behavior in a useful way, users may need clearer localization, confidence communication, or other forms of contextual explanation rather than a generic positive signal [4,9]. Third, timing matters. Decision support that appears after a case has effectively been interpreted is less likely to help and more likely to create noise or ambiguity. Fourth, workflow design matters. Some respondents suggested that the tool's value might be enhanced by workflow-level prioritization or routing features in addition to its use as an overlay during interpretation.

Subgroup Findings and Their Interpretation

Neuroradiologists were more likely than non-neuroradiologists to endorse the primary end point that false positive review time outweighed benefits, but this finding should be interpreted cautiously. It was the only subgroup result meeting nominal significance, and no exploratory subgroup differences remained significant after false discovery rate correction. Therefore, the most appropriate interpretation is descriptive and hypothesis generating rather than confirmatory. It is possible that subspecialty training influences tolerance for false positives, perceived incremental value, or expectations of tool performance, but this survey was not designed to establish those mechanisms.

Strengths and Limitations

A strength of this study is that it captured the views of radiologists exposed to the tool in routine clinical practice rather than in an artificial test setting. The inclusion of both structured survey items and free-text comments helped contextualize quantitative responses and identify concrete implementation barriers.

This study also has several limitations. First, the findings are based on self-reported perceptions rather than objective

XSL•FO
RenderX

measures of diagnostic performance, reading time, or patient outcomes. Second, the survey instrument was not formally validated, so items should be interpreted at face value rather than as established scales of latent constructs. Third, dichotomizing Likert responses improved interpretability but reduced response granularity. Fourth, the sample reflects a single organizational implementation and may not generalize to other workflows, prevalence environments, or AI products. Fifth, responses may be affected by selection bias and social desirability bias, particularly for items concerning vigilance or reliance. Finally, subgroup analyses were exploratory, and the study was not powered to make strong claims about differences by neuroradiology training.

Conclusions

Radiologists using an FDA-cleared ICH AI detection tool in a national teleradiology environment reported substantial false positive burden, limited perceived time savings, and strongly conditional trust in AI output. Self-reported reduced scrutiny after negative AI results was uncommon but endorsed by a minority of respondents. The findings suggest that the real-world value of radiology AI tools depends not only on whether they detect disease but also on whether they do so with sufficient specificity, appropriate timing, and workflow-compatible presentation to earn sustained clinician acceptance.

Data Availability

All data generated or analyzed during this study are included in this published article.

Funding

No funding was received for this work.

Authors' Contributions

Conceptualization: AJDG (lead), JKDG (supporting)
Data curation: JKDG (lead), AJDG (supporting)
Formal analysis: TAS (lead), AJDG (supporting)
Funding acquisition: None
Investigation: JKDG (lead), AJDG (supporting)
Methodology: AJDG (lead), JKDG (supporting)
Project administration: AJDG (lead)
Resources: AJDG
Supervision: AJDG
Validation: AJDG
Visualization: AJDG
Writing—original draft: AJDG (lead), JKDG (supporting)
Writing—review and editing: AJDG (lead), JKDG (supporting), TAS (supporting)

Conflicts of Interest

None declared.

References

1. Del Gaizo AJ, Osborne TF, Shahoumian T, Sherrier R. Deep learning to detect intracranial hemorrhage in a national teleradiology program and the impact on interpretation time. Radiol Artif Intell 2024 Sep;6(5):e240067. [doi: 10.1148/ryai.240067] [Medline: 39017032]
2. Pettet G, West J, Robert D, Khetani A, Kumar S, Golla S, et al. A retrospective audit of an artificial intelligence software for the detection of intracranial haemorrhage used by a teleradiology company in the United Kingdom. BJR Open 2024 Oct 04;6(1):tzae033 [FREE Full text] [doi: 10.1093/bjro/tzae033] [Medline: 39479271]
3. Rohren E, Ahmadzade M, Colella S, Kottler N, Krishnan S, Poff J, et al. Post-deployment monitoring of AI performance in intracranial hemorrhage detection by ChatGPT. Acad Radiol 2025 Oct;32(10):6104-6113 [FREE Full text] [doi: 10.1016/j.acra.2025.07.055] [Medline: 40796469]
4. Garcia GM, Young P, Dawood L, Elshikh M. Head-to-head comparison of 2 artificial intelligence computer-aided triage solutions for detecting intracranial hemorrhage on noncontrast head CT. AJNR Am J Neuroradiol 2026 Feb 03;47(2):376-385. [doi: 10.3174/ajnr.A8986] [Medline: 40957692]
5. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. BMJ 2022 May 18;377:e070904 [FREE Full text] [doi: 10.1136/bmj-2022-070904] [Medline: 35584845]
6. Scaringi JA, McTaggart RA, Alvin MD, Atalay M, Bernstein MH, Jayaraman MV, et al. Implementing an AI algorithm in the clinical setting: a case study for the accuracy paradox. Eur Radiol 2025 Jul;35(7):4347-4353. [doi: 10.1007/s00330-024-11332-z] [Medline: 39741216]

XSL•FO
RenderX

7. Sendelbach S, Funk M. Alarm fatigue: a patient safety concern. AACN Adv Crit Care 2013 Oct 01;24(4):378-386. [doi: 10.1097/nci.0b013e3182a903f9]
8. Ancker JS, Edwards A, Nosal S, Hauser D, Mauer E, Kaushal R. Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. BMC Med Inform Decis Mak 2017 Apr 10;17(1):36 [FREE Full text] [doi: 10.1186/s12911-017-0430-8] [Medline: 28395667]
9. Bernstein MH, Atalay MK, Dibble EH, Maxwell AW, Karam AR, Agarwal S, et al. Can incorrect artificial intelligence (AI) results impact radiologists, and if so, what can we do about it? A multi-reader pilot study of lung cancer detection with chest radiography. Eur Radiol 2023 Nov;33(11):8263-8269 [FREE Full text] [doi: 10.1007/s00330-023-09747-1] [Medline: 37266657]
10. Parasuraman R, Riley V. Humans and automation: use, misuse, disuse, abuse. Hum Factors 1997 Jun;39(2):230-253. [doi: 10.1518/001872097778543886]
11. Dzindolet MT, Peterson SA, Pomranky RA, Pierce LG, Beck HP. The role of trust in automation reliance. Int J Hum Comput Stud 2003 Jun;58(6):697-718. [doi: 10.1016/S1071-5819(03)00038-7]

Abbreviations

AI: artificial intelligence
CT: computed tomography
DECIDE-AI: Developmental and Exploratory Clinical Investigations of Decision Support Systems Driven by Artificial Intelligence
FDA: Food and Drug Administration
ICH: intracranial hemorrhage

Edited by A Kushniruk; submitted 25.Jan.2026; peer-reviewed by D Lyell, GL Baird; comments to author 19.Mar.2026; revised version received 13.Apr.2026; accepted 04.May.2026; published 02.Jun.2026.

Please cite as:
Del Gaizo AJ, Del Gaizo JK, Shahoumian TA
Radiologist Perceptions of an AI Tool for Intracranial Hemorrhage Detection in Teleradiology: Cross-Sectional Survey Study
JMIR Hum Factors 2026;13:e92145
URL: https://humanfactors.jmir.org/2026/1/e92145
doi:10.2196/92145
PMID:

©Andrew J Del Gaizo, Jackson K Del Gaizo, Troy A Shahoumian. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 02.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Preliminary Support for a Technology-Enhanced Suicide Prevention Intervention for College Students: Mixed Methods Preimplementation Study

Jocelyn I Meza[1*], PhD; Joan R Asarnow[1*], PhD; Natalia Jaramillo[1*], PhD; Juliane L Martinez[1*], BA; Candice Biernesser[2*], PhD, LCSW; Brandie George-Milford[2*], MA; David Brent[2], MD

[1]Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, 300 Medical Plaza Driveway Room 1235, Los Angeles, CA, United States

[2]Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, United States

[*]these authors contributed equally

Corresponding Author:
Jocelyn I Meza, PhD
Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, 300 Medical Plaza Driveway Room 1235, Los Angeles, CA, United States

Abstract

Background: Self-injurious thoughts and behaviors, including suicidal ideation, planning, attempts, and nonsuicidal self-injury, are a significant public health concern among college students. Community college students face elevated mental health risks and persistent barriers to traditional services, particularly those from underserved and ethnoracially minoritized backgrounds, underscoring the need for scalable, accessible digital mental health interventions.

Objective: This study conducted a mixed methods evaluation of the initial reactions of college students to a novel technology-enhanced suicide prevention intervention (TE-SPI), with particular focus on feedback regarding a mobile app prototype (BRITE) designed to support coping skill use and mood monitoring, as well as proposed nondemanding caring contact text messages.

Methods: A mixed methods approach using the technology acceptance model (TAM) was used to assess the TE-SPI among 20 ethnoracially diverse college students (mean age 22.8, SD 3.9) from a community college (n=10) and a 4-year university (n=10). Participants were shown the BRITE app prototype on a study iPad and asked to provide feedback on its perceived ease of use, usefulness, and relevance to student needs. Quantitative measures assessed initial impressions of satisfaction and usability, while qualitative analyses examined participants' reactions to the app prototype and preferences regarding potential caring contact text messages.

Results: Participants generally reported favorable initial impressions of the BRITE app prototype, with 90% (18/20) describing it as helpful and 85% (17/20) user-friendly; 90% (18/20) also indicated that they would recommend it to peers in distress. Qualitative feedback supported these findings, highlighting the app's perceived accessibility, relevance to student life, and potential usefulness during moments of distress. Participants also offered concrete suggestions for app refinements (eg, mood scale customization, gamification, and language options) and for useful caring contact messages they would want to receive. In total, 55% (11/20) of participants said that they would want to receive text messages as part of TE-SPI, while 25% (5/20) reported maybe or possibly wanting to receive messages.

Conclusions: Findings offer preliminary insight into how college students perceive the BRITE app prototype and proposed caring contact message, highlighting concrete feedback for app refinement prior to real-world testing. This study provides early user feedback that can inform further development of the TE-SPI model. Future research should examine actual use, acceptability, feasibility, and clinical impact among diverse college students.

(JMIR Hum Factors 2026;13:e79537) doi:10.2196/79537

KEYWORDS

suicide prevention; young adults; mobile intervention; college students; safety planning; caring contacts

XSL•FO
RenderX

Introduction

Background

Suicide is the second leading cause of death among college-age adults [1]. Additionally, self-injurious thoughts and behaviors (SITBs)—encompassing both thoughts and behaviors involving suicidal intent, as well as deliberate self-injury without intent to die—constitute major public health issues. Data show that in the United States, 11% of college students have experienced suicidal ideation, 5% have made a suicide plan, and 2% have attempted suicide in the previous 12 months [2], and self-harm ranked third for young people ages 10 - 24 years in the most recent Global Burden of Disease Survey [3]. Each year approximately 93,000 young adults (ages 18 - 24 years) present to emergency departments for SITBs [4]. Furthermore, suicide rates for this age group have remained particularly high over the last decade and have disproportionately increased among ethnoracially minoritized groups in the United States [1,5]. Previous research identifies elevated symptoms of anxiety and depression as key risk factors associated with the risk of suicide among college students [6,7], particularly if left untreated [8].

Mental Health Challenges in College Students

College students represent a diverse group of students with varying mental health profiles and needs depending on individual characteristics, socioecological characteristics, and college contexts underscoring the importance of considering student context and cultural relevance when developing mental health interventions for this population. Among the college student population, the subgroup of students attending community colleges reports higher traumatic life events, poorer physical health, and more severe psychological concerns compared to students at 4-year institutions [9-11]. Ethnoracially minoritized students and low-income populations, who make up a significant portion of the community college student body, are also less likely to engage with behavioral health services and are more likely to drop out of treatment [12-15]. Community college students who transfer to 4-year academic institutions also experience elevated risk for mental health outcomes, including SITBs, because they face several stressors as they navigate new campus climates, build social networks, and adjust to academic expectations at 4-year academic institutions [16,17]. In fact, data show that transfer students report significantly higher levels of depression, social anxiety, and educational and family-related distress compared to nontransfer students [18]. Contextualizing treatments to address the needs of community college students, including those who transfer to 4-year institutions, can lead to higher treatment engagement.

The diverse mental health challenges faced by community college students, along with limited access to resources, highlight the urgent need for innovative suicide prevention strategies. A few of the common barriers faced by community college campuses include financial restrictions, staff shortages, long waiting lists, vague and unstandardized crisis response guidelines, and limited student outreach [9,19,20]. In addition, some students choose not to report their mental health symptoms or seek professional support during a crisis as a result of perceived stigma, apprehension, or feeling that they may be a burden. To address this gap, novel evidence-based digital mental health interventions present a promising solution, particularly for underresourced community colleges.

Digital Mobile Apps for Suicide Prevention

Digital mental health interventions may bolster mental health care on college campuses and can promote skills that mitigate risk for SITBs [21-23]. Mental health apps have potential for offering timely support, especially given widespread reports of time constraints, perceived inconvenience associated with traditional face-to-face services, and the high levels of comfort and acceptance of technology among students [15,21,22,24]. Digital tools also have the potential to reduce suicide risk by providing discreet support for individuals struggling with mental health issues who may be hesitant to seek out resources and can offer a range of intervention components, including support for crisis intervention, mood tracking, and coping strategies, which can help users manage their symptoms [25]. Additionally, digital tools allow for continuous monitoring and support, offering timely interventions during moments of crisis. By integrating features such as direct links to crisis hotlines and emergency contacts, digital tools can ensure that users receive the necessary assistance promptly. From a research standpoint, user data collected can also offer objective insights on usability, guiding iterative updates and improvements. Overall, digital mental health tools can play a crucial role in suicide prevention by making resources more accessible and tailored to individual needs.

Despite the rapid growth of digital mental health tools, few are designed with the unique needs of college students in mind. In particular, rigorously tested apps for suicide prevention have primarily focused on adolescents who present to acute care settings. For example, BlueIce, an app with mood ratings and personalized features to improve mood (ie, using cognitive behavioral therapy and dialectical behavior therapy skills), was specifically codeveloped with adolescents with lived experience, and the phase 1 open uncontrolled trial found that nearly three-quarters of participants had reductions in self-harm at 12-week follow-up [26]. Similarly, not all evaluations of suicide prevention apps include ethnoracially minoritized samples, with the exception of the iBobbly app, a suicide prevention app for Indigenous Australian youths and young adults [27]. As Ramos et al [15] highlight, most existing app evaluation frameworks fail to account for critical diversity, equity, and inclusion factors such as literacy, cultural relevance, and technology access, which limits the effectiveness and equity of digital interventions. This gap underscores the need for more inclusive, accessible, and clinically guided digital tools. Indeed, community college students remain significantly underrepresented in suicide research, particularly around research that focuses on the development of digital mental health and suicide prevention tools, underscoring a significant gap and the urgency to address current SITB disparities for at-risk college students [28].

Technology-Enhanced Approach to Suicide Prevention

Combining the promise of digital interventions with a “human touch” has shown promise for increasing use of digital interventions, an important goal given that low use limits intervention exposure and benefits and has been a problem in

XSL•FO
RenderX

trials of prior interventions [29-32]. In an effort to optimize intervention use and value, we pilot-tested a technology-enhanced suicide prevention intervention (TE-SPI). Through TE-SPI, we aimed to adapt a previously validated safety planning and emotion regulation smartphone app, initially designed for adolescents [25,33,34], for use among college students. To increase engagement within this population, we added a new component, caring contact text messages, which are nondemanding messages of care designed to increase feelings of connection and support postcrisis [35-37]. Caring contacts have demonstrated high feasibility and acceptability among adults when delivered via email or text [38,39]. Although some data indicate a 1-year benefit in reducing self-reported suicide attempts [37,40], evidence remains mixed regarding its effect on suicide mortality or self-harm–related emergency department and hospital visits. Moreover, there is limited research on its effectiveness among diverse college student populations. Therefore, this project presents formative prototype evaluation data based on participant feedback to inform the adaptation and refinement of 2 synergistic intervention components for college students at risk for SITBs.

Technology Acceptance Model

The technology acceptance model (TAM) has been widely used to study the implementation of new technologies and digital tools, including pilot trials for mental health apps [41,42]. The TAM framework, based on the theory of reasoned action [43], posits that an individual's behavior (ie, use of a digital tool) is influenced by their attitudes and perceptions. We used TAM as a guiding framework to understand how college students perceived the components of a TE-SPI. More specifically, according to the TAM, 2 key factors predict an individual's engagement and use of new technology: perceived ease of use and perceived usefulness [41]. Perceived ease of use refers to how easily a user believes that they can learn and use the technology without difficulties, while perceived usefulness pertains to the belief that the new technology can enhance their daily routines [41]. However, an individual's perception of new technology can also be significantly shaped by their social environment and context [44]. For instance, perceptions may change based on friends' or family members' reactions and experiences with the digital tool, and community college students might find digital tools more beneficial if they are contextualized to their experiences and needs.

This Study

This mixed methods study aimed to explore the perceptions of college students of a TE-SPI designed to reduce distress and improve mental health well-being. We used the TAM framework to examine the perceived ease of use and perceived usefulness of the TE-SPI. The primary research questions guiding the study were:

- Question 1: What are community college students' perceptions of the TE-SPI features?
- Question 2: What is the perceived usability of the TE-SPI? (ie, what are the aspects and factors, in terms of functionality and utility, that students find useful?)
- Question 3: What aspects of the TE-SPI do students find more engaging?
- Question 4: What factors may influence the decisions of community college students to use the TE-SPI?

Methods

Research Design

Procedures and Participants

This substudy is part of a large research study examining the Screening and Treatment for Anxiety & Depression (STAND) system of care in a community college setting. The STAND system of care was developed as part of the University of California, Los Angeles (UCLA) Depression Grand Challenge to be a scalable approach to preventing, identifying, and treating anxiety and depression [45]. The STAND study was approved by the institutional review board at UCLA (IRB #22 - 0205). Participants were recruited from a local community college (East Los Angeles Community College [ELAC]) and a Transfer Student Center for community college students at a 4-year university (UCLA), enhancing student diversity over that afforded by a single community college site. Recruitment was from the fall of 2022 through the winter of 2023. Participants from ELAC were recruited through a listserv of students interested in participating in STAND-related research studies, and UCLA participants were recruited using outreach through informational presentations and flyer postings at the Transfer Student Center. Study eligibility criteria included (1) currently enrolled at ELAC or UCLA as a transfer student, (2) owning a smartphone, and (3) English proficiency. Eligible students interested in participating were screened on the phone and provided oral consent. Participants then scheduled an interview date with the study staff to complete the semistructured interview and brief surveys. All participants received a US $25 e-gift card for their participation in the study.

To ensure participant safety and confidentiality, participants were informed during the consent process that they could decline to answer any questions or stop the interview at any time without penalty due to the sensitive nature of the study topics (eg, distress, depression, and anxiety). Research staff conducting the interviews were trained to recognize signs of distress and to pause or discontinue the interview if participants became notably distressed. Research staff were also trained to assess the participant's well-being and offer additional support and mental health resources to all participants. Additionally, limits of confidentiality were clearly communicated to participants, including mandated reporting requirements related to risk of harm to self or others. A safety protocol was activated when participants disclosed current active suicidal ideation, particularly with intent or plan. The safety protocol included a risk assessment and safety planning by a licensed clinical psychologist, although there were no cases in this study that required the use of the safety protocol.

In total, 14 students from ELAC and 13 students from UCLA were eligible to participate in the study and provided oral consent. A total of 4 participants from ELAC and 3 from UCLA were excluded from the study due to failure to attend their scheduled semistructured interviews. Therefore, 20 participants were included in the analysis. Participants (N=20) in the study

XSL•FO
RenderX

were between the ages of 19 and 32 (mean 22.8, SD 3.9) years. The sample consisted of both community college students (n=10) at ELAC and transfer students at UCLA (n=10).

Data Collection

College students participated in the study in person or virtually using a secure videoconferencing platform (Zoom, UCLA Health System Protected). Participation consisted of completing a semistructured interview and 2 brief online surveys with a research staff. Semistructured interviews lasted up to 45 to 60 minutes each (average time was approximately 36, SD 12.8 minutes) and were audio-recorded and stored in a secure UCLA Box folder. Prior to starting the semistructured interview, participants were asked to complete the first brief online survey administered via Qualtrics inquiring about their demographic background, mental health use, and previous digital tool use for mental health concerns. Next, participants were introduced to the technology TE-SPI components, specifically (1) the BRITE app, which includes mood and distress ratings, a menu of distress tolerance and emotion regulation skills [25,33], and a coping or safety plan that students can turn to when needed to cope effectively (eg, when distress is high or increasing); and (2) caring contacts (brief nondemanding text messages aimed at providing connection to the intervention team and support [46]). As part of the prototype demonstration, participants were informed that both the BRITE app and caring contacts are designed to be introduced by a mental health provider. Next, the participants were given a study iPad and received a demonstration of how they would be onboarded to get access to the app (ie, after completing a safety plan with a clinician, although the development of a safety plan was not conducted as part of the demonstration). Then, students were given a tutorial on how they could create and individualize their account or profile to use the various features including (1) mood ratings, (2) use of coping activities, and (3) how to activate the support feature. Following this process, participants were given time to explore the app on the study iPad and were encouraged to ask questions regarding app features. After having time to review the app, participants were asked to share their perceptions and feedback on the app features. Next, the concept of caring contacts was presented with various examples, and participants were asked to share their perceptions and feedback on the BRITE app and caring contact TE-SPI components. Feedback for the clinician session to develop a safety plan was not included as part of the feedback, given its already established feasibility and acceptability in college samples [47]. Throughout the interview process, the research staff used a semistructured interview protocol to guide the participant through each of the questions. Following the interview, participants completed the second brief quantitative survey inquiring about their perceptions of the TE-SPI components and overall satisfaction questions. At the end of the study, all participants were sent information regarding campus and community mental health resources and received a US $25 e-gift card as compensation for participating in the study.

Ethical Considerations

This substudy was conducted separately from the STAND trial and reviewed and approved by the institutional review board at UCLA (IRB #22‐0244). All procedures complied with institutional and federal ethical standards. Informed consent was obtained from all participants prior to study participation. Participants were informed that their participation was voluntary and that they could decline to participate or withdraw from the study at any time without penalty. Additionally, to protect participants' privacy and confidentiality, all identifying information was removed from the study data prior to analysis.

Materials and Measures

TE-SPI Materials

BRITE App

The BRITE app includes (1) the development of a coping plan with a clinician, designed to support effective coping strategies that include evidence-based suicide prevention intervention tools; (2) mood and distress ratings; (3) a menu of distress tolerance and emotion regulation skills; and (4) a reaching out feature for support during distress. On the BRITE app, users are prompted to rate their mood or distress daily, but they can also self-initiate their own ratings. After entering a mood and distress rating into the app, users are provided with potentially helpful skills and activities (ie, mindfulness and distress tolerance skills) that they can complete independently to improve their mood or reduce their level of distress. If participants choose to complete the suggested activities, they are asked again to rate their mood and distress level after completing the suggested activities. Participants can personalize their profiles in the BRITE app by uploading photographs, videos, or links to websites and can update their coping plan with individualized warning signs, coping skills, and phone contacts for interpersonal support and clinician contact. Although participants were not asked to download the BRITE app onto their phones, they were given time to review all features of the tool via an iPad that was provided to them for use during the interview. See Kennard et al [25,33] for more details on the development of the digital tool (BRITE app), which was primarily developed for adolescents in acute care settings (ie, inpatient hospitalization).

Caring Contacts

Additionally, the TE-SPI model adds caring contacts in addition to the BRITE app to create a unified platform. Caring contact messages are brief, low-cost, nondemanding expressions of care—typically delivered via text, email, phone, or mail (eg, "Hey there, I'm wishing you well. I hope you're taking good care of yourself.")—and are traditionally sent by a care provider or care team to remind individuals that support is available and to promote connection [36,46]. As part of the TE-SPI model, the concept of caring contacts was introduced to participants. In this study, the study interviewer explained the approach for these messages as well as who and when these messages would be sent. Students were also provided with example messages such as "Hey, I've enjoyed working with you. Hope your day is going okay so far." Students were then asked about their perceptions of this approach, whether they would like to receive such messages, and were asked to share an example of a caring contact message that they would find meaningful and helpful during moments of distress.

XSL•FO
RenderX

Measures

Mental Health Status

Current symptoms of depression and anxiety were assessed using the Patient Health Questionnaire 2-item (PHQ-2) and the Generalized Anxiety Disorder 2-item (GAD-2 [47,48]). Both the PHQ-2 and GAD-2 measures were used to descriptively characterize the mental health symptom profile of the sample and were not used for hypothesis testing or subgroup analyses. The PHQ-2 is a self-report measure of depression symptoms over the last 2 weeks. The 2 items from PHQ-2 were the following: "Over the last 2 weeks, how often have you been bothered by having little interest or pleasure in doing things?" and "Over the last 2 weeks, how often have you been bothered by feeling down, depressed, or hopeless?" Response options ranged from 1=not at all to 4=nearly every day. The measure has good internal consistency and concurrent validity [47,49]. The GAD-2 is a self-report measure of anxiety symptoms over the last 2 weeks. The GAD-2 includes the following 2 items: "Over the last 2 weeks, how often have you been bothered by feeling nervous, anxious or on edge?" and "Over the last 2 weeks, how often have you been bothered by not being able to stop or control worrying?" Response options ranged from 1=not at all to 4=nearly every day. The measure has strong psychometric properties and has been found to be reliable [47,49]. Both items from the PHQ-2 and GAD-2 were summed separately, and a score of 3 was used as the cutoff for identifying possible cases of depression and generalized anxiety disorder, respectively [47,48,50]. The PHQ-2 (α=.826) and GAD-2 (α=.928) measures have strong internal consistency in our sample.

Feasibility and Acceptability

Feasibility and acceptability were assessed using both quantitative and qualitative indicators, including participant ratings of usability, satisfaction, perceived usefulness, engagement with app features, and qualitative feedback regarding BRITE app and caring contacts' relevance, accessibility, and suggested refinement.

Technology Use

To examine participants' overall technology use, we formulated questions to inquire about their access to a cell phone or smartphone, data plan, how often participants carry their phone, and specifics on how much time they spend on their mobile device (time measured in total minutes per day) for the following activities: texting, talking, emailing, surfing the web, playing games, social media apps, and mental health apps. Sample questions included "Do you have a smartphone?" and "How often do you carry your phone with you?." Additionally, participants were instructed to list the mobile apps that they frequently use and the type of features they found useful.

Digital Intervention Barriers

The Digital Intervention Barriers Scale-7 [44] was used to assess participants' perception of the TE-SPI, specifically their perceptions of engagement, motivations for use, and if the digital tool was adequate (not too long or too short). This measure asked students to "Please indicate the extent to which you agree or disagree with the following statements." Statements included (1) "I didn't understand the tasks or things I was supposed to do on the app.", (2) "I thought the app wasn't engaging.", (3) "It would be difficult to keep myself motivated to use the app.", and (4) "I thought the length of the app wasn't adequate (too long or too short)." The answer choices participants were able to choose from for all 4 questions were 1=totally disagree, 2=disagree, 3=neutral, 4=agree, and 5=totally agree. Participant responses were scored by obtaining a sum, with higher scores indicating more barriers to future use and engagement with the digital tool. Three items from the original Digital Interventions Barriers Scale-7 were omitted, given that students were not asked to download the app during this phase of the study. The omitted items were (1) "I had technical problems with my technology (eg, device, internet, platform).", (2) "I forgot to use the app.", (3) "The app didn't seem to be helping me." This measure has been found to have high internal consistency and criterion-related validity [44]. The internal consistency of the 4 items was high (α=.783).

Satisfaction With the BRITE App Component of TE-SPI

Using a questionnaire developed by the University of Pittsburgh [51], we assessed participants' satisfaction of the digital tool, by administering the following questions: (1) "If you started to feel really stressed and upset, do you think you would use the BRITE app?" with the following answer choice for participants to choose from 1=no, 2=possibly, 3=probably, 4=definitely, and 5=prefer not to answer; (2) "If a friend was in need of mental health care for depression or suicidal thoughts or behaviors, would you recommend the BRITE app to help them cope?" with the following answer choices: 1=definitely not, 2=no, I don't think so, 3=yes, I think so, 4=yes, definitely, and 5=prefer not to answer; (3) "How often do you think you would use the BRITE app if it were available to you during the month after it was introduced to you?" with the answer choices being: 1=never, 2=once or twice, 3=3 or 4 times, 4=5 or 10 times, and 5=more than 10 times; (4) "Do you think you would be more likely to use the BRITE app if you received messages reminding you of the app?" with the following answer choices: 1=no, I would find these messages annoying, 2=no, but the messages would not bother me, 3=maybe, 4=possibly, and 5=yes. After asking each of these questions, participants could provide additional written feedback or reasoning for their answer choices. The satisfaction with BRITE score was obtained by creating a composite score that summed all 4 questions in the measure.

Poststudy Satisfaction and Usability

To obtain more information regarding participants' experience with the TE-SPI components, they answered items from the Poststudy System Usability Questionnaire [42]. The Poststudy System Usability Questionnaire has been used in previous suicide prevention trials for adolescents after a hospitalization [34]. The questionnaire consists of 6 items that identify system usability. Participants rated questions on a 7-point Likert scale, ranging from 1=strongly disagree to 7=strongly agree. The following questions were asked: (1) "Overall, I am satisfied with how easy it is to use BRITE app."; (2) "The information provided in the BRITE app such as on-screen messages, and other documentation like links or pictures for the topic activities,

XSL•FO
RenderX

was clear."; (3) "The information provided in the BRITE app, such as notifications to rate my distress or mood with emojis, was clear."; (4) "I like interacting with the BRITE app program."; (5) "I needed to learn a lot of things before I could get going with the BRITE app."; and (6) "The information provided, such as the coping plan for BRITE users, was clear." This measure is reliable with a Cronbach α that spans from 0.83 to 0.96 [51,52]. A composite sum score was created, such that a higher score indicated higher satisfaction and perceived usability. An additional question was added to assess participants' perceptions of caring contacts. Specifically, participants were asked "Do you think you would be more likely to use the BRITE app if you received text messages from your mental health provider?," with the following answer options: 1=no, I would find them annoying, 2=no, but the messages would not bother me, 3=maybe, 4=possibly, and 5=yes.

Data Analysis

Quantitative Analysis

Descriptive analyses were conducted using SPSS (version 25; IBM Corp) software. We carried out descriptive statistics to obtain frequencies and percentages, means, and medians along with SDs and ranges for all quantitative measures.

Qualitative Analysis

For the qualitative analyses, all interview audio recordings were auto-transcribed using Microsoft Word software, and the transcripts were then reviewed individually by 2 independent researchers (NJ and JLM) to verify accuracy. Errors from the transcripts were fixed, and any identifiable participant information was removed. Additionally, the supervising researcher (NJ) checked each transcript for errors. The transcripts were then used for data coding.

A rapid qualitative coding approach was used from the health services literature [53], which prioritizes the efficient delivery of information to guide decision-making, particularly during preimplementation phases. First, the team developed a codebook with domains that were identified by the TAM framework. Then, using deductive and inductive strategies, coders identified additional codes and themes beyond the predefined TAM domains. After the codebook was finalized, 2 trained members independently applied domain analysis and created a summary template for each transcript. After all transcripts were coded and summarized, a matrix method was used to efficiently summarize and explore themes across all interview transcripts [54]. Themes were derived through comparison across structured summaries and matrix-based synthesis. This approach generates contextually rich information while being time efficient, and it facilitates continuous improvement and adaptation of intervention development [53].

The qualitative team did not use qualitative coding software, as the rapid matrix method was conducted using structured summary templates and team-based synthesis. The team conducted weekly meetings to review interview content, compare coding decisions and interpretations, resolve discrepancies through discussion, and refine the analytic approach. Using a rapid qualitative analysis process, each interviewer independently completed a standardized summary template following each interview (Multimedia Appendices 1 and 2), which included structured domains, initial content summaries, preliminary codes, and interpretive reflections [49]. These completed templates were compiled and reviewed by 2 researchers (NJ and JLM), who collaboratively synthesized responses across participants using a matrix method approach to identify key patterns and emerging themes. These themes were iteratively refined through weekly team discussions to compare summaries and interpretations and solve discrepancies to ensure consistency and representativeness across participants. Disagreements in coding were resolved through consensus-based discussion among the qualitative research team. A second review of the interviews for supporting quotes and thematic consistency was also carried out by the research team. Themes were finalized through iterative team discussions and comparison across structured matrix summaries to ensure consistency and alignment with participant responses. Throughout this process, analytic memos documenting observations and methodological reflections were completed after each interview and integrated into final thematic interpretation to enhance rigor and contextual grounding. Given the exploratory aims of the study, the sample size was considered sufficient to identify preliminary patterns in participant perceptions and inform future refinement of the TE-SPI. The data findings were considered sufficient to identify preliminary patterns relevant to intervention refinement. Furthermore, the analysis prioritized identification of salient patterns in participant feedback rather than quantification of theme prevalence.

The identified category themes were organized according to the TAM [41]. These themes helped focus the analysis, ensuring that key aspects of the data were systematically examined and aligned with the research questions. The research team defined and named the themes based on TAM (ie, perceived ease of use and perceived usefulness), then connected relevant quotes to these codes and organized them according to the model. The final step of the data analysis process was to produce a concise summary to address the research questions.

Results

Descriptive Analyses

Participants (N=20) in the study were between the ages 19 and 32 (mean 22.8, SD 3.9) years. The sample consisted predominantly of female participants (14/20, 70%). Participants were community college students (n=10) at ELAC and transfer students at UCLA (n=10). In terms of academic standing, 5% (1/20) of the participants were freshmen, 25% (5/20) were sophomores, 50% (10/20) were juniors, 5% (1/20) were seniors, and 1 participant preferred not to answer. For ethnicity, 55% (11/20) of participants identified as Hispanic or Latinx, and 45% (9/20) identified as non-Hispanic or Latinx; for race categories, 10% (2/20) identified as American Indian or Alaska Native, 25% (5/20) as Asian, 25% (5/20) as White, 20% (4/20) as multiracial, 10% (2/10) as other, and 2 students preferred not to answer the race category question. Data on participants' ethnicity and racial background were assessed separately. Participants who identified as Hispanic or Latinx may also be represented within the racial categories. Background information

XSL•FO
RenderX

was also collected, including previous treatment use and previous mental health app use. Half of the sample (10/20, 50%) reported that they had obtained counseling and therapy or seen a psychiatrist for their mental health in the past 12 months. Furthermore, participants (3/20, 15%) reported that they had used a digital tool for their mental health in the past. See Table 1 for additional demographic information.

Table . Participant demographics and characteristics (N=20)[a].

Variables	Values, n (%)
Age (years)	
19 - 25	16 (80)
26 - 32	4 (20)
Gender identity	
Male	5 (25)
Female	14 (70)
Nonbinary or third gender	1 (5)
Ethnicity	
Hispanic or Latine/x	11 (55)
Non-Hispanic or Latine/x	9 (45)
Race	
American Indian or Alaska Native	2 (10)
Asian	5 (25)
White	5 (25)
Other	2 (10)
Multiracial	4 (20)
Prefer not to answer	2 (10)
College standing	
Freshman	1 (5)
Sophomore	5 (25)
Junior	10 (50)
Senior	1 (5)
Prefer not to answer	3 (15)
University or community college student	
Transfer student	10 (50)
Community college	10 (50)
US born	
No	2 (10)
Yes	18 (90)
Mental health scores	
PHQ-2[b] (score of 3 and above)	4 (20)
GAD-2[c] (score of 3 and above)	8 (40)
Received mental health treatment	
Yes	10 (50)

[a]Ethnicity was assessed separately from race. Participants who identified as Hispanic/Latinx may also be represented within the racial categories.

[b]PHQ-2: Patient Health Questionnaire 2-Item.

[c]GAD-2: Generalized Anxiety Disorder 2-Item.

XSL•FO
RenderX

Regarding the mental health status of the participants in this study, data on self-reported symptoms indicated a mean composite score of 1.75 (SD 2.00) on the PHQ-2 (depression) and 2.45 (SD 1.82) on the GAD-2 (anxiety). A total of 4 (20%) participants were above the PHQ-2 screening threshold, suggesting possible elevated symptoms of depression and/or anhedonia, and 8 (40%) students scored above the GAD-2 threshold, indicating possible elevated anxiety symptoms. In total, 12 (60%) students scored above the screening threshold, indicating possible elevated symptoms for anxiety or depression. These results are presented to provide context regarding the sample's mental health profile and were not used to examine the differences in the perceptions of the intervention.

Perceived Ease of Use: BRITE App

Overview

This study examined initial user perceptions of the TE-SPI components using the TAM framework. Most participants perceived the BRITE app prototype as easy to interact with, with 75% (15/20) strongly agreeing and 5% (1/20) agreeing that they understood how to engage with it. Additionally, 85% (17/20) found the app features engaging (totally agreed: 13/20, 65% and agreed: 4/20, 20%). Potential motivation to use the app varied, with 50% (10/20) strongly agreeing, and 20% (4/20) agreeing that they would stay motivated to use it if they downloaded the app on their phone, while 15% (3/20) remained neutral, and 15% (3/20) disagreed. Regarding the app's feature set, 65% (13/20) strongly agreed, and 15% (3/20) agreed that it provided an adequate number of resources. See Table 2 for additional information on participant ratings.

Participants also rated their initial satisfaction with the app's ease of use: 35% (7/20) strongly agreed, 40% (8/20) somewhat agreed, and 20% (4/20) agreed that the app features were easy to use, while only 5% (1/20) remained neutral. Interaction with the app was well received, with 35% (7/20) strongly liking it, 40% (8/20) somewhat liking it, and 5% (1/20) disliking it. Regarding clarity, 60% (12/20) strongly agreed, and 30% (6/20) somewhat agreed that on-screen messages and activity links were clear. For distress and mood rating notifications, 45% (9/20) strongly agreed, 40% (8/20) somewhat agreed, and 10% (2/20) remained neutral about their clarity. Similarly, 70% (14/20) strongly agreed, and 20% (4/20) somewhat agreed that the personal coping plan information was clear, while 5% (1/20) found it unclear. See Table 3 to see more student feedback that was received on the app's perceived ease of use.

Table . Digital Intervention Barriers Scale (N=20)[a].

Question	Please indicate the extent to which you agree or disagree with the statements				
	Totally disagree, n (%)	Disagree, n (%)	Neutral, n (%)	Agree, n (%)	Totally agree, n (%)
I didn't understand the tasks or things I was supposed to do on the app.	15 (75)	1 (5)	—[b]	2 (10)	2 (10)
I thought the app wasn't engaging.	13 (65)	4 (20)	1 (5)	1 (5)	1 (5)
It would be difficult to keep myself motivated to use the app.	10 (50)	4 (20)	3 (15)	2 (10)	1 (5)
I thought the length of the app (ie, number of features) wasn't adequate (too long or too short).	13 (65)	3 (15)	3 (15)	—	1 (5)

[a]Three items from the original Digital Interventions Barriers Scale-7 were omitted, given that students were not asked to download the app during this phase of the study. The omitted items were (1) I had technical problems with my technology (eg, device, internet, and platform), (2) I forgot to use the app, and (3) The app didn't seem to be helping me.

[b]Not endorsed.

XSL•FO
RenderX

Table . Poststudy System Usability (N=20).

Question	Please rate the statement about your experiences with the app on a scale of 1 to 7 with 1 being strongly disagree and 7 being strongly agree.						
	Strongly disagree, n (%)	Somewhat disagree, n (%)	Disagree, n (%)	Neutral, n (%)	Agree, n (%)	Somewhat agree, n (%)	Totally agree, n (%)
Overall, I am satisfied with how easy it is to use the BRITE app.	—[a]	—	—	1 (5)	4 (20)	8 (40)	7 (35)
The information provided in the BRITE app, such as on-screen messages, and other documentation like links or pictures for the coping activities, was clear.	—	1 (5)	—	—	1 (5)	6 (30)	12 (60)
The information in the BRITE app, such as notification to rate my distress or mood with emojis, was clear.	—	—	—	2 (10)	1 (5)	8 (40)	9 (45)
I liked interacting with the BRITE app program.	—	—	1 (5)	3 (15)	1 (5)	8 (40)	7 (35)
I needed to learn a lot of things before I could get going with the BRITE app.	8 (40)	6 (30)	—	4 (20)	—	1 (5)	1 (5)
The information provided, such as the personal coping plan for the BRITE users, was clear.	—	1 (5)	—	—	1 (5)	4 (20)	14 (70)

[a]Not endorsed.

Look and Feel

During the semistructured interviews, participants were asked to provide qualitative feedback on the TE-SPI components. They shared their opinions on the BRITE app prototype aesthetic elements, such as the font, color scheme, and navigation (see Table 4for sample quotes and Multimedia Appendix 3 for images of the app prototype) as well as feedback on caring contact messages that they would find helpful and useful. The overall consensus we obtained based on student feedback was that the BRITE app was perceived as user-friendly, though some students expressed a desire for a broader range of emotions in the mood and distress rating scale. Some participants noted difficulty in identifying and connecting with the emotions represented by the tool's emojis. One student mentioned:

> *I think about the distress rating thing ... it's hard to quantify. If there were other ways expressing [the emoticon] ... or a number or a short sentence that you can put into categories to understand the displayed emotions.*

XSL•FO
RenderX

Table . Themes identified across domains[a].

Technology acceptance model components and subtheme	Sample quote	Participants mentioning related subtheme, n (%)
Perceived ease of use		
Look and feel	"I think the colors and the fonts are good ... they're really big ... And I like how it has like characters in it" [Transfer student]	16 (80)
Look and feel	"I think it was easy to navigate in general because you can just go on the menu, and then you can find exactly what you're looking for" [Transfer student]	11 (55)
Look and feel	"I think it's a fun app, it definitely feels like it could make an impact" [Transfer student].	2 (10)
Perceived accessibility	"If I don't have like a like a social support system that I can rely on at that moment, and if I feel like I need to somehow reframe my thoughts or ... just ease the mind a bit from like the negative stuff that's going on in my mind, I will use it personally" [Community college student].	1 (5)
Perceived accessibility	"I would use the [BRITE app] because it'll be more accessible and since I already carry my phone on a daily basis, I could use it while waiting for transportation, a class gap ..." [Transfer student].	6 (30)
Perceived usefulness		
Tool use in moments of distress	"Maybe just when I'm feeling down ... I can just go back to the app ... Also the stuff like breathing, sleeping, and exercising ... maybe it will just remind me to these things and to distress from everything" [Transfer student]	3 (15)
Tool use in moments of distress	"When I want to release my stress ... my emotions ... I want to tell someone how I feel ... no one wants to just look at a [digital tool] like when they're sad" [Community college student].	6 (30)
Tool use in moments of distress	"I think we believe that going to therapy is more than enough, but maybe this could be an additional tool to therapy ... You know, therapist aren't available 24 hours a day ... so this would be a helpful tool in those times" [Transfer student].	1 (5)
Interactive components	"I think about the distress rating thing ... it's hard to quantify. If there were other ways expressing the emoticons ... or a number or a short sentence that can put into categories to understand the displayed emotions" [Transfer student].	7 (35)
Interactive components	"I would definitely say ... It sounds like it would be helpful. Like I think the section where you can upload images and things that are important to you. I thought that it would be pretty useful" [Community college student, female, 25 years].	3 (15)
Interactive components	"Me personally, like I'm someone who uses my phone on dark mode, so to incorporate that it would be a big thing" [Community college student].	2 (10)

[a]Perspectives were included from a small number of participants (eg, 1 or 2), when they offered a unique perspective that added to the conceptualization of the overall theme.

To address this, several participants suggested changing the scale by incorporating numerical values or descriptive labels for the emoticons, which would help users more accurately select their current emotions. For example, one student shared:

Maybe adding more emotions besides these faces ... It doesn't look like there is a neutral face and I can't tell if one expression is crying or sleeping.

XSL•FO
RenderX

Participants also recommended additional features to enhance the app's functionality. These recommendations included a calendar and journal feature for tracking mood, distress, and customization options, such as adding an avatar. The app's lack of a dark mode, which is a visual display setting that changes the light background and dark text of an interface to a dark background with light text, was another point of concern. Participants expressed the importance of this feature for nighttime use to reduce eye strain. Their comments included:

> *I guess adding a dark mode, in case it's nighttime ... It won't sting my eyes and there are some people that get very sensitive to light.*

Another student noted: "Absolutely, these days, maybe like a dark mode interface could help." Overall, results on perceived ease of use highlight that this sample of students reported a high level of ease in using the BRITE app. They recommended that adding features like a dark mode and expanding the emotion rating scale would significantly enhance the tool's perceived ease of use.

Perceived Accessibility

Participants were also asked broadly about any sociocultural and developmentally related considerations of the BRITE app. In general, participants expressed that the overall feel of the tool felt culturally congruent and relevant to the college student context. Participants felt that young adult users could easily connect and relate to the content in the app. College students expressed that the language was straightforward and easy to understand; however, a few participants recommended adding a feature that allows users to translate the language in the app to the language they prefer to speak so that it could be more inclusive, especially for international students. Concerning this topic, a participant noted the following:

> *Maybe you could in the future ... translate the different words ... so that more people can feel inclined to use it. Maybe they don't speak English, you know ... so they want to have it in their own language ... So in the beginning, when you first get on the app maybe ask, "Please select your language."*

In addition to language, some participants discussed that the BRITE app was appropriate for college students, as it was easily accessible. A participant shared the following: "I would use the[app] because it'll be more accessible and since I already carry my phone on a daily basis, I could use it while waiting for transportation, a class gap ..."

The BRITE app's accessibility and convenience appeared to be particularly appealing to this student population. Participants shared examples of how the app could fit naturally into their college life routine, making it easy to use during gaps of time, such as waiting for transportation or between classes. This convenience factor highlights a significant advantage of mental health mobile apps, as they can be well integrated into daily habits without requiring students to set aside much dedicated time. It is possible that the app can increase use frequency by turning typically unproductive periods into opportunities for digital mental health content interaction. Overall, this participant feedback emphasized the potential advantages of digital tools available on phones as an avenue to promote mental health engagement among community college and transfer students.

Perceived Usefulness

Overview

Participants completed a usability questionnaire that provided valuable insights into their perspectives on the perceived usefulness of the TE-SPI components. In terms of use during moments of distress, half of the participants (10/20, 50%) indicated that they would definitely use the app if they felt distressed or upset. When asked if they would recommend the BRITE app to a friend needing mental health care for depression or more severe distress (ie, having thoughts of suicide), 50% (10/20) responded, "yes, I think so," and 40% (8/20) responded "yes, definitely." Regarding potential use in the first month of availability, 45% (9/20) of participants stated they would use the app 3 to 4 times, 35% (7/20) would plan to use it 5 to 10 times, and 15% (3/20) would expect to use it more than 10 times. When participants were asked if they would engage more with the BRITE app if they received caring contact messages, 55% (11/20) said "yes," 25% (5/20) said "possibly or maybe," and 20% (4/20) stated that the messages would not increase their overall motivation to use the app. When asked to elaborate on their ratings, one student mentioned:

> *We're are always connected to our phones, and at this age it is basically an extension of ourselves. I think reminders [text messages] would be helpful because notifications are a good way to get more engagement with the app.*

Another participant said: "If I was reminded, then I would have an inclination to return [to the app] since it gave me a reminder to go on," while another stated: "I tend to forget things, it [text messages] would help to keep on track." During the qualitative interviews, participants were also asked to elaborate on their perceptions of caring contacts by providing suggestions of caring contact messages that they would find useful (see Table S1 in Multimedia Appendix 4 for additional suggestions provided by participants). Overall, the survey feedback from students suggests a varied perception of the TE-SPI's perceived usefulness, with a clear indication that caring contact messages could enhance user engagement, although with some reservations regarding the frequency and content of the caring contact messages. See Table 5 to see more students' input on the perceived usefulness of the app.

XSL•FO
RenderX

Table . Satisfaction or qualitative questions (N=20).

Questions	Values, n (%)
If you start to feel really stressed and upset, do you think you would use the BRITE app?	
No	1 (5)
Possibly	4 (20)
Probably	5 (25)
Definitely	10 (50)
Prefer not to answer	—[a]
If a friend were in need of mental health care for depression or suicidal thoughts or behaviors, would you recommend BRITE to help them cope?	
Definitely not	—
No, I do not think so	1 (5)
Yes, I think so	10 (50)
Yes, definitely	8 (40)
Prefer not to answer	1 (5)
Do you think you would be more likely to use the BRITE app if you received messages reminding you of the app?	
No, I would find these messages annoying	4 (20)
No, but the messages would not bother me	—
Maybe	2 (10)
Possibly	3 (15)
Yes	11 (55)
How often do you think you would use the BRITE app if it were available to you during the month after it was introduced to you?	
Never	1 (5)
Once or twice	—
3 or 4 times	9 (45)
5 or 10 times	7 (35)
More than 10 times	3 (15)

[a]Not endorsed.

Tool Use in Moments of Distress

Participants offered insights into the usability of the app, particularly highlighting its potential as a support mechanism during times of emotional distress. One participant explained: “When I am at the point of you know getting down emotionally, I’d want to use it right before getting to that point.” Another student shared how there can be distressful moments in life and how the TE-SPI could help her get distracted from unhelpful and negative thoughts:

> *Sometimes there are moments where I need to de-stress or ... I need to distract myself from one specific thought or reoccurring thoughts, like flashbacks from the past and stuff ... But it [the BRITE app] would be a good distraction for me ... to distract my negative thoughts to something maybe useful.*

Another theme to emerge involved perceptions that the BRITE app can serve to augment and/or replace traditional treatment contacts and/or other forms of social support. For example, one participant noted:

> *I think we believe that going to therapy is more than enough, but maybe this could be an additional tool to therapy ... You know, therapist aren’t available 24 hours a day ... so this would be a helpful tool in those times.*

Another student stated:

> *If I don’t have like a social support system that I can rely on at that moment, and if I feel like I need to somehow reframe my thoughts or ... just ease the mind a bit from like the negative stuff that’s going on in my mind, I will use it personally.*

However, other students felt differently. Some stated that they would not want to engage with the tool if they were having low mood. One student stated:

> *When I want to release my stress ... my emotions ... I want to tell someone how I feel ... no one wants to just look at a [digital tool] like when they’re sad.*

Another participant explained that they would not choose to use the app during joyful moments, as it would be a reminder of times when they weren’t happy, stating: “I guess I wouldn’t

XSL•FO
RenderX

want to use it when I'm happy because it would remind me of all the times, I was not happy."

Overall, students highlighted various reasons as to when they would feel more inclined to use a digital mental health tool, with some students reporting that they would use it during moments of distress, while some reporting that they would avoid the tool when their mood and levels of distress improved. The consensus among participants suggests that they would use the tool sporadically, primarily as a tool for managing negative emotions. To exemplify this topic, a participant stated the following:

> *I could see myself using it once a week or so. I mean like I said, maybe I should be doing more and more daily mental health check-ins, but generally, I do like self-help when I'm like feeling bad. So, when I'm feeling good, I don't know if I would need it.*

The student feedback suggests that the TE-SPI components, including the BRITE app, may be a valuable tool when traditional social support is unavailable, as it helps reframe negative and unhelpful thoughts. However, the tool's utility appears to be situational, with students reporting that they would be less inclined to use it when they are in a positive state.

Interactive Components

Participants expressed a willingness to use the app to explore and engage with suggested coping activities such as practicing mindfulness, including an affective savoring the moment activity and breathing exercises, as well as tracking their mood and distress. A participant noted the following: "Yeah, I think this could be good, like a daily check-in and like resources here but I think I would be more inclined to use it if I could customize it more." Another participant highlighted the significance of the activities, stating:

> *I think breathing exercises would help a lot. And then ... maybe the meditation part. I think the activities would probably be the main reason why I use it [the BRITE app] and the mood tracker. I think tracking your mood is really important.*

Overall, most students reported that they would use the tool primarily to practice coping skills to reduce distress, with a special interest in activities focused on meditation or mindfulness and improving sleep quality.

Attitude Toward TE-SPI and Recommendations

During the qualitative interviews, participants were asked what would motivate them to use the app and if they would incorporate the tool into their daily lives. Many students answered that incentives and interactive games would further motivate them to engage with the digital tool. Furthermore, although the app currently includes badges that are administered based on participation, a few students suggested that incentives would be a great way to motivate and reinforce users to continuously use the tool since it could promote a sense of accomplishment and reward. For instance, a participant suggested that users could accumulate rewards over time based on tool use: "I think you would need something to get them [users] to use it more often ... gift cards or little things here and there ... maybe if they use the [BRITE app] this many times and it adds up over time, this is the amount they get."

Similarly, another participant shared:

> *I know people tend to get bored of apps. Unless it's like a social media app ... Maybe you could add games ... Maybe you could have a leaderboard or something.*

Some participants suggested the value of adding a live chat and journaling feature that could increase usability. For example, a participant stated:

> *I like to journal a lot. I do journal on my iPad, and I do write on my phone in the notes because I have everything connected, so can I just go back and forth. But that's what would motivate me to use it.*

Others expressed that a live chat feature for crisis support would be great for users, especially if they are unable to get a hold of their mental health provider, and the live chat feature could help them to connect to a crisis hotline volunteer who could help users in times of distress. An example of this feedback is: "... if you guys do end up adding the live chat, I think it will be a good app ... either just to talk to someone ... or ask for advice ..." Overall, the participants indicated that the integration of incentives and interactive features (eg, games, journaling, and live chat) would enhance their motivation to use the app.

Discussion

Principal Findings

This study contributes to the growing literature on digital mental health interventions by conducting a mixed methods evaluation of college students' perceptions of a novel, multicomponent, TE-SPI, intended for diverse college students. Unlike many mobile mental health apps that offer only standalone self-monitoring or psychoeducation, the BRITE app prototype was designed to integrate safety planning, personalized coping tools, and mood tracking, features rarely combined in a single app or digital tool, particularly for underserved student populations. A key distinction of the TE-SPI model is the integration of both the BRITE app with caring contact text messages, adding a "human touch" that may enhance perceived engagement and relevance. Using the TAM framework with the overarching aim of examining perceived usability and utility of the TE-SPI components and a qualitative rapid coding analysis methodology, our study identified salient themes related to student perceptions and preferences for optimal design features for community college students. Our study intentionally selected a diverse sample of community college students with varying levels of anxiety and depression to focus on those at elevated risk for developing future suicidal thoughts and behaviors, allowing us to examine how students at elevated risk perceived these resources during the prototype demonstration. Although participants reported possible elevated symptoms of depression and anxiety, this study was not designed or powered to examine whether symptom severity influenced perceptions of the intervention. Therefore, future research should explore whether differences in depression and anxiety symptoms are

XSL•FO
RenderX

associated with actual use, usability, and perceived usefulness of digital mental health tools.

Our findings align with prior work, suggesting that students value apps that are accessible, low-burden, and offer features such as mood tracking, skill reminders, and personalization [55]. Participants indicated a perceived willingness to use a tool like the BRITE app, particularly during times of stress, for activities such as mindfulness or meditation and mood tracking. However, they also noted that their anticipated use of this particular app would likely vary depending on their individual circumstances, such as being too busy with other responsibilities. This underscores the importance of designing features that align with users' daily needs and preferences, as well as considering strategies that may support sustained use in future real-world testing. The feedback from students regarding the app's look and feel emphasized the importance of user-friendliness and a desire for customization features, underscoring the observation that different students are likely to use the app differently and desire different features, including alternate approaches to presenting the mood monitoring descriptors, personalization features (eg, calendar, journaling, and dark mode options), interactive games to motivate app use, and live chat functionality that could link students to other resources such as the crisis hotline. Satisfaction ratings with the BRITE app components were high, as previously found in a sample of hospitalized suicidal adolescents [25]. Still, ongoing evaluation of user engagement metrics and satisfaction surveys will be important for future optimizations to ensure that the app remains relevant and effective in supporting the mental health of college students. Because participants in this study did not use the app over time, these ratings should be interpreted as reflecting initial impressions of the prototype rather than sustained engagement or real-world use.

Participants emphasized the importance of culturally responsive content and ease of integration into daily routines. Notably, the caring contacts were especially well received, echoing prior research suggesting that students benefit from brief, nonintrusive messages that convey connection and care [56]. Importantly, participants responded to example or proposed caring contact messages rather than receiving caring contacts as part of the intervention. Underscoring the value of a "human touch," participants felt it would be important to ensure that these messages felt authentic and human-driven, rather than automated or impersonal, to foster a sense of connection and trust among users. In addition, participants provided examples of caring contact messages that could be implemented in future caring contact interventions with this population. Participants' feedback regarding language translation options reflects the potential need to accommodate diverse linguistic backgrounds and highlights the importance of accessibility and inclusivity. Additionally, incorporating culturally and developmentally relevant content and resources could further enhance user engagement and relevance for diverse college students. It could also promote inclusivity within the app for ensuring that users feel both represented and supported in this digital space.

Limitations and Future Directions

While this study offers formative insights into the initial perceptions of 2 of the TE-SPI components (BRITE app and caring contacts) among ethnoracially diverse college students, several limitations should be noted. First, although the sample was diverse and sufficient to provide contextually rich preliminary feedback, it was relatively small and drawn from one segment of the transfer student or community college student population, which may limit generalizability to broader student populations. Second, participants did not download the BRITE app and did not actually receive the caring contact text messages over time as part of the study; rather, they interacted with a guided demonstration and provided feedback based on brief exposure to a prototype. Accordingly, findings reflect initial perceptions of the proposed intervention components rather than real-world feasibility, engagement, acceptability, or clinical benefit. This is particularly important, given prior findings that show app use tends to decline over time, even among students who initially express interest [56]. Additionally, the quantitative assessment in this study focused primarily on the BRITE app, with feedback on caring contacts derived largely from the qualitative interview data and only a single survey item. Therefore, conclusions regarding caring contacts are especially preliminary, and future research should evaluate this component using more robust quantitative approaches. Additional research is also needed to clarify student reactions to the clinician-guided coping or safety planning session during which the app was populated, and perceptions of this intervention component designed to offer a "human touch." Third, the study did not assess longitudinal app use or how mental health symptoms impacted perceptions of the intervention. Without data on repeated use, in-the-moment engagement, or clinical impact, conclusions about the intervention's effectiveness and acceptability in the real-world remain limited. This study also was not designed to estimate intervention effects or to provide effect size estimates for powering a future effectiveness trial. Instead, its primary contribution is formative and allowed for the identification of user-perceived barriers, facilitators, and priorities for refinement that can guide the next implementation phase. Incorporating ecological momentary assessment, via daily mood ratings in the app, and use analytics will be key to understanding how students engage with different features over time [57] and can help evaluate whether sustained use reduces suicide risk or improves mental health functioning among college students. Finally, while participants offered actionable recommendations to enhance personalization (eg, language translation, gamification, and live chat) during the qualitative interviews, these features were not part of the current prototype. Iterative development and testing of these enhancements may increase user engagement and cultural relevance, particularly for students from historically underserved backgrounds. As prior research has shown, users' trust and sustained engagement are influenced not only by app content and functionality but also by transparency about data use, perceived credibility, and institutional endorsement [58]. Future work should focus on establishing trust with participants on how data collection is stored to reinforce transparency and ethical standards, ultimately enhancing intervention relevance, acceptability, and long-term

XSL•FO
RenderX

impact on student mental health. Importantly, the qualitative findings highlight specific design and implementation issues, such as preference for customization features on the BRITE app or authentic caring contact text messages, that can be incorporated into the next iteration of the TE-SPI and evaluated prospectively in a more rigorous trial. More broadly, future studies should move beyond prototype feedback to test the intervention as delivered, including whether students download the app, engage with it over time, receive caring contact messages, and find the full provider-supported intervention acceptable and useful in real-world settings.

Conclusions

This study provides preliminary feedback on a TE-SPI featuring the BRITE app—a mobile app prototype supporting coping skills use and proposed caring contact text messages among a sample of college students, who often have high unmet mental health needs and low access to treatments. In line with recent recommendations [15], TE-SPI was designed to be low-burden, contextually relevant, and responsive to the specific challenges faced by structurally marginalized students, including limited access to culturally competent care and high academic stress. The high satisfaction and perceived usability reported in response to the prototype demonstration suggest that this approach may warrant further refinement and testing as a digital intervention that could complement traditional mental health services by providing immediate, accessible, and scalable support. Given the persistent challenges in accessing mental health care on college campuses, integrating digital approaches such as the study TE-SPI into campus service infrastructures—such as counseling centers, health centers, peer support programs, and crisis response services—may be worth exploring once the intervention has been tested under real-world conditions. Moreover, the TE-SPI was designed with considerations for cultural responsiveness, accessibility, and implementation in low-resource settings [59]. Future research is needed to test the TE-SPI in a larger, more diverse student population and to evaluate actual use, acceptability, feasibility, and its impact on risk for suicidal and self-harm thoughts and behaviors. Collaboration with universities, mental health professionals, and technology developers will be key in further developing the TE-SPI approach and scaling digital interventions for broad adoption. By leveraging digital solutions like TE-SPI, mental health care providers can extend their reach beyond clinical settings, offering timely and evidence-based support to students when and where they need it most. If effectively integrated into real-world mental health services, such approaches have the potential to become a critical component in improving mental well-being and reducing suicide risk in vulnerable college populations.

Acknowledgments

The authors would like to thank all the students who participated in their study and provided feedback on their technology-enhanced suicide prevention intervention tool. The authors also thank the developers of the BRITE app, including DB and Dr Beth Kennard, for graciously sharing their work and allowing their team to consider adaptations to serve our college student demographic. No generative AI tools were used in the writing, analysis, or preparation of this manuscript.

Funding

Funding support to Jocelyn Meza, Joan Asarnow, and Juliane Martinez came from Project 3 (project leads Jocelyn Meza and Joan Asarnow), a substudy of the ALACRITY Center Grant (P50MH126337-01; awarded Michelle G Craske and Kate Wolitzky-Taylor [multiple principal investigators]). Funding for the original BRITE app came from the ETUDES Center at the University of Pittsburgh (P50MH115838-02; awarded to David Brent [principal investigator]).

Data Availability

The dataset generated and analyzed in this study is not publicly available due to the sensitive nature of the data and the need to protect participant confidentiality. Data may be made available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: JIM, JRA
Methodology: JIM, JRA, NJ, JLM
Formal analysis: JIM, NJ, JLM
Validation: JIM, CB, BG-M, DB
Funding acquisition: JIM, JRA
Visualization: NJ, JLM
Writing—original draft: JIM, NJ, JLM
Writing—review and editing: JIM, JRA, NJ, JLM, CB, BG-M, DB

Conflicts of Interest

JRA receives grant support from the National Institute of Mental Health and the Patient-Centered Outcome Research Institute. She has served as a consultant on quality improvement for depression and suicide or self-harm prevention, and is currently on the Scientific Advisory Board of the Klingenstein Third Generation Foundation. DB has intellectual property licensed to the

XSL•FO
RenderX

University of Pittsburgh in the BRITE app. Ksana Health is developing a related app (in which we do not have intellectual property) based in part on BRITE, but we have a Small Business Innovation Research grant to do so, led by Ksana Health. The work on this paper was in part supported by NIMH P50 1158383.

Multimedia Appendix 1
Technology-enhanced suicide prevention intervention semistructured interview questions.
[DOCX File, 22 KB - humanfactors_v13i1e79537_app1.docx]

Multimedia Appendix 2
Technology-enhanced suicide prevention intervention interview rapid analysis coding template.
[DOCX File, 17 KB - humanfactors_v13i1e79537_app2.docx]

Multimedia Appendix 3
Distress mood rating and activity recommendation on the BRITE app.
[DOCX File, 492 KB - humanfactors_v13i1e79537_app3.docx]

Multimedia Appendix 4
Additional caring contacts text messages suggested by participants.
[DOCX File, 16 KB - humanfactors_v13i1e79537_app4.docx]

References

1. Web-based Injury Statistics Query and Reporting System (WISQARS). Centers for Disease Control and Prevention. 2020. URL: https://wisqars.cdc.gov/ [accessed 2024-09-13]
2. Stewart C. Past year suicidal and self-injurious behavior among U.S. college students 2024-2025. Statista. 2026. URL: https://www.statista.com/statistics/1126285/college-students-with-suicidal-and-self -injurious-behavior-us/ [accessed 2023-11-16]
3. GBD 2019 Disease and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2019 Nov;396(10262). [doi: 10.1016/S0140-6736(20)32226-1]
4. Arakelyan M, Freyleue S, Avula D, McLaren JL, O'Malley AJ, Leyenaar JK. Pediatric mental health hospitalizations at acute care hospitals in the US, 2009-2019. JAMA 2023 Mar 28;329(12):1000-1011. [doi: 10.1001/jama.2023.1992] [Medline: 36976279]
5. Martínez-Alés G, Jiang T, Keyes KM, Gradus JL. The recent rise of suicide mortality in the United States. Annu Rev Public Health 2022 Apr 5;43(1):99-116. [doi: 10.1146/annurev-publhealth-051920-123206] [Medline: 34705474]
6. Casey SM, Varela A, Marriott JP, Coleman CM, Harlow BL. The influence of diagnosed mental health conditions and symptoms of depression and/or anxiety on suicide ideation, plan, and attempt among college students: findings from the Healthy Minds Study, 2018-2019. J Affect Disord 2022 Feb 1;298(Pt A):464-471. [doi: 10.1016/j.jad.2021.11.006] [Medline: 34774646]
7. Moffett S, Javdani S, Miglin R, Sadeh N. Examining latent profiles of psychopathy in a mixed-gender sample of juvenile detainees. Personal Disord 2020 Jul;11(4):290-299. [doi: 10.1037/per0000370]
8. Busby DR, Horwitz AG, Zheng K, et al. Suicide risk among gender and sexual minority college students: the roles of victimization, discrimination, connectedness, and identity affirmation. J Psychiatr Res 2020 Feb;121:182-188. [doi: 10.1016/j.jpsychires.2019.11.013] [Medline: 31837538]
9. Katz DS, Davison K. Community college student mental health: a comparative analysis. Community Coll Rev 2014 Oct;42(4):307-326. [doi: 10.1177/0091552114535466]
10. Eisenberg D, Lipson SK, Posselt J. Promoting resilience, retention, and mental health. New Drctns for Student Srvcs 2016 Dec;2016(156):87-95. [doi: 10.1002/ss.20194]
11. Lipson SK, Zhou S, Abelson S, et al. Trends in college student mental health and help-seeking by race/ethnicity: findings from the National Healthy Minds Study, 2013-2021. J Affect Disord 2022 Jun 1;306:138-147. [doi: 10.1016/j.jad.2022.03.038] [Medline: 35307411]
12. Fry R, Cilluffo A. A rising share of undergraduates are from poor families, especially at less selective colleges. : Pew Research Center; 2019 URL: http://hdl.handle.net/10919/92670 [accessed 2026-08-30]
13. Lipson SK, Kern A, Eisenberg D, Breland-Noble AM. Mental health disparities among college students of color. J Adolesc Health 2018 Sep;63(3):348-356. [doi: 10.1016/j.jadohealth.2018.04.014] [Medline: 30237000]
14. Eustis EH, LoPresti J, Aguilera A, Schueller SM. Cultural responsivity in technology-enabled services: integrating culture into technology and service components. J Med Internet Res 2023 Oct 3;25:e45409. [doi: 10.2196/45409] [Medline: 37788050]

XSL•FO
RenderX

15. Ramos G, Hernandez-Ramos R, Taylor M, Schueller SM. State of the science: using digital mental health interventions to extend the impact of psychological services. Behav Ther 2024 Nov;55(6):1364-1379. [doi: 10.1016/j.beth.2024.04.004] [Medline: 39443071]
16. Xu D, Jaggars SS, Fletcher J, Fink JE. Are community college transfer students “a Good Bet” for 4-year admissions? Comparing academic and labor-market outcomes between transfer and native 4-year college students. J Higher Educ 2018 Jul 4;89(4):478-502. [doi: 10.1080/00221546.2018.1434280]
17. Daddona MF, Mondie-Milner C, Goodson J. Transfer student resources: keeping students once they enroll. J Coll Stud Ret 2021 Nov;23(3):487-506. [doi: 10.1177/1521025119848754]
18. Mehr KE, Daltry R. Examining mental health differences between transfer and nontransfer university students seeking counseling services. J College Stud Psychother 2016 Apr 2;30(2):146-155. [doi: 10.1080/87568225.2016.1140996]
19. Broton KM, Mohebali M, Lingo MD. Basic needs insecurity and mental health: community college students’ dual challenges and use of social support. Community Coll Rev 2022 Oct;50(4):456-482. [doi: 10.1177/00915521221111460]
20. Lipson SK, Phillips MV, Winquist N, Eisenberg D, Lattie EG. Mental health conditions among community college students: a national study of prevalence and use of treatment services. PS (Wash DC) 2021 Oct 1;72(10):1126-1133. [doi: 10.1176/appi.ps.202000437]
21. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869. [doi: 10.2196/12869] [Medline: 31333198]
22. Lattie EG, Lipson SK, Eisenberg D. Technology and college student mental health: challenges and opportunities. Front Psychiatry 2019;10:246. [doi: 10.3389/fpsyt.2019.00246] [Medline: 31037061]
23. Schueller SM, Hunter JF, Figueroa C, Aguilera A. Use of digital mental health for marginalized and underserved populations. Curr Treat Options Psych 2019 Sep;6(3):243-255. [doi: 10.1007/s40501-019-00181-z]
24. Kodish TA. Digital Mental Health Interventions for College Students of Color: Understanding Uptake and Enhancing Engagement [UCLA electronic theses and dissertations]. : eScholorship Publishing URL: https://escholarship.org/uc/item/7p773792 [accessed 2026-08-20]
25. Kennard BD, Goldstein T, Foxwell AA, et al. As Safe as Possible (ASAP): a brief app-supported inpatient intervention to prevent postdischarge suicidal behavior in hospitalized, suicidal adolescents. AJP 2018 Sep 1;175(9):864-872. [doi: 10.1176/appi.ajp.2018.17101151]
26. Stallard P, Porter J, Grist R. A smartphone app (BlueIce) for young people who self-harm: open phase 1 pre-post trial. JMIR Mhealth Uhealth 2018 Jan 30;6(1):e32. [doi: 10.2196/mhealth.8917] [Medline: 29382632]
27. Tighe J, Shand F, McKay K, Mcalister TJ, Mackinnon A, Christensen H. Usage and acceptability of the iBobbly app: pilot trial for suicide prevention in Aboriginal and Torres Strait Islander youth. JMIR Ment Health 2020 Dec 1;7(12):e14296. [doi: 10.2196/14296] [Medline: 33258782]
28. Knowles SE, Toms G, Sanders C, et al. Qualitative meta-synthesis of user experience of computerised therapy for depression and anxiety. PLoS One 2014;9(1):e84323. [doi: 10.1371/journal.pone.0084323] [Medline: 24465404]
29. Li J, Ning P, Cheng P, et al. Factors associated with dropout of participants in an app-based child injury prevention study: secondary data analysis of a cluster randomized controlled trial. J Med Internet Res 2021 Jan 29;23(1):e21636. [doi: 10.2196/21636] [Medline: 33512318]
30. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
31. Baskerville NB, Struik LL, Guindon GE, et al. Effect of a mobile phone intervention on quitting smoking in a young adult population of smokers: randomized controlled trial. JMIR Mhealth Uhealth 2018 Oct 23;6(10):e10893. [doi: 10.2196/10893] [Medline: 30355563]
32. Zhao Q, Arévalo SP, O’Brien J, Li W. Patterns of interpersonal violence experience and their association with mental health and substance use problems among women with incarceration histories: a multidimensional longitudinal study. J Fam Viol 2023 Nov;38(8):1639-1654. [doi: 10.1007/s10896-022-00453-y]
33. Kennard BD, Biernesser C, Wolfe KL, et al. Developing a brief suicide prevention intervention and mobile phone application: a qualitative report. J Technol Hum Serv 2015 Oct 1;33(4):345-357. [doi: 10.1080/15228835.2015.1106384] [Medline: 26977137]
34. Goldstein TR, Kennard BD, Porta G, et al. Bridging gaps in care following hospitalization for suicidal adolescents: As Safe As Possible (ASAP) and BRITE app. J Am Acad Child Adolesc Psychiatry 2025 May;64(5):612-624. [doi: 10.1016/j.jaac.2024.06.008] [Medline: 39032815]
35. Bero KM, Teresi GI, Porta G, Poling K, Brent DA, Goldstein TR. Treatment outcomes of an adolescent intensive outpatient program for depressed and suicidal youth. JAACAP Open 2025 Sep;3(3):576-588. [doi: 10.1016/j.jaacop.2024.10.006] [Medline: 40922758]
36. Motto JA. Suicide prevention for high-risk persons who refuse treatment. Suicide Life Threat Behav 1976;6(4):223-230. [doi: 10.1111/j.1943-278X.1976.tb00880.x] [Medline: 1023455]
37. Skopp NA, Smolenski DJ, Bush NE, et al. Caring contacts for suicide prevention: a systematic review and meta-analysis. Psychol Serv 2023 Feb;20(1):74-83. [doi: 10.1037/ser0000645] [Medline: 35420858]

XSL•FO
RenderX

38. Radin AK, Shaw J, Brown SP, et al. Comparative effectiveness of two versions of a caring contacts intervention in healthcare providers, staff, and patients for reducing loneliness and mental distress: a randomized controlled trial. J Affect Disord 2023 Jun 15;331:442-451. [doi: 10.1016/j.jad.2023.03.029] [Medline: 36963515]
39. Hebert LE, Fruhbauerova M, Evanson A, et al. Caring Texts, a strength-based, suicide prevention trial in 5 native communities: research design and methods. Contemp Clin Trials 2022 Dec;123:106966. [doi: 10.1016/j.cct.2022.106966] [Medline: 36252937]
40. Doupnik SK, Rudd B, Schmutte T, et al. Association of suicide prevention interventions with subsequent suicide attempts, linkage to follow-up care, and depression symptoms for acute care settings: a systematic review and meta-analysis. JAMA Psychiatry 2020 Oct 1;77(10):1021-1030. [doi: 10.1001/jamapsychiatry.2020.1586] [Medline: 32584936]
41. Mitchell KM, Holtz BE, McCarroll AM. Assessing college students' perceptions of and intentions to use a mobile app for mental health. Telemed J E Health 2022 Apr;28(4):566-574. [doi: 10.1089/tmj.2021.0106] [Medline: 34166099]
42. Davis FD. Technology acceptance model: TAM. In: Al-Suqri MN, Al-Aufi AS, editors. Information Seeking Behavior and Technology Adoption: IGI Global; 1989:205-219.
43. Masrom M. Technology acceptance model and e-learning. Presented at: 12th International Conference on Education; May 21-24, 2007.
44. Cohen A, Soffer T, Henderson M. Students' use of technology and their perceptions of its usefulness in higher education: international comparison. Comput Assist Learn 2022 Oct;38(5):1321-1331. [doi: 10.1111/jcal.12678]
45. STAND: screening & treatment for anxiety & depression. UCLA Depression Grand Challenge | STAND. URL: https://stand.ucla.edu/ [accessed 2026-08-30]
46. Motto JA, Bostrom AG. A randomized controlled trial of postcrisis suicide prevention. PS (Wash DC) 2001 Jun;52(6):828-833. [doi: 10.1176/appi.ps.52.6.828]
47. Stewart KL, Darling EV, Yen S, Stanley B, Brown GK, Weinstock LM. Dissemination of the safety planning intervention (SPI) to university counseling center clinicians to reduce suicide risk among college students. Arch Suicide Res 2020;24(sup1):75-85. [doi: 10.1080/13811118.2018.1531797] [Medline: 30300102]
48. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003 Nov;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
49. Kroenke K, Spitzer RL, Williams JBW, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007 Mar 6;146(5):317-325. [doi: 10.7326/0003-4819-146-5-200703060-00004] [Medline: 17339617]
50. Staples LG, Dear BF, Gandy M, et al. Psychometric properties and clinical utility of brief measures of depression, anxiety, and general distress: the PHQ-2, GAD-2, and K-6. Gen Hosp Psychiatry 2019;56:13-18. [doi: 10.1016/j.genhosppsych.2018.11.003] [Medline: 30508772]
51. Lewis JR. The System Usability Scale: past, present, and future. Int J Hum-Comput Interact 2018 Jul 3;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
52. Ramos G, Montoya AK, Hammons HR, Smith D, Chavira DA, Rith-Najarian LR. Digital Intervention Barriers Scale-7 (DIBS-7): development, evaluation, and preliminary validation. JMIR Form Res 2023 Apr 6;7:e40509. [doi: 10.2196/40509] [Medline: 37023417]
53. Hamilton AB, Finley EP. Qualitative methods in implementation research: an introduction. Psychiatry Res 2019 Oct;280:112516. [doi: 10.1016/j.psychres.2019.112516] [Medline: 31437661]
54. Vlachogianni P, Tselios N. Perceived usability evaluation of educational technology using the Post-Study System Usability Questionnaire (PSSUQ): a systematic review. Sustainability 2023 Aug 28;15(17):12954. [doi: 10.3390/su151712954]
55. Nevedal AL, Reardon CM, Opra Widerquist MA, et al. Rapid versus traditional qualitative analysis using the Consolidated Framework for Implementation Research (CFIR). Implement Sci 2021 Jul 2;16(1):67. [doi: 10.1186/s13012-021-01111-5] [Medline: 34215286]
56. Schueller SM, Neary M, Lai J, Epstein DA. Understanding people's use of and perspectives on mood-tracking apps: interview study. JMIR Ment Health 2021 Aug 11;8(8):e29368. [doi: 10.2196/29368] [Medline: 34383678]
57. Bautista J, Schueller SM. Understanding the adoption and use of digital mental health apps among college students: secondary analysis of a national survey. JMIR Ment Health 2023 Mar 22;10:e43942. [doi: 10.2196/43942] [Medline: 36947115]
58. Ryan K, Hogg J, Kasun M, Kim JP. Users' perceptions and trust in AI in direct-to-consumer mHealth: qualitative interview study. JMIR Mhealth Uhealth 2025 May 20;13:e64715. [doi: 10.2196/64715] [Medline: 40392584]
59. Joo E, Kononova A, Kanthawala S, Peng W, Cotten S. Smartphone users' persuasion knowledge in the context of consumer mHealth apps: qualitative study. JMIR Mhealth Uhealth 2021 Apr 13;9(4):e16518. [doi: 10.2196/16518] [Medline: 33847596]

Abbreviations

ELAC: East Los Angeles Community College
GAD-2: Generalized Anxiety Disorder 2-Item
PHQ-2: Patient Health Questionnaire 2-Item
SITB: self-injurious thought and behavior

XSL•FO
RenderX

STAND: Screening and Treatment for Anxiety & Depression
TAM: technology acceptance model
TE-SPI: technology-enhanced suicide prevention intervention
UCLA: University of California, Los Angeles

Edited by J Sarvestan; submitted 13.Aug.2025; peer-reviewed by A AL-Asadi, A Alshaher, M Pulier; revised version received 13.May.2026; accepted 25.May.2026; published 09.Sep.2026.

Please cite as:
Meza JI, Asarnow JR, Jaramillo N, Martinez JL, Biernesser C, George-Milford B, Brent D
Preliminary Support for a Technology-Enhanced Suicide Prevention Intervention for College Students: Mixed Methods Preimplementation Study
JMIR Hum Factors 2026;13:e79537
URL: https://humanfactors.jmir.org/2026/1/e79537
doi:10.2196/79537

© Jocelyn I Meza, Joan R Asarnow, Natalia Jaramillo, Juliane L Martinez, Candice Biernesser, Brandie George-Milford, David Brent. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 9.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Integrating Wearable Technology in Digital Therapy for People With Psychosis (SloMo): Mixed Methods Co-Design and Feasibility Study

Kathryn M Taylor[1,2], DClinPsy; Thomas Ward[1,2], PhD, DClinPsy; Matteo Cella[1,2], PhD, DClinPsy; Philippa Garety[1], CPsychol, PhD; Amy Grant[1,2]; Katie Philps[3], MSc; Anna Wojdecka[4], PhD; Melanie Flory[4], PhD; Amy Hardy[1,2], PhD, DClinPsy

[1]Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King's College London, Henry Wellcome Building, IoPPN, 16 De Crespigny Park, London, United Kingdom

[2]South London and Maudsley NHS Foundation Trust, London, United Kingdom

[3]Research and Development, Sussex Partnership NHS Foundation Trust, Sussex, United Kingdom

[4]Helen Hamlyn Centre for Design, Royal College of Art, London, United Kingdom

Corresponding Author:
Kathryn M Taylor, DClinPsy
Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King's College London, Henry Wellcome Building, IoPPN, 16 De Crespigny Park, London, United Kingdom

Abstract

Background: Digital health interventions for psychosis, like SloMo, leverage smartphone technology to help transfer learning from therapy to real-life situations. Usage relies on motivation, recall, and awareness. Wearable devices that track physiological signs of stress can boost engagement by encouraging the use of coping strategies when most needed.

Objective: This study aims to co-design an integrated user interface for wearable technology and the SloMo mobile app, and explore its usability, acceptability, engagement, and preliminary clinical outcomes for individuals with psychosis through a feasibility study.

Methods: A co-design team developed the wearable augmented SloMo app using the Double Diamond framework. The team included 10 experts by experience based across three UK National Health Service (NHS) mental health trusts, who had completed SloMo as part of a randomized controlled trial, and were purposively recruited to ensure a range of demographic backgrounds. Users tested the app for 8 consecutive weeks during a 5-month data collection period in 2018, providing feedback that guided iterative improvements. Self-reported engagement, usability, and acceptability were collected through quantitative ratings and interviews, evaluated through qualitative content analysis. Assessments of paranoia and well-being were completed pre- and post use, summarized descriptively, with change scores calculated and evaluated for reliable change.

Results: The Spire Stone clothing-adhered biosensor, which monitors respiratory rate and stress physiology, was chosen to augment SloMo. A minimum viable product of an integrated interface was developed. Feedback identified that users needed the technology to be easy to use, secure, and tailored to them through user control. Eight (8/10, 80%) users reported that the wearable prompted increased use of the SloMo app's features and other coping strategies. Self-reported engagement varied according to EBE user experience of the technology: higher use of the technology (n=3) was associated with higher usefulness (25% reduction in paranoia, SD 6%) and acceptability (84% satisfaction, SD 13%). Conversely, individuals that engaged with the device less (n=7) did not find it as useful (11% increase in paranoia, SD 17%) and were only modestly satisfied (63%, SD 20%).

Conclusions: Wearable technology is acceptable and can be integrated with a digitally supported therapy for psychosis to promote engagement and potentially support management of difficulties. However, further development work is needed to ensure acceptability and usefulness across a broader range of user needs.

Trial Registration: ISRCTN ISRCTN32448671; https://www.isrctn.com/ISRCTN32448671

(JMIR Hum Factors 2026;13:e91188) doi:10.2196/91188

KEYWORDS

human-centered design; mHealth; paranoia; schizophrenia; wearable electronic devices; psychotic disorders; digital health; feasibility studies; user-centered design; mobile health

XSL•FO
RenderX

Introduction

SloMo is a digitally supported cognitive behavioral therapy for psychosis (CBTp) that reduces paranoia [1]. It was developed to address the limitations of traditional psychological therapies for psychosis in relation to access, experience, and outcomes [2]. SloMo achieves this through targeting fast-thinking habits related to jumping to conclusions and belief inflexibility biases, which are established causal mechanisms in paranoia [3,4]. User experience is improved through co-design to leverage the capabilities of digital technologies [2,5]. SloMo consists of one-to-one sessions with a trained therapist, supported by a web-based therapy platform. Between sessions, the mobile app supports the individual to implement strategies independently, therefore providing continuity between therapy and daily life.

Between session practice, or "homework," is a fundamental component of CBTp [6], facilitating the generalization of therapeutic learning into real-life contexts. Such practice has shown to be predictive of successful outcomes in cognitive behavioral therapy (CBT) [7]. The availability, portability, and acceptability of smartphones means mobile apps are well placed to support skill practice in everyday life [8]. However, completion of tasks between sessions requires motivation, recall, understanding, and perceived benefits [9]. Difficulties in self-initiated behavior and planning can prevent effective use of strategies [10]. Relatedly, early experimental evaluations of SloMo showed its effects were moderated by cognitive difficulties, such as working memory problems, and negative symptoms [4].

Elevated levels of stress and anxiety present a further barrier to task completion, by detrimentally impacting functioning and cognition [11]. Early feasibility testing of SloMo demonstrated that individuals had greater difficulty remembering to use SloMo between sessions when distressed, and desired discrete prompts to help monitor their stress and encourage use of strategies [2]. User engagement or adherence, the extent to which users are willing to use and continue to use an intervention as intended, may influence effectiveness [12]. Consequently, approaches to improving engagement have been a significant focus to realizing the potential of digital health interventions and progress real-world implementation [13-15].

In psychosis, physiological parameters, such as heart rate variability, can provide an indicator of the onset of worsening paranoia [16]. Detection of elevated stress, anxiety, and paranoia could therefore prompt use of relaxation techniques or other coping strategies available via SloMo. However, consistent self-monitoring of stress can be challenging and effortful. Difficulties in identification of emotional states (ie, alexithymia) are common for people with psychosis [17]. Wearable devices provide a potential solution for detecting and managing stress through continuous passive monitoring of physiological signals. Accompanied by biofeedback and guidance via the technology or a linked device (eg, smartphone), this can support behavior change before stress escalates [18,19]. For SloMo users, notification of elevated physiological signs of stress could increase emotional awareness, providing opportunity to interrupt fast thinking habits, and support use of alternative slower thinking or coping strategies.

To be used and successfully integrated into daily life, wearable technologies must be usable, useful, and acceptable to the user [20]. Usability refers to the ease of use of the technology, including both the device itself and the app interface. Usefulness is the extent to which the device can achieve its intended purpose, ie, accurate monitoring and supporting self-management of difficulties. Acceptability encompasses factors that impact the user's willingness to use the technology, such as comfort, perceived value, and enjoyment. Poor usability risks impeding rather than augmenting outcomes [14]. This is of particular importance for individuals from marginalized groups who may be more likely to experience technology-related inequalities in access, use, and confidence [21]. For people with severe mental health needs, for whom privacy and data misuse concerns are common, security and trustworthiness of wearable technologies are especially important [22].

Inclusive, human-centered co-design (iHCD) is a method that places the user at the center of the design process: the needs of users representing a wide range of characteristics, abilities, and perspectives are explored, and design solutions incorporating these are iteratively tested and improved [23,24]. Transdisciplinary iHCD that includes experts by experience (EBE, ie, individuals, such as service users or carers, who have lived experience of the target problem), enhances adherence to digital interventions [25], promotes ethical usage [26], and is crucial for ensuring technologies are fit for purpose and meet the needs of diverse users [5]. This study adopted the Double Diamond process, a Human-Centered Design (HCD) framework using inclusive people-centered approaches [27,28].

The aims of this study were to (1) describe the development of a minimum viable product for an integrated user interface designed for wearable technology and the SloMo mobile app through co-design; and (2) conduct a feasibility study to explore usability, acceptability, engagement, and preliminary clinical outcomes of the developed technology, for individuals with psychosis.

Methods

This study was nested within the SlowMo randomized controlled trial (RCT) [1]. The first version of the therapy was "SlowMo," however spelling has since been updated to "SloMo" in line with feedback from individuals with lived experience of psychosis [5]. The RCT is referred to as SlowMo and the therapy as SloMo here forward.

Co-Design Team

A project team consisting of experts by experience (n=10), designers (n=4), software developers (n=5), clinicians (n=6), and researchers (n=5) participated in the co-design.

Ten individuals, who had previously received SloMo therapy as part of the trial and therefore had experience of the SloMo software, were invited to participate as EBE Co-Designers. In accordance with iHCD, purposive, and stratified sampling was applied with the aim of ensuring representation across trial sites,

XSL•FO
RenderX

ethnicity, age, and levels of technology confidence. EBEs were recruited from the 3 trial sites of London, Sussex, and Oxford. Eligible participants for the trial were: 18 years or older; had persistent (≥3 months) distressing paranoia; a diagnosis of schizophrenia spectrum psychosis; capacity to provide informed consent; and sufficient English to participate in trial processes. Participants were excluded if they had profound visual or hearing impairment, were unable to engage in assessments, were currently receiving psychological therapy for paranoia, or had a primary diagnosis of substance use disorder, personality disorder, organic syndrome, or learning disability. For further details of eligibility assessment, refer to Garety et al (2021) [1]. EBEs were invited to participate in this study following completion of all RCT research procedures (ie, 24-week follow-up assessment).

Characteristics of the EBEs are reported in Table 1. At baseline, persecutory ideation severity ranged from average (22) to very severe (62; see Measures: Green Paranoid Thought Scale [GPTS] [29]). Well-being scores were typically lower than the general population average (range 31 to 51; Warwick-Edinburgh Mental Wellbeing Scale [WEMWBS] [30]). Participant data from the RCT was extracted to facilitate description of the group and to inform preliminary understanding of who may benefit from SloMo augmentation with wearable technology. All EBEs completed all 8 sessions of SloMo therapy. Confidence and frequency of smartphone use (excluding calls) before SloMo were rated from 0 (Not at all; Never) to 100 (Totally; All the time). Confidence ranged from 11 to 95 (mean 67.0, SD 26.8), and frequency from 10 to 100 (mean 70.6, SD 27.6). In the RCT, a clinically observable response to SloMo therapy (defined as a 20% improvement in symptoms [31]) was seen for six (6/10; 60%) for paranoia (total GPTS), and two (2/10; 20%) for well-being (WEMWBS). The EBEs therefore represented a diverse subgroup of those who had completed SloMo therapy, with regards to demographics and technology attitudes and behaviors.

Table . Baseline demographic and clinical characteristics of EBE[a] co-designers (n=10).

Sociodemographic characteristic	Values
Age in years, mean (SD)	43.8 (11.0)
Sex, n (%)	
Male	8 (80.0)
Female	2 (20.0)
Ethnicity, n (%)	
Black, Black British, Black Welsh, Caribbean, or African	4 (40.0)
White	4 (40.0)
Asian, Asian British, Asian Welsh	1 (10.0)
Other ethnic group	1 (10.0)
GPTS[b], mean (SD)	
Total	76.1 (27.0)
Part A - referential ideas	37.8 (13.4)
Part B - persecution	38.3 (14.1)
WEMWBS[c], mean (SD)	
Total	39.7 (11.2)

[a]EBE: expert by experience co-designer.

[b]GPTS: Green et al [29] Paranoid Thought Scale.

[c]WEMWBS: Warwick-Edinburgh Mental Wellbeing Scale.

Intervention: SloMo

In SloMo therapy, an online therapy platform, accessed using a laptop or tablet, is used during one-to-one sessions with a trained therapist. A native Android mobile app supports strategy practice between sessions, with reminders and individualized content. An Android device, with the SloMo app installed, was provided to all participants for participation in the RCT and this study. This study included use of the mobile app software only, as EBEs had completed the full SloMo therapy prior to their involvement. Due to the early stage of development, access to the SloMo app was restricted to the supplied Android handset only, and was not available on the participant's own smartphone or alternative devices (eg, iOS).

The SloMo mobile app is described in detail elsewhere [2]. Key features include (1) a home screen where users can view personalized worries and safer thoughts; (2) a "Slow Down" function in which users are encouraged to notice worries, and use SloMo tips to help them to slow down to find ways of feeling safer; (3) a summary of therapy sessions and learning through "My Journey"; and (4) two behavioral coping enhancement activities: FlowMo, in which users are visually guided through an exercise to promote slow-paced breathing, with an expanding and shrinking bubble, and GoMo, a

XSL•FO
RenderX

distraction activity in which users are tasked with popping bubbles by tapping as they appear. FlowMo and GoMo were novel additions to the SloMo app, providing opportunity to intervene in the moment in response to elevated stress.

Ethical Considerations

The study was granted ethical approval through notice of substantial amendment of the RCT, which was reviewed and approved by the National Health Service Health Research Authority and London-Camberwell St. Giles Research Ethics Committee (16/LO/1862). After completion of all RCT procedures, eligible participants were provided with an information sheet and gave written, informed consent for participation in this substudy. Deidentified participant research data were held separately to personally identifiable information on secure servers accessible only to those in the research team. Participants were reimbursed £20 (GBP £1=US $1.43 as of February 1, 2018) for each meeting attended.

Procedure

Enrolment to the substudy commenced an average of 52 (SD 32) days after the SlowMo RCT 24-week follow-up assessment meeting, and 135 (SD 36) days after the posttherapy assessment.

Co-Design Procedures

The 4-phase Double Diamond framework [27], outlined below, was followed using a hybrid waterfall-agile approach, appropriate to the context of digital healthcare [5]. Following selection of the device, EBEs tested the technology for an 8-week period, contributing to all stages of the Double Diamond. The data collection period lasted 5 months, between February and July 2018.

Discover

The "Discover" phase is a divergent process focused on developing a broad understanding of the problem from multiple perspectives, while being informed by the existing evidence base. Initially, this phase consisted of "desk research" led by the designers, reviewing the existing market of wearable technologies to select an appropriate device for the augmentation of SloMo. Appropriateness was considered according to existing empirical evidence of acceptability to users, stress monitoring performance, feasibility of integration with SloMo, and compatibility with the previously identified user requirements for SloMo [2]. These prior co-design insights from SloMo therapy also informed the development of an initial wearable integration interface prototype. Designers, developers, and clinicians conducted testing of the wearable technology and prototype performance to identify improvements and bug fixes necessary before EBE testing.

EBEs were provided with the chosen wearable technology, shown how to use it, and asked to wear the device during the day over an 8-week period. Each EBE was provided a smartphone handset set-up with their own user profile from SloMo therapy (ie, tailored according to their own previously identified worries and safer thoughts). They were informed that they would receive feedback regarding stress levels through the mobile device. Handsets had a wireless internet connection for data transfer during the period of their involvement. Initial insights into attitudes and experience of the wearable were obtained through qualitative feedback at baseline and after one week of use. The EBEs also completed measures of paranoia, well-being, and were asked about attitudes towards monitoring stress and using technology.

EBEs completed the initial meeting with a study research worker at a time and location convenient to the individual, such as at home, their care team base, or at the research premises. Each meeting lasted up to one hour. If preferred, after initial set-up, interim qualitative feedback was collected remotely via telephone call. EBEs were asked at each meeting if they consented to audio recording, and written notes were produced by the research worker subsequently to summarize their feedback. If consent was not provided, the research worker took detailed written notes during the meeting. Questionnaire measures were completed on paper, with the support of the research worker as needed.

Define

The "Define" phase is a convergent process in which insights from the Discover phase were reviewed by the developers, designers, clinicians, and researchers, and synthesized to articulate user needs for the SloMo-wearable interface and a preliminary design brief for the solution. Throughout the co-design process, the study lead (AH) was responsible for the collation of EBE insights and ensuring user needs were refined accordingly, in collaboration with the interdisciplinary team.

Develop

The "Develop" phase is a divergent process in which a wide range of ideas for addressing the design brief are considered. Early-stage prototype solutions are iteratively developed, tested, and refined. Initial ideas in this phase were generated by the designers, software developers, and researchers. A series of software development sprints facilitated the iterative implementation of preferred solutions into the interface prototype, shaped by the ongoing collection of qualitative feedback from EBEs running in parallel. Decisions regarding which solutions should be kept were informed by both quantitative (ie, frequency of shared opinion) and qualitative content of feedback.

Deliver

In the "Deliver" phase, a convergent process is employed to develop the optimized solution into a higher fidelity prototype as the "minimum viable product" for further testing and release. The prototype was then tested and evaluated by the EBE group in a feasibility study. Usability and acceptability of the optimized interface were evaluated qualitatively and quantitatively through interviews and a user experience survey (UES). EBEs undertook reassessments of paranoia and well-being to support preliminary evaluation of the potential usefulness of the technology.

Measures

Stress Monitoring and Technology Use Attitudes

A semistructured interview was conducted during the Discover phase to understand the EBEs' attitudes to monitoring stress

XSL•FO
RenderX

and using technology (see Section 1 in Multimedia Appendix 1).

Usability and Acceptability

Semistructured Interview

Acceptability of the wearable device integration for SloMo was assessed at each follow-up meeting. Qualitative data regarding EBEs' views (ie, perceived helpfulness, concerns, and suggestions for improvement) and usage (ie, self-reported engagement; in what ways it is or is not being used, and likelihood of continued use) of the Spire Stone (Spire Health) and the SloMo app were elicited through a semistructured interview (Section 2 in Multimedia Appendix 1).

User Experience

User experience was assessed by the UES (see Multimedia Appendix 1; adapted from Ben-Zeev et al [32], as used in the SloMo RCT [1]). The UES is a 10-item measure consisting of 3 subscales assessing usefulness, usability, and enjoyment. Items are rated on a 0 (totally disagree) to 10 (totally agree) scale. Ratings are summed for each subscale and a percentage score calculated. The total score represents an overall measure of user satisfaction.

Usefulness

Paranoia

Paranoia was assessed using the GPTS [29]. The GPTS consists of two 16-item self-report subscales: part A—assessing referential ideas; part B—persecutory ideation. Items are rated on a 5-point scale (range 16 to 80 for each scale), with higher scores indicating greater difficulty. Internal reliability of the GPTS is considered excellent, with a reported Cronbach α of >0.9 [29,33].

Well-Being

Well-being was assessed using the WEMWBS [30], a 14-item self-report scale. Items are rated on a 5-point scale (range 14 to 70), with higher scores indicating better well-being. Very strong internal reliability has been reported (Cronbach α=0.93; [33]). A score of 42.9 or lower has been used to indicate poor psychological well-being [33].

Data Analysis

A combination of conventional and directed approaches to qualitative content analysis [34] was used to summarize usability and acceptability feedback. The lead author (AH) first read and reread all feedback for familiarization with the data. Using NVivo software (Lumivero), key details related to the EBE's user experience of the technology was highlighted. Preliminary codes were generated inductively and refined through iteration. Final codes were organized into categories according to whether they related to usability, acceptability, or usefulness. Frequency of participant responses for each code supported development decisions. Following identification of patterns during the initial inductive coding phase, qualitative feedback reflecting the extent of use of the technology was coded to indicate "lower," "mid," or "higher" engagement (see Table S1 in Multimedia Appendix 1).

Descriptive statistics were reported as frequency n (%) for categorical variables and mean (SD) for continuous variables. Change scores were calculated for paranoia and well-being outcomes between baseline and 8-week follow-up and evaluated for reliable change and a clinically observable response. The reliable change index (RCI) was calculated using the equation described by Jacobsen and Traux [35] (see Table S2 in Multimedia Appendix 1). A 20% reduction in symptoms was considered to represent a "clinically observable response," as has been used previously to signify at least "minimal improvement" [31]. An explorative subgroup analysis was conducted to investigate patterns of engagement according to EBE characteristics and outcomes, with mean GPTS, WEMWBS, and UES scores estimated, stratified by engagement level.

Results

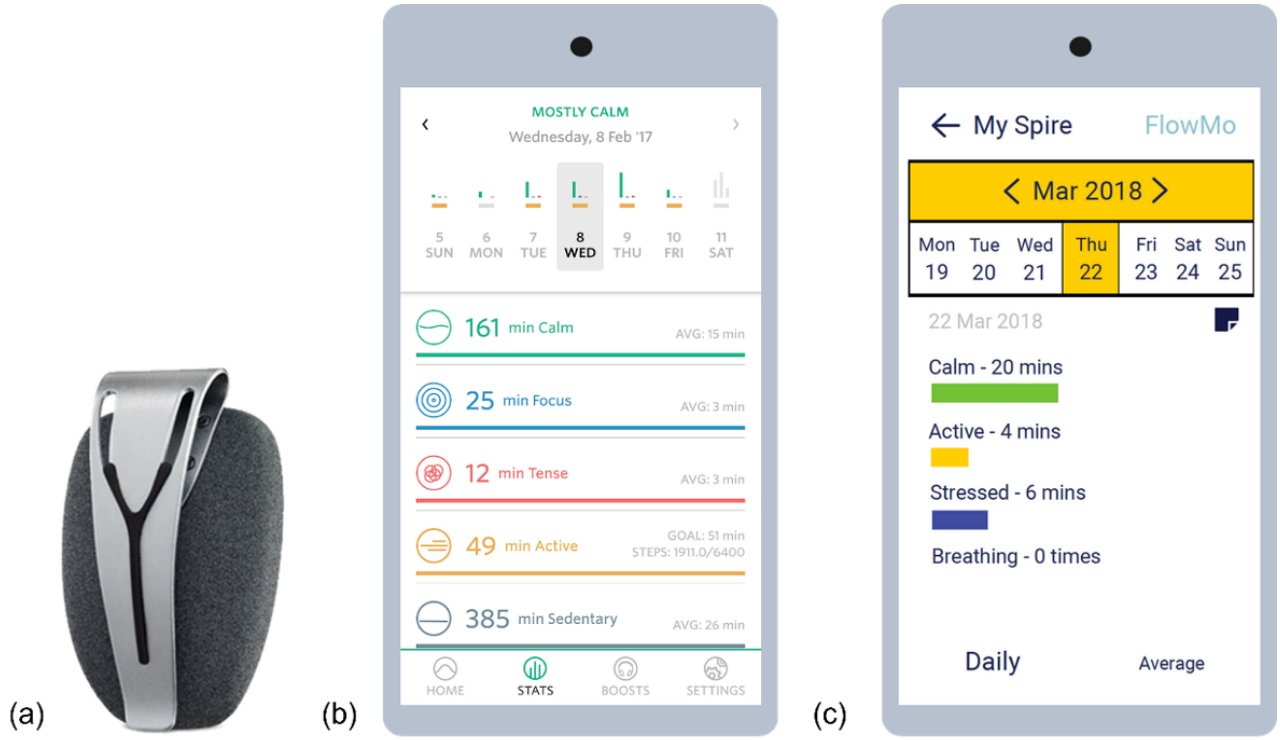
Discover

Wearable Technology

Desk research identified a range of wearable technologies that were easily accessible and commercially available. Considered solutions for the augmentation of SloMo included: wrist-worn bands and smartwatches, headsets, skin-adhered patches, in-earphones, clothing-embedded sensors, and smartphone software. The Spire Stone, a clothing-adhered biosensor commercially available until 2020, was selected to augment SloMo (Figure 1A) [36]. The Spire Stone had a publicly available API that enabled integration with the SloMo app, was financially feasible, and was considered discrete to wear relative to other devices.

XSL•FO
RenderX

Figure 1. (A) Spire clothing-adhered biosensor; (B) example nonadapted Spire smartphone app interface; (C) example SloMo app-My Spire interface.



The Spire Stone continuously monitored respiratory rate variability and compared this to the individual's median values after a period of calibration. Significant and sustained changes were interpreted as signals of cognitive-emotional state changes [37]. Feedback summarizing one-minute state intervals was presented to the wearer via a linked smartphone app, labeled as: "calm" (slow and consistent breathing), "focus" (moderately fast and consistent breathing), and "tense" (fast and erratic breathing) [18]. The device recorded the number of steps, inferring "active" compared to "sedentary" minutes. The wearer was also delivered in situ biofeedback notifications through device vibrations, to alert to sustained periods of states or sedentary behavior, aiming to facilitate self-regulation of stress levels. The Spire Stone was used previously to augment stress management interventions [18], and has demonstrated acceptability and usefulness [38]. Spire has reported strong correlations between their devices and "gold standard" medical device measurement of respiratory rate [39].

The nonadapted, commercially available Spire mobile app interface (Figure 1B) was also installed on the handsets provided to EBEs. Therefore, during the user testing, EBEs had the option to use either the Spire or SloMo wearable app, depending on their needs. Hereafter, we use "the technology" to refer to the integrated Spire Stone and linked SloMo wearable app.

Co-Design Feedback

Key themes regarding user needs related to acceptability, usability, and usefulness were identified from the EBEs' initial impressions of the technology. Illustrative quotes are presented in Table 2.

XSL•FO
RenderX

Table . Illustrative quotes representing EBE feedback on acceptability, usability, and usefulness for phase 1 (discover).

Quote ID	EBE[a]	Feedback
Acceptability		
1	2	Understanding can be useful […] validating your experiences: this is why you're feeling this way
2	3	I guess technology can make it more accurate to read how stressed you are and people who do not notice how they are feeling this could be a good way
3	8	Found the Spire app quite confusing to look at and said there were too many things going on, this put him off using it again
4	1	If someone was more stressed, it might annoy that person and highlight their anxieties for them
5	6	…Unhelpful to notice that she was feeling anxious as this would make her feel more anxious
6	1	Found the vibrations difficult to manage at work
7	3	Would use it more if [I] had more routine in [my] life, and at the moment [I'm] avoiding stressful situations and [do] not go out much so there is no point in wearing it when at home.
8	6	I trust you guys that you won't give me something with lasers in
9	2	[The Spire] looks like a microphone […] it might be recording
10	3	…can be hacked, and people can use if for harmful reasons to induce stress from the technology itself.
Usability		
11	10	There are too many alerts […] but it's not clear what the alert was for.
12	6	People who are worse at technology… or in the height of anxiety might not be able to know which are buttons or not, so could be made clearer.
Usefulness		
13	3	I don't notice that I'm tense when I'm doing certain things, but it appears that at times I'm flustered, which is helpful to know
14	4	When you are paranoid, you are in a vacuum so a bubble [notification] coming up breaks that vacuum
15	5	…like the simplicity [of FlowMo], puts me in a calm state […] you take a pill for a headache and a breath for anxiety.
16	7	[FlowMo] stops me going over the edge and calm down.
17	9	Maybe I'm just so used to feeling tense that it doesn't bother me that much
18	1	It wasn't that helpful to be told multiple times by the spire that you're 'feeling tense' and found it quite frustrating [it] was still 'feeling tense' after having done the breathing exercise

[a]EBE: expert by experience co-designer.

XSL•FO
RenderX

Acceptability

At baseline, most EBEs (n=7) perceived value in becoming more aware of emotions through monitoring of physical sensations (example quotes; Q 1 - 2). Three EBEs enjoyed being able to track progress and receive feedback and valued visually simpler displays to reduce confusion (Q3). Some (n=3), however, noted that increasing focus on stress levels through the technology could unhelpfully exacerbate difficulties (Q4-5). Two EBEs cited difficulty integrating into current routines, conversely either due to busyness (Q6) or limited activity levels (Q7). While the majority did not report concerns about the trustworthiness of the technology (Q8), 2 shared suspicions that there could be malicious intentions behind it (Q9-10). Nonetheless, both remained willing to continue using the device within the study.

Usability

Though most felt sufficiently competent in using the technology and reported they found it easy to use, two raised concerns that those less confident may have difficulties. After initial use, some reported a lack of clarity from the app and difficulty knowing what Spire Stone vibrations were signaling (n=2, Q11). One EBE described difficulty initiating the FlowMo breathing exercise when experiencing distress (Q12). Five reported issues with the technology that impacted use: the need for a second smartphone handset, poor battery life, and issues with synchronization between the Spire Stone and the app.

Usefulness

EBEs (n=6) described how using the technology to identify when they were feeling stressed supported self-management (Q13-14), including use of the SloMo app's wider features (n=6). Positive feedback was given on breathing exercises (n=6) in SloMo (FlowMo) or the Spire app, and GoMo (n=3), presenting an opportunity for relaxation, distraction, and prevention of worries worsening (Q15-16). Recommendations were given for the SloMo app to have further opportunities for relaxation and distraction, such as through calming sounds, music, or visuals, and games to have more variation (n=3). Monitoring and goal setting for activity levels within the Spire app were viewed as positive and motivating by some (n=3). While most (n=5) noted some accurate detection of difficult moments, 3 noted discrepancy and the technology overreporting stress, hypothesized to be due to desensitization by one EBE (Q17). In addition, the technology was noted to lack responsivity to efforts to reduce stress levels (n=2, Q18).

Define and Develop

Insights regarding user needs from the Discover phase were reviewed, synthesized and prioritized into 7 problems to be addressed through the co-design work (Table 3). Iterative development sprints enabled the testing and refinement of solutions within the SloMo wearable app interface (Table 3, Figure 1C).

Table . Redefined problems identified through phase 2 (define) and associated design solutions tested in phase 3 (develop).

Problem	Design solution to meet user requirements
Self-monitoring and self-management of stress can be challenging and effortful	The SloMo-Spire integration should be easy to use and access, through clear controls, navigation, and language. A diary function for note-taking and goal-setting supports users in self-management, monitoring, and motivation.
A focus on the experience of high stress can increase anxiety for users	Increased user controls enable the option to turn off "stress" state display and feedback. Notifications encourage users to recognize changes in emotion, rather than focusing solely on negative states.
Frequent notifications to alert to changes in emotional state through vibration lack clarity and meaning	Users can adjust notification time and frequency settings; users are provided with clear visual feedback on respiratory data. Vibration alerts have a corresponding visual, written app notification to prompt user action (ie, slowing down thoughts, FlowMo, or GoMo).
Feedback on emotional states that is misaligned with the user's own perception can be experienced as frustrating, confusing, and detrimental to trust	Simplification of data feedback to combine "focus" and "tense" states into one "stressed" state. SloMo app notifications neutrally prompt user reflections on current safety and distress to facilitate awareness and self-management.
Coping enhancement activities without variation and personalization offer limited opportunity for relaxation and/or distraction	Expansion of coping enhancement activities through the addition of sounds and music to FlowMo, and additional difficulty settings in GoMo.
Digital mental health intervention software may present challenges for those who are less technologically confident or experiencing acute anxiety	Users should be offered a clear demonstration, and additional support or repetition may be required to ensure confidence in use. The SloMo wearable app interface requires clear, easy-to-use buttons to activate coping activities.
Wearable technology can be a source of suspiciousness and paranoia	Users should be offered information regarding the purpose, function, and security of the technology, tailored according to individual concerns. This information should be easily accessible within the SloMo software.

Deliver

Qualitative Feedback

Illustrative quotes are presented in Table 4.

XSL•FO
RenderX

Table . Illustrative quotes representing EBE[a] feedback on acceptability, usability, and usefulness for phase 4 (deliver).

Quote ID	EBE	Feedback
Acceptability		
19	6	Was very excited to see her suggestions about relaxing noises being added […] really liked the forest noises in particular, as it reminded her of home
20	4	Finding it more enjoyable using the updated version […]using both levels on GoMo
21	5	Stopped using and made him worry that information might be broadcast to others […] He liked the updated app. The 'About SloMo' section reassured him [about] confidentiality and security, he would read for reassurance.
22	9	Has become habit to put on in the morning and play about with later on to see how he has been feeling
23	2	More positive about explaining the device and what it does to other people, and them being able to accept it
24	7	'Cause if the person's like me it would help them 100%. 'Cause like, I feel alright since I've been using it. It's helped me through all the bad times I've been using it, it's helped me.
Usefulness		
25	6	The app/device itself is not enough to remove the distress but it is great for helping notice her distress and take extra steps to alleviate it. […]
26	4	Spire will help clients deal with things better in between seeing their CPN or make them see CPN less often by supporting people in self-managing experiences
27	6	When you're on your own, it makes you feel like […] someone is actually there to say 'you can do these things.'
28	10	Learning [the] breathing without looking at the app […] can see it in his mind's eye and after a while won't need the app.
29	2	Found it the most frustrating that the app couldn't pick up on the distress his voices were expressing. He felt like the app made him look like a liar, as the distress experienced by the voices wasn't being picked up.

[a]EBE: Expert by Experience co-designer.

Acceptability

EBEs commented on liking the updated SloMo wearable app interface with regards to its simplicity and visual appeal (n=3). Changes made through the co-design process received positive feedback, specifically regarding the addition of relaxing sounds to FlowMo (n=2, Q19) and difficulty levels to GoMo (n=4, Q20) to better facilitate coping in response to wearable prompts. Some EBEs (n=4) expressed discomfort or inconvenience related to the wearable device type, with preference for alternatives (eg, wrist-worn and smartphone embedded). While 4 noted concerns related to intrusiveness, data privacy or app security, availability of in-app related information to provide clarity and reassurance was valued (Q21). EBEs described developments over time such as successful integration into daily life and habitual use (n=4, Q22), and becoming increasingly comfortable discussing the technology with loved ones (n=2, Q23). Six EBEs perceived that others with similar difficulties would be likely to benefit (Q24). The UES demonstrated overall positive ratings of acceptability (mean 72.7%, SD 28.1%) and enjoyment (mean 68.3%, SD 20.6%).

Usability

Several EBEs (n=4) commented on the ease of use of the technology. However, one user continued to feel they lacked technological skills or interest to benefit and wanted the app to

XSL•FO
RenderX

be more interactive or include 2-way communication with a person (eg, therapist) for support (P2). Limitations of the technology continued to impact experience, such as synchronization issues (n=2), inconvenience of requiring two handsets (n=1), and battery or charging issues (n=1). Responses to the UES demonstrated overall positive usability (mean 68.3%, SD 17.4%).

Usefulness

The majority of EBEs viewed the technology to be effective in facilitating self-management of difficulties (n=8) and reducing distress (n=4). While several noted the worn device did not solve the difficulty itself (n=3), it instead enabled them to become aware, prompting them to "do something differently," such as using the SloMo app's features (n=4), or engaging in other strategies (Q25). Four EBEs likened the app to an interim alternative to support from others (eg, a therapist, or community psychiatric nurse, Q26-27). Some described how they had begun to generalize learning to other contexts when not using the app (n=3, Q28). Three EBEs found there continued to be inaccuracies between the wearable feedback and their own perception, and for one this related specifically to the experience of hearing voices (Q29). Across the whole group, changes in outcomes were highly variable for both paranoia (total GPTS change mean −3.0, SD 18.1) and well-being (WEMWBS change mean −1.6, SD7.44).

Engagement

EBE feedback indicated varying levels of engagement with the technology through the co-design period. Three were categorized as "higher" usage, indicating they wore the Spire Stone and engaged with the technology for most days throughout the co-design period ("wearing the spire device every day from the moment she wakes up to moment she goes to bed," P4). Four indicated use for several days each week, or most weeks of the period and were considered "mid" usage. Three had limited engagement, using the technology for only a small proportion of the 8-week period ("didn't really use the app at all in last few weeks," P2).

EBE characteristics in relation to usage were explored to facilitate understanding of who may engage with and benefit from wearable augmentation of digital interventions, such as SloMo (Table S3 in Multimedia Appendix 1). Both female EBEs reported higher usage, however, each engagement group otherwise represented a range of ages, ethnicities, and smartphone confidence. As shown in Table 5, higher frequency users reported the most severe paranoia at baseline, while well-being scores were slightly lower for low frequency users at baseline.

Table . Mean pre- and postintervention scores for usefulness and acceptability outcome measures, by level of self-reported engagement.

Outcome, mean (SD)	Engagement level					
	Lower (n=3)		Mid (n=4)		Higher (n=3)	
	Baseline	8-week	Baseline	8-week	Baseline	8-week
Paranoia (GPTS[a] Total)	77.3 (25.4)	82.3 (22.8)	69.0 (27.6)	74.5 (5.5)	84.3 (35.9)	62.0 (21.2)
Paranoia: referential ideas (GPTS Part A)	38.0 (10.4)	41.3 (10.4)	33.0 (13.5)	33.25 (12.8)	44.0 (18.1)	30.7 (11.0)
Paranoia: persecutory ideas (GPTS Part B)	39.3 (15.0)	41.0 (12.5)	36.0 (14.1)	41.25 (12.9)	40.3 (18.8)	31.3 (10.2)
Well-being (WEMWBS)[b]	34.7 (3.2)	36.3 (2.5)	41.25 (4.5)	38.8 (5.2)	42.7 (7.4)	39.0 (21.2)
Satisfaction% (UES[c] Total)	—[d]	47.3 (22.0)	—	75.5 (4.4)	—	84.0 (13.0)
Enjoyment% (UES)	—	45.6 (18.4)	—	77.5 (5.7)	—	78.9 (20.4)
Usability% (UES)	—	47.5 (10.9)	—	73.8 (7.2)	—	81.7 (13.8)
Acceptability% (UES)	—	48.9 (42.3)	—	75.8 (13.4)	—	92.2 (6.9)

[a]GPTS: Green et al, [29] Paranoid Thought Scale (Subscale scores range from 16 - 80, total scores 32 - 160, higher scores indicate more severe symptoms).

[b]WEMWBS: Warwick-Edinburgh Mental Well-being Scale (scores range from 14 - 70, higher scores indicate better well-being).

[c]UES: User Experience Survey (scores reported as percentage agreement for each scale).

[d]Not applicable.

XSL•FO
RenderX

Level of engagement differed according to the technology's usefulness and acceptability to the user. Those who used the technology most also experienced improvement in total paranoia scores at 8-week follow-up (clinically observable response, n=2; reliable change; RC; n=1; see Table S3 in Multimedia Appendix 1). In contrast, a pattern was observed that lower and mid frequency users reported a slight deterioration, although this did not meet threshold for reliable change. Well-being scores remained consistent overall regardless of engagement, however one higher, and one mid frequency user reported a reliable deterioration in well-being. Similarly, acceptability data provided by UES responses aligned with self-reported usage, such that those that engaged more were also most satisfied, while low frequency users were least.

Discussion

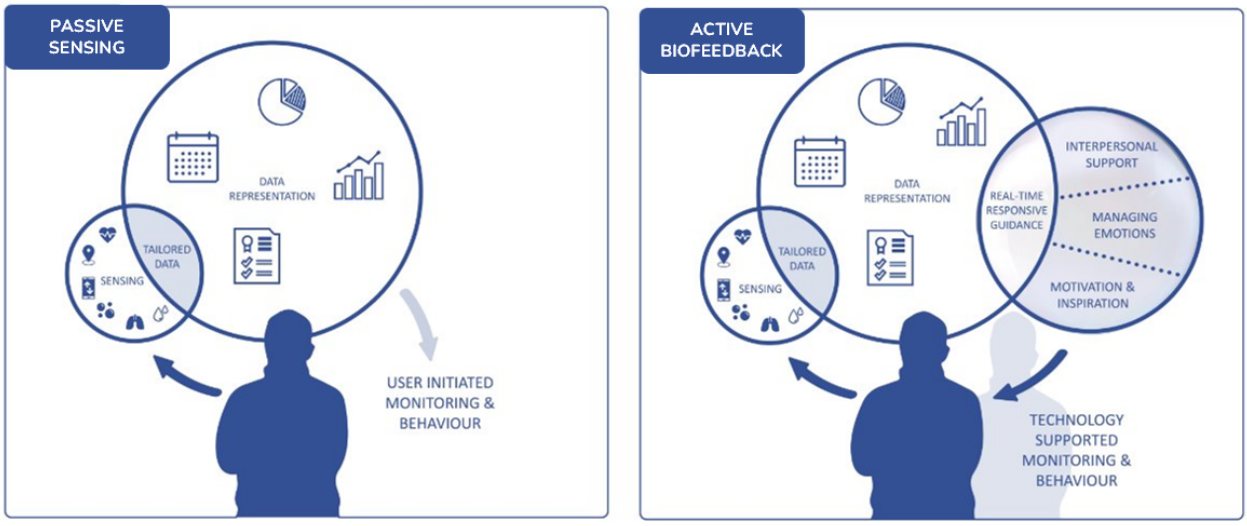
Principal Findings

This study is the first to explore the user experience of wearable technology integrated with an existing digital therapy for paranoia in people with psychosis. A minimal viable product was first developed, comprising a user interface that integrated the wearable technology with the SloMo mobile app. Most users found the developed technology to be usable, acceptable, and enjoyable to use (mean agreement>68%, SD 20%). Some EBEs reported using the technology frequently and had higher overall user satisfaction (agreement 84%, SD 13%), finding the wearable to be discrete, secure, robust, accurate, and useful, and reported reductions in paranoia. For other EBEs who described using the technology less often, modest user satisfaction was reported (47%, SD 22% to 76%, SD 4%) and there was no marked impact on paranoia. There was encouraging early evidence that, for some, wearables might promote engagement with a therapy app to support management of paranoia, at least in the short term, and even in the absence of therapist support.

The findings highlight the potential for responsive, wearable technology to improve engagement. Almost all EBEs (8/10, 80%) reflected that when using the Spire device, they also used the SloMo app more frequently. There are known challenges for engagement and retention in digital mental health interventions over time [40,41]. Strategies that are tailored and prompt intervention engagement when it is most needed may optimize effects and enhance subjective usefulness. The longer-term effectiveness of SloMo is not yet known and is currently being evaluated in an implementation-effectiveness study (ClinicalTrials.gov ID NCT06568081). The current study suggests that wearable integration may be a plausible way of encouraging ongoing use if engagement is found to decline after formal therapy completion.

For people with psychosis, to date, wearable technology has most commonly been used as a tool for assessment, monitoring, and diagnosis [42]. There is great interest in the potential of digital phenotyping (ie, passive sensing) through smartphones to detect and prevent relapse [43]. However, our inclusive, user-centered design research suggests that people want technology to provide secure, accessible, and responsive support to assist self-management, and are reluctant for this to involve the sharing of personal data without sufficient clarity and reassurance. This is consistent with the findings of a study exploring perspectives of people with psychosis on passive sensing, highlighting the importance of fully informed consent, choice, and autonomy [44]. Similarly, despite an initial focus on monitoring stress in this study, EBE feedback demonstrated this emphasis to be undesirable to some. Instead, a preference was shared for the wearable to help increase awareness of "calm" and "active" states, and the associated activities and strategies that support this. Wearable technology including active biofeedback, not solely passive monitoring, may be key to realizing the promise of digital health in innovating the treatment of psychosis (see Figure 2).

Figure 2. Wearable therapeutics offering biofeedback as a preferred method by EBE co-designers for supporting monitoring of symptoms and behavior change, relative to passive sensing only.



The participating EBEs represent a diverse group in relation to age, ethnicity, technological confidence, and paranoia severity. Engagement and user experience were not impacted by these factors. Notably, as a group, people who reported using the

XSL•FO
RenderX

technology most had the most severe paranoia at the outset. While some EBEs expressed concerns about the trustworthiness and safety of wearable technology, it did not deter them from continuing to use it. Reassurance, gaining familiarity over time, and availability of accessible written information regarding data privacy and security were viewed as helpful solutions to these concerns. Concerns regarding "service users feeling suspicious or paranoid" are cited as the most common barrier to mental health professionals recommending the use of digital mental health interventions (38). This study demonstrates that this concern may risk preventing access for those who could benefit.

In keeping with the early stage of the work, technological challenges were common, with the group reporting technical issues (7/10, 70%), usability problems (5/10, 50%), discomfort (4/10, 40%), and discrepancy with their own perceived emotional state (4/10, 40%). Advancements in wearable technologies may alleviate some of these challenges. However, an exacerbated discrepancy between self-reported and physiological stress levels in people with psychosis has been reported previously [45]. Biofeedback should be suggested cautiously, for example, "You may be feeling stressed; would you like to slow down?" This approach avoids damaging the person's trust in the technology.

Limitations and Future Directions

The EBE co-designers involved in this study represent a small, but diverse group of individuals with regards to age, ethnicity, technological confidence, and attitudes towards technology. All EBEs had engaged in SloMo therapy prior to involvement and had completed therapy with high rates of adherence. To this extent, the study's findings represent the views of people who engaged well, with prior knowledge of SloMo, and may differ from those who may be less engaged with digitally supported talking therapy. Furthermore, as the same group of EBEs were involved throughout the co-design process, acceptability ratings may be higher due to the implementation of their own recommendations for improvements. Conversely, EBEs were not involved in the initial selection of the Spire Stone wearable device. Consultation regarding user priorities for the device at an earlier stage may identify an alternative, more acceptable device to users and may facilitate improved engagement.

The experiences and outcomes reported were highly variable. Females were underrepresented in the group; however, both female EBEs used the technology frequently, reported high satisfaction, and benefitted in terms of paranoia improvements. Higher engagement and satisfaction with digital interventions for psychosis by female users has been noted previously [13], and was similarly found in the SlowMo RCT [1]. This indicates a clear need for development work to further optimize the wearable augmentation of SloMo to ensure acceptability for a wider range of people, particularly nonfemale users.

Baseline data was collected subsequent to the 24-week follow-up for the SlowMo RCT. In the RCT, participants typically experienced benefits to paranoia and well-being following SloMo therapy that were maintained at follow-up [1]. To this extent, most EBEs in the current study commenced participation having already achieved clinical benefit from SloMo therapy. Less engaged users experienced fluctuations in paranoia, as is expected in this population [46], though levels remained in keeping with the SlowMo treatment group average at 24-week follow-up. Furthermore, the fluctuation for these groups did not exceed the recommended minimally clinically significant response used in meta-analyses of CBTp [31]. Regardless, potential iatrogenic harms should continue to be explored in future work.

Exploratory post hoc investigations in relation to patterns of engagement were conducted based on coding of qualitative descriptions by the EBEs. The patterns observed correspond to the objective data collected in this study, such that use of a digital health technology is influenced by a complex system of interacting components including acceptability (eg, beliefs, knowledge, and affective attitude), usability, and usefulness [20]. Objective measurement of engagement (eg, technical usage data) and a priori operationalization of engagement levels would strengthen future similar investigations.

Reflective of the technology available at the time of study delivery, the reliability for "stress detection" was limited: the device could identify changes in breathing patterns but was insufficient in accurately determining whether these changes reflected genuine distress. Accordingly, the Spire Stone was reported as inaccurate in stress signal detection by several EBEs, impacting the perceived usefulness of the integrated technology. Advancements in wearable technology are rapidly developing [47,48] and may offer an improved solution that enhances trust, usefulness, and therefore the overall acceptability of mental health interventions.

Conclusions

This study demonstrates that wearable technology can be integrated into SloMo, a digitally supported therapy for psychosis. Involving those with lived experience in the development of digital mental health interventions is crucial to ensuring alignment with user needs. While standalone, passive monitoring of stress was not viewed as useful by EBEs, people were positive about the potential utility of integrating wearables with digital interventions to facilitate self-management of difficulties. People may use therapy apps more frequently when wearable devices are added, which is important because the experience and outcomes of current psychological therapies for psychosis need to be improved. However, wearable technology outcomes currently remain variable. Development work is needed to broaden acceptability, usability, and usefulness for a wide range of people.

Acknowledgments

We would like to thank Ed Matthews, Jonathan West, and Nassia Inglessis, Healthcare Lab, Helen Hamlyn CentreCentre for Design who contributed to the design of the first version of SloMo. We would also like to acknowledge Evolyst Limited who are

XSL•FO
RenderX

the software developers for SloMo version one. We are grateful to the research team members involved in the study and are indebted to the people with psychosis who have contributed to development.

The legal manufacturer of the SloMo mobile app is King's College London.

Funding

This work was primarily funded by the Wellcome Trust through their Pathfinder award (109586/Z/15/Z) and was assisted by two further funding bodies. First, the National Institute for Health Research and Medical Research Council have supported the research through their Efficacy and Mechanism Evaluation funding of the SlowMo trial (ISRCTN32448671). Second, the King's Commercialisation Institute has funded the development of an app, Mo, targeting well-being in the general population. This has assisted the work as several aspects of Mo's functionality have been repurposed for the SloMo wearable app. AH, MC and TW acknowledge funding from the Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: AH (lead), TW (supporting), PG (supporting)
Data curation: AH (lead), KT (supporting)
Formal analysis: KT (lead), AH (supporting)
Funding acquisition: AH, TW
Investigation: AH (lead), AW (supporting)
Methodology: AH (lead), TW (supporting)
Project administration: AH
Resources: AH
Software: AH (lead), AW (supporting), MF (supporting)
Supervision: AH (lead), TW (supporting), PG (supporting), MC (supporting)
Visualization: KT (lead), AH (equal), AW (equal)
Writing – original draft: KT (lead), AH (supporting)
Writing – review & editing: KT (lead), AH (equal), TW (supporting), MC (supporting), PG (supporting), AG (supporting), KP (supporting), AW (supporting), MF (supporting)

Conflicts of Interest

KT, AH, TW, and PG are cofounders of SloMo therapy. KT, AH, TW, PG, KP, and AG are investigators on the Wellcome Trust-funded SloMo2 implementation-effectiveness study. KT, AH, and TW are employees of King's College London who is the legal manufacturer of SloMo. The remaining authors have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Multimedia Appendix 1
Semi-structured interview topic guides; User Experience Survey; engagement coding description; Reliable Change Indicator calculation; individual EBE characteristics.
[DOCX File, 57 KB - humanfactors_v13i1e91188_app1.docx]

References

1. Garety P, Ward T, Emsley R, et al. Digitally supported CBT to reduce paranoia and improve reasoning for people with schizophrenia-spectrum psychosis: the SlowMo RCT. Efficacy Mech Eval 2021;8(11):1-90. [doi: 10.3310/eme08110]
2. Hardy A, Wojdecka A, West J, et al. How inclusive, user-centered design research can improve psychological therapies for psychosis: development of SlowMo. JMIR Ment Health 2018 Dec 5;5(4):e11222. [doi: 10.2196/11222] [Medline: 30518514]
3. Garety PA, Freeman D, Jolley S, et al. Reasoning, emotions, and delusional conviction in psychosis. J Abnorm Psychol 2005 Aug;114(3):373-384. [doi: 10.1037/0021-843X.114.3.373] [Medline: 16117574]
4. Garety P, Waller H, Emsley R, et al. Cognitive mechanisms of change in delusions: an experimental investigation targeting reasoning to effect change in paranoia. Schizophr Bull 2015 Mar;41(2):400-410. [doi: 10.1093/schbul/sbu103] [Medline: 25053650]
5. Hardy A, Taylor KM, Grant A, et al. Co-designing technology to improve psychological therapy for psychosis: SloMo, a blended digital therapy for fear of harm from others. Schizophr Res 2024 Dec;274:526-534. [doi: 10.1016/j.schres.2024.11.004] [Medline: 39579660]

XSL•FO
RenderX

6. Morrison AP, Barratt S. What are the components of CBT for psychosis? A Delphi study. Schizophr Bull 2010 Jan;36(1):136-142. [doi: 10.1093/schbul/sbp118] [Medline: 19880824]
7. Kazantzis N, Whittington C, Zelencich L, Kyrios M, Norton PJ, Hofmann SG. Quantity and quality of homework compliance: a meta-analysis of relations with outcome in cognitive behavior therapy. Behav Ther 2016 Sep;47(5):755-772. [doi: 10.1016/j.beth.2016.05.002] [Medline: 27816086]
8. Tang W, Kreindler D. Supporting homework compliance in cognitive behavioural therapy: essential features of mobile apps. JMIR Ment Health 2017 Jun 8;4(2):e20. [doi: 10.2196/mental.5283] [Medline: 28596145]
9. Dunn H, Morrison AP, Bentall RP. Patients' experiences of homework tasks in cognitive behavioural therapy for psychosis: a qualitative analysis. Clin Psychology and Psychoth 2002 Sep;9(5):361-369. [doi: 10.1002/cpp.344]
10. Cella M, Reeder C, Wykes T. Lessons learnt? The importance of metacognition and its implications for cognitive remediation in schizophrenia. Front Psychol 2015;6:1259. [doi: 10.3389/fpsyg.2015.01259] [Medline: 26388797]
11. Aas M, Dazzan P, Mondelli V, et al. Abnormal cortisol awakening response predicts worse cognitive function in patients with first-episode psychosis. Psychol Med 2011 Mar;41(3):463-476. [doi: 10.1017/S0033291710001170] [Medline: 20529412]
12. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 5;13(3):e52. [doi: 10.2196/jmir.1772] [Medline: 21821503]
13. Aref-Adib G, McCloud T, Ross J, et al. Factors affecting implementation of digital health interventions for people with psychosis or bipolar disorder, and their family and friends: a systematic review. Lancet Psychiatry 2019 Mar;6(3):257-266. [doi: 10.1016/S2215-0366(18)30302-X] [Medline: 30522979]
14. Eisner E, Faulkner S, Allan S, et al. Barriers and facilitators of user engagement with digital mental health interventions for people with psychosis or bipolar disorder: systematic review and best-fit framework synthesis. JMIR Ment Health 2025 Jan 20;12(1):e65246. [doi: 10.2196/65246] [Medline: 39832352]
15. Smith KA, Ward T, Lambe S, et al. Engagement and attrition in digital mental health: current challenges and potential solutions. NPJ Digit Med 2025 Jul 2;8(1):398. [doi: 10.1038/s41746-025-01778-w] [Medline: 40604240]
16. Schlier B, Krkovic K, Clamor A, Lincoln TM. Autonomic arousal during psychosis spectrum experiences: Results from a high resolution ambulatory assessment study over the course of symptom on- and offset. Schizophr Res 2019 Oct;212:163-170. [doi: 10.1016/j.schres.2019.07.046] [Medline: 31422861]
17. Ozdemir E, Xiao Z, Griffiths H, MacBeth A. Alexithymia in schizophrenia and psychosis vulnerability: a systematic review and meta-analysis. J Clin Psychol 2025 Jun;81(6):410-424. [doi: 10.1002/jclp.23788] [Medline: 40105143]
18. Smith EN, Santoro E, Moraveji N, Susi M, Crum AJ. Integrating wearables in stress management interventions: promising evidence from a randomized trial. Int J Stress Manag 2020;27(2):172-182. [doi: 10.1037/str0000137]
19. González Ramírez ML, García Vázquez JP, Rodríguez MD, Padilla-López LA, Galindo-Aldana GM, Cuevas-González D. Wearables for stress management: a scoping review. Healthcare (Basel) 2023 Aug 22;11(17):2369. [doi: 10.3390/healthcare11172369] [Medline: 37685403]
20. Perski O, Short CE. Acceptability of digital health interventions: embracing the complexity. Transl Behav Med 2021 Jul 29;11(7):1473-1480. [doi: 10.1093/tbm/ibab048] [Medline: 33963864]
21. Hardy A, Ward T, Emsley R, et al. Bridging the digital divide in psychological therapies: observational study of engagement with the SlowMo mobile app for paranoia in psychosis. JMIR Hum Factors 2022 Jul 1;9(3):e29725. [doi: 10.2196/29725] [Medline: 35776506]
22. Byrne S, Tohamy A, Kotze B, et al. Using a mobile health device to monitor physiological stress for serious mental illness: a qualitative analysis of patient and clinician-related acceptability. Psychiatr Rehabil J 2022 Sep;45(3):219-225. [doi: 10.1037/prj0000514] [Medline: 35298226]
23. Giacomin J. What Is human centred design? Des J 2014 Dec;17(4):606-623. [doi: 10.2752/175630614X14056185480186]
24. John Clarkson P, Coleman R. History of inclusive design in the UK. Appl Ergon 2015 Jan;46 Pt B:235-247. [doi: 10.1016/j.apergo.2013.03.002] [Medline: 23570838]
25. Killikelly C, He Z, Reeder C, Wykes T. Improving adherence to web-based and mobile technologies for people with psychosis: systematic review of new potential predictors of adherence. JMIR Mhealth Uhealth 2017 Jul 20;5(7):e94. [doi: 10.2196/mhealth.7088] [Medline: 28729235]
26. Doherty A, Bucci S, Kenny A, et al. Passive sensing at scale to transform understanding of poor mental health. Lancet Digit Health 2025 Mar;7(3):e172-e174. [doi: 10.1016/j.landig.2025.01.002] [Medline: 40015762]
27. Framework for Innovation. Design Council. 2019. URL: https://www.designcouncil.org.uk/our-resources/framework-for-innovation/ [accessed 2025-07-28]
28. West J, Fusari G, Raby E, et al. Developing the double diamond process for implementation. In: Barron D, Seemann K, editors. Presented at: Proceedings of the Fourth International Conference on Design4Health 2017; Dec 4-7, 2017; Melbourne, Australia p. 310-312 URL: https://researchonline.rca.ac.uk/3603/ [accessed 2026-09-03]
29. Green CEL, Freeman D, Kuipers E, et al. Measuring ideas of persecution and social reference: the Green et al. Paranoid Thought Scales (GPTS). Psychol Med 2008 Jan;38(1):101-111. [doi: 10.1017/S0033291707001638] [Medline: 17903336]
30. Tennant R, Hiller L, Fishwick R, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5(1):63. [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]

XSL•FO
RenderX

31. Bighelli I, Huhn M, Schneider-Thoma J, et al. Response rates in patients with schizophrenia and positive symptoms receiving cognitive behavioural therapy: a systematic review and single-group meta-analysis. BMC Psychiatry 2018 Dec 4;18(1):380. [doi: 10.1186/s12888-018-1964-8] [Medline: 30514268]
32. Ben-Zeev D, Brenner CJ, Begale M, Duffecy J, Mohr DC, Mueser KT. Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. Schizophr Bull 2014 Nov;40(6):1244-1253. [doi: 10.1093/schbul/sbu033] [Medline: 24609454]
33. Freeman D, Taylor KM, Molodynski A, Waite F. Treatable clinical intervention targets for patients with schizophrenia. Schizophr Res 2019 Sep;211:44-50. [doi: 10.1016/j.schres.2019.07.016] [Medline: 31326234]
34. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
35. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991 Feb;59(1):12-19. [doi: 10.1037//0022-006x.59.1.12] [Medline: 2002127]
36. System and method for monitoring respiration. Google Patents. 2014 Aug 14. URL: https://patents.google.com/patent/US20140228657A1/ko [accessed 2025-10-31]
37. Hernando A, Lazaro J, Gil E, et al. Inclusion of respiratory frequency information in heart rate variability analysis for stress assessment. IEEE J Biomed Health Inform 2016 Jul;20(4):1016-1025. [doi: 10.1109/JBHI.2016.2553578] [Medline: 27093713]
38. Alberts NM, Leisenring WM, Flynn JS, et al. Wearable respiratory monitoring and feedback for chronic pain in adult survivors of childhood cancer: a feasibility randomized controlled trial from the childhood cancer survivor study. JCO Clin Cancer Inform 2020 Nov;4(4):1014-1026. [doi: 10.1200/CCI.20.00070] [Medline: 33147073]
39. Holt M, Yule B, Jackson D, Zhu M, Moraveji N. Ambulatory monitoring of respiratory effort using a clothing-adhered biosensor. Presented at: 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA); Jun 11-13, 2018; Rome, Italy p. 1-6. [doi: 10.1109/MeMeA.2018.8438678]
40. Boucher EM, Raiker JS. Engagement and retention in digital mental health interventions: a narrative review. BMC Digit Health 2024;2(1):52. [doi: 10.1186/s44247-024-00105-9]
41. Baumel A, Edan S, Kane JM. Is there a trial bias impacting user engagement with unguided e-mental health interventions? A systematic comparison of published reports and real-world usage of the same programs. Transl Behav Med 2019 Nov 25;9(6):1020-1033. [doi: 10.1093/tbm/ibz147] [Medline: 31689344]
42. Fonseka LN, Woo BKP. Wearables in schizophrenia: update on current and future clinical applications. JMIR Mhealth Uhealth 2022 Apr 7;10(4):e35600. [doi: 10.2196/35600] [Medline: 35389361]
43. Bladon S, Eisner E, Bucci S, et al. A systematic review of passive data for remote monitoring in psychosis and schizophrenia. NPJ Digit Med 2025 Jan 27;8(1):62. [doi: 10.1038/s41746-025-01451-2] [Medline: 39870797]
44. Eisner E, Ball H, Ainsworth J, et al. Using passive sensing to predict psychosis relapse: an in-depth qualitative study exploring perspectives of people with psychosis. Schizophr Bull 2026 Jul 3;52(4):sbaf126. [doi: 10.1093/schbul/sbaf126] [Medline: 40840436]
45. Söder E, Clamor A, Kempkensteffen J, Moritz S, Lincoln TM. Stress levels in psychosis: do body and mind diverge? Biol Psychol 2018 Oct;138:156-164. [doi: 10.1016/j.biopsycho.2018.08.016] [Medline: 30142356]
46. Peters E, Crombie T, Agbedjro D, et al. The long-term effectiveness of cognitive behavior therapy for psychosis within a routine psychological therapies service. Front Psychol 2015;6:1658. [doi: 10.3389/fpsyg.2015.01658] [Medline: 26579041]
47. Elgendi M, Markov K, Liu H, et al. Wearable devices for anxiety assessment: a systematic review. Commun Med (Lond) 2026 Jan 9;6(1):20. [doi: 10.1038/s43856-025-01234-6] [Medline: 41513763]
48. Pinge A, Gad VR, Jaisighani D, Ghosh S, Sen S. Detection and monitoring of stress using wearables: a systematic review. Front Comput Sci 2024;6. [doi: 10.3389/fcomp.2024.1478851]

Abbreviations

CBT: cognitive behavioral therapy
CBTp: cognitive behavioral therapy for psychosis
EBE: expert by experience
GPTS: Green Paranoid Thought Scale
HCD: human-centered design
iHCD: inclusive human-centered design
RCI: reliable change index
RCT: randomized controlled trial
UES: User Experience Survey
WEMWBS: Warwick-Edinburgh Mental Wellbeing Scale

XSL•FO
RenderX

Edited by A Kushniruk; submitted 10.Jan.2026; peer-reviewed by HMS Ajmal, PJ Dunne; revised version received 24.Jul.2026; accepted 20.Aug.2026; published 02.Oct.2026.

Please cite as:
Taylor KM, Ward T, Cella M, Garety P, Grant A, Philps K, Wojdecka A, Flory M, Hardy A
Integrating Wearable Technology in Digital Therapy for People With Psychosis (SloMo): Mixed Methods Co-Design and Feasibility Study
JMIR Hum Factors 2026;13:e91188
URL: https://humanfactors.jmir.org/2026/1/e91188
doi:10.2196/91188

© Kathryn M Taylor, Thomas Ward, Matteo Cella, Philippa Garety, Amy Grant, Katie Philps, Anna Wojdecka, Melanie Flory, Amy Hardy. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 2.Oct.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Effects of a Group-Based Cognitive Behavioral Therapy for Sleep in Nursing Students: Pilot Randomized Controlled Trial

David Pérez-Manchón[1,2*], PhD; Clara Azpeleta[3], PhD; Beatriz Gal-Iglesias[4*], PhD; Cayetana Ruiz Zaldibar[1,2*], PhD

[1]Department of Nurisng, University Center for Health Sciences – HM Hospitals (CUHMED), Camilo Jose Cela University, Villanueva de la Cañada 28692, Madrid, Spain

[2]Instituto de Investigación Sanitaria HM Hospitales, Madrid, Spain

[3]Universidad Europea de Madrid, Faculty of Medicine, Health and Sports, Department of Medicine, Madrid, Spain

[4]Departamento de Ciencias Médicas Básicas, Facultad de Medicina, Universidad CEU San Pablo, CEU Universities, Madrid, Spain

[*]these authors contributed equally

Corresponding Author:
Cayetana Ruiz Zaldibar, PhD
Department of Nurisng, University Center for Health Sciences – HM Hospitals (CUHMED), Camilo Jose Cela University, Villanueva de la Cañada 28692, Madrid, Spain

Abstract

Background: Sleep quality is a strategic public health priority and a key factor in the study of circadian rhythms and chronotypes. Nursing students are a particularly vulnerable population group.

Objective: The objective of this study was to evaluate the effects of a brief group-based cognitive behavioral therapy for insomnia (CBT-I) intervention on sleep quality and circadian-related outcomes among first-year nursing students using ambulatory circadian monitoring.

Methods: This study is a 2-arm pilot randomized clinical trial conducted at a private university in Madrid, Spain, from October 2022 to March 2023. First-year nursing students aged 18 to 25 years were recruited in October 2022 using convenience sampling and randomly assigned to an experimental group (n=20) that received a cognitive-behavioral intervention using an active-constructive learning methodology to improve sleep quality and a control group (n=20) that followed their usual daily routine. Primary outcomes were objective sleep quality assessed using the Kronowise 3.0 ambulatory circadian monitoring system and subjective sleep quality assessed using the Pittsburgh Sleep Quality Index (PSQI). The objective Kronowise assessment was characterized by the continuous circadian rhythm and sleep parameters generated by the device. Participant satisfaction with the intervention was evaluated using an adapted 12-item Likert-type satisfaction questionnaire administered after the intervention. Measurements were collected preintervention and postintervention over 7-day monitoring periods. Adjusted analyses were performed using analysis of covariance, with treatment group as the fixed factor and baseline values, smoking status, school schedule, and coffee consumption as covariates. Treatment effects were summarized using regression coefficients (β), 95% CIs, and partial eta-squared (η^2). The study was conducted in accordance with the CONSORT (Consolidated Standards of Reporting Trials) statement.

Results: Forty students (n=34, 85% female; mean age 19.9, SD 1.8 y) were randomized equally into the intervention and control groups. Adjusted analyses showed generally small treatment effects for objective circadian rhythm and sleep outcomes. The largest treatment effect was observed for self-reported sleep latency (β=−19.78 min, 95% CI −33.83 to −5.74; partial η^2=0.194). Sleep duration (β=36.32 min, 95% CI −5.01 to 77.64; partial η^2=0.086) and sleep efficiency (β=8.75%, 95% CI −0.97 to 18.47; partial η^2=0.090) also favored the intervention group, although CIs indicated considerable uncertainty around these estimates. Participants reported high satisfaction with the intervention, with 65% (13/20) rating the program as “very good.”

Conclusions: This pilot randomized controlled trial provides preliminary evidence that a brief group-based CBT-I intervention may improve selected subjective sleep outcomes, particularly sleep latency, among first-year nursing students. Objective circadian rhythm and sleep outcomes showed generally small treatment effects. Given the exploratory nature of this pilot study, adequately powered randomized controlled trials with longer follow-up are needed to confirm these preliminary findings.

Trial Registration: ClinicalTrials.gov NCT05273086; https://clinicaltrials.gov/study/NCT05273086

International Registered Report Identifier (IRRID): RR2-https://doi.org/10.3390/ijerph192113886

(JMIR Hum Factors 2026;13:e85021) doi:10.2196/85021

KEYWORDS

nursing students; sleep quality; sleep hygiene; wearable; cognitive behavioral therapy

XSL•FO
RenderX

Introduction

Sleep is an essential physiological process for physical restoration and cognitive functioning. Sleep quality is closely linked to circadian rhythm regulation and is a key determinant of physical and psychological health. Adequate sleep quality is associated with well-being and daytime functioning, whereas poor sleep quality has been linked to adverse health outcomes, including anxiety, depression, hypertension, and metabolic disorders [1-3]. Because sleep quality reflects both behavioral and circadian processes, it is an important indicator for understanding sleep-wake regulation and overall circadian health [4].

It is estimated that 1 in 3 adults has a sleep disorder [5], and university students are a particularly vulnerable population due to academic demands, irregular schedules, and lifestyle-related factors that may disrupt healthy sleep patterns. Poor sleep quality in this population has been associated with fatigue, impaired concentration, emotional distress, and poorer academic performance [6]. Nursing students may be especially vulnerable because of the demanding academic workload, early clinical exposure, and the need to adapt to unfamiliar practice environments, all of which may increase stress and contribute to sleep disruption [7,8]. This vulnerability may be particularly relevant during the first academic year, when students are adapting to university life while facing increased academic and professional expectations.

Given the limitations of pharmacological approaches for sleep problems, nonpharmacological strategies have gained increasing attention [9]. Cognitive behavioral therapy (CBT) is supported by strong evidence as an effective intervention for improving sleep outcomes [10], and psychoeducational approaches have shown promise in promoting healthy sleep behaviors in nonclinical populations [11,12]. In university settings, psychological interventions have often focused on sleep-related symptoms associated with anxiety or depression rather than on sleep health promotion itself [13]. In addition, most studies have relied primarily on self-reported outcomes, with limited attention to objective indicators of circadian functioning such as motor activity, thermoregulation, and environmental synchronization (ES) [14]. This gap is particularly relevant for nursing students, who frequently experience sleep disturbances [15] and may benefit from behavioral interventions targeting maladaptive sleep-related habits and beliefs [16]. Although the efficacy of cognitive behavioral therapy for insomnia (CBT-I) is well established in clinical populations, less is known about its effects among nonclinical university students when delivered in group-based formats. In addition, most intervention studies in this field rely primarily on measures such as the Pittsburgh Sleep Quality Index (PSQI) [17]. Ambulatory circadian monitoring (ACM) offers the possibility of continuously assessing sleep-wake patterns and circadian regulation in real-world settings, providing a more comprehensive evaluation of behavioral intervention effects than self-report measures alone [18-21].

Although there are studies associated with poor sleep quality in nursing students [15], the evidence regarding the feasibility and acceptability of brief, group-based CBT-I interventions implemented as part of university health promotion programs for first-year nursing students remains limited. Besides, no previous studies have designed and evaluated interventions to improve sleep quality in this group using a wearable wristband device. Therefore, this pilot study aimed to explore the preliminary impact of a group-based CBT intervention on sleep quality among first-year nursing students. Specifically, the objectives were (1) to explore the preliminary effects of the intervention on sleep quality outcomes and (2) to assess participant satisfaction with the intervention. We hypothesized that, compared with the control group, participants receiving the intervention would show preliminary improvements in sleep quality from baseline to postintervention, as assessed through ACM and self-reported sleep quality.

Methods

Design

This study is a single-center, single-blind (data analysts), 2-arm pilot randomized clinical trial comparing 2 conditions: the intervention group, which receives a CBT-based intervention targeting sleep quality, and the control group, which receives no intervention. The randomization procedure was performed using a computer-generated random number sequence (Sealed Envelope software, version 1.21.0) and was assigned to participants' academic record numbers to randomly designate participants to 1 of the 2 groups (intervention or control group). Random permuted blocks were also used to reduce the predictability of the random sequence and guarantee a 1:1 allocation ratio. This sequence was stored in a password-protected table and was hidden from other researchers during the study.

The study is registered at ClinicalTrials.gov with reference number NCT05273086, and the study protocol has been published [22]. This study was performed according to the CONSORT (Consolidated Standards of Reporting Trials) statement (Checklist 1).

Study Setting, Recruitment, and Sampling

The study was conducted at a private university located in Madrid (Spain) during the 2022 to 2023 academic year. Participants were recruited in October 2022, and the study lasted from October 2022 to March 2023, when the postmeasure tests were carried out.

A convenience sampling method was used. Eligible students were informed about the study through an in-class presentation delivered by the research team, during which the study aims, procedures, duration, eligibility criteria, voluntary nature of participation, and confidentiality safeguards were explained. Students who expressed interest in participating were screened according to the inclusion and exclusion criteria. Those who met the eligibility criteria and agreed to participate provided written informed consent before baseline assessment and randomization.

The eligibility criteria of this study were as follows: the inclusion criteria were students aged between 18 and 25 years who were enrolled in the first year of a nursing degree during

XSL•FO
RenderX

the 2022 to 2023 academic year. Participants were excluded if they were diagnosed with a prior mental pathology, a sleep disorder, or both, and were treated with or without medication (hypnotics, sedatives, or melatonin); were taking medications to aid sleep; or were combining work and study.

Participants were not restricted to individuals with clinically diagnosed sleep disorders because the intervention was conceived as a universal sleep health promotion program for first-year nursing students. The objective was to promote healthier sleep-related habits and behaviors in a population exposed to academic demands and future shift-work schedules, rather than to diagnose or treat clinical or subclinical sleep disorders. This approach is consistent with current strategies aimed at promoting healthy sleep at the population level [23].

Sample Size Calculation

This study was designed as a pilot randomized controlled trial. Pilot studies are intended to assess feasibility and estimate key parameters for the design of a definitive trial, such as outcome variability and effect size estimates. Therefore, a formal a priori sample size calculation based on statistical power was not performed.

The methodological literature [24,25] on pilot studies suggests sample sizes ranging from 12 to 35 participants per group for this type of design. Although the protocol planned to recruit 60 participants, only 40 eligible students were enrolled during the recruitment period. This sample size also allows for the estimation of preliminary parameters with reasonable precision, in line with approaches focused on estimation rather than hypothesis testing. The findings of this study may inform the sample size calculation of future fully powered randomized controlled trials.

Intervention

Control Group

The control group continued with its normal routine. They did not attend any sessions; they only completed the pre-post measurements.

Intervention Group

Participants allocated to the intervention group received a structured group-based CBT-I–informed program specifically adapted for first-year nursing students within a university health promotion context. The intervention protocol was developed by the research team based on established cognitive behavioral sleep intervention principles and was previously described in the published study protocol [22]. The protocol integrated evidence-based components, including sleep psychoeducation, cognitive restructuring, behavioral stimulus control, and sleep hygiene strategies, delivered using active-constructive learning methodologies designed to promote participant engagement and self-reflection. The intervention group was subdivided into 2 groups of 10 participants to facilitate interaction and individualized discussion. The program consisted of 2 face-to-face 90-minute sessions delivered over 2 consecutive weeks during November and December 2022.

The first session focused on sleep psychoeducation and cognitive restructuring, including basic concepts of sleep physiology and circadian regulation, identification of biological and behavioral factors influencing sleep, and guided self-assessment of individual sleep patterns and habits. Participants were encouraged to reflect on their own sleep-related behaviors between sessions. The second session focused on behavioral skill development, including sleep hygiene education, stimulus control strategies, and practical behavioral adaptations tailored to participants' self-reported routines and baseline sleep assessments.

The intervention was delivered by a PhD candidate in biology, specialized in sleep neurophysiology, chronobiology, and circadian rhythms, and by a registered nurse (PhD) with expertise in health promotion and healthy lifestyle education. Both providers had training in sleep health and cognitive behavioral strategies for sleep. Intervention materials and procedures were reviewed by the research team before implementation to ensure consistency with the published protocol. Additional details regarding session content, intervention materials, and implementation procedures are provided in Multimedia Appendix 1.

Outcome Measures

Primary Outcomes

In accordance with the trial registration (ClinicalTrials.gov NCT05273086), the prespecified primary outcomes were objective sleep quality, assessed using the Kronowise 3.0 ACM system, and subjective sleep quality, assessed using the PSQI.

The Kronowise system provides a multidimensional assessment of sleep quality and circadian function through validated continuous circadian rhythm and sleep parameters, rather than a single global categorical score. Accordingly, the objective sleep outcome was characterized using these continuous parameters, which together describe sleep quality and circadian health.

In this study, sleep and circadian rhythms were monitored using the Kronowise 3.0 device (Kronohealth, S.L.) for 7 consecutive days (including weekends) in participants' homes during the preintervention period and for another 7 consecutive days during the postintervention period, in both the intervention and control groups, with a difference of 19 days between the 2 monitoring periods. The technical characteristics of the device have been described in several studies [14,26,27], and the analysis of its monitoring parameters was obtained from a population of 244 healthy volunteers aged 14 to 86 years using the Kronowizard platform and the Circadianware program, following the methodology of Madrid-Navarro et al [27] from the University of Murcia.

The primary signals recorded by the Kronowise system were surface temperature, acceleration, position, and exposure to light (TAPL). Peripheral skin temperature of the wrist is a recognized, validated circadian rhythm marker [28], suggesting that an increase in peripheral skin temperature is associated with drowsiness [29], whereas a decrease in core temperature is due to redistribution of heat from the interior to the periphery [30]. The triaxial motor acceleration shows the intensity of physical

XSL•FO
RenderX

activity and the time in movement, calculated as the sum of the absolute values of 300 acceleration vectors. Changes in posture reflect postural changes according to inclination axes (horizontality vs verticality) to identify immobility and the amplitude of movements during periods of activity vs rest. Light exposure is measured in 3 spectral bands (visible, blue [460 - 480 nm], and infrared [>700 nm]). All these variables were recorded at a sampling rate of 10 Hz (for acceleration and time in motion), 1 Hz (for skin temperature and light exposure), and 0.033 Hz for wrist position, generating 23 million raw data points over 7 days, condensed to 230,000, which were stored in 30-second stretches for later analysis. From these primary variables, the integrated variable TAPL is obtained as a modification of thermometry, activity, and body position, which expresses the level of general physical and mental activation and has been proved to be accurate for the estimation of individual sleep variables [20]. In TAPL, temperature is inverse to acceleration, position, and light exposure and, in global terms, acquires values between 0 and 1. Absolute values of 0 or close to 0 indicate low general activation (with high values of skin surface temperature), whereas high values close to or equal to 1 reflect high levels of activation corresponding to a maximum state of wakefulness, exposure to light, and movement (with low values of temperature). Given the objective of the investigation, TAPL V5 was used to obtain the mean value of this variable during the 5 consecutive hours of minimum values, that is, during sleep. TAPL V5 values range from 0.01 (5th percentile) to 0.07 (95th percentile), with a positive effect if it reaches lower values.

Circadian Parameters

The analysis of circadian parameters is related to sleep quality, and alterations may indicate chronodisruption [7]. All of these parameters range from 0 to 1, and their interpretation is based on previously published standardization procedures and normative values, including the 5th and 95th percentile reference values [20,21,26,27]. The interdaily stability (IS) regularity index, which is an indicator of the interdaily variability of the circadian rhythm, was calculated. The IS requires repeated measurements over several days because low values or values close to 0 indicate greater sleep fragmentation, whereas high values or values close to 1 indicate perfect stability. Relative amplitude normalized (RAN) is the normalized relative amplitude of the rhythm, calculated as the difference between M10 (average of the 10 consecutive hours with the maximum values of variables that rise during wakefulness) and L5 (average of the 5 consecutive hours with the minimum values of variables that fall during sleep), meaning that the higher the amplitude, the more robust the rhythm. The ES is the parameter related to the person's chronotype and indicates the degree of synchronization between the period of sleep or rest and the center of natural darkness, taking the official summertime of the country where the participant lives as a reference. In Spain, this central time of darkness during daylight saving time is 2 AM. Low values or values close to 0 indicate that the central hour of sleep is far from the central hour of darkness, whereas high values or values close to 1 indicate that the center of sleep is close to or coincides with the central hour of darkness. The main circadian rhythm robustness parameter is the circadian health score (CHS), described in 2019 [26], which integrates IS, RAN, and ES. CHS is the primary sleep quality indicator of the ACM technique and the main marker of circadian strength, which behaves inversely to chronodisruption when it has a value close to 1, implying a high-amplitude, nonfragmented, and stable rhythm.

Sleep Quality Parameters

Although sleep-wake parameters were detected with the above indices, specific sleep indicators were analyzed. These indicators were temperature 2 hours before sleep, temperature during sleep, exposure to total light and blue light 2 hours before sleep, sleep latency and actual sleep time, sleep activity time, number of awakenings, and sleep efficiency. Sleep latency refers to the time in minutes to sleep onset, actual sleep time is the time spent objectively asleep, and sleep activity time is expressed as the average time in seconds during which movement is detected. The number of awakenings is the average number of movements detected for ≥30 seconds per hour of the recorded sleep interval, and sleep efficiency refers to the percentage of time asleep as opposed to the time in bed.

Chronotype

Chronotype refers to categories defined by the differences in activity and alertness between morning and evening. These categories are determined in relation to the 2019 Munich test, which classifies them into extreme matutine, matutine, indefinite, vespertine, and extreme vespertine [31].

Subjective Sleep Quality

The PSQI [17] contains 19 items and 7 clinically important components related to sleep quality: subjective sleep quality, sleep latency, sleep duration, sleep efficiency, sleep disturbances, use of sleeping medication, and daytime dysfunction. Responses are reported based on a 5-point Likert-type scale ranging from 0 to 4. Overall, sleep scores of 5 or less are considered good quality, whereas scores of 5 or more are considered low quality. The PSQI version used is validated in the Spanish university framework by de la Vega et al [32] with a Cronbach α of 0.72. The questionnaire was delivered to all participants twice (before and immediately following the intervention).

Independent Variables

Independent variables were selected to characterize factors known to influence sleep quality and circadian regulation in young adults and university populations, including sociodemographic characteristics, lifestyle habits, physical activity, and contextual factors that may act as potential confounders or moderators of sleep-related outcomes. Satisfaction was additionally assessed to explore the acceptability of the intervention in this pilot study. Questionnaires on sociodemographic data and lifestyle factors were prepared ad hoc by the researchers.

First, sociodemographic data and anthropometric variables included date of birth, age, sex, and sleep habits during the study, including whether the participant slept alone or with someone. Participants reported their weight (kg) and height (cm) for BMI calculation. Second, lifestyle factors included

XSL•FO
RenderX

tobacco consumption (cigarettes per day), alcohol consumption (daily, weekly, and sporadic), drug consumption (yes/no and type), consumption of stimulating drinks (eg, coffee, sugary drinks, and energy drinks), and consumption of nonstimulating sweetened beverages. Third, physical activity was assessed using the simplified version of the International Physical Activity Questionnaire (IPAQ) [33]. The questionnaire consists of 7 questions assessing frequency, duration, and intensity of moderate and intense physical activity during the previous 7 days, as well as walking and sitting time during a working day. The questionnaire is classified into low, moderate, and high physical activity based on metabolic equivalent task minutes per week. The higher the score, the more physically active the profile. The questionnaire has been validated, with a mean reliability of 0.80 [33]. Fourth, satisfaction with the program was assessed using a questionnaire adapted from the proposal by Azpeleta et al [34], consisting of 12 questions with responses based on a 5-point Likert scale. Higher scores indicated greater satisfaction.

Data Collection

Data collection was conducted at 2 time points: baseline (preintervention) and immediately after completion of the intervention (postintervention). At baseline, eligible participants completed self-administered questionnaires assessing sociodemographic characteristics, lifestyle factors, physical activity, and sleep quality through the PSQI. Following questionnaire completion, participants were fitted with the Kronowise 3.0 wearable device and received standardized instructions for its use during the 7-day ambulatory monitoring period.

Participants wore the device continuously for 7 consecutive days, including weekdays and weekends, during their usual daily routines in their home environment. At the end of each monitoring period, devices were returned to the research team for data download and processing using the Kronowizard platform.

After baseline data collection, participants were randomized to either the intervention or control group. Following completion of the intervention period, both groups underwent a second 7-day monitoring period under the same conditions and completed the postintervention PSQI. Participants allocated to the intervention group additionally completed the satisfaction questionnaire immediately after the intervention. All assessments and device management procedures were coordinated by an independent investigator trained in study procedures and not involved in outcome analysis.

Ethical Considerations

This study was approved by the Research Ethics Committee of the Camilo José Cela University (code: 06 - 22-UCJC-Sleep) on April 1, 2022. All participants were informed of the duration and characteristics of the study, as well as its voluntary nature. No financial or nonfinancial compensation was provided to participants for study participation.

Participants were free to leave the study at any time, without further consequences. The utmost professionalism and absolute confidentiality were always ensured, in accordance with the European Regulation (EU) 2016/679 on the protection of natural persons and the processing and free movement of data, the Framework Law 3/2018 on the protection of personal data and guarantee of digital rights, and the Law 14/2007 on biomedical research. Only the main project researcher had access to the encrypted database, data processing, and free movement, in accordance with the Framework Law 3/2018 on the protection of personal data and guarantee of digital rights and Law 14/2007 on biomedical research. Only the main project researcher has access to the encrypted database.

Data Analysis

A descriptive analysis was made of all the variables included in the study. Categorical variables were analyzed using absolute and relative frequencies, whereas continuous variables were described using mean (SD) if the variable followed a normal distribution or median (IQR) if the variable followed a nonnormal distribution. Normality was checked using graphic methods (histogram) and statistical methods (Kolmogorov-Smirnov test or Shapiro-Wilk test).

A bivariate analysis was performed to compare the control group and the intervention group, using the Student *t* test (1-tailed) or the Mann-Whitney *U* test for continuous variables according to the distribution of the variable and the chi-square test or Fisher exact test for categorical variables when necessary. For pre-post comparison, the Student *t* test was used for paired samples, the Wilcoxon test was used for continuous variables, and the McNemar test was used for categorical variables. Effect size for parametric comparisons was estimated using Cohen *d* [35]. For nonparametric analyses, effect size was calculated as *r* based on the standardized *Z* statistic ($r = Z / \sqrt{N}$) proposed by Rosenthal and Rubin [36], where N represents the total sample size. Effect sizes were interpreted according to Cohen conventional thresholds for *d* (0 - 0.3 for small, >0.3 - 0.8 for moderate, and >0.8 for large effect) and *r* (0 - 0.3 for small, >0.3 - 0.5 for moderate, and >0.5 for large effect).

To account for baseline imbalances between groups, adjusted analyses were additionally performed using analysis of covariance. For each continuous sleep and circadian outcome, the postintervention value was entered as the dependent variable, treatment group (intervention vs control) as the fixed factor, and the corresponding baseline value of the outcome as a covariate. Smoking status, school schedule, and coffee consumption were additionally included as covariates because these variables differed between groups at baseline. Adjusted treatment effects (regression coefficients [β]), 95% CI, *P* values, and partial eta-squared (η^2) were estimated. Because baseline differences were identified between the intervention and control groups, the adjusted analysis of covariance models were considered the primary analytical approach for estimating treatment effects, whereas the unadjusted bivariate analyses were retained as complementary analyses.

In accordance with the trial registration, the prespecified primary outcomes were objective sleep quality, assessed using the Kronowise system, and subjective sleep quality, assessed using the PSQI. The continuous circadian rhythm and sleep parameters generated by the Kronowise system were analyzed to provide a detailed characterization of the objective sleep outcome. Given

XSL•FO
RenderX

the pilot nature of this trial, analyses of these individual parameters were considered exploratory and hypothesis-generating. Therefore, no formal adjustment for multiple comparisons was applied, and findings from these analyses should be interpreted as exploratory rather than confirmatory.

Results

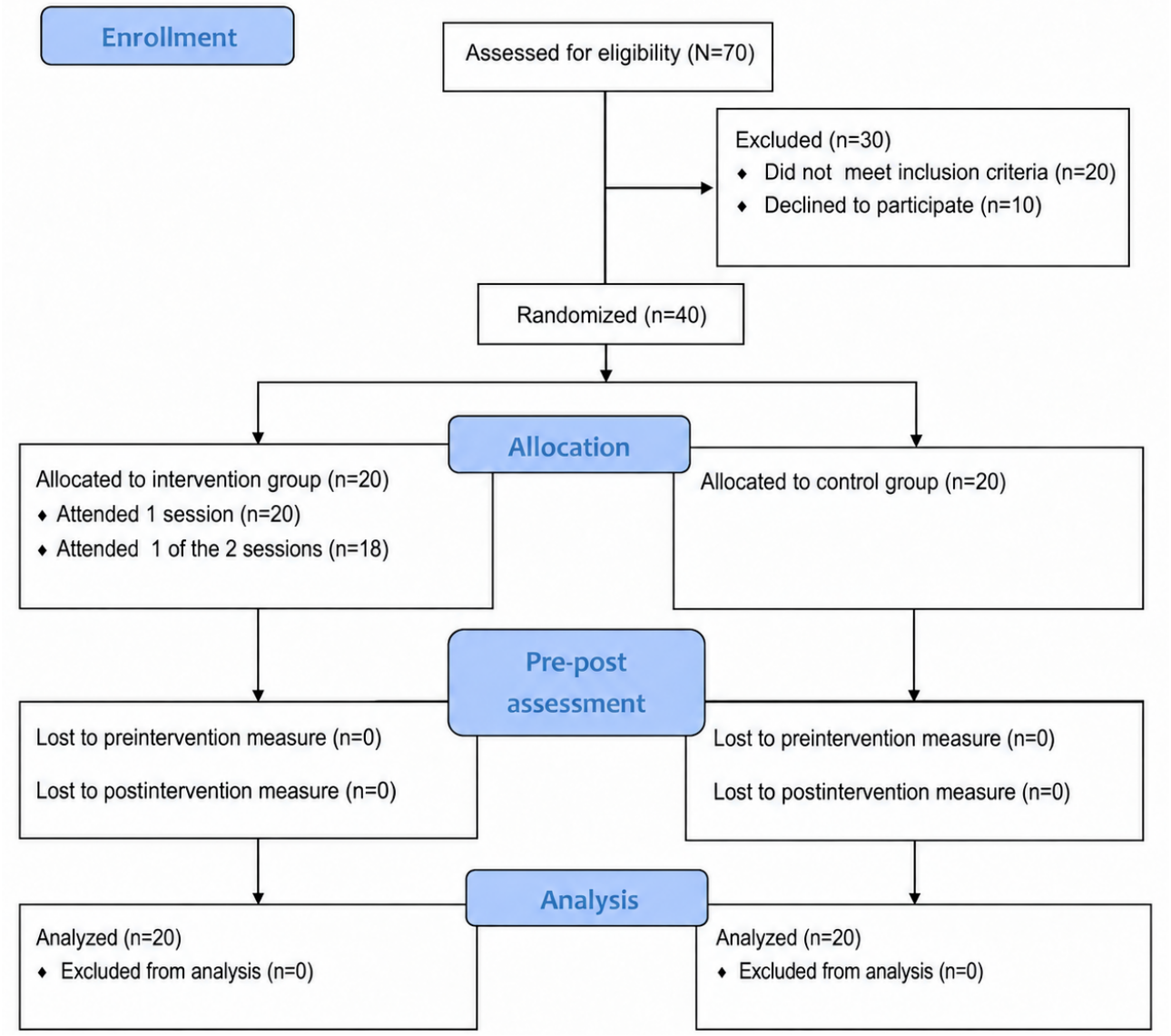
Figure 1 shows the CONSORT flow diagram of the study. A total of 70 students were assessed for eligibility. Twenty students were excluded because they did not meet the study criteria, most frequently due to melatonin use, combining work and study, or being outside the eligible age range. Of the 50 eligible students, 10 declined participation or did not complete the informed consent process. Reasons for nonparticipation were not systematically recorded. Consequently, 40 participants were enrolled and randomized. Of these, 20 were allocated to the intervention group (n=16, 80% female) and 20 to the control group (n=18, 90% female). The overall mean age was 19.9 (SD 1.8) years. The mean age was lower in the intervention group compared with the control group (mean 19.6, SD 1.8 vs mean 20.2, SD 1.8, respectively). Cohabitation of students during the study was mostly in a family setting (n=12, 60%) compared with 8 (40%) students residing outside the family environment, with no statistically significant difference.

The total sample was distributed in academic dedication schedules of 50% (20/40) morning shifts (8:30 AM to 1:30 PM) and 50% (20/40) in afternoon shifts (3:30 PM to 8:30 PM), with a statistically significant relationship (P=.01). The most prevalent BMI category in both groups was normal weight (31/40, 77.5%), with a mean of 23.1 (SD 3.8) kg/m^2 (Table 1).

Among 40 students, 37.5% (n=15) were active smokers, with a higher proportion of smokers in the intervention group than in the control group, a mostly daily consumption pattern (n=10, 66.7%), and a statistically significant median daily consumption (P=.04). A total of 67.5% (n=27) consumed alcohol on a regular basis, with no significant differences between the 2 research groups (Table 2). Almost 8 out of 10 participants reported daily coffee consumption, with higher consumption in the control group (18/20, 90%) than in the intervention group (12/20, 60%), which was statistically significant (P=.03), with a median consumption of close to 2 coffees per day in both groups. Almost 3 out of 10 students reported consuming energy drinks daily, with slightly higher consumption in the control group than in the intervention group (5/20, 25% vs 4/20, 20%). Overall, 45% of the participants were sedentary, with a median of 7 hours sitting per day. The sedentary condition was the highest proportion in both groups in terms of the type of physical activity they engaged in daily.

XSL•FO
RenderX

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram.



XSL•FO
RenderX

Table . Sociodemographic data and anthropometric variables.

Sociodemographic characteristics	Global (n=40)	Intervention group (n=20)	Control group (n=20)	*P* value
Age, mean (SD)	19.9 (1.8)	19.6 (1.8)	20.2 (1.8)	.29
Sex, n (%)				.66
Female	34 (85)	16 (80)	18 (90)	
Male	6 (15)	4 (20)	2 (10)	
Class schedule, n (%)				.01
Morning	20 (50)	14 (70)	6 (30)	
Afternoon	20 (50)	6 (30)	14 (70)	
Cohabitation during the study, n (%)				.23
Alone	12 (30)	5 (25)	7 (35)	
In a relationship	1 (2.5)	0 (0)	1 (5)	
With family	24 (60)	12 (60)	12 (60)	
Shared accommodation	3 (7.5)	3 (15)	0 (0)	
BMI (kg/m^2), mean (SD)	23.1 (3.8)	23.5 (4.3)	22.7 (3.2)	.52
BMI (kg/m^2), n (%)				.67
Low	1 (2.5)	0 (0)	1 (5)	
Normal	31 (77.5)	16 (80)	15 (75)	
Overweight	5 (12.5)	2 (10)	3 (15)	
Obese	3 (7.5)	2 (10)	1 (5)	

XSL•FO
RenderX

Table . Baseline sociodemographic characteristics and lifestyle factors of first-year nursing students participating in a pilot randomized controlled trial evaluating a group-based cognitive behavioral therapy for insomnia (CBT-I) intervention (Spain, 2022 - 2023)[a].

Lifestyle	Global (n=40)	Intervention group (n=20)	Control group (n=20)	*P* value
With a diagnosed illness, n (%)	6 (15)	2 (10)	4 (20)	.66
On pharmacological medication, n (%)	4 (10)	1 (5)	3 (15)	.61
Smokers, n (%)	15 (37.5)	8 (40)	7 (35)	.74
Consumption (n=15), n (%)				.63
Daily	10 (66.7)	5 (62.5)	5 (71.4)	
Weekly	4 (26.7)	2 (25)	2 (28.6)	
Monthly	1 (6.7)	1 (12.5)	0 (0)	
Tobacco (n=15), n (%)				.96
Factory made	7 (46.7)	4 (50)	3 (42.9)	
Rolled cigarettes	6 (40)	3 (37.5)	3 (42.9)	
Electronic devices	2 (13.3)	1 (12.5)	1 (14.3)	
Cigarettes per day, median (IQR); range	2 (0-10); 0-10	0 (0-3.5); 0-5	10 (0-10); 0-10	.04
Alcohol consumption, n (%)	27 (67.5)	13 (65)	14 (70)	.74
Alcohol type[b] (n=27), n (%)				
Beer	21 (77.8)	11 (84.6)	10 (71.4)	.68
Wine	15 (55.6)	8 (61.5)	7 (50)	.55
Spirits	13 (48.1)	7 (53.8)	6 (42.9)	.57
Liqueurs	4 (14.8)	3 (23.1)	1 (7.1)	.33
Alcohol volume (cm^3/mL) per week, median (IQR); range	300 (175-600); 50 - 2000	350 (212.5-800); 125 - 2000	275 (125-531.25); 50 - 1025	.31
Coffee consumption, n (%)	30 (75)	12 (60)	18 (90)	.03
Coffee consumption per day, median (IQR); range	0.5 (1-2); 1-4	1.5 (1-2.75); 1-4	1.5 (1-2); 1-3	.78
Daily energy drink consumption, n (%)	9 (22.5)	4 (20)	5 (25)	>.99
Energy drink types (n=9), n (%)				.46
Monster	6 (66.7)	2 (50)	4 (80)	
Caffeine	2 (22.2)	1 (25)	1 (20)	
Red Bull	1 (11.1)	1 (25)	0 (0)	
Daily consumption of non–energy-sweetened beverages, n (%)	12 (30)	5 (25)	7 (35)	.49
Beverages drink types (n=12), n (%)				>.99
Cola soft drinks	8 (66.7)	3 (60)	5 (71.4)	
Other soft drinks	4 (33.3)	2 (40)	2 (28.6)	
Hours seated per day, median (IQR); range	7 (5-9); 1-15	7.5 (5-8); 1-15	7 (4-10); (1-15)	.66
Physical activity type, n (%)				.57
Low or sedentary	18 (45)	8 (40)	10 (50)	

XSL•FO
RenderX

Lifestyle	Global (n=40)	Intervention group (n=20)	Control group (n=20)	*P* value
High	11 (27.5)	7 (35)	4 (20)	
Moderate	11 (27.5)	5 (25)	6 (30)	

[a]The total sample included 40 participants (20 per group). For behavior-specific variables, such as alcohol and tobacco consumption, data refer only to participants reporting the corresponding behavior; therefore, frequencies may not add up to the total sample. Continuous variables with nonnormal distributions are presented as median (Q1–Q3); range.

[b]For alcohol type, participants could report more than one category; therefore, frequencies and percentages do not necessarily add up to the total sample or 100%.

Adjusted treatment effects for circadian rhythm, objective sleep, and subjective sleep outcomes are presented in Table 3. For objective circadian rhythm parameters, adjusted treatment effects were uniformly small (partial η^2=0.004 - 0.047), and the corresponding 95% CIs were compatible with little or no intervention effect. Similarly, objective sleep and environmental outcomes showed generally small effect sizes (partial η^2=0.004 - 0.065), with CIs indicating considerable uncertainty around the estimated treatment effects. In contrast, subjective sleep outcomes showed larger adjusted effects. The intervention was associated with a shorter self-reported sleep latency (β=−19.78 min, 95% CI −33.83 to −5.74), corresponding to a large effect size (partial η^2=0.194). Sleep duration (β=36.32 min, 95% CI −5.01 to 77.64; partial η^2=0.086), sleep efficiency (β=8.75%, 95% CI −0.97 to 18.47; partial η^2=0.090), and the global PSQI score (β=−1.20, 95% CI −3.18 to 0.78; partial η^2=0.061) also favored the intervention group, although the CIs were compatible with a range of plausible effects. Unadjusted bivariate analyses are provided in Multimedia Appendix 1.

Objective sleep monitoring indicated that the predominant baseline chronotype in both groups was intermediate (ie, neither clearly morning nor evening type). Following the intervention, a shift toward a more evening-oriented chronotype was observed in both groups (Table 4).

Satisfaction with the intervention received and with the participation of the expert (Table 5) was mostly “very good” in all the items evaluated. Satisfaction with the intervention was very good for 63% of the survey and good for 35%.

XSL•FO
RenderX

Table . Adjusted treatment effects of the group-based cognitive behavioral therapy for insomnia (CBT-I) intervention on circadian rhythm, objective sleep, and subjective sleep outcomes[a].

Outcome	Adjusted treatment effect (IG[b] vs CG[c]), β (95% CI)	*P* value	Partial η^2
Circadian rhythm			
TAPL[d] V5	−0.004 (−0.030 to 0.021)	.73	0.004
Interdaily stability (IS)	−0.051 (−0.130 to 0.029)	.25	0.047
Relative amplitude (RAN)	−0.010 (−0.054 to 0.034)	.66	0.006
Environmental synchronization	0.042 (−0.032 to 0.116)	.26	0.037
Circadian health score (CHS)	0.009 (−0.036 to 0.053)	.69	0.005
Objective sleep outcomes			
Temperature 2 h before bedtime	−0.329 (−0.915 to 0.254)	.26	0.037
Temperature during sleep	0.217 (−0.234 to 0.668)	.34	0.027
White light exposure 2 h before bedtime	−0.680 (−3.184 to 1.824)	.59	0.009
Blue light exposure 2 h before bedtime	−0.035 (−0.162 to 0.093)	.59	0.009
Sleep latency	5.63 (−1.80 to 13.06)	.69	0.065
Total sleep time	−7.56 (−47.17 to 32.04)	.70	0.004
Sleep activity time	−0.208 (−0.125 to 0.540)	.21	0.045
Number of awakenings	−0.157 (−0.521 to 0.207)	.39	0.022
Wake after sleep onset	5.14 (−7.81 to 18.10)	.43	0.019
Sleep efficiency	−1.93 (−4.64 to 0.79)	.16	0.058
Results of PSQI[e]			
Sleep latency (min)	−19.78 (−33.83 to −5.74)	.007	0.194
Time in bed	−0.45 (−39.56 to 38.65)	.98	0.001
Sleep time	36.32 (−5.01 to 77.64)	.08	0.086
Sleep efficiency	8.75 (−0.97 to 18.47)	.08	0.090
Global PSQI score	−1.20 (−3.18 to 0.78)	.22	0.061

[a]Adjusted treatment effects were estimated using analysis of covariance models with the postintervention value as the dependent variable, treatment group as the fixed factor, and the corresponding baseline value, smoking status, school schedule, and coffee consumption as covariates. Partial η^2 was interpreted as <0.01 (trivial), 0.01 to <0.06 (small), 0.06 to <0.14 (moderate), and ≥0.14 (large).

[b]IG: intervention group.

[c]CG: control group.

[d]TAPL: temperature, acceleration, position, and exposure to light.

[e]PSQI: Pittsburgh Sleep Quality Index.

XSL•FO
RenderX

Table . Changes in chronotype distribution before and after a group-based cognitive behavioral therapy for insomnia (CBT-I) intervention among first-year nursing students participating in a pilot randomized controlled trial[a].

Chronotype	Preintervention		Postintervention	
	IG[b] (n=20), n (%)	CG[c] (n=20), n (%)	IG (n=20), n (%)	CG (n=20), n (%)
Extreme matutine	—[d]	—	—	—
Matutine	1 (5)	0 (0)	1 (5)	2 (10)
Undefined	12 (60)	12 (60)	8 (40)	8 (40)
Vespertine	4 (20)	5 (25)	8 (40)	4 (20)
Extreme vespertine	3 (15)	3 (15)	3 (15)	3 (15)

[a]Pre-post *P* values are IG=.30 and CG=.47.

[b]IG: intervention group.

[c]CG: control group.

[d]Not applicable.

Table . Results of satisfaction with the intervention[a].

Variables	Very good (n=20), n (%)	Good (n=20), n (%)	Regular (n=20), n (%)	Bad (n=20), n (%)	Very bad (n=20), n (%)
General assessment[b]					
Dissemination before commencing intervention	9 (45)	11 (55)	0 (0)	0 (0)	0 (0)
Briefing before commencing intervention	7 (35)	13 (65)	0 (0)	0 (0)	0 (0)
Explanation of study goals	16 (80)	4 (20)	0 (0)	0 (0)	0 (0)
Communication between participants and expert	15 (75)	5 (25)	0 (0)	0 (0)	0 (0)
Audiovisual media and materials used	14 (70)	6 (30)	0 (0)	0 (0)	0 (0)
Assessment of intervention expert[b]					
Clarity of presentation	18 (90)	2 (10)	0 (0)	0 (0)	0 (0)
Expert's knowledge on intervention content	15 (75)	5 (25)	0 (0)	0 (0)	0 (0)
Participant's motivation level	10 (50)	10 (50)	0 (0)	0 (0)	0 (0)
Adequate number of study activities	11 (55)	8 (40)	1 (5)	0 (0)	0 (0)
Counseling and support during the intervention	12 (60)	8 (40)	0 (0)	0 (0)	0 (0)
Technological resources and applications used	9 (45)	11 (55)	0 (0)	0 (0)	0 (0)
Overall satisfaction with the intervention	13 (65)	7 (35)	0 (0)	0 (0)	0 (0)

[a]Effect size *r*: 0-0.3 for low, >0.3-0.5 for moderate, and >0.5 for big.

[b]Participants responded to multiple satisfaction items; therefore, the frequencies within each response category do not add up to the total sample (n=20). For each individual item, responses add up to 20 participants.

XSL•FO
RenderX

Discussion

Principal Findings

The present pilot randomized controlled trial explored the preliminary effects of a brief group-based CBT-I–informed intervention delivered as a sleep health promotion program for first-year nursing students. Overall, the intervention showed more favorable effects on subjective sleep outcomes, particularly self-reported sleep latency, whereas objective circadian rhythm and sleep parameters assessed through ACM showed small and imprecise treatment effects. Participants also reported high satisfaction with the intervention, supporting the feasibility and acceptability of implementing this type of program within a university setting.

The pattern of findings observed in this study may be explained by the characteristics of both the intervention and the study population. Unlike most CBT-I trials, which recruit participants with clinically significant insomnia symptoms, this intervention was delivered as a universal health promotion strategy to first-year nursing students regardless of their baseline sleep status [10,12,13]. Consequently, participants may have had relatively limited objective circadian impairment at baseline, reducing the opportunity to detect measurable physiological improvements over the short intervention period. At the same time, increasing participants' awareness of sleep-related behaviors and promoting healthier sleep habits may have influenced their perception of sleep before producing detectable changes in objective circadian regulation [37].

The characteristics of the intervention itself should also be considered when interpreting these findings. Standard CBT-I programs are typically delivered over multiple sessions and incorporate individualized behavioral components, including stimulus control, sleep restriction, cognitive therapy, and the systematic review of sleep diaries, which are considered core mechanisms of treatment effectiveness [38,39]. By contrast, our intervention consisted of 2 face-to-face group sessions designed as a pragmatic university-based health promotion program rather than as an intensive therapeutic intervention. This lower treatment intensity may have been sufficient to increase participants' awareness of healthy sleep practices and influence their perception of sleep while being insufficient to induce sustained behavioral changes that could be reflected in objectively measured circadian outcomes during the relatively short follow-up period [40,41]. Furthermore, participant engagement with the behavioral recommendations outside the classroom was not formally assessed. Consequently, it remains unclear to what extent the absence of objective effects reflects the limited intensity of the intervention, suboptimal adherence to the recommended behavioral changes, or a combination of these factors, all of which are recognized determinants of CBT-I effectiveness [39,42].

The discrepancy between subjective and objective outcomes deserves particular attention. Participants in the intervention group reported improvements in sleep latency, with trends toward better perceived sleep duration and sleep efficiency, whereas objective monitoring did not demonstrate corresponding improvements. Similar discrepancies between subjective and objective sleep measures have been described previously in young adults and may reflect the fact that these instruments capture different aspects of sleep [37]. The discrepancy between subjective and objective outcomes should also be interpreted in light of the study design. Because participants were aware of their group allocation and the control group received no active intervention, subjective outcomes were more susceptible to expectancy effects, increased awareness of sleep behaviors, facilitator attention, and group interaction. Consequently, the observed improvements in subjective sleep outcomes may partly reflect nonspecific effects associated with participation in the intervention, in addition to the specific therapeutic components of CBT-I [43]. In contrast, objectively measured circadian outcomes are less directly influenced by participants' expectations and therefore provide a complementary perspective on intervention effects.

Our findings are broadly consistent with previous studies evaluating CBT-I and behavioral sleep interventions in university students. Randomized trials and systematic reviews have generally reported greater improvements in subjective sleep outcomes than in objective sleep measures, particularly when interventions are implemented in nonclinical student populations [13,40,44]. However, direct comparisons should be interpreted cautiously because studies differ substantially in intervention intensity, treatment duration, participant characteristics, and outcome assessment. Moreover, most previous studies have relied primarily on self-reported questionnaires or conventional actigraphy [13]. By incorporating multidimensional ACM alongside validated subjective measures, the present study provides a more comprehensive evaluation of behavioral intervention effects under real-world conditions and highlights the value of assessing both perceived and physiological dimensions of sleep [45].

Although the objective circadian outcomes did not demonstrate consistent intervention effects, these findings should not be interpreted as evidence that behavioral sleep interventions are ineffective in university populations. Rather, they suggest that a brief group-based health promotion program may be insufficient to modify physiological circadian regulation over a short period in generally healthy young adults. Future trials should evaluate whether greater intervention intensity, individualized behavioral support, reinforcement sessions, longer follow-up, and systematic assessment of participant adherence improve both subjective and objective sleep outcomes [44]. Such research may also help clarify the temporal relationship between changes in perceived sleep and objectively measured circadian outcomes.

One of the main strengths of this study is the combined assessment of subjective sleep quality and multidimensional objective circadian monitoring using a validated ambulatory wearable device under free-living conditions. This approach enabled continuous assessment of sleep and circadian function beyond self-reported outcomes—an aspect that has been infrequently incorporated into behavioral sleep intervention studies in university populations. In addition, the randomized controlled design and the high levels of participant satisfaction support the feasibility and acceptability of implementing sleep health promotion programs among first-year nursing students.

XSL•FO
RenderX

Overall, this pilot trial suggests that a brief, group-based, CBT-I–informed intervention is feasible, well-accepted, and may improve selected subjective aspects of sleep among first-year nursing students. However, measurable changes in objective circadian outcomes were not observed under the conditions evaluated. These findings support the need for adequately powered randomized controlled trials incorporating active control groups, longer follow-up periods, adherence assessment, and more intensive or individualized behavioral interventions to better establish the effectiveness of sleep health promotion programs in university populations.

Limitations

One of the main limitations of this study was the lack of blinding for participants and facilitators, which is common in behavioral interventions and may introduce bias. Additionally, the absence of an active control group may affect internal validity and potentially lead to an overestimation of the intervention effects. Although the use of a usual-care control enhances ecological validity in a university context, it does not account for nonspecific effects inherent to group-based interventions, such as therapist attention, peer interaction, and participant expectations. Furthermore, the face-to-face group format itself may have contributed to the observed outcomes through contextual or placebo effects, which cannot be ruled out.

The inclusion of an active control condition (eg, group-based education sessions without CBT-I components) would have allowed for more rigorous control over these nonspecific effects. Therefore, the findings should be interpreted with caution. In addition, the group-based format of the intervention did not allow for individualized tailoring based on the specific sleep characteristics identified during preintervention assessment. Although key aspects of sleep behavior were addressed, more personalized approaches may yield stronger effects.

Furthermore, although mediation analysis was prespecified in the study protocol, it was not performed due to the limited sample size and the absence of multiple follow-up assessments required to establish temporal relationships between potential mediators and outcomes. As a result, the mechanisms through which the intervention may have influenced sleep- and circadian-related outcomes could not be explored. Future, adequately powered trials that incorporate repeated measurements over time should investigate potential mediating pathways underlying intervention effects.

In addition, the study may be subject to selection bias. Participants were recruited from a single cohort of first-year nursing students at one university, and participation was voluntary. Consequently, students with greater interest in sleep-related issues or health-promoting behaviors may have been more likely to participate. This may limit the representativeness of the sample and reduce the generalizability of the findings to other student populations, academic disciplines, or educational settings.

Although participants were randomly allocated, some baseline differences were observed between groups in variables such as smoking status, school schedules, and caffeine consumption. While additional adjusted analyses were performed and yielded results consistent with the primary findings, residual confounding cannot be completely excluded given the small sample size characteristic of this pilot study.

Future research should aim to incorporate active control conditions, larger sample sizes, and more individualized interventions. Further studies are also needed to determine the clinical relevance of changes observed in circadian parameters and to validate multidisciplinary approaches based on individual sleep and circadian profiles. Finally, this pilot trial was not intended to address implementation at the institutional level. Future implementation studies are needed to evaluate scalability, resource requirements, and cost-effectiveness before broader adoption can be considered.

Conclusions

This pilot randomized controlled trial provides preliminary evidence that a brief group-based CBT-I intervention may improve selected subjective sleep outcomes among first-year nursing students. The largest treatment effect was observed for self-reported sleep latency, whereas objective circadian rhythm and sleep outcomes generally showed small treatment effects with wide CIs, reflecting considerable uncertainty around the estimated effects. Participants reported high levels of satisfaction with the intervention, supporting the feasibility and acceptability of this approach within a university setting. Given the exploratory nature and limited sample size of this pilot study, adequately powered randomized controlled trials with longer follow-up are needed to confirm these findings and to better define the magnitude and durability of the intervention effects.

Acknowledgments

We thank Universidad Camilo José Cela for financial support. The authors used ChatGPT (OpenAI) solely to assist with English-language editing and to improve the readability of the manuscript. No generative AI tools were used for study design, data collection, data analysis, interpretation of results, or generation of scientific conclusions. All scientific content, analyses, interpretations, and final decisions remain the sole responsibility of the authors.

Funding

This research was supported by a grant from Camilo José Cela University.

Data Availability

The data used during this study are available from the corresponding author upon reasonable request.

XSL•FO
RenderX

Authors' Contributions

Conceptualization: CRZ, DP-M
Formal analysis: CRZ, DP-M
Investigation: CA, BG-I
Methodology: CRZ, DP-M
Supervision: CRZ, DP-M
Writing – original draft: CRZ, DP-M
Writing – review & editing: CA, BG-I

Conflicts of Interest

None declared.

Multimedia Appendix 1
Unadjusted bivariate analyses of circadian rhythm parameters, sleep, and subjective sleep quality.
[DOCX File, 37 KB - humanfactors_v13i1e85021_app1.docx]

Checklist 1
CONSORT 2010 checklist.
[PDF File, 168 KB - humanfactors_v13i1e85021_app2.pdf]

References

1. Wittmann M, Dinich J, Merrow M, Roenneberg T. Social jetlag: misalignment of biological and social time. Chronobiol Int 2006;23(1-2):497-509. [doi: 10.1080/07420520500545979] [Medline: 16687322]
2. Ohayon M, Wickwire EM, Hirshkowitz M, et al. National Sleep Foundation's sleep quality recommendations: first report. Sleep Health 2017 Feb;3(1):6-19. [doi: 10.1016/j.sleh.2016.11.006] [Medline: 28346153]
3. Kessler RC, Berglund PA, Coulouvrat C, et al. Insomnia, comorbidity, and risk of injury among insured Americans: results from the America Insomnia Survey. Sleep 2012 Jun 1;35(6):825-834. [doi: 10.5665/sleep.1884] [Medline: 22654202]
4. Nelson KL, Davis JE, Corbett CF. Sleep quality: an evolutionary concept analysis. Nurs Forum 2022 Jan;57(1):144-151. [doi: 10.1111/nuf.12659] [Medline: 34610163]
5. Mellinger GD, Balter MB, Uhlenhuth EH. Insomnia and its treatment. Prevalence and correlates. Arch Gen Psychiatry 1985 Mar;42(3):225-232. [doi: 10.1001/archpsyc.1985.01790260019002] [Medline: 2858188]
6. Howell AJ, Jahrig JC, Powell RA. Sleep quality, sleep propensity and academic performance. Percept Mot Skills 2004 Oct;99(2):525-535. [doi: 10.2466/pms.99.2.525-535] [Medline: 15560340]
7. Silva M, Chaves C, Duarte J, Amaral O, Ferreira M. Sleep quality determinants among nursing students. Procedia Soc Behav Sci 2016 Feb;217:999-1007. [doi: 10.1016/j.sbspro.2016.02.090]
8. Chan CKL, So WKW, Fong DYT. Hong Kong baccalaureate nursing students' stress and their coping strategies in clinical practice. J Prof Nurs 2009;25(5):307-313. [doi: 10.1016/j.profnurs.2009.01.018] [Medline: 19751936]
9. Krystal AD, Prather AA, Ashbrook LH. The assessment and management of insomnia: an update. World Psychiatry 2019 Oct;18(3):337-352. [doi: 10.1002/wps.20674] [Medline: 31496087]
10. Koffel EA, Koffel JB, Gehrman PR. A meta-analysis of group cognitive behavioral therapy for insomnia. Sleep Med Rev 2015 Feb;19:6-16. [doi: 10.1016/j.smrv.2014.05.001] [Medline: 24931811]
11. Irish LA, Kline CE, Gunn HE, Buysse DJ, Hall MH. The role of sleep hygiene in promoting public health: a review of empirical evidence. Sleep Med Rev 2015 Aug;22:23-36. [doi: 10.1016/j.smrv.2014.10.001] [Medline: 25454674]
12. Murawski B, Wade L, Plotnikoff RC, Lubans DR, Duncan MJ. A systematic review and meta-analysis of cognitive and behavioral interventions to improve sleep health in adults without sleep disorders. Sleep Med Rev 2018 Aug;40:160-169. [doi: 10.1016/j.smrv.2017.12.003] [Medline: 29397329]
13. Friedrich A, Schlarb AA. Let's talk about sleep: a systematic review of psychological interventions to improve sleep in college students. J Sleep Res 2018 Feb;27(1):4-22. [doi: 10.1111/jsr.12568] [Medline: 28618185]
14. Blazquez A, Martinez-Nicolas A, Salazar FJ, Rol MA, Madrid JA. Wrist skin temperature, motor activity, and body position as determinants of the circadian pattern of blood pressure. Chronobiol Int 2012 Jul;29(6):747-756. [doi: 10.3109/07420528.2012.679328] [Medline: 22734575]
15. Yilmaz D, Tanrikulu F, Dikmen Y. Research on sleep quality and the factors affecting the sleep quality of the nursing students. Curr Health Sci J 2017;43(1):20-24. [doi: 10.12865/CHSJ.43.01.03] [Medline: 30595850]
16. Edinger JD, Means MK. Cognitive-behavioral therapy for primary insomnia. Clin Psychol Rev 2005 Jul;25(5):539-558. [doi: 10.1016/j.cpr.2005.04.003] [Medline: 15951083]
17. Buysse DJ, Reynolds 3rd CF III, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]

XSL•FO
RenderX

18. Ortiz‐Tudela E, Martinez‐Nicolas A, Albares J, et al. Ambulatory circadian monitoring (ACM) based on thermometry, motor activity and body position (TAP): a comparison with polysomnography. Physiol Behav 2014 Mar;126:30-38. [doi: 10.1016/j.physbeh.2013.12.009] [Medline: 24398067]
19. Mullington JM, Abbott SM, Carroll JE, et al. Developing biomarker arrays predicting sleep and circadian-coupled risks to health. Sleep 2016 Apr 1;39(4):727-736. [doi: 10.5665/sleep.5616] [Medline: 26951388]
20. Ortiz-Tudela E, Martinez-Nicolas A, Campos M, Rol M, Madrid JA. A new integrated variable based on thermometry, actimetry and body position (TAP) to evaluate circadian system status in humans. PLoS Comput Biol 2010 Nov 11;6(11):e1000996. [doi: 10.1371/journal.pcbi.1000996] [Medline: 21085644]
21. Martinez-Nicolas A, Martinez-Madrid MJ, Almaida-Pagan PF, Bonmati-Carrion MA, Madrid JA, Rol MA. Assessing chronotypes by ambulatory circadian monitoring. Front Physiol 2019;10:1396. [doi: 10.3389/fphys.2019.01396] [Medline: 31824327]
22. Ruiz-Zaldibar C, Gal-Iglesias B, Azpeleta-Noriega C, Ruiz-López M, Pérez-Manchón D. The effect of a sleep intervention on sleep quality in nursing students: study protocol for a randomized controlled trial. Int J Environ Res Public Health 2022 Oct 25;19(21):13886. [doi: 10.3390/ijerph192113886] [Medline: 36360766]
23. Morin CM, Benca R. Chronic insomnia. Lancet 2012 Mar 24;379(9821):1129-1141. [doi: 10.1016/S0140-6736(11)60750-2] [Medline: 22265700]
24. Whitehead AL, Julious SA, Cooper CL, Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. Stat Methods Med Res 2016 Jun;25(3):1057-1073. [doi: 10.1177/0962280215588241] [Medline: 26092476]
25. Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharm Stat 2005 Oct;4(4):287-291. [doi: 10.1002/pst.185]
26. Madrid-Navarro CJ, Puertas Cuesta FJ, Escamilla-Sevilla F, et al. Validation of a device for the ambulatory monitoring of sleep patterns: a pilot study on Parkinson's disease. Front Neurol 2019;10:356. [doi: 10.3389/fneur.2019.00356] [Medline: 31031690]
27. Madrid-Navarro CJ, Escamilla-Sevilla F, Mínguez-Castellanos A, et al. Multidimensional circadian monitoring by wearable biosensors in Parkinson's disease. Front Neurol 2018;9:157. [doi: 10.3389/fneur.2018.00157] [Medline: 29632508]
28. Sarabia JA, Rol MA, Mendiola P, Madrid JA. Circadian rhythm of wrist temperature in normal-living subjects. A candidate of new index of the circadian system. Physiol Behav 2008 Nov 28;95(4):570-580. [doi: 10.1016/j.physbeh.2008.08.005] [Medline: 18761026]
29. Raymann RJEM, Swaab DF, Van Someren EJW. Cutaneous warming promotes sleep onset. Am J Physiol Regul Integr Comp Physiol 2005 Jun;288(6):R1589-R1597. [doi: 10.1152/ajpregu.00492.2004] [Medline: 15677527]
30. Carskadon MA, Dement WC. Normal human sleep: an overview. In: Kryger MH, Roth T, Dement WC, editors. Principles and Practice of Sleep Medicine: Elsevier; 2011:16-26. [doi: 10.1016/B978-1-4160-6645-3.00002-5]
31. Juda M, Vetter C, Roenneberg T. The Munich ChronoType Questionnaire for Shift-Workers (MCTQShift). J Biol Rhythms 2013 Apr;28(2):130-140. [doi: 10.1177/0748730412475041] [Medline: 23606612]
32. de la Vega R, Tomé-Pires C, Solé E, et al. The Pittsburgh Sleep Quality Index: validity and factor structure in young people. Psychol Assess 2015 Dec;27(4):e22-e27. [doi: 10.1037/pas0000128] [Medline: 26653055]
33. Craig CL, Marshall AL, Sjöström M, et al. International Physical Activity Questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
34. Azpeleta C, Santos P, Sobrado A, Lesmes M, Gal B. Forcing a change: a learn-by-doing workshop on circadian rhythms to understand the complexities of human physiology. Adv Physiol Educ 2021 Sep 1;45(3):437-444. [doi: 10.1152/advan.00117.2020] [Medline: 34124954]
35. Cohen J. Statistical Power Analysis for the Behavioral Sciences: Routledge; 1988.
36. Rosenthal R, Rubin DB. Further issues in effect size estimation for one-sample multiple-choice-type data. Psychol Bull 1991;109(2):351-352. [doi: 10.1037/0033-2909.109.2.351]
37. Albinni B, Baker FC, Javitz H, et al. Morning perception of sleep, stress, and mood, and its relationship with overnight physiological sleep: findings from the National Consortium on Alcohol and Neurodevelopment in Adolescence (NCANDA) study. J Sleep Res 2023 Oct;32(5):e13886. [doi: 10.1111/jsr.13886] [Medline: 36941027]
38. Edinger JD, Arnedt JT, Bertisch SM, et al. Behavioral and psychological treatments for chronic insomnia disorder in adults: an American Academy of Sleep Medicine clinical practice guideline. J Clin Sleep Med 2021 Feb 1;17(2):255-262. [doi: 10.5664/jcsm.8986] [Medline: 33164742]
39. Walker J, Muench A, Perlis ML, Vargas I. Cognitive behavioral therapy for insomnia (CBT-I): a primer. Klin Spec Psihol 2022;11(2):123-137. [doi: 10.17759/cpse.2022110208] [Medline: 36908717]
40. Tadros M, Newby JM, Li S, Werner-Seidler A. A systematic review and meta-analysis of psychological treatments to improve sleep quality in university students. PLoS ONE 2025;20(2):e0317125. [doi: 10.1371/journal.pone.0317125] [Medline: 39946428]
41. Chandler L, Patel C, Lovecka L, et al. Improving university students' mental health using multi-component and single-component sleep interventions: a systematic review and meta-analysis. Sleep Med 2022 Dec;100:354-363. [doi: 10.1016/j.sleep.2022.09.003] [Medline: 36198252]

XSL•FO
RenderX

42. Fretz KM. Recommendations to bolster adherence in cognitive behavioral therapy for insomnia: a self-efficacy approach. Transl Behav Med 2023 Apr 15;13(4):206-211. [doi: 10.1093/tbm/ibac104] [Medline: 36688474]
43. Yeung V, Sharpe L, Geers A, Colagiuri B. Choice, expectations, and the placebo effect for sleep difficulty. Ann Behav Med 2020 Jan 24;54(2):94-107. [doi: 10.1093/abm/kaz030] [Medline: 31504091]
44. Lu YA, Lin HC, Tsai PS. Effects of digital sleep interventions on sleep among college students and young adults: systematic review and meta-analysis. J Med Internet Res 2025 May 12;27:e69657. [doi: 10.2196/69657] [Medline: 40354636]
45. Bonmati-Carrion MA, Middleton B, Revell V, Skene DJ, Rol MA, Madrid JA. Circadian phase assessment by ambulatory monitoring in humans: correlation with dim light melatonin onset. Chronobiol Int 2014 Feb;31(1):37-51. [doi: 10.3109/07420528.2013.820740] [Medline: 24164100]

Abbreviations

ACM: ambulatory circadian monitoring
CBT: cognitive behavioral therapy
CBT-I: cognitive behavioral therapy for insomnia
CHS: circadian health score
CONSORT: Consolidated Standards of Reporting Trials
ES: environmental synchronization
IPAQ: International Physical Activity Questionnaire
IS: interdaily stability
PSQI: Pittsburgh Sleep Quality Index
RAN: relative amplitude normalized
TAPL: temperature, acceleration, position, and exposure to light

Edited by S Law; submitted 30.Sep.2025; peer-reviewed by L Zhou, M Shimaoka; revised version received 21.Jul.2026; accepted 23.Jul.2026; published 16.Sep.2026.

Please cite as:
Pérez-Manchón D, Azpeleta C, Gal-Iglesias B, Ruiz Zaldibar C
Effects of a Group-Based Cognitive Behavioral Therapy for Sleep in Nursing Students: Pilot Randomized Controlled Trial
JMIR Hum Factors 2026;13:e85021
URL: https://humanfactors.jmir.org/2026/1/e85021
doi:10.2196/85021

© David Pérez-Manchón, Clara Azpeleta, Beatriz Gal-Iglesias, Cayetana Ruiz Zaldibar. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 16.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Using Intervention Mapping to Codevelop Orchid, a Digital Tool for Reproductive Life Planning: Development and Feasibility Study

Catherine Stewart[1], BSc, MSc; Helen Carr[1,2], MSc, MBBS; Maitri Shila Tursini[1], MSc, MD; Alice Howe[1,3], MSc, MBChB; Jennifer Hall[1], BSc, MSc, BMBS, PhD

[1]Reproductive Health Research Department, Institute for Women's Health, University College London, 74 Huntley Street, London, England, United Kingdom

[2]NHS Surrey Heartlands Health and Care Partnership, Guildford, United Kingdom

[3]Homerton University Hospital NHS Foundation Trust, London, England, United Kingdom

Corresponding Author:
Jennifer Hall, BSc, MSc, BMBS, PhD
Reproductive Health Research Department, Institute for Women's Health, University College London, 74 Huntley Street, London, England, United Kingdom

Abstract

Background: Most people make no health or lifestyle changes before pregnancy, missing a key opportunity to improve outcomes. Consequently, nearly half of UK pregnancies are unplanned, disproportionately affecting underserved groups and widening health inequalities. Digital health interventions (DHIs) offer promise but require systematic, theory-driven development to ensure effectiveness and real-world applicability.

Objective: This study aimed to use Intervention Mapping to codevelop and pilot Orchid, a novel DHI designed to support people of reproductive age to understand their pregnancy preferences and develop a reproductive life plan (RLP).

Methods: We used Intervention Mapping steps 1 - 4 to guide the systematic, theory-informed codevelopment of Orchid. A multidisciplinary planning group and codevelopment group of 21 members of the public contributed throughout. At step 1, previous research and a scoping review of existing RLPs informed the program goals and logic model of the problem. During step 2, we identified performance objectives and behavioral determinants to specify practical strategies for each target behavior. At step 3, we applied the Capability, Opportunity, Motivation–Behavior (COM-B) model and relevant behavior change techniques to guide intervention design. Finally, at step 4, Orchid was co-designed as a website and mobile app providing users with a pregnancy preference group and prediction of pregnancy, a dynamic RLP, tailored evidence-based information, and optional goal-setting features to support behavior change. Orchid was piloted between January and May 2025 to explore its feasibility and acceptability in health care settings. Interviews with users, nonusers, and health care professionals were conducted and quantitative data from users were collected.

Results: These findings indicate that implementation was feasible, and health care professionals found it acceptable to recommend Orchid to patients, though noted barriers including time constraints and competing priorities. Overall, 153 people signed up to Orchid; 68% (72/106) of eligible users received a pregnancy preference group and 27% (32/119) of eligible users completed a full RLP. Users were positive about Orchid, appreciating its content and design, noting that Orchid contained a wealth of information about reproductive health presented in an easy-to-understand manner. They valued the autonomy, convenience, and privacy afforded by the digital format, and found it acceptable to be recommended Orchid within a health care setting. Orchid uptake was lower than anticipated, and use was limited; this was partly expected given the short pilot period, but feedback also suggested targeted recruitment and navigation improvements could enhance uptake and engagement.

Conclusions: Orchid is the first co-designed DHI to support reproductive health across the life course. Its systematic development, theoretical foundation, strong user involvement, and positive pilot testing position it as a promising, scalable innovation to support reproductive health, deliver credible information in accessible formats, and promote preventative, community-based care across the National Health Service.

(JMIR Hum Factors 2026;13:e87650) doi:10.2196/87650

KEYWORDS

digital health intervention; reproductive life plan; reproductive health; preconception health; preconception care; contraception; intervention mapping

XSL•FO
RenderX

Introduction

Background

Within the United Kingdom, nearly half of all pregnancies are unplanned [1], and abortion rates are at an all-time high [2]. These trends suggest that reproductive health services are failing to meet people's needs. Contraceptive use remains inconsistent, with almost 1 in 5 women of reproductive age in England reporting using no contraception or unreliable methods [3], and the use of hormonal contraception is decreasing [4]. Unplanned pregnancies are disproportionately concentrated among underserved populations [5], including those living in poverty and individuals from ethnic minority groups, and are associated with a range of adverse outcomes such as preterm birth, involvement of children's social care services, financial adversity, and health inequalities [6]. As such, unplanned pregnancy remains a significant public health concern and a key driver of disparities in maternal and child health.

At the same time, the health of individuals entering into pregnancy (whether planned or unplanned) presents additional challenges. More than 90% of women in the United Kingdom have at least one modifiable risk factor associated with adverse pregnancy outcomes [7] including maternal and neonatal morbidity and mortality. Yet, the majority of women, and their partners, make no health or lifestyle changes in preparation for pregnancy [8]. This lack of planning and preparation represents a missed opportunity to optimize health before conception, improve pregnancy outcomes, and reduce longer-term risks for both parent and child. Contributing to this gap is the fact that discussions around pregnancy preferences and preconception health are not routine in health and social care settings [9,10], even for people with preexisting conditions, those taking medications, or those with social risk factors.

Furthermore, reproductive health discourse and interventions often overlook the role of men, despite evidence linking paternal nutrition, body composition, and lifestyle behaviors to sperm quality and offspring health outcomes [11,12]. This underscores the need for a more inclusive, holistic approach to reproductive health.

There is an unmet and expressed need among people of reproductive age for better support with their reproductive health [9,13]. Combining this need with the challenges outlined above highlights the importance of adopting a more proactive, person-centered approach to supporting reproductive decision-making. Reproductive life planning [14] offers such an approach, enabling individuals to reflect on their reproductive intentions, align these with their health and life circumstances, and access the information and care needed to achieve their goals.

While health care professionals have a vital role to play in delivering this support, additional approaches are needed to extend reach and accessibility. Digital health interventions (DHIs) offer a promising approach, with emerging evidence suggesting that they can overcome barriers of distance, accessibility, stigma, and judgment [15] and can improve preconception health [16]. However, evidence on their effectiveness in supporting contraception use remains limited [17,18].

The development of any new DHI should follow a systematic, theory-driven process to maximize its potential for effective real-world implementation. In line with Medical Research Council guidance on complex interventions [19,20], such development should be reported transparently to strengthen the evidence base, enhance reproducibility, and allow others to assess its relevance across contexts. Despite these imperatives, the intervention development phase remains relatively underreported, which poses issues with transparency, quality, and user-centeredness [21].

Aim

This paper aims to describe the systematic development and pilot testing of Orchid, a novel DHI to support people of reproductive age to understand their pregnancy preferences and develop a reproductive life plan (RLP).

Methods

Development of Orchid Using Intervention Mapping Framework

Our development of Orchid followed the Intervention Mapping (IM) framework [22], as used by other digital health programs [23].

Step 1: Needs Assessment

Establish and Work With a Planning Group

Our planning group, that met weekly, was comprised of 2 female researchers (JH and CS) who had led the development work underpinning Orchid, 3 female clinicians (HC and MST [2 general practitioners] and AH [sexual and reproductive health doctor]) affiliated with the team, and a male software developer.

While previous work was focused on women, early input from the codevelopment group highlighted that including men would be important both to reduce the disproportionate burden placed on women and to address the underrepresentation of men in reproductive health research. Therefore, we created a codevelopment group of 21 public members (males and females, aged 16 - 45 y, with diverse ethnicities, relationship statuses, sexuality, and family sizes, living across the United Kingdom). Members were recruited from existing Public and Patient Involvement groups and the school where Orchid was piloted.

Conduct Mixed Methods Needs Assessment

We conducted primary research on the feasibility and acceptability of different ways of assessing pregnancy preferences, followed by a scoping review, to explore the existing evidence on RLPs [24], to identify key goals for Orchid.

Primary Research

The Pregnancy Planning, Preparation, and Prevention Study (P3 Study) explored ways of assessing women's feelings and preferences regarding a future pregnancy [25]. Within its cohort, 80% of women expressed a preference for exploring their pregnancy preferences digitally [9]. This preference was consistent across different age groups and was further validated

XSL•FO
RenderX

through qualitative research involving in-depth discussions with women [13] and reflects findings from other studies [26]. Qualitative work with health care professionals [13] (HCPs) reinforced that a digital tool providing accessible reproductive health information could usefully extend support beyond clinical encounters. Although HCPs viewed discussions about pregnancy preferences as acceptable and valuable, many reported lacking the confidence, training, and time to deliver them effectively [13].

The P3 study also evaluated the psychometric performance and predictive validity of the UK version of the Desire to Avoid Pregnancy (DAP) scale [27] and found it to be a valid and reliable measure of women's desire to avoid pregnancy, as well as highly predictive of pregnancy within the next 12 months (sensitivity 78% and specificity 81% [27,28]).

Scoping Review

We carried out a scoping review of the published and gray literature to map the evidence on RLPs [24]. Most studies and gray literature originate from the United States, primarily within clinical settings. Across the literature, 21 different RLPs were identified, with many based on the 2006 Moos RLP [29]. Digital RLPs were scarce, with none currently available in the United Kingdom.

Acceptability of RLPs was consistently high across studies; users were overwhelmingly positive about being asked about reproductive life planning. However, evidence of effectiveness was limited. There was minimal evidence of the application of behavior change theory in the development of the RLPs, with only 3 tools (all digital) describing the inclusion of specific behavior change theories [30-32].

Identify Program Goals

The needs assessment helped to identify the program goals: to create a digital tool (app and website) to help people of reproductive age understand their pregnancy preferences, support them to develop and implement an RLP, and increase awareness of reproductive health. The aim is to support users to achieve their reproductive health goals, reduce unintended pregnancies, increase pregnancy preparations, and empower decision-making.

Create a Logic Model of the Problem

We created a logic model (Table S1 in Multimedia Appendix 1) of the problem from the findings of our scoping review and previous research.

Step 2: Selecting Behavioral Determinants and Performance Objectives

Informed by the results of our needs assessment and input from our codevelopment group, we identified specific performance objectives for subpopulations, based on reproductive preferences. For example, taking folic acid for those actively planning a pregnancy or consistent and correct contraceptive use for those wishing to avoid pregnancy. See Table S2 in Multimedia Appendix 1 for more examples.

Our scoping review identified 5 key behavioral determinants that are underlying drivers of behavior change: (1) knowledge [30], (2) self-efficacy [31,32], (3) social support [31], (4) intrinsic motivation [30], and (5) autonomy [30]. We explored the relationships between performance objectives and behavioral determinants and identified change objectives and practical strategies for each target behavior.

Step 3: Identifying Theory-Based Intervention Methods and Practical Applications

We applied the Capability, Opportunity, Motivation–Behavior (COM-B) model [33] when developing Orchid. This model was selected because each of the behavioral determinants we had identified (knowledge, self-efficacy, social support, intrinsic motivation, and autonomy) aligns with one or more of the COM-B components, providing a comprehensive framework for influencing behavior change within Orchid. We systematically mapped each component of the COM-B model to Orchid's proposed content and features, linking them to the specific objectives identified during the planning phase (Table 1). This ensured that theory directly informed Orchid's design, aligning practical features with the mechanisms expected to support behavior change. We also identified several behavior change techniques [34-37], including goal setting and reminders, feedback, habit stacking, action planning, information, and education. These techniques were incorporated into Orchid to support the COM-B framework, enhance engagement, and increase the likelihood of successful behavior change.

XSL•FO
RenderX

Table . Mapping the COM-B[a] model components as applied to Orchid against the preconception and contraception pathways.

Orchid intervention	COM-B components	Link to behavior change
Preconception		
Evidence-based questionnaire or risk assessment	• Psychological capability • Reflective motivation	• Builds awareness and knowledge of health status. • Prompts self-reflection and understanding of risk factors that may need addressing.
Informational content on health optimization	• Psychological capability • Reflective motivation	• Enhances users' knowledge and skills to make informed decisions. • Supports internal motivation by linking actions to outcomes.
Access to relevant resources	• Physical opportunity • Psychological capability	• Enhances opportunity by increasing access to information.
Find your nearest health care provider or support groups	• Social opportunity • Physical opportunity	• Enhances opportunity by increasing access to services and connecting users to support systems.
Goals and reminders for preconception behaviors	• Reflective motivation • Automatic motivation	• Supports planning and habit formation. • Reminders and prompts maintain engagement and drive action.
Positive reinforcement messaging	• Automatic motivation	• Builds positive associations, rewards behavior, and sustains engagement over time.
Contraception		
Personalized contraceptive questionnaire	• Psychological capability • Reflective motivation	• Builds understanding of contraceptive choices and helps users make informed decisions that fit their lifestyle, increasing likelihood of consistent use.
Educational content on contraception	• Psychological capability • Reflective motivation	• Enhances knowledge, increasing competence and confidence to choose and use contraception effectively.
Access to relevant resources	• Physical opportunity • Psychological capability	• Increases access to credible information and reduces barriers.
Goals and reminders for contraceptive use	• Automatic motivation • Reflective motivation	• Encourages habit formation and planning, reducing forgetfulness, and promoting consistent use.
Action planning for contraceptive scenarios	• Psychological capability • Reflective motivation	• Builds self-regulation and coping strategies, increasing users' ability to maintain contraceptive use across variable contexts.
Find your nearest health care provider	• Physical opportunity • Social opportunity	• Improves access to professional support and normalizes help-seeking behavior.

[a]COM-B: Capability, Opportunity, Motivation–Behavior model

Step 4: Program Development and Pilot Testing

Development of Orchid Prototype

Between November 2022 and February 2023, we codeveloped a prototype of Orchid. This incorporated the DAP scale and a digitized pregnancy prediction algorithm [28] to present users with a pregnancy preference group, their likelihood of pregnancy, and tailored resources. Throughout the process, our codevelopment group emphasized the importance of ensuring that Orchid is inclusive of all users, including men, and expressed a desire for Orchid to be available as both an app and a website to enhance accessibility.

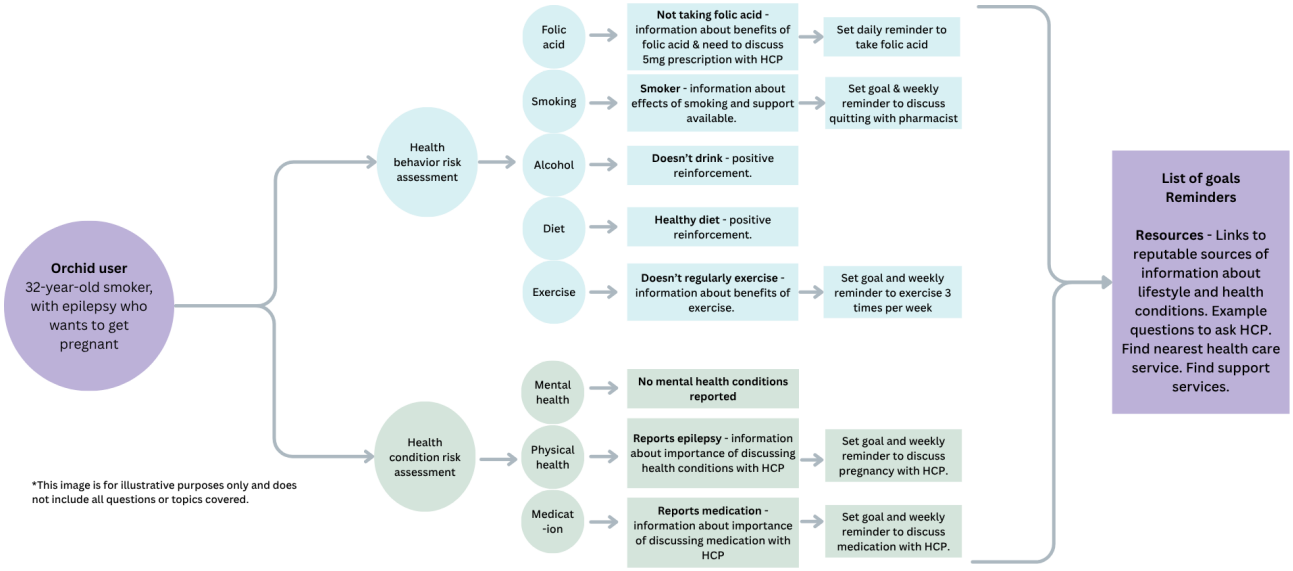
Development of Orchid Version 1

Starting with the Orchid prototype and guided by the COM-B model, program goals, and aims, we worked collaboratively with our program team and codevelopment group to build the RLP component of Orchid (Figure 1). Our public contributors worked with us throughout to enhance accessibility and

XSL•FO
RenderX

engagement. They played a central role in shaping the content and user pathways; their input led to substantial expansion of the scope to encompass sexual health, menstrual health, menopause, cancer awareness, and decision-making. The codevelopment of Orchid was an iterative process using a mixture of online group meetings (n=3) and surveys (n=10) from February to December 2024.

Figure 1. Walk-through of Orchid preconception Reproductive Life Plan pathway. HCP: health care professional.



Users are required to sign in to Orchid to protect their data and ensure confidentiality. They then answer a series of questions about their reproductive preferences (DAP Scale) [38]. Based on their answers, users are assigned a pregnancy preference group (eg, Planner, Preventer, and Pauser) and given a personalized pregnancy prediction. This prediction is generated by an algorithm (developed by JH) that combines DAP responses with selected sociodemographic characteristics to calculate the likelihood of becoming pregnant in the next year [39]. Full details on pregnancy preference group classification and the pregnancy prediction algorithm are provided in the Multimedia Appendix 1.

Users are then guided down a dynamic pathway of questions that adapts based on their answers to create a comprehensive RLP. Relevant, evidence-based information is provided along the pathway and signposted in the Resources section. As users move down the pathway, they can set optional to-do items and reminders, aimed at assisting users to take action to achieve their goals. For details, see Figure 2 for an overview of Orchid.

XSL•FO
RenderX

Figure 2. Orchid overview. DAP: Desire to Avoid Pregnancy scale.

XSL•FO
RenderX

Pilot Testing to Explore Acceptability and Feasibility

We conducted a pilot between January and May 2025 to explore the acceptability and feasibility of implementing Orchid in health care and educational settings. We collaborated with our codevelopment group to develop the ethics application for this study, codeveloping recruitment materials, such as posters and SMS text messages, participant information sheets, topic guides, and lesson plans. This paper focuses on the health care pilot; the school pilot is detailed separately [40].

Health Care Pilot

We piloted Orchid in 3 National Health Service (NHS) health care settings in Greater London: a GP practice, a sexual health clinic, and a women's health hub (WHH). The pilot ran for 5 months with GP and sexual health clinic sites launching in January 2025, followed by the WHH in February 2025. All study locations concluded activity at the end of May 2025.

Each site received a 1-hour training session from a clinician prior to the launch of Orchid, covering how and why to ask about pregnancy preferences, the principles of reproductive life planning and preconception health, and an introduction to Orchid and the pilot study. The content and style of the training program were based on Normalization Process Theory (NPT) [41,42] and COM-B [33]. The NPT constructs of context, mechanisms, and outcomes were considered, particularly aiming to present a coherent message that was cognitively engaging and included suggestions for collective action in that context. A presentation was adapted for each training site and an electronic copy of the slides was provided afterward. We delivered the training in person at 2 sites and online at the third. In total, 36 HCPs attended the training.

Patients (aged 18 - 50 y) were invited to use Orchid through 3 methods: text messages sent via Accurx (used only at the GP site), posters displayed in the waiting rooms (GP and sexual health clinic), and HCP recommendations during clinical consultations (all sites).

At the GP practice (the only site with a defined patient list), text messaging served as the primary recruitment method. Overall, 4456 eligible patients were identified and contacted in 2 batches: the first half received an initial invitation on January 8, 2025, and the second half on January 16, 2025. All patients were recontacted on February 7, 2025, followed by a targeted reminder sent on April 30, 2025, to 939 patients who had attended an appointment in the previous 2 months. In contrast, precise numbers of eligible patients could not be determined for the sexual health clinic or WHH, as neither holds a registered patient list. However, the sexual health clinic sees approximately 60 - 80 patients per day, while the WHH sees around 20 - 30 patients per week (with many ineligible for Orchid due to age criteria).

Orchid User Quantitative Data

Within Orchid, we collected quantitative data on uptake, completion of Orchid, sociodemographics of users, recruitment method, and location. Completion was defined as reaching the end of a personalized RLP pathway. Users were grouped according to their pregnancy preferences as follows: Planner (wants a pregnancy now), Undecided (has not decided whether they want children), Postponer (wants children in the future), Pauser (has at least 1 child and would like another in the future), Finisher (has all the children they want), and Preventer (does not want children). Quantitative data were analyzed using Stata/MP 18 (StataCorp).

Health Care Professional Interviews

HCPs from each of the 3 participating sites were recruited via email. To be eligible for an interview, HCPs were required to have attended the Orchid training.

Orchid User and Nonuser Interviews

We recruited Orchid users and nonusers through separate strategies to ensure representation of both groups. All registered Orchid users were invited to participate via email. Information about interviews was also included on posters displayed at sites. Nonusers were identified and invited through GP-sent Accurx text messages; patients received an initial message asking whether they had accessed Orchid, and those who responded "no" were asked about their reasons for nonuse and invited to take part in an interview. For details, see Figure S1 in Multimedia Appendix 1 for a flow diagram outlining the participant recruitment process.

All interviews were conducted via Zoom (Zoom Communications, Inc), audio-recorded with participant consent, and transcribed verbatim by a member of the research team. Following data familiarization, we collaboratively developed a shared coding framework for the user and nonuser interviews through group discussions, drawing on our research questions and early impressions from the transcripts. Using this common framework, user interviews were coded by CS and nonuser interviews by JH. In parallel, HCP interviews were analyzed using a separate coding framework informed by NPT [41], with coding undertaken by HC. To ensure analytic consistency, the full team met regularly to discuss emerging codes, compare interpretations, and refine frameworks as needed. These meetings supported iterative theme development and enabled us to explore differences and commonalities across user, nonuser, and HCP accounts, ensuring that all final themes were firmly grounded in the data. Qualitative data were analyzed in NVivo 14 (Lumivero).

Qualitative data collection concluded once all users, nonusers, and eligible HCPs who responded and consented had been interviewed.

Ethical Considerations

Ethical approval for the health care pilot was granted by the NHS London, Riverside Research Ethics Committee (Ref: 24/LO/0421). Informed consent was obtained from all Orchid users; they were provided with a participant information sheet and were required to provide consent before using Orchid. Informed consent was also obtained from all users, nonusers, and HCPs who participated in an interview. Participants who took part in interviews were given a £25 (US $33) voucher.

All user data collected through Orchid were handled in accordance with data protection regulations. Email addresses were stored separately from the main dataset, and all analytical

XSL•FO
RenderX

data were pseudonymized prior to analysis. All data were stored in the University College London Data Safe Haven and were only accessible to authorized members of the research team.

Results

All data presented are new and unique to this manuscript.

Health Care Professional Interviews

We conducted in-depth interviews with 10 HCPs. The NPT [41] framework was used to analyze the data, exploring the effectiveness of the training, the likelihood of adoption, implementation, and sustainment of Orchid. Outcome data were limited by the short-term nature of the pilot project, and we therefore focused on the intervention performance.

Training was feasible in all 3 clinical settings and was scheduled during each site's regular training program. In-person sessions were preferred to live online training, but HCPs agreed that online was sometimes more practical. Most HCPs found the training stimulating and understood the importance of the topic. New learning and increased confidence were commonly reported, and HCPs appreciated receiving detailed practical tips that they could immediately incorporate into their clinical practice. The inclusion of men in the RLP message was novel for most and several specifically commented on their cognitive engagement with the training, saying,

> *I thought the training was great, because it actually got me thinking differently.* [CC05]

Most HCPs said that being able to use Orchid within the training would have increased their confidence in sharing the tool with patients:

> [MP02]

Despite this, most HCPs found signposting patients to Orchid acceptable, although some demonstrated limited engagement; one HCP said,

> *I've discussed it with some patients, but I haven't found any leaflets... I'm kind of forgetting … about it.* [CC05]

Many HCPs felt that reminders or brief refresher sessions would be useful:

> *Just like a little nudge for people sometimes … just kind of brings it back up to the top of your memory.* [CC02]

and emphasized the need for more collective action to support sustained uptake of Orchid. Having a local champion at a site was seen as valuable in maintaining momentum, as illustrated by one HCP:

> *So [name] … works in the clinic, I know she just gets [the leaflets] out and puts them on the desk as well, so, so everybody hopefully knows about it and brings it up.* [WHH02]

HCPs found it feasible to incorporate opportunistic RLP and preconception information into certain consultations, particularly those relating to women's health, baby clinics, vaccination clinics, chronic disease clinics, and male fertility. One HCP noted

> *It doesn't take long … anything related to women's health is really easy.* [WHH01]

However, HCPs agreed that it was more challenging to introduce these topics in other consultations:

> *It's just difficult to talk contraception when somebody comes for something completely different.* [WHH01]

and many acknowledged that such discussions were not always possible:

> *You have your agenda of keeping the patient safe, and they have their agenda of what they want to get sorted out, and then this is like another thing to add into the mix.* [MP04]

Overall, views were mixed on the impact on consultation times; while some HCPs were hesitant to raise the topic for fear of it leading to a longer consultation, others acknowledged that Orchid could "save time maybe because … we can refer directly to the app and say, look, we have good information there" (MP01).

Importantly, where HCPs had mentioned Orchid to patients, the response was consistently positive:

> *I have recommended Orchid to a few patients, and they've been really interested in it.* [WHH02]

At the GP practice, text messages were used to promote Orchid. While HCPs were supportive of this method of recruitment, "I think the text message is absolutely excellent," they did share concerns about potential pitfalls:

> *Sometimes I do get patients worrying that the texts they're getting are scams and lots of people don't speak … or can't read English.* [MP04]

Orchid Users' Quantitative Data

In total, 153 people signed up to Orchid, of whom 109 (71%) were females and 44 (29%) were males. Most people found out about Orchid via a text message from their GP (102/153, 66.6%), followed by clinical consultation (19/153, 12.4%) and posters in waiting rooms (15/153, 9.8%). The most common reason for using Orchid was "I want reliable information" (48/153, 31.3%), followed by "just curious" (36/158, 23.5%). Sociodemographic data of Orchid users can be seen in Table 2.

XSL•FO
RenderX

Table . Orchid users' sociodemographic characteristics.

Demographic	Values
Age (n=136), mean (range)	33 (18 - 50)
Sex (n=153), n (%)	
Female	109 (71.2)
Male	44 (28.8)
Ethnicity (n=145), n (%)	
White	59 (40.7)
Black, Black British, Caribbean, or African	37 (25.5)
Asian or Asian British	18 (12.4)
Mixed or multiple ethnic groups	14 (9.7)
Other ethnic group	17 (11.7)
Relationship status (n=146), n (%)	
Married / or civil partnership	47 (32.2)
In a relationship, living with partner	23 (14.8)
In a relationship, not living with partner	36 (24.7)
Single	34 (23.3)
Other	6 (3.7)
Sexual orientation (n=142), n (%)	
Heterosexual	108 (76.1)
Bisexual	16 (11.3)
Gay or lesbian	10 (7.0)
Other	8 (5.6)
Education (n=135), n (%)	
Degree or above	76 (56.3)
Diploma in HE[a]	10 (7.4)
A Level or College	30 (22.2)
GCSE[b]	14 (10.4)
None	5 (3.7)
Employment (n=143), n (%)	
Employed full or part time	93 (65.0)
Looking after home or family	9 (6.3)
Long-term sick leave or disabled	11 (7.7)
Student, in full-time education	17 (11.9)
Unemployed	11 (7.7)
Other	2 (1.4)
English main language (n=153), n (%)	
Yes	117 (76.5)
No	36 (23.5)
Index of multiple deprivation quintile (n=102), n (%)	
1 (Most deprived 20% of areas)	47 (46.1)
2	44 (43.1)
3	6 (5.9)

XSL•FO
RenderX

Demographic	Values
4	2 (2.0)
5 (Least deprived 20% of areas)	3 (2.9)

[a]HE: higher education.

[b]GCSE: General Certificate of Secondary Education.

Of the 106 nonpregnant female users, 72 (68%) received a pregnancy preference group (Table 3). Of the 119 people who were eligible to complete an RLP pathway (3 pregnant females, 72 females who received a pregnancy preference group, and 44 males), 32 (26.9%) did so (Table 3). There was no difference in completion by ethnicity. More details of who was eligible for the different features can be found in the Multimedia Appendix 1.

Table . Percentage of Orchid users in each pregnancy preference group and Reproductive Life Plan pathway.

Groups	Orchid users in each group, n (%)
Pregnancy preference group[a] (n=72)	
Planner	26 (36.1)
Undecided	24 (33.3)
Postponer	11 (15.3)
Pauser	5 (6.9)
Finisher	4 (5.6)
Preventer	2 (2.8)
Reproductive Life Plan pathway[b] (n=66)	
Currently trying	7 (10.6)
1 year	12 (18.2)
2 - 4 years	8 (12.1)
5+ years	3 (4.6)
Not sure when	6 (9.1)
Not sure if	13 (19.7)
Don’t want	11 (16.6)
Currently pregnant or expecting	6 (9.1)

[a]Pregnancy preference groups: Planner (wants a pregnancy now), Undecided (has not decided whether they want children), Postponer (wants children in the future), Pauser (has at least one child and would like another in the future), Finisher (has all the children they want), and Preventer (does not want children).

[b]Reproductive Life Plan pathways: users who are currently pregnant or expecting, currently trying to conceive, users who would like to have a child within 1 year, within 2 - 4 years, or in 5+ years, users who want a child but are unsure when, are unsure if they want children, or do not want (more) children.

User and Nonuser Interviews

In total, 5 users and 2 nonusers were interviewed, and we also received written feedback (via Accurx text messaging) from 2 additional nonusers.

Background of Users and Nonusers Interviewed

A total of 4 users interviewed were female and one was male. Users ranged in age from 22 to 35 years, and none had children, although all expressed a desire to have them in the future. Several users had experience with reproductive health conditions, including polycystic ovary syndrome, irregular menstrual cycles, and fibroids. The 2 nonusers interviewed were both female and aged 40 - 49 years; 1 had children and did not want any more, while the other had never been pregnant and did not want children. We have no sociodemographic information on the 2 nonusers who provided written feedback via text message.

Digital Reproductive Health Information

While most users felt confident discussing reproductive health with HCPs, they noted that clinicians often lacked the time or sometimes the specific knowledge needed to answer all questions or provide tailored advice. As a result, users liked the idea of receiving information about reproductive health digitally and felt that digital tools such as Orchid would be a way of overcoming some of these barriers. Users viewed digital tools as a direct and convenient method of accessing information and found that digital content was often more personalized. Users also felt that digital tools would allow “much more productive conversations with a health professional” (U05), as it would

XSL•FO
RenderX

help "answer … questions before … seeing a doctor … and then know … what I'm looking for from the doctor" (U04).

Orchid Recruitment

The primary method of recruitment for the interviewed users was through text messages sent by their GP, with most agreeing that this was an effective approach as receiving the message directly from their GP made it feel official and trustworthy.

Most users chose to visit or download Orchid because they were intrigued by the tool's focus on providing information about reproductive health, with some expressing particular interest due to personal circumstances, such as wanting to learn more about preparing for pregnancy or managing reproductive health conditions. However, 1 user and 3 of the nonusers thought that Orchid was primarily for "women who are wanting to be pregnant" (U02), with the nonusers citing this as the reason why they had not used Orchid, as pregnancy was not something they currently wanted.

Orchid Use

Users were divided on whether they accessed Orchid through the app or the website. Website users typically did so because they were unaware that an app was available or preferred to explore the web version before committing to downloading the app.

Ongoing use of Orchid was limited, with most users reporting that they accessed it only once or twice, primarily to complete the questions and review relevant information. None of the users engaged with the goal or reminder features within the app, citing that goals were either not relevant to their current needs or they were unaware that they could set them. Some users did follow links to external "useful resources" [U05] to find out more information about topics discussed within Orchid. Users also indicated that they would return to the app in the future, particularly if their circumstances changed (eg, if they began actively trying for a pregnancy) and would then be more likely to use features such as goal setting and reminders.

Despite limited use, users reported learning new information from Orchid, particularly regarding preparing for pregnancy. For example, one user reported that Orchid had "chang[ed] my outlook on planning" (U04), while another noted that they had begun taking supplements because of the information provided in Orchid.

Orchid Feedback

Overall, users were positive about Orchid, appreciating both its content and design. They noted that the app contained a wealth of useful information about reproductive health, presented in an easy-to-understand manner. Users also praised the app's look and feel, describing the design as nice and simple, "well laid out" (U05) and "easy to navigate" (U01). They valued the adaptive flow of questions, which personalizes the experience based on users' answers, and appreciated the ability to "go back up the path and say … I've changed my mind" (U02), allowing Orchid to adjust as their needs or decisions changed.

In addition to the positive feedback, users also identified several areas where Orchid could be improved. Many wanted more varied content, such as images and videos, as well as more comprehensive explanations and additional topics. Some users found navigating the web-based version on a mobile phone challenging, and users reported difficulty locating where to download the app. Despite these issues, all users said that they would recommend Orchid to friends and family.

Discussion

Principal Findings

This study set out to codevelop and pilot Orchid, a novel DHI, to support people of reproductive age to understand their pregnancy preferences and form an RLP. Guided by the IM framework [22], we codeveloped Orchid through iterative co-design with end users and a multidisciplinary clinical academic team. The pilot study, conducted across 3 NHS settings, demonstrated that integrating Orchid into real-world clinical contexts was feasible and acceptable. HCPs were generally comfortable recommending Orchid to patients, and users valued its content and design. Although uptake was lower than anticipated, the data collected provide important insights for refinements needed for wider implementation.

Using IM [22] enabled transparent, evidence-based codevelopment grounded in behavioral theory and user needs. The structured process supported iterative refinement and ensured alignment between Orchid's purpose, content, delivery, and intended behavior change outcomes, avoiding assumption-led development. By mapping performance objectives and determinants to theory-based methods, grounding Orchid in COM-B [33], and incorporating behavioral determinants identified in the literature [24], we specified hypothesized mechanisms of action aligned with behavior change techniques shown to increase engagement [43]. This systematic approach enhances transparency and has been associated with improved DHI quality and effectiveness [44,45]. IM's emphasis on needs assessment and stakeholder involvement further ensured that Orchid was relevant, inclusive, and user-centered, which was reflected in strong acceptability among health professionals and diverse users [46,47].

Implementation within health care settings proved feasible, with training integrating well into existing sessions and HCPs generally finding it acceptable to recommend Orchid. However, barriers emerged, including time constraints, competing clinical priorities, and difficulty raising reproductive health in unrelated consultations, which are common concerns [13,48,49]. Users found it acceptable to be recommended Orchid and valued its digital format due to the convenience, autonomy, and privacy it affords, consistent with previous findings [13].

Uptake was lower than expected, a common challenge in digital health research [50,51]. Feedback from those who received a GP text message but did not sign up uncovered a perception that Orchid was primarily for people planning a pregnancy, despite our efforts to use neutral "reproductive health" messaging. This highlights the need for tailored recruitment strategies that clearly communicate relevance, though evidence on personalized recruitment remains mixed [52]. Digital recruitment remains cost-effective [50], suggesting optimization

XSL•FO
RenderX

is worthwhile [50]. HCPs valued the training but reported that the lack of a live demonstration limited their confidence in signposting, echoing wider evidence that successful DHI implementation requires practical exposure [53]. Addressing these factors is likely to improve recruitment in future roll-outs.

Engagement with Orchid during the pilot was limited, with infrequent use of goal-setting and reminder features. While partly expected due to the short pilot period, users reported barriers such as navigation challenges and not realizing Orchid could be downloaded as an app (where the goal-setting features are optimized). Despite this, users gained new knowledge and described ways in which Orchid informed their reproductive decision-making or led to behavior change, suggesting that even brief interventions can have a meaningful impact [54], as others have also found across a range of health conditions [51,55]. As part of IM step 4 [43], we will refine pathways through Orchid, reduce friction points, and expand features to support more sustained engagement, ahead of larger-scale evaluation (steps 5 and 6).

Existing RLP studies [56] demonstrate high acceptability but limited evidence of effectiveness [56], with most studies focusing on process outcomes, such as knowledge, rather than behavior change outcomes [24]. However, the OptimalMe study [57] found increased use of folic acid and reduced alcohol intake in women planning a pregnancy [57], while digital contraceptive interventions appear to be associated with increased uptake of effective contraception [58], especially post pregnancy, though this is not always the case [59]. As the first UK co-designed, theory-driven, digital RLP integrating reproductive preferences with both contraception and preconception pathways, Orchid fills longstanding gaps in scope, theory, and digital innovation.

As a feasibility and acceptability pilot, this study is limited. Key outcomes, including ongoing engagement, behavior change, and effectiveness, require evaluation in future stages. Additional limitations include the lack of sampling frames at 2 sites, preventing response rate calculation or assessment of generalizability, and uncertainty around exposure to different recruitment methods. The absence of a live app at the time of training may have reduced HCP signposting, and limited visibility of the app means we cannot draw conclusions about platform preference.

Orchid's systematic development, theoretical foundation, and extensive end-user involvement represent key strengths commonly associated with effective DHIs [19,20,33,44,45,54]. Together, these features position Orchid well to function as an effective intervention for improving reproductive health [60]. Our emphasis on inclusivity is reflected in the diversity of both the codevelopment group and pilot users, including people from more deprived areas, speakers of languages other than English, global majority populations, men, and nonbinary people. The varied pregnancy preferences among users show that Orchid serves diverse needs. A particularly notable strength is Orchid's explicit inclusion of men, a group often overlooked in preconception and reproductive life planning. The fact that nearly one-third of users were male highlights demand for male-inclusive reproductive health support and demonstrates Orchid's reach among underserved groups, aligning with emerging evidence that systematically codeveloped DHIs demonstrate greater acceptability and effectiveness [19,20,33,44,45,54].

Orchid has been designed to support national priorities for digital transformation and preventative care, reduction of unplanned pregnancy, and improvements in maternity outcomes. The diversity of pilot users supports our aim of equitable reach. The NHS 10-Year Plan for England [61] prioritizes digital innovation, prevention, and strengthened primary care, and Orchid aligns with each of these aims. By offering accessible, evidence-based reproductive health information and supporting people to articulate and act on their reproductive intentions, Orchid has the potential to enhance digital health literacy, reduce service pressures, and improve patient-provider interactions. By addressing inconsistent contraception use, lack of proactive reproductive life planning, and low awareness of preconception health actions, Orchid can contribute to reducing unplanned pregnancy and improving maternal and neonatal outcomes, as well as longer-term health and development of parents and children.

Conclusion

Orchid is the first DHI co-designed to support reproductive health across the life course in the United Kingdom. The development and pilot testing of Orchid demonstrate the value of using IM to create a theory-driven, codeveloped digital tool for reproductive health. The findings of our pilot suggest that Orchid is both feasible to implement and acceptable to health care providers and users, while highlighting areas for refinement to enhance engagement and relevance. Insights from the pilot will directly inform iterative improvements and the design of a larger-scale implementation evaluation study to evaluate effectiveness. With its evidence-based foundation, user-centered design, and alignment with national health priorities for prevention and digital transformation, Orchid holds promise as a scalable DHI to support reproductive decision-making, reduce unplanned pregnancy, and contribute to better maternal, parental, and child health outcomes across the NHS.

Acknowledgments

We would like to thank our codevelopment group for their vital contributions and insightful feedback throughout the development of Orchid. Their collaboration was instrumental in shaping this project. We also thank Roshni Wadhwa for her support in the development of Table 1.

XSL•FO
RenderX

Funding

The Orchid prototype development was funded by a University College London Therapeutic Acceleration Support Fund Grant (ref: 204841/Z/16/Z). The Orchid Version 1 development and pilot was funded by a Medical Research Council Gap Funder grant (ref: MR/Y503320/1).

Data Availability

The data generated during this pilot study are limited in scope and not sufficient to support broader secondary analyses.

Authors' Contributions

Conceptualization: JH (lead), CS (support)
Data curation: CS, HC, JH, MST (equal).
Formal analysis: CS, HC, JH (equal).
Funding acquisition: JH (lead), CS (support).
Investigation: CS, HC, JH, MST, AH (equal).
Methodology: JH (lead), CS (support), HC (support), MST (support), AH (support).
Project administration: JH (lead), CS (support).
Supervision: JH.
Validation: JH.
Visualization: CS.
Writing – original draft: CS (lead), HC (support), JH (support).
Writing – review & editing: CS, HC, JH, MST, AH (equal).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Orchid logic model, performance objectives and associated behavioral determinants, pregnancy preference group classifications, probability of pregnancy calculation, eligibility of Orchid features and Orchid recruitment flowchart.
[DOCX File, 141 KB - humanfactors_v13i1e87650_app1.docx]

References

1. Wellings K, Jones KG, Mercer CH, et al. The prevalence of unplanned pregnancy and associated factors in Britain: findings from the third National Survey of Sexual Attitudes and Lifestyles (Natsal-3). The Lancet 2013 Nov;382(9907):1807-1816. [doi: 10.1016/S0140-6736(13)62071-1]
2. Abortion statistics commentary, England and Wales: 2023. GOV.UK. URL: https://www.gov.uk/government/statistics/abortion-statistics-for-england-and-wales-2023/abortion-statistics-commentary-england-and-wales-2023 [accessed 2026-06-17]
3. Women's Reproductive Health Survey 2021 national pilot: contraception and abortion results. GOV.UK. 2025. URL: https://www.gov.uk/government/publications/womens-reproductive-health-survey-2021-national-pilot-contraception-and-abortion-results/womens-reproductive-health-survey-2021-national-pilot-contraception-and-abortion-results [accessed 2026-06-17]
4. McNee R, McCulloch H, Lohr PA, Glasier A. Self-reported contraceptive method use at conception among patients presenting for abortion in England: a cross-sectional analysis comparing 2018 and 2023. BMJ Sex Reprod Health 2025 Jul;51(3):186-190. [doi: 10.1136/bmjsrh-2024-202573]
5. Roberts L, Maxwell N, Elliott M. When young people in and leaving state care become parents: what happens and why? Child Youth Serv Rev 2019 Sep;104:104387. [doi: 10.1016/j.childyouth.2019.104387]
6. Nelson HD, Darney BG, Ahrens K, et al. Associations of unintended pregnancy with maternal and infant health outcomes: a systematic review and meta-analysis. JAMA 2022 Nov 1;328(17):1714-1729. [doi: 10.1001/jama.2022.19097] [Medline: 36318133]
7. Stephenson J, Schoenaker DA, Hinton W, et al. A wake-up call for preconception health: a clinical review. Br J Gen Pract 2021 May;71(706):233-236. [doi: 10.3399/bjgp21X715733] [Medline: 33926884]
8. Stewart C, Hall J. Pregnancy preparation among women and their partners in the UK: how common is it and what do people do? Womens Reprod Health 2024 Apr 2;11(2):429-442. [doi: 10.1080/23293691.2023.2271919]
9. Hall J, Chawla M, Watson D, et al. Addressing reproductive health needs across the life course: an integrated, community-based model combining contraception and preconception care. Lancet Public Health 2023 Jan;8(1):e76-e84. [doi: 10.1016/S2468-2667(22)00254-7] [Medline: 36603914]

XSL•FO
RenderX

10. Li Y, Kurinczuk JJ, Alderdice F, et al. Pre-pregnancy care in general practice in England: cross-sectional observational study using administrative routine health data. BMC Public Health 2025;25(1):1101. [doi: 10.1186/s12889-025-21728-1]
11. Kotelchuck M, Lu M. Father's role in preconception health. Matern Child Health J 2017 Nov;21(11):2025-2039. [doi: 10.1007/s10995-017-2370-4] [Medline: 28983715]
12. Huang JY, Low FM, Kee MZL, et al. More equitable preconception health: paternal life course opportunities for better pregnancy, child, and family outcomes. The Lancet 2026 Apr;407(10537):1459-1470. [doi: 10.1016/S0140-6736(26)00148-0]
13. Hall J, Carr H, Connolly A, Barrett G. How, when, and who should ask about pregnancy intentions in primary care? a qualitative study of primary healthcare professionals' preferences. BJGP Open 2024 Dec;8(4):BJGPO.2024.0148. [doi: 10.3399/BJGPO.2024.0148] [Medline: 39074879]
14. Morse JE, Moos MK. Reproductive life planning: raising the questions. Matern Child Health J 2018 Apr;22(4):439-444. [doi: 10.1007/s10995-018-2516-z] [Medline: 29512053]
15. Murray E. Web-based interventions for behavior change and self-management: potential, pitfalls, and progress. Med 2 0 2012;1(2):e3. [doi: 10.2196/med20.1741] [Medline: 25075231]
16. Nurleli N, Hadju V, Indriasari R, Areni IS, Masni M, Riskyani S. Effectiveness of digital health in improving pregnancy quality among preconception women: a systematic review. Narra J 2025 Aug;5(2):e993. [doi: 10.52225/narra.v5i2.993] [Medline: 40951455]
17. Aung B, Mitchell JW, Braun KL. Effectiveness of mHealth interventions for improving contraceptive use in low- and middle-income countries: a systematic review. Glob Health Sci Pract 2020 Dec 23;8(4):813-826. [doi: 10.9745/GHSP-D-20-00069] [Medline: 33361244]
18. Stephenson J, Bailey JV, Gubijev A, et al. An interactive website for informed contraception choice: randomised evaluation of Contraception Choices. Digit Health 2020 Jan;6. [doi: 10.1177/2055207620936435]
19. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021;374:n2061. [doi: 10.1136/bmj.n2061]
20. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008;337:a1655. [doi: 10.1136/bmj.a1655]
21. Duncan E, O'Cathain A, Rousseau N, et al. Guidance for reporting intervention development studies in health research (GUIDED): an evidence-based consensus study. BMJ Open 2020 Apr 8;10(4):e033516. [doi: 10.1136/bmjopen-2019-033516] [Medline: 32273313]
22. Fernandez ME, Ruiter RAC, Markham CM, Kok G. Intervention mapping: theory- and evidence-based health promotion program planning: perspective and examples. Front Public Health 2019;7:209. [doi: 10.3389/fpubh.2019.00209] [Medline: 31475126]
23. Hadjiconstantinou M, Schreder S, Brough C, et al. Using intervention mapping to develop a digital self-management program for people with type 2 diabetes: tutorial on MyDESMOND. J Med Internet Res 2020 May 11;22(5):e17316. [doi: 10.2196/17316] [Medline: 32391797]
24. Stewart C, Noh H, Pathak N, Hall JA. A scoping review of the format, content, effectiveness and acceptability of reproductive life planning tools. medRxiv. Preprint posted online on Jun 1, 2024. [doi: 10.1101/2024.05.31.24308272]
25. About the P3 study. Pregnancy Planning Preparation Prevention. URL: https://p3-study-ucl.co.uk/about [accessed 2026-06-17]
26. Daly MP, White J, Sanders J, Kipping RR. Women's knowledge, attitudes and views of preconception health and intervention delivery methods: a cross-sectional survey. BMC Pregnancy Childbirth 2022 Sep 24;22(1):729. [doi: 10.1186/s12884-022-05058-3] [Medline: 36151510]
27. Hall J, Barrett G, Rocca C. Evaluation of the Desire to Avoid Pregnancy scale in the UK: a psychometric analysis including predictive validity. BMJ Open 2022 Jul 25;12(7):e060287. [doi: 10.1136/bmjopen-2021-060287] [Medline: 35879004]
28. Hall JA, Barrett G, Stephenson JM, Edelman NL, Rocca C. Desire to Avoid Pregnancy scale: clinical considerations and comparison with other questions about pregnancy preferences. BMJ Sex Reprod Health 2023 Jul;49(3):167-175. [doi: 10.1136/bmjsrh-2022-201750]
29. Moos MK. Preconception health: where to from here? Womens Health Iss 2006;16(4):156-158 [FREE Full text]
30. Callegari LS, Nelson KM, Arterburn DE, et al. Development and pilot testing of a patient-centered web-based reproductive decision support tool for primary care. J Gen Intern Med 2021 Oct;36(10):2989-2999. [doi: 10.1007/s11606-020-06506-6] [Medline: 33538956]
31. Jagoda LM, Bigbee JL. Assessing the influences on rural women's reproductive life plans: a cross sectional descriptive study. Online J Rural Nurs Health Care 2016 Dec;16(2):27-57. [doi: 10.14574/ojrnhc.v16i2.398]
32. Batra P, Mangione CM, Cheng E, et al. A cluster randomized controlled trial of the MyFamilyPlan online preconception health education tool. Am J Health Promot 2018 May;32(4):897-905. [doi: 10.1177/0890117117700585] [Medline: 28391703]
33. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6(1):42. [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
34. Carey RN, Connell LE, Johnston M, et al. Behavior change techniques and their mechanisms of action: a synthesis of links described in published intervention literature. Ann Behav Med 2018. [doi: 10.1093/abm/kay078]

XSL•FO
RenderX

35. Michie S, Johnston M, Francis J, Hardeman W, Eccles M. From theory to intervention: mapping theoretically derived behavioural determinants to behaviour change techniques. Appl Psychol 2008 Oct;57(4):660-680 [FREE Full text] [doi: 10.1111/j.1464-0597.2008.00341.x]
36. Bailey RR. Goal setting and action planning for health behavior change. Am J Lifestyle Med 2019;13(6):615-618. [doi: 10.1177/1559827617729634] [Medline: 31662729]
37. Chuang CH, Weisman CS, Velott DL, et al. Reproductive life planning and contraceptive action planning for privately insured women: the MyNewOptions study. Perspect Sex Reprod Health 2019 Dec;51(4):219-227. [doi: 10.1363/psrh.12123] [Medline: 31820551]
38. Rocca CH, Ralph LJ, Wilson M, Gould H, Foster DG. Psychometric evaluation of an instrument to measure prospective pregnancy preferences: the desire to avoid pregnancy scale. Med Care 2019 Feb;57(2):152-158. [doi: 10.1097/MLR.0000000000001048] [Medline: 30550399]
39. Hall JA, Barrett G, Stephenson J, Rocca CH, Edelman N. Predictive ability of the Desire to Avoid Pregnancy scale. Reprod Health 2023 Sep 25;20(1):144. [doi: 10.1186/s12978-023-01687-9] [Medline: 37749640]
40. Tursini MS, Stewart C, Carr H, Howe A, Hall J. The educational pilot of orchid: a new digital health intervention for reproductive life planning and education. medRxiv. Preprint posted online on Dec 4, 2025. [doi: 10.64898/2025.12.04.25341624]
41. Murray E, Treweek S, Pope C, et al. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. BMC Med 2010 Oct 20;8(1):63. [doi: 10.1186/1741-7015-8-63] [Medline: 20961442]
42. May CR, Albers B, Bracher M, et al. Translational framework for implementation evaluation and research: a normalisation process theory coding manual for qualitative research and instrument development. Implement Sci 2022 Feb 22;17(1):19. [doi: 10.1186/s13012-022-01191-x] [Medline: 35193611]
43. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
44. El Barbari JS, Fikuart M, Beisemann N, et al. Improving medical photography in a level 1 trauma center by implementing a specialized smartphone-based app in comparison to the usage of digital cameras: prospective panel study. JMIR Form Res 2024 Jan 25;8:e47572. [doi: 10.2196/47572] [Medline: 38271087]
45. van Stralen MM, de Vries H, Mudde AN, Bolman C, Lechner L. The working mechanisms of an environmentally tailored physical activity intervention for older adults: a randomized controlled trial. Int J Behav Nutr Phys Act 2009 Dec 8;6(1):83. [doi: 10.1186/1479-5868-6-83] [Medline: 19995412]
46. Wiles LK, Kay D, Luker JA, et al. Consumer engagement in health care policy, research and services: a systematic review and meta-analysis of methods and effects. PLoS One 2022;17(1):e0261808. [doi: 10.1371/journal.pone.0261808] [Medline: 35085276]
47. Grindell C, Coates E, Croot L, O'Cathain A. The use of co-production, co-design and co-creation to mobilise knowledge in the management of health conditions: a systematic review. BMC Health Serv Res 2022;22(1). [doi: 10.1186/s12913-022-08079-y]
48. Steel A, Lucke J, Reid R, Adams J. A systematic review of women's and health professional's attitudes and experience of preconception care service delivery. Fam Pract 2016 Dec;33(6):588-595. [doi: 10.1093/fampra/cmw094] [Medline: 27650308]
49. Goossens J, De Roose M, Van Hecke A, Goemaes R, Verhaeghe S, Beeckman D. Barriers and facilitators to the provision of preconception care by healthcare providers: a systematic review. Int J Nurs Stud 2018 Nov;87:113-130. [doi: 10.1016/j.ijnurstu.2018.06.009] [Medline: 30096578]
50. Brøgger-Mikkelsen M, Ali Z, Zibert JR, Andersen AD, Thomsen SF. Online patient recruitment in clinical trials: systematic review and meta-analysis (preprint). JMIR Preprints. Preprint posted online on Jul 7, 2020. [doi: 10.2196/preprints.22179]
51. Gold N, Yau A, Rigby B, Dyke C, Remfry EA, Chadborn T. Effectiveness of digital interventions for reducing behavioral risks of cardiovascular disease in nonclinical adult populations: systematic review of reviews (preprint). J Med Internet Res. Preprint posted online on 2021. [doi: 10.2196/preprints.19688]
52. Thiblin E, Woodford J, Öhman M, von Essen L. The effect of personalised versus non-personalised study invitations on recruitment within the ENGAGE feasibility trial: an embedded randomised controlled recruitment trial. BMC Med Res Methodol 2022 Mar 6;22(1):65. [doi: 10.1186/s12874-022-01553-5] [Medline: 35249543]
53. Borges do Nascimento IJ, Abdulazeem H, Vasanthan LT, et al. Barriers and facilitators to utilizing digital health technologies by healthcare professionals. npj Digit Med 2023;6(1). [doi: 10.1038/s41746-023-00899-4]
54. Mair JL, Salamanca-Sanabria A, Augsburger M, et al. Effective behavior change techniques in digital health interventions for the prevention or management of noncommunicable diseases: An umbrella review. Ann Behav Med 2023 Sep 13;57(10):817-835. [doi: 10.1093/abm/kaad041] [Medline: 37625030]
55. Baradez C, Liska J, Brulle-Wohlhueter C, Pushkarna D, Baxter M, Piette J. Brief digital solutions in behavior change interventions for type 2 diabetes mellitus: a literature review. Diabetes Ther 2022 Apr;13(4):635-649. [doi: 10.1007/s13300-022-01244-w] [Medline: 35279813]
56. Hipp SL, Chung-Do J, McFarlane E. Systematic review of interventions for reproductive life planning. J Obstet Gynecol Neonatal Nurs 2019 Mar;48(2):131-139. [doi: 10.1016/j.jogn.2018.12.007] [Medline: 30664840]

XSL•FO
RenderX

57. Brammall BR, Garad RM, Teede HJ, Harrison CL. Evaluating preconception health and behaviour change in Australian women planning a pregnancy: the OptimalMe program, a digital healthy lifestyle intervention with remotely delivered coaching. Nutrients 2024 Jan 3;16(1):155. [doi: 10.3390/nu16010155] [Medline: 38201984]
58. Bewa JM. Digital interventions advancing equitable contraception. Johns Hopkins Bloomberg School of Public Health. 2025. URL: https://publichealth.jhu.edu/center-for-global-digital-health-innovation/digital-interventions-advancing-equitable-contraception [accessed 2026-06-17]
59. Stephenson J, Bailey JV, Blandford A, et al. An interactive website to aid young women's choice of contraception: feasibility and efficacy RCT. Health Technol Assess 2020;24(56):1-44. [doi: 10.3310/hta24560]
60. The role of digital health technologies in women's health, empowerment, and gender equality: project report. : World Health Organization; 2024 URL: https://www.who.int/europe/publications/i/item/WHO-EURO-2024-9234-4065-73153#:~:text=A%20growing%20body%20of%20evidence%20shows%20the%20digital,the%20rights%20and%20health%20care [accessed 2026-06-19]
61. Fit for the future: 10 Year Health Plan for England. NHS England. URL: https://www.england.nhs.uk/long-term-plan [accessed 2026-06-17]

Abbreviations

COM-B: Capability, Opportunity, Motivation–Behavior model
DAP Scale: Desire to Avoid Pregnancy Scale
DHI: digital health intervention
GP : general practitioner
HCP: health care professional
IM: Intervention Mapping
MRC: Medical Research Council
NHS: National Health Service
NPT: Normalization Process Theory
RLP: Reproductive Life Plan
WHH: women's health hub

Edited by A Kushniruk; submitted 12.Nov.2025; peer-reviewed by C Harrison, K Wilson-Mitchell, PA Okooboh; revised version received 24.Apr.2026; accepted 29.May.2026; published 02.Jul.2026.

Please cite as:
Stewart C, Carr H, Tursini MS, Howe A, Hall J
Using Intervention Mapping to Codevelop Orchid, a Digital Tool for Reproductive Life Planning: Development and Feasibility Study
JMIR Hum Factors 2026;13:e87650
URL: https://humanfactors.jmir.org/2026/1/e87650
doi:10.2196/87650

© Catherine Stewart, Helen Carr, Maitri Shila Tursini, Alice Howe, Jennifer Hall. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 2.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Examining the Effect of Online Engagement on Older Adults' Subjective Memory Capability: Cross-Sectional Path Analysis

Soohyoung Rain Lee[1*], PhD, MSW; Hang Liu[2*], MSW

[1]Wurzweiler School of Social Work, Yeshiva University, 2495 Amsterdam Ave, 905, New York, United States

[2]Selfhelp, New York, NY, United States

[*]all authors contributed equally

Corresponding Author:
Soohyoung Rain Lee, PhD, MSW
Wurzweiler School of Social Work, Yeshiva University, 2495 Amsterdam Ave, 905, New York, United States

Abstract

Background: The utility of online engagement in enhancing quality of life and mitigating social isolation among older adults is well documented. However, its relationship with cognitive functioning, particularly through online engagement, requires further exploration.

Objective: This study investigated whether active online engagement via the Virtual Senior Center (VSC) program was associated with subjective memory capability among older adults and whether subjective memory capability was associated with psychological well-being and loneliness.

Methods: This study included a cross-sectional sample of 53 homebound older adults participating in the VSC program, which offers diverse online classes to promote social interaction. Path analysis was conducted to examine the associations among online engagement, subjective memory capability, quality of life, and loneliness.

Results: Increased participation in VSC activities was associated with higher subjective memory capability (β=0.29, 95% CI 0.04-0.54; *P*<.02). Subjective memory capability was associated with better quality of life (β=0.29, 95% CI 0.04-0.54; *P*<.001) and lower loneliness (β=0.29, 95% CI 0.04-0.54; *P*<.003). No direct associations were observed between online engagement and quality of life or loneliness.

Conclusions: Subjective memory capability was associated with better quality of life and lower loneliness. Although online engagement had no direct association with quality of life and loneliness, the observed indirect pattern suggests that subjective memory capability may represent a psychologically meaningful pathway through which structured online engagement relates to well-being. These findings highlight the potential of digital platforms to complement traditional forms of socialization, particularly for older adults facing physical or geographic barriers to interaction.

(JMIR Hum Factors 2026;13:e73018) doi:10.2196/73018

KEYWORDS

online engagement; homebound older adults; subjective memory ability; social relationships; digital inclusion

Introduction

Background

In an increasingly digital society, online engagement, which is defined as active participation in digital platforms for social interaction, information sharing, and learning, has emerged as a key factor influencing the well-being of older adults. Given the potential of digital health technologies to mitigate social isolation and promote cognitive health, understanding the impact of online engagement on cognitive function is critical, particularly for homebound or mobility-challenged older adults [1,2].

Existing research has reported that active online engagement is associated with reduced feelings of isolation, which are prevalent among older adults, and has been examined in relation to cognitive functions such as memory. For example, older adults who use online platforms for social interaction have reported higher perceived quality of life and cognitive function [3,4]. However, there remains a notable gap in understanding the association between online engagement and subjective memory capability, defined as an individual's self-assessment of everyday memory functioning, and how these factors relate to overall well-being among older adults [5,6]. Many previous studies have focused on general internet use or access rather than participation in structured, program-based online engagement designed to promote sustained social and cognitive interaction [7,8]. As digital health and online community-based interventions continue to expand, understanding how structured engagement contexts relate to cognitive self-perception and well-being represents an important area for further investigation.

XSL•FO
RenderX

Given the growing interest in digital therapeutics and telehealth interventions, it is important to explore how online engagement contributes to cognitive self-perception and well-being among older adults. This study examined the association between participation in the Virtual Senior Center (VSC) program and subjective memory capability and social well-being among homebound older adults. Specifically, this research aimed to answer the following questions:

1. Does participation in a VSC influence older adults' subjective memory capability?
2. Does increased online engagement lead to improved social connectedness and well-being?
3. How does self-perceived memory capability mediate the relationship between online engagement and well-being?

It was hypothesized that greater participation in VSC activities would be positively associated with subjective memory capability and that subjective memory capability would be associated with higher quality of life and lower loneliness. This study contributes to the broader discussion of digital inclusion and cognitive health interventions by examining these relationships within a structured online engagement program for a homebound population. By focusing on subjective memory capability as a psychological mechanism, this study offers practical insights for designing digital health interventions and community-based programs targeting socially isolated older adults.

Literature Review

Research on internet communication technology in aging populations has highlighted the potential of online engagement to enhance social interaction, reduce isolation, and support psychological well-being [7,9]. Older adults who actively use online platforms for social interaction have reported higher life satisfaction, improved mental health, and better perceived cognitive outcomes than those who remain digitally disconnected [8,10].

Despite the recognized benefits of online engagement for mental health, studies examining its relationship with cognitive function and memory perception remain limited. Much of the literature has focused on the psychosocial aspects of general internet use, such as reducing loneliness and enhancing social connectedness [11,12]. Less attention has been paid to structured, program-based online engagement designed to promote sustained social and cognitive interaction. More recent studies suggest that such structured online engagement may be associated with cognitive benefits, including subjective memory capability and other indicators of cognitive self-perception [13,14].

The VSC program in this study was developed to provide older adults with structured online activities, including educational courses, group discussions, and interactive social experiences. Unlike passive internet browsing or unstructured social media use, the VSC emphasizes scheduled and socially interactive participation. As a community-based platform, it facilitates ongoing engagement among older adults who may face barriers to in-person interaction. Specifically, the facilitator leads each class and encourages participants to engage and lead the group. In contrast to studies focusing on general internet use, the VSC represents a structured program-based approach to online engagement, making it a useful context for examining how sustained participation in organized digital activities relates to cognitive self-perception and well-being.

Prior studies have reported that tailored digital interventions may support cognitive engagement among older adults. For example, the Internet-Based Conversational Engagement (I-CONECT) program found that structured online socialization significantly improved cognitive performance in older adults [13,14]. In a similar way, the VSC reflects an interactive, community-based approach that may be relevant to digital health interventions targeting older adults who face barriers to in-person engagement.

Theoretical Framework

This study is grounded in 2 complementary theoretical frameworks that help explain how online engagement may relate to cognitive self-perception and well-being in older adults. Cognitive reserve theory posits that engaging in intellectually stimulating activities helps build cognitive resilience, thereby maintaining cognitive function and delaying the onset of cognitive decline [15]. Research has shown that older adults who participate in structured cognitive and social activities, such as online discussions, memory training, and digital learning platforms, demonstrate better cognitive performance and self-perceived memory capabilities [13,14]. Digital interventions, including internet-based cognitive training and virtual social engagement, have been found to foster cognitive reserve, particularly among older individuals at risk of social isolation [1,2].

Complementing this perspective, socioemotional selective theory suggests that, as individuals age, they become more selective in their social interactions, prioritizing emotionally meaningful relationships and experiences [16]. Studies have demonstrated that older adults tend to prefer digital communication tools that facilitate socially enriching interactions, such as video calls and virtual communities, over passive internet browsing or impersonal interaction [3,4]. Online platforms such as the VSC cater to these preferences by providing structured, interactive environments where participants can engage in meaningful social and cognitive activities. Research on digital health interventions has highlighted that programs designed around socioemotional needs and cognitive engagement, such as the I-CONECT program, are particularly effective in maintaining cognitive function and improving emotional well-being in older populations [13,14] (see Figure 1 for the theoretical model).

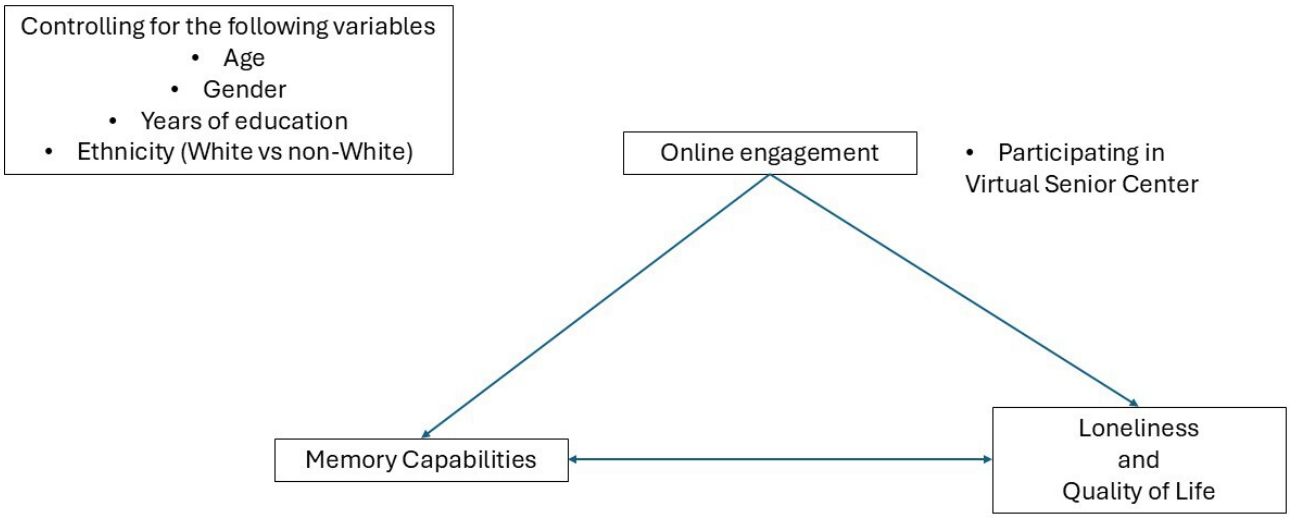
Given the increasing importance of digital inclusion, addressing disparities in digital access and engagement is critical [17]. Socioeconomic status, health conditions, age, and living arrangements all influence older adults' ability to benefit from online engagement [9,18-20]. This study examines the role of online engagement in supporting older adults who choose to age in place, particularly those with physical impairments that limit their access to in-person socialization. By facilitating virtual social interactions and providing access to cognitively stimulating activities, online engagement has the potential to

XSL•FO
RenderX

enhance subjective memory capability and overall quality of life for these individuals.

By integrating cognitive reserve theory and socioemotional selective theory, this study provides a comprehensive framework for understanding the dual role of digital engagement in promoting cognitive self-perception and emotional well-being among older adults.

Figure 1. Theoretical framework.



Methods

Participants and Recruitment

This study examined a cohort of 53 community-dwelling, homebound older adults participating in the VSC program, an online engagement platform designed to support social interaction and cognitive health. Participants were recruited through community aging services, older adult centers, and digital literacy outreach programs targeting older adults at risk of social isolation due to physical impairments, socioeconomic constraints, or mobility limitations.

These partnering organizations identified and referred individuals who experienced difficulty participating in in-person social activities, including those with limited mobility or chronic health conditions. Therefore, they represent a convenience sample of individuals enrolled in the VSC program during the study period.

To be eligible for the study, participants had to (1) be aged 55 years or older and residing in a community setting, (2) have access to a computer with internet connectivity (provided if necessary), (3) be enrolled in the VSC program and actively participating in its activities, and (4) report no clinically diagnosed severe cognitive impairments or neurodegenerative conditions (eg, Alzheimer disease).

Although the program primarily serves older adults aged 55 years or older, the final sample included participants ranging from 47 to 98 years, reflecting real-world enrollment patterns within the VSC program (ie, individuals with disabilities). This study focused on older adults who were dual eligible for Medicaid and Medicare, a population often characterized by low income, greater medical complexity, and increased health vulnerability [21-24]. To ensure equitable participation, participants were provided with computers; internet access when needed; and initial onboarding support, including assistance with logging into sessions and basic troubleshooting. When participants experienced technical difficulties, staff members could resolve the issues remotely without physically visiting their location to expedite the resolution process.

Study Design and Measurement

This study used a cross-sectional design, collecting data over a 6-month period (August 2017-January 2018) using a structured online questionnaire distributed through the VSC platform. Participants completed the survey after at least 3 months of engagement with the program to ensure that they had sufficient exposure to VSC activities before reporting on their experiences.

Ethical Considerations

This study was approved by the WCG institutional review board affiliated with Yeshiva University (IRB00023382). All participants provided informed consent prior to participation. Participant privacy and confidentiality were strictly maintained throughout the study, and all data were de-identified prior to analysis.

Measurement

Subjective Memory Capability

Participants' self-perceived memory function was assessed using the Metamemory in Adulthood (MIA) capacity subscale [25-27]. The MIA measures self-assessed memory performance in areas such as name recall, event tracking, and appointment remembering (eg, "I am good at remembering names," "I am good at remembering birth dates," and "I have no trouble keeping track of my appointments"). This validated measure has been widely used in aging research and has demonstrated moderate correlations with objective cognitive function [28].

Loneliness and Social Engagement

Loneliness was measured using the University of California, Los Angeles (UCLA), Loneliness Scale–Short Form [29,30],

XSL•FO
RenderX

which assesses perceived social connectedness and feelings of isolation (eg, "I feel lack of companionship" and "I feel left out"). Given that older adults in this study were homebound, the short version of the UCLA scale was selected to reduce respondent burden while maintaining validity [31].

Quality of Life

Quality of life was measured using the Older People's Quality of Life Questionnaire (eg, "I enjoy my life," "I am happy most of the time," and "I feel lucky compared to most people") [32-34]. This scale captures emotional well-being, social relationships, and life satisfaction, making it a comprehensive measure for assessing the impact of online engagement.

Online Engagement

Online engagement was quantified through the number of hours per week spent on the VSC platform, categorized into 1 to 3 hours, 4 to 6 hours, and 7 hours or more of participation. These data were extracted from VSC activity logs and reflect participation in structured online sessions (eg, classes and discussions) rather than passive internet use.

Rationale for Self-Reported Measure

While objective cognitive measures (eg, neuropsychological assessments) were not included in this study, research has shown that subjective memory capability is strongly associated with cognitive performance and can serve as a reliable predictor of cognitive decline [35,36]. This study focused on perceived cognitive capability rather than objective performance, consistent with research emphasizing subjective well-being and self-perception in the digital health context [8].

Statistical Analysis and Power Considerations

To enhance the robustness of the estimates, bias-corrected bootstrapping was used, a technique chosen for its ability to provide more accurate CIs in path analysis [37]. This procedure, based on 5000 bootstrap samples, has been shown to provide more accurate CI estimates in mediation and path models than percentile bootstrap methods [38,39]. Further, it allows for indirect effects within the model without type 1 errors, which are often associated with percentile bootstrap methods [37,40]. In the path analysis, standardized path coefficients and explained variance (R^2) were reported to describe the strength and explanatory contribution of the modeled associations [41,42]. Effect size indexes designed for mean differences (eg, Cohen *d*) were not applicable because the analyses focused on associations among continuous variables rather than group comparisons.

Results

Overview

Table 1 shows that the participants' average age was 77.86 (SD 11.55; range 47-98) years. The sample was approximately 72.1% (31/43) White individuals, 23.3% (10/43) African Americans, and 4.7% (2/43) Asian individuals. Only 1.9% (1/53) of the participants stated that they were Hispanic. Most participants (41/50, 82%) were female, and male individuals accounted for 18% (9/50). Participants' educational levels were as follows: 34% (16/47) had a 2-year college degree, 25.5% (12/47) had a master's degree, 25.5% (12/47) had a high school diploma, and 14.9% (7/47) had a bachelor's degree. Most participants maintained their social network or contact via phone (34/53, 64.2%), and 28.3% (15/53) reported engaging with others mainly through the VSC program. Only 7.5% (4/53) of the participants reported engaging with others in person.

The average score on the MIA capacity subscale [25] was 39.00 (SD 10.06; range 14-61). The average score on the UCLA Loneliness Scale was 7.69 (SD 3.60; range 0-12), and the mean score on the Older People's Quality of Life Questionnaire was 13.22 (SD 3.45; range 3-19).

XSL•FO
RenderX

Table . Sample demographics (N=53).

	Values
Age (y), mean (SD; range)	77.86 (11.55; 47 - 98)
Gender (n=50), n (%)	
Male	9 (18)
Female	41 (82)
Race (n=43), n (%)	
Asian	2 (4.7)
Black or African American	10 (23.3)
White	31 (72.1)
Educational level (n=47), n (%)	
High school diploma	12 (25.5)
2-y college degree	16 (34)
Bachelor's degree	7 (14.9)
Master's degree	12 (25.5)
Socialization tool, n (%)	
Phone	34 (64.2)
VSC[a]	15 (28.3)
In person	4 (7.5)
OPQOL[b] score (quality of life), mean (SD; range)	7.69 (3.60; 0-12)
UCLA[c] Loneliness Scale score (reverse scored), mean (SD; range)	13.22 (3.45; 3-19)

[a]VSC: Virtual Senior Center.

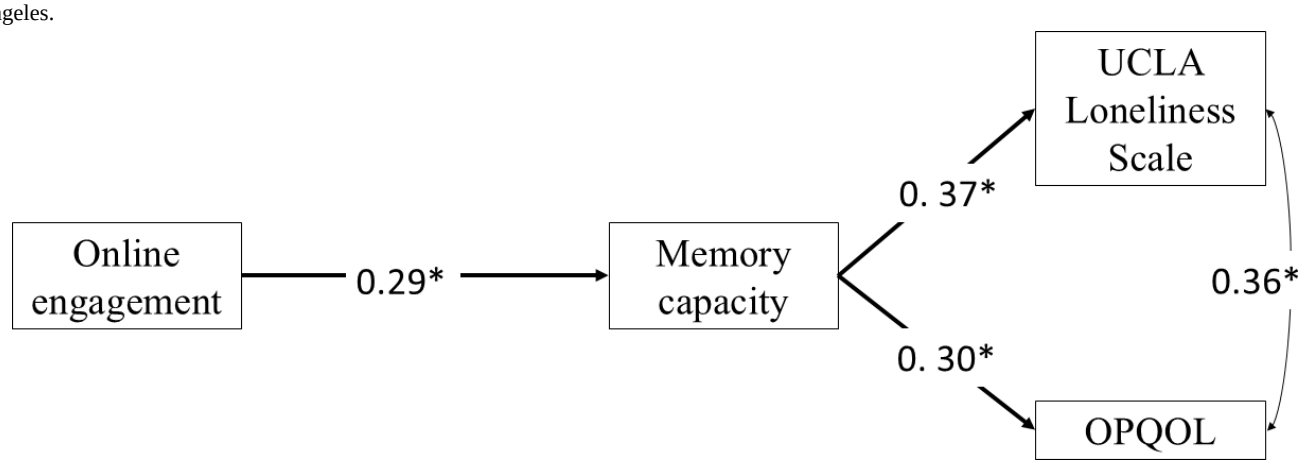
[b]OPQOL: Older People's Quality of Life Questionnaire.

[c]UCLA: University of California, Los Angeles.

Path Analysis

Path analysis revealed, as Figure 2 shows, that online engagement significantly increased older adults' subjective memory capability (β=0.29, 95% CI 0.08-0.50; *P*<.02). Individuals' memory capability was positively associated with both quality of life (β=0.90, 95% CI 0.09 - 0.57; *P*<.0001) and loneliness (β=0.37, 95% CI 0.18-0.57; *P*<.003). While demographic factors did not significantly influence participation in the VSC program, there were no direct effects on online engagement and quality of life and loneliness (refer to the full report in Table 2).

Figure 2. Path analysis with indirect effect. *P<.05. OPQOL: Older People's Quality of Life Questionnaire; UCLA: University of California, Los Angeles.



XSL•FO
RenderX

Table . Path analysis: effect of online engagement on loneliness and quality of life.

Variable	Coefficient (95% CI)	β (StdYX; 95% CI)
Age→online engagement	0.003 (−0.01 to 0.01)	0.05 (−0.19 to 0.29)
Gender→online engagement	1.19 (−0.09 to 2.48)	0.20 (−0.01 to 0.42)
Years of education→online engagement	0.02 (−0.08 to 0.13)	0.05 (−0.18 to 0.28)
Ethnicity→online engagement	1.77 (−2.26 to 5.18)	0.09 (−0.12 to 0.32)
Online engagement→memory capability	1.65[a] (0.44 to 2.86)	0.29[a] (0.08 to 0.50)
Memory capability→loneliness	0.07[a] (0.03 to 0.12)	0.37[a] (0.18 to 0.57)
Memory capability→quality of life	0.07[a] (0.02 to 0.12)	0.30[a] (0.09 to 0.50)

[a]*P*<.05.

Discussion

Interpreting the Role of Subjective Memory Capability

This study examined the association between structured online engagement and subjective memory capability among older adults participating in a VSC program. The findings suggest that increased participation in VSC activities was positively associated with subjective memory capability, which, in turn, was significantly linked to higher quality of life and reduced loneliness. However, online engagement was not directly associated with well-being measures, indicating that subjective memory capability plays a key mediating role.

A central finding of this study is the indirect effect of online engagement on well-being through subjective memory capability. While previous research has examined the benefits of online engagement for reducing social isolation and loneliness [8,11], our findings highlight a different pathway. Online engagement may be associated with cognitive self-perception, which may, in turn, relate to emotional and social well-being.

This aligns with prior studies showing that a positive self-view of memory capability can improve confidence in cognitive abilities, leading to greater social participation and engagement in intellectually stimulating activities [1,35]. The VSC program may provide an environment where older adults can reinforce their cognitive self-efficacy through structured interactions, online discussions, and virtual socialization.

Given the increasing emphasis on digital health interventions for aging populations, this study underscores the need for more tailored cognitive engagement programs that explicitly target memory self-perception and self-efficacy rather than focusing solely on social connectedness.

Comparisons With Prior Digital Interventions

Our findings align with results from the I-CONECT study that demonstrated that internet-based social interaction interventions can enhance cognitive functions in older adults [13,14]. Similarly, research on cognitive training apps and digital learning platforms has suggested that active digital engagement can reinforce cognitive reserve and maintain subjective cognitive function in aging populations [3,6].

However, this study differs in its approach by focusing on a blended intervention integrating social and cognitive engagement. Unlike traditional cognitive training programs emphasizing memory exercises or problem-solving tasks, the VSC model incorporates structured discussions, interactive virtual activities, and peer-to-peer engagement. This approach may provide a more sustainable and engaging format for older adults, particularly those who may not actively seek cognitive training but can benefit from embedded cognitive engagement within social participation. These findings suggest that digital health interventions targeting cognitive well-being should combine socialization, structured engagement, and cognitive stimulation to maximize effectiveness [7,8].

Implications for Digital Health and Aging Care

The results have several implications for digital health initiatives and gerontological practice. Structured online engagement programs such as the VSC could be incorporated into broader telemedicine and aging care strategies to support older adults in maintaining cognitive and social well-being. Given the growing prevalence of digital health platforms, integrating online engagement as a recognized intervention in aging services may help address social isolation and support cognitive well-being among older adults, particularly those who are homebound or facing mobility challenges [9,17].

Beyond integration into health care, the findings also emphasize the need for more user-centered digital platforms designed specifically for older adults. Digital interventions should prioritize features that enhance cognitive self-efficacy, such as interactive discussions, virtual memory exercises, and personalized engagement experiences. Ensuring accessibility and ease of use is equally critical as digital literacy and technological barriers remain challenges for many older adults. Addressing disparities in digital access through training programs, affordable technology, and support networks will be essential in expanding the reach and effectiveness of online engagement programs [18,19,43].

Limitations and Future Research

While this study provides valuable insights, several limitations should be considered. The cross-sectional design limits the ability to establish causality among online engagement, subjective memory capability, and well-being. Future research using longitudinal data would clarify how sustained digital participation influences cognitive and emotional outcomes over time. Additionally, while subjective memory capability is a

XSL•FO
RenderX

widely used measure in cognitive aging research, incorporating objective cognitive assessments would strengthen the findings and allow for a more comprehensive evaluation of the impact of online engagement [36,44].

The relatively small sample size presents another limitation as it restricts the generalizability of the findings to broader populations of older adults. Given the modest sample size, the findings should be interpreted as exploratory and hypothesis generating, and replication in larger samples is warranted. Future research should also explore comparative analyses of different online engagement models, including cognitive training applications, telehealth platforms, and interactive digital communities, to determine which approaches are most effective in supporting cognitive and social well-being.

Conclusions

This study examined the potential of structured online engagement programs to enhance cognitive self-perception, influencing emotional and social well-being among older adults. The findings reinforce the importance of subjective memory capability as a key mediator in the relationship between digital engagement and overall well-being. As digital health and telemedicine evolve, integrating interactive and cognitively engaging online experiences into aging care models may provide scalable solutions for promoting cognitive and social health in older populations. Expanding access to well-designed and inclusive digital programs will ensure that older adults can fully benefit from the advancements in digital health and online interventions.

Funding

The authors declared no financial support was received for this work.

Conflicts of Interest

None declared.

References

1. Glatt RM, Amos A, Merrill DA, et al. Neurocognitive effects of an online brain health program and weekly telehealth support group in older adults with subjective memory loss: a pilot study. Geriatrics (Basel) 2024 Mar 14;9(2):37. [doi: 10.3390/geriatrics9020037] [Medline: 38525754]
2. Ramos H, Alacreu M, Guerrero MD, Sánchez R, Moreno L. Lifestyle variables such as daily internet use, as promising protective factors against cognitive impairment in patients with subjective memory complaints. Preliminary results. J Pers Med 2021 Dec 14;11(12):1366. [doi: 10.3390/jpm11121366] [Medline: 34945838]
3. Benge JF, Kiselica AM, Aguirre A, et al. Technology use and subjective cognitive concerns in older adults. Arch Gerontol Geriatr 2023 Mar;106:104877. [doi: 10.1016/j.archger.2022.104877] [Medline: 36459914]
4. Cho H, Choi M, Lee H. Mobile internet use and life satisfaction among older adults: the moderating effect of living alone. J Appl Gerontol 2024 Jul;43(7):841-849. [doi: 10.1177/07334648231216383] [Medline: 37982396]
5. Ward AF. Supernormal: how the internet is changing our memories and our minds. Psychol Inq 2013 Oct;24(4):341-348. [doi: 10.1080/1047840X.2013.850148]
6. Firth JA, Torous J, Firth J. Exploring the impact of internet use on memory and attention processes. Int J Environ Res Public Health 2020 Dec 17;17(24):9481. [doi: 10.3390/ijerph17249481] [Medline: 33348890]
7. Hunsaker A, Hargittai E. A review of internet use among older adults. New Media Soc 2018 Oct;20(10):3937-3954. [doi: 10.1177/1461444818787348]
8. Stockwell S, Stubbs B, Jackson SE, Fisher A, Yang L, Smith L. Internet use, social isolation and loneliness in older adults. Ageing Soc 2021;41(12):2723-2746. [doi: 10.1017/S0144686X20000550]
9. Shi JG, Liu M, Fu G, Dai X. Internet use among older adults: determinants of usage and impacts on individuals' well-being. Comput Hum Behav 2023 Feb;139(C):107538. [doi: 10.1016/j.chb.2022.107538]
10. Tian Y, Zhang S, Wu R, Wang P, Gao F, Chen Y. Association between specific internet activities and life satisfaction: the mediating effects of loneliness and depression. Front Psychol 2018;9:1181. [doi: 10.3389/fpsyg.2018.01181] [Medline: 30050484]
11. Cotten SR, Anderson WA, McCullough BM. Impact of internet use on loneliness and contact with others among older adults: cross-sectional analysis. J Med Internet Res 2013 Feb 28;15(2):e39. [doi: 10.2196/jmir.2306] [Medline: 23448864]
12. Jun JS, Galambos C, Lee KH. Information and communication technology use, social support, and life satisfaction among Korean immigrant elders. J Soc Serv Res 2020;47(4):537-552. [doi: 10.1080/01488376.2020.1848969]
13. Wu CY, Yu K, Arnold SE, Das S, Dodge HH. Who benefited most from the internet-based conversational engagement RCT (I-CONECT)? Application of the personalized medicine approach to a behavioral intervention study. J Prev Alzheimers Dis 2024;11(3):639-648. [doi: 10.14283/jpad.2024.41] [Medline: 38706280]
14. Dodge HH, Yu K, Wu CY, et al. Internet-based conversational engagement randomized controlled clinical trial (I-CONECT) among socially isolated adults 75+ years old with normal cognition or mild cognitive impairment: topline results. Gerontologist 2024 Apr 1;64(4):gnad147. [doi: 10.1093/geront/gnad147] [Medline: 37935416]

XSL•FO
RenderX

15. Stern Y. What is cognitive reserve? Theory and research application of the reserve concept. J Int Neuropsychol Soc 2002 Mar;8(3):448-460. [Medline: 11939702]
16. Carstensen LL. Social and emotional patterns in adulthood: support for socioemotional selectivity theory. Psychol Aging 1992;7(3):331-338. [doi: 10.1037/0882-7974.7.3.331]
17. Kwok AOJ, Treiblmaier H. No one left behind in education: blockchain-based transformation and its potential for social inclusion. Asia Pacific Educ Rev 2022 Sep;23(3):445-455. [doi: 10.1007/s12564-021-09735-4]
18. Seifert A. The digital exclusion of older adults during the COVID-19 pandemic. J Gerontol Soc Work 2020;63(6-7):674-676. [doi: 10.1080/01634372.2020.1764687] [Medline: 32401181]
19. Yoon H, Jang Y, Vaughan PW, Garcia M. Older adults' internet use for health information: digital divide by race/ethnicity and socioeconomic status. J Appl Gerontol 2020 Jan;39(1):105-110. [doi: 10.1177/0733464818770772] [Medline: 29661052]
20. Ang S, Lim E, Malhotra R. Health-related difficulty in internet use among older adults: correlates and mediation of its association with quality of life through social support networks. Gerontologist 2021 Jul 13;61(5):693-702. [doi: 10.1093/geront/gnaa096] [Medline: 32744310]
21. Ornstein KA, Leff B, Covinsky KE, et al. Epidemiology of the homebound population in the United States. JAMA Intern Med 2015 Jul;175(7):1180-1186. [doi: 10.1001/jamainternmed.2015.1849] [Medline: 26010119]
22. Vu LN, Dean MJ, Mwamburi M, Au R, Qiu WQ. Executive function and mortality in homebound elderly adults. J Am Geriatr Soc 2013 Dec;61(12):2128-2134. [doi: 10.1111/jgs.12545] [Medline: 24479144]
23. Walsh EG, Clark WD. Managed care and dually eligible beneficiaries: challenges in coordination. Health Care Financ Rev 2002;24(1):63-82. [Medline: 12545599]
24. Xu D, Simpson VL. Subjective well-being, depression, and delays in care among older adults: dual-eligible versus Medicare-only beneficiaries. J Appl Gerontol 2022 Jan;41(1):158-166. [doi: 10.1177/07334648211000920] [Medline: 33736521]
25. Dixon RA, Hultsch DF. Structure and development of metamemory in adulthood. J Gerontol 1983 Nov;38(6):682-688. [doi: 10.1093/geronj/38.6.682] [Medline: 6630902]
26. Dixon RA, Hultsch DF. Metamemory and memory for text relationships in adulthood: a cross-validation study. J Gerontol 1983 Nov;38(6):689-694. [doi: 10.1093/geronj/38.6.689] [Medline: 6630903]
27. Dixon RA, Hultsch DF, Hertzog C. The Metamemory in Adulthood (MIA) questionnaire. Psychopharmacol Bull 1988;24(4):671-688. [Medline: 3249770]
28. Hertzog C, Dixon RA, Schulenberg JE, Hultsch DF. On the differentiation of memory beliefs from memory knowledge: the factor structure of the Metamemory in Adulthood Scale. Exp Aging Res 1987;13(1-2):101-107. [doi: 10.1080/03610738708259308] [Medline: 3678344]
29. Russell D, Peplau LA, Ferguson ML. Developing a measure of loneliness. J Pers Assess 1978 Jun;42(3):290-294. [doi: 10.1207/s15327752jpa4203_11] [Medline: 660402]
30. Pontinen HM, Swails JA. UCLA Loneliness Scale. In: Zeigler-Hill V, Shackelford T, editors. Encyclopedia of Personality and Individual Differences: Springer; 2017:1-3. [doi: 10.1007/978-3-319-28099-8_95-1]
31. Hays RD, DiMatteo MR. A short-form measure of loneliness. J Pers Assess 1987;51(1):69-81. [doi: 10.1207/s15327752jpa5101_6] [Medline: 3572711]
32. Bowling A. The psychometric properties of the Older People's Quality of Life Questionnaire, compared with the CASP-19 and the WHOQOL-OLD. Curr Gerontol Geriatr Res 2009;2009:298950. [doi: 10.1155/2009/298950] [Medline: 20168974]
33. Bowling A, Stenner P. Which measure of quality of life performs best in older age? A comparison of the OPQOL, CASP-19 and WHOQOL-OLD. J Epidemiol Community Health 2011 Mar;65(3):273-280. [doi: 10.1136/jech.2009.087668] [Medline: 20719807]
34. Bowling A, Hankins M, Windle G, Bilotta C, Grant R. A short measure of quality of life in older age: the performance of the brief Older People's Quality of Life questionnaire (OPQOL-brief). Arch Gerontol Geriatr 2013;56(1):181-187. [doi: 10.1016/j.archger.2012.08.012] [Medline: 22999305]
35. Zuniga KE, Mackenzie MJ, Kramer A, McAuley E. Subjective memory impairment and well-being in community-dwelling older adults. Psychogeriatrics 2016 Jan;16(1):20-26. [doi: 10.1111/psyg.12112] [Medline: 25737426]
36. Jonker C, Geerlings MI, Schmand B. Are memory complaints predictive for dementia? A review of clinical and population-based studies. Int J Geriatr Psychiatry 2000 Nov;15(11):983-991. [doi: 10.1002/1099-1166(200011)15:11<983::aid-gps238>3.0.co;2-5] [Medline: 11113976]
37. Hayes AF, Scharkow M. The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: does method really matter? Psychol Sci 2013 Oct;24(10):1918-1927. [doi: 10.1177/0956797613480187] [Medline: 23955356]
38. Bollen KA, Stine R. Direct and indirect effects: classical and bootstrap estimates of variability. Sociol Methodol 1990;20:115-140. [doi: 10.2307/271084]
39. Shrout PE, Bolger N. Mediation in experimental and nonexperimental studies: new procedures and recommendations. Psychol Methods 2002 Dec;7(4):422-445. [Medline: 12530702]
40. Fritz MS, Mackinnon DP. Required sample size to detect the mediated effect. Psychol Sci 2007 Mar;18(3):233-239. [doi: 10.1111/j.1467-9280.2007.01882.x] [Medline: 17444920]

XSL•FO
RenderX

41. Kenny DA. SEM: path analysis. DavidAKenny.net. 2011. URL: http://davidakenny.net/cm/pathanal.htm [accessed 2026-04-13]
42. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 1986 Dec;51(6):1173-1182. [doi: 10.1037//0022-3514.51.6.1173] [Medline: 3806354]
43. Seifert A, Hofer M, Rössel J. Older adults' perceived sense of social exclusion from the digital world. Educ Gerontol 2018;44(12):775-785. [doi: 10.1080/03601277.2019.1574415]
44. Gilewski MJ, Zelinski EM, Schaie KW. The Memory Functioning Questionnaire for assessment of memory complaints in adulthood and old age. Psychol Aging 1990;5(4):482-490. [doi: 10.1037/0882-7974.5.4.482] [Medline: 2278670]

Abbreviations

I-CONECT: Internet-Based Conversational Engagement
MIA: Metamemory in Adulthood
UCLA: University of California, Los Angeles
VSC: Virtual Senior Center

Edited by K Adapa; submitted 23.Feb.2025; peer-reviewed by A AL-Asadi, G Wallace, V Manchana; revised version received 19.Mar.2026; accepted 27.Mar.2026; published 06.May.2026.

Please cite as:
Lee SR, Liu H
Examining the Effect of Online Engagement on Older Adults' Subjective Memory Capability: Cross-Sectional Path Analysis
JMIR Hum Factors 2026;13:e73018
URL: https://humanfactors.jmir.org/2026/1/e73018
doi:10.2196/73018

© Soohyoung Rain Lee, Hang Liu. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 6.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

High Screen Exposure and Its Association With Physical and Mental Well-Being Among School-Going Children and Adolescents in Bangladesh: Cross-Sectional Study

Shahria Hafiz Kakon[1], MBBS, MPH; Tanjir Rashid Soron[2], MBBS, MD; Mohammad Sharif Hossain[1], MS; Biplob Hossain[1], MSc; Fahmida Tofail[1], MBBS, PhD; Rashidul Haque[1], MBBS, PhD

[1]icddr,b, 68, Shaheed Tajuddin Ahmed Sarani. Mohakhali, Dhaka, Bangladesh

[2]Tele psychiatry Research and Innovation Network Ltd, Dhaka, Bangladesh

Corresponding Author:
Shahria Hafiz Kakon, MBBS, MPH
icddr,b, 68, Shaheed Tajuddin Ahmed Sarani. Mohakhali, Dhaka, Bangladesh

Abstract

Background: In Bangladesh, as well as throughout the world, children's screen time has significantly increased. Children spend a lot of time on the internet and digital screens for entertainment, education, and communication, which has increased their daily screen time. However, the potential detrimental impacts of excessive screen time on children's mental, physical, and social health have drawn attention.

Objective: This study aimed to explore the effect of high exposure to screens on the health and mental well-being of school-going children and adolescents in Dhaka, Bangladesh.

Methods: This cross-sectional descriptive study was conducted from July 2022 to June 2024. A total of 420 school-going children and adolescents aged 6 to 14 years were enrolled from 3 English-language and 3 Bangla-language schools in Dhaka using a stratified random sampling technique. Anthropometric measurements, a semistructured questionnaire, and the Pittsburgh Sleep Quality Index, the Development and Well-Being Assessment scale, and the Strengths and Difficulties Questionnaire, all of which were validated in Bangla, were used to gather data. We considered students who were exposed to screens for less than 2 hours a day as the low-exposure group and those who were exposed for more than 2 hours a day as the high-exposure group.

Results: A total of 83.3% (350/420) of the students were in the high-exposure group, and their average screen time per day was 4.6 (SD 2.3) hours. Eye problems were reported by 35.7% (150/420) of the students, and a significant difference was found between the low- and high-exposure groups. In total, 96% (144/150) of the students with eye problems were from the high-exposure group, whereas 4% (6/150) were from the low-exposure group. Headaches were reported by 80% (336/420) of the students, and they were common in the high-exposure group (279/336, 83%). Moreover, students from the high-exposure group had a short duration and poor quality of sleep (mean 7.3, SD 1.4 hours), which was statistically significant. Furthermore, obesity was more predominant in the high-exposure group (P<.001). Our study revealed that, overall, 31% (130/420) of the students had at least one mental health problem and 9.8% (41/420) had more than one mental health problem using the Development and Well-Being Assessment scale, and mental health problems were greater in the high-exposure group than the low-exposure group. Although behavioral problems such as conduct issues (119/420, 28.3%) and peer difficulties (121/420, 28.8%) were observed among the participants, no statistically significant difference was found between the 2 groups.

Conclusions: A collaborative and coordinated multistage approach is essential to create effective and acceptable guidelines and policies for the optimum and positive use of digital screens for the children of Bangladesh. Further prospective studies on a larger scale can be conducted to determine the impacts of screen time on aspects of health.

(JMIR Hum Factors 2026;13:e73524) doi:10.2196/73524

KEYWORDS

screen time; high exposure; Strengths and Difficulties Questionnaire; Pittsburgh Sleep Quality Index; Development and Well-Being Assessment scale; Bangladesh

Introduction

The increasing availability and accessibility of digital devices have significantly influenced children's lifestyles worldwide. Screen time, defined as the amount of time spent using devices such as televisions, computers, tablets, and smartphones, has become an integral part of modern childhood [1]. While moderate and purposeful screen use can offer educational benefits and entertainment, excessive and uncontrolled screen

XSL•FO
RenderX

time is emerging as a major public health concern due to its potential negative impact on physical, mental, and social well-being, especially among school-going children. One of the primary concerns related to excessive screen time is its contribution to a sedentary lifestyle. Screen-based activities often replace physical activities, leading to decreased energy expenditure and increasing the risk of childhood obesity. Studies have demonstrated a strong correlation between prolonged screen time and higher BMI in children, with television viewing and digital gaming being notable contributors [2]. Furthermore, prolonged inactivity during screen use negatively affects cardiovascular and metabolic health, increasing the likelihood of developing conditions such as insulin resistance, hypertension, and dyslipidemia later in life [3]. Beyond weight-related issues, excessive screen time has been associated with musculoskeletal problems, such as poor posture and repetitive strain injuries. Extended use of handheld devices and computers can lead to neck, shoulder, and back pain, as well as long-term spinal alignment issues in growing children [4]. Sleep disturbances represent another critical concern in the context of excessive screen time. Exposure to the blue light emitted by screens, especially in the evening, can suppress melatonin production, disrupting circadian rhythms and delaying sleep onset. Insufficient and poor-quality sleep not only affects physical growth and immune function but also impairs cognitive performance and emotional regulation [5].

The implications of excessive screen time are not confined to physical health; it also significantly affects mental health and well-being. Research has indicated that more screen use is linked to higher rates of anxiety, depression, and emotional dysregulation among children and adolescents [6]. Higher durations of screen use are associated with increased risks of internalizing feelings such as sadness and withdrawal, as well as externalizing problems, including aggression and irritability [7]. The relationship between screen time and mental health is complicated, and it is frequently influenced by the context and goal of screen use. For example, passive activities such as binge-watching television are more strongly linked to adverse outcomes than active activities such as engagement with educational content or video calls that promote meaningful interactions [8]. The influence of these negative factors can cumulatively impact academic outcomes and overall health tremendously [9,10].

The social dimensions of screen time are equally important to consider. Excessive engagement with digital devices often reduces opportunities for face-to-face interactions, which are critical for developing communication and interpersonal skills during formative years. Prior studies suggest that excessive screen time often comes at the expense of time spent engaging in family activities, playing with peers, or participating in community events. Essential social skills such as empathy, communication, and conflict resolution—all of which are best acquired through face-to-face interaction—may be hampered by changes in behavior. Children who spend prolonged periods on screens may struggle with interpreting nonverbal cues and establishing meaningful connections, ultimately affecting their ability to build and sustain relationships [11]. Furthermore, exposure to inappropriate or violent content can normalize aggression and antisocial behaviors, further impairing social health outcomes [12].

The responsible use of technology plays a crucial role in children's education, creativity, communication, and skill development [13]. Certain educational programs can positively impact children, especially those from disadvantaged backgrounds [13]. However, excessive screen time can lead to neurobiological stimulation, affecting the hypothalamic-pituitary-adrenal axis and dopaminergic pathways, which are linked to reward processing [14]. Recognizing these risks, the World Health Organization (WHO) recommends no screen time for children under 2 years, a maximum of 1 hour a day for those aged 2 to 5 years, and less than 2 hours a day for children over 5 years [15].

However, a survey in Bangladesh found that 18.9% of children use screens for more than 6 hours daily and 48% do so for approximately 4 hours. Factors such as rapid urbanization, lack of recreational spaces, and safety concerns contribute to increased indoor activities, pushing children toward excessive screen use. With digital media becoming a central part of children's lives, it is crucial to understand its impact on their physical, mental, and social well-being. Thus, this study aims to examine the association between screen time and children's health, focusing on patterns of use associated with psychological challenges, to inform policies and interventions that promote balanced screen use and healthier lifestyles among school-going children in Dhaka, Bangladesh.

Methods

Study Design and Participants

This descriptive cross-sectional study was carried out from July 2022 to June 2024. Six schools were purposively chosen from the list of schools in the Dhaka North City Corporation and Dhaka South City Corporation. Following a screening process based on the students' exposure to digital screens and viability, 3 English-language and 3 Bangla-language schools were chosen. After obtaining permission from the school authority, a public notice was posted on the notice boards of the schools requesting students and their parents to schedule a meeting with the person conducting the research. Students who met the inclusion criteria—age between 6 and 14 years (grade 2 to grade 8) and having or not having access to gadgets such as portable gaming devices, smartphones, tablets, televisions, PCs, laptops, and gaming consoles—were chosen for the study. Following a thorough explanation of the study and provision of parents' informed written agreement, both eligible students and their parents were invited to participate. Students who had psychological impairments, including depression, anxiety, or attention-deficit/hyperactivity disorder, that impacted their well-being or who had certain preexisting mental health disorders were excluded.

We recruited 420 students in total, 70 (16.7%) from each school and 10 (2.4%) from each grade level (2 - 8), using a stratified random sampling method. The 420 participants were divided into 2 groups: one group had 70 (16.7%) students who spent less than 2 hours a day in front of devices, and the other group

XSL•FO
RenderX

included 350 (83.3%) students who spent more than 2 hours a day in front of devices. The WHO and other international health regulatory agencies established the cutoff of less than 2 hours a day for children older than 5 years [15]. To collect data at the school premises, a trained team consisting of 2 field research assistants and 1 field research supervisor was involved.

Data Collection Instruments

Overview

A semistructured pretested questionnaire was developed for both students and parents, and a pretest of the questionnaire was conducted on the field site on 30 nonstudy students. We conducted face-to-face interviews with parents and students at the school. It usually took 30 minutes for the students, and the interviews with the parents lasted approximately 60 minutes. We gathered information from parents about their children's screen time, digital screen use patterns, and sociodemographic factors. Additionally, using a self-reported questionnaire, information on typical physical health issues such as headaches, back discomfort, neck pain, and eye difficulties was gathered from the students. The comments from students about their physical health were first divided into 4 categories: "No," "Sometimes," "Often," and "Most of the time." These groupings were then merged into 2 broader categories: "No" and "Yes," where "Yes" included the responses "Sometimes," "Often," and "Most of the time."

The mental health status of the students was measured using the standardized Bangla-validated Development and Well-Being Assessment (DAWBA) scale. Social behavior and emotional problems were measured using the Bangla-validated Strengths and Difficulties Questionnaire (SDQ). Finally, the quality of sleep was measured using the Bangla-validated Pittsburgh Sleep Quality Index (PSQI). The standard pre-established parameters that apply to a healthy child were compared to the health metrics obtained from each scale.

Anthropometric Measurements

Students' height and weight were measured by a trained field research assistant, converted to *z*-scores, and then grouped based on the WHO BMI-for-age growth chart's cutoff points for those aged 5 to 19 years. These measurements were standardized before and during data collection. All BMI values, which were calculated by dividing the student's body weight by the square of their height, were categorized using the WHO cutoff criteria. According to the WHO BMI chart, the following categories apply to both boys and girls aged 6 to 14 years: underweight (<15 kg/m^2; –1 SD), normal weight (15 - 25 kg/m^2), overweight (25 - 30 kg/m^2; +1 SD), obese (30 - 40 kg/m^2; +2 SD), and severely obese (>40 kg/m^2; +4 SD) [16]. Three consecutive measures were taken, and the averaged value was used during data analysis.

SDQ Measure

The SDQ is a brief behavioral screening tool for students, where parents rate their children's psychological traits over the previous 6 months. It includes 5 categories: emotional symptoms, conduct issues, hyperactivity and attention deficiency, peer relationship problems, and prosocial behavior, each scored on a 3-point scale (0 - 2). The total score (maximum of 40) reflects overall behavioral challenges, whereas prosocial behavior is scored separately (maximum score of 10). The SDQ was translated into Bangla and validated in Bangladesh by Mullick and Goodman [17] in 2001.

PSQI Measure

The Bangla PSQI was used to evaluate overall sleep quality. Students completed this self-reported questionnaire, which assigns scores from 0 to 3 across various subcategories, including subjective sleep quality, sleep duration, sleep latency, habitual sleep efficiency, disruptions, sleep medicine use, and dysfunction during the day. A score of 5 or more indicates poor sleep quality, and the maximum score is 21 [18].

DAWBA Measure

To evaluate mental health problems, we used the DAWBA, a widely used questionnaire among children and adolescents aged 5 to 17 years. The structured questions and open-ended verbatim accounts were assessed by DAWBA raters to generate *International Classification of Diseases, 10th Revision*, or *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*, psychiatric diagnoses. In this analysis, we used the Bangla version of the DAWBA that was validated by Mullick and Goodman [19] in 2005.

Statistical Analysis

Data were imported into Microsoft Excel for statistical analysis, cleaned, and then exported into SPSS (version 20; IBM Corp) and Stata (version 15.1; StataCorp). Two main groups were formed among the participants: those with high exposure (>2 hours a day) and those with low exposure (<2 hours a day). A comparative analysis was then carried out across several variables, such as demographic characteristics, screen time, and health metrics. For categorical data, chi-square tests for independence were used to evaluate differences between groups, and for continuous data with nearly normal distributions, an unpaired 2-tailed *t* test for differences between proportions was used. A significance threshold of a *P* value of less than .05 was used to assess statistical significance.

Ethical Considerations

This study was approved by the institutional review board of icddr,b (protocol PR-22002). Confidentiality and anonymity were maintained throughout the study. Parental permission was obtained through informed written consent for all participants aged 6 to 14 years. Additionally, assent was obtained from students aged between 11 and 14 years. Every respondent received information in Bangla on their rights regarding their voluntary involvement in the study and their ability to leave the interview at any point while it was being conducted.

Results

Sociodemographic Characteristics of the Study Participants

The study included 420 students based on screening criteria, of whom 350 (83.3%) were categorized in the high-exposure group (using digital devices for more than 2 hours daily) and 70

XSL•FO
RenderX

(16.7%) were categorized in the low-exposure group. The students were aged between 6 and 14 years, with a mean age of 10.9 (SD 1.9) years, with just over half of them (n=213, 50.7%) being male. The average daily screen time of the parents was 3.9 (SD 2.7) hours, and the average daily screen time of the students was 4.6 (SD 2.3) hours according to their own statement. Table 1 presents detailed information on the sociodemographic characteristics of the study population.

Table . Sociodemographic characteristics of the study participants (N=420).

Characteristic	Overall	Low exposure (<2 h/d; n=70)	High exposure (>2 h/d; n=350)
Students' age (y), mean (SD)	10.9 (1.9)	11.2 (1.8)	10.9 (1.9)
Students' sex, n (%)			
Male	213 (50.7)	35 (50)	178 (58.9)
Female	207 (49.3)	35 (50)	172 (49.1)
Screen time–based information, mean (SD)			
Time spent by parents (h/d)	3.9 (2.7)	1.7 (1.5)	4.3 (2.7)
Time spent by the students (parents' statement; h/d)	4.5 (2.4)	1.1 (0.5)	5.1 (2.1)
Time spent by the students (students' statement; h/d)	4.6 (2.3)	1.2 (0.6)	5.2 (1.9)

Prevalence of Physical Symptoms by Screen Exposure Level

The prevalence of physical symptoms among study participants with high and low exposure is shown in Table 2. Of the 420 students included in the study, 150 (35.7%) had eye problems. Of these 150 affected students, 144 (96%) belonged to the high-exposure group, whereas only 6 (4%) were from the low-exposure group, showing a statistically significant association (P<.001). Among all participants, 80% (336/420) reported headaches, 65.7% (276/420) experienced indigestion, and 81.2% (341/420) reported abdominal pain. Additionally, 45.7% (192/420) reported neck pain, 31.9% (134/420) reported back pain, and 48.6% (204/420) reported changes in appetite. These symptoms were more prevalent in the high-exposure group, although the differences were not statistically significant. Hearing difficulties (14/420, 3.3%) and diabetes (4/420, 1%) were the least frequently reported symptoms among the high-exposure group.

XSL•FO
RenderX

Table . Prevalence of physical symptoms among study participants in the high- and low-exposure group (N=420).

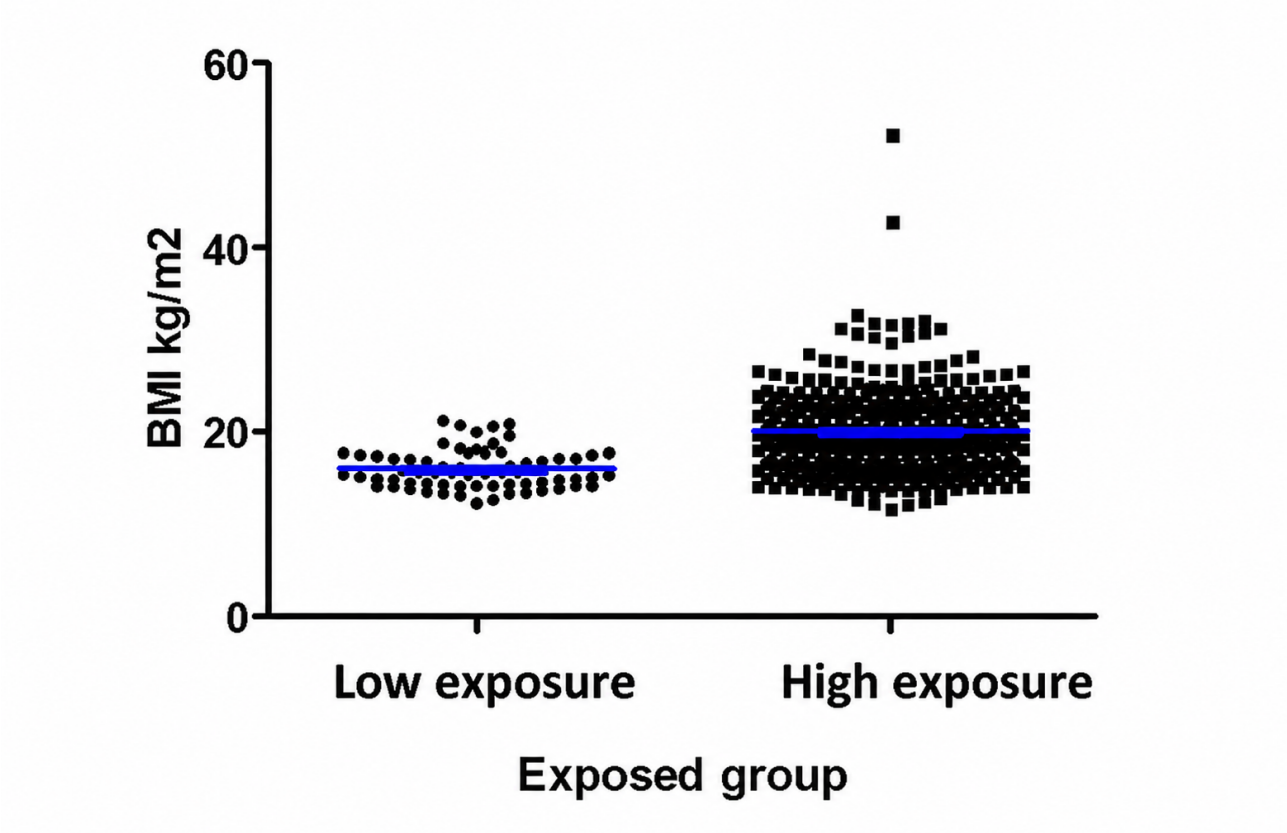
Characteristic	Overall, n (%)	Low exposure (<2 h/d; n=70), n (%)	High exposure (>2 h/d; n=350), n (%)	*P* value
Eye problems				.001
No	270 (64.3)	64 (91.4)	206 (58.9)	
Yes	150 (35.7)	6 (8.6)	144 (41.1)	
Hearing difficulties				.33
No	406 (96.7)	69 (98.6)	337 (96.3)	
Yes	14 (3.3)	1 (1.4)	13 (3.7)	
Indigestion				.78
No	144 (34.3)	25 (35.7)	119 (34)	
Yes	276 (65.7)	45 (64.3)	231 (66)	
Headache				.74
No	84 (20)	13 (18.6)	71 (20.3)	
Yes	336 (80)	57 (81.4)	279 (79.7)	
Neck pain				.43
No	228 (54.3)	41 (58.6)	187 (53.4)	
Yes	192 (45.7)	29 (41.4)	163 (46.6)	
Abdominal pain				.29
No	79 (18.8)	10 (14.3)	69 (19.7)	
Yes	341 (81.2)	60 (85.7)	281 (80.3)	
Back pain or any other musculoskeletal problem				.08
No	286 (68.1)	54 (77.1)	232 (66.3)	
Yes	134 (31.9)	16 (22.9)	118 (33.7)	
Diabetes				.37
No	416 (99)	70 (100)	346 (98.9)	
Yes	4 (1.0)	0 (0)	4 (1.1)	
Any changes in appetite				.60
No	216 (51.4)	38 (54.3)	178 (58.9)	
Yes	204 (48.6)	32 (45.7)	172 (49.1)	

Screen Time and Obesity by Level of Exposure

The mean BMI for the study population was 19.3 (SD 4.7) kg/m^2 (Figure 1). Students with high exposure to digital screens had an average BMI of 20.0 (SD 4.7) kg/m^2 compared to 15.96 (SD 2.08) kg/m^2 in the low-exposure group, which was statistically significant (P<.001). Additionally, students categorized as being overweight (42/420, 10.1%), obese (16/420, 3.8%), or severely obese (4/420 (0.9%) were predominant in the high-exposure group (P<.001) compared to the low-exposure group.

XSL•FO
RenderX

Figure 1. Screen time and obesity in the low- and high-exposure groups.



Association Between Screen Time and Behavior by Exposure Level (SDQ)

The SDQ analysis revealed that, of the 420 students, 119 (28.3%) exhibited conduct problems, 121 (28.8%) had peer relationship problems, 66 (15.7%) reported emotional issues, 73 (17.4%) experienced hyperactivity, and 28 (6.7%) exhibited prosocial behaviors. Overall, most students fell within the normal range for these domains. However, among students classified as borderline or abnormal, most were in the high-exposure group. For example, in the peer relations domain, 28.8% (121/420) of the students were classified as abnormal, of whom 87.6% (106/121) were in the high-exposure group, whereas only 12.4% (15/121) were in the low-exposure group. Despite these trends, none of the SDQ domains showed statistically significant associations with screen time. Table 3 illustrates the relationship between screen time and the SDQ among the low- and high-exposure groups.

XSL•FO
RenderX

Table . Correlation between child behavior and screen time in the low-exposure and high-exposure groups based on the Strengths and Difficulties Questionnaire (SDQ).

SDQ domains		Overall (n=420), n (%)	Low exposure (<2 h/d) (n=70), n (%)[a]	High exposure (>2 h/d) (n=350), n (%)[a]	*P* value
Emotional problems					>.99
	Normal	354 (84.3)	59 (16.7)	295 (83.3)	
	Borderline/abnormal	66 (15.7)	11 (16.7)	55 (83.3)	
Conduct problems					.96
	Normal	310 (71.7)	50 (16.6)	251 (83.4)	
	Borderline/abnormal	119 (28.3)	20 (16.8)	99 (83.2)	
Hyperactivity/inattention					.27
	Normal	347 (82.6)	61 (17.6)	286 (82.4)	
	Borderline/abnormal	73 (17.4)	9 (12.3)	64 (87.7)	
Peer relation problems					.13
	Normal	299 (71.2)	55 (18.4)	244 (81.6)	
	Borderline/abnormal	121 (28.8)	15 (12.4)	106 (87.6)	
Prosocial behavior					.86
	Normal	392 (93.3)	65 (16.6)	327 (83.4)	
	Borderline/abnormal	28 (6.7)	5 (17.9)	23 (82.1)	

[a]Percentages in these columns are calculated with the value in the "Overall" column as the denominator.

Association Between Screen Time and Sleep Quality by Exposure Level (PSQI)

The correlation between sleep and screen time in the low- and high-exposure groups according to the PSQI is shown in Table 4. The overall average sleep latency for all participants was 20.3 (SD 14.7) minutes, with the high-exposure group showing significantly longer latency (mean 21.1, SD 15.4 minutes) than the low-exposure group (mean 16.3, SD 10.1 minutes; *P*=.01). Additionally, the average sleep duration was significantly shorter in the high-exposure group (mean 7.3, SD 1.4 hours) than in the low-exposure group (mean 9.0, SD 1.4 hours; *P*<.001), indicating poorer sleep quality among students with higher screen exposure. An evaluation of the PSQI components revealed significant differences in sleep duration between the 2 groups (*P*<.001). Other components, such as breathing difficulties, bad dreams, sleep disturbances, and daytime dysfunction, were also more prevalent in the high-exposure group than in the low-exposure group.

Table . Correlation between sleep and screen time in the low- and high-exposure groups according to the Pittsburgh Sleep Quality Index (PSQI).

PSQI variable	Low exposure (<2 h/d; n=70)	High exposure (>2 h/d; n=350)	*P* value
Sleep latency (min), mean (SD)	16.3 (10.1)	21.1 (15.4)	.01
Sleep duration (h), mean (SD)	9.0 (1.4)	7.3 (1.4)	<.001
Waking up in the middle of the night, n (%)			.01
Not during the previous month	32 (45.7)	182 (52.0)	
Less than once a week	16 (22.9)	103 (29.4)	
Once or twice a week	18 (25.7)	36 (10.3)	
3 or more times a week	4 (5.7)	29 (8.3)	
Sleep duration (h), n (%)			<.001
>7	65 (92.9)	172 (49.1)	
6-7	5 (7.1)	118 (33.7)	
5-6	0 (0)	44 (12.6)	
<5	0 (0)	16 (4.6)	

XSL•FO
RenderX

Association Between Screen Time and Mental Health by Exposure Level

The correlation between student mental health and screen time in the low- and high-exposure groups according to the DAWBA scale is shown in Table 5. The DAWBA scale revealed that 31% (130/420) of the students reported at least one mental health issue, including specific phobia, generalized anxiety, depression, deliberate self-harm, hyperactivity, awkward behavior, and troublesome behavior. Additionally, 9.8% (41/420) of the students were identified as having more than one mental health problem. Notably, of the students with at least one mental health issue, 86.9% (113/130) were from the high-exposure group (screen use exceeding 2 hours per day). Similarly, among those with more than one mental health problem, 92.7% (38/41) belonged to the high-exposure group. The association between screen time and mental health problems showed borderline statistical significance, with a *P* value of .05.

Table . Correlation between student mental health and screen time in the low- and high-exposure groups according to the Development and Well-Being Assessment scale (N =420)[a].

	Overall, n (%)	Low exposure (<2 h/d; n=70), n (%)	High exposure (>2 h/d; n=350), n (%)	*P* value
No mental health problems	249 (59.3)	50 (71.4)	199 (56.9)	.05
At least one mental health problem	130 (31.0)	17 (24.3)	113 (32.3)	.05
More than one mental health problem	41 (9.8)	3 (4.3)	38 (10.9)	.05

[a]Mental health problems assessed: specific phobia, generalized anxiety, depression, deliberate self-harm, hyperactivity, awkward behavior, and troublesome behavior.

Discussion

Principal Results and Comparison With Prior Work

We found that a large majority of students (350/420, 83.3%) were exposed to digital screens for more than 2 hours per day and that higher screen exposure was strongly associated with poorer health outcomes. Nearly all students with eye problems (144/150, 96%) belonged to the high-exposure group (*P*<.001). Students with high screen time also had significantly higher BMI, with overweight (42/420, 10.1%) and obese (16/420, 3.8%) status concentrated almost entirely in this group (*P*<.001). In addition, high screen exposure was linked to poorer sleep, characterized by longer sleep latency and significantly shorter sleep duration (mean 7.3, SD 1.4 hours vs mean 9.0, SD 1.4 hours; *P*<.001). Finally, almost one-third of the students (130/420, 31%) had at least one mental health problem, with the vast majority (113/130, 86.9%) occurring among students with high screen exposure (*P*=.05).

In this study, the mean age of the participants was 10.9 (SD 1.9) years, with 49.3% (207/420) of the students being female. In a similar study, participants had a mean age of 10.9 (SD 1.16) years, and 47.7% were female [20]. Our study also found that the mean screen time per day was 4.6 (SD 2.3) hours from the students' perspective, which was similar to the parents' perspective (mean 4.5, SD 2.4 hours). Seguin et al [21] reported similar findings, which were a mean screen time of 5.9 (SD 3.4) hours per day. In addition, a systematic review showed that, overall, screen time among children aged 6 to 10 years and adolescents increased by 1.4 hours per day and 0.91 hours per day (in the year 2020-2021), respectively, compared to the pre–COVID-19 era [22]. The lockdown regimen during the pandemic brought a huge change to screen time duration and patterns in the daily routines of children. With regard to physical symptoms, 80% (336/420) of students in our study were found to have headaches. A recent study reported that there was a statistically significant association between increased screen time duration and headache, as well as dry eyes (*P*=.04 and *P*=.02, respectively) [23].

We found statistically significant associations between screen time duration and BMI. Similar findings were reported in Austria, where screen time was significantly associated with a higher BMI (*P*=.002) [24]. In contrast, a study in Pakistan demonstrated that there was no association between mean screen time and BMI (*P*=.64) [25]. The disparities may be attributable to various confounding factors such as food traditions and lifestyle.

In this study, as per the SDQ, in terms of prosocial behaviors, it was observed that only 6.7% (28/420) of the students were in the category of abnormal scores overall, with 7.1% (5/70) and 6.6% (23/350) in the lower- and higher-exposure groups, respectively. However, a study conducted on rural Chinese school-going children revealed that 10.5% of the participants belonged to the abnormal prosocial behavior category [26]. These differences may be due to other mental health problems as the community where the aforementioned study was conducted may have had various familial conflicts with lower levels of social support. In addition, previous research does show that screen time above the threshold is linked with increased abnormalities in behavior.

Regarding sleep habits and patterns, we found that duration of sleep was associated with screen time duration (*P*<.001). In a previous study in India, it was observed that both sleep efficiency and sleep duration were associated with screen time duration (*P*=.003 and *P*=.001, respectively) [27]. Another study reported that, among children who were exposed to screens for more than 2 hours a day, 61.8% had a PSQI score lower than 5 [28]. Multiple mechanisms suggest that screen time and media use are linked to sleep disturbances. There are several stimulating programs and contents that increase the activity level of the nervous system, inhibiting relaxation and leading

XSL•FO
RenderX

to anxiety and/or resulting in raised alertness and trouble falling asleep [29].

In the last decade, changes in screen use patterns and duration, type of content, type of device used, and parental monitoring have occurred. As a result, various types of studies have assessed the relationship between multiple aspects of screen time and/or device use and different facets of physical, mental, and social health in children, adolescents, and young adults, as well as academic performance outcomes and overall well-being. A study published in 2024 among children aged 6 to 10 years on screen addiction revealed that the group with high screen exposure showed statistically significant differences in screen addiction, distraction, and sedentary factors ($P<.05$) [30]. Another study reported that children who had greater exposure to screens, along with shorter sleep duration and lower physical activity levels, had poorer health-related quality of life [31]. In addition, a relevant study illustrated that children with exposure to screens for more than 2 hours per day had an increased likelihood of reported behavioral problems and poorer vocabulary [32].

It is worth mentioning that screen use widely varies across contexts. For instance, for academic purposes, screen time can exceed the recommended duration on certain days. This can be observed in the case of meeting deadlines or acquiring knowledge, where the screen is the only medium available. Another point to be kept in mind is the type of media being "overused" [33]. All in all, the context, platform, regimen, and secular trends should play a role in determining screen time duration.

Limitations

First, the 6 schools included in this study (3 in Bangla and 3 in English) may not adequately reflect the wider differences in resources between urban and rural environments. Second, there may have been recall bias because data on technological interactions were gathered through in-person interviews. Future studies on the effects of screen time ought to consider a wider range of settings and demographics. Study designs other than cross-sectional should be used, and extensive analyses involving other covariates should be conducted.

Strengths

Despite the aforementioned limitations, the study's strength lies in its innovative examination of the association between varying levels of screen exposure (low vs high) and the physical, mental, and social well-being of Bangladeshi students, particularly given that children with low exposure to digital screens are increasingly difficult to find. Furthermore, this study offers an insightful comparative analysis because it included students from Bangla- and English-language schools. The findings can inspire and inform parents, legislators, and educational authorities, emphasizing supporting students' healthier use of digital devices and more balanced lifestyles. More investigation is needed, possibly using longitudinal or experimental designs to examine causal relationships.

Conclusions

The findings of this study improve our knowledge of strategies to lessen the detrimental effects that excessive digital screen time has on children's mental health and well-being. Educators, parents, legislators, and children themselves must all be involved to create effective screen use guidelines. By holding workshops and seminars, schools can effectively teach children about responsible screen use and enforce a balanced screen time policy. To improve their children's well-being, parents must be aware of the dangers of excessive screen time as well as the advantages of active supervision. Accessible mental health resources, such as support groups and counseling, can also assist students in reducing the stress and anxiety associated with screen time. Additional research must be conducted to promote the development of comprehensive guidelines for screen time.

Acknowledgments

The authors would like to gratefully acknowledge their affiliation with icddr,b and its core donors, who provide unrestricted support to icddr,b for its operations and research, including the governments of Bangladesh and Canada. The authors are also grateful to the school authorities for their unconditional support. All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee support for the publication of this article.

Funding

This study was financially supported by the Bangladesh government's Mujib 100 Research Grant for Women in Science in Bangladesh award in collaboration with icddr,b (grant 01945).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

SHK and TRS contributed to conceptualization; analysis; and writing, revising, and finalizing the manuscript with the support of MSH, FT, and RH. All authors have read, revised, and approved the final version of the manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Qi J, Yan Y, Yin H. Screen time among school-aged children of aged 6-14: a systematic review. Glob Health Res Policy 2023 Apr 19;8(1):12. [doi: 10.1186/s41256-023-00297-z] [Medline: 37076910]
2. LeBlanc AG, Chaput JP, McFarlane A, et al. Active video games and health indicators in children and youth: a systematic review. PLoS One 2013;8(6):e65351. [doi: 10.1371/journal.pone.0065351] [Medline: 23799008]
3. Tremblay MS, LeBlanc AG, Kho ME, et al. Systematic review of sedentary behaviour and health indicators in school-aged children and youth. Int J Behav Nutr Phys Act 2011 Sep 21;8:98. [doi: 10.1186/1479-5868-8-98] [Medline: 21936895]
4. Straker L, Pollock C, Maslen B. Principles for the wise use of computers by children. Ergonomics 2009 Nov;52(11):1386-1401. [doi: 10.1080/00140130903067789] [Medline: 19851906]
5. Hale L, Guan S. Screen time and sleep among school-aged children and adolescents: a systematic literature review. Sleep Med Rev 2015 Jun;21:50-58. [doi: 10.1016/j.smrv.2014.07.007] [Medline: 25193149]
6. Twenge JM, Campbell WK. Associations between screen time and lower psychological well-being among children and adolescents: evidence from a population-based study. Prev Med Rep 2018;12:271-283. [doi: 10.1016/j.pmedr.2018.10.003] [Medline: 30406005]
7. Stiglic N, Viner RM. Effects of screentime on the health and well-being of children and adolescents: a systematic review of reviews. BMJ Open 2019 Jan 3;9(1):e023191. [doi: 10.1136/bmjopen-2018-023191] [Medline: 30606703]
8. Przybylski AK, Weinstein N. A large-scale test of the Goldilocks hypothesis. Psychol Sci 2017 Feb;28(2):204-215. [doi: 10.1177/0956797616678438] [Medline: 28085574]
9. Oberst U, Wegmann E, Stodt B, Brand M, Chamarro A. Negative consequences from heavy social networking in adolescents: the mediating role of fear of missing out. J Adolesc 2017 Feb;55:51-60. [doi: 10.1016/j.adolescence.2016.12.008] [Medline: 28033503]
10. Lissak G. Adverse physiological and psychological effects of screen time on children and adolescents: literature review and case study. Environ Res 2018 Jul;164:149-157. [doi: 10.1016/j.envres.2018.01.015] [Medline: 29499467]
11. Uhls YT, Michikyan M, Morris J, et al. Five days at outdoor education camp without screens improves preteen skills with nonverbal emotion cues. Comput Hum Behav 2014 Oct;39:387-392. [doi: 10.1016/j.chb.2014.05.036]
12. Krahé B, Möller I, Huesmann LR, Kirwil L, Felber J, Berger A. Desensitization to media violence: links with habitual media violence exposure, aggressive cognitions, and aggressive behavior. J Pers Soc Psychol 2011 Apr;100(4):630-646. [doi: 10.1037/a0021711] [Medline: 21186935]
13. Muppalla SK, Vuppalapati S, Reddy Pulliahgaru A, Sreenivasulu H. Effects of excessive screen time on child development: an updated review and strategies for management. Cureus 2023 Jun;15(6):e40608. [doi: 10.7759/cureus.40608] [Medline: 37476119]
14. Oswald TK, Rumbold AR, Kedzior SG, Moore VM. Psychological impacts of “screen time” and “green time” for children and adolescents: a systematic scoping review. PLoS One 2020;15(9):e0237725. [doi: 10.1371/journal.pone.0237725] [Medline: 32886665]
15. Choi EJ, King GK, Duerden EG. Screen time in children and youth during the pandemic: a systematic review and meta-analysis. Glob Pediatr 2023 Dec;6:100080. [doi: 10.1016/j.gpeds.2023.100080]
16. Freedman DS, Goodwin Davies AJ, Phan TLT, et al. Measuring BMI change among children and adolescents. Pediatr Obes 2022 Jun;17(6):e12889. [doi: 10.1111/ijpo.12889] [Medline: 35064761]
17. Mullick MS, Goodman R. Questionnaire screening for mental health problems in Bangladeshi children: a preliminary study. Soc Psychiatry Psychiatr Epidemiol 2001 Feb;36(2):94-99. [doi: 10.1007/s001270050295] [Medline: 11355451]
18. Islam MS, Rahman ME, Zubayer AA, et al. Investigating poor sleep quality and associated factors during the COVID-19 pandemic: a population-based survey in Bangladesh. Front Public Health 2021;9:724520. [doi: 10.3389/fpubh.2021.724520] [Medline: 34881215]
19. Mullick MSI, Goodman R. The prevalence of psychiatric disorders among 5-10 year olds in rural, urban and slum areas in Bangladesh: an exploratory study. Soc Psychiatry Psychiatr Epidemiol 2005 Aug;40(8):663-671. [doi: 10.1007/s00127-005-0939-5] [Medline: 16091858]
20. Belton S, Issartel J, Behan S, Goss H, Peers C. The differential impact of screen time on children's wellbeing. Int J Environ Res Public Health 2021 Aug 30;18(17):9143. [doi: 10.3390/ijerph18179143] [Medline: 34501733]
21. Seguin D, Kuenzel E, Morton JB, Duerden EG. School’s out: parenting stress and screen time use in school-age children during the COVID-19 pandemic. J Affect Disord Rep 2021 Dec;6:100217. [doi: 10.1016/j.jadr.2021.100217] [Medline: 34514458]
22. Trott M, Driscoll R, Irlado E, Pardhan S. Changes and correlates of screen time in adults and children during the COVID-19 pandemic: a systematic review and meta-analysis. EClinicalMedicine 2022 Jun;48:101452. [doi: 10.1016/j.eclinm.2022.101452] [Medline: 35615691]

XSL•FO
RenderX

23. Noori TA, Shalal OS, Kareem JM, Fadhil QH. Electronic devices' impact on ocular health. Al-Nisour J Med Sci 2024;6(1):157-165. [doi: 10.70492/2664-0554.1130]
24. Furthner D, Ehrenmueller M, Lanzersdorfer R, Halmerbauer G, Schmitt K, Biebl A. Education, school type and screen time were associated with overweight and obesity in 2930 adolescents. Acta Paediatr 2018 Mar;107(3):517-522. [doi: 10.1111/apa.14149] [Medline: 29131386]
25. Mukhtar S, Zubair R, Masroor S, Jaffar S, Anjum A, Naqvi S. Effect of screen-time on obesity: a cross-sectional study. Pak J Public Health 2023;13(2):75-79. [doi: 10.32413/pjph.v13i2.1114]
26. Wang H, Abbey C, Kennedy T, et al. The association between screen time and outdoor time on adolescent mental health and academic performance: evidence from rural China. Risk Manag Healthc Policy 2023;16:369-381. [doi: 10.2147/RMHP.S384997] [Medline: 36923495]
27. Saxena R, Parmar N, Kaur P, Allen T. Effect of screen-time on sleep pattern and dietary habits among college-going students in COVID-19 pandemic. Indian J Community Health 2021;33(1):65-74. [doi: 10.47203/IJCH.2021.v33i01.009]
28. Koreti S, Eske G, Saxena S. Impact of media on sleep quality of adolescents. Int J Pediatr Res 2017;4(12):721-726. [doi: 10.17511/ijpr.2017.i12.04]
29. Owens J, Maxim R, McGuinn M, Nobile C, Msall M, Alario A. Television-viewing habits and sleep disturbance in school children. Pediatrics 1999 Sep;104(3):e27. [doi: 10.1542/peds.104.3.e27] [Medline: 10469810]
30. Tekeci Y, Torpil B, Altuntaş O. The impact of screen exposure on screen addiction and sensory processing in typically developing children aged 6-10 years. Children (Basel) 2024 Apr 13;11(4):464. [doi: 10.3390/children11040464] [Medline: 38671681]
31. Wong CK, Wong RS, Cheung JP, et al. Impact of sleep duration, physical activity, and screen time on health-related quality of life in children and adolescents. Health Qual Life Outcomes 2021 May 12;19(1):145. [doi: 10.1186/s12955-021-01776-y] [Medline: 33980245]
32. McArthur BA, Tough S, Madigan S. Screen time and developmental and behavioral outcomes for preschool children. Pediatr Res 2022 May;91(6):1616-1621. [doi: 10.1038/s41390-021-01572-w] [Medline: 34012028]
33. K Kaye L, Orben A, A Ellis D, C Hunter S, Houghton S. The conceptual and methodological mayhem of "screen time". Int J Environ Res Public Health 2020 May 22;17(10):3661. [doi: 10.3390/ijerph17103661] [Medline: 32456054]

Abbreviations

DAWBA: Development and Well-Being Assessment
PSQI: Pittsburgh Sleep Quality Index
SDQ: Strengths and Difficulties Questionnaire
WHO: World Health Organization

Edited by A Kushniruk; submitted 06.Mar.2025; peer-reviewed by M Chukwuka; revised version received 13.Feb.2026; accepted 19.Feb.2026; published 04.May.2026.

Please cite as:
Kakon SH, Soron TR, Hossain MS, Hossain B, Tofail F, Haque R
High Screen Exposure and Its Association With Physical and Mental Well-Being Among School-Going Children and Adolescents in Bangladesh: Cross-Sectional Study
JMIR Hum Factors 2026;13:e73524
URL: https://humanfactors.jmir.org/2026/1/e73524
doi:10.2196/73524

© Shahria Hafiz Kakon, Tanjir Rashid Soron, Mohammad Sharif Hossain, Biplob Hossain, Fahmida Tofail, Rashidul Haque. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 4.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Telehealth Use Among Older Adults Receiving Home- and Community-Based Services: Cross-Sectional Analysis Using the National Core Indicators–Aging and Disabilities Survey

Dana P Urbanski[1], AuD, PhD; Romil R Parikh[2], MBBS, PhD; Benjamin W Langworthy[3], PhD; Jack M Wolf[3], PhD; Chanee D Fabius[4], PhD; Janette Dill[2], PhD; Eric Jutkowitz[5,6,7], PhD; Tetyana P Shippee[2], PhD

[1]College of Arts and Sciences, Indiana University Bloomington, 2631 E Discovery Pkwy, Bloomington, IN, United States

[2]Division of Health Policy & Management, School of Public Health, University of Minnesota, Minneapolis, MN, United States

[3]Division of Biostatistics & Health Data Science, School of Public Health, University of Minnesota, Minneapolis, MN, United States

[4]Department of Health Policy and Management, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

[5]Department of Health Services, Policy & Practice, School of Public Health, Brown University, Providence, RI, United States

[6]Center of Innovation in Long Term Services and Supports, Providence VA Medical Center, Providence, RI, United States

[7]Evidence Synthesis Program Center, Providence VA Medical Center, Providence, RI, United States

Corresponding Author:
Dana P Urbanski, AuD, PhD
College of Arts and Sciences, Indiana University Bloomington, 2631 E Discovery Pkwy, Bloomington, IN, United States

Abstract

Background: Telehealth was essential for maintaining care continuity during the COVID-19 pandemic, leading to its rapid adoption across the United States. Telehealth has been heralded as a strategy for improving health care access and reducing health disparities, especially for community-dwelling older adults who face significant barriers to in-person care. However, data on telehealth use among socially and financially vulnerable older adults are limited, and little is known about characteristics associated with telehealth use in this population.

Objective: Guided by the Systems Engineering Initiative for Patient Safety (SEIPS) 3.0 framework, this study examined factors associated with postpandemic telehealth use among older adults living at home and receiving publicly funded home- and community-based services (HCBS), considering HCBS receipt as an indicator of social and financial vulnerability.

Methods: This cross-sectional study included older adults aged 65 years or older living at home with available telehealth use data who participated in the 2021 - 2022 survey wave of the National Core Indicators-Aging and Disabilities Adult Consumer Survey. We used complete-case multivariable logistic regression, adjusting for sociodemographic and health-related factors with state-level random intercepts, to examine associations between telehealth use and covariates of interest (age, sex, race/ethnicity, zip code, rural-urban commuting area code, internet access, self-perceived overall health, medical transportation access, living alone, number of known non–Alzheimer disease and related dementias [ADRD] diagnoses, known ADRD diagnosis, and HCBS program/payer type). Based on the regression results, we estimated bivariate associations between internet access and key sociodemographic variables (age, sex, race/ethnicity, and zip code rural-urban commuting area) using the Pearson chi-square test. Findings were organized and interpreted through the SEIPS 3.0 framework.

Results: Of the 3680 participants, 1467 (40%) were telehealth users and 2213 (60%) were nonusers. Significantly lower odds of telehealth were observed for older adults in older age groups, males, Black individuals, those living in nonmetropolitan areas, and recipients of Older Americans Act services (odds ratios [OR] between 0.66 and 0.80). Individuals with more than one known non-ADRD diagnosis (OR 1.49, 95% CI 1.02 - 2.17) and those with an ADRD diagnosis (OR 1.33, 95% CI 1.07 - 1.66) had higher odds of telehealth use. Internet access was strongly associated with telehealth use (OR 2.51, 95% CI 2.15 - 2.92). Follow-up bivariate analyses between internet access and sociodemographic characteristics revealed that those of younger age, females, and White individuals had higher levels of internet access.

Conclusions: Differences in telehealth use among older HCBS recipients are associated with multiple individual, technological, and organizational factors. Interpreted through the SEIPS 3.0 framework, these findings underscore the importance of viewing telehealth use as the outcome of multiple features of the health care system. Future research should clarify the mechanisms driving variation in telehealth use to identify and address barriers to telehealth adoption among vulnerable older adults.

(JMIR Hum Factors 2026;13:e77522) doi:10.2196/77522

XSL•FO
RenderX

KEYWORDS

telehealth; older adults; home- and community-based services; aging in place; internet access

Introduction

During the COVID-19 pandemic, telehealth emerged as a key strategy for maintaining continuity of care amid lockdowns and social distancing measures [1-3]. Consequently, American patients and providers alike became more accepting of telehealth [4-7], contributing to a rapid increase in its use across the United States [8-10]. This surge in telehealth activity—defined as the use of telecommunications technology to provide health care remotely [11]—has since facilitated the continued integration and expansion of telehealth in United States health care delivery [9,12,13]. As its use has grown, telehealth has been increasingly viewed as a means of improving health care access and addressing potential disparities in health outcomes [14-17]. Proponents of telehealth emphasize its potential for significantly reducing cost and access barriers to care, especially for those who are medically vulnerable, geographically isolated, and socially marginalized [14-17]. For this reason, telehealth is seen as a particularly promising approach for improving access to both routine geriatric care and specialty care for chronic and complex conditions among community-dwelling older adults [17-19]. Across the United States, many older adults lack access to in-person care due to health and mobility-related challenges, combined with structural and sociodemographic factors that drive health disparities, including those related to race, ethnicity, socioeconomic status, and rurality [20-26]. For these individuals in particular, telehealth may be a promising strategy for improving access to health care and potentially reducing disparities [27].

Despite this promise, data on the demographics and characteristics of older telehealth users are limited, and it remains unclear whether telehealth improves access to care for older Americans. Available data indicate that, overall, telehealth use is relatively low among older adults, with only about one-third using telehealth [28,29]. Importantly, telehealth serves a wide range of functions beyond the management of diagnosed conditions. It is used for screening/triage, routine monitoring, medication review, mental health services, and supportive or social follow-up, services that are highly relevant for many older adults [1,6,19]. The fact that telehealth use remains relatively low among older adults despite these broad potential applications underscores the need to better understand the factors that shape telehealth use.

Broad, overarching analyses on telehealth use among older adults may conceal underlying differences in telehealth access and use driven by individual factors, such as age, sex, health status, and preference, as well as organizational and environmental factors, including health care systems, geographic location, and access to infrastructure that supports telehealth use (eg, broadband internet). These sets of factors, and their interactions, likely influence telehealth use among older adults. Recent data show that among American adults broadly, older age, male sex, rurality, and lack of internet access are negatively associated with telehealth use [30-32]. An emerging literature suggests these factors are also associated with telehealth use among older adults; however, older adults are also disproportionately affected by additional barriers to telehealth use, including inexperience with technology, lack of social support for managing telehealth technology, physical disability, and dementia [33-35]. More research is needed to identify the factors associated with telehealth use among older adults, with particular attention to conditions that contribute to potential disparities in access and use.

One such population is older adults who receive publicly funded home- and community-based services (HCBS). This group is particularly important to examine with respect to telehealth use because their substantial care and social needs make them likely to benefit from remote care while at the same time leaving them vulnerable to barriers that may limit telehealth use. HCBS are a broad range of services (such as personal care, homemaker, meal delivery, adult day care, and transportation services) designed for older adults with physical and/or cognitive functional impairments to receive care in their own homes or communities rather than in nursing homes or other institutional settings [36].

In the United States, HCBS are delivered through several publicly funded program types that vary in their structure, degree of integration with the formal health care system, and scope of benefits [37]. Major HCBS programs include Medicaid 1915 (c) waiver-funded HCBS, services authorized under the Older Americans Act (OAA) and administered by local Area Agencies on Aging (AAAs), managed long-term services and supports (MLTSS), and fully integrated care models such as the Program of All-Inclusive Care for the Elderly (PACE) [36]. Medicaid-funded HCBS programs, MLTSS, and PACE are directly connected to health care financing and typically include coverage of medical services alongside long-term services and supports. In contrast, OAA-funded services provide supportive services—such as meals, transportation, and referrals—rather than direct medical care.

More than 4.5 million older adults in the United States use HCBS [37]. This group represents a high-need population with complex health care needs, and their receipt of publicly funded HCBS serves as an indicator of their financial and social vulnerability [36]. At the same time, HCBS recipients are insured and already integrated into formal health care delivery systems, enabling examination of telehealth use among vulnerable older adults without the confounding factor of basic access to health insurance.

This study examines telehealth use among older adults who receive publicly funded HCBS and live in their own or family home. We approach this question using a human factors lens, asking not simply whether older adults adopt telehealth but how system elements at the individual, organizational, and environmental levels are associated with its use. To this end, we use the Systems Engineering Initiative for Patient Safety (SEIPS) 3.0 framework [38], a human factors model that conceptualizes health care outcomes, such as telehealth use, as emerging from interactions among persons, tools and

XSL•FO
RenderX

technology, environments, and organizations. In this study, SEIPS is used as an organizing framework to classify measured variables reflecting person characteristics (eg, age, sex, and health status), tools and technology (eg, internet access), environmental context (eg, rurality), and organizational features (eg, HCBS program type). This framework is well suited for understanding telehealth use among community-dwelling older adults receiving HCBS, whose social and financial vulnerability warrants careful attention to the individual, environmental, technological, and organizational factors that shape their interactions with the health care system.

We examine patterns of postpandemic telehealth use among older HCBS recipients using data from the National Core Indicators–Aging and Disabilities (NCI-AD) Adult Consumer Survey. It is important to note that the NCI-AD dataset reflects real-world reported telehealth use and does not indicate whether participants were explicitly offered telehealth by their providers or payers. This study therefore focuses on patterns of reported use among individuals with potential access through their existing service networks, providing insight into which subpopulations of older adults are—or are not—being reached by telehealth in practice. These findings can inform efforts to identify subpopulations of older adults who are underrepresented among telehealth users and the systems-level barriers that may limit their engagement with telehealth. Such data are essential for beginning to identify opportunities to expand telehealth among medically vulnerable older adults and ensure telehealth reaches its full potential for improving access to care.

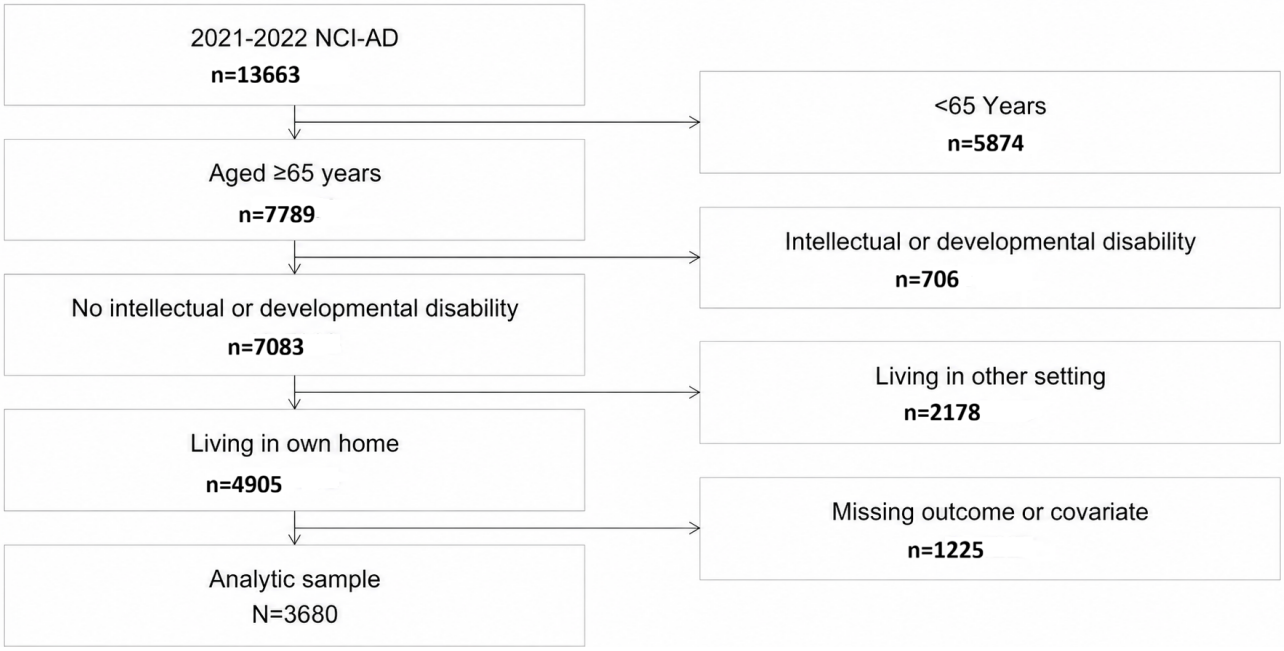
Methods

Data Source and Study Sample

This study used cross-sectional data from the 2021 - 2022 cycle of the NCI-AD Adult Consumer Survey, which aims to evaluate and monitor the effectiveness of publicly funded HCBS and other long-term care services and supports for older adults and individuals with disabilities. State participation in the NCI-AD is optional. Participating states administer NCI-AD to a representative sample of HCBS recipients each year. The NCI-AD survey consists of 2 sections. The first section gathers background information such as demographics, health conditions, and service use, typically sourced from state administrative records. The second section is a direct-conversation consumer interview and survey that collects person-centered data on the HCBS recipient's health, functional abilities, and care experiences. If an HCBS recipient cannot participate, a close proxy may complete portions of the survey that collect objective, observable information.

We focused on the 2021 - 2022 wave of the NCI-AD Adult Consumer Survey because it provided the most recent data and reflects telehealth use during the postpandemic period when telehealth use has largely stabilized [39]. The 2021 - 2022 survey included 13,663 participants from 15 states (Colorado, Delaware, Georgia, Indiana, Kansas, Michigan, Minnesota, Missouri, New Jersey, Ohio, Oklahoma, Tennessee, Texas, Washington, and Wisconsin). For the present analysis, we focused on recipients of publicly funded HCBS aged 65 years and older, who lived in their own or family home, did not have an intellectual or developmental disability, did not receive a nursing facility waiver, had nonmissing telehealth use data, and had complete covariate data (described below), resulting in a final analytic sample of 3680 participants. Figure 1 shows the detailed participant flow diagram. The percentage of missing data was below 10% for all individual items, including telehealth use. In total, 85% of participants with nonmissing telehealth data had complete data and were included in the complete-case regression analyses.

Figure 1. Flow diagram for construction of the analytic sample. NCI-AD: National Core Indicators–Aging and Disabilities.



XSL•FO
RenderX

Telehealth Use

The primary dependent variable of interest, telehealth use, was determined by an interview question that asked participants to report if they had ever communicated with health professionals via videoconferencing or telehealth. Possible responses were "yes," "no," "don't know," or "unclear/refused/no response." All "yes" responses were categorized as telehealth users, and "no" responses were considered nontelehealth users. All remaining responses were categorized as missing and excluded from the analyses.

Covariates

Covariates were selected based on the broader health care and telehealth literature and aligned with the SEIPS 3.0 human factors framework. As described previously, SEIPS 3.0 conceptualizes health care outcomes, such as telehealth use, as emerging from interactions among persons, tools and technology, environments, and organizations. Consistent with this framework, covariates were grouped according to these elements.

Person factors reflect the individual characteristics, capabilities, and experiences that shape how individuals interact with and navigate health care systems. This category included demographic characteristics (age, sex, and race/ethnicity); self-perceived overall health; lives alone; number of known non–Alzheimer disease and related dementias (ADRD) diagnoses; known ADRD diagnosis; and response type (self vs proxy response). We classified age into a categorical measure describing age group, including 65 - 74 years, 75 - 84 years, and ≥85 years; sex was categorized as female or male (no participants selected a sex other than male or female); and race and ethnicity were grouped as Black or African American, Hispanic or Latino, White, and multiple race/ethnicity. Age, sex, and race/ethnicity were populated from the background administrative section of the NCI-AD.

Self-perceived overall health was determined through the NCI-AD interview, with possible responses of poor, fair, good, very good, and excellent. The covariate "lives alone" indicates whether the participant lives independently, with participant interview responses categorized as "yes" for those living alone and "no" for those residing with others. The number of known non-ADRD diagnoses refers to the count of medical conditions excluding ADRD (including brain injury, cancer, chronic obstructive pulmonary disease, stroke, diabetes, hypertension or high blood pressure, heart disease, physical disability, and mental health diagnoses). ADRD was accounted for separately under the covariate known ADRD diagnosis, which included all individuals with a recorded ADRD diagnosis. ADRD and non-ADRD diagnoses were obtained from state administrative records when available or from participant reports, both linked to the NCI-AD survey dataset. Administrative records included provider-documented diagnoses (eg, *ICD-10* [*International Statistical Classification of Diseases, 10th Revision*] diagnosis codes or I4200 or I4800 Minimum Data Set fields). Finally, the covariate *proxy response* refers to whether the individual completed the interview independently through self-response or if a close proxy provided the answers on their behalf.

Tools and technology are the physical and digital resources that enable or constrain how individuals engage with health care systems. Within this framework, we examined internet access, which represents a foundational technological tool that enables older adults to interact with the health care system through telehealth. In the NCI-AD survey, internet access was defined as the availability of internet service at home, a coffee shop, or somewhere else the participant spends time during the day. This item was assessed via direct interview, with participant interview responses categorized as "no or sometimes" and "yes."

Environmental factors encompass the external conditions that shape how individuals interact with health care systems. These factors include geography, built environment, and community resources, which can either support or constrain care delivery and access. In this study, environmental factors included zip code rural-urban commuting area (RUCA) classification and medical transportation access. Zip code RUCA classification was used to approximate participants' geographic environment and degree of rurality and was classified into 4 levels, including metropolitan (areas with a population of 50,000 or more), micropolitan (areas with a population of between 10,000 and 49,999), small town (areas with a population under 10,000 that have commuting ties to an urban area), and rural (areas with no commuting ties to any urban area). Medical transportation access was assessed by the interview question, "Do you have transportation to get to medical appointments when you need to?" Participants could respond with "no," "sometimes," or "yes." Responses of "no" and "sometimes" were grouped to represent participants with at least some transportation barriers, while "yes" responses were retained as a separate category to represent those who have consistent access to transportation for in-person care. This variable was conceptualized as an environmental factor capturing the physical and community-level conditions that influence an individual's ability to access health care.

Organizational factors refer to the health care delivery organizations in which the patient journey occurs, characterized by workflows, policies, resource allocation, and interorganizational coordination that may influence patients' telehealth use. In this study, organizational factors were included as HCBS program type, which refers to the HCBS program administering or providing services as indicated by administrative records, classified into 3 categories, including PACE/MLTSS, OAA, and Medicaid. We combined PACE and MLTSS due to the small number of PACE participants (PACE: n=55; MLTSS: n=1259) in our sample and the shared characteristics of both programs. Both PACE and MLTSS operate under a capitated payment structure and are designed to serve older adults requiring long-term care. Since approximately 90% of our analytic sample reported receiving Medicare, consistent with United States census data [40], we focused on other payment types due to the low variance in Medicare enrollment.

Statistical Analyses

Descriptive statistics were used to summarize the analytic sample of participants with complete data for all included covariates. Bivariate associations between covariates and

XSL•FO
RenderX

telehealth use were assessed using chi-square tests and 2-sample 2-tailed *t* tests. We then estimated the adjusted associations between telehealth use and each of the covariates described using complete-case multivariable logistic regression, in which each coefficient represents the association of a given covariate with telehealth use while holding all other covariates constant. State was included as a random effect in the logistic regression model to account for potential variation in telehealth use across different states. To evaluate the impact of missing data, we conducted a sensitivity analysis using multiple imputation by chained equations [41,42], averaging results across 10 imputed datasets with no missingness (n=4786 per dataset).

Given the strong association between internet access and telehealth use in the complete-case regression analysis, we performed post hoc bivariate analyses on the analytic sample from the complete-case regression model to examine sociodemographic factors that may influence internet access as a potential precursor to telehealth use. We conducted bivariate analyses using chi-square tests to test the independence of internet access and the following factors: age, sex, race/ethnicity, and zip code RUCA classification.

All analyses were performed using R (version 4.4.2; R Project for Statistical Computing). $P<.05$ was considered statistically significant.

Ethical Considerations

This study used deidentified secondary data from the 2021-2022 NCI-AD Adult Consumer Survey. The study was submitted to the University of Minnesota Institutional Review Board and was deemed exempt from human subjects review (STUDY00010799). Because the present analysis used deidentified secondary data, additional informed consent from participants was not required. No compensation was provided to participants for this secondary analysis.

Results

Participant Characteristics

Table 1 presents the complete characteristics of the analytic sample. Percentages shown in Table 1 summarize the characteristics of the analytic sample and are not intended as population estimates; therefore, 95% CIs were not calculated for these descriptive values. The analytic sample included 3680 participants, with 1467 out of 3680 (40%) reporting ever using telehealth and 2213 out of 3680 (60%) reporting never using telehealth. Most participants were female (2667/3680, 72%) and most resided in metropolitan areas (2660/3680, 72%), while smaller proportions lived in micropolitan (528/3680, 14%), rural (219/3680, 6%), or small-town (273/3680, 7.4%) settings. The sample was racially diverse, with 1929 out of 3680 (52%) identifying as White, 858 out of 3680 (23%) as Black, 443 out of 3680 (12%) as Hispanic or Latino, and 450 out of 3680 (12%) as multiple racial/ethnic identities. Overall, 2144 out of 3680 (58%) reported having internet access, while 1536 out of 3680 (42%) reported no or intermittent access. In terms of health characteristics, 3161 out of 3680 (86%) participants did not have a documented ADRD diagnosis, and 2988 out of 3680 (81%) participants had more than one known non-ADRD diagnosis. Self-perception of health varied; most participants rated their overall health as fair (1519/3680, 41%) or good (1176/3680, 32%). About half of participants lived with at least one other person (1937/3680, 53%), and most reported having access to medical transportation (3395/3680, 92%). HCBS program enrollment was distributed among Medicaid (1630/3680, 44%), OAA programs (736/3680, 20%), and PACE/MLTSS (1314/3680, 36%).

Older HCBS recipients who were younger, female, had multiple diagnoses, lived with others, and had a known ADRD diagnosis had higher rates of telehealth use. In contrast, Black individuals and those from nonmetropolitan areas had lower rates of telehealth use compared to their counterparts. Participants with lower perceptions of their overall health tended to be more likely to use telehealth, though this relationship was not significant. Internet access showed a clear pattern with telehealth use: 49% (1047/2144) of those who reported having internet access were telehealth users, while only 27% (420/1536) of those with no or intermittent access reported using telehealth.

XSL•FO
RenderX

Table . Sample characteristics and predictor frequencies from National Core Indicators–Aging and Disabilities (NCI-AD) 2021 - 2022 by telehealth use. Participants are publicly funded home- and community-based services recipients aged 65 years and older without intellectual or developmental disabilities.

Characteristic	Overall (N=3680)	Telehealth use		*P* value[a]
		No, n=2213 (60)	Yes, n=1467 (40)	
Age (years), n (%)				<.001
65-74	1783 (48)	990 (56)	793 (44)	
75-84	1238 (34)	792 (64)	446 (36)	
≥85	659 (18)	431 (65)	228 (35)	
Sex, n (%)				.003
Female	2667 (72)	1564 (59)	1103 (41)	
Male	1013 (28)	649 (64)	364 (36)	
Race/ethnicity, n (%)				<.001
White	1929 (52)	1172 (61)	757 (39)	
Black	858 (23)	565 (66)	293 (34)	
Hispanic or Latino	443 (12)	223 (50)	220 (50)	
Multiple race/ethnicity	450 (12)	253 (56)	197 (44)	
Zip code RUCA[b], n (%)				<.001
Metropolitan	2660 (72)	1523 (57)	1137 (43)	
Micropolitan	528 (14)	350 (66)	178 (34)	
Small town	273 (7.4)	189 (69)	84 (31)	
Rural	219 (6.0)	151 (69)	68 (31)	
Internet access, n (%)				<.001
No or sometimes	1536 (42)	1116 (73)	420 (27)	
Yes	2144 (58)	1097 (51)	1047 (49)	
Self-perceived overall health, n (%)				.06
Poor	604 (16)	341 (56)	263 (44)	
Fair	1519 (41)	904 (60)	615 (40)	
Good	1176 (32)	737 (63)	439 (37)	
Very good	306 (8.3)	191 (62)	115 (38)	
Excellent	75 (2.0)	40 (53)	35 (47)	
Medical transportation access, n (%)				.10
No or sometimes	285 (7.7)	185 (65)	100 (35)	
Yes	3395 (92)	2028 (60)	1367 (40)	
Lives alone, n (%)				.008
No	1937 (53)	1125 (58)	812 (42)	
Yes	1743 (47)	1088 (62)	655 (38)	
Number of known non-ADRD[c] diagnoses				<.001
Mean (SD)	3.03 (1.62)	2.88 (1.60)	3.26 (1.63)	
Median (IQR)	3 (2-4)	3 (2-4)	3 (2-4)	
Number of known non-ADRD diagnoses (categorical), n (%)				<.001
0	166 (4.5)	121 (73)	45 (27)	
1	526 (14)	351 (67)	175 (33)	
More than 1	2988 (81)	1741 (58)	1247 (42)	

XSL•FO
RenderX

Characteristic	Overall (N=3680)	Telehealth use		*P* value[a]
		No, n=2213 (60)	Yes, n=1467 (40)	
Known ADRD diagnosis, n (%)				.001
No	3161 (86)	1935 (61)	1226 (39)	
Yes	519 (14)	278 (54)	241 (46)	
Program type, n (%)				<.001
Medicaid	1630 (44)	962 (59)	668 (41)	
OAA[d]	736 (20)	543 (74)	193 (26)	
PACE[e]/MLTSS[f]	1314 (36)	708 (54)	606 (46)	
Proxy response, n (%)				≥.99
No	2948 (80)	1773 (60)	1175 (40)	
Yes	732 (20)	440 (60)	292 (40)	

[a]Pearson chi-square test; Welch 2 sample *t* test

[b]RUCA: rural-urban commuting area.

[c]ADRD: Alzheimer disease and related dementias.

[d]OAA: Older Americans Act.

[e]PACE: Program of All-Inclusive Care for the Elderly.

[f]MLTSS: managed long-term services and supports.

Regression Analysis of Telehealth Use by SEIPS Domains

Table 2 presents the full logistic regression model results, with model estimates reported as odds ratios (ORs) and 95% CIs for telehealth use versus no telehealth use. Several significant associations between the covariates and telehealth use emerged from the analysis. Results from the sensitivity analysis using multiple imputation were consistent with those of the complete-case analysis, indicating that missing data did not substantially impact the observed associations.

XSL•FO
RenderX

Table . Results of the logistic regression model using complete-case analysis for telehealth use. Values are odds ratios (OR) and 95% CI for telehealth use versus no telehealth use.

Characteristic	Participants, n	OR[a] (95% CI)
Age (years)		
65-74	1783	—[b]
75-84	1238	0.72 (0.62-0.85)[c]
≥85	659	0.66 (0.53-0.81)[c]
Sex		
Female	2667	—
Male	1013	0.80 (0.68-0.94)[d]
Race/ethnicity		
White	1929	—
Black	858	0.72 (0.59-0.87)[c]
Hispanic or Latino	443	1.04 (0.77-1.39)
Multiple race/ethnicity	450	0.95 (0.76-1.20)
Zip code RUCA[e]		
Metropolitan	2660	—
Micropolitan	528	0.75 (0.60-0.93)[d]
Small town	273	0.67 (0.50-0.90)[d]
Rural	219	0.72 (0.52-1.00)[f]
Internet access		
No or sometimes	1536	—
Yes	2144	2.51 (2.15-2.92)[c]
Self-perceived overall health		
Poor	604	—
Fair	1519	0.91 (0.75-1.12)
Good	1176	0.85 (0.69-1.05)
Very good	306	0.89 (0.65-1.20)
Excellent	75	1.36 (0.81-2.27)
Medical transportation access		
No or sometimes	285	—
Yes	3395	1.15 (0.88-1.51)
Lives alone		
No	1937	—
Yes	1743	1.02 (0.88-1.19)
Number of known non-ADRD[g] diagnoses (categorical)		
0	166	—
1	526	1.26 (0.84-1.90)
More than 1	2988	1.49 (1.02-2.17)[f]
Known ADRD diagnosis		
No	3161	—
Yes	519	1.33 (1.07-1.66)[f]

XSL•FO
RenderX

Characteristic	Participants, n	OR[a] (95% CI)
Program type		
Medicaid	1630	—
OAA[h]	736	0.67 (0.47-0.98)[f]
PACE[i]/MLTSS[j]	1314	0.99 (0.74-1.32)
Proxy response		
No	2948	—
Yes	732	1.06 (0.86-1.30)

[a]OR: odds ratio.
[b]Not applicable.
[c]*P*<.001.
[d]*P*<.01.
[e]RUCA: rural-urban commuting area.
[f]*P*<.05.
[g]ADRD: Alzheimer disease and related dementias.
[h]OAA: Older Americans Act.
[i]PACE: Program of All-Inclusive Care for the Elderly.
[j]MLTSS: managed long-term services and supports.

Person factors showed several notable associations. Among older recipients of publicly funded HCBS, age was significantly associated with telehealth use, with older age associated with lower odds of telehealth use. Using adults aged 65-74 years as the reference group, those aged 75-84 years had an OR of 0.72 (95% CI 0.62 - 0.85), and those aged 85 years and older had an OR of 0.66 (95% CI 0.53 - 0.81). Sex was also significantly associated with telehealth use, with males having significantly lower odds of telehealth use relative to females (OR 0.80, 95% CI 0.68 - 0.94). Racial differences were observed as well: Black participants had significantly lower odds of using telehealth compared to White participants (OR 0.72, 95% CI 0.59 - 0.87), whereas Hispanic or Latino participants and those from other racial backgrounds had similar odds of telehealth use relative to White participants. Additionally, individuals with more than one known non-ADRD diagnosis had higher odds of using telehealth compared to those with no known diagnoses (OR 1.49, 95% CI 1.02 - 2.17). Those with a recorded ADRD diagnosis also had higher odds of telehealth use compared to individuals without an ADRD diagnosis (OR 1.33, 95% CI 1.07 - 1.66).

Technology and environmental factors were also associated with telehealth use. Individuals living in nonmetropolitan areas had lower odds of telehealth use compared to those in metropolitan areas (micropolitan: OR 0.75, 95% CI 0.60 - 0.93; small town: OR 0.67, 95% CI 0.50 - 0.90; rural: OR 0.72, 95% CI 0.52 - 1.00), although the rural estimate should be interpreted with caution because the upper limit of its CI approaches the null value. Internet access was strongly associated with telehealth use, with individuals reporting internet access exhibiting more than twice the odds of telehealth use compared with those with no or intermittent access (OR 2.51, 95% CI 2.15 - 2.92).

Organizational factors—specifically HCBS program type—were associated with differences in telehealth use. Participants receiving services through OAA showed lower odds of telehealth use compared with those enrolled in Medicaid (OR 0.67, 95% CI 0.47 - 0.98), while those receiving PACE/MLTSS had similar odds of telehealth use as Medicaid recipients.

No significant associations were observed between telehealth use and self-perceived overall health, medical transportation access, living alone, and response type (self vs proxy response).

Tools and Technology: Sociodemographic Correlates of Internet Access

Table 3 displays the frequencies of reported internet access, along with the *P* value from the omnibus chi-square test for each sociodemographic characteristic.

XSL•FO
RenderX

Table . Bivariate analyses of sociodemographic factors and internet access using Pearson chi-square test.

Characteristic	Internet access		*P* value[a]
	No or sometimes (n=1536)	Yes (n=2144)	
Age (years), n (%)			<.001
65-74	606 (34)	1177 (66)	
75-84	571 (46)	667 (54)	
≥85	359 (54)	300 (46)	
Sex, n (%)			.01
Female	1080 (40)	1587 (60)	
Male	456 (45)	557 (55)	
Race/ethnicity, n (%)			<.001
White	749 (39)	1180 (61)	
Black	370 (43)	488 (57)	
Hispanic or Latino	266 (60)	177 (40)	
Multiple race/ethnicity	151 (34)	299 (66)	
Zip code RUCA[b], n (%)			.06
Metropolitan	1076 (40)	1584 (60)	
Micropolitan	231 (44)	297 (56)	
Small town	128 (47)	145 (53)	
Rural	101 (46)	118 (54)	

[a]Pearson chi-square test.

[b]RUCA: rural-urban commuting area.

We conducted bivariate analyses to examine how person-level and environmental factors were associated with internet access, a key technology resource for telehealth. Our analysis revealed significant associations between internet access and age (*P*<.001), sex (*P*=.01), and race/ethnicity (*P*<.001). Participants aged 75-84 years and those aged 85 years and older were less likely to report internet access (667/1238, 54% and 300/659, 46%, respectively) than those aged 65-74 years (1177/1783, 66%). Male participants were also less likely than female participants to report having internet access (557/1013, 55% vs 1587/2667, 60%). Among racial and ethnic groups, Black and Hispanic or Latino participants (488/858, 57% and 177/443, 40%, respectively) were less likely to report internet access than White participants (1180/1929, 61%) and individuals from other racial backgrounds (66%, 299/450). No statistically significant association was found between zip code RUCA classification and internet access.

Discussion

Principal Findings

Guided by the SEIPS 3.0 framework, this study is, to our knowledge, the first to report differences in telehealth use among older HCBS recipients through a human factors lens. The SEIPS framework illustrates how health care outcomes emerge not from any single factor, but from the alignment—or misalignment—of people, technologies, organizations, and environments across the broader health care system and the patient journey within it. Applying this lens offers a systems-level perspective for identifying actionable system-level factors that may underlie variation in telehealth use.

The results of this study provide preliminary insight into the multiple system-level elements that may influence who is—and is not—being reached by telehealth. Importantly, the observed associations reflect estimates from a regression model that accounted for all included covariates, estimating each factor's association with telehealth use while adjusting for potentially confounding influences. These adjusted findings provide a clearer view of the factors that may enable or impede telehealth among older HCBS recipients.

Understanding these dynamics is particularly important as telehealth continues to expand, especially among older adults—a population recognized as a priority group for telehealth services due to their complex health needs and challenges [18,19,32,43-45]. Older adults—particularly those receiving publicly funded HCBS—are vulnerable to mobility issues, transportation barriers, chronic conditions, and poverty that make it challenging to access in-person care [46-51]. Health care organizations and policymakers have emphasized the importance of telehealth in geriatric populations for facilitating care access, improving health outcomes, reducing costs, and ultimately promoting health equity [15,52-55].

Taken together, our findings underscore the value of examining telehealth use as a product of multiple factors operating within the broader health care system. In the following sections, we

XSL•FO
RenderX

interpret our findings through the SEIPS 3.0 framework, organized by person, tools and technology, environment, and organizational factors.

Person Factors

In this study, age, sex, race/ethnicity, number of non-ADRD diagnoses, and the presence of an ADRD diagnosis were associated with differences in telehealth use among older HCBS recipients. It is important to note that compared to the broader population of community-dwelling older adults, HCBS recipients generally have more complex care needs and experience social and financial vulnerability. Consequently, the findings presented here extend prior evidence to older adults receiving HCBS.

Age was inversely associated with telehealth use, with individuals aged 85 years and older exhibiting the lowest odds of using telehealth. This finding is consistent with research demonstrating that older adults face a variety of barriers to adopting new technologies, including low digital literacy, technology-related anxiety, limited access to support, and health-related functional limitations [56-58]. Recent evidence demonstrates that these barriers hinder telehealth adoption in older adults [34] but have a more pronounced impact on those of more advanced age [59]. Consistent with our findings, several studies have similarly shown that, within older adult populations, telehealth users tend to be younger than nonusers [60,61].

We also observed sex differences in telehealth use, with males having significantly lower odds of using telehealth than females. This result aligns with other studies demonstrating that older females are more likely to use telehealth than their male counterparts [62-64]. The reasons for these differences are uncertain and may reflect broader sex differences in health care needs and use patterns among older adults [65,66], specifically those receiving HCBS [67,68]. Further research is needed to examine and explain sex differences in telehealth use among older recipients of publicly funded HCBS.

Importantly, our results revealed racial differences in telehealth use. In our sample of older HCBS recipients, Black older adults had significantly lower odds of using telehealth compared to White older adults, consistent with other studies showing lower telehealth use among Black older adults [33,69,70]. This finding may reflect systemic and structural barriers to telehealth access, including pervasive racial and ethnic disparities in access to technology and broadband internet, a relationship further supported by our bivariate analyses showing differences in internet access by race and ethnicity [71-73].

Beyond sociodemographic characteristics, HCBS recipients with more than one non-ADRD diagnosis had significantly higher odds of using telehealth compared to those with no known non-ADRD diagnoses. This finding aligns with prior research showing that individuals with multimorbidity—defined as the presence of 2 or more chronic conditions—are more likely to use telehealth than those with fewer health concerns [74,75]. Multimorbidity often requires ongoing communication with multiple health care providers, making telehealth a valuable option to reduce the burden of frequent in-person visits. Our results reinforce geriatric care recommendations emphasizing telehealth's potential to support older adults with chronic health conditions [45], particularly those who are homebound or have difficulty traveling to medical appointments [76]. This may be especially important for older HCBS recipients with multimorbidity, whose HCBS eligibility reflects high care needs and increased vulnerability to systemic and health-related barriers that limit access to in-person care. Future research should examine how telehealth can be integrated into the care of older HCBS recipients with multiple health conditions and how programmatic or policy changes could promote its use in this population.

Older HCBS recipients with a documented ADRD diagnosis had slightly higher odds of telehealth use than those without an ADRD diagnosis. This finding aligns with prior research showing that telehealth is generally well received and shows promise for people with ADRD and their caregivers [77-79]. For people living with dementia and their caregivers, telehealth can reduce the burden of travel to in-person health care appointments and facilitate more frequent assessments and follow-up care with health care providers [79]. Our results reinforce telehealth's potential value for older HCBS recipients with ADRD while also pointing to opportunities to expand its reach in this population. In our sample, only 46% (241/519) of older HCBS recipients with ADRD reported telehealth use—suggesting that while these individuals were slightly more likely to use telehealth than their counterparts, there remains an opportunity to increase uptake. Future research should explore how telehealth services can meet the unique needs of older HCBS recipients with ADRD.

Finally, several person factors were not significantly associated with telehealth use, including self-perceived overall health, living alone, and whether survey responses were provided by the participant or a proxy. These null findings suggest that, within this population, telehealth use may be more strongly patterned by sociodemographic characteristics and objective health indicators, such as documented diagnoses, rather than by subjective health perceptions or cohabitation status. Viewed together, our findings show that person-level characteristics and capabilities—core components of the SEIPS framework—may influence telehealth use among older adults receiving publicly funded HCBS.

Tools and Technology

Disparities in technology use—including telehealth—are often linked to gaps in broadband internet availability for vulnerable and marginalized older adults [80]. Our results underscore this relationship, revealing a strong association between internet access and telehealth use. In our sample, older adults with internet access had more than twice the odds of using telehealth compared to those with no or intermittent access. This finding aligns with prior research emphasizing the critical role of internet access in enabling telehealth use [28,81,82].

Our findings support the hypothesis that internet access functions as a core technological tool connecting people to health care systems. When absent or unreliable, the broader system's capacity to deliver care through telehealth may be compromised. As telehealth continues to expand, increasing broadband internet access is essential to promote equitable

XSL•FO
RenderX

access to telehealth services, especially for medically, socially, and financially vulnerable older adults [83]. Evidence suggests that state-level funding programs increase broadband availability, highlighting the potential impact of policy interventions on improving internet access [84]. Conversely, municipal and cooperative restrictions have been shown to decrease broadband availability, underscoring the role of regulatory barriers in limiting connectivity, such as state policies that restrict municipalities and electric cooperatives from providing broadband service—often justified as preventing publicly subsidized competition with private internet providers [84]. Targeted policy efforts, such as state and federal initiatives subsidizing broadband internet for low-income and rural populations, could facilitate more equitable telehealth access for vulnerable older adults.

Consistent with patterns observed in the person factors domain, internet access varied significantly by age, sex, and race/ethnicity. Older adults of more advanced age were less likely to have internet access than those aged 65-74 years, which may help explain the lower odds of telehealth use observed in this group. Additionally, males were less likely than females to report having internet access, aligning with our finding of lower odds of telehealth use among male participants. A prior review found mixed evidence regarding sex differences in internet use among older adults, with some studies reporting no sex effect and others showing higher internet access among older men than women [85]. Our findings differ from these results, possibly due to our focus on older recipients of publicly funded HCBS, a population disproportionately composed of women with care needs experiencing social and financial vulnerability [68]. Additional research is needed to characterize age- and sex-related patterns of internet access in this population and their relationship to telehealth use.

Our results highlight differences in internet access by race and ethnicity. In our sample of vulnerable older adults, Black and Hispanic or Latino individuals were less likely to report having internet access than White individuals and those from other racial backgrounds. This finding may help explain why Black individuals had lower odds of telehealth use than their White counterparts in our regression analysis. Interestingly, we found that Hispanic or Latino older adults were less likely to have internet access; however, their odds of telehealth use were comparable to White individuals. These results suggest differing impacts of internet access on telehealth use but generally align with extensive research documenting racial and ethnic disparities in internet access and use [26,71-73]. Addressing these gaps may require targeted programs that expand internet access among minoritized older adults, including those receiving publicly funded HCBS. Such initiatives could serve as precursors to telehealth use by ensuring vulnerable older adults have the technological tools they need to use telehealth.

Unexpectedly, internet access was not significantly associated with zip code RUCA classification. This result contrasts with extensive literature showing lower rates of broadband internet availability and access in rural communities compared to metropolitan areas [86-88]. Furthermore, this finding was unexpected given that in this sample, older HCBS recipients in nonmetropolitan settings demonstrated lower odds of telehealth use compared to those in metropolitan areas. The lack of association between internet access and zip code RUCA may reflect the smaller sample size of rural participants in our analyses. Future research should focus on characterizing internet access and digital disparities among older HCBS recipients, particularly those in traditionally underserved rural areas. Identifying and targeting specific communities with limited broadband internet access could help address gaps in telehealth use by ensuring older adults have access to foundational technology required for telehealth uptake.

Environmental Factors

We found that older adults residing in nonmetropolitan areas had lower odds of using telehealth compared to those in metropolitan areas. Our estimate for rural residents should be interpreted with caution, as the CI included the null value. Likewise, bivariate analyses showed no significant association between zip code RUCA classification and internet access. Together, these findings suggest that, within this sample, rurality alone may not capture the full range of environmental mechanisms shaping telehealth use. Nevertheless, rural older adults remain a population for whom telehealth is viewed as particularly beneficial, given the limited availability of specialty care in rural areas and the challenges of traveling long distances to receive services [89]. Although extensive prior evidence shows that rural older adults experience a persistent digital divide—including lower broadband availability, reduced device access, and lower rates of technology adoption compared to metropolitan peers [90-93]—our findings suggest that rurality alone may not directly translate into lower telehealth use within the HCBS population. This may indicate that other system components, such as person, technology, and organizational factors, help compensate for structural challenges based on environmental context. In this study, the absence of a strong rural effect could reflect offsetting influences—such as targeted initiatives to expand or prioritize telehealth access in local areas. This underscores the complexity of telehealth use as an outcome shaped by multiple system elements rather than a single environmental barrier.

Access to medical transportation was not significantly associated with telehealth use. This may reflect that transportation barriers do not clearly determine whether an older HCBS recipient uses telehealth as an alternative to in-person appointments. However, transportation remains an important environmental resource within the broader health care system, likely interacting with person and organizational factors to influence how older adults access and coordinate care. Taken together, the weak and/or null associations for environmental factors in this study suggest that other system variables may play a larger role in shaping telehealth access.

Organizational Factors

Older HCBS recipients enrolled in OAA had lower odds of telehealth use compared to those enrolled in Medicaid. This difference may reflect underlying organizational differences in how these programs connect participants to the health care system. As mentioned previously, OAA programs, administered by local AAAs, primarily provide supportive services such as home-delivered meals, transportation, and referrals rather than

XSL•FO
RenderX

direct medical care benefits. While these services are vital for many older adults, they do not directly address or translate to formal health care access. At most, OAA services are a "bridge" that supports better care access and coordination [94]. However, variability in local AAA resources and service delivery makes it difficult to draw generalizations about the relationship between OAA services and health care access. From these data, we also cannot determine whether OAA recipients differ meaningfully from other HCBS participants in their care needs, which could also contribute to differences in telehealth use. This uncertainty underscores how multiple system elements—such as person factors (eg, health needs) and organizational interfaces—may work together to shape telehealth use. Based on our findings, it is plausible that OAA programs provide a weaker interface between the person and the medical care system, offering essential social support but less integration with health care delivery pathways, which may result in lower telehealth use.

By contrast, Medicaid, PACE, and MLTSS programs are structured to deliver comprehensive health care with low out-of-pocket costs to eligible older adults, often including telehealth as a covered option. These organizational systems offer a stronger interface between the person and the health care system. Consequently, individuals enrolled in these programs may have clearer pathways to telehealth services compared to those enrolled in OAA.

These findings highlight the importance of considering HCBS recipients' organizational interfaces with the health care system and how they may offer or shape opportunities for telehealth use. Further research is needed to explore telehealth access among older HCBS recipients across different program types. Understanding these organizational dynamics is essential for designing system-level interventions that align HCBS program structures with the evolving role of telehealth.

Strengths and Limitations

Guided by the SEIPS 3.0 framework, this study leverages a unique dataset to examine factors associated with telehealth use among older adults receiving HCBS. This hard-to-reach population offers valuable insights into telehealth use among socially and financially vulnerable older adults, who are considered likely to benefit from these services [14-17]. However, the NCI-AD dataset has limitations, including a lack of information on the type and frequency of telehealth visits, preference for in-person versus telehealth care, telehealth satisfaction, and future intentions to use telehealth. Data regarding reasons for telehealth nonuse are also unavailable. In addition, the dataset does not capture whether participants were explicitly offered telehealth by their providers or payers, which may introduce variability in access opportunities that is not reflected in our analyses. Consequently, we could not analyze more nuanced aspects of telehealth use and nonuse within this sample. Additionally, the cross-sectional nature of the NCI-AD data limits the ability to draw causal inferences. Recall bias may also affect the accuracy of responses, as participants may not accurately report their telehealth and other health experiences. Furthermore, because the dataset primarily captures older adults receiving HCBS, our findings are most applicable to community-dwelling older adults and may not generalize to those residing in congregate settings. Future research should use longitudinal designs to better understand telehealth use and to examine how interactions across SEIPS components influence telehealth access, use, and effectiveness among older HCBS recipients and older adults more broadly.

Conclusion

This study uses the SEIPS 3.0 human factors framework to examine differences in telehealth use among older recipients of HCBS living at home, a socially and financially vulnerable population that may benefit greatly from telehealth services in reducing or removing barriers to in-person health care. Our results show that among older HCBS recipients, older individuals, males, those living in nonmetropolitan areas, Black individuals (relative to White), and OAA recipients (relative to those receiving Medicaid) had lower odds of using telehealth. Individuals with multiple diagnoses and those with a documented ADRD diagnosis had higher odds of telehealth use. Notably, internet access was strongly associated with telehealth use and, in bivariate analyses, was associated with similar sociodemographic factors as those linked to telehealth use, emphasizing the need for expanded broadband internet access to promote equitable telehealth use among vulnerable older adults. Viewed through the SEIPS model, the findings reveal differences in telehealth use as an outcome emerging from multiple system elements rather than any single determinant. Future research should identify and address the potential mechanisms underlying these differences to ensure telehealth fulfills its potential to improve access to care and promote health equity for older Americans.

Acknowledgments

We appreciate the collaboration and expertise of our colleagues at the Human Services Research Institute, Steph Giordano and Nilufer Isvan, as well as Wyatt Tarter's valuable statistical support. AI tools were not used.

Funding

This work was funded by the National Institute on Aging of the National Institutes of Health under grants 1R01AG069771-01, 1RF1AG069771-01, and 1R01AG060871.

XSL•FO
RenderX

Data Availability

The datasets analyzed during this study are not publicly available because participating state agencies own their state's data, but deidentified National Core Indicators-Aging and Disabilities data are available for secondary analysis by request through National Core Indicators.

Conflicts of Interest

Eric Jutkowitz is a co-founder and on the board of directors of Plans4Care, Inc, a digital health company that provides personalized dementia care on demand.

References

1. Andrews E, Berghofer K, Long J, Prescott A, Caboral-Stevens M. Satisfaction with the use of telehealth during COVID-19: an integrative review. Int J Nurs Stud Adv 2020 Nov;2:100008. [doi: 10.1016/j.ijnsa.2020.100008] [Medline: 33083791]
2. Bitar H, Alismail S. The role of eHealth, telehealth, and telemedicine for chronic disease patients during COVID-19 pandemic: a rapid systematic review. Digit Health 2021;7:20552076211009396. [doi: 10.1177/20552076211009396] [Medline: 33959378]
3. Wosik J, Fudim M, Cameron B, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Jun 1;27(6):957-962. [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]
4. Crouch J, Winters K, Zhang L, Stewart MW. Telehealth during the pandemic: patient perceptions and policy implications. J Nurs Scholarsh 2023 Jan;55(1):141-148. [doi: 10.1111/jnu.12832] [Medline: 36218219]
5. Walia B, Shridhar A, Arasu P, Singh GK. US physicians' perspective on the sudden shift to telehealth: survey study. JMIR Hum Factors 2021 Aug 12;8(3):e26336. [doi: 10.2196/26336] [Medline: 33938813]
6. Berry CA, Kwok L, Massar R, et al. Patients' perspectives on the shift to telemedicine in primary and behavioral health care during the COVID-19 pandemic. J Gen Intern Med 2022 Dec;37(16):4248-4256. [doi: 10.1007/s11606-022-07827-4] [Medline: 36167954]
7. Hong YR, Lawrence J, Williams D Jr, Mainous III A. Population-level interest and telehealth capacity of US hospitals in response to COVID-19: cross-sectional analysis of Google search and national hospital survey data. JMIR Public Health Surveill 2020 Apr 7;6(2):e18961. [doi: 10.2196/18961] [Medline: 32250963]
8. Neri AJ, Whitfield GP, Umeakunne ET, et al. Telehealth and public health practice in the United States-before, during, and after the COVID-19 pandemic. J Public Health Manag Pract 2022;28(6):650-656. [doi: 10.1097/PHH.0000000000001563] [Medline: 36037509]
9. Shaver J. The state of telehealth before and after the COVID-19 pandemic. Prim Care 2022 Dec;49(4):517-530. [doi: 10.1016/j.pop.2022.04.002] [Medline: 36357058]
10. Patel SY, Mehrotra A, Huskamp HA, Uscher-Pines L, Ganguli I, Barnett ML. Variation In telemedicine use and outpatient care during the COVID-19 pandemic In the United States. Health Aff (Millwood) 2021 Feb;40(2):349-358. [doi: 10.1377/hlthaff.2020.01786] [Medline: 33523745]
11. Frehse AE. Overview and history of telehealth. In: Telemedicine: Overview and Application in Pulmonary, Critical Care, and Sleep Medicine: Springer International Publishing; 2021:3-14. [doi: 10.1007/978-3-030-64050-7_1]
12. Saharkhiz M, Rao T, Parker-Lue S, Borelli S, Johnson K, Cataife G. Telehealth expansion and medicare beneficiaries' care quality and access. JAMA Netw Open 2024 May 1;7(5):e2411006. [doi: 10.1001/jamanetworkopen.2024.11006] [Medline: 38739388]
13. Chapter 7: mandated report: telehealth in medicare (June 2023 report). Medicare Payment Advisory Commission. 2023. URL: https://www.medpac.gov/document/chapter-7-mandated-report-telehealth-in-medicare-june-2023-report/ [accessed 2025-05-14]
14. Bailey JE, Gurgol C, Pan E, et al. Early patient-centered outcomes research experience with the use of telehealth to address disparities: scoping review. J Med Internet Res 2021 Dec 7;23(12):e28503. [doi: 10.2196/28503] [Medline: 34878986]
15. Shah DA, Sall D, Peng W, Sharer R, Essary AC, Radhakrishnan P. Exploring the role of telehealth in providing equitable healthcare to the vulnerable patient population during COVID-19. J Telemed Telecare 2024 Jul;30(6):1047-1050. [doi: 10.1177/1357633X221113711] [Medline: 35833345]
16. Agate S. Unlocking the power of telehealth: increasing access and services in underserved, urban areas. Harv J Hisp Policy 2017;29:85-96 [FREE Full text]
17. Ezeamii VC, Okobi OE, Wambai-Sani H, et al. Revolutionizing healthcare: how telemedicine is improving patient outcomes and expanding access to care. Cureus 2024 Jul;16(7):e63881. [doi: 10.7759/cureus.63881] [Medline: 39099901]
18. Dryden EM, Kennedy MA, Conti J, et al. Perceived benefits of geriatric specialty telemedicine among rural patients and caregivers. Health Serv Res 2023 Feb;58 Suppl 1(Suppl 1):26-35. [doi: 10.1111/1475-6773.14055] [Medline: 36054487]
19. Şahin E, Yavuz Veizi BG, Naharci MI. Telemedicine interventions for older adults: a systematic review. J Telemed Telecare 2024 Feb;30(2):305-319. [doi: 10.1177/1357633X211058340] [Medline: 34825609]
20. Tilhou AS, Jain A, DeLeire T. Telehealth expansion, internet speed, and primary care access before and during COVID-19. JAMA Netw Open 2024 Jan 2;7(1):e2347686. [doi: 10.1001/jamanetworkopen.2023.47686] [Medline: 38180762]

XSL•FO
RenderX

21. Cabañero-Garcia E, Martinez-Lacoba R, Pardo-Garcia I, Amo-Saus E. Barriers to health, social and long-term care access among older adults: a systematic review of reviews. Int J Equity Health 2025 Mar 12;24(1):72. [doi: 10.1186/s12939-025-02429-y] [Medline: 40075452]
22. Monnat SM. U.S. rural population health and aging in the 2020s. Public Policy Aging Rep 2025;35(1):3-9. [doi: 10.1093/ppar/prae031] [Medline: 40135186]
23. Krasniuk S, Crizzle AM. Impact of health and transportation on accessing healthcare in older adults living in rural regions. Transp Res Interdiscip Perspect 2023 Sep;21:100882. [doi: 10.1016/j.trip.2023.100882]
24. Lee K, Cassidy J, Mitchell J, Jones S, Jang SW. Documented barriers to health care access among Latinx older adults: a scoping review. Educ Gerontol 2023 Feb 1;49(2):81-95. [doi: 10.1080/03601277.2022.2082714]
25. McMaughan DJ, Oloruntoba O, Smith ML. Socioeconomic status and access to healthcare: interrelated drivers for healthy aging. Front Public Health 2020;8(231):231. [doi: 10.3389/fpubh.2020.00231] [Medline: 32626678]
26. Fabius CD, Parikh R, Wolf JM, et al. Racial and ethnic differences in unmet needs among older adults receiving publicly-funded home and community-based services. J Am Geriatr Soc 2024 Nov;72(11):3520-3529. [doi: 10.1111/jgs.19153] [Medline: 39210674]
27. McElroy JA, Day TM, Becevic M. The influence of telehealth for better health across communities. Prev Chronic Dis 2020 Jul 16;17(64):E64. [doi: 10.5888/pcd17.200254] [Medline: 32678060]
28. Okoye SM, Mulcahy JF, Fabius CD, Burgdorf JG, Wolff JL. Neighborhood broadband and use of telehealth among older adults: cross-sectional study of national survey data linked with census data. J Med Internet Res 2021 Jun 14;23(6):e26242. [doi: 10.2196/26242] [Medline: 34125071]
29. Spaulding EM, Fang M, Commodore-Mensah Y, Himmelfarb CR, Martin SS, Coresh J. Prevalence and disparities in telehealth use among US adults following the COVID-19 pandemic: national cross-sectional survey. J Med Internet Res 2024 May 10;26:e52124. [doi: 10.2196/52124] [Medline: 38728070]
30. Narcisse MR, Andersen JA, Felix HC, Hayes CJ, Eswaran H, McElfish PA. Factors associated with telehealth use among adults in the United States: findings from the 2020 national health interview survey. J Telemed Telecare 2024 Jul;30(6):993-1004. [doi: 10.1177/1357633X221113192] [Medline: 35892167]
31. Chang E, Penfold RB, Berkman ND. Patient characteristics and telemedicine use in the US, 2022. JAMA Netw Open 2024 Mar 4;7(3):e243354. [doi: 10.1001/jamanetworkopen.2024.3354] [Medline: 38517438]
32. Zhang D, Shi L, Han X, et al. Disparities in telehealth utilization during the COVID-19 pandemic: findings from a nationally representative survey in the United States. J Telemed Telecare 2024 Jan;30(1):90-97. [doi: 10.1177/1357633X211051677] [Medline: 34633882]
33. Lam K, Lu AD, Shi Y, Covinsky KE. Assessing telemedicine unreadiness among older adults in the United States during the COVID-19 pandemic. JAMA Intern Med 2020 Oct 1;180(10):1389-1391. [doi: 10.1001/jamainternmed.2020.2671] [Medline: 32744593]
34. Kalicki AV, Moody KA, Franzosa E, Gliatto PM, Ornstein KA. Barriers to telehealth access among homebound older adults. J Am Geriatr Soc 2021 Sep;69(9):2404-2411. [doi: 10.1111/jgs.17163] [Medline: 33848360]
35. Gallegos-Rejas VM, Thomas EE, Kelly JT, Smith AC. A multi-stakeholder approach is needed to reduce the digital divide and encourage equitable access to telehealth. J Telemed Telecare 2023 Jan;29(1):73-78. [doi: 10.1177/1357633X221107995] [Medline: 35733379]
36. Shippee TP, Bucy TI, Parikh RR, et al. Service utilization and consumer-reported unmet needs in medicaid home- and community-based services for older adults. J Aging Soc Policy 2025;37(5):817-831. [doi: 10.1080/08959420.2024.2422666] [Medline: 39679540]
37. Burns A, Wolk A, O'Malley Watts M, Mohamed M, Peña MT. A look at waiting lists for medicaid home- and community-based services from 2016 to 2024. KFF. 2024 Oct 31. URL: https://www.kff.org/medicaid/issue-brief/a-look-at-waiting-lists-for-medicaid-home-and-community-based-services-from-2016-to-2024 [accessed 2025-05-14]
38. Carayon P, Wooldridge A, Hoonakker P, Hundt AS, Kelly MM. SEIPS 3.0: human-centered design of the patient journey for patient safety. Appl Ergon 2020 Apr;84:103033. [doi: 10.1016/j.apergo.2019.103033] [Medline: 31987516]
39. Andino JJ, Eyrich NW, Boxer RJ. Overview of telehealth in the United States since the COVID-19 public health emergency: a narrative review. Mhealth 2023;9:26. [doi: 10.21037/mhealth-23-15] [Medline: 37492124]
40. Dual health insurance coverage declining for adults age 65 and over. United States Census Bureau. 2024 Apr 18. URL: https://www.census.gov/library/stories/2024/04/older-adults-health-coverage.html [accessed 2025-05-14]
41. Rubin DB. Multiple Imputation for Nonresponse in Surveys: John Wiley & Sons; 1987:1-258. [doi: 10.1002/9780470316696]
42. van Buuren S. Flexible Imputation of Missing Data, 2nd edition: Chapman & Hall/CRC Press; 2018:1-444. [doi: 10.1201/9780429492259]
43. Lillicrap L, Hunter C, Goldswain P. Improving geriatric care and reducing hospitalisations in regional and remote areas: the benefits of telehealth. J Telemed Telecare 2021 Aug;27(7):397-408. [doi: 10.1177/1357633X19881588] [Medline: 31645171]
44. Morley JE. Editorial: telehealth and geriatrics. J Nutr Health Aging 2021;25(6):712-713. [doi: 10.1007/s12603-021-1643-1] [Medline: 34179920]

XSL•FO
RenderX

45. Johns T, Huot C, Jenkins JC. Telehealth in geriatrics. Prim Care 2022 Dec;49(4):659-676. [doi: 10.1016/j.pop.2022.04.009] [Medline: 36357069]
46. Syed ST, Gerber BS, Sharp LK. Traveling towards disease: transportation barriers to health care access. J Community Health 2013 Oct;38(5):976-993. [doi: 10.1007/s10900-013-9681-1] [Medline: 23543372]
47. Suntai Z, Won CR, Noh H. Access barrier in rural older adults' use of pain management and palliative care services: a systematic review. Am J Hosp Palliat Care 2021 May;38(5):494-502. [doi: 10.1177/1049909120959634] [Medline: 32959683]
48. Gong N, Meng Y, Hu Q, et al. Obstacles to access to community care in urban senior-only households: a qualitative study. BMC Geriatr 2022 Feb 14;22(1):122. [doi: 10.1186/s12877-022-02816-y] [Medline: 35164708]
49. Twardzik E, Guralnik JM, Falvey JR. Community mobility among older adults who are socioeconomically disadvantaged: addressing the poverty penalty. Phys Ther 2024 Oct 2;104(10):pzad182. [doi: 10.1093/ptj/pzad182] [Medline: 38157292]
50. Ganguli I, Orav EJ, Lii J, Mehrotra A, Ritchie CS. Which medicare beneficiaries have trouble getting places like the doctor's office, and how do they do it? J Gen Intern Med 2023 Jan;38(1):245-248. [doi: 10.1007/s11606-022-07615-0] [Medline: 35469357]
51. Wolfe MK, McDonald NC, Holmes GM. Transportation barriers to health care in the United States: findings from the national health interview survey, 1997–2017. Am J Public Health 2020 Jun;110(6):815-822. [doi: 10.2105/AJPH.2020.305579] [Medline: 32298170]
52. Masterson Creber R, Dodson JA, Bidwell J, et al. Telehealth and health equity in older adults with heart failure: a scientific statement from the American Heart Association. Circ Cardiovasc Qual Outcomes 2023 Nov;16(11):e000123. [doi: 10.1161/HCQ.0000000000000123] [Medline: 37909212]
53. Bakas T, Sampsel D, Israel J, et al. Using telehealth to optimize healthy independent living for older adults: a feasibility study. Geriatr Nurs (Lond) 2018 Sep;39(5):566-573. [doi: 10.1016/j.gerinurse.2018.04.002]
54. Gentry MT, Lapid MI, Rummans TA. Geriatric telepsychiatry: systematic review and policy considerations. Am J Geriatr Psychiatry 2019 Feb;27(2):109-127. [doi: 10.1016/j.jagp.2018.10.009] [Medline: 30416025]
55. Esfandiari E, Miller WC, Ashe MC. The effect of telehealth interventions on function and quality of life for older adults with pre-frailty or frailty: a systematic review and meta-analysis. J Appl Gerontol 2021 Nov;40(11):1649-1658. [doi: 10.1177/0733464820983630] [Medline: 33402043]
56. Kebede AS, Ozolins LL, Holst H, Galvin K. Digital engagement of older adults: scoping review. J Med Internet Res 2022 Dec 7;24(12):e40192. [doi: 10.2196/40192] [Medline: 36477006]
57. Kim HN, Freddolino PP, Greenhow C. Older adults' technology anxiety as a barrier to digital inclusion: a scoping review. Educ Gerontol 2023 Dec 2;49(12):1021-1038. [doi: 10.1080/03601277.2023.2202080]
58. Schroeder T, Dodds L, Georgiou A, Gewald H, Siette J. Older adults and new technology: mapping review of the factors associated with older adults' intention to adopt digital technologies. JMIR Aging 2023 May 16;6:e44564. [doi: 10.2196/44564] [Medline: 37191976]
59. Turcotte S, Bouchard C, Rousseau J, et al. Factors influencing older adults' participation in telehealth interventions for primary prevention and health promotion: a rapid review. Australas J Ageing 2024 Mar;43(1):11-30. [doi: 10.1111/ajag.13244] [Medline: 38014903]
60. Jaffe DH, Lee L, Huynh S, Haskell TP. Health inequalities in the use of telehealth in the United States in the lens of COVID-19. Popul Health Manag 2020 Oct;23(5):368-377. [doi: 10.1089/pop.2020.0186] [Medline: 32816644]
61. Choi NG, DiNitto DM, Marti CN, Choi BY. Telehealth use among older adults during COVID-19: associations with sociodemographic and health characteristics, technology device ownership, and technology learning. J Appl Gerontol 2022 Mar;41(3):600-609. [doi: 10.1177/07334648211047347] [Medline: 34608821]
62. Ryskina KL, Shultz K, Zhou Y, Lautenbach G, Brown RT. Older adults' access to primary care: gender, racial, and ethnic disparities in telemedicine. J Am Geriatr Soc 2021 Oct;69(10):2732-2740. [doi: 10.1111/jgs.17354] [Medline: 34224577]
63. Sheng X, Martirosyan Y, Hossain KS, Felix R, Singh A. Unpacking the influencing factors of telehealth usage among older consumers. Health Mark Q 2024;41(1):95-111. [doi: 10.1080/07359683.2023.2287927] [Medline: 38032235]
64. Haimi M, Sergienko R. Adoption and use of telemedicine and digital health services among older adults in light of the COVID-19 pandemic: repeated cross-sectional analysis. JMIR Aging 2024 Apr 24;7:e52317. [doi: 10.2196/52317] [Medline: 38656768]
65. Cameron KA, Song J, Manheim LM, Dunlop DD. Gender disparities in health and healthcare use among older adults. J Womens Health (Larchmt) 2010 Sep;19(9):1643-1650. [doi: 10.1089/jwh.2009.1701] [Medline: 20695815]
66. Song J, Chang RW, Manheim LM, Dunlop DD. Gender differences across race/ethnicity in use of health care among medicare-aged Americans. J Womens Health (Larchmt) 2006 Dec;15(10):1205-1213. [doi: 10.1089/jwh.2006.15.1205] [Medline: 17199461]
67. Che RP, Cheung MC. Factors associated with the utilization of home and community-based services (HCBS) among older adults: a systematic review of the last decade. J Gerontol Soc Work 2024;67(6):776-802. [doi: 10.1080/01634372.2024.2342455] [Medline: 38616618]
68. Weaver RH, Roberto KA. Home and community-based service use by vulnerable older adults. Gerontologist 2017 Jun 1;57(3):540-551. [doi: 10.1093/geront/gnv149] [Medline: 26608335]

XSL•FO
RenderX

69. Shachar C, Engel J, Elkin PL. Seizing the moment for telehealth policy and equity. Health Affairs. 2021 Sep 13. URL: https://www.healthaffairs.org/content/forefront/seizing-moment-telehealth-policy-and-equity [accessed 2025-05-14]
70. Eberly LA, Kallan MJ, Julien HM, et al. Patient characteristics associated with telemedicine access for primary and specialty ambulatory care during the COVID-19 pandemic. JAMA Netw Open 2020 Dec 1;3(12):e2031640. [doi: 10.1001/jamanetworkopen.2020.31640] [Medline: 33372974]
71. Alhussein M, Patel B, Liu X. Closing the gap: technology access and telehealth use among older adults in the U.S. medicare beneficiaries. Telemat Inform Rep 2023 Dec;12:100103. [doi: 10.1016/j.teler.2023.100103]
72. Yu RP, Ellison NB, McCammon RJ, Langa KM. Mapping the two levels of digital divide: internet access and social network site adoption among older adults in the USA. Inf Commun Soc 2016 Oct 2;19(10):1445-1464. [doi: 10.1080/1369118X.2015.1109695]
73. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: internet use patterns, eHealth literacy, and attitudes toward computer/internet use. J Med Internet Res 2013 May 2;15(5):e93. [doi: 10.2196/jmir.2645] [Medline: 23639979]
74. Price EAB, Hamiduzzaman M, McLennan V, Williams C, Flood V. Virtual care appointments and experience among older rural patients with chronic conditions in New South Wales: an analysis of existing survey data. Int J Environ Res Public Health 2024 Dec 17;21(12):1678. [doi: 10.3390/ijerph21121678] [Medline: 39767517]
75. Choi SL, Hites L, Bolland AC, Lee J, Payne-Foster P, Bissell K. Telehealth uptake among middle-aged and older Americans during COVID-19: chronic conditions, social media communication, and race/ethnicity. Aging Ment Health 2024;28(1):160-168. [doi: 10.1080/13607863.2022.2149696] [Medline: 36450359]
76. Moo LR, Schwartz AW. The urgent need for rigorous studies of telehealth for older adults who are homebound. JAMA Netw Open 2021 Sep 1;4(9):e2113451. [doi: 10.1001/jamanetworkopen.2021.13451] [Medline: 34499141]
77. Wang N, Seale M, Chen J. Availability and use of telehealth services among patients with ADRD enrolled in traditional medicare vs. medicare advantage during the COVID-19 pandemic. Front Public Health 2024;12:1346293. [doi: 10.3389/fpubh.2024.1346293] [Medline: 38476485]
78. Liang J, Aranda MP. The use of telehealth among people living with dementia-caregiver dyads during the COVID-19 pandemic: scoping review. J Med Internet Res 2023 May 25;25:e45045. [doi: 10.2196/45045]
79. Yi JS, Pittman CA, Price CL, Nieman CL, Oh ES. Telemedicine and dementia care: a systematic review of barriers and facilitators. J Am Med Dir Assoc 2021 Jul;22(7):1396-1402. [doi: 10.1016/j.jamda.2021.03.015] [Medline: 33887231]
80. Curtis ME, Clingan SE, Guo H, Zhu Y, Mooney LJ, Hser YI. Disparities in digital access among American rural and urban households and implications for telemedicine-based services. J Rural Health 2022 Jun;38(3):512-518. [doi: 10.1111/jrh.12614] [Medline: 34355427]
81. Quinton JK, Ong MK, Vangala S, et al. The association of broadband internet access and telemedicine utilization in rural Western Tennessee: an observational study. BMC Health Serv Res 2021 Aug 3;21(1):765. [doi: 10.1186/s12913-021-06746-0] [Medline: 34344377]
82. Hunter I, Lockhart C, Rao V, Tootell B, Wong S. Enabling rural telehealth for older adults in underserved rural communities: focus group study. JMIR Form Res 2022 Nov 4;6(11):e35864. [doi: 10.2196/35864] [Medline: 36331533]
83. Rodriguez JA, Clark CR, Bates DW. Digital health equity as a necessity in the 21st century cures act era. JAMA 2020 Jun 16;323(23):2381-2382. [doi: 10.1001/jama.2020.7858] [Medline: 32463421]
84. Whitacre B, Gallardo R. State broadband policy: impacts on availability. Telecomm Policy 2020 Oct;44(9):102025. [doi: 10.1016/j.telpol.2020.102025]
85. Hunsaker A, Hargittai E. A review of internet use among older adults. New Media Soc 2018 Oct;20(10):3937-3954. [doi: 10.1177/1461444818787348]
86. Drake C, Zhang Y, Chaiyachati KH, Polsky D. The limitations of poor broadband internet access for telemedicine use in rural America: an observational study. Ann Intern Med 2019 Sep 3;171(5):382-384. [doi: 10.7326/M19-0283] [Medline: 31108509]
87. Zahnd WE, Bell N, Larson AE. Geographic, racial/ethnic, and socioeconomic inequities in broadband access. J Rural Health 2022 Jun;38(3):519-526. [doi: 10.1111/jrh.12635] [Medline: 34792815]
88. Fealy S, McLaren S, Seaman CE, et al. Exploring the sociodemographic and health-related determinants of telehealth use among a cohort of older australians during the COVID-19 pandemic: repeated cross-sectional study. JMIR Aging 2024 Aug 23;7:e58594. [doi: 10.2196/58594] [Medline: 39178035]
89. Butzner M, Cuffee Y. Telehealth interventions and outcomes across rural communities in the United States: narrative review. J Med Internet Res 2021 Aug 26;23(8):e29575. [doi: 10.2196/29575] [Medline: 34435965]
90. Choi EY, Kanthawala S, Kim YS, Lee HY. Urban/rural digital divide exists in older adults: does it vary by racial/ethnic groups? J Appl Gerontol 2022 May;41(5):1348-1356. [doi: 10.1177/07334648211073605] [Medline: 35196918]
91. Lee HY, Kanthawala S, Choi EY, Kim YS. Rural and non-rural digital divide persists in older adults: internet access, usage, and attitudes toward technology. Gerontechnology 2021 Jan 1;20(2):1-9 [FREE Full text] [doi: 10.4017/gt.2021.20.2.32-472.12]

XSL•FO
RenderX

92. Falvey JR, Sun N, Miller MJ, Pravdo A, Mullins CD. Demystifying the digital divide: disparities in telerehabilitation readiness among older adults in the United States. Arch Phys Med Rehabil 2024 Jul;105(7):1255-1261. [doi: 10.1016/j.apmr.2024.03.009] [Medline: 38554795]
93. Ofei-Dodoo S, Medvene LJ, Nilsen KM, Smith RA, DiLollo A. Exploring the potential of computers to enrich home and community-based services clients' social networks. Educ Gerontol 2015 Mar 4;41(3):216-225. [doi: 10.1080/03601277.2014.950491]
94. Browdie R, Whitmire MP, Blancato R. Advocacy and the Older Americans Act: the environment is very different these days. Generations 2021;45(4):1-5 [FREE Full text]

Abbreviations:

AAAs: Area Agencies on Aging
ADRD: Alzheimer disease and related dementias
HCBS: home- and community-based services
***ICD-10*:** *International Statistical Classification of Diseases, 10th Revision*
MLTSS: managed long-term services and supports
NCI-AD: National Core Indicators–Aging and Disabilities
OAA: Older Americans Act
OR: odds ratio
PACE: Program of All-Inclusive Care for the Elderly
RUCA: rural-urban commuting area
SEIPS: Systems Engineering Initiative for Patient Safety

Edited by P Santana-Mancilla; submitted 14.May.2025; peer-reviewed by AG Reyes, P Sockolow; revised version received 17.Mar.2026; accepted 19.Mar.2026; published 02.Jun.2026.

Please cite as:
Urbanski DP, Parikh RR, Langworthy BW, Wolf JM, Fabius CD, Dill J, Jutkowitz E, Shippee TP
Telehealth Use Among Older Adults Receiving Home- and Community-Based Services: Cross-Sectional Analysis Using the National Core Indicators–Aging and Disabilities Survey
JMIR Hum Factors 2026;13:e77522
URL: https://humanfactors.jmir.org/2026/1/e77522
doi:10.2196/77522

© Dana P Urbanski, Romil R Parikh, Benjamin W Langworthy, Jack M Wolf, Chanee D Fabius, Janette Dill, Eric Jutkowitz, Tetyana P Shippee. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 2.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Comparing Images of Depression in Mass Media and AI-Generated Pictures: Mixed Methods Study

Nuria Saladie, PhD; Carolina Llorente, PhD; Renée Joosten, BSc; Gema Revuelta, MD, PhD

Science, Communication and Society Studies Centre, Department of Medicine and Life Sciences, Universitat Pompeu Fabra, Carrer Doctor Aiguader 88, Barcelona, Spain

Corresponding Author:
Gema Revuelta, MD, PhD
Science, Communication and Society Studies Centre, Department of Medicine and Life Sciences, Universitat Pompeu Fabra, Carrer Doctor Aiguader 88, Barcelona, Spain

Abstract

Background: Images play an important role in reducing stigma related to mental health, which often is distorted in the media. In recent years, generative artificial intelligence (AI) has been used to generate images related to mental health. However, first reports suggest that AI-generated images do not depict mental health conditions accurately. In-depth studies on the topic of mental health representations in AI-generated images are still missing.

Objective: The main objective of this study is to analyze and compare the visual representation of depression in mass media and in AI-generated images.

Methods: The methodologies used were discussion groups (15 participants) and a quasi-experimental online survey (792 interviewees), aimed at people with depression and young people.

Results: The results showed that both the images used in the media and those generated by AI reproduced stereotypes and stigmas about depression. However, participants considered AI-generated pictures to be more stereotypical, stigmatizing, and more likely to have a negative impact on people with depression. In contrast, media images were considered more appropriate, realistic, inclusive, and that better reflected the relationship between gender and depression. Statistically significant differences were observed between the control and test groups in both people with depression and young people (P<.001), indicating that when people were aware of what images were AI-generated, they tended to reject them to a greater extent.

Conclusions: Considering the current trend toward the widespread use of AI in mental health communication, it is crucial to promote closer collaborations between science journalists, AI developers, and mental health experts, including patients' associations, as well as a shift toward user-participatory AI design.

(JMIR Hum Factors 2026;13:e81230) doi:10.2196/81230

KEYWORDS

mental health; depression; artificial intelligence; mass media; mental health communication

Introduction

Background

The worldwide depression incidence increased during 2020 due to the COVID-19 pandemic, although before then, depression already featured as a leading burden of disease globally [1]. According to the World Health Organization, approximately 280 million people have depression worldwide, the equivalent of 5% of the adult population [2]. Young people are among the most affected groups [3-5], with depression rates having doubled from prepandemic estimates [6]. Spain was one of the countries in Europe most affected by COVID-19 infections, complications, and deaths, with young people being among the groups (together with frontline health care workers) with higher prevalence rates of mental health symptoms [7]. The prevalence of depression in Spain increased from 3.06% prepandemic to 12% postpandemic [8]. At an informational level, Google searches for mental health disorders have tripled for some pathologies compared with 2016, which highlights that society needs, and is actively searching for, information on mental health [9].

Mental health stigma and discrimination have been described as having worse consequences than the mental health conditions themselves do [10]. Higher levels of self-stigma in people with mental health issues result in lower treatment adherence [11], which is related to the fact that internalized stigma and treatment-related stigma are negatively associated with seeking help [12]. In Spain, the stigmatization of mental health conditions is prevalent in stereotypes, prejudices, and discrimination associated with people with mental health issues, as shown by a recent systematic review [13].

When studying mental health stigmatization, it is necessary to pay attention to the images that accompany the text, especially

XSL•FO
RenderX

since multiple studies conclude that negative media representations of mental health lead to mental health stigma [14-16]. In light of this, framing images positively or negatively can significantly affect the way audiences react to messages [17]. Positive visual information can reduce stigma related to mental health treatments; improve information, understanding, and effectiveness in identifying symptoms; promote empowerment; and foster better communication between patients, family members, and medical professionals [18,19].

Against this background, it is evident that the media play a key role: they can contribute not only to the creation and perpetuation of stereotypes and prejudices but also to the reduction of stigma surrounding mental health by portraying realistic, balanced, or even positive counterstereotypical examples [20]. In reality, however, mental health coverage by the media is insufficient and often resorts to stigmatizing archive images depicting hopelessness [9]. An analysis of a wide range of media revealed that descriptions of mental illness are often distorted by misinformation, exaggeration, and inaccuracies [14]. Similarly, the media, including news and social media, were found to display mental health issues in a negative light, although an increasing number of balancing stories appeared, mainly in magazines [15]. Images of depression on the social media platform Tumblr mostly depicted crying women, were often very dark-colored, and represented gloominess and sadness [21]. On Instagram, images of public self-disclosure of mental health issues express emotional distress, calls for help, and displays of vulnerability [22].

In this context, artificial intelligence (AI) could become a mental health ally. Studies have reported potential benefits of integrating AI in health communication, which include improving access to health information, supporting health behavior change, addressing diverse community needs [23], or boosting the reach and effectiveness of public health communication strategies [24]. Particularly for mental health, studies show that AI presents opportunities to enhance mental health care, for example, by serving as a complementary tool to bridge gaps in conventional mental health services, especially in areas with limited resources or high social stigma [25], developing algorithms to analyze images and texts shared on social media to diagnose or predict depression [26], or supporting the mental health of health professionals with AI chatbots [27]. However, AI use in mental health also presents some ethical challenges [25,28], as it can make systematic errors in its decision-making processes leading to unfair outcomes, also called bias [29]. First reports suggest that AI-generated images do not depict mental health conditions accurately [30,31] and that they frequently reflect cultural stereotypes and historical visual tropes, including gender biases and stigma [32]. Nonetheless, in-depth studies on the topic of mental health representations in AI-generated images are still missing.

From a theoretical perspective, the social impact of media and AI-generated images can be explained through complementary approaches from media effects research, social psychology, and technology acceptance studies. Framing theory posits that visual elements, such as color, lighting, posture, and scene context, shape how mental health is interpreted, guiding audience attributions of realism, stigma, vulnerability, or social support [17,33,34]. However, media effects are not uniform: reception studies emphasize that individuals actively interpret images through their personal experiences, beliefs, emotional states, and social identities, which may modulate responses across age, gender, and lived experience with depression [35,36]. From a psychosocial perspective, social representations and emotional anchoring processes contribute to the stabilization of shared visual stereotypes of depression in collective memory [37,38]. Finally, differential evaluations of media versus AI-generated images can be interpreted through the technology acceptance model (TAM), which proposes that perceived usefulness and perceived ease of use shape behavioral intentions toward technology [39] and that these perceptions may vary across demographic groups, such as gender and age, due to differences in prior technological experience [40], as well as attitudinal biases toward AI-generated products [41,42].

Study Objectives

The main objective of this study is to analyze and compare the visual representation of depression in mass media and in AI-generated images. To do so, we explore the perceptions of 2 strategic population groups: people with depression and young people. The specific research questions (RQs) are as follows:

- RQ1. How is the representation of depression in media and AI-generated images perceived?
- RQ2. How do the opinions of people with depression differ from the opinions of young people regarding the depiction of depression in media and AI-generated images?
- RQ3. How do people with depression recommend better illustrating depression, and how do these recommendations compare to those by young people?
- RQ4. How do people with depression, as well as young people, receive recommendations from mental health associations regarding how to illustrate depression?

Methods

Two types of methods were implemented: one qualitative (discussion groups) and one quantitative (survey). A mixed methods approach was used because of the greater depth and breadth of information that it can offer compared to using 1 type of method alone [43].

Discussion Groups

In total, 3 discussion groups were organized involving 15 people (13 women and 2 men): one was with people with depression and two with young people. The distribution of the discussion groups organized is available in Multimedia Appendix 1.

Each discussion group had a tailored script approved by the Advisory Board of the project (formed by a mixed group of professionals from the fields of clinical psychology, science journalism, and representatives from mental health associations). A slide show was prepared for discussion groups containing 64 images, of which 30 were found in the media, another 30 were AI-generated, and 4 were related to recommendations from mental health associations.

Media images were obtained from 3 different Spanish newspapers by providing their inbuilt search engine with the

XSL•FO
RenderX

keyword “depression” (in Spanish, “depresión”) on December 18, 2023. The 10 most recent images were selected for each newspaper, spanning from December 15, 2023, to January 4, 2019. The 3 newspapers selected were *La Razón*, *El País*, and *Eldiario.es*. These newspapers represent different ideologies, from more conservative to more left-leaning ideologies, and traditional and online-native media [44]. Out of the 30 images from the media, 22 included women, 7 included men, and 1 was an ungendered illustration. In 6 images, the figure of the person was silhouetted against the light. Eleven images were of celebrities. On 1 occasion, 2 people interacted with each other, while all other pictures contained either a single person or people not interacting. In 2 instances, images depicted a woman holding a baby. Only 1 image was in black and white, and the rest were in color. These images are available in Multimedia Appendix 2.

AI-generated images were obtained from 4 different tools in June 2023: Deep Dream Generator (Stable Diffusion 1.5), DeepAI (Stable Diffusion 1.5), OpenAI (DALL·E 2), and Dream Studio (Stable Diffusion 2.1). Images were created using the prompt “depression.” Standard parameters were applied to generate the images, with no particular filters or prompt settings. Out of the 30 AI-generated images, 10 depicted women, 11 depicted men, and 9 were unknown or ungendered. The most repeated type of image was that of a person sitting down, either with their hands to their head or face, or holding their knees with their arms, looking down (18 images). Eleven pictures were in black and white, and 19 were in color. In addition, 3 were illustrations, and 27 represented realistic pictures. These images are available in Multimedia Appendix 3.

Images recommended by mental health associations were obtained from the guides from the Confederación Salud Mental España, “Style guide on mental health for the media” [45], and from Cochrane, “Choosing images for sharing evidence: a guide” [46].

Discussion groups took place between the 11th and 18th of January 2024. Conversations were registered and later transcribed. The coding and qualitative analysis were carried out via the qualitative research support program ATLAS.ti (version 22).

Ethical Considerations

Before commencing work, the research project was evaluated by a panel of independent experts of the Spanish Foundation for Science and Technology (FECYT), who deemed the data management protocol as adequate. Following this evaluation, a bioethics committee and an ethical protocol were not considered necessary because no high-risk interventions were involved. Before conducting the discussion groups, informed consent was obtained from all participants, who provided it voluntarily and with the full knowledge of the facts.

All appropriate measures have been taken to comply with applicable regulations on the protection of personal data, including the General Data Protection Regulation and the Spanish Organic Law 3/2018 of December 5 on the Protection of Personal Data and Guarantee of Digital Rights [47]. At Universitat Pompeu Fabra, the Institutional Commission for the Ethical Review of Projects was responsible for the ethical assessment and monitoring of research involving human participants, human biological samples, or personal data (including health and clinical data). As this study did not involve any of these elements, ethical review and approval were not required according to institutional and national regulations.

Survey

Based on the qualitative analysis of the discussion groups, a quasi-experimental survey was developed. The insights contributed by discussion groups’ participants offered contextual depth and nuance regarding lived experiences and perspectives. Specific topics that appeared in the discussion groups and that were later transferred to the survey included, among others, how images could be more inclusive, better illustrate the relationship between gender and depression, or could negatively affect people with depression. Based on the most commonly voiced opinions regarding how the images could be improved, a list of options was prepared for the survey asking respondents to select the phrases that could help improve the way depression was illustrated.

The survey consisted of 14 questions and was validated twice before being shared: first, by the Advisory Board of the project; and second, by discussion group participants. The survey is available in Multimedia Appendix 4.

The survey had 2 versions: a test and a control (a mechanism specific to quasi-experimental surveys). In the test version, respondents knew which images were from the media and which ones were AI-generated; in the control version, respondents were not told the difference.

Target respondents were people with depression and young people. The first 7 questions of the survey asked respondents to select up to 5 images out of the 30 presented (15 from the media and 15 AI-generated). The remaining 7 questions were either multiple choice, scale, or open text questions.

The survey was sent to target respondents via an online opt-in panel. The sample was representative of Spanish society in terms of population distribution, encompassing all autonomous communities, age groups, genders, education levels, and population sizes. The data collection period was from the 22nd to 24th of May 2024. A total of 792 responses were obtained—390 from people with depression and 402 from young people. These numbers surpassed the minimum required to obtain a statistical robustness of 95% with an error margin of 5% (achieved at 385 responses for each target group).

Regarding the screening of participants, the first question of the survey asked them whether a doctor had diagnosed them with any mental health condition, and a list of 6 options was offered, including depression. Participants who ticked that option were labeled under “People with depression.” People from this group were not restricted by their age because their defining trait for the purposes of the study was the fact that they had been diagnosed with depression. On the other hand, people who had not been diagnosed with any mental health issues and were 18 to 29 years old were labeled under “Young people.”

XSL•FO
RenderX

Raw data were cleaned in Excel. For statistical analysis, all data were imported into the R software (R Core Team). The data were initially tested for normality, and the corresponding tests were run (Pearson chi-square test, Fisher exact test, logistic regressions, Mann-Whitney *U* test, and Kruskal-Wallis test, specified in text). For all analyses, a statistical threshold of α=.05 was used. Violin plots were used to better represent the full distribution of nonparametric data.

Results

The results are presented following the study's research questions.

How Is the Representation of Depression in Media and AI-Generated Images Perceived?

Both media and AI-generated images reproduce stereotypes and stigma about depression. However, respondents perceive AI-generated images more negatively than media images. In particular, AI-generated images are considered to be even more stereotypical, reproducing stigmas of marginalization or social exclusion and negatively affecting people with depression. In contrast, media images are considered relatively more appropriate, realistic, inclusive, and that represent the relationship between gender and depression better.

How Do the Opinions of People With Depression Differ From the Opinions of Young People Regarding the Depiction of Depression in Media and AI-Generated Images?

To answer this question, we present the results in 2 categories, depending on whether the questions were framed (1) positively or (2) negatively.

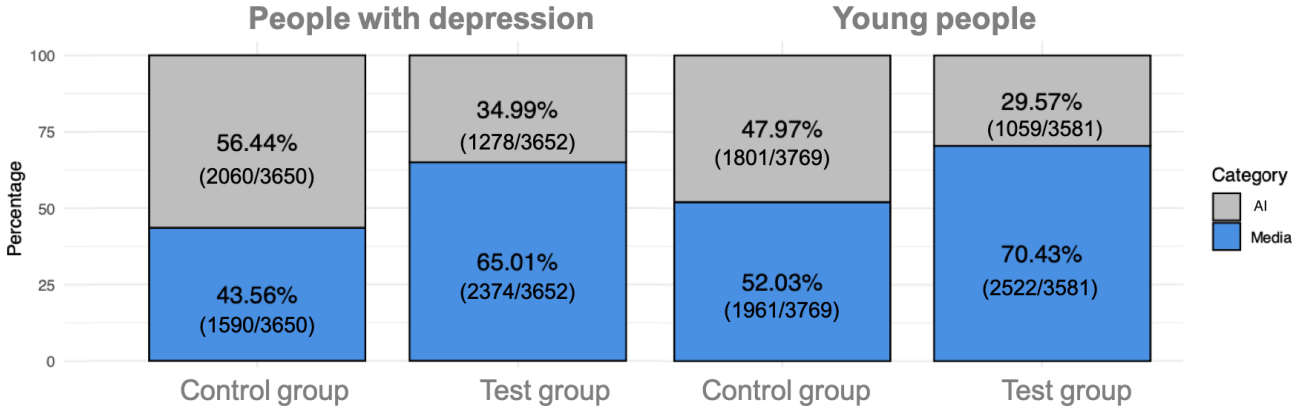
Responses to Positively Framed Questions

The questions that were framed positively asked which images were considered more appropriate, realistic, inclusive, and better represented the relationship between gender and depression.

A nonparametric Pearson chi-square analysis was carried out to examine differences between the test and control groups among people with depression. There were statistically significant differences between the 2 groups (P<.001), with the test group (ie, people who knew which images were from the media and which were AI-generated) choosing media images more frequently (2374/3652, 65.01% chose media; 1278/3652, 34.99% chose AI images). On the other hand, the control group (those who did not know whether images were from the media or AI) selected a higher percentage of AI-generated images (1590/3650, 43.56% chose media; 2060/3650, 56.44% chose AI images).

The same trend was observed in responses from young people (P<.001)—while the test group chose media images more frequently (2522/3581, 70.43% chose media; 1059/3581, 29.57% chose AI), the ratio in the control group was much more balanced (1961/3769, 52.03% chose media; 1801/3769, 47.97% chose AI; Figure 1).

Figure 1. Distribution of answers for positively framed questions regarding media or artificial intelligence (AI)–generated images, segregated by group among people with depression and young people.

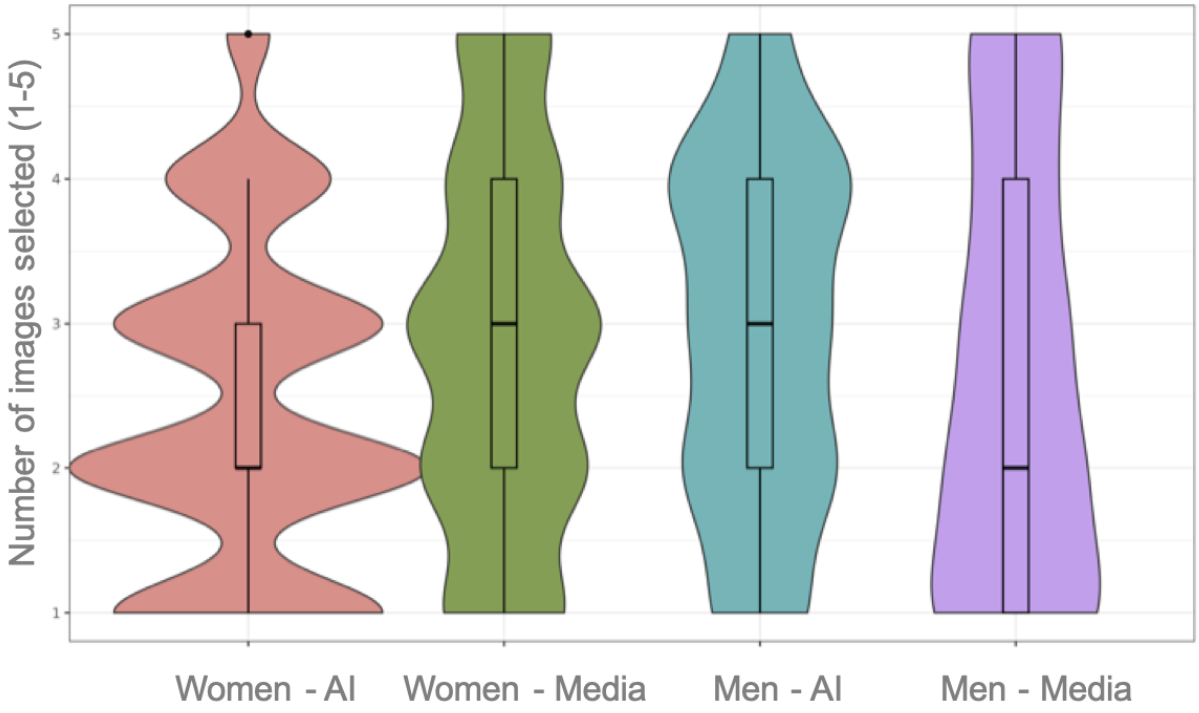


The top 3 images selected for positively framed questions among people with depression were all from the media for the test group, whereas the control group included 2 out of 3 that were AI-generated. Among young people, the same was observed: while the test group chose only media images in the top 3 positions, the control group had a more balanced selection. Among people with depression, the first images selected both by the test and control groups represented women sitting down, hiding their face in their arms. Among young people, both the test and control groups selected in the first place the same image, showing a woman holding a baby. Images are available as Multimedia Appendices 2 and 3.

Responses to positively framed questions from people with depression were crossed with their demographic data. In terms of sex, no significant differences were found regarding which images were more appropriate (χ^2_1=1.785; P=.18; ϕ=.02), more inclusive (χ^2_1=0.289; P=.59; ϕ=0), and better represented the relationship between gender and depression (χ^2_1=3.430; P=.06; ϕ=.04). However, when they were asked about which images represented depression more realistically, women chose media images more frequently (χ^2_1=7.877; P=.005; ϕ=.06) Figure 2).

XSL•FO
RenderX

Figure 2. Violin plot for the question "Which images represent depression more realistically?" The Y axis represents the amount of images selected, from 1 to 5. The median for each column is marked with a thicker black line. We can observe how, out of the 5 possible options to select, women had a median of 2 artificial intelligence (AI) images (widest curve of the first column), as opposed to a median of 3 media images (widest curve of the second column). Differences among men were not statistically significant.



Regarding age groups, no significant differences were found in their preference for media or AI-generated images when asked about what images were more realistic (χ^2_5=6.439; *P*=.27; V=0.03) and appropriate (χ^2_5=7.4996; *P*=.19; V=0.04) to represent depression. However, when asked about which images better represented the relationship between gender and depression, 25- to 29-year-olds and 40- to 50-year-olds showed significant differences in selecting media images more frequently (χ^2_1=13.211; *P*<.001; ϕ=.11 and χ^2_1=28.9; *P*<.001; ϕ=.08, respectively). Likewise, when asked about which images were more inclusive, significant differences were found among 35- to 39-year-olds, who chose media images more frequently (χ^2_1=13.111; *P*=.002; ϕ=.17).

The same tests were conducted for young people. In terms of sex, statistically significant differences were observed when asked which were more appropriate to represent depression, with women choosing media images more frequently (χ^2_1=15.085; *P*<.001; ϕ=.09). There were also significant differences when asked about which images were more realistic; in this case, both men and women preferred media images, but women chose media images at a much higher rate (χ^2_1=152.4932; *P*<.001; ϕ=.13) than men did (χ^2_1=4.5533; *P*=.03; ϕ=.04).

Responses to Negatively Framed Questions

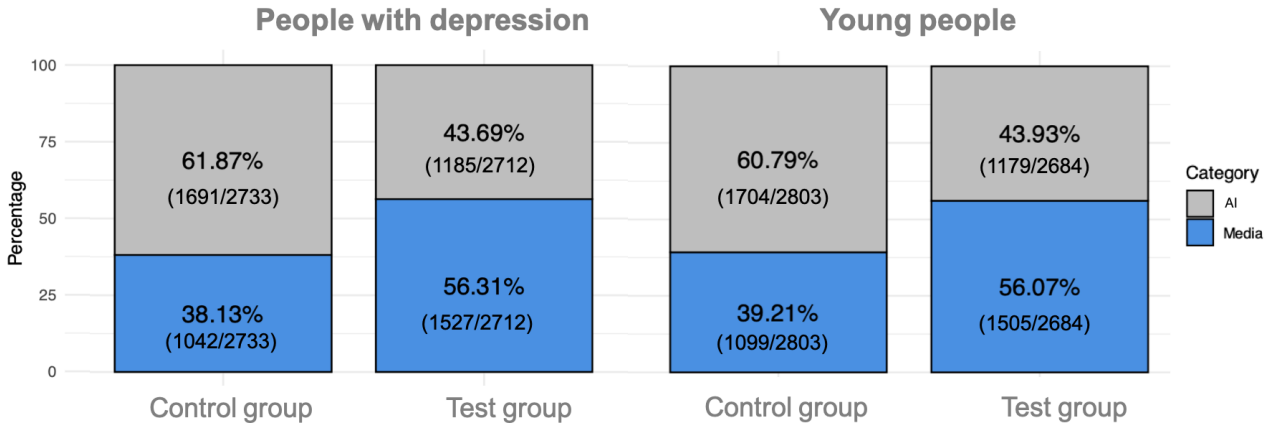
The questions that were framed negatively asked which images were considered more stereotyping, which images reproduced marginalization or social exclusion stigmas, and which images could affect people with depression more negatively.

A nonparametric Pearson chi-square test was used to examine differences between the test and control groups among people with depression. There were statistically significant differences between the 2 groups (*P*<.001): while the test group chose media images more frequently (1527/2712, 56.31% chose media; 1185/2712, 43.69% chose AI), the control group prioritized AI-generated images (1042/2733, 38.13% chose media; 1691/2733, 61.87% chose AI). Although these results follow the same pattern as those observed in positively framed questions, here the test group selected AI-generated images at a higher ratio.

A similar trend was observed among young people's responses. There were statistically significant differences between the test and control groups (*P*<.001)—while the test group selected media images more frequently (1505/2684, 56.07% chose media; 1179/2684, 43.93% chose AI), the control group preferred AI-generated images (1099/2803, 39.21% chose media; 1704/2803, 60.79% chose AI; Figure 3).

XSL•FO
RenderX

Figure 3. Distribution of answers for negatively framed questions regarding media or artificial intelligence (AI)–generated images, segregated by group among people with depression and young people.



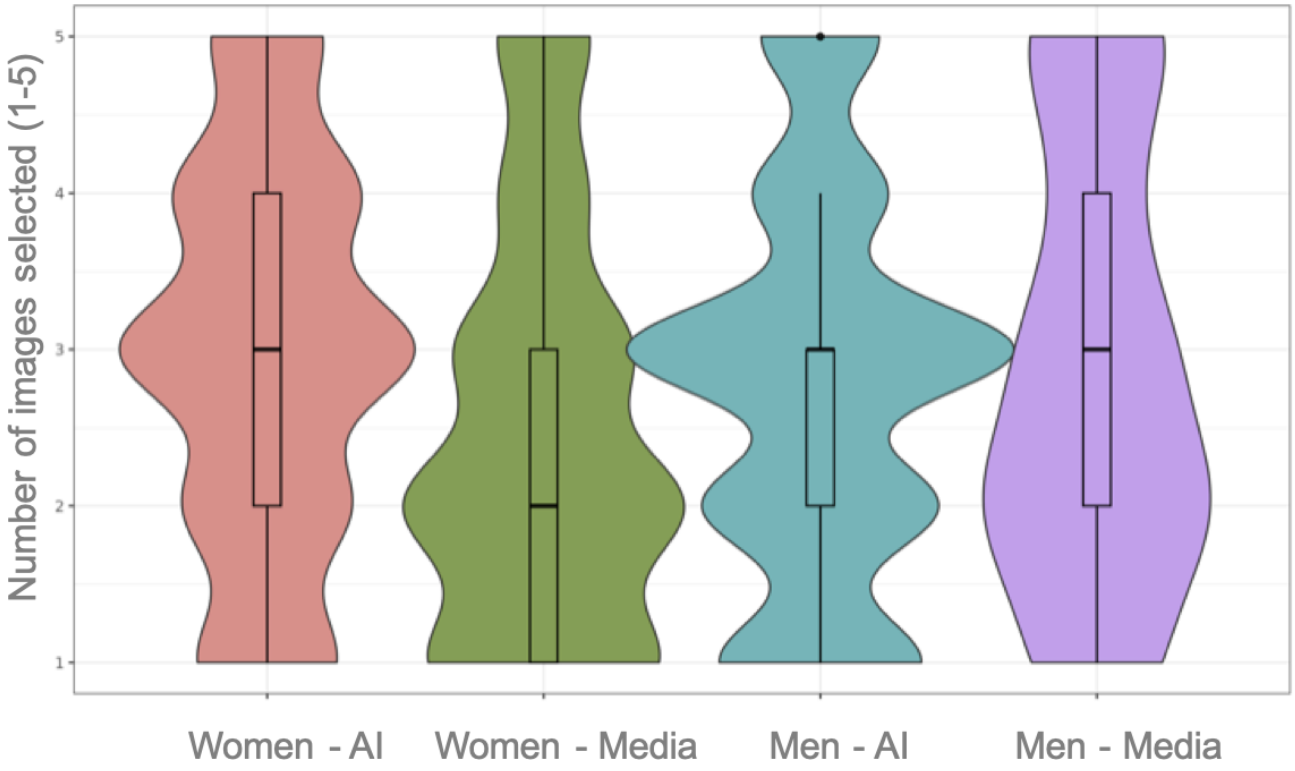
The top 3 images selected for negatively framed questions among people with depression were all from the media among the test group, whereas the control group selected only AI-generated images in the top 3 positions. Among young people, the top 3 images for negatively framed questions followed a similar pattern: all the top 3 images among the test group were from the media, while 2 out of the top 3 images of the control group were AI-generated. Among people with depression, the test group selected in the first place a media image of a woman sitting down, with her elbows to her knees, holding her head with her hands, looking down. The control group selected an AI-generated image of a young man sitting in the street, holding his knees with his arms, facing down. Among young people, the test group selected in the first place a media image of a back-lit silhouetted woman taking a pill. This control group chose the same image as the control group of people with depression. Images are available as Multimedia Appendices 2 and 3.

Responses to negatively framed questions from people with depression were crossed with their demographic data. No significant differences were found on the basis of sex regarding which images were more stereotyped (χ^2_1=0.02806; *P*=.87; ϕ=0), reproduced more stigmas (χ^2_1=0.59149; *P*=.44; ϕ=0) or were more likely to have a negative impact on people with depression (χ^2_1=0.51907; *P*=.47; ϕ=0). Likewise, no significant differences were found on the basis of age for those same questions (χ^2_5=2.8295, *P*=.73, V=0; χ^2_5=10.114, *P*=.07, V=0.05; and χ^2_5=8.6723, *P*=.12, V=0.05; respectively).

The same analysis was conducted for young people's responses. Significant differences were found between sexes (χ^2_1=9.6555; *P*=.002; ϕ=.07), with women choosing AI images more frequently when asked which images were more stereotyped (χ^2_1=18.2141; *P*<.001; ϕ=.04; Figure 4).

XSL•FO
RenderX

Figure 4. Violin plot for the question "Which images are more stereotyped?" The Y axis represents the amount of images selected, from 1 to 5. The median for each column is marked with a thicker black line. We can observe how, out of the 5 possible options to select, women had a median of 3 artificial intelligence (AI) images (widest curve of the first column), as opposed to a median of 2 media images (widest curve of the second column). Differences among men were not statistically significant.



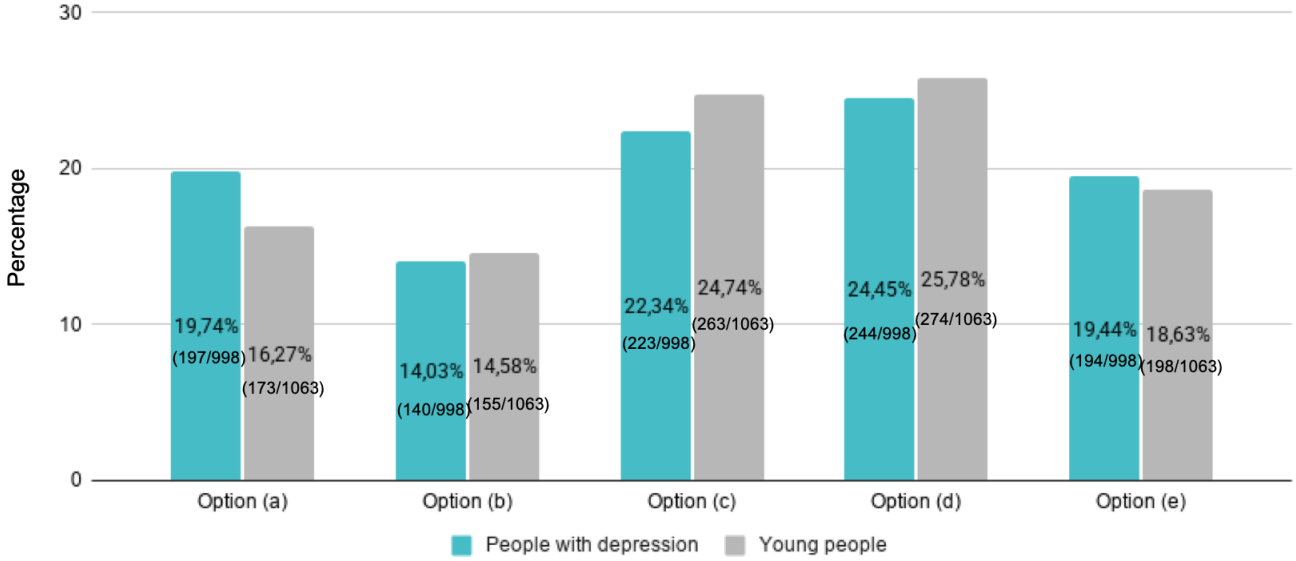
How Do People With Depression Recommend Better Illustrating Depression, and How Do These Recommendations Compare to Those by Young People?

Respondents were asked to select their preferred sentences to help improve the way depression was depicted and were given the following five options: (a) show people with depression who are going about their life, spending time with family, friends, at work, having fun, at an everyday activity, and forth; (b) show more diversity of people in the images (different skin color, gender, age, functional diversity, types of bodies, etc); (c) show that help is available and that there are options for therapy, support, and accompaniment; (d) consider that depression has many degrees and that a single image cannot illustrate all it represents; and (e) consult with mental health specialists when choosing images.

Both people with depression and young people selected option (4) in the first place, "consider that depression has many degrees and that a single image cannot illustrate all it represents" (people with depression: 244/998, 24.45%; young people: 274/1063, 25.78%). In second place, both groups also agreed with option (3), "show that help is available and that there are options for therapy, support, and accompaniment" (people with depression: 223/998, 22.34%; young people: 263/1063, 24.74%). In third place, people with depression chose option (1), "show people with depression who are going about their life, spending time with family, friends, at work, having fun, at an everyday activity, and so forth" (197/998, 19.74%), while young people chose option (5), "consult with mental health specialists when choosing images" (198/1063, 18.63%; Figure 5).

XSL•FO
RenderX

Figure 5. Double bar chart with respondents' opinions (options a to e) regarding recommendations. Options: (a) show people with depression who are going about their life, spending time with family, friends, at work, having fun, at an everyday activity, and so forth; (b) show more diversity of people in the images (different skin color, gender, age, functional diversity, types of bodies, etc); (c) show that help is available and that there are options for therapy, support, and accompaniment; (d) consider that depression has many degrees and that a single image cannot illustrate all it represents; and (e) consult with mental health specialists when choosing images.



Among people with depression, there were no statistically significant differences between the test and control groups ((a) odds ratio [OR]=0.96, 95% CI 0.62-1.47; *P*=.84; (b) OR=0.89, 95% CI 0.58-1.38; *P*=.62; (c) OR=1.27, 95% CI 0.82-1.98; *P*=.29; (d) OR=1.01, 95% CI 0.64-1.58; *P*=.98; (e) OR=0.69, 95% CI 0.45-1.05; *P*=.08), nor when segregated by sex ((a) OR=1.22, 95% CI 0.76-1.99; *P*=.41; (b) OR=1.40, 95% CI 0.86-2.30; *P*=.17; (c) OR=1.29, 95% CI 0.79-2.08; *P*=.31; (d) OR=1.15, 95% CI 0.69-1.89; *P*=.59; (e) OR=0.74, 95% CI 0.46-1.19; *P*=.22). However, differences were observed among age groups for option (a), with 40- to 50-year-olds (OR=2.44, 95% CI 1.21-4.90; *P*=.004) and 25- to 29-year-olds (OR=3.73, 95% CI 1.18-11.78; *P*=.01) being more likely to select this option than people older than 51 years. Among young people, there were no statistically significant differences between the test and control groups ((a) OR=0.96, 95% CI 0.62-1.47; *P*=.84; (b) OR=0.89, 95% CI 0.58-1.38; *P*=.62; (c) OR=1.27, 95% CI 0.82-1.98; *P*=.29; (d) OR=1.01, 95% CI 0.64-1.58; *P*=.98; (e) OR=0.69, 95% CI 0.45-1.05; *P*=.08). However, there were significant differences between sexes, with female respondents choosing more frequently than male respondents options (b) (OR=2.14, 95% CI 1.19-3.98; *P*=.01) and (c) (OR=1.79, 95% CI 1.06-3.02; *P*=.03).

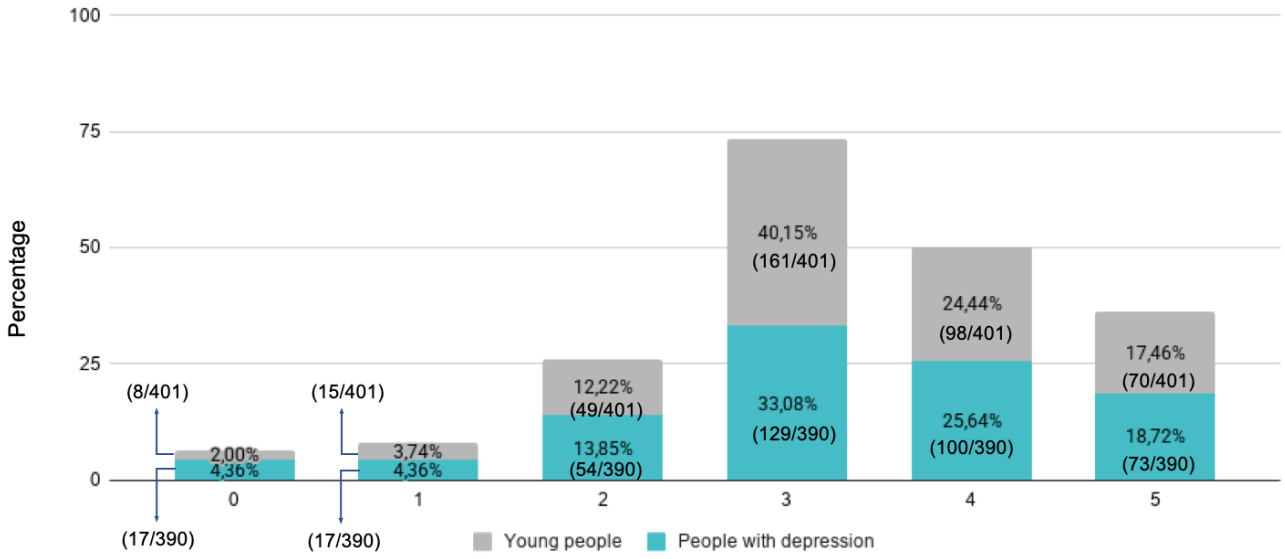
How Do People With Depression, as Well as Young People, Receive Recommendations From Mental Health Associations Regarding How to Illustrate Depression?

Respondents were offered a scale from 0 to 5 to indicate the appropriateness of the images recommended by mental health associations (0="images are very inadequate to represent depression" and 5="images are very adequate to represent depression").

Among people with depression, 44.36% (173/390) responded with ratings between 4 and 5. The most common single rating was 3 (129/390, 33.08%). The average (mean) was 3.27 (SD 1.28), and the median and mode were 3 (IQR 3-4). Young people responded in a very similar pattern: 41.90% (168/401) selected ratings between 4 and 5. The most common single rating was also 3 (161/401, 40.15%). The average (mean) was slightly higher than among people with depression, with 3.34 (SD 1.13), and the median and mode were also 3 (Figure 6).

XSL•FO
RenderX

Figure 6. Stacked bar chart with responses (scale 0 - 5) with respondents' opinions regarding the appropriateness of images recommended by mental health associations.



There were no significant differences among people with depression depending on group (median test group 4, IQR 4-5 vs median control group 4, IQR 4-5; W=19134; *P*=.91), sex (median males 4, IQR 3-5 vs median females 4, IQR 4-5; W=14224; *P*=.19) or region (χ^2_{15}=18.936; *P*=.22). Likewise, for young people, no differences were found depending on group (median test group 4, IQR 4-5 vs median control group 4, IQR 4-5; W=20726.5; *P*=.57), sex (median males 4, IQR 4-5 vs median females 4, IQR 4-5; W=11397; *P*=.08) or region (χ^2_{16}=7.731; *P*=.96). These results were validated with the Mann-Whitney U test and the Kruskal-Wallis test. Although all age groups consider these recommendations positively, people with depression between the ages of 35 to 39 years do so the most.

Discussion

Principal Findings

This study analyzed media and AI-generated images representing depression and complemented that with an exploration of the perceptions of people with depression and young people regarding those images. This study also analyzed how recommendations from mental health associations regarding how depression should be illustrated are perceived by people with depression and young people. This research offers valuable insights into the quality and appropriateness of media and AI-generated images, indicating overt preferences among respondents. The integration of a quasi-experimental survey, including test and control groups, ensures the robustness of the results.

The first research question we set ourselves to address was "How is the representation of depression in media and AI-generated images perceived?" The results indicate that the representation of depression is perceived more negatively in images generated by AI than in images from the media. That is, AI-generated images are perceived as more stereotyping, that they reproduce stigmas of marginalization or social exclusion, and that they affect people with depression more negatively. In contrast, compared with AI-generated images, media images are perceived to be more appropriate, realistic, inclusive, and to better represent the relationship between gender and depression.

However, it must be noted that both media and AI-generated images reproduce stereotypes and stigma about depression. Most media and AI images include characteristics that associations suggest avoiding, such as depicting people alone; without social interactions; staring into space; lying on the floor; covering their faces; or in long, dark hallways [45].

The second research question was "How do the opinions of people with depression differ from the opinions of young people regarding the depiction of depression in media and AI-generated images?" In short, their answers follow a similar pattern in most questions. Let us take a deeper look into each of the groups.

According to the survey results, people with depression tend to favor media images over AI-generated images. This trend is observed particularly strongly in questions that are framed positively (in this case, asking which images are considered more appropriate, realistic, inclusive, and better represent the relationship between gender and depression). People with depression from the test group (ie, those who know which images are from the media and which are AI-generated) clearly prioritize media images over AI-generated ones in positively framed questions. In other words, people from the control group (ie, those who do not know which images are from the media and which are AI-generated) select AI-generated images at a higher ratio than those in the test group. Compared with young people's responses, the same trend is observed: young people from the test group also prioritize media images, or avoid AI-generated images, in positively framed questions, while the difference is less pronounced in the control group. This supports

XSL•FO
RenderX

the idea that when people know which images are AI-generated, they tend to reject them, indicating a negative preconception about AI. This finding coincides with previous research observing a prejudice against AI-generated products [41,42]. However, another interpretation could be that test group participants could identify the more stereotypical images because they had been told in advance which ones were AI-generated.

There are also negatively framed questions, asking which images are considered more stereotyping, reproduce marginalization or social exclusion stigmas, or can affect people with depression more negatively. For these types of questions, people with depression from the test group choose media images more frequently than AI-generated ones, while people from the control group prioritize AI-generated images over media ones. This pattern is the same among young people, who also tend to select a higher ratio of AI-generated images when they cannot distinguish them from the media ones. Interestingly, the differences between the test and control groups do not seem to be influenced by demographic variables, such as sex, age, or region.

It is relevant to note how the framing of a question affects its responses, coinciding with previous literature indicating that audiences' reactions depend on the framing of images [17]. In this study, the results obtained from positively and negatively framed questions followed the same pattern: the test groups always prioritized media images, while the control group favored AI-generated images more frequently. However, in negatively framed questions, control groups select AI-generated images at a higher ratio than test groups do. This could indicate that when people know which images are generated by AI, they tend to associate them more strongly with negative concepts, such as stereotypes or stigma. This interpretation aligns with previous research, as mentioned before, observing a prejudice against AI-generated products [41,42].

The third research question of the study was as follows: "How do people with depression recommend better illustrating depression, and how do these recommendations compare to those by young people?" This question is particularly relevant if we consider that nowadays, AI-generated images do not depict mental health conditions accurately [30,31]; thus, the need for recommendations becomes clear. The survey results indicate that people with depression and young people share similar opinions regarding the proposed recommendations, as they both select the same options in first and second place ("consider that depression has many degrees and that a single image cannot illustrate all it represents," followed by "show that help is available and that there are options for therapy, support, and accompaniment"). For these recommendations, no statistically significant differences are observed among people with depression, either by test and control groups or when segregated by demographic variables.

Among young people, no statistically significant differences are observed between test and control groups; however, when segregated by sex, women choose the second recommendation ("show that help is available and that there are options for therapy, support, and accompaniment") more frequently than men do. This result is supported by the literature showing that men are less inclined than women to seek help for mental health issues [48,49].

The fourth and final research question to address was as follows: "How do people with depression, as well as young people, receive recommendations from mental health associations regarding how to illustrate depression?" Both respondent groups perceive these recommendations positively: from a scale of 0‐5 asking about the appropriateness of the images recommended, the average among people with depression is 3.27 (SD 1.28), which is slightly lower than the 3.34 (SD 1.13) among young people. These results indicate that both people with depression and young people approve of the recommendations provided by mental health associations. This positive result can be interpreted as validation of the recommendations.

The observed gender- and age-related differences in attitudes toward technology use can be interpreted through established theoretical frameworks, such as the TAM. The TAM proposes that perceived usefulness and perceived ease of use shape behavioral intentions, and these perceptions may vary across demographic groups due to differences in prior technological experience [39,40]. This model provides a theoretical basis for understanding why demographic characteristics can lead to systematic differences in technology acceptance, as observed in this study.

In conclusion, there is currently a clear trend toward the widespread use of AI in journalism [50], as well as in science communication [51]. It is not difficult to imagine a future in which AI-generated images will start to appear in the media, in the same way as stock images do today. In this scenario, it is of utmost importance to consider social perceptions and opinions, particularly those of the groups most affected by the use of the new technology. Moreover, and particularly because negative media representations of mental health lead to mental health stigma [14-16], the findings of our study underscore the need for closer collaboration among journalists, AI developers, mental health experts, and patient associations, as well as a shift toward user-participatory AI design. With this study, we aimed to contribute to these considerations by highlighting social perceptions regarding the use of AI tools in the media and providing enough data to guide a more responsible way forward for the benefit of all.

Limitations

Despite the insights offered by this study, some limitations warrant consideration. First, we base our results on a selection of 30 images. Although it is clear that a limit to the number of images is needed to conduct the study, this same limit means that opinions and perceptions are inextricably linked to the images selected. It is possible that a different pool of images could produce dissimilar outcomes.

Second, the authors note that both AI technology and the ethical protocols that guide it evolve rapidly [52], as do people's sentiments about it [53]. This could mean that, in a few years, AI-generated images could reproduce fewer stereotypes, or that people's perceptions of AI-generated images have shifted toward a more positive outlook. With this in mind, this study would

XSL•FO
RenderX

benefit from being repeated over time to monitor whether its results are sustained over time.

Third, the qualitative sample of the discussion groups is not gender-balanced (13 women, 2 men). Since the qualitative phase informed the quantitative one, the authors acknowledge that this gender imbalance may have biased the development of the survey questions.

Fourth, the authors note that the predominantly Spanish cultural context of the study may limit the transferability of results to other cultural environments, where different social norms and practices may be at play.

Fifth, there is a potential confound with study populations. The group "Young people" did not include individuals with depression; however, the group "People with depression" was not restricted by age and therefore also included young people. This constitutes a potential confound of the study populations.

Finally, a methodological limitation of the study is the use of a simplified prompt ("depression") to generate the AI images. This single-word prompt could have increased the probability that generative models defaulted to widely learned, stereotypical visual representations of depression present in their training data. Therefore, the observed tendency for AI-generated images to appear "more stereotypical" could actually be a reflection of the prompting strategy rather than an inherent shortcoming of the generative systems themselves.

Conclusions

Academic literature had reported the impact of AI on mental health communication, but the effectiveness, limitations, and ethical issues of using AI in the communication of mental health, particularly depression, were yet to be studied. Combining qualitative and quantitative approaches, our study offers novel insights into how people with depression and young people perceive the use of AI to generate images of depression. The results show that although both media and AI-generated images are considered to reproduce stereotypes about depression, respondents tend to reject AI-generated images, indicating a certain bias against AI-produced pictures. These results hold substantial implications in the current drift toward the widespread use of AI in mental health communication, highlighting the need to build bridges among science journalists, AI developers, mental health experts, and patient associations.

Acknowledgments

We would like to acknowledge Gerard Romero for his technical assistance with statistical analysis.

As disclosed in the Methods sections, the images generated by artificial intelligence used for this study were obtained from 4 different artificial intelligence tools: Deep Dream Generator, DeepAI, OpenAI, and Dream Studio.

Funding

This research was supported by the Spanish Foundation for Science and Technology (FECYT)

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: CL (supporting), GR (lead)
Data curation: NS
Formal analysis: CL (supporting), GR (supporting), NS (lead)
Funding acquisition: CL (lead), GR (equal)
Investigation: CL (supporting), NS (lead)
Methodology: CL (supporting), GR (lead), NS (supporting), RJ (supporting)
Project administration: NS
Resources: GR
Supervision: GR
Validation: CL (equal), GR (lead), NS (supporting)
Visualization: GR (supporting), NS (lead)
Writing – original draft: NS (lead), RJ (supporting)
Writing – review & editing: CL (supporting), NS (lead), GR (equal)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Distribution of the discussion groups organized, specifying their date, time, profile of participants, mode of attendance, and number and gender of participants.

XSL•FO
RenderX

[DOCX File, 13 KB - humanfactors_v13i1e81230_app1.docx]

Multimedia Appendix 2
Media images.
[DOCX File, 16 KB - humanfactors_v13i1e81230_app2.docx]

Multimedia Appendix 3
Artificial intelligence (AI)–generated images.
[DOCX File, 24118 KB - humanfactors_v13i1e81230_app3.docx]

Multimedia Appendix 4
Surveys.
[DOCX File, 23 KB - humanfactors_v13i1e81230_app4.docx]

References

1. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021 Nov 6;398(10312):1700-1712. [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
2. Depression. World Health Organization. 2025. URL: https://www.who.int/news-room/fact-sheets/detail/depression [accessed 2025-07-18]
3. Daniali H, Martinussen M, Flaten MA. A global meta-analysis of depression, anxiety, and stress before and during COVID-19. Health Psychol 2023 Feb;42(2):124-138. [doi: 10.1037/hea0001259] [Medline: 36802363]
4. León CN, Olivera ES. Revisión sistemática sobre la salud mental en el contexto COVID-19 [Article in Spanish]. Rev Invest Univ Le Cordon Bleu 2022;9(1):61-78. [doi: 10.36955/RIULCB.2022v9n1.006]
5. Son C, Hegde S, Smith A, Wang X, Sasangohar F. Effects of COVID-19 on college students' mental health in the United States: interview survey study. J Med Internet Res 2020 Sep 3;22(9):e21279. [doi: 10.2196/21279] [Medline: 32805704]
6. Racine N, McArthur BA, Cooke JE, Eirich R, Zhu J, Madigan S. Global prevalence of depressive and anxiety symptoms in children and adolescents during COVID-19: a meta-analysis. JAMA Pediatr 2021 Nov 1;175(11):1142-1150. [doi: 10.1001/jamapediatrics.2021.2482] [Medline: 34369987]
7. Zhang SX, Chen RZ, Xu W, et al. A systematic review and meta-analysis of symptoms of anxiety, depression, and insomnia in Spain in the COVID-19 crisis. Int J Environ Res Public Health 2022 Jan 17;19(2):1018. [doi: 10.3390/ijerph19021018] [Medline: 35055841]
8. Ayuso-Mateos JL, Morillo D, Haro JM, Olaya B, Lara E, Miret M. Changes on depression and suicidal ideation under severe lockdown restrictions during the first wave of the COVID-19 pandemic in Spain: a longitudinal study in the general population. Epidemiol Psychiatr Sci 2023 Sep 1;32:e55. [doi: 10.1017/S2045796023000677] [Medline: 37655398]
9. Revuelta G, Llorente C, Gonzalo C, Moena O. Informe quiral 2022-2023 la gravedad de la salud mental: comunicación y percepción [Report in Spanish]. : Fundació Vila Casas; 2023. [doi: 10.5281/zenodo.8177359]
10. Thornicroft G, Mehta N, Clement S, et al. Evidence for effective interventions to reduce mental-health-related stigma and discrimination. Lancet 2016 Mar 12;387(10023):1123-1132. [doi: 10.1016/S0140-6736(15)00298-6] [Medline: 26410341]
11. Carrara BS, Ventura CAA. Self-stigma, mentally ill persons and health services: an integrative review of literature. Arch Psychiatr Nurs 2018 Apr;32(2):317-324. [doi: 10.1016/j.apnu.2017.11.001] [Medline: 29579531]
12. Clement S, Schauman O, Graham T, et al. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. Psychol Med 2015 Jan;45(1):11-27. [doi: 10.1017/S0033291714000129] [Medline: 24569086]
13. Zamorano S, Sáez-Alonso M, González-Sanguino C, Muñoz M. Social stigma towards mental health problems in Spain: a systematic review. Clínica y Salud 2023 Mar 13;34(1):23-34. [doi: 10.5093/clysa2023a5]
14. Klin A, Lemish D. Mental disorders stigma in the media: review of studies on production, content, and influences. J Health Commun 2008;13(5):434-449. [doi: 10.1080/10810730802198813] [Medline: 18661386]
15. Ma Z. How the media cover mental illnesses: a review. Health Educ 2017 Jan 3;117(1):90-109. [doi: 10.1108/HE-01-2016-0004]
16. Stuart H. Media portrayal of mental illness and its treatments: what effect does it have on people with mental illness? CNS Drugs 2006;20(2):99-106. [doi: 10.2165/00023210-200620020-00002] [Medline: 16478286]
17. Sontag JM. Visual framing effects on emotion and mental health message effectiveness. J Commun Healthc 2018 Jan 2;11(1):30-47. [doi: 10.1080/17538068.2018.1435017]
18. Broussard B, Radkins JB, Compton MT. Developing visually based, low-literacy health education tools for African Americans with psychotic disorders and their families. Community Ment Health J 2014 Aug;50(6):629-636. [doi: 10.1007/s10597-013-9666-7] [Medline: 24633539]

XSL•FO
RenderX

19. Unger JB, Cabassa LJ, Molina GB, Contreras S, Baron M. Evaluation of a fotonovela to increase depression knowledge and reduce stigma among Hispanic adults. J Immigr Minor Health 2013 Apr;15(2):398-406. [doi: 10.1007/s10903-012-9623-5] [Medline: 22485012]
20. Ramasubramanian S, Yadlin‐Segal A. Stigma: media influence on. In: Rössler P, Hoffner CA, Zoonen L, editors. The International Encyclopedia of Media Effects, 1st edition: Wiley; 2017:1-10. [doi: 10.1002/9781118783764]
21. Hussain SA. Is this what depression looks like? Visual narratives of depression on social media. Visual Stud 2020 May 26;35(2-3):245-259. [doi: 10.1080/1472586X.2020.1779611]
22. Manikonda L, De Choudhury M. Modeling and understanding visual attributes of mental health disclosures in social media. 2017 May 2 Presented at: CHI '17; May 6-11, 2017. [doi: 10.1145/3025453.3025932]
23. Sezgin E, Kocaballi AB. Era of generalist conversational artificial intelligence to support public health communications. J Med Internet Res 2025 Jan 20;27:e69007. [doi: 10.2196/69007] [Medline: 39832358]
24. Faus M, Alonso F, Fernández C, et al. Use of big data, artificial intelligence and other emerging technologies in public health communication campaigns: a systematic review. Rev Commun Res 2025;13:31-48. [doi: 10.52152/RCR.V13.3]
25. Wang X, Zhou Y, Zhou G. The application and ethical implication of generative AI in mental health: systematic review. JMIR Ment Health 2025 Jun 27;12:e70610. [doi: 10.2196/70610] [Medline: 40577783]
26. Babu NV, Kanaga EGM. Sentiment analysis in social media data for depression detection using artificial intelligence: a review. SN Comput Sci 2022;3(1):74. [doi: 10.1007/s42979-021-00958-1] [Medline: 34816124]
27. Baek G, Cha C, Han JH. AI chatbots for psychological health for health professionals: scoping review. JMIR Hum Factors 2025 Mar 19;12:e67682. [doi: 10.2196/67682] [Medline: 40106346]
28. Rahsepar Meadi M, Sillekens T, Metselaar S, van Balkom A, Bernstein J, Batelaan N. Exploring the ethical challenges of conversational AI in mental health care: scoping review. JMIR Ment Health 2025 Feb 21;12:e60432. [doi: 10.2196/60432] [Medline: 39983102]
29. Ferrara E. Fairness and bias in artificial intelligence: a brief survey of sources, impacts, and mitigation strategies. Sci 2023;6(1):3. [doi: 10.3390/sci6010003]
30. King M. Harmful biases in artificial intelligence. Lancet Psychiatry 2022 Nov;9(11):e48. [doi: 10.1016/S2215-0366(22)00312-1] [Medline: 36244361]
31. Putland E, Chikodzore-Paterson C, Brookes G. Artificial intelligence and visual discourse: a multimodal critical discourse analysis of AI-generated images of "Dementia". Soc Semiotics 2025 Mar 15;35(2):228-253. [doi: 10.1080/10350330.2023.2290555]
32. Flathers M, Smith G, Wagner E, Fisher CE, Torous J. AI depictions of psychiatric diagnoses: a preliminary study of generative image outputs in Midjourney V.6 and DALL-E 3. BMJ Ment Health 2024 Dec 4;27(1):e301298. [doi: 10.1136/bmjment-2024-301298] [Medline: 39632121]
33. McCombs ME, Shaw DL. The evolution of agenda-setting research: twenty-five years in the marketplace of ideas. J Commun 1993 Jun;43(2):58-67. [doi: 10.1111/j.1460-2466.1993.tb01262.x]
34. Entman RM. Framing: toward clarification of a fractured paradigm. J Commun 1993 Dec;43(4):51-58. [doi: 10.1111/j.1460-2466.1993.tb01304.x]
35. Bauer MW. Controversial medical and agri-food biotechnology: a cultivation analysis. Public Underst Sci 2002 Apr;11(2):93-111. [doi: 10.1088/0963-6625/11/2/301] [Medline: 14621673]
36. Niederdeppe J, Fowler EF, Goldstein K, Pribble J. Does local television news coverage cultivate fatalistic beliefs about cancer prevention? J Commun 2010 Jun 1;60(2):230-253. [doi: 10.1111/j.1460-2466.2009.01474.x] [Medline: 20563221]
37. Marková I. The epistemological significance of the theory of social representations. J Theory Soc Behav 2008 Dec;38(4):461-487. [doi: 10.1111/j.1468-5914.2008.00382.x]
38. Lazarus RS. Thoughts on the relations between emotion and cognition. Am Psychol 1982;37(9):1019-1024. [doi: 10.1037/0003-066X.37.9.1019]
39. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
40. Venkatesh V, Morris MG. Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. MIS Q 2000 Mar;24(1):115-139. [doi: 10.2307/3250981]
41. Chamberlain R, Mullin C, Scheerlinck B, Wagemans J. Putting the art in artificial: aesthetic responses to computer-generated art. Psychol Aesthet Creat Arts 2017;12(2):177-192. [doi: 10.1037/aca0000136]
42. Tubadji A, Huang H, Webber DJ. Cultural proximity bias in AI-acceptability: the importance of being human. Technol Forecast Soc Change 2021 Dec;173:121100. [doi: 10.1016/j.techfore.2021.121100]
43. Almalki S. Integrating quantitative and qualitative data in mixed methods research—challenges and benefits. J Educ Learn 2016;5(3):288. [doi: 10.5539/jel.v5n3p288]
44. López G, Valera Ordaz L. La información sobre la Monarquía española en los nuevos medios digitales: Eldiario.es y Vozpopuli.com [Article in Spanish]. adComunica 2016;6:65-81. [doi: 10.6035/2174-0992.2013.6.5]
45. Fábregas M, Tafur A, Guillén A, Bolaños L, Méndez JL, Sevilla P. Guía de estilo sobre salud mental para medios de comunicación: las palabras sí importan [Report in Spanish]. : Confederación SALUD MENTAL ESPAÑA; 2018 URL: https://consaludmental.org/publicaciones/Guia-estilo-salud-mental.pdf [accessed 2025-07-18]

XSL•FO
RenderX

46. Choosing images for sharing evidence: a guide. : Cochrane Knowledge Translation and Cochrane UK; 2020 URL: https://www.cochrane.org/sites/default/files/uploads/Images%20Checklist%20for%20Evidence%20Dissemination%20-%20Final%2C%20version%201.0_1.pdf [accessed 2025-07-18]
47. Spain Organic Law 3/2018 on Data Protection and Digital Rights (LOPDGDD). URL: https://theartofservice.com/frameworks/spain-organic-law-3-2018-on-data-protection-and-digital-rights-lopdgdd [accessed 2026-04-14]
48. Liddon L, Kingerlee R, Barry JA. Gender differences in preferences for psychological treatment, coping strategies, and triggers to help-seeking. Br J Clin Psychol 2018 Mar;57(1):42-58. [doi: 10.1111/bjc.12147] [Medline: 28691375]
49. Seidler ZE, Dawes AJ, Rice SM, Oliffe JL, Dhillon HM. The role of masculinity in men's help-seeking for depression: a systematic review. Clin Psychol Rev 2016 Nov;49:106-118. [doi: 10.1016/j.cpr.2016.09.002] [Medline: 27664823]
50. Stray J. Making artificial intelligence work for investigative journalism. Digit Journal 2019;7(8):1076-1097. [doi: 10.1080/21670811.2019.1630289]
51. Schäfer MS. The Notorious GPT: science communication in the age of artificial intelligence. J Sci Commun 2023;22(2). [doi: 10.22323/2.22020402]
52. Shukla S. Principles governing ethical development and deployment of AI. Int J Eng Bus Manag 2024;8(2):26-46. [doi: 10.22161/ijebm.8.2.5]
53. Lima G, Kim C, Ryu S, Jeon C, Cha M. Collecting the public perception of AI and robot rights. Proc ACM Hum-Comput Interact 2020 Oct 14;4(CSCW2):1-24. [doi: 10.1145/3415206]

Abbreviations

AI: artificial intelligence
OR: odds ratio
RQ: research question
TAM: technology acceptance model

Edited by A Choudhury; submitted 24.Jul.2025; peer-reviewed by A Amanna, Y Guo; revised version received 10.Dec.2025; accepted 07.Jan.2026; published 14.Apr.2026.

Please cite as:
Saladie N, Llorente C, Joosten R, Revuelta G
Comparing Images of Depression in Mass Media and AI-Generated Pictures: Mixed Methods Study
JMIR Hum Factors 2026;13:e81230
URL: https://humanfactors.jmir.org/2026/1/e81230
doi:10.2196/81230

© Nuria Saladie, Carolina Llorente, Renée Joosten, Gema Revuelta. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 14.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Feasibility and Acceptability of a Prevention-Focused Screener for Perinatal Depression Risk: Mixed Methods Cohort Study

Tamar Krishnamurti[1,2], PhD; Samantha Rodriguez[1], MS; Leah Cope[3], MSW; Lara Lemon[2], PhD, PharmD; Priya Gopalan[4], MD; Cara Nikolajski[3], MPH, PhD; Hyagriv Simhan[2], MS, MD; Kelly Williams[3], MPH, PhD

[1]General Internal Medicine, University of Pittsburgh School of Medicine, 200 Meyran Avenue, Parkvale Building, Suite 300, Pittsburgh, PA, United States

[2]Obstetrics, Gynecology, and Reproductive Sciences, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

[3]UPMC Center for High Value Healthcare, Pittsburgh, PA, United States

[4]Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

Corresponding Author:
Tamar Krishnamurti, PhD
General Internal Medicine, University of Pittsburgh School of Medicine, 200 Meyran Avenue, Parkvale Building, Suite 300, Pittsburgh, PA, United States

Abstract

Background: More than 20% of perinatal women experience depression, with suicide being a leading cause of maternal death in the United States. Professional societies emphasize the need to identify those at risk of developing perinatal depression to better target preventive care delivery during pregnancy.

Objective: We evaluated receptivity to a machine learning–based predictive screener designed to identify women in the first trimester of pregnancy who were asymptomatic but were at risk for developing moderate to severe depression symptoms later in pregnancy.

Methods: Our participants were adult pregnant women with negative first-trimester depression (Patient Health Questionnaire–9) screens at 1 of 4 obstetric practices. Of the 810 women who were clinically eligible, 787 were successfully contacted via their patient portal. Of these, 289 (36.7%) viewed the screener and 255 (88.2%) completed the 6-question predictive screener. In total, 51 (20%) were identified by the screener as being at risk for developing perinatal depression. Participants were asked a series of follow-up questions regarding the acceptability of the predictive screener and desired preventive resources. Chi-square tests were used to compare demographic characteristics, perceived benefits and concerns, and desired resources between those identified as at risk for depression and those who were not. Differences in acceptability ratings between the two risk groups were determined using nonparametric Mann-Whitney *U* tests.

Results: On a 5-point Likert scale of agreement, participants found the screener questions easy to complete (median score 5, IQR 5-5) and felt comfortable sharing their answers with their obstetric care providers (median 5, IQR 4-5). Key perceived benefits of completing the screener included opportunities to seek preventive care (75/255, 29.4%) and to receive education on depression risk (66/255, 25.9%). Primary concerns about knowing one's risk of future depression included worrying about developing depression (90/255, 35.3%) and a lack of prevention opportunities (39/255, 15.3%). Desired preventive resources included counseling (197/255, 77.3%), mind-body interventions (166/255, 65.1%) such as exercise, and prenatal classes or support groups (81/255, 31.8%).

Conclusions: Participants found the screener acceptable and felt comfortable receiving it through their patient portal. Specific preventive care options were commonly endorsed, several of which are scalable and evidence based. A minority of participants voiced addressable concerns about knowing their risk of developing depression in the future.

(JMIR Hum Factors 2026;13:e81638) doi:10.2196/81638

KEYWORDS

prediction algorithm; machine learning; preventive care; predictive screening; patient-centered care; depression; pregnancy; postpartum; mental health

Introduction

More than 20% of US mothers experience depression during pregnancy and the postpartum period, which is consistent with global rates [1,2]. Untreated depression can contribute to long-term health problems [3-5], including cognitive and emotional challenges for children and unhealthy maternal behaviors, such as alcohol or drug use, insufficient physical

XSL•FO
RenderX

activity, and poor diet [6,7]. Furthermore, 1 in 5 women with perinatal depression expresses thoughts of self-harm, with suicide being a leading cause of maternal death [8]. While 40% to 50% of women with perinatal depression have a history of depression, the remainder experience depression onset for the first time during their pregnancy or in the early postpartum period [9,10]. Currently, the United States Preventive Services Task Force recommends early identification of perinatal women at risk of developing depression to facilitate preventive rather than reactive care [11-13]. However, implementing this recommendation requires accurate identification of risk prior to depression onset. This identification is particularly challenging for those without a history of depression, who may not recognize and self-report the onset of early symptoms.

There are documented factors associated with perinatal depression, including pregnancy-related stressors (eg, financial instability, fear of the birth experience, and body image concerns); psychosocial stressors (eg, lack of social support, adverse life events, history of anxiety, and food insecurity); physical health conditions (eg, gestational diabetes and hypertension); and race or ethnicity (eg, Black and Latinx individuals). Nevertheless, little is known about which combination of factors most accurately predicts future risk of developing perinatal depression. Moreover, current perinatal depression screening tools administered during prenatal care are designed to quantify active depression symptoms and do not identify risk factors for future depressive episodes [14,15].

In response to this gap, we developed an interpretable machine learning–based predictive screener that can be administered directly to patients in the first trimester to identify those at risk of developing depression for the first time during pregnancy [16,17]. This 6-question predictive screener was designed to identify individuals who will develop new-onset depression among those who are currently asymptomatic and have no history of depression. The screener was co-designed with clinicians and people with lived experience through an iterative process, including focus groups with providers (eg, obstetricians, family medicine physicians, and behavioral health professionals) and virtual community engagement studios with pregnant and postpartum women, the details of which are described in a prior publication [14]. The predictive screener was shown to have a high degree of accuracy (area under the receiver operating characteristic curve=0.93), sensitivity (0.90), and specificity (0.81) [17]. However, the ultimate usefulness of this risk screening tool will be in its adoption into clinical workflow and its acceptability to the patients completing it. To fully close the loop, successful implementation also requires the provision of preventive care in response to a positive screener score.

A review encompassing 19 systematic reviews and meta-analyses and 152 randomized controlled trials across 26 countries identified therapy-based or physical activity interventions as having a small to moderate effect on the severity of perinatal depression symptoms [18]. Interpersonal and cognitive behavioral therapy are similarly effective for perinatal depression prevention [19-22]. Earlier therapeutic intervention can more effectively address perinatal mental health risk. With developments in predictive screening, there is an opportunity to tailor preventive care to address immediate, pressing risk factors (eg, food insecurity, labor and delivery fears, or lack of social support), while also developing strategies for long-term resilience to manage new and ongoing stressors. However, to date, patient-preferred preventive care options for those at risk of developing perinatal depression have not been documented.

The majority of women receiving prenatal care in our clinical setting access their health care information through a patient portal. As such, deployment of prenatal risk screeners through the patient portal is our standard workflow and is increasingly common in obstetric practices nationally [23-26]. Here, we examined the feasibility of deploying our predictive depression screener through our digital patient portal, as well as patients' receptivity to completing it. To do this, we deployed the screener to women receiving first-trimester prenatal care in our health care system, followed by a survey on their experience with completing it. We also solicited their preferences for preventive care resources.

Methods

Study Design and Setting

Participants were women (aged ≥18 years) receiving prenatal care from 1 of 4 University of Pittsburgh Medical Center obstetric practices. Electronic health records in the participating practices were screened for study eligibility on a weekly basis between October 2024 and January 2025. Eligibility criteria included being in the first trimester of pregnancy and screening negative for active first-trimester depressive symptoms using the Patient Health Questionnaire–9 [15].

Eligible pregnant women were sent an invitation via their patient portal to complete the 6-question predictive screener and subsequent survey questions related to the screener. The invitation contained a unique survey link that expired at the end of the first trimester. Upon accepting the link and completing informed consent, a pregnant woman was considered a study participant. Participants were not provided with a risk score. This allowed for independent validation of the screener in this population through passive observation of depressive symptoms and diagnoses over the course of the pregnancy. However, mental health and other pregnancy resources were provided to all participants, regardless of risk, to reflect best clinical practice.

Measures

Feasibility of screening was assessed by measuring (1) the proportion of pregnant women eligible from the larger clinical population (*eligibility rate*), (2) the number of eligible individuals who engaged with the screener study invitation (*reach*), and (3) the number of completed surveys among those who accessed the survey (*completion rate*).

Acceptability was assessed with a series of Likert-scale questions (1=strongly disagree to 5=strongly agree), which included participant evaluations of screener question clarity ("The screener questions were clear"), ease of completion ("The screener was easy to complete"), and comfort with and honesty in disclosing responses ("I would be comfortable sharing my answers with my provider" and "I would have no concerns about answering the questions honestly," respectively).

XSL•FO
RenderX

Participants were asked about the usefulness of the screener as part of routine prenatal care, including the benefits of knowing the results of the screener ("What would be the *benefit(s) to you personally* of knowing you are likely to develop depression at some point during your pregnancy?"), their preferred method of receiving screener results ("If you completed a screener like this as part of your prenatal care, *how* would you prefer to receive your results?"), perceived helpfulness of knowing depression risk ("How *helpful* would it be to know whether you are likely to develop depression at some point during your pregnancy?"), and any concerns they might have if they were to learn that they were likely to develop depression at some point during their pregnancy ("What *concern(s) would you personally have* about knowing you are likely to develop depression at some point during your pregnancy?"). Participants were also asked to endorse from a prespecified list and/or offer open-ended options for desired preventive care resources ("If you learned you were likely to develop depression at some point in your pregnancy, which of the following resources would you be *most likely to use* if available to you at *low or no-cost*?") and offer any additional thoughts or comments they have about the survey or depression prevention ("We welcome any other thoughts you have about your participation in this survey or preventing depression in pregnancy."). Deployment feasibility and acceptability of the screener were the primary outcomes of this study. Secondary analyses examined differences in attitudes between high- and low-risk pregnant women.

Statistical Analysis

Descriptive analyses used means and SDs for continuous variables (eg, maternal age), medians and IQRs for ordinal variables (eg, Likert-scale responses), and distributions for categorical variables. The study sample was stratified by screener score into those with and without risk of developing moderate to severe depression symptoms. Differences in survey responses by risk category were assessed using chi-square tests for categorical and binary responses. The distributions of continuous and Likert-scale responses were assessed for normality using the D'Agostino-Pearson test. Differences between normally distributed responses were evaluated using 2-sample 2-tailed *t* tests, and nonnormally distributed responses were evaluated using Mann-Whitney *U* tests. Effect size was calculated using Cliff *d*, a robust nonparametric method appropriate for ordinal data such as Likert-scale data [27]. An α level of .05 was used for statistical significance. All respondents completed every survey item, resulting in no missing data. Analyses were performed using Python (version 3.12.4).

Rapid inductive thematic coding was applied to open-ended survey responses by 2 coders, each with qualitative coding experience and one with formal training in rapid qualitative analysis. After reading all open-ended survey responses, the coders cocreated a codebook. Responses were then independently coded. Once independent coding was complete, codes were compared, and discrepancies were discussed until consensus (ie, consistent classification) was reached.

Ethical Considerations

All study procedures were approved by the University of Pittsburgh Institutional Review Board (STUDY24020098) and are reported according to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for reporting observational studies [28]. We conducted pretests of the survey with pregnant or recently pregnant colleagues, gathering feedback on the design and clarity of the instructions and survey questions. Participants provided digital consent for study participation and were informed that they could withdraw their consent at any time. They received a US $20 gift card for completing the screener and follow-up survey, which took approximately 10 minutes. All participant data were stored in a confidential and secure manner and access was restricted to authorized members of the research team.

Results

Survey Participants: Reach and Risk

During the study period, 1925 women received prenatal care at one of the participating obstetric practices. Of these 1925 women, 810 (42.1%) had an active patient portal, were at a gestational age of <14 weeks, and had scored <5 on the Patient Health Questionnaire–9 measure of active depression symptoms (*eligibility rate*). Of the 810 sent an invitation through their patient portal, only 2.8% (n=23) did not access it. In total, 36.7% (289/787) who accessed the invitation opened the study link (*reach*) and, of those, 88.2% (255/289) completed it (*completion rate*).

All participants were female with a mean age of 32 (SD 4.5) years. Most were White (217/255, 85.1%) and non-Hispanic (232/255, 91%), married (196/255, 76.9%), and had commercial insurance (196/255, 76.9%). Full demographic information is available in Table 1. Participants were demographically representative of the larger eligible population.

XSL•FO
RenderX

Table . Demographic characteristics of participants (n=255).

Demographics	Study sample
Age (years), mean (SD)	32 (4.5)
Gestational age (days), mean (SD)	71 (11.0)
Nulliparous, n (%)	111 (43.5)
Household income (US $), n (%)	
<25,000	23 (9)
25,000-54,999	29 (11.4)
55,000-99,999	37 (14.5)
100,000-149,999	80 (31.4)
>150,000	85 (33.3)
Prefer not to answer or missing	1 (0.4)
Poverty line[a], n (%)	
Above	193 (75.7)
Below	47 (18.4)
Uncertain[b]	15 (5.9)
Race[c], n (%)	
American Indian or Alaska Native	1 (0.4)
Asian	10 (3.9)
Black	16 (6.3)
White	217 (85.1)
Multiracial	9 (3.5)
Unknown	0 (0)
Other	5 (2)
Prefer not to answer or missing	5 (2)
Ethnicity, n (%)	
Hispanic or Latinx	11 (4.3)
Non-Hispanic	232 (91)
Unknown	3 (1.2)
Prefer not to answer or missing	9 (3.5)
Marital status, n (%)	
Single or never married	37 (14.5)
Married or domestic partnership	205 (80.4)
Divorced or separated	6 (2.4)
Other	7 (2.7)
Employment status, n (%)	
Employed full-time	172 (67.4)
Employed part-time	27 (10.6)
Stay-at-home parent	29 (11.4)
Seeking employment	7 (2.7)
Not working	14 (5.5)
Other	6 (2.4)
Insurance type, n (%)	

XSL•FO
RenderX

Demographics	Study sample
Commercial	196 (76.9)
Medicaid	44 (17.2)
Self-pay or other	15 (5.9)

[a]These thresholds are calculated using income and household size and are based on 215% of the federal poverty line in 2024, which is a standard threshold for receipt of federal or state assistance programs.

[b]Uncertain poverty line labels are for participants whose selected income bracket and household size contain the poverty line value for the given household size.

[c]Race options are not mutually exclusive; therefore, totals exceed 100%.

On the basis of established cutoff scores [17], 80% (204/255) of participants were identified by the predictive screener as having no or minimal depression risk and 20% (51/255) were identified as at risk of developing moderate to severe depression at a later stage of pregnancy.

Acceptability

The predictive screener demonstrated high acceptability among participants. Most agreed or strongly agreed (236/255, 92.5%) that the screener questions were clear (median 5, IQR 4-5), and 97.3% (248/255) reported that the questions were easy to complete (median 5, IQR 5-5). Most (236/255, 92.5%) agreed or strongly agreed that they would feel comfortable sharing the results of the screener with their obstetric care team providers (median 5, IQR 4-5). Similarly, the majority (247/255, 96.9%) agreed or strongly agreed that they would have no concerns about answering the questions honestly (median 5, IQR 5-5).

Table 2 presents acceptability metrics stratified by depression risk category, with those at risk of developing moderate-to-severe depression symptoms being less likely to endorse strong agreement with comfort and ease of honesty in responding to the screener.

Table . Median (IQR) responses to acceptability questions rated on a 5-point Likert scale.

Acceptability statements	No or minimal depression risk[a], median (IQR)	Moderate to severe depression risk[b], median (IQR)	Mann-Whitney *U* statistic	Effect size (Cohen *d*)	*P* value
The screener questions were clear.	5 (4-5)	5 (4-5)	4413.5	−0.15	.04
The screener was easy to complete.	5 (5-5)	5 (4-5)	4337.5	−0.17	.07
I would be comfortable sharing my answers with my provider.	5 (4-5)	4 (4-5)	3939.0	−0.24	.01
I would have no concerns about answering the questions honestly.	5 (5-5)	5 (4-5)	4218.5	−0.19	.04

[a]No or minimal depression risk: n=204.

[b]Moderate to severe depression risk: n=51.

Perspectives on Receiving Depression Screening Scores

Overall, 80% (204/255) reported that the screener was helpful or extremely helpful (median 4, IQR 4-5). Participants largely voiced a preference for receiving screener results through the patient portal (150/255, 58.8%), compared to receiving results in person (76/255, 29.8%) or by phone (13/255, 5.1%). Perspectives did not differ by depression risk.

Perceived Benefits and Risks of Depression Risk Screening

When asked about potential benefits of receiving their depression risk screening score, participants most frequently cited seeking appropriate preventive resources and/or treatments (63/204, 30.9%) and the opportunity to receive education from their obstetric health provider (53/204, 26%). One survey participant wrote that the feedback could help them "reduce risk factors associated with postpartum depression and postpartum psychosis."

Despite generally positive attitudes toward score disclosure, participants also expressed concerns. The most frequently reported concerns included increased anxiety or worry about developing depression (71/204, 34.8%) and worry about the inability to prevent depression (34/204, 16.7%). Several open-ended responses were offered, including concern about parenting ability and emotional connection to children if depression were to occur, premature medication use, and false positives.

Participants at risk of developing moderate to severe depression were significantly more likely than those with no or minimal depression risk to endorse the concern "fear that your health care provider will view you negatively" (5.9% vs 0.5%; *P*=.03). However, at-risk participants were also more likely to cite the ability to identify factors that might be affecting their mood as

XSL•FO
RenderX

a benefit (17.6% vs 5.9%; *P*=.01). Table 3 presents the complete breakdown of perceived benefits and risks.

Table . Breakdown of perceived benefits and risks, stratified by depression risk.

Benefits or concerns	Participants with no depression risk[a], n (%)	Participants with depression risk[b], n (%)	*P* value[c]
Benefits			
Allows you to seek appropriate preventive resources (ie, community resources and/or preventive treatments)	63 (30.9)[d]	12 (23.5)	.39
Provides an opportunity to receive education on pregnancy and postpartum depression, symptoms, and/or warning signs from your health care provider	53 (26.0)	13 (25.5)	.99
Provides information to discuss with your support system (eg, partner, friends, and family) about your needs	29 (14.2)	8 (15.7)	.96
Offers insight into why you might be feeling a certain way	26 (12.7)	4 (7.8)	.47
Helps you identify factors that might be affecting your mood	12 (5.9)	9 (17.6)	*.01*[e]
Gives you and your health care provider an opportunity to discuss depression risk	15 (7.4)	4 (7.8)	.99
Another benefit (open-ended)	1 (0.5)	0 (0.0)	.99
Concerns			
Increased anxiety and/or worry about developing depression	71 (34.8)	19 (37.3)	.87
Worry about your ability to prevent depression	34 (16.7)	5 (9.8)	.32
Fear that the results will be reported to child protective services	10 (4.9)	7 (13.7)	.052
Fear that others (eg, partner, friends, and family) will view you negatively	8 (3.9)	5 (9.8)	.18
Inability to access the right preventive resources	10 (4.9)	3 (5.9)	.99
Concern that your health care provider will doubt your parenting abilities	8 (3.9)	1 (2.0)	.80
Fear that your health care provider will view you negatively	1 (0.5)	3 (5.9)	*.03*
Concern that your provider will not know the appropriate education or resources to offer you	1 (0.5)	2 (3.9)	.19
Another concern (open-ended)	6 (2.9)	1 (2.0)	.99

[a]Participants with no depression risk: n=204.

[b]Participants with depression risk: n=51.

[c]*P* values are determined using the chi-square test to compare perceived benefits and concerns between those with and without first-time onset depression in our sample.

[d]Perceived benefits and concerns are not mutually exclusive; therefore, totals exceed 100% of participants.

[e]Italicized p-values indicate a statistically significant difference between those with no depression risk and those with depression risk.

Desired Preventive Care Resources

When asked about desired preventive health resources, participants endorsed multiple options. Preferred resources included counseling (197/255, 77.3%), with 61.8% (126/197) endorsing in-person therapy and 41.6% (82/197) endorsing app-based delivery. Mind- or body-based interventions (166/255, 65.1%) were frequently endorsed, with “yoga, exercise classes, or gym membership” (114/166, 68.7%) being a highly desired intervention, followed by prenatal classes or support groups (81/255, 31.8%).

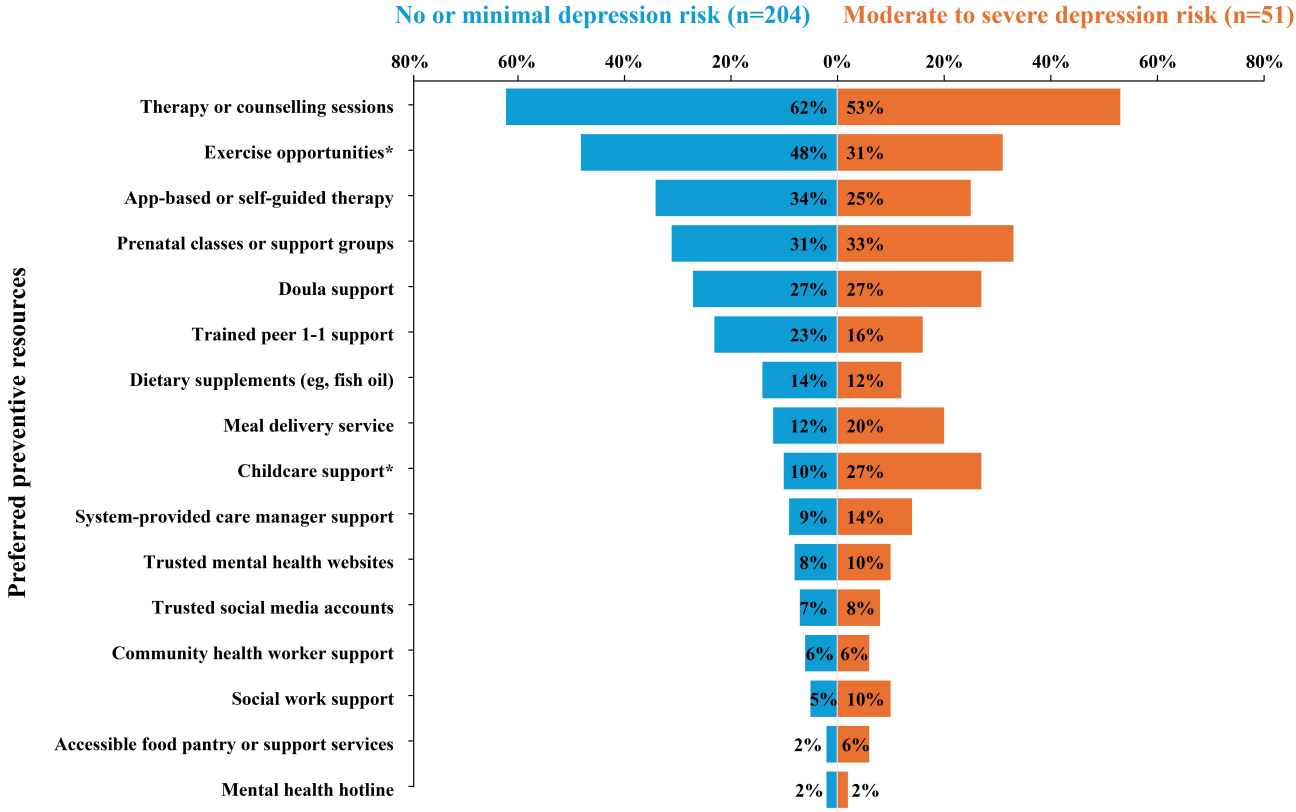
Participants (n=61) shared several resources not otherwise listed in open-ended responses. From most to least frequently listed, these included mental health resources for their spouse or support person (eg, “Education for family/spouses so they can possibly help the individual battling depression”), self-care opportunities (eg, “massage”), practical support such as cleaning services or meal preparation (eg, “A service similar to task rabbit that is geared towards pregnancy needs. Perhaps people that could help identify inexpensive sources of second hand furniture/items, assembly help, specific errand support, meal prep/planning for once baby arrives, etc.”), parental leave,

XSL•FO
RenderX

increased medical touchpoints during pregnancy, additional mental health education (eg, books and fact sheets), medication for depression prevention, structural changes (eg, improved health care access and transportation), access to a live mental health support person ("Mental health support chat forums with a licensed counselor"), a more holistic approach to health care, and additional faith or spiritual support.

Figure 1 shows preferred resources stratified by depression risk. Notably, less than 10% (20/204) of those with no or minimal depression risk expressed interest in childcare support compared to more than 25% (14/51) of those at risk of developing moderate to severe depression (χ^2_1=9.52; n=255; *P*=.002). Although highly endorsed by both risk groups, participants with no or minimal depression risk expressed greater interest in exercise opportunities (χ^2_1=3.94; n=255; *P*=.05).

Figure 1. Preferred resources stratified by depression risk. *Substantial differences in resource desirability.



Open-Ended Feedback

After completing the survey, 55.3% (141/255) of participants shared optional thoughts about screener participation and about preventing depression in pregnancy. The most frequent open-ended comments expressed appreciation for the study or confirmed that participants would find a screener of this kind helpful. Additional responses ranged from suggestions for the screener or resources (eg, "A big area of concern for working moms is their jobs. I felt that the questions did not address the concern as much.") to specific stressors (eg, "I think one of the biggest stressors that can contribute greatly to mood in the first trimester is the viability and health/wellness of the baby, i.e., risk of miscarriage or is there any genetic or chromosomal issue?") and barriers to care ("I feel like cost is one of the biggest barriers to seeking help, just ahead of the stigma that still exists."). One participant reported feeling that a discussion of depression may exacerbate depressive symptoms. Those at risk of developing moderate to severe depression more frequently expressed the need for help and social support during pregnancy, while those with no or minimal depression risk largely highlighted the need for depression- and anxiety-specific resources.

Discussion

Principal Findings

This study demonstrates the acceptability of a machine learning–based predictive screener designed to support perinatal depression prevention. Predictive screening is an emerging field, and, to the best of our knowledge, this is the first study to examine receptivity to a predictive mental health screener, to query pregnant women about their desired preventive care options, and to examine differences in desired resources by predicted risk level. While machine learning algorithms are being newly applied to mental health prediction, their ultimate success depends on adoption into clinical workflow, which requires patient engagement and a feasible implementation process [29]. Here, we highlight an approach to evaluating receptivity to such predictive screeners while testing the feasibility of sending the screener through our health care system's digital patient portal.

In this study, pregnant women who completed the screener largely reported positive screener experiences, with over 90% agreeing that it was clear, easy, and comfortable to respond to and that they had no concerns about responding honestly. More

XSL•FO
RenderX

than 80% thought it would be helpful to know their risk of developing depression symptoms during pregnancy, regardless of their risk score. Those at risk of developing moderate to severe depression were slightly more likely to report feeling uncomfortable sharing their honest responses with an obstetric health care provider, suggesting that the screener questions touch on risk factors that are challenging to disclose. Perceived benefits and risks of the screener were similar between the no- or minimal-risk and at-risk groups, with 2 exceptions: those at risk of developing depression reported greater value in the screener for helping them understand contributors to their mood. They also reported greater fear about being judged by their provider. These anticipated challenges may be grounded in stigma and prior experiences of discrimination [30,31].

Deploying the screener through the patient portal reached a little more than one-third of the eligible pregnant women receiving prenatal care in the participating obstetric practices. This suggests that in-person screening may need to be offered as part of a clinical workflow to deploy the screener effectively to a majority of the patient population. However, among those who initiated the screening survey, engagement was encouraging, with almost 90% of participants completing the entire screening process. Thus, if the eligible patient reach is optimized through multiple outreach channels, the screener could have a high completion rate.

Survey participants identified specific perceived benefits and concerns associated with learning about being at risk of developing depression symptoms. The desire for therapy, whether in person or through an app, is reflective of the current best practices for preventive care, which include cognitive behavioral therapy and interpersonal therapy. Notably, participants also reported a desire for exercise opportunities and support groups, echoing research showing the protective effects of physical activity and social support on depression [18,21,32,33]. Child care was highly desired among the study population at risk of developing moderate to severe depression, suggesting that direct tangible support may impact the trajectory of depression risk. Financial stressors during pregnancy have previously been associated with poor outcomes for both mothers and infants. Findings from recent studies of guaranteed income programs for perinatal women show promising evidence of improvements in maternal mental health and infant developmental outcomes associated with alleviating financial stress [32,34-36].

This study contributes to a burgeoning literature on the importance of providing preventive mental health care during pregnancy and the postpartum period [37]. The perinatal period is uniquely vulnerable to depression due to a combination of hormonal changes, novel stressors (eg, pregnancy-related anxiety), and environmental factors [38]. However, up to half of the cases remain undiagnosed [39]. Although women without a psychiatric history are at lower baseline risk, they are more likely to be overlooked when they do experience depression, with lower rates of both depression screening and referral to care [40]. One solution to address the lack of diagnosis is more frequent, routine depression screening [13]. Care navigators dedicated to monitoring mental health symptoms throughout pregnancy, for example, have been shown to improve depressive symptoms among the patients they serve [41,42]. However, to date, there have been no screening tools available to support those professionals in identifying risk among individuals not actively reporting depressive symptoms [11]. If successfully implemented, predictive screeners could make the work of such care navigators more targeted and efficient.

Limitations

This study has several limitations. Eligible participants were limited to those who could be contacted through an active patient portal. Although most of our prenatal patients have an active patient portal account, it is possible that these portals are not frequently accessed. Other studies suggest that individuals with limited health literacy are less likely to use patient portals to access health information [43-45]. Thus, the moderate response rate may introduce selection bias if individuals more familiar with portal-based screening were more likely to participate, whereas those with lower health literacy were less likely to engage. There were some differences in responses by participant risk level. Although these differences seem intuitive (eg, participants with greater depression risk were more likely to find the screener helpful in identifying influences on mood yet were more concerned about potential stigma from providers), they should be interpreted with caution given the modest sample size of the at-risk group and the number of comparisons conducted. The study sample was generally representative of the patient population of the academic medical system in which it was conducted, and the survey was developed using a patient-centered design approach with a diverse population. However, the results cannot be generalized to other health care systems in the United States or to clinics with more diverse patient populations. While the screener showed high accuracy within the sample on which it was built [17], further work is needed to establish independent external validation. Additional research must address how to best implement this screener within our health care system, including a testing workflow to fully close the loop on preventive care. A future goal would be to conduct validation and acceptability research of the screener in non-US populations.

Conclusions

This study is a first step toward understanding how a machine learning–based prevention-focused predictive depression screener can be integrated into a clinical workflow for preventive care. We found that only one-third of the eligible pregnant women in our clinics completed the screener via their digital patient portal. Ultimately, deploying preventive screening tools will require a hybrid digital and in-person screening approach to comprehensively screen all pregnant women who may be at risk of developing perinatal depression. Our risk screener was designed to identify individuals who may need additional evaluation and care from a trained provider. As such, it is intended as a supplement to current prenatal care, allowing for earlier identification of those at risk rather than serving as a replacement for professional mental health evaluation. Our findings suggest that the screener was acceptable to those who completed it. Moreover, participants voiced clear preferences for certain forms of preventive care in response to learning their risk. Participants' preference for receiving predictive depression

XSL•FO
RenderX

screening results through the patient portal suggests a promising approach to expanding access to mental health assessment among individuals already engaged with digital health platforms. Our findings also suggest some next steps on how to inform the clinical workflow to support a more personalized approach to preventive care in pregnancy, leveraging the most endorsed preventive care resources (eg, counseling, peer support groups, exercise, and childcare support) to address the individual risk factors identified by the predictive screener.

Acknowledgments

The authors gratefully acknowledge Sarah Burns, MSW, LSW, at the University of Pittsburgh for assistance with survey preparation.

Funding

This study was supported by grants from the National Institute of Mental Health (R34MH130950 and R01MH139097). The funder had no role in study design, data collection and analysis, the decision to publish, or preparation of the manuscript.

Data Availability

Study data will be made available upon reasonable request to the corresponding author.

Authors' Contributions

Conceptualization: TK, KW
Data curation: SR, LL
Formal analysis: SR, LC
Funding acquisition: TK
Methodology: TK, KW
Project administration: SR, LC
Resources: LL
Supervision: TK
Visualization: SR
Writing—original draft: TK
Writing—review and editing: all authors

Conflicts of Interest

TK and HS hold equity in Naima Health LLC. All other authors declare no other conflicts of interest.

References

1. Nidey N, Tabb KM, Carter KD, et al. Rurality and risk of perinatal depression among women in the United States. J Rural Health 2020 Jan;36(1):9-16. [doi: 10.1111/jrh.12401] [Medline: 31602705]
2. Al-Abri K, Edge D, Armitage CJ. Prevalence and correlates of perinatal depression. Soc Psychiatry Psychiatr Epidemiol 2023 Nov;58(11):1581-1590. [doi: 10.1007/s00127-022-02386-9] [Medline: 36646936]
3. O'Connor TG, Heron J, Glover V, Alspac Study Team. Antenatal anxiety predicts child behavioral/emotional problems independently of postnatal depression. J Am Acad Child Adolesc Psychiatry 2002 Dec;41(12):1470-1477. [doi: 10.1097/00004583-200212000-00019] [Medline: 12447034]
4. Slomian J, Honvo G, Emonts P, Reginster JY, Bruyère O. Consequences of maternal postpartum depression: a systematic review of maternal and infant outcomes. Womens Health (Lond) 2019;15:1745506519844044. [doi: 10.1177/1745506519844044] [Medline: 31035856]
5. Jahan N, Went TR, Sultan W, et al. Untreated depression during pregnancy and its effect on pregnancy outcomes: a systematic review. Cureus 2021 Aug;13(8):e17251. [doi: 10.7759/cureus.17251] [Medline: 34540477]
6. Alhusen JL, Ayres L, DePriest K. Effects of maternal mental health on engagement in favorable health practices during pregnancy. J Midwifery Womens Health 2016;61(2):210-216. [doi: 10.1111/jmwh.12407] [Medline: 26849176]
7. Tong VT, Farr SL, Bombard J, D Angelo D, Ko JY, England LJ. Smoking before and during pregnancy among women reporting depression or anxiety. Obstet Gynecol 2016 Sep;128(3):562-570. [doi: 10.1097/AOG.0000000000001595] [Medline: 27500342]
8. Chin K, Wendt A, Bennett IM, Bhat A. Suicide and maternal mortality. Curr Psychiatry Rep 2022 Apr;24(4):239-275. [doi: 10.1007/s11920-022-01334-3] [Medline: 35366195]
9. Biaggi A, Conroy S, Pawlby S, Pariante CM. Identifying the women at risk of antenatal anxiety and depression: a systematic review. J Affect Disord 2016 Feb;191:62-77. [doi: 10.1016/j.jad.2015.11.014] [Medline: 26650969]
10. Di Florio A, Forty L, Gordon-Smith K, et al. Perinatal episodes across the mood disorder spectrum. JAMA Psychiatry 2013 Feb;70(2):168-175. [doi: 10.1001/jamapsychiatry.2013.279] [Medline: 23247604]

XSL•FO
RenderX

11. US Preventive Services Task Force, Curry SJ, Krist AH, et al. Interventions to prevent perinatal depression: US Preventive Services Task Force recommendation statement. JAMA 2019 Feb 12;321(6):580-587. [doi: 10.1001/jama.2019.0007] [Medline: 30747971]
12. ACOG Committee Opinion No. 757 Summary: screening for perinatal depression. Obstet Gynecol 2018;132(5):1314-1316. [doi: 10.1097/AOG.0000000000002928] [Medline: 30629563]
13. Screening and diagnosis of mental health conditions during pregnancy and postpartum: ACOG Clinical Practice Guideline No. 4. Obstet Gynecol 2023 Jun 1;141(6):1232-1261. [doi: 10.1097/AOG.0000000000005200] [Medline: 37486660]
14. Cox J. Use and misuse of the Edinburgh Postnatal Depression Scale (EPDS): a ten point "survival analysis". Arch Womens Ment Health 2017 Dec;20(6):789-790. [doi: 10.1007/s00737-017-0789-7] [Medline: 29101480]
15. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x]
16. Williams K, Nikolajski C, Rodriguez S, et al. Incorporating end-user perspectives into the development of a machine learning algorithm for first time perinatal depression prediction. J Am Med Inform Assoc 2025 Jul 1;32(7):1186-1198. [doi: 10.1093/jamia/ocaf086] [Medline: 40493528]
17. Krishnamurti T, Rodriguez S, Wilder B, Gopalan P, Simhan HN. Predicting first time depression onset in pregnancy: applying machine learning methods to patient-reported data. Arch Womens Ment Health 2024 Dec;27(6):1019-1031. [doi: 10.1007/s00737-024-01474-w] [Medline: 38775822]
18. Singh B, Olds T, Curtis R, et al. Effectiveness of physical activity interventions for improving depression, anxiety and distress: an overview of systematic reviews. Br J Sports Med 2023 Sep;57(18):1203-1209. [doi: 10.1136/bjsports-2022-106195] [Medline: 36796860]
19. Pettman D, O'Mahen H, Blomberg O, Svanberg AS, von Essen L, Woodford J. Effectiveness of cognitive behavioural therapy-based interventions for maternal perinatal depression: a systematic review and meta-analysis. BMC Psychiatry 2023 Mar 29;23(1):208. [doi: 10.1186/s12888-023-04547-9] [Medline: 36991389]
20. Li X, Laplante DP, Paquin V, Lafortune S, Elgbeili G, King S. Effectiveness of cognitive behavioral therapy for perinatal maternal depression, anxiety and stress: a systematic review and meta-analysis of randomized controlled trials. Clin Psychol Rev 2022 Mar;92:102129. [doi: 10.1016/j.cpr.2022.102129] [Medline: 35123346]
21. O'Connor E, Senger CA, Henninger ML, Coppola E, Gaynes BN. Interventions to prevent perinatal depression: evidence report and systematic review for the US Preventive Services Task Force. JAMA 2019 Feb 12;321(6):588-601. [doi: 10.1001/jama.2018.20865] [Medline: 30747970]
22. Jiang X, Li H, Wang D, Shan L, Wang F, Kang Y. Efficacy of nondrug interventions in perinatal depression: a meta-analysis. Psychiatry Res 2022 Nov;317:114916. [doi: 10.1016/j.psychres.2022.114916] [Medline: 37000462]
23. Implementing telehealth in practice: ACOG Committee Opinion Summary, number 798. Obstet Gynecol 2020 Feb;135(2):493-494. [doi: 10.1097/AOG.0000000000003672] [Medline: 31977794]
24. Holder K, Feinglass J, Niznik C, Yee LM. Use of electronic patient messaging by pregnant patients receiving prenatal care at an academic health system: retrospective cohort study. JMIR Mhealth Uhealth 2024 Apr 29;12:e51637. [doi: 10.2196/51637] [Medline: 38686560]
25. Clarke JR, Gibson M, Savaglio M, Navani R, Mousa M, Boyle JA. Digital screening for mental health in pregnancy and postpartum: a systematic review. Arch Womens Ment Health 2024 Aug;27(4):489-526. [doi: 10.1007/s00737-024-01427-3] [Medline: 38557913]
26. Ukoha EP, Yee LM. Use of electronic patient portals in pregnancy: an overview. J Midwifery Womens Health 2018 May;63(3):335-339. [doi: 10.1111/jmwh.12761] [Medline: 29758124]
27. Meissel K, Yao ES. Using Cliff's delta as a non-parametric effect size measure: an accessible web app and R tutorial. Pract Assess Res Eval 2024 Jan 22;29(1):2. [doi: 10.7275/pare.1977]
28. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007 Oct 20;370(9596):1453-1457. [doi: 10.1016/S0140-6736(07)61602-X] [Medline: 18064739]
29. Nilsen P, Svedberg P, Nygren J, Frideros M, Johansson J, Schueller S. Accelerating the impact of artificial intelligence in mental healthcare through implementation science. Implement Res Pract 2022;3:26334895221112033. [doi: 10.1177/26334895221112033] [Medline: 37091110]
30. Andrada A. The legacy of stigma: American single mothers. In: Living Legacies of Social Injustice: Power, Time and Social Change, 1st edition: Routledge; 2023:73-90. [doi: 10.4324/9781003311928]
31. Andrada A. 'Woman' and (un)partnered mother. In: Intersections of Mothering: Feminist Accounts, 1st edition: Routledge; 2019:44-56. [doi: 10.4324/9780429430411-4]
32. Seward N, Garman E, Davies T, et al. The importance of social support and food insecurity in improving outcomes for perinatal depression: a causal mediation analysis of the AFFIRM-SA trial. J Affect Disord 2025 Jul 1;380:552-560. [doi: 10.1016/j.jad.2025.03.159] [Medline: 40154807]
33. Physical activity and exercise during pregnancy and the postpartum period: ACOG Committee Opinion, number 804. Obstet Gynecol 2020;135(4):e178-e188. [doi: 10.1097/AOG.0000000000003772] [Medline: 32217980]

XSL•FO
RenderX

34. Reesor-Oyer L, Cepni AB, Lee CY, Zhao X, Hernandez DC. Disentangling food insecurity and maternal depression: which comes first? Public Health Nutr 2021 Nov;24(16):5506-5513. [doi: 10.1017/S1368980021000434] [Medline: 33517950]
35. Nishimura HM, Snguon S, Moen M, Dean LT. Guaranteed income and health in the United States and Canada: a scoping review. Epidemiol Rev 2025 Jan 10;47(1):mxaf003. [doi: 10.1093/epirev/mxaf003] [Medline: 40114543]
36. Rizvi A, Kearns M, Dignam M, et al. Effects of guaranteed basic income interventions on poverty-related outcomes in high-income countries: a systematic review and meta-analysis. Campbell Syst Rev 2024 Jun;20(2):e1414. [doi: 10.1002/cl2.1414] [Medline: 38887375]
37. Radoš SN, Ganho-Ávila A, Rodriguez-Muñoz MF, et al. Evidence-based clinical practice guidelines for prevention, screening and treatment of peripartum depression. Br J Psychiatry 2025 Jun 26;227(5):1-12. [doi: 10.1192/bjp.2025.43] [Medline: 40566968]
38. Rajendran B, Ibrahim SU, Ramasamy S. Maternal and neonatal risk factors associated with perinatal depression-a prospective cohort study. Indian J Psychol Med 2024 Jan;46(1):24-31. [doi: 10.1177/02537176231176405] [Medline: 38524952]
39. Sudhanthar S, Sheikh ZE, Thakur K. Postpartum depression screening: are we doing a competent job? BMJ Open Qual 2019;8(4):e000616. [doi: 10.1136/bmjoq-2018-000616] [Medline: 31673639]
40. McNicholas E, Boama-Nyarko E, Julce C, et al. Understanding perinatal depression care gaps by examining care access and barriers in perinatal individuals with and without psychiatric history. J Womens Health (Larchmt) 2023 Oct;32(10):1111-1119. [doi: 10.1089/jwh.2022.0306] [Medline: 37582274]
41. Byatt N, Brenckle L, Sankaran P, et al. Effectiveness of two systems-level interventions to address perinatal depression in obstetric settings (PRISM): an active-controlled cluster-randomised trial. Lancet Public Health 2024 Jan;9(1):e35-e46. [doi: 10.1016/S2468-2667(23)00268-2] [Medline: 38176840]
42. Lemon LS, Chaney J, Young M, et al. Evaluating impact of population health outreach and an interdepartmental approach to perinatal depression. Gen Hosp Psychiatry 2026;99:24-29. [doi: 10.1016/j.genhosppsych.2026.01.003] [Medline: 41512404]
43. Coughlin SS, Stewart JL, Young L, Heboyan V, De Leo G. Health literacy and patient web portals. Int J Med Inform 2018 May;113:43-48. [doi: 10.1016/j.ijmedinf.2018.02.009] [Medline: 29602432]
44. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016 Oct 4;18(10):e264. [doi: 10.2196/jmir.6349] [Medline: 27702738]
45. Tieu L, Schillinger D, Sarkar U, et al. Online patient websites for electronic health record access among vulnerable populations: portals to nowhere? J Am Med Inform Assoc 2017 Apr 1;24(e1):e47-e54. [doi: 10.1093/jamia/ocw098] [Medline: 27402138]

Abbreviations

STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by J Sarvestan; submitted 05.Aug.2025; peer-reviewed by S Alves; revised version received 19.Mar.2026; accepted 20.Mar.2026; published 05.May.2026.

Please cite as:
Krishnamurti T, Rodriguez S, Cope L, Lemon L, Gopalan P, Nikolajski C, Simhan H, Williams K
Feasibility and Acceptability of a Prevention-Focused Screener for Perinatal Depression Risk: Mixed Methods Cohort Study
JMIR Hum Factors 2026;13:e81638
URL: https://humanfactors.jmir.org/2026/1/e81638
doi:10.2196/81638

© Tamar Krishnamurti, Samantha Rodriguez, Leah Cope, Lara Lemon, Priya Gopalan, Cara Nikolajski, Hyagriv Simhan, Kelly Williams. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 5.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Using a Digital-Based Mindfulness Curriculum to Enhance Healthy Aging Outcomes in Community-Dwelling Older Adults in Taiwan: Mixed Methods Feasibility Study

Yu-Rung Wang[1*], PhD; Pei-Lun Hsieh[2*], PhD; Chia-Hsiu Chang[3*], PhD; Chia-Chi Hsiao[4*], PhD; Mei-Lien Hu[5*], NP, MSN

[1]Department of Nursing and Center for Smart Healthcare Education, Chang Gung University of Science and Technology, Chiayi Campus, Puzi, Taiwan

[2]Department of Nursing, National Taichung University of Science and Technology, Taichung, Taiwan

[3]Department of Nursing, Hungkuang University, Taichung, Taiwan

[4]Chang Gung Medical Foundation, Chiayi Chang Gung Memorial Hospital, Puzi, Taiwan

[5]Department of Nursing, Linkou Chang Gung Memorial Hospital, 5 Fuxing St, Guishan District, Taoyuan City, Taiwan

[*]all authors contributed equally

Corresponding Author:
Mei-Lien Hu, NP, MSN
Department of Nursing, Linkou Chang Gung Memorial Hospital, 5 Fuxing St, Guishan District, Taoyuan City, Taiwan

Abstract

Background: Taiwan's status as a super-aged society underscores the need for efficient strategies to promote healthy aging. While the benefits of mindfulness-based interventions for sleep and mental health are evident, there is a shortage of cultural adaptations for Taiwan's older adults. Current courses mainly focus on mindfulness-based stress reduction, while neglecting mindfulness-based elder care, and mindful sustainable aging principles. However, the abstract nature of some mindfulness concepts, combined with limited digital support and localized resources, makes it challenging for older adults to engage effectively.

Objective: To enhance healthy aging outcomes in older community-dwelling adults in Taiwan, this study developed an 8-week theory-based mindfulness curriculum that combined the mindfulness-based stress reduction, mindfulness-based elder care, and mindful sustainable aging frameworks with digital health support.

Methods: This research employed a mixed methods feasibility study design. We used the modified Delphi method in phase 1 to reach agreement on curriculum content and digital material selection. Ten older adults participated in the pilot study, which comprised phase 2. We used quantitative measurements to assess mindfulness, sleep quality, cognition, heart rate variability, perceptions of aging, healthy aging perspectives, and physical function and activity at baseline, at weeks 4 and 8, and at a 1-month follow-up. Qualitative interviews were conducted to gather insights into participants' experiences.

Results: Participants achieved significant improvements in mindfulness, sleep quality, aging perception, healthy aging outlook, and physical function during the study (all $P<.05$). Cognition, heart rate variability, and physical activity levels showed no significant changes throughout the study. The qualitative data supported these findings, as participants selected simple practices, such as mindful breathing, which they practiced daily to enhance their emotional well-being and social connections. Participants valued the digital learning materials for aiding their practice; however, some older adults with reading difficulties faced challenges accessing the content.

Conclusions: Using the Delphi method resulted in an acceptable and feasible theory-based digitally supported mindfulness program that improved several indicators of healthy aging in older adults in Taiwan. Our findings need to be validated using longer trials to assess cognitive and physiological effects. Furthermore, digital accessibility requires further development.

(JMIR Hum Factors 2026;13:e84161) doi:10.2196/84161

KEYWORDS

mindfulness; older adults; healthy aging; Delphi method; mixed methods; digital learning

Introduction

Taiwan is rapidly becoming a super-aged society, prompting the government to develop long-term care policies aligned with the World Health Organization's (WHO) Decade of Healthy Aging framework [1,2]. Current community programs in Taiwan emphasize physical exercise and cognitive training but lack adequate organized support for mental health and attitudes toward aging [3-5]. The WHO model defines healthy aging as the maintenance of functional abilities alongside the

XSL•FO
RenderX

management of ongoing health conditions and external environmental challenges. Recent research suggests that healthy aging involves 3 key elements—an individual's wellness status, their ability to adapt, and their capacity to resist challenges—all while accounting for the mental and social aspects of aging [2,4,5]. Supporting older adults requires interventions that address both physical and psychological factors to help preserve their intrinsic capacity and functional ability.

According to Kabat-Zinn [6], mindfulness practice enables individuals to experience the present moment without judgment. Research has indicated that mindfulness practice can improve physical and mental health across different groups [7-11]. Moreover, mindfulness-based stress reduction (MBSR) has been shown to enable participants to reduce stress symptoms, improve sleep quality, and achieve better mental health outcomes [9-11]. Although older adults who participate in mindfulness-based interventions (MBIs) have demonstrated better sleep quality and higher life satisfaction, such programs have not shown positive effects on cognitive functions [12-14]. Mindfulness practices thus have the potential to help seniors during their aging journey, but they require consideration of the particular needs of older adults.

As older adults are still able to learn and develop skills, multiple frameworks have been developed for mindfulness practice during senior years. For example, the mindfulness-based elder care (MBEC) program aims to teach frail older adults and their caregivers to practice mindfulness [15-17]. The mindful sustainable aging (MSA) framework unites successful aging principles with disengagement, transcendence, and activity theory to help older adults maintain their outward activities while developing inner reflection and resilience [18,19]. These approaches, together with the core stress-reduction practices of MBSR [6,9,10], establish a strong theoretical foundation for mindfulness programs that focus on emotional regulation, the acceptance of aging, and the development of purpose for older adults. Community-based programs targeting this population therefore require interventions that combine these particular frameworks to achieve their goals.

The majority of mindfulness programs for older adults in Taiwan use MBSR as their primary format; however, these programs do not directly implement MBEC and MSA principles and thus may not address specific needs of older populations. Older participants may have difficulty understanding abstract concepts such as the "nine attitudes of mindfulness" (ie, nonjudging, patience, beginner's mind, trust, nonstriving, acceptance, letting go, gratitude, and generosity) [20]. Since older adults may lack classroom experience and have limited memory capacity, teachers can facilitate their learning through practical, hands-on activities rather than abstract examples. Additionally, the lack of suitable, culturally adapted, age-friendly materials for older Taiwanese adults can contribute to poor adherence after course completion.

Taiwan's digital health agenda encourages the development of digital learning programs that can be integrated into community-based services for older adults [21]. About half of Taiwanese adults aged65 years and above use the internet, and many are familiar with LINE (LY Corporation), a widely used communication and social media application [21]. Digital platforms incorporating reminder systems may support older adults in monitoring medical treatment schedules and making health-related behavioral changes [13]. However, developers should ensure that digital solutions are accessible to users who have reading difficulties, mistrust of web-based information, or difficulty operating digital systems.

Existing research has demonstrated a lack of mindfulness programs that are culturally adapted for Taiwanese older adults and that integrate basic digital tools while adhering to established principles of mindfulness and healthy aging. To address this gap, this study adopted a mixed methods approach and conducted a feasibility study. Our research team developed an 8-week mindfulness program for community-dwelling older adults in Taiwan, integrating components of MBSR, MBEC, and MSA. Home-practice support was provided through LINE and YouTube. The study had two primary aims: (1) to apply a modified Delphi approach to achieve expert consensus regarding curriculum content and digital educational resources and (2) to assess the program's feasibility, participant acceptance, and initial impact on mindfulness levels, sleep quality, cognitive ability, heart rate variability (HRV), perceptions of aging, physical activity, and overall physical functioning. Based on established evidence-based approaches, an initial digital mindfulness program framework incorporating proven methods was created and adapted to support Taiwanese older adults in managing the psychosocial and functional aspects of healthy aging through digital health support.

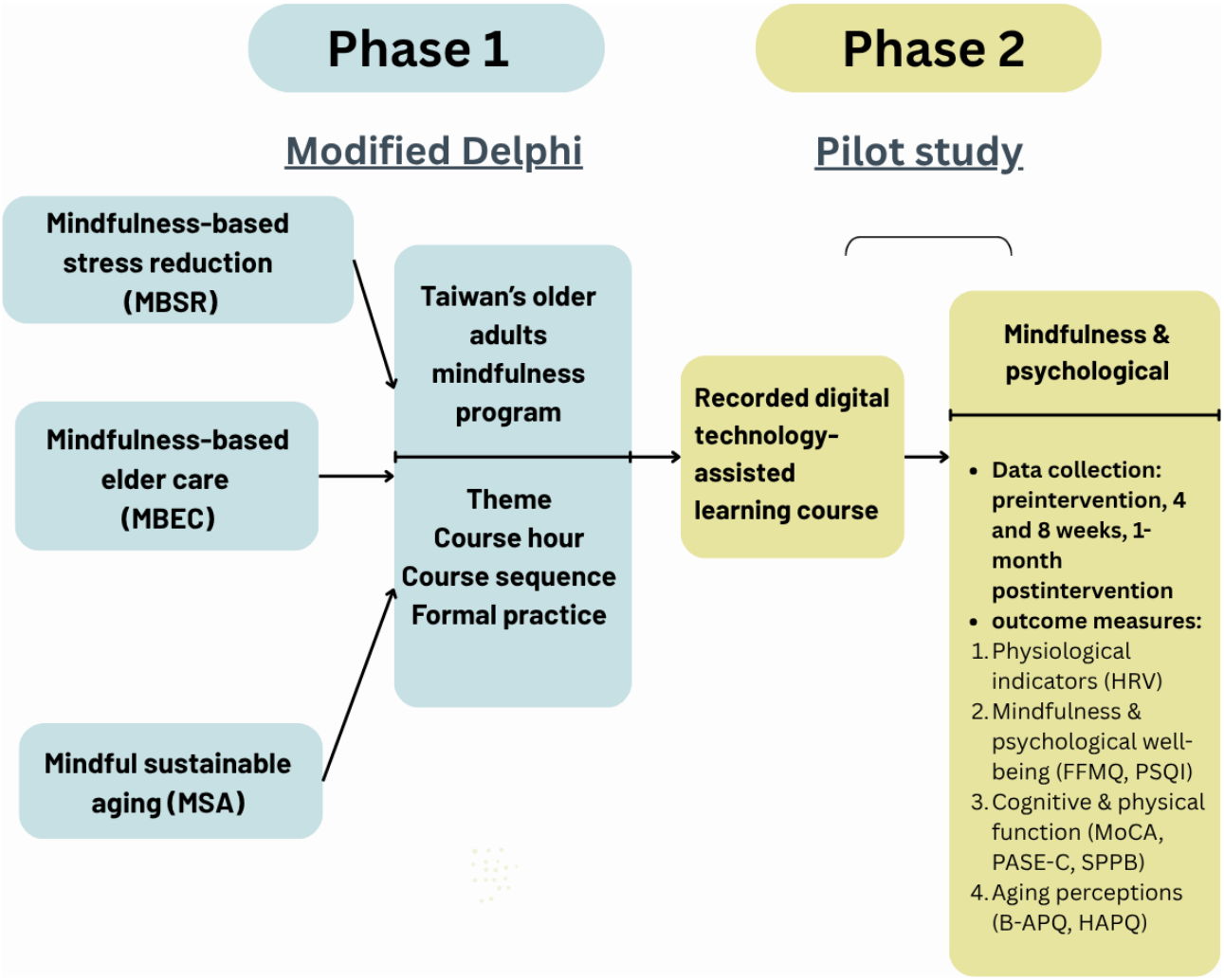
Methods

Study Design

The mixed methods feasibility design was based on Creswell and Plano Clark [22] (Figure 1). Phase 1 used a modified Delphi process to develop and validate a theory-based, digitally supported mindfulness curriculum tailored for community-dwelling older adults. Phase 2 consisted of a single-group pilot study with embedded qualitative interviews (QUAN→qual) to evaluate feasibility and acceptability and explore the curriculum's preliminary effects on healthy aging outcomes.

XSL•FO
RenderX

Figure 1. Conceptual and methodological framework of the 2-phase study: Delphi curriculum development and pilot intervention. B-APQ: Brief Aging Perceptions Questionnaire; FFMQ: Five Facet Mindfulness Questionnaire; HAPQ: Healthy Aging Perspectives Questionnaire; HRV: heart rate variability; MoCA: Montreal Cognitive Assessment; PASE-C: Chinese Physical Activity Scale for the Elderly; PSQI: Pittsburgh Sleep Quality Index; SPPB: Short Physical Performance Battery.



Phase 1: Delphi Curriculum Development

Theoretical Framework and Curriculum Design

We used a modified Delphi approach to determine the appropriate format and content of a mindfulness-based program for community-dwelling older adults. The curriculum incorporated 3 theoretical frameworks: MBSR [6], which teaches mindfulness practices and stress reduction for daily life; MBEC [15], which focuses on mindfulness practice in gerontological care settings; and MSA [19], which promotes mindfulness to help individuals improve adaptability and resilience in later life. The curriculum development process also included the creation of digital learning resources. These resources supported participants between in-person program sessions, facilitated adherence to home practices, and enhanced the intervention's ecological validity.

Participants

Experts were purposefully recruited from professional networks and academic institutions. Eligible participants included certified mindfulness instructors, physicians, nurses, physical and occupational therapists, schoolteachers, health educators, and case managers with a minimum of 2 years of experience in mindfulness instruction or elder care (Table S1 in Multimedia Appendix 1).

Procedure

The Delphi process was administered via mailed questionnaires across 3 sequential rounds conducted over approximately 6 months, with each round lasting approximately 2 months, including time for qualitative feedback synthesis and item revision. During the first round, the experts reviewed the curriculum materials and provided qualitative suggestions, which were synthesized and transformed into structured items. In the subsequent 2 rounds, experts rated these items quantitatively and offered additional feedback. This iterative process continued until consensus was achieved. The finalized 8-week curriculum and digital learning materials were then implemented and evaluated in phase 2 at a community care station in Chiayi City.

Data Analysis

Descriptive statistics were calculated for the Delphi survey responses. For each curriculum item, the median, IQR, and percentage of experts rating the item as "suitable" or "very

XSL•FO
RenderX

suitable" were calculated. Consensus was defined a priori as an IQR≤1.0 and at least 80% agreement among experts [22]. Items that did not reach consensus were revised according to experts' comments and resubmitted in the next round. This process continued until response stability was achieved.

Phase 2: Pilot Intervention Study (QUAN→Qual)

Quantitative Study

The second phase comprised a single-group pilot intervention involving 10 community-dwelling older adults. Participants completed standardized assessments at 4 time points: baseline (P0), week 4 (P1), week 8 (P2), and 1-month postintervention (P3). The repeated-measures longitudinal design allowed examination of both the immediate and long-term effects of the intervention on healthy aging outcomes.

Digital Support Component

The intervention combined in-person instruction with digital and communication tools to support participants in their mindfulness practice between sessions.

Two platforms were used:

- YouTube content: short instructor-recorded videos and audio-guided sessions demonstrating core practices—such as mindful breathing, body scans, gentle stretching, and mindful walking—were made available through YouTube. Participants could access these materials at home to support independent practice and reinforce what they had learned in class (Figure 2).
- LINE reminders and group chat: A LINE group was created for the class to facilitate communication. Brief practice reminders and encouraging messages were sent to participants up to 3 times daily (morning, noon, and evening), along with links to relevant YouTube content and simple emotional support cards. The group also provided a means for participants to ask questions, share experiences, provide mutual support, and maintain engagement between sessions (Figure 3).

This hybrid format combined the social advantages of in-person sessions with accessible, age-appropriate digital tools already familiar to participants. The perceived usefulness of these digital components was assessed through postprogram qualitative feedback; objective engagement metrics, such as video view counts or message read rates, were not collected.

XSL•FO
RenderX

Figure 2. Older adult using digital learning materials on the YouTube platform.



XSL•FO
RenderX

Figure 3. LINE app group for mindful practice and real-time interaction.



Qualitative Component: In-Depth Interviews

One week after completing the 8-week program, all participants were invited to take part in face-to-face, in-depth interviews to discuss their experiences with the curriculum and digital support materials. Interviews were conducted in a private room at the community care station by the course instructor, who was trained in mindfulness facilitation and qualitative interviewing. Sessions lasted approximately 30 minutes and followed a semistructured interview guide (Multimedia Appendix 2), encompassing questions regarding participants' personal experiences, any benefits they experienced, and their recommendations for improvements.

To minimize social desirability bias arising from the instructor-participant relationship, participants were assured that their responses would not affect their continued participation or access to services. They were encouraged to share both positive and negative experiences and reminded that they could decline to answer any question. With participants' consent, the interviews were audio-recorded and transcribed verbatim in Mandarin; field notes were taken during or immediately after each session. Transcripts were reviewed for accuracy prior to analysis.

Participants

Participants were recruited from a community care station in Chiayi City offering frailty prevention and disability-delay programs. Recruitment was initiated through announcements at Long-Term Care Service C Stations, followed by referrals from social workers and case managers. These stations are community centers that provide primary long-term care by

XSL•FO
RenderX

integrating disability prevention strategies, connecting preventive care with active aging programs to reduce physical decline, promoting health and well-being, and improving quality of life [1].

The eligibility criteria were as follows: (1) aged 65 years or older, (2) able to communicate in Mandarin or Taiwanese and participate in group discussions, (3) owning and regularly using a smartphone and being familiar with the LINE application, and (4) willing and able to attend all scheduled sessions during the intervention period. The exclusion criteria included (1) a prior diagnosis of cognitive impairment (eg, dementia), (2) severe hearing, speech, or communication difficulties that would prevent participation in group activities, and (3) severe or unstable physical conditions that would make participation unsafe or impractical. Ten older adults were ultimately enrolled in the pilot study.

Ethical Considerations

Each participant signed a written consent form prior to study participation. We received ethical approval for this study from the Chang Gung Medical Foundation institutional review board (number 202400087B0). Before data collection, the research team explained the study objectives, procedures, participants' rights, and measures to protect confidentiality and research data to all eligible participants. Participation was voluntary, and participants were assured they could discontinue at any stage of the study without penalty or negative impact. Original paper documents were kept in locked cabinets at the study institution. After the removal of personally identifiable information, research data were maintained in institutional cloud-based storage systems and on password-protected computers, and access was restricted to approved members of the research team. Data will be preserved until 2027 and subsequently destroyed in accordance with institutional data management requirements. A signed copy of the informed consent form was provided to each participant. To acknowledge their time and contribution, participants received a NT$100 gift voucher, approximately US $3, for each completed study-related task.

Measures

Quantitative Outcome Measures

Quantitative data were collected at all 4 time points (P0-P3) using the following validated instruments:

- Mindfulness: Taiwanese version of the Five Facet Mindfulness Questionnaire (T-FFMQ; 39 items, higher scores indicate greater mindfulness) [7,23].
- Cognitive function: Taiwanese version of the Montreal Cognitive Assessment (MoCA; score: 0 - 30, higher scores indicate better cognition) [24,25].
- Sleep quality: Chinese Pittsburgh Sleep Quality Index (PSQI; global score: 0 - 21, higher scores indicate poorer sleep) [26,27].
- Perceptions of aging: Chinese Brief Aging Perceptions Questionnaire (B-APQ) [28-30].
- Healthy aging perspectives: Healthy Aging Perspectives Questionnaire (HAPQ) [31].
- Physical activity: Chinese Physical Activity Scale for the Elderly (PASE-C) [32].
- Physical function: Short Physical Performance Battery (SPPB; score: 0 - 12) [33].
- HRV: standard deviation of normal-to-normal (SDNN) and low frequency (LF)/high frequency (HF) ratio measured with a Taiwan Food and Drug Administration–approved BEATINFO heart rate sensor and application.

We selected these measures to capture multiple domains of healthy aging, including psychological, cognitive, and physical dimensions. Multimedia Appendix 3 provides detailed descriptions of each instrument and its psychometric properties.

Data Analysis

Quantitative and Qualitative Data Analysis

Quantitative and qualitative data were analyzed separately before integration. Descriptive statistics (mean, SD, frequency, and percentage) were used to characterize participants' demographic data. Because of the small sample size and nonnormal distribution of the repeated-measures data, we employed the Friedman test for within-group comparisons across the 4 time points. For statistically significant results, post hoc pairwise comparisons were conducted to determine which time points showed meaningful changes. Statistical significance was set at $P<.05$ (2-tailed). All analyses were performed using SPSS Statistics version 25 (IBM Corporation). All 10 participants completed the 8-week intervention, all 4 quantitative assessments, and the postprogram interviews. Given no missing outcome data, analyses were conducted using complete cases (n=10) without the need for imputation.

Qualitative Analysis

We conducted thematic analysis following Braun and Clarke's 6-phase framework (familiarization, coding, theme development, review, definition, and reporting) [34]. Two researchers individually coded the transcripts and resolved discrepancies through discussion. Codes were organized into subthemes and overarching themes reflecting participants' experiences with the curriculum and digital practice materials. We applied investigator triangulation and peer debriefing methods, maintaining a coding decision audit trail to enhance trustworthiness. The selected quotes effectively represent the themes and use pseudonyms to protect participants' identities.

Integration of Quantitative and Qualitative Findings

At the interpretation stage, quantitative and qualitative findings were merged to develop a comprehensive understanding of the feasibility, acceptability, and preliminary effects of the mindfulness curriculum.

Results

Phase 1: Delphi Consensus

The modified Delphi process yielded strong expert consensus on the 8-week mindfulness curriculum (Tables 1 and 2). Qualitative feedback from round 1 was synthesized and used to revise items in subsequent rounds. In rounds 2 and 3, all items received strong consensus, with median scores of 4 and IQRs ≤1.0 across sessions. By round 3, agreement remained high, with at least 88% of experts rating each session as "suitable" or "very suitable" (Table 2). These results supported the

XSL•FO
RenderX

curriculum's content validity and expert acceptability, including its digital learning materials, and justified advancing to pilot feasibility testing.

XSL•FO
RenderX

Table . Eight-week mindfulness curriculum for older adults.

Week	Theme	Course outline	Formal practice	Digital learning materials
Introduction	Orientation	Course introduction, mindfulness breathing, awareness exercises, group building, and instruction on using digital course videos (LINE/YouTube)	Mindfulness breathing	—[a]
1	Mindfulness awareness experience	Mindful breathing, intention meditation, body scan, 5-sense awareness, and mindful stretching	Mindful breathing, body scan, mindful stretching, and yoga	Mindful breathing and body scan
2	Perception and creative response	Mindful breathing, sharing of home practice, body scan, mindful stretching, mindful eating, and 9-point practice	Body scan, mindful stretching, yoga, and mindful eating	Mindful breathing, body scan, mindful stretching (standing), and mindful eating
3	Experiencing the joy and strength of the present moment	Mindful breathing, mindful stretching, sharing of home practice, “awareness triangle” for appreciating good things, mindful eating and drinking	Breathing observation, mindful stretching, and mindful eating and drinking	Mindful breathing, body scan, mindful eating, and mindful stretching (lying down)
4	Habits and perspectives: how interpretations shape experience	Mindful breathing, mindful stretching, sharing of home practice, application of mindfulness in daily life, recognizing thoughts, sharing pleasant events, in-class exchange, and loving-kindness meditation	Mindful breathing, mindful stretching, yoga, and loving-kindness meditation	Mindful breathing, body scan, and loving-kindness meditation
5	Discovering space and possibilities in choice	Mindful breathing, mindful stretching, non-judgmental awareness practice, midterm review and sharing, awareness of unpleasant events and mind-body connection, mindful walking, in-class exchange, and loving-kindness meditation	Mindful breathing, mindful walking, and loving-kindness meditation	Mindful breathing, body scan, mindful walking, and loving-kindness meditation
6	Stress and automatic pilot (STOP)	Mindful breathing, mindful stretching, sharing of home practice and daily application of mindfulness, mindful eating, understanding stress and stressors, body scan, and STOP practice (stop, take a breath, observe, and proceed), and loving-kindness meditation	Mindful breathing, body scan, STOP practice, and loving-kindness meditation	Mindful breathing, body scan, mindful eating, and mindful stretching
7	Integrating mindfulness into daily life	Mindful breathing, mindful stretching, home-practice sharing, mindful eating, mindful listening and communication, in-class exchange, mindful walking and empathy practice, and loving-kindness meditation	Breathing observation, mindful walking, mindful listening, and loving-kindness meditation	Mindful breathing, body scan, mindful eating, mindful stretching, and mindful listening and communication

XSL•FO
RenderX

Week	Theme	Course outline	Formal practice	Digital learning materials
8	A new life with mindfulness	Mindful breathing, mindful stretching, sharing of home practice, body scan, group sharing, application of mindfulness in daily life, loving-kindness meditation, and graduation ceremony	Mindful breathing, body scan, mindful stretching, yoga, and loving-kindness meditation	Mindful breathing, body scan, mindful eating, and mindful life

[a]Not applicable.

Table . Delphi results from rounds 2 and 3.

Session[a]	Round 2		Round 3	
	Median (IQR)	Agreement (%)	Median (IQR)	Agreement (%)
Introduction	4 (3-4)	88.33	4 (3.25-4)	88.75
Session 1	4 (4-4)	92.59	4 (4-4)	94.32
Session 2	4 (4-4)	95.15	4 (4-4)	91.67
Session 3	4 (4-4)	96.67	4 (4-4)	95.31
Session 4	4 (4-4)	96.67	4 (4-4)	97.60
Session 5	4 (4-4)	98.79	4 (4-4)	99.11
Session 6	4 (4-4)	96.11	4 (4-4)	97.60
Session 7	4 (4-4)	97.04	4 (4-4)	98.75
Session 8	4 (4-4)	94.55	4 (4-4)	99.43

[a]Items were rated on an ordinal scale; the results are presented as median (IQR) and percentage agreement (rated as “suitable” or “very suitable”). Consensus was defined a priori as IQR≤1.0 and agreement ≥80%.

Phase 2: Pilot Study

Recruitment, Retention, and Missing Data

Ten older adults participated in the pilot study. All completed the 8-week mindfulness program, the quantitative assessments at all 4 time points (P0-P3), and the postprogram qualitative interview. There was no attrition, and quantitative analyses were conducted using complete cases (n=10).

Participant Characteristics

All 10 participants were community-dwelling older women with an average age of 77.4 (SD 6.2) years. The majority were retired and had attained a junior high school education or below (n=7, 70%), with monthly incomes under NT $20,000 (US $645; n=9, 90%). Most had experienced widowhood (n=7, 70%) and practiced Buddhism while living with either children or spouses (n=7, 70%). They had multiple chronic health conditions, including musculoskeletal disorders (n=5, 50%), hypertension (n=4, 40%), diabetes (n=3, 30%), and cardiovascular diseases (n=3, 30%). All participants maintained complete independence in their activities of daily living (mean 100, SD 0; Table 3).

XSL•FO
RenderX

Table . Participant characteristics (N=10).

Characteristic and item	Value
Age (y), mean (SD)	77.4 (6.20)
Gender	
Female	10 (100)
BMI, mean (SD)	26.25 (3.82)
ADL[a], mean (SD)	100 (0)
Education level, n (%)	
Junior high school and below	7 (70)
Senior school and above	3 (30)
Currently employed, n (%)	
Yes	1 (10)
No	9 (90)
Religion, n (%)	
No	1 (10)
Yes	9 (90)
Marital status, n (%)	
Married	2 (20)
Single/widower/divorced	8 (80)
Monthly income, n (%)	
NT $20,000 (US $645) and below (low income)	9 (90)
NT $20,000 (US $645) and above	1 (10)
Chronic conditions, n (%)	
Hypertension	4 (40)
Diabetes	3 (30)
Hyperlipidemia	2 (20)
Cardiovascular disease	3 (30)
Musculoskeletal disorders	5 (50)
Other	2 (20)
Living alone, n (%)	3 (30)
Yes	3 (30)
No	7 (70)

[a]ADL: activities of daily living.

Quantitative Outcomes

Table 4 presents medians and IQRs for all outcomes across the 4 time points (P0-P3). Friedman tests revealed statistically significant changes over time in mindfulness (FFMQ), sleep quality (PSQI), aging perceptions (B-APQ), healthy aging perspectives (HAPQ), and physical function (SPPB). No significant changes were observed in cognitive function (MoCA), physical activity (PASE-C), or HRV indices (SDNN and LF/HF ratio).

Mindfulness levels also increased over time. Median FFMQ scores rose from 139.50 (IQR 102.25-155.75) at baseline to 150.00 (IQR 141.75-156.25) at week 8 and 155.00 (IQR 136-164.25) at follow-up (χ^2_3=11.57, *P*=.009). Post hoc comparisons indicated significantly higher mindfulness at follow-up than at baseline (P3>P0), suggesting sustained improvement in mindfulness skills after the intervention.

XSL•FO
RenderX

Table . Changes in healthy aging outcomes across 4 time points.

Variable[a]	Baseline (P0) median (IQR)	Week 4 (P1) median (IQR)	Week 8 (P2) median (IQR)	Follow-up (P3) median (IQR)	χ^2 (df)	*P* value	Post hoc
HRV[b]- SDNN[c]	33.15 (21.08-83.42)	25.70 (17.46-73.57)	156.84 (36.10-215.47)	28.78 (19.63-120.74)	1.20 (3)	.75	—[d]
HRV-LF/HF ratio[e]	0.93 (0.54-2.02)	2.05 (1.11-4.63)	2.40 (1.21-4.18)	0.88 (0.69-3.99)	6.60 (3)	.09	—
FFMQ[f]	139.50 (102.25-155.75)	135.00 (124.5-164)	150.00 (141.75-156.25)	155.00 (136-164.25)	11.57 (3)	.009	P3>P0
MoCA[g]	29.00 (26.75-30.00)	29.50 (28.50-30.00)	30.00 (28.50-30.00)	30.00 (28.75-30.00)	3.98 (3)	.26	—
PSQI[h]	8.50 (2.75-11.25)	9.00 (2.75-10.25)	3.50 (1.00-6.25)	2.50 (1.00-4.50)	15.03 (3)	.002	P3>P0, P2>P0
B-APQ[i]	3.25 (2.96-4.06)	3.63 (3.03-4.53)	3.58 (3.42-4.22)	4.23 (3.57-4.80)	10.88 (3)	.01	P3>P0
HAPQ[j]	3.14 (3.01-3.22)	3.68 (3.30-4.23)	3.78 (3.52-3.94)	4.13 (4.09-4.69)	20.70 (3)	<.001	P3>P0, P2>P0
PASE-C[k]	105.16 (89.67-150.36)	91.36 (61.25-165.01)	67.02 (57.35-139.94)	118.52 (73.19-127.96)	2.03 (3)	.57	—
SPPB[l]	12.00 (11.50-12.00)	11.50 (11.00-12.00)	12.00 (11.75-12.00)	12.00 (12.00-12.00)	8.37 (3)	.04	—

[a]PSQI: lower scores indicate better sleep quality. FFMQ, B-APQ, HAPQ, and SPPB: higher scores indicate more favorable outcomes.
[b]HRV: heart rate variability.
[c]SDNN: standard deviations of normal-to-normal.
[d]Not applicable.
[e]LF/HF ratio: low frequency/high frequency ratio.
[f]FFMQ:Five Facet Mindfulness Questionnaire.
[g]MoCA: Montreal Cognitive Assessment.
[h]PSQI: Pittsburgh Sleep Quality Index.
[i]B-APQ: Brief Aging Perceptions Questionnaire.
[j]HAPQ**:** Healthy Aging Perspectives Questionnaire.
[k]PASE-C: Physical Activity Scale for the Elderly.
[l]SPPB: Short Physical Performance Battery.

Sleep quality improved markedly: median PSQI scores decreased from 8.50 (IQR 2.75-11.25) at baseline to 3.50 (IQR 1.00-6.25) at week 8 and 2.50 (IQR 1.00-4.50) at follow-up (χ^2_3=15.03, *P*=.002). Post hoc comparisons showed significantly lower scores at P2 and P3 than at P0, indicating progressively better sleep.

Perceptions of aging and healthy aging perspectives also became more positive. Median B-APQ scores increased from 3.25 (IQR 2.96-4.06) at baseline to 4.23 (IQR 3.57-4.80) at follow-up (χ^2_3=10.88, *P*=.01), with post hoc tests indicating higher scores at P3 than P0. Median HAPQ scores increased from 3.14 (IQR 3.01-3.22) at baseline to 3.78 (IQR 3.52-3.94) at week 8 and 4.13 (IQR 4.09-4.69) at follow-up (χ^2_3=20.70, *P*<.001), with significantly higher scores at P2 and P3 compared with P0.

Physical function remained high and showed small but statistically significant variation over time: median SPPB scores were 12.00 (IQR 11.50-12.00) at baseline, dipped slightly to 11.50 (IQR 11.00-12.00) at week 4, and returned to 12.00 (IQR 11.75-12.00) at week 8 and to 12.00 (IQR 12.00-12.00) at week 12 (χ^2_3=8.37, *P*=.04).

By contrast, cognitive function (MoCA) remained consistently high (≈29.0) at all time points, and median PASE-C and HRV indices (SDNN and LF/HF ratio) fluctuated without a clear trend. None of these variables changed significantly over time (all *P*>.05).

Qualitative Outcomes

Thematic analysis of the postprogram interviews revealed 5 main themes, each with several subthemes, describing the physical and psychological sensations experienced after participating in the digitally supported mindfulness program (Table 5):

- Enhanced physical and psychological well-being: participants reported greater relaxation, better sleep, and improved emotional regulation.
- Transformation of daily routines: mindfulness practices gradually shifted from occasional use to habitual strategies for coping with discomfort and stress in everyday life.
- Memorable practices and preferred activities: simple, accessible practices (eg, deep breathing, bed-based stretching, and mindful eating) were viewed as most useful and sustainable.

XSL•FO
RenderX

- Positive life changes and social connectedness: participants described adopting healthier behaviors, shortening periods of negative emotions, and experiencing reduced loneliness through group interactions.
- Program evaluation and suggestions: the overall satisfaction was high. Digital tools (LINE and YouTube) were perceived as helpful, although some participants with reading difficulties needed family assistance to access online content and requested continued classes.

Taken together, the qualitative findings indicate that participants experienced meaningful benefits from the digitally supported mindfulness program, including improved relaxation, emotional regulation, and social connectedness. Accessible practices centered on breathing and yoga were most consistently adopted into daily life. Although digital tools were widely valued for supporting home practice, a subset of participants reported accessibility challenges, highlighting the need for age-friendly digital solutions.

XSL•FO
RenderX

Table . Themes, subthemes, and illustrative quotes from qualitative interviews.

Theme and subtheme	Illustrative quote
Enhanced physical and psychological well-being	
Improved relaxation and sleep	• *Sometimes when I cannot fall asleep, I just take deep breaths and stretch my legs in bed. Soon, I feel more comfortable and can fall asleep.* (participant C) • *Prior to participating in this course, I experienced significant difficulties with sleep; however, following the practice sessions, my sleep quality has shown gradual improvement.* (participant H)
Emotional stabilization and stress relief	• *When I am busy and feel upset, I just take a few deep breaths. It really helps me relax, and my mood improves.* (participant A)
Transformation of daily routines	
From passive learning to habitual practice	• *At first, I often forgot. But when I saw the YouTube videos in the LINE group, I remembered and did the exercises at home.* (participant F)
From reactive to proactive coping	• *When my chest hurt, I first thought about doing deep breathing. I tried hard to breathe in, and after a while, I felt better.* (participant B)
Memorable practices and preferred activities	
Breathing as a core practice	• *Deep breathing is the most useful. I can do it anywhere, and it calms me quickly.* (participant H)
Bed-based and stretching exercises	• *I like the bed exercises the most. Even when it's cold, I can do them under a blanket. They really loosen my body.* (participant D) • *During my stretching exercises, I have started to focus closely on the sensations in my body with each movement, especially to pinpoint the exact areas where I feel exertion.* (participant E)
Mindful eating and slow chewing	• *Now I chew more slowly and carefully. It helps me taste the food and prevents me from choking.*"(participant G)
Positive life changes and social connectedness	
Adoption of healthier behaviors	• When I experience a negative mood, I practice mindful walking in the park to help alleviate it. (participant F) • *I eat slower, drink more water, and keep moving every day. It's a real change for me.* (participant E)
Improved emotional regulation	• *When bad things happen, I shorten the time I feel sad or angry. I try not to hold onto it for too long.* (participant D)
Reduced loneliness and enhanced social ties	• *Coming here and practicing with others makes me feel less lonely. At home, I would think too much.* (participant H)
Program evaluation and suggestions	
High satisfaction and appreciation	• *I really enjoy these classes. The teacher's voice is very soothing.* (participant G)
Role of digital technology	• *I don't read well, so I rarely check the YouTube videos myself. My daughter helps me open them.* (participant I)
Desire for continuity	• *I hope you can come again regularly. If we keep practicing together, we won't forget.* (participant A)

XSL•FO
RenderX

Discussion

Principal Findings

In this mixed methods feasibility study, we developed and piloted an 8-week digitally supported mindfulness curriculum integrating MBSR, MBEC, and MSA frameworks for community-dwelling older adults in Taiwan. Assessments at 4 time points revealed significant improvements in mindfulness and sleep quality, along with more positive attitudes toward aging. However, physical function showed only a slight (but significant) improvement, and no significant changes were observed in cognitive function, physical activity, or HRV indices. The quantitative data were supported by the qualitative interview results, which indicated that participants developed more effective relaxation and emotional regulation techniques and became more capable in daily activities and interpersonal relationships. Overall, the findings indicated that the program was successful, with participants providing positive feedback on both culturally adapted content and digital support components.

The program integrated MBSR with MBEC and MSA frameworks while adapting content to the cultural and educational contexts of older adults in Taiwan. Mindfulness interventions offered via structured courses and digital platforms require careful cultural adaptation [35], and research highlights the importance of incorporating cultural values, language, social interaction styles, and accessible technology to enhance participant engagement and overall effectiveness [36-38]. In this study, we used 3 strategies to address reported barriers related to low literacy and abstract mindfulness concepts: simplified language, locally relevant exercises, and embodied practices. Participants described the LINE reminders and YouTube videos as useful digital tools that facilitated between-session practice. However, some participants reported the need for family assistance to access and use the materials due to limited reading ability. These findings highlight both the potential and practical challenges of digital support for older adults and underscore the importance of age-friendly, accessible interface design.

The observed improvement in sleep quality is consistent with participants' descriptions of using breathing techniques and gentle stretching to manage sleeplessness. By reducing presleep cognitive arousal and facilitating relaxation, these practices likely contributed to the marked decline in PSQI scores. This finding aligns with previous research showing that mindfulness training improves sleep quality by supporting emotional regulation and reducing stress-related hyperarousal symptoms, which can affect older adults [39].

Improvements in aging perceptions and healthy aging perspectives reflect the cognitive reappraisal and acceptance techniques emphasized by both the MBEC and MSA approaches. Mindfulness practices enable individuals to distance themselves from unpleasant thoughts, cultivate self-compassion, and recognize meaningful moments throughout their day. Participants reported experiencing shorter episodes of negative emotions as they learned to manage difficult situations, which align with an adaptive aging orientation. This transition may have significant health implications, as individuals who believe they are aging well tend to preserve their physical abilities and remain active in later life [2,4].

Factors Contributing to Unchanging Variables

Median SPPB scores remained high throughout and showed a small but statistically significant improvement, whereas physical activity levels assessed by the PASE-C did not change significantly. This outcome is consistent with previous studies demonstrating that mindfulness training produces variable improvements in physical movement quality, balance, and mobility [14,40,41]. The curriculum focused on teaching safety and awareness during light exercise rather than increasing overall exercise duration; consequently, changes in short mindful movement sessions—such as bed-based stretching and mindful walking—are unlikely to be captured by the PASE-C, which primarily measures the frequency of leisure activities and household tasks. Although mindfulness training does not aim to increase activity duration, it can facilitate improvement in movement quality and physical awareness during these exercises.

Despite participants' numerous positive subjective reports, no statistically significant improvement was observed in cognitive function as measured by MoCA. This finding is likely attributable to the ceiling effect. Participants showed normal cognitive abilities at baseline, as their MoCA scores were already near maximum and concentrated within a narrow range. The MoCA test fails to detect small changes in executive function and attention that mindfulness may produce in cognitively healthy older adults under specific conditions. Future research should include a more diverse group of participants who remain under observation for an extended period while using various cognitive tests to measure the effectiveness of interventions [12].

The HRV indices (SDNN and LF/HF ratio) did not show significant variations. However, HRV can fluctuate widely owing to factors such as the testing environment, breathing patterns, body position, measurement timing, and concurrent medications. Meta-analytic research indicates that mindfulness practices have small and variable effects on HRV, which are influenced by vagal nerve function [42,43]. Future studies should aim to determine when physiological changes occur by employing larger sample sizes, longer observation periods, and standardized resting-state measurement protocols.

Limitations

Several study limitations should be noted. First, the small sample size (N=10) and single-site recruitment reduce the statistical power and generalizability of the findings. Second, since the majority of older adults at the Long-Term Care Service C Stations are female, all participants were women, restricting the applicability of findings to older men [44]. Third, the PASE-C instrument lacked the sensitivity to detect changes in mindful light-intensity movements, which likely contributed to the null finding in physical activity levels. Finally, although digital tools supported program engagement, several participants with limited literacy required family assistance to access online materials,

XSL•FO
RenderX

emphasizing the need to create digital solutions that are both age-friendly and inclusive.

Links With Prior Evidence, Implications, and Future Directions

Our findings support previous studies showing that mindfulness programs benefit older adults' sleep quality and mental health [12] by demonstrating that similar gains are achievable within a digitally supported, community-based format that also enhances aging-related perceptions and functional outcomes. This hybrid learning model—combining weekly in-person sessions with LINE reminders and instructor-recorded videos—offers a potentially cost-effective approach to sustaining home practice and maintaining engagement in real-world settings, addressing common challenges to program adherence [13].

Based on this pilot study, findings should be interpreted as preliminary evidence of feasibility and perceived usefulness rather than as a basis for immediate large-scale implementation. The program structure may nonetheless inform future evaluations at Long-Term Care Service C Stations and other community-based aging services, serving as a potential model for incorporating brief mindfulness practices—such as mindful breathing, chair-based stretching, and brief body scans—into regular group activities. A train-the-trainer approach supported by a structured manual (including session scripts, safety cues, and culturally relevant examples) may facilitate standardized delivery across sites. Importantly, suggestions regarding age-friendly digital interfaces (eg, larger text, simplified navigation, structured screens, and short audio/video materials delivered via familiar platforms) were derived from participants' reports of accessibility barriers (eg, limited reading ability and reliance on family assistance) rather than from formal usability testing; therefore, these design implications should be validated in future work using established usability evaluation methods.

Future research should assess program effectiveness and scalability using adequately powered randomized controlled trials with longer follow-up, objective activity monitoring, more sensitive cognitive measures, standardized protocols, and hybrid effectiveness-implementation designs that assess both outcomes and adoption metrics (eg, reach, fidelity, and cost). Formal usability evaluation (eg, task-based testing and validated usability measures) and objective engagement metrics should also be incorporated to optimize the digital support components.

Conclusion

This study makes a dual contribution: it demonstrates expert consensus on the content validity of a culturally adapted, theory-grounded mindfulness curriculum for community-dwelling older adults in Taiwan, and it provides preliminary evidence of the program's feasibility, acceptability, and beneficial effects across psychological, sleep-related, and aging perception domains. The hybrid delivery format was generally well-received and appeared to support home practice. Importantly, this work extends existing research by showing that mindfulness benefits are not only confined to sleep and mental health but also extend to how older adults perceive and engage with aging. Additional trials are needed to confirm these preliminary effects, refine assessment methods, and establish conditions for scalability.

Acknowledgments

The authors gratefully acknowledge the participation of older adults and the support of staff from the Long-Term Care Service C Stations in Chiayi City, whose contributions made this study possible.

Funding

This research was funded by the National Science and Technology Council, Taiwan (grant number 113-2410-H-255-001-MY1).

Data Availability

The data supporting this study's findings are available from the corresponding author upon reasonable request. Access to the datasets is subject to ethical considerations and requires appropriate approval.

Authors' Contributions

YRW (lead), MLH (equal)
Data curation: YRW
Formal analysis: PLH (lead), CHC (supporting)
Funding acquisition: YRW
Investigation: YRW (lead), MLH (equal), PLH (supporting)
Methodology: YRW (lead), MLH (equal)
Project administration: YRW (lead), MLH (equal), CCH (supporting)
Resources: CCH
Supervision: MLH
Validation: MLH (lead), PLH (equal)
Visualization: CHC (lead), CCH (supporting)
Writing—original draft: YRW (lead), MLH (supporting)
Writing—review and editing: YRW (lead), PLH (equal), MLH (supporting)

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Expert characteristics.
[DOCX File, 18 KB - humanfactors_v13i1e84161_app1.docx]

Multimedia Appendix 2
Semistructured interview guide.
[DOCX File, 21 KB - humanfactors_v13i1e84161_app2.docx]

Multimedia Appendix 3
Research measurements.
[DOCX File, 18 KB - humanfactors_v13i1e84161_app3.docx]

References

1. Ministry of Health and Welfare. Long-term care plan 2.0. URL: https://english.ey.gov.tw/News3/9E5540D592A5FECD/332a78c0-0c8e-4064-bd71-8c22477dae75 [accessed 2026-04-15]
2. World health organization. UN decade of healthy ageing: plan of action 2021–2030. 2020. URL: https://cdn.who.int/media/docs/default-source/decade-of-healthy-ageing/decade-proposal-final-apr2020-en.pdf?sfvrsn=b4b75ebc_28
3. Ministry of Health and Welfare. World Health Organization's new version - Integrated Care for Older People. 2020. URL: https://www.mohw.gov.tw/cp-4631-54550-1.html [accessed 2026-04-15]
4. Menassa M, Stronks K, Khatmi F, et al. Concepts and definitions of healthy ageing: a systematic review and synthesis of theoretical models. EClinicalMedicine 2023 Feb;56:101821. [doi: 10.1016/j.eclinm.2022.101821] [Medline: 36684393]
5. Wang YR, Lee HF, Hsieh PL, Chang CH, Chen CM. Relationship between physical activity and perceptions of ageing from the perspective of healthy ageing among older people with frailty with chronic disease: a cross-sectional study. BMC Nurs 2023 Sep 16;22(1):319. [doi: 10.1186/s12912-023-01481-9] [Medline: 37716946]
6. Kabat-Zinn J. Mindfulness-based interventions in context: past, present, and future. Clin Psychol Sci Pract 2003;10(2):144-156. [doi: 10.1093/clipsy.bpg016]
7. Baer RA, Smith GT, Hopkins J, Krietemeyer J, Toney L. Using self-report assessment methods to explore facets of mindfulness. Assessment 2006 Mar;13(1):27-45. [doi: 10.1177/1073191105283504] [Medline: 16443717]
8. Didonna F. Clinical Handbook of Mindfulness: Springer; 2009.
9. Kabat-Zinn J. Full Catastrophe Living: How to Cope with Stress, Pain and Illness Using Mindfulness Meditation: Piatkus; 2013. URL: https://archive.org/details/fullcatastrophel0000kaba_w6g8 [accessed 2026-04-15]
10. Santorelli SF, Kabat-Zinn J, Blacker M, Meleo-Meyer F, Koerbel L. Mindfulness-based stress reduction (MBSR) authorized curriculum guide. : Center for Mindfulness in Medicine, Health Care, and Society (CFM), University of Massachusetts Medical School; 2017 URL: https://www.rightdecisions.scot.nhs.uk/media/2105/mbsr-curriculum-guide-2017.pdf [accessed 2026-04-15]
11. Sun Y, Liu F, Wang Y, Zhou M, Yang B, Cao F. Mindfulness improves health worker's occupational burnout: the moderating effects of anxiety and depression. Int Arch Occup Environ Health 2021 Aug;94(6):1297-1305. [doi: 10.1007/s00420-021-01685-z] [Medline: 33787983]
12. Lannon-Boran C, Hannigan C, Power JM, Lambert J, Kelly M. The effect of mindfulness-based intervention on cognitively unimpaired older adults' cognitive function and sleep quality: a systematic review and meta-analysis. Aging Ment Health 2024;28(1):23-35. [doi: 10.1080/13607863.2023.2228255] [Medline: 37485984]
13. Denkova E, Barry J, Zanesco AP, Rooks J, Rogers SL, Jha AP. Online mindfulness training for older adults during the COVID-19 pandemic: a randomized controlled trial using a multi-method assessment approach. Aging Ment Health 2024;28(1):130-141. [doi: 10.1080/13607863.2023.2242301] [Medline: 37702364]
14. Weber M, Schnorr T, Morat M, Morat T, Donath L. Effects of mind-body interventions involving meditative movements on quality of life, depressive symptoms, fear of falling and sleep quality in older adults: a systematic review with meta-analysis. Int J Environ Res Public Health 2020 Sep 9;17(18):6556. [doi: 10.3390/ijerph17186556] [Medline: 32916879]
15. Bee LM. Mindfulness-based elder care: communicating mindfulness to frail elders and their caregivers. In: Didonna F, editor. Clinical Handbook of Mindfulness: Springer; 2009:431-445. [doi: 10.1007/978-0-387-09593-6_24]
16. McBee L. I am sure to grow old: mindfulness-based elder care. In: Baer RA, editor. Mindfulness-Based Treatment Approaches, 2nd edition: Academic Press; 2014:239-265. [doi: 10.1016/B978-0-12-416031-6.00011-6]
17. Tsai CC, Lee HL, Wu CS, Chen PY, Chen TW, Chen MF. The efficacy of a mindfulness-based exercise program in older residents of a long-term care facility in Taiwan. Geriatr Nurs 2023;50:227-233. [doi: 10.1016/j.gerinurse.2023.01.021] [Medline: 36805954]

XSL•FO
RenderX

18. Nilsson H. The four-dimensional model of mindful sustainable aging: a holistic alternative. J Relig Spiritual Aging 2023 Oct 2;35(4):430-454. [doi: 10.1080/15528030.2022.2083743]
19. Nilsson H, Bülow PH, Kazemi A. Mindful sustainable aging: advancing a comprehensive approach to the challenges and opportunities of old age. Eur J Psychol 2015 Aug;11(3):494-508. [doi: 10.5964/ejop.v11i3.949] [Medline: 27247673]
20. Chems-Maarif R, Cavanagh K, Baer R, Gu J, Strauss C. Defining mindfulness: a review of existing definitions and suggested refinements. Mindfulness 2025 Jan;16(1):1-20. [doi: 10.1007/s12671-024-02507-2]
21. Social and Family Affairs Administration. Strategies to address the challenges of a super-aging society. 2022. URL: https://www.sfaa.gov.tw/sfaa/page/5fQ [accessed 2026-04-15]
22. Jünger S, Payne SA, Brine J, Radbruch L, Brearley SG. Guidance on conducting and reporting Delphi studies (CREDES) in palliative care: recommendations based on a methodological systematic review. Palliat Med 2017 Sep;31(8):684-706. [doi: 10.1177/0269216317690685] [Medline: 28190381]
23. Huang FY, Wu CW, Bhikshu HM, Shih GH, Chao YP, Dai CT. Validation of the Taiwanese version of the Five Facet Mindfulness Questionnaire (T-FFMQ) [Article in Chinese]. Psychol Test 2015;62(3):231-260. [doi: 10.7108/PT.201509_62(3).0003]
24. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
25. Tsai CF, Lee WJ, Wang SJ, Shia BC, Nasreddine Z, Fuh JL. Psychometrics of the Montreal Cognitive Assessment (MoCA) and its subscales: validation of the Taiwanese version of the MoCA and an item response theory analysis. Int Psychogeriatr 2012 Apr;24(4):651-658. [doi: 10.1017/S1041610211002298] [Medline: 22152127]
26. Buysse DJ, Reynolds CF III, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
27. Tsai PS, Wang SY, Wang MY, et al. Psychometric evaluation of the Chinese version of the Pittsburgh Sleep Quality Index (CPSQI) in primary insomnia and control subjects. Qual Life Res 2005 Oct;14(8):1943-1952. [doi: 10.1007/s11136-005-4346-x] [Medline: 16155782]
28. Barker M, O'Hanlon A, McGee HM, Hickey A, Conroy RM. Cross-sectional validation of the Aging Perceptions Questionnaire: a multidimensional instrument for assessing self-perceptions of aging. BMC Geriatr 2007 Apr 26;7(9):9. [doi: 10.1186/1471-2318-7-9] [Medline: 17462094]
29. Sexton E, King-Kallimanis BL, Morgan K, McGee H. Development of the Brief Ageing Perceptions Questionnaire (B-APQ): a confirmatory factor analysis approach to item reduction. BMC Geriatr 2014 Apr 9;14:44. [doi: 10.1186/1471-2318-14-44] [Medline: 24716631]
30. Wang YR, Lee HF, Chen CM. Validating a Brief Aging Perception Questionnaire (B-APQ) for older persons with chronic disease in Taiwan. Aging Ment Health 2021 Jun;25(6):1143-1150. [doi: 10.1080/13607863.2020.1734914] [Medline: 32133880]
31. Wang YR, Lee HF, Hsieh PL, Chen CM. Development of the Healthy Aging Perspectives Questionnaire among Older adults with chronic disease in Taiwan. Health Soc Care Community 2023 Apr 18:8238748. [doi: 10.1155/2023/8238748]
32. Ku PW, Sun WJ, Chang CY, Chen LJ. Reliability and validity of the Chinese version of the Physical Activity Scale for the Elderly [Article in Chinese]. SER Rev Med Soc 2013 Sep 30;15(3):309-319. [doi: 10.5297/ser.1503.006]
33. Guralnik JM, Simonsick EM, Ferrucci L, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol 1994 Mar;49(2):M85-M94. [doi: 10.1093/geronj/49.2.m85] [Medline: 8126356]
34. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
35. Naderbagi A, Loblay V, Zahed IUM, et al. Cultural and contextual adaptation of digital health interventions: narrative review. J Med Internet Res 2024 Jul 9;26:e55130. [doi: 10.2196/55130] [Medline: 38980719]
36. Haidar A, Kalantzis M, Nallajerla M, Vela A, Adler SR, Burnett-Zeigler I. A qualitative examination of perceptions of mindfulness among Arab, Middle Eastern, and North African Americans: implications for cultural adaptation of mindfulness-based interventions. Glob Adv Integr Med Health 2024;13:27536130241247074. [doi: 10.1177/27536130241247074] [Medline: 38655072]
37. McCarren L, Kuoppamäki S. Design preferences, routines, and well-being of older adults using voice-guided digital mindfulness: qualitative interview study. JMIR Hum Factors 2025 Jun 3;12:e67533. [doi: 10.2196/67533] [Medline: 40460434]
38. Vasudevan MK, Besevli C, Madapura Nagaraj P, Zhong S, Muniyandi M, Obrist M. Co-producing mindful tactile technologies for healthy ageing: cross-cultural insights from India and the UK. Digit Health 2025;11:20552076251349909. [doi: 10.1177/20552076251349909] [Medline: 40534893]
39. Black DS, O'Reilly GA, Olmstead R, Breen EC, Irwin MR. Mindfulness meditation and improvement in sleep quality and daytime impairment among older adults with sleep disturbances: a randomized clinical trial. JAMA Intern Med 2015 Apr;175(4):494-501. [doi: 10.1001/jamainternmed.2014.8081] [Medline: 25686304]

XSL•FO
RenderX

40. Lenze EJ, Voegtle M, Miller JP, et al. Effects of mindfulness training and exercise on cognitive function in older adults: a randomized clinical trial. JAMA 2022 Dec 13;328(22):2218-2229. [doi: 10.1001/jama.2022.21680] [Medline: 36511926]
41. Tang YY, Fan Y, Lu Q, et al. Long-term physical exercise and mindfulness practice in an aging population. Front Psychol 2020;11:358. [doi: 10.3389/fpsyg.2020.00358] [Medline: 32300317]
42. Brown L, Rando AA, Eichel K, et al. The effects of mindfulness and meditation on vagally mediated heart rate variability: a meta-analysis. Psychosom Med 2021;83(6):631-640. [doi: 10.1097/PSY.0000000000000900] [Medline: 33395216]
43. Guendelman S, Lutz M, Koenig J, Bayer M, Dziobek I. Effects of a mindfulness intervention on emotion differentiation and heart rate variability. Front Hum Neurosci 2025;19:1515334. [doi: 10.3389/fnhum.2025.1515334] [Medline: 40672745]
44. Social and Family Affairs Administration. 2021 project on analyzing demand, resource integration, and service utilization in elderly welfare. URL: https://www.sfaa.gov.tw/sfaa/list/detail/5eZ/5PA

Abbreviations

B-APQ: Brief Aging Perceptions Questionnaire
FFMQ: Five Facet Mindfulness Questionnaire
HAPQ: Healthy Aging Perspectives Questionnaire
HF: high frequency
HRV: heart rate variability
LF: low frequency
MBEC: mindfulness-based elder care
MBI: mindfulness-based intervention
MBSR: mindfulness-based stress reduction
MoCA: Montreal Cognitive Assessment
MSA: mindful sustainable aging
PASE-C: Physical Activity Scale for the Elderly
PSQI: Pittsburgh Sleep Quality Index
SDNN: standard deviation of normal-to-normal
SPPB: Short Physical Performance Battery
WHO: World Health Organization

Edited by A Kushniruk; submitted 20.Sep.2025; peer-reviewed by R Singh, YL Lee; revised version received 12.Mar.2026; accepted 25.Mar.2026; published 13.May.2026.

Please cite as:
Wang YR, Hsieh PL, Chang CH, Hsiao CC, Hu ML
Using a Digital-Based Mindfulness Curriculum to Enhance Healthy Aging Outcomes in Community-Dwelling Older Adults in Taiwan: Mixed Methods Feasibility Study
JMIR Hum Factors 2026;13:e84161
URL: https://humanfactors.jmir.org/2026/1/e84161
doi:10.2196/84161

© Yu-Rung Wang, Pei-Lun Hsieh, Chia-Hsiu Chang, Chia-Chi Hsiao, Mei-Lien Hu. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 13.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Adaptation and Acceptability of a Low-Intensity Cognitive Behavioral Therapy App to Support Low Mood and Worry Management in Female Forces Veterans: Mixed Methods Study

Paul Farrand[1], PhD; Andy Bacon[2], PGCert, PSC; Melika Janbakhsh[1], PhD; Natalie Flay[1], MSc; Elizabeth Turnbull[1], MSc; Jonathan Baker[3], MA

[1]Clinical Education, Development and Research (CEDAR), Psychology, Faculty of Health and Life Sciences, University of Exeter, Perry Road, Exeter, United Kingdom

[2]Westminster Centre for Research in Veterans, University of Chester, Cheshire, United Kingdom

[3]Iona Mind Inc, Wilmington, DE, United States

Corresponding Author:
Paul Farrand, PhD
Clinical Education, Development and Research (CEDAR), Psychology, Faculty of Health and Life Sciences, University of Exeter, Perry Road, Exeter, United Kingdom

Abstract

Background: Mental health help-seeking barriers experienced by female forces veterans result in them being underserved and underrepresented. Efforts are therefore required to adapt interventions for female veterans to enhance acceptability and maximize engagement. Given a smaller number and wider geographical distribution of female veterans, targeting adaptation efforts at a digital mobile phone app based on cognitive behavioral therapy (CBT) has potential for greatest impact to improve access to a scalable evidence-based psychological therapy.

Objective: This study aimed to examine the adaptation of a low-intensity CBT app to support low mood and worry management in female forces veterans and examine acceptability and usability.

Methods: Using a mixed methods methodology, this study comprises a focus group of female forces veterans to inform adaptation with extracted themes used as the basis of an adaptation framework. Following adaptation, a wider sample of female veterans was recruited to use the app and complete the mHealth App Usability Questionnaire to determine acceptability, usability, and usefulness.

Results: Two main areas were identified as requiring adaptation to maximize acceptability and usability. While using imagery and quotes to reflect the armed forces was initially found helpful to initiate engagement, it was considered that continued reference to the armed forces should be dropped when progressing through the app. Most app features were found acceptable; however, adaptations were requested to the content and structure of signposting information, navigation, and the way progress was monitored. No adaptations were required, however, regarding the CBT techniques used, with specific app features motivating engagement. Following the adaptation, there were good levels of acceptability, usability, and usefulness.

Conclusions: Involving female forces veterans as part of an intervention adaptation process has promise to improve acceptability and engagement with a digital CBT mobile phone intervention. Ensuring that the intervention represented the transition from serving to female forces veteran is of particular significance in enhancing acceptability.

(JMIR Hum Factors 2026;13:e84365) doi:10.2196/84365

KEYWORDS

female; veterans; armed forces; adaptation; acceptability; low-intensity CBT; Ecological Validity Framework; mixed methods

Introduction

Increased efforts are being made to recruit females into the UK Armed Forces who currently account for 11.5% of Regular and 16% of Reservists [1]. However, while government policy is seeking to increase this number to 30% by 2030, intake dropped by 1% in 2024 [1]. Despite significant efforts being made to integrate females into active service, female forces veterans remain underserved [2]. Difficulties are experienced because of support systems primarily developed for males with limited tailoring to the distinct needs of female forces veterans [3].

Both male and female forces veterans face barriers accessing mental health support such as stigma, concerns they will not be understood by health care providers, and a culture of stoicism [4]. However, female veterans frequently report additional gendered adversity, such as military sexual trauma and

XSL•FO
RenderX

emotional bullying [5]. Furthermore, they may need to navigate tensions between military culture and their feminine identity [6]. In other words, managing disparities between military culture and feminine norms, discrimination experienced both during and after service, lack of awareness, and particularly branding that reinforces male veteran stereotypes [7] are additional barriers to accessing mental health support faced by female forces veterans.

These barriers can, in turn, shape female forces veterans' help-seeking behavior [8]. Reports indicate that while female forces veterans are more likely to seek statutory adult mental health services than male veterans, they are less likely to access veteran-specific services [9]. Consequently, female veterans' lower engagement with veteran-specific services can limit their access to tailored support addressing areas such as finance, housing, employment, and welfare [10].

Additionally, compared with male veterans (16.7% and 13.5%), the prevalence of depression and anxiety among female veterans is 25.8% and 25.5%, respectively [11]. Experiencing these stressors during transition into civilian life is of particular concern as this is recognized as a particularly vulnerable period [12]. During this transition, female forces veterans are required to negotiate identity changes, gender-related, and societal expectancies to fit with feminine norms and practical difficulties related to housing and employment [13]. Together, the reduced utilization of veteran-focused mental health services, accompanied by higher rates of depression and anxiety experienced by female forces veterans, may contribute to the existing treatment gap [14]. This is of concern, given a higher prevalence of common mental health difficulties amongst female forces veterans [15].

To reduce the treatment gap, low-intensity cognitive behavioral therapy (CBT) has been implemented in statutory adult mental health services, such as NHS (National Health Service) Talking Therapies for anxiety and depression [16]. Low-intensity CBT improves access via flexible and scalable goal-oriented and time-limited self-help interventions [17]. The approach has already been demonstrated to resonate with values associated with self-management held by the armed forces [18], with flexibility more likely to enhance help-seeking and engagement [19]. Furthermore, acceptability has been demonstrated where intervention adaptations are made to tailor language and imagery to target veterans [20].

Low-intensity CBT also provides greater ability to enhance flexibility through digital delivery [21]. This is particularly important for armed forces veterans, given the growing acceptance of digital mental health support [22]. However, while mobile apps are being developed, for example, to promote recognition of mental health problems and support help-seeking among armed forces veterans [23], they remain largely biased toward male veterans. Challenges to generalizability, therefore, endure, resulting in female forces veterans at risk of remaining underserved [24].

Using a mixed methods methodology [25], this study reports on the adaptation of the Iona app [21] for female forces veterans. Adopting a systematic approach, the Ecological Validity Framework [26] (EVF) serves to guide culturally sensitive adaptations for specific groups by considering 8 dimensions: language, persons, metaphors, content, concepts, goals, methods, and context. Making adaptations to these dimensions offers promise to maximize acceptability to improve engagement and access to a scalable mental health approach [27].

Methods

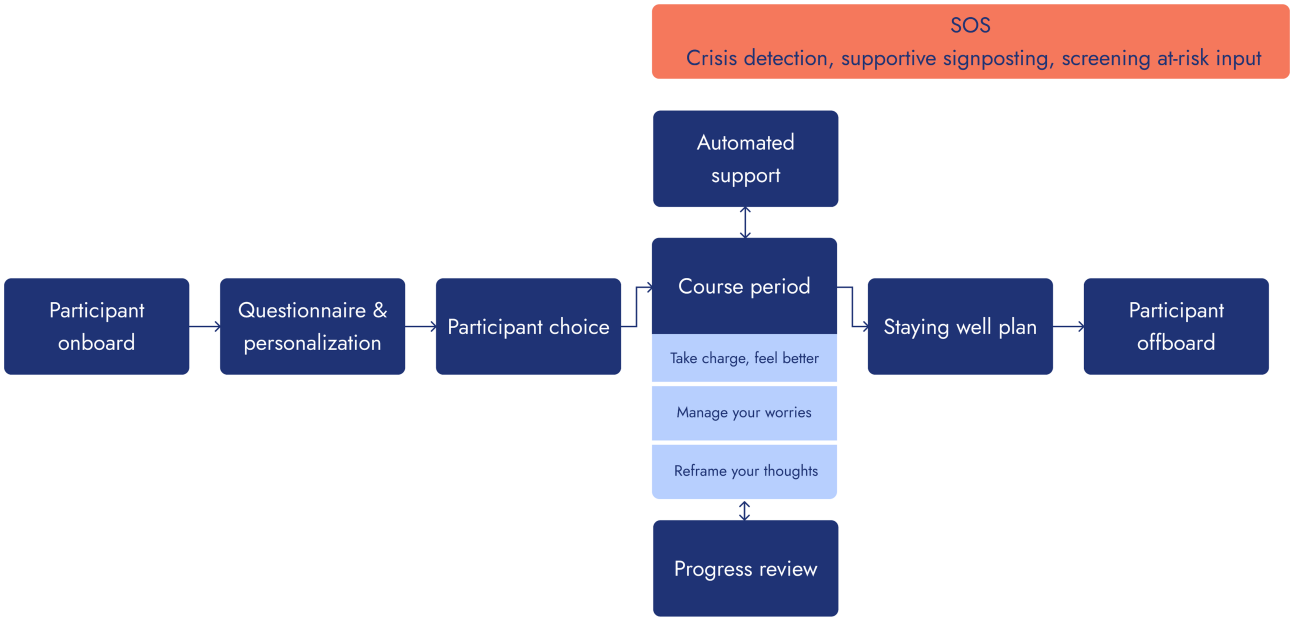
Intervention

The Iona Mind app (Iona) for the management of mild to moderate low mood [28] or excessive worry [21] informed the intervention adopted for adaptation (see Figure 1 for user flow through Iona).

After onboarding, the user is asked to complete the Patient Health Questionnaire-8 [29] and Generalized Anxiety Disorder-7 Scale [30] patient-reported outcome measures to assess the severity of low mood and worry symptoms, respectively. They then engage with an interactive CBT Five Areas model [31] before progressing through the intervention-specific factor techniques presented as 6 educational modules. Where the user scores mild to moderate symptom severity for both low mood and worry symptoms, they are directed to the technique corresponding with the highest symptom severity. Where symptom severity is identical, however, users are recommended the technique for low mood, although choice still can be exercised. An "SOS" button is available on every screen for users who either score within the severe range during onboarding or show signs of deterioration during use. If the person is seeking immediate help, the "SOS" page contains NHS Emergency contact details.

XSL•FO
RenderX

Figure 1. User flow through the Iona app.



Where treatment for low mood is recommended, the user is introduced to the behavioral activation [32] or cognitive restructuring [33] techniques derived from low-intensity CBT [34]. Alternatively, where symptoms associated with worry are prominent, the Worry Management technique is proposed [21]. To facilitate choice, a description of each technique is available on request. Based on the evidence base, all techniques are informed by protocols recommended by the National Institute for Health and Care Excellence, Behavioral Activation [35,36], alongside Cognitive Restructuring for low mood [37] and Worry Management for excessive worry [38]. All techniques are included within the most commonly adopted written self-help interventions [39], used by NHS Talking Therapies for anxiety and depression services [40]. Engagement with the app is maintained through an omnipresent chatbot to establish a "therapeutic approach" [41]. This promotes collaborative empiricism to encourage the user to explore and reflect upon outcomes arising from engagement for themselves [42].

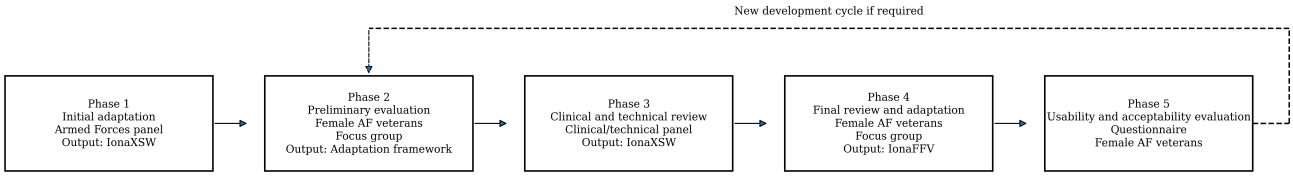
Intervention Adaptation

The EVF [26] was adopted to systematically inform adaptation from the original Iona app to that specifically for Iona Ex-Servicewomen (IonaXSW). The app was later currently called "Iona Female Forces Veterans" (IonaFFV) based on results from phase 2. To ensure that cultural and contextual factors associated with the female forces veteran's community are represented, the adaptation process focused on dimensions of the EVF [43] related to language, metaphor, context, and content. While dimensions related to concepts and methods were also adapted with respect to language, the theoretical rationale and technique-specific factors maintained high fidelity to the underlying CBT model. These dimensions were used to explicitly inform the focus group topic guide to ensure that structure was provided to ensure adequate coverage.

Adaptation Process

The adaptation process comprised 5 phases to adapt Iona and examine its acceptability for female forces veterans (Figure 2). Across the 5 phases, 4 panels were established to provide different perspectives and contributed in varied ways with each panel building upon the work of the others [44].

Figure 2. Adaptation process. AF: Armed Forces; IonaFFV: Iona Female Forces Veterans; IonaXSW: Iona Ex-Servicewomen.



During the adaptation process, an Armed Forces Panel was established to guide the initial adaptations, focusing on the context and content dimensions of the EVF. Phase 1 aimed to tailor the Iona app for female forces veterans and support the development of IonaXSW. After initial adaptations, a female forces veterans panel used their lived experience to refine the app further in phase 2. Data were analyzed to create an adaptation framework, which confirmed acceptable phase 1 adaptations and identified areas requiring additional modification. The framework was then provided to the clinical or technical panel in phase 3, who made further changes to develop the IonaFFV app. In phase 4, the app was returned to the same female forces veterans panel from phase 2 to assess the acceptability of the adaptations and ensure that changes based on the adaptation framework were effectively implemented. Finally, in phase 5, the IonaFFV app was released online to a wider group of female forces veterans to explore acceptability.

XSL•FO
RenderX

Methodology

Study Design

To adapt the Iona app to maximize acceptability for female forces veterans, an exploratory sequential mixed methods research design comprising focus groups (phases 2 and 4) and a questionnaire to examine acceptability and usability (phase 5) were adopted [45].

Phase 1

Participants

The Armed Forces Panel comprised a male armed forces officer veteran who held senior NHS roles associated with armed forces veterans and study author (AB), a female officer forces veteran holding national roles in veterans' charities, a male NHS Talking Therapies for anxiety and depression [16] armed forces veteran service lead, and a female regional commissioner for armed forces veterans.

Recruitment

Opportunistic sampling [46] based on personal contacts was held by AB. Participants contacted AB with preferences to organize a meeting time. The panel comprised 5 female forces veterans, 3 aged between 45 and 54 years and 2 aged between 35 and 44 years. Members of the panel had served in the Army (n=3), Royal Navy (n=1), and Royal Air Force (n=1).

Methods

Panel members were provided with a link to the unadapted Iona app and asked to review it before the meeting. A 1-hour online meeting, facilitated by AB and recorded for analysis, was held to gather initial feedback on areas requiring adaptation to better meet the needs of female forces veterans. Following the meeting, the recording was transcribed by a researcher (NF), who also made preliminary notes to develop an initial list of suggested adaptations. This list was then reviewed and refined in discussion with another researcher (PF). Once agreed upon, the final list of general adaptations was sent to the Iona app developer (JB) to create IonaXSW.

Phase 2

Participants

The panel comprised 5 female forces veterans, 3 aged between 45 and 54 years and 2 aged between 35 and 44 years. Participants had served in the Army (n=3), Royal Navy (n=1), and Royal Air Force (n=1).

Recruitment

The female forces veterans panel, which served as the focus group, was recruited using snowball nonprobability convenience sampling [47]. Study advertisements, including a link to participant information, were initially posted on social media platforms that engage female forces veterans. Individuals who provided consent and contacted the researcher were invited to share the advertisement on their own social media accounts to further extend recruitment. Interested participants then indicated their availability for the focus group, which was scheduled with the first 5 participants for whom a suitable date could be confirmed. Participants were reimbursed only for travel expenses to attend the focus groups.

Methods

Prior to the focus group, all members were sent a link to IonaXSW and asked to work through it for 4 weeks, with support to download provided by the researcher if requested. Based on the focus group schedule informed by dimensions of the EVF, a 3-hour audio-recorded focus group was held face-to-face facilitated by 2 members of the research team (PF and NF). A reflective journal was used to log the progress, obstacles, and success encountered running the focus group [48]. The topic guide comprised open-ended questions to gain an initial impression of IonaXSW, address each dimension covered by the EVF [43], and explore the ability to use and engage with each app component. Following the focus group, the audio recording was transcribed and analyzed to inform a second adaptation framework.

Analysis

The focus group discussion was transcribed by NF with the transcript uploaded into NVivo (version 12.0; QSR International) to support manifest content analysis [49]. Analysis was based on articulated data [50] where NF read the transcript, coded meaning units using line-by-line inductive coding, and developed a codebook. During coding, concepts and important portions of raw text were recorded with a word or short phrase used to represent the meaning of a specific text segment. To develop a deeper and more nuanced understanding of the findings, the codebook was then discussed with PF with variations in coding discussed and refined until consensus was reached. This resulted in new codes being added until new concepts emerged, or similar concepts were grouped together to form categories and subcategories. To enhance trustworthiness, field notes and audit trails were taken to ensure that preconceived perspectives were not imposed on the data, with a checklist adopted to inform all phases of analysis [51].

Phase 3

The clinical or technical panel comprised a male app developer (JB) for Iona Mind, and a male expert (PF) in low-intensity CBT apps who holds national expert advisory positions in low-intensity CBT and digital technologies. They used the adaptation framework developed in phase 2 to develop IonaFFV. The panel ensured that all adaptations maintained fidelity with the underlying low-intensity CBT clinical protocols, and software remained conformant to appropriate software requirements.

Phase 4

Overview

Once adapted, the same participants who comprised the female forces veterans focus group in phase 2 were sent a link to IonaFFV and asked to engage with it for 4 weeks before participating in a second focus group. The topic guide for this session was designed to gather overall impressions of IonaFFV and to evaluate how effectively the adaptations proposed in phase 2 had been implemented, with reference to each dimension of the EVF. In addition, selected questions from previous topic

XSL•FO
RenderX

guides on low-intensity CBT techniques [52] were incorporated to explore participant perceptions of different app components and their support preferences.

Analysis

Consistent with phase 2, the focus group discussion was transcribed and uploaded into NVivo (version 12.0) to support manifest content analysis [49]. Again, trustworthiness was established through discussions between NF and PF to reach consensus in coding [51].

Phase 5

Participants

A total of 21 participants completed the demographic questionnaire, with a mean age of 49 years. Participants had left the armed forces between 1 and 43 years previously ($\bar{x}$=18.71), although 1 participant did not report their age (Table 1). The majority identified as female (n=20, 95%), White British (n=20, 95%), and as having served as a regular in the Navy (n=11, 52%).

Table . Participant demographics.

Characteristics	Frequency
Age range (years), n (%)	
25 - 34	1 (5)
35 - 44	5 (25)
45 - 54	7 (35)
55 - 64	5 (25)
65 - 74	2 (10)
Gender, n (%)	
Female	20 (95)
Nonbinary	1 (5)
Ethnicity, n (%)	
White British	20 (95)
White and Black African	1 (5)
Service, n (%)	
Navy (regular)	11 (52)
Army (regular)	8 (38)
RAF[a] (regular)	1 (5)
Navy (reserve)	1 (5)
Months served	
Average	144.4
Minimum	12
Maximum	372
Reason for discharge, n (%)	
Planned	14 (67)
Unplanned (medical, compulsory, and family related)	7 (33)
Deployed, n (%)	
Yes	14 (67)
No	7 (33)
Years since left service, n (%)	
Average	18.71
Minimum	1
Maximum	73

[a]RAF: Royal Air Force.

XSL•FO
RenderX

Recruitment

Permission was obtained to promote a study advertisement on the social media platforms used by charitable and community organizations providing support to British female forces veterans. Additionally, members of the female forces veterans focus group posted the study advertisement through relevant social media platforms they had membership and sent it to personal contacts. The study advertisement stated that they should be female forces veterans aged 18 years or older, not currently receiving mental health support, and had access to a smartphone with internet access. Potential participants consented online and were automatically sent a link to download IonaFFV with onboarding details. To promote recruitment, potential participants were offered the opportunity to enter a prize draw for a £100 (US $133) electronic voucher.

Methods

Following download, participants were asked to trial the app for 2 weeks, after which they were sent a link to complete the stand-alone patient version of the mHealth App Usability Questionnaire (MAUQ) on the Qualtrics platform [53]. This measure adopts 18 items to explore usability and acceptability across 3 subscales: Ease of Use, Interface and Satisfaction, and Usefulness. However, as IonaFFV was used in a research study focused on adapting an app for female forces veterans with participants who did not present with depression or anxiety and developed for use outside of health care services, it was considered that 3 questions were inappropriate. Consequently, "I would use this app again," "The app improved my access to healthcare services," and "The app helped me manage my health effectively" were removed.

Participants also completed a series of Iona-related questions. These included four 7-point Likert scale items, added after phase 4, to assess participants' views on language, signposting information, and the SOS button. In addition, participants responded to a separate 7-point Likert scale question regarding human support. Those who selected Strongly Agree, Agree, Somewhat Agree, or Neither Agree nor Disagree were directed to answer 4 follow-up multiple-choice questions exploring different aspects of receiving human support alongside the app; multiple responses were permitted. At the end of the survey, all participants were invited to provide additional comments.

Analysis

For each subscale of the MAUQ, responses to the relevant items were averaged per participant and reported as median (IQR). All items were rated on a 7-point Likert scale, ranging from 1 (Strongly Agree) to 7 (Strongly Disagree). For analysis of the subscale items, responses of 1-3 were classified as agreement, 4 as neutral, and 5-7 as disagreement [54]. Percentages for each response category are reported for each question.

Additional questions examining attitudes toward specific adaptations to IonaFFV arising from phase 4 were rated using 7-point Likert-scale questions alongside 4 multiple-choice questions regarding preferences regarding type of human support. The Likert-scale questions ranged from 1 (Strongly Agree) to 7 (Strongly Disagree), with means and SDs calculated for each question. Proportions (%) were reported for the multiple-choice questions, with a multiple response frequency analysis conducted to examine the distribution of selected options regarding preferences for different aspects of human support [55]. As participants could select up to 6 responses for 3 questions and up to 5 for 1 question, frequencies and percentages were calculated for each response option.

Ethical Considerations

Ethical approval was obtained for the focus groups (reference: 2611056) and online acceptability (reference: 5930558) studies from the University of Exeter psychology department. To ensure informed consent, all participants were provided with participant information sheets, which included information on the aims of the study, right to withdraw, data storage, and management processes. Participants were informed that their participation was entirely voluntary, and they could withdraw from the study without any negative consequences. No identifying information about participants is included. All data were reviewed to ensure that individuals cannot be identified, and no materials containing identifiable features were used. As all data were anonymized prior to analysis, participant consent for publication of identifying information was not required. Phases 2 and 4 focus group participants were offered compensation for travel costs, and phase 5 questionnaire participants were offered the opportunity to enter a prize draw for a US $133 electronic voucher.

Results

Overview

Overall, feedback provided by the female veterans focus group combined with responses added to the free-text response box included within the MAUQ [53] questionnaire administered in phase 5 highlighted that IonaFFV was well received. Awareness that the app had been adapted for female forces veterans was positively responded to.

> *I'm still using it. It has helped a lot in the last 6 weeks or so. Please don't take it away. I can dip in when I'm finding things tough and knowing it's for female veterans is comforting.*

Adaptations appeared to promote continued engagement with IonaFFV as participants worked through each module.

> *It made me want to continue to see what was coming next and where it was going and all that.*

However, there was 1 disconfirming case reported on the free response text box.

> *The suggested activities are not relevant when you are actually struggling. I doubt anyone who has had an actual mental health episode was consulted. The activities are ok for when feeling a bit low, but not when you can't get out of bed and are struggling to function.*

Phases 2 and 4

Analysis is presented in 2 parts. First, reflecting the attitudes expressed toward IonaXSW by the focus group during phase 2, alongside preferences regarding support. Second, attitudes

XSL•FO
RenderX

expressed in phase 4 regarding the extent to which adaptations have been successfully implemented during phase 3 to create IonaFFV, alongside identifying outstanding areas requiring further adaptation. Categories identified as emerging were associated with "Context," "App features," and "No adaptations needed" (Table 2).

Table . Categories and subcategories associated with adaptation.

Armed forces context	App features	No adaptations required
Label	SOS page	CBT[a] model
Graphics	Navigation	Contextual factors
Quotes	Monitoring progress	Mapping users
Avatar	N/A[b]	N/A

[a]CBT: cognitive behavioral therapy.

[b]N/A: not applicable.

Armed Forces Context

This category captured the attitude of participants toward features of the Iona app that have been adapted to represent female forces veterans. These were grouped into label, graphics, quotes, and avatar.

Label

All participants in the focus group expressed disapproval toward using the "ex-servicewoman" label that was initially proposed by the Armed Forces Panel during intervention adaptation phase 1. They highlighted that this label does not accurately represent who they are and is not consistent with how they identify.

> *I really don't like the term 'ex-servicewoman'. I don't want to be viewed as 'ex' anything, but rather who I am now. I'm a female veteran and that's what I would like to be called.*

While it was agreed that the term "ex-servicewoman" was not appropriate, there was wider discussion among participants regarding a preference for what other labels should be used. While all focus group members preferred the term "veteran" to "ex-servicewomen," further discussion highlighted that some still considered it to be too vague.

> *I don't like veteran because it doesn't stipulate what you're a veteran of. You could be a veteran of working in an ice cream van.*

Participants further discussed using the term "female" versus "women" and armed forces, forces, or military. Consensus was finally reached regarding the term "female forces veteran." Discussions around the language adopted here highlight the importance of this EVF dimension when undertaking adaptations for female forces veterans.

Graphics

Participants expressed a complex set of attitudes toward the use of imagery related to the armed forces. There was general acceptance toward using the triservice colors representing the army, navy, and air force. Participants agreed that having the triservice color enhances the immediate association of the app with the armed forces (Figure 3A).

> *Keep the armed forces colors because people with military connections will pick up on that.*

However, it was considered that using armed forces graphics was justified only when restricted to the start of the app. Participants emphasized that as they progressed through the app, the focus of the content needed to shift toward their identity as female forces' veterans. Several participants also discussed that using these graphics within IonaXSW was extensive and could potentially lead to the app being considered "too official," as this was a military platform. It was considered that this could lead to female forces veterans being reluctant to engage initially or maintain engagement with the IonaXSW app.

> *If I could kind of have the benefit of, like, working through the app but without it feeling like too official, so using the image less so it's like 'this is unofficial', you know, then maybe that.*

Another concern participants expressed regarding overusing armed forces–related graphics was that this could result in the app being seen as developed for male rather than female veterans. That is, participants recognized a lack of female veterans' representation in the app.

> *And I think I just worry that if it's all branded, then it's kind of, it's like a bloke thing.*

XSL•FO
RenderX

Figure 3. Adaptations to Iona Female Forces Veterans.



Quotes

All participants agreed that using quotes from female forces veterans within IonaXSW was a good addition to promote motivation and continued engagement. They also agreed that locating these quotes near the start of the app was effective (Figure 3B). However, participants proposed that using quotes throughout IonaXSW should not be limited to the armed forces only but adapted to recognize the transition from the armed forces to civilians.

> *The idea of using quotes was good, however they should not be military because you are transitioning out of the force and want to blend in.*

One solution suggested by a participant was to facilitate the personalization of quotes by enabling users to add their own.

> *I think it's quite nice [to add your own quotes] and you can go back and then have a look at, and, you know, read over them and think oh, actually yeah.*

Consistent with this solution, one participant suggested adding a tab specifically designed to enable users to add their own quotes and organize them into different categories.

> *The motivational quotes are great. There should be a tab full of them with space to enter your own on a pinboard of gratitude, motivation, inspiration, and positivity.*

However, one participant noted that the user may need to exercise caution when allowing others to see their app with personalized quotes, given the unique armed forces sense of humor which could potentially cause offense to others.

> *What you have to remember is the forces have got their own sick sense of humor and I think sometimes our morbid sense of humor comes, comes through, and it is again a very fine line.*

Avatars

Within phase 2, the female forces veterans were unanimous in their dislike of the avatars chosen to introduce the modules included within IonaXSW. Participants expressed concern that the avatars did not appropriately capture the context dimension of the EVF. Although the appearance and voice adopted for the avatar were initially chosen in phase 1 by members of the Armed Forces Panel to be representative of female forces veterans, this was not considered to be the case in phase 2.

> *You look at them [the avatar] and you think, 'Actually, you don't look like you've ever served'. You know the person that you're looking at has to be believable as well I think.*

However, when the same voice was used without the accompanying avatar to introduce different components of the CBT techniques, only the audio was appreciated.

> *I liked the sound of the audio reading it [technique components] out, but would prefer not to see the face and just listen to it.*

App Features

App features refer to attitudes held regarding general features included within IonaXSW to support help seeking, aid navigation and usability, and motivate continued engagement.

SOS Page

Participants highlighted the importance of having contact details for organizations to manage risk or access wider sources of armed forces–specific support. To aid navigation, attention was directed toward providing a relevant structure. This involved moving the helplines to manage high risk provided at the top, followed by Veterans Welfare and Female Veterans Support. Participants also discussed exclusion of any contact details that may become out-of-date.

> *It's very important to ensure that it [link] works and that those listed don't give users what's termed as the "dockyard run around," where they take your details but ultimately don't help you and pass you on.*

Consequently, links to regional organizations were excluded, and contact details for NHS Op COURAGE provided in the event details for regional support were requested. Some participants also recognized that not all users may be aware of the services offered by included organizations. It was therefore

XSL•FO
RenderX

suggested to "include a brief description of what each charity provides."

Navigation

Overall, participants reported IonaFFV as being easy to engage with and that progress through the educational modules was straightforward.

> *Generally, I found it easy to move through the different parts [of IonaFFV] and get to where I wanted to go.*

However, uncertainty was expressed regarding the ability to navigate between the different low-intensity CBT techniques if they wanted to.

> *Can, I ask, is there a back button? And you know if, if, at the moment, I'm going down one course [low-intensity technique], because that's what I selected, but can you go back to select another one? I haven't quite found that out.*

Monitoring

Features adopted in IonaFFV to support users in recognizing their progress and achievements after completing specific tasks were generally well received by participants in phase 4.

> *I do like having the progress shown, and the praise for completing things.*

While participants appreciated knowing when a module had been completed, the use of graphics representing fireworks and confetti, along with the collection of badges to represent completion, was considered inappropriate for the target audience.

> *Get rid of the fireworks 'pop up'. It's a bit patronizing; but yes, to having something that recognizes your achievement. Likewise, the concept of getting badges is great, but the look is a bit childish and patronizing.*

No Adaptations Needed

Three main screens and interactive functionality were positively responded to by all focus group members in phase 2 and did not need further adaptation. This potentially indicates that initial adaptations in these areas made to accommodate female forces veterans were compatible with the content, methods, and context dimensions of the EVF.

CBT Model

Participants identified the "Here and Now" CBT model as helping the user understand the way in which a mental health difficulty can affect different areas of their life.

> *I liked the diagram [CBT model] where all the different areas associated with a problem were linked together. It helped show how different things are linked and sorting one out by using a treatment [CBT technique] can help the others.*

Additionally, being able to complete the CBT model themselves was also found helpful. This enabled participants to personalize the app to their experience and motivated them to remain engaged.

> *I liked that you filled in the boxes for yourself, so it reflected the problems you were having, and this made it more likely you wanted to carry on using it.*

Enhancing Contextual Factors

All participants agreed that presenting a screen with the names of trusted armed forces veterans' organizations and a university at the start of IonaFFV engagement helped them establish a sense of trust and raised their expectations regarding the app's effectiveness (Figure 3C).

> *Seeing that screen with the armed forces and University logos was good as it made me feel I could trust the app and that it may help.*

Mapping Users

All participants positively evaluated the screen designed to simulate the location of other app users (Figure 3D). One main benefit of this screen was to support an awareness that other female forces veterans were also using the app. This, in turn, led to motivating participants and supporting continued engagement.

> *I didn't realize there was that many people using it around my area. If there's that many people using it, it motivates me to maybe use it more because others are clearly using it as well.*

It was also considered that this screen would enhance engagement by creating a sense of community between female forces veterans. This sense of community helped participants recognize that their difficulties were shared by others, which may have reduced feelings of isolation and encouraged help-seeking behavior.

> *Adds a sense of community. Encourages you to believe you're not the only one [female forces veteran] facing mental health difficulties. In this sense it motivates you.*

The potential benefit of developing a map to enable users to contact each other highlights the importance of peer support in promoting mental health and the need for having this conversation.

> *Would it be possible to contact people who have started using the app? I would be really interested in knowing if I could meet them.*

Support Preferences

All focus group members highlighted the importance of having some form of human support available. They suggested that support could promote and maintain engagement with the app and be beneficial, especially where users are struggling.

> *Some human help might be useful, especially if after two weeks you've stopped engaging with the app as you are having problems.*

While providing face-to-face support was highlighted as a potential opportunity to further enhance and maintain

XSL•FO
RenderX

engagement, it was also considered that some participants may need it to be offered.

> *Face to face is a level of support that people ask for, but they don't actually know how to ask, so being offered it would make it easier, especially when struggling.*

Telephone support also emerged as an acceptable support modality. However, some considerations were raised regarding ways to provide and access it.

> *Regarding telephone calls, many don't answer calls from strangers, so if this were to be offered it might work better if a text was sent prior to a telephone call to advise that it [the call] is imminent.*

Furthermore, providing peer support from other female forces veterans was raised as having considerable potential with good levels of acceptability.

> *...peer to peer support is very underused but has a positive impact for us female veterans if the right person with similar experiences to us was in the role.*

Phase 5 (Acceptability)

Median (IQR) values across each of the MAUQ subscales were as follows: ease of use 1.80 (1.20 - 2.80), interface and satisfaction 2.50 (1.54 - 3.33), and usefulness 2.75 (1.75 - 4.00). Note that the lower values indicate higher agreement. Each subscale was further examined to determine percentage agreement with each statement (Table 3). Responses to some items were missing, resulting in varied sample sizes across items (n=19 - 20). Overall, most participants agreed with each statement, indicating that the app was highly acceptable, except for "I could use the app even when the Internet connection was poor or not available," rated by 63% of participants as neutral or disagreed.

XSL•FO
RenderX

Table . Agreement with each statement.

	Strongly agree/agree	Neutral	Strongly disagree/disagree
Ease of use, n (%)			
The app was easy to use.	17 (85)	0 (0)	3 (15)
It was easy for me to learn and use the app.	19 (95)	1 (5)	0 (0)
The navigation was consistent when moving between screens.	19 (95)	0 (0)	1 (5)
The app interface allowed me to use all the functions offered by the app (such as entering information, responding to reminders, and viewing information).	18 (90)	0 (0)	2 (10)
Whenever I made a mistake using the app, I could recover easily and quickly.	12 (60)	7 (35)	1 (5)
Interface and satisfaction, n (%)			
I like the interface of the app.	17 (85)	2 (10)	1 (5)
The information in the app was well organized, so I could easily find the information I needed.	14 (70)	2 (10)	4 (20)
The app acknowledged and provided information to let me know the progress of my action.	18 (90)	1 (5)	1 (5)
I feel comfortable using this app in social settings.	14 (74)	3 (16)	2 (10)
The amount of time involved in using this app has been fitting (adequate) for me.	13 (69)	1 (5)	5 (26)
Overall, I am satisfied with this app.	12 (60)	1 (5)	7 (35)
Usefulness, n (%)			
The app would be useful for my health and well-being.	11 (58)	2 (10)	6 (32)
This app has all the functions and capabilities I expected it to have.	14 (70)	4 (20)	2 (10)
I could use the app even when the Internet connection was poor or not available.	7 (37)	9 (47)	3 (16)
The app provides an acceptable way to receive mental well-being services and allows me to track my own activities.	15 (75)	2 (10)	3 (15)

Using a 7-point Likert scale (1=Strong Agree to 7=Strongly Disagree) a descriptive analysis was undertaken on questions related to 5 main areas associated with IonaFFV. Participants considered the language engaging (mean 1.85, SD 0.87), SOS button was required in the event of a mental health crisis (mean 1.95, SD 1.47), and the signposting information included within IonaFFV was helpful (mean 2.50, SD 1.38). However, it was considered that signposting details to regional organizations should be provided (mean 2.10, SD 1.15). Participants also indicated a need for human support to engage with the app (mean 1.90, SD 1.18). In follow-up questions about human support, where participants could select more than 1 option, there was a preference for weekly scheduled support provided by a professional or voluntary or community sector staff over the telephone (Table 4).

XSL•FO
RenderX

Table . Support preferences (N=19).

Questions and responses		Frequency, n (%)
Type of human support		
	Weekly contact with a professional	10 (53)
	Weekly contact with voluntary or community organization staff	10 (53)
	Weekly contact with a peer	9 (47)
	On-demand contact with a peer	8 (42)
	On-demand contact with a professional	5 (26)
	On-demand contact with voluntary or community organization staff	4 (21)
Ways human support should be offered		
	Telephone	14 (74)
	Videoconference	13 (68)
	Text	11 (58)
	Face-to-face	8 (42)
	Email	8 (42)
Types of organization(s) to offer human support		
	NHS (veteran)	16 (84)
	NHS (nonveteran)	15 (79)
	Voluntary and community sector (veteran)	12 (63)
	Voluntary and community sector (nonveteran)	9 (47)
	Workplace assistance program	8 (42)
Areas for human support to cover[a]		
	Motivation	14 (78)
	Process accountability (support engagement with CBT[b] techniques)	14 (78)
	Goal setting	12 (67)
	Signposting opportunities	10 (56)
	Expectations	9 (50)
	Performance monitoring	8 (44)

[a]N=18.

[b]CBT: cognitive behavioral therapy.

Interestingly, while there was least preference for weekly support to be provided by peers, using peers was identified as the most preferred in the event support was provided on demand. Where support was provided, it was considered that it should be focused on motivating continued engagement and supporting engagement with techniques used within the app as required.

Discussion

Principal Findings

Overall, adopting CBT as the basis of an app to support low mood and worry management in female forces veterans was acceptable. Participants found that the CBT model adapted to reflect female forces veterans both aided understanding of their emotional difficulties and helped tailor the app to them as a distinct group. Furthermore, there was good engagement with techniques used to support low mood and worry management, with initial engagement further enhanced by highlighting use of IonaFFV by other users at the start of engagement and the contextual factors. However, participants differ from most previous research studies with respect to the following characteristics: from the Navy, underwent unplanned discharge, and deployed. Although the average age is consistent with previous research examining access to the Op COURAGE high-intensity service [10].

Applying the EVF [43] to inform the development of an artificial intelligence–supported CBT app for the management of low mood and worry was, however, important to enhance acceptability and usefulness among female forces veterans. Analysis of phases 2 and 4 highlighted the need to make

XSL•FO
RenderX

adaptations to ensure that IonaFFV specifically represented their transition from serving into civilian life. Focusing on the EVF context dimension highlighted the need to redirect adaptations from serving to female forces veterans to adequately capture this transition while maximizing acceptability. This was exemplified by reactions to the term "ex-servicewomen," which was not considered to represent a new civilian identity while acknowledging their past service in the armed forces. This highlights the importance of the language dimension of the EVF when considering adaptations for female forces veterans.

Of interest, "ex-servicewomen" was the term originally proposed during phase 1 of the development process comprising senior members of the armed forces community and service providers. Furthermore, while phase 1 was used to identify the graphics and voices for use by the avatars, these were not considered to represent female forces veterans as a group and therefore removed during development. Adaptations were also required to enhance usability with respect to the structure of the SOS page, enhance navigation through the app where there was a desire to change CBT technique, and enhance the way in which progress was captured.

Comparison With Previous Research

Consistent with previous research, the findings of this study reinforce the significant role that serving in the armed forces plays in shaping the identities of both male and female veterans [56]. The results emphasize the particular challenges faced by female forces veterans in negotiating their identities during the transition to civilian life [6]. This highlights the need for tailored approaches that acknowledge their unique experiences in the armed forces. While female forces veterans are often "hypervisible" during their military careers, they may simultaneously suppress behaviors traditionally considered feminine [7]. Failure to recognize these adaptive strategies supports the need identified in this study for appropriate modifications to IonaFFV. Implementing adaptations was recognized to improve acceptability by better reflecting the experiences of female forces veterans as they transition from life in the armed forces to becoming a civilian [6].

While the initial adaptations to IonaXSW used graphics aligned with the armed forces to enhance engagement, participants emphasized the importance of gradually reducing the focus on the armed forces as they progressed through the app to better reflect civilian life [7]. They also recommended replacing quotes associated with the armed forces context with those drawn from civilian life. A lack of recognition of the transition from military to civilian identity could have hindered the process of "rediscovery" that supports successful reintegration [7]. Furthermore, failing to change the term "ex-servicewomen" [1,7], proposed in adaptation phase 1 to female forces veterans, would have further reflected a transition oversight. However, the preference to maintain "veteran" is not consistent with previous research [57], where it is traditionally considered to reflect previous generations of male veterans [57]. Potentially, these differences have arisen from study demand characteristics, given that recruitment materials adopted the term "veteran" [58].

Features of IonaXSW that did not require adaptation are consistent with those commonly identified as acceptable with CBT [59]. For example, the CBT model both enhanced understanding regarding the nature of the mental health difficulty faced and aided an understanding of the way the low-intensity CBT technique may help [60]. Furthermore, consistent with a metasynthesis of computerized CBT, personalization of CBT-specific factor content enhanced relevance [61], promoting connection and engagement. The screen simulating the location of other female forces veterans also served to normalize mental health support [62] and encouraged lateral social comparisons [63].

While the chatbot successfully maintained user engagement with IonaXSW, consistent with findings from CBT-based self-help interventions more broadly, participants continued to express a preference for human support [64]. Among the various forms of support considered acceptable, the most preferred option was scheduled weekly sessions with a professional from a veteran-specific NHS service. This finding suggests potential benefits of integrating support for IonaFFV through services such as Op COURAGE [10]. However, despite the overall acceptability of this service [10], its evaluation included only male veterans, which limits the generalizability of the findings to female armed forces veterans.

Alternatively, nearly half of the participants preferred on-demand peer support, indicating potential advantages similar to those observed with peer-support groups [65]. Implementing IonaFFV alongside support from female forces veterans within veteran-specific services [9] could serve to enhance help-seeking and address existing barriers to engagement [66]. Moreover, as low-intensity CBT techniques can improve access to care, adoption could enable services to manage increased demand without extending waiting times [67]. While IonaFFV incorporates in-app support to promote engagement [68], incorporating peer support could mitigate low levels of engagement with a stand-alone app following download [69].

Strengths and Limitations

Adopting an approach toward adaptation comprising focus groups to guide initial modifications and using the MAUQ [53] to assess acceptability offers a means to evaluate how well these adaptations are acceptable to a broader group of female forces veterans. This methodology could help address challenges related to the limited generalizability of focus group data [70]. However, the small number of participants completing the MAUQ limited further analysis regarding the acceptability of IonaFFV across different demographic groups. Specifically, only a single participant had served in the Royal Air Force, identified as non-White, or had experience as a reservist. Recruitment challenges for phase 5 may have been influenced by the requirement for participants to engage with IonaFFV for 4 weeks before completing the questionnaire. Therefore, future research should consider strategies to enhance recruitment and questionnaire completion before assessing the acceptability and usability of IonaFFV and generalizing study results [71].

It may also have been of benefit to collect additional information in phase 5 regarding the service experiences of female forces veterans. Female forces veterans experience higher levels of

XSL•FO
RenderX

in-service adversity than their male colleagues, including exposure to military sexual trauma [72]. Although IonaFFV was explicitly designed for female forces veterans experiencing low mood or difficulties with worry management, some questionnaire responses may have been influenced by past service-related adversity. This possibility was suggested by a free-text questionnaire response which indicated that the participant may have been experiencing moderate to severe mental health difficulties. Gathering more comprehensive data on participants' service experiences could have supported a deeper interpretation of the findings and informed further adaptations, such as incorporating broader signposting information.

Conclusions

There is a clear need to recognize female forces veterans as a distinct group from their male peers. While sharing many values and traditions common across the armed forces, they also hold a collective identity that differs from that of male veterans [73]. Addressing the current inequities in access to, and appropriateness of, mental health services for female veterans [8,14] requires a focus on enhancing the acceptability of veteran-specific services. In particular, using dimensions of the EVF to structure the focus group topic guide was valuable to identify the types of adaptations necessary to maximize acceptability. The EVF also helped highlight how these dimensions shift during the transition from serving to armed forces veteran, offering insight into how mental health services can be better tailored for female forces veterans. When applied to the development of a CBT-based mobile phone app for managing low mood and worry, the EVF informed adaptations that produced a digital mental health approach with strong usability, acceptability, and perceived usefulness among female forces veterans.

Acknowledgments

The authors are grateful to members of the Armed Forces Panel, the female forces veterans comprising the focus group and those completing the acceptability questionnaire for participating in this study, and to the female veterans' charities for supporting recruitment into phase 5 of the study.

The authors attest that there was no use of generative artificial intelligence (AI) technology in the generation of text, figures, or other informational content of this manuscript

Funding

This paper has been produced with funding from the Office for Veterans' Affairs in April 2023 as part of an independent research initiative. The views, findings, and conclusions expressed herein are those of the authors and do not reflect the policies or positions of the UK Government.

Data Availability

mHealth App Usability Questionnaire data are available only on request from the authors.

Authors' Contributions

PF conceptualized and designed the study with AB providing theoretical input informing the background of the paper and armed forces context and supporting recruitment. NF, MJ, and ET were involved in data analysis, with JB undertaking adaptations to the Iona app, drafting the section describing the app, and clarifying any uncertainties regarding the app description. PF wrote the initial draft of the paper, with all authors contributing to the development of the paper, interpretation of the analysis, and editing of the final manuscript.

Conflicts of Interest

JB is a full-time employee, and PF is on a paid sabbatical with Iona Mind Inc from the University of Exeter ending October 2025. All other authors have confirmed that they have no conflicts of interest to declare.

References

1. UK Armed Forces biannual diversity statistics: April 2024. Gov.uk. URL: https://www.gov.uk/government/statistics/uk-armed-forces-biannual-diversity-statistics-april-2024/uk-armed-forces-biannual-diversity-statistics-april-2024 [accessed 2025-09-09]
2. Croak B, Rafferty L, Sharp ML, et al. Barriers and facilitators to successful transition to civilian life for ex‐servicewomen: the perspective of service providers and policymakers. Health Soc Care Community 2025 Jan;2025(1). [doi: 10.1155/hsc/2327743]
3. Godier-McBard LR, Gillin N, Fossey MJ. "Treat everyone like they're a man": stakeholder perspectives on the provision of health and social care support for female veterans in the UK. Health Soc Care Community 2022 Nov;30(6):e3966-e3976. [doi: 10.1111/hsc.13790] [Medline: 35289437]

XSL•FO
RenderX

4. Campbell GM, Perry MP, Milford J, Murphy D. Personalising veteran healthcare: recognising barriers to access for minority and under-represented groups of veterans. BMJ Mil Health 2024 Sep 20;170(5):446-450. [doi: 10.1136/military-2024-002768] [Medline: 38897640]
5. Hendrikx LJ, Williamson V, Murphy D. Adversity during military service: the impact of military sexual trauma, emotional bullying and physical assault on the mental health and well-being of women veterans. BMJ Mil Health 2023 Oct;169(5):419-424. [doi: 10.1136/bmjmilitary-2021-001948] [Medline: 34697241]
6. Demers AL. From death to life: female veterans, identity negotiation, and reintegration into society. J Humanist Psychol 2013;53(4):489-515. [doi: 10.1177/0022167812472395]
7. Sharp ML, Croak B, Khan R, et al. SUSTAIN: Identifying and Examining the Barriers and Facilitators to Ex-Servicewomen Making a Successful and Sustainable Transition to Civilian Life in the UK: King's College London; 2025. URL: https://kcmhr.org/pdf/Sustain_ExServiceWomen_Transition_Report.pdf [accessed 2025-09-11] [doi: 10.1155/hsc/2327743]
8. Randles R, Finnegan A. Veteran help-seeking behaviour for mental health issues: a systematic review. BMJ Mil Health 2022 Feb;168(1):99-104. [doi: 10.1136/bmjmilitary-2021-001903] [Medline: 34253643]
9. Godier-McBard LR, Wood A, Kohomange M, Cable G, Fossey M. Barriers and facilitators to mental healthcare for women veterans: a scoping review. J Ment Health 2023 Oct;32(5):951-961. [doi: 10.1080/09638237.2022.2118686] [Medline: 36062860]
10. Finnegan A, Salem K, Green N, Ainsworth-Moore L, Ghomi M. Evaluation of the NHS England 'Op COURAGE' High Intensity Service for military veterans with significant mental health problems. BMJ Mil Health 2025 Feb;171(1):51-58. [doi: 10.1136/military-2023-002385]
11. Finnegan A, Randles R. Prevalence of common mental health disorders in military veterans: using primary healthcare data. BMJ Mil Health 2023 Dec;169(6):523-528. [doi: 10.1136/bmjmilitary-2021-002045]
12. Mattocks KM, Shenette LL, Walker L, Maher N, Haskell S. Understanding women's transitions from military to Department of Veterans Affairs Care. J Womens Health (Larchmt) 2024 Sep 1;33(9):1185-1197. [doi: 10.1089/jwh.2024.0022]
13. Smith A, Rafferty L, Croak B, et al. A systematic review of military-to-civilian transition, the role of gender. PLoS One 2025;20(2):e0316448. [doi: 10.1371/journal.pone.0316448]
14. Iversen A, Nikolaou V, Greenberg N, et al. What happens to British veterans when they leave the armed forces? Eur J Public Health 2005 Apr;15(2):175-184. [doi: 10.1093/eurpub/cki128] [Medline: 15755777]
15. Jones N, Jones M, Greenberg N, Phillips A, Simms A, Wessely S. UK military women: mental health, military service and occupational adjustment. Occup Med (Lond) 2020 Jun 20;70(4):235-242. [doi: 10.1093/occmed/kqaa019]
16. National Collaborating Centre for Mental Health. NHS Talking Therapies for Anxiety and Depression Manual: National Collaborating Centre for Mental Health; 2024. URL: https://www.england.nhs.uk/wp-content/uploads/2018/06/nhs-talking-therapies-manual-v7.1-updated.pdf [accessed 2026-03-29]
17. Lee CT, Harty S, Adegoke A, Palacios J, Gillan CM, Richards D. The effectiveness of low-intensity psychological interventions for comorbid depression and anxiety in patients with long-term conditions: a real-world naturalistic observational study in IAPT integrated care. Int J Behav Med 2024 Oct;31(5):730-740. [doi: 10.1007/s12529-023-10215-9] [Medline: 37697142]
18. Farrand P, Jeffs A, Bloomfield T, Greenberg N, Watkins E, Mullan E. Mental health service acceptability for the armed forces veteran community. Occup Med (Lond) 2018 Aug 11;68(6):391-398. [doi: 10.1093/occmed/kqy086] [Medline: 29912388]
19. Powell T, Garratt K, Burnett H. Veterans: access to health services. House of Commons Library. 2024. URL: https://researchbriefings.files.parliament.uk/documents/CBP-10126/CBP-10126.pdf [accessed 2025-09-11]
20. Farrand P, Mullan E, Rayson K, Engelbrecht A, Mead K, Greenberg N. Adapting CBT to treat depression in Armed Forces Veterans: qualitative study. Behav Cogn Psychother 2019 Sep;47(5):530-540. [doi: 10.1017/S1352465819000171] [Medline: 30894234]
21. Farrand P, Raue PJ, Ward E, Repper D, Baker J, Areán P. Use and engagement with low-intensity cognitive behavioral therapy techniques used within an app to support worry management: content analysis of log data. JMIR Mhealth Uhealth 2024 Jan 10;12:e47321. [doi: 10.2196/47321] [Medline: 38029300]
22. Shore JH, Aldag M, McVeigh FL, Hoover RL, Ciulla R, Fisher A. Review of mobile health technology for military mental health. Mil Med 2014 Aug;179(8):865-878. [doi: 10.7205/MILMED-D-13-00429] [Medline: 25102529]
23. Parkes S, Croak B, Brooks SK, et al. Evaluating a smartphone app (MeT4VeT) to support the mental health of UK Armed Forces Veterans: feasibility randomized controlled trial. JMIR Ment Health 2023 Aug 28;10:e46508. [doi: 10.2196/46508] [Medline: 37639295]
24. Fitzke RE, Bouskill KE, Sedano A, et al. Barriers and facilitators to behavioral healthcare for women veterans: a mixed-methods analysis of the current landscape. J Behav Health Serv Res 2024 Apr;51(2):164-184. [doi: 10.1007/s11414-023-09862-3] [Medline: 37798569]
25. McBride KA, MacMillan F, George ES, Steiner GZ. The use of mixed methods in research. In: Liamputtong P, editor. Handbook of Research Methods in Health Social Sciences: Springer:978-981. [doi: 10.1007/978-981-10-5251-4_97]

XSL•FO
RenderX

26. Bernal G, Bonilla J, Bellido C. Ecological validity and cultural sensitivity for outcome research: issues for the cultural adaptation and development of psychosocial treatments with Hispanics. J Abnorm Child Psychol 1995 Feb;23(1):67-82. [doi: 10.1007/BF01447045] [Medline: 7759675]
27. Heim E, Kohrt BA. Cultural adaptation of scalable psychological interventions: a new conceptual framework. Clin Psychol Eur 2019;1(4). [doi: 10.32872/cpe.v1i4.37679]
28. Farrand P, Raue PJ, Ward E, Healey S, Baker J. Low-intensity behavioral activation smartphone app to support low mood: description, engagement and log data analysis. Acad Ment Health Well Being 2025;2(2). [doi: 10.20935/MHealthWellB7785]
29. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026]
30. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
31. Williams C, Garland A. A cognitive–behavioural therapy assessment model for use in everyday clinical practice. Adv Psychiatr Treat 2002 May;8(3):172-179. [doi: 10.1192/apt.8.3.172]
32. Farrand P, McGahey S, Woodford J. Lift Your Low Mood: University of Exeter; 2022. URL: https://cedar.exeter.ac.uk/v8media/facultysites/hls/cedar/resources/Lift_Your_Low_Mood_(Colour).pdf [accessed 2025-09-10]
33. Farrand P, Woodford J, Jackson K. Unhelpful Thoughts: Challenging and Testing Them Out: University of Exeter; 2019. URL: https://cedar.exeter.ac.uk/v8media/facultysites/hls/cedar/resources/2018CLES041_CEDAR_Unhelpful_Thoughts_(Interactive).pdf [accessed 2025-09-10]
34. Farrand P. Low-Intensity CBT Skills and Interventions: A Practitioner's Manual, 2nd edition: SAGE Publications; 2024. [doi: 10.4135/9781036232641]
35. Hopko DR, Lejuez CW, Ruggiero KJ, Eifert GH. Contemporary behavioral activation treatments for depression: procedures, principles, and progress. Clin Psychol Rev 2003 Oct;23(5):699-717. [doi: 10.1016/s0272-7358(03)00070-9] [Medline: 12971906]
36. Addis ME, Jacobson NS, Martell CR. Depression in Context: Strategies for Guided Action: WW Norton; 2001.
37. Depression in adults: treatment and management. National Institute for Health and Clinical Excellence. URL: https://www.nice.org.uk/guidance/ng222 [accessed 2025-09-12]
38. Borkovec TD, Wilkinson L, Folensbee R, Lerman C. Stimulus control applications to the treatment of worry. Behav Res Ther 1983;21(3):247-251. [doi: 10.1016/0005-7967(83)90206-1] [Medline: 6615390]
39. Farrand P, Dawes A, Doughty M, et al. Development and application of criteria to evaluate written CBT self-help interventions adopted by Improving Access to Psychological Therapies services. Cogn Behav Therap 2022;15(e28). [doi: 10.1017/S1754470X22000241]
40. NHS talking therapies for anxiety and depression manual. National Collaborating Centre for Mental Health. 2024. URL: https://www.england.nhs. uk/wp-content/uploads/2018/06/nhs-talking-therapies-manual a.-v7.1-updated.pdf [accessed 2025-09-11]
41. Tee J, Kazantzis N. Collaborative empiricism in cognitive therapy: A definition and theory for the relationship construct. Clin Psychol Sci Pract 2011;18(1):47-61. [doi: 10.1111/j.1468-2850.2010.01234.x]
42. Kazantzis N, Beck JS, Dattilio FM, Dobson KS, Rapee RM. Collaborative empiricism as the central therapeutic relationship element in CBT: an expert panel discussion at the 7th International Congress of Cognitive Psychotherapy. Int J Cogn Ther 2013 Dec;6(4):386-400. [doi: 10.1521/ijct.2013.6.4.386]
43. Bernal G, Sáez‐Santiago E. Culturally centered psychosocial interventions. J Community Psychol 2006 Mar;34(2):121-132. [doi: 10.1002/jcop.20096]
44. O'Connor TM, Perez O, Beltran A, et al. Cultural adaptation of "Healthy Dads, Healthy Kids" for Hispanic families: applying the ecological validity model. Int J Behav Nutr Phys Act 2020 Apr 21;17(1):52. [doi: 10.1186/s12966-020-00949-0] [Medline: 32316983]
45. Takona JP. Research design: qualitative, quantitative, and mixed methods approaches / sixth edition. Qual Quant 2024 Feb;58(1):1011-1013. [doi: 10.1007/s11135-023-01798-2]
46. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015 Sep;42(5):533-544. [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
47. Leighton K, Kardong-Edgren S, Schneidereith T, Foisy-Doll C. Using social media and snowball sampling as an alternative recruitment strategy for research. Clin Simul Nurs 2021 Jun;55:37-42. [doi: 10.1016/j.ecns.2021.03.006]
48. Heslop CW, Burns S, Lobo R, McConigley R. Developing a framework for community-based sexual health interventions for youth in the rural setting: protocol for a participatory action research study. BMJ Open 2017 May 30;7(5):e013368. [doi: 10.1136/bmjopen-2016-013368] [Medline: 28559453]
49. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
50. Massey OT. A proposed model for the analysis and interpretation of focus groups in evaluation research. Eval Program Plann 2011 Feb;34(1):21-28. [doi: 10.1016/j.evalprogplan.2010.06.003] [Medline: 20655593]

XSL•FO
RenderX

51. Elo S, Kääriäinen M, Kanste O, Pölkki T, Utriainen K, Kyngäs H. Qualitative content analysis: a focus on trustworthiness. SAGE Open 2014;4(1):215824401452263. [doi: 10.1177/2158244014522633]
52. Svedin F, Brantnell A, Farrand P, et al. Adapting a guided low-intensity behavioural activation intervention for people with dementia and depression in the Swedish healthcare context (INVOLVERA): a study protocol using codesign and participatory action research. BMJ Open 2021 Jul 16;11(7):e048097. [doi: 10.1136/bmjopen-2020-048097] [Medline: 34272221]
53. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500. [doi: 10.2196/11500] [Medline: 30973342]
54. Teong LF, Khor BH, Radion Purba K, et al. A mobile app for triangulating strategies in phosphate education targeting patients with chronic kidney disease in Malaysia: development, validation, and patient acceptance. Healthcare (Basel) 2022;10(3):535. [doi: 10.3390/healthcare10030535]
55. Vokey JR. Multiway frequency analysis for experimental psychologists. Can J Exp Psychol 2003 Sep;57(3):257-264. [doi: 10.1037/h0087429] [Medline: 14596481]
56. Cooper L, Caddick N, Godier L, Cooper A, Fossey M. Transition from the military into civilian life: an exploration of cultural competence. Armed Forces Soc 2018;44(1):156-177. [doi: 10.1177/0095327X16675965]
57. Burdett H, Woodhead C, Iversen AC, Wessely S, Dandeker C, Fear NT. “Are You a Veteran?” Understanding of the term “Veteran” among UK ex-service personnel. Armed Forces Soc 2013 Oct;39(4):751-759. [doi: 10.1177/0095327X12452033]
58. Jones G. Exploring the psychological health and wellbeing experiences of female veterans transitioning from military to civilian environments. Dissertation.: University of Manchester; 2018 URL: https://research.manchester.ac.uk/en/studentTheses/exploring-the-psychological-health-and-wellbeing-experiences-of-f/ [accessed 2026-03-29]
59. Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa TA. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression. JAMA Psychiatry 2019 Jul 1;76(7):700. [doi: 10.1001/jamapsychiatry.2019.0268]
60. Redhead S, Johnstone L, Nightingale J. Clients’ experiences of formulation in cognitive behaviour therapy. Psychol Psychother 2015 Dec;88(4):453-467. [doi: 10.1111/papt.12054] [Medline: 25573018]
61. Knowles SE, Toms G, Sanders C, et al. Qualitative meta-synthesis of user experience of computerised therapy for depression and anxiety. PLoS One 2014;9(1):e84323. [doi: 10.1371/journal.pone.0084323]
62. Pescosolido BA, Medina TR, Martin JK, Long JS. The “backbone” of stigma: identifying the global core of public prejudice associated with mental illness. Am J Public Health 2013 May;103(5):853-860. [doi: 10.2105/AJPH.2012.301147]
63. Gerber JP, Wheeler L, Suls J. A social comparison theory meta-analysis 60+ years on. Psychol Bull 2018 Feb;144(2):177-197. [doi: 10.1037/bul0000127] [Medline: 29144145]
64. Hanson K, Webb TL, Sheeran P, Turpin G. Attitudes and preferences towards self-help treatments for depression in comparison to psychotherapy and antidepressant medication. Behav Cogn Psychother 2016 Mar;44(2):129-139. [doi: 10.1017/S1352465815000041]
65. Helgeson VS, Gottlieb BH. Support groups. In: Cohen S, Underwood LG, Gottlieb BH, editors. Social Support Measurement and Intervention: A Guide for Health and Social Scientists: Oxford University Press; 2000:221-245. [doi: 10.1093/med:psych/9780195126709.003.0007]
66. Campbell GM, Williamson V, Murphy D. “A hidden community”: the experiences of help-seeking and receiving mental health treatment in U.K. women veterans. a qualitative study. Armed Forces Soc 2025 Jan;51(1):22-45. [doi: 10.1177/0095327X231182140]
67. Farrand P. Low intensity cognitive behavioral therapy: revolution not evolution. In: Farrand P, editor. Low-Intensity CBT Skills and Interventions: A Practitioner’s Manual: SAGE Publications; 2025:5-20. [doi: 10.4135/9781036232641.n1]
68. Farrand P, Woodford J. Impact of support on the effectiveness of written cognitive behavioural self-help: a systematic review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2013 Feb;33(1):182-195. [doi: 10.1016/j.cpr.2012.11.001] [Medline: 23238023]
69. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
70. Vicsek L. Issues in the analysis of focus groups: generalisability, quantifiability, treatment of context and quotations. TQR 2010;15(1):122-141. [doi: 10.46743/2160-3715/2010.1143]
71. Fan W, Yan Z. Factors affecting response rates of the web survey: a systematic review. Comput Human Behav 2010 Mar;26(2):132-139. [doi: 10.1016/j.chb.2009.10.015]
72. Herriott C, Campbell G, Godier-McBard L, Wood A, Murphy D. Defining military sexual trauma: establishing parameters and perspectives. Eur J Psychotraumatol 2024;15(1):2390759. [doi: 10.1080/20008066.2024.2390759] [Medline: 39149943]
73. Crompvoets S. The health and wellbeing of female veterans: a review of the literature. J Mil Vet Health 2011;19(2):25-31 [FREE Full text]

Abbreviations

CBT: cognitive behavioral therapy
EVF: Ecological Validity Framework

XSL•FO
RenderX

IonaFFV: Iona Female Forces Veterans
IonaXSW: Iona Ex-Servicewomen
MAUQ: mHealth App Usability Questionnaire
NHS: National Health Service

Edited by A Kushniruk; submitted 18.Sep.2025; peer-reviewed by LR Godier-McBard, M Hirschi; revised version received 14.Jan.2026; accepted 19.Feb.2026; published 29.Apr.2026.

Please cite as:
Farrand P, Bacon A, Janbakhsh M, Flay N, Turnbull E, Baker J
Adaptation and Acceptability of a Low-Intensity Cognitive Behavioral Therapy App to Support Low Mood and Worry Management in Female Forces Veterans: Mixed Methods Study
JMIR Hum Factors 2026;13:e84365
URL: https://humanfactors.jmir.org/2026/1/e84365
doi:10.2196/84365

© Paul Farrand, Andy Bacon, Melika Janbakhsh, Natalie Flay, Elizabeth Turnbull, Jonathan Baker. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 29.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Mental Health Profiles Based on Self-Regulation and Technology Use in the Digital Era in a Spanish-Speaking Sample: Latent Profile Analysis

Angélica Garzón Umerenkova[1], PhD; Marisleidy Alba Cabañas[2], PhD; Erika Andrea Malpica-Chavarria[1], PhD

[1]School of Psychology, Fundación Universitaria Konrad Lorenz, Bogotá, Colombia

[2]Business School, Fundación Universitaria Konrad Lorenz, Cra. 9 Bis #No. 62 - 43, Bogotá, Colombia

Corresponding Author:
Marisleidy Alba Cabañas, PhD
Business School, Fundación Universitaria Konrad Lorenz, Cra. 9 Bis #No. 62 - 43, Bogotá, Colombia

Abstract

Background: The widespread use of digital technologies has raised growing concerns about their impact on mental health. While self-regulation has been proposed as a protective factor, little is known about how distinct psychological profiles based on self-regulatory and technology use patterns relate to psychological distress. Person-centered approaches, such as latent profile analysis, may offer deeper insights, particularly in underrepresented populations.

Objective: This study aimed to identify latent psychological profiles based on self-regulation, nomophobia (fear of being without a phone), and problematic use of the internet and social media (defined by behavioral symptoms), to examine their associations with general psychological distress and the presence of emotional symptoms in a Colombian sample. Additionally, the predictive roles of age and gender in class membership were explored.

Methods: Participants were recruited through a convenience sampling strategy aimed at ensuring heterogeneity of the sample in terms of age and gender. A total of 453 participants aged 12 to 57 years (mean 21.03, SD 8.41 years; 257/453, 56.7% female) completed validated measures of self-regulation (Abbreviated Self-Regulation Questionnaire), nomophobia (Nomophobia Questionnaire), internet and social media use (MULTICAGE-TIC, a multidomain screening questionnaire based on the CAGE framework), and psychological distress (General Health Questionnaire-12). Latent profile analysis was conducted using standardized scores of continuous variables. Model fit was assessed using the Bayesian information criterion, entropy, and bootstrapped likelihood ratio test. Differences in psychological distress scores across latent classes were examined through variance analysis (ANOVA) and regression models. A multinomial logistic regression tested the predictive value of age and gender for class membership.

Results: The optimal solution revealed 4 distinct latent profiles (entropy=0.85). Class 1 showed high self-regulation and low problematic technology use, displaying the lowest psychological distress scores. Class 2 presented moderate levels across all indicators but the highest level of psychological distress. Classes 3 and 4 showed mixed patterns. Class 3 (higher information and communication technology [ICT] use and lower self-regulation) exhibited lower distress than class 2, whereas class 4 (younger individuals with low self-regulation and moderately high ICT use) showed higher distress than class 3. Psychological distress differed significantly across profiles (ANOVA, $P<.001$). Age and gender predicted class membership. Older males were more likely to belong to class 1, and younger females were more likely to be classified into classes 3 and 4.

Conclusions: Latent profile analysis identified distinct configurations of digital behavior, self-regulation, and psychological distress. Self-regulation consistently differentiated profiles with lower distress scores, suggesting its relevance for understanding how individuals manage ICT use. These findings support the value of person-centered approaches to characterize heterogeneous patterns of technology-related behaviors. The study provides evidence from a Spanish-speaking sample, offering a novel perspective on psychological distress and problematic technology use in contexts that remain underrepresented in the literature.

(JMIR Hum Factors 2026;13:e77167) doi:10.2196/77167

KEYWORDS

cyberpsychology; self-regulation; problematic technology use; smartphone addiction; nomophobia; Colombia; latent profile analysis

XSL•FO
RenderX

Introduction

Background

The global use of information and communication technologies (ICTs) has grown exponentially over the past decade, with more than 67% of the world's population connected as of 2023 [1]. In Latin America, countries, such as Brazil, Argentina, and Colombia, report daily usage exceeding 9 hours, driven by the widespread adoption of mobile devices and the growing popularity of social media platforms [1-3].

In parallel with this digital expansion, studies have reported a rise in problematic behaviors associated with technology use, with documented consequences for both physical and psychological health [4]. Although terms, such as problematic use, digital addiction, and technology dependence, are frequently adopted, there is still no consensus regarding their clinical classification. The literature distinguishes subtypes, including problematic internet use, social media use, smartphone use [5], gaming disorder, and nomophobia [6], which reflect the multidimensional nature of the phenomenon [7]. Some authors argue that problematic technology use shares features with behavioral addictions, social anxiety, and sleep disturbances [8] or compulsive habits [9,10]. However, others suggest that its consequences do not reach the severity of substance addictions and therefore recommend distinguishing among abuse, dependence, and problematic use [11].

Among these, nomophobia has emerged as a relevant subtype of problematic technology use, with direct psychological effects (anxiety, stress, and insomnia). Defined as the irrational fear of being without a mobile phone [6,12], nomophobia is associated with anxiety, insomnia, dependency, and greater psychological vulnerability [13].

This raises the question of which personal protective factors can mitigate the consequences of problematic technology use. Self-regulation, understood as the ability to manage impulses and behaviors, has been studied as a protective factor against both substance-related and behavioral addictions. Evidence suggests that higher self-regulation is associated with greater psychological well-being, healthier digital habits, and reduced procrastination [14-16]. Baumeister and Heatherton [17] proposed that self-regulation failure occurs when individuals are unable to inhibit immediate impulses in favor of long-term goals. In the digital context, this translates to difficulty controlling screen time or compulsive smartphone use, despite the obvious negative consequences.

Empirical evidence supports the notion that deficits in self-regulation are linked to both problematic technology use and broader mental health problems [18-20]. Conversely, higher self-regulation is associated with greater psychological well-being, healthier digital habits, and reduced procrastination [21-24].

Most previous studies have addressed problematic technology use, focusing on average trends and assuming population homogeneity, which obscures subgroup differences. This gap may be explained by the predominant use of variable-centered approaches [25]. In contrast, person-centered approaches, such as latent profile analysis (LPA), enable the identification of specific profiles (subgroups characterized by combinations of self-regulation, ICT use, and nomophobia), thereby revealing differential patterns of psychological distress. For example, it has been demonstrated that self-regulation profiles among students were significantly associated with variations in well-being [26] and personality profiles influenced mental health differently during the pandemic [27].

Aim

This study aimed to identify distinct psychological profiles through LPA, based on the levels of self-regulation, nomophobia, and internet and social media use. By adopting a person-centered approach, the study seeks to uncover subgroups characterized by combined patterns of problematic technology use and self-regulation.

The identified profiles will be examined in relation to age and sex as covariates to determine whether these sociodemographic variables predict membership in specific profiles. In addition, the profiles will be compared in terms of their differential levels of psychological distress, thereby providing a more detailed understanding of whether distinct patterns of problematic technology use and self-regulation have varying impacts on distress scores. This, in turn, provides a basis for designing differentiated and personalized interventions.

Finally, by incorporating a Colombian sample, this study contributes to addressing the scarcity of empirical evidence in Spanish-speaking populations, particularly in Latin America, where digital growth has not been matched by equal progress in preventive and clinical research [28].

Methods

Ethical Considerations

The project was previously approved by the Bioethics Committee of the Konrad Lorenz University Foundation (number: 07-23). This study was conducted in accordance with the ethical principles established in Law 1090 of 2006, particularly Article 13 [29], which states that all professional conduct in psychology must be based on the core principles of the profession, including beneficence and nonmaleficence. In line with these principles, the study used self-report instruments without applying procedures that could pose harm or risk to participants.

This consent process ensured participant autonomy by providing full study information, the opportunity to ask questions, and the option to participate voluntarily. Throughout the research process, participant responses were anonymized, and data were reported in aggregate form to prevent identification of individuals. All identifying data, such as gender, education level, and individual test responses, were replaced with codes. Participation in the study was voluntary and involved no financial compensation, and participants could withdraw at any time without consequences.

Participants

Participants were recruited through a convenience sampling strategy aimed at ensuring heterogeneity of the sample in terms

XSL•FO
RenderX

of age and gender, utilizing both digital and institutional channels. Specifically, the research team distributed the study invitation via social media platforms (Facebook and WhatsApp) and during in-person and virtual classes at 3 universities in Colombia. Individuals interested in participating accessed the study link hosted on Jotform, which included the eligibility criteria and the informed consent form. Recruitment was conducted directly by the research team members. The inclusion and exclusion criteria are presented in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Age between 12 and 60 years
- Colombian nationality
- Possession of a smartphone with internet access
- Ability to provide informed consent

Exclusion criteria

- Diagnosed with a severe condition that limits the ability to answer self-report questionnaires
- Inability to read or understand Spanish
- Lack of internet connectivity that prevents completing an online survey
- Withdrawal of informed consent during participation

Instruments

The MULTICAGE-TIC (a multidomain screening questionnaire based on the CAGE framework) [30] is a questionnaire designed to identify problematic behaviors associated with the use of ICT. Its design is based on the MULTICAGE CAD-4 [31], a screening instrument that assesses compulsive behaviors related to both substance and nonsubstance use and has been widely used to measure behavioral addictions. The MULTICAGE-TIC evaluates 5 areas of ICT use: mobile phones, internet, social media, video games, and instant messaging. A subscale with 4 items represents each area, with each scoring 4 - 8 points. Responses are dichotomous (no=1 point; yes=2 points). According to the suggested interpretation of the authors, usage is classified as nonproblematic (when no response or 1 response is affirmative; 4 - 5 points), at risk (when 2 responses are affirmative; 6 points), and problematic (when 3-4 responses are affirmative; 7 - 8 points). For this study, the areas of internet use and social media use (each with 4 items) were considered representative of overall ICT usage. In previous studies, the instrument has demonstrated adequate reliability levels, with McDonald ω and Cronbach α coefficients above .70 across all subscales. Additionally, its scores have been found to correlate significantly with symptoms of prefrontal dysfunction and psychological distress [32].

The Nomophobia Questionnaire (NMP-Q) [6] measures nomophobia, which is defined as a situational phobia caused by a lack of availability of a smartphone or the thought of not having it, not being able to use it, or losing it. It contains 20 items and 4 dimensions: “communication inability” (6 items), “loss of connection” (5 items), “inability to access information” (4 items), and “renunciation of convenience” (5 items), which can explain 69.5% of the variance. The internal consistency of the factors is between 0.81 and 0.94. Responses are recorded on a 7-point Likert scale, ranging from “1” (strongly disagree) to “7” (strongly agree). The total score ranges from 20 to 140. The Spanish version of the test, which has been validated in a Colombian population [33], was used.

The General Health Questionnaire-12 (GHQ-12) is a brief screening scale designed to detect potential indicators of psychological distress and the risk of mental disorders in the general population [34]. The version used was validated for the Colombian population [35]. It assesses aspects such as emotional symptoms, stress, anxiety, and overall psychological functioning. The scale consists of 12 items with a 4-point Likert response format (range 0 - 3). Higher scores on the GHQ-12 reflect greater psychological distress and poorer mental health. Previous studies have shown a unidimensional structure and significant associations with variables such as experiential avoidance and life satisfaction [35].

The Abbreviated Self-Regulation Questionnaire (CAR-abr) was initially developed by Brown [14]. Its abbreviated Spanish version, which has been validated in a Colombian population [15], was used. It assesses individuals’ general self-regulatory capacity. It comprises 17 items, with Likert-scale response options organized into 4 subscales. The scale has demonstrated solid psychometric performance in terms of both the fit to the Rasch model and the quality of its items and response categories. Previous studies have reported internal consistency values ranging from 0.75 to 0.80, as well as evidence of structural and content validity [15].

Procedure

The project was previously approved by the Bioethics Committee of the Konrad Lorenz University Foundation (number: 07 - 23). The protocol and consent are consistent with the Colombian Code of Ethics [29], which guarantees the anonymous and confidential treatment of the data collected. With institutional authorization, participants were recruited through social networks and public schools in the city. Participants provided informed consent before answering the questionnaires, and guardian assent was obtained for minors.

XSL•FO
RenderX

The instruments were applied virtually and voluntarily, without remuneration.

Data Analysis

All analyses were conducted in RStudio (version 2024.12.1+563; Posit).

To address the possibility of common method bias arising from same-source, same-time self-reports, we planned to conduct 2 complementary tests. First, we applied the Harman single-factor test using a polychoric exploratory factor analysis with all items (self-regulation, nomophobia, internet use, social media use, and GHQ-12 score). Second, we conducted a series of confirmatory factor analyses comparing the following three models: (1) a trait model specifying the 5 latent constructs, (2) a 1-factor model in which all items loaded on a single factor, and (3) an unmeasured latent method construct model that incorporated both the trait factors and a latent method factor. Analyses were performed using the packages psych (for polychoric exploratory factor analysis), lavaan (for confirmatory factor analysis estimation), and semTools (for reliability and method-factor modeling).

All psychological and behavioral indicators were included in an LPA, a person-centered modeling approach designed to identify subgroups of individuals based on similar patterns of responses [36]. The LPA was conducted using 4 continuous indicators: self-regulation, nomophobia, internet use, and social media use. These variables were standardized before analysis to ensure comparability. Models with 1-4 profiles were estimated using the tidyLPA package in R.

Multiple statistical criteria, including the Bayesian information criterion (BIC), Lo-Mendell-Rubin likelihood ratio test (bootstrapped likelihood ratio test [BLRT]), class size, posterior probability, and entropy, guided model selection. The 4-profile model showed the lowest BIC and acceptable entropy (0.85), supporting its selection as the optimal solution. This model yielded 4 distinct psychological profiles based on the interplay between digital behavior and self-regulation.

To evaluate between-group differences in psychological distress across profiles, we first conducted an ANOVA. Pairwise comparisons were examined using Tukey post hoc tests. Then, psychological distress scores were modeled using an ordinary least squares linear regression, with latent classes entered as dummy-coded predictors (class 1 as the reference category). Additionally, a multinomial logistic regression model was estimated, with age and gender as covariates, to predict latent class membership. All visualizations and analyses were performed in R using packages such as ggplot2, nnet, and tidyLPA. Given the exploratory nature of this study and the person-centered approach adopted, no a priori hypotheses were specified regarding the number or structure of latent profiles.

Results

Participant Characteristics

The study recruited 453 participants located in Colombia. Participants were 196 males (43.3%) and 257 females (56.7%), reporting an age between 12 and 57 years, with an average age of 21.03 years (SD 8.41 years). In terms of age distribution, 221 participants were between 12 and 17 years old, 145 were between 18 and 25 years old, and 87 were above 26 years old. Moreover, 70.2% (318/453) of participants were elementary and high school students, while the remaining 29.8% (135/453) had completed technical or professional studies. The majority were single (423/453, 93.4%) and resided in urban areas (417/453, 92.1%). Sociodemographic data are presented in Table 1.

XSL•FO
RenderX

Table . Sociodemographic data.

Variable	Value (N=453), n (%)
Sex	
Female	257 (56.7)
Male	196 (43.3)
Marital status	
Single	423 (93.4)
Married	30 (6.6)
Income[a]	
1	18 (4.0)
2	87 (19.2)
3	291 (64.2)
4	46 (10.2)
5	11 (2.4)
Education level[b]	
Elementary school	84 (18.5)
High school	234 (51.7)
Technical education	54 (11.9)
Undergraduate education	58 (12.8)
Postgraduate education	23 (5.1)

[a]Lower income (1) to higher income (5).

[b]Education level indicates the final level of education.

Psychometric Properties

We obtained internal consistency values. For internet use, the McDonald ω coefficient was 0.61, composite reliability (CR) was 0.76, and average variance extracted (AVE) was 0.45. The social media use scale yielded robust psychometric indicators, with a McDonald ω coefficient of 0.75, a CR of 0.86, and an AVE of 0.62, exceeding the recommended threshold.

The NMP-Q showed the strongest indicators among the instruments evaluated. The McDonald ω coefficient and CR were 0.96 and 0.97, respectively, and both were in the excellent range. Additionally, the AVE reached 0.62, indicating adequate convergent validity. Factor loadings were consistently high (0.65-0.91), confirming the internal coherence and explanatory power of the scale.

The total score of the CAR-abr was used as a global indicator of self-regulation, and the values obtained demonstrated high reliability (ω=0.89; CR=0.90), supporting its internal consistency. Nevertheless, the AVE was 0.35, which is below the minimum criterion, suggesting that items do not share enough common variance to ensure strong convergent validity. Despite this, the reliability level suggested that the measure is stable, although it may be multidimensional in its factorial structure.

The GHQ-12 demonstrated adequate internal consistency, with a total McDonald ω coefficient of 0.75 and a CR of 0.76. The AVE was 0.45, which is slightly below the recommended threshold of 0.50, indicating limitations in convergent validity. This pattern indicates that, although the scale is reliable, some items contribute less common variance, which may reflect the multidimensional nature of psychological distress symptoms.

Common Method Bias

The Harman single-factor test showed that the first factor accounted for 25.78% of the variance, which is well below the conservative threshold of 50% - 60%, indicating the absence of a dominant single factor [37]. In the confirmatory factor analyses, the trait model demonstrated good fit (comparative fit index [CFI]=0.959; Tucker-Lewis index [TLI]=0.957), whereas the 1-factor model showed substantially poorer fit (CFI=0.853; TLI=0.847; root mean square error of approximation [RMSEA]=0.181; standardized root mean square residual=0.155). The unmeasured latent method construct model did not yield any meaningful improvement over the trait model (ΔCFI=0.0001; ΔRMSEA=0.0000), and the variance attributable to the method factor was only 3.1% [38]. Taken together, these results suggest that although a small amount of variance may be attributable to method effects, common method bias did not substantively distort the relationships among the study constructs.

Descriptive Statistics of Indicator Variables

The variables used to estimate the latent profiles included self-regulation, nomophobia, internet use, and social media use. Self-regulation showed a mean score of 58.21 (SD 9.58), while the nomophobia score ranged from 20 to 140 points, with a

XSL•FO
RenderX

mean of 61.58 (SD 27.95). For "internet use" and "social media use," usage has been classified as nonproblematic (when no response or 1 response is affirmative; 4 - 5 points), at risk (when 2 responses are affirmative; 6 points), and problematic (when 3-4 responses are affirmative; 7 - 8 points). Details are provided in Table 2.

Table . Descriptive statistics.

Variable	Mean (SD)	Minimum value	Maximum value
Self-regulation	58.21 (9.58)	25	84
Nomophobia	61.58 (27.95)	20	140
Internet use	5.86 (1.30)	4	8
Social media use	5.45 (1.41)	4	8

Associations Between Variables

Following previous guidelines [39], correlations below 0.30 were considered small, those at 0.30-0.50 were considered moderate, and those above 0.50 were considered large. In psychological and health sciences, even small associations may be meaningful at the population level [40]. The results of the associations were in the expected direction. Correlations between variables revealed moderate relationships between nomophobia and internet use (r=0.378) and social media use (r=0.361). There were significant negative correlations between self-regulation and internet use (r=−0.314), and between self-regulation and social media use (r=−0.211). Moreover, higher levels of self-regulation were associated with lower use of these ICTs.

Higher GHQ-12 scores indicated more psychological distress. A negative moderate correlation was found between self-regulation and psychological distress scores (r=−0.506), suggesting that higher self-regulation was associated with lower psychological distress. Nomophobia showed a small positive correlation with psychological distress (r=0.261). Moreover, there were positive correlations between internet use and psychological distress (r=0.327), and between social media use and psychological distress (r=0.188) (Table 3), indicating that more problematic technology use correlates positively with higher distress scores.

Table . Correlation values between variables.

Pair	Correlation	Bootstrapped 95% CI[a]
Self-regulation versus nomophobia	−0.208[b]	−0.301 to −0.113
Self-regulation versus internet use	−0.314[b]	−0.405 to −0.218
Self-regulation versus social media use	−0.211[b]	−0.310 to −0.114
Self-regulation versus psychological distress	−0.506[b]	−0.564 to −0.442
Nomophobia versus internet use	0.378[b]	0.290 to 0.465
Nomophobia versus social media use	0.361[b]	0.270 to 0.445
Nomophobia versus psychological distress	0.261[b]	0.164 to 0.361
Internet use versus social media use	0.592[b]	0.519 to 0.658
Internet use versus psychological distress	0.327[b]	0.238 to 0.411
Social media use versus psychological distress	0.188[b]	0.086 to 0.285

[a]With 1000 iterations in the percentile method.

[b]Correlation is significant (P<.001; 2-tailed).

LPA (Model Without Covariates)

An LPA was conducted to test solutions ranging from 1 to 4 classes. The 4-class model was identified as the best fit based on the BIC. Models with 1 to 4 classes were evaluated using fit indices such as the BIC, entropy, and the BLRT.

The 4-class model showed the lowest BIC (4710), an acceptable entropy (0.85), and a significant improvement over the 3-class model (BLRT P<.01), indicating adequate class separation. Additionally, all classes represented more than 5% of the sample (minimum=17%), suggesting a stable and well-differentiated solution. Details are provided in Table 4.

XSL•FO
RenderX

Table . Fit statistics and class distribution for competing latent profile analysis models.

Class model[a]	LogLik[b]	AIC[c]	BIC[d]	SABIC[e]	Entropy	BLRT[f] *P* value	n_min (%)[g]	n_max (%)[h]
1	−2569	5154	5187	5161	1.000	—[i]	100	100
2	−2391	4807	4861	4819	0.817	.0099	34.4	65.6
3	−2356	4748	4822	4765	0.745	.0099	26.5	43.0
4[j]	−2285	4615	4710	4637	0.852	.0099	17.2	36.9

[a]Number of latent profiles or classes estimated in the model.
[b]LogLik: log-likelihood.
[c]AIC: Akaike information criterion.
[d]BIC: Bayesian information criterion.
[e]SABIC: sample size–adjusted BIC.
[f]BLRT: bootstrapped likelihood ratio test.
[g]Percentage of the sample assigned to the smallest class.
[h]Percentage of the sample assigned to the largest class.
[i]Not applicable.
[j]Selected model.

The entropy value of the 4-class model was 0.852, indicating good class separation. The BLRT yielded *P*<.01 for model comparisons, supporting the inclusion of additional classes.

Description of the Resulting Profiles or Classes

Figure 1 presents the standardized mean scores (*z* scores) for each latent profile, representing the deviation of each class mean from the overall sample mean. For instance, class 1 scored about 1 SD above the sample mean in self-regulation (+1.0 SD) while remaining nearly 1 SD below the mean in nomophobia and ICT use (–1 SD). In contrast, class 2 showed slightly below-average scores across all variables, including self-regulation, nomophobia, and ICT use (–0.3 to –0.5 SD). Class 3 profiles had self-regulation about 1 SD below the mean (–1 SD), and nomophobia and ICT use more than 1 SD above the mean (+1 to +1.5 SD). Finally, class 4 exhibited moderately low self-regulation (–0.5 SD) in combination with average levels of ICT use and nomophobia (0 SD).

Figure 1. Four-class solution of the latent profile analysis (*z* scores across study variables).



The qualitative interpretation of the latent classes suggested that class 1 was characterized by high self-regulation, low nomophobia, and low ICT use. Class 2 exhibited a more balanced but generally lower profile, characterized by slightly below-average levels of self-regulation, nomophobia, and ICT use. Class 3 was characterized by low self-regulation combined

XSL•FO
RenderX

with high nomophobia and ICT use. Finally, class 4 presented a mixed profile, with moderately low self-regulation and average levels of nomophobia and ICT use. These standardized differences highlight that the most distinctive contrast was present between class 1 (high regulation, low ICT use) and class 3 (low regulation, high ICT use), whereas classes 2 and 4 represented more intermediate or mixed profiles.

Analysis With Predictor Covariates (Age and Sex)

A multinomial model with covariates was also used. Age (as a continuous variable) and sex (1=female; 2=male) were included as latent class membership predictor covariates (class 1 as reference; see Table 5). Age significantly increased the odds of belonging to class 4 (β=−0.70, SE=0.20; *z*=−3.43; *P*<.001; odds ratio [OR] 0.50, 95% CI 0.34-0.74), indicating that younger participants were more likely to belong to this profile. The association with class 3 had a *P* value of .07, and no effect was observed for class 2 (*P*=.55). Regarding gender, males were less likely than females to belong to class 2 (β=−0.68, SE=0.26; *z*=−2.67; *P*=.007; OR 0.51, 95% CI 0.31-0.83). No other gender contrasts reached statistical significance (*P*>.29). Overall, demographic covariates showed selective but meaningful differentiation of latent profiles.

Table . Results of the multinomial model with age and sex as predictors of latent class membership.

Comparison (reference: class 1)	Predictor[a]	β (log-odds, standardized)[b]	SE	*z*	*P* value	OR[c] (95% CI)
Class 2 versus class 1	Age_z	−0.068	0.115	−0.59	.55	0.93 (0.75-1.17)
Class 2 versus class 1	Male versus female	−0.680	0.255	−2.67	.007	0.51 (0.31-0.83)
Class 3 versus class 1	Age_z	−0.252	0.137	−1.84	.07	0.78 (0.59-1.02)
Class 3 versus class 1	Male versus female	−0.277	0.263	−1.06	.29	0.76 (0.45-1.27)
Class 4 versus class 1	Age_z	−0.695	0.203	−3.43	<.001	0.50 (0.34-0.74)
Class 4 versus class 1	Male versus female	0.277	0.280	0.99	.32	1.32 (0.76-2.28)

[a]Age was z-standardized (mean 0, SD 1); sex was coded as 1=female and 2=male.

[b]Coefficients represent multinomial log-odds relative to class 1.

[c]OR: odds ratio.

External Validation Using the Psychological Distress Variable

An ANOVA test was conducted to evaluate whether there were significant differences in psychological distress scores across the latent classes identified in the LPA model (Table 6).

Table . ANOVA results.

Source	Degrees of freedom	Sum of squares	Mean square	*F* test (*df*)	*P* value
Class (factor)	3	2212	737.4	18.32	.001[a]
Residuals	449	18,069	40.2	—[b]	—

[a]Statistically significant differences in psychological distress levels across the latent classes.

[b]Not applicable.

Subsequently, a Tukey post hoc test was performed to examine pairwise differences (Table 7).

XSL•FO
RenderX

Table . Tukey post hoc test results.

Class comparison	Mean difference (est.)	95% CI	Adjusted *P* value[a]
Class 2 versus class 1	5.08	3.09 to 7.06	<.001[b]
Class 3 versus class 1	1.68	−0.43 to 3.79	.17
Class 4 versus class 1	4.62	2.38 to 6.86	<.001[b]
Class 3 versus class 2	−3.40	−5.68 to −1.12	<.001[b]
Class 4 versus class 2	−0.46	−2.86 to 1.95	.96
Class 4 versus class 3	2.94	0.44 to 5.45	.01[c]

[a]*P* values were adjusted using the Tukey method.

[b]*P*<.001.

[c]*P*<.05.

The Tukey test results indicated that class 2 differed from class 1 by an average of 5.08 points in psychological distress scores, suggesting that class 2 experienced more distress. Class 4 showed an average difference of 4.62 points compared with class 1, indicating high psychological distress. Class 3 had lower distress than class 2, with a significant difference of –3.40 points. Additionally, class 4 exhibited more distress than class 3, with a significant difference of 2.94 points. Pairwise comparisons confirmed differences across classes.

Finally, a regression model was also estimated to predict psychological distress based on latent class membership (Table 8).

Table . Regression model results[a].

Coefficient[b]	Estimate	SE	*t* test (*df*)	*P* value
Intercept (class 1)	10.35	0.49	21.09 (449)	<.001[c]
Class 2	5.08	0.77	6.59 (449)	<.001[c]
Class 3	1.68	0.82	2.05 (449)	.04[d]
Class 4	4.62	0.87	5.31 (449)	<.001[c]

[a]Higher values reflect greater psychological distress (General Health Questionnaire-12).

[b]Adjusted R^2=10.3%. Coefficients indicate mean differences in distress scores relative to class 1.

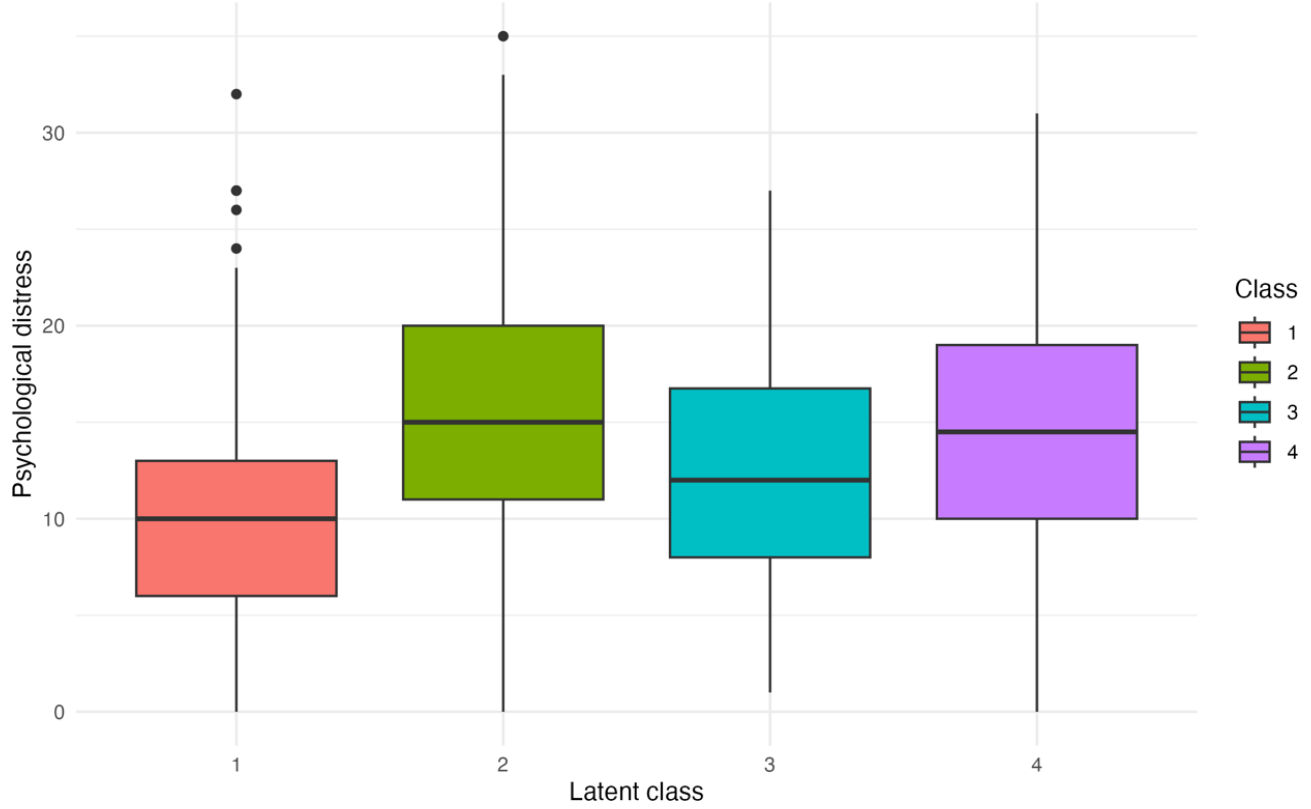
[c]*P*<.001.

[d]*P*<.05.

Classes 2 and 4 showed higher psychological distress scores than class 1, indicating more average psychological symptoms (since higher scores on this scale reflect increased symptoms or distress). Class 3 showed a slight deterioration, although the effect was more moderate. Details are presented in Figure 2.

XSL•FO
RenderX

Figure 2. Psychological distress comparison across latent profile analysis classes.



The results of combining the 4 classes, along with age, sex, and psychological distress, are shown in Table 9.

Table . Integrated results of profiles (classes 1 - 4), age, sex, and psychological distress.

Class	Nomophobia	Problematic technology use	Self-regulation	Age[a]	Sex[a]	Psychological distress
1	Low (–1 SD)	Low (–1 SD)	High (+1 SD)	Balanced distribution	More men ↑	Lowest psychological distress
2	Below average (–0.3 to –0.5 SD)	Below average (–0.3 to –0.5 SD)	Below average (–0.3 to –0.5 SD)	Balanced distribution	More women ↑	Highest psychological distress
3	High (+1 to +1.5 SD)	High (+1 to +1.5 SD)	Low (–1 SD)	Younger	More women ↑	Significantly lower distress than class 2; not significantly different from class 1
4	Average (0 SD)	Average (0 SD)	Moderately low (–0.5 SD)	Younger ↑↑	More women ↑↑	Significantly higher distress than classes 1 and 3

[a]Arrows indicate relative predominance in group composition based on descriptive comparisons. A single arrow indicates a higher proportion relative to other classes, whereas a double arrow indicates a more pronounced predominance. These symbols are descriptive and do not represent statistical significance or effect size.

Discussion

Principal Results

This study contributes to the field of cyberpsychology by providing empirical evidence on psychological profiles associated with problematic technology use and their relationship with psychological distress. Through a person-centered approach, we identified differentiated profiles characterized by distinct patterns of self-regulation, nomophobia, and ICT use, highlighting both consistent associations and apparent contradictions with prior research.

The LPA identified 4 distinct classes that significantly differed in distress scores. Class 1, which was characterized by high self-regulation and low problematic technology use, showed the lowest psychological distress. In contrast, classes 2 and 4 exhibited higher psychological distress, which was significantly worse than that of class 1. Class 2 was especially relevant because, despite reporting relatively low ICT use and moderate-to-low nomophobia, it exhibited the highest

XSL•FO
RenderX

psychological distress, suggesting that other unmeasured factors may contribute to this vulnerability. Meanwhile, class 4, involving the youngest individuals, showed low self-regulation combined with high ICT use and nomophobia, also resulting in high psychological distress. Interestingly, the distress in class 3 was significantly lower than that in class 2, was comparable to that in class 1, and was lower than that in class 4. These findings indicate that the relationship between problematic technology use and distress is not monotonic, and the interplay of multiple factors must be considered.

Comparison With Prior Work

There are some important aspects to consider. First, the results confirm that self-regulation is a central protective factor for mental health, even in the presence of high ICT use or nomophobia. This is consistent with studies indicating that it is not the absolute amount of use that explains psychological distress, but rather individuals' ability to regulate their digital behaviors [41,42]. Accordingly, class 1, with high self-regulation, showed the lowest psychological distress, while profiles with low self-regulation (classes 3 and 4) experienced greater distress.

Second, class 2 concentrated on individuals with the highest psychological distress despite lower ICT use and lower levels of nomophobia. This apparently counterintuitive finding suggests that other factors not assessed in this study, such as neuroticism, loneliness, low self-esteem, lack of social support, and dysfunctional coping strategies [43-45], may influence the outcomes. The identification of this subgroup underscores the value of LPA, which, unlike variable-centered approaches, can reveal hidden profiles and inform the design of tailored interventions.

Third, the results suggest that age plays a protective role, but it is not linear. While prior research has emphasized adolescents and young adults as especially vulnerable to problematic social media use and nomophobia [46,47], our findings reveal nuances. Class 3, which consisted mainly of young adults with low self-regulation and high ICT use, did not show severe psychological distress, while class 4, which also consisted of younger individuals with a similar usage pattern, showed psychological distress. One possible explanation is that, in class 3, intensive ICT use may have been socially normalized in the age groups, reducing its negative effects [48,49]. In contrast, class 4 combined low self-regulation with more compulsive or maladaptive usage patterns, resulting in significantly more distress.

Fourth, gender showed selective effects. Men were less likely to belong to class 2 compared with class 1, while women were proportionally more represented in classes 3 and 4. These profiles combined higher ICT use and higher nomophobia, which aligns with previous research reporting higher levels of nomophobia and smartphone-related stress among women [49,50].

Finally, the profile comparison reinforces that problematic technology use is a multifactorial phenomenon that cannot be explained solely by usage intensity. A central contribution of this study is the analysis of a Colombian sample, which remains underrepresented in the international literature, thereby expanding evidence from Latin American contexts. A notable finding is that class 2 exhibited the highest psychological distress despite moderate ICT use, suggesting that additional individual or contextual factors—beyond usage frequency—may contribute to psychological outcomes. In addition, young women showed higher levels of nomophobia, greater intensive social media use, and a statistical association with increased psychological distress.

In this context, the role of nomophobia deserves attention. Although initially conceptualized as an emerging phenomenon linked to negative psychological outcomes, the profiles identified in this study suggest that its effects may not operate in a linear manner. For instance, class 3 showed high nomophobia and intensive ICT use combined with only moderate distress levels. This pattern suggests that nomophobia may interact with individual characteristics or contextual factors rather than exerting uniform effects across users. Understanding nomophobia as a component within broader behavioral configurations may help explain why similar levels of ICT engagement are associated with different psychological outcomes across groups.

Limitations

Among the main limitations of this study is the exclusive use of self-report measures, which may have caused bias in the assessment of both technology use and psychological distress. To evaluate common method bias, we conducted the Harman single-factor test and a series of confirmatory factor analyses, which suggested minimal common method bias effects. Nevertheless, future research should incorporate external corroborating measures or social desirability scales to enhance validity.

Although the sample was diverse, a more representative sample of older adults would improve the generalizability of the findings. Additionally, longitudinal designs are necessary to establish causal relationships among the studied variables. Future studies should also incorporate qualitative methodologies to complement quantitative data and evaluate interventions that strengthen self-regulation as a preventive strategy in digital mental health.

Moreover, other psychological and contextual factors, such as resilience, coping strategies, social support, and specific motivations for technology use, should be considered in future analyses, as they may provide valuable insights into psychological vulnerability to problematic technology use [51]. Cross-cultural comparative studies are also recommended, as the dynamics of ICT use and its mental health implications may vary depending on sociocultural contexts. Complementing self-reports with objective indicators (eg, screen time and activity logs) can be considered to enhance measurement accuracy. Ultimately, analyzing the differentiated impact of various ICT types, such as social media, video games, and educational platforms, may provide further insights into their distinct implications for psychological well-being.

XSL•FO
RenderX

Conclusions

The results of this study offer a comprehensive, data-driven perspective on how different patterns of technology use and personal resources shape distinct psychological risk profiles. Strengthening self-regulation has emerged as a promising pathway to promote healthier ICT use, particularly among younger populations and within educational or community contexts.

Our findings underscore the need for differentiated and personalized interventions that enhance self-regulation and promote adaptive technology use, with special attention to vulnerable groups, such as class 2 (individuals with high psychological distress despite low ICT use) and class 4 (young individuals with low self-regulation and high nomophobia). Future studies should expand this framework by incorporating additional psychological and contextual factors, such as social support, neuroticism, and coping strategies, to better explain heterogeneity in the psychological consequences of technology use patterns.

Authors' Contributions

AGU contributed to the study idea, design, analysis, and initial writing. MAC contributed to analysis, discussion writing, draft revision, review, and the final version in English. EAM-C contributed to data collection and draft revision. All authors contributed to the article and approved the submitted version.

Conflicts of Interest

None declared.

References

1. Global offline population steadily declines to 2.6 billion people in 2023. International Telecommunication Union. 2023. URL: https://www.itu.int/itu-d/reports/statistics/2023/10/10/ff23-internet-use/ [accessed 2024-08-19]
2. Marco general de los medios en españa [Article in Spanish]. : Asociación para la Investigación de Medios de Comunicación (AIMC); 2022 URL: https://www.aimc.es/a1mc-c0nt3nt/uploads/2022/01/marco2022.pdf [accessed 2026-01-28]
3. Digital 2024: global overview report. : DataReportal; 2024 URL: https://datareportal.com/reports/digital-2024-global-overview-report [accessed 2026-01-28]
4. Shannon H, Bush K, Villeneuve PJ, Hellemans KG, Guimond S. Problematic social media use in adolescents and young adults: systematic review and meta-analysis. JMIR Ment Health 2022 Apr 14;9(4):e33450. [doi: 10.2196/33450] [Medline: 35436240]
5. Chemnad K, Aziz M, Belhaouari SB, Ali R. The interplay between social media use and problematic internet usage: four behavioral patterns. Heliyon 2023 May;9(5):e15745. [doi: 10.1016/j.heliyon.2023.e15745] [Medline: 37159716]
6. Yildirim C, Correia AP. Exploring the dimensions of nomophobia: development and validation of a self-reported questionnaire. Comput Human Behav 2015 Aug;49:130-137. [doi: 10.1016/j.chb.2015.02.059]
7. Basel AM, McAlaney J, Skinner T, Pleva M, Ali R. Defining digital addiction: key features from the literature. Psihologija 2020;53(3):237-253. [doi: 10.2298/PSI191029017A]
8. Chen Y, Li S, Tian Y, Li D, Yin H. Problematic social media use may be ruining our sleep: a meta-analysis on the relationship between problematic social media use and sleep quality. Int J Ment Health Addiction 2024:1-36. [doi: 10.1007/s11469-024-01407-9]
9. Marks I. Behavioural (non-chemical) addictions. Br J Addict 1990 Nov;85(11):1389-1394. [doi: 10.1111/j.1360-0443.1990.tb01618.x] [Medline: 2285832]
10. Potenza M. Perspective: behavioural addictions matter. Nature New Biol 2015 Jun 25;522(7557):S62. [doi: 10.1038/522S62a] [Medline: 26107100]
11. Chóliz Montañés M, Villanueva Silvestre V. Evaluación de la adicción al móvil en la adolescencia [Article in Spanish]. Rev Esp Drogodepend 2011;36(2):165-183 [FREE Full text]
12. Daraj LR, AlGhareeb M, Almutawa YM, Trabelsi K, Jahrami H. Systematic review and meta-analysis of the correlation coefficients between nomophobia and anxiety, smartphone addiction, and insomnia symptoms. Healthcare (Basel) 2023 Jul 19;11(14):2066. [doi: 10.3390/healthcare11142066] [Medline: 37510507]
13. Notara V, Vagka E, Gnardellis C, Lagiou A. The emerging phenomenon of nomophobia in young adults: a systematic review study. Addict Health 2021 Apr;13(2):120-136. [doi: 10.22122/ahj.v13i2.309] [Medline: 34703533]
14. Brown JM. Self-regulation and the addictive behaviors. In: Miller WR, Heather N, editors. Treating Addictive Behaviors, 2nd edition: Plenum Press; 1998:61-73. [doi: 10.1007/978-1-4899-1934-2_5]
15. Garzón-Umerenkova A, de la Fuente Arias J. Evaluación de la autorregulación comportamental: un estudio de validación [Article in Spanish]. Electron J Res Educ Psychol 2024;22(63):347-372. [doi: 10.25115/ejrep.v22i63.9438]
16. de la Fuente J, Cubero I, Sánchez-Amate MC, Peralta FJ, Garzón A, Fiz Pérez J. The adolescent's competency for interacting with alcohol as a determinant of intake: the role of self-regulation. Front Psychol 2017;8:1800. [doi: 10.3389/fpsyg.2017.01800] [Medline: 29123492]

XSL•FO
RenderX

17. Baumeister RF, Heatherton TF. Self-regulation failure: an overview. Psychol Inq 1996 Jan;7(1):1-15. [doi: 10.1207/s15327965pli0701_1]
18. Akınci T. Determination of predictive relationships between problematic smartphone use, self-regulation, academic procrastination and academic stress through modelling. IJPE 2021 Feb 1;17(1):35-53. [doi: 10.29329/ijpe.2021.329.3]
19. Jiang Z, Liu J, Zhao Y, Lv C, Zhu L. Problematic mobile phone use and cognitive failures: the mediating role of emotional distress and the moderating role of coping style. Curr Psychol 2023 Sep;42(27):24025-24034. [doi: 10.1007/s12144-022-03392-5]
20. Khan NA, Khan AN, Moin MF. Self-regulation and social media addiction: a multi-wave data analysis in China. Technol Soc 2021 Feb;64:101527. [doi: 10.1016/j.techsoc.2021.101527] [Medline: 34898757]
21. Garzón Umerenkova A, de la Fuente Arias J, Gil Flores J. Relationship between academic procrastination, well-being, and grades: the mediational role of self-regulation and bad habits. J Posit Psychol Wellbeing 2022;6(2):1247-1262 [FREE Full text]
22. Garzón Umerenkova A, Gil Flores J, de la Fuente J. Demographic, academic and personal traits associated with three types of procrastination in university students [Article in Spanish]. Bordon Revista De Pedagogia 2020;72(1):49-65. [doi: 10.13042/Bordon.2020.01.69513]
23. Hong W, Liu RD, Ding Y, Sheng X, Zhen R. Mobile phone addiction and cognitive failures in daily life: the mediating roles of sleep duration and quality and the moderating role of trait self-regulation. Addict Behav 2020 Aug;107:106383. [doi: 10.1016/j.addbeh.2020.106383] [Medline: 32200196]
24. Mann T, de Ridder D, Fujita K. Self-regulation of health behavior: social psychological approaches to goal setting and goal striving. Health Psychol 2013 May;32(5):487-498. [doi: 10.1037/a0028533] [Medline: 23646832]
25. Collins LM, Lanza ST. Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences: John Wiley & Sons, Inc; 2009. [doi: 10.1002/9780470567333]
26. Chen J, Lin CH, Chen G. Extramural ICT factors impact adolescents' academic performance and well-being differently: types of self-regulated learners also matter. Educ Inf Technol 2024 Oct;29(15):20459-20491. [doi: 10.1007/s10639-024-12642-x]
27. Li M, Ahmed MZ, Hiramoni FA, Zhou A, Ahmed O, Griffiths MD. Mental health and personality traits during COVID-19 in China: a latent profile analysis. Int J Environ Res Public Health 2021 Aug 17;18(16):8693. [doi: 10.3390/ijerph18168693] [Medline: 34444441]
28. Digital 2024: Colombia. : DataReportal; 2024 URL: https://datareportal.com/reports/digital-2024-colombia [accessed 2025-08-19]
29. Ley 1090 de 2006. Función Pública. 2006. URL: https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=66205 [accessed 2025-08-19]
30. Pedrero Pérez EJ, Ruiz Sánchez de León JM, Rojo Mota G, et al. Tecnologías de la Información y la Comunicación (TIC): abuso de Internet, videojuegos, teléfonos móviles, mensajería instantánea y redes sociales mediante el MULTICAGE-TIC [Article in Spanish]. Adicciones 2018;30(1):19-32. [doi: 10.20882/adicciones.806]
31. Pedrero Pérez EJ, Rodríguez Monje MT, Gallardo Alonso F, Fernández Girón M, Pérez López M, Chicharro Romero J. Validación de un instrumento para la detección de trastornos de control de impulsos y adicciones: el MULTICAGE CAD-4 [Article in Spanish]. Trastor Adict 2007 Dec;9(4):269-278. [doi: 10.1016/S1575-0973(07)75656-8]
32. Pedrero-Pérez E, Morales-Alonso S, Gallardo-Arriero V, Blázquez-Rollón L, Ruiz-Sánchez JM. Impulsivity-compulsivity axis in the abuse of information and communication technologies (ICT) from the perspective of the Research Domain Criteria Project (RDoC). Behav Psychol 2021 Sep 1;29(2):399-415. [doi: 10.51668/bp.8321211s]
33. Bernal Acevedo JD, García Bolaños C, Garzón Umerenkova A. Propiedades psicométricas del Cuestionario de Nomofobia (NMP-Q) en una muestra de adolescentes y adultos colombianos [Thesis] [Article in Spanish]. : Fundación Universitaria Konrad Lorenz; 2024 URL: https://biblioteca.konradlorenz.edu.co/bib/25005 [accessed 2026-01-28]
34. Goldberg DP. The Detection of Psychiatric Illness by Questionnaire: A Technique for the Identification and Assessment of Non-Psychotic Psychiatric Illness: Oxford University Press; 1972.
35. Ruiz FJ, García-Beltrán DM, Suárez-Falcón JC. General Health Questionnaire-12 validity in Colombia and factorial equivalence between clinical and nonclinical participants. Psychiatry Res 2017 Oct;256:53-58. [doi: 10.1016/j.psychres.2017.06.020] [Medline: 28623768]
36. Williams GA, Kibowski F. Latent class analysis and latent profile analysis. In: Jason LA, Glenwick DS, editors. Handbook of Methodological Approaches to Community-Based Research: Qualitative, Quantitative, and Mixed Methods: Oxford University Press; 2016:143-152. [doi: 10.1093/med:psych/9780190243654.003.0015]
37. Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 2003 Oct;88(5):879-903. [doi: 10.1037/0021-9010.88.5.879] [Medline: 14516251]
38. Chin WW, Thatcher JB, Wright RT. Assessing common method bias: problems with the ULMC technique. MIS Q 2012 Sep 1;36(3):1003-1019. [doi: 10.2307/41703491]
39. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd edition: Lawrence Erlbaum Associates; 1988. URL: https://utstat.utoronto.ca/~brunner/oldclass/378f16/readings/CohenPower.pdf [accessed 2026-01-28]

XSL•FO
RenderX

40. Evans JD. Straightforward Statistics for the Behavioral Sciences: Brooks/Cole Publishing Company; 1996.
41. Troll ES, Friese M, Loschelder DD. How students' self-control and smartphone use explain their academic performance. Comput Human Behav 2021 Apr;117:106624. [doi: 10.1016/j.chb.2020.106624]
42. van Deursen AJAM, Bolle CL, Hegner SM, Kommers PAM. Modeling habitual and addictive smartphone behavior. Comput Human Behav 2015 Apr;45:411-420. [doi: 10.1016/j.chb.2014.12.039]
43. Charzyńska E, Sussman S, Atroszko PA. Profiles of potential behavioral addictions' severity and their associations with gender, personality, and well-being: a person-centered approach. Addict Behav 2021 Aug;119:106941. [doi: 10.1016/j.addbeh.2021.106941] [Medline: 33915392]
44. Servidio R. Fear of missing out and self-esteem as mediators of the relationship between maximization and problematic smartphone use. Curr Psychol 2023;42(1):232-242. [doi: 10.1007/s12144-020-01341-8]
45. Hasani J, Chashmi SJE, Firoozabadi MA, Noory L, Turel O, Montag C. Cognitive emotion regulation mediates the relationship between big-five personality traits and internet use disorder tendencies. Comput Human Behav 2024 Mar;152:108020. [doi: 10.1016/j.chb.2023.108020]
46. Lam LT. Risk factors of Internet addiction and the health effect of internet addiction on adolescents: a systematic review of longitudinal and prospective studies. Curr Psychiatry Rep 2014 Nov;16(11):508. [doi: 10.1007/s11920-014-0508-2] [Medline: 25212714]
47. Lozano-Blasco R, Robres AQ, Sánchez AS. Internet addiction in young adults: a meta-analysis and systematic review. Comput Human Behav 2022 May;130:107201. [doi: 10.1016/j.chb.2022.107201]
48. Anderson M, Jiang J. Teens, social media and technology 2018. Pew Research Center. 2018. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2025-08-19]
49. Sahu M, Gandhi S, Sharma MK, Marimuthu P. Mobile phone addiction among children and adolescents: a systematic review. J Addict Nurs 2019;30(4):261-268. [doi: 10.1097/JAN.0000000000000309] [Medline: 31800517]
50. Argumosa-Villar L, Boada-Grau J, Vigil-Colet A. Exploratory investigation of theoretical predictors of nomophobia using the Mobile Phone Involvement Questionnaire (MPIQ). J Adolesc 2017 Apr;56:127-135. [doi: 10.1016/j.adolescence.2017.02.003] [Medline: 28231484]
51. Wickord LC, Quaiser-Pohl C. Psychopathological symptoms and personality traits as predictors of problematic smartphone use in different age groups. Behav Sci (Basel) 2022 Jan 25;12(2):20. [doi: 10.3390/bs12020020] [Medline: 35200272]

Abbreviations

AVE: average variance extracted
BIC: Bayesian information criterion
BLRT: bootstrapped likelihood ratio test
CAR-abr: Abbreviated Self-Regulation Questionnaire
CFI: comparative fit index
CR: composite reliability
GHQ-12: General Health Questionnaire-12
ICT: information and communication technology
LPA: latent profile analysis
NMP-Q: Nomophobia Questionnaire
OR: odds ratio
RMSEA: root mean square error of approximation
TLI: Tucker-Lewis index

Edited by A Kushniruk, Y Quintana; submitted 08.May.2025; peer-reviewed by C Annan, I Arpaci, P Larionow; revised version received 03.Dec.2025; accepted 14.Dec.2025; published 11.Feb.2026.

Please cite as:
Garzón Umerenkova A, Alba Cabañas M, Malpica-Chavarria EA
Mental Health Profiles Based on Self-Regulation and Technology Use in the Digital Era in a Spanish-Speaking Sample: Latent Profile Analysis
JMIR Hum Factors 2026;13:e77167
URL: https://humanfactors.jmir.org/2026/1/e77167
doi:10.2196/77167

© Angélica Garzón Umerenkova, Marisleidy Alba Cabañas, Erika Andrea Malpica-Chavarria. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 11.Feb.2026. This is an open-access article distributed under the terms of the

XSL•FO
RenderX

Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Online Tiered Screening for Mental Health Problems Among Refugees in Sweden: Validation Study

Jennifer Meurling[1], MSc; Elisabet Rondung[1], PhD; Youstina Demetry[2], MSc; Anahita Geranmayeh[2], MSc; Anna Leiler[1], PhD; Gerhard Andersson[3], PhD; Anna Bjärtå[1], PhD

[1]Department of Education, Psychology and Social Work, Mid Sweden University, Östersund, Sweden

[2]Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden

[3]Department of Behavioral Sciences and Learning, Linköping University, Linköping, Sweden

Corresponding Author:
Jennifer Meurling, MSc
Department of Education, Psychology and Social Work
Mid Sweden University
Kunskapens väg 1
Östersund, 831 24
Sweden
Phone: 46 101428589
Email: jennifer.meurling@miun.se

Abstract

Background: Refugees and asylum-seekers commonly experience numerous adverse and traumatic events and are therefore at increased risk of mental health problems. Despite the high need for mental health interventions, services tend to be underused by refugees and asylum-seekers, and various barriers compromise access. Digital, efficient screening, adapted for these groups, could facilitate initial assessment and increase accessibility to mental health services. We developed an internet-based tiered screening procedure (i-TAP) aiming to identify clinically relevant symptoms of major depressive disorder (MDD), anxiety disorder, posttraumatic stress disorder, and insomnia disorder among individuals with a refugee background. The i-TAP is an adaptive procedure with 3 tiers aiming to identify general mental distress in Tier 1, differentiate between symptoms in Tier 2, and assess the severity of symptoms in Tier 3. Each tier additionally functions as a gateway to further assessment, as a negative outcome terminates the procedure.

Objective: The purpose of this study was to evaluate the diagnostic test accuracy of the i-TAP, using structured clinical assessments as the reference standard.

Methods: In this prospective study, 70 adult participants with a refugee background, literate in Arabic, Dari, Farsi, or Swedish, and residing in Sweden, completed the i-TAP on tablets and participated in a subsequent structured diagnostic interview.

Results: It has been shown that the i-TAP could identify 91.7% (33/36) of individuals assessed with any psychiatric disorder, and correctly identified 82.1% of all positive cases of MDD, anxiety disorder, posttraumatic stress disorder, and insomnia disorder, with few false negative assessments. Overall test accuracy of the i-TAP ranged between 77.1% and 84.3%, depending on disorder. The tiered design could reduce item burden while maintaining accuracy. A vast majority of participants rated the user experience as positive. In this sample, 36/70 (51.4%) individuals were assessed with one or more psychiatric disorders and comorbidity was high.

Conclusions: The i-TAP may be a valid, efficient, and feasible screening tool for the identification of common psychiatric disorders among individuals with a refugee background in Sweden. The i-TAP could be implemented as a first screener in various settings, including online and in-person clinical practices. The digital, adaptive, multilingual format could facilitate early assessment and increase the availability of mental health services for refugees and asylum-seekers.

(JMIR Hum Factors 2026;13:e82763) doi:10.2196/82763

KEYWORDS

asylum-seekers; diagnostic test accuracy; digital mental health; eHealth; mHealth; online assessment; refugees; tiered screening; validation

XSL•FO
RenderX

Introduction

Background

According to the United Nations High Commissioner for Refugees (UNHCR), there were an estimated 123.2 million forcibly displaced people in the world by the end of 2024 [1]. This number is expected to rise due to ongoing conflicts and wars globally. UNHCR reserves the term refugees for forcibly displaced people who, due to violence, war, conflict, or persecution, have crossed international borders to seek safety in another country, and the term asylum-seeker for people who have applied for refugee status and are awaiting a decision [2]. In Sweden, asylum-seekers have come mainly from Syria, Afghanistan, Iraq, Eritrea, Somalia, and Iran in the years 2015-2025 [3]. Since 2022, refugees from Ukraine have been treated under the European Union Temporary Protection Directive.

In addition to having fled their homes, many refugees and asylum-seekers have endured numerous stressful and potentially traumatic events, increasing the risk of psychological distress and various mental disorders [4-6]. When arriving in their host countries, refugees are further exposed to postmigration stressors, such as language barriers, discrimination, long asylum processes, and economic as well as social challenges, all of which are associated with poor mental health [4,6,7].

Posttraumatic stress disorder (PTSD), major depressive disorder (MDD), and anxiety disorders are highly prevalent among refugees [8,9], with even higher rates observed among asylum-seekers [10,11]. Symptoms persist over time [9,12] and often co-occur [11]. Furthermore, insomnia disorder is highly prevalent [13,14] and various sleep disturbances are reported as problematic [15]. Although prevalence rates vary between studies of refugee mental health, consensus in the field establishes significantly elevated levels of psychiatric symptoms and disorders among refugee groups, including asylum-seekers, compared to the general population [8,9,16]. Despite the high prevalence of mental health issues, mental health services tend to be underused [17,18]. This discrepancy, or treatment gap, can be understood through barriers frequently reported by both help-seeking refugees and asylum-seekers, and service providers. These barriers include language obstacles and the need to use interpreters, difficulties navigating the health care system, and fear of stigma [17,19], along with practical issues and financial concerns [20], all of which affect accessibility to relevant services.

Given that the mental health of individuals who have fled their homelands is affected by adversity experienced before, during, and after the flight [6], the increased risk of mental health problems persists over time, also long after resettlement [9,12]. This underscores the importance of identifying those in need of interventions, regardless of the duration of displacement. In a recent review [21], systematic mental health screening with at least 2 screening occasions is recommended for recently resettled refugees and asylum-seekers. Early and repeated mental health screening could increase access to and provision of treatment, ultimately reducing unnecessary suffering from untreated symptoms. This emphasizes the need for validated, efficient, and feasible methods to identify common mental health problems among individuals who have fled their homelands.

Offering internet-based mental health screening could increase accessibility of mental health services by addressing several of the aforementioned barriers. For instance, internet-based and smartphone-assisted options provide flexibility and privacy, can be time-saving, and can easily be offered in multiple languages [22,23]. Although the literature on this topic is limited, digital mental health screening tools have previously been used among refugees and asylum-seekers [24], with evidence supporting their acceptability, feasibility, and validity for these populations [25,26]. Furthermore, a digital format enables the use of adaptive hierarchical screening models, which allow simultaneous screening of multiple psychiatric symptoms and can reduce the overall respondent burden while maintaining high accuracy [27,28].

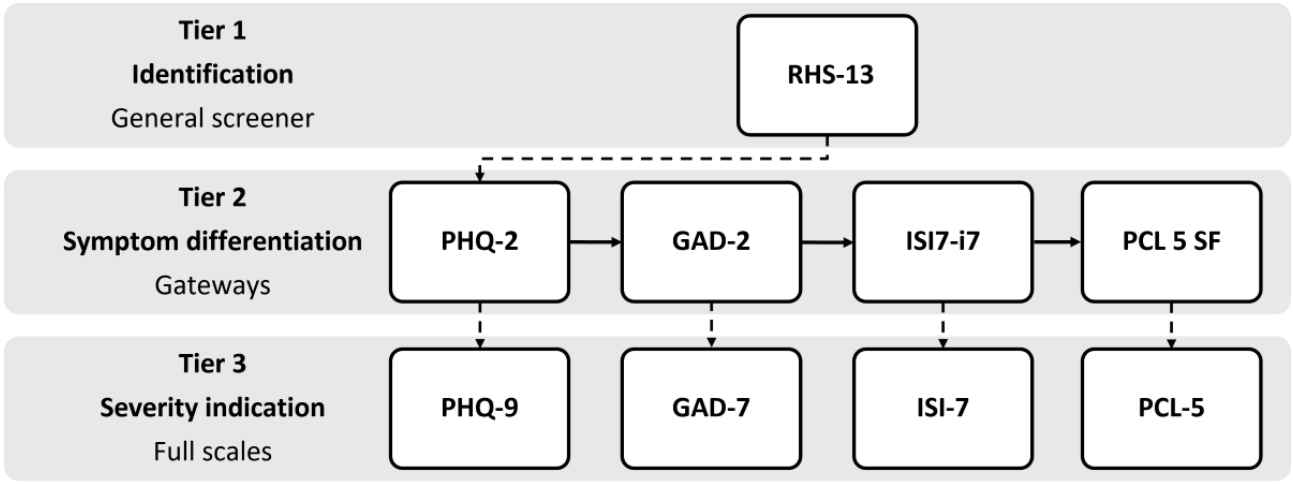
Recognizing the potential technical as well as social advantages of online mental health screening for refugees and asylum-seekers, we set out to design and evaluate an internet-based tiered assessment procedure (i-TAP). The i-TAP is tailored for refugee groups residing in Sweden, and specifically designed to identify clinically relevant symptoms of depression, anxiety, PTSD, and insomnia.

The i-TAP

In a previous study by Meurling et al [28], the instruments and the procedure constituting the i-TAP were evaluated, yielding a model designed for optimal psychometric performance, along with promising results regarding the efficiency, including a reduced item burden and accuracy of the procedure. The i-TAP is a 3-tiered screening procedure that adapts to the respondent's answers in each tier. The first tier is highly sensitive, aiming to identify general mental distress while preventing further assessment of individuals without psychiatric symptoms. Thus, a negative outcome in Tier 1 terminates the procedure, while a positive outcome forwards the respondent to Tier 2. The second tier comprises brief symptom scales and differentiates between symptoms of depression, anxiety, PTSD, and insomnia. It also serves a gateway function, forwarding individuals with a positive outcome on one or more scales to further assessment in Tier 3, while terminating the procedure for respondents with a negative outcome on all Tier 2 scales. In the third tier, the full scales corresponding to the outcome of Tier 2 are presented to the respondent, with the purpose of indicating severity and identifying clinically relevant symptoms of depression, anxiety, PTSD, and insomnia. The items already answered in Tier 2 have been removed from each full scale in Tier 3. The tiered design of the i-TAP with the included instruments is depicted in Figure 1. In both the initial study [28] and this one, the i-TAP was provided in Arabic, Dari, Farsi, English, and Swedish, based on the largest groups of refugees and asylum-seekers in Sweden at the time the study commenced [29].

XSL•FO
RenderX

Figure 1. Illustration of the internet-based tiered screening procedure, i-TAP, with the symptom scales used in each tier. GAD-2: Generalized Anxiety Disorder-2; GAD-7: Generalized Anxiety Disorder-7; Insomnia Severity Index-7, item 7; ISI-7: Insomnia Severity Index-7; PCL 5: PTSD (posttraumatic stress disorder) Checklist for DSM-5 (Diagnostic and Statistical Manual of Mental Disorders [Fifth Edition]); PCL-5 SF: PTSD (posttraumatic stress disorder) Checklist for DSM-5 (Diagnostic and Statistical Manual of Mental Disorders [Fifth Edition]) Short Form; PHQ-2: Patient Health Questionnaire-2; PHQ-9: Patient Health Questionnaire-9; RHS-13: Refugee Health Screener-13.



In Meurling et al [28], the Tier 1 and Tier 2 scales were evaluated using the Tier 3 scales as the reference standard. We also applied receiver operating characteristic analysis to identify gateway cutoffs for optimal performance. This resulted in an accurate screening tool (total precision 86%) that could identify most moderate symptoms of PTSD, depression, anxiety, and/or insomnia (96.2%). These results and the fact that only the first 2 tiers of the model were tested prompted further evaluation of the full 3-tiered model. The developmental process of the i-TAP is described in further detail in Meurling et al [28].

Aim

The aim of this study is to test the performance of the full i-TAP. Specifically, we seek to evaluate the criterion validity by assessing diagnostic test accuracy of the full tiered model, using structured clinical assessments of psychiatric disorders as the reference standard. Our objective is to determine the extent to which the i-TAP is able to identify MDD, anxiety disorder, PTSD, and insomnia disorder among individuals with a refugee background, residing in Sweden. We want to emphasize that the heightened risk for these disorders follows experiences of flight and resettlement. The target group for the i-TAP is therefore focused on the refugee experience, irrespective of current legal status. Accordingly, we henceforward use the term individuals with a refugee background to describe our study population.

Methods

Study Context and Design

This study and the development of the i-TAP are part of the SAHA project, a collaboration between Mid Sweden University, Linköping University, and Karolinska Institute, with the overarching aim of developing digital mental health solutions, including tailored interventions, for individuals with a refugee background. We conducted this prospective cross-sectional validation study with individuals with a refugee background residing in Sweden, collecting data from June to October 2022.

Recruitment and Participants

Participants were recruited through convenience sampling in various nonclinical settings (adult education for immigrants, language cafés, nongovernmental organizations, and at an asylum housing facility). Recruitment took place on site with the help of local personnel and interpreters. After being informed about the study orally, interested individuals could approach immediately or later via email or phone. The inclusion criteria were refugee background, aged 18 years or older, literacy in Arabic, Dari, Farsi, Swedish, or English, and currently living in Sweden. Being a refugee or having a refugee background was herein specified as having fled to Sweden due to war, conflict, persecution, threat, or similar reasons, and was self-defined by the participants. This criterion was carefully explained during recruitment and verified before starting each study procedure, to ensure inclusion of participants with a refugee experience. Individuals were thus included regardless of residence status, and we have not differentiated analyses based on that.

Data Collection and Material

Overview

Data collection was carried out by 11 persons trained in clinical psychology and with practical experience of psychiatric assessment, herein referred to as psychologists. The assessments were conducted in each respondent's preferred language. Five of the psychologists were bilingual and 6 psychologists used authorized interpreters via telephone to conduct the assessments. To facilitate translation, the assessment material was sent to interpreters in advance. All psychologists lived in Sweden and were fluent in Swedish. To ensure methodological equivalence, all psychologists received training in the full procedure prior to data collection.

Data collection included digital screening with the i-TAP, followed by a structured clinical interview. The whole procedure was carried out by a psychologist on a single occasion in a private space adjacent to each recruitment site. Each occasion

XSL•FO
RenderX

lasted a maximum of 2 hours and was preceded by oral and written informed consent from the participant.

Digital Screening

Participants completed the i-TAP independently, online via tablets. For this study, we also included background questions and feedback questions regarding the experience of completing the i-TAP. The psychologists (and telephone interpreters) were available to answer any questions but were instructed not to sit next to or help the participants fill in the survey. The results of the i-TAP were blinded to the psychologists.

Structured Diagnostic Interview

The structured diagnostic interview targeted the same disorders as the i-TAP, namely insomnia, PTSD, MDD, and anxiety disorders (panic disorder, agoraphobia, social anxiety disorder, and generalized anxiety disorder [GAD]). Before terminating the clinical interview, the psychologist invited the participant to add any information or raise other topics of their choice. To conclude the procedure, the psychologist summarized the diagnostic interview and gave participants feedback and recommendations based on their assessment. All participants received a brief written summary of the clinical assessment, recommendations (if applicable), a folder with contact information to available local and national health care services, and a SEK 99 (US $9.7) gift card to a supermarket. We offered to guide participants in help-seeking, for example, by showing websites, finding opening hours and contact information, and exploring maps. If an immediate risk was identified, this was handled according to a predefined safety protocol described below. To ensure confidentiality, data were pseudonymized. In cases where the procedure was terminated before completion, participants were also offered the information folder and gift card.

Each clinical interview was discussed at team meetings, referred to as clinical conferences, before determining the final clinical assessment of each participant. The psychologists were still blinded to the results of the i-TAP during these discussions. The clinical conferences promoted both discussion and unanimity in the final assessments.

Index Test

The i-TAP builds on 56 items from 5 validated scales, described in detail below. The symptom scales were selected based on psychometric properties, cross-cultural validity or previous use in refugee populations (or a combination thereof). All scales have previously been evaluated by us for the current population [28,30]. For a detailed description of the scales, translation, and psychometric properties, please see our previous publication from the SAHA project [28]. All scales use Likert response alternatives. For the analyses in this study, responses were dichotomized using cutoffs on each scale. Considering that i-TAP aims for high diagnostic sensitivity yet being an efficient screening procedure, we have applied the sensitivity model proposed in our first study [28] for this study. That is, we have applied cutoffs to identify as many true positives as possible without compromising accuracy.

The following instruments and cutoff values for moderate symptoms were used: for identification of general psychological distress in Tier 1, we used the Refugee Health Screener-13 (RHS-13) [31,32] with cutoff ≥11. For symptoms of MDD, we used the Patient Health Questionnaire-2 (PHQ-2) [33], with cutoff ≥2 in Tier 2 and the Patient Health Questionnaire-9 (PHQ-9) [34] with cutoff ≥10 in Tier 3. Symptoms of anxiety disorders (including panic disorder, social anxiety disorder, and GAD), were measured using the Generalized Anxiety Disorder-2 (GAD-2) [35], with cutoff ≥2 in Tier 2 and the Generalized Anxiety Disorder-7 (GAD-7) [36,37] with cutoff ≥10 in Tier 3. Selection of the anxiety scales was based on very good psychometric properties when screening for GAD, panic disorder, and social anxiety disorder [37], which enables the use of 1 instead of 3 separate anxiety scales, reducing the item burden for the respondents. This aligns with the aim of the i-TAP to efficiently detect clinically relevant symptoms of a range of anxiety disorders. Symptoms of PTSD were measured with the PTSD Checklist for *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders* [*Fifth Edition*]) Short Form (PCL-5 SF) [38], using cutoff ≥5 in Tier 2 and the PTSD Checklist for *DSM-5* (PCL-5) [39] using cutoff ≥32 in Tier 3. To identify symptoms of insomnia, we used the Insomnia Severity Index-7 (ISI-7) [40], item 7 with cutoff ≥2 in Tier 2 and the full scale in Tier 3. For ISI-7, we previously used a conservative cutoff for detecting moderate to severe symptoms of insomnia (≥14). However, 2 more sensitive cutoffs have been proposed, ≥8 and ≥11 [41,42]. We evaluated the performance of all 3 cutoffs and concluded that cutoff 11 was best suited for the purpose of the i-TAP. Please see Table S1 in Multimedia Appendix 1 for detailed results of this analysis. All Tier 2 instruments are composed of items from the Tier 3 instruments. To avoid posing the same questions twice, these items were not repeated in Tier 3. Instead, we included the item responses from Tier 2 in the analyses of the full scales in Tier 3.

To assess user experience of the i-TAP, we formulated 3 items about answering questions about mental health online. The first assessed if participants had a negative or positive attitude (1-10 with 1=very negative and 10=very positive), the second assessed if the i-TAP was difficult or easy to use (1-10 with 1=very difficult and 10=very easy), and the third assessed if the questions in the i-TAP were difficult or easy to understand (1-5 with 1=very difficult, 2=difficult, 3=neither difficult nor easy, 4=easy, and 5=very easy to understand).

Reference Standard

The reference standard was a clinical diagnosis of MDD, anxiety (panic disorder, agoraphobia, social anxiety disorder, and GAD), PTSD, or insomnia according to the clinical assessment. The assessment of clinical diagnoses relied on a structured diagnostic interview, including sections investigating insomnia disorder, trauma experience, PTSD, MDD, panic disorder, agoraphobia, social anxiety disorder, and GAD, in the presented order. Structured diagnostic interviews are generally accepted as the gold, or reference standard in clinical research [43].

The principal instrument of the structured diagnostic interview was the Mini International Neuropsychiatric Interview 7.0.0 (M.I.N.I.) [44]. M.I.N.I. is a short, structured psychiatric

XSL•FO
RenderX

interview based on *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*) and *ICD-10* (*International Statistical Classification of Diseases, Tenth Revision*) diagnostic criteria and has demonstrated excellent test and interrater reliability [45]. It was originally designed for epidemiological and multicenter studies and is now used extensively in both research and clinical settings, as well as in global mental health settings [46] and with refugee populations [16,47]. The M.I.N.I. is designed to be sensitive [44] and has repeatedly been proven so in comparison with expert clinical diagnoses [43,46]. In this study, we used the diagnostic sections for PTSD, major depressive episode, panic disorder, agoraphobia, social anxiety disorder, and GAD. No official validated translations of the M.I.N.I. were available. For Arabic, we used a previously evaluated translation [46], and the bilingual psychologists made translations to Dari and Farsi. As insomnia is not included in the M.I.N.I., we designed a module for this specific diagnosis comprising 6 questions responding to the diagnostic criteria in *DSM-5* [48], matching the format in the M.I.N.I. to facilitate administration. We formulated brief questions to simplify interpreter-assisted assessments. The complete insomnia assessment is depicted in Multimedia Appendix 1. Furthermore, we adapted the M.I.N.I. to the target population by replacing the suicidality module and the criterion A assessment for PTSD (see details below).

To investigate potentially traumatic events (PTEs), we used the Refugee Trauma History Checklist (RTHC), a measure developed for adult refugees and previously validated in a sample of Syrian refugees in Sweden [49]. The RTHC is designed to assess the most common refugee-related PTEs, before, during, and after flight, in a nonintrusive way. The checklist is composed of 2 questions, where the respondent is asked if they have experienced any of 7 or other PTEs before and after they left their home (war at close quarters, forced separation from family or close friends, loss or disappearance of family members or loved ones, physical violence or assault, witnessing physical violence or assault, torture, sexual violence, and other frightening situation where you felt your life was in danger). Participants were not asked to elaborate or describe reported events further. We used the RTHC as a replacement for item H1 in the M.I.N.I., which corresponds to the criterion A assessment for PTSD in *DSM-5*; thus, a positive RTHC outcome regarding any traumatic life event was followed by continued PTSD assessment with the M.I.N.I. This adaptation was made against previous research pointing toward the importance of broadening the trauma criterion and description in *DSM-5* for individuals with a nonwestern background [50].

To assess suicidality, we used the Columbia Suicide Severity Rating Scale Screening Version (C-SSRS Screen) [51,52] as a structured interview. The C-SSRS Screen has 6 items with a similar structure to the M.I.N.I., is validated as an initial step to identify individuals with short-term suicide risk [51], and has previously been used in refugee populations [53]. A positive outcome was handled according to a predefined safety protocol, including immediate consultation with relevant health services, together with participants, for thorough suicide risk evaluation. If suggested, we offered to accompany the participant to the care facility. We included suicidality as part of the clinical assessment to identify any individuals in need of immediate clinical attention; however, suicidality was not part of this validation study. Against this, and to further adapt the structured assessment to the target group, we decided to replace the extensive and sometimes linguistically complicated suicidality module in the M.I.N.I. with a shorter, more comprehensible screener.

Data Treatment and Analysis

Sample size requirements were estimated based on expected prevalence and desired sensitivity. Based on previous studies of refugee mental health [8,9,16], we expected a high prevalence (>30%) of moderate symptoms of each disorder and an even higher prevalence based on total outcomes from the i-TAP. According to Bujang and Adnan [54], a sample size of 67 is adequate to determine accuracy (80% power and significance level <.05) based on a minimum prevalence of 30%, and the aim of 80% sensitivity we set for the i-TAP.

The structured diagnostic interview was assessed and summarized as positive (“diagnosis”) or negative (“no diagnosis”) for each diagnosis. A diagnosis is herein defined as fulfilling diagnostic criteria for a psychiatric disorder according to clinical assessment with M.I.N.I. and the structured clinical interview for insomnia. Panic disorder, agoraphobia, social anxiety disorder, and GAD in M.I.N.I. were collapsed into one variable, anxiety disorder, for the analysis.

Descriptive measures were summarized as frequencies and proportions or means and SDs depending on variable type. Internal consistency was calculated for the full scales included in the i-TAP using Cronbach alpha. Intercorrelations were assessed using Pearson *r*.

We tested the criterion validity of the i-TAP by comparing the outcome of the complete i-TAP procedure for each symptom category (depression, anxiety, PTSD, and insomnia) with the corresponding clinical diagnoses according to the reference standard. We used Cohen κ as a statistically adjusted measure of concordance between 2 measures [55]. Interpretation of κ values was guided by Landis and Koch [56] suggested labels: <0 no agreement, 0-0.20 slight, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 substantial, and 0.81-1.0 perfect agreement. Diagnostic test accuracy, including sensitivity, specificity, positive and negative predictive values [57], and the number of false negatives and false positives, was calculated for each tier and symptom category, using clinical diagnosis from the structured clinical interview as the reference standard.

All individuals had complete data on all scales and were included in the analysis. Analyses were performed using IBM SPSS (version 29).

Ethical Considerations

The study protocol has been approved by the Swedish Ethical Review Authority (2020-00214). This study was planned and performed in accordance with the ethical standards of the 1975 Helsinki declaration and its amendments (World Medical Association, 2001). Written informed consent was obtained from all participants, in their preferred language, prior to study participation. Participants were carefully informed that

XSL•FO
RenderX

participation was voluntary and data confidential, that they could terminate their participation at any time, and that participation could not affect the asylum process. To protect the privacy of participants, data are presented at the group level, and potentially identifying details regarding participants or data collection have been omitted.

Results

Sample

In total, 85 individuals registered interest to participate in the study; however, 6 were lost due to nonattendance. Out of the 79 that initiated assessments, 7 did not complete for reasons such as low literacy, not having a refugee background, or practical circumstances. We completed assessments with 72 participants. Due to technical issues, data were incomplete for 2 respondents, rendering a final sample of 70 participants. Against the power calculation, our sample size of 70 was deemed sufficient.

Clinical and Sociodemographic Characteristics

Out of the 70 analyzed assessments, 31 were conducted in Arabic, 14 in Dari, 15 in Farsi, and 10 in Swedish. Participants were 31 females and 39 males, between 19 and 69 years old (mean 39.1, SD 13.16), and 31.4% (22/70) lived at an asylum facility, 50% (35/70) had their own housing, and 18.6% (13/70) lived together with family or friends, see Table 1 for sociodemographic characteristics of the sample.

XSL•FO
RenderX

Table 1. Sample characteristics of participants (N=70).

Characteristic	Participants, n (%)
Nationality	
Syria	25 (35.7)
Afghanistan	19 (27.1)
Iran	12 (17.1)
Palestine	5 (7.1)
Other (n<5)[a]	9 (12.9)
Sex	
Female	31 (44.3)
Male	39 (55.7)
Marital status	
Single	17 (24.3)
Married or partner	44 (62.9)
Divorce or separated	6 (8.6)
Widowed	2 (2.9)
Other	1 (1.4)
Time in Sweden	
<1 year	10 (14.3)
1-3 years	17 (24.3)
4-6 years	19 (27.1)
7-10 years	24 (34.3)
Education	
University Master of Arts	10 (14.3)
University Bachelor of Arts	10 (14.3)
High school	25 (35.7)
Primary school	19 (27.1)
Vocational qualification	3 (4.3)
Other	3 (4.3)
Residence permit	
No residence permit	19 (27.1)
Temporary	18 (25.7)
Permanent	33 (47.1)

[a]Iraq, Eritrea, Tajikistan, Somalia, Sudan, and Kurdistan.

Each scale in the i-TAP demonstrated excellent internal consistency, with Cronbach α of 0.93 for RHS-13, 0.92 for PHQ-9, 0.92 for GAD-7, 0.96 for PCL-5, and 0.92 for ISI-7. All scales were intercorrelated, *r* ranging from 0.70 to 0.89 (all *P*<.001). Regarding symptom assessment, the mean for RHS-13 was 21.61 (SD 13.95; 95% CI 18.29-24.92), for PHQ-9, 9.86 (SD 7.20; 95% CI 8.14-11.57), for GAD-7, 8.33 (SD 6.43; 95% CI 6.79-9.86), for PCL-5, 29.99 (SD 22.74; 95% CI 24.56-35.41), and the mean for ISI-7 was 11.64 (SD 7.96; 95% CI 9.75-13.54).

Prevalence of clinical diagnoses of MDD, anxiety disorder, PTSD, and insomnia disorder was high in the sample, see Table 2 for prevalence estimates for each disorder by the i-TAP and the clinical assessment, respectively. All participants reported experiencing at least one potentially traumatic event. According to the structured diagnostic interview, 51.4% (36/70) met diagnostic criteria for one or more disorders.

XSL•FO
RenderX

Table 2. Prevalence estimates assessed with the structured diagnostic interview and the i-TAP[a], respectively (N=70)[b].

Disorder	Clinical assessment (diagnosis), n (%)	i-TAP (moderate symptoms), n (%)
Major depressive disorder	20 (28.6)	33 (47.1)
Anxiety disorder (any)[c]	17 (24.3)	25 (35.7)
Panic disorder	5 (7.1)	—[d]
Agoraphobia	0 (0)	—
Social anxiety disorder	3 (4.3)	—
Generalized anxiety disorder	14 (20.0)	—
PTSD[e]	20 (28.6)	27 (38.6)
Insomnia disorder	26 (37.1)	32 (45.7)
Any[f]	36 (51.4)	41 (58.6)

[a]i-TAP: internet-based tiered screening procedure.

[b]Multiple diagnoses per individual are possible. Clinical assessment=assessment results of the structured diagnostic interview, i-TAP=individuals who scored positive in Tier 1, Tier 2, and above the cutoff for moderate symptoms in Tier 3.

[c]Positive case for any anxiety disorder.

[d]Not applicable.

[e]PTSD: posttraumatic stress disorder.

[f]Positive case for any disorder.

Comorbidity was generally high. The clinical assessment showed that 23 individuals (32.9%) fulfilled criteria for 2 or more diagnoses, of which 8 individuals met criteria for 3 diagnoses, and 8 for all 4 diagnoses. Comorbidity rates per identified disorder are depicted in Table 3. Among the 13 individuals who fulfilled criteria for one diagnosis only, insomnia was most prevalent (n=7).

Table 3. Comorbidity of disorders diagnosed with the structured diagnostic interview (N=70)[a].

Diagnosis	Number of participants	Comorbid MDD[b], n (%)	Comorbid anxiety disorder, n (%)	Comorbid PTSD[c], n (%)	Comorbid insomnia disorder, n (%)
MDD	20	—[d]	—	—	—
Anxiety disorder	17	12 (17.1)	—	—	—
PTSD	20	14 (20.0)	12 (17.1)	—	—
Insomnia disorder	26	15 (21.4)	13 (18.6)	13 (18.6)	—

[a]Percentages are calculated based on the total sample (N=70).

[b]MDD: major depressive disorder.

[c]PTSD: posttraumatic stress disorder.

[d]Not applicable.

Concordance of the i-TAP and the Structured Diagnostic Interview

The i-TAP identified 33 of the 36 (91.7%) individuals with a clinical diagnosis according to the structured diagnostic interview. The i-TAP correctly identified 19/20 cases of MDD, 13/17 cases of anxiety, 16/20 cases of PTSD, and 21/27 cases of insomnia disorder (ie, true positives) in the current sample. κ values indicated moderate concordance between the i-TAP and the clinical assessment for the specific disorders, and diagnostic test accuracy was between 77.1% and 78.6% for all 4 diagnoses. Regarding overall concordance, accuracy was 84.3% for the identification of any disorder and the overall concordance was substantial, see Table 4 for detailed results.

XSL•FO
RenderX

Table 4. Concordance of the i-TAP[a] and the clinical assessment, calculated for each i-TAP disorder with clinical diagnosis according to the structured diagnostic interview as the reference standard.

i-TAP diagnosis	n_{pos}[b]	Cohen κ	Sensitivity, % (95% CI)	Specificity, % (95% CI)	PPV[c], % (95% CI)	NPV[d], % (95% CI)	Accuracy, % (95% CI)
Depression	33	.56	95.0 (85.5-100)	72.0 (59.6-84.5)	57.6 (40.7-74.5)	97.3 (92.0-100)	78.6 (68.9-88.2)
Anxiety	25	.46	76.5 (56.3-96,7)	77.4 (66.1-88.7)	52.0 (32.4-71.6)	91.1 (82.8-99.4)	77.1 (67.3-86.9)
PTSD[e]	27	.53	80.0 (62.5-97.5)	78.0 (66.5-89.5)	59.3 (40.8-77.8)	90.7 (82.0-99.3)	78.6 (68.9-88.2)
Insomnia	32	.53	80.8 (64.6-95.4)	75.0 (62.2-87.8)	65.6 (49.1-82.1)	86.8 (76.0-97.6)	77.1 (67.3-86.94)
Any[f]	41	.68	91.7 (82.69-100)	76.5 (62.25-90.75)	80.5 (68.4-92.6)	89.7 (71.1-97.5)	84.3 (75.8-92.8)

[a]i-TAP: internet-based tiered screening procedure.

[b]n_{pos}: number of positive cases for each diagnosis.

[c]PPV: positive predictive value.

[d]NPV: negative predictive value.

[e]PTSD: posttraumatic stress disorder.

[f]Positive case for any diagnosis.

Precision of the i-TAP

To investigate to what level the i-TAP was able to correctly assess the prevalence of disorders on an individual level, we have calculated the number of correct assessments (ie, true positives and true negatives) performed by the i-TAP over all 4 disorders and within each individual. That is, each participant could have 0 to 4 correct assessments of diagnosis or no diagnosis.

The i-TAP could correctly assess 54.3% (n=38) of all participants. These individuals were thus correctly positively or negatively assessed on all 4 disorders. Fourteen (20.0%) individuals were correctly assessed on 3 of the 4 disorders, of which 11 (78.6%) of the failed assessments were false positives. The remaining 18 participants had 2 or less correctly identified assessments. Equally for these individuals, a vast majority of the failed assessments were false positives (77.1%). A total of 8 individuals were assessed as false positives by the i-TAP without any diagnosis.

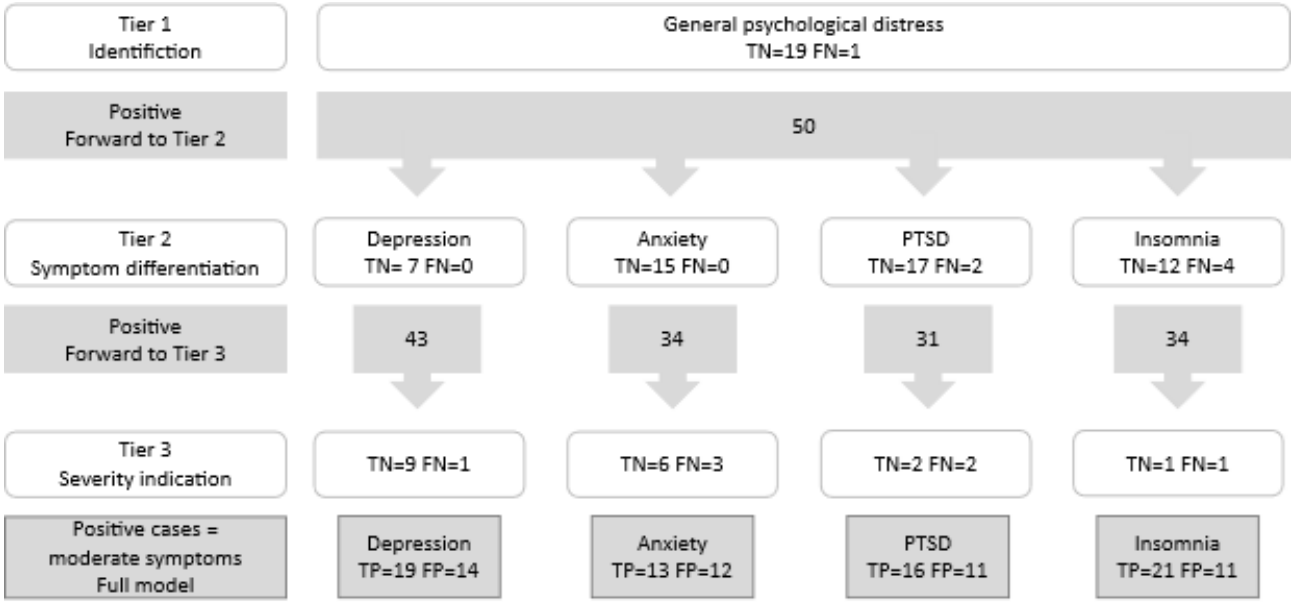
Regarding false negative assessments, a total of 11 individuals were affected as the i-TAP failed to identify 1 case of MDD, 4 cases of GAD, 4 cases of PTSD, and 5 cases of insomnia disorder. However, only 3 individuals with a diagnosis according to the reference standard were entirely missed, as the other 8 false negative cases were identified as positive on another diagnosis.

A Tiered Model

To calculate the tiered model, we used all data (all respondents answered all scales) to simulate the effects of the tiers in the i-TAP. In Tier 1, identification of symptoms, 50 individuals screened positive, indicating psychological distress in 71.4% (50/70) of the participants. Tier 1 showed very high sensitivity (97.2%), low specificity (55.9%), and correctly excluded 19 individuals (ie, true negatives) from further assessment; however, 1 individual with clinical anxiety diagnosis was falsely excluded (ie, a false negative). Tier 2, differentiation between symptoms, also exhibited high sensitivity (range 78.6%-100%) and specificity ranging between 54.0% and 74.0% for the 4 disorders. See Figure 2 for an illustration of the culling effects of the hierarchical design. Detailed results of the psychometric evaluation of each tier and disorder are depicted in Table S1 in Multimedia Appendix 2.

XSL•FO
RenderX

Figure 2. The culling effects of the hierarchical design of the internet-based tiered screening procedure (i-TAP). FN: false negatives (incorrectly excluded cases); FP: false positives (incorrectly identified cases); PTSD: posttraumatic stress disorder; TN: true negatives (correctly excluded cases); TP: true positives (correctly identified cases).



User Experience

A majority of the sample had a positive attitude toward answering questions about mental health online, with 51% (36/70) scoring 10/10 (very positive; mean 8.41, SD 2.08; range 3-10) and only 4.3% (3/70) scoring in the negative range (ie, <5). Moreover, the method was perceived as easy (mean 8.47, SD 1.98; range 3-10), with 50% (35/70) scoring 10/10 (very easy), whereas 5.7% (4/70) scored it somewhat difficult (ie, <5). Participants scored the questions as easy to understand, with 61.4% (43/70) scoring 5/5 (very easy; mean 4.37, SD 0.90; range 1-5) and only 1 (1.4%) participant answered that it was very difficult (<3). There were no significant differences between language groups.

Discussion

Principal Findings

The results from this study show that the i-TAP was very efficient in detecting individuals assessed as having a clinical diagnosis (91.7%) and could correctly identify a total of 82.1% of all positive cases of MDD, anxiety disorder, PTSD, and insomnia disorder, while keeping a low rate of false negatives. The overall accuracy of the i-TAP, ranging between 77.1% and 78.6% for the specific diagnoses and rising to 84.3% for any diagnosis, could be considered commendable, considering the multisymptom screening function and the emphasis on sensitivity. Furthermore, the psychometric performance of the i-TAP is comparable to the performance of the individual instruments included [41,50,58,59], showing that a hierarchical design does not compromise accuracy. In line with previous findings [9,15], our results show a sample with high prevalence of MDD, anxiety disorders, PTSD, and insomnia disorder, as well as a high rate of comorbid disorders. The higher prevalence rates found by the i-TAP compared to the clinical assessment were expected, since the use of self-rating instruments is known to overestimate prevalence [60-62] as they identify symptoms, not disorders. Since the i-TAP is a screening tool, we have prioritized sensitivity over specificity in this validation study. This comes at the cost of a higher number of false positives, which have affected the overall diagnostic test accuracy negatively, and which contribute to the overestimation of prevalence. However, prioritizing specificity would increase the risk of overlooking individuals in need. Weighing sensitivity against specificity for the i-TAP is based on the aim to identify as many individuals as possible, and is built upon the standpoint that, in this case, false positives are less harmful than missing someone in need of mental health services. Against this aim, the i-TAP performed very well. However, the relationship between sensitivity and specificity must be considered for each setting and context, with regard to the purpose and resources for implementation of the i-TAP.

Using a tiered hierarchical model for screening enables the reduction of the number of items an individual responds to, referred to as item burden. Reducing the item burden reduces the time and effort required to complete a screening, thereby increasing the probability of both initiating and completing a screening. Using the i-TAP, an individual with no symptoms would have a reduction in item burden with 69.8% compared to a screening using the full scales, and an individual with, for example, clinically relevant symptoms of anxiety would have 37.2% reduction in item burden. However, for the individuals with comorbid symptoms of all disorders, there is, in fact, an increased item burden, adding the RHS-13 [32] in Tier 1. Despite this, we have weighed the benefits of including the RHS-13 as higher than the costs, in this case, in item burden for individuals with symptoms of all 4 diagnoses. The RHS-13 is a suitable first screener, with culturally sensitive items including somatic symptoms, designed for the general refugee population [31]. In this study, it was shown to be highly sensitive, detecting all but one individual and case of clinical diagnosis, thus serving the purpose of the first tier superbly.

XSL•FO
RenderX

The high prevalence of comorbid diagnoses observed in this sample, as evidenced in several prior studies, for example [11,13], underscores the importance of broad initial screening for refugees. Moreover, the high comorbidity might affect the accuracy of the i-TAP for the specific disorders. Several items and symptoms are similar or shared between instruments and diagnoses, making differential diagnosis difficult in early screening. Regarding anxiety specifically, a positive outcome on the i-TAP can imply any, as well as comorbid, anxiety disorders. Furthermore, the i-TAP falsely identified 8 individuals without any disorder. Hence, the i-TAP cannot be used as a diagnostic tool. Nevertheless, the i-TAP could correctly identify a clear majority of the positive cases for each diagnosis, and all but 3 individuals with a clinical diagnosis. Thus, even if not always precise, the i-TAP can be considered a reliable screening tool for the identification of individuals in need of mental health services.

Regarding the user experience of the i-TAP, results were promising, with a vast majority reporting being positive toward the format and finding the questions easy to complete, indicating acceptability of the i-TAP; however, this needs further investigation.

Clinical Implications

The results from this study show that the i-TAP could be implemented as a first screener in various clinical settings, such as primary and psychiatric care, schools, health programs, and online mental health services, for individuals with a refugee background in Sweden. In practice, the i-TAP could be useful as a guide to assessment, for symptom-specific interventions, or to tailor support and treatment. For example, combining the i-TAP with a tailored internet-based cognitive behavioral therapy intervention [63], where treatment module provision is based on symptoms, would facilitate automatic adaptation without requiring help-seeking individuals to answer lengthy questionnaires unnecessarily.

In a setting where differentiation of symptoms is less important, the i-TAP could serve as a general screener to identify individuals with any clinically relevant symptoms, and would do so with higher specificity and efficiency compared to using the RHS-13 or full scales only. Applying either approach, the i-TAP's tiered design could prove especially beneficial in settings with resource and time constraints [25,27]; however, unsuitable in settings where complete data on measures are essential (eg, clinical trials).

Concerning implementation, it is critical to emphasize that the i-TAP is not designed to be a diagnostic instrument and should not be used as such in any context. It is intended as a screening tool, requiring further clinical assessment by a mental health professional in the case of a positive outcome. Moreover, the i-TAP is highly sensitive, and acknowledging the stigma around mental health reported by refugees, we want to highlight the risk of false positive results causing unnecessary distress. Thus, when reporting results back to a patient or participant, it should be done with both cultural and psychometric sensitivity in mind.

Finally, completing the i-TAP requires reading proficiency and digital literacy, and would thus not be a suitable option for all help-seeking individuals with a refugee background; however, adding audio files could be considered to increase applicability. Against this, and the fact that the i-TAP missed 3 individuals with a disorder, we emphasize the need for flexibility and individual adaptation of assessment procedures to promote equity in care.

Limitations

Some methodological limitations need consideration. The selection of participants was neither randomized nor consecutive, resulting in a convenience sample, which raises questions about bias, representativeness, and generalization. Given the sampling method and self-selection, there is a risk that individuals with mental health problems were more likely to participate, potentially resulting in a higher prevalence compared with the broader population of individuals with a refugee background in Sweden. However, relative to this population, our sample included a higher proportion of asylum-seekers (no residence permit) and individuals residing in asylum facilities—established risk factors for mental ill health [10,11]—which likely contributed to the high prevalence found in this study. The literacy requirement excluded illiterate individuals, who thus are not represented in the sample. Nevertheless, regarding the distribution of sex, age, and countries of origin, the sample was found representative of the Swedish adult population of refugees and asylum-seekers at the time of data collection [29]. While the sample mainly comprised individuals from South West Asia, thereby limiting the generalizability of the findings to these groups, it is likely that the i-TAP could be applicable to other refugee populations currently residing in Sweden, such as Ukrainians, as well as in other contexts. However, this warrants future studies.

Another limitation regards language proficiency as we did not control for literacy. However, the results point to a general understanding of the questions, indicating sufficient language proficiency in the sample. Furthermore, about half of the structured diagnostic interviews were conducted with telephone interpreters, the most common interpretation modality within health care in Sweden [64]. Although clinicians, patients, and interpreters have been found to prefer in-person interpretation for building rapport and trust [65,66], a systematic review found no differences in patient satisfaction between modes of interpretation [67]. Validated translations of the M.I.N.I. would have increased reliability and validity, but were not available. This is the common reality in both research and practice with linguistically diverse groups, and interpreters have assisted M.I.N.I. assessments in previous studies [68,69]. Sharing the Swedish and Arabic translations with interpreters did facilitate the interview, but it was not always possible. Nevertheless, the use of interpreters should not, but always can, influence communication quality and ultimately the clinical assessment [9], and thus pose a limitation in this study.

We did not conduct an interrater reliability check as it was not deemed ethically justifiable to film or ask participants to partake twice, given the vulnerability of the population and study settings [70,71]. It would also pose a risk to recruitment. Measures taken to counter this limitation were: psychologists with previous experience of the M.I.N.I., training, authorized

XSL•FO
RenderX

interpreters, assessment conferences, and using structured and validated material. We argue that the range of psychologists and interpreters is a strength in this study, reducing the risk of individual bias in assessments.

Using the M.I.N.I. as the reference standard comes with some limitations. Previous studies show that standardized diagnostic interviews result in more diagnoses when compared to clinical psychiatric assessments [43], thus, the results on prevalence should be interpreted with some caution. Regarding the validation of the i-TAP, we however argue that the sensitivity of the M.I.N.I. [43,44] is suitable, as the objective was to develop and evaluate a sensitive screening tool with the aim of identifying clinically relevant symptoms.

Conclusion

Our results suggest that the online tiered assessment procedure, i-TAP, can identify clinical depression, anxiety, PTSD, and insomnia efficiently and with high accuracy among individuals with a refugee background residing in Sweden. Implementation of the i-TAP would enable simultaneous screening of the most common psychiatric disorders among refugees and asylum-seekers, reduce the item burden for symptom-free individuals and circumvent decisions on what to screen for. This, together with the digital, multilingual format of the i-TAP, underlines its potential as a feasible, accessible screening tool, with the potential to bridge several of the reported barriers to mental health care.

Acknowledgments

The authors are grateful to all participants for their time and engagement in the study. We thank Karin Christiansen and her team, as well as Lina Andersson for valuable collaborations and assistance with recruitment. We also acknowledge the contributions of the psychologists Helen Ahmadpourian, Olivia Erling, Tove Forsström, Hania Kutabi, Agnes Lorenz, Hajar Mohammadi, and Thilda Persson. For translations of the material, we thank Amjad AbuLeil, Farzaneh Sianpour, and Sohrab Soltani. Finally, we thank George Vlaescu for providing technical support.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to continued analyses and reports but are available from the corresponding author on reasonable request.

Funding

This study was funded by the Swedish Research Council (2018-05827). The funders did not partake in or influence the data collection, analysis, or reporting of the findings. Open access funding provided by Mid Sweden University.

Authors' Contributions

Conceptualization: AB, JM
Methodology: AB, JM, ER, AL
Investigation: AB, JM, ER, YD, AG
Formal analysis: JM
Project administration: JM
Resources: JM
Supervision: AB, ER
Funding acquisition: GA
Writing – original draft: AB, JM, ER
Writing – review & editing: AB, JM, ER, YD, AG, AL, GA

Conflicts of Interest

None declared.

Multimedia Appendix 1
Assessment material for insomnia disorder and detailed psychometric evaluation of the Insomnia Severity Index-7.
[DOCX File , 40 KB - humanfactors_v13i1e82763_app1.docx]

Multimedia Appendix 2
Extended psychometric evaluation of the internet-based tiered screening procedure.
[DOCX File , 38 KB - humanfactors_v13i1e82763_app2.docx]

References

1. Global trends report - forced displacement in 2024. United Nations High Commissioner for Refugees. 2025. URL: https://wwwunhcrorg/global-trends-report-2024 [accessed 2026-01-07]

XSL•FO
RenderX

2. Convention and protocol relating to the status of refugees. United Nations High Commissioner for Refugees. 2025. URL: https://www.unhcr.org/3b66c2aa102010 [accessed 2026-01-07]
3. Asylsökande under året efter medborgarskapsland och kön [Asylum-seekers in the year, by country of citizenship and sex]. SCB. 2025. URL: https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101P/AsylsokandeN/ : [accessed 2026-01-07]
4. Bogic M, Ajdukovic D, Bremner S, Franciskovic T, Galeazzi GM, Kucukalic A, et al. Factors associated with mental disorders in long-settled war refugees: refugees from the former Yugoslavia in Germany, Italy and the UK. Br J Psychiatry 2012;200(3):216-223 [FREE Full text] [doi: 10.1192/bjp.bp.110.084764] [Medline: 22282430]
5. Turrini G, Purgato M, Ballette F, Nosè M, Ostuzzi G, Barbui C. Common mental disorders in asylum seekers and refugees: umbrella review of prevalence and intervention studies. Int J Ment Health Syst 2017;11:51 [FREE Full text] [doi: 10.1186/s13033-017-0156-0] [Medline: 28855963]
6. Porter M, Haslam N. Predisplacement and postdisplacement factors associated with mental health of refugees and internally displaced persons: a meta-analysis. JAMA 2005;294(5):602-612. [doi: 10.1001/jama.294.5.602] [Medline: 16077055]
7. Malm A, Tinghög P, Narusyte J, Saboonchi F. The refugee post-migration stress scale (RPMS) - development and validation among refugees from Syria recently resettled in Sweden. Confl Health 2020;14:2 [FREE Full text] [doi: 10.1186/s13031-019-0246-5] [Medline: 31921332]
8. Fazel M, Wheeler J, Danesh J. Prevalence of serious mental disorder in 7000 refugees resettled in western countries: a systematic review. Lancet 2005;365(9467):1309-1314. [doi: 10.1016/S0140-6736(05)61027-6] [Medline: 15823380]
9. Blackmore R, Boyle JA, Fazel M, Ranasinha S, Gray KM, Fitzgerald G, et al. The prevalence of mental illness in refugees and asylum seekers: a systematic review and meta-analysis. PLoS Med 2020;17(9):e1003337 [FREE Full text] [doi: 10.1371/journal.pmed.1003337] [Medline: 32956381]
10. Posselt M, McIntyre H, Ngcanga M, Lines T, Procter N. The mental health status of asylum seekers in middle- to high-income countries: a synthesis of current global evidence. Br Med Bull 2020;134(1):4-20. [doi: 10.1093/bmb/ldaa010] [Medline: 32409820]
11. Leiler A, Bjärtå A, Ekdahl J, Wasteson E. Mental health and quality of life among asylum seekers and refugees living in refugee housing facilities in Sweden. Soc Psychiatry Psychiatr Epidemiol 2019;54(5):543-551. [doi: 10.1007/s00127-018-1651-6] [Medline: 30580381]
12. Borho A, Viazminsky A, Morawa E, Schmitt GM, Georgiadou E, Erim Y. The prevalence and risk factors for mental distress among Syrian refugees in Germany: a register-based follow-up study. BMC Psychiatry 2020;20(1):362. [doi: 10.1186/s12888-020-02746-2] [Medline: 32641014]
13. Al-Smadi AM, Tawalbeh LI, Gammoh OS, Ashour A, Tayfur M, Attarian H. The prevalence and the predictors of insomnia among refugees. J Health Psychol 2019;24(8):1125-1133. [doi: 10.1177/1359105316687631] [Medline: 28810381]
14. Basishvili T, Eliozishvili M, Maisuradze L, Lortkipanidze N, Nachkebia N, Oniani T, et al. Insomnia in a displaced population is related to war-associated remembered stress. Stress Health 2012;28(3):186-192. [doi: 10.1002/smi.1421] [Medline: 22282401]
15. Richter K, Baumgärtner L, Niklewski G, Peter L, Köck M, Kellner S, et al. Sleep disorders in migrants and refugees: a systematic review with implications for personalized medical approach. EPMA J 2020;11(2):251-260 [FREE Full text] [doi: 10.1007/s13167-020-00205-2] [Medline: 32549917]
16. Patanè M, Ghane S, Karyotaki E, Cuijpers P, Schoonmade L, Tarsitani L, et al. Prevalence of mental disorders in refugees and asylum seekers: a systematic review and meta-analysis. Glob Ment Health (Camb) 2022;9:250-263 [FREE Full text] [doi: 10.1017/gmh.2022.29] [Medline: 36618716]
17. Satinsky E, Fuhr DC, Woodward A, Sondorp E, Roberts B. Mental health care utilisation and access among refugees and asylum seekers in Europe: a systematic review. Health Policy 2019;123(9):851-863 [FREE Full text] [doi: 10.1016/j.healthpol.2019.02.007] [Medline: 30850148]
18. Due C, Green E, Ziersch A. Psychological trauma and access to primary healthcare for people from refugee and asylum-seeker backgrounds: a mixed methods systematic review. Int J Ment Health Syst 2020;14(1):71 [FREE Full text] [doi: 10.1186/s13033-020-00404-4] [Medline: 32944067]
19. Kiselev N, Morina N, Schick M, Watzke B, Schnyder U, Pfaltz MC. Barriers to access to outpatient mental health care for refugees and asylum seekers in Switzerland: the therapist's view. BMC Psychiatry 2020;20(1):378 [FREE Full text] [doi: 10.1186/s12888-020-02783-x] [Medline: 32680485]
20. Byrow Y, Pajak R, Specker P, Nickerson A. Perceptions of mental health and perceived barriers to mental health help-seeking amongst refugees: a systematic review. Clin Psychol Rev 2020;75:101812. [doi: 10.1016/j.cpr.2019.101812] [Medline: 31901882]
21. Maillet L, Goudet A, Godbout I, Ntanda GM, Laliberté G, Desjardins F, et al. Measuring the optimal time interval between arrival and first mental health evaluation's for refugees in Québec: a scoping review. Adm Policy Ment Health 2023;50(4):563-575 [FREE Full text] [doi: 10.1007/s10488-023-01257-y] [Medline: 36881163]
22. Whitehead L, Talevski J, Fatehi F, Beauchamp A. Barriers to and facilitators of digital health among culturally and linguistically diverse populations: qualitative systematic review. J Med Internet Res 2023;25:e42719 [FREE Full text] [doi: 10.2196/42719] [Medline: 36853742]

XSL•FO
RenderX

23. Liem A, Natari RB, Jimmy, Hall BJ. Digital health applications in mental health care for immigrants and refugees: a rapid review. Telemed J E Health 2021;27(1):3-16. [doi: 10.1089/tmj.2020.0012] [Medline: 32498658]
24. Donnelly O, Leavey G. Screening tools for mental disorders among female refugees: a systematic review. J Child Adolesc Trauma 2022;15(2):209-219 [FREE Full text] [doi: 10.1007/s40653-021-00375-9] [Medline: 35600514]
25. Morina N, Ewers SM, Passardi S, Schnyder U, Knaevelsrud C, Müller J, et al. Mental health assessments in refugees and asylum seekers: evaluation of a tablet-assisted screening software. Confl Health 2017;11:18 [FREE Full text] [doi: 10.1186/s13031-017-0120-2] [Medline: 29026438]
26. Vindbjerg E, Sonne C, Silove D, Bibby H, Hall J, Momartin S, et al. Evaluating the procedural validity of the multilingual computer assisted self interview (MultiCASI) in a refugee population. Psychiatry Res 2022;312:114529. [doi: 10.1016/j.psychres.2022.114529] [Medline: 35398659]
27. Batterham PJ, Calear AL, Sunderland M, Carragher N, Christensen H, Mackinnon AJ. Hierarchical screening for multiple mental disorders. J Affect Disord 2013;151(1):229-236. [doi: 10.1016/j.jad.2013.05.085] [Medline: 23806587]
28. Meurling J, Rondung E, Leiler A, Wasteson E, Andersson G, Richards D, et al. An online tiered screening procedure to identify mental health problems among refugees. BMC Psychiatry 2023;23(1):7 [FREE Full text] [doi: 10.1186/s12888-022-04481-2] [Medline: 36597066]
29. Statistik om asylärenden. Migrationsverket. 2022. URL: https://www.migrationsverket.se/Om-Migrationsverket/Statistik.html2022 [accessed 2026-01-07]
30. Bjärtå A, Leiler A, Ekdahl J, Wasteson E. Assessing severity of psychological distress among refugees with the Refugee Health Screener, 13-item version. J Nerv Ment Dis 2018;206(11):834-839 [FREE Full text] [doi: 10.1097/NMD.0000000000000886] [Medline: 30256332]
31. Hollifield M, Toolson E, Verbillis-Kolp S, Farmer B, Yamazaki J, Woldehaimanot T, et al. Effective screening for emotional distress in refugees: the Refugee Health Screener. J Nerv Ment Dis 2016;204(4):247-253. [doi: 10.1097/NMD.0000000000000469] [Medline: 26825376]
32. Hollifield M, Verbillis-Kolp S, Farmer B, Toolson EC, Woldehaimanot T, Yamazaki J, et al. The Refugee Health Screener-15 (RHS-15): development and validation of an instrument for anxiety, depression, and PTSD in refugees. Gen Hosp Psychiatry 2013;35(2):202-209. [doi: 10.1016/j.genhosppsych.2012.12.002] [Medline: 23347455]
33. Kroenke K, Spitzer R, Williams J. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
34. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
35. Sapra A, Bhandari P, Sharma S, Chanpura T, Lopp L. Using Generalized Anxiety Disorder-2 (GAD-2) and GAD-7 in a primary care setting. Cureus 2020;12(5):e8224 [FREE Full text] [doi: 10.7759/cureus.8224] [Medline: 32582485]
36. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
37. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007;146(5):317-325. [doi: 10.7326/0003-4819-146-5-200703060-00004] [Medline: 17339617]
38. Zuromski KL, Ustun B, Hwang I, Keane TM, Marx BP, Stein MB, et al. Developing an optimal short-form of the PTSD checklist for DSM-5 (PCL-5). Depress Anxiety 2019;36(9):790-800 [FREE Full text] [doi: 10.1002/da.22942] [Medline: 31356709]
39. Weathers FW, Keane TM, Palmieri PA, Marx BP, Schnurr PP. The PTSD Checklist for DSM-5 (PCL-5). 2013. URL: https://www.ptsd.va.gov/professional/assessment/adult-sr/
40. Bastien C, Vallières A, Morin C. Validation of the insomnia severity index as an outcome measure for insomnia research. Sleep Med 2001;2(4):297-307. [doi: 10.1016/s1389-9457(00)00065-4] [Medline: 11438246]
41. Morin C, Belleville G, Bélanger L, Ivers H. The insomnia severity index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 2011;34(5):601-608 [FREE Full text] [doi: 10.1093/sleep/34.5.601] [Medline: 21532953]
42. Morin CM, Colecchi C, Stone J, Sood R, Brink D. Behavioral and pharmacological therapies for late-life insomnia: a randomized controlled trial. JAMA 1999;281(11):991-999. [doi: 10.1001/jama.281.11.991] [Medline: 10086433]
43. Rettew DC, Lynch AD, Achenbach TM, Dumenci L, Ivanova MY. Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. Int J Methods Psychiatr Res 2009;18(3):169-184 [FREE Full text] [doi: 10.1002/mpr.289] [Medline: 19701924]
44. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33;quiz 34. [Medline: 9881538]
45. Lecrubier Y, Sheehan D, Weiller E, Amorim P, Bonora I, Sheehan KH, et al. The mini international neuropsychiatric interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. Eur Psychiatr 2020;12(5):224-231. [doi: 10.1016/s0924-9338(97)83296-8]

XSL•FO
RenderX

46. Karnouk C, Böge K, Lindheimer N, Churbaji D, Abdelmagid S, Mohamad S, et al. Development of a culturally sensitive Arabic version of the Mini International Neuropsychiatric Interview (M.I.N.I.-AR) and validation of the depression module. Int J Ment Health Syst 2021;15(1):24 [FREE Full text] [doi: 10.1186/s13033-021-00447-1] [Medline: 33736659]
47. Wiechers M, Strupf M, Bajbouj M, Böge K, Karnouk C, Goerigk S, et al. Empowerment group therapy for refugees with affective disorders: results of a multicenter randomized controlled trial. Eur Psychiatry 2023;66(1):e64 [FREE Full text] [doi: 10.1192/j.eurpsy.2023.2431] [Medline: 37458215]
48. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). Washington: American Psychiatric Association; 2013.
49. Sigvardsdotter E, Nilsson H, Malm A, Tinghög P, Gottvall M, Vaez M, et al. Development and preliminary validation of Refugee Trauma History Checklist (RTHC)-a brief checklist for survey studies. Int J Environ Res Public Health 2017;14(10):1175 [FREE Full text] [doi: 10.3390/ijerph14101175] [Medline: 28976937]
50. Verhey R, Chibanda D, Gibson L, Brakarsh J, Seedat S. Validation of the posttraumatic stress disorder checklist - 5 (PCL-5) in a primary care population with high HIV prevalence in Zimbabwe. BMC Psychiatry 2018;18(1):109 [FREE Full text] [doi: 10.1186/s12888-018-1688-9] [Medline: 29685117]
51. Bjureberg J, Dahlin M, Carlborg A, Edberg H, Haglund A, Runeson B. Columbia-suicide severity rating scale screen version: initial screening for suicide risk in a psychiatric emergency department. Psychol Med 2021;52(16):1-9 [FREE Full text] [doi: 10.1017/S0033291721000751] [Medline: 33766155]
52. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. The Columbia-suicide severity rating scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry 2011;168(12):1266-1277 [FREE Full text] [doi: 10.1176/appi.ajp.2011.10111704] [Medline: 22193671]
53. Rondung E, Leiler A, Sarkadi A, Bjärtå A, Lampa E, Löfving SG, et al. Feasibility of a randomised trial of teaching recovery techniques (TRT) with refugee youth: results from a pilot of the Swedish UnaccomPanied yOuth Refugee Trial (SUPpORT). Pilot Feasibility Stud 2022;8(1):40 [FREE Full text] [doi: 10.1186/s40814-022-00998-1] [Medline: 35164865]
54. Bujang MA, Adnan TH. Requirements for minimum sample size for sensitivity and specificity analysis. J Clin Diagn Res 2016;10(10):YE01-YE06 [FREE Full text] [doi: 10.7860/JCDR/2016/18129.8744] [Medline: 27891446]
55. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960 Apr 01;20(1):37-46. [doi: 10.1177/001316446002000104]
56. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(1):159-174. [doi: 10.2307/2529310]
57. Mandrekar JN. Simple statistical measures for diagnostic accuracy assessment. J Thorac Oncol 2010;5(6):763-764 [FREE Full text] [doi: 10.1097/JTO.0b013e3181dab122] [Medline: 20502268]
58. Plummer F, Manea L, Trepel D, McMillan D. Screening for anxiety disorders with the GAD-7 and GAD-2: a systematic review and diagnostic metaanalysis. Gen Hosp Psychiatry 2016;39:24-31. [doi: 10.1016/j.genhosppsych.2015.11.005] [Medline: 26719105]
59. Urtasun M, Daray FM, Teti GL, Coppolillo F, Herlax G, Saba G, et al. Validation and calibration of the Patient Health Questionnaire (PHQ-9) in Argentina. BMC Psychiatry 2019;19(1):291 [FREE Full text] [doi: 10.1186/s12888-019-2262-9] [Medline: 31533674]
60. Turner SW, Bowie C, Dunn G, Shapo L, Yule W. Mental health of Kosovan Albanian refugees in the UK. Br J Psychiatry 2018;182(5):444-448. [doi: 10.1192/bjp.182.5.444]
61. Engelhard IM, van den Hout MA, Weerts J, Arntz A, Hox JJCM, McNally RJ. Deployment-related stress and trauma in Dutch soldiers returning from Iraq. Prospective study. Br J Psychiatry 2007;191:140-145. [doi: 10.1192/bjp.bp.106.034884] [Medline: 17666498]
62. Thombs BD, Kwakkenbos L, Levis AW, Benedetti A. Addressing overestimation of the prevalence of depression based on self-report screening questionnaires. CMAJ 2018;190(2):E44-E49 [FREE Full text] [doi: 10.1503/cmaj.170691] [Medline: 29335262]
63. Lindegaard T, Seaton F, Halaj A, Berg M, Kashoush F, Barchini R, et al. Internet-based cognitive behavioural therapy for depression and anxiety among Arabic-speaking individuals in Sweden: a pilot randomized controlled trial. Cogn Behav Ther 2021;50(1):47-66 [FREE Full text] [doi: 10.1080/16506073.2020.1771414] [Medline: 32603632]
64. Tolkutredningen 2019. Almega-Språkföretagen. 2019. URL: https://www.regeringen.se/contentassets/626ccb432337479d9c5968efebb5186a/almega-sprakforetagen.pdf [accessed 2026-01-07]
65. Ruiz A, Chen J, Brown TT, Cai X, Hernandez Fernandez P, Rodriguez HP. Comparing remote and in-person interpretation experiences for clinicians and Spanish-speaking patients with limited English proficiency: a mixed methods study. BMJ Open Qual 2025;14(2):e003227 [FREE Full text] [doi: 10.1136/bmjoq-2024-003227] [Medline: 40340979]
66. Price EL, Pérez-Stable EJ, Nickleach D, López M, Karliner LS. Interpreter perspectives of in-person, telephonic, and videoconferencing medical interpretation in clinical encounters. Patient Educ Couns 2012;87(2):226-232 [FREE Full text] [doi: 10.1016/j.pec.2011.08.006] [Medline: 21930360]
67. Joseph C, Garruba M, Melder A. Patient satisfaction of telephone or video interpreter services compared with in-person services: a systematic review. Aust Health Rev 2018;42(2):168-177. [doi: 10.1071/AH16195] [Medline: 30021688]

XSL•FO
RenderX

68. Heeren M, Mueller J, Ehlert U, Schnyder U, Copiery N, Maier T. Mental health of asylum seekers: a cross-sectional study of psychiatric disorders. BMC Psychiatry 2012;12:114 [FREE Full text] [doi: 10.1186/1471-244X-12-114] [Medline: 22900706]
69. Hocking DC, Mancuso SG, Sundram S. Development and validation of a mental health screening tool for asylum-seekers and refugees: the STAR-MH. BMC Psychiatry 2018;18(1):69 [FREE Full text] [doi: 10.1186/s12888-018-1660-8] [Medline: 29548315]
70. Deps P, Rezende I, Andrade M, Collin S. Ethical issues in research with refugees. Ethics Med Public Health 2022;24:100813. [doi: 10.1016/j.jemep.2022.100813]
71. International ethical guidelines for health-related research involving humans. Council for International Organizations of Medical Sciences WHO. 2016. URL: https://cioms.ch/publications/product/international-ethical-guidelines-for-health-related-research-involving-humans/ [accessed 2026-01-07]

Abbreviations

C-SSRS Screen: Columbia Suicide Severity Rating Scale Screening Version
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
GAD: generalized anxiety disorder
GAD-2: Generalized Anxiety Disorder-2
GAD-7: Generalized Anxiety Disorder -7
ICD-10: International Statistical Classification of Diseases, Tenth Revision
ISI-7: Insomnia Severity Index
i-TAP: internet-based tiered assessment procedure
MDD: major depressive disorder
M.I.N.I.: Mini International Neuropsychiatric Interview
PCL-5: PTSD (posttraumatic stress disorder) Checklist for DSM-5 (Diagnostic and Statistical Manual of Mental Disorders [Fifth Edition])
PCL-5 SF: PTSD (posttraumatic stress disorder) Checklist for DSM-5 (Diagnostic and Statistical Manual of Mental Disorders [Fifth Edition]) Short Form
PHQ-2: Patient Health Questionnaire-2
PHQ-9: Patient Health Questionnaire-9
PTE: potentially traumatic event
PTSD: posttraumatic stress disorder
RHS-13: Refugee Health Screener -13
RTHC: Refugee Trauma History Checklist
UNHCR: United Nations High Commissioner for Refugees

Edited by A Kushniruk; submitted 21.Aug.2025; peer-reviewed by DC Hocking, N Kiselev; comments to author 01.Dec.2025; accepted 02.Jan.2026; published 29.Jan.2026.

Please cite as:
Meurling J, Rondung E, Demetry Y, Geranmayeh A, Leiler A, Andersson G, Bjärtå A
Online Tiered Screening for Mental Health Problems Among Refugees in Sweden: Validation Study
JMIR Hum Factors 2026;13:e82763
URL: https://humanfactors.jmir.org/2026/1/e82763
doi:10.2196/82763
PMID:

©Jennifer Meurling, Elisabet Rondung, Youstina Demetry, Anahita Geranmayeh, Anna Leiler, Gerhard Andersson, Anna Bjärtå. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 29.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Investigating the Impact of Lived Experience Stories on Self-Harm, Mood, and Help-Seeking Intentions: Web-Based Between-Participants Experimental Study in Individuals With Recent Self-Harm

Jennifer Ferrar[1], PhD; Lizzy Winstone[2], PhD; Ian Penton-Voak[1,3], PhD; Lucy Biddle[2,3], PhD; Paul Moran[2,3], MD; Lydia Grace[4], PhD; Becky Mars[2,3], PhD

[1]School of Psychological Science, University of Bristol, Bristol, United Kingdom

[2]Centre for Academic Mental Health, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

[3]NIHR Biomedical Research Centre, University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, England, United Kingdom

[4]Samaritans, Surrey, United Kingdom

Corresponding Author:
Jennifer Ferrar, PhD
School of Psychological Science
University of Bristol
12a Priory Road
Bristol, BS8 1TU
United Kingdom
Phone: 44 1174558807
Email: jennifer.ferrar@bristol.ac.uk

Abstract

Background: Lived experience stories are often used on formal help sites as a support resource for individuals who self-harm. While self-harm–related internet use provides an alternative for individuals who are not yet ready or are unwilling or unable to access support offline, it has also been shown to unintentionally reinforce self-harm behavior. There are several components that might influence whether a lived experience story is perceived as helpful, unhelpful, or potentially harmful, and the evidence supporting that these encourage help-seeking in the reader is limited.

Objective: This study is part of a mixed methods project that aimed to investigate how variations in help-seeking messages contained within online lived experience stories are interpreted by and psychologically impact those with a history of self-harm.

Methods: Individuals with a recent history of self-harm were recruited via newsletters, social media, and websites run by the university and mental health charities to take part in an online experiment. During the experiment, participants were randomized to read stories that either mentioned (1) self-help strategies, (2) seeking help from informal and formal sources, or (3) did not mention help-seeking. Help-seeking intentions, mood, entrapment, and expectations of future self-harm were measured, and participants provided feedback on the stories.

Results: There was limited evidence for an effect of story type on help-seeking intentions ($F_{2, 230}$=4.2; P=.02; η^2=0.25), and clearer evidence for an effect of story type on negative affect ($F_{2, 230}$=4.02; P=.02; η^2=0.10; adjustment for age, gender, and help-seeking history included). Participants in the “self-help” condition (n=83) reported lower negative affect after reading the stories compared to participants in the “no help” condition (n=80; mean difference=–3.97, 95% CI –7.72 to –0.22; P=.04) and the “informal/formal” help condition (n=75; mean difference=–3.70, 95% CI –7.55 to 0.14; P=.06). A key criticism of the stories was that they were unrelatable, but this sentiment was less prevalent among those in the “no help” condition. Key positives were that the stories included a realistic but hopeful outlook of recovery (less prevalent in the “informal/formal help” condition) and were supportive (less prevalent in the “no help” condition).

Conclusions: While the inclusion of self-help strategies in a lived experience story reduced its impact on negative affect, the inclusion of self-help strategies or seeking help from others did not encourage help-seeking intentions. Making stories relatable, authentic, and providing multiple strategies for support might be key to encouraging help-seeking, but more research is needed.

XSL•FO
RenderX

(JMIR Hum Factors 2026;13:e71280) doi:10.2196/71280

KEYWORDS

help-seeking behavior; internet-based intervention; mental health recovery; personal narratives; self-injurious behavior

Introduction

Individuals experiencing self-harm or suicidal thoughts often seek support via the internet. Self-harm–related internet use can be both helpful and harmful [1,2]. Self-harm–related internet content can foster a sense of community, thereby reducing feelings of loneliness and isolation [3-6], and provide practical advice and moral support on how to manage self-harm [4,7-10]. In addition, self-harm–related internet use provides an alternative for individuals who are not yet ready, or may be unwilling, or unable to access support offline [11-14].

However, there is also a link between self-harm–related internet use and higher levels of suicidal intent [12,15,16], although the direction of effects is not clear. Due to a lack of regulation, content encouraging self-harm is easily accessible online [2,17,18]. In addition, content or environments intended to discourage self-harm could unintentionally reinforce self-harm behavior, for example, via normalization or competition [18,19]. Further, the categorization of self-harm content as either harmful or helpful is not straightforward. A recent systematic review illustrates the complex nature of engagement with self-harm content. Whether the content negatively or positively affects the individual engaging with it depends on the content, the individual, and the time and space in which it is encountered [2].

The internet provides a wide range of self-harm content, including information about self-harm behavior, self-help strategies, lived experience stories, images of self-harm, signposting to help, and dialogue [2,13,18,20,21]. They all come with limitations; for example, research suggests that information-giving can often be impersonal, disengaging, and "insufficient in moments of crisis" and that signposting can often be redundant, as many searching for help online are doing so because of barriers to engaging with signposted services [13]. The impact of self-harm images, such as scars, can be mixed, reminding some of recovery while triggering others [2]. Among those who access self-harm–related internet content, there seems to be a preference for real-time help (eg, "live chat"), self-help tools, and interaction with others, particularly via lived experience content [13,22].

Lived experience stories are often used on formal help sites. They are viewed as personal and engaging and can validate the crisis for readers, while also offering reassurance that they are not alone in their struggle and that recovery is possible [13,22,23]. However, there are several components that might influence whether a lived experience story is perceived as helpful, unhelpful, or potentially harmful. Stories appearing on forums, even those that are run by trained moderators, might be triggering for some people, for example, by including details of methods, injury, or graphic imagery. Stories appearing on websites run by charities and health care organizations usually include a personal narrative of help-seeking and recovery. However, the evidence supporting that these encourage help-seeking in the reader is limited [22,23]. Further, the careful regulation of content might be perceived as overly censored and lead some to search for unrefined content in covert and riskier spaces [2]. Even unregulated content posted by a lay individual is curated in a different respect, for example, on social media platforms where users are often presenting themselves in an idealized way [24,25].

This study is part of a mixed methods project that aimed to investigate how variations in help-seeking messages contained within online lived experience stories are interpreted by and psychologically impact those with a history of self-harm. This study used an experimental design to investigate the psychological impact of stories with different types of help-seeking messages, including those mentioning self-help strategies, informal/formal help-seeking (ie, friends, family, or professionals), and no help-seeking. We sought to assess if the type of help-seeking referenced in a lived experience story had an impact on help-seeking intentions, mood, expectations of future self-harm behaviors, as well as psychological factors implicated within the suicidal pathway, such as sense of entrapment (ie, lack of escape from unbearable thoughts, feelings, or situations) [26,27]. It also incorporated an exploratory, qualitative component in which participants provided open-ended feedback on the stories to assess their acceptability as potential interventions. The other study in this mixed methods project conducted focus groups with a subset of the participant sample used in this study (after they took part in the online experiment) to further explore this topic. The results from the focus group, published before this paper was written, demonstrated that although it is a complex relationship, individuals with recent self-harm experience could positively engage with lived experiences stories, for example, if they reflected a wide range of personal journeys and avoided the use of stigmatizing language [23].

Methods

Overview

This paper presents the quantitative and qualitative results of a self-administered online experiment [23] which explored how people with self-harm experience and interpret lived experience accounts as a means of support. The protocol [28] was preregistered ahead of data collection [29]. For results of the follow-up focus groups, see Winstone et al [23] for further discussion.

Ethical Considerations

Ethical approval was granted by the University of Bristol School of Psychological Science Research Ethics Committee (reference number 10504). Before beginning the online experiment, participants were presented with the participant information sheet, which included researcher contact details. Participants were invited to contact the researcher with any questions

XSL•FO
RenderX

regarding the study, before providing informed consent. Participants were able to navigate away from the survey while awaiting researcher response and resume participation once ready. Participants were then asked to confirm (via tick boxes) that they met the eligibility criteria and that they were providing informed consent to take part in the online study. Participants who completed the online survey were offered the option to enter a prize draw to win one of four £50 (US $67.50) Love2shop vouchers. All aspects of the General Data Protection Regulation and the Data Protection Act 2018 were adhered to. Anonymous study data have been backed up on a secured University of Bristol network drive and transferred to a designated University of Bristol Research Data Storage Facility for long-term archiving. The study data will be kept for a minimum of 15 years. The data sheet has been made open using the University of Bristol Research Data Repository.

Recruitment

The study was advertised by the research team and mental health charities on their web pages, social media accounts, and in newsletters sent out to individuals who signed up to hear about regular research/charity activity. The study was also advertised on self-harm/mental health-specific forums, on Reddit, RecoverYourLife, and the National Self-Harm Network website. The advertisements contained a link to the experiment, which was programmed and hosted online using Gorilla Experiment Builder [30]. Participants needed access to a device and an internet connection, but could access the task from any device that allowed them to access any web browser. After confirming eligibility and providing consent, and completing a self-administered baseline questionnaire, participants were randomly assigned (using Gorilla's balanced randomization feature) to one of the following three story conditions: (1) self-help strategies, (2) informal/formal help-seeking, (3) no help-seeking.

We conducted an a priori power analysis using G*Power (version 3.1.9.7; Heinrich-Heine-Universität Düsseldorf) [31], which determined that a sample of 246 participants would provide adequate (80%) power for our main analysis (1-way ANOVA) to detect a small to medium effect size (f=0.2) for our primary outcome (differences in help-seeking intentions) among the 3 story groups.

The study was conducted according to the revised Declaration of Helsinki (2013) and the 1996 International Council for Harmonisation Guidelines for Good Clinical Practice E6(R2). Further details about anticipated risks and our strategy to mitigate any risks can be found under "Ethical Considerations and Informed Consent" and "Safety" in the study protocol [28]. To be eligible for the study, participants had to be aged 16 years or older, fluent in English, currently resident in the United Kingdom, have engaged in self-harm within the past year, and be able to provide informed consent. Before beginning the study, participants had to provide informed written consent. Safeguarding resources were signposted to participants at the start and at the end of the study. Participants were asked to download a copy (or make note of) the safeguarding resources before beginning the study to ensure that all participants had access to the resources regardless of how far they progressed in the study. Participants were informed that it would take approximately 30 minutes to complete the survey. It was mandatory to answer all questions during the experiment, but participants were reminded that they could stop taking part at any time. Participants were also informed that the study was anonymous and that all data from participants who completed the study would be made public at the end of the study. After taking part, participants were debriefed (and again presented with safeguarding resources) and were offered the opportunity to enter a prize draw to win one of four £50 (US $67.50) Love2shop vouchers.

Materials

A total of 3 template stories were used in this study (Multimedia Appendix 1). The same 3 templates were presented to all participants, but a portion of text was modified (removed and/or inserted) to reflect the different experimental conditions: self-help strategies, informal/formal help-seeking, and no help-seeking. The template stories were chosen from an initial shortlist of 10 real-life stories curated by BM. These were selected from a variety of online sources, including the websites of third-sector organizations, health care organizations, web-based forums, and personal blogs. This approach was used to ensure that they were not solely representative of the more polished stories that often appear on formal help sites. For ethical reasons, stories were excluded if they encouraged self-harm, described self-harm methods in detail, or did not have a recovery-oriented/hopeful outlook. Each of the 10 stories was independently ranked (in order of preference) and discussed by BM and 3 of the study authors (JF, IP-V, and LB). The 3 lived experience stories were then chosen following the ranking and discussion to be used as the template stories. The only key difference across the experimental conditions was whether text concerning self-help strategies, informal/formal help-seeking, or no help-seeking (but a positive/hopeful message) was inserted into the stories. These inserts were drawn from the original 10 stories selected by BM.

Measures

Participant Characteristics (Assessed at Baseline)

Participants reported their age in years, their gender, sexual orientation, and ethnicity. They also answered questions on self-harm thoughts (past year frequency), passive and active suicidal thoughts (in lifetime and recency), nonsuicidal self-harm (in lifetime, recency, and past year frequency), and suicide attempts (in lifetime and recency). To measure help-seeking history, participants completed a modified version of the Actual Help-Seeking Questionnaire (AHSQ) [32,33]. For further details, please refer to the protocol [28].

Outcome Measures (Assessed Post Randomization)

Primary Outcome: Help-Seeking Intentions

To measure help-seeking intentions, participants completed a modified version of the General Help Seeking Questionnaire (GHSQ) [34]. A total GHSQ score was computed by summing responses across all help-source items, excluding the "I would not seek help from anyone" item, in line with prior research [35,36]

XSL•FO
RenderX

Secondary Outcomes: Affect, Entrapment, and Future Self-Harm

Participants completed the Positive and Negative Affect Schedule (PANAS) [37]. They completed the 4-item Entrapment Scale Short Form (E-SF) [38]. They also answered 2 questions about future self-harm, rated on a 5-point scale. First, how likely they would be to hurt themselves on purpose in the future after reading the stories, and second, how likely it was that they could see a time in the future without self-harm.

Open-Ended Feedback on the Lived Experience Stories

Participants were asked to provide open-ended feedback on the lived experience stories presented to them during the study. Specifically, they were asked, "Was there anything you found helpful/unhelpful about the stories? If so, please explain."

Attention Check

To ensure that only data from participants who were paying attention during the study were included in the analysis, one of the items among the outcome measures asked participants to select 1 on a scale from 0 to 4. Participants who selected a number other than 1 had their data removed prior to the analysis [39].

Procedure

Participants completed the questionnaires on demographics, current/past self-harm thoughts and behaviors, and help-seeking history. They were then randomized, using Gorilla's balanced randomization feature, to 1 of the 3 conditions (ie, self-help, informal/formal help-seeking, or no help-seeking). Participants were presented with and asked to read the 3 lived experience stories, with amended text based on the condition that they were assigned to. Participants were not aware of the condition they had been assigned to, nor were they informed that there were different variations of the stories. The stories were presented in the same fixed order for all participants. They then completed the outcome measures in the following order: (1) free text opinions of the lived experience stories, (2) mood, (3) entrapment, (4) thoughts of self-harm/suicide, and (5) help-seeking intentions.

Analysis Plan

Data Screening

To maintain data quality, each participant's data were screened against the following criteria prior to analysis: (1) each participant must have completed the entire study and (2) must have correctly answered the eligibility questions and the attention check question within the experiment. The data were assessed for normality using skewness and kurtosis statistics, and were assessed to ensure that they met the assumptions for ANOVA testing.

Data Analysis

Descriptive statistics were used to summarize the participant characteristics and outcome measures in each group. An ANOVA was used to test if the type of help-seeking mentioned in the lived experience story influenced the primary outcome of interest (ie, total score on the GHSQ). A series of ANOVAs was used to test if the type of help-seeking mentioned in the lived experience story influenced the secondary outcome measures (ie, total scores on PANAS, E-SF, and the 2 items assessing future self-harm). Where the type of help-seeking mentioned in the stories impacted any of the outcomes, post hoc pairwise comparisons were conducted using Tukey's Honestly Significant Difference (HSD) test to control for multiple comparisons.

This study was powered to detect a main effect of the type of lived experience story; therefore, any interactions were exploratory. Where any of the analyses indicated a main effect of type of lived experience story on the outcome measures, then "story type × self-harm history" analyses of covariance (ANCOVAs) were used to assess if the influence of type of lived experience story on the outcome measures differed by self-harm history. The self-harm history variable was defined as the recency of self-harm behavior in the past year. In cases where a participant had a history of both nonsuicidal self-harm and attempting suicide, the recency score (indicating "either in the past week" or "more than a week ago, but in the past year") for whichever behavior was more recent was used. Where any interactions existed, they were explored in post hoc analyses using 2-tailed *t* tests. As prior research demonstrates that gender [40,41], age [40], and prior help-seeking experience [42] impact future help-seeking, all analyses were run both unadjusted and adjusted for age, gender, and a continuous variable of help-seeking history (ie, total score on the AHSQ).

In the protocol, we had specified that exploratory analyses would be guided by the qualitative results. Findings from focus groups suggested that lived experience stories can have a positive impact on readers, but that this effect may be moderated by age; for example, older participants struggled emotionally when reading about someone else's recovery that occurred by a certain age [23]. Therefore, we have included a tertiary analysis, exploring interactions between story type and age on any primary and secondary outcomes where main effects were found.

Participant opinions on the lived experience stories were analyzed using content analysis. After reviewing the data, specifically noting that many responses were brief and lacked richness, it was decided that instead of conducting thematic analysis (as originally specified in the protocol), it would be more appropriate to conduct content analysis. Established guidelines for conducting a conventional concept analysis were followed [43]. No preconceived categories were imposed; instead, key concepts were allowed to emerge inductively from the data. The data were imported into NVivo (version 14; QSR International) [44] for coding. Initial open coding was conducted independently by JF and JS (see Acknowledgments section), who identified salient thoughts and concepts within participants' responses. JF then refined the coding scheme by examining conceptual relationships and consolidating codes where responses reflected similar underlying ideas. LW subsequently coded 10% of the responses, adding additional codes when deemed appropriate. Intercoder reliability was then assessed and calculated to be around 99%. JF reviewed any amendments, and any disagreements were discussed and resolved. Finally, JF refined and renamed codes to ensure that each label accurately captured the underlying concept.

XSL•FO
RenderX

Results

The data that form the basis of the results presented here are available from the University of Bristol's Research Data Repository [45].

Participant Characteristics (Assessed at Baseline)

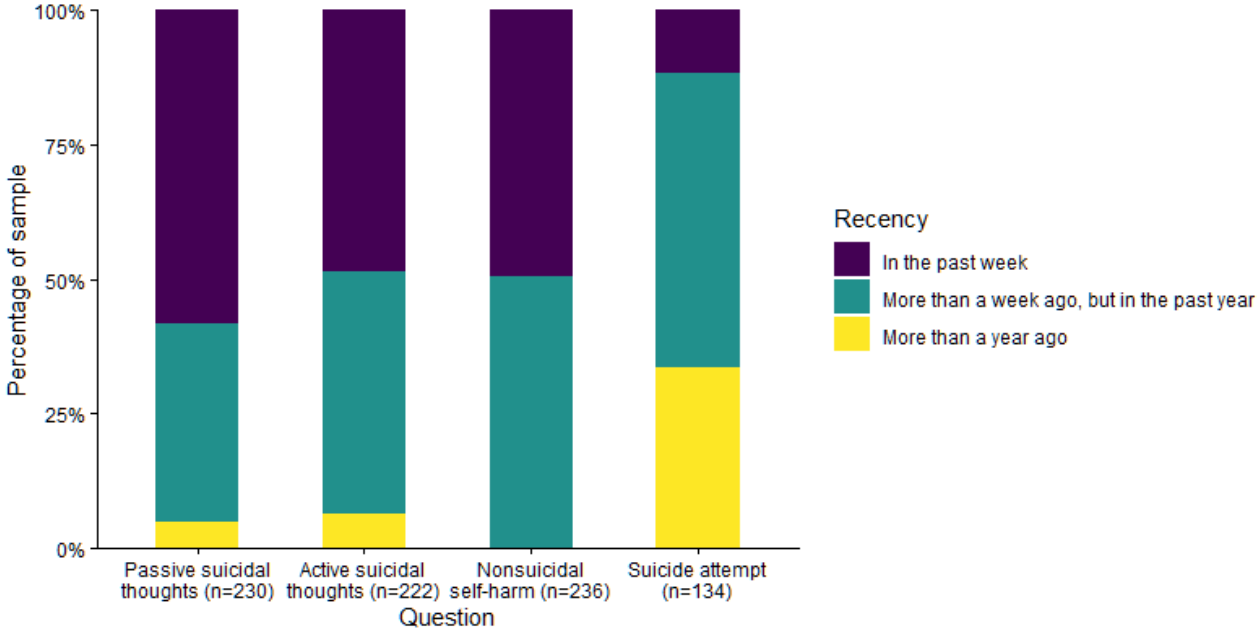
Demographics

Data collection ceased when there were complete datasets from 246 participants. One participant was identified as completing the study twice (based on IP address) and was removed from the dataset. Of the remaining 244 participants, data cleaning procedures identified 2 participants who failed the attention check and 6 who did not meet the eligibility criterion of having hurt themselves on purpose in the past year. Therefore, data from 238 (97.5% of our target sample size) participants were included in the analysis. Participants were primarily female (188/238, 78.9%), White (207/238, 86.9%), and were aged between 16 and 59 (mean 23.2, SD 9.4) years. Table S1 in Multimedia Appendix 2 provides further details on participant demographics across the experimental conditions, as well as a breakdown by experimental condition. The experiment ran for 28 weeks (from late March to early October 2022). In total, 381 individuals visited the link for the experiment, 379 provided informed consent, 361 passed the screening, and 340 were randomized.

Self-Harm Thoughts and Behaviors

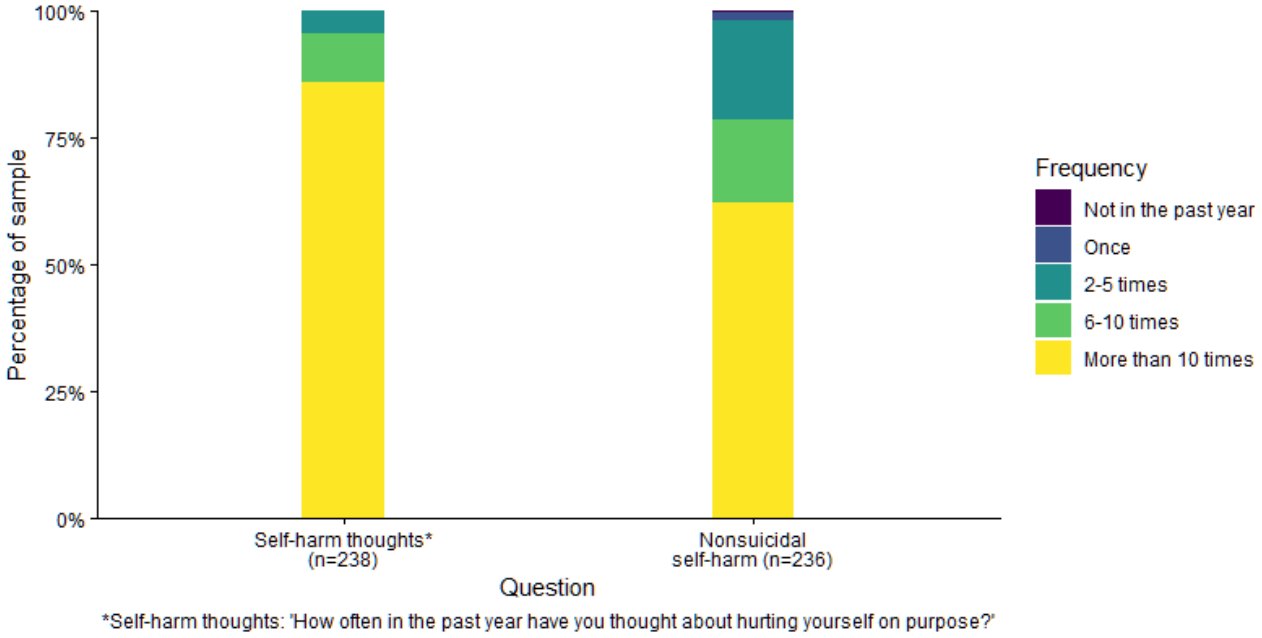
Out of 238 participants, 236 (99.2%) had engaged in nonsuicidal self-harm, 230 (96.6%) had thoughts that life was not worth living (ie, passive suicidal thoughts), and 222 (93.3%) had experienced suicidal ideations (ie, active suicidal thoughts) during their lifetime. Just over half the sample (n=134, 56.3%) had made a suicide attempt. Figures 1 and 2 display information about the recency and past-year frequency of self-harm thoughts and behaviors in the sample.

Figure 1. Recency of suicidal thoughts and self-harm behavior.



XSL•FO
RenderX

Figure 2. Frequency of self-harm thoughts and behavior (in the past year).



In terms of the type of self-harm methods used, 105 (44.1%) out of 238 participants reported swallowing pills or something poisonous, 109 (45.8%) reported burning themselves, and 226 (94.9%) reported cutting themselves. Using another method not listed was reported by 109 (45.8%) participants. These were screened by JF and BM and then categorized (for full details, please refer to Table S2 in Multimedia Appendix 2). Behaviors not considered to be self-harm (eg, tattooing, eating behaviors, or indirect self-harm such as unsafe sex) were removed; however, all participants reporting these behaviors also reported at least one other self-harm method, and so no participants were removed from the analysis. In terms of the number of self-harm methods used, 100 (42%) out of 238 participants reported having used 2 self-harm methods during their lifetime, with 45 (18.9%), 62 (26.1%), and 31 (13%) reporting 1, 3, and 4 or more methods, respectively.

Help-Seeking History

When reflecting on help-seeking over their lifetime, not seeking help from anyone was reported by 39 (16.4%) out of 238 participants. While some of these participants responded affirmatively to other items on the scale, 30 (12.6%) of those participants did not report any form of help-seeking. For a breakdown of the prevalence of accessing each source of informal and formal help, please refer to Table S3 in Multimedia Appendix 2.

Outcome Measures (Assessed Post Randomization)

Overview

Descriptive statistics are reported for the 5 outcome measures by experimental condition in Table 1.

Table 1. Descriptive statistics for outcomes.

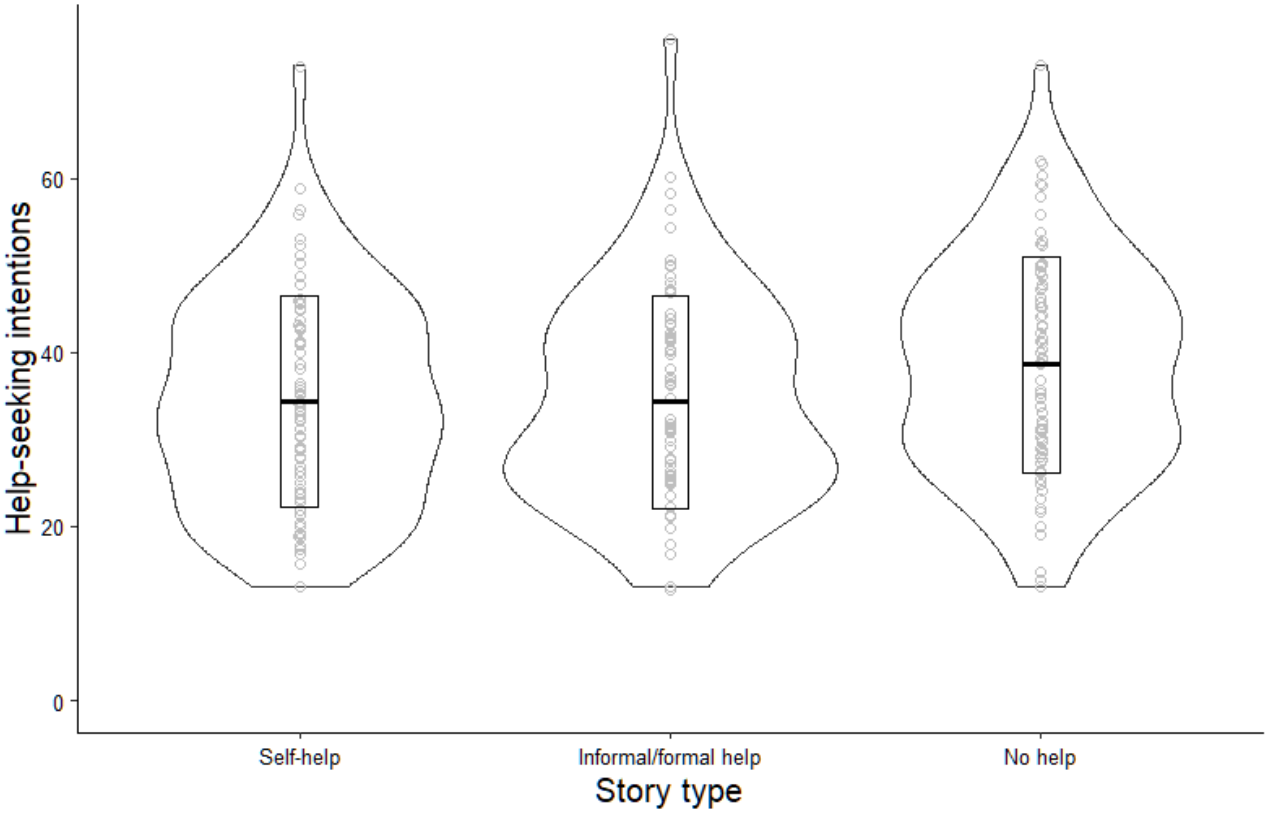
	Self-help	Informal/formal help	No help
Primary outcomes, mean (SD; range)			
Help-seeking intentions (higher scores indicate greater intention to seek help),	34.3 (12.2; 13-73)	34.3 (12.3; 13-76)	38.6 (12.4; 13-73)
Secondary outcomes, mean (SD; range)			
Positive affect	18.4 (6.0; 10-36)	18.0 (6.5; 10-43)	18.8 (6.0; 10-35)
Negative affect	24.8 (10.4; 10-48)	28.5 (10.6; 11-49)	28.7 (9.5; 10-49)
Entrapment	15.1 (4.9; 4-20)	15.0 (4.4; 4-20)	15.6 (4.1; 5-20)
Future self-harm, median (IQR)			
"After reading these stories, how likely is it that you will hurt yourself on purpose in the future?" (0=much less likely; 4=much more likely)	2 (2-2)	2 (2-2)	2 (2-3)
"Can you see a time in the future without self-harm?" (0=not at all likely; 4=very likely)	2 (1-3)	1 (0-3)	1 (1-3)

XSL•FO
RenderX

Help-Seeking Intentions

Due to the skewed nature of the data, the variable for help-seeking intentions was transformed using a square-root transformation. There was moderate evidence for an effect of story type on help-seeking intentions ($F_{2,\ 235}$=3.3; P=.04; η^2=0.03). Adjusting for age, gender, and help-seeking history provided stronger evidence for an effect of story type on help-seeking intentions ($F_{2,\ 230}$=4.2; P=.02; η^2=0.25; the full results of the model using transformed and untransformed data can be found in Tables S1 and S2 in Multimedia Appendix 3). None of the pairwise comparisons (Tukey's HSD on the unadjusted means) reached the conventional significance threshold. However, 2 comparisons approached significance, suggesting a trend toward group differences that were consistent with the overall effect. Participants in the "no help" condition reported marginally higher levels of help-seeking intentions after reading the stories compared to participants in the "self-help" condition (mean difference=–0.36, 95% CI –0.75 to –0.03; P=.08) and the "informal/formal help" condition (mean difference=–.36, 95% CI –0.74 to –0.02; P=.07). The impact of story type on help-seeking intentions is illustrated in Figure 3 (using untransformed data to aid interpretability).

Figure 3. Impact of story type on help-seeking intentions.



An ANCOVA was conducted to determine whether self-harm history (recency of self-harm) interacted with story type to influence help-seeking intentions. Evidence for this interaction was not found ($F_{2,\ 232}$=0.28; P=.76). An ANCOVA was conducted to explore if age interacted with story type to influence help-seeking intentions. Evidence for this interaction was also not found ($F_{2,\ 232}$=1.25; P=.29). However, the former only achieved 15% power, and the latter 41% power.

Affect, Entrapment, Future Self-Harm

Due to the skewed nature of the data, attempts to normalize the distribution of all secondary outcomes, using square-root and cube transformations, were made. Despite these efforts, visual inspection of residuals and formal tests of normality indicated that the residuals remained nonnormal. Therefore, the following analyses were conducted using untransformed data.

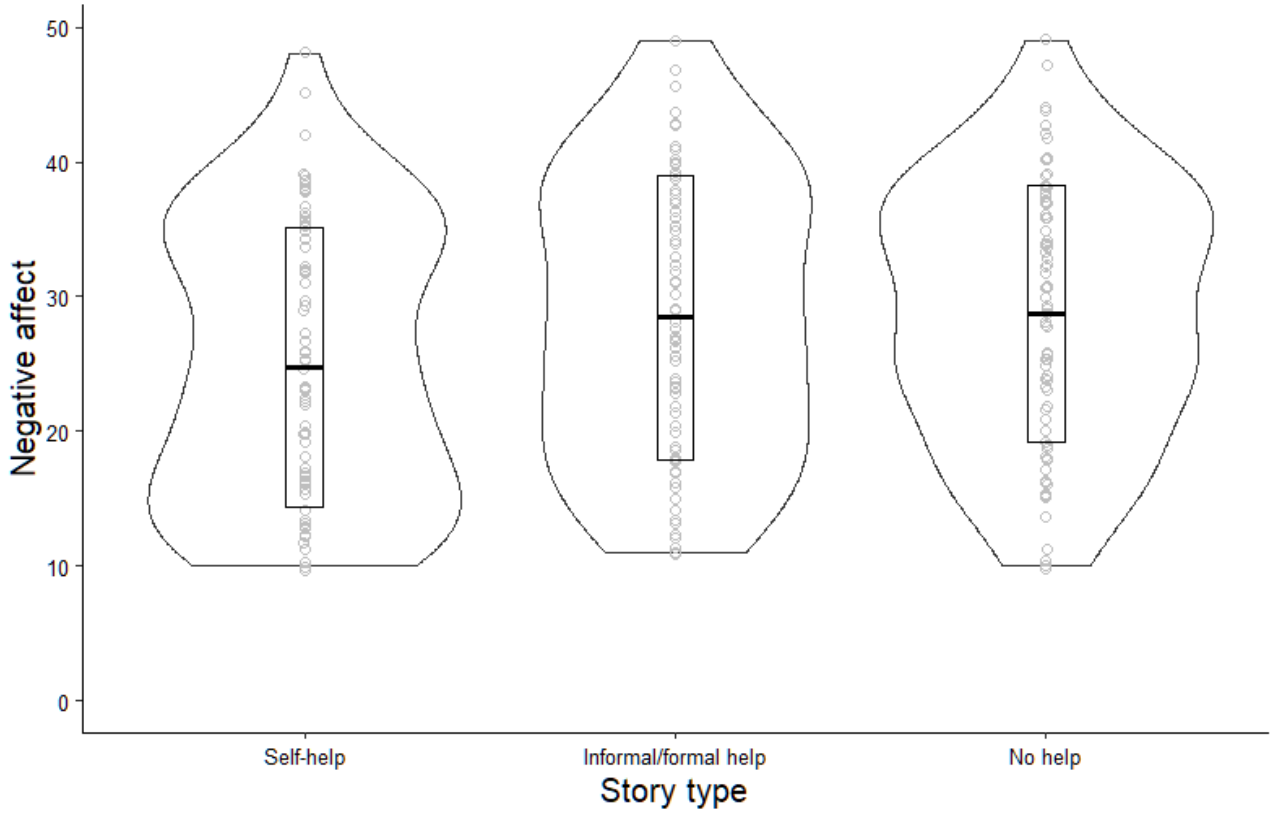
We found no evidence for an effect of story type on positive affect ($F_{2,\ 235}$=0.32; P=.73; η^2=0.002) or entrapment ($F_{2,\ 235}$=0.33; P=.72; η^2=0.003). There was also no evidence for an effect of story type on future self-harm (likelihood of self-harm after reading stories: $F_{2,\ 235}$=2.40; P=.09; η^2=0.02; likelihood of a future without self-harm: $F_{2,\ 235}$=0.94; P=.39; η^2=0.008). The pattern of results was unchanged following adjustment for age, gender, and help-seeking history (the full results of the model can be found in Tables S3 and S5-S7 in Multimedia Appendix 3).

There was evidence for an effect of story type on negative affect ($F_{2,\ 235}$=3.83; P=.02; η^2=0.03). Adjusting for age, gender, and help-seeking history increased the strength of association between story type and negative affect ($F_{2,\ 230}$=4.02; P=.02; η^2=0.10; the full results of the model can be found in Table S4 in Multimedia Appendix 3). Pairwise comparisons (Tukey's HSD on the unadjusted means) demonstrated that participants in the "self-help" condition reported lower levels of negative affect after reading the stories compared to participants in the

XSL•FO
RenderX

"no help" condition (mean difference=–3.97, 95% CI –7.72 to –0.22; *P*=.04) and the "informal/formal help" condition (mean difference=–3.70, 95% CI –7.55 to 0.14; *P*=.06). Figure 4 presents the impact of story type on negative affect.

Figure 4. Impact of story type on negative affect.



An ANCOVA was conducted to explore whether self-harm history interacted with story type to influence negative affect. Although there was strong evidence for an effect of self-harm history on negative affect ($F_{1, 232}$=28.94; *P*<.001), evidence for an interaction with story type was not found ($F_{2, 232}$=0.94; *P*=.39). An ANCOVA was also conducted to explore if age interacted with story type to influence negative affect. Although there was strong evidence for an effect of age on negative affect ($F_{1, 232}$=16.85; *P*<.001), evidence for an interaction was not found ($F_{2, 232}$=1.06; *P*=.35). However, the former only achieved 22% power, and the latter 26% power.

Open-Ended Feedback on the Lived Experience Stories

Approximately 87.8% (209/238) of the sample provided written feedback on the stories. Limitations and advantages of the lived experience stories, categorized by the type of story, are reported in Tables S1 and S2 in Multimedia Appendix 4. Below, we have reported noticeable differences in the themes across conditions.

Differences in Perceived Limitations of the Stories Across Conditions

The Recovery Messaging Was Problematic

Within the comments, messages of hope, positivity, and recovery were described as annoying, frustrating, and overwhelming. Some participants expressed dislike toward references to counting the number of days without self-harm. Other comments referred to the difficulty in looking to the future, as that felt dismissive and more challenging than taking it day by day. A participant explained,

> *I don't care if I'm fine by the time I'm 21. I've ruined my teenage years with this, and I can't get that back. Yeah, I hope for a happier future, but when people say that everything is fine because they recovered (...) they're ignoring how horrific the emotions were when they were struggling. Recovery doesn't magically erase the past.* [Participant in "informal/formal help" condition]

Another participant expressed frustration that the narrators in the stories seemed to be working toward recovery due to feeling stigmatized and pressured to conform, rather than of their own accord. The "no help" condition stands out in that more participants appeared to struggle with the inclusion of recovery messaging compared to the other story conditions. This may be because the recovery message appeared without a clear indication of how recovery occurred, as comments reflected the following sentiments,

> *Sometimes words can get lost, eg, "things get better." The term has been used so many times that I have almost become desensitized to it and what it means.* [Participant in "no help" condition]
>
> *Doesn't say how to get better.* [Participant in "no help" condition]

XSL•FO
RenderX

The Stories Were Unrelatable

While some participants in all conditions commented on the stories being unrelatable, this was less pronounced in the "no help" condition. This is plausibly a result of the stories in this condition providing less concrete detail about the narrator's experience, and thus, there being less to be able to relate to. Stories in the "self-help" condition were criticized more than stories in the other conditions for neglecting to acknowledge that "one size does not fit all." This was mostly reflected in the criticism that the stories were too specific, usually because of a focus on younger ages. However, some participants in the "self-help" condition agreed with the sentiment that "the people in the stories seemed to have a better idea of why they started self-harming and why they continued to do it, which I don't have."

The most prevalent reasons why the stories were unrelatable referred to the stories having unrealistic portrayals of self-harm, and instilling feelings of competition and thus feelings of isolation in readers. They also commented that they felt pressure to have "stronger urges" and "scars" as the narrators had. Participants in all conditions found the stories unrealistic to a similar degree; however, more references about the stories seeming sanitized or inauthentic were made by participants in the "self-help" condition. Participants in the "self-help" and "informal/formal help" conditions commented on the stories, making them feel isolated much more than those in the "no help" condition. Several participants emphasized that there is a very competitive nature to self-harm at both extremes:

> *It can sometimes make you feel as if you are not bad enough to deserve help.* [Participant in "self-help" condition]

> *All people who seemed to be able to stop with help. It made me feel bad for still struggling.* [Participant in "informal/formal help" condition]

The Specific Support Mentioned in the Stories (or Lack Thereof) Was Problematic

The stories in the "self-help" condition were criticized for not mentioning professional or medical support. All conditions were criticized for offering nonpractical advice. This was more predominant in the "self-help" condition, followed by the "informal/formal help" condition, and much less in the "no help" condition. The type of self-help methods mentioned were mainly criticized for being methods that readers had already tried and found unsuccessful in aiding recovery; there were also a couple of comments criticizing the focus on short-term coping mechanisms, instead of long-term recovery.

> *After trying the methods they said helped them, I eventually slipped back to self harming. I don't know what else to do.* [Participant in "self-help" condition]

> *They didn't give much detail on what actually helped them other than immediate coping mechanisms.* [Participant in "self-help" condition]

The type of informal/formal help methods mentioned were mainly criticized, as participants felt the stories made accessing those kinds of support sound much easier than it is. For example, a participant stated,

> *Sometimes it may not be as easy to access support services, whilst barriers to access are likely to vary, it felt like the people in the stories were able to somewhat easily access support - the way it should be I guess.* [Participant in "informal/formal help" condition]

Another stated,

> *I'm a first-generation immigrant, mental health was never something to talk about in my household because of cultural reasons. If I had gone to my parents, I know I wouldn't get the same sympathy just frustration.* [Participant in "informal/formal help" condition]

This made readers who did not have supportive friends/family or who had struggled to access formal support feel isolated and distressed,

> *I guess it makes me sad to hear about people having their boyfriend, friends and family around to help them when I feel like I have no one.* [Participant in "informal/formal help" condition]

Differences in Perceived Advantages of the Stories Across Conditions

While the "informal/formal help" condition did not stand out in the number of criticisms received, participants in this condition did seem to have fewer positive things to say about the stories. This was apparent in relation to comments about the stories having a realistic but hopeful outlook of recovery and comments about the authenticity of the stories.

The Stories Included a Realistic But Hopeful Outlook on Recovery

Participants in all conditions appreciated hearing the stories from those who recovered, and particularly the emphasis on recovery being a nonlinear process that takes time. Only the "self-help" stories were praised for highlighting that the path to recovery is not a one-size-fits-all approach, and that readers will find different coping mechanisms and recovery strategies helpful. As mentioned earlier, participants in the "self-help" condition were also more likely to criticize the stories for failing to do exactly this. Only the stories in the "self-help" and "no help" conditions were praised for the discussion of relapses, despite these themes appearing in all conditions.

The Stories Were Relatable

The number of positive comments about the relatability of the stories was similar across conditions. However, there were some differences among the conditions when looking at the subthemes of relatability. Although positive comments about the authenticity of the stories were made about all of the stories, this was more prevalent in the "self-help" and "no help" conditions. All of the stories were commended for accurately portraying the emotions experienced in self-harm, but the stories in the "self-help" and "no help" conditions were further

XSL•FO
RenderX

commended for relaying "lived experience" and seeming "honest and truthful." One participant explained that,

> *It can be hard to accept advice from people that don't fully understand or appreciate just how dark the thoughts can be and how strong these urges and negative thoughts can take over your mind.* [Participant in "no help" condition]

We note that there are inconsistencies between the unrelatable and relatable themes, which we will elaborate on in the discussion.

The Stories Were Supportive

Unsurprisingly, participants in the "no help" condition had fewer positive remarks about the theme of support compared to participants in the conditions that focused on help-seeking. However, this difference was not as substantial, as some participants in this condition felt that hearing others' stories reminded them that they were not alone. It is also worth noting that this feeling of "camaraderie" was reported a similar number of times across conditions. Other aspects that fell under the theme of support seen in the other conditions were: participants in the "self-help" condition appreciated hearing what specific methods worked for others, as it gave them hope, generally, but also ideas about new coping mechanisms to try. Participants in the "informal/formal help" condition appreciated the emphasis on the importance of accessing support for recovery. The stories highlighted to them how helpful it was to have both informal and formal help at the start, but also throughout recovery. The stories could help to instill hope and confidence,

> *It does get better once you find the right support. There are people who do care about you, you just don't know it.* [Participant in "informal/formal help" condition]

While some participants who had negative or no help-seeking experiences in the past found it difficult to read about others having success, some participants found it reassuring to hear about other people successfully accessing support, as there is concern that accessibility can be restricted.

Discussion

Main Findings

In this study of individuals with recent self-harm experience, we were unable to find clear evidence of an association between the type of content of lived experience stories and help-seeking intentions, entrapment, or positive affect. However, we did find evidence that reading stories about self-help strategies led to a lower prevalence of negative emotions, compared to reading stories about seeking help from others, or stories that made no mention of help-seeking.

One potential explanation for why we did not see an increase in help-seeking intentions after reading a story that emphasizes help-seeking might be due to the relatability of the stories. The open-ended feedback suggested that the stories that mentioned help-seeking were seen as less relatable than the stories that did not mention help-seeking. This demonstrates that it is not only the characteristics of the narrator (eg, age and gender) that the reader might not relate to. They might also struggle to relate to the actions and consequences of help-seeking experienced by the narrator. First, about 12.6% (30/238) of participants reported no history of help-seeking. Others who recounted negative experiences when they sought help (eg, ineffective self-help strategies, negative reactions from family and friends, and inaccessible formal support) reported feeling distressed (ie, feeling like a failure, alone, and hopeless) when they read about others' positive experiences. Further, only the stories that mentioned help-seeking seemed to evoke a feeling that the reader's situation was less serious than the narrator's. As perceived need is an important component of help-seeking decision-making [46], it is feasible that the stories that mentioned help-seeking set a benchmark for when help-seeking is needed. In our focus group study, we similarly found that being able to identify with the narrator of a lived experience story—made more challenging by including specific details of age or circumstances—was "an important factor in encouraging help-seeking through the recognition of need in others and thus oneself" [23].

Findings need to be interpreted in light of some limitations. First, the sample was self-selecting, so it is possible it may not be representative of the wider population; further, there was a considerable degree of attrition, and the reasons for this are unknown. Second, we did not measure help-seeking intentions before and after the exposure, so we cannot determine if there was an overall increase across all participants. Third, the study was powered to detect a main effect of the type of lived experience story, and tests of interactions were exploratory and underpowered. These post hoc findings therefore need to be interpreted with caution, and larger-scale research is needed to understand the role of moderators, such as self-harm history and age. Fourth, as we only measured help-seeking intentions, we do not know if these intentions would manifest into help-seeking behavior. Therefore, future research would benefit from recruiting a larger sample, capturing reasons for deciding not to participate or withdrawing from the study, and including a follow-up assessment to measure whether intentions translated into behavior.

It is worth noting that some of the themes that emerged from the open-ended feedback questions contradicted one another. For example, the self-help stories seemed to arouse thoughts about self-harm and recovery being very individualistic experiences, but they were both praised for doing this well and critiqued for doing it poorly. Similarly, the same aspect of a story could be a positive that made the story more relatable for one participant, and a negative addition that made the story less relatable for another. While some attempts were made to reconcile contradictions like this when interpreting the data, we were conservative in our approach. First, because we felt that the contradictions were the result of a broad focus. Feedback was collected from a large number of individuals, and that feedback could be directed toward various aspects of the stories. Second, the contradictions in themselves may be meaningful, as they indicate how individualistic the experience of self-harm and recovery can be, and emphasize that a delicate balance needs to be found in these stories if hoping to apply them broadly.

XSL•FO
RenderX

Overall, there was broad consistency between the themes that emerged from the open-ended feedback collected in this study and those that emerged during the focus group discussions with 5.5% (13/238) of the original sample [23]. Both studies highlighted the importance of lived experience stories to be relatable, but that a careful balance had to be struck between specificity and ambiguity. Both studies also highlighted the importance of lived experience stories to feel authentic (avoiding both dramatization and sanitization) and that authenticity is a key factor in how relatable a story is to the reader. Previous research on mental health recovery narratives also emphasizes the importance of authenticity to readers [47,48]. A systematic review on peer support for self-harm found that shared experience was an important component in making the support they received feel genuine [49]. While the template stories were written by those with lived experience, it was up to the research team's discretion about how to modify them for the purpose of this study, and this might have impacted the relatability or other aspects of the stories. Therefore, we suggest using a co-design approach in future work.

In line with the findings of Winstone et al [23], this study also highlighted the importance of lived experience stories to offer practical advice and to portray the journey to recovery realistically. This finding might be specific to self-harm recovery, as another study, which more broadly focused on participants with any mental health concern and excluded narratives that described harmful behavior (eg, self-harm), found participants reported a preference for narratives about upward trajectories toward recovery over narratives that focused on only snapshots of the journey or nonlinear journeys [47].

Results from the focus groups suggest that reading the stories had a generally positive impact on mood [23]. We did not measure baseline mood in this study, and we cannot assume that if we did that the results would support the qualitative findings, as the focus group sample may have been subject to sampling bias, as participants who felt the stories negatively impacted their mood might have been less likely to enroll in the follow-up qualitative study. Nonetheless, this study did demonstrate that participants who read stories about self-help strategies reported lower levels of negative emotions, compared to the other two groups. The open-ended feedback suggested that reading about self-help strategies that were successful for others left some readers feeling hopeful that recovery might be possible for them too. It also gave them ideas of new strategies to try, inspiring them to act toward achieving their own recovery. It is possible that these positives may have increased their self-efficacy [50] to recover, thereby increasing their mood and decreasing feelings of needing to rely on others for help. While this interpretation is speculative, further investigation is warranted since an unintended negative consequence may be a decreased likelihood of seeking help from others.

The results from this study suggest that readers might react more negatively to themes of recovery in the absence of practical advice on how to recover. The results from this study also add the caveat that it would be beneficial to mention various help-seeking strategies within the same story. This might prevent the reader from feeling demoralized (and avoid seeking help) after reading about someone else successfully obtaining help via a strategy that was unsuccessful for the reader or is a strategy that they are unwilling to try.

Our findings, consistent with the findings of Winstone et al [23], demonstrate that lived experience stories can help readers to "feel validated and help them to better understand their own feelings and experiences" and manage "feelings of isolation, knowing that others are experiencing similar feelings, and engaging in self-harm as a coping mechanism" [23]. However, the results from both studies also highlight that reading about others' experiences can sometimes foster a sense of competition and thereby increase feelings of isolation. Previous research has also highlighted the competitive element of self-harm [19,42,49] and found that individuals sometimes purposefully seek out material that instills feelings of competition when experiencing self-harm urges [42].

One way to prevent lived experience stories from triggering a sense of competition in readers is by avoiding words such as "clean" or other language that might cause readers to feel stigmatized or under pressure to recover. Focus group participants in the follow-up study further explained that stigma can arise when using these terms, because they can unfairly attach moralistic judgments to a medical issue [23]. In addition, it might be helpful to explicitly refer, in lived experience stories, to the competitive nature of self-harm. For example, narrators can emphasize that they are relaying only one out of many stories (that not everyone's self-harm will have the same causes, that not everyone will experience it in the same way, and that not everyone will recover in the same way), but that everyone's experience is valid. Winstone et al [23] further concluded that avoiding numeric references within a story, such as the number of days without self-harm, might help to reduce competition.

The lived experience stories under study were carefully selected to be recovery-oriented and exclude details that might be triggering for someone who self-harms (eg, detailed descriptions of self-harm methods). While this was an important consideration to mitigate risk in this study, it does mean that the results may not generalize to lived experience stories that include more triggering content that one might find on unregulated forums. This will need to be kept in mind if using the results of this study to inform practice.

Further, despite attempts to exclude triggering content from the stories, the results demonstrated that some participants (across all conditions) reported emotional distress as a result of reading the stories. Some participants were negatively impacted by the mention of the specific self-harm method used (ie, cutting). While lived experience stories on forums often include descriptive detail about methods and injuries [51], those hosted on more formal websites (eg, run by charities and health care organizations) usually exclude these kinds of details [22]. However, some participants were negatively impacted by other details that might be found in stories hosted on formal websites, such as hearing about others struggling with self-harm, being reminded of how they harm themselves, or how one's self-harm might impact their loved ones. It was also reported that reading about others still struggling with self-harm made readers feel hopeless that their situation would ever meaningfully improve.

XSL•FO
RenderX

Therefore, future research, using a co-design approach, is needed to identify elements that make a story harmful.

Conclusions

Lived experience stories of self-harm are commonly found on the internet, including on formal help sites, blog posts, and forums. The results of this study provide guidance for those posting or moderating such content. The findings taken together illustrate that the impact of lived experience stories on individuals with a recent history of self-harm is complicated. While the inclusion of help-seeking in a story can help readers to feel validated, less isolated, less helpless, and more hopeful that recovery is possible for them, there is a fine line to tread. The results indicate that the experience of self-harm and the pathway to recovery are very personal, so too many personal details about either experience can make the stories feel unrelatable. However, there are certain shared experiences of self-harm that can be included to ensure a lived experience story feels authentic. Therefore, the results suggest that stories should aim to foster both a sense of individuality as well as community. For example, stories should focus less on the characteristics of narrators, but more on their emotions and experience of self-harm, to which readers can more easily relate. Stories should also propose multiple strategies for achieving recovery, and emphasize that recovery will take time and that there may be setbacks along the way. Future experimental work is needed to confirm whether these adaptations to lived experience stories about help-seeking can motivate individuals to seek help for their self-harm. Future experimental work should also incorporate suggestions on how to limit bias and improve the methodology used in this study. Lived experience stories of self-harm that find the right balance of many of these nuances have the potential to encourage those struggling with self-harm urges to engage in self-help practices or reach out to others for support.

Acknowledgments

This study was funded by the National Institute for Health and Care Research (NIHR) Biomedical Research Centre at the University Hospitals Bristol and Weston National Health Service Foundation Trust and the University of Bristol. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. BM is funded by a Medical Research Foundation fellowship (MRF-058-0017-F-MARS-C0869). LW is supported by the NIHR School for Public Health Research (SPHR) (NIHR204000). PM is funded by the University Hospitals Bristol and Weston National Health Service Foundation Trust and the University of Bristol. PM and LB are part-funded by the NIHR Applied Research Collaboration (ARC) West. The funders were not involved in the study design; collection, analysis, and interpretation of the data; writing of the report; or decision to submit the paper for publication. We would like to thank Jenna Selvey for assisting with coding the open-ended feedback on the lived experience stories.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Story templates.
[DOCX File , 33 KB - humanfactors_v13i1e71280_app1.docx]

Multimedia Appendix 2
Participant characteristics.
[DOCX File , 33 KB - humanfactors_v13i1e71280_app2.docx]

Multimedia Appendix 3
Outcome measures.
[DOCX File , 32 KB - humanfactors_v13i1e71280_app3.docx]

Multimedia Appendix 4
Open-ended feedback.
[DOCX File , 40 KB - humanfactors_v13i1e71280_app4.docx]

References

1. Boyce N. Pilots of the future: suicide prevention and the internet. Lancet 2010;376(9756):1889-1890. [doi: 10.1016/s0140-6736(10)62199-x] [Medline: 21155079]
2. Brennan C, Saraiva S, Mitchell E, Melia R, Campbell L, King N, et al. Self-harm and suicidal content online, harmful or helpful? A systematic review of the recent evidence. JPMH 2022;21(1):57-69. [doi: 10.1108/jpmh-09-2021-0118]
3. Baker D, Fortune S. Understanding self-harm and suicide websites: a qualitative interview study of young adult website users. Crisis 2008;29(3):118-122. [doi: 10.1027/0227-5910.29.3.118] [Medline: 18714907]

XSL•FO
RenderX

4. Harris IM, Roberts LM. Exploring the use and effects of deliberate self-harm websites: an internet-based study. J Med Internet Res 2013;15(12):e285 [FREE Full text] [doi: 10.2196/jmir.2802] [Medline: 24362563]
5. Shaw LH, Gant LM. In defense of the internet: the relationship between internet communication and depression, loneliness, self-esteem, and perceived social support. Cyberpsychol Behav 2002;5(2):157-171 [FREE Full text] [doi: 10.1089/109493102753770552] [Medline: 12025883]
6. Whitlock JL, Powers JL, Eckenrode J. The virtual cutting edge: the internet and adolescent self-injury. Dev Psychol 2006;42(3):407-417. [doi: 10.1037/0012-1649.42.3.407] [Medline: 16756433]
7. Lai MH, Maniam T, Chan LF, Ravindran AV. Caught in the web: a review of web-based suicide prevention. J Med Internet Res 2014;16(1):e30 [FREE Full text] [doi: 10.2196/jmir.2973] [Medline: 24472876]
8. Luxton DD, June JD, Kinn JT. Technology-based suicide prevention: current applications and future directions. Telemed J E Health 2011;17(1):50-54. [doi: 10.1089/tmj.2010.0091] [Medline: 21214382]
9. Robinson J, Cox G, Bailey E, Hetrick S, Rodrigues M, Fisher S, et al. Social media and suicide prevention: a systematic review. Early Interv Psychiatry 2016;10(2):103-121. [doi: 10.1111/eip.12229] [Medline: 25702826]
10. Scherr S, Reinemann C. First do no harm: cross-sectional and longitudinal evidence for the impact of individual suicidality on the use of online health forums and support groups. Comput Hum Behav 2016;61:80-88. [doi: 10.1016/j.chb.2016.03.009]
11. Frost M, Casey L. Who seeks help online for self-injury? Arch Suicide Res 2016;20(1):69-79. [doi: 10.1080/13811118.2015.1004470] [Medline: 25706352]
12. Mars B, Heron J, Biddle L, Donovan JL, Holley R, Piper M, et al. Exposure to, and searching for, information about suicide and self-harm on the internet: prevalence and predictors in a population based cohort of young adults. J Affect Disord 2015;185:239-245 [FREE Full text] [doi: 10.1016/j.jad.2015.06.001] [Medline: 26150198]
13. Biddle L, Derges J, Goldsmith C, Donovan JL, Gunnell D. Online help for people with suicidal thoughts provided by charities and healthcare organisations: a qualitative study of users' perceptions. Soc Psychiatry Psychiatr Epidemiol 2020;55(9):1157-1166 [FREE Full text] [doi: 10.1007/s00127-020-01852-6] [Medline: 32157324]
14. Greidanus E, Everall RD. Helper therapy in an online suicide prevention community. Br J Guid Couns 2010;38(2):191-204. [doi: 10.1080/03069881003600991]
15. Padmanathan P, Biddle L, Carroll R, Derges J, Potokar J, Gunnell D. Suicide and self-harm related internet use. Crisis 2018;39(6):469-478 [FREE Full text] [doi: 10.1027/0227-5910/a000522] [Medline: 29848080]
16. Biddle L, Derges J, Goldsmith C, Donovan JL, Gunnell D. Using the internet for suicide-related purposes: contrasting findings from young people in the community and self-harm patients admitted to hospital. PLoS One 2018;13(5):e0197712 [FREE Full text] [doi: 10.1371/journal.pone.0197712] [Medline: 29795637]
17. Biddle L, Donovan J, Hawton K, Kapur N, Gunnell D. Suicide and the internet. BMJ 2008;336(7648):800-802 [FREE Full text] [doi: 10.1136/bmj.39525.442674.AD] [Medline: 18403541]
18. Lewis SP, Seko Y. A double-edged sword: a review of benefits and risks of online nonsuicidal self-injury activities. J Clin Psychol 2016;72(3):249-262. [doi: 10.1002/jclp.22242] [Medline: 26613372]
19. Susi K, Glover-Ford F, Stewart A, Knowles Bevis R, Hawton K. Research review: viewing self-harm images on the internet and social media platforms: systematic review of the impact and associated psychological mechanisms. J Child Psychol Psychiatry 2023;64(8):1115-1139. [doi: 10.1111/jcpp.13754] [Medline: 36940718]
20. McDermott E, Roen K, Piela A. Hard-to-reach youth online: methodological advances in self-harm research. Sex Res Soc Policy 2013;10(2):125-134. [doi: 10.1007/s13178-012-0108-z]
21. Daine K, Hawton K, Singaravelu V, Stewart A, Simkin S, Montgomery P. The power of the web: a systematic review of studies of the influence of the internet on self-harm and suicide in young people. PLoS One 2013;8(10):e77555 [FREE Full text] [doi: 10.1371/journal.pone.0077555] [Medline: 24204868]
22. Cohen R, Rifkin-Zybutz R, Moran P, Biddle L. Web-based support services to help prevent suicide in young people and students: a mixed-methods, user-informed review of characteristics and effective elements. Health Soc Care Community 2022;30(6):2404-2413 [FREE Full text] [doi: 10.1111/hsc.13819] [Medline: 35514202]
23. Winstone L, Mars B, Ferrar J, Moran P, Penton-Voak I, Grace L, et al. Investigating how people who self-harm evaluate web-based lived experience stories: focus group study. JMIR Ment Health 2023;10:e43840 [FREE Full text] [doi: 10.2196/43840] [Medline: 36719729]
24. Bechmann A, Lomborg S. Mapping actor roles in social media: different perspectives on value creation in theories of user participation. New Media Soc 2012;15(5):765-781. [doi: 10.1177/1461444812462853]
25. Bailey ER, Matz SC, Youyou W, Iyengar SS. Authentic self-expression on social media is associated with greater subjective well-being. Nat Commun 2020;11(1):4889 [FREE Full text] [doi: 10.1038/s41467-020-18539-w] [Medline: 33024115]
26. O'Connor RC, Portzky G. The relationship between entrapment and suicidal behavior through the lens of the integrated motivational-volitional model of suicidal behavior. Curr Opin Psychol 2018;22:12-17. [doi: 10.1016/j.copsyc.2017.07.021] [Medline: 30122271]
27. Willliams JMG. Suicide and Attempted Suicide: Understanding the Cry of Pain. 2nd ed. London: Penguin Books; 2002.
28. Ferrar J, Penton-Voak I, Biddle L, Moran P, Mars B. 2022.03.14 SH study 1 protocol V1.0.pdf. Open Science Framework. 2022. URL: https://osf.io/qtgvp [accessed 2025-08-26]

XSL•FO
RenderX

29. Ferrar J, Penton-Voak I, Biddle L, Moran P, Mars B. Investigating the impact of lived experience stories on self harm, mood, and help-seeking intentions. OSF Registries. 2025. URL: https://osf.io/cesda [accessed 2025-08-26]
30. Anwyl-Irvine AL, Massonnié J, Flitton A, Kirkham N, Evershed JK. Gorilla in our midst: an online behavioral experiment builder. Behav Res Methods 2020;52(1):388-407 [FREE Full text] [doi: 10.3758/s13428-019-01237-x] [Medline: 31016684]
31. Faul F, Erdfelder E, Buchner A, Lang A. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
32. Rickwood D, Deane FP, Wilson CJ, Ciarrochi J. Young people's help-seeking for mental health problems. Aust e-J Advanc Ment Health 2014;4(3):218-251. [doi: 10.5172/jamh.4.3.218]
33. Rickwood D, Braithwaite V. Social-psychological factors affecting help-seeking for emotional problems. Soc Sci Med 1994;39(4):563-572. [doi: 10.1016/0277-9536(94)90099-x] [Medline: 7973856]
34. Wilson MG, Morley JE. Impaired cognitive function and mental performance in mild dehydration. Eur J Clin Nutr 2003;57 Suppl 2:S24-S29. [doi: 10.1038/sj.ejcn.1601898] [Medline: 14681710]
35. Gondek D, Kirkbride JB. Predictors of mental health help-seeking among polish people living the United Kingdom. BMC Health Serv Res 2018;18(1):693 [FREE Full text] [doi: 10.1186/s12913-018-3504-0] [Medline: 30189870]
36. Ashworth E, Saini P. Increasing help-seeking intentions for mental health difficulties in early adolescence: the role of cumulative promotive factors. Advers Resil Sci 2024;5(4):429-445. [doi: 10.1007/s42844-024-00137-0]
37. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
38. De Beurs D, Cleare S, Wetherall K, Eschle-Byrne S, Ferguson E, B O'Connor D, et al. Entrapment and suicide risk: the development of the 4-item entrapment scale short-form (E-SF). Psychiatry Res 2020;284:112765 [FREE Full text] [doi: 10.1016/j.psychres.2020.112765] [Medline: 31945600]
39. Oppenheimer DM, Meyvis T, Davidenko N. Instructional manipulation checks: detecting satisficing to increase statistical power. J Exp Soc Psychol 2009;45(4):867-872. [doi: 10.1016/j.jesp.2009.03.009]
40. Mackenzie CS, Gekoski WL, Knox VJ. Age, gender, and the underutilization of mental health services: the influence of help-seeking attitudes. Aging Ment Health 2006;10(6):574-582. [doi: 10.1080/13607860600641200] [Medline: 17050086]
41. Biddle L, Gunnell D, Sharp D, Donovan JL. Factors influencing help seeking in mentally distressed young adults: a cross-sectional survey. Br J Gen Pract 2004;54(501):248-253 [FREE Full text] [Medline: 15113490]
42. Murray CD, Fox J. Do internet self-harm discussion groups alleviate or exacerbate self-harming behaviour? Aust e-J Advanc Ment Health 2014;5(3):225-233. [doi: 10.5172/jamh.5.3.225]
43. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
44. QSR International. NVivo. 2023. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC3930250/ [accessed 2026-01-07]
45. Ferrar J, Mars B. Online stories of self-harm Open. University of Bristol research data repository. 2026. URL: https://data.bris.ac.uk/data/ [accessed 2026-03-03]
46. Mojtabai R, Olfson M, Mechanic D. Perceived need and help-seeking in adults with mood, anxiety, or substance use disorders. Arch Gen Psychiatry 2002;59(1):77-84. [doi: 10.1001/archpsyc.59.1.77] [Medline: 11779286]
47. Ng F, Charles A, Pollock K, Rennick-Egglestone S, Cuijpers P, Gillard S, et al. The mechanisms and processes of connection: developing a causal chain model capturing impacts of receiving recorded mental health recovery narratives. BMC Psychiatry 2019;19(1):413 [FREE Full text] [doi: 10.1186/s12888-019-2405-z] [Medline: 31864315]
48. Rennick-Egglestone S, Ramsay A, McGranahan R, Llewellyn-Beardsley J, Hui A, Pollock K, et al. The impact of mental health recovery narratives on recipients experiencing mental health problems: qualitative analysis and change model. PLoS One 2019;14(12):e0226201 [FREE Full text] [doi: 10.1371/journal.pone.0226201] [Medline: 31834902]
49. Abou Seif N, John-Baptiste Bastien R, Wang B, Davies J, Isaken M, Ball E, et al. Effectiveness, acceptability and potential harms of peer support for self-harm in non-clinical settings: systematic review. BJPsych Open 2022;8(1):e28 [FREE Full text] [doi: 10.1192/bjo.2021.1081] [Medline: 35034666]
50. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]
51. Self-harm discussion and support. Recovery Your Life. 2023. URL: https://www.recoveryourlife.com/forum/forumdisplay.php?f=29 [accessed 2023-09-27]

Abbreviations

AHSQ: Actual Help-Seeking Questionnaire
ANCOVA: analysis of covariance
E-SF: 4-item Entrapment Scale Short Form
GHSQ: General Help Seeking Questionnaire
HSD: Honestly Significant Difference
PANAS: Positive and Negative Affect Schedule

XSL•FO
RenderX

Edited by A Kushniruk; submitted 08.May.2025; peer-reviewed by A Pitman, JA Cox; comments to author 14.Jul.2025; revised version received 04.Dec.2025; accepted 02.Jan.2026; published 04.Mar.2026.

Please cite as:
Ferrar J, Winstone L, Penton-Voak I, Biddle L, Moran P, Grace L, Mars B
Investigating the Impact of Lived Experience Stories on Self-Harm, Mood, and Help-Seeking Intentions: Web-Based Between-Participants Experimental Study in Individuals With Recent Self-Harm
JMIR Hum Factors 2026;13:e71280
URL: https://humanfactors.jmir.org/2026/1/e71280
doi:10.2196/71280
PMID:

©Jennifer Ferrar, Lizzy Winstone, Ian Penton-Voak, Lucy Biddle, Paul Moran, Lydia Grace, Becky Mars. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 04.Mar.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Use of Digital Peer Support for Employee Well-Being: Retrospective Analysis Across Five Large Employers

Farbod Sedaghati[1], PhD; Anya Stetsenko[1], PhD; Ilayda Ozsan McMillan[1], PhD; Harpreet Nagra[1], PhD; Zara Dana[1], PhD

Supportiv, Berkeley, CA, United States

Corresponding Author:
Harpreet Nagra, PhD
Supportiv
2222 Harold Way
Berkeley, CA, 94704
United States
Phone: 1 800 845 0015
Email: harpreet@supportiv.com

Abstract

Background: Anonymous, 24/7 digital peer support (DPS) offers a scalable solution to support employees' emotional well-being. Understanding sociobehavioral factors, such as timing of engagement and the impact of shared resources, can help employers and employee assistance programs (EAPs) integrate digital tools to better support workforce well-being.

Objective: This study examines (1) outcomes via overall sentiment changes among DPS users; (2) sociobehavioral differences between employees who accessed real-time, anonymous DPS during versus outside business hours; (3) differences in sentiment outcomes based on time of use; and (4) the impact of in-session resource sharing on sentiment improvement.

Methods: Using OpenAI's large language model (LLM) GPT-4o-mini with a few-shot learning approach, 24,818 anonymous chat conversations from 13,879 employees at 5 large employers were evaluated for subclinical sentiment variables, including loneliness, sadness, stress, anxiety, depression, despair, helplessness, and optimism.

Results: Distinct activity patterns were observed between employees during and outside business hours, with a median user age of 36 years. During business hours, employees reported higher baseline stress (Δ1.6%), whereas outside business hours, baseline depression (Δ1.7%) and loneliness (Δ1.2%) were higher. After DPS use, LLM-derived negative sentiment scores decreased (loneliness 46% reduced, sadness 45%, stress 46%, anxiety ~39%, depression ~40%, despair ~40%, and helplessness ~38%), and optimism increased (~77%). Outside business hours, users engaged more (~68% of all sessions), remained in sessions 19% longer, and discussed 13% more topics, whereas users during business hours reported greater improvements in depression (Δ2.3%), helplessness (Δ1.9%), and loneliness (Δ0.5%). Resource sharing was associated with greater improvements in loneliness (Δ2.9%), stress (Δ1.3%), anxiety (Δ1.9%), depression (Δ7.3%), despair (Δ3.0%), helplessness (Δ2.8%), and optimism (Δ8.8%), but not sadness.

Conclusions: DPS complements employers' EAPs by addressing employee engagement gaps, reducing barriers to mental health care, and promoting emotional well-being among the workforce.

(JMIR Hum Factors 2026;13:e90431) doi:10.2196/90431

KEYWORDS

digital peer support; employee benefits; employee wellness; workforce well-being; employee assistance programs

Introduction

Background

Workplace emotional well-being is a pressing concern, as increasing numbers of employees suffer from burnout and chronic stress [1,2]. These issues lead to additional workplace problems, including absenteeism, reduced productivity, and mental health concerns [3]. Current mental health tools and employee assistance programs (EAPs) have expanded significantly in scope, evolving from a focus on mental health and substance use to include broader aspects of emotional well-being, such as family health and workplace stress [4]. However, the effectiveness of EAPs has been mixed. A 2017 systematic review concluded that EAPs benefit functioning and presenteeism but have mixed effects on absenteeism and

XSL•FO
RenderX

engagement [5]. Due to the continued shortage of mental health workers, novel strategies are necessary to meet the growing demand for employee emotional support [6].

The integration of digital technology has enabled more immediate, cost-effective, and accessible forms of employee support, such as digital peer support (DPS), to be incorporated into traditional counseling-focused EAP infrastructures [7]. DPS offers an innovative approach to complement traditional EAPs, providing support on demand, in real time, by peers with shared experiences rather than clinicians, thereby promoting social connectedness, significant sentiment improvements, and emotional regulation [7,8]. However, how employees engage with DPS has yet to be studied.

Sociobehavioral Factors

Demographic variables such as age, gender, culture, technology use, and internet access significantly affect the engagement and effectiveness of digital interventions [9-12]. Barriers such as limited technological proficiency, especially among older adults or individuals with low digital abilities, may reduce user engagement and limit program success [13]. Prior research indicates that women are more likely than men to seek mental health support and use digital mental health tools [14-17]. These differences highlight the importance of tailoring support strategies to the needs of different groups. While mobile apps and chatbots using cognitive behavioral techniques are effective in reducing symptoms of depression and anxiety, evidence shows that digital tools are more effective when supplemented with human interaction, whether through peer support or moderated participation [9].

Additionally, promoting awareness of and trust in EAP resources through culturally sensitive education has been linked to enhanced utilization among diverse employee populations [13,18]. Addressing these sociobehavioral challenges through user-centered design, sensitive outreach, and anonymous technology can improve accessibility and inclusivity across demographic groups [13,19].

Digital Peer Support

Research shows that DPS improves emotional well-being, self-efficacy, and social connection and empathy, all of which contribute to positive organizational outcomes [14]. DPS has emerged as a critical modality for delivering emotional health and wellness services through technological platforms [7,20]. It encompasses both synchronous and asynchronous support delivered via chat, forums, video, mobile apps, and other digital communication tools, allowing users to flexibly seek mental health support [20,21]. It offers higher accessibility than more traditional services, as most individuals own a smartphone, including those from lower socioeconomic backgrounds or with serious mental illness [22-24]. Additionally, DPS has fewer barriers than in-person services, which may involve scheduling, transportation, mobility, or discomfort with face-to-face interactions [19,25], as well as tele- or mobile clinical mental health services that may require insurance, appointments, and delays due to formal clinical assessments before care can begin. As a result, DPS facilitates broader participation, particularly among populations that are underserved or less likely to engage in typical care [19,25].

In addition to accessibility, DPS services can be available 24/7, allowing continuity of care and reinforcing self-efficacy and emotional regulation, thereby filling temporal gaps left by typical mental health services [7]. Peer-delivered digital interventions have been shown to improve both psychiatric and medical self-management by empowering individuals to actively participate in their care within a supportive, nonclinical environment [20,21,25]. Reviews of DPS programs, including those for public safety workers, highlight reductions in stress and trauma [26-28], while workplace-based engagement is associated with increased belonging, accountability, and a sense of community [20,21,25]. DPS is also unique in its ability to integrate lived experience into its design, development, and delivery [21,29,30]. The involvement of peers and their lived experiences increases the relevance and acceptability of interventions [31].

The timing of user engagement is affected by a variety of sociobehavioral factors. Users who seek DPS after work hours may experience heightened levels of emotional distress, loneliness, insomnia, or anxiety, and may engage in more spontaneous or emotionally charged exchanges, whereas interactions during business hours tend to be more structured [7,20,32]. Additional barriers such as privacy concerns, stigma, and work-hour constraints can also reduce engagement during regular business hours, highlighting the need for after-hours availability of support [18,33]. Engagement patterns are also affected by demographic variables such as age, gender, employment status, education level, and housing conditions [34,35]. Preferences for specific technologies, such as text-based versus visual platforms, may vary depending on users' mental health conditions and comfort levels [36]. For instance, some individuals may avoid video-based support and prefer text-only interactions, seeking privacy, cultural alignment, or emotional safety [37].

DPS has also demonstrated measurable improvements in emotional well-being. Services report rapid reductions in sadness (57.5%) and loneliness (55.0%), with sentiment changes emerging within minutes of engagement [8]. These benefits may be understood through relational frameworks such as social penetration theory, which suggests that trust and intimacy in relationships deepen over time, especially when individuals feel secure in their anonymity and in control over self-presentation [38]. These dynamics are often heightened during after-hours use, when users may feel more emotionally vulnerable and open, forming meaningful peer connections [32,38]. Taken together, these findings point to the critical role of timing, sociobehavioral context, and technological preferences in shaping DPS engagement.

Role of Resource Sharing in DPS

A central mechanism of DPS effectiveness is resource sharing, which promotes mental health literacy, reduces stigma, and encourages help-seeking behaviors [27,28]. Sharing success stories and testimonials has been shown to increase employee engagement within EAPs, supporting a workplace culture that values mental health and well-being [27]. Workplace-related

XSL•FO
RenderX

resource sharing also enhances coping, particularly in managing remote work, and reduces isolation [26-28]. DPS also offers timely, subclinical support during a shortage of behavioral health clinicians, helping users access resources when needed without waitlists [6,27]. Access to tailored resources may be especially valuable for underserved or marginalized populations, helping to reduce equity gaps in mental health care [13].

This study examines how sociobehavioral factors, such as the timing of support access, levels of perceived emotional distress, and in-session resource sharing, shape employee engagement and sentiment outcomes within a DPS environment. By analyzing user data across 5 large employers, we aim to identify key trends that can inform the design and implementation of digital emotional support services within workplace well-being programs.

Study Objectives

This study examines how timing and in-session support influence sentiment outcomes within an anonymous, synchronous DPS service, Supportiv. Specifically, we investigate (1) overall sentiment changes across users of DPS; (2) sociobehavioral differences between users who accessed DPS during versus outside business hours; (3) differences in sentiment outcomes based on the timing of service use (during vs outside business hours); and (4) the impact of targeted in-session resources on users' sentiment changes.

Methods

Supportiv DPS Service Model

Supportiv is an anonymous, synchronous, small-group, live-chat DPS service that is available on demand, 24/7. Employees can access the service through links from their employer portals, health plans, EAPs, or related channels. This study analyzes data from employees whose organizations offered Supportiv in addition to their existing EAP programs.

Users voluntarily access Supportiv's anonymous peer-to-peer support service through an informed self-selection process. Engagement begins with a user-initiated response to the free-text prompt, "What's your struggle?" Before accessing services, individuals are required to review and actively consent to the Terms of Service and Privacy Policy, which include consent for the use of their inherently anonymous data—anonymous from the point of collection—for research and service improvement purposes. No personally identifiable information, personal health information, or Health Insurance Portability and Accountability Act (HIPAA)–protected health information is collected.

To reduce the risk of indirect user identification through free-text content, all transcripts were audited using automated methods followed by manual processing before analysis. Potential personal identifiers, including names, locations, or contact information, as well as contextual details that could allow for identification, were omitted from the text. Data were stored and analyzed in deidentified form, and access was limited to authorized personnel. All data handling complies with General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA) requirements, including lawfulness, fairness, transparency, data minimization, purpose limitation, and integrity and confidentiality.

The platform is designed according to data protection by design and by default principles. To mitigate the risk of indirect identification in free-text content, all chats are continuously processed through automated filtering systems that remove explicit identifiers such as names, contact information, social media handles, websites, and physical addresses. In addition, trained human moderators review chats in real time and remove any remaining potentially identifying or narrowing contextual information as needed. These safeguards are applied universally across all chats as part of routine platform operations, not solely for research purposes. Access to research data is restricted to authorized personnel. Together, these safeguards reduce the risk of direct or indirect identification and ensure compliance with applicable privacy regulations.

Supportiv uses artificial intelligence (AI)–driven natural language processing (NLP) to match users with other concurrent users experiencing similar life challenges, forming small, synchronous groups of no more than 5 peers. Each group is facilitated in real time by a trained subclinical human moderator. Users are automatically placed into a chat session that may include other participants and the moderator. Moderators facilitate peer-to-peer discussion, maintain psychological safety by enforcing community guidelines, and share contextually relevant, useful resources identified through AI to match users' expressed needs. Moderators also guide collaborative problem-solving and, in cases of crisis, promptly refer users to professional crisis services, ensuring comprehensive support.

Study Design and Participant Enrollment

During the study period, 40,431 total chat sessions were recorded between January 2023 and December 2024. This retrospective observational study analyzed 24,818 anonymous, live user chat sessions from 13,879 employees after excluding sessions with fewer than 3 messages. Data were securely stored in the Supportiv database. Users consisted of employees from 5 large companies that implemented Supportiv's digital peer-to-peer support service to complement their existing EAPs. The main objective was to evaluate the service's impact on employee well-being through sentiment outcomes derived from chat session transcript analysis. The primary outcomes were within-session sentiment changes across 8 key dimensions: loneliness, sadness, stress, anxiety, depression, despair, helplessness, and optimism. Secondary descriptive outcomes included engagement metrics, use patterns, and resource-sharing outcomes. In addition to anonymous chat content, limited metadata, such as user location at the city level, derived from nonidentifying technical data, were available to support contextual analysis and geographic segmentation. No personally identifiable information or personal health information was collected.

Ethical Considerations

As this study used a large language model (LLM) to identify sentiment scores associated with user language in DPS live chats, several ethical considerations must be addressed [39]. Concerns related to privacy are mitigated by the use of

XSL•FO
RenderX

anonymous text data and the absence of user data collection. Additionally, sentiment scores were used to assess within-session emotional benefits rather than to make clinical diagnoses. All chats were facilitated by trained subclinical human moderators to support accurate sentiment interpretation and constructive DPS sessions. This retrospective study utilized operational data and did not involve direct human participant research. It was determined to be exempt from human participant research review by the Alpha Institutional Review Board (protocol IRBEMP042025).

Conversation Setting

Each participant was matched to a chat within approximately 30 seconds of responding in free text to a single prompt: "What's your struggle?" This response, referred to as the "user-chat-session struggle," determined whether the user was placed into a live, ongoing small-group chat discussion on the same topic or into a 1-on-1 session with a moderator, to which others with similar struggles could later be added. Each chat could include up to 5 users and a trained subclinical professional moderator.

Moderators facilitated discussions using motivational interviewing–based OARS (open-ended questions, affirmations, reflective listening, and summaries) techniques, including open-ended questions, affirmations, reflective statements, summarizations, and structured information sharing [40]. These strategies supported conversation flow, encouraged engagement, and promoted collaborative problem-solving. To maintain continuity, moderators provided a summary for new participants joining mid-session. Moderators played a key role not only in guiding discussions and fostering a psychologically safe environment but also in reviewing AI-recommended resources, such as practical articles and informational materials tailored to users' expressed needs, to ensure contextual sensitivity before sharing them with users.

Sentiment Analysis

Each user chat was analyzed independently, even if the same user participated in multiple sessions. These individual sessions were referred to as "user-chat sessions." Although some users returned for multiple sessions, the service's anonymous nature required that each chat experience be analyzed independently, without linking sessions to the same individual. This approach ensured that each interaction was assessed in isolation, providing a clearer and more user-focused analysis of sentiment changes per intervention.

To analyze the target sentiments (loneliness, sadness, stress, anxiety, depression, despair, helplessness, and optimism), OpenAI's LLM model, GPT-4o-mini, was leveraged using a few-shot learning technique. This method involves presenting the model with a limited number of labeled examples within the prompt to guide its outputs [41]. Previous research has demonstrated that this approach matches or exceeds the performance of state-of-the-art models [8]. Despite agreement with expert human ratings established in previous work, the LLM-derived sentiment scores should not be regarded as psychometric instruments for clinical diagnoses. These scores are indicators of expressed emotional content rather than measures of clinical symptom severity.

A user session was defined by the initial "user-chat-session struggle," consisting of a brief free-text response of up to 300 characters, along with all subsequent messages exchanged during the chat. When a user joined a new chat session at a later time, they submitted a new struggle prompt, initiating a separate and independent session that was analyzed separately. Sentiment changes were tracked for each user within a session. Every message was assessed for sentiment intensity on a 1-10 scale, where a score of "1" indicated low intensity, "5" indicated moderate intensity, and "10" indicated high intensity for the relevant sentiment. This scale provided sufficient granularity to detect subtle variations while maintaining interpretability. The use of sentiment scoring with this model has been previously validated and reported in a prior study [8] (Multimedia Appendix 1), demonstrating strong agreement with expert human assessments. The model's performance is continuously monitored for false positives and false negatives, and flagged misclassifications are reviewed to iteratively fine-tune the model. Importance sampling is employed to prioritize messages with high uncertainty and those near the model's decision boundary. This strategy allows human experts to review and further refine the model's detection of language that may involve linguistic nuances, such as indirect, metaphorical, or sarcastic expressions.

Examples illustrating scores of 1, 5, and 10 are provided in Multimedia Appendix 1. Messages not suitable for sentiment analysis, such as clarifying or irrelevant statements, were excluded (eg, "Or a skill," "What do you mean," "At what time?"). To reduce misclassification, the model was instructed to refrain from assigning scores when insufficient information was available to infer sentiment content.

Due to the DPS service's voluntary live chat model, validated multiitem self-report instruments such as the 9-item Patient Health Questionnaire (PHQ-9), 7-item Generalized Anxiety Disorder (GAD-7), or 10-item Perceived Stress Scale (PSS-10) were not administered to users. Although ecological momentary assessments and visual analog scales can support in-the-moment measurement, they are typically used in structured research protocols and rely on repeated prompts and user incentives. By contrast, peer support environments prioritize user autonomy and natural conversational flow; therefore, repeated prompts may be disruptive. Additionally, many symptom instruments rely on self-report, which is prone to bias, social pressure, reduced introspection, distortion, and response fatigue. In this context, LLM-assigned sentiment scores were used as conversationally embedded measures rather than validated instruments. These scores were based on the language used by participants. Our previous study demonstrated that this approach is comparable to state-of-the-art models [8]. Additionally, rather than indicating diagnostic severity, sentiment scores were used to assess within-session changes. This is consistent with prior research using NLP to evaluate digital interventions at scale [42-44]. These scores were not intended to represent diagnostic severity; instead, they provide scalable indicators of expressed emotional content to assess within-session sentiment changes. Future validation work may assess how NLP-derived indicators

XSL•FO
RenderX

align with standardized measures (such as PHQ-9, GAD-7, and PSS-10).

Inclusion and Exclusion Criteria

To ensure a meaningful assessment of sentiment trends, specific criteria were applied when selecting chat sessions for analysis. Only conversations in which users engaged beyond a single message were included, as single-message interactions did not provide sufficient data to track sentiment changes over time.

For each user chat session, 8 emotions were screened to determine whether the session contained an interpretable sentiment change. To ensure the analysis focused on sessions with meaningful sentiment expression, a threshold score of 5 was established. For negative emotions (loneliness, sadness, stress, anxiety, depression, helplessness, or despair), sessions were included if the initial score was greater than 5, whereas for the positive emotion (optimism), a score below 5 was required. The initial emotional state was defined using the highest sentiment score from the user's first 3 messages, recognizing that individuals often begin conversations with brief or casual exchanges before revealing more meaningful emotional content and underlying struggles. This approach captured users who eventually expressed their emotional concerns while filtering out sessions lacking sufficient sentiment depth for analysis.

Sentiment Analysis Across All User Chat Conversations

To evaluate sentiment states over time across user chat sessions, the sentiment data were standardized to allow comparison despite variation in conversation duration. Each conversation was normalized to a relative scale (from 0 to 1) using bins of size 0.01. If a message with a sentiment score was present within a bin, the score was assigned to that bin. This normalization process allowed sentiment trends to be compared across conversations.

After aggregating the binned data for each sentiment, a linear trend line was fitted. The slope of this line indicated the direction and magnitude of sentiment change, reflecting the service's effectiveness in supporting well-being. The goodness of fit was evaluated using the coefficient of determination (R^2). Finally, CIs were calculated to represent the standard error of the data, indicating variability around the mean for each conversation duration bin.

Propensity Matching

Propensity matching was used to create comparable cohorts for each research question. Users within the cohorts were matched by training a logistic regression classifier using the following features: user-chat-session topic, device type, user country, number of users in the chat, conversation duration, engagement rate (number of characters typed by the user per unit of conversation time), and initial sentiment score. This matching ensured a high level of alignment between users in the 2 cohorts at the start of the chat by incorporating both demographic and behavioral metrics.

To ensure meaningful matching, conversation duration and engagement rate were transformed into categorical bins using predefined ranges. For conversation duration, conversations in each cohort were divided into the following bins based on length (in minutes): "0-5," "5-10," "10-20," "20-45," "45-60," and ">60." Similarly, for engagement rate (measured as the number of characters typed by the primary user per unit of time), bins were created to categorize conversations based on characters typed per unit time: "0-10," "10-20," "20-30," "30-40," "40-50," and ">50."

The logistic regression model produced a propensity score for each user chat session, representing the likelihood of assignment to one cohort. We used 25% of the SD of all propensity scores as the caliper (radius of search) for matching. For each user chat in 1 cohort, the closest match in the other cohort was identified using a k-nearest neighbor classifier within the caliper. Propensity matching was conducted separately for each trend analyzed.

The topics for user chat struggles were categorized into the following themes (alphabetically): anxiety, caregiving/parenting, depression, financial/social determinants of health, identity, loneliness, mental health conditions, physical health/medical, relationship issues, social connection and belonging, stress, suicide/self-harm, trauma, and work/productivity/burnout. These categories were identified through manual inspection of chats to determine common themes across the population.

Analyzing Employee User Experiences

To gain a comprehensive understanding of employee users and characterize their engagement with the service, several key components were analyzed. First, although the service is anonymous, users may voluntarily provide demographic information and optional feedback at the end of a chat session. Analysis of these data offers valuable insights into the study population.

The primary target group consisted of employees from 5 large employers across varied industry sectors, including retail, media, and health care. As these organizations differed in job roles and workforce structure, engagement patterns were analyzed across heterogeneous contexts, enhancing ecological validity and reducing bias. User engagement patterns were examined throughout the day to determine whether usage varied between business and nonbusiness hours. For this analysis, business hours were defined as Monday through Friday from 8:00 AM to 5:00 PM local time, while all other periods, including evenings, nights, and weekends, were categorized as nonbusiness hours.

A core advantage of the Supportiv service is its ability to deliver highly personalized resources tailored to users' stated struggles and needs. Supportiv enables moderators to share coping tools, self-help materials, and other resources, as well as links to relevant employee benefits or community resources, specifically tailored to a user's needs based on context. These resources include various formats such as articles, videos, audio files, tools, worksheets, exercises, community resources, and client-specific materials. Resource sharing is driven by a 2-stage AI recommendation system. The AI model retrieves relevant items that match users' expressed needs from a resource library and then ranks the retrieved items to ensure that the most

XSL•FO
RenderX

appropriate and valuable options are prioritized. Moderators review and select from this AI-curated list before sharing resources with users. The analysis examined the types of resources recommended by AI and shared by moderators, as well as how users interacted with them.

Lastly, although all target users were employees utilizing Supportiv through sponsored links (via their EAPs, employee benefits portals, or other employer-specific URLs), moderators may refer employees to additional relevant EAP services based on their expressed needs. These EAP-related metrics were analyzed to better understand how employees used integrated support services and how DPS functioned within the organizational benefits system.

Sentiment Trends

Variables Analyzed

Two primary variables were independently analyzed for their influence on user sentiment trends: (1) time of engagement and (2) in-session resource sharing.

Time of Engagement

We hypothesized that users engaging with the service outside business hours would exhibit different sentiment outcomes compared to those using it during business hours. To evaluate this, chat sessions initiated during business hours were compared with those started outside business hours, with the user's time zone used to convert the system's default time to local time for cohort classification.

Resource Sharing

We also hypothesized that receiving a shared resource would promote positive sentiment trends, regardless of the type of resource or the extent of user interaction with it.

For both variables, sentiment evolution throughout each chat session was analyzed, and propensity matching was employed to create 2 comparable cohorts for each target variable. This matching procedure ensured a fair comparison between groups to effectively assess how the effects of time of engagement and in-session resource sharing shape sentiment outcomes.

Statistical Analysis

Statistical analyses were conducted to (1) compare differences between groups for specific features, (2) evaluate the significance of predicted sentiment trends, and (3) compare sentiment improvement metrics (slopes) between target cohorts. Statistical significance was defined as $P<.05$, and all tests were 2-tailed. Independent contractors were engaged to develop the analysis plan and conduct statistical analyses to minimize commercial bias. Effect sizes for pairwise comparisons were computed using Cohen *d*.

All analyses were performed using Python 3.9.6 (Python Foundation) with the following libraries: pandas 2.3.3, NumPy 2.0.2, SciPy 1.13.1, scikit-learn 1.6.1, statsmodels 0.14.6, matplotlib 3.9.4, and seaborn 0.13.2.

Topic Modeling

We utilized a topic-modeling framework based on LLM-driven categorization to assign topics and subtopics to text data. After reviewing the conversations, a list of topics was manually defined. All texts were then processed using GPT-4o-mini with a structured prompt to categorize them into the predefined categories. Classification accuracy was manually verified, and, when necessary, categories were corrected manually.

Results

Chat Statistics

A total of 40,431 chat sessions were recorded during the study period. Sessions containing only 1 or 2 messages were excluded, as they lacked sufficient context to evaluate sentiment trends or changes over time. After applying these criteria, 24,818 chat sessions from 13,879 employees with at least three messages were included in the analysis.

User Reported Struggle Topics

Users often reported more than 1 concern during their chat sessions. The most common struggles were relationship issues (7531/24,818, 30.34%), trauma (2663/24,818, 10.73%), workplace concerns (2650/24,818, 10.68%), stress (2074/24,818, 8.36%), loneliness (1757/24,818, 7.08%), depression (1570/24,818, 6.33%), and financial/social determinants of health (1565/24,818, 6.31%). Additional struggles included support system challenges (1160/24,818, 4.67%), anxiety (1098/24,818, 4.42%), mental health conditions (723/24,818, 2.91%), identity concerns (638/24,818, 2.57%), caregiving/parenting (617/24,818, 2.49%), physical health/medical concerns (399/24,818, 1.61%), and suicide/self-harm (373/24,818, 1.50%).

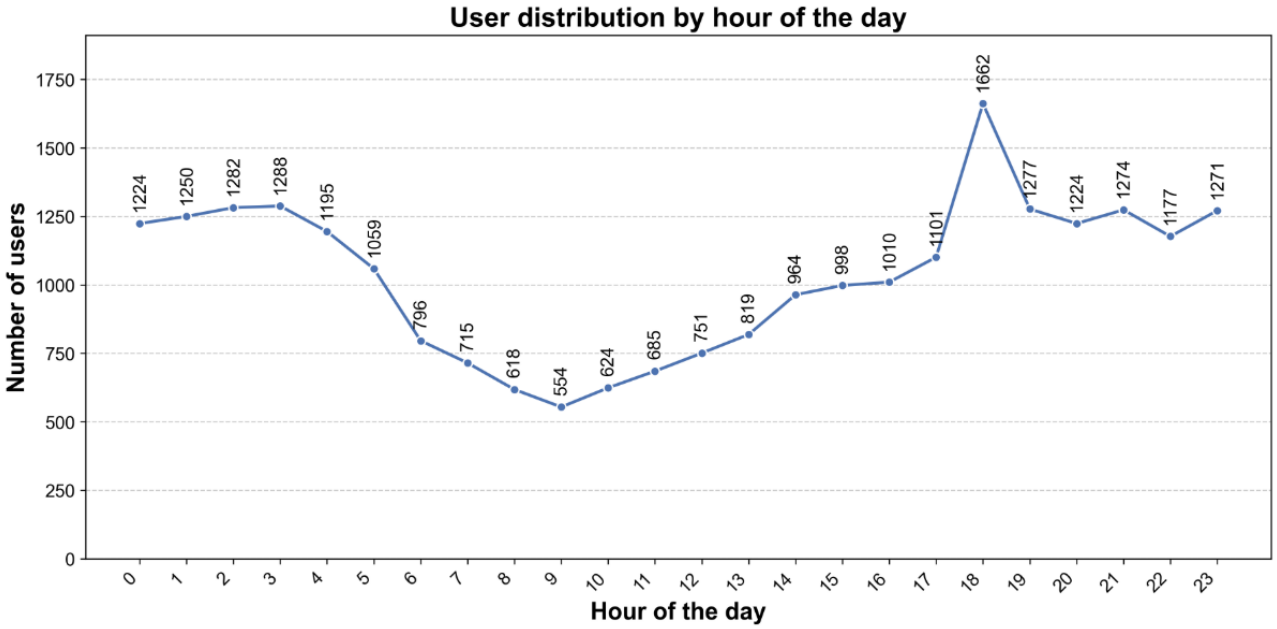
User Activity

On average, users spent 37.97 (SD 35.51) minutes in peer group chats. The average number of messages sent by the primary user was 24.42 (SD 29.45). Variation in chat duration, as reflected by time spent and number of messages, indicates that the analysis includes both users who engaged in extended conversations and those who participated for shorter periods but met the minimum inclusion criteria. To better capture user participation intensity, a metric termed "characters per chat duration" was calculated, with an average of 70.65 (SD 1394.24) characters per minute. The high SD in this metric may be attributed to differences in users' communication styles. While some users type long, paragraph-style messages, others send multiple short messages consisting of only a few words at a time.

Figure 1 illustrates the number of users throughout the day. User activity decreases during late-night hours, increases steadily as business hours begin, and peaks toward the end of the day before declining at night. Notably, user engagement was higher outside business hours, suggesting that employees may be more likely to use DPS during their personal time, when they are away from job responsibilities and have more space to share their thoughts and concerns without distraction.

XSL•FO
RenderX

Figure 1. Distribution of conversation start times based on local time over a 24-hour period. The graph shows the frequency of user chat sessions throughout the day.



Study Population

Demographics

Employees from all global regions accessed the DPS service, with the majority of conversations (16,720/20,976, 79.71%, chats with reported location information) located in the United States. Most users accessed the service via mobile devices (19,854/24,817, 80%), followed by desktop (4467/24,817, 18%) and tablets (496/24,817, 2%). Although the service is anonymous, users have the option to voluntarily share demographic information, including age, race, and gender. Out of 24,818 anonymous chat conversations, 8647 (34.84%) users disclosed demographic information, which is a relatively high response rate compared with similar digital mental health services [45]. While demographic data were not available for all users, the information that was voluntarily provided offered valuable insights and was incorporated into the propensity matching algorithm to enhance comparative sentiment analyses among similar users.

Table 1 presents the distribution of user ages, grouped into 3-year increments. Overall, the median age of the users is 36 years, with a mean of ~45 years, indicating that DPS attracts a broad adult population. Individuals aged 37-57 generally access psychological services less frequently [46]. However, the accessibility and anonymity of Supportiv's DPS model may make it a more viable option for this group by offering an easier point of entry to mental health support.

In conversations with users who disclosed their gender, 5738 (66.36%) identified as female, 2538 (29.35%) as male, and 371 (4.29%) as nonbinary. This distribution aligns with the literature showing that, despite using digital services more frequently [47], men are significantly less likely to seek mental health support than women [48]. The use of DPS by nonbinary users may reflect its inclusive and anonymous design, which can reduce barriers often encountered in traditional clinical settings. This inclusivity highlights the potential for DPS services to foster psychological safety.

In terms of race and ethnicity among conversations with those who provided their ethnicity/race demographic information (n=8914), 4759 (53.39%) identified as White, followed by African American (n=1469, 16.48%), Hispanic (n=1290, 14.47%), and Asian (n=810, 9.09%) users, with the remainder reporting other minority groups (n=309, 3.47%, American Indian or Alaska Native; n=171, 1.92%, Native Hawaiian or Other Pacific Islander; and n=106, 1.19%, Middle Eastern or North African). This diversity suggests that DPS services are effective in reaching historically underrepresented groups. Anonymous access and culturally neutral peer interaction may help reduce barriers to accessing support services.

XSL•FO
RenderX

Table 1. Distribution of users by age.[a]

User age (years)	Count, n
13-15	239
16-18	954
19-21	833
22-24	691
25-27	616
28-30	630
31-33	582
34-36	640
37-39	275
40-42	484
43-45	238
46-48	266
49-51	466
52-54	194
55-57	367
58-60	169
61-63	172
64-66	106
67-69	149
70-72	91
73-75	85
76-78	139

[a]Demographic data were obtained from users who voluntarily disclosed age information (8386/24,818, 33.79%).

EAP Inflows and Outflows

Tables 2 and 3 provide perspectives on the relationship between DPS and EAP utilization, reflecting users referred to DPS through their EAP programs (inflow) and those referred to EAP programs after using DPS (outflow). These sessions were initiated by employees with varying job roles from 5 large employers representing multiple industries, including retail, media, and health care. Table 2 (EAP inflows) presents the distribution of users who accessed the service through their organization's EAP, categorized by struggle topic. This table shows how employees are reaching DPS and aims to identify the most common concerns prompting employees with existing EAP benefits to use DPS services. Of 24,818 total user chat sessions, 5482 (22.09%) were initiated through users' EAP programs. Table 3 (EAP outflows) illustrates the distribution of users who were referred to their organization's EAP services through DPS sessions, also categorized by struggle topic. This reflects continuity of care from DPS chat sessions—based on moderator assessment or users expressing need—to additional employer-provided services. Of 24,818 user chat sessions, 2720 (10.96%) were referred to EAP services after DPS sessions.

Notably, the proportions of inflow and outflow vary across topics, suggesting that certain concerns may be more likely to drive employees to seek support through DPS, while others necessitate referrals to EAP resources. These insights can help refine targeted support strategies across both DPS and EAP services, ensuring that users receive effective care that bridges gaps in traditional care. For example, topics such as relationship issues, loneliness, stress, and trauma are among the most common concerns and may show higher referral rates, indicating opportunities to strengthen collaboration between DPS and EAPs for these specific issues.

XSL•FO
RenderX

Table 2. Employee users who accessed the Supportiv service through their organization's EAP[a] program referral.

Struggle topic	Count (N=24,818), n	EAP (N=5482), n (%)[b]
Anxiety	1098	363 (33.06)
Loneliness	1757	513 (29.20)
Physical health/medical	399	116 (29.07)
Suicide/self-harm	373	107 (28.69)
Trauma	2663	659 (24.75)
Relationship issues	7531	1840 (24.43)
Identity	638	140 (21.94)
Mental health conditions	723	157 (21.72)
Depression	1570	324 (20.64)
Social connection and belonging	1160	233 (20.09)
Stress	2074	387 (18.66)
Caregiving/parenting	617	87 (14.10)
Work/productivity/burnout	2650	356 (13.43)
Financial/Social Determinants of Health	1565	200 (12.78)

[a]EAP: employee assistance program.

[b]Percentages derived as follows: per struggle topic, (EAP n / Count n) × 100.

Table 3. Employee users who were referred back to EAP[a] resources after participating in a DPS[b] chat session.

Struggle topic	Count (N=24,818), n	EAP (N=2720), n (%)[c]
Anxiety	1098	207 (18.85)
Depression	1570	252 (16.05)
Caregiving/parenting	617	83 (13.45)
Mental health conditions	723	97 (13.42)
Suicide/self-harm	373	50 (13.40)
Trauma	2663	342 (12.84)
Loneliness	1757	210 (11.95)
Stress	2074	235 (11.33)
Social connection and belonging	1160	120 (10.34)
Relationship issues	7531	764 (10.14)
Physical health/medical	399	40 (10.03)
Identity	638	43 (6.74)
Financial/Social Determinants of Health	1565	105 (6.71)
Work/productivity/burnout	2650	172 (6.49)

[a]EAP: employee assistance program.

[b]DPS: digital peer support.

[c]Percentages derived as follows: per struggle topic, (EAP n / Count n) × 100.

Shared Resources

A key feature of the Supportiv service is its ability to share contextually relevant, tailored resources with users seeking emotional support, enhancing the user experience and the effectiveness of the DPS service on sentiment outcomes. In total, 33,632 resources were shared across 24,818 chat sessions, averaging 1.36 resources per session, indicating that at least one relevant resource was shared per session. Among these resources, 16,157 were articles, 8813 were videos, 5269 were employer-specific benefits, 2581 were worksheets, 810 were inspirational quotes, and 2 were audio files. Table 4 shows the distribution of resource types across various struggle topics, with an average of 2402 resources shared per topic.

XSL•FO
RenderX

Table 4. Distribution of resource types shared with users across various struggle topics.

Struggle topic	Blog (N=16,157), n	Video (N=8813), n	Employee specific (N=5269), n	Worksheet (N=2581), n	Quote (N=810), n	Audio (N=2), n	Count (N=33,632), n
Relationship issues	5338	3089	1314	983	332	2	11,058
Trauma	1866	956	590	382	74	0	3868
Work/productivity/burnout	1556	749	629	170	68	0	3172
Stress	1347	736	496	269	55	0	2903
Loneliness	1174	649	332	195	80	0	2430
Depression	974	516	365	151	47	0	2053
Financial/Social Determinants of Health	871	382	442	64	27	0	1786
Social connection and belonging	651	364	227	91	35	0	1368
Anxiety	753	440	265	99	35	0	1592
Mental health conditions	453	261	154	37	15	0	920
Identity	380	217	104	47	11	0	759
Caregiving/parenting	361	219	158	45	14	0	797
Physical health/medical	219	128	118	19	8	0	492
Suicide/self-harm	214	107	75	29	9	0	434

User Struggles

The 5 most common struggle topics expressed by users during business hours (total chats n=7995) were relationship issues (n=2249, 28.13%), workplace concerns (n=1029, 12.87%), financial or Social Determinants of Health (n=744, 9.31%), trauma (n=734, 9.18%), and stress (n=732, 9.16%). By contrast, outside business hours (total chats n=16,823), users most frequently reported relationship issues (n=5282, 31.40%), trauma (n=1929, 11.47%), workplace concerns (n=1621, 9.64%), stress (n=1342, 7.98%), and loneliness (n=1313, 7.80%). These patterns suggest that, while some concerns are consistently reported throughout the day, workplace and financial concerns are more prevalent during business hours, when work-related stressors affect users. Outside business hours, loneliness and trauma are more commonly reported, indicating that users are more open to engaging in personal discussions during nonworking hours.

User Activity

Several user activity metrics were analyzed for the 2 cohorts in the business-hour study. These included conversation duration (time from chat initiation to exit), typing activity (number of typed characters per minute as an indicator of engagement), total messages and characters sent, and the number of topics discussed during the chat.

As shown in Figure 2A, a significant difference (P<.001) was observed in conversation duration between user chat sessions. The mean conversation duration was 33.63 (SD 32.72) minutes for the “within-business-hours” cohort and 40.03 (SD 36.59) minutes for the “outside-business-hours” cohort, indicating that users in the “outside-business-hours” group stayed on the service 19% longer than those in the “within-business-hours” group.

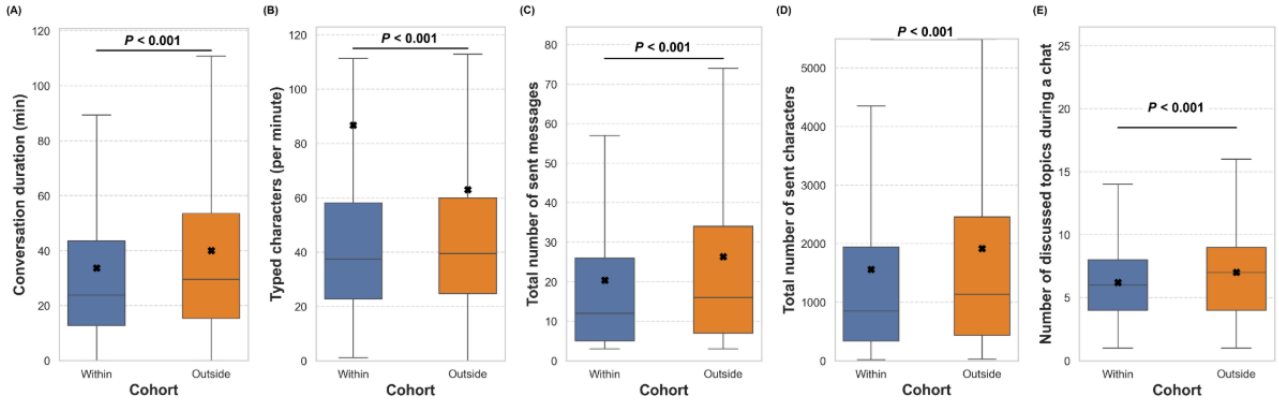
User participation, measured by typing activity, also differed significantly between groups (P<.001; Figure 2B). During business hours, users typed an average of 86.75 (SD 2316.61) characters per minute, compared with 63.00 (SD 563.27) characters per minute outside business hours. This suggests that users during business hours typed 38% faster.

Another key engagement metric is the number of messages and characters sent. Users in the “within-business-hours” group sent an average of 20.36 (SD 25.55) messages and 1559.61 (SD 2084.01) characters, whereas those in the “outside-business-hours” group sent 26.35 (SD 30.95) messages and 1917.05 (SD 2406.78) characters, representing increases of 29% and 23%, respectively. These differences were significant between the 2 groups (both P<.001; Figures 2C and 2D).

Users discussed an average of 6.21 (SD 3.28) topics in chat sessions during business hours, compared with 7.01 (SD 3.53) topics outside business hours, representing a 13% increase outside business hours (P<.001; Figure 2E). Taken together, these findings indicate differences in user engagement patterns between the 2 cohorts. Users who used DPS outside business hours remained in chats longer, exchanged more messages, and covered more topics, suggesting that they were more likely to engage at a deeper and more personal level when work-related stressors, distractions, or time constraints were not present.

XSL•FO
RenderX

Figure 2. User activity in the "within" versus the "Outside" business hours groups are separated based on (A) conversation duration, (B) characters sent per minute, (C) number of sent messages, (D) number of sent characters, and (E) number of discussed topics, by the user. The value denoted by "x" is the mean.



Sentiment Outcomes and Trends

Within-Session Sentiment Trajectories Across DPS Engagement

Changes in users' sentiment outcomes for loneliness, sadness, stress, anxiety, depression, despair, helplessness, and optimism were analyzed from the beginning to the end of each moderated, peer-to-peer live chat session to assess whether engaging with DPS was associated with improved sentiment scores. These trends provide an overview of the service's general sentiment support and serve as a reference point for interpreting more targeted results in subsequent sections.

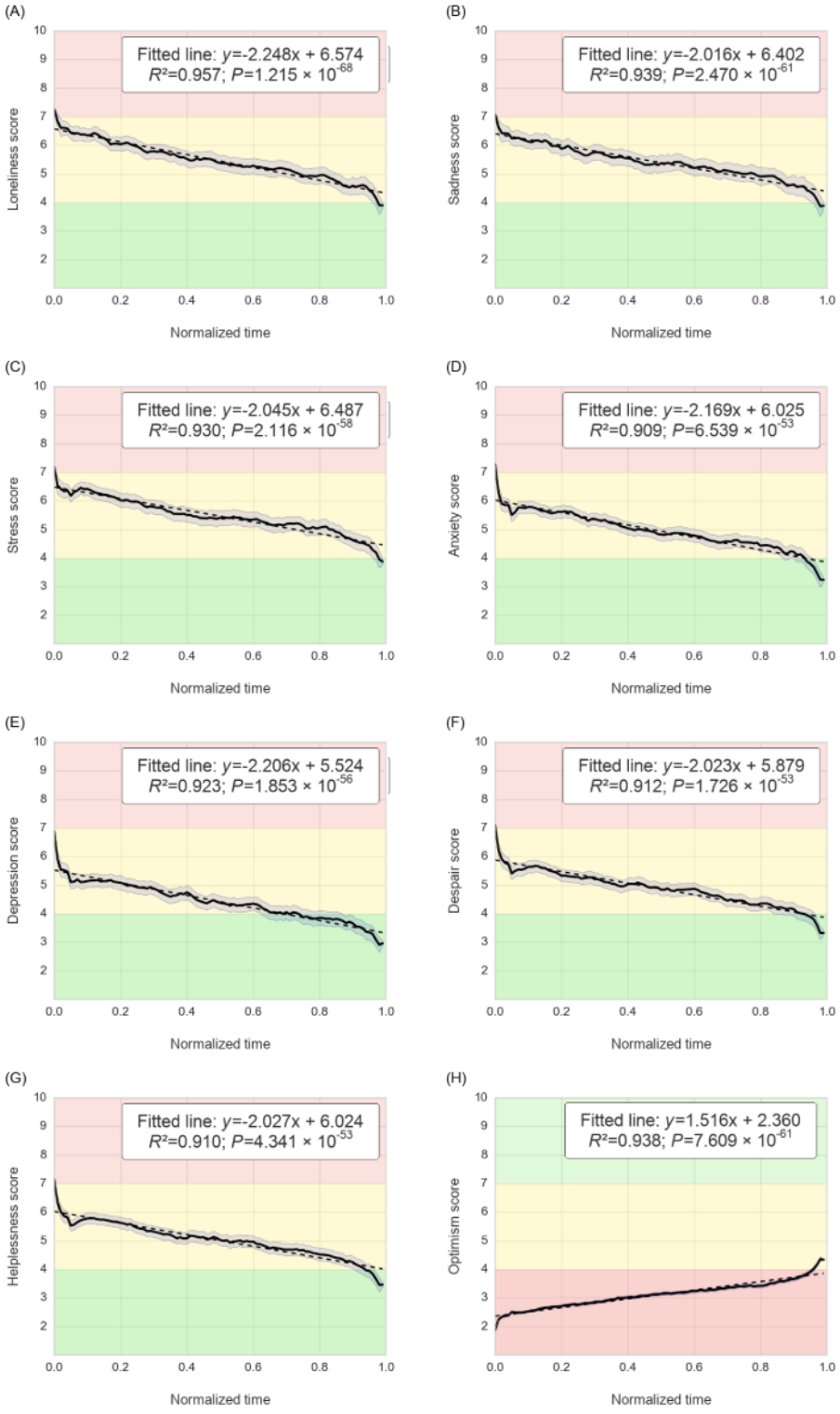
As shown in Figure 3, the 8 time-series graphs illustrate how users' sentiment states evolved throughout the course of the live chat sessions. Each graph displays one of the key dimensions (loneliness, sadness, stress, anxiety, depression, despair, helplessness, and optimism) with a fitted linear regression line, where the x-axis represents normalized session time and the y-axis represents the respective sentiment scores on a 10-point scale. These plots capture trends across all users, offering a holistic picture of Supportiv's influence on sentiment change, independent of contextual factors such as session timing or receipt of a resource.

The results reveal consistently improved sentiment states across all 8 dimensions. Notably, all negative sentiments (loneliness, sadness, stress, anxiety, depression, despair, and helplessness) declined steadily over the course of a session, while the positive sentiment of optimism increased. The negative sentiments showed strong, pronounced downward trends. As not all conversations included all sentiments, the number of conversations per sentiment analysis differed. The greatest changes were observed in loneliness, which was reduced by 46% (n=11,492; R^2=0.957; P<.001; Figure 3A); sadness, by 45% (n=10,741; R^2=0.939; P<.001; Figure 3B); and stress, by 46% (n=13,928; R^2=0.93; P<.001; Figure 3C). These were followed by anxiety, which was reduced by about 39% (n=9654; R^2=0.909; P<.001; Figure 3D); depression, by about 40% (n=5585; R^2=0.923; P<.001; Figure 3E); despair, by about 40% (n=6197; R^2=0.912; P<.001; Figure 3F); and helplessness, by about 38% (n=7742; R^2=0.910; P<.001; Figure 3G), highlighting the consistency of sentiment relief across users. The downward trends in negative sentiment scores suggest that users' expressed emotional language became less indicative of negative affect over DPS chats. By contrast, optimism increased by about 77% (Figure 3H). This sustained upward trend (n=23,491; R^2=0.938; P<.001) suggests that users not only experienced relief from distress during DPS conversations but also left sessions feeling more hopeful and supported. Notably, these changes occurred smoothly over time, with minimal noise or volatility, as shown by the narrow CIs in most graphs.

Taken together, these findings indicate that DPS use was consistently associated with improvements in within-session sentiment scores. The strength and stability of the observed trends confirm improved sentiments following DPS use. Overall, users began sessions with high-intensity negative affect and concluded within the neutral-to-positive range by the end of the sessions, illustrating that engagement with DPS is associated with favorable shifts in sentiment scores.

XSL•FO
RenderX

Figure 3. Sentiment outcomes trends throughout live chat peer-to-peer conversation progression. Sentiment changes were observed as conversations progressed. The solid black line shows the emotion score across the conversation, the dashed line represents the fitted line, and the shaded area represents the SEM. The fitted line information is provided in the box, which includes the slope (how fast the emotion is reduced per normalized conversation time), the intercept (the approximate initial emotion score), and the goodness of fit. (A–G) The green area (1-4) represents "low" levels of emotion, yellow (4-7) signifies "moderate" levels, and red (7-10) indicates "high" momentary emotion. (H) The red area represents low optimism (1-4), the yellow area represents neutral optimism (4-7), and the green area represents positive optimism (7-10).



Business Hour Effect

Consistent with previous research [7], user engagement revealed that nearly two-thirds of sessions occurred outside business hours, with 16,823 chats outside business hours and 7995 during local business hours (Figure 1). Table 5 presents the initial baseline sentiment states of users who accessed the service during business hours compared with those who did so outside business hours. Baseline scores were similar for most sentiments, although a few showed differences. Users who

XSL•FO
RenderX

accessed the DPS service during business hours reported slightly higher stress levels (P<.001), suggesting that work-related stressors may contribute to increased distress during business hours. Conversely, users who accessed DPS after business hours reported higher levels of loneliness (P=7.63 × 10^{-3}), possibly due to increased isolation during evening and nighttime hours. No significant differences were found in anxiety (P=.54), despair (P=.23), helplessness (P=.58), sadness (P=.26), or optimism (P=.15) scores. These results suggest that employees may seek support with different baseline sentiment states depending on the time of engagement.

Figure 4 compares sentiment score improvements between users who initiated chat sessions during local business hours (within) and those who did so outside business hours (outside) across 8 sentiment domains: anxiety, depression, despair, helplessness, loneliness, sadness, stress, and optimism. Users who utilized DPS outside business hours showed significantly greater reductions in depression (41.2% vs 38.9%; P=1.0 × 10^{-2}; d=0.0197), helplessness (34.7% vs 32.8%; P=2.8 × 10^{-2}; d=0.0058), and loneliness (32.2% vs 31.7%; P=2.5 × 10^{-2}; d=0.0166), suggesting that support outside work hours was associated with greater changes in sentiment scores, although the effect sizes were small. No statistically significant differences were observed for anxiety (P=.10), despair (P=.92), sadness (P=.45), stress (P=.62), or optimism (P=.10), and changes in sentiment scores were similar for users during and after business hours. These results also suggest that engagement during business hours may serve as an important source of immediate support and validation for employees, allowing them to process their thoughts and feelings while helping sustain concentration during the workday.

Table 5. Comparison of baseline sentiment scores between business hours and nonbusiness hours users.

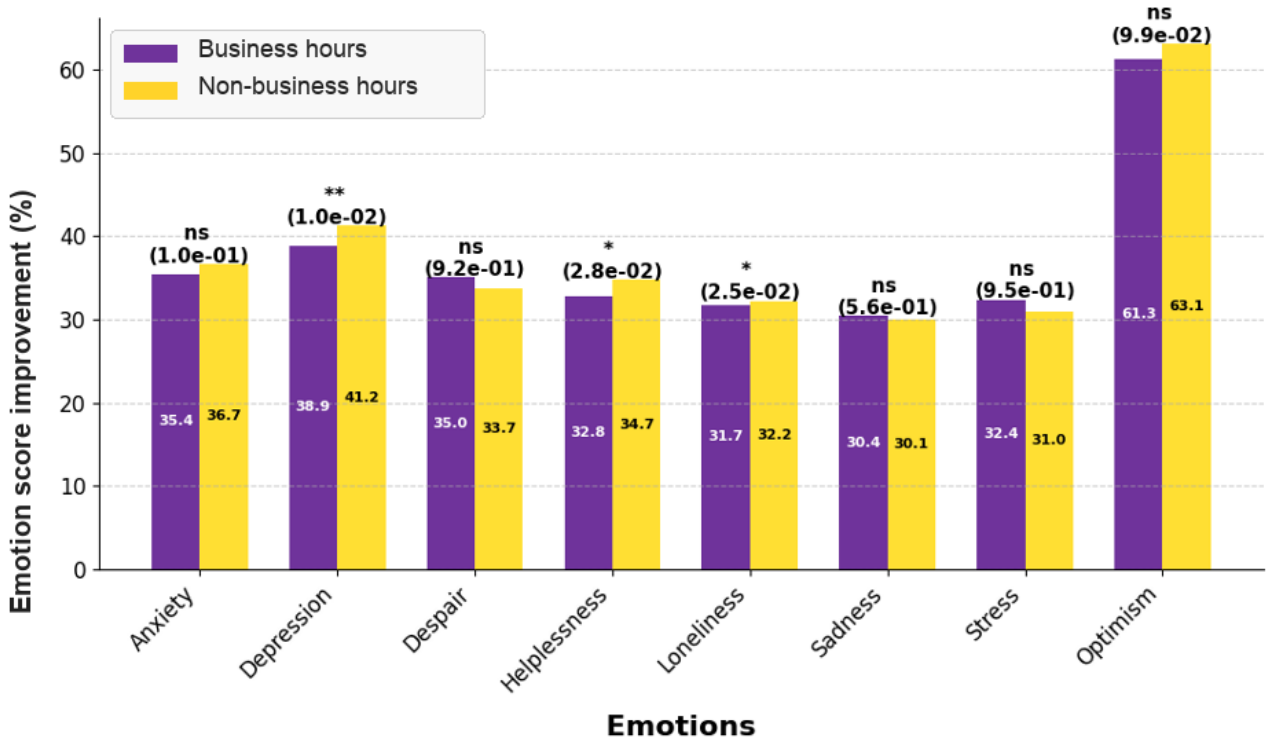
Factor	Anxiety	Depression	Despair	Helplessness	Loneliness	Sadness	Stress	Optimism
Within business hours, mean (SD)	6.24 (2.40)	5.17 (2.29)	5.82 (2.31)	6.17 (2.33)	6.66 (1.99)	6.62 (1.82)	6.96 (1.85)	3.12 (1.64)
Outside business hours, mean (SD)	6.25 (2.29)	5.26 (2.16)	5.88 (2.19)	6.17 (2.27)	6.74 (2.00)	6.58 (1.83)	6.85 (1.86)	3.10 (1.65)
P value	.54[a]	.04[b]	.23[a]	.58[a]	7.63 × 10^{-3}[c]	.28[a]	<.001	.15[a]

[a]Not significant.

[b]P<.05.

[c]P<.01.

Figure 4. Emotion score improvements by time-of-service use: users engaged during business hours experienced greater reductions in depression, helplessness, and loneliness. No significant differences were found for anxiety, despair, sadness, stress, or optimism. The improvement percentages were calculated based on the slope of the fitted trend line. P values are indicated in brackets. *P<.05. **P<.01.

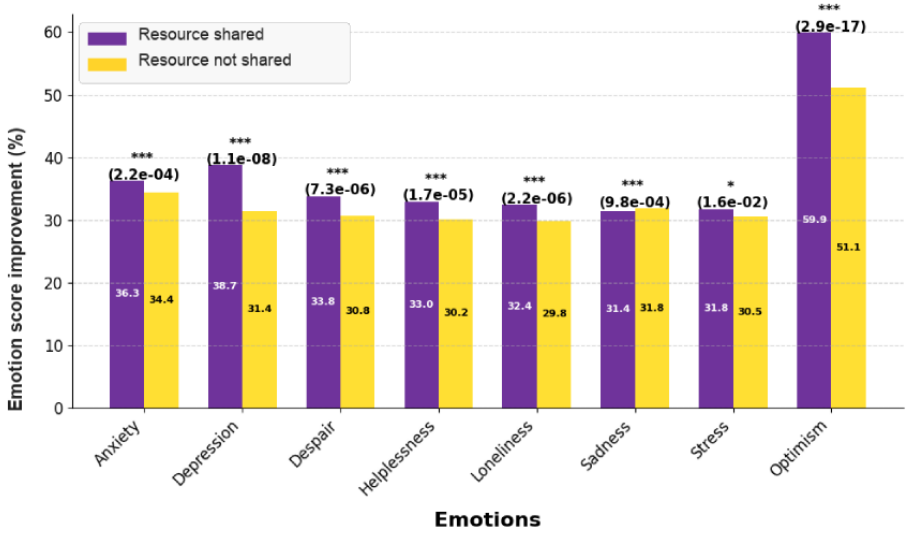


XSL•FO
RenderX

Resource Sharing Effect

Figure 5 compares sentiment score improvements between users who received at least one resource during their chat session (with) and those who did not (without) across 8 dimensions: anxiety, depression, despair, helplessness, loneliness, sadness, stress, and optimism. Users who received a resource showed significantly greater changes in almost all sentiment states: anxiety (36.3% vs 34.4%; P<.001; d=−0.0003), depression (38.7% vs 31.4%; P<.001; d=−0.0830), despair (33.8% vs 30.8%; P<.001; d=−0.0468), helplessness (33.0% vs 30.2%; P<.001; d=−0.0752), loneliness (32.7% vs 29.8%; P<.001; d=−0.0789), stress (31.8% vs 30.5%; $P=1.6\times10^{-2}$; d=−0.0342), and optimism (59.9% vs 51.1%; P<.001; d=−0.0145). Interestingly, users who did not receive a resource showed a slightly greater reduction in sadness (31.8% vs 31.4%; P<.001; d=−0.0287), although the difference was negligible. Overall, these findings highlight an association between resource sharing and improved sentiment scores, particularly in alleviating negative sentiment states and enhancing optimism. Therefore, resource provision serves as a supportive tool when incorporated into DPS services.

Figure 5. Comparison of emotion score improvements between users who received at least one resource and those who did not. Resource sharing was associated with significantly greater reductions in anxiety, depression, despair, helplessness, loneliness, and stress, and gains in optimism, while sadness slightly favored the nonreceiving group. The improvement percentages were calculated based on the slope of the fitted trend line. *P* values are indicated in brackets. *P<.05; ***P<.001.



Discussion

Principal Findings

This study offers novel insights into the role of timing, in-chat behaviors, and structural barriers in shaping user engagement and sentiment outcomes in a DPS service offered as a stand-alone service or integrated into EAPs. The findings advance understanding of when and how employees use DPS services, how these interactions guide sentiment trajectories, and which factors affect access to mental health support.

Barriers to EAP Utilization

Despite the potential of DPS to fill critical gaps in care, broader EAP engagement remains limited due to persistent systemic barriers. The most pressing among these is a widespread lack of awareness about the scope and availability of EAP services, leading to underutilization even in organizations with robust offerings [49]. Research shows that, without sustained promotional efforts, EAP usage rates remain below 10%, while programs with active education and engagement campaigns have reported utilization rates closer to 30% [50,51]. Additionally, internalized stigma may deter individuals from seeking help [52]. Employees may interpret help-seeking as a signal of personal failure or professional inadequacy, especially in high-performance cultures. National surveys indicate declining employee comfort with employer-sponsored mental health services, with only 52% of employees reporting comfort accessing EAP resources in 2022, down from 64% in 2021 and 67% in 2020 [53]. Addressing these gaps requires shifts at the organizational level, including visible leadership endorsement of mental health resources, normalization of mental health care, and integration of DPS into wellness initiatives. These strategies can help develop EAPs from crisis tools into preventive support systems and increase their perceived legitimacy [54].

Sociobehavioral Factors Influencing Use

Beyond timing, additional sociobehavioral factors also influenced engagement with DPS. Although many users identified as female and reported White ethnicity, as expected for mental health services [55], DPS attracted a larger proportion of male and Black, Indigenous, and People of Color users than is typically observed in such services. The age distribution also reflected a broad user base, with a median age of 36 years and a mean of approximately 45 years. The accessibility and anonymity of DPS may offer a solution for adults who are underserved by traditional mental health care. These patterns highlight the broader demographic reach of DPS compared with typical services. The flexible, anonymous nature of DPS mitigates certain barriers associated with traditional mental health care, such as time constraints, privacy concerns, and stigma. Users were more likely to engage outside work hours,

XSL•FO
RenderX

likely due to an increased sense of psychological safety for vulnerable self-disclosure when outside of work environments. This pattern supports theories of readiness to disclose, which suggest that individuals are more likely to seek help in environments they perceive as emotionally and socially safe [56,57]. Therefore, timing is not only an operational variable but also a sociobehavioral signal of trust, readiness, and psychological safety. Designing digital services that accommodate different levels of readiness would better enable users to engage based on their comfort and needs.

Discussion Topics

The most commonly discussed topic, by a clear margin, was relationship issues, which appeared in nearly one-third of all live chat sessions. However, differences in topic prevalence emerged when inflows from and outflows to EAP sources were examined. Most frequently, users accessed DPS services through their EAP (inflows) to discuss anxiety, loneliness, physical or mental health, and suicide or self-harm (Table 2). By contrast, users who sought DPS services and were subsequently referred to their EAP programs (outflows) most frequently discussed anxiety, depression, and caregiving or parenting (Table 3). Resource distribution across discussion topics showed that sessions centered on relationship issues received the greatest number of resources, reflecting the high volume of conversations on this topic (Table 4). Trauma, workplace concerns, and stress followed, aligning with their discussion frequency.

Sentiment Outcomes

Significant improvements in sentiment scores were observed across all investigated sentiment states. The trajectories for loneliness, sadness, stress, anxiety, depression, despair, helplessness, and optimism revealed consistent and statistically significant improvements across all 8 dimensions (Figure 3). Notably, optimism showed the most pronounced change (n=23,491, ~77%; R^2=0.938; P<.001), indicating that DPS use was associated with increased feelings of hopefulness and support. The greatest reductions in negative sentiment states were observed in loneliness (n=11,492, 46%; P<.001), stress (n=13,928, 46%; P<.001), and sadness (n=10,741, 45%; P<.001), suggesting that DPS use was associated with nearly halving users' initial distress.

Impact of Timing of Use

Consistent with previous research [7], user engagement revealed that nearly two-thirds of sessions occurred outside business hours. This high level of engagement after hours may reflect limited access to traditional in-person or typical tele-mental health support, as well as employees' greater openness to sharing personal struggles during periods of privacy or reduced work-related stressors.

Users discussed different topics during business hours compared with outside work hours. During business hours, the most common topics were relationship issues, workplace concerns, financial or social determinants of health, trauma, and stress, whereas outside work hours, relationship issues were discussed most frequently, followed by trauma, workplace concerns, stress, and loneliness. Although concerns such as relationship issues were prevalent throughout the day, these differences suggest that users tend to focus more on workplace concerns and social determinants of health during the business day, while outside work hours they are more likely to explore more personal issues such as trauma, stress, and loneliness.

Differences in user activity based on time of use also emerged. Users were more likely to engage for longer periods (Figure 2A), send a higher number of messages and typed characters (Figures 3C and 3D), and discuss more topics (Figure 2E) outside business hours. By contrast, users typed faster during work hours (Figure 2B), suggesting that time constraints related to work responsibilities may limit engagement. Overall, users were able to participate in deeper conversations outside work hours, indicating the need for services that are accessible beyond the typical 9-to-5 window.

Baseline sentiment states differed by time of engagement between the 2 cohorts (Table 5). Users who engaged during business hours reported significantly higher baseline stress, potentially due to situational demands (eg, work or other daytime stressors) [58,59]. By contrast, users who accessed DPS after business hours reported higher baseline levels of depression and loneliness, aligning with research suggesting increased social isolation during evening and nighttime hours, when there are fewer opportunities for social interaction [60,61]. Meanwhile, other sentiment outcomes did not differ by time of day, suggesting that they may be more stable and less influenced by situational factors [62].

These findings suggest 2 distinct sentiment profiles: engagement during business hours is characterized by acute, reactive distress, whereas after-hours engagement reflects more chronic vulnerability. Improvements in sentiment states support these profiles. Depression, helplessness, and loneliness were reduced at higher rates with DPS use outside work hours, suggesting deeper personal engagement (Figure 4). Comparable reductions in anxiety, despair, sadness, stress, and optimism across the 2 groups indicate that these dimensions may be less affected by timing [58-61]. Furthermore, DPS supported psychological relief independent of timing, reinforcing the need for around-the-clock accessibility.

Impact of In-Session Resource Sharing

Another significant factor influencing sentiment improvements was resource sharing within DPS sessions. Sessions that included at least one resource, such as coping strategies or other psychoeducational materials, were associated with greater positive sentiment changes in anxiety, depression, despair, helplessness, loneliness, stress, and optimism (Figure 5). This finding aligns with the peer support literature, which emphasizes the therapeutic value of shared knowledge and validation of distress [63]. Peer support has been described as "a system of giving and receiving help founded on key principles of respect, shared responsibility, and an agreement of what is helpful" [64]. In digital formats, where users may initially feel disconnected or skeptical, resource sharing may help deepen engagement, enhance perceived value, and reinforce feelings of self-efficacy.

Interestingly, users who did not receive a resource showed slightly greater improvements in sadness, although this difference was minimal. This pattern may suggest that

XSL•FO
RenderX

individuals with higher sadness were less likely to seek additional support materials, or that focusing on problem-solving interventions may temporarily heighten awareness of one's struggles. These nuances highlight the importance of tailoring support strategies to individuals, suggesting that resource sharing may be most impactful for those navigating more acute sentiment states, while others may benefit more from affirmation and increased social connection.

Limitations and Future Research

While the study provides valuable insights, several limitations should be noted. The observational design prevents causal inference regarding the relationship between session characteristics and sentiment outcomes. An important limitation is that the primary outcome—sentiment change—was based on LLM-derived sentiment scores rather than clinical psychometric instruments. Although prior work and post hoc analyses have shown alignment between LLM-derived sentiment scores and expert human ratings, this study did not examine the psychometric properties of the sentiment scoring approach used. Given the large dataset, human annotation and validation were not conducted. However, post hoc analyses on a subgroup of conversational transcripts confirmed alignment between the model and expert human interpretation. Therefore, the reported sentiment score changes are not diagnostic or absolute measures of emotional change, but rather model-estimated shifts in sentiment within DPS sessions. Despite their relevance in understanding how DPS may support users and potentially encourage engagement with additional clinical care, these results should not be interpreted as indicators of absolute sentiment severity and are not equivalent to self-reported or clinician-scored outcomes. They do not reflect psychiatric symptom severity.

Additionally, limited demographic data restrict analyses of how variables such as age, gender, cultural background, or job role may shape engagement patterns or session efficacy. Although more than half of the users who reported demographic information identified as female and White—a pattern common across digital mental health platforms—DPS attracted a higher proportion of male and Black, Indigenous, and People of Color users than is typically observed. However, the limited demographic data prevented detailed subgroup analyses of sentiment outcomes. Future research should adopt randomized controlled trials or mixed methods designs to identify causal mechanisms underlying DPS effectiveness, while also prioritizing more comprehensive demographic reporting to better capture the service's impact across diverse populations.

Demographic variables were analyzed for only 8647 out of 24,818 (34.84%) conversations with users who voluntarily disclosed their information. The optional disclosure of personal information introduced selection bias, limiting generalizability to a broader population. A large proportion of individuals may have preferred to withhold information due to privacy, stigma, or other vulnerability-related concerns, and individuals from certain demographic backgrounds may be more likely to face stigma or distrust the health care system, potentially skewing the findings. Therefore, these results may not fully represent the demographic distribution of the broader population; however, they provide valuable insight into user characteristics. The analyzed subset remains sizable and informative regarding the distribution of users in real-world digital mental health settings.

Longitudinal tracking could also reveal whether sentiment improvements persist over time and whether early engagement predicts long-term EAP utilization or mental health outcomes. Moreover, segmenting user data by demographic and occupational variables could clarify subgroup differences, such as among shift workers, caregivers, parents, or underrepresented racial and ethnic groups. Further investigation of digital literacy and privacy preferences as moderating factors could help guide the design of more inclusive digital support systems. Finally, further exploration of resource sharing is warranted. Examining resource types, relevance, and user feedback could help refine the resources provided. Such analyses could also inform best practices for moderator facilitation and content curation within DPS, optimizing its function as a bridge between nonclinical and clinical support systems.

Practical Implications

The implications of these findings for workplace mental health strategy are substantial. Organizations can benefit from prioritizing awareness-raising efforts for EAP offerings through multimodal, repeated communications, especially during employee onboarding and at moments of organizational stress [54]. Embedding DPS as a core EAP component could expand reach and reduce barriers to support. Leadership endorsement, testimonials, and peer ambassador programs may help reduce stigma and foster a culture of openness [54]. Moreover, training moderators in sentiment attunement and evidence-based resource sharing may enhance user engagement and, in turn, increase the perceived benefits of DPS in improving sentiment state changes [9]. Integrating DPS within EAPs is a scalable approach that could reduce stigma associated with mental health care and improve employee well-being. This approach aligns well with a modern workforce that requires immediate, accessible, and flexible support tools.

Conclusions

This study underscores the importance of 24/7 access to mental health support, especially for employees who experience emotional distress outside standard business hours. However, because the primary measures were based on LLM-derived sentiment scores, conclusions about effectiveness remain exploratory. Around-the-clock DPS can bridge gaps in care when traditional services are unavailable, helping prevent crisis escalation and demonstrating organizational commitment to employee well-being. To improve EAP utilization, organizations should invest in consistent awareness efforts and destigmatize help-seeking. Leadership endorsement, integration of mental health into everyday communications, and collaboration with employee resource groups can broaden reach and foster a culture of support.

In-session resource sharing demonstrates the value of a hybrid approach that combines AI recommendations with trained peer support moderators to deliver targeted, practical tools. Incorporating psychoeducational resources within sessions may

XSL•FO
RenderX

increase user engagement, emotional regulation, and self-efficacy. Furthermore, tailoring resources to users' expressed needs in real time influences their effectiveness.

Finally, effective mental health strategies must be inclusive and flexible, adapting to diverse schedules, privacy needs, and digital comfort levels. By aligning accessibility with flexibility, employers can offer effective digital mental health interventions to employees. Overall, this study contributes to the growing evidence positioning DPS as an effective, scalable, and equitable tool for mental health across a diverse user base.

Acknowledgments

This work was sponsored by Supportiv. The authors thank Alejandro Martinez of Stanford University for providing the original peer support training for Supportiv moderators and Dylan Dudek, MS, for assistance in completing this manuscript.

Funding

This research was funded internally by Supportiv. No external funding was received.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to their proprietary nature, but may be available from Supportiv upon reasonable request. Please contact Supportiv at info@supportiv.com.

Conflicts of Interest

All authors are current or former employees of Supportiv. The authors have no further conflicts of interest to disclose.

Multimedia Appendix 1
Examples provided to interpret emotion intensities in a few-shot manner.
[DOCX File , 17 KB - humanfactors_v13i1e90431_app1.docx]

References

1. Karmali M. Over half of American employees have used AI to take workplace training, according to new data. Moodle. URL: https://moodle.com/us/news/ai-for-workplace-training-in-america/ [accessed 2025-12-01]
2. Workplace stress. Occupational Safety and Health Administration (OSHA). URL: https://www.osha.gov/workplace-stress [accessed 2025-12-01]
3. Maulik P. Workplace stress: a neglected aspect of mental health wellbeing. Indian J Med Res 2017 Oct;146(4):441-444 [FREE Full text] [doi: 10.4103/ijmr.IJMR_1298_17] [Medline: 29434056]
4. Long T, Cooke FL. Advancing the field of employee assistance programs research and practice: a systematic review of quantitative studies and future research agenda. Human Resource Management Review 2023 Jun;33(2):100941. [doi: 10.1016/j.hrmr.2022.100941]
5. Joseph B, Walker A, Fuller-Tyszkiewicz M. Evaluating the effectiveness of employee assistance programmes: a systematic review. European Journal of Work and Organizational Psychology 2017 Sep 06;27(1):1-15. [doi: 10.1080/1359432x.2017.1374245]
6. Stringer H. Mental health care is in high demand. psychologists are leveraging tech and peers to meet the need. American Psychological Association. 2024 Jan. URL: https://www.apa.org/monitor/2024/01/trends-pathways-access-mental-health-care [accessed 2025-12-01]
7. Nagra H, Mines RA, Dana Z. Exploring the impact of digital peer support services on meeting unmet needs within an employee assistance program: retrospective cohort study. JMIR Hum Factors 2025 Feb 25;12:e68221 [FREE Full text] [doi: 10.2196/68221] [Medline: 39998863]
8. Dana Z, Nagra H, Kilby K. Role of synchronous, moderated, and anonymous peer support chats on reducing momentary loneliness in older adults: retrospective observational study. JMIR Form Res 2024 Oct 25;8:e59501 [FREE Full text] [doi: 10.2196/59501] [Medline: 39453688]
9. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]
10. Krukowski RA, Ross KM, Western MJ, Cooper R, Busse H, Forbes C, et al. Digital health interventions for all? Examining inclusivity across all stages of the digital health intervention research process. Trials 2024 Jan 30;25(1):98 [FREE Full text] [doi: 10.1186/s13063-024-07937-w] [Medline: 38291539]
11. Jagemann I, Stegemann M, von Brachel R, Hirschfeld G. Gender differences in preferences for mental health apps in the general population - a choice-based conjoint analysis from Germany. BMC Psychiatry 2024 Oct 14;24(1):682 [FREE Full text] [doi: 10.1186/s12888-024-06134-y] [Medline: 39402505]

XSL•FO
RenderX

12. Williams KDA, Wijaya C, Stamatis CA, Abbott G, Lattie EG. Insights into needs and preferences for mental health support on social media and through mobile apps among Black male university students: exploratory qualitative study. JMIR Form Res 2022 Aug 31;6(8):e38716 [FREE Full text] [doi: 10.2196/38716] [Medline: 36044261]
13. Employer's guide to diversity, equity, and inclusion strategy. Employers Council. 2021 Aug. URL: https://www.employerscouncil.org/wp-content/uploads/2021/08/EC-Guide-to-DEI.pdf [accessed 2025-12-01]
14. Kannisto KA, Korhonen J, Adams CE, Koivunen MH, Vahlberg T, Välimäki MA. Factors associated with dropout during recruitment and follow-up periods of a mHealth-based randomized controlled trial for Mobile.Net to encourage treatment adherence for people with serious mental health problems. J Med Internet Res 2017 Feb 21;19(2):e46 [FREE Full text] [doi: 10.2196/jmir.6417] [Medline: 28223262]
15. Crisp DA, Griffiths KM. Participating in online mental health interventions: who is most likely to sign up and why? Depress Res Treat 2014;2014:790457 [FREE Full text] [doi: 10.1155/2014/790457] [Medline: 24804089]
16. Kenny R, Dooley B, Fitzgerald A. Ecological momentary assessment of adolescent problems, coping efficacy, and mood states using a mobile phone app: an exploratory study. JMIR Ment Health 2016 Nov 29;3(4):e51 [FREE Full text] [doi: 10.2196/mental.6361] [Medline: 27899340]
17. Smail-Crevier R, Powers G, Noel C, Wang J. Health-related internet usage and design feature preference for e-mental health programs among men and women. J Med Internet Res 2019 Mar 18;21(3):e11224 [FREE Full text] [doi: 10.2196/11224] [Medline: 30882361]
18. Keyworth C, Alzahrani A, Pointon L, Hinsby K, Wainwright N, Moores L, et al. Barriers and enablers to accessing support services offered by staff wellbeing hubs: a qualitative study. Front Psychol 2022 Nov 15;13:1008913 [FREE Full text] [doi: 10.3389/fpsyg.2022.1008913] [Medline: 36457933]
19. Sanders CK, Scanlon E. The digital divide is a human rights issue: advancing social inclusion through social work advocacy. J Hum Rights Soc Work 2021 Mar 19;6(2):130-143 [FREE Full text] [doi: 10.1007/s41134-020-00147-9] [Medline: 33758780]
20. Fortuna KL, Venegas M, Umucu E, Mois G, Walker R, Brooks JM. The future of peer support in digital psychiatry: promise, progress, and opportunities. Curr Treat Options Psychiatry 2019 Sep;6(3):221-231 [FREE Full text] [doi: 10.1007/s40501-019-00179-7] [Medline: 33796435]
21. Fortuna KL, Solomon P, Rivera J. An update of peer support/peer provided services underlying processes, benefits, and critical ingredients. Psychiatr Q 2022 Jun;93(2):571-586 [FREE Full text] [doi: 10.1007/s11126-022-09971-w] [Medline: 35179660]
22. Young AS, Cohen AN, Niv N, Nowlin-Finch N, Oberman RS, Olmos-Ochoa TT, et al. Mobile phone and smartphone use by people with serious mental illness. Psychiatr Serv 2020 Mar 01;71(3):280-283 [FREE Full text] [doi: 10.1176/appi.ps.201900203] [Medline: 31744429]
23. Luther L, Buck BE, Fischer MA, Johnson-Kwochka AV, Coffin G, Salyers MP. Examining potential barriers to mHealth implementation and engagement in schizophrenia: phone ownership and symptom severity. J Technol Behav Sci 2022 Mar 31;7(1):13-22 [FREE Full text] [doi: 10.1007/s41347-020-00162-7] [Medline: 35434267]
24. Torous J, Wisniewski H, Liu G, Keshavan M. Mental health mobile phone app usage, concerns, and benefits among psychiatric outpatients: comparative survey study. JMIR Ment Health 2018 Nov 16;5(4):e11715 [FREE Full text] [doi: 10.2196/11715] [Medline: 30446484]
25. Fortuna KL, Naslund JA, LaCroix JM, Bianco CL, Brooks JM, Zisman-Ilani Y, et al. Digital peer support mental health interventions for people with a lived experience of a serious mental illness: systematic review. JMIR Ment Health 2020 Apr 03;7(4):e16460 [FREE Full text] [doi: 10.2196/16460] [Medline: 32243256]
26. Rayland A, Andrews J. From social network to peer support network: opportunities to explore mechanisms of online peer support for mental health. JMIR Ment Health 2023 Feb 28;10:e41855 [FREE Full text] [doi: 10.2196/41855] [Medline: 36853738]
27. Fallon P, Jaegers LA, Zhang Y, Dugan AG, Cherniack M, El Ghaziri M. Peer support programs to reduce organizational stress and trauma for public safety workers: a scoping review. Workplace Health Saf 2023 Nov 13;71(11):523-535. [doi: 10.1177/21650799231194623] [Medline: 37702192]
28. Yeo G, Fortuna KL, Lansford JE, Rudolph KD. The effects of digital peer support interventions on physical and mental health: a review and meta-analysis. Epidemiol Psychiatr Sci 2025 Feb 13;34:e9. [doi: 10.1017/s2045796024000854]
29. The importance of connection in the workplace. Circles. 2017 Dec 17. URL: https://www.circles.com/resources/social-connection-in-workplace-key-to-employee-engagement-and-wellbeing [accessed 2025-12-01]
30. Ganesh K. CultureMonkey. URL: https://www.culturemonkey.io/employee-engagement/diversity-in-the-workplace/ [accessed 2025-12-02]
31. Perslon L. The comprehensive guide to employee assistance programs. NPA Benefits. URL: https://npabenefits.com/employee-assistance-program/ [accessed 2025-12-01]
32. Mikolajczak-Degrauwe K, Slimmen SR, Gillissen D, de Bil P, Bosmans V, Keemink C, et al. Strengths, weaknesses, opportunities and threats of peer support among disadvantaged groups: a rapid scoping review. Int J Nurs Sci 2023 Oct;10(4):587-601 [FREE Full text] [doi: 10.1016/j.ijnss.2023.09.002] [Medline: 38020843]

XSL•FO
RenderX

33. Yeo G, Loo G, Oon M, Pang R, Ho D. A digital peer support platform to translate online peer support for emerging adult mental well-being: randomized controlled trial. JMIR Ment Health 2023 Apr 18;10:e43956 [FREE Full text] [doi: 10.2196/43956] [Medline: 36756843]
34. Yao Y, Zhang S, Li A. Effects of educational attainment and housing condition on self-rated health in old age: heterogeneity and tendency in China. Front Public Health 2021 Jan 6;9:774364 [FREE Full text] [doi: 10.3389/fpubh.2021.774364] [Medline: 35071161]
35. Yue D, Ponce NA, Needleman J, Ettner SL. The relationship between educational attainment and hospitalizations among middle-aged and older adults in the United States. SSM Popul Health 2021 Sep;15:100918 [FREE Full text] [doi: 10.1016/j.ssmph.2021.100918] [Medline: 34568538]
36. Lim C, Fuchs C, Torous J. Integrated digital mental health care: a vision for addressing population mental health needs. IJGM 2024 Feb;Volume 17:359-365. [doi: 10.2147/ijgm.s449474]
37. Bettis AH, Burke TA, Nesi J, Liu RT. Digital technologies for emotion-regulation assessment and intervention: a conceptual review. Clin Psychol Sci 2022 Jan;10(1):3-26 [FREE Full text] [doi: 10.1177/21677026211011982] [Medline: 35174006]
38. Social penetration theory. EBSCO Information Services, Inc. URL: https://www.ebsco.com/research-starters/communication-and-mass-media/social-penetration-theory [accessed 2025-12-02]
39. Blease C, Rodman A. Generative artificial intelligence in mental healthcare: an ethical evaluation. Curr Treat Options Psych 2024 Dec 09;12(1):5. [doi: 10.1007/s40501-024-00340-x]
40. Miller W, Rollnick S. Enhancing Motivation for Change in Substance Use Disorder Treatment: Updated 2019 (Treatment Improvement Protocol (TIP) Series, No. 35). Rockville (MD): Substance Abuse and Mental Health Services Administration (US); 2019.
41. Madaan A, Zhou S, Alon U, Yang Y, Neubig G. Language models of code are few-shot commonsense learners. 2022 Presented at: The 2022 Conference on Empirical Methods in Natural Language Processing; December 7-11, 2022; Abu Dhabi, United Arab Emirates p. 1384-1403.
42. Malgaroli M, Hull TD, Zech JM, Althoff T. Natural language processing for mental health interventions: a systematic review and research framework. Transl Psychiatry 2023 Oct 06;13(1):309 [FREE Full text] [doi: 10.1038/s41398-023-02592-2] [Medline: 37798296]
43. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. Psychol Med 2019 Jul;49(9):1426-1448. [doi: 10.1017/S0033291719000151] [Medline: 30744717]
44. Zhang T, Schoene AM, Ji S, Ananiadou S. Natural language processing applied to mental illness detection: a narrative review. NPJ Digit Med 2022 Apr 08;5(1):46 [FREE Full text] [doi: 10.1038/s41746-022-00589-7] [Medline: 35396451]
45. The magic number: how to optimize and improve your survey response rate. Kantar. URL: https://www.kantar.com/inspiration/research-services/what-is-a-good-survey-response-rate-pf [accessed 2025-12-02]
46. Katona C, Livingston G. Impact of screening old people with physical illness for depression? The Lancet 2000 Jul;356(9224):91-92. [doi: 10.1016/s0140-6736(00)02466-1]
47. Fortuna KL, Divatia S, Neupane S, Geiger P, Bohm A. “As soon as I start trusting human beings, they disappoint me, and now I am going to get on an app that someone could hack. I really do not want to take that chance”: barriers and facilitators to digital peer support implementation into community mental health centers. Front Digit Health 2023 Jul 20;5:1130095 [FREE Full text] [doi: 10.3389/fdgth.2023.1130095] [Medline: 37547620]
48. Terlizzi E, Zablotsky B. Mental health treatment among adults: United States, 2019 (data brief no 380). National Center for Health Statistics. Hyattsville, MD: National Center for Health Statistics; 2020. URL: https://www.cdc.gov/nchs/products/databriefs/db380.htm [accessed 2026-03-30]
49. Croft J, Parks A, Whillans A. Why workplace well-being programs don’t achieve better outcomes. Harvard Business Review. 2024 Oct 18. URL: https://hbr.org/2024/10/why-workplace-well-being-programs-dont-achieve-better-outcomes [accessed 2025-12-02]
50. Attridge M. The current state of employee assistance programs in the United States: a research-based commentary. IJSRP 2023 Aug 24;13(8):74-91. [doi: 10.29322/IJSRP.13.08.2023.p14010]
51. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate - service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]
52. Ahad A, Sanchez-Gonzalez M, Junquera P. Understanding and addressing mental health stigma across cultures for improving psychiatric care: a narrative review. Cureus 2023 May 26;15(5):e39549 [FREE Full text] [doi: 10.7759/cureus.39549] [Medline: 37250612]
53. Employees say workplaces are offering fewer mental health services in 2022, according to APA Poll May 22, 2022. American Psychiatric Association. URL: https://www.psychiatry.org/news-room/news-releases/apa-poll-employee-mental-health [accessed 2025-12-02]
54. Stigma, prejudice and discrimination against people with mental illness. American Psychiatric Association. URL: https://www.psychiatry.org/patients-families/stigma-and-discrimination [accessed 2025-12-02]
55. Terlizzi E, Zablotsky B. Symptoms of anxiety and depression among adults: United States, 2019 and 2022. In: National Health Statistics Reports. Hyattsville MD: National Center for Health Statistics; Sep 2020.

XSL•FO
RenderX

56. Zhang C, Cui C, Yao Q. “I” am willing to disclose, but “we” are unwilling: the impact of self-construal on individuals’ willingness to disclose. PRBM 2021 Dec;Volume 14:1929-1945. [doi: 10.2147/prbm.s336223]
57. Rickwood DJ, Mazzer KR, Telford NR. Social influences on seeking help from mental health services, in-person and online, during adolescence and young adulthood. BMC Psychiatry 2015 Mar 7;15(1):40. [doi: 10.1186/s12888-015-0429-6]
58. Sonnentag S, Fritz C. Recovery from job stress: the stressor-detachment model as an integrative framework. J Organiz Behav 2014 Apr 11;36(S1):S72-S103. [doi: 10.1002/job.1924]
59. Jin T, Hong S, Liu J, Liu L, Jia S, Shen G. Emotion regulation strategy moderates the impact of emotional labor on burnout in psychiatric nurses: a cross-sectional study. BMC Nurs 2025 Jul 01;24(1):687 [FREE Full text] [doi: 10.1186/s12912-025-03487-x] [Medline: 40597196]
60. van Roekel E, Scholte RHJ, Engels RCME, Goossens L, Verhagen M. Loneliness in the daily lives of adolescents. The Journal of Early Adolescence 2014 Sep 24;35(7):905-930. [doi: 10.1177/0272431614547049]
61. Vogel DHV, Krämer K, Schoofs T, Kupke C, Vogeley K. Disturbed experience of time in depression-evidence from content analysis. Front Hum Neurosci 2018 Feb 20;12:66 [FREE Full text] [doi: 10.3389/fnhum.2018.00066] [Medline: 29515385]
62. Huffman JC, Legler S, Millstein RA, Gomez-Bernal F, Celano CM, Chung W, et al. Does timeframe adjustment of the Life Orientation Test-Revised assess optimism as a state?: Data from the PEACE-III trial in patients with heart disease. J Posit Psychol 2019 Feb 12;14(6):799-806 [FREE Full text] [doi: 10.1080/17439760.2019.1579355] [Medline: 32983245]
63. Shalaby RAH, Agyapong VIO. Peer support in mental health: literature review. JMIR Ment Health 2020 Jun 09;7(6):e15572 [FREE Full text] [doi: 10.2196/15572] [Medline: 32357127]
64. Mead S, Hilton D, Curtis L. Peer support: a theoretical perspective. Psychiatr Rehabil J 2001;25(2):134-141. [doi: 10.1037/h0095032] [Medline: 11769979]

Abbreviations

AI: artificial intelligence
CCPA: California Consumer Privacy Act
DPS: digital peer support
EAP: employee assistance program
GAD-7: 7-item Generalized Anxiety Disorder
GDPR: General Data Protection Regulation
HIPAA: Health Insurance Portability and Accountability Act
LLM: large language model
NLP: natural language processing
OARS: open-ended questions, affirmations, reflective listening, and summaries
PHQ-9: 9-item Patient Health Questionnaire
PSS-10: 10-item Perceived Stress Scale

Edited by J Sarvestan; submitted 27.Dec.2025; peer-reviewed by G Cameron; comments to author 27.Jan.2026; revised version received 12.Mar.2026; accepted 13.Mar.2026; published 07.Apr.2026.

Please cite as:
Sedaghati F, Stetsenko A, Ozsan McMillan I, Nagra H, Dana Z
Use of Digital Peer Support for Employee Well-Being: Retrospective Analysis Across Five Large Employers
JMIR Hum Factors 2026;13:e90431
URL: https://humanfactors.jmir.org/2026/1/e90431
doi:10.2196/90431
PMID:

©Farbod Sedaghati, Anya Stetsenko, Ilayda Ozsan McMillan, Harpreet Nagra, Zara Dana. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 07.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Risk Factors for Noninitiation and Dropout in Blended Therapy in Inpatient Psychiatric Patients: Retrospective Cohort Study

Nikita Gupta[1], Dr med; Fabian Gardin[1], Dr med; Thomas Berger[2], Prof Dr; Wolfram Kawohl[1,3,4], Prof Dr Med

[1]Clienia AG, Oetwil am See, Switzerland
[2]University of Bern, Bern, Switzerland
[3]University of Zurich, Zurich, Switzerland
[4]University of Nicosia, Nicosia, Cyprus

Corresponding Author:
Nikita Gupta, Dr med
Clienia AG
Schlösslistrasse 8
Oetwil am See, 8618
Switzerland
Phone: 41 789106034
Email: gupta.n44@gmail.com

Related Article:

This is a corrected version. See correction statement: https://humanfactors.jmir.org/2026/1/e111245/

Abstract

Background: Blended therapy (BT) combines digital applications with face-to-face treatment and has become an increasingly important component of psychiatric care. Evidence indicates that BT can achieve outcomes comparable to or even superior to those of traditional face-to-face therapy. Despite certain advantages, routine implementation of BT remains challenging, and clinical practice suggests that while some inpatients engage with BT, many either discontinue early or do not initiate its use at all. To better understand these patterns, this multicentric, retrospective observational study investigates factors associated with noninitiation and dropout among inpatients who are offered BT.

Objective: In this study, data from 278 inpatients were analyzed to examine the influence of sociodemographic variables, comorbidities, and symptom severity on the uptake and continued use of BT. The objective was to identify predictors of noninitiation and dropout.

Methods: Multivariable logistic regression models were conducted to identify significant predictors of noninitiation and dropout among inpatients using the transdiagnostic, cognitive behavioral therapy–based electronic mental health platform Minddistrict, which offers modules targeting psychoeducation, cognitive restructuring, and behavioral activation. Data were collected from 2 psychiatric hospitals between January 2020 and May 2024. The sample consisted predominantly of patients diagnosed with depression (182/278, 65.7%) and posttraumatic stress disorder (61/278, 21.9%), alongside various comorbid conditions.

Results: The findings indicate distinct patterns of association for noninitiation and dropout. Of the 278 patients, only 5 (1.8%) completed all the assigned modules, and one-third of the patients never initiated the platform at all. Specifically, increasing age was linked to a lower risk of noninitiation (odds ratio [per year age difference] 0.98, 95% CI 0.96-1.00; *P*=.01), while the presence of a comorbid anxiety disorder was associated with a reduced risk of dropout (odds ratio 0.23, 95% CI 0.08-0.66; *P*=.007). Several variables showed no association with either noninitiation or dropout across all analyses, including sex, overall symptom severity, and certain comorbidities such as personality disorders and depression.

Conclusions: In this preselected inpatient sample, uptake of BT was very limited. Older age was associated with lower noninitiation, and comorbid anxiety disorders were associated with a lower likelihood of dropout. These findings may help inform future prospective studies on how BT can be introduced and supported more effectively in inpatient psychiatric care. As access to BT was granted selectively by therapists, the results should be interpreted as predictors of engagement within a selected sample rather than general predictors of BT uptake among all psychiatric inpatients.

XSL•FO
RenderX

(JMIR Hum Factors 2026;13:e85816) doi:10.2196/85816

KEYWORDS

blended therapy; noninitiation; dropout; inpatients; psychiatric; predictors

Introduction

Blended therapy (BT) is defined as the combination of digital applications, such as computer-based or internet-based programs, with face-to-face therapy. This combination of therapy modalities can be implemented in various forms. A substantial body of literature indicates that BT can achieve outcomes comparable or even superior to those of traditional standalone face-to-face therapy [1,2].

Despite its numerous advantages, implementing BT in routine practice remains challenging. This is due to various factors, including the high resource demands associated with the use of BT [3,4].

Online interventions are often characterized by high dropout and low adherence rates [5,6], highlighting the need to systematically investigate the underlying mechanisms behind it, as adherence can be a relevant factor for therapeutic success in prior research on treatment outcomes [7,8]. Understanding these factors is essential for building a comprehensive knowledge of how to optimize and implement this BT effectively. Evidence suggests that dropout is not a random phenomenon but is influenced by specific factors, particularly sociodemographic characteristics and symptom severity [9,10]. For example, Wu et al [10] examined predictors of noninitiation and dropout in blended cognitive behavioral therapy and found that both sociodemographic and clinical variables, such as depression severity and early digital engagement, significantly influenced adherence. Similarly, a meta-analysis by Karyotaki et al [9] and earlier findings by Christensen et al [11] confirmed that younger age, limited knowledge about psychological treatments, and greater symptom severity are associated with lower adherence in internet-based interventions. It should be noted that there is limited scientific literature specifically addressing BT, as most of the existing literature focuses on the area of unguided and guided self-help. Nonetheless, it is reasonable to assume that the use of BT also involves diverse patient-specific and diagnosis-specific influencing factors.

In the real-life implementation of BT at the psychiatric hospital *Clienia Schlössli AG*, many patients were offered blended treatment, but only a few actually started and completed the modules. This showed a relevant gap between initial setup and active participation in the treatment. This discrepancy is noteworthy, especially because a prior qualitative study conducted at the same hospital suggested that patients who engaged with BT were generally satisfied with its implementation [4]. This may indicate that acceptability among users does not necessarily translate into broad uptake or sustained use in routine inpatient care. Identifying risk factors for a low uptake of BT and/or low adherence could be therefore crucial to optimize the selection of suitable patients for this therapeutic approach. Adherence, defined as the completion of assigned modules on the platform, is associated with better treatment outcomes for depression [12]. In contrast, metrics such as the number of log ins or the total time spent on the platform have shown no correlation with improved outcomes [13]. Therefore, this study emphasizes module completion as a key indicator of engagement and its potential association with improved therapeutic outcomes. A distinctive feature of this study is its examination of therapy noninitiation, a phenomenon of clinical relevance that is exemplified by the case discussed earlier. While most studies on adherence primarily address dropout, referring to participants who begin but do not complete therapy, patients who never appropriately start therapy are often excluded from analysis. This exclusion may result in the loss of valuable information about a potentially important subgroup of patients.

A more targeted and personalized application of BT in selected patient groups may promote more efficient use of clinical resources. This is particularly relevant given that previous studies have reported that the implementation of BT can require additional time, effort, and resources [3,4,14]. Therefore, a part of optimizing resources should be the careful selection of patients who can truly benefit from its application.

The population examined in this study consists of patients receiving inpatient treatment. This is particularly relevant because the evidence base for BT in inpatient settings remains limited. Although only a few studies have examined its use in this context, the implementation of BT in inpatient care represents a valuable extension of this therapeutic approach.

The primary objective is to examine potential links between the noninitiation and dropout of BT and various influencing factors, building on findings from previous studies investigating similar phenomena.

Methods

BT Intervention

The study is a multicentric, retrospective observational study. We analyzed data from patients who used the electronic mental health platform *Minddistrict* at 2 psychiatric hospitals during their inpatient treatment. *Minddistrict* is a platform founded in the Netherlands in 2008, and it provides a transdiagnostic catalog structured into modules, self-help training, and diaries, drawing on cognitive behavioral therapy principles. The platform's modules primarily target psychoeducation, cognitive restructuring, and behavioral activation. It allows for individualized treatment by assigning suitable interventions tailored to each patient's needs. Patients can monitor behavioral patterns and track therapy progress through diaries or independently select interventions from the self-help catalog. *Minddistrict* additionally offers various communication channels, including video calls and chat functions. Access to *Minddistrict* was granted to selected and interested patients by their therapists, who created user accounts on their behalf. After the account creation, patients could engage with the platform

XSL•FO
RenderX

through therapist-guided modules, which required therapists to unlock new content after certain steps were completed. Additionally, patients had the option to use diaries, where they could document and reflect on their experiences independently and without therapist involvement. *Minddistrict* additionally offers various communication channels, including video calls and chat functions. The use of the video calls and diary features was excluded from this study because these components do not involve structured BT interventions.

Inclusion and Exclusion Criteria

The analysis includes inpatients from 2 psychiatric hospitals: *Clienia Schlössli* and *Clienia Littenheid*, which accessed the electronic mental health platform between January 2020 and May 2024. Platform use was reassessed in August 2024 to capture a comprehensive overview of module engagement and log in frequency. Although it is possible that some patients completed additional modules or increased use after this cutoff, such cases are considered unlikely. Patients were excluded if no modules had been assigned by their therapists. Patient accounts with multiple log ins and without module assignments were excluded, as we assumed that the platform was only used for video calls rather than for structured BT in these cases.

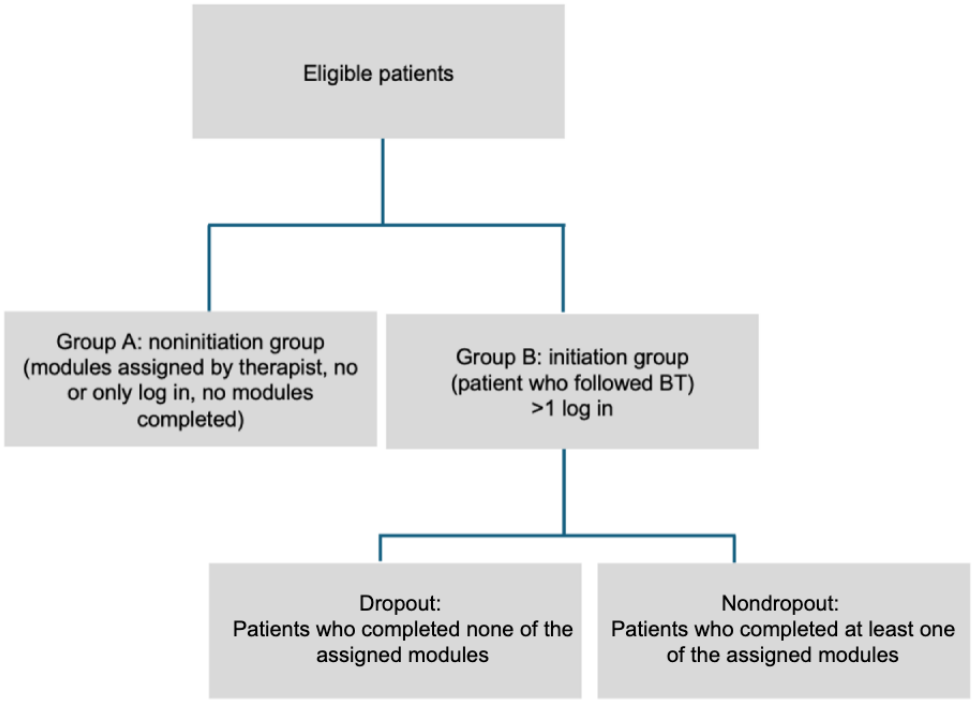
Division of Patient Groups

As outlined in the Introduction section, this study investigates noninitiation, defined as the failure to begin the intervention, and dropout, defined as the failure to complete any of the assigned therapeutic modules after platform initiation. To enable this analysis, patients were categorized into the following groups:

- Group A: patients who never logged in or only logged in once while not completing any modules
- Group B: patients who logged in once and completed one or more modules
- Dropout: patients who completed none of their assigned modules
- Nondropout: patients who completed at least one of their assigned modules

A schematic overview of the patient categorization process is presented in Figure 1.

Figure 1. Flowchart of patient classification into noninitiation and initiation groups. BT: blended therapy.



Outcome Measures

Overview

The data sources for this study included electronic medical records containing routinely collected datasets at the time of patient admission. Additionally, data on user activity on the platform are obtained directly from the *Minddistrict*'s database. An overview of all collected data is presented in Textbox 1.

XSL•FO
RenderX

Textbox 1. Overview of the collected data.

Sociodemographic characteristics

- Age
- Gender
- Marital status
- Educational level
- Employment status
- Setting (inpatient vs outpatient)

Assessments for symptom severity

- Brief Symptom Checklist
- Health of the Nation Outcome Scales

Diagnosis

- Main diagnosis
- Comorbid diagnoses

Platform use

- Number of log ins to the platform
- Number of modules assigned by the therapist
- Number of modules completed by the patient

Assessment of Symptom Severity

Brief Symptom Checklist

The Brief Symptom Inventory is a standardized self-report questionnaire used as a psychometric instrument to assess psychological distress. It consists of 53 items rated on a 5-point Likert scale and covers 9 symptom dimensions, including somatization, obsessiveness, interpersonal insecurity, depressiveness, anxiety, aggressiveness or hostility, phobic anxiety, paranoid thinking, and psychoticism. The Brief Symptom Checklist (BSCL) is the licensed German adaptation of the Brief Symptom Inventory. In the clinical setting, assessment is based on the total score of the BSCL, which was documented at admission and discharge. This allows the evaluation of symptom change over the course of inpatient treatment. Internal consistency (Cronbach α) ranged from satisfactory to very good across subscales, with the highest values for depressiveness and the Global Severity Index. Test-retest reliability after 1 week has been satisfactory to good across most dimensions [15].

Health of the Nation Outcome Scales

The Health of the Nation Outcome Scales (HoNOS) is a clinician-rated scale used to assess 12 items covering behavior, impairment, symptoms, and social functioning, including work and daily life. Each item is rated from 0 to 4, with a maximum score of 48, with higher scores indicating higher problem burden. The scale is routinely completed at hospital admission and discharge to track symptom changes. HoNOS is considered a practical tool for assessing the severity of mental health problems in inpatient settings. Existing literature supports its good interrater and test-retest reliability [16].

Ethical Considerations

This study (project ID: 2024-00768) has been reviewed and approved by the Zurich Ethics Committee. This research was conducted as a retrospective cohort study; therefore, active informed consent was not required by the ethics committee. Patient data were pseudonymized prior to analysis. Identifying information was stored separately from the analytic dataset and accessible only to authorized study personnel.

Data Analysis

Descriptive statistics include means and SDs for continuous variables, as well as numbers and percentages for categorical variables. Multivariable logistic regression models were used to identify factors associated with the extent of platform tool use based on a predefined set of variables, including age, sex, education, employment, civil status, disease severity (HoNOS or BSCL), treatment setting, and main diagnosis. Missing values were handled using multiple imputation by chained equations with the *mice* package in R [17]. Appropriate imputation methods were applied depending on the variable type, including logistic, polytomous, proportional odds regressions, and predictive mean matching. A total of 70 datasets were imputed for the whole dataset with 30 iterations, and 45 datasets were imputed for the stationary setting subset with 15 iterations. All analyses were carried out with R (version 4.4.3; R Foundation for Statistical Computing), and RMarkdown (Posit Software) was used for dynamic reporting.

XSL•FO
RenderX

Results

Descriptive Analysis

Distribution of Patient Characteristics

Descriptive statistics for the full inpatient sample (N=278) are presented in Table 1. The mean age was 45.45 (SD 15.14) years, and the sample consisted of 112 (40.3%) patients who were male. Regarding civil status, most patients were either single (96/236, 40.7%) or married or cohabiting (92/236, 38.9%). In terms of educational background, the largest proportion had completed an apprenticeship or vocational school (114/263, 43.3%), followed by university or higher education (54/263, 20.5%). Employment status varied considerably, with 94 (34.4%) out of 273 patients unemployed, 67 (24.5%) patients employed full-time, and 60 (21.9%) patients receiving disability or retirement benefits.

Most patients were diagnosed with depression (182/278, 65.7%) or posttraumatic stress disorder (61/278, 21.9%). Comorbidity was common, with a mean of 2.56 (SD 1.34) diagnoses per patient. The most frequent comorbid conditions were posttraumatic stress disorder (82/278, 29.6%), substance use disorder (65/278, 23.5%), and personality disorder (55/278, 19.9%).

Disease severity decreased substantially over the course of inpatient treatment. Mean HoNOS scores improved from 18.04 (SD 4.54) at admission to 9.22 (SD 4.42) at discharge, and BSCL scores decreased from 78.36 (SD 33.57) to 42.79 (SD 31.64).

XSL•FO
RenderX

Table 1. Descriptive characteristics of the patient sample (N=278).

Characteristics	Values	Missing
Sex (male), n (%)	112 (40.3)	0 (0)
Age (years), mean (SD)	45.45 (15.14)	0 (0)
Civil status, n (%)		42 (15.1)
Single	96 (40.7)	
In a registered partnership	1 (0.4)	
Married or cohabiting	92 (39)	
Divorced	44 (18.6)	
Widowed	3 (1.3)	
Educational level, n (%)		15 (5.4)
Not completed school education	4 (1.5)	
Compulsory school education	41 (15.6)	
Apprenticeship or full-time vocational school	114 (43.3)	
Higher vocational school or technical college	46 (17.5)	
Secondary school (Matura level)	4 (1.5)	
University or higher education institution	54 (20.5)	
Employment status, n (%)		5 (1.8)
Full-time employed	67 (24.5)	
Part-time employed	52 (19)	
Unemployed or not employed	94 (34.4)	
Disability insurance (IV)[a], old-age pension (AHV)[b], or another pension form	60 (22)	
Primary diagnosis, n (%)		1 (0.4)
Depression	182 (65.7)	
Bipolar disorders	9 (3.2)	
Personality disorder	8 (2.9)	
Pain-related disorder and somatoform disorder	1 (0.4)	
Posttraumatic stress disorder	61 (22)	
Anxiety or fear-related disorders	4 (1.4)	
Eating disorder	2 (0.7)	
Adaptation disorder	6 (2.2)	
Obsessive-compulsive disorder	0 (0)	
Schizophrenia or other primary psychotic disorders	1 (0.4)	
Disorders due to substance use	2 (0.7)	
Other	1 (0.4)	
Comorbidity		
Number of diagnoses, mean (SD)	2.56 (1.34)	0 (0)
Depression, n (%)	250 (90.3)	1 (0.4)
Personality disorder, n (%)	55 (19.9)	1 (0.4)
Substance use disorder, n (%)	65 (23.5)	1 (0.4)
Posttraumatic stress disorder, n (%)	82 (29.6)	1 (0.4)
Anxiety disorder, n (%)	36 (13)	1 (0.4)

[a]IV: Invalidenversicherung (disability insurance).

[b]AHV: Alters- und Hinterlassenenversicherung (national old-age pension scheme).

XSL•FO
RenderX

Distribution of Platform Use Parameters

On average, patients logged into the system 5.9 (SD 14.4; median 2.0, IQR 1.0-5.8) times. The mean number of modules assigned was 3.8 (SD 3.4; median 3.0, IQR 1.0-5.0), while the mean number of completed modules was 0.9 (SD 2.0; median 0.0, IQR 0.0-1.0), corresponding to a mean module completion rate of 24% (0.9/3.8). Notably, only 5 (1.8%) out of 278 patients completed all assigned modules. In terms of group distribution, 94 (33.8%) patients were in group A (noninitiation group), and 184 (66.2%) patients initiated the platform. Among those who initiated the platform, 94 (51.1%) completed none of their assigned modules (dropout), while 90 (48.9%) completed at least one assigned module (nondropout).

Regression Models With Imputed Data

Logistic Regression Model on Factors Potentially Associated With Noninitiation

In the logistic regression model assessing factors associated with noninitiation of the online platform among inpatients (N=278), age was the only variable significantly associated with noninitiation. Specifically, increasing age was linked to a lower likelihood of noninitiation (odds ratio [OR]=0.98, 95% CI 0.96-1.00; *P*=.01). While the OR is presented as 1, the significant *P* value indicates an actual effect, with the actual OR likely being marginally below 1 but not captured due to rounding. The odds decreased by 2% for each 1-year increase in age, suggesting that older patients were more likely to engage with the platform. For an age difference of 10 years, this would indicate an OR of 0.8. All other variables were not significantly associated with noninitiation (Table 2).

Table 2. Logistic regression model of factors potentially associated with noninitiation.

Parameter	Odds ratio (95% CI)	*P* value
Sex (male)	1.05 (0.60-1.84)	.86
Age (difference per year)	0.98 (0.96-1.00)	.01
Depression	1.24 (0.51-3.03)	.63
Personality disorder	0.68 (0.34-1.37)	.28
Substance use disorder	1.34 (0.73-2.46)	.35
Posttraumatic stress disorder	1.22 (0.66-2.26)	.52
Anxiety disorder	1.12 (0.52-2.40)	.77
HoNOS[a] entry	0.98 (0.93-1.05)	.59
BSCL[b] entry	1.04 (0.94-1.15)	.41

[a]HoNOS: Health of the Nation Outcome Scales.

[b]BSCL: Brief Symptom Checklist.

Logistic Regression Model on Factors Potentially Associated With Dropout

In the logistic regression model assessing factors associated with dropout, anxiety disorder emerged as a statistically significant predictor. Patients diagnosed with an anxiety disorder had significantly lower odds of dropping out (OR 0.23, 95% CI 0.08-0.66; *P*=.007), indicating a higher likelihood of continued engagement with the platform. In addition, 2 nonsignificant trends were observed. First, patients with a personality disorder showed a tendency toward lower dropout rates (OR 0.50, 95% CI 0.21-1.19; *P*=.11), suggesting a potential association that may need further investigation. Second, there was a weak trend indicating that higher symptom severity at admission, as measured by the HoNOS score, might be associated with increased dropout risk (OR 1.07, 95% CI 0.99-1.15; *P*=.08), though this finding did not reach statistical significance (Table 3).

XSL•FO
RenderX

Table 3. Logistic regression model of factors potentially associated with the dropout of the online platform.

Parameter	Odds ratio (95% CI)	*P* value
Sex (male)	0.88 (0.44-1.74)	.70
Age (difference per year)	0.98 (0.96-1.01)	.15
Depression	1.1 (0.36-3.31)	.86
Personality disorder	0.5 (0.21-1.19)	.11
Substance use disorder	1.25 (0.57-2.73)	.57
Posttraumatic stress disorder	0.88 (0.40-1.95)	.74
Anxiety disorder	0.23 (0.08-0.66)	.007
HoNOS[a] entry	1.07 (0.99-1.15)	.08
BSCL[b] entry	1.05 (0.93-1.19)	.39

[a]HoNOS: Health of the Nation Outcome Scales.

[b]BSCL: Brief Symptom Checklist.

Discussion

Summary and Interpretation of Findings

This study identified 2 significant predictors of engagement with BT among inpatients: increasing age was associated with a lower risk of noninitiation, while comorbid anxiety disorder was associated with a lower risk of dropout. These findings are particularly noteworthy as they differ depending on whether noninitiation or dropout is examined, and these associations were observed in a heterogeneous patient population encompassing a broader range of diagnoses than that included in a previous study [10], which was limited to patients with anxiety disorders and depression. In contrast, our study examines a more heterogeneous patient population with a wider range of diagnoses, allowing a more comprehensive analysis of the BT's applicability across different conditions. The statistical analysis reveals significant findings, which interestingly differ depending on whether we examine the issue of noninitiation or dropout.

Previous studies have demonstrated that nonparticipation in general psychotherapy does not occur at random over time. The risk is particularly high at the beginning of the therapeutic process, often even before the first therapy session [18]. This finding is consistent with our study, where approximately one-third of the participants (94/278, 33.8%) were classified as belonging to the noninitiation group. This pattern suggests that not initiating a therapy may be driven not by the intervention itself but by other underlying factors. Therefore, it is essential to monitor this vulnerable phase closely and provide enhanced support precisely at this critical phase.

A striking finding of this study is the overall low uptake of the intervention. Despite being offered to a preselected group of interested patients, only 5 (1.8%) out of 278 patients completed all the assigned modules. One-third of patients never initiated the platform at all, and of those who did, most did not complete their assigned modules, resulting in a mean module completion rate of approximately 24% (mean 0.9, SD 2.0 out of 3.8, SD 3.4 assigned modules). These findings are consistent with the broader literature on digital interventions, which has repeatedly documented high rates of nonuse and early disengagement [5,6].

Several factors may help explain this low overall uptake. First, the absence of standardized follow-up procedures likely played a relevant role. Without systematic monitoring of patient engagement, low levels of platform use may have gone unnoticed or unaddressed by therapists [4]. Second, the lack of standardized criteria for module assignment may have contributed to low completion rates. Therapists in a qualitative study conducted at the same institution reported uncertainty about which modules were suitable for which patients and emphasized that active integration and follow-up are essential for meaningful engagement and that module assignment alone is insufficient unless supported by structured guidance [4].

Furthermore, the findings suggest that increasing age has a protective effect against noninitiation, indicating that older patients may be more likely to begin the intervention after modules have been assigned by the therapist. Previous research investigating internet-based cognitive behavioral therapy has shown that older adults tend to complete more modules and remain engaged over longer periods compared to younger users [19-22]. These studies suggest that older individuals may benefit from greater persistence, show more stable use patterns, and show potentially higher motivation, resulting in better adherence and lower dropout rates in digital interventions. One hypothesis could be that older individuals may adhere slightly better to digital interventions because they are more likely to be retired and therefore may have more time available to engage with therapy. However, this explanation does not apply to this study, as all participants were treated under inpatient conditions and therefore had access to comparable time resources. This suggests that factors other than time availability may underlie the higher adherence observed among older patients. A possible explanation is that younger users may approach digital interventions in a more exploratory way while lacking a strong initial commitment to the intervention. If immediate benefits are not apparent, they may disengage quickly. In contrast, older adults may begin treatment with more realistic expectations and greater persistence. Their decision to engage with digital therapy, despite potential barriers such as unfamiliarity with technology or initial skepticism, may be due to a high baseline level of intrinsic motivation. Younger individuals, for whom

XSL•FO
RenderX

digital tools are more embedded in daily life, may face lower entry thresholds, resulting in a more heterogeneous group with varying levels of motivation, including those with only low therapeutic interest. Additionally, older adults who agree to participate in digital treatment may possess other characteristics associated with treatment success, such as higher education, more cognitive resources, or greater health-related self-efficacy.

Our findings indicate that patients with comorbid anxiety disorders in inpatient settings have a lower likelihood of discontinuing therapy. Anxiety disorders frequently cooccur with other psychiatric conditions, making it important to investigate their potential influence on the chance of dropouts. The prior evidence regarding the relationship between comorbid anxiety and psychotherapy dropout has been mixed and inconsistent. Studies have examined anxiety as a predictor of dropout both in general psychotherapy settings and specifically in the context of internet-based interventions. Some studies suggest that comorbid anxiety is associated with lower dropout rates [23], while others report the opposite, indicating a higher likelihood of early termination [24]. The evidence remains inconclusive. This raises the question of why patients with anxiety disorders in our study were less likely to drop out. One possible hypothesis relates to personality characteristics. Evidence suggests that individuals with anxiety disorders more frequently display cluster C personality traits [25]. These traits are associated with high conscientiousness, and a strong sense of duty can have a link to improved treatment adherence. Such individuals may be more willing to follow therapeutic guidance and continue engaging with assigned modules due to their conscientious traits. Another possible explanation for the lower dropout rate among patients with comorbid anxiety is that the structured and predictable nature of BT aligns well with their needs. The clear framework and guidance of online modules may support adherence by complementing traits commonly associated with anxiety. Additionally, the platform's transdiagnostic content may target essential mechanisms of anxiety, such as rumination and avoidance, which may increase the perceived personal relevance of the material. In addition, anxious patients also tend to closely monitor their own progress, which may support ongoing participation, particularly in digital formats that involve self-tracking or structured exercises. It is also possible that individuals with anxiety disorders require additional effort to engage in social situations, including those encountered during face-to-face psychotherapy. psychotherapy. Working on a digital platform may reduce the potentially distressing effects of direct interpersonal interaction, which could make digital therapy a more appealing and manageable option for them. Finally, the higher frequency of contact points in BT may lead to a sense of safety and continuity that helps reduce uncertainty.

Implications of the Findings

The results of this study could lead to 2 possible approaches. On the one hand, it might be beneficial to provide additional support to patients with a high risk of dropout or noninitiation to ensure successful completion of the therapy, for instance, by giving special attention to these patients. On the other hand, one could argue that these patients may not benefit from this therapy form and that alternative therapeutic approaches should be prioritized for them. Considering the wide range of available therapy options, certain approaches may simply be better suited for specific conditions. Wentzer et al [26] aimed to develop a tool to systematically assess factors critical for successful outcomes in BT, focusing on aspects such as computer access, internet skills, acute crises, and patient motivation. Similarly, Karyotaki et al [27] applied a personalized medicine approach in a systematic review to improve therapy allocation for patients with depression. Using sociodemographic characteristics and baseline symptom severity as predictors, the authors developed an interactive decision-support tool to guide treatment allocation among guided digital therapy, unguided digital therapy, and treatment as usual. While these studies concentrated on either practical barriers or depression-specific treatment optimization, this study extends this approach by including a broader patient population across multiple diagnoses. Such efforts are essential to optimize both treatment outcomes and resource use. In the future, standardized tools could help predict dropout risk, thereby enabling more targeted and effective implementation of BT. Accurate identification of suitable patients may enhance therapeutic efficiency and reduce health care costs. However, the findings of this study should be interpreted within their specific context. Our analysis focuses exclusively on adherence to BT, without assessing treatment outcomes. It is conceivable that some patients completed modules without having any clinical benefit, for example, to meet expectations from their therapists. Nevertheless, previous research has shown that homework adherence is positively associated with therapeutic success [28]. In line with this, we assume that module completion in BT also contributes to improved outcomes. By identifying risk factors for noninitiation and dropout, this study offers valuable insights into an underexplored area of digital mental health. It supports a more personalized and cost-efficient approach to psychiatric care by helping to match BT more precisely to patients who are likely to benefit. It expands the existing evidence base and contributes to the advancement of targeted, evidence-based treatment strategies.

Limitations

Our findings must be interpreted considering several limitations. A limitation is the heterogeneity of the data, collected over an extended period (January 2020 to May 2024), with considerable variability in therapist training, application methods, and patient selection. No standardized procedures were applied, and information on selection criteria is missing. Importantly, as access to the Minddistrict platform was granted selectively by therapists to interested patients, no systematic data were collected on patients who were not offered BT during the study period. This precludes any formal comparison between included and nonincluded patients, and the extent and direction of the resulting selection bias therefore remains unknown. Future prospective studies should consider collecting data on all potentially eligible patients, regardless of whether they are ultimately offered BT, to allow for such comparisons.

Moreover, therapist-related variables are incomplete. Prior research shows that the extent and consistency of guidance significantly affect adherence [29,30] and that therapists' attitudes toward digital tools play a role in patient engagement. The therapeutic relationship is also a known protective factor

XSL•FO
RenderX

against dropout [31]. The absence of standardized training and varying application strategies may have contributed to differences in adherence. Additional influencing factors such as self-efficacy, expectations, digital affinity, and therapeutic alliance were not assessed. Finally, while the study reveals associations between predictors and dropout or noninitiation, it does not clarify underlying mechanisms. Qualitative studies could offer further insights, and future prospective studies with standardized protocols and more homogeneous samples are recommended. These limitations should be considered when interpreting the findings.

Acknowledgments

The authors would like to thank the staff of *Clienia AG* for assistance with data collection.

Funding

Financial support was provided by *Schwyn Stiftung*.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: NG
Data curation: NG
Formal analysis: NG
Supervision: TB, WK
Validation: TB, WK
Investigation: NG
Methodology: NG, FG
Project administration: NG, WK
Resources: NG
Visualization: NG
Writing—original draft: NG
Writing—review and editing: NG, FG, TB, WK

Conflicts of Interest

None declared.

References

1. Romijn G, Batelaan N, Koning J, van Balkom A, de Leeuw A, Benning F, et al. Acceptability, effectiveness and cost-effectiveness of blended cognitive-behavioural therapy (bCBT) versus face-to-face CBT (ftfCBT) for anxiety disorders in specialised mental health care: a 15-week randomised controlled trial with 1-year follow-up. PLoS One 2021 Nov 12;16(11):e0259493 [FREE Full text] [doi: 10.1371/journal.pone.0259493] [Medline: 34767575]
2. Berger T, Krieger T, Sude K, Meyer B, Maercker A. Evaluating an e-mental health program ("deprexis") as adjunctive treatment tool in psychotherapy for depression: results of a pragmatic randomized controlled trial. J Affect Disord 2018 Feb;227:455-462. [doi: 10.1016/j.jad.2017.11.021] [Medline: 29154168]
3. Mol M, van Genugten C, Dozeman E, van Schaik DJ, Draisma S, Riper H, et al. Why uptake of blended internet-based interventions for depression is challenging: a qualitative study on therapists' perspectives. J Clin Med 2019 Dec 30;9(1):91 [FREE Full text] [doi: 10.3390/jcm9010091] [Medline: 31905834]
4. Gupta N, Leuba S, Seifritz E, Berger T, Kawohl W. Resources, support, and integration as potential barriers and facilitators to the implementation of blended therapy in the routine care of inpatients: a qualitative study. Front Psychiatry 2024 Dec 17;15:1417784 [FREE Full text] [doi: 10.3389/fpsyt.2024.1417784] [Medline: 39742329]
5. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
6. de Graaf LE, Huibers MJ, Riper H, Gerhards SA, Arntz A. Use and acceptability of unsupported online computerized cognitive behavioral therapy for depression and associations with clinical outcome. J Affect Disord 2009 Aug;116(3):227-231. [doi: 10.1016/j.jad.2008.12.009] [Medline: 19167094]
7. Loh A, Leonhart R, Wills CE, Simon D, Härter M. The impact of patient participation on adherence and clinical outcome in primary care of depression. Patient Educ Couns 2007 Jan;65(1):69-78. [doi: 10.1016/j.pec.2006.05.007] [Medline: 17141112]

XSL•FO
RenderX

8. El Alaoui S, Ljótsson B, Hedman E, Kaldo V, Andersson E, Rück C, et al. Predictors of symptomatic change and adherence in internet-based cognitive behaviour therapy for social anxiety disorder in routine psychiatric care. PLoS One 2015 Apr 20;10(4):e0124258 [FREE Full text] [doi: 10.1371/journal.pone.0124258] [Medline: 25893687]
9. Karyotaki E, Kleiboer A, Smit F, Turner DT, Pastor AM, Andersson G, et al. Predictors of treatment dropout in self-guided web-based interventions for depression: an 'individual patient data' meta-analysis. Psychol Med 2015 Apr 17;45(13):2717-2726. [doi: 10.1017/s0033291715000665]
10. Wu MS, Chen SY, Wickham RE, Leykin Y, Varra A, Chen C, et al. Predicting non-initiation of care and dropout in a blended care CBT intervention: impact of early digital engagement, sociodemographic, and clinical factors. Digit Health 2022 Oct 26;8:20552076221133760 [FREE Full text] [doi: 10.1177/20552076221133760] [Medline: 36312847]
11. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
12. Fuhr K, Schröder J, Berger T, Moritz S, Meyer B, Lutz W, et al. The association between adherence and outcome in an internet intervention for depression. J Affect Disord 2018 Mar 15;229:443-449. [doi: 10.1016/j.jad.2017.12.028] [Medline: 29331706]
13. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
14. Davies F, Shepherd HL, Beatty L, Clark B, Butow P, Shaw J. Implementing web-based therapy in routine mental health care: systematic review of health professionals' perspectives. J Med Internet Res 2020 Jul 23;22(7):e17362 [FREE Full text] [doi: 10.2196/17362] [Medline: 32706713]
15. Franke GH. Brief Symptom Checklist. Bern, Switzerland: Hogrefe Publishing Group; 2017.
16. Andreas S, Harfst T, Dirmaier J, Kawski S, Koch U, Schulz H. A psychometric evaluation of the German version of the 'Health of the Nation Outcome Scales, HoNOS-D': on the feasibility and reliability of clinician-performed measurements of severity in patients with mental disorders. Psychopathology 2007;40(2):116-125. [doi: 10.1159/000098492] [Medline: 17215598]
17. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Softw 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]
18. Fernandez E, Salem D, Swift JK, Ramtahal N. Meta-analysis of dropout from cognitive behavioral therapy: magnitude, timing, and moderators. J Consult Clin Psychol 2015 Dec;83(6):1108-1122. [doi: 10.1037/ccp0000044] [Medline: 26302248]
19. Hobbs MJ, Joubert AE, Mahoney AE, Andrews G. Treating late-life depression: comparing the effects of internet-delivered cognitive behavior therapy across the adult lifespan. J Affect Disord 2018 Jan 15;226:58-65. [doi: 10.1016/j.jad.2017.09.026] [Medline: 28963865]
20. Hobbs MJ, Mahoney AE, Andrews G. Integrating iCBT for generalized anxiety disorder into routine clinical care: treatment effects across the adult lifespan. J Anxiety Disord 2017 Oct;51:47-54. [doi: 10.1016/j.janxdis.2017.09.003] [Medline: 28926805]
21. Pabst A, Löbner M, Stein J, Luppa M, Kersting A, König HH, et al. Internet-based cognitive behavior therapy only for the young? A secondary analysis of a randomized controlled trial of depression treatment. Front Psychiatry 2020 Jul 24;11:735 [FREE Full text] [doi: 10.3389/fpsyt.2020.00735] [Medline: 32848915]
22. Mewton L, Sachdev PS, Andrews G. A naturalistic study of the acceptability and effectiveness of internet-delivered cognitive behavioural therapy for psychiatric disorders in older Australians. PLoS One 2013 Aug 12;8(8):e71825 [FREE Full text] [doi: 10.1371/journal.pone.0071825] [Medline: 23951253]
23. van Dijk DA, Deen ML, van den Boogaard TM, Ruhé HG, Spijker J, Peeters FP. Prevalence and prediction of dropout during depression treatment in routine outpatient care: an observational study. Eur Arch Psychiatry Clin Neurosci 2023 Aug;273(5):1151-1161 [FREE Full text] [doi: 10.1007/s00406-022-01499-1] [Medline: 36253582]
24. van Bronswijk SC, Lemmens LH, Huibers MJ, Arntz A, Peeters FP. The influence of comorbid anxiety on the effectiveness of Cognitive Therapy and Interpersonal Psychotherapy for Major Depressive Disorder. J Affect Disord 2018 May;232:52-60. [doi: 10.1016/j.jad.2018.02.003] [Medline: 29477584]
25. Brandes M, Bienvenu OJ. Personality and anxiety disorders. Curr Psychiatry Rep 2006 Aug;8(4):263-269. [doi: 10.1007/s11920-006-0061-8] [Medline: 16879789]
26. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen JE. Mixing online and face-to-face therapy: how to benefit from blended care in mental health care. JMIR Ment Health 2016 Feb 09;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]
27. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FM, Furukawa TA, Cuijpers P, Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry 2021 Apr 01;78(4):361-371 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]
28. Mausbach BT, Moore R, Roesch S, Cardenas V, Patterson TL. The relationship between homework compliance and therapy outcomes: an updated meta-analysis. Cognit Ther Res 2010 Oct;34(5):429-438 [FREE Full text] [doi: 10.1007/s10608-010-9297-z] [Medline: 20930925]

XSL•FO
RenderX

29. Zarski AC, Lehr D, Berking M, Riper H, Cuijpers P, Ebert DD. Adherence to internet-based mobile-supported stress management: a pooled analysis of individual participant data from three randomized controlled trials. J Med Internet Res 2016 Jun 29;18(6):e146 [FREE Full text] [doi: 10.2196/jmir.4493] [Medline: 27357528]
30. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
31. Sharf J. Psychotherapy Dropout: A Meta-analytic Review of Premature Termination. New York, NY: Adelphi University; 2007.

Abbreviations

BSCL: Brief Symptom Checklist
BT: blended therapy
HoNOS: Health of the Nation Outcome Scales
OR: odds ratio

Edited by A Kushniruk; submitted 13.Oct.2025; peer-reviewed by J Diemer; comments to author 08.Apr.2026; accepted 29.May.2026; published 29.Jun.2026.

Please cite as:
Gupta N, Gardin F, Berger T, Kawohl W
Risk Factors for Noninitiation and Dropout in Blended Therapy in Inpatient Psychiatric Patients: Retrospective Cohort Study
JMIR Hum Factors 2026;13:e85816
URL: https://humanfactors.jmir.org/2026/1/e85816
doi:10.2196/85816
PMID:

©Nikita Gupta, Fabian Gardin, Thomas Berger, Wolfram Kawohl. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 29.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Wellness4Teachers Supportive Text Messaging Program Association With Observed Improvement in Psychological Symptoms Among Teachers in Canada: Single-Arm Prospective Longitudinal and Naturalistic Controlled Study

Belinda Agyapong[1,2], MEd, PhD; Yifeng Wei[1], PhD; Pamela Brett-MacLean[1], PhD; Vincent Israel Opoku Agyapong[1,2], MD, PhD

[1]Faculty of Medicine & Dentistry, University of Alberta, Edmonton, AB, Canada

[2]Dalhousie University, Halifax, NS, Canada

Corresponding Author:
Belinda Agyapong, MEd, PhD
Faculty of Medicine & Dentistry
University of Alberta
4-142A Katz Group Centre for Research
11315- 87 Ave NW
Edmonton, AB, T6G 2H5
Canada
Phone: 1 7804922487
Email: bagyapon@ualberta.ca

Abstract

Background: Wellness4Teachers is a daily supportive text messaging program implemented to mitigate psychological distress and support teachers' well-being in Canada.

Objective: This study aimed to assess whether Wellness4Teachers can reduce the severity and prevalence of psychological issues and improve resilience among teachers in Alberta, Newfoundland and Labrador, and Nova Scotia.

Methods: This study used a single-arm prospective longitudinal design with a secondary exploratory naturalistic contemporaneous comparison. Teachers self-subscribed to the program for 6 months and completed web-based surveys. Longitudinal outcomes were assessed at baseline, 6 weeks, 3 months, and 6 months. The naturalistic comparison included an intervention group (IG), which received messages for 6 months and completed 6-month assessments between March 1 and June 30, 2023, and a control group (CG), which enrolled and completed baseline assessments during the same period before receiving messages. Standardized scales were used for assessing psychological symptoms. Likely generalized anxiety disorder (GAD), likely major depressive disorder (MDD), resilience, burnout, and stress were assessed using the Generalized Anxiety Disorder-7 (GAD-7) scale, the Patient Health Questionnaire-9 (PHQ-9), the Brief Resilience Scale, the Maslach Burnout Inventory-Educator Survey, and the Perceived Stress Scale-10 (PSS-10), respectively. Data were analyzed using SPSS. Paired- and independent-sample *t* tests were used to compare longitudinal changes and groups, respectively. McNemar and chi-square tests were used to examine categorical variables. Missing follow-up data were addressed using last observation carried forward, with sensitivity analyses comparing baseline scores with those at 6 weeks, 3 months, and 6 months. The group type (IG vs CG) was entered as a variable together with all demographic and work-related variables into 9 separate logistic regression models.

Results: In the longitudinal analysis, PSS-10, Maslach Burnout Inventory-Emotional Exhaustion subscale, PHQ-9, PHQ-9 item 3 (sleep disturbance), PHQ-9 item 9 (suicidal ideation/self-harm), and GAD-7 scores decreased significantly at 6 months, by 5.50% (mean difference [MD]=1.29, 95% CI 0.86-1.72), 4.48% (MD=1.67, 95% CI 0.43-2.92), 18.58% (MD=2.10, 95% CI 1.39-2.80), 17.54% (MD=0.30, 95% CI 0.15-0.44), 37.50% (MD=0.09, 95% CI 0.02-0.17), and 13.52% (MD=1.30, 95% CI 0.62-1.99), respectively. Likely MDD decreased from 55.2% (n=116) to 41.0% (n=86), a 14.2% reduction. In the exploratory contemporaneous comparison, the IG (n=166) had lower PSS-10 (6.04%; MD=–1.41, 95% CI –2.17 to –0.65), PHQ-9 (21.16%; MD=–2.33, 95% CI –3.59 to –1.06), and GAD-7 (22.48%; MD=–2.19, 95% CI –3.39 to –0.98) scores than the CG (n=152). Likely MDD and GAD prevalence were 20.6% and 20.3% lower, respectively. Adjusted analyses showed IG status was associated

XSL•FO
RenderX

with lower odds of high stress (odds ratio [OR] 0.42, 95% CI 0.19-0.95), likely MDD (OR 0.47, 95% CI 0.28-0.81), and likely GAD (OR 0.37, 95% CI 0.21-0.64).

Conclusions: Wellness4Teachers was associated with improvements in several psychological outcomes over 6 months and may offer a promising approach to support teachers' psychological well-being.

International Registered Report Identifier (IRRID): RR2-10.2196/37934

(JMIR Hum Factors 2026;13:e92142) doi:10.2196/92142

KEYWORDS

anxiety; burnout; depression; educators; resilience; stress; teachers; text messages; Wellness4Teachers

Introduction

Background

Teachers play diverse roles in educating students beyond simple curriculum delivery. Teachers also serve as mentors, counselors, and facilitators of social and emotional learning and work to ensure an enriched, supportive learning environment that fosters students' overall development and well-being [1]. As important a role as teachers have, their work exposes them to numerous occupational stressors, which can result in teachers feeling overwhelmed. Key stressors include demanding workloads, challenging student behaviors and classroom management, and poor working conditions (insufficient resources and lack of support from the administration) [2]. These factors can contribute to high levels of stress, which can lead to burnout, decreased job satisfaction, and psychological health impacts such as anxiety and depression [3,4]. Chronic stress occurs when persistent external pressures challenge an individual's ability to cope effectively, leading to significant physical and emotional strain, which can manifest in various physical, emotional, and psychological symptoms [5-8].

When stress becomes chronic or unmanageable, it can evolve into burnout—a state of exhaustion, cynicism, and reduced effectiveness that impacts mental and physical health [9]. Burnout is characterized by 3 dimensions: emotional exhaustion (feeling overextended physically and emotionally), cynicism (a cold, detached attitude toward work and people at work), and inefficacy (increased sense of inadequacy) [9,10]. Among teachers, burnout is a significant and growing concern that can exacerbate emotional distress and psychological difficulties, contributing to a lack of professional fulfillment [7]. Burnout can also impact teachers' health, lifestyle, coping ability, and the quality of education they provide, as well as their personal functioning or lifestyle [10]. High stress, high burnout, and low coping ability among teachers have been associated with poor student outcomes, and teachers with high levels of anxiety also struggle with classroom management [11]. Individuals vary in how they respond to stress and burnout, and resilience acts as a protective factor or buffer, helping them to cope with stress, anxiety, and depression. Among the general population, research has documented a moderate negative relationship between resilience and psychological distress [12]. Higher resilience has been reported to be significantly associated with reduced stress, improved mental health, and better well-being [13].

Resilience is a complex concept that refers to the adaptive capacity to withstand, adapt to, and recover from adversity, stress, or trauma [14]. It can be developed at various levels, including individual, community, and institutional levels, and encompasses a range of adaptive or positive processes, including coping strategies (problem-solving and emotional regulation), social support (relational connections and counseling), cognitive restructuring or positive reframing (including finding meaning), creative expression, and practicing self-compassion that can enhance resilience by fostering strength and adaptability [15]. Variation in how individuals respond to similar stressors suggests that resilience can be inhibited by the detrimental effects of stress, adversity, or biological factors, and personal history, coping mechanisms, and psychological makeup can influence differing levels of resilience [14]. Resilience requires self-awareness and internal coherence, which are activated by personal development; thus, it requires continuous cultivation throughout one's lifespan and opportunities for growth [16].

Depression and anxiety are prevalent mental health issues that can significantly impact the well-being and professional effectiveness of teachers [5]. Published research reported that teachers with depression self-reported higher levels of perceived stress, anxiety, and lower quality of life [17]. Higher levels of depressive symptoms have been documented among teachers compared to other professionals, with factors such as job adversity and lack of support being key contributors [18]. Anxiety symptoms are prevalent among teachers, ranging from 38% to 41.2%, and are usually exacerbated among those working in high-stress environments [5]. The impact of anxiety and depression on educators extends far beyond their well-being and affects their professional performance, productivity, competence, and student achievement [19]. Based on this, it is important to support teachers' psychological well-being, as doing so enhances their professional role and, invariably, students' achievement.

Recently, there has been growing evidence of the use and feasibility of mobile technology in closing psychological treatment gaps by fostering resilience and enhancing overall health and well-being [20]. The use of eHealth technology has been popular over the years, even among the older population, where therapist-guided internet-delivered cognitive behavioral therapy (CBT) significantly lowered the scores on the Patient Health Questionnaire-9 (PHQ-9) for depression symptoms and Generalized Anxiety Disorder-7 (GAD-7) for anxiety symptoms in the intervention group (IG) compared to the control group (CG) at postintervention [21]. Various mobile health apps, such as Psychological First Aid and the Behavioural Health Disaster Response App, have also been adopted and assessed for their efficacy, feasibility, and acceptability as interventions during

XSL•FO
RenderX

disasters [22]. Several interventions have also been explored, particularly to address burnout and stress, which usually lead to anxiety or depression among teachers. However, many of these interventions are limited by time, cost, long wait times, and geographic location barriers to access [19]. The most popularly studied interventions were mindfulness-based interventions alone or in combination with CBT or the behavioral and cognitive-behavioral domains [19,23]. For example, Cultivating Awareness and Resilience in Education (CARE for Teachers) is a mindfulness-based professional development program designed to promote teachers' social and emotional competence and improve the quality of classroom interactions. It had significant positive effects on adaptive emotion regulation, mindfulness, and psychological distress [24]. Though mindfulness and other interventions reported positive outcomes, they presented the issue of time and geographical location constraints for some teachers [19], and only a few online programs are focusing on the mental health of teachers with limited evidence of effectiveness [25]. Digital mental health support, specifically supportive text messaging interventions, has been reported to mitigate symptoms of anxiety and depression among patients with mental health problems who are receiving formal psychiatric care [26], among young adults [27], and among first responders [28]. Text message–delivered health behavior intervention also led to a significant reduction in glycated hemoglobin in adults with type 2 diabetes mellitus [29]. Additionally, text-message interventions have been shown to be effective in promoting vaping cessation among young adults [30].

Wellness4Teachers Program

Based on existing evidence of the effectiveness of mobile text messaging interventions in diverse contexts, the Wellness4Teachers program was developed and implemented to support teachers in Canada during the 2022 to 2023 academic year. Wellness4Teachers [31] is a self-subscription daily supportive text message intervention program proposed to address stress, burnout, anxiety, and depression, build resilience, and improve professional fulfillment in educators. Teachers can subscribe to Wellness4Teachers by texting "TeachWell" to a designated phone number and receive daily CBT-based text messages with or without embedded links to mental health literacy resources or information for 6 months. The Wellness4Teachers application sent unidirectional, noninteractive psychological well-being messages to teachers' mobile phones. Subscribers were informed of the noninteractive nature of the supportive messaging program through the first message they received after subscription. The daily supportive messages delivered through the Wellness4Teachers program were crafted by psychiatrists, mental health therapists, and psychologists and reviewed by education specialists. The messages are based on CBT principles, and teachers received different messages daily from a bank of messages for 6 months. Sample messages are found below; other examples can be accessed in the protocol [31].

> *Take it one day at a time. Do not expect too much too soon. Set realistic, doable daily goals. Take part in activities that make you feel better. Seek out people who are supportive and helpful.*
>
> *Stress and burnout can lead to poor physical and mental well-being, and it may affect one's health. Try to manage your time effectively and take breaks when needed.*
>
> *Mistakes are facts of life; the response to error counts. Turn your adversities into opportunities. Try new things that will bring positive outcomes.*

Study Aim

This study assesses whether the Wellness4Teachers daily supportive text messaging program can reduce the severity and prevalence of stress, burnout, anxiety, and depression symptoms and improve resilience among teachers in Alberta, Newfoundland and Labrador, and Nova Scotia.

Methods

Study Setting

Canada is a North American country with 10 provinces and 3 territories and a population of 40,769,890 in January 2024 [32]. This study was conducted in 3 provinces: Newfoundland and Labrador, Alberta, and Nova Scotia. Newfoundland and Labrador Schools is a group tasked with educating all English-speaking students at kindergarten to grade 12 public schools and has over 10,000 employees [33,34]. Nova Scotia has approximately 10,000 public school teachers represented by the Nova Scotia Teachers Union [35], while Alberta has 32,523 full- and part-time teachers [36].

Study Design

This study used longitudinal and naturalistic controlled designs. The longitudinal component examined outcomes in Wellness4Teachers subscribers who completed surveys at baseline and 6 months. An intention-to-treat analysis was conducted, where the last follow-up observation—either 6 weeks or 3 months—was carried forward to the 6-month mark for subscribers who did not complete the final 6-month survey. The last-observation-carried-forward (LOCF) approach was selected to maximize the use of available longitudinal data. However, this approach assumes stable outcomes after dropout and may underestimate variability or introduce bias, warranting cautious interpretation of the estimated intervention effects. The naturalistic comparison was constructed using contemporaneously collected outcome assessments rather than by prospectively following an untreated CG for 6 months. The IG comprised subscribers who had enrolled approximately 6 months earlier, received the Wellness4Teachers daily supportive text messages for 6 months, and completed their 6-month outcome assessment between March 1 and June 30, 2023. The naturalistic CG comprised new subscribers who enrolled between March 1 and June 30, 2023, and completed their baseline assessment before receiving any supportive text messages. Thus, outcome measurements for both groups were obtained during the same calendar period, although the groups differed in elapsed time since program enrollment and intervention exposure. This design was intended to provide a contemporaneous untreated reference group while reducing potential seasonal and other calendar-time influences. It was not intended to constitute a conventional longitudinal controlled

XSL•FO
RenderX

trial in which intervention and control participants are followed over matched elapsed periods.

Ethical Considerations

The study protocol was approved by the University of Alberta Reviews and Ethics Board (Pro00086163) and the Dalhousie University Ethics Review Board (REB #2023-6840), and the study was conducted in accordance with ethical principles for medical research involving human subjects as outlined in the Declaration of Helsinki. Informed consent was implied when subscribers completed the online survey and submitted responses, as approved by the ethics review boards. Given the noninteractive, minimal-risk nature of the digital intervention, implied consent was considered ethically appropriate. Participants were informed that enrollment and continued participation indicated their voluntary consent to participate in the study.

Data Collection

The published study protocol provides a detailed description of data collection procedures and the survey instruments [31]. The Wellness4Teachers support program was available to all teachers in the participating provinces and was advertised by the teachers' associations or unions in these provinces. Teachers in these provinces could enroll in the Wellness4Teachers program by texting "TeachWell" to a specified telephone number. The program was self-subscription, and participation in the survey was voluntary.

The first message from the Wellness4Teachers program welcomes subscribers to the service and invites them to complete a baseline survey. Subscribers receive additional invitations via links in text messages to complete follow-up surveys at 6 weeks, 3 months, and 6 months. Teachers who subscribed could decide to opt in or out of completing the associated survey at baseline or at other time points (6 weeks, 3 months, and 6 months). Participation in the Wellness4Teachers service to receive daily supportive text messages was independent of survey completions, and subscribers could opt out at any time by texting "Stop TeachWell" to the designated phone number. Program messages were delivered by a computer program at 9:00 AM Mountain Time each day. The Wellness4Teachers surveys captured various sociodemographic and work-related data provided as categorical variables (age, sex at birth, relationship status, school setting, major role in school, average class size, and teaching experience in years) and standardized scales for assessing psychological symptoms, including the Perceived Stress Scale-10 (PSS-10; a score of ≥27 indicates high stress) [37], the GAD-7 (a score of ≥10 indicates likely generalized anxiety disorder) [38], the PHQ-9 (a score ≥10 indicates likely major depressive disorder [MDD]) [39], the Brief Resilience Scale (BRS; mean scores ranging from 1.00 to 2.99 indicate low resilience and a score of ≥3.00 suggests normal to high resilience) [40], and the Maslach Burnout Inventory-Educator Survey (MBI-ES) [41,42]. The MBI-ES contains a 22-item instrument that measures emotional exhaustion (MBI-EE subscale scores ≥27 indicate high emotional exhaustion), depersonalization (MBI-DP subscale scores ≥13 indicate high depersonalization), and personal accomplishment (MBI-PA subscale). The psychometric properties of the PSS-10, MBI-ES, and BRS have been described in a related publication [7].

The GAD-7 questionnaire has a Cronbach α of 0.92 and is used to assess the self-reported levels of anxiety in respondents in the 2 weeks prior to assessment [38]. The PHQ-9 has a Cronbach α of 0.89 and is used to assess the severity of depression symptoms [39]. Data for the longitudinal trial were collected between September 1, 2022, and April 1, 2024, and data for the naturalistic controlled study were collected between March 1, 2023, and June 30, 2023. The survey took, on average, 15 minutes to complete, and subscribers were not offered incentives for completing the surveys.

Protocol Deviation

This study deviated from the protocol published in *JMIR Research Protocols* (2022) [31] in some respects. First, the study was expanded to include a third province, and the research ethics board approval was subsequently updated under a revised approval number. Second, the sample size calculation was modified from that specified in the published protocol to reflect the revised study design and anticipated recruitment. These deviations were made during study implementation in response to changes in the study context, feasibility, and available data, and are reported here to ensure transparency regarding the distinction between the prespecified protocol and the final study design. The planned qualitative key informant interviews have not yet been conducted and therefore do not form part of the present evaluation. When feasibly implemented, the qualitative evaluation will be published separately in a subsequent manuscript with its relationship to this study described accordingly.

Outcome Measures

Primary outcome measures of interest included the psychological scale scores and symptom prevalence rates across specific evaluation time points on the psychological scales (PSS-10, MBI-ES subscales, PHQ-9, third question in PHQ-9 assessing severity of sleep problems, ninth question in PHQ-9 assessing severity of suicidal thoughts, GAD-7, and the BRS) at baseline and 6 months for the longitudinal study and between the IG and the CG in the naturalistic trial. Other primary outcome measures of interest for this study were differences in the prevalence of psychological symptoms (high stress, high emotional exhaustion, high depersonalization, low professional accomplishment, likely MDD, sleep disturbance, suicidal ideation/thoughts of self-harm, likely generalized anxiety disorder, and low resilience) from baseline to 6 months for the longitudinal study and between the IG (6-month prevalence) and the CG (baseline prevalence) for the naturalistic trial.

A secondary outcome of interest was to determine whether, when controlling for demographic variables, the daily supportive text messages would contribute to lower odds of psychological symptoms in subscribers for participants in the naturalistic controlled study.

Hypothesis

We hypothesized that for subscribers to Wellness4Teachers who complete surveys at the 2 time points, there would be at

XSL•FO
RenderX

least 20% lower mean scores on the psychological scales at 6 months compared with their baseline scores. We also hypothesized that subscribers to Wellness4Teachers in the IG would have at least 20% lower mean scores on the psychological scales compared with corresponding mean scores for the CG participants. These hypotheses are based on the greater than 20% reduction in anxiety symptom scores from baseline to 3 months in the general population subscribers of Text4Hope [43] and the more than 20% decrease in depression symptom scores in the treatment group compared to the CG observed in two randomized controlled trials in Ireland and Canada [44,45]. We further hypothesized that the prevalence of each psychological symptom would be 20% lower at 6 months compared to the baseline for the longitudinal study and 20% lower in the IG compared with the CG in the naturalistic controlled study. Finally, we hypothesize that being in the IG would independently predict the absence of each psychological symptom when all the demographic variables are controlled for in a logistic regression model.

Sample Size Considerations

For the longitudinal study, we determined using an online script [46,47] that a sample size of 82 would be sufficient to detect differences in the baseline and sixth-month prevalence for emotional exhaustion with an 80% power, assuming a 90% prevalence of emotional exhaustion at baseline, a hypothesized difference of 20% between the baseline and sixth-month prevalence, a 2-sided significance of 5% for detecting a difference, a correlation between paired observations of 0%, and after applying continuity correction.

For the naturalistic controlled study, to detect a 20% difference in the prevalence of emotional exhaustion between the IG and the CG, based on a power of 80% (β=.2) and a 2-sided significance level of α=.05, and assuming a 90% prevalence of emotional exhaustion at baseline for the CG, we estimated using an online script [48] that a sample size of 73 per group will be sufficient.

Statistical Analysis

Data were analyzed using SPSS for Windows (version 28; IBM Corp) [49]. Descriptive statistics were used to summarize participants' characteristics and psychological scale scores. For the longitudinal study, sociodemographic characteristics and the prevalence and severity (mean scores) of psychological symptoms were examined between the participants who completed the baseline survey and at least one follow-up survey and participants who completed only the baseline survey, using a chi-square or the independent-sample *t* test. A comparison of baseline psychological scores between follow-up completers and noncompleters was conducted to assess potential attrition bias and determine whether participants who remained in the study differed in their initial psychological status from those who did not complete any follow-up survey.

The paired sample *t* test and the McNemar test, respectively, were used to assess differences in mean scores of each psychological scale and the prevalence of each psychological symptom from baseline to 6 months for subscribers who completed surveys at 2 time points, including at baseline. The LOCF approach, consistent with the prespecified analytic plan, was used to impute 6-month outcomes for subscribers who completed only the 6-week and/or 3-month follow-up surveys. Sensitivity analyses were conducted using complete-case paired-samples *t* tests to compare mean psychological scale scores at baseline with those at 6 weeks, 3 months, and 6 months, including only participants with available data at both time points to assess within-participant changes over time.

For the naturalistic controlled study, demographic characteristics were examined between the 2 study groups, IG and CG. For the main study outcomes, we examined the prevalence of clinically meaningful psychological indicators between the IG and CG to ascertain if there were differences using the chi-square test. In addition, the differences in mean scores for each psychological scale between the IG and CG were also examined using independent-sample *t* tests. Results were summarized by numbers and percentages and compared by chi-square analysis with a 2-tailed criterion (α<.01 after Bonferroni correction). Bonferroni correction was applied by dividing the traditional statistically significant level of .05 by the total number of variables being compared. Percentage changes in mean scores were calculated by dividing the mean score difference by the baseline mean score and multiplying it by a hundred. There was no imputation of missing data, except for the longitudinal analysis, where we applied the LOCF, and the total numbers reported represent the total responses recorded for each variable.

Independent-sample *t* tests and chi-square tests were conducted using baseline data to respectively compare the mean psychological scale scores and sociodemographic characteristics for participants who completed the follow-up surveys at each time point with those who did not complete the follow-up survey for the given time point.

To assess the independent impact of Wellness4Teachers daily supportive messages on the presence of psychological symptoms on subscribers at 6 months, all demographic predictors and "IG" variables were imputed into separate binary logistic regression models. The odds ratios (ORs) from the regression models were examined to determine the associations between "receipt of the daily support" and the likelihood for participants to self-report symptoms consistent with the psychological symptoms while controlling for sociodemographic variables (gender, ethnicity, relationship status, and housing status) in the respective models. Correlation analyses were performed before the regression analysis to rule out strong correlations (r_s≥0.7) among predictor variables.

Results

Longitudinal Study Outcomes

Figure 1 shows that response rates for the baseline and for 6 months applying an intention-to-treat analysis were 43.29% (803/1855) and 12% (222/1855), respectively. Of the 1855 subscribers of Wellness4Teachers who completed the baseline survey, only 222 subscribers completed the baseline and at least one follow-up survey and were included in the longitudinal outcome analysis.

XSL•FO
RenderX

Figure 1. Study flowchart.

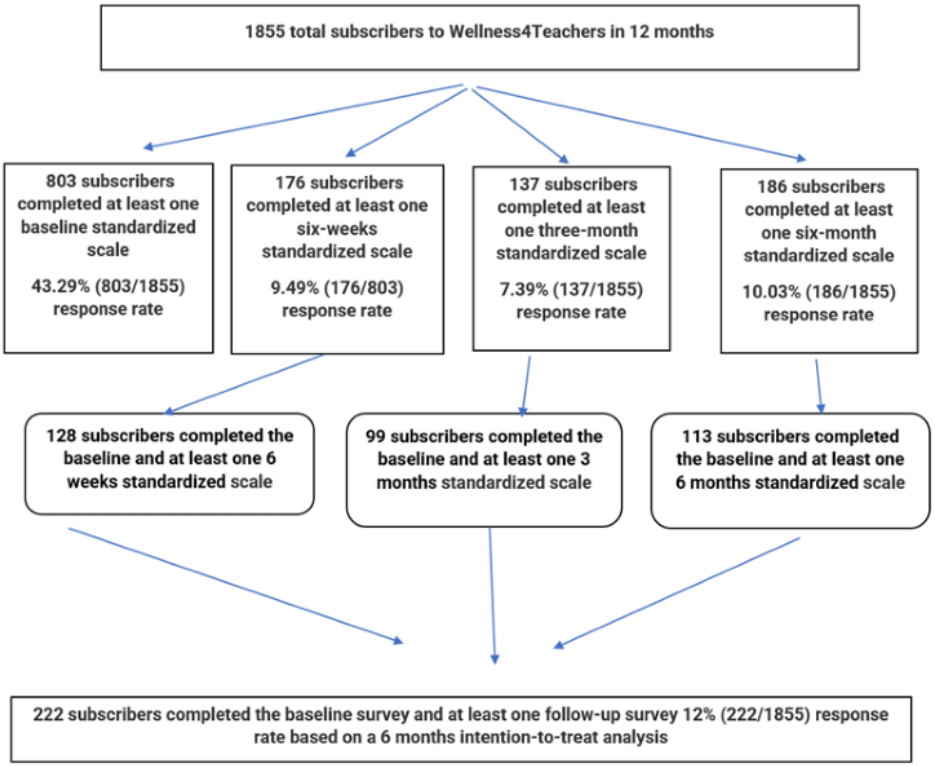


Table S1 in Multimedia Appendix 1 compared the distribution of demographic and work-related variables and the prevalence and severity of psychological symptoms for participants who completed both baseline and at least one follow-up survey and participants who completed only the baseline survey.

Applying a Bonferroni-corrected significant *P* value of ≤.01 (derived by dividing 0.05 by the total number of variables and approximating to 2 decimal places) for the demographic and work-related variables and the psychological variables, Table S1 in Multimedia Appendix 1 indicates that participants who completed the baseline and at least one follow-up survey were similar with respect to all the demographic and work-related characteristics and psychological variables.

For subscribers who completed the baseline and at least one follow-up survey, a paired-samples *t* test was run to assess the changes in mean scores on psychological scales, as shown in Tables 1-4.

Applying a Bonferroni corrected significant *P* value of ≤.01, Table 1 indicates that based on last follow-up observation carried forward, the mean scores on the PSS-10, the PHQ-9, the third and ninth question of the PHQ-9 the GAD-7 and the MBI-EE subscale were significantly lower by 5.51%, 18.58%, 17.54%, 37.5%, 13.52% and 4.84% at 6 months compared to baseline respectively for subscribers who completed the baseline survey and at least one follow-up survey. Using a Bonferroni-corrected significance threshold of *P*≤.01, sensitivity analyses of paired data demonstrated statistically significant changes from baseline for the PSS-10, PHQ-9, and PHQ-9 item 3 (sleep disturbance), at 6 weeks (4.28%, 10.79%, and 14.97%, respectively; Table 2), and there was no significance for any psychological issues at 3 months (Table 3). Among subscribers who completed both the baseline and 6-month surveys, paired-data sensitivity analysis showed statistically significant reductions in mean scores for PSS-10, the MBI-EE subscale, the PHQ-9, PHQ-9 item 3 (sleep disturbance), and GAD-7 (5.87%, 7.77%, 24.98%, 17.06%, and 19.98%, respectively; Table 4). Notably, independent-sample *t* test and chi-square test comparing the baseline psychological scale scores and the sociodemographic characteristics respectively between follow-up completers and noncompleters across each time point (Table S2 in Multimedia Appendix 2 and Table S3 in Multimedia Appendix 3) revealed no statistically significant differences between the 2 groups.

Changes in the prevalence of psychological symptoms of subscribers who completed the baseline and at least one follow-up survey were assessed using the McNemar test, as shown in Table 5.

XSL•FO
RenderX

Table 1. Changes in psychological scale scores from baseline to 6 months: last follow-up observation carried forward results.

Measure	Sample size, n	Baseline score, mean (SD)	Sixth-month score, mean (SD)	Change from baseline (%)	Mean difference (95% CI)	*P* value	*t* test (*df*)	Effect size (Cohen *d*)
PSS-10[a]	216	23.43 (3.02)	22.13 (3.17)	–5.51	1.29 (0.86 to 1.72)	<.001	5.89 (215)	0.40
MBI-EE[b] subscale	215	34.53 (10.54)	32.86 (11.85)	–4.84	1.67 (0.43 to 2.92)	.01	2.66 (214)	0.18
MBI-DP[c] subscale	215	8.05 (6.27)	7.97 (5.93)	–0.90	0.08 (–0.64 to 0.80)	.82	(0.23 214)	0.02
MBI-PA[d] subscale	214	35.30 (6.83)	35.31 (7.06)	+0.02	0.01 (–0.82 to 0.80)	.98	–0.02 (213)	–0.002
PHQ-9[e]	210	11.30 (5.89)	9.20 (5.61)	–18.58	2.10 (1.39 to 2.80)	<.001	5.87 (209)	0.41
PHQ-9 item 3 (severity of sleep disturbance)	210	1.71 (1.01)	1.42 (1.01)	–17.54	0.30 (0.15 to 0.44)	<.001	4.10 (209)	0.28
PHQ-9 item 9 (suicidal ideation/thoughts of self-harm)	210	0.24 (0.59)	0.15 (0.47)	–37.50	0.09 (0.02 to 0.17)	<.001	2.38 (209)	0.16
GAD-7[f]	210	9.61 (5.71)	8.30 (5.56)	–13.53	1.30 (0.62 to 1.99)	<.001	3.78 (209)	0.26
BRS[g]	209	3.04 (0.94)	3.12 (0.93)	+2.63	0.08 (–0.18 to 0.03)	.14	–1.50 (208)	–0.10

[a]PSS-10: Perceived Stress Scale-10.

[b]MBI-EE: Maslach Burnout Inventory-Emotional Exhaustion.

[c]MBI-DP: Maslach Burnout Inventory-Depersonalization.

[d]MBI-PA: Maslach Burnout Inventory-Personal Accomplishment.

[e]PHQ-9: Patient Health Questionnaire-9.

[f]GAD-7: Generalized Anxiety Disorder-7.

[g]BRS: Brief Resilience Scale.

Table 2. Changes in psychological scale scores from baseline to 6 weeks: a paired-sample analysis.

Measure	Sample size, n	Baseline score, mean (SD)	Sixth-week score, mean (SD)	Change from baseline (%)	Mean difference (95% CI)	*P* value	*t* test (*df*)	Effect size (Cohen *d*)
PSS-10[a]	128	23.38 (2.88)	22.38 (2.88)	4.28	1.00 (0.49 to 1.51)	<.001	3.87 (127)	2.93
MBI-EE[b] subscale	122	34.84 (10.66)	34.29 (11.11)	2.73	0.95 (–1.51 to 3.42)	.47	0.73 (121)	8.49
MBI-DP[c] subscale	123	7.98 (6.27)	7.57 (5.72)	15.91	1.27 (–0.37 to 2.91)	.38	0.88 (122)	5.22
MBI-PA[d] subscale	123	36.75 (6.18)	37.09 (5.95)	0.93	–0.34 (–2.02 to 1.34)	.87	0.17 (122)	5.33
PHQ-9[e]	118	11.03 (5.51)	9.85 (5.22)	10.79	1.19 (0.26 to 2.11)	.01	2.54 (117)	5.08
PHQ-9 item 3 (severity of sleep disturbance)	118	1.67 (1.02)	1.42 (1.00)	14.97	0.25 (0.06 to 0.44)	.01	2.66 (117)	1.04
PHQ-9 item 9 (suicidal ideation/thoughts of self-harm)	118	0.21 (0.58)	0.14 (0.46)	33.33	0.07 (–0.03 to 0.17)	.17	1.38 (117)	0.53
GAD-7[f]	117	9.92 (5.40)	9.51 (5.66)	4.13	0.41 (–0.44 to 1.26)	.34	0.95 (116)	4.65
BRS[g]	115	2.96 (0.91)	3.03 (0.90)	2.03	–0.06 (–0.19 to 0.06)	.33	–0.98 (114)	0.68

[a]PSS-10: Perceived Stress Scale-10.

[b]MBI-EE: Maslach Burnout Inventory-Emotional Exhaustion.

[c]MBI-DP: Maslach Burnout Inventory-Depersonalization.

[d]MBI-PA: Maslach Burnout Inventory-Personal Accomplishment.

[e]PHQ-9: Patient Health Questionnaire-9.

[f]GAD-7: Generalized Anxiety Disorder-7.

[g]BRS: Brief Resilience Scale.

XSL•FO
RenderX

Table 3. Changes in psychological scale scores from baseline to 3 months: a paired-sample analysis.

Measure	Sample size, n	Baseline score, mean (SD)	Third-month score, mean (SD)	Change from baseline (%)	Mean difference (95% CI)	*P* value	*t* test (*df*)	Effect size (Cohen *d*)
PSS-10[a]	99	23.10 (3.05)	22.43 (3.10)	2.90	0.67 (.02 to 1.31)	.04	2.05 (98)	3.24
MBI-EE[b] subscale	95	35.31 (9.93)	33.64 (11.09)	4.70	1.66 (–0.22 to 3.55)	.08	1.75 (94)	9.25
MBI-DP[c] subscale	95	8.28 (6.65)	8.00 (6.32)	3.38	0.28 (–0.76 to 1.32)	.59	0.54 (94)	5.11
MBI-PA[d] subscale	95	34.95 (6.26)	34.86 (6.81)	0.23	0.08 (–1.04 to 1.20)	.88	0.15 (94)	5.50
PHQ-9[e]	93	11.02 (6.04)	9.77 (5.62)	11.34	1.25 (0.23 to 2.27)	.02	2.43 (92)	4.95
PHQ-9 item 3 (severity of sleep disturbance)	93	1.67 (1.00)	1.41 (0.96)	15.57	0.26 (0.04 to 0.47)	.02	2.39 (92)	1.04
PHQ-9 item 9 (suicidal ideation/thoughts of self-harm)	93	0.18 (0.51)	0.15 (0.53)	16.67	–0.03 (–0.07 to 0.13)	.52	0.65 (92)	0.48
GAD-7[f]	93	9.30 (5.85)	8.18 (5.29)	12.04	1.12 (0.20 to 2.04)	.02	2.41 (92)	4.48
BRS[g]	92	2.99 (0.96)	3.14 (0.94)	5.02	–0.15 (–0.30 to –0.00)	.04	–2.04 (91)	0.72

[a]PSS-10: Perceived Stress Scale-10.

[b]MBI-EE: Maslach Burnout Inventory-Emotional Exhaustion.

[c]MBI-DP: Maslach Burnout Inventory-Depersonalization.

[d]MBI-PA: Maslach Burnout Inventory-Personal Accomplishment.

[e]PHQ-9: Patient Health Questionnaire-9.

[f]GAD-7: Generalized Anxiety Disorder-7.

[g]BRS: Brief Resilience Scale.

Table 4. Changes in psychological scale scores from baseline to 6 months: a paired-sample analysis.

Measure	Sample size, n	Baseline score, mean (SD)	Sixth-month score, mean (SD)	Change from baseline (%)	Mean difference (95% CI)	*P* value	*t* test (*df*)	Effect size (Cohen *d*)
PSS-10[a]	113	23.33 (2.63)	21.96 (3.18)	5.87	1.37 (–0.76 to 1.98)	<.001	4.43 (112)	3.29
MBI-EE[b] subscale	110	34.09 (10.34)	31.44 (11.84)	7.77	2.65 (0.90 to 4.41)	<.001	3.00 (109)	9.29
MBI-DP[c] subscale	110	8.27 (5.98)	7.68 (5.82)	7.13	0.59 (–0.50 to 1.68)	.29	1.07 (109)	5.77
MBI-PA[d] subscale	109	34.72 (7.13)	35.04 (7.34)	0.92	–0.32 (–1.60 to 0.96)	.62	–0.50 (108)	6.75
PHQ-9[e]	105	11.61 (6.12)	8.70 (5.60)	24.98	2.90 (1.86 to 3.95)	<.001	5.52 (104)	5.39
PHQ-9 item 3 (severity of sleep disturbance)	105	1.70 (1.01)	1.42 (1.04)	17.06	0.29 (0.09 to 0.49)	.01	2.83 (104)	1.04
PHQ-9 item 9 (suicidal ideation/thoughts of self-harm)	105	0.28 (0.66)	0.16 (0.44)	39.29	0.11 (–0.01 to 0.24)	.06	1.87 (104)	0.63
GAD-7[f]	104	9.41 (5.32)	7.54 (4.92)	19.98	1.88 (0.88 to 2.87)	<.001	3.73 (103)	5.13
BRS[g]	103	3.00 (0.92)	3.17 (0.92)	6.00	–0.18 (–0.34 to –0.02)	.03	–2.19 (102)	0.83

[a]PSS-10: Perceived Stress Scale-10.

[b]MBI-EE: Maslach Burnout Inventory-Emotional Exhaustion.

[c]MBI-DP: Maslach Burnout Inventory-Depersonalization.

[d]MBI-PA: Maslach Burnout Inventory-Personal Accomplishment.

[e]PHQ-9: Patient Health Questionnaire-9.

[f]GAD-7: Generalized Anxiety Disorder-7.

[g]BRS: Brief Resilience Scale.

XSL•FO
RenderX

Table 5. McNemar test results indicating changes in the prevalence of psychological symptoms among Wellness4Teachers intervention group subscribers from baseline to 6 months.

Psychological symptoms	Prevalence		Change from baseline (%)	*P* value
	Baseline, n (%)	6 months, n (%)		
High stress	32 (14.8)	19 (8.8)	6.0	.03
High emotional exhaustion	166 (77.2)	151 (70.2)	7.0	.02
High depersonalization	52 (24.2)	48 (22.3)	1.9	.64
Low professional accomplishment	80 (37.4)	80 (37.4)	0.0	>.99
Likely major depressive disorder	116 (55.2)	86 (41.0)	14.2	>.001
Sleep disturbance	183 (87.1)	169 (80.5)	6.6	.04
Suicidal ideation/thoughts of self-harm	36 (17.1)	23 (11.0)	6.1	.04
Likely generalized anxiety disorder	96 (45.7)	78 (37.1)	8.6	.03
Low resilience	95 (45.5)	84 (40.2)	5.3	.18

Applying a Bonferroni-corrected significant *P* value of ≤.01, Table 5 indicates that the prevalence of likely MDD was significantly lower by 14.2% at 6 months compared to baseline for subscribers who completed the baseline survey and at least one follow-up survey. There was no significant difference in prevalence for the other variables at baseline and 6 months.

Naturalistic Controlled Study

The total number of teachers who completed the Wellness4Teachers survey either at baseline or 6 months between March 1 and June 30, 2023, was 318, with 152 (47.8%) in the CG and 166 (52.2%) in the IG, as illustrated in Table 6. Table S4 in Multimedia Appendix 4 illustrates the demographic characteristics of respondents in the CG and IG. However, it is worth noting that not all the participants who completed the follow-up surveys completed the demographic survey.

XSL•FO
RenderX

Table 6. Independent-sample t test comparing psychological outcomes between the intervention group (IG) and control group (CG) participants.

Psychological scale/subscale	CG score (n=152), mean (SD)	IG score (n=166), mean (SD)	Difference in mean scores (%)	Mean difference (95% CI)	*P* value	*t* test (*df*)	Effect size (Hedges *g*)
PSS-10[a]	23.32 (3.16)	21.92 (3.67)	–6.04	–1.41 (–0.65 to –2.17)	<.001	3.65 (316)	0.41
MBI-EE[b] subscale	34.59 (10.91)	31.36 (12.32)	–9.37	–3.24 (–0.66 to –5.81)	.01	2.47 (316)	0.28
MBI-DP[c] subscale	8.57 (6.54)	7.77 (5.72)	–9.45	–0.81 (–0.55 to –2.16)	.24	1.17 (316)	0.13
MBI-PA[d] subscale	34.28 (7.08)	35.55 (7.34)	+3.70	+1.27 (2.87 to 0.32)	.12	–1.57 (316)	–0.18
PHQ-9[e]	11.01 (5.97)	8.69 (5.47)	–21.16	–2.33 (–1.06 to –3.59)	<.001	3.63 (316)	0.41
PHQ-9 item 3 (severity of sleep problems)	1.57 (1.15)	1.45 (1.05)	–8.28	–0.13 (0.12 to 0.37)	.31	1.03 (316)	0.12
PHQ-9 item 9 (severity of suicidal thoughts)	0.20 (0.49)	0.13 (0.38)	–40.0	0.08 (0.02 to 0.17)	.12	1.57 (316)	0.18
GAD-7[f]	9.74 (5.80)	7.55 (5.12)	–22.48	–2.19 (–0.98 to –3.39)	<.001	3.57 (316)	0.40
BRS[g]	3.17 (0.88)	3.29 (0.91)	+3.79	+0.12 (0.32 to 0.07)	.22	–1.24 (316)	–0.14

[a]PSS-10: Perceived Stress Scale-10.

[b]MBI-EE: Maslach Burnout Inventory-Emotional Exhaustion.

[c]MBI-DP: Maslach Burnout Inventory-Depersonalization.

[d]MBI-PA: Maslach Burnout Inventory-Personal Accomplishment.

[e]PHQ-9: Patient Health Questionnaire-9.

[f]GAD-7: Generalized Anxiety Disorder-7.

[g]BRS: Brief Resilience Scale.

Applying a Bonferroni corrected significant *P* value of ≤.01, Table S4 in Multimedia Appendix 4 indicates that participants in the IG and CG were similar with respect to all the demographic and work-related characteristics, except for teaching experience, where the proportion of teachers with 5 years or less teaching experience was double for the CG compared with the IG, and the proportion of teachers with more than 20 years of teaching experience was double for the IG compared to the CG.

Table 6 shows the results of an independent-sample *t* test assessing differences in mean scores between the IG and CG on scales for measuring the various psychological symptoms.

Applying a Bonferroni-corrected significance level of *P*≤.01, Table 6 indicates that the mean scores on the PSS-10, MBI-EE subscale, the PHQ-9, and the GAD-7 were significantly lower in the IG by 6.04%, 9.37%, 21.16%, and 22.48% compared to the CG.

The impact of the Wellness4Teachers program on subscribers was further assessed by using the chi-square test to compare the prevalence of psychological symptoms among the IG and CG participants during the same designated period, as shown in Table 7.

XSL•FO
RenderX

Table 7. Differences in the prevalence of psychological symptoms among subscribers in the intervention group (IG) and control group (CG).

Psychological symptom	Prevalence, n (%)		Change from baseline (%)	Chi-square (*df*)	*P* value
	CG (baseline; n=152)	IG (6 months; n=166)			
High stress	26 (17.1)	17 (10.2)	–6.9	3.20 (1)	.07
High emotional exhaustion	120 (78.9)	115 (69.3)	–9.6	3.85 (1)	.05
High depersonalization	40 (26.3)	31 (18.7)	–7.6	2.67 (1)	.10
Low professional accomplishment	104 (68.4)	124 (74.7)	–6.3	1.54 (1)	.21
Likely major depressive disorder	90 (59.2)	64 (38.6)	–20.6	13.56 (1)	<.001
Sleep disturbance	116 (76.3)	131 (78.9)	2.6	0.31 (1)	.58
Suicidal ideation/thoughts of self-harm	26 (17.1)	18 (10.8)	–6.3	2.61 (1)	.11
Likely generalized anxiety disorder	79 (52.0)	51 (30.7)	–20.3	14.83 (1)	<.001
Low resilience	61 (40.1)	55 (34.9)	–5.2	1.58 (1)	.20

Applying a Bonferroni-corrected significant *P* value of ≤.01, Table 7 indicates that the prevalence of likely high depersonalization, MDD, and generalized anxiety disorder was significantly lower, by 7.6%, 20.6%, and 20.3%, respectively, for the IG compared to the CG. There was no statistically significant difference between the IG and the CG concerning other psychological symptoms.

To assess the impact of the Wellness4Teachers intervention on the reduced likelihood of psychological symptoms being present among subscribers while controlling for demographic and work-related variables, we entered the group type (IG vs CG) as a variable together with all demographic and work-related variables into nine separate logistic regression models. The key findings from the 9 models are summarized in Table 8.

Table 8. Odds of psychological symptoms among wellness4teachers subscribers in the control group compared to the intervention group after controlling for demographic and work-related variables.

Psychological variables of interest	Odds ratio (95% CI)	*P* value
High stress	0.42 (0.19-0.95)	.03
High emotional exhaustion	0.54 (0.29-1.03)	.06
High depersonalization	0.72 (0.36-1.43)	.35
Low professional accomplishment	0.70 (0.39-1.26)	.24
Likely major depressive disorder	0.47 (0.28-0.81)	.006
Sleep disturbance	0.72 (0.38-1.34)	.30
Suicidal ideation/thoughts of self-harm	0.50 (0.22-1.11)	.09
Likely generalized anxiety disorder	0.37 (0.21-0.64)	<.001
Low resilience	1.66 (0.95-2.90)	.74

Table 8 shows key results from separate logistic regression models, which assessed the independent contribution of the IG variable on the reduced presence of each of the 9 psychological symptoms in participants while controlling for sociodemographic variables.

The results in Table 8 demonstrate that the IG variable was a significant predictor (*P*≥.05) of high stress, likely MDD, and likely generalized anxiety disorder among the study participants when controlling for all demographic and work-related variables. However, the IG variable did not significantly predict the reduced presence of other psychological symptoms when controlling for demographic and work-related variables.

In terms of high stress, the model was not statistically significant (χ^2_{17}=18.53; n=237; *P*=.36), explaining only between 6.6% (Cox and Snell R^2) and 12.3% (Nagelkerke R^2) of the variance and correctly classified 87.1% (207/237) of all cases. However, controlling for all demographic and work-related variables, the "treatment group" contributed significantly to the model (Wald=4.38). The IG was 0.42 times less likely to present with high stress during the defined study period than CG (OR 0.42, 95% CI 0.19-0.95). This indicates that participants in the CG were 2.37 times more likely to meet the cutoff threshold for High Stress than participants in the IG, controlling for the demographic and work-related variables.

Concerning likely MDD, the model was not significant (χ^2_{17}=28.06; n=272; *P*=.04), explaining between 9.8% (Cox and Snell R^2) and 13.1% (Nagelkerke R^2) of the variance and correctly classifying 50.4% (137/272) of all cases. Controlling for all demographic and work-related variables, the "treatment group" contributed significantly to the model (Wald=7.64). The IG was 0.47 times less likely to present with likely MDD during

XSL•FO
RenderX

the defined study period than CG (OR 0.47, 95% CI 0.28-0.81). This indicates that participants in the CG were 2.13 times more likely to meet the cutoff threshold for likely MDD than participants in the IG, controlling for the demographic and work-related variables.

For likely GAD, the regression model was statistically significant (χ^2_{17}=41.76; n=272; P<.001), explaining between 14.2% (Cox and Snell R^2) and 19.1% (Nagelkerke R^2) of the variance and correctly classified 66.2% (180/272) of all cases. Controlling for all demographic and work-related variables, the "treatment group" contributed significantly to the model (Wald=12.52). The IG was 0.37 times less likely to present with likely generalized anxiety disorder during the defined study period than CG (OR 0.37, 95% CI 0.21-0.64). This indicates that participants in the CG were 2.70 times more likely to meet the cutoff threshold for likely generalized anxiety disorder than participants in the IG, controlling for the demographic and work-related variables.

Discussion

Principal Findings

Supporting teachers' mental health and well-being is vital, not only because of the immense contribution they make to students' lives but also because of the indirect impact their poor psychological well-being has on students' well-being and academic achievements [5,50]. This study aimed to evaluate whether the Wellness4Teachers daily supportive text messaging program could reduce the prevalence and severity of psychological issues while enhancing resilience among teachers in Alberta, Newfoundland and Labrador, and Nova Scotia. To achieve this objective, we conducted both longitudinal and naturalistic controlled studies comparing psychological outcomes for subscribers of Wellness4Teachers. Although the naturalistic controlled design enhances ecological validity, it does not support definitive causal inference because there was no randomization. While regression analyses adjusted for measured confounders, residual confounding from unmeasured factors remains possible and may still influence the findings. Consequently, the findings should be interpreted as evidence of real-world association rather than causality.

The longitudinal findings suggest that participation in Wellness4Teachers may be associated with improvements in psychological well-being over time. Consistency across longitudinal and sensitivity analyses provides some support for the observed patterns, although fewer outcomes remained significant in complete-case analyses after correcting for multiple comparisons. These differences highlight the importance of considering missing data, attrition, and multiple testing when interpreting changes in psychological outcomes.

Most Wellness4Teachers subscribers signed up for the program at the beginning of the school year, with follow-up evaluations occurring in the middle of the school year. Without any intervention, the natural trajectory would be for symptoms of anxiety and depression to increase in the teachers as the school year progressed. In the randomized controlled study that evaluated predictors of anxiety and depressive symptoms among teachers, it was reported that anxiety and depressive symptoms increased for all teachers over the school year [51]. Thus, the significant reductions in symptoms of stress, anxiety, depression, disturbed sleep, and suicidal thoughts, as per the analysis using the last follow-up observation carried forward, provide indications of improvements in psychological symptoms observed among teachers participating in the Wellness4Teachers program.

The contemporaneous comparison suggests that participation in Wellness4Teachers was associated with more favorable psychological outcomes across several indicators. Adjusted analyses further support these associations, although the naturalistic, nonrandomized design limits causal interpretation. The absence of significant differences in other outcomes also suggests that potential benefits may vary across psychological domains, warranting further investigation in controlled studies. The outcome of severity measures in this study aligns with those of the Text4Hope youth program, where the mean scores on the PSS-10, PHQ-9, and GAD-7 were significantly lower for the IG compared to the CG by 8.6%, 25.7%, and 23.6%, respectively [27]. The results of the naturalistic controlled study also align with the results of greater symptom reduction for depression in the IG compared to the CG observed in the Text4Hope study and 2 randomized controlled trials in Ireland and Canada [27,45]. In addition, the prevalence outcomes in this study are partially consistent with longitudinal study findings from the Text4Hope [52] and Text4PTSI programs [28].

Beyond statistical significance, the observed reductions in PHQ-9, GAD-7, and PSS-10 scores may have meaningful practical implications, as improvements in depression, anxiety, and perceived stress are likely to enhance teachers' emotional well-being, concentration, work engagement, interpersonal relationships, and overall professional functioning. This may potentially contribute to improved classroom effectiveness and job satisfaction, ultimately benefiting students in terms of their overall academic performance.

The consistency of the positive impacts of supportive text messaging interventions across diverse programs, populations, and contexts highlights the potential for the Wellness4Teachers intervention as an approach for improving mental health and promoting psychological well-being among teachers in Canada and internationally. Furthermore, given the high prevalence of stress, anxiety, and depression among teachers globally and the generally busy schedule and widespread geographical distribution of teachers [5], interventions that are accessible, easy to sign up for, and have no geographical barriers are needed to address these psychological symptoms and promote well-being among teachers. For example, time constraints and geographic location may be factors restraining some teachers from accessing human-delivered and in-person mindfulness-based interventions with or without yoga or CBT [19]. The probable association between participants in the Wellness4Teachers program and psychological symptom reduction in this study inspires hope, as the program does not require significant time commitments from teachers and has no geographic location barriers in addressing symptoms of stress, anxiety, and depression. Therefore, it allows widespread

XSL•FO
RenderX

adoption, contextualization, and implementation in high- and low-resource settings and in multiple jurisdictions to mitigate mental health symptoms among teachers. One study suggested that text-based therapy is 42.2% the cost of traditional services and offers much-reduced wait times [53]. The ResilienceNHope suite of programs, including the Wellness4Teachers program, may be a cost-effective intervention that can help close the psychological treatment and mental health literacy gaps among diverse populations worldwide [20]. The automation of the Wellness4Teachers program, coupled with computer-based delivery and limited human resource involvement, makes it a suitable population-level intervention that can be adopted by governments and education policy directors and managers to support tens of thousands of teachers simultaneously.

Notwithstanding the perceived positive impacts of the Wellness4Teachers program in mitigating symptoms of stress, anxiety, and depression among teachers, the results displayed for this study show that the intervention did not lead to statistically significant reductions in the severity or prevalence of the two burnout syndromes (depersonalization and personal accomplishment) or low resilience in subscribers in either the longitudinal study or the naturalistic study. This suggests that while Wellness4Teachers is a promising tool for addressing depression and anxiety symptoms among teachers, additional measures may be needed to address burnout syndromes and low resilience symptoms among teachers. However, mean scores for the CG and IG on the BRS were 3.17 and 3.29, respectively, which both fall within the score range for normal resilience. Practical measures such as reduced class sizes, enhanced support from school administration, reduced workload, support with the management of disruptive student behavior, and enhanced remuneration, which may address identified causes of increased burnout among teachers [5], need to be implemented alongside the Wellness4Teachers program.

Strengths and Limitations

The study has some limitations worth mentioning. First, the naturalistic study design used does not follow the conventional randomization of participants into the two study groups (IG and CG) at baseline. Nonetheless, the IG and CG participants were similar in all demographic and work-related characteristics, except for teaching experience. Furthermore, the logistic regression analysis used in our study controlled for any differences in sociodemographic and work-related characteristics while assessing the impact of the Wellness4Teachers program in reducing the prevalence of psychological symptoms in the naturalistic study. In addition, the comparison was deliberately structured so that outcomes for 6-week program completers were compared with those of new enrollees assessed at baseline during the same time period. Thus, both groups were situated within the same broader social, educational, and environmental context at the time of assessment. Consequently, time-varying influences such as seasonal effects, academic-calendar fluctuations, contemporaneous workplace pressures, policy changes, or other historical events would be expected to affect both groups concurrently rather than systematically favoring one group. The principal distinction between the groups was that one had received 6 weeks of the supportive text-messaging intervention, whereas the contemporaneous baseline group had not yet been exposed to the intervention.

Second, although outcome assessments for the intervention and control groups in the naturalistic study were obtained during the same calendar period, the groups were not observed over matched elapsed time intervals. The IG had received 6 months of supportive text messages before assessment, whereas the naturalistic CG was assessed at program entry before receiving the intervention. Contemporaneous assessment reduces potential academic-calendar time influences, and other environmental influences operating at the time outcomes were measured, but it does not control for maturation, regression to the mean, differential selection into the program over time, or other unmeasured factors associated with time since enrollment. Consequently, differences between the groups cannot be attributed solely to the intervention and should not be interpreted as causal treatment effects. The naturalistic comparison is therefore considered supportive and contextual to the primary prospective longitudinal findings rather than equivalent to a randomized or longitudinally matched controlled comparison.

Third, the relatively low follow-up survey completion rate and the LOCF approach used to analyze the longitudinal data assume stability of outcomes following the last available observation and may underestimate variability, particularly when participant outcomes continue to change after dropout. Accordingly, findings based on LOCF should be interpreted with appropriate caution, as this approach cannot eliminate potential bias arising from attrition or informative missingness. To address this issue, complete-case sensitivity analyses were conducted using paired-samples analyses at each follow-up time point. Although the findings were broadly consistent across approaches, this agreement does not establish robustness, as both methods may be biased by informative missingness. For example, selection bias may have occurred because data were collected through web-based questionnaires, potentially excluding teachers who lacked access to or were unable or unwilling to use smartphones or online platforms. Furthermore, the absence of statistically significant baseline differences between follow-up completers and noncompleters does not establish comparability or exclude attrition bias. Therefore, the findings should be interpreted cautiously. Fourth, the self-rated scales used to assess stress, anxiety, and depression symptomatology, although validated and standardized, are not diagnostic.

Notwithstanding these limitations, this study provides vital insights and evidence on the probable usefulness of the Wellness4Teachers program in supporting educators' mental well-being and alleviating their mental health burden.

Conclusions

The Wellness4Teachers program was associated with improvements in several psychological outcomes among teachers over 6 months. Findings from the longitudinal and exploratory naturalistic controlled analyses suggest that daily supportive text messaging may offer a promising approach to supporting teachers' psychological well-being. However, the observational, nonrandomized design limits causal inference, and residual confounding cannot be excluded. Further randomized or well-designed quasi-experimental studies are

XSL•FO
RenderX

needed to confirm these associations and assess the program's effectiveness, feasibility, and cost-effectiveness. Future research should also examine implementation within educational systems and potential effects on student well-being and academic outcomes, which were not assessed in this study.

Acknowledgments

The authors thank the Newfoundland and Labrador Teachers Union, Alberta Teachers Association, Edmonton Catholic School Board, and Nova Scotia Teachers Union for promoting the Wellness4Teachers program among their teachers. We are also grateful to all the teachers who participated in the program and completed the surveys.

Funding

The Alberta Mental Health Foundation funded this research.

Data Availability

Dataset supporting the conclusions of this study will be made available from the corresponding author upon reasonable request.

Generative AI Use

The authors declare that generative AI was not used in the preparation of this manuscript.

Authors' Contributions

Conceptualization: BA
Methodology: BA, VIOA
Formal analysis: BA, VIOA
Investigation: BA, YW, PB-M, VIOA
Data curation: BA, VIOA
Software: BA, YW, PB-M, VIOA
Resources: VIOA
Project administration: BA
Supervision: YW, PB-M
Funding acquisition: VIOA
Validation: BA, YW, PB-M, VIOA.
Writing–original draft: BA
Writing–review & editing: BA, YW, PB-M, VIOA

Conflicts of Interest

VIOA is the founder and principal investigator of the ResilienceNHope messaging suite of programs. VIOA and BA volunteer as board members of the Global Psychological eHealth Foundation, a not-for-profit organization that oversees the ResilienceNHope suite of programs. BA volunteers as the president and chief executive officer of the Global Psychological eHealth Foundation. VIOA and BA receive no financial or other compensation from the Global Psychological eHealth Foundation and have no financial conflicts of interest regarding this article. PB-M and YW declare no conflicts of interest.

Multimedia Appendix 1
Baseline demographic and psychological characteristics of participants who completed the baseline and at least one follow-up survey and the participants who completed only the baseline survey.
[DOCX File , 20 KB - humanfactors_v13i1e92142_app1.docx]

Multimedia Appendix 2
Comparison of baseline psychological scale score.
[DOCX File , 26 KB - humanfactors_v13i1e92142_app2.docx]

Multimedia Appendix 3
Sociodemographic and professional characteristics follow-up.
[DOCX File , 55 KB - humanfactors_v13i1e92142_app3.docx]

Multimedia Appendix 4
Distribution of demographic and work-related characteristics in the control group and intervention group.

XSL•FO
RenderX

[DOCX File , 17 KB - humanfactors_v13i1e92142_app4.docx]

References

1. Cox J. What is the role of a teacher? ThoughtCo. 2020. URL: https://www.thoughtco.com/what-is-the-role-of-a-teacher-2081511 [accessed 2025-05-12]
2. Chennoufi L, Ellouze F, Cherif W, Mersni M, M'rad MF. Stress and burnout among Tunisian teachers. Encephale 2012;38(6):480-487. [doi: 10.1016/j.encep.2011.12.012] [Medline: 23200614]
3. Hadi AA, Naing NN, Daud A, Nordin R, Sulong MR. Prevalence and factors associated with stress among secondary school teachers in Kota Bharu, Kelantan, Malaysia. Southeast Asian J Trop Med Public Health 2009;40(6):1359-1370. [Medline: 20578472]
4. Hooker S, Olson-Bullis B, Levin A, Ziegenfuss JY, Margolis KL, Rossom RC. Depression and anxiety among K-12 teachers in the United States: a systematic review. Sch Psychol 2025;40(4):445-461. [doi: 10.1037/spq0000666] [Medline: 39264662]
5. Agyapong B, Obuobi-Donkor G, Burback L, Wei Y. Stress, burnout, anxiety and depression among teachers: a scoping review. Int J Environ Res Public Health 2022;19(17):10706 [FREE Full text] [doi: 10.3390/ijerph191710706] [Medline: 36078422]
6. Parslow R, Jorm AF, Christensen H, Broom DH, Strazdins L, D' Souza RM. The impact of employee level and work stress on mental health and GP service use: an analysis of a sample of Australian government employees. BMC Public Health 2004;4:41 [FREE Full text] [doi: 10.1186/1471-2458-4-41] [Medline: 15456518]
7. Agyapong B, da Luz Dias R, Wei Y, Agyapong VIO. Burnout among elementary and high school teachers in three Canadian provinces: prevalence and predictors. Front Public Health 2024;12:1396461 [FREE Full text] [doi: 10.3389/fpubh.2024.1396461] [Medline: 38737860]
8. Arvidsson I, Leo U, Larsson A, Håkansson C, Persson R, Björk J. Burnout among school teachers: quantitative and qualitative results from a follow-up study in southern Sweden. BMC Public Health 2019;19(1):655 [FREE Full text] [doi: 10.1186/s12889-019-6972-1] [Medline: 31142318]
9. Maslach C, Schaufeli WB, Leiter MP. Job burnout. Annu Rev Psychol 2001;52:397-422. [doi: 10.1146/annurev.psych.52.1.397] [Medline: 11148311]
10. Maslach C, Leiter MP. The Truth About Burnout: How Organizations Cause Personal Stress and What to do About it. New Jersey: John Wiley & Sons; 2000.
11. Herman KC, Hickmon-Rosa J, Reinke WM. Empirically derived profiles of teacher stress, burnout, self-efficacy, and coping and associated student outcomes. J Posit Behav Interv 2017;20(2):90-100. [doi: 10.1177/1098300717732066]
12. Jeamjitvibool T, Duangchan C, Mousa A, Mahikul W. The association between resilience and psychological distress during the COVID-19 pandemic: a systematic review and meta-analysis. Int J Environ Res Public Health 2022;19(22):14854 [FREE Full text] [doi: 10.3390/ijerph192214854] [Medline: 36429573]
13. Kermott CA, Johnson RE, Sood R, Jenkins SM, Sood A. Is higher resilience predictive of lower stress and better mental health among corporate executives? PLoS One 2019;14(6):e0218092 [FREE Full text] [doi: 10.1371/journal.pone.0218092] [Medline: 31185049]
14. Rutter M. Implications of resilience concepts for scientific understanding. Ann N Y Acad Sci 2006;1094:1-12. [doi: 10.1196/annals.1376.002] [Medline: 17347337]
15. Southwick SM, Bonanno GA, Masten AS, Panter-Brick C, Yehuda R. Resilience definitions, theory, and challenges: interdisciplinary perspectives. Eur J Psychotraumatol 2014;5 [FREE Full text] [doi: 10.3402/ejpt.v5.25338] [Medline: 25317257]
16. Herrman H, Stewart DE, Diaz-Granados N, Berger EL, Jackson B, Yuen T. What is resilience? Can J Psychiatry 2011;56(5):258-265. [doi: 10.1177/070674371105600504] [Medline: 21586191]
17. Besse R, Howard K, Gonzalez S, Howard J. Major depressive disorder and public school teachers: evaluating occupational and health predictors and outcomes. J Appl Biobehav Res 2015;20(2):71-83. [doi: 10.1111/jabr.12043]
18. Schonfeld IS, Bianchi R. Burnout and depression: two entities or one? J Clin Psychol 2016;72(1):22-37. [doi: 10.1002/jclp.22229] [Medline: 26451877]
19. Agyapong B, Brett-MacLean P, Burback L, Agyapong VIO, Wei Y. Interventions to reduce stress and burnout among teachers: a scoping review. Int J Environ Res Public Health 2023;20(9):5625 [FREE Full text] [doi: 10.3390/ijerph20095625] [Medline: 37174145]
20. Agyapong B, Shalaby R, Wei Y, Agyapong VIO. Can ResilienceNhope, an evidence-based text and email messaging innovative suite of programs help to close the psychological treatment and mental health literacy gaps in college students? Front Public Health 2022;10:890131 [FREE Full text] [doi: 10.3389/fpubh.2022.890131] [Medline: 36388394]
21. Titov N, Dear BF, Ali S, Zou JB, Lorian CN, Johnston L, et al. Clinical and cost-effectiveness of therapist-guided internet-delivered cognitive behavior therapy for older adults with symptoms of depression: a randomized controlled trial. Behav Ther 2015;46(2):193-205. [doi: 10.1016/j.beth.2014.09.008] [Medline: 25645168]
22. Ezeonu NA, Hertelendy AJ, Adu MK, Kung JY, Itanyi IU, Dias RDL, et al. Mobile apps to support mental health response in natural disasters: scoping review. J Med Internet Res 2024;26:e49929 [FREE Full text] [doi: 10.2196/49929] [Medline: 38520699]

XSL•FO
RenderX

23. von der Embse N, Ryan SV, Gibbs T, Mankin A. Teacher stress interventions: a systematic review. Psychol Sch 2019;56(8):1328-1343. [doi: 10.1002/pits.22279]
24. Jennings PA. Impacts of the CARE for teachers program on teachers? social and emotional competence and classroom interaction. J Educ Psychol 2017;109(7):1010 [FREE Full text]
25. Ungar P, Schindler A, Polujanski S, Rotthoff T. Online programs to strengthen the mental health of medical students: a systematic review of the literature. Med Educ Online 2022;27(1):2082909 [FREE Full text] [doi: 10.1080/10872981.2022.2082909] [Medline: 35642839]
26. Noble JM, Vuong W, Surood S, Urichuk L, Greenshaw AJ, Agyapong VIO. Text4Support mobile-based programming for individuals accessing addictions and mental health services-retroactive program analysis at baseline, 12 weeks, and 6 months. Front Psychiatry 2021;12:640795 [FREE Full text] [doi: 10.3389/fpsyt.2021.640795] [Medline: 34122173]
27. Agyapong B, Shalaby R, Vuong W, Gusnowski A, Surood S, Greenshaw AJ, et al. Text4Hope effectiveness in reducing psychological symptoms among young adults in Canada: longitudinal and naturalistic controlled program evaluation. J Clin Med 2023;12(5):1942 [FREE Full text] [doi: 10.3390/jcm12051942] [Medline: 36902729]
28. Obuobi-Donkor G, Shalaby R, Eboreime E, Agyapong B, Phung N, Eyben S, et al. Text4PTSI: a promising supportive text messaging program to mitigate psychological symptoms in public safety personnel. Int J Environ Res Public Health 2023;20(5):4215 [FREE Full text] [doi: 10.3390/ijerph20054215] [Medline: 36901235]
29. Hong Q, Zhang X, Guo M, Wen Z, Tang Q, Zhou J, et al. Effectiveness of text message-delivered health behaviour intervention on HbA1c change in adults with type 2 diabetes mellitus: a systematic review and meta-analysis of randomised controlled trials. Health Psychol Rev 2026;20(2):386-419. [doi: 10.1080/17437199.2025.2595126] [Medline: 41368984]
30. Graham A, Amato MS, Cha S, Jacobs MA, Bottcher MM, Papandonatos GD. Effectiveness of a vaping cessation text message program among young adult e-cigarette users: a randomized clinical trial. JAMA Intern Med 2021;181(7):923-930 [FREE Full text] [doi: 10.1001/jamainternmed.2021.1793] [Medline: 33999133]
31. Agyapong B, Wei Y, da Luz Dias R, Agyapong VIO. Burnout and associated psychological problems among teachers and the impact of the wellness4Teachers supportive text messaging program: protocol for a cross-sectional and program evaluation study. JMIR Res Protoc 2022;11(7):e37934 [FREE Full text] [doi: 10.2196/37934] [Medline: 35834305]
32. Canada's population estimates: strong population growth in 2023. Statistics Canada. 2024. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/240327/dq240327c-eng.htm [accessed 2026-10-22]
33. Newfoundland and Labrador English School District Information. URL: https://tinyurl.com/3cnryyt2 [accessed 2023-06-20]
34. NLSchools. 2024. URL: https://www.nlschools.ca/about/districtoverview.jsp [accessed 2024-10-06]
35. Nova Scotia teachers recognized for making a difference. Nova Scotia Teachers Union. 2015. URL: https://nstu.ca/the-nstu/communications/the-media-room/education-news?c=nova-scotia-teachers-recognized-for-making-a-difference [accessed 2022-01-10]
36. The Alberta Teachers' Association. Teacher count: province only 38 teachers ahead. ATA Releases Teacher Count Statistics 2024;39(6):1 [FREE Full text]
37. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983;24(4):385-396. [doi: 10.2307/2136404]
38. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
39. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
40. Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The Brief Resilience Scale: assessing the ability to bounce back. Int J Behav Med 2008;15(3):194-200. [doi: 10.1080/10705500802222972] [Medline: 18696313]
41. Maslach C, Jackson SE, Leiter MP. Maslach Burnout Inventory. In: Evaluating Stress: A Book of Resources. Lanham, Maryland: Scarecrow Education; 1997:191-218.
42. Maslach C, Jackson SE, Schwab R. MBI: Educators Survey. California: Mind Garden Inc; 2022.
43. Agyapong VIO, Hrabok M, Vuong W, Shalaby R, Noble JM, Gusnowski A, et al. Changes in stress, anxiety, and depression levels of subscribers to a daily supportive text message program (Text4Hope) during the COVID-19 pandemic: cross-sectional survey study. JMIR Ment Health 2020;7(12):e22423 [FREE Full text] [doi: 10.2196/22423] [Medline: 33296330]
44. Agyapong VI, Juhás M, Mrklas K, Hrabok M, Omeje J, Gladue I, et al. Randomized controlled pilot trial of supportive text messaging for alcohol use disorder patients. J Subst Abuse Treat 2018;94:74-80. [doi: 10.1016/j.jsat.2018.08.014] [Medline: 30243421]
45. Agyapong VIO, Juhás M, Ohinmaa A, Omeje J, Mrklas K, Suen VYM, et al. Randomized controlled pilot trial of supportive text messages for patients with depression. BMC Psychiatry 2017;17(1):286 [FREE Full text] [doi: 10.1186/s12888-017-1448-2] [Medline: 28768493]
46. Sample size calculator. Survey Monkey. URL: https://www.surveymonkey.com/learn/research-and-analysis/sample-size-calculator/ [accessed 2026-10-22]
47. Dhand NK. Sample Size Calculator for Comparing Paired Proportions. 2014. URL: https://statulator.com/SampleSize/ss2PP.html [accessed 2026-10-22]

XSL•FO
RenderX

48. Kane SP. Sample Size Calculator. Determines the minimum number of subjects for adequate study power. ClinCalc. 2019. URL: https://clincalc.com/stats/samplesize.aspx [accessed 2026-10-22]
49. Release notes: IBM® SPSS® Statistics 28. IBM. 2022. URL: https://tinyurl.com/5bbwcdxn [accessed 2026-10-22]
50. Emeljanovas A, Sabaliauskas S, Mežienė B, Istomina N. The relationships between teachers' emotional health and stress coping. Front Psychol 2023;14:1276431 [FREE Full text] [doi: 10.3389/fpsyg.2023.1276431] [Medline: 38054175]
51. Peele M, Wolf S. Predictors of anxiety and depressive symptoms among teachers in Ghana: evidence from a randomized controlled trial. Soc Sci Med 2020;253:112957 [FREE Full text] [doi: 10.1016/j.socscimed.2020.112957] [Medline: 32251932]
52. Agyapong VI, Hrabok M, Shalaby R, Vuong W, Noble JM, Gusnowski A, et al. Text4Hope: receiving daily supportive text messages for 3 months during the COVID-19 pandemic reduces stress, anxiety, and depression. Disaster Med Public Health Prep 2022;16(4):1326-1330. [doi: 10.1017/dmp.2021.27] [Medline: 33551009]
53. Hull TD, Mahan K. A study of asynchronous mobile-enabled SMS text psychotherapy. Telemed J E Health 2017;23(3):240-247. [doi: 10.1089/tmj.2016.0114] [Medline: 27797646]

Abbreviations

BRS: Brief Resilience Scale
CBT: cognitive behavioral therapy
CG: control group
GAD-7: Generalized Anxiety Disorder-7
IG: intervention group
LOCF: last-observation-carried-forward
MBI-DP: Maslach Burnout Inventory-Depersonalization
MBI-EE: Maslach Burnout Inventory-Emotional Exhaustion
MBI-ES: Maslach Burnout Inventory-Educator Survey
MBI-PA: Maslach Burnout Inventory-Personal Accomplishment
MDD: major depressive disorder
OR: odds ratio
PHQ-9: Patient Health Questionnaire-9
PSS-10: Perceived Stress Scale-10

Edited by A Kushniruk, S Law; submitted 25.Jan.2026; peer-reviewed by I Muhammad; comments to author 23.Jul.2026; revised version received 19.Oct.2026; accepted 19.Oct.2026; published 05.Nov.2026.

Please cite as:
Agyapong B, Wei Y, Brett-MacLean P, Agyapong VIO
Wellness4Teachers Supportive Text Messaging Program Association With Observed Improvement in Psychological Symptoms Among Teachers in Canada: Single-Arm Prospective Longitudinal and Naturalistic Controlled Study
JMIR Hum Factors 2026;13:e92142
URL: https://humanfactors.jmir.org/2026/1/e92142
doi:10.2196/92142
PMID:42896217

©Belinda Agyapong, Yifeng Wei, Pamela Brett-MacLean, Vincent Israel Opoku Agyapong. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 05.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Objective Monitoring of Tablet Use–Related Optical Exposure and Its Association With Axial Length in Preschool Children: Cross-Sectional Intelligent Monitoring Study

Yidong Zhu[1*], MPH; Hao Chen[1*], PhD; Senlin Lin[2,3*], MSc; Hong Jiang[4*], PhD; Mingdao Zhang[5], MEng; Yi Sun[1], MPH; Chenshu Li[1], MPH; Yingnan Jia[1], PhD

[1]Preventive Medicine and Health Education Department, School of Public Health, Fudan University, 130 Dong An Road, Shanghai, China

[2]Shanghai Eye Diseases Prevention & Treatment Center, Shanghai, China

[3]Shanghai Eye Hospital, School of Medicine, Tongji University, Shanghai, China

[4]Department of Maternal, Child and Adolescent Health, School of Public Health, Fudan University, Shanghai, China

[5]Midea Group (China), Nanchang, Jiangxi, China

[*]these authors contributed equally

Corresponding Author:
Yingnan Jia, PhD
Preventive Medicine and Health Education Department, School of Public Health, Fudan University, 130 Dong An Road, Shanghai, China

Abstract

Background: In recent years, the global prevalence of myopia among children has continued to rise. The preschool years represent a critical period for visual development, and the widespread adoption of electronic screens among young children has brought increasing attention to pediatric visual health. However, the association between visual environmental exposures related to screen use—such as screen brightness and ambient illuminance—and the risk of myopia in preschool children has not been thoroughly investigated.

Objective: This monitoring study aimed to investigate the association between electronic screen brightness, ambient illuminance, and axial length in preschool children to provide evidence-based support for developing screen brightness usage recommendations for this population.

Methods: This cross-sectional monitoring study was conducted between March and July 2023 in Shanghai, China, involving 2 representative samples of kindergarten children aged 3 to 6 years. Each participant was provided with a tablet preinstalled with intelligent monitoring software, which continuously and objectively recorded real-time data on screen time and screen brightness over a consecutive 7-day period. In addition, comprehensive data collection encompassed standardized ophthalmic assessments, high-precision ambient illuminance measurements, simulated laboratory lighting evaluations, and parental questionnaires. Associations between ambient illuminance, screen brightness, and axial length were analyzed using multivariable linear regression and restricted cubic spline models.

Results: Of the 199 children included in the total sample, 124 (62.3%) were boys, and 75 (37.7%) were girls. After adjustment for demographic characteristics, parental myopia, and screen use behaviors, the median ambient illuminance during tablet use was significantly inversely associated with axial length (β=–0.13, 95% CI –0.22 to –0.04; P=.006). A nonlinear dose-response relationship was identified between median screen brightness and axial length ($P_{nonlinearity}$=.004), with axial elongation accelerating beyond approximately 27 cd/m² and peaking around 56 cd/m². Boys (P<.001) and greater height (P=.33) were also significantly associated with longer axial length.

Conclusions: Higher ambient illuminance during tablet use is associated with shorter axial length in preschoolers, whereas screen brightness exhibits a nonlinear effect on axial elongation. This study highlights the importance of optimizing both environmental lighting and device settings to protect visual health in young children, providing empirical support for guidelines on safe digital device use and ambient lighting conditions in early childhood.

(JMIR Hum Factors 2026;13:e79266) doi:10.2196/79266

KEYWORDS

screen brightness; ambient illuminance; axial length; preschool children; visual health

XSL•FO
RenderX

Introduction

The rising global prevalence of myopia among children and adolescents has become a significant public health challenge [1]. Early intervention is critical, as a younger age of onset is strongly associated with a higher risk of developing high myopia [2,3]. Axial length represents a key biomarker for evaluating myopia severity and predicting disease progression, with abnormal elongation being strongly associated with myopia advancement [4]. The preschool stage (age 3 - 5 y) is a critical period for refractive development. During this time, the characteristics of visual environmental exposure can profoundly influence the onset and progression of myopia.

In our digitalized society, electronic screens have become a dominant component of children's visual environment. A large-scale epidemiological study conducted across 38 countries on 6 continents revealed that 60% to 93% of children and adolescents are exposed to screens for more than 2 hours daily [5], with preschool children showing a trend of earlier initial screen exposure and higher usage frequency [6,7]. This underscores the urgent need to evaluate the potential impact of screen exposure on visual health and myopia risk in preschoolers. Population-based studies have preliminarily demonstrated the association between screen time and myopia. A systematic review and meta-analysis published in *The Lancet Digital Health* confirmed that the independent duration of smartphone and tablet use as well as the cumulative effect of screen time from computers were significantly associated with an increased risk of myopia [8]. A study from Japan further identified a relationship between prolonged screen time and abnormal axial length elongation in young children [9]. A cross-sectional study from the Netherlands revealed a significant positive correlation between the increased frequency of uninterrupted smartphone usage lasting over 20 minutes and increased scleral elasticity ratio as well as the anterior chamber depth-to-lens thickness ratio [10]. Although existing studies have investigated the association between screen time and visual development, research on the impact of screen brightness, a critical optical parameter, on visual health remains absent and warrants further exploration.

Modern mobile devices are commonly equipped with ambient light sensors that enable automatic screen brightness adjustment, a dynamic feature that may influence the visual system through two mechanisms: first, the adaptation of screen brightness to ambient light levels may affect visual function; and second, excessively high ambient light levels may induce dry eye symptoms, while abnormally elevated screen brightness could increase the burden of glare perception [11-13]. Evidence from animal models indicates that low illumination levels can induce myopia, whereas increased light exposure may activate retinal dopaminergic activity (primarily via the D2 receptor pathway), thereby regulating retinal gene expression and suppressing the biological signaling pathways that promote axial elongation [14]. Furthermore, ergonomic research has identified an inverted U-shaped relationship between display brightness and visual performance, fatigue, and discomfort, suggesting that both excessively low and high brightness levels compromise visual comfort and performance [15]. Benedetto et al [11] demonstrated that increased screen brightness significantly exacerbates visual fatigue. However, epidemiological evidence regarding the impact of screen brightness on eye axial elongation in children remains lacking.

Therefore, the primary aim of this study was to collect real-time objective exposure data through intelligent monitoring technology and then examine the association between electronic screen brightness, ambient illuminance, and axial length in preschool children. These findings may contribute to the development of evidence-based guidelines regarding screen use and myopia prevention in young children.

Methods

Study Design and Population

This cross-sectional monitoring study was conducted in two districts of Shanghai, China: Xuhui District (urban) and Pudong New Area (suburban), from March to July 2023. A random sampling method was used to select a kindergarten from each of the 2 districts. For each kindergarten, a cluster sampling method was used to select 5 classes, in which the average number of children in each class was 30, from the junior, middle, and senior grades. A total of 199 preschool children participated in the study, each of whom was provided with a tablet for a 1-week period. Before participation, parental consent was obtained during the eligibility screening phase. During the monitoring period, participants were instructed to maintain their typical screen usage habits, with the provided tablets being strictly limited to use by child participants only, thereby ensuring the integrity of exposure data collection. The study enrollment criteria were established as follows. Inclusion criteria were as follows: (1) preschool children currently registered in the kindergarten; (2) aged 3 to 6 years at the time of recruitment; and (3) whose parents or other caregivers have allowed them to use screen devices, such as televisions, mobile phones, tablets, and computers. Exclusion criteria were as follows: (1) presence of diagnosed organic ocular pathologies, (2) demonstrated inability to complete refractive error assessments, and (3) absence of legally authorized guardian consent.

Measures

Vision Screening

Axial length, corneal curvature, and refractive errors were measured using standardized ophthalmic instruments. Spherical and cylindrical refractive errors were assessed with the KR-800 auto-refractor (Topcon Corporation), while axial length and corneal curvature were measured using the IOL Master 700 (Carl Zeiss Meditec AG). To ensure measurement accuracy, examinations were conducted sequentially, starting with the right eye, followed by the left eye. If children demonstrated poor cooperation, measurements were repeated. All examiners received prior training, and instruments were regularly calibrated to maintain precision.

Screen Exposure Monitoring Technology

The *Healthy Screen Viewing for Children (HSVC)* application, developed in collaboration with Huawei Corporation, was installed on dedicated tablets provided to participants [16]. This

XSL•FO
RenderX

software automatically recorded screen usage data, including application type, screen brightness, and duration of use, at 1-minute intervals over a 7-day monitoring period. Equipped with built-in ambient light sensors, the tablets dynamically adjusted screen brightness according to real-time ambient lighting conditions. The HSVC application concurrently recorded the corresponding screen luminance values (in cd/m²) along with precise time stamps for each measurement. All collected data were securely transmitted to a cloud server for subsequent analysis.

Simulated Lighting Environment Setup

To convert the recorded screen brightness values into corresponding ambient illuminance levels for each child, we established a response model between screen brightness and horizontal illuminance under controlled laboratory conditions. This model allowed us to estimate the ambient illuminance during each child's tablet use based on the screen brightness data captured by the HSVC application. A controlled experiment was conducted in a professional lighting laboratory equipped with blackout conditions, where a response model between horizontal illuminance and screen brightness was constructed. The lighting environment was simulated using a warm-white light source with a correlated color temperature of 3000 K and a color rendering index (Ra) >80, to reflect typical nontask lighting conditions found in residential settings. The relative spectral power distribution of the light source is provided in Multimedia Appendix 1. This setup was designed to ensure both environmental realism and comfort.

The self-luminous device was operated under its default automatic brightness adjustment mode and fixed on a standard platform, with the screen center positioned at a height of 75 cm above the ground. A CL-500A spectroradiometer (KONICA MINOLTA) was used to record illuminance levels, with the sensor placed horizontally at the center of the desktop, approximately 30 cm in front of the screen. In accordance with the BS ISO/CIE 8995-1:2025 standard, horizontal illuminance was adopted as the primary photometric parameter due to its experimental stability and representativeness in evaluating screen-based visual environments. A total of 32 paired measurements of screen brightness and horizontal illuminance were collected by systematically adjusting ambient light intensity. A significant nonlinear relationship was observed between screen brightness and horizontal illuminance. As illustrated in Multimedia Appendix 2, the fitted quadratic regression model was as follows:

$$E=0.199x2-11.862x+177.930$$

where x denotes the screen brightness (unit: cd/m²) and E represents the corresponding horizontal illuminance (unit: lux). The model exhibited a high goodness of fit (R^2=0.988).

On the basis of the illuminance requirements specified in BS ISO/CIE 8995-1:2025 for *computer work only and young children* (300 lux) and *general classroom activities* (500 lux), the experimental data were classified into three ambient lighting categories: low illuminance (<300 lux), representing relaxed or recreational settings; moderate illuminance (300 - 500 lux), suitable for light visual tasks or display-based activities; and high illuminance (>500 lux), corresponding to functionally enhanced or brightly lit environments.

Covariates

A self-administered online questionnaire was distributed to parents to collect demographic information, including the child's age, sex, caregiver characteristics, only-child status, parental education levels, and household income. Additionally, the questionnaire collected relevant covariates based on existing literature, such as parental myopia status and children's daily behaviors (eg, duration of indoor and outdoor activities and sedentary time). Furthermore, the questionnaire assessed screen time across all media types, addressing the limitations of the HSVC in capturing data from television, personal computers, and other nontablet devices.

Statistical Analysis

Categorical variables were analyzed through frequency distributions (%), whereas continuous variables were presented according to their distribution. The normality of all continuous variables was assessed using the Shapiro-Wilk test. The Shapiro-Wilk test indicated that axial length and height were normally distributed (P>.05), whereas the other continuous variables deviated significantly from normality. Normally distributed data were presented as mean (SD), and nonnormally distributed data were presented as median (25th percentile and 75th percentile). Univariate analyses used independent *t* tests for binary comparisons (eg, sex differences) and 1-way ANOVA for multigroup comparisons (eg, age category). To evaluate monotonic relationships between variables, the Spearman rank correlation coefficient (ρ) was used. Spearman correlation coefficients quantified bivariate associations between axial length and continuous physiological or behavioral measures (eg, height, weight, and activity durations). Before model fitting, key regression assumptions were verified. The homoscedasticity assumption was evaluated using a Breusch-Pagan test, and the normality of residuals was formally assessed using the Shapiro-Wilk test; no substantial violations were detected (P>.05). Multivariable linear regression models adjusted for covariates (eg, age, sex, parental myopia, screen time, and activity levels) to assess independent effects of ambient illuminance and tablet brightness on axial length. Restricted cubic spline (RCS) models with 3 knots explored nonlinear dose-response relationships. Statistical significance was set at P<.05 (2 tailed), and analyses were performed using R software (version 4.3.2; R Foundation for Statistical Computing).

Ethical Considerations

Before the participants underwent screen exposure monitoring and questionnaire surveys, we informed their parents (legal guardians) of the entire process of participation in the surveys as well as the benefits and risks involved, and the legal guardians signed the informed consent. As a participation acknowledgment, children received art supply kits valued at US $15 per set. This study conformed to the ethical guidelines of the 1975 Declaration of Helsinki and was approved by the Ethics Committee for Medical Research at the School of Public Health, Fudan University (IRB#2023-11-1088). When extracting data from the Huawei Research Platform and the online

XSL•FO
RenderX

questionnaire library for analysis, all data were deidentified, and an anonymous study ID was used as an identifier for each participant.

Results

Participant Characteristics

Complete demographic and axial length data were available for all 199 participants. As presented in Table 1, 62.3% (n=124) of the participants were boys, and the majority (n=144, 72.4%) were aged between 4 years 9 months and 5 years 6 months. The participants' caregivers were predominantly their mothers (n=106, 53.3%), and more than half of the parents held a bachelor's degree (fathers: n=113, 56.8%; and mothers: n=113, 56.8%). The prevalence of maternal myopia was significantly higher (n=113, 56.8%) compared to paternal myopia (n=108, 54.3%), and maternal myopia was found to be significantly associated with a longer axial length (*P*=.01).

XSL•FO
RenderX

Table . Demographic characteristics and comparison of axial length between groups (N=199).

Variable	Values, n (%)	Axial length (mm), mean (SD)	*T* test/*F* test[a] (*df*)	*P* value[b]
Sex			6.184 (197)	<.001
Boy	124 (62.3)	22.68 (0.653)		
Girl	75 (37.7)	22.07 (0.717)		
Age			3.163 (2)	.04
4 y 9 mo to 5 y 6 mo	144 (72.4)	22.52 (0.747)		
4 y 3 mo to 4 y 9 mo	38 (19.0)	22.32 (0.682)		
3 y 3 mo to 4 y 3 mo	17 (8.5)	22.11 (0.692)		
Caregiver			1.254 (3)	.29
Mother only	106 (53.3)	22.52 (0.789)		
Father only	11 (5.5)	22.58 (0.749)		
Grandparents	78 (39.2)	22.33 (0.662)		
Other family members	4 (2.0)	22.67 (0.664)		
Education of father			1.571 (2)	.21
Secondary school or less	22 (11.0)	22.30 (0.755)		
Senior high school	64 (32.2)	22.36 (0.768)		
Bachelor’s	113 (56.8)	22.53 (0.715)		
Education of mother			1.065 (2)	.35
Secondary school or less	32 (16.1)	22.27 (0.742)		
Senior high school	54 (27.1)	22.48 (0.812)		
Bachelor’s	113 (56.8)	22.48 (0.700)		
Total annual household income (CNY)			–1.503 (197)	.13
<200,000 (US $27,800)	84 (42.2)	22.36 (0.719)		
≥200,000 (US $27,800)	115 (57.8)	22.52 (0.749)		
Only-child status			–0.389 (197)	.70
Yes	101 (50.8)	22.43 (0.714)		
No	98 (49.2)	22.47 (0.766)		
Current marital status of parents			–1.697 (197)	.09
Married	192 (96.5)	22.43 (0.734)		
Single/long-term separation	7 (3.5)	22.91 (0.767)		
Whether the father is nearsighted			–0.498 (197)	.62
No	91 (45.7)	22.42 (0.735)		
Yes	108 (54.3)	22.47 (0.744)		
Whether the mother is nearsighted			–2.470 (197)	.01
No	86 (43.2)	22.30 (0.750)		

XSL•FO
RenderX

Variable	Values, n (%)	Axial length (mm), mean (SD)	*T* test/*F* test[a] (*df*)	*P* value[b]
Yes	113 (56.8)	22.56 (0.713)		
Whether the child was ill in the past year			–1.167 (197)	.24
Yes	173 (86.9)	22.42 (0.740)		
No	26 (13.1)	22.61 (0.721)		

[a]*T* tests were used for binary comparisons, and the *F* statistic from a 1-way ANOVA was used for multigroup comparisons.

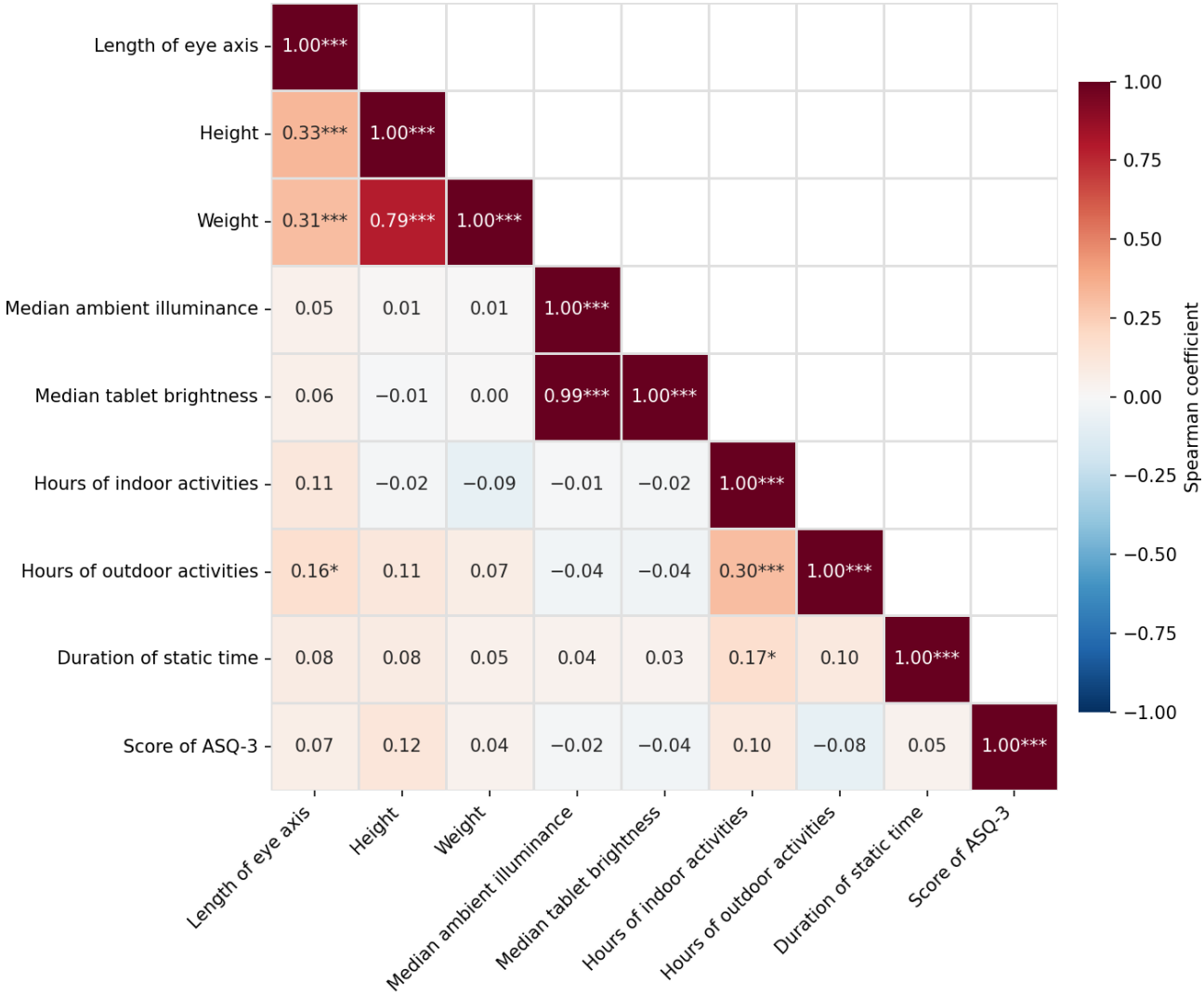
[b]*P* values were derived from independent *t* tests (for binary variables such as sex) or 1-way ANOVA (for multicategory variables such as age group). *P*<.05 was considered statistically significant.

Univariate analyses revealed significant differences in axial length based on sex, with boys having a greater mean axial length (22.68, SD 0.653 mm) compared to girls (22.07, SD 0.717 mm; *P*<.001). Additionally, significant differences in axial length were observed across different age groups (*P*=.04). However, no significant associations were found between axial length and variables, such as caregiver type, parental education level, household income, or only-child status (*P*>.05).

Correlation Analyses

The Spearman rank correlation analysis revealed several significant associations between certain demographic and behavioral factors and axial length. Specifically, height (ρ=0.33; *P*<.001), weight (ρ=0.31; *P*<.001), and outdoor activity duration (ρ=0.16; *P*=.03), all demonstrated statistically significant positive correlations with axial length (Figure 1). Notably, these relationships persisted even after adjusting for potential confounders, including parental myopia status and digital device exposure parameters.

Figure 1. Spearman rank correlation heatmap of key variables and axial length (N=199). Note: **P*<.05, ***P*<.01, ****P*<.001. ASQ-3: Ages and Stages Questionnaires, Third Edition.



XSL•FO
RenderX

Associations Between Ambient Illuminance or Tablet Brightness and Axial Length

As presented in Tables 2 and 3, multivariable linear regression demonstrated that higher median ambient illuminance during tablet use was inversely associated with axial length (β=–0.13, 95% CI –0.22 to –0.04; *P*=.006), after adjusting for sex, age, height, parental myopia, total screen time, and behavioral factors (Table 2). Specifically, the median ambient illuminance was 134.77 lux (P25=67.74 lux and P75=221.65 lux). Boy sex (*P*<.001) and greater height (*P*=.33) also significantly predicted axial length. However, there was no significant association between tablet brightness and axial length (Table 3). Although the median tablet brightness was 55 cd/m^2 (P25=47 cd/m^2 and P75=63.25 cd/m^2), it did not exhibit a linear association with axial length (Table 3).

Table . Associations between ambient illuminance and axial length (N=199)[a].

Variables	β[b] (95% CI)	*P* value
Height (cm)	0.16 (0.01 to 0.31)	.03
Weight (kg)	0.05 (–0.08 to 0.18)	.46
Sex		
Boy	Reference	—[c]
Girl	–0.59 (–0.79 to –0.40)	<.001
Age		
4 y 9 mo to 5 y 6 mo	Reference	—
4 y 3 mo to 4 y 9 mo	–0.01 (–0.28 to 0.25)	.92
3 y 3 mo to 4 y 3 mo	–0.12 (–0.51 to 0.26)	.52
Whether the mother is nearsighted		
No	Reference	—
Yes	0.12 (–0.07 to 0.31)	.20
Whether the father is nearsighted		
No	Reference	—
Yes	0.09 (–0.09 to 0.28)	.33
Total screen time used (ks)	0.03 (–0.06 to 0.12)	.56
Median ambient illuminance (lux)	–0.13 (–0.22 to –0.04)	.006
Hours of indoor activities (min)	0.05 (–0.05 to 0.14)	.31
Hours of outdoor activities (min)	0.05 (–0.04 to 0.14)	.26
Duration of static time (min)	–0.01 (–0.10 to 0.08)	.85
Score of ASQ-3[d] (pts)	0.07 (–0.02 to 0.16)	.13

[a]Results are derived from multivariable linear regression models. The dependent variable is axial length (mm).

[b]β represents the regression coefficient.

[c]Not applicable.

[d]ASQ-3: Ages & Stages Questionnaires, Third Edition.

XSL•FO
RenderX

Table . Associations between tablet brightness and axial length (N=199)[a].

Variables	β[b] (95% CI)	*P* value
Height (cm)	0.17 (0.02 to 0.32)	.02
Weight (kg)	0.04 (–0.09 to 0.17)	.51
Sex		
Boy	Reference	—
Girl	–0.58 (–0.78 to –0.39)	<.001
Age		
4 y 9 mo to 5 y 6 mo	Reference	—
4 y 3 mo to 4 y 9 mo	–0.01 (–0.28 to 0.26)	.94
3 y 3 mo to 4 y 3 mo	–0.10 (–0.49 to 0.29)	.60
Whether the mother is nearsighted		
No	Reference	—
Yes	0.12 (–0.07 to 0.31)	.22
Whether the father is nearsighted		
No	Reference	—
Yes	0.10 (–0.09 to 0.29)	.32
Total screen time used (ks)	0.03 (–0.06 to 0.12)	.51
Median tablet brightness (cd/m²)	–0.08 (–0.17 to 0.01)	.09
Hours of indoor activities (min)	0.05 (–0.05 to 0.14)	.35
Hours of outdoor activities (min)	0.06 (–0.04 to 0.15)	.24
Duration of static time (min)	–0.02 (–0.11 to 0.07)	.67
Score of ASQ-3[c] (pts)	0.07 (–0.03 to 0.16)	.17

[a]Results are derived from multivariable linear regression models. The dependent variable is axial length (mm).

[b]β represents the regression coefficient.

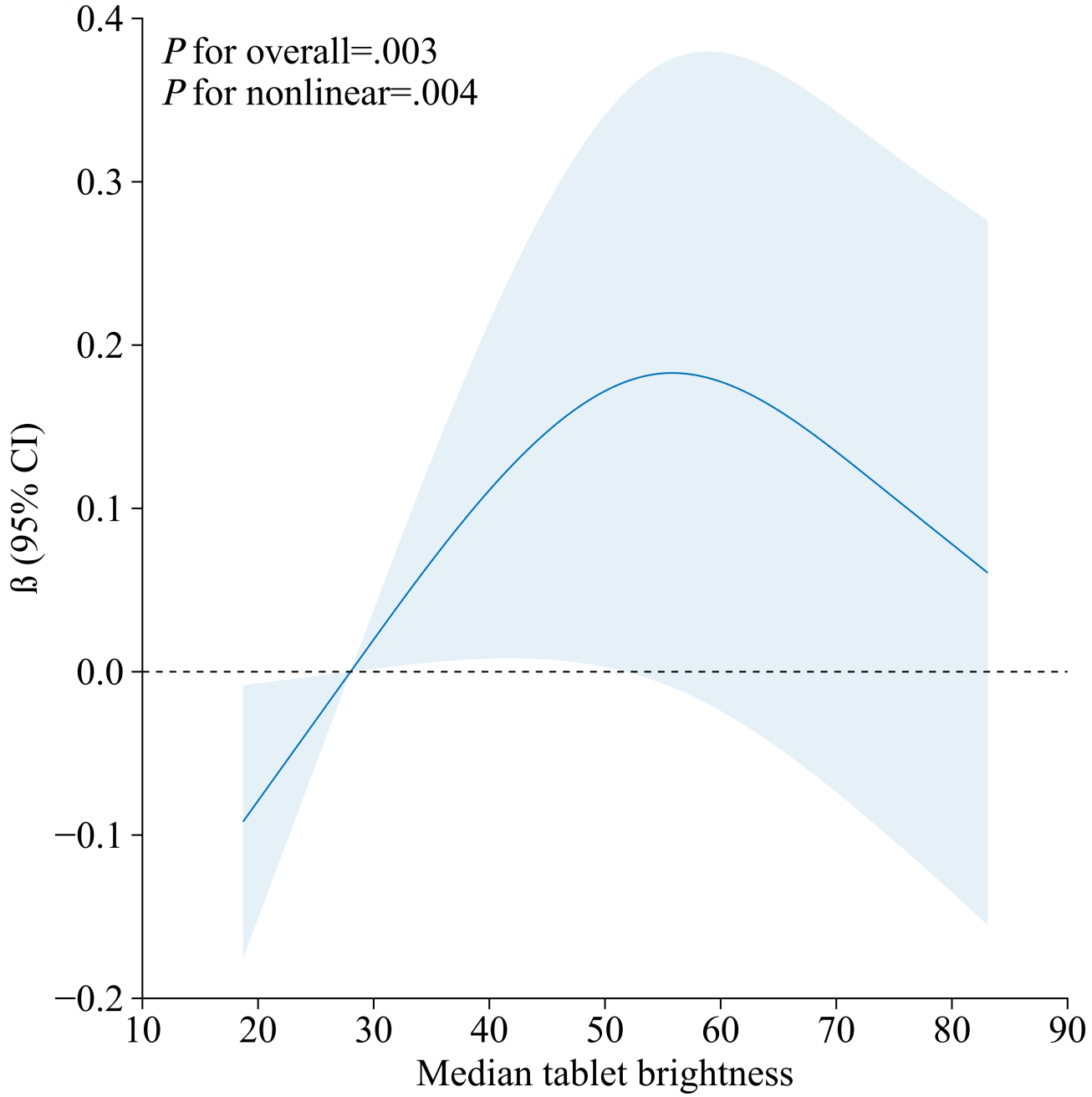
[c]ASQ-3: Ages and Stages Questionnaires, Third Edition.

Dose-Response Relationships Between Ambient Illuminance or Tablet Brightness and Axial Length

The dose-response relationships of tablet brightness and ambient illuminance with axial length were examined using RCS, as shown in Figures 2 and 3. Multivariable linear regression models incorporating RCS were fitted with 3 knots placed at the 10th, 50th, and 90th percentiles of the exposure distribution [17]. All models were adjusted for sociodemographic factors (eg, boy or girl, age, height, weight, and parental myopia), behavioral factors (indoor and outdoor activity duration, and sedentary time), and developmental characteristics (score from Ages and Stages Questionnaires, Third Edition). A significant nonlinear association was identified between median tablet brightness and axial length ($P_{\text{nonlinearity}}$=.004). The coefficient β represents the marginal change in axial length (in mm) per unit increase in exposure. The fitted curve indicated a negative slope at brightness levels below 27 cd/m², suggesting a reduction in the rate of axial elongation. This was followed by a transitional phase (27 - 56 cd/m²) where the curve became positive and increased, peaking at approximately 56 cd/m². Beyond this point, the fitted curve gradually decreased. In contrast, no significant nonlinear relationship was found between ambient illuminance and axial length ($P_{\text{nonlinearity}}$=.17), with the curve remaining approximately flat across the measured range (Figure 3).

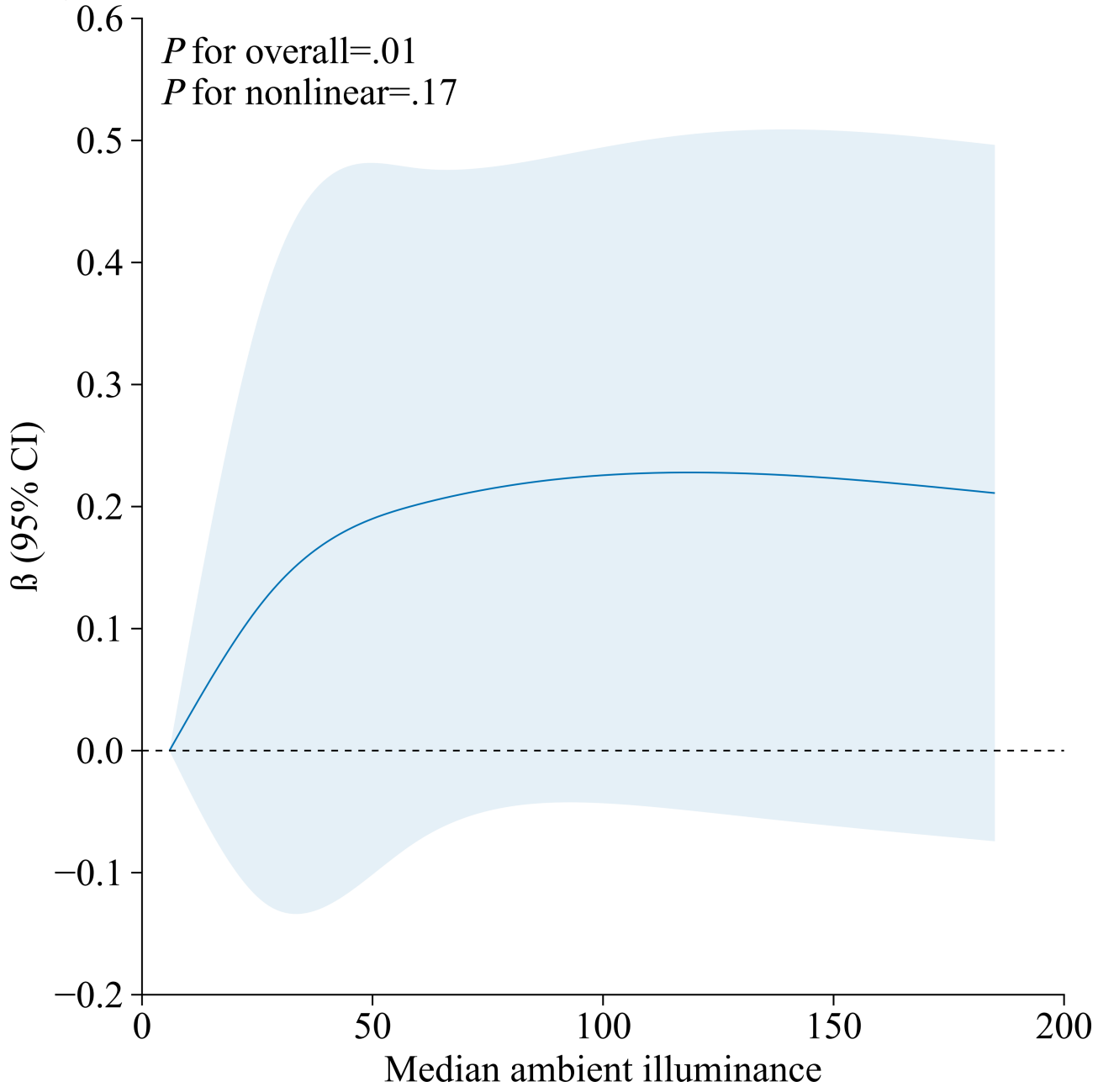
XSL•FO
RenderX

Figure 2. Dose-response relationship between tablet brightness and axial length (N=199). The solid line represents the estimated adjusted regression coefficient β, and the blue area indicates the 95% CI. Knots were placed at the 10th, 50th, and 90th percentiles of the median tablet brightness.



XSL•FO
RenderX

Figure 3. Dose-response relationship between ambient illuminance and axial length (N=199). The solid line represents the estimated adjusted regression coefficient β, and the blue area indicates the 95% CI. Knots were placed at the 10th, 50th, and 90th percentiles of the median ambient illuminance.



Discussion

Principal Findings

On the basis of continuous 7-day real-time monitoring data collected via the HSVC application from 199 preschool children aged 3 to 6 years in 2 kindergartens in Shanghai, this study investigated the associations among screen luminance, ambient illuminance during device use, and axial length in early childhood, addressing a critical gap in the existing literature regarding the impact of dynamic screen optical parameters on visual development in children. Our findings reveal a significant inverse association between higher ambient illuminance during tablet use and shorter axial length, as well as a nonlinear dose-response relationship between tablet luminance and axial length. These results elucidate the link between digital device optical exposure and axial elongation, providing robust empirical evidence to inform public health recommendations on screen use practices and ambient lighting standards in preschool children.

Relationship Between Ambient Illuminance and Axial Length

Our study identified a significant inverse association between ambient illuminance during tablet use and axial length in preschool children, aligning with findings from animal models [14]. Experimental studies in chicks and rhesus monkeys have established that light intensity modulates myopic progression [18-21]. Multiple investigations further demonstrate that increased light exposure may exert its effect by activating retinal dopaminergic activity—primarily through D2 receptor pathways—thereby regulating retinal gene expression and suppressing pro-axial elongation signaling [22-25]. The retinal

XSL•FO
RenderX

dopamine system is thus considered a potent inhibitor of ocular growth [26].

The present results align with several domestic studies investigating the role of light environment in myopia prevention. For instance, a study on elevated classroom lighting found that increasing illuminance significantly slowed axial elongation and reduced the incidence of myopia among school-aged children [27]. Similarly, a cluster randomized trial in Shanghai using objective monitoring showed that increased outdoor time and sustained high-intensity light exposure delayed axial elongation [28]. Together, these studies support a protective role of light exposure against axial growth. Our study extends this evidence from "macroenvironmental" settings to "microenvironments during screen use," refining the understanding of light exposure effects.

Complementing this mechanistic context, it is important to recognize that not all studies of indoor optical environments have detected an effect on refractive development. For instance, a school-based comparison of high-daylight versus low-daylight classrooms reported no overall difference in 6-month changes in refractive error or axial length, with effects limited only to a baseline subgroup characterized by short axial length [29]. Similarly, an independent population-based analysis found no association between exposure to ambient nighttime lighting during infancy and later myopia [30]. These inconsistent outcomes may be attributable to methodological and contextual differences. The former study used classroom-level daylight factors—a distal and nonspecific metric—which may not accurately capture retinal light exposure during near-device use. Furthermore, the study was conducted in a specific ethnic and educational setting (Korean schoolchildren), where genetic predispositions and academic pressures may modulate the effects of light exposure. The latter study examined nighttime lighting exposure in infancy. This differs fundamentally from daytime ambient light during visual tasks in its spectral composition, exposure timing, and biological mechanisms. Furthermore, the cohort primarily comprised White participants, who have a distinct genetic background from Asian populations. These methodological and demographic differences may collectively account for the heterogeneity observed regarding the association between light exposure and myopia.

It is noteworthy that typical indoor illuminance experienced by humans rarely exceeds 1000 lux and is usually much lower, often within the range of 100 to 500 lux [31]. According to standards set by the International Organization for Standardization and the International Commission on Illumination, as well as China's national guidelines (GB 7793 - 2010) on classroom lighting, the recommended threshold for children's indoor ambient illuminance is 300 lux [32,33]. In our findings, the median ambient illuminance during tablet use for 174 children was below 300 lux, with an overall median of 134.77 (IQR 67.74-221.65) lux, indicating that the ambient lighting conditions when participants used electronic devices were generally lower than internationally recommended standards. Nevertheless, we still observed a robust association between ambient illuminance and axial length, underscoring the critical role of environmental lighting in ocular development. Suboptimal lighting may fail to provide adequate visual comfort and could contribute to visual fatigue or other adverse effects on visual development. The widespread prevalence of insufficient ambient lighting during device use is concerning, as it represents a potentially modifiable environmental risk factor for myopia in children. Given the increasing reliance on digital devices in early childhood, our findings highlight the urgent need for evidence-based guidelines on ambient lighting for children to mitigate potential risks to visual health and development.

However, our results diverge from some previous reports, suggesting that ambient illuminance has no significant effect on visual fatigue [34]. This discrepancy suggests that the mechanisms underlying short-term visual discomfort and long-term regulation of axial growth may differ and warrant separate investigation. While short-term visual fatigue is likely associated with transient visual stimuli and ocular muscle strain, long-term axial elongation is influenced by a complex interplay of factors, including light-mediated regulation of ocular signaling pathways. From the perspective of long-term axial development, our study provides novel evidence for the impact of ambient illuminance on children's visual health.

Relationship Between Tablet Brightness and Axial Length

Using RCS with knots at the 10th, 50th, and 90th percentiles of the exposure distribution, we identified a significant nonlinear association between median tablet brightness and axial length after adjustment for sociodemographic, behavioral, and developmental covariates ($P_{nonlinearity}$=.004). At lower levels of tablet brightness (with a median tablet brightness <27 cd/m²), increases in brightness are accompanied by a deceleration in axial length growth ($\beta<0$), suggesting that very dim screens may accelerate eye growth, whereas moderate brightness increases could help mitigate excessively rapid axial elongation. When tablet brightness exceeds a certain threshold (median tablet brightness >27 cd/m²), further increases in brightness result in an acceleration of axial elongation, with excessively high screen brightness stimulating abnormal axial growth. When tablet brightness reaches a critical threshold (median tablet brightness=56 cd/m²), the rate of axial elongation peaks. Excessive brightness can cause glare or strong light stimuli, disrupting normal physiological eye development and accelerating axial elongation, which may worsen myopia progression [13]. In this study, the median tablet brightness was 56 cd/m² (P25=47 cd/m² and P75=63.25 cd/m²), falling within the critical range that may influence ocular growth.

These findings align with the "parabolic or inverted U-shaped dose-response curve" described in previous ergonomic studies regarding screen brightness, visual performance, and discomfort symptoms [15]. At very low brightness, children may reduce viewing distance to maintain visibility, increasing accommodative demand and potentially stimulating axial elongation. Moderately increased brightness may improve legibility and allow for more comfortable viewing distances, corresponding to the observed negative marginal effect below 27 cd/m² [35,36]. Within the midrange, heightened brightness during sustained near-work may provide stronger visual signals associated with axial elongation, resulting in a positive marginal

XSL•FO
RenderX

effect that peaks near 56 cd/m² [35,37]. At even higher brightness, discomfort and incipient glare may prompt self-regulatory behaviors—such as microbreaks, blinking, or adjustment of posture or gaze—that could attenuate the marginal effect, consistent with the decline in the fitted curve at high brightness levels [37-39]. This nonmonotonic pattern aligns with ergonomic literature in which display brightness often exhibits nonlinear relationships with both performance and discomfort [15,40]. To our knowledge, this is the first study to quantitatively characterize a nonlinear dose-response relationship between objectively measured tablet brightness and axial length in children, addressing a significant evidence gap in this field. This finding is also consistent with the "light dose-axial plasticity" effect reported in domestic studies on optical interventions. For instance, a randomized clinical trial in 10 primary schools in Shanghai demonstrated that repeated low-level red light therapy reduced the annual axial elongation rate in children from 0.47 mm to 0.30 mm over 12 months, alongside a significant decrease in new-onset myopia [41]. Although repeated low-level red light represents a therapeutic intervention fundamentally distinct from daily screen-based illuminance, its outcomes clearly indicate that the mode and dosage of light stimulation can structurally influence axial elongation. This lends biological plausibility to the nonlinear association between screen brightness and axial length observed under naturalistic conditions in our study and further supports the notion that optical parameters play a modifiable and nonlinear role in axial eye growth.

More importantly, the study identifies a relatively "safe" brightness range for children's use of mobile devices, within which axial elongation remains stable and the impact on myopia progression is minimal. To translate these key findings into practical application, device manufacturers should implement child-specific brightness-limiting features that cap maximum screen luminance below 56 cd/m², incorporating ambient-adaptive safeguards to prevent exceedance of this threshold and thus reduce the risk of excessive axial elongation. Caregivers should be advised to maintain screen brightness within the empirically derived safe range identified in this study and to encourage regular, timed breaks during prolonged use to minimize potential adverse effects on visual development. These results provide robust scientific evidence to support public health guidelines for optimal screen brightness settings in both educational and household environments.

Limitations

Several limitations of this study should be acknowledged. First, the cross-sectional nature of this study precludes any causal inferences. Future prospective studies are needed to determine whether reduced ambient illuminance directly accelerates axial elongation and to clarify causal relationships among relevant factors. Second, our sample was limited to children in Shanghai, potentially restricting the generalizability of our findings to populations with different screen usage habits. Future research should extend to larger and more diverse cohorts across various regions and populations to enhance external validity. Additionally, although we attempted to control for covariates, such as the duration of outdoor activities and unmeasured confounders—including genetic predisposition and dietary factors—may partially account for the observed associations. Future research should address these factors to comprehensively elucidate the determinants affecting axial eye development in children. Finally, as cycloplegic refraction was not performed in this study, the accuracy of refractive error measurements was limited. Therefore, axial length was selected as the primary outcome. In future research, we will refine the methodology by incorporating cycloplegic refraction to enable a more comprehensive evaluation of the effects of ambient illuminance on children's visual health.

Conclusions

This monitoring study demonstrates a significant association between both ambient illuminance and tablet screen brightness with ocular axial length in preschool children, with a nonlinear dose-response relationship observed for screen brightness. These findings underscore the importance of optimizing both environmental lighting and device settings to safeguard early childhood visual development and inform evidence-based guidelines for digital device use in preschool children.

Acknowledgments

The authors would like to extend our sincere appreciation to Huawei for developing the intelligent monitoring software used to assess children's screen exposure in this study. The authors are also deeply grateful to the kindergarten teachers from Pudong New District and Xuhui District for their invaluable assistance, as well as to all the participants and their families for their contributions to this research.

Funding

This work was supported by the National Natural Science Foundation of China (grant 82304260); Key Discipline and Project of High-Quality Development of Public Health of School of Public Health, Fudan University–Jiading District Health Commission (grant GWGZLXK-2023 04); and Shanghai Municipal Health Commission (grant GWV-10.2-YQ23). The funders had no involvement in the study design, data collection, analysis, interpretation, or the writing of the manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

XSL•FO
RenderX

Authors' Contributions

Conceptualization: YJ (lead), HC (equal), HJ (supporting), SL (supporting), YZ (supporting)
Formal analysis: HC (lead), YZ (equal), MZ (supporting)
Funding acquisition: YJ (lead), HC (equal), HJ (supporting)
Investigation: YZ (lead), HC (equal), YS (supporting), CL (supporting), MZ (supporting)
Methodology: YJ (lead), HC (equal), HJ (supporting), SL (supporting), YZ (supporting)
Project administration: YJ (lead), MZ (equal), HJ (supporting)
Resources: YJ (lead), MZ (equal), HJ (supporting)
Supervision: YJ (lead), HJ (equal), SL (supporting), HC (supporting)
Writing – original draft: YZ (lead), MZ (supporting)
Writing – review & editing: YJ (lead), HC (equal), HJ (supporting), SL (supporting), YZ (supporting), MZ (supporting), YS (supporting), CL (supporting)
HJ is the senior author for this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Relative spectral power distribution of the light source (correlated color temperature=3000 K).
[DOCX File, 84 KB - humanfactors_v13i1e79266_app1.docx]

Multimedia Appendix 2
Nonlinear response relationship between screen luminance and horizontal illuminance of the self-luminous device.
[DOCX File, 93 KB - humanfactors_v13i1e79266_app2.docx]

References

1. Dolgin E. The myopia boom. Nature 2015 Mar 19;519(7543):276-278. [doi: 10.1038/519276a] [Medline: 25788077]
2. Chua SYL, Sabanayagam C, Cheung YB, et al. Age of onset of myopia predicts risk of high myopia in later childhood in myopic Singapore children. Ophthalmic Physiologic Optic 2016 Jul;36(4):388-394. [doi: 10.1111/opo.12305]
3. Pärssinen O, Kauppinen M. Risk factors for high myopia: a 22-year follow-up study from childhood to adulthood. Acta Ophthalmol 2019 Aug;97(5):510-518. [doi: 10.1111/aos.13964] [Medline: 30460746]
4. Haarman AEG, Enthoven CA, Tideman JWL, Tedja MS, Verhoeven VJM, Klaver CCW. The complications of myopia: a review and meta-analysis. Invest Ophthalmol Vis Sci 2020 Apr 9;61(4):49. [doi: 10.1167/iovs.61.4.49] [Medline: 32347918]
5. Tremblay MS, Barnes JD, González SA, et al. Global matrix 2.0: report card grades on the physical activity of children and youth comparing 38 countries. J Phys Act Health 2016 Nov;13(11 Suppl 2):S343-S366. [doi: 10.1123/jpah.2016-0594] [Medline: 27848745]
6. Rideout V, Robb MB. The common sense census: media use by kids age zero to eight, 2020. : Common Sense Media; 2020 URL: https://www.commonsensemedia.org/sites/default/files/research/report/2020_zero_to_eight_census_final_web.pdf [accessed 2025-12-10]
7. Tooth LR, Moss KM, Mishra GD. Adherence to screen time guidelines among families in Australia with children of different ages. JAMA Pediatr 2022 May 1;176(5):517-519. [doi: 10.1001/jamapediatrics.2021.6382] [Medline: 35157014]
8. Foreman J, Salim AT, Praveen A, et al. Association between digital smart device use and myopia: a systematic review and meta-analysis. Lancet Digit Health 2021 Dec;3(12):e806-e818. [doi: 10.1016/S2589-7500(21)00135-7] [Medline: 34625399]
9. Terasaki H, Yamashita T, Yoshihara N, Kii Y, Sakamoto T. Association of lifestyle and body structure to ocular axial length in Japanese elementary school children. BMC Ophthalmol 2017 Jul 12;17(1):123. [doi: 10.1186/s12886-017-0519-y] [Medline: 28697750]
10. Enthoven CA, Polling JR, Verzijden T, et al. Smartphone use associated with refractive error in teenagers. Ophthalmology 2021 Dec;128(12):1681-1688. [doi: 10.1016/j.ophtha.2021.06.016] [Medline: 34245754]
11. Benedetto S, Carbone A, Drai-Zerbib V, Pedrotti M, Baccino T. Effects of luminance and illuminance on visual fatigue and arousal during digital reading. Comput Human Behav 2014 Dec;41:112-119. [doi: 10.1016/j.chb.2014.09.023]
12. Rosenfield M. Computer vision syndrome: a review of ocular causes and potential treatments. Ophthalmic Physiol Opt 2011 Sep;31(5):502-515. [doi: 10.1111/j.1475-1313.2011.00834.x] [Medline: 21480937]
13. Yu H, Akita T, Koga T, Sano N. Effect of character contrast ratio of tablet PC and ambient device luminance ratio on readability in low ambient illuminance. Displays 2018 Apr;52:46-54. [doi: 10.1016/j.displa.2018.03.002]
14. Norton TT. What do animal studies tell us about the mechanism of myopia—protection by light? Optom Vis Sci 2016 Sep;93(9):1049-1051. [doi: 10.1097/OPX.0000000000000917] [Medline: 27362614]

XSL•FO
RenderX

15. Lin C, Ji Z, Lin Y. Optimum display luminance and contrast polarity of desktop head-up display under office lighting level based on visual ergonomic study. Ergonomics 2024 Nov;67(11):1491-1503. [doi: 10.1080/00140139.2024.2339439] [Medline: 38629123]
16. Chen H, Sun Y, Luo S, et al. Association of screen content with early development among preschoolers in Shanghai: 7-day monitoring study with auto intelligent technology. J Med Internet Res 2025 Mar 5;27:e65343. [doi: 10.2196/65343] [Medline: 40053802]
17. Austin PC, Fang J, Lee DS. Using fractional polynomials and restricted cubic splines to model non-proportional hazards or time-varying covariate effects in the cox regression model. Stat Med 2022 Feb 10;41(3):612-624. [doi: 10.1002/sim.9259] [Medline: 34806210]
18. Karouta C, Ashby RS. Correlation between light levels and the development of deprivation myopia. Invest Ophthalmol Vis Sci 2014 Dec 9;56(1):299-309. [doi: 10.1167/iovs.14-15499] [Medline: 25491298]
19. Smith EL III, Hung LF, Huang J. Protective effects of high ambient lighting on the development of form-deprivation myopia in rhesus monkeys. Invest Ophthalmol Vis Sci 2012 Jan 25;53(1):421-428. [doi: 10.1167/iovs.11-8652] [Medline: 22169102]
20. Lan W, Feldkaemper M, Schaeffel F. Intermittent episodes of bright light suppress myopia in the chicken more than continuous bright light. PLoS One 2014;9(10):e110906. [doi: 10.1371/journal.pone.0110906] [Medline: 25360635]
21. Rucker F, Britton S, Taylor C. Color and Temporal Frequency Sensitive Eye Growth in Chick. Invest Ophthalmol Vis Sci 2018 Dec 3;59(15):6003-6013. [doi: 10.1167/iovs.18-25322] [Medline: 30572345]
22. Stone RA, Lin T, Laties AM, Iuvone PM. Retinal dopamine and form-deprivation myopia. Proc Natl Acad Sci U S A 1989 Jan;86(2):704-706. [doi: 10.1073/pnas.86.2.704] [Medline: 2911600]
23. Iuvone PM, Tigges M, Fernandes A, Tigges J. Dopamine synthesis and metabolism in rhesus monkey retina: development, aging, and the effects of monocular visual deprivation. Vis Neurosci 1989;2(5):465-471. [doi: 10.1017/s0952523800012360] [Medline: 2577263]
24. McCarthy CS, Megaw P, Devadas M, Morgan IG. Dopaminergic agents affect the ability of brief periods of normal vision to prevent form-deprivation myopia. Exp Eye Res 2007 Jan;84(1):100-107. [doi: 10.1016/j.exer.2006.09.018] [Medline: 17094962]
25. Feldkaemper M, Schaeffel F. An updated view on the role of dopamine in myopia. Exp Eye Res 2013 Sep;114:106-119. [doi: 10.1016/j.exer.2013.02.007] [Medline: 23434455]
26. Iuvone PM, Tigges M, Stone RA, Lambert S, Laties AM. Effects of apomorphine, a dopamine receptor agonist, on ocular refraction and axial elongation in a primate model of myopia. Invest Ophthalmol Vis Sci 1991 Apr;32(5):1674-1677. [Medline: 2016144]
27. Hua WJ, Jin JX, Wu XY, et al. Elevated light levels in schools have a protective effect on myopia. Ophthalmic Physiol Opt 2015 May;35(3):252-262. [doi: 10.1111/opo.12207] [Medline: 25913873]
28. He X, Sankaridurg P, Wang J, et al. Time outdoors in reducing myopia: a school-based cluster randomized trial with objective monitoring of outdoor time and light intensity. Ophthalmology 2022 Nov;129(11):1245-1254. [doi: 10.1016/j.ophtha.2022.06.024] [Medline: 35779695]
29. Suh YW, Ha SG, Kim SH. Effect of classroom illuminance on the development and progression of myopia in school children. Korean J Ophthalmol 2022 Jun;36(3):194-201. [doi: 10.3341/kjo.2021.0170] [Medline: 35067020]
30. Gwiazda J, Ong E, Held R, Thorn F. Myopia and ambient night-time lighting. Nature 2000 Mar 9;404(6774):144-144. [doi: 10.1038/35004663] [Medline: 10724158]
31. Norton TT, Siegwart JT. Light levels, refractive development, and myopia--a speculative review. Exp Eye Res 2013 Sep;114:48-57. [doi: 10.1016/j.exer.2013.05.004] [Medline: 23680160]
32. BS ISO/CIE 8995-1:2025 light and lighting — lighting of work places — part 1: indoor. : International Organization for Standardization & CIE; 2025 URL: https://www.iso.org/standard/76342.html [accessed 2025-12-10]
33. GB/t 36876-2018. hygienic requirements of lighting design and setting in primary and middle schools. : State Administration for Market Regulation, Standardization Administration of China; 2018 URL: https://www.chinesestandard.us/products/gbt36876-2018 [accessed 2025-12-10]
34. Chen JG, Wu SK, Chiu HP, Tu CN, Liu CH. Evaluation of three tablet computers at two levels of ambient illumination. Int J Hum Comput Interact 2016 May 3;32(5):394-401. [doi: 10.1080/10447318.2016.1151614]
35. Zhou Y, Shi H, Chen QW, Ru T, Zhou G. Investigation of the optimum display luminance of an LCD screen under different ambient illuminances in the evening. Appl Sci 2021;11(9):4108. [doi: 10.3390/app11094108]
36. Shieh KK, Lee DS. Preferred viewing distance and screen angle of electronic paper displays. Appl Ergon 2007 Sep;38(5):601-608. [doi: 10.1016/j.apergo.2006.06.008] [Medline: 17049333]
37. Sheedy JE, Smith R, Hayes J. Visual effects of the luminance surrounding a computer display. Ergonomics 2005 Jul 15;48(9):1114-1128. [doi: 10.1080/00140130500208414] [Medline: 16251151]
38. Yin Z, Liu B, Hao D, Yang L, Feng Y. Evaluation of VDT-induced visual fatigue by automatic detection of blink features. Sensors (Basel) 2022 Jan 25;22(3):916. [doi: 10.3390/s22030916] [Medline: 35161662]
39. Miyake-Kashima M, Dogru M, Nojima T, Murase M, Matsumoto Y, Tsubota K. The effect of antireflection film use on blink rate and asthenopic symptoms during visual display terminal work. Cornea 2005 Jul;24(5):567-570. [doi: 10.1097/01.ico.0000151564.24989.38] [Medline: 15968162]

XSL•FO
RenderX

40. Muhamad N, Amali NAN. Digital display preference of electronic gadgets for visual comfort: a systematic review. Iran J Public Health 2023 Aug;52(8):1565-1577. [doi: 10.18502/ijph.v52i8.13396] [Medline: 37744537]
41. He X, Wang J, Zhu Z, et al. Effect of repeated low-level red light on myopia prevention among children in China with premyopia: a randomized clinical trial. JAMA Netw Open 2023 Apr 3;6(4):e239612. [doi: 10.1001/jamanetworkopen.2023.9612] [Medline: 37099298]

Abbreviations

HSVC: Healthy Screen Viewing for Children
RCS: restricted cubic spline

Edited by A Kushniruk; submitted 19.Jun.2025; peer-reviewed by R Doron, W Yang; revised version received 18.Nov.2025; accepted 01.Dec.2025; published 28.Jan.2026.

Please cite as:
Zhu Y, Chen H, Lin S, Jiang H, Zhang M, Sun Y, Li C, Jia Y
Objective Monitoring of Tablet Use–Related Optical Exposure and Its Association With Axial Length in Preschool Children: Cross-Sectional Intelligent Monitoring Study
JMIR Hum Factors 2026;13:e79266
URL: https://humanfactors.jmir.org/2026/1/e79266
doi:10.2196/79266

© Yidong Zhu, Hao Chen, Senlin Lin, Hong Jiang, Mingdao Zhang, Yi Sun, Chenshu Li, Yingnan Jia. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 28.Jan.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Good Clinical Practice Guidance for Line Materials, Filtration, and Light Protection in Intravenous Medication Administration: Modified Delphi Consensus Study

Andrew Dickman[1], BSc, MSc, DPharm, MRPharmS; Penelope Tuffin[2], PGDipHospPharm, MPallCare; Rania Al-Jaber[3], BSc, RPh, MSc; Mona El-Harmeel[4], RPh, MBA; Jennifer Schneider[5], PhD; Irene Taladriz-Sender[6], DPharm; Adam Sutherland[7], PhD, MRPharmS; Robert Terkola[8], MSc, PhD, aHPh; James Waterson[9], BA, MHEc, MMedEd

[1]Liverpool University Hospitals NHS Foundation Trust, Liverpool, United Kingdom

[2]Fiona Stanley Hospital, Bethesda Hospital and Western Australia Country Health Service, Australia

[3]King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia

[4]Children's Cancer Hospital, Cairo, Egypt

[5]School of Medicine and Public Health Faculty of Medicine and Health, Newcastle, Australia

[6]Hospital General Universitario Gregorio Marañón, Madrid, Spain

[7]School of Pharmacy and Medical Sciences, University of Bradford, Bradford, United Kingdom

[8]Paracelsus Medical University, Salzburg, Austria

[9]Becton Dickinson and Company, Dubai, United Arab Emirates

Corresponding Author:
James Waterson, BA, MHEc, MMedEd
Becton Dickinson and Company
11F Blue Bay Tower
Dubai, 25229
United Arab Emirates
Phone: 971 566035154
Email: james.waterson@bd.com

Abstract

Background: There is an unmet need for reliable medication stability information to avoid suboptimal administration of intravenous medications, which may lead to reduced efficacy of therapy and potential patient harm through medication degradation or incompatibilities that cause vascular access issues.

Objective: This study aimed to develop evidence-informed guidance for pharmacists and nurses on the use of administration line materials, in-line filtration, and light protection during storage and intravenous administration of medications commonly used in critical care and oncology.

Methods: An initial list of 181 medications was compiled in consultation with pharmacist stakeholders from critical care and oncology specialties. A modified Delphi study was conducted over 3 rounds with a panel of 8 expert pharmacists selected for their clinical expertise, professional experience, and geographic location to ensure representation of diverse health care settings and medication administration practices. Panelists anonymously ranked statements on a 5-point Likert scale developed from a review of the literature with respect to the requirement for a specific administration line material for each medication; the need for, and type of, filtration required during administration; and the need for light protection during storage and administration of medications. After each round, items achieving 80% consensus were finalized. Those that did not achieve consensus were carried forward to the next round. This iterative approach allowed panelists to reconsider their ratings based on emerging group consensus and additional evidence shared by panelists between rounds.

Results: A total of 1044 administration and storage requirements were assessed for a final list of 174 medications. In round 1, consensus was reached for 613 (58.7%) statements, with every statement being addressed and scored by the panelists. Panelists provided additional evidence sources for their decisions, and these were distributed to all panel members for round 2. All items were addressed and scored by the panelists. By the conclusion of round 3, consensus had been achieved for 697 (66.8%) statements, with all items being addressed and scored by the panelists.

XSL•FO
RenderX

Conclusions: This study developed consensus-based recommendations for the selection of administration line materials, the use of in-line filtration, and light protection for the administration and storage of a range of medications administered intravenously. The guidance will aid medication stability and efficacy and promote good clinical practice; it currently underpins a prototype bedside app for correct intravenous administration line selection for nursing and pharmacy staff in critical care and oncology units.

(JMIR Hum Factors 2026;13:e88333) doi:10.2196/88333

KEYWORDS

administration; consensus; panels; reviews; protection; panelists

Introduction

Pharmacists are considered, by both medical and nursing staff, to be the arbiters of administration knowledge for intravenous (IV) medications. Knowledge of administration line material types and their associated requirements, including light protection and filtration, has been reported to be poor, even among nurses working in specialist disciplines [1]. This is despite approximately 40% of nursing time being devoted to medication administration [2]. In a survey of Spanish nurses, 80.8% of respondents reported having "a good level of knowledge on the preparation and administration of IV medicine," although 88.1% also reacted positively to statements on the need to improve their skills [3].

While pharmaceutical manufacturers provide guidance on the storage and protection of parental medications, there is generally much less information available to the end user after reconstitution or following compounding processes. Furthermore, preparation of IV medicines takes place in many locations across facilities, and information should be provided for their mixing and administration across many medical specialties. Some health systems provide medications compounded in the pharmacy or by a commercial compounder, in which cases, the package inserts may be unavailable for consultation by the nursing staff who administer the medication. Conversely, in many health systems, most IV medication preparation is undertaken by nurses in the clinical area without direct pharmacist support [4], and often package inserts or supporting guidelines do not offer information on storage and handling after reconstitution. For example, there may be an assumption that medicines prepared on the ward will be administered immediately, but this is not always the case.

An extensive survey of sterile compounding errors in American hospital pharmacies indicated that wrong preparation techniques, including the attachment of an incorrect administration line, were identified in 26% of compounding errors [5]. A key contributory factor to these findings was the lack of standardized processes. The need for systems-focused interventions, including standardized pathways, has been recognized as a key concept in driving efficiency and sustainability in cancer care [6]. A substantial component of such tools would be dedicated to improved compounding processes, final-product preparation, labeling, and the provision of information [7].

There are many variables to consider when developing standard practice guidelines for medication preparation and administration, and these differ depending on where and how the medication is prepared for administration. However, with the assumption that most preparation occurs at the bedside and that, once a medication is reconstituted or diluted, members of the health care team need a practical reference guide for IV medication stability, the requirements for such a clinical reference fall broadly into the following 3 categories: the need for filtration, the need for protection from light during storage and administration, and the material composition of administration systems—syringes, bags, and lines. These factors are essential for maintaining medication stability and effectiveness, yet this information is often missing from standard drug references.

Additionally, there are 3 potential medication-material interactions: absorption into the administration line, adsorption onto the administration line, and leaching or migration of materials into medication solutions. Reduced efficacy of medications due to product degradation is a concern because it has the potential to directly impact patient care and management and affect treatment outcomes associated with regimens and individual medications. Vascular access issues, patient condition, or delayed laboratory results can cause delays in medication administration in both critical care and oncology treatment centers [8].

Such delays may be important in terms of the integrity of the compounded solution if it is sensitive to light exposure or affected by contact with administration line material.

The deleterious effect of light on some compounds, such as parenteral nutrition, is well known [9], but the effect of light on other medications is less well established. Approximately 30% of the degradation of clear medication solutions due to light may occur in the line or tubing of the infusion set [10], and this would logically increase at very low flow rates.

There is also an aspect of pharmaceutical waste in these circumstances, whereby medications that are not administered may be discarded because of uncertainty regarding their stability outside the controlled environments in which they are stored and prepared.

There is an unmet need for reliable and unambiguous medication stability information to support clinical decision-making at the bedside. This study aimed to develop guidance for different types of administration lines, in-line filtration, and light protection during storage and administration through a review by subject matter experts of available evidence for 181 IV medications commonly used in critical care and oncology.

XSL•FO
RenderX

Methods

Overview

The stability of IV admixtures is fairly well studied; however, the literature on this topic is vast and often contradictory. Therefore, we sought expert consensus to inform pragmatic information for clinicians and conducted a 3-round Delphi study. The Delphi approach is considered a powerful method that can deliver agreement in areas of uncertainty or where definitive evidence is lacking, and it allows expert opinion to emerge through consensus [11-13].

We formed a panel of 8 subject matter experts, identified for their professional experience, geographic location, patient population served, and compounding processes and workflows used in practice. The panel represented practice and expertise from Europe, the United Kingdom, and Australia, and practice in hospitals, manufacturing units, and academia. The panel also represented pediatric practice, critical care, oncology and palliative care. The panel members were selected by the lead author (AD) and study facilitator (JW), but the panel was otherwise anonymous. Initial agreement to participate was secured by email exchange and a formal letter of commitment. AD acted as chair of the panel and oversaw data analysis and the mitigation of disagreements in later stages.

Our modified Delphi process ran over 3 rounds. After 2 rounds, a further review was conducted by panel members who were considered outliers with respect to consensus. They were shown the anonymized scores of other panel members who had reached consensus. In addition, the panel chairperson (AD) and cochairperson (PT) reviewed these results and had casting votes. The reason for this approach was that, while local practice was actively sought as a constructive contribution to the Delphi panel, it might also act as an anomaly if practitioners had not been exposed to recent evidence regarding an individual medication. In this way, it was hoped that the Delphi panel process would itself facilitate consensus by sharing knowledge and challenging assumptions. Outlier selections by only one panel member were subject to the casting votes of the chair and cochair, but, as noted later, this only extended to exposing the panel member to the scores of other (still anonymized) Delphi panel members and encouraging further review of their selection. This modification of the Delphi technique is in line with current thinking on the process, which advocates “controlled feedback” [14].

Panel members were presented with an evidence dossier of 181 IV medications (presented by their recommended international nonproprietary names) and a library of literature to support assessment and opinion development. The following references were used to build the dossier:

1. Stabilis.org (independent pharmacy advice system) [15]
2. Medicines and Light, Oslo University [10]
3. University of Illinois at Chicago College of Pharmacy, Drug Information Group, with the 2023 update used upon publication [16,17]
4. Manufacturer package inserts
5. Individual independent scientific studies addressing specific medications and research on material compatibility, filtration, and light protection requirements

Sources 1 to 3 were selected because they were either recently published or updated frequently. Source 4 was selected as these are the evidence presented to and accepted by regulatory bodies. The independent scientific studies (n=5) were obtained via individual medication name searches undertaken in PubMed, Google Scholar, and under searches using specific search terms related to administration line materials. Evidence sources were evaluated according to time since publication and the strength of their conclusions. Where multiple sources were available, literature that most closely met the above criteria was selected. A sample of the evidence dossier is given in Multimedia Appendix 1.

Delphi panel members were instructed to work independently to determine their own opinions regarding the administration requirements of each medication with regard to material selection of the administration line, the need for protection from light during storage and administration, and the need for filtration of the product during administration. They were advised to use real-world assumptions to guide their deliberations: there may be delays between compounding and administration due to compounding processes, such as batching and preparation ahead of time, and patient-related factors (such as vascular access issues) could lead to an extended period during which the infusion would be exposed to light and possible administration line material interactions.

The threshold for consensus across the 8 panel members was set at 80% for each question asked. After each round of questions, those medications and their administration characteristics that reached 80% consensus were removed from further rounds. In the third (and final) round, the chair approached those panelists who were outliers in their responses. These panelists were shown the anonymized decisions of other panel members and asked whether they wished to reconsider their responses. Four medication profiles were resolved using this process, with 1 panelist per medication (3 individual panelists overall) being identified as an outlier.

The role of the chairperson was to oversee data analysis, identify and remove items that had reached consensus at each round, and field comments, questions, and queries from panel members. Panel members were also encouraged to identify and present new evidence if, within their scope of experience, they felt it was relevant, but the chairperson made decisions on the inclusion of this new data.

Ethical Considerations

Given the absence of human participants or their data and the noncontentious nature of the questions being asked, the study was undertaken as a service evaluation. All variables in the worksheets were vendor-neutral. Panel members received reimbursement for their time. Individual participation by panel members was undertaken under the supervision of the Ethics and Compliance Department of Becton, Dickinson and Company and according to the Becton Dickinson Global Standards for Interactions with Healthcare Professionals, Healthcare

XSL•FO
RenderX

Organizations, and Government Officials (effective March 1, 2022) and the Becton Dickinson Global Interactions with Healthcare Professionals Policy (effective October 1, 2019). Reimbursement and the expected duration of the work involved were agreed upon through four standard operating procedures for engagement of health care professionals according to this policy:

1. Calculation of a reimbursement fair market value (hourly rate) according to profession, qualification, standing, and geography
2. Cosigning of a master services agreement stating the fair market value and general conditions of engagement
3. A cosigned statement of work stating the task to be undertaken and the expected duration of the activity
4. A cosigned privacy agreement stating the rights of panelists regarding data protection under the General Data Protection Regulation: Regulation (European Union) 2016-679 of the European Parliament and of the Council of April 27, 2016

Delphi Procedure

Panelists were provided with 3 worksheets created in Microsoft Excel (version 17). Each worksheet considered a different aspect of the administration or storage of the medication—material of the final administration line, protection from light, or the need for terminal filtration at the point of administration. Panelists were presented with recommendations from the evidence dossier. Because there was likely to be variation in practice and experience, and because for many medicines, answers to the questions could be ambiguous (eg, morphine could be refrigerated or not with no impact on the stability of the end product), a binary "Yes/No" response would not have captured those medicines for which storage or administration materials were critical and those for which these factors were less important. Therefore, panelists were asked to rate the strength of their decision on a 5-point Likert-type scale. This scale ranged from 1 ("strongly disagree") to 5 ("strongly agree"). A score of 3 indicated no strong opinion. We chose a 5-point Likert scale because there is no absolute definition for the tools a Delphi panel should use, and we were interested in identifying disagreement as much as consensus. However, it was important to give panelists distinct and clear statements for evaluation [11,17,18].

Definition of Consensus

Classically, Delphi panel methods accept consensus at ≥70% for individual-item removal from subsequent rounds [18]. This is commonly because there is a dearth of empirical evidence in areas where the technique is applied [19]. Given the importance of appropriate administration practice, the accepted threshold for consensus in this study was set at 80%. Likert weighting was used as the first point of assessment of the overall panel response. This required median scores of 1 to 2 (negative) or 4 to 5 (positive) on the Likert scale to be considered a collectively valid response to the statement being reviewed. The degree of consensus was calculated using the SD of the panel members' scores. Percentage consensus was then derived using the Excel formula:

Consensus percentage=(1–SD/2)×100

In practice, it was found that, for items where the median did not reflect a strong opinion (median >2 and median <4), the SD was also always >0.4, giving a consensus score of less than 80%.

After the first round, items with ≥80% consensus were removed from review. Comments from panel members in round 1 were reviewed by the chairperson and cochairperson, which led to the removal of several medications. For example, moxetumomab was deleted based on its removal from the European market in 2021 and from the US market in 2023 [20].

Panel members also offered further evidence sources for their score selection in round 1. The chairperson and cochairperson added these to the second-round review. Therefore, the evidence base grew and could alter opinions and gain more consensus through subsequent rounds. For the remaining items, each member's scores were returned to them in a workbook with instructions to review them as in round 1, and to apply any change they felt was appropriate to their score.

After round 2, items with ≥80% consensus were removed from further analysis. Items were identified for which a maximum of 2 panel members' determinations were outside the consensus (ie, outliers). Mini-review sheets were created for individual panel members whose scores for individual medications were outliers, and all panel members' scores (aggregated while maintaining anonymity) were revealed to the outlier panel member. The chairperson and cochairperson also rated these medications, effectively using their casting votes for these medications only.

After this final round of review, the process was closed, with items with >80% consensus being carried forward as clinical end-user guidance.

Results

Following the removal of medications no longer marketed, 174 medications were assessed. Of the 1044 IV medication administration requirements assessed, there was ≥80% consensus for 613 (58.71%) items after round 1, rising to 697 (66.76%) items after 3 rounds. The number of definitive and nondefinitive recommendations achieving ≥80% consensus for each attribute assessed is shown in Table 1.

XSL•FO
RenderX

Table 1. Number of definitive and nondefinitive recommendations (N=1044)[a].

Recommendation category	Definitive recommendations (≥80% consensus; n=697), n (%)	Nondefinitive recommendations (<80% consensus; n=347), n (%)
Di(2-ethylhexyl) phthalate–free polyvinyl chloride administration set	132 (75.86)	42 (24.14)
Polyethylene administration set	123 (70.69)	51 (29.31)
Polypropylene administration set	93 (53.45)	81 (46.55)
In-line filtration		
0.2-µm filter	16 (9.20)	10 (5.75)
1.2-µm filter	2 (1.15)	1 (0.57)
0.2- to 1.2-µm filter	1 (0.57)	2 (1.15)
0.2- to 5-µm filter	1 (0.57)	0 (0)
15-µm filter	0 (0)	2 (1.15)
No filter required	131 (75.29)	8 (4.60)
Light protection during storage		
Yes	76 (43.68)	41 (23.57)
Not required	26 (14.94)	31 (17.81)
Light protection during administration		
Yes	8 (4.60)	62 (35.64)
Not required	88 (50.59)	16 (9.17)

[a]For full output by each assessed medication, refer to Multimedia Appendix 2.

Table 2 illustrates how consensus and no-consensus conditions were determined. Some consensus issues were due to highly conflicting (and sometimes emerging) evidence. This highlighted how the final consensus opinion could diverge strongly from the initial evidence statement presented to the panelists for review. This was most commonly seen in the sections related to light protection during administration and filtration.

XSL•FO
RenderX

Table 2. Medication evidence statements with panelist comments and consensus alignment from rounds 1 and 2 of the Delphi panel.

Presented statements	Medications	Sample panelist comments	Consensus alignment (%)
Light protection during administration			
Yes			
	Temsirolimus (Torisel)	"reference 234"; "torisel should be protected from excessive room light and sunlight"	100.0
	Leucovorin (calcium folinate)	"injectable drugs guide protect from light"; "however AHFS [American Hospital Formulary Service]: stable for 24 h when at RT[a], protected from light"; "however evidence that reconstituted solutions of drug are not adversely affected by exposure to room light (Calcium leucovorin)"	61.8
	Filgrastim	"no mention...Injectable drugs guide...also stabilis studies under light and reference 275: diluted neupogen can be stored at RT for up to 24h, no mention of protect from light"	39.3
	Necitumumab	"Store protected from light. Brief exposure to ambient light is acceptable while preparation and administration is taking place. If delay in admin, protect from light. Score=4 to err on side of caution"	39.3
	Decitabine	"this is not stable drug- no mention of light in monographs but 1 h room temp. Within 15 minutes of reconstitution, the concentrate (in 10 ml of sterile water for injections) must be further diluted with cold (2 °C-8 °C) infusion fluids. This prepared diluted solution for intravenous infusion can be stored at 2 °C-8 °C for up to a maximum of 3 hours, followed by up to 1 hour at room temperature (20 °C-25 °C) before administration."	37.1
	Naxitamab-gqgk	"protect until time of use. FDA advice per document 2020/761171	31.3
	Daunorubicin	"reference 5"	29.3
	Bendamustine	"ref 5: stability of diluted solutions affected by light"; "the final admixture is stable for 24 hours when stored refrigerated (2-8 °C or 36-47 °F) or for 3 hours when stored at room temperature (15-30 °C or 59-86 °F) and room light. Administration of TREANDA must be completed within this period"; "source:https://www.accessdata.fda.gov/drugsatfda_docs/label/2008/022303lbl.pdf. (err on side of caution due to 3h)"	26.9
	Chlorphenamine	"no reference, usually give over short time: use immediately after dilution, so protect from light not relevant Administered as a short infusion/bolus so wouldn't be unwrapped until the moment of use usually [AS]"	25.5
	Cisplatin	"Cisplatin 0.15 mg/mL in sodium chloride IV infusion 0.9% is chemically stable for 24 hours when stored at room temperature and protected from light. https://medsinfo.com.au/product-information/document/DBL_Cisplatin_Injection_PI. Also Baertschi reference"	25.5
	Doxorubicin liposomal	"no specific reference but because of doxorubicin component, err on side of caution"	25.5
	Bevacizumab	"No information stating need light protection during admin"	23.6
	Bortezomib	"reference 5 states reconstituted solutions affected by light but Velcade states does not need to be protected from light when admin: also not longer than 8h when exposed to room light (change from 5 to 3)"	19.6
	Trastuzumab emtansine	"infusions over 30 to 90 min; only information I can find says if not immediately administer, store in refrigerator; refrigerators may or not protect from light…"	16.3
	Digoxin	"no reference states protect from light Administered as a short infusion/bolus so wouldn't be unwrapped until the moment of use usually"	10.2
	Meropenem	"no reference to light. Injectable drugs guide Administered as a short infusion/bolus so wouldn't be unwrapped until the moment of use usually"	10.2
	Amiodarone hydrochloride	"Illinois Chicago reference states does not need to be protected from light during admin BUT avoid exposure to direct sunlight (3 because of sunlight reference): however Tonnesen reference refers to amiodarone as light sensitive critical care drug"	9.9
	Midazolam hydrochloride	"no reference to protect from light injectable drugs guide. No outcome related data to support this"	5.7
Not required			

XSL•FO
RenderX

Presented statements	Medications	Sample panelist comments	Consensus alignment (%)
	Daratumumab	"reference mentions followed by 15 h including infusion time at RT and room light"	25.5
	Daunorubicin liposomal	"based on daunorubicin protected from light, erring on side of caution would say protect this formulation from light"	25.5
	Dexamethasone	"given over short time analyzed it myself"	25.5
	Etoposide	"micromedex stability studies indicating stability under fluorescent light"	25.5
Filtration no or filter size during administration			
1.2	Infliximab	"reference Hospital Pharmacy Ipema et al"	81.4
0.2	Tagraxofusp	"reference 233: elzonris refers to using 0.2-µm filter"	81.4
0.2	Avelumab	"Aust Inject Handbook and also reference 29 state use 0.2-µm low protein binding filter"	62.7
Not required	Atezolizumab	"reference 26: use of in-line filter membranes is optional"	61.8
1.2	Asparaginase (*Escherichia coli*)	"reference mentions use of 5-µm filter"; "loss of potency with 0.2 µm"; "this spreadsheet states 1.2 µm-perhaps better to be 5 µm?"; "also, this product may no longer be available? Changed to 4 due to filter size"	55.1
Not required	Asparaginase Erwinia	"no mention of use of filter score remains at 5"	39.3
0.2	Cemiplimab	"reference Hospital Pharmacy Ipema et al"	26.9
0.2	Dinutuximab	"reference 103 recommended that a 0.22-µm in-line filter during infusion"	26.9
0.2	Blinatumomab	"product information states to use 0.2-µm filter (reference 263) ipema et al Hospital Pharmacy (7 day not required)"	25.5
Not required	L-asparaginase	"filter may be used if fibrous particles 5 µm and no loss of potency, some loss of potency with 0.2 µm ref 154"	25.5
Not required	Moxetumomab	"no reference to use of filter"	25.5
0.2	Magnesium sulphate	"If diluted below super-saturated solution (50%) then no filter required in practice"	23.6
0.2	Clofarabine	"SPC [Summary of Product Characteristics]: Clofarabine-ratio pharm 1 mg/ml Concentrate for Solution for Infusion be filtered through a sterile 0.2-micron syringe filter and then diluted with a 9 mg/ml (0.9%) sodium chloride infusion solution)"	16.3

[a]RT: room temperature.

Discussion

Principal Findings

We have generated an extensive list of medications, with recommended administration sets and storage and handling requirements, that can be applied at the bedside to potentially improve access to high-quality medication information. The Delphi panel method proved to be a useful tool for creating consensus-based guidance for clinicians administering IV medications, many of which might be considered specialized therapies. The study goal was achieved by creating a distilled and easy-to-understand guide. The guidance is comprehensive for the parameters reviewed, and its basis in published evidence and professional expertise should provide clinicians with confidence in any tool within which it is subsequently deployed. Furthermore, by distilling a diverse range of sources into a single output, we potentially mitigate some of the idiosyncrasies commonly present in localized clinician assessments.

There is potential for this tool to reduce avoidable medication waste and unnecessary storage, which supports health care initiatives aimed at improving sustainability and efficiency. This is additionally important given the financial burden of medication waste on health care systems and patients [19].

The panel made a collaborative decision during a final open meeting on the required frequency of review of the medication list, based on likely changes in the evidence base, new administration line materials, and new medications entering the market. This was set at 24 months, which is within the review timing of other medication review sites, such as the National Injectable Medicines Guide in the United Kingdom [21]. The assessment of new medications, either requiring an interim decision between formal reviews or addition to the "pending" list for future panel review, is possible based on local experience and the review of dashboard sites, such as the UK National Institute for Health Research Innovation Observatory [22].

Limitations

There are limitations to the study. Only the most common administration line materials and those most readily available to clinicians were considered to ensure real-world guidance. New medications may be introduced, while established medications may be withdrawn, and the evidence base for their administration requirements is changing. Line material is also

XSL•FO
RenderX

evolving. For example, thermoplastic polyurethane is expected to replace polyvinyl chloride (manufactured without di(2-ethylhexyl) phthalate). No study can be exhaustive, but a review period of 24 months was considered by the panel to be sufficient to cover these changes in the environment. It is also difficult to cover, in any study or review, all regions and countries. There was no panel member from North America, but there was a strong reliance on evidence presented to the Food and Drug Administration.

There was no consensus for just over one-third of the medications assessed. Therefore, the global nature of the panel highlighted a diversity of opinion and reflects the issues faced by compounding pharmacists and nursing staff. These unresolved medications can be prioritized in the planning of the next Delphi panel (ie, at the 24-month review period). The potential for machine learning to assist in synthesizing evidence dossiers for these medications could include incorporating evidence from languages other than English. A wider catchment of generic manufacturers' package inserts and extracts from industry publications could also aid in the creation of consensus for these medications. Our general advice for these medications is to follow a conservative approach for light protection during storage, and to consider the duration of infusion when determining whether light protection is needed during administration. For filtration, there is emerging evidence that higher-concentration infusions used at lower rates [23], and the use of 1-lumen for multiple infusions may contribute to vascular access occlusion, and that the use of 0.2-µm filters may reduce this risk [24,25]. Reviewing medications at high and low concentrations, and the effect of concentration on the need for specific line materials, filtration, and light protection, could be included in subsequent reviews.

The dossier of evidence was prepared by JW and AD. Given the extent and variety of the evidence reviewed, it is inevitable that omissions were made. This was evident in the panelists' recommendations to incorporate new medications or remove those that were no longer actively marketed. In the future, these dossiers should be coproduced by a broader panel of experts, independent of the Delphi panel.

A further important limitation lies in the Delphi process. Delphi exercises are labor-intensive, and there is evidence that cycles of consensus reach diminishing returns once they extend beyond 3 cycles [26,27]. There is also the potential for bias within the panel; however, its members are selected, and it is impossible to ensure that the outcome would be the same if panel membership were different. Therefore, we must reiterate that this tool is not a universal consensus but is representative of the views of the 8 subject matter experts who took part.

Generic manufacturers are allowed—and often do use—different inactive excipients in their formulations. Excipients can influence properties, such as solubility, pH, particulate formation, adsorption to plastic, or light sensitivity. Therefore, 2 generics labeled as "the same drug" may behave differently. The ultimate goal would be to consult each generic's prescribing information or stability study, as uniform performance across all brands cannot be expected. In this work, we had to narrow down the scope of validity to the sources we have used in our dossier.

Recommendations for Future Research

While we have produced an outline framework for decision support for these medications, it has not been tested for utility in real-world settings. Therefore, we recommend a program of study evaluating the utility of this tool in health care practice environments. There are also myriad ways in which this can be deployed, either as simple lookup tables in low-resource settings or through integration into electronic patient records and prescribing systems.

Conclusions

We feel that our approach of restricting ourselves to the fundamental information required by bedside clinicians was vindicated, as casting the net wider to include other parameters, such as concentration and solvent, would arguably overcomplicate the output we were seeking from the study and any tools derived from it. Our experience here matched that of other authors who found that the "temptation to be comprehensive by including every theoretically possible scenario may be counterproductive. Also, being selective allows larger numbers of cues to be considered explicitly, which should be an advantage in terms of reliability" [17]. As noted previously under the "working assumptions" in the Methods section, panel members were instructed to use the most conservative scenarios for each medication being assessed; for example, higher-concentration solutions would be assessed as part of the requirements for filtration, administration line material, and light protection.

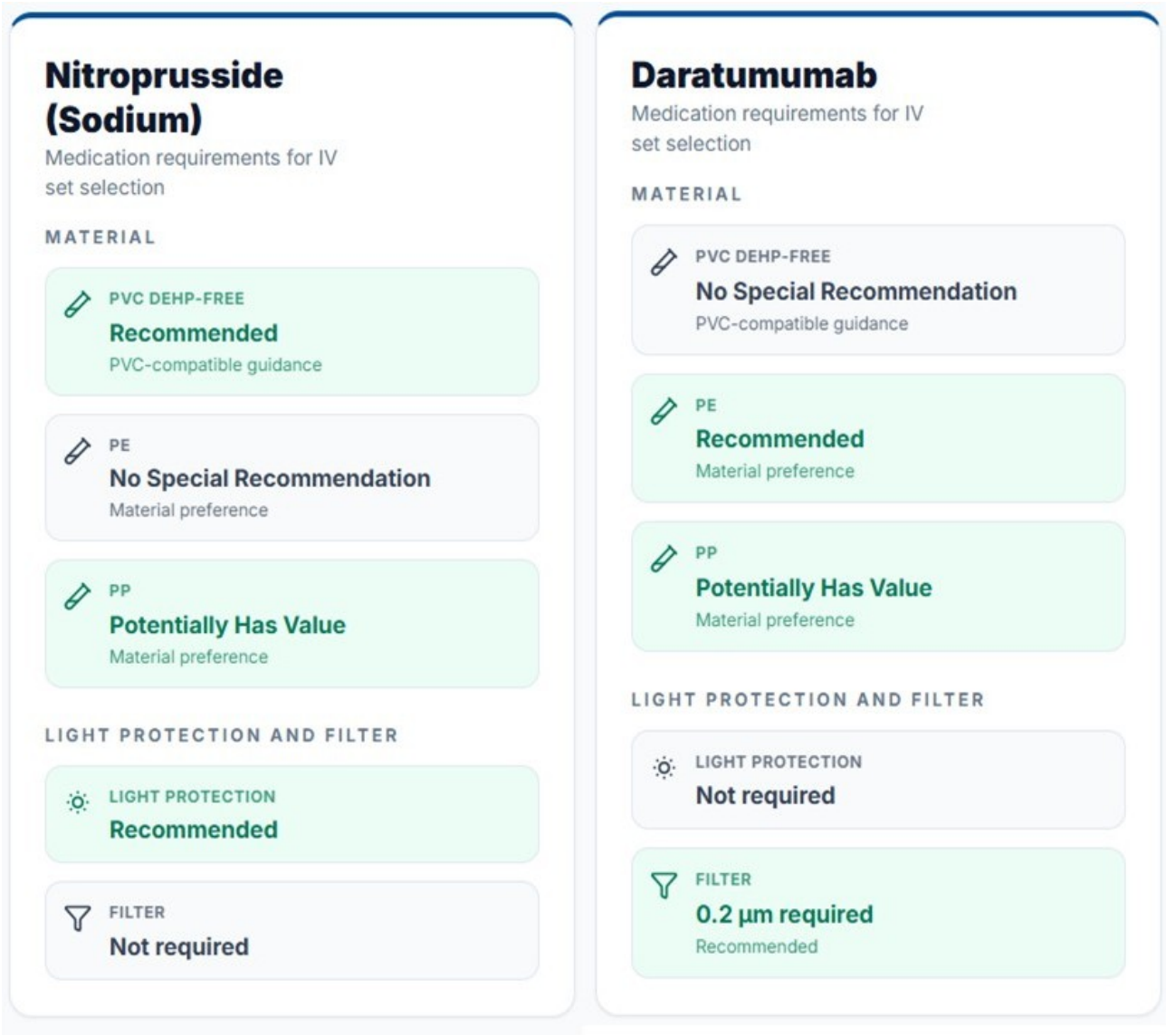
One immediate step that could use the results of the consensus study would be to label sterile compounded products with information for the end user, such as "protect from light during storage and administration."

The output of the study is also being used to underpin algorithms for a Python-based prototype bedside app through which clinicians can identify the correct administration line characteristics required for both primary and secondary administration after entering the medication name into the app. Such a tool could also help address the perennial problems of clinicians spending considerable amounts of time looking for information and guidance or addressing knowledge and information deficits with recalled facts gathered from past experience, including informal and formal teaching, interactions with colleagues, and the traditions of care within the unit where they work [28]. This was one reason we prioritized simplicity in our project plan and method, as we aimed to ensure that the results could be simply disseminated to the point of care. Given the 2022 study's findings that knowledge of administration line material type and characteristic requirements (light protection and filtration), even among specialist nurses, was poor [1]; the immense (and well-established) challenges associated with cascading information through organizations [29]; and the fact that end users would be predominantly nurses (the largest component of the tertiary health care workforce), such a point-of-care line selection tool could be very valuable [30].

XSL•FO
RenderX

We hope to fully evaluate this bedside tool in a study within the next year. In its current unreleased form, a medication can be selected on a smartphone app from a drop-down list of the medications on which consensus was secured, and the required properties of the line in terms of material, light protection, and filtration are displayed (Figure 1). This will need to be extended to the inventory of administration lines and IV infusion pumps available to the end user to direct them to the most appropriate options for the administration of the medication.

Figure 1. Screenshots of a prototype app to be used bedside for selecting administration line characteristics.



Acknowledgments

The authors acknowledge the contribution of Dr Eng Khalil Rafiq Al-Battat to the development of the initial prototype administration line selection tool shown in Figure 1.

Funding

The study was sponsored by Becton, Dickinson and Company. The study did not specifically address any product manufactured or distributed by Becton, Dickinson and Company, and was undertaken as a clinical and quality-of-care improvement exercise. The evidence dossiers were compiled by the Medical Affairs Department of Becton, Dickinson and Company. The dossiers, as noted previously, were compiled from the scientific literature, pharmaceutical manufacturers' package inserts, and independent pharmacy advice systems. No evidence was drawn from Becton, Dickinson and Company internal sources. The Medical Affairs Department operates as a distinct arm outside the commercial operations of the company.

XSL•FO
RenderX

Authors' Contributions

Study planning, conception, and design were undertaken by AD, PT, and JW. Data acquisition and dissemination were undertaken by JW. Initial manuscript drafting was undertaken by AD, PT, JW, RT, and AS. A Delphi panel review of each evidence monograph was undertaken by PT, AS, RT, RA-J, ME-H, JS, and IT-S. Review of the final manuscript and revisions in response to peer review were undertaken by AD, JW, PT, AS, RT, RA-J, ME-H, JS, and IT-S.

Conflicts of Interest

JW declares no conflicts of interest other than his employment in the Medical Affairs Department of Becton, Dickinson and Company. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
Sample evidence dossier.
[DOCX File , 20 KB - humanfactors_v13i1e88333_app1.docx]

Multimedia Appendix 2
Consensus recommendations after a 3-round Delphi study.
[DOCX File , 131 KB - humanfactors_v13i1e88333_app2.docx]

References

1. Al-Jaber R, Samuda N, Chaker A, Waterson J. Critical care nurses' knowledge of correct line types for administration of common intravenous medications: assessment and intervention study. JMIR Form Res 2022 Apr 26;6(4):e36710 [FREE Full text] [doi: 10.2196/36710] [Medline: 35471247]
2. Armitage G, Knapman H. Adverse events in drug administration: a literature review. J Nurs Manag 2003 Mar;11(2):130-140. [doi: 10.1046/j.1365-2834.2003.00359.x] [Medline: 12581401]
3. Márquez-Hernández VV, Fuentes-Colmenero AL, Cañadas-Núñez F, Di Muzio M, Giannetta N, Gutiérrez-Puertas L. Factors related to medication errors in the preparation and administration of intravenous medication in the hospital environment. PLoS One 2019 Jul 24;14(7):e0220001 [FREE Full text] [doi: 10.1371/journal.pone.0220001] [Medline: 31339914]
4. Albsoul-Younes A, Gharaibeh L, Murtaja AA, Al-Lawama M, Badran E, Hassan A, et al. Impact of implementation of a collaborative pharmacist-nurse work model on intravenous medication preparation errors: a quasi-study design. J Neonatal Nurs 2025 Apr;31(2):101616 [FREE Full text] [doi: 10.1016/j.jnn.2025.101616]
5. ISMP guidance and tools. ECRI. URL: https://ismp.org/resources/ismp-survey-provides-insights-pharmacy-sterile-compounding-systems-and-practices [accessed 2023-09-19]
6. Ferrara L, Otto M, Aapro M, Albreht T, Jonsson B, Oberst S, et al. How to improve efficiency in cancer care: dimensions, methods, and areas of evaluation. J Cancer Policy 2022 Dec;34:100355 [FREE Full text] [doi: 10.1016/j.jcpo.2022.100355] [Medline: 36007873]
7. Reece KM, Lozano MA, Roux R, Spivey SM. Implementation and evaluation of a gravimetric i.v. workflow software system in an oncology ambulatory care pharmacy. Am J Health Syst Pharm 2016 Feb 01;73(3):165-173. [doi: 10.2146/ajhp150169] [Medline: 26796911]
8. Keers RN, Williams SD, Cooke J, Ashcroft DM. Causes of medication administration errors in hospitals: a systematic review of quantitative and qualitative evidence. Drug Saf 2013 Nov;36(11):1045-1067 [FREE Full text] [doi: 10.1007/s40264-013-0090-2] [Medline: 23975331]
9. Chessex P, Laborie S, Nasef N, Masse B, Lavoie JC. Shielding parenteral nutrition from light improves survival rate in premature infants. JPEN J Parenter Enteral Nutr 2017 Mar;41(3):378-383. [doi: 10.1177/0148607115606407] [Medline: 26376662]
10. Tønnesen H. Medicines and Light-Ensuring Patient Safety: Photostability of Parenteral Products. Oslo, Norway: Gunn-Therese Lund Sørland; 2020.
11. Stewart D, Gibson-Smith K, MacLure K, Mair A, Alonso A, Codina C, et al. A modified Delphi study to determine the level of consensus across the European Union on the structures, processes and desired outcomes of the management of polypharmacy in older people. PLoS One 2017 Nov 20;12(11):e0188348 [FREE Full text] [doi: 10.1371/journal.pone.0188348] [Medline: 29155870]
12. Avella JR. Delphi panels: research design, procedures, advantages, and challenges. Int J Dr Stud 2016;11:305-321. [doi: 10.28945/3561]
13. McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. Int J Clin Pharm 2016 Jun;38(3):655-662 [FREE Full text] [doi: 10.1007/s11096-016-0257-x] [Medline: 26846316]
14. Nasa P, Jain R, Juneja D. Delphi methodology in healthcare research: how to decide its appropriateness. World J Methodol 2021 Jul 20;11(4):116-129 [FREE Full text] [doi: 10.5662/wjm.v11.i4.116] [Medline: 34322364]
15. Stability and compatibility of drugs. Stabilis. URL: https://www.stabilis.org/ [accessed 2023-09-15]

XSL•FO
RenderX

16. University of Illinois at Chicago College of Pharmacy, Drug Information Group. Light-sensitive injectable prescription drugs. Hosp Pharm 2014 Feb;49(2):136-163 [FREE Full text] [doi: 10.1310/hpj4902-136] [Medline: 24623868]
17. Elste JM, Ipema HJ, Denton C, Munir F, Alomari R, Dazy A, et al. Light-sensitive injectable prescription drugs-2022. Hosp Pharm 2023 Oct;58(5):448-475. [doi: 10.1177/00185787221133804] [Medline: 37711411]
18. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, et al. Consensus development methods, and their use in clinical guideline development. Health Technol Assess 1998;2(3):i-iv, 1-88 [FREE Full text] [Medline: 9561895]
19. West LM, Diack L, Cordina M, Stewart D. Applying the Delphi technique to define 'medication wastage'. Eur J Hosp Pharm 2015 Mar 19;22(5):274-279. [doi: 10.1136/ejhpharm-2014-000593]
20. Dear healthcare provider letter. U.S. Food & Drug Administration. URL: https://www.fda.gov/media/164425/download [accessed 2026-05-01]
21. Jones J, Hunter D. Consensus methods for medical and health services research. BMJ 1995 Aug 05;311(7001):376-380 [FREE Full text] [doi: 10.1136/bmj.311.7001.376] [Medline: 7640549]
22. Keeling S, Burfield R, Proudlove C, Scales K. The injectable medicines guide website. Br J Nurs 2010;19(19):S25-S28 [FREE Full text] [doi: 10.12968/bjon.2010.19.Sup9.79309] [Medline: 21042244]
23. Kia A, Waterson J, Bargary N, Rolt S, Burke K, Robertson J, et al. Determinants of intravenous infusion longevity and infusion failure via a nonlinear model analysis of smart pump event logs: retrospective study. JMIR AI 2023 Sep 13;2:e48628 [FREE Full text] [doi: 10.2196/48628] [Medline: 38875535]
24. Fonzo-Christe C, Bochaton N, Kiener A, Rimensberger PC, Bonnabry P. Incidence and causes of infusion alarms in a neonatal and pediatric intensive care unit: a prospective pilot study. J Pediatr Pharmacol Ther 2020;25(6):500-506 [FREE Full text] [doi: 10.5863/1551-6776-25.6.500] [Medline: 32839653]
25. Shimoyama S, Takahashi D, Arai S, Asami Y, Nakajima K, Ikeda K, et al. A large amount of microscopic precipitates are inevitably injected during infusion therapy without an in-line filter. Oxf Med Case Reports 2022 Feb 19;2022(2):omab134 [FREE Full text] [doi: 10.1093/omcr/omab134] [Medline: 35198221]
26. Diamond IR, Grant RC, Feldman BM, Pencharz PB, Ling SC, Moore AM, et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. J Clin Epidemiol 2014 Apr;67(4):401-409. [doi: 10.1016/j.jclinepi.2013.12.002] [Medline: 24581294]
27. Niederberger M, Spranger J. Delphi technique in health sciences: a map. Front Public Health 2020 Sep 22;8:457 [FREE Full text] [doi: 10.3389/fpubh.2020.00457] [Medline: 33072683]
28. Klein G. Naturalistic decision making. Hum Factors 2008 Jun 01;50(3):456-460. [doi: 10.1518/001872008x288385]
29. Rycroft-Malone J, Harvey G, Seers K, Kitson A, McCormack B, Titchen A. An exploration of the factors that influence the implementation of evidence into practice. J Clin Nurs 2004 Nov;13(8):913-924. [doi: 10.1111/j.1365-2702.2004.01007.x] [Medline: 15533097]
30. Thompson BW. The transforming effect of handheld computers on nursing practice. Nurs Adm Q 2005;29(4):308-314. [doi: 10.1097/00006216-200510000-00004] [Medline: 16260994]

Abbreviations

IV: intravenous

Edited by A Kushniruk, J Sarvestan; submitted 25.Nov.2025; peer-reviewed by A Dainty; comments to author 20.Apr.2026; revised version received 21.May.2026; accepted 26.May.2026; published 10.Jun.2026.

Please cite as:
Dickman A, Tuffin P, Al-Jaber R, El-Harmeel M, Schneider J, Taladriz-Sender I, Sutherland A, Terkola R, Waterson J
Good Clinical Practice Guidance for Line Materials, Filtration, and Light Protection in Intravenous Medication Administration: Modified Delphi Consensus Study
JMIR Hum Factors 2026;13:e88333
URL: https://humanfactors.jmir.org/2026/1/e88333
doi:10.2196/88333
PMID:

©Andrew Dickman, Penelope Tuffin, Rania Al-Jaber, Mona El-Harmeel, Jennifer Schneider, Irene Taladriz-Sender, Adam Sutherland, Robert Terkola, James Waterson. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 10.Jun.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information,

XSL•FO
RenderX

a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Development of a Blended Physical Activity Intervention for Office Employees Using Intervention Mapping: Intervention Development Study

Yan Sun[1,2], PhD; Alison YT Ou[3], PhD; Tony Lei Shi[4], PhD; Yang Gao[1], PhD; Martin S Hagger[5,6,7,8], PhD; Zhennan Xiong[1], MSocSc; Xiang-yan Chen[9], PhD; Shirley SM Fong[10], PhD; Linda L.D. Zhong[11], PhD; Yao Jie Xie[12], PhD; Julien S. Baker[13], PhD

[1]Academy of Wellness and Human Development, Faculty of Arts and Social Sciences, Hong Kong Baptist University, Hong Kong, China (Hong Kong)

[2]Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

[3]Faculty of Arts and Social Sciences, Hong Kong Baptist University, Hong Kong, China (Hong Kong)

[4]School of General Education, Beijing Normal-Hong Kong Baptist University, Zhu Hai, Guangdong, China

[5]Department of Psychological Science, University of California, Merced, Merced, CA, United States

[6]Health Sciences Research Institute, University of California, Merced, Merced, CA, United States

[7]School of Applied Psychology, Griffith University, Brisbane, Mount Gravatt, Australia

[8]Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland

[9]Division of Science, Engineering and Health Studies (SEHS), The Hong Kong Polytechnic University College of Professional and Continuing Education, Hong Kong, China (Hong Kong)

[10]Department of Health and Physical Education, The Education University of Hong Kong, Hong Kong, China (Hong Kong)

[11]School of Biological Sciences, Nanyang Technological University, Singapore, Singapore

[12]School of Nursing, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[13]Neurology and Endocrinology Departments, Medical Research Centre, Ningbo No.2 Hospital, Ningbo, Zhejiang, China

Corresponding Author:
Yang Gao, PhD
Academy of Wellness and Human Development
Faculty of Arts and Social Sciences
Hong Kong Baptist University
935, AAB, Baptist University Road Campus
Hong Kong
China (Hong Kong)
Phone: 852 34117770
Fax: 852 34117770
Email: gaoyang@hkbu.edu.hk

Abstract

Background: Insufficient engagement in moderate to vigorous physical activity (MVPA) is a significant risk factor for major noncommunicable diseases, including cardiovascular diseases (eg, coronary heart disease and stroke), type 2 diabetes, and several cancers. Physical inactivity accounts for an estimated 3.2 million deaths annually. Office employees, due to their sedentary and desk-based work patterns, are particularly vulnerable to low MVPA levels, which negatively affect health and work productivity. The COVID-19 pandemic exacerbated these issues, further reducing MVPA levels due to lockdowns and work-from-home policies. Although numerous interventions have aimed to promote MVPA, many lack theoretical grounding, stakeholder involvement, or systematic development frameworks. A theory- and evidence-informed approach is warranted to target modifiable determinants and mechanisms, improve coherence and replicability, and enhance effectiveness and scalability.

Objective: This study aimed to design an evidence-based blended intervention to improve MVPA, health, and productivity among office employees, guided by the intervention mapping (IM) framework.

Methods: Following the first 4 steps of the IM framework, we developed a 12-week blended intervention and created a website named “Smartly Active for Health and Life.” Step 1 involved a needs assessment through a literature review and focus group discussions (n=12 office employees) to identify barriers, facilitators, and determinants of MVPA. Step 2 defined intervention

XSL•FO
RenderX

outcomes, performance objectives, and change objectives based on the social ecological model. Step 3 integrated theory- and evidence-based behavior change techniques with delivery methods, resulting in a 3-group randomized controlled trial design: (1) a blended intervention group (web-based sessions and e-workshops), (2) a web-only intervention group, and (3) an active control group. Step 4 developed and refined intervention materials and protocols through a pilot study (n=18). Key behavior change techniques included problem-solving, goal setting, self-monitoring, and habit formation.

Results: The pilot study demonstrated the intervention's acceptability, feasibility, and adaptability. Participants reported increased motivation, improved knowledge of MVPA benefits, and greater confidence in goal setting and self-regulation. Feedback led to refinements such as improved website navigation and interactive content. The intervention is expected to enhance MVPA levels, health, and work productivity by addressing both motivational and volitional phases of behavior change while supporting habitual behaviors.

Conclusions: The IM framework offered a systematic and iterative approach for developing a blended MVPA intervention tailored to the needs of office employees. This study highlights the importance of theory-driven and stakeholder-informed approaches in designing scalable and sustainable health promotion programs. The intervention has the potential to address the critical public health challenge of physical inactivity in office settings and serves as a model for future behavior change interventions.

International Registered Report Identifier (IRRID): RR2-10.1186/s12889-020-09128-z

(JMIR Hum Factors 2026;13:e87328) doi:10.2196/87328

KEYWORDS

physical activity; intervention mapping; behavior change; theory and evidence-based; intervention development; intervention design

Introduction

Physical inactivity is a leading risk factor for global mortality, contributing significantly to common noncommunicable diseases and increased health care burdens [1]. Regular moderate to vigorous physical activity (MVPA) mitigates such risks, improving physical and mental health, work productivity, and overall well-being [2-4]. World Health Organization (WHO) guidelines recommend that adults aged 18 to 64 years engage in 150 to 300 minutes of MVPA per week, but approximately 1.8 billion adults (31% of the global adult population) do not achieve this recommendation [5,6]. A systematic review revealed that working adults engage in an average of 14.7 minutes of MVPA daily, with office employees exhibiting the most sedentary behavior due to their desk-based work and daily tasks [7]. In the COVID-19 pandemic (2020 to 2022), there was a marked decrease in adults' MVPA levels due to lockdowns, closures, and limited mobility, with office employees particularly affected by work-from-home setups, leading to reductions in commuting and leisure activities [8-10].

Many theory-based and non–theory-based interventions have been developed to increase MVPA levels among adults. However, the effectiveness of such interventions remains unclear due to the substantial heterogeneity in their designs and components [11]. Furthermore, many interventions have been developed without reference to sufficient empirical evidence and have frequently failed to include key stakeholders' perspectives [12]. There is evidence that behavioral change interventions based on theoretical frameworks, empirical evidence, and stakeholder engagement are potentially more effective than those lacking a theoretical basis [13-15]. Nevertheless, few well-designed theory-based interventions have been conclusively demonstrated to be effective.

Intervention mapping (IM) is regarded as a systematic framework that guides effective decision-making throughout the development, implementation, and evaluation of evidence-based health promotion and disease prevention programs. It is a stepwise process that uses theory, empirical evidence, and community feedback to develop interventions that are effective, feasible, sustainable, and customized to the needs of the target population [16]. The six steps of IM comprise (1) conducting a needs assessment; (2) specification of intervention outcomes, objectives, and change objectives; (3) selection of theory- and evidence-based change methods and practical modes of delivery; (4) creation and refining of materials and a protocol; (5) development of an implementation plan; and (6) design of an evaluation plan.

There is an urgent need for interventions to increase MVPA levels using the IM framework to bridge existing research gaps and enrich the evidence base. While the protocol and results of the main randomized controlled trial (RCT) examining the effectiveness of the intervention have been published elsewhere [17,18], this paper focuses on the development process of the blended intervention. Specifically, it describes the application of IM steps 1 to 4 (needs assessment, intervention outcomes and objectives, intervention design, and intervention production). The final stages of the IM framework, intervention implementation and evaluation, will be detailed in separate publications.

Methods

Overview

Steps 1 to 4 in this paper involved developing and refining the intervention protocol and materials, with each step building on the results of the previous step. A book on the IM approach [16], which outlines the planning of health promotion programs, served as the foundation for the development of the intervention.

XSL•FO
RenderX

The methodology and tasks of steps 1 to 4 are detailed in Multimedia Appendix 1.

Step 1: Needs Assessment

The first step of the IM framework involved a needs assessment to identify the health problem, justify the promotion of MVPA, and determine the underlying determinants in the target population. This provided a solid foundation for designing an evidence-based and culturally appropriate intervention [16].

Planning Group

A planning group was formed at the start of the process, comprising researchers with expertise in epidemiology, statistics, health promotion, psychology, and sports science, as well as office employees from local companies. These stakeholders ensured all steps of the intervention were well aligned with workplace realities. Weekly meetings focused on planning and brainstorming, transitioning to biweekly sessions for progress tracking. However, due to challenges presented by the COVID-19 pandemic restrictions, managers had to be excluded from the group.

Literature Review

A comprehensive literature review was conducted to better understand the prevalence and effects of insufficient MVPA on health and workplace productivity, as well as the most common barriers and facilitators identified in working adults. The review also explored the evidence on theory- and evidence-based change methods and delivery modes for intervention design, which has been published elsewhere [19]. Six academic databases (Scopus, PsycInfo, PubMed, Embase, Web of Science, and SPORTDiscus) were searched for original studies and systematic reviews published in English before July 30, 2022. Search terms included topics like “physical activity,” “working populations,” “health outcomes,” and “theory-based interventions.” Governmental reports from organizations such as the WHO were also consulted to provide global context. Detailed search strategies can be found in Multimedia Appendix 2.

Focus Group Discussions

Focus group discussions were conducted to complement findings from the literature review and gain detailed insights into the barriers, facilitators, and preferences for interventions specific to insufficiently active office employees. This group was intentionally selected to ensure the study directly addressed the needs of individuals likely to benefit most from tailored interventions. Two focus groups (n=12) were conducted to explore gender-specific barriers and facilitators to MVPA [20,21]. Participants were intentionally recruited via online surveys. Of the 12 participants, 6 (50%) were female, with a mean age of 29 (SD 3.08) years. A total of 9 (75%) held a bachelor’s degree or higher, and 3 (25%) had a diploma or lower educational qualification. The participants provided informed consent to be video recorded during discussions. The focus groups, moderated via Zoom (Zoom Video Communications, Inc) due to COVID-19 pandemic restrictions, lasted 60 to 100 minutes and were facilitated by a trained qualitative researcher, with a second researcher present for note-taking and follow-up questions. The video recordings were transcribed verbatim by 2 trained investigators (AYTO and YS). The transcripts were cross-checked with the video recordings to ensure accuracy. The transcripts were reviewed and analyzed, and the research questions were used as a framework to categorize codes and derive themes from the narrative data [22]. Any discrepancies in interpretation were collaboratively discussed until a consensus was reached.

Step 2: Intervention Outcomes and Objectives

Step 2 focused on defining the desired intervention outcomes and specifying performance objectives (POs) based on the needs assessment conducted in step 1. The planning group developed a matrix of behavioral change objectives by linking these POs with the underlying determinants of MVPA identified earlier, ensuring that the intervention addressed the key factors influencing MVPA behavior.

Step 3: Intervention Design

The purpose of step 3 was to design theory- and evidence-based interventions to promote MVPA, addressing the determinants and objectives identified in step 2. This step involved selecting behavior change methods, linking them to practical applications, and specifying their implementation. The effective theories, behavior change techniques (BCTs; eg, goal setting and self-monitoring), and delivery modes (eg, web-based platforms) were identified from the literature review [19,23,24]. These findings informed the intervention’s design to ensure alignment with evidence-based practices. The planning group held additional meetings to finalize the intervention’s structure, components, and sequence, ensuring feasibility and cultural relevance within workplace settings.

Step 4: Intervention Production

Step 4 focused on preparing the intervention materials, messages, and activities, followed by a 12-week pilot study with 18 participants (6 per group, excluded from the main study) to evaluate the readability of the materials, message execution, and acceptability, feasibility, and adaptability of the intervention. The pilot study also assessed participant recruitment and outcome measures, while collecting feedback from participants and implementers. On the basis of the pilot results, the intervention materials, protocols, and implementation plans were refined to optimize delivery and ensure smooth execution.

The original study protocol [17], developed before the COVID-19 pandemic, had to be revised due to challenges caused by the pandemic. Between June and August 2022, despite reaching out to approximately 2000 companies, none agreed to participate, mainly due to concerns about COVID-19 transmission and disruptions in work routines. As a result, the original cluster-RCT design was replaced with an individualized RCT. Key changes to the study design are summarized in Multimedia Appendix 3.

Ethical Considerations

This study received ethics approval from the Research Ethics Committee of Hong Kong Baptist University (HASC/17-18/0529). Electronic informed consent was obtained from the 18 participants included in the pilot study after they were informed of the study objectives, procedures, potential

XSL•FO
RenderX

risks and benefits, data use, confidentiality protections, and their right to withdraw. Identifiable information was stored separately from research data and will be destroyed after study completion. Deidentified datasets will be retained for research purposes and may be available from the corresponding author upon reasonable request. No identifiable participant information or images are included in this manuscript or supplementary materials. Participants received HK $50 (HK $1=US $0.13 as of June 17, 2026) coupons for each completed evaluation session and a personalized physical activity and health tracking report as part of the study benefits.

Results

Step 1: Needs Assessment

The literature review analyzed 183 articles spanning observational, intervention-based, and systematic review studies. Key themes identified barriers and facilitators to MVPA participation, categorized into personal, interpersonal, and environmental levels following the social ecological model [16]. These insights, outlined in Table 1, were enriched by focus group discussions to create a tailored intervention addressing the needs of inactive office employees.

Table 1. Facilitators of and barriers to moderate to vigorous physical activity (MVPA) participation identified from literature reviews and focus groups.

	Facilitators of MVPA participation	Barriers to MVPA participation
Personal level [25-30]	• Knowledge of the benefits of MVPA (physical and mental health and disease prevention) • Self-monitoring (eg, track progress and record workouts) • Time availability (time management) • Self-efficacy and confidence • Improve sleep quality • Improve self-esteem and sense of achievement • Workplace well-being and quality of life • Being active with children	• Lack of knowledge and awareness (importance of participating in MVPA) • Lack of time or competing demands on time • Lack of motivation (eg, lack of interest and low self-efficacy) • Lack self-monitoring and goal setting • Health limitations or concerns (eg, health conditions, physical limitations, and injuries) • Psychological barriers (eg, mental health concerns, perceived lack of energy, fatigue, and concerns about sweating) • Safety concerns • Minimize exposure to infection during the COVID-19 pandemic • Unpredictable breaks and shifts (eg, certain life changes or events)
Interpersonal level [25-27,31]	• Body image and societal beauty standards • Religious and cultural norms • Mentoring and coaching (professional supervision) • Social ties • Positive support (social, emotional, and psychological)	• Lack of social support • Negative social influences (eg, peer pressure and lack of positive role models)
Environmental level [25,26,28,32]	• Accessibility of resources (eg, facilities, supportive environments, and professional supervision) • Leadership and institutional support • Supportive policies (eg, reminders for work breaks)	• Limited access to facilities, environment, and space • Lack of professional-development opportunities • Lack of institutional support • Lack of financial resources • Inclement weather (eg, rain and extreme heat) • COVID-19 pandemic policy restrictions (eg, closure of sports venues and exercise facilities to reduce infection risk)
Additional findings from focus groups	• Young, energetic, and confident • Improve work productivity • Master motor and athletic skills • Interesting and engaging • Incentives • Policies and flexible work arrangements • Environmental modifications	• No additional information was added from focus groups.

Step 2: Intervention Outcomes and Objectives

In step 2, outcomes and POs were defined, leading to the creation of a matrix of change objectives. The primary outcome was expected to be an increase in MVPA levels. The secondary outcomes were expected to be improvements in health and work productivity. The POs developed in Table 2 were aligned with the determinants identified in step 1 to construct a matrix of change objectives (Multimedia Appendix 4). For example, PO5 (Table 3) focuses on action planning and highlights determinants such as knowledge, self-efficacy, and social support.

XSL•FO
RenderX

Table 2. Performance objectives (POs) for promoting moderate to vigorous physical activity (MVPA) levels among inactive office employees.

Code	Performance objective
PO1	Possess knowledge of MVPA and its impacts
PO2	Generate motivation to promote MVPA
PO3	Monitor personal MVPA levels
PO4	Set long-term and short-term MVPA goals
PO5	Make specific action plans
PO6	Develop self-monitoring skills and capabilities (eg, progress feedback)
PO7	Evaluate goals and revise action plans
PO8	Address identified barriers with coping plans
PO9	Identify prompts and cues for being active
PO10	Prevent habit reversal (relapse prevention)
PO11	Engage in and maintain sufficient MVPA in daily life
PO12	Serve as positive role models and leaders
PO13	Seek support and motivation from family, friends, and colleagues to engage in MVPA

Table 3. Example of change objectives for performance objective PO5.

Determinant	Change objective
Knowledge	Provide employees with tools and strategies for effective time management, and training and guidance on setting specific, measurable, achievable, relevant, and time-bound goals aligned with their desired outcomes
Self-efficacy	Enhance employees' belief in their ability to create and execute action plans by enhancing confidence and providing support through skill-building exercises and positive reinforcement
Self-monitoring	Assist individuals in developing self-monitoring strategies to track their progress and adherence to action plans, such as using activity trackers, journals, or digital apps
Social support	Enhance peer interaction, foster a more supportive and collaborative environment, and encourage action planning with peers, colleagues, and family

Step 3: Intervention Design

We designed a 3-group RCT consisting of a web-based intervention group, a blended intervention (web-based intervention combined with e-workshops) group, and an active control group. Web-based interventions are widely adopted for their accessibility, convenience, and ability to provide personalized content; however, these interventions are often hindered by high dropout rates and limited interactivity, which may reduce user engagement and effectiveness [33]. In contrast, workplace-based live workshops offer direct knowledge transfer, support, and interaction, which can enhance motivation and engagement [34-37]. By integrating these methods, the blended intervention aimed to combine the strengths of both approaches, improving its overall effectiveness. Additionally, insights from focus group discussions were considered when designing the intervention. Participants expressed a preference for textual content supported by a few simple and essential images, which informed the design of the web-based materials.

After specifying the change objectives, appropriate theoretical change methods were selected from the literature review [19]. To address the intention-behavior gap, our intervention incorporated dual-process theories, combining reflective (eg, deliberate self-regulation) and automatic (eg, habit formation) components [38,39]. Three complementary strategies, motivation, self-regulation, and habit formation, were integrated to enhance MVPA levels and sustained engagement. These strategies were operationalized using BCTs to ensure coherence, usability, and effectiveness [40,41]. Regarding the web-based intervention content, a detailed breakdown of the 6 sessions, including the timelines, questions asked, personalized feedback, and BCTs used, was provided (Multimedia Appendix 5). For example, in session 3 (week 5), action planning (1.4) was implemented by encouraging participants to articulate their MVPA goals by answering specific prompts, such as "What activity will you do?", "Where will you do it?", and "When will you do it?" Participants were further instructed to document their plans and place them in a visible location as regular reminders. Additionally, feedback on the evaluation and adjustment of action plans (eg, reviewing goal progress and refining prompts) helped reinforce participants' motivation and habit development.

The control group was designed as an active control, incorporating 3 BCTs: information about health consequences (5.1), social and environmental consequences (5.3), and emotional consequences (5.6) [42]. A total of 18 essays containing general, nonpersonalized information were provided to participants. Three essays were delivered every 2 weeks across 6 sessions. The essays covered a variety of topics, such as the recommended MVPA levels, the health and productivity

XSL•FO
RenderX

benefits of MVPA, the importance of time management for staying active, strategies for injury prevention, and the advantages of leading an active and healthy lifestyle. The content was factual and designed to be informative, focusing on general knowledge rather than individualized guidance.

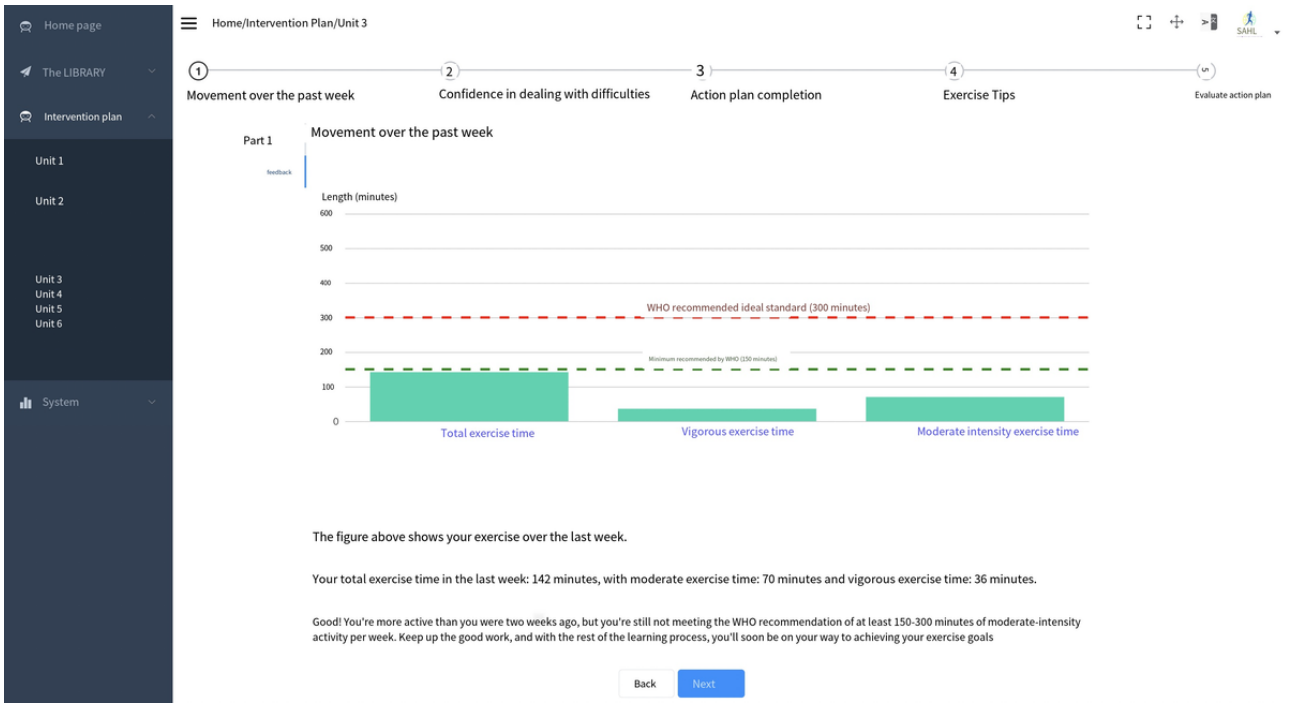
The topics in the e-workshops for the blended intervention included understanding MVPA, building self-efficacy, establishing habits, setting goals, developing action plans, identifying barriers, creating coping strategies, using prompts, and fostering social support. Each topic was carefully designed to address participants' needs and challenges, providing a comprehensive framework to encourage long-term engagement in MVPA. To enhance the effectiveness of these workshops, 4 additional BCTs were applied: problem-solving (1.2), instruction on how to perform the behavior (4.1), reattribution (4.3), and mental rehearsal of successful performance (15.2) [42]. For instance, problem-solving was integrated into sessions focused on identifying barriers and creating coping strategies, helping participants develop "if-then" plans (eg, "If I have only 15 minutes during lunch, then I will take a brisk walk around the office building"). Similarly, mental rehearsal was used to support participants in visualizing success when setting goals or enacting action plans, boosting their confidence and motivation. By embedding these BCTs within the core topics, the intervention ensured that theoretical strategies were effectively applied to practical, real-life situations. This design not only reinforced behavioral changes but also provided interactive and personalized elements to address participants' unique challenges.

A 12-week intervention period was selected, as it aligns with durations commonly used in prior web-based physical activity studies. This timeframe is generally considered sufficient for participants to adapt to new exercise routines and form habits, which research suggests may take a median of 66 days [43,44]. A 12-week follow-up period was included to assess the maintenance of behavioral changes and monitor any health outcomes resulting from the intervention. These collaborative design efforts enhanced the practicality of the intervention and participants' motivation to sustain their plans.

Step 4: Intervention Production

We developed materials and messages for the web-based intervention that were informed by the outcomes of the first 3 steps and organized them into 6 distinct sessions. These sessions were deployed every 2 weeks via the website named "Smartly Active for Health and Life." For reference, a screenshot of session 3 from the "Intervention" subsection has been translated from Traditional Chinese using YouDao Translate (NetEase Youdao; Figure 1). The website was structured into 2 main subsections: an "Intervention" subsection and a "Library" subsection. The participants in the blended intervention group and the web-based intervention group were able to access both subsections, whereas the participants in the control group were only able to access the "Library" subsection. Eighteen essays were included in the "Library" subsection, with 3 essays available in each session. The e-workshops were delivered in weeks 2, 4, and 8.

Figure 1. A screenshot of session 3 in the "Intervention" subsection on the website (translated from Traditional Chinese by "YouDao" translator).



The planning group scrutinized and refined all materials before launching the pilot study. Recruitment posters were disseminated on social media platforms, including WeChat (Tencent Holdings Ltd), Instagram (Meta Platforms, Inc), and Facebook (Meta Platforms, Inc). A total of 18 participants were recruited for the pilot, of whom 15 (83.3%) completed the study and provided feedback. The retention and engagement rates were 83.3% (15/18) and 73% (92/126), respectively, demonstrating the intervention's feasibility and acceptability. Participants rated the intervention highly in terms of usefulness (13/15, 86.7%),

XSL•FO
RenderX

enjoyability (13/15, 86.7%), and time acceptability (14/15, 93.3%).

Participant feedback led to minor improvements, such as adding navigation aids to the website and redesigning the "Library" section by replacing long text blocks with clickable links to enhance usability and engagement. Observations during trial implementation also captured implementers' perceptions, focusing on trialability, observability, and potential challenges under COVID-19 pandemic conditions. On the basis of these findings, strategies such as implementer training plans were developed to improve the intervention's acceptability, implementation, and sustainability.

Discussion

Principal Findings

In this study, we developed a blended intervention to increase MVPA among physically inactive office employees using IM steps 1 to 4. The main outputs were (1) a logic model of the problem and a set of key determinants identified through the needs assessment, (2) POs and change objectives mapped to these determinants, (3) a set of theory- and evidence-informed intervention components and materials delivered in a blended format, and (4) feedback during the optimization and pilot phase used to refine the content and delivery of the intervention.

Our selection of IM as our guiding framework significantly influenced our findings and led us to gain critical insights into the final outcomes and the development of our interventions [16]. In step 1, we conducted a needs assessment primarily using a literature review and focus group discussions. We also formed a planning group comprising professionals from various disciplines and key stakeholders. This led to the generation of a comprehensive, acceptable, and resource-effective intervention plan. It also promoted consensus among the members, leading to a unified approach to the implementation of the intervention. Moreover, the needs assessment revealed knowledge gaps about MVPA and attitudes that could potentially hinder engagement in MVPA, which shaped the educational and motivational components of our intervention. The information gathered was vital in tailoring the intervention to meet the unique needs of our target population, thereby enhancing its relevance and value. In step 2, we ensured that the change objectives aligned closely with the identified determinants and intervention goals. This alignment was crucial to the success of the intervention as it ensured that the approach was focused on achieving meaningful outcomes while meeting specific needs. In step 3, we translated the broad goals and objectives identified in the previous steps into concrete strategies for change. These strategies formed the backbone of our intervention, providing a clear road map for its implementation. In step 4, intervention components and materials were produced and subjected to several rounds of feedback and revision. This iterative process ensured the quality and appropriateness of our intervention. Additionally, the pilot study results indicated high levels of acceptance and satisfaction among the participants, suggesting that our intervention was well received and had achieved their intended goals. Our study aligns with a growing body of research that has used the IM framework for developing disease prevention interventions [15]. For instance, a previous study used IM to create a workplace health promotion intervention targeting enhanced presenteeism. It was found that the systematic nature of IM enabled the unique needs and preferences of workers to be addressed, which resonates with our own results [45]. Nevertheless, although IM is a systematic process for intervention development, outcomes can vary widely depending on numerous factors. This reinforces the need for continued research to refine and optimize health interventions in diverse settings.

The strengths of this study lie in its systematic and theory-driven approach. Using the IM framework, we integrated stakeholder insights and behavioral theories to design an intervention supporting both the initiation and maintenance of MVPA. The flexibility of the framework during the COVID-19 pandemic allowed adaptations to online sessions and home-based exercise, ensuring feasibility and relevance in dynamic conditions.

However, despite these strengths, several limitations should be noted. First, the IM framework, while robust and comprehensive, is highly resource-intensive, requiring significant time, expertise, and financial resources, making it challenging to replicate in resource-limited settings [12,46,47]. Second, the needs assessment relied on a small focus group sample (n=12), limiting the representativeness of findings and potentially overlooking diverse perspectives [45]. Additionally, since participants were individuals who did not meet WHO's MVPA guidelines, the findings may not generalize to populations already achieving recommended activity levels. Third, COVID-19 pandemic–related adaptations, while ensuring feasibility, altered the original intervention design, limiting its generalizability to nonpandemic contexts. For instance, workplace-level components could not be implemented due to restrictions.

Our findings have important implications for advancing health behavior change interventions in both research and practice. From a research perspective, applying the IM framework strengthened the rigor and transparency of our study while facilitating replication and exploration of mechanisms that contribute to intervention efficacy. From a practical perspective, the blended intervention outlined in this study offers a feasible road map for office employees to incorporate more MVPA into their routines, thereby improving health, productivity, and organizational outcomes. Additionally, these insights could guide policymakers in creating supportive environments that encourage MVPA in workplace settings.

Future research should address the broader applicability of IM-designed interventions across diverse populations and health behaviors. Specifically, studies should explore whether such interventions are more effective than other approaches in areas like disease management, mental health, and among vulnerable groups, such as children and older adults. Incorporating multilevel strategies that address organizational and environmental components, alongside individual-level approaches, could improve ecological validity. For instance, school-based interventions could benefit from engaging stakeholders, such as parents and teachers, to scale and adapt programs effectively. Similarly, implementing cluster RCTs in real-world settings is crucial to assessing the relevance, scalability, and sustainable impact of such interventions.

XSL•FO
RenderX

Conclusions

This study demonstrates the value of the IM framework as a systematic, stepwise tool for designing blended interventions to promote MVPA among office employees. The framework's ability to integrate evidence and stakeholder insights while adapting to dynamic contexts, such as the COVID-19 pandemic, highlights its versatility and practical relevance. These findings underscore the framework's potential for developing scalable, long-term health promotion programs to improve both individual well-being and workplace productivity.

Future research should expand on this work by testing IM-based interventions in diverse populations and real-world contexts to assess scalability and broader applicability. By maintaining a continuously updated evidence base, the IM framework could enhance the design and impact of interventions across changing circumstances.

Funding

This study received support from the General Research Fund, University Grants Committee of Hong Kong (project number 12609919). The funding body had no role in study design, collection, management, analysis, and interpretation of data, writing of the manuscript, or the decision to submit the manuscript for publication. The funding body had no ultimate authority over any of these activities.

Data Availability

Data sharing is not applicable to this paper as no datasets were generated or analyzed during this study.

Authors' Contributions

Conceptualization: YG (lead), MSH (equal).
Methodology: YG (lead), YS (equal), MSH (equal).
Investigation: YS (lead), AYTO (equal), ZX (equal), TLS (equal), XC (supporting), SSMF (supporting), YJX (supporting).
Data curation: AYTO (lead), ZX (equal), TLS (equal).
Project administration: YS (lead), YG (equal).
Resources: SSMF, XC, YJX.
Supervision: YG (lead), MSH (equal).
Validation: JSB (lead), YS (supporting).
Visualization: YS (lead), ZX (supporting).
Writing – original draft: YS.
Writing – review & editing: YG (lead), JSB (supporting), XC (supporting), SSMF (supporting), YJX (supporting), TLS (supporting), LLDZ (supporting).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Intervention mapping steps and tasks.
[PNG File , 65 KB - humanfactors_v13i1e87328_app1.png]

Multimedia Appendix 2
Search strategy for the literature review.
[DOCX File , 19 KB - humanfactors_v13i1e87328_app2.docx]

Multimedia Appendix 3
Protocol adjustments due to the COVID-19 pandemic.
[DOCX File , 15 KB - humanfactors_v13i1e87328_app3.docx]

Multimedia Appendix 4
A matrix of change objectives for each performance objective cross-referenced with determinants.
[DOCX File , 21 KB - humanfactors_v13i1e87328_app4.docx]

Multimedia Appendix 5
Timeline, questions, personalized feedback, and behavior change techniques involved in 6 web-based sessions in 2 intervention groups.

XSL•FO
RenderX

[DOCX File , 21 KB - humanfactors_v13i1e87328_app5.docx]

References

1. Noncommunicable diseases. World Health Organization. URL: https://www.who.int/en/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2026-06-13]
2. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob Health 2018 Oct;6(10):e1077-e1086. [doi: 10.1016/s2214-109x(18)30357-7]
3. Warburton DE, Bredin SS. Reflections on physical activity and health: what should we recommend? Can J Cardiol 2016 Apr;32(4):495-504. [doi: 10.1016/j.cjca.2016.01.024] [Medline: 26995692]
4. Pedersen BK, Saltin B. Exercise as medicine - evidence for prescribing exercise as therapy in 26 different chronic diseases. Scand J Med Sci Sports 2015 Dec;25 Suppl 3:1-72. [doi: 10.1111/sms.12581] [Medline: 26606383]
5. Physical activity. World Health Organization. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/physical-activity [accessed 2026-06-11]
6. Physical activity. World Health Organization. URL: https://www.who.int/health-topics/physical-activity#tab=tab_2 [accessed 2026-06-15]
7. Prince SA, Elliott CG, Scott K, Visintini S, Reed JL. Device-measured physical activity, sedentary behaviour and cardiometabolic health and fitness across occupational groups: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2019 Apr 02;16(1):30 [FREE Full text] [doi: 10.1186/s12966-019-0790-9] [Medline: 30940176]
8. Fukushima N, Machida M, Kikuchi H, Amagasa S, Hayashi T, Odagiri Y, et al. Associations of working from home with occupational physical activity and sedentary behavior under the COVID-19 pandemic. J Occup Health 2021 Jan;63(1):e12212 [FREE Full text] [doi: 10.1002/1348-9585.12212] [Medline: 33683779]
9. Prince SA, Rasmussen CL, Biswas A, Holtermann A, Aulakh T, Merucci K, et al. The effect of leisure time physical activity and sedentary behaviour on the health of workers with different occupational physical activity demands: a systematic review. Int J Behav Nutr Phys Act 2021 Jul 20;18(1):100 [FREE Full text] [doi: 10.1186/s12966-021-01166-z] [Medline: 34284795]
10. Ráthonyi G, Kósa K, Bács Z, Ráthonyi-Ódor K, Füzesi I, Lengyel P, et al. Changes in workers' physical activity and sedentary behavior during the COVID-19 pandemic. Sustainability 2021 Aug 24;13(17):9524. [doi: 10.3390/su13179524]
11. Lock M, Post D, Dollman J, Parfitt G. Efficacy of theory-informed workplace physical activity interventions: a systematic literature review with meta-analyses. Health Psychol Rev 2021 Dec;15(4):483-507. [doi: 10.1080/17437199.2020.1718528] [Medline: 31957559]
12. Hazelzet E, Houkes I, Bosma H, de Rijk A. Using intervention mapping to develop 'Healthy HR' aimed at improving sustainable employability of low-educated employees. BMC Public Health 2021 Jun 29;21(1):1259 [FREE Full text] [doi: 10.1186/s12889-021-11278-7] [Medline: 34187445]
13. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
14. McEwan D, Beauchamp MR, Kouvousis C, Ray CM, Wyrough A, Rhodes RE. Examining the active ingredients of physical activity interventions underpinned by theory versus no stated theory: a meta-analysis. Health Psychol Rev 2019 Mar;13(1):1-17. [doi: 10.1080/17437199.2018.1547120] [Medline: 30412685]
15. Garba RM, Gadanya MA. The role of intervention mapping in designing disease prevention interventions: a systematic review of the literature. PLoS One 2017 Mar 30;12(3):e0174438 [FREE Full text] [doi: 10.1371/journal.pone.0174438] [Medline: 28358821]
16. Bartholomew-Eldredge LK, Markham CM, Ruiter RA, Fernandez ME, Kok G, Parcel GS. Planning Health Promotion Programs: An Intervention Mapping Approach. Hoboken, NJ: John Wiley & Sons; 2016.
17. Sun Y, Wang A, Yu S, Hagger MS, Chen X, Fong SS, et al. A blended intervention to promote physical activity, health and work productivity among office employees using intervention mapping: a study protocol for a cluster-randomized controlled trial. BMC Public Health 2020 Jun 25;20(1):994 [FREE Full text] [doi: 10.1186/s12889-020-09128-z] [Medline: 32586293]
18. Sun Y, Gao Y, Ou AY, Wang X, Xie Y, Chen XY, et al. Effectiveness of a blended intervention to promote physical activity among office employees: randomized controlled trial. J Med Internet Res 2026 May 22;28:e80249 [FREE Full text] [doi: 10.2196/80249] [Medline: 42172646]
19. Sun Y, Gao Y, Yu S, Wang A, Ou X, Tao D, et al. Promoting physical activity among workers: a review of literature and future directions on developing theory-based interventions. Int J Environ Res Public Health 2022 Oct 20;19(20):13594 [FREE Full text] [doi: 10.3390/ijerph192013594] [Medline: 36294174]
20. Guest G, Namey E, McKenna K. How many focus groups are enough? Building an evidence base for nonprobability sample sizes. Field Methods 2016 Apr 28;29(1):3-22. [doi: 10.1177/1525822X16639015]
21. Morgan DL. Focus Groups as Qualitative Research. Thousand Oaks, CA: SAGE Publications; 1997.

XSL•FO
RenderX

22. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
23. McKenzie JF, Neiger BL, Thackeray R. Planning, Implementing, and Evaluating Health Promotion Programs: A Primer. San Francisco, CA: Pearson/Benjamin Cummings; 2009.
24. Hagger M, Cameron L, Hamilton K, Hankonen N, Lintunen T. The Handbook of Behavior Change. Cambridge, UK: Cambridge University Press; 2020.
25. Planchard JH, Corrion K, Lehmann L, d'Arripe-Longueville F. Worksite physical activity barriers and facilitators: a qualitative study based on the transtheoretical model of change. Front Public Health 2018 Nov 15;6:326 [FREE Full text] [doi: 10.3389/fpubh.2018.00326] [Medline: 30525018]
26. Leese C, Abraham K, Smith BH. Narrative review – barriers and facilitators to promotion of physical activity in primary care. Lifestyle Med 2023 May 19;4(3):e81. [doi: 10.1002/lim2.81]
27. Safi A, Cole M, Kelly AL, Zariwala MG, Walker NC. Workplace physical activity barriers and facilitators: a qualitative study based on employees physical activity levels. Int J Environ Res Public Health 2022 Aug 01;19(15):9442 [FREE Full text] [doi: 10.3390/ijerph19159442] [Medline: 35954798]
28. Hunter JR, Gordon BA, Bird SR, Benson AC. Perceived barriers and facilitators to workplace exercise participation. Int J Workplace Health Manag 2018 Oct 11;11(5):349-363. [doi: 10.1108/IJWHM-04-2018-0055]
29. Physical activity reduces risk of 13 types of cancer. Just one year of regular exercise can reduce the risk. Hong Kong Anti-Cancer Society. 2017. URL: https://hkacs.org.hk/en/medicalnews.php?id=215 [accessed 2026-06-17]
30. Survey findings on attitudes towards physical exercises in Hong Kong. Hong Kong Institute of Asia-Pacific Studies at Chinese University of Hong Kong. 2016. URL: https://tinyurl.com/2dtvcnr8 [accessed 2026-06-15]
31. Choi J, Cho J, Shin NM, Tsoh J. Exploring barriers to and facilitators of physical activity among Korean American women. West J Nurs Res 2021 Sep;43(9):817-827. [doi: 10.1177/0193945920980453] [Medline: 33317420]
32. Survey says nearly 70% of Hongkongers change their exercise and eating habits due to COVID-19. Hong Kong Anti-Cancer Society. 2020 Dec 21. URL: https://hkacs.org.hk/en/medicalnews.php?id=239 [accessed 2026-06-14]
33. Schröer S, Haupt J, Pieper C. Evidence-based lifestyle interventions in the workplace--an overview. Occup Med (Lond) 2014 Jan;64(1):8-12. [doi: 10.1093/occmed/kqt136] [Medline: 24280187]
34. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
35. Hattie J, Timperley H. The power of feedback. Rev Educ Res 2007 Mar;77(1):81-112. [doi: 10.3102/003465430298487]
36. Johnson DW, Johnson RT, Smith KA. Cooperative learning returns to college what evidence is there that it works? Change 2010 Mar 25;30(4):26-35. [doi: 10.1080/00091389809602629]
37. Kolb DA. Experiential Learning: Experience as the Source of Learning and Development. London, UK: Pearson Education; 2014.
38. Strack F, Deutsch R. Reflective and impulsive determinants of social behavior. Pers Soc Psychol Rev 2004;8(3):220-247 [FREE Full text] [doi: 10.1207/s15327957pspr0803_1] [Medline: 15454347]
39. Rebar AL, Dimmock JA, Jackson B, Rhodes RE, Kates A, Starling J, et al. A systematic review of the effects of non-conscious regulatory processes in physical activity. Health Psychol Rev 2016 Dec;10(4):395-407. [doi: 10.1080/17437199.2016.1183505] [Medline: 27118430]
40. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
41. Michie S, Johnston M. Theories and techniques of behaviour change: developing a cumulative science of behaviour change. Health Psychol Rev 2012;6(1):1-6. [doi: 10.1080/17437199.2012.654964]
42. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
43. Joseph RP, Durant NH, Benitez TJ, Pekmezi DW. Internet-based physical activity interventions. Am J Lifestyle Med 2014 Jan;8(1):42-68 [FREE Full text] [doi: 10.1177/1559827613498059] [Medline: 25045343]
44. Lally P, van Jaarsveld CH, Potts HW, Wardle J. How are habits formed: modelling habit formation in the real world. Eur J Soc Psychol 2009 Jul 16;40(6):998-1009. [doi: 10.1002/ejsp.674]
45. Ammendolia C, Côté P, Cancelliere C, Cassidy JD, Hartvigsen J, Boyle E, et al. Healthy and productive workers: using intervention mapping to design a workplace health promotion and wellness program to improve presenteeism. BMC Public Health 2016 Nov 25;16(1):1190 [FREE Full text] [doi: 10.1186/s12889-016-3843-x] [Medline: 27884132]
46. McEachan RR, Lawton RJ, Jackson C, Conner M, Lunt J. Evidence, theory and context: using intervention mapping to develop a worksite physical activity intervention. BMC Public Health 2008 Sep 22;8:326 [FREE Full text] [doi: 10.1186/1471-2458-8-326] [Medline: 18808709]
47. Golsteijn RH, Bolman C, Volders E, Peels DA, de Vries H, Lechner L. Development of a computer-tailored physical activity intervention for prostate and colorectal cancer patients and survivors: OncoActive. BMC Cancer 2017 Jun 26;17(1):446 [FREE Full text] [doi: 10.1186/s12885-017-3397-z] [Medline: 28651586]

XSL•FO
RenderX

Abbreviations

BCT: behavior change technique
IM: intervention mapping
MVPA: moderate to vigorous physical activity
PO: performance objective
RCT: randomized controlled trial
WHO: World Health Organization

Edited by A Kushniruk; submitted 10.Nov.2025; peer-reviewed by E Neter, A Woodward; comments to author 31.Mar.2026; revised version received 12.May.2026; accepted 29.May.2026; published 14.Jul.2026.

Please cite as:
Sun Y, Ou AYT, Shi TL, Gao Y, Hagger MS, Xiong Z, Chen XY, Fong SSM, Zhong LL, Xie YJ, Baker JS
Development of a Blended Physical Activity Intervention for Office Employees Using Intervention Mapping: Intervention Development Study
JMIR Hum Factors 2026;13:e87328
URL: https://humanfactors.jmir.org/2026/1/e87328
doi:10.2196/87328
PMID:

©Yan Sun, Alison YT Ou, Tony Lei Shi, Yang Gao, Martin S Hagger, Zhennan Xiong, Xiang-yan Chen, Shirley SM Fong, Linda L.D. Zhong, Yao Jie Xie, Julien S. Baker. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 14.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Investigating Context in How Clinicians Make Sense of AI: Qualitative Interview Study

Simone Mingels[1,2*], MSc; Hannah Piehl[1,2*], MA; Rianne Fijten[1,2], PhD; Ricky Janssen[3,4], PhD; Bart Penders[3], PhD

[1]Department of Radiation Oncology, GROW Research Institute for Oncology and Reproduction, Maastricht University Medical Centre+, Paul-Henri Spaaklaan 1, Maastricht, Limburg, The Netherlands

[2]Maastro, Maastricht Radiation Oncology Institute, Maastricht, The Netherlands

[3]Department of Health, Ethics & Society, Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, The Netherlands

[4]Department of Health, Ethics & Society, Mental Health and Neuroscience Research Institute (MHeNs), Maastricht University, Maastricht, The Netherlands

[*]these authors contributed equally

Corresponding Author:
Simone Mingels, MSc
Department of Radiation Oncology, GROW Research Institute for Oncology and Reproduction, Maastricht University Medical Centre+, Paul-Henri Spaaklaan 1, Maastricht, Limburg, The Netherlands

Abstract

Background: AI-driven clinical decision support systems (AI-CDSS) are increasingly being introduced to enhance clinical decision-making, yet their adoption in health care remains limited. Research has examined technical, organizational, and contextual barriers to adoption, but less attention has been paid to how knowledge, values, and responsibility around AI-CDSS are constructed and negotiated in practice.

Objective: This study examines how clinicians and AI experts make sense of AI-CDSS by reanalyzing qualitative interview data through a feminist and constructivist science and technology studies (STS) perspective. Whereas the original study examined informational needs, this reanalysis asks how participants construct relevant knowledge, evaluate the value of AI-CDSS, and negotiate appropriate use and responsibility in clinical practice.

Methods: The authors conducted a reanalysis of semistructured interviews held in 2024 with clinicians (n=8) and AI experts (n=8) in the Netherlands. The original study focused on informational needs for AI-CDSS. For this analysis, a deductive coding framework based on European Commission guidelines for trustworthy AI was first applied, and the material was then inductively reorganized into 5 empirical themes. Subsequently, feminist and constructivist STS concepts were used as sensitizing concepts in an abductive analysis, and these concepts were brought into dialogue with the empirical themes to examine the sociotechnical relations underlying participants' accounts.

Results: *Situated knowledge* shows that the informational needs of clinicians are not universal but relate to their roles and specific clinical contexts, with understanding oriented less toward technical transparency than toward actionable guidance, relevance, and integration into decision-making. *Standpoint theory* illuminates differences between stakeholder priorities: clinicians emphasize information on how to use and act on AI outputs within existing workflows, whereas AI experts focus more on providing technical characteristics and performance metrics. *Tinkering* captures how clinicians experiment with AI systems in practice to develop situated, lay forms of expertise. *Situated value* highlights that perceptions of an AI-CDSS's worth are relational, with clinicians valuing novel, context-sensitive insights and developers prioritizing performance metrics. Finally, *response-ability* reveals how trust and accountability are largely delegated to institutional structures, while clinicians report limited legal, organizational, and structural support to respond meaningfully to uncertainty or errors when using AI-CDSS.

Conclusions: Reanalyzing the same interview material through an abductive STS approach makes sociotechnical dimensions visible that were not the focus of the original analysis. Rather than treating informational requirements as universal or adoption as primarily a matter of technical performance and communication, the findings show how knowledge, value, and responsibility are constituted through situated professional practices and institutional arrangements. Processes and considerations regarding how clinicians make sense of AI-CDSS are not universal but are shaped by social, cultural, and institutional contexts. This analytical perspective underscores the need for context-sensitive, multistakeholder strategies to support meaningful and responsible AI adoption in clinical practice.

(JMIR Hum Factors 2026;13:e99117) doi:10.2196/99117

KEYWORDS

AI; decision support; science and technology studies; human-computer interaction; sociotechnical systems; adoption

XSL•FO
RenderX

Introduction

AI is becoming an integral part of health care, with the potential to revolutionize clinical decision-making. By predicting outcomes and offering treatment options tailored to individual patients, AI-driven clinical decision support systems (AI-CDSS) promise a more personalized and efficient approach [1-3]. However, despite its potential, the adoption of AI-CDSS in clinical practice remains slow [4].

This slow or lack of adoption has been studied thoroughly by examining factors such as technical performance, data quality, workflow integration, and regulatory compliance [5]. These studies have highlighted that AI can be difficult for clinicians to interpret and explain, and that its integration often demands significant adjustments to established clinical routines [6].

Considerations about algorithmic bias, legal liability, and ethical concerns, as well as uncertainty about the effects of these systems regarding roles and responsibilities, further complicate the adoption of AI in health care [7-9]. While these studies provide valuable insights into social, organizational, and contextual factors, alongside technical characteristics, they often tend to conceptualize these dimensions as determinants, barriers, or facilitators of adoption. This overlooks the ways in which trust, professional identity, and medical practices shape how AI is understood and used, as well as processes through which stakeholders negotiate what constitutes useful, trustworthy, and appropriate AI in clinical practice. Addressing these dimensions requires approaches that explore not only the functionality of AI systems but also how they fit into the social and institutional environments of health care to contribute to meaningful forms of adoption.

Human-computer interaction (HCI) and technology adoption research are among the fields that have increasingly sought to address AI as a sociotechnical system. A sociotechnical perspective holds that social and technological dimensions are inseparable and must be examined in their mutual constitution rather than in isolation [10]. However, despite this shift toward relating social and technical factors to each other, HCI operationalizes social context through structured frameworks, such as the technology acceptance model [11] and the unified theory of acceptance and use of technology [12], as well as determinants intended to support implementation or adoption. As a result, much of the literature on clinical decision support systems (CDSS) continues to rely on structured frameworks or models to incorporate social and contextual factors [13-15]. In this paper, we turn to perspectives from other disciplines that foreground questions of power, governance, and contestation in the development and use of AI-based systems in health care—questions that are less often the primary object of technology adoption–oriented research [16,17].

Science and technology studies (STS) provides a distinct perspective by examining the relationship between science, technology, and the societal structures within which they are developed and deployed. STS is an interdisciplinary research field exploring how scientific knowledge and technological advancements are shaped by, and in turn shape, cultural, political, and economic contexts [18]. Closely related to critical branches of the public understanding of science (PUS), STS highlights that science and technology are intertwined with historical, societal, and cultural dimensions and cannot be viewed separately from the sociotechnical surroundings within which they are constructed. Moreover, STS centers on how "scientific and technological configurations of illness and treatment are themselves charged with social meaning and importance" (p. 469) [19]. Therefore, understanding AI adoption involves not only identifying social, technical, or organizational factors that influence implementation but also revolves around questions such as: Who decides what counts as *working well* for an AI-CDSS? and What does *adoption* or *correct use* mean?

These questions should not be confined to social scientists or STS researchers but should also be taken up by stakeholders directly involved in the development and implementation of AI-CDSS as a prerequisite for responsible, equitable, and reflective deployment. Accordingly, Wagman and Parks [16] call for closer engagement between STS and HCI, which they argue can offer "ways of thinking about the complexities of the social, user identities, and power dynamics that can enrich the process of conceptualizing, designing, and developing social machines" (p. 4).

A substantial body of scholarship in STS and the sociology of health has examined digital health technologies by focusing on how they reconfigure relationships, responsibilities, and practices within health care settings [17,20-22]. We aim to extend these approaches to AI systems, whose opacity, scalability, and perceived autonomy introduce distinct challenges that warrant renewed examination of clinical responsibility and trust [23,24].

In our previous study, clinicians were interviewed about their informational needs when using AI-CDSS [25]. An underlying assumption in that study was that providing clinicians with more and *better* information would raise awareness and ensure safe and appropriate use of AI-CDSS. This reflects what STS and PUS scholars refer to as the *deficit model*, which assumes that hesitancy or mistrust toward an emerging technology is the result of a lack (ie, a deficit) of knowledge [26]. However, this perspective oversimplifies the dynamics at play. Trust in AI is not built solely by closing a knowledge gap but is also shaped by factors such as perceived relevance, institutional support, and the alignment of technology with existing practices and values [27-29]. Recognizing the limitations of the deficit model underscores the need for approaches that examine how information is constructed, translated, challenged, and negotiated in specific contexts.

Through an STS lens, this study reanalyzed interview material to examine how clinicians and AI experts construct requirements and concerns regarding AI-CDSS in clinical practice. Rather than aiming to deliver concrete implementation strategies, design guidelines, or policy recommendations, our objective was to unpack the sociotechnical conditions that shape perceptions of knowledge, value, and responsibility. In this process, STS concepts provided conceptual resources for interpreting patterns in the data without dictating our empirical findings.

XSL•FO
RenderX

Methods

Study Design

This qualitative study reanalyzes interview data to demonstrate how empirical material collected under a different analytical orientation can reveal additional sociotechnical dimensions. The aim of the original research was to explore the informational needs of clinicians regarding the use of AI-driven recommendations in medical decision-making [25]. The data were originally gathered from two perspectives: (1) user-centered preferences of clinicians and (2) technical and safety requirements identified by AI experts. AI experts were determined as individuals with direct experience in developing and/or implementing AI systems for use in health care. The selection of the 2 groups in the original research was based on the assumption that the informational needs may go beyond the expressed preferences of clinicians and require expert-identified information for the informed and safe use of AI in clinical settings. This paper was written in accordance with the Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist (Checklist 1) [30].

Research Group

SM is a PhD student in Clinical Data Science (CDS) at Maastricht University. Her research focuses on how to communicate information about AI to health care professionals and patients, with the aim of supporting understanding of AI-based tools. HP is a PhD student at CDS and is affiliated with the Department of Health, Ethics and Society (HES). Her research explores the trustworthiness of AI in health care, focusing on how negotiations around and disruptions of trust shape the implementation of AI. RF is an assistant professor at CDS. Her research explores the application of AI in the health care domain, with an emphasis on the various aspects that influence the clinical adoption of AI, such as trust, informational needs, effectiveness, and more. RJ is an assistant professor at HES. Her research explores how new digital technologies and medical diagnostics contribute to care practices and the changing roles of, and relationships between, patients, health care providers, and health services more broadly. BP is an associate professor of Biomedicine and Society at HES. His research examines how credibility, integrity, and collaboration shape trustworthy knowledge, and how scientific norms evolve under institutional and technological change.

Reflexivity and Research Process

This reanalysis was initiated after SM and HP collaborated on the original study [25]. Throughout the initial analysis phase, the research team (SM and HP) observed significant thematic depth in the data. While not relevant to the original research question, it warranted a dedicated investigation from a distinct theoretical perspective.

To explore dimensions that were intentionally excluded from the original research scope, we investigate how clinicians' understanding of AI-CDSS extends beyond the information requirements constructed by clinicians and AI experts. We initially focused on themes of trust in AI, which is the main area of expertise of one researcher (HP). However, during the reanalysis process, we extended the analytical focus beyond trust by drawing on STS concepts to take broader sociotechnical perspectives into account—which affect, but also go beyond, trust in technology.

While one set of researchers (HP, RJ, and BP) had previous experience applying STS concepts to empirical material, the remainder (SM and RF) were unfamiliar with STS. Furthermore, over half of our team works in a department that has a generally optimistic and positivist stance toward the dissemination and implementation of AI, which might have shaped the research process and the reanalysis of the data.

Study Participants

The original study used a combination of convenience and snowball sampling to recruit participants between July and November 2024. Potential participants were identified through the research team's professional networks and targeted LinkedIn (Microsoft Corp) searches, with invitations and study information folders distributed via email or personal messages. Due to interview participants' time constraints, 8 individuals per group were set as the target for recruitment rather than basing recruitment on theoretical saturation. The original study established inclusion criteria requiring clinicians to be active practitioners in the Netherlands, while excluding any potential participants with less than 2 years of relevant work experience. Clinicians were not required to have any prior knowledge of AI systems. AI experts were included based on direct professional involvement in the development, evaluation, or implementation of AI systems in university and/or health care settings. All participants were based in tertiary care settings, either academic medical centers or specialized institutes, where clinicians practiced and AI experts developed, implemented, or evaluated AI systems. An extensive overview of the participants is provided in Table 1.

XSL•FO
RenderX

Table . Recruitment and participation information, demographic and professional characteristics, and interview settings of the 16 participants interviewed for the original study between August and November 2024 in the Netherlands.

Characteristics	Clinicians (n=8)	AI experts (n=8)
Recruitment and participation		
Total participating/total invited, n/N (%)	8/26 (31)	8/11 (73)
Declined—lack of time, n	8	0
Declined—no reason provided, n	10	3
Demographics		
Gender, n		
Men	6	5
Women	2	3
Median age (y), (IQR)	44 (29-59)	37 (28-48)
Professional roles		
Median work experience (y), (IQR)	14 (2-22)	11 (3-21)
Specialties/roles, n		
Internal medicine	2	—[a]
Oncology	2	—
Gastroenterology	2	—
Cardiology	1	—
Radiology	1	—
Medical imaging AI	—	3
Innovation and implementation management	—	3
Data infrastructure and FAIR[b] principles	—	2
Previous professional AI experience, n		N/A[c]
Prediction models	5	
Imaging	2	
Generative AI	1	
Interview setting		
Interview language, n		
Dutch	8	6
English	0	2
Format, n		
Online	4	6
Face-to-face	4	2

[a]Not applicable.

[b]FAIR: findability, accessibility, interoperability, and reusability.

[c]N/A: not available (subject matter experts).

Interview Data

SM conducted semistructured interviews with 8 clinicians and 8 AI experts between August and November 2024. Interviews were held either online through Microsoft Teams or face-to-face at the hospital or university and lasted between 30 and 60 minutes. While most sessions were conducted in Dutch, 2 interviews with AI experts were held in English. The interviews were transcribed verbatim, and all Dutch interviews were translated into English. No field notes were made during the interviews.

The interview guide contained questions eliciting participants' reflections on their experiences with AI applications in health care settings and their involvement in the development of AI systems for clinical use (Textbox 1). Clinicians were asked what information they would ideally want to receive, while AI experts were asked what information they believe should be

XSL•FO
RenderX

communicated to end users. In this context, clinicians are regarded as the end users of the AI-CDSS. In the final segment of the interview, all participants were presented with 3 examples of documentation tools—Model Card [31], Model Facts [32], and the TRIPOD-AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis—Artificial Intelligence) checklist [33,34]—which served as prompts for discussion about clarity, usefulness, and relevance in supporting transparency in AI-CDSS. These materials functioned as discussion prompts to explore participants' perceptions of possible transparency formats but were not themselves the focus of the analysis. The interview guide and examples (Multimedia Appendices 1 and 2) were based on a Bayesian Network from a public repository to reflect moderate model complexity, with full construction details described in Mingels et al [25]. No repeat interviews were conducted.

Textbox 1. Outline of the interview guide topics for clinicians and AI experts used in the original research [25].

Clinicians

- Demographic information (age, gender, job description, and work experience)
- Experience with AI in care settings
- Currently received information about AI models
- Information they would prefer to receive
- Examples of the Model Card, TRIPOD-AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis—Artificial Intelligence) checklist, and Model Facts were shown, and participants were requested to provide feedback on the amount, structure, and comprehensiveness of the information

AI experts

- Demographic information (age, gender, job description, and work experience)
- Experience developing AI for care settings
- Provided information concerning AI models
- Information they think they should provide
- Examples of the Model Card, TRIPOD-AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis—Artificial Intelligence) checklist, and Model Facts were shown, and participants were requested to provide feedback on the amount, structure, and comprehensiveness of the information

While we recognize that the original interview guide and interviews were designed from a post-positivist perspective, focusing on practical requirements and outcomes, the richness of the qualitative data nevertheless provided insights into how participants engaged in making sense of AI-CDSS in clinical practice. We acknowledge that the original research design constrained the extent of theoretical engagement and that the interview guide did not explicitly explore the dimensions we discuss in this publication. However, the depth and variability of responses allowed us to identify themes relevant to social, institutional, and moral dimensions of AI adoption, which go beyond the original focus on informational needs. This illustrates that even studies oriented toward practical outcomes inevitably capture the complexities of health care AI, highlighting the value of reinterpreting data through a theoretical lens different from that originally intended.

Data Analysis

This interpretive study combines deductive and inductive coding strategies. Initially, deductive codes based on the European Commission's guidelines for trustworthy AI [35] were collaboratively applied by SM and HP to all interview transcripts using ATLAS.ti Web (version 25.0.1; ATLAS.ti Scientific Software Development GmbH). This provided insights into how participants linked specific informational needs to the European Commission's conceptualization of trustworthiness, as notions of trust appeared both organically and implicitly throughout the interview data.

Following this, SM and HP collaboratively reviewed and restructured the coded material, carrying out an inductive thematic analysis [36]. Several thematic areas were created from this reorganization of the data. The themes were information overload vs information deficit, trust through experimentation, understanding the output, institutional trust and responsibility, and clinical relevance.

During the reorganization, one researcher (HP) identified close affinities with feminist and constructivist STS concepts. Constructivist concepts assume that scientific knowledge and technological design are socially and materially constructed, and are therefore a product of their surrounding social and material processes. Feminist STS critically examines power relations that shape knowledge production and the development of technology, questioning whose perspectives are included and what assumptions guide them.

Following extensive discussions with the supervisory team familiar with STS (BP and RJ), SM and HP linked the themes to STS concepts to provide analytic structure and to make them accessible to audiences less familiar with feminist and constructivist STS. Specifically, situated knowledge was used to examine how knowledge claims depend on social and professional position; standpoint theory to examine differences

XSL•FO
RenderX

in perspectives and epistemic priorities across stakeholder groups; tinkering to examine experimentation and practical knowledge-making in use; situated value to examine how the worth of AI-CDSS is constituted in relation to practices and priorities; and response-ability to examine how responsibility is distributed and enacted within institutional arrangements. These concepts were introduced after inductive thematic reorganization and were used abductively to interpret and reconsider empirical patterns. Their relation to the interview material is described in the "Results" section.

In summary, our analytical process followed a 3-step sequence: initial deductive coding mapped data to the European Commission's trustworthy AI framework; inductive thematic reorganization then clustered these codes into core empirical themes; and finally, abductive theorization, following principles of abductive reasoning, integrated prior theoretical knowledge of STS concepts to make sense of the practical complexities and tensions encountered in the empirical material. The 2 frameworks used in the analysis served different epistemic and analytical functions. The European Commission's framework for trustworthy AI was used deductively as an initial normative coding framework, providing predefined dimensions through which participants' accounts could be systematically examined. This approach does not share the epistemic commitments of feminist and constructivist STS, which instead informed the subsequent interpretive phase. Abductive thematic analysis is distinct from both inductive and deductive approaches. It specifically addresses scenarios where empirical observations are brought into iterative dialogue with theoretical literature to interpret complex phenomena [37,38]. In this study, we viewed these inductively created themes through STS sensitizing concepts to unpack the underlying sociotechnical dynamics. The analysis consequently treats the 2 frameworks as sequential rather than epistemologically integrated: the first structured the initial empirical analysis, while the second enabled a critical reinterpretation of what those empirical patterns meant in practice.

Ethical Considerations

Ethical approval for the original study was granted by the Faculty of Health, Medicine, and Life Sciences—Research Ethics Committee (FHML/HDT/2024.022). Before the interviews, all individuals were provided with detailed information regarding the study's objectives and data management protocols. Written informed consent was obtained for both participation and audio recording. Additionally, all individuals specifically consented to the reuse of their data for related research. To protect participant privacy, all transcripts were pseudonymized and stripped of identifiers. No member checking was conducted, to avoid influencing participants' perspectives in potential follow-up research.

Results

Overview

We selected the concepts of situated knowledge, standpoint theory, tinkering, situated value, and response-ability because they provide guidance for analyzing how knowledge, authority, and value are coproduced in sociotechnical systems, and illuminate the often-hidden social, institutional, and moral dimensions of AI adoption. This allowed us to interpret participants' experiences not merely as practical challenges but as reflections of broader sociotechnical dynamics, including power relations, professional roles, and organizational constraints. Our analysis does not aim to list themes but to explore how clinicians and AI experts make sense of AI systems *in practice*. In each section, we first explain the STS concept and then follow with representative examples from our data. The aim of the quotations is not to provide statistical representation but to demonstrate how interpretations are grounded in participants' accounts. The STS concepts did not, therefore, determine what was found; rather, they provided conceptual resources for interpreting patterns from the empirical material.

Situated Knowledge

A central concern across the interviews was how much and what kinds of information clinicians need about AI-CDSS used in practice. This was expected, given that the interviews focused specifically on clinicians' informational needs. However, the desired level of information varied across clinicians, depending on individual background and the specific clinical and technical context in which the AI system was being used. While some clinicians described a lack of access to useful information about the workings of the AI model or how it was developed and validated, others experienced the provided information as too technical and not grounded in clinical practice. Similarly, AI experts reported struggles in deciding what information to communicate to clinicians.

> *I think clinicians should know all the shortcomings that are there, but you also do not want to overwhelm them with the information that they cannot use.* [Expert 3]

Our data reveal that the information clinicians want to receive to *understand* AI systems is shaped by their professional roles, prior experience with AI, and the specific clinical contexts in which AI-CDSS are used.

> *And I think that for us simple doctors, to put it that way, we are not going to be able to do without AI specialists if we want to work with this [AI model]. You will not be able to understand everything. You won't have to understand everything either.* [Clinician 2]

The notion of *situated knowledge* was initially introduced by Haraway [39], challenging the belief that there is one single universal truth. Instead, they put forward that all knowledge is partial (incomplete), situated (produced from a specific positioning in the world), and multiple (different perspectives can coexist). This means that knowledge does not exist in a value-free space but is shaped by the social and cultural structures it is embedded in. By *situating* knowledge, one aims to show from which perspectives and viewpoints knowledge comes. This creates possibilities for accountability, as understanding the origin and cultural context of a knowledge claim enables it to be held to account.

XSL•FO
RenderX

What counts as relevant information for one clinician is excessive or irrelevant to another. These differences show that there is no *universally fixed* informational package that enables clinicians to make sense of an AI system. This introduces tensions in epistemic coordination, where information is defined in different ways and only becomes meaningful in relation to the practices, responsibilities, and decisions of those engaging with it.

Standpoint Theory

Within the interviews, clinicians emphasized that understanding how to interpret and act on the AI-CDSS' output is even more important than grasping the inner workings of a model. The *usefulness* of an AI system's recommendation depends on the clinician's ability to integrate it meaningfully into their decision-making process. Clinicians stressed that they do not want to be handed an output, but rather to understand what to do with it. This meant needing a sense of the AI system's limitations, its range of applicability, and how much weight to give it relative to their own judgment to be able to act on the output.

> *That you have a good model that is also clinically relevant. And with all due respect, but an information specialist should not be able to judge that. He is not trained for that. And a clinician is not trained to develop that model, I think you have to do it together in the development phase.* [Clinician 2]

Feminist standpoint theories, as originally coined by Harding [40] and Collins [41], build on the idea of situated knowledge by emphasizing the importance of making explicit where knowledge claims come from. They argue that knowledge originates from certain cultural and social positionings—or standpoints—and is therefore riddled with values and influenced by power relations. To avoid reinforcing epistemic inequalities (which are differences in whose knowledge and experiences are recognized, valued, and trusted), it is important to clarify the standpoint of knowledge claims. Standpoint theory emphasizes centering perspectives from less-heard or underrepresented groups, explicitly acknowledging context, and actively pursuing balance between standpoints.

Clinicians emphasized understanding how to interpret outputs and integrate them into decision-making, while AI experts focused on model performance metrics and hyperparameters. Information that is useful to AI experts is not always the same as information that is useful to clinicians.

> *The more you go into that statistic, the more difficult it becomes of course, see, the doctors know what an AUC is and the higher the better. They also know that it is never 1. That 0.5 is random. But then it stops. That is that then. I think that majority even think that that applies to all the patients they then throw into that model that it is not even at subgroup level that some predict better, others don't.* [Expert 2]

Tinkering

In the interviews, clinicians reported that initial engagement with AI-CDSS often happens through experimentation. In cases where structured onboarding or formal training was lacking, clinicians developed their own strategies to make sense of the AI system. This included entering different inputs into the system to observe how the output would change and was carried out both before using the AI in clinical practice at all—with fictional information entries—and with a real-world patient. For example, if an 80-year-old patient seems more vital than the average 80-year-old, one clinician stated that they would type an age of 70 years into the AI-CDSS to find out how a more vital patient would respond to the intended treatment. This was not a formal validation procedure, but a way for clinicians to learn how the system's output changes across different situations and how these relate to their own clinical assessment.

> *Age is just a number, isn't it? (...) There are people in their 60s who function at the level of a 70-plus or even an 80-something (...) You are actually more vital than your age suggests, so that's all interpretation. And I think you should be able to leave that to a clinician.* [Clinician 8]

AI experts also reported that clinicians wish to test how the AI system behaves with data collected within their own institution (local validation). This distinct mode of engagement allows clinicians to explore how the system responds to their specific data, workflows, and clinical routines.

> *I think doctors are especially good at external validation and generalizability, but also yes. I actually think that they would also like to test it internally themselves, want to pilot.* [Expert 5]

Knorr [42] introduces the practice of *tinkering* in the context of a natural science laboratory, where the researchers' "projective openness and the material character of operation and outcome" (p. 368) are part of the research process. Similarly, Mol et al [43] define tinkering as "attentive experimentation" (p. 13), a practice of trying out and adapting. Through this practice, experimentation is connected to the processes of valuation. Tinkering can involve adapting, adjusting, and improvising, which results in a form of reasoning in which appropriateness can be constructed through practice. This means that processes of valuation are not predetermined or defined in advance, opening up possibilities for alternative forms and standpoints in knowledge creation.

These findings highlight different forms of experimentation, or tinkering, through which clinicians test the AI system's behavior and draw conclusions about when and how to (mis)use it. Tinkering refers not simply to experimentation or trial-and-error, but to the practices and negotiations through which clinicians develop an understanding of when and how AI outputs become meaningful.

Situated Value

Participants emphasized that for AI-CDSS to be meaningfully adopted in clinical practice, they must contribute something beyond what clinicians already know. Some participants argued that if the AI-CDSS merely replicates clinical reasoning using the same variables, without introducing new perspectives or insights, its perceived value is limited. In these cases, AI systems were seen as adding complexity to an already complicated

XSL•FO
RenderX

system. On the other hand, participants argued that AI should not aim to replace human expertise but rather augment it.

> *There are a lot of tools, but actually no real applied decision models that we can use in the clinic. (...) Certainly not as a substitute for a choice. So yes, you do have certain models that can give you a bit of direction, but not really to make a choice.* [Clinician 5]

From the clinicians' standpoint, an AI-CDSS must generate actionable insights to be considered useful within clinical practice. Furthermore, our data indicate that AI experts tend to prioritize performance metrics and are not necessarily aware of different forms of valuation, which is further reinforced by publishing and reporting standards.

> *Some doctors say, if I have to look something up in another system, I'm not going to use it. Others say, well, if it does help me, then I will and it will cost me a little more. So, there is also different parties between different parties. So, you can never make something that is good for everyone. But you do want, yes, not that you lose everyone on some points. (...) In the end you have 3 men who all do not use it for other reasons, so you want to have them all mapped out. And some can be solved by doing technical things. And others you have to tackle by education or information. Whether convincing and motivating people is often another thing.* [Expert 2]

Heuts and Mol [44] draw from the work of Boltanski and Thévenot [45] on economies of worth and introduce the notion of *situated value*. This idea highlights that the worth of something is not fixed, but dependent on the context in which it is encountered. In their study of what counts as a good tomato, Heuts and Mol [44] interviewed *tomato experts* in the Netherlands and found that the same tomato could be valued in very different ways depending on which qualities played a role in its use. These differences in valuing are not simply in conflict with each other but can exist alongside one another. This reveals that multiple notions of what counts as *good* or *desirable* can be true at the same time. Similar to situated knowledge [39], the concept of situated value suggests that, instead of one singular, universal measure, values can be plural, relational, and shaped by the practices and intentions of those involved.

Interviewed clinicians judged an AI system's usefulness based on whether it added something meaningful to their own clinical practice. For some, value meant gaining new insights; for others, it meant reducing uncertainty in decision-making. These different valuations contrast with those of AI experts, among whom higher performance metrics were the priority in constructing a *good* system. Therefore, value is not an inherent property of the AI-CDSS itself but is shaped by context, use cases, and the professional priorities of those encountering it.

Response-Ability

Clinicians often raised the concern that the AI-CDSS output might be incorrect. Health care professionals are held accountable for the consequences of clinical decisions, including those that involve or are influenced by AI. Most interviewed clinicians argued that they lack technical expertise or resources to independently assess the AI system, which meant they had to place trust in others, often in the developers of the AI system. In addition, trust is also delegated to institutional structures, such as expert review boards or internal regulatory bodies, responsible for evaluating and approving AI systems for use in clinical environments. The lack of clear legal rules, standard training, or strong organizational guidance limits clinicians' capacity to act on AI outputs accordingly.

> *(...) But the Medical Device Regulation looks very much at, yes, say privacy things and reliability of the tool, but there is actually no clinical validation step to it. There are no guidelines for that and it is often very rattling. So, I'm actually looking for clearer guidelines about yes, just like you have to do phase 1 2 3 studies, about medicines of what do you have to do now before you can really use something reliable in practice. (...)* [Clinician 4]

However, even AI experts have stated that they sometimes struggle to fully understand the functioning of an AI system, especially when contextual knowledge is lacking or when the AI system itself is difficult to interpret.

> *In the end, you have a responsibility as a doctor to do the right thing, but within the framework of using it well, so to speak. When you use a model, and it doesn't work, you can eventually say, yes, but I still used the model, so to speak, you know. If you do not get it, you cannot really make an informed decision if you do not understand the information.* [Expert 5]

Response-ability can be understood as the "capacity to respond" [46]. Originally introduced by Haraway [47], the concept describes an approach to finding ways of responding to situations, which is particularly relevant in contexts where the consequences or implications of certain actions are not immediately obvious. Response-ability invites reflection and thinking beyond causal reactions to explore alternative forms of response. These alternative forms of engagement should acknowledge the complex and interconnected nature of engagement. Therefore, things gain their meaning through interactions with other human and nonhuman entities. From this perspective, response-ability is not about accounting for every possible consequence. Instead, it calls for paying attention to possibilities, harms, and outcomes that might have more significant implications than expected.

These concerns illustrate how response-ability plays out in practice. Responding here is not only about reacting to an error when it occurs but also about *being able* to respond when AI outputs are uncertain, contested, or difficult to interpret. These concerns also depend on whether institutions provide the conditions and support needed for clinicians to act on uncertainty.

Discussion

STS Concepts as Sensitizing Lenses

The findings of this study highlight that the adoption and meaningful use of AI-CDSS in practice are not determined only

XSL•FO
RenderX

by technical performance but also by the interplay of social, organizational, and moral relations. Our results do not challenge the well-established observation that technology adoption is shaped by social and organizational context, as demonstrated in existing frameworks of HCI, implementation science, and technology adoption research. This study extends these traditions by examining how clinicians' and AI experts' engagement with AI systems is shaped by the situated character of knowledge, the standpoint from which they operate, the ways they experiment and adapt tools in practice, the values they assign to AI outputs, and the organizational structures that they believe enable or constrain their capacity to respond. These findings resonate with broader insights from STS and health care technology research and extend them by highlighting how AI systems intersect with professional roles, institutional norms, and experiential expertise. Our contribution is not to claim that the general observation that technologies are shaped by their contexts is novel. Rather, this study applies feminist and constructivist STS concepts as sensitizing lenses to a specific contemporary case: the implementation of AI-CDSS, where questions of algorithmic authority, responsibility, and trust introduce particular challenges beyond conventional technology adoption.

Context Matters

Participants described different informational needs in relation to their professional role, experience with AI, and clinical context. We interpret this variation through the concept of situated knowledge [39]. In practice, this could mean that information about the AI-CDSS needs to be specifically tailored to relevant stakeholders. Different users may benefit from different formats or levels of explanation depending on their role, expertise, and the decisions they need to make. Moreover, accountability emerges not simply as a property of the system, but as an effect of how information is articulated. This includes, for example, information about how data for the model were collected, as well as reflections on the structure and presentation of information to clinicians.

The situated character of how clinicians make sense of AI aligns with ethnographic and STS-informed studies on CDSS and medical technologies, which demonstrate that technologies are implemented differently across contexts and that adoption depends on the interplay between technology, organizational routines, and social interactions [48-53]. The variability we observed confirms that successful AI implementation requires attention to local practices, material infrastructure, and workflow integration.

Multistakeholder Perspectives

Clinicians and AI experts emphasized different aspects of AI-CDSS when describing what they needed to understand to evaluate the model. We interpret these differences through standpoint theory [40,41]. Standpoint theory provides a framework for interpreting these differences as situated forms of knowing, rather than simply as differences in information preferences. AI experts have a different *standpoint* due to their professional roles and likely require different information from clinicians concerning AI-CDSS, such as performance metrics and hyperparameters. From the clinicians' standpoint, a recommendation is only useful if it can be applied in practice, which highlights that integration extends to material processes, such as the availability of treatments, resources, and workflows in the clinical setting.

This echoes Mørk et al [54], who show that power relations within and across communities of practice shape whose knowledge is recognized and valued. Similarly, Breuer et al [55] demonstrate that engineers' imaginaries influence health care technology design and engagement. These insights suggest that centering multiple standpoints is crucial to creating AI systems that are both usable and socially legitimate. Making these standpoints explicit and reflecting on them can help ensure that AI-CDSS are designed with the needs of their end users in mind to avoid overlooking expertise that comes from working directly with patients. Without this, they risk missing the realities of clinical care, rendering them not applicable in daily practice.

Space for Experimentation

Participants' descriptions of testing, interpreting, adapting, and locally validating AI-CDSS point to forms of work that extend beyond formal training and evaluation activities. Interpreted through the lens of tinkering, these activities make visible the often-unacknowledged work involved in making AI systems workable in practice [43]. Rather than relying on technical explanations, clinicians developed what might be described as lay expertise [56,57] based on their direct experience with the system's output. This perspective draws attention to how practical knowledge about AI-CDSS can develop through experimentation and adjustment in use. This kind of knowledge is situated in clinical practice and grounded in relevance to specific cases, rather than in the inner workings or technical background of the AI system. The understanding that develops through experimentation may shape whether and how clinicians choose to rely on the AI system in practice.

The practice of tinkering resonates with prior research emphasizing the importance of incremental adaptation and situated experimentation in technology adoption [43,48,58]. Tinkering shows that knowledge about AI is coproduced through use, rather than passively acquired from technical documentation. This emphasizes the importance of looking beyond formal training or documentation and of highlighting the ways in which knowledge about AI is produced through use by nonexpert users, as well as of providing technical infrastructures for experimental learning and collaboration [59]. By applying this concept, we highlight the invisible labor involved in making AI systems workable in clinical practice, including the often-unacknowledged effort clinicians invest in testing, interpreting, adapting, and integrating AI systems beyond formal training and evaluation settings [60,61].

Diverging Valuations

Participants differed in what they considered valuable about AI-CDSS, which suggests that the value attributed to AI-CDSS was related to participants' professional roles and the purposes for which they anticipated using or evaluating the technology. We interpret this through the notion of situated value [44], showing that the perceived usefulness of AI-CDSS varies among stakeholders. What counts as *good* or *useful* AI depends on the

XSL•FO
RenderX

perspective from which it is evaluated. This does not mean that one notion of value is superior to another, but rather that multiple, sometimes divergent, notions of value can be true at the same time. The differences in participants' accounts suggest that AI-CDSS development needs to accommodate multiple professional perspectives on what constitutes useful and meaningful AI. Involving clinicians throughout the entire life cycle of the AI-CDSS, from conception until deployment and even afterward, and accepting that different perspectives and values shape the development of AI systems in divergent ways would be one way of doing that.

This aligns with STS analyses that argue that technology's value is relational and context-dependent [16,62]. Recognizing these divergent valuations highlights the need for continuous involvement of multiple stakeholders throughout the AI life cycle, from design to deployment, ensuring that tools align with real-world clinical priorities rather than abstract technical ideals.

Importance of Structural and Institutional Factors

Participants described uncertainty about responsibility in situations where AI outputs could be incorrect, while also identifying gaps in training, organizational guidance, and institutional support. We interpret these accounts through response-ability, which draws attention to the conditions that shape actors' capacity to respond to AI-supported decisions [47]. Response-ability helps us interpret these accounts as pointing to the distributed conditions under which responsibility can be enacted, and it captures clinicians' reliance on institutional structures and expert oversight when interacting with AI-CDSS. This entails anticipating possible problems and knowing how to act when perceived uncertainty is high. Clinicians rely on others to assess AI-CDSS, but these groups might lack necessary context or technical understanding. This is further complicated by a lack of clear legal frameworks, standardized training, and organizational guidance, with clinicians forced to navigate complex forms of delegated trust without adequate resources or support throughout the AI life cycle. Participants' accounts suggest that explainability, validation, or certification alone may not address all of the concerns they associated with responsibility and appropriate use, particularly where training, organizational guidance, or institutional support were perceived as limited.

Uncertainties about the ability to respond to an AI system from the clinicians' perspectives parallel feminist STS critiques of algorithmic authority, which show that technical expertise alone cannot account for the social, ethical, and political conditions that enable safe AI use [16,63]. From this perspective, safe and fair use of AI depends not only on how accurate the model is and what model information is communicated to clinicians but also on structural and institutional arrangements, which are necessary for clinicians' use (or rejection) of AI.

Limitations

The results of this study should be interpreted in light of several limitations, some of which are inherent to the original research design. First, the use of convenience and snowball sampling in the original study likely introduced selection bias, as individuals who volunteer to participate in AI research are often more technologically engaged or positive about the topic than the general clinical population. Consequently, our findings on *situated value* and *tinkering* may primarily reflect the sense-making of lead users or enthusiasts, rather than capturing clinicians more distanced from digital transformation. The findings should not be interpreted as representative of all clinicians' perspectives or as capturing the experiences of those with limited engagement with AI-CDSS. As this is an interpretive qualitative study, our aim is not statistical generalization but to examine how meanings, values, and responsibilities surrounding AI-CDSS are constructed within specific professional and institutional contexts.

Second, the fixed sequence in providing participants with the Model Cards, TRIPOD-AI, and Model Facts may have introduced order effects, such as primacy bias. The heterogeneity of the specialties among the participant group also means that while we could identify broad themes of *situated knowledge*, we lacked the disciplinary depth to map out specific medical specialties in detail. Future research should examine in more detail how clinical workflows, organizational contexts, and specialty-specific forms of expertise shape the interpretation and integration of AI systems.

Finally, the original study was grounded in a post-positivist research stance, which constrained the depth of theoretical engagement. In particular, the research question foregrounded informational needs, a framing that aligns with instrumental and problem-solving approaches but does not constitute a central object of inquiry within STS. Moreover, in the initial interview design, participants were presented with multiple examples of reporting standards and asked to evaluate them. This approach diverges from STS methodologies, which typically privilege processual, narrative, and practice-oriented accounts over evaluative or opinion-based questioning. Finally, the interview guide itself mirrored these limitations by not exploring dimensions such as power relations, epistemic assumptions, and institutional constraints more explicitly.

Conclusions

This study suggests that the adoption and meaningful use of AI-CDSS cannot be understood solely in terms of technical performance or informational provision. Instead, it is shaped by the social, political, and moral characteristics of the world they are supposed to become part of. These characteristics, and the recognition of diverse clinical and organizational contexts, are central to understanding technology adoption. As health care environments continuously evolve, our study provides a snapshot of how clinicians and AI experts in the Dutch context interpret and engage with AI-CDSS, highlighting the context-dependent and negotiated nature of AI uptake. To this end, feminist and constructivist STS concepts—situated knowledge, standpoint theory, tinkering, situated value, and response-ability—can provide a lens to examine more closely these *hidden* factors that are often ignored or only briefly discussed.

This methodological shift changes what is analytically foregrounded in the same interview material. For example, the original study analyzed clinicians' and AI experts' statements primarily in terms of the information they considered necessary.

XSL•FO
RenderX

This analysis interprets these same accounts as expressing different forms of situated knowledge and professional standpoints, and therefore different criteria for what makes information meaningful. Similarly, participants' accounts of experimentation can be read not only as a means of developing trust but also as practices through which clinicians develop situated knowledge about when and how AI outputs become usable, including the often-unacknowledged labor involved in testing and integrating the system. The divergence between clinicians' emphasis on actionable clinical augmentation and AI experts' emphasis on performance metrics, validation, and reporting requirements further illustrates that the value of an AI-CDSS is not assessed according to a single criterion but is negotiated across different professional practices and priorities. Finally, participants' accounts of responsibility reveal that the capacity to respond to uncertain or contested AI outputs depends not only on the properties of the system or the information provided but also on institutional and organizational conditions.

In practical terms, these findings suggest that developing and implementing AI-CDSS requires continuous engagement with end users, attention to local workflows and decision-making practices, and the creation of institutional structures that enable accountability. Understanding AI adoption involves not only identifying social, technical, and organizational factors that influence implementation but also examining how decisions about appropriate performance and use are made in practice. This raises questions such as who determines what counts as an AI-CDSS working well and what constitutes appropriate use in a particular clinical setting. Developing detailed implementation frameworks or policy recommendations would require additional research specifically focused on intervention design, organizational implementation processes, or regulatory evaluation, which was beyond the scope of this qualitative interpretive study. However, incorporating STS perspectives can guide developers, policymakers, and clinicians to design AI systems that are socially legitimate, ethically grounded, and that can be integrated into clinical practice in a meaningful manner. This approach underscores that responsible AI adoption in health care depends as much on understanding the social and institutional conditions of use as on technical performance or communication strategies.

Acknowledgments

The authors thank the clinicians and AI experts who participated in the interviews. Their contributions provided essential insights into the use of AI in clinical contexts.

Generative AI Use

SM and HP declare the use of generative AI (GenAI) in the research and writing process. According to the GAIDeT taxonomy [64], the following tasks were delegated to GenAI tools under full human supervision: (1) proofreading and editing and (2) reformatting. The GenAI tools used were ChatGPT (GPT-4) and Gemini 3. Responsibility for the final manuscript lies entirely with the authors. GenAI tools are not listed as authors and do not bear responsibility for the final outcomes. The remaining authors (RJ, RF, BP) did not use generative AI.

Funding

This research was funded by the Dutch Research Council (NWO) under the grant number NGF.1607.22.040.

Data Availability

The datasets gathered and analyzed during the study are not publicly available due to the presence of potentially identifiable information within the interview transcripts.

Authors' Contributions

Conceptualization: SM, HP
Data curation: SM, HP
Formal analysis: SM, HP
Investigation: SM
Methodology: SM, HP, RF, RJ, BP
Project administration: SM, HP
Resources: SM, HP, RF, BP
Supervision: RF, RJ, BP
Writing – original draft: SM, HP
Writing – review & editing: SM, HP, RF, RJ, BP

Conflicts of Interest

SM, HP, and RF are affiliated with a department where AI-based systems are developed. RF receives research grants related to the development and implementation of AI. In addition, SM, HP, and RF knew some of the respondents interviewed for the primary research before the study. Outside the submitted work, RF's research group receives funding from public-private research

XSL•FO
RenderX

consortia, including contributions from Janssen-Cilag and Takeda. These companies had no role in this study. RJ and BP declare no conflicts of interest.

Multimedia Appendix 1
Interview guide clinicians and interview guide AI experts.
[DOCX File, 23 KB - humanfactors_v13i1e99117_app1.docx]

Multimedia Appendix 2
Examples of Model Card, TRIPOD-AI, and Model facts.
[DOCX File, 77 KB - humanfactors_v13i1e99117_app2.docx]

Checklist 1
COREQ checklist.
[PDF File, 433 KB - humanfactors_v13i1e99117_app3.pdf]

References

1. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial intelligence and the implementation challenge. J Med Internet Res 2019 Jul 10;21(7):e13659. [doi: 10.2196/13659] [Medline: 31293245]
2. Goirand M, Austin E, Clay-Williams R. Implementing ethics in healthcare AI–based applications: a scoping review. Sci Eng Ethics 2021 Sep 3;27(5):61. [doi: 10.1007/s11948-021-00336-3] [Medline: 34480239]
3. Ankolekar A, van der Heijden B, Dekker A, et al. Clinician perspectives on clinical decision support systems in lung cancer: implications for shared decision-making. Health Expect 2022 Aug;25(4):1342-1351. [doi: 10.1111/hex.13457] [Medline: 35535474]
4. Gama F, Tyskbo D, Nygren J, Barlow J, Reed J, Svedberg P. Implementation frameworks for artificial intelligence translation into health care practice: scoping review. J Med Internet Res 2022 Jan 27;24(1):e32215. [doi: 10.2196/32215] [Medline: 35084349]
5. Nair M, Svedberg P, Larsson I, Nygren JM. A comprehensive overview of barriers and strategies for AI implementation in healthcare: mixed-method design. PLoS One 2024;19(8):e0305949. [doi: 10.1371/journal.pone.0305949] [Medline: 39121051]
6. Smith H, Downer J, Ives J. Clinicians and AI use: where is the professional guidance? J Med Ethics 2024 Jun 21;50(7):437-441. [doi: 10.1136/jme-2022-108831] [Medline: 37607805]
7. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019 Oct 29;17(1):195. [doi: 10.1186/s12916-019-1426-2] [Medline: 31665002]
8. Petersson L, Larsson I, Nygren JM, et al. Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden. BMC Health Serv Res 2022 Jul 1;22(1):850. [doi: 10.1186/s12913-022-08215-8] [Medline: 35778736]
9. Hanna MG, Pantanowitz L, Jackson B, et al. Ethical and bias considerations in artificial intelligence/machine learning. Mod Pathol 2025 Mar;38(3):100686. [doi: 10.1016/j.modpat.2024.100686] [Medline: 39694331]
10. Cherns A. The principles of sociotechnical design. Hum Relat 1976 Aug;29(8):783-792. [doi: 10.1177/001872677602900806]
11. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep 1;13(3):319-340. [doi: 10.2307/249008]
12. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012 Mar 1;36(1):157-178. [doi: 10.2307/41410412]
13. Hussain MI, Reynolds TL, Mousavi FE, Chen Y, Zheng K. Thinking together: modeling clinical decision-support as a sociotechnical system. AMIA Annu Symp Proc 2017;2017:969-978. [Medline: 29854164]
14. Salwei ME, Carayon P. A sociotechnical systems framework for the application of artificial intelligence in health care delivery. J Cogn Eng Decis Mak 2022 Dec;16(4):194-206. [doi: 10.1177/15553434221097357] [Medline: 36704421]
15. Schaffernak I, Cecil J, Kleine AK, Lermer E. Sociotechnical influences on the adoption and use of AI-enabled clinical decision support systems in ophthalmology: a theory-based interview study. BMC Health Serv Res 2025 Oct 22;25(1):1398. [doi: 10.1186/s12913-025-13620-w] [Medline: 41126217]
16. Wagman KB, Parks L. Beyond the command: feminist STS research and critical issues for the design of social machines. Proc ACM Hum Comput Interact 2021;5(CSCW1). [doi: 10.1145/3449175]
17. Henwood F, Marent B. Understanding digital health: productive tensions at the intersection of sociology of health and science and technology studies. Sociol Health Illn 2019 Oct;41(S1):1-15. [doi: 10.1111/1467-9566.12898] [Medline: 31599984]
18. Sismondo S. An Introduction to Science and Technology Studies, 2nd edition: Wiley-Blackwell; 2009.
19. Knopes J. Science, technology, and human health: the value of STS in medical and health humanities pedagogy. J Med Humanit 2019 Dec;40(4):461-471. [doi: 10.1007/s10912-019-09551-3]

XSL•FO
RenderX

20. Hardon A, Moyer E. Medical technologies: flows, frictions and new socialities. Anthropol Med 2014;21(2):107-112. [doi: 10.1080/13648470.2014.924300] [Medline: 25175289]
21. Robinson JH. Bringing the pregnancy test home from the hospital. Soc Stud Sci 2016 Oct;46(5):649-674. [doi: 10.1177/0306312716664599] [Medline: 28948881]
22. Vezyridis P, Timmons S. E-infrastructures and the divergent assetization of public health data: expectations, uncertainties, and asymmetries. Soc Stud Sci 2021 Aug;51(4):606-627. [doi: 10.1177/0306312721989818] [Medline: 33499771]
23. Ananny M, Crawford K. Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. New Media Soc 2018 Mar;20(3):973-989. [doi: 10.1177/1461444816676645]
24. Elish MC. Moral crumple zones: cautionary tales in human-robot interaction. Engag Sci Technol Soc 2019;5:40-60. [doi: 10.17351/ests2019.260]
25. Mingels S, Piehl H, Therrien M, et al. Understanding clinicians' informational needs for AI-driven clinical decision support systems: qualitative interview study. JMIR Med Educ 2026 Mar 12;12:e85228. [doi: 10.2196/85228] [Medline: 41834065]
26. Wynne B. Misunderstood misunderstanding: social identities and public uptake of science. Public Underst Sci 1992 Jul;1(3):281-304. [doi: 10.1088/0963-6625/1/3/004]
27. Piehl HS, Janssen R, Penders B, Fijten RRR. Between map and maze: reframing trust in healthcare AI. AI Soc 2026;41(7):6925-6935. [doi: 10.1007/s00146-026-02997-9] [Medline: 42482956]
28. Sturgis P, Allum N. Science in society: re-evaluating the deficit model of public attitudes. Public Underst Sci 2004 Jan;13(1):55-74. [doi: 10.1177/0963662504042690]
29. Engel N, Krumeich A. Valuing simplicity: developing a good point of care diagnostic. Front Sociol 2020;5:37. [doi: 10.3389/fsoc.2020.00037] [Medline: 33869444]
30. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
31. Mitchell M, Wu S, Zaldivar A, et al. Model cards for model reporting. Proc Conf Fairness Accountability Transpar 2019 Jan 29:220-229. [doi: 10.1145/3287560.3287596]
32. Sendak MP, Gao M, Brajer N, Balu S. Presenting machine learning model information to clinical end users with model facts labels. NPJ Digit Med 2020;3(1):41. [doi: 10.1038/s41746-020-0253-3] [Medline: 32219182]
33. Collins GS, Dhiman P, Andaur Navarro CL, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open 2021 Jul 9;11(7):e048008. [doi: 10.1136/bmjopen-2020-048008] [Medline: 34244270]
34. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024 Apr 18;385:q902. [doi: 10.1136/bmj.q902] [Medline: 38636956]
35. European Commission, High-Level Expert Group on Artificial Intelligence. Ethics guidelines for trustworthy AI. : Publications Office of the European Union; 2019 Apr 8 URL: https://data.europa.eu/doi/10.2759/346720 [accessed 2026-10-14]
36. Braun V, Clarke V. Thematic analysis. In: Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D, Sher KJ, editors. APA Handbook of Research Methods in Psychology, Vol 2: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological: American Psychological Association; 2012:57-71. [doi: 10.1037/13620-004]
37. Khurshid F, Veen M, Thompson J, Hegazi I. Navigating thematic analysis: practical strategies grounded in abductive reasoning. Teach Learn Med 2026;38(1):106-115. [doi: 10.1080/10401334.2025.2475098] [Medline: 40172437]
38. van Hulst M, Visser EL. Abductive analysis in qualitative research. Public Adm Rev 2025 Mar;85(2):567-580. [doi: 10.1111/puar.13856]
39. Haraway D. Situated knowledges: the science question in feminism and the privilege of partial perspective. Feminist Stud 1988;14(3):575-599. [doi: 10.2307/3178066]
40. Harding S. After the neutrality ideal: science, politics, and "strong objectivity". Soc Res 1992;59(3):567-587 [FREE Full text]
41. Collins PH. Comment on Hekman's "Truth and method: feminist standpoint theory revisited": where's the power? Signs (Chic) 1997 Jan;22(2):375-381. [doi: 10.1086/495162]
42. Knorr KD. Tinkering toward success: prelude to a theory of scientific practice. Theory Soc 1979;8(3):347-376. [doi: 10.1007/BF00167894]
43. Mol A, Moser I, Pols J, editors. Care in Practice: On Tinkering in Clinics, Homes and Farms: transcript Verlag; 2010. [doi: 10.14361/transcript.9783839414477]
44. Heuts F, Mol A. What is a good tomato? A case of valuing in practice. Valuat Stud 2013;1(2):125-146. [doi: 10.3384/vs.2001-5992.1312125]
45. Boltanski L, Thévenot L. On Justification: Economies of Worth: Princeton University Press; 2006. [doi: 10.2307/j.ctv1zm2tzm]
46. Draude C. "Boundaries do not sit still" from interaction to agential intra-action in HCI. In: Kurosu M, editor. Human-Computer Interaction Design and User Experience Thematic Area, HCI 2020, Held as Part of the 22nd International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part I: Springer International Publishing; 2020:20-32. [doi: 10.1007/978-3-030-49059-1_2]

XSL•FO
RenderX

47. Haraway DJ. Awash in urine: DES and Premarin® in multispecies response-ability. WSQ 2012 Mar;40(1-2):301-316. [doi: 10.1353/wsq.2012.0005]
48. Pope C, Halford S, Turnbull J, Prichard J, Calestani M, May C. Using computer decision support systems in NHS emergency and urgent care: ethnographic study using normalisation process theory. BMC Health Serv Res 2013 Mar 23;13(1):111. [doi: 10.1186/1472-6963-13-111] [Medline: 23522021]
49. Suchman LA. Plans and Situated Actions: The Problem of Human-Machine Communication: Cambridge University Press; 1987.
50. Wyatt S. Non-users also matter: the construction of users and non-users of the internet. In: Oudshoorn N, Pinch T, editors. How Users Matter: The Co-Construction of Users and Technology: The MIT Press; 2003:67-79. [doi: 10.7551/mitpress/3592.003.0006]
51. Oudshoorn NEJ, Pinch T. Introduction: how users and non-users matter. In: How Users Matter: The Co-Construction of Users and Technology: The MIT Press; 2003:1-25. [doi: 10.7551/mitpress/3592.003.0002]
52. Berg M. Rationalizing Medical Work: Decision Support Techniques and Medical Practices: The MIT Press; 1997.
53. Timmermans S, Berg M. The practice of medical technology. Sociol Health Illn 2003;25:97-114. [doi: 10.1111/1467-9566.00342] [Medline: 14498932]
54. Mørk BE, Hoholm T, Ellingsen G, Edwin B, Aanestad M. Challenging expertise: on power relations within and across communities of practice in medical innovation. Manage Learn 2010 Nov;41(5):575-592. [doi: 10.1177/1350507610374552]
55. Breuer S, Braun M, Tigard D, Buyx A, Müller R. How engineers' imaginaries of healthcare shape design and user engagement: a case study of a robotics initiative for geriatric healthcare AI applications. ACM Trans Comput Hum Interact 2023;30(2):1-33. [doi: 10.1145/3577010]
56. Collins HM, Evans R. Rethinking Expertise: The University of Chicago Press; 2007. [doi: 10.7208/chicago/9780226113623.001.0001]
57. Akrich M, Rabeharisoa V. On the multiplicity of lay expertise: an empirical and analytical overview of patient associations' achievements and challenges. In: Eyal G, Medvetz T, editors. The Oxford Handbook of Expertise and Democratic Politics: Oxford University Press; 2023:103-133. [doi: 10.1093/oxfordhb/9780190848927.013.5]
58. Greenhalgh T, Wherton J, Shaw S, Papoutsi C, Vijayaraghavan S, Stones R. Infrastructure revisited: an ethnographic case study of how health information infrastructure shapes and constrains technological innovation. J Med Internet Res 2019 Dec 19;21(12):e16093. [doi: 10.2196/16093] [Medline: 31855184]
59. Papoutsi C, Wherton J, Shaw S, Morrison C, Greenhalgh T. Putting the social back into sociotechnical: case studies of co-design in digital health. J Am Med Inform Assoc 2021 Feb 15;28(2):284-293. [doi: 10.1093/jamia/ocaa197] [Medline: 33043359]
60. Star SL, Strauss A. Layers of silence, arenas of voice: the ecology of visible and invisible work. Comput Support Coop Work 1999 Mar;8(1-2):9-30. [doi: 10.1023/A:1008651105359]
61. Davies SR, Holmer C. Care, collaboration, and service in academic data work: biocuration as 'academia otherwise'. Inf Commun Soc 2024 Mar 11;27(4):683-701. [doi: 10.1080/1369118X.2024.2315285]
62. Giardullo P. Theoretical inputs for empirical outputs: the lenses of science and technology studies from digital media for exploring algorithms in healthcare. In: Miele F, Giardullo P, editors. Reframing Algorithms: STS Perspectives to Healthcare Automation: Springer International Publishing; 2024:19-44. [doi: 10.1007/978-3-031-52049-5_2]
63. Kubes T. A critique of human-centred AI: a plea for a feminist AI framework (FAIF). AI Soc 2026 Apr;41(4):2827-2838. [doi: 10.1007/s00146-025-02556-8]
64. Suchikova Y, Tsybuliak N, Teixeira da Silva JA, Nazarovets S. GAIDeT (Generative AI Delegation Taxonomy): a taxonomy for humans to delegate tasks to generative artificial intelligence in scientific research and publishing. Account Res 2026 Apr;33(3):2544331. [doi: 10.1080/08989621.2025.2544331] [Medline: 40781729]

Abbreviations

AI-CDSS: AI-driven clinical decision support systems
CDS: Clinical Data Science
CDSS: clinical decision support system
COREQ: Consolidated Criteria for Reporting Qualitative Research
HCI: human-computer interaction
HES: Department of Health, Ethics and Society
PUS: public understanding of science
STS: science and technology studies
TRIPOD-AI: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis—Artificial Intelligence

XSL•FO
RenderX

Edited by S Law; submitted 22.Apr.2026; peer-reviewed by A Sutcliffe, Z Yu; revised version received 26.Aug.2026; accepted 25.Sep.2026; published 13.Nov.2026.

Please cite as:
Mingels S, Piehl H, Fijten R, Janssen R, Penders B
Investigating Context in How Clinicians Make Sense of AI: Qualitative Interview Study
JMIR Hum Factors 2026;13:e99117
URL: https://humanfactors.jmir.org/2026/1/e99117
doi:10.2196/99117

© Simone Mingels, Hannah Piehl, Rianne Fijten, Ricky Janssen, Bart Penders. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 13.Nov.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

A Log-Level Data-Driven Precision Education Tool for Pediatrics Trainees: Human-Centered Development and Validation Study

Alexander Fidel[1]; Mark V Mai[2], MHS, MD; Naveen Muthu[2], MD; Adam C Dziorny[3], MD, PhD

[1]Center for Healthcare Quality and Analytics, Children's Hospital of Philadelphia, 734 Schuylkill Ave, Philadelphia, PA, United States

[2]Department of Pediatrics, School of Medicine, Emory University, Atlanta, GA, United States

[3]Department of Pediatrics, School of Medicine, University of Rochester, Rochester, NY, United States

Corresponding Author:
Alexander Fidel
Center for Healthcare Quality and Analytics, Children's Hospital of Philadelphia, 734 Schuylkill Ave, Philadelphia, PA, United States

Abstract

Background: Exposure to patients and clinical diagnoses drives learning in graduate medical education (GME). Measuring practice data, how each trainee experiences that exposure, is critical to planned learning processes including the assessment of trainee needs. We previously developed and validated an automated system to accurately identify resident provider-patient interactions.

Objective: In this follow-up study, we use human-centered design methods to meet two objectives: (1) to understand trainees' planned learning needs and (2) to design, build, and validate the usability and use of a tool based on our automated resident provider-patient interaction system to meet these needs.

Methods: We collected data from 2 institutions new to the American Medical Association's "Advancing Change" initiative, using a mixed methods approach with purposive sampling. First, interviews and formative prototype testing yielded qualitative data that we analyzed in several coding cycles. We built interview guides to collect data required for a work domain assessment, learning use case elicitation, and ultimately design requirement identification. We structured coding efforts within 2 existing theoretical models. Feature prioritization matrix analysis then transformed qualitative analysis outputs into actionable prototype elements that were refined through formative usability methods. Finally, qualitative data from a summative usability test validated the final prototype with measures of usefulness, usability, and intent to use. We used quantitative methods (eg, time on task and task completion rate in summative testing).

Results: We represented the GME work domain assessment through process-map-design artifacts that provide target opportunities for intervention. Of the identified decision-making opportunities, trainee-mentor meetings stood out as optimal for delivering reliable practice-area information. We designed a "midpoint" report for the use case of such meetings. We arrived at a final prototype through formative testing and design iteration. This final version showed 5 essential visualizations. Summative usability testing resulted in high performance in subjective and objective metrics. Insufficient baseline data were captured to draw comparative conclusions in a formal evaluation against existing tools or workarounds to support planned learning. However, the prevailing reported absence of tools and the ad hoc nature of approaches that do exist strongly imply an unmet need for the type of usable summary method delivered in our tool. We collected data from June 2021 through September 2023. Eight resident physicians composed the validation sample, including 4 (50%) residents from the Children's Hospital of Philadelphia and 4 (50%) residents from the University of Rochester Medical Center.

Conclusions: We describe the multisite development of a tool providing visualizations of log-level electronic health record data, using human-centered design methods. Delivered at an identified point in GME, the tool is ideal for fostering the development of master adaptive learners. The resulting prototype is validated with high performance on a summative usability test. Additionally, the design, development, and assessment process may be applied to other tools and topics within clinical informatics.

(JMIR Hum Factors 2026;13:e79952) doi:10.2196/79952

KEYWORDS

planned learning; human-centered design; HCD; decision support; experience tracking; human-computer interaction; cognitive ergonomics; human factors engineering

XSL•FO
RenderX

Introduction

Experiential learning through direct patient care is an important paradigm in graduate medical education (GME) [1]. Readily available measures of patient exposure are necessary to quantify gaps in trainee experience [2]. Such objective tools can provide data to guide a trainee's individualized education plan and enhance future learning [3,4]. The reflective practice and precision education (PE) conceptual models build on the Master Adaptive Learner (MAL) model, pointing to patient exposure data powerfully informing the planning phases of trainee education [5-7]. Planned learning in this context amounts to educational experiences that are structured by the institution, but also extends to intentional, goal-directed behavior arising organically from metacognition surrounding practice. Planned learning is integrated into the residency or fellowship curriculum either by program directors or the learners themselves. It can be the decision to take one elective over another, but it can also be the decision to look out for and seize opportunities to get exposure to a certain patient population or procedure. It is distinct from informal, spontaneous learning that might occur during daily patient care without specific predefined objectives or evaluation methods. This intentionality is anchored in the MAL model, which holds that material learning occurs by cycling through states of planning, learning, assessing, and adjusting.

Using practice data, trainees may situate their experience within their cohorts, identifying targeted opportunities for learning [6]. Although manually tracked case logs by trainees themselves show promise in providing such data, those methods are labor-intensive and limited in their ability to scale [8,9]. Several technology-based systems have been developed to identify trainees' patient experiences automatically and accurately across rotations [10-12]. PE of this kind in GME has recently been advanced as a data-driven approach for improving personalization and efficiency of learning, assessment, and feedback by using longitudinal learner data and analytics. In this framing, programs move beyond aggregate case logs or episodic evaluations toward fruitful cycles of the MAL [13,14]. And while reliable characterization of each trainee's clinical experiences at deep granularity has emerged as a key PE enabler, it remains unknown how to optimally deliver this information to trainees [15,16].

GME trainees need targeted delivery of specific information and knowledge to enhance their educational experience. *Educational* decision support (EDS) deals with such information [17]. We posit that EDS systems are as necessary as clinical decision support (CDS) systems that support clinical duties. To this end, authors initially developed and validated TRAILS (Trainee Individualized Learning System), an automated software platform that uses electronic health record (EHR) metadata—particularly audit logs, note authorship, order placement, and care team assignments—to identify clinically meaningful resident provider-patient interactions across multiple care settings for medical education tracking and assessment [11,18]. We used EHR audit log time stamps and algorithmically calculated interaction intervals by summing time gaps between consecutive events to derive total time-in-chart, a computed metric with strong predictive power to add to raw metadata features. Data cleaning steps like exclusion of data when total time-in-chart is less than 1 hour ensured the models accurately distinguished genuine clinical interactions from incidental chart access. We then applied logistic regression classifiers trained on resident-validated resident provider-patient interactions.

To be useful and usable, EDS output like that of TRAILS must deliver the right information to the right subset of users, at the right time, through the right information channels, and in the most usable format [19]. Traditional technology-centered design (TCD) approaches are insufficient, with outcomes of increased complexity for users, elevated error rates, poor intention to use, and abandoned adoption [20,21]. TCD approaches fit an interface to work-as-imagined and engage users late in the process, if at all. The alternative, human-centered design (HCD), begins by assessing user goals, tasks, abilities, and cognition prior to design ideation. By first identifying contextual needs through requirements elicitation with end users, followed by iterative prototyping and formative evaluation, HCD interface creation instead fits work-as-done, and better results follow: simplified use, low error rates, high affinity, and eventually, high adoption [22,23].

The objective of this study was to follow an HCD approach to design and develop a user interface overlaying an existing automated EDS system, and to measure the summative usability of the design product. Specifically, we sought to: assess the trainee learning work domain and overall context for system implementation; use this knowledge to inform initial designs; apply formative and summative evaluation techniques to the tool; use mixed methods for data collection consistent with HCD standards. We hypothesized that this approach would generate a prototype, meeting acceptable benchmarks for technological acceptance, predicting eventual adoption. The endorsement of such a prototype, assessed to have good usability, would support that our approach could be generalized to other medical applications. The HCD approach we targeted was synthesized across diverse health innovations and health IT including CDS [24-26]. It is also supported by systematic reviews in broader human-computer interaction or software-engineering contexts [27,28]. These suggest that any one viable HCD method is transferable beyond a single specialty or use case.

Methods

Study Design

We performed a prospective mixed methods study within 3 distinct phases of data collection (Table 1) and analysis. Participants were physicians from 2 large academic institutions: the Children's Hospital of Philadelphia (CHOP) and the University of Rochester Medical Center (URMC). Annual trainee cohort sizes are over 500 and over 900, respectively [29,30]. These programs cultivate cultures of planned learning by embedding residents in mentored research projects and data-driven quality improvement teams where they critically reflect on performance metrics, identify specific gaps in outcomes, and develop targeted interventions that close those gaps through iterative, supervised cycles of implementation and

XSL•FO
RenderX

assessment [31,32]. We conducted the study over a 2-year period, from June 2021 to September 2023.

Table . Data collection methods.

Principal data collection method	Participant recruitment[a]	Moderation[b] and setup	Data
Semistructured interview	Trainees and individuals responsible for trainee coaching and program-level planned learning.	Participant was asked questions from role-appropriate interview guide	• Recorded and transcribed session • Moderator notes
Prototype walkthrough interview	Trainees and faculty mentors	Participant was given most recent prototype iteration of planned learning tool and asked to share screen during walkthrough	• Recorded and transcribed session • Moderator notes
Summative usability test	Residents	Participant was given final prototype of planned learning support tool, the "MV[c] midpoint report" and directed to complete test on REDCap[d] (Vanderbilt University) using the report	• Time-stamped, REDCap-collected answers to task questions • Responses to TAM[e]-based Likert psychometrics assessing PU[f], PEOU[g], and scenario realism.

[a]All samples purposive; interview phase saturation was reached succeeding first session when no new data could be synthesized in analysis; repeat participation in or across phases was disallowed.

[b]Each session in all methods was conducted with 1 participant at a time with 1 moderator over Teams (Microsoft) or Zoom (Zoom Communications, Inc) over 1 hour.

[c]MV: minimum viable.

[d]REDCap: Research Electronic Data Capture.

[e]TAM: technology acceptance model.

[f]PU: perceived usability.

[g]PEUO: perceived ease of use.

For coding paradigms in all qualitative research, a 4-member coding team reached cooperative consensus. All had full access to the complete qualitative datasets. Each team member independently reviewed first-cycle coding of all raw data prior to weekly consensus sessions. In these meetings, the team interrogated discrepant interpretations and pursued dialogical intersubjectivity until group consensus was reached. We also refined codebooks in this way, revising, collapsing, or generating new codes, categories, and themes as warranted.

Work Domain Assessment

The first phase of an HCD process, work domain assessment, identifies individuals' roles and details the overall context or work system in which a prototype tool would be used [33,34]. We conducted semistructured interviews with a purposive sample of stakeholders involved in trainee-planned learning. Collecting data from all roles involved is essential to characterize a sociotechnical system of clinical work accurately [35]. Role-specific interview guides covered 2 subjects: trainee sensemaking and decision-making in practice-area learning; mentor or program perspectives on entrustment ladders and how practice-area gaps are identified, discussed, and closed (Multimedia Appendix 1).

We coded the corpus of qualitative data (Table 2). Notably, we borrowed constructs in the Systems Engineering Initiative for Patient Safety to create early cycle topic codes [35,36]. From early cycle coding analysis, we created process maps visually representing dynamics of planned learning.

XSL•FO
RenderX

Table . Qualitative analysis methods used in the study.

Coding type	Purpose	Output
Descriptive coding [37]	Identify topics in the interview corpus.	Table of quotes from interviews and their labels
Provisional coding [36]	Deductively organize topics into categories using a "start list" code book from SEIPS[a] [35]. Improve the design team's grasp of users [38,39].	Expanded table with labels fit to topics and categories
Process coding and causation coding [40,41]	Abstract dynamism or change in topics or categories into graphical representations: SEIPS topics acting as barriers or facilitators to learning; mental models acting on decisions; evolving intentions, choices, objectives, values, perspectives, needs, desires, and agency.	Process maps explaining contexts, conditions, interactions, and consequences of planned learning
Inductive coding [42]	Abstract process map and code table data into questions our users ask, answerable with trainee EHR[b] metadata filtered on clinically meaningful rPPIs[c] and encounter context fields required by that question (eg, care setting or elective label).	Column in feature prioritization matrix
Magnitude coding [43]	Assigns each user question an impact score: numerical value (from 1 to 5 where 5="high impact") showing impact an informed answer would have to planned learning weighing use of the described learnings and degree of relevance to scenarios described in interviews.	Column in feature prioritization matrix

[a]SEIPS: Systems Engineering Initiative for Patient Safety.

[b]EHR: electronic health record.

[c]rPPI: resident provider patient interaction.

Formative Usability Test

To begin formative testing, we required an initial prototype. We initiated further coding processes (Table 1), first inferring key user questions, such as "Will I see important diagnoses if I choose elective X?" These questions formed the basis of an important design artifact in HCD, the feature prioritization matrix (FPM; Figure 1).

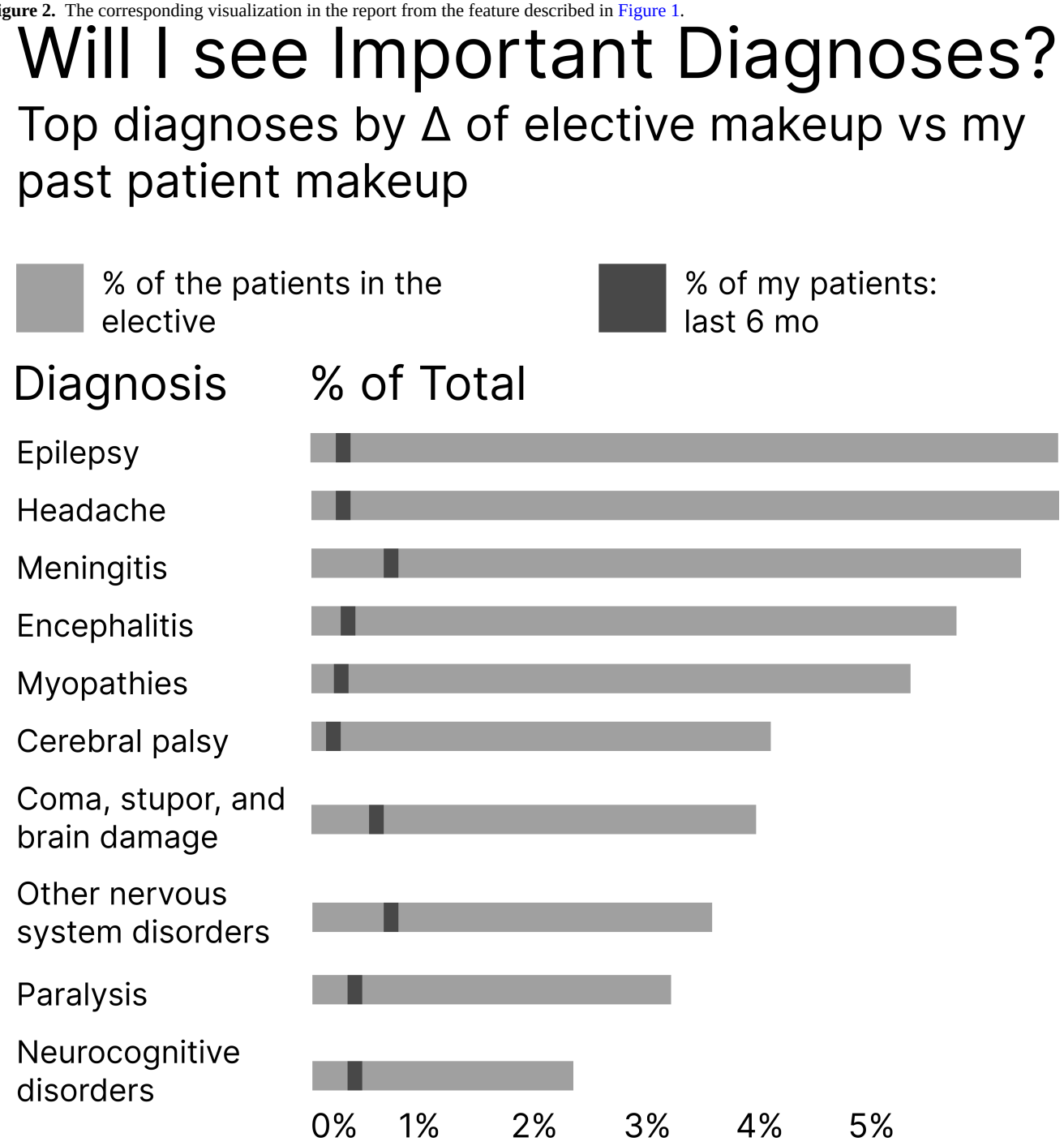
Figure 1. An example of a single record from the feature prioritization matrix.

Visualization Title	Report Section (question answered by graph)	What is displayed? (Metadata displayed and analytic logic)	Date Range for Data Pull	Impact Score (1=low; 5= high)
Top diagnoses by Δ of elective makeup vs my past patient makeup	Will I see important diagnoses (if I choose this elective)?	Count of all diags seen by same-year-as-user trainees. Group by diags. Show as a % of total. Show count of diags seen by the user-trainee as a comparison.	Last 6 Months	4

We produced an initial prototype of a planned learning support tool using the FPM from a collaborative unmoderated design session informed by reviewing recently completed process maps. We reached consensus on data visualizations that would allow trainees to identify or close gaps in experience (Figure 2). This action is the core of planned learning, and we formed the first report with visualizations supporting it.

XSL•FO
RenderX

Figure 2. The corresponding visualization in the report from the feature described in Figure 1.



We then iterated on the design in interview walkthroughs with users. With each walkthrough, newly elicited features and corresponding visualizations were added to the FPM and report. Using an impact metric from the completed FPM (Multimedia Appendix 2) and descriptions of required data for each visualization, we determined which features the current system architecture could feasibly support. We then produced a final version of our planned learning tool, the "minimum viable (MV) midpoint report," including only the most impactful, feasible visualizations (Multimedia Appendix 3).

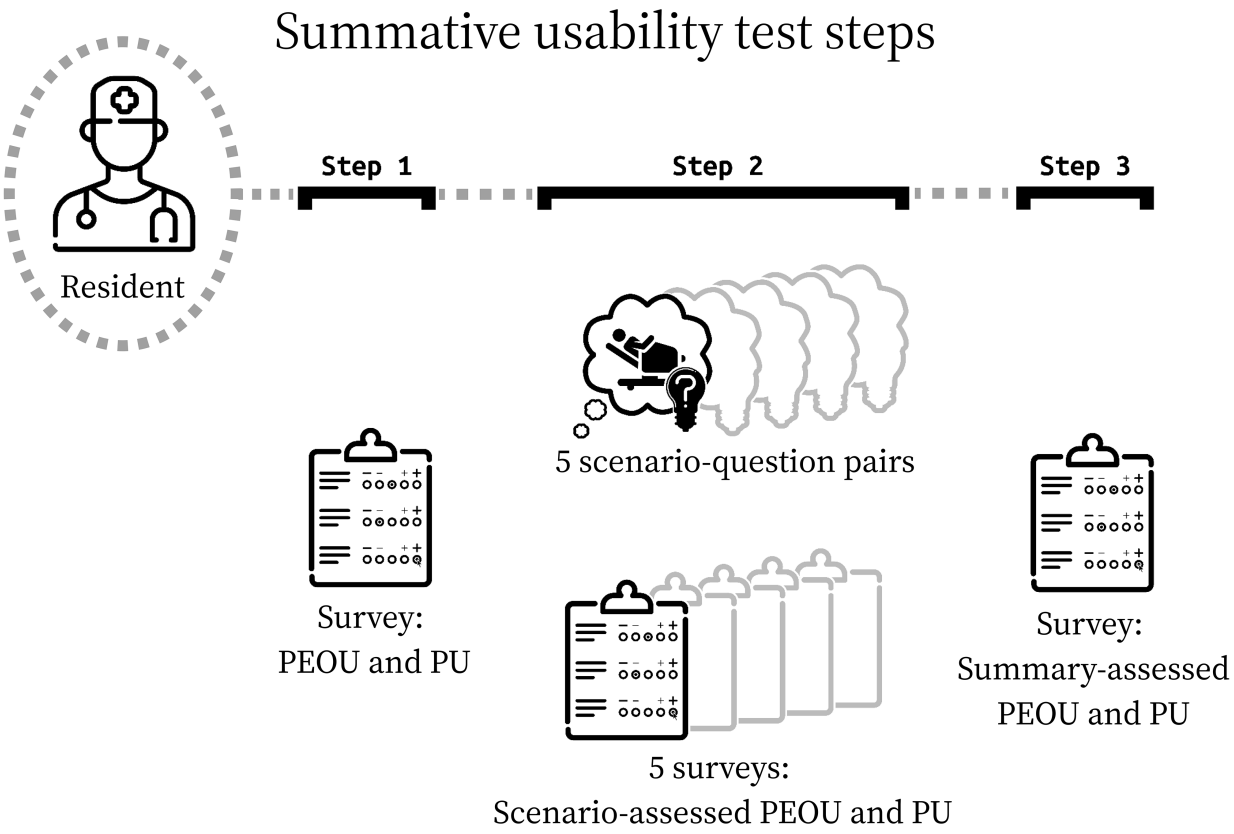
Summative Usability Test

Whereas formative usability testing is meant to expose design flaws and drive prototype iteration forward, summative usability testing is meant to validate a prototype that designers theorize is ready for deployment. We developed a summative usability test to assess usability and use of the "MV midpoint report" and measure performance on MAL planning-related tasks using the report (Multimedia Appendix 4). Participants completed 5 tasks (Figure 3). For each task, participants used the report to answer one of the following key user questions:

1. What are the diagnoses to which I have received the most exposure?
2. In what clinical environments am I seeing them?
3. What gaps in diagnosis exposure could I be filling?
4. What is my exposure to acuity and complex care?
5. In each elective available to me, will I see important diagnoses?

XSL•FO
RenderX

Figure 3. Summative usability test procedure for a single session, where a resident participant moves through the 3 steps, guided by a moderator. PEOU: perceived ease of use; PU: perceived usability.



In this way, the final cycle of analysis and design in the HCD process would be complete (Multimedia Appendix 5).

For each task, we developed a scenario with realistic clinical data and a single complementary test question to which an MV report feature offered an answer. The purpose of each scenario-question pair was to assess usability and use. Four of the 5 scenarios were paired with a multiple-choice question. A fifth scenario was paired with a free-text response. Participants in this usability test were thus assessed with these 5 scenario-question pairs.

Participants were given the MV report in advance of the test for optional review. Participants were told to suspend disbelief and answer questions in the test as though the report were their own, aggregating their own practice area data and encountering the described scenarios during residency.

Participants completed a pretest questionnaire that asked postgraduate year of the participant, whether and how they kept records of their practice areas, and if so, perceived usability (PU) and perceived ease of use (PEOU) of that system. This quantified the usability of participants' baseline approaches to tracking practice-area exposure (if any) for comparison against our tool. PU and PEOU questions were adapted from the development of another successfully implemented CDS tool (Table 3) [44].

XSL•FO
RenderX

Table . Measures within the summative test and the group of user beliefs which they assess.

Measure[a]	Score group
This scenario is important in finding learning or entrustment gaps and resolving them.	Test realism
This scenario is realistic in the search for learning or entrustment gaps and resolving them.	Test realism
In this scenario, the system was easy to use.	PEOU[b]
In this scenario, the system allowed me to perform tasks efficiently.	PEOU
Overall, I am satisfied with how the system is designed in this scenario.	PEOU
In this scenario, the system provided useful features.	PU[c]
In this scenario, the system provided useful information.	PU
In this scenario, the system is an improvement over what I would have used before.	PU

[a]Measures were assessed on a Likert scale from 1 (strongly disagree) to 9 (strongly agree). Participants were assessed after each scenario and then once more in a summary assessment at the end of the test.

[b]PEUO: perceived ease of use.

[c]PU: perceived usability.

We report these descriptive statistics for the groups of measures that assessed PU and PEOU immediately after scenarios, PU and PEOU summarily at the very end of the test, and for the group assessing the realism of the scenarios. Scenario-question pairs were scored for correctness and used to calculate the task completion rate (TCR) by dividing correct answers by total answered questions. We also report average task completion time as the amount of time spent from a scenario-question load page to the submission of a *correct* response.

Ethical Considerations

This study protocol was reviewed by the institutional review boards (IRBs) of both CHOP (IRB 18-014866) and URMC (study 00006717) and determined to meet exemption criteria. We obtained a waiver of written informed consent from both IRBs. Participant data were deidentified and stored by random identifier to protect subject privacy. The participants were not compensated for their participation.

Results

Work Domain Assessment Results

We interviewed 8 participants (Table 4). We deployed initial descriptive and provisional coding efforts. Starting from the top of the codebook, 4 People, Environment, Tools, and Tasks scan categories were decomposed into 34 topics which yet further broke down into 57 subtopics. Multimedia Appendix 5 links key quotations from the interviews for each of the People, Environment, Tools, and Tasks categories and traces their impact forward through all study phases.

Table . Participants for all study phases.

Phase	Purpose	Sample
Work domain assessment (interview)	Understand work system or planned learning decisions	2 trainees, 5 program directors, and 1 administrator
Formative usability testing (prototype walk-through)	Iterate report design	6 mentors and 1 resident
Summative usability testing	Validate EDS[a] tool (MV[b] midpoint report) with end users	8 residents

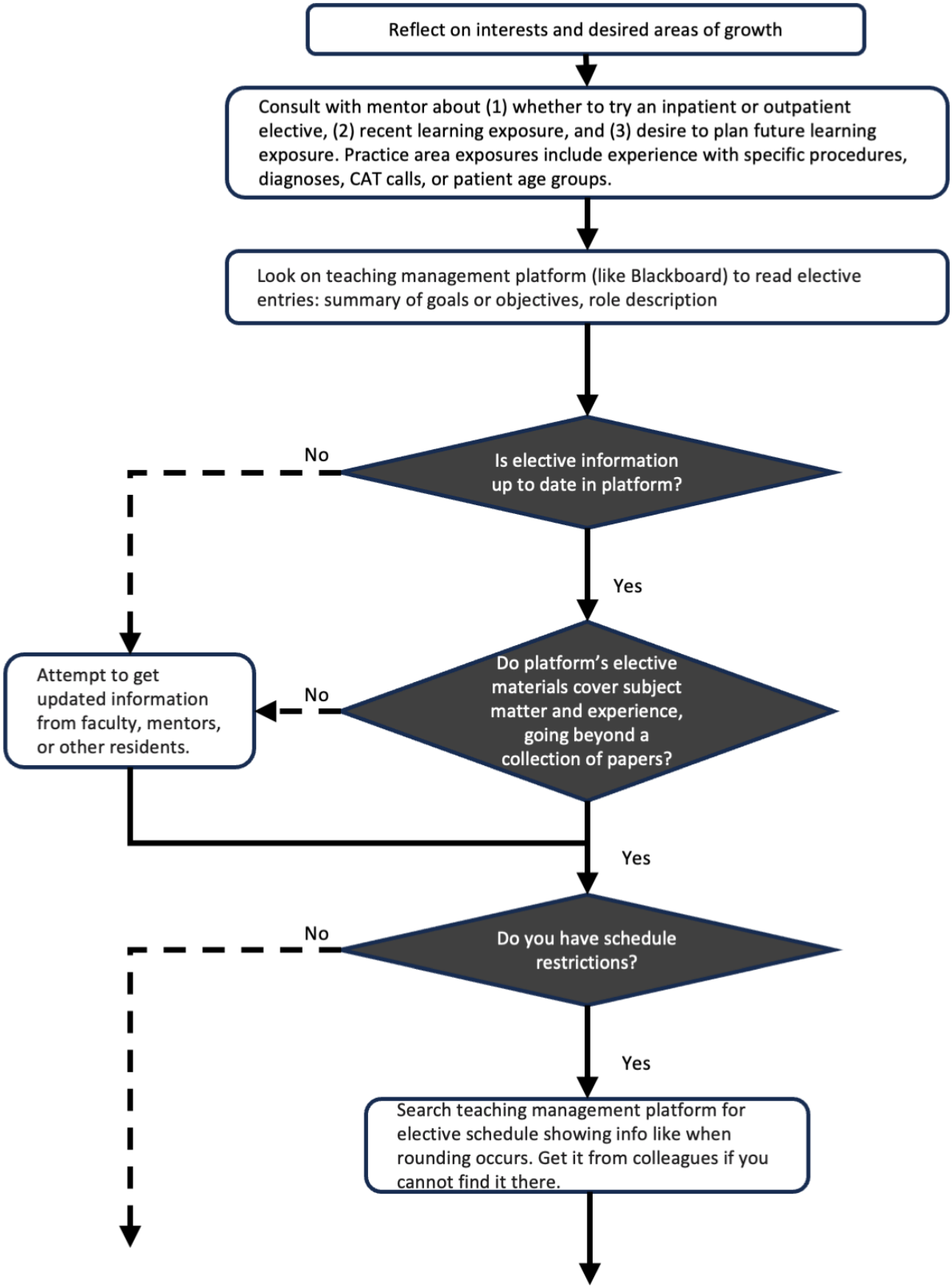
[a]EDS: educational decision support.

[b]MV: minimum viable.

Through our process or causation coding and process map analysis, we produced 7 maps describing processes across 3 roles: trainees (Figure 4), trainee mentors or program directors, and program administrators. The documents mapped 52 decisions or forks in the processes with 128 other steps, actions, or changes in the mental model (Multimedia Appendix 6).

XSL•FO
RenderX

Figure 4. A segment of 1 of the 7 process maps. Diamonds denote decisions and rectangles stand for events, actions, or changes in mental model. CAT: Critical Activation Team.



We first established that mentors and program directors had fewer existing assessment workflows than trainees. While their leadership and expertise amplified potential impact they could achieve with better practice data, trainees had a far greater

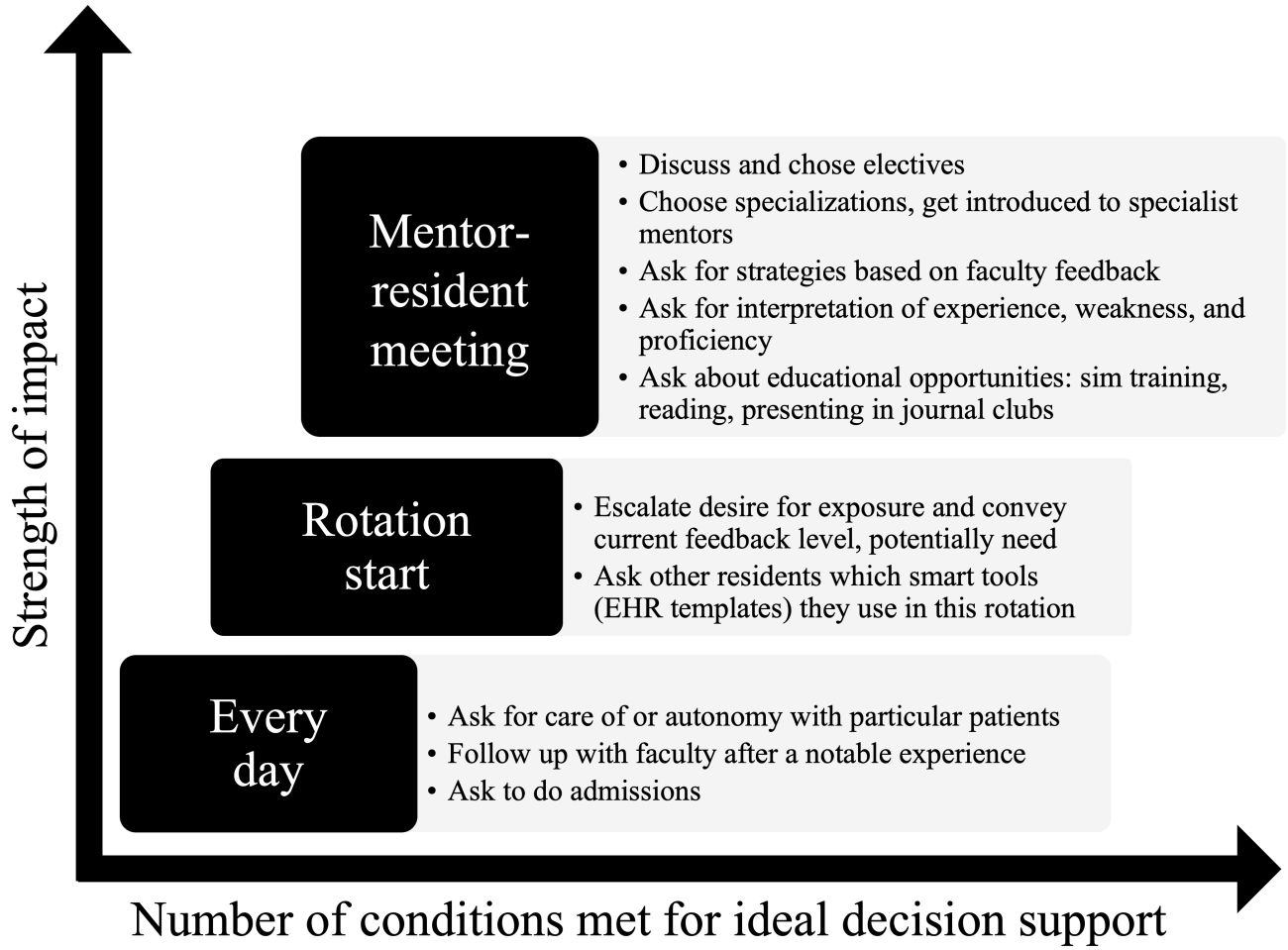
XSL•FO
RenderX

volume of opportunity to identify gaps and close them for themselves. Residents, in particular, had the highest number of planning and assessment opportunities. While other populations could still be influential secondary users of an EDS, residents emerged as optimal primary users of a decision support tool targeting planned learning.

Reviewing the process maps, we determined that clinic and rotations were high-inertia settings with high mental stimulation for our trainee users. The heavy demands on cognitive resources meant that planning in these situations would be secondary to a primary use case that also emerged from the data.

Alternatively, trainee-mentor meetings, conducted much less frequently, stood out as a key target experience. By the time trainees meet with their mentors, they have usually formed some idea of a learning state gap but lack the expert system knowledge required for planning its closure. The MAL model holds that planned learning occurs by cycling through states of planning, learning, assessing, and adjusting. These meetings offered a calm moment when trainees and mentors cooperatively identify gaps, make plans to close them, or assess those efforts. This is the critical work of the planning and assessing phases. This trainee-mentor meeting environment had other benefits. It also allowed for contingency plans and provided a structure for replanning in future meetings, both of which are optimal planning practices from the standpoint of human factors engineering and cognitive science (Figure 5).

Figure 5. Settings where a report of practice area data could be leveraged and planned learning opportunities for each. These are plotted against qualities the research team used to select the target setting of the report. EHR: electronic health record.

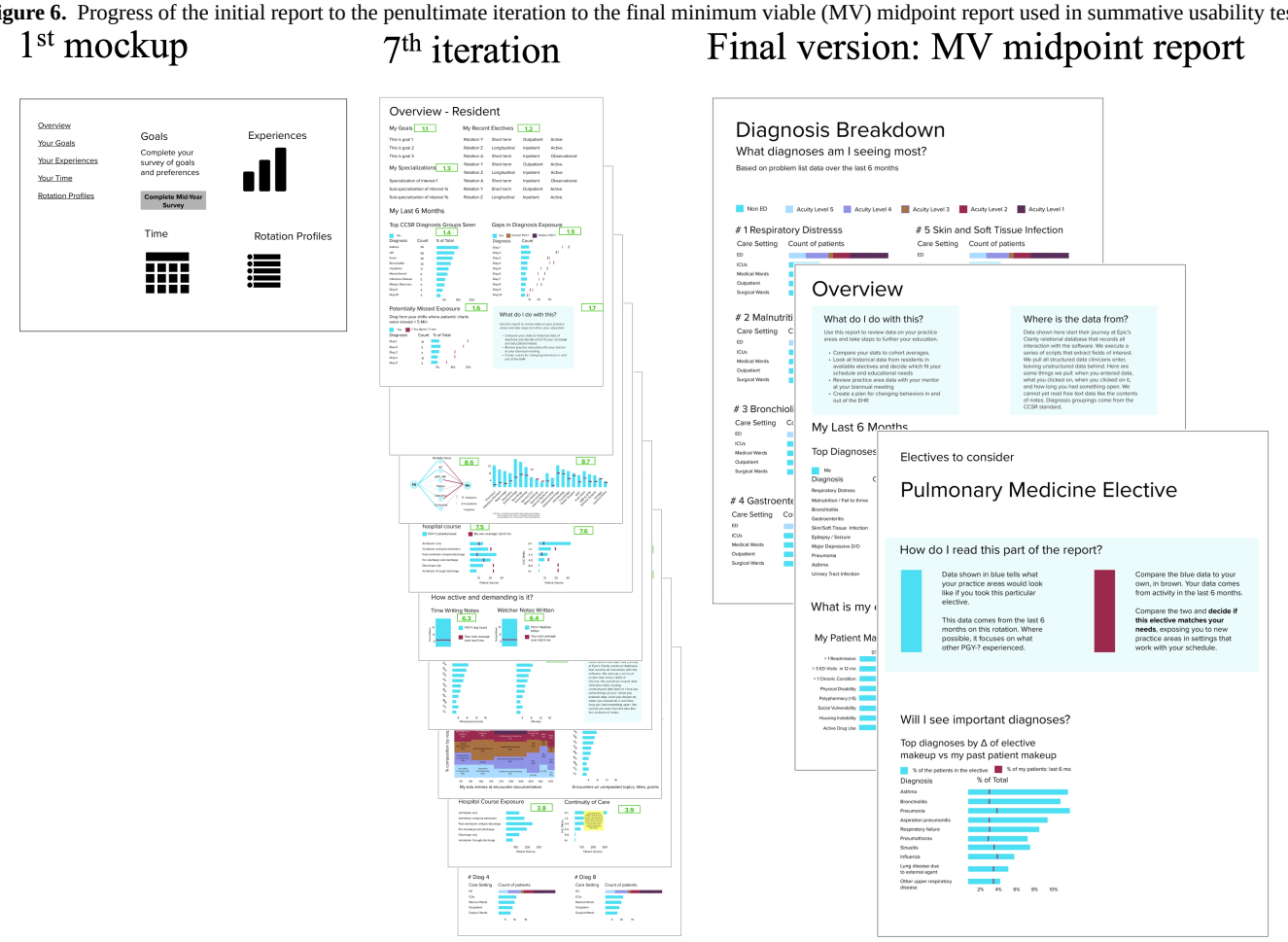


Formative Usability Test Results

Based on the results of our workflow assessment, we designed an initial report containing a single page of placeholder visualizations. Seven participants (6 program directors and 1 resident) completed the formative usability test, iterating from this initial version. Iterations to the midpoint report were made following each test, adding new and altering extant features. The seventh version of the report following the final formative interview contained 51 distinct visualizations of trainee past and potential future data, spanning nine pages (Figure 6). The complete FPM contained a corresponding 51 records (Multimedia Appendix 6). Following feature reduction, the MV midpoint report contained 5 visualizations over 3 pages, assuming one elective under consideration where an additional page would be added for each elective.

XSL•FO
RenderX

Figure 6. Progress of the initial report to the penultimate iteration to the final minimum viable (MV) midpoint report used in summative usability test.



Summative Usability Test Results

Eight resident physician participants in total completed the summative test; 4 each from CHOP and the URMC (n=4, 50%). The TCR across all 5 scenario-question pairs, defined by the selection of the correct answer to a scenario's accompanying multiple-choice question, was 78%. Mean task completion time was 2 minutes and 39 seconds (SD 2 minutes and 30 seconds). Results broken out by each of the 5 scenario-question pair tasks (Table 5) show that completion-time variance was primarily driven by task 1 which showed a much longer mean time and wider range than the other tasks. This may have included orientation time with the structure of the test as well as the MV midpoint report. Tasks 3 and 4 most negatively impacted overall TCR. This may be attributed to more salient, earlier visualizations in the report being used in favor of more setting-specific data breakdowns required for these questions. The summative usability test data and capture plan did not allow for more thorough error type analysis.

Table . Summative test results by task (N=8).

Task	Completion time (min:s), mean (SD)	Task completion rate (%)
1	04:31 (4:18)	88
2	01:40 (0:36)	88
3	03:28 (1:24)	63
4	02:43 (1:00)	50
5	01:19 (0:48)	100

The pretest questionnaire provided baseline findings. Seven of the 8 participants reported no systematic method for tracking their practice-area exposure. Only 1 participant reported using an EHR-based workaround (described in a free text survey field as "saving patient reports on Epic"), and rated this approach 5/9 for both PU and PEOU (scores 6 - 9 considered positive). PU and PEOU scores for the MV midpoint report were high (Table 6). The median scores for scenario-assessed PU and scenario-assessed PEOU were 8 (IQR 5-9) and 8 (IQR 6-9), respectively. Scenario realism received a median score of 8 (IQR 6-9). Distinct from the scenario-assessed PU and PEOU that followed each of the 5 scenarios, a single summary questionnaire assessed the same measures at the end of the test.

XSL•FO
RenderX

The median summaries for PU and PEOU were 8 (IQR 8-9) and 7.5 (IQR 7-8.25), respectively.

Table . Performance of the education decision support tool in terms of usability and scenarios in terms of task fidelity in summative usability testing (N=8).

Score group	Median[a] (IQR)
Scenario-assessed (per-task) measures	
Perceived usefulness	8 (5-9)
Perceived ease of use	8 (6-9)
Summary (holistic post-test) measures	
Perceived usefulness	8 (8-9)
Perceived ease of use	7.5 (7-8.25)
Test realism	8 (6-9)

[a]Scale from 1 (nonperformant) to 9 (performant).

Discussion

Principal Findings

In this study, we describe the HCD process used to develop and assess a user interface to an EDS system that we find likely to be adopted. Like many designs for complex sociotechnical systems, an initial phase to better understand the implementation context led to substantial changes in our design goal. We shifted from a tool that trainees might use regularly to a tool for longer-term planning in the context of trainee-mentor meetings and potential choice of future elective rotations. We subsequently followed a typical iterative design and formative testing process. Upon achieving a stable "MV" design, we conducted our summative testing. Given that our goal in summative testing is to predict future adoption of the tool by trainees, we supplement typical usability measures of task completion, efficiency, and perceived satisfaction with an assessment of perceived usefulness and intent to use, consistent with the technology acceptance model (TAM) [45].

Systems have been judged above average in their usability when TAM-based questionnaires reach 80% [46]. High TAM usability scores signify intent to use a tool, and intent is the highest predictor of adoption [45]. This predictive factor is strengthened by high TCR [47]. Unfortunately, it is known in usability research that meta-analysis has not set completion rate benchmarks [48]. Deference, however, to TCR of similar tools at testing is standard practice in their place [49]. In comparable analytics software from education or clinical data analysis in sensemaking, reflective contexts (as opposed to CDS for real-time use preceding immediate, nonreversible patient action), we find similar TCR [46,50,51]. Matching this standard, we may draw the conclusion that our scores suggest the likely adoption of the tool.

The Accreditation Council for Graduate Medical Education requires trainees to experience a range of diagnoses during graduate medical training [1]. Using an EDS system, we can facilitate appropriate diversity of patient experiences and better quantify outcomes in GME using informatics tools [52-54]. Previous efforts to attribute trainee patient experiences have been limited in scope and dissemination. Several automated systems rely solely on EHR data to link clinical practice directly with education to personalized, timely outcomes [10,16,53]. However, these systems do not attribute patient experiences across clinical contexts or institutions. Dashboards to facilitate the feedback of clinical data to both GME and to continuous professional development have had mixed effectiveness, citing development without proper learning system stakeholder involvement and the absence of facilitated precepting to understand their learning gaps [55,56]. Our tool follows established HCD methods to include those users in interaction conception and make precepting intuitive.

With the usability and PU demonstrated in the summative test, we will implement the MV report to assist trainees with a number of tasks: visualizations of their top diagnosis types seen; their gaps in exposure compared to their peers and historical comparisons; their exposure to complex care patients; and most importantly, diagnosis makeup of electives or clinics that trainees could choose to further their exposure in weak areas of knowledge. The report will be delivered to trainees and their mentors before scheduled meetings biannually and will be discussed by the pair. Among other decisions that the report may inform, it will simplify the critical choice of electives or clinics that trainees can choose to better close their training gaps.

Limitations

The formative testing sample over-represented mentors or program leaders relative to trainees; however, the final MV report was evaluated with resident end users in summative usability testing.

The study is limited by the sample size at 2 institutions, which may not represent the breadth of GME experience. Because of this, projected adoption based on high scores on PU and PEOU assessments may not be generalizable, and ongoing measurement will be necessary during implementation.

The level of realism achieved in the simulations required participants to suspend disbelief. All participants had experience and context to create clear mental settings from experience in the medical education system, though the fidelity achieved by descriptive text in immersing participants in that reality is low. In addition, the level of realism was tempered by technological limitations. We could not use actual trainee data from each

XSL•FO
RenderX

specific participant or generalized trainee data from each site. Reports were populated by data that did not map to participants' clinical experience, instead showing a generalized portrait of an "average" resident validated by clinical experts, requiring further suspension of disbelief.

Future work will improve validity deficiencies by using trainee-specific data in usability testing in a multicenter study with a greater sample drawn from more institutions, additionally including error type analysis of failed tasks. Exploration should also attempt to draw insights from beyond the EHR to triangulate them from multiple datasets including those from sources like training management systems, health care workforce management platforms, and other self-reported data. Similar efforts are already underway to support professional development [57,58]. A follow-up longitudinal study should examine click-stream engagement metrics, long-term post-deployment satisfaction via psychometrics, and downstream effects, such as changes in elective selection patterns, increased exposure in previously underrepresented diagnoses, and improvements in competency-based assessments.

Conclusions

We iteratively developed and performed usability testing on a 5-visualization report displaying trainees' aggregate practice area data from the EHR. Our results indicate a high likelihood of the report's adoption as an effective tool in graduate medical education, aligning with the MAL planning phase tasks and allowing them to be completed in a timely manner. We assessed use and usability with instruments and results that predict significant future adoption. Evolving research will examine downstream effects on entrustment following implementation. This study provides a successful user interface for functions executed from an educational decision support system and offers a foundation for further research of clinical and educational applications of this system.

Acknowledgments

The authors would like to acknowledge the program leadership, administration, chief residents, and trainees in the pediatrics residency programs at the Children's Hospital of Philadelphia and the University of Rochester Medical Center.

Funding

This study was partially funded through the American Medical Association's "Accelerating Change in Medical Education" program (ACD and MVM). The funder did not have a role in the conceptualization, design, data collection, analysis, or publication preparation. The authors have no other funding sources to declare.

Data Availability

The datasets generated and/or analyzed during this study are not publicly available due to data privacy restrictions established by our institutional review boards but are available from the corresponding author upon reasonable request.

Authors' Contributions

ACD and MVM conceptualized the study and designed the protocol along with NM and AF. AF, MVM, and ACD recruited participants and conducted interviews. All authors participated in the human-centered design process and analyzed the interview results. AF conducted summative usability testing and prepared the draft manuscript. All authors read, edited, and approved the final manuscript.

Conflicts of Interest

NM reports equity ownership in Phrase Health, a clinical workflow analytics company. The other authors declare no conflicts of interest.

Multimedia Appendix 1
Work domain assessment interview guides.
[DOCX File, 43 KB - humanfactors_v13i1e79952_app1.docx]

Multimedia Appendix 2
Feature prioritization matrix.
[PDF File, 129 KB - humanfactors_v13i1e79952_app2.pdf]

Multimedia Appendix 3
Final prototype minimum viable (MV) midpoint report.
[PDF File, 81 KB - humanfactors_v13i1e79952_app3.pdf]

Multimedia Appendix 4

XSL•FO
RenderX

Summative usability test.
[PDF File, 72 KB - humanfactors_v13i1e79952_app4.pdf]

Multimedia Appendix 5
Quotes and coding leading to summative tasks.
[PDF File, 67 KB - humanfactors_v13i1e79952_app5.pdf]

Multimedia Appendix 6
Process maps.
[PDF File, 2605 KB - humanfactors_v13i1e79952_app6.pdf]

References

1. Common program requirements. ACGME - Accreditation Council for Graduate Medical Education. URL: https://www.acgme.org/programs-and-institutions/programs/common-program-requirements/ [accessed 2026-01-31]
2. Rhee DW, Chun JW, Stern DT, Sartori DJ. Experience and education in residency training: capturing the resident experience by mapping clinical data. Acad Med 2022 Feb 1;97(2):228-232. [doi: 10.1097/ACM.0000000000004162] [Medline: 33983144]
3. Mylopoulos M, Brydges R, Woods NN, Manzone J, Schwartz DL. Preparation for future learning: a missing competency in health professions education? Med Educ 2016 Jan;50(1):115-123. [doi: 10.1111/medu.12893] [Medline: 26695471]
4. Bucalon B, Shaw T, Brown K, Kay J. State-of-the-art dashboards on clinical indicator data to support reflection on practice: scoping review. JMIR Med Inform 2022 Feb 14;10(2):e32695. [doi: 10.2196/32695] [Medline: 35156928]
5. Cutrer WB, Miller B, Pusic MV, et al. Fostering the development of master adaptive learners: a conceptual model to guide skill acquisition in medical education. Acad Med 2017 Jan;92(1):70-75. [doi: 10.1097/ACM.0000000000001323] [Medline: 27532867]
6. Janssen A, Kay J, Talic S, et al. Electronic health records that support health professional reflective practice: a missed opportunity in digital health. J Healthc Inform Res 2023 Jan;6(4):375-384. [doi: 10.1007/s41666-022-00123-0] [Medline: 36744083]
7. Schumacher DJ, Santen SA, Pugh CM, Burk-Rafel J. Foreword: the next era of assessment and precision education. Acad Med 2024 Apr 1;99(4S Suppl 1):S1-S6. [doi: 10.1097/ACM.0000000000005609] [Medline: 38109655]
8. Langdorf MI, Montague BJ, Bearie B, Sobel CS. Quantification of procedures and resuscitations in an emergency medicine residency. J Emerg Med 1998;16(1):121-127. [doi: 10.1016/s0736-4679(97)00252-7] [Medline: 9472773]
9. Sequist TD, Singh S, Pereira AG, Rusinak D, Pearson SD. Use of an electronic medical record to profile the continuity clinic experiences of primary care residents. Acad Med 2005 Apr;80(4):390-394. [doi: 10.1097/00001888-200504000-00017] [Medline: 15793025]
10. Levin JC, Hron J. Automated reporting of trainee metrics using electronic clinical systems. J Grad Med Educ 2017 Jun;9(3):361-365. [doi: 10.4300/JGME-D-16-00469.1] [Medline: 28638518]
11. Mai MV, Orenstein EW, Manning JD, Luberti AA, Dziorny AC. Attributing patients to pediatric residents using electronic health record features augmented with audit logs. Appl Clin Inform 2020 May;11(3):442-451. [doi: 10.1055/s-0040-1713133] [Medline: 32583389]
12. Wang MD, Rosner BI, Rosenbluth G. Where is the digitally silent provider? Development and validation of a team-centered electronic health record attribution model for supervising residents. Acad Med 2023 Jan 1;98(1):62-66. [doi: 10.1097/ACM.0000000000004978] [Medline: 36576768]
13. Burk-Rafel J, Triola MM. Precision medical education: institutional strategies for successful implementation. Acad Med 2025 Jun 1;100(6):655-660. [doi: 10.1097/ACM.0000000000005980] [Medline: 39889718]
14. Desai SV, Burk-Rafel J, Lomis KD, et al. Precision education: the future of lifelong learning in medicine. Acad Med 2024 Apr 1;99(4S Suppl 1):S14-S20. [doi: 10.1097/ACM.0000000000005601] [Medline: 38277444]
15. Bojic I, Mammadova M, Ang CS, et al. Empowering health care education through learning analytics: in-depth scoping review. J Med Internet Res 2023 May 17;25:e41671. [doi: 10.2196/41671] [Medline: 37195746]
16. Janssen A, Coggins A, Tadros J, Quinn D, Shetty A, Shaw T. Using electronic health data to deliver an adaptive online learning solution to emergency trainees: mixed methods pilot study. JMIR Med Educ 2025 Dec 17;11(1):e65287. [doi: 10.2196/65287] [Medline: 41406416]
17. Kotsiantis SB. Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades. Artif Intell Rev 2012 Apr;37(4):331-344. [doi: 10.1007/s10462-011-9234-x]
18. Mai MV, Muthu N, Carroll B, Costello A, West DC, Dziorny AC. Measuring training disruptions using an informatics based tool. Acad Pediatr 2023;23(1):7-11. [doi: 10.1016/j.acap.2022.03.006] [Medline: 35306187]
19. Osheroff JA, Teich J, Levick D, et al. Improving Outcomes with Clinical Decision Support An Implementer's Guide, 2nd edition: Himss Publishing; 2012. [doi: 10.4324/9780367806125]

XSL•FO
RenderX

20. Woods D, Winograd WTT, Woods DD. Challenges for human-centered design. In: Flanagan J, Huang T, Jones P, Kasif S, editors. Human-Centered Systems: Information, Interactivity, and Intelligence: National Science Foundation; 1997. URL: https://www.researchgate.net/publication/261303273_Woods_DD_Winograd_WTT_Challenges_for_Human-Centered_Design_1997_In_Human-Centered_Systems_Information_Interactivity_and_Intelligence_J_Flanagan_T_Huang_P_Jones_S_Kasif_Eds_National_Science_Foundation [accessed 2023-05-09]
21. Boy GA. Human-centered design of complex systems: an experience-based approach. Des Sci 2017;3:e8. [doi: 10.1017/dsj.2017.8]
22. Nguyen TH, Cunha PP, Rowland AF, Orenstein E, Lee T, Kandaswamy S. User-centered design and evaluation of clinical decision support to improve early peanut introduction: formative study. JMIR Form Res 2023 Aug 22;7:e47574. [doi: 10.2196/47574] [Medline: 37606983]
23. Sauro J. A Practical Guide to Measuring Usability 72 Answers to the Most Common Questions about Quantifying the Usability of Websites and Software: Measuring Usability LLC; 2010. URL: https://measuringu.com/wp-content/uploads/2017/05/QuantitativeUsabilityTestOnline.pdf [accessed 2026-02-10]
24. Bayor AA, Li J, Yang IA, Varnfield M. Designing clinical decision support systems (CDSS)—a user-centered lens of the design characteristics, challenges, and implications: systematic review. J Med Internet Res 2025 Jun 20;27:e63733. [doi: 10.2196/63733] [Medline: 40540451]
25. Göttgens I, Oertelt-Prigione S. Application of human-centered design approaches in health research and innovation: a narrative review of current practices. JMIR mHealth uHealth 2021 Dec 6;9(12):e28102. [doi: 10.2196/28102] [Medline: 34874893]
26. Busse TS, Jux C, Laser J, et al. Involving health care professionals in the development of electronic health records: scoping review. JMIR Hum Factors 2023 Jul 10;10:e45598. [doi: 10.2196/45598] [Medline: 37428524]
27. Brhel M, Meth H, Maedche A, Werder K. Exploring principles of user-centered agile software development: a literature review. Inf Softw Technol 2015 May;61:163-181. [doi: 10.1016/j.infsof.2015.01.004]
28. Silva da Silva T, Martin A, Maurer F, Silveira M. User-centered design and agile methods: a systematic review. Presented at: AGILE Conference 2011 (AGILE 2011); Aug 7-13, 2011. [doi: 10.1109/AGILE.2011.24]
29. Graduate medical education statistics. Children's Hospital of Philadelphia. URL: https://www.chop.edu/about-us/annual-report-of-our-financials/graduate-medical-education-statistics [accessed 2026-01-31]
30. Graduate medical education. University of Rochester Medical Center. URL: https://www.urmc.rochester.edu/education/graduate-medical-education [accessed 2026-01-31]
31. Pursuing excellence in clinical learning environments. University of Rochester Medical Center. URL: https://www.urmc.rochester.edu/education/graduate-medical-education/pursuing-excellence [accessed 2026-01-31]
32. Why choose CHOP's Pediatrics Residency Program. Children's Hospital of Philadelphia. URL: https://www.chop.edu/pediatric-residency-program/why-choose-chop-s-pediatric-residency-program [accessed 2026-01-31]
33. Johnson CM, Johnson TR, Zhang J. A user-centered framework for redesigning health care interfaces. J Biomed Inform 2005 Feb;38(1):75-87. [doi: 10.1016/j.jbi.2004.11.005] [Medline: 15694887]
34. Ratwani RM, Savage E, Will A, et al. Identifying electronic health record usability and safety challenges in pediatric settings. Health Aff (Millwood) 2018 Nov;37(11):1752-1759. [doi: 10.1377/hlthaff.2018.0699] [Medline: 30395517]
35. Holden RJ, Carayon P. SEIPS 101 and seven simple SEIPS tools. BMJ Qual Saf 2021 Nov;30(11):901-910. [doi: 10.1136/bmjqs-2020-012538] [Medline: 34039748]
36. Saldaña J. Provisional coding. In: The Coding Manual for Qualitative Researchers, 4th edition: SAGE Publications; 2021:144-147.
37. Saldaña J. Descriptive coding. In: The Coding Manual for Qualitative Researchers, 4th edition: SAGE Publications; 2021:87-88.
38. Wooldridge AR, Carayon P, Hundt AS, Hoonakker PLT. SEIPS-based process modeling in primary care. Appl Ergon 2017 Apr;60:240-254. [doi: 10.1016/j.apergo.2016.11.010] [Medline: 28166883]
39. Larsen EP, Silvestro E, Ferro DF, et al. Using human factors principles to redesign a 3D lab workflow during the COVID-19 pandemic. 3D Print Med 2022 Nov 12;8(1):34. [doi: 10.1186/s41205-022-00161-9] [Medline: 36371509]
40. Saldaña J. Process coding. In: The Coding Manual for Qualitative Researchers, 4th edition: SAGE Publications; 2021:96-100.
41. Saldaña J. Causation coding. In: The Coding Manual for Qualitative Researchers, 4th edition: SAGE Publications; 2021:163-175.
42. Saldaña J. The Coding Manual for Qualitative Researchers, 4th edition: SAGE Publications; 2021.
43. Saldaña J. Magnitude coding. In: The Coding Manual for Qualitative Researchers, 4th edition: SAGE Publications; 2021:72-77.
44. Utidjian LH, Hogan A, Michel J, et al. Clinical decision support and palivizumab: a means to protect from respiratory syncytial virus. Appl Clin Inform 2015;6(4):769-784. [doi: 10.4338/ACI-2015-08-RA-0096] [Medline: 26767069]
45. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172. [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
46. Roa Romero Y, Tame H, Holzhausen Y, et al. Design and usability testing of an in-house developed performance feedback tool for medical students. BMC Med Educ 2021 Jun 23;21(1):354. [doi: 10.1186/s12909-021-02788-4] [Medline: 34162382]

XSL•FO
RenderX

47. Broekhuis M, van Velsen L, Hermens H. Assessing usability of eHealth technology: a comparison of usability benchmarking instruments. Int J Med Inform 2019 Aug;128:24-31. [doi: 10.1016/j.ijmedinf.2019.05.001] [Medline: 31160008]
48. Mator JD, Lehman WE, McManus W, et al. Usability: adoption, measurement, value. Hum Factors 2021 Sep;63(6):956-973. [doi: 10.1177/0018720819895098] [Medline: 31934799]
49. Lewis JR, Sauro J. Item benchmarks for the System Usability Scale. J Usability Stud 2018 May;13(3):158-167 [FREE Full text]
50. Bellmann L, Wiederhold AJ, Trübe L, Twerenbold R, Ückert F, Gottfried K. Introducing attribute association graphs to facilitate medical data exploration: development and evaluation using epidemiological study data. JMIR Med Inform 2024 Jul 24;12(1):e49865. [doi: 10.2196/49865] [Medline: 39046780]
51. Morales KAS. Design heuristics for instructor dashboard interfaces rooted in self-determination theory [Dissertation]. : Iowa State University; 2022 URL: https://www.proquest.com/docview/2711008057 [accessed 2026-01-31]
52. Arora VM. Harnessing the power of big data to improve graduate medical education: Big idea or bust? Acad Med 2018 Jun;93(6):833-834. [doi: 10.1097/ACM.0000000000002209] [Medline: 29538103]
53. Schumacher DJ, Martini A, Holmboe E, et al. Developing resident-sensitive quality measures: engaging stakeholders to inform next steps. Acad Pediatr 2019 Mar;19(2):177-185. [doi: 10.1016/j.acap.2018.09.013] [Medline: 30268426]
54. Weinstein DF. Optimizing GME by measuring its outcomes. N Engl J Med 2017 Nov 23;377(21):2007-2009. [doi: 10.1056/NEJMp1711483] [Medline: 28976869]
55. Hauer KE, Iverson N, Quach A, Yuan P, Kaner S, Boscardin C. Fostering medical students' lifelong learning skills with a dashboard, coaching and learning planning. Perspect Med Educ 2018 Oct;7(5):311-317. [doi: 10.1007/s40037-018-0449-2] [Medline: 30187391]
56. Pizzuti C. Integrating eHealth data analytics into continuing professional development for medical practitioners: an ecosystemic perspective. Front Med (Lausanne) 2025;12:1553479. [doi: 10.3389/fmed.2025.1553479] [Medline: 40224628]
57. Janssen A, Nazir M. Technology-enabled workplace learning through rethinking electronic health records to support performance feedback: protocol for a mixed methods study. JMIR Res Protoc 2025 May 23;14(1):e66824. [doi: 10.2196/66824] [Medline: 40406851]
58. Bucalon B, Whitelock-Wainwright E, Williams C, et al. Thought leader perspectives on the benefits, barriers, and enablers for routinely collected electronic health data to support professional development: qualitative study. J Med Internet Res 2023 Feb 16;25:e40685. [doi: 10.2196/40685] [Medline: 36795463]

Abbreviations

CDS: clinical decision support
CHOP: Children's Hospital of Philadelphia
EDS: educational decision support
EHR: electronic health record
FPM: feature prioritization matrix
GME: graduate medical education
HCD: human-centered design
IRB: institutional review board
MAL: Master Adaptive Learner
MV: minimum viable
PE: precision education
PEOU: perceived ease of use
PU: perceived usability
rPPI: resident provider patient interaction
TAM: technology acceptance model
TCR: task completion rate
TRAILS: Trainee Individualized Learning System
URMC: University of Rochester Medical Center

Edited by A Kushniruk; submitted 01.Jul.2025; peer-reviewed by CL O'Brien, C Liu, KH Lin; revised version received 30.Dec.2025; accepted 18.Jan.2026; published 23.Feb.2026.

Please cite as:
Fidel A, Mai MV, Muthu N, Dziorny AC
A Log-Level Data-Driven Precision Education Tool for Pediatrics Trainees: Human-Centered Development and Validation Study
JMIR Hum Factors 2026;13:e79952
URL: https://humanfactors.jmir.org/2026/1/e79952
doi:10.2196/79952

XSL•FO
RenderX

© Alexander Fidel, Mark V Mai, Naveen Muthu, Adam C Dziorny. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 23.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Shifting Skills in Robot-Assisted Surgery: Case Study With Interdisciplinary Perspectives for Human Factors and Career Research

Thomas Ellwart*, Prof Dr; Julia Birke*, MSc; Johanna T Paul, MSc

Department of Business Psychology, Trier University, Universitätsring 15, Trier, Rheinland-Pfalz, Germany

*these authors contributed equally

Corresponding Author:
Julia Birke, MSc
Department of Business Psychology, Trier University, Universitätsring 15, Trier, Rheinland-Pfalz, Germany

Abstract

Background: The introduction of STARA (smart technology, AI, robotics, and algorithms) changes work processes and the application of skills, depending on specific task design within a given context.

Objective: From a Human Factors perspective, this study uses empirical data from a case study to demonstrate how different task designs influence the application of theater nurses' skills during robot-assisted surgeries. Moreover, from a transdisciplinary perspective, the data elucidate that STARA-related task design extends beyond Human Factors' perspectives, encompassing cross-level effects between task configurations of different surgical units and individual evaluations regarding career motives. Finally, the study aims to integrate interdisciplinary perspectives by outlining theoretical, methodological, and practical implications for both STARA-related Human Factors task design in health care and Sustainable Career Research.

Methods: Empirical data derive from an embedded single-case study design within a hospital. The case study illustrates field experiences from a focus group interview (n=5) and postsurgery surveys (n=36) on Human Factors–related variables (objective task variety, perceived workload, perceived application of professional skills), and variables from Career Research (perceived career sustainability, growth needs).

Results: Higher task variety and skill application during surgeries correlate with long-term career sustainability evaluations. Moreover, career-related growth needs relate to interindividual differences in STARA task evaluations. Interdisciplinary perspectives with respect to theoretical models, practical implications, and methodological challenges are discussed.

Conclusions: The introduction of STARA (smart technology, AI, robotics, and algorithms) changes work processes and the application of skills, depending on specific task design within a given context.

(JMIR Hum Factors 2026;13:e87735) doi:10.2196/87735

KEYWORDS

delivery of health care; nontechnical skills; technical skills; robotic surgical procedures; sustainable career; growth needs

Introduction

Background

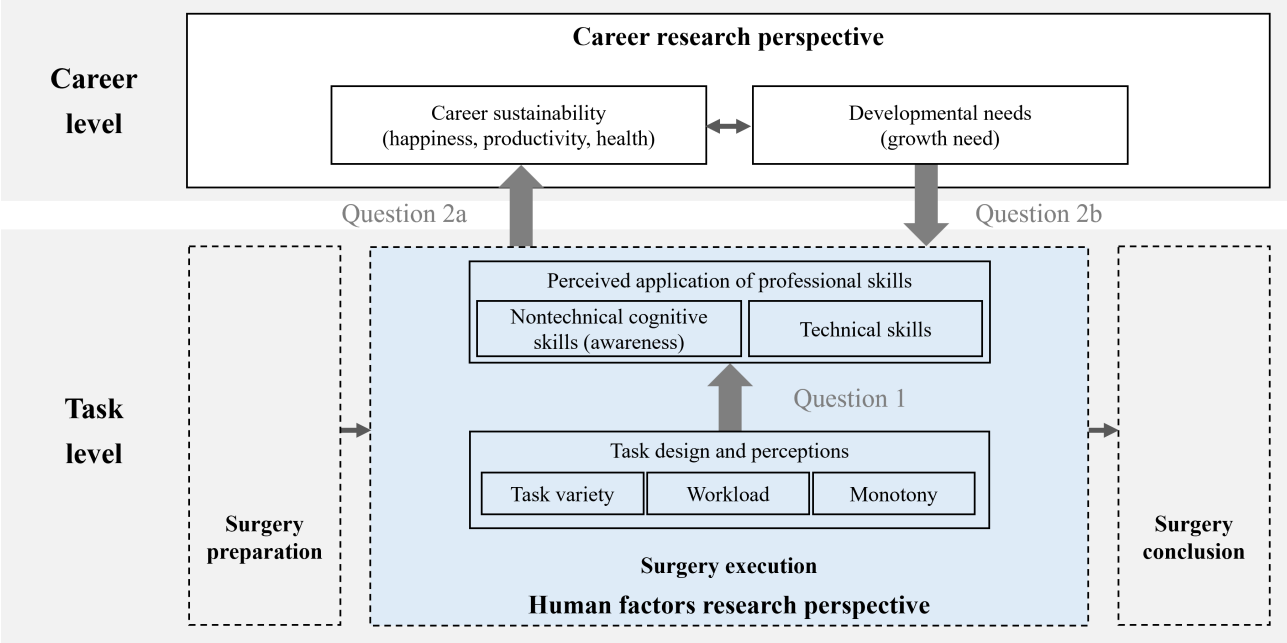
The implementation of smart technologies, AI, robotics, and algorithms (smart technology, AI, robotics, and algorithms; STARA) has substantial interdisciplinary implications in the health care sector. In Human Factors, the focus lies on designing work systems that are human-centered, efficient, effective, and resilient. In the field of human resource management, the emphasis is, for example, on questions of sustainable career development. One prominent example, among many, is the da Vinci Surgical System (Intuitive Surgical Inc), which has significantly transformed the landscape of surgery and, in particular, the role of operating room nurses who assist surgeons during procedures [1]. During surgical execution, the surgeon operates the robotic system from a console, using hand controls to manipulate instruments while viewing the surgical field in detail. However, the surgeon no longer stands face-to-face at the table with the theater nurse and does not interact directly. As a result, nurses' tasks are modified, and task designs can differ across units within a hospital. This results in a varying number of tasks that nurses perform during robot-assisted surgery (RAS), thereby impacting workload, monotony, and the application of professional skills during these tasks. In that respect, the project underpinning the case study focused on a Human Factors perspective to elaborate on the following research question: "How are different task-specific demands in surgical workflows related to perceived skill application?" (see Question 1 in Figure 1). Specifically, nontechnical cognitive skills were investigated to predict safety and resilience of the work-system [2,3]. However, expanding the extensively studied

XSL•FO
RenderX

Human Factors perspective, technological work design goes beyond the task-specific focus. Human resource-oriented career research, for instance, examines the effects of technological advancements on career intentions, yet typically does not incorporate the task-specific theoretical and methodological approaches established in Human Factors Research. This gap reveals considerable interdisciplinary potential between Human Factors and Career Research and may be illustrated by quotes from preparatory interviews with nurses in this study. These interviews indicated that changes in skill application during surgery influence individual career evaluations and, consequently, the sustainability of career projections. As one participant noted, "I did not choose a career as a theater nurse to be bored by watching robotic arms during surgery," summarizing this perception. Against this background, the present study goes beyond the Human Factors perspective and questions "How do task-specific demands and perceived skill application align with reflection of career sustainability?" and vice versa "How do career-related needs relate to situational evaluations of task-specific demands and perceived skill application?" in order to illustrate the relationship between task-level design of robot-assisted surgeries and nurses' career evaluation (see Questions 2a and 2b in Figure 1).

Overall, this study seeks to empirically capture and conceptually discuss the interdisciplinary potential between Human Factors and Career Research, using the example of RAS. Career Research investigates the changing work tasks and skill requirements within the STARA environment [4-6]. Reflecting the career-related experiences of theater nurses in the case study, we focus on the concept of career sustainability [7]. Thus, we aim to integrate task-level concepts from Human Factors into models of Career Sustainability, emphasizing the potential for an interdisciplinary examination.

Figure 1. Task and career level perspectives of theatre nurses in RAS.



Study Aims

Empirical Human Factors Perspective: Task Design and Skill Application in RAS

Using exploratory data from a case study, we aim to examine, from a Human Factors perspective, how different task-specific demands in surgical workflows of RAS are related to perceived skill application. We explore task-specific concepts from Human Factors Research within a socio-technical environment, highlighting functional task design as essential for ensuring human-centered work systems [8,9]. This focus includes task-related variables such as task variety, perceived workload, and monotony, alongside the perceived application of professional skills during surgical situations (Question 1).

Empirical Transdisciplinary Perspective: Human Factors and Career Evaluation in RAS

Through a case study, we aim to illustrate that STARA-related changes in tasks of theater nurses (eg, surgery execution) go beyond Human Factors' questions and address cross-level effects between task-level related changes and individual perceptions at career level (Question 2). In a field sample, we explore nurses' reflections on career sustainability after RAS with different degrees of task variety and skill application (Question 2a). Vice versa, the influence of individual career-related needs on the Human Factors' variables is explored with respect to interindividual differences in task evaluation (Question 2b).

Interdisciplinary Perspectives: Theoretical, Methodological, and Practical Implications

Following the exploratory case study, theoretical, methodological, and practical implications for Career Research and Human Factors are presented. The data collected contributes to a better understanding of the impact and potential effectiveness of career development and workplace interventions. In 5 implications, we elaborate on the addition of variables at the task-level in existing frameworks for sustainable careers, reflect approaches to human-centered work design, and discuss methodological issues.

XSL•FO
RenderX

The subsequent section briefly introduces the theoretical key concepts from Human Factors and Career Research to elaborate the empirical research questions. It then outlines the study's contribution, followed by a presentation of the empirical study in the method and results sections and, in the discussion, a more in-depth reflection on the interdisciplinary perspectives.

Theoretical Context

STARA and Human Factors Research

Overview

Since the 1980s, the "ironies of automation" have been a central topic in the multidisciplinary field of Human Factors [10]. Instead of simplifying tasks, automation can critically change the demands placed on human operators by reducing active application of skills, leading to more passive tasks such as monitoring [8]. Higher automation levels can diminish valuable hands-on expertise due to limited engagement with manual tasks, which can in turn lead to deterioration of skills required for problem-solving and quick decision-making in high-stakes situations [2,11]. This issue becomes more pronounced with the rise of STARA, where technology interacts interdependently with individual workers or teams–in industry as well as in health care [12]. Subsequently, the need for designing automated work environments which balance technological support with opportunities for skill application and development becomes especially important, as system resilience relies heavily on the operator's situational awareness (SA) adaptability and skill application [13]. In this study, RAS serves as an empirical example for the shift in professional skills due to the introduction of a STARA-technology. Its introduction into surgical practice redefines work and task processes, disrupts established workflows, and necessitates the redistribution of primary and secondary tasks between technology and professional staff [11,14]. But does technology such as RAS adequately represent the broader concept of STARA? In this study, we use the term STARA to refer to the broader category of smart technologies, AI, robotics, and algorithms that are not just automations of manual work but are characterized by different Levels of Autonomy (LoA) [11].

Level of Autonomy in STARA Environments

RAS represents a specific type of STARA, as it mediates human work through a semiautonomous technological system embedded in a complex, multiteam environment [15]. Current generations of da Vinci systems are positioned at lower LoA (eg, level 1) [16,17], functioning in a master-slave configuration and therefore unable to operate autonomously. Nevertheless, underlying AI and algorithms have supportive functions in enabling the system to execute essential secondary tasks, such as image stabilization, motion scaling [15], and augmented reality overlays integrating preoperative imaging data into the live surgical view [18-20]. Following Endsley [11], current RAS systems would be positioned at a relatively low level of executive task support. Determining the level of autonomy at the task level is crucial for comparing the effects of different STARA technologies in health care. Existing frameworks provide useful distinctions between different task types (eg, monitoring, decision-making, and execution) and their respective levels of autonomy, thereby enabling a more nuanced assessment of technological capabilities and human-autonomy interaction [15,21].

The introduction of RAS technology in particular, and STARA technologies more generally, alters the primary tasks of nurses; however, changes in subsequent task characteristics such as variety are driven not only by the LoA of the technology itself but also by the technology-independent design and allocation of secondary tasks. In the present study, surgical teams applied the same RAS system differently, which led to varying task demands for the nurses. The following sections examine how technology-induced changes in task-specific demands affect skill application and broader career-related outcomes in health care settings.

Task-Specific Demands

A central focus of Human-Factor-oriented work design is the work-task itself, with one of its core constructs being task variety. Task variety is defined as the number of distinct tasks an operator must perform within a given job [22]. Research has shown that higher task variety is associated with subjective performance, job satisfaction due to more stimulating and interesting tasks, reduced monotony, increased mental alertness, and greater engagement [8,23]. In STARA environments, technology can take on tasks which were previously performed by humans and thereby reduce task variety [1]. These STARA-related task changes will influence another core construct, namely perceived workload. More specifically, during a RAS with reduced task variety, theater nurses perceive less workload than during non-RAS procedures [24,25]. In practice, however, whether a decrease, invariance, or even increase of task variety and workload will occur after the introduction of RAS depends on unit-specific work design and (secondary and primary) task allocations to the theater nurses [25]. If tasks are delegated to technology without compensatory task design, this subsequently results in decreased involvement of nurses and relates to higher monotony or lower application of professional skills. The latter phenomenon represents the key difference between surgical teams in this study, which was investigated with respect to the effects on professional skill application.

Technical and Nontechnical Professional Skills and System Resilience

In this paper, we use the umbrella term of professional skills and differentiate technical and nontechnical skills [3]. Technical skills often refer to trained knowledge and behaviors required to perform specific tasks. They are present across situations, acquired through training, and form an essential element of an individual's professional competencies [26]. Additionally, Human Factors also considers nontechnical skills, which refer to more generic behavioral, emotional, or cognitive activities shown in a specific situation [27]. Nontechnical skills can further be divided into two subgroups: cognitive (eg, SA) and social (eg, teamwork) nontechnical skills [27].

Nontechnical cognitive skills are significant predictors of safety and resilience in socio-technical work systems [2,3]. In particular, the nontechnical skill of SA is defined as a 3-level hierarchical construct. It consists of the perception of information and stimuli from the environment (I Perceiving),

XSL•FO
RenderX

the understanding of the meaning of this information and stimuli (II Understanding), and the anticipation of their state in the future (III Projecting) [28,29]. With a high level of SA, any changes in a situation can be noticed and, for example, communicated to the team [30], which highlights its importance for team resilience. Regarding the health care sector, maintaining SA or its levels can prevent errors in the work process, for example, by recognizing changes in vital signs, thus ensuring patient safety [29,31]. SA is, therefore, primarily intended to prevent critical situations from arising or escalating [32] and is of utmost importance in maintaining resilient systems [2].

From a Human Factors perspective, the effects of unit-specific task design and its effects during surgeries were of interest. We evaluated how task demands (ie, objective task variety, subjective workload, and perceived monotony) during RAS surgeries influence nurses' application of professional skills (Question 1).

Sustainable Careers: Task Demands, Professional Skills, and Needs

The case of RAS underpins the need for a trans- and interdisciplinary approach: at the task-level, it focuses on Human-Factors questions about optimal and resilient work design. At the career level, it raises pivotal issues of how technologically driven change shapes long-term professional identities and career evaluations. Whilst the former sections examine these interdependencies by analyzing task-specific demands and skill application, the following section focuses on the broader consequences for career sustainability in STARA-mediated health care settings.

Concepts and Models of Sustainable Careers

Most scholars define sustainable careers as "the sequence of an individual's different career experiences, reflected through a variety of patterns of continuity over time, crossing several social spaces, and characterized by individual agency, herewith providing meaning to the individual" [33]. More specifically, the theoretical model of De Vos et al [7] provides three key indicators that distinguish sustainable from unsustainable careers, namely health, happiness, and productivity. Crucial for perceived sustainable careers is a dynamic person-career fit over time, described as the ability to adapt and evolve with changing needs and environmental demands [34]. Chin et al [5] differentiate four dimensions (resourceful, flexible, renewable, and integrative) to capture the dynamic nature of career sustainability in fast-paced environments. The dimension of "renewability" highlights task redesign due to STARA, requiring employees to reorient their role perceptions, knowledge, and skills which also applies to the example of RAS as explained below.

Professional Skills and Task-Specific Design

On a task-level, "renewability" implies that workers evaluate the opportunities to apply, update, or learn role-related professional skills in the light of STARA task-change. In our case study, da Vinci robots fundamentally change the way theater nurses can participate in the workflow. A perceived loss of role-identity creating activities on task-level should therefore influence their evaluations on the career level. If the nurses are hindered in applying their professional skills during the surgery and anticipate this particular task design as a future scenario, they might experience a sense of role incongruence and frustration, leading to the anticipation of a less sustainable career.

However, Chin et al [5] do not elaborate or define specific task-level variables in their model on sustainable careers. Other literature on Career Research has differentiated between tasks within a job [6] or followed a task-based approach to investigate the effects of automation on skill requirements [35]. However, despite the detailed listing of tasks there is a lack of linkage, for example through integration into subsequent processes and analysis of preliminary and secondary tasks. Furthermore, the measurement of skills or competencies tends to be on a much more generic level in studies of Career Research [36,37]. Human Factors, on the other hand, investigates perceived application of professional skills on a "fine-grained" task-level of analysis in specific contexts, which could add to the understanding of career sustainability evaluation and model development.

Complementing the antecedents of sustainable career development in the model by Chin et al [5], a task-level perspective (task variety, workload, or complexity) may contribute to the individual and organizational level of the present theory. To our knowledge, no studies have been conducted on this cross-level mechanism between variations in the task-specific design of a use case and long-term evaluations of career sustainability. Thus, we explore in the case study how task-specific demands and perceived skill application align with reflections on career sustainability (Question 2a).

Career-Related Needs and Situational Perception of Workload and Skill Application

Several researchers elaborate on individual needs or motivation with respect to career evaluation. For example, Chin et al [5] point out that the needs of workers keep changing over their life courses, leading to interindividual differences in desires for sustainable careers. De Vos et al [7] indicate that the fulfillment of personal needs contributes to perceived career satisfaction or success. Especially, the sustainability indicator of happiness relates to how well the career aligns with one's values and needs, for example, the need for personal growth [7,38]. The opportunity to develop and apply new skills contributes to perceived growth and is also considered a facet of subjective career success [39]. In contrast, frustration through a mismatch between personal career needs and developmental opportunities has been linked to higher turnover intentions and job dissatisfaction [40]. However, career-related needs have not yet been connected to evaluations of a specific STARA-related task change. For example, workers with a high need for personal growth would evaluate tasks more critically if STARA hinders the application of skills rather than fostering development. On the other hand, workers with low growth needs would evaluate the objectively similar task less critically and therefore differ from their colleagues within the same unit. Thus, we explore in the case study how career-related needs correspond to situational evaluations of workload and skill application (Question 2b).

XSL•FO
RenderX

Study Contribution

The contribution of this paper is threefold. Firstly, the transdisciplinary multilevel perspective (person, task, and organization) enriches conceptual frameworks for both STARA-related Career Research and Human Factors. Career Research may benefit, for example, if taxonomies or task-level concepts of STARA design from Human Factors are integrated into career-related models to allow differentiated predictions on job threats and career decisions. Conversely, Human Factors will benefit from Career Research if task-designers recognize effects of interindividual differences in career-related variables on the evaluation of STARA design.

Secondly, the paper elaborates on theoretical and methodological perspectives for interdisciplinary STARA research. For example, the variance of task variety, workload, or skill-application between two embedded units raises questions of the appropriate levels of analysis for research models predicting career evaluations.

Thirdly, the paper contributes to applied questions of career development and Human Factors design on multiple levels. Organizations in health care and other STARA-affected fields (macro-level) are urgently looking for highly committed professionals (microlevel) for jobs in a drastically changed task-environment (meso-level). Thus, workplace interventions addressing task design and the consideration of interindividual differences (eg, growth needs) will contribute to retaining specialists in key professions with STARA work tasks that promote both the desired efficiency gains in productivity and personal development goals.

Methods

Background and Setting of the Case Study

In the project underpinning the case study, da Vinci assisted surgery was applied in two units (urological and visceral surgeries) of a hospital. Despite the similar surgical context, the number of tasks (task variety) for theater nurses differed between operating units due to unit-specific task design. Thus, this setting can be classified as embedded (multiple units) single-case design [41] and is illustrated through qualitative and quantitative data.

The underlying project had an application-oriented Human Factors character and aimed to explore qualitatively and quantitatively the effects of different task-specific designs in surgical workflows. A qualitative semistructured focus group interview (n=5) was conducted prior to the quantitative data collection (n=36). The aim of the interviews was to gain a qualitative understanding of the tasks and challenges during RAS and to explain the planned quantitative data collection in the units. Extending beyond a Human Factors focus, career-related experiences were also reported in these interviews. Both perspectives, Human Factors and Career, were subsequently examined empirically and exploratorily in the quantitative phase. This quantitative part of the case study was not designed as a hypothesis-testing study but provides insightful transdisciplinary perspectives.

Methods

Part One: Semistructured Focus Group Interview

The qualitative interview data of the case study derived from a semistructured focus group [42] involving five nurses experienced in the da Vinci X Surgery System (Intuitive Surgical Inc). All had a background in urology, with four also having experience in visceral surgery. The group included two operating department practitioners and three theater nurses. The focus group took place in a room adjacent to the operating theater of the hospital featured in this case study. The Human Factors question (Question 1) prompted was “What changes does the da Vinci robot cause in the different phases of surgery compared to laparoscopic surgery?” During the focus group interview, various aspects were discussed, such as changes in team collaboration (processes), changes in one’s own work and role (tasks, requirements, and skills), occurrence and management of problems, disruptions and critical situations, as well as general concerns, wishes, or ideas for the future design of RAS. Statements from the in-depth focus group interview were documented through field notes and assessed using thematic analysis [43] but without an elaborate coding strategy. Results are reported in a thematic structure later in this paper to illustrate experiences of nurses in terms of Human Factors and career-related perspectives in line with the research questions formulated above. Career-related aspects were not directly elicited but were reported by participants as outcomes of task-related work design and were further explored in the quantitative part.

Part Two: Postsurgery Surveys

The quantitative data of the case study was collected over a period of 12 weeks from June 2024 to September 2024. Theater nurses (including four participants from the interview) were surveyed following each RAS. They responded to task-specific items related to the surgery execution phase and to overarching scales on career sustainability and developmental needs. Variables and measures are displayed in Table 1. Approximately 28 RAS took place during this period, with a total of 44 returned questionnaires after surgery (79% response rate). Four data sets were excluded due to missing data. Four data sets were excluded because task variety was not quantifiable due to role switch or fluctuation during a procedure, resulting in n=36 data sets. Because some nurses had two or more da Vinci surgeries over the course of the study, seven participants completed the survey three times or more, and five participants completed it twice. In the current exploratory correlational analysis, the partially dependent data was treated as independent data points.

XSL•FO
RenderX

Table . Variables and measures of quantitative survey[a].

Variable	Measures
Task variety	Number of specific tasks assigned to theater nurses during the surgery execution phase and was coded by authors following Rau et al [44].
Subjective workload	Three items from NASA-TLX[b] [45] representing mental load, temporal load, and effort on a 15-point rating scale.
Subjective monotony	Two context-specific items on a 6-point rating scale, relating to underchallenge and boredom in DIN EN ISO[c] 10075 - 1 [46].
Perceived application of nontechnical cognitive skills (SA[d] I-III)	Perceived application of nontechnical cognitive skills, measured by behaviorally anchored items adopted from established behavioral observation scales by Schreyer et al [30]. SA I: perceiving, SA II: understanding, SA III: projecting.
Perceived application of technical skills	Single-item, "I was able to apply my skills and abilities to the surgical process" on a 6-point rating scale.
Perceived career sustainability	Participants were asked to reflect: "Imagine that the surgery evaluated here would be a typical day at work for future theatre nurses. In light of this, how do you rate the following career related statements?." Following De Vos et al [7] 3 indicators of sustainable careers were reflected (satisfaction, productivity, and health) using single items.
Growth need	Six items from the JDS[e] originally developed by Hackman and Oldham [47].

[a]All items are accessible in Multimedia Appendix 1.

[b]NASA-TLX: NASA Task Load Index.

[c]DIN EN ISO: Deutsches Institut für Normung Europäische Norm International Organization for Standardization.

[d]SA: situational awareness.

[e]JDS: Job Diagnostic Survey.

Ethical Considerations

The study adheres to ethical and legal requirements. Participation was voluntary, and informed consent was collected anonymously prior to participation. Participants received no form of compensation for their partaking in both parts of the study. Both parts of the study received prior ethical approval from the hospital's Clinical Ethics Committee (CEC) and the Staff Representative Council. The study complies with the General Data Protection Regulation (GDPR); participants' privacy and confidentiality of their data were maintained.

Results

Part One: Qualitative Data From the Focus Group Interview

Question 1

Nurses described varied experiences with workload, monotony, and skill application throughout RAS. During the phase of surgery execution (primary task), some reported a notably lower workload, largely due to the assistant surgeon taking over instrument handling. This reduction in engagement led to feelings of monotony and even boredom among theater nurses, who described this phase as repetitive. The reduction of involvement through instrument handling was associated with decreased SA. This can be seen as problematic since in RAS, nontechnical skills such as quick decision-making, cooperation, and quick adaptation to technical difficulties were noted as crucial. One aspect from the interview should be emphasized here and is reflected in the first implication at the end of this paper. It concerns the different evaluation of preparatory and concluding tasks (secondary tasks) in relation to the actual core task of executing the surgery. The monotony reported during surgery execution contrasted with heightened demands in surgery preparation and surgery conclusion. In these phases, staff often faced time pressures, especially in coordinating with anesthesia teams, whose shorter patient preparation times for RAS increase the demand on theater nurses. In response, nurses expressed a desire for additional support personnel. The lack of such support frequently led to overload, especially after undocking the robotic system, when multiple tasks needed immediate attention.

Question 2

Nurses also expressed complex views on how their involvement with RAS influences their job and career projections. The initial enthusiasm for working with advanced technology was reported to fade over time, with some indicating frustration due to the monotony during surgery execution. Additionally, although there was no concern over potential job replacement by the robotic system, questions were raised about future staffing needs: for example, whether two surgical nurses would remain necessary per RAS procedure. The lack of substantial, fulfilling tasks in RAS has led to a sense of diminished responsibility among some staff, with a desire for roles that offer more significant nontechnical skill applications and career development opportunities. While physical demands were not seen as problematic, some participants reported experiencing psychological stress due to workload peaks in phases of surgery

XSL•FO
RenderX

preparation and surgery conclusion. The potential for increased productivity was noted, especially if procedural bottlenecks, such as moving and preparing the robotic system, were addressed. Some staff members mentioned actively trying to reduce their involvement in RAS by transitioning to hospitals with lower frequencies in order to gain enough variety in their schedules to prevent RAS routine monotony.

Part Two: Quantitative Data From Postsurgery Surveys

Table 2 displays the nonparametric correlations (Spearman ρ) between the variables.

XSL•FO
RenderX

Table . Correlations (Spearman rho) among and descriptive statistics of case study variables[a].

Variables	Task variety	Perceived workload	Perceived monotony	Nontech SA[b] I	Nontech SA II	Nontech SA III	Nontech skill mean	Technical skill	Satisfaction	Productivity	Health	Growth need
Task level												
Task variety[c]												
ρ	1	0.13	–0.40[d]	0.43[d]	0.39[d]	–0.01	0.46[e]	0.39[d]	0.04	0.27	0.37[d]	–0.01
P value	—[f]	.44	.02	.008	.02	.98	.005	.02	.81	.12	.03	.95
Perceived workload[g]												
ρ	0.13	1	–0.50[e]	0.60[h]	0.21	–0.26	0.40[d]	0.08	0.13	0.17	0.47[e]	–0.48[e]
P value	.44	—	.002	<.001	.22	.13	.02	.65	.44	.32	.004	.003
Perceived monotony[i]												
ρ	–0.40[d]	–0.50[e]	1	–0.75[h]	–0.26	–0.07	–0.53[h]	–0.27	–0.08	–0.28	–0.4[d]	0.31
P value	.02	.002	—	<.001	.13	.70	<.001	.11	.64	.10	.01	.06
Nontech SA I[j]												
ρ	0.43[d]	0.60[h]	–0.75[h]	1	0.38[d]	0.09	0.75[h]	0.44[e]	0.10	0.26	0.50[e]	–0.34[d]
P value	.008	<.001	<.001	—	.02	.60	<.001	.007	.58	.13	.002	.04
Nontech SA II[k]												
ρ	0.39[d]	0.21	–0.26	0.38[d]	1	0.39[d]	0.85[h]	0.60[h]	0.13	0.15	0.42[d]	–0.48[e]
P value	.02	.22	.13	.02	—	.02	<.001	<.001	.45	.39	.01	.003
Nontech SA III[l]												
ρ	–0.01	–0.26	–0.07	0.09	0.39[d]	1	0.41[d]	0.59[h]	0.23	0.31	0.05	–0.09
P value	.98	.13	.70	.60	.02	—	.01	<.001	.17	.07	.77	.61
Nontech skill mean[m]												
ρ	0.46[e]	0.40[d]	–0.53[d]	0.75[h]	0.85[h]	0.41[d]	1	0.69[d]	0.15	0.28	0.57[h]	–0.48[e]
P value	.005	.02	<.001	<.001	<.001	.01	—	<.001	.39	.10	<.001	.003
Technical skill[n]												
ρ	0.39[d]	0.08	–0.27	0.44[e]	0.60[h]	0.59[h]	0.69[h]	1	0.28	0.33[d]	0.30	–0.30
P value	.02	.65	.11	.007	<.001	<.001	<.001	—	.10	.048	.08	.07
Career level												
Satisfaction[o]												
ρ	0.04	0.13	–0.08	0.10	0.13	0.23	0.15	0.28	1	.83[h]	0.50[e]	–0.30
P value	.81	.44	.64	.58	.45	.17	.39	.10	—	<.001	.002	.08
Productivity[p]												
ρ	0.27	0.17	–0.28	0.26	0.15	0.31	0.28	0.33[d]	.83[h]	1	0.59[h]	–0.18

XSL•FO
RenderX

Variables	Task variety	Perceived workload	Perceived monotony	Nontech SA[b] I	Nontech SA II	Nontech SA III	Nontech skill mean	Technical skill	Satisfaction	Productivity	Health	Growth need
P value	.12	.32	.10	.13	.39	.07	.10	.048	<.001	—	<.001	.29
Health[q]												
ρ	0.37[d]	0.47[e]	–0.4[d]	0.50[e]	0.42[d]	0.05	0.57[h]	0.30	0.50[e]	0.59[h]	1	–0.42[d]
P value	.03	.004	.01	.002	.01	.77	<.001	.08	.002	<.001	—	.01
Growth need[r]												
ρ	–0.01	–0.48[e]	0.31	–0.34[d]	–0.48[e]	–0.09	–0.48[e]	–0.30	–0.30	–0.18	–0.42[d]	1
P value	.95	.003	.06	.04	.003	.61	.003	.07	.08	.29	.01	—

[a]n=36, variables: "Nontech" refers to perceived application of nontechnical cognitive skills during a surgery, "Technical skill" refers to perceived application of technical skills during a surgery, "Satisfaction", "Productivity", and "Health" correspond to perceived career sustainability.

[b]SA: Situational Awareness.

[c]Mean (SD): 3.2 (0.66).

[d]*P*<.05.

[e]*P*<.01.

[f]Not applicable.

[g]Mean (SD): 5.24 (2.77).

[h]*P*<.001.

[i]Mean (SD): 3.59 (1.82).

[j]Mean (SD): 4.43 (0.93).

[k]Mean (SD): 4.92 (0.97).

[l]Mean (SD): 4.81 (1.01).

[m]Mean (SD): 4.69 (0.69).

[n]Mean (SD): 3.31 (1.91).

[o]Mean (SD): 3.77 (1.07).

[p]Mean (SD): 3.72 (1.06).

[q]Mean (SD): 3.00 (1.22).

[r]Mean (SD): 4.70 (1.07).

Question 1

Correlates indicate a relationship between objective task variety of theater nurses and perceived application of professional skills. The more tasks involved, the lower perceived monotony (ρ=−0.40; *P*=.02), and the higher perceived application of nontechnical (ρ=0.46; *P*=.005) as well as technical skills (ρ=0.39; *P*=.02). A similar pattern emerges for subjectively perceived workload during the surgical task.

Question 2

Results indicate that higher objective task variety and perceived workload during a surgical task are related to more positive evaluation of health-related career prospects (for task variety: ρ=0.37; *P*=.03 and for workload: ρ=0.47; *P*=.004). Similarly, the higher the perceived application of nontechnical and professional skills during a surgery the more positive reported health and productivity related career prospects appear. Moreover, results show a relationship between individual growth need and the individual evaluation of task characteristics and perceived skill application. Participants with high growth need report lower workload during surgery (ρ=−0.48; *P*=.003) and less perceived application of nontechnical (ρ=−0.48; *P*=.003) and technical skills (ρ=−0.30; *P*=.07).

Discussion

This study investigated how task-specific demands shape theater nurses' perceived professional skill application during RAS (Question 1), and how these task-level changes relate to broader evaluations of career sustainability and vice versa (Question 2).

Qualitative results showed that while the surgical execution phase reduced workload, it also heightened monotony and lowered situation awareness, whilst pre- and postoperative preparations led to overload due to time pressure and insufficient staffing. The quantitative data confirmed that in the execution phase higher task variety correlated with higher application of nontechnical and technical skills (Question 1). Moreover, task design and skill application during RAS were linked to nurses' career reflections in the qualitative data. The survey showed higher task variety and skill application during surgeries were correlated with long-term career sustainability evaluations. Moreover, career-related growth needs correlate with

XSL•FO
RenderX

interindividual differences in STARA task evaluations (Question 2).

Limitations of the Case Study

Firstly, it is important to note that the presented correlational data can only be interpreted to a limited extent because of data independence, context-specific unvalidated scales, and small sample size. Repeated observations from the same nurses were included deliberately, as task variety differs situationally across procedures, capturing meaningful within-person variation–, though ensuring fully distinct evaluations remains challenging. Spearman ρ was chosen as it evaluates monotonic relationships regardless of distributional properties [48] with more robust type I error control than Pearson *r* [49,50]. Furthermore, besides task variety and career-related needs, additional contextual factors (eg, duration of the surgery) may affect perceptions of monotony and skill application. However, for exploratory reasons, the data allows for cautious reflections regarding task-specific work design in the field of Human Factors (Question 1). A decrease in task variety during RAS correlates with monotony and reduced application of professional skills. As task variety fluctuates between units, it is the responsibility of surgical team leaders to create conditions which increase task variety through more possibilities to apply nontechnical and technical skills. Task designs with higher applications of nontechnical cognitive skills enable faster and more resilient responses in case of disruptions. Moreover, this contributes to the maintenance and development of individual skills as a requirement of humane work design [8,51].

Examining the relationships between task-level and career evaluations in RAS provides interdisciplinary insights (Question 2). It is evident that both the specific task design of RAS (eg, task variety) and the associated application of professional skills are related to individual evaluations of career prospects. Furthermore, interindividual differences must be considered, as variance in personal growth need covaries with different evaluations of task characteristics and career prospects [1].

The case study gives practical insights with high ecological validity but is empirically limited due to a small sample size, situationally changing context conditions, or changing team compositions. For example, the small number of nurses working with da Vinci in a hospital are surveyed and observed on multiple measurement points. However, depending on the work plan or the motivation to participate, the within measurements vary greatly and range from only one instance of participation to up to 4 sessions per person. This results in a small between sample with varying within measurement points. Moreover, this data is confounded through the complexity and dynamic of the work setting, such as personnel changes during a surgery or “outlier” situations which need to be coded and removed from the statistical evaluation. For this reason, STARA’s interdisciplinary Career Research requires quantitative multilevel methods that are robust against the violation of prerequisites, as well as a broad mix of qualitative and multivariate quantitative methods.

Theoretical, Practical, and Methodological Implications

The following section outlines the implications for Human Factors and interdisciplinary perspectives. These implications extend beyond the specific application of RAS and are discussed more broadly under the framework of STARA.

Human Factors’ Perspectives to Support Skill Application in STARA Environments

From a Human Factors’ perspective, the study’s findings emphasize that skill application in RAS is not determined by technology implementation alone, but rather by how surgical workflows are specifically designed and managed within different units of a hospital. Distinct task configurations across RAS procedures shape task-specific demands and are directly related to theater nurses’ perceived application of their professional skills. This highlights that leadership-driven work design and task allocation decisions, rather than STARA itself, play a critical role in whether skills are maintained, reduced, or expanded. Even within the same technological environment, variations between organizational units can create differences in opportunities for professional skill application. The consequences affect both broader system resilience and individual career projections.

First, professional skill application is closely tied to resilience, particularly through the maintenance of SA. In this case study, preserving nontechnical skills is not limited to perceiving and understanding or monitoring procedural steps (SA levels I and II), but also includes the critical ability to anticipate future disruptions or complications (SA level III). This anticipatory skill is essential for adaptive responses during unexpected events and directly contributes to resilience and safety of work systems. Examples are grounded in both health care and sociotechnical team research. In health care, projecting future states is critical for the early identification of complications and safety threats, thereby strengthening patient safety [52]. Furthermore, within sociotechnical team research, SA extends beyond immediate procedural coordination to encompass risk awareness regarding technological vulnerabilities, system limitations, and potential disruptions introduced by technologies themselves. This broader perspective is particularly relevant in STARA environments, where human actors must continuously perceive, understand, and project technology-related risks across a wide range of tasks [53].

Secondly, ongoing automation of tasks initiates a decrease in middle skilled jobs, such as theater nurses, leading to the individual use of career transition strategies [4]. This very effect became evident in the qualitative part of the case study when nurses mentioned transitioning from high-volume RAS sites to hospitals with lower frequencies. To avoid these negative STARA effects, human-centered design principles should be explicitly considered during the implementation of technologies. For example, Parker and Grote [8] underline the importance of task-variety, as well as human autonomy, skill-use, social interaction, or feedback. These factors can contribute to shaping the workplace and, thus indirectly, the careers of workers in STARA contexts in a way that is sustainable, healthy, productive, and beneficial to personal development. For example, lower task-variety in one subtask may be compensated

XSL•FO
RenderX

with additional primary or secondary tasks so that employees remain involved and can develop personal competencies. In the present case study, lower task-variety during surgery could be complemented by implementing mixed reality technologies in RAS to inform and involve nurses during inactive periods [54,55]. A prospective strategy would focus not only on skill training, but on fostering proactive work as well as task design. Providing competencies and resources for joint optimization of changing workplaces can foster the adaptation to STARA-induced changes [56]. Employees who perceive STARA as an opportunity, rather than a threat, are more likely to engage in job crafting [57], which may also support sustainable careers [5]. Designing human-centered STARA environments therefore requires preserving meaningful opportunities for both technical and nontechnical skill application to promote long-term career sustainability.

Consequently, task design should be understood as a strategic lever for fostering both sustainable careers and system resilience.

Interindividual Differences in Skill Application and Career-Related Needs Are Relevant for Evaluating STARA Work Systems

Human Factors as well as Career Research can benefit from a differential approach focusing on interindividual differences.

Human Factors Research should also consider that there will be interindividual variations in the evaluation of task-characteristics within a given STARA work environment. Evaluations of STARA work systems go beyond task characteristics (see Question 2b). In our case study, we found evidence that interindividual differences in individual growth needs correlate with subjective task evaluations (but not with the objective task variety). Nurses with high growth needs seemed to perceive generally lower workload and skill application than nurses with lower needs. Theoretically, the importance of individual motives on the assessment of work systems is not completely new (Hackman and Oldham [58] for work satisfaction, motivation, and turnover). STARA modifies task characteristics, but individual motives and interindividual differences in task evaluation are often neglected in Human Factors. Numerous motives and personality traits related to STARA work environments can be identified in organizational and Career Research, which may affect the evaluation of the system design in a specific work context (eg, need for autonomy or personal growth [38]). In this context, an interdisciplinary career perspective offers numerous starting points for Human Factors Research and theory development. Differential work design can be used to consider varying development needs based on these findings of interindividual differences in motives and traits.

For Career Research it is important to acknowledge that the perception of skill application in a particular task is highly individual. Beyond that, the interindividual perceptions also relate to the evaluation of indicators for sustainable careers [7]. Question 2a of our case study focused on perceived application of technical and nontechnical skills relevant to system resilience [3] and its link to career reflections. Nurses reporting higher application of role-related nontechnical skills reflected positively on career health, whereas perceived application of technical skills correlated with career productivity. This link between interindividual different perceptions of task-related skills and career evaluations should be considered in future Career Research as theories mention the complexity of objective environmental conditions and subjective evaluations resulting in a wide variety of individual reflections on whether one's career is sustainable or not [7]. For future investigations, Human Factors offers systematic measures of technical and nontechnical skills [30] which can also be applied in practice for resilient task design, personnel development, and career counseling.

Intraindividual Changes Are a Significant Temporal Variable for Theory and Practice

Theories of career sustainability underline the importance of temporal dynamics and intraindividual variation over time from a methodological and conceptual standpoint [5,7,34].

In the short term, research on career sustainability may investigate the effects of time and quality of individual experiences with STARA technology. In this respect, Human Factors Research provides measures and evidence about user experience [59]. User experience does not only have a quantitative temporal focus (ie, the duration of experience with a STARA technology), but also a qualitative one. For example, quality depends on whether STARA experiences fulfill effort and performance expectations [12]. Thus, time and quality of user experience may add technology-related variables to the individual level of models on career sustainability, stimulating future research on STARA implementation.

In the long term, attitudes towards STARA may vary throughout an individual's lifetime. As shown in gerontological research [60], goals related to personal growth and new experiences decrease with age. This variable is related to growth needs and was negatively correlated with perceived career sustainability in our study. Future research may investigate how shifts in personal goals or other motives impact career evaluations, especially in STARA environments.

Human Factors Task Characteristics Expand Models of Career Sustainability

STARA includes an overwhelming variety of technical applications with different implications for workers' career sustainability across diverse contexts. For example, Chin et al [5] focus on manufacturing workers facing full automation of the entire work task correlated with job insecurity and fear of job losses. In contrast, in our case study of RAS, the automation affected one specific subtask regarding SA. The two examples differ essentially in the extent of STARA-induced changes and subsequent career assessments but become comparable on task-level characteristics. Present conceptual frameworks of career sustainability [5] could benefit from expanding the individual and organizational levels to a task-level with different variables such as LoA, task type, or task variety.

First, LoA [61] indicates the degree of autonomy with which STARA technology performs tasks independently from human control. Low degrees of technical autonomy, such as in our RAS example, can have very different effects on career sustainability compared to high degrees of automation, where

XSL•FO
RenderX

machines generate and implement decisions without human involvement. Second, automation through STARA may affect different task types like monitoring, planning, decision-making, or execution [11]. In one context, automation can be related to an executing task in manufacturing [5]. In another context, STARA might automate a monitoring task, leaving decision and execution tasks to humans with different consequences for career evaluations. Third, task-variety can be used to investigate STARA-induced changes between or within contexts. As shown in the case study, nurses' evaluations of career sustainability can even differ when working in the same context with the same technology but in different surgical work units. Nurses with higher objective task variety seemed to reflect generally more positively about anticipated health, one indicator of sustainable careers. Furthermore, task-level analysis also involves the composition of primary and secondary tasks. In the case study, we focused on the variety of tasks nurses performed during the primary task of surgery execution. In addition to this core task, there are preliminary and concluding secondary tasks that might shift in their task characteristics due to STARA and thus influence career evaluations.

The task-level variables mentioned could be further supplemented, for example, by STARA-relevant team-level variables such as team cohesion or team trust [12]. As has been requested in previous studies on sustainable careers [5], task and team-level variables from Human Factors could enrich the framework as relevant predictors, mediators, or moderators, and inspire future studies.

Interdisciplinary Perspectives on STARA Require Conceptual and Methodological Clarity

The former 4 implications display valuable synergies of interdisciplinary research by integrating person, task, and technology characteristics into frameworks of sustainable career as well as Human Factors research. This added value was also highlighted in the case study, which will now be supplemented by conceptual and methodological desiderata for future research.

When interdisciplinary concepts are integrated into models, it must be ensured that variables with the same name do not represent different content or operationalization. In our case study, the perceived application of technical and nontechnical skills was a variable connecting Human Factors with Career Research. Both disciplines have an extensive body of research on skill-related concepts, measures, and models (eg, recruitment diagnostic, and lifelong learning in Career Research; skills for system resilience, and problem solving in Human Factors). However, we were faced with the challenge that even within the discipline of Human Factors, researchers applied terminologies and definitions heterogeneously, which calls urgently for conceptual clarifications. For example, some authors refer to nontechnical skills as skills [27] some as competencies [62]. This issue, referred to as Jingle-Jangle-Fallacy, hinders STARA-related research, because interdisciplinarity increases language diversity [63].

Besides conceptual clarity, our study underlines the careful distinction between objective and subjective evaluations on a measurement level. For example, it is necessary to carefully distinguish between the objective expression of professional skills in a STARA context and the subjective self-evaluations as conceptualized in this study. Consequently, for fruitful interdisciplinary research, it is crucial that concepts, definitions, and operationalization are clearly described and that future work identifies existing inconsistencies (for methods see [63]).

Conclusion

This paper highlights the importance of conceptually and empirically linking task characteristics to career sustainability in health care-related STARA work contexts and beyond. It is the way new technologies are implemented into existing work processes, and not technology itself, that changes task characteristics. This, in turn, affects both short- and long-term employee career experiences in the health-care sector. Research on sustainable careers in particular can benefit from the expansion of existing theories to include highly task- and context-specific variables that can explain inter- and intraindividual differences in career evaluation. Thus, cross-disciplinary insights can and should enrich Career Research's theoretical models, empirical investigations, and practical interventions with constructs and measures.

Acknowledgments

We thank all participants for taking part in this study. The authors would also like to thank the anonymous reviewers and the editor for their constructive comments and valuable suggestions, which substantially improved the quality and clarity of this manuscript.

No generative AI was used in the preparation or revision of this manuscript.

Funding

The publication was funded / supported by the Open Access Fund of Universität Trier.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Supervision: TE.

XSL•FO
RenderX

Project Administration: JB.
Investigation: JTP.
Conceptualization: TE, JB.
Methodology: TE, JB, JTP.
Formal Analysis: TE, JTP, JB.
Writing - Original Draft Preparation: TE, JB, JTP.
Writing - Review & Editing: TE, JB, JTP.
Visualization: TE, JB, JTP.

Conflicts of Interest

JTP declares a part-time employment at the hospital during data collection, which did not influence the research process or its outcomes. The other authors declare no conflicts of interest.

Multimedia Appendix 1
Measures from Part Two of the Case Study: Quantitative Data from Postsurgery Surveys.
[DOCX File, 37 KB - humanfactors_v13i1e87735_app1.docx]

References

1. Randell R, Greenhalgh J, Hindmarsh J, et al. How do team experience and relationships shape new divisions of labour in robot-assisted surgery? A realist investigation. Health (London) 2021 Mar;25(2):250-268. [doi: 10.1177/1363459319874115] [Medline: 31522572]
2. Endsley MR. Supporting human-AI teams:transparency, explainability, and situation awareness. Comput Human Behav 2023 Mar;140:107574. [doi: 10.1016/j.chb.2022.107574]
3. Wachs P, Righi AW, Saurin TA. Identification of non-technical skills from the resilience engineering perspective: a case study of an electricity distributor. Work 2012;41 Suppl 1(S1):3069-3076. [doi: 10.3233/WOR-2012-0565-3069] [Medline: 22317186]
4. Chen Q, Wang JA, Ou R, Sun J, Chang LC. Disruptive technologies and career transition strategies of middle-skilled workers. Career Dev Int 2019 Sep 19;24(5):475-490. [doi: 10.1108/CDI-06-2018-0172]
5. Chin T, Li G, Jiao H, Addo F, Jawahar IM. Career sustainability during manufacturing innovation: a review, a conceptual framework and future research agenda. Career Dev Int 2019;24(6):509-528. [doi: 10.1108/CDI-02-2019-0034]
6. Hirschi A. The fourth industrial revolution: issues and implications for career research and practice. Career Dev Q 2018 Sep;66(3):192-204. [doi: 10.1002/cdq.12142]
7. De Vos A, Van der Heijden B, Akkermans J. Sustainable careers: towards a conceptual model. J Vocat Behav 2020 Mar;117:103196. [doi: 10.1016/j.jvb.2018.06.011]
8. Parker SK, Grote G. Automation, algorithms, and beyond: why work design matters more than ever in a digital world. Appl Psychol 2022 Oct;71(4):1171-1204. [doi: 10.1111/apps.12241]
9. Tegtmeier P, Weber C, Sommer S, Tisch A, Wischniewski S. Criteria and guidelines for human-centered work design in a digitally transformed world of work: findings from a formal consensus process. Int J Environ Res Public Health 2022 Nov 23;19(23):23. [doi: 10.3390/ijerph192315506] [Medline: 36497580]
10. Bainbridge L. Ironies of automation. In: Analysis, Design and Evaluation of Man–Machine Systems: Pergamon; 1983:129-135. [doi: 10.1016/B978-0-08-029348-6.50026-9]
11. Endsley MR. From here to autonomy. Hum Factors 2017 Feb;59(1):5-27. [doi: 10.1177/0018720816681350] [Medline: 28146676]
12. Ellwart T, Schauffel N. Human-autonomy teaming in ship inspection: psychological perspectives on the collaboration between humans and self-governing systems. In: Johansson TM, Dalaklis D, Fernández JE, Pastra A, Lennan M, editors. Smart Ports and Robotic Systems: Springer International Publishing; 2023:343-362. [doi: 10.1007/978-3-031-25296-9_18]
13. Romero D, Stahre J. Towards the resilient operator 5.0: the future of work in smart resilient manufacturing systems. Procedia CIRP 2021;104:1089-1094. [doi: 10.1016/j.procir.2021.11.183]
14. Catchpole K, Cohen T, Alfred M, et al. Human factors integration in robotic surgery. Hum Factors 2024 Mar;66(3):683-700. [doi: 10.1177/00187208211068946] [Medline: 35253508]
15. O'Neill T, McNeese N, Barron A, Schelble B. Human-autonomy teaming: a review and analysis of the empirical literature. Hum Factors 2022 Aug;64(5):904-938. [doi: 10.1177/0018720820960865] [Medline: 33092417]
16. Haidegger T, Speidel S, Stoyanov D, Satava RM. Robot-assisted minimally invasive surgery—surgical robotics in the data age. Proc IEEE 2022;110(7):835-846. [doi: 10.1109/JPROC.2022.3180350]
17. van den Berg J, Miller S, Duckworth D, et al. Superhuman performance of surgical tasks by robots using iterative learning from human-guided demonstrations. : IEEE Presented at: 2010 IEEE International Conference on Robotics and Automation; May 3-8, 2010; Anchorage, AK p. 2074-2081. [doi: 10.1109/ROBOT.2010.5509621]

XSL•FO
RenderX

18. Giannone F, Felli E, Cherkaoui Z, Mascagni P, Pessaux P. Augmented reality and image-guided robotic liver surgery. Cancers (Basel) 2021 Dec 14;13(24):6268. [doi: 10.3390/cancers13246268] [Medline: 34944887]
19. Asadizeidabadi A, Hosseini S, Vetshev F, Osminin S, Hosseini S. Comparison of da Vinci 5 with previous versions of da Vinci and Sina: a review. Laparosc Endosc Robot Surg 2024 Jun;7(2):60-65. [doi: 10.1016/j.lers.2024.04.006]
20. Volonté F, Buchs NC, Pugin F, et al. Augmented reality to the rescue of the minimally invasive surgeon. The usefulness of the interposition of stereoscopic images in the Da Vinci™ robotic console. Int J Med Robot 2013 Sep;9(3):e34-e38. [doi: 10.1002/rcs.1471] [Medline: 23239589]
21. Parasuraman R, Sheridan TB, Wickens CD. A model for types and levels of human interaction with automation. IEEE Trans Syst, Man, Cybern A 2000 May;30(3):286-297. [doi: 10.1109/3468.844354]
22. Morgeson FP, Humphrey SE. The Work Design Questionnaire (WDQ): developing and validating a comprehensive measure for assessing job design and the nature of work. J Appl Psychol 2006 Nov;91(6):1321-1339. [doi: 10.1037/0021-9010.91.6.1321] [Medline: 17100487]
23. Humphrey SE, Nahrgang JD, Morgeson FP. Integrating motivational, social, and contextual work design features: a meta-analytic summary and theoretical extension of the work design literature. J Appl Psychol 2007 Sep;92(5):1332-1356. [doi: 10.1037/0021-9010.92.5.1332] [Medline: 17845089]
24. Burtscher MJ, Koch A, Weigl M. Intraoperative teamwork and occupational stress during robot-assisted surgery: an observational study. Appl Ergon 2024 Nov;121:104368. [doi: 10.1016/j.apergo.2024.104368] [Medline: 39146909]
25. Totonchilar S, Aarabi A, Eftekhari N, Mohammadi M. Examining workload variations among different surgical team roles, specialties, and techniques: a multicenter cross-sectional descriptive study. Perioper Med 2024;13(1):1. [doi: 10.1186/s13741-023-00356-6]
26. Nasir ANB, Ali DF, Noordin MKB, Nordin MSB. Technical skills and non-technical skills: predefinition concept. Presented at: International Engineering and Technology Education Conference (IETEC '11); Jan 16-19, 2011 URL: https://www.researchgate.net/publication/259782791_Technical_skills_and_non-technical_skills_predefinition_concept [accessed 2026-08-25]
27. Flin R, O'Connor P. Safety at the Sharp End: A Guide to Non-Technical Skills: CRC Press; 2017. [doi: 10.1201/9781315607467]
28. Endsley MR. Toward a theory of situation awareness in dynamic systems. Hum Factors 1995 Mar;37(1):32-64. [doi: 10.1518/001872095779049543]
29. Green B, Parry D, Oeppen RS, Plint S, Dale T, Brennan PA. Situational awareness - what it means for clinicians, its recognition and importance in patient safety. Oral Dis 2017 Sep;23(6):721-725. [doi: 10.1111/odi.12547] [Medline: 27447437]
30. Schreyer J, Koch A, Herlemann A, et al. RAS-NOTECHS: validity and reliability of a tool for measuring non-technical skills in robotic-assisted surgery settings. Surg Endosc 2022 Mar;36(3):1916-1926. [doi: 10.1007/s00464-021-08474-2] [Medline: 33844085]
31. Brennan PA, Holden C, Shaw G, Morris S, Oeppen RS. Leading article: what can we do to improve individual and team situational awareness to benefit patient safety? Br J Oral Maxillofac Surg 2020 May;58(4):404-408. [doi: 10.1016/j.bjoms.2020.01.030]
32. Stubbings L, Chaboyer W, McMurray A. Nurses' use of situation awareness in decision-making: an integrative review. J Adv Nurs 2012 Jul;68(7):1443-1453. [doi: 10.1111/j.1365-2648.2012.05989.x] [Medline: 22429250]
33. Van der Heijden B, De Vos A. Sustainable careers: introductory chapter. In: Handbook of Research on Sustainable Careers: Edward Elgar Publishing; 2015:1-19. [doi: 10.4337/9781782547037.00006]
34. Schweitzer L, Lyons S, Smith CJ. Career sustainability: framing the past to adapt in the present for a sustainable future. Sustainability 2023;15(15):11800. [doi: 10.3390/su151511800]
35. Autor DH, Levy F, Murnane RJ. The skill content of recent technological change: an empirical exploration. Q J Econ 2003 Nov 1;118(4):1279-1333. [doi: 10.1162/003355303322552801]
36. Akkermans J, Brenninkmeijer V, Huibers M, Blonk RWB. Competencies for the contemporary career: development and preliminary validation of the career competencies questionnaire. J Career Dev 2013;40(3):245-267. [doi: 10.1177/0894845312467501]
37. Hirschi A, Nagy N, Baumeler F, Johnston CS, Spurk D. Assessing key predictors of career success: development and validation of the career resources questionnaire. J Career Assess 2018;26(2):338-358. [doi: 10.1177/1069072717695584]
38. Chin T, Jawahar IM, Li G. Development and validation of a career sustainability scale. J Career Dev 2022 Aug;49(4):769-787. [doi: 10.1177/0894845321993234]
39. Shockley KM, Ureksoy H, Rodopman OB, Poteat LF, Dullaghan TR. Development of a new scale to measure subjective career success: a mixed-methods study. J Organiz Behav 2016 Jan;37(1):128-153. [doi: 10.1002/job.2046]
40. Chen T, Chang P, Yeh C. A study of career needs, career development programs, job satisfaction and the turnover intentions of R&D personnel. Career Dev Int 2004 Jun 1;9(4):424-437. [doi: 10.1108/13620430410544364]
41. Yin RK. Case Study Research and Applications: Design and Methods, 6th edition: Sage; 2018. URL: https://study.sagepub.com/yin6e [accessed 2026-08-25]

XSL•FO
RenderX

42. Krueger RA, Casey MA. Focus Groups.: SAGE Publications Inc; 2014. URL: https://www.lehmanns.de/shop/schulbuch-lexikon-woerterbuch/29679328-9781483365244-focus-groups [accessed 2026-05-06]
43. Ahmed SK, Mohammed RA, Nashwan AJ, et al. Using thematic analysis in qualitative research. Journal of Medicine, Surgery, and Public Health 2025 Aug;6:100198. [doi: 10.1016/j.glmedi.2025.100198]
44. Rau R, Schweden F, Hoppe J, Hacker W. Verfahren zur Tätigkeitsanalyse und -gestaltung bei mentalen Arbeitsanforderungen (TAG-MA): Asanger Verlag; 2021.
45. Hart SG, Staveland LE. Development of NASA-TLX (task load index): results of empirical and theoretical research. Adv Psychol 1988:139-183. [doi: 10.1016/S0166-4115(08)62386-9]
46. Standardization, communication, documentation [Webpage in German]. DIN Media. URL: https://www.dinmedia.de/de/normung-kommunikation-dokumentation/BVFR010 [accessed 2026-05-06]
47. Hackman JR, Oldham GR. Work Redesign: Addison-Wesley Publishing Company; 1980. URL: https://www.scribd.com/document/217232528/Hackman-Oldham-Work-Redesign [accessed 2026-08-25]
48. Hauke J, Kossowski T. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. Quaest Geogr 2011 Jun 1;30(2):87-93. [doi: 10.2478/v10117-011-0021-1]
49. Bishara AJ, Hittner JB. Testing the significance of a correlation with nonnormal data: comparison of Pearson, Spearman, transformation, and resampling approaches. Psychol Methods 2012 Sep;17(3):399-417. [doi: 10.1037/a0028087] [Medline: 22563845]
50. de Winter JCF, Gosling SD, Potter J. Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: a tutorial using simulations and empirical data. Psychol Methods 2016 Sep;21(3):273-290. [doi: 10.1037/met0000079] [Medline: 27213982]
51. DIN EN ISO 9241-210:2020-03. dinmedia. URL: https://www.dinmedia.de/de/normung-kommunikation-dokumentation/BVFR010 [accessed 2026-05-06]
52. Fairbanks RJ, Wears RL, Woods DD, Hollnagel E, Plsek P, Cook RI. Resilience and resilience engineering in health care. Jt Comm J Qual Patient Saf 2014 Aug;40(8):376-383. [doi: 10.1016/s1553-7250(14)40049-7] [Medline: 25208443]
53. Birke J, Ellwart T, Sörgel J. Risk and benefit awareness as individual skills in socio-technical systems. Presented at: 22nd European Congress of Work and Organizational Psychology; May 21-24, 2025.
54. Yoshida S, Sasaki A, Sato C, et al. A novel approach to surgical instructions for scrub nurses by using see-through–type head-mounted display. Comput Inform Nurs 2015 Aug;33(8):335-338. [doi: 10.1097/CIN.0000000000000153] [Medline: 26018576]
55. Baashar Y, Alkawsi G, Wan Ahmad WN, Alomari MA, Alhussian H, Tiong SK. Towards wearable augmented reality in healthcare: a comparative survey and analysis of head-mounted displays. Int J Environ Res Public Health 2023 Feb 22;20(5):3940. [doi: 10.3390/ijerph20053940] [Medline: 36900951]
56. Yang C, Jiang P. The effect of employee STARA awareness on job crafting: exploring the moderating role of positive stress mindset. J Managerial Psychol 2025 Feb 20;40(2):211-225. [doi: 10.1108/JMP-07-2023-0369]
57. Zhang W, Guan X, Zhou X, Lu J. The effect of career adaptability on career planning in reaction to automation technology. Commun Dis Intell 2019 Oct 16;24(6):545-559. [doi: 10.1108/CDI-05-2018-0135]
58. Hackman JR, Oldham GR. Motivation through the design of work: test of a theory. Organ Behav Hum Perform 1976 Aug;16(2):250-279. [doi: 10.1016/0030-5073(76)90016-7]
59. Venkatesh V, Thong J, Xu X, University of Arkansas, Hong Kong University of Science and Technology, The Hong Kong Polytechnic University. Unified theory of acceptance and use of technology: a synthesis and the road ahead. J Assoc Inf Syst 2016;17(5):328-376. [doi: 10.17705/1jais.00428]
60. Buchinger L, Richter D, Heckhausen J. The development of life goals across the adult life span. J Gerontol B Psychol Sci Soc Sci 2022 May 5;77(5):905-915. [doi: 10.1093/geronb/gbab154] [Medline: 34417801]
61. Parasuraman R, Sheridan TB, Wickens CD. Situation awareness, mental workload, and trust in automation: viable, empirically supported cognitive engineering constructs. J Cogn Eng Decis Mak 2008 Jun;2(2):140-160. [doi: 10.1518/155534308X284417]
62. Sharma A, Kim TE. Exploring technical and non-technical competencies of navigators for autonomous shipping. Marit Policy Manag 2022 Aug 18;49(6):831-849. [doi: 10.1080/03088839.2021.1914874]
63. Hanfstingl B, Oberleiter S, Pietschnig J, Tran US, Voracek M. Detecting jingle and jangle fallacies by identifying consistencies and variabilities in study specifications - a call for research. Front Psychol 2024;15:1404060. [doi: 10.3389/fpsyg.2024.1404060] [Medline: 39282677]

Abbreviations

CEC: Clinical Ethics Committee
GDPR: General Data Protection Regulation
LoA: Level of Autonomy
RAS: robot-assisted surgery
SA: Situational Awareness

XSL•FO
RenderX

STARA: Smart Technology, AI, robotics, and algorithms

Edited by A Kushniruk, S Law; submitted 14.Nov.2025; peer-reviewed by D Menychtas, P Taiwo; revised version received 08.May.2026; accepted 23.Jun.2026; published 22.Sep.2026.

Please cite as:
Ellwart T, Birke J, Paul JT
Shifting Skills in Robot-Assisted Surgery: Case Study With Interdisciplinary Perspectives for Human Factors and Career Research
JMIR Hum Factors 2026;13:e87735
URL: https://humanfactors.jmir.org/2026/1/e87735
doi:10.2196/87735

© Thomas Ellwart, Julia Birke, Johanna T Paul. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 22.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Health Care Workers' Perspectives on Virtual Reality in Mental Health: Qualitative Interview Study

Anna Szczegielniak[1], MD, PhD; Jakub Możaryn[2], MEng, PhD; Łukasz Kunert[1], MD, PhD; Szymon Florek[1], MD, PhD; Robert Pudlo[1], MD, PhD

[1]Department of Psychoprophylaxis, Medical University of Silesia, Poniatowskiego 15, Katowice, Poland

[2]Institute of Automatic Control and Robotics, Warsaw University of Technology, Warsaw, Poland

Corresponding Author:
Anna Szczegielniak, MD, PhD
Department of Psychoprophylaxis, Medical University of Silesia, Poniatowskiego 15, Katowice, Poland

Abstract

Background: Virtual reality (VR) offers immersive, interactive environments, with growing evidence supporting its use in mental health diagnostics and therapy. Despite the increasing availability of VR equipment in clinical settings, integration into routine mental health care remains limited. Understanding health care workers' attitudes, knowledge, and perceived barriers is essential for bridging the gap between technological availability and clinical implementation.

Objective: This study aimed to investigate health care workers' attitudes toward VR in mental health treatment and identify factors influencing its adoption in settings where VR equipment is already available.

Methods: A qualitative study design was used, combining structured one-to-one interviews with a supporting narrative literature review. The narrative review informed the interview design and contextualized findings within the existing evidence base. Structured interviews were conducted with 16 mental health professionals (n=12, 75% women; n=4, 25% men) working in psychiatric and mental health settings in one city in southern Poland where VR equipment was available for patient care. Descriptive statistics summarized closed-ended question responses, and thematic analysis was applied to open-ended questions.

Results: Of the 16 participants, 10 (62.5%) had heard of VR technology before and held a favourable view of it, but only 6 (37.5%) were aware of its applications in mental health care. Participants aged under 35 years (6/16) rated the value of VR in therapy more positively than those aged over 50 years (3/16). Among those with prior VR exposure, 85.7% (6/7) expressed confidence in its effectiveness compared with 44.4% (4/9) of those without prior exposure. The most frequently mentioned benefits included interactivity, the ability to individualize therapeutic activities to patient needs, isolated visual and auditory interactions, and an increased sense of patient involvement and control. High costs were identified as the primary barrier by 75% (12/16), and 68.8% (11/16) expressed a need for comprehensive training. Women rated their subjective VR knowledge lower than men despite demonstrating comparable objective knowledge during interviews. VR equipment was described as being restricted within certain service areas. Legal and insurance constraints limited the use of home-based VR.

Conclusions: Access to VR equipment alone is insufficient to drive adoption in mental health services. Effective integration requires targeted education, comprehensive training, clear institutional guidelines on equipment responsibility and insurance, and supportive infrastructure. Addressing barriers and knowledge gaps identified in this study is essential for realizing VR's therapeutic potential in routine clinical practice.

(JMIR Hum Factors 2026;13:e80851) doi:10.2196/80851

KEYWORDS

virtual reality; mental health; health care workers; qualitative research; technology adoption; barriers; attitudes

Introduction

Virtual Reality Applications in Mental Health

Virtual reality (VR) technology has evolved substantially over the past 2 decades, offering immersive, interactive environments with growing applications in mental health diagnostics and therapy [1-3]. VR enables clinicians to create controlled, reproducible therapeutic scenarios that would be difficult or impossible to replicate in traditional clinical settings. Clinical VR tools have advanced the prevention, assessment, and treatment of conditions including posttraumatic stress disorder [4], anxiety disorders [5-7], and depression [3,8].

The therapeutic potential of VR extends across a broad range of mental health conditions. VR-based interventions have demonstrated efficacy for anxiety disorders through VR exposure therapy, which allows patients to confront fear-inducing stimuli in a safe, controlled environment [6,9,10]. Meta-analyses have confirmed that VR interventions produce

XSL•FO
RenderX

significant reductions in symptoms of anxiety and depression [3,8]. Beyond exposure therapy, VR has shown promise for social skill training in schizophrenia [11], pain management [12], and broader psychotherapeutic applications [13].

One factor in assessing the viability of VR therapy is treatment adherence. A meta-analysis of 46 studies on VR therapy for anxiety disorders found an overall dropout rate of 16%, lower than estimated rates for in vivo therapy and cognitive behavioral therapy (CBT) [14]. Similarly, a review of CBT in VR treatments found no significant difference in dropout likelihood between VR and traditional modalities [15], suggesting that VR-based approaches are at least as acceptable to patients as conventional treatments.

Accessibility remains a critical consideration. Digital mental health tools, including VR, must be designed to accommodate diverse populations, including those with disabilities and limited technological literacy [16,17]. The digital divide—unequal access to digital mental health resources—disproportionately affects people with serious mental illness in community settings [18]. Addressing these accessibility concerns is essential for the equitable implementation of VR in mental health services.

Research on Health Care Workers' Attitudes and Barriers Conducted to Date

A growing body of research has examined health care professionals' perspectives on VR in clinical practice. Recent years have seen a substantial increase in studies investigating clinician attitudes, acceptance, and implementation barriers [19-23]. A systematic review by Bin et al [24] synthesized evidence on mental health professionals' attitudes toward VR therapies, whereas Felnhofer et al [25] identified key barriers to adoption from the perspective of clinical psychologists and psychotherapists, including concerns about technical complexity, insufficient evidence, and lack of training. Recent reviews [26,27] of VR interventions for psychosis demonstrated the expanding scope of VR applications in severe mental illness and the corresponding need to understand clinician readiness.

Studies focused on cognitive behavioral therapists have provided important insights. Lindner et al [19] conducted a cross-sectional survey and found generally positive attitudes toward VR therapy, with therapists recognizing its potential not only for anxiety disorders but also for other conditions. However, therapists' decisions about future VR use were more strongly influenced by negative attitudes—concerns and fears—than by positive attitudes, suggesting that addressing clinician apprehensions may be more effective than simply promoting benefits [19,28,29]. Therapists' main concerns included doubts about the real-world transfer of therapeutic gains, worries about the quality of the virtual environment, and practical issues such as the potential impact of VR goggles on the therapeutic relationship and eye contact [20,30].

The successful integration of new technologies into health care depends substantially on the knowledge and training of health care workers. Adequate training enhances acceptance and effective use [31]. Designated champions—personnel with specific training and a clear institutional mandate—can significantly facilitate technology adoption within services [32,33]. Wray et al [20] identified that, despite the unambiguous success of VR treatments for anxiety disorders, few therapists use them in practice, pointing to a gap between evidence and implementation driven by barriers at multiple levels: individual (lack of training and unfamiliarity), institutional (cost and infrastructure), and systemic (lack of guidelines and insurance coverage).

Health care professionals' attitudes toward adopting technological innovations are influenced by factors including perceived usefulness, ease of use, and potential to enhance patient care [34]. Positive attitudes are often linked to the technology's perceived benefits and alignment with clinical workflows [35]. However, resistance can occur due to concerns about disrupting established practices or skepticism about effectiveness. Engaging health care workers in development and implementation processes can lead to more favorable attitudes [36]. Health care providers have reported generally favorable attitudes toward VR despite limited previous clinical use and perceived barriers to implementation [21]. Clinician-focused research further highlights the importance of practical guidance and institutional support for successful VR adoption [22].

Despite the potential benefits, several barriers hinder widespread VR adoption in mental health care. Cost is a significant factor as acquiring and maintaining VR equipment can be expensive. Safety concerns arise regarding the potential for cybersickness or exacerbation of symptoms in vulnerable populations. Ethical considerations include ensuring patient privacy, obtaining informed consent, and addressing the risk of overreliance on technology at the expense of human interaction [37,38]. Addressing these barriers requires comprehensive strategies encompassing policy development, training, and ongoing evaluation.

Aims of This Study

The present study was conducted in mental health settings in Poland where VR equipment was already available for patient care. This unique context—investigating staff who have physical access to VR but may not use it routinely—enabled examination of whether access alone is sufficient to drive adoption. We aimed to (1) explore health care workers' attitudes, knowledge, and perceptions regarding VR in mental health care; (2) identify barriers to and facilitators of VR adoption in settings where equipment is already available; and (3) examine how demographic factors such as age, gender, and professional experience relate to attitudes toward VR. We hypothesized that mere availability of VR equipment would be insufficient to ensure its integration into clinical practice and that training, institutional support, and prior personal experience with VR would be key factors shaping adoption.

Methods

Study Design

This study used a qualitative design combining structured one-to-one interviews with a supporting narrative literature review to examine health care workers' attitudes and knowledge regarding VR in mental health care. The narrative review

XSL•FO
RenderX

informed the interview design and contextualized findings within the existing evidence base. The structured interview guide was developed based on themes identified in the literature, including VR awareness, perceived benefits and barriers, training needs, and attitudes toward technology adoption in clinical settings.

Participants

This study targeted health care workers actively engaged in mental health services. A total of 16 participants were recruited, comprising psychiatrists (n=7, 43.8%), mental health nurses (n=3, 18.8%), psychologists (n=2, 12.5%), CBT psychotherapists (n=2, 12.5%), physical therapists (n=1, 6.3%), and social workers (n=1, 6.3%). Of the 16 participants, 12 (75%) were women, and 4 (25%) were men. Most (n=11, 68.8%) had experience working only with adult patients. Participants were recruited across mental health care settings in one city in southern Poland, including a general psychiatry hospital unit, ambulatory care, and a community mental health center. All selected participants worked in services that offered patients the opportunity to participate in VR-based therapy and were involved in the care of patients taking part in a therapeutic program using VR.

Of 22 interview invitations, 6 (27.3%) individuals declined due to conflicting work commitments. The sample included individuals with diverse experience levels, from newly trained professionals to senior practitioners, to capture a range of perspectives. Demographically, efforts were made to ensure representation across gender and age groups to account for cultural and systemic influences on attitudes toward VR. .

Data Collection

The structured interview collected responses from professionals in therapeutic fields about their experiences, perceptions, and knowledge regarding VR tools in therapy. The interview guide (Multimedia Appendix 1) covered the following domains: (1) VR familiarity—participants' prior exposure to and subjective knowledge of VR technology, (2) perceived advantages and limitations—benefits and drawbacks of VR in therapeutic settings, (3) barriers and support needs—factors hindering VR adoption and resources required for implementation, and (4) future use—likelihood of integrating VR tools into regular clinical practice.

The interview comprised both closed-ended questions (eg, rated on Likert-type scales) and open-ended questions designed to elicit detailed perspectives. Each interview lasted approximately 20 to 30 minutes and was conducted entirely by a single researcher (AS) to ensure consistency and reduce potential interviewer bias. Interview themes and sample questions are shown in Textbox 1.

Textbox 1. Interview questions regarding virtual reality (VR) use in clinical practice.

VR familiarity

- How would you rate your knowledge of VR technology in general?
- Have you previously used VR tools in other settings (eg, entertainment, games)?
- Have you previously used VR tools in your patient care practices?

VR in a therapeutic context

- In your opinion, which disorder group, out of the following, could VR tools be used successfully?
- In your opinion, which disorder group, out of the following, should VR tools not be used at all?
- In your opinion, what is the biggest strength of the VR environment in a therapeutic context?

Barriers and support needs

- Do you have concerns regarding data security, patient privacy, and the storage of personally identifiable information within VR platforms?
- Would marketing resources or patient orientation materials be helpful in addressing potential patient resistance or reluctance to try VR?
- Would the provision of dedicated IT support personnel be a critical enabler for implementing VR in your workflow?

Future use

- How likely do you believe it is that VR technology will become a routinely implemented tool in your clinical environment within the next 1 to 5 years?
- How likely are you to begin using VR tools in your patient work within the next year?
- What support do you need to implement VR tools into your regular work?

Data Analysis

With participants' consent, face-to-face interviews were transcribed verbatim. At the beginning of each interview, participants were informed about the study's purpose, data security measures, and their right to withdraw at any time without providing a reason. The authors identified and discussed data saturation; after reaching this point, no additional interviews were scheduled.

Data from closed-ended questions were analyzed using descriptive statistical methods to identify patterns and trends across demographic subgroups. Given the small sample size (N=16), all analyses were exploratory and descriptive; no formal

XSL•FO
RenderX

inferential statistical tests were performed. Percentages are reported alongside absolute numbers throughout.

Open-ended question transcripts were analyzed using thematic analysis following the 6-phase approach described by Braun and Clarke [39]. Each transcript was carefully examined, and relevant passages were categorized during an initial, primarily deductive coding process based on the interview domains. Subcategories were then developed inductively from the data, and significant text passages were assigned to these subcategories in a second coding process.

The narrative literature review was conducted separately to inform interview design and provide context for findings. Relevant literature was sourced from PubMed, PsycInfo, and Scopus using search terms related to VR, mental health, and health care workers' attitudes (publications up to March 2026). Inclusion criteria focused on peer-reviewed studies examining VR use in mental health care, with emphasis on adoption, clinician perspectives, and implementation barriers.

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki. It is part of a larger project titled "An attempt to identify factors influencing the effectiveness of physical activity in virtual reality conditions in the treatment of anxiety and depressive disorders" approved by the Bioethics Committee of the Medical University of Silesia in Katowice under the number PCN/CBN/0052/KBI/106/22. This study involved voluntary, structured interviews with health care professionals about their professional opinions on VR technology. All participants provided informed verbal consent before taking part, including consent to audio recording and transcription. Participants were assured of anonymity and informed of their right to withdraw at any time without consequence. All data were stored securely, and identifying information was removed during transcription.

Results

Participant Characteristics

Sixteen mental health professionals participated in the structured interviews. Table 1 summarizes participant demographics. The detailed recruitment process is presented in Figure 1.

Table . Participant demographics (N=16).

Profession	Frequency, n (%)	Gender[a]	Experience (y)[b]
Psychiatrists	7 (43.8)	5 women; 2 men	5 - 25
Mental health nurses	3 (18.8)	3 women	10 - 30
Psychologists	2 (12.5)	1 woman; 1 man	3 - 15
CBT[c] psychotherapists	2 (12.5)	2 women	5 - 10
Physical therapists	1 (6.3)	1 woman	20
Social workers	1 (6.3)	1 man	8

[a]Total: 12 women; 4 men.

[b]Mean experience was 14.70 (SD 11.91) years.

[c]CBT: cognitive behavioral therapy.

XSL•FO
RenderX

Figure 1. Flow diagram of participant recruitment. CBT: cognitive behavioral therapy; VR: virtual reality.

XSL•FO
RenderX

VR Awareness and Knowledge

Participants rated their subjective VR knowledge with a mean of 2.44 (SD 1.26) out of 5 (where 1 indicated complete unfamiliarity with VR technology and 5 indicated extensive knowledge), suggesting moderate self-assessed familiarity. No clear patterns were observed between VR knowledge ratings and demographic factors such as age or years of experience.

A gender-based pattern was noted: women rated their subjective VR knowledge lower than men despite demonstrating comparable objective knowledge when asked factual questions about VR functionality during interviews. When women were asked general questions about VR equipment appearance and use (assessing objective knowledge), their responses were at the same level as those of men. This mismatch between subjective self-assessment and objective knowledge among women may reflect broader gendered patterns in technology confidence and warrants further investigation.

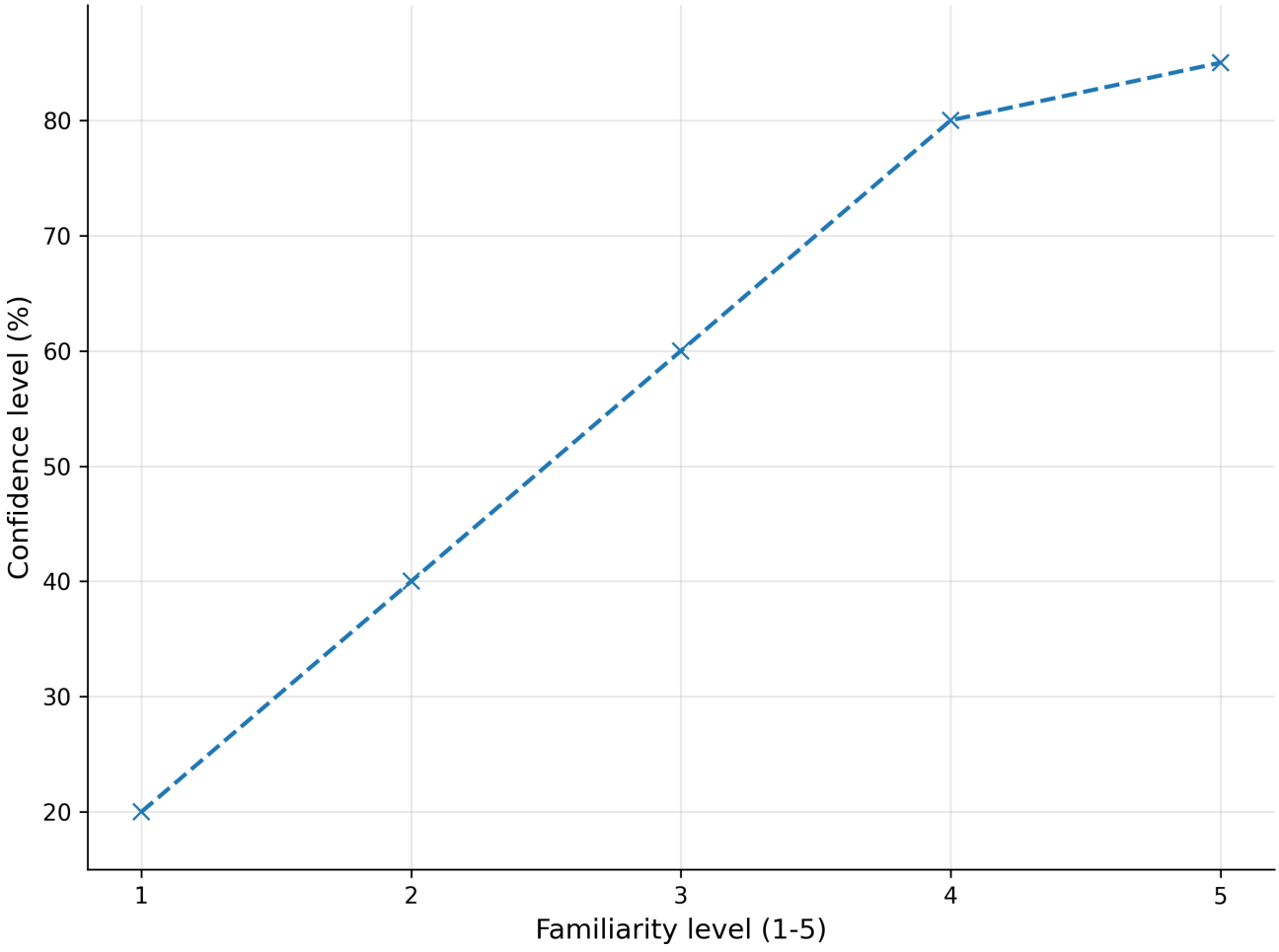
A positive relationship was observed between personal use of VR outside the workplace (eg, for gaming or entertainment) and self-rated VR knowledge, suggesting that personal familiarity enhances confidence (Figure 2). Higher self-rated VR knowledge was associated with a greater expressed likelihood of adopting VR tools in clinical practice, underscoring the importance of exposure and training.

One participant, a nurse aged 56 years, offered an illustrative example. Although she had never used VR herself, her son had. After hearing about his experiences, she concluded that VR could be a valuable relaxation tool for patients in a hospital ward, especially those with limited opportunities to go for walks or connect with nature. This suggests that even indirect exposure—through family members or peers—can shape positive attitudes toward VR.

Notably, mainly younger participants reported having had nonclinical contact with VR. Participants also mentioned that even when using VR outside clinical settings, it was typically done under supervision. They expressed concerns about the equipment being expensive and requiring oversight, which contributed to a feeling of limited independence with the technology. None of the 16 participants owned VR equipment for personal entertainment or gaming at home.

Overall, 87.5% (14/16) had heard about the VR project operating in their institution. However, a smaller proportion, 25.0% (4/16), had actively used VR tools in a clinical setting. These findings indicate high awareness of the internal institutional initiative, which contrasts sharply with the staff's limited pre-existing knowledge of psychiatric VR applications and their minimal hands-on clinical experience. When asked about practical applications, most respondents pointed to relaxation aspects of VR—nature scenes, breathing exercises, and guided mindfulness. In total, 31.3% (5/16) of the participants specifically mentioned exposure therapy in the context of phobias.

Figure 2. Correlation between virtual reality (VR) familiarity and confidence, emphasizing the role of exposure in shaping attitudes toward VR adoption.



XSL•FO
RenderX

VR in the Therapeutic Context

Participants noted that VR's ability to provide interactive, personalized, and engaging therapy was a strong motivator for its potential adoption. VR was seen as particularly promising for patient groups with conditions such as anxiety or depression but less applicable for those with severe cognitive impairments.

One participant, a psychiatrist aged 32 years, stated that nearly all patients, regardless of diagnosis, could benefit from individually tailored VR therapy. She noted the following:

> *My experience is that patients are eager to participate in "novelty" classes. They are drawn to technological solutions and enjoy engaging in games that can assist with cognitive rehabilitation, even if they don't always recognize it as therapy.*

Another participant, a physiotherapist aged 51 years, emphasized the importance of VR tools for individuals with psychiatric conditions:

> *VR can help eliminate barriers to physical activity, especially for patients with mental health disorders who often engage in low levels of physical exercise. This lack of activity can lead to chronic conditions and more pronounced side effects from medication, ultimately affecting their quality of life and increasing mortality rates. VR-based activities could provide an opportunity to integrate physical activity into their lives beyond the gym or sports, which are often inaccessible to these patients.*

Among all participants, 31.3% (5/16)—3 psychiatrists, 1 psychologist, and 1 physiotherapist, all with prior VR experience—stated that no group of patients should be excluded from potential VR use. The remaining 68.8% (11/16) identified schizophrenia (during acute episodes), psychotic disorders, and advanced dementia as conditions for which VR should be avoided, citing concerns about exacerbating psychotic symptoms or causing confusion.

Practitioners expressed largely positive attitudes toward VR's potential to enhance mental health treatment. The most frequently mentioned benefits included interactivity, the possibility of individualizing therapeutic activities according to patient needs, isolated interactions in visual and auditory perception, and an increased sense of patient involvement and control. A total of 68.8% (11/16) believed that VR could complement traditional therapies, and 62.5% (10/16) identified its adaptability as a strength.

Despite this optimism, notable knowledge gaps were evident. In total, 43.8% (7/16) of participants perceived VR primarily as a gaming tool, reflecting misconceptions about its clinical applications. A lack of understanding of VR safety protocols and evidence-based effectiveness also contributed to hesitancy among some participants.

Barriers to VR Adoption

Several barriers to VR adoption in therapeutic settings were identified. Financial constraints emerged as the most prominent obstacle, with 75% (12/16) identifying the cost of purchasing and upgrading VR equipment as the greatest disadvantage. Notably, none of these participants could specify the actual cost of a single headset available at their institution; their estimates were, on average, €500 to €1000 higher than the actual cost (€1=US $1.15 as of August 5, 2026). This discrepancy suggests that perceptions of cost may be inflated relative to reality.

A particularly striking finding concerned the sense that VR equipment was "locked" or restricted within certain service areas. Participants reported that even when VR equipment was technically accessible, there was often no clear information regarding who was responsible in the event of damage. This ambiguity about responsibility created anxiety among staff and limited their willingness to use the equipment freely. A female psychiatrist aged 30 years highlighted a specific issue: for legal reasons surrounding equipment insurance, patients could not take VR goggles home. This limited the ability to maintain the positive therapeutic effects observed during ward-based VR sessions and reduced the likelihood of VR being incorporated into ongoing therapeutic programs.

Staff described wanting more independence with the equipment—the ability to use it during individual therapy sessions or bring it to patients' rooms (eg, for bedridden individuals)—but felt constrained by institutional rules, unclear protocols, and insufficient guidance. Several participants noted that although written materials about the VR program were provided, participants lacked the time to seek answers independently. Joint training sessions had been limited: only 2 were held over a 6-month period during working hours, making it difficult for many staff members to participate.

Consequently, comprehensive training emerged as the secondary barrier, with 68.8% (11/16) of participants expressing a need for thorough education. Other significant obstacles included ethical and safety concerns, raised by 37.5% (6/16), such as patient privacy, risk of retraumatization, and potential overreliance on VR technology; infrastructure limitations, including limited access to dedicated space and technical support; and digital literacy concerns, noted by 25% (4/16), who observed that without proper digital education, some patients may struggle to use VR independently.

Most participants (12/16, 75%) reported that access to VR equipment at work and involvement in caring for patients in a VR-based therapeutic program had not substantially changed their attitudes toward VR. They explained that the information provided about available therapeutic options had been superficial, lacking the full range of activities offered to patients. There was no information on whether the therapeutic protocol could be modified or whether the equipment could be used in patients' rooms or during one-on-one therapy sessions.

Facilitators of VR Adoption

When asked about the likelihood of starting to use VR tools in patient work over the following year, participants rated their chances at a mean of 3.25 out of 5 (where 1 was "very likely" and 5 was "definitely not"). Only 6.3% (1/16) selected "very likely," citing positive experiences during an ongoing ward-based VR therapeutic program.

Exploratory subgroup analysis indicated that participants aged under 35 years (6/16) rated the value of VR in therapy more

XSL•FO
RenderX

positively than those aged over 50 years (n=3; mean 3.83, SD 0.75 vs 3.00, SD 0.00 on a 5-point scale, with higher scores reflecting more favorable attitudes). Professionals with fewer than 5 years of experience (5/16) also reported a somewhat greater likelihood of starting to use VR within the next year than those with more than 15 years of experience (6/16; mean 2.80, SD 1.10 vs 3.33, SD 1.21, respectively, with lower scores indicating greater likelihood). The latter group more often cited concerns about disrupting established practices. Among those with prior VR exposure (whether personal or professional), approximately 85.7% (6/7) expressed strong confidence in its effectiveness, compared with approximately 44.4% (4/9) of those without such exposure. This pattern underscores the potential importance of hands-on experience in shaping positive perceptions.

These subgroup patterns should be interpreted cautiously given the small sample size (N=16) and are presented as exploratory, descriptive observations rather than statistically significant differences.

In response to the question "What support do you need to implement VR tools into your regular work?" participants highlighted several priorities: support from institutional leadership, education for all department staff to encourage active participation, adequate administrative and logistical support, and high-quality visualization programs enabling immersion. A total of 18.8% (3/16) of the participants specifically indicated the need for funding covering not only a one-time equipment purchase but also the systematic development and evaluation of therapeutic programs.

A female physiotherapist aged 51 years suggested that VR-based therapeutic activities should be presented to patients as additional relaxation options rather than strict treatments, emphasizing the importance of making the experience enjoyable. Other participants recommended organizing more afternoon activities, allowing patients to engage independently without staff oversight to help them feel more comfortable. They also highlighted the need to inform patients about VR's therapeutic mechanisms and evidence base.

Discussion

Principal Results

This qualitative study of 16 mental health professionals working in settings where VR equipment was available for patient care revealed a consistent pattern: physical access to VR technology alone was insufficient to drive its integration into routine clinical practice. While most participants (10/16, 62.5%) had heard of VR before the project and expressed broadly positive attitudes toward its therapeutic potential, only 37.5% (6/16) were aware of VR's mental health applications, and only 25.0% (4/16) had actively engaged with VR in a clinical capacity. This study identified a constellation of barriers—including knowledge gaps, financial concerns, insufficient training, unclear institutional guidelines, and legal and insurance constraints—that collectively prevented adoption despite equipment availability.

Several qualitative observations were particularly noteworthy. Staff described VR equipment as being "locked" within certain areas of the service, with no clear framework for independent clinical use. Anxiety about responsibility for expensive equipment deterred staff from allowing patients to use it freely. A gender-based pattern in VR knowledge self-assessment was observed, with women rating their knowledge lower than men despite demonstrating comparable factual understanding, suggesting that confidence rather than competence may be a barrier for some professionals. Younger and less experienced professionals were descriptively more open to VR, whereas those with prior personal VR exposure—even outside clinical settings—expressed markedly greater confidence in its effectiveness.

Comparison With Prior Research

Our findings are broadly consistent with the growing literature on clinician perspectives on VR in mental health care. Felnhofer et al [25] identified similar barriers among clinical psychologists and psychotherapists, including concerns about technical complexity, insufficient evidence, and lack of training. Lindner et al [19] found that CBT therapists held generally positive attitudes toward VR but that negative concerns—particularly about real-world transfer of therapeutic gains—more strongly predicted future use intentions than positive attitudes. This aligns with our finding that even staff with access to VR and generally favorable views hesitated to adopt it in the face of unresolved practical concerns.

The importance of prior exposure and familiarity identified in our study is consistent with findings by Wray et al [20], who noted that despite the unambiguous success of VR anxiety treatments, very few therapists use VR in practice, largely due to unfamiliarity and lack of training opportunities. Cushnan et al [22] similarly found that clinicians' perspectives on immersive tools were shaped by hands-on experience and institutional support. Our finding that 85.7% (6/7) of participants with prior VR exposure expressed confidence in its effectiveness compared with 44.4% (4/9) without exposure reinforces the critical role of experiential learning in technology acceptance.

The misconception that VR is primarily a gaming tool, expressed by 43.8% (7/16) of our participants, is consistent with findings that clinicians often lack awareness of VR's clinical applications [40]. At the same time, participants in our study suggested that framing VR as entertaining might paradoxically encourage patients to engage with therapeutic tools. This tension between clinical credibility and patient engagement represents an important consideration for implementation strategies.

Our study extends existing research by focusing specifically on staff working in settings where VR is already physically present but not integrated into routine practice. Previous studies have generally examined attitudes in the abstract or in settings where VR is being newly introduced. The finding that access alone is insufficient—and that institutional, legal, and training frameworks are necessary to translate access into use—adds a new dimension to the implementation literature. This is consistent with the work by Berardi et al [31], who identified systemic barriers to digital technology implementation in mental health, and by Chung et al [41], who emphasized the importance of clinician and service manager engagement.

XSL•FO
RenderX

Recent research highlights familiarity and exposure as key factors in mental health professionals' adoption of VR. Prior experience with VR and greater knowledge of VR-based therapy have been associated with a greater likelihood of future clinical use [20,42]. Moreover, recent findings similarly indicate greater interest in therapeutic VR among professionals with prior VR experience [25,43]. Structured VR training has also been shown to improve therapists' knowledge, skills, and confidence in using VR [43]. Furthermore, younger mental health professionals may show greater interest in therapeutic VR than older professionals. By contrast, a large recent study [25] found that years of professional experience were not significantly associated with interest in VR. In the exploratory sample under consideration, younger and less experienced professionals appeared more receptive to the use of VR. We should interpret these findings with caution, given the small subgroup sizes.

Limitations

This study has several limitations. First, the sample of 16 participants was drawn from mental health services in a single city in southern Poland, limiting generalizability to other geographic, cultural, or health care system contexts. Second, the use of structured interviews, while ensuring consistency, may have limited the depth of qualitative exploration compared with semistructured or in-depth interview approaches. Third, the small sample size precludes formal statistical inference; all quantitative findings are descriptive and exploratory, and subgroup patterns (by age, gender, and years of experience) should not be interpreted as statistically significant differences. Fourth, potential selection bias may have influenced participation: staff with a preexisting interest in technology may have been more inclined to take part, potentially inflating positive attitudes relative to the broader workforce. Fifth, the rapidly evolving nature of VR technology means that findings may become outdated as new, more affordable and user-friendly VR systems enter the market, necessitating ongoing research. Finally, despite individual studies showing positive effects of VR therapies, there remains insufficient scientific evidence regarding the optimal use and target populations for these tools [44], which may influence clinicians' willingness to adopt VR even when it is available.

Conclusions

There was a broadly positive outlook on VR use in mental health care among the health care professionals interviewed in this study. However, significant barriers related to knowledge, cost, training, institutional guidelines, and legal constraints must be addressed before VR can be effectively integrated into routine clinical practice. Free access to VR equipment may foster more positive attitudes, but a lack of knowledge about how to use it—combined with anxiety about equipment responsibility and the absence of clear protocols—decreases willingness to implement VR in daily therapeutic routines and may hinder its introduction as a permanent tool in mental health services.

Financial constraints extend beyond initial equipment purchase. It is equally important to establish clear rules for equipment use, create appropriate insurance conditions, and ensure systemic access to training and evaluation of therapeutic programs. Targeted education, comprehensive training, and supportive institutional infrastructure are essential to overcome the barriers identified in this study and fully realize VR's potential as a therapeutic tool. Future research should include larger, multisite samples; use semistructured interview methods to capture richer qualitative data; and investigate the effectiveness of specific implementation strategies—such as VR champion programs, structured training curricula, and institutional policy frameworks—in translating equipment access into routine clinical use.

A visual abstract summarizing the main study findings, implementation barriers, and integration requirements is presented in Multimedia Appendix 2.

Funding

This research did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability

The data that support the findings of this study are available from the corresponding author, AS, on reasonable request.

Authors' Contributions

Conceptualization: AS, RP
Data curation: AS, ŁK, SF
Formal analysis: JM
Investigation: AS, ŁK, SF
Methodology: AS, RP
Project administration: RP
Resources: AS, RP
Software: JM
Supervision: RP
Validation: AS, LK, SF
Visualization: JM
Writing—original draft: AS, JM
Writing—review and editing: AS, JM, LK, SF, RP

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide used in this study.
[DOCX File, 21 KB - humanfactors_v13i1e80851_app1.docx]

Multimedia Appendix 2
Visual Abstract summarizing health care workers' perspectives, key implementation barriers, and requirements for successful virtual reality integration in mental health care.
[PNG File, 2943 KB - humanfactors_v13i1e80851_app2.png]

References

1. Gregg L, Tarrier N. Virtual reality in mental health: a review of the literature. Soc Psychiatry Psychiatr Epidemiol 2007 May;42(5):343-354. [doi: 10.1007/s00127-007-0173-4] [Medline: 17431528]
2. Freeman D, Reeve S, Robinson A, et al. Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychol Med 2017 Oct;47(14):2393-2400. [doi: 10.1017/S003329171700040X] [Medline: 28325167]
3. Fodor LA, Cote CD, Cuijpers P, Szamoskozi , David D, Cristea IA. The effectiveness of virtual reality based interventions for symptoms of anxiety and depression: a meta-analysis. Sci Rep 2018 Jul 9;8(1):10323. [doi: 10.1038/s41598-018-28113-6] [Medline: 29985400]
4. Rizzo A, Shilling R. Clinical virtual reality tools to advance the prevention, assessment, and treatment of PTSD. Eur J Psychotraumatol 2017;8(sup5):1414560. [doi: 10.1080/20008198.2017.1414560] [Medline: 29372007]
5. Glegg SM, Levac DE. Barriers, facilitators and interventions to support virtual reality implementation in rehabilitation: a scoping review. PM R 2018 Nov;10(11):1237-1251.e1. [doi: 10.1016/j.pmrj.2018.07.004] [Medline: 30503231]
6. Krzystanek M, Surma S, Stokrocka M, et al. Tips for effective implementation of virtual reality exposure therapy in phobias-a systematic review. Front Psychiatry 2021;12:737351. [doi: 10.3389/fpsyt.2021.737351] [Medline: 34621197]
7. Shahid S, Kelson J, Saliba A. Effectiveness and user experience of virtual reality for social anxiety disorder: systematic review. JMIR Ment Health 2024 Feb 8;11:e48916. [doi: 10.2196/48916] [Medline: 38329804]
8. Emmelkamp PM, Meyerbröker K. Virtual reality therapy in mental health. Annu Rev Clin Psychol 2021 May 7;17:495-519. [doi: 10.1146/annurev-clinpsy-081219-115923] [Medline: 33606946]
9. Bani Mohammad E, Ahmad M. Virtual reality as a distraction technique for pain and anxiety among patients with breast cancer: a randomized control trial. Palliat Support Care 2019 Feb;17(1):29-34. [doi: 10.1017/S1478951518000639] [Medline: 30198451]
10. Maples-Keller JL, Bunnell BE, Kim SJ, Rothbaum BO. The use of virtual reality technology in the treatment of anxiety and other psychiatric disorders. Harv Rev Psychiatry 2017;25(3):103-113. [doi: 10.1097/HRP.0000000000000138] [Medline: 28475502]
11. Park KM, Ku J, Choi SH, et al. A virtual reality application in role-plays of social skills training for schizophrenia: a randomized, controlled trial. Psychiatry Res 2011 Sep 30;189(2):166-172. [doi: 10.1016/j.psychres.2011.04.003] [Medline: 21529970]
12. Wiederhold BK, Wiederhold MD. Virtual Reality Therapy for Anxiety Disorders: Advances in Evaluation and Treatment: American Psychological Association; 2005. [doi: 10.1037/10858-000]
13. Bell IH, Pot-Kolder R, Rizzo A, et al. Advances in the use of virtual reality to treat mental health conditions. Nat Rev Psychol 2024;3:552-567. [doi: 10.1038/s44159-024-00334-9]
14. Benbow AA, Anderson PL. A meta-analytic examination of attrition in virtual reality exposure therapy for anxiety disorders. J Anxiety Disord 2019 Jan;61:18-26. [doi: 10.1016/j.janxdis.2018.06.006] [Medline: 30646997]
15. Dilgul M, Martinez J, Laxhman N, Priebe S, Bird V. Cognitive behavioural therapy in virtual reality treatments across mental health conditions: a systematic review. Consort Psychiatr 2020 Sep 2;1(1):30-46. [doi: 10.17650/2712-7672-2020-1-1-30-46] [Medline: 38680386]
16. Duggin A. What we mean when we talk about accessibility. United Kingdom Government. 2016. URL: https://accessibility.blog.gov.uk/2016/05/16/what-we-mean-when-we-talk-about-accessibility-2/ [accessed 2025-05-19]
17. Bunyi J, Ringland KE, Schueller SM. Accessibility and digital mental health: considerations for more accessible and equitable mental health apps. Front Digit Health 2021;3:742196. [doi: 10.3389/fdgth.2021.742196] [Medline: 34713206]
18. Kozelka EE, Acquilano SC, Al-Abdulmunem M, et al. Documenting the digital divide: identifying barriers to digital mental health access among people with serious mental illness in community settings. SSM Ment Health 2023 Dec;4:100241. [doi: 10.1016/j.ssmmh.2023.100241]
19. Lindner P, Miloff A, Zetterlund E, Reuterskiöld L, Andersson G, Carlbring P. Attitudes toward and familiarity with virtual reality therapy among practicing cognitive behavior therapists: a cross-sectional survey study in the era of consumer VR platforms. Front Psychol 2019;10:176. [doi: 10.3389/fpsyg.2019.00176] [Medline: 30800086]

XSL•FO
RenderX

20. Wray TB, Kemp JJ, Larsen MA. Virtual reality (VR) treatments for anxiety disorders are unambiguously successful, so why are so few therapists using it? Barriers to adoption and potential solutions. Cogn Behav Ther 2023 Nov;52(6):603-624. [doi: 10.1080/16506073.2023.2229017] [Medline: 37376984]
21. Preston AM, Brown L, Padala KP, Padala PR. Veterans Affairs health care provider perceptions of virtual reality: brief exploratory survey. Interact J Med Res 2022 Sep 2;11(2):e38490. [doi: 10.2196/38490] [Medline: 36053568]
22. Cushnan J, McCafferty P, Best P. Clinicians' perspectives of immersive tools in clinical mental health settings: a systematic scoping review. BMC Health Serv Res 2024 Sep 18;24(1):1091. [doi: 10.1186/s12913-024-11481-3] [Medline: 39294727]
23. Skeva R, Gregg L, Jay C, Pettifer S. Views of practitioners and researchers on the use of virtual reality in treatments for substance use disorders. Front Psychol 2021;12:606761. [doi: 10.3389/fpsyg.2021.606761] [Medline: 34093303]
24. Bin S, Alrashdi DH, Whitehead T, Riches S, Drini E. Mental health professionals' attitudes towards virtual reality therapies: a systematic review. J Technol Behav Sci 2025;11:146-169. [doi: 10.1007/s41347-025-00499-x]
25. Felnhofer A, Pfannerstill F, Gänsler L, et al. Barriers to adopting therapeutic virtual reality: the perspective of clinical psychologists and psychotherapists. Front Psychiatry 2025;16:1549090. [doi: 10.3389/fpsyt.2025.1549090] [Medline: 40171310]
26. Schroeder AH, Bogie BJM, Rahman TT, Thérond A, Matheson H, Guimond S. Feasibility and efficacy of virtual reality interventions to improve psychosocial functioning in psychosis: systematic review. JMIR Ment Health 2022 Feb 18;9(2):e28502. [doi: 10.2196/28502] [Medline: 35179501]
27. Spark J, Pot-Kolder R, Dzafic I, Nelson B, Byrne LK, Lum JAG. Virtual reality for the treatment of positive symptoms of psychosis: a meta-analysis of trials. Curr Treat Options Psych ;12(1). [doi: 10.1007/s40501-025-00350-3]
28. Harned MS, Dimeff LA, Woodcock EA, Contreras I. Predicting adoption of exposure therapy in a randomized controlled dissemination trial. J Anxiety Disord 2013 Dec;27(8):754-762. [doi: 10.1016/j.janxdis.2013.02.006] [Medline: 23538148]
29. Harned MS, Dimeff LA, Woodcock EA, et al. Exposing clinicians to exposure: a randomized controlled dissemination trial of exposure therapy for anxiety disorders. Behav Ther 2014 Nov;45(6):731-744. [doi: 10.1016/j.beth.2014.04.005] [Medline: 25311284]
30. Segal R, Bhatia M, Drapeau M. Therapists' perception of benefits and costs of using virtual reality treatments. Cyberpsychol Behav Soc Netw 2011;14(1-2):29-34. [doi: 10.1089/cyber.2009.0398] [Medline: 21329440]
31. Berardi C, Antonini M, Jordan Z, Wechtler H, Paolucci F, Hinwood M. Barriers and facilitators to the implementation of digital technologies in mental health systems: a qualitative systematic review to inform a policy framework. BMC Health Serv Res 2024 Feb 26;24(1):243. [doi: 10.1186/s12913-023-10536-1] [Medline: 38408938]
32. Deighan MT, Ayobi A, O'Kane AA. Social virtual reality as a mental health tool: how people use VRChat to support social connectedness and wellbeing. In: CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems: Association for Computing Machinery; 2023. [doi: 10.1145/3544548.3581103]
33. Shiner CT, Croker G, McGhee J, Faux SG. Perspectives on the use of virtual reality within a public hospital setting: surveying knowledge, attitudes, and perceived utility among health care professionals. BMC Digit Health 2024;2:18. [doi: 10.1186/s44247-024-00076-x]
34. Freeman D, Lambe S, Kabir T, et al. Automated virtual reality therapy to treat agoraphobic avoidance and distress in patients with psychosis (gameChange): a multicentre, parallel-group, single-blind, randomised, controlled trial in England with mediation and moderation analyses. Lancet Psychiatry 2022 May;9(5):375-388. [doi: 10.1016/S2215-0366(22)00060-8] [Medline: 35395204]
35. Gonçalves R, Pedrozo AL, Coutinho ES, Figueira I, Ventura P. Efficacy of virtual reality exposure therapy in the treatment of PTSD: a systematic review. PLoS One 2012;7(12):e48469. [doi: 10.1371/journal.pone.0048469] [Medline: 23300515]
36. Groom V, Bailenson JN, Nass C. The influence of racial embodiment on racial bias in immersive virtual environments. Soc Influ 2009;4(3):231-248. [doi: 10.1080/15534510802643750]
37. He J, Wu Y. Application of digital interactive display design based on computer technology in VR film. Mob Inf Syst 2022 Oct 10;2022:1-7. [doi: 10.1155/2022/8462037]
38. Wiebe A, Kannen K, Selaskowski B, et al. Virtual reality in the diagnostic and therapy for mental disorders: a systematic review. Clin Psychol Rev 2022 Dec;98:102213. [doi: 10.1016/j.cpr.2022.102213] [Medline: 36356351]
39. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
40. Tarr B, Slater M, Cohen E. Synchrony and social connection in immersive virtual reality. Sci Rep 2018 Feb 27;8(1):3693. [doi: 10.1038/s41598-018-21765-4] [Medline: 29487405]
41. Chung OS, Robinson T, Johnson AM, et al. Implementation of therapeutic virtual reality into psychiatric care: clinicians' and service managers' perspectives. Front Psychiatry 2022;12:791123. [doi: 10.3389/fpsyt.2021.791123] [Medline: 35058823]
42. Gomez Bergin AD, Allison AM, Hazell CM. Understanding public perceptions of virtual reality psychological therapy using the Attitudes Towards Virtual Reality Therapy (AVRT) scale: mixed methods development study. JMIR Ment Health 2024 Jan 12;11:e48537. [doi: 10.2196/48537] [Medline: 38214958]

XSL•FO
RenderX

43. Kouijzer MM, Koenis LA, Huizinga D, Kelders SM, Bouman YH, Kip H. Virtual reality implementation in mental health care is a marathon, not a sprint: qualitative longitudinal study of a virtual reality training program. JMIR Ment Health 2026 Mar 10;13:e83453. [doi: 10.2196/83453] [Medline: 41805629]
44. Selaskowski B, Wiebe A, Kannen K, et al. Clinical adoption of virtual reality in mental health is challenged by lack of high-quality research. NPJ Ment Health Res 2024 May 16;3(1):24. [doi: 10.1038/s44184-024-00069-8] [Medline: 38755289]

Abbreviations

CBT: cognitive behavioral therapy
VR: virtual reality

Edited by A Kushniruk; submitted 17.Jul.2025; peer-reviewed by B Ozel, R Murolo, T Ong; revised version received 19.Apr.2026; accepted 29.May.2026; published 29.Sep.2026.

Please cite as:
Szczegielniak A, Możaryn J, Kunert Ł, Florek S, Pudlo R
Health Care Workers' Perspectives on Virtual Reality in Mental Health: Qualitative Interview Study
JMIR Hum Factors 2026;13:e80851
URL: https://humanfactors.jmir.org/2026/1/e80851
doi:10.2196/80851

© Anna Szczegielniak, Jakub Możaryn, Łukasz Kunert, Szymon Florek, Robert Pudlo. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 29.Sep.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Interactive, Personalized Patient Decision Aid for COVID-19 Vaccination in Canada: User-Centered Design Approach

Doriane Etienne[1,2,3], MSc, MBA; Patrick M Archambault[1,2,4], MD, MSc; Isaac I Bogoch[5], MD, MS; Christine T Chambers[6,7], PhD; Andrea D Chittle[8], MD, MPH, CCFP; Juliette Demers[9], BSc; S Michelle Driedger[10], PhD; Ève Dubé[3], PhD; Marie-Pierre Gagnon[2,11], PhD; Teresa Gavaruzzi[12], PhD; Anik Giguère[1,2], MA, MSc, PhD; Nathalie Grandvaux[13,14], PhD; Kelly Grindrod[15], BScPharm, PharmD, MSc; Hina Hakim[1,2,3], BDS, MSc, PhD; Samira Jeimy[16], MD, PhD; Jason Kindrachuk[17], PhD; Annie LeBlanc[1,2], PhD; Shannon E MacDonald[18], PhD; Ruth Ndjaboue[19,20], PhD; Magniol Noubi[1,2,3], BSc; Rita Orji[21], PhD; Jean-Sébastien Paquette[1,2], MD, MSc; Elizabeth Parent[1], MA; Jean-Sébastien Renaud[1,2], PhD; Beate Sander[22], RN, MBA, MEcDev, PhD; Monica Taljaard[23], PhD; Dana Tannenbaum Greenberg[24], BA; Marie-Claude Tremblay[1], MA, PhD; Sabina Vohra-Miller[25], MSc; Vivian A Welch[26], PhD; Holly O Witteman[1,2,3], PhD

[1]Department of Family and Emergency Medicine, Faculty of Medicine, Laval University, Québec City, QC, Canada
[10]College of Community and Global Health, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, Canada
[11]Faculty of Nursing Sciences, Laval University, Québec, QC, Canada
[12]Department of Medical and Surgical Sciences, University of Bologna, Bologna, Italy
[13]Department of Biochemistry and Molecular Medicine, Faculty of Medicine, Université de Montréal, Montréal, QC, Canada
[14]Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Montréal, QC, Canada
[15]School of Pharmacy, Faculty of Science, University of Waterloo, Waterloo, ON, Canada
[16]Division of Clinical Immunology and Allergy, Department of Medicine, Western University, London, ON, Canada
[17]Department of Medical Microbiology & Infectious Diseases, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, Canada
[18]Faculty of Nursing, University of Alberta, Edmonton, AB, Canada
[19]Faculty of Letter and Human Sciences, School of Social Work, Université de Sherbrooke, Sherbrooke, QC, Canada
[2]VITAM – Centre de recherche en santé durable, Université Laval, 2480 Chemin de la Canardière, Québec City, QC, Canada
[20]Centre de Recherche sur le Vieillissement, Sherbrooke, QC, Canada
[21]Faculty of Computer Science, Dalhousie University, Halifax, NS, Canada
[22]Health Systems and Policy Research Collaborative Centre, University Health Network, Toronto, ON, Canada
[23]Methodological and Implementation Research, Ottawa Hospital Research Institute, Ottawa, ON, Canada
[24]Diabetes Action Canada, Toronto, ON, Canada
[25]Dalla Lana School of Public Health, Division of Clinical Public Health, University of Toronto, Toronto, ON, Canada
[26]Bruyere Health Research Institute, Ottawa, ON, Canada
[3]Research Centre of the CHU de Québec, Laval University, Québec City, QC, Canada
[4]Centre de Recherche Intégrée Pour un Système Apprenant en Santé et Services Sociaux du Centre Intégré de Santé et de Services Sociaux de Chaudière-Appalaches, Lévis, QC, Canada
[5]Divisions of General Internal Medicine & Infectious Diseases, Toronto General Hospital, University Health Network, Toronto, ON, Canada
[6]Departments of Psychology and Neuroscience & Pediatrics, Faculties of Science and Medicine, Dalhousie University, Halifax, NS, Canada
[7]Centre for Pediatric Pain Research, IWK Health, Halifax, NS, Canada
[8]Department of Family Medicine, McMaster University, Hamilton, ON, Canada
[9]Department of Biochemistry, Faculty of Science, McGill University, Montréal, QC, Canada

Corresponding Author:
Doriane Etienne, MSc, MBA
Department of Family and Emergency Medicine, Faculty of Medicine, Laval University, Québec City, QC, Canada

Abstract

Background: The COVID-19 pandemic highlighted the need for practical digital health tools to support informed decision-making amid rapidly evolving evidence and widespread misinformation.

Objective: We iteratively developed and refined VaxDA-C19, a bilingual (English and French) web-based patient decision aid designed to support informed decision-making in Canada about COVID-19 vaccination. VaxDA-C19 integrates interactive and

XSL•FO
RenderX

personalized features aimed to enhance vaccine confidence, reduce cognitive overload, and respond to diverse informational needs.

Methods: VaxDA-C19 was developed using an iterative, user-centered design approach. Throughout the development process, we involved a citizen panel, health care professionals, user experience designers, and scientific experts to guide refinements. We also conducted usability testing sessions with adults in Canada, using semistructured interviews, comparative testing, and think-aloud protocols with thematic analysis. We ultimately conducted 4 design cycles in total with adults in Canada (users) and expert reviewers (experts). Cycle 1 involved 9 people (9 users), cycle 2 involved 22 people (22 users), cycle 3 involved 9 people (3 users and 6 experts), and cycle 4 involved 9 people (9 experts).

Results: In cycle 1, user feedback guided design decisions about how to present quantitative information and technical vaccine descriptions more simply. In cycle 2, while most users (9/11, 82%) favored in-depth explanations of vaccine development, a few raised concerns about content that could be perceived as politically charged. Cycle 3 identified usability improvements, including more explicit navigation controls, simplified medical terminology, and optimized interactive components (avatars and sliders). Expert reviews in cycle 4 refined linguistic consistency, mobile responsiveness, content transparency, and scientific accuracy, emphasizing explicit instructional guidance and bilingual accessibility.

Conclusions: Our iterative process produced a personalized, bilingual digital decision aid to support evidence-informed, values-congruent decisions about COVID-19 vaccination. A randomized controlled trial will further evaluate VaxDA-C19's impact on vaccination intentions, knowledge retention, emotional responses, decisional conflict, and decisional regret. If it proves effective, the patient decision aid may also be used as a platform to support other vaccine decisions, namely, influenza, measles, shingles, pertussis, and potentially other emerging infectious diseases.

(JMIR Hum Factors 2026;13:e86283) doi:10.2196/86283

KEYWORDS

COVID-19 vaccination; patient decision aid; web-based tool; iterative human-centered design; digital health; health literacy; user engagement; public health communication; data visualization; personalized avatar

Introduction

The COVID-19 pandemic exposed challenges in public health decision-making driven by rapidly evolving scientific evidence and widespread misinformation [1-8]. People had to make complex choices, such as whether to accept vaccination or adopt preventive behaviors, often under conditions of uncertainty [4,9-13]. These challenges highlighted the need for decision support tools that empower users to make values‐congruent choices informed by the best available evidence [1,10,14-21].

Patient decision aids support shared decision-making by making decisions explicit, providing balanced and evidence-based information about options, and helping individuals clarify their values [22]. Robust evidence from meta-analyses and systematic reviews, including a comprehensive Cochrane review, demonstrates that patient decision aids significantly improve knowledge, accuracy of risk perceptions, and alignment between decisions and personal values, while also reducing decisional conflict [23,24]. In the context of vaccination, a systematic review and meta-analysis of 5 randomized controlled trials (total n=2158) found that patient decision aids for vaccine decisions had significant positive effects on both vaccine intentions (odds ratio 1.89, 95% CI 1.20‐2.97) and uptake (odds ratio 1.77, 95% CI 1.25‐2.52) [25].

Emerging evidence suggests that interactive and personalized digital tools can encourage behavior change among individuals who have not yet engaged in recommended practices, reinforce existing behaviors, and provide reassurance [26-29]. Such multiple purposes may be particularly critical during periods of crisis, when individuals seek validation for their actions and guidance for future decisions [30,31].

As COVID-19 vaccines move toward regular annual or biannual vaccines in Canada, similar to influenza vaccines, there is a need for patient decision aids that can be easily updated to reflect updated vaccines and recommendations. To this end, we created VaxDA-C19, a web-based vaccine (Vax) patient decision aid (DA) designed to support COVID-19 (-C19) vaccination decisions in Canada. VaxDA-C19 is designed to support multiple languages and to provide information in a layered way, meaning that people who want to quickly access basic information can view brief bullet points, while those who want to know more can expand sections, get more details, and access original references. Our goals were to provide clear, evidence-based information on vaccination options, benefits, and risks to oneself and others while reducing cognitive overload. This paper describes VaxDA-C19's design and development process.

Methods

Overview and Theoretical Frameworks

We developed VaxDA-C19 using an iterative, user-centered design approach. Our overall design process followed the International Patient Decision Aids Standards [32]. The patient decision aid was structured according to the Ottawa Decision Support Framework [33]. The web-based design and development approach combined human-centered design principles [34] and a Scrum framework [35]. Human-centered design was implemented via early and ongoing involvement of users and stakeholders, iterative prototyping, and cycle-by-cycle usability feedback to enhance content layout, navigation, and overall user experience (UX) [34]. Scrum served as a structure for delivering small, incremental updates through brief

XSL•FO
RenderX

development cycles, with backlog prioritization and periodic review sessions to ensure that implementation aligned with new evidence and user input [35]. We tailored Scrum for this project by synchronizing review points with usability testing rounds. These rounds included checkpoints to validate content and evidence for modifications to vaccine information or recommendations. We also used streamlined remote meetings to accommodate the team and respond to evolving guidelines.

Throughout the development process, we involved a citizen panel, health care professionals, UX designers, and scientific experts to guide refinements. We also conducted usability testing sessions with adults in Canada, using semistructured interviews, comparative testing, and think-aloud protocols with thematic analysis. We ultimately conducted 4 design cycles in total with adults in Canada and expert reviewers. Cycle 1 involved 9 people (all prospective users), cycle 2 involved 22 people (all prospective users), cycle 3 involved 9 people (3 prospective users and 6 experts), and cycle 4 involved 9 people (9 experts).

Citizen Panel

To help guide the overall project, a citizen panel of 6 individuals, comprising people from a range of age groups, languages, educational backgrounds, health statuses, and geographic regions (Alberta, Saskatchewan, Ontario, and Québec) provided feedback about the design and development of the patient decision aid and other interventions (reported separately) as well as the user testing methods and results. In the context of COVID-19 restrictions in place at the beginning of the project, the citizen panel was assembled from existing contacts with patient and citizen partners across a range of prior projects and life circumstances. This group met by Zoom (Zoom Video Communications) for 2 hours at a time to discuss ongoing project design, development, and study planning. This group served as an advisory panel, distinct from user testing with previously unknown members of the general public, described below [34].

Expert Involvement

Health care professionals and research experts, including infectious disease researchers, epidemiologists, primary care providers, and 2 UX designers (internal and external experts), participated in the iterative process as expert reviewers. The designers' expertise covered interaction design and usability evaluation (UX research and testing) as well as visual and information design. The internal designer was part of the senior author's (HOW) human-centered digital health research team and supported participatory design activities, including planning and conducting user testing in field settings or in a human-computer interaction laboratory. The external designer is a UX lead specializing in design systems, including the development and evolution of interface components. Depending on their exact area of expertise, the designers' roles were to support content presentation, interface coherence, and usability of the web-based patient decision aid [34].

Ethical Considerations

This study received ethics approval from the "Comité d'éthique de la Recherche en Sciences de la Santé" of the Université Laval (approval: 2020 - 222). All participants received detailed study information and provided written informed consent before participation. We offered user testing participants US $28.80 for their time and to help defray internet costs. We removed all identifying details of participants from reports and ensured the confidentiality of data according to institutional policies.

Recruitment for User Testing

In light of COVID-19 restrictions in place during the bulk of this project, we used online tools to conduct our user tests. Specifically, we recruited user testing participants through tailored Facebook advertisements, then conducted user testing sessions using Zoom, with participants sharing their screen with the interviewer. In line with the objectives of human-centered design, the recruitment process involved a nonprobability convenience sample to support rapid, iterative refinement. The findings were intended to inform design improvements rather than statistical generalization [36]. Eligible participants for user testing were adults (≥18 years) residing in Canada, able to communicate in English or French, and possessing basic computer skills. Individuals without internet access or unable to provide informed consent were excluded. We therefore selected Facebook options to advertise to people who were 18 years and older across Canada and who indicated an education level of either an associate degree (eg, 2-year postsecondary diploma), high school graduate, or unspecified. We did not exclude people who had other educational statuses but used tailored recruitment in an effort to include people across a range of backgrounds, as our previous experiences suggested that people who respond to such advertisements are often more highly educated. We stopped recruiting for each cycle after reaching thematic saturation in comments and usability issues identified.

Overall Design of the Intervention

We used Figma design software to iteratively design the overall structure and appearance of VaxDA-C19. As described further in the Results section, the patient decision aid ultimately included an introductory section in which the patient decision aid is presented, and the user selects the age group of the person for whom a decision is required (themself or their child), their province or territory of residence (which determines which vaccines are available), and the decision they are making (whether or not to get a primary series of vaccines, whether or not to get a booster dose, which vaccine to get, or which booster to get). The user is then invited to create an avatar representing themself or their child. This avatar is used in the next section, which describes the individual-level benefits and risks of the vaccine. VaxDA-C19 offers layered, straightforward summaries of evidence tailored to the user's selected age, group, province or territory, and decision, including vaccine availability and eligibility based on current authorizations and guidelines. It presents individual benefits and harms in both text and visual risk formats, with links to public health sources and peer-reviewed evidence. The top of the page provides a summary, and the user can click to view more detailed explanations below the summary. On the next page, the user can choose to create up to 8 more avatars representing people around them or their child or to have additional avatars auto-generated. They may then view or skip a brief (2.5-minute)

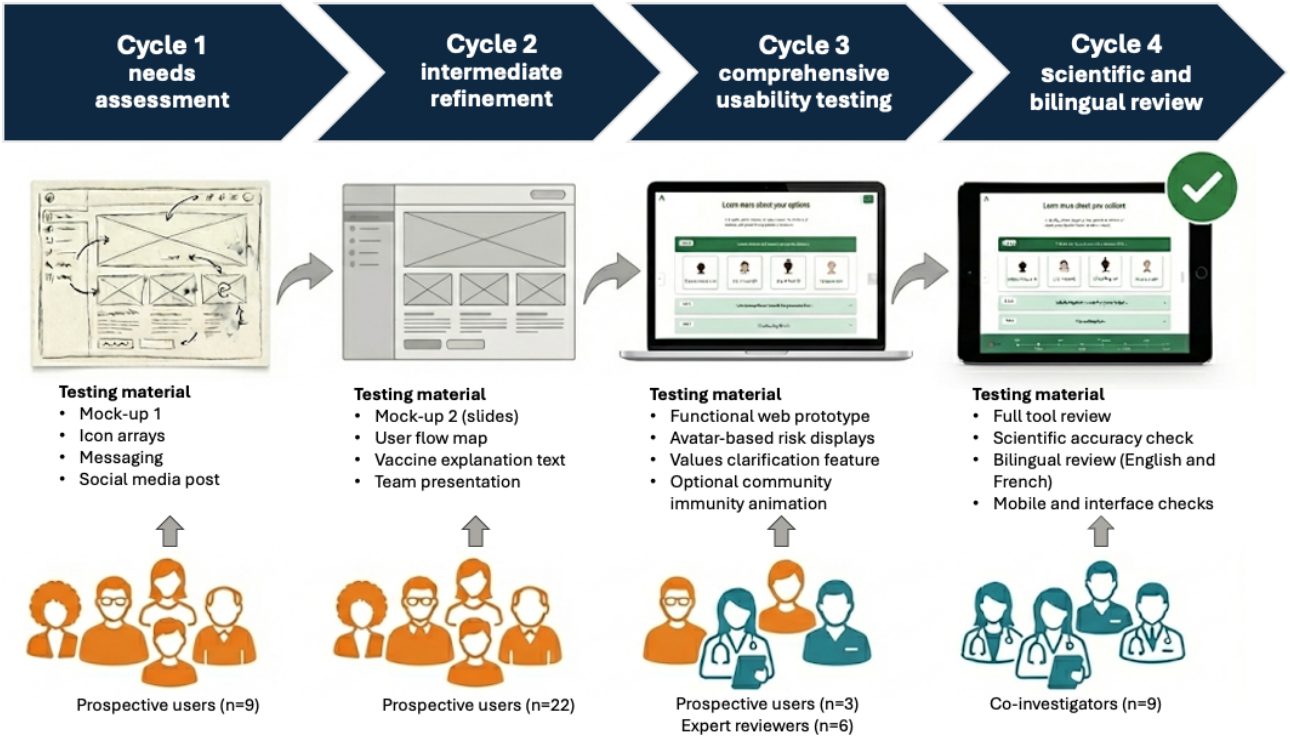
XSL•FO
RenderX

animation showing how vaccination may protect those around us [37]. This optional brief animation explains community immunity and can be viewed or skipped. At the same time, core evidence on individual risks and benefits, along with the rest of the patient decision aid, remains accessible at all times. Following this optional component, the user is invited to reflect on what matters to them, relevant to the decision, using a values clarification method previously developed and tested in the context of treatment decisions and influenza vaccination decisions [38]. Finally, the user has the option to read answers to frequently asked questions (FAQs) and to learn more about our team. A short walk-through video is available on YouTube [39].

Design and Development Approach

Our multidisciplinary team followed a human-centered design over 4 iterative cycles, summarized in Figure 1 [34]. We reported our patient decision aid development process using the DEVELOPTOOLS reporting checklist, provided in Checklist 1 [40]. As per the International Patient Decision Aids Standards [32], our design process was structured in multiple testing cycles, each involving small to medium user samples (3 - 31 participants, no overlap between cycles). This aligns with cognitive ergonomics standards [41-43], which state that iterative testing with small samples efficiently identifies approximately 85% of usability issues per cycle. After each cycle, the team met to review findings and make decisions about modifications. The citizen panel met separately throughout the project to offer advice from their viewpoints.

Figure 1. Iterative development cycles of VaxDA-C19, a web-based patient decision aid.



Data Collection and Analysis

We used Zoom to conduct user testing interviews. We evaluated whether VaxDA-C19 met its communication goals, and then applied a rapid thematic analysis [44,45] to qualitative data from interviews, comparative testing, and think-aloud sessions. Two trained research associates (EP and HH) independently coded transcripts, first extracting functional requirements (ie, needed features, content adjustments, and navigation flows), and then flagging any usability issues. We resolved disagreements through discussion with the lead author (DE) and principal investigator (HOW) until we reached a consensus. Finally, DE, EP, and HOW organized the agreed functional requirements and identified bugs into a clear, detailed plan to guide the development team. We reported the qualitative components in line with the SRQR (Standards for Reporting Qualitative Research) and provided the completed checklist in Checklist 2 [46]. We documented cycle-by-cycle decisions and changes to support transparency.

Iterative Cycles

Across all cycles, we presented participants with increasingly higher-fidelity versions of VaxDA-C19 prototypes. We asked participants to summarize the information, identify points of confusion or unanswered questions, and react and respond to wording.

Cycle 1

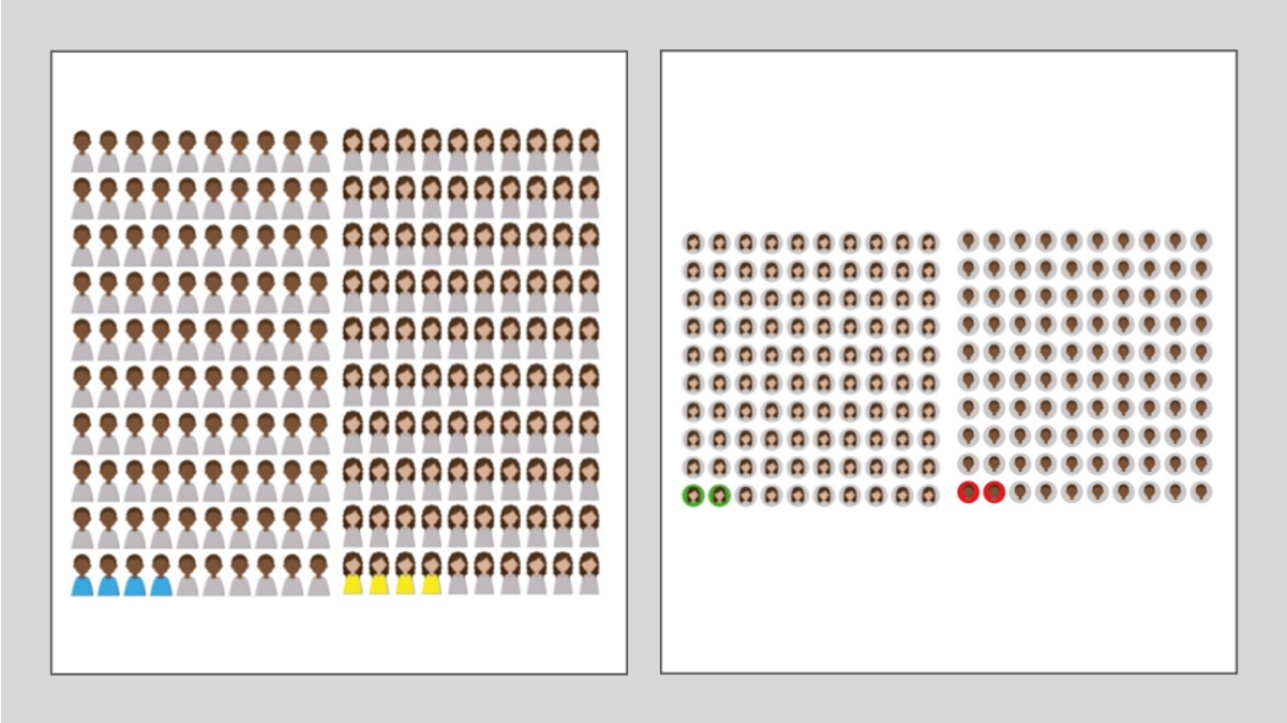
For the needs assessment, we developed an initial content mock-up of VaxDA-C19 focused on primary COVID-19 vaccination (first-dose decision). We based the mock-up's content on contemporaneous recommendations from the Canadian National Advisory Committee on Immunization, the Institut National de Santé Publique du Québec, and Health Canada authorizations. Feedback from the citizen panel and content experts addressed clarity, scientific accuracy, and relevance. We then recruited participants (n=9) via tailored Facebook advertisements and conducted individual

XSL•FO
RenderX

semistructured Zoom interviews, whose interview guides are shown in full in Multimedia Appendices 1 and 2. In each session, we asked participants to interpret icon arrays, react to hypothetical provider messaging, and consider a social media post.

For the icon arrays, as shown in Figure 2, we offered 2 icon array styles for risk communication: avatars with colored shirts and avatars with circular backdrops, and asked participants to compare these, focusing on how well or poorly these visual aids conveyed risk data.

Figure 2. Icon arrays used during the first cycle.



For the provider messaging, we asked people who were able to receive messages from their primary care provider,

> *Imagine your primary care health professional (e.g., family doctor, nurse practitioner) sends you a message like, "I got my COVID-19 vaccine when I was eligible. Now that you are eligible, I'd like to support you by answering any questions you might have. Here's a link with some information: link. Please send me a message if you have more questions." How would that make you feel? How might you react or respond?*

We asked participants who were unable to receive messages to respond to a similar message, but imagining it was during an in-person visit. Full details are shown in Multimedia Appendix 1.

The proposed social media post aimed to explain why COVID-19 vaccines moved so rapidly through development, for use in a potential image-based post, such as an Instagram carousel. The proposed English text is present in Textbox 1. Full text is available in Multimedia Appendix 2.

XSL•FO
RenderX

Textbox 1. English content for the social media intervention (Instagram carousel) explaining the accelerated development of COVID-19 vaccines.

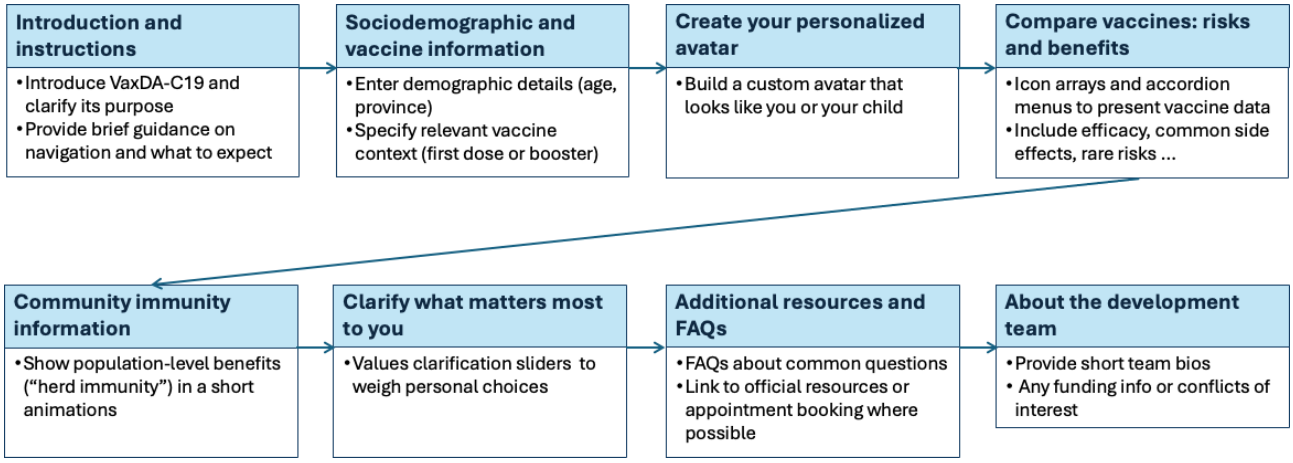
- 9 women can't make a baby in 1 month, no matter how well they work together!
- But when a bunch of the world's best scientists work together, with enough research funding, motivated by a pandemic that is disrupting lives around the world, vaccine development can happen faster than usual.
- Making a new vaccine usually involves months or years of waiting time. Waiting to get more research funding. Waiting for enough people to sign up for studies that make sure the vaccines are safe and work well. Waiting for approvals from independent regulatory authorities before starting to produce the vaccine for distribution.
- Usually, this waiting happens because the need for the vaccine isn't considered urgent enough for everyone to drop everything else and make the new vaccine everyone's top priority. But COVID-19 has been a different story. It has been a priority around the world.
- To make COVID-19 vaccines, they cut out a lot of the waiting. But they didn't cut corners. All the safety checks are still being done. When COVID-19 vaccines are approved by Health Canada (Canada's independent agency, staffed by top-notch scientists and not affiliated to any political party) you can trust that they are safe and they work.
- We are Canadian scientists, doctors, nurses, and experts in health and vaccines. None of us works for pharmaceutical companies. We don't accept pharmaceutical funds either. We will get the COVID-19 vaccines for ourselves as soon as they are approved by Health Canada and available to us. Once they have been tested and approved for children, our children will get the vaccines, too.
- Still have questions about COVID-19 vaccines and don't know where to ask? We want to use our scientific training to help answer your questions. Send us your question here. We will do our very best to answer as many as we can.

Cycle 2

Based on results from cycle 1, after identifying initial user needs, we developed the first version of content using Google Slides to create mock-ups of a proposed website. These mock-ups served as early prototypes, allowing us to conceptualize the user interface and gather feedback from the citizen panel on design, usability, and information presentation prior to investing substantial time in web development. Figure 3 maps the user flow. It began with a page of welcome text and instructions, then a page in which people indicate the age and province of residence of the person for whom a vaccine is being considered, then a page in which people build an avatar representing that person, then a page showing the individual-level risks and benefits of the vaccine for that person (using the avatar previously created to illustrate statistics), then a section showing the population-level benefits of vaccination, then a values clarification method to help people align their decision with what matters to them, and finally, FAQs, information about where to obtain a vaccine, references supporting the content, and information about our team.

We then recruited new participants (n=22) via tailored Facebook advertisements and conducted a second round of individual semistructured interviews on Zoom, again applying thematic analysis (Multimedia Appendices 3 and 4). In each interview, we asked: "In your own words, what was this text about?," "Why do you think these vaccines arrived so quickly?," and "What important questions about COVID-19 vaccines remain unanswered?" During these sessions, we used A/B testing to compare designs and content presentations of the team working on VaxDA-C19 and related projects. A/B testing, or split testing, involves presenting 2 versions (A and B) of a webpage or app to different user groups to determine which version performs better based on predefined metrics [47]. In this case, participants viewed 2 versions of the team presentation. Version A included only drawings, while version B included photographs and the names of research team members. Full details of the 2 designs are available in Multimedia Appendix 4.

Figure 3. User flow representation of our web-based patient decision aid, VaxDA-C19. FAQ: frequently asked question.



XSL•FO
RenderX

Cycle 3

Cycle 3 moved VaxDA-C19 from static mock-ups into a fully functional web prototype, and then tested it with end users and experts. Based on results from cycle 2, we added 5 interactive features. These included having users create a personalized avatar to promote personalized user engagement [48,49] and to take advantage of the Proteus effect, a documented effect in which people identify with their avatar [50,51]. VaxDA-C19 then uses that personalized avatar in icon arrays to facilitate visual risk comparisons [52-54]. It also offers users the option to view a 2-minute visualization incorporating their avatar, previously developed by our team to illustrate how community immunity (herd immunity) may contribute to protection [48,49]. The visualization makes use of 4 theories or frameworks: the health belief model [55], Gestalt visual principles [56], the cognitive theory of multimedia learning [57,58], and the affect heuristic [59,60]. We also added a values clarification feature, as per International Patient Decision Aid Standards, to help users align their vaccine decision with their values [61,62]. Finally, we used accordion menus (menus that show a brief title and allow people to click to open for full information) to enable users to decide for themselves how much or how little information they wish to read about the available vaccines [63]. We built the prototype in Vue.js (HTML, CSS, and JavaScript) [64] with JSON-templated content. This architecture enabled rapid updates as new vaccine data and guidelines evolved.

Next, we conducted a third round of online user testing, again recruiting new participants (n=3) via tailored Facebook advertisements. Through individual semistructured interviews on Zoom, we used think-aloud protocols focused on multidevice compatibility, navigation, and overall usability (technical details in Multimedia Appendix 5).

Following user testing, the lead author (DE) presented the prototype at the 18th Biennial European Conference of the Society for Medical Decision-Making during the Society for Medical Decision-Making Core Course: Introduction to shared decision-making and patient decision aids. During the session, attendees engaged in a discussion in which attendees shared insights from their respective fields. As part of this discussion, the lead author (DE) had the opportunity to show the attendees the progress of our work and to get their feedback. This additional expert feedback (n=5) addressed some direction on medical content and interface navigation. In addition, an external UX designer (n=1) delivered targeted feedback after testing to complement the internal designer's expertise.

While this cycle was underway, evolving public health guidelines suggested 4 distinct vaccination decisions: initial COVID-19 vaccination for those aged 12 years and older, booster doses for those aged 12 years and older, initial vaccination for children aged 6 months to 11 years, and booster doses for children aged 6 months to 11 years. We therefore restructured VaxDA-C19 accordingly.

Cycle 4

In cycle 4, we finalized VaxDA-C19, such that it might be ready for testing in an online randomized controlled trial. We updated its content to reflect the most recent Health Canada-authorized COVID-19 vaccines and national recommendations from the National Advisory Committee on Immunization. We then invited project coinvestigators to conduct an expert review; 9 experts provided detailed feedback. The expert reviewers were project coinvestigators with multidisciplinary expertise relevant to patient decision aids and COVID-19 vaccination, including digital health, human-centered design, virology, vaccinology, infectious disease epidemiology, primary care, public health, risk communication, health equity, and knowledge translation. Evaluation criteria encompassed scientific accuracy, linguistic clarity, content completeness, and user journey across all 4 vaccination decisions.

Based on this feedback, we prioritized key modifications, including removing irrelevant vaccine selection options for specific age groups, clarifying instructional text to specify required selections more clearly, incorporating contact information and funder logos on the homepage, and resolving layout inconsistencies, notably mobile interface challenges. Textbox 2 presents a synthesis of the key insights and considerations that emerged from our iterative design and development process.

Textbox 2. Key points from the design and development process.

Recruitment strategies: Recruitment through tailored social media advertisements (Facebook), complemented by internal dissemination via research team networks, effectively reached a diverse group of participants, including variations in educational level, age, language, and geographic location. Future studies could potentially benefit from additional strategies, such as direct collaboration with community-based organizations.

Iterative usability testing: Short, iterative usability cycles (4 cycles, sample sizes from 3 to 22 participants each) allowed rapid identification and correction of usability issues. Participants' repeated feedback emphasizes the need to simplify interactive elements (eg, avatar creation, icon arrays, and value clarification sliders), which may have contributed to enhanced clarity and ease of use across testing cycles.

Flexible content management: Implementing JSON-based templates allowed efficient updates to rapidly evolving vaccine guidelines, easing the task of ensuring ongoing scientific accuracy. Early involvement of content experts contributed to the accuracy, clarity, and scientific rigor of the web-based patient decision aid.

Multidisciplinary input: Feedback from diverse interested parties (patients, health care professionals, user experience designers, and researchers) contributed a wide range of perspectives on the clinical relevance, usability, and accuracy of VaxDA-C19.

Table 1 provides an overview of each cycle, along with the major modifications we implemented.

XSL•FO
RenderX

Table . Cycle summaries.

Cycle	Objective	Sample size, n	Procedure used	Main modifications
1	Needs assessment	9 users	• Semistructured interviews • Thematic analysis • Comparative testing	• Clarified initial content
2	Intermediate refinement	22 users	• Google Slides mock-ups • Comparative testing • Thematic analysis	• Added avatar, accordion, icon arrays, visualizations, values clarification feature
3	Comprehensive usability testing	3 users+6 experts	• Think-aloud protocols • Functional prototype testing	• Multidevice optimization • Refined navigation • Specified 4 distinct vaccination decisions
4	Scientific and bilingual review	9 experts	• Expert review • User journey analysis	• Updated content • Included French translation • Improved interface and instructional text

Results

Participant Characteristics

Across the cycles, 45 participants responded to our user testing recruitment efforts, of which 34 completed one iterative user testing phase (cycle 1: n=9, cycle 2: n=22, and cycle 3: n=3). Prospective participants who did not complete the testing did so because they did not wish to finish the study (n=2), were unavailable for any of the times scheduled for testing (n=4), responded after we had already reached saturation in findings (n=2), or for other reasons (n=3). Among the 34 participants, the mean age was 49 (SD 16) years. The majority of our study participants were women, Canadian-born, and English speakers. In total, 11/34, 32% of participants reported an ethnocultural identity other than White. Detailed demographics are shown in Table 2.

XSL•FO
RenderX

Table . User testing participant demographics across testing cycles (N=34).

Category	Cycle 1 (n=9)	Cycle 2 (n=22)	Cycle 3 (n=3)
Age (years), mean (SD)	47 (16)	50 (16)	52 (21)
Birthplace, n (%)			
Canada	7 (78)	21 (95)	3 (100)
Other	2 (22)	1 (5)	0 (0)
Preferred languages, n (%) (multiple answers allowed)			
French	6 (67)	5 (23)	0 (0)
English	3 (33)	18 (82)	3 (100)
Other	0 (0)	0 (0)	0 (0)
Ethnocultural groups, n (%) (multiple answers allowed)			
Asian—Central (eg, Kazakhstani and Uzbekistani)	0 (0)	0 (0)	0 (0)
Asian—East (eg, Chinese, Japanese, and Korean)	0 (0)	3 (14)	0 (0)
Asian—South (eg, Indian, Pakistani, and Sri Lankan)	0 (0)	1 (5)	0 (0)
Asian—South-East (eg, Malaysian, Filipino, and Vietnamese)	0 (0)	0 (0)	1 (33)
Black—African (eg, Ghanaian, Kenyan, and Somali)	2 (22)	1 (5)	0 (0)
Black—Caribbean (eg, Barbadian and Jamaican)	0 (0)	2 (9)	0 (0)
Black—North American (eg, Canadian and American)	0 (0)	0 (0)	0 (0)
First Nations	0 (0)	0 (0)	0 (0)
Indigenous or aboriginal person from outside of Canada (eg, Native American, Maori, and Quechua)	0 (0)	0 (0)	0 (0)
Inuit	0 (0)	0 (0)	0 (0)
Latin American (eg, Chilean, Mexican, and Salvadorian)	0 (0)	0 (0)	0 (0)
Metis (Metis Nation in Canada)	0 (0)	1 (5)	0 (0)
Middle Eastern (eg, Egyptian, Iranian, and Lebanese)	0 (0)	0 (0)	0 (0)
North African (eg, Moroccan and Tunisian)	0 (0)	0 (0)	0 (0)
White or European (eg, English, Italian, Portuguese, and Russian)	1 (11)	6 (27)	1 (33)
White or North American (eg, Canadian and American)	6 (67)	10 (45)	1 (33)
Other	0 (0)	0 (0)	0 (0)
Prefers not to answer	0 (0)	0 (0)	0 (0)
Disability status, n (%)			
No disability	7 (78)	15 (68)	3 (100)
At least 1 disability	2 (22)	6 (27)	0 (0)
Prefers not to answer	0 (0)	1 (5)	0 (0)
Gender identity, n (%)			
Woman	7 (78)	12 (55)	3 (100)

XSL•FO
RenderX

Category	Cycle 1 (n=9)	Cycle 2 (n=22)	Cycle 3 (n=3)
Man	2 (22)	10 (45)	0 (0)
Indigenous or other cultural gender minority identity (eg, two-spirit)	0 (0)	0 (0)	0 (0)
Something else (eg, nonbinary and gender-fluid)	0 (0)	0 (0)	0 (0)
Prefers not to answer	0 (0)	0 (0)	0 (0)
Highest education level completed, n (%)			
Some elementary school (completed or not)	0 (0)	0 (0)	0 (0)
High school diploma	1 (11)	2 (9)	0 (0)
Apprenticeship or trade certificate or diploma	0 (0)	0 (0)	0 (0)
College or polytechnical school certificate or diploma	1 (11)	7 (32)	0 (0)
University degree, bachelor level or below	3 (33)	10 (45)	2 (67)
University graduate degree (master or doctorate level)	4 (44)	2 (9)	1 (33)
Prefers not to answer	0 (0)	1 (5)	0 (0)
Do not know	0 (0)	0 (0)	0 (0)
Sexual orientation, n (%)			
Straight or heterosexual	8 (89)	17 (77)	3 (100)
Lesbian, gay, or homosexual	1 (11)	1 (5)	0 (0)
Other (options included bisexual, something else, and prefers not to answer)	0 (0)	4 (18)	0 (0)
Health status, n (%)			
Poor	1 (11)	2 (9)	0 (0)
Fair	0 (0)	3 (14)	0 (0)
Good	2 (22)	8 (36)	0 (0)
Very good	5 (56)	7 (32)	2 (67)
Excellent	1 (11)	2 (9)	1 (33)
Public-facing occupation, n (%)			
Yes	2 (22)	5 (23)	0 (0)
No	7 (78)	16 (73)	3 (100)
Prefers not to answer	0 (0)	1 (5)	0 (0)
Provinces and territories, n (%)			
Alberta	1 (11)	0 (0)	0 (0)
British Columbia	0 (0)	1 (5)	0 (0)
Manitoba	0 (0)	4 (18)	0 (0)
New Brunswick	0 (0)	1 (5)	0 (0)
Newfoundland and Labrador	0 (0)	0 (0)	0 (0)
Nova Scotia	0 (0)	0 (0)	0 (0)
Ontario	2 (22)	11 (50)	3 (100)
Prince Edward Island	0 (0)	0 (0)	0 (0)
Quebec	6 (67)	5 (23)	0 (0)

XSL•FO
RenderX

Category	Cycle 1 (n=9)	Cycle 2 (n=22)	Cycle 3 (n=3)
Saskatchewan	0 (0)	0 (0)	0 (0)
Northwest Territories	0 (0)	0 (0)	0 (0)
Nunavut	0 (0)	0 (0)	0 (0)
Yukon	0 (0)	0 (0)	0 (0)
Residing in a rural community, n (%)			
No	—[a]	17 (77)	2 (67)
Yes	—	2 (9)	1 (33)
Do not know	—	3 (14)	0 (0)

[a]Not available.

Cycle 1

Main Findings

Icon-Array Interpretation

All participants (n=9) correctly identified the icon arrays as representing a "group of people" and understood that the colored icons indicated numerical or probabilistic information. However, some participants struggled initially with the precise probabilistic meaning of each icon. Regarding layout preferences, 5 participants preferred the arrays using circular backgrounds, finding them "easier to count" due to clearer spacing and structure. In total, 3 participants favored the shirt-colored arrays because they appreciated seeing "more character detail," and 1 participant found both layouts equally clear but mentioned that the shirt-colored icons seemed to visually represent more data because of their elongated shape.

Provider-Sent Message Framing

When participants reacted to the hypothetical vaccine recommendation message from their primary care provider (Multimedia Appendix 1), 5 of 9 participants expressed neutral feelings toward the message. In total, 2 of 9 participants indicated that receiving such a message would encourage them to consult their provider with further questions. One participant (1/9) found the message reassuring, whereas another (1/9), who had expressed prior vaccine concerns, felt "pressured and uncomfortable."

Vaccine-Confidence Social Media Text

All participants (9/9) appreciated the explanation provided about the rapid vaccine development process (Multimedia Appendix 2). Nevertheless, several participants raised concerns regarding the text's accessibility and readability. Specifically, 5 of 9 participants thought the text was appropriate mainly for highly educated people or those who already trust the authorities. In total, 3 of 9 participants indicated that the text was too long. Conversely, 3 of 9 participants found this message reassuring. Another 3 of 9 participants (1 English-speaking and 2 French-speaking) reported confusion with the opening sentence, unsure how it related to the subsequent content. Additionally, 1 participant expressed concerns about insufficient transparency regarding potential vaccine side effects.

Citizen Panel Recommendations

The citizen panel (n=6) recommended incorporating relatable real-world data into the patient decision aid, automated vaccine selections based on provincial guidelines, and helped to refine visual branding with a logo.

Changes for the Next Cycle

Following the cycle 1 findings summarized in Table 3, we implemented several targeted modifications to improve the patient decision aid. We explicitly clarified that each icon represented 1 individual in a given population to address confusion around probabilistic visualization. Considering mixed feedback on icon-array layouts, we retained both designs for further testing.

We revised the health care provider message to a more neutral tone, addressing concerns about pressure or discomfort. Finally, to improve the vaccine-confidence content, the structure was reorganized to enhance clarity. We simplified technical language, clarified the opening sentence, and incorporated more transparent information about potential side effects, in line with participants' expectations regarding tone, accessibility, and completeness.

XSL•FO
RenderX

Table . Summary of thematic feedback from cycle 1 user testing (n=9) and changes made following citizen panel recommendations.

Theme and key feedback or issue		Participants, n (%)	Resulting modifications
Avatar arrays			
	Interpreted as a "group of people"; needed a clearer explanation that each icon represents 1 individual in a population	9 (100)	Added explanatory text indicating probabilistic meaning
Icon array design preferences			
	Circular array for better spacing	5 (56)	Refined visual layout, used more spacing
	Shirt-colored array for character detail	3 (33)	None (optional alternate design retained for future testing)
Vaccine question framing			
	Neutral feelings about the question	5 (56)	Adjusted question to more neutral wording
	Would consult a health care provider	2 (22)	None; consistent with decision aid concept
	Increased reassurance	1 (11)	None
	Felt pressured, uncomfortable	1 (11)	Wording revised for less prescriptive tone
Vaccine explanation text			
	Effective explanation of rapid vaccine development	9 (100)	Retained main points; reorganized structure
	Perceived as too technical or lengthy	5 (56)	Simplified text and terminology
	Reassuring	3 (33)	Maintained neutral tone
	Confusion with the first sentence	3 (33)	Clarified opening sentence
	Insufficient details on side effects	1 (11)	Emphasized side-effect transparency in revised draft

Cycle 2

Main Findings

Overview

During this second round of user testing, participants (n=22) reviewed 2 content variants designed to explain the rapid development of COVID-19 vaccines. The objective was to assess the clarity, effectiveness, and comprehensiveness of each version. Version A contained only illustrations, while version B included photographs and researchers' names (Multimedia Appendix 4). We randomly assigned 11 participants to version A and 11 to version B. Participants answered questions on message interpretation and identified important unanswered questions about COVID-19 vaccination missing from the explanations (Multimedia Appendix 3).

Reactions to the VaxDA-C19 Team Presentation

Participants expressed polarized opinions about version B. In total, 2 participants appreciated the inclusion of researchers' names and photos, interpreting it as a sign of credibility and transparency. One participant remarked that it was "good to see a Manitoba doctor" involved, while another said that the real faces "gave some credibility." Others expressed suspicion and discomfort, perceiving the design as overly promotional or politically biased. Participants describe the content using terms like "political PR," "spin," or sarcastic phrases such as "a miracle that will solve everything." Several participants misinterpreted the visuals, believing these scientists developed the vaccines. Some participants perceived the section as an attempt to distance pharmaceutical companies from the vaccine's origins and described this as reassuring or misleading.

Interpretation of Vaccine Explanation Text

In response to the question, "Can you tell me what this text was about?," most participants described an explanation of how vaccines were developed (version A: 7/11 and version B: 9/11). In total, 5 participants in each group thought that the section highlighted the scientists behind the vaccine development. A small number of participants said that the message aimed to counter vaccine skepticism or clarify funding sources. Table S1 in Multimedia Appendix 6 summarizes how participants interpreted the vaccine explanation text.

Reactions to the Vaccine Explanation Text

During this round of testing, participants (n=22) provided varied emotional reactions and interpretations regarding the explanatory text on rapid COVID-19 vaccine development. In total, 2 participants perceived a clear division between the scientists or authorities and the public, illustrated by expressions of suspicion or detachment such as "The people who created it are vouching for it," and referring to authorities as "Them."

XSL•FO
RenderX

Some participants identified important information gaps, particularly the lack of details on vaccine effectiveness, side effects, and characteristics of clinical trial participants. For instance, a participant remarked that the vaccine was presented as "a miracle that will solve everything," while another expressed nuanced skepticism: "I know exactly how it's done, and I take some and leave some [...] I'm a bit perplexed." In total, 2 participants specifically indicated confusion or misunderstanding about the sentence stating, "Other vaccines are still being tested in places with fewer COVID-19 cases." One described needing more time to understand the phrase clearly, while the other explicitly found it ambiguous.

One participant strongly criticized the text as too political and not scientific, disputing claims about vaccine availability and the accuracy of funding statements. Another felt that the text insufficiently emphasized the societal consequences of refusing vaccination, remarking strongly that, "It's basically a crime if people are not taking this seriously."

Finally, one participant noted perceived contradictions within the explanation, particularly regarding how previous coronavirus research aligns with the rapid vaccine development and funding during the pandemic. This participant also highlighted inconsistencies in public health messages from different provincial health authorities. Table S2 in Multimedia Appendix 6 summarizes the main emotional reactions and comprehension issues identified.

Unanswered Questions

Table S3 in Multimedia Appendix 6 summarizes the types of questions participants felt remained unanswered after reading the vaccine explanation text. These focus on vaccine mechanisms, COVID-19–specific information, distribution and eligibility, safety, and research.

Changes for the Next Cycle

Building on participant feedback in cycle 2, we made several targeted changes to the next iteration of the VaxDA-C19 decision aid to enhance its clarity, transparency, and trustworthiness. The photos and names of scientists were moved to a dedicated "About Us" section to reduce perceptions of promotional content and clarified their roles to reduce confusion. While keeping the explanatory text detailed, we simplified complex terms to increase accessibility. In response to recurring concerns and unanswered questions about vaccine technology, safety, variants, and side effects, we also expanded the FAQs section.

Cycle 3

Main Findings

During cycle 3, we conducted a third round of user testing (n=3) and expert consultation (n=6) to assess navigation, usability, and comprehension across devices. Table S4 in Multimedia Appendix 6 presents a thematic summary of participant feedback.

Navigation and Usability Improvements

Participants (users: n=3 and experts: n=6) identified significant navigation issues, notably difficulties finding the "Next" button, which impaired smooth transitions between sections. They also recommended clearer labeling and explicit instructions to clarify interactions with pages involving comparative vaccine selection and slider functionalities.

Information Transparency and Accessibility

Participants emphasized the need for simplified medical terminology and a more transparent presentation of vaccine efficacy and side effects data. They noted that clear, immediate comparative data would be crucial, particularly important for parents evaluating COVID-19 vaccination for children.

Interactive Features Optimization

Interactive elements, such as avatar creation and values clarification sliders, received mixed responses. Participants initially perceived avatar creation positively, but 1 participant found creating multiple avatars to form a community as unnecessary or excessively time-consuming. One participant had a preference for factual information rather than interactive elements or animations. Sliders were occasionally confusing due to unclear functionality interactions, dull colors, and poorly defined medical terms.

Insights From External Experts

Five health care professionals and medical decision-making researchers provided feedback at a medical decision-making conference. This feedback included recommendations for improved mobile navigation, structured vaccine selection menus, refined content spacing, and modal placements. A subsequent external reviewer (n=1) with expertise in UX recommended adding a step-by-step user guide and informative pop-ups to support user interactions.

Changes for the Next Cycle

Following cycle 3, we implemented targeted modifications to improve navigation, content clarity, and overall UX in VaxDA-C19. We enhanced button visibility and repositioned key navigation elements, adding clear user instructions to support a smoother transition. To increase content accessibility and transparency, we simplified medical terminology and added comparative data on vaccine efficacy and side effects directly within relevant sections. To streamline interactivity, we introduced an avatar bypass option and clarified the functionality of values clarification sliders through revised instructions and enhanced visual design. External expert recommendations were also applied to improve mobile navigation, refine spacing, structure vaccine menus more clearly, and integrate contextual guidance through informative pop-ups.

Cycle 4

Main Findings

Cycle 4 consisted of a comprehensive expert review (n=9), focusing on usability, clarity, bilingual implementation, and overall application refinements as shown in Table S5 in Multimedia Appendix 6.

High-Priority Usability Improvements

Experts (n=9) identified several issues affecting clarity and UX. These included irrelevant vaccine comparisons for children younger than 12 years of age, unclear user instructions,

XSL•FO
RenderX

confusing avatar placement, navigation difficulties on the mobile interface, and incomplete French implementation. Experts also suggested clarifying the decision-making context consistently across the web-based application (eg, clearly indicating options such as "I don't know") and recommended optimizations for animations and interactive sliders to reduce confusion and enhance usability.

Moderate-to-Low Priority Feedback

Additional recommendations included improving free navigation across sections, clearly displaying institutional logos and contact information, enhancing visual spacing for provincial selection buttons, and refining icon-array color schemes. Experts also proposed custom speed and pause features for animations, reducing overall text density, and simplifying vaccine comparison visuals. Due to technical and feasibility constraints, we recorded moderate-to-low priority feedback for future updates. Textbox 3 summarizes the key points from the design and development process.

Textbox 3. Key points from usability testing results.

Navigation and interface usability: Participants consistently highlighted the need for clear, intuitive, and easily navigable interfaces. Regular usability checks using small iterative samples allowed the research team to quickly identify critical navigation barriers. Participants recommended explicit labeling of navigation buttons (such as the "Next" button), logical repositioning of interactive elements, and improved visibility of control. These adjustments may improve user satisfaction and reduce cognitive load.

Information transparency and accessibility: Participants frequently emphasized the importance of clear, concise, and transparent information on vaccine efficacy and side effects. The research team anticipated that simplifying medical language and providing immediate comparative information could enhance user engagement, comprehension, and trust. Participants identified addressing transparency around sensitive health decisions, such as pediatric vaccination, as important to increase the usability and credibility of the patient decision aid.

Interactive features optimization: Participants suggested explicitly defining medical terms, clarifying instructions for using interactive elements (avatars and values clarification sliders), optimizing slider functionalities, and offering optional bypass features for less nonessential interactive steps. The research team incorporated these recommendations to reduce perceived cognitive burden and potentially enhance user experience, though further validation remains necessary.

Responsive content adjustments: User feedback identified specific gaps in information (eg, vaccine efficacy, technology, clinical trial, and distribution logistics) and highlighted areas susceptible to misunderstanding or skepticism (eg, funding sources and researchers' roles). Participants recommended expanding and refining frequently asked questions, clearly explaining scientific roles, and transparently communicating funding and clinical processes. The research team implemented these content adjustments to address user skepticism, improve comprehensiveness, and enhance perceived relevance.

Multidisciplinary and expert reviews: Feedback from multidisciplinary perspectives, including health care professionals, user experience specialists, and citizen panels, helped refine the web-based patient decision aid's usability and scientific content. The research team ranked experts' recommendations according to their overall importance and implemented all high and medium priorities, such as improved mobile navigation, a restructured vaccine selection process, and clearer scientific explanations. These refinements aimed to enhance clarity, accessibility, and user acceptance, although further formal validation of their impact is required.

Implementation of the Final Version of VaxDA-C19

VaxDA-C19 evolved from an initial Google Slides mock-up into a fully functional, bilingual web-based application (Figures 4 and 5). Key technical improvements include transitioning to a responsive HTML5 and JavaScript architecture, with real-time content synchronization (Multimedia Appendix 5). Major features include automated visualizations for herd immunity, refining comparative interactive features, and enhanced icon arrays. These refinements ensure that VaxDA-C19 remains human-centered, responsive to evolving vaccination guidelines, while improving usability and user trust.

XSL•FO
RenderX

Figure 4. Visual representation of the VaxDA-C19's choices page.

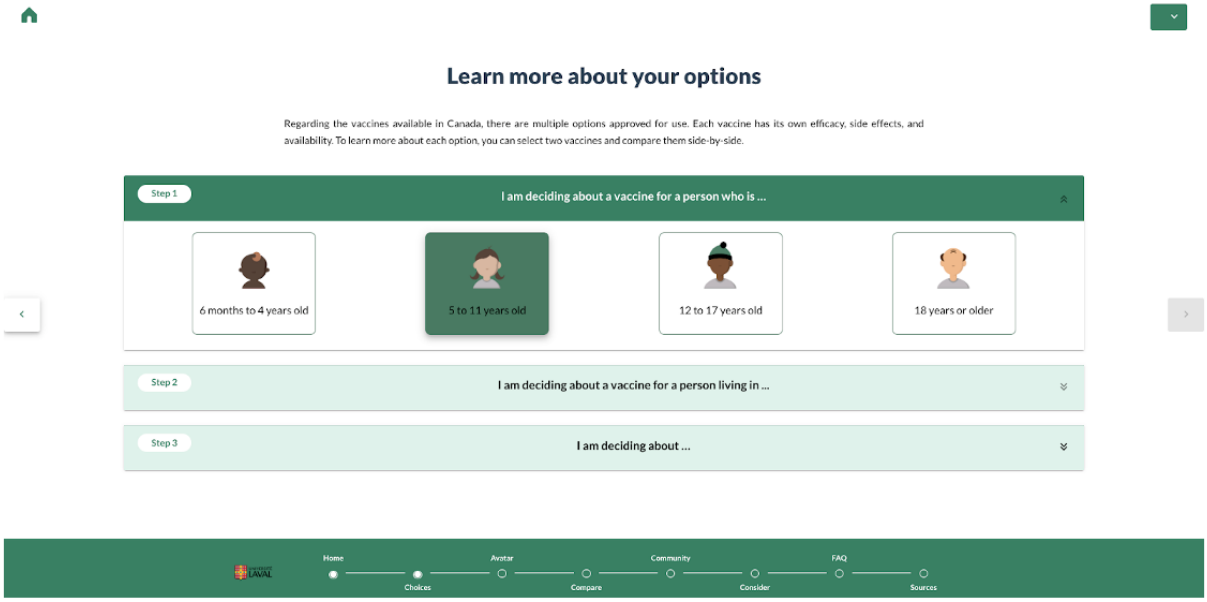
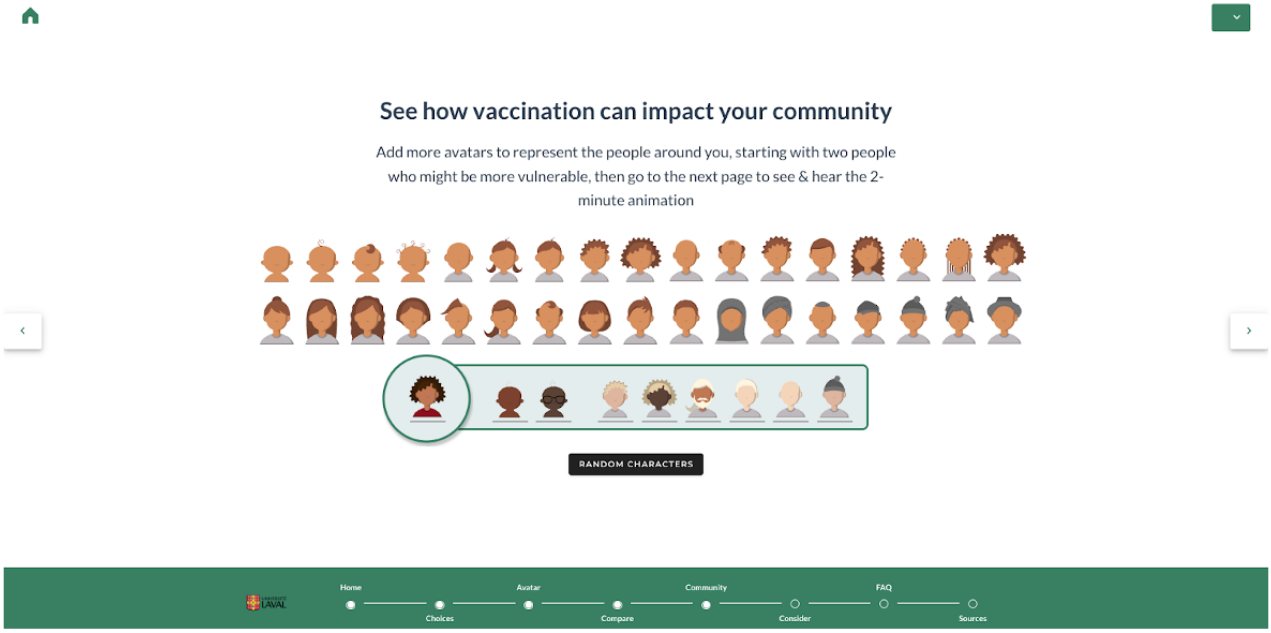


Figure 5. Selected views from the final version of VaxDA-C19.



Discussion

Principal Findings

This study aimed to iteratively develop and refine a bilingual (English and French) web-based patient decision aid (VaxDA-C19) to help people in Canada make evidence-informed, values-congruent decisions about COVID-19 vaccination. VaxDA-C19 includes interactive components such as customizable avatars, personalized icon arrays constructed out of those avatars, an optional, animated, theory-based visualization of community immunity, and a values clarification method. We included these features based on their capacity to facilitate effective communication, optimize cognitive load, enhance empathy, and provide clear, evidence-based information on vaccine options, benefits, and risks [32,48,49,52,61-63]. We present 3 principal findings from this work.

First, a recurrent theme in participant feedback was the need for more detailed and transparent information about vaccine

XSL•FO
RenderX

efficacy, variant coverage, immunity duration, and the inclusivity of clinical trials. To address these concerns, enhance transparency, and hopefully reinforce trust in the information presented, rather than offering a brief list of references in a separate section at the end of the document or website, as is commonly done in patient decision aids, we integrated explicit links to peer-reviewed studies and health authority documents immediately following each statement. We used the word "citation" for each reference and linked directly to open-access resources whenever possible. We also designed visual strategies, such as including images of scientists, to increase credibility [65-67]; however, responses to this approach varied. Some participants found this approach reassuring, while others perceived it as manipulative or politically motivated. To ultimately mitigate these concerns, we removed images and prioritized emotional neutrality and transparent, nonjudgmental communication [1,4,15,68-70]. This approach aligns with the trust determination theory, which emphasizes how credibility, inclusivity, and clarity foster public confidence and reduce skepticism [71]. Similarly, the extended parallel process model highlights how fear, distrust, and perceived governmental or institutional biases influence message reception, potentially diminishing the effectiveness of health communication [72].

Second, rather than seeking an elusive one-size-fits-all information presentation, presenting a simple overview of essential points and allowing people to easily access more detail when they want it may better serve a wider range of information needs. People with lower health literacy, who often express greater vaccine hesitancy, find more transparent explanations of risks and benefits particularly helpful [68,73-75]. Some participants in our study similarly emphasized the importance of clear, simplified explanations, particularly for individuals with varying levels of health literacy and also for people who lack the time or interest to delve deeply into information about vaccines. However, as noted earlier in our first point, other participants wanted far more detail. These tensions between some people's desires for comprehensive information and others' desires for more succinct information highlighted the complexity inherent in balancing simplicity and visual appeal with informational accuracy [76,77]. To address these competing priorities, we revised medical terminology, restructured content for better readability, and ultimately presented a summary of key evidence to all users, followed by accordion menus, which allow users to expand sections to read more about a given topic. By giving people control over the amount of detail they receive, we hope VaxDA-C19 will serve the needs of people who want a summary of essential points as well as people whose questions are not answered by simplified public health messages. Future research should continue to examine optimizing content presentation to provide sufficient detail without overwhelming users [78].

Third and finally, specific challenges associated with developing a patient decision aid for COVID-19 vaccination included a novel pathogen that launched a global pandemic, novel vaccine technology, continuously evolving evidence about vaccine effectiveness and adverse events following immunization, changing vaccine products and authorizations, and evolving recommendations for use. The fast-evolving COVID-19 vaccination landscape required proactive content management. We used flexible structures like JSON templates to enable timely updates aligned with new evidence and public health recommendations. This technical adaptability, combined with sufficient time commitments on the part of team members, should help us ensure that the VaxDA-C19 will remain scientifically accurate, while also facilitating version control, scalability, and potential future reuse for patient decision aids about other vaccine-preventable diseases.

Comparisons With Existing Literature

The COVID-19 pandemic led to the rapid development of digital patient decision aids aimed at addressing vaccine hesitancy, which intensified due to misinformation and information overload (infodemic) [79]. Several countries, including France [80], the Netherlands [81], and Australia [82], introduced digital tools to support informed decision-making. Compared to static informational approaches, these interactive decision aids improved user knowledge, reduced decisional conflict, and increased satisfaction with the decision-making process [79,81]. Despite these benefits, major challenges persist, including potential biases in content presentation, variations in complexity, and inconsistent accessibility, all of which affect usability and effectiveness [79,81].

International evaluations identified key factors that contribute to the success of digital patient decision aids. Co-design approaches that involve health professionals and target populations enhance relevance and usability. Regular updates ensure content accuracy and alignment with evolving public health guidelines. User-friendly interfaces and personalization based on individual risk assessments further support engagement and informed decision-making [79-82]. Certain digital patient decision aids have also integrated practical features, such as direct scheduling links for vaccination appointments, to facilitate action [79,81]. Similarly, in our study, multiple cycles of qualitative usability testing within an iterative, human-centered approach helped identify and address navigation, comprehension, and usability issues early in development. Small-scale usability tests effectively detected key issues, allowing for rapid refinements. Cognitive ergonomics literature supports this strategy, indicating that explanatory texts provide more explicit contextual cues and facilitate information processing, particularly for individuals with lower health literacy [83-85].

Limited linguistic and cultural adaptability restricts the reach of many existing digital patient decision aids, particularly among culturally and linguistically diverse populations [79,81]. VaxDA-C19 aims to address this limitation initially through its bilingual content in English and French, the official languages of Canada. Following evaluation in an online randomized controlled trial, we also plan to translate it into other major languages spoken in Canada.

Although digital patient decision aids may influence vaccine intentions, evidence directly linking these tools to increased vaccine uptake remains incomplete. Ongoing randomized controlled trials, such as SMART-DA in Korea [86], aim to assess their impact on key outcomes, including vaccine intentions, knowledge retention, decisional conflict, and

XSL•FO
RenderX

emotional responses. Findings suggest that patient decision aids improve preventive health behaviors by enhancing decision quality rather than by applying persuasive messaging [87,88].

Our study aligned with previous research, suggesting that effective communication in patient decision aids such as VaxDA-C19 requires careful consideration of emotions, trust, and perceived biases [1,4,10,89]. Emotional responses play an important role in how individuals interpret vaccine-related messages [1,15]. In our study, some participants expressed skepticism, anxiety, or distrust, particularly if they perceived wording as being politically charged, overly promotional, ambiguous, or institutionally biased. This finding aligns with other studies [90], and these reactions overall align with Sandman's risk=hazard + outrage model, which posits that public outrage or mistrust can overshadow objective risk assessments, leading to defensive responses such as rejection or disengagement [4,18,90,91]. Preexisting attitudes toward vaccines may also influence people's reactions to interventions such as VaxDA-C19. Future research should incorporate such considerations into study design.

Our study also confirmed that intuitive navigation, visual clarity, and balanced content presentation played essential roles in usability. These findings align with existing literature, which demonstrates that interactive and personalized elements improve engagement and comprehension in digital health interventions [52,92,93].

Strengths and Limitations

Our study has 2 main limitations. First, while recruiting user testing participants through tailored Facebook advertisements allowed us to reach people living in different parts of Canada with diverse backgrounds and who are online and therefore potential users of a web-based patient decision aid, no recruitment platform can perfectly represent the broader population. We note that our user testing participants were, as a group, slightly older than the broader group of Facebook users in Canada. Specifically, among Facebook users in Canada, the largest age group (around 25%) is between 25 and 34 years, with the next largest group (19%) being 35 to 44 years [94]. We may have recruited more people older than these groups due to increasing interest in COVID-19 vaccines with age. We note that the age group we recruited was approximately representative of the desired population for a patient decision aid about COVID-19 vaccines. Older people in Canada were already receiving strong recommendations to obtain the vaccine, whereas people in the age range of our user testing participants were likely to be both candidates themselves and may also be parents of children who were candidates for COVID-19 vaccination at the time of the study. Our study participants also included more people with higher levels of education than the general population in Canada. People with more education may seek more information about health compared to people with less education. Additionally, we note that, as a group, participants included a substantial proportion of people without a college or university degree. In other words, although there were many people with higher levels of education in our study, there were also many people without such education. No one intervention can be everything to everyone; our goal with this intervention was to provide a middle ground between the simplified information available on public health websites and the detailed information buried in technical documents and academic literature. Because vaccination campaigns are most successful when they reach as many people as possible, an approach with different components for different subpopulations (eg, an online patient decision aid for people with enough motivation to seek more information, community-based education meetings for people who trust community leaders, and simple information for people who have no questions and simply want to know where to get their vaccine) is likely to achieve the best population-level protection. Second, the challenge of balancing comprehensive content with cognitive load remains potentially unresolved, particularly in adapting information for users with varying levels of health literacy. While brief, simple messaging can be most effective in some contexts and for some people, the resulting lack of detail and missing explanations can be a problem for others. Variability in patient decision aid formats, development methodologies, and user preferences suggests that no single approach fits all contexts [25,95]. We therefore chose to prioritize user control over information presentation, trusting that many people will know what they need. Nevertheless, additional research should investigate these aspects further to refine content presentation and improve the effectiveness of digital patient decision aids.

Our study also has 2 main strengths. First, we used a rigorous iterative design based on human-centered methodologies, a multidisciplinary approach, bilingual validation, and structured engagement of a range of interested parties, including user testing, a citizen panel, UX interviews, and expert reviews. The collaboration between a diversity of researchers, UX specialists, members of the public, and health care professionals helped refine the patient decision aid, contributing to its scientific reliability and usability. User testing participants from different backgrounds offered different responses and reactions, providing a range of feedback to which we were able to respond. Complementing the contributions of these members of the public who participated in user testing and saw VaxDA-C19 at a single time point, members of the citizen panel saw VaxDA-C19's evolution over time and were able to comment at a higher level on how well or poorly the design was responding to its original goals as well as the evolving pandemic context. A wide range of experts brought insights from related projects, literature, and guidelines and also lent their expertise to verify VaxDA-C19's content and design. Systematically integrating all these people's feedback on an ongoing basis allowed for continuous adjustments. Second, we believe that VaxDA-C19 fills an unmet need by bridging the gap between highly technical vaccine information and highly simplified public health messaging. As COVID-19 vaccines continue to be offered to people in Canada, we hope to help support high-quality decision-making, thus contributing to the health of people in Canada and to public health more broadly. The platform itself may also be adapted to other vaccines.

We plan to assess the effects of VaxDA-C19 in an online randomized controlled trial (ISRCTN91743566). Specifically, participants who are eligible to receive COVID-19 vaccines will be randomly assigned to 1 of 3 arms: a control group

XSL•FO
RenderX

receiving no additional information, a usual information group directed to the Public Health Agency of Canada's COVID-19 vaccination web pages, or an intervention group receiving access to VaxDA-C19. We will assess effects on the primary outcome of vaccination intentions and secondary outcomes: knowledge, emotions, decisional conflict, vaccine uptake, and decisional regret. We will also examine effects on trust in the information provided, comparing only between the usual information and VaxDA-C19. Finally, we will conduct a follow-up study 3 - 6 months later to ascertain whether or not participants received COVID-19 vaccines.

Conclusions

VaxDA-C19 is an online patient decision aid designed and developed to help people in Canada make evidence-informed, values-congruent decisions about COVID-19 vaccines on an ongoing basis. An iterative, human-centered design approach and a multidisciplinary team bringing diverse perspectives led to a flexible application that we hope will prove useful for years to come.

Acknowledgments

The authors gratefully acknowledge the contributions of all the team members who contributed greatly to this research project as a whole but were unable to accept the invitation to participate on this particular paper as coauthors due to time constraints, including citizen panel members who chose not to accept the invitation to be coauthors: Ross McCreery, Samantha Meyer, Linda Kim, Jean Légaré, and Melissa McLeod. The authors thank Magniol Noubi and Dan Emmanuel Marie Krecoum for programming the web application and Selma Chipenda Dansokho for constructive comments throughout the project. Finally, the authors thank all study participants for helping them identify ways to improve education, communication, and decision-making in health care. The authors declare the use of generative artificial intelligence in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to generative artificial intelligence tools under full human supervision: literature search and systematization, visualization, proofreading and editing, and translation. The generative artificial intelligence tools used were Consensus AI, Perplexity, ChatGPT 4o, Gemini 3, DeepL, and Grammarly. Responsibility for the final manuscript lies entirely with the authors. Generative artificial intelligence tools are not listed as authors and do not bear responsibility for the final outcomes. DE used Consensus AI and Perplexity for literature searches and systematization. ChatGPT 4o and Gemini 3 were used for visualization, while DeepL and Grammarly were used for proofreading, editing, and translation.

Funding

This study was funded by the Canadian Institutes of Health Research and the Canadian Immunization Research Network (VR5-172668; principal investigator: HOW). The funders had no role in determining the study design, the data collection or analysis plans, the decision to publish, or the preparation of this manuscript. DE received doctoral student support from the VITAM Research Centre in Sustainable Health. HOW is funded by a Tier 2 Canada Research Chair in Human-Centred Digital Health.

Data Availability

All data generated or analyzed during this study are included in this published paper and are available on demand.

Authors' Contributions

DE, HOW, and ED contributed to the design of the study. DE, EP, HH, and HOW contributed to data collection. DE conducted data analysis and interpretation. DE and HOW drafted the first version of the paper with an early revision. PMA, IIB, CTC, ADC, JD, SMD, ED, MPG, TG, AG, NG, KG, HH, SJ, JK, AL, SEM, RN, MN, RO, JSP, EP, JSR, BS, MT, DTG, MCT, SV-M, and VAW critically revised the paper and approved the final version for submission for publication. DE and HOW had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Conflicts of Interest

PMA declares research support from AstraZeneca for an unrelated trial of an antiviral agent [96]. IIB consults to the Weapons Threat Reduction Program at Global Affairs Canada. ADC is a paid employee of the Public Health Agency of Canada, in Sexually Transmitted and Blood-Borne Infection Surveillance. KG declares speaker honouria from Apotex, the Candian Society of Hospital Pharmacists and the Ontario Pharmacy Association.

M-PG is Associate Editor of Journal of Medical Internet Research at the time of this publication. HOW is Associate Editor of JMIR Human Factors at the time of this publication.

Other authors declare no conflict of interest.

Multimedia Appendix 1

User testing interview guide—cycle 1.

[PDF File, 227 KB - humanfactors_v13i1e86283_app1.pdf]

XSL•FO
RenderX

Multimedia Appendix 2
Proposed social media text on vaccine confidence—cycle 1.
[PDF File, 166 KB - humanfactors_v13i1e86283_app2.pdf]

Multimedia Appendix 3
User testing interview guide—cycle 2.
[PDF File, 127 KB - humanfactors_v13i1e86283_app3.pdf]

Multimedia Appendix 4
Vaccine explanation text storyboard used in cycle 2.
[PDF File, 219 KB - humanfactors_v13i1e86283_app4.pdf]

Multimedia Appendix 5
Technical development and version control workflow for VaxDA-C19.
[PDF File, 93 KB - humanfactors_v13i1e86283_app5.pdf]

Multimedia Appendix 6
Qualitative findings from user testing and expert review.
[PDF File, 193 KB - humanfactors_v13i1e86283_app6.pdf]

Checklist 1
DEVELOPTOOLS reporting checklist.
[PDF File, 143 KB - humanfactors_v13i1e86283_app7.pdf]

Checklist 2
SRQR reporting checklist for qualitative studies.
[PDF File, 204 KB - humanfactors_v13i1e86283_app8.pdf]

References

1. Lavoie K, Gosselin-Boucher V, Stojanovic J, et al. Understanding national trends in COVID-19 vaccine hesitancy in Canada: results from five sequential cross-sectional representative surveys spanning April 2020–March 2021. BMJ Open 2022 Apr;12(4):e059411. [doi: 10.1136/bmjopen-2021-059411]
2. Fieselmann J, Annac K, Erdsiek F, Yilmaz-Aslan Y, Brzoska P. What are the reasons for refusing a COVID-19 vaccine? A qualitative analysis of social media in Germany. BMC Public Health 2022 Apr 27;22(1):846. [doi: 10.1186/s12889-022-13265-y] [Medline: 35484619]
3. Lazarus JV, Wyka K, White TM, et al. A survey of COVID-19 vaccine acceptance across 23 countries in 2022. Nat Med 2023 Feb;29(2):366-375. [doi: 10.1038/s41591-022-02185-4] [Medline: 36624316]
4. Pourrazavi S, Fathifar Z, Sharma M, Allahverdipour H. COVID-19 vaccine hesitancy: a systematic review of cognitive determinants. Health Promot Perspect 2023;13(1):21-35. [doi: 10.34172/hpp.2023.03] [Medline: 37309435]
5. Dionne M, Rochette L, Hamel D, Dube È. Change in intention and hesitancy regarding COVID-19 vaccines in a cohort of adults in Quebec during the pandemic. Hum Vaccin Immunother 2024 Dec 31;20(1):2309006. [doi: 10.1080/21645515.2024.2309006] [Medline: 38347660]
6. Pai M. Covidization of research: what are the risks? Nat Med 2020 Aug 1;26(8):1159-1159. [doi: 10.1038/s41591-020-1015-0]
7. Riccaboni M, Verginer L. The impact of the COVID-19 pandemic on scientific research in the life sciences. PLoS ONE 2022;17(2):e0263001. [doi: 10.1371/journal.pone.0263001] [Medline: 35139089]
8. Rawlings L, Looi JCL, Robson SJ. Economic considerations in COVID-19 vaccine hesitancy and refusal: a survey of the literature. Econ Rec 2022 Jun;98(321):214-229. [doi: 10.1111/1475-4932.12667] [Medline: 35937100]
9. Graeber D, Schmidt-Petri C, Schröder C. Attitudes on voluntary and mandatory vaccination against COVID-19: evidence from Germany. PLoS ONE 2021;16(5):e0248372. [doi: 10.1371/journal.pone.0248372] [Medline: 33970933]
10. Lin C, Tu P, Beitsch LM. Confidence and receptivity for COVID-19 vaccines: a rapid systematic review. Vaccines (Basel) 2020 Dec 30;9(1):16. [doi: 10.3390/vaccines9010016] [Medline: 33396832]
11. COVID-19 Vaccination Coverage Survey (CVCS): cycle 2 full report. Public Health Agency of Canada. 2021. URL: https://www.canada.ca/en/public-health/services/publications/vaccines-immunization/covid-19-vaccination-coverage-survey/full-report-cycle-2.html [accessed 2024-07-24]
12. Dubé E, Gagnon D, Pelletier C. COVID-19 vaccination in 5-11 years old children: drivers of vaccine hesitancy among parents in Quebec. Hum Vaccin Immunother 2022 Jan 31;18(1). [doi: 10.1080/21645515.2022.2028516]

XSL•FO
RenderX

13. Barry MJ, Härter M, Househ M, Steffensen KD, Stacey D. What can we learn from rapidly developed patient decision aids produced during the covid-19 pandemic? BMJ 2022 Sep 29;378:e071530. [doi: 10.1136/bmj-2022-071530] [Medline: 36175019]
14. Griffith J, Marani H, Monkman H. COVID-19 vaccine hesitancy in Canada: content analysis of tweets using the theoretical domains framework. J Med Internet Res 2021 Apr 13;23(4):e26874. [doi: 10.2196/26874] [Medline: 33769946]
15. Kamal A, Hodson A, Pearce JM. A rapid systematic review of factors influencing COVID-19 vaccination uptake in minority ethnic groups in the UK. Vaccines (Basel) 2021 Oct 1;9(10):1121. [doi: 10.3390/vaccines9101121] [Medline: 34696228]
16. Honora A, Wang KY, Chih WH. How does information overload about COVID-19 vaccines influence individuals' vaccination intentions? The roles of cyberchondria, perceived risk, and vaccine skepticism. Comput Human Behav 2022 May;130:107176. [doi: 10.1016/j.chb.2021.107176] [Medline: 35013641]
17. Lowe M, Harmon SHE, Kholina K, Parker R, Graham JE. Public health communication in Canada during the COVID-19 pandemic. Can J Public Health 2022 Dec;113(Suppl 1):34-45. [doi: 10.17269/s41997-022-00702-z] [Medline: 36329358]
18. Avelino-Silva VI, Ferreira-Silva SN, Soares MEM, et al. Say it right: measuring the impact of different communication strategies on the decision to get vaccinated. BMC Public Health 2023 Jun 16;23(1). [doi: 10.1186/s12889-023-16047-2]
19. Zhao S, Hu S, Zhou X, et al. The prevalence, features, influencing factors, and solutions for COVID-19 vaccine misinformation: systematic review. JMIR Public Health Surveill 2023 Jan 11;9:e40201. [doi: 10.2196/40201] [Medline: 36469911]
20. Canada's COVID-19 immunization plan: saving lives and livelihoods. Public Health Agency of Canada. 2020. URL: https://open.canada.ca/data/info/c9771ee9-fcc1-42e7-ad5a-14caf684cb70 [accessed 2024-07-03]
21. Katapally TR. A global digital citizen science policy to tackle pandemics like COVID-19. J Med Internet Res 2020 May 26;22(5):e19357. [doi: 10.2196/19357] [Medline: 32408267]
22. Patient decision aids research group. Ottawa Hospital Research Institute. URL: https://decisionaid.ohri.ca/index.html [accessed 2025-03-09]
23. Stacey D, Légaré F, Lewis K, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017 Apr;2017(4). [doi: 10.1002/14651858.CD001431.pub5]
24. Stacey D, Lewis KB, Smith M, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2024 Jan 29;1(1):CD001431. [doi: 10.1002/14651858.CD001431.pub6] [Medline: 38284415]
25. Vujovich-Dunn C, Kaufman J, King C, et al. A systematic review and meta-analysis of effectiveness of decision aids for vaccination decision-making. Vaccine (Auckl) 2021 Jun;39(28):3655-3665. [doi: 10.1016/j.vaccine.2021.05.021]
26. Marzo RR, Chen HWJ, Abid K, et al. Adapted digital health literacy and health information seeking behavior among lower income groups in Malaysia during the COVID-19 pandemic. Front Public Health 2022 Sep 14;10:998272. [doi: 10.3389/fpubh.2022.998272]
27. Montagni I, Mabchour I, Tzourio C. Digital gamification to enhance vaccine knowledge and uptake: scoping review. JMIR Serious Games 2020 May 18;8(2):e16983. [doi: 10.2196/16983] [Medline: 32348271]
28. Hakim H, Driedger SM, Gagnon D, et al. Digital gamification tools to enhance vaccine uptake: scoping review. JMIR Serious Games 2024 Feb 29;12:e47257. [doi: 10.2196/47257] [Medline: 38421688]
29. Dubé E, Gagnon D, Vivion M. Optimizing communication material to address vaccine hesitancy. Can Commun Dis Rep 2020 Feb 6;46(2-3):48-52. [doi: 10.14745/ccdr.v46i23a05] [Medline: 32167088]
30. Stacey D, Ludwig C, Archambault P, et al. Decisions and decisional needs of Canadians from all provinces and territories during the COVID-19 pandemic: population-based cross-sectional surveys. JMIR Public Health Surveill 2023 Mar 21;9:e43652. [doi: 10.2196/43652] [Medline: 36688986]
31. Brar Prayaga R, Prayaga RS. Mobile fotonovelas within a text message outreach: an innovative tool to build health literacy and influence behaviors in response to the COVID-19 pandemic. JMIR Mhealth Uhealth 2020 Aug 10;8(8):e19529. [doi: 10.2196/19529] [Medline: 32716894]
32. International Patient Decision Aids Standards (IPDAS) Collaboration. URL: http://ipdas.ohri.ca/what.html [accessed 2024-07-08]
33. Stacey D, Légaré F, Boland L, et al. 20th Anniversary Ottawa Decision Support Framework: Part 3 overview of systematic reviews and updated framework. Med Decis Making 2020 Apr;40(3):379-398. [doi: 10.1177/0272989X20911870] [Medline: 32428429]
34. Witteman HO, Vaisson G, Provencher T, et al. An 11-Item Measure of User- and Human-Centered Design for Personal Health Tools (UCD-11): development and validation. J Med Internet Res 2021 Mar 16;23(3):e15032. [doi: 10.2196/15032] [Medline: 33724194]
35. Rising L, Janoff NS. The Scrum software development process for small teams. IEEE Softw 2000;17(4):26-32. [doi: 10.1109/52.854065]
36. Ahmed SK. How to choose a sampling technique and determine sample size for research: a simplified guide for researchers. Oral Oncol Rep 2024 Dec;12:100662. [doi: 10.1016/j.oor.2024.100662]
37. Hakim H, Bettinger JA, Chambers CT, et al. An online application to explain community immunity with personalized avatars: a randomized controlled trial. medRxiv. Preprint posted online on Oct 18, 2024. [doi: 10.1101/2024.10.18.24314709]

XSL•FO
RenderX

38. Witteman HO, Julien AS, Ndjaboue R, et al. What helps people make values-congruent medical decisions? Eleven strategies tested across 6 studies. Med Decis Making 2020 Apr;40(3):266-278. [doi: 10.1177/0272989X20904955] [Medline: 32428426]
39. Lab W. VaxDA-C19 v5 demo. YouTube. 2025 Dec 22. URL: www.youtube.com/watch?v=lmb5olf-fDo [accessed 2026-03-24]
40. Witteman HO, Maki KG, Vaisson G, et al. Systematic development of patient decision aids: an update from the IPDAS collaboration. Med Decis Making 2021 Oct;41(7):736-754. [doi: 10.1177/0272989X211014163] [Medline: 34148384]
41. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. 1993 Presented at: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems; Apr 24-29, 1993. [doi: 10.1145/169059.169166]
42. Nielsen J. Why you only need to test with 5 users. Nielsen Norman Group. 2000. URL: https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ [accessed 2025-03-11]
43. Alwashmi MF, Hawboldt J, Davis E, Fetters MD. The iterative convergent design for mobile health usability testing: mixed methods approach. JMIR Mhealth Uhealth 2019 Apr 26;7(4):e11656. [doi: 10.2196/11656] [Medline: 31025951]
44. Joe J, Chaudhuri S, Le T, Thompson H, Demiris G. The use of think-aloud and instant data analysis in evaluation research: exemplar and lessons learned. J Biomed Inform 2015 Aug;56:284-291. [doi: 10.1016/j.jbi.2015.06.001] [Medline: 26071683]
45. Austrian J, Mendoza F, Szerencsy A, et al. Applying A/B testing to clinical decision support: rapid randomized controlled trials. J Med Internet Res 2021 Apr 9;23(4):e16651. [doi: 10.2196/16651] [Medline: 33835035]
46. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251. [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
47. Quin F, Weyns D, Galster M, Silva CC. A/B testing: a systematic literature review. J Syst Softw 2024 May;211:112011. [doi: 10.1016/j.jss.2024.112011]
48. Etienne D, Archambault P, Aziaka D, et al. A personalized avatar-based web application to help people understand how social distancing can reduce the spread of COVID-19: cross-sectional, observational, pre-post study. JMIR Form Res 2022;7:e38430. [doi: 10.2196/38430]
49. Hakim H, Bettinger JA, Chambers CT, et al. A web application about herd immunity using personalized avatars: development study. J Med Internet Res 2020 Oct 30;22(10):e20113. [doi: 10.2196/20113] [Medline: 33124994]
50. Yee N, Bailenson J. The Proteus Effect: the effect of transformed self-representation on behavior. Human Comm Res 2007 Jul;33(3):271-290. [doi: 10.1111/j.1468-2958.2007.00299.x]
51. Yee N, Bailenson JN, Ducheneaut N. The Proteus Effect: implications of transformed digital self-representation on online and offline behavior. Commun Res 2009 Jan;36(2):285-312. [doi: 10.1177/0093650208330254]
52. Zikmund-Fisher BJ, Witteman HO, Dickson M, et al. Blocks, ovals, or people? Icon type affects risk perceptions and recall of pictographs. Med Decis Making 2014 May;34(4):443-453. [doi: 10.1177/0272989X13511706] [Medline: 24246564]
53. Witt JK. An objective measure of decisional clarity to assess decision aid effectiveness in situations with equipoise: a randomized trial. Med Decis Making 2022 Aug;42(6):822-831. [doi: 10.1177/0272989X221085489] [Medline: 35297721]
54. Evergreen S. Icon array. Better Evaluation. URL: https://www.betterevaluation.org/methods-approaches/methods/icon-array [accessed 2025-03-10]
55. Rosenstock IM. Historical origins of the health belief model. Health Educ Monogr 1974 Dec;2(4):328-335. [doi: 10.1177/109019817400200403]
56. Wertheimer M. A brief introduction to Gestalt, identifying key theories and principles. Psychol Forsch 1923;4:301-350. [doi: 10.1007/BF00410640]
57. Mayer RE. Cognitive Theory of Multimedia Learning: Cambridge University Press; 1997.
58. Sorden SD. The cognitive theory of multimedia learning. In: Handbook of Educational Theories: Information Age Publishing; 2012.
59. Slovic P, Finucane ML, Peters E, Macgregor DG. The affect heuristic. In: The Construction of Preference:434-453. [doi: 10.1017/CBO9780511618031.024]
60. Slovic P, Peters E, Finucane ML, Macgregor DG. Affect, risk, and decision making. Health Psychol 2005 Jul;24(4S):S35-S40. [doi: 10.1037/0278-6133.24.4.S35] [Medline: 16045417]
61. Witteman HO, Chipenda Dansokho S, Exe N, Dupuis A, Provencher T, Zikmund-Fisher BJ. Risk communication, values clarification, and vaccination decisions. Risk Anal 2015 Oct;35(10):1801-1819. [doi: 10.1111/risa.12418] [Medline: 25996456]
62. Witteman HO, Ndjaboue R, Vaisson G, et al. Clarifying values: an updated and expanded systematic review and meta-analysis. Med Decis Making 2021 Oct;41(7):801-820. [doi: 10.1177/0272989X211037946] [Medline: 34565196]
63. Dionne M, Sauvageau C, Etienne D, Kiely M, Witteman H, Dubé E. Development of promising interventions to improve human papillomavirus vaccination in a school-based program in Quebec, Canada: results from a formative evaluation using a mixed methods design. JMIR Form Res 2024 Jul 8;8:e57118. [doi: 10.2196/57118] [Medline: 38976317]
64. Introduction. Vue.js. URL: https://vuejs.org/guide/introduction.html [accessed 2025-03-11]
65. Kim HM, Saffer AJ, Liu W, et al. How public health agencies break through COVID-19 conversations: a strategic network approach to public engagement. Health Commun 2022 Sep;37(10):1276-1284. [doi: 10.1080/10410236.2021.1886393] [Medline: 33591839]

XSL•FO
RenderX

66. Ford C, MacKay M, Thaivalappil A, McWhirter J, Papadopoulos A. COVID-19 vaccine communications on Instagram and vaccine uptake in young adults: a content assessment and public engagement analysis. Emerg Adulthood 2024 Apr;12(2):224-235. [doi: 10.1177/21676968231222439]
67. Dharma AAS. Exploring the communication, information, and education strategies of the Indonesian FDA for COVID-19 vaccine development. ERUDITIO - Indonesian J Food Drugs Saf 2024 May 2;4(1):17-32. [doi: 10.54384/eruditio.v4i1.179]
68. Saleem SM, Jan SS. Navigating the infodemic: strategies and policies for promoting health literacy and effective communication. Front Public Health 2023;11:1324330. [doi: 10.3389/fpubh.2023.1324330] [Medline: 38283287]
69. Cooper S, Gadanya MA, Kaawa-Mafigiri D, et al. Using social media to build confidence in vaccines: lessons from community engagement and social science research in Africa. BMJ 2024 Jan 16;384:e075564. [doi: 10.1136/bmj-2023-075564] [Medline: 38228329]
70. Enria L, Dwyer H, Marchant M, et al. Political dimensions of misinformation, trust, and vaccine confidence in a digital age. BMJ 2024 Jun 20;385:e079940. [doi: 10.1136/bmj-2024-079940] [Medline: 38901859]
71. An Overview of Risk Communication Communicating in Risk, Crisis, and High Stress Situations: Wiley; 2021:33-67. [doi: 10.1002/9781119081753]
72. Pang Z, Ma X. What drives the vaccination intention against COVID-19? Application of EPPM, TAM, and theories of risk assessment. Hum Vaccin Immunother 2023 Dec 31;19(1):1-13. [doi: 10.1080/21645515.2023.2180969] [Medline: 36919459]
73. Evergreen rapid review on COVID-19 vaccine attitudes and uptake in Canada: update 11. Public Health Agency of Canada. 2022. URL: https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/canadas-reponse/summaries-recent-evidence/evergreen-rapid-review-vaccine-attitudes-uptake-update-11.html [accessed 2024-09-16]
74. Fenta ET, Tiruneh MG, Delie AM, et al. Health literacy and COVID-19 vaccine acceptance worldwide: a systematic review. SAGE Open Med 2023;11:20503121231197869. [doi: 10.1177/20503121231197869] [Medline: 37823070]
75. Song I, Lee SH. COVID-19 vaccine refusal associated with health literacy: findings from a population-based survey in Korea. BMC Public Health 2023 Feb;23(1). [doi: 10.1186/s12889-023-15182-0]
76. Xiong C, Sarvghad A, Goldstein DG, Hofman JM, Demiralp Ç. Investigating perceptual biases in icon arrays. Presented at: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; Apr 30 to May 6, 2022. [doi: 10.1145/3491102.3501874]
77. Fansher M, Adkins TJ, Lalwani P, et al. Icon arrays reduce concern over COVID-19 vaccine side effects: a randomized control study. Cogn Res 2022 Dec;7(1):1-9. [doi: 10.1186/s41235-022-00387-5]
78. Capurro G, Maier R, Jardine C, Tustin J, Driedger SM. The good, the bad, and the ugly: a qualitative evaluation of web-based COVID-19 vaccine communication in Canada. Hum Vaccin Immunother 2025 Dec 31;21(1). [doi: 10.1080/21645515.2025.2515658]
79. Kiss A, Zhang Q, Carley M, et al. Quality of patient decision aids to support the public making COVID-19 decisions: an online environmental scan. Patient Educ Couns 2023 Sep;114:107797. [doi: 10.1016/j.pec.2023.107797] [Medline: 37244134]
80. Vincent YM, Frachon A, Allaire A, et al. Development of a patient decision aid for COVID-19 vaccination with the Comirnaty vaccine. Fam Pract 2022 May 28;39(3):486-492. [doi: 10.1093/fampra/cmab156]
81. Preuhs K, Bussink-Voorend D, van Keulen HM, et al. An online tailored COVID-19 vaccination decision aid for Dutch citizens: development, dissemination, and use. JMIR Form Res 2024 Oct 30;8:e56390. [doi: 10.2196/56390] [Medline: 39475719]
82. Decision aid: should I get the COVID-19 vaccine? NCIRS. URL: https://ncirs.org.au/decision-aid-should-i-get-covid-19-vaccine [accessed 2025-03-16]
83. Cui L, Lo S, Liu Z. The use of visualizations to improve Bayesian reasoning: a literature review. Vision (Basel) 2023 Mar 2;7(1):17. [doi: 10.3390/vision7010017] [Medline: 36977297]
84. Batanero C, Álvarez-Arroyo R. Teaching and learning of probability. ZDM Math Educ 2024 Feb;56(1):5-17. [doi: 10.1007/s11858-023-01511-5]
85. Bancilhon M, Wright A, Ha S, Crouser RJ, Ottley A. Why combining text and visualization could improve Bayesian reasoning: a cognitive load perspective. Presented at: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems; Apr 23-28, 2023. [doi: 10.1145/3544548.3581218]
86. Lee M, Oh B, Yoon NH, Kim S, Jung YI. The evaluation of web-based communication interventions to support decisions about COVID-19 vaccination among patients with underlying medical conditions: protocol for a randomized controlled trial. JMIR Res Protoc 2023 Jan 19;12:e42837. [doi: 10.2196/42837] [Medline: 36599054]
87. Anraad C, van Empelen P, Ruiter RAC, Rijnders M, van Groessen K, van Keulen HM. Promoting informed decision making about maternal pertussis vaccination: the systematic development of an online tailored decision aid and a centering-based group antenatal care intervention. Front Public Health 2024 Feb;12. [doi: 10.3389/fpubh.2024.1256337]
88. Rager JB, Schmidt KK, Schwartz PH. Discordant care and decision quality: patients' reasons for not receiving their initial test of choice in colorectal cancer screening. Med Decis Making 2024 Aug;44(6):705-714. [doi: 10.1177/0272989X241262278] [Medline: 39056287]
89. Yenew C, Dessie AM, Gebeyehu AA, Genet A. Intention to receive COVID-19 vaccine and its health belief model (HBM)-based predictors: a systematic review and meta-analysis. Hum Vaccin Immunother 2023 Dec 31;19(1):2207442. [doi: 10.1080/21645515.2023.2207442] [Medline: 37170620]

XSL•FO
RenderX

90. Malecki KMC, Keating JA, Safdar N. Crisis communication and public perception of COVID-19 risk in the era of social media. Clin Infect Dis 2021 Feb 16;72(4):697-702. [doi: 10.1093/cid/ciaa758] [Medline: 32544242]
91. Communicating with patients about COVID-19 vaccination. World Health Organization. URL: https://iris.who.int/bitstream/handle/10665/340751/WHO-EURO-2021-2281-42036-57837-eng.pdf [accessed 2026-03-10]
92. Zikmund-Fisher BJ, Witteman HO, Fuhrel-Forbis A, Exe NL, Kahn VC, Dickson M. Animated graphics for comparing two risks: a cautionary tale. J Med Internet Res 2012 Jul 25;14(4):e106. [doi: 10.2196/jmir.2030] [Medline: 22832208]
93. Bonner C, Trevena LJ, Gaissmaier W, et al. Current best practice for presenting probabilities in patient decision aids: fundamental principles. Med Decis Making 2021 Oct;41(7):821-833. [doi: 10.1177/0272989X21996328] [Medline: 33660551]
94. Canada Facebook users by age 2025. Statista. URL: https://www.statista.com/statistics/863754/facebook-user-share-in-canada-by-age/ [accessed 2025-04-15]
95. Bruel S, Leclercq T, Ginzarly M, Botelho-Nevers E, Frappé P, Gagneux-Brunon A. Patient decision aid in vaccination: a systematic review of the literature. Expert Rev Vaccines 2020 Apr;19(4):305-311. [doi: 10.1080/14760584.2020.1742111] [Medline: 32163307]
96. Efficacy and safety of tozorakimab in patients hospitalised for viral lung infection requiring supplemental oxygen (TILIA). ClinicalTrials.gov. 2026 Mar 18. URL: https://clinicaltrials.gov/study/NCT05624450 [accessed 2026-03-24]

Abbreviations

FAQ: frequently asked question
SRQR: Standards for Reporting Qualitative Research
UX: user experience

Edited by J Sarvestan; submitted 22.Oct.2025; peer-reviewed by LS Boland, O Ajayi; revised version received 16.Jan.2026; accepted 20.Jan.2026; published 16.Apr.2026.

Please cite as:
Etienne D, Archambault PM, Bogoch II, Chambers CT, Chittle AD, Demers J, Driedger SM, Dubé È, Gagnon MP, Gavaruzzi T, Giguère A, Grandvaux N, Grindrod K, Hakim H, Jeimy S, Kindrachuk J, LeBlanc A, MacDonald SE, Ndjaboue R, Noubi M, Orji R, Paquette JS, Parent E, Renaud JS, Sander B, Taljaard M, Tannenbaum Greenberg D, Tremblay MC, Vohra-Miller S, Welch VA, Witteman HO
Interactive, Personalized Patient Decision Aid for COVID-19 Vaccination in Canada: User-Centered Design Approach
JMIR Hum Factors 2026;13:e86283
URL: https://humanfactors.jmir.org/2026/1/e86283
doi:10.2196/86283

© Doriane Etienne, Patrick M Archambault, Isaac I Bogoch, Christine T Chambers, Andrea D Chittle, Juliette Demers, S Michelle Driedger, Ève Dubé, Marie-Pierre Gagnon, Teresa Gavaruzzi, Anik Giguère, Nathalie Grandvaux, Kelly Grindrod, Hina Hakim, Samira Jeimy, Jason Kindrachuk, Annie LeBlanc, Shannon E MacDonald, Ruth Ndjaboue, Magniol Noubi, Rita Orji, Jean-Sébastien Paquette, Elizabeth Parent, Jean-Sébastien Renaud, Beate Sander, Monica Taljaard, Dana Tannenbaum Greenberg, Marie-Claude Tremblay, Sabina Vohra-Miller, Vivian A Welch, Holly O Witteman. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 16.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Enhancing Understanding and Acceptance of Equipment Localization: Mixed Methods Study With Clinic Staff and Potential Patients

Angela Fiedler, MSc, Dr rer pol; Sabine Patzl, MSc; Melissa Schütz, MSc; Astrid Schütz, Prof Dr

Department of Psychology, University of Bamberg, Markusplatz 3, Bamberg, Germany

Corresponding Author:
Angela Fiedler, MSc, Dr rer pol
Department of Psychology, University of Bamberg, Markusplatz 3, Bamberg, Germany

Abstract

Background: Digital technologies, such as equipment localization systems, can help clinics use mobile devices more efficiently. Their successful implementation, however, depends not only on technical feasibility but also on how staff and patients perceive and understand these systems.

Objective: This research used 2 complementary studies to (1) obtain an initial picture of clinic staff attitudes toward the localization of vacuum-assisted closure (VAC) pumps and related concerns and (2) examine whether a simple layout change in a privacy policy (using guiding questions vs standard text) is associated with greater subjective understanding and acceptance among potential patients.

Methods: In study 1, 38 employees of a German clinic completed a short survey assessing their comfort with and perceived usefulness of VAC pump localization and answered an open-ended question about reservations or concerns. Quantitative responses were analyzed descriptively, and free-text answers were coded using qualitative content analysis. In study 2, 498 participants from an online sample took part in a preregistered experiment. They were randomly assigned to read either a standard privacy policy information sheet or an otherwise identical version supplemented with guiding questions. Subjective understanding of the information and acceptance of the policy were then assessed and analyzed using rank-based regression models controlling for sociodemographic covariates.

Results: Clinic staff in study 1 generally reported high levels of comfort (mean 7.34, SD 2.75) and perceived usefulness (mean 7.29, SD 2.69) regarding localization on 0 - 10 scales. Concerns centered mainly on implementation feasibility, technical reliability, costs, and possible additional workload, rather than on privacy. In study 2, subjective understanding was slightly higher in the guiding-question layout condition than in the standard layout condition (mean 3.37, SD 0.63 [n=248] vs mean 3.24, SD 0.68 [n=250]); this difference was also significant in the rank-based regression model (b=0.13, SE=0.05, t=2.57; P=.01), and better understanding was associated with higher acceptance of the policy, explaining about 13.8% of the variance in acceptance scores.

Conclusions: The exploratory findings suggest that, in the context of VAC pump localization, clinic staff generally view equipment tracking positively while still raising practical concerns that should be addressed during implementation. For potential patients, relatively small changes in the layout of privacy information—such as adding guiding questions—may support subjective understanding and willingness to accept data collection.

(JMIR Hum Factors 2026;13:e79583) doi:10.2196/79583

KEYWORDS

digital health; informed consent; health care staff; understanding; implementation; patient privacy; real-time location systems

Introduction

In large clinical settings, medical equipment must remain mobile to ensure that resources are available at the right place and time. However, locating mobile assets often requires time-consuming coordination and manual documentation, resulting in inefficiencies. Digital solutions—particularly real-time location systems (RTLSs), which are local infrastructures that identify and track the position of tagged objects or people via beacons distributed throughout the location of interest—offer promising avenues for improving workflows and resource allocation by providing real-time data on equipment use and availability [1-3].

One clinical example in which location-based data collection may be especially valuable involves vacuum-assisted closure (VAC) pumps, which are widely used in wound therapy. Although VAC therapy is well established, the manual process of tracking, cleaning, and redistributing pumps poses logistical challenges. When the stock is low, staff often need to call other units or request short-term rentals. Real-time localization could support more efficient allocation by enabling automated

XSL•FO
RenderX

documentation, identifying device availability, and reducing cost as well as administrative burden [4].

Although the potential benefits of RTLSs include improved workflow, increased safety, and reduced cost [5-7], their successful implementation depends on more than technical infrastructure. On the one hand, health care workers are more likely to adopt new technologies if the benefits are clearly understood and they are actively involved in the implementation process [8]. Conversely, concerns about increased workload, privacy, and surveillance may hinder adoption [9,10].

On the other hand, patients' privacy concerns will be more salient if location data can be linked to individuals. Thus, from a patient perspective, systems that track movement raise critical questions about data security, consent, and control over personal information [11]. Beyond data protection in a narrow sense, RTLS infrastructures can also enable fine-grained monitoring of individuals' movements, which has raised broader concerns about surveillance, autonomy, and potential changes in workplace and care relationships in hospital and long-term care settings [9-11].

Although privacy policies and consent forms are designed to soothe privacy concerns and enable informed decisions, studies have consistently found that these documents are often too complex and that patients struggle to comprehend them—especially when they are presented in dense, text-heavy formats [12,13]. Health communication research has indicated that the layout of consent materials can significantly affect comprehension. Structured documents with headings, bullet points, and guiding questions have been shown to improve readability and facilitate understanding [14,15]. Because document layout directly shapes user trust, informed decision-making, and willingness to give consent, thoughtful designs are pivotal for digital solutions (such as RTLSs) in health contexts.

This research investigated the perspectives of both clinic staff and patients on location-based data collection in health care, using the example of VAC pump localization. In study 1, we administered a survey to assess clinic staff's perceived comfort with and perceived usefulness of device localization and to identify any concerns related to its implementation. In study 2, we used an experimental design to test whether improving the layout of a privacy policy—featuring a guiding-question format as opposed to plain text—could enhance potential patients' subjective understanding and whether a better understanding, in turn, could predict greater acceptance. Together, the 2 studies were designed to inform how digital health tools can be introduced in ways that are transparent, user-centered, and conducive to adoption. Rather than offering a broad theoretical advance, this research provides exploratory, context-specific evidence that links RTLS implementation concerns with consent and privacy-communication design in the concrete case of VAC pump localization.

Methods

Study 1

Participants

In a clinic in southern Germany, 2 wards were selected for recruitment, and volunteers were invited to participate. The final sample consisted of 38 clinic employees. Participants had a mean age of 33.18 (SD 11.92) years and reported working in the clinic for an average of 13.40 (SD 11.66) years. Most participants identified as female (n=27), followed by male (n=7), and a small number of participants preferred not to disclose their gender (n=4). Regarding professional roles, the majority worked in nursing (n=32), while the remaining participants reported working in other roles (n=3), in wound management (n=2), or as a physician (n=1).

Ethical Considerations

This study was conducted in accordance with the ethical principles of the Declaration of Helsinki and was approved by the Ethics Committee of the University of Bamberg, Germany (dossier/review number: 2023-11/23). All participants provided informed consent prior to participation. Participation was voluntary, and participants could discontinue at any time without penalty. Participants were informed about data handling during the consent process. Privacy and confidentiality were protected throughout the study; no directly identifying personal data are reported in this manuscript; study data were analyzed and reported in aggregate form only; and any contact information collected for the raffle was stored separately from the survey responses. As compensation, participants were offered the opportunity to win 1 of 10 gift vouchers worth €10 (US $11.70) each for the city of Bamberg, Germany.

Procedure and Measures

Before data collection, we obtained consent from the clinic's work council. To minimize the disruption of routine clinical workflows, a printed questionnaire was distributed in 2 wards, where staff could complete it whenever their schedules permitted. The completed questionnaires were collected in sealed and nontransparent boxes to ensure anonymization. The survey remained available for a total of 29 days, from March to April 2024.

Attitudes toward the localization of VAC pumps were assessed with 2 self-developed items. Perceived comfort with localization was measured with the item: "How comfortable do you feel about the idea of localizing VAC pumps?" Responses were provided on an 11-point scale ranging from 0 (very uncomfortable) to 10 (very comfortable). Perceived usefulness was assessed with the item: "How helpful do you think the localization of VAC pumps would be?" Participants responded on the same 11-point scale ranging from 0 (not helpful at all) to 10 (very helpful). To minimize the burden during routine clinical work, these brief items were used as exploratory, descriptive indicators (not validated scales).

In addition, participants were asked an open-ended question: "Please describe any doubts or concerns you have about the localization of items or the project as a whole." Responses were

XSL•FO
RenderX

analyzed using qualitative content analysis. Thematic coding was conducted by the research team to group responses on the basis of content similarity.

Because of item-level missing responses, sample sizes vary slightly across study 1 outcomes; we report outcome-specific n values based on available cases for each item.

Study 2

Preregistration

The analysis was preregistered on the Open Science Framework, and an anonymized version can be accessed by following the link in the references under files [16]. Any deviations from the preregistration are transparently reported and justified.

Participants

Participants were recruited online via the following three approaches: (1) snowball sampling, (2) calls posted in the University of Bamberg's student forums, and (3) an announcement distributed through the university's mailing list to individuals from the general population who had signed up to be informed about studies. As an incentive, participants were offered the opportunity to win one of forty €10 (US $11.70) gift vouchers for the city of Bamberg, Germany. Alternatively, psychology students could receive course credit for completing the study. A total of 498 individuals completed the online survey between November 2024 and February 2025. Of these, 308 identified as female, 182 as male, 6 as diverse, and 2 preferred not to disclose their gender. The mean age was 41.12 (SD 18.9) years. Regarding educational background, 38 participants reported basic education, 258 reported intermediate education, and 199 reported higher education, while 3 observations were missing. As this sample size exceeded the originally preregistered sample size of 310, we drew a random subsample of 155 female and 155 male participants for an additional analysis as a robustness check. Participants in the subsample had a mean age of 41.45 (SD 18.64) years. Regarding educational background, 23 participants reported basic education, 150 reported intermediate education, and 136 reported higher education. One participant did not provide information on their education.

Outlier Detection and Exclusion Criteria

To ensure data quality, careless responding was assessed using the relative speed index (RSI), which was calculated as the inverse of each participant's total survey completion time relative to the group average. Participants with an RSI≥3, indicating exceptionally fast completion times, were to be excluded. However, no participants exceeded this threshold; thus, no exclusions were made based on the RSI.

Ethical Considerations

This study was conducted in accordance with the ethical principles of the Declaration of Helsinki and was approved by the Ethics Committee of the University of Bamberg, Germany (dossier/review number: 2025-05/39). All participants provided informed consent prior to participation. Participation was voluntary, and participants could discontinue at any time without penalty. Participants were informed about data handling during the consent process. Privacy and confidentiality were protected throughout the study; no directly identifying personal data are reported in this manuscript; study data were analyzed and reported in aggregate form only; and any contact information collected for the raffle was stored separately from the survey responses. As compensation, participants were offered the opportunity to win 1 of 40 gift vouchers worth €10 (US $11.70) each for the city of Bamberg, Germany. Psychology students could alternatively receive course credit for completing the study.

Measures

Subjective Understanding

Subjective understanding was assessed with 5 self-developed items rated on a 4-point Likert scale (1=strongly disagree, 4=strongly agree), including sample items such as "I feel that I understood the points listed." We deliberately chose a 4-point scale without a neutral midpoint to discourage "safe" or noncommittal responses and to elicit clearly directional judgments about participants' perceived understanding, while also keeping the response format cognitively simple in this applied setting. Responses were aggregated into a single composite score. Internal consistency was high (Cronbach α=0.86, 95% CI 0.837-0.877; McDonald ω=0.86, 95% CI 0.812-0.911). An exploratory factor analysis confirmed unidimensionality, with all factor loadings ranging from 0.68 to 0.81—well above the predefined threshold of 0.40.

Acceptance

Acceptance was measured with a single item: "How likely is it that you would agree to this privacy policy?" Participants responded on a slider scale ranging from 1 (representing 0%) to 101 (representing 100%). For the analyses, the scale was rescaled to range from 0 to 100.

Design

Participants were randomly assigned to one of two types of information sheets. The two versions presented identical content; however, one version included guiding questions for each paragraph (eg, "What data are processed when the vacuum pumps are tracked?" and "Who has access to your data?") to support comprehension. The information sheets (in German) were also uploaded to the Open Science Framework along with this preregistration. Participants were then asked to report their subjective understanding of the information sheet content and their acceptance. To test hypothesis 1—that including guiding questions in the layout of a privacy policy information sheet enhances subjective understanding as compared with a standard layout—we conducted a rank-based regression analysis. The dependent variable was subjective understanding, and the independent variable was the treatment condition (standard vs improved layout). Gender, age, education, and reading time were included as covariates to control for potential confounds.

To test hypothesis 2—that increased subjective understanding would lead to greater acceptance of the privacy policy—we conducted a rank-based regression analysis. Acceptance served as the dependent variable, and subjective understanding was included as the key predictor. Gender, age, and education were entered as covariates to control for potential confounding effects.

XSL•FO
RenderX

To further examine the relationship between layout, subjective understanding, and acceptance, we conducted an exploratory mediation analysis. We tested whether the effect of the layout condition (standard vs guiding questions) on acceptance was mediated by subjective understanding. Subjective understanding was modeled as the mediator and acceptance as the outcome. Both variables were analyzed in their original metric, with acceptance ranging from 0 to 100. We specified two linear regression models. In the mediator model, subjective understanding was regressed on layout condition, controlling for gender, age, education, and reading time. In the outcome model, acceptance was regressed on layout condition and subjective understanding, controlling for gender, age, and education (reading time was not included here to remain consistent with the main regression specification for acceptance). Indirect (average causal mediation effect [ACME]), direct (average direct effect [ADE]), and total effects were estimated using the mediation package in R with nonparametric bootstrap CIs based on 5000 simulations. Because this analysis relies on linear modeling and was not preregistered, we report it as exploratory and interpret it cautiously, using it to complement (but not replace) the rank-based regression analyses.

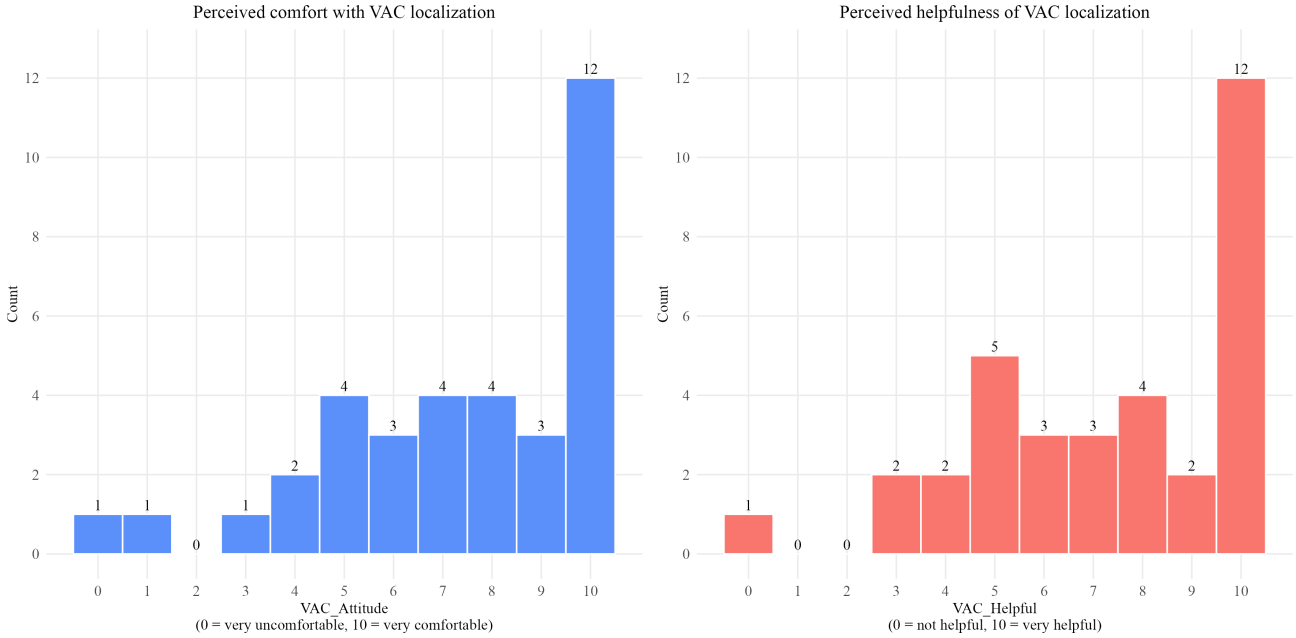
Results

Study 1

Perceived Comfort and Usefulness of VAC Pump Localization

Participants generally expressed positive attitudes toward the localization of VAC pumps. On a scale ranging from 0 (very uncomfortable) to 10 (very comfortable), the average level of comfort with the idea of localization was 7.34 (SD 2.75; n=35), see Figure 1.

Figure 1. Staff perceptions of vacuum-assisted closure (VAC) pump localization. The histograms show the distributions of responses regarding perceived comfort (left) and perceived helpfulness (right) of localizing VAC pumps. Responses were given on a scale ranging from 0 to 10, with higher values indicating greater comfort or helpfulness.



Similarly, the perceived usefulness of localization was rated positively (mean 7.29, SD 2.69; n=34) on a scale ranging from 0 (not helpful at all) to 10 (very helpful). A visual inspection of the distributions indicated a slight positive skew, with the most frequent response being the highest value (10) for both items.

Staff Concerns Regarding Localization in Clinical Contexts

A total of 18 statements describing concerns about the localization of items in clinical practice were categorized across 5 themes. The most frequently mentioned concern was related to implementation feasibility (n=7), followed by technical doubts (n=3). Other participants questioned whether localization would be helpful, specifically for VAC pumps (n=2). Additional concerns included increased costs (n=2) and added workload for nursing staff (n=2).

Study 2

Layout and Subjective Understanding

We hypothesized that including guiding questions in the layout of a privacy policy information sheet would enhance subjective understanding compared with a standard layout. A rank-based regression analysis was conducted with subjective understanding as the dependent variable; treatment condition (standard vs improved layout) as the predictor; and gender, age, education, and reading time as covariates. The primary analysis was conducted on the full sample, and the preregistered random subset consistent with the preregistered sample size was analyzed as a robustness check.

In the full sample, the analysis revealed a significant effect of treatment (t=2.57; P=.01), indicating higher subjective understanding in the improved layout condition compared with the standard layout (Table 1). Age was a significant positive predictor (t=3.91; P<.001), indicating that older participants

XSL•FO
RenderX

reported a higher subjective understanding. No significant effects were found for gender, education, or reading time. The model explained 3.7% of the variance in subjective understanding (R^2=0.037), and the reduction in dispersion test was significant (P=.006), indicating that the model provided a better fit than an intercept-only model. As a robustness check, the same analysis in the preregistered subset showed a marginal effect of treatment (t=1.93; P=.06) and a significant positive effect of age (t=1.99; P=.048); no significant effects were found for gender, education, or reading time. The model explained 2.5% of the variance in subjective understanding (R^2=0.025), and the reduction in dispersion test was not significant (P=.27).

Table . Rank-based regression predicting subjective understanding[a].

Predictor	Estimate (SE)	t	P value
Full sample (n=498)			
Intercept	3.1539 (0.1216)	25.93	<.001
Treatment	0.1273 (0.0495)	2.57	.01
Male	−0.0076 (0.0525)	−0.15	.89
Age	0.0054 (0.0014)	3.91	<.001
Education (intermediate)	−0.0011 (0.0967)	−0.01	.99
Education (higher)	0.0274 (0.0971)	0.28	.78
Reading time	-9×10^{-6} (7×10^{-6})	−1.28	.20
Preregistered random gender-balanced subsample (n=310)			
Intercept	3.3496 (0.1513)	22.14	<.001
Treatment	0.1185 (0.0615)	1.93	.06
Male	0.012 (0.0631)	0.19	.85
Age	0.0034 (0.0017)	1.99	.04
Education (intermediate)	−0.1176 (0.1223)	−0.96	.34
Education (higher)	−0.0869 (0.1217)	−0.71	.48
Reading time	-9×10^{-6} (7×10^{-6})	−1.25	.21

[a]Subjective understanding was predicted using rank-based regression. Treatment was coded as 0=standard and 1=improved layout. Reading time was measured in seconds. For the treatment comparison, subjective understanding in the full sample was a mean of 3.24 (SD 0.68; n=250) in the standard layout condition and a mean of 3.37 (SD 0.63; n=248) in the improved layout condition. In the preregistered gender-balanced subsample, the corresponding values were mean 3.26 (SD 0.64; n=160) and mean 3.40 (SD 0.61; n=150), respectively. Reported t values are coefficient test statistics from the rank-based regression model.

Subjective Understanding and Acceptance

To examine whether higher subjective understanding would support greater acceptance of the privacy policy (hypothesis 2), we conducted a rank-based regression with acceptance as the dependent variable and subjective understanding as the key predictor. Gender, age, and education were included as covariates. Again, the primary analysis was conducted on the full sample; the preregistered random subset consistent with the preregistered sample size was analyzed as a robustness check.

The full-sample model revealed a significant positive effect of subjective understanding on acceptance (Table 2; t=11.33; P<.001), indicating that participants who reported better understanding of the privacy policy also expressed a greater willingness to accept it. No significant effects were found for gender, age, or education (all Ps>.32). The model explained approximately 13.8% of the variance in acceptance (robust R^2=0.138), and the reduction in dispersion test was significant (P<.001), suggesting that the model provided a better fit than an intercept-only model. As a robustness check, the corresponding analysis in the preregistered subset likewise showed a significant positive effect of subjective understanding on acceptance (t=8.90; P<.001), whereas gender, age, and education were not significant. The model explained approximately 16% of the variance in acceptance (robust R^2=0.16), and the reduction in dispersion test was significant (P<.001).

XSL•FO
RenderX

Table . Rank-based regression predicting acceptance[a].

Predictor	Estimate (SE)	*t*	*P* value
Full sample			
Intercept	39.58 (5.03)	7.86	<.001
Understanding	13.32 (1.17)	11.33	<.001
Male	1.62 (1.63)	0.99	.32
Age	−0.01 (0.04)	−0.21	.84
Education (intermediate)	0.07 (2.99)	0.02	.98
Education (higher)	−0.46 (3.00)	−0.15	.88
Preregistered random gender-balanced subsample			
Intercept	40.36 (6.29)	6.42	<.001
Understanding	12.82 (1.44)	8.90	<.001
Male	0.57 (1.86)	0.31	.76
Age	0.05 (0.05)	1.06	.29
Education (intermediate)	−0.73 (3.58)	−0.20	.84
Education (higher)	0.49 (3.56)	0.14	.89

[a]Acceptance of the privacy policy was predicted using rank-based regression. Subjective understanding was entered as a continuous predictor.

Exploratory Mediation Analysis

In an exploratory step, we examined whether the effect of layout (standard vs guiding questions) on acceptance was mediated by subjective understanding. Using linear regression models with nonparametric bootstrapping, the indirect effect of layout on acceptance via subjective understanding (ACME) was small but statistically significant (ACME=2.26, 95% CI 0.64-4.32; *P*=.006) on a 0 - 100 acceptance scale. The direct effect of layout on acceptance when controlling for subjective understanding (ADE) was not statistically significant (ADE=–1.48, 95% CI –6.68 to 3.66; *P*=.58), and the total effect of layout on acceptance was also nonsignificant (total effect=0.79, 95% CI –4.67 to 6.13; *P*=.78). Overall, this pattern of effects is consistent with an indirect pathway from layout to acceptance via subjective understanding, with no evidence of a remaining direct effect when subjective understanding is included.

Discussion

Study 1

Given the single-site convenience sample, our findings should be interpreted as context-specific and descriptive. The results of study 1 indicate that the staff appeared to be open to the idea of localizing VAC pumps to improve workflows and support them in their daily work. Their concerns were predominantly technical and infrastructural in nature, with no explicit articulation of privacy concerns about the localization of VAC pumps. This finding is encouraging, as privacy concerns have been identified as a potential hindrance to the successful implementation of such systems [9,10]. Nevertheless, it is essential that patients agree to the localization, as it involves their personal data. Ensuring ethical and informed consent necessitates not only patients' acceptance of these systems but also their comprehension of the relevant information provided to them. Consequently, study 2 was designed to gain an understanding of the conditions under which potential patients can subjectively comprehend and accept information about VAC pump localization, with a particular emphasis on the impact of the guiding-question format compared with a plain text format.

Study 2

Our results provide support for the hypothesis that incorporating guiding questions into privacy policy layouts enhances subjective understanding. This aligns with research showing that simplified and well-structured consent materials—using plain language, headings, and clearer layouts—improve comprehension [14,17-19]. While the variance explained in subjective understanding was modest, this is consistent with the fact that we manipulated only a single structural feature of the information sheet and did not account for other plausible determinants of perceived understanding, such as general cognitive ability and health literacy [20].

Interestingly, age emerged as a significant predictor in both analyses, with older participants reporting a greater subjective understanding regardless of condition. Several explanations could account for older participants reporting a greater subjective understanding of privacy policies. One possible explanation could be that older people have had more experience with health care. Older participants may have more extensive experience with health care settings and procedures, making them more confident in their understanding of health care–related documentation, including privacy policies. Another possible explanation could be their familiarity with traditional documentation. Older adults might be more accustomed to reading formal documents, such as consent forms and privacy policies, in their original formats, whereas younger participants might expect more interactive or digital presentations (eg, videos or app-like interfaces) [21]. An additional reason might be a

XSL•FO
RenderX

self-reporting bias. The finding might reflect different calibrations in self-assessments rather than actual differences in understanding. Younger participants may be more critical of their own understanding, whereas older participants may overestimate their understanding in the sense of an illusion of explanatory depth [22]. Finally, the finding may have resulted from greater concerns about privacy among older adults [23]. Older participants might invest more attention in privacy-related content due to heightened privacy concerns, resulting in a better subjective understanding. This counterintuitive finding highlights the importance of testing information design approaches with diverse age groups rather than making assumptions about digital or information literacy based on age alone.

The data also supported our second hypothesis, demonstrating that higher subjective understanding significantly predicted greater acceptance. This relationship explained 13.8% of the variance in acceptance scores (robust R^2=0.138), and the corresponding analysis in the preregistered subset yielded a comparable pattern (robust R^2=0.16). This indicates that subjective understanding is a meaningful predictor of acceptance in this context. Conceptually, this aligns with work on data-intensive technologies—such as crowd-counting systems in smart cities—where better explanations and clearer communication are associated with higher trust and acceptance [24]. In our study, the exploratory mediation analysis yielded a small but statistically significant indirect effect of layout on acceptance via subjective understanding, while the total effect of layout on acceptance was not statistically significant. This pattern is compatible with the interpretation that the layout manipulation primarily affects acceptance via perceived understanding rather than exerting an additional direct influence; however, given the exploratory nature of the mediation analysis and the reliance on self-reported understanding, this inference should be treated cautiously. In general, our results are consistent with broader evidence that design-driven improvements in understanding are only one part of a larger "privacy calculus," in which perceived benefits, trust, and risk perceptions jointly shape acceptance [25,26].

Building on study 1, which examined health care providers' perceptions of VAC pump localization systems, in study 2, we shifted our focus to how structural elements of privacy policies affect user comprehension and acceptance. While study 1 highlighted clinical staff concerns, study 2 provided evidence that an improved layout can positively influence how patients perceive privacy information and, via subjective understanding, relate to acceptance judgments. At the same time, study 2 assessed subjective understanding rather than objective comprehension, and acceptance was measured as a single-item willingness judgment; thus, future studies should test whether similar layout manipulations improve performance-based comprehension and decision-relevant outcomes.

General Discussion

Principal Findings

We examined attitudes toward RTLS-based equipment localization in health care, focusing on the example of VAC pump localization in 1 German clinic. Across 2 complementary studies, we explored how clinical staff perceive such technologies and how information sheet layout may influence patient understanding and acceptance.

Study 1 provided initial insights into the perspectives of clinic employees. Overall, participants expressed generally positive attitudes toward the idea of VAC pump localization, reporting high levels of comfort and perceived usefulness. These findings are consistent with prior work on RTLSs, which emphasizes perceived usefulness, workflow support, and safety gains as key drivers of staff acceptance [3,8]. At the same time, some employees voiced concerns about technical feasibility, implementation logistics, and the potential for an increased workload—issues that echo previous evaluations of RTLSs in hospital contexts, where organizational readiness and human-factors considerations often determine whether technically viable systems are used consistently in practice [3,10].

Study 2 experimentally tested whether the layout of a privacy policy influences subjective understanding and its association with acceptance. Controlling for gender, age, education, and reading time, a layout incorporating guiding questions significantly improved subjective understanding compared with a standard format. These results align with health-communication research, demonstrating that structured consent documents with headings, bullet points, and guiding prompts improve readability and comprehension, especially when baseline literacy is heterogeneous [13,14,17,19]. Among the covariates, only age showed a small positive association with subjective understanding, indicating that older participants felt slightly more confident in their understanding (as discussed above). Moreover, higher subjective understanding was associated with greater acceptance of the policy. These findings are consistent with prior evidence showing that structured consent materials—such as those using headings and guiding prompts—can improve comprehension and user engagement [14,15].

Taken together, the findings underscore the dual importance of content and context in the introduction of data-driven technologies in clinical environments. While technical feasibility and staff buy-in are essential for implementation success, the way information is presented also plays a critical role in shaping patients' perceptions and willingness to engage with such systems. The studies provide complementary perspectives on the implementation of location-tracking technologies in health care settings. Study 1 revealed the nuanced attitudes and operational concerns of frontline staff toward VAC pump localization, and study 2 demonstrated how information sheet layout can impact patients' understanding and acceptance of such systems. This complementarity is particularly valuable because even though the staff members in study 1 were very positive about the implementation of a localization system, patients might express concerns about their privacy, which study 2 then addressed by testing communication approaches that could mitigate these very concerns. Even though the material that participants in study 2 received was addressed to patients, it might be just as helpful for staff to understand the system. For example, the guiding-question format, which improved subjective comprehension among potential patients, might also

XSL•FO
RenderX

be beneficial for addressing the operational concerns expressed by the clinic staff in study 1. Applying structured communication approaches in training sessions and staff information materials may help mitigate concerns about technical feasibility and implementation logistics. By improving staff members' understanding of how the localization system works and what benefits it offers, their acceptance and willingness to use the system may increase. Greater staff acceptance, in turn, may enable more consistent use in everyday practice and thereby enhance efficiency in clinical workflows—for example, by reducing the time needed to locate equipment. In addition, confident and well-informed staff are likely to convey more trust to patients, thereby positively shaping patients' perceptions and acceptance of the system. Thus, integrating the insights from both studies into a comprehensive communication strategy might simultaneously address staff concerns and patient privacy considerations, thereby facilitating more effective implementation of RTLS technologies in health care settings. Together, the 2 studies provide a more comprehensive picture than either study could have offered in isolation—combining a rich contextual understanding with experimental evidence on how to address key implementation barriers.

Limitations

Several limitations should be noted. The sample size in study 1 was relatively small and limited to a single clinical institution, thereby potentially reducing the generalizability of the findings. Given the voluntary nature of participation, staff who held stronger concerns may have been less likely to respond, which could have biased attitudes in a more positive direction. Moreover, attitudes toward VAC pump localization were assessed with 2 single self-developed items (comfort and perceived usefulness), which precludes an assessment of internal consistency and may have reduced measurement reliability. Because these items were not validated and we did not estimate a measurement model, the study 1 quantitative results should be interpreted as descriptive indicators in this specific setting rather than as precise or generalizable estimates of staff acceptance. We chose this brief format to keep the survey feasible in a busy clinical environment and obtain an initial indication of staff attitudes, but future studies should use multi-item scales and larger, multisite samples to provide more robust and generalizable evidence. Furthermore, the acceptance of object localization may depend on the specific type of item that is being tracked. While VAC pumps were seen as unproblematic, other objects—particularly those more closely associated with staff or patient movement—may raise greater concerns, including fears of increased surveillance or performance monitoring, which our data do not capture [9-11]. Generalization to other use cases should therefore be approached with caution.

In study 2, the measures relied on self-reported understanding and a single-item acceptance judgment rather than behavioral outcomes. Additionally, while the layout manipulation in study 2 targeted structural design, other features, such as language complexity and institutional trust, were not varied and may also influence understanding [13]. Moreover, participants' subjective understanding was the sole focus of the assessment. Consequently, the study lacks the necessary data to determine whether participants' comprehension was genuinely enhanced by the improved explanation or whether they merely perceived an improvement in understanding. Recent work in the context of location tracking in a different setting has shown that objective understanding, assessed with comprehension questions, can be more strongly related to acceptance—via increased project-specific trust—than subjective understanding alone [24]. Against this background, our reliance on subjective understanding may have attenuated the observed associations and could partly explain why the estimated indirect effect of layout on acceptance via understanding was small but statistically significant. Future studies should therefore combine subjective ratings with performance-based comprehension measures and, where possible, behavioral indicators of information use and technology acceptance.

Practical Implications

The results point to several actionable recommendations. First, incorporating clear, question-based formatting into privacy communications can support perceived understanding, as evidenced by study 2. Such formatting could be implemented by redesigning consent forms, digital interfaces, and educational materials throughout the health care facility.

Second, early engagement with clinic personnel—especially those responsible for operational logistics—is essential to address concerns and increase buy-in. This engagement should include structured opportunities for staff to provide feedback during all implementation phases, from initial planning to postdeployment evaluation. Implementation could include comprehensive training programs that not only focus on technical operation but also provide context about how the technology integrates with existing workflows and benefits patient care. Such information is especially important given that staff voiced concerns about potential increases in workload in study 1.

Future Directions

Future research could build on these findings in several ways. Longitudinal studies may help assess whether improved understanding leads to sustained trust and actual use over time. In addition, experimental variations in communication elements—such as video-based or interactive explanations—could shed more light on how to optimize privacy-related materials for clinical audiences [5]. Critically, future studies should directly compare subjective understanding with objective comprehension to determine whether layout- or format-driven changes translate into demonstrable knowledge gains. Building on these initial directions, future research should examine whether the age-related differences we observed in subjective understanding translate into actual differences in comprehension. Studies incorporating brief knowledge assessments (eg, multiple-choice or true or false items targeting key policy elements such as what data are collected, who can access them, and retention periods) alongside self-reported measures would provide valuable insights into this question, as earlier research found differences between objective and subjective understanding and their relationships with acceptance [24]. Where feasible, such work should also incorporate

XSL•FO
RenderX

decision-relevant outcomes (eg, consent decisions or information-seeking behavior) rather than relying solely on self-report.

Additionally, comparative research across different health care institutions could help identify how contextual variables, such as organizational culture and existing technology infrastructure, influence staff attitudes toward location-tracking technologies. Such research would expand upon our single-institution findings and provide a more generalizable understanding of implementation factors.

Conclusion

As health care systems continue to adopt digital technologies, understanding how such innovations are communicated and perceived is becoming increasingly important. Within the specific context of VAC pump localization in one clinic and a hypothetical patient scenario, this research highlights 2 practical dimensions of successful implementation: the design of communication materials and the perspectives of clinical staff who are expected to engage with new systems. We showed that simple layout adjustments can significantly improve subjective understanding and are indirectly related to acceptance via perceived understanding. In our clinic sample, staff respondents generally welcome innovations such as equipment localization when they are presented transparently and aligned with clinical needs. At the same time, implementation must account for practical concerns on the ground. Together, these findings underscore the value of integrating human-centered design principles—in both communication and system planning—to ensure that digital innovations are not only functional but are also trusted and embraced by those who use them.

Acknowledgments

During manuscript preparation, we used ChatGPT (OpenAI) to assist with language editing, phrasing, and preparation of R code. All artificial intelligence (AI)–generated or AI-assisted output was carefully reviewed, revised, and verified by the authors. The authors remain fully responsible for the accuracy, originality, and integrity of the final manuscript, including all references and citations.

Funding

The authors declare that financial support was received for the research, authorship, and/or publication of this study. The project on which this report is based was funded by the German Federal Ministry of Research, Technology and Space under the funding code 16KISA079K. The responsibility for the content of this publication lies with the authors. The project was also funded by the European Union-NextGenerationEU. The views and opinions expressed are solely those of the authors and do not necessarily reflect the views of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them. The publication of this study was supported by the Open Access publication fund of the University of Bamberg.

Data Availability

The datasets generated and analyzed for our study are available from the authors upon request.

Authors' Contributions

AF, SP, MS, and AS conceptualized the study. AF, SP, and MS conducted data curation and analysis. AF wrote the main manuscript text and prepared the tables and figures. SP, MS, and AS substantively revised the manuscript. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Kamel Boulos MN, Berry G. Real-time locating systems (RTLS) in healthcare: a condensed primer. Int J Health Geogr 2012 Jun 28;11(1):25. [doi: 10.1186/1476-072X-11-25] [Medline: 22741760]
2. Motamedi A, Soltani MM, Hammad A. Localization of RFID-equipped assets during the operation phase of facilities. Adv Eng Inform 2013 Oct;27(4):566-579. [doi: 10.1016/j.aei.2013.07.001]
3. Overmann KM, Wu DTY, Xu CT, Bindhu SS, Barrick L. Real-time locating systems to improve healthcare delivery: a systematic review. J Am Med Inform Assoc 2021 Jun 12;28(6):1308-1317. [doi: 10.1093/jamia/ocab026] [Medline: 33682009]
4. Regolini J, Frosini F, Ciagli E, et al. A simple room localization method to find technology in a big trauma center. Presented at: World Congress on Medical Physics and Biomedical Engineering 2018; Jun 3-8, 2018; Prague, Czech Republic p. 321-325. [doi: 10.1007/978-981-10-9035-6_58]
5. Kim MH, Park YS. Use of a real-time locating system in infection control. Infect Chemother 2024 Dec;56(4):427-431. [doi: 10.3947/ic.2024.0043] [Medline: 39431344]

XSL•FO
RenderX

6. Morissette N, Shabah A. CHRONOS project: the transformation of time-dependent clinical trajectories into intelligent ones using an innovative technological solution. Prehosp Disaster Med 2023 May;38(S1):s49-s50. [doi: 10.1017/S1049023X23001620]
7. Ng GY, Ong BC. Contact tracing using real-time location system (RTLS): a simulation exercise in a tertiary hospital in Singapore. BMJ Open 2022 Oct 3;12(10):e057522. [doi: 10.1136/bmjopen-2021-057522] [Medline: 36192104]
8. Waldie P, Attef M, Trudel C, Chan ADC, Wallace B. Implementation and use of real-time location systems in hospital environments – rapid scoping review of human factors considerations. Presented at: Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care; Mar 26-29, 2023; Orlando, Florida, USA p. 164-168. [doi: 10.1177/2327857923121038]
9. Fisher JA, Monahan T. Evaluation of real-time location systems in their hospital contexts. Int J Med Inform 2012 Oct;81(10):705-712. [doi: 10.1016/j.ijmedinf.2012.07.001] [Medline: 22857790]
10. Grigorovich A, Kulandaivelu Y, Newman K, et al. Factors affecting the implementation, use, and adoption of real-time location system technology for persons living with cognitive disabilities in long-term care homes: systematic review. J Med Internet Res 2021 Jan 20;23(1):e22831. [doi: 10.2196/22831] [Medline: 33470949]
11. Shen N, Bernier T, Sequeira L, et al. Understanding the patient privacy perspective on health information exchange: a systematic review. Int J Med Inform 2019 May;125:1-12. [doi: 10.1016/j.ijmedinf.2019.01.014] [Medline: 30914173]
12. Siberry V, Ratwani R, Wesley D. Mobile health app privacy policies: what patients want vs. understand. Presented at: Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care; Mar 20-23, 2022; New Orleans, Louisiana, USA p. 156-161. [doi: 10.1177/2327857922111031]
13. Simonds VW, Buchwald D. Too dense and too detailed: evaluation of the health literacy attributes of an informed consent document. J Racial Ethn Health Disparities 2020 Apr;7(2):327-335. [doi: 10.1007/s40615-019-00661-1] [Medline: 31823337]
14. Bühn S, Huppertz E, Weise A, et al. The effects of modifying elements of written informed consent forms for elective surgical or invasive procedures: a systematic review. Patient Educ Couns 2023 Feb;107:107576. [doi: 10.1016/j.pec.2022.107576] [Medline: 36455317]
15. Geier C, Adams RB, Mitchell KM, Holtz BE. Informed consent for online research-is anybody reading?: Assessing comprehension and individual differences in readings of digital consent forms. J Empir Res Hum Res Ethics 2021 Jul;16(3):154-164. [doi: 10.1177/15562646211020160] [Medline: 34029168]
16. Trust and understanding location data sharing in clinical context. Open Science Framework (OSF). URL: https://osf.io/gj97m/files/osfstorage?view_only=047079787d3744fe904f81d73ebb5a7f [accessed 2025-06-17]
17. Kim EJ, Kim SH. Simplification improves understanding of informed consent information in clinical trials regardless of health literacy level. Clin Trials 2015 Jun;12(3):232-236. [doi: 10.1177/1740774515571139] [Medline: 25701156]
18. Nebeker C, Gholami M, Kareem D, Kim E. Applying a digital health checklist and readability tools to improve informed consent for digital health research. Front Digit Health 2021;3:690901. [doi: 10.3389/fdgth.2021.690901] [Medline: 34713167]
19. Coleman E, O'Sullivan L, Crowley R, et al. Preparing accessible and understandable clinical research participant information leaflets and consent forms: a set of guidelines from an expert consensus conference. Res Involv Engagem 2021 May 18;7(1):31. [doi: 10.1186/s40900-021-00265-2] [Medline: 34006326]
20. Ownby RL, Acevedo A, Goodman K, Caballero J, Waldrop-Valverde D. Health literacy predicts participant understanding of orally-presented informed consent information. Clin Res Trials 2015;1(1):15-19. [doi: 10.15761/CRT.1000105] [Medline: 26767117]
21. Kassam I, Ilkina D, Kemp J, Roble H, Carter-Langford A, Shen N. Patient perspectives and preferences for consent in the digital health context: state-of-the-art literature review. J Med Internet Res 2023 Feb 10;25(1):e42507. [doi: 10.2196/42507] [Medline: 36763409]
22. Rozenblit L, Keil F. The misunderstood limits of folk science: an illusion of explanatory depth. Cogn Sci 2002 Sep 1;26(5):521-562. [doi: 10.1207/s15516709cog2605_1] [Medline: 21442007]
23. Harris MT, Blocker KA, Rogers WA. Older adults and smart technology: facilitators and barriers to use. Front Comput Sci 2022;4. [doi: 10.3389/fcomp.2022.835927]
24. Waclawek T, Fiedler A, Schütz M, Schütz A. Crowd-counting technology within the Smart City context: understanding, trust, and acceptance. Front Psychol 2024;15:1423837. [doi: 10.3389/fpsyg.2024.1423837] [Medline: 39624374]
25. Fox G, Clohessy T, van der Werff L, Rosati P, Lynn T. Exploring the competing influences of privacy concerns and positive beliefs on citizen acceptance of contact tracing mobile applications. Comput Human Behav 2021 Aug;121:106806. [doi: 10.1016/j.chb.2021.106806]
26. Frangopoulou MS, van der Laan LN, Ebbers W. The privacy calculus in the context of novel health technology for diagnosing and tracking infectious diseases: the role of disease severity and technology's evidence base for effectiveness in adoption and voluntary health data-sharing. Technol Soc 2024 Sep;78:102616. [doi: 10.1016/j.techsoc.2024.102616]

Abbreviations

ACME: average causal mediation effect
ADE: average direct effect

XSL•FO
RenderX

RSI: relative speed index
RTLS: real-time location system
VAC: vacuum-assisted closure

Edited by A Kushniruk; submitted 24.Jun.2025; peer-reviewed by MBR Mahmoud, R Marshall; revised version received 11.Feb.2026; accepted 21.Feb.2026; published 27.Apr.2026.

Please cite as:
Fiedler A, Patzl S, Schütz M, Schütz A
Enhancing Understanding and Acceptance of Equipment Localization: Mixed Methods Study With Clinic Staff and Potential Patients
JMIR Hum Factors 2026;13:e79583
URL: https://humanfactors.jmir.org/2026/1/e79583
doi:10.2196/79583

© Angela Fiedler, Sabine Patzl, Melissa Schütz, Astrid Schütz. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 27.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Learning From the Adoption of a Readmission Clinical Decision Support Tool: Group Model Building Approach

Nina Rachel Sperber[1,2], PhD; Sarah Elizabeth Haas[1], PhD; Jiaxin Gao[3], MS; Samantha Hamelsky[3]; Theresa Kiki-Teboum[3]; Afraaz Malick[3]; Rishab Pulugurta[3], BS; Jacqueline Rodriguez[3]; Hana Shafique[4], BA, BS; Eden Singh[4], BA; Kriti Vasudevan[3]; Scott Rockart[5], PhD; David Gallagher[6], MD; Adam Johnson[2,7], MPH, MD

[1]Department of Population Health Sciences, Duke University School of Medicine, Box 90046, Durham, NC, United States

[2]Durham VA Health Care System, Durham, NC, United States

[3]Trinity College of Arts & Sciences, Duke University, Durham, NC, United States

[4]Duke University School of Medicine, Durham, NC, United States

[5]The Fuqua School of Business, Duke University, Durham, NC, United States

[6]Department of Medicine, Duke University School of Medicine, Durham, NC, United States

[7]Division of Vascular and Endovascular Surgery, Department of Surgery, Duke University School of Medicine, Durham, NC, United States

Corresponding Author:
Nina Rachel Sperber, PhD
Department of Population Health Sciences, Duke University School of Medicine, Box 90046, Durham, NC, United States

Abstract

Background: Computerized clinical decision support (CDS) has the potential to improve patient outcomes by offering evidence-based guidance at the point of care—enhancing guideline adherence and diagnostic accuracy—and supporting system-level outcomes by enabling predictive analytics for more efficient resource planning. Prior work has identified factors that affect adoption, such as clinicians' expectations of usefulness, ease of use, alignment with workflows, and resources to support utilization. However, CDS adoption is not static and changes according to dynamic systems of behaviors and workflows, requiring a deeper understanding of how evolving conditions affect implementation and outcomes.

Objective: To explore the dynamic factors influencing CDS adoption, we examined the implementation of the "Unplanned readmission model version 1," developed by Epic Medical Records System, at Duke University Health System, using group model building and system dynamics modeling.

Methods: We first conducted group model-building workshops with staff (case managers, physical and occupational therapists, hospitalist faculty physicians, and resident physicians) who participate in decisions about discharging patients. Study team members guided participants to identify and connect variables in causal loop diagrams. We coded workshop transcripts in software designed for system dynamics analysis to identify themes, aggregated them into a causal loop diagram, and reviewed them with participants to converge on a common model. A team member applied equations to the pathways and tested data to simulate conditions leading to full, limited, or no adoption of a tool.

Results: We identified key balancing loops driven by external pressure (eg, Centers for Medicare & Medicaid Services penalties) that motivated initial adoption and reinforcing loops based on perceived internal benefits to sustain use. While institutional incentives led to early training and tool use, efforts declined due to staff turnover, competing priorities (eg, COVID-19), and workflow changes. Reinforcing loops emerged when staff described clinical utility, such as improved discharge planning and team communication. However, staff also suggested that these loops were often weak due to difficulty linking the use of the tool to outcomes in real time. Simulation modeling showed that while strong external pressure and rapid training led to initial success, interest in using the tool waned as workflows improved and readmission rates approached Centers for Medicare & Medicaid Services goals. When conflicting priorities were introduced, adoption stalled earlier, and fewer staff were trained. In contrast, when internal motivation was strengthened by reducing the amount of evidence needed to perceive success, individual interest remained high even as institutional attention declined, sustaining tool use and further reducing readmissions.

Conclusions: External pressure to improve can be a strong motivator for initial adoption, but in the face of conflicting demands for attention, it may fall short of sustained long-term tool use. Tools are more likely to have extensive and sustained use when those using the tools can perceive internal benefits.

(JMIR Hum Factors 2026;13:e87522) doi:10.2196/87522

XSL•FO
RenderX

KEYWORDS

clinical informatics; implementation science; risk prediction; clinical decision support; system science

Introduction

Computerized clinical decision support (CDS) tools, health IT systems that deliver patient-specific recommendations during clinical workflows, have the potential to transform health care delivery. By delivering evidence-based guidance at the point of care, these tools enhance clinical decision-making, improve patient outcomes, and enable health care organizations to anticipate future resource needs through predictive analytics, leading to more efficient planning and resource allocation [1-4]. A targeted literature review published in the *Interactive Journal of Medical Research* found that CDS systems positively impacted quality assurance in 69% of the studies analyzed and provided clinical benefits in 41% compared to usual care [5]. Quality assurance was supported through improved adherence to clinical guidelines, better diagnostic support, monitoring practices, and improved screening and treatments. Clinical benefits included improved diagnostic accuracy, clinical decision-making, and treatment selection. The effectiveness of CDS tools is closely linked to their successful adoption at the bedside.

Understanding the factors that influence clinicians' adoption of CDS tools is essential for designing effective implementation strategies. The Unified Theory of Acceptance and Use of Technology, developed in 2003, has identified 4 key concepts that could influence new information technology adoption: performance expectancy (belief that using the system will help one improve job performance), effort expectancy (ease associated with the system including individual perceptions and design features), social influence (perceived social pressure to use the system), and facilitating conditions (perceived organizational and technical support for system use) [6]. Lui et al [7] found that performance expectancy and effort expectancy were significant predictors of CDS adoption and that these beliefs depended on individuals having autonomy and agency. A meta-analysis that examined predictors of health care practitioners' intention to use artificial intelligence–enabled CDS through the lens of the Unified Theory of Acceptance and Use of Technology found that facilitating conditions influenced performance and effort expectancy and, in turn, intention to use artificial intelligence–enabled CDS [8]. A recent study by Wang et al [9] additionally demonstrated the importance of top management support, in which senior administrators recognize the importance of the information system and are actively engaged in its implementation, for adoption by positively shaping not only performance and effort expectancies but also social influence. Other work has shown that clinicians are more likely to adopt tools that they view as accurate and useful when these tools integrate with existing workflows and have resources available to act on the risk scores [10,11]. Conversely, tools perceived as irrelevant and interruptive to workflows have led to "alert fatigue" and low utilization rates [12,13]. While this work sheds light on how multilevel factors affect CDS adoption, there is additionally a need to understand the dynamics of the systems within which CDS functions, with behaviors and workflows that change over time; this information is important for developing strategies to sustain adoption [14].

Systems-aligned precision medicine describes CDS strategies that account for complex and dynamic systems of care, not only for the patient but also for the clinician engaged in decision-making [15]. To understand the dynamic complexity of factors that affect adoption, we developed a study to explore facilitators and barriers in the adoption of a single use case, the implementation of "Unplanned readmission model version 1," developed by Epic Medical Records System, at Duke University Health System using system dynamics modeling [16]. Readmissions have been identified as an essential quality indicator and linked to institutional financial incentives through the Centers for Medicare & Medicaid Services (CMS) Hospital Readmission Reduction Program [17,18]. Due to the significant impact on patients and institutional stakeholders, many readmission risk scores have been developed to target high-risk patients with an improved discharge process [19,20]. We engaged interest holders in developing a system dynamics model by conducting group-based modeling sessions with health system staff actively involved in the discharge process. Through these sessions, we aimed to uncover factors that inhibit or promote the use of the readmission risk score and identify potential solutions to enhance its adoption and effectiveness. After completing the workshops, we built a mathematical version of the model to explore the logic captured in the causal loop diagram. This type of model, known as a simulation model, provides an opportunity to determine if the loops can create the behaviors reported in the workshops and under what conditions they may create other, not yet observed, behaviors.

Methods

Ethical Considerations

The protocol was reviewed by the Duke Health institutional review board and determined to be exempt from further institutional review board review (Pro00118266). The exemption was granted under category 2, covering minimal-risk research involving survey or interview procedures, and was consistent with institutional policy and applicable federal regulations. Because the study was deemed exempt, written informed consent was not required; participants were provided with study information and informed that participation was voluntary. No direct identifiers were collected, and the data were managed in accordance with Duke Health policies on privacy and confidentiality. Participants did not receive compensation for participation.

Setting

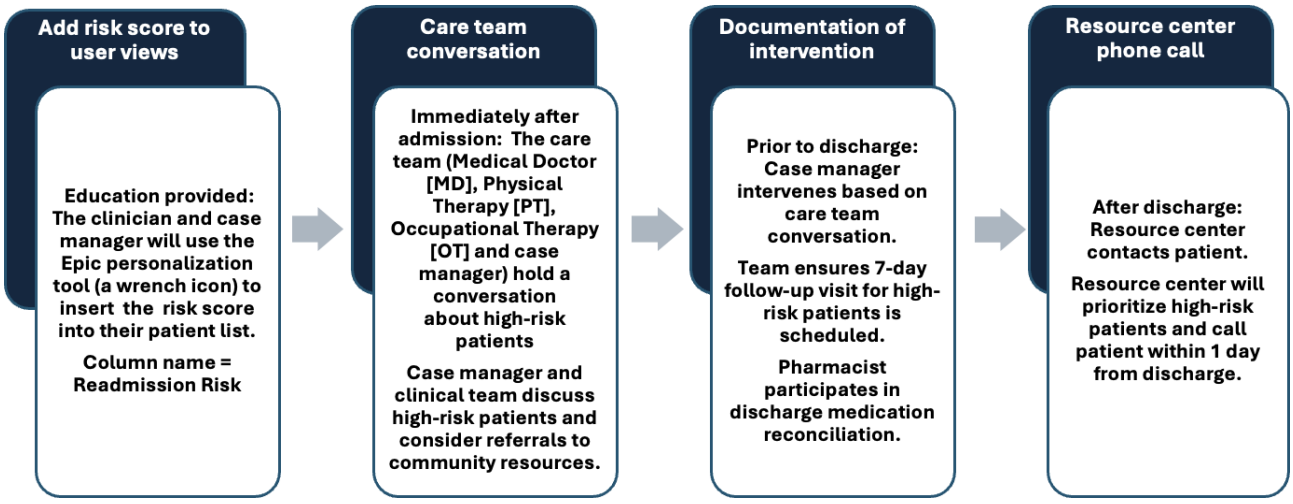
In 2017, Duke Health implemented Epic's "Unplanned readmission model version 1" (hereafter referred to as the Epic readmission risk score) in response to the need to reduce unplanned readmissions. CMS considers readmissions a key quality indicator and thus implemented the Hospital Readmissions Reduction Program to incentivize institutions to

XSL•FO
RenderX

reduce their readmission rates [18]. Duke University Hospital (DUH) is a quaternary academic hospital located in Durham, North Carolina, providing high-acuity care across specialties. The hospital serves a diverse patient population from Durham County, Wake County, and surrounding areas in North Carolina.

At a hospital like DUH, with almost 1200 beds and an average of more than 100 discharges a day, it would be impossible and unnecessary to apply all discharge resources to every patient. Hospitals like DUH need to stratify their time and resources based on patients who have the highest risk of being readmitted. The Epic readmission risk score was introduced at DUH in November 2017 to proactively identify patients at high risk of readmission and implement targeted interventions to reduce readmission rates [16]. This CDS tool analyzes Epic's electronic health record data to generate a readmission risk score, ranging from 0 to 100, along with corresponding color codes: red for high risk, yellow for medium risk, and green for low risk. Predictors such as patient age, laboratory variables, clinical diagnosis variables, medication numbers and classes, order types, and utilization variables are used to populate the score. This score and color code are then displayed within the patient's electronic health record, allowing clinicians to make informed decisions about postdischarge care and follow-up. Figure 1 shows the proposed integration into clinical workflows with multidisciplinary review at discharge and the implementation of close follow-up and additional services upon discharge.

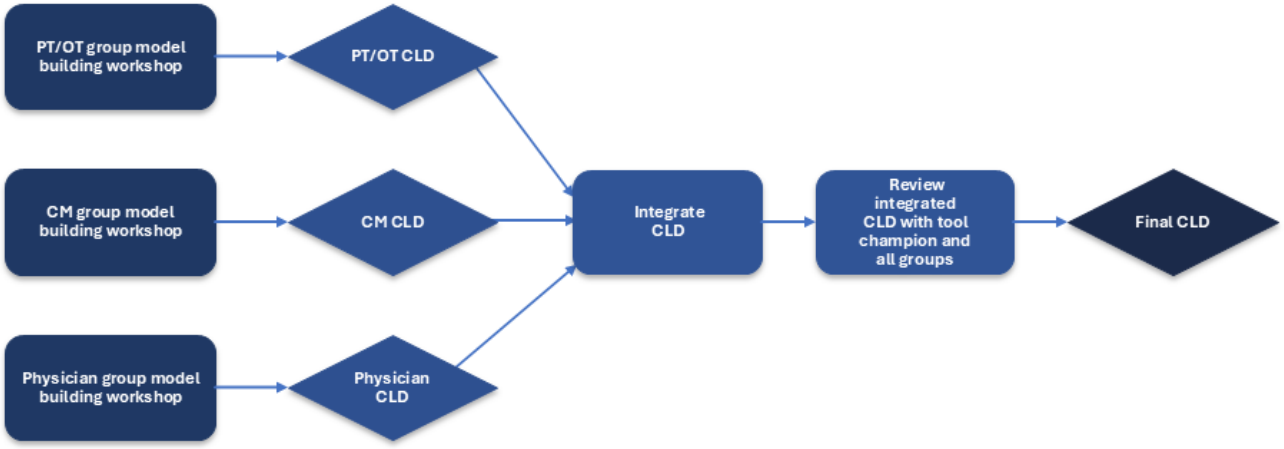
Figure 1. Integration of the Epic readmission risk score into clinical workflows used during training and the initial implementation.



Project Design

To explore adoption in this setting, we used group model building, a process that engages stakeholders in developing a system dynamics model. We performed 3 group model-building workshops with small groups of staff involved in the discharge process at DUH. These staff included case managers, physical and occupational therapists, and internal medicine and general surgery resident physicians. The workshops aimed to understand the factors influencing the uptake and sustained use of the Epic readmission risk score. In each session, the team facilitated the production of a causal loop diagram meant to represent the utilization of the Epic readmission risk score at the hospital, from the participants' perspectives. A final review session was held with members from all groups, as well as the chief medical officer, who initially led the implementation of the tool, and faculty hospitalist physicians. Figure 2 provides a visual representation of how our causal loop diagram was developed by the workshop participants and facilitators.

Figure 2. Study diagram. PT/OT: physical therapy/occupational therapy, CM: case manager, CLD: causal loop diagram.



Before the sessions, a script was created detailing the intended agenda of the workshops. The team adopted a process for variable elicitation and causal loop diagramming from *Scriptapedia*, a handbook of scripts for developing structured

XSL•FO
RenderX

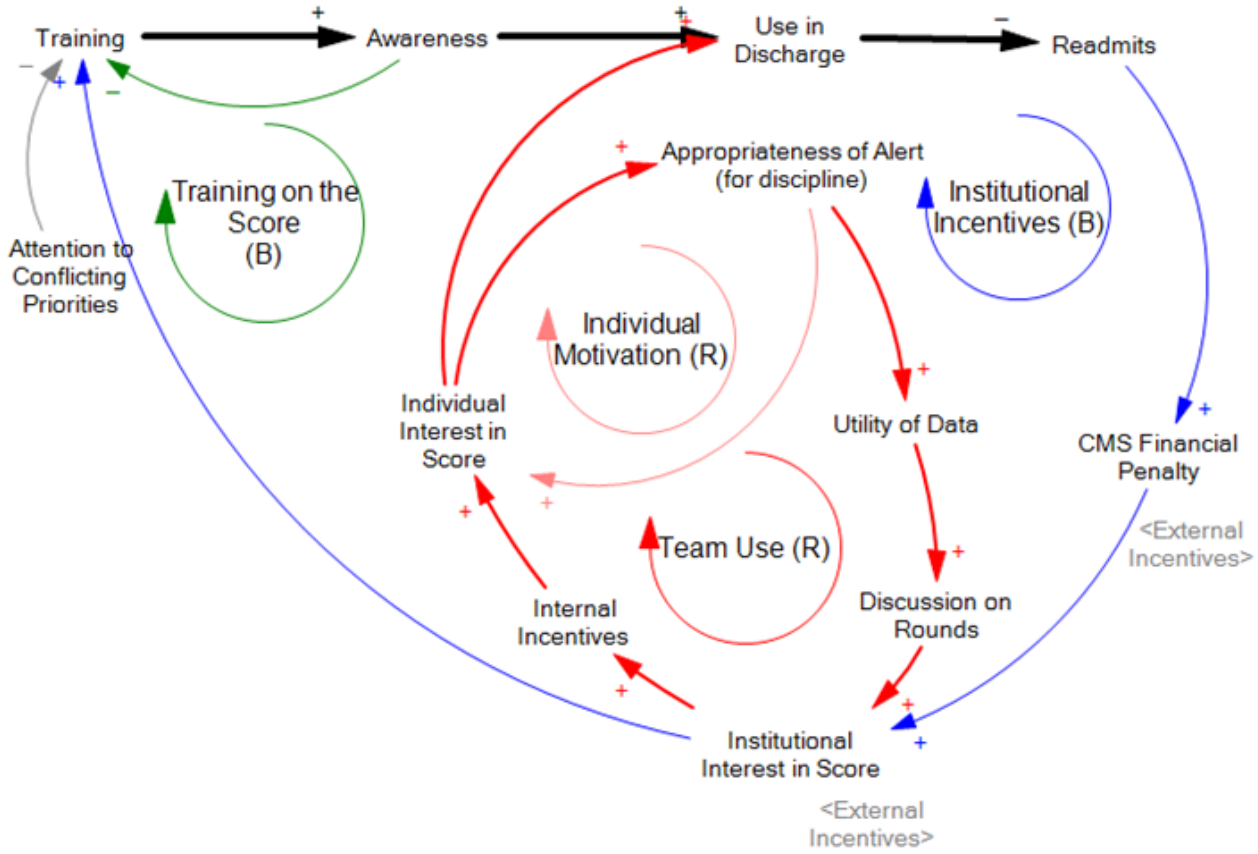
group model-building sessions [21]. In this structure of group modeling, there is a core modeling team that includes facilitators who lead the discussion, a recorder who monitors the audio recording while documenting important themes, and a "wall builder" who draws content on a whiteboard [22]. In these workshops, the facilitators (primary and secondary) first introduced themselves, the problem of declining use of the risk score, and the objective of constructing a causal loop diagram. The facilitators then prompted participants to describe their roles in the discharge process and discuss their experiences, or lack thereof, with the tool. The facilitators invited participants to contribute, by speaking or writing, suggestions about conditions that have affected tool use and worked with the group to identify variables and feedback loops. While the "wall builder" recorded the variables and diagram on a whiteboard, participants additionally had the opportunity to "take the pen" and contribute to the diagram directly. These workshops were transcribed, and the transcriptions were then coded in the DynamicVu (DynamicVu) software for key variables and causal links to match quotes from participants to parts of the causal loop diagram [23]. The study team worked together to integrate data from the 3 workshops into a single causal loop diagram, which was subsequently converted into a mathematical simulation model to assess internal validity, utilizing Vensim PLE (Ventana Systems, Inc). A final review session was conducted with a few members of each group, along with the chief medical officer at DUH, who led the initial implementation of the tool in 2017.

Results

Overview

We successfully completed the 3 interest holder workshops with the active participation of 5 to 10 clinicians per workshop. We then held a final group workshop with representatives from each group with active participation. As we explored one tool, we developed a set of dynamic hypotheses based on workshop participants' insights. The analysis revealed 3 key themes that explain the dynamics observed in the organization's efforts to lower the readmission rate. The three themes included institutional-level learning and reinforcement, individual-level learning and reinforcement, and the training and support of a champion (Figure 3).

Figure 3. Simplified causal loop diagram developed from the workshops directly. CMS: Centers for Medicare & Medicaid Services.



Organizational Response to Lower Readmission Rates

The first feedback loop identified was the organization's response to the external desire to lower the readmission rate ("Institutional Incentives" loop). This initiative to implement the Epic unplanned readmission score, driven by the institution's general concern for reducing readmissions, began prior to these modeling workshops. The tool was initially implemented through a pilot project that included an evaluation component led by a multidisciplinary team comprising members from hospital medicine, case management, clinical operations, biostatistics, and informatics, who encouraged training programs on the tool [16]. The hope was that this training would lead to increased use and thus better discharge plans for patients with a high risk of 30-day readmission. The initial push for tool use led to more training and awareness.

XSL•FO
RenderX

Our workshops led us to believe that while some staff still use the tool, many are unaware of its presence or capabilities, with usage generally declining over time rather than becoming widespread. We explored with staff potential explanations for the lack of widespread awareness and use of the tool, resulting in a balancing feedback loop, in which initial training that aimed at equipping staff with the necessary skills to achieve the desired outcomes dropped off, and other circumstances distracted attention away from continued training and use. We identified several reasons why training may have declined. These included normal turnover and events such as the COVID-19 pandemic, which placed other constraints on training. Additional changes, including arguably positive ones, changed the workflow by incorporating alternative practices and tools that may have been inspired by and substituted for the tool. The case managers described how the pilot program to introduce the score had changed their behavior by providing a visual cue to collect more information for specific patients indicated as "high risk" for readmission with a red alert. Some participants suggested there would be increased adoption if factors included in the algorithm aligned with their conceptual model of risk in these patients. Over time, however, they internalized the criteria associated with the alert, and those characteristics themselves became the cues. For example, one case manager said:

> *I think that the … [alert] started out like great in the beginning because it was the startup of a new system. So, it was like, hey, this is your visual to know that we have to do something. And maybe now we lost part of that visual because now we know what we're looking for. Practice had changed since the introduction of the score to collect additional information for patients flagged as "high risk" for readmission. … it just became part of our workflow of knowing in our assessment, like of room to assess more in depth, right, if that makes sense. It gave us the tools to then ask more in-depth questions, but then that became part of our assessment process.*

This organizational-level balancing loop indicates that, over time, as the readmission rate reaches a certain threshold, the institutional interest, driven by external pressure from the CMS penalty, will eventually drop [18]. The level of institutional interest, in turn, would affect individual staff members' interest in using the score. As one resident said, "… if the institution is saying we need to reduce readmissions at the end, I'm going to do everything I can to reduce readmissions ..."

Reinforcement Through Perceived Benefits

Over the long term, however, as the initiative transitions from the initial training phase to broader institutionalization, the efforts of these champions may be partly or wholly replaced by individual or organizational learning processes that reinforce tool use (see "Individual Motivation" and "Team Use" loops, which are reinforcing). These feedback loops emerged from the hope that staff would recognize the benefits of the training and the resulting improvements in readmission rates.

At both the individual and team levels, reinforcement can come from staff using the tools, observing the tool's appropriateness for their work, and becoming more committed to its use. Clinicians discussed how evidence of the effectiveness of the tool for their practice is an important driver for adoption. They indicated that they encounter many scores in their field and said that those they want to use "have been shown to be effective" and "provide obvious, clear direct value." In this case, they want to know what contributes to the score and "if any of that relates back to discharge." One resident explained:

> *It's on par with a new risk score where we don't actually know- like you're telling us that this is going to work or this is going to do this, but we have no backing or no evidence of that. So, I think that's another aspect to consider is like, yeah, we could look at it, but we don't actually know how much this makes a difference, how impactful this truly is.*

Staff said that they would value the data generated from the score if they contained information that would affect their decisions in real time. One physician said:

> *The only scores that we have time to spend on in medicine when we're running around 8000 times a day are the ones that directly affect our decision-making in that exact moment.*

They would then incorporate the information into the discussion during multidisciplinary rounds, which aligns with institutional interests. One case manager described a push for bringing multidisciplinary teams together at the same time as the introduction of the readmission risk tool:

> *There's been a push for pulling the multidisciplinary team together to improve communication. So, there's been lots of work all happening at the same time, and then we've got the hospital care hub ... part of the interdisciplinary ... case reviews and rounds.*

These reinforcing loops were based on the expectation that visible positive outcomes would motivate continued engagement and institutionalization of the practices introduced during the training. A medical doctor said:

> *If the institution is saying we need to reduce readmissions at the end, I'm going to do everything I can to reduce readmissions if I'm incentivized to do so … If using this helps, then I will keep using. So then that gets into clinical utility.*

A physical therapist suggested a similar sentiment, saying that their use would depend on evidence that the score had clinical utility for the kinds of factors that they assess to prevent discharge. For example, "I think it would be helpful if we knew a little bit more about … what's contributing to the score and if there's any … factors … that relate back to discharge ..."

Risk for Reinforcement

We additionally recognized areas of risk that could prevent this reinforcement from taking over. In our setting, the reinforcing loops could be weak and thus unable to sustain the program at a high level. The "Individual Motivation" loop may be weak because readmission is a rare event and often not directly managed by the discharging team, so team members are unlikely to see the immediate impact of their actions by observing fewer readmissions. In addition, the clear causal connection between

XSL•FO
RenderX

the recommended interventions and the intended outcome may be difficult to observe in real time. On the other hand, length of stay was much more directly observed by discharging clinicians, and efforts to reduce length of stay may come into conflict with efforts to reduce readmission, which can at times increase length of stay. Therefore, the natural reinforcing loops are weak, and external reinforcement through training and awareness campaigns is required. However, the ongoing work of champions and reinforcement can be disrupted in various ways, leading to a decline in tool awareness and use over time. As people change roles and take on new tasks, the ongoing active support of champions is likely to diminish, leading to a drop in training and awareness, especially as other changes crowd out efforts to support the use of the tool. Reinforcement relies on individuals being able to observe the tool's value in their work or, at higher administrative levels, recognizing the value of the tool and creating incentives. One medical doctor described how awareness of a tool would decline over time if it was perceived as inappropriate for their workflow and not easily integrated into the work of new team members:

> *... that institutional knowledge you have to maintain over those cycles. And that means that it has to be, like directly applicable, because I'm helping set up my interns who will then help set up someone for another four years, right. So, if like, if I don't know about it now, then no one... [will] four years from now.*

Additionally, participants expressed difficulty in observing the tool's value at an individual level for low-probability events (readmission is a relatively low probability event, making it harder to see the tool's signal). Resistance to broader changes in the workflow for how decisions are made could also impede the use of the tool. For example, participants observed barriers as the institution encouraged moving to a more consultative, multidisciplinary group process. As one medical doctor said:

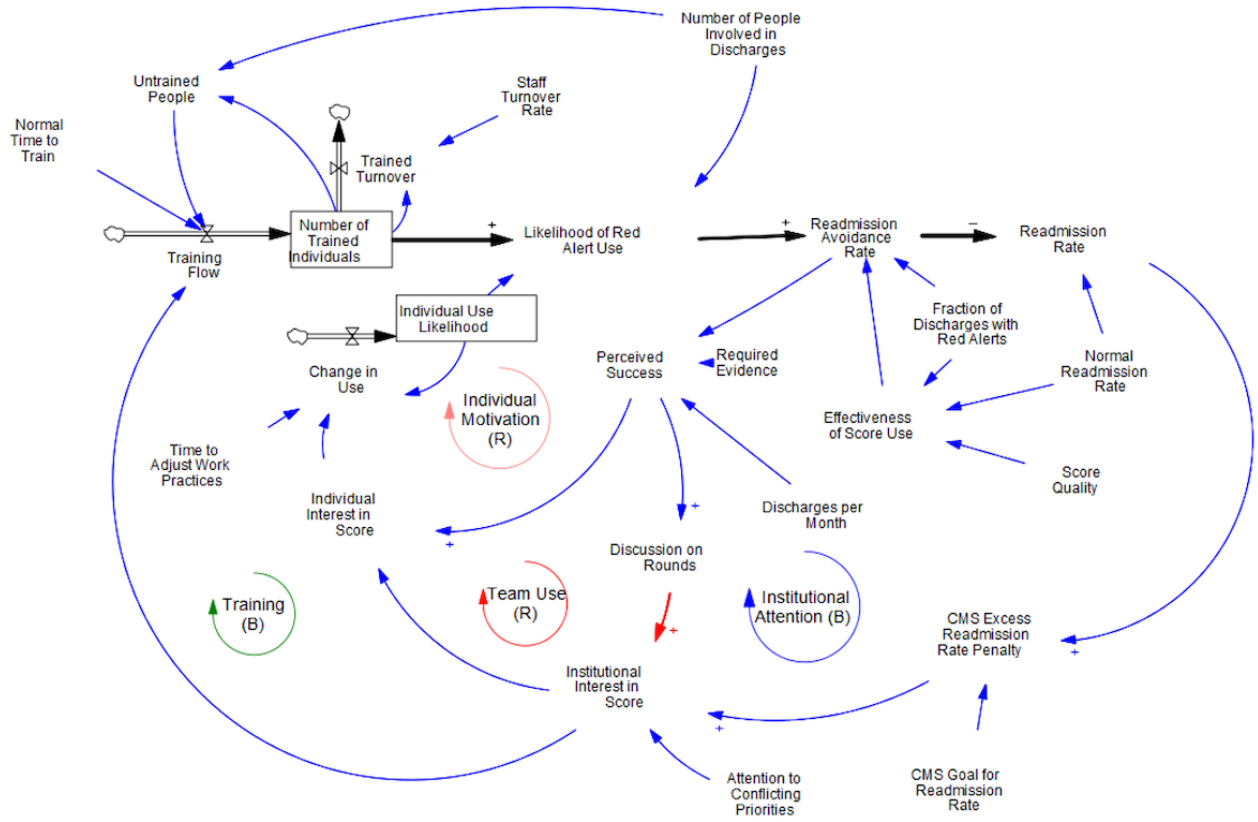
> *I think one thing is hierarchical decision-making is a barrier in the sense that like if the case managers or the residents are the target utilizers for the tool, but we're not the primary stakeholders in decision-making, then it's … kind of futile like I, I guess to make it very concrete. It may be difficult for individuals or organizations to recognize the value of the tools, and it may be challenging to find ways to ensure continued use.*

Internal Validity Test With Simulation Modeling

A mathematical simulation model was developed from the initial causal relationships (Figure 4; see Multimedia Appendix 1 for equations). The primary objective in building the mathematical model was to adhere as closely as possible to the causal-loop diagram. To that end, the equations leverage the simplest possible "accounting" or probabilistic logic wherever feasible. As an example of an "accounting" logic, "Awareness" is modeled as the "Number of Trained Individuals," with that number increasing when people are trained and decreasing when trained staff depart. As an example of probabilistic logic, the "Likelihood of Red Alert Use" is modeled as the fraction of people involved in discharges who are trained (the Number of Trained Individuals divided by the Number of People Involved in Discharges), multiplied by the fraction of discharges during which trained individuals pay attention to the score. The product of these 2 variables should capture the fraction of people involved in each discharge who are both trained to use the score and are actively looking at the score.

XSL•FO
RenderX

Figure 4. Simulation model derived from the causal loop diagram. CMS: Centers for Medicare & Medicaid Services.



The “Readmission Avoidance Rate” is central to the model, which captures the reduction in readmissions. This equation leverages probabilistic logic. The “Readmission Avoidance Rate” is the product of 3 parameters: the “Likelihood of Red Alert Use,” the “Fraction of Discharges with Red Alerts,” and the “Effectiveness of Score Use.” In the best possible case, 3 factors are true. First, every red alert is given full attention (the Likelihood of Red Alert Use equals 1). Second, everyone and only those who would later be readmitted are given a red alert (the Fraction of Discharges with Red Alerts equals the Normal Discharge Rate with no false positives). Third, attention to the score leads to interventions that are entirely effective in avoiding readmission (the Effectiveness of Score Use equals 1). If these conditions are true, the “Readmission Avoidance Rate” will equal the “Normal Readmission Rate,” and readmissions will drop to zero.

Understanding this logic around avoiding readmissions takes us most of the way to understanding how the model captures workshop themes of balancing institutional incentives and training on the score loops and the 2 reinforcing loops of individual interest and team use. As derived from the workshop discussions, the 2 institutional level loops are strong initially, when readmissions are high enough to drive institutional interest in the score, leading to training and individual attention to the score. In the simulation model, the training loop is a balancing loop because if training is successful, the use of the tool will lead to lower readmissions and thus the penalty to fall, resulting in a drop in institutional interest and, with it, training. The simulation model similarly recognizes that a reduction in the financial penalty reduces institutional interest in and attention to the score with a similar logic, except that, rather than the reduced readmission rate leading to a fall in training, it leads to a fall in individual interest in the score.

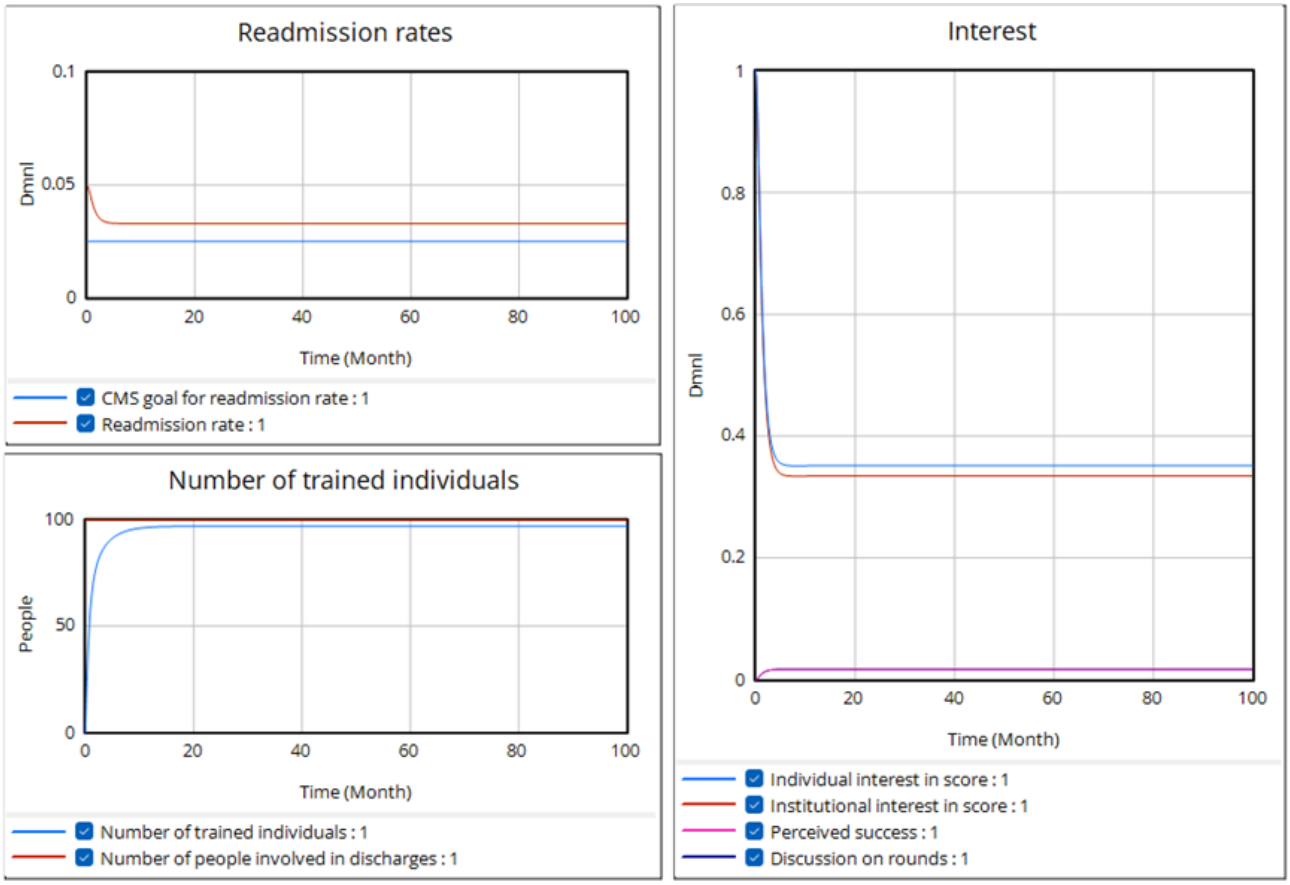
We grounded the connection between the readmission rate and institutional interest in the score using the CMS “Excess Readmission Rate Penalty.” When institutional readmission rates exceed the CMS goal, CMS payments to institutions decrease, driving institutional interest in tools (in this case the score) that can lower the rate to below the CMS goal for the Readmission Rate. For simplicity, we have assumed that this penalty depends on the fractional amount by which the institution’s readmission rate exceeds the CMS goal (This is both a simplified calculation of the CMS penalty and a change in wording from the CMS original word “score” to “penalty” to avoid using the word “score” for multiple different ideas in the model.).

We can see the behavior of these 2 balancing loops in a run of the model (Figure 5) set, so that these are strong loops, and the other (reinforcing) loops are weak. Here, training is quick (all untrained people are trained within 1 month), and turnover is low (only 1% per month), so the number of qualified individuals can rise quickly and, if so, stay high. Red alerts are set to match the regular readmission rate with no false positives, so attention to the score always leads to avoiding readmission. In addition, we have assumed that the institution has little need to address conflicting priorities. With this favorable set of parameters, the

XSL•FO
RenderX

high initial readmission rate relative to the CMS goal leads to high institutional interest that drives training and individual interest in the score. With trained people using the score regularly, the readmission rate falls rapidly. The number of qualified individuals never quite rises to equal the number of individuals involved in discharges because of the delay in training as staff depart.

Figure 5. Simulation run with strong balancing loops. The equations and parameters for this run are given in Multimedia Appendix 1. CMS: Centers for Medicare & Medicaid Services.

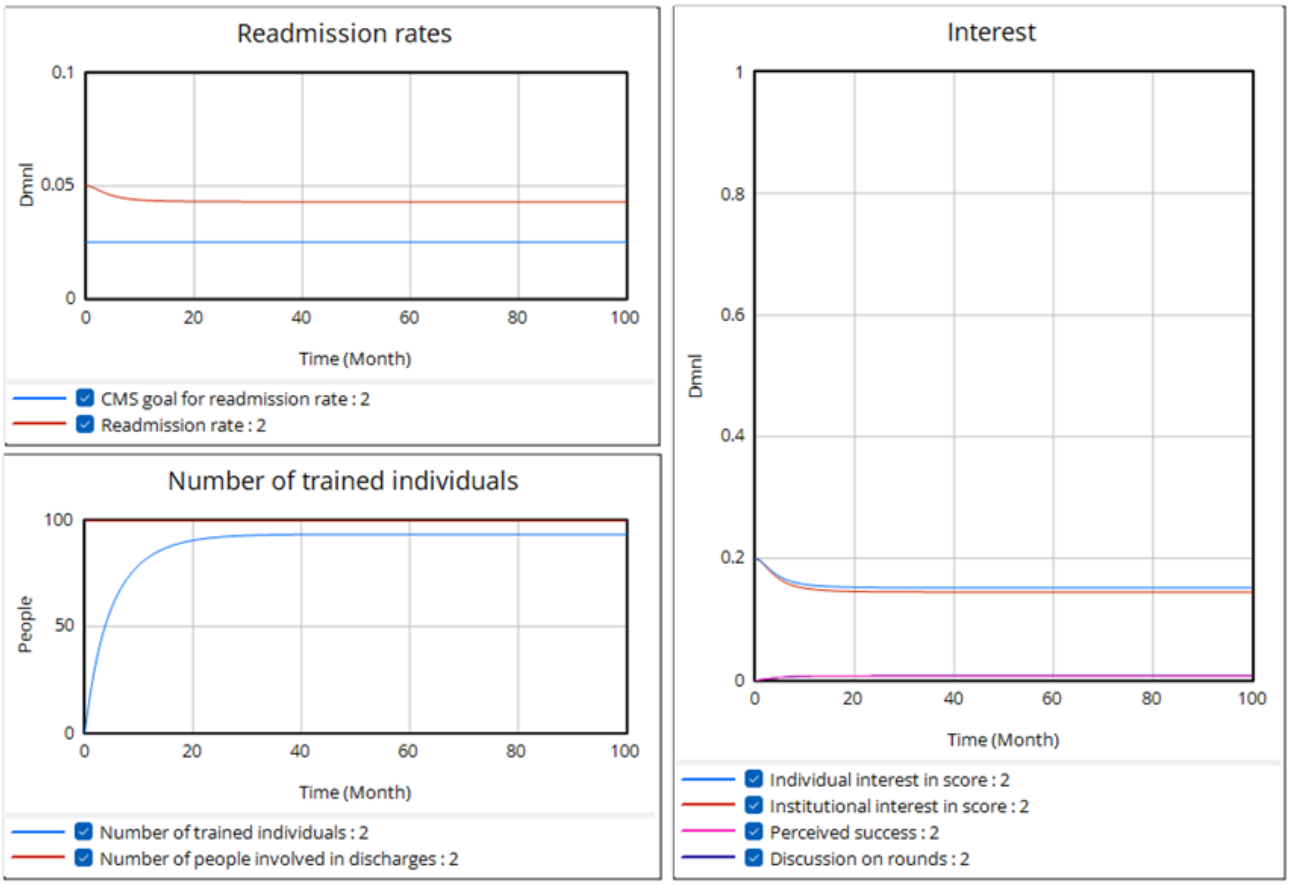


Additionally, institutional and individual interest in the score decreases as the readmission rate approaches the CMS goal. While we see a highly successful introduction of the tools with widespread and effective use, external pressure can only drive the effort so far. As the institution gets better, we can expect balancing loops to stop short of complete success because the organization relies on some continued external pressure (provided by high readmission rates) to sustain interest.

As is likely apparent from the discussion above, many factors could limit the effectiveness of these 2 balancing loops. A second run (Figure 6) shows one of these factors affecting both loops. Imagine a scenario where, as hospital staff discussed, the institution faces substantially higher conflicting priorities such as a pandemic or a stronger focus on reducing length of stay. These alternative priorities, which may have a more proximal impact than hospital readmissions, could reasonably divert attention and resources from training on the readmission risk score, thereby increasing the gap between the number of trained individuals and the number of people involved in discharge. While the readmission rate decreases as a result of the initial push for training and education, institutional interest will drop sooner with conflicting priorities, leading to reduced training and diminished individual interest in the score. Although still successful, the program results in fewer people being trained and a smaller reduction in readmissions. The shortfall is even more significant because the organization relies on continued pressure (provided by even higher readmission rates) to maintain interest.

XSL•FO
RenderX

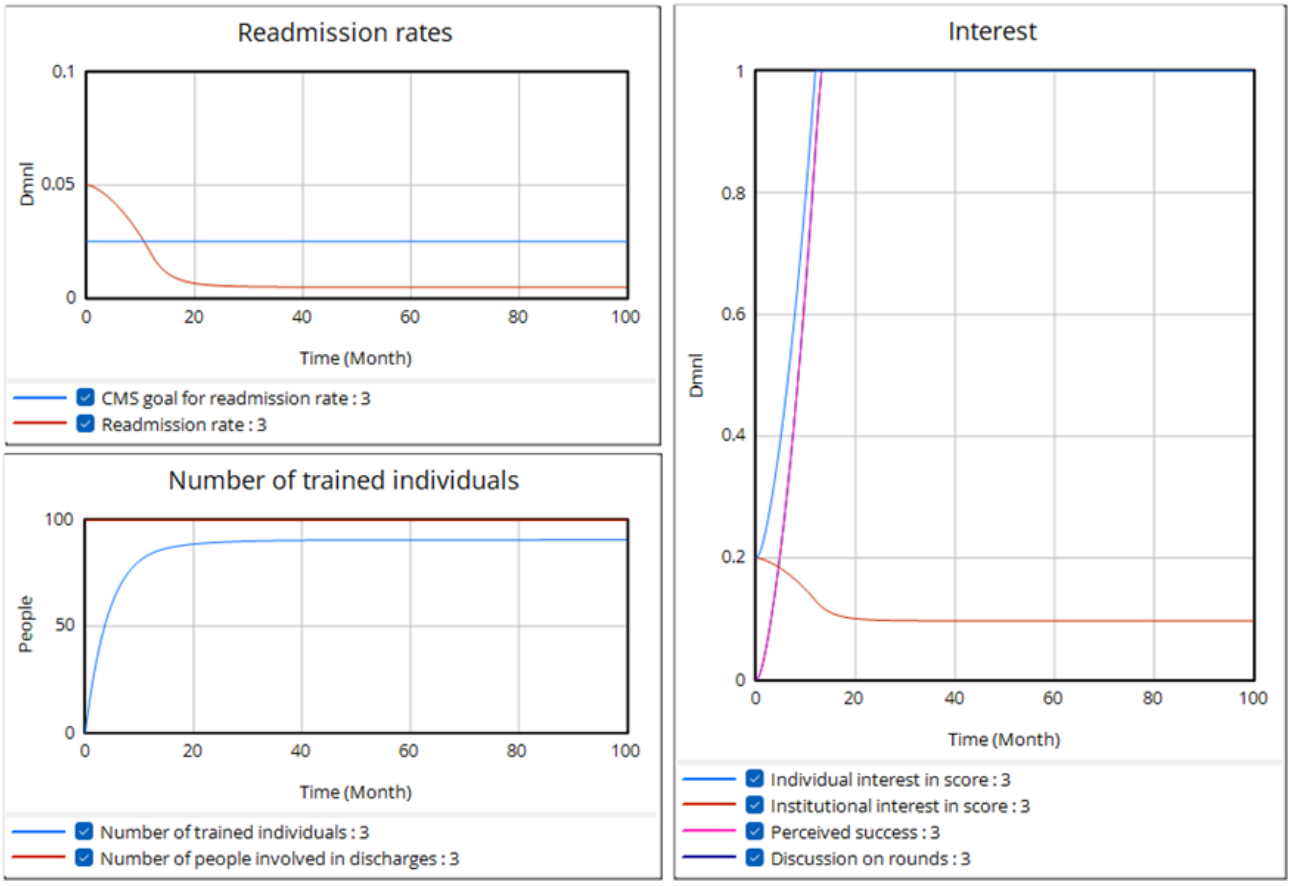
Figure 6. Simulation run with greater attention to conflicting priorities weakening the balancing loops. In this run, the Attention to Conflicting Priorities has been raised from 1 to 5 with no other changes to the model equations or parameters from the previous run. CMS: Centers for Medicare & Medicaid Services.



Having set the parameters (notably raising attention to conflicting priorities) so that the balancing loops alone lead to limited success, we then explore conditions under which the reinforcing loops could take over and drive more significant use of the tool and improvement in the readmission rate (Figure 7). The main change in this run is to reduce the Required Evidence, which refers to how many discharged patients each month must avoid readmission for individuals and groups involved in discharges to perceive tool success. We were mainly trying to capture how evident it is that unplanned readmissions were avoided, as we heard in the group model building (GMB) sessions. In this case, the readmission rate falls below the CMS goal, causing a drop in institutional interest. Despite the loss of external pressure from the CMS, the program continues to be driven by the excitement generated internally from perceived tool success. Institutional interest is even lower than before, and thus so is training, though it is buoyed by internal pressure from use on rounds. Still, individual interest is far higher, which offsets the drop in training.

XSL•FO
RenderX

Figure 7. Run with reduced required evidence strengthening the reinforcing loops. In this run, the Required Evidence has been reduced from 100 persons/month to 3.5 persons/month with no other changes to the model equations or parameters from the previous run. CMS: Centers for Medicare & Medicaid Services.



Discussion

Principal Findings

The causal loop diagram represents hypotheses that emerged from workshops regarding the adoption of a readmission risk prediction tool. The analysis of these feedback loops provided insights into the organizational dynamics influencing the success and sustainability of the initiative to reduce readmission rates. The findings inspired the development of a mathematical model to explore the logic of these loops in greater depth.

The balancing loops identified during the workshops highlighted challenges in maintaining the momentum of training and the need for ongoing reinforcement, both "facilitating conditions," to achieve the desired reduction in readmission rates [6]. Eagerness to reduce readmission rates to meet an external goal can only take adoption so far. Although feedback loops tied to external goals are intended to close the gap between current readmission rates and benchmark targets, the narrowing of that gap can paradoxically reduce effort. This effort refers to the ongoing commitment of time, training, and behavioral adaptation required for successful tool integration into clinical workflows. Depending on how quickly the effort to close the gap diminishes, as it might do rapidly in the face of strong external pressure for attention elsewhere, effort may fall too low before the gap is closed, and the tool may be used far less than its potential.

For sustained success, we believe the implementation of the readmission risk score requires additional input. The workshops provided indications that these additional inputs, in the form of reinforcing loops, would likely come from individuals across several disciplines perceiving internal benefits from the tool. Similarly, Venkatesh et al [6] suggested that while institutional influence, such as the perception that management wants staff to use the tool, tends to motivate new technology use in the early stages of implementation, over time, employees will rely less on these external drivers as they gain experience and integrate it with their workflow. If these internal reinforcing loops are strong, the process is likely to sustain itself without additional training or input from the tool developer. However, there is a risk of relying too heavily on the education of clinical team members and getting them to incorporate it into their own workflow. If the reinforcing loops are not strong, there are implementation strategies that could provide additional input. For example, a monitoring and evaluation system could function as an integral control mechanism for the Institutional Attention balancing loop that maintains tool use even when performance improves by reminding people of the improvements that could be lost without continued efforts [24,25]. Interventions might include a champion who could respond to feedback by adjusting support, such as by reinforcing the Training loop, to reduce long-term drift from desired tool use or an electronic health record–integrated dashboard that provides immediate feedback on key metrics to address the need for evidence about the CDS within the Individual Motivation and Team Use loops [26,27]. The key lesson then is that widespread and sustained adoption is far more likely if those who use the tool are able to see a

XSL•FO
RenderX

positive impact in daily practice on patient care decisions and outcomes.

By using a group model-building approach, we uncovered how these constructs interact over time through feedback loops at our institution. Our simulation model showed that while external pressure can drive initial adoption, it often leads to balancing loops that limit long-term success. Competing priorities that arise only magnify the effect of these balancing loops by directing attention away from initiatives to promote use. In contrast, reinforcing loops—driven by perceived internal benefits and team-level learning—can sustain tool use even after institutional interest declines. However, these reinforcing loops are fragile and require conditions such as visible tool impact, low required evidence thresholds, and integration into daily workflows to take hold. This dynamic systems perspective highlights the importance of designing implementation strategies that not only address user perceptions but also anticipate and support the evolving organizational and social conditions in which CDS tools are used. Modeling system behavior with different types of CDS tools and across diverse institutional contexts will help build a more generalizable theory of how system structure shapes adoption. For example, tools that provide more immediate feedback, such as real-time alerts for clinical deterioration, may be less vulnerable to competing priorities than tools with delayed feedback. To advance this work, it will also be important to build protocols to measure the effectiveness of the GMB and derived solutions: this could include measuring changes in staff understanding of the problem as a result of participating in the GMB and changes in implementation outcomes as a result of solutions implemented based on GMB insights. While this model explores the dynamics of implementing one type of CDS in one institutional setting, it provides a foundation for future research to extend toward a broader systems-aligned precision medicine framework.

Conclusions

We have developed a system dynamics model to approximate the driving and impeding forces related to the adoption of a readmission risk tool at a quaternary academic hospital through group model-building approaches. While we examined one specific outcome of CDS tool use, where there remains uncertainty about the widespread adoption of the tool into clinical workflow, the workshops and modeling point to the potential for a range of different outcomes in different settings. The decline in adoption of this specific score likely relates to the rare nature of the event, as there is limited real-world experience of the impact of the risk model on day-to-day experiences, combined with a decrease in active training and promotion due to competing priorities with the COVID-19 pandemic and an additional focus on length of stay. Understanding the implementation of these risk algorithms in a dynamic system will help developers iterate and improve implementation moving forward. Further research is planned to test the model predictions against a wider range of cases and settings, calibrate the model to test for external validity of the arguments, and refine the lessons.

Acknowledgments

Microsoft's Copilot (GPT-4; Microsoft Corp.) was used to summarize meeting transcripts related to the internal validity test with simulation modeling. All interpretation and analyses were conducted by the authors.

Funding

This work was supported by Bass Connections through the Health Policy and Innovation Theme at Duke University.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Equations used in the simulation model.
[DOCX File, 17 KB - humanfactors_v13i1e87522_app1.docx]

References

1. Chen Z, Liang N, Zhang H, et al. Harnessing the power of clinical decision support systems: challenges and opportunities. Open Heart 2023 Nov 28;10(2):e002432. [doi: 10.1136/openhrt-2023-002432] [Medline: 38016787]
2. Moja L, Polo Friz H, Capobussi M, et al. Implementing an evidence-based computerized decision support system to improve patient care in a general hospital: the CODES study protocol for a randomized controlled trial. Implement Sci 2016 Jul 7;11(1):89. [doi: 10.1186/s13012-016-0455-x] [Medline: 27389248]
3. Arruda-Olson AM, Afzal N, Priya Mallipeddi V, et al. Leveraging the electronic health record to create an automated real-time prognostic tool for peripheral arterial disease. J Am Heart Assoc 2018 Dec 4;7(23):e009680. [doi: 10.1161/JAHA.118.009680] [Medline: 30571601]
4. Kwan JL, Lo L, Ferguson J, et al. Computerised clinical decision support systems and absolute improvements in care: meta-analysis of controlled clinical trials. BMJ 2020 Sep 17;370:m3216. [doi: 10.1136/bmj.m3216] [Medline: 32943437]

XSL•FO
RenderX

5. Grechuta K, Shokouh P, Alhussein A, et al. Benefits of clinical decision support systems for the management of noncommunicable chronic diseases: targeted literature review. Interact J Med Res 2024 Nov 27;13:e58036. [doi: 10.2196/58036] [Medline: 39602213]
6. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003 Sep 1;27(3):425-478. [doi: 10.2307/30036540]
7. Liu S, Reese TJ, Kawamoto K, Del Fiol G, Weir C. A theory-based meta-regression of factors influencing clinical decision support adoption and implementation. J Am Med Inform Assoc 2021 Oct 12;28(11):2514-2522. [doi: 10.1093/jamia/ocab160] [Medline: 34387686]
8. Dingel J, Kleine AK, Cecil J, Sigl AL, Lermer E, Gaube S. Predictors of health care practitioners' intention to use AI-enabled clinical decision support systems: meta-analysis based on the Unified Theory of Acceptance and Use of Technology. J Med Internet Res 2024 Aug 5;26:e57224. [doi: 10.2196/57224] [Medline: 39102675]
9. Wang J, Zhou Y, Tan K, Yu Z, Li Y. Acceptance of artificial intelligence clinical assistant decision support system to prevent and control venous thromboembolism among healthcare workers: an extend Unified Theory of Acceptance and Use of Technology model. Front Med 2025;12:1475577. [doi: 10.3389/fmed.2025.1475577]
10. Wohlgemut JM, Pisirir E, Stoner RS, et al. A scoping review, novel taxonomy and catalogue of implementation frameworks for clinical decision support systems. BMC Med Inform Decis Mak 2024 Nov 1;24(1):323. [doi: 10.1186/s12911-024-02739-1] [Medline: 39487462]
11. Miller A, Moon B, Anders S, Walden R, Brown S, Montella D. Integrating computerized clinical decision support systems into clinical work: a meta-synthesis of qualitative research. Int J Med Inform 2015 Dec;84(12):1009-1018. [doi: 10.1016/j.ijmedinf.2015.09.005] [Medline: 26391601]
12. Ancker JS, Edwards A, Nosal S, et al. Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. BMC Med Inform Decis Mak 2017 Apr 10;17(1):36. [doi: 10.1186/s12911-017-0430-8] [Medline: 28395667]
13. Backman R, Bayliss S, Moore D, Litchfield I. Clinical reminder alert fatigue in healthcare: a systematic literature review protocol using qualitative evidence. Syst Rev 2017 Dec 13;6(1):255. [doi: 10.1186/s13643-017-0627-z] [Medline: 29237488]
14. Venkatesh V. Adoption and use of AI tools: a research agenda grounded in UTAUT. Ann Oper Res 2022 Jan;308(1-2):641-652. [doi: 10.1007/s10479-020-03918-9]
15. Kahkoska AR, Freeman NLB, Hassmiller Lich K. Systems-aligned precision medicine-building an evidence base for individuals within complex systems. JAMA Health Forum 2022 Jul 1;3(7):e222334. [doi: 10.1001/jamahealthforum.2022.2334] [Medline: 36219007]
16. Gallagher D, Zhao C, Brucker A, et al. Implementation and continuous monitoring of an electronic health record embedded readmissions clinical decision support tool. J Pers Med 2020 Aug 26;10(3):103. [doi: 10.3390/jpm10030103] [Medline: 32858890]
17. Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care: advantages and limitations. Arch Intern Med 2000 Apr 24;160(8):1074-1081. [doi: 10.1001/archinte.160.8.1074] [Medline: 10789599]
18. Hospital Readmissions Reduction Program (HRRP). CMS. URL: https://www.cms.gov/medicare/payment/prospective-payment-systems/acute-inpatient-pps/hospital-readmissions-reduction-program-hrrp [accessed 2025-03-11]
19. Artetxe A, Beristain A, Graña M. Predictive models for hospital readmission risk: a systematic review of methods. Comput Methods Programs Biomed 2018 Oct;164:49-64. [doi: 10.1016/j.cmpb.2018.06.006] [Medline: 30195431]
20. Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. JAMA 2011 Oct 19;306(15):1688-1698. [doi: 10.1001/jama.2011.1515] [Medline: 22009101]
21. Hovmand P, Rouwette E, Andersen DF, et al. Scriptapedia: a handbook of scripts for developing structured group model building sessions. Presented at: 2011 International System Dynamics Conference; Jul 24-28, 2011 URL: https://proceedings.systemdynamics.org/2011/proceed/papers/P1404.pdf [accessed 2026-03-23]
22. Hodgson AM. Hexagons for systems thinking. Eur J Oper Res 1992 May;59(1):220-230. [doi: 10.1016/0377-2217(92)90019-6]
23. Farr WW, Allen SD, Tomoaia‐Cotisel A, Hovmand PS. Documenting the modeling process with a standardized data structure described and implemented in DynamicVu. Syst Dyn Rev 2022 Jul;38(3):264-291. [doi: 10.1002/sdr.1716]
24. Chen R. A comprehensive analysis of PID control applications in automation systems: current trends and future directions. High Sci Eng Technol 2024;97:126-132. [doi: 10.54097/6q4xxg69]
25. Proportional-integral-derivative control. ScienceDirect Topics. URL: https://www.sciencedirect.com/topics/engineering/proportional-integral-derivative-control [accessed 2026-02-11]
26. Morena AL, Gaias LM, Larkin C. Understanding the role of clinical champions and their impact on clinician behavior change: the need for causal pathway mechanisms. Front Health Serv 2022;2:896885. [doi: 10.3389/frhs.2022.896885]
27. Randell R, Alvarado N, McVey L, et al. Requirements for a quality dashboard: lessons from national clinical audits. AMIA Annu Symp Proc 2020;2019:735-744. [Medline: 32308869]

XSL•FO
RenderX

Abbreviations

CDS: clinical decision support
CMS: Centers for Medicare & Medicaid Services
DUH: Duke University Hospital
GMB: group model building

Edited by A Kushniruk; submitted 11.Nov.2025; peer-reviewed by AJ Umeano, K Khashu; revised version received 21.Feb.2026; accepted 21.Feb.2026; published 29.Apr.2026.

Please cite as:
Sperber NR, Haas SE, Gao J, Hamelsky S, Kiki-Teboum T, Malick A, Pulugurta R, Rodriguez J, Shafique H, Singh E, Vasudevan K, Rockart S, Gallagher D, Johnson A
Learning From the Adoption of a Readmission Clinical Decision Support Tool: Group Model Building Approach
JMIR Hum Factors 2026;13:e87522
URL: https://humanfactors.jmir.org/2026/1/e87522
doi:10.2196/87522

© Nina Rachel Sperber, Sarah Elizabeth Haas, Jiaxin Gao, Samantha Hamelsky, Theresa Kiki-Teboum, Afraaz Malick, Rishab Pulugurta, Jacqueline Rodriguez, Hana Shafique, Eden Singh, Kriti Vasudevan, Scott Rockart, David Gallagher, Adam Johnson. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 29.Apr.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX